

Trabajo Fin de Grado  
Grado en Ingeniería de las Tecnologías de  
Telecomunicación

Big data y Analítica de Datos. Aplicaciones en  
Movilidad, Logística y Transportes.

Autor: Manuel Jerez González

Tutor: Fernando Guerrero López

Cotutor: Jesús Racero Moreno

**Dpto. Organización Industrial y Gestión de Empresas I**  
**Escuela Técnica Superior de Ingeniería**  
**Universidad de Sevilla**

Sevilla, 2022





Proyecto Fin de Carrera  
Grado en Ingeniería de las Tecnologías de Telecomunicación

# **Big data y Analítica de Datos. Aplicaciones en Movilidad, Logística y Transportes.**

Autor:

Manuel Jerez González

Tutor:

Fernando Guerrero López

Profesor titular

Dpto. Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2023



Proyecto Fin de Carrera: Big data y Analítica de Datos. Aplicaciones en Movilidad, Logística y Transportes.

Autor: Manuel Jerez González

Tutor: Fernando Guerrero López

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2023

El Secretario del Tribunal



*A mi familia*

*A mis maestros*





# Agradecimientos

---

Resulta muy difícil condensar en pocas palabras todo lo vivido durante los casi 5 años que han pasado entre que entré en la ETSI por primera vez en mi vida y ahora que presento este Trabajo de Fin de Grado, donde ha pasado una pandemia y un año viviendo en el que considero mi segunda casa, Italia; pero afortunadamente la mayor parte de la gente con la que cuento hoy en mi vida también estaba en ella o apareció durante los primeros meses de la carrera.

En primer lugar, agradecer a mis padres y a mi hermana por la confianza y el apoyo que me brindan día a día, por apostar tanto por mí y por mi educación; y al resto de mi familia por el orgullo con el que siempre hablan de mí y por el cariño que me muestran cada vez que estamos juntos; especialmente a mis abuelos, los que están y los que ya no.

A mis amigos del colegio, con los que tantas cosas buenas he vivido y que me han visto pasar por tantísimas etapas de mi vida, buenas y malas, y siempre han estado ahí; y a los que considero mi otra familia después de tantos años de relación.

A todas las personas que la ETSI ha puesto en mi vida: los *presos*, con los que hemos aprendido a base de mucho trabajo y muchas horas en Discord lo que era esto de estudiar *teleco*, y a los de la *Maldición Virtual*, que han aparecido este año en mi vida y con los que tengo la “suerte” de compartir casi todas las horas del día en el máster; con ello es todo mucho más llevadero.

Por último, a mi tutor; Fernando Guerrero, por permitirme hacer este TFG, por permitirme el ritmo de trabajo que necesitase y libertad absoluta para trabajar en este campo en el que tanto he aprendido, a Jesús Racero por las indicaciones y ayudas prestadas durante la etapa final del trabajo; y a Rubén García, Director Gerente de TUSSAM, por su disponibilidad y amabilidad en todo momento y por facilitarme los datos que he necesitado de cara a realizar este trabajo.

*Manuel Jerez González*

*Sevilla, 2023*



# Resumen

---

El Big Data se ha consolidado como uno de los paradigmas tecnológicos más importantes de nuestro tiempo, y gracias a la consolidación de otras herramientas desarrolladas a la par como la Inteligencia Artificial o el despliegue de 5G pretende revolucionar cómo vivimos. Por otro lado, el concepto de Smart City se apoya en todas estas herramientas para mejorar la vida de los habitantes de las ciudades, que cada vez son y serán más durante este siglo.

En este contexto, el uso de estas tecnologías despierta un gran interés en el campo del transporte público. El análisis de datos de transporte, tanto estáticos como en tiempo real puede mejorar la planificación y eficiencia del sistema de transporte público, reducir los tiempos de espera, mejorar la seguridad y reducir el impacto ambiental, ofreciendo alternativas más sostenibles y deseables que el coche.

Bajo estas premisas, en este documento se pretende realizar un estudio de las tecnologías de recolección y procesamiento de datos en Sevilla en contraposición con ciudades que destacan en este ámbito como Londres, así como un análisis del servicio prestado por la empresa pública de autobuses, TUSAM.



# Abstract

---

Big Data has established itself as one of the most important technological paradigms of our time, and thanks to the consolidation of other tools developed at the same time, such as Artificial Intelligence or the deployment of 5G, it aims to revolutionize the way we live. On the other hand, the Smart City concept relies on all these tools to improve the lives of the inhabitants of the cities, which are and will be more and more during this century.

In this context, the use of these technologies arouses great interest in the field of public transport. The analysis of transportation data, both static and in real time, can improve the planning and efficiency of the public transportation system, reduce waiting times, improve safety, and reduce environmental impact, offering more sustainable and desirable alternatives to the car.

Under these premises, this document aims to carry out a study of data collection and processing technologies in Seville as opposed to cities that stand out in this field such as London, as well as an analysis of the service provided by the public bus company, TUSAM.



# Índice

Agradecimientos .....	ix
Resumen .....	xi
Abstract .....	xiii
Índice .....	xv
Índice de Tablas .....	xviii
Índice de Figuras .....	xx
Índice de Códigos .....	xxiii
Notación .....	xxv
<b>1 Introducción .....</b>	<b>1</b>
1.1. <i>Objetivos</i> .....	2
1.2. <i>Estado del arte</i> .....	2
<b>2 Big Data .....</b>	<b>11</b>
2.1. <i>Características del Big Data</i> .....	11
2.2. <i>Fases en el uso de los datos</i> .....	12
2.2.1. Recogida de la Información .....	13
2.2.2. Almacenamiento de la Información.....	13
2.2.3. Procesamiento .....	14
2.2.4. Análisis.....	16
2.2.4.1. Análisis Descriptivo .....	16
2.2.4.2. Análisis Diagnóstico .....	16
2.2.4.3. Análisis Predictivo .....	17
2.2.4.4. Análisis Prescriptivo .....	17
2.3. <i>Tecnologías para el trabajo con datos</i> .....	17
2.3.1. Fuentes y formatos de los datos .....	17
2.3.1.1. Datos GTFS .....	18
2.3.2. Procesado de los datos.....	19
2.3.3. Visualización de los datos.....	20
<b>3 Smart Cities.....</b>	<b>22</b>
3.1. <i>El papel del Big Data en las Smart Cities</i> .....	24
3.2. <i>Smart Cities: Panorama actual</i> .....	24
3.3. <i>Mobility as a Service</i> .....	26
3.1.1. Aplicaciones MaaS.....	27
<b>4 Casos de uso del Big Data en Transportes.....</b>	<b>29</b>
4.1. <i>El caso de Londres y Transport for London</i> .....	29
4.1.1. Recolección de datos.....	31
4.1.1.1. La Oyster Card .....	31
4.1.1.2. El caso del metro y el rastreo WiFi .....	32

4.1.1.3	El autobús y sus problemas asociados.....	34
4.1.2	TfL y Cloud Computing.....	44
4.1.3	Usos de los datos recopilados.....	45
4.1.3.1	Para los usuarios.....	46
4.1.3.2	Para la empresa.....	46
4.1.4	Beneficios de la estrategia digital de Transport for London.....	47
4.2.	<i>El caso de Sevilla</i> .....	49
4.2.1	El servicio de autobús urbano, TUSSAM.....	50
4.2.1.1	Sistemas de Ayuda a la Explotación (SAE) y prioridad semafórica.....	50
4.2.1.2	Sistema de información audiovisual.....	51
4.2.1.3	Planificación de viajes y comunicación con los usuarios.....	51
4.2.1.4	Otros sistemas implementados.....	52
4.2.2	Metro de Sevilla.....	53
4.2.2.1	Pago con tarjeta de crédito.....	53
4.2.2.2	Control automático de aforo.....	54
4.2.2.3	Control en tiempo real de la calidad del aire.....	54
4.2.2.4	Planificación de viajes y comunicación con los usuarios.....	55
4.2.3	Otros medios de transporte.....	56
4.2.3.1	Autobuses metropolitanos.....	56
4.2.3.2	Cercanías RENFE.....	56
4.2.3.3	Bicicletas municipales SEVICI.....	57
4.2.4	Ecosistema de datos abiertos.....	57
<b>5</b>	<b>Análisis de la operativa de TUSSAM a través del big data.....</b>	<b>59</b>
5.1.	<i>Fuentes de datos</i> .....	60
5.1.1	API de TUSSAM.....	60
5.1.2	Datos de la plataforma ArcGIS.....	61
5.1.3	Datos GTFS.....	62
5.1.3.1	Valores de mayor relevancia para nuestro trabajo.....	62
5.2.	<i>Preparación de los datos</i> .....	63
5.2.1.	Inferencia de horarios.....	63
5.2.2.	Generación de <i>shapes.txt</i> .....	65
5.2.3.	Completado de <i>trips.txt</i> y validación del <i>feed</i> .....	67
5.3.	<i>Cálculo de estadísticas de paradas</i> .....	69
5.3.1	Combinación de horarios y líneas asociadas.....	69
5.3.2	División por tipo de servicio y ordenación por tiempo.....	69
5.3.3	Generación de estadísticas.....	70
5.3.4	Representación de los resultados.....	72
5.4.	<i>Cálculo de estadísticas de líneas</i> .....	73
5.4.1	Estadísticas de viajes.....	74
5.4.2	Estadísticas de líneas.....	74
5.4.2.1	Cálculo del servicio en horas punta.....	75
5.4.3	Representación de los resultados.....	76
5.5.	<i>Resultados</i> .....	77
5.5.1	Presentación de resultados.....	77
5.5.2	Análisis de los resultados.....	77
5.5.2.1	Propuestas de mejora para la red de transportes de Sevilla.....	84
<b>6</b>	<b>Conclusiones y líneas de mejora.....</b>	<b>86</b>
6.1.	<i>Conclusiones</i> .....	86
6.2.	<i>Líneas de mejora</i> .....	87
<b>Anexo A</b> .....		<b>88</b>
<b>Anexo B</b> .....		<b>94</b>
<b>Referencias</b> .....		<b>97</b>





# ÍNDICE DE TABLAS

---

Tabla 1. Ejemplo de entradas en la base de datos generadas en el piloto de rastreo WiFi en el metro. [67]	33
Tabla 2. Endpoints de la API de TUSSAM.	61
Tabla 3. Cambio en el formato de horarios de madrugada para trabajar en Python.	63
Tabla 4. Muestra del fichero <i>stopTimesComp.txt</i> .	69
Tabla 5. Muestra de <i>avgFreqComplete.csv</i> .	71
Tabla 6. Muestra de <i>frecuencias.csv</i> .	72
Tabla 7. Muestra de la tabla resultado de ejecutar <i>statsViajes.ipynb</i> .	74
Tabla 8. Muestra de una de las tablas resultado de ejecutar <i>statsLineas.ipynb</i> .	75



# ÍNDICE DE FIGURAS

---

Figura 1. Estructura del Cloud Computing y del Edge Computing.[23]	15
Figura 2. Tipos, complejidad, y valor de los análisis.[25]	16
Figura 3. Dimensiones de la smart city. [41]	23
Figura 4. Puntuación de Sevilla en las diferentes dimensiones analizadas en el IESE Cities in Motion Index 2020. [40]	25
Figura 5. Elementos de la movilidad como servicio.[50]	26
Figura 6. Logos de TfL (primero) y de los diferentes servicios que opera. [61]	30
Figura 7. Procedimiento para almacenar la información en el programa piloto.	33
Figura 8. Interfaz de iBus para los conductores de autobuses, con diferentes avisos para los pasajeros.	35
Figura 9. Ciudades donde se han realizado investigaciones sobre la inferencia de origen y destino entre 1994 y 2021. [79]	37
Figura 10. Ejemplo de inferencia de destinos entre diferentes rutas. [76]	38
Figura 11. Algoritmo de inferencia de origen.[75]	40
Figura 12. Algoritmo de inferencia de origen. [75]	41
Figura 13. Ejemplo de uso de sensores infrarrojos activos y pasivos. [88] [89]	43
Figura 14. Estrategia de gestión de los activos de TfL. [98]	47
Figura 15. Gráfico de la distribución de beneficios estimados gracias a la estrategia digital de TfL.	48
Figura 16. Reparto modal de movilidad en Sevilla según la obligatoriedad del desplazamiento.[104]	49
Figura 17. Información en tiempo real del tiempo de paso de la línea 41 en la parada 38 el sábado 3 de septiembre de 2022. El tiempo mostrado en Google Maps (izquierda) es diferente del mostrado en Moovit (centro) y la web oficial de TUSSAM (derecha).	52
Figura 18. Información de horario previsto del tiempo de paso de la línea 41 en la parada 38 el sábado 3 de septiembre de 2022. El horario mostrado en Google Maps (izquierda) es igual al de Moovit (centro), pero diferente al de la web de TUSSAM (derecha).	52
Figura 19. Ejemplo de consulta de la media de las condiciones climáticas de todos los vehículos a través de la web. [131]	55
Figura 20. Ejemplos de conversación con bot de metro de Sevilla, solicitando un comando existente (izquierda) y realizando una pregunta genérica (derecha).	56
Figura 21. Evolución de usuarios de SEVICI. [104]	57
Figura 22. Flujo de trabajo desarrollado en la aplicación realizada.	60
Figura 23. Ejemplo de uso del analizador de paquetes para acceder a la URL de una petición.	60
Figura 24. Plataforma de peticiones ArcGIS REST Services.	66
Figura 25. Método para corregir ordenación de horarios.	70
Figura 26. Distribución de los viajes en Sevilla según hora de inicio.	75

Figura 27. Correlación entre la distancia al Prado de San Sebastián y el tiempo medio de paso en laborables.	78
Figura 28. Aumento de los tiempos de espera en fines de semana respecto del tiempo de paso en laborables.	79
Figura 29. Correlación entre el número de paradas y la velocidad de cada línea.	80
Figura 30. Correlación entre la distancia recorrida y la velocidad de cada línea.	81
Figura 31. Porcentaje de reducción de los tiempos medios de paso entre las 07:30 y 09:00 en días laborables.	81
Figura 32. Porcentaje de reducción de los tiempos medios de paso en horas punta.	82
Figura 33. Porcentaje de mejora de las frecuencias de paso en horas punta.	82
Figura 34. Porcentaje de rutas que mejoran la velocidad en horas punta.	83
Figura 35. Porcentaje de rutas que mejoran sus servicios en las distintas horas punta.	84
Figura 36. Porcentaje de rutas que mejoran o empeoran sus servicios en horas punta.	84



# ÍNDICE DE CÓDIGOS

---

Código 5.1. Notebook <i>inferStopTimes.ipynb</i> para inferencia de tiempos.	65
Código 5.2. Notebook <i>genShapes.ipynb</i> para la generación de <i>shapes.txt</i> .	67
Código 5.3. Notebook <i>addShapesToTrips.ipynb</i> para la asignación de <i>shape_id</i> a cada ruta.	68
Código 5.4. Notebook para la validación del feed.	68
Código 5.5. Notebook <i>displayStops.ipynb</i> para la generación y guardado de mapas de paradas.	73
Código 5.6. Notebook <i>displayLinesMap.ipynb</i> para la generación y guardado de mapas relativos a la velocidad de las líneas.	77
Código A. 1. Notebook <i>stopTimesForStats.ipynb</i> .	88
Código A. 2. Notebook <i>sortAndSplitByDay.ipynb</i> .	89
Código A. 3. Notebook <i>statsParadas.ipynb</i> .	93
Código A. 4. Notebook <i>combStops.ipynb</i> .	93
Código B. 1. Notebook <i>statsViajes.ipynb</i> .	94
Código B. 2. Notebook <i>statsLineas.ipynb</i> .	96
Código B. 3. Notebook <i>compareLabsPeak.ipynb</i> .	96





TUSSAM	Transportes Urbanos de Sevilla, Sociedad Anónima Municipal
SQL	Structured Query Language
XML	Extensible Markup Language
(GEO)JSON	(Geographic) JavaScript Object Notation
CSV	Comma-Separated Values
HDFS	Hadoop Distributed File System
ODS	Objetivos de Desarrollo Sostenible
TIC	Tecnologías de la Información y la Comunicación
Wi-Fi, WiFi	Wireless Fidelity
IoT	Internet of Things
IMD	International Institute for Management Development
IESE	Instituto de Estudios Superiores de la Empresa
SaaS	Software as a Service
MaaS	Mobility as a Service
EMT	Empresa Municipal de Transportes
VTC	Vehículo de Turismo con Conductor
TfL	Transport for London
JJOO	Juegos Olímpicos
API	Application Programming Interface
NFC	Near Field Communication
MAC	Media Access Control
GPS	Global Positioning System
CCTV	Closed-Circuit Television
CDN	Content Delivery Network
AWS	Amazon Web Services
CTAN	Consortio de Transportes de Andalucía
RENFE	Red Nacional de Ferrocarriles Españoles
SAE	Sistema de Ayuda a la Explotación
GTFS	General/Google Transit Feed Specification
URL	Uniform Resource Location
NaN	Not a Number
HTML	Hypertext Markup Language



# 1 INTRODUCCIÓN

---

*Cualquier cosa que se mide y observa, mejora*

*- Bob Parsons -*

En los últimos tiempos, la tecnología ha ido introduciéndose en nuestras vidas, pasando de formar parte de los escritorios de algunas oficinas a acompañarnos en todo lo que hacemos en nuestro día a día. Esta tendencia ha hecho que se simplifiquen muchas tareas cotidianas, pero también que la dependencia que tengamos de los diferentes dispositivos sea cada vez mayor. En paralelo, estos equipos y sus aplicaciones se han encargado de recopilar información sobre todos sus usuarios, generando una cantidad de datos nunca vista en la historia de la humanidad.

La generación de esta gran cantidad de información ha hecho que emerjan nuevos conceptos como el Big Data, y también nuevas profesiones relacionadas con el análisis y la ingeniería de datos, a través de las cuales las empresas quieren beneficiarse del uso de estos. Esto se debe a que los datos en sí mismos no aportan ningún valor, pero sí lo tienen, y mucho, los patrones, modelos y, en definitiva, los conocimientos que podemos extraer a partir del análisis de los mismos. Los datos, correctamente almacenados y utilizados tienen un gran valor en la sociedad de hoy día, que muchos han comparado con el petróleo o el oro en tiempos pasados. Hoy en día, las grandes corporaciones tecnológicas como Apple, Microsoft, Google, Meta o Amazon, que se lucran a costa de la explotación de nuestra información personal, se encuentran entre las empresas más grandes del mundo. [1]

La recopilación de los datos, realizada con el permiso, pero también con el desconocimiento de gran parte de la sociedad, ha hecho, gracias al consecuente tratamiento de estos, que la experiencia de usuario de la tecnología se adapte mucho mejor a las necesidades de cada individuo, con una gran personalización de la información que recibimos. Sin embargo, el uso de estos datos por parte de las grandes empresas también ha sido motivo de preocupación. A pesar de la controversia generada por los diferentes escándalos relacionados con el uso de los datos de los usuarios y la creciente preocupación social por la privacidad, la recopilación y el uso de datos por parte de las grandes empresas continúa.

La correcta explotación de estas cantidades tan grandes de datos también abre un horizonte de posibilidades que no podemos dejar escapar. La recolección, almacenamiento, procesamiento y tratamiento de estos datos nos está permitiendo transformar y optimizar prácticamente todos los procesos de nuestra vida; incluyendo las industrias, los negocios o las comunicaciones; hasta el punto de ser uno de los pilares de la *cuarta revolución industrial*. [2] Este fenómeno también puede afectar a la forma en que nos comunicamos con el entorno urbano en el que vivimos, y la forma en que las ciudades se adaptan a las necesidades de sus habitantes.

En nuestro día a día ya son muchos los ejemplos de aplicaciones que usamos y se sirven de los datos recopilados para predecir el tiempo que tardaremos hasta un destino, la combinación de transporte público más conveniente

en un determinado momento del día o cómo evitar atascos en un trayecto por carretera. A pesar de estos ejemplos, el Big Data puede traer muchas más mejoras que impacten directamente a la eficiencia de la red de transportes de las ciudades y, directamente, a sus ciudadanos.

## 1.1. Objetivos

- Analizar el concepto de Big Data, así como conceptos relacionados con los tipos y usos de los datos y los análisis que estas tecnologías permiten hacer.
- Analizar el concepto de Smart City, los objetivos que persigue, sus retos y la relación de este término con el Big Data.
- Analizar el concepto de la movilidad como servicio, así como su relación las *Smart Cities* y el estado en el que se encuentra este paradigma en la actualidad.
- Analizar cómo el Big Data puede transformar y ha transformado los sistemas de transporte de algunas ciudades; donde destaca el caso de Londres.
- Analizar el estado del uso de diferentes tecnologías de recolección y análisis de datos de transportes en la ciudad de Sevilla en contraposición con Londres; así como la capacidad de la ciudadanía de acceder a estos.
- Estudiar la operativa de la empresa de Transportes Urbanos de Sevilla, TUSAM, a través del uso de técnicas de ciencia y análisis de datos.

## 1.2. Estado del arte

En la actualidad, son ya muchas las ciudades que han decidido apoyarse en el uso de Big Data con el objetivo de optimizar sus sistemas de transporte urbano. La recolección de estos datos se ha apoyado fundamentalmente en el uso de sensores en los diferentes tipos de vehículos y estaciones, pero también en los dispositivos móviles de los usuarios de los transportes y la información extraída de diversas redes sociales. La mayor parte de estos datos se han utilizado con el objetivo de servir a los usuarios de información útil para sus trayectos y la planificación de los viajes; no obstante, el uso de la información junto con técnicas de predicción también permite a las empresas de transportes anticipar posibles fallos del servicio y minimizar el efecto de estos, así como controlar elementos relacionados con la seguridad. [3] En la actualidad, la mayor parte de los operadores proveen a sus ciudadanos de aplicaciones que permiten dar a conocer la ruta óptima para llegar de un punto a otro. Además de eso, en muchos casos la información para planificar la ruta se complementa de actualizaciones en tiempo real, permite ser adaptada a necesidades específicas del usuario y combina información proveniente de distintos medios de transporte. Proveer de este servicio requiere de la recolección, de gran cantidad de datos, así como de su procesamiento en tiempo real.

De todos los usos del Big Data en transporte público, probablemente el más interesante sea el relacionado con el estudio de los movimientos de los usuarios, que permite a los operadores planificar el servicio prestado basándose en previsiones de la demanda según los movimientos y tendencias observadas a través de los datos. Gracias a esto, los servicios ofrecidos se vuelven mucho más eficientes, permitiendo adaptar el transporte público a las necesidades reales de los usuarios. En este sentido, existen ejemplos concretos que nos permiten ejemplificar la revolución que supone el Big Data para las ciudades, y más concretamente, para el transporte. Durante el confinamiento ocasionado por la crisis del COVID-19, la ciudad de Nueva York se vio obligada a cerrar el metro por las noches; y para no dejar de dar servicio a aquellos usuarios que lo necesitaban, en cuestión de días fueron capaces de analizar la demanda y planificar nuevas rutas de autobuses que cubrieran el servicio que se iba a interrumpir. [4] Al igual que en Nueva York, otras ciudades como Seúl o han puesto en marcha programas parecidos para adaptar sus servicios de transporte a las necesidades basándose en el Big Data; lo cual se vuelve crítico en megaciudades como las mencionadas. Para realizar esta tarea, las fuentes de datos pueden ser muy diversas; pudiendo incluir encuestas, datos de tarjetas de crédito y tarjetas de transporte, movimientos de coches o taxis, llamadas de teléfono, datos del sistema de posicionamiento global (GPS), etc. [5]

Además de las acciones que los diferentes operadores públicos de transporte llevan a cabo, otra tendencia

relacionada con la aplicación del Big Data en transporte es la Movilidad como Servicio o MaaS; que consiste en la integración de información de diferentes operadores en una misma aplicación para proveer a los usuarios de estas un portal unificado para acceder a la información de transporte. En general, la Movilidad como Servicio es un campo que sigue aún en exploración y desarrollo, y fundamentalmente se lleva a cabo por empresas privadas aunque con la colaboración de los operadores públicos, que facilitan datos de transporte a estas. Los grados de integración son diversos y dependen de las aplicaciones y ubicaciones donde actúan; permitiendo desde tan sólo planificar un trayecto multimodal o consultar un precio hasta integrar los distintos transportes en una sola tarifa o pagar los servicios de transporte directamente a través de la aplicación aun siendo de un tercero. [6]

Un elemento diferenciador de muchos sistemas de transporte es, además, la publicación de todos los datos con los que cuentan al público, con el objetivo de favorecer la transparencia, pero también de permitir la innovación y el desarrollo de aplicaciones por parte de otros que complementen la oferta de información que dan los propios operadores. Dentro de esta categoría, destaca la ciudad de Londres, que además de implementar en su sistema de transporte una gran cantidad de tecnologías que permiten la recopilación de datos y el uso de esta información para mejorar su operación y la información provista a los ciudadanos; disponen de un portal público y una API donde todo el mundo puede entrar y realizar consultas para ver información que incluye no sólo datos del funcionamiento de los transportes públicos, sino también de accidentes de tráfico, rutas ciclistas o calidad del aire, entre otras cosas. [7]

En España, el Ministerio de Transportes, Movilidad y Agenda Urbana ha realizado varios estudios haciendo uso del Big Data relacionados con la movilidad en todo el país desde 2017; en sustitución de métodos como las encuestas, basándose en los datos obtenidos de los registros de los móviles de los ciudadanos. El primer estudio, realizado en el verano de ese año, se encargó de estudiar la movilidad interprovincial de nuestro país. [8] Este estudio se ha ido repitiendo y ampliando a estudiar los distintos movimientos de personas por la geografía española, realizándose de forma diaria desde el 1 de enero de 2022. No obstante, estos datos sólo evalúan los movimientos de las personas, y no el uso que se hace del transporte público, que es el tema que nos ocupa. A pesar de ello, el propio Ministerio dice que lo hace con el objetivo de ofrecer un mejor Sistema Nacional de Transportes, con el objetivo de hacerlo más sostenible. [9]

Además de estos estudios, las diferentes comunidades autónomas y localidades hacen estudios de diversos tipos relacionados con asuntos de interés para el transporte en las ciudades. En este sentido, además de la información relacionada con horarios en tiempo real de autobuses y metros, en nuestro país destacan los servicios que permiten a los ciudadanos conocer la ocupación de plazas de aparcamiento en tiempo real en sus ciudades, realizados en ciudades como Barcelona, Ronda, Cáceres o Pamplona. Muchos de estos conjuntos se encuentran agrupados en un portal de datos abiertos creado por el Ministerio de Asuntos Económicos y Transformación Digital, lo que permite a todo el mundo consultarlos y construir aplicaciones que hagan uso de ellos. También en este portal se pueden encontrar conjuntos de datos e informes elaborados tanto por autonomías como por empresas públicas. [10]

Por último, si miramos el estado de estas tecnologías en Sevilla, aún queda mucho trabajo por hacer en todos los campos. En 2016 se creó el Portal de Datos Abiertos de la ciudad, con una gran variedad de conjuntos de datos relacionados con todas las instituciones municipales. No obstante, la sección de datos relacionada con el transporte lleva sin ser actualizada desde ese mismo año, por lo que resultan inútiles en la actualidad. [11] Esto supone que el trabajo con datos relacionados con los transportes en nuestra ciudad sea mucho más complejo y dependa del uso de aplicaciones privadas relacionadas con la movilidad como servicio, como Moovit o el propio Google Maps, que sí ofrecen APIs para acceder a sus datos, aunque no de forma abierta ni gratuita. Por otro lado, la propia legislación en España no permite que los datos recopilados por entidades públicas sean empleados con fines comerciales. Por tanto, la posibilidad de recopilar datos a partir de los teléfonos móviles en los distintos vehículos o los pagos con tarjetas de transporte para mejorar la operativa de la empresa queda en un estado legal algo delicado. [12]



# 2 BIG DATA

---

*Crear oportunidades significa mirar donde otros no lo han hecho*  
- Mark Cuban -

**A** pesar de ser un término muy presente en nuestro día a día, el término Big Data no fue acuñado hasta el año 2005, momento en el que Roger Magoulas lo usó para referirse a un gran conjunto de datos difícil de administrar. Antes de este momento, sí existían ya ordenadores y máquinas destinadas al procesamiento de los datos; sin embargo, no fue hasta entrado el S. XXI que se empezó a definir el término. [13]

Si atendemos a la traducción literal del inglés, podemos ver que se refiere a “grandes datos”. Podríamos decir que el término Big Data se refiere al análisis de cantidades de datos lo suficientemente grandes como para no poderse utilizar las técnicas y enfoques tradicionales del análisis de datos. De este modo, son el conjunto de técnicas que nos permiten extraer valor de cantidades de datos grandes. A pesar de ello, el término definido se encuentra en constante cambio y no tiene muy delimitadas sus fronteras, y es que también podemos utilizar el Big Data para referirnos no sólo a las técnicas; sino también a los diferentes elementos que nos permiten la recolección, almacenamiento y procesamiento de la información de este tipo.

Diversas definiciones se han intentado dar para explicar el término, relacionadas con diversos aspectos relacionados a este. Además de las mencionadas, referidas a la tecnología, también se habla del Big Data como la capacidad de generar señales que nos permitan anticiparnos a eventos, como una “cultura” que afecta nuestros procesos diarios, como una revolución que permite extraer mucho más y mejor conocimiento de la información generada por nuestro entorno; y como una oportunidad de negocio para muchas empresas, que ven en el Big Data la posibilidad de aumentar sus beneficios, ya sea ofreciendo servicios relacionados con esto o utilizando los datos para mejorar sus procesos empresariales. [14]

En cualquier caso, las diferentes definiciones coinciden en definir el Big Data con un conjunto de características que, aunque también se encuentran en expansión, nos permiten conocer mejor el término.

## 2.1. Características del Big Data

Todas las definiciones del Big Data incluyen tres características fundamentales, conocidas como 3Vs: [15]

- **Volumen:** Como ya se ha especificado anteriormente, la característica fundamental del Big Data se refiere a la cantidad de información que se debe manejar, que hace imposible usar los métodos tradicionales relacionados con el análisis de datos. Además de ser el elemento diferenciador, también es el mayor reto, puesto que se debe hacer frente en las diferentes fases de procesamiento de los datos a una gran cantidad de información. Estos datos provienen de una gran cantidad de fuentes diferentes,

que pueden ser internas o externas a la aplicación que estemos utilizando; y es que en muchas ocasiones será necesario hacer uso de información externa para complementar la propia.

- **Velocidad:** En los últimos tiempos, nuestra percepción y exigencias respecto de la velocidad a la que necesitamos la información se ha vuelto mucho más exigente. Cuando hablamos de velocidad en el Big Data, nos referimos a la capacidad de los sistemas para transportar y procesar los datos a una gran velocidad, cercana al tiempo real. Para ello, es necesario contar con sistemas y algoritmos con una gran capacidad de computación; así como con tecnologías de telecomunicación lo suficientemente rápidas como para que las transmisiones de datos tarden el menor tiempo posible. Muchos dispositivos y aplicaciones requieren reaccionar a eventos casi en el mismo instante en el que estos ocurren. En este tipo de casos de uso, la velocidad de transmisión y procesamiento de los datos se vuelve una dimensión crítica.
- **Variedad:** Una peculiaridad del Big Data y que le separa del análisis de datos tradicional, además del tamaño, se refiere a los tipos de datos. En este tipo de tecnologías, la información se puede extraer de prácticamente cualquier fuente, lo que hace que sea necesario un trabajo de procesar y estructurar la información disponible para poder trabajar con ella. Esto hace que los enfoques de bases de datos relacionales, con tipos de datos fijos, no sean suficientes para trabajar con el Big Data. Además de los datos estructurados, nos encontramos con datos semiestructurados y no estructurados. Esta falta de homogeneidad supone un reto más para el Big Data.

Para muchos, sin embargo, estas tres características no son suficientes para definir el concepto de Big Data hoy en día. Es por eso que, en la mayor parte de los textos actuales se habla de “5 Vs”, añadiendo la veracidad y el valor a las ya mencionadas previamente. [16]

- **Veracidad:** Debido a la gran cantidad de datos recibidos y a la diversidad de fuentes de donde se reciben los datos, es posible que parte de ellos sea errónea. Por tanto, esta característica se refiere a que las diferentes técnicas de procesamiento del Big Data deben asegurarse de que la información introducida en el sistema es correcta. De ello depende el correcto funcionamiento de lo que hagamos posteriormente con los datos, y es que, si la información de la que disponemos para tomar decisiones es errónea o sólo parcialmente correcta, las actuaciones que hagamos se verán fuertemente comprometidas.
- **Valor:** El objetivo último en el uso del Big Data es analizar esta información con el objetivo de extraer de ella información útil para aplicarlo en el mundo real. Por tanto, esta característica es fundamental, y se refiere a la capacidad de sacar valor de aquellos datos ya procesados y almacenados. Evidentemente, esta característica depende fuertemente de las anteriores; ya que, si la información es errónea, no hay cantidad suficiente de la misma como para analizarla, o no es lo suficientemente variada como para representar nuestro alrededor, difícilmente podremos conseguir algo de ella. En definitiva, esta característica se refiere a la explotación de la información, con el objetivo de aumentar el valor de aquello para lo que estemos aplicando el Big Data.

Por último, cabe destacar que existen definiciones alternativas o complementarias a las anteriores, que incluyen otra serie de características que deben estar también asociadas al Big Data. La visualización, entendida como la capacidad de plasmar las grandes cantidades de datos manejados en gráficos y figuras legibles para el ser humano, que permitan un análisis intuitivo de estos; la variabilidad, entendida como la capacidad de los modelos de adaptarse a los cambios en la información o en la forma en la que esta se recibe; o la viabilidad, referida a la capacidad de las empresas o usuarios de manejar los grandes volúmenes de datos recibidos; son algunas características repetidas en diferentes modelos que contemplan más de “5 Vs” para referirse al Big Data.

## 2.2. Fases en el uso de los datos

El trabajo con Big Data requiere especificar diferentes fases según las operaciones que se deben realizar y las tecnologías implicadas, que van desde la identificación de fuentes y recopilación de información hasta la aplicación de las conclusiones extraídas a nuestra aplicación o negocio. En este caso, tampoco existe una definición exacta de las diferentes fases que debemos aplicar para trabajar correctamente con los datos, pero diferenciaremos las etapas más fundamentales del proceso.



### 2.2.1. Recogida de la Información

En primer lugar, es necesario que la aplicación que busca utilizar el Big Data para mejorar sus procesos identifique sus fuentes de información. Esto incluye cualquier lugar de donde se pueda extraer información que pueda resultar de utilidad en posteriores análisis. La detección de las fuentes adecuadas de información es necesaria en tanto que se busca evitar recibir información que no sea útil pero también que no sea errónea. Como ya hemos visto, una peculiaridad del Big Data es la gran cantidad de información que se recibe y la variedad de formatos de la misma. De este modo, la información se puede encontrar en bases de datos de la propia entidad o aplicación, se puede leer de sensores u otros elementos que reciban información del mundo real, se puede realizar una monitorización de diferentes procesos o realizar encuestas y formularios. Además, la información puede no ser generada directamente por la propia empresa, sino también apoyarse en datos obtenidos de redes sociales o de repositorios públicos de información. La variedad de fuentes de información hace que el posterior procesamiento resulte crítico, pero también hace que nos encontremos diferentes tipos de datos, según la facilidad con los que estos pueden ser utilizados: [17]

- **Datos estructurados:** Son aquellos datos que ya se habían utilizado en el análisis de datos tradicional. Están ya estructurados o formateados y por tanto se adaptan a un modelo, lo que hace sencillo el trabajo con ellos. Además, se pueden utilizar programas como Excel o SQL para visualizarlos fácilmente, ya sea en formato de tablas o gráficos. Ejemplos de este tipo de datos pueden ser hojas de cálculo con estadísticas, resultados de formularios web o, en general, bases de datos relacionales. Esta estructura tan definida de los datos hace que sean más sencillos de utilizar y gestionar, pero también que la información sea más interdependiente y que presente cierta rigidez y resistencia a la variabilidad de la información. Por tanto, existen fuentes de información que no pueden ser definidas y almacenadas en este tipo de estructuras.
- **Datos semiestructurados:** Son datos con un cierto nivel de estructura, pero que no cuentan con la regularidad suficiente como para ser considerados como el tipo anterior. Los formatos XML o JSON son el mejor ejemplo de este tipo de datos, donde existe un esquema para escribir la información, pero esta no tiene que adaptarse a ningún patrón, además de poder contener una jerarquía entre los diferentes contenidos. Este tipo de datos, así como los no estructurados, suponen un mayor reto para el análisis de datos; dado que no son tan previsibles como los anteriores y la organización de los mismos puede ser confusa.
- **Datos no estructurados:** Son datos que no tienen una estructura definida, ya sea porque acaban de ser recopilados y no han sido procesados para conseguir estructurarlos o porque no es posible hacerlo. En este caso se incluyen todas las fuentes de información que no tienen un formato específico que permite almacenar la información de forma estandarizada o jerarquizada, dado que no comparten patrones o modelos. Este tipo de datos incluye imágenes, audios y vídeos; correos electrónicos o incluso papeles escritos a mano. Los datos no estructurados son uno de los mayores retos en el campo del Big Data, dado que estos no pueden ser abordados con los enfoques tradicionales, y en muchas ocasiones resulta difícil extraer información valiosa de ellos. El entendimiento de estos datos es, sin embargo, crítico, dado que más del 80% de la información que se genera hoy día es de este tipo, pero sólo el 18% de las empresas son capaces de obtener información de valor de ella, lo que no supone sólo una pérdida de oportunidades, pero también una pérdida de recursos, invertidos en almacenar esta información, pero sin que sea utilizada. [18]

Una vez hecha la recopilación de toda la información, esta debe ser clasificada, categorizada y filtrada, con el objetivo de hacer más sencillo el trabajo posterior de uso de la misma. Podemos hablar de que existe una etapa de procesamiento de la información, orientada a preparar los datos para que su manejo sea lo más sencillo posible posteriormente.

### 2.2.2. Almacenamiento de la Información

El gran volumen de información que se maneja al trabajar con el Big Data, sumado a la variedad de formatos y tipos en los que los datos recopilados se presentan, hace que el almacenamiento de la información sea un asunto crítico y un reto tecnológico. Además de esa flexibilidad y la gran capacidad para almacenar esos datos, los sistemas relacionados con el guardado de datos deben garantizar otra serie de características.

Relacionados con la capacidad de almacenamiento, es necesario que el sistema optimice el espacio lo máximo

posible, así como que presente una gran escalabilidad ante un aumento o disminución en la necesidad de recursos. Además, el sistema debe ser flexible y compatible con los diferentes tipos de datos, permitiendo el almacenamiento y la gestión de diferentes archivos. Como ya hemos visto, la velocidad es fundamental en el uso del Big Data, así que el acceso debe ser sencillo y rápido. Por último, debe garantizarse que el almacenamiento es seguro frente amenazas o ataques externos, pero también robusto y disponible en todo momento para el acceso a la información.

Una opción para el diseño y la implementación de un sistema de almacenamiento eficiente es el almacenamiento por niveles. De esta forma, la información se estructura de forma jerárquica, de forma que aquellos datos de más valor y más accedidos se ubican en lugares que garanticen la seguridad, disponibilidad, y velocidad de acceso a estos; mientras que los datos que tienen menos valor se ubican en lugares con menos recursos dedicados. Esto, además de ser eficiente, permite ahorrar costes, al dar diferente tratamiento a los archivos según su importancia y necesidad. Además, los diferentes niveles o lugares donde se almacena la información pueden utilizar tecnologías físicas diferentes, de forma que cada una se adapte mejor a la necesidad de los recursos almacenados.

No obstante, no existe ninguna prescripción sobre dónde debería almacenarse la información, sino que cada empresa o servicio que haga uso del Big Data tiene ante él una gran cantidad de opciones disponibles. En la actualidad, las bases de datos relacionales son la opción más popular para el almacenamiento de la información, sin embargo, estas presentan limitaciones, dado que no dan soporte a los datos semiestructurados y no estructurados. Frente a este tipo de bases de datos, nos encontramos con sistemas basados en archivos; que nos permiten almacenar todo tipo de datos. Esto nos permite guardar los datos tal cual fueron obtenidos de la fuente, sin necesitar un procesamiento previo para adaptarlos a esquemas relacionales. [19]

Además de decidir el formato de contenido que se guardará en nuestro sistema, se debe decidir el “lugar” en el que se guardarán todos los datos. En la actualidad, las opciones de almacenamiento en la nube son las más populares, dado que nos permiten ahorrar costes y acceder cómodamente a la información desde cualquier parte; pero también, dependiendo de la solución contratada, nos ofrece una gran seguridad, disponibilidad y accesibilidad, así como gran facilidad para comenzar a trabajar. En contraposición a este tipo de infraestructura, existen otras opciones, como el Hadoop Distributed File System (HDFS), un sistema de archivos distribuido que permite almacenar cualquier tipo de archivo, que será duplicado y repartido en diferentes nodos para garantizar una gran tolerancia a los fallos. El HDFS no es la única opción para guardar los datos en un sitio “físico” sino que existen otras opciones. [20]

Para tomar la decisión respecto del tipo de contenedor para el almacenamiento de los datos, así como para decidir dónde almacenarlos, no se debe tener en cuenta sólo el coste, sino también el tipo de información a almacenar, así como el uso que se vaya a hacer de esta; como se verá en el siguiente apartado.

### 2.2.3. Procesamiento

Tras realizar el almacenamiento de la información “en bruto”, esta debe ser procesada. En esta fase, podemos diferenciar dos tipos de procesamiento. El primero tiene que ver con el procesamiento de los datos previamente mencionado, en la etapa de recopilación.

Una vez realizada la recolección de información debe realizarse un trabajo de procesamiento orientado a conocer si la información almacenada es correcta, mediante el contraste con otras fuentes o la búsqueda de anomalías. Dentro de esta etapa, nos encontramos también con un trabajo en los datos orientado a hacer la información más legible y sencilla de acceder; traduciendo los datos semiestructurados o no estructurados a formatos que sean más accesibles y cómodos de utilizar. Por último, la información puede ser clasificada y etiquetada, con el objetivo de que el acceso y el trabajo posterior con los datos sea más sencillo.

Después de completar este proceso de revisión de la información, podemos comenzar con el tratamiento de los datos con el objetivo de extraer información valiosa a partir de todo lo recopilado. En este punto, la frontera entre el procesamiento y el análisis de la información es difusa, en tanto que debemos trabajar con los datos usando algoritmos y modelos complejos, estadística y otras técnicas como la inteligencia artificial para extraer información “oculta” a partir de lo recopilado, y obtener resultados que nos permitan estudiar el pasado, presente y las tendencias futuras de nuestro negocio, el entorno, y la repercusión de nuestras posibles decisiones que se podrían tomar en base a las métricas conocidas. En este sentido, existen técnicas de computación que, aunque de propósito general, son tremendamente útiles y muy utilizadas para el trabajo con Big Data:

- **Cloud Computing:** Es uno de los elementos fundamentales que ha hecho que el Big Data esté tan presente en nuestros días. Es un “software como servicio” que permite utilizar, a través de un portal o consola, servicios de computación extremadamente potentes. Este servicio es muy útil, dado que antes esa capacidad de computación estaba limitada a pocas empresas, en ubicaciones concretas y a unos costes muy elevados. Gracias al *cloud computing*, cualquiera que lo desee puede acceder a máquinas remotas desde cualquier lugar, con grandes capacidades de procesamiento y a costos reducidos. Además de ello, los diferentes operadores de *cloud computing* ofrecen junto a este servicio otra serie de productos y tecnologías que pueden resultar útiles para el trabajo con Big Data, como máquinas virtuales, bases de datos o sistemas de archivos, todo también en la nube. El *cloud computing* nos permite usar con mayor facilidad la inteligencia artificial y el *machine learning*, debido a la alta cantidad de recursos que estas técnicas requieren para trabajar. De esta forma, podemos usar el *cloud computing* para procesar información recopilada y transformarla en estructuras más sencillas de utilizar o estandarizarla, pero también para analizar la información y realizar predicciones. Debido a la potencia de la que disponemos, además, estas operaciones pueden realizarse en tiempo real. [21]
- **Edge Computing:** Es una arquitectura que distribuye el cloud computing a más ubicaciones de la red, cercanas al dispositivo final, lo que permite reducir la latencia de acceso a los datos, haciendo más cercanas las operaciones al tiempo real. Estos dispositivos pueden realizar las mismas tareas que ya se han comentado en el cloud computing, aunque la capacidad de procesamiento de la que disponen es, en general, menor. No obstante, siguen siendo una gran solución para el trabajo con el Big Data, sirviendo como lugares de recopilación y agregación de datos, almacenamiento y procesamiento de estos. En aplicaciones que requieran muy poca latencia, puede ser una opción mejor que el uso de servicios cloud, dado que además de aumentar la velocidad, reduce la congestión en el interior de la red, invoca los dispositivos más potentes sólo cuando es necesario y descentraliza la toma de decisiones. [22]

### Cloud Computing vs Edge Computing

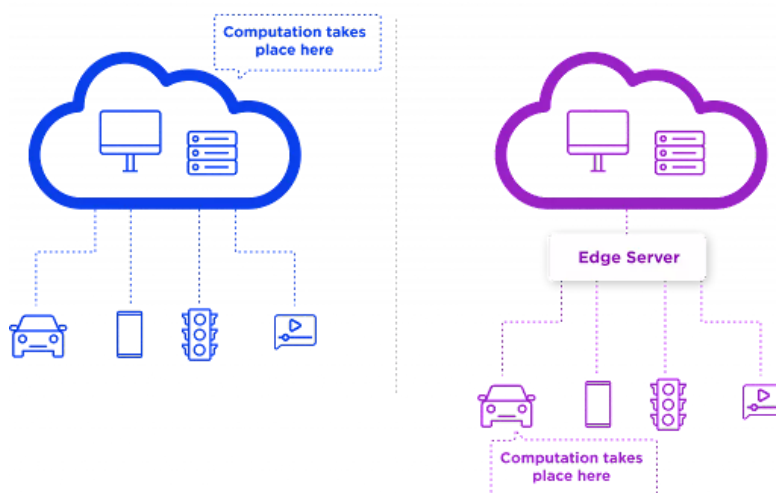


Figura 1. Estructura del Cloud Computing y del Edge Computing.[23]

- **Hadoop:** Es un *framework* de software libre ampliamente utilizado en el trabajo con el Big Data que ofrece servicios relacionados con diferentes etapas del trabajo con los datos, permitiendo el almacenamiento y la ejecución de aplicaciones. En primer lugar, se apoya en la tecnología HDFS para almacenar grandes cantidades de datos en forma de archivos, lo cual permite el trabajo con datos estructurados y no estructurados. Además de eso, permite el procesamiento de esos grandes volúmenes de datos usando MapReduce, que atomiza las tareas y las reparte entre los nodos que forman la red. En ambos casos, son servicios escalables, donde sumando nodos al trabajo podemos aumentar la cantidad de información que se puede almacenar y procesar. Hadoop ofrece otra serie de funcionalidades, como una interfaz gráfica para el acceso a las aplicaciones o bases de datos distribuidas, tanto relacionales como no relacionales, que también pueden ser de utilidad en el trabajo con el Big Data. [24]

## 2.2.4. Análisis

Una vez realizado todo el trabajo con datos, es necesario extraer conclusiones. Los datos y la información, de forma aislada, no tienen ningún significado ni aplicación útil para las empresas. Por tanto, es necesario, una vez procesados los datos y extraída la información útil, analizar los resultados, con el objetivo de realizar cambios en los procesos del negocio para mejorar los resultados. El uso de Big Data en las empresas puede repercutir en prácticamente todos los niveles del negocio, ya sea con el objetivo de transformar procesos internos, cambiar el tipo de relación con los clientes, lanzar nuevos productos o campañas y, en definitiva, cualquier cosa que añada valor y dé sentido a todo el trabajo realizado previamente.

Este análisis se puede realizar en diferentes dimensiones, según el enfoque y los objetivos que se pretendan conseguir en cada etapa. Cada una de las dimensiones tiene diferentes perspectivas, por tanto, la complejidad y el valor del análisis generado es también variable.

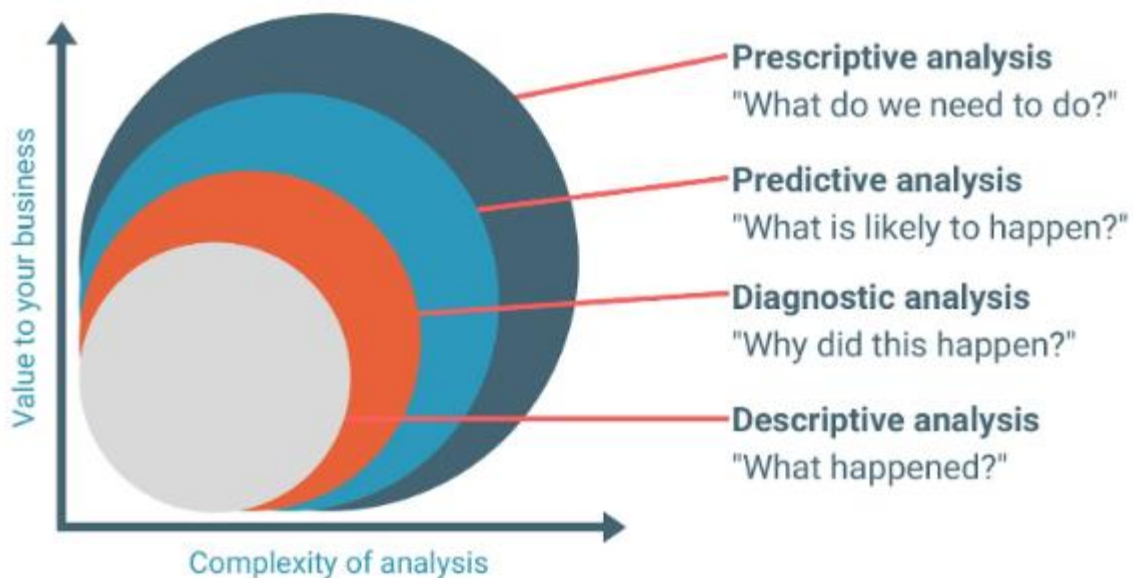


Figura 2. Tipos, complejidad, y valor de los análisis.[25]

### 2.2.4.1. Análisis Descriptivo

Este tipo de análisis busca analizar qué es lo que ha pasado en aquellos procesos que se están estudiando. Se limita, por tanto, a describir la realidad, almacenando y utilizando los datos históricos de la empresa. Gracias a este análisis podemos estudiar qué ha ocurrido hasta ahora, conociendo estadísticas del negocio. Permite, gracias a la visualización de las diferentes métricas, conocer las tendencias de la empresa, las cosas que han funcionado, y las que no; y sirve para realizar un acercamiento inicial al objetivo último que nos hemos propuesto, sea cual sea este.

### 2.2.4.2. Análisis Diagnóstico

Este tipo de análisis está directamente relacionado con el anterior; y es que si el descriptivo busca responder a la pregunta de "¿qué ocurrió?", el análisis diagnóstico busca entender las razones por las que los eventos analizados anteriormente ocurrieron, así como entender cómo se ha llegado a la situación actual. Un análisis diagnóstico preciso puede ofrecernos las respuestas necesarias para evitar que se repitan problemas pasados en el futuro, así como para intentar repetir fórmulas de éxito.

Este análisis, así como el anterior, se basan en el uso de la información ya existente, y por tanto son menos complejos de realizar y consumen menos recursos; pero también el valor extraído es más bajo, dado que no siempre podemos extraer información útil de cara al futuro de los servicios ofrecidos; más allá de intentar no repetir los errores que se hayan podido producir anteriormente, o corregir tendencias negativas. [26]

### 2.2.4.3. Análisis Predictivo

Como ya hemos analizado antes, la gran ventaja del Big Data es que el uso de cantidades tan grandes de datos, combinados con técnicas de análisis novedosas, permite extraer información o conclusiones imposibles de conseguir mediante el uso de enfoques tradicionales. En este tipo de análisis, el trabajo con patrones, tendencias o modelos encontrados en la información revisada permite realizar predicciones sobre eventos que pueden ocurrir en el futuro. Esto nos permite conocer las tendencias del mercado en el que nos encontramos y el posible funcionamiento de nuestro negocio en el futuro; así como adelantarnos a posibles sucesos negativos que puedan afectar a nuestro trabajo. La precisión de las predicciones realizadas dependerá del tipo y la calidad de herramientas que se utilicen, que pueden ser muy diversas dependiendo del tipo de información que se maneja; así como de la calidad en la información presente en nuestro sistema. Además, las predicciones serán más certeras dependiendo del momento en el que se encuentre el entorno de funcionamiento de nuestro producto. [25]

### 2.2.4.4. Análisis Prescriptivo

Igual que los análisis descriptivo y diagnóstico, el análisis prescriptivo tiene una gran dependencia respecto del análisis predictivo. En este tipo de análisis, el objetivo es conocer las posibles consecuencias que tendrían diferentes decisiones que se pudieran tomar, y de este modo ver qué es lo que más conviene al negocio en el que estamos trabajando. Al igual que en el caso anterior, las técnicas de simulación y de predicción tienen una gran importancia en esta fase, puesto que se quiere estudiar el impacto de decisiones que podrían tomarse. Además, deben tenerse en cuenta factores externos y que no se pueden controlar; pero que podrían afectar también a nuestra empresa.

Este tipo de análisis es el más complejo, dado que se apoya en una serie de predicciones del futuro y en la predicción del impacto de decisiones que aún no se han tomado; pero también es el más valioso, ya que, una vez analizados los posibles escenarios, la empresa puede decidir implementar algunas de esas posibles acciones que han visto que serían más beneficiosas. Junto con el análisis predictivo, supone el mayor valor en el uso del Big Data, dado que la gran cantidad de información, junto con las técnicas de computación tan avanzadas disponibles en la actualidad, nos permiten realizar predicciones sobre nuestro entorno mucho más certeras y precisas. [27]

## 2.3. Tecnologías para el trabajo con datos

A continuación se presentarán algunas de las tecnologías más populares para el trabajo con datos, repasando la recopilación, procesado y la posterior presentación de resultados y visualización de estos. Dentro de las distintas tecnologías mencionadas también se presentan las que se emplearán en la posterior aplicación práctica desarrollada, relacionada con el trabajo con datos de transporte.

### 2.3.1. Fuentes y formatos de los datos

Dada la propia naturaleza del Big Data, la procedencia de los datos para realizar trabajos de este tipo puede provenir de una gran cantidad de fuentes diferentes, en formatos y especificaciones muy diversas; y será tarea del analista o programador estudiar esos datos y organizarlos para poder extraer de ellos información valiosa. En general, cuando el trabajo con datos se realiza dentro de una organización con el objetivo de mejorar alguna parte del trabajo que realiza, muchos de los datos provendrán de los propios sistemas de la empresa; pudiendo incluir información de sensores (si hablásemos de una aplicación de Internet de las Cosas, por ejemplo), estadísticas de uso de los clientes o usuarios, etc. Esta información puede ser, además, complementada con informaciones externas que pudieran resultar de utilidad. De la misma forma, si un particular quisiera hacer un análisis con Big Data de cualquier tema, podría acceder a fuentes públicas de información; o solicitar a las distintas organizaciones generadoras de datos algún tipo de acceso a la información.

En la actualidad, existen gran cantidad de repositorios de información accesibles para todo el mundo y con información de todo tipo. En los últimos años, con la gran demanda por parte de los ciudadanos a las instituciones, muchos gobiernos y entidades públicas han creado portales de transparencia donde se publica información relacionada con las empresas gubernamentales. Por otro lado, también muchas empresas han puesto a disposición de los usuarios APIs a través de las cuales acceder a los datos de las organizaciones, ya sea de

forma gratuita con un registro previo o realizando un pago mensual o por uso. De la misma forma, grandes corporaciones tecnológicas como Google cuentan con una gran variedad de posibilidades para acceder a muchos de los datos que gestiona a través de todas sus aplicaciones y servicios; con modelos de pago similares a los mencionados previamente. Además de esto, existen muchos repositorios y webs mantenidos por particulares y enfocados en distintos aspectos con el objetivo de facilitar el acceso a los datos, proporcionando un punto de acceso único a muchos conjuntos de datos relacionados entre sí o del mismo tipo.

Aparte de las propias herramientas que los dueños de los datos ponen a disposición de quien esté interesado para acceder a su información, existen empresas que han desarrollado distintos *softwares* para el trabajo con datos, pudiendo generar, almacenar, visualizar o compartir estos datos; y que por tanto sirven como portales para que los usuarios accedan a consultarlos. En este caso, no nos referimos a plataformas como Dropbox, Google Drive o GitHub que permiten compartir cualquier tipo de archivo, sino plataformas especializadas en el tratamiento de los datos. En línea con el trabajo realizado, destaca la aplicación ArcGIS.

ArcGIS es un software que permite ver, editar, crear y analizar datos geográficos, mapas e información espacial. Está compuesto de varias aplicaciones y herramientas; y es el mayor software geográfico del mercado actualmente. Además de las aplicaciones de escritorio que permiten manejar distintos tipos de mapas y catálogos de funcionalidades, también cuenta con una plataforma online que permite visualizar sobre un mapa, una API REST para realizar consultas y una interfaz para descargar en diferentes formatos la información deseada. Por último, también cuenta con una librería oficial de Python que permite programar muchas de las tareas que también se pueden hacer en el propio software. [28]

Por otro lado, resulta importante hablar de la estandarización o normalización de los datos. En muchos casos de uso, saber cómo almacenarlos y tratarlos resulta de vital importancia. En este sentido, la estandarización de datos consiste en el proceso de adaptar o ajustar la información recopilada para que cumpla ciertas características de un modelo previamente elaborado; en general por organizaciones distintas a la propia. El hacer esto facilitará que el entendimiento de los datos, las relaciones entre ellos, su acceso y su uso sea más sencillo; pero además, a la hora de compartir el trabajo con otros o que otros hagan uso de nuestra información, simplemente con referenciar el modelo empleado será sencillo para los demás trabajar en el proyecto. La estandarización de datos, por tanto, ayuda no sólo a organizar la información, sino a trabajar con ella y a compartirla. [29] La estandarización de los datos es una actividad que puede darse de muchas formas distintas. Dos formas comunes de elaborar un modelo son que varias instituciones de un mismo ámbito se pongan de acuerdo para elaborar un estándar relacionado con esos datos con las que ambas trabajan, con el objetivo de elaborar una norma a la que todos se atengan; o bien que una o varias empresas se pongan de acuerdo para trabajar sus datos de una forma, y que sucesivamente otras empresas comiencen a adoptarlo de forma autónoma, convirtiéndose en un estándar *de facto*. El caso del estándar de transporte por excelencia, que se presenta a continuación, se corresponde al segundo caso.

### 2.3.1.1. Datos GTFS

*General Transit Feed Specification* (GTFS) es un formato de especificación e intercambio de datos de transporte nacido de la colaboración entre Google y TriMet, una agencia de transportes de Portland en 2005. El objetivo de la compañía era poner a disposición de todo el público los datos de transporte de la ciudad; pero hacerlo en lugares más accesibles que la propia web de la empresa, como podía ser Google Maps. En un principio no se desarrolló un formato común de especificación; sino que simplemente se buscaba que las agencias compartiesen sus datos de operación con Google. Durante el año 2006 más agencias se fueron interesando por estas actividades, lo que provocó que finalmente Google lanzase el *Google Transit Feed Specification*; donde muchas empresas de transporte colaboraron en la especificación del estándar, que fue una adaptación del formato empleado por TriMet el año anterior; con el objetivo de que fuera fácil para las ciudades adherirse a este formato de expresión de los datos de transporte. En 2013 el formato estaba ya en uso en cientos de ciudades alrededor del mundo, y actualmente es, de facto, el estándar empleado para la distribución de datos de transporte; se emplea en una gran cantidad de aplicaciones y ha ido evolucionando hasta proveer también de especificaciones para la actualización de datos en tiempo real. La transversalidad de este tipo de archivos llevó a los creadores a transformar el significado de la letra G del nombre del estándar, pasando a representar la palabra *General*. [30] Es decisión de cada uno de los operadores de transporte de cada ciudad o país si utilizar GTFS como el formato de especificación de los datos de su compañía, así como si hacerlos públicos o no. El mayor repositorio de *feeds* GTFS es mantenido por una organización sin ánimo de lucro llamada MobilityData; que cuenta con más de

1800 conjuntos de datos, incluyendo algunos del País Vasco, Madrid, Barcelona, Comunidad Valenciana o RENFE. [31]

La información contenida en un *feed* GTFS se corresponde con el servicio prestado en una semana estándar dentro de un periodo determinado, definiendo cuándo se encuentran activos cada uno de los posibles viajes cada día de la semana, agrupando los días de la semana por tipo y estableciendo el periodo de validez de esos datos; por ejemplo, porque el servicio cambie de estructura según la estación del año en la que se encuentre. Además de eso, se definen posibles excepciones al calendario establecido.

La referencia habla de 17 archivos diferentes, todos de tipo CSV para facilitar la lectura, comprensión y el trabajo sobre los mismos; donde todos ellos deben encontrarse en una carpeta comprimida. De ellos, no todos son obligatorios ni resultan de interés para el caso en el que nos encontramos. A continuación, se detallan los archivos por los cuales se ha compuesto el *feed* de transporte empleado en nuestro caso, con una breve descripción de la información más relevante que contienen. La referencia completa del formato se puede encontrar en su página oficial. [32]

- *agency.txt*: Contiene información relativa a la agencia de transportes de la cual se ha elaborado el conjunto de datos. Esta información incluye nombre, zona horaria, la URL a la web o información de contacto.
- *calendar.txt*: Contiene información relativa a los días de servicio de la agencia; estableciendo identificadores de servicio para los distintos días, los días de la semana que atañen y las fechas de comienzo y fin del *feed* especificado.
- *calendar\_dates.txt*: Permite definir excepciones en fechas de los servicios expresados en el archivo anterior.
- *feed\_info.txt*: Contiene información sobre la entidad publicadora del *feed*; que puede o no coincidir con la agencia de transportes a la que se refiere.
- *trips.txt*: Define un identificador para cada viaje, indicando el tipo de día donde se desarrolla según lo especificado en *calendar.txt*, haciendo referencia a la línea, sentido y destino a la que se refiere.
- *routes.txt*: Define todas las rutas (entendiendo una ruta como una posible variación de una línea) que realiza la agencia de transportes; indicando en cada entrada la línea a la que pertenece, el sentido de la ruta, el nombre y su identificador y el tipo de vehículo que emplea.
- *shapes.txt*: Contiene información de puntos geográficos de referencia de cada una de las rutas que componen el servicio de transportes; lo que permite trazar sobre un mapa los servicios ofrecidos. Puede contener también información relativa a la distancia recorrida por el vehículo en los diferentes puntos de la ruta.
- *stop\_times.txt*: Cada entrada de este archivo contiene la hora de llegada y salida del transporte para cada viaje y en cada parada del sistema. Es, por tanto, el archivo más crítico y extenso del sistema, dado que contiene la información más relevante del mismo. Es importante hacer algunas matizaciones al respecto. La primera es que, en nuestro caso, las horas de llegada y salida siempre coinciden. Esto se debe a que la empresa no especifica un doble horario para la llegada y la salida del transporte. Además, en el *feed* elaborado por TUSSAM la mayor parte de las filas de este documento no especificaban ningún punto temporal, por lo que estos datos han tenido que ser inferidos para realizar los cálculos pertinentes.
- *stops.txt*: Contiene información relativa a las paradas del sistema; indicando el identificador de la parada, su nombre y localización

### 2.3.2. Procesado de los datos

Los datos obtenidos de las distintas fuentes empleadas en un proyecto requieren, en general, una comprobación de la corrección de los mismos y una preparación de estos para tenerlos en un formato adecuado para trabajar, lo cual se conoce como preprocesado o preparado de los mismos. Esta preparación suele incluir, en general, el completado de información inexistente, la corrección de posibles errores o limpieza del conjunto de datos y el cambio en el formato de los ficheros o expresión de los datos. Posteriormente, ya con los datos preparados, será mucho más sencillo para el analista o programador el hacer uso de ellos, obteniendo estadísticas y otros datos a partir de la computación de los mismos.

Todas las tareas de preparación y procesado de datos, así como el acceso a la información (ya sea en bases de datos o ficheros) puede hacerse con una gran variedad de lenguajes de programación distintos, donde cada uno de ellos aporta diferentes ventajas e inconvenientes. La mejor opción para realizar el trabajo dependerá, por tanto, del tipo de almacenamiento que se haga de los datos, los resultados que se quieran conseguir, los tipos de datos que se manejen, la propia preferencia del programador o la dificultad de la tarea. A continuación se repasan los más comunes:

- **Python:** Es uno de los lenguajes de programación más populares para el desarrollo de software, pero también destaca como el más popular para otros campos como el trabajo con datos o la inteligencia artificial. Cuenta con una gran cantidad de bibliotecas y herramientas orientados específicamente al procesado de datos, así como a la visualización de los mismos. Además, permite utilizar el entorno de programación conocido como *Jupyter Notebook*, muy apropiado para el trabajo con datos por la facilidad para depurar y visualizar resultados intermedios durante el trabajo con estos.
- **Java:** En la actualidad, es uno de los lenguajes más utilizados para el desarrollo de aplicaciones de escritorio, desarrollo web y de aplicaciones móviles. No obstante, cuenta con herramientas que nos facilitan el trabajo con datos.
- **Scala:** Es un lenguaje de programación de alto rendimiento que se utiliza para el procesamiento de datos en tiempo real, cada vez más popular para el desarrollo de aplicaciones y estudios con Big Data.
- **R:** Es un lenguaje de programación orientado a la estadística y las matemáticas utilizado para el análisis de datos, y cuenta con una amplia gama de paquetes y herramientas para el análisis y la visualización de estos.
- **MatLab:** Es un lenguaje de programación utilizado para el procesamiento de señales, simulación numérica y el análisis de datos. Si bien Matlab no es tan común en el mundo de Big Data como Python, todavía se utiliza en algunos campos para el procesamiento de datos. Además, cuenta con la posibilidad de instalar distintos *plugins* para incluir funciones que podrían resultar de interés para el trabajo según los tipos de datos manejados.

Además, aunque no sea un lenguaje de programación, SQL destaca como tecnología para la consulta y trabajo de datos en bases de datos relacionales. [33]

### 2.3.3. Visualización de los datos

Cada uno de los lenguajes de programación anteriormente mencionados cuentan con diferentes herramientas para la visualización de los datos. Por ejemplo, MatLab cuenta con distintas funciones que permiten dibujar distintos tipos de gráficos, visualizarlos e interactuar con ellos. A pesar de ello, de cara a la visualización de resultados y datos destacan otra serie de tecnologías, que no son necesariamente parte de lenguajes de programación.

En primer lugar, Python no sólo destaca como un lenguaje que sirva para trabajar con los datos, sino también nos permite visualizarlos de forma cómoda, con una gran variedad de librerías enfocadas a esta tarea. Dentro de estas, la más famosa es Matplotlib, que facilita la creación de gráficos de distintos tipos, sencillos y que no permiten la interactividad; pero sí una gran capacidad de personalización; lo que la hace ideal para el dibujo de gráficos bidimensionales y con pocos datos, en general numéricos. No obstante, para tareas más complejas, esta librería se queda corta, haciéndose más adecuadas otras como Plotly, que sí permiten la interacción con los gráficos, abriendo un puerto al que conectarse a través del navegador para verlos correctamente. Para el caso que tratamos, además, puede resultar de interés la representación de datos geográficos, que en Python se puede realizar de forma sencilla con la librería Folium, que provee de una interfaz en Python para la librería Leaflet de JavaScript, enfocada en el trabajo con mapas interactivos a través de interfaces web. [34] [35]

Saliendo del ecosistema de Python, y en general de los lenguajes de programación, nos encontramos con diversas alternativas que son también muy populares para la visualización de datos y la generación de figuras. Dentro de estas alternativas nos encontramos con los distintos procesadores de hojas de cálculo, donde destaca Microsoft Excel como aplicación más popular y completa. Esta nos permite importar datos, generar tablas a partir de ellas e implementar lógica dentro de estas para que realicen tareas que resulten convenientes.

Por último, destacamos aplicaciones relacionadas con el *Business Intelligence* o inteligencia del negocio, que



nos permiten importar datos y generar informes visuales, utilizados en entornos empresariales para tareas como las presentaciones de resultados. Dentro de este sector, las dos más famosas son Microsoft Power BI y Tableau, que también permiten la actualización de los datos en tiempo real, la predicción de estos o la publicación de los datos en formato web de forma rápida.

Además de todas las alternativas anteriormente mencionadas, existen entornos específicos para la publicación y distribución de datos y la visualización de los mismos; como es el caso de ArcGIS para los datos geográficos. Por último, cabe mencionar que el diseño y programación web también son herramientas que se pueden emplear para la generación de informes o entornos de difusión de la información; aunque no sean herramientas propias del Big Data sino paradigmas generalistas que pueden ser empleados por analistas de datos para la difusión y visualización de los datos.

## 3 SMART CITIES

---

*Damos forma a nuestros edificios, luego ellos nos dan forma a nosotros.*

*- Winston Churchill -*

Las proyecciones actuales nos invitan a pensar que para el año 2050 la población mundial será de 9 mil millones de personas, de las cuales casi el 70% lo harán en ciudades. [36] Esta tendencia, no obstante, no es nueva, ya que durante los últimos años la población urbana ha ido aumentando de forma sostenida. Además, la preocupación por el medio ambiente y el uso de las tecnologías han ido influyendo cada vez más en nuestro día a día, lo que también afecta al entorno en el que vivimos. Este hecho, sumado a la cantidad de recursos que las urbes necesitan y la contaminación que generan, ha creado la necesidad de transformar la vida en las ciudades, con el objetivo de hacerlas más eficientes y sustentables, para, en definitiva, hacer sostenible la vida en ellas. Esta necesidad de transformación se ve potenciada por la creación de los Objetivos de Desarrollo Sostenible, que en su dimensión 11 trazan una serie de metas para las ciudades, tales como el acceso a servicios básicos, el transporte seguro y sostenible o la reducción del impacto ambiental de las ciudades. [37]

Trazar el origen del término *smart city* resulta complicado, dado que ya existen referencias a este a finales de los años 90 en Malasia, haciendo referencia a planes de desarrollo para la ciudad. [38] Sin embargo, podemos pensar que la referencia a este término distaba mucho de lo que pensamos ahora sobre él. En la actualidad, hablamos de *smart cities* para referirnos al uso de diferentes aplicaciones de las tecnologías de la información y la comunicación (TIC) en entornos urbanos con el objetivo de administrar de forma más eficaz los recursos de la ciudad. El objetivo último de esta gestión, y por tanto del desarrollo inteligente de las ciudades, es el aumento de la calidad de vida de sus ciudadanos, así como garantizar el uso sostenible de los recursos de las urbes. Estos objetivos generan una serie de desafíos, que afectan a muchas dimensiones de la vida diaria de la ciudad. A pesar de que los diferentes informes y análisis sobre las *smart cities* discrepan sobre el número de elementos que hacen a una ciudad inteligente, podemos agruparlos en los siguientes: [39] [40]

- **Entorno Inteligente:** Es una de las motivaciones fundamentales de las *smart cities*. Se refiere a todo lo relacionado con la gestión eficiente de la energía, la calidad del agua y la reducción de la emisión de gases nocivos para el medio ambiente. En este sentido, se incluyen desafíos como el uso de contadores conectados que realicen lecturas en tiempo real y otros sensores, la distribución de la generación de energía, la gestión de los residuos, la creación de edificios autosustentables y de espacios verdes.
- **Desarrollo humano:** Como ya se ha especificado antes, además de la preocupación por el medio ambiente, el objetivo de las *smart cities* es mejorar la calidad de vida de los ciudadanos. En este sentido,

los desafíos están relacionados con los servicios públicos, dotando de servicios sanitarios y de seguridad a los habitantes, combinando ambos elementos con el uso de las TIC. Además, debe favorecerse la educación, incorporando en esta el uso de las tecnologías. Por último, el desarrollo de las *smart cities* depende de la aceptación y conocimiento de sus habitantes, por lo que se necesita la complicidad de estos para desarrollar los planes de transformación de la ciudad.

- **Gobierno inteligente:** El desarrollo de las *smart cities* requiere de una legislación que impulse y proteja los planes necesarios para transformar las ciudades, así como la protección de los datos personales de los ciudadanos en entornos conectados. El gobierno se hace absolutamente imprescindible dado que abarca el resto de las dimensiones de la ciudad; requiriendo de una gestión eficiente, y del desarrollo de políticas útiles. También se busca favorecer la transparencia entre los mandatarios y los habitantes, la participación activa de los ciudadanos en las decisiones que se toman en la ciudad y la protección de los derechos y libertades.
- **Economía inteligente:** El desarrollo de la economía en la ciudad es fundamental para favorecer la creación de empleo, la llegada de empresas y la mejora en los servicios ofrecidos a los ciudadanos. La economía de las *smart cities* busca favorecer el emprendimiento, la innovación y el desarrollo; permitiendo que las necesidades de la ciudad sean abastecidas por empresas que atraigan inversiones. Además, se incluyen en esta dimensión elementos como el teletrabajo o la búsqueda de una alta productividad por parte de los trabajadores. El impulso del comercio online seguro forma parte también de uno de los desafíos del plano económico de las ciudades conectadas.
- **Movilidad inteligente:** Una de las claves para el desarrollo de la *smart city* es la búsqueda de un sistema de movilidad inteligente, capaz de anticiparse y reaccionar ante eventos. En este sentido, es necesario el uso de las TIC para analizar el uso del transporte en las ciudades, que permita planificar y desarrollar nuevos servicios y operar los existentes de forma inteligente. La movilidad inteligente no sólo implica el uso de las tecnologías en los transportes ya existentes, sino también el desarrollo de sistemas de transporte públicos lo suficientemente útiles para la mayoría de los ciudadanos. Además, con el objetivo de reducir las emisiones, busca favorecerse el uso de bicicletas y los vehículos eléctricos, con la consecuente necesidad de carriles-bici y puntos de recarga para estos vehículos; y de la compartición de vehículos. Es importante que los transportes públicos, al igual que el resto de servicios, sean accesibles, inteligentes y sostenibles. [40]



Figura 3. Dimensiones de la smart city. [41]

Además de las dimensiones previamente especificadas, existe un elemento transversal a todas ellas, que es crítico para la consecución de los objetivos que se plantean al desarrollar una ciudad inteligente, y es el de la tecnología. En relación con esta dimensión, necesitamos de redes de telecomunicaciones fuertes, que permitan a los

ciudadanos estar conectados en todo momento. El despliegue de redes 3G/4G/5G, cobertura móvil, puntos de acceso público Wi-Fi, fibra óptica para los negocios y particulares; junto con el uso de dispositivos inteligentes como *wereables* y otros sensores, o el desarrollo de aplicaciones móviles que permitan la interacción de los ciudadanos con su entorno son elementos fundamentales para el desarrollo correcto de smart cities. [40]

### 3.1. El papel del Big Data en las Smart Cities

Como ya se ha mencionado, el desarrollo tecnológico de la ciudad es fundamental para la correcta implementación de diferentes acciones que tienen como objetivo convertir las urbes en lugares inteligentes. En este sentido, la gran cantidad de datos generada todos los días por los dispositivos que ya utilizan los ciudadanos es muy útil; sin embargo, también es importante la inversión en otra serie de equipos que generen más información relevante para su análisis y explotación con el objetivo de hacer la ciudad más sostenible.

Es aquí donde el papel del Internet de las Cosas (Internet of Things, IoT) cobra una importancia fundamental, ya que dotando a los elementos que se encuentran en las calles, edificios y casas de la ciudad de estos dispositivos se pueden recabar una gran cantidad de datos que permiten conocer casi cualquier parámetro relevante que pueda afectar a los habitantes, así como permitir un estudio de diversos procesos presentes en la ciudad con el fin de mejorarlos. Además de cámaras, utilizadas para conocer el estado del tráfico en diferentes vías, la congestión de personas en una zona y la posible videovigilancia para aumentar la seguridad; y los sensores para medir diferentes parámetros medioambientales y relacionados con los recursos, las ciudades también pueden usar datos de los vehículos inteligentes, que pueden comunicarse con otros coches y con los sistemas de la ciudad para intercambiar información. Esto, sumado a los datos recopilados de los propios ciudadanos a través de sus teléfonos inteligentes, y el conocimiento previo existente sobre la ciudad supone la primera fase del trabajo con Big Data, y, en conjunción con otras técnicas, puede transformar el entorno urbano, con la meta de conseguir los objetivos fundamentales de las smart cities: una mayor sostenibilidad y un aumento de la calidad de vida de los ciudadanos.

Como ya se ha dicho, la instalación de sensores junto con la recolección de otros datos es muy importante, ya que son la base para posteriores análisis. Sin embargo, para que el uso del IoT tenga sentido, las ciudades también requieren de una red de comunicación que permita la transferencia de los datos recopilados al lugar donde se almacenen y procesen casi en tiempo real. De este modo, la red debe ser muy robusta, en tanto que debe soportar la transferencia de una gran cantidad de información, de muchos dispositivos, con una disponibilidad continuada; y muy rápida, ya que la transferencia de información, o las acciones que se deben comunicar tienen que ocurrir de forma instantánea, o lo más cerca posible de ello. Además del desarrollo de estas redes, se deben tener contempladas otras etapas del trabajo propio con Big Data, como son el tipo y lugar de almacenamiento, o el tipo de procesamiento de la información que se va a realizar. En este sentido, las opciones de almacenamiento y procesamiento en la nube son tan válidas como el uso del *edge computing*, pudiendo optarse también por soluciones híbridas. Por último, se requiere del desarrollo de diferentes algoritmos, que, apoyándose en la inteligencia artificial, sean capaces de recibir los datos y tomar decisiones sobre la ciudad, ejecutándose en los sistemas definidos previamente.

Son muchos los ejemplos en los que el uso de datos de la ciudad puede ser aprovechado para hacer de esta un lugar mejor para vivir. De este modo, el uso de cámaras para aumentar la seguridad y la instalación de GPS a patrullas de policías, ambulancias o bomberos, semáforos regulados en función del tráfico existente en una zona, sensores para analizar el consumo energético o planificar posibles picos de demanda, para conocer fugas en tuberías u otros dispositivos, o para conocer la posible contaminación del agua o el aire son aplicaciones del Big Data posibles en las ciudades de nuestros días. [42]

### 3.2. Smart Cities: Panorama actual

La cantidad de desafíos y ambiciones que presentan las smart cities hacen que en muchos casos suene como algo utópico. A pesar de ello, sí que han existido y existen proyectos que han servido como “laboratorio” para probar el concepto de smart city. La ciudad austríaca de Aspern, a 14km de Viena, es el mejor ejemplo, dado que ya es un lugar funcional, con habitantes y que sigue hoy día desarrollándose, con el objetivo de llegar en 2028 a 20000 habitantes. La ciudad, donde los coches están prohibidos, cuenta con líneas de autobús, un servicio

público de bicicletas y cuenta con un metro en construcción y un tren planificado. [43] La ciudad, creada como un entorno de pruebas por diferentes empresas, se define como el mayor proyecto de investigación sobre energía en Europa, y combina el uso de técnicas novedosas y eficientes de construcción de edificios, el uso de energías renovables, una red eléctrica inteligente (o *smart grid*), y tecnologías orientadas a las personas para conseguir una generación, distribución, almacenamiento y uso óptimo de los recursos energéticos. [44] Además de la ciudad de Aspern, existen otros ejemplos, como la ciudad de Masdar, en construcción, en los Emiratos Árabes Unidos; o la fracasada ciudad de Dongtan, en China. Sin embargo, estos proyectos se refieren a ciudades construidas desde cero, y es que no existen ciudades ya existentes que hayan podido ser adaptadas completamente al modelo de smart city por ahora.

No obstante, en muchos lugares sí existen proyectos que tienen como objetivo transformar y digitalizar el entorno urbano, con el objetivo de que este sea un mejor lugar para sus ciudadanos, incorporando algunos elementos que sí son propios de las smart cities, que buscan mejorar aspectos concretos de la vida en la ciudad más allá de conseguir una transformación completa de la misma. A partir de la incorporación de estas medidas, las ciudades se convertirán, poco a poco, en más inteligentes. En el caso de Sevilla, ya está en marcha un proyecto, denominado “Horizonte Sevilla Inteligente” que tiene como objetivos la conservación del patrimonio histórico, el avance hacia la eficiencia energética, la simplificación de gestiones administrativas, el análisis de diferentes métricas medioambientales y mejorar la seguridad y el transporte en la ciudad. [45] En esta línea, ya se han realizado proyectos a modo de prueba, utilizando la Semana Santa como escenario, dotando a la ciudad de cámaras de ultra alta definición para reconocer grandes aglomeraciones y posibles problemas; o reconocimiento de personas, iluminación inteligente o balizas para generar notificaciones en caso de emergencias. [46]

A pesar de que no existan smart cities completamente funcionales en nuestros días, existen diversos índices, entre los que destacan el Smart City Index, desarrollado por el *International Institute for Management Development* (IMD), y el Instituto de Estudios Superiores de la Empresa (IESE) Cities in Motion Index, que regularmente estudian las diferentes dimensiones de una smart city y cómo estas son llevadas a la práctica en diferentes lugares del mundo. En el caso del primer estudio, en 2021, Singapur ocupa el primer lugar de forma consecutiva respecto de años anteriores, siendo Zurich y Oslo las dos ciudades que completan las primeras tres posiciones. En este ranking la ciudad española mejor posicionada es Bilbao, que ocupa la décima posición. [47] En el segundo estudio, realizado en 2020, las tres primeras posiciones son ocupadas por Londres, Nueva York y París, en ese orden. En este documento, la mejor ciudad española es Madrid, en la posición 25. La ciudad de Sevilla, que sólo está presente en el segundo de los estudios, ocupando la posición 81. [40]

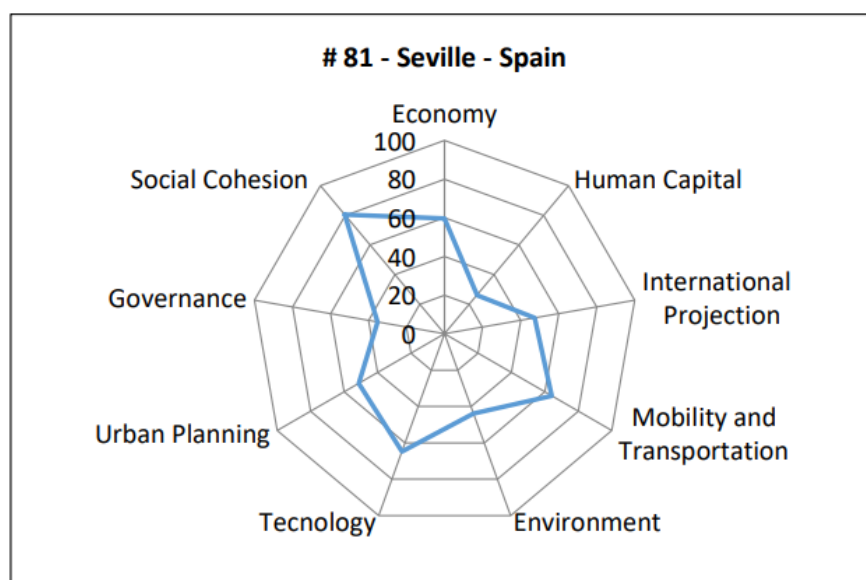


Figura 4. Puntuación de Sevilla en las diferentes dimensiones analizadas en el IESE Cities in Motion Index 2020. [40]

### 3.3. Mobility as a Service

En los últimos años, se ha desarrollado una tendencia cada vez más utilizada por empresas y usuarios relacionadas con la contratación de servicios informáticos para facilitar tareas del día a día. Dentro de estas soluciones, destaca el Software como Servicio o SaaS; en el que una entidad desarrolla y ofrece programas y se encarga de toda la gestión y el mantenimiento del software, así como de actualizar el mismo y dar soporte a posibles problemas. Estos servicios son contratados por entidades, que a cambio pagan una cuota mensual o de pago por uso, generalmente. La ventaja fundamental que comporta es que no requiere que las empresas desarrollen sus propias aplicaciones y las mantengan, lo que supone un gran ahorro económico y de tiempo, así como permitir gran escalabilidad y capacidad gracias a las tecnologías de cloud computing existentes actualmente. [48]

El desarrollo de la tendencia a ofrecer los productos como servicios adaptados a la necesidad de los ciudadanos o las empresas también se ha extendido a la movilidad, y se puede enmarcar dentro de la búsqueda de una movilidad inteligente, dimensión de las smart cities ya mencionada anteriormente. Así nace la Movilidad como Servicio o MaaS. Este proyecto tiene como objetivo principal la transformación de la movilidad dentro de las ciudades, pasando de una movilidad personalizada donde predomina el uso del coche a un sistema que se adapta a las necesidades de los usuarios, ofreciendo una gran variedad de alternativas en función del entorno y de sus propias preferencias, creando valor para los usuarios. Este cambio de paradigma permite ahorrar recursos a los ciudadanos, al no existir la necesidad de tener uno o varios vehículos en propiedad; pero también reduce la emisión de gases contaminantes a la atmósfera. [49]

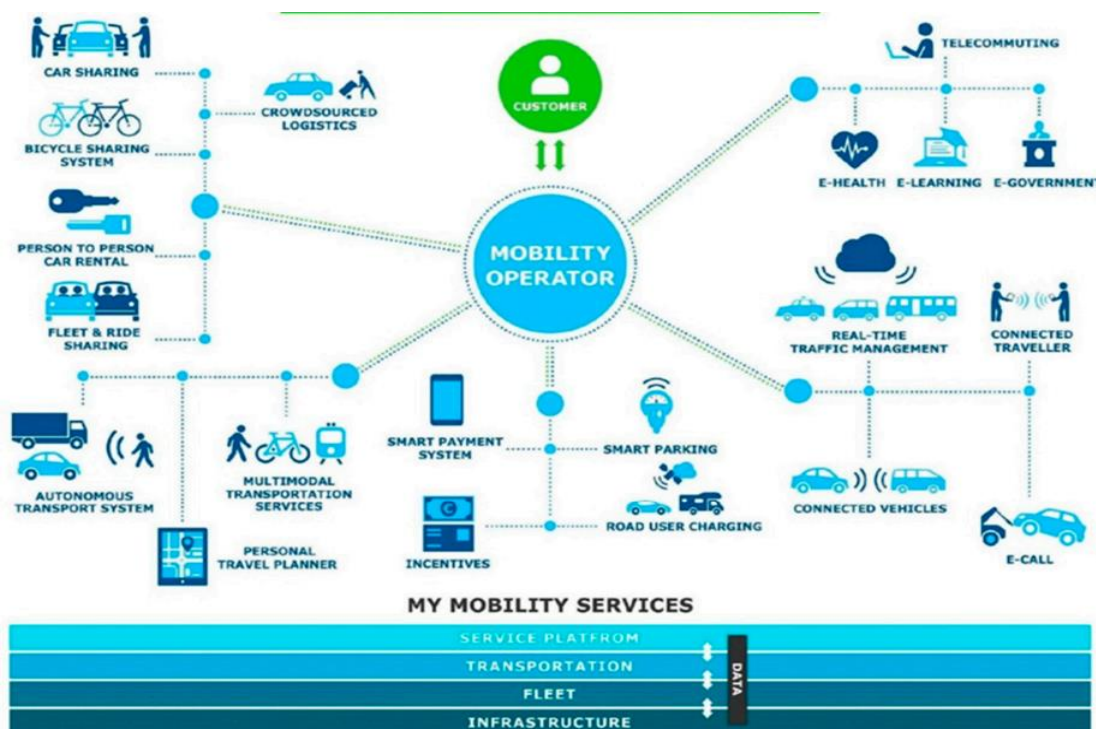


Figura 5. Elementos de la movilidad como servicio.[50]

Existen varios elementos centrales dentro de la movilidad como servicio, recogidos en la figura 5. El primero de todos es la integración de todos los modos de transporte en una sola plataforma, a partir de la cual se puedan recibir alternativas y combinaciones que incluyan el uso del transporte público, bicicletas, taxis y cualquier otro posible medio de transporte disponible en la ciudad; incluyendo también caminar. Para esto, las diferentes entidades que operan servicios de transporte en una ciudad, ya sean públicos o privados, deben integrarse de algún modo en la plataforma desarrollada; y debe garantizarse una unidad en los métodos de pago utilizados en todo el sistema; ya sea utilizando el pago por uso o con una suscripción mensual. [51] Además, con el objetivo de desalentar la compra de vehículos privados, debe favorecerse tanto compartir coche como el alquiler de vehículos, ya sean automóviles, motos, bicicletas y patinetes. Este elemento debe estar integrado de igual modo en la plataforma común de los diferentes servicios, permitiendo que sea una opción más para los usuarios.

Esta unificación de servicios y la fuerte intermodalidad requiere poner a disposición de la ciudadanía una aplicación o web que incluya un planificador de viajes que permita incluir preferencias de usuario; relacionadas con el tiempo, métodos a evitar, distancia máxima andando, etc. La simplicidad y la accesibilidad, tanto al planificar un viaje como al tener en un mismo lugar y con un mismo método de pago todos los medios de transporte disponibles son una clave del desarrollo de un correcto sistema de MaaS. [50] Otros servicios tecnológicos, como el manejo del tráfico en tiempo real para evitar rutas con atascos, la actualización en tiempo real de las mejores rutas para ir de un lugar a otro, o la búsqueda de lugares libres donde aparcar también deben estar integrados en el MaaS.

La movilidad como servicio requiere apoyarse, inevitablemente, en el Big Data; con el objetivo de intercambiar información entre los vehículos, la infraestructura, las empresas de transporte y los usuarios. Esto incluye la recopilación de información para conocer el estado y las necesidades de la red en tiempo real o el acceso a los datos de usuario para ofrecer información de interés para sus rutas. La infraestructura necesaria para desarrollarlo requiere por tanto de una red de sensores y elementos de recolección de información; tecnologías de comunicación, almacenamiento y procesamiento de los datos; y una gran seguridad, disponibilidad y escalabilidad en los diferentes sistemas implicados.

En muchas ocasiones, el desarrollo de un ecosistema de movilidad de este tipo en una ciudad o área metropolitana requiere también de un replanteamiento de la red de transportes en sí misma, con el objetivo de hacerla más eficiente y de ofrecer mejores opciones a los usuarios finales. Por ello, no sólo es necesario el desarrollo tecnológico de la ciudad, sino también el compromiso de gobiernos, empresas públicas y empresas privadas, con el objetivo de integrar todos los servicios para ofrecer un mejor producto final. [52]

Además, alineadas con la sostenibilidad, muchas ciudades están buscando electrificar su flota de transporte público y desalentar el uso de vehículos de combustión; lo que ha hecho aparecer más recientemente el concepto de eMaaS, la movilidad eléctrica como servicio. No obstante, desde un punto de vista de diseño e implantación; ambos modelos requieren de todas las características mencionadas anteriormente, relacionadas con el enfoque en los usuarios, la intermodalidad, la unificación de plataformas y pagos, el despliegue tecnológico necesario y el uso de Big Data. [50]

### 3.1.1 Aplicaciones MaaS

En la actualidad, algunas ciudades ya han comenzado a implementar elementos de la movilidad como servicio. Dentro de los diferentes proyectos, Helsinki, capital de Finlandia, fue la primera en desarrollar una aplicación que incluyese no sólo los diferentes medios de transporte público sino también modos privados para viajar de un lugar a otro, como el alquiler o la compartición de coches y bicicletas. Esta aplicación móvil, que también permite vincular una tarjeta de crédito y realizar el pago por uso de transporte o con una tarifa mensual marcó el punto de inicio; y también la referencia para el resto de las ciudades en el objetivo de desarrollar sistemas MaaS. [52] La aplicación, denominada Whim, permite vincular el calendario personal, para recibir notificaciones basadas en la ubicación de nuestros eventos. Posteriormente a su lanzamiento en Helsinki se ha expandido a otras ciudades de Europa, como Viena o Birmingham; y a Tokio. [53]

En otros lugares de nuestro continente se han desarrollado también proyectos exitosos de aplicaciones MaaS. En Viena, la aplicación WienMobil ofrece a los usuarios una gran cantidad de opciones para viajar por la ciudad; y gracias a acuerdos con empresas como Car2Go, DriveNow o Europcar, estos servicios pueden ser reservados y pagados a través de la aplicación. Existe también una tarifa anual de 365 euros que permite, a través de la aplicación, hacer uso de los servicios públicos de transporte. La ciudad de Hanover, en Alemania, también permite las opciones mencionadas anteriormente; realizándose un cargo mensual en el banco correspondiente a los transportes utilizados a lo largo de cada mes. [54] También en Alemania, el proyecto Qixxit pretendía unificar en una sola aplicación todos los medios de transporte del país, incluyendo trenes, barcos y aviones; integrando también los pagos en la plataforma. [55] Esta aplicación, no obstante, no se encuentra disponible al público actualmente.

En nuestro país, el proyecto más sólido de movilidad como servicio se comenzó a desarrollar en la capital, con una aplicación denominada MaaS Madrid. En la ciudad, existían en 2019 más de 40 empresas relacionadas con el transporte, incluyendo aquellas de compartir vehículos, aparcamientos, enchufes para recarga de vehículos eléctricos además de los servicios de transporte tradicionales. Esto llevó a la Empresa Municipal de Transportes (EMT) a desarrollar una aplicación agregadora de todos los servicios, con funcionalidades de cálculo de rutas en función de las preferencias, un mapa sobre el que ver diferentes servicios cercanos, indicaciones en tiempo

real, reservas y pagos. Para realizar esto, integrando tantos actores privados en una única pasarela de pago, desarrolló el servicio EMT Pay, una pasarela de pagos propia que podía comunicarse con los diferentes servicios para ofrecer esta funcionalidad a los usuarios. Además, la aplicación requería usar las diferentes APIs de las empresas de las que ofrecía los servicios. [56] Desafortunadamente, el desarrollo de esta aplicación no llegó a completarse debido a la escasa popularidad que tuvo, con tan sólo 1000 descargas, y actualmente no está disponible para los ciudadanos. [57]

En febrero de 2023, la empresa pública INDRA en colaboración con la Empresa Municipal de Transportes de Valencia anunció la creación de la primera aplicación pública de movilidad como servicio, con 6 localidades de 6 comunidades autónomas diferentes en su fase inicial; pero abierta a otras que se quieran unir. El proyecto verá la luz a finales de 2024, contará con 11 millones de euros de inversión y buscará revolucionar el sector de la movilidad en España, integrando en una misma plataforma un planificador de viajes multimodal y la posibilidad de acceder tanto a medios de transporte públicos como privados. [58]

Además de las iniciativas llevadas a cabo por empresas públicas, ayuntamientos y estados, existen otras aplicaciones que tienen como objetivo el desarrollo de la movilidad como servicio de forma global; aunque estas aún se encuentran lejos del objetivo de unificar en un solo lugar toda la información y todos los pagos necesarios para desplazarse de un lugar a otro de la ciudad. Así, en la aplicación de Google Maps, además de los trayectos en transporte público, también se ofrecen opciones relacionadas con el alquiler de bicicletas municipales o privadas, así como patinetes eléctricos o motos de alquiler, dependiendo de la disponibilidad de estas en la ciudad y en la zona donde el usuario se encuentre. Desafortunadamente, no se permite el alquiler directamente dentro de la aplicación, sino que se debe abrir la de la empresa que gestione el medio de transporte en concreto. De igual modo, se detalla el precio estimado utilizando diferentes empresas de Vehículos de transporte con Conductor (VTC) o coche de alquiler según la zona, pero tampoco puede reservarse a través de la aplicación de Google. La aplicación Moovit funciona de una forma similar, ofreciendo rutas en transporte público y también con otras empresas como Uber o ShareNow, aunque no permite solicitarlas, sino que se abre la aplicación de la empresa correspondiente.

Por otro lado, Meep es una aplicación que sí nace con el objetivo de convertirse en una plataforma Maas; estando disponible en muchos lugares de España, así como en Portugal, Marruecos, Malta y Chipre. [59] Esta aplicación es la que más soluciones combina de las mencionadas anteriormente, permitiendo consultar la ubicación de bicicletas eléctricas, patinetes y motos de una gran cantidad de empresas; así como ofreciendo opciones para llegar al destino en transporte público o taxi. También integra algunos aparcamientos públicos indicando su precio; y cargadores de coches eléctricos de algunas empresas. En este caso, se permite el pago a través de la aplicación dependiendo de la empresa y la ciudad, existiendo acuerdos entre la aplicación y los operadores sólo en algunos casos. En aquellos casos donde no está disponible, es necesario abrir la aplicación correspondiente para reservar o pagar. En los tres casos explicados, tampoco el metro o el autobús pueden ser pagados en las aplicaciones, teniendo que realizar el pago en el vehículo o estación.



## 4 CASOS DE USO DEL BIG DATA EN TRANSPORTES

---

*Una ciudad avanzada no es en la que los pobres pueden moverse en coche, sino una en la que incluso los ricos utilizan el transporte público*

*- Enrique Peñalosa -*

Como ya se ha comentado, el Big Data es un elemento que puede aportar un gran valor a aquellas empresas, instituciones o entidades que pongan en práctica tecnologías de recopilación y análisis de datos y sean capaces de basar sus estrategias y decisiones en las conclusiones extraídas a partir del uso de estas. De esta forma, las entidades públicas pueden incluir el uso del Big Data en los diferentes servicios ofrecidos para que estos funcionen de forma más eficaz y con un menor coste operativo; lo que tendrá siempre a la larga un beneficio para los ciudadanos y supondrá un ahorro de dinero público que puede ser destinado a la mejora de los servicios.

En el caso del transporte público, son cada vez más las ciudades que están implementando diferentes sistemas con el objetivo de conocer las rutas preferidas y el comportamiento de los usuarios a lo largo de las redes de transporte para hacer los servicios más eficientes y la información difundida más relevante para los ciudadanos. Basar las decisiones estratégicas de las compañías en evidencias observadas y analizadas con Big Data no sólo hace que estas se tomen en base a una información de mayor calidad, sino que los sistemas modernos de recopilación de datos son más eficientes que el método tradicional de encuestar a los usuarios; en tanto que no es necesario personal encargado de ello y el procesamiento puede realizarse en tiempo real. En las grandes ciudades, donde los diferentes transportes; así como los usuarios y los condicionantes a los que estos están sometidos suponen tal cantidad de información, el uso de Big Data con el objetivo de mejorar al máximo los sistemas de movilidad urbana se vuelven especialmente críticos.

### 4.1. El caso de Londres y Transport for London

Londres, la capital de Reino Unido, es una de las ciudades más pobladas del mundo, con una población en 2018 de alrededor de 9 millones de personas. Esto la convierte también en la ciudad más poblada de Europa occidental, y la más poblada de la Unión Europea hasta la salida del país británico. Además de esto, es uno de los centros empresariales más grandes del mundo. Al día, se realizan 5 millones de trayectos en metro, y 24 contando el resto de los transportes de la ciudad, que hacen que a lo largo más de 1300 millones de pasajeros hayan utilizado el transporte subterráneo. En horas punta, 543 vehículos se encuentran en funcionamiento por debajo de la ciudad de Londres. [60] Estas circunstancias, sumadas a las expectativas de crecimiento de su población en los próximos años; y a las grandes cantidades de turistas que visitan Londres cada día, hacen que el correcto

funcionamiento de la red de transportes de la ciudad sea crítico.

Transport for London (TfL) es la empresa encargada de la gestión de los transportes de Londres y su área metropolitana. Sus servicios no se ciñen sólo a la administración de autobuses, metro y tranvías; también gestionan ferries, trenes de cercanías, taxis para personas con necesidades especiales, estaciones de autobús de larga distancia, bicicletas públicas y el telecable del río Támesis. Además de la operación de todos estos transportes públicos, TfL también expide las licencias de taxi, gestiona las carreteras de la ciudad; lo cual incluye las obras que haya que hacer o los 6300 semáforos presentes en la ciudad, los carriles-bici y en general, colabora con el Ayuntamiento para planificar las calles de la ciudad, con el objetivo de hacerla amigable para los peatones. Dentro de todos estos servicios, destaca la operación del transporte público; con 402km de metro repartidos en 11 líneas de metro y 272 estaciones y una flota de 9300 autobuses utilizados en 675 líneas y con 50 estaciones y más de 19000 paradas a las que dan servicio, [60] no sólo por los volúmenes de vehículos, distancias y pasajeros que maneja, sino por el intento de hacer este servicio cada vez más popular entre la población, desplazando el uso de vehículos privados. Es en estos servicios, además de en el resto de los servicios ferroviarios, donde más inversiones se han hecho para aprovechar el Big Data en el día a día.

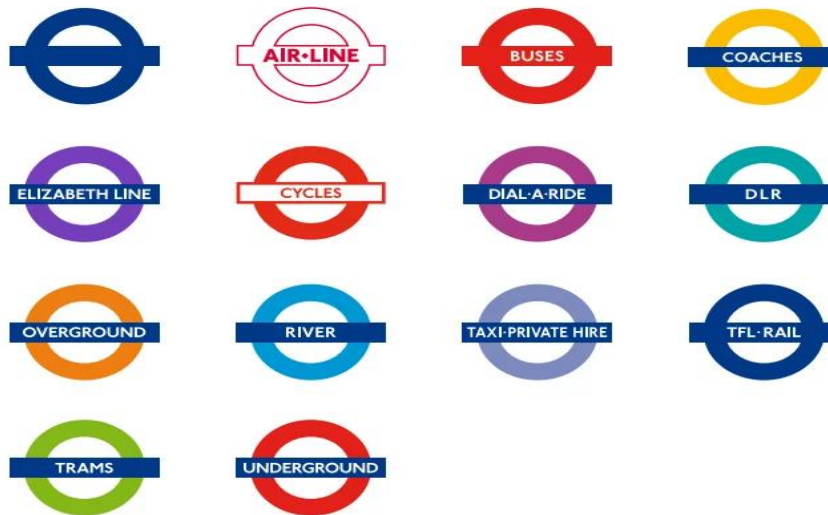


Figura 6. Logos de TfL (primero) y de los diferentes servicios que opera. [61]

La decisión de transformar TfL para convertirla en un referente de empresa pública que usa el Big Data para mejorar servicios de cara al público se tomó en 2007, momento en el que se lanzaron algunos widgets para seguir el estado del metro, intentando que los usuarios lo comprobasen los fines de semana, cuando el servicio se reducía. Al mismo tiempo, se intentaba mejorar el servicio del metro de Londres, en el contexto de los Juegos Olímpicos que se celebrarían unos años después en la ciudad; optando por formas innovadoras de realizarlo. Ese mismo año, en 2007, también se abren al público los diferentes conjuntos de datos disponibles para TfL y se lanzan otros widgets para seguir en directo noticias de viajes y planificar trayectos. En 2009 se lanza un área para desarrolladores en la web de TfL, y en 2010 se lanza el London Datastore, con información en tiempo real del transporte, así como de otras dimensiones de la ciudad. La razón por la cual se decidió abrir al público todo el conjunto de datos disponible fue al mismo tiempo un compromiso de la empresa pública con la transparencia; pero también con la intención de que desarrolladores de todo el mundo pudieran, de forma libre, crear aplicaciones que aportasen valor a los ciudadanos; mientras al mismo tiempo TfL continuaba desarrollando sus propias aplicaciones o páginas web para ofrecer servicios de información de calidad a los londinenses.

Con los Juegos Olímpicos celebrados en Londres en 2012 en el horizonte, durante este año y el anterior, las primeras APIs fueron lanzadas al público, orientadas a la planificación de trayectos y la localización y estimación de tiempos de llegada de buses y metros a paradas determinadas; además de un portal para acceder a los datos generados por este evento. [62]

En la actualidad, más de 80 conjuntos de datos están disponibles para todo el mundo, accesibles a través de una API unificada, con más de 12000 desarrolladores registrados y 650 aplicaciones usando los datos de TfL, entre las que se encuentran algunas tan importantes como Google Maps o Waze. Actualmente, el 42% de los

londinenses utilizan alguna aplicación que trabaja con los datos abiertos de TfL, mientras el 83% utilizan la web de la compañía para planificar sus viajes. [63]

A lo largo de los años, TfL ha ido desarrollando diferentes programas piloto antes de implementar sistemas para la recolección de datos, ha necesitado migrar sus datos y servidores entre diferentes proveedores de servicios cloud y ha generado beneficios que afectan no sólo al funcionamiento de la empresa, sino también a la ciudad de Londres; que se traducen en un ahorro económico, pero también en una mejor calidad de vida, gracias a un acceso más cómodo al transporte público.

A pesar de que TfL provee de información relacionada con otros temas y gestiona una gran cantidad de servicios en la red de transportes de Londres, el análisis se centrará en los asuntos relacionados con el transporte público; más concretamente con el servicio de autobús y metro, que vertebran la ciudad, aunque, a efectos prácticos, este análisis es extensible al resto de servicios ferroviarios, que utilizan mecanismos de recolección de datos iguales a los del metro. Se analizará cómo se recogen los datos y se hace frente a problemas asociados; y los usos que se les dan a los datos y beneficios de la implementación realizada por parte de TfL.

#### 4.1.1 Recolección de datos

Como ya se ha hablado anteriormente, para comenzar a trabajar con Big Data es necesario realizar una recopilación de una gran cantidad de datos, relacionados con todos los parámetros que pueden aportar valor al ser analizados. En el caso del transporte público, esto comporta diferentes aspectos. Por un lado, es necesario conocer toda la información relacionada con los vehículos y las infraestructuras a las que se da servicio; lo que incluye la localización de los vehículos, y su estado, así como el de las paradas o estaciones a las que los vehículos se dirigen. No obstante, si se desea dar una información de calidad, y personalizada, también es necesario conocer las rutas que suelen hacer los pasajeros, el tipo de viaje que hacen o avisarlos de posibles alertas. La forma tradicional de recopilación de información en el transporte público era el uso de encuestas, con el objetivo de conocer las tendencias de los pasajeros o sus preferencias a la hora de viajar. Con el desarrollo de la tecnología, existen técnicas mucho más eficientes, que no requieren el empleo de tiempo por parte de los pasajeros, y permiten extraer información mucho más valiosa.

##### 4.1.1.1 La Oyster Card

A pesar de que la decisión de transformar TfL se tomó en 2007, uno de los elementos fundamentales que han permitido la recopilación masiva de datos relacionados con el transporte público se tomó en 2003; con la implantación de la Oyster Card en el transporte metropolitano de Londres. Esta tarjeta, concebida como un abono de transportes que permitía a los usuarios del transporte público acceder de forma más cómoda a los vehículos, pagando al usarlos usando el modelo *pay-as-you-go*; también realizaba una recogida de información sobre los usos de la tarjeta en los distintos tipos de transporte, lo cual posteriormente ha permitido que la información histórica recopilada de las diferentes tarjetas pueda ser analizada para conocer información relevante sobre el servicio. En 2012, se cifra que más del 80% de los londinenses usaban esta tarjeta para pagar en el transporte público. En 2014, de forma paralela, se comenzó a implementar en los diferentes vehículos la posibilidad de pagar directamente con una tarjeta de crédito o débito, a través de un pago *contactless*, aprovechando el desarrollo de la tecnología NFC. A pesar de la popularidad que sigue comportando el uso de la Oyster Card, durante el verano de 2018 el pago directamente con este más moderno sistema superó al uso del abono de transportes tradicional, tendencia que se mantiene hasta nuestros días. Además, la posibilidad de pagar con el móvil ha acrecentado esta diferencia de uso, mucho más popular entre los jóvenes. A pesar de ello, el uso de la Oyster Card es muy relevante actualmente, dado que también provee de determinados descuentos en el transporte a determinados grupos de edad o segmentos de la población. [64]

La información recopilada en autobuses y trenes ligeros, donde sólo hace falta pagar al entrar, incluye el momento del día y la fecha en el que se ha realizado la transacción; la localización (entendida como la parada donde se ha subido el pasajero) y el número de vehículo y la ruta a la que se ha accedido. En el caso del metro, así como del resto de servicios ferroviarios, hace falta pasar la tarjeta por el lector al entrar y salir de las estaciones. Esto permite que el sistema almacene información de las localizaciones de entrada y salida, además de la fecha y hora de ambos eventos. Además de los datos recopilados con el trayecto realizado por el usuario, otra información útil es almacenada de cara a conocer estadísticas internas de la empresa. Así, al utilizar la Oyster Card en un autobús se conocerá el número de pasajeros totales que han pasado por ese vehículo; y en el metro, el número de viajes realizados en ese servicio, así como el número de usuarios utilizando esa estación.

Toda esta información es recopilada en la base de datos de TfL, que posteriormente se ofrece al público con la plataforma de datos abiertos ya explicada anteriormente. [65] Esta misma información se almacena si el pago se realiza con tarjeta de crédito o débito.

Además, si la Oyster Card o tarjeta bancaria están asociadas a una cuenta en la web de TfL, otra información personal será almacenada; incluyendo el nombre, la dirección, el teléfono, el historial de viajes realizados y el de recargas realizadas en la tarjeta Oyster. Esta recopilación de datos sobre los viajes de los usuarios permite que reciban avisos sobre servicios que usan frecuentemente; así como notificaciones sobre nuevas opciones de viajes, cambios en las tarifas o en los términos y condiciones del viaje. La información personal se almacena en los servidores de TfL, mientras que algunos datos relacionados con los viajes se encuentran en la tarjeta utilizada. Al acceder al portal de datos de TfL, la información y estadísticas sobre viajes que se ofrece está completamente despersonalizada, de forma que los datos personales de los usuarios del transporte público no son de dominio público. [66]

Como se ha explicado, el uso de la Oyster Card o del pago *contactless* tiene un papel fundamental en la recogida de datos por parte de TfL. Sin embargo, se requieren de muchos otros sistemas para tener una visión completa del sistema de transportes de Londres, que permita a la empresa que lo gestiona optimizar sus procesos a través del Big Data.

#### 4.1.1.2 El caso del metro y el rastreo WiFi

Como se ha dicho anteriormente, en el metro y el resto de los servicios ferroviarios administrados por TfL (a excepción del tranvía) es necesario pasar la tarjeta por un lector tanto al entrar en la estación donde se va a coger el transporte como en la estación de salida al finalizar el viaje, quedando registrados los puntos por donde ha comenzado el servicio y donde ha finalizado. Esto es extremadamente útil y supone una gran ventaja respecto de los transportes donde sólo se pasa la tarjeta al entrar, dado que el sistema puede conocer el destino final del viajero fácilmente, conociendo así sus viajes más frecuentes y también el número de viajeros que hacen determinados trayectos. No obstante, este sistema no permite conocer los diferentes lugares por los que el usuario ha pasado entre la entrada y la salida del sistema.

En el caso de ciudades como Londres, donde la red es tremendamente densa, es habitual que los usuarios del transporte público necesiten realizar transbordos de una a otra línea a lo largo de un trayecto entre dos estaciones. Además, dependiendo de las zonas donde se realice el trayecto, es posible que existan muchas opciones diferentes de transbordos que pueden llevar a un viajero a su destino final. En este caso, nos encontramos con una limitación del pago con Oyster Card o tarjeta *contactless*, que no pueden conocer esta información, en tanto que en los lugares de intercambio no es habitual (aunque en determinadas estaciones sí es necesario) pasar la tarjeta por uno de los lectores.

Tratando de solucionar este problema, y de conocer el comportamiento de los usuarios dentro de las estaciones una vez entrados en la red de transportes, TfL lanzó durante 30 días de noviembre y diciembre de 2016 un programa piloto para realizar un rastreo WiFi en 54 estaciones de la red, ubicadas en el centro de Londres y en las líneas más usadas. En ese momento, ya el 97% de las estaciones contaban con conexión WiFi pública, de forma que los usuarios de la red podían optar por esta opción, dado que en muchos lugares no estaban disponibles los datos móviles. Este programa tenía como meta trazar los movimientos de los usuarios entre esas estaciones a través del uso de la capacidad WiFi de los teléfonos móviles. Los objetivos eran:

1. Comprobar si la información generada de esta forma podría servir de utilidad para la planificación de trayectos y la elusión de congestión.
2. Entender los movimientos de los usuarios dentro de las estaciones para conocer las necesidades de los clientes y manejar eventos inesperados de forma más eficiente.
3. Entender cómo los usuarios utilizan el metro para cambiar horarios o diseñar estaciones de forma más eficiente.
4. Conocer la visibilidad de los diferentes emplazamientos publicitarios dentro de las estaciones. [67]

La característica empleada por el programa piloto fue aprovechar los mensajes utilizados en el protocolo WiFi, así como los módems ya instalados en las diferentes estaciones. En este tipo de conectividad, los dispositivos con la característica WiFi encendida envían periódicamente un mensaje conocido como *probe request* que sirve

para descubrir redes alrededor suyo. Al enviar este mensaje, incluyen información entre la que se encuentra la dirección MAC, que sirve para identificar de forma unívoca al dispositivo.

Una vez recibido el valor de la dirección esta pasaría dos veces por una función de hash. La primera de ellas, usando el mecanismo de la pimienta, es decir, almacenando el valor resultante a partir de la dirección y de un valor aleatorio igual en todos los elementos que se hagan pasar por la función. A partir de la salida de la primera función y de una cadena aleatoria diferente para cada dirección MAC, se haría uso de una nueva función de hash, con el mecanismo de la sal; generando el resultado almacenado en la base de datos.

El valor resultante, junto con el día, la hora y la ubicación del punto de acceso al que se envió el mensaje, eran almacenados en una base de datos. De esta forma, TfL podría procesar esas entradas creadas a partir de los mensajes de los terminales de los usuarios, y a partir de ellos, reconstruir rutas de viaje, diferenciando además entre los mensajes el lugar por el que un usuario entró y salió de la estación; así como los diferentes lugares por los que pasó a lo largo de un trayecto. [68]

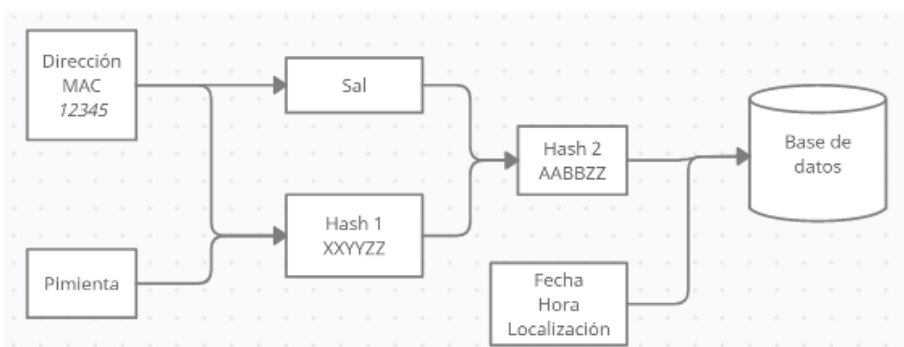


Figura 7. Procedimiento para almacenar la información en el programa piloto.

Para realizar este trabajo, sin embargo, fue necesario realizar algunas tareas relacionadas con la protección de los datos de los ciudadanos. Además de la anonimización de las direcciones MAC a través de las técnicas de *hashing*, este programa no permitió la identificación de la persona accediendo y moviéndose por las redes de túneles de metro, ni tampoco la información quedaba asociada de ninguna manera a la tarjeta con la que se había pagado el transporte público; aun así, era una forma de conocer datos de cómo los usuarios, de forma anónima, se movían por la red, lo cual permitió alcanzar los objetivos anteriormente especificados. Además, los usuarios que no quisieran ser rastreados podrían desactivar la característica WiFi en sus móviles, de forma que estos no enviarían los mensajes de rastreo, y por tanto no serían contabilizados. [69] Además de esto, TfL realizó una campaña previa de publicidad para explicar la importancia de este programa y los resultados que esperaban del mismo; con el objetivo de que los ciudadanos no se sintiesen vigilados y no desactivasen el WiFi de sus teléfonos móviles. [67]

Hashed MAC address	Date	Time	Access point
xIjx7F893lL4jO	01/12/2016	07:12:22	London Bridge 01
xIjx7F893lL4jO	01/12/2016	07:13:22	London Bridge 15
xIjx7F893lL4jO	01/12/2016	07:14:06	London Bridge 10
xIjx7F893lL4jO	01/12/2016	07:17:04	Bank 10
xIjx7F893lL4jO	01/12/2016	07:17:26	Bank 30
xIjx7F893lL4jO	01/12/2016	07:20:26	Old Street 06
xIjx7F893lL4jO	01/12/2016	07:20:50	Old Street 02

Tabla 1. Ejemplo de entradas en la base de datos generadas en el piloto de rastreo WiFi en el metro. [67]

Dentro de los diferentes datos recopilados, sólo aquellos datos de terminales que estuvieran registrados para poder conectarse al WiFi público de la red de metro eran almacenados y analizados. Esto se debe a que, cuando un dispositivo móvil envía mensajes de tipo *probe request*, oculta su dirección MAC real a los módems, generando direcciones aleatorias para evitar ser rastreados. En el caso de que los dispositivos sí estuvieran registrados, sí envían su dirección MAC real, para intentar conectarse a la red. De este modo, todos los datos correspondientes a dispositivos no autenticados eran eliminados tan pronto como fuera posible. Además, una

vez realizado el programa piloto, la información tuvo que limpiarse, eliminando patrones de movimiento no concluyentes o que no permitieran dibujar ningún trayecto.

Después de un mes de trabajo, 509 millones de mensajes fueron recibidos, de los que se pudieron reconstruir más de 42 millones de viajes completos de 5 millones de dispositivos diferentes. Gracias a estos datos, se pudieron sacar los siguientes beneficios:

- Se observa cómo, en estaciones con varias líneas y que suponen grandes lugares de intercambio, la ocupación reflejada a partir de los datos WiFi siempre era mucho mayor que la obtenida a partir de los datos de usuarios entrando o saliendo del metro con la Oyster Card o las tarjetas bancarias.
- Permitir, a través del planificador de viajes de TfL, mostrar la ocupación de las estaciones y vehículos de los que debe hacer uso a lo largo de su viaje en tiempo real, añadiendo también el tiempo que se tarda en realizar el transbordo de un tren a otro en función del aforo del lugar.
- Entender la reacción de los usuarios cuando un fallo en la red de transportes provoca que un determinado servicio entre dos puntos no esté disponible, observando los lugares a los que estos se desplazan para tomar el transporte alternativo y permitiendo planificar y reforzar estas alternativas en caso de necesidad.
- Entender la reacción de los usuarios ante eventos en las estaciones y permitir que la respuesta sea más segura y efectiva.
- En zonas donde existen muchas opciones diferentes para trasladarse entre dos puntos, conocer las rutas más populares.
- Realizar una mejor planificación de los servicios ofrecidos, realizar diseños y actualizaciones de las estaciones en función de las necesidades mostradas por los patrones de movimiento de los usuarios, así como modificar los horarios de las diferentes líneas en caso de necesidad.
- Conocer los espacios publicitarios más y menos visitados, permitiendo ajustar los precios de estos, así como conocer las rutas más comunes para poder ofrecer espacios publicitarios a los mismos vendedores a lo largo de una ruta, de forma que los usuarios de esa ruta vean de forma repetida el mismo anuncio. Esto permite una maximización de los beneficios obtenidos por publicidad, que pueden ser invertidos en una mejor red de transportes. [68]

El 8 de julio de 2019, después de analizar los datos recopilados a partir del programa piloto e introducir algunos cambios, el sistema de rastreo WiFi fue implementado a lo largo de toda la red de metro de Londres donde la conexión WiFi estaba disponible; de forma que, en la actualidad, este procedimiento se sigue realizando, permitiendo a TfL continuar explotando los beneficios previamente mencionados. [70] Además, este tipo de rastreo fue utilizado posteriormente en un programa para conocer los aforos en los autobuses, como se detalla posteriormente.

De este modo, la recopilación de datos por parte de TfL en el metro y el resto de los servicios ferroviarios está absolutamente parametrizado. Por un lado, TfL es capaz de conocer los viajes que realizan los usuarios de forma regular, permitiendo ofrecer información personalizada a estos. Por otro lado, la empresa es capaz de recopilar datos anónimos acerca del estado de los vehículos y estaciones, lo cual se traduce en más información para los usuarios del transporte, así como permite un análisis interno para mejorar los servicios ofrecidos.

#### **4.1.1.3 El autobús y sus problemas asociados**

Como ya se ha mencionado anteriormente, en el caso de la red de autobús y tranvías de Londres el pago sólo se realiza al entrar al vehículo, y no al salir. Al pasar la tarjeta por el lector se cobraría directamente el importe, y el usuario podría bajarse del transporte cuando lo quisiera. Esto comporta dos problemas fundamentales. Por un lado, es imposible saber, a diferencia de lo que ocurría en el caso del transporte ferroviario, el lugar donde cada usuario termina su viaje, lo cual se traduce en una mayor dificultad para ofrecerle actualizaciones de su viaje de y noticias personalizadas que resulten de su interés. Además de eso, es imposible saber con exactitud el número de personas utilizando el servicio en un determinado momento; como si era posible hacer en el metro, donde se conocen cuándo los usuarios entran o salen del servicio. Por último, en el caso del autobús y el tranvía, que comparten vías urbanas con el resto de los vehículos públicos y privados que transitan las carreteras, el tráfico se convierte en un factor crítico para el buen funcionamiento de los servicios, siendo necesario conocer el tiempo

estimado de llegada con exactitud y de forma dinámica.

#### 4.1.1.3.1 iBus y la localización automática de vehículos

Para solucionar el problema relacionado con el tiempo estimado y la localización del autobús, TfL y Siemens anunciaron en 2006 el desarrollo un sistema conocido como iBus; implementado definitivamente en los autobuses de Londres entre los años 2007 y 2009; año en que se completó la instalación en todos los vehículos que prestan el servicio. Originalmente, este sistema tenía como objetivo mejorar la forma de transmitir la siguiente parada a través de anuncios visuales, intentando integrar a las personas sordas, aunque en la actualidad se ha convertido en una forma de mejorar la comunicación con los pasajeros del transporte público y conocer con precisión la localización y el tiempo estimado de llegada de un vehículo. [71] iBus es un sistema de localización automática de vehículos que provee a cada autobús de un ordenador basado en Windows, que contiene información de todas las líneas y paradas de la ciudad, permitiendo a través de mensajes sonoros y visuales, la comunicación de las siguientes paradas; así como de otros mensajes previamente guardados que pueden resultar de interés para los viajeros, como el final de la línea o la necesidad de liberar hueco para una silla de ruedas; habiendo más de 30000 anuncios diferentes posibles. [72]

Además del ya mencionado ordenador, cada autobús cuenta con un GPS, un cuentakilómetros, un velocímetro, un sensor de velocidad rotacional y un giroscopio. La combinación de la información generada por estos sensores, junto con el uso del ordenador y las diferentes líneas y paradas de autobús que conforman la red de Londres permiten no sólo que el autobús sea capaz de detectar la próxima parada y anunciarlo en el vehículo; sino también conocer la localización de forma mucho más precisa, con un radio de 10 metros; y el tiempo de llegada a la siguiente parada con un margen de 10 segundos. Esto supuso un gran avance de cara a la planificación de viajes, dado que en muchas paradas de autobús existían ya pantallas que permitían conocer de forma casi perfecta la estimación de tiempo que le faltaba a cada uno de los autobuses gracias a la nueva implementación. Además, en 2011, se lanzó un servicio de pago que, gracias a iBus, permitía conocer el tiempo estimado de llegada de las diferentes líneas a una parada enviando un SMS con el código de la parada deseada. Este sistema ya no está operativo; pudiéndose consultar la misma información a través de la aplicación y la web de TfL o de las diferentes compañías que hacen uso de sus datos.

Además del sistema de localización, los nuevos avisos y la generación de datos sobre los tiempos estimados de llegada a un destino; el despliegue de iBus también supuso el despliegue de un nuevo sistema de radio que mejoraba el anterior, permitiendo a los conductores de autobuses comunicarse con el centro de control de forma más rápida y con comunicaciones de más calidad, además de permitir que las comunicaciones fueran a toda la flota, sólo a un grupo de conductores o a uno determinado. Este mejorado sistema de radio, combinado con la mayor precisión en la localización, supuso un avance en la seguridad, dado que, en caso de una emergencia, el autobús podría comunicarse con mayor eficacia y el centro de control podría conocer la localización con mucha más exactitud y rapidez; actuando sobre el incidente de forma más rápida. [73]



Figura 8. Interfaz de iBus para los conductores de autobuses, con diferentes avisos para los pasajeros.

El método para calcular la localización y el tiempo estimado hasta la siguiente parada hace uso de los diferentes elementos de iBus. Por un lado, el GPS permite conocer la posición con un radio de 100 metros; pero esto no es suficiente, debido a que la precisión se puede ver afectada por el movimiento del vehículo y por los edificios



aledaños. De este modo, además de la posición del GPS, se contabilizan la distancia recorrida desde la última parada (a partir del cuentakilómetros), las paradas por las que el autobús ya ha transcurrido previamente (conocidas gracias al momento donde el autobús abrió y cerró las puertas), los giros que se han realizado (conocidos a partir de los giroscopios) y la ruta conocida que el autobús está realizando. A partir de todos esos datos, el sistema es capaz de conocer la posición del autobús con un margen de 10 metros; siendo la media del error de precisión de 3 metros. [73] Una vez conocida la localización del autobús, esta información es enviada al sistema central, que, a partir de otros datos como los del tráfico puede actualizar el tiempo estimado de llegada, tanto en el sistema de datos abiertos como en las paradas con contadores luminosos.

Por último, gracias a iBus, el sistema central puede utilizar la información de la localización del autobús para alterar el funcionamiento de los semáforos; regulando el tráfico de forma más eficiente y dando prioridad al transporte público. La monitorización de la localización de los autobuses con el objetivo de cambiar los semáforos no es un procedimiento exclusivo de Londres; otras ciudades también lo han probado. En el caso de iBus se establecen cuatro niveles de prioridad; que los autobuses pueden solicitar en función de las necesidades. Se ha demostrado, a través de diferentes pruebas, que la priorización de autobuses usando los semáforos para regular a los vehículos que van tarde respecto de sus horarios planificados puede ser efectiva y tiene grandes beneficios para el servicio prestado a la ciudadanía. [74]

#### 4.1.1.3.2 Inferencia de origen y destino

Como se ha mencionado en múltiples ocasiones, al utilizar el autobús o el tranvía sólo es necesario pagar al subirse al vehículo y no al finalizar el trayecto; de forma que resulta imposible conocer la ruta que hace un usuario determinado sin utilizar técnicas más avanzadas. Esto hace que, a diferencia de lo que ocurre con los usuarios del metro, sea complicado ofrecer noticias personalizadas sobre los trayectos que un cliente utiliza de forma regular. Para solucionar este problema se utiliza el método de la inferencia de origen y destino, el cual aplica sólo al transporte en autobús en la ciudad de Londres. [75]

Con esta técnica, basándose en una serie de parámetros, se intenta determinar el recorrido que realizó un usuario en autobús, a partir de los siguientes medios de transporte utilizados. Este método depende fuertemente de algunas suposiciones, desarrollados en diferentes estudios desde el año 2007, adoptadas por las diferentes implementaciones llevadas a cabo en Londres y otras ciudades; que permiten aplicar el método a ciertos trayectos. La primera de las suposiciones tiene que ver con la distancia recorrida para realizar un transbordo; asumiendo que el usuario se bajará en la parada de su primer transporte más cercana al lugar donde tomará el siguiente transporte. La segunda suposición tiene que ver con la simetría de los viajes, dando por hecho que el destino al final del día de un usuario será cercano al lugar donde se tomó el primer transporte ese mismo día. [76] Además de eso, la asociación de viajes es imposible de realizar si los viajes no se realizan con el mismo método de pago o si se utiliza un medio de transporte privado al ir de un lugar a otro; y se introduce una limitación de distancia andando entre el final de un trayecto y el comienzo del siguiente. [77] Algunas de estas suposiciones no han sido utilizadas en todos los estudios empleados, o se han empleado variaciones de las mismas; por ejemplo, respecto de la distancia máxima andando que puede darse entre el final de un trayecto y el comienzo del siguiente.

Dicho todo esto, es importante recalcar que este método no permite determinar el trayecto de un usuario en todos los viajes que realice; y en ocasiones, tampoco permite determinar con exactitud la parada de autobús en la que finalizó una ruta, pudiendo confundirla con otra parada cercana. A pesar de que este sistema ha sido implementado en Londres, es aplicable a muchas otras ciudades del mundo donde se han realizado estudios similares, algunos basados en las conclusiones extraídas de la capital británica. [78]





Figura 9. Ciudades donde se han realizado investigaciones sobre la inferencia de origen y destino entre 1994 y 2021. [79]

Para trabajar con este método, lo primero que se debe hacer es conocer el lugar donde el usuario tomó el autobús para comenzar su viaje, dado que esta información no es registrada directamente por los sistemas de TfL; siendo necesario inferir el origen de este. Cuando un usuario pasa la tarjeta por el lector, la información registrada, además de los datos de la tarjeta, sólo tiene que ver con el vehículo y la ruta correspondientes. A través del sistema iBus, que almacena el historial de ubicaciones de los distintos autobuses, se puede conocer en qué lugar entró el usuario, lo cual se realiza cruzando las diferentes horas a las que un autobús pasó por cada parada, y el momento en el que se realizó el pago en el terminal habilitado para ello. No obstante, esto sólo funciona si la lectura de la tarjeta se realizó entre la llegada de autobús a la parada y su salida; o con un margen de tiempo determinado entre el momento de pago y la parada más cercana.

Existe un parámetro, denominado el error de inferencia de origen, que se refiere a la diferencia entre el momento del pago y el momento de salida del autobús de la parada donde se subió el pasajero. Si el pasajero se subió antes de que el autobús saliera, este valor será cero. Cuando el error supera una tolerancia determinada se generará un fallo, de forma que el proceso se detendrá, sin determinar el origen, y no se podrá determinar tampoco destino del viaje. A pesar de los posibles errores, los diferentes lugares donde esta técnica se ha implementado muestran una gran efectividad; infiriendo más del 94% de los orígenes en pruebas realizadas en Londres [76] y más del 97% en Boston. [78]

Si el origen de la ruta se ha encontrado, se intenta descubrir el destino del trayecto realizado en autobús. Para ello, es necesario determinar una localización objetivo que sirva como referencia para obtener la posible parada donde el usuario se bajó del autobús. La necesidad de encontrar una referencia para inferir el destino implica que, si el viaje realizado en autobús es el único registro de viaje presente para esa tarjeta en un determinado día, por tanto, será imposible determinar el destino del usuario, en tanto que no existe una localización objetivo. Una vez descartada esta opción, existen dos casos diferentes, según si existe un otro viaje más tarde durante ese día o si es el trayecto en autobús actual es el último de la jornada.

Si nos encontramos en el primer caso, donde existe un trayecto posterior a lo largo del día, se establece como localización objetivo el lugar donde se ha utilizado la tarjeta en siguiente ocasión. En el caso de que sea una estación de metro o tren, la localización es conocida; si es otro autobús, será necesario inferir el origen de la nueva ruta para conocer la localización objetivo. Una vez establecida esta ubicación, se intentará encontrar la parada más cercana de la ruta previa, confirmando con los datos de iBus si el vehículo paró o no en ese lugar. Si en la parada encontrada no se realizó una apertura de puertas que el sistema de iBus registrase quiere decir que no paró allí; por lo que se probará con la siguiente parada más cercana. Este algoritmo se repite hasta un límite, típicamente de 1km, alrededor del destino objetivo. Si no se encuentra ninguna opción viable, el algoritmo fallará. Como ya se ha comentado, este método no está exento de posibles imprecisiones. Por ejemplo, un usuario podría bajarse una parada antes de la más cercana del autobús para realizar una gestión y luego caminar a su siguiente destino; ante lo cual el sistema podría no ser capaz de identificar la parada con exactitud. [76] En el ejemplo de la figura 11, se muestran dos inferencias de destinos para este caso de trayectos sucesivos, y uno adicional al final de la tercera ruta, referido al caso del último trayecto del día que se desarrollará posteriormente.

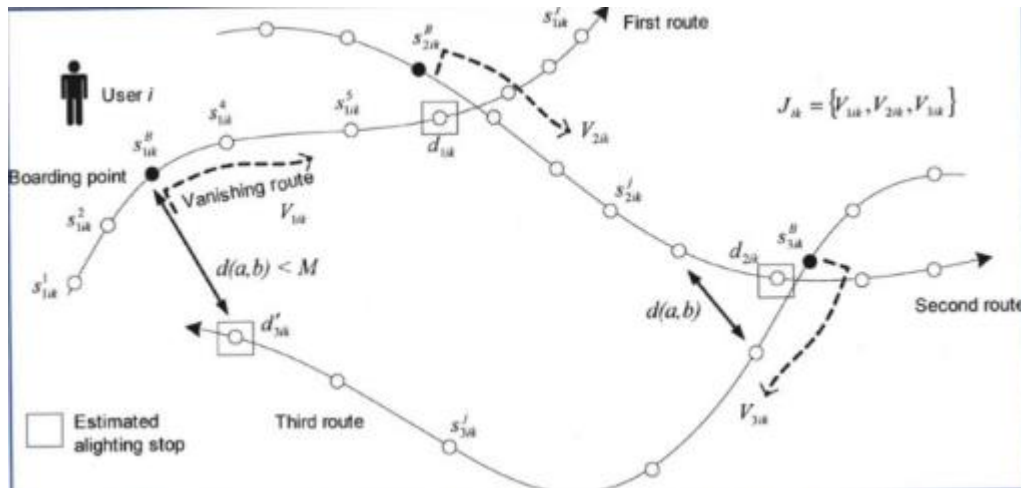


Figura 10. Ejemplo de inferencia de destinos entre diferentes rutas. [76]

Versiones más avanzadas de este algoritmo permiten en la actualidad, haciendo uso de los datos de iBus, conocer si es posible que un usuario se desplazase andando del final de un trayecto al comienzo del siguiente en el tiempo transcurrido entre el final de un viaje y el comienzo del siguiente; añadiendo una capa más de comprobaciones a la selección del posible final de un trayecto. De este modo, una vez inferido el posible destino, se comprueba la hora a la que el autobús llegó a la ubicación, y la hora de la siguiente lectura de la Oyster Card o tarjeta de crédito; y se determina si es posible que un usuario se desplazase en ese tiempo de un lugar a otro, considerando una velocidad máxima determinada de movimiento que, a pesar de variar en diferentes estudios realizados, se encuentra cerca de los 5km/h. [77] Si no fuera posible desplazarse de un punto a otro en ese tiempo, se comprobaría la siguiente parada más cercana, como en el caso anterior. En este caso, la suposición de que el usuario no se desplaza haciendo uso de un transporte privado entre los dos puntos, como un coche o bicicleta, tiene especial relevancia. [80]

En el segundo caso, donde el trayecto realizado en autobús es el último del día, el destino inferido será el más cercano a la parada desde la que se realizó el primer viaje del día; basándose en la simetría de los viajes asumida anteriormente. En el caso de la figura 10, un usuario comienza su día, desde su casa, tomando un autobús del que ha sido posible inferir el origen. Tras realizar una serie de tareas a lo largo del día, y desplazarse en otras rutas, vuelve a su casa haciendo uso del tercer trayecto; bajándose en una parada cercana a su casa. Este destino será el inferido por el método aplicado, en tanto que las demás paradas del autobús que ha tomado como último transporte del día son más lejanas a la primera parada utilizada esa jornada. En este caso también se incluye un límite, estableciendo una distancia máxima entre el último destino del día y el primer origen. De esta forma, si la distancia entre la parada más cercana de la ruta 3 y la primera de la ruta 1 estuviera más lejos del límite marcado, el sistema fallaría. [80]

Por último, es importante subrayar que este método es fuertemente dependiente de los datos grabados por los diferentes sistemas de TfL. En consecuencia, si no existiera información sobre la localización de las diferentes paradas que intervienen en un viaje, del vehículo empleado o de los lugares y tiempos de las diferentes paradas sería imposible determinar las rutas utilizadas por los pasajeros con la aplicación de este algoritmo.

La aplicación de este método en Londres ha arrojado resultados positivos, infiriendo más del 57% de los destinos de las rutas, cifra que sube hasta casi el 80% en determinadas líneas de autobús. Estos datos fueron posteriormente comprobados y validados con una gran encuesta realizada dentro de las diferentes líneas estudiadas, donde se encontró que el 65% de los intercambios totales habían sido inferidos. [77]

Posteriores estudios han trabajado orientados a conocer si el uso consecutivo de dos medios de transporte forma parte o no de un mismo viaje que requiriese un intercambio. Para determinar si dos trayectos son parte o no de un mismo viaje, se considera que no lo son si la actividad realizada entre ambos usos del transporte era un objetivo en sí mismo que motivaba el primer viaje. Si lo era, entonces los dos usos del transporte forman parte de viajes separados, en tanto que el primero tenía como objetivo realizar una actividad y el segundo no. Por ejemplo, si al realizar un transbordo entre dos usos de autobús el usuario parase en un kiosco a recargar el saldo en su tarjeta Oyster, podrían considerarse ambas rutas parte del mismo trayecto, y la parada para la recarga no era un destino en sí mismo. Por el contrario, si al realizar un trayecto andando de una parada de autobús a otra

el usuario parase en una oficina de correos a enviar un paquete, esta actividad sí resultaría un objetivo en sí mismo; siendo dos viajes diferentes los realizados antes y después.

Para determinar esto, aunque es confuso y complicado, se tiene en cuenta el tiempo transcurrido entre los dos usos, y la diferencia media de tiempo a partir de datos anteriores de otros usuarios (o del propio usuario en días anteriores); influyendo la distancia que separa las paradas y la velocidad al caminar, así como posibles caminos que se hayan podido seguir y el tiempo “perdido” respecto del estimado en caminar de un lugar a otro. [81]

Los siguientes diagramas muestran el algoritmo de funcionamiento de este método. Para el primero de ellos, referido al origen, se asume que el medio de transporte escogido es el autobús, en tanto que para el metro es conocido el origen y en el tranvía no se utiliza este método. Para el segundo diagrama, que tiene como objetivo explicar cómo se calcula el destino, se asume que el modo de transporte es el autobús (para tranvías el método no está definido, y para metro y otros servicios ambos puntos son conocidos), que existen otros registros en ese día y que la localización ha sido inferida. Como ya se ha especificado, en los casos de que no existan más registros para ese mismo día sobre otros usos de la tarjeta o no exista un origen inferido, no existe la posibilidad de encontrar el destino.

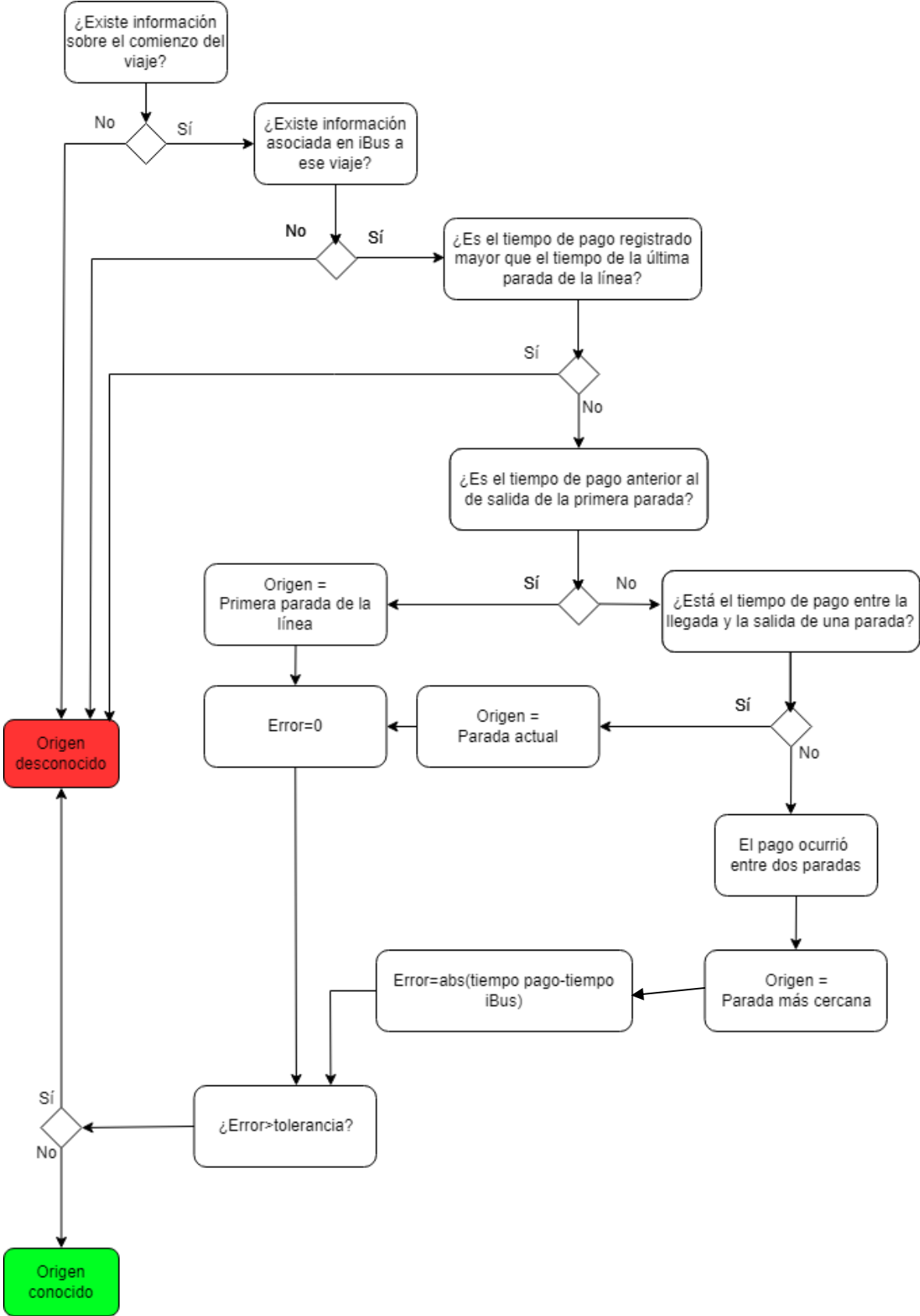


Figura 11. Algoritmo de inferencia de origen.[75]

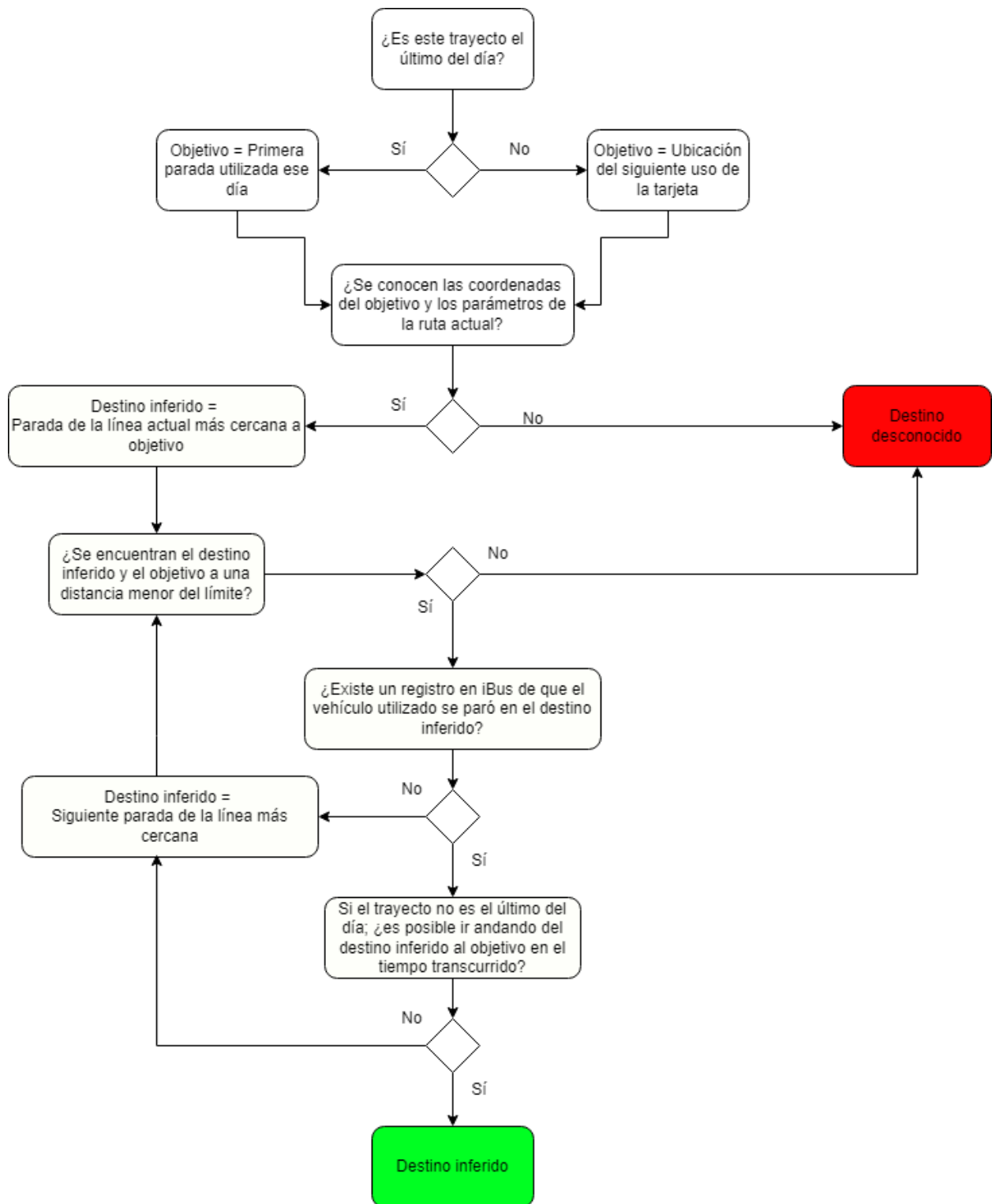


Figura 12. Algoritmo de inferencia de origen. [75]

#### 4.1.1.3.3 Cálculo del aforo en autobuses

Como se ha comentado, otro de los problemas asociados a que el pago en autobuses sólo se realice al entrar es la imposibilidad de conocer de forma sencilla el número de personas en el interior del vehículo en un momento determinado. En el metro, era posible conocer en qué parada entraba y salía cada usuario, pudiendo conocerse el estado de la red de forma global (aunque no fuera posible determinar la cantidad de personas en un tren o una zona de la estación hasta la implementación del conteo automático con WiFi); pero en el autobús no ocurre esto. Para intentar solucionar este problema, en el año 2018, TfL desarrolló un programa piloto, probando diferentes técnicas para contar de forma automática los pasajeros.

A pesar del piloto desarrollado en 2018, otros trabajos previos ya se habían desarrollado en esta línea. En 2014, TfL instaló pantallas en la planta inferior de los famosos autobuses de dos niveles, que permitían ver el número y la posición de los asientos libres en la planta de arriba. Para ello, se hizo uso de las cámaras de seguridad del circuito cerrado de televisión (CCTV) ya existente en estos vehículos; a lo que se añadió un software de procesamiento de la imagen, que permitía diferenciar asientos vacíos de aquellos ocupados, y en base a eso, conocer y mostrar en pantalla la información, de forma que el pasajero no tuviera que subir las escaleras y encontrar que todos los sitios estuvieran ocupados. [82] Este sistema, a pesar de ser una buena opción, tenía algunas debilidades, como que podría no identificar a una persona en caso de que esta tuviera colores muy similares a los de los asientos. Además de este precedente, el piloto realizado en el autobús incorporó elementos ya probados en 2016 con las pruebas del conteo de pasajeros en el metro a través de WiFi.

El programa de conteo automático de pasajeros de autobús fue lanzado en junio de 2018, y duró tres meses. Durante este tiempo, siete autobuses en cinco líneas diferentes fueron dotados de diferentes sistemas de conteo de pasajeros; con el objetivo de conocer las diferencias en las mediciones e identificar las mejores maneras de implementar el conteo automático. Al igual que en el caso del conteo automático en estaciones y vehículos del metro, los objetivos del programa eran proveer a los usuarios de información respecto del aforo de los transportes de los que tenían que hacer uso en la interfaz de planificación de rutas; así como mejorar la planificación de futuras líneas e inversiones y minimizar los efectos de posibles emergencias y interrupciones en el servicio. [83]

En total, 5 tipos de tecnologías para contar fueron empleadas para conocer el aforo de los autobuses en tiempo real de forma automática. La primera de ellas consistió en la colocación de cámaras en el suelo en las entradas y salidas del autobús; de forma que estas podían diferenciar, gracias a un software de análisis de imagen, los pasajeros entrando y saliendo del vehículo. La segunda tecnología, en la misma línea que el programa desarrollado en 2014 en los autobuses de dos plantas, consistió en utilizar las imágenes del CCTV ya instalado en los autobuses para, de nuevo usando un software de procesamiento de imagen, contar el número de personas en el vehículo, diferenciando las formas de estos.

La tercera de las tecnologías empleadas fue, al igual que en el metro, el rastreo WiFi. Este método, de nuevo, utilizaba los mensajes de tipo *probe request* del protocolo para conocer las MAC en el rango de actuación; permitiendo saber cuántas de estas son detectadas en el autobús y los lugares en donde suben y bajan al vehículo. Al igual que en el piloto realizado en 2016 en el metro, los datos eran anonimizados y no permitían identificar a cada usuario. No obstante, no existen detalles públicos sobre si el procedimiento para recopilar y procesar los datos era igual al realizado dos años antes o las posibles diferencias entre ambas pruebas. Existen otros estudios relacionados con el conteo de pasajeros en autobuses usando sensores WiFi de bajo coste [84] o apoyándose también en el uso de los datos del GPS [85]; pero no tenemos información sobre el método utilizado en Londres para realizar esta tarea.

Los dos últimos tipos de conteo automático son los más utilizados en otras ciudades del mundo para realizar una estimación de pasajeros a bordo del autobús. Por un lado, tenemos el conteo basado en peso, muy popular debido a que ya se utilizaban sensores de este tipo para medir la carga en las diferentes suspensiones de los vehículos; los sensores existían ya; pero para otro propósito. En este caso, el problema viene dado porque la fuerza aplicada en las suspensiones varía a lo largo de un trayecto, lo que hace que el peso se tenga que medir cuando el vehículo se encuentra detenido antes y después de una parada, comparando el peso del vehículo antes y después de abrir las puertas. A pesar de que no lo podemos saber con seguridad, podemos pensar que este tipo de sensor es el empleado en el programa piloto al que nos referimos, dado que su uso no requiere, en general, de la instalación de sensores nuevos.

Una alternativa a este tipo de medición basada también en el peso son los sistemas WIM, que consisten en la instalación de sensores en la calzada (o en los raíles en el caso de querer conocer el peso en trenes y tranvías) y

permiten, conociendo el peso del autobús y el conductor, su combustible inicial, la distancia recorrida y el consumo previo; conocer la estimación de personas a bordo del mismo a partir del peso medio de un humano. Este sistema puede tener muchas otras aplicaciones, como la medición de la ocupación de una carretera, el pesaje de vehículos en movimiento o mercancías, la distancia entre vehículos consecutivos o la velocidad; y los sensores con los que contará dependerá de la aplicación. En el caso que nos ocupa, el sistema WIM requeriría de sensores de peso; junto con un procesador y software que pueda recibir la información de los sensores y procesarla, alguna forma de comunicación con el exterior para enviar los datos y una fuente de alimentación. Este sistema podría instalarse en las paradas o en otras zonas del trayecto del autobús, permitiendo la medición del peso a lo largo de la ruta. Como estos sensores se encuentran directamente en la calzada, su instalación implica un mayor coste que en el caso anterior, aunque las aplicaciones son mucho más diversas. [86]

La última de las tecnologías utilizadas en el programa piloto fue el uso de sensores de conteo de pasajeros instalados en las puertas. De nuevo, no se detalla qué tipo de sensores fueron empleados durante el programa, así que más adelante se detallan los diferentes y más usados en la industria actualmente:

- **Sensores infrarrojos:** Los de tipo activo funcionan con un transmisor y receptor colocados enfrentados actuando como un interruptor, de forma que cuando se corta el haz de luz se realiza la detección; o con un reflector de forma que tanto transmisor como receptor se encuentran juntos. En general, se necesita más de un dispositivo a dos distancias diferentes de la puerta, y se colocan de forma que los rayos viajan paralelos al suelo; de forma que se puede conocer si el pasajero está entrando o saliendo según qué interruptor se activa primero. Además, se pueden instalar a varias alturas para tener más seguridad en la medición; actuando como una “barrera”. Los sensores de tipo pasivo, por otro lado, crean un espacio de detección que permite, a través de los cambios en los reflejos de la luz captados por el sensor generados por los diferentes objetos en la zona y sus movimientos, conocer si están entrando o saliendo los pasajeros. [87] En el caso de la figura 13, podemos ver cómo a partir del movimiento del usuario se genera un tipo de onda detectada por el sensor, que permite conocer la dirección en la que está caminando.

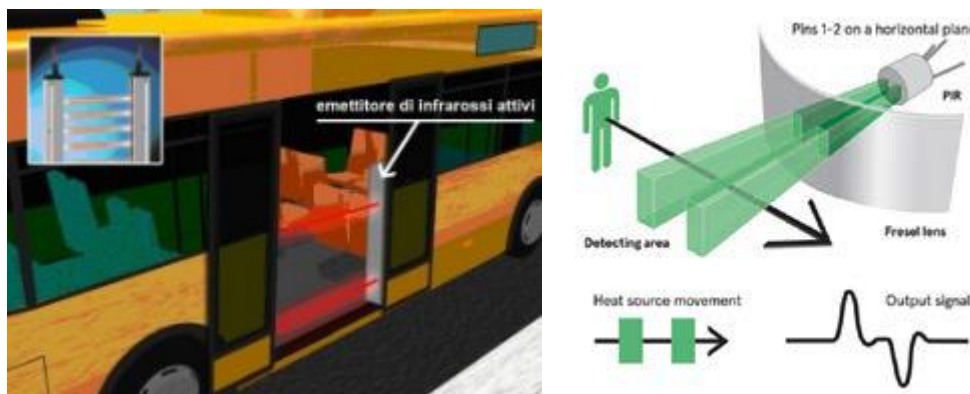


Figura 13. Ejemplo de uso de sensores infrarrojos activos y pasivos. [88] [89]

- **Alfombra contadora:** Es una estructura de metal cubierta de una capa de goma que se instala en las entradas o salidas; en definitiva, en lugares de paso de personas, que activan un botón al pisarse. Dado que el botón se encuentra en el centro, los extremos de este dispositivo no permiten detectar a los pasajeros, lo que hace que su colocación sea crítica para su correcto funcionamiento. Además de esta limitación, cuando los transportes están muy llenos de personas, el sensor puede ser pulsado debido a pasajeros que se encuentran de pie en el vehículo; por lo que no es el método óptimo si se prevé un gran aforo. Su instalación es más compleja y costosa; pero no requiere mantenimiento y la precisión trabajando en condiciones óptimas es mayor que utilizando sensores infrarrojos. [90]
- **Sensores ópticos OPS:** Son dispositivos más complejos que permiten distinguir el movimiento de los pasajeros al entrar y salir por las puertas de los vehículos; guardando información relativa a la fecha y la hora. Permiten una mayor precisión, así como mayor simplicidad en la instalación y el mantenimiento. Cuentan con una fuente de alimentación y una interfaz de comunicación con el exterior, un microcontrolador para procesar los datos y los diferentes sensores; que dependen del tipo de sensor y el modelo. En general, cuentan con LEDs y cámaras estereoscópicas que permiten la identificación de volúmenes; gracias a las cuales pueden realizar el conteo de pasajeros. [88]



A diferencia de otros programas anteriores desarrollados por TfL, en este caso no existen informes oficiales sobre los resultados obtenidos y las conclusiones extraídas del estudio. Después de casi 4 años desde la finalización del programa piloto, el sistema de conteo automático de pasajeros en autobús no se ha implementado de forma generalizada en la red, no sabemos si porque los datos obtenidos siguen siendo analizados o porque el programa no arrojó los resultados esperados. A pesar de ello, el conteo automático y con precisión de pasajeros tiene muchos beneficios para los sistemas de transportes que lo implementen. Además de los objetivos de TfL de conocer en tiempo real la ocupación y poder ofrecerlo a los usuarios, la planificación de nuevas líneas y optimización de las existentes o la posibilidad de atajar emergencias con mayor efectividad; otros beneficios ya analizados en el caso del metro como el reajuste de los precios de los espacios publicitarios en función de la visibilidad de los mismos se podrían obtener gracias a este sistema. Por último, el análisis de los pasajeros en un vehículo en contraposición con aquellos que han pagado el *ticket* resulta muy útil en los sistemas de transporte donde no existe vigilancia al pagar, y podría permitir conocer el porcentaje de usuarios que se están montando en el autobús o tranvía de forma irregular y, en su caso, reformar el sistema de pagos en los vehículos.

#### 4.1.2 TfL y Cloud Computing

Desde el año 2007 que TfL comenzó a publicar los datos recopilados a través del transporte público para todo el mundo, dando comienzo a toda la transformación digital que ha desarrollado desde entonces, son muchos los cambios y las nuevas tecnologías disponibles para trabajar y procesar grandes cantidades de datos. Sin ir más lejos, la fundación de Amazon Web Services (AWS) se produjo en 2006; mientras que el lanzamiento de Google Cloud en 2008. Desde entonces, una gran cantidad de nuevos servicios están disponibles, lo que ha hecho que los diferentes servicios requeridos por Transport for London hayan experimentado grandes cambios a lo largo de su historia.

Para la empresa de transportes, la clave de su estrategia digital siempre ha sido la voluntad de hacer públicos todos los datos posibles para todo el mundo, dado su compromiso con la transparencia y el favorecimiento de la investigación y el emprendimiento para que los ciudadanos puedan beneficiarse de productos útiles para ellos. Desde el lanzamiento del proyecto de digitalización en 2007, TfL desarrolló una red de distribución de contenidos (CDN) para facilitar el acceso a los datos y a su web. El principal problema de este tipo de red es que la escalabilidad era nula; teniendo que estar sobre capacitada con el objetivo de poder lidiar con picos de demanda, que podían llegar a multiplicarse por 20 en días de nieve, colapsando el sistema. [91] Además de ello, la necesidad de contar con más recursos de los requeridos normalmente implicaba un coste mucho mayor.

Además de la escalabilidad, con el paso del tiempo una gran cantidad de requisitos nuevos comenzaron a surgir, como el manejo y procesamiento de grandes cantidades de datos con el objetivo de proveer información de calidad en tiempo real y personalizada según las necesidades de los usuarios. Con este objetivo, en febrero de 2012 la empresa comenzó a buscar alternativas que le llevaron a las soluciones en la nube ofrecidas por AWS; completando la migración en 2014. [91]

La opción de mover sus diferentes datos y servicios a la nube era un paso lógico gracias a la gran cantidad de beneficios que esto comportaba. Además de la escalabilidad requerida, que era el problema principal, otras características como la gran resistencia a los fallos y disponibilidad, la velocidad en la transmisión y el procesamiento de los datos, la flexibilidad ofrecida para trabajar con esta y los reducidos costes de operación y mantenimiento aportaban gran valor a TfL, no sólo económico sino también funcional. Además, las necesidades de los ciudadanos de recibir información en tiempo real y personalizada podían ser atajadas con estas características. [92] Dentro de las diferentes opciones en el mercado, donde las predominantes son Microsoft Azure y Google Cloud Platform, además de Amazon Web Services, la decisión de TfL de optar por esta última estuvo basada en diferentes elementos. En primer lugar, su gran flexibilidad y compatibilidad con diferentes tecnologías y sistemas operativos que estaban siendo usados hasta la fecha en los diferentes departamentos de la empresa, la seguridad ofrecida en el almacenamiento de los datos, la elasticidad y escalabilidad prácticamente sin un límite superior de capacidad y unos precios más competitivos que la competencia basados en el modelo pago por uso; un gran beneficio respecto de la CDN anterior, que también tenía unos costes de mantenimiento asociados. [7]

Gracias a la asociación con AWS la nueva web de TfL pudo ser lanzada. Esta web contaba con muchas de las necesidades requeridas por los usuarios, como la adaptabilidad de la web según el dispositivo (necesaria debido al auge de los dispositivos móviles), la personalización y el uso de la localización para facilitar la planificación de viajes. Esta nueva web contaba también con el primer planificador multimodal de la compañía, permitiendo



ver trayectos que combinasen diferentes medios de transporte, como el metro, autobús, tranvías y ferris; y también permitía consultar en tiempo real el estado de todas las líneas de los medios de transporte; características que se han mantenido hasta nuestros días. El uso de AWS permitió realizar predicciones en tiempo real del tiempo hasta la llegada de un vehículo con 10 segundos de margen y 0 de latencia; realizando más de 100.000 predicciones cada 30 segundos. Durante los primeros meses con el servicio funcionando completamente en AWS, la API unificada de TfL fue lanzada, de forma que los datos eran accesibles de una forma más sencilla para los desarrolladores; dado que anteriormente estos datos debían ser accedidos a través de canales diferentes según los datos a los que se quisiera acceder. Además, gracias al lanzamiento de esta tecnología se lanzaron también los primeros conjuntos de datos en tiempo real, accesibles también a partir de la API unificada. [93] Los diferentes datos que TfL maneja y distribuye también se encuentran en los diferentes servidores de AWS. Además, en 2016 la empresa pública de transportes se asoció con Preservica, una empresa de preservación de documentos, con el objetivo de almacenar los más de 140.000 archivos de los que dispone, digitales y físicos digitalizados, en la nube. Estos documentos contienen información relativa al transporte de Londres desde 1857, incluyendo documentos sobre la Segunda Guerra Mundial o los diferentes Juegos Olímpicos celebrados en la capital británica, mapas, información sobre trabajadores, modelados o investigaciones. [94]

El traslado de los servicios a los servidores de Amazon también ha permitido que TfL despliegue sus servicios realizando el modelo de despliegue azul/verde; gracias al cual el servicio no debe interrumpirse al lanzar un nuevo producto o funcionalidad en la web. En este tipo de despliegue, el tráfico se va dirigiendo poco a poco desde la versión anterior, denominada azul; a la nueva, conocida como verde; parecida a la anterior, pero con el nuevo servicio ya implementado, gracias al uso de un balanceador de cargas. Además del beneficio obvio de no necesitar realizar un mantenimiento, también permite que las dos versiones cohabiten para que, si en algún momento se detecta algún fallo, se pueda redirigir todo el tráfico a la versión anterior. Otro beneficio directo del uso de AWS es que ha permitido trabajar con un entorno en arcoíris, pasando de 4 a 19 entornos diferentes de los servicios para que los departamentos y grupos dentro de la empresa puedan desarrollar diferentes funcionalidades y enfocarse en diferentes etapas del desarrollo de estas, sin que los diferentes proyectos se limiten entre ellos, y permitiendo ser desarrollados en diferentes sistemas operativos y sólo pagando por los momentos en los que se está trabajando, gracias al modelo de pago por uso de AWS. [93]

Gracias al salto de calidad dado por los diferentes servicios digitales de TfL al trasladarlos a la nube, durante los 6 primeros meses desde el lanzamiento oficial en 2014 el planificador de viajes de TfL aumentó un 10%, el uso en dispositivos móviles aumentó superando al realizado en ordenadores y la satisfacción de los usuarios creció hasta el 87%, con el 76% de los londinenses utilizando la web oficial para planificar sus trayectos en transporte público. [92]

A pesar de que el ecosistema de Amazon Web Services es el predominante dentro de la compañía, entre 2010 y 2014, los diferentes datos recopilados por TfL estuvieron almacenados en los servicios en la nube de Microsoft Azure. Posteriormente estos datos se movieron a AWS. A pesar de ello, el lanzamiento del pago sin contacto en el metro directamente con tarjeta de crédito fue posible gracias a Microsoft SQL Server, un servicio dentro de Azure. [95] Actualmente, Azure sigue siendo utilizado en algunos departamentos de la compañía, y las herramientas de trabajo de Microsoft son predominantes en TfL. Algunos ejemplos son aplicaciones como Microsoft Teams, o, relacionados con el trabajo, con datos, PowerBI, que permite visualizar y comunicar estadísticas de forma cómoda, tanto para uso interno como con los ciudadanos.

### 4.1.3 Usos de los datos recopilados

A pesar de que algunos ya se han mencionado anteriormente, la implementación de las diferentes tecnologías de recopilación y procesamiento de información ha generado una gran cantidad de usos prácticos en el día a día, que crean beneficios no sólo para los ciudadanos, sino también para la propia empresa. Además de eso, la publicación de todos los datos de forma pública y gratuita a través de una API unificada hace que trabajar con la información de TfL sea sencillo, lo que ha permitido el desarrollo de una gran cantidad de aplicaciones que usan estos datos y sirven a los ciudadanos para conocer información sobre los diferentes servicios de TfL por canales no oficiales. Aparte de los desarrolladores, los datos pueden ser utilizados por diferentes estudios que tengan como objetivo analizar el sistema de transportes de Londres; o incluso, a partir de los datos y actuaciones de la capital británica, intentar mejorar sistemas de transportes de otras ciudades.

#### 4.1.3.1 Para los usuarios

- Conocer en tiempo real cuánto le falta a un determinado vehículo para llegar a una estación, tanto en señales en estaciones y algunas paradas como en la web, así como su ubicación.
- Conocer el aforo de una estación, y cómo este puede influir al realizar un transbordo de un lugar a otro, modificando el tiempo en realizar un traslado de un andén a otro.
- Ofrecer a los usuarios de TfL diferentes opciones de rutas entre dos lugares, combinando diferentes medios de transporte y ofreciendo una cantidad de opciones personalizables para emplear sólo algunos medios de transporte, andar o utilizar la bicicleta una determinada distancia o hacer uso de otros servicios de TfL. Ofrecer otra información útil referida a los trayectos, como la accesibilidad o el aforo a las estaciones.
- Ofrecer una cuenta de usuario única para todos los servicios de TfL que permita consultar el balance de la Oyster Card después de un viaje, identificar usuarios habituales de determinados servicios y ofrecer un servicio personalizado, con notificaciones de interés y que no resulte invasivo; y ofrecer descuentos personalizados en caso de que estos sean aplicables.
- Avisar de posibles retrasos o emergencias en las rutas elegidas y ofrecer recomendaciones y alternativas; así como información y actualizaciones en tiempo real.
- Una gran variedad de aplicaciones, tanto oficiales como de terceros, que permiten la existencia de servicios de nicho que no serían posibles si los datos no fueran públicos.
- Recibir anuncios personalizados en función de las rutas que el usuario utiliza.
- En el pasado, TfL desarrolló un programa piloto junto con Twitter, creando cuentas para los distintos servicios que enviaban alertas con posibles cambios en el servicio a la bandeja de mensajes privados de los usuarios que lo solicitasen. Este programa luego no fue implementado de forma oficial, aunque existen cuentas no oficiales que realizan una labor parecida. [96] De la misma forma, en el pasado un *bot* de Facebook estuvo disponible para realizar consultas a través de Facebook Messenger, pudiendo solicitar el estado de las líneas, el tiempo estimado de llegada de un transporte a una parada o pedir mapas con las líneas disponibles. Este servicio tampoco se encuentra ya operativo.

#### 4.1.3.2 Para la empresa

- Mejorar el uso de las vías de atención al cliente, analizando las quejas de forma más eficiente y permitiendo formar al personal con el objetivo de mejorar la relación con el público.
- Reducir la presión de las oficinas de atención al cliente, ofreciendo información fiable, accesible de forma sencilla para todo el mundo.
- Conocer en tiempo real emergencias, picos de demanda y posibles interrupciones del servicio, para que el personal de TfL pueda reaccionar a ellos en el menor tiempo posible, así como entender las reacciones de los usuarios ante estas circunstancias y sus alternativas preferidas en caso de suspensiones de servicio, para poder reforzarlas.
- Conocer las líneas, rutas y tramos más utilizados, encontrar las necesidades del sistema en cuanto a horarios de los transportes y actuar si fuera necesario; conocer opciones preferidas para viajar a lugares con muchas alternativas de transporte y reforzarlas; y planificar de mejor forma eventos multitudinarios como grandes conciertos y partidos de fútbol.
- Entender los movimientos de los usuarios dentro de las estaciones, permitiendo aumentar la seguridad, conocer las necesidades de estas y desplegando el personal de atención en lugares estratégicos para una mejor atención a los clientes.
- Mejoras en el diseño de las estaciones e inversión del dinero en aquellos lugares y formas en que sea más necesario, a partir del análisis de los movimientos de los usuarios en el interior de estas.
- Mejor venta de los espacios publicitarios, adaptando los precios según la visibilidad real de los mismos y permitiendo el aumento de los beneficios de la empresa por publicidad.

#### 4.1.3.2.1 Mantenimiento predictivo

Los anteriores usos han sido explicados a lo largo de este documento gracias a los diferentes sistemas implementados y programas piloto realizados por TfL. Sin embargo, el mantenimiento predictivo es un uso de los datos crítico para la compañía y que no ha sido desarrollado todavía.

El uso de los datos con el objetivo de predecir si un determinado sistema va a fallar o no es un punto clave en la estrategia digital de TfL, dado que los gastos asociados a los fallos en estos y su mantenimiento ocupan el 59% del presupuesto de la compañía y más de la mitad de los retrasos son debidos a estos. [97] Los fallos no incluyen únicamente los vehículos, sino que también pueden ser referidos a estaciones o carreteras. Si estos gastos pudieran reducirse gracias al uso de todos los datos recopilados, la situación económica de TfL mejoraría enormemente, así como la satisfacción de los usuarios con el servicio. Con el objetivo de evitar los fallos, se podrían realizar tareas de mantenimiento rutinario, pero esto supondría gastos asociados a realizar más operaciones de revisión de las necesarias. Para que estos algoritmos funcionen, no obstante, es necesario contar con un gran sistema de monitorización y transmisión en tiempo real de la operación de los vehículos que, combinado con otras fuentes de información, permita trabajar con esos datos. [98]

Diferentes proyectos relacionados con el análisis de datos se están llevando a cabo en paralelo para desarrollar algoritmos que puedan predecir con alta precisión cuándo un vehículo puede fallar. Los diferentes algoritmos se encargan de analizar el historial de fallos de los vehículos, con el objetivo de encontrar patrones y factores de riesgo que puedan hacer pensar que el sistema puede fallar. Esto los ha llevado a tener en cuenta cuatro factores fundamentales, que son el tipo de vehículo, el historial de fallos y reparaciones realizadas, la operación del vehículo (elementos como el tiempo de uso, la estación de partida del vehículo, o posibles comportamientos del vehículo durante su operación) y factores externos, donde destaca el tiempo atmosférico. Además de determinar cuándo puede fallar un vehículo, el algoritmo intenta determinar el tipo de fallo, el impacto que puede tener y los costes asociados. [99]

El objetivo último de TfL al realizar estos análisis es la de reducir los costes operativos, migrando de un sistema de reparaciones basado en los fallos a uno completamente predictivo; creando calendarios de reparaciones para cada vehículo basados en el riesgo de que estos fallen y los costes asociados al fallo, y reducir el número de fallos aumentando la satisfacción, eficiencia y seguridad de los clientes al usar el transporte público. La predicción de los errores también permite a TfL fijar unos porcentajes asumibles de fallos y costes durante el servicio. [98]

Esta estrategia se enmarca en un plan de gestión de los servicios lanzado en 2016 por TfL, que no se queda en el plano de la gestión del transporte público sino también de carreteras o calles. Esta estrategia incluye, en primer lugar, la reparación de muchos activos y la renovación de otros, así como la mejora y adaptación de sistemas antiguos a las necesidades actuales. Una vez la flota de vehículos se encuentre actualizada, es necesaria la monitorización de los mismos, conseguida gracias a los dispositivos de recolección de la información desplegados a lo largo de la red y el procesamiento de los mismos; desarrollando los algoritmos previamente explicados para realizar mantenimientos predictivos basados en el riesgo de fallo. Estos algoritmos se encuentran en continuo desarrollo, con el objetivo de mejorarlos y refinarlos para que los costes operativos sean cada vez más bajos. [98]

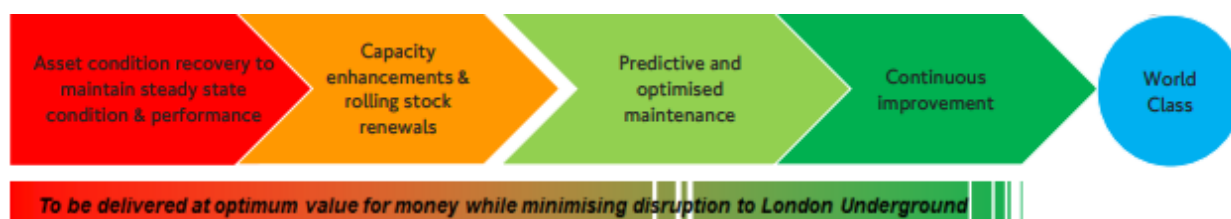


Figura 14. Estrategia de gestión de los activos de TfL. [98]

#### 4.1.4 Beneficios de la estrategia digital de Transport for London

La estrategia digital desarrollada por TfL desde el año 2007 ha traído una gran cantidad de beneficios a la ciudad de Londres. La publicación de todos los datos de forma gratuita y su fácil acceso a través de la API unificada ha permitido crear un círculo virtuoso donde todos los elementos que intervienen en la cadena de servicios del transporte se benefician y retroalimentan. Este ciclo comienza con los datos públicos y de libre acceso, gracias

al cual grandes empresas como Google o Waze pueden hacer uso de la información y mejorar la información que se le da a los usuarios. Además, pequeños desarrolladores también pueden crear aplicaciones más específicas o que resuelvan problemas para segmentos más pequeños de la población. Esto repercute directamente en los ciudadanos, que pueden hacer uso de estas aplicaciones. Al facilitar el acceso a la información sobre el transporte público, más gente lo utilizará, y la satisfacción de los usuarios será mayor. Esto supone un beneficio económico directo para TfL, que puede destinar ese dinero a mejorar otros servicios, aumentar la cantidad de datos recopilados y que nuevos conjuntos estén disponibles para los desarrolladores. [63]

Un estudio llevado a cabo en 2017 por Deloitte [63] afirma que la estrategia digital de TfL y el uso de datos públicos generaba un ahorro económico de entre 90 y 130 millones de libras al año; cifra que podemos pensar que ha aumentado en los 5 años que han transcurrido. Los ahorros debidos a la cancelación del sistema SMS se refieren al ahorro en los bolsillos de los ciudadanos, que debían pagar para recibir alertas mediante SMS con anterioridad a creación de las alertas en la aplicación. Además, se observa cómo los beneficios se reparten entre los diferentes actores de la ciudad, repercutiendo positivamente tanto en la empresa, los ciudadanos, como en Londres de forma general.

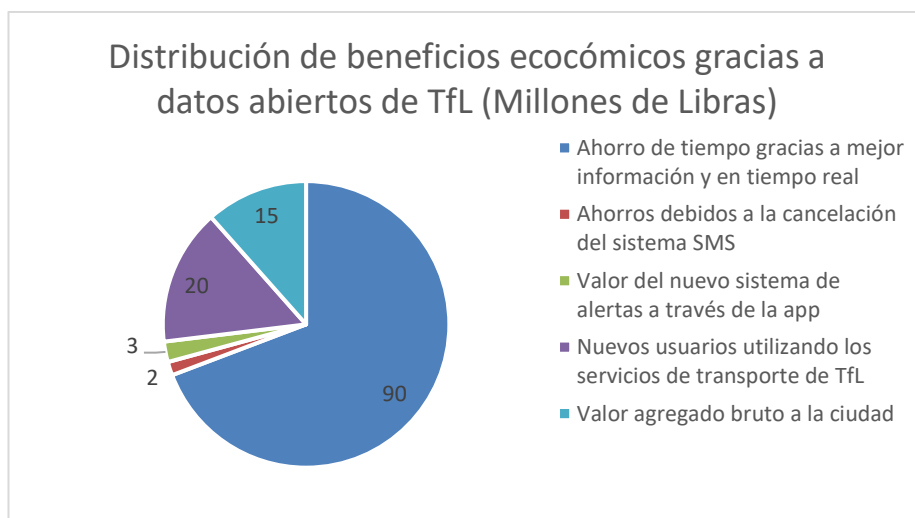


Figura 15. Gráfico de la distribución de beneficios estimados gracias a la estrategia digital de TfL.

Además de los beneficios económicos directos, existen muchos otros beneficios de forma indirecta. En el mismo estudio, se afirma que más de 500 puestos de trabajo directos se han creado gracias al proceso de digitalización de TfL, además de otros empleos indirectos. Otros beneficios indirectos para TfL incluyen la reducción de las llamadas a atención al cliente, y gracias a los más de 13000 desarrolladores registrados y las más de 600 aplicaciones que hacen uso de los datos han permitido un ahorro indirecto a TfL, dado la empresa misma no ha tenido que desarrollar ni tiene que operar estos programas y sólo ha colaborado puntualmente con algunos trabajos. Además, los datos recopilados por los desarrolladores de las aplicaciones de terceros son enviados a TfL, lo que permite a la empresa darles uso o beneficiarse de ellos.

Además de los beneficios para TfL, los ciudadanos de Londres han experimentado una mejoría en los servicios y la información comunicada por parte de la empresa pública de transportes. Esto ha repercutido de forma directa en el desarrollo de un estilo de vida más saludable para los londinenses, con una mejor red de carriles bici y bicicletas públicas y el soporte de rutas caminando combinadas con el transporte público en el planificador de viajes de la web y la aplicación de TfL. Además, el aumento del uso del transporte público y la progresiva disminución en el uso de coches privados ha supuesto la reducción de la emisión de gases contaminantes, que no sólo afecta de forma positiva a la salud de los ciudadanos, sino también a la situación del planeta.

Por último, el desarrollo de una estrategia digital tan grande, que no incluye sólo a TfL sino a todas las áreas de la ciudad de Londres, hace que esta ciudad se haya convertido en una referencia a nivel de transporte inteligente y en el uso del Big Data para mejorar la ciudad. La alcaldía de Londres aspira a fomentar el talento y las habilidades digitales, utilizar estos para mejorar los servicios públicos, explorar nuevas alianzas en el sector tecnológico, liderar la transformación digital en todo el Reino Unido y, en definitiva, crear en Londres un ecosistema de excelencia tecnológica; lo que actualmente y en futuro supondrá un aliciente para la investigación y la creación de empleo y empresas. [100]

## 4.2. El caso de Sevilla

La ciudad de Sevilla cuenta con una red de transportes donde cada servicio está gestionado por una empresa distinta. Por un lado, los servicios de autobuses urbanos y tranvía son gestionados por TUSSAM, las líneas de autobús interurbanas por el Consorcio de Transportes de Andalucía (CTAN), los trenes de cercanías por RENFE, el sistema de bicicletas públicas por SEVICI [101] y el metro por un consorcio de empresas, donde la mayoría la tiene la empresa Globalvia. [102] Este fenómeno genera dos problemas principales. El primero de ellos se encuentra asociado a la inexistencia de una integración tarifaria entre los servicios, que tampoco pueden pagarse haciendo uso del mismo método de pago; y el segundo problema es un desequilibrio entre los servicios prestados por las diferentes organizaciones, al igual que una gran diferencia entre las prestaciones y características de los diferentes medios de transporte. Este último factor hace que el grado de integración de los transportes públicos con las diferentes tecnologías de Big Data sea diferente según el caso, y también hace la información mucho más inaccesible para los usuarios; tanto para la planificación de viajes, al no existir una aplicación oficial, pública e integrada con todos los medios para ello, como para el acceso de desarrolladores a la información a través de los diferentes medios oficiales. Las diferentes organizaciones mencionadas gestionan 45 líneas de bus urbano diurnas, a las que se suman 10 nocturnas y un tranvía; una línea de metro; una red de autobuses metropolitanos que conectan la capital con las localidades cercanas [101] y la tercera flota de bicicletas públicas más grande del país, a lo que se suma la red de carriles bici con más kilómetros por habitante de España. [103]

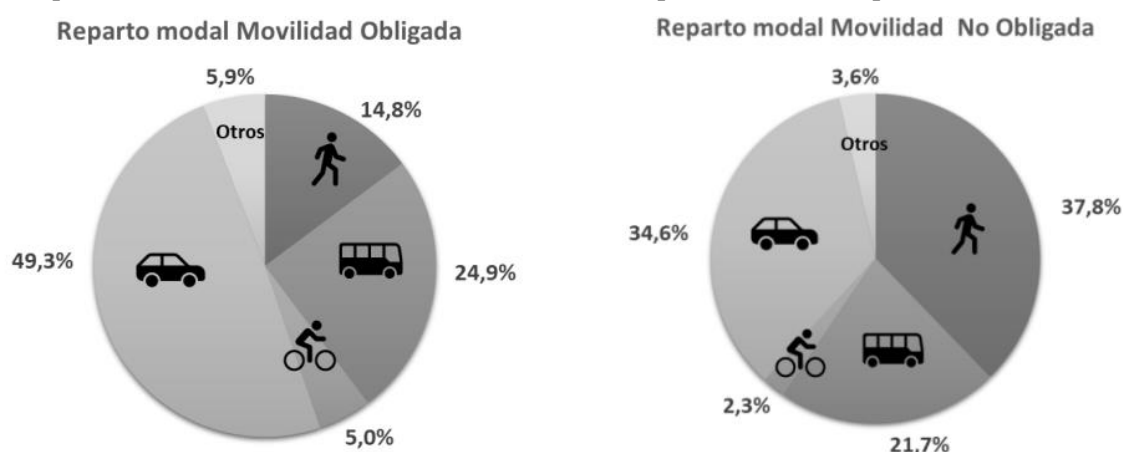


Figura 16. Reparto modal de movilidad en Sevilla según la obligatoriedad del desplazamiento. [104]

A pesar de estos servicios disponibles a la ciudadanía, entre las grandes ciudades de nuestro país es en Sevilla donde mayor uso se hace del coche, existiendo un 53% de la población que sufre al menos un atasco a la semana y a la dificultad de encontrar aparcamiento gratuito en la ciudad. [105] No podemos encontrar un motivo concreto para este suceso, pero sí podemos ver cómo una encuesta elaborada en 2017 revelaba que más del 80% de los sevillanos consideraban los precios de los autobuses caros o muy caros; y casi la mitad de la población consideraba la red de autobuses el principal problema de la ciudad. [106]

Dejando de lado la percepción del excesivo precio de los títulos de transporte y la inexistente intermodalidad, el 60% de las líneas urbanas ofrecidas por TUSSAM son radiales, [101] lo que hace que los viajes entre lugares alejados del centro de la ciudad sean mucho menos eficientes en comparación con el uso de vehículos privados, siendo el tiempo medio de viaje en autobús de 27.3 minutos frente a los 21.3 usando transportes privados. [104] La ineficiencia de los autobuses urbanos, además de por la tipología de la red, se debe a la existencia de pocos carriles exclusivos para autobuses; que componen el 12.7% de la red frente al 18.4% en Barcelona o el 21.1% en Valencia, [107] y el mal uso que el resto de ciudadanos hacen de estos, aparcando y circulando por ellos; además de contar con rutas excesivamente largas y con una gran cantidad de paradas. [104] Estos motivos hacen que la velocidad comercial del servicio de TUSSAM sea de 13.1km/h, por debajo de la media nacional de 13.8km/h. [101]

Además de lo comentado respecto del autobús, el estado del servicio del resto de transportes públicos desalienta de igual forma el uso de estos. Por un lado, el metro de Sevilla se encuentra en un estado de paralización absoluta, habiéndose ejecutado sólo 1 de las 4 líneas planificadas a lo que hay que sumar la peor relación de habitantes por kilómetro del país y el menor número de estaciones de metro de España. Desde su inauguración, en 2009, no ha recibido ningún tipo de actualización. [108] El servicio de cercanías, conformado por 5 líneas, tiene unos

precios desorbitados en comparación con el resto de transportes, además de no existir una integración con el resto de transportes, un número reducido de estaciones y una muy baja cadencia en las diferentes líneas. [101] Por último, el uso de autobuses metropolitanos se encuentra en una tendencia de uso descendente desde 2003, lo que se debe a la poca fiabilidad en los tiempos de paradas, la baja velocidad comercial e inexistencia de carriles exclusivos y a la poca conectividad que tiene con otros servicios como el metro. [101] Además, parte de las líneas ofrecidas por esta modalidad coinciden con la única línea de metro disponible.

Diversos métodos, desarrollados en el Plan de Movilidad Urbana Sostenible del Ayuntamiento de Sevilla se han propuesto con el objetivo de mejorar los diferentes servicios de transporte de la ciudad, buscando una transformación completa para el año 2030. Este plan, que no se circunscribe al transporte público, sino a todos los asuntos relacionados con la movilidad, incluyendo el tráfico y la movilidad peatonal, propone una reorganización completa de las paradas y una reordenación total de las líneas de autobús; el desarrollo de una red provisional de autobuses de alta velocidad cubriendo el recorrido del metro, una mejor conexión con los autobuses metropolitanos y la integración de todos los servicios bajo un único sistema de transportes. [104]

Como ya se ha comentado, además de existir un desequilibrio en precios y en la integración de servicios, el hecho de que cada una de las empresas gestione vehículos y tipos de transporte de forma independiente, hace que la integración con tecnologías de recopilación de información, uso e información se encuentren en diferentes estados. Además de eso, el intercambio y la integración de los datos recopilados por los diferentes operadores en plataformas únicas gestionadas por entidades públicas es inexistente, por lo que la planificación de viajes intermodales depende enteramente de empresas privadas que desarrollen sus propias aplicaciones para ello. En el plan descrito anteriormente se detallan algunas mejoras en esta línea, y algunos programas piloto y pequeñas mejoras han sido ya integradas, como la prioridad semafórica en los autobuses o el control de aforo. [104]

A continuación, se analizarán los diferentes mecanismos relacionados con las tecnologías de la información empleados en los diferentes servicios, con el objetivo de conocer qué elementos se están utilizando ya en el sistema de transportes públicos de Sevilla, así como conocer qué margen de mejora existe y cómo se podrían utilizar mejor los elementos que ya se encuentran en la actualidad instalados en los vehículos y estaciones.

#### **4.2.1 El servicio de autobús urbano, TUSSAM**

El servicio de autobuses urbanos se encuentra gestionado por la empresa pública TUSSAM, fundada en 1975. Esta empresa ha sido pionera en muchas ocasiones, siendo una de las primeras en incluir un bonobús en sus tarifas e instalando un SAE por primera vez en España, en 1982 y 1986 respectivamente. [109] A pesar de ello, la operación de TUSSAM actualmente ha quedado algo obsoleta, aprovechando tan sólo parcialmente algunos sistemas instalados en los vehículos, existiendo una plataforma de comunicación con los clientes con algunos fallos y sin intercambiar información con las principales aplicaciones de navegación.

##### **4.2.1.1 Sistemas de Ayuda a la Explotación (SAE) y prioridad semafórica**

Conocemos como SAE al conjunto de dispositivos y sistemas que permiten a las empresas conocer y controlar los recursos disponibles en una red de transporte público en tiempo real. Estos dispositivos, instalados en todos los vehículos, están formados por una serie de elementos que sirven para recopilar información; donde destaca un GPS, junto con tecnologías que permiten comunicar la información recopilada a un centro de control para que la información pueda ser utilizada allí y distribuida al resto de los ciudadanos en la forma que mejor se considere conveniente. [110]

El SAE empleado por TUSSAM cuenta con un sensor GPS y con un odómetro, que permite conocer la posición del vehículo en tiempo real, con un margen de 25 metros cada segundo, [111] y de sistemas de radio que permiten comunicar al centro de control la posición. Además de eso, cuenta con un ordenador de a bordo, que permite al conductor conocer información relevante sobre el estado del autobús, como si existe algún problema o avería y gestionarla; enviar y recibir mensajes a la central y llamar a un superior, o conocer el tiempo de retraso o adelanto que tiene respecto del horario teórico para cambiar su forma de conducir según lo que se requiera en el momento. Gracias al SAE, además, la información de próxima parada puede ser reproducida en el autobús y el tiempo estimado hasta las siguientes paradas de la línea puede ser calculado y comunicado a los usuarios del transporte público. [112]

Una opción muy útil para aumentar la velocidad comercial de los autobuses es el uso de la prioridad semafórica

en los cruces de la ciudad. En el caso de TUSAM, en 2009 fueron introducidos los primeros sistemas de este tipo en 10 intersecciones, 4 dedicadas a los autobuses y 6 para el tranvía. [107] Posteriormente, en 2016, la empresa lanzó tres nuevas líneas de alta velocidad conectando el centro de Sevilla con las zonas este y norte de la ciudad, y con la Isla de La Cartuja, aunque esta última línea fue cancelada posteriormente. [113] Estas nuevas líneas también contaban con este sistema de prioridad, de forma que el tiempo perdido por paradas en semáforos se redujese al máximo. [111]

El funcionamiento de estos semáforos prioritarios se consigue gracias al GPS instalado en el SAE del autobús y sus posibilidades de comunicación. Cuando el autobús se sitúa a 200 metros del cruce, se realiza una petición al sistema central, solicitando la prioridad. Una vez confirmada esa petición, se produce una transición, de forma que si el cruce se encontraba en verde se mantiene de este modo hasta que el autobús comience a alejarse; pero si estaba en rojo deben cambiarse a ámbar y rojo los otros semáforos de la intersección, y poner en rojo los pasos de peatones si los hubiera. Una vez realizada esta transición, se pone en verde el semáforo para nuestro autobús, permitiendo que no se detenga en el cruce. [111]

#### 4.2.1.2 Sistema de información audiovisual

El sistema de información audiovisual empleado por TUSAM, denominado BUSI, consiste en un ordenador, conectado a una serie de altavoces colocados en distintos lugares del vehículo junto con varias pantallas (de 1 a 3 dependiendo de la longitud del autobús) que sirven para reproducir de forma cíclica información de actualidad, como noticias, eventos deportivos y culturales, efemérides o preguntas de cultura general, que se actualiza cada día. Además de esto, en las diferentes pantallas se anuncia la proximidad de la siguiente parada, junto con un mapa de la ruta para conocer las próximas a esta; así como anuncios relevantes para los pasajeros, como una avería, un desvío, un final alternativo de la ruta, una parada más larga, etc. [114] El conductor, al igual que una consola para controlar el SAE, cuenta con una consola destinada al control de BUSI, reproduciendo los mensajes determinados en caso de necesidad. [112]

En paralelo a toda esta información, el canal sirve también como un espacio publicitario que permite a los anunciantes reproducir anuncios de forma cíclica en las pantallas, siendo un canal que puede alcanzar una gran cantidad de impresiones, y que permite la localización de los anuncios en rutas determinadas, de forma que los comerciales sean relevantes para los usuarios que los están visualizando y generen un alto impacto en los viajeros. [115]

#### 4.2.1.3 Planificación de viajes y comunicación con los usuarios

La compañía TUSAM, al igual que se revisará posteriormente con el Metro, tiene dos portales oficiales a través de los cuales puede consultarse información en tiempo real sobre el estado de los diferentes servicios. El primero de los dos canales es la web oficial, en la que pueden consultarse noticias y avisos de última hora sobre eventos relacionados con la red que puedan afectar a las rutas; además de otras noticias relacionadas con la compañía. Gracias a esta sección, si un servicio debe suspenderse o modificarse, los interesados pueden conocerlo al instante. [116] Además de esto, en la web existe una funcionalidad determinada “Red de Líneas” a través de la cual se puede consultar información en tiempo real sobre las diferentes líneas a las que TUSAM da servicio, los vehículos activos y su ubicación, la información de próximas llegadas en cada parada y los horarios planificados de las líneas, útil para conocer los horarios iniciales y finales del día y los tiempos de partida de los autobuses desde las cabeceras. [117] Además de esta información, útil para los viajeros en el momento de viajar en autobús o tranvía por la ciudad, es posible recargar un bonobús a través de la web, realizando el pago con tarjeta de crédito o Bizum; [118] y también en la web se puede encontrar más información relativa a las operaciones de la empresa; como tarifas, balances, políticas, etc. [116]

La otra de las opciones que pone TUSAM a disposición de la ciudadanía para la planificación de viajes es la aplicación oficial, que cuenta con más funcionalidades útiles para ello. Además de conocer el estado de las líneas, sus paradas, los horarios planificados y en tiempo real, se pueden crear alertas para recibir una notificación cuando un autobús de una línea se encuentre a un tiempo determinado de una parada elegida; permitiendo que estas alarmas sólo funcionen en algunos días de la semana y a unas horas concretas. También a través de la aplicación es posible consultar el saldo restante en tarjetas de TUSAM y recargarlas con los mismos métodos de pago que en la web. [118] En la aplicación también encontramos una sección denominada “Seguimiento de viaje” pero no se encuentra en funcionamiento.



Un punto negativo que destacar es que TUSSAM no ofrece en ninguno de los dos espacios explicados un planificador de ruta propio, que ofrezca indicaciones al usuario de cómo llegar hasta una parada, qué línea coger o qué transbordos son necesarios y dónde. Si bien tanto en la web como en la aplicación para móviles existe una sección para ello, esta se limita a abrir un enlace o la aplicación de Google Maps, con las indicaciones solicitadas. Este hecho no sólo no es una desventaja competitiva para la empresa, dado que Google Maps puede recomendar opciones preferidas que no incluyan el uso del autobús, sino que tampoco es profesional, dado que Google Maps cuenta con unos horarios teóricos de paso de los autobuses que no coinciden con los horarios oficiales que se encuentran en la web de TUSSAM; y estos tiempos tampoco se actualizan en tiempo real con los datos oficiales. A diferencia de Google Maps, otras aplicaciones como Moovit sí cuentan con tiempos de paso en tiempo real iguales a los que ofrece TUSSAM en su aplicación y web, a pesar de que los horarios teóricos tampoco son correctos. Esta última afirmación se ha comprobado a través de la comparación en diferentes líneas, días de la semana y paradas.

Además de a través de la aplicación y la web, en aproximadamente el 10% de las paradas de autobús de la ciudad se encuentran instaladas pantallas denominadas “terminales de información en parada” que muestran información en tiempo real del tiempo que le falta al autobús más próximo de cada línea para llegar; junto con posibles avisos de última hora. [119]

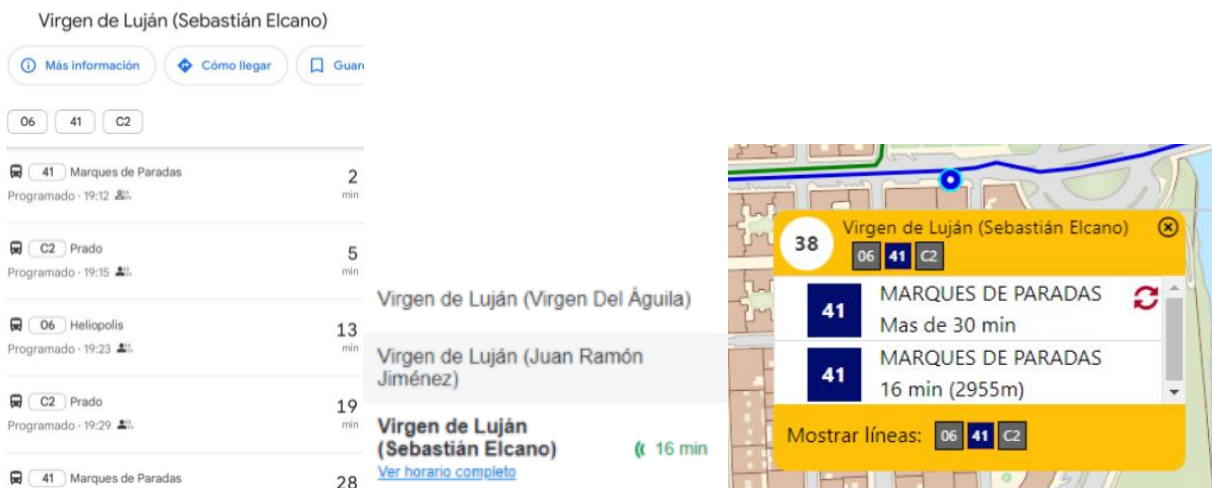


Figura 17. Información en tiempo real del tiempo de paso de la línea 41 en la parada 38 el sábado 3 de septiembre de 2022. El tiempo mostrado en Google Maps (izquierda) es diferente del mostrado en Moovit (centro) y la web oficial de TUSSAM (derecha).

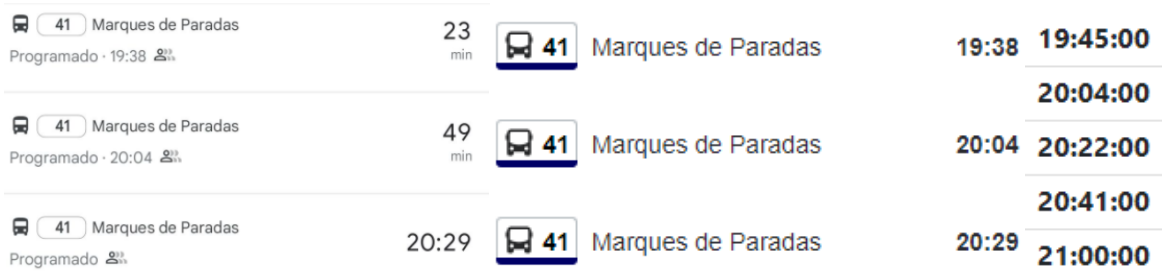


Figura 18. Información de horario previsto del tiempo de paso de la línea 41 en la parada 38 el sábado 3 de septiembre de 2022. El horario mostrado en Google Maps (izquierda) es igual al de Moovit (centro), pero diferente al de la web de TUSSAM (derecha).

#### 4.2.1.4 Otros sistemas implementados

Además de las tecnologías de la información explicadas anteriormente, existen otras iniciativas puestas en marcha por el operador de autobuses urbanos de Sevilla:

- **Retrovisores digitales:** Consiste en la sustitución de los espejos retrovisores empleados en la mayor parte de vehículos de carretera hasta la fecha por un conjunto de varias cámaras en los laterales del autobús, conectadas a pantallas en la cabina del conductor; que permiten tener un mayor ángulo de



visión, eliminan los puntos ciegos y aumentan la visibilidad en situaciones de lluvia, nieve o noche; eliminando los deslumbramientos y aumentando la seguridad de los pasajeros, otros vehículos y peatones. [120] Actualmente se encuentra en 24 autobuses de los 408 de la flota. [121]

- **Wifi público:** Lanzado en 2019, y actualmente disponible en todos los vehículos de la flota, consiste en la instalación de un *router* a bordo de cada autobús con 2 tarjetas 4G, que proveen de conexión a internet a los usuarios, que para acceder al servicio deben iniciar sesión tras conectarse, con una red social o la cuenta de la aplicación de TUSSAM. La conexión a internet también es utilizada por algunos sistemas del autobús para comunicarse con el exterior. [122] En base a la información disponible, esta tecnología no se usa para ningún otro tipo de finalidad, como puede ser el rastreo WiFi para conocer el aforo en tiempo real.
- **E-Ticketing:** En 2009, TUSSAM implementó en todos sus autobuses la posibilidad de pagar utilizando diferentes tipos de bonobús, siendo una opción más rápida que el pago en metálico. TUSSAM no detalla tampoco en este caso si se registra algún tipo de información relativa al usuario, el vehículo, la ubicación o la ruta a la que está accediendo. [119]
- **Sistema CCTV:** En 2021, TUSSAM completó la instalación de cámaras de videovigilancia en todos los vehículos de la flota, permitiendo a la central acceder en tiempo real a las imágenes y también almacenarlas. [123] De nuevo en este caso no se tiene constancia de que las imágenes también se usen para conocer el aforo de cada autobús o tranvía.

En base a los últimos sistemas mencionados y al estudio realizado sobre el transporte público en Londres, es interesante recalcar que podrían utilizarse elementos ya existentes en los vehículos e infraestructura de TUSSAM para otras finalidades. Así, podría recopilarse información a partir del uso del bonobús para conocer las rutas más utilizadas como forma alternativa y más eficiente al uso de encuestas. De la misma forma, se podría estudiar el uso de técnicas de inferencia de origen y destino para conocer los viajes más frecuentes que combinan el uso de dos o más líneas. Este último estudio sería mucho más eficiente si se extendiese también utilizando datos del metro y el resto de transportes metropolitanos, con las tarjetas ya existentes que sirven como método de pago en las diferentes modalidades de transporte o con una hipotética nueva tarjeta para un sistema unificado de transportes en toda la ciudad y sus alrededores.

Del mismo modo, usar el rastreo WiFi para controlar el aforo en tiempo real en cada vehículo sería una forma muy eficiente de conocer las líneas y tramos más utilizados, así como las paradas con más afluencia de pasajeros; y no requeriría la instalación de nuevos dispositivos al contarse en los autobuses con *routers* funcionales. Otra opción para realizar este conteo de pasajeros de forma automática sería mediante el uso de software de reconocimiento de imagen a partir de las grabaciones del sistema de cámaras CCTV. Todas las medidas propuestas anteriormente servirían para dar respuesta a algunas de las mejoras propuestas en el Plan de Movilidad Sostenible desarrollado por el Ayuntamiento. [104]

## 4.2.2 Metro de Sevilla

A pesar de que el Metro de Sevilla fue inaugurado en 2009, existe una gran opacidad respecto de los diferentes sistemas de recopilación de información utilizados, no detallándose información sobre el SAE, el control del aforo, la información recopilada al realizar pagos o los servidores utilizados. Preguntados por esta circunstancia, la empresa detalla que “no facilitan datos técnicos de sus sistemas”. A pesar de ello, son varias las innovaciones que este servicio público ha incluido en los últimos años, con el objetivo de ofrecer mejor información a sus usuarios, tener un mejor control de sus instalaciones y vehículos y hacer más accesible el transporte a los ciudadanos.

### 4.2.2.1 Pago con tarjeta de crédito

En septiembre de 2020, el Metro de Sevilla fue el primer metro en España que admitió el pago con tarjeta de crédito o débito directamente en las barreras de seguridad, gracias a una colaboración con el Banco Santander y Mastercard. Al realizar la entrada en el servicio de este modo, no sólo existe un ahorro de tiempo, al no tener que pasar por las taquillas de venta de billetes; sino que la tarifa aplicada es la correspondiente al pago con un bonometro, por lo que también existe un ahorro económico en comparación con los billetes ocasionales. Para usar este servicio, simplemente es necesario pasar la tarjeta por los lectores a la entrada y la salida del metro, para que se pueda computar el trayecto realizado, el número de saltos y la tarifa que se debe cargar. [124] Este

sistema de pago, ideal para turistas o viajeros ocasionales, y ya existente en otros lugares del mundo como Londres o Nueva York, sigue siendo único en los diferentes metros de nuestro país. En otros lugares de España, se pueden realizar pagos con tarjeta en la taquilla, pero no acceder directamente a los andenes con la tarjeta de crédito.

Un año después del lanzamiento de esta funcionalidad, en septiembre de 2021, Sevilla volvió a transformar la forma de acceder al metro, siendo el primero de España en contar con “postpago inteligente”. [125] De este modo, el pago con tarjeta de crédito se puede realizar de dos formas. Por un lado, utilizando la modalidad anterior, pero con una transformación. Ahora, varias personas pueden acceder a la vez, pasando la tarjeta tantas veces como sea necesario por el lector tanto en la entrada como en la salida (que debe realizarse en la misma estación). Además de eso, el cobro se realiza al día siguiente, de forma que no sólo se aplica la tarifa del bonometro, que es más favorable; sino que, en el caso de un viajero único, si el importe a cobrar fuera mayor que el precio de la tarifa de viajes ilimitados durante un día, se le aplicaría este último precio, de forma que siempre recibiese el trato más favorable. [126]

La segunda forma de pago requiere registrarse en la plataforma habilitada para ello, indicando la tarjeta de crédito con la que se accederá a la estación. De este modo, el pago de todos los viajes realizados se realiza al final de cada mes, aplicando la mejor tarifa disponible dentro de todas las posibles para que el importe pagado sea siempre el menor para el usuario. Además, la web del Metro de Sevilla permite conocer los viajes realizados hasta el momento, consultar una estimación del precio a pagar al final del mes y realizar otras gestiones como descargar facturas. En este caso, si se realizasen varios pagos a la vez con la misma tarjeta de crédito para permitir el acceso a varios pasajeros, a estos se les aplicarán la tarifa del bonometro directamente. [126]

#### **4.2.2.2 Control automático de aforo**

A finales de 2020, con el objetivo de vigilar los niveles de aforo en las estaciones de metro, y velar por un servicio adecuado a las necesidades marcadas por la crisis del COVID-19, Metro de Sevilla instaló en todas las estaciones sensores de aforo, permitiendo conocer en tiempo real la ocupación en los andenes y el resto de espacios de las instalaciones. Gracias a esta tecnología, el centro de control puede conocer en tiempo real el nivel de ocupación, medio en tres niveles diferentes, con el objetivo de modificar el servicio en función de las necesidades para hacer frente a grandes afluencias de pasajeros. [127] Esta información, no obstante, no es pública para los usuarios del transporte, sino que sólo se hace uso de ella a nivel interno para la planificación de la cadencia de los trenes. El conteo se realiza a través de más de 200 cámaras instaladas a lo largo de las diferentes estaciones, que permiten no sólo la vigilancia de los espacios sino la detección automática del aforo a través de un software de reconocimiento de imagen, que actualiza de forma automática un panel que muestra la ocupación en tiempo real al centro de control. [128]

#### **4.2.2.3 Control en tiempo real de la calidad del aire**

También enmarcado en una serie de acciones destinadas a luchar contra el COVID-19 y a hacer las instalaciones del Metro de Sevilla más seguro, en julio de 2021 la empresa comenzó la instalación de dispositivos en todos los vehículos de la flota con el objetivo de monitorizar la calidad del aire dentro de estos. Los objetivos de este proyecto eran diversos. Por un lado, conocer parámetros como la concentración de dióxido de carbono, humedad o temperatura; analizándolos en tiempo real y generando avisos para que el personal pudiera actuar en caso de que existieran niveles anormales de algunos de los valores. [129] El otro de los objetivos era, en colaboración con la Universidad de Sevilla, estudiar diferentes sistemas de ventilación para los vehículos y estaciones, con el objetivo de optimizar el proceso de renovación del aire y mantener condiciones ambientales lo más óptimas posibles. [130]

Condiciones ambientales en el interior de los trenes*	
Día	11/09/2022
Hora	18:56
Temperatura (°C)	23
Humedad (%)	51
CO <sub>2</sub> (ppm)	688
Renovación de aire	Excelente

*\* Promedio de los últimos 15 minutos.*

Figura 19. Ejemplo de consulta de la media de las condiciones climáticas de todos los vehículos a través de la web. [131]

A diferencia de la medición de la ocupación, los datos de la calidad del aire sí son públicos para los ciudadanos. En primera instancia, esta información se podía consultar escaneando códigos QR en los vehículos, que permitían acceder a la información climática de ese tren en tiempo real. Poco después del lanzamiento de esta medida, además, se incluyeron en la página web oficial del servicio y en la aplicación espacios habilitados para consultar la media de todas las mediciones de los vehículos en funcionamiento. [130]

#### 4.2.2.4 Planificación de viajes y comunicación con los usuarios

Actualmente, Metro de Sevilla dispone de tres plataformas diferentes a través de las cuales se puede consultar el estado de los servicios para planificar un trayecto en este medio de transporte. En primer lugar, al igual que TUSAM, cuenta con una aplicación para móviles y una web oficial; donde pueden consultarse el tiempo estimado de viaje entre dos paradas, consultar el tiempo que le falta por llegar a los trenes en las dos direcciones en una parada determinada, calcular el precio del viaje entre dos estaciones y otras informaciones como mapas, noticias, precios, etc. Además, se permite la recarga de bonómetros a través de ambos portales, y la consulta del saldo restante. En la aplicación móvil, además, se pueden añadir tarjetas, consultando el histórico de viajes de un bonometro; y añadir paradas a favoritos para acceder rápidamente a la información más útil sobre la misma. Además de todo esto, como ya se ha dicho anteriormente, se pueden consultar las condiciones ambientales del metro y realizar operaciones relacionadas con el servicio Tap&Go. En la aplicación móvil, además, también nos encontramos la opción de consultar el histórico de viajes realizados con un bonometro, algo no disponible a través de la web.

Además de los dos canales “tradicionales” de información, en noviembre de 2019 la empresa lanzó un *bot* de Whatsapp, al cual se le pueden realizar diferentes peticiones para obtener información similar a la que se puede encontrar en la web y la aplicación. De este modo, a través del chat se puede solicitar el precio y la duración de un trayecto, obtener direcciones en Google Maps hasta una parada, obtener información en tiempo real de los tiempos de llegada de los trenes, consultar el saldo de tarjetas o recibir información sobre cómo utilizar el sistema Tap&Go. Para utilizarlo, simplemente se debe abrir el chat con este *bot*, y escribir uno de los comandos existentes para obtener la información deseada. [132] Además, se pueden realizar otras preguntas no estipuladas en los comandos, y el *bot* intentará dar respuesta a ellas. No obstante, esta última forma de solicitar información no funciona en todos los casos, y algunas informaciones se encuentran desactualizadas.

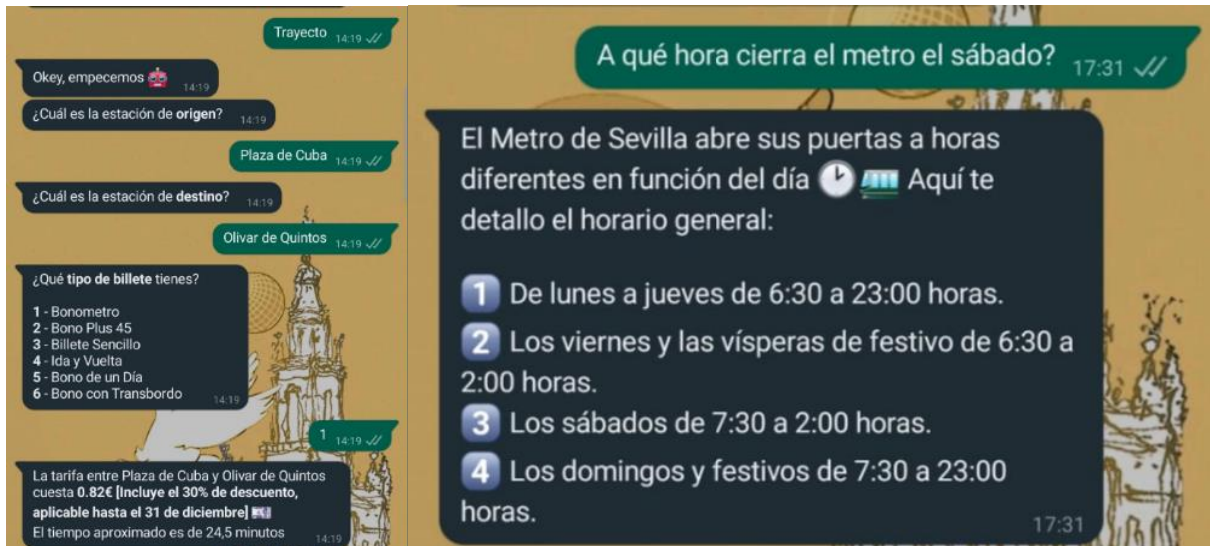


Figura 20. Ejemplos de conversación con bot de metro de Sevilla, solicitando un comando existente (izquierda) y realizando una pregunta genérica (derecha).

### 4.2.3 Otros medios de transporte

Además del Metro y los autobuses urbanos, en Sevilla existen otros tres servicios públicos mayoritarios que dan cobertura a los viajeros. Aunque estas opciones no son tan relevantes como las anteriores, es interesante comentar el estado de esos servicios, y las tecnologías que emplean.

#### 4.2.3.1 Autobuses metropolitanos

De los diferentes servicios existentes en la ciudad y sus alrededores, los autobuses metropolitanos son los que se encuentran en una situación más precaria, tecnológicamente hablando; cuya gestión corresponde al Consorcio de Transportes de Andalucía. [133] A pesar de que todos los vehículos sí que cuentan con sistemas de e-ticketing, pudiéndose pagar de forma más sencilla y barata con un bonobús; los autobuses no cuentan con SAE, un elemento central de la gestión de un sistema de transportes por carretera, ya que sin ellos es imposible conocer la localización del autobús y proveer a los pasajeros del tiempo real restante para que el transporte llegue a una parada determinada; lo que hace imposible que los pasajeros puedan planificar correctamente sus trayectos y reaccionar ante posibles incidencias del servicio. [101]

Además de eso, existe una aplicación oficial que permite acceder a la información de los diferentes consorcios de Andalucía. Esta aplicación, no obstante, cuenta con menos de 100.000 descargas [134] frente a los más de 8 millones de habitantes de la comunidad autónoma; [135] y sólo permite ver los horarios planificados en cada parada, las líneas disponibles, conocer las formas disponibles para ir de un lugar a otro y consultar un tablón de noticias, además de otra información del Consorcio. No obstante, no permite conocer ni recargar el saldo de un bonobús, ni recibir alertas en tiempo real sobre ningún tipo de servicio. Del mismo modo, el resto de las aplicaciones de movilidad, tales como Google Maps o Moovit, ofrecen a los usuarios la posibilidad de conocer los horarios planificados, pero no recibir actualizaciones en tiempo real de este tipo de transporte.

#### 4.2.3.2 Cercanías RENFE

El caso de los diferentes trenes de cercanías es similar al comentado con respecto de los autobuses metropolitanos. Este transporte es muy poco utilizado en nuestra provincia; debido varios factores, entre los que se encuentran pocas líneas disponibles y zonas muy desatendidas, frecuencias de paso muy bajas y un alto precio. [101] Este servicio se encuentra apoyado por una web oficial [136] y una aplicación, ambas gestionadas por RENFE, que permite conocer los horarios de las líneas y planificar viajes entre dos paradas, indicando los transbordos necesarios. En el caso de la aplicación, también pueden consultarse avisos de última hora sobre las líneas en funcionamiento.

### 4.2.3.3 Bicicletas municipales SEVICI

El servicio de bicicletas públicas de Sevilla fue lanzado en el año 2008, y en 2021 contó con cerca de 30000 usuarios. [137] Su uso, al igual que en otras muchas ciudades, depende de un abono contratado, que puede ser de una semana o de un año. En ambos casos, la primera media hora de uso de la bicicleta es gratuita, con un precio por minuto estipulado superados los 30 minutos. [138] Este servicio, a pesar del gran desplome de usuarios que ha experimentado desde su lanzamiento, [137] ha transformado completamente su servicio en los últimos años; poniendo en servicio una aplicación desde la que gestionar todas las operaciones relacionadas con el alquiler de bicicletas.

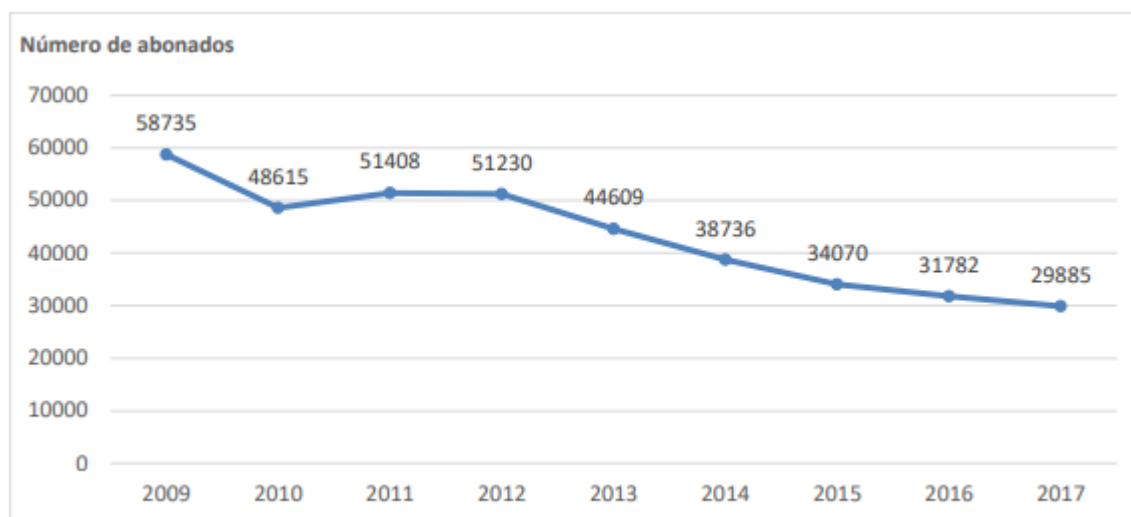


Figura 21. Evolución de usuarios de SEVICI. [104]

Antes de este lanzamiento, cada usuario disponía de una tarjeta, de forma que cuando quisiera alquilar una bicicleta, debía acercarse a un lector, seleccionar el vehículo que quisiera alquilar e introducir un PIN. Además de la lentitud de realizar las operaciones y del mal funcionamiento de las pantallas y teclados habilitados para ello, en ocasiones también era necesario esperar a que otros hicieran sus operaciones, en tanto que sólo existía un punto de alquiler en cada parada. Con el lanzamiento de la nueva aplicación, todo se realiza desde el móvil, eliminando el uso de los terminales físicos en las paradas, poniendo a disposición de los usuarios un mapa con las estaciones cercanas y el número de bicicletas y puestos de aparcamiento disponibles en estos. Además, esta aplicación hace uso de una gran cantidad de datos, que también pone a disposición de los usuarios, ya que al alquilar las bicicletas se pueden comprobar las valoraciones que otros usuarios han hecho de ellas, lo que facilita que los compradores del servicio no utilicen bicicletas en mal estado y sea más sencillo para la empresa el mantenimiento de las mismas.

Aparte de esta información, a través de la aplicación se pueden conocer las tarifas contratadas por el usuario y el historial de pagos; el historial de viajes y su precio y las estadísticas de uso; como cuantos viajes por semana se han realizado, el acumulado de viajes totales o la duración media de estos.

De nuevo, SEVICI no detalla si la información de los viajes realizados por el usuario o la bicicleta empleada en cada viaje se almacenan; pero sí sabemos que las horas del trayecto sí quedan registradas. Al igual que se ha comentado en el caso de TUSAM, sería interesante que esa información se conociese y fuera utilizada para mejorar la calidad de los servicios; reforzando estaciones donde existe una mayor demanda o estudiando la implementación de líneas de autobús en trayectos frecuentes en muchos usuarios como alternativa al uso de la bicicleta.

### 4.2.4 Ecosistema de datos abiertos

El estado de los datos abiertos en la ciudad de Sevilla se encuentra incompleto y en un estado algo confuso. Empezando por esta segunda característica, los datos publicados no se encuentran en un único lugar, centralizado y con un acceso común a los datos, sino que se encuentran gestionados por dos instituciones diferentes, al igual que ocurre con la propia operación de los transportes. Por un lado, la Junta de Andalucía, y por otro, el Ayuntamiento de Sevilla.

En el primer caso, la institución autonómica puso en 2020 en marcha un portal de datos abiertos de los distintos departamentos, con el objetivo de favorecer la transparencia, la innovación y la difusión de información relativa a proyectos llevados a cabo por la Junta de Andalucía. [139] Dentro de este proyecto, existen 8 conjuntos de datos relativos al transporte, donde el más relevante está referido al Consorcio de Transportes de Andalucía. [140] En este conjunto de datos, podemos descargar en formato CSV toda la información relativa a los autobuses metropolitanos de la comunidad, así como de otros servicios gestionados por este organismo (aunque en Sevilla no existen); esto incluye formas de contacto, calendarios de servicio, rutas, horarios de paradas, localizaciones de paradas, etc. Además, se pone a disposición de la ciudadanía una API pública, que permite acceder no sólo a esa información, sino también a otra como pueden ser los municipios a los que da cobertura el Consorcio, noticias relacionadas con los transportes o lugares de interés en cada población. [141] No obstante, después examinar este servicio, se ha comprobado cómo algunos datos se encuentran incompletos, existiendo provincias para las que algunos datos sí se encuentran disponibles y para otras no.

Por otro lado, el Ayuntamiento de Sevilla también cuenta con su propio portal de datos abiertos, con más de 500 conjuntos de datos disponibles; de los cuales algo más de una veintena se encuentran relacionados con los transportes de la ciudad. Este portal, sin embargo, se encuentra en un estado bastante precario debido al abandono que ha sufrido desde hace muchos años. Conjuntos de datos como el estado del tráfico o la operativa de TUSAM llevan sin ser actualizados desde 2015 y 2016 respectivamente; y esta es la línea general relativa a todos los conjuntos de datos de transporte; aunque sí que se publican y actualizan regularmente otros conjuntos de datos relativos a otras áreas de la gestión de la ciudad. [11]

## 5 ANÁLISIS DE LA OPERATIVA DE TUSSAM A TRAVÉS DEL BIG DATA

---

A lo largo de todo este documento se han comentado los beneficios de aplicar el Big Data para el análisis de la movilidad, así como la posibilidad de que estos datos se encuentren publicados en abierto para todo el mundo. A continuación, se realizará un análisis de la operativa de TUSSAM a partir de datos facilitados por la propia empresa. Para ello, se utilizará el lenguaje de programación Python; utilizando Jupyter Notebooks como entorno de desarrollo. Previamente en este documento se han ofrecido otras alternativas para el procesado de datos; no obstante, la sencillez de Python, la existencia de una gran cantidad de documentación y las librerías disponibles justifican el uso de este lenguaje. Por otro lado, el uso de Jupyter Notebooks se debe a la facilidad para visualizar los datos durante la propia ejecución de los códigos y el desarrollo de los mismos. Para el análisis empleado se ha utilizado el formato de datos GTFS. También se ha accedido a información a través de otras fuentes, como son la API de la empresa y mapas publicados en la plataforma de datos geográficos ArcGIS. El objetivo de realizar esta tarea será el obtener estadísticas relativas al servicio prestado por el transporte en la ciudad.

Además de las tecnologías mencionadas anteriormente, se ha hecho uso de una serie de librerías en Python para facilitar el trabajo, que se detallan a continuación.

- ***numpy***: Librería ampliamente usada para el manejo de vectores y matrices y para realizar operaciones con ellos.
- ***pandas***: Librería usada para el análisis y manipulación de grandes cantidades de datos estructurados, ofreciendo estructuras, funciones y herramientas de análisis altamente eficientes.
- ***gtfs-kit***: Librería empleada para el análisis y procesamiento de datos de transporte público en formato GTFS.
- ***folium***: Empleada para el dibujado y visualización de datos espaciales sobre mapas interactivos.
- Además de estas librerías, se han usado otras librerías incluidas en Python, que son ***json***, ***math*** y ***datetime***

En la siguiente figura se puede observar una introducción al flujo de trabajo realizado en el desarrollo de la aplicación. En primer lugar se partirán de los datos GTFS originales proporcionados por TUSSAM; la cual deberá preprocesarse infiriendo los horarios no anunciados, así como completarse utilizando datos de la plataforma ArcGIS. Una vez hecho esto dispondremos de un *feed* completo y válido; a partir del cual haremos operaciones para obtener estadísticas de las paradas y las líneas a las que se da servicio desde la empresa

municipal de transportes. Estos cálculos darán como productos, en ambos casos, una serie de ficheros CSV con las estadísticas empleadas; y a la vez también mapas para visualizar estas estadísticas. Toda esa información será publicada en una web; y a la vez, servirá para realizar el análisis pertinente.

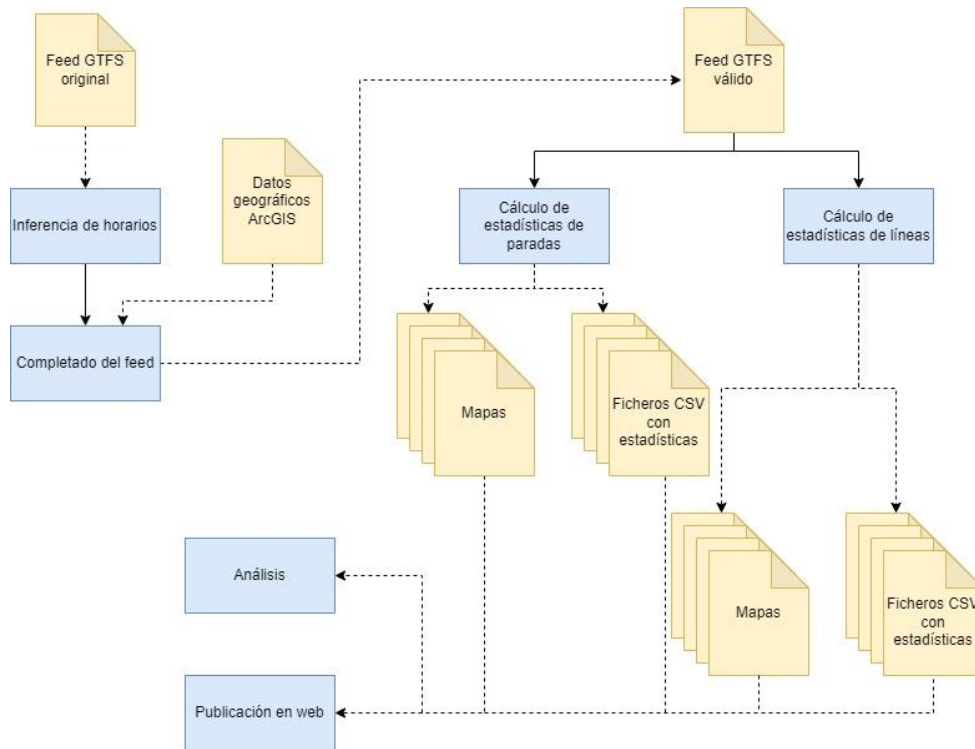


Figura 22. Flujo de trabajo desarrollado en la aplicación realizada.

## 5.1. Fuentes de datos

### 5.1.1 API de TUSSAM

TUSSAM no ha puesto al servicio de la ciudadanía una API donde se puedan realizar consultas y que permita a terceros construir aplicaciones haciendo uso de los datos de la empresa. A pesar de que sería deseable que esto estuviera permitido, sí que existe una forma de hacer uso de la API que utiliza TUSSAM para proporcionar información a la ciudadanía a través de su web. Para ello, es necesario acceder al servicio de Red de Líneas de la empresa, que permite obtener información en tiempo real sobre el transporte de la ciudad; más concretamente, permite ver las líneas activas, recorridos, posiciones de vehículos en tiempo real, horarios y estimaciones de tiempos de paso de los autobuses.

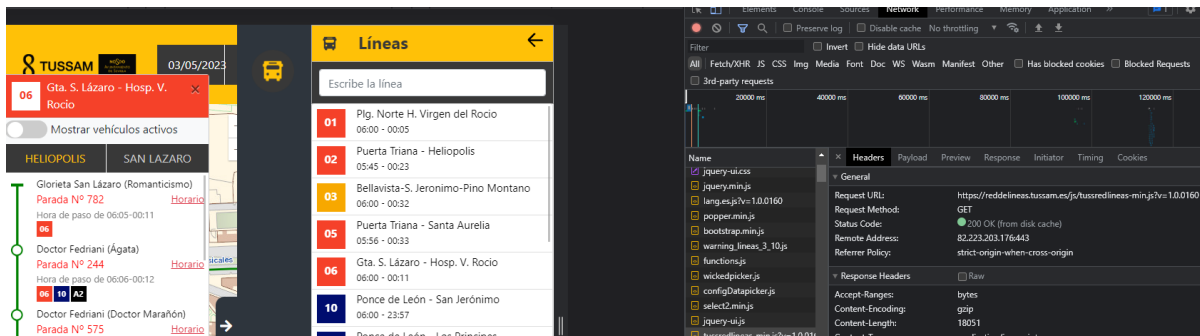


Figura 23. Ejemplo de uso del analizador de paquetes para acceder a la URL de una petición.

Una vez abierta esta web, es necesario abrir un analizador de paquetes, como WireShark o las herramientas de desarrollador que se encuentran disponibles en todos los navegadores, en la opción de red. Si bien este método



no es del todo apropiado y requiere mayor trabajo, es la forma más directa que tenemos de acceder a los datos provistos por una fuente oficial.

Navegando por las distintas posibilidades que ofrece la página con la herramienta abierta se intercambiarán mensajes entre nuestro equipo y distintos servidores; y analizando los paquetes intercambiados, podemos ver que algunos de ellos están dirigidos a URLs comenzadas por *reddelneas.tussam.es/api/infotus-ui*. Este es el enlace básico de la API REST de la que se puede hacer uso para acceder a la información de TUSSAM con este sistema. No obstante, la empresa no cuenta con ninguna documentación relativa a la misma, en tanto que no se considera algo abierto a todo el mundo. Por este motivo, acceder a los *endpoints* de la misma resulta una tarea laboriosa. A partir de analizar el tráfico generado al navegar por todas las opciones que ofrece el servicio, se han encontrado los siguientes *endpoints*. Los accesos a cada uno de estos enlaces proporcionan, en formato JSON, la información detallada en la figura 29. Resulta importante destacar que, en general, sólo se puede acceder a información relativa al día actual y a los dos siguientes. Además, toda la información se actualiza en tiempo real y provee información de la situación actual o especificada en una fecha. De este modo, si se quisiera utilizar algún enlace para conocer el estado de alguna línea o parada no activa en ese momento, se recibirá un mensaje similar a un estado de error; tal y como si la información no existiera.

Enlace	Información accedida
<a href="https://reddelneas.tussam.es/api/infotus-ui/buses/codigo_linea">https://reddelneas.tussam.es/api/infotus-ui/buses/codigo_linea</a>	Cada entrada indica la posición, expresada en latitud y longitud de los autobuses en funcionamiento en la línea indicada con <i>codigo_linea</i> ; así como el sentido que están recorriendo y la distancia total desde el comienzo de la ruta.
<a href="https://reddelneas.tussam.es/api/infotus-ui/horarios/codigo_parada/codigo_linea/DD-MM-AAAATHH:MM:SS">https://reddelneas.tussam.es/api/infotus-ui/horarios/codigo_parada/codigo_linea/DD-MM-AAAATHH:MM:SS</a>	Indica los horarios de paso previstos para la línea expresada en <i>codigo_linea</i> para la parada indicada con <i>codigo_parada</i> , en el día especificado.
<a href="https://reddelneas.tussam.es/API/infotus-ui/tiempos/codigo_parada">https://reddelneas.tussam.es/API/infotus-ui/tiempos/codigo_parada</a>	Devuelve el nombre y la localización de la parada indicada con <i>codigo_parada</i> , así como la estimación de tiempo y distancia de cada línea que pasa por la misma.
<a href="https://reddelneas.tussam.es/API/infotus-ui/nodosLinea/codigo_linea/codigo_sentido/DD-MM-AAAATHH:MM:SS">https://reddelneas.tussam.es/API/infotus-ui/nodosLinea/codigo_linea/codigo_sentido/DD-MM-AAAATHH:MM:SS</a>	Devuelve información relativa a las paradas activas de la línea indicada en <i>codigo_linea</i> y sentido <i>codigo_sentido</i> para la fecha y hora indicada. Esta información incluye la ubicación, horario de inicio y fin del servicio, código de la parada o nombre de esta.

Tabla 2. Endpoints de la API de TUSSAM.

Debido a la falta de documentación comentada y que los resultados sobre una misma consulta pueden variar según la fecha o la hora en la que se realice, dado que se actualizan en tiempo real, **en el análisis final realizado no he hecho uso de la API para obtener información**. No obstante, como aproximación inicial sí se hizo uso de la API.

### 5.1.2 Datos de la plataforma ArcGIS

Para el caso que nos atañe, en ArcGIS existen dos conjuntos de datos útiles de donde podemos obtener información relativa a la empresa de pública TUSSAM. El primero de ellos **contiene un conjunto de puntos que permite dibujar sobre el mapa las distintas líneas que componen la red de autobuses urbanos; junto con información de estas líneas como su nombre, código, sentido o destino**. El otro muestra una información similar a la obtenida con la información de paradas a través de la API, como es la ubicación, el código, nombre o líneas y sentidos de todas las paradas del sistema. Como se explicará más adelante, de los dos conjuntos se ha hecho uso tan sólo del primero; del que se ha descargado toda la información a través de la herramienta de

peticiones de ArcGIS Online.

De nuevo, como ocurría con la API de TUSSAM, la disponibilidad pública de estos datos no está anunciada ni publicitada a través de ningún medio oficial.

### 5.1.3 Datos GTFS

Para realizar nuestro análisis, **la mayor parte de los datos empleados proceden o están derivados del *feed* GTFS correspondiente a la empresa TUSSAM**. Este conjunto de archivos está elaborado por la propia empresa pública de transportes, y para acceder a él tuve que ponerme en contacto con la propia empresa, dado que aunque estos datos son de dominio público no se encuentran publicados en ningún lugar. Los datos relativos a TUSSAM que se encuentran publicados en el Portal de Datos Abiertos mencionado anteriormente se corresponden con el formato GTFS, pero no están actualizados desde el año 2016. [11]

Estos datos proporcionados incluyen todos los ficheros mencionados en el apartado 2 del presente documento, y todos ellos son necesarios para que la librería empleada para el procesado de datos pueda validar el conjunto de datos. No obstante, la información contenida en los ficheros *feed\_info.txt* y *agency.txt* no se ha usado, en tanto que contiene información genérica de la empresa de transportes que no resulta de utilidad para el trabajo realizado.

#### 5.1.3.1 Valores de mayor relevancia para nuestro trabajo

Dentro de todo el conjunto de tablas y campos que define el estándar, en nuestro caso sólo haremos uso de algunos de ellos. Para facilitar la comprensión de los códigos explicados a continuación y los resultados obtenidos es conveniente explicar brevemente los campos más importantes y de los que mayor uso hacemos a lo largo de la aplicación práctica desarrollada:

- *stop\_name*: Nombre de una parada, definiendo en general la calle en la que se encuentra.
- *stop\_id*: Número natural identificador de una parada, coincide con el identificador que se da a esta en la aplicación de la empresa de transportes.
- *route\_id*: Número natural que identifica una línea de autobuses. En general, este identificador no tiene por qué coincidir con el número por el cual se nombra a la línea. En algunos casos sí son el mismo valor; pero en otros, donde las líneas llevan letras en su nombre, esta coincidencia no se respeta.
- *route\_short\_name*: Es el nombre que recibe la línea. Como se ha especificado anteriormente, puede o no coincidir con el *route\_id*.
- *direction\_id*: Valor binario que sirve para indicar el sentido de la línea. Si este valor es 0, indica que la ruta va en su sentido de ida; si el valor es 1, indica que es el sentido de vuelta. De este modo, el par *route\_id* y *direction\_id* permiten identificar la línea y la dirección que lleva un vehículo.
- *trip\_headsign*: Indica el destino de cada par ruta/dirección. Equivale al mensaje que se muestra en el frontal de los autobuses para advertir a la ciudadanía del sentido del vehículo.
- *service\_id*: Sirve para especificar qué días de la semana da servicio una ruta. De este modo, a cada viaje que tiene lugar dentro de una línea se le puede indicar si se realiza de lunes a jueves, viernes, sábado o domingo; puesto que así es como tiene agrupados los viajes la empresa municipal.
- *trip\_id*: Identifica cada uno de los viajes a los que da servicio una línea. Un viaje se refiere al recorrido que realiza un vehículo desde que sale de una de las cabeceras de la línea hasta que llega al final de la misma; y cada viaje está asociado a un tipo de servicio, una ruta y una dirección.
- *shape\_id*: Sirve para indicar en un solo valor la ruta y dirección asociada a un viaje. En general, este valor se emplea para ligar un viaje con las coordenadas en el mapa que sigue el trayecto del autobús, pudiendo dibujarlo fácilmente sobre un plano de la ciudad.

Por último, es necesario indicar que **estos datos proporcionados por TUSSAM no se corresponden con medidas reales sino con la planificación del servicio por parte de la empresa**, por tanto, el análisis realizado es un análisis estático del servicio proporcionado por la empresa durante los meses correspondientes a la primavera de 2023, sin contar los posibles servicios especiales dados en fechas señaladas como la Feria de Abril

o la Semana Santa; o eventos especiales como partidos de fútbol. TUSSAM no proporciona información de otro tipo, como podrían ser los horarios reales medidos en un periodo anterior, la demanda de cada vehículo en cada parada (número de pasajeros recogidos) o la propia ocupación de los autobuses. Además de esto, los datos GTFS recibidos son incompletos, por lo que métricas como la velocidad por sectores de cada línea no son posibles de calcular.

**Con esta información podemos hacer una evaluación de los servicios planificados por TUSSAM, y más concretamente de la oferta que la empresa pone a disposición de las personas en los diferentes días de la semana.** No podemos hacer, por tanto, evaluaciones relacionadas con el cumplimiento de esas planificaciones, ni tampoco de la atención o no de la demanda existente o la existencia de servicios innecesarios o que requieren refuerzos, en tanto que esos datos no son proporcionados por la compañía.

## 5.2. Preparación de los datos

Antes de comenzar con el trabajo sobre los datos con el objetivo de obtener estadísticas e información sobre la operativa de TUSSAM, ha sido necesario realizar una serie de tareas sobre el *feed* GTFS, con el objetivo de completar información que no se encontraba disponible para parte de los datos o añadir información inexistente en los ficheros recibidos.

### 5.2.1. Inferencia de horarios

La especificación GTFS define que los horarios de líneas cuyo servicio comienza un día pero termina el siguiente deben expresarse con formatos de hora posteriores a las 24:00:00; como si fueran horas extra del día donde comenzó el servicio. De esta forma, si una ruta comenzase a las 23:50 del lunes y finalizase a las 01:00 del día siguiente esta hora de finalización se expresaría como las 25:00 del propio lunes. Desafortunadamente, la librería utilizada en Python para el manejo de horarios es incompatible con este formato; por lo que previamente a realizar la inferencia de horarios para cada parada era necesario realizar una actualización de los horarios. Para ello, sería posible recorrer todo el conjunto de datos con algún script en Python, de forma que se fueran leyendo y sustituyendo los horarios expresados del formato incompatible al general de 24 horas. Sin embargo, debido al gran volumen de filas del fichero *stop\_times.txt* la mejor opción era hacer uso de las funciones de búsqueda y reemplazo de cualquier editor de texto.

Hora en formato GTFS	Hora reemplazada para trabajar en Python	Hora en formato GTFS	Hora reemplazada para trabajar en Python
25:00	01:00	28:00	04:00
26:00	02:00	29:00	05:00
27:00	03:00	30:00	06:00

Tabla 3. Cambio en el formato de horarios de madrugada para trabajar en Python.

Como se ha comentado anteriormente, el fichero recibido de TUSSAM relativo a los horarios de las paradas de autobús contaba con una gran cantidad de filas sin información temporal. **Para calcular las estadísticas deseadas, por tanto, ha sido necesario realizar una inferencia de los mismos.** Tomando como referencia los tiempos más cercanos anterior y posterior, el método más preciso para realizar la inferencia sería calibrar los tiempos en función de las distancias entre distintas paradas. No obstante, esa información no la tenemos disponible tampoco en el *feed* GTFS de la empresa. Por tanto, la solución por la que se ha optado ha sido dividir el tiempo transcurrido entre dos paradas conocidas de forma igualitaria entre todas las paradas intermedias de las cuales hay que inferir un horario. Aunque este método no es del todo preciso en tanto que los trayectos no son idénticos y están afectados por factores externos, sí que es un método lo suficientemente correcto como para obtener estadísticas relativas al funcionamiento del sistema; y se encuentra expresado matemáticamente a continuación. En la expresión,  $t_n$  representa el tiempo de la parada  $n$  dentro de un intervalo de paradas con tiempo desconocido siendo  $n$  el número del orden de esta parada,  $t_{\text{inicial}}$  el tiempo de la última parada conocida;  $\Delta t$  el tiempo transcurrido entre las paradas cuyos tiempos sí conocemos y  $N$  el número total de paradas a las que

calcular el tiempo.

$$t_n = t_{inicial} + n * \frac{\Delta t}{N + 1}$$

En el código 5.1 mostrado en las dos próximas páginas se encuentra el notebook empleado para realizar esta tarea. La ejecución de los dos primeros bloques se refiere a la importación de librerías y del propio conjunto de datos. Posteriormente, en las siguientes tres celdas, se realiza todo el procesado de los datos, generando un nuevo *dataframe* sobre el que se realiza la inferencia. Para realizar las operaciones, se obtiene una lista de los horarios disponibles con sus índices; y se itera sobre aquellos sin tiempos asignados; calculando el tiempo transcurrido entre los puntos anterior y siguiente conocidos y el número de paradas que existen en ese intervalo. Por último, se sustituyen los valores calculados y se genera un nuevo archivo respetando el formato GTFS. También se muestran ejemplos de los datos en el estado previo y posterior a la ejecución. El resto de campos, con valores iguales no numéricos (NaN), no serán utilizados para nuestros análisis.

```
#Import the gtfs-kit module
import gtfs_kit as gk
import numpy as np
import math
import datetime
```

```
#En primera instancia lo haremos repartiendo tiempos de forma igualitaria
path = 'TUSSOK.zip'

#Read the feed with gtfs-kit
feed = (gk.read_feed(path, dist_units='km'))
feed.stop_times
```

	trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type	shape_dist_traveled
0	3600101004	07:00:00	07:00:00	233	6717	NaN	NaN	NaN	NaN
1	3600101004	NaN	NaN	251	6902	NaN	NaN	NaN	NaN
2	3600101004	07:04:00	07:04:00	252	7102	NaN	NaN	NaN	NaN
3	3600101005	07:04:00	07:04:00	252	0	NaN	NaN	NaN	NaN
4	3600101005	NaN	NaN	225	301	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
505943	4412707023	NaN	NaN	571	4499	NaN	NaN	NaN	NaN
505944	4412707023	22:05:00	22:05:00	223	6210	NaN	NaN	NaN	NaN
505945	4412707023	NaN	NaN	507	7889	NaN	NaN	NaN	NaN
505946	4412707023	NaN	NaN	512	8775	NaN	NaN	NaN	NaN
505947	4412707023	22:19:00	22:19:00	535	9400	NaN	NaN	NaN	NaN

505948 rows × 9 columns

```
#Nos quedamos solo con los campos que nos interesan
df = feed.stop_times.drop(['stop_headsign', 'pickup_type', 'drop_off_type', 'shape_dist_traveled'],axis=1)
dfConditions = df.isna()
#Sacamos una lista de los que tienen tiempos asignados
dataTimes = df.dropna()

timeIndexes = []
for row in dataTimes.index:
    timeIndexes.append(row)

prevTime = datetime.datetime.strptime(df.iloc[timeIndexes[0]]['arrival_time'], '%H:%M:%S')
nextTime = datetime.datetime.strptime(df.iloc[timeIndexes[1]]['arrival_time'], '%H:%M:%S')
```

```
i=0
j=1
newTimes = []
for index, row in df.iterrows():
    if(dfConditions.iloc[index]['arrival_time']):
        timeElapsed = (nextTime-prevTime)/(timeIndexes[i]-timeIndexes[i-1])
        myTime = prevTime + j*timeElapsed
        myTime = myTime.replace(microsecond=0)
        myTime = myTime.time()
        j=j+1
        newTimes.append(myTime.strftime("%H:%M:%S"))

    else:
        j=1
        if(i==len(timeIndexes)-1):
            newTimes.append(nextTime.strftime("%H:%M:%S"))
        else:
            prevTime = datetime.datetime.strptime(df.iloc[timeIndexes[i]]['arrival_time'], '%H:%M:%S')
            nextTime = datetime.datetime.strptime(df.iloc[timeIndexes[i+1]]['arrival_time'], '%H:%M:%S')
            newTimes.append(prevTime.strftime("%H:%M:%S"))
            i=i+1
```

```
df['arrival_time'] = newTimes
df['departure_time'] = newTimes
```

```
feed.stop_times['arrival_time'] = df['arrival_time']
feed.stop_times['departure_time'] = df['departure_time']
feed.stop_times
```

	trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	pickup_type	drop_off_type	shape_dist_traveled
0	3600101004	07:00:00	07:00:00	233	6717	NaN	NaN	NaN	NaN
1	3600101004	07:02:00	07:02:00	251	6902	NaN	NaN	NaN	NaN
2	3600101004	07:04:00	07:04:00	252	7102	NaN	NaN	NaN	NaN
3	3600101005	07:04:00	07:04:00	252	0	NaN	NaN	NaN	NaN
4	3600101005	07:05:00	07:05:00	225	301	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
505943	4412707023	22:00:00	22:00:00	571	4499	NaN	NaN	NaN	NaN
505944	4412707023	22:05:00	22:05:00	223	6210	NaN	NaN	NaN	NaN
505945	4412707023	22:09:40	22:09:40	507	7889	NaN	NaN	NaN	NaN
505946	4412707023	22:14:20	22:14:20	512	8775	NaN	NaN	NaN	NaN
505947	4412707023	22:19:00	22:19:00	535	9400	NaN	NaN	NaN	NaN

505948 rows × 9 columns

```
feed.stop_times.to_csv(r'stop_times_fixed.txt', index=False)
```

Código 5.1. Notebook *inferStopTimes.ipynb* para inferencia de tiempos.

### 5.2.2. Generación de *shapes.txt*

El *feed* proporcionado por TUSSAM no contaba con el tipo de archivo *shapes.txt* que he decidido añadir por completitud y con el objetivo de poder utilizar, si fuera necesario, para realizar representaciones geográficas de la geometría de la red de transportes urbanos. Para ello, he hecho uso de la plataforma

ArcGIS previamente comentada. Más concretamente, he utilizado la funcionalidad de peticiones a la API a través de la interfaz web; que me ha permitido descargarme un fichero JSON con la información necesaria para la construcción de este archivo.

A través de la plataforma, he solicitado realizar una petición de todos los objetos del conjunto de datos (en este caso, indicado con la condición *where 1=1*) y recibiendo como salida todos los campos. Por último, en el formato de datos de salida, he seleccionado el tipo GEOJSON; un tipo de fichero igual que el JSON pero con los valores de latitud y longitud expresados en grados y decimales, respetando el formato empleado por la propia especificación GTFS. Cabe destacar que esta aplicación nos permite controlar una gran cantidad de variables y opciones para realizar las peticiones; pero estas no son necesarias y se han omitido de la figura 30 al no ser relevantes para nuestro caso.

**El resultado de ejecutar esta petición ha sido un JSON; donde cada uno de los objetos contenidos tiene dos partes fundamentales;** por un lado, una sucesión ordenada de puntos geográficos que permiten ir trazando las ubicaciones por las que pasa esa ruta, y otra parte con las propiedades de la ruta, como su nombre, código, destino, identificador de ruta o el color asociado a esta.

**Query: TUSSAM\_Lineas\_0 (ID: 0)**

Where:	1=1
Object IDs:	
Time:	
Input Geometry:	
Geometry Type:	Envelope
Input Spatial Reference:	
Spatial Relationship:	Intersects
Result Type:	None
Distance:	0.0
Units:	Meters
Relation:	
Return Geodetic:	<input type="radio"/> True <input checked="" type="radio"/> False
Out Fields:	*
Format:	GEOJSON
<input type="button" value="Query (GET)"/> <input type="button" value="Query (POST)"/>	

Figura 24. Plataforma de peticiones ArcGIS REST Services.

Con los datos descargados podemos crear el fichero deseado *shapes.txt* en Python de forma simple. Para ello, haciendo uso de la librería JSON, **cargamos los datos recibidos y vamos recorriendo cada uno de los features de los que estaba compuesto el conjunto;** creando distintas listas con los valores deseados. Por último, es creado un *dataframe* de pandas que se guarda como CSV con el nombre del fichero GTFS.

```
import pandas as pd
import json

f = open('data/shapes.json')
data = json.load(f)['features']
shape_id=[]
shape_pt_lat=[]
shape_pt_lon=[]
shape_pt_sequence=[]
for e in data:
    i=0
    for coord in e['geometry']['coordinates']:
        shape_id.append(e['properties']['Ruta'])
        shape_pt_lon.append(coord[0])
        shape_pt_lat.append(coord[1])
        shape_pt_sequence.append(i)
        i=i+1

df = pd.DataFrame(list(zip(shape_id, shape_pt_lat, shape_pt_lon, shape_pt_sequence)),
                  columns = ['shape_id', 'shape_pt_lat', 'shape_pt_lon', 'shape_pt_sequence'])
df
```

```
df.to_csv(r'shapes.txt', index=False)
```

Código 5.2. Notebook *genShapes.ipynb* para la generación de *shapes.txt*.

### 5.2.3. Completado de *trips.txt* y validación del *feed*

Para terminar con la preparación de los datos y el completado del feed GTFS recibido, **procedemos a añadir a cada uno de los viajes el identificador de su correspondiente forma (que es el mismo identificador que la ruta) en el fichero *shapes.txt***. Estos identificadores de ruta se caracterizan por estar directamente relacionados con la línea a la que dan soporte; asignando a cada una un número según la siguiente fórmula; siendo *direction\_id* igual a 0 o 1 según si la ruta se corresponde con el viaje de ida o de vuelta de esa línea. Esta expresión que se detalla debajo no es estándar; sino que se ha observado que se respeta después de observar los datos empleados por TUSSAM.

$$shape\_id = 100 * route\_id + 11 + direction\_id$$

De este modo, podemos recorrer el conjunto de viajes definido en *trips.txt*, asignando a cada una su *shape\_id* correspondiente sólo fijándonos en los valores de esa misma fila, siguiendo la fórmula mencionada anteriormente.

```
#Import the gtfs-kit module
import gtfs_kit as gk
path = 'TUSS_SHAPES.zip'

#Read the feed with gtfs-kit
feed = (gk.read_feed(path, dist_units='km'))
```

```
feed.trips
```

	route_id	service_id	trip_id	trip_headsign	direction_id	block_id	shape_id
0	1	INVFES20	3600101004	POLIGONO NORTE	0	NaN	NaN
1	1	INVFES20	3600101005	HOSPITAL V.ROCIO	1	NaN	NaN
2	1	INVFES20	3600101006	POLIGONO NORTE	0	NaN	NaN
3	1	INVFES20	3600101007	HOSPITAL V.ROCIO	1	NaN	NaN
4	1	INVFES20	3600101008	POLIGONO NORTE	0	NaN	NaN
...	...	...	...	...	...	...	...
27377	127	INNVIE20	4412707019	BELLAVISTA	1	NaN	NaN
27378	127	INNVIE20	4412707020	SANTA JUSTA	0	NaN	NaN
27379	127	INNVIE20	4412707021	BELLAVISTA	1	NaN	NaN
27380	127	INNVIE20	4412707022	SANTA JUSTA	0	NaN	NaN
27381	127	INNVIE20	4412707023	BELLAVISTA	1	NaN	NaN

```
df = feed.trips
listTrips = []
for ind, row in df.iterrows():
    listTrips.append(int(row['route_id'])*100+10+1+int(row['direction_id']))
df['shape_id'] = listTrips
df
```

	route_id	service_id	trip_id	trip_headsign	direction_id	block_id	shape_id
0	1	INVFES20	3600101004	POLIGONO NORTE	0	NaN	111
1	1	INVFES20	3600101005	HOSPITAL V.ROCIO	1	NaN	112
2	1	INVFES20	3600101006	POLIGONO NORTE	0	NaN	111
3	1	INVFES20	3600101007	HOSPITAL V.ROCIO	1	NaN	112
4	1	INVFES20	3600101008	POLIGONO NORTE	0	NaN	111
...	...	...	...	...	...	...	...
27377	127	INNVIE20	4412707019	BELLAVISTA	1	NaN	12712
27378	127	INNVIE20	4412707020	SANTA JUSTA	0	NaN	12711
27379	127	INNVIE20	4412707021	BELLAVISTA	1	NaN	12712
27380	127	INNVIE20	4412707022	SANTA JUSTA	0	NaN	12711
27381	127	INNVIE20	4412707023	BELLAVISTA	1	NaN	12712

27382 rows × 7 columns

```
df.to_csv(r'trips.txt', index=False)
```

Código 5.3. Notebook *addShapesToTrips.ipynb* para la asignación de *shape\_id* a cada ruta.

Después de realizar todo el proceso de preparación de los datos, para terminar con esta tarea, **hay que crear un nuevo fichero .zip que contenga todos los archivos modificados y anteriores, de forma que se cumpla con la especificación GTFS**. Una vez hecho este fichero, podemos validar el *feed* para comprobar que efectivamente los cambios que hemos hecho continúan respetando el formato de referencia GTFS y que no hay ningún error. Esto lo hacemos con un breve *script* de Python, haciendo uso de la función de validación de la librería *gtfs\_kit*, la librería de referencia que estamos empleando para hacer el análisis de los datos de transporte. Haciendo uso de ella, observamos en el código 5.4 los resultados, que muestran tres avisos en tanto que en el fichero *feed\_info.txt* no se encuentran definidos el lenguaje, correo electrónico y enlace de contacto. No obstante, estos avisos no son incompatibles con el trabajo a realizar; por tanto, podemos concluir aquí con el preparado de los datos.

```
#Importamos el módulo
import gtfs_kit as gk

#Definimos la ruta
path = 'TUSSAM_COMP.zip'

#Leemos los datos y los validamos
feed = (gk.read_feed(path, dist_units='km'))
feed.validate()
```

	type	message	table	rows
0	warning	Unrecognized column default_lang	feed_info	[]
1	warning	Unrecognized column feed_contact_email	feed_info	[]
2	warning	Unrecognized column feed_contact_url	feed_info	[]

Código 5.4. Notebook para la validación del feed.

Una vez terminada esta fase contamos con un *feed* GTFS con toda la información necesaria para realizar un estudio y evaluación del sistema de transporte de la ciudad; analizando y presentando datos para las distintas paradas y líneas que se ponen a disposición de la ciudadanía. Para ello, en primer lugar tendremos que procesar los datos generando información relevante a partir de ellos; y luego presentarlos y extraer conclusiones.



### 5.3. Cálculo de estadísticas de paradas

En primer lugar, trabajaremos con las estadísticas relacionadas con las paradas. El objetivo en esta etapa será conseguir, a partir de los datos presentes en el *feed* GTFS completo preparado en el apartado anterior, valores relevantes relacionados con las paradas del sistema de transportes de la ciudad. Más concretamente, buscamos conocer la frecuencia media de paso de cada una de las líneas por cada una de las paradas, según el día de la semana en el que nos encontremos. A partir de ahí, buscaremos obtener la frecuencia media general de cada parada, y el incremento de tiempo que se sufre en cada parada los sábados y los domingos, es decir, el empeoramiento del tiempo de paso que se produce los días no laborables. Todos estos datos se volcarán en una serie de ficheros que se irán construyendo de forma progresiva, con el objetivo de poder hacer uso de los datos posteriormente para el presentado de los resultados y su inclusión en este mismo documento.

Para realizar todas estas tareas y obtener los datos explicados anteriormente hemos hecho uso de tres Jupyter Notebooks distintos; y cada uno de ellos proporciona como resultados distintos ficheros que serán almacenados en el directorio del proyecto *results/frequencias*. Los códigos completos empleados en esta etapa se encuentran incluidos completamente en el Anexo A del presente documento.

#### 5.3.1 Combinación de horarios y líneas asociadas

En primer lugar, se ha generado un documento en formato CSV con la información de partida que necesitaremos para realizar el análisis, en el fichero *stopTimesForStats.ipynb*. Este fichero cuenta con más de 500000 entradas, correspondientes a la combinación de todos los horarios de todas las paradas y todas las líneas para los cuatro tipos de servicios prestados por la empresa. Para ello, se ha partido del fichero *stop\_times.txt* de los datos GTFS, del cual se han copiado los campos *trip\_id*, *arrival\_time* y *stop\_id*. A estas columnas se les han añadido columnas para identificar la ruta, dirección y nombre del destino al que se refieren; uniendo todo en una misma tabla. Por último, se han ido recorriendo los valores de *trip\_id* en la tabla *trips.txt* del conjunto de datos, completando la tabla nueva y asociando a cada valor del viaje su correspondiente línea y dirección. Por último, se ha almacenado toda esta información en un archivo CSV, denominado *stopTimesComp.txt* y guardado en el directorio *data*. El aspecto del resultado obtenido se muestra en la siguiente figura.

<i>trip_id</i>	<i>arrival_time</i>	<i>stop_id</i>	<i>route_id</i>	<i>direction_id</i>	<i>trip_headsign</i>
3600101004	07:00:00	233	1	0.0	POLIGONO NORTE
3600101004	07:02:00	251	1	0.0	POLIGONO NORTE
3600101004	07:04:00	252	1	0.0	POLIGONO NORTE
3600101005	07:04:00	252	1	1.0	HOSPITAL V.ROCIO
3600101005	07:05:00	225	1	1.0	HOSPITAL V.ROCIO

Tabla 4. Muestra del fichero *stopTimesComp.txt*.

#### 5.3.2 División por tipo de servicio y ordenación por tiempo

Una vez finalizado el fichero anterior, resulta interesante el dividir los datos según el día de servicio al que pertenecen para hacer más sencillo trabajar con ellos. Además, si la tabla de salida se encuentra ordenada bajo ciertos criterios, nos será mucho más sencillo hacer los cálculos relacionados con la frecuencia de paso de los vehículos por las paradas. Para ello hacemos uso del Notebook *sortAndSplitByDay.ipynb*; que genera como salida 4 ficheros en formato CSV correspondientes a los datos segmentados por cada tipo de servicio (según el día en el que se producen) y ordenados.

En este código, una vez cargados los datos GTFS y el fichero generado en la etapa anterior, creamos un nuevo *dataframe* donde la tabla está ordenada según el identificador de la parada, la dirección, ruta y el tiempo de llegada; en ese orden de prioridad. Esto permite que se encuentren juntos y ordenados en la tabla los horarios de cada una de las líneas en cada parada. Posteriormente, cuando tengamos que computar el tiempo transcurrido entre dos pasos de vehículos, podremos hacerlo directamente a partir de esta ordenación; salvando una excepción

que se explicará más adelante, relacionada con los servicios que se prestan durante todo el día pero también llegan a la medianoche del día siguiente.

Una vez realizada la ordenación, usamos la tabla *trips.txt* para filtrar y generar nuevas tablas por cada tipo de servicio. Sabiendo que TUSSAM define cuatro tipos de servicio correspondientes a de lunes a jueves, viernes, sábados y festivos; lo que hacemos es filtrar la tabla original según si cada uno de los *trip\_id* presentes en ella pertenece a cada uno de los tipos de servicio. De este proceso de filtrado surgen 4 nuevos ficheros con datos, los cuales están ordenados siguiendo el criterio anterior. Estos ficheros servirán como entrada al último código empleado para la generación de estadísticas relacionadas con las paradas; que es el que generará los datos que luego presentaremos y analizaremos.

### 5.3.3 Generación de estadísticas

El código empleado para la generación de los resultados es el más complejo de los estudiados hasta ahora debido a la gran cantidad de datos y ficheros que maneja. Este código se encuentra en *statsParadas.ipynb*, y se divide en distintas etapas. **Primero se encarga de procesar y generar estadísticas para cada uno de los tipos de día, luego genera una tabla completa con toda la información y por último una tabla resumen con los datos más relevantes de cada parada.**

En el código, después de realizar las importaciones de las librerías, se define la función *computeFreqs*. Esta recibe como parámetros la ruta del fichero de entrada, que se corresponde con cada uno de los CSVs generados en el apartado anterior y la ruta donde se escribirá el fichero de salida. La misma tabla que se escribe en la ruta indicada es devuelta por la función. En ella, se agrupan las combinaciones únicas de identificadores de parada, ruta y dirección, y para cada una de estas claves se recuperan los tiempos asociados a ellas. Como estos valores ya fueron ordenados anteriormente, no necesitamos reordenarlos; pero sí corregir un problema observado después de ejecutarlo y analizar los resultados. Al estar ordenados de forma creciente, si una parada recibe servicio de un trayecto un día por la mañana y continúa recibiendo este servicio hasta pasadas las 12 de la noche del día siguiente, la hora posterior a las 00:00 aparecerá la primera de la lista, lo cual implica que los tiempos transcurridos entre cada paso por la parada no serán correctos. Para solucionar este problema, sabiendo que las rutas diurnas de TUSSAM comienzan a las 06:00, se ha optado por alterar el orden de la lista, de forma que todas las horas asociadas a un trayecto que ocurren antes de esa hora se mueven al final de la lista. Esta forma de hacerlo no altera los casos de las rutas nocturnas, por lo que es buena para nuestro caso. En la siguiente imagen se ejemplifica el algoritmo seguido para corregirlo.

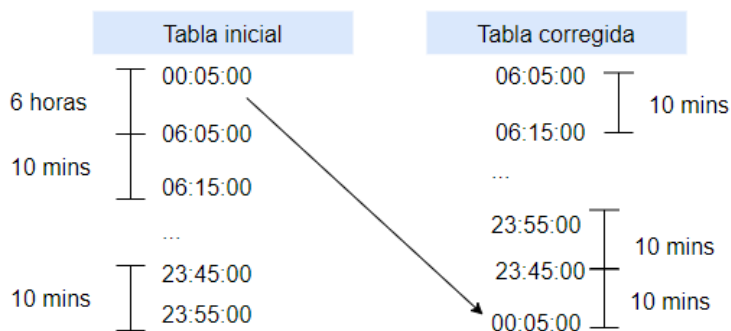


Figura 25. Método para corregir ordenación de horarios.

Por último, se calcula la diferencia entre tiempos consecutivos de la lista; y en el caso de que algún tiempo se haya computado en negativo (lo cual ocurre si el tiempo está entre las 23:XX y las 00:XX) se corrige la diferencia. Una vez hecho esto, se calcula la media de todos los tiempos calculados (en segundos) y se divide entre 60 para convertirlo a minutos. Ese resultado se escribe en una tabla que cuenta con las columnas *stop\_id*, *route\_id*, *direction\_id*, *trip\_headsign* y *avgFrequency*, en el último de los campos. Esta función se llama en 4 ocasiones, calculando todo lo anterior para los cuatro tipos de servicios prestados por TUSSAM. La expresión matemática utilizada para calcular los tiempos medios de paso de cada línea por cada parada se muestra a continuación, donde N es el número total de pasos de cada línea por la parada y los distintos  $t_i$  los horarios de estos pasos.

$$t_{avg} = \frac{1}{N-1} \sum_{i=1}^N t_{i+1} - t_i$$

Una vez obtenidos los resultados para cada tipo de día, pasamos a generar un único fichero con toda la información calculada anteriormente. Este tendrá los mismos campos relativos a la identificación de paradas y rutas; pero con cinco columnas para expresar las frecuencias de los cuatro tipos de servicio y la frecuencia media semanal. Además, añadiremos un campo que indicará el nombre de la parada. Para ello, creamos una copia de la tabla de los días laborables (ya que esta cuenta con todas las rutas que existen, algunas no tienen servicio los fines de semana) y le añadimos primero las nuevas columnas, que estarán vacías. A partir de ahí, iremos recorriendo con un bucle cada una de las entradas, añadiendo primero el valor del nombre de la parada (obtenido del GTFS) y luego añadiendo a cada columna la frecuencia correspondiente. Si una frecuencia no estuviera disponible, añadiremos un 0. Por último, calculamos la frecuencia media semanal, ponderando cada tipo de servicio por el número de días que se da en la semana:

$$avg_{week} = 4 * avg_{lab} + avg_{vie} + avg_{sab} + avg_{fest}$$

El resultado de esta parte de la ejecución es una tabla como la siguiente, con una entrada por cada terna parada, ruta y dirección que existe en el flujo GTFS.

stop_name	stop_id	route_id	direction_id	trip_headsign	avgFreqLab	avgFreqVie	avgFreqSab	avgFreqFest	avgFreqWeekly
Campana (Sierpes)	1	27	0.0	PLAZA DEL DUQUE	8.181102	8.181102	9.122807	12.844156	8.981782
Campana (Sierpes)	1	27	1.0	SEVILLA ESTE	8.349593	8.349593	9.20354	12.844156	9.113666
Menéndez Pelayo (Puerta Carmona)	2	1	1.0	HOSPITAL V.ROCIO	9.265487	9.265487	19.398148	17.051724	11.825329
Menéndez Pelayo (Puerta Carmona)	2	21	0.0	PLAZA DE ARMAS	13.289474	13.289474	15.605128	16.754386	14.115269
Menéndez Pelayo (Puerta Carmona)	2	83	1.0	PRADO	7.902256	7.902256	15.065714	12.3825	9.565642
...	...	...	...	...	...	...	...	...	...
Puerta Jerez	3006	60	1.0	SAN BERNARDO	9.436364	9.436364	10.71134	10.565217	9.779768
Prado San Sebastián	3007	60	1.0	SAN BERNARDO	9.436364	9.436364	10.71134	10.565217	9.779768
San Bernardo	3008	60	1.0	SAN BERNARDO	9.436364	9.436364	10.71134	10.565217	9.779768
San Bernardo	3009	60	0.0	PLAZA NUEVA	9.267857	9.267857	10.602041	10.451613	9.627563
San Bernardo	3009	60	1.0	SAN BERNARDO	9.436364	9.436364	10.71134	10.565217	9.779768

Tabla 5. Muestra de *avgFreqComplete.csv*.

Por último, crearemos una tabla con el tiempo medio de paso de cualquier autobús por una parada para cada día, así como el incremento porcentual que se da en los fines de semana y festivos con respecto de los días laborables; para tener una medida de cómo empeora el servicio los días de menor actividad. Para ello, en primer lugar creamos una nueva tabla con los nombres e identificadores de cada parada a la cual añadimos una serie de columnas vacías que posteriormente rellenaremos. Por otro lado, creamos otra tabla con las ternas de identificadores de parada, ruta y dirección.

Una vez hecho esto, volvemos a leer los ficheros con los datos de horarios de cada parada en cada día, y vamos almacenando en distintas variables el número de viajes de cada servicio que se hacen y el tiempo medio asociado a ese servicio para cada día. Cuando completamos esto para cada parada, añadimos a la nueva tabla creada el tiempo medio general de cada día de la semana; el cual está ponderado según la cantidad de viajes de cada línea que pasan por allí:

$$avg_{total} = \frac{1}{total\_viajes} \sum_{i=1}^N n\_viajes_i * freq_i$$

Para finalizar este código, se calcula el porcentaje de incremento del tiempo en cada parada de sábado y domingo con respecto de los días de diario. En algunos casos este valor es negativo, debido a que existe mayor frecuencia de paso los fines de semana que los de diario. En general, esto se debe a que la parada ofrece un mayor número de servicio de rutas nocturnas que diurnas. Además, existen casos de rutas especiales que no dan servicio los fines de semana. En ese caso, el valor del incremento será nulo.

Posteriormente, en el código del notebook *combStops.ipynb* añadimos a la tabla mostrada en la figura 34 las coordenadas de cada parada, para el posterior análisis de los datos.

	stop_name	stop_id	avgFreqLab	avgFreqVie	avgFreqSab	avgFreqFest	increaseSab	increaseFest
	Campana (Sierpes)	1	8.264011	8.264011	9.162997	12.844156	10.878331	55.422788
	Menéndez Pelayo (Puerta Carmona)	2	11.013345	11.416946	18.618760	17.008641	69.056357	54.436647
	Menéndez Pelayo (Puerta de La Carne)	3	10.569357	10.826993	18.069714	16.377407	70.963227	54.951783
	Menéndez Pelayo (Juzgados)	4	15.586138	15.931432	20.212818	21.337733	29.684584	36.901992
	Avda. María Luisa (Teatro Lope de Vega)	5	8.950534	9.292929	15.764239	17.707886	76.126247	97.841675
	...	...	...	...	...	...	...	...
	Archivo de Indias	3005	9.436364	9.436364	10.711340	10.565217	13.511312	11.962805
	Puerta Jerez	3006	9.436364	9.436364	10.711340	10.565217	13.511312	11.962805
	Prado San Sebastián	3007	9.436364	9.436364	10.711340	10.565217	13.511312	11.962805
	San Bernardo	3008	9.436364	9.436364	10.711340	10.565217	13.511312	11.962805
	San Bernardo	3009	9.351358	9.351358	10.656413	10.508111	13.955780	12.369896

Tabla 6. Muestra de *frecuencias.csv*.

### 5.3.4 Representación de los resultados

Una vez obtenidos los resultados anteriores, además de observarlos en bruto podemos representarlos para que nos sean más fáciles de analizar. **Para ello, representaremos sobre distintos mapas de la ciudad cada uno de los valores calculados** al final del apartado anterior, que se corresponden con la muestra de la figura 34; **donde cada parada tendrá colores asociados en función del tiempo medio de paso de los autobuses. Estos mapas estarán en formato HTML**, lo cual permitirá la interacción de los usuarios con ellos y que sean publicados en una web con todos los resultados. Para esta tarea, utilizaremos la librería de Python llamada *Folium*; que permite crear mapas con distintos estilos y añadir marcadores personalizables, generando un fichero de salida en HTML. La generación de los resultados se ha hecho en el código *displayStops.ipynb*.

Una vez importadas las librerías a utilizar, en primer lugar creamos el mapa, el cual está centrado en unas coordenadas del centro de la ciudad y tendrá un nivel de zoom para que de un vistazo inicial se pueda ver aproximadamente toda la ciudad. Además, usaremos un mapa con versión oscura para que los colores sean realizados. Después de esto, importamos el *feed* GTFS para poder acceder a las coordenadas de las paradas.

A continuación, definimos dos funciones que nos ayudarán con la tarea de crear el mapa y representarlo adecuadamente. La primera de ellas es para asociar un color a cada parada, de forma que el punto en el mapa será más verde cuanto menor sea el tiempo (o el incremento porcentual de tiempos) y más rojo cuanto mayor sea ese tiempo. Esta función será llamada desde la segunda que hemos definido, que sirve para añadir los marcadores al mapa. A esta segunda función se le pasa como parámetro el nombre de la columna de *frecuencias.csv* que queremos representar, el objeto donde hemos guardado el mapa con *Folium* y el *dataframe* obtenido del *feed* GTFS con la tabla de las paradas, que contiene las coordenadas. La tabla con la información calculada y la que contiene las coordenadas se unirán en una sola; y esta se irá recorriendo fila por fila para asociar un color y añadir al mapa cada punto a representar. En la última celda del documento se llama a esta función, se guarda el mapa en un archivo HTML y se representa en el Notebook añadiéndole un título. Este código lo ejecutaremos varias veces cambiando el nombre de la columna que queremos representar; con el objetivo de guardar los ficheros para su posterior publicación. En el siguiente código se muestra un ejemplo del resultado obtenido.

```

import folium
import pandas as pd
import gtfs_kit as gk

folium_map = folium.Map(location=[37.384642436284395, -5.971477021097102],
                        zoom_start=14,
                        tiles="CartoDB dark_matter")

path = 'TUSSAM_COMP.zip'
feed = (gk.read_feed(path, dist_units='km'))
stops = feed.stops

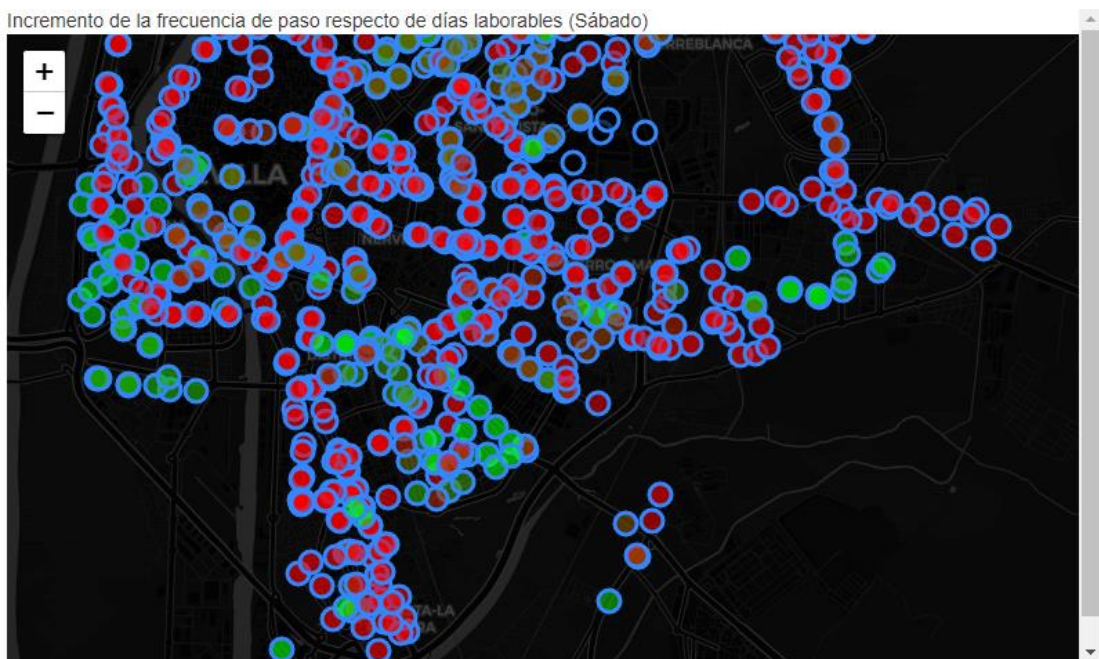
def generate_color(value):
    red_value = int(min(255, max(0, (value) * 7.125)))
    green_value = int(min(255, max(0, (30 - value) * 7.125)))
    blue_value = 0
    color = "#{:02X}{:02X}{:02X}".format(red_value, green_value, blue_value)
    return color

def addMarkers(plottingColumn,map,stopsDf):
    df = pd.read_csv('results/frecuencias/frecuencias.csv')
    stopsDf['stop_id'] = stopsDf['stop_id'].astype(int)
    stops = pd.merge(df,stopsDf,how="left",on="stop_id")[['stop_id','stop_name_x',plottingColumn,'stop_lat','stop_lon']].copy()
    stops.rename(columns={"stop_name_x": "stop_name"}, inplace = True)
    for index,row in stops.iterrows():
        c = generate_color(row[plottingColumn])
        if (plottingColumn.startswith('increase')):
            string = '<h7><b>'+row['stop_name']+'</b></h7>'+</h5> </br></br>Incremento de tiempo: </h5> <h4>'+str(round(row[plottingColumn],2))+'%</h4>'
        else:
            string = '<h7><b>'+row['stop_name']+'</b></h7>'+</h5> </br></br>Frecuencia media de paso [minutos]: </h5> <h4>'+str(round(row[plottingColumn],2))+'</h4>'
        marker = folium.CircleMarker(location=[row['stop_lat'], row['stop_lon']],radius = 8,popup=string, fill=True, fill_color=c, fill_opacity=0.6)
        marker.add_to(map)
    return map

s1 = addMarkers('increaseSab',folium_map,stops)
s1.get_root().html.add_child(folium.Element('Incremento de la frecuencia de paso respecto de días laborables (Sáb)'))
s1.save('results/frecuencias/maps/increaseSab.html')

s1

```



Código 5.5. Notebook displayStops.ipynb para la generación y guardado de mapas de paradas.

## 5.4. Cálculo de estadísticas de líneas

A continuación se procede a explicar el procedimiento seguido para el cálculo de las estadísticas de las líneas y rutas del sistema de autobuses de la ciudad de Sevilla. Para ello, se hace uso de los códigos incluidos en *statsViajes.ipynb* y *statsLineas.ipynb*, los cuales se encuentran de forma completa en el Anexo B de este documento. Posteriormente se explicará el código empleado para la representación de los resultados en un formato similar al que se empleó en el caso anterior.



### 5.4.1 Estadísticas de viajes

Como paso previo al cálculo que vamos a realizar, **vamos a computar las estadísticas de cada viaje registrado en el feed GTFS**, haciendo uso de la función `compute_trip_stats` ya implementada en la librería `gtfs-kit`. Esta función nos resultará de gran ayuda para nuestro trabajo y nos ahorrará el tiempo de calcular manualmente algunos valores relacionados con las líneas. Esta función, dado un conjunto GTFS sin errores y con el fichero `shapes.txt` incluido, nos devolverá un `dataframe` con parámetros de cada viaje del sistema, como la distancia recorrida, la hora de inicio y final del servicio, la primera y última parada o la velocidad que lleva. Es por esto que, antes de comenzar a trabajar, se generó este fichero con las formas de las líneas. Además, incluye otra información más relacionada con la propia identificación de los viajes, como la ruta o el sentido al que pertenece.

Una vez obtenido esta tabla con los resultados de las estadísticas de los viajes, no obstante; **observamos datos extraños, los cuales podrían ser considerados outliers**, es decir, valores anormales e inconsistentes con el resto de la serie de datos. Estos, en nuestros datos, incluyen velocidades superiores a los 40 km/h, alcanzando velocidades de más de 130 km/h; y líneas de servicio con tiempos totales de menos de 5 minutos. Como decisión para eliminar estos fenómenos, la decisión tomada ha sido eliminar del `dataframe` todos los viajes que alcanzasen velocidades de más de 50km/h (velocidad máxima de vehículos por la ciudad) y con trayectos totales de duración menor de 6 minutos. De esta forma, haciendo una breve comprobación de la velocidad media calculada para todas las rutas, obtenemos 13.8 km/h; algo superior a la velocidad comercial real anunciada por TUSSAM a principios de 2023 para el año anterior, que fue de 12.7 km/h. [142] Con estos errores detectados y solucionados, el resultado ha sido escrito en varios ficheros para su posterior procesamiento con el objetivo de calcular las estadísticas generales para las líneas; donde cada uno de los ficheros se corresponde con los distintos días en los que se da servicio a lo largo de la semana.

trip_id	route_id	route_short_name	route_type	direction_id	shape_id	num_stops	start_time	end_time	start_stop_id	end_stop_id	is_loop	duration	distance	speed	trip_headsign
4100101004	1	01	3	0	111	21	06:06:00	06:36:00	430	252	0	0.500000	9.033379	18.066758	POLIGONO NORTE
3900103004	1	01	3	0	111	21	06:19:00	06:50:00	430	252	0	0.516667	9.033379	17.483960	POLIGONO NORTE
4400103004	1	01	3	0	111	21	06:19:00	06:50:00	430	252	0	0.516667	9.033379	17.483960	POLIGONO NORTE
3900105004	1	01	3	0	111	21	06:30:00	07:01:00	430	252	0	0.516667	9.033379	17.483960	POLIGONO NORTE
4400105004	1	01	3	0	111	21	06:30:00	07:01:00	430	252	0	0.516667	9.033379	17.483960	POLIGONO NORTE
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3608701053	87	B3	3	1	8712	17	22:55:00	23:15:00	923	406	0	0.333333	5.049066	15.147199	SANTA CLARA
4108701049	87	B3	3	1	8712	17	23:20:00	23:41:00	923	406	0	0.350000	5.049066	14.425904	SANTA CLARA
3908701047	87	B3	3	1	8712	17	23:23:00	23:42:00	923	406	0	0.316667	5.049066	15.944421	SANTA CLARA
4408701047	87	B3	3	1	8712	17	23:23:00	23:42:00	923	406	0	0.316667	5.049066	15.944421	SANTA CLARA
3608701055	87	B3	3	1	8712	17	23:33:00	23:53:00	923	406	0	0.333333	5.049066	15.147199	SANTA CLARA

Tabla 7. Muestra de la tabla resultado de ejecutar `statsViajes.ipynb`.

### 5.4.2 Estadísticas de líneas

**El objetivo una vez calculados los valores relevantes de los viajes será agrupar y calcular esos valores medios para las distintas rutas a las que se da servicio en la empresa TUSSAM;** también según los tipos de día en los que se realiza cada trabajo. Para ello, haremos uso del código `statsLineas.ipynb`; donde todas las tareas se realizan en una función auxiliar definida que se encarga de realizar todos los cálculos y construir y guardar el resultado. Esta función es llamada con dos parámetros, indicando el fichero de entrada con los datos obtenidos en el apartado anterior y la ruta del fichero que servirá como salida.

En la función auxiliar definida, en primer lugar se obtiene una lista de todas las líneas (separadas según su sentido) con informaciones relativas a la forma asociada en `shapes.txt`, el nombre que recibe popularmente y el destino al que se dirige. Posteriormente, sobre esa línea se irá iterando, buscando las filas del fichero de entrada que pertenecen a ella; calculando los tiempos de servicio inicial y final de cada línea y el tiempo medio entre salidas de cada trayecto, empleando una fórmula similar a la empleada para calcular los tiempos medios de paso en cada parada.

$$t_{salidas\_avg} = \frac{1}{N-1} \sum_{i=1}^N t_{salida\ i+1} - t_{salida\ i}$$

El método para calcular las diferencias de tiempos es igual que el empleado en el cálculo de las estadísticas de las paradas; calculando el tiempo medio en segundos y posteriormente la media pasándola a minutos. Es importante notar que, aunque ya se explicó en el cálculo de los tiempos medios de paso por las paradas, aquí

también es necesario incluir una corrección que suma un día a los tiempos negativos, que se darán en el caso de que computemos la diferencia entre horas antes y después de la medianoche.

Después de trabajar con las horas de servicio de cada línea calculamos la duración media de cada ruta en minutos, el número medio de paradas (es necesario que sea número medio en tanto que hay líneas que modifican su recorrido dependiendo de la hora o el día en el que se encuentre, cambiando así el número de paradas). Por último, se calcula la velocidad media de cada ruta. Esta tabla de resultados se guarda como un CSV en la ruta específica y se devuelve como resultado de la función. Se muestra un ejemplo del resultado a continuación.

route_id	direction_id	trip_headsign	shape_id	route_short_name	startOfService	endOfService	avgDepartureTime	avgDuration	nStops	avgSpeed
1	0	POLIGONO NORTE	111	01	06:06:00	23:11:00	19.711538	39.132075	21.000000	13.939080
1	1	HOSPITAL V.ROCIO	112	01	06:05:00	23:12:00	19.377358	40.351852	24.981481	10.635908
2	0	CARTUJA	211	02	05:46:00	23:02:00	11.908046	59.750000	32.454545	14.706386
2	1	HELIOPOLIS	212	02	05:52:00	23:02:00	11.704545	57.089888	30.719101	16.110876
3	0	BELLAVISTA	311	03	06:00:00	22:47:00	19.365385	71.301887	30.000000	16.708949
...	...	...	...	...	...	...	...	...	...	...
119	1	TORREBLANCA	11912	B4	06:30:00	22:36:00	26.833333	64.324324	44.756757	13.000878
125	0	PRADO	12511	38A	06:00:00	23:01:00	72.928571	32.733333	14.000000	15.618773
125	1	PITAMO P.OLAVIDE	12512	38A	06:33:00	22:33:00	73.846154	31.571429	13.000000	16.477911
127	0	SANTA JUSTA	12711	LS	09:00:00	21:25:00	20.135135	36.473684	12.000000	15.370585
127	1	BELLAVISTA	12712	LS	09:11:00	22:00:00	20.236842	36.512821	10.000000	16.985352

Tabla 8. Muestra de una de las tablas resultado de ejecutar *statsLineas.ipynb*.

#### 5.4.2.1 Cálculo del servicio en horas punta

Por último, antes de pasar a la representación de los resultados, repetiremos todos los cálculos anteriores para únicamente las horas punta, con el objetivo de estudiar el servicio prestado por TUSSAM en los momentos de mayor número de desplazamientos del día. En esta ciudad, estos momentos de mayor tensión se corresponden con las horas desde las 8:00 hasta las 9:00 y desde las 14:00 hasta las 15:00, donde se inician el 15.6% de todos los viajes realizados a lo largo de un día. [101]

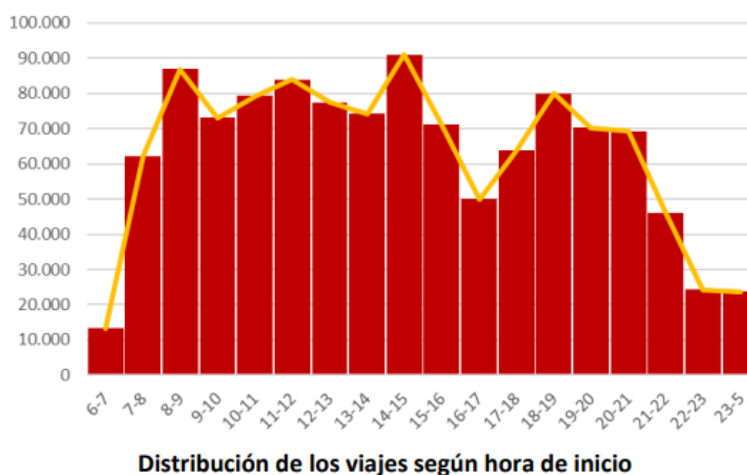


Figura 26. Distribución de los viajes en Sevilla según hora de inicio.

Para calcular las estadísticas en horas punta, partiremos del resultado obtenido en *statsViajes.ipynb*, que producía ficheros para los viajes realizados en cada uno de los distintos tipos de día de la semana. A partir del fichero de texto con los resultados para los días laborables nos quedamos únicamente con los viajes introducidos entre las 7:30 y las 9:00 y las 13:30 y 15:00 (en ficheros distintos, para comparar también los resultados entre los servicios ofrecidos en las dos mayores horas punta); y ejecutaremos el código mostrado en el apartado 5.4.2; para posteriormente representar los resultados y poder compararlos. Para generar un fichero CSV con la información de los días laborables en general y de cada uno de los picos de demanda y poder contrastar los resultados se ha empleado el fichero *compareLabsPeak.ipynb*, incluido en el Anexo B.

### 5.4.3 Representación de los resultados

Al igual que en la representación de los resultados asociados a las paradas de autobús, en este caso **utilizaremos Folium para representar sobre distintos mapas interactivos en formato HTML todas las rutas del sistema, con códigos de colores según la velocidad que alcanzan estas rutas**. Esta tarea se encuentra hecha en el código `displayLinesMap.ipynb`. En primer lugar crearemos el mapa con los mismos parámetros que anteriormente, importamos el `feed` GTFS y definimos una función para establecer el código de colores análoga a la que escribimos anteriormente; alterando la lógica de asignación de verde y rojo, en tanto que ahora los valores altos de velocidad deben corresponderse con colores más verdes.

Después, definimos la función encargada de añadir al mapa las distintas rutas. En este caso, cada línea no se representa como un conjunto aislado de puntos, sino como dos listas de coordenadas, una lista para la latitud y otra para la longitud. Para obtener estos valores, accedemos a la tabla de `shapes` GTFS que se pasa como parámetro con cada uno de los identificadores de la forma que se incluyeron en la tabla de la figura 36. Después, según la velocidad media de esa ruta, asignamos un color, y por último definimos el mensaje que aparecerá en el mapa al pulsar en cada línea y añadimos todos los puntos al mapa usando la función de la librería de dibujado de mapas llamada `Polyline`, que dibujará líneas continuas del color especificado pasando por todos los puntos indicados como parámetro.

Después de llamar a esta función, guardaremos el mapa en formato HTML y lo dibujamos como resultado del Notebook. Con el objetivo de dibujar también sobre las líneas la frecuencia de salidas de cada una de las rutas, se ha usado el mismo notebook, alternando la forma en la que se asignan los colores para asignar el verde a los menores tiempos y basándose en el valor `avgDepartureTime` en vez de `avgSpeed` de la tabla para calcularlo. No obstante, para este último caso, se han excluido las rutas nocturnas de los mapas representados, al tener cadencias muy definidas y no ser servicios esenciales para el día a día de la ciudad.

```
import folium
import pandas as pd
import gtfs_kit as gk

folium_map = folium.Map(location=[37.384642436284395, -5.971477021097102],
                        zoom_start=14,
                        tiles="CartoDB dark_matter")

path = 'TUSSAM_COMP.zip'
feed = (gk.read_feed(path, dist_units='km'))
shapes = feed.shapes

def generate_color(value):
    green_value = int(min(255, max(0, (value) * 7.125)))
    red_value = int(min(255, max(0, (30 - value) * 7.125)))
    blue_value = 0
    color = "#{:02X}{:02X}{:02X}".format(red_value, green_value, blue_value)
    return color

def addLines(map, shapesDf, routesFile):
    dfRoutes = pd.read_csv(routesFile)

    for index, row in dfRoutes.iterrows():
        shape_id = str(row['shape_id'])
        lat = shapesDf.loc[(shapesDf['shape_id'] == shape_id), 'shape_pt_lat'].tolist()
        long = shapesDf.loc[(shapesDf['shape_id'] == shape_id), 'shape_pt_lon'].tolist()
        speed = row['avgSpeed']
        c = generate_color(speed)

        string = '<h7><b>Línea ' + str(row['route_short_name']) + ': ' + row['trip_headsign'] + '</b></h7>' + '</h5> </br></br>Velocidad '
        shape = []
        for i in range(len(lat)):
            shape.append([lat[i], long[i]])
        folium.PolyLine(shape, color=c, popup=string).add_to(map)

    return map
```




```

m = addLines(folium_map,shapes,'results/routes/routesFest.csv')
m.get_root().html.add_child(folium.Element('Velocidad media de las rutas (Festivos)'))
m.save('results/routes/maps/speedFest.html')

```

m

Velocidad media de las rutas (Festivos)



Código 5.6. Notebook displayLinesMap.ipynb para la generación y guardado de mapas relativos a la velocidad de las líneas.

## 5.5. Resultados

### 5.5.1 Presentación de resultados

Para la **presentación de los resultados, la consulta de estos y la posibilidad de interactuar con los distintos mapas se utiliza el servicio *GitHub Pages***, que permite de forma gratuita el alojamiento de páginas webs, aportando además una gran facilidad de cara a la publicación de cambios, que se hace de forma automática al subir los cambios al repositorio asociado a la web. Esta web servirá como web del proyecto y permitirá que cualquier ciudadano pueda consultar los resultados obtenidos, abriendo así muchos de los datos al público, lo cual es novedoso en esta ciudad, en tanto que, como ya se ha comentado, existe poca transparencia relativa a la publicación de los datos de transporte. Además de los mapas, a través de este portal se pueden consultar las distintas tablas de resultados obtenidas de los códigos incluidos en los Anexos A y B. Para hacer esto último, se ha hecho uso de un repositorio de GitHub que permite incluir directamente tablas en CSV a la web utilizando JavaScript, así como realizar búsquedas en las tablas o cambiar el orden de estas. El repositorio empleado se encuentra en [este enlace](#).

La web creada se puede consultar a través de [este enlace](#).

### 5.5.2 Análisis de los resultados

Atendiendo a las frecuencias medias de paso, vemos que existe una distribución muy desigual de los tiempos medios de paso según la zona de la ciudad en la que nos encontremos. Si bien en el Centro y barrios residenciales alrededor de este, como Los Remedios, Triana, NerviÓN o La Macarena nos encontramos con tiempos de paso medios que oscilan entre los 9 y los 11 minutos los días laborables, siendo valores que se encuentran en los más bajos de los calculados. En general este tiempo aumenta según vamos aumentando la distancia con el centro de

la ciudad. Además, otras zonas como la Isla de La Cartuja, a pesar de contar con pocas líneas de autobús dirigidas hasta esa zona, cuenta con tiempos medios de paso de menos de 7 minutos y medio. Otra zona que a pesar de la distancia se encuentra con buenos tiempos medios de paso es Pino Montano, que acumula en este barrio las 3 paradas con los tiempos de paso más bajos en los días laborables y las dos con autobuses más frecuentes los sábados. Respecto de los domingos, las paradas con mejores tiempos de paso son aquellas que dan servicio al tranvía.

La nota negativa de la observación de estos mapas se centra en algunos barrios o zonas de la ciudad, que se encuentran infraconectadas. Es el caso de Tablada, Sevilla Este, Torreblanca, Santa Clara o el Tiro De Línea. En todos estos casos, encontramos una gran cantidad de paradas con tiempos superiores de espera de más de 13/14 minutos; llegando incluso a tiempos medios de más de 20 y 30 minutos. En días laborables, podríamos decir que los comportamientos observados entran dentro de lo esperado, salvo para los casos concretos mencionados anteriormente. Especialmente relevante es el caso de la Universidad Pablo de Olavide, que recibe frecuencias de paso de casi un autobús a la hora, siendo un servicio que se podría mejorar con transportes exprés a otros nodos de la ciudad. En esta ubicación, no obstante, existe servicio de Metro.

Si estudiamos la relación entre los tiempos medios de espera en las paradas con respecto de la distancia al centro, en un gráfico de dispersión como el que se muestra en la siguiente figura, vemos que efectivamente existe una tendencia que relaciona la distancia con los tiempos medios de paso. Si bien es cierto que la mayor parte de paradas de la red se encuentran en general por debajo de los 30 minutos de paso, conforme aumenta la distancia al centro encontramos puntos mucho más dispersos y con tiempos de espera sensiblemente mayores. En el gráfico, la distancia se ha medido como la distancia euclídea de cada parada del sistema al Prado de San Sebastián, sobre la latitud y la longitud asociadas a cada punto. Estudiando esta misma relación para el resto de días de la semana, esta tendencia se mantiene; aunque en general los tiempos de paso son mayores, existiendo mucha más dispersión en aquellos lugares más alejados del centro de la ciudad. De esta gráfica cabe destacar que los puntos situados exactamente en los 60 minutos se corresponden con los autobuses nocturnos, que cuentan con esa frecuencia las noches laborables, y que los puntos que reciben tiempos aún mayores de espera son los servidos únicamente por las líneas C6A y C6B, 38A y 35; que van a Valdezorras y el Gordillo para las primeras, la Universidad Pablo de Olavide la segunda y Palmas Altas en la última de ellas; todas ubicaciones muy alejadas del centro de la ciudad.

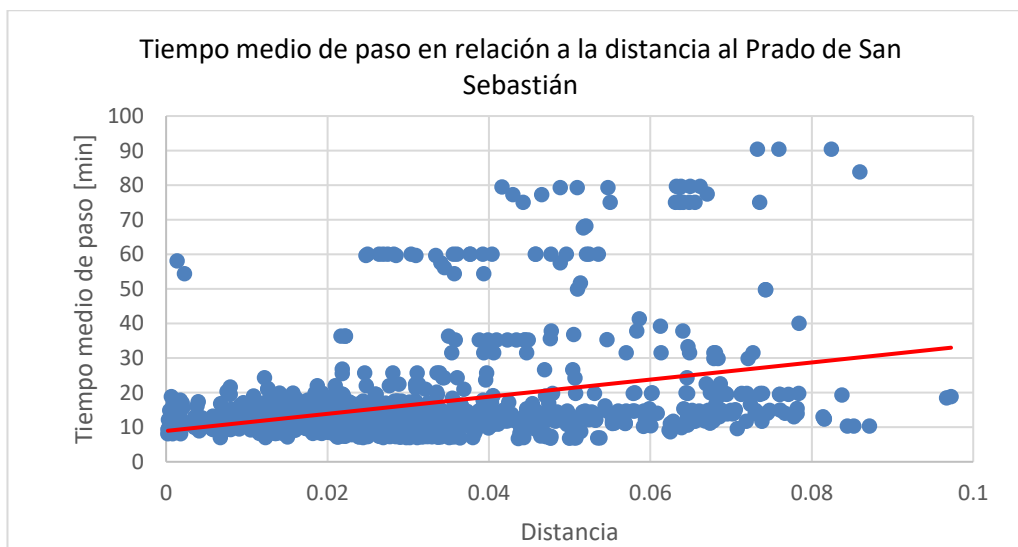


Figura 27. Correlación entre la distancia al Prado de San Sebastián y el tiempo medio de paso en laborables.

Los resultados del análisis se vuelven mucho más dramáticos cuando observamos los mismos resultados para días no laborables; y estudiamos el porcentaje de incremento de tiempos observado. Para ello, miramos en primer lugar los datos relativos a los sábados; que en general arrojan un empeoramiento generalizado de todos los tiempos medios de paso, que se hace más evidente en las zonas que antes destacábamos por su buen funcionamiento, donde los tiempos aumentan en muchos casos más de un 50%, llegando en algunas paradas hasta tiempos de más del doble. En aquellas paradas donde el tiempo medio de paso ya era alto, los incrementos de tiempo en porcentaje no se hacen tan malos, debido a que el servicio que se da ya es deficiente de forma habitual. En algunos casos observamos mejorías, las cuales tienen que ver con la pertenencia de esas paradas a

líneas nocturnas, las cuales los fines de semana ofrecen mucho mayor servicio que los días laborables. En este sentido, si observamos el caso de los domingos vemos que el servicio ofrecido por TUSSAM es claramente peor que en los sábados; con tiempos de paso de en torno al doble en la mayor parte de las paradas de la red que en días laborables. De nuevo en este caso, el mayor incremento se da en las zonas que reciben un servicio mejor durante la semana. Observando esto en un gráfico, vemos como efectivamente, tanto los sábados como los domingos, la mayor parte de las paradas que se encuentran por encima de la línea de tendencia son aquellas que menores tiempos de espera tienen los días laborables. Además, los incrementos de espera los domingos son sensiblemente mayores que los sábados, con más de un 20% de diferencia, en general, entre ambas rectas de mejor ajuste.

### Incremento del tiempo de paso [%] respecto del tiempo de paso en días laborables

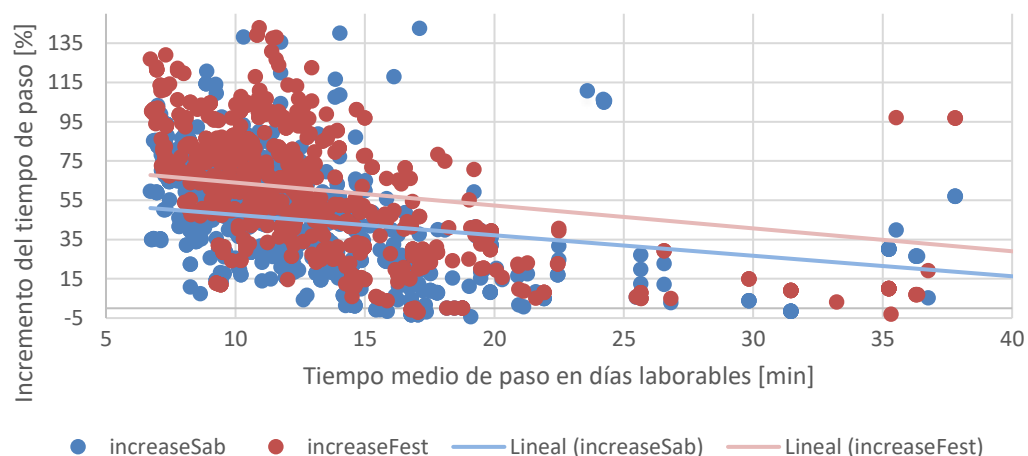


Figura 28. Aumento de los tiempos de espera en fines de semana respecto del tiempo de paso en laborables.

Si atendemos ahora al repaso de las líneas según sus tiempos medios entre salidas, observamos de nuevo que los tiempos irán aumentando conforme nos alejamos del centro de la ciudad. Es especialmente relevante notar que, en general, se observan tiempos entre salidas mucho mayores entre aquellas líneas que tienen origen o destino fuera del límite marcado por la SE-30 y la Ronda Urbana Norte, a excepción de las líneas que se dirigen a Pino Montano. Para las líneas interiores a los límites antes especificados sí que tenemos tiempos medios de menos de 10 minutos en general; y con cadencias mucho menores en las horas punta. La línea con mayor frecuencia los días laborables es la que conecta Pino Montano con el centro de la ciudad, que es la número 12. La siguiente con mayor cadencia es el C1, una de las circulares que vertebran la ciudad. Respecto de los sábados, la línea 27 que conecta La Campana con Sevilla Este es la más frecuente. Los domingos y festivos, el tranvía es la línea con un menor tiempo medio entre salidas.

Comparando de nuevo los resultados obtenidos en días laborables con aquellos correspondientes los fines de semana vemos que se observa la misma tendencia; con los mayores cambios a peor en las líneas con menores tiempos durante la semana y en las líneas exprés mencionadas anteriormente. Aunque en los mapas no se representan, las grandes beneficiadas y que reducen más sus tiempos los viernes y sábados en comparación con el resto de los días son las rutas nocturnas; que en general salen con el doble de frecuencia en estas noches, donde la demanda es, evidentemente, mucho mayor. Además de ellas, la línea especial para el Aeropuerto de Sevilla mantiene su frecuencia independientemente del día de la semana estudiado.

Observando los mapas relacionados con las líneas también observamos el carácter fuertemente radial del sistema de transportes gestionado por TUSSAM, con gran variedad de alternativas para moverse entre los barrios que quedan en el centro de la ciudad o sus alrededores, así como para ir de prácticamente cualquier punto de la ciudad al centro de Sevilla; pero muy pocas opciones para moverse entre distintos barrios de la periferia, lo que hace que estos movimientos de personas sean mucho más ineficientes; y distancias geográficamente cercanas se hagan muy distantes si tienen que recorrer en transporte público.

Si estudiamos los valores relacionados con la velocidad de cada una de las líneas, observamos resultados contrarios a los vistos anteriormente. En este caso, las que mejor desempeño tienen son aquellas que desarrollan

parte de la ruta alejados del centro de la ciudad. Esta circunstancia no es casual, en tanto que aquellas líneas más exteriores se ven menos afectadas por los frecuentes atascos que tienen lugar en Sevilla, circularán por vías de mayor velocidad, en algunos casos carreteras no urbanas, y también encontrarán menos semáforos. Además, se observa viendo los mapas que las líneas nocturnas (aquellas cuyo nombre empieza por A) obtienen velocidades mucho más altas que el resto, de nuevo debido al menor tráfico a estas horas y a que la cantidad de personas que utilizan este servicio es mucho menor, lo que implica menor tiempo en cada parada. De hecho, la línea más rápida de todo el sistema es la A4, una de las nocturnas, siendo la única que supera los 35 km/h. En contra posición, la línea C4 en dirección a Plaza de Armas es la más lenta, con una velocidad de algo más de 6.4 km/h de media. Esta línea, que se caracteriza por circular alrededor del centro, se ve afectada por la topología tan peculiar de esta zona de la ciudad.

En general, vemos que el día de la semana que más rápidas son las líneas es los domingos, hecho que puede venir influenciado por la menor cantidad de tráfico en las calles y menos cantidad de personas a las que transportar, con casi 16.2 km/h. Esta velocidad es algo superior a la que se alcanzan los sábados, que se reduce hasta los 15.3 km/h. El resto de días de la semana, donde las calles son mucho más caóticas, se reduce la velocidad media, con 14.4 km/h de lunes a jueves y marcando el peor dato los viernes, con 14.2 km/h.

Aunque podríamos pensar que existe una correlación entre el número de paradas y la velocidad, de forma que a mayor número de paradas la velocidad se hace menor, podemos ver en un simple gráfico que esta relación no es tan evidente. Si bien es cierto que las rutas con velocidades más altas tienen menos de la mitad de paradas que la línea que más tiene, aquellas con un mayor número de paradas alcanzan velocidades similares e incluso superiores a la media. Trazando una recta de tendencia, vemos que esta sí que indica que a mayor número de paradas la velocidad decrece, pero esta variación no es muy grande. Este mismo comportamiento se observa si buscamos la relación entre la velocidad y la duración de los viajes. Respecto de este último valor, el de la duración, la línea que menos duración tiene es el tranvía T1, con un recorrido algo superior a los 12 minutos, mientras que la línea más larga es la número 3, que llega hasta casi los 80 minutos en hacer su recorrido completo.

En contraposición a lo anterior, si realizamos la misma evaluación de correlación entre la distancia recorrida por las líneas y su velocidad, sí que observamos una clara tendencia; de forma que al aumentar la distancia recorrida también lo hace la velocidad media de la línea. Esta observación refrenda el pensamiento inicial; y es que cuando las líneas se alejan del centro de la ciudad, su velocidad es en general mayor.

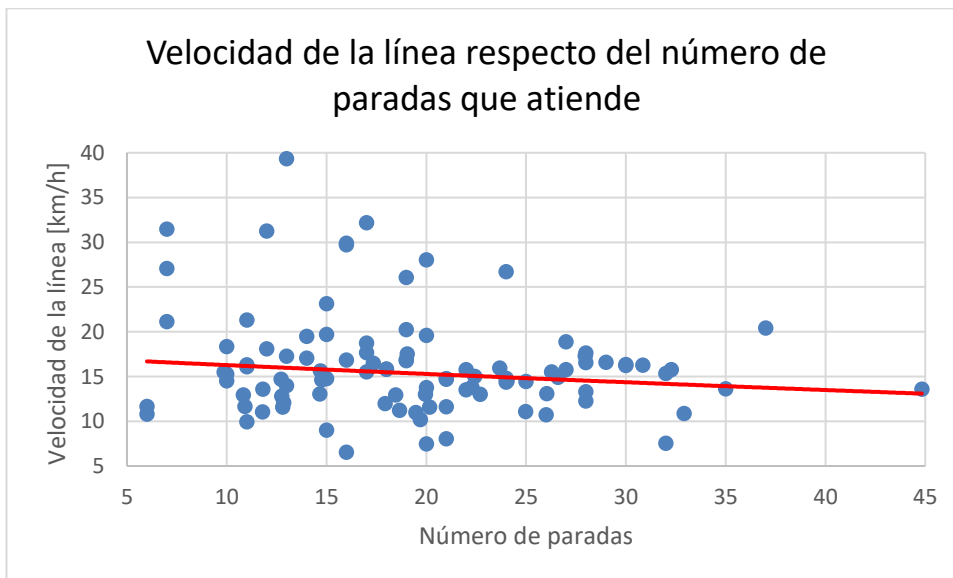


Figura 29. Correlación entre el número de paradas y la velocidad de cada línea.

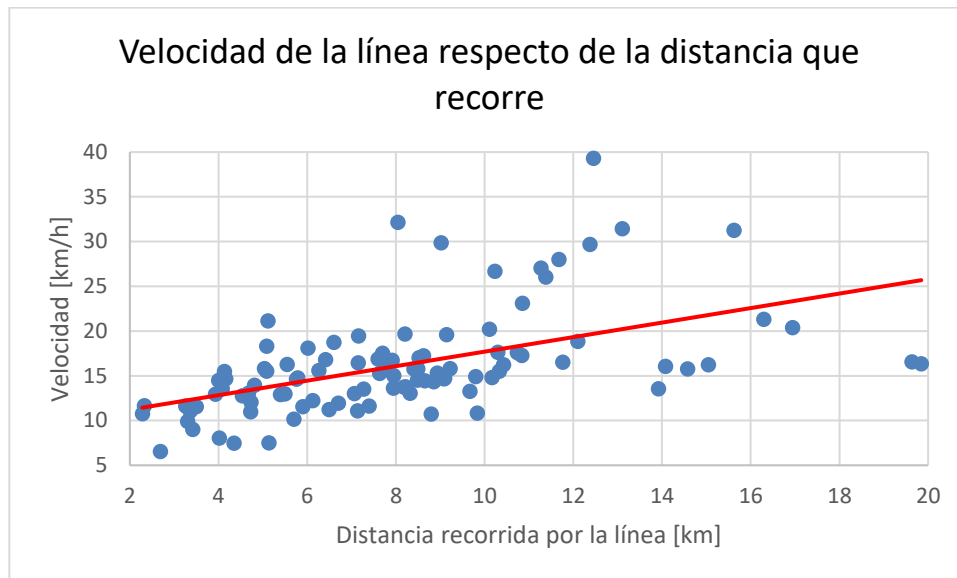


Figura 30. Correlación entre la distancia recorrida y la velocidad de cada línea.

Por último, contrastaremos los niveles de servicio ofrecidos por las distintas líneas en media durante todo el día con los ofrecidos durante las horas de mayor demanda, que son de 08:00 a 09:00 y de 14:00 a 15:00.

Atendiendo primero a la evaluación realizada para el servicio que se presta en la primera hora punta de la jornada, observamos que la mayor parte de líneas reducen los tiempos de espera entre autobuses de una misma ruta. En este segmento del día, la ruta con mayor frecuencia de paso es el C2, que tiene en torno a 4.8 minutos de espera entre pasos del mismo vehículo, siendo la única línea con tiempos de paso inferiores a los 5 minutos. Además, hasta 14 recorridos distintos obtienen tiempos de espera inferiores a los 6 minutos, algo que no ocurría a lo largo del día, con 6.75 minutos para la ruta que menos tardaba en pasar, en media. Si observamos la mejoría del servicio en términos relativos, la línea que mejores resultados obtiene es la que comunica el Prado de San Sebastián con la Universidad Pablo de Olavide, que reduce sus tiempos de espera medios en un 114% y 125% para los dos sentidos; dando un servicio más aceptable que el resto del día para la comunicación con este centro educativo; con tiempos de paso a esta hora de en torno a los 20 minutos. Por otro lado, sí que existen algunas rutas que aumentan los tiempos medios de espera en esta hora punta, aunque en general el aumento es de menos del 5%. La única línea que supera este dato es la número 25, donde aumenta el tiempo de espera en casi un 10%. Por último, cabe mencionar que la mejora media de los tiempos en estas horas es de un 20.55%, pasando de un tiempo medio de algo menos de 15 minutos a uno de menos de 12 minutos y medio.

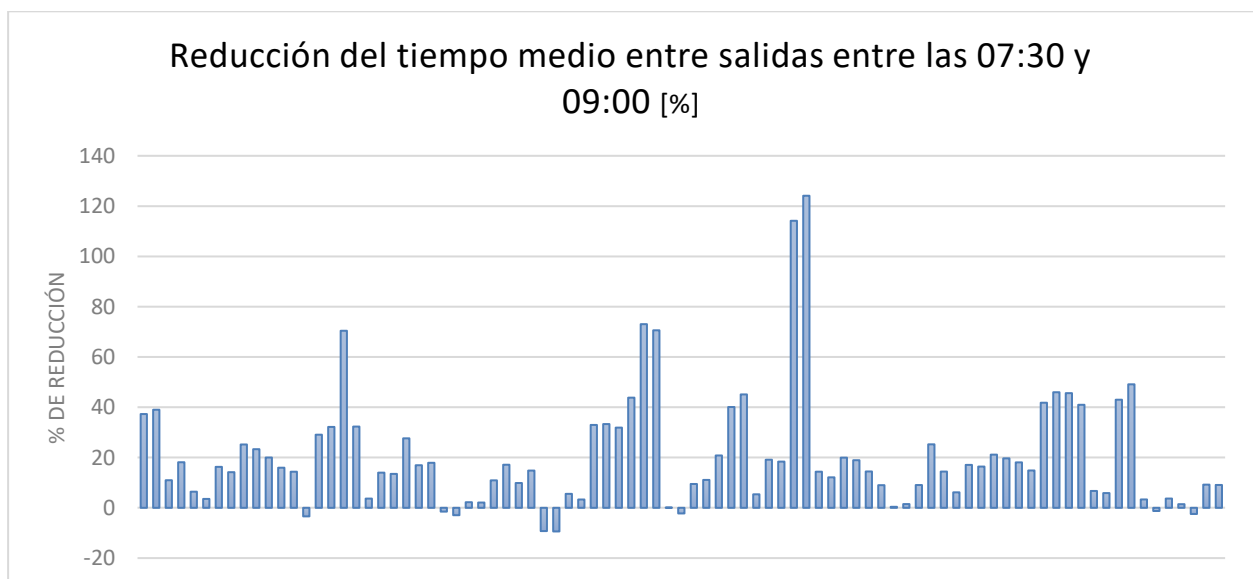


Figura 31. Porcentaje de reducción de los tiempos medios de paso entre las 07:30 y 09:00 en días laborables.

Si estudiamos el mismo comportamiento para la hora punta del final de la jornada laboral, la correspondiente a



las 14:00, observamos que en la mayor parte de las líneas, el comportamiento es aproximadamente similar a lo ya visto para las horas iniciales del día. En este caso, destaca el aumento de la frecuencia de paso del trayecto desde el Prado de San Sebastián hasta la Pablo de Olavide, que tiene un 190% mayor frecuencia a estas horas que en el resto del día, coincidiendo con un horario de comienzo de las clases para muchos estudiantes universitarios. El sentido opuesto, sin embargo, no experimenta un crecimiento tan grande. Por otro lado, observamos que en estas horas del día, el autobús especial al aeropuerto experimenta una gran reducción, de casi el 50%, en los tiempos de espera, cosa que no ocurría en el caso anterior. Si atendemos a los valores más bajos, vemos que por la tarde prácticamente ninguna ruta empeora su servicio, y los que lo hacen, es en un porcentaje de tiempo muy pequeño. En esta franja horaria, la línea con menores tiempos entre pasos es la número 13 en sus dos sentidos, con tiempos de 5.43 minutos, existiendo 9 trayectos con tiempos de paso inferiores a los 6 minutos, y con una reducción general de todos los tiempos del 18%, sensiblemente menor a la observada en la otra hora punta del día.

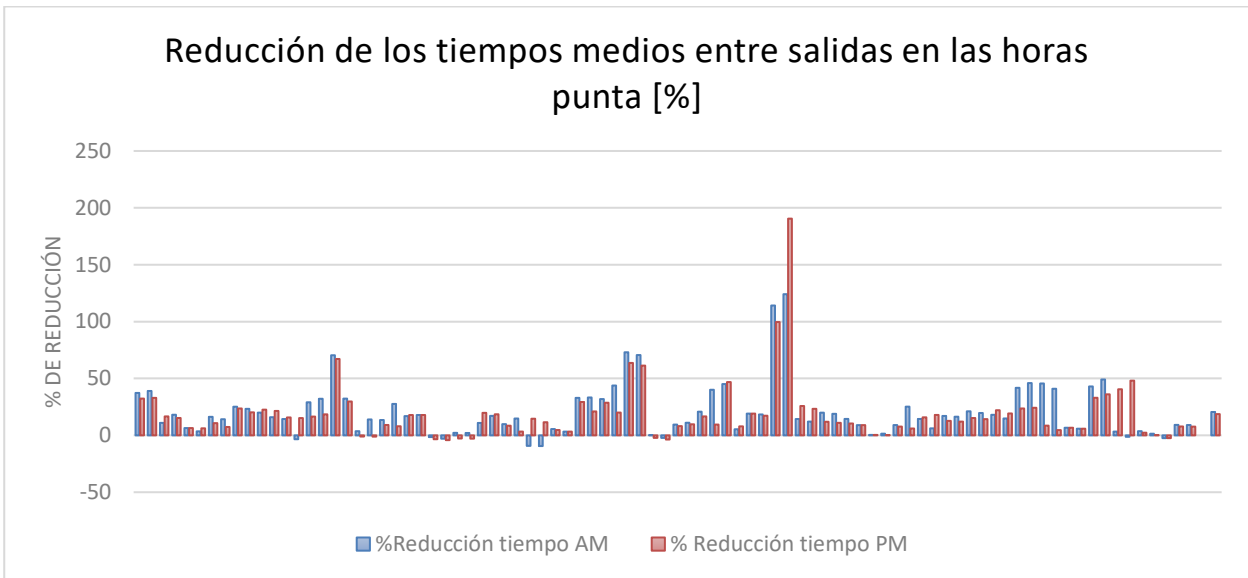


Figura 32. Porcentaje de reducción de los tiempos medios de paso en horas punta.

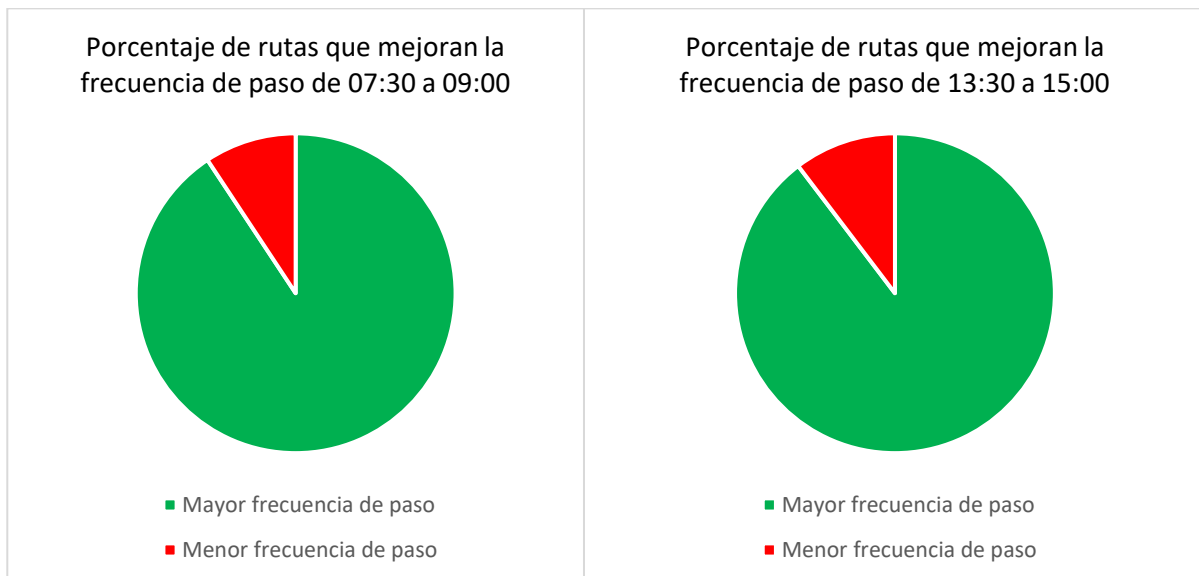


Figura 33. Porcentaje de mejora de las frecuencias de paso en horas punta.

El aumento de la oferta de autobuses en estos segmentos del día, en cualquier caso, no tiene por qué significar una mejora en el servicio si, debido al tráfico, el tiempo que se tarda en alcanzar el destino es mayor. Para evaluar esto, atendemos a la velocidad media de las rutas en estas horas. Eliminadas las rutas que no dan servicio en estas horas, la velocidad media a lo largo del día es 13.4 km/h. En la primera de las horas punta, esta se reduce hasta los 12.73 km/h, y en la hora punta correspondiente al mediodía vuelve a reducirse hasta los 12.32 km/h;

perdiendo más de 1km/h de velocidad con respecto de la media del día. Esto supone un empeoramiento de un 5.2% y un 8.5%, respectivamente.

Si estudiamos la velocidad línea por línea, vemos que tan sólo 16 de las 87 rutas que existen mejoran su velocidad entre las 08:00 y las 09:00; mientras que sólo una lo hace en la segunda hora punta del día, siendo la número 39 en dirección a la Hacienda San Antonio. En el caso del segmento horario de la mañana, existen dos líneas con velocidades sensiblemente superiores al resto del día, que son la 32 y la 52 dirigidas al Polígono Sur y a la Barriada Padre Pío Palmete. Esta misma línea dirigida al Polígono Sur, no obstante, es la que mayor reducción de la velocidad sufre de 14:00 a 15:00, siendo un 23% más lento que el resto del día. En el extremo contrario, las líneas 12 y 13 en sus sentidos desde Pino Montano hasta el centro sufren una reducción de la velocidad de más del 25% en ambos casos por la mañana, siendo las más afectadas a esta hora. A la hora punta correspondiente al mediodía la pérdida de velocidad de las líneas es mucho más dramática, dado que todas las líneas obtienen peores velocidades medias, y tan sólo dos trayectos mantienen la misma velocidad que el resto del día.

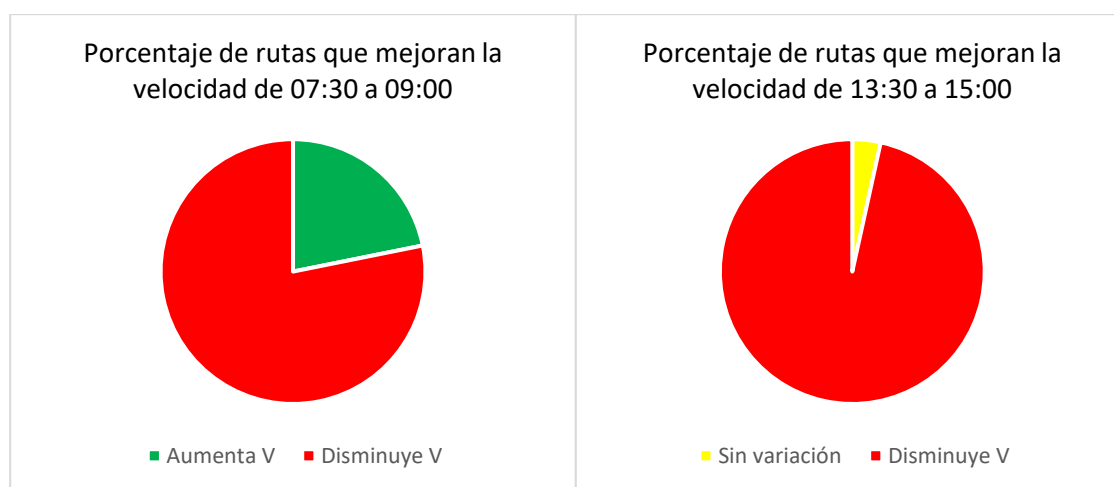


Figura 34. Porcentaje de rutas que mejoran la velocidad en horas punta.

Por último, comparamos si los servicios prestados en las dos franjas horarias mostradas son, en general, mejor o peor que el que se da el resto del día. Para ello, sumaremos el porcentaje de reducción del tiempo con el del incremento de la velocidad; con el objetivo de estudiar si la mejora que se hace al reducir los tiempos de espera compensa el aumento del tiempo que se emplea en realizar un recorrido. Esto lo repetiremos, de nuevo, para ambos segmentos horarios, con el objetivo de estudiar qué rutas prestan peores servicios en estas horas punta que el resto del día. La evaluación final también se encuentra disponible en el portal web creado para la publicación de los resultados. Existen 18 trayectos que empeoran su funcionamiento tanto en la hora punta de la mañana como en la del mediodía. De estas, destacan para mal las líneas B4, 20, 25, 26, 30 y 43, que empeoran sus resultados en ambos sentidos de la línea. Además, existen otros 7 trayectos que dan peor servicio en la hora punta de las 8 de la mañana que el resto del día, pero mejoran en el servicio de las 14:00; y 13 rutas que dan un servicio peor por las tardes pero sí mejoran por las mañanas. Las otras 49 rutas analizadas (en esta etapa no se han analizado todas las rutas, sino sólo aquellas que dan servicio durante las horas punta de los días laborables, excluyendo servicios de sólo fines de semana o nocturnos) en esta parte de la investigación ofrecen mejores servicios para las dos horas punta. Respecto de las mayores mejoras y empeoramientos, ya se ha comentado que la línea con mejores mejoras en ambos tramos horarios es la número 38 dirigida a la Pablo de Olavide; mientras que la que peores servicios ofrece en estos momentos con respecto del resto del día es la número 25, que conecta el Prado con Rochelambert.

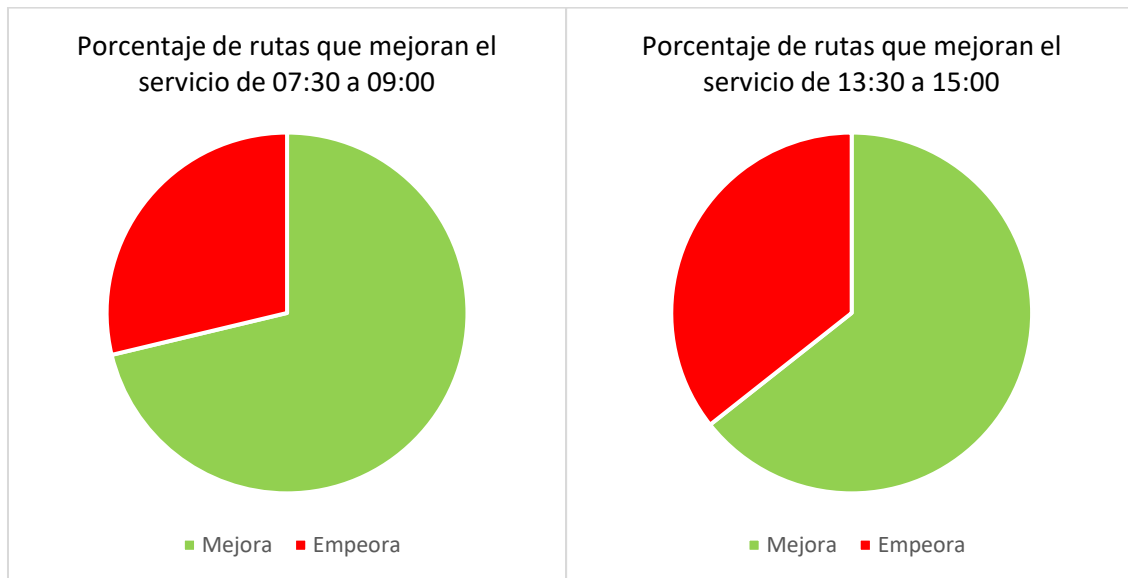


Figura 35. Porcentaje de rutas que mejoran sus servicios en las distintas horas punta.

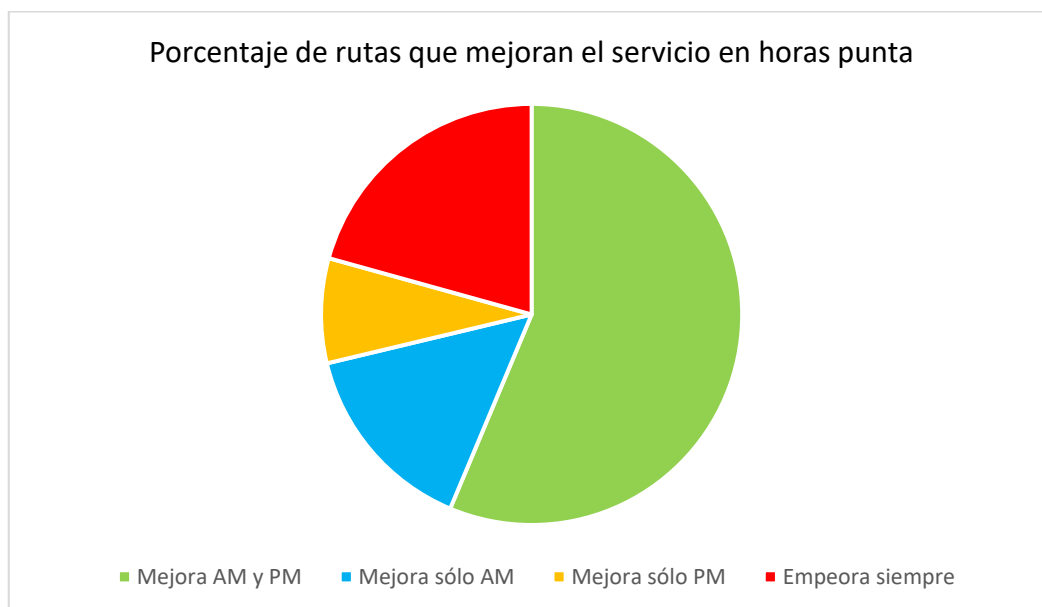


Figura 36. Porcentaje de rutas que mejoran o empeoran sus servicios en horas punta.

### 5.5.2.1 Propuestas de mejora para la red de transportes de Sevilla

En primer lugar, atendiendo a lo estudiado en relación con el carácter radial de la red, una de las propuestas de mejora más evidentes del sistema con el que contamos en la ciudad es la puesta en marcha de líneas de autobuses circulares exteriores al centro de la ciudad; o la creación de líneas que permitiesen la movilidad de forma más directa entre zonas periféricas o dentro de algunas de ellas que son lo suficientemente grandes como para recibir este servicio. Además, el fomento del transporte público en zonas periféricas que, en muchas ocasiones, también son coincidentes con las que niveles socioeconómicos más bajos tienen se hace importante, como forma de impulso de estos barrios, abriéndolos a las oportunidades que proporciona el resto de la ciudad. Después de observar que las líneas son más rápidas cuanto más se alejan del centro, podemos decir que estas posibles nuevas líneas podrían ser las más veloces del sistema de TUSSAM.

Para esta tarea, la realización de un estudio de la movilidad entre estas zonas periféricas se hace necesario e imprescindible; con el objetivo de medir en qué ubicaciones e incluso horarios tendría sentido la implementación de nuevos servicios de transporte. Este estudio podría y debería apoyarse no sólo en encuestas, sino también en técnicas más avanzadas para medir la movilidad, como el análisis de movimientos según las conexiones de teléfonos móviles a según qué celdas de la red o, al igual que se hace en Londres, a partir de estudiar los



movimientos de los usuarios analizando sus pagos con tarjetas de transporte o el rastreo WiFi. Incluso una posible reordenación de todo el sistema de transportes de la ciudad sería apropiada y justificaría un estudio con técnicas tan avanzadas como las explicadas.

A pesar de que queda fuera del estudio realizado, la necesidad de la ciudad de tener un sistema unificado de transportes que incluya los servicios de trenes de cercanías y el propio metro se hace importante, también como forma de acercar a las localidades de alrededor de la ciudad a Sevilla. La multimodalidad existente en muchas otras ciudades debería complementarse con autobuses urbanos que acercasen a los ciudadanos a los servicios de raíles más cercanos como nodos de comunicación, aun siendo gestionados por distintas empresas, tales como la antigua línea LC, ya desaparecida; que conectaba la Isla de la Cartuja con la parada de metro de Blas Infante.

En el plano relacionado con la velocidad de cada línea, los trabajos deberían estar destinados a aumentar la velocidad de las líneas en aquellas zonas de la ciudad más céntricas, con mayor tráfico, semáforos, personas en movimiento y menos cantidad de carriles, con velocidades más bajas. Aunque evidentemente nunca se alcanzarán velocidades como aquellas a las que se llegaría con una red de metro acorde a la ciudad, sí existen opciones para mejorar la velocidad comercial del transporte en autobuses. Muchas de ellas ya se han comentado previamente en este documento, como la prioridad semafórica o el uso de carriles dedicados sólo para transportes públicos. Estas propuestas de aumento de la velocidad también se vuelven aplicables al tranvía.

Respecto al estudio realizado comparando el servicio medio prestado a lo largo del día en comparación con las horas punta, vemos que en general la mejora del servicio es bastante sustancial, con la mayoría de líneas mejorando el servicio prestado por una reducción bastante grande de los tiempos de espera. A pesar de ello, observamos que las velocidades de funcionamiento de los autobuses sufren los atascos tan frecuentes en la ciudad a estas horas, lo que provoca una reducción generalizada de las velocidades medias de las distintas rutas, que se hace más grave de 14:00 a 15:00. Esto hace que lo comentado en el párrafo anterior, orientado a la mejora de las velocidades de las rutas, se hace aún más importante. En este análisis también hemos visto que existen una serie de rutas que ofrecen peor servicio en algunas de las horas punta o incluso en los dos tramos de mayor demanda de transportes. Es necesario realizar un análisis, entendiéndolo por qué estas rutas prestan un servicio peor cuando debería reforzarse el mismo, estudiando la ocupación de estas líneas en estas horas y viendo si es necesario que sean reforzadas.

## 6 CONCLUSIONES Y LÍNEAS DE MEJORA

---

**E**ste trabajo tenía dos objetivos fundamentales. En primer lugar, buscaba analizar el estado de las tecnologías de recopilación y procesamiento de datos en la movilidad; en mi caso centrado en el transporte público. Después, la definición y ejecución de una aplicación relacionada con la ciencia de datos en movilidad.

### 6.1. Conclusiones

Una vez terminado el trabajo, creo que ambos objetivos se han cumplido. Por un lado, analizando el exitoso ecosistema de datos abiertos de transportes de Londres, y comprobando cómo efectivamente ha repercutido de forma positiva en la ciudad; no sólo en el plano de los transportes sino también de la economía, la innovación y el empleo. Para hacer esto hemos estudiado la transformación, impulsada por el Big Data, llevada a cabo por TfL, así como el uso que se ha dado de la información recopilada y el impacto que ha tenido en la operación de la red de transportes; lo cual ha repercutido en la economía, infraestructuras y el empleo de la ciudad. En contraposición a este sistema se han estudiado las tecnologías de información y recogida de datos puestas en marcha en Sevilla, así como las carencias existentes en la ciudad en este sentido. Se han comprobado que además de la nula integración entre modos de transporte, la transparencia relativa a la operación y la obtención de información resultan difíciles.

Por último, se ha realizado un estudio de la operativa del transporte público en Sevilla a través de los datos. Esto se ha hecho a partir del conjunto GTFS de datos de la empresa TUSSAM, y, usando Python, se ha completado, obtenido estadísticas relativas al funcionamiento de la empresa y analizado utilizando no sólo herramientas de visualización de datos en Python, sino también con Excel. A partir del análisis se han realizado diversas propuestas de mejora. Además de eso, se ha creado un portal que pone a disposición de la ciudadanía los resultados extraídos.

A nivel personal, me ha resultado muy interesante realizar este trabajo. Por la parte de investigación, he aprendido mucho en lo relativo a la gestión del transporte público y a la gran cantidad de posibilidades que la tecnología puede ofrecer a este sector; siempre con el objetivo de dar mejor servicio a la gente que vive en la ciudad. Me apena también ver las carencias que se presentan en nuestra ciudad, donde muchas tecnologías no se utilizan, la comunicación que se hace a la ciudadanía es bastante deficiente y que la información que se recopila no se pone a disposición del público a pesar de la gran cantidad de beneficios que esto podría conllevar. Además, he manejado muchos documentos oficiales, algo a lo que no estaba acostumbrado, tanto en inglés como en castellano; lo cual seguro que me resultará útil para el futuro.

Respecto de la parte práctica, he aprendido a manejar flujos de transporte que siguen la especificación GTFS, ampliamente usada en el sector, y toda la terminología propia de este sistema y del transporte público. También, gracias a la aplicación, he aprendido a utilizar herramientas relacionadas con los datos geográficos; tanto la plataforma ArcGIS como el procesado y representación de datos sobre un mapa interactivo en Python; así como adquirir mucha más destreza de la que tenía anteriormente en la librería *Pandas* y en el desarrollo web; aprendiendo a utilizar los despliegues directamente a través de GitHub Pages.

He descubierto en el Big Data un sector de aplicación de mis conocimientos que no creo que se enseñe en la carrera, con grandes oportunidades laborales y en el que me ha gustado iniciarme; en el que no me importaría trabajar en el futuro. Tampoco descarto, después de este trabajo, el dedicarme al sector de transporte y, si fuera posible, a seguir poniendo mis conocimientos al servicio de la ciudad.

## 6.2. Líneas de mejora

- Debido al carácter estático del *feed* GTFS, los datos calculados y los resultados obtenidos representan una foto estática del sistema de transportes por autobuses de la ciudad de Sevilla en un momento concreto, correspondiente con la primavera del año 2023. Una posible ampliación y mejora del proyecto sería adaptar el cálculo de las distintas estadísticas utilizando la API pública de TUSSAM; o la actualización de estas estadísticas cada vez que se publicase una nueva versión de los documentos GTFS.
- Si existiese la posibilidad de conocer la comparación entre los datos teóricos planificados y los datos reales, se podría evaluar el desempeño del servicio y ver en dónde y por qué se produce el malfuncionamiento, si lo hubiera, de la red de transportes. Otros datos como la cantidad de gente que accede a cada vehículo en cada parada o en qué horas también sería interesante, de cara a conocer el dibujo completo del servicio.
- Sería interesante realizar una automatización de todas las tareas realizadas de cara a procesar la información, de forma que dando como entrada un *feed* con ciertas características, se obtuvieran como resultado las métricas más importantes, tablas y mapas. Si esto no fuera posible, al menos que se automatizasen parte de las tareas, como el propio preparado de los datos. En relación con lo anterior, los distintos códigos empleados pueden ser empleados por cualquier persona con un *feed* GTFS en tanto que se apoyan en la norma definida por Google, pero una automatización de parte de las tareas y la encapsulación de las mismas en funciones de una librería haría mucho más simple trabajar con estos datos.
- A pesar de que muchas de las estadísticas y tablas mostradas en el análisis se han hecho con Excel y siendo esta herramienta muy común en el trabajo con datos, se podrían desarrollar funciones y Notebooks que permitiesen calcular y visualizar dentro de Python gráficas de forma dinámica, haciendo uso de librerías como *Plotly*.
- Si las instituciones responsables lo permitiesen, sería interesante integrar en un mismo análisis el resto de los servicios de movilidad que atienden a la ciudad, como son el Metro de Sevilla, los autobuses metropolitanos y los servicios de Cercanías de RENFE, para tener una visión completa de la red de transportes de la ciudad.

# ANEXO A

A continuación, se incluye el código de los ficheros *stopTimesForStats.ipynb*, *sortAndSplitByDay.ipynb*, *statsParadas.ipynb* y *combStops.ipynb*.

```
import gtfs_kit as gk
import numpy as np
import pandas as pd

path = 'TUSSAM_COMP.zip'
feed = gk.read_feed(path, dist_units='km')
newDF = feed.stop_times[['trip_id', 'arrival_time', 'stop_id']]
newDF['route_id'] = pd.Series(dtype='int')
newDF['direction_id'] = pd.Series(dtype='int')
newDF['trip_headsign'] = pd.Series(dtype='string')

#Creamos un nuevo DF que nos sirva como base para computar estadísticas de la
#paradas y líneas
#El nuevo DF tendrá información relativa a la parada, tiempo y ruta asociada.
listTrips = newDF['trip_id'].unique()
for trip in listTrips:

    info = feed.trips.loc[feed.trips['trip_id']==trip].iloc[0]
    dir = info['direction_id']
    route = info['route_id']
    headsign = info['trip_headsign']

    newDF.loc[newDF['trip_id']==trip, 'direction_id'] = dir
    newDF.loc[newDF['trip_id']==trip, 'route_id'] = route
    newDF.loc[newDF['trip_id']==trip, 'trip_headsign'] = headsign

newDF.to_csv(r'data/stopTimesComp.txt', index=False)
```

Código A. 1. Notebook *stopTimesForStats.ipynb*.

```
import numpy as np
import pandas as pd
import gtfs_kit as gk

path = 'TUSSAM_COMP.zip'
feed = (gk.read_feed(path, dist_units='km'))

#Ordenamos los tiempos por paradas, linea, ruta y tiempo, siguiendo esta
jerarquia
df = pd.read_csv('data/stopTimesComp.txt')
df = df.sort_values(by=['stop_id', 'direction_id', 'route_id', 'arrival_time'])
#Guardamos los tiempos ordenados en distintos ficheros según el día de servicio
al que pertenecen.
#Para ello, buscamos qué servicios se prestan cada día y filtramos el feed
festivos = feed.trips.loc[feed.trips['service_id'] == 'INVFES20']['trip_id']
festivos = [int(e) for e in festivos]
dfFest = df[df.trip_id.isin(festivos)]
dfFest.to_csv('data/stopTimesFest.txt', index=False)

laborables = feed.trips.loc[feed.trips['service_id'] == 'INVLAB20']['trip_id']
laborables = [int(e) for e in laborables]
dfLab = df[df.trip_id.isin(laborables)]
dfLab.to_csv('data/stopTimesLab.txt', index=False)

sabados = feed.trips.loc[feed.trips['service_id'] == 'INVSAB20']['trip_id']
sabados = [int(e) for e in sabados]
dfSab = df[df.trip_id.isin(sabados)]
dfSab.to_csv('data/stopTimesSab.txt', index=False)

viernes = feed.trips.loc[feed.trips['service_id'] == 'INVVIE20']['trip_id']
viernes = [int(e) for e in viernes]
dfVie = df[df.trip_id.isin(viernes)]
dfVie.to_csv('data/stopTimesVie.txt', index=False)
```

Código A. 2. Notebook *sortAndSplitByDay.ipynb*.

```

import numpy as np
import pandas as pd
import datetime
import gtfs_kit as gk
import warnings
warnings.filterwarnings("ignore", category=FutureWarning)

#Recibe como primer parametro la ruta del CSV generado anteriormente con los
datos de un día
#y como segundo parámetro la ruta del CSV con los resultados calculados.
#Devuelve
def computeFreqs(inFile,outFile):
    #Import data
    df = pd.read_csv(inFile)
    #Obtenemos la combinación de claves sobre las que queremos calcular las
estadísticas (parada, ruta y dirección)
    claves = df.groupby(["stop_id", "route_id",
"direction_id", 'trip_headsign']).size().reset_index().drop([0],axis=1)
    claves['avgFrequency'] = 0
    for index, row in claves.iterrows():
        stop_id = row['stop_id']
        route_id = row['route_id']
        direction_id = row['direction_id']
        trip_headsign = row['trip_headsign']
        times = df.loc[(df['route_id'] == route_id) & (df['stop_id'] ==
stop_id) & (df['direction_id'] == direction_id), 'arrival_time'].tolist()

        #Reordenamos los tiempos de la lista empezando por las 6am; hora de
comienzo del servicio
        #Para ello, cogemos los tiempos anteriores a las 6 y los ponemos al
final de la lista
        #Este método no afecta si no hay tiempos anteriores a las 6 o si todos
los tiempos
        #fueran anteriores a las 6 (nocturnos)

        timesFirst = [t for t in times if t >= '06:00:00']
        timesLast = [t for t in times if t < '06:00:00']

        times = timesFirst + timesLast

        timedeltas = [(datetime.datetime.strptime(times[i], '%H:%M:%S')-
datetime.datetime.strptime(times[i-1], '%H:%M:%S')).total_seconds() for i in
range(1, len(times))]
        #En el caso de computar tiempos entre 23:xx y 00:xx obtenemos tiempos
negativos que tenemos que corregir
        #sumando los segundos de un día completo
        for i in range(len(timedeltas)):
            if (timedeltas[i]<0):
                timedeltas[i]=timedeltas[i]+24*60*60
    claves.at[index, 'avgFrequency'] = np.mean(timedeltas)/60

```

```

    claves.to_csv(outFile, index=False)
    return claves
freqLab =
computeFreqs('data/stopTimesLab.txt','results/frequencies/avgFreqLab.csv')
freqVie =
computeFreqs('data/stopTimesVie.txt','results/frequencies/avgFreqVie.csv')
freqSab =
computeFreqs('data/stopTimesSab.txt','results/frequencies/avgFreqSab.csv')
freqFest =
computeFreqs('data/stopTimesFest.txt','results/frequencies/avgFreqFest.csv')
#Buscamos los nombres de las paradas para el fichero general de frecuencias
#Para ello, primero importamos el feed
path = 'TUSSAM_COMP.zip'
feed = (gk.read_feed(path, dist_units='km'))
stops = feed.stops
freqComplete = freqLab[['stop_id', 'route_id',
'direction_id', 'trip_headsign', 'avgFrequency']].copy()
freqComplete.insert(0, 'stop_name', '')
freqComplete.rename(columns={'avgFrequency': 'avgFreqLab'}, inplace=True)
freqComplete['avgFreqVie'] = ""
freqComplete['avgFreqSab'] = ""
freqComplete['avgFreqFest'] = ""
#We add the information about the other days to the complete df, as well as
name for the stop
for index, row in freqComplete.iterrows():
    stop_id = str(row['stop_id'])
    route_id = row['route_id']
    direction_id = row['direction_id']

    stop_name = stops.loc[stops['stop_id'] == stop_id]['stop_name'].tolist()
    freqComplete.at[index, 'stop_name'] = stop_name[0]
    stop_id = row['stop_id']

    #We complete with data of the other days
    freqVieStop = freqVie.loc[(freqVie['stop_id'] == stop_id) &
(freqVie['route_id'] == route_id) & (freqVie['direction_id'] ==
direction_id)]['avgFrequency'].tolist()
    #That day might not have service, thus we need to change to 0
    if not freqVieStop:
        freqVieStop.append(0)
    freqComplete.at[index, 'avgFreqVie'] = freqVieStop[0]

    freqSabStop = freqSab.loc[(freqSab['stop_id'] == stop_id) &
(freqSab['route_id'] == route_id) & (freqSab['direction_id'] ==
direction_id)]['avgFrequency'].tolist()
    if not freqSabStop:
        freqSabStop.append(0)
    freqComplete.at[index, 'avgFreqSab'] = freqSabStop[0]

```

```

    freqFestStop = freqFest.loc[(freqFest['stop_id'] == stop_id) &
(freqFest['route_id'] == route_id) & (freqFest['direction_id'] ==
direction_id)][['avgFrequency']].tolist()
    if not freqFestStop:
        freqFestStop.append(0)
    freqComplete.at[index, 'avgFreqFest'] = freqFestStop[0]
#We need to check this to compute well the average time
freqComplete['avgFreqWeekly'] =
(4*freqComplete['avgFreqLab']+freqComplete['avgFreqVie']+freqComplete['avgFreqS
ab']+freqComplete['avgFreqFest'])/7
freqComplete.to_csv('results/frequencies/avgFreqComplete.csv', index=False)
#Now we create a file with the most relevant values:
#We will have the avg time en each stop per day and the time increase on
#saturdays and sundays to working days

resultDf = pd.DataFrame(freqComplete.loc[:, ['stop_name',
'stop_id']].drop_duplicates(), columns=['stop_name', 'stop_id'])

keysDf = pd.DataFrame(freqComplete.loc[:, ['stop_id', 'route_id',
'direction_id']].drop_duplicates(), columns=['stop_id', 'route_id',
'direction_id'])

resultDf['avgFreqLab'] = 0
resultDf['avgFreqVie'] = 0
resultDf['avgFreqSab'] = 0
resultDf['avgFreqFest'] = 0
resultDf['increaseSab'] = 0
resultDf['increaseFest'] = 0

dfLab = pd.read_csv('data/stopTimesLab.txt')
dfVie = pd.read_csv('data/stopTimesVie.txt')
dfSab = pd.read_csv('data/stopTimesSab.txt')
dfFest = pd.read_csv('data/stopTimesFest.txt')

#We iterate through all the stops
for index,row in resultDf.iterrows():
    stop_id = row['stop_id']
    routesOfStop = freqComplete.loc[freqComplete['stop_id'] == stop_id]
    auxLab = pd.DataFrame({'nTrips': pd.Series(dtype='float64'), 'time':
pd.Series(dtype='float64')})
    auxVie = pd.DataFrame({'nTrips': pd.Series(dtype='float64'), 'time':
pd.Series(dtype='float64')})
    auxSab = pd.DataFrame({'nTrips': pd.Series(dtype='float64'), 'time':
pd.Series(dtype='float64')})
    auxFest = pd.DataFrame({'nTrips': pd.Series(dtype='float64'), 'time':
pd.Series(dtype='float64')})

    #We will compute the total avg value as a weigthed sum
    for iRoute,rowRoute in routesOfStop.iterrows():
        route_id = rowRoute['route_id']

```



```

direction_id = rowRoute['direction_id']
#We need to get the number of trips each day for each route
nLab = len(dfLab.loc[(dfLab["stop_id"]==stop_id) &
(dfLab["route_id"]==route_id) & (dfLab["direction_id"]==direction_id)])
nVie = len(dfVie.loc[(dfVie["stop_id"]==stop_id) &
(dfVie["route_id"]==route_id) & (dfVie["direction_id"]==direction_id)])
nSab = len(dfSab.loc[(dfSab["stop_id"]==stop_id) &
(dfSab["route_id"]==route_id) & (dfSab["direction_id"]==direction_id)])
nFest = len(dfFest.loc[(dfFest["stop_id"]==stop_id) &
(dfFest["route_id"]==route_id) & (dfFest["direction_id"]==direction_id)])
#Store this values
auxLab = auxLab.append({'nTrips': nLab, 'time':
rowRoute['avgFreqLab']}, ignore_index=True)
auxVie = auxVie.append({'nTrips': nVie, 'time':
rowRoute['avgFreqVie']}, ignore_index=True)
auxSab = auxSab.append({'nTrips': nSab, 'time':
rowRoute['avgFreqSab']}, ignore_index=True)
auxFest = auxFest.append({'nTrips': nFest, 'time':
rowRoute['avgFreqFest']}, ignore_index=True)

#Compute the avg for each day
resultDf.loc[resultDf['stop_id'] == stop_id, 'avgFreqLab'] =
sum(auxLab['nTrips'] * auxLab['time'])/auxLab['nTrips'].sum()
resultDf.loc[resultDf['stop_id'] == stop_id, 'avgFreqVie'] =
sum(auxVie['nTrips'] * auxVie['time'])/auxVie['nTrips'].sum()
resultDf.loc[resultDf['stop_id'] == stop_id, 'avgFreqSab'] =
sum(auxSab['nTrips'] * auxSab['time'])/auxSab['nTrips'].sum()
resultDf.loc[resultDf['stop_id'] == stop_id, 'avgFreqFest'] =
sum(auxFest['nTrips'] * auxFest['time'])/auxFest['nTrips'].sum()

#And increase percentage
resultDf['increaseSab'] = (resultDf['avgFreqSab']/resultDf['avgFreqLab']-1)*100
resultDf['increaseFest'] = (resultDf['avgFreqFest']/resultDf['avgFreqLab']-
1)*100
resultDf.to_csv('results/frequencies/frequencies.csv', index=False)

```

Código A. 3. Notebook *statsParadas.ipynb*.

```

import pandas as pd
import gtfs_kit as gk
path = 'TUSSAM_COMP.zip'

stops = (gk.read_feed(path, dist_units='km')).stops
stops = stops[['stop_id', 'stop_lat', 'stop_lon']].copy()
stops['stop_id'] = stops['stop_id'].astype(float)

stopsExc = pd.read_csv('results/frequencies/frequencies.csv')
r = pd.merge(stopsExc, stops, on="stop_id")
r.to_csv('results/frequencies/freqWithCoord.csv')

```

Código A. 4. Notebook *combStops.ipynb*.

# ANEXO B

A continuación, se incluye el código de los ficheros *statsViajes.ipynb* y *statsLineas.ipynb*.

```
#Import the gtfs-kit module
import gtfs_kit as gk
import pandas as pd
#Declare the directory path for the GTFS zip file
path = 'TUSSAM_COMP.zip'
#Read the feed with gtfs-kit
feed = (gk.read_feed(path, dist_units='km'))

tripStats = gk.trips.compute_trip_stats(feed)
tripsFiltered = feed.trips[['trip_id', 'trip_headsign']].copy()
tripStats = pd.merge(tripStats, tripsFiltered, on=['trip_id'])

#Parece que algunas líneas tienen errores, ya que muestran velocidades de más
de 130km/h.
#Eliminaremos las filas con posibles errores, estableciendo que todas las
líneas deben durar al menos 6 minutos.
#También eliminamos las filas con más de 50km/h de velocidad media.

tripStats.drop(tripStats[tripStats['duration'] < 0.1].index, inplace = True)
tripStats.drop(tripStats[tripStats['speed'] > 50].index, inplace = True)
print(tripStats[['speed']].mean())
tripStats.to_csv('data/tripComp.txt', index=False)

#Obtenemos una velocidad comercial teórica de aproximadamente 13.8 km/h;
ligeramente superior a la velocidad
#comercial alcanzada en 2022.

#Lo siguiente que haremos será crear distintos archivos con los trips de cada
tipo de día.

festivos = feed.trips.loc[feed.trips['service_id'] == 'INVFES20']['trip_id']
dfFest = tripStats[tripStats.trip_id.isin(festivos)]
dfFest.to_csv('data/tripsFest.txt', index=False)

laborables = feed.trips.loc[feed.trips['service_id'] == 'INVLAB20']['trip_id']
dfLab = tripStats[tripStats.trip_id.isin(laborables)]
dfLab.to_csv('data/tripsLab.txt', index=False)

sabados = feed.trips.loc[feed.trips['service_id'] == 'INVSAB20']['trip_id']
dfSab = tripStats[tripStats.trip_id.isin(sabados)]
dfSab.to_csv('data/tripsSab.txt', index=False)

viernes = feed.trips.loc[feed.trips['service_id'] == 'INVVIE20']['trip_id']
dfVie = tripStats[tripStats.trip_id.isin(viernes)]
dfVie.to_csv('data/tripsVie.txt', index=False)
```

Código B. 1. Notebook *statsViajes.ipynb*.

```

import pandas as pd
import gtfs_kit as gk
import datetime
import numpy as np

def computeRouteStats(inFile,outFile):
    df = pd.read_csv(inFile)
    claves = df.groupby(["route_id",
"direction_id", 'trip_headsign', 'shape_id', 'route_short_name']).size().reset_index().drop([0],axis=1)
    claves['startOfService'] = 0
    claves['endOfService'] = 0
    claves['avgDepartureTime'] = 0
    claves['avgDuration'] = 0
    claves['nStops'] = 0
    claves['avgSpeed'] = 0

    for index,row in claves.iterrows():

        route_id = row['route_id']
        direction_id = row['direction_id']

        #startOfService es el primer startTime
        #endOfService es el último startTime
        times = df.loc[(df['route_id'] == route_id) & (df['direction_id'] ==
direction_id), 'start_time'].tolist()
        claves.at[index,'startOfService'] = min(times)
        claves.at[index,'endOfService'] = max(times)

        #AvgDepartureTime es la media de tiempos transcurridos entre start_time
        #Calculamos diferencias de tiempo y aplicamos una posible corrección
        como en el caso de las paradas. Luego calculamos la media y convertimos a
        minutos
        timedeltas = [(datetime.datetime.strptime(times[i], '%H:%M:%S')-
datetime.datetime.strptime(times[i-1], '%H:%M:%S')).total_seconds() for i in
range(1, len(times))]
        for i in range(len(timedeltas)):
            if (timedeltas[i]<0):
                timedeltas[i]=timedeltas[i]+24*60*60
        claves.at[index,'avgDepartureTime'] = np.mean(timedeltas)/60

        #avgDuration es la media de la duración del recorrido (x60 para
        minutos)
        durations = df.loc[(df['route_id'] == route_id) & (df['direction_id']
== direction_id), 'duration'].tolist()
        claves.at[index,'avgDuration'] = np.mean(durations)*60

        #nStops numero medio de paradas de la ruta (hay rutas con recorridos
        variables en función de la hora)

```

```

        numStops = df.loc[(df['route_id'] == route_id) & (df['direction_id'] ==
direction_id), 'num_stops'].tolist()
        claves.at[index, 'nStops'] = np.mean(numStops)
        #avgSpeed es la media de duracion del recorrido (en km/h)
        speeds = df.loc[(df['route_id'] == route_id) & (df['direction_id'] ==
direction_id), 'speed'].tolist()
        claves.at[index, 'avgSpeed'] = np.mean(speeds)

        claves.to_csv(outFile, index=False)
    return claves

computeRouteStats('data/tripsFest.txt', 'results/routes/routesFest.csv')
computeRouteStats('data/tripsLab.txt', 'results/routes/routesLab.csv')
computeRouteStats('data/tripsVie.txt', 'results/routes/routesVie.csv')
computeRouteStats('data/tripsSab.txt', 'results/routes/routesSab.csv')

```

Código B. 2. Notebook *statsLineas.ipynb*.

```

import pandas as pd
df = pd.read_csv('results/routes/routesLab.csv')
dfAM = pd.read_csv('results/routes/routesLabPeakAM.csv')
dfPM = pd.read_csv('results/routes/routesLabPeakPM.csv')

dfCompare =
df[['route_id', 'direction_id', 'trip_headsign', 'shape_id', 'route_short_name', 'av
gDepartureTime', 'avgSpeed']].copy()
dfCompare = dfCompare.rename(columns={'avgDepartureTime':
'avgDepartureTimeDay', 'avgSpeed': 'avgSpeedDay'})

dfCompare =
pd.merge(dfCompare, dfAM, on=['route_id', 'direction_id', 'trip_headsign', 'shape_id
', 'route_short_name'])
dfCompare = dfCompare.drop(columns=['startOfService',
'endOfService', 'avgDuration', 'nStops'])
dfCompare = dfCompare.rename(columns={'avgDepartureTime': 'avgDepartureTimeAM',
'avgSpeed': 'avgSpeedAM'})

dfCompare =
pd.merge(dfCompare, dfPM, on=['route_id', 'direction_id', 'trip_headsign', 'shape_id
', 'route_short_name'])
dfCompare = dfCompare.drop(columns=['startOfService',
'endOfService', 'avgDuration', 'nStops'])
dfCompare = dfCompare.rename(columns={'avgDepartureTime': 'avgDepartureTimePM',
'avgSpeed': 'avgSpeedPM'})

dfCompare.to_csv('results/routes/comparingRoutesPeakHours.csv')

```

Código B. 3. Notebook *compareLabsPeak.ipynb*.

## REFERENCIAS

- [1 L. G. Alcalde, “Business Insider,” 16 Junio 2022. [Online]. Available: <https://www.businessinsider.es/10-empresas-valiosas-mundo-microsoft-lider-956705>. [Accessed 24 Julio 2022].
- [2 Iberdrola, [Online]. Available: <https://www.iberdrola.com/innovacion/cuarta-revolucion-industrial>. [Accessed 24 Julio 2022].
- [3 APTA, “Leveraging Big Data in the Transportation Industry,” [Online]. Available: <https://www.apta.com/wp-content/uploads/Big-Data-Policy-Brief.pdf>. [Accessed 03 Mayo 2023].
- [4 J. Surico, “MIT Technology Review,” 29 Abril 2021. [Online]. Available: <https://www.technologyreview.es/s/13320/el-big-data-esta-iniciando-la-revolucion-del-transporte-publico-que-tanto-necesitamos>.
- [5 Seoul Metropolitan Government, “Seoul Urban Solutions Agency,” [Online]. Available: <https://susa.or.kr/en/Big-Data-in-Transportation-Policy?ckattempt=1>. [Accessed 01 Mayo 2023].
- [6 W. L. e. al., “MaaS development and APPs,” in *Big Data and Mobility as a Service*, 2022.
- [7 R. Shah, *Innovation through Open Data at Transport for London - London Innovation Series 2017*, Londres, 2017.
- [8 Ministerio de Transportes, Movilidad y Agenda Urbana, “TLE Observatorio,” [Online]. Available: <https://observatoriotransporte.mitma.gob.es/estudio-experimental>. [Accessed 03 Mayo 2023].
- [9 Ministerio de Transportes, Movilidad y Agenda Urbana, [Online]. Available: <https://www.mitma.gob.es/ministerio/proyectos-singulares/estudio-de-movilidad-con-big-data>. [Accessed 01 Mayo 2023].
- [10 Ministerio de Asuntos Económicos y Transformación Digital, “datos.gob.es,” [Online]. Available: <https://datos.gob.es/en/noticia/main-mobility-datasets-datosgobes>. [Accessed 03 Mayo 2023].
- [11 Ayuntamiento de Sevilla, “Portal de Datos Abiertos,” [Online]. Available: <http://datosabiertos.sevilla.org/data/?groups=transporte>. [Accessed 24 Julio 2022].
- [12 R. García, Interviewee, *Director Gerente, TUSSAM*. [Interview]. Noviembre 2022.
- [13 EGOS BI, “EGOS BI,” [Online]. Available: <https://www.egosbi.com/historia-del-big-data/#:~:text=En%202005%2C%20Roger%20Mougalas%20de,herramientas%20tradicionales%20de%20inteligencia%20empresarial..> [Accessed 25 Julio 2022].
- [14 PowerData, “PowerData,” [Online]. Available: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/381767/7-definiciones-de-big-data>. [Accessed 25 Julio 2022].

- [1 Indra, “Indra,” [Online]. Available: <https://www.indracompany.com/es/blogneo/cuantas-v-deberia-big-5> data. [Accessed 25 Julio 2022].
- [1 Telefónica, “Blog Think Big,” [Online]. Available: <https://empresas.blogthinkbig.com/5-v-big-6> data/#:~:text=Las%20V%20del%20Big,velocidad%2C%20veracidad%2C%20variedad%20y%20valor. [Accessed 25 Julio 2022].
- [1 Hitachi Vantara, [Online]. Available: <https://www.hitachivantara.com/es-latam/insights/faq/what-are-7> different-data-types.html#:~:text=Los%20datos%20semiestructurados%20se%20refieren,de%20datos%20de%20la%20aplicaci%C3%B3n. [Accessed 26 Julio 2022].
- [1 B. Gitensein, “Why Unstructured Data Is Your Organization’s Best-Kept Secret,” *Geek Wire*, 2021. 8]
- [1 PowerData, [Online]. Available: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/que-es-big-data-9> hadoop-y-para-que-sirve. [Accessed 27 Julio 2022].
- [2 Cybiant, [Online]. Available: <https://www.cybiant.com/resources/big-data-storage-methods/>. [Accessed 27 0] Julio 2022].
- [2 G. Sharma, “IEEE Computer Society,” [Online]. Available: <https://www.computer.org/publications/tech-1> news/trends/big-data-and-cloud-computing. [Accessed 27 Julio 2022].
- [2 A. Banafa, “BBVA Open Mind,” [Online]. Available: <https://www.bbvaopenmind.com/tecnologia/mundo-2> digital/edge-computing-paradigma/. [Accessed 27 Julio 2022].
- [2 Fauna, [Online]. Available: <https://fauna.com/blog/edge-computing-vs-cloud-computing-whats-the-3> difference. [Accessed 29 Julio 2022].
- [2 Amazon Web Services, [Online]. Available: <https://aws.amazon.com/es/elasticmapreduce/details/hadoop/>. 4] [Accessed 27 Julio 2022].
- [2 Kurios, [Online]. Available: <https://www.kurios.la/blog/los-4-tipos-de-data-analytics>. [Accessed 27 Julio 5] 2022].
- [2 Universidad de Alcalá, [Online]. Available: <https://www.master-data-scientist.com/caracteristicas-data-6> analytics/. [Accessed 27 Julio 2022].
- [2 Gamco, [Online]. Available: <https://gamco.es/tipos-de-analisis-realizan-con-big-data/>. [Accessed 27 Julio 7] 2022].
- [2 GISGeography, [Online]. Available: <https://gisgeography.com/what-is-arcgis/>. [Accessed 18 Marzo 2023]. 8]
- [2 Ayuware, “Ayuware Blog,” [Online]. Available: <https://www.ayuware.es/blog/como-estandarizar-9> datos/#:~:text=La%20estandarizaci%C3%B3n%20de%20datos%2C%20en,o%20personas%20que%20di spongan%20de. [Accessed 06 Mayo 2023].
- [3 B. Goldstein, *Beyond Transparency: Open Data and the Future of Civic Innovation*, 2013. 0]

- [3] MobilityData, “Mobility Database,” [Online]. Available: <https://database.mobilitydata.org/>. [Accessed 06 Mayo 2023].
- [3] Google Developers, “Google Transit,” [Online]. Available: <https://developers.google.com/transit/gtfs?hl=es-419>. [Accessed 18 Marzo 2023].
- [3] Epitech, “Epitech Spain - Blog Tecnológico,” [Online]. Available: <https://www.epitech-it.es/lenguajes-3-programacion-big-data/>. [Accessed 3 Mayo 2023].
- [3] J. Zapata. [Online]. Available: <https://joserzapata.github.io/courses/python-ciencia-datos/visualizacion/>. [Accessed 03 Mayo 2023].
- [3] J. L. G. Grandes, “MappingGIS,” [Online]. Available: <https://mappinggis.com/2022/09/leaflet-utilizando-5-leaflet-con-python/>. [Accessed 03 Mayo 2023].
- [3] Naciones Unidas, “Revision of World Urbanization Prospects,” Nueva York, 2018. 6]
- [3] Asamblea General de las Naciones Unidas, “Memoria del Secretario General sobre la Labor de la Organización,” Nueva York, 2015. 7]
- [3] M. Fernández, Descifrar las smart cities: ¿Qué queremos decir cuando hablamos de smart cities?, Caligrama 8] Editorial, 2016.
- [3] D. Sikora-Fernandez, “Factores de desarrollo de las ciudades inteligentes,” 2017. 9]
- [4] IESE, “IESE Cities in Motion Index 2020,” 2020. 0]
- [4] Zigurat Global Institute of Technology, [Online]. Available: <https://www.e-zigurat.com/blog/en/urban-1-development-policy-smart-cities/>. [Accessed 28 Julio 2022].
- [4] G. D. Maayan, “Dataversity,” [Online]. Available: <https://www.dataversity.net/how-big-data-impacts-2-smart-cities/#:~:text=Big%20data%20can%20have%20an,%2C%20city%20budgets%2C%20and%20more.&text=Smart%20cities%20must%20provide%20security,predict%20the%20exact%20crime%20location..> [Accessed 29 Julio 2022].
- [4] eSmartCity, “eSmartCity.es,” [Online]. Available: <https://www.esmartcity.es/2018/04/27/aspern-ciudad-3-inteligente-sin-coches-laboratorio-urbano-estilo-vida-sostenible-14-kilometros-viena>. [Accessed 28 Julio 2022].
- [4] Aspern Smart City Research, [Online]. Available: <https://www.ascr.at/en/>. [Accessed 28 Julio 2022]. 4]
- [4] Ayuntamiento de Sevilla, “Horizonte Sevilla Inteligente - Dossier Informativo,” 2021. 5]
- [4] DigitalES, “Sevilla se estrena como ‘Smart City’ esta Semana Santa,” 2018. 6]

[4 IMD, “Smart City Index,” 2021.  
7]

[4 Oracle, [Online]. Available: [https://www.oracle.com/es/applications/what-is-saas/#:~:text=El%20software%20como%20servicio%20\(SaaS,con%20un%20sistema%20de%20pago.](https://www.oracle.com/es/applications/what-is-saas/#:~:text=El%20software%20como%20servicio%20(SaaS,con%20un%20sistema%20de%20pago.)  
8] [Accessed 05 Septiembre 2022].

[4 Iberdrola, [Online]. Available: <https://www.iberdrola.com/innovacion/maas-movilidad-sostenible.>  
9] [Accessed 5 Septiembre 2022].

[5 J. R. R. G. e. al, “State of the Art of Mobility as a Service (MaaS) Ecosystems and Architectures—An  
0] Overview of, and a Definition, Ecosystem and System Architecture for electric Mobility as a Service (eMaaS),” 2019.

[5 MaaS Alliance, [Online]. Available: <https://maas-alliance.eu/homepage/what-is-maas/>. [Accessed 5  
1] Septiembre 2022].

[5 Deloitte Review, “The rise of Mobility as a Service,” 2017.  
2]

[5 “Whim,” [Online]. Available: <https://whimapp.com/about-whim/>. [Accessed 5 Septiembre 2022].  
3]

[5 Velco, [Online]. Available: <https://velco.tech/en/3-european-cities-maas-service-mobility/>. [Accessed 5  
4] Septiembre 2022].

[5 Qixxit, “LinkedIn,” [Online]. Available: 5.  
5]

[5 Enrique Diego Bernardo, “eSmartCity,” [Online]. Available:  
6] <https://www.esmartcity.es/comunicaciones/comunicacion-movilidad-como-servicio-madrid-maas-madrid.>  
[Accessed 5 Septiembre 2022].

[5 F. Peinado, “El País,” [Online]. Available:  
7] [https://elpais.com/ccaa/2018/12/19/madrid/1545246072\\_989686.html](https://elpais.com/ccaa/2018/12/19/madrid/1545246072_989686.html). [Accessed 5 Septiembre 2022].

[5 INDRA, [Online]. Available: [https://www.indracompany.com/en/noticia/indra-develops-first-mobility-  
8\] service-maas-platform-six-spanish-cities-six-autonomous.](https://www.indracompany.com/en/noticia/indra-develops-first-mobility-service-maas-platform-six-spanish-cities-six-autonomous.) [Accessed 03 Mayo 2023].

[5 Meep, [Online]. Available: <https://meep.app/es/ciudades/>. [Accessed 5 Septiembre 2022].  
9]

[6 Transport For London, [Online]. Available: [https://tfl.gov.uk/corporate/about-tfl/what-we-  
0\] do?intcmp=2582.](https://tfl.gov.uk/corporate/about-tfl/what-we-do?intcmp=2582.) [Accessed 8 Agosto 2022].

[6 London Transport Museum, [Online]. Available:  
1] <https://www.ltmuseum.co.uk/collections/stories/design/evolution-roundel.> [Accessed 8 Agosto 2022].

[6 B. Hogge, “Transport For London, Get Set, Go!,” 2016.  
2]



- [6 Deloitte, “Assessing the value of TfL’s open data and digital partnerships,” 2017.  
3]
- [6 H. Low, “BBC News,” [Online]. Available: <https://www.bbc.com/news/uk-england-london-58197631>.  
4] [Accessed 10 Agosto 2022].
- [6 Q. Fu, “Understanding the travel patterns on London Rail network: The use of Oyster Card Data,” Leeds,  
5] 2012.
- [6 Transport For London, [Online]. Available: [https://tfl.gov.uk/corporate/privacy-and-cookies/contactless-](https://tfl.gov.uk/corporate/privacy-and-cookies/contactless-payment)  
6] payment. [Accessed 10 Agosto 2022].
- [6 Transport For London, “Review of the TfL WiFi pilot,” Londres, 2017.  
7]
- [6 Transport For London, [Online]. Available: [https://tfl.gov.uk/corporate/privacy-and-cookies/wi-fi-data-](https://tfl.gov.uk/corporate/privacy-and-cookies/wi-fi-data-collection)  
8] collection. [Accessed 10 Agosto 2022].
- [6 Information Commissioner's Office, “Wi-Fi location analytics,” 2016.  
9]
- [7 J. O'Malley, “WIRED,” [Online]. Available: [https://www.wired.co.uk/article/london-underground-wifi-](https://www.wired.co.uk/article/london-underground-wifi-tracking#:~:text=In%20the%20autumn%20of%202016,people%20were%20using%20the%20tube)  
0] tracking#:~:text=In%20the%20autumn%20of%202016,people%20were%20using%20the%20tube.  
[Accessed 10 Agosto 2022].
- [7 BBC News, [Online]. Available: [http://news.bbc.co.uk/2/hi/uk\\_news/england/london/4616530.stm](http://news.bbc.co.uk/2/hi/uk_news/england/london/4616530.stm).  
1] [Accessed 11 Agosto 2022].
- [7 BBC News, [Online]. Available: <http://news.bbc.co.uk/1/hi/england/london/4991326.stm>. [Accessed 11  
2] Agosto 2022].
- [7 S. Reeds, “Intelligent Transport,” [Online]. Available: [https://www.intelligenttransport.com/transport-](https://www.intelligenttransport.com/transport-articles/688/the-ibus-system-serving-as-the-backbone-for-londons-bus-network/)  
3] articles/688/the-ibus-system-serving-as-the-backbone-for-londons-bus-network/. [Accessed 11 Agosto  
2022].
- [7 Transportation Research Group, University of Southampton, “THE WAY AHEAD FOR LONDON’S  
4] BUS PRIORITY AT TRAFFIC SIGNALS,” Southampton, 2009.
- [7 J. B. Gordon, “Intermodal Passenger Flows on London’s Public Transport Network,” Massachusetts, 2012.  
5]
- [7 W. Wang, “Bus passenger origin-destination estimation and travel behavior using automated data collection  
6] systems in London, UK,” Massachusetts, 2010.
- [7 W. Wang, “Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data  
7] Collection Systems,” Massachusetts, 2011.
- [7 R. Dumas, “Analyzing Transit Equity Using Automatically Collected Data,” Massachusetts, 2015.  
8]
- [7 J. O. Mohammed Mohammed, “Origin-destination inference in public transportation systems: A  
9] comprehensive review,” International Journal of Transportation Science and Technology, 2022.

- [8 J. B. G. e. al., “Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and 0] Vehicle Location Data,” 2013.
- [8 N. N. e. al, “Activity detection and transfer identification for public transit fare card data,” Springer 1] Science+Business Media, Nueva York, 2015.
- [8 A. George, “Wired,” [Online]. Available: <https://www.wired.com/2014/06/london-double-decker-seats/>. 2] [Accessed 15 Agosto 2022].
- [8 Transport For London, [Online]. Available: <https://tfl.gov.uk/info-for/media/press-releases/2018/june/tfl-to-trial-automatic-passenger-counting-on-buses#:~:text=TfL%20is%20beginning%20a%20three,time%20with%20the%20most%20accuracy..> 3] [Accessed 15 Agosto 2022].
- [8 U. M. e. al., “Occupancy Estimation using WiFi: A Case Study for Counting Passengers on Busses,” 2019. 4]
- [8 A. H. e. al., “Estimating bus passenger volume based on a Wi-Fi scanner survey,” Tokyo, 2020. 5]
- [8 D. e. al, “Weigh-in-motion sensors and traffic monitoring systems - State of the art and development trends,” 6] 2020.
- [8 GisGeography, “GisGeography,” [Online]. Available: <https://gisgeography.com/passive-active-sensors-remote-sensing/>. [Accessed 15 Agosto 2022]. 7]
- [8 I. P. e. al, “Intelligent Transport,” [Online]. Available: <https://www.intelligenttransport.com/transport-articles/3116/automatic-passenger-counting-systems-for-public-transport/>. [Accessed 15 Agosto 2022]. 8]
- [8 Revista Española de Electronica, [Online]. Available: <https://www.redeweb.com/articulos/adaptando-la-tecnologia-de-sensor-infrarrojo-pasivo-pir-a-nuevas-aplicaciones/>. [Accessed 15 Agosto 2022]. 9]
- [9 US Department of Transportation, Automatic Vehicle Monitoring, 1982. 0]
- [9 G. Wentworth, “Transport for London creates an open data exosystem with Amazon Web Services,” 2015. 1]
- [9 P. Young, *Taking Transport for London into the Cloud - AWS Government Education and Nonprofits Symposium*, Londres, 2014. 2]
- [9 D. Mewett, *Transport for London - AWS Summit London*, Londres, 2015. 3]
- [9 Preservica, [Online]. Available: <https://preservica.com/resources/press-releases/transport-for-london-tfl-extends-its-use-of-the-cloud-with-new-digital-preservation-platform.> 4] [Accessed 19 Agosto 2022].
- [9 Microsoft, “Microsoft News,” [Online]. Available: <https://news.microsoft.com/en-gb/2015/03/17/six-months-of-contactless-journeys-on-londons-transport-network-powered-by-microsoft-technology/>. 5] [Accessed 19 Agosto 2022].
- [9 S. Gutierrez, “TfL Digital Blog,” [Online]. Available: <https://blog.tfl.gov.uk/2016/08/17/twitter-pilot->

- 6] weekend-travel-advice-direct-messages/. [Accessed 22 Agosto 2022].
- [9 T. Macaulay, “Computer World,” [Online]. Available:  
7] <https://www.computerworld.com/article/3427521/how-tfl-is-using-predictive-analytics-to-keep-the-underground-moving.html#:~:text=Predictive%20maintenance%20at%20TfL,on%20the%20track%20takes%20place..> [Accessed 22 Agosto 2022].
- [9 Transport For London, “London Underground Asset Management Strategy Summary,” Londres, 2016.  
8]
- [9 K. Scott, “Tech Insight,” [Online]. Available: <https://techinsight.net/2022/04/predictive-maintenance-london-underground/>. [Accessed 22 Agosto 2022].
- [1 Greater London Authority, “Smarter London Together,” Londres, 2018.  
00  
]
- [1 Ayuntamiento de Sevilla, “Plan de Movilidad Urbana Sostenible del municipio de Sevilla - Diagnóstico,”  
01 2019.  
]
- [1 Globalvia, [Online]. Available: <https://www.globalvia.com/es/inicio/movilidad/concesion/metro-de-sevilla/>. [Accessed 10 Septiembre 2022].  
02  
]
- [1 OCU, [Online]. Available: <https://www.ocu.org/coches/bicicletas-electricas/noticias/mejores-ciudades-para-bicis>. [Accessed 10 Septiembre 2022].  
03  
]
- [1 Ayuntamiento de Sevilla, “Plan de Movilidad Urbana Sostenible 2030 del municipio de Sevilla,” 2019.  
04  
]
- [1 OCU, [Online]. Available: <https://www.ocu.org/coches/coches/noticias/movilidad-urbana>. [Accessed 10  
05 Septiembre 2022].  
]
- [1 FACUA, [Online]. Available: <https://www.facua.org/es/noticia.php?Id=11158>. [Accessed 10 Septiembre  
06 2022].  
]
- [1 Ministerio de Fomento, “El transporte urbano y metropolitano en España,” 2019.  
07  
]
- [1 A. G. Reyes, “El metro de Sevilla es el peor de España,” *ABC de Sevilla*, 12 Octubre 2014.  
08  
]
- [1 TUSSAM, “Conoce la historia de TUSSAM,” [Online]. Available:  
09 <https://www.tussam.es/es/tussam/historia/conoce-la-historia-de-tussam>. [Accessed 14 Septiembre 2022].  
]

[1 C. d. P. H. e. al., “Impacto de los sistemas de apoyo a la explotación (SAE) en la mejora de los servicios de  
10 transporte público urbano,” 2009.

]

[1 M. Ruesga, “Movilidad implantará la prioridad semafórica en más líneas de Tussam,” *Diario de Sevilla*, 20  
11 Septiembre 2016.

]

[1 TUSSAM, “Manual de Operaciones Capítulo 1,” 2021.

12

]

[1 Redacción, “Tussam suprime la lanzadera de la Cartuja,” *Diario de Sevilla*, 10 Enero 2019.

13

]

[1 Encuentro Urbano, [Online]. Available: [http://www.encuentrourbano.es/nuestros-soportes/canales-  
14 informativos-en-autobuses-urbanos/](http://www.encuentrourbano.es/nuestros-soportes/canales-14-informativos-en-autobuses-urbanos/). [Accessed 12 Septiembre 2022].

]

[1 TUSSAM, [Online]. Available: <https://www.tussam.es/es/publicidad/bussi>. [Accessed 12 Septiembre  
15 2022].

]

[1 “Tussam,” [Online]. Available: <https://www.tussam.es/es>. [Accessed 14 Septiembre 2022].

16

]

[1 “Red de Líneas TUSSAM,” [Online]. Available: <https://reddelineas.tussam.es/>.

17

]

[1 TUSSAM, [Online]. Available: [https://www.tussam.es/es/noticias/tussam-incorpora-el-pago-online-  
18 bizum-0](https://www.tussam.es/es/noticias/tussam-incorpora-el-pago-online-18-bizum-0). [Accessed 14 Septiembre 2022].

]

[1 TUSSAM, “Catálogo de Servicios,” 2020.

19

]

[1 Azimut Bus Solutions, [Online]. Available: [https://www.azimutbussolutions.com/blog/por-que-necesito-  
20 un-retrovisor-digital-para-autobus/](https://www.azimutbussolutions.com/blog/por-que-necesito-20-un-retrovisor-digital-para-autobus/). [Accessed 14 Septiembre 2022].

]

[1 R.S., “Tussam eleva a 24 su flota de autobuses urbanos de gas natural,” *Diario de Sevilla*, 28 Febrero 2020.

21

]

[1 Redacción, “[https://www.diariodesevilla.es/sevilla/wifi-gratuito-llega-autobuses-Tussam-Metrocentro-  
22 video\\_0\\_1396660457.html](https://www.diariodesevilla.es/sevilla/wifi-gratuito-llega-autobuses-Tussam-Metrocentro-22-video_0_1396660457.html),” *Diario de Sevilla*, 1 Octubre 2019.

]

[1 TUSSAM, “Memoria anual 2021,” 2021.

23

]

[1 Metro de Sevilla, “Noticias Metro de Sevilla,” [Online]. Available: <https://www.metro-sevilla.es/noticias/43>. [Accessed 11 Septiembre 2022].

]

[1 Metro de Sevilla, “Noticias Metro de Sevilla,” [Online]. Available: <https://www.metro-sevilla.es/noticias/61>. [Accessed 11 Septiembre 2022].

]

[1 Metro de Sevilla, [Online]. Available: <https://www.metro-sevilla.es/tap-and-go>. [Accessed 11 Septiembre 2022].

]

[1 Metro de Sevilla, “Noticias Metro de Sevilla,” [Online]. Available: <https://www.metro-sevilla.es/noticias/44>. [Accessed 11 Septiembre 2022].

]

[1 M. Ávila, “Invertia - El Español,” [Online]. Available: [https://www.elespanol.com/invertia/observatorios/movilidad/20200928/gran-hermano-metro-sevilla-sistema-covid-capitales/523948336\\_0.html](https://www.elespanol.com/invertia/observatorios/movilidad/20200928/gran-hermano-metro-sevilla-sistema-covid-capitales/523948336_0.html). [Accessed 11 Septiembre 2022].

[1 FIWOO, [Online]. Available: <https://www.fiwoo.eu/fiwoo-y-sosteco-realizan-pruebas-con-un-nuevo-dispositivo-iot-en-el-metro-de-sevilla/>. [Accessed 11 Septiembre 2022].

]

[1 Sevilla Actualidad, “Sevilla Actualidad,” [Online]. Available: <https://www.sevillaactualidad.com/sevilla/197266-la-calidad-del-aire-en-los-trenes-del-metro-se-podra-consultar-por-internet/>. [Accessed 11 Septiembre 2022].

[1 Metro de Sevilla, [Online]. Available: <https://www.metro-sevilla.es/calidad-aire>. [Accessed 11 Septiembre 2022].

]

[1 Metro de Sevilla, “Noticias Metro de Sevilla,” [Online]. Available: <https://www.metro-sevilla.es/noticias/29>. [Accessed 12 Septiembre 2022].

]

[1 Consorcio de Transporte Metropolitano del Área de Sevilla, [Online]. Available: <https://www.consorciotransportes-sevilla.com/>. [Accessed 15 Septiembre 2022].

]

[1 Google Play, [Online]. Available: <https://play.google.com/store/apps/details?id=com.baolian.saldoconsorcioandalucia>. [Accessed 15 Septiembre 2022].

[1 Instituto de Estadística y Cartografía de Andalucía, “Padrón Municipal de Habitantes,” [Online]. Available: [https://www.juntadeandalucia.es/institutodeestadisticaycartografia/badea/operaciones/consulta/anual/6675?CodOper=b3\\_128&codConsulta=6675](https://www.juntadeandalucia.es/institutodeestadisticaycartografia/badea/operaciones/consulta/anual/6675?CodOper=b3_128&codConsulta=6675). [Accessed 15 Septiembre 2022].

[1 RENFE, [Online]. Available: <https://www.renfe.com/es/es/cercanias/cercanias-sevilla/horarios>. [Accessed

36 15 Septiembre 2022].

]

[1 M. Ruesga, “El desplome de usuarios obliga a Sevici a estudiar un servicio más atractivo”. *Diario de Sevilla*.

37

]

[1 SEVICI, [Online]. Available:

38 <https://www.sevici.es/es/tutorial/groups?tab=RIDE&type=vls#Los%2030%20primeros%20minutos%20s>

] on%20gratuitos. [Accessed 17 Septiembre 2022].

[1 Junta de Andalucía, “Sobre el Proyecto - Datos Abiertos,” [Online]. Available:

39 <https://www.juntadeandalucia.es/datosabiertos/portal/proyecto>. [Accessed 17 Septiembre 2022].

]

[1 Junta de Andalucía, “Datos Abiertos,” [Online]. Available:

40 [https://www.juntadeandalucia.es/datosabiertos/portal/dataset?vocab\\_field\\_activity\\_sector=Transporte&\\_v](https://www.juntadeandalucia.es/datosabiertos/portal/dataset?vocab_field_activity_sector=Transporte&vocab_field_activity_sector_limit=0&tags_limit=0)

]ocab\_field\_activity\_sector\_limit=0&tags\_limit=0. [Accessed 17 Septiembre 2022].

[1 Red de Consorcios de Transportes de Andalucía, “Portal de Datos Abiertos,” [Online]. Available:

41 <http://api.ctan.es/>. [Accessed 17 Septiembre 2022].

]

[1 Ayuntamiento de Sevilla, “sevilla.org,” [Online]. Available:

42 [https://www.sevilla.org/actualidad/noticias/tussam-gano-12-8-millones-de-viajeros-en-2022-tras-las-](https://www.sevilla.org/actualidad/noticias/tussam-gano-12-8-millones-de-viajeros-en-2022-tras-las-mejoras-del-servicio-el-incremento-de-la-oferta-las-lineas-expres-y-la-rebaja-de-las-tarifas-multiviajes#:~:text=Mientras%2C%20la%20velocidad%20comercial%20)

]mejoras-del-servicio-el-incremento-de-la-oferta-las-lineas-expres-y-la-rebaja-de-las-tarifas-multiviajes#:~:text=Mientras%2C%20la%20velocidad%20comercial%20. [Accessed 26 Abril 2023].

[1 Universidad Autónoma de Madrid, [Online]. Available: [https://www.iic.uam.es/innovacion/big-data-](https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/)

43 [caracteristicas-mas-importantes-7-v/](https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/). [Accessed 25 Julio 2022].

]

[1 Universidad Complutense de Madrid, [Online]. Available: [https://www.masterbigdataucm.com/que-es-](https://www.masterbigdataucm.com/que-es-big-data/)

44 [big-data/](https://www.masterbigdataucm.com/que-es-big-data/). [Accessed 25 Julio 2022].

]

[1 T-Systems, [Online]. Available: <https://www.t-systemsblog.es/big-data-almacenamiento-datos-niveles/>.

45 [Accessed 26 Julio 2022].

]

[1 Tokio School, [Online]. Available: <https://www.tokioschool.com/noticias/fases-big-data/>. [Accessed 27

46 Julio 2022].

]

[1 “masgrande,” [Online]. Available: <https://masgrande.site/geografia/ciudades-mas-grandes-de-europa/>.

47 [Accessed 8 Agosto 2022].

]

[1 Transport For London, “Transport For London,” [Online]. Available: [https://tfl.gov.uk/corporate/privacy-](https://tfl.gov.uk/corporate/privacy-and-cookies/oyster-card)

48 [and-cookies/oyster-card](https://tfl.gov.uk/corporate/privacy-and-cookies/oyster-card). [Accessed 10 Agosto 2022].

]

[1 S. Dodson, “The Guardian,” [Online]. Available:

49 <https://www.theguardian.com/technology/2008/feb/28/research.transport>. [Accessed 11 Agosto 2022].  
]

[1 Equipo Editorial Upnify, [Online]. Available: <https://upnify.com/es/blog/saas-la-oportunidad-dentro-del-mercado-tecnologico.html>. [Accessed 05 Septiembre 2022].  
]

[1 A. García, “Movilidad Conectada,” [Online]. Available: <https://movilidadconectada.com/2018/06/14/los-7-principios-del-mobility-as-a-service-maas/>. [Accessed 5 Septiembre 2022].  
]