



**FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES**

**GRADO EN  
MARKETING E INVESTIGACIÓN DE MERCADOS**

**“REGLAS DE ASOCIACIÓN. APLICACIÓN PRÁCTICA EN  
LA CESTA DE LA COMPRA DE LOS CONSUMIDORES.”**

Trabajo Fin de Grado presentado por **Dña. Cristina Pérez González**, en el curso académico 2017/2018 en la convocatoria ordinaria de septiembre, siendo el tutor del mismo Prof. **Dr. D. Manuel Jesús Sánchez Franco**.

Vº. Bº., del tutor:

Alumna:

Dr. D. Manuel J. Sánchez Franco

Dña. Cristina Pérez González

Sevilla, 20 de julio de 2018

## **RESUMEN EJECUTIVO**

En la era digital en la que nos encontramos, las empresas cada vez más cuentan con grandes cantidades de datos de sus clientes. Esto ha sido posible gracias al avance exponencial de la tecnología y al abaratamiento del almacenamiento de datos. Todo ello ha provocado la necesidad en las empresas de aprender a tratar estos datos y poder obtener de ellos información útil que les permita conocer mejor aún a sus clientes. Existen multitud de técnicas de análisis de datos que permiten obtener la información oculta en las grandes bases de datos, información que puede proporcionar a la empresa una gran ventaja competitiva en el mercado. En este estudio vamos a profundizar en una de esas técnicas, las reglas de asociación, mostrando al lector el potencial que tiene una técnica tan sencilla como esta.

### **Palabras clave:**

Reglas de asociación, análisis de la cesta de la compra, big data, análisis de datos, cesta de la compra

## **ABSTRACT**

In the current digital era, companies increasingly have amounts of data from their customers. This has been possible thanks to the exponential advance of technology and the cheapening of data storage. For this reason, the companies need to learn about data manipulation and to obtain useful information that allows them get to know their customers better. There are multitude of data analysis techniques that allow to obtain information hidden in large databases, information that can provide the company a great competitive advantage in the market. In this study, we will delve into one of these techniques, the association rules, showing the reader the potential of a simple technique like this.

### **Keywords:**

Association rules, market basket analysis, big data, MBA, data analysis



# ÍNDICE DE CONTENIDOS

<b>1.- INTRODUCCIÓN</b>	<b>8</b>
<b>2.- REGLAS DE ASOCIACION</b>	<b>10</b>
2.1.- CONCEPTO DE REGLAS DE ASOCIACIÓN	10
2.1.1- ANÁLISIS DE LA CESTA DE LA COMPRA	13
2.2.- INDICADORES CLAVE DE LAS REGLAS DE ASOCIACIÓN	14
2.3.- ETAPAS DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN	18
<b>3.- CASO PRÁCTICO: ANÁLISIS DE LA CESTA DE LA COMPRA</b>	<b>21</b>
3.1.- RECOGIDA DE DATOS	21
3.2.- ANÁLISIS DESCRIPTIVO DE LA MUESTRA	22
3.3.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA	35
3.3.1.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "SEXO"	36
3.3.2.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "EDAD"	37
3.3.3.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "ESTADO CIVIL"	39
3.4.- ANÁLISIS DE LA CESTA DE LA COMPRA POR SEXO: REGLAS DE ASOCIACIÓN	42
3.4.1.- OBSERVACIONES PREVIAS	42
3.4.2.- APLICACIÓN DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN: REGLAS DE ASOCIACIÓN POSITIVAS	43
3.4.3.- APLICACIÓN DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN: REGLAS DE ASOCIACIÓN NEGATIVAS	49
<b>4.- CONCLUSIONES</b>	<b>54</b>
<b>BIBLIOGRAFÍA</b>	<b>58</b>
<b>ANEXO I: ENCUESTA</b>	<b>61</b>

## ÍNDICE DE TABLAS

TABLA 2.1: DISTRIBUCIÓN DE LA MATRIZ DE DATOS	12
TABLA 2.2: EJEMPLO ILUSTRATIVO	13
TABLA 3.1: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "SEXO"	24
TABLA 3.2: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "EDAD"	25
TABLA 3.3: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "MIEMBROS DE LA UNIDAD FAMILIAR"	26
TABLA 3.4: NÚMERO DE HOGARES PROYECTADOS SEGÚN SU TAMAÑO	26
TABLA 3.5: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "ESTADO CIVIL"	27
TABLA 3.6: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "CREENCIAS RELIGIOSAS"	28
TABLA 3.7: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "NACIONALIDAD"	29
TABLA 3.8: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "NIVEL DE ESTUDIOS"	30
TABLA 3.9: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "OCUPACIÓN"	31
TABLA 3.10: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "ÁMBITO DE RESIDENCIA"	32
TABLA 3.11: DISTRIBUCIÓN DE FRECUENCIAS VARIABLE "SUPERMERCADO"	33
TABLA 3.12: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "SEXO=MUJER"	36
TABLA 3.13: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "EDAD=18-30 AÑOS"	37
TABLA 3.14: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "EDAD=31-65 AÑOS"	38
TABLA 3.15: DATOS SOBRE POBLACIÓN JOVEN EMANCIPADA	39
TABLA 3.16: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=CASADO/A"	40
TABLA 3.17: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=SOLTERO/A"	41
TABLA 3.18: DISTRIBUCIÓN DE FRECUENCIAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=DIVORCIADO/A"	42
TABLA 3.19: RESULTADOS DE LA REGLA DE ASOCIACIÓN POSITIVA {LECHE, PAN} -> {...}	44
TABLA 3.20: RESULTADOS DE LA REGLA DE ASOCIACIÓN POSITIVA {LECHE, PAN, PAPEL HIGIÉNICO}->{TOALLITAS}	46
TABLA 3.21: RESULTADOS DE LA REGLA DE ASOCIACIÓN POSITIVA {LECHE, CARNE DE POLLO} -> {...}	46
TABLA 3.22: RESULTADOS DE LA REGLA DE ASOCIACIÓN {..., ...} -> {COCA-COLA}	47
TABLA 3.23: RESULTADOS DE LA REGLA DE ASOCIACIÓN NEGATIVA {LECHE, PAN} -> {...}	50
TABLA 3.24: RESULTADOS DE LA REGLA DE ASOCIACIÓN NEGATIVA {CARNE DE POLLO} -> {...}	51
TABLA 3.25: RESULTADOS DE LA REGLA DE ASOCIACIÓN NEGATIVA {PAN} -> {CARNE DE POLLO}	53

## ÍNDICE DE GRÁFICOS

GRÁFICO 3.1: GRÁFICO DE BARRAS VARIABLE "SEXO"	24
GRÁFICO 3.2: GRÁFICO DE SECTORES VARIABLE "EDAD"	25
GRÁFICO 3.3: GRÁFICO DE BARRAS VARIABLE "MIEMBROS DE LA UNIDAD FAMILIAR"	26
GRÁFICO 3.4: GRÁFICO DE SECTORES VARIABLE "ESTADO CIVIL"	27
GRÁFICO 3.5: GRÁFICO DE SECTORES VARIABLE "CREENCIAS RELIGIOSAS"	28
GRÁFICO 3.6: GRÁFICO DE SECTORES VARIABLE "NACIONALIDAD"	29
GRÁFICO 3.7: GRÁFICO DE BARRAS VARIABLE "NIVEL DE ESTUDIOS"	30
GRÁFICO 3.8: GRÁFICO DE SECTORES VARIABLE "OCUPACIÓN"	31
GRÁFICO 3.9: GRÁFICO DE SECTORES VARIABLE "ÁMBITO DE RESIDENCIA"	32
GRÁFICO 3.10: GRÁFICO DE BARRAS VARIABLE "SUPERMERCADO"	33
GRÁFICO 3.11: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "SEXO=MUJER"	37
GRÁFICO 3.12: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "EDAD=18-30 AÑOS"	38
GRÁFICO 3.13: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "EDAD=31-65 AÑOS"	39
GRÁFICO 3.14: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=CASADO/A"	40
GRÁFICO 3.15: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=SOLTERO/A"	41
GRÁFICO 3.16: GRÁFICO DE BARRAS DE PRODUCTOS ORDENADOS DE MAYOR A MENOR FRECUENCIA VARIABLE "ESTADO CIVIL=DIVORCIADO/A"	42
GRÁFICO 3.17: GRÁFICO ASOCIACIONES POSITIVAS DE {LECHE, PAN} -> {...}	45
GRÁFICO 3.18: GRÁFICO ASOCIACIONES POSITIVAS DE {LECHE, CARNE DE POLLO} -> {...}	47
GRÁFICO 3.19: GRÁFICO ASOCIACIONES POSITIVAS DE {...} -> {COCA-COLA}	48
GRÁFICO 3.20: GRÁFICO ASOCIACIONES NEGATIVAS DE {LECHE, PAN} -> {...}	51
GRÁFICO 3.21: GRÁFICO ASOCIACIONES NEGATIVAS DE {CARNE DE POLLO} -> {...}	52

## 1.- INTRODUCCIÓN

Hoy en día vivimos lo que se conoce como la era de la información o la era digital, un período de la historia de la humanidad que va muy ligado a las tecnologías y la comunicación. La tecnología forma parte de la vida cotidiana de todo ser humano. El ejemplo más claro es internet; cualquier cosa que queramos hacer, *e.g.*, comprar ropa, reservar un hotel, reservar un vuelo, pedir comida a domicilio, incluso hacer la compra podemos hacerlo a través de internet.

La era de la información supone pues una gran oportunidad para las empresas debido a que a través de la red pueden recopilar enormes cantidades de datos y almacenarlos. Posteriormente esos datos los analizan y los transforman en información útil para sus negocios. Esta necesidad por parte de las empresas del tratamiento de los datos es lo que ha generado una nueva ciencia, nuevas metodología y herramientas para analizar grandes masas de datos como es el Big Data o la minería de datos. Y en suma, el análisis de los datos se realiza con el objetivo clave de conocer al cliente, *i.e.*, qué quiere, cómo lo quiere, cuándo lo quiere supone una gran oportunidad para las empresas de aumentar sus beneficios.

Gracias a la minería de datos y a la informática se han podido desarrollar metodologías y algoritmos predictivos que permiten conocer patrones de comportamiento ocultos del individuo que con otras técnicas es imposible conocer. Esto ayuda a las empresas a entender cómo se comportan sus clientes en el momento de realizar una compra. Una vez conocido su patrón de comportamiento las empresas se adaptan mejor a él con el fin de conseguir la máxima satisfacción, confianza y compromiso del cliente y a su vez aumentar sus beneficios.

Existen multitud de herramientas actualmente para llevar a cabo este tipo de análisis. En este trabajo nos vamos a centrar en una de ellas, las reglas de asociación, una metodología que consideramos con un gran potencial. El estudio comienza con un marco teórico donde se explica con todo detalle y de forma muy sencilla lo necesario para que el lector pueda comprender esta técnica. Posteriormente se llevará a cabo un caso práctico con datos reales donde aplicaremos esta técnica y se expondrán finalmente las conclusiones obtenidas.

*"Sin análisis de big data, las empresas están ciegas y sordas."*

Geofferey Moore.



## 2.- REGLAS DE ASOCIACION

### 2.1.- CONCEPTO DE REGLAS DE ASOCIACIÓN

Se definen las reglas de asociación como una técnica que permite el descubrimiento de relaciones de asociación o correlaciones entre un conjunto de elementos recogidos en una base de datos. Normalmente, se expresan en forma de reglas que muestran las condiciones de valor de los atributos que ocurren con frecuencia juntas en un conjunto de datos dado. Dichas relaciones de asociación o correlaciones es lo que se denomina un patrón.

El concepto de reglas de asociación se popularizó tras la publicación de un artículo de *Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases"*, publicado en 1993. En su estudio, los autores introdujeron las reglas de asociación para descubrir las relaciones entre las transacciones de productos registrados por los sistemas de punto de venta en los supermercados.

Amplie algunos términos para entender mejor el concepto. Una asociación es simplemente la concurrencia de dos o más cosas. Las hamburguesas por ejemplo pueden estar asociadas con los refrescos, pan de hamburguesa o las patatas fritas. Se dice entonces que existe una asociación positiva; *i.e.*, la presencia de algunos elementos implica la presencia de otros elementos dentro de la misma transacción. También es posible que se dé la situación contraria. Una asociación negativa implica que la presencia de algunos elementos puede hacer poco probable o nula la presencia de otros elementos en una misma transacción. Un inversor interesado en comprar una finca rústica no solo le interesa saber la ubicación del terreno, si no también encontrar problemas de calidad ambiental, uso de recursos naturales o problemas de naturaleza política. Asimismo, una transacción es un conjunto de elementos que ocurren a la vez en una observación. Una unidad de transacción muy habitual es una cesta de la compra, un conjunto de productos que se compran al mismo tiempo. Otro ejemplo de transacción también puede ser el conjunto de páginas web que un usuario visita durante una sesión.

Una regla expresa la implicación de un conjunto de artículos a condición de otro conjunto dentro de una transacción con una fuerza, ya sea fuerte o débil. La asociación entre hamburguesa y pan de hamburguesa se expresa de la siguiente forma: {hamburguesa} → {pan de hamburguesa}. Las reglas de asociación pueden expresar también la relación de

múltiples elementos: {hamburguesa, queso cheddar, pepinillos} → {pan de hamburguesas, patatas fritas, refresco de cola, helado de vainilla}.

En suma, la técnica de reglas de asociación es muy fácil de entender. Sin embargo, no siempre arrojan resultados útiles. A continuación, se ilustra los tres tipos de resultado más habituales generados por las reglas de asociación.

### **RESULTADOS FACTIBLES**

Estos resultados contienen información útil y de calidad. Normalmente los patrones que se identifican no son difíciles de justificar y son totalmente factibles de aplicación. Algunas de ellas nos pueden orientar a cómo colocar los productos dentro de un establecimiento o cuáles deberían ser los productos sobre los que se puede aplicar una promoción para incrementar las ventas tanto de dicho producto como de sus productos relacionados. Por ejemplo: si el cliente compra espárragos, también compra caramelos de menta.

### **RESULTADOS TRIVIALES**

Los resultados triviales son conocidos por cualquier profesional familiarizado con el negocio al que se dedica. Es decir, un usuario no se plantea la compra de un producto sin otro. La compra de pinturas requiere de la necesidad de comprar también pinceles o rodillos para su aplicación o al comprar un coche el usuario sabe que necesita contratar un seguro para poder circular. Es por ello que es habitual que los concesionarios al efectuarse la compra de un coche le ofrezcan al usuario la contratación de un seguro.

### **RESULTADOS IMPOSIBLES**

Estos resultados suelen carecer de explicación alguna. La información que proporciona no ayuda a entender el comportamiento del consumidor, más bien suele relacionarse con casos puntuales que ocurren en un momento determinado. Quizás la aplicación de otros tipos de técnicas o análisis ayuden a entender mejor este tipo de resultados y si pueden llegar a tener o no una aplicación comercial.

Una vez expuesto al lector que es una regla de asociación y sus principales elementos y posibles resultados, procedemos a realizar una definición teórica de la misma.

Siguiendo la definición propuesta por Agrawal, Imieliński y Swami (1993) una regla de asociación podemos definirla formalmente como sigue (se ilustra en la tabla 2.1):

- $I = \{i_1, i_2, \dots, i_n\}$  es un conjunto de "n" atributos llamados elementos o ítems.
- $D = \{t_1, t_2, \dots, t_n\}$  es un conjunto de transacciones llamada base de datos donde cada transacción  $t$  o tupla corresponde a un cliente. Cada transacción en  $D$  tiene un identificador de transacción único y contiene un subconjunto de los elementos de  $I$ .
- Dentro de cada transacción,  $A_i = v$  es un elemento o t-item, donde  $v$  puede tomar los valores "0" y "1". Si  $v=1$  indica la presencia del t-ítem en la transacción correspondiente, si  $v=0$  representa la ausencia del t-ítem en esa transacción.

**Tabla 2.1: Distribución de la matriz de datos**

ID TRANSACCIÓN	I <sub>1</sub>	I <sub>2</sub>	I <sub>3</sub>	... I <sub>N</sub>
T <sub>1</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>n</sub>
T <sub>2</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>n</sub>
T <sub>3</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>n</sub>
...T <sub>N</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>n</sub>

Fuente: elaboración propia

La regla de asociación, por tanto, se define como una aplicación de la forma:

$$X \rightarrow Y, \text{ donde } X, Y \subseteq I$$

Cada regla está compuesta por dos conjuntos diferentes de artículos, también conocidos como conjuntos de elementos,  $X$  e  $Y$ , donde  $X$  se llama *antecedente o lado izquierdo* (LHS) e  $Y$  *consecuente o lado derecho* (RHS).

Planteamos un ejemplo ilustrativo para facilitar su comprensión. Supongamos una pequeña base de datos de un supermercado:

**Tabla 2.2: Ejemplo ilustrativo**

ID TRANSACCIÓN	LECHE	PATATAS	AGUA	CEBOLLA	PAN
1	1	1	0	0	1
2	1	1	1	0	1
3	1	1	1	1	1
4	1	1	1	1	0
5	0	0	1	0	0

*Fuente: elaboración propia*

El conjunto de elementos es  $I = \{\text{leche, patatas, agua, cebolla, pan}\}$ . En la tabla podemos observar una pequeña base de datos donde se han realizado 5 transacciones y en cada una de ellas se indica con "1" o "0" si el ítem está o no presente en la transacción. Un ejemplo de regla de asociación basada en el ejemplo ilustrativo anterior puede ser  $\{\text{agua, patatas}\} \rightarrow \{\text{leche}\}$ , lo que significa que si un cliente compra agua y patatas juntos es probable que también compre leche.

Este tipo de reglas de asociación basadas en variables binarias recibe el nombre de **reglas de asociación categóricas**. Es el tipo más común de regla de asociación porque su origen fue el analizar los datos del carro de la compra de forma binaria, es decir, un cliente compró ("1") o no compró ("0") un producto en particular. Sin embargo, también es posible derivar reglas de asociación en variables continuas (escala tipo Likert). A este tipo de reglas de asociación se le denomina **reglas de asociación cuantitativas**. Mientras que en una regla de asociación categórica se relaciona el valor de una variable categórica con un valor de otra variable categórica, en una regla de asociación cuantitativa se relaciona un valor de una variable categórica con una estadística de resumen de una variable continua. Autores como Aumann y Lindell han desarrollado una teoría estadística sobre este tipo de reglas de asociación.

### **2.1.1- ANÁLISIS DE LA CESTA DE LA COMPRA**

Las reglas de asociación es una técnica posible de aplicar en múltiples sectores tales como la bioinformática, la ciencia nuclear, la inmunología o la geofísica. Sin embargo, el sector más frecuente de aplicación es en el sector comercial, concretamente con la aplicación del análisis de la cesta de la compra y sobre el cuál se centra especialmente este trabajo. Procedemos a

definir qué es este tipo de análisis, en qué consiste y a qué tipo de conclusiones puede llegar el investigador con su aplicación.

El **análisis de la cesta de la compra** (en inglés "*market basket analysis*") es una metodología que estudia la composición de una cesta de la compra de productos comprados durante una única transacción. El objetivo es identificar qué productos o categoría de productos tienden a comprarse conjuntamente descubriendo así patrones o relaciones de compra que antes no eran tan evidentes o se desconocían y puedan ser explotadas comercialmente. Además, permite a los investigadores a identificar aquellos elementos que se compran simultáneamente, con qué frecuencia se compran y evaluar el grado de relación.

Con los resultados que aporta esta metodología es posible generar una base de datos acerca de cómo se comporta el consumidor dentro de un establecimiento. Esta información se puede usar para determinar dónde colocar los productos dentro del establecimiento, así como para ayudar a administrar el inventario al personal. Otra aplicación comercial es la emisión de promociones o cupones descuento, el valor nominal del cupón y el tiempo de distribución utilizando la información obtenida de las cestas del mercado. En el ámbito online, es habitual que se capturen las compras que realizan sus clientes a tiempo real para, al cruzarlo con datos históricos de las ventas realizada o con datos de comportamientos similares, proporcionarle al cliente recomendaciones de productos similares al que está buscando o incluso ofrecer productos complementarios para su compra. Un ejemplo de ello puede ser el caso ilustrativo expuesto anteriormente en la tabla 2.2 o el famoso ejemplo de la asociación entre la cerveza y los pañales expuesto por la cadena de supermercados Wal-Mart en los años 90. No obstante, ha habido muchas aplicaciones más allá del sector comercial. En el ámbito empresarial el análisis de la cesta de compra puede ayudar a la toma de decisiones en relación con los beneficios de los empleados, el comportamiento disfuncional de los empleados o la identificación del talento emprendedor.

## **2.2.- INDICADORES CLAVE DE LAS REGLAS DE ASOCIACIÓN**

**Tres indicadores** son los que se deben de tener en cuenta al evaluar las posibles asociaciones obtenidas a través del análisis de la cesta de la compra. Estos indicadores son: **lift**, **soporte y confianza**. La importancia de estos indicadores radica en la información complementaria y

no redundante que estos aportan sobre las reglas obtenidas. También es importante el orden en el que se deben obtener estos indicadores tal y como sigue:

## **LIFT**

El lift nos indica si existe realmente una asociación o no entre uno o varios productos y si esa asociación es positiva o negativa. El motivo por el que se calcula el lift primero es porque se obtiene un indicador muy similar a la prueba de significación en estadística pudiendo así saber si la presencia de una asociación es significativa o ha sido producto del azar.

En su forma matemática el lift se define como  $\frac{P(X \cap Y)}{P(X) * P(Y)}$ , es decir, la probabilidad de que se dé la compra de los dos productos conjuntamente (es decir, que estén asociados) entre la probabilidad de que se dé la compra de los dos productos independientemente el uno del otro (es decir, que no estén asociados).

El lift puede arrojarnos tres posibles resultados:

- Si el lift es igual o muy próximo a uno quiere decir que el numerador es muy similar al denominador, lo cual sugiere que los productos son independientes uno de otro y que su presencia conjunta se ha debido al puro azar.
- Si el lift es menor a uno nos indica que el numerador es más pequeño que el denominador, es decir, la probabilidad de que ambos productos estén asociados es menor que la probabilidad de que ambos productos no estén relacionados lo cual indica que los productos de la asociación considerada guardan una relación negativa (la presencia de X implica la ausencia de Y). También podríamos decir que dichos productos son sustitutivos.
- Si el lift es superior a uno nos indica que el numerador es mayor que el denominador, es decir, la probabilidad de que ambos productos estén asociados es mayor que la probabilidad de que ambos productos no estén relacionados lo cual indica que los productos de la asociación considerada guardan una relación positiva (la presencia de X implica la presencia de Y). También podríamos decir que dichos productos son complementarios.

## SOPORTE

Un elemento  $X$  en una base de datos  $D$  tiene un soporte, denotado como  $\text{supp}(X)$ . El soporte en una regla de asociación  $X \rightarrow Y$  es la proporción de transacciones en  $D$  que contienen tanto  $X$  como  $Y$  (ambos productos coocurren). Es decir,

$$\text{supp}(X) = P(X \cap Y) = \frac{\text{número de transacciones que contienen tanto } X \text{ como } Y}{\text{número total de transacciones}}$$

El valor del soporte se encuentra entre cero y uno y suele expresarse en porcentaje.

De nuevo, en el caso del análisis de la cesta de la compra nos interesa que el soporte de una asociación sea el valor más grande posible, pues cuanto mayor sea nos indica que existe una mayor probabilidad de que la asociación considerada se repita en futuras transacciones.

Se debe tener en cuenta, no obstante, que este indicador presenta un inconveniente. En presencia de un conjunto de datos grande el soporte no es de gran utilidad. En esa situación, el valor del soporte será bastante bajo debido a dos cosas:

1. Por un mayor número de transacciones que además implicarán a muchos más elementos.
2. Puede suceder que los soportes de diferentes reglas de asociación sean muy similares entre sí al tratarse de valores tan bajos. Por lo tanto, en el estudio nos será poco útil para distinguir la fuerza de una regla de asociación de la fuerza de otra.

## CONFIANZA

La confianza es el último indicador que se debe calcular. Este indicador nos orienta sobre la precisión de la asociación que hemos considerado. La confianza de una regla de asociación es la probabilidad de que dicha regla se cumpla para una nueva transacción que contenga el/los productos del lado izquierdo de la condición. Si lo definimos de forma teórica viene a ser:

$$C(X \rightarrow Y) = P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{\text{número de transacciones que contienen } X \text{ e } Y}{\text{número de transacciones que contiene } X}$$

También puede expresarse como:  $\frac{\text{supp}(X)}{P(X)}$

Al igual que el soporte el valor de la confianza se encuentra entre cero y uno y suele expresarse en porcentaje.

Una ventaja que presenta la confianza frente al soporte es que no se ve afectada por el tamaño del conjunto de datos ya que solo considera aquellas transacciones donde estén involucrados X e Y. Además (como se ha expresado anteriormente) dado que la confianza también puede calcularse como el soporte por la probabilidad de X, el valor de la confianza en casi todos los casos va a ser mayor que el valor del soporte. Entonces cuando varias reglas de asociación tengan un valor del soporte muy similar, es muy probable que se diferencien mucho en los valores de confianza. Por lo tanto, en comparación con el soporte, la confianza es capaz de detectar con mayor claridad diferencias en las fortalezas de las reglas de asociación.

Tras la explicación de los distintos indicadores para evaluar una regla de asociación es el momento de preguntarse, ¿qué valores son los más adecuados para cada indicador? Obviamente va a depender por un lado de la contextualización del estudio que se esté realizando y por otro lado del criterio del investigador. Si lo contextualizamos a nuestro caso, el estudio del análisis de la cesta de la compra:

- Nos interesa encontrar asociaciones con un valor del **lift mayor o menor que uno**. Las asociaciones con un lift igual o muy próximo a uno, tal y como lo hemos descrito anteriormente, nos indica que no existe ninguna asociación entre los productos y que su presencia se debe al puro azar. Por tanto, no tendría ningún sentido aplicar en ellas técnicas comerciales porque no se produciría ningún incremento en el volumen de ventas de los productos.
- El valor del **soporte** no tiene por qué ser muy elevado. En la práctica se suele usar **valores** tales como **0,01, 0,10, 0,20, etc.** El elegir uno y otro va a depender de las características del negocio y de los datos para que nos aporte un resultado significativo y útil.
- De manera similar ocurre con la **confianza**. Es cierto que cuanto más alta sea será mejor para obtener una conclusión fiable del análisis. En la práctica se suelen usar **valores** entorno **al valor 0,8** pero un valor tan alto aplicado en una base de datos pequeña va a resultar demasiado restrictivo; quizás con un valor a 0,2, 0,3... es suficiente. De nuevo va a depender de las características del negocio y de los datos.



## 2.3.- ETAPAS DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN

En este apartado procedemos a describir cuáles son las etapas que se han de llevar a cabo para realizar un análisis de reglas de asociación o análisis de la cesta de la compra. Estas etapas se han de llevar a cabo cuidadosamente y con riguroso orden para obtener unos resultados fiables y correctos.

### 1. Determinar la idoneidad del análisis de la cesta de la compra.

Como se ha explicado en el apartado 1 de esta sección, el análisis de la cesta de la compra se utiliza para encontrar asociaciones entre variables categóricas, generalmente binarias. También es posible aplicarlo a un conjunto de datos de variables continuas recopilados en una escala tipo Likert, por ejemplo. Sin embargo, no podríamos aplicar este tipo de análisis en conjuntos de datos recopilados en variables continuas solamente porque no sería posible establecer asociaciones como las que propone esta técnica. En este sentido, es estrictamente necesario que haya al menos una variable categórica para poder establecer una asociación. Es decir, como primer aspecto a tener en cuenta es el **tipo de variable** sobre el que se va a aplicar la regla de asociación.

Por otro lado, se ha de tener en cuenta el **objetivo del investigador** al aplicar este tipo de análisis. El análisis de la cesta de la compra tiene un gran potencial para identificar relaciones multinivel, de contingencia y dinámicas. Para aplicarlo no es necesario que el conjunto de datos cumpla supuestos como normalidad o linealidad, entre otros. Además, los resultados que aporta este tipo de análisis pueden comunicarse fácilmente a una audiencia practicante. Si el fin del investigador es alguno de los mencionados anteriormente, el análisis de la cesta de la compra es una buena alternativa para llevar a cabo este tipo de análisis exploratorios. Sin embargo, si el fin del investigador es conseguir fundamentos para validar una teoría, el análisis de la cesta de la compra no es apropiado para ello debido a su naturaleza exploratoria.

### 2. Definir las transacciones.

Desde el punto de vista del marketing una transacción típica consiste en un conjunto de productos comprados por un cliente en un determinado establecimiento o sitio web y en un momento determinado. Estas transacciones pueden considerarse durante un período como sea un mes, un trimestre o un año creando así múltiples niveles de análisis. La recopilación de

estas transacciones a través de un programa informático nos ayudará a construir observaciones para estudiar el comportamiento de los individuos dentro de un supermercado e identificar posibles reglas de asociación entre productos o categorías de productos que se desconocen.

### **3. Recopilación de los datos.**

El análisis de la cesta de la compra fue diseñado originariamente para su uso con grandes conjuntos de datos. Es posible que los investigadores recolecten los datos por sus propios medios a través, por ejemplo, de encuestas pero debido al esfuerzo, el tiempo y los problemas que de accesibilidad a las fuentes de datos necesarias es muy habitual que utilicen datos recopilados por terceros. Lo más probable es que el investigador busque acuerdos con organizaciones que proporcionarán datos a cambio de realizar análisis y presentar resultados a quienes proporcionaron dichos datos.

### **4. Verificar los requisitos del análisis de la cesta de la compra.**

#### **4.1.- Tamaño de la muestra**

Para aplicar el análisis de la cesta de la compra es necesario tener un número suficientemente grande de transacciones; de lo contrario, el investigador encontraría pocas reglas de asociación, y si las hubiera los valores del lift serían muy próximos a uno.

Afortunadamente es bastante sencillo cumplir con este requisito debido a la gran cantidad de reservas de datos con las que cuentan las organizaciones debido a la era de la información en la que vivimos actualmente. En el ámbito del marketing no es inusual que los estudios incluyan tamaños de muestra que van desde decenas de miles a millones de transacciones, aunque también se han utilizado muestras tan pequeñas como unos cientos de transacciones.

En cualquier caso, es importante que el investigador explique en detalle la fuente de la que ha obtenido los datos utilizados en el análisis para que los lectores puedan estar completamente informados con respecto a la medida en que las reglas de asociación pueden generalizarse a otras configuraciones.

#### **4.2.- Verificar que las "no respuestas".**

Otro aspecto a tener en cuenta antes de iniciar el análisis es que la ausencia de respuesta en algunos ítems, las "no respuestas", no sean sustancialmente significativas. Para ello, el investigador puede ayudarse de técnicas exploratorias como tablas de frecuencia para identificar los ítems que no obtuvieron u obtuvieron muy pocas respuestas por parte de los individuos. De esta manera, el investigador podrá optar por eliminar el/los ítem/s de la investigación o considerar esos ítems teniendo especial cautela al interpretar los datos de lift, soporte y confianza.

Por tanto, confirmado que las no respuestas no son sustantivamente significativas el investigador puede proceder a realizar el análisis.

#### **5. Obtención de las reglas de asociación y sus índices.**

Llegados a este punto, el investigador debe ingresar el conjunto de transacciones en un programa apropiado para extraer tanto las reglas de asociación como sus índices correspondientes descritos en el apartado 2: lift, soporte y confianza. Para estos últimos, el investigador deberá especificarlos valores de corte del soporte y la confianza para poder concluir si la presencia de una asociación es significativa o por el contrario no lo es. Esto resulta ser bastante útil con vistas a que el programa no produzca un número muy grande de asociaciones al tener un conjunto de transacciones considerable.

Existen varios paquetes de software disponibles en el mercado que nos permiten realizar el análisis de la cesta de la compra, por ejemplo, IBM SPSS Modeler, SAS Enterprise Miner, Excel, Python... En nuestro caso para el desarrollo de este trabajo utilizaremos el software estadístico R.

#### **6. Interpretación de las reglas de asociación obtenidas.**

La última etapa del análisis de la cesta de la compra es interpretar las reglas de asociación obtenidas a través del software utilizado. Tanto la etapa 5 como la 6 podrá verla el lector más

detallada y con una explicación más extendida en el apartado 3.4 del presente trabajo donde realizaremos un caso práctico utilizando el software estadístico RStudio.

### **3.- CASO PRÁCTICO: ANÁLISIS DE LA CESTA DE LA COMPRA**

En este apartado se llevará a cabo un caso práctico detallado sobre el análisis de la cesta de la compra a partir de una base de datos real. Para llevarlo a cabo utilizaremos dos softwares informáticos: *SPSS Statistics* y *RStudio*.

#### **3.1.- RECOGIDA DE DATOS**

Para llevar a cabo este caso práctico contamos con una base de datos real obtenida a través de una encuesta elaborada por tres compañeras de cuarto curso del grado de Marketing e Investigación de Mercados de la Universidad de Sevilla. Dicha encuesta fue redactada con la herramienta *Google Form*, la cual puede consultar el lector en el anexo I del presente documento.

El objetivo que se pretendía conseguir con la redacción de esta encuesta era, por un lado, conocer al encuestado realizando preguntas que respondieran a su sexo, edad, estado civil, entre otras, contando así con 9 items que lo describen. De esta manera, podamos construir un perfil de cliente completo. Por otro lado, queríamos conocer qué productos había comprado dicho cliente recientemente; por ello se le solicita al encuestado que indique los productos que había comprado en el supermercado en los últimos 15 días confeccionando así una cesta de la compra del cliente.

La encuesta estuvo disponible para su realización durante dos semanas y **se consiguieron 128 respuestas** de las cuáles fueron **válidas** para llevar a cabo este caso práctico **126 respuestas**.

### **3.2.- ANÁLISIS DESCRIPTIVO DE LA MUESTRA**

Procedemos a analizar la muestra de este estudio en función de las variables: Sexo, Edad, Estado Civil, Miembros del hogar, Religión, Nacionalidad, Nivel de estudios, Ocupación, Residencia y Supermercado. Los distintos valores que puedan tomar cada una de las variables van a condicionar la composición de la cesta de la compra del individuo. Por ello, importante conocer cuál es la utilidad de cada una de ellas dentro de este estudio.

El contenido de la cesta de la compra va a depender si el individuo que la realiza es un hombre o una mujer. Hoyer y MacInnis (2012) afirman que existen diferencias de conducta entre ambos sexos; mientras que las mujeres son más minuciosas en examinar los estímulos de marketing que perciben y toman decisiones con base a los atributos del producto, los hombres procesan más selectivamente los estímulos buscando la simplicidad. Es por ello que las mujeres van a implicarse más en el proceso de compra planificándolas y acudiendo a distintos supermercados para comprar los productos que necesitan en función de diferentes estímulos: descuentos, promociones, ofertas, etc. Para ellas ir de compras supone una actividad placentera y estimulante. Sin embargo, para los hombres ir de compras es una simple tarea. Van a tener una menor implicación en el proceso de compra; no van a planificarlas, van a actuar por impulso y probablemente acudan a un único supermercado donde puedan adquirir todos los productos que necesitan.

Otro aspecto importante que puede condicionar la cesta de la compra de un individuo es la configuración del hogar. La edad, el estado civil, los miembros de la unidad familiar y la ocupación del individuo son variables que nos van a permitir conocer aspectos importantes sobre el perfil del encuestado para comprender por qué realizan unas compras y no otras. Existen diversos estudios sobre el hogar que establecen relaciones entre las variables que estamos considerando y el comportamiento de compra. Exponemos algunos ejemplos. Mägi (2003) señala que un mayor volumen de compra puede estar relacionado con un mayor tamaño de la unidad familiar. En consecuencia, los hogares más grandes dispondrán de menos tiempo para realizar sus compras y tenderán a concentrarlas en un menor número de supermercados o incluso en uno solo.

East, Harris, Willson y Lomax (1995) y Mägi (2003) también exponen en sus estudios que edad del responsable de la compra se encuentra relacionada directamente con el reparto del gasto en diversos supermercados. East, Harris, Willson y Lomax (2000) aportan una posible

explicación a esta causa, los consumidores de mayor edad (como caso extremo personas que estén jubiladas) disponen de más tiempo libre y son capaces de invertir más tiempo en realizar la compra en distintos supermercados simultáneamente.

El ámbito de residencia del individuo también es un aspecto a considerar en el comportamiento de compra. Una persona que viva en una ciudad tendrá una mayor disponibilidad de supermercados para poder realizar sus compras y podrá planificarlas según su conveniencia de acudir a un establecimiento o a otro. En cambio, una persona que viva en un entorno rural cuenta con un abanico de establecimientos más reducido para realizar sus compras y por ello deberán desplazarse si desea adquirir cierto tipo de productos los cuales no tiene a su disposición en su localidad. Deberá invertir más tiempo en planificar y realizar la compra.

La variable nacionalidad era interesante contemplarla por conocer si al tener distintos individuos nacionalidades diferentes los productos y los supermercados a los que acuden a realizar sus compras también lo son. Algo parecido puede ocurrir con rasgos más personales del individuo como puede ser la educación o la creencia religiosa. Si el individuo ha sido educado en un entorno de sostenibilidad y respeto al medio ambiente, probablemente tenderá a comprar productos BIO o acudir a supermercados que compartan esta filosofía. En cuanto a su creencia religiosa hay que tener en cuenta de que existen multitud de religiones y algunas de ellas imponen restricciones alimenticias. Por ejemplo, el islam no permite a sus creyentes el consumir carne de cerdo o bebidas alcohólicas o el budismo donde generalmente sus creyentes llevan una dieta vegetariana y no deben consumir raíces como zanahorias, patatas o cebolla. Como podemos intuir, la religión también se ha de tener en consideración porque va a condicionar la cesta de la compra del individuo.

Para analizar todas estas variables de las que hemos hablado utilizaremos el software informático *SPSS Statistics*. Dado que todas las variables que vamos a analizar en este apartado son cualitativas, solo podremos obtener de ellas las tablas de frecuencia y la moda.

## ▪ SEXO

Contamos con una muestra de 126 encuestados, en la cual el papel de la mujer es bastante representativo. Concretamente son 102 mujeres las que han contestado a nuestra encuesta representando un 81% del tamaño de la muestra. El número de hombres ha sido bastante menos de lo esperado siendo solo 24 los que han contestado a la encuesta (19%).

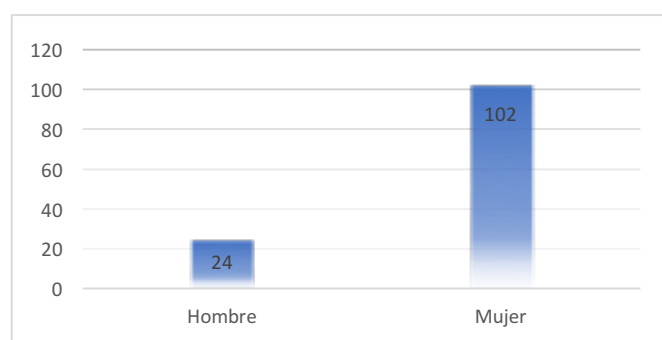
**Tabla 3.1: Distribución de frecuencias variable "Sexo"**

		Sexo del encuestado			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Hombre	24	19,0	19,0	19,0
	Mujer	102	81,0	81,0	100,0
Total		126	100,0	100,0	

*Fuente: elaboración propia con los resultados obtenidos en SPSS*

En el gráfico de barras que sigue se puede observar una representación de los datos obtenidos en la tabla 3.1. Nótese la diferencia significativa comentada anteriormente entre ambos sexos en cuanto a la participación en este estudio.

**Gráfico 3.1: gráfico de barras variable "Sexo"**



*Fuente: elaboración propia con los resultados obtenidos en SPSS*

Este resultado, aunque resulte llamativo, concuerda con la realidad existente; las mujeres son las que más acuden al supermercado a realizar la compra. De hecho, el barómetro del CIS de mayo de 2017 afirma que de las respuestas de 2.479 personas mayores de 18 años a las que se le realizó la encuesta el 42,3% de las mujeres afirman realizar ellas la compra del hogar frente al 20,9% de los hombres que afirman realizarla ellos.

#### ▪ **EDAD**

Para la variable **Edad** se establecieron cuatro rangos considerando a cada uno de ellos como un segmento concreto de la población. Consideramos:

- Adolescentes a "menos de 18 años".
- Jóvenes a "entre 18 y 30 años"
- Adultos a "entre 31 y 65 años"
- Y Tercera edad a "más de 65 años"

Entre la población encuestada contamos con un importante número de personas consideradas "Jóvenes" siendo un 55,6% del total (70 individuos) seguido de los "Adultos" siendo un 43,7% (55 individuos). Solo un "Adolescente" rellenó nuestra encuesta y no contamos con ninguna respuesta de personas con edad superior a 65 años.

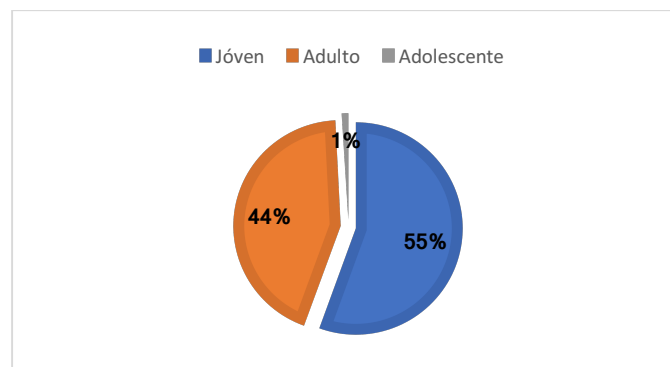
**Tabla 3.2: Distribución de frecuencias variable "Edad"**

		Edad del encuestado			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Jóven</b>	70	55,6	55,6	55,6
	<b>Adulto</b>	55	43,7	43,7	99,2
	<b>Adolescente</b>	1	0,8	0,8	100,0
<b>Total</b>		126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

En el siguiente gráfico de sectores puede verse de una forma más clarificadora lo comentado anteriormente.

**Gráfico 3.2: gráfico de sectores variable "Edad"**



Fuente: elaboración propia con los resultados obtenidos en SPSS



## ▪ NUMERO DE MIEMBROS DE LA UNIDAD FAMILIAR

Entre la población encuestada se encontró que la unidad familiar predominante es de 4 miembros en un 41,3%, seguido de unidades de 3 miembros con un 29,4%, con un 13,5% la unidad familiar de 2 miembros y las unidades familiares de 1 miembro y de 5 o más miembros con un 8,7% y 7,1% respectivamente.

**Tabla 3.3: distribución de frecuencias variable "Miembros de la unidad familiar"**

Número de miembros de la unidad familiar					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>1 miembro</b>	11	8,7	8,7	8,7
	<b>2 miembros</b>	17	13,5	13,5	22,2
	<b>3 miembros</b>	37	29,4	29,4	51,6
	<b>4 miembros</b>	52	41,3	41,3	92,9
	<b>5 miembros o más</b>	9	7,1	7,1	100,0
	Total	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

Resulta llamativo que el número de miembros con mayor representación en nuestra encuesta sea de 4 miembros si nos basamos en los últimos datos recogidos por el INE.

Según sus últimos datos publicados sobre "Número de hogares proyectados según su tamaño", el número de miembros que predomina es el de 2 personas y parece tener una tendencia creciente con el paso del tiempo.

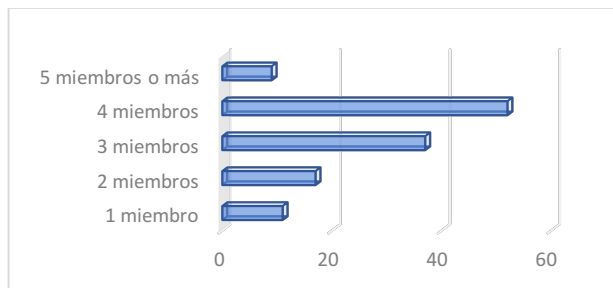
**Tabla 3.4: Número de hogares proyectados según su tamaño**

	AÑO 2016	AÑO 2031
<b>TOTAL DE HOGARES</b>	18.378.691	19.281.354
<b>HOGARES DE 1 PERSONA</b>	4.611.129	5.522.762
<b>HOGARES DE 2 PERSONAS</b>	5.617.423	6.192.325
<b>HOGARES DE 3 PERSONAS</b>	3.857.493	3.818.805
<b>HOGARES DE 4 PERSONAS</b>	3.242.387	2.958.098
<b>HOGARES DE 5 O MÁS PERSONA</b>	1.050.259	789.363

Fuente: INE

En el siguiente gráfico de barras se pueden observar las proporciones de cada uno de las opciones expresadas según su frecuencia.

**Gráfico 3.3: gráfico de barras variable "Miembros de la unidad familiar"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ ESTADO CIVIL

En cuanto al estado civil de los encuestados, resulta estar representado por una mayoría de personas solteros/as en un 56,3% (71 personas), seguido de un 34,1% de personas casadas, un 3,2% de personas divorciadas y un 2,4% de personas separadas.

En la encuesta se incluyó una opción con el nombre "Otro" donde el encuestado podría escribir su estado civil si no se consideraba representado por ninguno de los ítems propuestos. Este ítem representa en la muestra un 4% (está sombreada en amarillo en la tabla que sigue). Los datos introducidos por los 5 encuestados que rellenaron este apartado no representan realmente un estado civil reconocido legalmente; por tanto, no los consideraremos como tales en el estudio y lo mantendremos con la categoría "Otro".

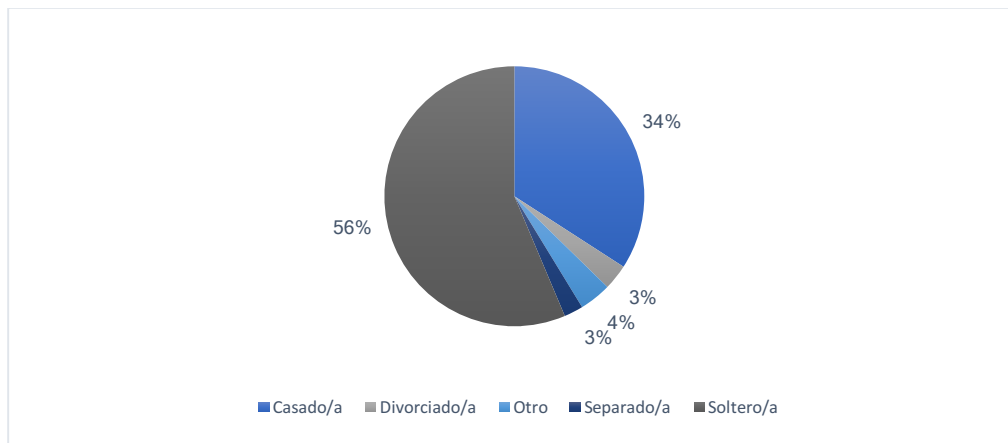
**Tabla 3.5: Distribución de frecuencias variable "Estado civil"**

Estado civil del encuestado					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Casado/a</b>	43	34,1	34,1	34,1
	<b>Divorciado/a</b>	4	3,2	3,2	37,3
	<b>Otro</b>	5	4,0	4,0	41,3
	<b>Separado/a</b>	3	2,4	2,4	43,7
	<b>Soltero/a</b>	71	56,3	56,3	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

En el siguiente gráfico de sectores se pueden observar las proporciones de cada uno de las opciones expresados en porcentaje (%).

**Gráfico 3.4: Gráfico de sectores variable "Estado civil"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ RELIGIÓN

El 54,8% de las personas (69 individuos) encuestadas se consideran católicos frente a un 25,4% que se considera ateo o no creyente en ninguna religión. Un 3,2% de la población cree en otras religiones no representadas en los ítems y un 16,7% prefiere no contestar a esta pregunta.

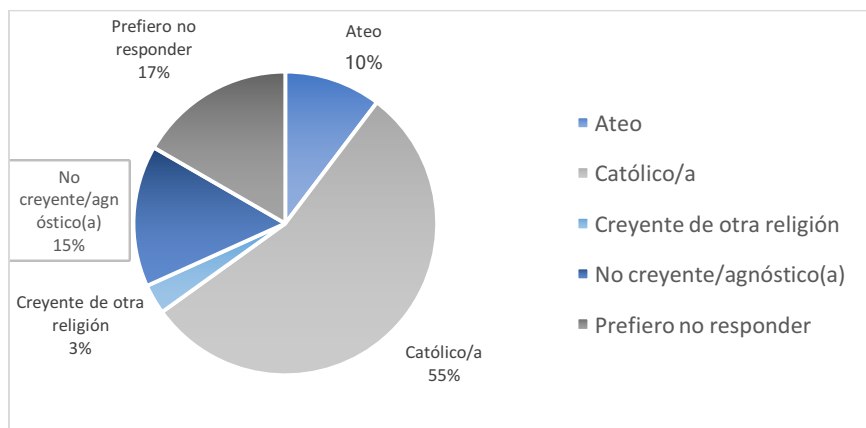
**Tabla 3.6: Distribución de frecuencias variable "Creencias religiosas"**

Creencias religiosas del encuestado					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ateo	13	10,3	10,3	10,3
	Católico/a	69	54,8	54,8	65,1
	Creyente de otra religión	4	3,2	3,2	68,3
	No creyente/agnóstico(a)	19	15,1	15,1	83,3
	Prefiero no responder	21	16,7	16,7	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

En el gráfico 3.5 de sectores se pueden observar sus proporciones expresadas en porcentaje (%).

**Gráfico 3.5: Gráfico de sectores variable "Creencias religiosas"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ NACIONALIDAD

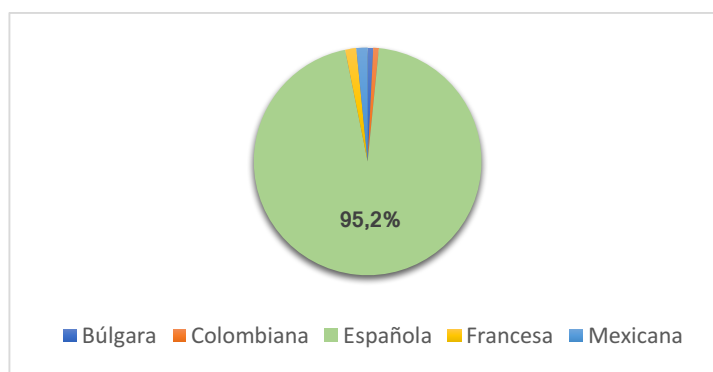
Prácticamente, la totalidad de la población encuestada es de nacionalidad española con un 95,2% tal y como se muestra en la tabla de frecuencia y el gráfico de sectores que sigue. Otras nacionalidades recogidas son: búlgara (0,8%), francesa (1,6%), colombiana (0,8%) y mexicana (1,6%).

**Tabla 3.7: Distribución de frecuencias variable "Nacionalidad"**

Nacionalidad del encuestado					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Búlgara</b>	1	,8	,8	,8
	<b>Colombiana</b>	1	,8	,8	1,6
	<b>Española</b>	120	95,2	95,2	96,8
	<b>Francesa</b>	2	1,6	1,6	98,4
	<b>Mexicana</b>	2	1,6	1,6	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

**Gráfico 3.6: Gráfico de sectores variable "Nacionalidad"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ NIVEL DE ESTUDIOS

En términos de educación más de la mitad de los encuestados cuentan con estudios universitarios siendo un 51,6%, seguido del 23,8% de individuos que cuentan con estudios de formación profesional. El 24,6% restante cuenta con estudios de menor importancia: bachillerato (11,1%), secundaria (7,1%), primaria (5,6%) y sin estudios (0,8%).

**Tabla 3.8: Distribución de frecuencias variable "Nivel de estudios"**

		Nivel de estudios del encuestado			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Bachillerato</b>	14	11,1	11,1	11,1
	<b>Estudios universitarios</b>	65	51,6	51,6	62,7
	<b>Formación profesional</b>	30	23,8	23,8	86,5
	<b>Primaria</b>	7	5,6	5,6	92,1
	<b>Secundaria</b>	9	7,1	7,1	99,2
	<b>Sin estudios</b>	1	,8	,8	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

**Gráfico 3.7: Gráfico de barras variable "Nivel de estudios"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ OCUPACIÓN

Se le preguntó también a los encuestados sobre su ocupación. El 48,4% se encuentran trabajando frente a un 15,1% que están en situación de desempleo. El 34,1% son estudiantes y el 2,4% están jubilados.

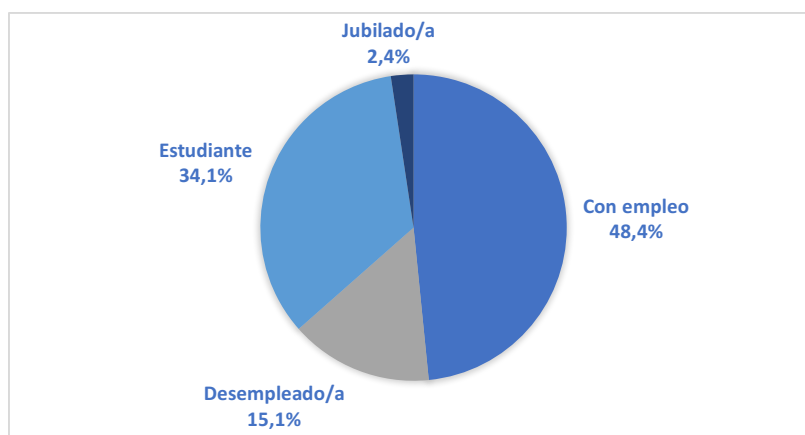
**Tabla 3.9: Distribución de frecuencias variable "Ocupación"**

Ocupación del encuestado					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Con empleo</b>	61	48,4	48,4	48,4
	<b>Desempleado/a</b>	19	15,1	15,1	63,5
	<b>Estudiante</b>	43	34,1	34,1	97,6
	<b>Jubilado/a</b>	3	2,4	2,4	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

Su distribución sectorial puede verse en la gráfica 3.8 representada en porcentaje (%).

**Gráfico 3.8: Gráfico de sectores variable "Ocupación"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

## ▪ ÁMBITO DE RESIDENCIA

Nos resulta de interés en este estudio conocer donde reside el encuestado. Este ítem se ha dividido en dos opciones: en un entorno urbano (una ciudad, un pueblo grande...) o en un entorno rural (pueblo pequeño, aldea, pueblo dormitorio...).

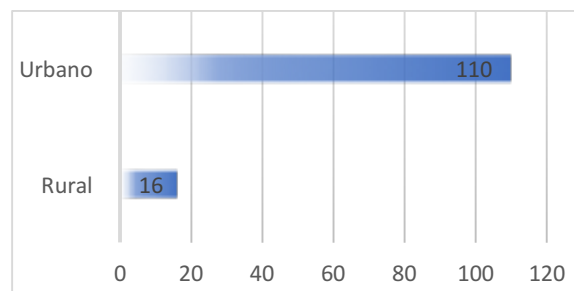
Entre la población encuestada se encontró que el 87,3% vive en un entorno urbano frente al 12,7% que vive en un entorno rural.

**Tabla 3.10: Distribución de frecuencias variable "Ámbito de residencia"**

Ámbito de residencia del encuestado					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Válido</b>	<b>Rural</b>	16	12,7	12,7	12,7
	<b>Urbano</b>	110	87,3	87,3	100,0
	<b>Total</b>	126	100,0	100,0	

Fuente: elaboración propia con los resultados obtenidos en SPSS

**Gráfico 3.9: Gráfico de sectores variable "Ámbito de residencia"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

### ▪ SUPERMERCADO

La variable "Supermercado" era el único ítem presente en la encuesta con opción a respuesta múltiple, es decir, el encuestado podía seleccionar más de una opción.

Del 100% de los encuestados una importante mayoría manifiesta que compra habitualmente en Mercadona (90,5%), seguido de Supermercados Día con un 34,9% y Carrefour con un 31%. Existe un 22,2% que afirma comprar en otros supermercados diferentes a los propuestos. Algunas de las respuestas han sido Erosky, Covirán, Aldi, entre otros, incluso las pequeñas tiendas de Ultramarinos o Desavíos que se pueden encontrar en su zona residencial.

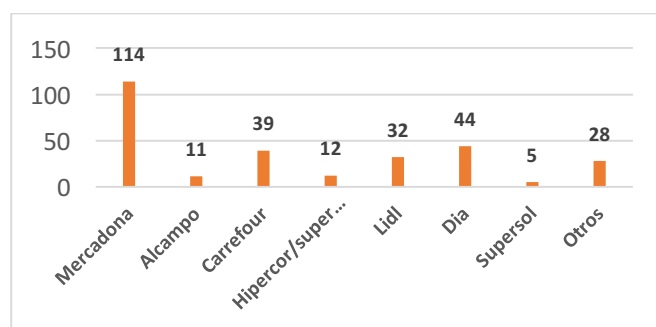
**Tabla 3.11: Distribución de frecuencias variable "Supermercado"**

		Frecuencia	Porcentaje de casos
<b>Supermercado en el que compra en indiv.</b>	<b>Mercadona</b>	114	90,5%
	<b>Alcampo</b>	11	8,7%
	<b>Carrefour</b>	39	31,0%
	<b>Hipercor/supercor</b>	12	9,5%
	<b>Lidl</b>	32	25,4%
	<b>Día</b>	44	34,9%
	<b>Supersol</b>	5	4,0%
	<b>Otros</b>	28	22,2%
<b>Total</b>		-	-

Fuente: elaboración propia con los resultados obtenidos en SPSS



**Gráfico 3.10: Gráfico de barras variable "Supermercado"**



Fuente: elaboración propia con los resultados obtenidos en SPSS

Una vez realizado este análisis descriptivo de los resultados obtenidos de cada una de las variables consideradas en el estudio, podemos exponer al lector algunas observaciones.

Contamos con algunas variables en las que podemos eliminar alguna de sus opciones por no tener una representación significativa dentro de la muestra y, por tanto, no las tendremos en consideración en este estudio porque no tendremos suficientes fundamentos empíricos para poder extrapolarlo al conjunto.

Una de ellas es la variable "Sexo". En nuestra muestra solo contamos con 24 respuestas de hombres frente a las 102 respuestas de mujeres obtenidas. Ante la poca representatividad de la presencia de hombres en la muestra consideramos conveniente eliminarlos de nuestro estudio a la hora de analizar esta variable.

Igual nos ocurre con la variable "Nacionalidad" y "Ámbito de residencia". El 95,2% de los encuestados son de nacionalidad española, con lo cual deseamos el resto de nacionalidades que no tiene una representación significativa en la población encuestada.

En la variable "Ámbito de residencia" el 87,3% de la población encuestada vive en un entorno urbano. Al contar con un número de respuestas tan significativo frente a un 12,7% que corresponde al entorno rural, deseamos esta última y no se tendrá en cuenta para el resto del estudio.

También añadir dos observaciones sobre la variable "Edad" y "Nivel de estudios". En ambas contamos con una sola respuesta para una de las opciones propuestas. Estas opciones

consideramos conveniente que sean eliminadas del estudio ya que no van a aportar información relevante al estudio y tampoco va ser posible extrapolar los resultados obtenidos a este grupo debido a su poca representatividad en la muestra. Por tanto, en la variable "Edad" eliminamos la opción "Adolescente", con lo cual nos quedamos con un rango de edad de la muestra de entre 18 y 65 años y en la variable "Nivel de estudios" eliminamos la opción "Sin estudios" considerando así que todos los encuestados tienen al menos estudios primarios.

En resumen, la muestra con la que contamos para este estudio es: mujeres de nacionalidad española y que viven en un entorno urbano. Los individuos cuentan con una edad de entre 18 y 65 años y su familia se compone mayoritariamente de 4 miembros. Están casados, son de religión católica, tienen algún tipo de cualificación académica mayoritariamente estudios universitarios y actualmente se encuentran con empleo. Los individuos encuestados manifiestan que realizar sus compras habitualmente en Mercadona.

### **3.3.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA**

En este apartado procederemos a realizar un análisis de la cesta de la compra aportada por los individuos encuestados. La analizaremos de forma que extraeremos los productos más frecuentemente comprados en función de las variables: "Sexo", "Edad" y "Estado civil". Hubiese sido interesante incluir en este análisis la variable "Supermercado" pero se nos plantean dos problemas:

1. La forma en la que hemos recogido esta variable en la encuesta no nos permite relacionarla con los productos comprados sino con el individuo, es decir, no hemos preguntado dónde ha comprado los productos que nos facilita en la encuesta, simplemente hemos recogido dónde realiza sus compras el individuo con mayor frecuencia.
2. Las respuestas obtenidas en relación con la variable "Supermercado" no se encuentran ordenadas según la preferencia del comprador al elegir las opciones. Por ello tampoco podemos restringir el análisis a un solo supermercado o solo al primer supermercado que nos proporciona la muestra porque estaríamos aportando conclusiones que no son reales.

Estos dos inconvenientes citados se podrían solventar configurando de otra forma la pregunta donde se recoge la variable "Supermercado".

Volviendo al inicio de este apartado, realizaremos el análisis comentado utilizando el software estadístico *RStudio*. Nosotros hemos decidido usar *RStudio* por su comodidad y facilidad a la hora de manejar y obtener el tipo de resultado que estamos buscando. Puede consultar los comandos que se han utilizado para la obtención de los resultados en el Anexo 2 del presente documento.

### 3.3.1.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "SEXO"

Estos son los resultados obtenidos en relación a la variable "Sexo" sobre los productos comprados con mayor frecuencia:

**Tabla 3.12: Distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable "Sexo=Mujer"**

Sexo: Mujer	
Producto	Frecuencia
Leche	37
Carne de pollo	33
Queso	30
Pan	26
Huevos	25
Coca-cola	23
Tomate frito	21
Yogurt	21

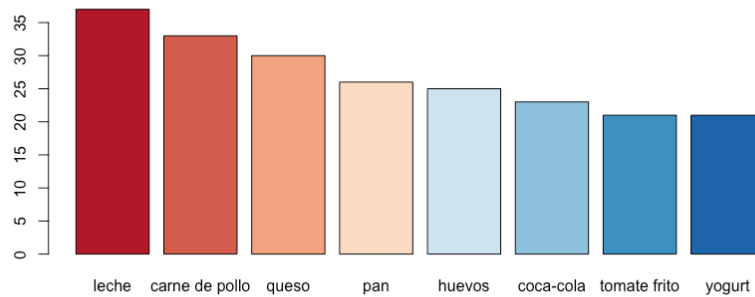
*Fuente: elaboración propia con los resultados obtenidos en RStudio*

Como se puede observar en la tabla, los dos productos más frecuentemente comprado por parte de las mujeres son los productos "leche" y "carne de pollo", dos bienes considerados de primera necesidad y por tanto es lógico que nos salga en nuestros resultados al igual que "pan" y "huevos" que se encuentra en la posición cuatro y cinco respectivamente.

Resulta llamativo que el producto "coca-cola" aparezca también entre los primeros puestos de frecuencia de compra antes que "agua" por ejemplo. En este hecho se puede observar lo que apunta el informe "Brand Footprint 2018" de la consultora Kantar Worldpanel donde afirma que "la enseña de este popular refresco sigue siendo la marca más elegida por los hogares españoles" además de ser una de las más conocidas a nivel global. Realmente, al recoger

estos datos desconocemos si este producto "coca-cola" realmente sea el producto original de la marca Coca-Cola, pero al menos podemos afirmar que el refresco de cola tiene una destacada presencia en la cesta de la compra de los españoles.

**Gráfico 3.11: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable "Sexo=Mujer"**



Fuente: elaboración propia con los resultados obtenidos en R

### 3.3.2.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "EDAD"

Para este apartado contamos con dos rangos de edad representativos de la muestra obtenida. Empezaremos con el rango de edad entre 18-30 años.

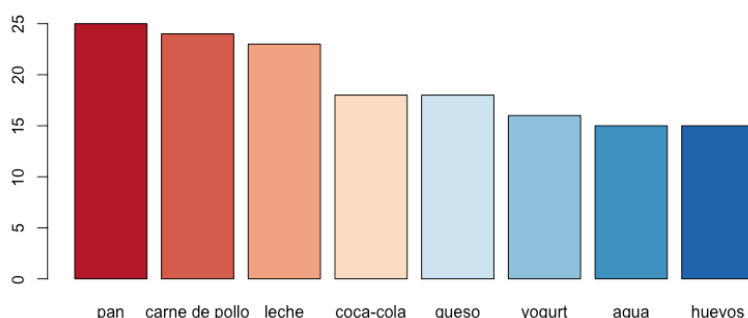
**Tabla 3.13: distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable "Edad=18-30 años"**

Edad=18-30 años	
Producto	Frecuencia
Pan	25
Carne de pollo	24
Leche	23
Coca-cola	18
Queso	18
Yogurt	16
Agua	15
Huevos	15

Fuente: elaboración propia con los resultados obtenidos en RStudio

Podemos ver que el ranking de productos obtenidos es muy parecido al del apartado anterior, y ya adelanto al lector que los resultados sucesivos también lo son. "Pan", "Carne de pollo" y "Leche" ocupan las primeras posiciones y "Coca-cola" o refresco de cola también continúa entre los primeros puestos.

**Gráfico 3.12: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable "Edad=18-30 años"**



Fuente: elaboración propia con los resultados obtenidos en R

Para el rango de edad entre 31-65 años vemos los resultados en la tabla que sigue (Tabla 3.14).

**Tabla 3.14: Distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable "Edad=31-65 años"**

Edad=31-65 años	
Producto	Frecuencia
Leche	24
Carne de pollo	18
Queso	18
Yogurt	15
Huevos	14
Pan	14
Detergente	13
Tomate frito	13

Fuente: elaboración propia con los resultados obtenidos en RStudio

Continúan los productos "Carne de pollo" y "Leche" entre los primeros puestos. Parece que el producto "Coca-cola" o refresco de cola no es un producto habitual en la cesta de la compra

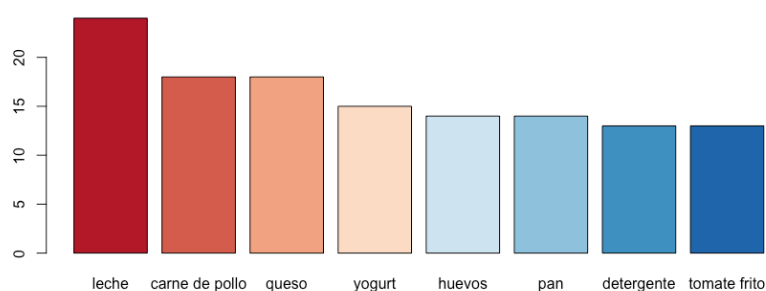
de este rango de edad, pero continúan estando presentes "Huevos" y "Pan". La aparición del producto "Detergente" como uno de los más frecuentes lo podemos asociar, con respecto al rango analizado anteriormente, a que al ser personas con mayor edad consideramos que estén emancipadas y por ello se deben ocupar de más labores del hogar. Las personas en la edad de entre los 18-30 años es muy probable que aún vivan con sus familiares y no les sea necesario realizar esta compra. Esta explicación propuesta se apoya en un informe del Observatorio de Emancipación del Consejo de la Juventud de España. En él se expone que en 2017 solo el 19,4% de los jóvenes españoles ha podido emanciparse en el año 2017. Registran en el estudio una caída en la emancipación en todos los rangos de edad, siendo algo superior de entre las personas de 30 a 34 años con un descenso del 4,88%. Puede el lector observar los datos del informe en la tabla 3.15.

**Tabla 3.15: Datos sobre población joven emancipada**

<b>POBLACIÓN JOVEN EMANCIPADA</b>				
	<b>Total</b>	<b>16-24 años</b>	<b>25-29 años</b>	<b>30-34 años</b>
<b>NÚMERO DE PERSONAS EMANCIPADAS</b>	1.257.620	238.340	1.019.280	2.092.727
<b>VARIACIÓN INTERANUAL</b>	-2,28%	-3,23%	-2,06%	-4,88%
<b>TASA DE EMANCIPACIÓN</b>	19,4%	5,9%	41,0%	72,3%

*Fuente: Observatorio de Emancipación del Consejo de la Juventud de España*

**Gráfico 3.13: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable "Edad=31-65 años"**



*Fuente: elaboración propia con los resultados obtenidos en R*

### 3.3.3.- ANÁLISIS DESCRIPTIVO DE LA CESTA DE LA COMPRA SEGÚN LA VARIABLE "ESTADO CIVIL"

En este caso tendremos en consideración los valores "Casado/a", "Soltero/a" y "Divorciado/a". No tendremos en cuenta el valor "Separado/a" puesto que al realizar el análisis no obtenemos compras frecuentes de alguno de los productos recogidos, todos con los que contamos en nuestra base de datos respecto a este valor solo han sido comprados una vez.

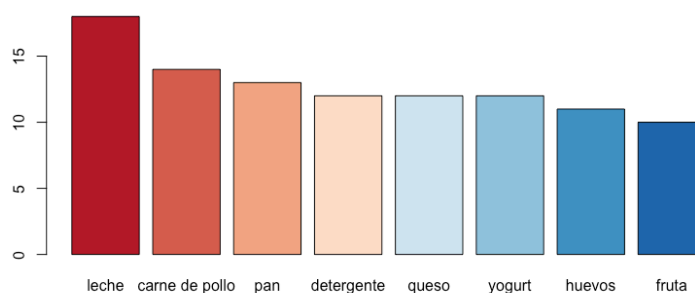
La cesta de la compra de las personas que manifiestan estar casados/as no dista mucho de lo mencionado ya en los casos anteriores, "leche", "carne de pollo" y "pan" entre los primeros puestos de frecuencia de compra al tratarse de productos de primera necesidad y aparece también el producto "detergente" en la cuarta posición. La razón por la cual consideramos que aparece es que las personas casadas normalmente viven ya fuera del hogar familiar y deben ocuparse de más tareas del hogar y necesitan realizar para ello la compra de ese producto.

**Tabla 3.16: Distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable "Estado Civil=Casado/a"**

Estado Civil=Casado/a	
Producto	Frecuencia
Leche	18
Carne de pollo	14
Pan	13
Detergente	12
Queso	12
Yogurt	12
Huevos	11
Fruta	10

Fuente: elaboración propia con los resultados obtenidos en RStudio

**Gráfico 3.14: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable "Estado Civil=Casado/a"**



Fuente: elaboración propia con los resultados obtenidos en R

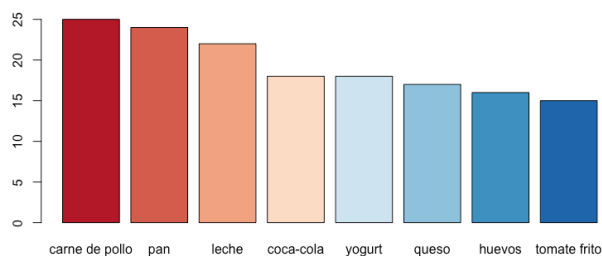
En las personas solteras también predominan los productos “carne de pollo”, “pan” y “leche” y aparece de nuevo el producto “Coca-cola” en cuarto lugar.

**Tabla 3.17: Distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable “Estado Civil=Soltero/a”**

Estado Civil=Soltero/a	
Producto	Frecuencia
Carne de pollo	25
Pan	24
Leche	22
Coca-cola	18
Yogurt	18
Queso	17
Huevos	16
Tomate frito	15

Fuente: elaboración propia con los resultados obtenidos en RStudio

**Gráfico 3.15: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable “Estado Civil=Soltero/a”**



Fuente: elaboración propia con los resultados obtenidos en R

Sin embargo, en las personas divorciadas nos encontramos que no siguen el mismo patrón que en los casos anteriores. Podemos observar que siguen estando presentes productos básicos como “leche” o “carne de pollo” y adquieren presencia productos de comida precocinada como “bacalao congelado” o canelones”.

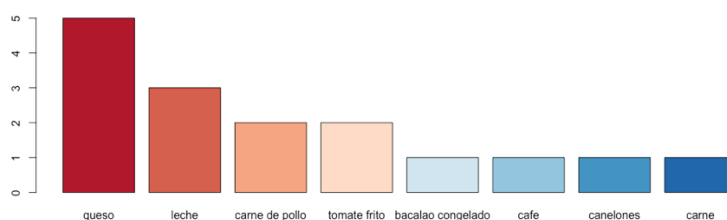


**Tabla 3.18: Distribución de frecuencias de productos ordenados de mayor a menor frecuencia variable "Estado Civil=Divorciado/a"**

Estado Civil=Divorciado/a	
Producto	Frecuencia
Queso	5
Leche	3
Carne de pollo	2
Tomate frito	2
Bacalao congelado	1
Café	1
Canelones	1
Carne	1

Fuente: elaboración propia con los resultados obtenidos en RStudio

**Gráfico 3.16: Gráfico de barras de productos ordenados de mayor a menor frecuencia variable "Estado Civil=Divorciado/a"**



Fuente: elaboración propia con los resultados obtenidos en R

Como conclusión de todo el análisis realizado podemos decir que en su mayoría los productos más frecuentemente comprados por los individuos encuestados son productos de primera necesidad como es "leche", "pan", "carne de pollo", "huevos", "queso", entre otros. Destacar la abundante presencia del producto "Coca-cola" en la cesta de la compra de los encuestados.

### **3.4.- ANÁLISIS DE LA CESTA DE LA COMPRA POR SEXO: REGLAS DE ASOCIACIÓN**

#### **3.4.1.- OBSERVACIONES PREVIAS**

Previamente a realizar el análisis de las reglas de asociación que puedan existir en nuestra muestra, repasaremos brevemente el apartado 2.3 donde se describen los pasos y los requisitos que se deben cumplir para poder aplicar este tipo de análisis.

En la aplicación del análisis de la cesta de la compra pretendemos obtener reglas de asociación entre productos comprados por los individuos que han contestado a nuestra encuesta. No contamos con ninguna idea o teoría preestablecida, sino que pretendemos explorar en las respuestas aportadas por los individuos encuestados y encontrar posibles patrones de compra. Además, todas las variables con las que contamos para realizar el análisis son de tipo categórica o cualitativa.

Los datos han sido recopilados por medios propios a través de una encuesta online. A través de ella hemos obtenido datos sobre las compras realizadas por los individuos los 15 días previos a la realización de la encuesta. Esta es por tanto nuestra transacción.

Contamos con 126 respuestas válidas. No es un tamaño demasiado grande, pero lo consideramos suficiente para el estudio que pretendemos realizar. Además, la tasa de "no respuesta" entre las respuestas válidas es nula.

Por estas razones consideramos idóneo la aplicación del análisis de la cesta de la compra a la muestra obtenida al cumplir con los requisitos necesarios anteriormente descritos.

### **3.4.2.- APLICACIÓN DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN: REGLAS DE ASOCIACIÓN POSITIVAS**

Antes de iniciar el análisis, hemos de considerar sobre qué variable cualitativa base de datos queremos basar el análisis. En nuestro caso vamos a aplicarlo en una base de datos considerada en función de la variable "Sexo". Tal y como hemos mencionado en anteriores apartados, para este estudio solo vamos a considerar al sexo "mujer" puesto que del sexo "hombre" contamos con muy poca representación en nuestra muestra y por tanto no obtendremos unos resultados fiables que pudiésemos extrapolar a la población masculina. En cambio, el 81% de la muestra está constituida por mujeres, lo que sí nos permitirá obtener un resultado más fiable.

Una vez filtrada y construida la base de datos sobre la que aplicar el análisis procedemos a valorar los productos que van a condicionar la regla de asociación. Una posibilidad puede ser elegir alguno de los productos más frecuentemente comprados por los individuos encuestado

que hemos analizado en el apartado anterior como puede ser el pan y la leche. Decidimos elegir este tipo de productos porque pensamos que al ser un producto que se compra con gran frecuencia es más fácil encontrar un patrón de compra que se repita y que dicho patrón tenga una fiabilidad más alta. Además, pensamos que va a resumir muy bien nuestra muestra. Basándonos en los resultados obtenidos en el apartado 3.3, los productos más frecuentemente comprados fueron leche, pan y carne de pollo. Entonces vamos a definir la regla de asociación: "Si compran leche y pan, también compran..." que expresado de forma teórica sería: {Leche, Pan} -> {...}

Los resultados de las posibles asociaciones los obtendremos mediante un algoritmo diseñado en RStudio que nos realizará los cálculos necesarios y que el lector podrá consultar en el anexo 3. Al introducir los productos leche y pan en la variable vector "PRODUCTO\_LHS" obtenemos el siguiente resultado:

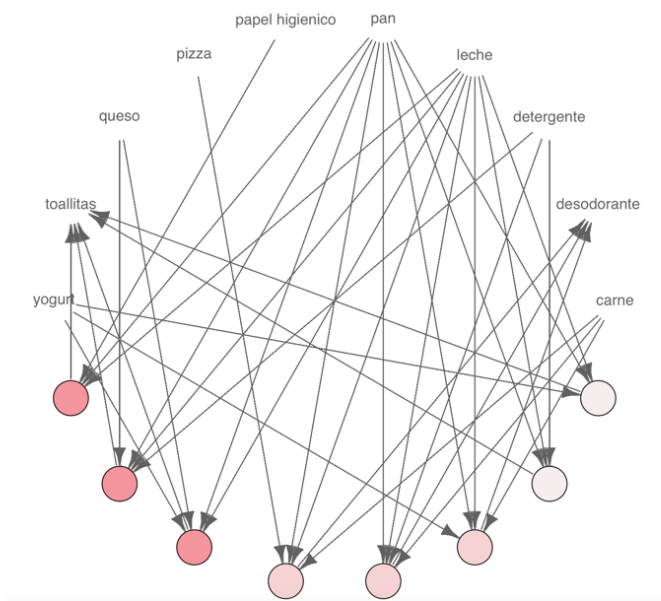
**Tabla 3.19: Resultados de la regla de asociación positiva {Leche, Pan} -> {...}**

	<b>LHS</b>	<b>RHS</b>	<b>SOPORTE</b>	<b>CONFIANZA</b>	<b>LIFT</b>
<b>1</b>	{leche, pan, papel higienico}	{toallitas}	0,01960784	1,0000000	25,5
<b>2</b>	{detergente, leche, pan, queso}	{toallitas}	0,01960784	1,0000000	25,5
<b>3</b>	{leche, pan, queso, yogurt}	{toallitas}	0,01960784	1,0000000	25,5
<b>4</b>	{carne, leche, pan, pizza}	{desodorante}	0,01960784	1,0000000	20,4
<b>5</b>	{carne, detergente, leche, pan}	{desodorante}	0,01960784	1,0000000	20,4
<b>6</b>	{carne, leche, pan, yogurt}	{desodorante}	0,01960784	1,0000000	20,4
<b>7</b>	{detergente, leche, pan}	{toallitas}	0,01960784	0,6666667	17,0
<b>8</b>	{leche, pan, yogurt}	{toallitas}	0,01960784	0,6666667	17,0

*Fuente: elaboración propia con los resultados obtenidos en R*

Se pueden observar dos resultados muy fuertes para la regla que hemos propuesto. El individuo que compra "pan" y "leche" es muy probable que compre "toallitas" o bien que compre "desodorante". Si nos centramos en los indicadores (explicados en el apartado 2.2) los valores resultantes del lift son 25.5 para "toallitas" y 20,4 para "desodorante", lo que nos indica que realmente existe una asociación entre los productos que componen la regla. Es cierto que el lift que corresponde a "toallitas" es 5,4 puntos más alto que el que corresponde al resultado "desodorante" pero ambos son resultados elevados a considerar a la hora de tomar decisiones sobre estos productos.

**Gráfico 3.17: Gráfico asociaciones positivas de {Leche, Pan} -> {...}**<sup>1</sup>



*Fuente: elaboración propia con los resultados obtenidos en R*

El valor del soporte en esta asociación es aceptable teniendo en cuenta que la base de datos es pequeña pero es suficiente para asegurar que este comportamiento se repetirá en futuras compras. Además, el valor de la confianza es el mejor que podríamos obtener siendo una regla con un 100% de precisión de manera que existe una probabilidad muy alta de que la regla propuesta se cumpla para una nueva transacción.

En el proceso de obtención del patrón anterior, se puede detectar otra regla de asociación a considerar. Los resultados pueden verse en la tabla 3.20 y es: {Leche, Pan, Papel Higiénico} -> {toallitas}. Este patrón presenta un lift alto de 25,5 puntos con lo cual no es una asociación producto del azar, sino que existe una asociación real entre este grupo de productos. El soporte de esta asociación es igual que el anterior, de aproximadamente un 2% y una confianza del 100% de precisión. De manera que hay una probabilidad muy alta de que la regla propuesta se cumpla para una nueva transacción.

<sup>1</sup> Interpretación del gráfico de asociaciones: las flechas unen a los productos entre los que existe una asociación. Cada círculo representa una asociación; las flechas que entran en el círculo son los LHS y las que salen son las RHS. El tamaño del círculo representa la cantidad de asociaciones de ese tipo. El color del círculo será más intenso o más claro si la asociación entre los productos es más fuerte o más débil respectivamente.

**Tabla 3.20: Resultados de la regla de asociación positiva {Leche, Pan, Papel Higiénico} -> {toallitas}**

	<b>LHS</b>	<b>RHS</b>	<b>SOPORTE</b>	<b>CONFIANZA</b>	<b>LIFT</b>
<b>1</b>	{leche, pan, papel higienico}	{toallitas}	0,01960784	1,0000000	25,5
<b>2</b>	{leche, pan, papel higienico}	{detergente}	0,01960784	1,0000000	6,8
<b>3</b>	{leche, pan, papel higienico}	{yogurt}	0,01960784	1,0000000	6,0
<b>4</b>	{leche, pan, papel higienico}	{queso}	0,01960784	1,0000000	3,4

*Fuente: elaboración propia con los resultados obtenidos en R*

Otra regla de asociación que proponemos, basándonos en los productos básicos que más manifiestan haber comprado nuestros encuestados, es {Leche, Carne de pollo} -> {...}. En la siguiente tabla podemos visualizar los resultados obtenidos asignando estos productos a la variable vector "PRODUCTO\_LHS".

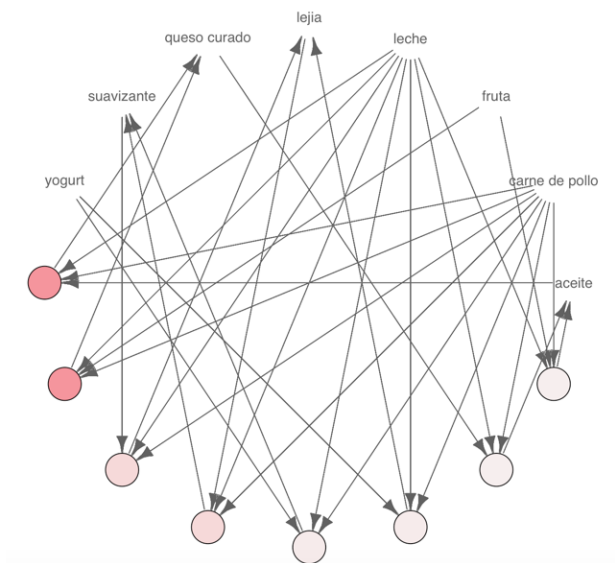
**Tabla 3.21: Resultados de la regla de asociación positiva {Leche, Carne de pollo} -> {...}**

	<b>LHS</b>	<b>RHS</b>	<b>SOPORTE</b>	<b>CONFIANZA</b>	<b>LIFT</b>
<b>1</b>	{aceite, carne de pollo, leche}	{queso curado}	0,01960784	1,0000000	25,5
<b>2</b>	{carne de pollo, fruta, leche}	{queso curado}	0,01960784	1,0000000	25,5
<b>3</b>	{carne de pollo, leche, suavizante}	{lejía}	0,01960784	1,0000000	14,57
<b>4</b>	{carne de pollo, leche, lejía}	{suavizante}	0,01960784	1,0000000	14,57
<b>5</b>	{carne de pollo, leche, yogurt}	{suavizante}	0,01960784	0,6666667	9,71
<b>6</b>	{carne de pollo, leche, yogurt}	{lejía}	0,01960784	0,6666667	9,71
<b>7</b>	{carne de pollo, leche, queso curado}	{aceite}	0,01960784	1,0000000	9,27
<b>8</b>	{carne de pollo, fruta, leche}	{aceite}	0,01960784	1,0000000	9,27

*Fuente: elaboración propia con los resultados obtenidos en R*

Se descubre una asociación entre los productos "leche" y "carne de pollo" con el producto "queso curado". Esta asociación cuenta con un lift de 25,5 puntos que nos indica la existencia de una asociación entre los productos. Además "leche", "carne de pollo" con el producto "lejía" son productos que también se encuentran asociados. En este caso cuenta con un lift de 14,57 puntos. Ambas asociaciones tienen una confianza del 100% de probabilidad de que se cumpla en nuevas transacciones.

**Gráfico 3.18: Gráfico asociaciones positivas de {Leche, Carne de pollo} -> {...}**



Fuente: elaboración propia con los resultados obtenidos en R

Estos patrones de compra que hemos propuesto han sido fruto de condicionar el lado izquierdo de la regla de asociación (LHS). Recordar al lector que la estructura de una regla de asociación se compone de dos partes: el lado izquierdo o antecedente (LHS) y el lado derecho o consecuente (RHS). Vamos a proceder ahora a condicionar el lado derecho de la regla de asociación (RHS) y observar si puede existir alguna combinación de productos de la que resulte la compra del producto "Coca-cola". En este caso vamos a establecer el producto "Coca-cola" como factor consecuente de la regla de asociación. Elegimos este por los resultados tan llamativos que hemos obtenido en el apartado 3.3.

**Tabla 3.22: Resultados de la regla de asociación {..., ...} -> {coca-cola}**

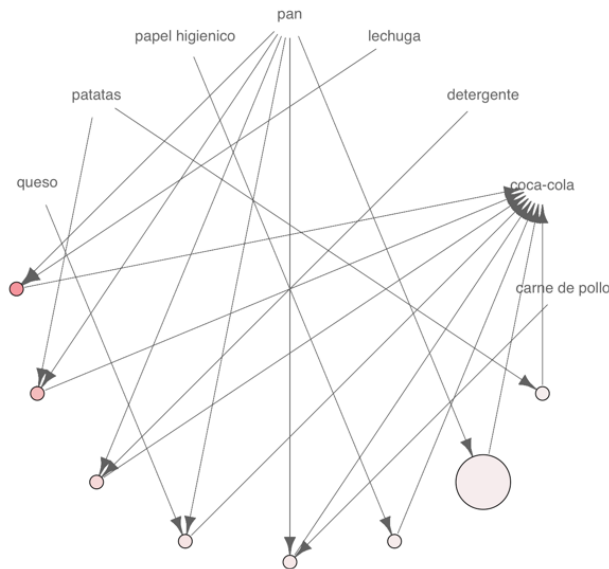
	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{lechuga, pan}	{coca-cola}	0,01960784	0,6666667	11,3
2	{pan, patatas}	{coca-cola}	0,01960784	0,5000000	8,5
3	{detergente, pan}	{coca-cola}	0,01960784	0,3333333	5,67
4	{pan, queso}	{coca-cola}	0,01960784	0,2500000	4,25
5	{carne de pollo, pan}	{coca-cola}	0,01960784	0,2222222	3,78
6	{papel higienico}	{coca-cola}	0,01960784	0,2000000	3,40
7	{pan}	{coca-cola}	0,01960784	0,1923077	3,27
8	{patatas}	{coca-cola}	0,01960784	0,1818182	3,09

Fuente: elaboración propia con los resultados obtenidos en R

Como resultado hemos obtenido un patrón que relaciona "lechuga" y "pan" con "coca-cola". Esta asociación presenta un lift de 11,3 puntos. No es un valor tan alto como en el caso

anterior, pero representa en todo caso que existe una asociación entre estos productos. La confianza de esta regla de asociación es del 66,6%, solo en el 66% de las nuevas transacciones que aparezca "lechuga" y "pan" aparecerá también "coca-cola".

**Gráfico 3.19: Gráfico asociaciones positivas de {...} -> {coca-cola}**



*Fuente: elaboración propia con los resultados obtenidos en R*

Las reglas de asociación que hemos expuesto se tratan de reglas de asociación positivas, es decir, la presencia de uno o un grupo de productos implica la presencia de otro producto. Esto es así porque el lift de todas las asociaciones es superior a uno. Que el lift de una regla de asociación sea mayor o menor con respecto a otra va a depender de la fuerza de la asociación. Cuando mayor sea el valor del lift mayor será la fuerza de la asociación entre los productos. Cabe observar que el resultado de los indicadores de las reglas que hemos seleccionado tiene un valor del lift y la confianza bastante alto pero el soporte es bajo. Esto se debe que los productos no tienen una alta tasa de aparición en el conjunto de datos, pero realmente sí están asociados entre sí. Las reglas que presentan los indicadores de esta forma son las que realmente son interesantes de obtener porque son impredecibles o difíciles de identificar en otro tipo de investigaciones.

En suma, los productos que componen este tipo de reglas de asociación se dice que son productos complementarios entre sí por el hecho de que la compra de un producto implica la compra del otro producto. Por este motivo es habitual que se apliquen sobre ellos técnicas comerciales que permitan incrementar las ventas de los productos implicados y aumentar los beneficios del supermercado.

### **3.4.3.- APLICACIÓN DEL ANÁLISIS DE REGLAS DE ASOCIACIÓN: REGLAS DE ASOCIACIÓN NEGATIVAS**

Hasta ahora hemos analizado asociaciones positivas entre productos, es decir, la presencia de uno o un grupo de productos implica la presencia de otro producto o grupo de productos. Sin embargo, también puede darse la situación opuesta, tal y como hemos expuesto ya al lector en este estudio. Puede ocurrir que la presencia de uno o un grupo de productos implica la ausencia de otro producto o grupo de productos. Este aspecto también es esencial conocerlo a la hora de tomar decisiones comerciales sobre dichos productos. Si anteriormente hablamos de productos complementarios entre sí al presentarse en ellos una asociación positiva, en este caso vamos a hablar de productos que son sustitutivos entre sí ya que al comprarse "unos" no se van a comprar "otros" exponiendo pues un efecto negativo (por ejemplo, un cliente que compra Pepsi no va a comprar Coca-Cola). Cuando se da este tipo de asociación negativa, lo normal es que no se apliquen técnicas comerciales sobre estos productos de forma conjunta porque no se va a conseguir incrementar sus volúmenes de venta al ser estos excluyentes entre sí.

Dada su importancia, queremos detectar también si existe alguna asociación negativa en la base de datos que estamos trabajando. Para detectar este tipo de reglas el indicador determinante para ello va a ser el lift. Si este es inferior a uno implica una asociación negativa entre productos (véase apartado 2.2). En estos casos los valores tanto del soporte como de la confianza van a ser mucho menores que en los cálculos anteriores en RStudio porque van a ser patrones de comportamiento con poca probabilidad de que se cumplan en futuras compras<sup>2</sup>.

Volvemos a considerar la asociación {Leche, Pan} -> {...} y comprobaremos si, mediante los cálculos con RStudio, existe alguna relación negativa con estos productos. Incluiremos una pequeña modificación en el algoritmo para poder realizar este análisis. En vez de ordenar el lift de las transacciones de forma descendente como hemos hecho en el apartado anterior, lo ordenaremos de forma ascendente. Con todo ello, obtenemos los siguientes resultados:

---

<sup>2</sup> Se ha de tener en cuenta que el algoritmo que hemos diseñado en RStudio no diferencia entre reglas positivas o negativas. RStudio nos proporciona solo, en este caso, reglas positivas.



**Tabla 3.23: Resultados de la regla de asociación negativa {Leche, Pan} -> {...}**

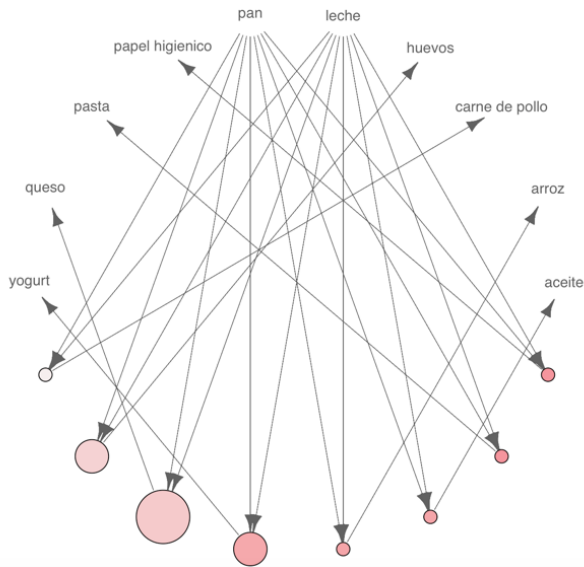
	LHS	RHS	SOPORTE	CONFIANZA	LIFT
<b>1</b>	{leche, pan}	{carne de pollo}	0,01960784	0.1666667	0,52
<b>2</b>	{leche, pan}	{huevos}	0.02941176	0.2500000	1,02
<b>3</b>	{leche, pan}	{queso}	0.03921569	0,3333333	1,13
<b>4</b>	{leche, pan}	{yogurt}	0.02941176	0,2500000	1,50
<b>5</b>	{leche, pan}	{arroz}	0,01960784	0.1666667	1,55
<b>6</b>	{leche, pan}	{aceite}	0,01960784	0.1666667	1,55

Fuente: elaboración propia con los resultados obtenidos en R

En la transacción 1 podemos observar una regla negativa con los productos del lado izquierdo de la condición: {Leche, Pan}  $\neq$  {Carne de pollo}. El lift de este patrón es inferior a 1 (0,52 puntos) lo cual nos indica que la relación que existe entre estos productos es negativa, es decir, el individuo que compra de leche y pan no va a comprar carne de pollo. El valor del soporte es bajo (2%) al igual que el valor de la confianza que se sitúa en un 16,7%. Esto es debido a la relación negativa que existe entre los productos. Al ser un patrón de comportamiento que no se va a repetir nunca o casi nunca en el futuro es lógico obtener un porcentaje tan bajo comparado con los casos anteriores de relaciones positivas.

Resulta destacable la relación que ha surgido entre estos tres productos, *i.e.*, "leche", "pan" y "carne de pollo". Aun siendo los productos más comprados por los encuestados conforman al parecer dos grupos totalmente excluyentes entre sí. Aprovechamos esta situación para destacar la gran importancia de que el tomador de decisiones dentro del supermercado deba tener conocimiento sobre estos patrones de comportamiento. Lo que cabría pensar es que los productos más frecuentemente comprados pudieran conformar un grupo completo sobre el cual podríamos aplicar acciones comerciales e incrementar las ventas de los productos. Este hecho nos indica todo lo contrario, se deben aplicar acciones comerciales sobre "leche" y "pan" y sobre "carne de pollo" de forma independiente al no existir una relación directa.

**Gráfico 3.20: Gráfico asociaciones negativas de {Leche, Pan} -> {...}**



Fuente: elaboración propia con los resultados obtenidos en R

Ante esta situación que se nos ha presentado vamos a proceder a analizar si existe alguna relación negativa con el producto carne de pollo siguiendo el modelo que hemos planteado para el caso anterior y comprobar si realmente esta relación negativa existe sobre estos productos. Recordar al lector que en el apartado anterior hemos considerado "leche" y "carne de pollo" como los antecedentes de una regla de asociación. De demostrarse realmente que esta relación es excluyente, la regla expuesta anteriormente no será válida.

**Tabla 3.24: Resultados de la regla de asociación negativa {Carne de pollo} -> {...}**

	LHS	RHS	SOPORTE	CONFIANZA	LIFT
1	{carne de pollo}	{patatas}	0,01960784	0.06060606	0.56
2	{carne de pollo}	{pescado}	0.02941176	0.09090909	0,76
3	{carne de pollo}	{leche}	0.09803922	0,30303030	0,84
4	{carne de pollo}	{picos}	0,01960784	0.06060606	0,88
5	{carne de pollo}	{atún}	0,03921569	0.12121212	0,88
6	{carne de pollo, fruta}	{leche}	0,01960784	0.33333333	0,92
7	{carne de pollo, huevos}	{leche}	0.02941176	0.33333333	0,92
8	{carne de pollo}	{papel higienico}	0.02941176	0.09090909	0,93
9	{carne de pollo, queso}	{leche}	0,03921569	0,36363636	1,00

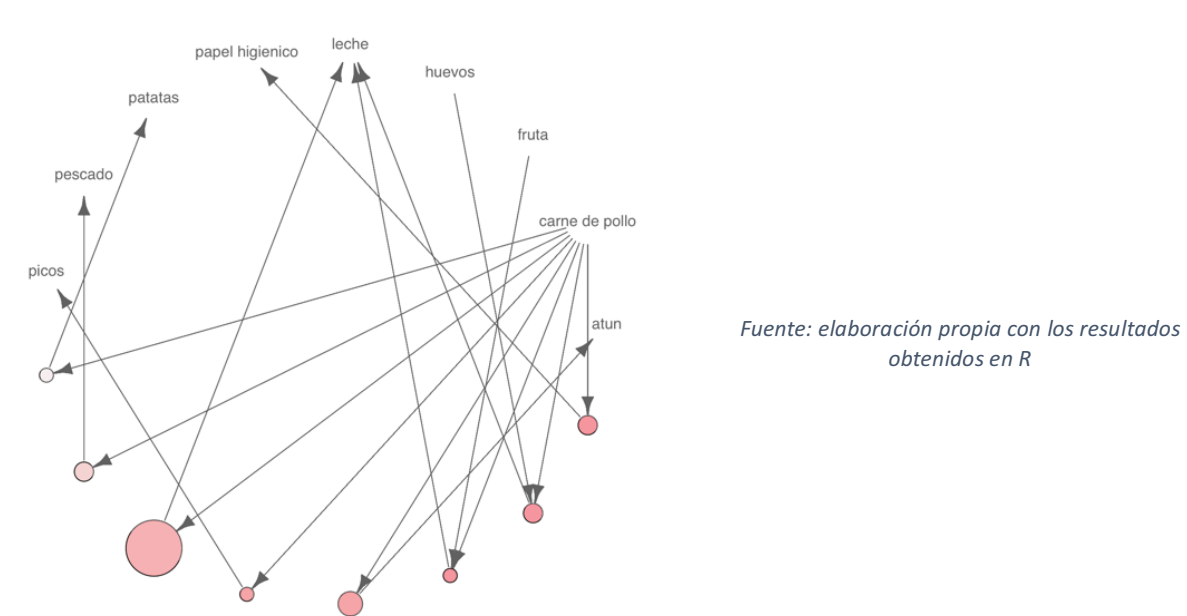
Fuente: elaboración propia con los resultados obtenidos en R

En los resultados podemos observar que no está muy clara la relación negativa entre los productos "leche" y "carne de pollo" (transacción 3) al tener un valor del lift próximo a 1.

Incluso podemos observar que si el individuo compra carne de pollo y otro producto más diferente a "pan" la relación entre "carne de pollo" y "leche" sigue siendo próximo a 1 con una confianza entorno a un 9-36%.

Sí podemos afirmar con mayor certeza la relación negativa que existe entre "carne de pollo" y "patatas". Vemos que el lift en este caso se encuentra en 0,56 puntos y con una confianza de tan solo un 6% de probabilidad de comprar el producto "patatas" cuando se compra "carne de pollo". A estos productos se recomendará que no se apliquen técnicas comerciales de forma conjunta al tratarse de productos que presentan una relación excluyente y por ello no se podrán incrementar sus ventas conjuntamente.

**Gráfico 3.21: Gráfico asociaciones negativas de {Carne de pollo} -> {...}**



Comprobamos entonces si existe alguna posible asociación negativa entre "pan" y "carne de pollo". En este caso exigimos al algoritmo diseñado en RStudio que la variable vector PRODUCTO\_LHS sea "pan" y que la variable vector PRODUCTO\_RHS sea "carne de pollo" para así obtener los resultados de solo esas asociaciones.

**Tabla 3.25: Resultados de la regla de asociación negativa {Pan} -> {Carne de pollo}**

	<b>LHS</b>	<b>RHS</b>	<b>SOPORTE</b>	<b>CONFIANZA</b>	<b>LIFT</b>
<b>1</b>	{pan}	{carne de pollo}	0,08823529	0.3461538	1,07
<b>2</b>	{coca-cola, pan}	{carne de pollo}	0,01960784	0.4000000	1,24
<b>3</b>	{pan, pizza}	{carne de pollo}	0,01960784	0.4000000	1,24
<b>4</b>	{pan, yogurt}	{carne de pollo}	0,01960784	0.4000000	1,24
<b>5</b>	{jamon york, pan}	{carne de pollo}	0,01960784	0.5000000	1,55
<b>6</b>	{fruta, pan}	{carne de pollo}	0.02941176	0.5000000	1,55
<b>7</b>	{huevos, pan}	{carne de pollo}	0.02941176	0.5000000	1,55

*Fuente: elaboración propia con los resultados obtenidos en R*

Como podemos observar en los resultados los productos "pan" y "carne de pollo" no guardan ninguna asociación entre ellos porque el lift de las asociaciones obtenidas están muy próximos a 1. Recordar al lector que, una asociación con un lift igual o próximo a uno nos indica que no existe ninguna relación entre los productos y que su presencia conjunta se debe al puro azar. Además, contamos con una confianza elevada entre el 30 y el 50%.

Con todo este estudio realizado podemos concluir que los productos "leche", "pan" y "carne de pollo" no guardan ninguna asociación entre ellos, ni positiva ni negativa. Pero la combinación entre ellos sí da como resultados asociaciones positivas fuertes con otro tipo de productos como hemos visto en el apartado 3.4.2. Es un tanto lógico que se obtenga este resultado puesto que los tres productos son están dentro del grupo de productos de primera necesidad.

## 4.- CONCLUSIONES

Las reglas de asociación son un enfoque metodológico que se originó en el campo del marketing y que actualmente se está extendiendo su utilización en diferentes disciplinas científicas. Existen dos razones por las cuales se está aumentando el uso de esta técnica. Una de ellas es la sencillez de su aplicación sobre un conjunto de datos en comparación con otro tipo de análisis. No es necesario que se cumplan requisitos de normalidad o linealidad, ni siquiera que los datos estén ordenados. La otra razón es el gran valor de los resultados obtenidos a través de esta metodología.

Técnicas como esta que hemos aplicado en este estudio pueden proporcionar a la empresa información muy relevante sobre el comportamiento de sus clientes a la hora de realizar la compra en un supermercado. Por ello, podemos englobar a las reglas de asociación como un tipo de análisis dentro del Big Data que nos permite realizar un análisis predictivo de nuestros clientes sobre sus compras. Lo que realmente conseguimos con esta técnica es conocer sus patrones de comportamiento de manera que se identifiquen combinaciones de productos relacionados positivamente y así incentivar su compra con el uso de técnicas comerciales y aumentar los beneficios de las empresas. De igual forma podemos encontrarnos con combinaciones de productos que guarden entre ellos una relación negativa. Si en el caso anterior la presencia de un producto implica la presencia de otro, en este caso la presencia de un producto implica la ausencia de otro.

La finalidad con la que la empresa puede utilizar estos datos es muy diversa, pero resulta muy útil para implementar estrategias de venta cruzada, realizar recomendaciones a los clientes sobre productos relacionados, o bien promocionar estos colocándolos muy cerca unos de otros en páginas web, catálogos o estanterías de supermercado.

En la base de datos del caso práctico que hemos llevado a cabo en este estudio hemos encontrado patrones de comportamiento que son de interés. Se ha observado en el análisis previo de la muestra que tenemos una importante presencia de la mujer. Por ello el análisis de la cesta de la compra que se ha centrado en el sexo femenino. Además, en la cesta de la compra de las mujeres encuestadas lo que más predomina son productos básicos como puede ser "leche", "carne de pollo", "pan", "queso" o "huevos". Destacar la presencia del producto "Coca-cola" entre los productos más frecuentemente comprados. Desconocemos si el producto es realmente original de la marca Coca-cola o se trata de un refresco de cola de una marca

de distribuidor, pero al menos podemos afirmar que este producto está muy presente en la cesta de la compra de las mujeres.

Nos hemos centrado en estos productos para encontrar posibles patrones de comportamiento porque son los más comprados. Además, creemos que la fiabilidad de los indicadores de la asociación será mayor. Se ha observado una asociación positiva importante entre los productos "leche", "pan" y "toallitas" y "leche", "pan" y "desodorante", dos reglas de asociación que cuentan con un lift bastante alto y que por tanto nos permite concluir que los dos conjuntos de productos mantienen una relación positiva entre ellos. También contamos con una confianza del 100% de probabilidad de que de que se elija el producto del lado derecho cuando se elige el producto del lado izquierdo para ambas asociaciones.

Otra relación positiva encontrada en el caso práctico de la que resulta el producto "toallitas" es la combinación "leche", "pan" y "papel higiénico". El lift de esta asociación nos indica que también guardan una relación fuerte y con una confianza del 100%. Aparece en este caso una asociación de productos de higiene personal.

Lo mismo ocurre con la asociación "leche", "carne de pollo" y "suavizante" con el producto "lejía". También guardan una relación fuerte y podemos relacionarlo con la limpieza o higiene del hogar. Por otro lado, "leche" y "carne de pollo" también guardan una relación importante con el producto "queso curado".

Dado que nos llamó mucho la atención la presencia del producto "Coca-cola" en nuestro análisis de la cesta de la compra, quisimos averiguar si existía alguna relación de interés con este producto. En el análisis detectamos una asociación positiva de "pan" y "lechuga" con el producto "Coca-cola" al igual que los productos "pan" y "patatas". Ambas asociaciones cuentan con un lift elevado y la confianza se aproxima a un 60%.

Al igual que consideramos relaciones positivas entre los productos de nuestra base de datos, también queríamos detectar si existía alguna relación negativa a considerar. A través de ello detectamos que los productos "leche", "pan" y "carne de pollo" son productos independientes, no guardan ningún tipo de asociación entre ellos. Si lo pensamos detenidamente es lógico este resultado puesto que estos tres productos son de primera necesidad.

Solo hemos observado una regla de asociación negativa entre los productos “carne de pollo” y “patatas”. El lift que relaciona a estos productos es inferior a uno y cuenta con una confianza del 6%.

Llegados a este punto consideramos necesario exponer las limitaciones de este estudio. Hemos realizado una exposición muy general sobre los productos que están relacionados entre sí descubriendo los patrones de compra existentes en los individuos encuestados. Bien es cierto que la muestra considerada no es demasiado grande y además no lo hemos focalizado en un único supermercado. Sólo hemos considerado la cesta de la compra de los últimos 15 días del individuo sin distinguir el establecimiento en el que ha realizado la compra. Quizás el estudio sería de mayor utilidad si se hubiese centrado en la cesta de la compra de un solo supermercado. También destacar que solo nos hemos centrado en el segmento de las mujeres. Nos parecía interesante porque contábamos con una gran representatividad en la muestra y por ello los resultados obtenidos sí podrían ser extrapolados a la población con mayor fiabilidad. Por las razones anteriores nos gustaría finalizar este apartado proponiendo algunas líneas de investigación futuras:

- Se podría aplicar el mismo estudio focalizado en los hombres. En nuestro caso no hemos podido realizarlo por su poca representatividad en la muestra. Este trabajo sería bastante interesante puesto que hay estudios que avalan que el hombre cada vez más está teniendo un papel más significativo en la compra familiar. Por ello es necesario conocer cómo se comportan a la hora de realizar la compra porque actualmente son los grandes desconocidos para las empresas de alimentación.
- También consideramos interesante realizar este estudio recogiendo la edad exacta del individuo. Sólo con observar el análisis descriptivo de la cesta de la compra en el apartado 3.3 podemos deducir que las reglas de asociación que se obtendrán serán muy diferentes a las obtenidas en el presente estudio.
- La mayor utilidad que le apreciamos a este estudio es focalizarlo en un supermercado en concreto. Hay que tener en cuenta que cada supermercado tiene un perfil de cliente diferente y una variedad de productos diferentes. Por ello se podrían descubrir oportunidades de negocio que podrían aumentar sus ventas. Además, si se estudiasen varios por separado, se podría exponer una comparativa sobre la cesta de la compra de sus clientes.

- Podemos utilizar este estudio no solo para conocer qué compra el individuo que acude al supermercado, también es posible analizar y conocer a los clientes que acuden. Proponemos crear diferentes perfiles de clientes asociados a las categorías de producto. Por ejemplo, los que compran productos de las categorías "fruta", "carne" y "dietética deportiva" podemos establecer que son personas que practican ejercicio físico. Otro ejemplo podría ser, los clientes que compran "platos preparados", "pizzas" y "productos congelados" siguen un perfil de hombres solteros, de 34 años y que viven solos. De esta forma el tomador de decisiones conocería mucho mejor qué público es el que acude a su establecimiento. En estos casos se podría plantear una vez conocido el perfil del cliente que compra determinados productos adaptar mejor el *packaging* del producto o realizar una comunicación o promoción más personalizada.



## **BIBLIOGRAFÍA**

- Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Acm sigmod record* (Vol. 22, No. 2, pp. 207-216). ACM.
- Aguinis, H., Forcum, L. and Joo, H. (2013). Using Market Basket Analysis in Management Research. *Journal of Management*, 39(7), pp.1799-1824.
- Chapman, C. & Feit, E.. (2015). Association Rules for Market Basket Analysis. En *R for marketing research and analytics* (339-360). USA: Springer.
- Chen, C., Lan, G., Hong, T., & Lin, Y. (2013). Mining high coherent association rules with consideration of support measure. *Expert Systems with Applications*, 40(16), 6531-6537.
- Chen, M., Chao, C., & Wu, K. (2012). Pattern filtering and classification for market basket analysis with profit-based measures. *Expert Systems*, 29(2), 170-182.
- Consejo de la Juventud de España (2017) Observatorio de emancipación- 1er Semestre (2).
- EAST, R.; HAMMOND, K.; HARRIS, P. Y LOMAX, W. (2000). «First-Store Loyalty and Retention», *Journal of Marketing Management*, 16 (4), 307–325.
- EAST, R.; HARRIS, P.; WILLSON, G. Y LOMAX, W. (1995). «Loyalty to Supermarkets», *The International Review of Retail, Distribution and Consumer Research*, 5 (1), 99–109.
- Giudici, P. & Figini, S. (2009). Market basket analysis. En *Applied data mining for business and industry* (175-191). Chichester, U.K: Wiley.
- Hoyer, W. D., MacInnis, D. J., & Pieters, R. (2010). *Comportamiento del consumidor*. México: Cengage Learning.
- Kantar Wordpanel. (2018). A global ranking of the most chosen consumer brands (6) Recuperado de <https://www.kantarworldpanel.com/brand-footprint-ranking/#/download>
- Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, 85 (International Conference on Computational Modelling and Security, CMS 2016), 78-85.
- Larose, D. & Larose, C. (2014). Association Rules. En *Discovering Knowledge in Data: An Introduction to Data Mining* (247-264). Hoboken: Wiley.

Linoff, G. & Berry, M. (2011). Market Basket Analysis and Association Rules. En Data mining techniques: for marketing, sales, and customer relationship management (287-319). Indianapolis, Ind: Wiley.

Mägi, A. W. (2003). «Share of Wallet in Retailing: The Effects of Customer Satisfaction, Loyalty Cards and Shopper Characteristics», Journal of Retailing, 79 (2), 97-106.

Maresh, J. M. L. (2012). Association models for market basket analysis, customer behaviour analysis and business intelligence solution embedded with apriori concept. International Journal of Research in Finance and Marketing, 2(1), 40-53.

Manero, C. B., Caraballo, N. M., Valero, P. G., & Figueras, M. S. (2008). Análisis del comportamiento variado de compra de los hogares españoles: una aproximación bayesiana. Estadística española, 50(169), 479-521.

Olson, D. (2017). Market Basket Analysis. En Descriptive data mining (29-41). Singapore: Springer Nature.

Rai, N., Susheel, J. & Anurag, J. (2013, Septiembre) "Mining positive and negative association rule from frequent and infrequent pattern based on imlms\_ga." International Journal of Computer Applications 77, pp.48-52.

Raorane AA, Kulkarni RV & Jitkar BD. Association Rule – Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences 2012:1(2): pp.19-27.

Rocha Hernández, J.A., Rodríguez Herrera, D.M. & Rodríguez Rodríguez, J.E. (2016, Abril) A research comparative among association rules algorithms. Visión electrónica, 2.

Ruiz Jimenez, M.D. (2010). Modelado formal para representación y evaluación de reglas de asociación (tesis doctoral). E.T.S. de Ingenierías informática y de telecomunicación. Universidad de Granada, Granada, España.

Santana Sepúlveda, J.S., & Mateos Farfán, E., (2014) El arte de programar en R: un lenguaje para la estadística. Progreso, Jiutepec, Morelos, México. Instituto Mexicano de Tecnología del Agua.

Schmitt, J. (2010). Drawing association rules between purchases and in-store behavior: An extension of the market basket analysis. Advances in Consumer Research, 37, 899.

Szymkowiak, M., Klimanek, T., & Józefowski, T. (2018). Applying market basket analysis to official statistical data. Econometrics / Ekonometria, 22(1), 39-57.

Yen-Liang, C., Tang, K., Ren-Jie, S., & Ya-Han, H. (2005). Market basket analysis in a multiple store environment. Decision Support Systems, 40(2), 339-354.

Zekić-Sušac M. & Has A.. (2015). Data Mining as Support to Knowledge Management in Marketing. Business Systems Research Journal, 6(2), pp.18-30.

Zekic-Susac, M., & Has, A. (2015). Discovering market basket patterns using hierarchical association rules. *Croatian Operational Research Review*, 6(2), 475-487.

Zhang, C. & Zhang, S. (2002). *Association rule mining: models and algorithms*. Berlin New York: Springer.

## **ANEXO I: ENCUESTA**

### **CESTA DE LA COMPRA**

El objetivo del siguiente formulario es la recopilación de datos para un trabajo de fin de grado sobre las últimas compras realizadas en el núcleo familiar. Pedimos la mayor precisión posible por su parte.

Muchas gracias.

Grado en Marketing e Investigación de Mercados en la Universidad de Sevilla.

**\*Obligatorio**

#### **Datos personales.**

##### **1. Sexo: \***

*Marca solo un óvalo.*

- Mujer
- Hombre

##### **2. Edad: \***

*Marca solo un óvalo.*

- Menos de 18 años
- 18-30 años
- 31-65 años
- Más de 65 años

##### **3. Nacionalidad: \***

*Marca solo un óvalo.*

- Española
- Otro: \_\_\_\_\_

##### **4. Estado civil: \***

*Marca solo un óvalo.*

- Casado/a
- Soltero/a
- Separado/a
- Divorciado/a
- Viudo/a

- Otro: \_\_\_\_\_

**5. Número de miembros de la unidad familiar \***

*Marca solo un óvalo.*

- 1 miembro
- 2 miembros
- 3 miembros
- 4 miembros
- 5 miembros o más

**6. Ocupación: \***

*Marca solo un óvalo.*

- Estudiante
- Con empleo
- Desempleado/a
- Jubilado/a

**7. Nivel de estudios: \***

*Marca solo un óvalo.*

- Sin estudios
- Primaria
- Secundaria
- Bachillerato
- Formación profesional
- Estudios universitarios

**8. Ámbito de residencia: \***

*Marca solo un óvalo.*

- Urbano (Ciudad, pueblo grande, etc)
- Rural (Aldea, pueblo pequeño, etc)

**9. Creencias religiosas: \***

*Marca solo un óvalo.*

- Católico/a
- Creyente de otra religión
- No creyente/agnóstico(a)
- Ateo
- Prefiero no responder

10. **¿En qué supermercados suele realizar normalmente sus compras?** \* *Selecciona todos los que correspondan.*

- Mercadona.
- Alcampo.
- Carrefour.
- Hipercor/supercor
- Lidl.
- Dia.
- Supersol.
- Otro: \_\_\_\_\_

11. **Escriba a continuación los productos que haya comprado en los últimos 15 días en el supermercado.** \*

No necesitamos que especifique la marca ni la cantidad que haya comprado, pero si el tipo de producto adquirido. Por ejemplo: pack de 6 cervezas sin alcohol, yogurt de vainilla, leche semidesnatada sin lactosa, pan sin gluten, queso en loncha o queso curado, pack de botellas o latas de coca-cola, pechuga o filete de pollo... Intente introducir al menos 7 productos.

---

---

---

---