



Facultad de Ciencias Económicas y Empresariales

ICADE

**Optimización de la Estrategia de Ventas de un  
Supermercado: Análisis de Cesta de la Compra**

Autor: Marta Echevarría Casasola

Directora: Anitha Srinivasan

MADRID | junio 2024

## **Resumen**

El presente trabajo de investigación se enfoca en la aplicación de técnicas avanzadas de análisis de datos para aumentar las ventas de una cadena de supermercados minoristas. En un contexto marcado por la era digital y la proliferación de sistemas de recopilación de información, la autora se propone resolver un caso de consultoría estratégica mediante el análisis de cesta de la compra, técnica que identifica asociaciones entre productos comprados conjuntamente. Tras una revisión exhaustiva de literatura académica e informes empresariales para entender el sector de los supermercados minoristas y los factores clave del éxito, se selecciona el análisis de cesta de la compra como técnica más adecuada para el estudio. Haciendo uso del algoritmo Apriori, y con apoyo de la herramienta R Studio, se revela la existencia de 11 reglas de asociación que permiten identificar combinaciones frecuentes de artículos comprados conjuntamente (e.g., relación entre la compra de plátanos y la compra de otros productos como la leche o el aguacate). Estas relaciones identificadas permiten al negocio optimizar sus inventarios, desarrollar estrategias de marketing dirigidas, crear promociones personalizadas y mejorar la distribución de los productos en el supermercado. Todo ello contribuyendo significativamente al objetivo final de incrementar las ventas de la empresa.

## **Palabras clave**

Análisis de cesta de la compra, reglas de asociación, algoritmo Apriori, supermercado minorista, estrategia de ventas.

## **Abstract**

This research work focuses on the application of advanced data analysis techniques to increase sales of a retail supermarket chain. In a context marked by the digital era and proliferation of information gathering systems, the author proposes to address a strategic consulting case through market basket analysis, a technique that identifies associations between jointly purchased products. After a detailed review of academic literature and business reports to understand the retail supermarket industry and its key success factors, market basket analysis is selected as the most appropriate technique for the study. Using Apriori algorithm, and supported by R Studio software, 11 association rules are discovered that allow the identification of frequent combinations of items purchased together (e.g., relationship between the purchase of bananas and the purchase of other products such as milk or avocado). These identified relationships allow businesses to optimize their inventories, develop targeted marketing strategies, create personalized promotions and improve product distribution in the supermarket. All significantly contributing to the final goal of boosting the sales of the company.

## **Key words**

Market basket analysis, association rules, Apriori algorithm, retail supermarket, sales strategy.

## Índice de contenidos

1.	Introducción.....	6
1.1	Contextualización del tema.....	6
1.2	Objetivo del trabajo.....	6
1.3	Estructura del trabajo.....	7
1.4	Metodología del trabajo.....	7
2.	Exposición del Caso de Negocio.....	8
3.	Objetivo del Caso de Negocio .....	9
3.1	Análisis del Mercado Minorista de Supermercados .....	9
3.2	Claves del Éxito en Supermercados .....	10
4.	Metodologías de Análisis de Datos.....	12
5.	Análisis de Cesta de la Compra .....	14
5.1	Recopilación de datos.....	14
5.2	Procesamiento de datos .....	15
5.3	Aplicación del algoritmo Apriori .....	18
6.	Resultados del Análisis.....	22
7.	Aplicaciones Prácticas para el Negocio.....	25
8.	Conclusiones.....	30
9.	Bibliografía.....	35
10.	Anexo .....	37
10.1	Código R.....	37
10.2	Resultados del Análisis en R.....	49

## Índice de figuras

Figura 1: Diagrama entidad-relación del sistema .....	15
Figura 2: Distribución del tamaño de las transacciones de los clientes .....	21
Figura 3: Plano de distribución propuesto del supermercado del cliente .....	25

## Índice de anexos

Anexo 1: Contenido de los datasets .....	49
Anexo 2: Top 15 productos más populares .....	50
Anexo 3: Número de establecimientos únicos .....	50
Anexo 4: Número de productos únicos.....	50
Anexo 5: Número de transacciones únicas .....	50
Anexo 6: Transacciones por establecimiento .....	51
Anexo 7: Establecimiento más concurrido.....	51
Anexo 8: Alcance de los datos (número de familias con transacciones) .....	51
Anexo 9: Número de transacciones en las que está incluido cada artículo.....	51
Anexo 10: Distribución de la duración de las 30 campañas de marketing .....	51
Anexo 11: Gasto medio de los clientes.....	52
Anexo 12: Distribución de los cupones .....	52
Anexo 13: Frecuencia de compra media de los clientes .....	52
Anexo 14: Dataset de transacciones .....	52
Anexo 15: Tamaño de las compras.....	52
Anexo 16: Artículos más frecuentes (términos absolutos y relativos) .....	53
Anexo 17: Itemsets comprados al menos 75 veces.....	53
Anexo 18: Reglas de asociación.....	54
Anexo 19: Evaluación de las reglas.....	54
Anexo 20: Reglas maximales .....	55
Anexo 21: Reglas redundantes .....	55

# **1. Introducción**

## **1.1 Contextualización del tema**

La creciente importancia de la recopilación y análisis de datos en el panorama actual es innegable, constituyéndose como un pilar fundamental en la toma de decisiones estratégicas. El constante avance de las tecnologías ha facilitado la generación instantánea de enormes volúmenes de datos, sobre todo gracias a la proliferación de dispositivos digitales en la vida cotidiana. Desde el uso de redes sociales hasta la realización de transacciones financieras, cada interacción humana genera datos que pueden ser aprovechados para obtener información de valor.

En este contexto, el desarrollo de técnicas analíticas cada vez más sofisticadas, como el Machine Learning, el análisis de series temporales o el análisis de imágenes, ha abierto un mundo de posibilidades sin límites. Estas técnicas permiten identificar patrones complejos, predecir tendencias futuras y tomar decisiones informadas en tiempo real. El potencial transformador del análisis de datos en diversos campos, como la medicina, el marketing o las finanzas, es ya una realidad.

Adicionalmente, el reciente período de prácticas realizadas en el área de consumo de una consultora estratégica alemana ha sido revelador para la autora de este trabajo de investigación. Durante este tiempo, ha sido testigo del impacto tangible del análisis de datos en la formulación de estrategias empresariales efectivas. Esta experiencia ha fortalecido su fascinación por el mundo del Big Data y su determinación por aplicar técnicas analíticas avanzadas en un contexto de consultoría real.

## **1.2 Objetivo del trabajo**

El objetivo principal del presente trabajo de investigación es por tanto resolver un caso de consultoría estratégica utilizando una técnica avanzada de análisis de datos como es el análisis de cesta de la compra. Este análisis busca descubrir la relación que existe entre productos que se compran conjuntamente, y se utilizará para resolver un caso de negocio de consultoría estratégica donde el objetivo del cliente, un retailer de supermercados minoristas, es aumentar sus ventas en los próximos años. Se ha utilizado R Studio como herramienta de apoyo para realizar el análisis.

### **1.3 Estructura del trabajo**

Para lograr cumplir el objetivo expuesto anteriormente, la metodología de la investigación ha sido la siguiente: Se proporcionó la base de datos extraída de Kaggle: *Dunnhumby - The Complete Journey*, como punto de partida del proyecto, y se decidió integrarla como parte de un caso de consultoría estratégica. Previamente a resolver el caso, se realizó una revisión de literatura académica, consulta de páginas web e informes empresariales para entender el sector de los supermercados minoristas, los factores clave del éxito y técnicas de análisis de datos utilizadas por empresas del sector. Una vez contextualizada y encaminada la investigación se procedió a realizar un análisis exploratorio de datos para comprender más a fondo la información proporcionada y, a continuación, se aplicó el algoritmo Apriori propio de este tipo de análisis. Como última parte de la investigación, se estudiaron los resultados obtenidos y se desarrollaron conclusiones y sugerencias con el objetivo de potenciar las compras del minorista protagonista del caso de negocio.

### **1.4 Metodología del trabajo**

En cuanto a la forma de estructurar la investigación, el trabajo se ha organizado en varios apartados con la finalidad de proporcionar una visión completa y coherente del proceso de análisis y resolución del caso planteado. Comienza un capítulo que expone brevemente en qué consiste el caso de negocio, sentando así las bases de la investigación. Le sigue un segundo capítulo de análisis del objetivo que busca contextualizar el caso y entender cómo funciona el sector de los supermercados. El tercer capítulo expone la técnica seleccionada, análisis de cesta de la compra, y demuestra su utilidad para cumplir con el objetivo de la investigación. Continúa la parte central del análisis, la aplicación de la técnica de análisis de datos, la cual se divide en tres fases consecutivas: recopilación de datos, procesamiento de la información y aplicación del algoritmo Apriori. Tras este capítulo, se recopilan los resultados obtenidos en el análisis y se ofrecen recomendaciones estratégicas y sugerencias de próximos pasos a seguir.

## **2. Exposición del Caso de Negocio**

Se considera que el equipo de consultores de una empresa de consultoría estratégica alemana ha sido recientemente contactado por un cliente del sector minorista que busca impulsar sus ventas en los años venideros.

Según la información revelada, el cliente es un retailer, propietario de varios establecimientos tipo supermercado que visitan con frecuencia 2.500 familias de la región en la que se encuentran ubicados. La cadena de supermercados cuenta con más de 500 establecimientos y comercializa más de 90.000 productos diferentes (Gonen, 2019).

Hace dos años, con el impulso de las nuevas tecnologías y la aparición de herramientas de recopilación y análisis de datos, el cliente implementó un software de recogida de información con el que se recaban datos de todas las transacciones realizadas por los consumidores. Esto le permite conocer el número de transacciones realizadas en un día, monitorizar el inventario de productos de cada establecimiento o llevar el registro de los clientes, entre otros aspectos.

### **3. Objetivo del Caso de Negocio**

Como se ha comentado anteriormente, el cliente establece como objetivo del proyecto aumentar las ventas de su negocio, lo cual indica una clara necesidad de implementar estrategias efectivas que impulsen el crecimiento y la rentabilidad de la empresa.

Es esencial para dar solución a este problema y poder entender qué técnicas son las adecuadas para ello, adentrarse en el contexto de los supermercados minoristas y explorar qué factores determinan el éxito en ventas de estos.

#### **3.1 Análisis del Mercado Minorista de Supermercados**

En la actualidad, el sector de los supermercados minoristas está experimentando cambios significativos para adaptarse a las nuevas dinámicas del mercado, marcadas por la evolución en las preferencias del consumidor y los avances tecnológicos. Se trata, además, de una industria altamente competitiva, en la que tener una estrategia óptima de ventas resulta clave para destacar entre la amplia oferta existente y garantizar la fidelidad de los clientes (Aull et al., 2022).

Según el estudio de tendencias del sector retail de alimentos realizado por Oracle en 2023, el sector destaca por el impacto significativo que la pandemia ha tenido en la industria alimentaria. Esta crisis, además de provocar desafíos y complicaciones, actuó como catalizador para la rápida adopción de innovaciones que han transformado radicalmente la forma en que los consumidores interactúan con los productos alimenticios. Entre estas innovaciones, destacan el crecimiento del e-commerce y la demanda de servicios de entrega a domicilio.

Adicionalmente, el cambio constante de preferencias de los consumidores representa un desafío continuo para los supermercados, que deben mantenerse al tanto de las nuevas tendencias y anticiparse a las demandas de los clientes prácticamente a tiempo real. Entre estas tendencias emergentes, se encuentra el creciente interés por parte de los consumidores de adoptar estilos de vida más saludables y abogar por prácticas comerciales más sostenibles. Este cambio se refleja en una demanda cada vez mayor de alimentos frescos y opciones orgánicas, y nuevas prácticas como la búsqueda de proveedores que compartan valores de responsabilidad social y ambiental (Delberghe et al., 2024).

Paralelamente, en un contexto marcado por el constante progreso tecnológico, la digitalización y el análisis de datos están desempeñando un papel fundamental en la transformación de aspectos críticos del negocio de supermercados. Entre estos aspectos destacan, la gestión eficiente de stocks, donde las herramientas de análisis predictivo están permitiendo a los supermercados optimizar sus inventarios y evitar obsolescencia y escasez de productos, y las estrategias de fidelización de clientes, donde se consigue que los establecimientos minoristas personalicen ofertas y promociones en función de las preferencias de sus consumidores, mejorando así la experiencia de compra y fortaleciendo la lealtad del cliente.

Este cambio de paradigma hacia una gestión más eficiente y orientada hacia el cliente está redefiniendo el panorama competitivo del sector y abriendo nuevas oportunidades de crecimiento e innovación.

### **3.2 Claves del Éxito en Supermercados**

Entre los factores y claves que determinan el éxito en ventas de un supermercado o establecimiento del tipo encontramos:

- **Experiencia de cliente:** Es esencial garantizar una experiencia de compra satisfactoria para los clientes, ya que esta puede fortalecer la fidelización y potenciar las ventas. Factores como la selección cuidadosa de productos y la disposición estratégica de los artículos en tienda pueden tener un impacto significativo en las decisiones de compra de los clientes. Además, la atención personalizada o un servicio rápido pueden marcar la diferencia en la percepción del cliente sobre la calidad del establecimiento. En este sentido, aunque el auge de las tecnologías dirija el mundo hacia los pedidos online, se conoce que el cliente sigue valorando poder ir al establecimiento a realizar su compra, por lo que es esencial para los supermercados seguir potenciando las tiendas físicas y la búsqueda de la experiencia de cliente perfecta. (Aversa et al., 2024).
- **Estrategia de marketing y publicidad:** Muchos establecimientos utilizan estrategias como la aplicación de promociones para incentivar la compra o la creación de paquetes de productos para aumentar las ventas cruzadas. Implementar promociones individuales como el 2x1 busca generar un sentido de urgencia en los clientes y motivarlos a realizar una compra en el momento. Por su

lado, los paquetes de productos aprovechan la teoría de la complementariedad, que se basa en la idea de que los clientes tienden a adquirir productos complementarios cuando se ofrecen juntos como parte de un paquete. Esta estrategia puede aumentar el valor del ticket medio de compra y fomentar la fidelización al ofrecer una experiencia de compra conveniente.

- **Gestión óptima de inventario:** Garantizar que se cuenta con suficiente stock de productos demandados o anticiparse a las necesidades de los clientes también es una estrategia dirigida a incrementar las ventas en cualquier negocio minorista. Una gestión eficiente del inventario no solo asegura que los productos estén disponibles cuando los clientes los necesiten, sino que también puede tener un impacto significativo en la rentabilidad del negocio (e.g., al evitar la escasez de productos populares, se reduce el riesgo de pérdidas de ventas debido a la falta de disponibilidad). Adicionalmente, una buena gestión de inventario puede ayudar al negocio a minimizar sus costes, reduciendo la obsolescencia y la necesidad de almacenar grandes cantidades de producto.
- **Personalización de la experiencia:** La estrategia de ofrecer sugerencias y recomendaciones a cada perfil de cliente se considera otro de los factores clave del éxito. Es esencial para ello comprender las preferencias y los comportamientos de compra de los consumidores. Esto puede incluir desde el seguimiento de historiales de compras hasta el análisis de datos demográficos. Con esta información, los establecimientos pueden crear perfiles de cliente detallados y ofrecer experiencias de compra altamente personalizadas, aumentando así la satisfacción del cliente y las tasas de conversión.

Una vez entendido el contexto de mercado y conocido el terreno de juego de los supermercados minoristas, se deben explorar y adoptar técnicas o herramientas que consigan impulsar eficazmente los factores clave de éxito previamente identificados.

## 4. Metodologías de Análisis de Datos

Dentro de las posibles técnicas de análisis de datos que podrían utilizarse para potenciar las ventas de una cadena de supermercados, como el análisis de segmentación de clientes, los modelos de predicción de demanda o los modelos de recomendación, se ha escogido el análisis de cesta de la compra (market basket analysis) como herramienta óptima para cumplir con el objetivo establecido.

El análisis de cesta de la compra o análisis de asociación es una técnica de minería de datos que busca identificar aquellas compras de productos que suelen ocurrir conjuntamente (e.g., las personas que compran aceite suelen comprar sal o aquellas que compran champú también pueden comprar acondicionador). El objetivo del análisis es entender qué relación hay entre los elementos, por lo tanto, es un método de extracción de datos que intenta encontrar los conjuntos de artículos que suelen estar asociados entre sí con regularidad durante las transacciones.

Es importante comprender qué se entiende por cesta de la compra. Se trata del conjunto de productos o artículos que el consumidor elige durante su visita a un establecimiento tipo retail (como un supermercado o una tienda de ropa). Este conjunto de artículos también se conoce como "set de artículos" o "itemset" y engloba todos los productos adquiridos en una misma compra, registrados bajo un solo ticket.

El propósito del análisis de la cesta de la compra es por tanto comprender mejor el comportamiento del cliente mediante la evaluación de patrones, tendencias y relaciones entre los productos comprados en una sola visita. En el ámbito de la venta al por menor y el comercio electrónico, esta técnica se utiliza con frecuencia permitiendo a los establecimientos comprender los hábitos de compra de los clientes y optimizar las estrategias de ventas. Al identificar las combinaciones de productos más frecuentes, las empresas pueden tomar decisiones estratégicas como la ubicación de productos en tienda, la creación de promociones personalizadas e incluso el desarrollo de nuevos productos que satisfagan las demandas específicas de los clientes.

El análisis de cesta de la compra se realiza recopilando datos detallados sobre las transacciones de los clientes, incluyendo el tipo de producto, la fecha de compra y el identificador de la transacción, entre otros, lo que permite identificar vínculos y

asociaciones entre diferentes productos y revela patrones de compra recurrentes en los consumidores.

Aunque el análisis de la cesta de la compra es comúnmente empleado para comprender los hábitos de consumo en el sector minorista, su utilidad se extiende a otros ámbitos, incluyendo la detección de reclamaciones fraudulentas en seguros o el desarrollo de sistemas de recomendación para la venta cruzada de productos, como los implementados por gigantes del comercio electrónico Amazon y Netflix. Este enfoque analítico se ha mostrado esencial para la identificación de patrones y tendencias que permiten optimizar procesos y mejorar la experiencia del usuario en una gran variedad de industrias (Inigo, 2018).

Como se ha evidenciado a lo largo de este capítulo, a través de la herramienta de análisis de datos seleccionada se consiguen abordar completamente los factores que aseguran el éxito de una cadena de supermercados expuestos anteriormente.

## **5. Análisis de Cesta de la Compra**

Se dividirá el análisis en tres fases consecutivas: una primera fase de recopilación de datos, una segunda fase de procesamiento de la información y una tercera fase de aplicación del algoritmo Apriori.

### **5.1 Recopilación de datos**

Una base de datos es una estructura organizada de información que sirve para recopilar, almacenar y gestionar datos de diversas naturalezas y formatos.

En la actualidad, contar con una base de datos extensa es determinante para cualquier organización, pues se utiliza como instrumento para administrar eficazmente los procesos del negocio, facilitando la organización y el acceso a la información empresarial. Este sistema permite organizar y acceder a información crítica de manera eficiente, asegurando que los datos relevantes estén disponibles cuando se necesiten.

Entre los numerosos beneficios de contar con un sistema de base de datos se encuentran por tanto, la agilidad en la toma de decisiones, la eficiencia operativa y la integración de datos de múltiples fuentes. Al centralizar la información, se elimina además la redundancia y se asegura la consistencia de los datos, obteniendo una visión más clara y precisa del negocio. Esto, a su vez, permite tomar decisiones informadas de manera más rápida y efectiva.

Para la construcción de este sistema, es fundamental desarrollar tablas específicas que representen de manera precisa y estructurada los diversos aspectos del negocio, como productos, clientes o proveedores. Además, es importante definir las relaciones existentes entre las diferentes tablas, con el objetivo de garantizar una integración coherente y completa de la información.

Para este business case, el cliente proporciona una base de datos construida a partir de la implantación de un software interno de extracción de información. Este conjunto de datos contiene más de 250.000 transacciones que realizaron 2.500 hogares que son compradores frecuentes de los supermercados del cliente durante los últimos dos años. Incluye todas las compras realizadas por cada hogar, recogiendo la cantidad de unidades compradas de cada producto, el nombre y tipo del mismo, el local donde se realizó la

compra, e incluso el precio de venta de cada artículo. Además, se proporciona información demográfica del consumidor, así como el historial de contactos de marketing directo, datos sobre las campañas lanzadas o sobre los descuentos aplicados en cada transacción.

La información no está contenida en una sola tabla, sino que la base de datos se compone de ocho hojas de cálculo diferentes que se relacionan entre sí mediante campos comunes. En la figura I se presenta una visualización que muestra las diversas tablas proporcionadas, junto con los campos que las componen, y cómo se interrelacionan entre sí para conformar una base de datos integral y detallada.

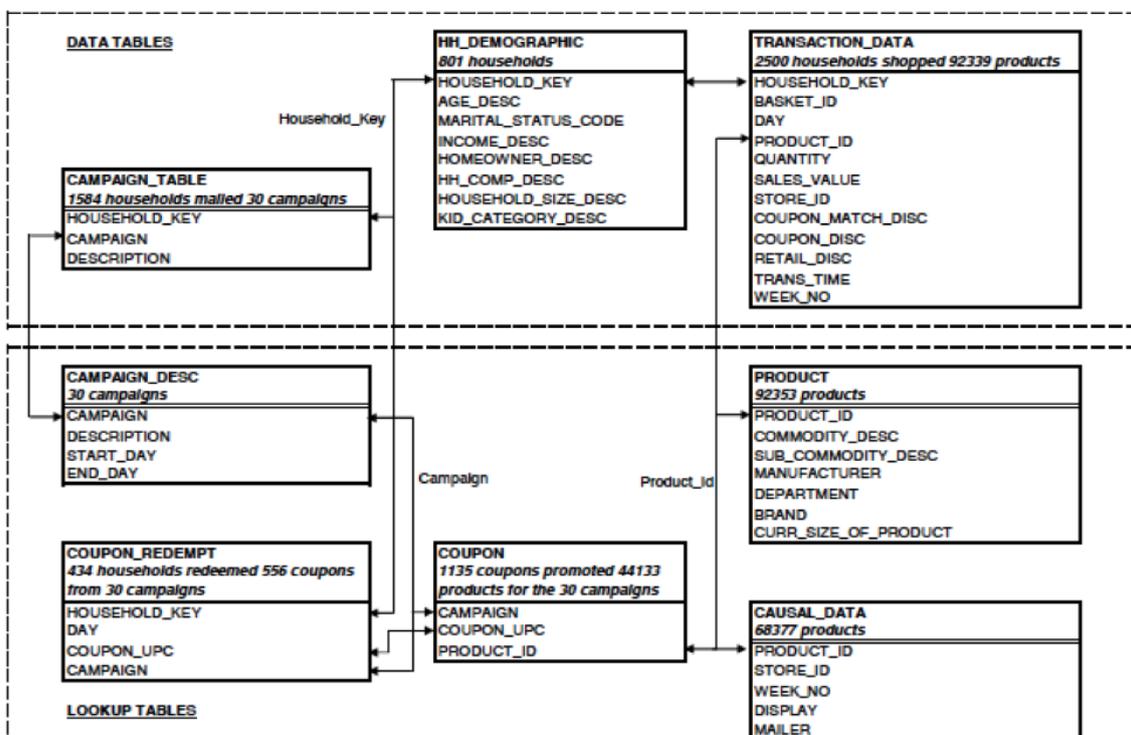


Figura 1: Diagrama entidad-relación del sistema

Fuente: Github

## 5.2 Procesamiento de datos

Los datasets descargados ya están limpios y revisados por el cliente, por lo que se supone que estos son de alta integridad, confiables y faltos de errores.

Debido a la complejidad de los cálculos y el alto número de combinaciones posibles, se recurre al uso de softwares especializados para llevar a cabo este tipo de análisis. En este

caso, se utilizará la herramienta de programación R Studio, entorno de desarrollo para el lenguaje R. Se trata de un software libre diseñado para tareas estadísticas y de visualización que destaca por su versatilidad y potencia. Es por su utilidad en el campo de la estadística, su flexibilidad y el conocimiento previo de la autora sobre el lenguaje, que se ha escogido como instrumento de apoyo para resolver el caso de negocio.

Además de la información facilitada por el cliente al inicio del proyecto, se necesita comprender en detalle la situación operativa de la empresa para poder así definir una estrategia óptima. Es esencial investigar cuántos productos son comercializados por el minorista, identificar los más demandados entre ellos, determinar la cantidad de establecimientos que opera o evaluar el nivel de afluencia de estos, entre otros aspectos relevantes.

En este contexto, el análisis se enfocará en entender a fondo las características del negocio y los patrones de compra de los clientes, en lugar de profundizar en aspectos de tipo demográficos.

Con este objetivo, se realizará un análisis exploratorio previo a la aplicación del algoritmo Apriori apoyándose también en la herramienta R Studio. El análisis exploratorio de datos o AED busca descubrir las relaciones y características no observables a simple vista en un conjunto de datos. Es una herramienta utilizada para comprender la mejor manera de manipular la información a la hora de resolver un problema, además de permitir descubrir patrones o verificar asunciones.

Filtrando el dataset del cliente, se puede conocer que la empresa cuenta con exactamente 582 establecimientos, en los que comercializa 92.353 productos diferentes. Estos datos sugieren una cadena de supermercados de gran escala, pues mantener una red tan extensa de locales requiere recursos significativos y una presencia de marca establecida en las regiones donde se encuentra. Además, su extenso catálogo de productos sugiere una oferta diversificada y posiblemente una clientela con perfil variado. Estos indicadores apuntan a una empresa grande con presencia significativa en el mercado.

En este sentido, el establecimiento 367 destaca como el más activo en términos de volumen de transacciones, con un total de 75.573 registros. Se encuentra a una distancia considerable del segundo establecimiento más concurrido, el 406, que acumula 59.331 transacciones. Este hallazgo sugiere que el establecimiento 367 es un punto importante

para la actividad comercial de la marca, con un volumen significativamente mayor de transacciones en comparación con otros establecimientos. Esto puede indicar una mayor demanda de productos, una mejor ubicación geográfica o una estrategia de marketing más efectiva. De la misma forma, es un target claro a la hora de priorizar la implementación de las estrategias de venta.

Teniendo en cuenta todos los establecimientos del cliente, sus diez productos más populares o demandados estos dos últimos años han sido por orden: plátanos, gasolina regular sin plomo, leche entera, huevos, pan de molde, pan de hamburguesa, batido de chocolate, fresas, pimiento verde y pan de perrito caliente. Podemos inferir de esta información que el cliente opera en múltiples sectores, ya que los productos mencionados son variados y no pertenecen todos a una única categoría. Además, la popularidad de ciertos productos como plátanos, leche entera o huevos sugiere una demanda constante de alimentos básicos y comunes en la dieta diaria. La comercialización de productos como gasolina regular sin plomo indica que el cliente podría tener operaciones que van más allá de la venta minorista de alimentos, posiblemente incluyendo servicios relacionados con el transporte como las gasolineras. Por último, parece que los establecimientos se localizan en zonas donde las hamburguesas y hot dogs son algunos de los platos más populares.

Al reorganizar los datos, se puede observar que el gasto promedio de los clientes de la cadena de supermercados en cada visita es de 29,14 euros<sup>1</sup>. Esta información sugiere que los clientes tienden a realizar compras de tamaño reducido, dejando ver que se trata de establecimientos frecuentados para adquirir productos del día a día o realizar compras repentinas de última hora. Este patrón de gasto refleja, por tanto, una preferencia del consumidor por visitas más frecuentes en lugar de compras masivas, resaltando la importancia de mantener un inventario bien surtido de artículos de uso cotidiano, así como de productos atractivos que puedan incentivar compras no planificadas.

El análisis revela también que el minorista lanzó un total de 30 campañas durante el período de tiempo en que se recopilaron los datos. Estas campañas tienen una duración media de 47 días, con la campaña más corta durando 32 días y la más larga extendiéndose a 161 días. Esto sugiere una tendencia del cliente a lanzar campañas relativamente breves,

---

<sup>1</sup> Se desconoce la unidad de medida real

pero de manera constante, lo que puede indicar una estrategia de marketing centrada en mantener la atención del consumidor a través de promociones frecuentes y variadas.

La base de datos proporcionada incluye información sobre los cupones utilizados por los clientes. Al analizar y filtrar estos datos, se ha identificado que el gasto promedio por compra es de aproximadamente 27 euros cuando no se utiliza un cupón, mientras que esta cifra se eleva a una media de 68 euros cuando se aplica un cupón. A su vez, se observa que los descuentos permiten a los clientes ahorrar, en promedio, unos 2.5 euros por transacción. Este análisis sugiere que el uso de cupones incentiva a los clientes a realizar compras significativamente mayores y por ello debería ser una estrategia interesante que considerar. Esta tendencia puede indicar que los clientes aprovechan los cupones para adquirir productos adicionales o de mayor valor que normalmente no comprarían sin el incentivo del descuento (Haijie, 2022).

Para continuar con el estudio de los patrones de compra de los clientes, se ha examinado la frecuencia con la que los consumidores realizan sus compras. En promedio, los clientes de la cadena de supermercados visitan una de sus tiendas aproximadamente cada 16 días, es decir 2 veces al mes. Esta información no solo revela la regularidad con la que los clientes visitan un establecimiento, sino que permite conocer la frecuencia con la que necesitan reabastecerse los productos.

Por último, se ha observado que, a pesar de contar con información demográfica de 2.500 familias, los datos de transacciones que se utilizarán en este análisis están disponibles únicamente para 801 de los hogares de la muestra, lo que limita el alcance de las familias de las cuales podemos obtener información sobre sus hábitos de compra.

Se concluye gracias a este análisis, que el cliente posee una cadena de supermercados de gran escala, con una presencia significativa en el mercado, gran cantidad de establecimientos y una gama diversificada productos. Su clientela parece ser variada y parece que se utiliza como recurso para compras más pequeñas y frecuentes.

### **5.3 Aplicación del algoritmo Apriori**

Una vez procesada y comprendida la información, y con un entendimiento más profundo sobre el negocio del cliente, se requiere descubrir las reglas de asociación que predominan en el conjunto de datos.

Se ha decidido proceder el análisis utilizando el algoritmo Apriori debido a su eficiencia, flexibilidad y facilidad de implementación, aunque también se consideraron otros algoritmos diseñados con el mismo propósito de detectar conjuntos de elementos frecuentes, como el FP-growth o el Eclat.

El algoritmo Apriori es una regla de asociación que se utiliza para identificar relaciones o asociaciones en un conjunto de datos concreto. Estas asociaciones se representan mediante reglas del tipo "si {X}, entonces {Y}", donde X e Y representan conjuntos de elementos. Estos conjuntos de elementos o itemsets pueden estar formados por uno o más artículos, siendo el término de la izquierda el antecedente (LHS) y el de la derecha el consecuente (RHS) de la regla de asociación.

El algoritmo se compone de dos fases consecutivas. En la primera fase, se identifican los elementos que tienen una frecuencia superior a un límite establecido por el usuario con el objetivo de eliminar la información de poca relevancia. En la segunda fase, se lleva a cabo un análisis bottom-up para convertir ese conjunto de elementos relevantes en varias reglas de asociación. El algoritmo va agregando elementos gradualmente y verifica si cumplen con los requisitos establecidos para integrar o frenar la regla (Amat, 2018).

El algoritmo hace uso de cuatro parámetros fundamentales que permiten entender su funcionamiento e interpretar las reglas de asociación identificadas:

- **Frecuencia:** Muestra el número de veces que los productos de un itemset aparecen de manera conjunta en un conjunto de transacciones. Se calcula sumando el número de transacciones que contienen un itemset concreto.
- **Soporte:** Se trata de un indicador de frecuencia relativa que muestra el número de veces que un determinado itemset aparece en un conjunto de transacciones. Se calcula dividiendo la frecuencia de un itemset concreto entre el número total de transacciones.
- **Confianza:** Indica la probabilidad con la que dos grupos de artículos (antecedente y consecuente) se compran conjuntamente. Se calcula dividiendo el soporte de la unión de dos itemsets entre el soporte del antecedente.
- **Elevación:** Mide el aumento en la probabilidad de compra del artículo o artículos consecuentes dado el antecedente, en comparación con la probabilidad de compra

del consecuente si el antecedente no estuviera presente. Si el valor del indicador es mayor que 1, la presencia del antecedente aumenta la probabilidad de compra del consecuente. Se calcula dividiendo la confianza entre antecedente y consecuente entre el soporte del consecuente. Se utiliza para evaluar las reglas de asociación identificadas.

El usuario determina límites para filtrar los datos haciendo uso de estos parámetros, basándose para ello en su propio criterio y comprensión de la información. La selección de estos límites es fundamental para garantizar que el análisis se enfoque en los elementos más relevantes y significativos para resolver el caso planteado.

Para esta fase del análisis, sólo se utilizará el dataset "*datos\_transacciones*", en concreto las columnas que se utilizarán son:

**BASKET\_ID:** Identificador único de once dígitos asignado a cada transacción.

**PRODUCT\_ID:** Identificador único, de longitud variable, de cada artículo comercializado.

Se genera un nuevo conjunto de datos denominado "*transacciones*", el cual contiene únicamente la información del identificador de la transacción y los productos adquiridos en ella. Este conjunto se convierte así en un registro histórico que documenta cada compra realizada por los clientes en el pasado. Para aplicar el algoritmo Apriori, es necesario transformar este conjunto de datos en una estructura transaccional, es decir, una matriz binaria donde cada fila representa una transacción individual y cada columna identifica un artículo específico o elemento involucrado en dicha transacción, conteniendo el valor 1, si el artículo se ha comprado en esa transacción, o 0 si no se ha hecho.

Al analizar el dataset del cliente con R Studio, se observa que la mayoría de los consumidores adquieren entre 5 y 10 productos en una sola transacción, y que el 90% de ellos compra como máximo 24 artículos a la vez. Esto sugiere, confirmando los resultados del análisis exploratorio, que la mayoría de las compras son de tamaño reducido, con una proporción minoritaria de compras más grandes. En la figura II se observa que las transacciones son en general de pocos artículos, acumulándose la mayoría de ellas en el rango 1-10 artículos por compra.

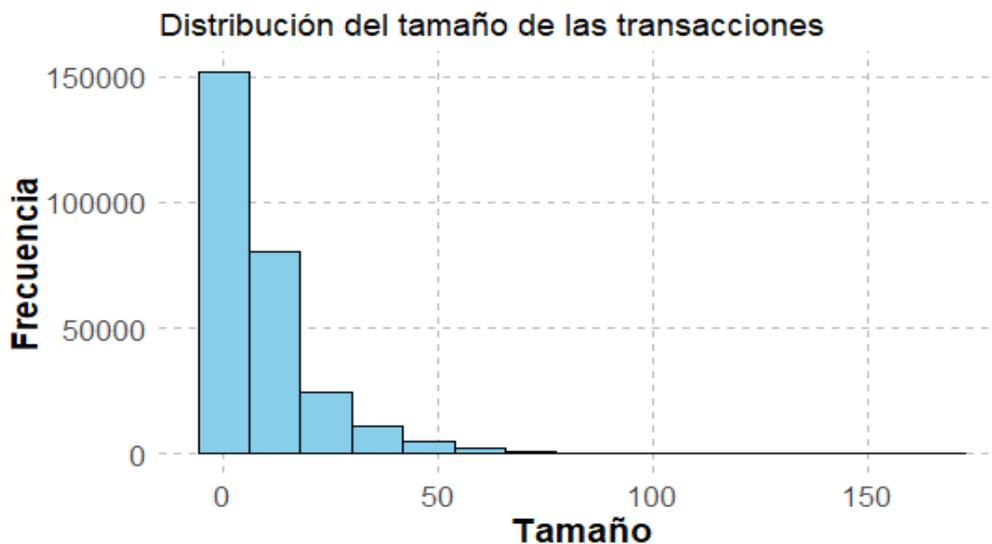


Figura 2: Distribución del tamaño de las transacciones de los clientes

Fuente: Elaboración propia con datos de Kaggle

Como primera etapa del análisis, es fundamental determinar qué elementos superan un límite de soporte establecido. Este umbral de soporte se define con el objetivo de identificar conjuntos de elementos que se presenten con una frecuencia suficiente en el conjunto de datos para considerarlos relevantes desde una perspectiva comercial.

En este contexto, y habiendo analizado el número de veces que se repite cada artículo en las transacciones estudiadas, se ha fijado el umbral de aparición de un mismo artículo en al menos 75 transacciones, lo que asegura la captura de conjuntos de elementos pertinentes y frecuentes en las transacciones de los clientes. Esta selección deja en total 15.206 itemsets que superan el umbral de soporte definido, el cual se ha establecido en 0,0002712634. Con este filtrado se generan varios conjuntos de elementos, estando la mayor parte de ellos, en un 48% de las ocasiones, formados por un único artículo, mientras que el 46% se compone de dos artículos, quedando el resto de las transacciones conformadas por conjuntos de 3 y 4 artículos.

La segunda etapa del análisis se centra en crear las reglas de asociación, necesitando para ello establecer una confianza mínima para que una regla se considere válida. En este caso se define como 70% por razones de fiabilidad y relevancia, asegurando que la probabilidad de que una regla sea cierta sea lo suficientemente alta y que estemos considerando relaciones realmente relevantes para el análisis.

## 6. Resultados del Análisis

R Studio genera una tabla donde se pueden observar varios parámetros de las reglas resultantes: lhs (antecedente), rhs (consecuente), support (soporte), confidence (confianza), lift (elevación) y count (conteo). Se trata de reglas de asociación con el mismo significado: Si el artículo o artículos del lado izquierdo (antecedente) está presente en una transacción, el producto o productos del lado derecho (consecuente) también lo está con cierta probabilidad.

De esta forma, se identifican 11 reglas de asociación diferentes que superan el soporte y confianza establecidos. Cuatro de las reglas están compuestas por dos elementos y las siete restantes por tres artículos distintos. Las reglas o relaciones identificadas en orden de más a menos probable son las siguientes:

- Si el cliente compra zumo sin carbohidratos 50% natural, entonces también comprará un pack de plátanos.
- Si el cliente compra un paquete de 6 latas de bebida gaseosa y un paquete de pastelitos, entonces también comprará un pack de plátanos.
- Si el cliente compra una tarjeta regalo electrónica, entonces también comprará un producto indeterminado del que no se recoge información.
- Si el cliente compra un paquete de 6 latas de bebida gaseosa y pan de molde, entonces también comprará un pack de plátanos.
- Si el cliente compra leche normal y un paquete de 6 latas de bebida gaseosa, entonces también comprará un pack de plátanos.
- Si el cliente compra yogurt de dos marcas o tipologías diferentes, también comprará un pack de plátanos.
- Si el cliente compra yogurt de dos marcas o tipologías diferentes, también comprará yogurt de otro tercer tipo diferente.
- Si un cliente compra aguacates y leche normal, también comprará un pack de plátanos.

- Si un cliente compra comida de perro enlatada de un tipo, también comprará comida de perro enlatada de otro tipo.
- Si un cliente compra carne congelada y pan de perrito caliente, también comprará pan de hamburguesa.
- Si un cliente compra un tipo de revista, también comprará otro tipo de revista.

Esta lista de 11 asociaciones identifica un conjunto robusto de reglas de asociación que sucede más del 70% de las veces y que sirve para proporcionar valiosos insights sobre los patrones de compra de los clientes de la cadena de supermercados.

Adicionalmente, se observa que las 11 reglas identificadas son maximales, es decir, ya se han incluido en ellas todas las combinaciones posibles dentro de una misma regla, y que no existen reglas de asociación redundantes (Amat, 2018).

Una vez determinadas, es útil analizar la elevación de las reglas obtenidas para comprender mejor el impacto de la compra del antecedente en la decisión de adquirir también el consecuente. Como se ha visto anteriormente, la elevación nos permite medir la fuerza de la asociación entre los productos de una regla, indicando si la presencia de un artículo aumenta significativamente la probabilidad de que el cliente también compre el otro. Al calcular la elevación del conjunto de reglas, se advierte que todas ellas tienen valores superiores a seis, siendo la regla de compra de comida de perro la que posee una elevación más alta. Dado que todas las elevaciones superan el valor uno, se puede concluir que existe una asociación positiva entre los antecedentes y consecuentes identificados. A mayor elevación, más fuerte es la relación, lo que indica que los productos del antecedente y el consecuente se compran juntos con más frecuencia de la esperada por azar. Se observa, por ejemplo, que el cliente es 8,6 veces más propenso a comprar un pack de plátanos si ha comprado ya zumo sin carbohidratos 50% natural, en comparación con la probabilidad de comprar plátanos solo.

La alta elevación detectada en las reglas de asociación implica una fuerte correlación entre la compra de ciertos productos. Se observa, por tanto, que la decisión de comprar un artículo en función de si se ha comprado otro juega un papel fundamental para los clientes del supermercado. Esta relación revela potencial para influir en las decisiones de

compra de los clientes y lanzar una estrategia de ventas que incentiven a los consumidores a percibir estos artículos como complementarios, promoviendo su compra simultánea.

## 7. Aplicaciones Prácticas para el Negocio

Los resultados obtenidos en el capítulo anterior permitirán definir una estrategia de ventas que potencie los ingresos de la cadena de supermercados, pues estas reglas no solo revelan las relaciones intrínsecas entre los diferentes productos, sino que también destacan oportunidades estratégicas para optimizar campañas de marketing, mejorar la gestión de inventarios y crear promociones cruzadas. Implementar acciones basadas en estas asociaciones permitirá al cliente no solo incrementar sus ventas, sino también mejorar la satisfacción y fidelización de sus clientes, fortaleciendo su posición competitiva en el mercado.

Una de las primeras acciones a considerar, es la optimización de la disposición de productos en tienda. Conociendo los patrones de compra de los consumidores del supermercado, se podrían colocar juntos los productos complementarios para fomentar compras adicionales.

En este sentido y después de analizar en profundidad las diversas reglas de asociación identificadas y sus interrelaciones, se sugiere una reorganización estratégica de los supermercados del cliente.

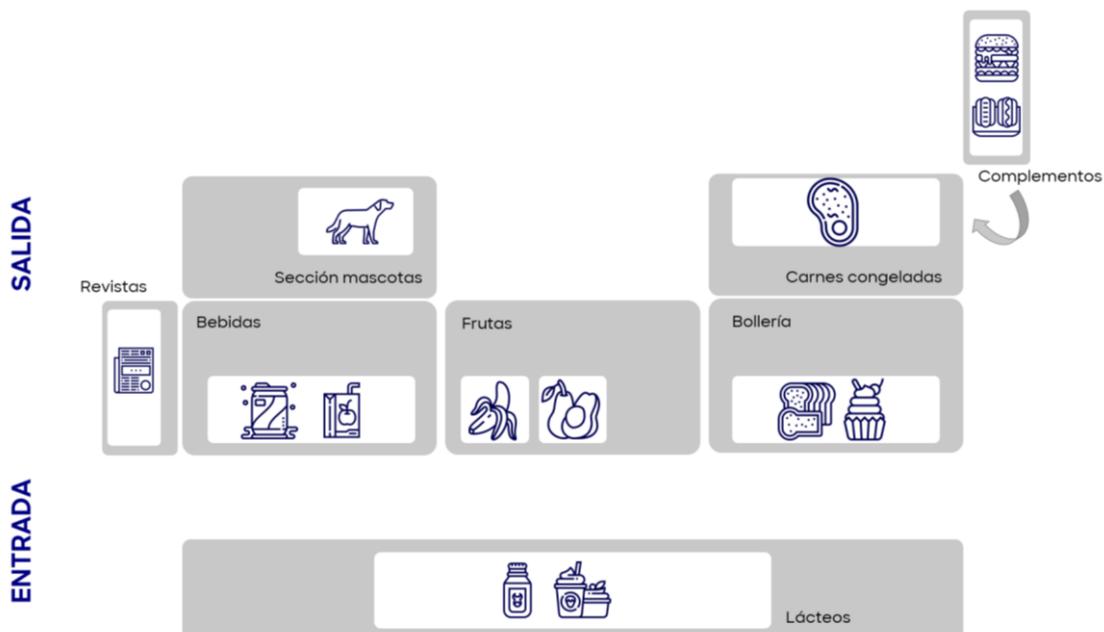


Figura 3: Plano de distribución propuesto del supermercado del cliente

Fuente: Elaboración propia

Como se observa en la figura III, los plátanos deberían ubicarse en un área visible y de fácil acceso, dada su popularidad, lo que busca promover e incentivar su venta. Junto a ellos, debería situarse el aguacate, además de incluir en la zona una sección con las demás frutas. El objetivo es que el cliente que compre plátanos también aproveche para comprar aguacates. En una sección adyacente a las frutas, se deberían disponer las bebidas, dentro de las que se encuentran muchas tipologías, como los zumos y los refrescos en lata (e.g., bebida gaseosa). En otra área, enfrente de estas, se podrían exhibir los lácteos, especialmente la leche y los yogures de diversas marcas. Enfrente de los lácteos y adyacentes también a las frutas, se podrían colocar productos de bollería, como el pan de molde y los pastelitos.

Por otro lado, la sección dedicada a las mascotas podría agrupar todos los alimentos para animales, incluyendo diferentes tipos y marcas de comida, organizados por tipo de animal. De la misma forma, se deberían concentrar en una sección todos los tipos y marcas de revista, a poder ser a la salida del establecimiento, donde los clientes ya hacen cola y se entretienen leyendo antes de pagar.

Finalmente, sería recomendable ubicar cerca de la sección de carnes congeladas un área con diversos acompañamientos y productos complementarios, entre los que se encontraría el pan de perrito caliente y el pan de hamburguesa.

Esta disposición estratégica de los productos no solo facilita las compras cruzadas y promueve asociaciones visuales entre los artículos, sino que también mejora la experiencia del cliente en la tienda. Al encontrar rápidamente los elementos que buscan, los consumidores tienen más tiempo para explorar otros productos y tomar decisiones de compra adicionales. Esta eficiencia en la búsqueda, y la exposición visual de productos complementarios aumenta tanto las ventas, como la satisfacción del cliente, favoreciendo así la rentabilidad general del supermercado.

Las reglas de asociación identificadas permiten también la personalización de ofertas y promociones. En base a los hábitos de compra de los clientes, se pueden dirigir promociones específicas a segmentos de clientes con preferencias similares.

Al conocer que quien compra carne congelada y pan de perrito caliente, también comprará pan de hamburguesa, no solo se pueden colocar estos productos en la misma estantería del supermercado, sino que se podría crear un pack oferta, como por ejemplo un 50% en

pan de hamburguesa si se compra pan de perrito caliente, panes de perrito caliente y hamburguesa al 30% si se compra carne congelada, o un descuento del 20% por comprar los tres productos juntos.

Del mismo modo, se pueden crear packs promocionales que incluyan una variedad de productos relacionados, como yogures de diferentes sabores, revistas de distintos géneros o latas de comida para perros de distintas marcas. Por ejemplo, se puede implementar una promoción de 3x2, donde al comprar dos artículos similares se regala un tercero, o aplicar un 50% de descuento en una marca diferente del mismo producto. Al ofrecer una combinación de productos atractiva a un precio reducido, se incentiva al cliente a adquirir más artículos y explorar nuevas opciones, lo que contribuye a aumentar las ventas.

Otra palanca que se puede activar para fomentar el aumento de ventas es la utilización del análisis de cesta de la compra para prever la demanda de productos y optimizar el inventario en función de las tendencias de compra de los clientes. Se evita así, la escasez de productos populares y se reduce el exceso de inventario en productos menos demandados.

El artículo más demandado y con fuerte asociación a otros elementos, se advierte que es el plátano, por lo que asegurar una constante reposición de este es una estrategia clave. Además, dado que el plátano es un producto perecedero, es importante ajustar adecuadamente la cantidad comercializada diariamente para minimizar las pérdidas por mal estado.

En el análisis exploratorio realizado antes de aplicar el algoritmo Apriori, se ha detectado que ciertos productos, como las fresas, la gasolina regular sin plomo, los huevos, el batido de chocolate y el pimiento verde, aunque no incluidos en las reglas de asociación, son altamente demandados por los consumidores. Esta información es esencial para determinar qué inventario debe priorizarse y asegurar que estos productos estén siempre disponibles en la tienda, garantizando así un flujo constante de ventas.

En relación también con la previsión de la demanda, sería útil garantizar que la disponibilidad de los productos asociados sea similar, lo que permitiría satisfacer la demanda de los clientes que buscan itemsets relacionados. Por ejemplo, si un cliente compra zumo sin carbohidratos 50% natural y encuentra siempre los packs de plátanos cerca, es más probable que compre ambos productos juntos. Por lo tanto, mantener un

inventario adecuado de ambos artículos y asegurar su reposición constante ayudará a maximizar las ventas. Además, monitorear tendencias de compra como esta y ajustar los niveles de inventario en consecuencia, garantizará que los clientes siempre encuentren los productos que desean, lo que puede servir para aumentar la satisfacción del cliente y fomentar la fidelidad a la marca.

Conocer la información sobre los productos más populares y las tendencias de compra actuales, permite también desarrollar campañas de marketing más efectivas. Se pueden de esta forma, destacar los productos más populares y promover la compra conjunta de productos complementarios. Se proponen tres campañas de marketing a partir de los insights generados.

Una estrategia de marketing creada a partir de la información de las reglas de asociación podría ser crear un "combo de desayuno completo" que incluya los productos más populares del supermercado, entre los que se encuentran los plátanos, la leche o el pan de molde. Esta iniciativa puede incentivar la compra al presentar una solución conveniente y atractiva para el desayuno de los clientes. Además, se podría promocionar este combo como una opción saludable y equilibrada para comenzar el día, resaltando los beneficios nutricionales de cada producto. Esta estrategia no solo aumentaría las ventas de los productos individuales, sino que también mejoraría la percepción del supermercado como un destino para compras completas y convenientes, al ser, según se ha estudiado anteriormente, un lugar común para realizar las compras cotidianas.

Una táctica de ventas similar, podría ser la creación de una campaña promocional llamada "Fiesta americana en casa", centrada en traer la emoción y diversión de las celebraciones del 4 de julio de Estados Unidos directamente a los hogares de los clientes. Se podrían promocionar durante todo el mes paquetes especiales que incluyan carne congelada, pan de perrito caliente y pan de hamburguesa, todo ello diseñado para facilitar la preparación de clásicos platos americanos en casa. Además de ofrecer productos de manera individual, se puede destacar el valor añadido de adquirir todo el pack para poder disfrutar de una experiencia americana completa, aumentando el atractivo de la campaña con descuentos por la compra del pack completo o añadiendo regalos adicionales, como una bandera americana. Es interesante puntuar al cliente que el pack satisface los gustos de toda la familia al ofrecer variedad de opciones, cumpliendo con los más pequeños que prefieran comer hot dogs y, a su vez, con los amantes de las hamburguesas.

Siguiendo la nueva tendencia estudiada de cuidar la alimentación y preocuparse por mantener hábitos de consumo saludables, otra estrategia interesante a lanzar, basada en las reglas de asociación detectadas, una campaña de marketing para crear batidos nutritivos y orgánicos. El objetivo es mostrar al cliente que, con leche, plátanos y aguacate, tres elementos altamente correlacionados, puede crear el batido nutritivo perfecto para cuidar su alimentación y su salud cada mañana. Esta estrategia consiste en crear un pack de tres artículos que invita al cliente a comprarlos conjuntamente. Podrían añadirse al pack promociones que incentiven todavía más la compra o incluir una botella como regalo, donde poder llevar el batido a todos lados. Proponiendo una receta saludable y nutritiva, además de fácil y accesible, se consigue captar la atención de un consumidor cada vez más preocupado por su salud y, como consecuencia, se incentivan las ventas del establecimiento. Esta campaña, como punto adicional, está alineada con los estándares y valores de salud y bienestar que en los últimos años las empresas de diferentes industrias están aplicando como propios.

El análisis de cesta de la compra permite como se ha podido observar, conocer las relaciones ocultas entre productos comercializados en el supermercado del cliente, ayudando a definir estrategias de gestión de inventario, definición de promociones personalizadas, previsión de demanda de productos populares y desarrollo de estrategias de marketing dirigidas. Estas iniciativas potencian e impulsan las ventas de la empresa, consiguiendo que el cliente mejore su rentabilidad en el corto-medio plazo y que el equipo de consultores consiga ofrecer una solución efectiva a su problema.

## 8. Conclusiones

En el presente trabajo de investigación se ha abordado un caso de negocio enfocado en identificar estrategias para incrementar las ventas de una cadena minorista de supermercados. El análisis se ha realizado utilizando una técnica de minería de datos conocida como análisis de cesta de la compra, la cual se centra en descubrir las relaciones de asociación existentes entre productos comprados conjuntamente. Este enfoque ha permitido identificar patrones de asociación en los datos del cliente, proporcionando una base sólida para definir estrategias comerciales que impulsen las ventas y ofrezcan una comprensión más profunda del comportamiento de compra de los clientes.

Después de un estudio exhaustivo, se ha concluido que:

- El cliente opera una extensa cadena de supermercados con una fuerte presencia en el mercado, numerosos establecimientos y una amplia variedad de productos. La base de clientes es diversa y los supermercados se posicionan como lugar donde realizar compras más pequeñas y frecuentes.
- Se han identificado once reglas de asociación significativas en el conjunto de datos proporcionado, todas con una probabilidad de ocurrencia superior al 70%.
- Todas las reglas de asociación identificadas poseen una elevación muy superior a uno, lo que indica una fuerte correlación en la decisión de comprar un artículo en función de si se ha comprado otro. Esto revela un gran potencial para influir en las decisiones de compra mediante estrategias de ventas que incentiven la compra de artículos complementarios.
- Existe una distribución óptima del espacio en los establecimientos que fomenta las compras cruzadas y promueve asociaciones visuales entre los artículos, mejorando la experiencia del cliente en tienda y aumentando las ventas.
- Ciertos productos, como los plátanos, tienen una alta demanda y fuerte asociación con otros artículos, lo que implica la necesidad de una gestión de inventario más dinámica y ajustada a las tendencias de compra.
- Las reglas de asociación facilitan la personalización de ofertas y promociones, permitiendo combinar productos complementarios que los clientes suelen comprar juntos y aplicar descuentos que incentiven la compra.

- El análisis de cesta de la compra permite prever la demanda de productos, asegurando que los productos más populares siempre estén disponibles y que no falten productos complementarios.
- Se podrían lanzar varias campañas de marketing para incrementar las ventas del supermercado, como la creación de paquetes promocionales (e.g., combo de desayuno completo).
- Los resultados y estrategias obtenidos no solo sirven para impulsar las ventas del negocio, sino que también mejoran la satisfacción del cliente y fomentan la fidelidad a la marca.

La aplicación práctica de los resultados obtenidos puede conducir a mejoras sustanciales en la rentabilidad y eficiencia operativa de la cadena de supermercados. Implementar estrategias basadas en las reglas de asociación identificadas no solo aumentará las ventas, sino que también mejorará la experiencia del cliente al proporcionar una oferta más personalizada y acorde con sus hábitos de compra.

A pesar de los resultados positivos obtenidos, durante la realización de este trabajo de investigación se han encontrado algunas limitaciones. Entre las más importantes se encuentra la falta de precisión de algunos de los datos de la muestra proporcionada. Se ha observado un nivel de detalle insuficiente en la descripción de ciertos artículos, lo cual puede afectar a la exactitud de las asociaciones identificadas y, por ende, a la efectividad de las estrategias empresariales propuestas.

Una vez realizado este trabajo, se sugieren varias líneas de investigación o cuestiones que podrían plantearse en el futuro, entre las que destacan: ¿Es factible crear un modelo de previsión de la demanda que pueda anticiparse eficazmente a las necesidades y comportamientos de los consumidores?, ¿cuál es el impacto a largo plazo de las estrategias basadas en el análisis de cestas de compra?, o ¿cuál sería la forma más eficaz de segmentar a los clientes del supermercado para poder personalizar las estrategias de marketing y ventas?

En conclusión, el análisis de cesta de la compra ha demostrado ser una herramienta poderosa para entender y optimizar los patrones de compra en una cadena de supermercados. La implementación de estrategias basadas en estos hallazgos puede ofrecer ventajas competitivas significativas, mejorando tanto la satisfacción del cliente

como la rentabilidad del negocio. Las futuras investigaciones en las áreas propuestas tienen el potencial de profundizar aún más en este campo y aportar nuevas perspectivas y soluciones innovadoras para el sector minorista.

## **Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado**

**ADVERTENCIA:** Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Marta Echevarría Casasola, estudiante de Doble grado en Administración de Empresas y Análisis de Negocios de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Optimización de la Estrategia de Ventas de un Supermercado: Análisis de Cestas de Compra", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente

de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 05/06/2024

Firma: *Marta Echevarría Casasola*

## 9. Bibliografía

Amat, J. (2018) Reglas de Asociación y algoritmo apriori con R. Enlace: [https://cienciadedatos.net/documentos/43\\_reglas\\_de\\_asociacion#Introducción](https://cienciadedatos.net/documentos/43_reglas_de_asociacion#Introducción)

Aull, B., Kuijpers, D., & Läubli, D. (2022). The state of grocery retail around the world. *McKinsey & Company*.

Aversa, J., & Fares, O. H. (2024). Meet the ‘new consumer’: How shopper behaviour is changing in a post-inflation world. *The Conversation*. Enlace: <https://theconversation.com/meet-the-new-consumer-how-shopper-behaviour-is-changing-in-a-post-inflation-world-231671>

Benites, L. (2021) Análisis de la Cesta de la compra: Definición, Ejemplos. *Statologos*. Enlace: <https://statologos.com/analisis-de-la-canasta-de-mercado/>

Delberghe, C., Delbarre, A., Vissers, D., Laizet, F., Läubli, D., Kleis, A., & Vallöf, R. (2024). Signs of Hope—The State of Grocery Retail 2024: Europe. *EuroCommerce and McKinsey & Company*.

*Estrategias a seguir para conseguir el éxito en tu supermercado*. (2001). *Panatta*. Enlace: <https://panatta.es/blog/articulos/consejos-para-exito-en-tu-supermercado/>

EuroCommerce, & McKinsey & Company. (2021). Disruption and Uncertainty: The State of Grocery Retail 2021 - Europe. *EuroCommerce and McKinsey & Company*.

Fernandez, R. (2020). ¿Qué es el Análisis Exploratorio de Datos o EDA? Unipython. Enlace: [https://unipython.com/que-es-el-analisis-exploratorio-de-datos/?utm\\_content=cmp-true](https://unipython.com/que-es-el-analisis-exploratorio-de-datos/?utm_content=cmp-true)

Food, F. (2023). ¿Qué estrategias utilizan los supermercados para aumentar sus ventas? - *Financial Food*. Enlace: <https://financialfood.es/que-estrategias-utilizan-los-supermercados-para-aumentar-sus-ventas/>

Gonen, F. (2019). Dunnhumby - The Complete Journey. [Base de datos]. *Kaggle*. Enlace: <https://www.kaggle.com/datasets/frtgnn/dunnhumby-the-complete-journey/data>

Haijie. (2022). Dunnhumby Exploratory data analysis. *Kaggle*. Enlace: <https://www.kaggle.com/code/haijie/dunnhumby-exploratory-data-analysis>

Inigo. (2023). Venta cruzada: significado, estrategias y 3 ejemplos exitosos - Michel Miró. *Michel Miró*. Enlace: <https://michelmiro.com/venta-cruzada-significado-estrategias-y-3-ejemplos-exitosos/>

Jackley, M. (2023). 10 tendencias del sector retail de alimentos en 2023. *Oracle*. Enlace: <https://www.oracle.com/es/retail/grocery-industry-trends/#:~:text=10%20tendencias%20del%20sector%20retail%20de%20alimentos%20en,ampliados%20de%20entrega%20y%20suscripci%C3%B3n%20.%20M%C3%A1s%20elementos>

*Market basket Analysis using R*. (2018). *Datacamp*. Enlace: <https://www.datacamp.com/tutorial/market-basket-analysis-r>

Rodríguez, D. (2018). Reglas de asociación y market-basket analysis. *Analytics Lane*. Enlace: <https://www.analyticslane.com/2018/08/31/reglas-de-asociacion-market-basket-analysis/>

Smartbridge. (2024). Market Basket Analysis 101: Anticipating Customer Behavior. *Smartbridge*. Enlace: <https://smartbridge.com/market-basket-analysis-101/>

*T-Test, Chi-Square, ANOVA, Regression, Correlation*. . . (s. f). *Datatab*. Enlace: <https://datatab.es/tutorial/market-basket-analysis>

*Una breve introducción al análisis de la cesta de la compra: reglas de asociación*. (2020). *ICHI.PRO*. Enlace: <https://ichi.pro/es/una-breve-introduccion-al-analisis-de-la-cesta-de-la-compra-reglas-de-asociacion-275202586591438>

Kapur, N. (2023). What Is Cross-Selling? A Guide to Boosting Sales Through Add-Ons. *Salesforce*. Enlace: <https://www.salesforce.com/blog/cross-selling/>

## 10. Anexo

### 10.1 Código R

```
## Código R - Análisis de Cesta de la Compra
```

```
## Trabajo fin de grado Marta Echevarría Casasola - Optimización de la Estrategia de  
Ventas de un Supermercado: Análisis de Cesta de la Compra
```

```
# Limpiar el entorno de trabajo y la consola
```

```
rm(list=ls())
```

```
cat("\014")
```

```
# Instalar paquetes
```

```
install.packages("dplyr")
```

```
install.packages("ggplot2")
```

```
install.packages("magrittr")
```

```
install.packages("tsibble")
```

```
install.packages("forecast")
```

```
install.packages("skimr")
```

```
# Cargar librerías
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(magrittr)
```

```
library(tsibble)
```

```
library(forecast)
```

```
library(skimr)
```

```
library(tidyverse)
```

```
library(arules)
```

```
# Importar datasets
```

```
campañas1 <- read.csv("campaign_desc.csv")
```

```
campañas2 <- read.csv("campaign_table.csv")
```

```
causal <- read.csv("causal_data.csv")
```

```
cupones_canjeados <- read.csv("coupon_redempt.csv")
```

```
cupones <- read.csv("coupon.csv")
```

```
demografía <- read.csv("hh_demographic.csv")
```

```
productos <- read.csv("product.csv")
```

```
datos_transacciones <- read.csv("transaction_data.csv")
```

```
# Análisis exploratorio
```

```
# Contenido de los datasets
```

```
summary(campañas1)
```

```
summary(campañas2)
```

```
summary(causal)
```

```
summary(cupones_canjeados)
```

```
summary(cupones)
```

```
summary(demografia)
```

```
summary(productos)
```

```
summary(datos_transacciones)
```

```
# Top 15 productos más populares
```

```
productos_populares <- datos_transacciones %>%
```

```
  group_by(PRODUCT_ID) %>%
```

```
  summarise(Frecuencia = n()) %>%
```

```
  arrange(desc(Frecuencia))
```

```
head(productos_populares, 15)
```

```
IDs_a_buscar <- c(1082185, 6534178, 1029743, 995242, 1106523, 981760, 1133018,  
883404, 951590, 826249, 840361, 1127831, 995785, 1098066, 908531)
```

```
top15 <- subset(productos, PRODUCT_ID %in% IDs_a_buscar)
```

```
print(top15)
```

```
# Número de establecimientos únicos
```

```
num_establecimientos <- datos_transacciones %>%
```

```
  distinct(STORE_ID) %>%
```

```
  n_distinct()
```

```
print(num_establecimientos)
```

```
# Número de productos únicos
```

```
num_productos <- productos %>%
```

```
  distinct(PRODUCT_ID) %>%
```

```
  n_distinct()
```

```
print(num_productos)
```

```
# Número de transacciones únicas
```

```
num_transacciones_unicas <- datos_transacciones %>%
```

```
  distinct(BASKET_ID) %>%
```

```
  nrow()
```

```
print(num_transacciones_unicas)
```

```
# Establecimiento más concurridos
```

```
transacciones_por_establecimiento <- datos_transacciones %>%
```

```
  group_by(STORE_ID) %>%
```

```
  summarise(num_transacciones = n()) %>%
```

```
  arrange(desc(num_transacciones))
```

```
establecimiento_mas_transacciones <-
```

```
transacciones_por_establecimiento$STORE_ID[1]
```

```
num_transacciones_max <- transacciones_por_establecimiento$num_transacciones[1]
```

```

print(establecimiento_mas_transacciones)

print(num_transacciones_max)

# Alcance de los datos (número de familias de las que se tienen transacciones)

transacciones_list <- unique(datos_transacciones$household_key)

demografia_list <- unique(demografia$household_key)

# Función para encontrar elementos comunes

common_member <- function(a, b) {

  a_set <- unique(a)

  b_set <- unique(b)

  common_elements <- intersect(a_set, b_set)

  if (length(common_elements) > 0) {

    return(common_elements)

  } else {

    print("No common elements")

  }

}

trans_data_demo_data_list <- common_member(transacciones_list, demografia_list)

length(trans_data_demo_data_list)

```

```
# Número de transacciones en las que está incluido cada artículo
```

```
transaction_counts <- datos_transacciones %>%
```

```
  group_by(PRODUCT_ID) %>%
```

```
  summarise(transaction_count = n_distinct(BASKET_ID)) %>%
```

```
  arrange(desc(transaction_count))
```

```
print(transaction_counts)
```

```
# Distribución de la duración de las 30 campañas de marketing
```

```
campañas1$duracion= campañas1$END_DAY - campañas1$START_DAY
```

```
summary(campañas1$duracion)
```

```
# Gasto medio de los clientes
```

```
cesta <- datos_transacciones %>%
```

```
  group_by(BASKET_ID) %>%
```

```
  summarise(
```

```
    SALES_VALUE = sum(SALES_VALUE),
```

```
    COUPON_DISC = sum(COUPON_DISC),
```

```
    COUPON_MATCH_DISC = sum(COUPON_MATCH_DISC)
```

```
  ) %>%
```

```
  mutate(Use_coupon = COUPON_DISC != 0)
```

```
head(cesta)
```

```

valor_medio_cesta <- mean(cesta$SALES_VALUE, na.rm = TRUE)

print(valor_medio_cesta)

# Distribución de los cupones

Use_coupon <- cesta %>%

  group_by(Use_coupon) %>%

  summarise(

    SALES_MEAN = mean(SALES_VALUE),

    COUPON_DISC_MEAN = mean(COUPON_DISC),

    COUPON_MATCH_DISC_MEAN = mean(COUPON_MATCH_DISC),

    Nb_coupon = n()

  )

# Porcentaje de cestas con cupones

Use_coupon <- Use_coupon %>%

  mutate(`%_coupon` = round((Nb_coupon / sum(Nb_coupon)) * 100, 2))

print(Use_coupon)

# Frecuencia de compra media de los clientes

veces_semana <- datos_transacciones %>%

  group_by(household_key, WEEK_NO, DAY) %>%

```

```

summarise(SALES_VALUE = sum(SALES_VALUE),

          RETAIL_DISC = sum(RETAIL_DISC),

          COUPON_DISC = sum(COUPON_DISC),

          COUPON_nb = sum(COUPON_DISC > 0)) %>%

ungroup()

frecuencia_visitas <- veces_semana %>%

group_by(household_key) %>%

summarise(DAY_max = max(DAY),

          DAY_min = min(DAY),

          DAY_count = n())

frecuencia_visitas <- frecuencia_visitas %>%

mutate(frequency = round((DAY_max - DAY_min) / DAY_count, 1))

frecuencia_media <- round(mean(frecuencia_visitas$frequency), 1)

print(frecuencia_media)

# Algoritmo A priori

# Dataset de transacciones

transacciones <- read.transactions(file = "transaction_data.csv",

                                  format = "single",

```

```

        sep = ",",

        header = TRUE,

        cols = c("BASKET_ID", "PRODUCT_ID"),

        rm.duplicates = TRUE)

transacciones

colnames(transacciones)[1:5]

rownames(transacciones)[1:5]

inspect(transacciones[1:5])

# Tamaño del las compras

tamanyos <- size(transacciones)

summary(tamanyos)

quantile(tamanyos, probs = seq(0,1,0.1))

# Representación gráfica de la distribución de tamaños

ggplot(data = data.frame(tamanyos), aes(x = tamanyos)) +

  geom_histogram(fill = "skyblue", color = "black", bins = 15) +

  labs(title = "Distribución del tamaño de las transacciones",

        x = "Tamaño",

```

```

y = "Frecuencia") +

# Cambiar el estilo del fondo del gráfico

theme_minimal() +

# Personalizar la apariencia de los ejes y la leyenda

theme(axis.text = element_text(size = 12),

      axis.title = element_text(size = 14, face = "bold"),

      legend.text = element_text(size = 12)) +

# Añadir una línea de cuadrícula en el fondo

theme(panel.grid.major = element_line(color = "gray", linetype = "dashed"),

      panel.grid.minor = element_blank())

# Artículos más frecuentes (términos absolutos y relativos)

frecuencia_items <- itemFrequency(x = transacciones, type = "relative")

frecuencia_items %>% sort(decreasing = TRUE) %>% head(5)

frecuencia_items <- itemFrequency(x = transacciones, type = "absolute")

frecuencia_items %>% sort(decreasing = TRUE) %>% head(5)

IDSfrecuentes <- c(1082185, 6534178, 1029743, 995242, 1106523)

top5 <- subset(productos, PRODUCT_ID %in% IDSfrecuentes)

print(top5)

```

```

# Itemsets comprados al menos 75 veces

soporte <- 75 / dim(transacciones)[1]

itemsets <- apriori(data = transacciones,

                    parameter = list(support = soporte,

                                     minlen = 1,

                                     maxlen = 20,

                                     target = "frequent itemset"))

summary(itemsets)

# Reglas de asociación

soporte <- 75 / dim(transacciones)[1]

reglas <- apriori(data = transacciones,

                  parameter = list(support = soporte,

                                   confidence = 0.70,

                                   # Se especifica que se creen reglas

                                   target = "rules"))

summary(reglas)

inspect(sort(x = reglas, decreasing = TRUE, by = "confidence"))

```

```

# Evaluación de las métricas

metricas <- interestMeasure(reglas, measure = c("coverage", "fishersExactTest"),
                             transactions = transacciones)

metricas

quality(reglas) <- cbind(quality(reglas), metricas)

# inspect(sort(x = reglas, decreasing = TRUE, by = "confidence"))

df_reglas <- as(reglas, Class = "data.frame")

df_reglas %>% as.tibble() %>% arrange(desc(confidence)) %>% head()

# Reglas maximales

reglas_maximales <- reglas[is.maximal(reglas)]

reglas_maximales

inspect(reglas_maximales[1:10])

# Reglas redundantes

reglas_redundantes <- reglas[is.redundant(x = reglas, measure = "confidence")]

reglas_redundantes

```

## 10.2 Resultados del Análisis en R

### Anexo 1: Contenido de los datasets

```
> summary(campañas1)
DESCRIPTION          CAMPAIGN          START_DAY          END_DAY
Length:30            Min.   : 1.00        Min.   :224.0       Min.   :264.0
Class :character     1st Qu.: 8.25       1st Qu.:360.0       1st Qu.:405.8
Mode  :character     Median :15.50       Median :470.0       Median :502.0
                          Mean  :15.50       Mean  :463.9        Mean  :510.5
                          3rd Qu.:22.75     3rd Qu.:584.0      3rd Qu.:640.2
                          Max.  :30.00       Max.  :659.0        Max.  :719.0

> summary(campañas2)
DESCRIPTION          household_key          CAMPAIGN
Length:7208          Min.   : 1           Min.   : 1.00
Class :character     1st Qu.: 644         1st Qu.: 9.00
Mode  :character     Median :1263         Median :14.00
                          Mean  :1269         Mean  :15.66
                          3rd Qu.:1924       3rd Qu.:20.00
                          Max.  :2500         Max.  :30.00

> summary(causa1)
PRODUCT_ID          STORE_ID          WEEK_NO          display
Min.   : 26190      Min.   : 286       Min.   : 9.0      Length:36786524
1st Qu.: 928850     1st Qu.: 329       1st Qu.: 32.0     Class :character
Median :1312148     Median : 369       Median : 56.0     Mode  :character
Mean  : 3512237     Mean  : 3234       Mean  : 55.3
3rd Qu.: 5730125     3rd Qu.: 421       3rd Qu.: 78.0
Max.  :18244453     Max.  :34280       Max.  :101.0

mailler
Length:36786524
Class :character
Mode  :character

> summary(cupones_canjeados)
household_key      DAY          COUPON_UPC          CAMPAIGN
Min.   : 1         Min.   :225.0       Min.   :1.000e+10   Min.   : 1.00
1st Qu.: 588       1st Qu.:458.2     1st Qu.:1.000e+10   1st Qu.:13.00
Median :1396       Median :532.0     Median :5.234e+10   Median :14.00
Mean  :1303        Mean  :528.2      Mean  :4.123e+10    Mean  :15.55
3rd Qu.:2004       3rd Qu.:605.0    3rd Qu.:5.430e+10   3rd Qu.:18.00
Max.  :2500        Max.  :704.0      Max.  :5.898e+10    Max.  :30.00

> summary(cupones)
COUPON_UPC          PRODUCT_ID          CAMPAIGN
Min.   :1.000e+10   Min.   : 25671      Min.   : 1.00
1st Qu.:1.000e+10   1st Qu.: 959303    1st Qu.:13.00
Median :1.000e+10   Median :1411451    Median :14.00
Mean  :2.198e+10    Mean  :4651277     Mean  :15.86
3rd Qu.:5.111e+10   3rd Qu.:8357538   3rd Qu.:18.00
Max.  :5.999e+10    Max.  :18148540    Max.  :30.00

> summary(demografia)
AGE_DESC          MARITAL_STATUS_CODE INCOME_DESC          HOMEOWNER_DESC
Length:801        Length:801          Length:801          Length:801
Class :character   Class :character     Class :character     Class :character
Mode  :character   Mode  :character     Mode  :character     Mode  :character

HH_COMP_DESC      HOUSEHOLD_SIZE_DESC KID_CATEGORY_DESC   household_key
Length:801        Length:801          Length:801          Min.   : 1
Class :character   Class :character     Class :character     1st Qu.: 596
Mode  :character   Mode  :character     Mode  :character     Median :1218
                          Mean  :1235
                          3rd Qu.:1914
                          Max.  :2499

> summary(productos)
PRODUCT_ID          MANUFACTURER          DEPARTMENT          BRAND
Min.   : 25671      Min.   : 1           Length:92353        Length:92353
1st Qu.: 970628     1st Qu.: 328         Class :character     Class :character
Median :1621091     Median :1094         Mode  :character     Mode  :character
Mean  : 5328353     Mean  :1739
3rd Qu.: 9704770     3rd Qu.:2264
Max.  :18316298     Max.  :6477

COMMODITY_DESC      SUB_COMMODITY_DESC    CURR_SIZE_OF_PRODUCT
Length:92353        Length:92353        Length:92353
Class :character     Class :character     Class :character
Mode  :character     Mode  :character     Mode  :character
```

```
> summary(datos_transacciones)
household_key   BASKET_ID         DAY         PRODUCT_ID
Min.   : 1   Min.   :2.698e+10   Min.   : 1.0   Min.   : 25671
1st Qu.: 656   1st Qu.:3.041e+10   1st Qu.:229.0   1st Qu.: 917459
Median :1272   Median :3.276e+10   Median :390.0   Median : 1028816
Mean   :1272   Mean   :3.403e+10   Mean   :388.8   Mean   : 2891435
3rd Qu.:1913   3rd Qu.:4.013e+10   3rd Qu.:553.0   3rd Qu.: 1133018
Max.   :2500   Max.   :4.231e+10   Max.   :711.0   Max.   :18316298

QUANTITY        SALES_VALUE        STORE_ID        RETAIL_DISC
Min.   : 0.0   Min.   : 0.000   Min.   : 1   Min.   :-180.0000
1st Qu.: 1.0   1st Qu.: 1.290   1st Qu.: 330   1st Qu.: -0.6900
Median : 1.0   Median : 2.000   Median : 372   Median : -0.0100
Mean   : 100.4   Mean   : 3.104   Mean   : 3143   Mean   : -0.5387
3rd Qu.: 1.0   3rd Qu.: 3.490   3rd Qu.: 422   3rd Qu.: 0.0000
Max.   :89638.0   Max.   :840.000   Max.   :34280   Max.   : 3.9900

TRANS_TIME      WEEK_NO          COUPON_DISC      COUPON_MATCH_DISC
Min.   : 0   Min.   : 1.00   Min.   :-55.93000   Min.   :-7.700000
1st Qu.:1308   1st Qu.: 33.00   1st Qu.: 0.00000   1st Qu.: 0.000000
Median :1613   Median : 56.00   Median : 0.00000   Median : 0.000000
Mean   :1562   Mean   : 56.22   Mean   :-0.01642   Mean   :-0.002919
3rd Qu.:1843   3rd Qu.: 80.00   3rd Qu.: 0.00000   3rd Qu.: 0.000000
Max.   :2359   Max.   :102.00   Max.   : 0.00000   Max.   : 0.000000
```

Anexo 2: Top 15 productos más populares

```

          PRODUCT_ID MANUFACTURER DEPARTMENT BRAND COMMODITY_DESC SUB_COMMODITY_DESC
6678      826249      69 GROCERY Private BAKED BREAD/BUNS/ROLLS HAMBURGER BUNS
8282      840361      69 GROCERY Private EGGS EGGS - LARGE
13160     883404      69 GROCERY Private BAKED BREAD/BUNS/ROLLS MAINSTREAM WHITE BREAD
16016     908531      69 GROCERY Private FLUID MILK PRODUCTS CHOCOLATE MILK
20974     951590     910 GROCERY National BAKED BREAD/BUNS/ROLLS MAINSTREAM WHITE BREAD
24251     981760      69 GROCERY Private EGGS EGGS - X-LARGE
25755     995242      69 GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
25814     995785      2 PRODUCE National PEPPERS-ALL PEPPERS GREEN BELL
29658     1029743     69 GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
35577     1082185      2 PRODUCE National TROPICAL FRUIT BANANAS
37360     1098066     69 GROCERY Private BAKED BREAD/BUNS/ROLLS HOT DOG BUNS
38263     1106523     69 GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
40601     1127831     5937 PRODUCE National BERRIES STRAWBERRIES
41176     1133018     69 GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
57222     6534178     69 KIOSK-GAS Private COUPON/MISC ITEMS GASOLINE-REG UNLEADED

# A tibble: 15 x 2
  PRODUCT_ID Frecuencia
  <int> <int>
1 1082185 29778
2 6534178 19820
3 1029743 14430
4 995242 12542
5 1106523 9826
6 981760 9126
7 1133018 7315
8 883404 6904
9 951590 6138
10 826249 5505
11 840361 5338
12 1127831 5271
13 995785 5174
14 1098066 5040
15 908531 5028

          CURR_SIZE_OF_PRODUCT
6678      12 OZ
8282      1 DZ
13160     20 OZ
16016     20 OZ
20974     1 DZ
24251     1 DZ
25755     1 DZ
25814     48-54 CT
29658     1 GA
35577     40 LB
37360     11 OZ
38263     1 GA
40601     16 OZ
41176     1 GA
57222     1 GA

```

Anexo 3: Número de establecimientos únicos

```
> print(num_establecimientos)
[1] 582
```

Anexo 4: Número de productos únicos

```
> print(num_productos)
[1] 92353
```

Anexo 5: Número de transacciones únicas

```
> print(num_transacciones_unicas)
[1] 276484
```

Anexo 6: Transacciones por establecimiento

```
> transacciones_por_establecimiento
# A tibble: 582 × 2
  STORE_ID num_transacciones
  <int>     <int>
1     367     75573
2     406     59331
3     356     49332
4     381     45268
5     292     45240
6     343     41348
7     361     41101
8    31782     40949
9    32004     40064
10     427     39373
# i 572 more rows
```

Anexo 7: Establecimiento más concurrido

```
> print(establecimiento_mas_transacciones)
[1] 367
> print(num_transacciones_max)
[1] 75573
```

Anexo 8: Alcance de los datos (número de familias con transacciones)

```
> length(trans_data_demo_data_list)
[1] 801
```

Anexo 9: Número de transacciones en las que está incluido cada artículo

```
> print(transaction_counts)
# A tibble: 92,339 × 2
  PRODUCT_ID transaction_count
  <int>     <int>
1    1082185     29778
2    6534178     19820
3    1029743     14430
4     995242     12542
5    1106523     9826
6     981760     9126
7    1133018     7315
8     883404     6904
9     951590     6138
10    826249     5505
# i 92,329 more rows
```

Anexo 10: Distribución de la duración de las 30 campañas de marketing

```
> summary(campañas1$duracion)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  32.0   32.0   37.5   46.6   54.5  161.0
```

Anexo 11: Gasto medio de los clientes

```
> print(valor_medio_cesta)
[1] 29.1426
```

Anexo 12: Distribución de los cupones

```
> print(Use_coupon)
# A tibble: 2 × 6
  Use_coupon SALES_MEAN COUPON_DISC_MEAN COUPON_MATCH_DISC_MEAN Nb_coupon %_coupon
  <lg1>      <dbl>      <dbl>      <dbl>      <dbl>      <int>      <dbl>
1 FALSE      26.7         0           0           0         259733     93.9
2 TRUE       67.6        -2.54       -0.452      16751      6.06
```

Anexo 13: Frecuencia de compra media de los clientes

```
> print(frecuencia_media)
[1] 15.5
```

Anexo 14: Dataset de transacciones

```
> colnames(transacciones)[1:5]
[1] "1000002" "1000029" "1000050" "1000057" "1000059"
> rownames(transacciones)[1:5]
[1] "26984851472" "26984851516" "26984896261" "26984905972" "26984945254"
>
> inspect(transacciones[1:5])
  items                                transactionID
[1] {1004906, 1033142, 1036325, 1082185, 8160430} 26984851472
[2] {1043142, 1085983, 1102651, 6423775, 826249, 9487839} 26984851516
[3] {842930, 897044, 920955, 937406, 981760} 26984896261
[4] {1022843, 1048462, 1071333, 833715, 866950} 26984905972
[5] {1131351, 824399, 923972} 26984945254
```

Anexo 15: Tamaño de las compras

```
> summary(tamanyos)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000   5.000   9.388 12.000 168.000
> quantile(tamanyos, probs = seq(0,1,0.1))
 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
 1   1   1   2   3   5   7  10  15  24 168
```

Anexo 16: Artículos más frecuentes (términos absolutos y relativos)

```
> frecuencia_items <- itemFrequency(x = transacciones, type = "relative")
> frecuencia_items %>% sort(decreasing = TRUE) %>% head(5)
  1082185    6534178    1029743    995242    1106523
0.10770243 0.07168588 0.05219109 0.04536248 0.03553913
> frecuencia_items <- itemFrequency(x = transacciones, type = "absolute")
> frecuencia_items %>% sort(decreasing = TRUE) %>% head(5)
1082185 6534178 1029743 995242 1106523
  29778    19820    14430    12542    9826
```

```
> print(top5)
  PRODUCT_ID MANUFACTURER DEPARTMENT  BRAND  COMMODITY_DESC  SUB_COMMODITY_DESC
25755      995242           69  GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
29658      1029743           69  GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
35577      1082185            2  PRODUCE National TROPICAL FRUIT BANANAS
38263      1106523           69  GROCERY Private FLUID MILK PRODUCTS FLUID MILK WHITE ONLY
57222      6534178           69  KIOSK-GAS Private COUPON/MISC ITEMS GASOLINE-REG UNLEADED
  CURR_SIZE_OF_PRODUCT
25755
29658           1 GA
35577           40 LB
38263           1 GA
57222
```

Anexo 17: Itemsets comprados al menos 75 veces

```
> summary(itemsets)
set of 15206 itemsets

most frequent items:
1082185 1029743 995242 981760 1106523 (other)
  2327    598    385    357    317    20088

element (itemset/transaction) length distribution:sizes
  1  2  3  4
7311 6928 963 4

  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  1.000  1.000  2.000  1.583  2.000  4.000

summary of quality measures:
  support      count
Min.   :0.0002713  Min.   : 75.0
1st Qu.:0.0003255  1st Qu.: 90.0
Median :0.0004268  Median : 118.0
Mean   :0.0007148  Mean   : 197.6
3rd Qu.:0.0006655  3rd Qu.: 184.0
Max.   :0.1077024  Max.   :29778.0

includes transaction ID lists: FALSE

mining info:
  data ntransactions  support confidence
transacciones      276484 0.0002712634      1

apriori(data = transacciones, parameter = list(support = soporte, minlen = 1, maxlen = 20, target = "frequent itemset"))
```

Anexo 18: Reglas de asociación

```

> summary(reglas)
set of 11 rules

rule length distribution (lhs + rhs):sizes
2 3
4 7

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000  2.000   3.000   2.636   3.000   3.000

summary of quality measures:
  support      confidence      coverage      lift      count
Min.   :0.0002713  Min.   :0.7034  Min.   :0.0002930  Min.   :  6.595  Min.   : 75.0
1st Qu.:0.0002966  1st Qu.:0.7095  1st Qu.:0.0003689  1st Qu.:  7.772  1st Qu.: 82.0
Median :0.0003110  Median :0.7570  Median :0.0004268  Median :  8.597  Median : 86.0
Mean   :0.0003854  Mean   :0.7934  Mean   :0.0004830  Mean   :264.230  Mean   :106.5
3rd Qu.:0.0003490  3rd Qu.:0.8963  3rd Qu.:0.0004919  3rd Qu.:384.875  3rd Qu.: 96.5
Max.   :0.0010814  Max.   :0.9259  Max.   :0.0011936  Max.   :1360.650  Max.   :299.0

mining info:
  data ntransactions      support confidence
transacciones      276484 0.0002712634      0.7

call
apriori(data = transacciones, parameter = list(support = soporte, confidence = 0.7, target = "rules"))

> inspect(sort(x = reglas, decreasing = TRUE, by = "confidence"))
  lhs      rhs      support      confidence      coverage      lift      count
[1] {949257} => {1082185} 0.0002712634 0.9259259 0.0002929645 8.597075 75
[2] {6632283, 892728} => {1082185} 0.0002712634 0.9259259 0.0002929645 8.597075 75
[3] {1021164} => {5978648} 0.0010814369 0.9060606 0.0011935591 87.377489 299
[4] {6632283, 951590} => {1082185} 0.0003110487 0.8865979 0.0003508340 8.231921 86
[5] {1029743, 6632283} => {1082185} 0.0003218993 0.7876106 0.0004087036 7.312839 89
[6] {5582789, 5586942} => {1082185} 0.0002929645 0.7570093 0.0003870025 7.028712 81
[7] {7466806, 8011397} => {8011282} 0.0003110487 0.7107438 0.0004376383 694.379114 86
[8] {1028816, 1106523} => {1082185} 0.0003725351 0.7103448 0.0005244426 6.595439 103
[9] {10285187} => {10285149} 0.0003255161 0.7086614 0.0004593394 1360.649606 90
[10] {1012587, 1098066} => {826249} 0.0003797688 0.7046980 0.0005389100 35.392864 105
[11] {1077133} => {894439} 0.0003001982 0.7033898 0.0004267878 682.372049 83

```

Anexo 19: Evaluación de las reglas

```

# A tibble: 6 × 8
  rules      support confidence coverage lift count coverage.1 fishersExactTest
<chr>      <dbl>      <dbl>      <dbl> <dbl> <int>      <dbl>      <dbl>
1 {949257} => {1082185} 0.000271 0.926 0.000293 8.60 75 0.000293 3.98e-65
2 {6632283,892728} => {1082185} 0.000271 0.926 0.000293 8.60 75 0.000293 3.98e-65
3 {1021164} => {5978648} 0.00108 0.906 0.00119 87.4 299 0.00119 0
4 {6632283,951590} => {1082185} 0.000311 0.887 0.000351 8.23 86 0.000351 1.53e-70
5 {1029743,6632283} => {1082185} 0.000322 0.788 0.000409 7.31 89 0.000409 9.66e-64
6 {5582789,5586942} => {1082185} 0.000293 0.757 0.000387 7.03 81 0.000387 1.05e-55

```

Anexo 20: Reglas maximales

```
> reglas_maximales <- reglas[is.maximal(reglas)]
> reglas_maximales
set of 11 rules
> inspect(reglas_maximales[1:10])
  lhs                rhs                support  confidence coverage  lift      count coverage
[1] {1021164}          => {5978648} 0.0010814369 0.9060606 0.0011935591 87.377489 299 0.0011935591
[2] {1077133}          => {894439} 0.0003001982 0.7033898 0.0004267878 682.372049 83 0.0004267878
[3] {949257}           => {1082185} 0.0002712634 0.9259259 0.0002929645 8.597075 75 0.0002929645
[4] {10285187}        => {10285149} 0.0003255161 0.7086614 0.0004593394 1360.649606 90 0.0004593394
[5] {7466806, 8011397} => {8011282} 0.0003110487 0.7107438 0.0004376383 694.379114 86 0.0004376383
[6] {6632283, 892728} => {1082185} 0.0002712634 0.9259259 0.0002929645 8.597075 75 0.0002929645
[7] {6632283, 951590} => {1082185} 0.0003110487 0.8865979 0.0003508340 8.231921 86 0.0003508340
[8] {1029743, 6632283} => {1082185} 0.0003218993 0.7876106 0.0004087036 7.312839 89 0.0004087036
[9] {5582789, 5586942} => {1082185} 0.0002929645 0.7570093 0.0003870025 7.028712 81 0.0003870025
[10] {1012587, 1098066} => {826249} 0.0003797688 0.7046980 0.0005389100 35.392864 105 0.0005389100
fishersExactTest
[1] 0.000000e+00
[2] 2.366703e-224
[3] 3.980518e-65
[4] 6.341051e-280
[5] 1.247517e-233
[6] 3.980518e-65
[7] 1.531781e-70
[8] 9.657633e-64
[9] 1.050199e-55
[10] 5.345535e-142
```

Anexo 21: Reglas redundantes

```
> reglas_redundantes
set of 0 rules
```