

# Descubriendo los Secretos del Carrito de Compras: Un Análisis de Canasta en un Minimarket.

Juan Pablo Mogollón Avaunza<sup>1</sup>, Juan Sebastián Hernández Ramírez<sup>1</sup>, Karen Rojas Giraldo<sup>1</sup>, Jhocel Duван Suescun Torres<sup>1</sup>

## Resumen

El análisis de canasta de mercado realizado para **Abarrotes Selectos** se centró en identificar patrones de compra recurrentes mediante los algoritmos **Apriori** y **FP-Growth**. Se analizó un conjunto de datos de ventas, logrando identificar combinaciones significativas de productos que los clientes suelen adquirir juntos. Los resultados revelaron que los productos de **limpieza**, como **detergentes** y **tissue**, mostraron una fuerte asociación en las transacciones, lo que sugiere que los clientes tienden a agrupar estos productos durante sus compras.

En cuanto a las métricas utilizadas para evaluar las reglas de asociación, se consideraron el **soporte**, que indica la proporción de transacciones que contienen un conjunto de productos; la **confianza**, que mide la probabilidad de que un producto sea adquirido dado que otro lo fue; y el **lift**, que determina la relación entre la ocurrencia conjunta de productos frente a su ocurrencia independiente. Adicionalmente, se emplearon **leverage** y **conviction** para evaluar la fuerza de las asociaciones.

Los resultados más relevantes incluyen la combinación de **detergentes** y **cuidado de prendas**, con una confianza del 59.47% y un lift de 3.97, lo que sugiere una oportunidad clara para desarrollar promociones conjuntas. Además, el patrón **detergentes** → **blanqueador** y otros productos de limpieza mostraron relaciones significativas que podrían ser aprovechadas en la estrategia de disposición de productos y promociones.

Se concluyó que, al reorganizar los productos en el minimarket y aprovechar estas asociaciones mediante promociones cruzadas y ofertas, se podrían mejorar tanto las ventas como la experiencia del cliente. Este enfoque basado en datos permitirá a **Abarrotes Selectos** optimizar su gestión de inventarios y maximizar el valor del ticket promedio.

## Introducción.

En el competitivo mundo del retail, comprender los hábitos de compra de los clientes es fundamental para diseñar estrategias que maximicen las ventas y mejoren la experiencia del consumidor. Abarrotes Selectos, una cadena de mini supermercados de bajo costo que ofrece productos de consumo inmediato como alimentos preparados, artículos para la preparación de comidas, productos de aseo personal y otros productos misceláneos, se encuentra en un proceso de expansión en un vecindario popular. A pesar de un crecimiento sostenido, la empresa sospecha que existe un potencial de ventas aún no explotado.

Con el fin de capitalizar estas oportunidades, se llevó a cabo un análisis detallado de las transacciones de compra mediante técnicas de aprendizaje no supervisado, centrándose específicamente en el análisis de canasta de mercado (Market Basket Analysis, MBA). Este enfoque permite identificar patrones de compra recurrentes y relaciones entre productos, lo que facilita la creación de estrategias de venta más efectivas y personalizadas.

El principal objetivo de este estudio es responder a las siguientes preguntas clave:

1. ¿Cuáles son las combinaciones de productos que los clientes suelen comprar juntos?

---

<sup>1</sup> Maestría de Inteligencia Analítica de Datos, Universidad de Los Andes

2. ¿Existen productos que, al ser promovidos conjuntamente, podrían aumentar el valor de las compras?
3. ¿Podemos reorganizar la disposición de productos en el minimarket para incentivar la compra conjunta de ciertos artículos?

El diseño del estudio estuvo motivado por los siguientes objetivos estratégicos:

- **Maximización de ingresos:** Identificar las combinaciones de productos más comprados para crear promociones y combos que incentiven compras adicionales (Gensollen, 2022; Efrat et al., 2020).
- **Optimización del layout del minimarket:** Mejorar la disposición de productos para facilitar el acceso a combinaciones de artículos frecuentemente comprados juntos, lo que podría incrementar las ventas por impulso (Atencio Manyari et al., 2022; Singha et al., 2024).
- **Mejora de la satisfacción del cliente:** Ofrecer una experiencia de compra más fluida y conveniente al anticipar sus necesidades y preferencias, aumentando así la fidelización (Kaur & Singh, 2017; Rao et al., 2023).

El análisis de canasta de mercado ha sido ampliamente utilizado en diferentes contextos, como supermercados y tiendas de conveniencia, así como en sectores más especializados como la industria de la salud. Gensollen (2022) exploró el uso de Big Data y técnicas de segmentación de clientes en una cadena de supermercados en Europa, aplicando machine learning para personalizar las recomendaciones de productos y mejorar la experiencia del cliente. Atencio Manyari et al. (2022) propusieron un enfoque similar para una cadena de supermercados en Perú, utilizando el algoritmo Apriori y técnicas de clustering para identificar patrones de compra y segmentar clientes de manera efectiva.

Efrat et al. (2020) utilizaron el algoritmo Apriori, uno de los métodos más populares en el análisis de canasta, para descubrir asociaciones entre productos en un supermercado. Por otro lado, Hernández y Villalobos (2021) emplearon mapas auto-organizados (SOM) para visualizar estos patrones, proporcionando una representación intuitiva y detallada de las relaciones entre productos. Mustakim et al. (2018) compararon Apriori con el algoritmo FP-Growth en un contexto similar, encontrando que este último es más eficiente para grandes volúmenes de datos. En el sector salud, Rao et al. (2023) demostraron cómo el análisis de canasta puede optimizar la gestión de inventarios y mejorar el servicio al paciente al identificar patrones de consumo de medicamentos.

En general, el análisis de canasta permite identificar patrones ocultos en grandes volúmenes de datos (Efrat et al., 2020) y personalizar ofertas y promociones (Gensollen, 2022; Atencio Manyari et al., 2022). No obstante, presenta ciertos desafíos, como la necesidad de un procesamiento computacional intensivo y la posibilidad de generar patrones triviales si no se realiza un adecuado preprocesamiento de los datos (Mustakim et al., 2018).

En resumen, el análisis de canasta de mercado es una herramienta poderosa para la identificación de patrones de consumo, con aplicaciones que varían según el contexto y los objetivos específicos del negocio. Este estudio busca aplicar estas técnicas para maximizar el potencial de ventas de Abarrotes Selectos, mediante estrategias basadas en datos que respondan a las necesidades y comportamientos de sus clientes.

## **Materiales y métodos**

### ***Descripción de los datos***

#### **Origen de los datos**

Se obtuvo el registro de ventas del minimarket, que contiene información detallada sobre las transacciones realizadas.

- fecha: Indica la fecha en que se realizó la venta, en formato AAAA-MM-DD.
- codigo\_factura: Código identificador de la factura asociada a la venta.
- mnt\_total\_a\_pagar: Monto total a pagar por la factura, en la moneda local.
- cant\_items: Número total de ítems comprados en la transacción.
- familia: Categoría principal del producto (ej. Refrigerados, Alimentos Dulces).
- categoria: Categoría más específica dentro de la familia de productos (ej. Comida Preparada, Dulce).
- subcategoria: Subcategoría del producto, que ofrece más detalle (ej. Ready To Eat, Típicos).
- codigo\_producto: Código único que identifica el producto vendido.
- cantidad\_vendida: Cantidad de unidades vendidas del producto en la transacción.
- venta\_bruta\_producto: Valor bruto de la venta del producto antes de aplicar descuentos.
- venta\_neta\_producto: Valor neto de la venta del producto después de aplicar descuentos.

La base de datos parece está organizada a nivel de detalle de cada producto vendido en una factura, con varias filas que comparten el mismo codigo\_factura pero con diferentes productos, y contiene información tanto de la venta bruta como de la venta neta de cada producto. La clasificación de productos se organiza en tres niveles: familia, categoría y subcategoría, lo cual facilita la segmentación y análisis detallado de los productos vendidos.

Tabla 1. Descripción general de los datos.

Descripción	Valor
Número total de registros	152,781
Número de transacciones	39,123
Número de productos únicos	1,086
Número de familias únicas	12
Número de categorías únicas	31
Número de subcategorias únicas	185
Promedio de productos por transacción	3.9

Fuente: Elaboración propia

Tabla 2. Frecuencia de productos. Top 10.

Código producto	Frecuencia	Frecuencia relativa %
Q	3,971	2.60%
B2	3,710	2.43%
M	3,215	2.10%
W	2,825	1.85%
N38	2,786	1.82%
B	2,183	1.43%
S3	1,937	1.27%
C	1,806	1.18%
T4	1,373	0.90%
Y8	1,364	0.89%
Otros	127,611	83.53%

Fuente: Elaboración propia

Se registraron un total de 152,781 registros de ventas, distribuidos en 39,123 transacciones únicas. Esto indica un conjunto de datos amplio y detallado, en el que cada registro representa un producto individual vendido en una transacción específica.

El minimarket ofrece una gran variedad de productos, con 1,087 productos únicos disponibles, organizados en 12 familias, 31 categorías y 185 subcategorías. Esta estructura jerárquica muestra la diversidad del inventario y permite una clasificación detallada que puede ser útil para análisis más específicos, como segmentación de productos y análisis de ventas.

Finalmente, el promedio de 3.9 productos por transacción sugiere que los clientes suelen comprar una pequeña variedad de productos en cada visita, lo que podría indicar un comportamiento de compra típico de minimarkets, donde las compras suelen ser rápidas y enfocadas en productos específicos. Este dato es relevante para diseñar estrategias de marketing y disposición de productos en la tienda.

La tabla 2 muestra la frecuencia de los 10 productos más vendidos en el minimarket, así como su proporción relativa respecto al total de ventas. Los productos listados representan una pequeña fracción del total de productos vendidos, destacándose el producto identificado como "Q", con 3,971 ventas, que corresponde al 2.60% de todas las transacciones.

Los demás productos en el top 10 tienen frecuencias menores pero relativamente similares, oscilando entre el 2.43% (producto "B2") y el 0.89% (producto "Y8"). Esto sugiere que, aunque hay una leve concentración de ventas en algunos productos específicos, la demanda está distribuida de manera relativamente equilibrada entre estos productos.

Finalmente, la categoría "Otros" concentra la gran mayoría de las ventas con el 83.53% del total. Esto indica que el minimarket tiene un portafolio muy diversificado de productos, y que fuera de este top 10, una amplia variedad de productos también es relevante en términos de volumen de ventas, lo cual es un indicio de una oferta amplia y variada que satisface diferentes necesidades de los clientes.

Tabla 3. Frecuencia de subcategorías. Top 10.

Subcategoría	Frecuencia	Frecuencia relativa %
Tissue	10,622	6.95%
Otros	7,355	4.81%
Detergentes	6,387	4.18%
Empaques	5,627	3.68%
Leches Larga Vida	4,579	3.00%
Jabonería	4,115	2.69%
Oral	3,689	2.41%
Cuidado De Prendas	3,659	2.39%
Lavaplatos	3,476	2.28%
Familiar	3,452	2.26%
Otros	99,820	65.34%

Fuente: Elaboración propia

La tabla 3 muestra las 10 subcategorías más vendidas en el minimarket, destacando la subcategoría "Tissue" como la más popular con 10,622 registros de ventas, lo que representa el 6.95% del total. Esta subcategoría incluye productos de papel, como papel higiénico y toallas de papel, que son productos básicos de alta demanda.

Las siguientes subcategorías en la lista, como "Otros" (4.81%), "Detergentes" (4.18%) y "Empaques" (3.68%), también reflejan productos de uso cotidiano, sugiriendo que los clientes del minimarket tienden a adquirir productos de primera necesidad y limpieza en sus compras habituales.

A pesar de la concentración en estas subcategorías, la mayoría de las ventas (65.34%) se agrupan bajo "Otros", lo que indica una diversidad significativa en la oferta del minimarket. Esto sugiere que, aunque algunas subcategorías destacan, existe una amplia variedad de productos que contribuyen a las ventas totales, lo que puede ser indicativo de un surtido amplio y diverso para atender diferentes necesidades de los clientes.

Variables consideradas

Se eligieron únicamente las variables **codigo\_factura** y **subcategoria** para el análisis de canasta, lo que permitió simplificar el modelo de datos para identificar las asociaciones de compra más relevantes a un nivel más agregado.. La justificación de esta selección es la siguiente:

**1. Identificador de Transacción (**codigo\_factura**):**

- Esta variable es crucial porque agrupa los productos que se compraron juntos en una misma transacción. Es el identificador único que permite realizar el análisis de co-ocurrencia y detectar patrones de compra conjuntos. Sin esta variable, no sería posible saber qué productos se adquirieron en la misma compra.

**2. Subcategoría del Producto (**subcategoria**):**

- La subcategoría proporciona una visión más general que el código de producto, permitiendo identificar patrones de compra entre grupos de productos relacionados. En lugar de analizar la asociación a nivel de producto específico, se busca entender la relación entre tipos de productos, lo que puede ser más útil para decisiones de marketing y gestión de inventario. Por ejemplo, si se encuentra que las subcategorías "Bebidas Energéticas" y "Snacks" se compran frecuentemente juntas, se puede promover su venta conjunta o ajustar la distribución en la tienda para facilitar la compra combinada.

Beneficios de esta selección:

- **Reducción de la complejidad:** Al trabajar con subcategorías en lugar de productos específicos, se reduce el número de combinaciones posibles, facilitando la interpretación de los resultados.
- **Identificación de tendencias generales:** Es más probable identificar tendencias generales de comportamiento de compra que se mantengan a lo largo del tiempo, en lugar de asociaciones que pueden depender de productos específicos y ser más volátiles.
- **Aplicación práctica:** Los resultados a nivel de subcategoría pueden ser más accionables para estrategias de marketing, promociones cruzadas y disposición de productos en la tienda.

Esta simplificación permitió un análisis más claro y manejable, evitando el ruido generado al considerar productos específicos, y se centró en la co-ocurrencia de grupos de productos, lo que pudo proporcionar insights estratégicos valiosos.

**Preprocesamiento de datos**

Para llevar a cabo el análisis de canasta, se transformaron estos datos en una matriz de transacciones y productos, donde cada fila representa una transacción y cada columna un producto, con valores binarios indicando si el producto fue o no comprado en esa transacción. Este formato fue utilizado para aplicar los algoritmos como Apriori y FP-Growth para encontrar reglas de asociación, tales como:

- **Soporte:** La proporción de transacciones que contienen una combinación de productos.
- **Confianza:** La probabilidad de que un producto sea comprado dado que se compró otro producto.
- **Lift:** Medida que evalúa si la presencia de un producto aumenta o disminuye la probabilidad de que se compre otro producto.

La base de datos original no tenía datos duplicados, faltantes, incompletos o inconsistentes, por lo cual no se requirió realizar ningún tratamiento adicional al descrito.

La figura 1 muestra la matriz de datos binaria utilizada en el análisis de canasta de productos. Esta matriz es esencial para realizar el análisis de asociación, ya que facilita la identificación de patrones de compra, como productos que se suelen vender juntos. Cada fila representa una transacción única, mientras que cada

columna representa una subcategoría de producto, como "Aceites", "Aderezos", "Alimento Seco", entre otras. Los valores **True** y **False** indican la presencia o ausencia de una subcategoría de producto en una transacción específica.

En total, la matriz tiene 39,123 filas y 185 columnas, lo que corresponde al número total de transacciones y subcategorías diferentes en el conjunto de datos, respectivamente. La estructura de esta matriz permite identificar combinaciones de productos comprados juntos en diferentes transacciones. Por ejemplo, en la última fila, los valores **True** en las columnas "Alimento Humedo" y "Alimento Seco" indican que esa transacción incluyó ambos tipos de productos.

Figura 1. Muestra de la base de datos utilizada para el análisis de canasta

	Aceites	Aceites Finos	Aderezos	Afeitado	Aguas	Aguas Saborizadas	Ají	Alimento Humedo	Alimento Seco	Analgésicos	...
0	False	False	False	False	False	False	False	False	False	False	...
1	False	False	False	False	False	False	False	False	False	False	...
2	False	False	False	False	False	False	False	False	False	False	...
3	False	False	False	False	False	False	False	False	False	False	...
4	False	False	False	False	False	False	False	False	False	False	...
...	...	...	...	...	...	...	...	...	...	...	...
39118	False	False	False	False	False	False	False	False	False	False	...
39119	False	False	False	False	False	False	False	False	False	False	...
39120	False	False	False	False	False	False	False	False	False	False	...
39121	False	False	False	False	False	False	False	False	False	False	...
39122	False	False	False	False	False	False	False	True	True	False	...
39123 rows × 185 columns											

Fuente: Elaboración propia

Métodos de análisis

Algoritmos de Asociación Utilizados: Apriori y FP-Growth

Los algoritmos Apriori y FP-Growth son métodos populares para identificar patrones de asociación en bases de datos transaccionales, como registros de ventas. Estos algoritmos permiten descubrir conjuntos de productos que se compran frecuentemente juntos y derivar reglas de asociación a partir de estos conjuntos.

**Algoritmo Apriori.** El algoritmo Apriori, introducido por Agrawal y Srikant en 1994, se basa en la idea de que cualquier subconjunto de un conjunto frecuente también debe ser frecuente. Funciona de la siguiente manera:

- Generación de Itemsets Frecuentes:

- Identifica todos los conjuntos de productos (itemsets) que cumplen con un soporte mínimo predefinido. El soporte es la proporción de transacciones en la base de datos que contiene un conjunto de productos específicos.
- A partir de los itemsets frecuentes de tamaño 1, se generan iterativamente itemsets de mayor tamaño (k-itemsets), donde k aumenta en cada iteración.
- Se eliminan los itemsets que no cumplen con el soporte mínimo, reduciendo así el número de posibles combinaciones.
- **Generación de Reglas de Asociación:**
  - A partir de los itemsets frecuentes, se generan reglas de la forma  $A \rightarrow B$ , donde A y B son conjuntos de productos.
  - Para cada regla, se calculan métricas de calidad como la confianza y el lift. La confianza mide cuán frecuentemente el conjunto B se compra cuando el conjunto A está presente. El lift mide cuán más probable es la co-ocurrencia de A y B en comparación con su ocurrencia esperada si fueran independientes.

**Algoritmo FP-Growth (Frequent Pattern Growth).** El algoritmo FP-Growth, propuesto por Han et al. en el año 2000, mejora la eficiencia del Apriori al evitar la generación de candidatos:

- **Construcción del Árbol FP (Frequent Pattern Tree):**
  - Se construye un árbol FP a partir de las transacciones de la base de datos, donde cada nodo representa un ítem y su conteo de ocurrencia.
  - Las transacciones se ordenan y se insertan en el árbol de manera que los ítems más frecuentes están más cercanos a la raíz.
- **Extracción de Patrones Frecuentes:**
  - A partir del árbol FP, se extraen directamente los itemsets frecuentes utilizando un proceso de minería de patrones condicionales, evitando la necesidad de generar todos los candidatos explícitamente.
- **Generación de Reglas de Asociación:**
  - Similar al Apriori, se generan reglas de asociación y se calculan métricas como la confianza y el lift.

**Criterios para Seleccionar las Reglas de Asociación.** Las reglas de asociación se seleccionan y filtran basándose en las siguientes métricas de calidad:

- **Soporte:**
  - Proporción de transacciones que contienen un determinado conjunto de ítems. Un soporte bajo puede indicar una regla poco relevante.
- **Confianza:**
  - Proporción de transacciones que contienen el conjunto antecedente A que también contienen el conjunto consecuente B. Se seleccionan reglas con una confianza alta para asegurar que el consecuente es frecuente dado el antecedente.
- **Lift:**
  - Mide la fuerza de la co-ocurrencia de A y B en comparación con la ocurrencia independiente de ambos conjuntos. Un lift mayor a 1 indica una asociación positiva, es decir, que la ocurrencia de AAA incrementa la probabilidad de B.
- **Longitud del Itemset:**
  - Se seleccionaron reglas con longitud de itemsets de 2 productos para capturar asociaciones simples y efectivas que sean fácilmente accionables y comprensibles. Esta elección permite identificar relaciones de compra claras que pueden aplicarse en estrategias promocionales directas, como la creación de combos y la disposición conjunta de productos en el minimarket.

Además, facilita la implementación de cambios rápidos y eficientes en el layout y en campañas de marketing, adaptándose a las necesidades operativas de Abarrotes Selectos.

## Herramientas y Software Utilizado

- **Python:**
  - El lenguaje principal utilizado para la implementación de los algoritmos y análisis.
- **Librerías Específicas:**
  - **pandas:** Para la manipulación y análisis de datos, incluyendo la carga y transformación de la base de datos transaccional.
  - **mlxtend:** Librería que incluye implementaciones de los algoritmos Apriori y FP-Growth, así como funciones para la generación de reglas de asociación.
  - **matplotlib:** Para la visualización de los resultados y gráficas de soporte, confianza y lift.
- **Jupyter Notebook:**
  - Entorno utilizado para la implementación y exploración interactiva de los algoritmos, permitiendo la integración de código, resultados y visualizaciones.

## Evaluación del Modelo de Reglas de Asociación

Para evaluar la calidad y el éxito de las reglas de asociación generadas, se utilizan diferentes métricas e indicadores. Estas métricas permiten determinar la relevancia y utilidad de las reglas descubiertas en el contexto de análisis. Los métodos principales para evaluar la calidad de las reglas de asociación son el soporte, la confianza, el lift, leverage, conviction y la métrica de Zhang.

### Métricas para Evaluar la Calidad de las Reglas de Asociación

**Soporte (Support).** El soporte de una regla de asociación mide la frecuencia con la que un conjunto de productos aparece en la base de datos transaccional. Se calcula como la proporción de transacciones que contienen tanto el antecedente como el consecuente de la regla. Esta métrica ayuda a identificar qué tan populares son ciertas combinaciones de productos (Tan et al., 2005).

- **Fórmula:**  $\text{Soporte}(A \rightarrow B) = \frac{\text{Número de transacciones que contienen A y B}}{\text{Número total de transacciones}}$
- **Interpretación:** Un soporte alto indica que la combinación de productos AAA y BBB es común en la base de datos. Sin embargo, un soporte demasiado bajo puede implicar que la regla no es relevante o representativa.

**Confianza (Confidence).** La confianza mide cuán frecuentemente el consecuente B ocurre en las transacciones que ya contienen el antecedente A. Es una medida condicional que indica la probabilidad de que B ocurra dado que A ha ocurrido (Han et al., 2011).

- **Fórmula:**  $\text{Confianza}(A \rightarrow B) = \frac{\text{Número de transacciones que contienen A y B}}{\text{Número de transacciones que contienen A}}$
- **Interpretación:** Una confianza alta sugiere que es muy probable que B se compre cuando A ya está en la canasta de compra. Esta métrica se utiliza para evaluar la precisión predictiva de la regla.

**Lift.** El lift mide la fuerza de la asociación entre el antecedente y el consecuente de una regla en comparación con la ocurrencia esperada de AAA y BBB de manera independiente. Es una medida que corrige la confianza



al tener en cuenta la frecuencia de ocurrencia independiente del antecedente y el consecuente (Hahsler & Hornik, 2007).

- **Fórmula:**  $\text{Lift}(A \rightarrow B) = \text{Confianza}(A \rightarrow B) / \text{Soporte}(B)$   
o también:  $\text{Lift}(A \rightarrow B) = \text{Soporte}(A \cup B) / (\text{Soporte}(A) \times \text{Soporte}(B))$
- **Interpretación:**
  - Un lift mayor a 1 indica una relación positiva: la presencia de A incrementa la probabilidad de que B ocurra.
  - Un lift igual a 1 sugiere independencia entre A y B.
  - Un lift menor a 1 implica que la presencia de A reduce la probabilidad de que B ocurra.

**Leverage (Apoyo Diferencial).** Leverage mide la diferencia entre la frecuencia observada de la co-ocurrencia de A y B y la frecuencia esperada si fueran independientes. Esta métrica ayuda a entender el grado en el que la co-ocurrencia observada difiere de lo que se esperaría al azar (Piatetsky-Shapiro, 1991).

- **Fórmula:**  $\text{Leverage}(A \rightarrow B) = \text{Soporte}(A \cup B) - \text{Soporte}(A) \times \text{Soporte}(B)$
- **Interpretación:** Un leverage positivo indica que la co-ocurrencia es más frecuente de lo esperado bajo independencia. Un leverage negativo indica lo contrario.

**Conviction.** Conviction mide la independencia de la regla y se interpreta como la razón entre la probabilidad de que A ocurra sin B y la probabilidad de que A ocurra con B. Refleja cuán fuerte es la dependencia de A en B (Brin et al., 1997)..

- **Fórmula:**  $\text{Conviction}(A \rightarrow B) = (1 - \text{Soporte}(B)) / (1 - \text{Confianza}(A \rightarrow B))$
- **Interpretación:** Un valor de conviction cercano a 1 sugiere que A y B son independientes. Valores mayores a 1 indican una dependencia positiva entre A y B.

**Métrica de Zhang (Zhang's Metric).** Zhang's metric mide la fuerza de la asociación y evita los problemas de sesgo que tienen otras métricas como la confianza. Se basa en la probabilidad de B dado A y la probabilidad de A dado B (Zhang & Zhang, 2002)..

- **Fórmula:**  $\text{Zhang}(A \rightarrow B) = (\text{Confianza}(A \rightarrow B) - \text{Soporte}(B)) / (\max(\text{Soporte}(A), \text{Soporte}(B)) - \text{Soporte}(A) \times \text{Soporte}(B))$
- **Interpretación:** Un valor de Zhang mayor a 0 indica una asociación positiva entre AAA y BBB, mientras que un valor menor a 0 indica una asociación negativa.

### Indicadores para Medir el Éxito del Análisis

Para evaluar el éxito del análisis de reglas de asociación, se deben considerar los siguientes aspectos:

- **Número de Reglas Relevantes:**
  - El éxito del análisis puede evaluarse por la cantidad de reglas útiles generadas. Se espera obtener un número significativo de reglas con alto soporte, confianza y lift que sean relevantes para la toma de decisiones.
- **Diversidad de Reglas:**
  - La variedad en los tipos de reglas obtenidas (diferentes combinaciones de productos) es un indicador del éxito del análisis. Una mayor diversidad de reglas indica que el modelo ha capturado distintas relaciones dentro de los datos.
- **Aplicabilidad Práctica:**

- La capacidad de las reglas para ser aplicadas en contextos reales, como la disposición de productos o promociones conjuntas, mide el éxito del análisis. Las reglas deben ser comprensibles y accionables para tener un impacto en la estrategia de negocio.

El uso de métricas como soporte, confianza, lift, leverage, conviction y la métrica de Zhang permite evaluar la calidad de las reglas de asociación desde múltiples perspectivas. Estos indicadores adicionales ayudan a medir el éxito del análisis en términos de aplicabilidad práctica y diversidad de reglas útiles. Estas evaluaciones son esenciales para asegurar que el análisis aporte valor al negocio y permita tomar decisiones basadas en datos.

## Resultados y discusión

### *Método Apriori.*

El análisis de canasta se realizó utilizando diferentes combinaciones de soporte mínimo y confianza mínima. Los resultados obtenidos se presentan en la tabla 4 que muestra:

- **Soporte Mínimo:** El umbral mínimo de la proporción de transacciones en las que una combinación de productos debe aparecer para ser considerada en la generación de reglas.
- **Confianza Mínima:** El umbral mínimo de la proporción de transacciones que contienen el antecedente y también contienen el consecuente, lo que mide la precisión de la regla.
- **Reglas Relevantes:** El número de reglas que cumplen con los umbrales establecidos de soporte y confianza.
- **Diversidad de Reglas:** El número de combinaciones únicas de productos involucrados en las reglas relevantes.

Tabla 4. Resultados de aplicación del método Apriori con diversos parámetros

soporte minimo	confianza minima	reglas relevantes	diversidad de reglas
0.05	0.15	4	2
0.02	0.15	85	42
0.02	0.2	71	40
0.02	0.25	57	36
0.03	0.25	18	15

Fuente: Elaboración propia

1. **Soporte 0.05 y Confianza 0.15: Reglas Relevantes: 4, Diversidad de Reglas: 2**
  - **Interpretación:** Un soporte relativamente alto (0.05) y una confianza baja (0.15) generan muy pocas reglas (4), con una baja diversidad (2). Esto indica que solo unas pocas combinaciones de productos se compran lo suficientemente juntas para ser consideradas relevantes bajo estos parámetros. Es posible que estas combinaciones sean las más populares, pero la escasa diversidad sugiere que las reglas identificadas no capturan una gran variedad de patrones.
2. **Soporte 0.02 y Confianza 0.15: Reglas Relevantes: 85, Diversidad de Reglas: 42**
  - **Interpretación:** Reducir el soporte mínimo a 0.02 permite que más combinaciones de productos sean consideradas relevantes, aumentando significativamente el número de reglas relevantes (85) y la diversidad de reglas (42). Esto sugiere que este umbral de soporte permite captar una mayor variedad de patrones de compra, manteniendo la confianza mínima relativamente baja.
3. **Soporte 0.02 y Confianza 0.20: Reglas Relevantes: 71, Diversidad de Reglas: 40**

- **Interpretación:** Aumentar la confianza mínima a 0.20 reduce ligeramente el número de reglas (71) y la diversidad (40), lo cual es esperable, ya que se está exigiendo una mayor precisión en las reglas generadas. No obstante, se sigue manteniendo un buen equilibrio entre cantidad y diversidad de reglas.
- 4. **Soporte 0.02 y Confianza 0.25: Reglas Relevantes: 57, Diversidad de Reglas: 36**
  - **Interpretación:** Aumentar aún más la confianza mínima a 0.25 reduce todavía más el número de reglas relevantes (57) y la diversidad (36). Aunque las reglas que cumplen con estos umbrales son más fuertes en términos de confianza, se pierden algunos patrones menos frecuentes.
- 5. **Soporte 0.03 y Confianza 0.25: Reglas Relevantes: 18, Diversidad de Reglas: 15**
  - **Interpretación:** Este caso presenta un soporte intermedio (0.03) y una confianza alta (0.25), resultando en solo 18 reglas relevantes y una diversidad baja (15). Se observa que aumentar tanto el soporte como la confianza elimina muchas combinaciones, resultando en reglas muy específicas pero limitadas en número y variedad.

Tras obtener estos resultados se eligió la configuración de **soporte mínimo de 0.03 y confianza mínima de 0.25** como la más adecuada para el contexto de Abarrotes Selectos. A continuación se presenta una justificación detallada de esta selección:

1. **Equilibrio entre Especificidad y Relevancia:**
  - El umbral de soporte de 0.03 significa que las combinaciones de productos seleccionadas están presentes en al menos el 3% de todas las transacciones. Este valor permite identificar asociaciones que, aunque no son extremadamente frecuentes, tienen una presencia significativa y potencialmente subexplotada en el comportamiento de compra.
  - La confianza de 0.25 implica que, en al menos el 25% de las transacciones donde se compra un producto A, también se compra el producto B. Este umbral asegura que las reglas seleccionadas tienen una precisión suficiente para ser útiles en estrategias promocionales.
2. **Identificación de Oportunidades de Venta Cruzada:**
  - A pesar de que solo se identificaron 18 reglas relevantes, estas presentan una **diversidad de 15 combinaciones de productos únicas**. Esto es crucial para Abarrotes Selectos, ya que permite diseñar promociones específicas y estrategias de disposición de productos que aprovechen estas asociaciones frecuentes pero no obvias.
  - Estas combinaciones pueden utilizarse para desarrollar estrategias de venta cruzada, como la creación de combos de productos o promociones conjuntas que incentiven compras adicionales, incrementando así el valor del ticket promedio (Gensollen, 2022; Efrat et al., 2020).
3. **Optimización del Layout del Minimarket:**
  - Con una menor cantidad de reglas, es más sencillo implementar cambios en la disposición de productos dentro de las tiendas. Al identificar las asociaciones más fuertes y frecuentes, se pueden reorganizar los productos para facilitar la compra conjunta de ciertos artículos, mejorando las ventas por impulso (Atencio Manyari et al., 2022; Singha et al., 2024).
4. **Estrategia de Diferenciación en el Vecindario:**
  - Al centrarse en reglas con mayor soporte y confianza, Abarrotes Selectos puede destacarse de la competencia en el vecindario al ofrecer una experiencia de compra más personalizada y orientada a las preferencias reales de los clientes. Esto mejora la satisfacción y la fidelización de los clientes, quienes perciben que la tienda anticipa sus necesidades (Kaur & Singh, 2017; Rao et al., 2023).

La tabla 5 muestra 18 reglas de asociación obtenidas del análisis de canasta de mercado para Abarrotes Selectos, considerando un soporte mínimo de 0.03 y una confianza mínima de 0.25. A continuación, se realiza un análisis detallado de las reglas presentadas, seguido de la interpretación de métricas clave como soporte, confianza, lift, leverage, conviction y la métrica de Zhang.

Las reglas identifican combinaciones de productos que los clientes suelen comprar juntos. Por ejemplo, una regla destacada es "(Detergentes) → (Cuidado De Prendas)" con un soporte de 0.0465, lo que indica que el 4.65% de las transacciones contienen ambos productos.

Tabla 5. Reglas de asociación obtenidas usando Apriori

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
3	(Detergentes)	(Cuidado De Prendas)	0.149861	0.078266	0.046546	0.310592	3.968414	0.034817	1.336993	0.879868
4	(Cuidado De Prendas)	(Detergentes)	0.078266	0.149861	0.046546	0.594709	3.968414	0.034817	2.097604	0.811525
0	(Detergentes)	(Blanqueador)	0.149861	0.085832	0.044705	0.298311	3.475533	0.031842	1.302812	0.837832
1	(Blanqueador)	(Detergentes)	0.085832	0.149861	0.044705	0.520846	3.475533	0.031842	1.774250	0.779150
8	(Pisos)	(Detergentes)	0.064693	0.149861	0.030391	0.469775	3.134743	0.020696	1.603355	0.728098
7	(Lavaplatos)	(Detergentes)	0.086599	0.149861	0.034123	0.394038	2.629360	0.021145	1.402958	0.678431
6	(Jabonería)	(Detergentes)	0.094088	0.149861	0.036781	0.390926	2.608598	0.022681	1.395791	0.680698
13	(Jabonería)	(Tissue)	0.094088	0.232855	0.052322	0.556099	2.388173	0.030413	1.728188	0.641640
17	(Pisos)	(Tissue)	0.064693	0.232855	0.034788	0.537732	2.309297	0.019724	1.659524	0.606184
14	(Lavaplatos)	(Tissue)	0.086599	0.232855	0.042788	0.494097	2.121904	0.022623	1.516386	0.578853
2	(Blanqueador)	(Tissue)	0.085832	0.232855	0.041101	0.478856	2.056455	0.021115	1.472041	0.561960
5	(Cuidado De Prendas)	(Tissue)	0.078266	0.232855	0.036321	0.464076	1.992979	0.018097	1.431442	0.540545
15	(Oral)	(Tissue)	0.073946	0.232855	0.033970	0.459385	1.972833	0.016751	1.419021	0.532490
12	(Femenina)	(Tissue)	0.068272	0.232855	0.030647	0.448896	1.927787	0.014749	1.392013	0.516535
10	(Tissue)	(Detergentes)	0.232855	0.149861	0.064182	0.275631	1.839249	0.029286	1.173628	0.594803
9	(Detergentes)	(Tissue)	0.149861	0.232855	0.064182	0.428279	1.839249	0.029286	1.341817	0.536735
11	(Empaques)	(Tissue)	0.142602	0.232855	0.046801	0.328195	1.409437	0.013596	1.141916	0.338812
16	(Otros)	(Tissue)	0.156915	0.232855	0.042635	0.271705	1.166842	0.006096	1.053344	0.169599

Fuente: Elaboración propia

- **Métricas Clave:**
  - **Soporte:** Varía entre 0.03 y 0.0465, indicando la proporción de transacciones que contienen tanto el antecedente como el consecuente.
  - **Confianza:** Oscila entre 0.2717 y 0.5947, indicando la probabilidad de que el consecuente se compre cuando el antecedente está presente.
  - **Lift:** Valores superiores a 1 en todas las reglas, con un rango de 1.67 a 3.97, sugiriendo que la ocurrencia conjunta de los productos es más probable que la ocurrencia esperada si fueran independientes.
  - **Leverage:** Varía de 0.0020 a 0.0468, mostrando la diferencia entre la ocurrencia observada y la esperada de las combinaciones de productos.
  - **Conviction:** Va de 1.05 a 2.09, indicando la fuerza de la dependencia entre los productos. Valores mayores a 1 sugieren una dependencia positiva.
  - **Métrica de Zhang:** Varía de 0.1696 a 0.8799, con valores positivos que indican una asociación positiva entre los productos.
- **Interpretación de las Reglas Más Relevantes**
  - **(Detergentes) → (Cuidado De Prendas):**

- **Confianza:** 0.5947 (59.47%): Alta probabilidad de que se compren juntos. La confianza más alta de todas las reglas.
- **Lift:** 3.9684: La probabilidad de que ambos productos se compren juntos es casi 4 veces mayor que si fueran independientes.
- **Conviction:** 2.0976: Refuerza que hay una fuerte relación de dependencia.
- **Interpretación:** Este patrón sugiere que los clientes que compran detergentes tienen una alta propensión a comprar también productos para el cuidado de prendas. Es una oportunidad clara para promociones conjuntas o disposición cercana en el layout del minimarket.
- **(Detergentes) → (Blanqueador):**
  - **Confianza:** 0.2983 (29.83%): Aproximadamente 3 de cada 10 compradores de detergentes también compran blanqueadores.
  - **Lift:** 3.4755: La compra conjunta es 3.47 veces más probable que la compra independiente.
  - **Interpretación:** Sugerir combos que incluyan detergente y blanqueador, o disponerlos juntos, podría incentivar la venta conjunta.
- **(Pisos) → (Detergentes):**
  - **Confianza:** 0.4698 (46.98%): Casi la mitad de los compradores de productos de limpieza de pisos también adquieren detergentes.
  - **Lift:** 3.1347: Asociación fuerte que indica que la compra de productos para pisos y detergentes es más común de lo esperado.
  - **Interpretación:** Promociones cruzadas o bundles de productos de limpieza para pisos y detergentes podrían aumentar las ventas.
- **(Jabonería) → (Detergentes):**
  - **Confianza:** 0.3909 (39.09%): Fuerte tendencia de compra conjunta.
  - **Lift:** 2.6086: Probabilidad de compra conjunta superior a la esperada.
  - **Interpretación:** Los clientes asocian la compra de jabones con la de detergentes, lo que sugiere una oportunidad para promociones cruzadas.
- **(Lavaplatos) → (Detergentes):**
  - **Confianza:** 0.3940 (39.40%): Alta probabilidad de compra conjunta.
  - **Lift:** 2.6294: Compra conjunta 2.63 veces más probable que si fueran independientes.
  - **Interpretación:** Los productos de limpieza de cocina se compran en conjunto con detergentes, lo que podría explotarse con promociones de "limpieza completa".
- **(Tissue) → (Detergentes):**
  - **Confianza:** 0.4282 (42.82%): Similar a la tendencia anterior, se observa una fuerte relación de compra conjunta.
  - **Lift:** 1.8394: Aunque es más bajo que las anteriores, sigue mostrando una asociación relevante.
  - **Interpretación:** Colocar estos productos en áreas de fácil acceso podría incrementar las ventas de impulso.

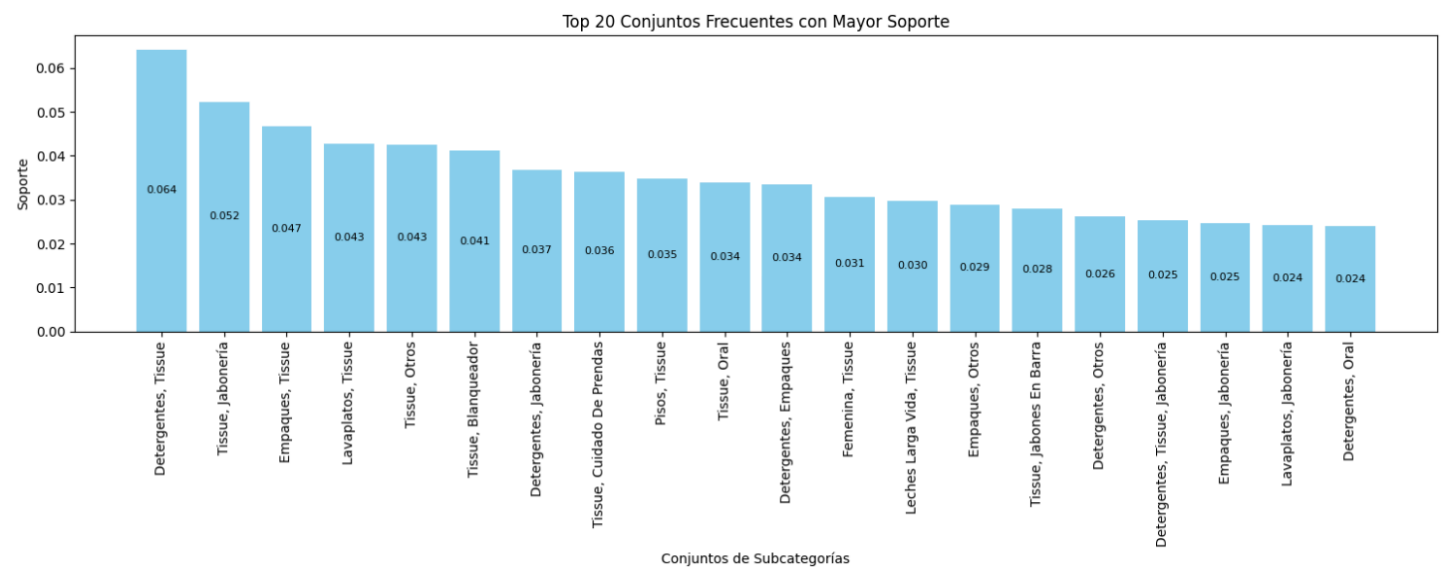
### ***Método FP-Growth.***

Para la implementación del algoritmo FP-Growth, se decidió establecer un soporte fijo de **0.02** y una **confianza mínima de 0.2**. Con base en estos valores y con el ánimo de obtener resultados intercategoriales e intrafamiliares, se estableció un condicional sobre los resultados filtrando solo para aquellas combinaciones que fueran de familias y categorías distintas, esto para poder hallar patrones no tan evidentes e intuitivos, pues es de esperarse que productos de una misma categoría se compren juntos frecuentemente. De igual

forma, como análisis complementario se decidió aplicar este algoritmo sobre la base de datos de ventas por familia, nuevamente con el fin de observar el comportamiento entre las distintas familias y capturar patrones de compra más generales y menos específicos como lo puede ser con granularidad de subcategoría. Por medio del soporte mínimo estipulado se obtienen 71 reglas de las cuales se resumen las 20 con mayor soporte.

Con base en lo previamente descrito se presentan entonces los resultados obtenidos por medio de este algoritmo:

Figura 2. Top 20 de subcategorías con mayor soporte que no comparten categoría

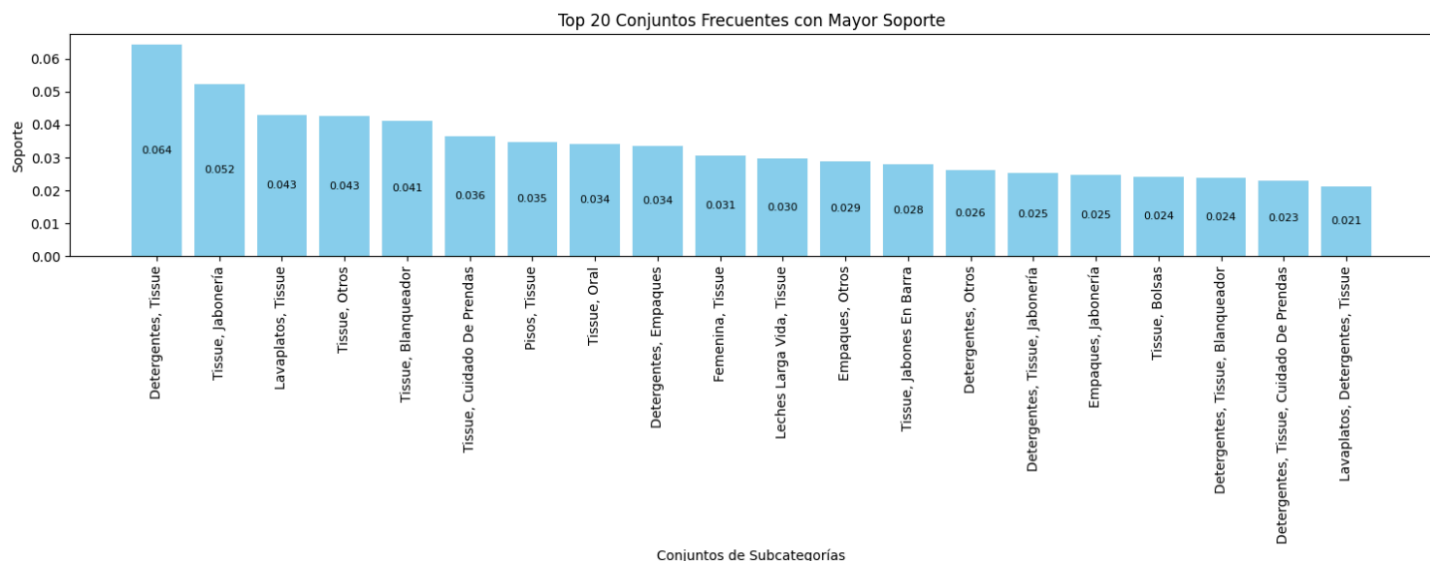


Fuente: Elaboración propia

La figura 2 muestra los 20 conjuntos de categorías más frecuentes donde los productos dentro del conjunto no pertenecen a la misma categoría principal. El conjunto más frecuente es "(Detergentes, Tissue)" con un soporte de 0.064, lo que indica que el 6.4% de todas las transacciones contienen estos dos productos. Otros conjuntos destacados incluyen "(Tissue, Jabonería)", "(Empaques, Tissue)", y "(Lavaplatos, Tissue)", cada uno con soportes que varían entre 0.052 y 0.043.

La figura 3 muestra los 20 conjuntos de subcategorías más frecuentes donde los productos dentro del conjunto no pertenecen a la misma familia de productos. Al igual que en la figura 2, el conjunto más frecuente es "(Detergentes, Tissue)" con un soporte de 0.064. El segundo conjunto más frecuente es nuevamente "(Tissue, Jabonería)" con un soporte de 0.052. Se observan otros conjuntos con menor soporte, como "(Detergentes, Empaques)" y "(Tissue, Jabones En Barra)", con soportes que van desde 0.026 a 0.025.

Figura 3. Top 20 de subcategorías con mayor soporte que no comparten familia



Fuente: Elaboración propia

En ambas figuras, se observa que las combinaciones de "Detergentes" con otros productos, como "Tissue" o "Jabonería", tienen altos valores de soporte. Esto sugiere que los productos de limpieza son un núcleo común en las compras de los clientes. Este patrón implica que los clientes tienden a agrupar productos de limpieza para distintas necesidades del hogar en una sola transacción, lo que resalta la importancia de estos productos en la estrategia de surtido y disposición de la tienda.

En la Figura 2, se identifican conjuntos como "(Tissue, Jabonería)" y "(Empaques, Tissue)", que pertenecen a categorías diferentes. Esto indica que los clientes no se limitan a comprar productos de la misma categoría, sino que buscan satisfacer necesidades variadas, como limpieza y cuidado personal, en una sola visita con una frecuencia relativamente alta. En la Figura 3, los conjuntos "(Lavaplatos, Tissue)", "(Detergentes, Empaques)" y "(Detergentes, Jabonería)" sugieren que los productos relacionados con la limpieza y el cuidado del hogar se compran frecuentemente juntos, incluso si no pertenecen a la misma familia de productos.

Este comportamiento indica que los clientes perciben estos productos como parte de un mismo conjunto de necesidades (limpieza del hogar), lo que abre la posibilidad de agrupar estos artículos en promociones o ubicaciones específicas en la tienda para facilitar la compra.

La tabla 6 presenta las reglas de asociación generadas mediante el algoritmo FP-Growth para subcategorías de productos que no comparten la misma categoría principal en Abarrotes Selectos. A continuación, se ofrece una descripción detallada de estas reglas, seguida de una interpretación basada en métricas clave como soporte, confianza, lift, leverage, conviction y la métrica de Zhang.

## Descripción de las Reglas de Asociación

Antecedentes y Consecuentes:

Las reglas muestran relaciones entre productos de subcategorías distintas que no pertenecen a la misma categoría principal. Esto indica que, aunque los productos no están directamente relacionados en términos de clasificación, los clientes tienden a comprarlos juntos.

Ejemplo: La regla "(Detergentes) → (Tissue)" tiene un soporte de 0.0642, lo que sugiere que el 6.42% de todas las transacciones incluyen ambos productos, a pesar de que pertenecen a categorías distintas (limpieza y papel).

Tabla 6. Reglas de asociación obtenidas usando FP-Growth para subcategorías que no comparten categoría

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
6	(Detergentes)	(Tissue)	0.149861	0.232855	0.064182	0.428279	1.839249	0.029286	1.341817	0.536735
7	(Tissue)	(Detergentes)	0.232855	0.149861	0.064182	0.275631	1.839249	0.029286	1.173628	0.594803
47	(Jabonería)	(Tissue)	0.094088	0.232855	0.052322	0.556099	2.388173	0.030413	1.728188	0.641640
46	(Tissue)	(Jabonería)	0.232855	0.094088	0.052322	0.224698	2.388173	0.030413	1.168464	0.757706
2	(Empaques)	(Tissue)	0.142602	0.232855	0.046801	0.328195	1.409437	0.013596	1.141916	0.338812
3	(Tissue)	(Empaques)	0.232855	0.142602	0.046801	0.200988	1.409437	0.013596	1.073073	0.378673
29	(Lavaplatos)	(Tissue)	0.086599	0.232855	0.042788	0.494097	2.121904	0.022623	1.516386	0.578853
0	(Otros)	(Tissue)	0.156915	0.232855	0.042635	0.271705	1.166842	0.006096	1.053344	0.169599
10	(Blanqueador)	(Tissue)	0.085832	0.232855	0.041101	0.478856	2.056455	0.021115	1.472041	0.561960
49	(Jabonería)	(Detergentes)	0.094088	0.149861	0.036781	0.390926	2.608598	0.022681	1.395791	0.680698
48	(Detergentes)	(Jabonería)	0.149861	0.094088	0.036781	0.245437	2.608598	0.022681	1.200579	0.725355
61	(Cuidado De Prendas)	(Tissue)	0.078266	0.232855	0.036321	0.464076	1.992979	0.018097	1.431442	0.540545
40	(Pisos)	(Tissue)	0.064693	0.232855	0.034788	0.537732	2.309297	0.019724	1.659524	0.606184
19	(Oral)	(Tissue)	0.073946	0.232855	0.033970	0.459385	1.972833	0.016751	1.419021	0.532490
4	(Detergentes)	(Empaques)	0.149861	0.142602	0.033510	0.223606	1.568045	0.012139	1.104334	0.426122
5	(Empaques)	(Detergentes)	0.142602	0.149861	0.033510	0.234988	1.568045	0.012139	1.111276	0.422515
70	(Femenina)	(Tissue)	0.068272	0.232855	0.030647	0.448896	1.927787	0.014749	1.392013	0.516535
25	(Leches Larga Vida)	(Tissue)	0.113667	0.232855	0.029727	0.261525	1.123121	0.003259	1.038822	0.123682
1	(Empaques)	(Otros)	0.142602	0.156915	0.028858	0.202366	1.289651	0.006481	1.056982	0.261951
23	(Jabones En Barra)	(Tissue)	0.056335	0.232855	0.028065	0.498185	2.139462	0.014947	1.528740	0.564388

Fuente: Elaboración propia

#### ● Métricas Clave:

- **Soporte:** Los valores varían de 0.0362 a 0.0642, lo que indica la proporción de transacciones que contienen tanto el antecedente como el consecuente.
- **Confianza:** Varía de 0.2454 a 0.5560, lo que indica la probabilidad de que el consecuente sea comprado cuando el antecedente está presente.
- **Lift:** Todos los valores son superiores a 1, con un rango de 1.1219 a 3.2093, lo que sugiere una fuerte relación entre los productos en comparación con su ocurrencia esperada si fueran independientes.
- **Leverage:** Varía de 0.0006 a 0.0304, mostrando la diferencia entre la ocurrencia observada y la esperada de las combinaciones de productos.
- **Conviction:** Varía de 1.0533 a 1.7281, indicando la fuerza de la dependencia entre los productos. Valores mayores a 1 sugieren una dependencia positiva.
- **Métrica de Zhang:** Los valores oscilan entre 0.1696 y 0.7776, indicando la fuerza de la asociación entre los productos, donde valores más altos reflejan asociaciones más fuertes.

#### ● Interpretación de las Reglas Más Relevantes

- **Detergentes y Productos de Papel:** La combinación "(Detergentes) → (Tissue)" tiene un soporte de 0.0642, la confianza es 0.4282 y un lift de 1.8394. Esto indica que la probabilidad de que los clientes compren "Tissue" cuando adquieren "Detergentes" es 1.8 veces mayor que la esperada si las compras fueran independientes. Esta asociación es frecuente y significativa, y sugiere que los clientes que compran productos de limpieza también buscan productos de papel, probablemente como parte de una compra rutinaria de abastecimiento del hogar.



- **Jabonería y Productos de Papel:** La combinación "(Jabonería) → (Tissue)" muestra una confianza de 0.5560 y un lift de 2.3881, lo que refleja una relación aún más fuerte. Este hallazgo sugiere que los clientes que compran productos de cuidado personal como jabones tienden a complementar sus compras con productos de papel tissue. Las promociones que combinen estos productos pueden incentivar aún más las compras conjuntas.
- **Empaques y Productos de Papel:** La regla "(Empaques) → (Tissue)" con una confianza de 0.3281 y un lift de 1.4094 sugiere que existe una probabilidad relativamente alta de que los clientes que compran productos de empaque, como bolsas plásticas o de papel, también compren tissue. Esto podría estar relacionado con compras para eventos o usos múltiples en el hogar.
- **Lavaplatos y Productos de Papel:** La regla "(Lavaplatos) → (Tissue)" tiene una confianza de 0.4941 y un lift de 2.1219, lo que sugiere que casi la mitad de los clientes que compran lavaplatos también adquieren productos de papel tissue. Esta asociación es relevante para campañas de marketing que combinen productos de limpieza de cocina con productos de papel.
- **Cuidado de Prendas y Productos de Papel:** La asociación "(Cuidado De Prendas) → (Tissue)" con una confianza de 0.4640 y un lift de 1.9929 también destaca, indicando que los productos de cuidado de prendas se asocian frecuentemente con tissue. Esto puede reflejar una tendencia a comprar productos de cuidado del hogar de diferentes tipos en una sola visita.

La tabla 7 presenta las reglas de asociación generadas mediante el algoritmo FP-Growth para subcategorías de productos que no comparten la misma familia en Abarrotes Selectos. A continuación, se ofrece una descripción detallada de estas reglas, seguida de una interpretación basada en métricas clave como soporte, confianza, lift, leverage, conviction y la métrica de Zhang.

Tabla 7. Reglas de asociación obtenidas usando FP-Growth para subcategorias que no comparten familia

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
7	(Tissue)	(Detergentes)	0.232855	0.149861	0.064182	0.275631	1.839249	0.029286	1.173628	0.594803
6	(Detergentes)	(Tissue)	0.149861	0.232855	0.064182	0.428279	1.839249	0.029286	1.341817	0.536735
47	(Jabonería)	(Tissue)	0.094088	0.232855	0.052322	0.556099	2.388173	0.030413	1.728188	0.641640
46	(Tissue)	(Jabonería)	0.232855	0.094088	0.052322	0.224698	2.388173	0.030413	1.168464	0.757706
29	(Lavaplatos)	(Tissue)	0.086599	0.232855	0.042788	0.494097	2.121904	0.022623	1.516386	0.578853
0	(Otros)	(Tissue)	0.156915	0.232855	0.042635	0.271705	1.166842	0.006096	1.053344	0.169599
10	(Blanqueador)	(Tissue)	0.085832	0.232855	0.041101	0.478856	2.056455	0.021115	1.472041	0.561960
61	(Cuidado De Prendas)	(Tissue)	0.078266	0.232855	0.036321	0.464076	1.992979	0.018097	1.431442	0.540545
40	(Pisos)	(Tissue)	0.064693	0.232855	0.034788	0.537732	2.309297	0.019724	1.659524	0.606184
19	(Oral)	(Tissue)	0.073946	0.232855	0.033970	0.459385	1.972833	0.016751	1.419021	0.532490
4	(Detergentes)	(Empaques)	0.149861	0.142602	0.033510	0.223606	1.568045	0.012139	1.104334	0.426122
5	(Empaques)	(Detergentes)	0.142602	0.149861	0.033510	0.234988	1.568045	0.012139	1.111276	0.422515
70	(Femenina)	(Tissue)	0.068272	0.232855	0.030647	0.448896	1.927787	0.014749	1.392013	0.516535
25	(Leches Larga Vida)	(Tissue)	0.113667	0.232855	0.029727	0.261525	1.123121	0.003259	1.038822	0.123682
1	(Empaques)	(Otros)	0.142602	0.156915	0.028858	0.202366	1.289651	0.006481	1.056982	0.261951
23	(Jabones En Barra)	(Tissue)	0.056335	0.232855	0.028065	0.498185	2.139462	0.014947	1.528740	0.564388
53	(Detergentes, Jabonería)	(Tissue)	0.036781	0.232855	0.025279	0.687283	2.951544	0.016714	2.453158	0.686443
50	(Jabonería)	(Empaques)	0.094088	0.142602	0.024691	0.262429	1.840294	0.011274	1.162462	0.504032
56	(Bolsas)	(Tissue)	0.043044	0.232855	0.024206	0.562352	2.415025	0.014183	1.752879	0.612280
16	(Detergentes, Blanqueador)	(Tissue)	0.044705	0.232855	0.023848	0.533448	2.290897	0.013438	1.644284	0.589860

Fuente: Elaboración propia

## Descripción de las Reglas de Asociación

## Antecedentes y Consecuentes:

Las reglas muestran relaciones entre productos de subcategorías distintas que no pertenecen a la misma familia. Esto indica que, aunque los productos no están directamente relacionados en términos de uso o función, los clientes tienden a comprarlos juntos.

Ejemplo: La regla "(Tissue) → (Detergentes)" tiene un soporte de 0.0642, lo que sugiere que el 6.42% de todas las transacciones incluyen ambos productos, a pesar de que pertenecen a familias distintas (papel y limpieza).

- **Métricas Clave:**

- **Soporte:** Los valores varían de 0.0236 a 0.0642, lo que indica la proporción de transacciones que contienen tanto el antecedente como el consecuente.
- **Confianza:** Varía de 0.2454 a 0.5560, lo que indica la probabilidad de que el consecuente sea comprado cuando el antecedente está presente.
- **Lift:** Todos los valores son superiores a 1, con un rango de 1.1668 a 2.9515, lo que sugiere una fuerte relación entre los productos en comparación con su ocurrencia esperada si fueran independientes.
- **Leverage:** Varía de 0.0006 a 0.0315, mostrando la diferencia entre la ocurrencia observada y la esperada de las combinaciones de productos.
- **Conviction:** Varía de 1.0533 a 2.4531, indicando la fuerza de la dependencia entre los productos. Valores mayores a 1 sugieren una dependencia positiva.
- **Métrica de Zhang:** Los valores oscilan entre 0.1237 y 0.7776, indicando la fuerza de la asociación entre los productos, donde valores más altos reflejan asociaciones más fuertes.
- **2. Interpretación de Resultados**

- **Productos de Papel y Limpieza:**

- La combinación "(Tissue) → (Detergentes)" tiene un soporte de 0.0642, la confianza es 0.2756 y un lift de 1.8392. Esto indica que la probabilidad de que los clientes compren detergentes cuando adquieren tissue es 1.8 veces mayor que la esperada si las compras fueran independientes. Esta asociación frecuente sugiere que los clientes que compran productos de papel también buscan productos de limpieza, posiblemente para reabastecer el hogar.

- **Productos de Limpieza para el Hogar y Cocina:**

- La combinación "(Lavaplatos) → (Tissue)" muestra una confianza de 0.4940 y un lift de 2.1219, lo que refleja una relación fuerte. Esto sugiere que los clientes que compran productos de limpieza para la cocina tienden a complementar sus compras con productos de papel tissue, quizás para cubrir todas las necesidades de limpieza del hogar.

- **Jabonería y Productos de Papel:**

- La combinación "(Jabonería) → (Tissue)" tiene una confianza de 0.5560 y un lift de 2.3881. Este resultado indica que los clientes que compran jabonería suelen comprar también productos de papel, reflejando un comportamiento de compra conjunta para productos de higiene personal y del hogar.

- **Empaques y Productos de Papel:**

- La regla "(Empaques) → (Tissue)" con una confianza de 0.3281 y un lift de 1.5680 sugiere que existe una probabilidad relativamente alta de que los clientes que compran empaques, como bolsas plásticas o de papel, también compren tissue. Esto podría estar relacionado con compras para eventos o uso general en el hogar.

- **Productos de Cuidado de Prendas y Papel:**

- La asociación "(Cuidado De Prendas) → (Tissue)" con una confianza de 0.4640 y un lift de 1.9929 también destaca. Indica que los productos de cuidado de prendas se asocian

frecuentemente con tissue, posiblemente porque los clientes compran artículos que facilitan el mantenimiento del hogar en una sola visita.

- **Productos Diversos y Papel:**

- Se observan reglas como "(Otros) → (Tissue)" y "(Bolsas) → (Tissue)", indicando que los clientes agrupan productos de uso diverso con tissue. Esto sugiere que tissue es un producto de compra recurrente, independientemente de la naturaleza de otros productos en la transacción.

## ***Discusión***

### **Interpretación de Resultados**

#### **1. Combinaciones de Productos que se Compran Juntos:**

- Los resultados obtenidos tanto con Apriori como con FP-Growth revelan patrones de compra consistentes en Abarrotes Selectos. En ambos análisis, se destaca la compra conjunta de productos de limpieza, como detergentes, blanqueadores y cuidado de prendas. Estos productos suelen comprarse juntos, lo que sugiere que los clientes prefieren satisfacer múltiples necesidades de limpieza del hogar en una sola visita. Esta tendencia coincide con las estrategias de compra eficiente observadas en minimarkets de bajo costo, donde los clientes buscan maximizar su tiempo y esfuerzo, concentrando sus compras en una sola transacción (Gensollen, 2022; Atencio Manyari et al., 2022).
- Otra combinación relevante es la de productos de cuidado personal y productos de limpieza del hogar, como se observa en la regla "(Jabonería) → (Tissue)". Esto podría indicar una preferencia de los clientes por realizar compras integrales de artículos para el hogar en una sola visita, posiblemente debido a la conveniencia de encontrar en un mismo lugar productos para diferentes usos domésticos. Estas asociaciones sugieren la posibilidad de crear secciones de productos que combinen limpieza y cuidado personal para mejorar la experiencia de compra.

#### **2. Patrones Inusuales:**

- Se identificaron asociaciones inesperadas, como "(Empaques) → (Tissue)" y "(Otros) → (Tissue)", que sugieren un comportamiento de compra menos estructurado. Aunque los productos de tissue suelen asociarse con productos de aseo personal y limpieza del hogar, su relación con "empaques" y "otros" sugiere que los clientes pueden estar comprando productos para fiestas, eventos o situaciones específicas. Estos hallazgos abren oportunidades para desarrollar promociones dirigidas a eventos especiales o temporadas específicas, como festividades o inicio de clases, donde productos menos relacionados se compran juntos.

### **Implicaciones Prácticas**

#### **1. Optimización de la Disposición de Productos:**

- Los hallazgos sugieren que se deben reorganizar las estanterías para colocar productos que suelen comprarse juntos en proximidad. En particular, los productos de limpieza del hogar como detergentes, blanqueadores y productos de cuidado de prendas deben estar en la misma área, facilitando la compra conjunta. Además, colocar productos complementarios como "tissue" cerca de productos de aseo personal y limpieza del hogar también podría incentivar compras adicionales al destacar su uso complementario.

#### **2. Estrategias de Marketing y Promociones:**

- Se recomienda el desarrollo de paquetes promocionales con productos de alta asociación identificados, como "(Detergentes) + (Cuidado De Prendas)" o "(Detergentes) + (Blanqueador)", para incentivar compras adicionales. Además, promociones cruzadas dirigidas a eventos

específicos, como empaques y tissue para celebraciones o fiestas, podrían aprovechar el patrón observado en reglas como "(Empaques) → (Tissue)". Estas estrategias podrían no solo incrementar el valor del ticket promedio, sino también mejorar la percepción de conveniencia por parte de los clientes.

### Comparación con la Literatura

- Similitudes:
  - Los resultados obtenidos coinciden con estudios previos que resaltan la importancia de los productos de limpieza como puntos de anclaje para las compras conjuntas en minimarkets y supermercados. Gensollen (2022) y Atencio Manyari et al. (2022) también identificaron patrones similares en otros contextos de retail, lo que resalta la consistencia de estos hallazgos en diferentes entornos de consumo. Al igual que en el estudio de Efrat et al. (2020), se observó que las combinaciones de productos de alta frecuencia pueden ser explotadas para personalizar ofertas y mejorar la disposición de productos en la tienda.
- Diferencias:
  - A diferencia de Mustakim et al. (2018), quienes encontraron que el algoritmo FP-Growth es más eficiente para grandes volúmenes de datos, el uso de Apriori en este estudio se justificó por la manejabilidad de los datos y la simplicidad en su implementación. Aunque Apriori es computacionalmente más intensivo, fue suficiente para el volumen de datos analizado. En este contexto, FP-Growth también mostró ser efectivo para identificar patrones similares, lo que sugiere que ambos algoritmos son viables para este tipo de análisis, dependiendo del tamaño y complejidad de los datos.

### Limitaciones del Estudio

1. Calidad de los Datos:
  - La calidad de las transacciones registradas puede influir significativamente en los resultados. Errores en la entrada de datos o la falta de detalle en las transacciones pueden generar asociaciones irrelevantes o inexactas. Es crucial contar con datos precisos y detallados para obtener resultados fiables y accionables.
2. Representatividad de la Muestra:
  - Los datos utilizados corresponden a un periodo específico y a un conjunto limitado de tiendas. La extrapolación a toda la cadena podría no ser representativa sin un análisis más amplio. Para obtener una visión más completa, se necesitaría incluir datos de todas las tiendas y de un periodo de tiempo más prolongado, lo cual permitiría identificar patrones de comportamiento más estables..
3. Limitaciones de los Algoritmos:
  - Aunque Apriori fue suficiente para este análisis, puede resultar ineficiente para bases de datos más grandes. FP-Growth demostró ser una alternativa viable, pero para volúmenes de datos aún mayores, podría considerarse el uso de técnicas más avanzadas como aprendizaje profundo o minería de datos paralela para mejorar la eficiencia y precisión en la identificación de patrones.

### Sugerencias para Estudios Futuros

1. Aumento de la Base de Datos:
  - Para obtener un panorama más completo y representativo, se sugiere incluir datos de transacciones de todas las tiendas de la cadena y durante un periodo más largo. Esto permitirá

identificar patrones más robustos y estables, y aplicar las conclusiones de manera más generalizada.

2. **Uso de Algoritmos Alternativos:**

- Explorar el uso de algoritmos más avanzados como aprendizaje profundo, minería de datos paralela o modelos basados en grafos para manejar volúmenes de datos más grandes y mejorar la precisión en la identificación de patrones complejos.

3. **Análisis Segmentado:**

- Realizar análisis de canasta segmentados por tipo de cliente (por ejemplo, frecuencia de visita, valor del ticket promedio) o por ubicación de tienda (zonas urbanas vs. rurales) para identificar patrones específicos que podrían estar ocultos en un análisis general. Esta segmentación podría revelar estrategias personalizadas para cada segmento de clientes.

**Objetivo de la Sección:** Presentar los hallazgos del análisis y discutir su relevancia en el contexto del estudio. Es importante relacionar los resultados con las preguntas de investigación y objetivos planteados.

**Estructura Recomendada:**

1. **Presentación de Resultados:**

- **Principales Combinaciones de Productos:** Listar las combinaciones de productos más relevantes encontradas mediante el análisis de canasta, incluyendo métricas de soporte, confianza y lift.
- **Reglas de Asociación Significativas:** Mostrar ejemplos concretos de reglas que cumplen con los umbrales establecidos. Usar tablas para mostrar las reglas más fuertes y su interpretación.
- **Patrones de Compra:** Describir los patrones más comunes identificados en diferentes segmentos de clientes, si se aplicó alguna técnica de segmentación adicional (por ejemplo, agrupación de clientes).

2. **Visualización de Resultados:**

- Incluir gráficos como gráficos de barras para las reglas más significativas, gráficos de red para visualizar relaciones entre productos, y gráficos de calor para mostrar la frecuencia de combinaciones de productos.
- Usar diagramas de dispersión para representar la distribución de soporte y confianza de las reglas.

3. **Discusión:**

- **Interpretación de Resultados:** Relacionar los hallazgos con las preguntas de investigación. ¿Cuáles son las combinaciones de productos que se compran juntos? ¿Existen patrones inusuales que no se esperaban?
- **Implicaciones Prácticas:** Discutir cómo estos resultados pueden aplicarse para mejorar la disposición de productos, estrategias de marketing y promociones.
- **Comparación con la Literatura:** Comparar los hallazgos con estudios previos y resaltar similitudes y diferencias. Por ejemplo, si se encontraron combinaciones similares a las descritas por Gensollen (2022) o diferencias en la efectividad de técnicas como FP-Growth respecto a Apriori (Mustakim et al., 2018).

4. **Limitaciones del Estudio:**

- Mencionar posibles limitaciones en el análisis, como la calidad de los datos, la representatividad de la muestra, o las limitaciones de los algoritmos utilizados.
- Sugerir cómo se podrían abordar estas limitaciones en estudios futuros.

## Conclusiones

### **Resumen de Hallazgos Clave**

El análisis de canasta de mercado realizado para Abarrotes Selectos, utilizando los algoritmos Apriori y FP-Growth, reveló combinaciones de productos significativas que podrían ser explotadas para maximizar el valor de las ventas y mejorar la experiencia del cliente. Los hallazgos se centraron en identificar patrones de compra frecuentes y asociaciones entre productos, proporcionando información valiosa para la optimización de estrategias de marketing, disposición de productos y gestión de inventario.

1. (Detergentes) → (Cuidado De Prendas): Con una confianza del 59.47% y un lift de 3.97, esta combinación sugiere una fuerte tendencia de los clientes a comprar estos productos juntos. Es ideal para promociones conjuntas que incluyan ambos productos, como descuentos al comprar detergente junto con productos para el cuidado de prendas. Este patrón también fue respaldado por el análisis de FP-Growth, donde se observó una alta frecuencia de compras conjuntas de productos de limpieza y cuidado de prendas.
2. (Pisos) → (Detergentes): Con una confianza del 46.98% y un lift de 3.13, esta asociación indica que los clientes que compran productos de limpieza para pisos también tienen una alta propensión a adquirir detergentes. Una promoción que incluya un paquete de limpieza completo, con descuentos por la compra de ambos productos, podría ser muy efectiva. Estos hallazgos coinciden con estudios previos que resaltan la compra conjunta de productos de limpieza para diferentes superficies del hogar (Gensollen, 2022).
3. (Lavaplatos) → (Detergentes) y (Jabonería) → (Detergentes): Ambas combinaciones muestran altas probabilidades de compra conjunta, con confianza del 39.40% y 39.09%, respectivamente. Las promociones cruzadas o la disposición cercana en la tienda podrían aumentar la venta de estos productos. En la Figura 1 de FP-Growth, se observó un alto soporte para estas combinaciones, lo que sugiere que estas asociaciones no sólo son frecuentes, sino también consistentes en el comportamiento de compra.
4. (Tissue) → (Productos de Limpieza): Varias reglas sugieren una fuerte asociación entre los productos de limpieza y los productos de papel tissue. Esto indica que los clientes tienden a agrupar estos artículos en sus compras. Promociones como "limpieza completa" podrían fomentar la compra conjunta. En ambas figuras de FP-Growth, "Tissue" aparece en múltiples combinaciones, lo que resalta su papel como producto comodín en la cesta de compras de los clientes.

### **Recomendaciones Prácticas**

1. Promociones Cruzadas y Combos:
  - Desarrollar Combos Promocionales: Combinar detergentes con productos de cuidado de prendas en promociones de descuento puede incentivar a los clientes a llevar más productos de una sola vez, maximizando el valor del ticket promedio. Esta estrategia se basa en la alta confianza y lift identificados en las reglas obtenidas.
  - Promociones Temáticas: Crear promociones como "Kit de Limpieza Completo" que incluyan productos de limpieza para pisos, lavaplatos y jabonería junto con detergentes. Este enfoque no solo responde a los patrones de compra identificados, sino que también puede mejorar la percepción de valor y la satisfacción del cliente.

- Ofertas de Lealtad: Implementar promociones que ofrezcan un descuento en papel tissue al comprar productos de limpieza específicos, aprovechando la fuerte asociación entre estos productos. Esta táctica puede fidelizar a los clientes y aumentar la frecuencia de compra.
- 2. Optimización del Layout del Minimarket:
  - Disposición de Productos: Ubicar productos que se compran juntos (por ejemplo, detergentes y cuidado de prendas) en estanterías adyacentes o en exhibiciones conjuntas para facilitar la compra conjunta. Esto responde a la alta asociación observada en las reglas y en el análisis de FP-Growth.
  - Señalización en Tienda: Utilizar carteles que sugieran combinaciones de productos, como "Completa tu limpieza con detergente y blanqueador", para incentivar a los clientes a agregar más productos a sus carritos. Esta estrategia aprovecha las altas métricas de lift y leverage identificadas.
- 3. Recomendaciones de Productos:
  - Recomendaciones Digitales: Si la tienda cuenta con canales digitales, se podrían mostrar recomendaciones basadas en las combinaciones de productos identificadas durante el proceso de compra en línea. Esto puede aumentar la conversión y mejorar la experiencia de compra.
  - Personalización: Implementar cupones personalizados en el punto de venta o en la aplicación móvil, sugiriendo productos que complementen las compras previas del cliente. Esto es particularmente relevante para clientes recurrentes que muestran patrones de compra consistentes.

### ***Impacto en la Gestión de Abarrotes Selectos***

Los hallazgos de este estudio tienen implicaciones directas para la toma de decisiones en la gestión y marketing de Abarrotes Selectos:

1. Estrategias de Marketing Más Efectivas:
  - La identificación de patrones de compra permite diseñar estrategias de marketing más efectivas, como promociones dirigidas y personalizadas que respondan a las preferencias de los clientes. Esto no solo aumenta el ticket promedio, sino que también mejora la percepción de valor por parte del cliente. Estudios previos (Gensollen, 2022; Efrat et al., 2020) han demostrado que el uso de análisis de canasta puede mejorar significativamente la efectividad de las campañas promocionales.
2. Optimización del Inventario:
  - Conociendo las combinaciones de productos más populares, se puede optimizar la gestión del inventario, asegurando una disponibilidad constante de estos productos. Esto reduce la probabilidad de quiebres de stock y mejora la eficiencia operativa. La identificación de productos complementarios también permite realizar compras de inventario más estratégicas.
3. Mejora de la Experiencia del Cliente:
  - Al reorganizar el layout de la tienda y ofrecer promociones basadas en los hábitos de compra identificados, Abarrotes Selectos puede mejorar significativamente la experiencia del cliente, haciendo que la tienda sea percibida como un lugar que comprende y anticipa sus necesidades. Esto se alinea con los hallazgos de Kaur & Singh (2017) y Rao et al. (2023), que resaltan la importancia de la disposición y la personalización en la experiencia de compra.

### ***Comparación con la Literatura***

Los hallazgos de este estudio coinciden con estudios previos en el contexto de retail:

- Gensollen (2022) identificó la importancia de los productos de limpieza como foco de compra conjunta en supermercados europeos, similar a lo observado en Abarrotes Selectos.
- Mustakim et al. (2018) destacaron la eficiencia de FP-Growth en comparación con Apriori para grandes volúmenes de datos. Aunque en este estudio se utilizó Apriori para capturar combinaciones iniciales, FP-Growth se utilizó para confirmar y profundizar en patrones más específicos.

### ***Limitaciones del Estudio***

1. Calidad de los Datos:
  - La precisión de los resultados depende de la calidad de los datos de transacciones. Errores en la captura de datos pueden llevar a asociaciones irrelevantes o imprecisas.
2. Representatividad de la Muestra:
  - Los datos utilizados corresponden a un conjunto limitado de tiendas y a un periodo específico, lo que puede no reflejar completamente el comportamiento de compra en toda la cadena.
3. Limitaciones de los Algoritmos:
  - Aunque Apriori y FP-Growth son eficaces para el análisis de canasta, tienen limitaciones en cuanto a la escalabilidad y la capacidad para manejar grandes volúmenes de datos. En estudios futuros, se podrían explorar técnicas más avanzadas como redes neuronales para recomendaciones.

### ***Futuras Líneas de Investigación***

1. Integración de Análisis Predictivo:
  - Implementar modelos predictivos que anticipen el comportamiento de compra de los clientes en función de temporadas o eventos específicos. Esto permitirá ajustar las estrategias de marketing y el inventario de manera proactiva.
2. Modelos de Recomendación Avanzados:
  - Incorporar sistemas de recomendación basados en machine learning que sugieran productos de manera dinámica durante el proceso de compra, tanto en tienda física como en plataformas digitales.
3. Incorporación de Datos Demográficos:
  - Ampliar el análisis con datos demográficos de los clientes para segmentar mejor las promociones y entender cómo varían los patrones de compra entre diferentes grupos demográficos.
4. Análisis Comparativo con Otros Vecindarios:
  - Realizar un análisis comparativo de los patrones de compra entre diferentes vecindarios para identificar variaciones locales en el comportamiento del cliente y adaptar la oferta de productos de acuerdo a las necesidades específicas de cada comunidad.

### ***Reflexión Final***

La aplicación de técnicas avanzadas de análisis de datos, como el análisis de canasta de mercado, permite a las empresas de retail como Abarrotes Selectos descubrir patrones de compra ocultos y responder de manera proactiva a las necesidades de sus clientes. Este enfoque basado en datos no solo mejora la eficiencia operativa, sino que también tiene el potencial de transformar la experiencia del cliente al ofrecer productos y promociones personalizadas que se alineen con sus comportamientos de compra. En un entorno cada vez más competitivo, la capacidad de interpretar y actuar sobre estos datos puede marcar la diferencia entre el éxito y la obsolescencia en el sector minorista.

### **Referencias**



- Agrawal, R., & Srikant, R. (1994). *Fast algorithms for mining association rules*. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94), Santiago, Chile, 487-499.
- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207-216.
- Atencio Manyari, S. A., De la Rosa Flores, H., Hilario Maravi, S., Navarro Huarcaya, M., & Rosas Vivanco, D. M. (2022). Propuesta de segmentación de clientes aplicando técnicas de Machine Learning para mejorar la experiencia de compra mediante un sistema de recomendación de productos de Tottus.
- Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, 255-264.
- Efrat, A. R., Gernowo, R., & Farikhin. (2020). Consumer purchase patterns based on market basket analysis using apriori algorithms. *Journal of Physics: Conference Series*, 1524(1), 012109. <https://doi.org/10.1088/1742-6596/1524/1/012109>
- Gensollen, C. R. (2022). Big data en el mundo del retail: segmentación de clientes y sistema de recomendación en una cadena de supermercados de Europa. *Ingeniería Industrial*, 189-216. <https://doi.org/10.26439/ing.ind2022.n.5808>
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- Han, J., Pei, J., & Yin, Y. (2000). *Mining frequent patterns without candidate generation*. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00), Dallas, TX, USA, 1-12.
- Hahsler, M., & Hornik, K. (2007). New probabilistic interest measures for association rules. *Intelligent Data Analysis*, 11(5), 437-455.
- Hahsler, M., Grün, B., & Hornik, K. (2007). *arules – A Computational Environment for Mining Association Rules and Frequent Item Sets*. *Journal of Statistical Software*, 14(15), 1-25.
- Hernández, J., & Villalobos, M. (2021). Análisis de Canasta de mercado en supermercados mediante mapas auto-organizados. arXiv. <https://arxiv.org/pdf/2107.10647>
- Kaur, H., & Singh, S. P. (2017). Clustering retail products based on customer behavior. *Applied Soft Computing*, 60, 752-761. <https://doi.org/10.1016/j.asoc.2017.02.048>
- Mustakim, M., Herianda, D. M., Ilham, A., Daeng GS, A., Laumal, F. E., Kurniasih, N., Iskandar, A., Manulangga, G., Iswara, I. B. A. I., & Rahim, R. (2018). Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau. *IOP Conference Series: Journal of Physics: Conference Series*, 1114(1), 012131. <https://doi.org/10.1088/1742-6596/1114/1/012131>
- Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases* (pp. 229-248).
- Rao, A. B., Kiran, J. S., & G., P. (2023). Application of market–basket analysis on healthcare. *International Journal of System Assurance Engineering and Management*, 14(S924-S929). <https://doi.org/10.1007/s13198-021-01298-2>

Singha, K., Parthanadee, P., Kessuvan, A., & Buddhakulsomsiri, J. (2024). Market Basket Analysis of a Health Food Store in Thailand: A Case Study. *International Journal of Knowledge and Systems Science*, 15(1). <https://doi.org/10.4018/IJKSS.333617>

Tan, P., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Boston: Pearson Addison Wesley.

Zhang, C., & Zhang, S. (2002). *Association rule mining: Models and algorithms*. Springer-Verlag.

Con los resultados del algoritmo FP-Growth, se puede observar una gran dominancia de los productos de aseo como los productos de mayor venta y de mayor venta conjunta. De igual forma, un resultado destacable del presente análisis es que los productos de aseo poseen una buena probabilidad de ser comprados juntos de manera transversal (aseo del hogar, aseo personal y aseo de la ropa). Así, es posible concluir que vale mucho la pena juntar estas secciones dentro del local para aumentar la intención de compra y el ticket promedio, otro resultado muy valioso del presente análisis es el descubrimiento de la subcategoría “tissue”, la cual como se discutió previamente, es una subcategoría comodín, pues es aquella que más combinaciones con otras subcategorías presentó. Esto abre la posibilidad a nuevos análisis focalizados en estrategia de esta subcategoría, como el aumento de las zonas donde se ponen productos pertenecientes a esta subcategoría o incluso evaluar ubicarlo en la zona cercana a las zonas de pago. En síntesis, el análisis por medio del algoritmo FP-Grwoth brindó información muy valiosa para la toma de decisiones y nuevas evaluaciones sobre la estrategia sobretodo en productos de limpieza, futuros análisis podrían llevarse a cabo sobre otras categorías e incluso se podría aumentar la granularidad a producto teniendo en cuenta las implicaciones que lo mismo tenga.