

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

Juan Pablo Mogollón Avaunza¹, Juan Sebastián Hernández Ramírez¹, Karen Rojas Giraldo¹, Jhocel Duvan Suescun Torres¹

Resumen

El análisis de canasta de mercado en Abarrotes Selectos, utilizando los algoritmos Apriori y FP-Growth, reveló combinaciones de productos significativas que los clientes suelen adquirir juntos. Destacaron los productos de limpieza, como detergentes y tissue, que mostraron una fuerte asociación, lo que sugiere que los clientes prefieren agrupar estos artículos en sus compras.

Las métricas utilizadas para evaluar las reglas de asociación, como soporte, confianza y lift, confirmaron combinaciones importantes. Por ejemplo, la relación entre detergentes y productos para el cuidado de prendas mostró una confianza del 59.47% y un lift de 3.97, lo que indica una clara oportunidad para promociones conjuntas. Además, se encontraron asociaciones relevantes entre detergentes y blanqueadores, que pueden aprovecharse para mejorar la disposición de productos y estrategias promocionales.

Se concluyó que reorganizar los productos en la tienda y utilizar promociones cruzadas basadas en estas asociaciones podría 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, alineando su oferta con los patrones de compra identificados.

Introducción.

En el competitivo entorno del retail, comprender los hábitos de compra de los clientes es esencial 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 artículos misceláneos, se encuentra en un proceso de expansión en un vecindario popular. Aunque la empresa ha experimentado un crecimiento sostenido, sospecha que existe un potencial de ventas aún no explotado.

Para capitalizar estas oportunidades, se llevó a cabo un análisis detallado de las transacciones de compra utilizando técnicas de aprendizaje no supervisado, específicamente el análisis de canasta de mercado (Market Basket Analysis, MBA). Esta metodología permite identificar patrones de compra recurrentes y relaciones entre productos, facilitando la creación de estrategias de venta más efectivas y personalizadas.

El estudio se enfoca en responder a las siguientes preguntas clave:

- ¿Cuáles son las combinaciones de productos que los clientes suelen comprar juntos?
- ¿Existen productos que, al ser promovidos conjuntamente, podrían aumentar el valor de las compras?
- ¿Podemos reorganizar la disposición de productos en el minimarket para incentivar la compra conjunta de ciertos artículos?

El diseño del estudio estuvo motivado por objetivos estratégicos como la maximización de ingresos, la optimización del layout del minimarket y la mejora de la satisfacción del cliente. Estos objetivos se alinean con la literatura existente, que ha explorado el uso de técnicas de MBA para personalizar las recomendaciones de productos y mejorar la experiencia del cliente (Gensollen, 2022; Efrat et al., 2020; Atencio Manyari et al., 2022; Singha et al., 2024).

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. 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) emplearon el algoritmo Apriori, uno de los métodos más populares en el

¹ Maestría de Inteligencia Analítica de Datos, Universidad de Los Andes

análisis de canasta, para descubrir asociaciones entre productos en un supermercado. Asimismo, 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. Hernández y Villalobos (2021) emplearon mapas auto-organizados (SOM) para visualizar estos patrones, proporcionando una representación intuitiva y detallada de las relaciones entre productos.

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 de Abarrotes Selectos, un minimarket en expansión, el cual contiene información detallada sobre 152,781 transacciones realizadas en 39,123 compras únicas. Los datos incluyen campos como fecha de venta, código de factura, monto total, cantidad de ítems, familia, categoría, subcategoría, código de producto, cantidad vendida, y valores brutos y netos de cada venta. Cada registro representa un producto individual vendido en una transacción específica, facilitando un análisis detallado de patrones de compra.

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 subcategorías 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

El inventario del minimarket está compuesto por 1,086 productos únicos, organizados en 12 familias, 31 categorías y 185 subcategorías, lo que refleja una gran variedad de productos y permite segmentaciones detalladas para el análisis de ventas. En promedio, se compran 3.9 productos por transacción, lo que sugiere

compras rápidas y específicas, típicas de este formato de tienda. El análisis inicial mostró que la mayoría de las ventas (83.53%) se agrupan bajo la categoría "Otros", indicando una oferta diversificada (Tablas 1 y 2) .

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

Preprocesamiento de datos y variables consideradas

Para el análisis de canasta de mercado (Market Basket Analysis, MBA), se seleccionaron dos variables clave: `codigo_factura` y `subcategoría`. Esta elección simplificó el modelo de datos y permitió identificar asociaciones de compra relevantes a nivel de subcategorías, facilitando decisiones estratégicas para la disposición de productos y promociones cruzadas. La transformación de los datos en una matriz binaria de transacciones y productos permitió aplicar los algoritmos Apriori y FP-Growth para detectar patrones de co-ocurrencia de productos.

La base de datos original no presentaba datos duplicados o faltantes, por lo que no fue necesario realizar tratamientos adicionales. La matriz final, con 39,123 transacciones y 185 subcategorías, facilitó la identificación de combinaciones de productos comprados conjuntamente.

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

Se utilizaron los algoritmos Apriori y FP-Growth para descubrir reglas de asociación. El algoritmo Apriori identifica conjuntos de productos (itemsets) frecuentes en las transacciones y genera reglas de la forma $A \rightarrow B$, donde A y B son conjuntos de productos. Las métricas de evaluación incluyen soporte (proporción de transacciones que contienen la combinación), confianza (probabilidad de que B sea comprado dado A) y lift (medida de la fuerza de la co-ocurrencia comparada con su ocurrencia esperada).

El algoritmo FP-Growth, por su parte, construye un árbol de patrones frecuentes a partir de las transacciones, evitando la generación de candidatos y mejorando la eficiencia en comparación con Apriori. Ambas técnicas permiten generar reglas de asociación útiles para identificar patrones de compra en el minimarket.

Criterios de Selección de Reglas de Asociación

Las reglas de asociación se seleccionaron basándose en métricas como soporte, confianza y lift. Se priorizaron reglas con alto soporte y confianza para asegurar su relevancia y precisión. Se eligieron itemsets de 2 productos para capturar asociaciones simples y efectivas, facilitando la aplicación de estrategias promocionales y la disposición conjunta de productos.

Evaluación de la Calidad de las Reglas

Se evaluó la calidad de las reglas generadas mediante el uso de métricas adicionales como leverage, conviction y la métrica de Zhang, que ayudan a entender mejor la fuerza y relevancia de las asociaciones descubiertas. Se consideró el número y la diversidad de reglas relevantes, así como su aplicabilidad práctica para la toma de decisiones en el negocio, como la disposición de productos o promociones cruzadas.

Herramientas y Software Utilizado

El análisis se realizó en Python utilizando las librerías `pandas` para la manipulación de datos, `mlxtend` para implementar los algoritmos Apriori y FP-Growth, y `matplotlib` para la visualización de los resultados. Jupyter Notebook se utilizó como entorno interactivo para la implementación y exploración de los algoritmos.

Resultados y discusión

Método Apriori.

El análisis de canasta se realizó ajustando diferentes combinaciones de soporte mínimo y confianza mínima para identificar patrones de compra. Los resultados se resumen en la Tabla 4, que detalla:

- **Soporte Mínimo:** Proporción mínima de transacciones en las que debe aparecer una combinación de productos para ser considerada.
- **Confianza Mínima:** Proporción mínima de transacciones que contienen el antecedente y el consecuente, lo que mide la precisión de la regla.
- **Reglas Relevantes:** Número de reglas que cumplen con los umbrales establecidos de soporte y confianza.
- **Diversidad de Reglas:** Cantidad de combinaciones únicas de productos 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

- **Soporte 0.05 y Confianza 0.15:** Se generan pocas reglas (4) con baja diversidad (2), indicando solo unas pocas combinaciones de productos populares.
- **Soporte 0.02 y Confianza 0.15:** Mayor número de reglas (85) y diversidad (42), capturando una amplia variedad de patrones de compra.
- **Soporte 0.02 y Confianza 0.20:** Se reduce ligeramente el número de reglas (71) y diversidad (40), equilibrando cantidad y precisión.
- **Soporte 0.02 y Confianza 0.25:** Menos reglas relevantes (57) y diversidad (36), indicando combinaciones más precisas pero menos frecuentes.
- **Soporte 0.03 y Confianza 0.25:** Menor cantidad de reglas (18) y diversidad (15), mostrando combinaciones específicas pero limitadas.

Se eligió la configuración de soporte 0.03 y confianza 0.25 como la más adecuada para el contexto de Abarrotes Selectos, debido a su equilibrio entre especificidad y relevancia de las reglas generadas.

Justificación de la Selección:

- **Equilibrio entre Especificidad y Relevancia:** Las combinaciones seleccionadas están presentes en al menos el 3% de las transacciones, capturando asociaciones significativas pero no triviales. La confianza del 25% asegura precisión suficiente para estrategias promocionales.
- **Oportunidades de Venta Cruzada:** Las 18 reglas relevantes presentan 15 combinaciones únicas, útiles para promociones específicas y estrategias de disposición de productos.
- **Optimización del Layout del Minimarket:** Menor cantidad de reglas facilita la implementación de cambios en la disposición de productos, favoreciendo la compra conjunta e impulsiva.
- **Diferenciación en el Vecindario:** Focalizarse en reglas con mayor soporte y confianza mejora la experiencia de compra personalizada, aumentando la fidelización de clientes.

Reglas de asociación y combinaciones de productos

La tabla 5 presenta las 18 reglas obtenidas con soporte 0.03 y confianza 0.25. Un ejemplo destacado es la regla "(Detergentes) → (Cuidado De Prendas)", con un soporte de 0.0465 y una confianza de 0.5947, lo que indica que el 4.65% de las transacciones incluyen ambos productos y que en el 59.47% de las compras de detergentes también se adquieren productos para el cuidado de prendas.

Métricas Clave:

- Soporte: Varía entre 0.03 y 0.0465, mostrando la frecuencia de co-ocurrencia.
- Confianza: Oscila entre 0.2717 y 0.5947, indicando la probabilidad de compra conjunta.
- Lift: Valores de 1.67 a 3.97, sugiriendo asociaciones más fuertes de lo esperado.
- Leverage: De 0.0020 a 0.0468, indicando la diferencia entre ocurrencia observada y esperada.
- Conviction: De 1.05 a 2.09, reflejando la dependencia entre productos.
- Métrica de Zhang: De 0.1696 a 0.8799, valores positivos que indican asociaciones favorables.

Interpretación de Reglas Relevantes:

1. (Detergentes) → (Cuidado De Prendas): Confianza alta (0.5947) y lift (3.97) indican una fuerte asociación, ideal para promociones conjuntas.
2. (Detergentes) → (Blanqueador): Confianza moderada (0.2983) y lift (3.47), sugiriendo promociones cruzadas o disposición conjunta.
3. (Pisos) → (Detergentes): Fuerte relación (confianza 0.4698, lift 3.13) indica potencial para bundles de productos de limpieza.
4. (Jabonería) → (Detergentes): Probabilidad de compra conjunta alta (confianza 0.3909, lift 2.61), favoreciendo promociones cruzadas.
5. (Lavaplatos) → (Detergentes): Alta probabilidad de compra conjunta (confianza 0.3940, lift 2.63), sugiriendo promociones de "limpieza completa".
6. (Tissue) → (Detergentes): Relación positiva (confianza 0.4282, lift 1.84) con potencial para incrementar ventas de impulso.

	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

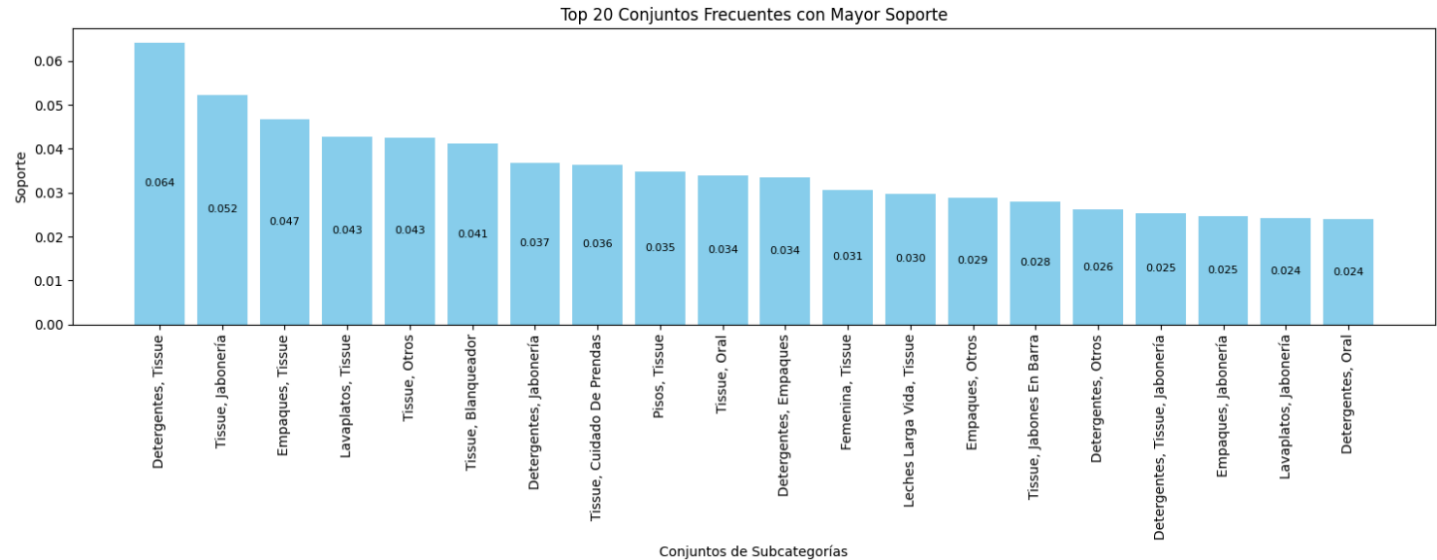
Fuente: Elaboración propia

Estas asociaciones permiten identificar oportunidades estratégicas en la disposición de productos y el diseño de promociones para Abarrotes Selectos, mejorando así la experiencia de compra y aumentando el valor de las transacciones.

Método FP-Growth.

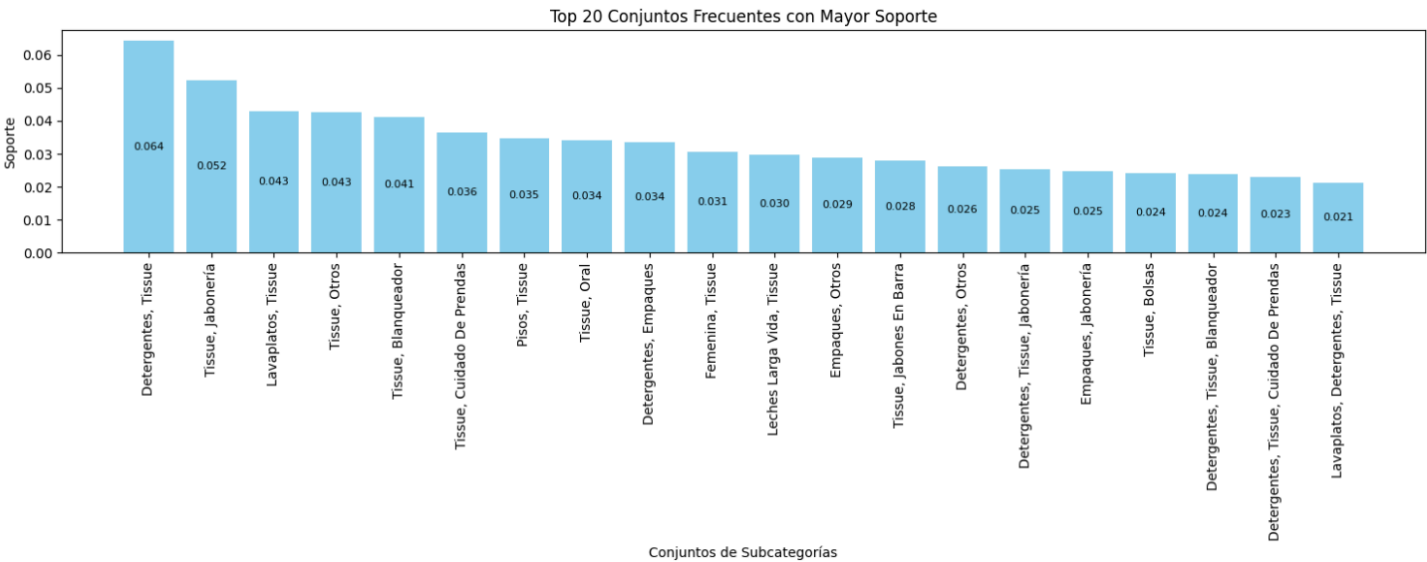
Para implementar el algoritmo FP-Growth, se estableció un soporte mínimo de 0.02 y una confianza mínima de 0.2. Con el fin de identificar patrones de compra menos evidentes, se filtraron las combinaciones de productos que no comparten la misma familia o categoría principal. También se aplicó el análisis a nivel de familias de productos para capturar patrones generales de compra. El análisis resultó en 71 reglas de asociación, de las cuales se destacan las 20 con mayor soporte.

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



Las figuras 2 y 3 muestran los 20 conjuntos de subcategorías y familias más frecuentes, respectivamente, donde los productos no pertenecen a la misma categoría o familia. En ambas figuras, las combinaciones de "Detergentes" con productos como "Tissue" o "Jabonería" presentan altos soportes, indicando que los productos de limpieza son un núcleo común en las compras de los clientes. Esto sugiere que los consumidores tienden a agrupar productos de limpieza para distintas necesidades del hogar en una sola transacción. La combinación más frecuente, "(Detergentes, Tissue)", aparece en el 6.4% de las transacciones, seguida de conjuntos como "(Tissue, Jabonería)" y "(Empaques, Tissue)", con soportes que van del 5.2% al 4.3%.

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



Fuente: Elaboración propia

Este comportamiento resalta la importancia de estos productos en la estrategia de surtido y disposición de la tienda, abriendo la posibilidad de agruparlos en promociones o ubicaciones específicas para facilitar la compra. Además, se observan patrones similares en las familias de productos, como la relación frecuente entre productos de limpieza del hogar y de cocina, lo que indica que los clientes prefieren realizar compras integrales de artículos de limpieza.

Análisis de Reglas de Asociación

La tabla 6 presenta las reglas de asociación generadas para subcategorías que no comparten la misma categoría. Un ejemplo destacado es la regla "(Detergentes) → (Tissue)", con un soporte de 0.0642 y un lift de 1.8394, indicando que los clientes que compran detergentes tienen 1.8 veces más probabilidades de comprar tissue en la misma transacción. Otras asociaciones significativas incluyen "(Jabonería) → (Tissue)", con una confianza de 0.5560 y un lift de 2.3881, sugiriendo que los clientes que compran productos de cuidado personal también suelen adquirir productos de papel.

Las combinaciones frecuentes entre productos de limpieza y papel reflejan un comportamiento de compra rutinario, donde los clientes agrupan productos de mantenimiento del hogar. La combinación "(Lavaplatos) → (Tissue)" también muestra un patrón relevante con un lift de 2.1219, lo que sugiere que casi la mitad de los clientes que compran lavaplatos también adquieren tissue. Estas asociaciones son clave para diseñar promociones efectivas y estrategias de distribución en la tienda.

La tabla 7 presenta las reglas de asociación para subcategorías que no comparten la misma familia de productos. Se destaca la regla "(Tissue) → (Detergentes)", con un soporte de 0.0642 y un lift de 1.8392, que sugiere que la compra de tissue incrementa significativamente la probabilidad de adquirir detergentes. Estas

reglas reflejan un comportamiento de compra orientado a satisfacer múltiples necesidades del hogar en una sola visita.

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

Tabla 7. Reglas de asociación obtenidas usando FP-Growth para subcategorías 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

Las combinaciones frecuentes entre productos de limpieza y papel reflejan un comportamiento de compra rutinario, donde los clientes agrupan productos de mantenimiento del hogar. La combinación "(Lavaplatos) → (Tissue)" también muestra un patrón relevante con un lift de 2.1219, lo que sugiere que casi la mitad de los clientes que compran lavaplatos también adquieren tissue. Estas asociaciones son clave para diseñar promociones efectivas y estrategias de distribución en la tienda.

El análisis FP-Growth permitió identificar patrones significativos de compra entre productos de diferentes categorías y familias. Los resultados presentados en las tablas 6 y 7 proporcionan información valiosa para optimizar la disposición de productos y diseñar campañas de marketing enfocadas en promover la compra conjunta de productos relacionados. Las estrategias basadas en estas asociaciones pueden mejorar la experiencia de compra del cliente y aumentar el valor promedio de las transacciones en Abarrotes Selectos.

Discusión

1. Interpretación de Resultados

- El análisis realizado con Apriori y FP-Growth en Abarrotes Selectos identificó patrones de compra consistentes. Se observó una fuerte asociación entre productos de limpieza como detergentes, blanqueadores y cuidado de prendas, lo que sugiere que los clientes buscan satisfacer múltiples necesidades de limpieza en una sola visita, maximizando su tiempo y esfuerzo (Gensollen, 2022; Atencio Manyari et al., 2022). También se detectó una combinación frecuente de productos de cuidado personal y limpieza del hogar, como la regla "(Jabonería) → (Tissue)", lo que indica una preferencia por compras integrales de artículos para el hogar.
- Además, se encontraron patrones inusuales, como "(Empaques) → (Tissue)", que sugieren compras para eventos específicos, abriendo oportunidades para promociones dirigidas a festividades o temporadas especiales.

2. Implicaciones Prácticas

- Los resultados sugieren la reorganización de estanterías para colocar productos que se compran juntos en proximidad, como detergentes, blanqueadores y productos de cuidado de prendas. También se recomienda desarrollar paquetes promocionales basados en las combinaciones identificadas, como "(Detergentes) + (Cuidado De Prendas)", para incentivar compras adicionales. Asimismo, promociones cruzadas para eventos específicos podrían aprovechar las asociaciones observadas, como empaques y tissue para celebraciones, incrementando así el valor del ticket promedio y la percepción de conveniencia del cliente.

3. Comparación con la Literatura

- Los hallazgos son consistentes con estudios previos que resaltan la importancia de los productos de limpieza como puntos de anclaje para las compras conjuntas en retail (Gensollen, 2022; Atencio Manyari et al., 2022). Al igual que en Efrat et al. (2020), se observó que las combinaciones de productos frecuentes pueden explotarse para personalizar ofertas y mejorar la disposición de productos. A diferencia de Mustakim et al. (2018), se utilizó Apriori debido al tamaño manejable de la base de datos, aunque FP-Growth también demostró ser efectivo.

4. Limitaciones del Estudio

- La calidad de los datos puede influir en los resultados, ya que errores o falta de detalle en las transacciones pueden generar asociaciones irrelevantes. Además, los datos utilizados corresponden a un periodo específico y a un conjunto limitado de tiendas, por lo que los resultados podrían no ser representativos de toda la cadena. Aunque Apriori fue adecuado para este análisis, podría no ser eficiente para bases de datos más grandes. FP-Growth es una alternativa viable, pero para volúmenes aún mayores se podrían explorar técnicas más avanzadas.

5. Sugerencias para Estudios Futuros

- Se recomienda ampliar la base de datos para incluir transacciones de todas las tiendas y durante un periodo más prolongado, lo que permitiría identificar patrones más robustos. Además, se sugiere explorar algoritmos avanzados como aprendizaje profundo o minería de datos paralela para mejorar la eficiencia y precisión en la identificación de patrones. Un análisis segmentado por tipo de cliente o ubicación de tienda podría revelar estrategias personalizadas para diferentes segmentos de clientes.

- En resumen, los hallazgos de este estudio proporcionan insights valiosos para optimizar la disposición de productos y desarrollar estrategias de marketing efectivas en Abarrotes Selectos, contribuyendo a maximizar las ventas y mejorar la experiencia de compra de los clientes.

Conclusiones

Hallazgos Clave

El análisis de canasta de mercado realizado para Abarrotes Selectos reveló combinaciones de productos significativas que pueden maximizar el valor de las ventas y mejorar la experiencia del cliente. Se identificaron asociaciones fuertes como "(Detergentes) → (Cuidado De Prendas)" con una confianza del 59.47% y un lift de 3.97, indicando que los clientes suelen comprar estos productos juntos, lo que sugiere oportunidades para promociones conjuntas. También se observó una tendencia similar con "(Pisos) → (Detergentes)", que podría aprovecharse con paquetes de limpieza completos. Otras combinaciones como "(Lavaplatos) → (Detergentes)" y "(Jabonería) → (Detergentes)" refuerzan la posibilidad de reubicar productos en la tienda para facilitar la compra conjunta.

Recomendaciones Prácticas

- **Promociones Cruzadas y Combos:** Crear combos promocionales que incluyan detergentes con productos de cuidado de prendas o kits de limpieza completos para aumentar el valor del ticket promedio.
- **Optimización del Layout:** Colocar productos que se compran juntos, como detergentes y cuidado de prendas, en estanterías adyacentes y usar señalización para incentivar compras combinadas.
- **Recomendaciones Digitales y Personalización:** Implementar recomendaciones basadas en compras anteriores en canales digitales y personalizar cupones en puntos de venta para aumentar la conversión.

Impacto en la Gestión de Abarrotes Selectos

Los hallazgos permiten diseñar estrategias de marketing más efectivas, como promociones dirigidas y personalizadas que aumenten el ticket promedio y mejoren la percepción de valor del cliente. Además, ayudan a optimizar el inventario y reducir quiebres de stock, mejorando la eficiencia operativa y la experiencia de compra al anticipar las necesidades de los clientes.

Comparación con la Literatura

Los resultados coinciden con estudios previos que destacan la importancia de los productos de limpieza en la compra conjunta (Gensollen, 2022). Aunque se utilizó Apriori para combinaciones iniciales, FP-Growth confirmó patrones más específicos, validando la efectividad de ambos algoritmos para este tipo de análisis.

Limitaciones

- **Calidad de los Datos:** Errores en la captura de datos pueden llevar a asociaciones imprecisas.
- **Representatividad:** Los datos se limitaron a un conjunto específico de tiendas y periodos, lo que puede no reflejar completamente el comportamiento de compra en toda la cadena.
- **Escalabilidad de los Algoritmos:** Apriori y FP-Growth pueden tener limitaciones con grandes volúmenes de datos, por lo que se recomienda explorar técnicas más avanzadas en futuros estudios.

Futuras Líneas de Investigación

- **Integración de Análisis Predictivo:** Desarrollar modelos que anticipen el comportamiento de compra según temporadas.
- **Modelos de Recomendación Avanzados:** Usar machine learning para recomendaciones dinámicas.
- **Datos Demográficos:** Incluir datos demográficos para mejorar la segmentación y personalización de promociones.

- **Análisis Comparativo:** Comparar patrones de compra entre vecindarios para adaptar la oferta según las necesidades locales.

Reflexión Final

El análisis de canasta de mercado ofrece a Abarrotes Selectos una herramienta poderosa para descubrir patrones de compra y adaptar estrategias de manera proactiva. En un entorno competitivo, la capacidad de actuar sobre estos datos es clave para mejorar la experiencia del cliente y la eficiencia operativa, asegurando el éxito y la sostenibilidad 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.