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

Juan Sebastián Hernández Ramírez

Jhocel Duvan Suescun Torres

Karen Rojas Giraldo

Juan Pablo Mogollón Avaunza

Universidad de los Andes

Curso de Aprendizaje No Supervisado

Maestría en Inteligencia Analítica de Datos

Bogotá, Colombia

02/09/2024

**Resumen**

El proyecto busca poder analizar los patrones de compra de los clientes. La identificación de combinaciones de productos frecuentemente adquiridos juntos y la segmentación de los clientes según sus hábitos de compra, no solo permitirá ajustar la disposición de los productos en el minimarket sino que también va a poder permitir impulsar promociones que ayudaran a aumentar las ventas. Este proyecto hará una contribución significativa a la optimización de las operaciones.

**Contenido**

**Pág.**

[Introducción 9](#_Toc175991019)

[Marco Teórico 10](#_Toc175991020)

[Descripción detallada de los datos 11](#_Toc175991021)

[Propuesta metodológica 12](#_Toc175991022)

[Referencias 13](#_Toc175991023)

Introducción

En el competitivo mundo del retail, conocer los hábitos de compra de los clientes es fundamental para diseñar estrategias que maximicen las ventas. Nuestro minimarket, situado en un vecindario popular, ha visto un crecimiento constante, pero sospechamos que hay potencial no explotado en nuestras ventas. Para capitalizar estas oportunidades, hemos decidido llevar a cabo un análisis profundo de las transacciones mediante técnicas de aprendizaje no supervisado, enfocándonos en el análisis de canasta.

Pregunta/Problema a Resolver

El principal objetivo de nuestro proyecto es responder a las siguientes preguntas clave:

1. ¿Cuáles son las combinaciones de productos que los clientes suelen comprar juntos?
2. ¿Existen productos que, al ser promovidos juntos, 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?

Para abordar estas preguntas, nos centraremos en **tareas de clustering y asociación**. El análisis de canasta se enfocará en descubrir reglas de asociación entre productos, mientras que el clustering podría ayudar a identificar segmentos de clientes con comportamientos de compra similares. Si encontramos que los datos son demasiado complejos o numerosos, también podríamos considerar técnicas de **reducción de dimensión** para simplificar la interpretación.

Motivación

Maximización de ingresos: Al identificar qué productos se compran juntos, se pueden crear promociones y combos que incentiven compras adicionales.

Optimización del layout del minimarket: Mejorar la disposición de productos para facilitar el acceso a combinaciones de productos frecuentemente comprados juntos, lo que podría incrementar las ventas por impulso.

Mejora de la satisfacción del cliente: Ofrecer una experiencia de compra más fluida y conveniente al anticipar sus necesidades y preferencias.

Clasificación del Problema

Este problema pertenece principalmente a la **tarea de asociación** (análisis de canasta). Sin embargo, puede incluir elementos de **clustering** para segmentar clientes y entender patrones de compra más detallados.

Marco Teórico

* **Revisión preliminar de antecedentes en la literatura**. Revisión de literatura nacional e internacional, para enriquecer el proyecto con lo que ya se haya escrito sobre el tema que se va a tratar.
* Hernández, J., & Villalobos, M. (2021). *Análisis de Canasta de mercado en supermercados mediante mapas auto-organizados*. arXiv. <https://arxiv.org/pdf/2107.10647>

Este trabajo presenta un enfoque para el análisis de canasta en supermercados utilizando mapas auto-organizados (SOM), una técnica de aprendizaje no supervisado. Este enfoque se centra en identificar patrones de compra mediante la visualización de datos de transacciones. En comparación con nuestro proyecto, ambos comparten el objetivo de descubrir combinaciones de productos que se compran juntos. Mientras que el artículo utiliza mapas auto-organizados para visualizar patrones, nuestro enfoque combina análisis de canasta con clustering y considera la posibilidad de utilizar reducción de dimensión.

Este estudio refuerza la importancia de utilizar métodos avanzados de análisis para mejorar la disposición de productos y las estrategias de marketing en el retail, lo que puede ser complementario o incorporado en nuestra metodología propuesta.

* 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>

El artículo titulado "Clustering retail products based on customer behavior" publicado en ScienceDirect aborda cómo se pueden agrupar productos minoristas en función del comportamiento de los clientes utilizando técnicas de clustering como K-means. Este método es similar al que proponemos en nuestro proyecto, que también utiliza clustering para segmentar productos y clientes basándose en patrones de compra. Sin embargo, mientras que el estudio se enfoca principalmente en la segmentación de productos, nuestro enfoque combina clustering con análisis de canasta para descubrir reglas de asociación entre productos.

Descripción detallada de los datos

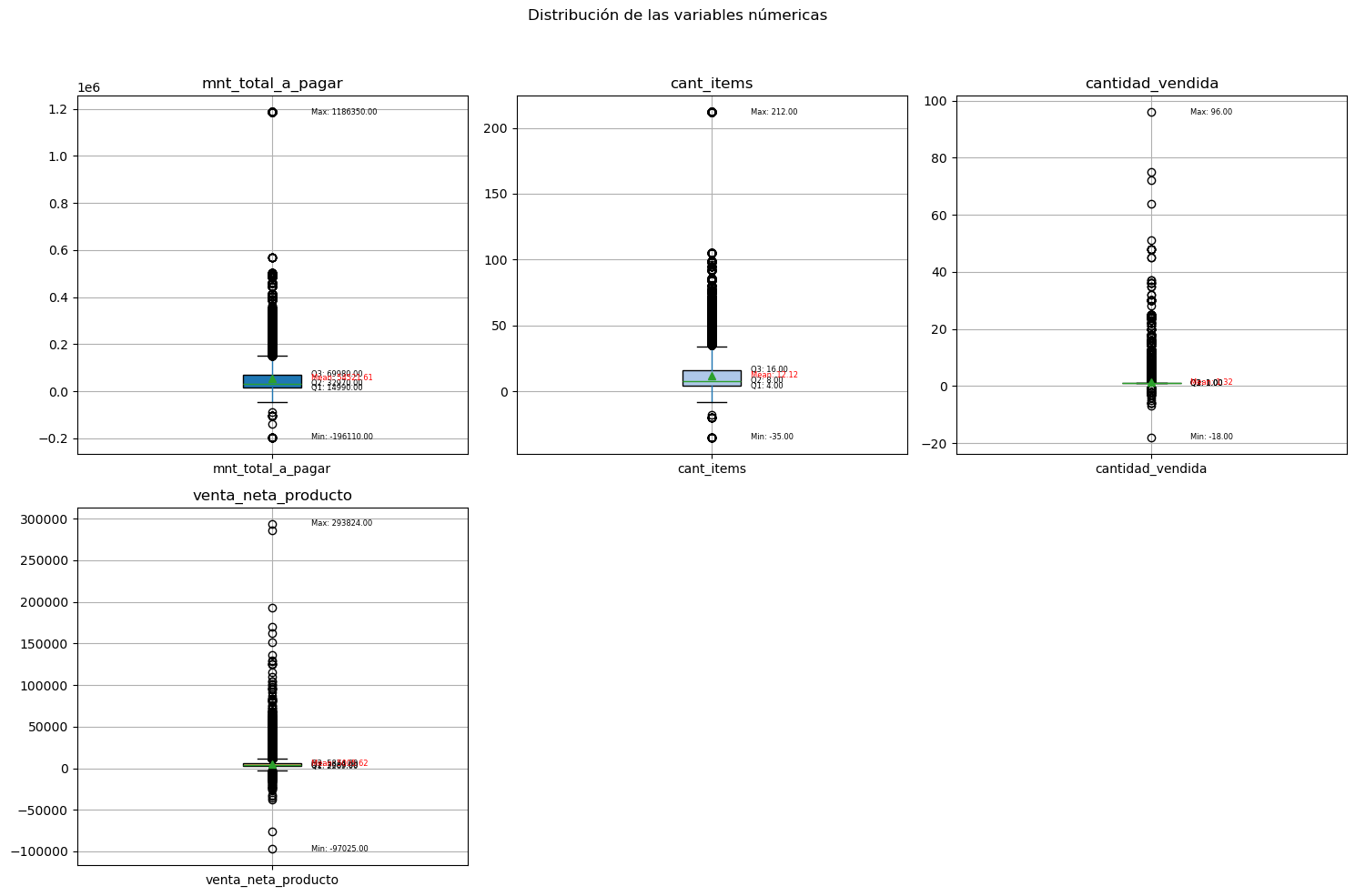
Se cuenta con una base de datos con 152.781 registros y 11 columnas que comprenden de:

* Fecha: fecha de la factura
* codigo\_factura: código de la factura
* mnt\_total\_a\_pagar: Valor total a pagar
* cant\_items: Cantidad de productos
* familia: 11 grupos al que pertenece el articulo (Aseo, alimentos sal, bebidas, alimentos dulces, refrigerados, empaques y bazar, commodities, congelados, mascotas, farmacia, perecederos, cigarrillos)
* categoría: Se encuentran 31 categorías
* subcategoria: Se encuentran 186 subcategorías.
* codigo\_producto: Código del producto.
* cantidad\_vendida Cantidad de unidades vendidas del artículo.
* venta\_bruta\_producto: Valor bruto del artículo
* venta\_neta\_producto:. Valor neto del artículo.

Una revisión preliminar a la base de datos:

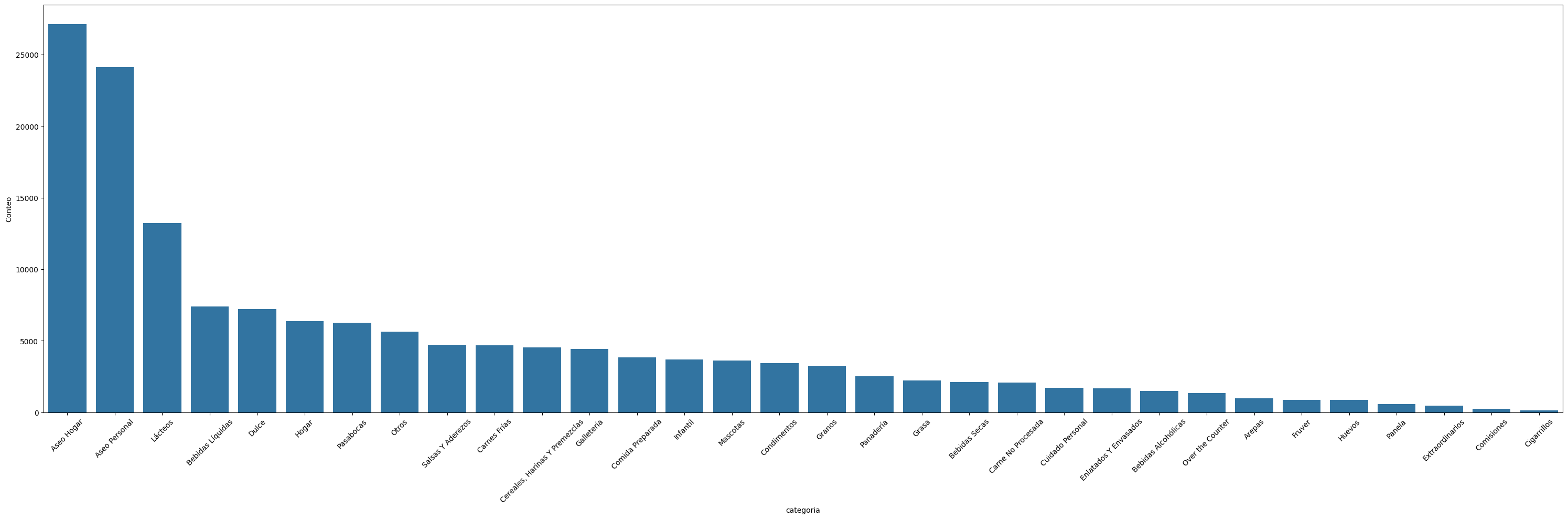
****

**Diagrama de cajas y bigotes para las variables numéricas:**



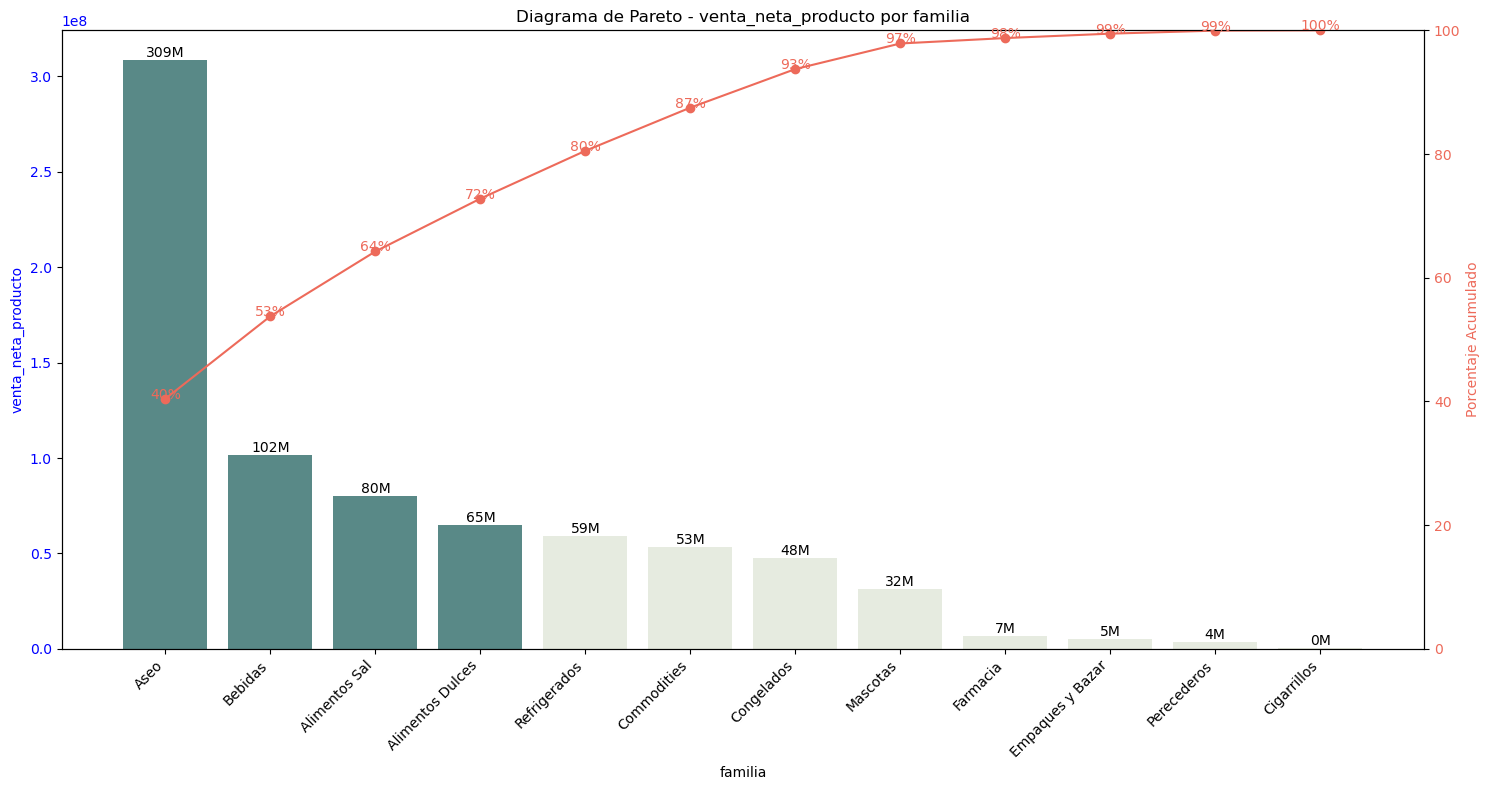
* mnt\_total\_a\_pagar: Se puede observar que a mayoría de los valores están concentrados cerca de la parte baja del rango, con algunos valores atípicos que se extienden mucho más allá del tercer cuartil, mientras que hay algunos valores negativos que alcanzan aproximadamente -200.000. Estos valores negativos representan devoluciones.
* cant\_items: Se puede observar que la mediana es baja, por lo que se podría decir que la mayoría de las observaciones tienen una cantidad baja de ítems. También se encuentran varios valores atípicos en la parte superior, que muestran pedidos con una cantidad inusualmente alta de ítems.
* cantidad\_vendida: La mayoría de las observaciones están concentradas en la parte baja del rango, con varios valores atípicos hacia el extremo superior.
* venta\_neta\_producto: Se puede observar una alta concentración de datos en la parte baja, con un rango extenso de valores atípicos positivos.

**Distribución por categoría:**

****

Se puede observar que la categoría aseo hogar y aseo personal son las compras más frecuentes, por su parte lácteos, bebidas liquidas y dulce tienen una frecuencia regular pero no tan alta comparada con las dos categorías de aseo.

**Diagrama de Pareto:**

****

Se puede observar que la categoría aseo es la más dominante, con un total de 309 millones en ventas netas, lo que representa el 40% del total acumulado, la categoría de bebidas sigue con 102 millones, y junto con aseo, cubre el 53% del total acumulado, otras categorías como alimentos sal, alimentos dulces y refrigerados también contribuyen significativamente, alcanzando juntos un 80% del total acumulado.

Propuesta metodológica

En esta sección debe mostrarse cómo el proyecto incorpora aprendizaje no supervisado y qué algoritmo planean utilizar. (Este algoritmo puede ser preliminar y pueden modificarlo para la entrega final, si ven que a lo largo del curso incorporan nuevas herramientas que pueden resultar más apropiadas)

Referencias

* 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