**Logotipo

Descripción generada automáticamente**

**Predicción de la elasticidad de la demanda de un grupo de referencias del negocio de papel higiénico de una cadena de supermercados de Estados Unidos de América**

Laura Melisa Gaviria Tapias

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Nombres completos, Título académico más alto

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2025

|  |  |
| --- | --- |
| **Cita** | (Gaviria Tapias, 2025) |
| **Referencia**  **Estilo APA 7 (2020)** | Gaviria Tapias (2023). Predicción de la elasticidad de la demanda de un grupo de referencias del negocio de papel higiénico de una cadena de supermercados de Estados Unidos de América Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteIX.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

Tabla de contenido

[Resumen 7](#_Toc198850533)

[Abstract 8](#_Toc198850534)

[1. Descripción del problema 9](#_Toc198850535)

[1.1. Problema de negocio 10](#_Toc198850536)

[1.2. Aproximación desde la analítica de datos 11](#_Toc198850537)

[1.3. Origen de los datos 12](#_Toc198850538)

[1.4. Métricas de desempeño 12](#_Toc198850539)

[2. Objetivos 14](#_Toc198850540)

[2.1. Objetivo general 14](#_Toc198850541)

[2.2. Objetivos específicos 14](#_Toc198850542)

[3. Datos 15](#_Toc198850543)

[3.1. Datos originales 15](#_Toc198850544)

[3.2. Datsets 16](#_Toc198850545)

[3.3. Analítica descriptiva 17](#_Toc198850546)

[Referencias 23](#_Toc198850547)

**Lista de tablas**

[Tabla 1 - Campos de los dataset utilizados 18](#_Toc198847986)

[Tabla 2 - Estadísticas descriptivas del dataset original, eliminando los 0 de variables move y price 19](#_Toc198847987)

**Lista de figuras**

[Ilustración 1- Distribución de variable price antes de limpieza 18](#_Toc198850477)

[Ilustración 2 -Distribución de variable price después de limpieza 18](#_Toc198850478)

[Ilustración 3 - Distribución de variable move (Cantidad ventas) después de limpieza 18](#_Toc198850479)

[Ilustración 4 - relación entre variable categórica sale (tipo de promoción) y move (cantidad de ventas) 19](#_Toc198850480)

[Ilustración 5 - relación entre variable categórica sale(tipo de promoción) y profit (ganancias) 20](#_Toc198850481)

[Ilustración 6 - relación entre variable categórica sale (tipo de promoción) y precio. 20](#_Toc198850482)

[Ilustración 7 - Tendencia de las ventas a lo largo del tiempo 21](#_Toc198850483)

[Ilustración 8 - Tendencia de las ganancias a lo largo del tiempo 21](#_Toc198850484)

[Ilustración 9 - Tendencia del precio a lo largo del tiempo 21](#_Toc198850485)

**Siglas, acrónimos y abreviaturas**

**APA** American Psychological Association

**ERIC** Education Resources Information Center

**Esp.** Especialista

**MP** Magistrado Ponente

**MSc** Magister Scientiae

**Párr.** Párrafo

**PhD** Philosophiae Doctor

**PBQ-SF** Personality Belief Questionnaire Short Form

**PostDoc** PostDoctor

**UdeA** Universidad de Antioquia

# Resumen

La gestión eficiente de precios y promociones en el sector minorista es un factor crítico para la rentabilidad, especialmente en mercados competitivos como el de supermercados en Estados Unidos de América. El problema central radica en la necesidad de comprender con precisión cómo la demanda de productos esenciales, como el papel higiénico, reacciona a los cambios en el precio. Las decisiones actuales, a menudo basadas en análisis históricos simples o intuición, no logran capitalizar completamente el potencial de optimización de ingresos y cuota de mercado.

Esta monografía tuvo como objetivo principal predecir la elasticidad de la demanda de un grupo de referencias de papel higiénico para una cadena de supermercados de EE. UU., utilizando técnicas avanzadas de *machine learning*. Para ello, se empleó una aproximación sistemática que incluyó una revisión exhaustiva de la literatura para identificar modelos predictivos de alta precisión. Se realizó una analítica descriptiva detallada sobre un dataset histórico de transacciones, originado por la Universidad de Chicago Booth y Dominick’s (1989-1994), para caracterizar las variables clave como precio, promociones y volumen de ventas. Posteriormente, se entrenaron y evaluaron cuatro modelos de regresión seleccionados (Decision tree, Gradient Boosting, LSTM, MOMENT (T5)) utilizando métricas como el Error Cuadrático Medio (MSE), el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R2).

**Resultado y conclusiones, se pondrían luego cuando evalue los modelos**.

<https://github.com/LMelisaGaviria/Monografia_Proyeccion_Demanda-Precio.git>

*Palabras clave*: elasticidad, precio, demanda, *retail*, papel higiénico, *machine learning*.

# Abstract

Efficient management of pricing and promotions in the retail sector is a critical factor for profitability, especially in competitive markets like U.S. supermarkets. The core problem lies in the need to precisely understand how the demand for essential products, such as toilet paper, reacts to price changes. Current decisions, often based on simple historical analysis or intuition, fail to fully capitalize on the potential for optimizing revenue and market share.

This academic work main objective was to predict the demand elasticity of a group of toilet paper SKUs for a U.S. supermarket chain using advanced machine learning techniques. To achieve this, a systematic approach was employed, which included an exhaustive literature review to identify high-precision predictive models. A detailed descriptive analysis was performed on a historical transaction dataset, originally compiled by the University of Chicago Booth and Dominick's (1989-1994), to characterize key variables such as price, promotions, and sales volume. Subsequently, four selected regression models (Decision Tree, Gradient Boosting, LSTM, and MOMENT (T5)) were trained and evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R2).

*Keywords***:** Elasticity, Price, Demand, Retail, Toilet Paper, Machine Learning.

# Descripción del problema

El mercado global de productos de consumo masivo, incluyendo el papel higiénico, se caracteriza por una creciente sensibilidad al precio por parte de los consumidores, especialmente en economías desarrolladas como la de Estados Unidos de América (Nielseniq, 2023). Factores como la inflación y la fluctuación del poder adquisitivo han impulsado a los consumidores a buscar activamente productos accesibles y promociones atractivas, lo que ha llevado a un aumento en la migración hacia marcas propias y formatos de descuento (Nielseniq, 2023). En este entorno dinámico, comprender cómo responde la demanda ante variaciones en el precio es un factor crítico para la rentabilidad de las cadenas de supermercados.

La **elasticidad precio de la demanda**, definida como el cambio porcentual en la cantidad demandada de un bien o servicio frente a un cambio porcentual en su precio (Rebollar Rebollar, Herdandez Martínez, Guzmán Soria, Rebollar Rebollar , & González Razo, 2021), es una métrica fundamental para optimizar estrategias de fijación de precios, diseño de promociones y gestión de inventarios (Bendhi, 2025). Sin embargo, en el canal *retail*, las decisiones comerciales sobre precios y promociones a menudo se basan en análisis históricos simplificados, sin incorporar modelos analíticos que permitan anticipar con precisión el comportamiento del consumidor (Bendhi, 2025) .Adicionalmente, los enfoques econométricos tradicionales, si bien son valiosos, presentan limitaciones en la captura de la no linealidad y la alta dimensionalidad de los datos de consumo actuales (Bendhi, 2025).

Ante esta realidad, la aplicación de técnicas de analítica avanzada y modelosde *machine learning* surgen como una ventaja competitiva, ya que permiten capturar relaciones complejas y patrones ocultos en los datos (Bendhi, 2025). Diversos estudios recientes han demostrado la efectividad y aplicabilidad de estas herramientas para optimizar decisiones de precios y promociones en entornos reales:

* (Membreño, 2024), por ejemplo, aplicó modelos de *machine learning* y econométricos para analizar la elasticidad de la demanda en la industria aérea, evidenciando la superioridad de los enfoques de aprendizaje automático frente a la complejidad y no linealidad del comportamiento del consumidor.
* (Kubišta, 2020) utilizó redes neuronales para evaluar el comportamiento de los clientes y estimar la elasticidad del precio en un conjunto de datos de comercio electrónico.
* (Nieto, 2019) desarrolló un modelo dinámico de *Pricing* para el canal *Retail* en la industria alimentaria, empleando algoritmos de series de tiempo supervisados para predecir la demanda, maximizar ganancias y analizar el efecto de las promociones.

En este contexto, la presente monografía se propone abordar el desafío de la predicción de la elasticidad de la demanda para un grupo de referencias de papel higiénico en una cadena de supermercados de Estados Unidos de América**.** Al aplicar modelos de aprendizaje automático, se busca generar predicciones más precisas y robustas, fortaleciendo así la toma de decisiones comerciales basadas en datos para optimizar las estrategias de precios y promociones en este segmento de productos esenciales.

## Problema de negocio

El negocio minorista, particularmente en el sector de supermercados, opera en un entorno altamente competitivo donde los minoristas a menudo se impulsan entre sí a frecuentes recortes de precios (y bajos márgenes de ganancia) (Vaja, 2015).Para una cadena de supermercados, la gestión eficiente de los precios y las promociones de productos esenciales como el papel higiénico es crucial para la rentabilidad y el mantenimiento de su cuota de mercado. Actualmente, las decisiones comerciales relacionadas con la fijación de precios y la planificación de promociones para estas referencias clave a menudo se basan en el análisis de datos históricos de ventas y en la intuición de los gestores, un enfoque que enfrenta limitaciones crecientes ante la complejidad del mercado moderno.

Este enfoque, aunque tradicional, presenta limitaciones significativas. Sin una comprensión precisa de la elasticidad de la demanda, lo que podría llevar a estrategias de pricing equivocadas que pueden derivar en:

* Ineficacia en las promociones: Las promociones pueden no generar el impacto esperado en el volumen de ventas o en la rentabilidad si la respuesta del consumidor al cambio de precio no se anticipa correctamente.
* Gestión de inventario ineficiente**:** Una predicción imprecisa de la demanda resultante de cambios de precios puede llevar a excesos de inventario (costos de almacenamiento) o escasez (pérdida de ventas).
* Reacciones tardías al mercado: La falta de un modelo predictivo robusto impide una respuesta ágil a los cambios en el comportamiento del consumidor o a las acciones de la competencia.
* Pérdida de cuota de mercado: La falta de alineación entre el precio y el valor percibido por los consumidores puede generar una respuesta negativa, incentivando a buscar alternativas en la competencia. Esta migración de la demanda conlleva directamente a una pérdida sostenida de cuota de mercado a largo plazo.

En síntesis, el problema de negocio radica en la necesidad de optimizar las decisiones de precios y promociones para el papel higiénico a través de una comprensión profunda y predictiva de la elasticidad de su demanda. La implementación de modelos de *machine learning* para predecir esta elasticidad ofrecería a la cadena de supermercados una herramienta valiosa para tomar decisiones comerciales más informadas y basadas en datos, lo que se traduciría en una mejora directa de sus márgenes de beneficio, la satisfacción del cliente y su posición competitiva en el mercado.

## Aproximación desde la analítica de datos

El abordaje de este estudio se centró en una aproximación sistemática desde la analítica de datos, combinando la revisión de literatura con la aplicación práctica de modelos predictivos para la estimación de la elasticidad de la demanda. Este enfoque metodológico se estructura en las siguientes fases clave:

En primer lugar, se realizó una exhaustiva revisión de literatura especializada en el campo de la predicción de la demanda y la elasticidad precio. El objetivo fue identificar y comprender los modelos de regresión y las técnicas de *machine learning* que han demostrado la mejor precisión y aplicabilidad en contextos similares al del negocio minorista. Esta fase permitió un entendimiento profundo de las metodologías existentes y sus métricas de evaluación de desempeño, tales como el Error Cuadrático Medio (MSE), el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R2), con el fin de seleccionar hasta cuatro modelos candidatos para una evaluación posterior.

Posteriormente, se llevó a cabo una analítica descriptiva detallada del conjunto de datos proporcionado por la cadena de supermercados. Esta fase fue crucial para caracterizar las variables independientes clave, como el precio y las promociones, e identificar patrones, tendencias y comportamientos relevantes que puedan influir en la demanda de las referencias de papel higiénico. La comprensión de estas características iniciales permitió una preparación de datos adecuada y una mejor interpretación de los resultados de los modelos predictivos.

Finalmente, se realizó la evaluación y entrenamiento de los cuatro modelos de regresión o *machine learning* previamente seleccionados. Cada modelo será aplicado al conjunto de datos, y su desempeño predictivo de la demanda será rigurosamente evaluado utilizando las métricas definidas (MSE, MAE y R2). Este proceso permitirá identificar el modelo con la mayor precisión y robustez para la predicción de la elasticidad de la demanda, sentando las bases para su potencial aplicación en la optimización de estrategias comerciales.

## Origen de los datos

El dataset utilizado en este proyecto fue construido por la Escuela de negocios Booth de la Universidad de Chicago en colaboración con la antigua cadena de supermercados Dominick’s en Estados Unidos de América para investigar en la gestión de estanterías y Pricing (CHICAGO BOOTH Kilts Center for Marketing, 2013). Para la construcción de los datos se tomaron muestras aleatorias de más de 25 categorías en las diferentes tiendas de la cadena Dominick´s durante los años 1989 y 1994, el dataset se encuentra en la sección de *Research and Data* de la página web de *Kilt Center for Marketing* de la Universidad de Chicago.

## Métricas de desempeño

Para que el modelo se considerara exitoso y viable para su implementación en un entorno empresarial real, se establecieron y se cumplieron los siguientes umbrales mínimos:

Métricas de *Machine Learning*:

* MSE y MAE: Se considera un buen modelo un valor de MSE y MAE menor o igual al 7% del valor promedio de la demanda
* R2: El Coeficiente de Determinación (R2) para que el modelo pueda ser considerado bueno debe ser igual o superior a 0.75.

# Objetivos

## Objetivo general

Evaluar la precisión de 4 modelos de machine learning para la predicción de la demanda en un conjunto de referencias de papel higiénico aplicadas a una fuente de datos proveniente de una cadena de supermercados de Estados Unidos de América.

## Objetivos específicos

* Identificar, según la revisión de literatura, los modelos de regresión con mayor precisión en la predicción de la demanda a partir de variables como precios y promociones. Esto incluye la comparación de sus métricas (Error Cuadrático Medio - MSE, Error Absoluto Medio - MAE, y Coeficiente de Determinación - R2) para seleccionar hasta cuatro modelos candidatos.
* Realizar analítica descriptiva para el conjunto de datos de la cadena de supermercados de Estados Unidos de América. Esto incluye la identificación de patrones, tendencias y comportamientos de las variables independientes (Precio, Promociones) para caracterizar su influencia en la demanda de las referencias de papel higiénico.
* Entrenar y evaluar los cuatro modelos seleccionados con el conjunto de datos, utilizando las métricas definidas (MSE, MAE y R2) para identificar el modelo con el mejor desempeño predictivo de la demanda.
* Analizar la influencia relativa de las variables precio y promociones en la predicción de la elasticidad de la demanda de las referencias de papel higiénico, utilizando el modelo con mejor desempeño para cuantificar su impacto en el comportamiento del consumidor.
* Proponer un modelo predictivo validado y escalable de elasticidad de la demanda, derivado del análisis y las métricas obtenidas, que pueda ser implementado en el canal retail para optimizar las estrategias de precios y promociones en el negocio de papel higiénico.

# Datos

## Datos originales

El presente trabajo de monografía tiene como objetivo analizar la elasticidad precio de la demanda de productos de papel higiénico (Bathroom Tissue) utilizando el conjunto de datos Dominicks-Manual-and-Codebook\_KiltsCenter. Este dataset, proporcionado por el Kilts Center for Marketing de la Universidad de Chicago, contiene información real detallada sobre precios, volúmenes de venta y características de productos comercializados en la cadena de supermercados Dominick's.

El conjunto de datos cuenta con 2 bases de datos:

1. UPC (Universal product code) corresponde datos de información de los productos, códigos, descripción, tamaño. Contiene los datos maestros (upctti.cvs) tiene 128 filas y 6 columnas con un peso de 7KB
2. Archivo de movimientos(wtti.csv), contiene data de ventas semanales por UPC, de 5 años, precios, promociones, margen de ganancia, tienda, otros. Cuenta con 11 columnas y 1159016 filas.

**Descripción de las variables:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Descripción** | **Tipo** | ***dataset*** |
| UPC | *Universal Product code* | numérico | wtti.csv |
| STORE | Número de tienda | numérico | wtti.csv |
| WEEK | Número de semana | numérico | wtti.csv |
| MOVE | Cantidad vendida | numérico | wtti.csv |
| PRICE | Precio *retail* | numérico | wtti.csv |
| QTY | Número de artículos juntos | numérico | wtti.csv |
| PROFIT | Margen de ganancia | numérico | wtti.csv |
| SALE | Código de promoción (B(Bono), C(Cupón), S(reducción simple de precio)) | Texto | wtti.csv |
| OK | 1 para datos válidos, 0 datos sospechosos numérico | Binario | wtti.csv |
| UPC | *Universal Product code* | numérico | upctti.cvs |
| DESCRIP | Nombre descriptivo de la UPC | Texto | upctti.cvs |
| SIZE | Tamaño | Texto | upctti.cvs |
| CASE | Cantidad por caja | numérico | upctti.cvs |
| NITEM | Número del Item | numérico | upctti.cvs |

Tabla 1 - Campos de los dataset utilizados

El conjunto de datos original presentaba limitaciones debido a su antigüedad y posible incompletitud en la captura de los efectos promocionales. Para abordar estas limitaciones y aumentar el conjunto de datos para un análisis más robusto, se implementó un proceso de síntesis de datos. Este proceso implicó la generación de datos sintéticos que reflejaran las propiedades estadísticas del conjunto de datos original, extendiendo así el alcance del conjunto de datos, particularmente en términos de cobertura temporal.

## Datsets

Para la construcción del dataset se realizaron los siguientes procedimientos:

Se crearon datos sintéticos utilizando Python y las bibliotecas pandas, numpy y scikit-learn. Los pasos clave de la metodología de síntesis de datos se describen a continuación:

1. Inicialización del DataFrame Sintético. Se inicializó un DataFrame de pandas (synthetic\_data) con 1’000,000 filas para servir como base para los datos sintéticos.
2. Generación de Variables Categóricas. Las variables categóricas *upc, sale y store* se generaron mediante el muestreo de las distribuciones empíricas observadas en el conjunto de datos original (*df\_move\_cleaned*). Esto aseguró que los datos sintéticos retuvieran la composición categórica de los datos reales.
3. Generación de la Variable Temporal. Para extender el alcance temporal de los datos, la variable *week* se generó utilizando una distribución aleatoria uniforme, comenzando desde la semana inmediatamente posterior a la última semana en el conjunto de datos original. Este enfoque simuló semanas futuras, permitiendo el análisis de tendencias más allá del marco temporal del conjunto de datos original.
4. Generación de Variables Numéricas. Las variables numéricas *move, price, qty y profit* se generaron condicionalmente en función de la variable sale y, en algunos casos, la variable week. Esta generación condicional tenía como objetivo capturar las relaciones entre las actividades promocionales y estas métricas clave.
5. Se emplearon distribuciones de probabilidad específicas (p. ej., gamma, log-normal, normal) para simular la variabilidad y las características de distribución de estas variables. Los parámetros de estas distribuciones a veces se ajustaron para reflejar posibles tendencias temporales.
6. Concatenación de Conjuntos de Datos. El conjunto de datos sintético (*synthetic\_data*) se concatenó con el conjunto de datos original (*df\_move\_cleaned*) para crear un conjunto de datos combinado (*combined\_data*). Este conjunto de datos combinado se utilizó para análisis posteriores.
7. En los pasos anteriores también se realizó la limpieza de los datos: eliminación o reemplazo de datos *null, nan* y 0.
8. Tratamiento de outliers también fue necesario antes de definir las distribuciones y hacer los análisis descriptivos de las variables numéricas y crear sus datos sintéticos.

## Analítica descriptiva

El conjunto de datos se sometió a varios pasos de preprocesamiento para garantizar su idoneidad para el análisis. Estos pasos incluyeron la limpieza de datos, el manejo de valores faltantes y la síntesis de datos. Los datos originales tenían las siguientes características:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estadística** | **move** | **price** | **qty** | **profit** |
| count | 11599016 | 11599016 | 11599016 | 11599016 |
| mean | 75.57 | 2.10 | 1.03 | 16.64 |
| std | 294.02 | 1.84 | 0.30 | 11.15 |
| min | 1.00 | 0.19 | 1.00 | -99.88 |
| 25% | 12.00 | 0.99 | 1.00 | 9.77 |
| 50% | 28.00 | 1.48 | 1.00 | 16.86 |
| 75% | 63.00 | 2.59 | 1.00 | 22.49 |
| max | 69394.00 | 838.16 | 7.00 | 99.94 |

Tabla 2 - Estadísticas descriptivas del dataset original, eliminando los 0 de variables move y price

* Limpieza de Datos. Se eliminaron las columnas irrelevantes o redundantes del conjunto de datos para agilizar el análisis y reducir la dimensionalidad. Se abordaron las inconsistencias en las variables categóricas estandarizando las representaciones. Específicamente, los valores nulos en la variable ‘*sale*’, inicialmente representados como '*nan*' y '*NaN*', se unificaron bajo la etiqueta "no\_aplica".
* Se hizo una identificación inicial de valores extremos y su tratamiento que permitió identificar mejor las distribuciones de las principales variables.

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 1- Distribución de variable price antes de limpieza

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 2 -Distribución de variable price después de limpieza

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 3 - Distribución de variable move (Cantidad ventas) después de limpieza

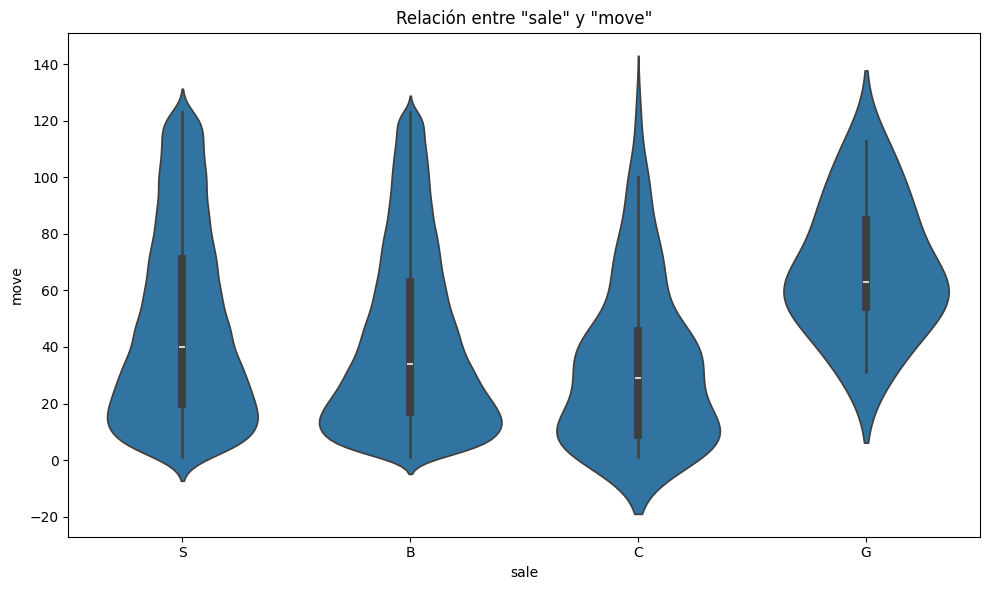
* Manejo de Valores Faltantes. Se tuvo 2 casos: para las ventas que registraban un valor cero y precio cero se determinó eliminar estos datos y para la variable sale donde el campo era vacío se trataron como una categoría separada.
* Síntesis de Datos. Para aumentar el conjunto de datos y abordar las limitaciones relacionadas con su alcance temporal, se generaron datos sintéticos utilizando una metodología descrita en la comprensión de los datos. Este proceso implicó el muestreo de las distribuciones de los datos originales para crear nuevos puntos de datos que preservaran las propiedades estadísticas del conjunto de datos original al tiempo que extendían su cobertura temporal.

Ilustración - relación entre variable categórica sale (tipo de promoción) y move (cantidad de ventas)

* + La estategia promocional G tiene el mayor número de movimientos(ventas) y también una mediana más alta.
  + Las estategias promocionales S y B presentan distribuciones similares: más concentradas entre 10 y 60 movimientos(ventas), aunque con algunos valores altos.
  + La estategia C promocional tiene una distribución más ancha en la parte inferior, lo que indica más valores bajos de "move" (ventas).

Gráfico

Descripción generada automáticamente

Ilustración - relación entre variable categórica sale(tipo de promoción) y profit (ganancias)

* + Las estrategias promocionales S y B muestran una dispersión mayor y presentan valores negativos, lo cual indica que hay pérdidas asociadas a estas ventas.
  + C y G tienen distribuciones más concentradas, especialmente C, lo que sugiere ganancias más estables.
  + La estrategia promocional G presenta los valores más altos de "profit", lo cual puede indicar que es la venta más rentable.

Gráfico

Descripción generada automáticamente

Ilustración - relación entre variable categórica sale (tipo de promoción) y precio.

* + Las estrategias promocionales S y B presentan una distribución más amplia en precios, con algunos valores extremos (precios altos).
  + La estrategia C parece tener precios más bajos y estables, concentrados en torno a 1.
  + G también tiene precios bajos y poco dispersos, pero levemente superiores a C.

También se realizaron análisis de las tendencias para asegurarse que los datos sintéticos siguieran estas mismas tendencias de las variables numéricas.

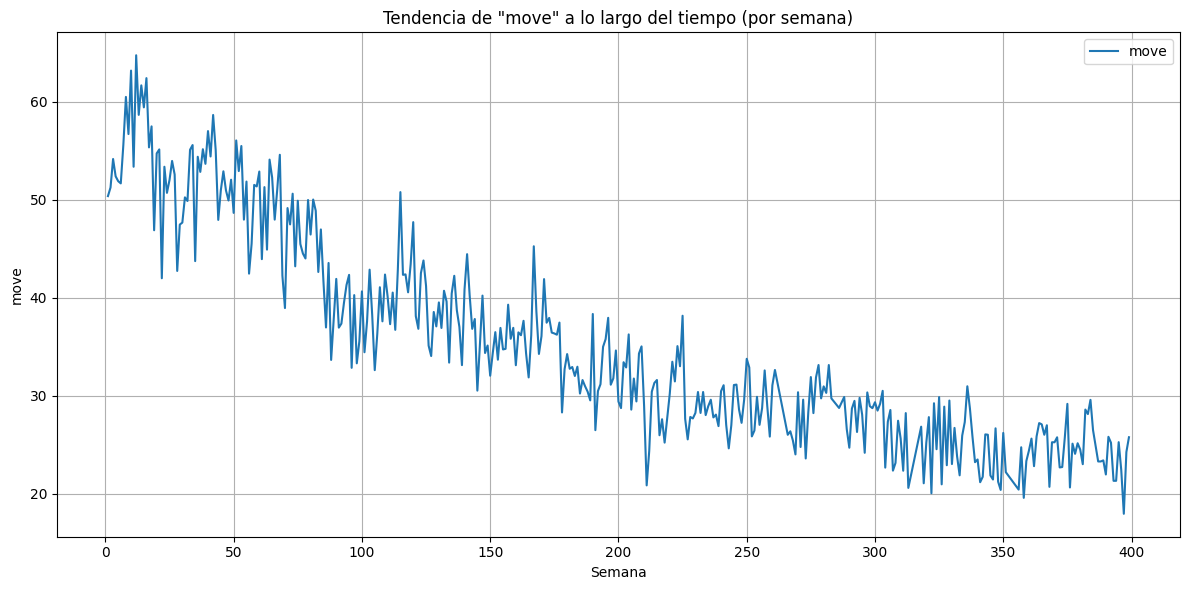


Ilustración - Tendencia de las ventas a lo largo del tiempo

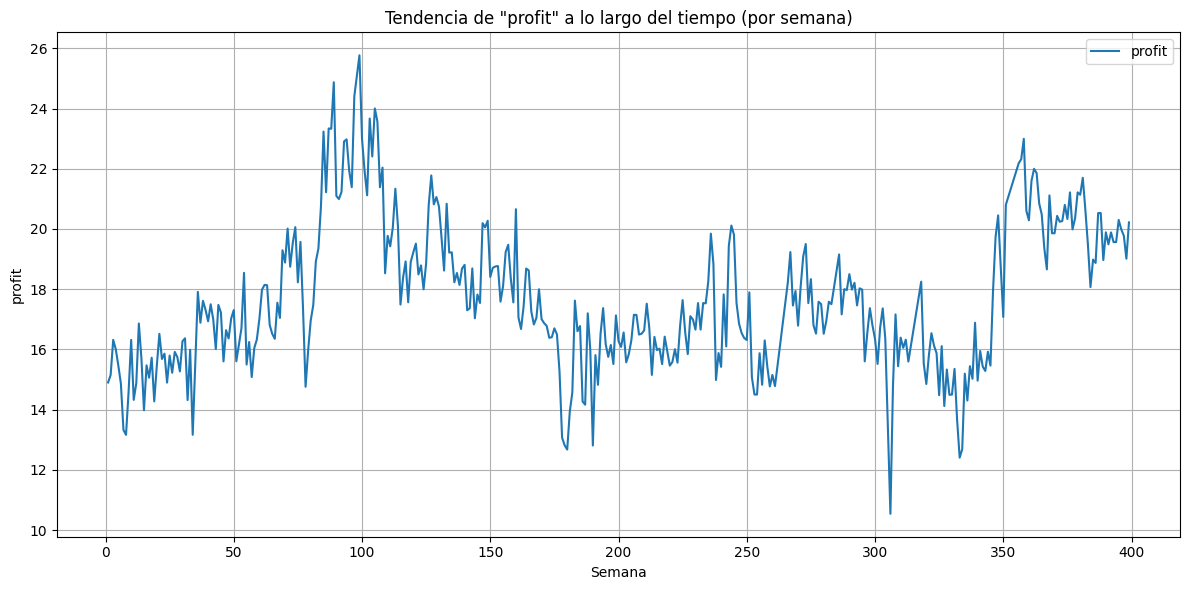


Ilustración - Tendencia de las ganancias a lo largo del tiempo

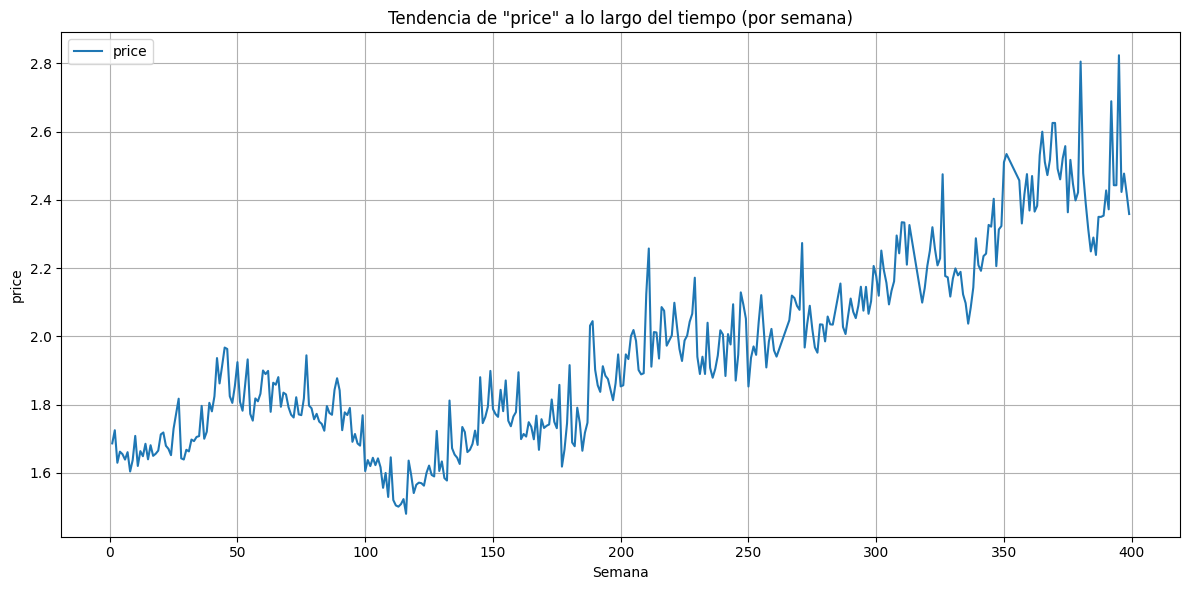


Ilustración - Tendencia del precio a lo largo del tiempo

* Análisis de Valores Atípicos. Se utilizaron diagramas de caja para identificar visualmente posibles valores atípicos en las variables numéricas (*move, price, qty, profit*). Si bien se observaron valores atípicos, se tomó la decisión de retenerlos en el conjunto de datos, considerando su potencial relevancia para capturar eventos extremos o variabilidad en los datos.
* Transformación de Variables. Se calculó una nueva variable, value, que representa el valor total de las ventas, multiplicando *price, move y profit*. Las variables numéricas se estandarizaron utilizando el *StandardScaler* de la biblioteca *scikit-learn*. Esta transformación escaló las variables para tener una media de 0 y una desviación estándar de 1, facilitando la comparación y mejorando el rendimiento de ciertas técnicas analíticas.
* Además, se realizó la creación de variables *dummy* de la variable categórica *sale* para posteriormente utilizarla en los modelos.

# Referencias

Bendhi, M. R. (2025). Demand Elasticity Prediction in Retail Using MOMENT: A T5 - Based Time Series Model. *International Journal of Science and Research (IJSR), 14*.

CHICAGO BOOTH Kilts Center for Marketing. (2013, Julio). Dominick’s Data Manual. *Dominicks-Manual-and-Codebook\_KiltsCenter*. Chicago Illinois.

Kubišta, M. (2020). *A simulation based analysis of price elasticity of demand.* Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague.

Membreño, M. (2024). *Price Elasticity of Air Travel Demand Using Econometrics and Machine Learning to Scale Up Sustainable Aviation Fuels.* Tesis de maestría, Massachusetts Institute of Technology, Cambridge.

Nielseniq. (2023). *nielseniq*. Retrieved from https://nielseniq.com/global/es/insights/report/2023/fmcg-pulse-evolucion-del-consumo-masivo-en-america-latina-en-el-primer-trimestre-de-2023/

Nieto, S. C. (2019). *Dynamic Pricing in the Retail Sector and their Elasticity Effects.* Universidad Pontificia Comillas, Madrid.

Rebollar Rebollar, S., Herdandez Martínez, J., Guzmán Soria, E., Rebollar Rebollar , E., & González Razo, F. (2021). LA ELASTICIDAD PRECIO DE LA DEMANDA: CASO TEORICO NO CORROBORADO. *Revista Mexicana de Agronegocios, 48*.

Vaja, B. R. (2015). RETAIL MANAGEMENT. *International Journal of Research and Analytical Reviews*, 22-28.