



POWERED BY SIMPLON.CO By Educamás

## **CIENCIAS DE DATOS**

### **“Introducción a la Inteligencia Artificial”**

#### **CORTE 15**

#### **Equipo N°1**

**Aarón Mateo Tocora Jiménez<sup>1</sup>**

**Shelly Mayra Alejandra Gaviria Botina<sup>2</sup>**

**Andrés Felipe Mejía Carrillo<sup>3</sup>**

**Joel Valencia<sup>4</sup>**

#### **Docente:**

**David Felipe Rincón Carrillo**

## **PROYECTO ANÁLISIS DE VENTAS HISTÓRICAS DE UN SUPERMERCADO**

**BOGOTÁ D.C**

**Agosto 24 de 2024**

## **Introducción**

En el sector retail, especialmente en los supermercados, la competencia por captar la atención de los clientes es intensa, y las promociones y descuentos estratégicos se han convertido en herramientas clave para impulsar las ventas, especialmente durante fechas especiales y festivos. Sin embargo, estas iniciativas deben ser cuidadosamente gestionadas para evitar el riesgo de canibalizar ventas futuras o no generar el retorno esperado. Además, su impacto varía significativamente entre los diferentes departamentos de un supermercado.

Por ejemplo, una promoción en el departamento de textiles podría no tener el mismo efecto que en el de electrodomésticos, al igual que un festivo podría influir de manera distinta en las ventas de abarrotes comparado con frutas y verduras.

El presente análisis se enfocará en comprender y evaluar el impacto de los festivos y promociones pasadas mediante el análisis de datos históricos de ventas, identificando patrones y tendencias clave. A partir de estos insights, se desarrollará un modelo predictivo para proyectar las ventas futuras basadas en la efectividad de promociones anteriores.

Además, se darán recomendaciones estratégicas al equipo comercial y de marketing para planificar campañas promocionales que maximicen las ventas y minimicen los riesgos asociados.

Finalmente, este estudio permitirá optimizar los recursos invertidos en marketing y descuentos, identificando las mejores prácticas y los periodos más rentables para la ejecución de promociones. De esta manera, el supermercado podrá tomar decisiones más informadas y basadas en datos, incrementando la efectividad de sus campañas promocionales y mejorando el rendimiento general de ventas en fechas clave para cada tienda.

## **Objetivos Específicos:**

### **1. Análisis Exploratorio de Datos:**

- Realizar un análisis exhaustivo de los datos para entender mejor las características y el comportamiento de los clientes.
- Identificar patrones y tendencias en los datos de compra.

### **2. Desarrollo del Modelo de IA:**

- Implementar un modelo de machine learning para proyectar las ventas semanales del siguiente año
- Evaluar el rendimiento del modelo utilizando métricas adecuadas.

### **3. Mejora de la Estrategia de descuentos:**

- Utilizar las predicciones del modelo para generar recomendaciones al equipo comercial y de mercadeo.

### **4. Visualización de Resultados:**

- Diseñar y producir visualizaciones de datos en un dashboard para comunicar los hallazgos del análisis y las predicciones del modelo a los interesados.
- Explicar los resultados de manera sencilla y clara, considerando que el gerente del área de mercadeo no es muy bueno en temas de datos.

## **Metodología**

Se realizó una recopilación de datos en la cual se dispone de tres bases de datos en formato csv, el cual contiene las ventas semanales de 45 tiendas y 99 departamentos, Adicionalmente, se cuenta con información del tamaño y tipo de las tiendas, y de información adicional que permite identificar si hubo descuentos, fue festivo, e incluso data relacionada con la temperatura promedio, el precio del combustible, la tasa de desempleo y el índice de precios del consumidor por cada semana y tienda.

A continuación, se realiza una breve descripción en el significado de cada dato utilizado y encontrado:

Stores data-set.csv	Sales data-set	Features data set.csv
Store: ID de tienda Type: Tipo de tienda (A, B, C...) Size: Tamaño de tienda	Store: ID de tienda Dept: ID del departamento Date: Fecha Weekly_Sales: Ventas semanales IsHoliday: Marca de festivo	Store: ID de tienda Date: Fecha Temperature: Temperatura promedio en F° Fuel_Price: Precio de la gasolina MarkDown1 a 5: Descuentos y Promociones CPI: Índice de precios del consumidor Unemployment: Tasa de desempleo IsHoliday: Marca de festivo

Denominamos así a cada base de datos para una mejor identificación y correlación.

Para realizar la visualización y carga de datos utilizamos Google Colab, verificando las distintas etapas del procesamiento, iniciando por la limpieza de datos en la cuales se verifico en cada Datframe, los valores nulos, duplicados, formato de fechas y la eliminación de datos que no fueran consistentes con las ventas.

Priorizando así mantener la integridad de los datos, evitando sesgos y errores en los resultados del análisis.

En la integración de datos y filtrado, se conjugo la información de las tiendas, ventas y factores externos, eliminado registros que no tuvieran relación con las ventas, agrupando fechas por periodos semanales para que en los resultados nos permita consolidar el modelo predictivo y evitar problemas de colinealidad.

En el análisis descriptivo se identificaron a través de gráficos, patrones en las fechas especiales de mayor consumo o que se realizan promociones para estimular las compras de los clientes, como es a finales de año, detallando el impacto de los descuentos y registrando en la comparación anual, el cambio a lo largo de los años.

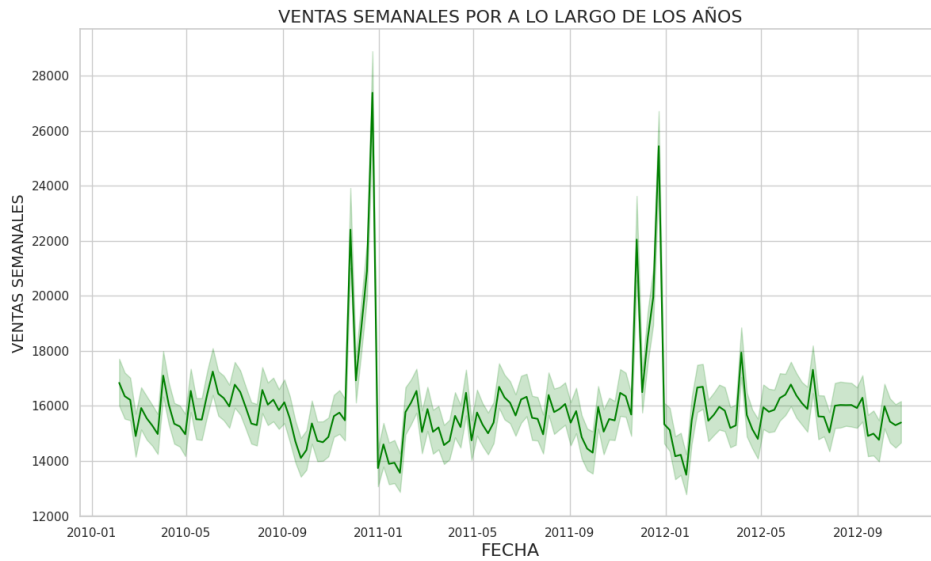
#### Conclusiones y resultados:

En el desarrollo de modelo agrupamos y agregamos los datos por semana y tienda para obtener una tabla en un formato más adecuado para el análisis las cuales se almacenaron en el documento df.join1 generando a partir de estas la predicción de modelos, esto permite resumir las ventas y otras variables clave, facilitando la identificación de patrones y tendencias relevantes para el modelo.

Teniendo en cuenta que el modelo de varianza de la variable dependiente es predecible a partir de la variable independiente, determinando así que el valor más cercano a 1 explica la variabilidad de datos, se obtiene un R2 score por 0.5027391591364971 en el entrenamiento.

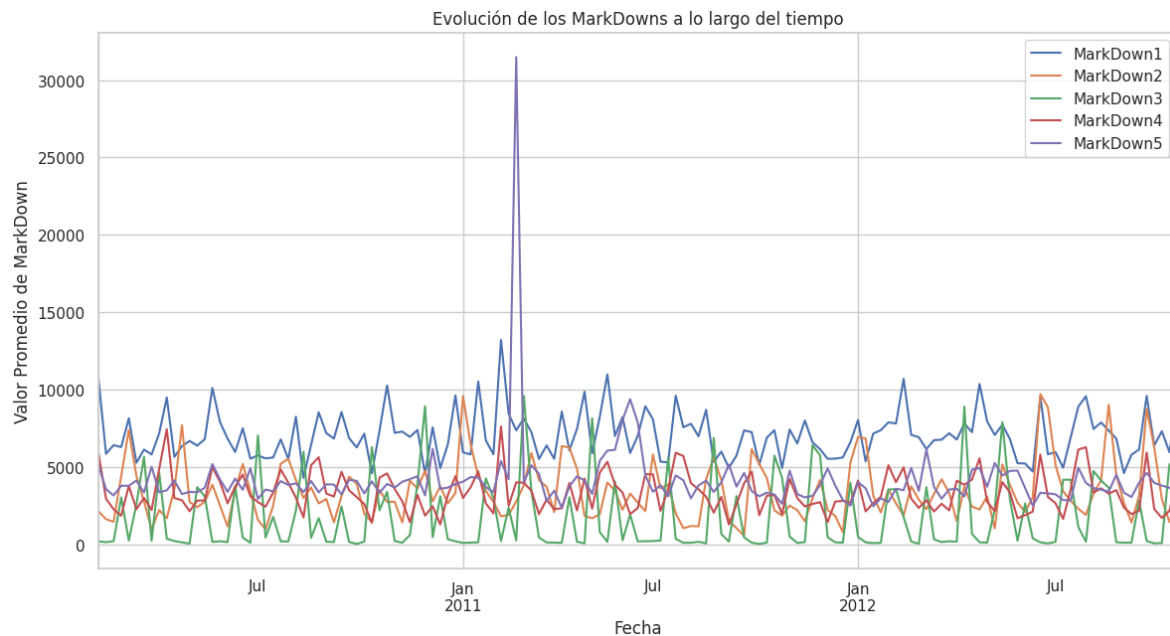
Vemos un cambio considerable en la prueba de la base df.join1 al 100% de un R2 0.558253580272379 acercándonos a la variabilidad de los datos del modelo.

Opcionalmente se prueba el modelo polinómico con un  $R^2$  Polinómico: 0.229419428605373, demuestra que a pesar de su complejidad no logra capturar algunas características lineales, dando a entender que el modelo lineal en este caso tiene un mayor desempeño.



Se analiza de 2010 a 2012, donde se evidencia un pico de crecimiento en 2010, 09 al año 2011,01 al igual que en el año 2011.09 al 2012,01, lo que indica un aumento de ventas en vísperas de año, y un incremento gradual de ventas en tres años, indicando una sostenibilidad en ventas del supermercado.

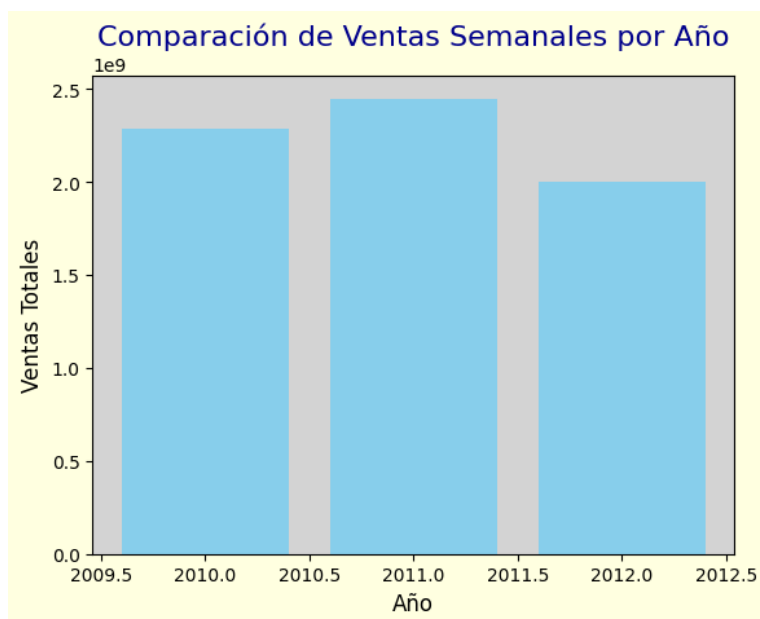
Evolución de descuentos a lo largo del tiempo:



Teniendo en cuenta las promociones realizadas se observa un patrón recurrente en ciertas fechas del año que pueden coincidir con descuentos en fechas especiales para fomentar las ventas en el consumidor.

Cada Markdown da a conocer un patrón de descuentos lo cual nos indica que algunos productos podrían tener promociones más frecuentes o de mayor tamaño, esto varía según el artículo, por lo cual se evidencia que se están ajustando los precios en forma del mercado o su competencia, pero sin una estrategia.

Comparación de ventas semanales por año:



Se observa una disminución en el año 2012 comparadas con el anterior año (2011), mostrando un incremento en el año 2010 y 2011, alcanzando sus mayores ventas en el año 201, aunque se observa una disminución parece no afectar la estabilidad de las ventas durante este tiempo.

