Programa
Data Scientist

skills.tech

Pronóstico de ventas WALMART

Secciones.

1. Introducción
2. Investigación
3. Presentación de los datos
4. Análisis exploratorio
5. Metodología utilizada para el desarrollo del proyecto
6. Hallazgos realizados
7. Resultados finales del modelo
8. Conclusiones y retos
9. Referencias y fuentes

1.Introducción

Introducción.

Walmart es una corporación multinacional minorista que cuenta con una amplia cadena de hipermercados. En este proyecto se trabajarán con un dataset que incluye las ventas semanales de 45 tiendas, días festivos, temperatura, CPI, precio de la gasolina, tasa de desempleo y las fechas en las cuales se registraron las ventas.

Problema.

Walmart no cuenta con una herramienta de análisis descriptivo que permita visualizar un modelo de predicciones para la demanda, que le permita establecer estrategias efectivas y optimas para elaborar un plan a largo plazo en el futuro y realizar tomas de decisiones.
La falta de esta herramienta puede generar problemas de desabastecimiento, sobrestock, ineficiencia en la cadena de suministro, costos adicionales, y disminución en la satisfacción del cliente.

Introducción.

Walmart es una corporación multinacional minorista que cuenta con una amplia cadena de hipermercados. En este proyecto se trabajarán con un dataset que incluye las ventas semanales de 45 tiendas, días festivos, temperatura, CPI, precio de la gasolina, tasa de desempleo y las fechas en las cuales se registraron las ventas.

Problema.

Walmart no cuenta con una herramienta de análisis descriptivo que permita visualizar un modelo de predicciones para la demanda, que le permita establecer estrategias efectivas y optimas para elaborar un plan a largo plazo en el futuro y realizar tomas de decisiones.
La falta de esta herramienta puede generar problemas de desabastecimiento, sobrestock, ineficiencia en la cadena de suministro, costos adicionales, y disminución en la satisfacción del cliente.

Introducción.

Objetivos del Proyecto

- Analizar si las ventas se encuentran afectadas por factores de tiempo y espacio.
- Analizar cómo puede aumentar las ventas semanales la inclusión de un día festivo.
- Desarrollar un modelo predictivo de alta precisión, utilizando técnicas de machine learning y series de tiempo.
- Predecir las ventas semanales de toda la cadena.
- Que el modelo desarrollado contribuya significativamente a la planificación y procesos de tomas de decisiones que se traduzcan en una mejoría para la eficiencia y la rentabilidad de la empresa, así como a su capacidad para satisfacer las necesidades de sus clientes.

Introducción.

Objetivos del Proyecto

- Analizar si las ventas se encuentran afectadas por factores de tiempo y espacio.
- Analizar cómo puede aumentar las ventas semanales la inclusión de un día festivo.
- Desarrollar un modelo predictivo de alta precisión, utilizando técnicas de machine learning y series de tiempo.
- Predecir las ventas semanales de toda la cadena.
- Que el modelo desarrollado contribuya significativamente a la planificación y procesos de tomas de decisiones que se traduzcan en una mejoría para la eficiencia y la rentabilidad de la empresa, así como a su capacidad para satisfacer las necesidades de sus clientes.

Walmart

Introducción.

Objetivos del Proyecto

- Analizar si las ventas se encuentran afectadas por factores de tiempo y espacio.
- Analizar cómo puede aumentar las ventas semanales la inclusión de un día festivo.
- Desarrollar un modelo predictivo de alta precisión, utilizando técnicas de machine learning y series de tiempo.
- Predecir las ventas semanales de toda la cadena.
- Que el modelo desarrollado contribuya significativamente a la planificación y procesos de tomas de decisiones que se traduzcan en una mejoría para la eficiencia y la rentabilidad de la empresa, así como a su capacidad para satisfacer las necesidades de sus clientes.

Walmart

Introducción.

Objetivos del Proyecto

- Analizar si las ventas se encuentran afectadas por factores de tiempo y espacio.
- Analizar cómo puede aumentar las ventas semanales la inclusión de un día festivo.
- Desarrollar un modelo predictivo de alta precisión, utilizando técnicas de machine learning y series de tiempo.
- Predecir las ventas semanales de toda la cadena.
- Que el modelo desarrollado contribuya significativamente a la planificación y procesos de tomas de decisiones que se traduzcan en una mejoría para la eficiencia y la rentabilidad de la empresa, así como a su capacidad para satisfacer las necesidades de sus clientes.

Walmart

Introducción.

Objetivos del Proyecto

- Analizar si las ventas se encuentran afectadas por factores de tiempo y espacio.
- Analizar cómo puede aumentar las ventas semanales la inclusión de un día festivo.
- Desarrollar un modelo predictivo de alta precisión, utilizando técnicas de machine learning y series de tiempo.
- Predecir las ventas semanales de toda la cadena.
- Que el modelo desarrollado contribuya significativamente a la planificación y procesos de tomas de decisiones que se traduzcan en una mejoría para la eficiencia y la rentabilidad de la empresa, así como a su capacidad para satisfacer las necesidades de sus clientes.

Walmart

Introducción.

Beneficios del Proyecto:

- Contar con un modelo de machine learning permita predecir la demanda de los productos semanalmente con precisión.
- Disponer de una herramienta que les permita mejorar la planificación y toma de decisiones orientadas la gestión de la cadena de suministro.
- Ahorro en costos de almacenamiento y adquisición de productos.
- Optimización de estrategias de marketing.
- Mejorar la satisfacción del cliente.

Introducción.

Beneficios del Proyecto:

- Contar con un modelo de machine learning permita predecir la demanda de los productos semanalmente con precisión.
- Disponer de una herramienta que les permita mejorar la planificación y toma de decisiones orientadas la gestión de la cadena de suministro.
- Ahorro en costos de almacenamiento y adquisición de productos.
- Optimización de estrategias de marketing.
- Mejorar la satisfacción del cliente.

Introducción.

Beneficios del Proyecto:

- Contar con un modelo de machine learning permita predecir la demanda de los productos semanalmente con precisión.
- Disponer de una herramienta que les permita mejorar la planificación y toma de decisiones orientadas la gestión de la cadena de suministro.
- Ahorro en costos de almacenamiento y adquisición de productos.
- Optimización de estrategias de marketing.
- Mejorar la satisfacción del cliente.

Introducción.

Beneficios del Proyecto:

- Contar con un modelo de machine learning permita predecir la demanda de los productos semanalmente con precisión.
- Disponer de una herramienta que les permita mejorar la planificación y toma de decisiones orientadas la gestión de la cadena de suministro.
- Ahorro en costos de almacenamiento y adquisición de productos.
- Optimización de estrategias de marketing.
- Mejorar la satisfacción del cliente.

Introducción.

Beneficios del Proyecto:

- Contar con un modelo de machine learning permita predecir la demanda de los productos semanalmente con precisión.
- Disponer de una herramienta que les permita mejorar la planificación y toma de decisiones orientadas la gestión de la cadena de suministro.
- Ahorro en costos de almacenamiento y adquisición de productos.
- Optimización de estrategias de marketing.
- Mejorar la satisfacción del cliente.

Introducción.

Retos:

- **Análisis de datos:** Detectar las técnicas aplicadas para el análisis de los datos y realizar su limpieza adecuada de para procesarlos de manera efectiva.
- **Preparación de los datos:** El dataset contiene columnas que representan factores altamente relevantes para el modelo predictivo que se busca desarrollar, como los son los días festivos y la temperatura local por cada tienda, lo que afectaría la complejidad para la generación del modelo de predicción de ventas.
- **Desarrollo del modelo:** Debido a la complejidad del dataset, investigar y desarrollar modelos predictivos para determinar el cual generaría la mayor confiabilidad y precisión que cumpla con el objetivo del proyecto.
- **Presentación de resultados:** Presentar los resultados del proyecto de una manera clara, sencilla y efectiva a las partes interesadas de Walmart, demostrando que el modelo de predicción es funcional y aplicable para los objetivos de la empresa con este proyecto.

Introducción.

Retos:

- **Análisis de datos:** Detectar las técnicas aplicadas para el análisis de los datos y realizar su limpieza adecuada de para procesarlos de manera efectiva.
- **Preparación de los datos:** El dataset contiene columnas que representan factores altamente relevantes para el modelo predictivo que se busca desarrollar, como los son los días festivos y la temperatura local por cada tienda, lo que afectaría la complejidad para la generación del modelo de predicción de ventas.
- **Desarrollo del modelo:** Debido a la complejidad del dataset, investigar y desarrollar modelos predictivos para determinar el cual generaría la mayor confiabilidad y precisión que cumpla con el objetivo del proyecto.
- **Presentación de resultados:** Presentar los resultados del proyecto de una manera clara, sencilla y efectiva a las partes interesadas de Walmart, demostrando que el modelo de predicción es funcional y aplicable para los objetivos de la empresa con este proyecto.

Introducción.

Retos:

- **Análisis de datos:** Detectar las técnicas aplicadas para el análisis de los datos y realizar su limpieza adecuada de para procesarlos de manera efectiva.
- **Preparación de los datos:** El dataset contiene columnas que representan factores altamente relevantes para el modelo predictivo que se busca desarrollar, como los son los días festivos y la temperatura local por cada tienda, lo que afectaría la complejidad para la generación del modelo de predicción de ventas.
- **Desarrollo del modelo:** Debido a la complejidad del dataset, investigar y desarrollar modelos predictivos para determinar el cual generaría la mayor confiabilidad y precisión que cumpla con el objetivo del proyecto.
- **Presentación de resultados:** Presentar los resultados del proyecto de una manera clara, sencilla y efectiva a las partes interesadas de Walmart, demostrando que el modelo de predicción es funcional y aplicable para los objetivos de la empresa con este proyecto.

Introducción.

Retos:

- **Análisis de datos:** Detectar las técnicas aplicadas para el análisis de los datos y realizar su limpieza adecuada de para procesarlos de manera efectiva.
- **Preparación de los datos:** El dataset contiene columnas que representan factores altamente relevantes para el modelo predictivo que se busca desarrollar, como los son los días festivos y la temperatura local por cada tienda, lo que afectaría la complejidad para la generación del modelo de predicción de ventas.
- **Desarrollo del modelo:** Debido a la complejidad del dataset, investigar y desarrollar modelos predictivos para determinar el cual generaría la mayor confiabilidad y precisión que cumpla con el objetivo del proyecto.
- **Presentación de resultados:** Presentar los resultados del proyecto de una manera clara, sencilla y efectiva a las partes interesadas de Walmart, demostrando que el modelo de predicción es funcional y aplicable para los objetivos de la empresa con este proyecto.

2.Investigación

Investigación.

Contexto del proyecto.

Se ha obtenido un Dataset con datos históricos de 45 tiendas de Walmart ubicadas en diferentes regiones, con el registro de las ventas por periodo semanal desde el 05/02/2010 hasta el 02/11/2012. Incluye escenarios relacionados a ciertos eventos y días festivos donde se debe analizar cómo estos factores afectan a las ventas y como se correlacionan entre si.

Días festivos a evaluar:

- **Super Bowl:** 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
- **Labour Day:** 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
- **Thanksgiving:** 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
- **Christmas:** 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

Link del dataset: https://www.kaggle.com/datasets/rutuspatel/walmart-dataset-retail

Investigación.

Proyectos similares investigados.

Retail Sales Forecasting
Short term forecasting to optimize in-store inventories.

Conjunto de datos que contiene muchos datos históricos de ventas de una tienda minorista.
Contiene muchas (Productos) SKU y muchas tiendas.
El reto de este proyecto era encontrar el mejor modelo de predicción de la demanda con el fin de optimizar la cantidad de inventario que deben tener las tiendas para reducir el costo de capital circulante, operativos, evitar pérdidas en ventas, clientes insatisfechos y mala reputación de la marca

Referencias: https://www.kaggle.com/datasets/tevecsystems/retail-sales-forecasting

Investigación.

Proyectos similares investigados.

Bigmart Sales Prediction.

Bigmart Sales Prediction es un problema de regresión en el que se deben analizar y predecir las ventas de Bigmart basándose en varios aspectos del dataset. El objetivo es construir un modelo predictivo y descubrir las ventas de cada producto en su respectiva tienda.
Este proyecto se trabajó con la recopilación de datos de ventas de un año con 1559 productos en 10 tiendas de diferentes ciudades. Con ciertos atributos definidos de cada producto y tienda.

Referencias: https://www.hackersrealm.net/post/bigmart-sales-prediction-analysis-using-python#:~:text=Bigmart%20Sales%20Prediction%20is%20a,product%20at%20their%20respective%20store.

Investigación.

Proyectos similares investigados.

An End-to-End Project on Time Series Analysis and Forecasting with Python.

Un proyecto que consiste en explicar cómo las series de tiempo se utilizan ampliamente para datos no estacionarios, como los económicos, meteorológicos, bursátiles y de ventas al por menor. Utilizando un dataset que incluye detalles de ventas de una tienda a lo largo de sucursales en los Estados Unidos y Canadá.

Referencias: https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

Investigación.

Enfoque de Solución.

1- Carga de datos

2- Entendimiento de los datos

3- EDA

4- Limpieza de datos

5- Preprocesamiento de los datos

6- Selección de Modelo

7- Segmentación de Datos

8- Evaluación de Modelos

3. Presentación de los Datos

Presentación de los Datos.

Walmart Dataset (Retail).

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	5/2/2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12/2/2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	2/19/2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	2/26/2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	5/3/2010	1554806.68	0	46.50	2.625	211.350143	8.106
...	...	...	...	...	...	...	...	...
6430	45	9/28/2012	713173.95	0	64.88	3.997	192.013558	8.684
6431	45	5/10/2012	733455.07	0	64.89	3.985	192.170412	8.667
6432	45	12/10/2012	734464.36	0	54.47	4.000	192.327265	8.667
6433	45	10/19/2012	718125.53	0	56.47	3.969	192.330854	8.667
6434	45	10/26/2012	760281.43	0	58.85	3.882	192.308899	8.667

6435 rows × 8 columns

Fuente: https://www.kaggle.com/datasets/rutuspatel/walmart-dataset-retail

4. EDA

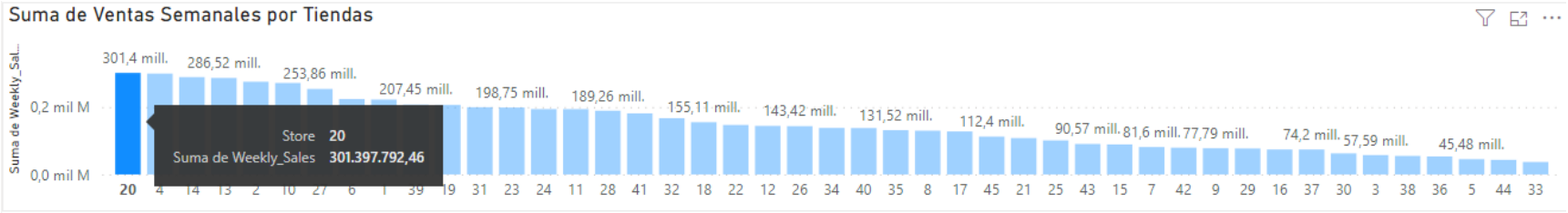
EDA

Propósito

- Identificar patrones y tendencias que se encuentren presentes los datos.
- Evaluar la distribución y la variabilidad de los datos.
- Identificar relaciones y correlaciones entre las variables.
- Comunicar las respuestas a interrogantes de la compañía a través de los resultados del análisis, de manera clara y efectiva a través de visualizaciones y resúmenes estadísticos.
- Identificar las variables más importantes y relevantes para el análisis y la creación del modelo de predicción.
- Identificar posibles problemas de datos, como datos faltantes o inconsistentes.
- Proporcionar información valiosa para la selección y ajuste de los modelos de predicción.

EDA

Interrogantes respecto a Walmart respondidas a través del EDA



Store	Suma de Weekly_Sales
20	301,397,792.46
4	299,543,953.38
14	288,999,911.34
13	286,517,703.80
2	275,382,440.98
10	271,617,713.89
27	253,855,916.88
6	223,756,130.64
1	222,402,808.85
39	207,445,542.47
19	206,634,862.10
31	199,613,905.50
23	198,750,617.85
24	194,016,021.28

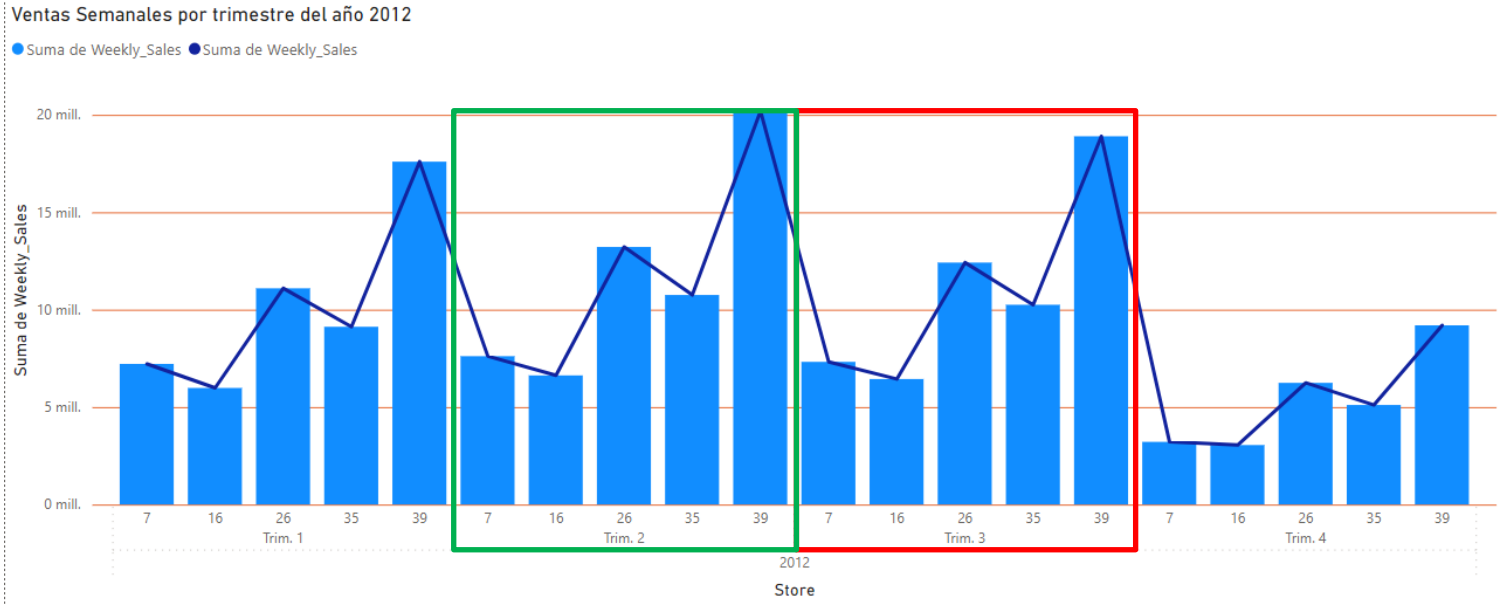
¿Qué tienda tiene el máximo de ventas?

La tienda número 20 presenta la máximo de ventas por un monto de $ 301.397.792,46

EDA

Interrogantes respecto a Walmart respondidas a través del EDA

¿Qué tienda/s tiene una buena tasa de crecimiento trimestral en el tercer trimestre de 2012?

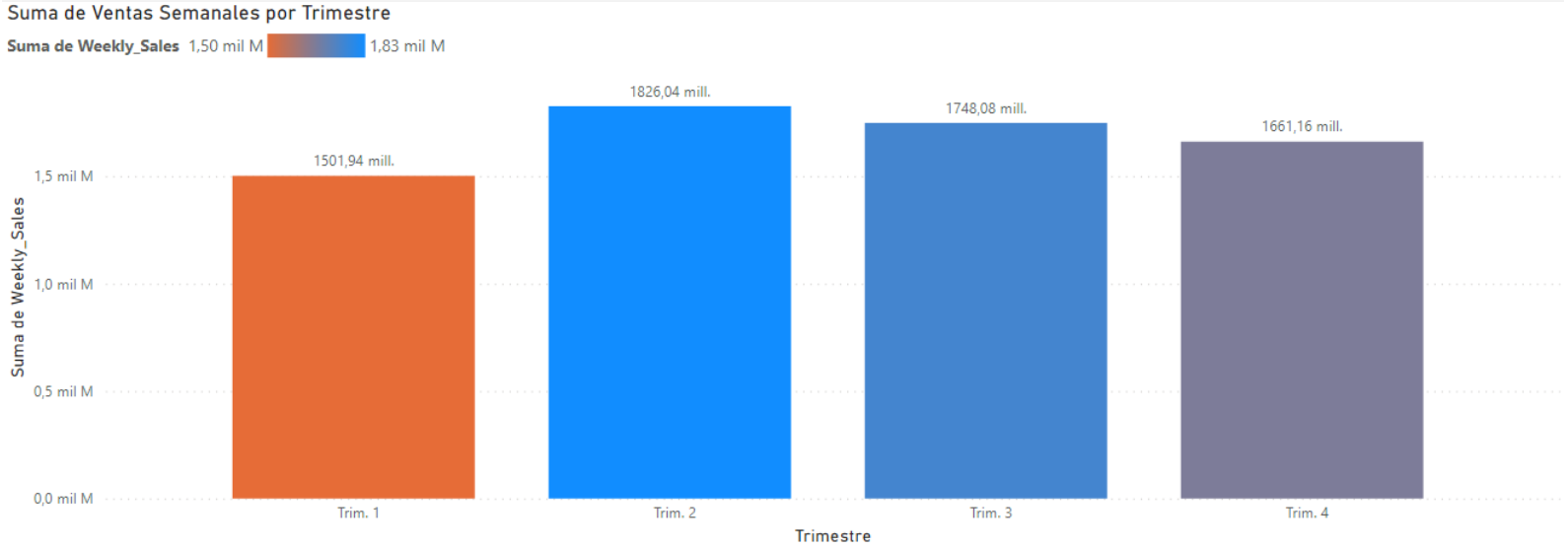


Ninguna tienda presenta crecimiento en el tercer trimestre del año 2012

EDA

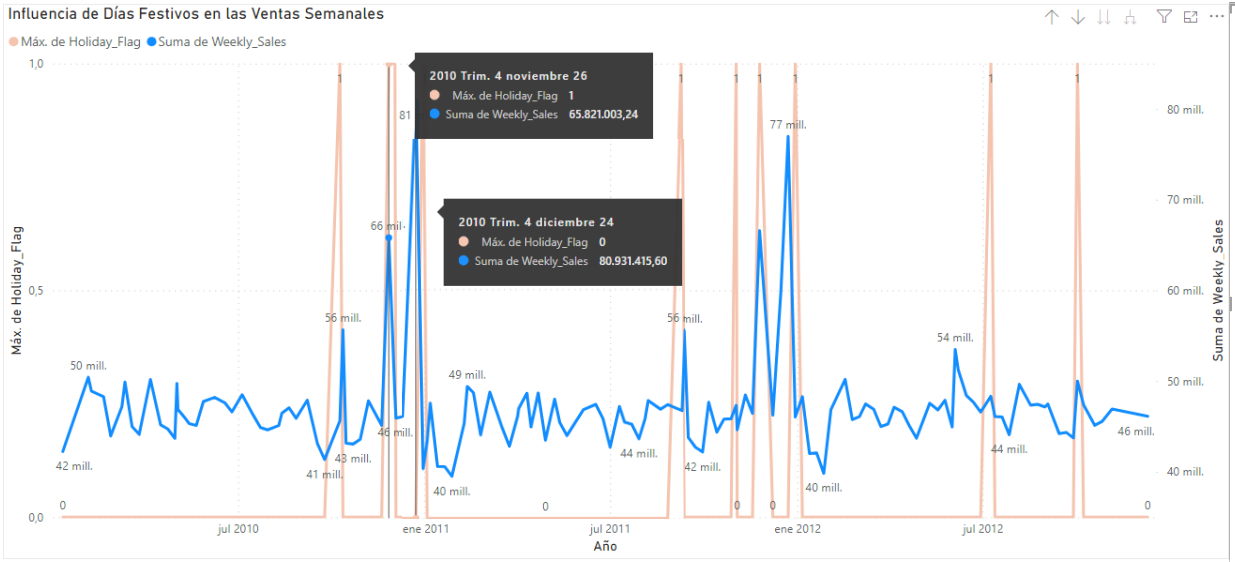
Interrogantes respecto a Walmart respondidas a través del EDA

En términos generales, si se evalúan las ventas semanales por trimestre de todos los años, el mayor ingreso por ventas se percibe en el segundo trimestre.



EDA

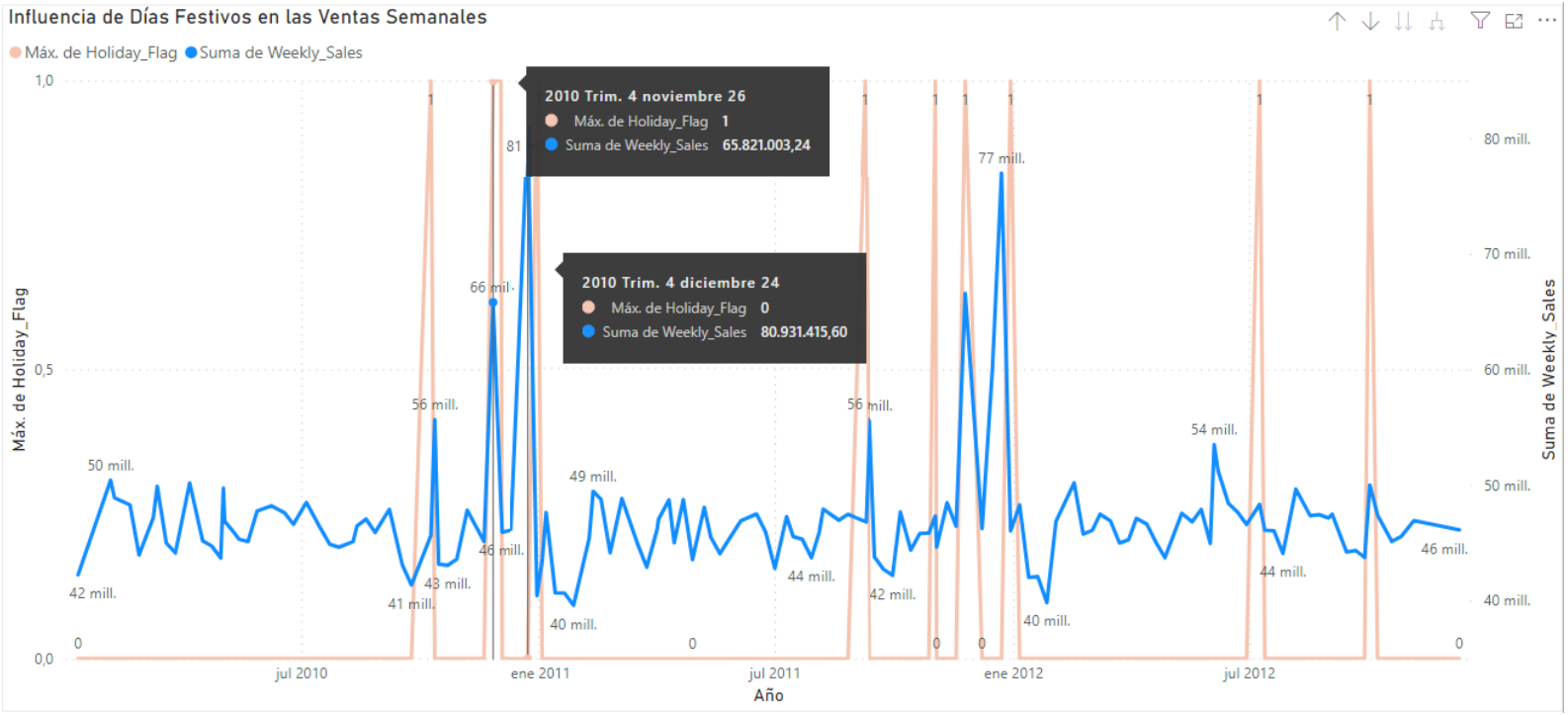
Algunas festividades tienen un impacto negativo en las ventas. Averigüe los días festivos que tienen ventas más altas que las ventas medias en temporada no festiva para todas las tiendas juntas



Año	Mes	Día	Máx. de Holiday_Flag	Suma de Weekly_Sales
2010	octubre	9	1	45,634,397.84
2010	noviembre	26	1	65,821,003.24
2010	diciembre	2	1	48,336,677.63
2010	diciembre	31	1	40,432,519.00
2011	septiembre	9	1	46,763,227.53
2011	noviembre	2	1	47,336,192.79
2011	noviembre	25	1	66,593,605.26
2011	diciembre	30	1	46,042,461.04
2012	julio	9	1	48,330,059.31
2012	octubre	2	1	50,009,407.92
2010	enero	10	0	42,239,875.87
2010	febrero	4	0	50,423,831.26
2010	febrero	7	0	48,917,484.50
2010	febrero	19	0	48,276,993.78
2010	febrero	26	0	43,968,571.13
2010	marzo	9	0	47,194,257.61
2010	marzo	12	0	49,909,027.88
2010	marzo	19	0	44,988,974.64
2010	marzo	26	0	44,133,961.05
2010	abril	6	0	50,188,543.12
2010	abril	16	0	45,183,667.08
2010	abril	23	0	44,734,452.56
2010	abril	30	0	43,705,126.71
2010	mayo	2	0	49,750,740.50
Total			**1**	**6,737,218,987.11**

EDA

Gráfica de relación entre días festivos y su influencia en las ventas semanales.

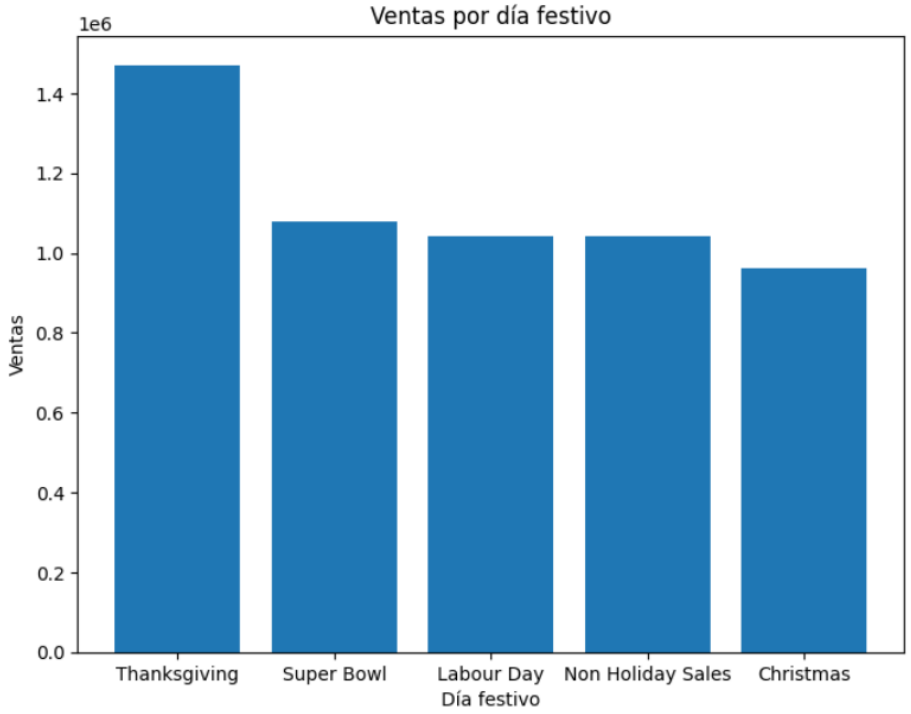


EDA

Año	Mes	Día	Máx. de Holiday_Flag	Suma de Weekly_Sales
2010	octubre	9	1	45,634,397.84
2010	noviembre	26	1	65,821,003.24
2010	diciembre	2	1	48,336,677.63
2010	diciembre	31	1	40,432,519.00
2011	septiembre	9	1	46,763,227.53
2011	noviembre	2	1	47,336,192.79
2011	noviembre	25	1	66,593,605.26
2011	diciembre	30	1	46,042,461.04
2012	julio	9	1	48,330,059.31
2012	octubre	2	1	50,009,407.92
2010	enero	10	0	42,239,875.87
2010	febrero	4	0	50,423,831.26
2010	febrero	7	0	48,917,484.50
2010	febrero	19	0	48,276,993.78
2010	febrero	26	0	43,968,571.13
2010	marzo	9	0	47,194,257.61
2010	marzo	12	0	49,909,027.88
2010	marzo	19	0	44,988,974.64
2010	marzo	26	0	44,133,961.05
2010	abril	6	0	50,188,543.12
2010	abril	16	0	45,183,667.08
2010	abril	23	0	44,734,452.56
2010	abril	30	0	43,705,126.71
2010	mayo	2	0	49,750,740.50
Total			**1**	**6,737,218,987.11**

Tabla de ventas semanales registradas en los días festivos.

EDA

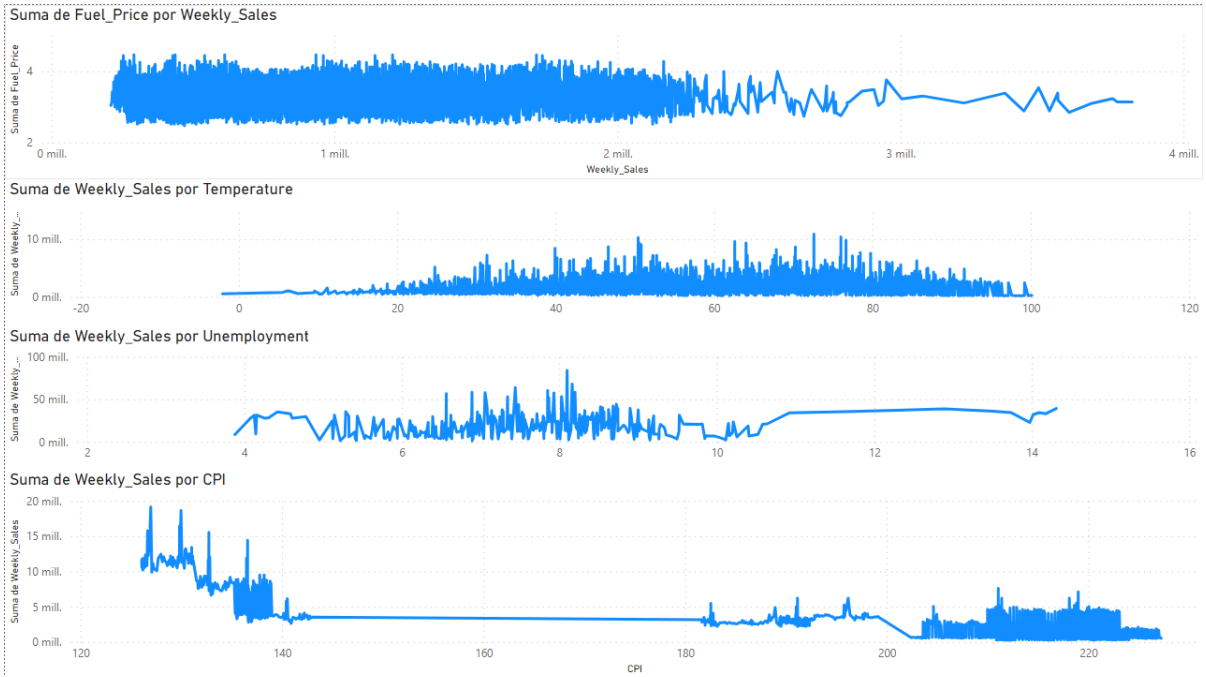


- **Super Bowl:** 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
- **Labour Day:** 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
- **Thanksgiving:** 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
- **Christmas:** 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

Thanksgiving	1471273.43
Super Bowl	1079127.99
Labour Day	1042427.29
Non Holiday Sales	1041256.38
Christmas	960833.11

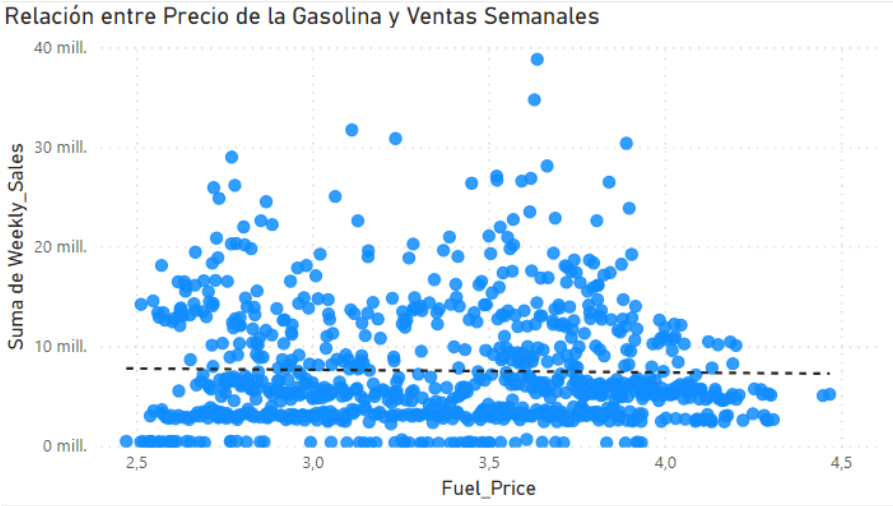
EDA

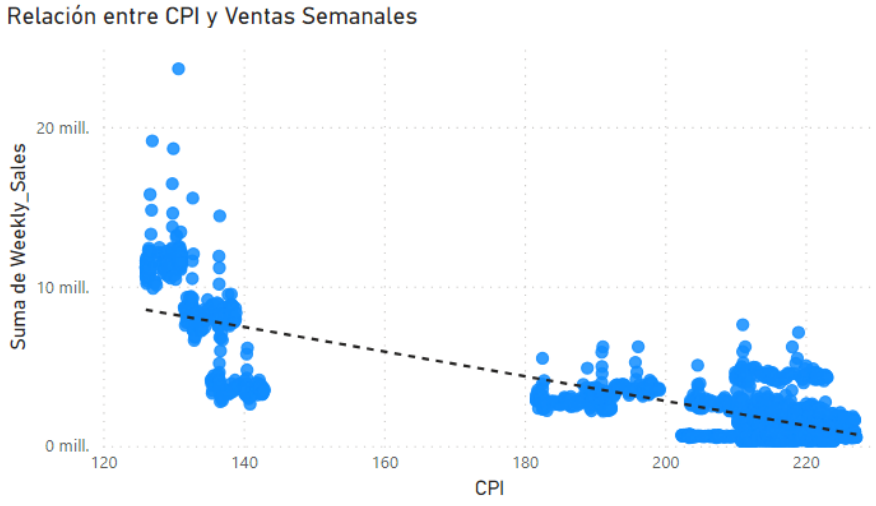
Evaluación de correlaciones entre variables.



EDA

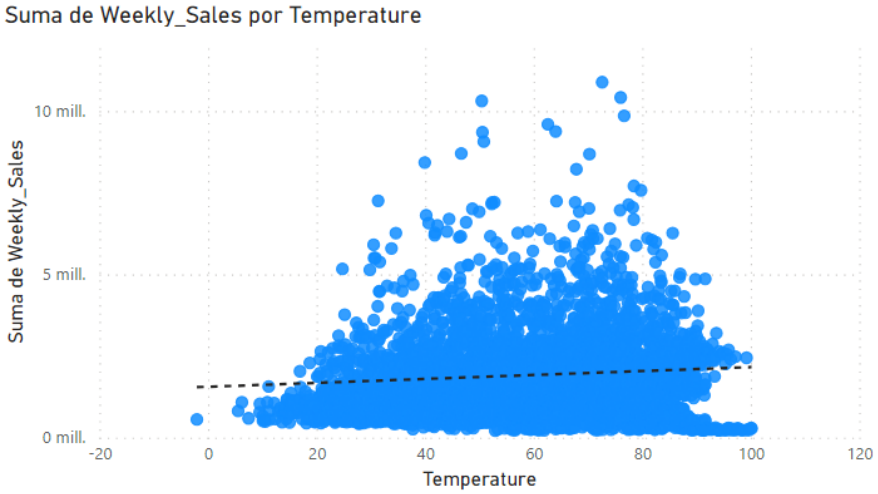
Evaluación de correlaciones entre variables.

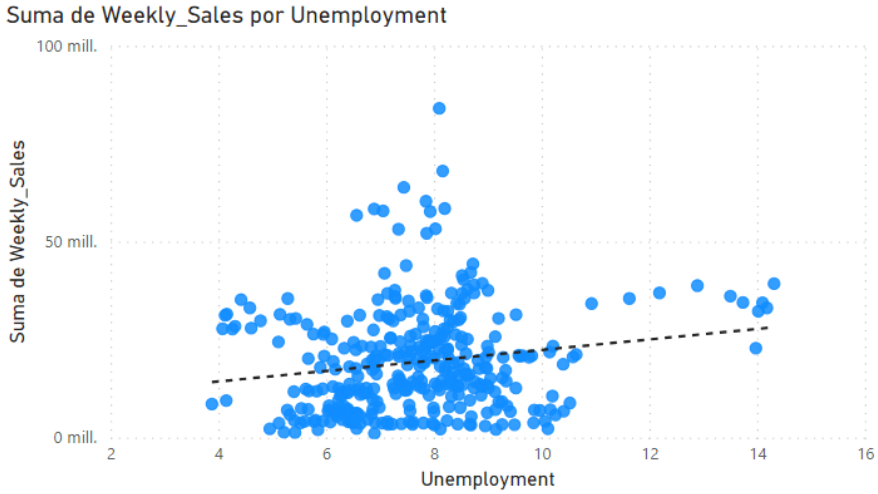




EDA

Evaluación de correlaciones entre variables.

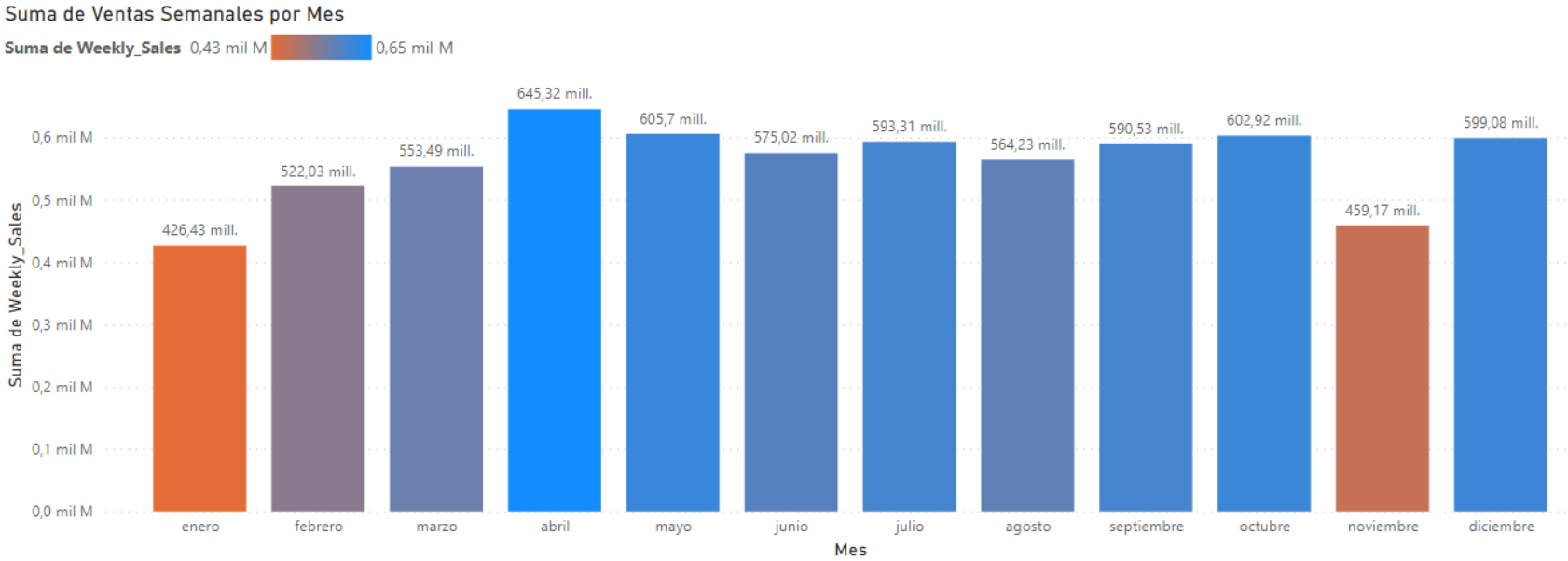




EDA

Ventas semanales mensuales.

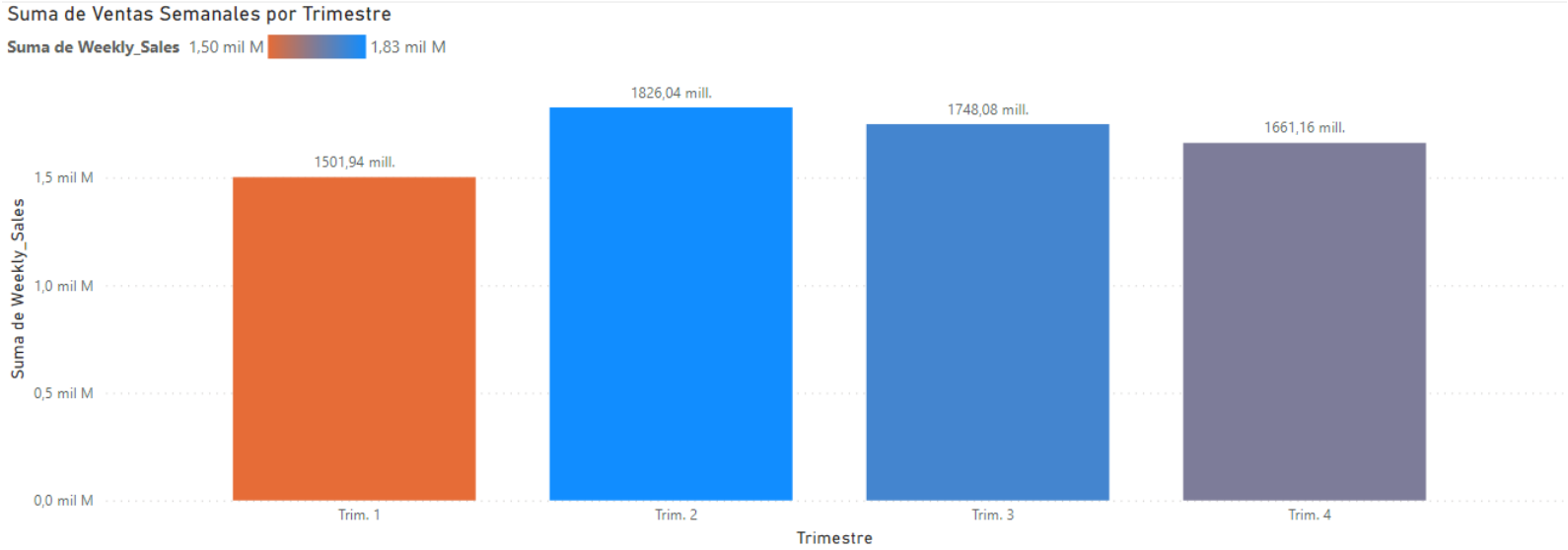
- ❑ Mayor cantidad de transacciones percibidas en abril.



EDA

Ventas semanales trimestrales.

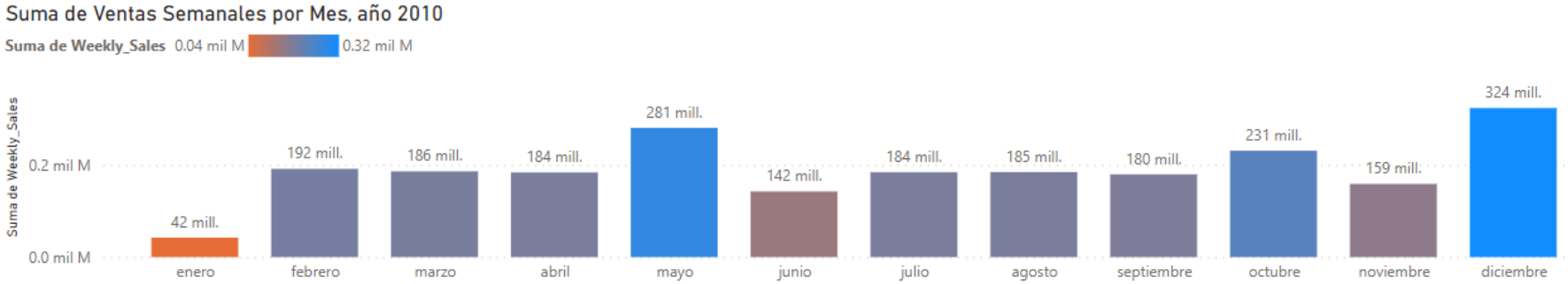
- ❑ Mayor cantidad de transacciones percibidas en el segundo trimestre.
- ❑ Calculo semestral (Trim. 1 + Trim. 2) = 3328 mill. (Trim. 3 + Trim. 4) = 3409 mill.
  Mayora cantidad de ventas percibidas en los últimos semestres del año.



EDA

Análisis de Ventas por año.

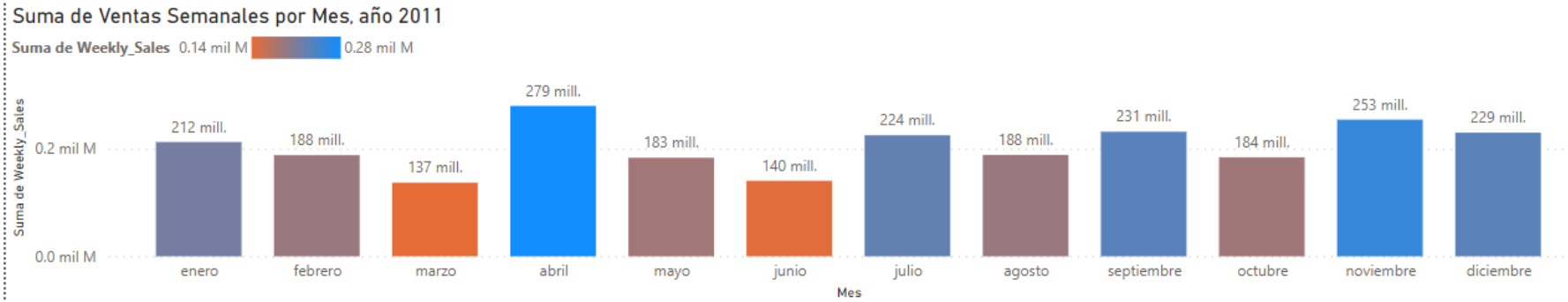
2010



EDA

Análisis de Ventas por año.

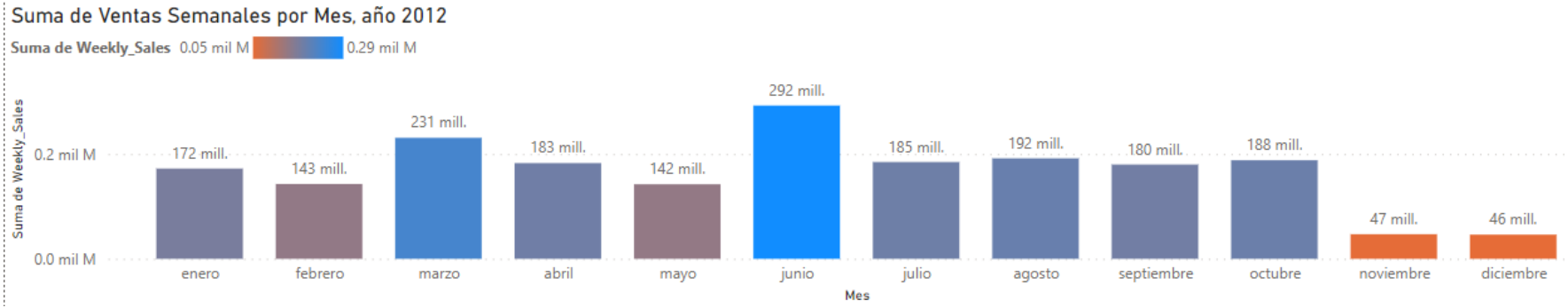
2011



EDA

Análisis de Ventas por año.

2012



5. Metodología

Metodología.

Herramientas aplicadas en el Desarrollo del proyecto.

Análisis de datos y EDA.

python

Power BI

jupyter

Pandas

NumPy

Metodología.

Desarrollo y evaluación de modelos de predicción.

6. Hallazgos realizados

Metodología.

Aplicación del modelo NEURALPROPHET.

¿Qué es?

Neural Prophet es una biblioteca de código abierto en Python que se basa en la popular biblioteca Prophet de Facebook para el análisis de series de tiempo y predicción.

A diferencia de Prophet, que utiliza un modelo aditivo no lineal para la predicción de series de tiempo, Neural Prophet utiliza redes neuronales para modelar relaciones no lineales y dependencias de datos.

Metodología.

Aplicación del modelo NEURALPROPHET.

Preparación de los datos para su aplicación al modelo.

```
dfall = df[['Date','Weekly_Sales']]
dfall = dfall.groupby('Date').sum()
dfall.reset_index(inplace = True)
```

✓ 0.0s

```
dfall.columns=['ds','y'] #Rename columns
dfall
```

✓ 0.0s

	ds	y
0	2010-01-10	42239875.87
1	2010-02-04	50423831.26
2	2010-02-07	48917484.50
3	2010-02-19	48276993.78
4	2010-02-26	43968571.13
...	...	...
138	2012-10-08	47403451.04
139	2012-10-19	45122410.57
140	2012-10-26	45544116.29
141	2012-11-05	46925878.99
142	2012-12-10	46128514.25

143 rows × 2 columns

Ajuste de los datos, separados en conjuntos de entrenamiento y prueba.

	MAE	RMSE	Loss	RegLoss	epoch
0	16788988.00	20670908.00	0.454866	0.0	0
1	17086124.00	21219810.00	0.468279	0.0	1
2	17018066.00	21225080.00	0.469398	0.0	2
3	17099928.00	21325842.00	0.465587	0.0	3
4	16978056.00	20794144.00	0.462879	0.0	4
...	...	...	...	...	...
477	2672168.75	4418870.50	0.029745	0.0	477
478	2649822.00	4192379.50	0.028390	0.0	478
479	2731744.00	4430820.00	0.029792	0.0	479
480	2617372.25	4074822.25	0.028080	0.0	480
481	2663931.75	4128645.50	0.028401	0.0	481

482 rows × 5 columns

Metodología.

Aplicación del modelo REGRESIÓN LINEAL

¿Qué es?

Un modelo de predicción con regresión lineal es un modelo estadístico que se utiliza para predecir el valor de una variable dependiente (y) a partir de una o más variables independientes (x) que tienen una relación lineal con la variable dependiente.

Metodología.

Aplicación del modelo REGRESIÓN LINEAL

Preparación y ajuste de los datos para su aplicación al modelo.

```
# Linear Regression :
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
X = df_clean[['Store','Fuel_Price','CPI','Unemployment','Day','Month','Year']]
Y = df_clean['Weekly_Sales']
X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size=0.2)
```

```
sc = StandardScaler()
x_train = sc.fit_transform(X_train)
x_test = sc.fit_transform(X_test)
```

Aplicación del modelo.

```
print('Linear Regression:')
print()
reg = LinearRegression()
reg.fit(X_train, Y_train)
Y_pred = reg.predict(X_test)
print('Train_Accuracy:',reg.score(X_train, Y_train)*100)
print('Test_Accuracy:',r2_score(Y_test, Y_pred)*100)
mape = metrics.mean_absolute_percentage_error(Y_test, Y_pred).round(4)
mae = metrics.mean_absolute_error(Y_test, Y_pred).round(4)
mse = metrics.mean_squared_error(Y_test, Y_pred).round(4)
r2 = r2_score(Y_test, Y_pred).round(4)
print(f"MAPE: {(mape)*(100)}%")
print(f"MAE: {(mae)}")
print(f"MSE: {(mse)}")
print(f"R2: {(r2)}")
#sns.scatterplot(x=Y_pred,y=Y_test)
#plt.figure(figsize=(7,5), dpi=75)
sns.regplot(x=Y_test,y=Y_pred)

import warnings
warnings.filterwarnings('ignore')
```

Metodología.

Aplicación del modelo NEURALPROPHET.

Métricas Evaluadas

MAPE (Mean Absolute Percentage Error)
Es una medida de error relativo que se utiliza comúnmente para evaluar la precisión de los modelos de pronóstico y series temporales.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

```
MAPE: 3.61%
```

Metodología.

Aplicación del modelo NEURALPROPHET.

Métricas Evaluadas

MAPE (Mean Absolute Percentage Error)
Es una medida de error relativo que se utiliza comúnmente para evaluar la precisión de los modelos de pronóstico y series temporales.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

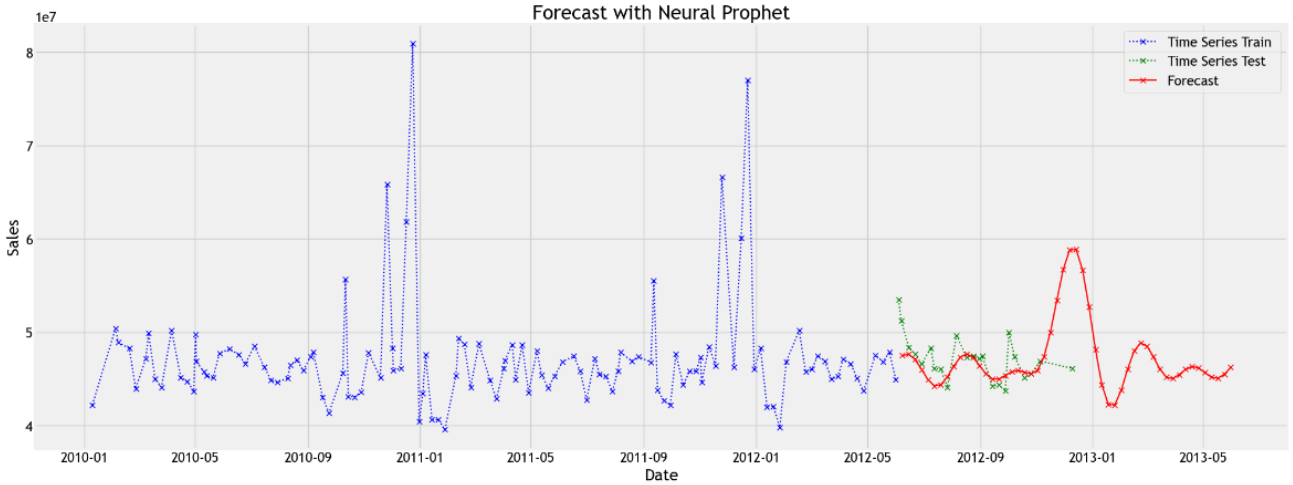
```
MAPE: 3.61%
```

7. Resultados Finales.

Metodología.

Aplicación del modelo NEURALPROPHET.

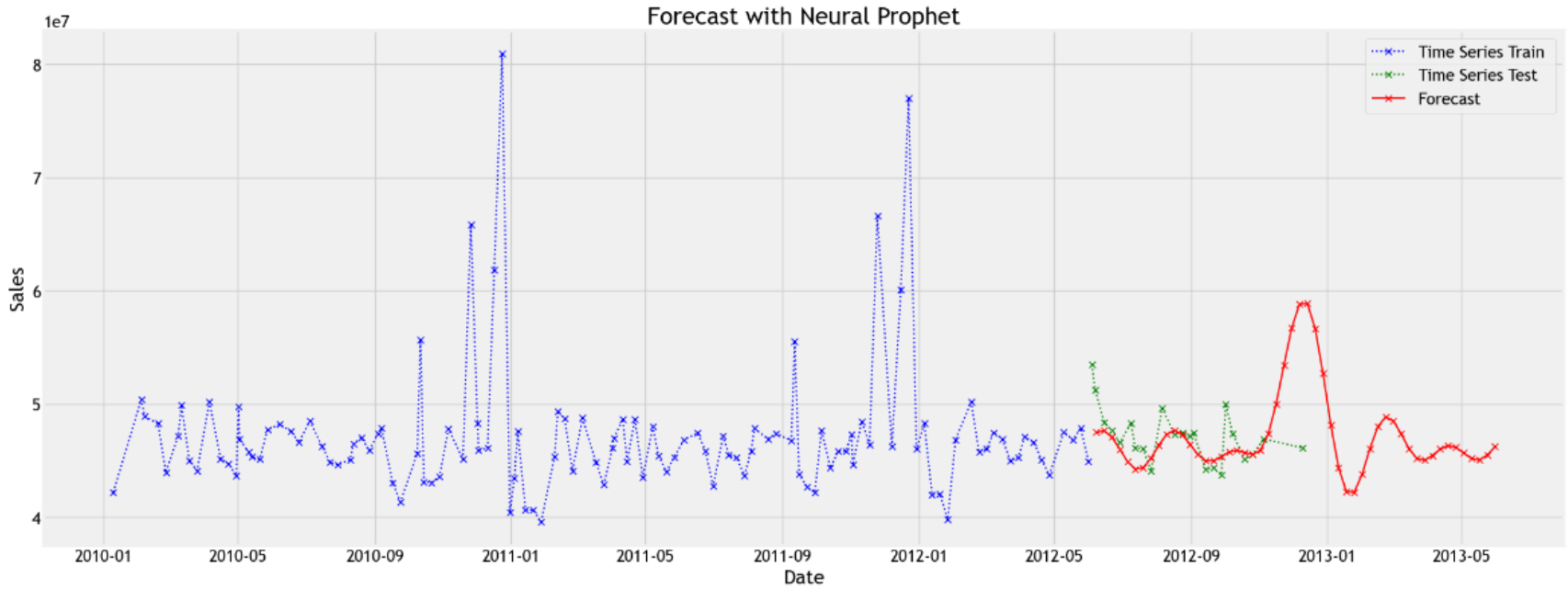
Predicción del Modelo – Pronóstico 1 año



MAPE: 3.61%

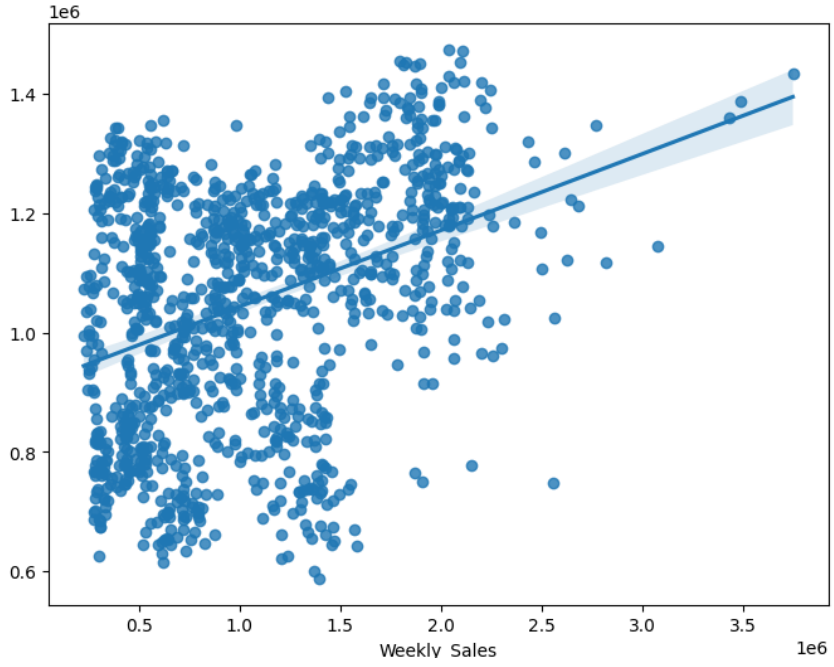
Metodología.

Predicción del Modelo – Pronóstico 1 año



Metodología.

Aplicación del modelo REGRESIÓN LINEAL




```
MAPE: 63.1299999999995%
```

Metodología.

Aplicación del modelo REGRESIÓN LINEAL

```
MAPE: 3.61%
```

Neural Prophet

scikit learn

```
MAPE: 63.129999999999995%
```

8. Conclusiones y Retos.

Conclusiones y Retos.

Conclusiones.

- El modelo que presenta un menor MAPE es NEURALPROPHET, por lo cual, dados los modelos aplicados en este Proyecto, resulta ser el mas adecuado para su consideración como herramienta para la predicción de la demanda.

- Es necesaria mayor capacitación para la búsqueda y experimentación de modelos aplicables a series de tiempo con el fin de encontrar un modelo en el cual se pueda realizar un buen ajuste de los datos y que su confiabilidad sea alta para ser comparado con NEURALPROPHET y definir cual funciona como herramienta para predecir la demanda de Walmart según la época del año.

- El desarrollo de este tipo de herramientas es vital para Walmart ya que le permitirá como compañía, mejorar los procesos de toma de decisiones, de planificación y gestión logística sin contratiempos, y sin generar costos extras.

Conclusiones y Retos.

Retos.

- La investigación, desarrollo y aplicación de modelos de Machine Learning para series, fue el principal reto afrontado durante la ejecución de este proyecto, ya que era requerido más tiempo del planificado para la investigación de tutoriales, entendimiento de conceptos y ejecución de pruebas.

- El uso de herramientas para el desarrollo de los notebooks fue el resto secundario, ya que al inicio se intento trabajar con Google Colab, pero hubo conflictos para desarrollar el modelo de NeuralProphet, por lo que se tuvo que desarrollar de manera local, iterar en diferentes versiones de Python y sus librerías, hasta obtener el entorno virtual adecuado para poder trabajarlo.

- No hubo posibilidad de aplicar el modelo de **Tensorflow** – **Regresión Lineal** ya que no se pudo comprender en su totalidad el tutorial para ser aplicado al escenario trabajado en el proyecto, en su lugar se aplico el modelo de Scikit Learn.

9. Referencias y Fuentes.

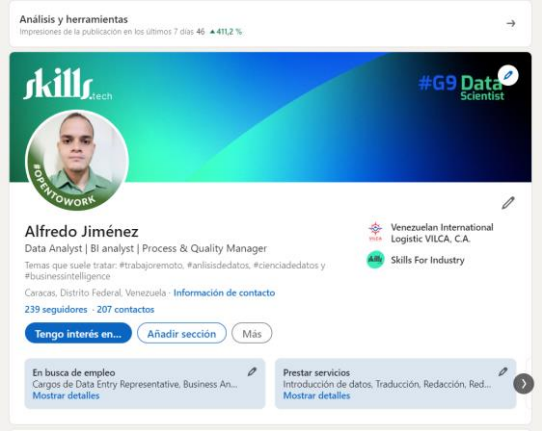
Referencias y Fuentes.

Retos.

- Aslan Ahmedov. (2022). **Walmart Sales Forecasting** https://www.kaggle.com/code/aslanahmedov/walmart-sales-forecasting#Random-Forest-Regressor .
- Oskar Triebe. (2021). **Neural Prophet.** https://neuralprophet.com/contents.html.
- Pierre-Louis Danieau. (2021). **Walmart Sales Forecasting** https://www.kaggle.com/code/pierrelouisdanieau/walmart-sales-forecasting/notebook.
- Divyajeet Thakur. (2021). **Walmart Sales Prediction.** https://www.kaggle.com/datasets/divyajeetthakur/walmart-sales-prediction **Dataset**
- Devashree Madhugiri. (2022). **5 Python Libraries for Time-Series Analysis.** https://www.analyticsvidhya.com/blog/2022/05/5-python-libraries-for-time-series-analysis/

Contacto

Contactos.



https://www.linkedin.com/in/alfredo-jim%C3%A9nez-985241145/

https://github.com/Faraves

Alfredo Jimenez

Caracas, Venezuela, ZIP code: 1010 | +58 424 143 5070 | alfredo.jimenezu95@gmail.com | LinkedIn: http://bit.ly/3ILUZYR

Sumario

Experiencia en gestión y documentación de procesos, implementación, control y desarrollo de Sistema de Gestión de Calidad basado en la normas ISO 9001:2015, gestión y análisis de conjuntos de datos, KPI's, y métricas empresariales.
Creo en el trabajo en equipo, colaborar con áreas multidisciplinarias en el desarrollo de proyectos, el aprendizaje continuo de habilidades técnicas y blandas que contribuyan a mejorar mi desempeño laboral, lograr objetivos empresariales y poder colaborar con mi equipo de trabajo.

Experiencia

VILCA C.A Distrito Capital, Caracas / Venezuela **Freight Forwarder.**
Coordinador de Calidad Sep, 2021 – Apr, 2023 Responsable de:
- Implementación de un Sistema de Gestión de Calidad bajo la norma ISO 9001:2015.
- Desarrollo de KPI's y métricas de negocio.
- Análisis de datos extraídos de los KPI's y métricas de negocio para la creación de reportes.
- Análisis y documentación de procesos.
- Auditoría Interna del Sistema de Gestión de Calidad.

AUTOMERCADOS PLAZAS Distrito Capital, Caracas/Venezuela Supermarket Chain.
Analista de Procesos y Calidad Ago, 2019 – Sept, 2021 Responsable de:
- Revisión de Procesos de Negocios.
- Documentación de Procesos.
- Desarrollo de Manuales de Normas y Procedimientos.
- Q/A Testing.
- Implementación de Procesos de Control de Calidad.

Education

Skills.Tech México Diplomado en Data Scientist. Sept, 2022 - Abr, 2023 Un programa diseñado para desarrollar las habilidades necesarias para impulsar una carrera en Ciencias de Datos. Orientado a un enfoque 100% para resolver casos de negocio en compañía de expertos activos en la industria.

- *Cursos Relevantes:* Introducción a la Ciencia de Datos, Python aplicado al análisis de Datos, SQL (Postgres), Power BI, AWS.

Universidad Nacional Experimental Simón Bolívar Venezuela TSU organización Empresarial. Abr, 2016 - Dic, 2019 Una carrera de 3 años orientada al desarrollo de un profesional que aplique sus conocimientos para optimizar la productividad y el mantenimiento de los bienes, equipos e infraestructura de un negocio,optimizando el uso y aprovechamiento de los recursos en los procesos productivos, la conservación y mejora de la infraestructura existente.
- *Cursos Relevantes:* Administración de Empresas, Logística, Administración de Sistemas de Calidad, Álgebra Lineal, Estadística, Economía, Contabilidad.

Adicional:

- Idiomas: Spanish (Nativo), Inglés (Intermedio - B1)
- Habilidades Técnicas: SQL(Postgres), Python (Pandas, Matplotlib), Power BI, Excel, Google Suite, Office, VS Code, Jupyter Notebooks, Google Colab.
- Habilidades Blandas: Orientación a la resolución de problemas, Pensamiento Crítico, Trabajo en equipo, Proactivo, Adaptabilidad, Storytelling, Ética de trabajo, Servicio al Cliente, Autodidacta.