

Descripción del Proyecto

Este proyecto analiza los factores que afectan las ventas semanales de tiendas minoristas utilizando tres conjuntos de datos: **Features_data**, **Sales_data** y **Stores_data**, que contienen información sobre eventos externos, ventas semanales y características de las tiendas, respectivamente. El objetivo es identificar tendencias, patrones estacionales y factores que influyen en las ventas.

Objetivos

Analizar cómo las variables como el tamaño de la tienda, días festivos, precios de combustible y tasas de desempleo afectan las ventas.

Explorar patrones y tendencias en los datos a través de visualizaciones gráficas.

Desarrollar un modelo predictivo basado en regresión lineal y regresión múltiple.

Descripción de archivos usados para el análisis

Stores_data_set.csv: Contiene información sobre el tamaño y la clasificación de cada tienda.

Sales_data_set.csv: Registra las ventas semanales de diferentes tiendas y departamentos

Sores_data_set.csv: Contiene información sobre el tamaño y tipo de tiendas.

Análisis Exploratorio

Analizamos la distribución de algunas de las variables clave, como la temperatura en diferentes fechas, el precio del combustible, el IPC (índice de precios al consumidor) y el desempleo para obtener una visión clara sobre la variabilidad de estas variables y su posible impacto en las ventas de las tiendas minoristas."

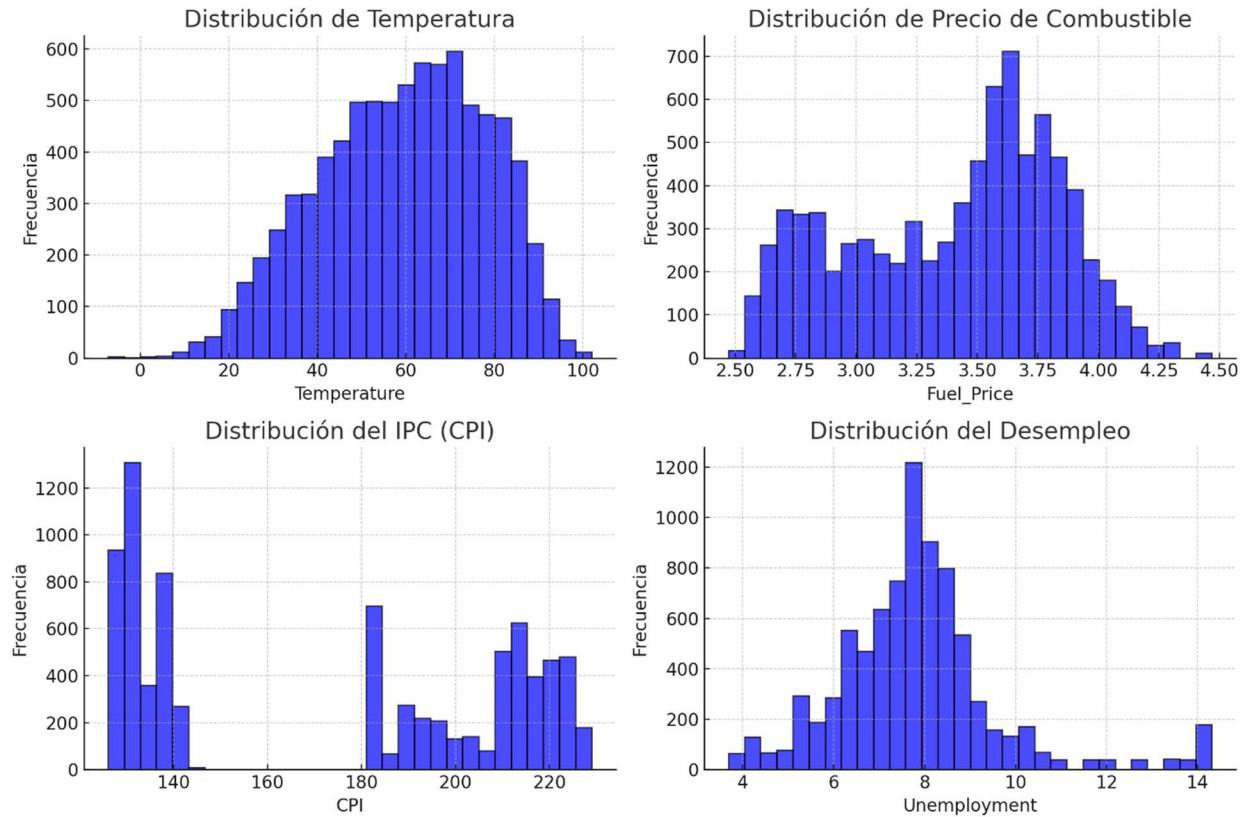


Figura 1. *Distribución de variables clave en las ventas minoristas: temperatura, IPC, precio del combustible y tasa de desempleo.*

En la Figura 1 se pueden observar las siguientes características:

- Temperatura: Distribución normal con un rango entre 0 y 100.
- Precio del Combustible: Tiene varias modas, lo que sugiere fluctuaciones en los precios a lo largo del tiempo.
- CPI (Índice de Precios al Consumidor): Presenta múltiples picos, indicando variaciones en la inflación durante el periodo analizado.
- Tasa de Desempleo: Los valores de desempleo varían entre el 4% y el 14%, con una mayor concentración alrededor del 8%.

Tendencia de Ventas por Semana

Para analizar el comportamiento de las ventas a lo largo del tiempo, se generó una gráfica que muestra su evolución semanal desde el 2010 hasta el 2012.

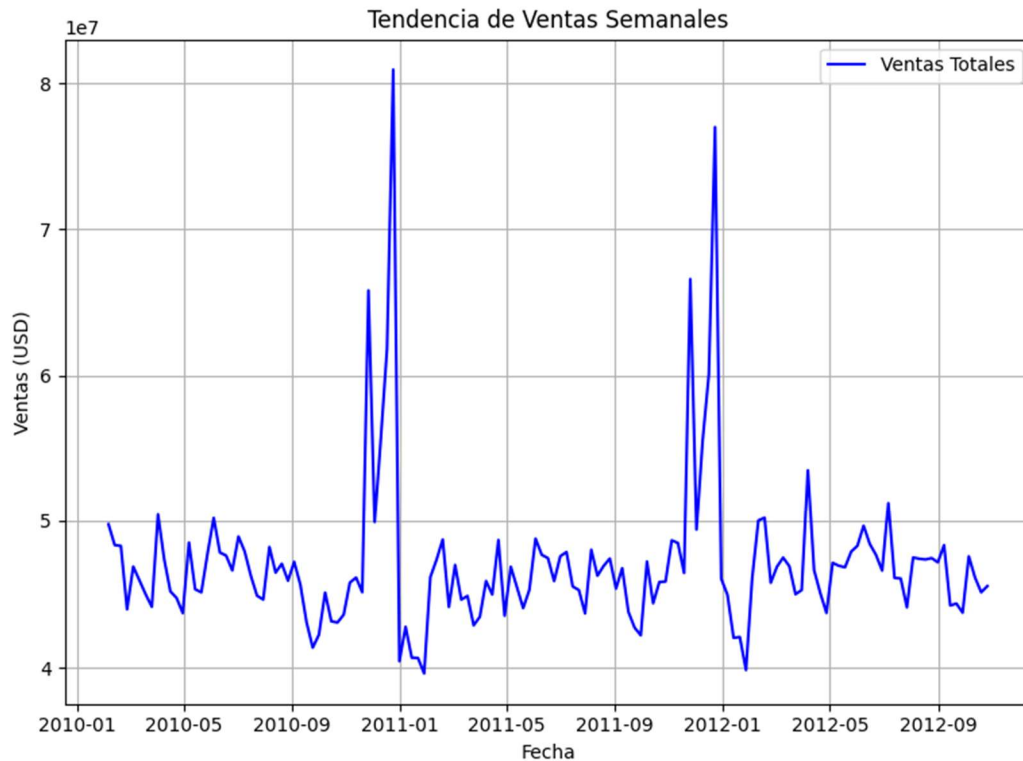


Figura 2. Numero de ventas semanales por fecha.

En la Figura 2 se observan cuatro picos principales en el período 2010-2012, correspondientes a los meses de noviembre y diciembre de los años 2011 y 2012. El pico más pronunciado ocurre en diciembre, reflejando un aumento significativo en las ventas.

Este comportamiento se debe a eventos estacionales como Black Friday en noviembre y las festividades de diciembre, incluyendo Navidad y Año Nuevo, que generan un aumento en las compras.

Tras visualizar la evolución de las ventas en el tiempo, realizaremos un análisis más detallado para evaluar el impacto de distintas variables en el número de ventas.

Análisis sobre la relación que existe entre los días festivos y el número de ventas

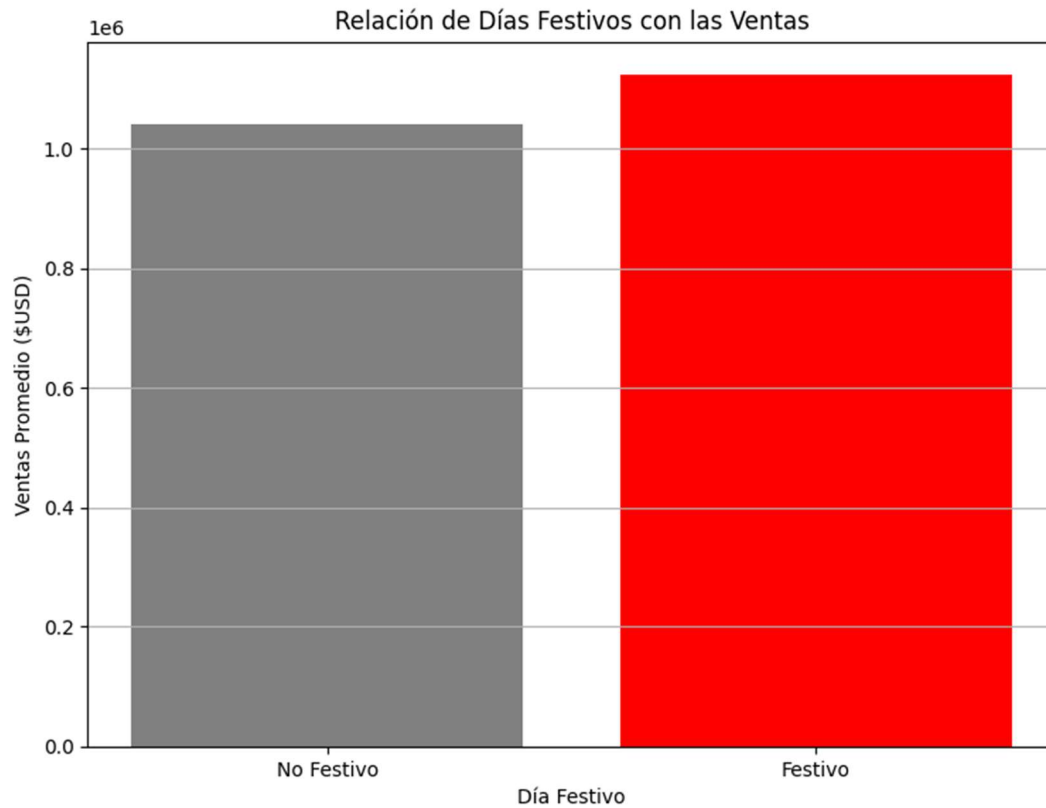


Figura 3. Relación entre los días festivos y el numero de ventas

En la Figura 3 se observa un ligero incremento en el promedio de ventas durante los días festivos, lo cual es consistente con el comportamiento estacional identificado en la tendencia de ventas semanales.

Aunque la diferencia no es drástica, sí es estadísticamente significativa y respalda la influencia de los días festivos en las ventas.

Relación entre el precio de combustible por día y la cantidad de ventas.

En esta sección analizamos la relación entre el precio del combustible y el número de ventas diarias para determinar si existe una correlación entre ambas variables.

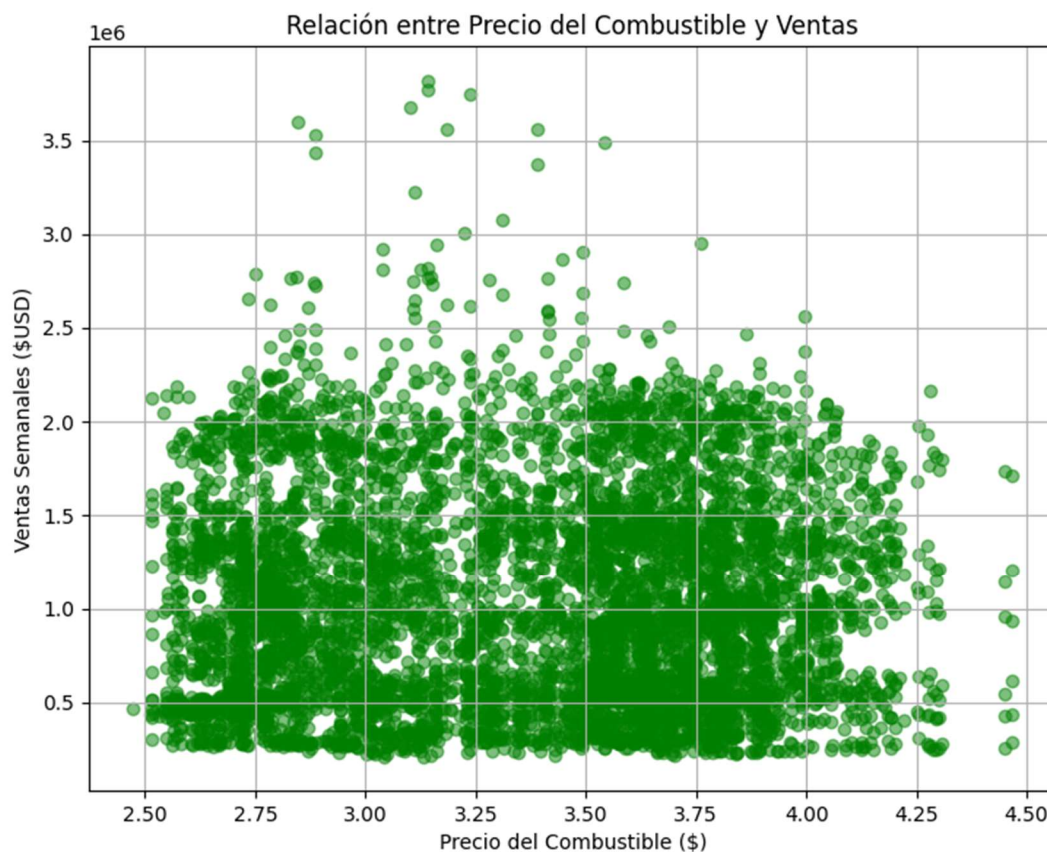


Figura 4. Precio de combustible contra ventas semanales.

En la Figura 4 no muestra una relación clara entre el precio del combustible y las ventas semanales. Aunque se observa cierta dispersión en los datos, los valores de ventas están distribuidos de manera relativamente uniforme en todo el rango de precios del combustible. Esto sugiere que el precio del combustible no tiene un impacto directo y fuerte en las ventas de las tiendas minoristas.

Interpretación

Los consumidores de estas tiendas parecen no ser altamente sensibles a los cambios en el precio del combustible al realizar sus compras.

Otros factores, como el tamaño de la tienda o su ubicación, podrían estar influyendo más significativamente en el volumen de ventas.

Ventas por mes

Para tener una visualización más clara sobre cómo se comportan el número de ventas, generamos un histograma con el promedio de ventas por mes.

En la Figura 2 se identifican picos de ventas en los meses de noviembre y diciembre. Para analizar este patrón con mayor detalle, generamos la siguiente gráfica, que muestra la variación de las ventas promedio por mes.

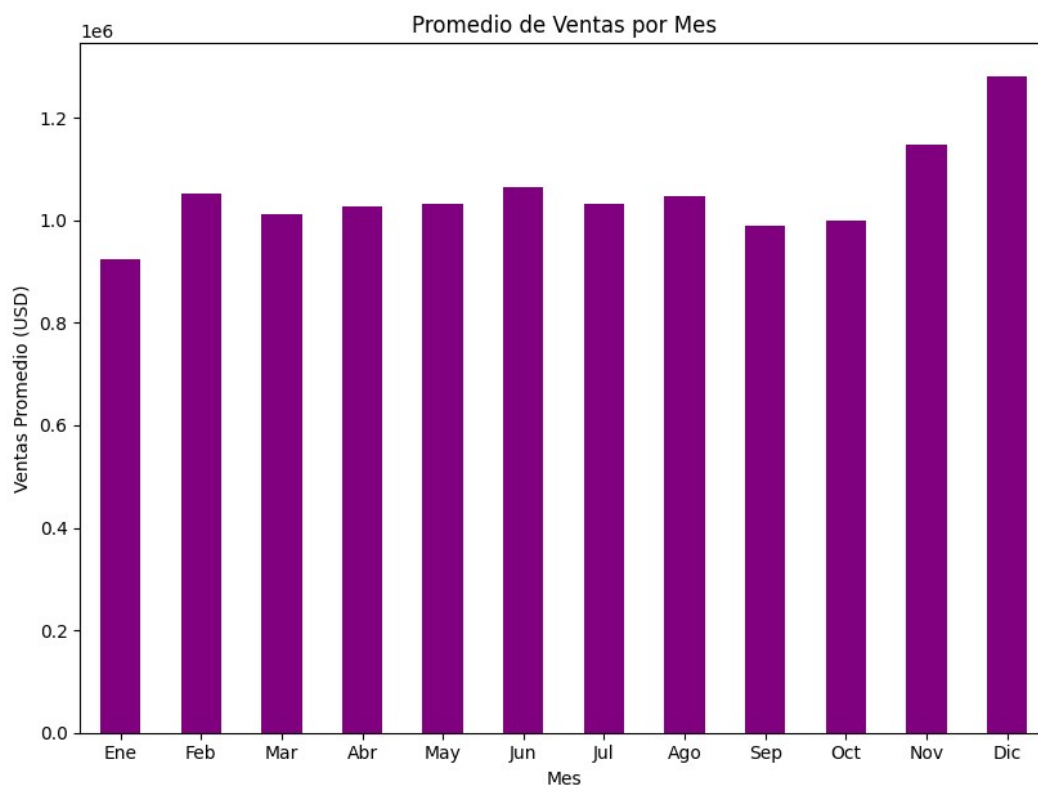


Figura 5. Histograma de la distribución de ventas por mes a lo largo del año

Como se observa en la Figura 5, las ventas aumentan significativamente en noviembre y alcanzan su punto máximo en diciembre, probablemente debido a eventos comerciales como Black Friday y las compras navideñas.

Por otro lado, enero registra el nivel más bajo de ventas, lo que podría deberse a la reducción del consumo después de la temporada navideña.

Regresión Lineal

Dado que variables como el precio del combustible o los días festivos no parecen tener un impacto significativo en el número de ventas, analizamos el tamaño de la tienda, ya que intuitivamente debería ser un factor más determinante en la cantidad de ventas.

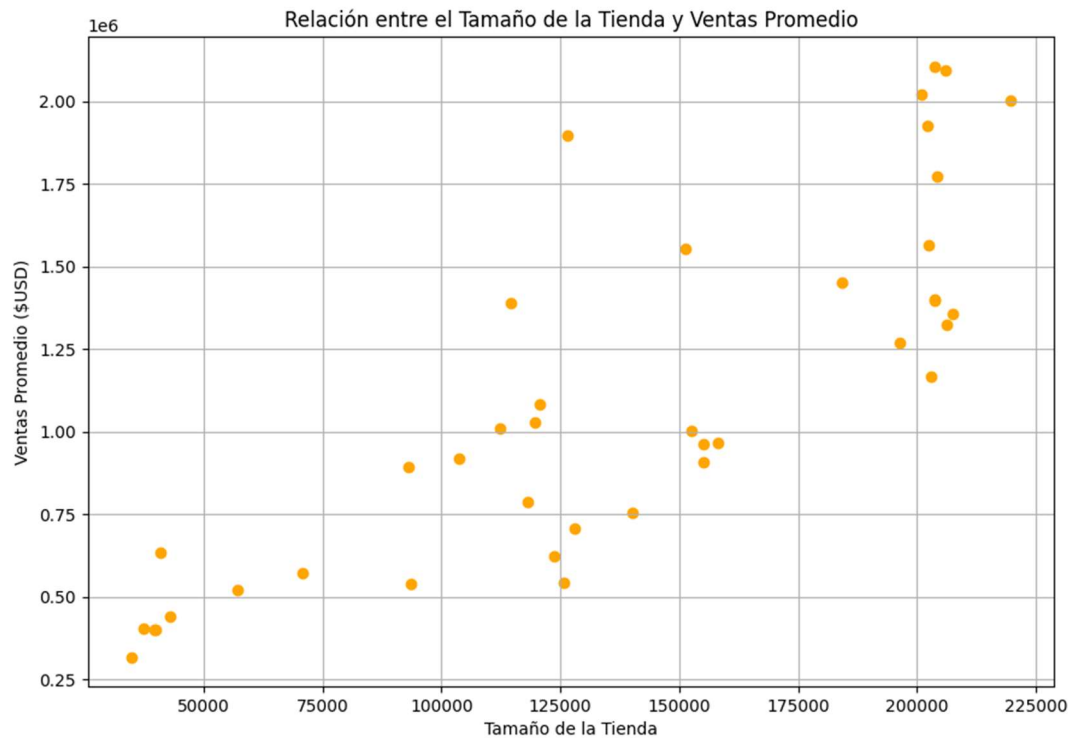
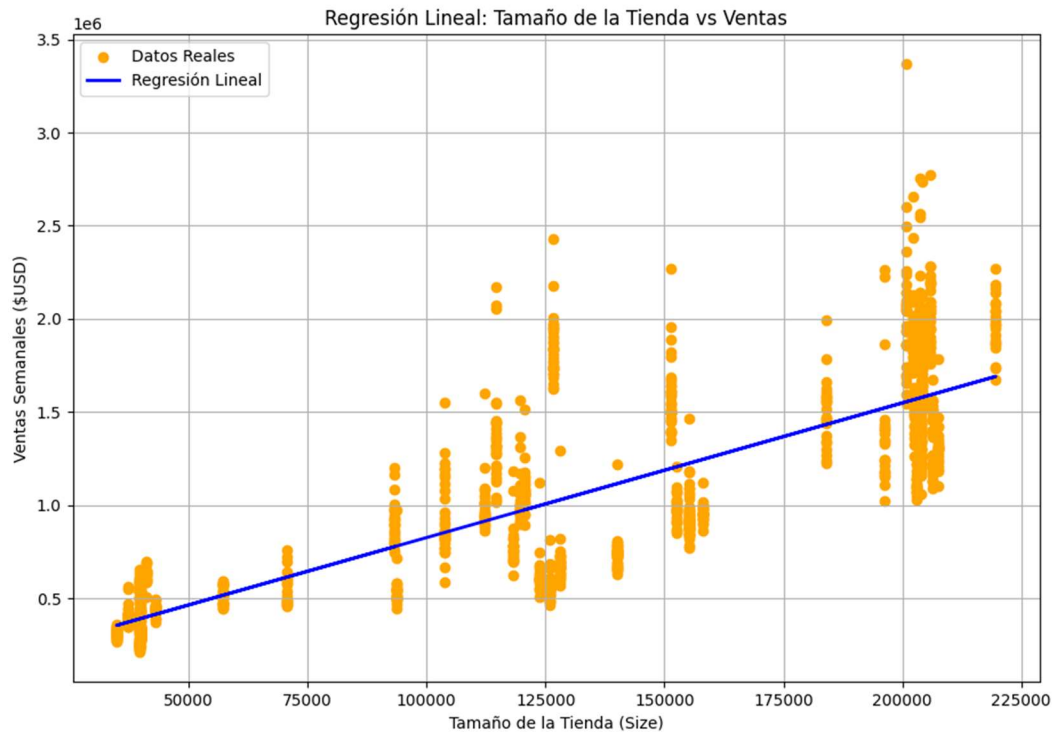


Figura 6. Datos que reflejan la relación entre el número de ventas y el tamaño de la tienda.

En la Figura 6 se observa una relación entre el tamaño de la tienda y el número de ventas, lo que indica que las tiendas más grandes tienden a registrar mayores ventas en promedio.

Dado que la relación es evidente, aplicamos una regresión lineal para cuantificar la tendencia y generar un modelo que nos permita predecir las ventas de tiendas cuyo dato

del tamaño no está presente en el dataset.



Se obtuvieron los siguientes parámetros para el modelo de regresión lineal:

MAE (Error Absoluto Medio): 247,939.37

El error absoluto medio (MAE) indica que, en promedio, las predicciones se desvían en aproximadamente \$248,000 de las ventas reales.

MSE (Error Cuadrático Medio): 1.96×10^{15}

Penaliza los errores grandes más que los pequeños, pero su magnitud es difícil de interpretar directamente.

R^2 (Coeficiente de Determinación): 0.6689 (~67%)

El tamaño de la tienda explica el 66.89% de la variabilidad de las ventas semanales.

Un 66.89% de explicación es bastante decente, y aunque el modelo basado solo en el tamaño de la tienda explica una proporción significativa de la variabilidad en las ventas, existen otros factores relevantes. Por ello, extendemos el análisis a una regresión múltiple incorporando más variables para mejorar la precisión del modelo.

Regresión múltiple

Para mejorar el modelo de predicción, incorporamos variables adicionales como el precio del combustible (Fuel Price), el índice de precios al consumidor (CPI), la tasa de desempleo (Unemployment) y la indicación de si la semana es festiva o no (IsHoliday). Esto nos permitirá evaluar si estas variables contribuyen a mejorar el ajuste del modelo.

Metrica	Regresión Simple	Regresión múltiple
MAE	247,939.37	246,368.28
MSE	1.96×10^{15}	1.02×10^{13}
R ²	0.6689	0.6780

Variable	Coeficiente
Size	7.71
Fuel_price	-12,634.00
CPI	2,116.89
Unemployment	-17,773.89 (
IsHoliday	86,617.26

Conclusiones:

El modelo múltiple mejora ligeramente la predicción, reduciendo el error absoluto medio y el error cuadrático medio. El R² pasó de 0.6689 a 0.6780, lo que indica que las nuevas variables explican un poco mejor las ventas. Sin embargo, la mejora no es muy grande, lo que sugiere que Size sigue siendo el principal factor determinante.

El tamaño de la tienda sigue siendo la variable más relevante, con un coeficiente de 7.17, lo que confirma su fuerte influencia en el número de ventas.

El coeficiente del precio del combustible es -12,634, lo que sugiere una ligera relación negativa: a medida que el precio del combustible aumenta, las ventas tienden a disminuir. Sin embargo, el impacto de esta variable es relativamente pequeño.

El coeficiente del CPI (2,176.89) indica una relación positiva con las ventas, lo que sugiere que un aumento en el índice de precios al consumidor está asociado con un incremento en las ventas. Esto podría estar relacionado con ajustes en los precios o en el comportamiento del consumidor.

El coeficiente del desempleo (-17,773.89) indica que un aumento en la tasa de desempleo está correlacionado con una reducción en las ventas, lo cual es consistente con la lógica económica: cuando hay más desempleo, los consumidores tienen menor poder adquisitivo

El coeficiente de IsHoliday (86,617.28) sugiere que las semanas festivas tienen un impacto positivo y significativo en las ventas, lo que es coherente con el comportamiento del comercio minorista