

Presentación proyecto final

Di Marco



Contenido



Abstract



Dataset



Variables



Preguntas de interés




Análisis Exploratorio de Datos e Insights




Modelado

Abstract

Este estudio se enfoca en el análisis de un conjunto de datos de un Superstore para comprender los factores que influyen en las ventas y rentabilidad del negocio.



El dataset incluye una amplia gama de variables relacionadas con transacciones, como las ventas, cantidad de productos vendidos, descuento aplicado, regiones de ventas, entre otros.



El análisis de este dataset proporcionará información valiosa para la toma de decisiones estratégicas, permitiendo al Superstore optimizar sus operaciones y maximizar sus ganancias.

Dataset



El set de datos contiene información detallada sobre 9994 transacciones realizadas dentro del Superstore.



Cada fila del dataset corresponde a una transacción única, mientras que las columnas contienen información detallada sobre cada transacción y los clientes involucrados.



Entre las variables se incluyen detalles como las ventas, la cantidad de productos vendidos, los descuentos aplicados, así como información sobre los clientes, como su segmento, estado y región.

Variables

Row ID: Identificador único para cada fila en el dataset.	Order ID: Identificador único para cada orden de compra.	Ship Mode: Modo de envío seleccionado para la orden.	Segment: Segmento al que pertenece el cliente (por ejemplo, consumidor, corporativo, oficina en casa).	Country: País donde se realizó la orden.	City: Ciudad de destino de la orden.
State: Estado de destino de la orden.	Postal Code: Código postal de la ubicación de envío.	Region: Región geográfica de destino de la orden.	Product ID: Identificador único para cada producto.	Category: Categoría a la que pertenece el producto (por ejemplo, muebles, suministros de oficina, tecnología).	Sub-Category: Subcategoría del producto.
Product Name: Nombre del producto.	Sales: Monto total de ventas de la orden.	Quantity: Cantidad de productos vendidos en la orden.	Discount: Descuento aplicado a la orden.	Profit: Ganancia obtenida de la orden.	Customer ID: Identificador único para cada cliente.
Customer Name: Nombre del cliente.		Order Date: Fecha en que se realizó la orden.	Ship Date: Fecha de envío de la orden.		

Preguntas de interés



¿CUÁLES SON LOS PRODUCTOS MÁS VENDIDOS?



¿QUÉ CATEGORÍA DE PRODUCTOS GENERA LOS MAYORES INGRESOS Y GANANCIAS?



¿CUÁL ES LA TENDENCIA DE LAS GANANCIAS A LO LARGO DEL TIEMPO?



¿QUÉ SEGMENTO DE CLIENTES TIENE EL MAYOR IMPACTO EN LAS VENTAS?

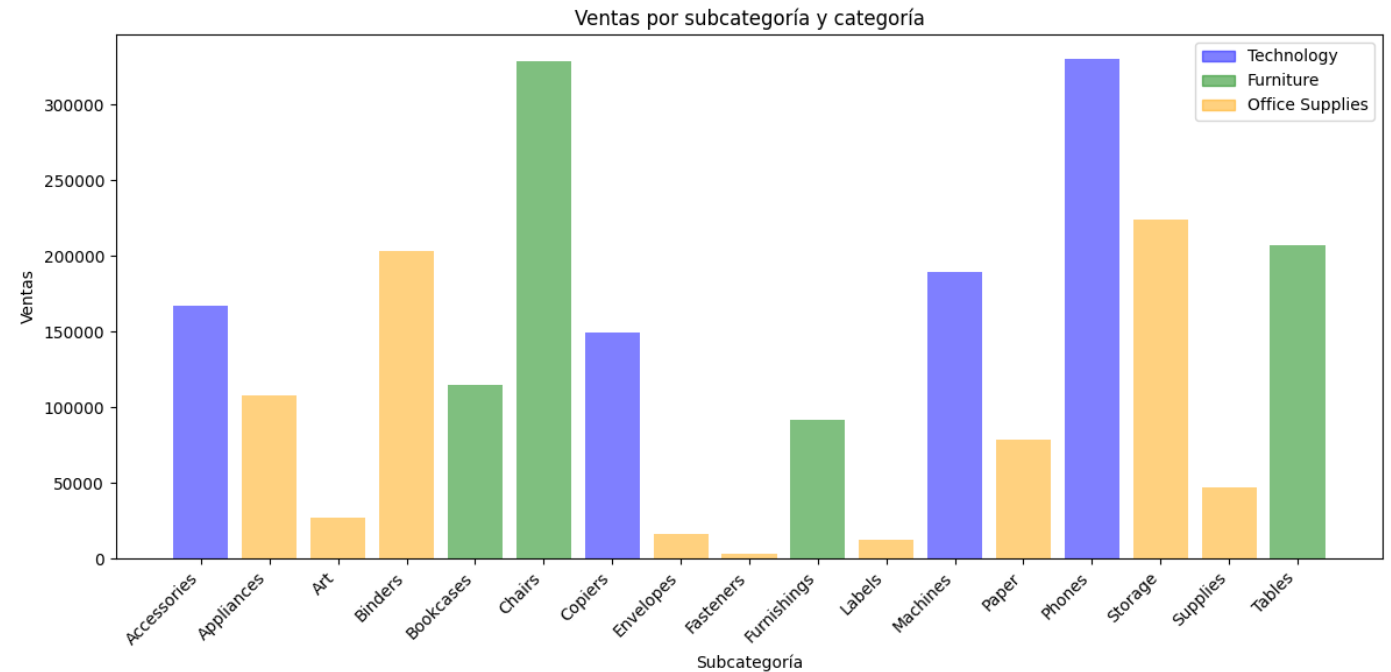
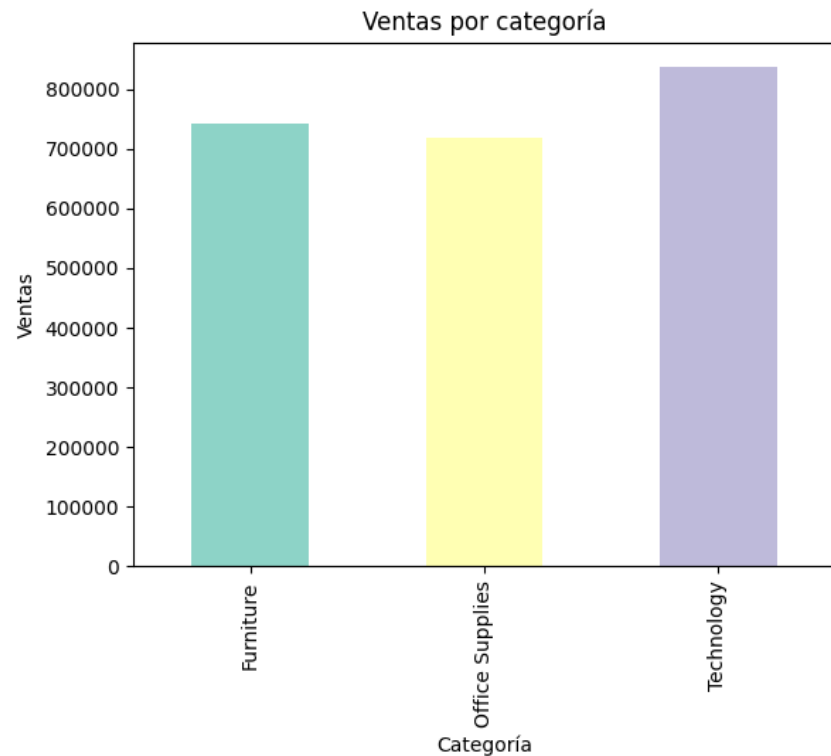


¿EXISTE ALGUNA RELACIÓN ENTRE LAS VENTAS Y LOS DESCUENTOS?

Análisis exploratorio de datos

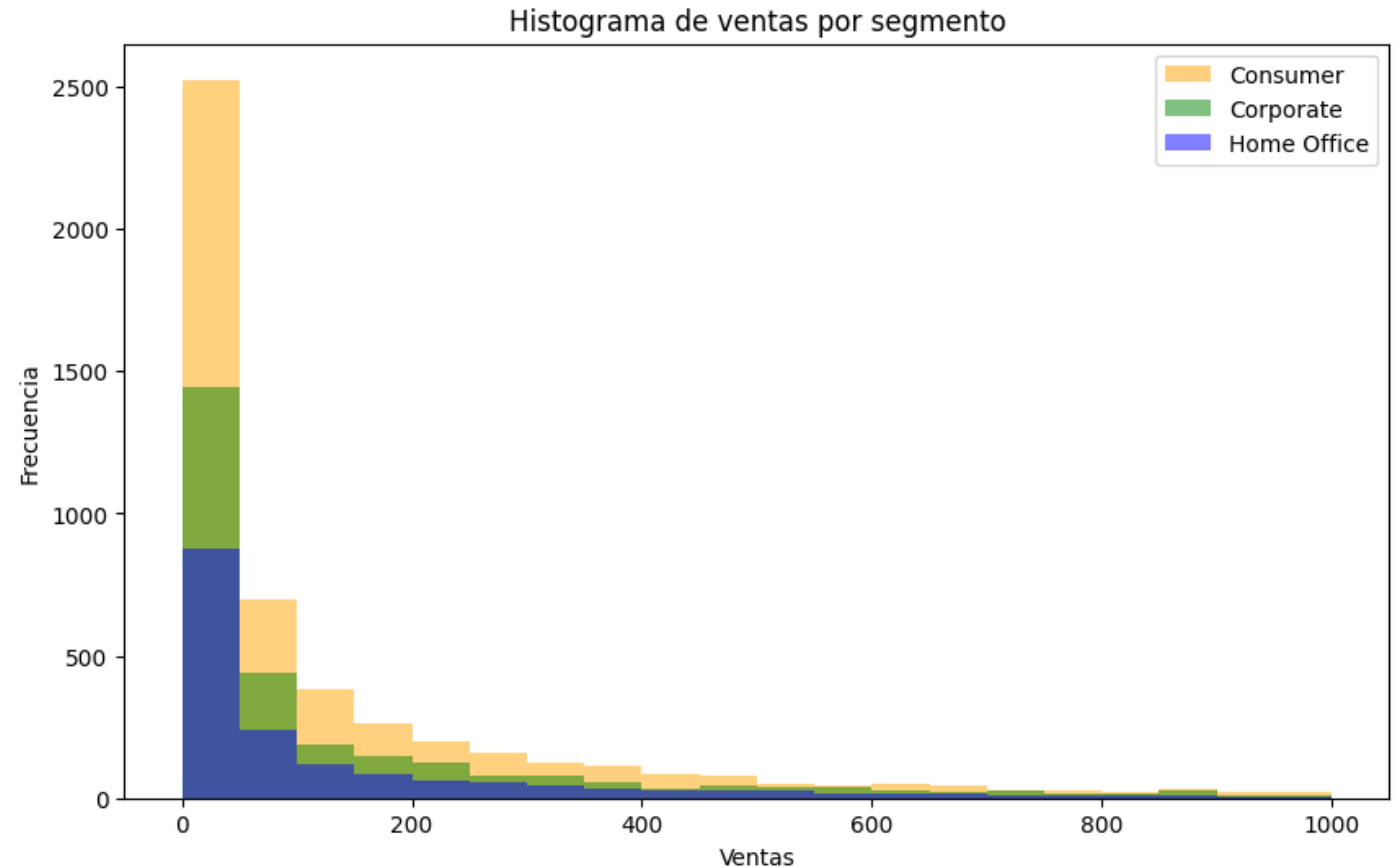
Ventas por categoría y subcategoría

Se observa que las mayores ventas se dan en la categoría "Tecnología", mientras que sillas y celulares se destacan como las subcategorías de productos mas vendidos.



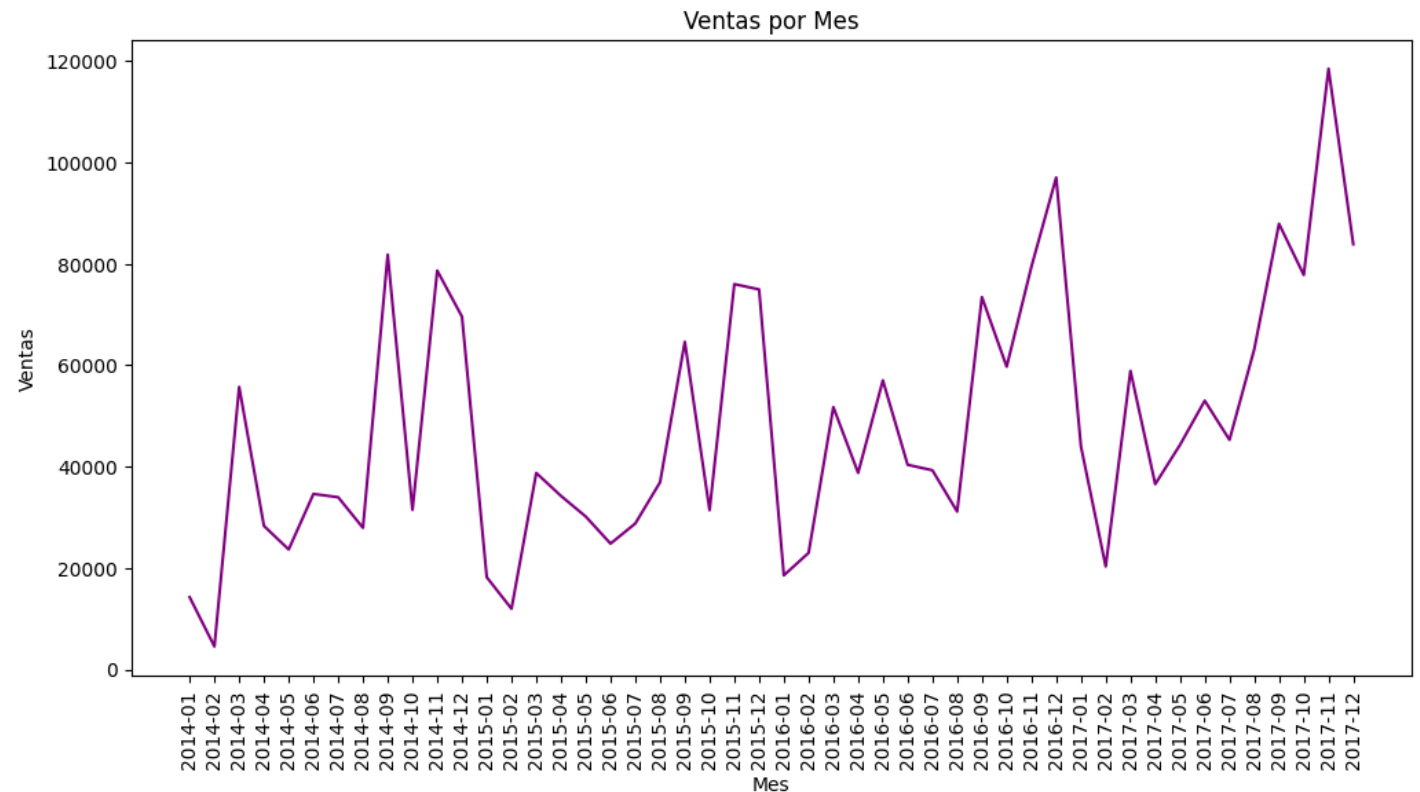
Ventas por segmento

Se puede apreciar que el segmento "Consumer" es el que tiene el mayor número de ventas en comparación con los segmentos "Corporate" y "Home Office". Esto sugiere que el segmento de consumidores individuales es el más influyente en términos de generación de ingresos para el negocio.



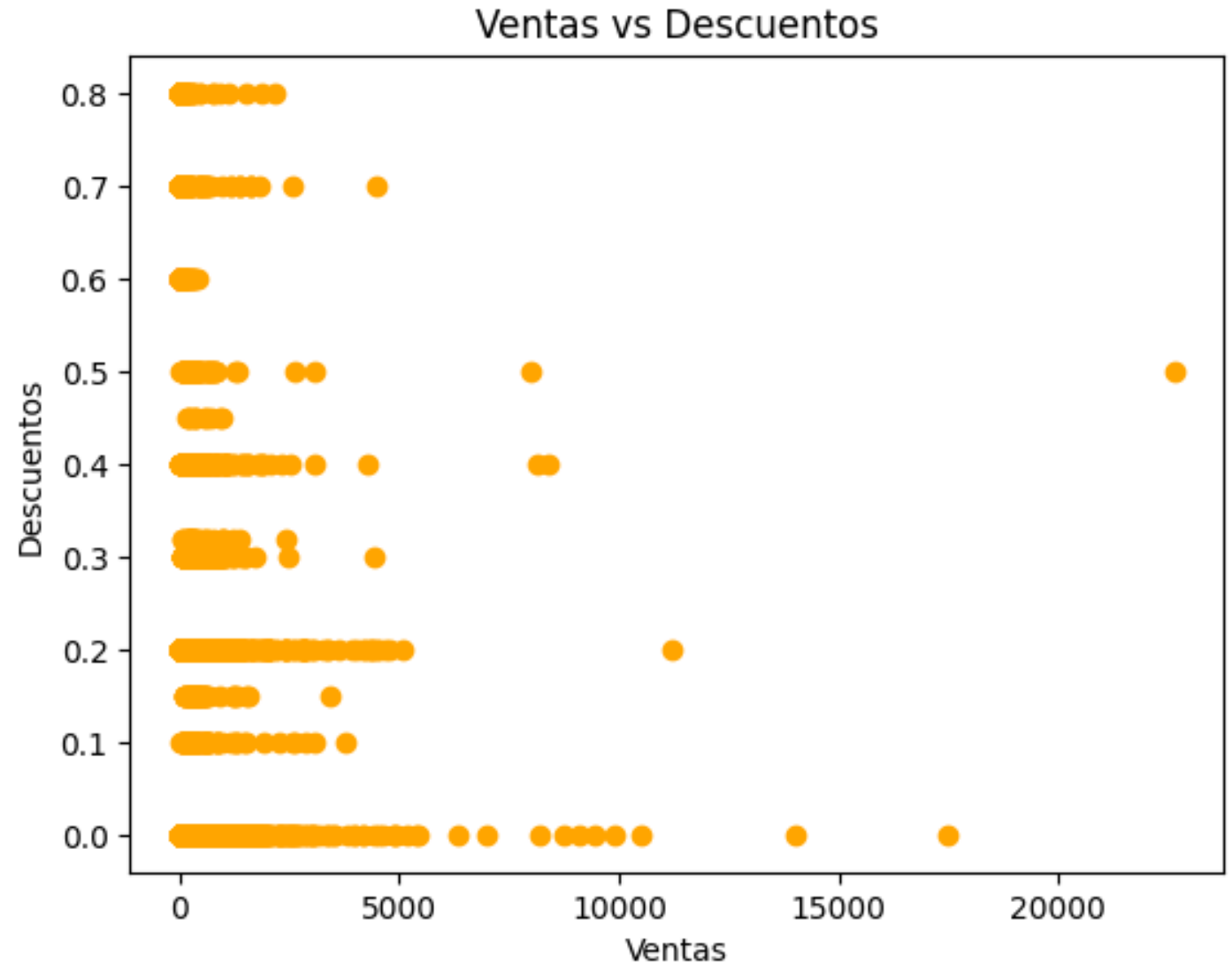
Ventas por mes

Se puede notar una tendencia general de aumento en las ganancias durante ciertos períodos del año. Esta tendencia sugiere que las ganancias pueden estar influenciadas por factores estacionales o eventos específicos que afectan las ventas en el supermercado. Por ejemplo, se observan picos de ganancias hacia fin de año, probablemente debido a la temporada de festividades y compras navideñas.



Ventas vs Descuentos

No se observa una relación lineal clara entre los descuentos ofrecidos y las ventas generadas en el supermercado.



Insight

1

El segmento de clientes "Consumer" es el más influyente en las ventas totales del supermercado: Este insight sugiere que dirigir estrategias de marketing y atención al cliente hacia este segmento puede ser crucial para maximizar las ventas y la rentabilidad del negocio.

2

Las ganancias muestran una tendencia estacional: Se observa una tendencia de aumento en las ganancias durante ciertos períodos del año, seguida de períodos de disminución. Esto indica la importancia de ajustar las estrategias comerciales según las fluctuaciones estacionales y eventos específicos que afectan las ventas.

3

Las subcategorías de productos "Sillas" y "Teléfonos" son las más vendidas: Este insight destaca la importancia de estos productos en las ventas totales del supermercado y sugiere que podría ser beneficioso enfocar recursos adicionales en estas subcategorías para maximizar las ventas y satisfacer la demanda de los clientes.

4

Relación entre descuentos y ventas: Aunque no se observa una relación lineal entre descuentos y ventas en el análisis, sería interesante investigar más a fondo cómo los descuentos afectan el comportamiento de compra de los clientes y cómo pueden utilizarse estratégicamente para impulsar las ventas.

Modelado

PROPÓSITO

El propósito del modelado en este proyecto es utilizar técnicas de aprendizaje automático para predecir las ganancias en función de diversas variables como las ventas, la cantidad de productos vendidos, el descuento aplicado, la región, la categoría del producto, el modo de envío, etc. El modelado nos permite comprender mejor los factores que influyen en las ganancias y proporcionar información valiosa para la toma de decisiones comerciales estratégicas.

Algoritmos utilizados



SVR



LinearRegression



RandomForestRegressor



DecisionTreeRegressor

Evaluación del modelo



La técnica utilizada para evaluar el modelo se basa en el coeficiente de determinación (R^2). Este coeficiente mide la calidad de las predicciones del modelo en relación con los valores reales de la variable que estamos tratando de predecir. Un R^2 cercano a 1 indica un buen ajuste del modelo a los datos, lo que significa que las predicciones son bastante precisas.



En el caso del proyecto, se calcula el R^2 para cada modelo ajustado (SVR, Regresión Lineal, Random Forest Regressor y Decision Tree Regressor) utilizando los datos de prueba

Resultados

-SVR (Support Vector Regressor): Tiene el valor R^2 más alto (0.855), lo que sugiere que este modelo explica aproximadamente el 85.5% de la variabilidad en la variable objetivo ("Profit"). Esto indica que el SVR podría ser el modelo más adecuado para hacer predicciones precisas sobre la ganancia.

- Linear Regression (Regresión Lineal): Tiene un valor R^2 de 0.812, lo que indica que explica alrededor del 81.2% de la variabilidad en la variable objetivo. Aunque es ligeramente menor que el SVR, sigue siendo un buen desempeño y puede ser considerado como una alternativa válida.

Conclusiones

El análisis detallado de los datos del Superstore ha proporcionado valiosos insights sobre las ventas, el comportamiento del cliente y los factores que influyen en la rentabilidad. Se identificaron las categorías y subcategorías de productos más vendidos, se segmentaron los clientes según su comportamiento de compra, y se evaluaron las tendencias de ganancias a lo largo del tiempo.

"Además, el proceso de modelado ha demostrado ser efectivo en la predicción de las ganancias basado en diversas variables relacionadas con las transacciones en el supermercado. Los modelos de regresión utilizados, especialmente el SVR (Support Vector Regressor) y el RandomForestRegressor, han mostrado un alto nivel de precisión en la predicción de las ganancias, lo que sugiere su utilidad para la planificación financiera y la toma de decisiones estratégicas.