

PRONÓSTICO DE VENTAS MENSUALES EN TIENDAS ROSSMANN

DESDE EXPLORACIÓN DE DATOS HASTA MODELOS DE MACHINE LEARNING.

UN ANÁLISIS PREDICTIVO BASADO EN DATOS REALES DE VENTAS

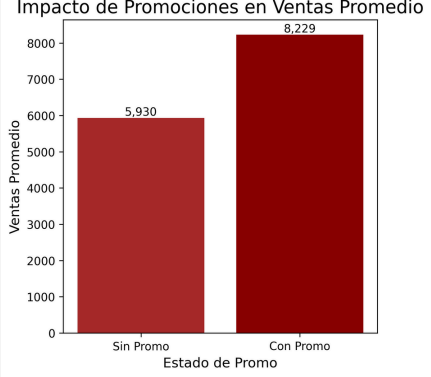
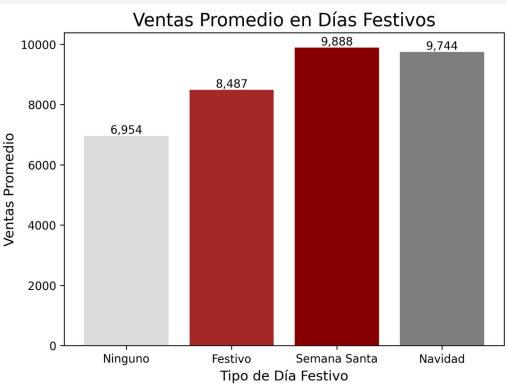
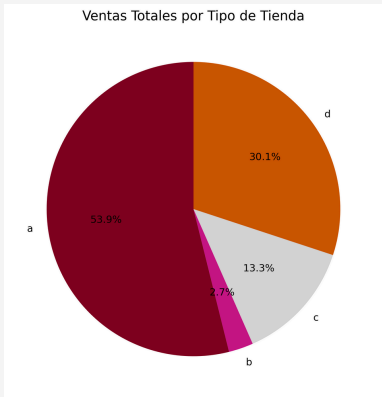
ARITZEL MURRAY

PROF: JUAN MARCOS CASTILLO, PHD

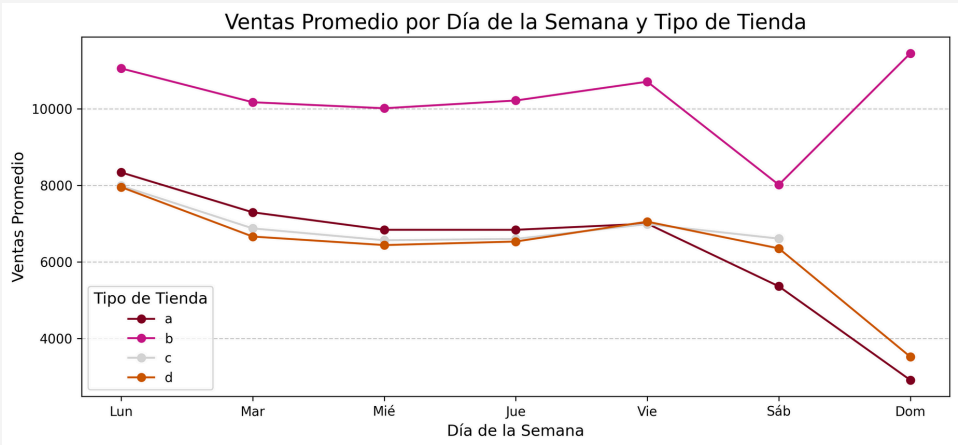
A nivel personal escogí este dataset Rossmann Stores, con más de 1,110 tiendas en Europa, porque tenía el interés en predicción de ventas.

El objetivo es desarrollar un modelo predictivo basado en datos reales para estimar las ventas mensuales de las tiendas Rossmann en los próximos meses. Esto ayudará a identificar patrones clave, evaluar el impacto de factores como promociones y estacionalidad, y proporcionar una herramienta práctica para planificar recursos y aumentar la competitividad.

Insights valiosos: Seleccionando data

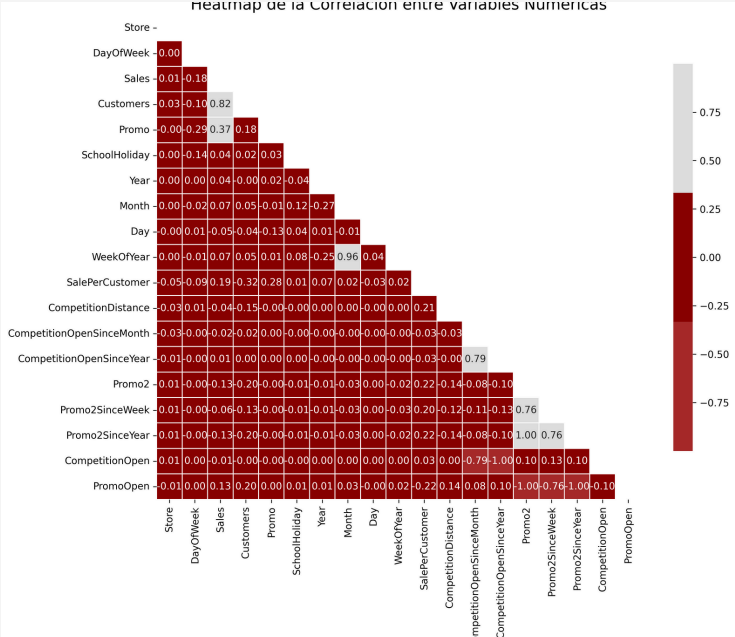


*Los días festivos y las promociones generan un impacto en las ventas.



*Las ventas aumentan los días laborables, especialmente los lunes, mientras que los domingos registran una caída notable debido al cierre de la mayoría de las tiendas. Las tiendas **tipo b**, al abrir todos los días, logran incrementar sus ventas promedio los domingos frente a otros tipos de tienda.

Identificación de Variables Relevantes para el Análisis Predictivo"



Con este heatmap de correlación se identifiquo que **Customers** y seguido de **Promo** tienen una relación más fuertes con **Sales**.

Con estas variables identificadas como las más relevantes, se procedió a construir modelos predictivos, empezando por regresiones lineales y múltiples para explorar su capacidad de estimar **Sales** a partir de los patrones encontrados.

PRONÓSTICO DE VENTAS MENSUALES EN TIENDAS ROSSMANN

USOS DE DIFERENTES MODELOS PREDICTIVOS

REGRESIÓN LINEAL

R2	RMSE	MAE
0.680	1,758.59	1,271.24

Utilizando únicamente **Customers** como variable independiente.

REGRESIÓN MÚLTIPLE

R2	RMSE	MAE
0.727	1,620.14	1,155.38

Utilizando **Customers y Promo** como variables independientes.

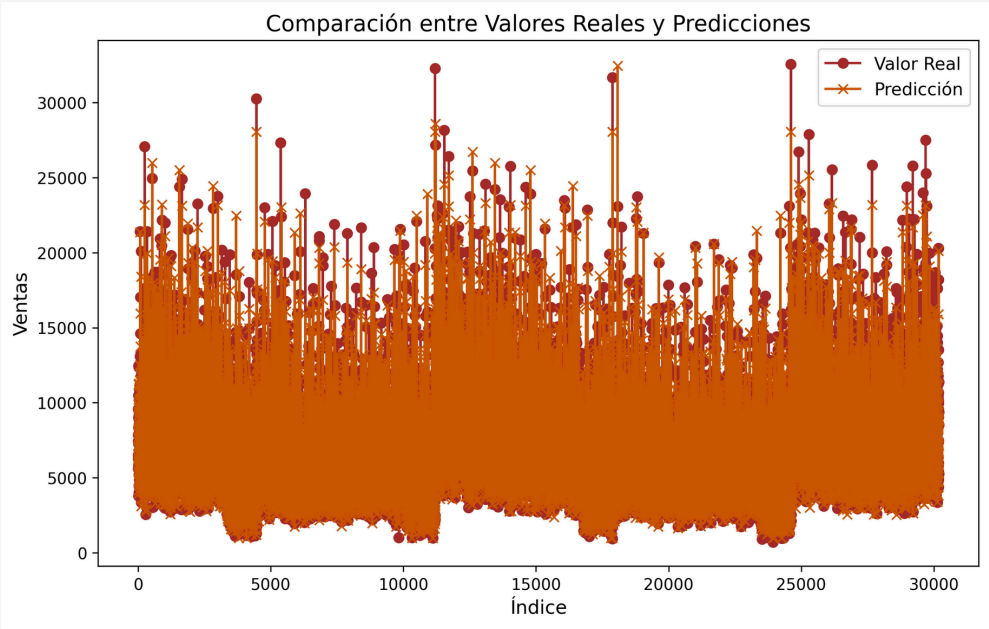
Dada la precisión relativamente baja obtenida con los modelos de regresión lineal y múltiple, se decidió utilizar varios algoritmo, de cual se escogió **XGBRegressor**

XGB REGRESSOR

Conjunto	R2	RMSE	MAE	MAPE
Prueba	0.9546	661.92	455.33	6.77%
Entrenamiento	0.9580	637.37	441.77	6.60%
Julio 2015 (No Entrenado)	0.9374	734.70	527.63	7.40%

se agregaron algunas variables adicionales como: **DayOfWeek,, StoreType, Assortment, StateHoliday, CompetitionDistance** como variablese independientes.

Valores reales vs Predicción



Estos resultados reflejan la solidez del modelo XGBRegressor para predecir ventas en tiendas Rossmann. La combinación de una precisión alta y errores bajos, incluso en datos no entrenados, valida su utilidad como herramienta para decisiones estratégicas basadas en datos.

Aprendizajes: identificar patrones clave y a seleccionar las variables más relevantes como Customers y Promo, utilizando herramientas como heatmaps de correlación. Esto me permitió construir modelos predictivos progresivamente, desde regresiones lineales y múltiples hasta XGBRegressor, optimizado con GridSearchCV.

Validé el modelo en datos no entrenados, como julio de 2015, asegurando su capacidad de generalización.

Este proceso me enseñó a conectar análisis descriptivo y predictivo para resolver problemas reales de forma efectiva.