



PROYECTO MACHINE LEARNING:

Predicción de ventas

Guillem Pañella Valiente

Tutores: Alberto Romero, Ivan Codero, Lucas Zamora

Escuela The Bridge

Data Science Online Madrid Mayo 2024 - Diciembre 2024

ÍNDICE DE CONTENIDO

1. Introducción, problema de negocio y elección de bases de datos
2. Exploración de datos y entendimiento de la base de datos
3. Machine Learning
 - 3.1. Modelos Machine Learning diarios
 - 3.2. Modelos Machine Learning semanales
4. Conclusiones y mejoras del modelo



PROBLEMA DE NEGOCIO

"Trayectoria" del año 2015 con los datos de 2011-2014

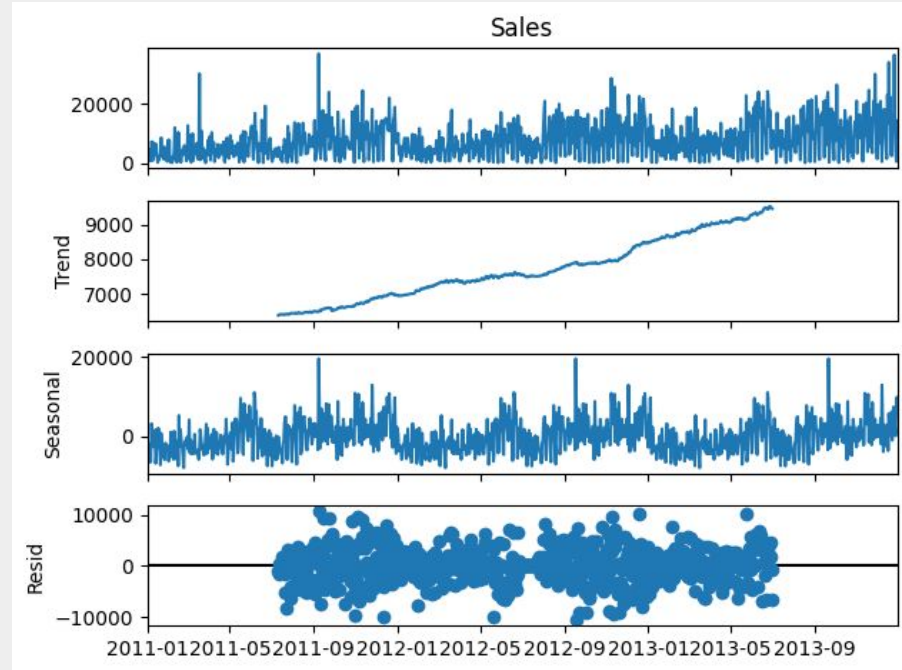
- Cálculo de gastos
- Previsión de picos/valles
- Inversiones y resultados a realizar

Base de datos: Order Sales de SuperStore (2011-2014)

- Bibliotecas utilizadas:
- Pandas
- Numpy
- Matplotlib
- statsmodels.tsa.statespace (SARIMA, ARIMA, adfuller, seasonal_decompose)
- sklearn.metrics,
- pmdarima

EXPLORACIÓN DE DATOS

Curva de ventas diaria, tendencia, estacionariedad y residual (diario train)



Alta seasonality y tendencia a la alza de los datos



Test de ADF diario (Stationarity)

La estacionariedad es una propiedad fundamental en series temporales que indica que las propiedades estadísticas de la serie (como la media, la varianza y la autocorrelación) no cambian a lo largo del tiempo.

Train ADF Test Statistic:
p-value: 0.015521291890962964

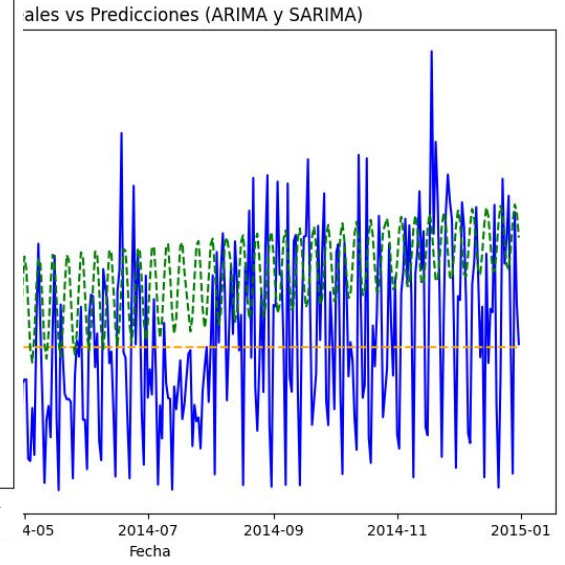
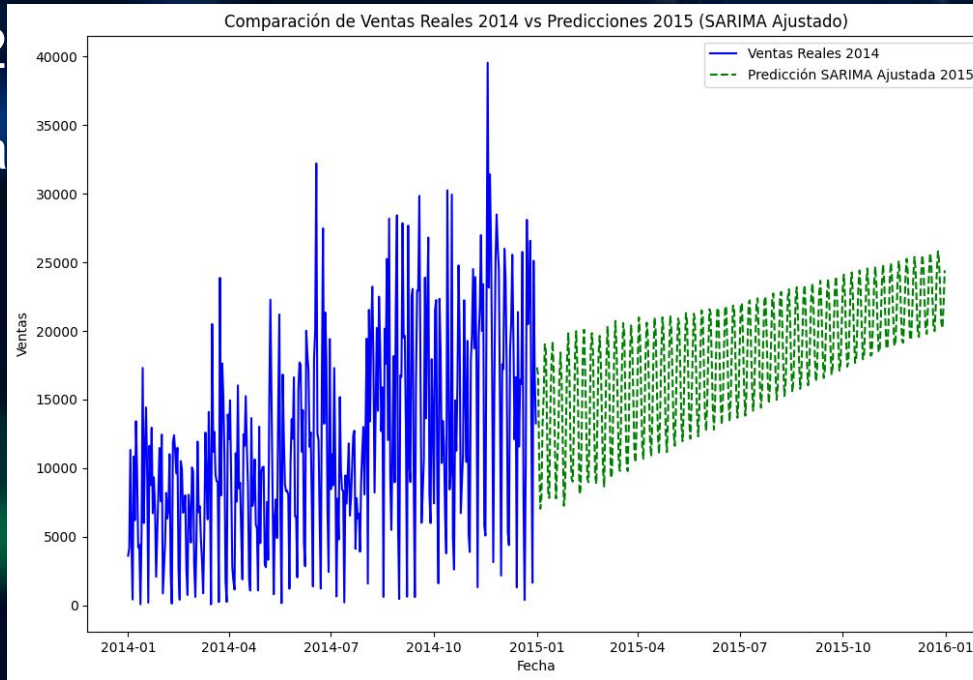
Test ADF Test Statistic:
p-value: 0.3938535201782304

- En los datos de entrenamiento, la serie parece ser estacionaria (p-value < 0.05).
- En los datos de prueba, la serie no es estacionaria (p-value > 0.05), por lo que necesitarías hacer un tratamiento (diferenciación u otras técnicas) para hacerla estacionaria.

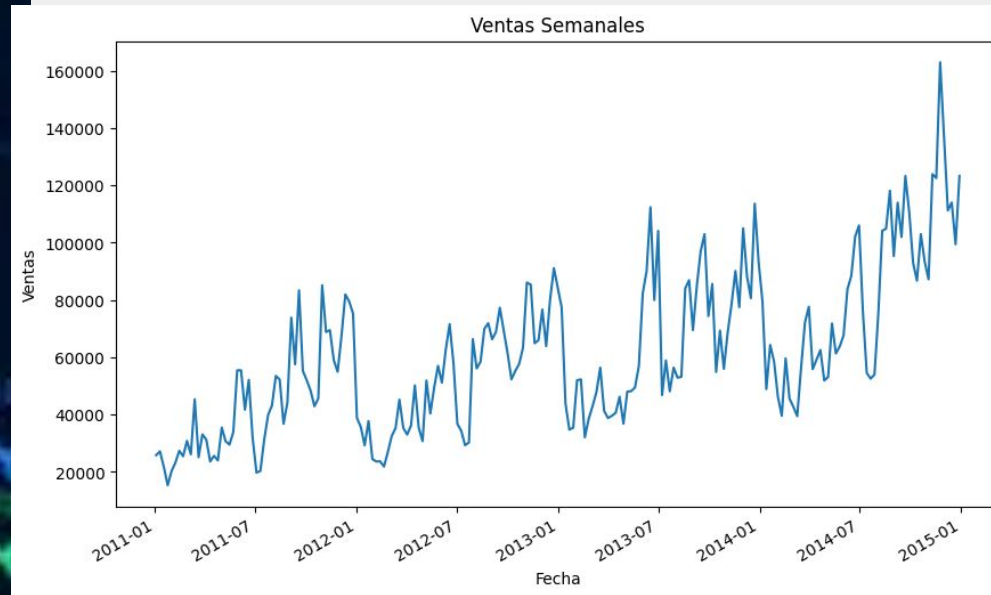
RESULTADOS MODELO (DIARIO)

- Entrenamiento con 2011-2014
- Evaluación del test 2014-2015

ARIMA - RMSE: 7736.27,
MAE: 6444.52
SARIMA - RMSE: 9776.02,
MAE: 6208.95

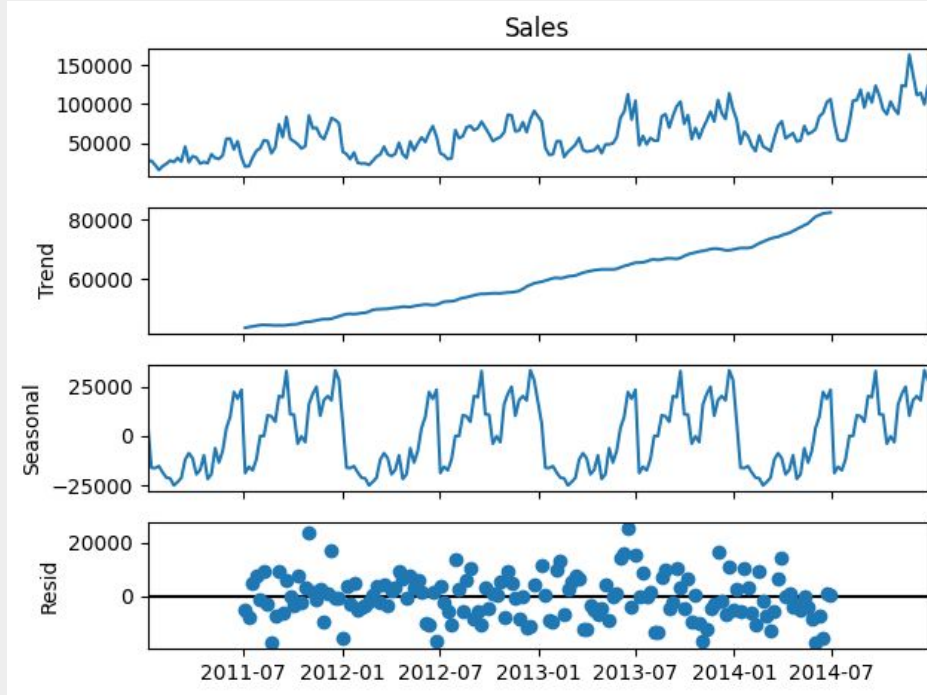


Y si planteamos el análisis de forma semanal?



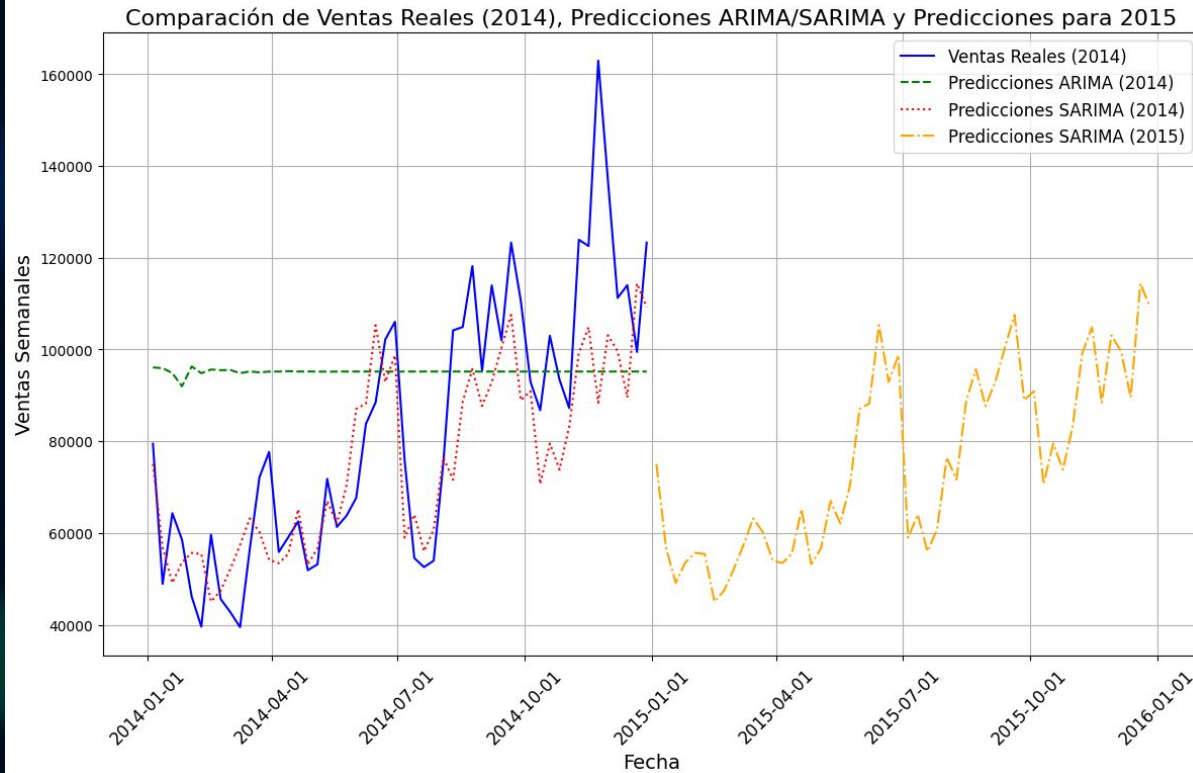
EXPLORACIÓN DE DATOS

Curva de ventas diaria, tendencia, seasonality y residual (semanal train)



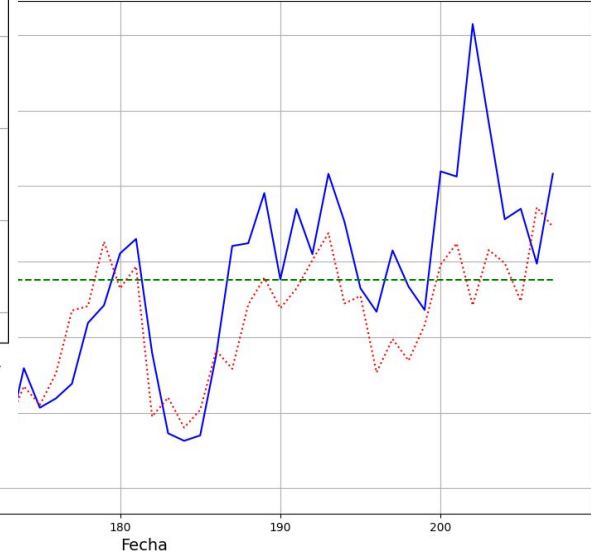
Alta seasonality y tendencia a la alza de los datos

RESULTADOS MODELO (SEMANALES)

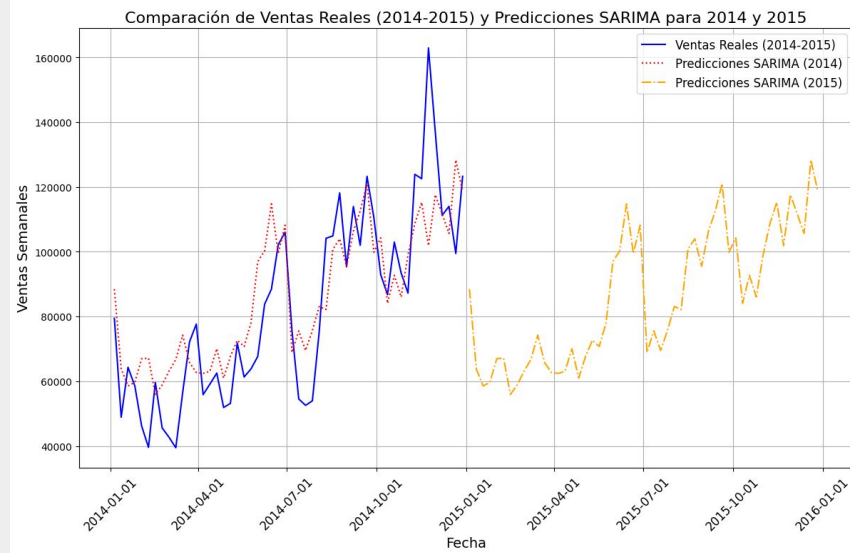
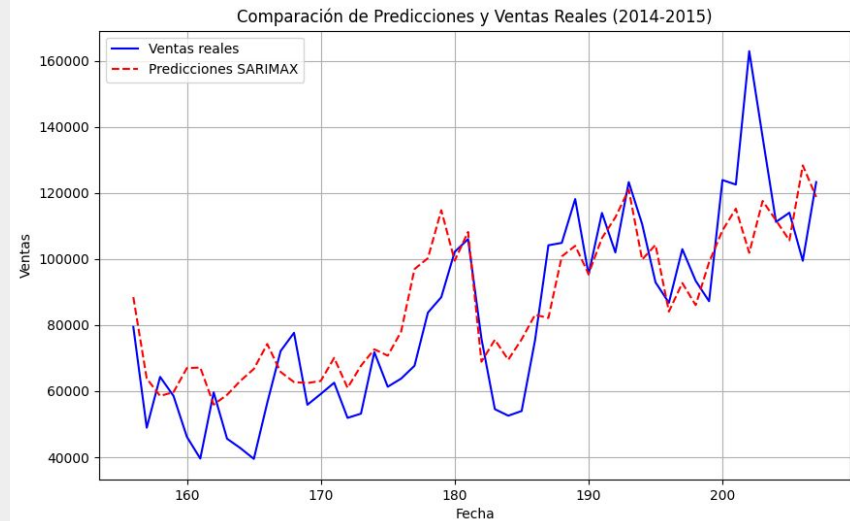


MSE: 31312.63,
675.77
RMSE: 17778.95,
159.39

Ventas Reales y Predicciones ARIMA/SARIMA



MEJORA DE PARÁMETROS (2014)



CONCLUSIONES

- El modelo, tanto diario como semanal, da bien las ventas globales del año pero tiene mayores dificultades para captar los picos/valles diarios de las cifras
- El modelo semanal capta mucho mejor la estacionalidad pero no acaba de adaptar la tendencia para el 2014 que tiene un crecimiento mayor que los años 2011-2013
- El problema de stationary se soluciona pero lo que ha dado más dificultades es igualar el número de semanas entre años
- Solo se ha trabajado con ARIMA y SARIMA como modelos pero hubiera podido estar bien trabajar con Prophet para captar mejor la estacionalidad y tendencia de ventas



GRACIAS