Springboard Capstone Project 2

Time Series Forecasting for Sales Order

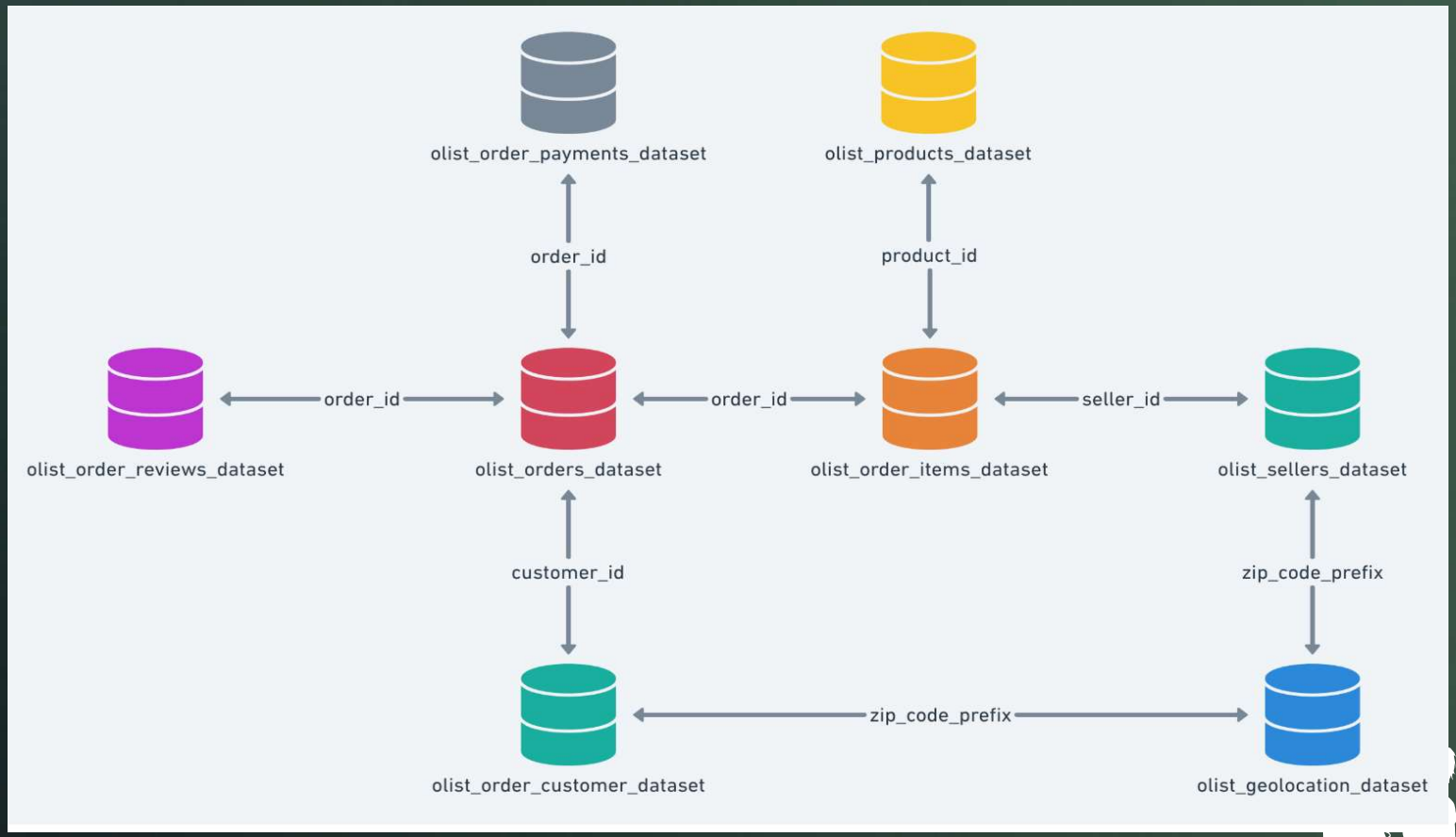
Presented by: Cuili Huang

Agenda

1. Introduction
2. Data Wrangling
3. Time Series Analysis
4. Time Series Forecasting (Base Model)
5. Tune Hyperparameters
6. Conclusion

Introduction

Brazilian E-Commerce Company Olist Datasets



Data Wrangling

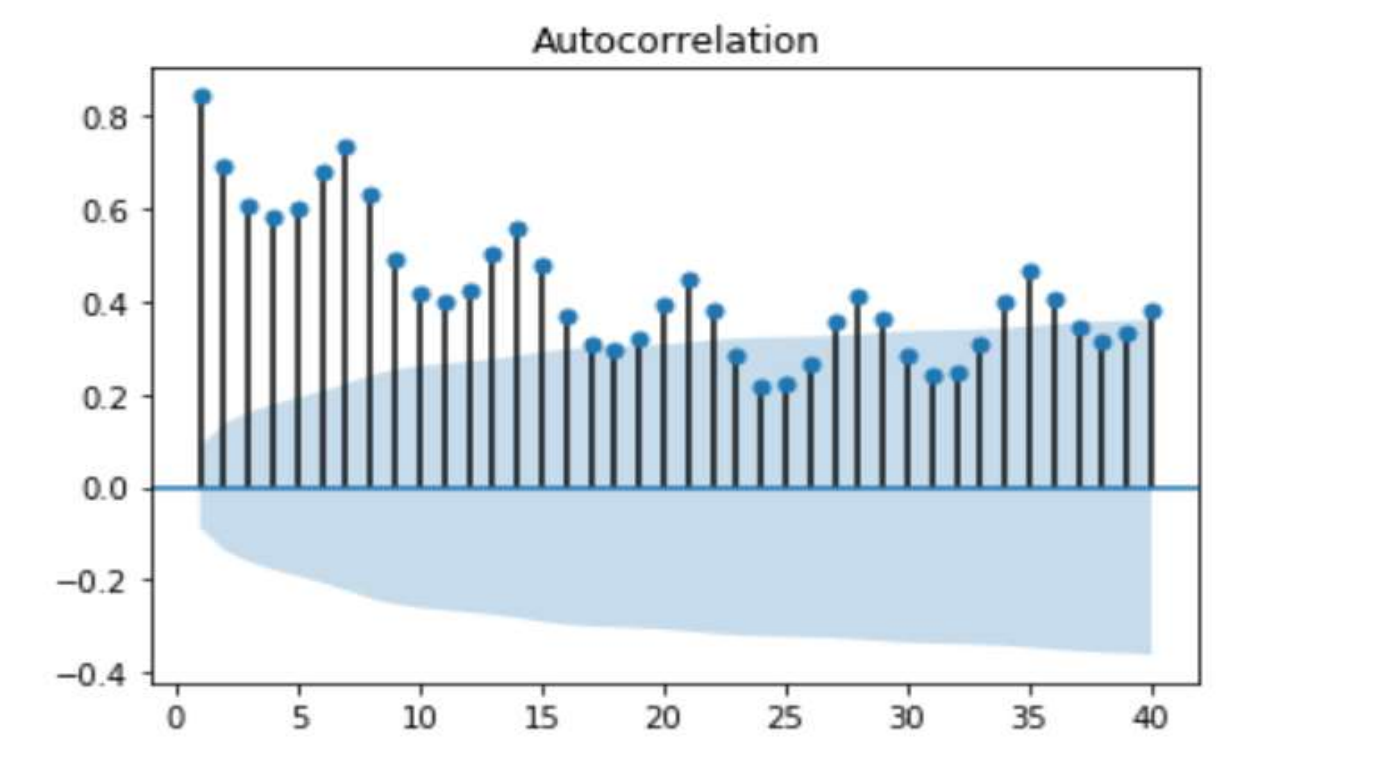
Changed datetime format

Aggregated ‘order_id’ by day

Handled missing values and outliners

Time Series Analysis

Plot autocorrelation function (ACF) to discover seasonality

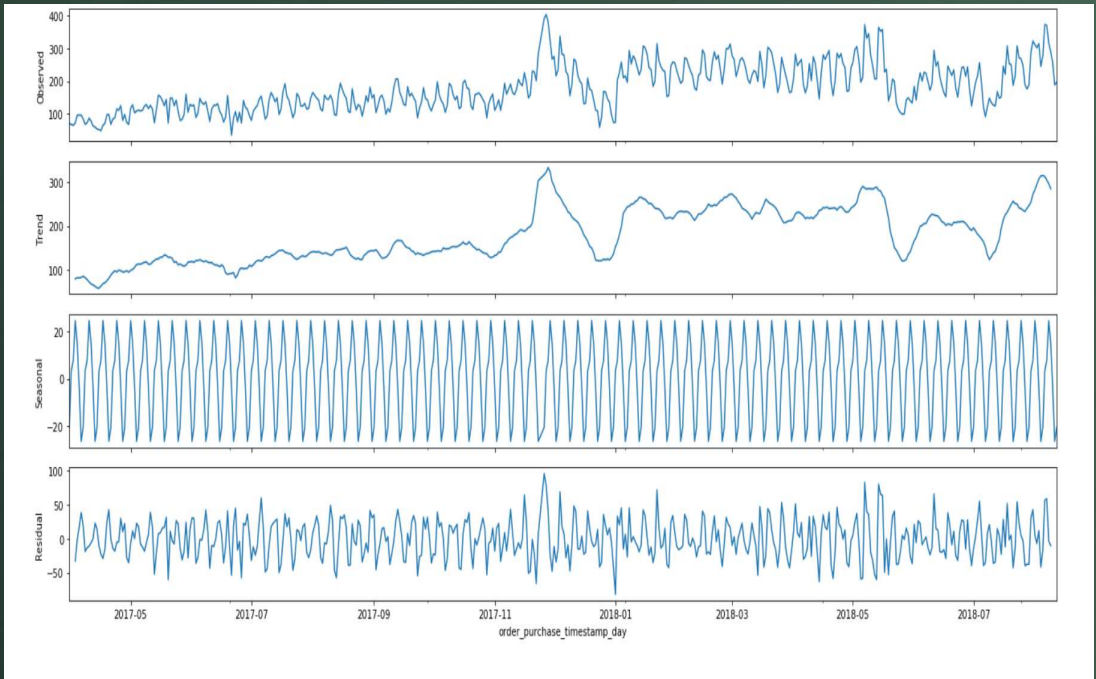


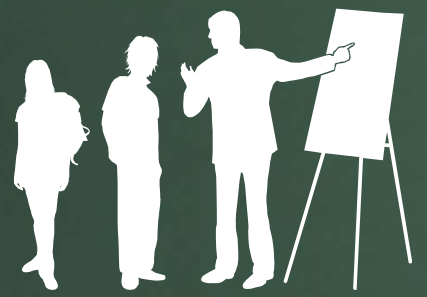
Time Series Analysis

Decompose the time series:

1. Slightly upward trend;

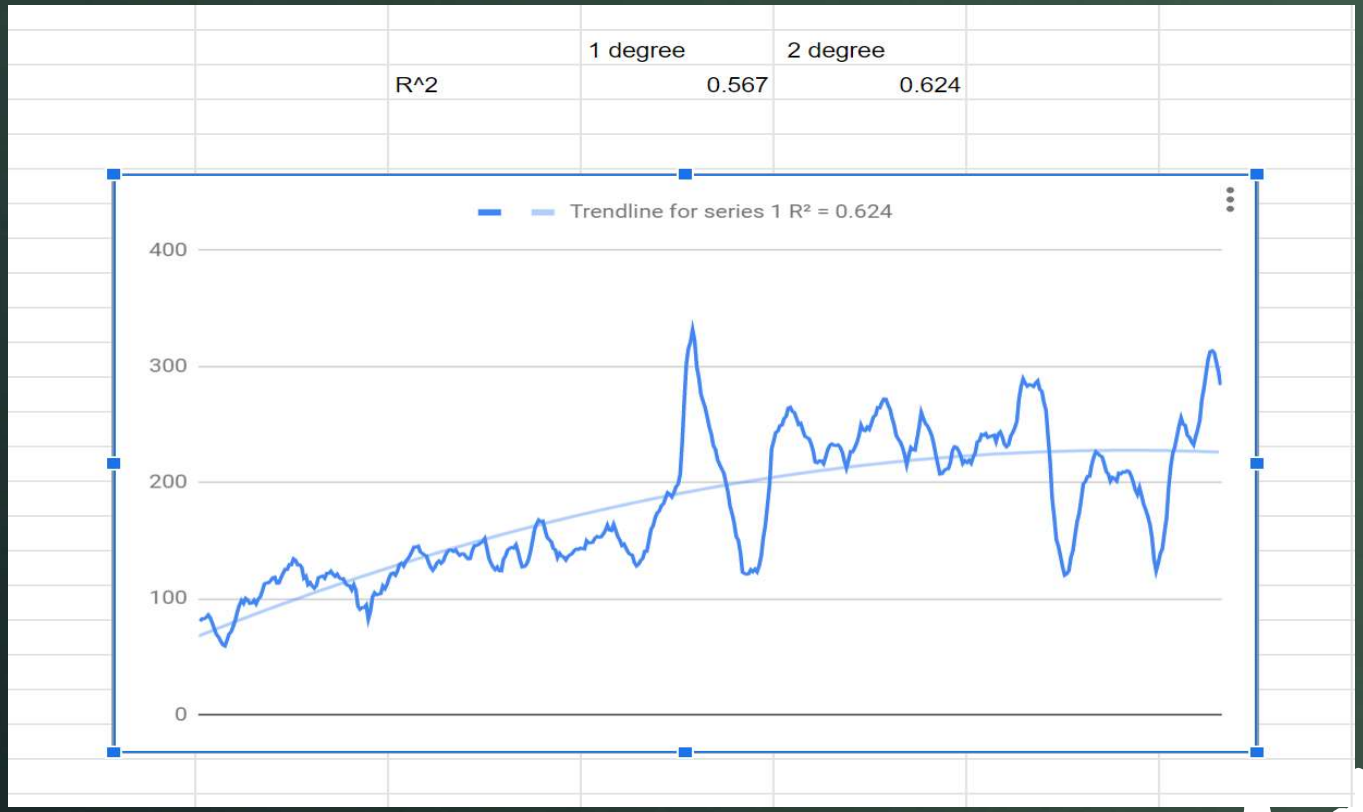
2. Strong seasonality;

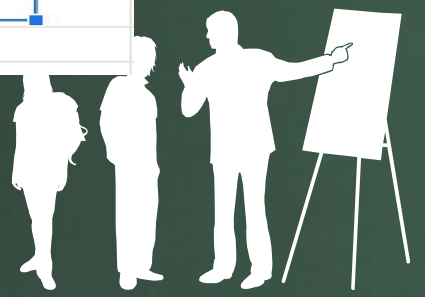




Time Series Analysis

- Trendline research for trend data





Time Series Forecasting (Base Model)

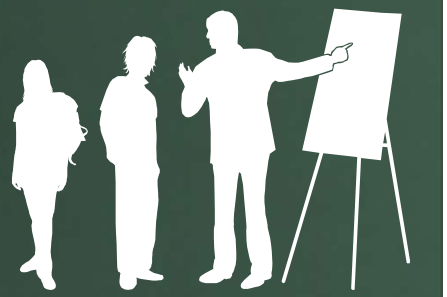
1. Use SARIMAX model;

2. Use the first 80% of data as training dataset;

3. Performed one-step-ahead forecast on the latter 20% of data;

4. Evaluation metric: Symmetric Mean Absolute Percentage Error (SMAPE); base model SMAPE: 6.78%;

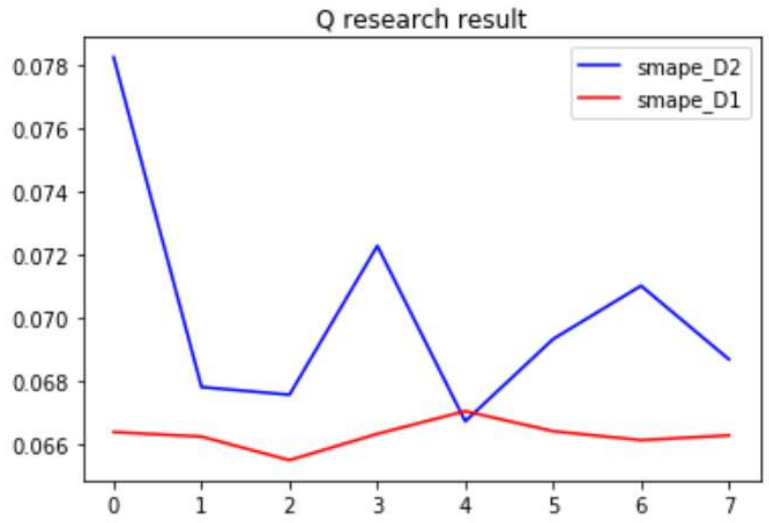
$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{|A_t| + |F_t|}$$



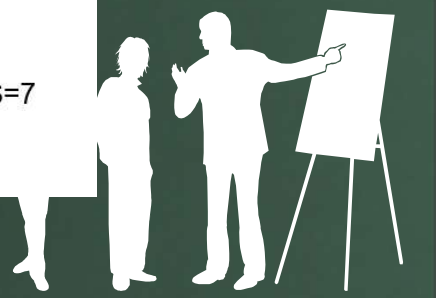
Time Series Forecasting

Tune Hyperparameters

1. P=7, D=2, S=7, search for Q;
2. P=7, D=1, S=7, search for Q;

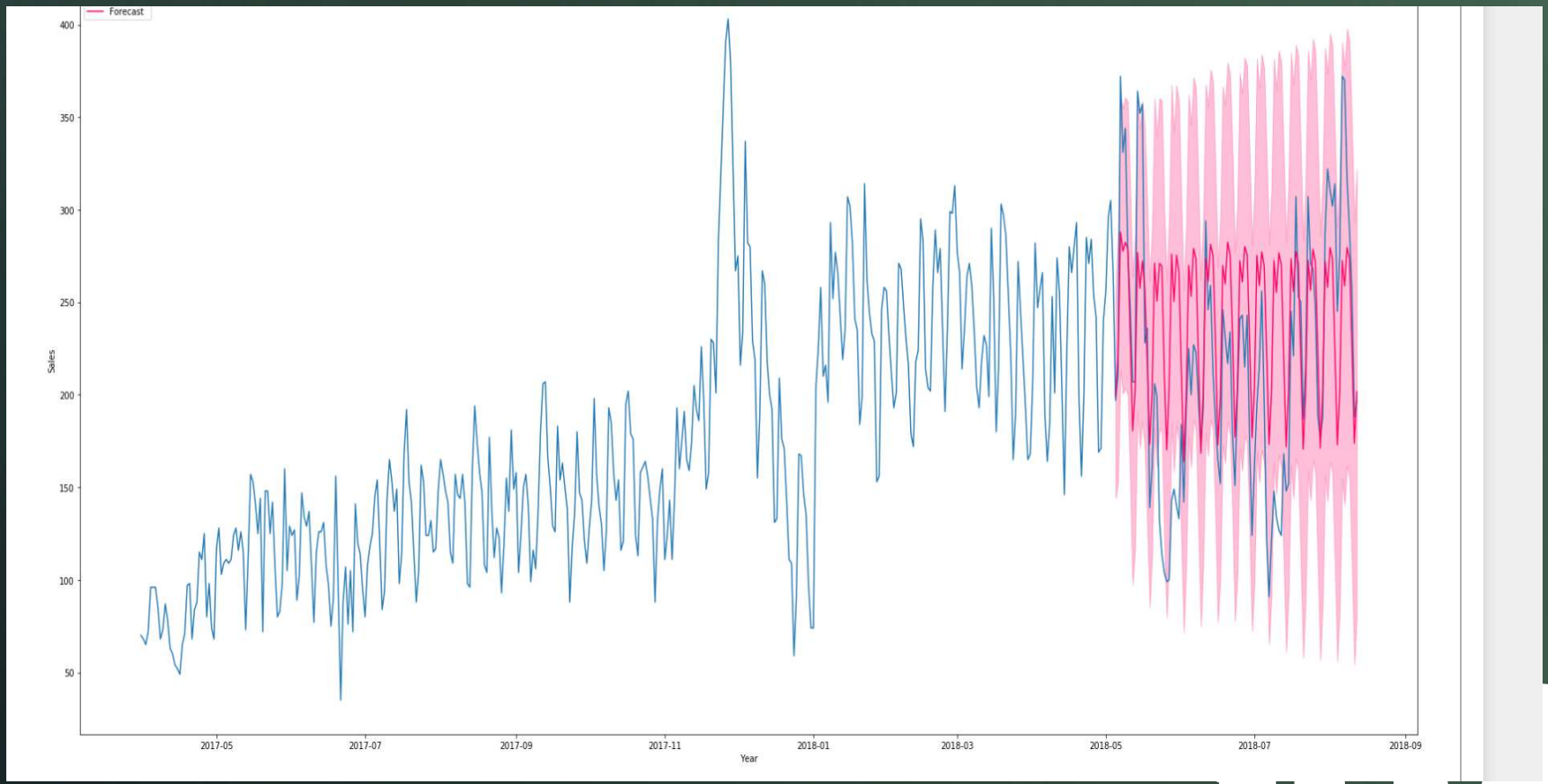


Based on the research above, the optimal values for the hyperparameters are P=7, D=1, Q=2, S=7



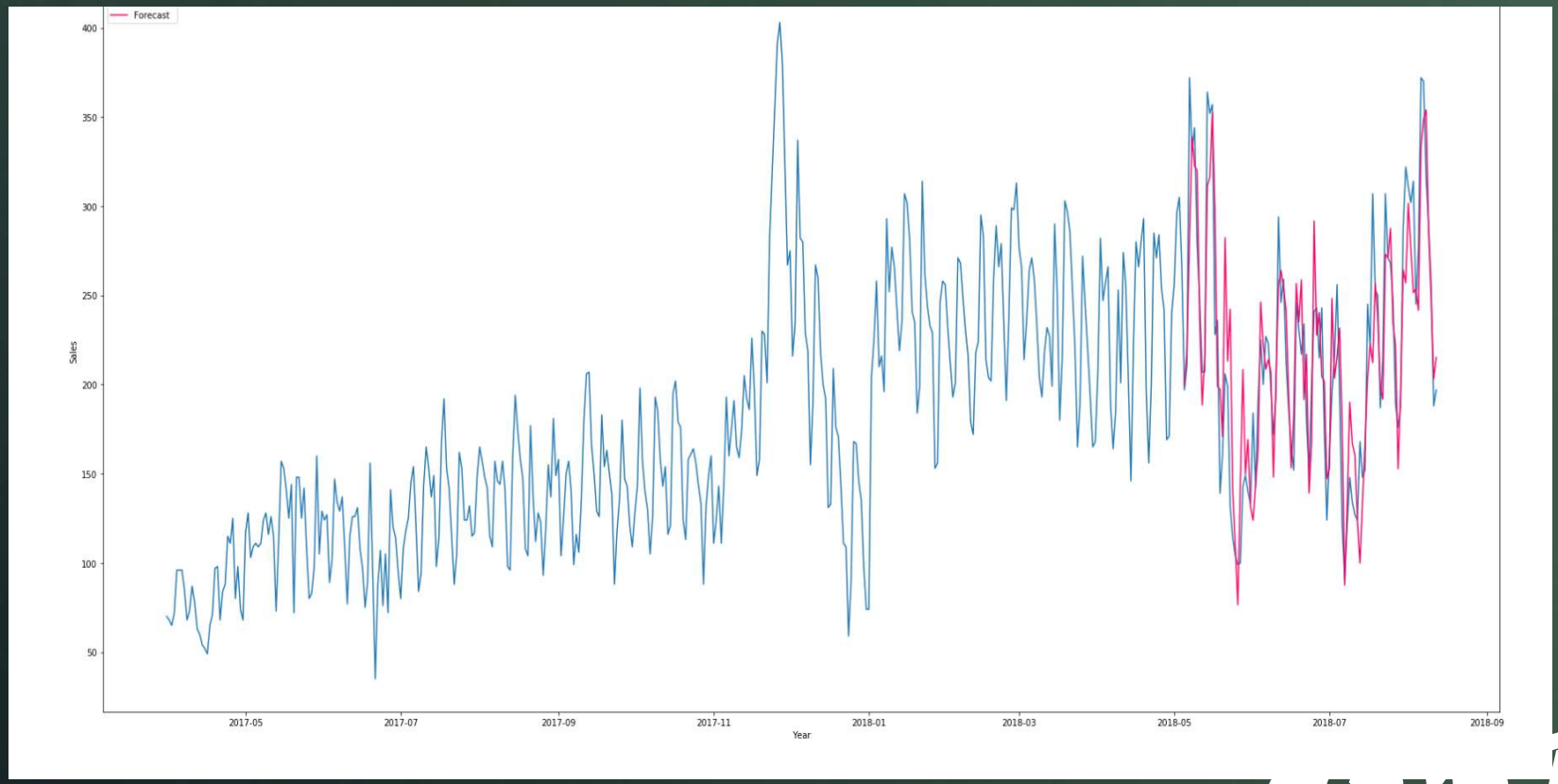
Conclusion

Perform forecast for the next 100 days



Conclusion

Perform one-step-ahead forecast



Conclusion

Summary

1. The sales data display a slight upward trend and strong seasonality;

2. Although trendline research suggested D=2, hyperparameters research suggested that the SARIMAX model performed better when D=1;

3. Perform one-step-ahead prediction will yield more accurate results comparing to few-step-ahead prediction;


QUESTION