Sales Forecast

Elizabeth Sebastian

Contents

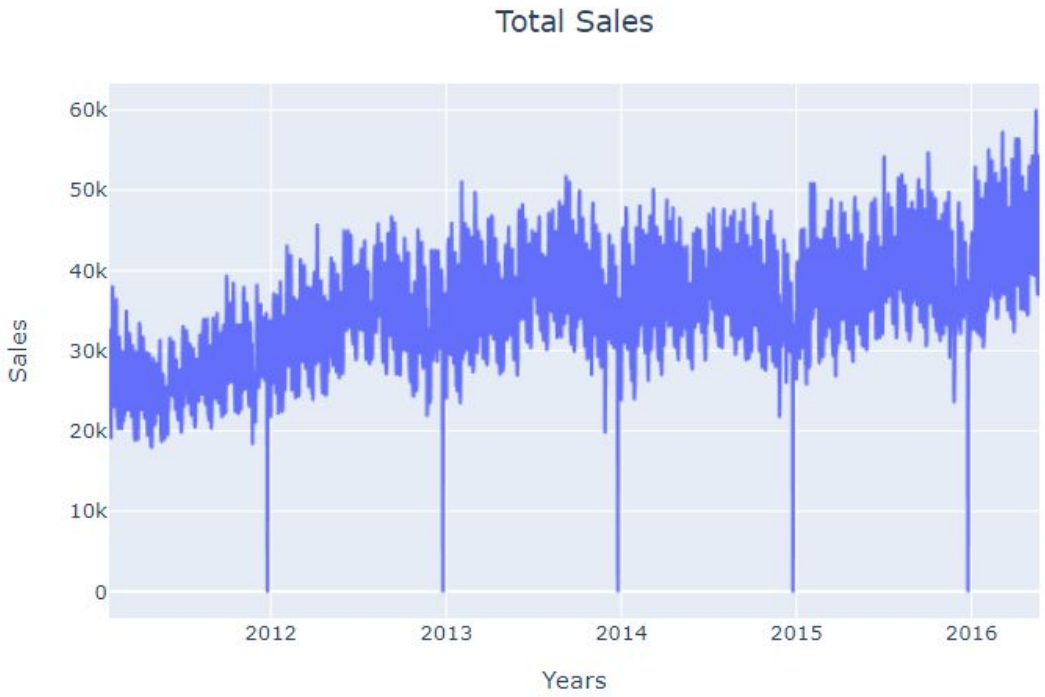
Background

Sales forecasting is essentially involves predicting your future sales/profit, based on the sales that your dealership has on order. The purpose of this is to give you an insight into your margins so that you are able to manage your business more efficiently. This estimation certainly helps different companies to increase their revenues.

Problem Statement

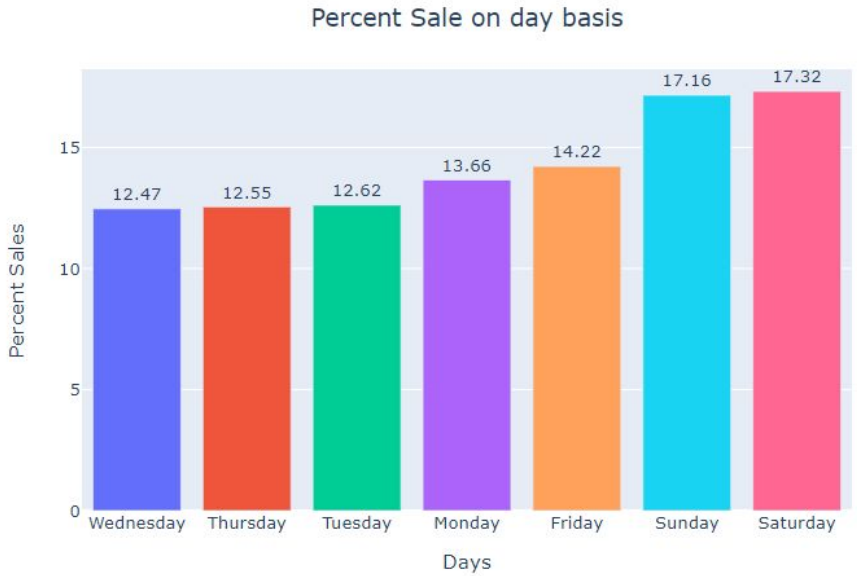
The main objective is to estimate or predict the total sales of Walmart retail goods at stores in various locations for the next 28-days. The prediction is based on 5 years of historical daily total sales data. This project intended to evaluate and compare a number of statistical forecasting algorithms on the given data. SMAPE (symmetric mean absolute percentage error) is used to quantify the accuracy of all forecasts and to compare different forecasting algorithms. A lower SMAPE means higher accuracy.

Data

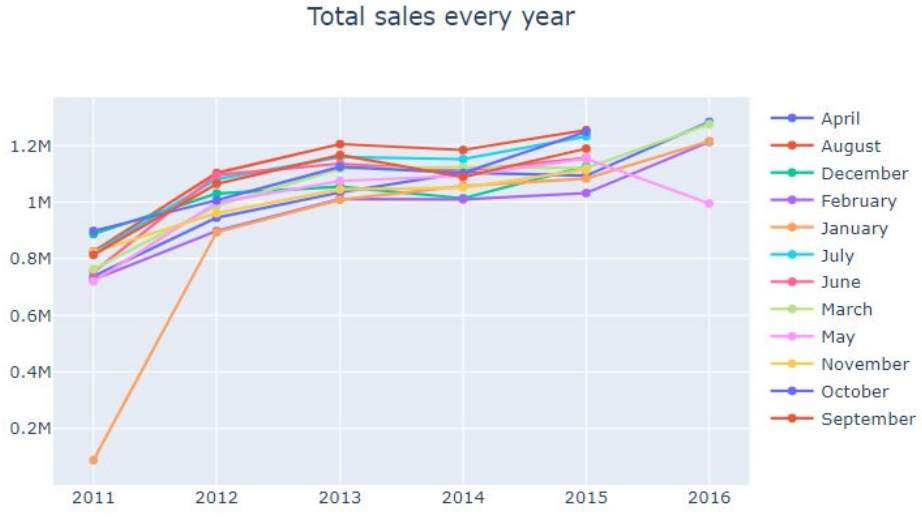


- 1941 days columns/features, ~5 years of data
- Christmas the sales greatly fall less than 50
- Demand is growing every year

EDA

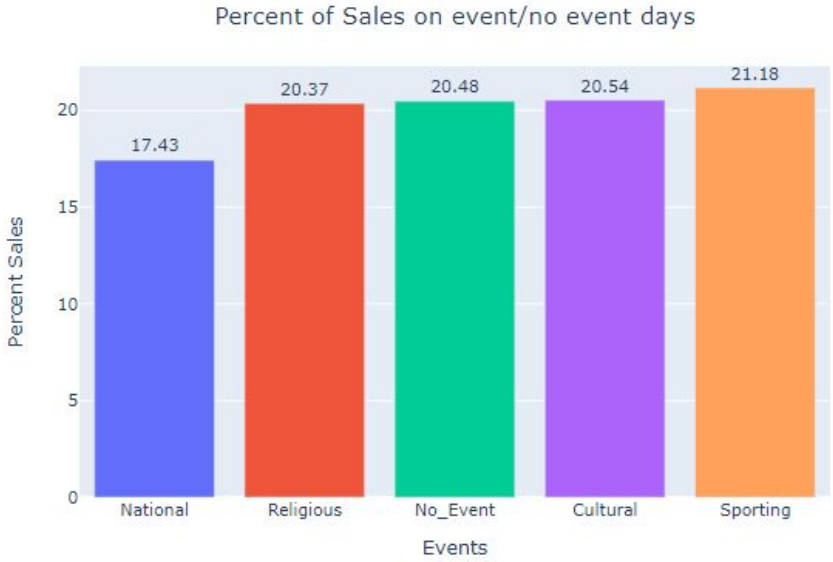


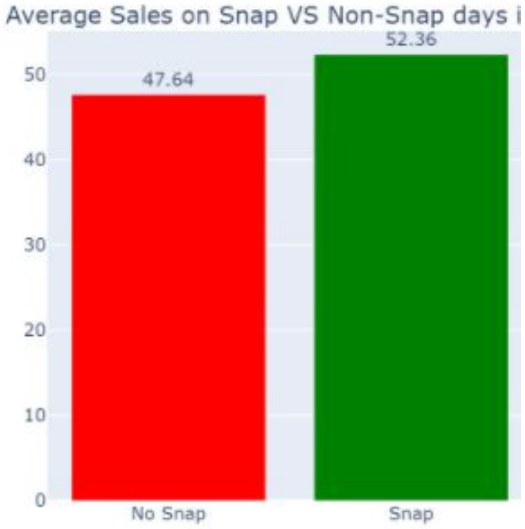
- Huge number of sales happening during the weekends.



- Since 2012, August rates the highest sales about 1.2M products July either ranks 2nd or 3rd on total sales. At 2016, the sales numbers have increased almost 20%.

EDA





- The sales happened during sporting event times are slightly more and that happened on National event days are little lower.

- Sales on Snap days are high than Non-Snap days.

Temporal_train_test_split

Models using SKtime

Naive/
Seasonal Naive

Exponential
Smoothing

Auto ETS

Auto Arima

XGBoost

LightGBM

Ensemble

BATS/TBATS

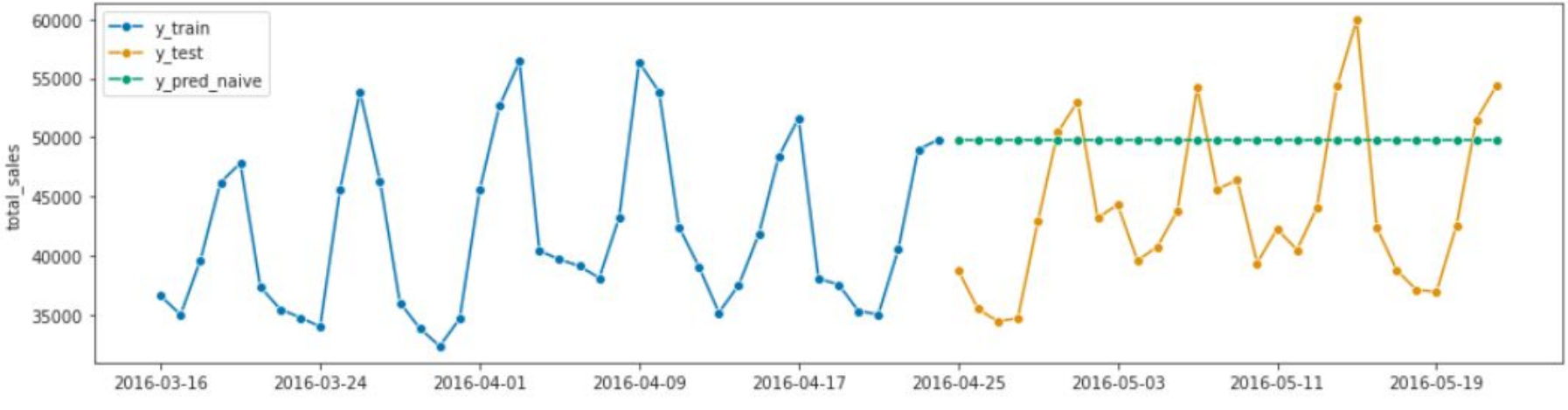
Theta

FB Prophet

Random Forest

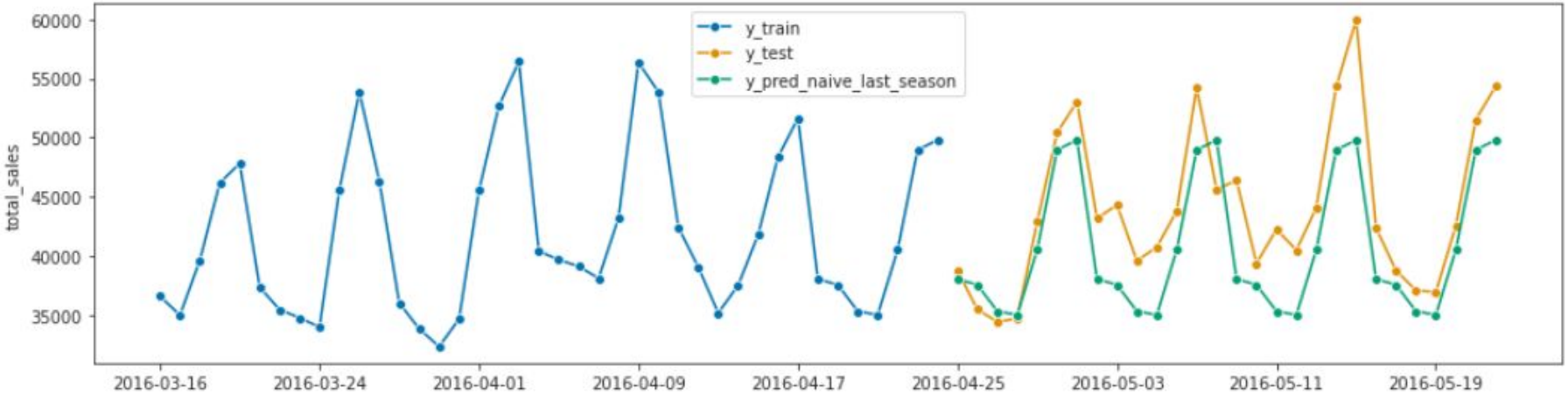
Naive Model

- Predicting the last value
- SMAPE Loss = 0.173581



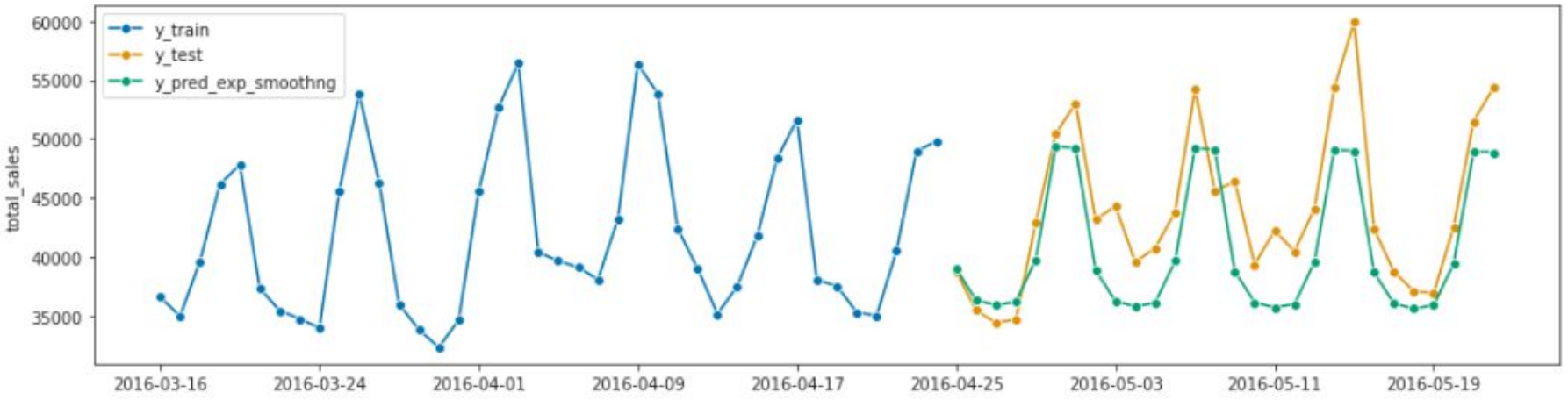
Seasonal Last Seasonal Naive Model

- Predicting the last season
- SMAPE Loss = 0.087535



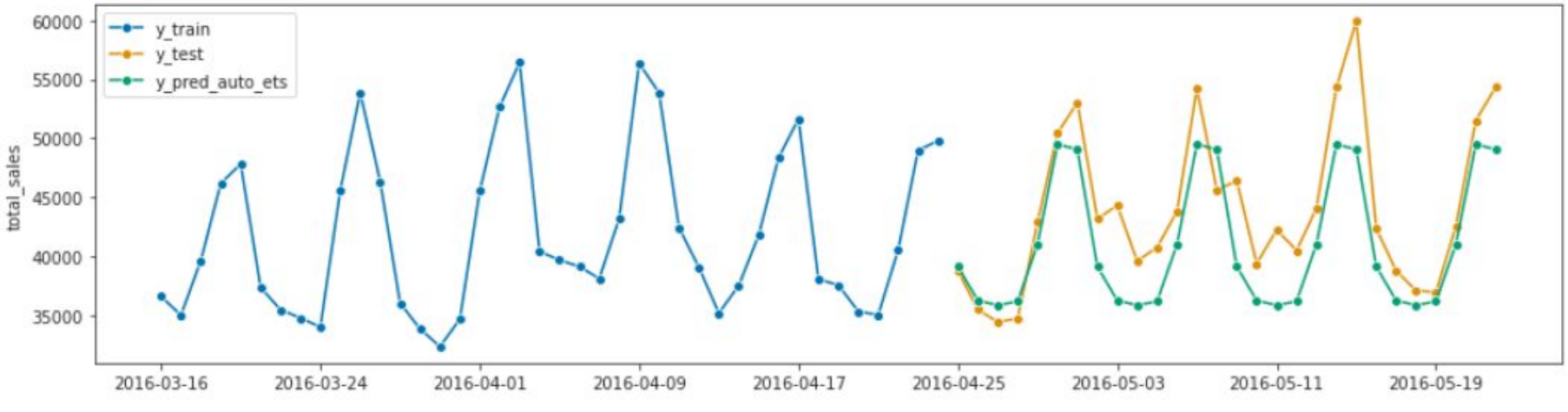
Exponential Smoothing

- Uses Exponential Window Function
- SMAPE Loss = 0.08931



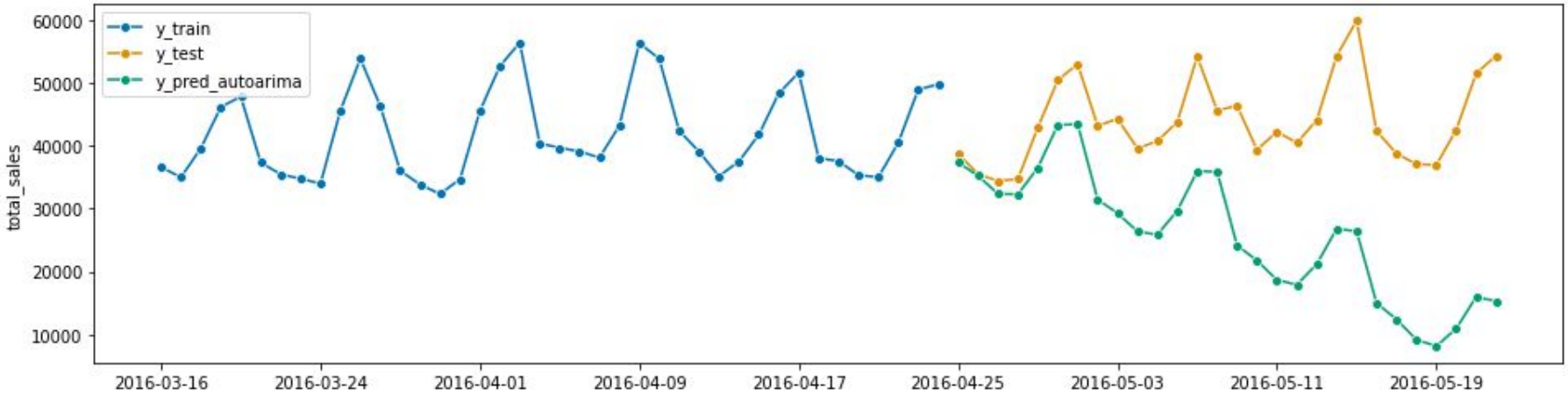
AUTO ETS

- Automated Exponential Smoothing
- Automatic Model Selection
- SMAPE Loss = 0.081176



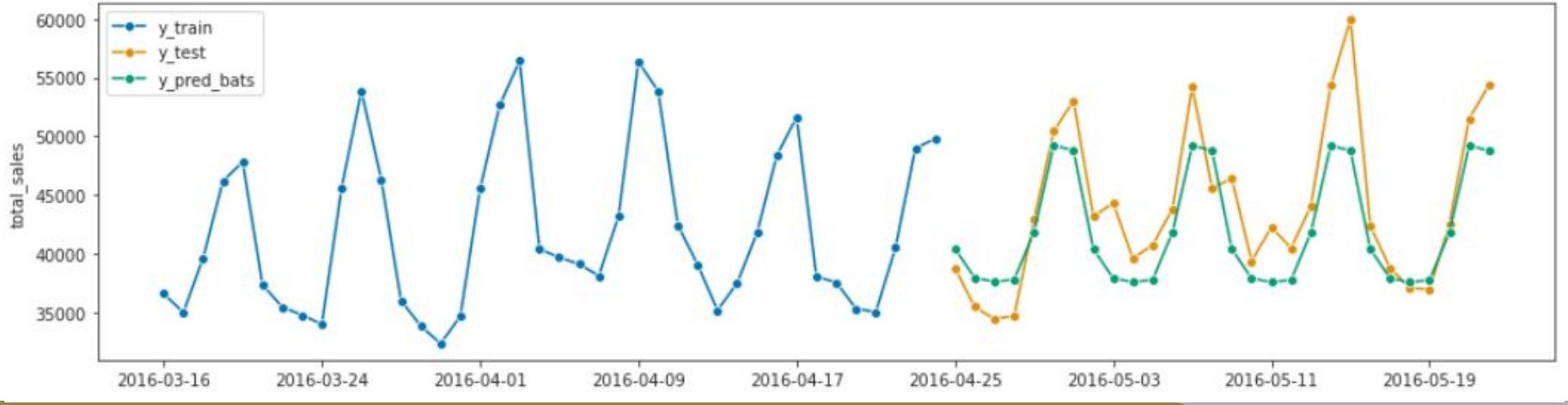
Auto ARIMA

- Autoregressive Integrated Moving Average
- SMAPE Loss = 0.571929



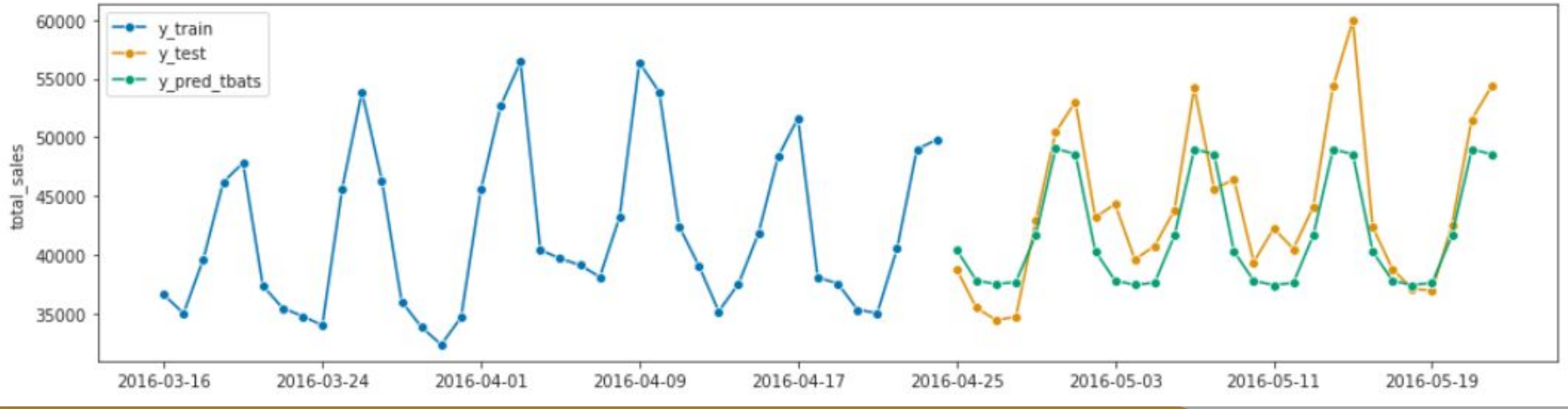
BATS

- Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components
- Fit the best performing model
- SMAPE Loss = 0.070344



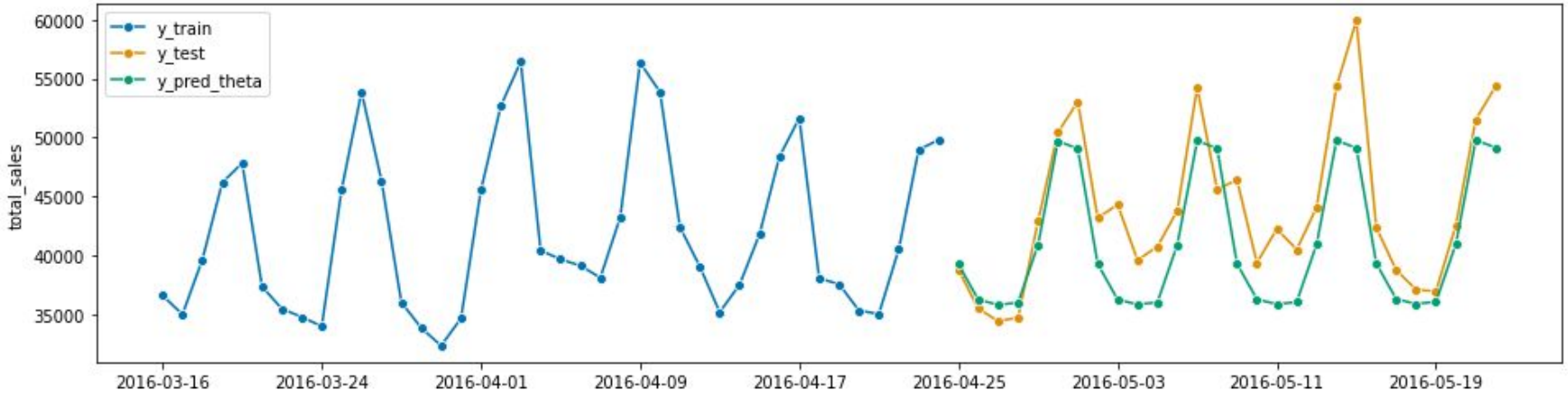
TBATS

- Exponential smoothing state space model with Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components
- Automatically fit the best performing model
- SMAPE Loss = 0.072434



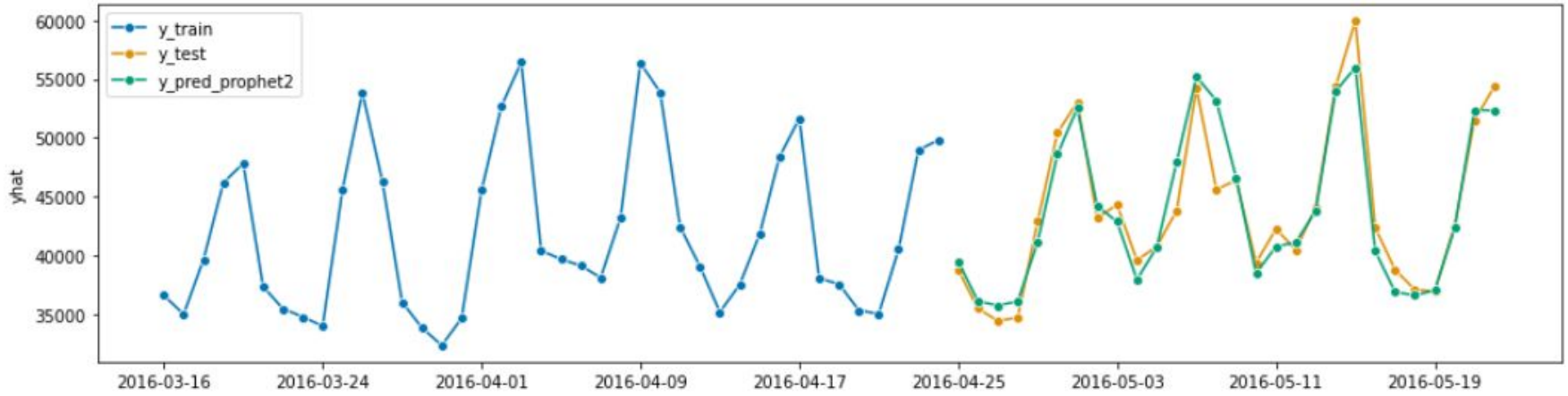
ThetaForecaster

- Equivalent to simple exponential smoothing (SES) with drift.
- SMAPE Loss = 0.080372



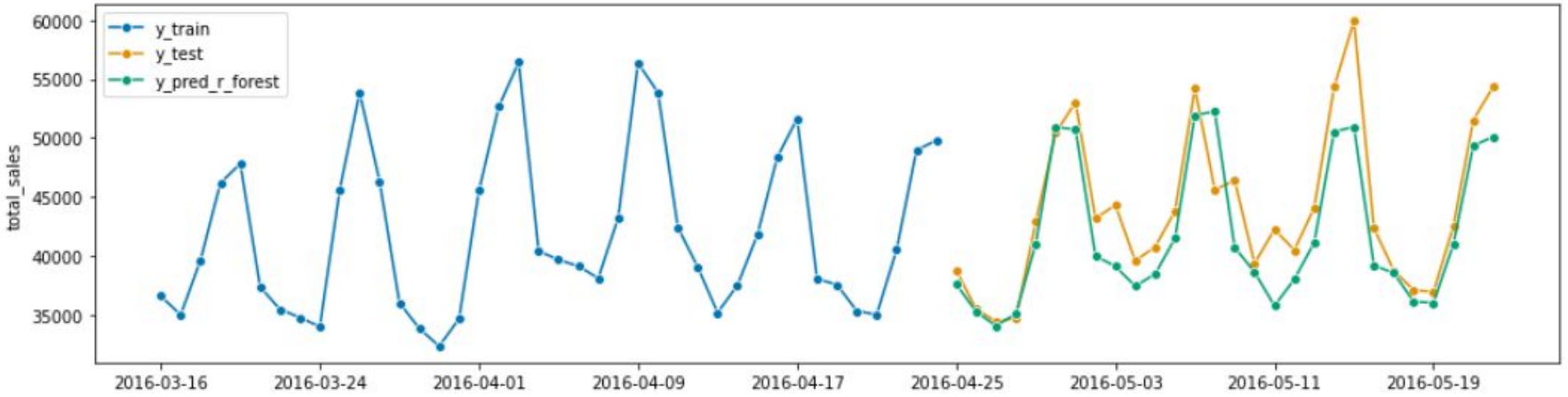
Prophet

- The best performing model
- Tried with 2 different API(SKtime, FBprophet)
- Both the API produced similar result.
- SMAPE Loss = 0.031872



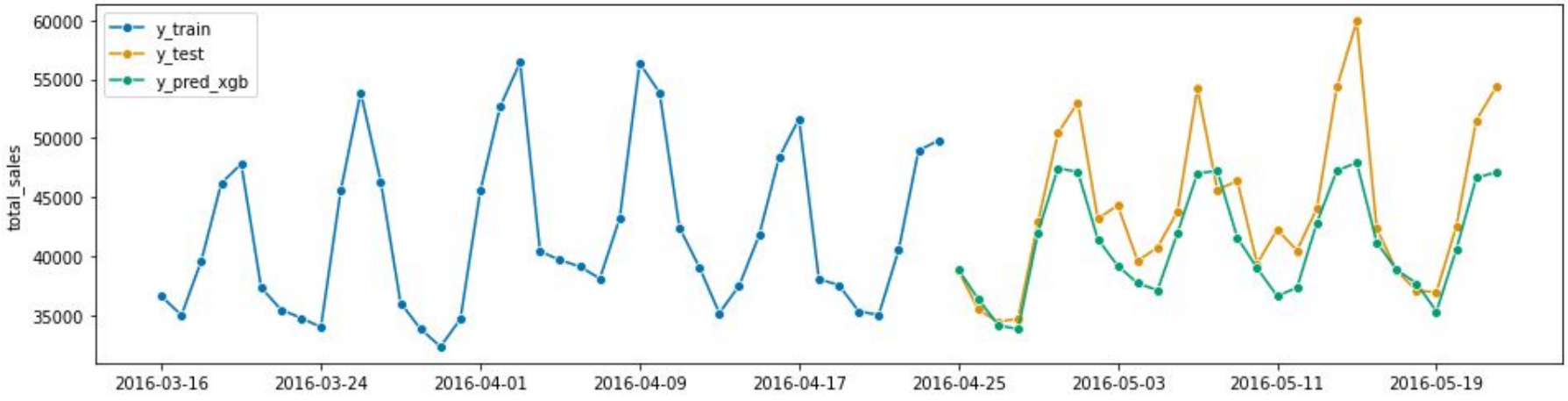
Random Forest

- Construct a multitude of decision trees and output mean prediction of individual trees
- SMAPE Loss = 0.060069



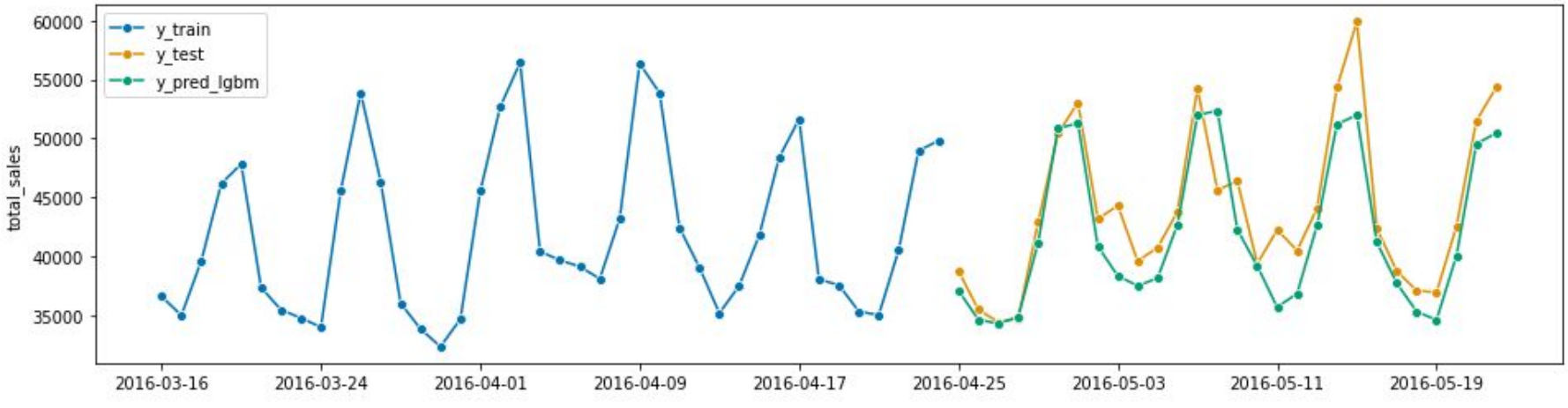
XGBRegressor

- Extreme Gradient Boosting
- SMAPE Loss = 0.067767



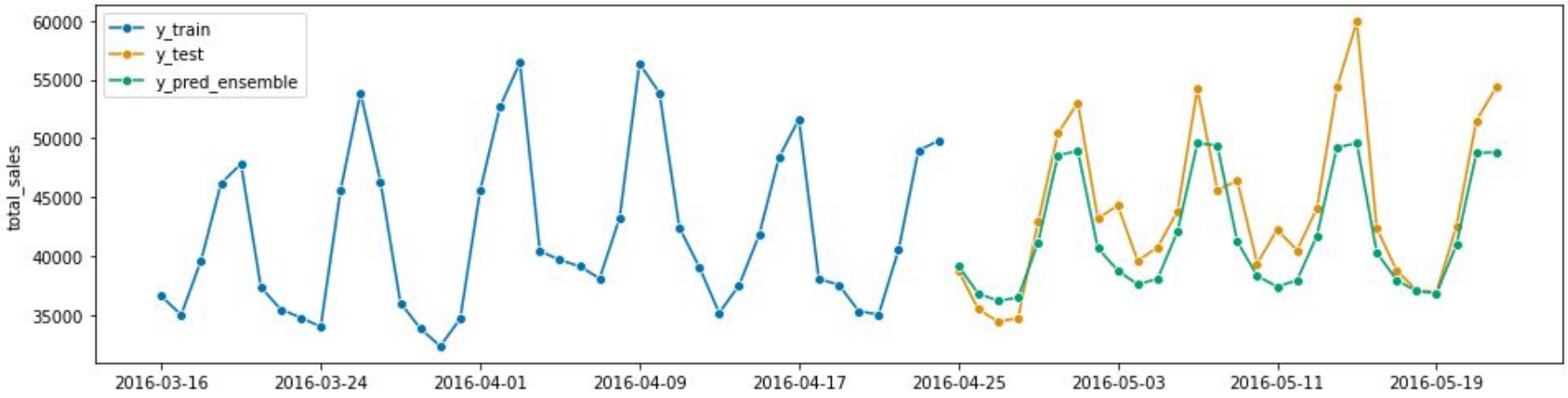
LGBM Regressor

- Light Gradient Boosting Machine
- Similar to XG Boost
- SMAPE Loss = 0.060069

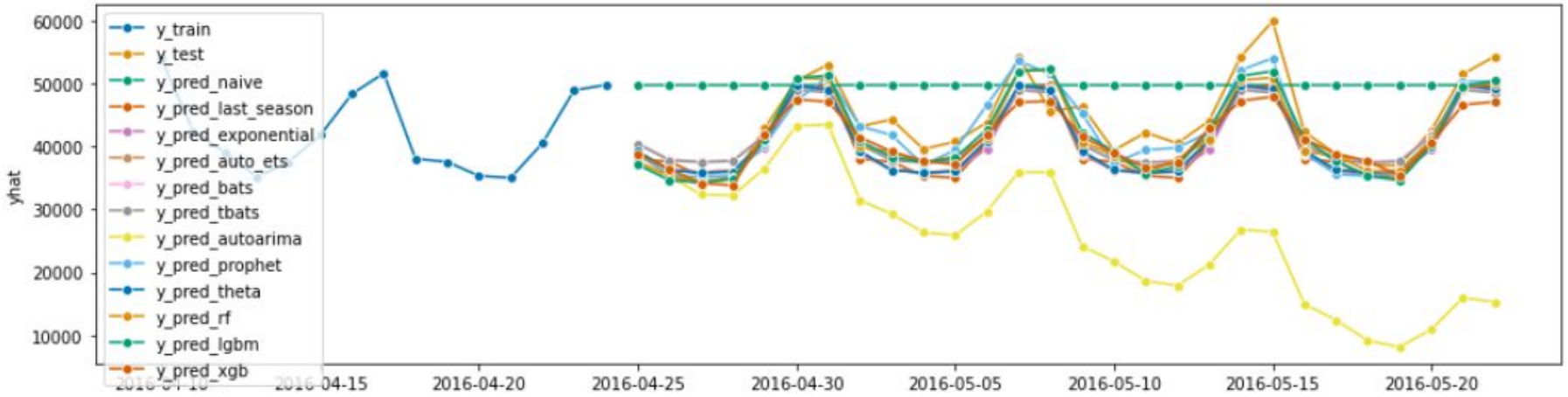


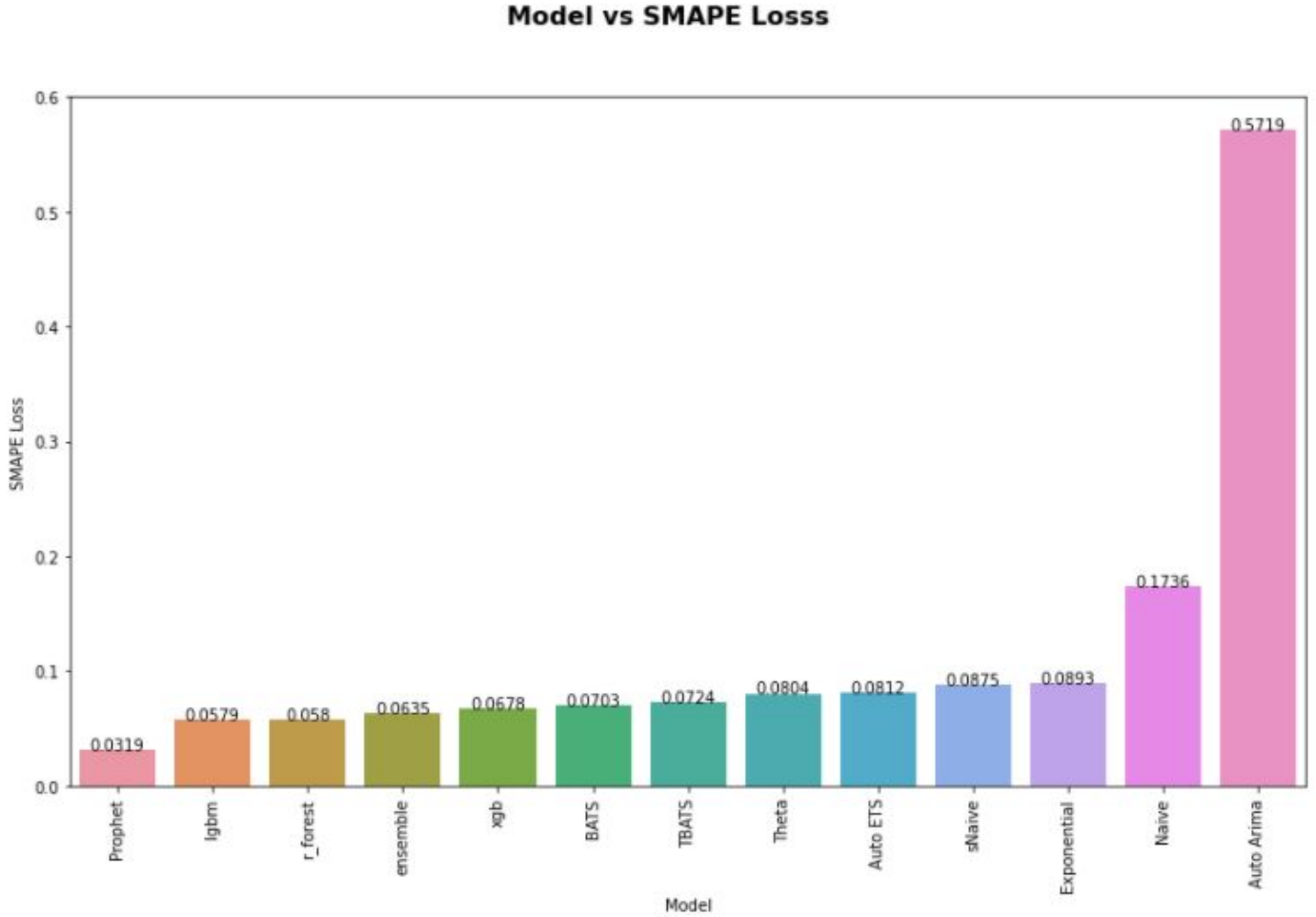
SKtime Ensembler

- Ensemble multiple forecasting algorithms
- SMAPE Loss = 0.063478



Compare Different Models




Model vs SMAPE Losss
SMAPE Loss
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.0319
0.0579
0.058
0.0635
0.0678
0.0703
0.0724
0.0804
0.0812
0.0875
0.0893
0.1736
0.5719
Prophet
lgbm
r_forest
ensemble
xgb
BATS
TBATS
Theta
Auto ETS
sNaive
Exponential
Naive
Auto Arima
Model

Conclusion

- Successfully predict the total sales of Walmart retail goods at stores in various locations for the next 28-days.
- Evaluated and compared a number of statistical forecasting algorithms on the given data.