Retail Giant sales forecasting- Time series analysis

By - Akarsh Tyagi

Problem Statement

- Global Mart is an online supergiant store that has worldwide operations. This store takes orders and delivers across the globe and deals with all the major product categories (segment) — consumer, corporate and home office and 7 major markets.
- On combining market and segment we get 21 different market segments.
- As a sales manager for this store, we have to forecast the sales of the products for the next 6 months.
- Not all of the 21 market segments are important from the store's point of view. We need to find out the most consistently profitable market-segment from the given data and forecast the sales and demand for that single market-segment only.

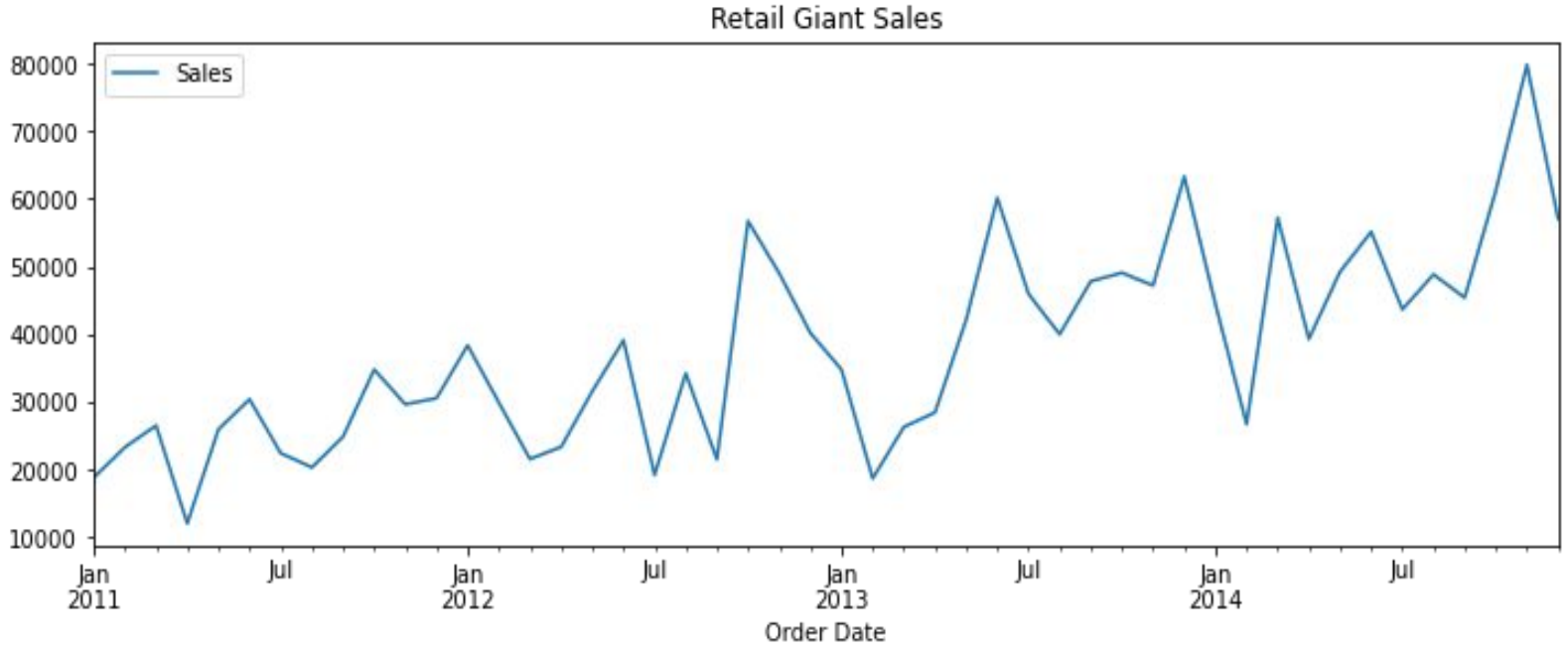
Best market-segment

- In order to find most consistently profitable market-segment we need to make a pivot table on the column names market-segment and order dates.
- Then we will be splitting the data into train and test set(42:6).
- We find the coefficient of variation of Profit on the train data.
- Now the best best market-segment will be with the least CoV value.

Comparing The coefficient of variance

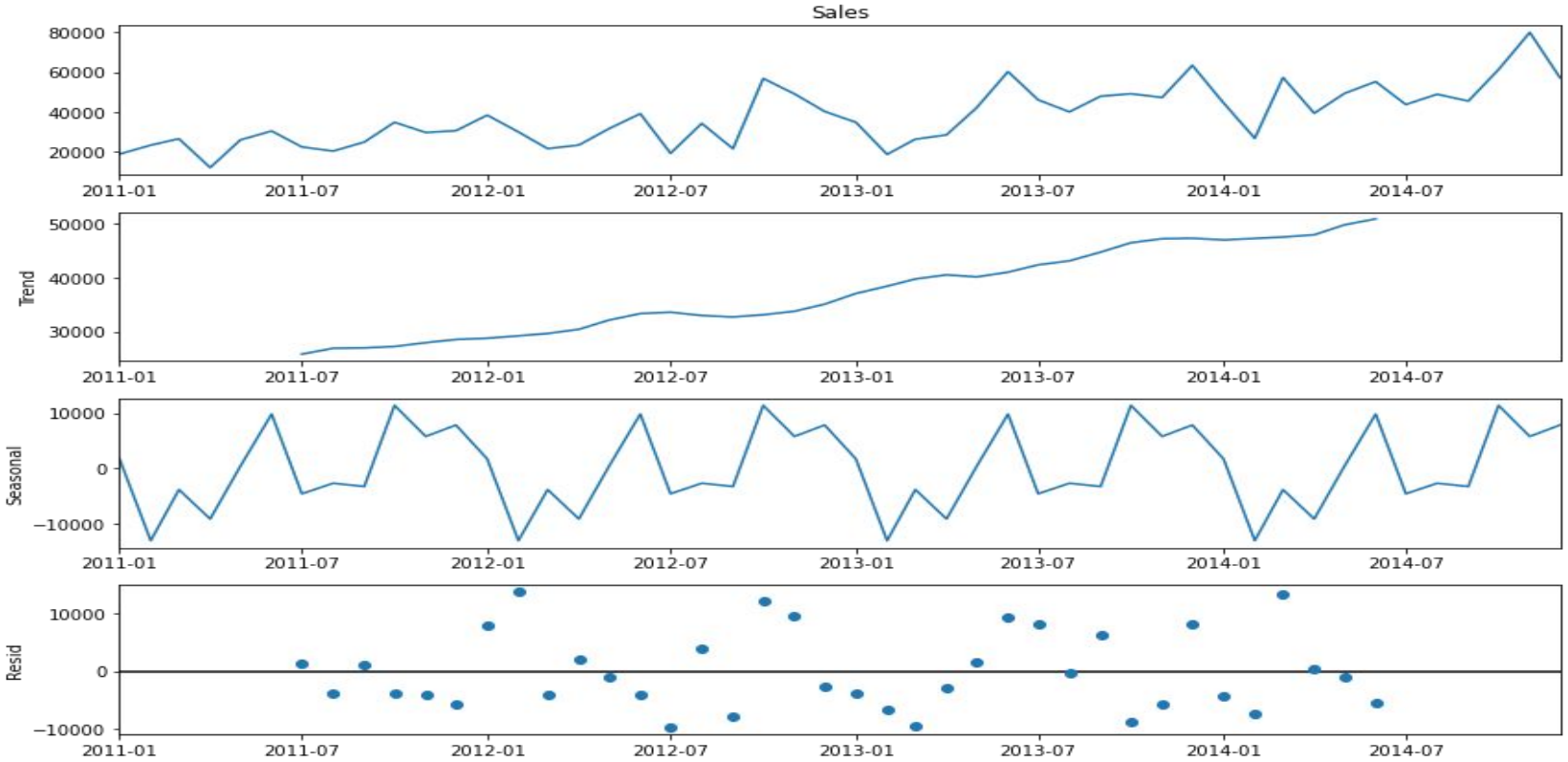
- On comparing Coefficient of variation among 21 different market segments.
- We get that APAC Consumer is having lowest Coefficient of Variation.
- Hence we find that APAC Consumer is most consistently profitable market segment. We will forecast for the same
- In probability theory and statistics, the **coefficient of variation** (**CV**), also known as **relative standard deviation** (**RSD**), is a standardized measure of dispersion of a probability distribution or frequency distribution. It is often expressed as a percentage, and is defined as the ratio of the standard deviation to the mean

Market-Segment	CoV
APAC Consumer	0.529061
APAC Corporate	0.536476
EU Consumer	0.602430
LATAM Consumer	0.692059
EU Corporate	0.730829
LATAM Corporate	0.892870
EU Home Office	0.949443
APAC Home Office	1.020441
US Consumer	1.022779
US Corporate	1.084821
US Home Office	1.137656
LATAM Home Office	1.183872
Canada Consumer	1.267099
Africa Consumer	1.326234
Canada Corporate	1.812884
Africa Corporate	1.914675
Africa Home Office	2.037337
Canada Home Office	2.420662
EMEA Consumer	2.684648
EMEA Corporate	6.432058
EMEA Home Office	7.828128



Time series Analysis

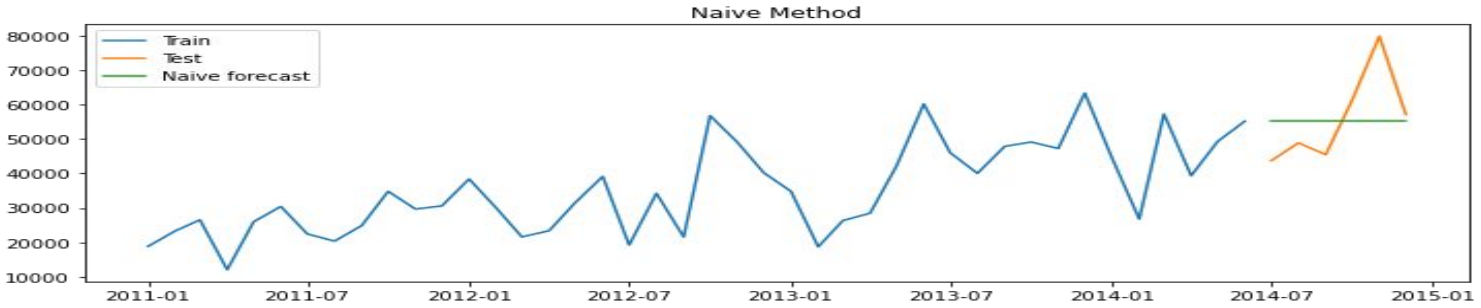
Plotting time series data

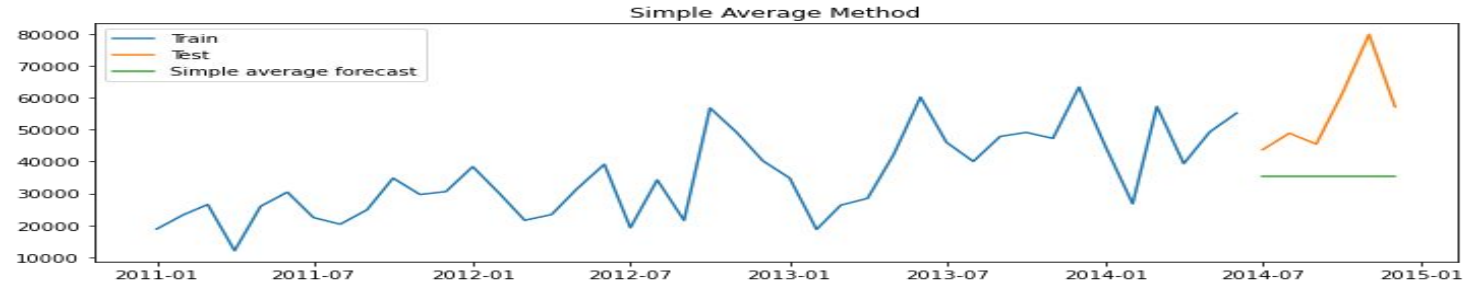


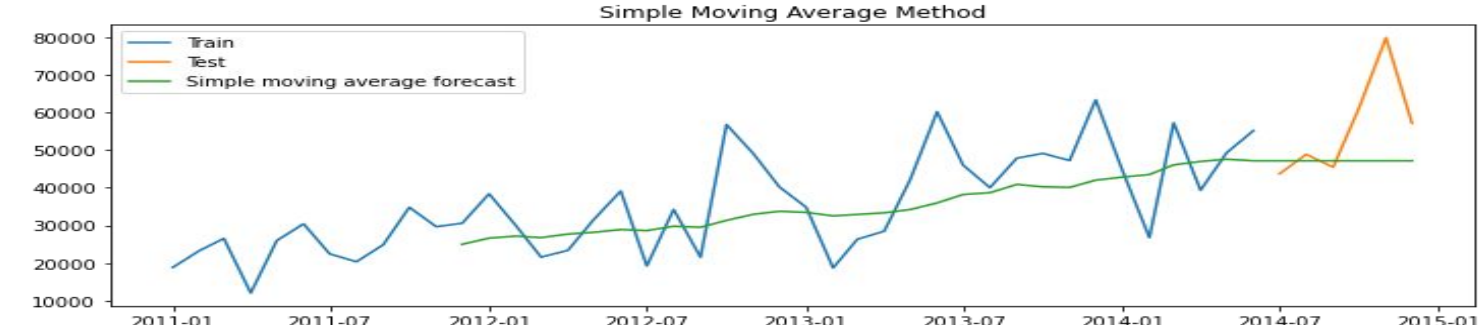
Decomposition of Data

Analysis on Decomposition

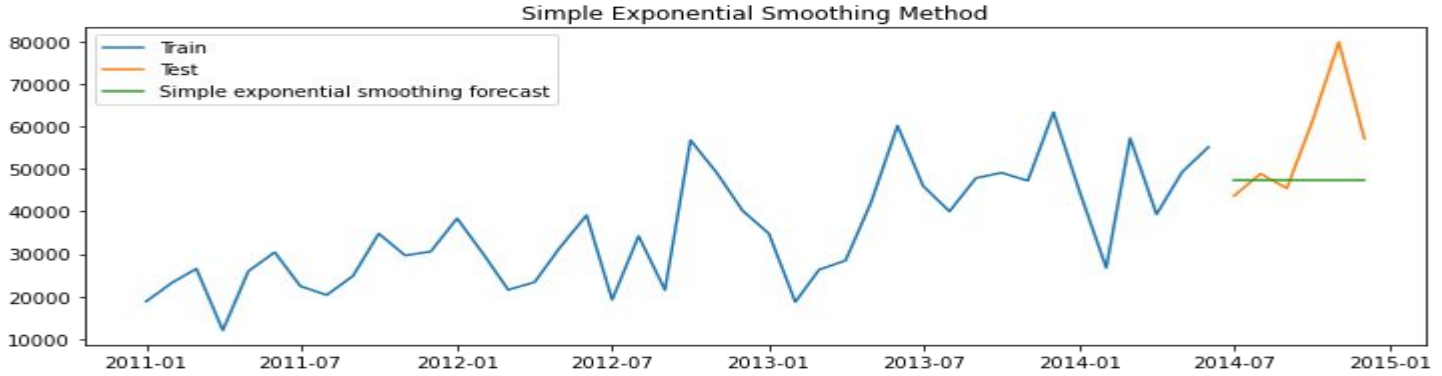
- The data has an increasing trend. The data shows seasonal behavior for every 12 months.
- Our data has more than 10 rows of data so so Naïve, Simple Average, Simple Moving Average, AR, MA, ARMA models will not be optimum techniques for forecasting.
- Simple ES is better than above methods but it won't be able to forecast the trend.
- Holt's method will be better than the above methods for predicting trend also.
- Holt Winter's method among Exponential Smoothing models and SARIMA method among Auto Regressive models will give most optimum forecast as it predicts trend and seasonality both.

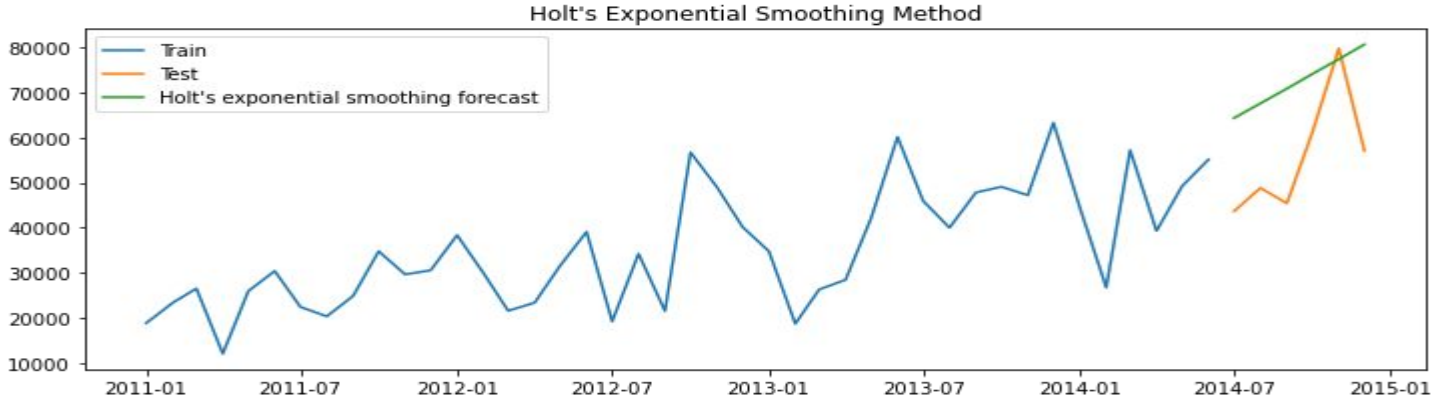




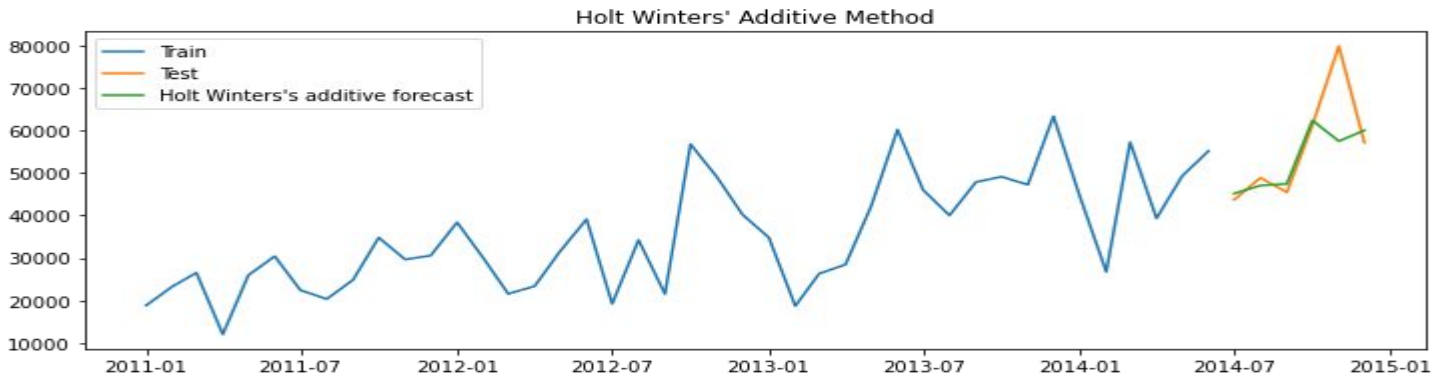


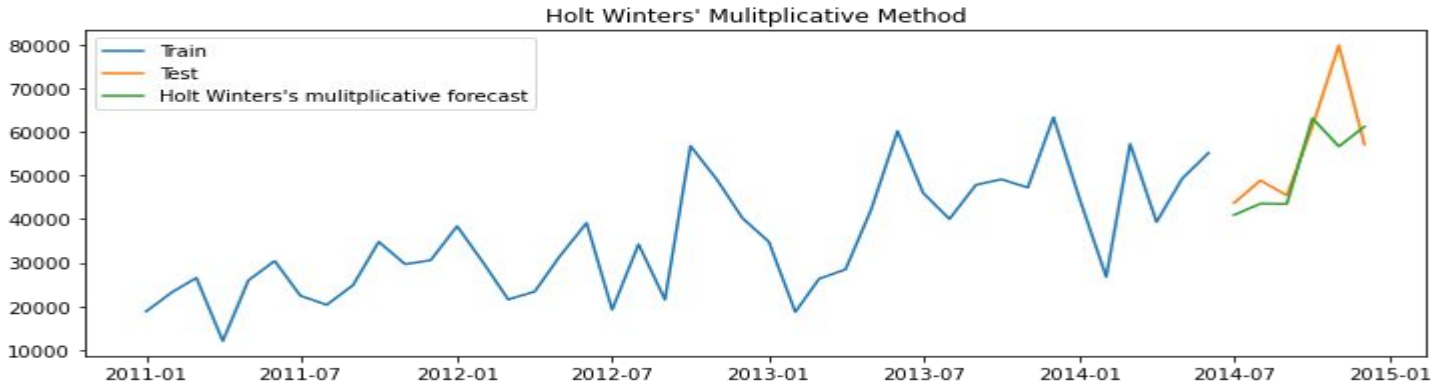
Simple time series methods





Exponential smoothing methods

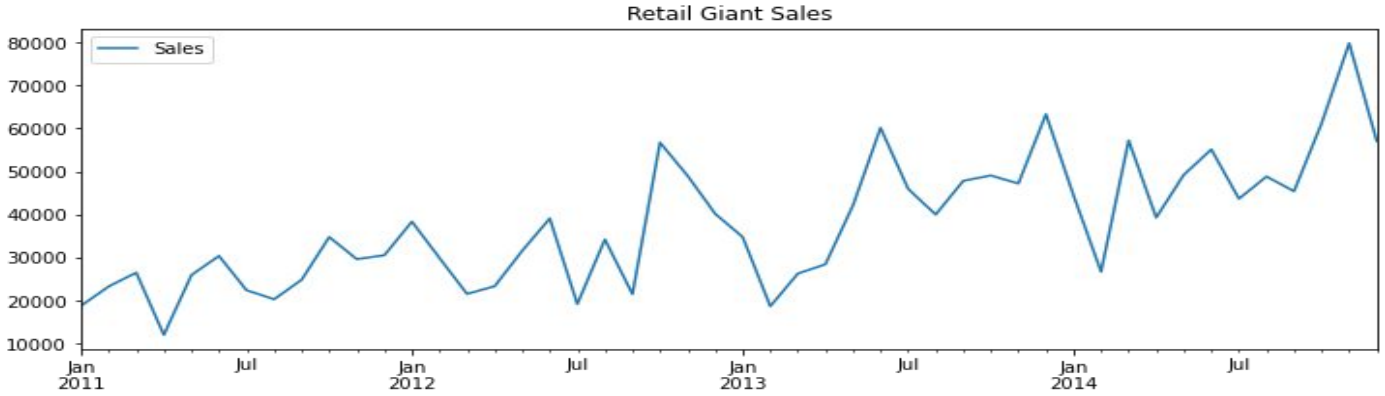




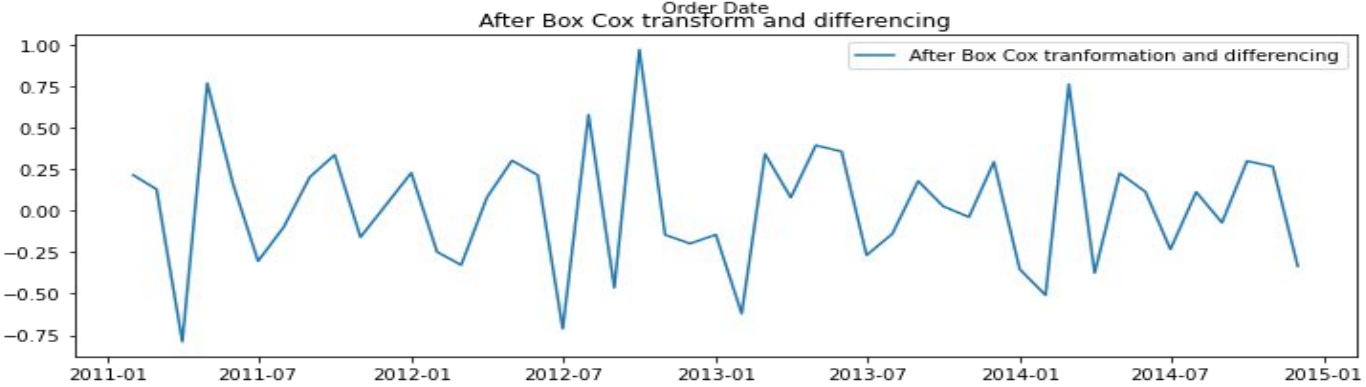
Holt Winters' additive method with trend and seasonality

Converting Non Stationary to Stationary

- The time series we have is non stationary.
- For Auto Regressive methods we convert the time series to stationary by making variance constant using box-cox transformation and removing trend by differencing.
- After we test the data with ADF and KPSS tests which shows that data is converted to stationary successfully.

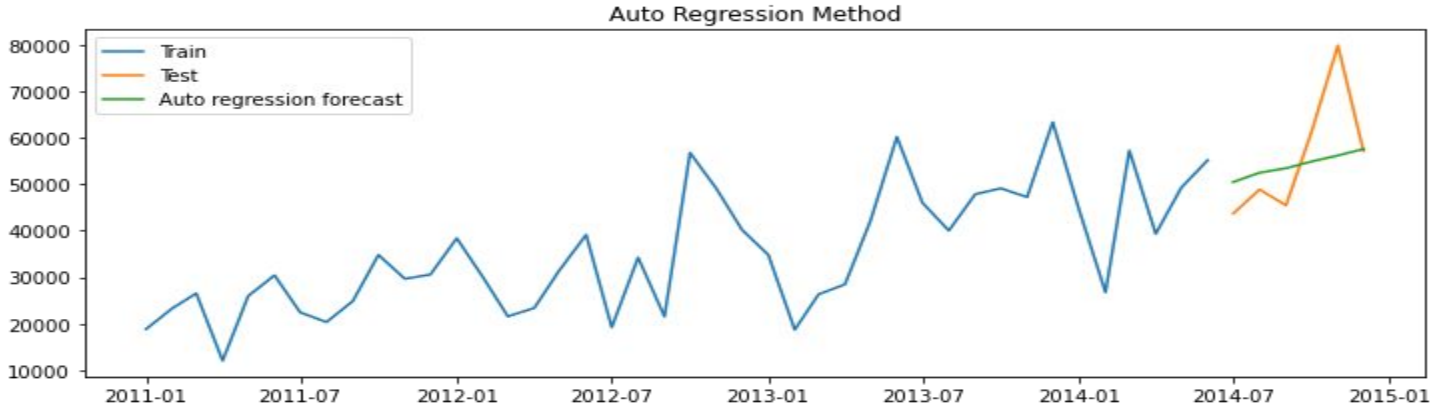


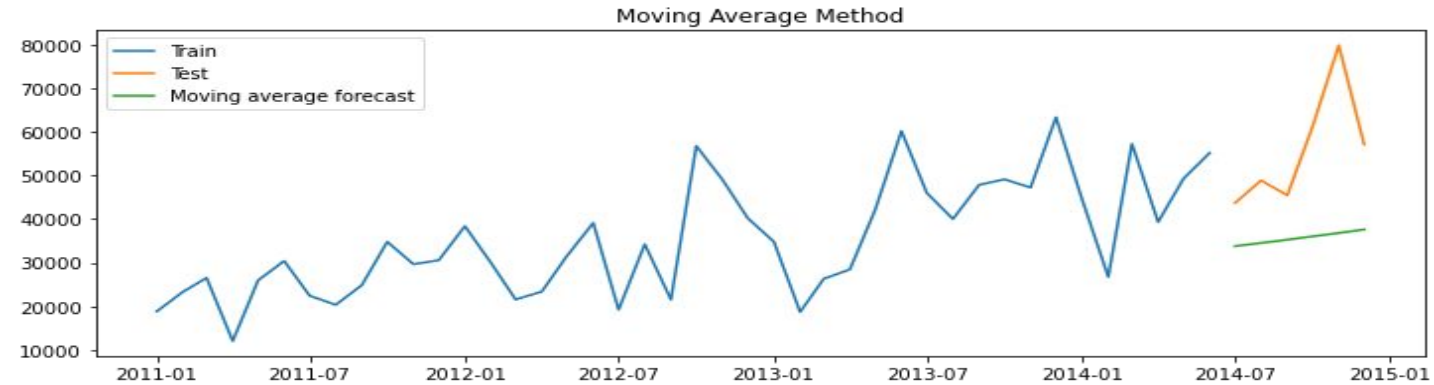
Non-Stationary Time series



Stationary Time series after Box Cox transformation and differencing

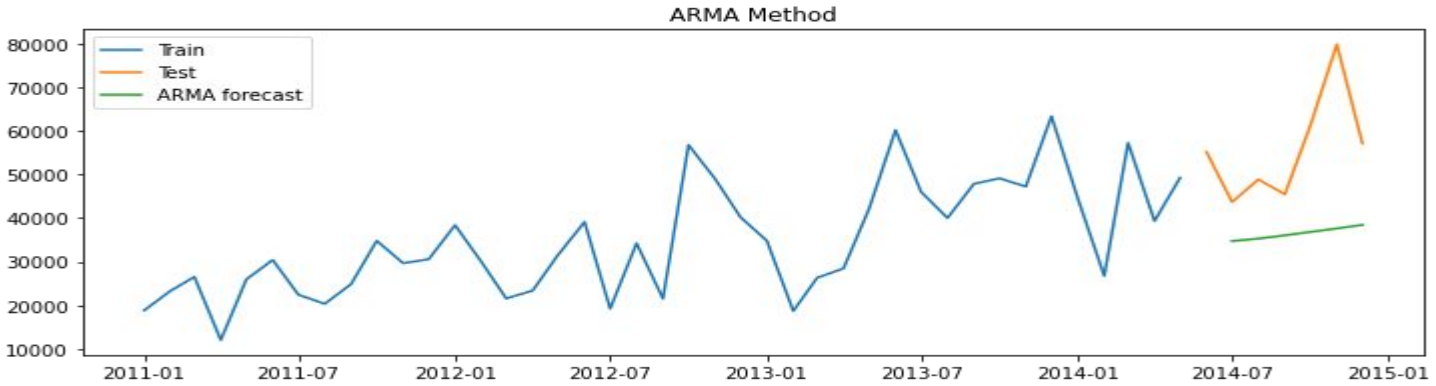
Stationary Vs Non Stationary Plots

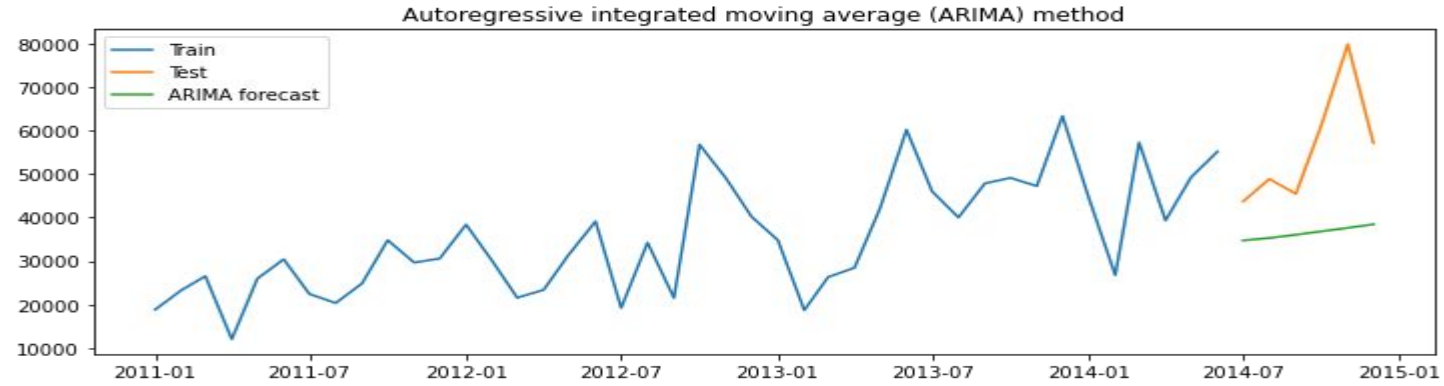




Auto Regressive methods:

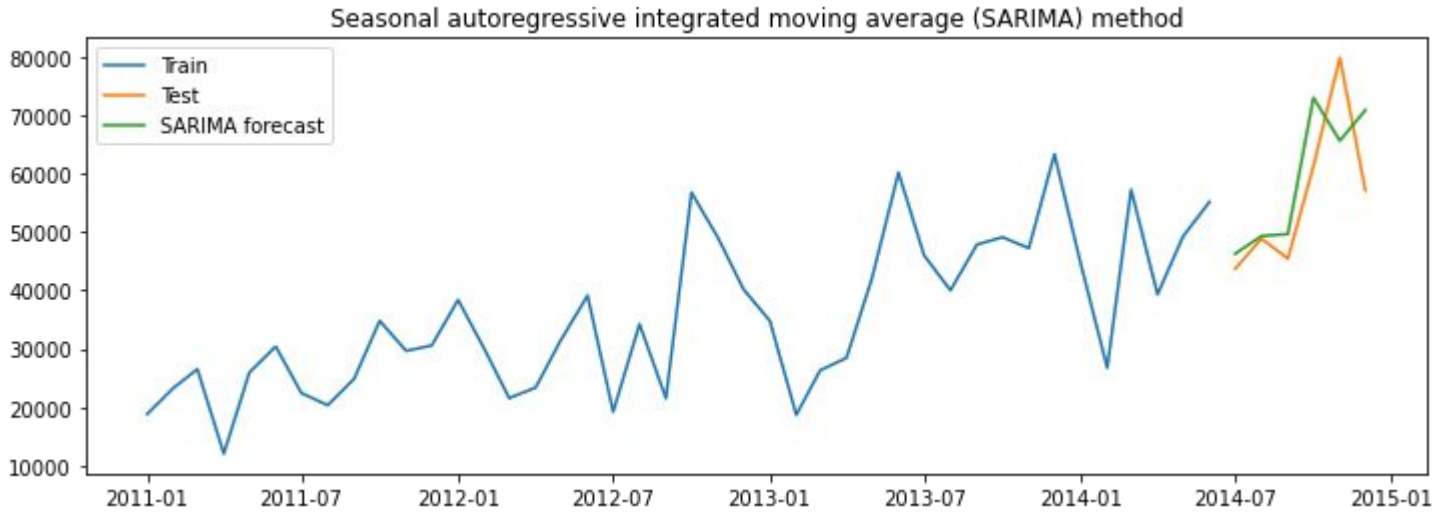
AR and MA





Auto Regressive methods:

ARMA and ARIMA



Auto Regressive methods:

SARIMA

Comparing Plots of Models

- Naïve, Simple Average, Simple Moving Average, Simple Exponential Smoothing plots does not show the trend on forecasted test values.
- Holt's plot is showing slight trend but not showing any seasonality.
- AR plot is showing better forecast than MA, ARMA, ARIMA but is not showing seasonality.
- Holt Winter's and SARIMA plots are showing trend and seasonality both on forecasted test data.

Comparing RMSE and MAPE values of different models

	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple average method	24146.06	34.34
0	Simple moving average forecast	15192.01	16.10
0	Simple exponential smoothing forecast	15011.49	15.99
0	Holt's exponential smoothing method	18976.37	34.57
0	Holt Winters' additive method	9309.63	7.73
0	Holt Winters' multiplicative method	9977.52	10.12
0	Autoregressive (AR) method	10985.28	13.56
0	Moving Average (MA) method	23360.02	33.93
0	Autoregressive moving average (ARMA) method	22654.30	32.40
0	Autoregressive integrated moving average (ARIM...	22654.30	32.40
0	Seasonal autoregressive integrated moving aver...	9617.12	12.88

Comparing RMSE and MAPE values of different models

- We see that on our data Simple Average, MA, ARMA, ARIMA methods have higher MAPE values which is not desirable.
- Naïve, SMA, SES, AR, Holt's, SARIMA have MAPE values are smaller than above mentioned models and all are in same range of (11 – 17.5).
- Holt – Winter's Additive method is having lowest RMSE and MAPE values fro smoothing technique and for Auto regressive methods SARIMA method performs the best.

Conclusion

- Comparing Plots, RMSE and MAPE values we can conclude that Holt-Winter's additive method is the optimum technique which works the best Among the smoothing methods for forecasting of sales.
- Among the auto regressive methods SARIMA method works best for the forecasting of sales.
- As These model forecasts trend and seasonality both as seen in plot and have the lowest RMSE and MAPE value.

Thank you