RETAIL GIANT SALES FORECASTING

NAVADEEP R JAGATHKARI

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 — consumer, corporate and home office.

As a sales manager for this store, we have to forecast the sales of the products for the next 6 months, so that you have a proper estimate and can plan your inventory and business processes accordingly.

If we check the entries in the dataset, we will see that the store caters to 7 different geographical market segments and 3 major customer segments. Based on these, there are **21 unique "Market-Segments"** for which the sales forecasts can be made.

So, not all of these 21 market segments are important from the store's point of view. You need to find out the most consistently profitable market-segment from the above and forecast the sales and demand for that single market-segment only.

Problem Approach

As an analyst, our job is to analyze the data available and find out the most consistent market_segment and to forecast the time-series.

Since this is a time series forecasting problem. There are various forecasting methods are available, here we are checking all the methods and conclude the best forecasting for the dataset.

Steps to Approach:

1. Read, clean and understand the dataset.
2. Converting the Order Data column to month_year format.
3. Split the dataset, Calculate CoV.
4. Create new dataframe extracting Sales of only most consistent market_segment data.
5. Model Building and Evaluation.
   - Checking all the forecasting methods.
6. Conclusion.

EDA

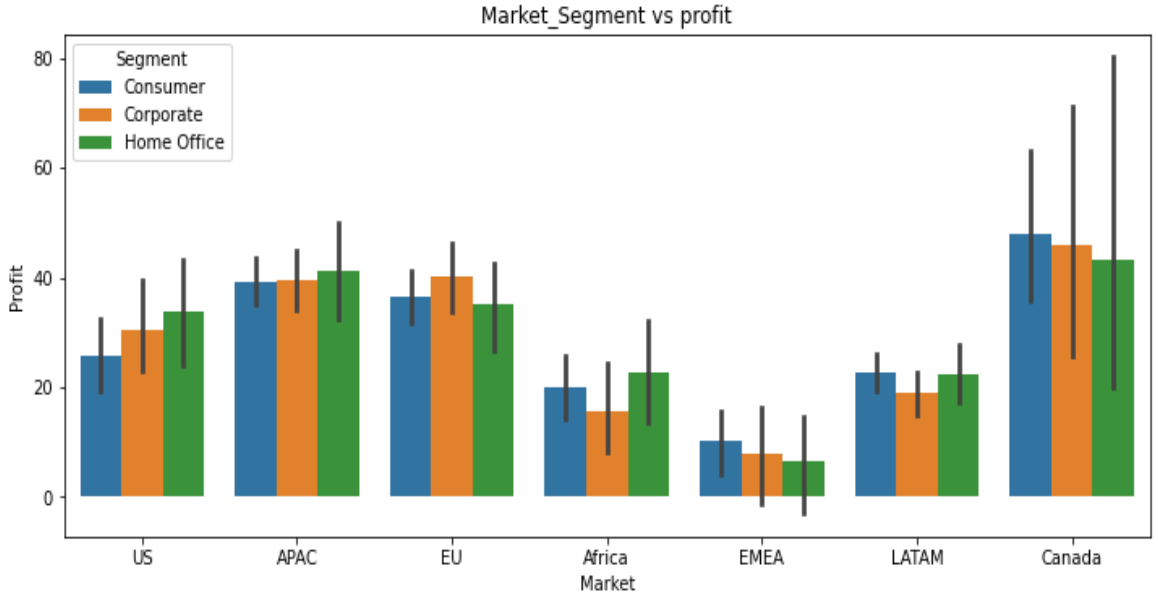
- Assign the Market and Segment columns to new column ‘Market_Segment’
- After combining we are getting 21 unique Market_Segments.
- Visualizing Market_Segment vs Sales and Market_Segment vs Profit.
- Change Order Date to Month-Year format.
- Derive monthly aggregated transaction data using pivot table.
- Train-test split
- CoV

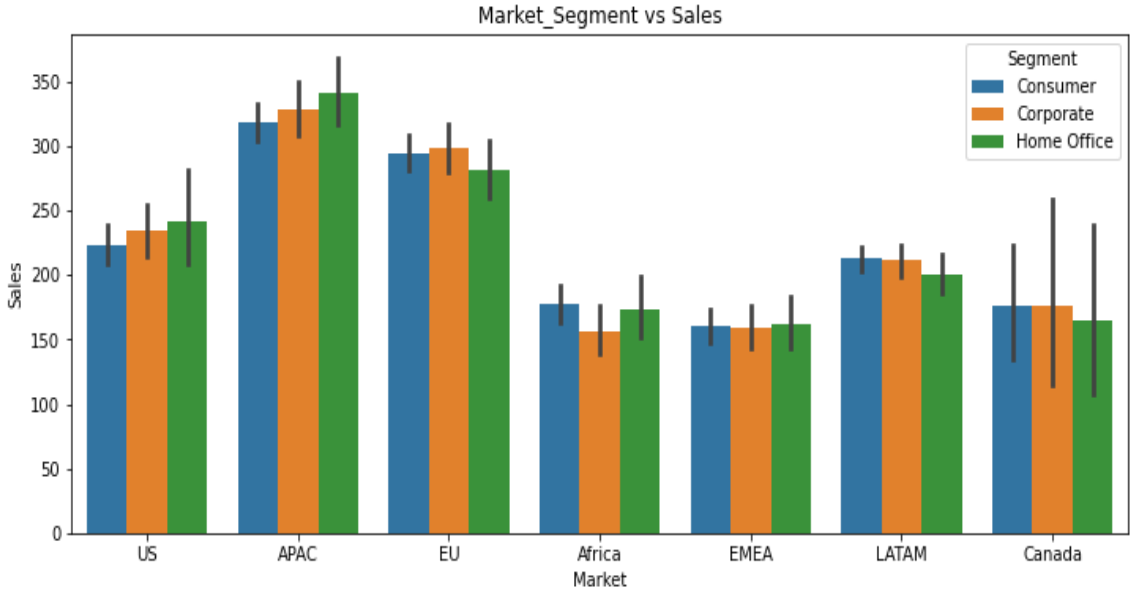
21 Unique Market_Segments

APAC_Consumer	5699
LATAM_Consumer	5321
US_Consumer	5191
EU_Consumer	5186
APAC_Corporate	3283
EU_Corporate	3077
LATAM_Corporate	3053
US_Corporate	3020
EMEA_Consumer	2538
Africa_Consumer	2381
APAC_Home Office	2020
LATAM_Home Office	1920
US_Home Office	1783
EU_Home Office	1737
EMEA_Corporate	1574
Africa_Corporate	1312
EMEA_Home Office	917
Africa_Home Office	894
Canada_Consumer	202
Canada_Corporate	110
Canada_Home Office	72

- The list of all the Market_Segments that we obtained.
- *APAC_Cosumer* is the most consistent.
- *Canada_Home* Office is least consistent.

Market_Segment vs Profit & Sales



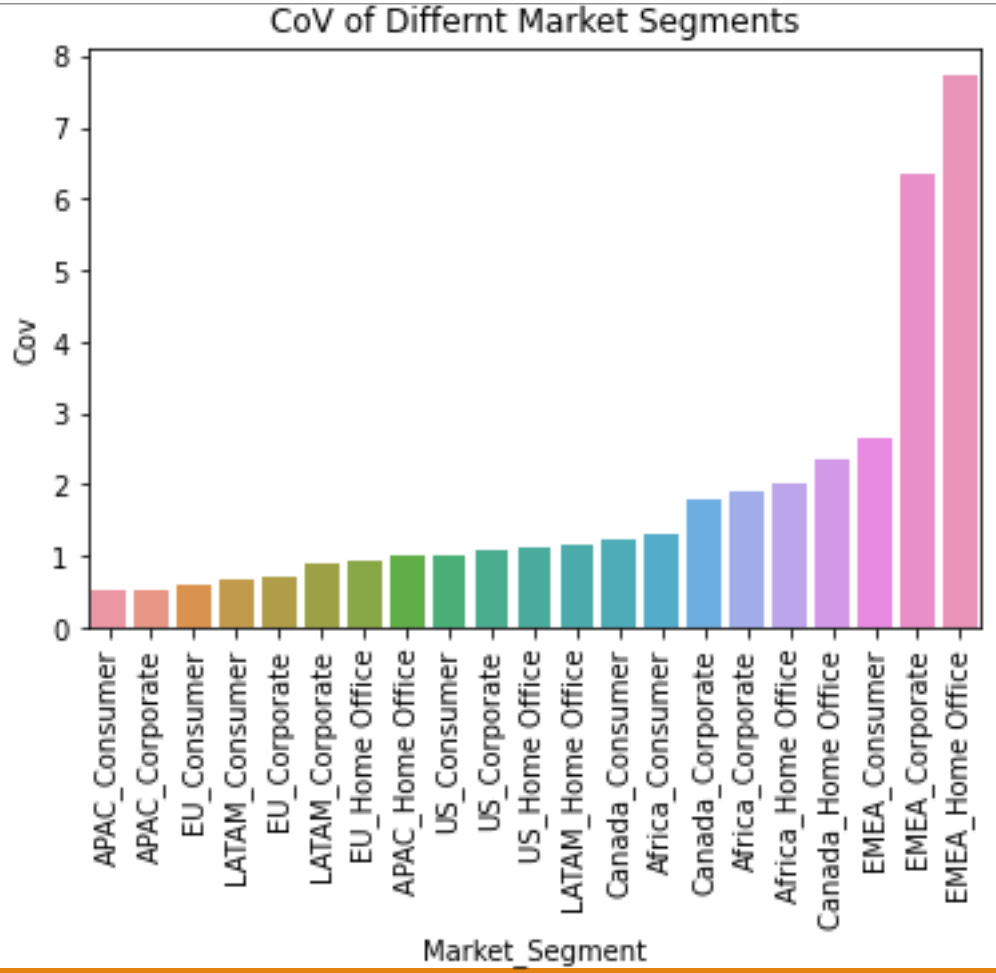


- In Market_segment vs profit, Canada has more profit followed by APAC

- APAC market_segment is having most sales.

Values of CoV of Market_Segments

	Market_Segment	Mean	Std	Cov
0	APAC_Consumer	4400.894243	2300.457687	0.522725
1	APAC_Corporate	2574.919807	1364.837734	0.530051
12	EU_Consumer	3699.977143	2202.282289	0.595215
15	LATAM_Consumer	2295.555697	1569.632686	0.683770
13	EU_Corporate	2216.299429	1600.336696	0.722076
16	LATAM_Corporate	1122.633016	990.360880	0.882177
14	EU_Home Office	1224.456536	1148.627937	0.938072
2	APAC_Home Office	1511.088314	1523.508658	1.008219
18	US_Consumer	2686.740912	2715.031412	1.010530
19	US_Corporate	1754.199083	1880.200775	1.071829
20	US_Home Office	1132.065762	1272.476439	1.124030
17	LATAM_Home Office	818.398941	957.275713	1.169693
6	Canada_Consumer	225.987632	282.555788	1.250315
3	Africa_Consumer	957.707000	1254.932072	1.310351
7	Canada_Corporate	90.980294	162.493114	1.786025
4	Africa_Corporate	412.617571	780.566850	1.891744
5	Africa_Home Office	377.221071	759.322203	2.012937
8	Canada_Home Office	118.003750	279.632866	2.369695
9	EMEA_Consumer	423.960286	1124.552711	2.652495
10	EMEA_Corporate	182.642643	1160.698430	6.355024
11	EMEA_Home Office	84.231366	651.283095	7.732073

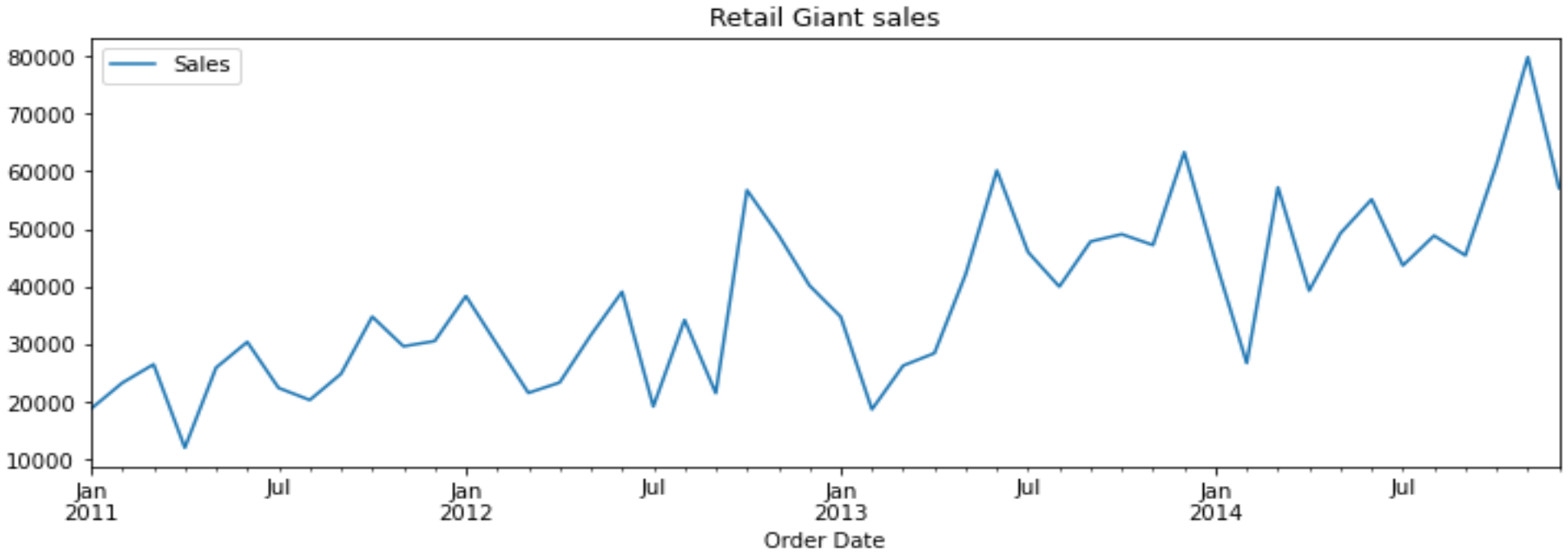


APAC _consumer Market_Segment

The reason APAC _consumer is more consistent and profitable market_segments:

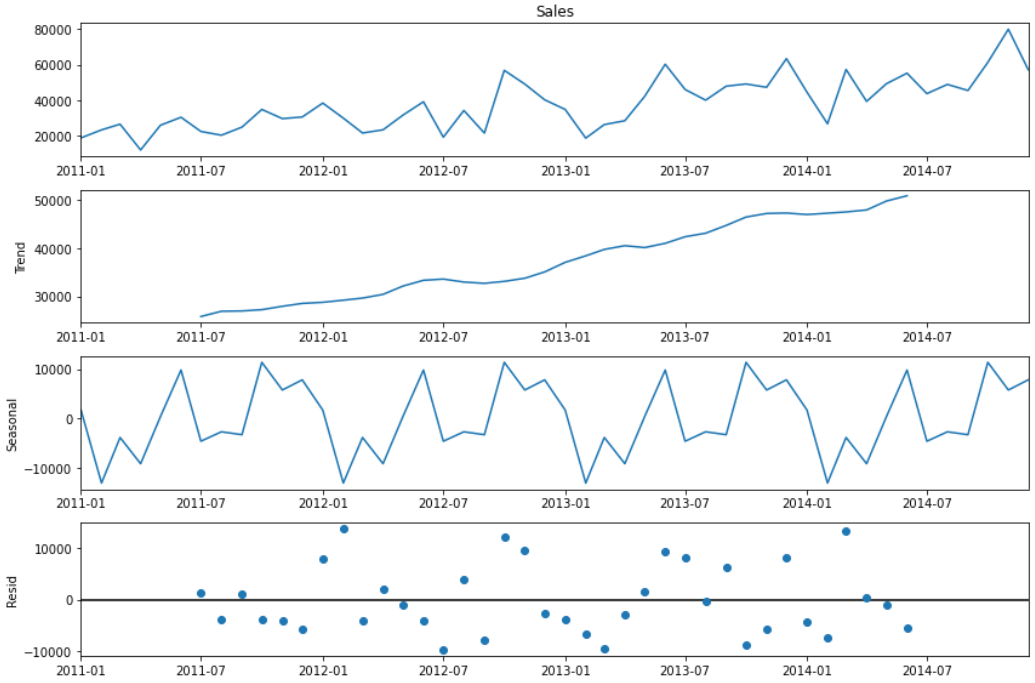
- In the above slides we saw the sorted list of all unique market_segments.
- Then compared the graphs for market_segments vs profit and market_segments vs sales, the APAC_consumer has out performed in both the possibilities.
- CoV(Coefficient of Variance) = std/mean.
- Lesser the CoV better the Profit of the market_segment, APAC_Consumer has least CoV.
- These are the reasons APAC_consumer is our ideal market_segment for further forecasting analysis.

Time Series Analysis – Sales of APAC_Consumer

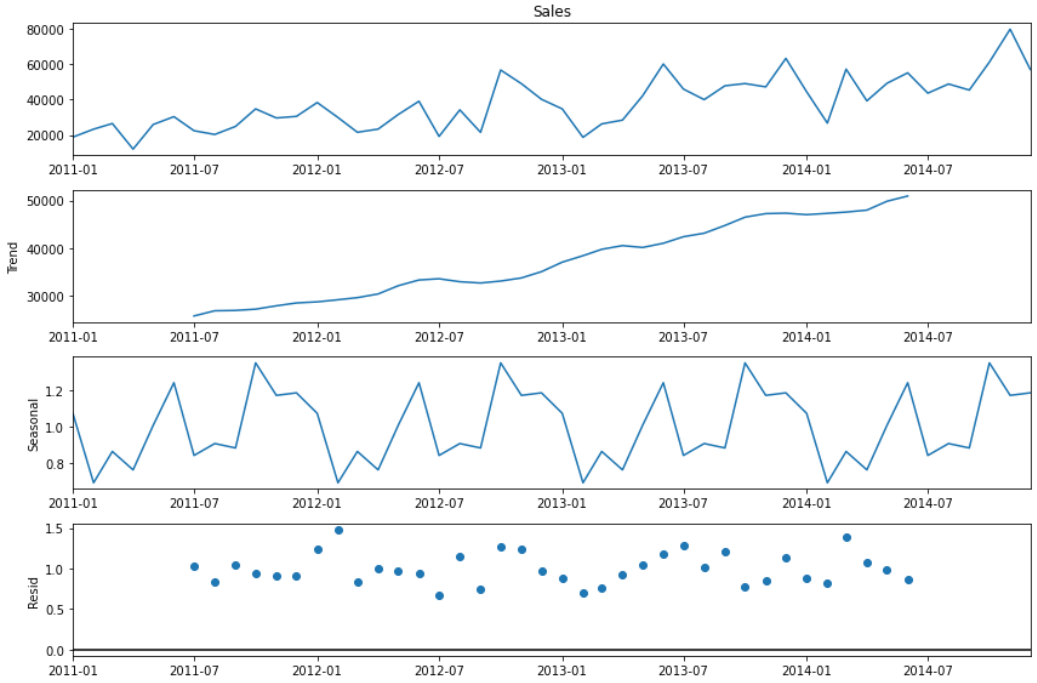


- This graph is basic plot of data which we got.
- There is a increasing trend, a small kind of seasonality after the months of January, which show a trough.

Time series Decomposition



Decomposition_Additive

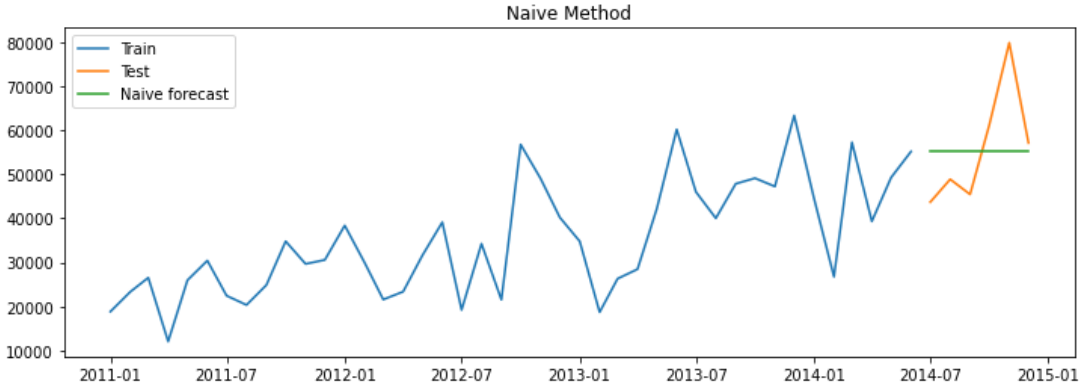


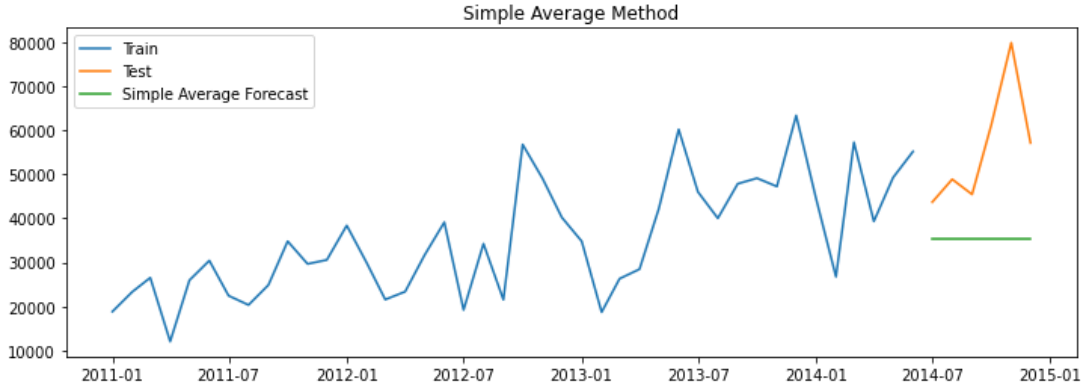
Decomposition_Multiplicative

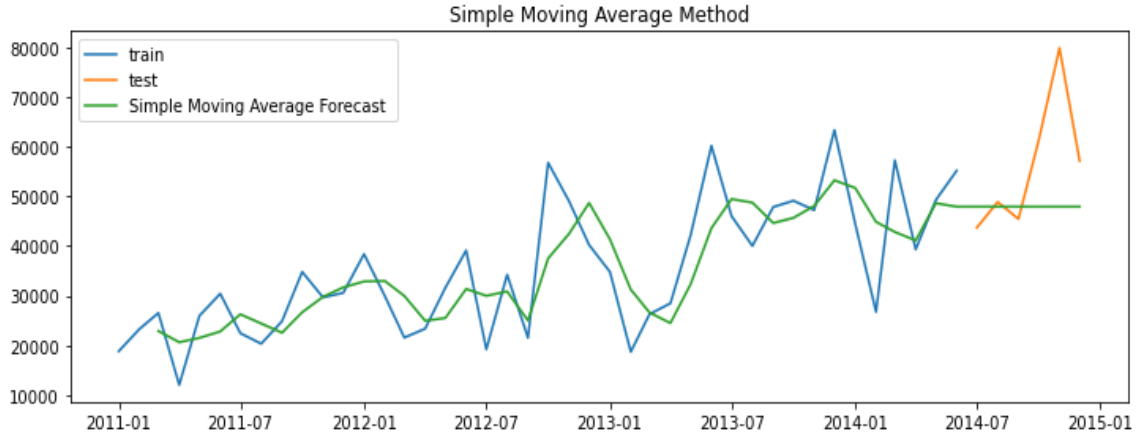
Steps of Model Building and Evaluating

- Forecasting Methods
  - Simple exponential smoothing
  - Holt's exponential smoothing
  - Holt-Winters' exponential smoothing - Additive
  - Holt-Winters' exponential smoothing – Multiplicative
  - AR model
  - MA model
  - ARMA model
  - ARIMA model
  - SARIMA mode
- Evaluating Method.
- Plotting the method to understand the forecasting obtained.
- Calculating the RMSE and MAPE.

Basic Forecasting Models



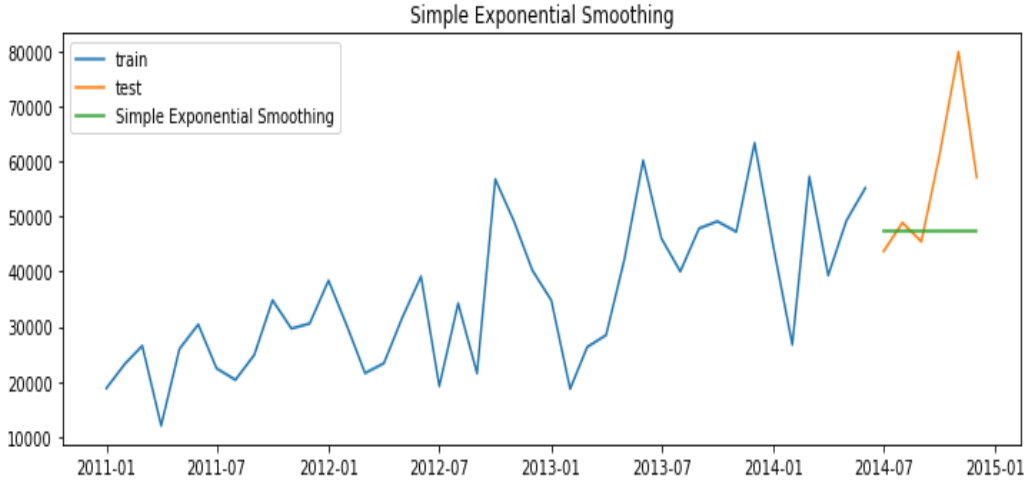


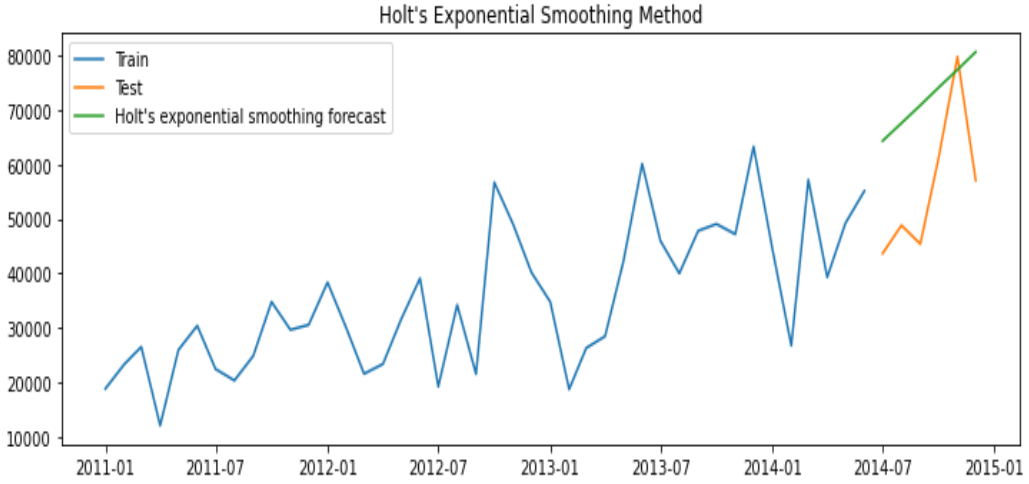


	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple Average Method	24146.06	34.34
0	Simple Moving Average Method	14756.73	15.82

- Here SMA method has least MAPE, which can be best forecast among the three.

Smoothing Models

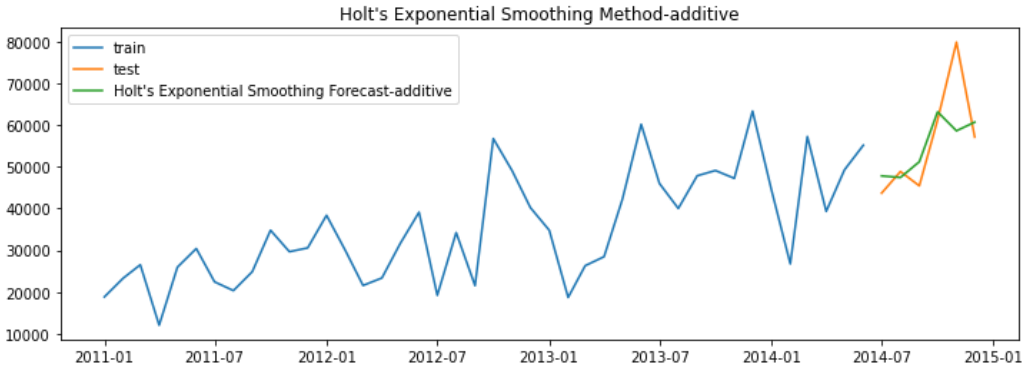


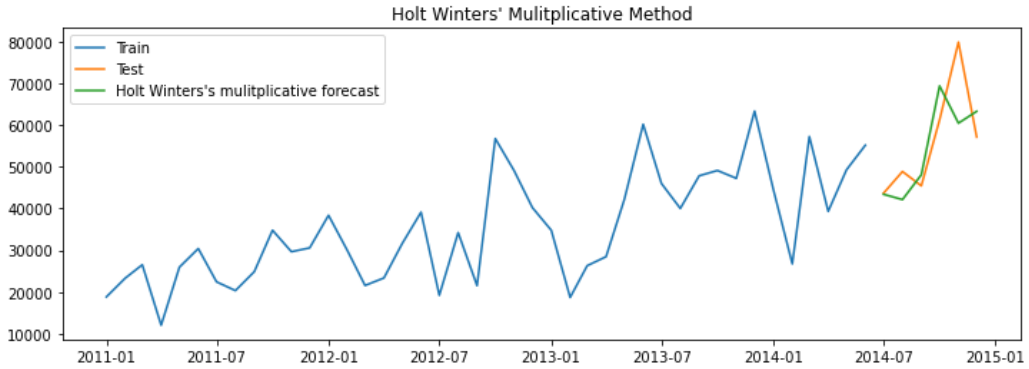


	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple Average Method	24146.06	34.34
0	Simple Moving Average Method	14756.73	15.82
0	Simple Exponential Smoothing	15011.49	15.99
0	Holt's Exponentail Smoothing	18976.37	34.57

- Holt’s model here has forecasted better trend.
- As of now, Simple Exponential Smoothing is having least MAPE value.

Holt's Winter Exponential Smoothing

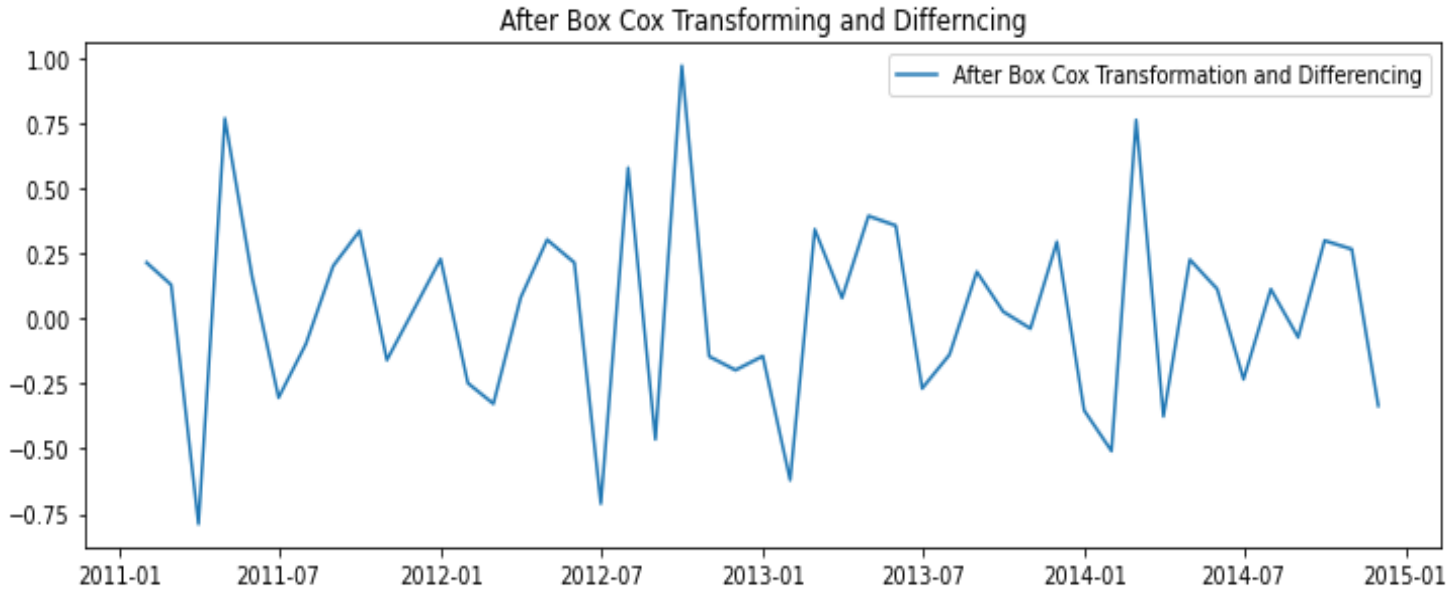




	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple Average Method	24146.06	34.34
0	Simple Moving Average Method	14756.73	15.82
0	Simple Exponential Smoothing	15011.49	15.99
0	Holt's Exponentail Smoothing	18976.37	34.57
0	Holt Winters' additive method	9306.82	10.17
0	Holt Winters' multiplicative method	9423.23	11.43

- Holt's winter model has forecasted trend and seasonality both.
- Holt's Winter Additive Method is the best forecasting model.

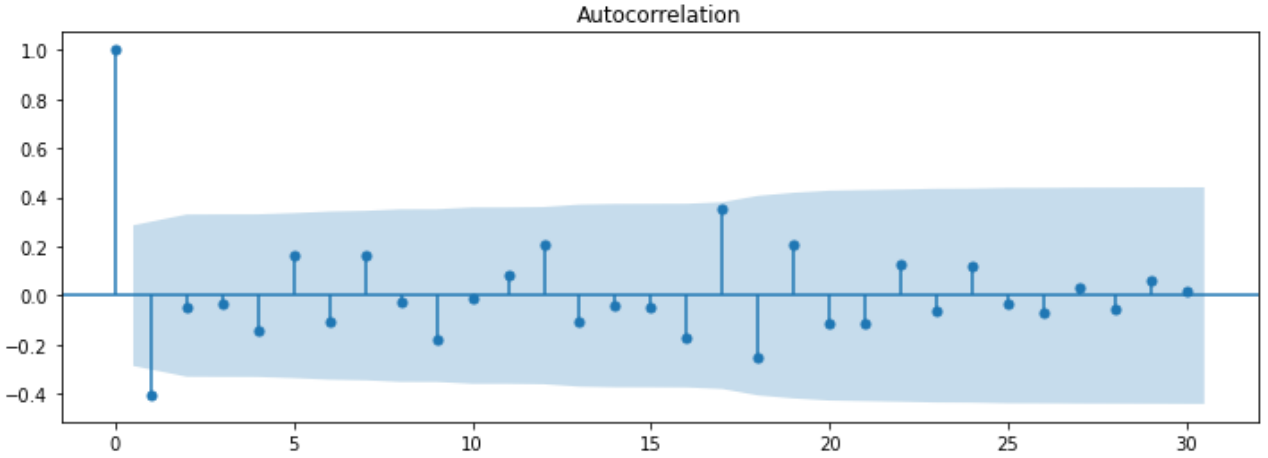
After Boxcox Transformation and Differencing



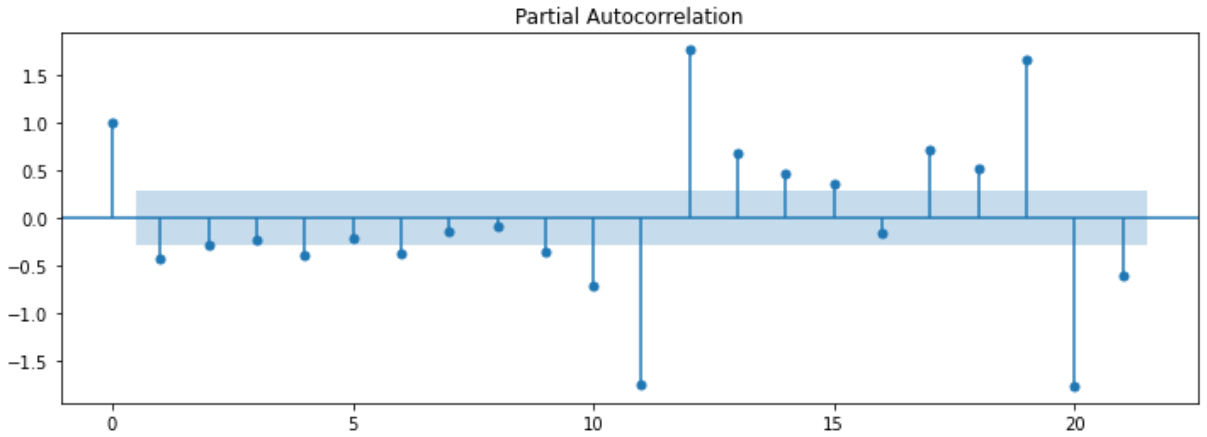
- This plot is gives Stationarity for the series after ADF and KPSS tests.
- Here the mean of the series is also taken around to zero.
- This is used to forecast Auto Regressive models.

Auto Correlation Functions

ACF



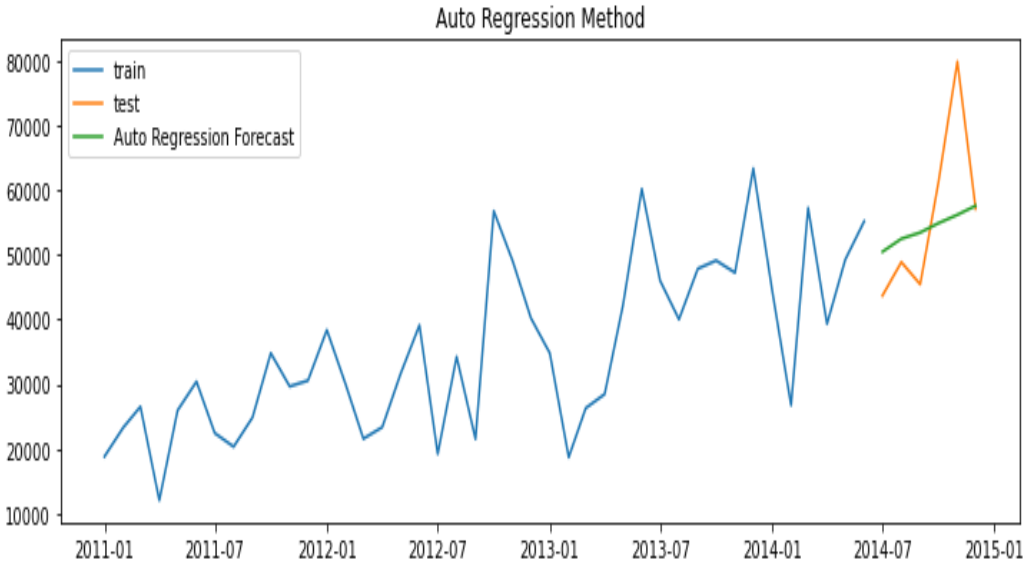
PACF

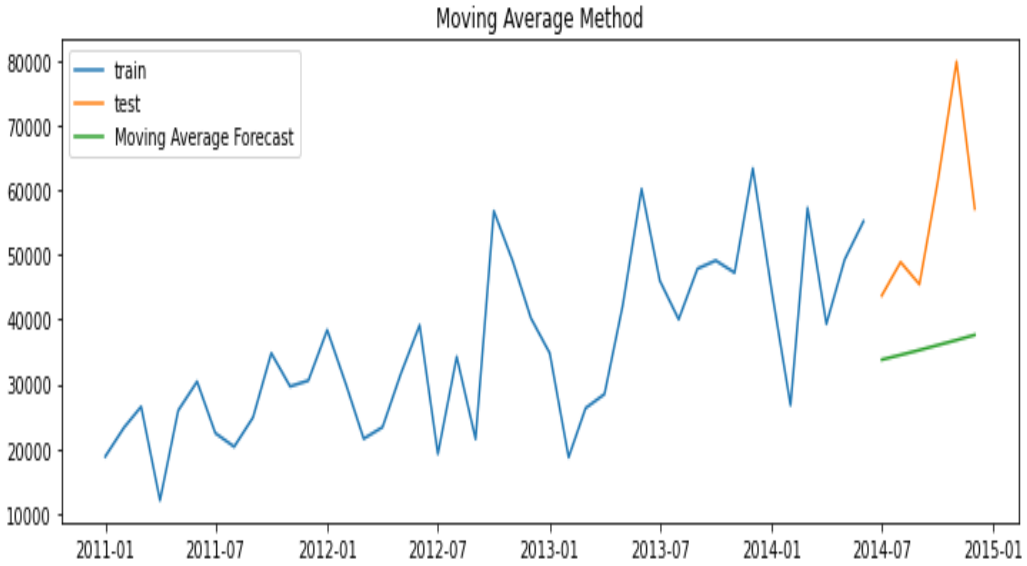


•Autocorrelation helps us to know how a variable is influenced by its own lagged values.

•The autocorrelation function tells about the correlation between an observation with its lagged values. It helps you to determine which lag of the observation is influencing it the most.

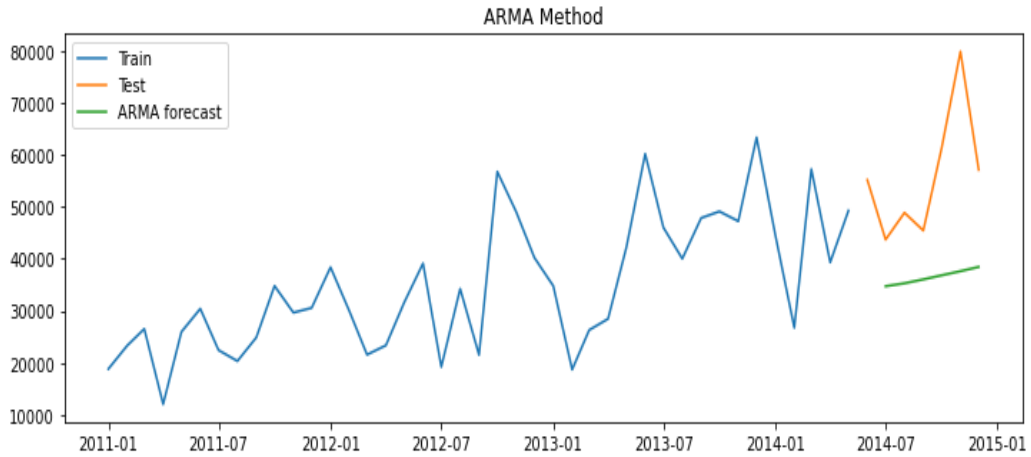
Auto Regressive and Moving Average

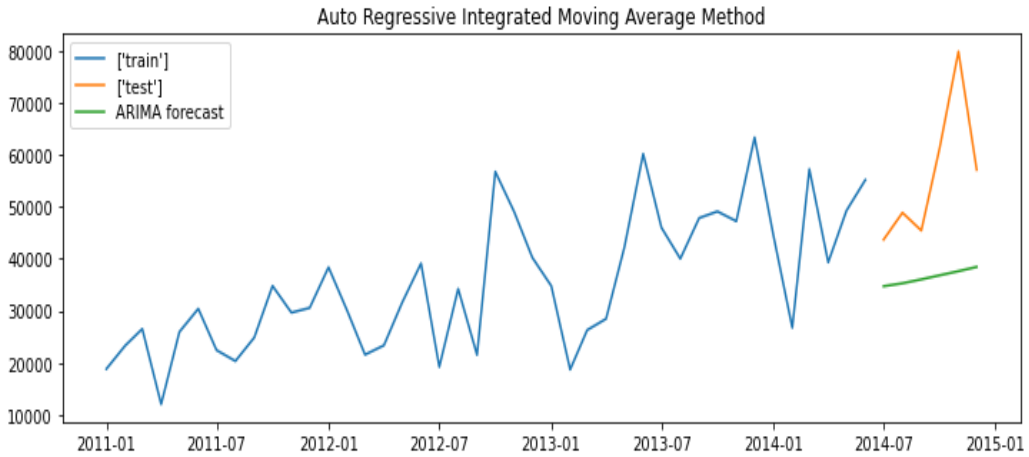




- AR model is better than MA model
- The Green line In MA model has came down than the mean of the forecast.

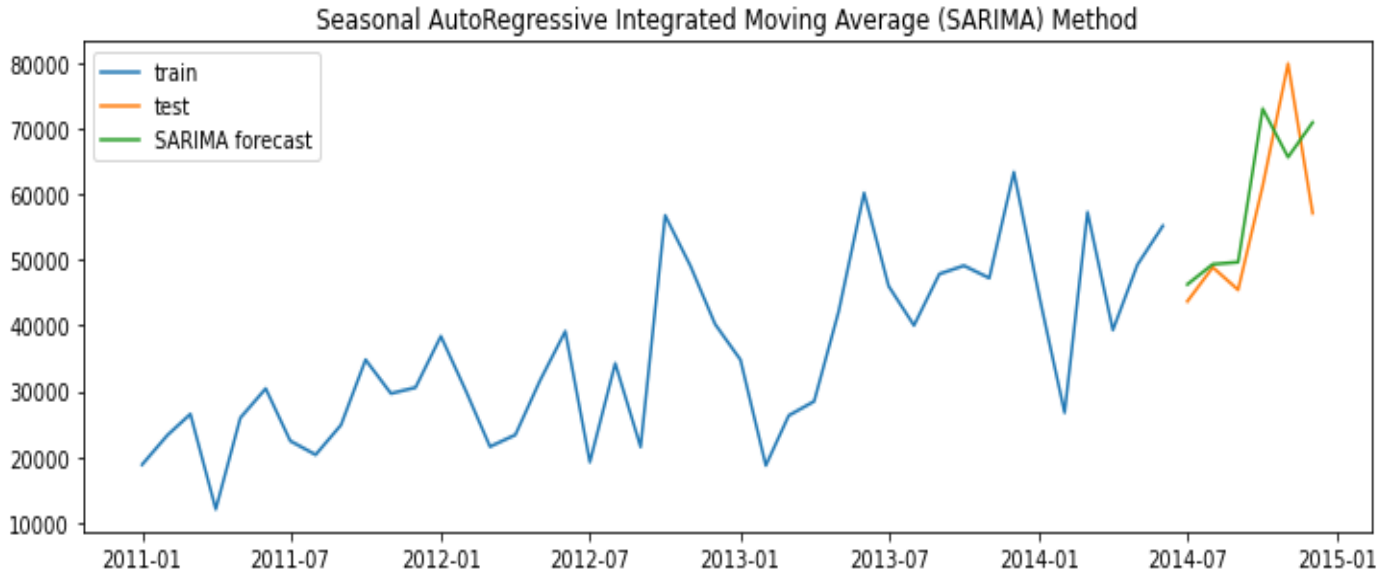
ARMA and ARIMA





- ARMA and ARIMA is giving almost same trend and level.
- The mean, RMSE and MAPE values are same.

SARIMA(Seasonal Auto Regressive Integrated Moving Average)



- The SARIMAX is used to forecast not just level and trend but also seasonality.
- The RMSE and MAPE of ARIMA and SARIMA is same except the seasonality here.

RMSE and MAPE

	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple Average Method	24146.06	34.34
0	Simple Moving Average Method	14756.73	15.82
0	Simple Exponential Smoothing	15011.49	15.99
0	Holt's Exponentail Smoothing	18976.37	34.57
0	Holt Winters' additive method	9306.82	10.17
0	Holt Winters' multiplicative method	9423.23	11.43
0	Autoregressive (AR) method	10985.28	13.56
0	Moving Average(MA) Method	23360.02	33.93
0	Auto Regressive Moving Average (ARMA) Method	22654.32	32.40
0	Autoregressive integrated moving average (ARIM...	22654.32	32.40
0	Seasonal autoregressive integrated moving aver...	9617.98	12.88

In all of the Auto Regressive Models SARIMA has performed well.

- Overall, the Holt's winter additive model is performing well of all the forecasting models.
- Followed by Holt's winter Multiplicatie model.

Conclusion

- ✓The best forecast of the model is concluded based on level trend and seasonality.
- ✓The Holt's Winter method and SARIMA method give seasonality constraint which is one of the most important element while forecasting. We got the best results for this dataset based on the same methods.
- ✓Overall the MAPE value of Holt's winter additive method got the least of all.
- ✓We conclude that, *"Holt Winters Additive" method* is the best forecasting method in the smoothing techniques
- ✓And *SARIMA* is the best method in ARIMA set of techniques.