Retail Giant Sales Forecasting Assignment

by Pratik Chakraborty

Business Problem

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, you 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.

The store dataset has the following 5 attributes and their data description is as given below:

| Attributes	Description
Order-Date	The date on which the order was placed
Segment	The segment to which the product belongs
Market	The market to which the customer belongs
Sales	Total sales value of the transaction
Profit	Profit made on the transaction

Objective: Forecast the sales of the products for the next 6 months

Approach

- The initial data preparation and use of EDA finding out the most profitable and consistent market segments.
- Time Series modelling using classical decomposition and ARIMA to forecast.
- The optimum technique from the flow chart that might work best for the sales forecast.
- Model Evaluation finding the MAPE value.

21 unique market segments

- The store caters to 7 different geographical market segments and 3 major customer segments.

- Market_Segment column created which contains 21 unique "Market-Segments" for which the sales forecasts can be made.APAC_Consumer has the highest count as per plot.

21 Unique

The market to which the customer belongs

The segment to which the product belongs
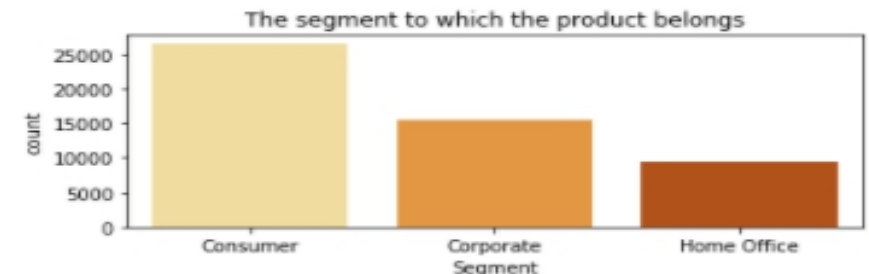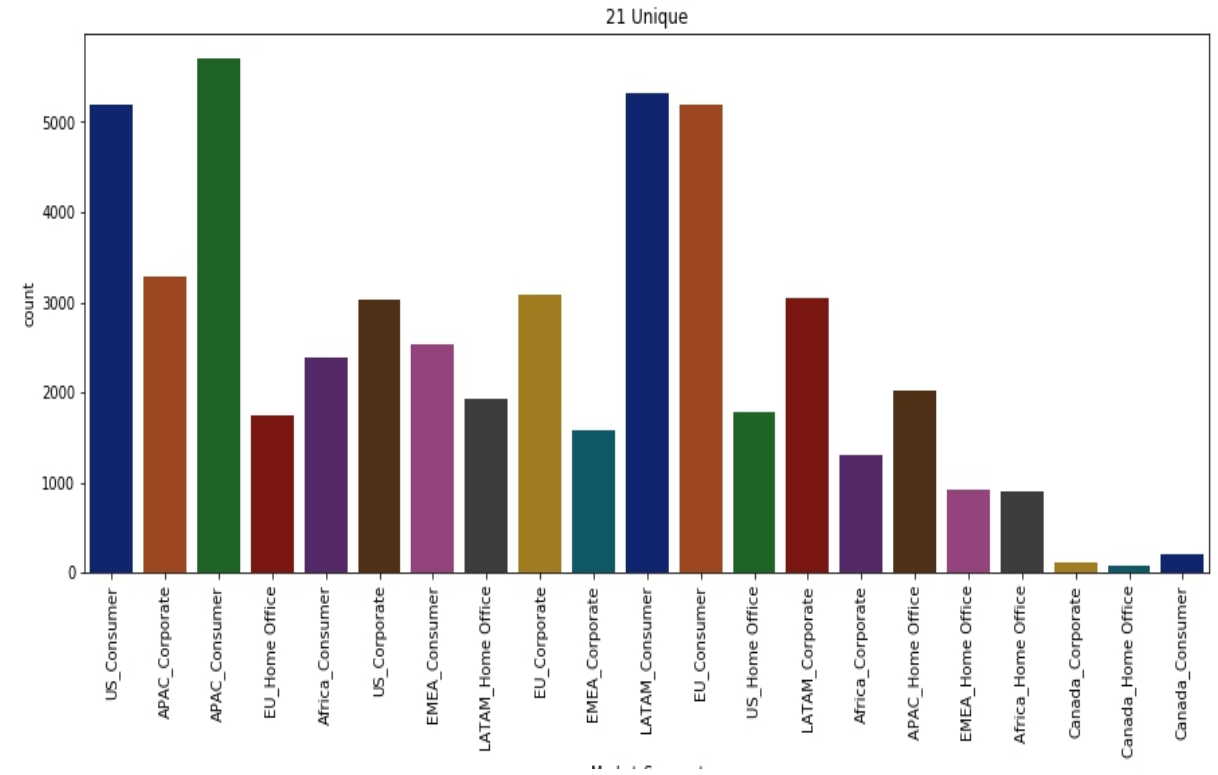
CoV on the profit

- Coefficient of variation is a ratio of the standard deviation to mean.

- Checking the standard deviation and the mean calculated on profit for all the 21 market segments and compare.

- The most profitable market segment by comparing the CoV values is **APAC_Consumer**

- The reason why a market segment "**APAC_Consumer**" is the most profitable market segment as cov value is **0.522725** least coefficient of variation

	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

```
1  # The least  variation in the profits
2  least_COV =min(CoV["CoV"])
3  least_COV
```

0.5227250555218098

Time series

Retail Giant Sales
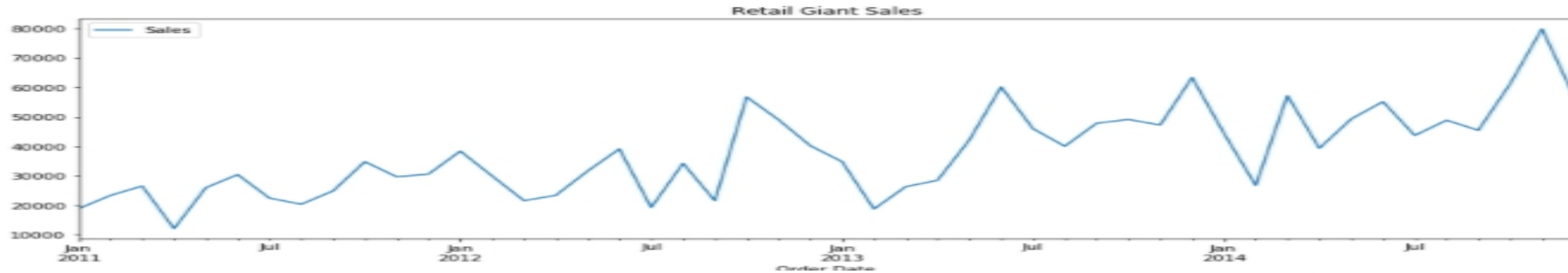
Time series Decomposition

- Trend, Seasonality, and residuals are the various components time series can be splited

- Individual components can be added to get the time-series data

- Individual components can be multiplied to get the time-series data

Building and evaluating time series forecast

- Naive method
- Simple average method
- Simple moving average forecast
- Simple exponential smoothing forecast
- Holt's exponential smoothing method
- Holt Winters' additive method
- Holt Winters' multiplicative method
- Autoregressive (AR) method
- Moving Average (MA) method
- Autoregressive moving average (ARMA)
- Autoregressive integrated moving average (ARIMA)
- Seasonal autoregressive integrated moving average (SARIMA)

	Method	RMSE	MAPE
0	Naive method	12355.97	17.47
0	Simple average method	24146.06	34.34
0	Simple moving average forecast	14756.73	15.82
0	Simple exponential smoothing forecast	14765.75	15.83
0	Holt's exponential smoothing method	18976.37	34.57
0	Holt Winters' additive method	8942.57	8.84
0	Holt Winters' multiplicative method	9976.32	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	22654.32	32.40
0	Autoregressive integrated moving average (ARIM...	22654.32	32.40
0	Seasonal autoregressive integrated moving aver...	9617.20	12.88

- **Naive method:** Estimating technique in which the last period's actuals are used as this period's forecast, without adjusting them or attempting to establish causal factors.
The forecast for the next six months is the same value as the last observation of the blue line.

- **Simple average method:** The average unit cost is calculated by multiplying the total of these unit costs simply by the number of receiving. Green line forecasted has no trend or seasonality while train and test data had both trend and seasonality.

- **Simple moving average forecast:** An arithmetic moving average calculated by adding recent prices and then dividing that figure by the number of time periods in the calculation average.
Helps to reduce unsystematic noise in the data.
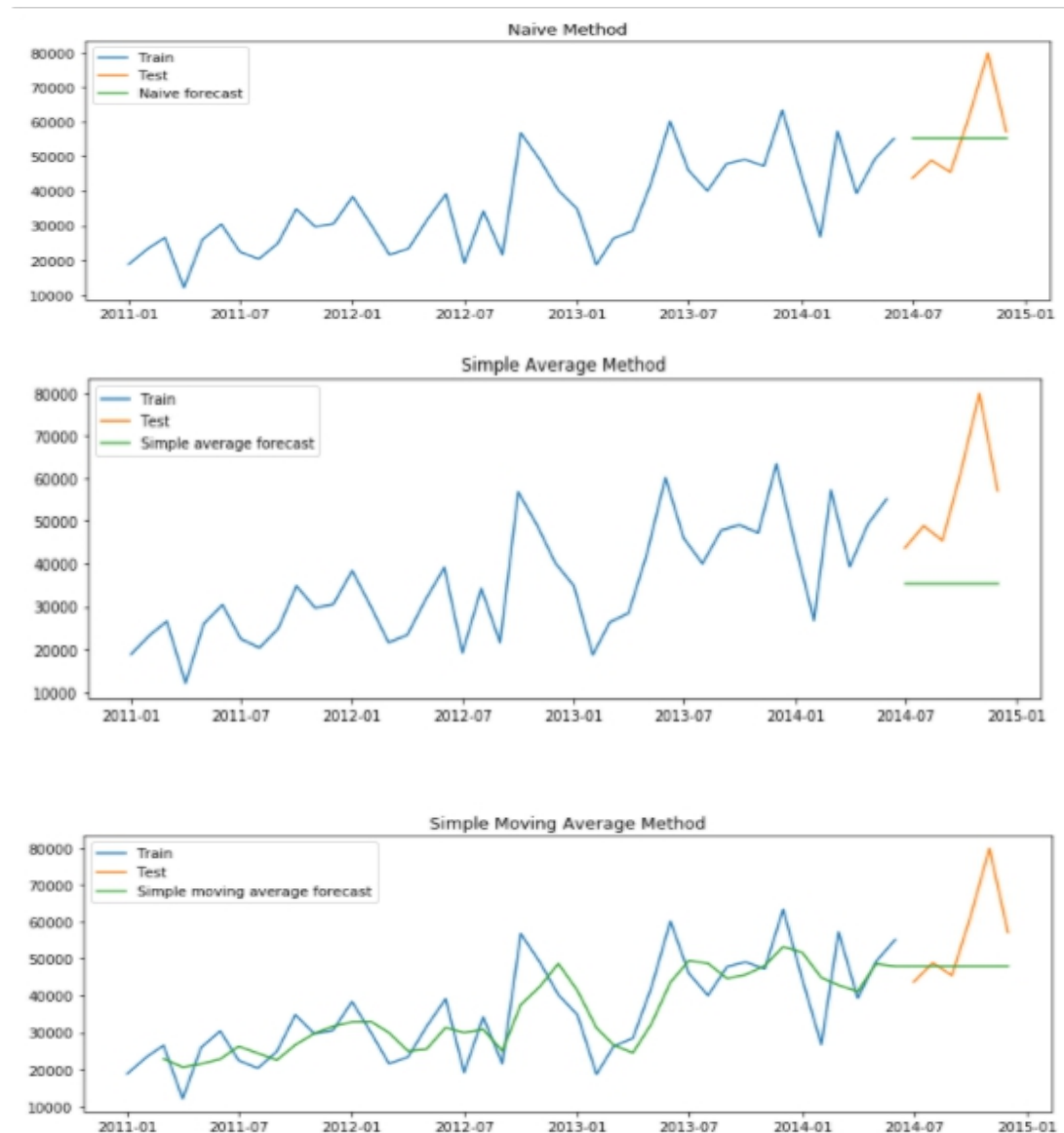
- **Simple exponential smoothing:** A time series forecasting method for univariate data without a trend or seasonality.
  It can be seen that the level of a time series is captured.

- **Holt's Exponential Smoothing:** The trend itself is updated over time through the second equation, where the trend is expressed as the difference between the last two smoothed values.
  It can be seen that the forecast is a straight line, sloping upwards as Holt's method captured both trend and level.

- **Holt Winters' additive method:** The method produces exponentially smoothed values for the level of the forecast, the trend of the forecast, and the seasonal adjustment to the forecast.
  It can be seen that seasonal variation and the trend is captured.

- **Holt Winters' multiplicative method:** This method is best for data with trend and with seasonality that increases over time. It results in a curved forecast that reproduces the seasonal changes in the data.
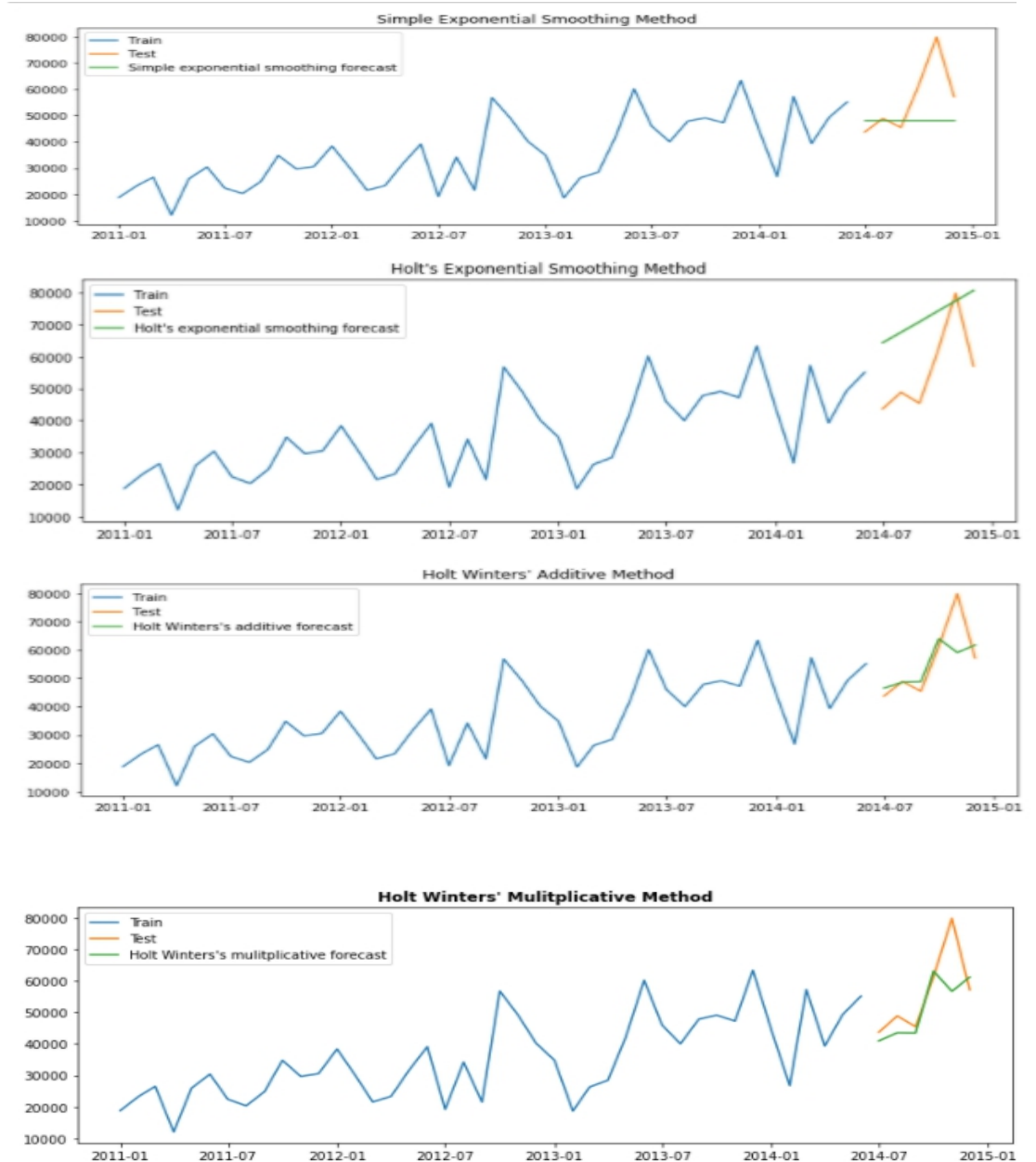
The least RMSE and MAPE values among all the methods done above is Holt Winters' additive method.

- **Auto regression method (AR):** Predict future values based on past values.
  Seasonality is cannot be seen in this plot but trent is captured in the forecast.

- **Moving average method (MA):** Useful for forecasting long-term trends. You can calculate it for any period of time.
  Captured trend but not seasonality in the forecast as per the plot. Too many errors as RMSE and MAPE values are high

- **Auto regression moving average method (ARMA):** Provide a parsimonious description of a stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the second for the moving average (MA).
  ARMA model captured Trend but no seasonality
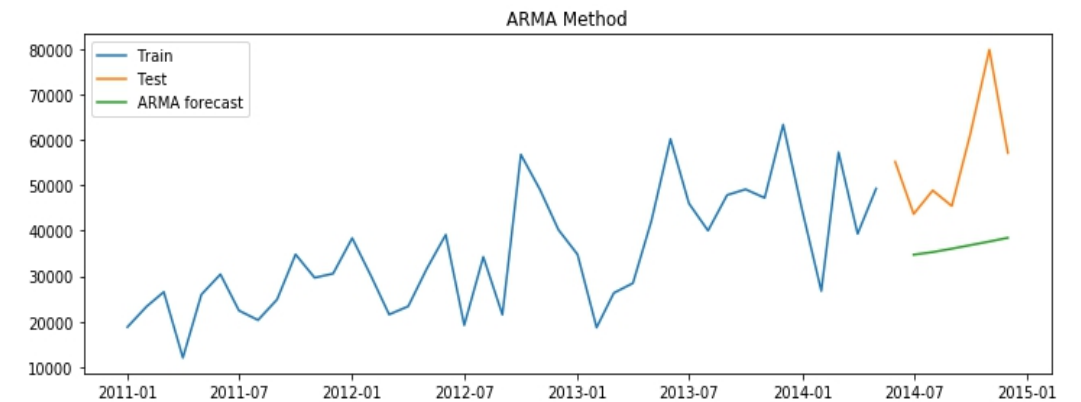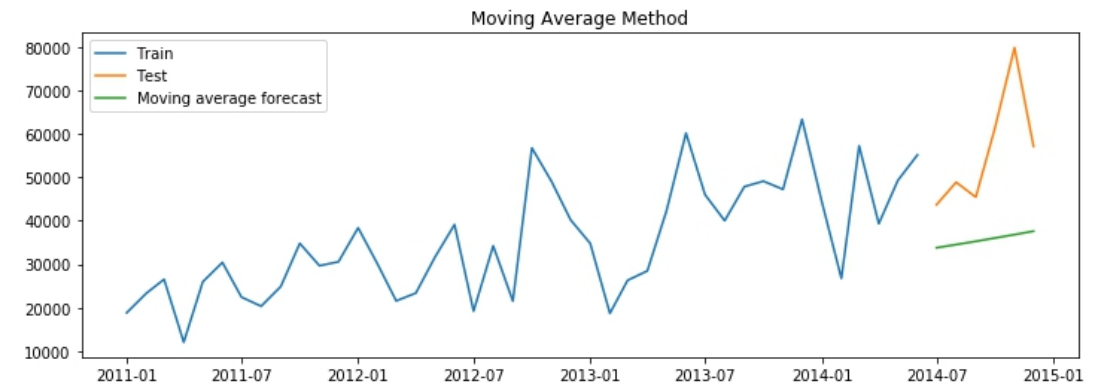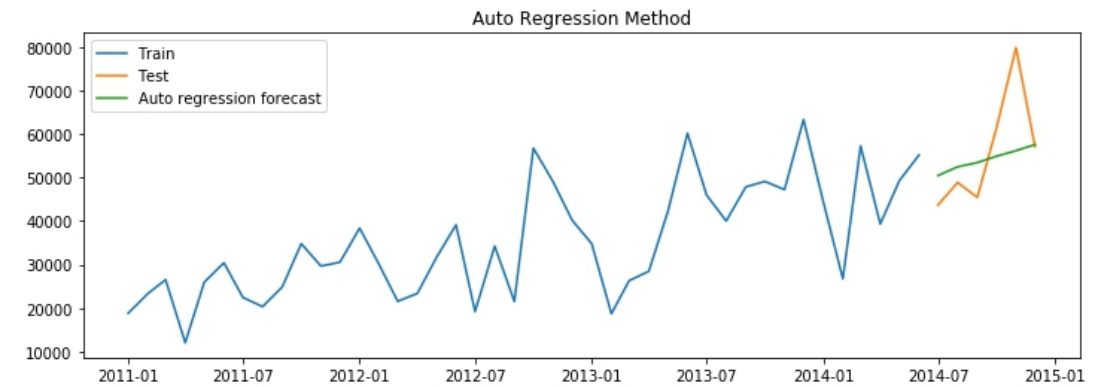
- **Auto regressive integrated moving average (ARIMA):** A statistical analysis model that uses time series data to either better understand the data set or to predict future trends.ARIMA model has three parameters p,d,q. 'd' is the differencing parameter.
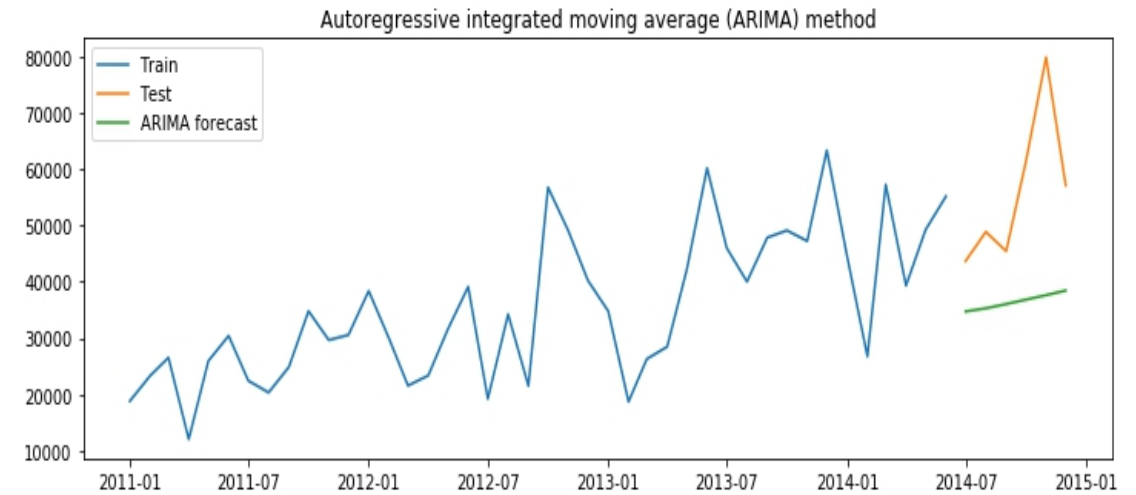
- **Seasonal auto regressive integrated moving average (SARIMA):** Seasonal auto regressive integrated moving average (SARIMA). It has six parameters along with seasonality.
It can be seen in the plot that the forecast captured both trend and seasonality
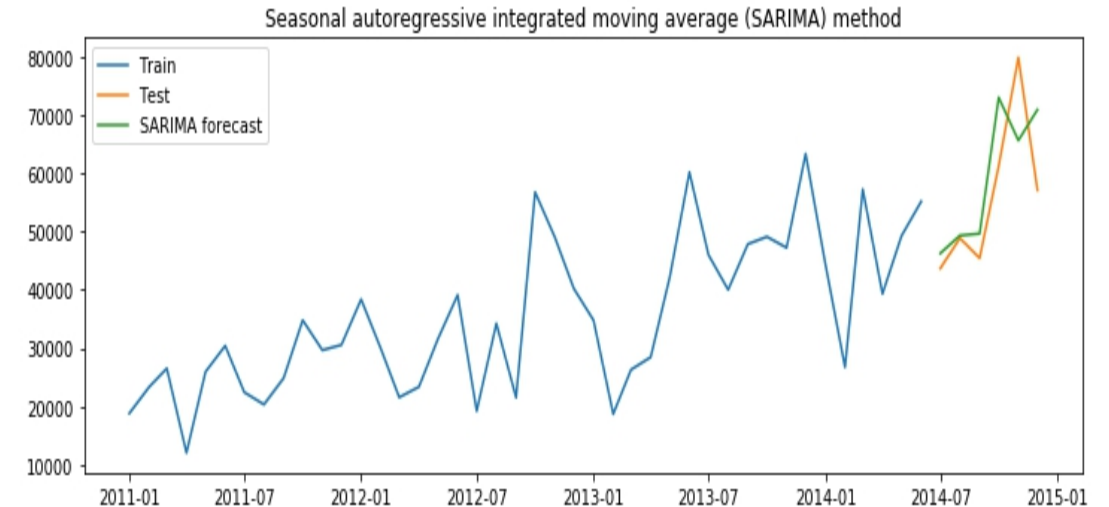
The least RMSE and MAPE values among all the method done in the ARIMA above is SARIMA method.
SARIMA method is able to predict the sales closer to the actual values

Conclusion

- There are 21 unique market Segments.
- The most profitable market segment by comparing the CoV values is APAC_Consumer
- The least RMSE and MAPE values among all forecasting method in the smoothing techniqueis is Holt Winters' additive method.
- The least RMSE and MAPE values among all the method done in ARIMA set of techniques is SARIMA method means the best method.