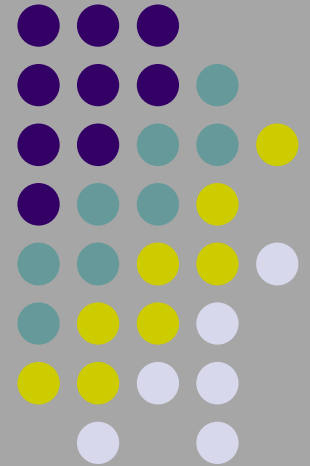




# Global Mart Sales Forecasting Case Study

---

*Presented by*  
*Nikhil Jojen*





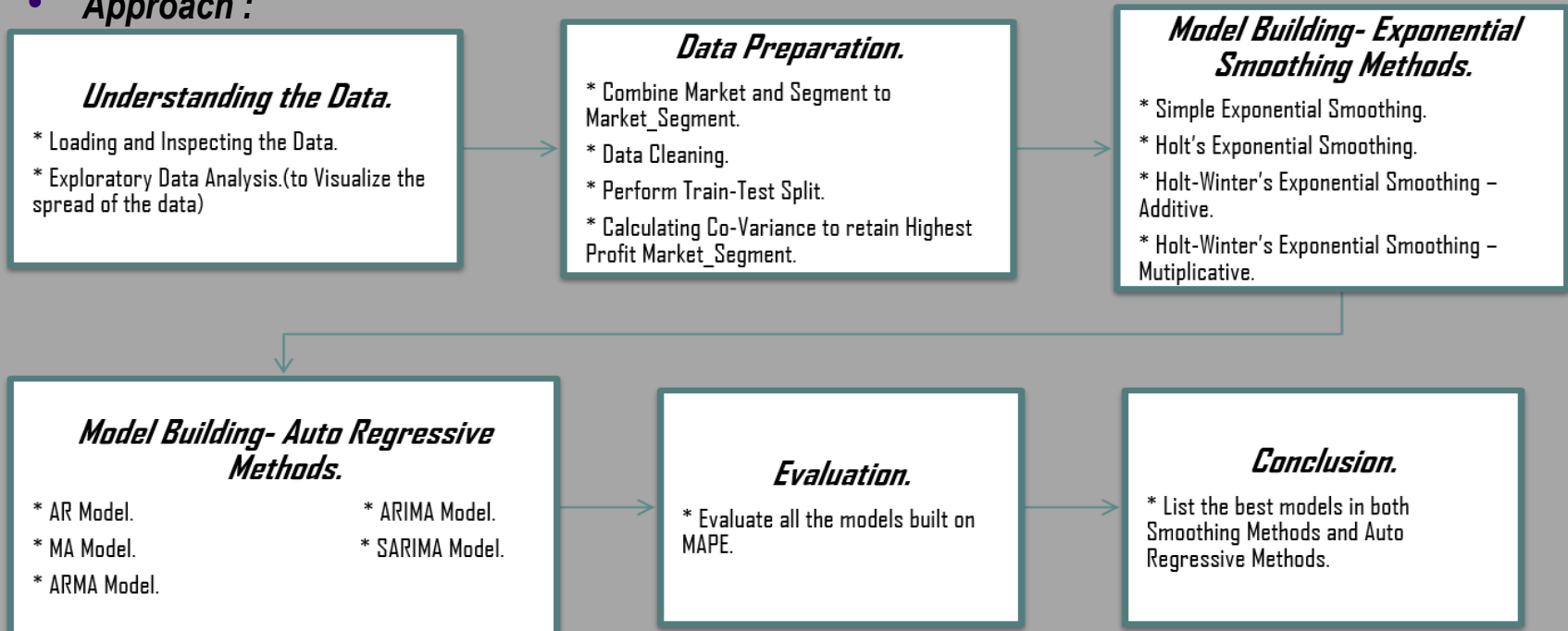
# *Business Goal and Approach*

- **Business Goal :**

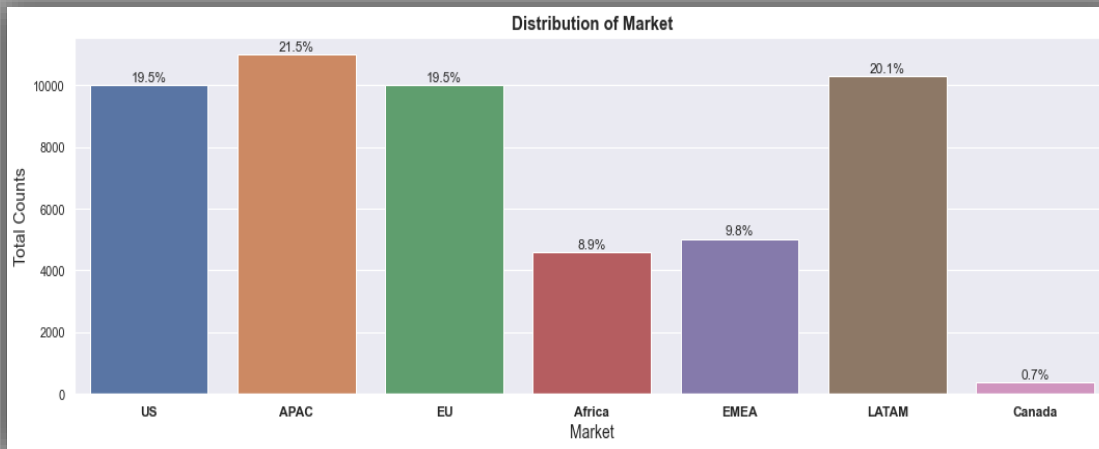
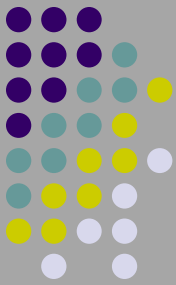
Global Mart is an online supergiant store that has worldwide operations. This store takes orders and delivers across the globe and deals with all the major product categories — consumer, corporate and home office.

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.

- **Approach :**

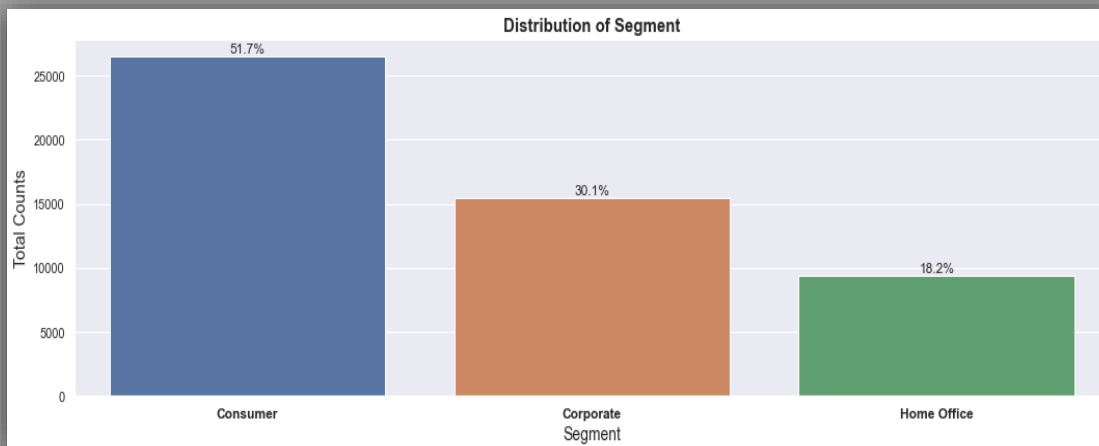


# *Exploratory Data Analysis.*



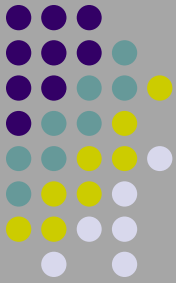
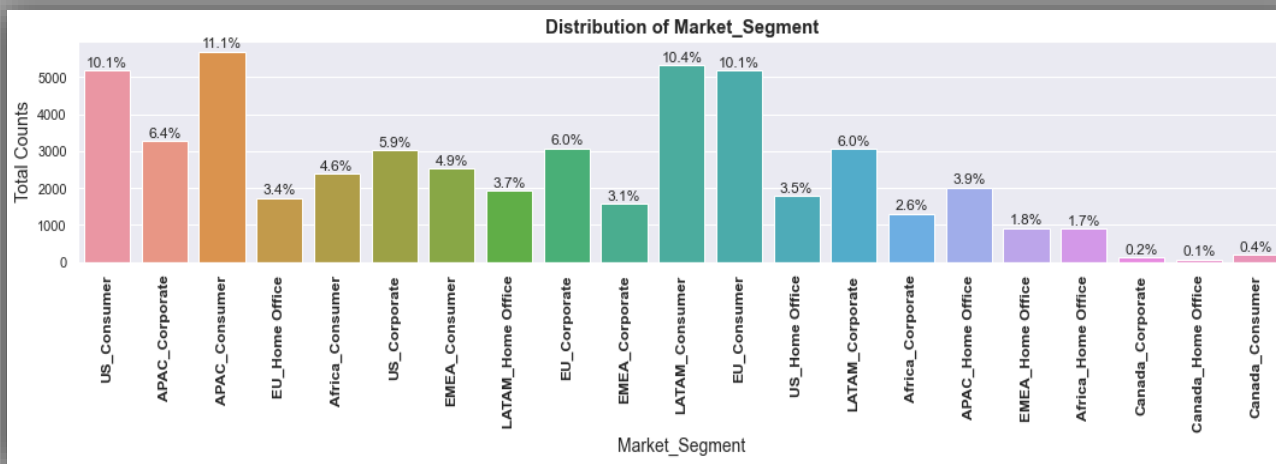
## ***Market Distribution.***

- The maximum percentage of Markets are from APAC followed by LATAM, EU and US.
- Canada has the least percentage of Markets when compared with the data given.



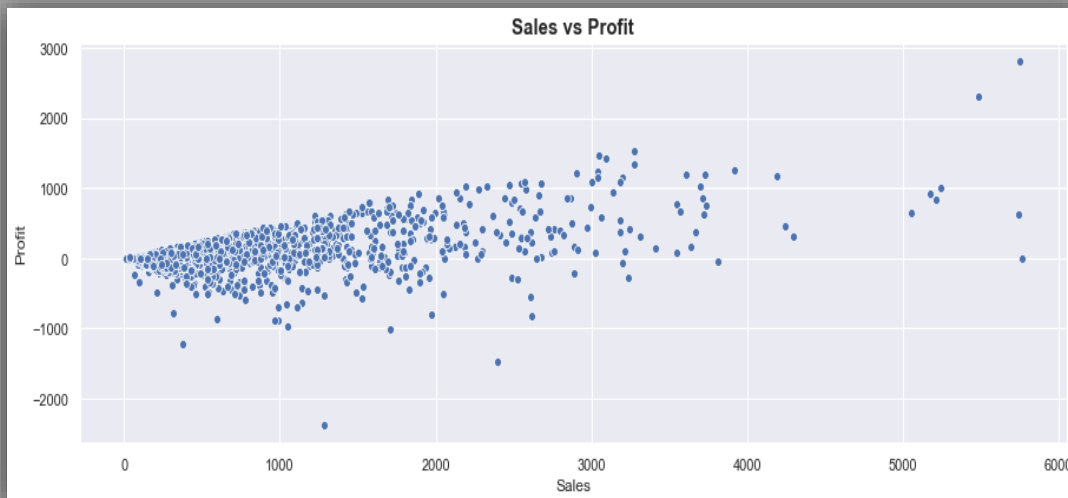
## ***Segment Distribution.***

- The maximum percentage of Segments are from Consumer.



## ***Market\_Segment Distribution.***

- Here we can observe that the maximum percentage of Market\_Segment is from APAC\_Consumer followed by LATAM\_Consumer.
- Market Segments in Canada have the lowest count when compared with the others.

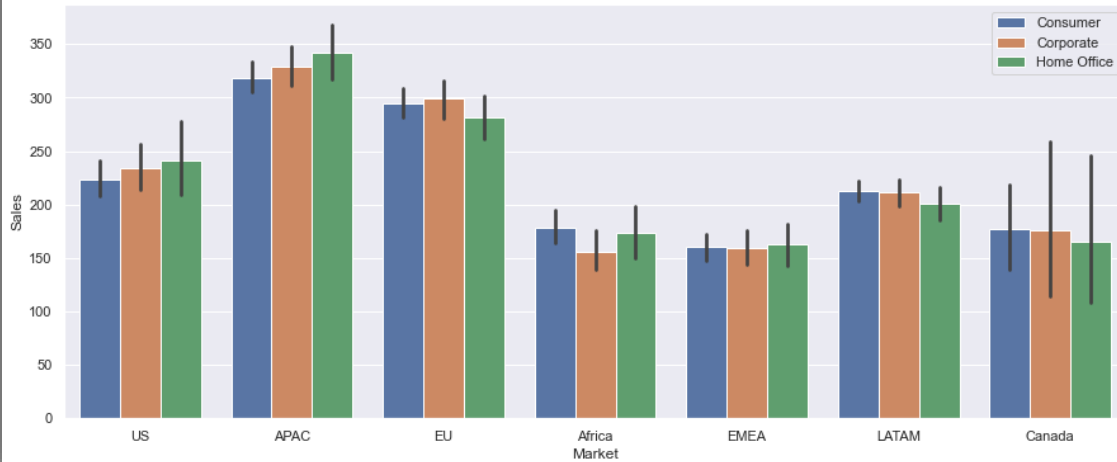


## ***Sales Vs Profit(APAC Market\_Segment).***

- As we can see from the graph, as the sales increases the profit made increases. Hence we can say that Profit is directly proportional to Sales.



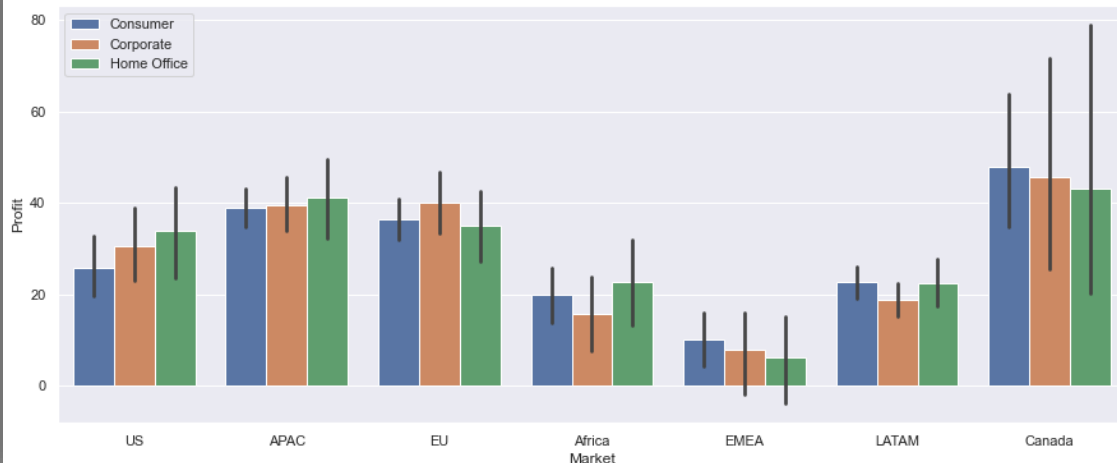
Market Vs Sales



## Market Vs Sales.

- The Maximum sales were from the Market 'APAC' followed by 'EU'.
- The sales are good from Market 'Canada' even though the total number of Markets are very less.

Market Vs Profit

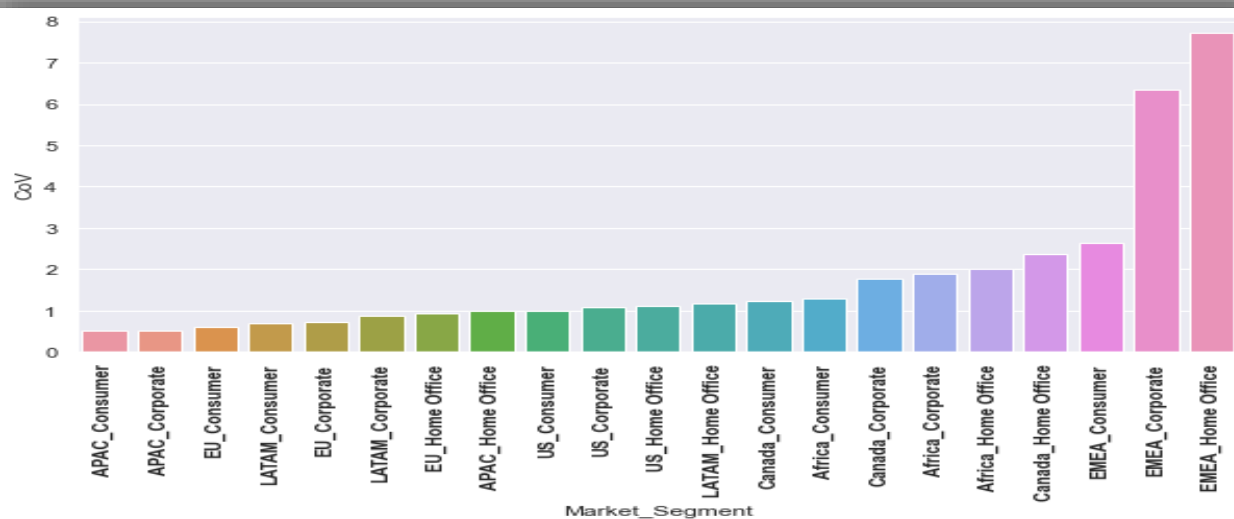


## Market Vs Profit.

- We can see maximum profit were made by Market 'Canada' .
- Even though having large number of Markets in 'APAC' the profit made by them is lower than that of Canada.

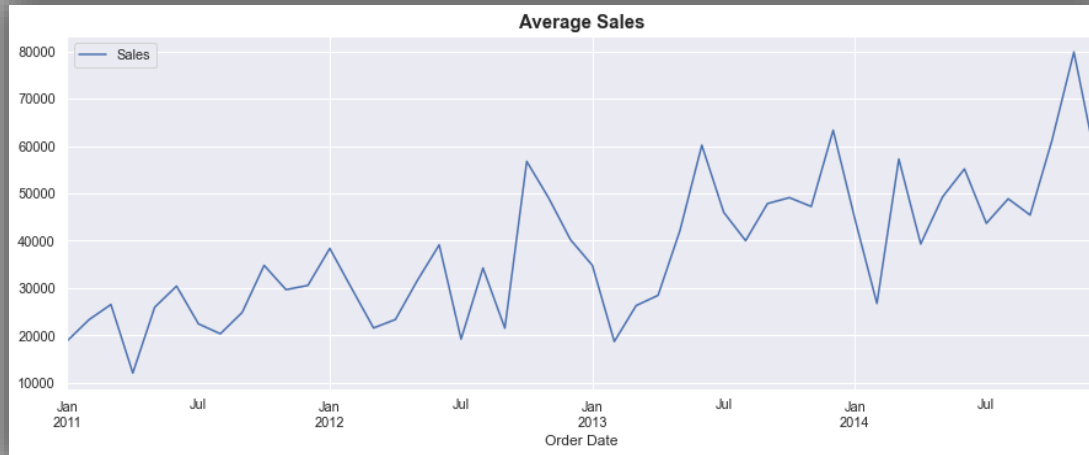
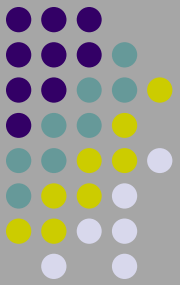
# Data Preparation.

- We combined the Market and Segment columns to one Market\_Segment column for building the model.
- We dropped both Market and Segment columns from the Data-frame.
- Defined a dataset to get aggregated value of profit for each Market\_Segment.
- Split the Data into Train and Test.
- Calculated the Co-Variance for the 21 Market\_Segments.
- From the table we can see that the Lowest Cov(Covariance) is 0.522725, which corresponds to APAC\_Consumer. Hence the most profitable Market\_Segment is APAC\_Consumer.
- Since we were asked to forecast for the most profitable Market\_Segment, we will be using APAC\_Consumer for model building.



	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

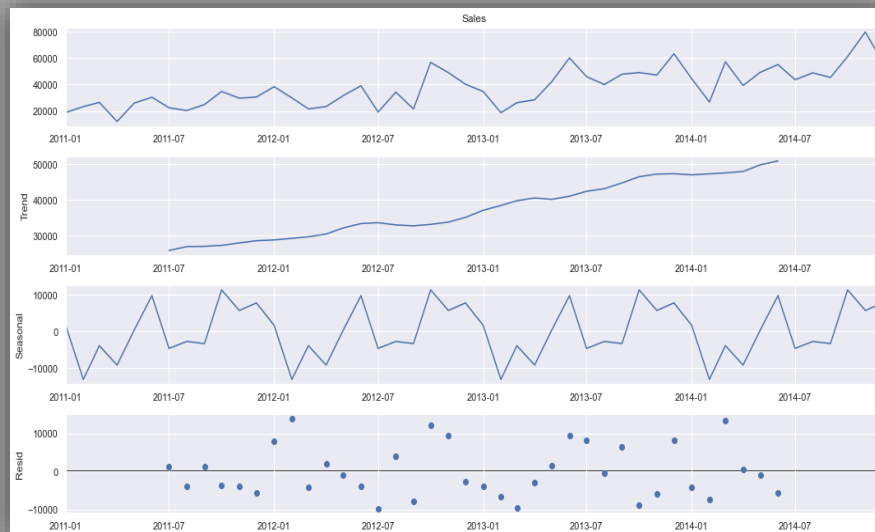
# Model Building.



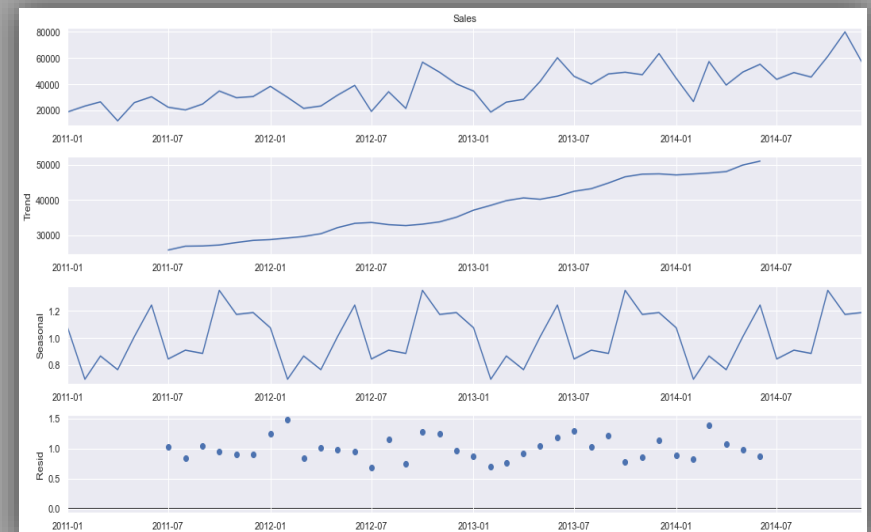
## *Average Sales Spread across the Years.*

- It is quite evident that the trend is steadily increasing.
- The seasonality extract shows a slight dip in sales during the month of January and July every year.
- In general, sales are higher in the middle of the year.

## *Additive Decomposition.*

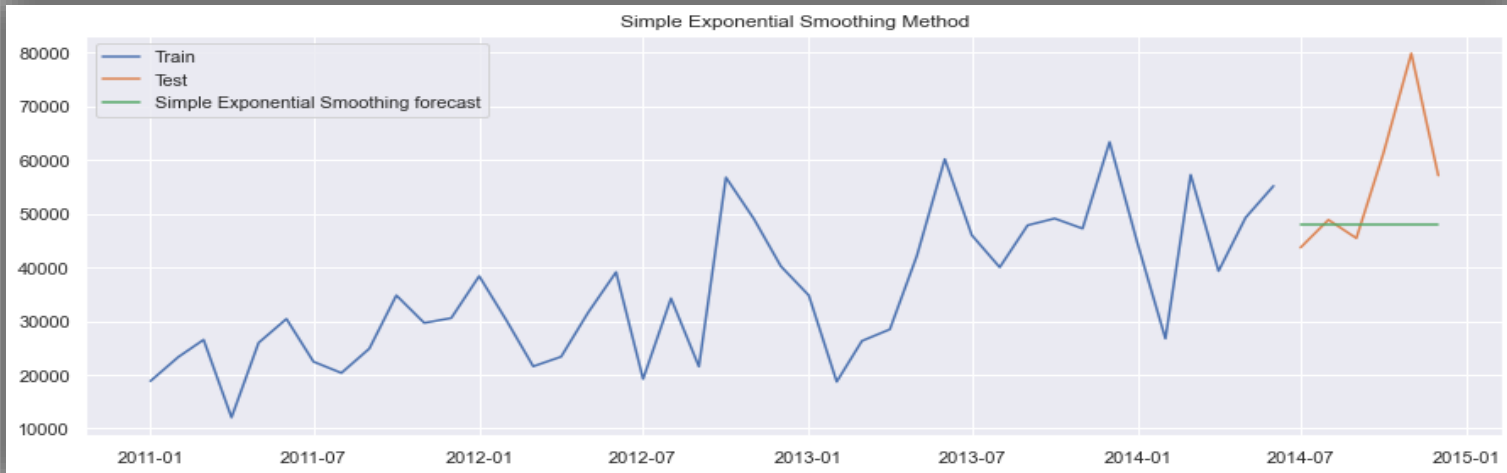


## *Multiplicative Decomposition.*

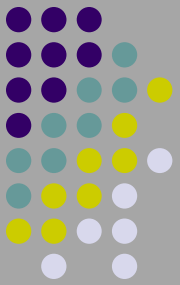
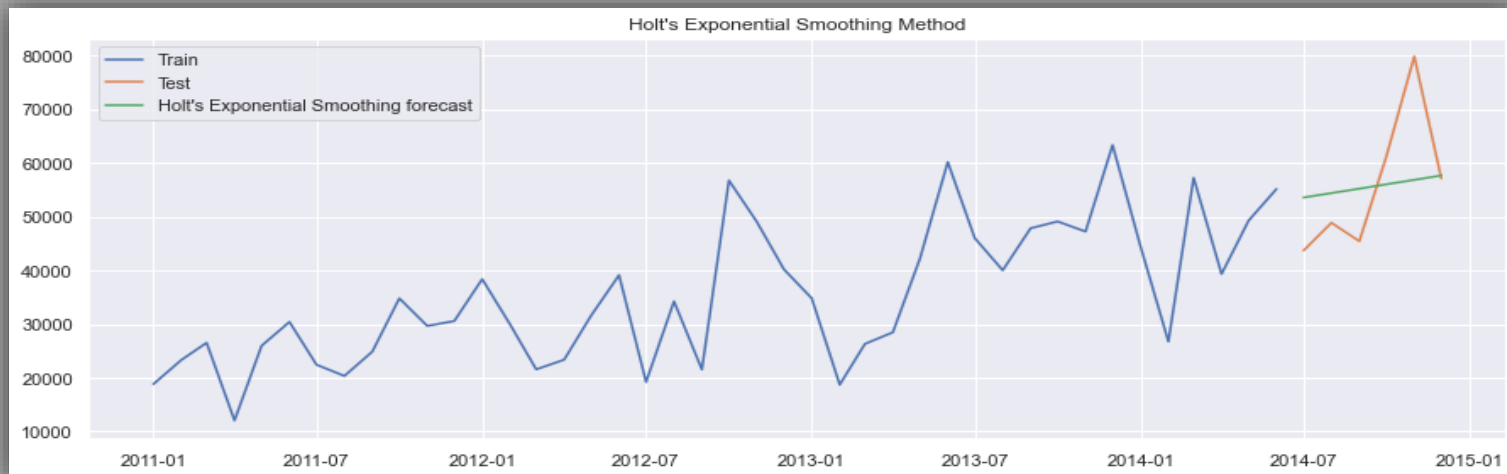


# *Smoothing Techniques.*

## *Simple Exponential Smoothing Method.*

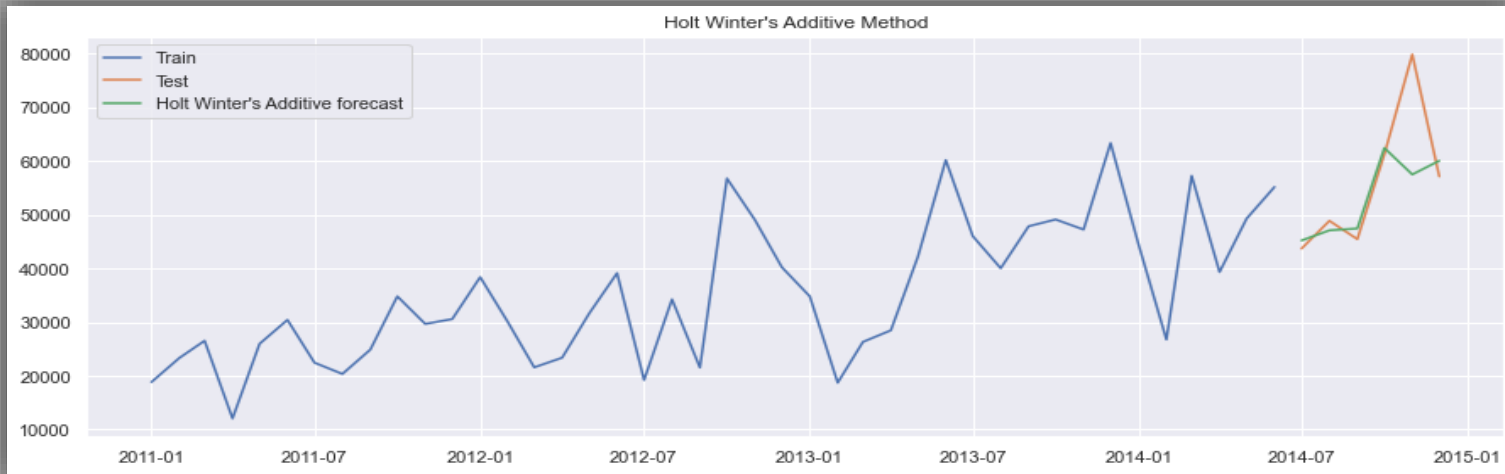


## *Holt's Exponential Smoothing Method.*

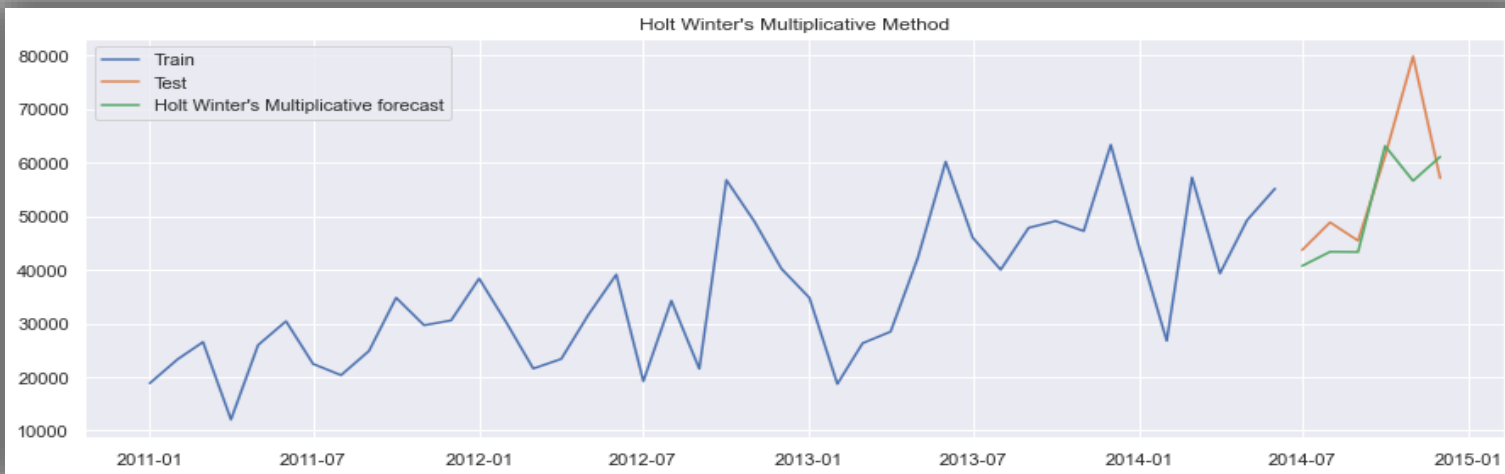




## ***Holt-Winter's Additive Method.***



## ***Holt-Winter's Multiplicative Method.***



# ARIMA Set of Techniques.

## Augmented Dickey-Fuller(ADF) Test.

ADF Statistic: -3.376024

Critical Values @ 0.05: -2.93

p-value: 0.011804

## Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test.

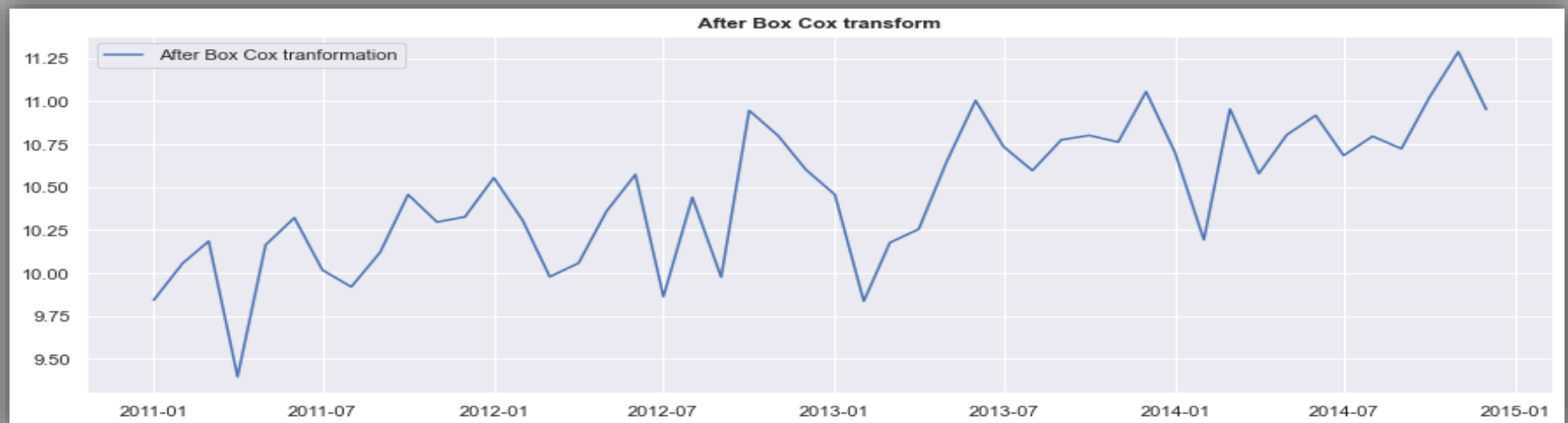
KPSS Statistic: 0.577076

Critical Values @ 0.05: 0.46

p-value: 0.024720

## Stationary Vs Non-Stationary Test Series.

- ADF Test.
  - × *p-value is 0.011804 < 0.05, Fail to Reject the Alternate Hypothesis(H1). Hence, the series is stationary.*
- KPSS Test.
  - × *p-value is 0.024720 < 0.05, Fail to Reject the Alternate Hypothesis(H1). Hence, the series is not stationary.*
- *Since, both the Tests doesn't prove the Time Series to be Stationary, we will Perform Box-Cox Transformation and Differencing, to get a Stationary Time Series.*





## *After BoxCox Transformation And Differencing.*

### *Augmented Dickey-Fuller(ADF) Test.*

ADF Statistic: -4.535011

Critical Values @ 0.05: -2.95

p-value: 0.000170

### *Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test.*

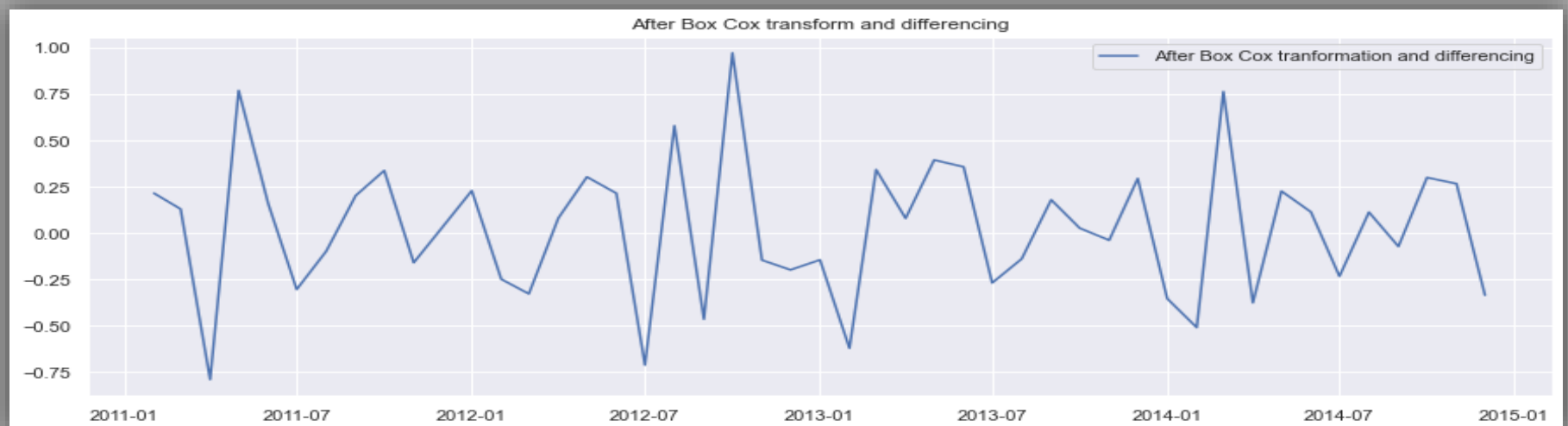
KPSS Statistic: 0.156456

Critical Values @ 0.05: 0.46

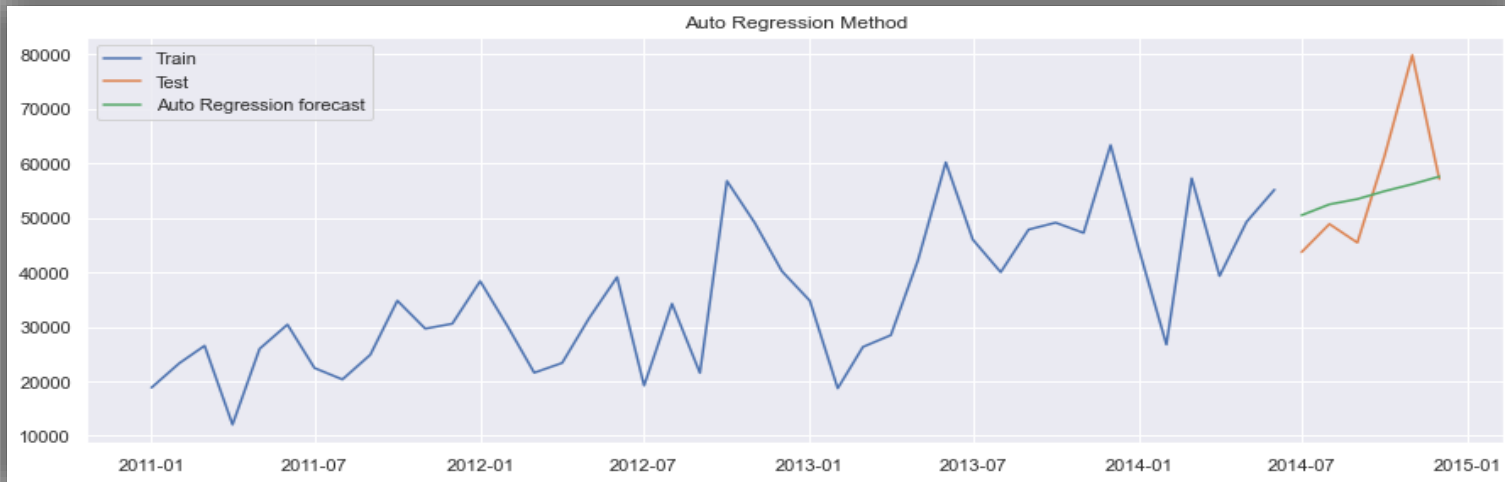
p-value: 0.100000

### *Stationary Vs Non-Stationary Test Series.*

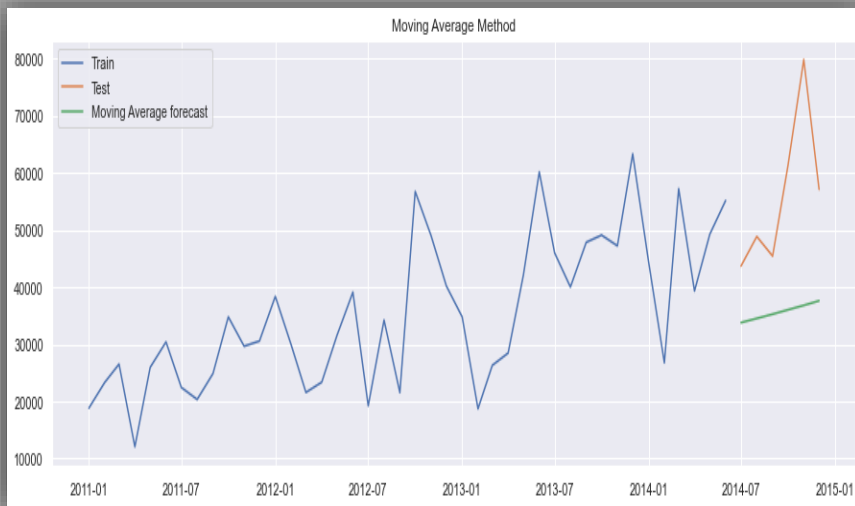
- *ADF Test.*
  - × *p-value is 0.00017 < 0.05, Fail to Reject the Alternate Hypothesis(H1). Hence, the series is stationary.*
- *KPSS Test.*
  - × *p-value is 0.100 > 0.05, Fail to Reject the Null Hypothesis(H0). Hence, the series is stationary.*
- *After the Box-Cox Transformation and Differencing the model now has become stable for the Auto Regressive Methods.*



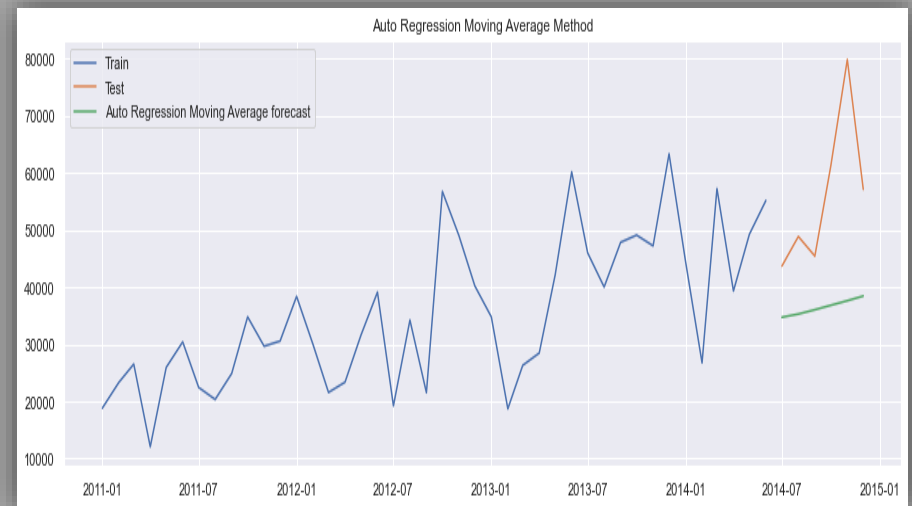
## ***Auto Regressive(AR) Method.***



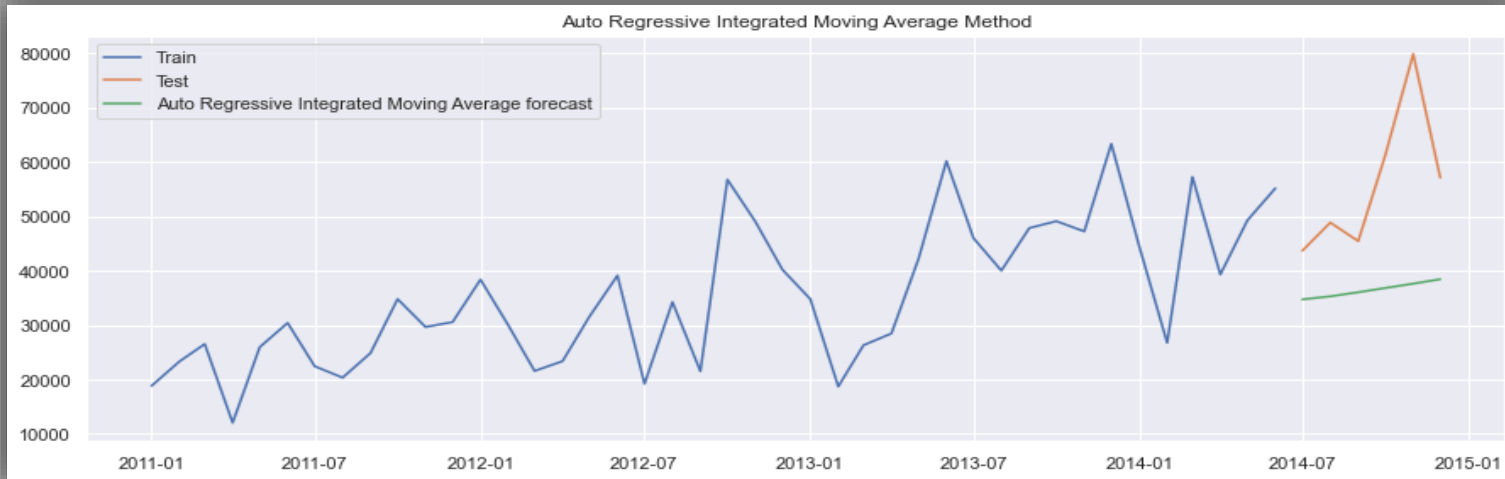
## ***Moving Average(MA) Method.***



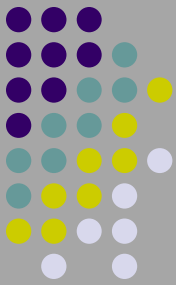
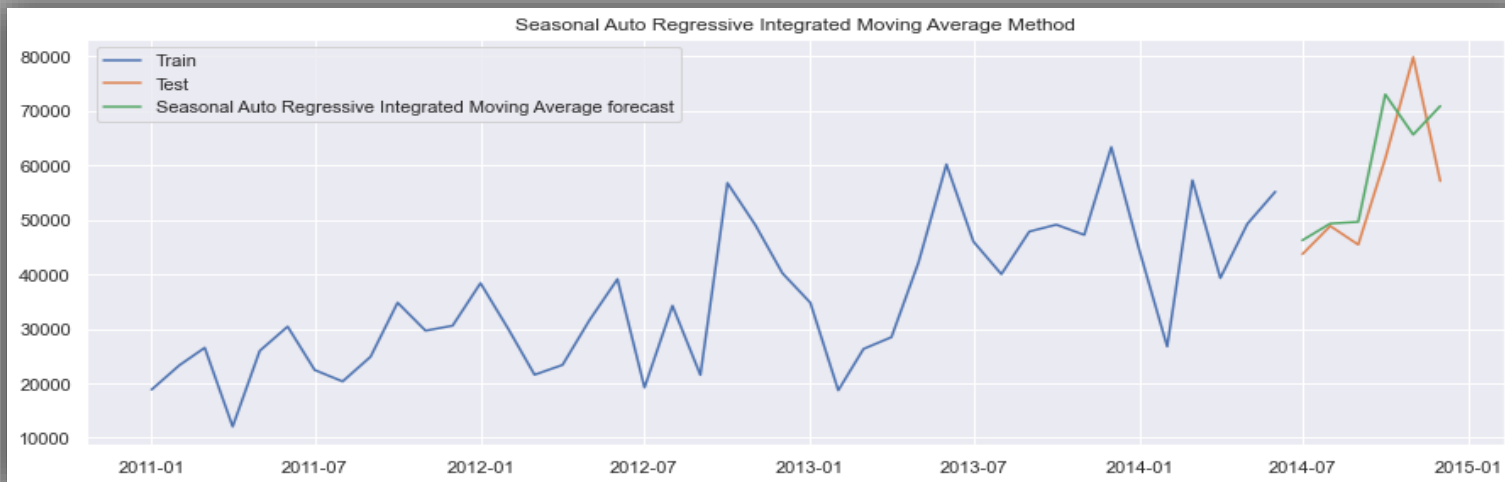
## ***Auto Regressive Moving Average(ARMA) Method.***



## ***Auto Regressive Integrated Moving Average(ARIMA) Method.***



## ***Seasonal Auto Regressive Integrated Moving Average(SARIMA) Method.***



# *Evaluation.*



	Method	RMSE	MAPE
0	Simple exponential smoothing forecast	14764.66	15.83
0	Holt's Exponential Smoothing Method	11402.24	15.63
0	Holt Winter's Additive Method	9304.48	7.76
0	Holt Winter's Multiplicative Method	10021.79	10.30
0	Auto Regression Method	10985.28	13.56
0	Moving Average Method	23360.02	33.93
0	Auto Regression Moving Average Method	22654.32	32.40
0	Auto Regressive Integrated Moving Average Method	22654.32	32.40
0	Seasonal Auto Regressive Integrated Moving Ave...	9612.36	12.86

# *Conclusion.*

- *The most profitable segment is APAC\_Consumer segment as it has the lowest COV value of all segments which means that it has high sales.*
- *In the smoothing technique, Holt Winter's Additive Method is a better forecasting method.*
- *In the set of ARIMA techniques, SARIMA is the most effective.*

The RMSE and MAPE values of Holt Winters' additive method and SARIMA are the lowest. Nonetheless, looking at the forecast plot, we can see that SARIMA has much better forecasting. Consequently, if we consider only smoothing techniques, Holt Winter's additive forecasting method will be best, while from the several ARIMA methods SARIMA is the best. For this case study, SARIMA forecasting method is the best approach.





**THANK YOU**