

# Retail Giant Sales Forecasting Assignment

Ankita Patel

DSC42, Business Analyst

# Problem Statement - I

- Global Mart is an online supergiant store with worldwide operations, taking orders in the product categories — consumer, corporate and home office.
- The company wants to forecast the sales of the products for the next 6 months for that market segment which is the best and most consistent in terms of profitability.
- This will help in planning the inventory and business processes.

# Problem Statement - II

- Model Building and Evaluation
  - data Understanding
  - finding the most profitable segments
  - forecasting sales and demand for each profitable segments
  - recommendation for inventory management

# Steps of Analysis

## Data Preparation

- For the given dataset, the order date format is changed to get a monthly aggregated transaction data. The market and segment columns are concatenated to get the time series data consisting of order date, sales, profit, market segment.
- The “Market” attribute has 7-factor levels representing the geographical sector and the “Segment” attribute tells which of the 3 segments that customer belongs to, making it a total of 21 unique market segments.
- For these market segments, the coefficient of variation (CoV) is calculated and compared on the respective profits. The most profitable market segment is arrived at based on the CoV value.

# Modelling & Evaluation

- For forecasting the sales, all the techniques in the Smoothing and ARIMA set of methods are applied.
- The sales for the concerned market segment is plotted, and the data is decomposed to get insights from the trend, seasonal and residual components. Based on these insights, the best suited smoothing and ARIMA technique will be concluded.
- All the models in the smoothing and ARIMA set of techniques (except ARIMAX and SARIMAX) will be applied and the sales will be forecasted for the next 6 months. These plots will be compared along with the MAPE values which will be added in the same table.

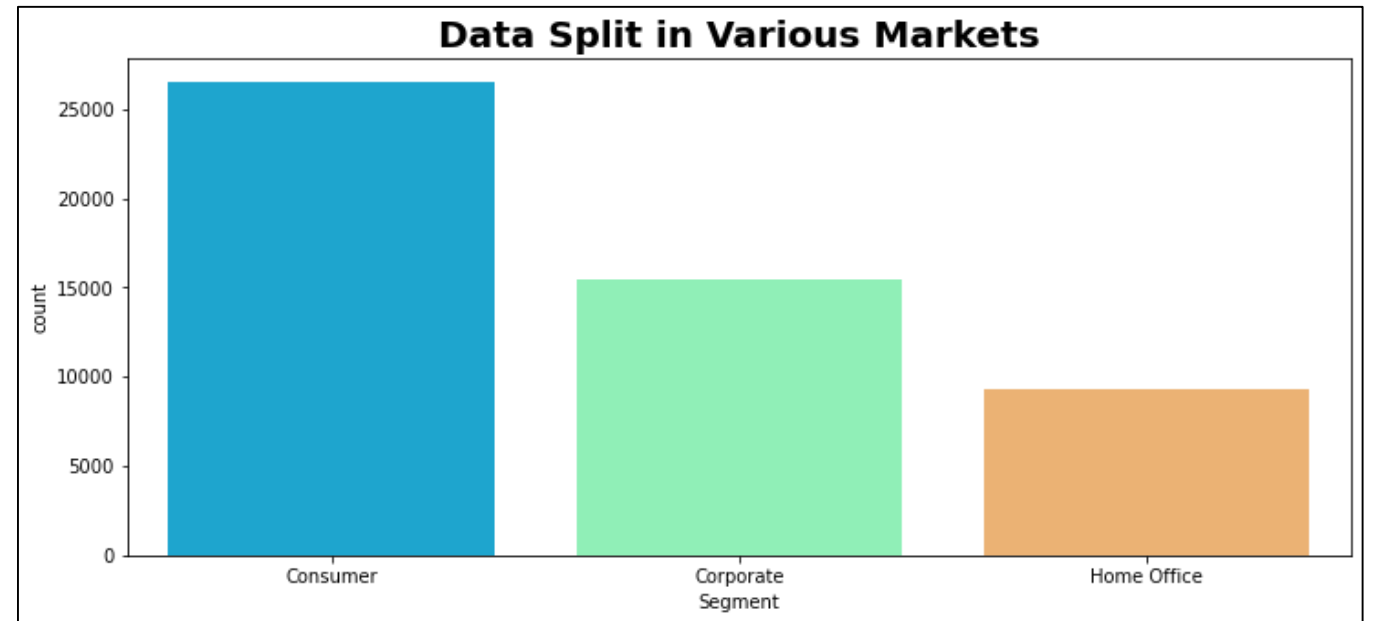
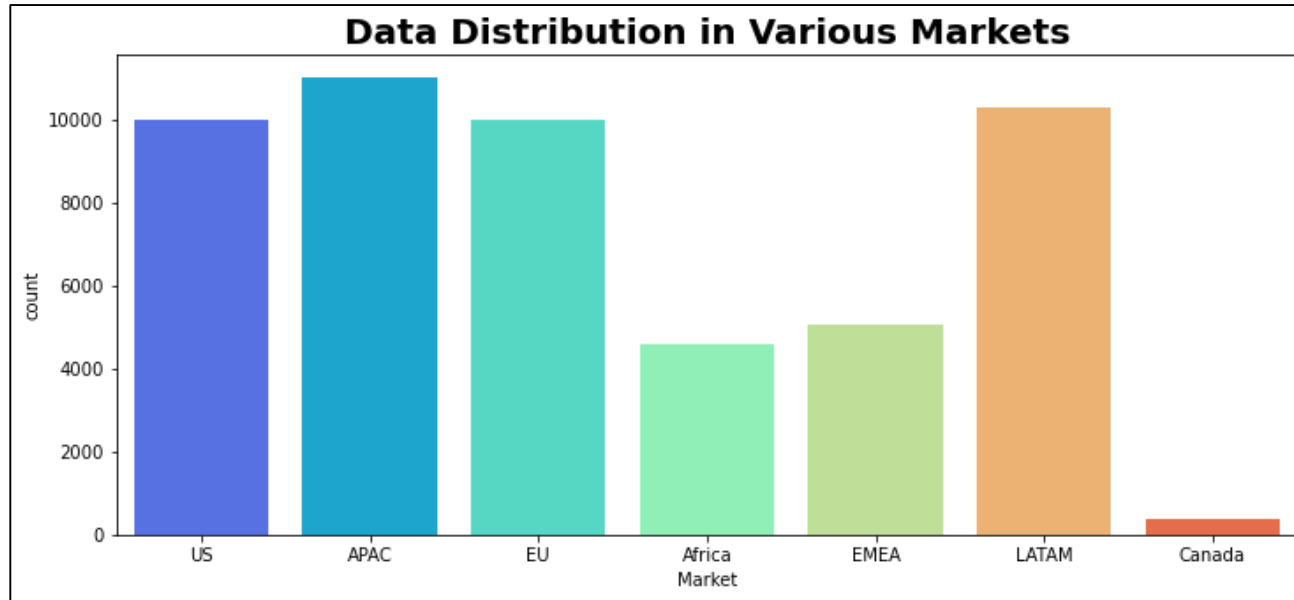
# Model Development and Evaluation – Steps

After preparing the dataset for the 21 market segments, calculating the CoV and finding the most profitable market segment, the prediction model is developed as follows:

- Drop the rest of the 20 market segments and keep the most profitable market segment.
- The data is split such that there are 42 months in train set and 6 months in the test set.
- Time-series data is decomposed to get the trend, seasonality and residual parts.
- The best suited smoothing and ARIMA technique will be decided based on the insights from above.
- All the below methods (not limited to) will be applied, and observation will be noted if hypothesis formed from the flowchart was correct or not based on the forecast plots and their MAPE values:
  - Simple exponential smoothing
  - Holt's exponential smoothing
  - Holt-Winters' exponential smoothing – Additive
  - Holt-Winters' exponential smoothing – Multiplicative
  - AR, MA, ARMA
  - ARIMA, SARIMA
- The smoothing techniques whose sales' forecast is closest to the actual values and whose MAPE values is the least among all the methods done above will be concluded as the best method.

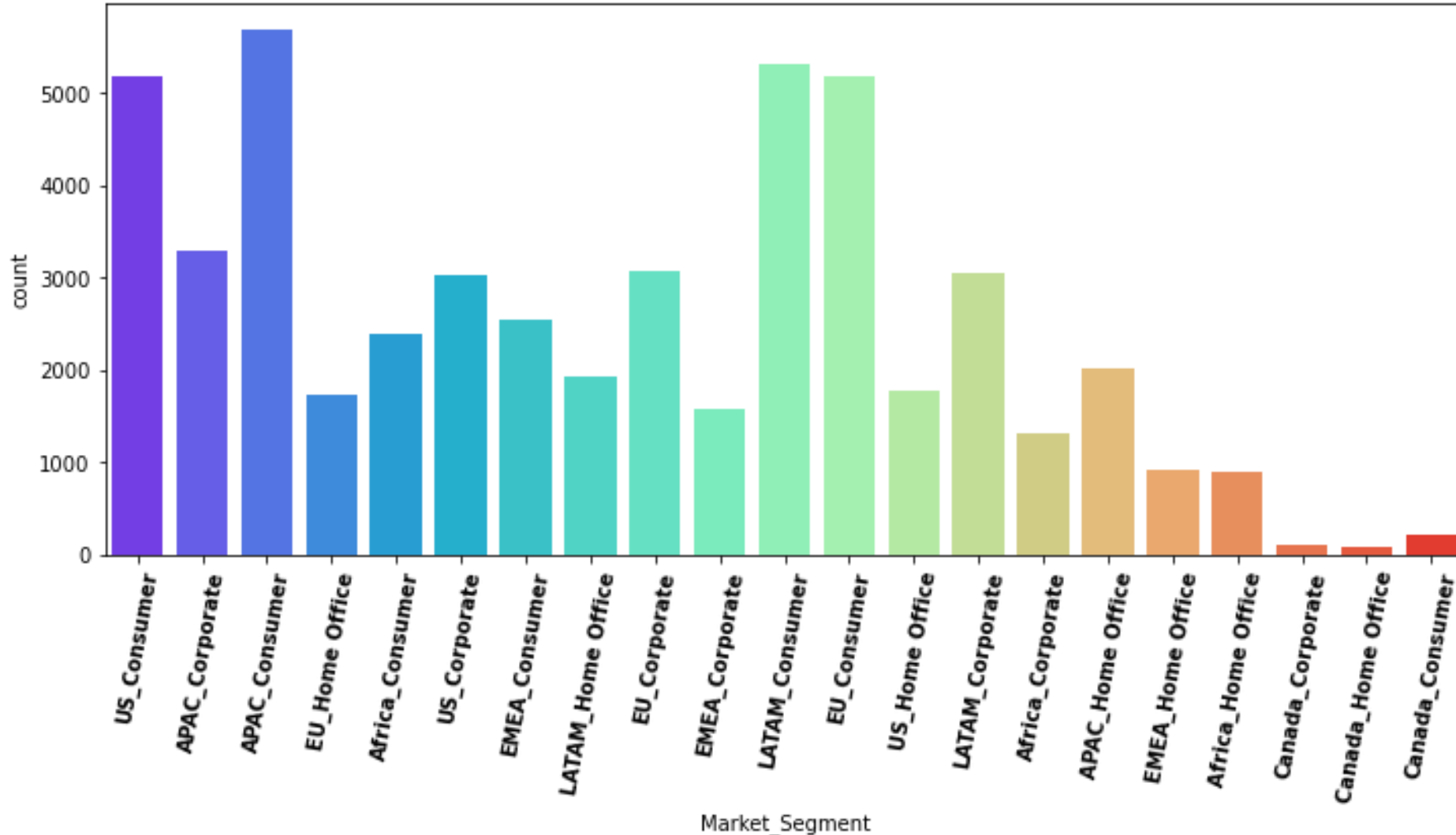
# Data Distribution

Market wise  
&  
Segment wise



# Market Segments in the Data

Data Split in Various Markets Segments

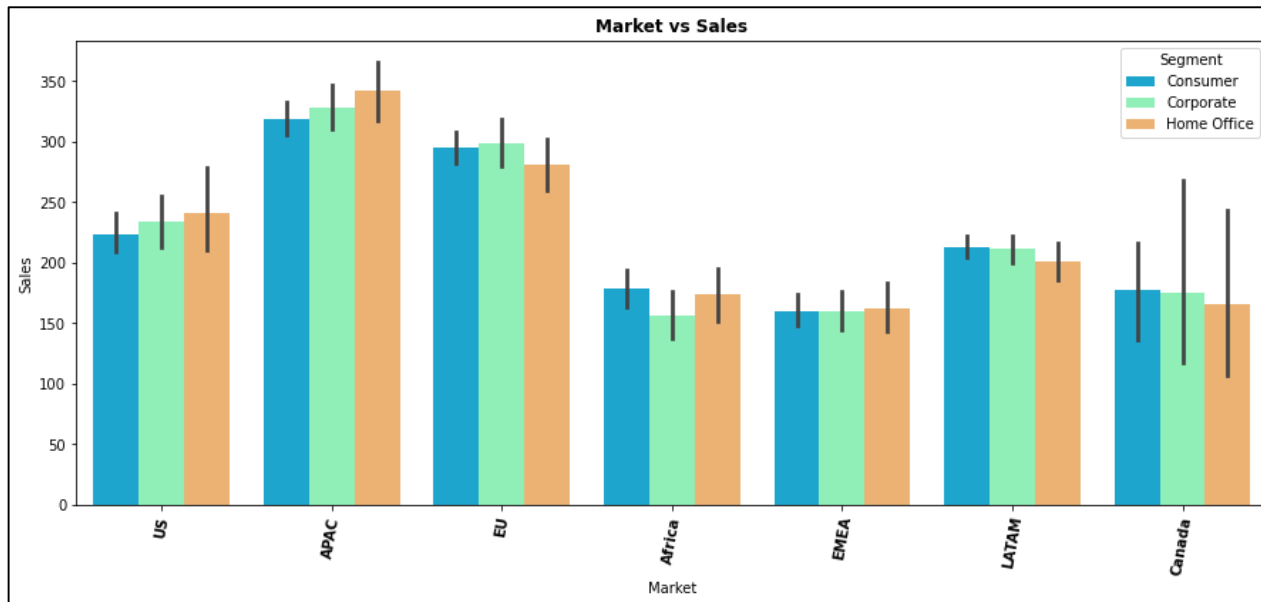
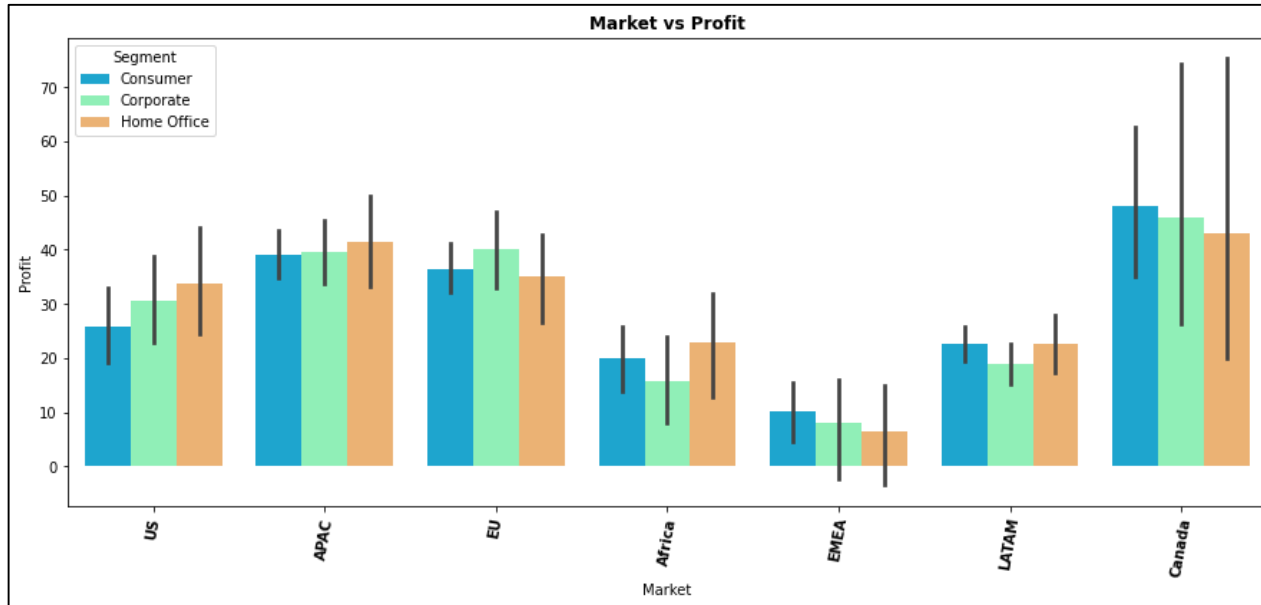


Observations:

We can see from the plot APAC\_Consumer has the highest count



# EDA – Bivariate Analysis



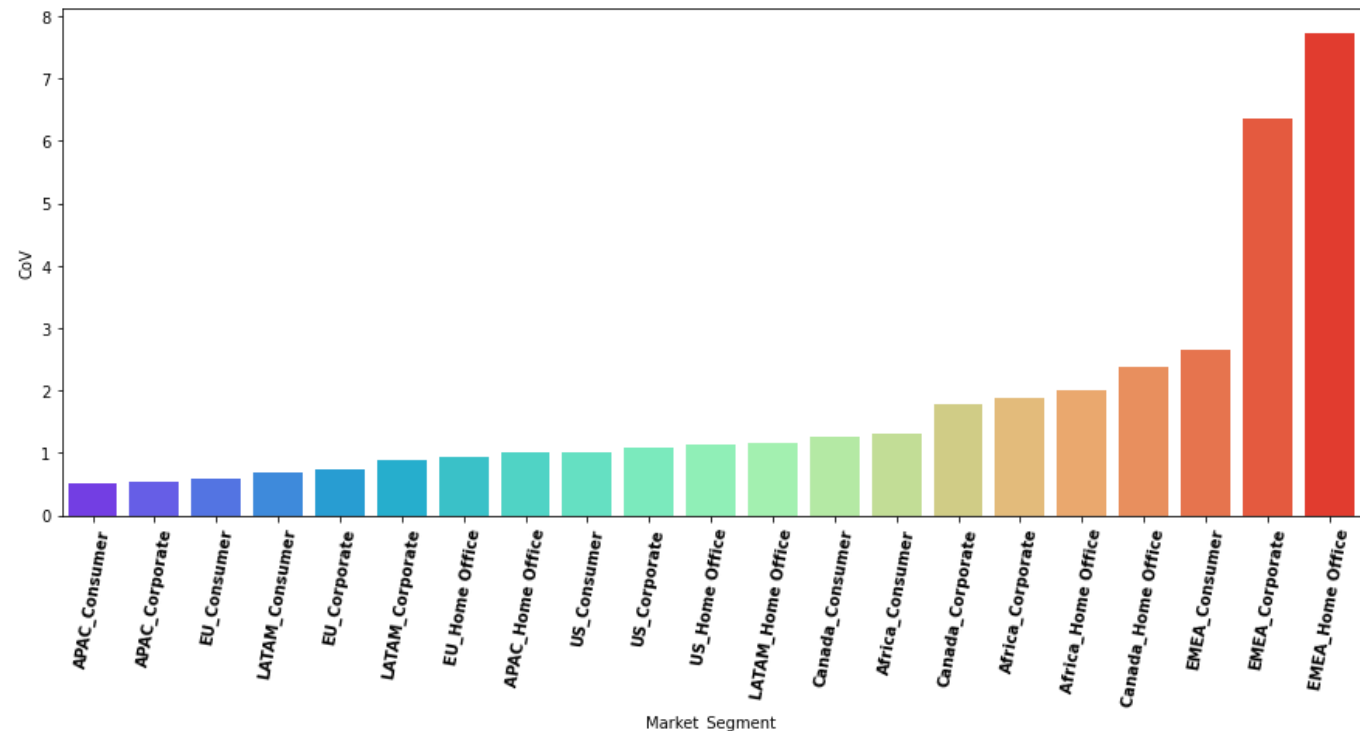
## Observations:

- Canada and APAC markets with the segments Consumer, Corporate and Home Office have more profit
- Africa and EMEA are with less profits
- APAC Market has higher sales in all 3 segments
- Africa and EMEA are with less sales too

# The most profitable Market Segment

	Market_Segment	Mean	Stdev	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

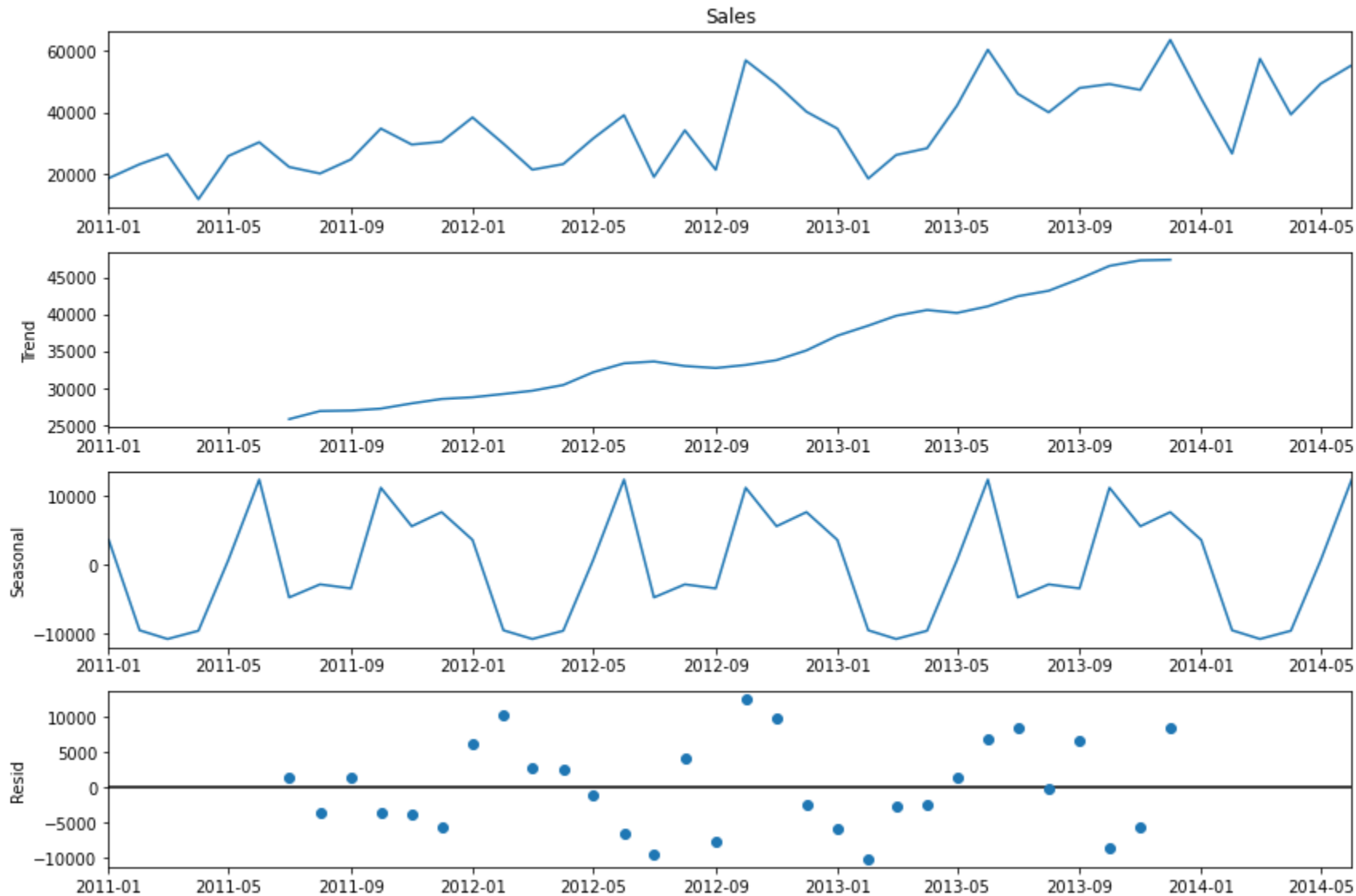
## Coefficient of Variation (CoV) on Profit



# Coefficient of Variation (CoV)

- Coefficient of Variation (CoV) is the ratio of the standard deviation to the mean - useful statistic for comparing the degree of variation from one data series to another, even if the means are very different from one another
- Significance: Lower the CoV value of a series, higher the stability and lower the fluctuations.
- The table alongside shows the comparison of values for the CoV calculated on the profit for all the 21 market segments
- It can be observed that the APAC\_Consumer is the market segment with least CoV, meaning that is the segment which is consistently most profitable
- The sales forecast will be done for this segment for a period of 6 months

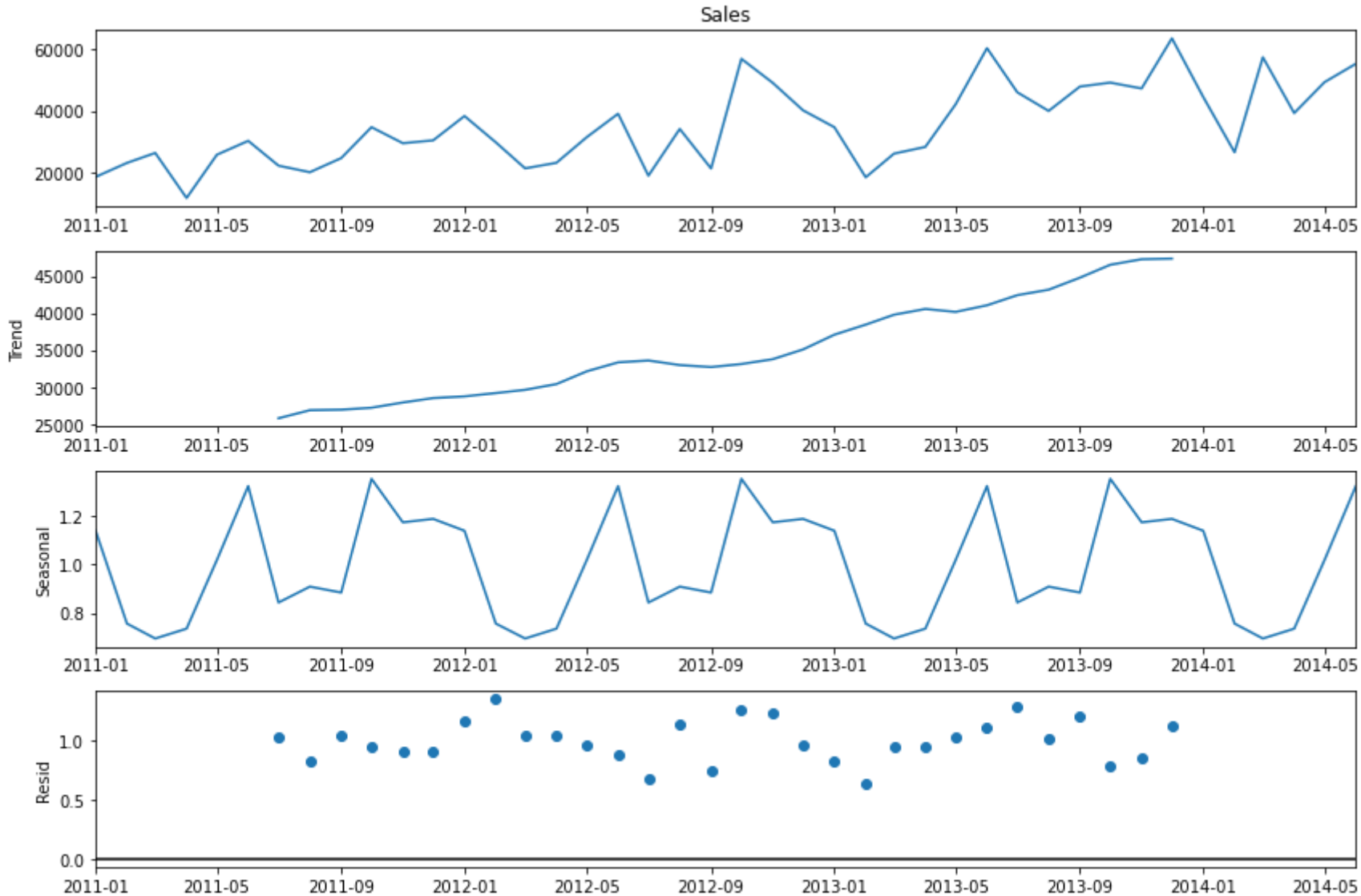
# Time Series Decomposition - Additive



## Observations:

- Upward Trend
- Yearly seasonality in the data

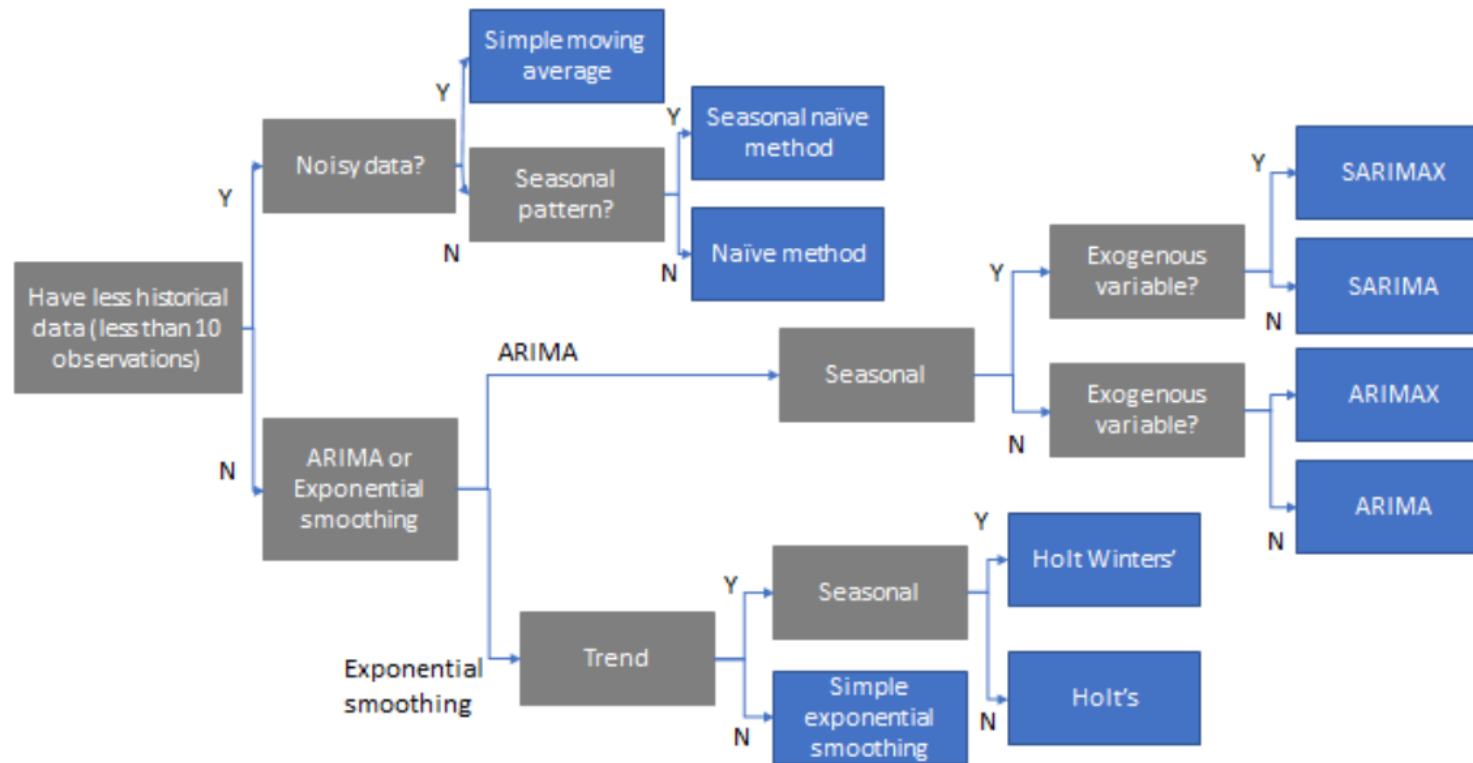
# Time Series Decomposition - Multiplicative



## Observations:

- Upward Trend
- Yearly seasonality in the data

# Choosing the Right Time Series Method



According to the flow chart of choosing the right method

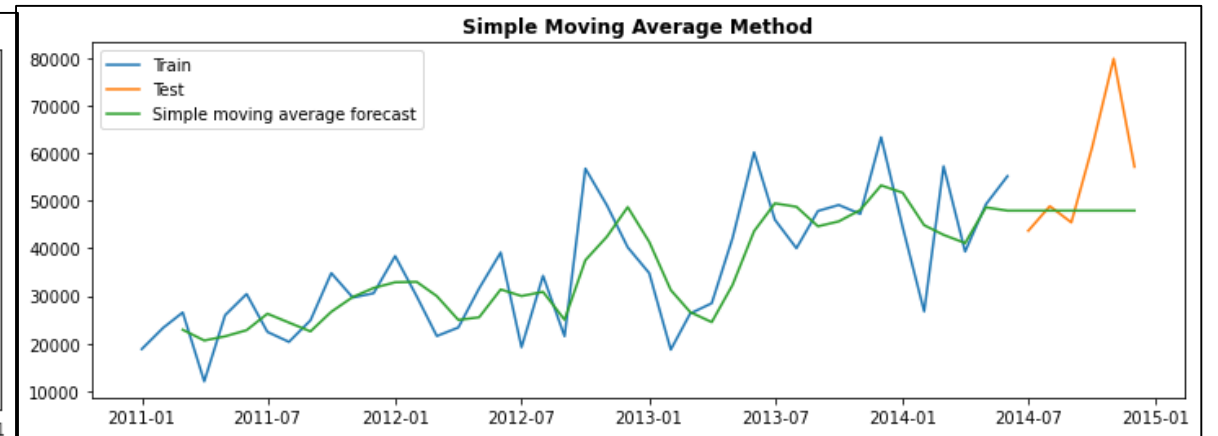
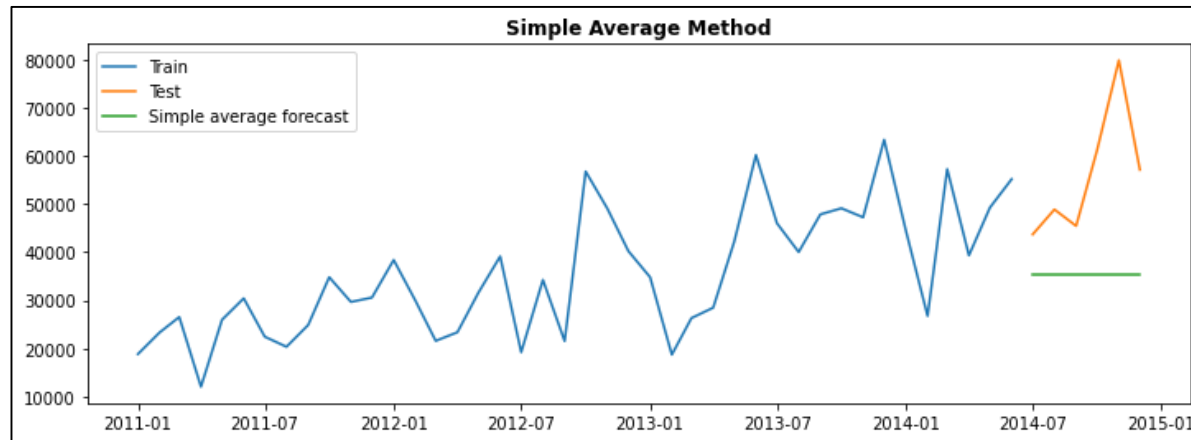
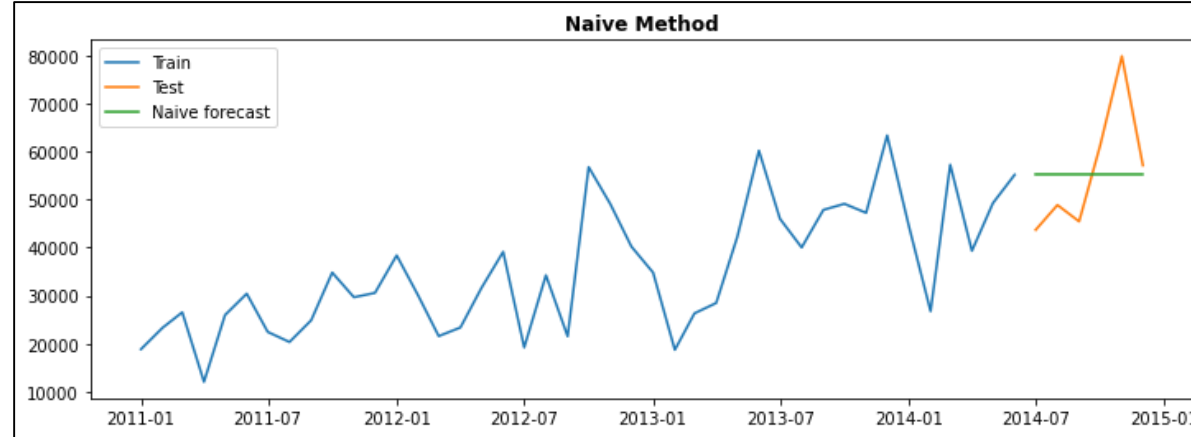
1. Data is more than 10 observations

➤ ARIMA or Exponential Smoothing

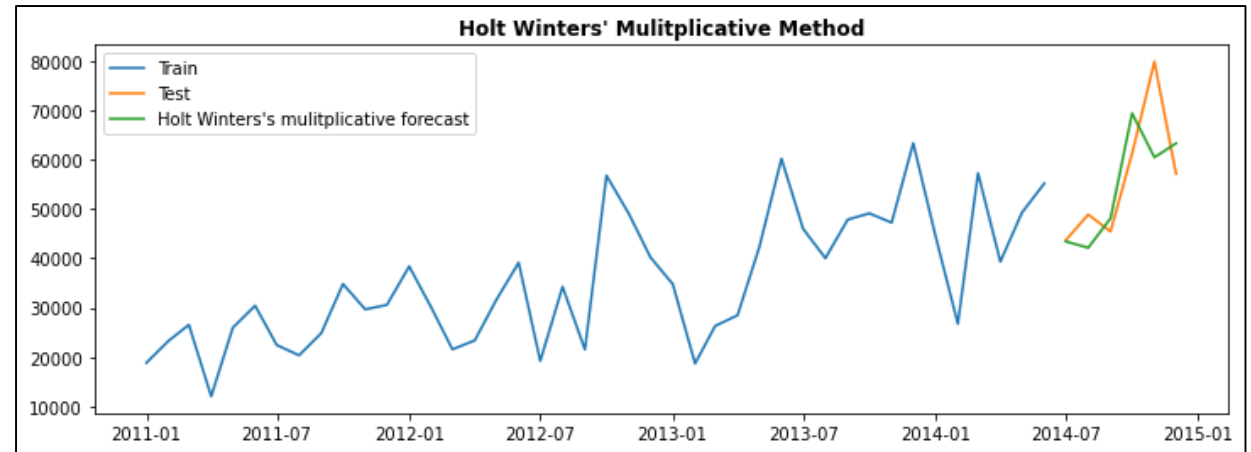
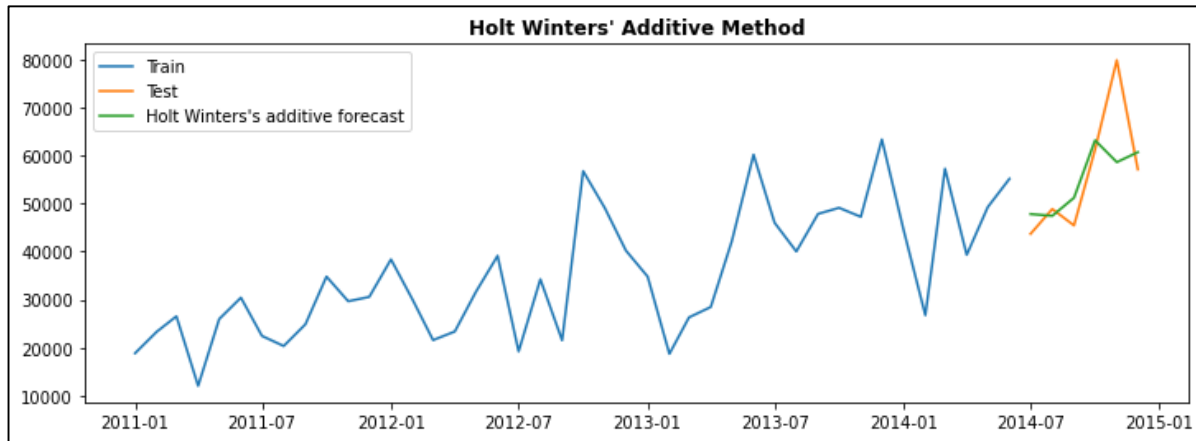
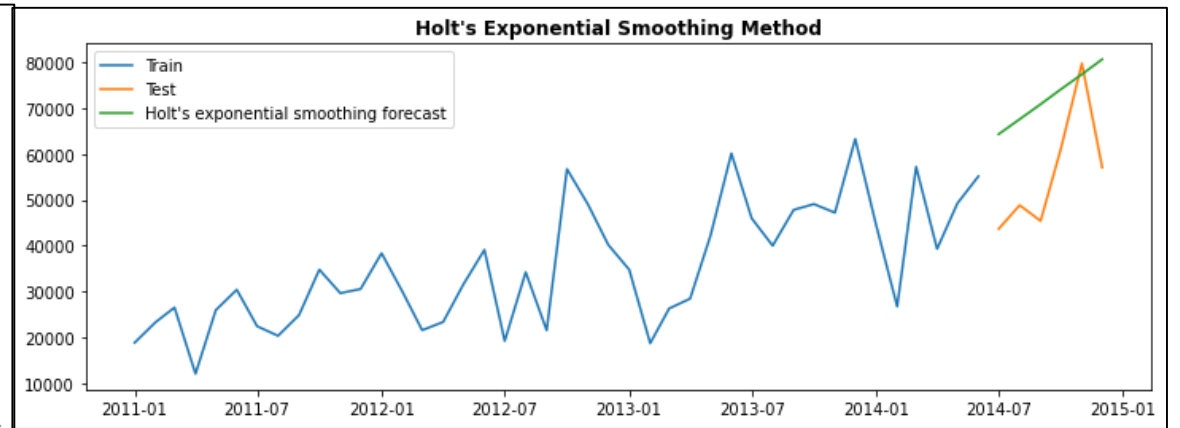
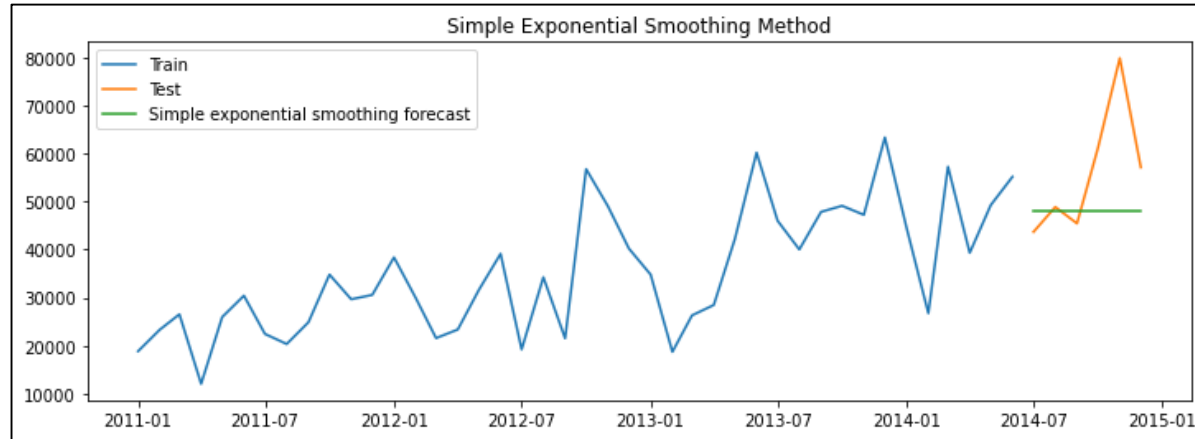
2. If the data has no exogenous variables and the level, trend and seasonality need to be captured, the Holt-Winters' exponential smoothing technique / SARIMA works best.

Hence, the flow chart makes a great sense in this case as shared in the subsequent slides.

# Time Series Model



# Time Series Model - Continued



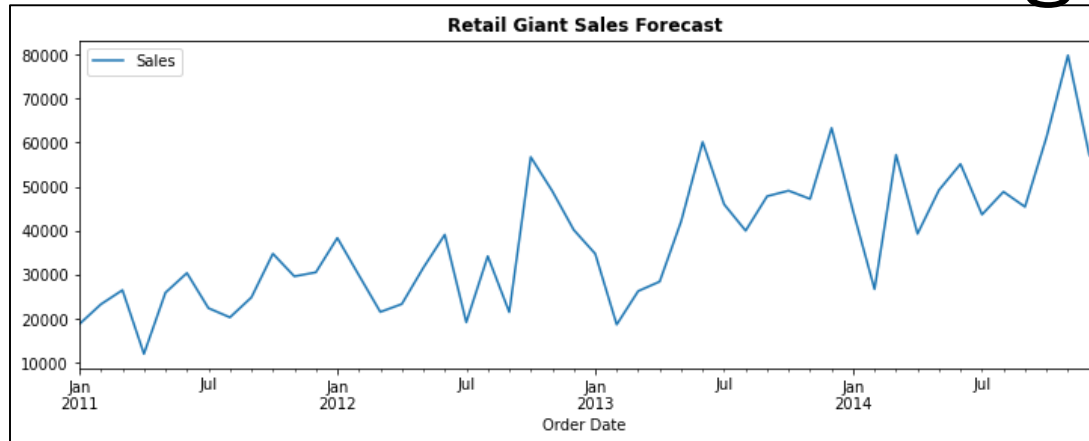


# Comparison of Sales Forecast with Smoothing Techniques

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	14627.34	15.74
0	Holt's exponential smoothing method	18976.37	34.57
0	Holt Winters' additive method	9306.82	10.17
0	Holt Winters' multiplicative method	9423.23	11.43

Out of all the smoothing techniques used,  
"Holt Winters additive method"  
is the best forecasting method with a  
MAPE value of 10.17

# Auto Regressive Methods

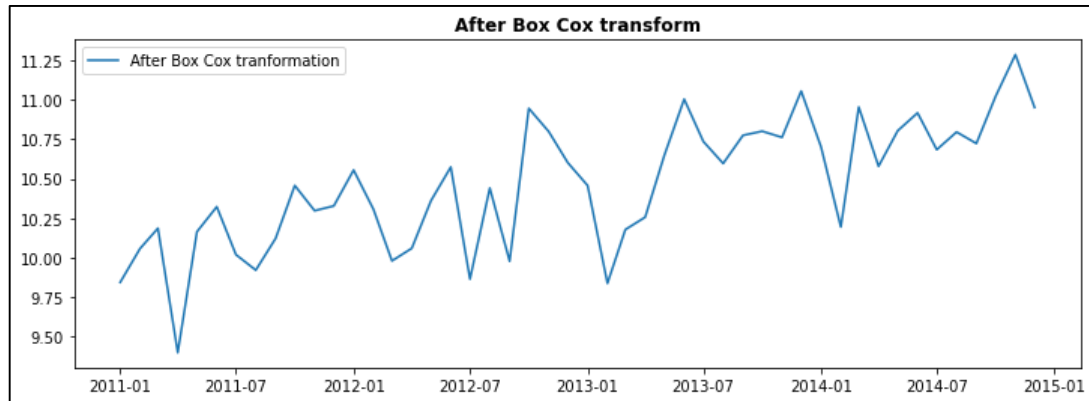


Performing ADF and KPSS it is observed that Time series data is not stationary.

ADF Statistic: -4.535011

Critical Values @ 0.05: -2.95

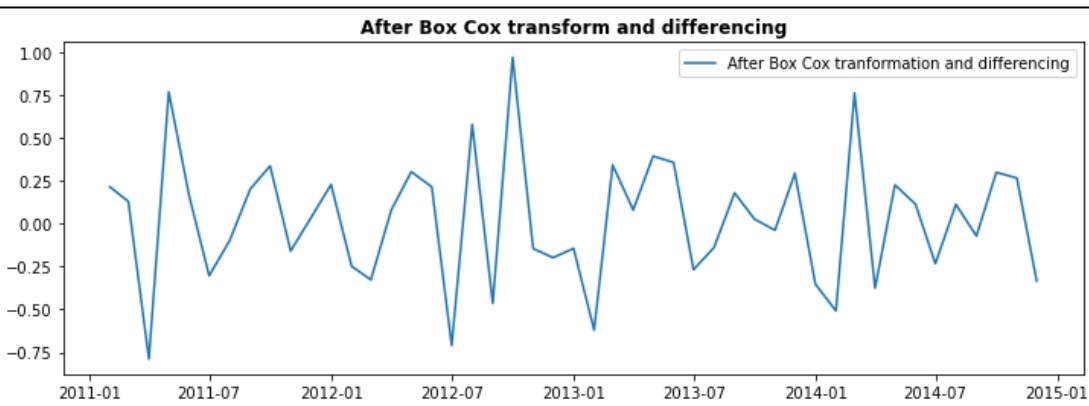
p-value: 0.000170



KPSS Statistic: 0.156456

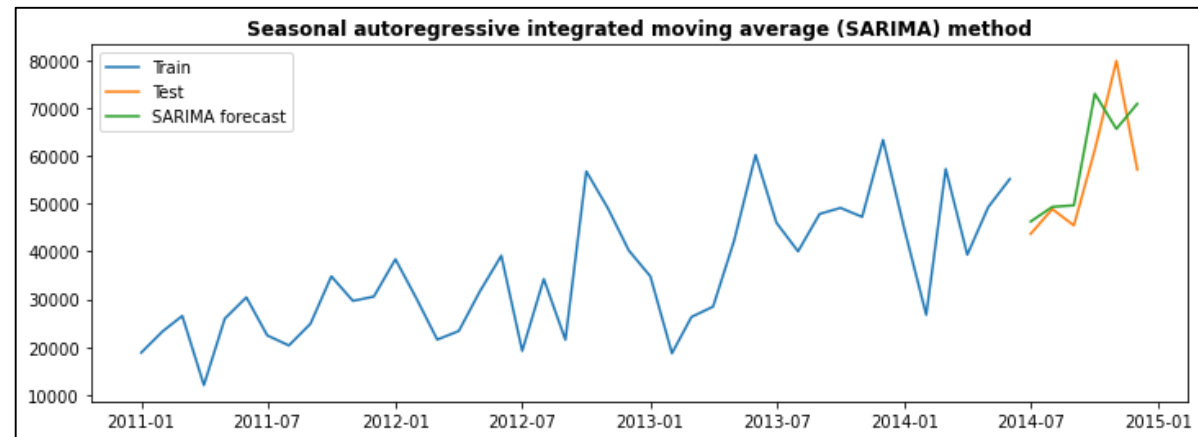
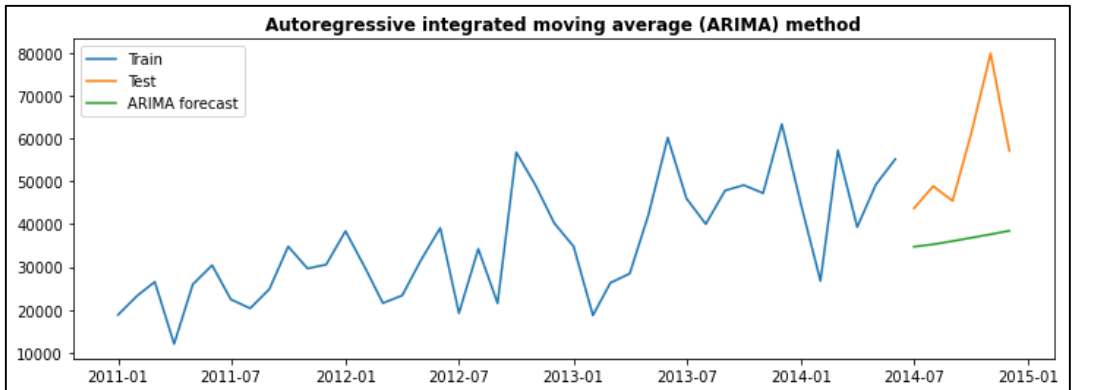
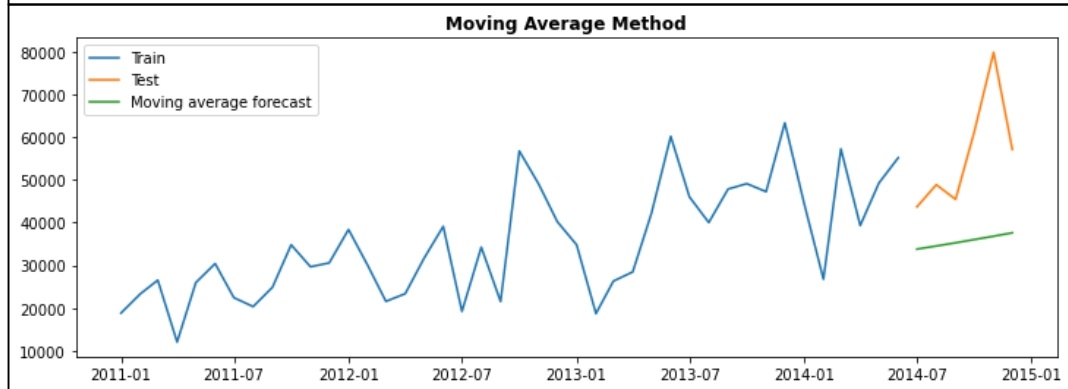
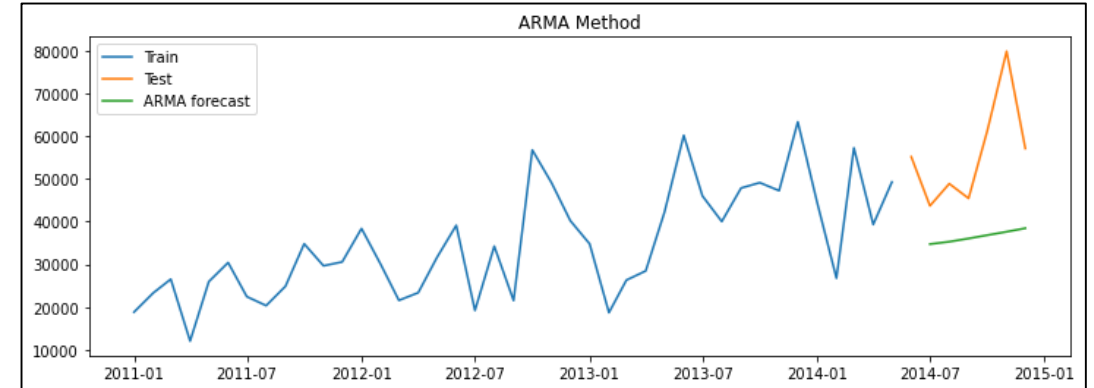
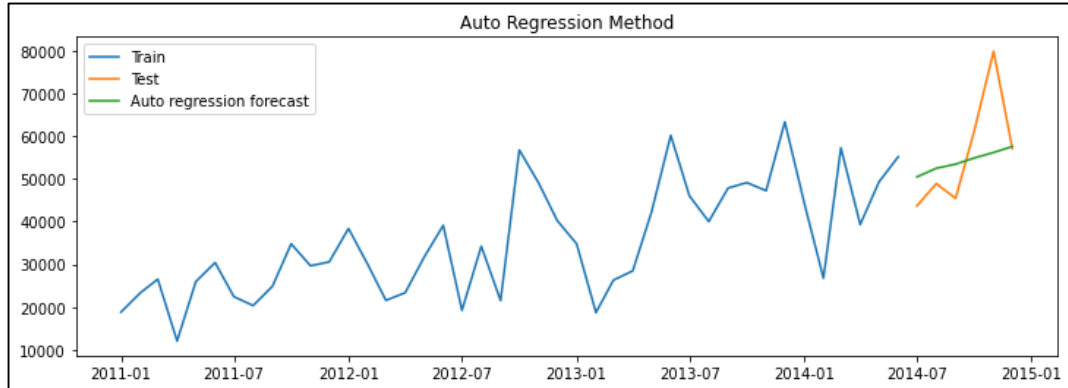
Critical Values @ 0.05: 0.46

p-value: 0.100000



Hence, needed to do Box cox and differentiating to make the data stationary.

# Auto Regression Method



# Comparison of Sales Forecast with ARIMA Techniques

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	14627.34	15.74
0 Holt's exponential smoothing method	18976.37	34.57
0 Holt Winters' additive method	9306.82	10.17
0 Holt Winters' multiplicative method	9423.23	11.43
0 (AR) Auto Regressive method	10985.28	13.56
0 (MA) Moving Average method	23360.02	33.93
0 (ARMA) Auto Regressive MA method	22654.32	32.40
0 (ARIMA) AR Integrated MA method	22654.32	32.40
0 (SARIMA) Seasonal AR Integrated MA method	9616.66	12.87

Out of all the ARIMA techniques used,  
"SARIMA - Seasonal Auto Regressive Integrated Moving Average" is the best forecasting method with a  
MAPE value of 12.87

Method	RMSE	MAPE
Naive method	12355.97	17.47
Simple average method	24146.06	34.34
Simple moving average forecast	14756.73	15.82
Simple exponential smoothing forecast	14627.34	15.74
Holt's exponential smoothing method	18976.37	34.57
Holt Winters' additive method	9306.82	10.17
Holt Winters' multiplicative method	9423.23	11.43
(AR) Auto Regressive method	10985.28	13.56
(MA) Moving Average method	23360.02	33.93
(ARMA) Auto Regressive MA method	22654.32	32.40
(ARIMA) AR Integrated MA method	22654.32	32.40
(SARIMA) Seasonal AR Integrated MA method	9616.66	12.87

- Smoothing techniques: "Holt Winters additive method" is the best forecasting method with a MAPE value of 10.17
- ARIMA techniques: "SARIMA - Seasonal AR Integrated Moving Average" is the best method with MAPE value of 12.87

# Conclusion

