Retail-Giant Sales Forecasting Case Study

Submitted by:

Group Name: Team Insight

1. Kaushal Kashyap
2. Manjiri Paranjape
3. Snehil Gupta
4. Ravinder Gill

Business Understanding

- ***Global Mart***, an online store super giant having worldwide operations.
- Delivers across the globe and deals with all the major product categories -Technology, Furniture, Office Supplies
- The store caters to 7 different markets namely Africa, APAC, Canada, EMEA, EU, LATAM, US
- Deals with 3 major customer segments – Corporate, Consumer and Home Office

Problem and Goal Analysis

Problem:

- As operations manager, need to forecast the sales and the demand for the next 6 months.
- Needed to manage the revenue and inventory accordingly.

Goal:

- Need to find out 2 most profitable (and consistent) segment from these 21.
- Forecast the sales and the demand for these segments for the next 6 months.
- To forecast the future sales and demand, by using time stamped data.
- Create model using Classic decomposition and Auto ARIMA method.
- Come up with the most suitable model for the segment.

Data Understanding

Data Source: Data is provided in two ***.csv file*** format:

1. Global Superstore → Transaction level data with 24 attributes related to each such transaction.
2. Data Dictionary → File containing the meaning and explanations for 24 attributes.

Data Understanding:

1. Market attribute has 7-factor levels representing the geographical market sector that the customer belongs to.
2. The "Segment" attribute tells which of the 3 segments that customer belongs to.
3. Profit
4. Quantity
5. Sales
6. Order date

Methodology

Analysis Type	Operations Performed	Methodology/ Tools used
Data Preparation	• Data Preparation: First convert the Date in to uniform format • Segment the data in to market and Segment • Aggregate the segmented data based on Order Date Market and Segment	▪ R - studio
Modelling	• Two Most profitable segments obtained are then modelled into time series. • Two types of modelling techniques are used: Classical Decomposition, ARIMA Modelling • For Classical Decomposition smoothing techniques are used.	▪ ADF ▪ KPSS ▪ ACF and PACF ▪ ARMA ▪ Auto ARIMA
Model Evaluation	• Model is then evaluated based on MAPE value	▪ MAPE

Data Preparation and Cleaning

For Data cleaning and preparation following steps were performed in sequence.

1. First off all check for NA and duplicate values: NA values are only present in PINCODE columns.
2. Convert the date in to correct format and eliminate the date value keeping only month and year.
3. Segment the data based on Market, Segment, Order date along with Profit, Sales and Quantity.
4. Aggregate the data on bases of Order date and arrange the data in ascending order.
5. Aggregate another data set based on Market and Segment and calculate the COV and arrange the data is ascending order of Profit values.
6. This will give the two most profitable segment.

Tools Used:

- R - language

COV for Data Selection

	Market	Segment	sum_profit	sum_quantity	sum_sales	cv
1	APAC	Consumer	222817.560	21414	1816753.70	4.206702
2	EU	Consumer	188687.707	19541	1529716.24	4.718084
3	US	Consumer	134119.209	19521	1161401.34	9.389450
4	APAC	Corporate	129737.235	12142	1078466.31	4.231301
5	EU	Corporate	123393.980	11635	920008.28	4.776482
6	LATAM	Consumer	120632.932	19853	1133847.03	5.438845
7	US	Corporate	91979.134	11608	706146.37	7.616929
8	APAC	Home Office	83445.254	7670	690524.11	4.633390
9	EU	Home Office	60748.054	6597	488364.54	4.923759
10	US	Home Office	60298.679	6744	429653.15	6.280008
11	LATAM	Corporate	57875.421	11499	645252.48	5.789517
12	Africa	Consumer	47772.099	5503	423766.81	7.351006
13	LATAM	Home Office	43135.134	7174	385505.65	5.336331
14	EMEA	Consumer	25532.574	5871	406745.18	14.441103
15	Africa	Corporate	20686.965	2943	204938.95	9.334133
16	Africa	Home Office	20412.567	2118	155067.45	6.264113
17	EMEA	Corporate	12499.134	3515	250571.39	22.038317
18	Canada	Consumer	9677.700	454	35719.11	2.093766

- Coefficient of Variation (COV) of the Profit, is used for all 21 market segments.
- Based on COV factor the 2 most profitable segments are chooses.
- The two most Profitable Segments are:
- APAC Consumer → COV = 4.2
- EU Consumer → COV = 4.7
- For these two segment time series are created using Classical decomposition and Auto ARIMA method.

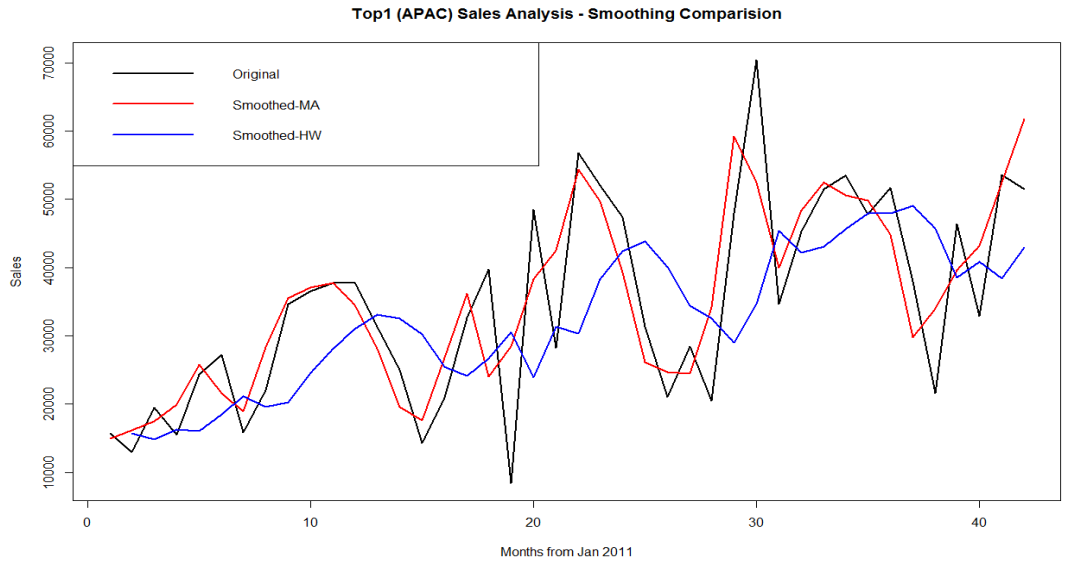
Modelling the Time Series

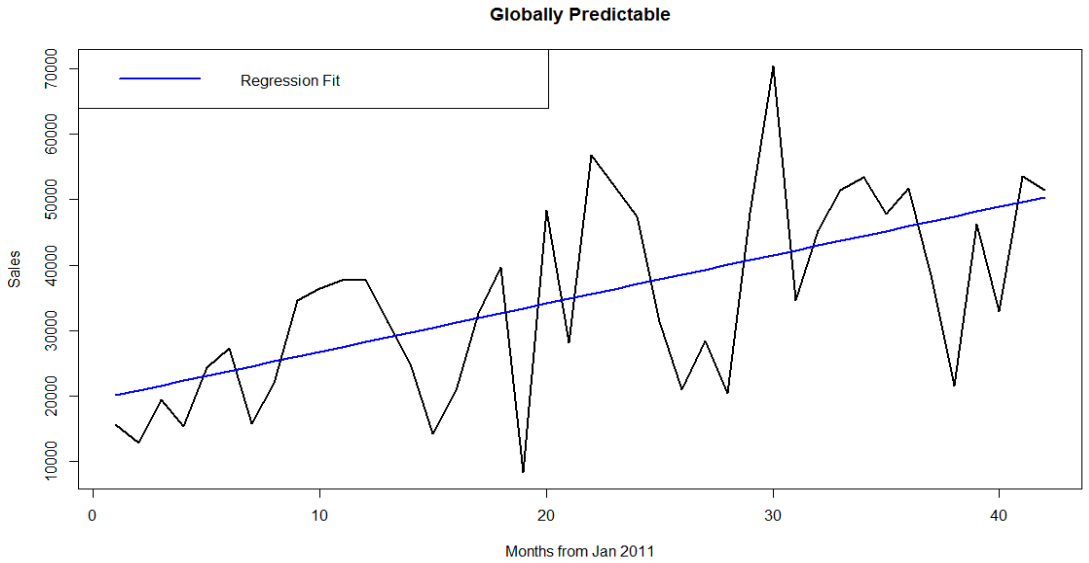
1. Based on COV of the Profit for all 21 market segments the 2 most profitable segments are obtained.
2. The two 2 most profitable and consistently segments are:
    1. APAC Consumer → Sales and Quantity
    2. EU Consumer → Sales and Quantity
3. Forecast the sales and quantity for the next 6 months.
4. Technique used for forecasting
    1. Classical decomposition
    2. Auto ARIMA for forecasting.
5. For Classical decomposition method smoothing of series is done using
    1. Moving Average Smoothing
    2. Holt – Winters method of exponential smoothing
6. While modelling various techniques are used to check the presence of White noise and stationarity.
    1. ADF
    2. KPSS

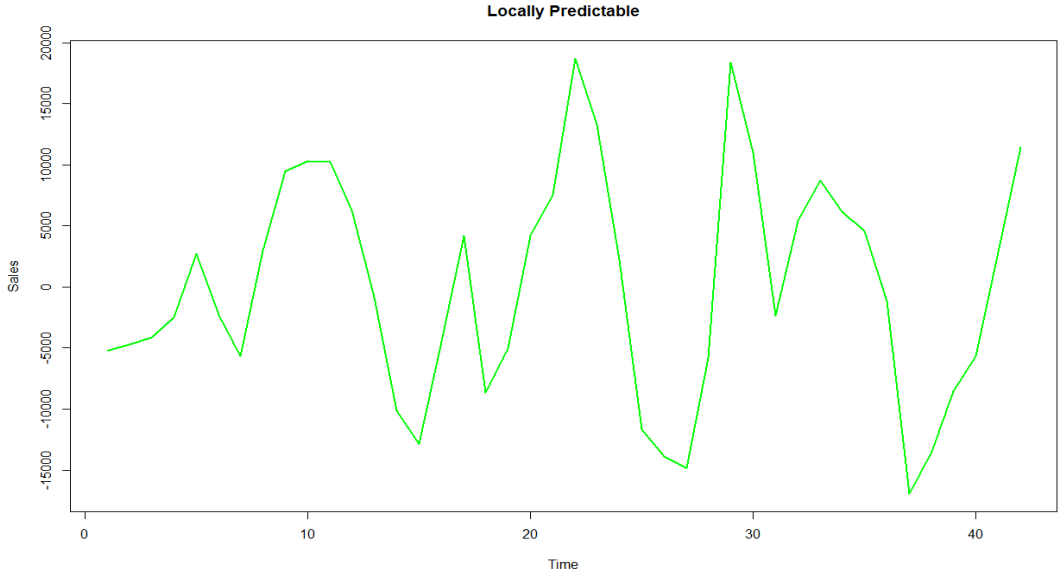
Model Evaluation

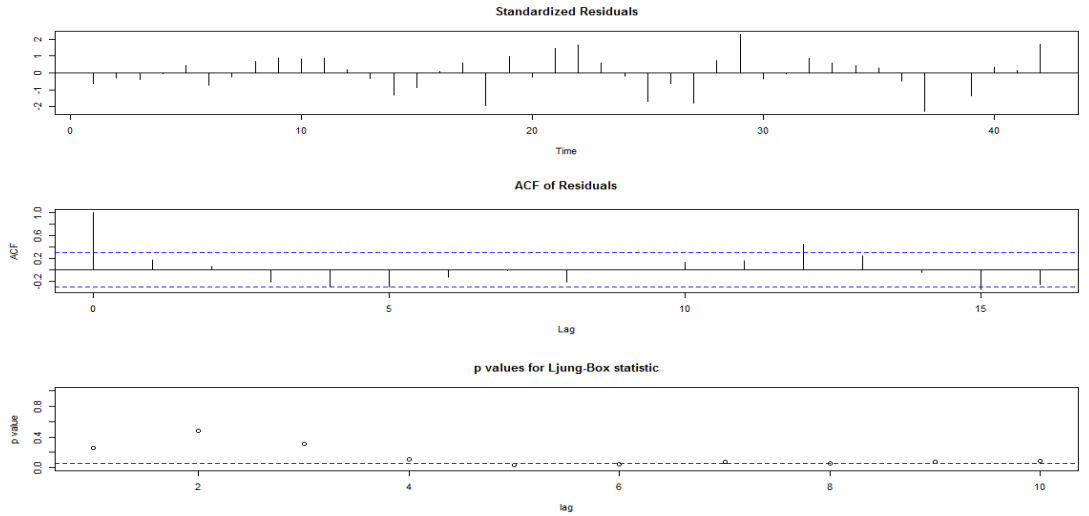
1. During modelling total 8 time series are prepared.
    1. APAC Consumer Sales
    2. APAC Consumer Quantity
    3. EU Consumer Sales
    4. EU Consumer Quantity
2. All the above mentioned 4 time series are modelled using two techniques Classical decomposition and Auto ARIMA. This create 8 time series for future forecast.
3. To test the accuracy of the forecasted value, last 6 months values were separated out the from initial dataset, after aggregating the transaction level data into the monthly data.
4. Then next 6 months forecast was done using the out-of-sample figures.
5. And the results were compared based on MAPE values.

Modelling Top1Sales i.e. APAC Consumer Sale

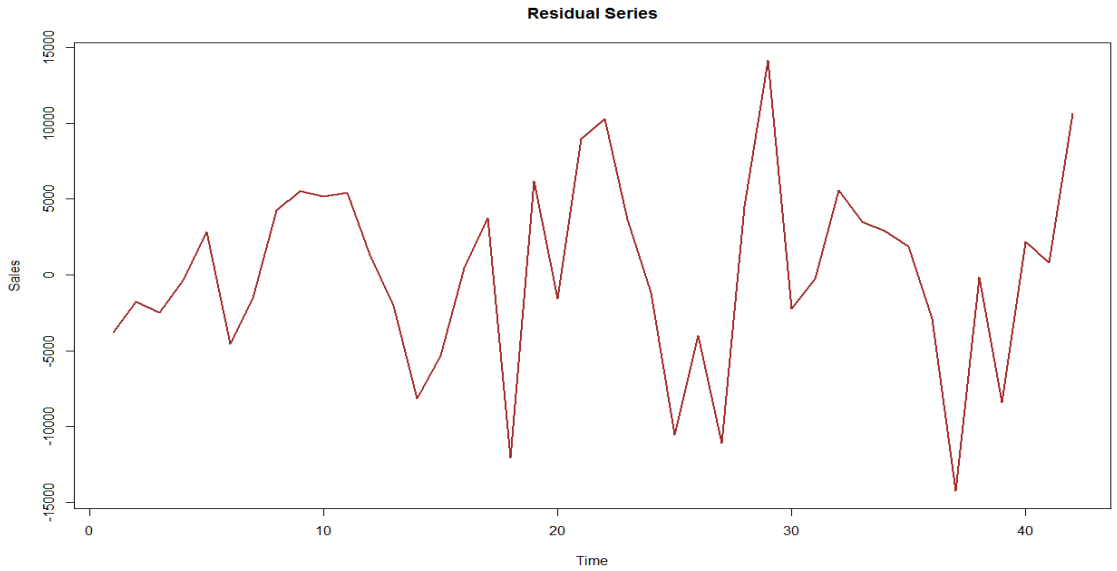


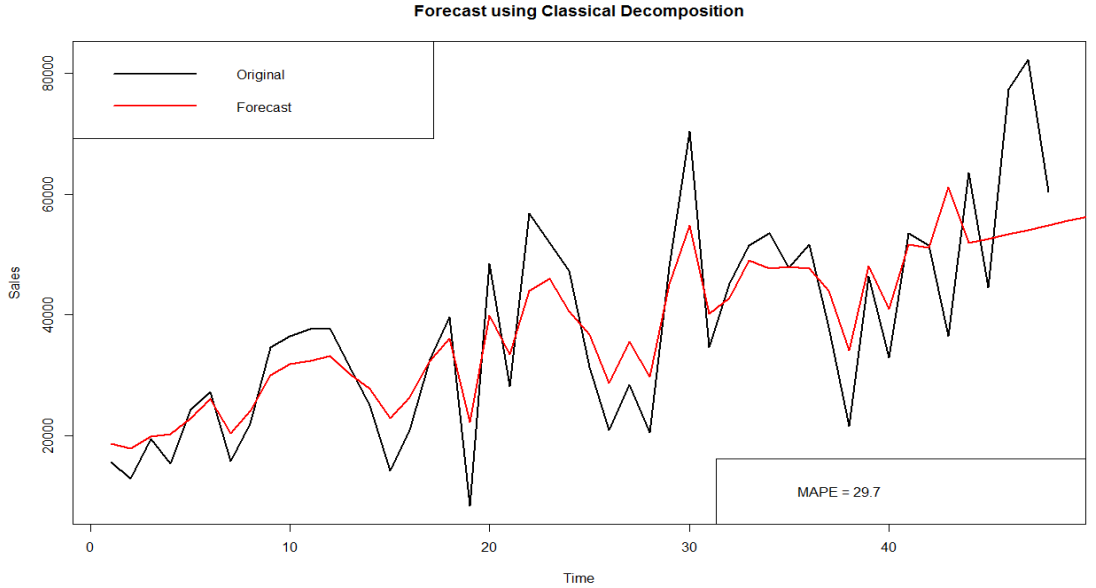


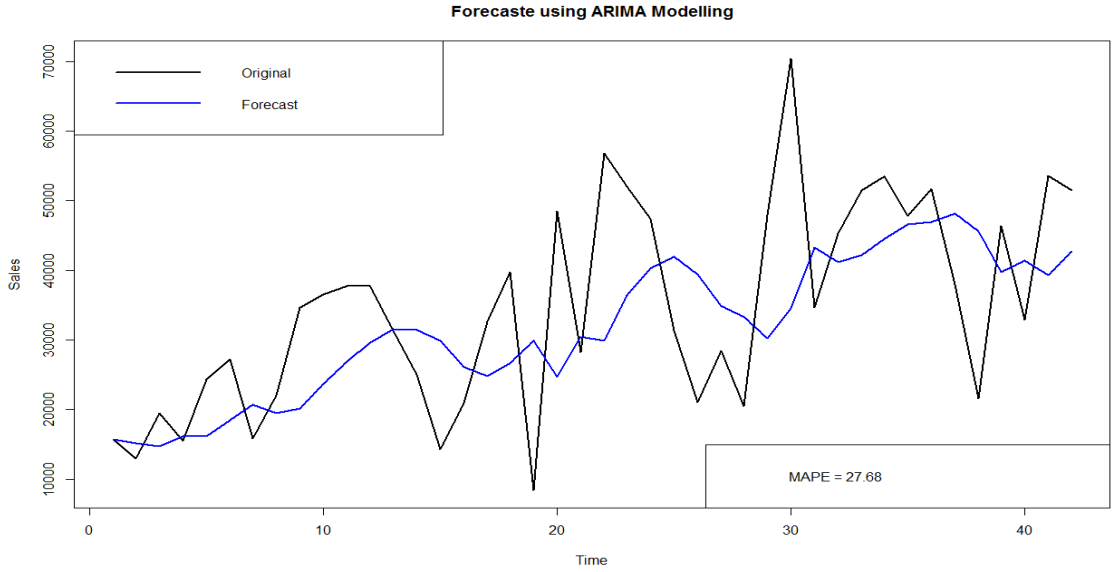


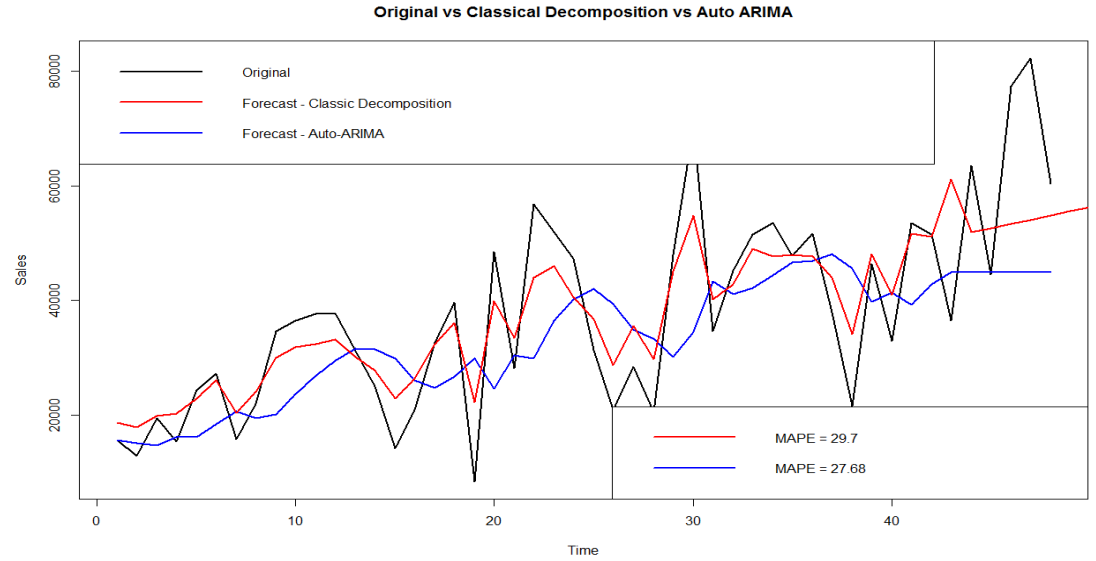


Modelling Top1Sales i.e. APAC Consumer Sale

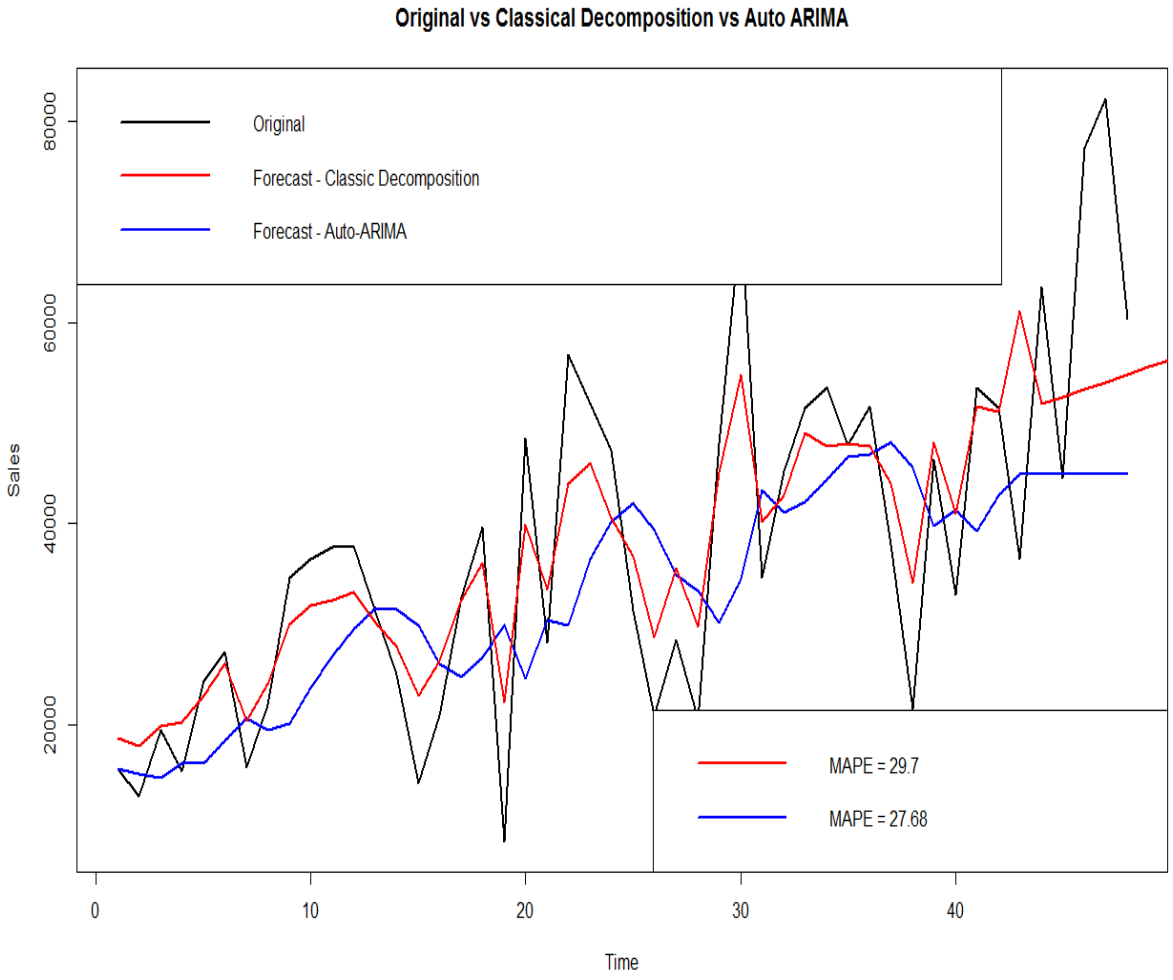








Top1Sales - Comparison
Classical Decomposition and ARIMA Method

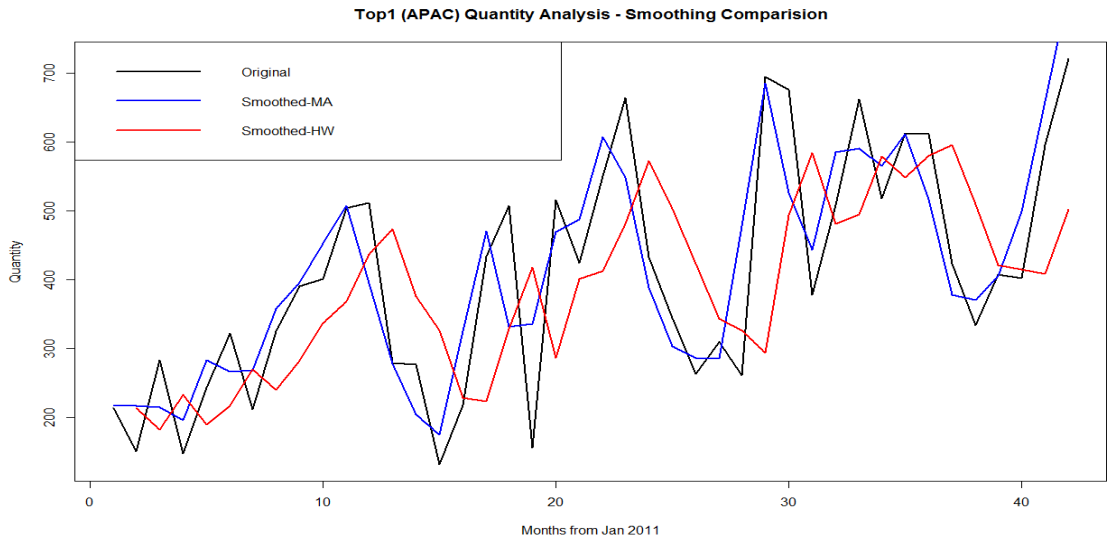


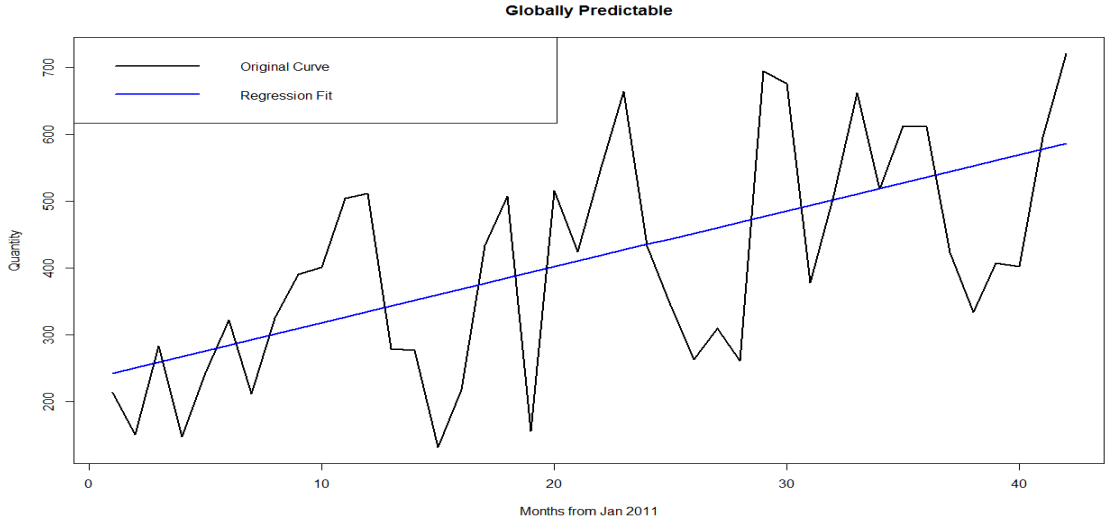
1. Classical Decomposition :
    a) ARIMA: (0,0,1)
    b) MAPE: 29.7

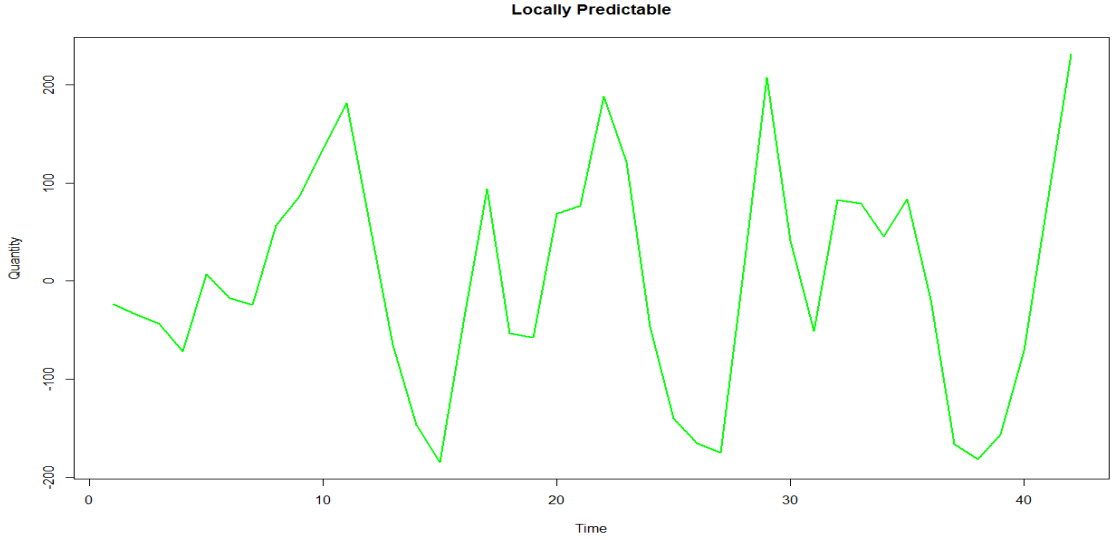
2. Auto ARIMA Method :
    a) ARIMA: (0,1,1)
    b) MAPE: 27.68

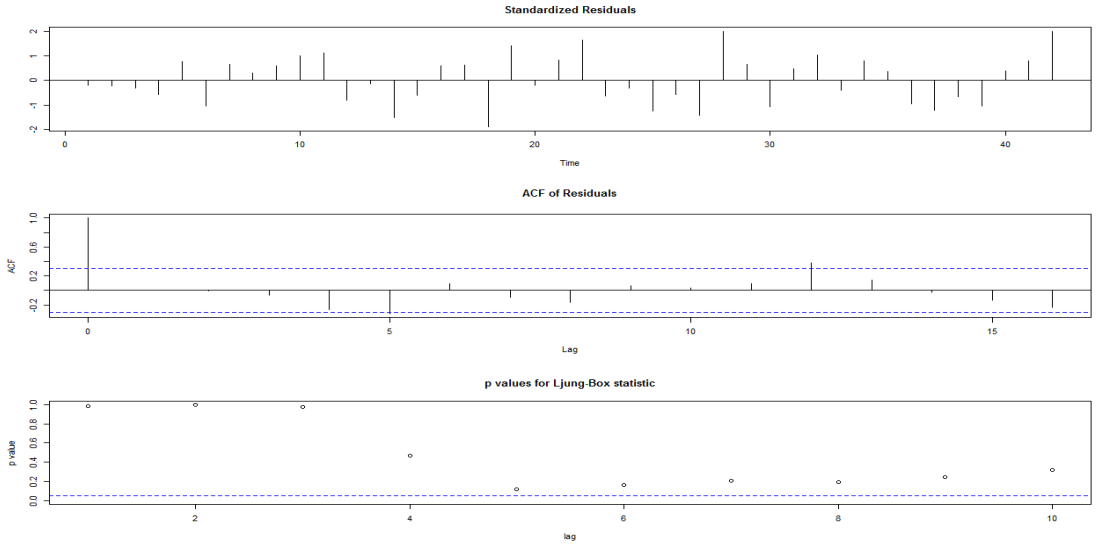
- Conclusion: Since Auto ARIMA shows low MAPE value, Hence it seems to be a good option.

Modelling Top1Qty i.e. APAC Consumer Quantity

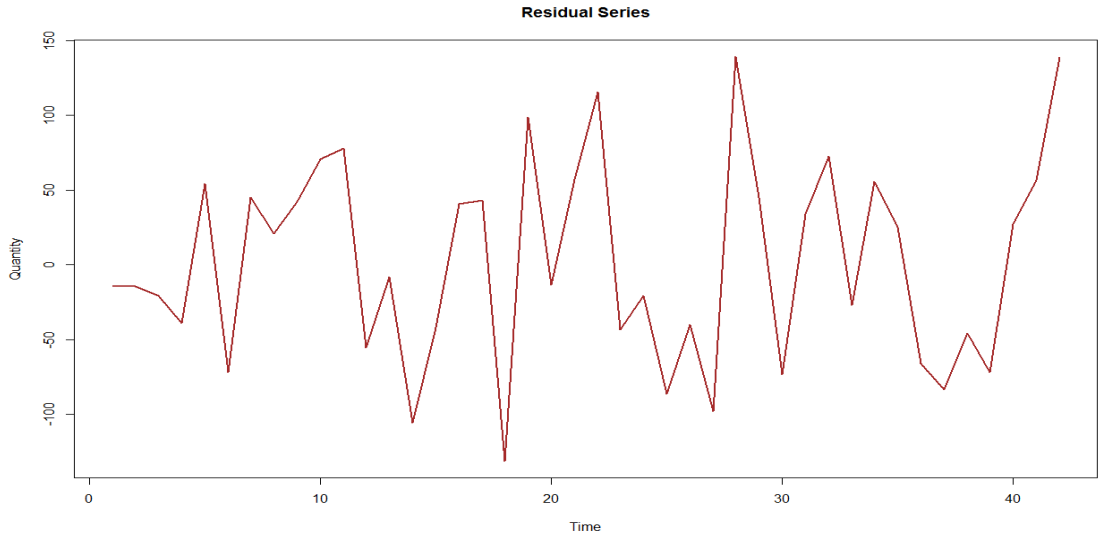


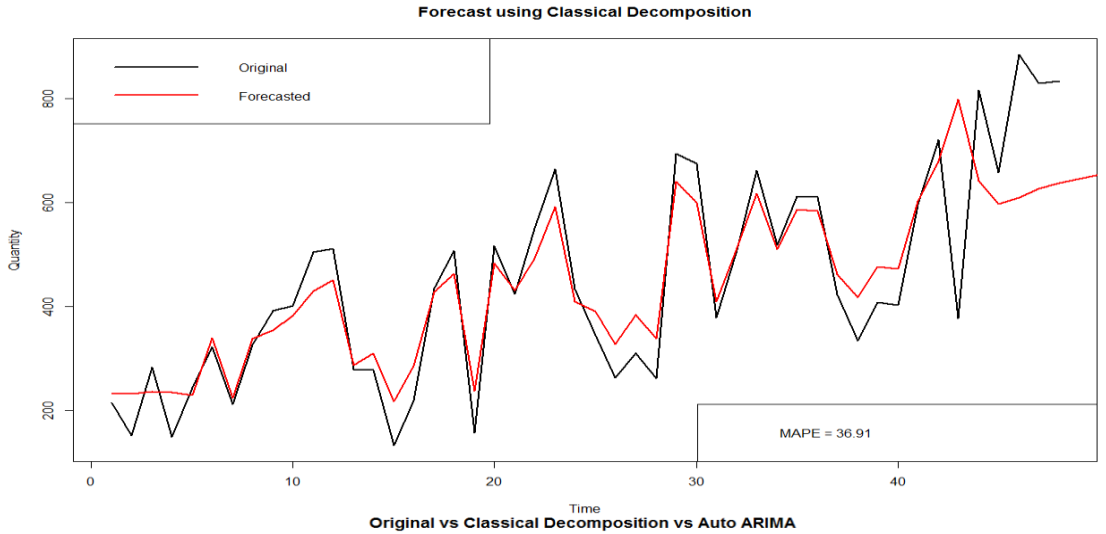


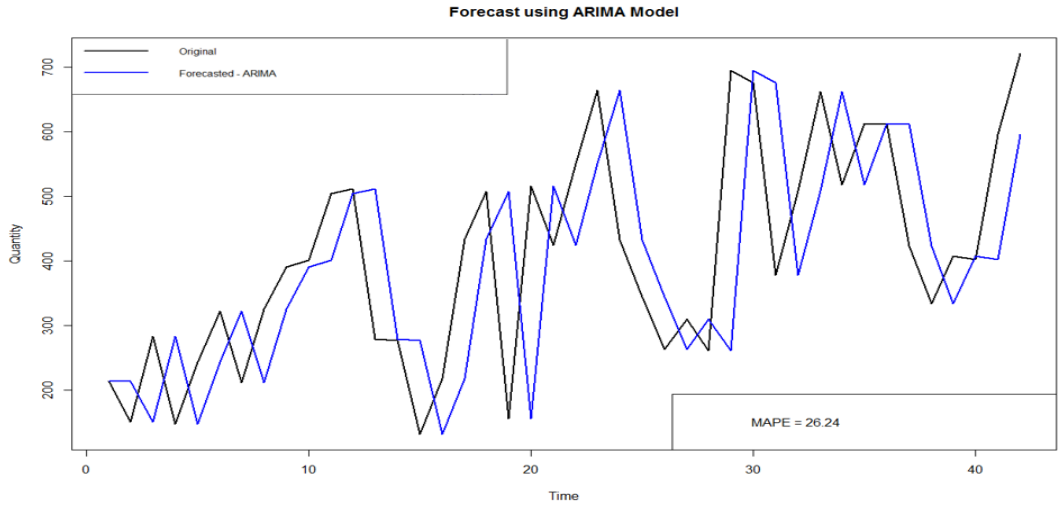


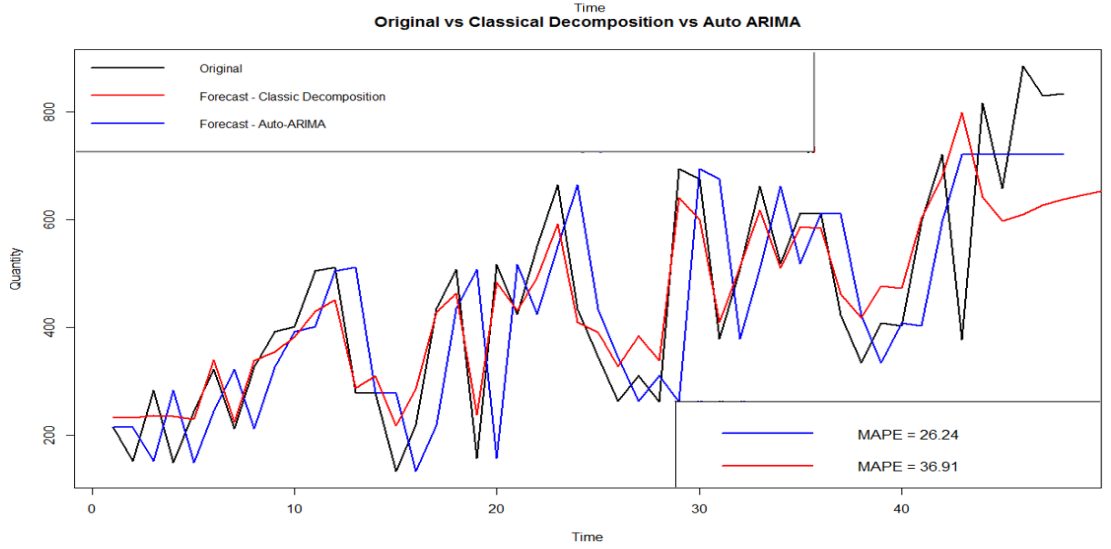


Modelling Top1Qty i.e. APAC Consumer Quantity

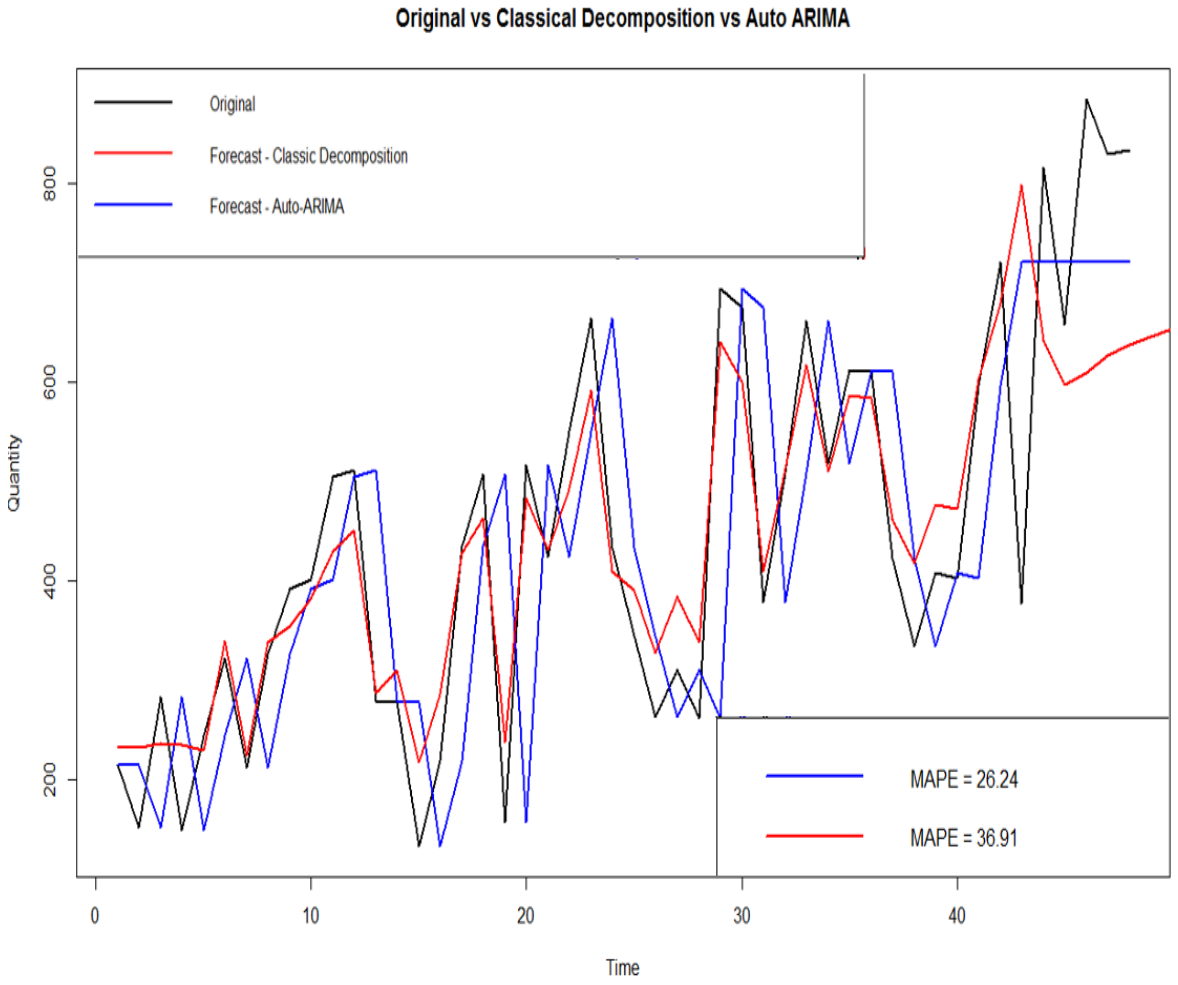








Top1Qty - Comparison
Classical Decomposition and ARIMA Method

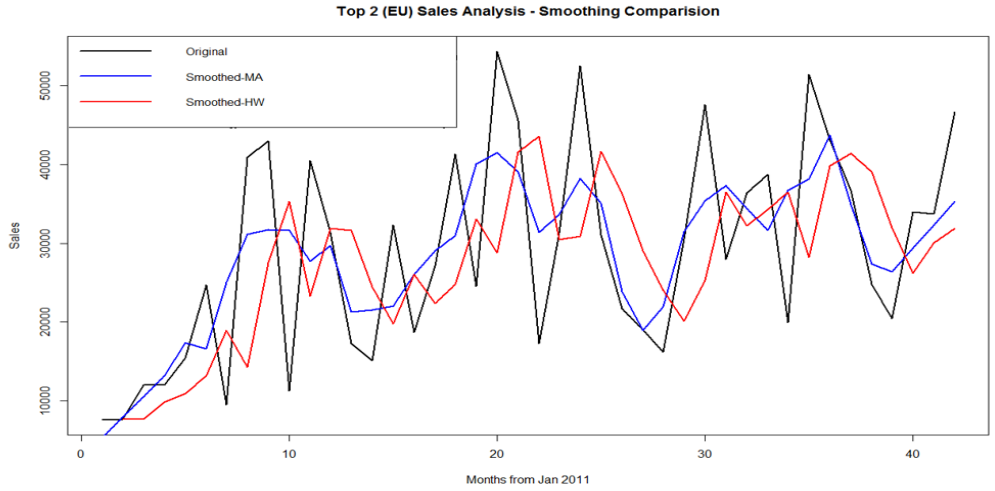


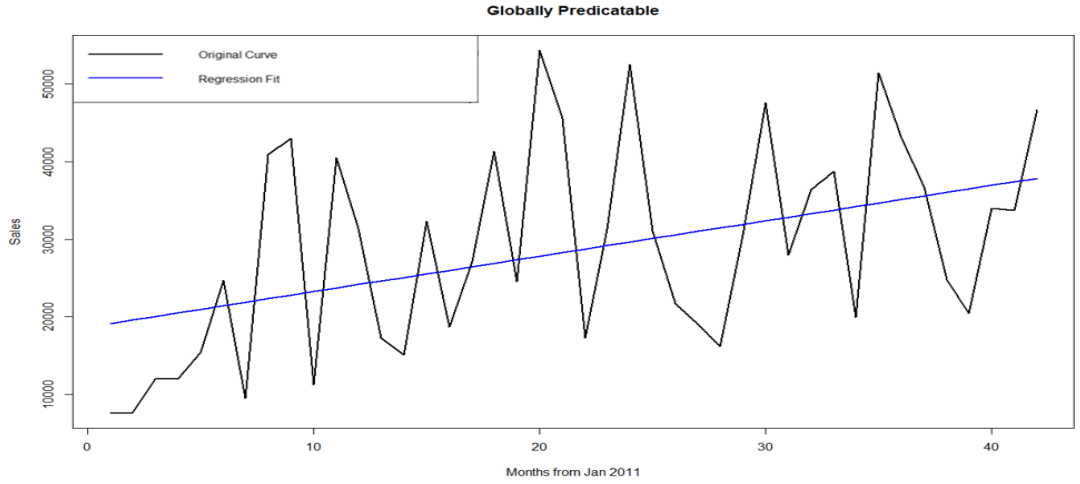
1. Classical Decomposition :
   a) ARIMA: (2,0,1)
   b) MAPE: 36.9

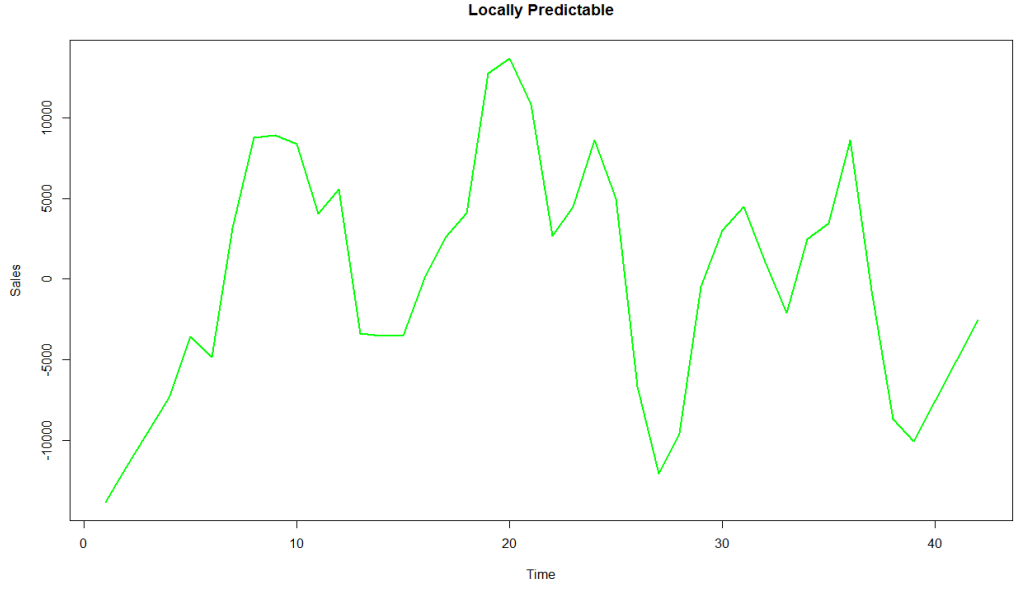
2. Auto ARIMA Method :
   a) ARIMA: (0,1,0)
   b) MAPE: 26.24

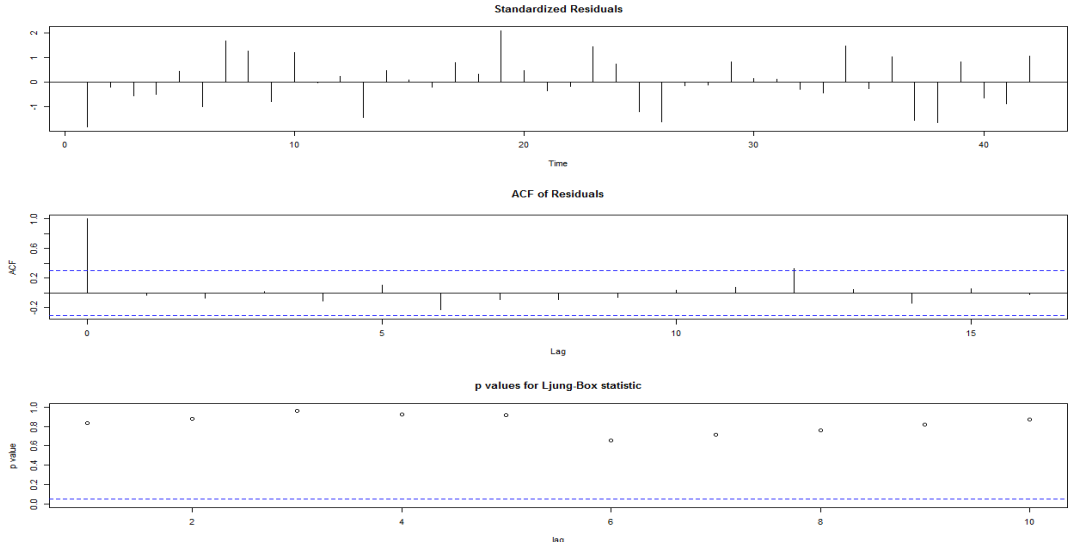
- Conclusion: Since Auto ARIMA shows low MAPE value, Hence it seems to be a good option.

Modelling Top2Sales i.e. EU Consumer Sales

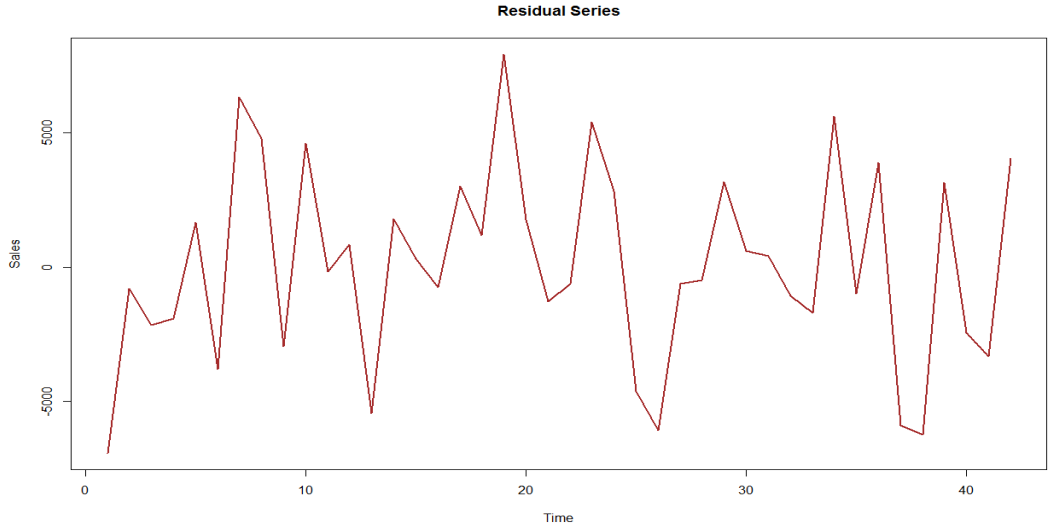


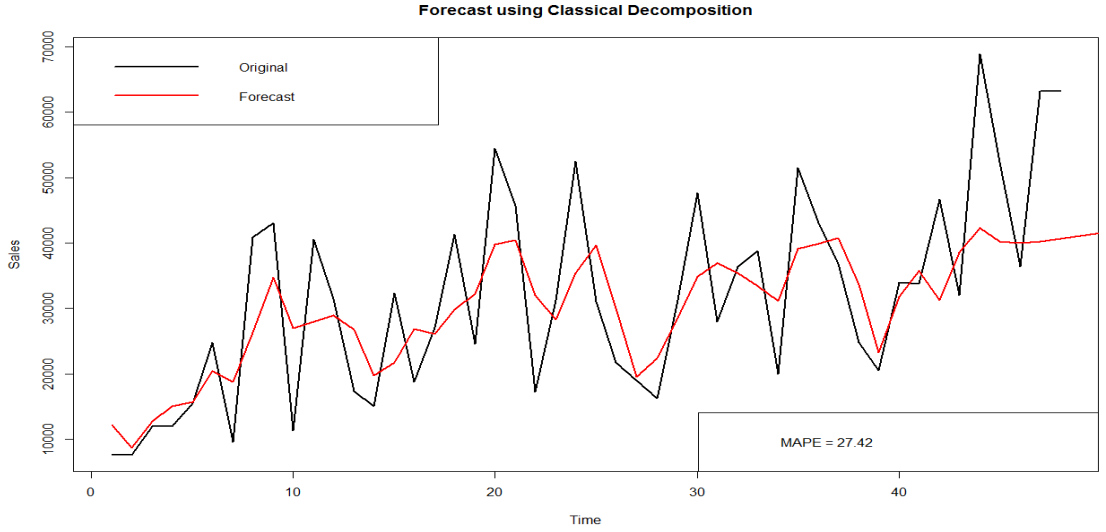


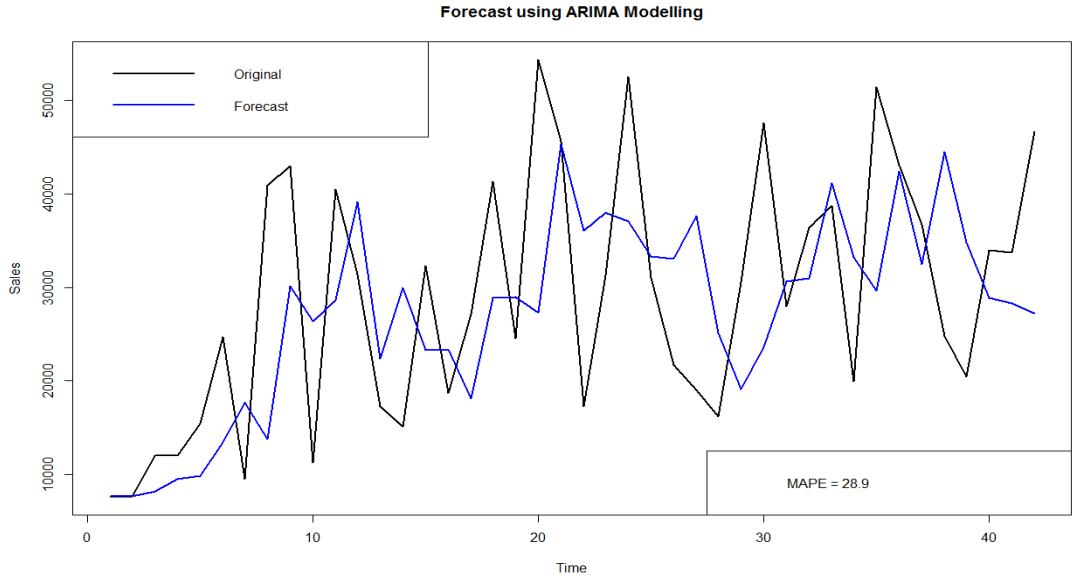


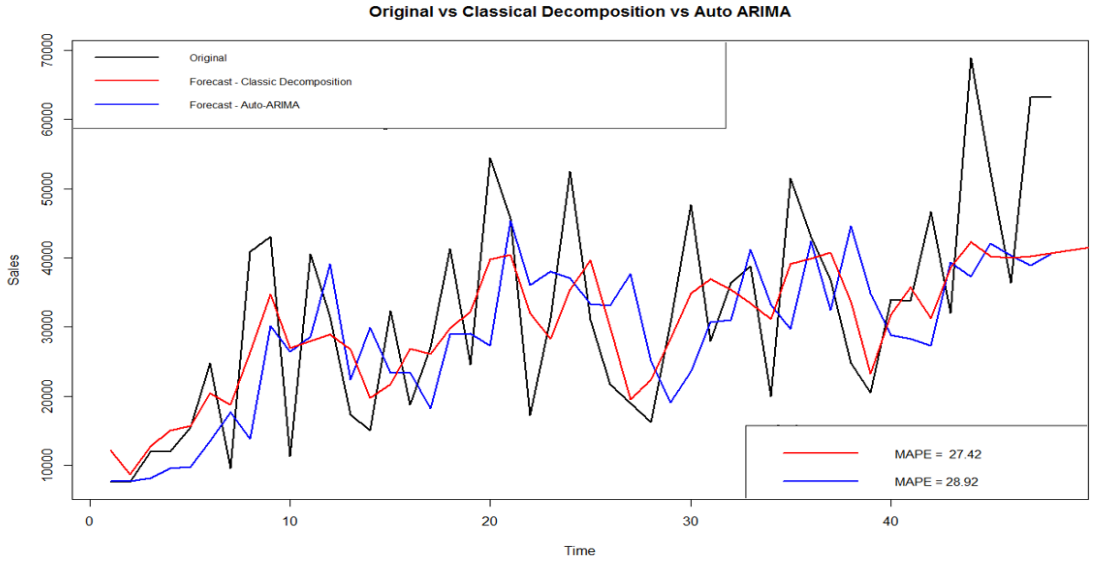


Modelling Top2Sales i.e. EU Consumer Sales

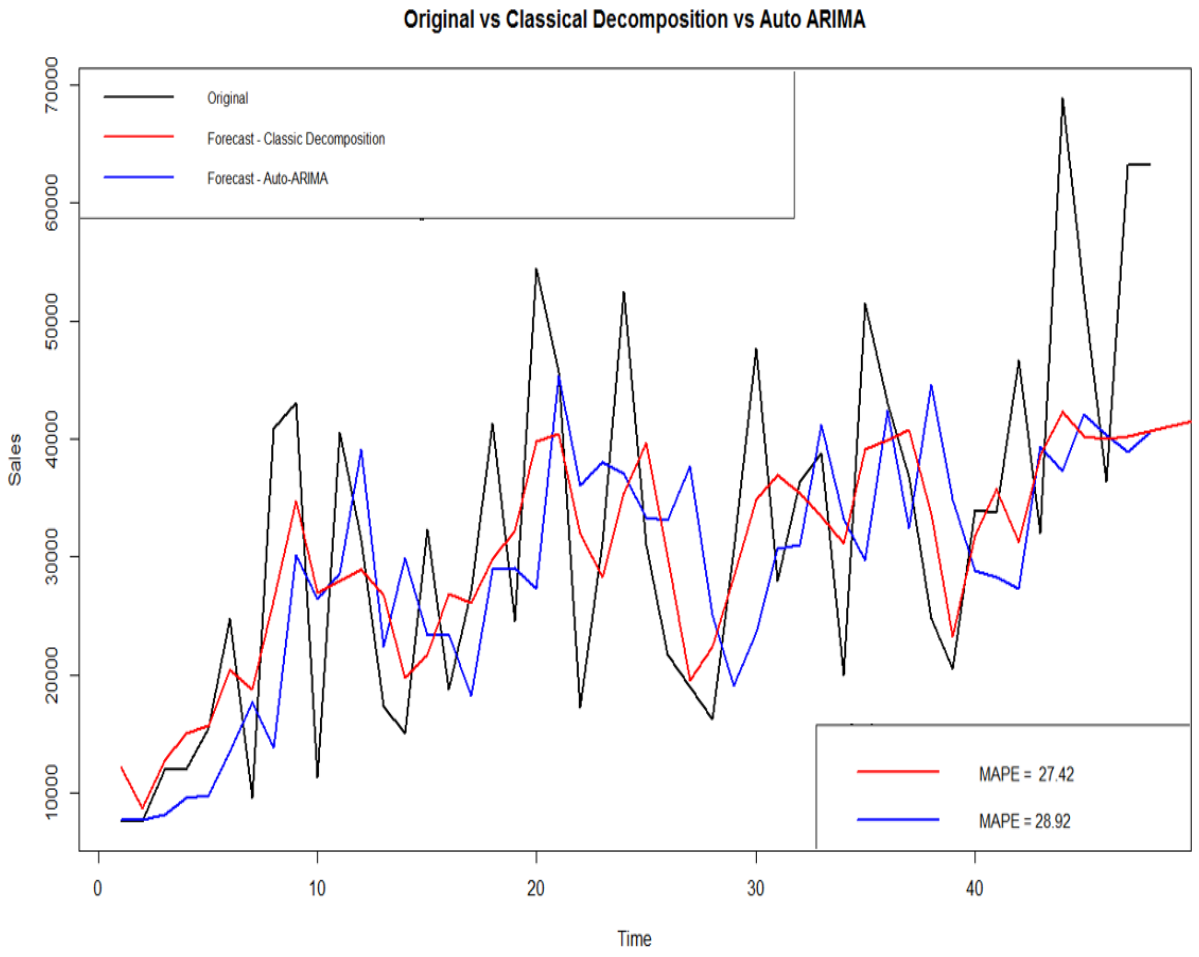








Top2Sales - Comparison
Classical Decomposition and ARIMA Method

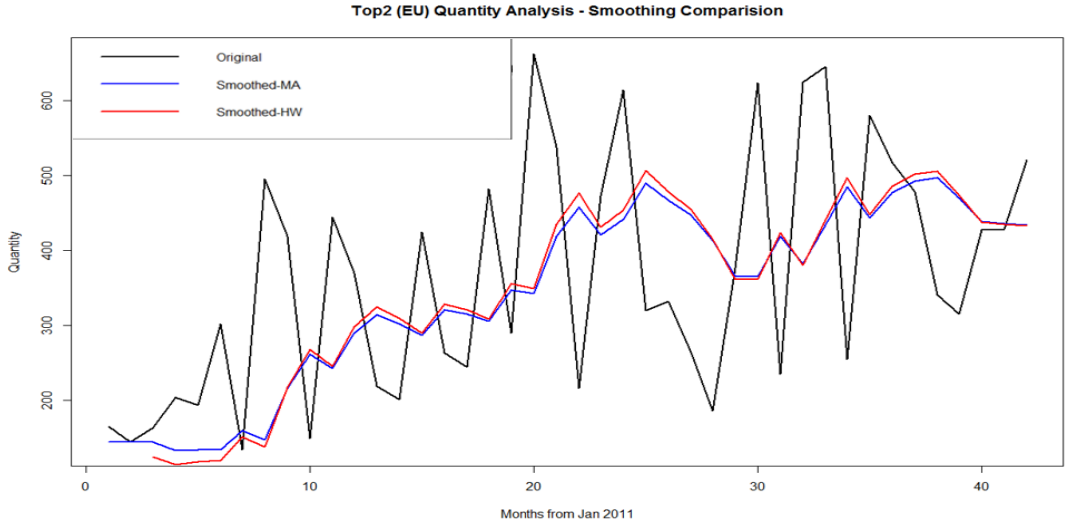


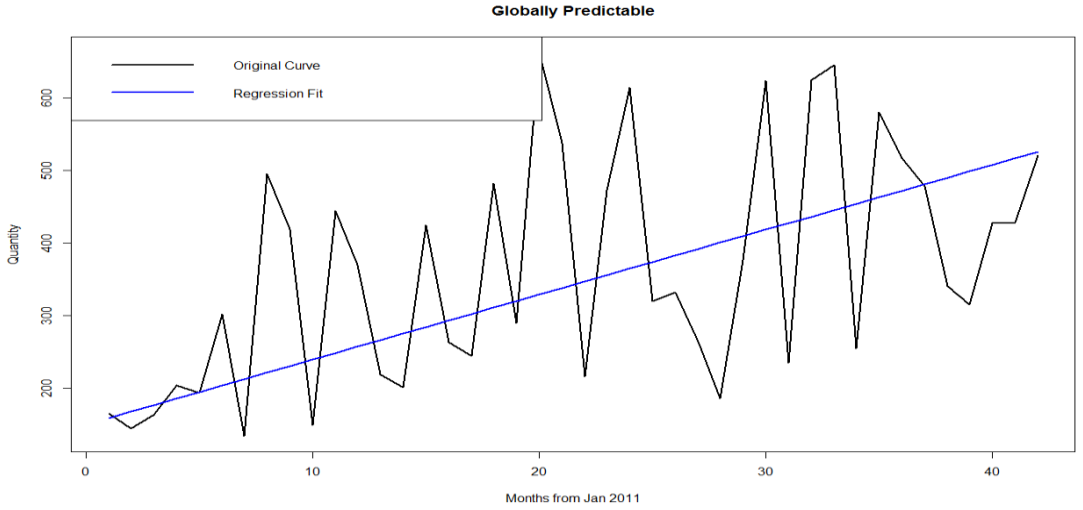
1. Classical Decomposition :
    a) ARIMA: (1,0,2)
    b) MAPE: 27.42

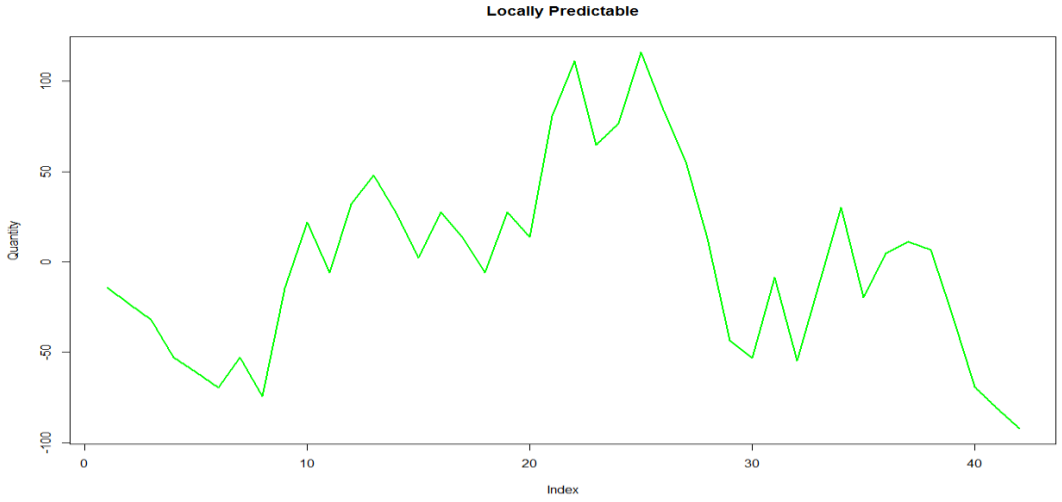
2. Auto ARIMA Method :
    a) ARIMA: (2,1,0)
    b) MAPE: 28.92

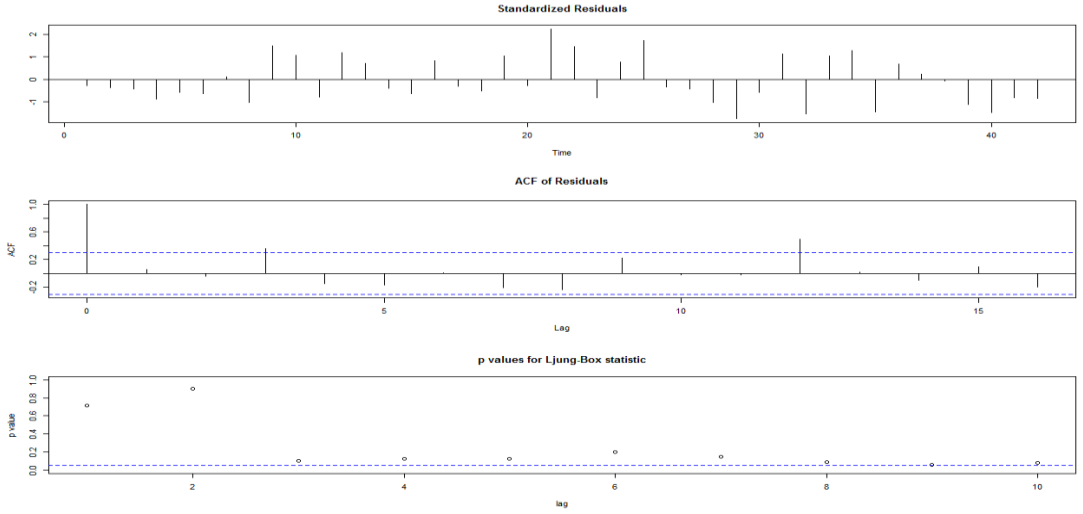
- Conclusion: Since classical decomposition shows low MAPE value, Hence it seems to be a good option.

Modelling Top2Qty i.e. EU Consumer Quantity

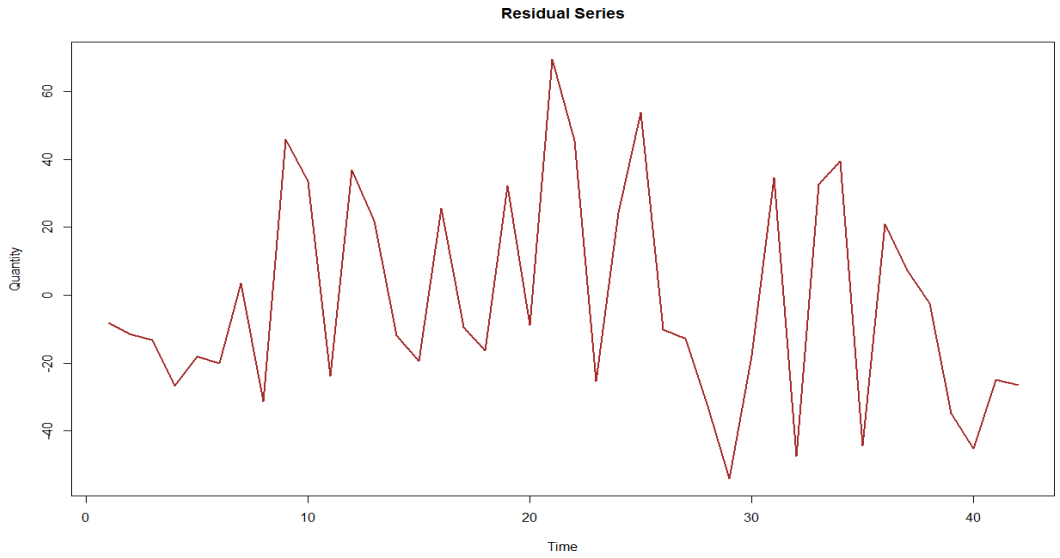


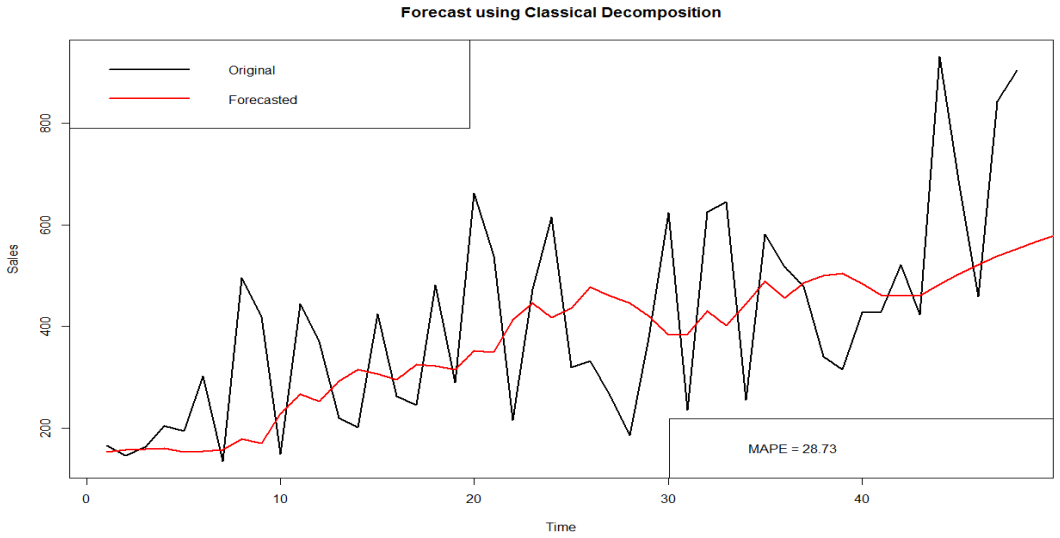


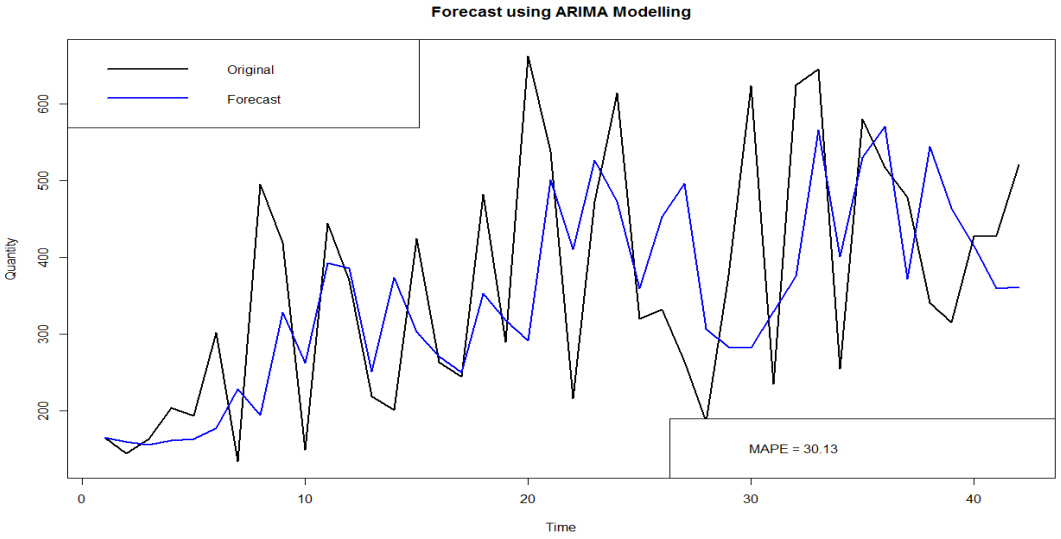


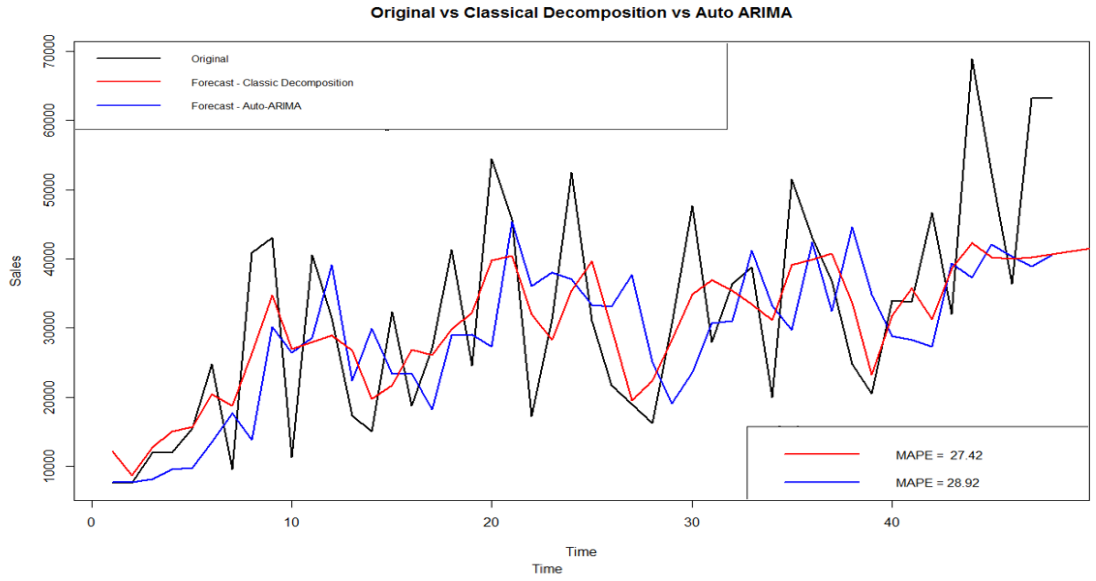


Modelling Top2Qty i.e. EU Consumer Quantity

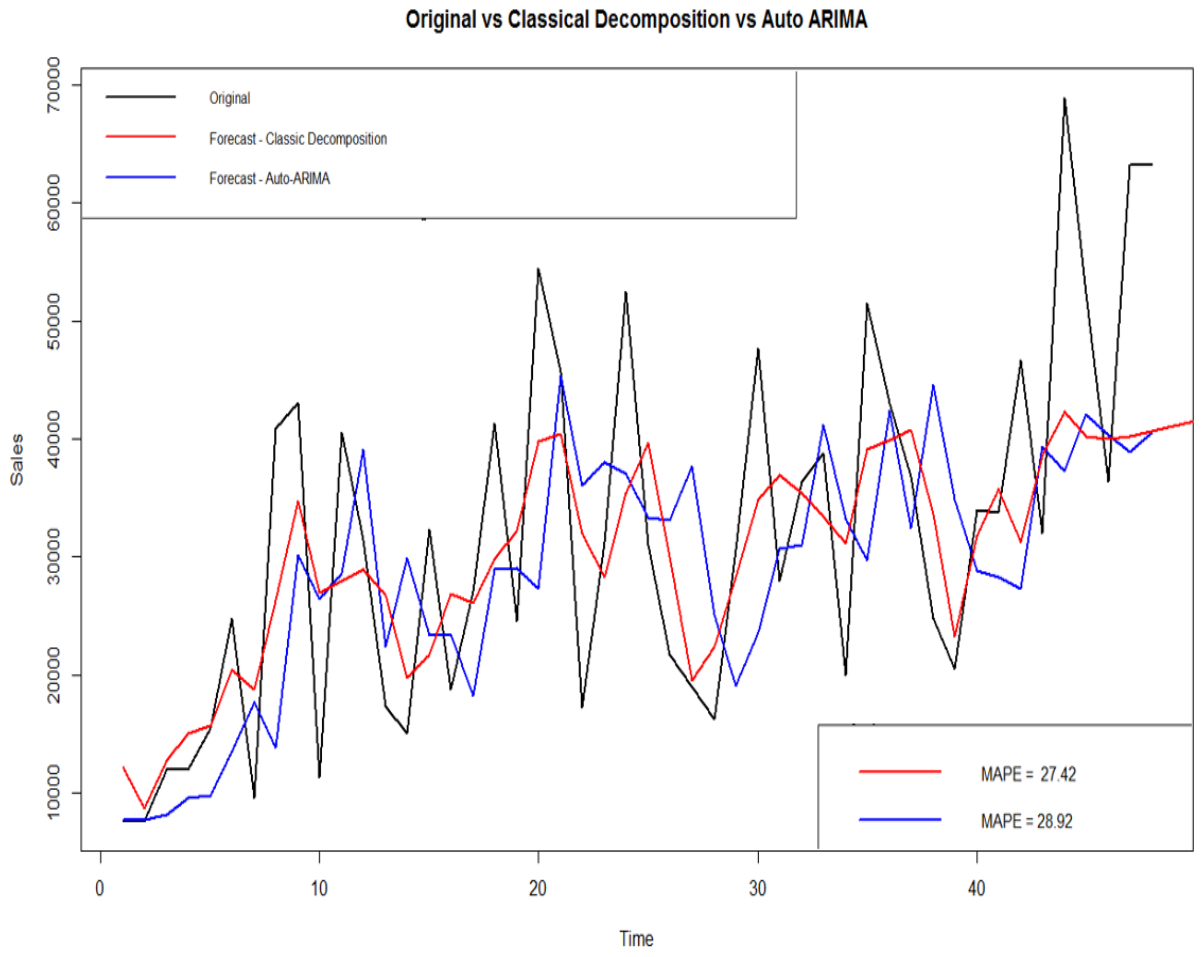








Top2Qty - Comparison
Classical Decomposition and ARIMA Method



1. Classical Decomposition :
    a) ARIMA: (1,0,0)
    b) MAPE: 28.73

2. Auto ARIMA Method :
    a) ARIMA: (2,1,0)
    b) MAPE: 30.13

- Conclusion: Since Classical decomposition shows low MAPE value, Hence it seems to be a good option.

Analysis and Conclusion

1. For APAC - Consumer:

- Both market segments i.e. Sales and Quantity, Auto ARIMA method seems to be a better choice to forecast future sales and demand, since it has low MAPE value.

Whereas

2. For EU - Consumer:

- Both market segments i.e. Sales and Quantity, Classical Decomposition comes across as a better choice to forecast future sales and demand, since it has low MAPE value.