

RETAIL GIANT SALES FORECASTING CASE STUDY

Group:DSWarriors

Group Members:

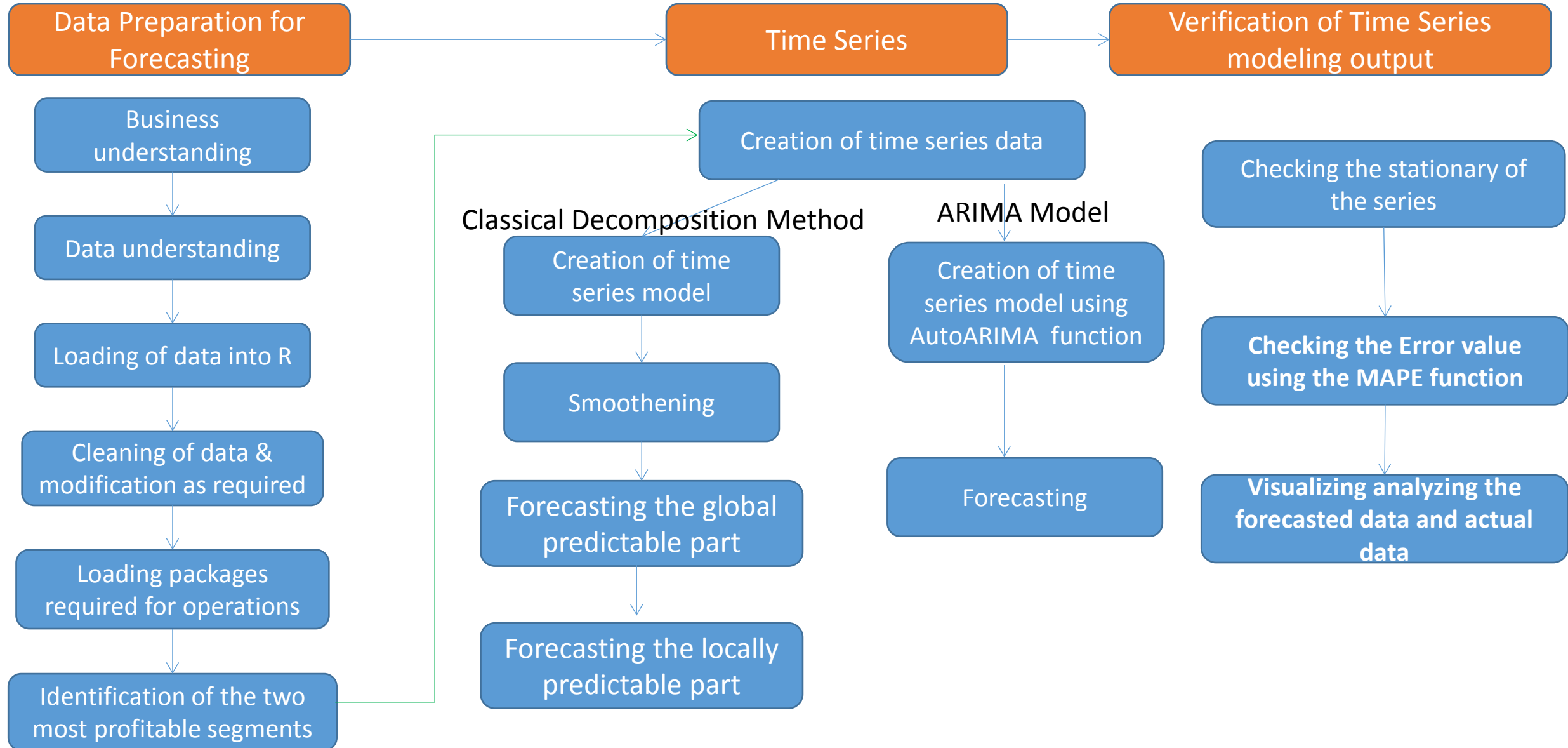
1. Sriram Pala
2. Meenakshi Lakshmanan
3. Gaurav Sachdeva
4. Parth Samanta

- The online super store giant "Global Mart" operates worldwide and takes orders and delivers across the globe
- There are 3 unique Customer Segments [1]"Consumer" [2]"Corporate" [3]"Home Office"
- There are 7 unique Market Segments as follows:
 - 1. Africa - African Continent 2. APAC - Asia Pacific/Asia Central 3. Canada – Canada 4. EMEA - Europe, Middle East and Africa 5. EU - European Union 6. LATAM - Latin America 7. US - United States of America

Business Objective of the study:

- Online giant needs to manage its revenue and inventory well for the next 6 months.
- So the The sales/operations team, want to finalize the forecasting plan for the next 6 months which will help them to that
- So the business objective is to forecast the sales for the two most consistently profitable segments for the next 6 months

Goal: Forecast the Sales and demand of the identified two consistently profitable product and marker combination segment for the next 6 months



A work flow was created as shown in the previous diagram keeping in mind the business objectives and limitations .
The first task for the team was Understanding Data.

1. Understanding Data : The following data sets were given

1. Data Set: The excel file has the data corresponding to sales, demand, order details (date, place. Etc.,)
2. Data Dictionary: Data dictionary which explain the data in the excel file

After loading the data into R, the below operations were done before going into analysis

2. Cleaning of Data:

1. Removal of the Row id which is similar to the serial number
2. Verifying any null values in the data file
3. Checking if there are any duplicate values in the data

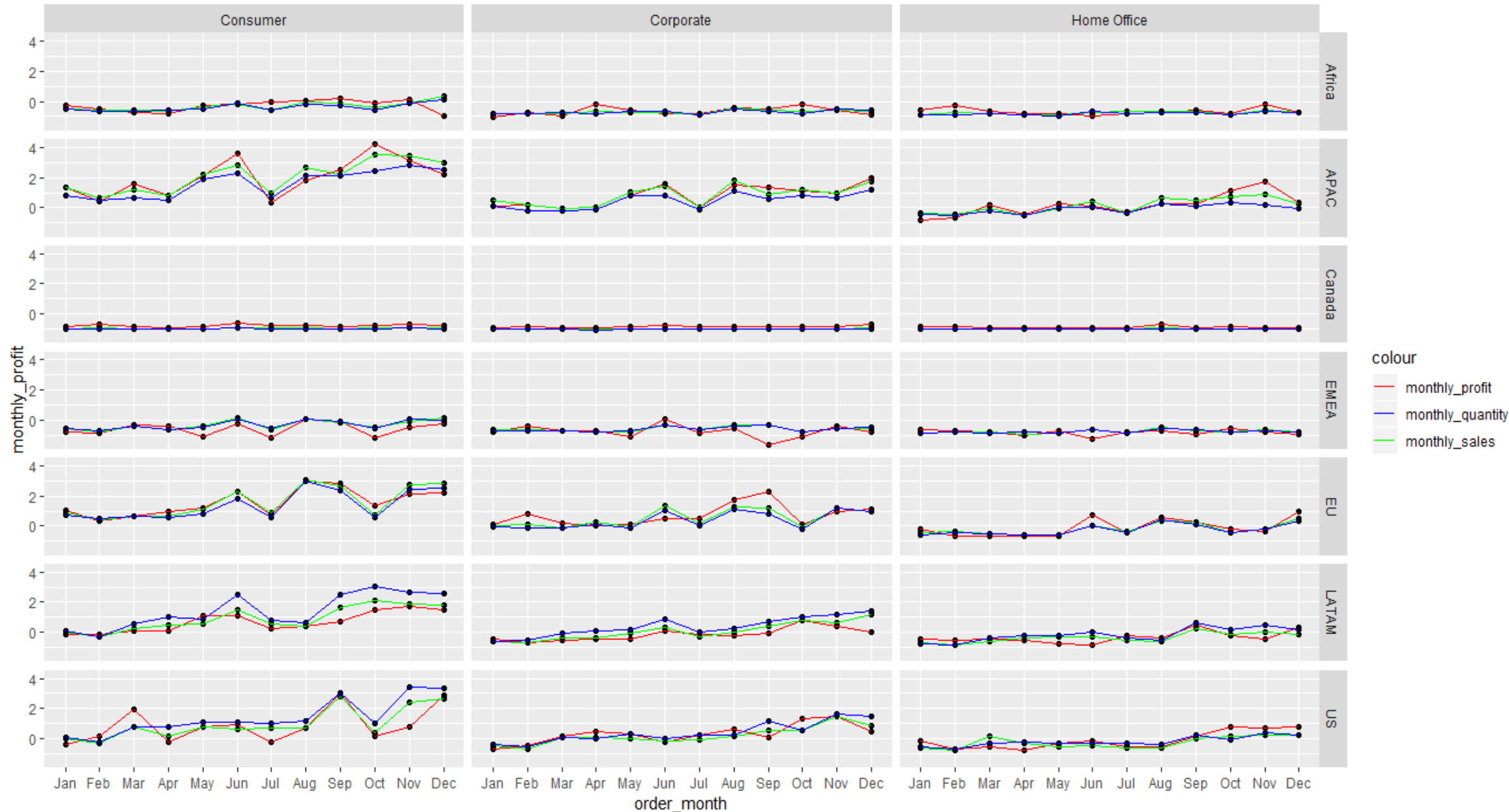
3. Data Preparation:

1. Extracted order month and year from the order date
2. Aggregating the data based in the profit per month per segment.
3. Identifying the top two segments based on the consistency in the profitability

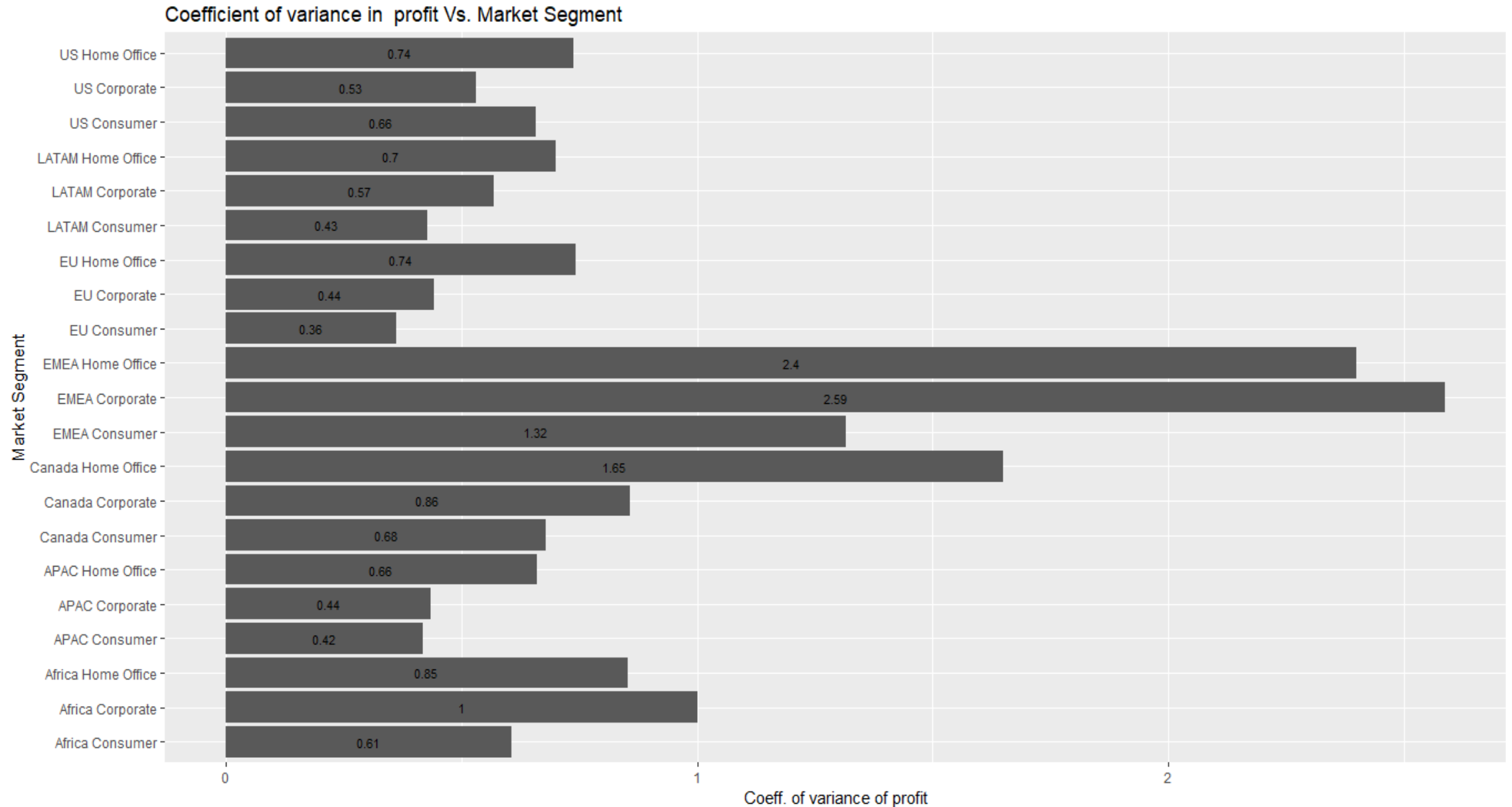
Tools Used :

- Rstudio is used for data cleaning and modelling

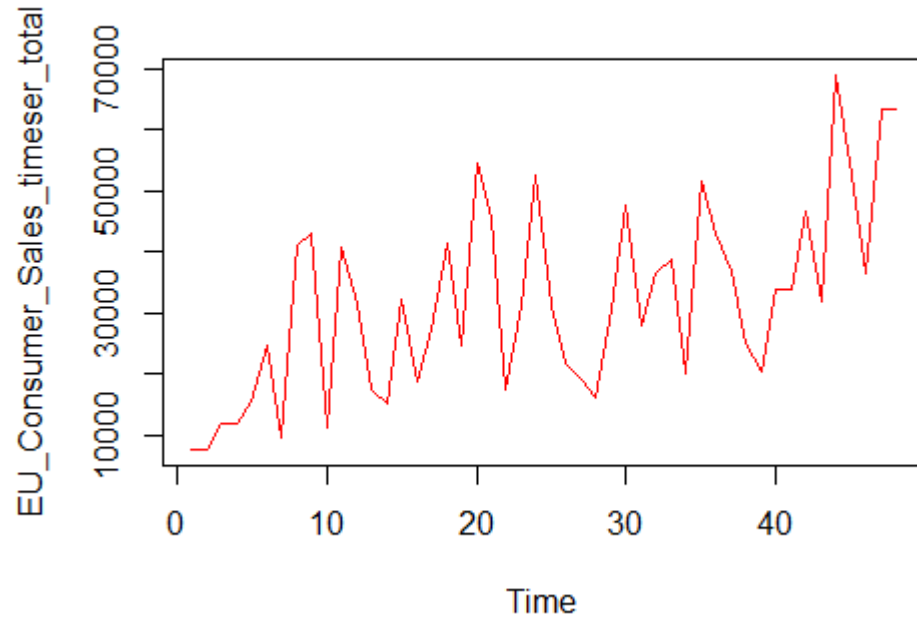
Exploratory Data Analysis for Profit,Sales,QtyUpGrad



Determining Coefficient of Variation for Market Segment

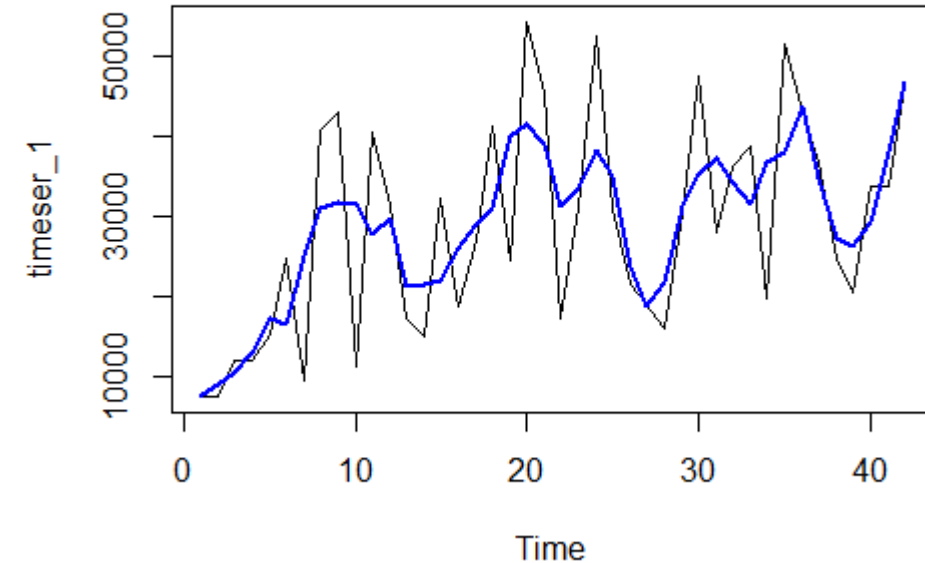


EU Consumer Sales Time Series

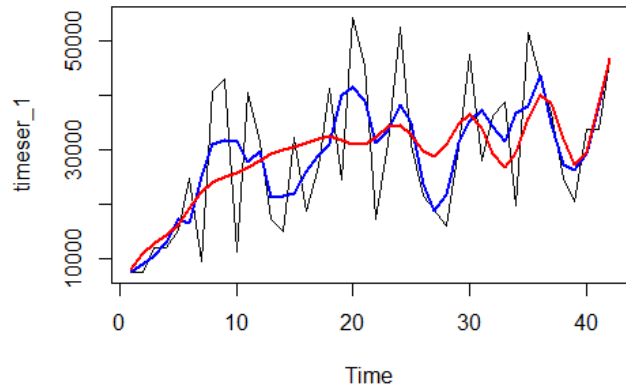


KPSS Test for Level Stationarity data: timeser_1 KPSS
 Level = 0.57717, Truncation lag parameter = 3, p-value = 0.02471

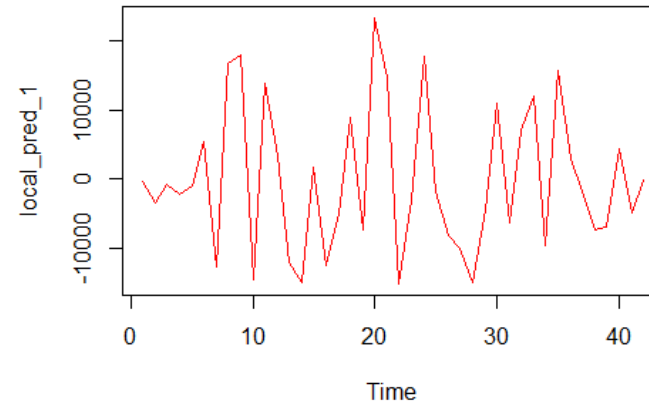
The original data series is Not Stationary, so it needs to be smoothed to remove seasonality and trend



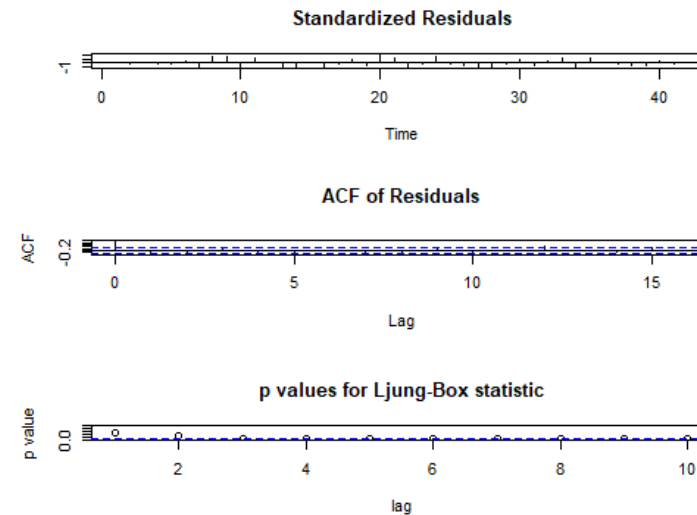
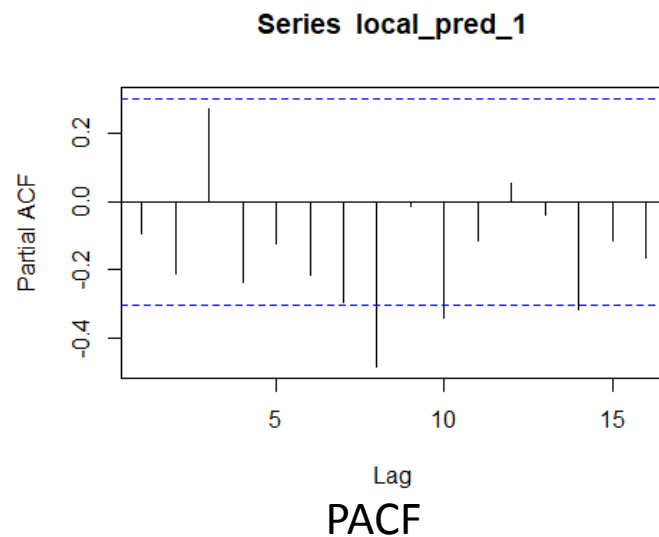
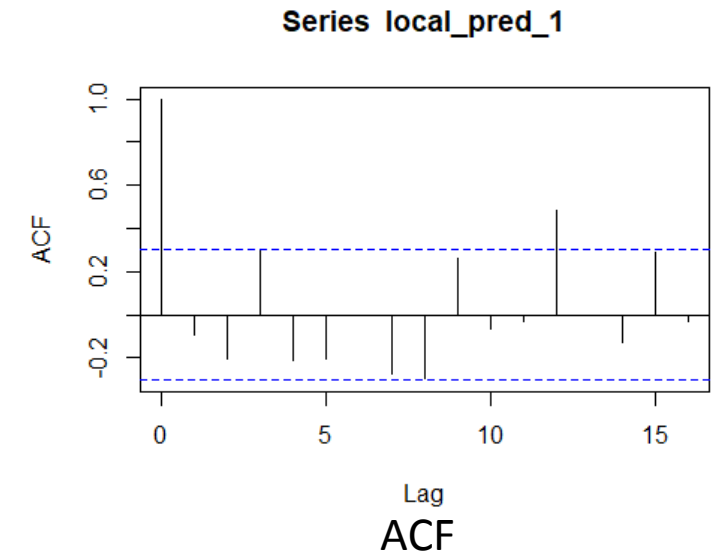
The series After Smoothing



Global Prediction Plot – Red line

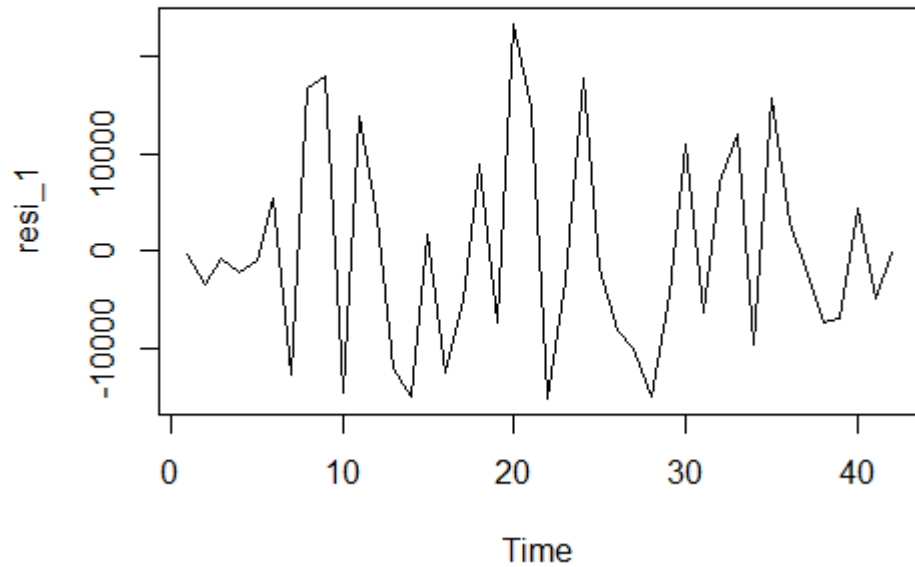


Locally Predicted Plot



AUTO-ARIMA

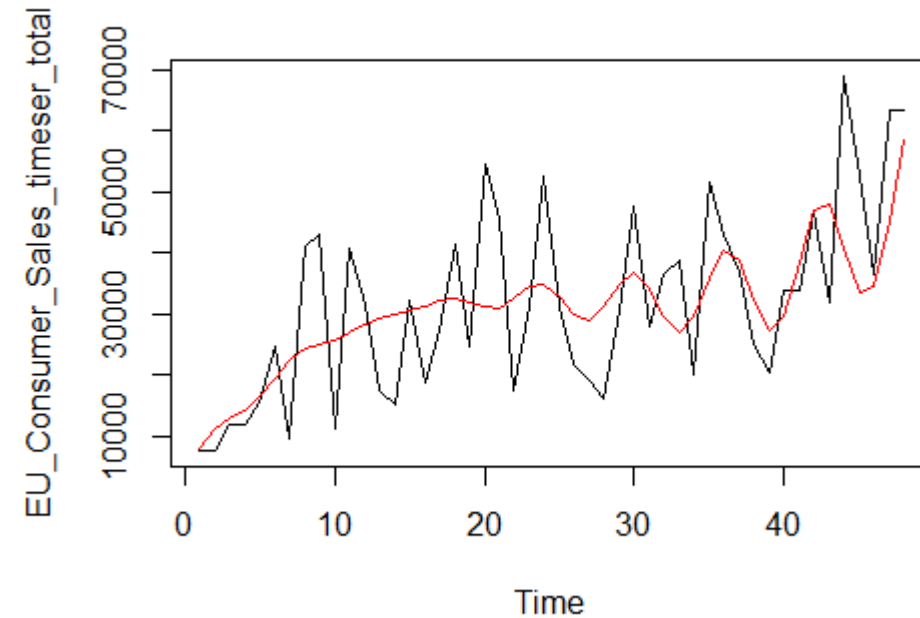
Residual test for the locally predicted part to check white noise



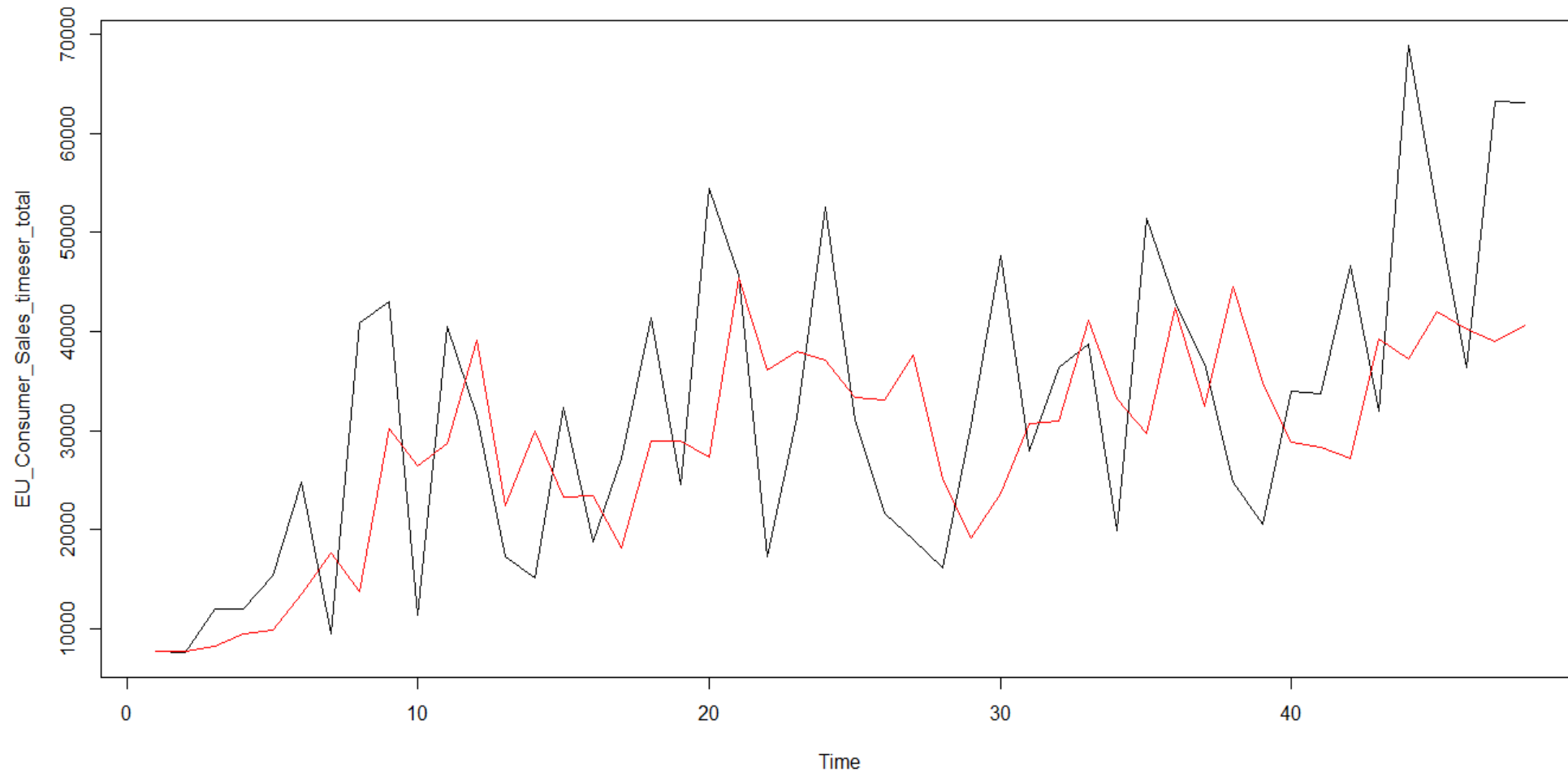
KPSS Test for Level Stationarity data: resi_1 KPSS Level = 0.038183, Truncation lag parameter = 3, p-value = 0.1

From both the tests it is clear that the series is now stationary

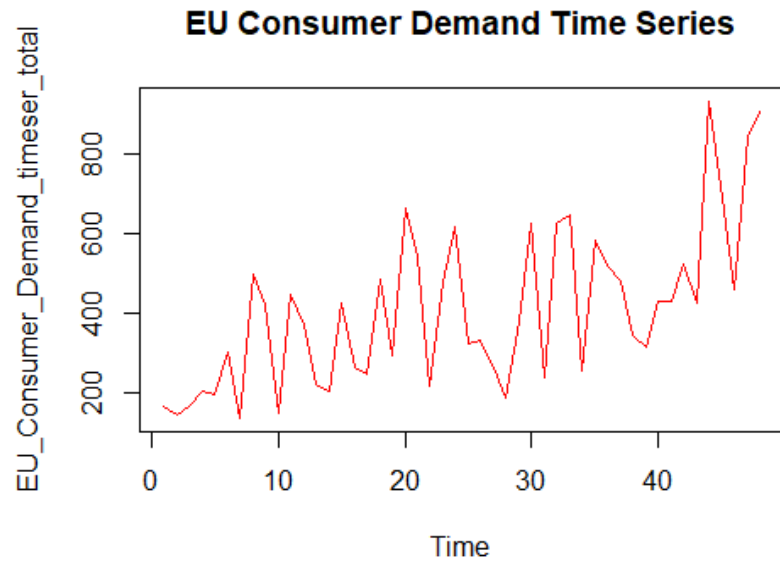
MAPE value of the globally predicted part
28.04386



Fitting the final data

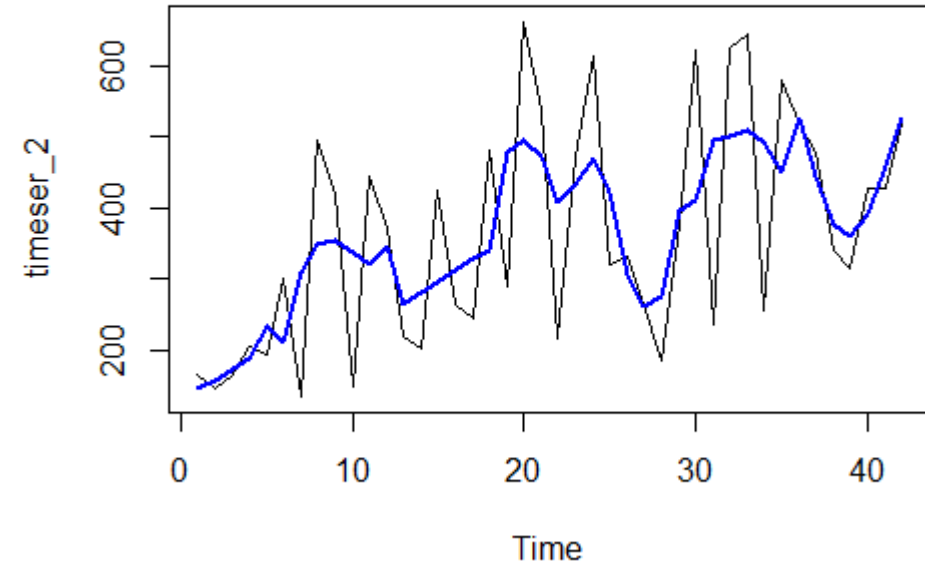


MAPE value->28.9226

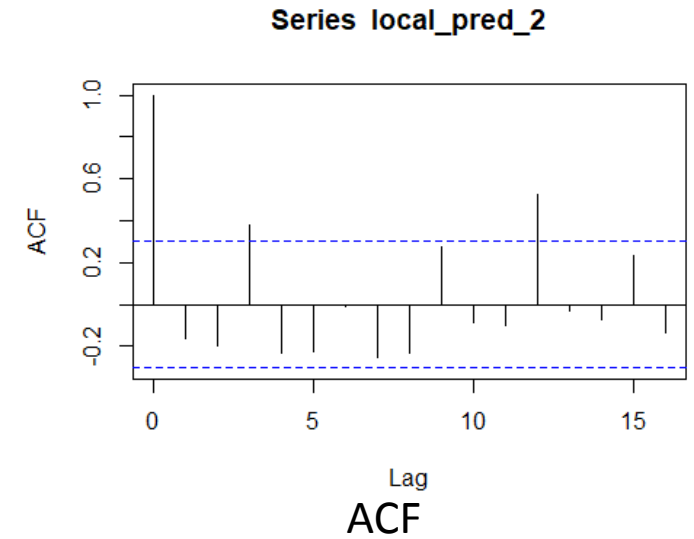
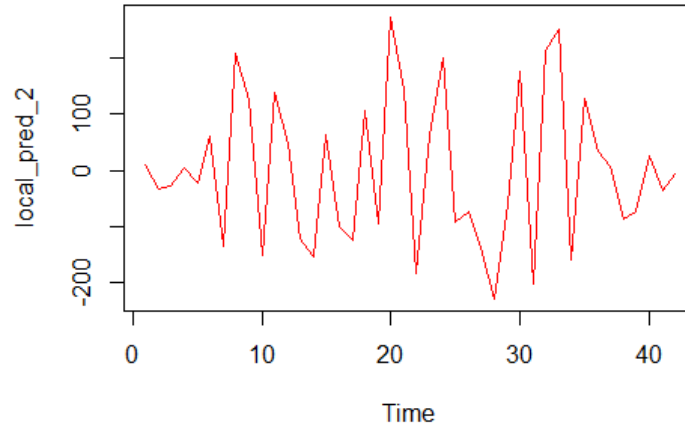
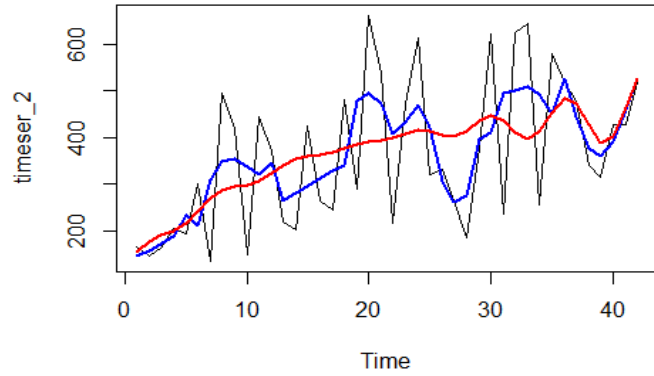


KPSS Test for Level Stationarity data: timereser_2 KPSS Level = 0.68981, Truncation lag parameter = 3, p-value = 0.01447

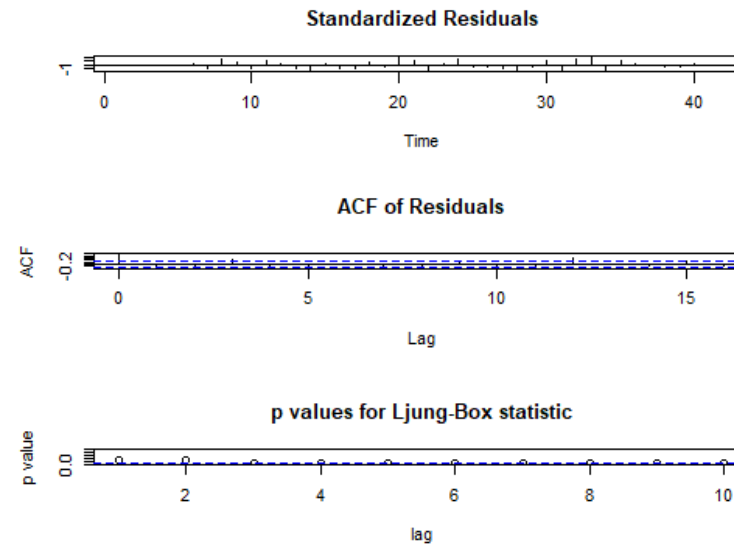
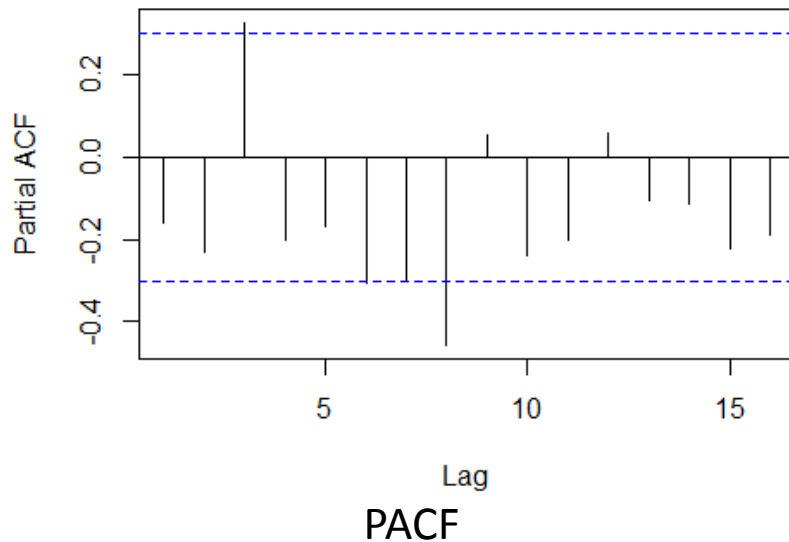
The original data series is Not Stationary, so it needs to be smoothed to remove seasonality and trend



The series After Smoothing

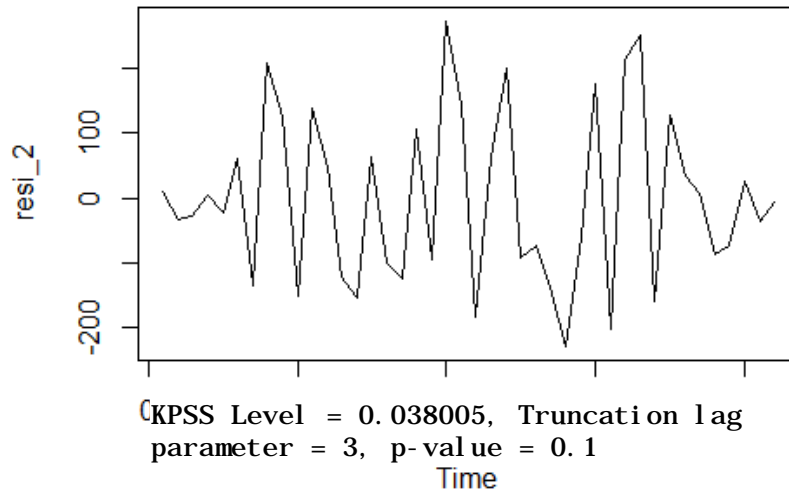


Series local_pred_2



AUTO-ARIMA

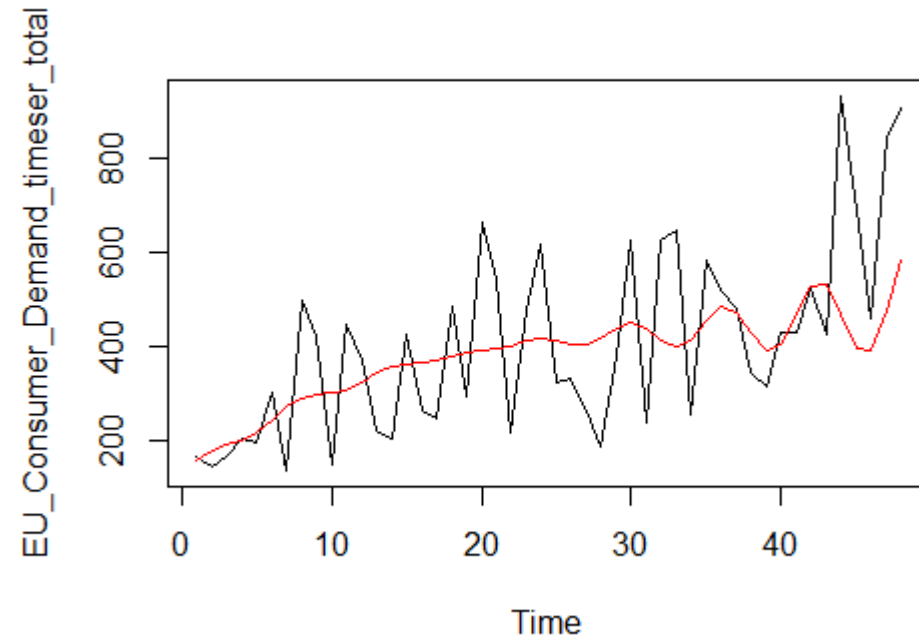
Residual test for the locally predicted part to check white noise



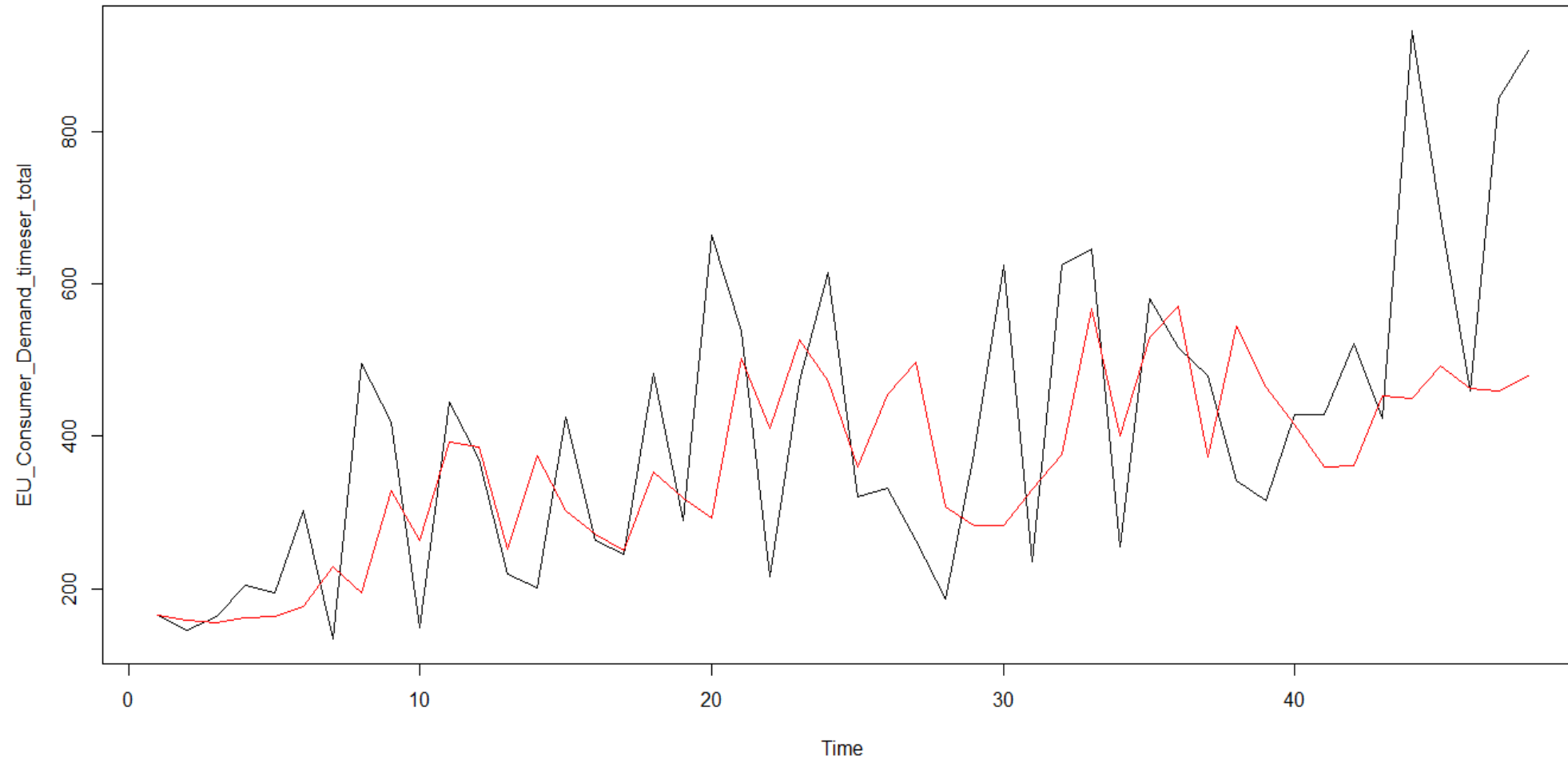
From both the tests it is clear that the series is now stationary

MAPE value of the globally predicted part

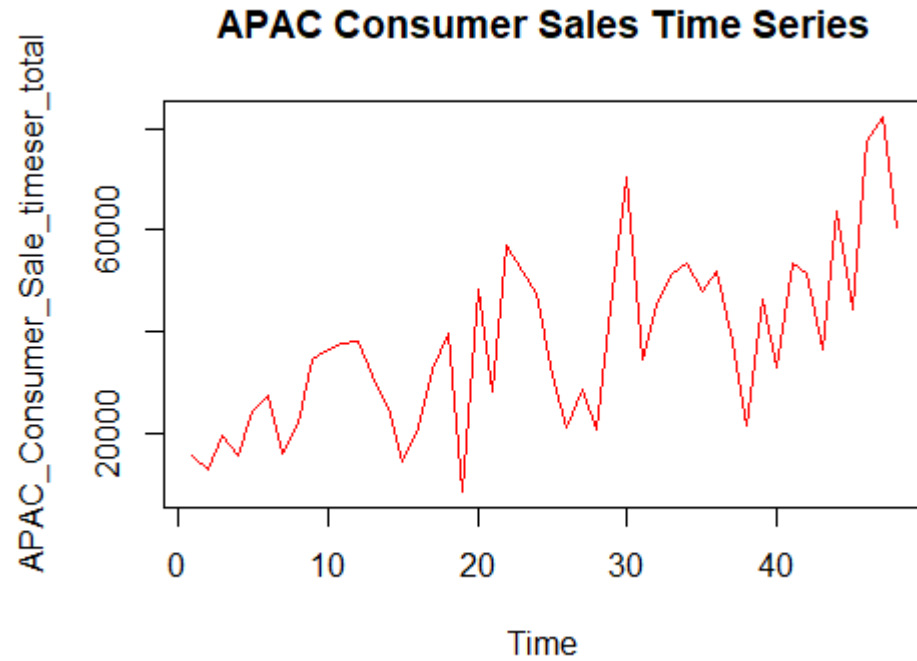
35.31684



Fitting the final data

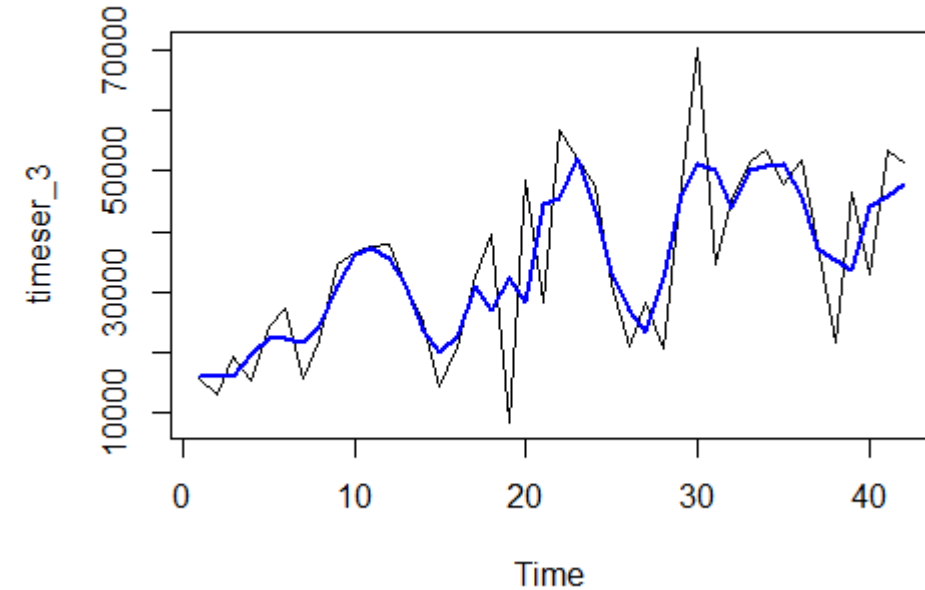


MAPE value->30.13319

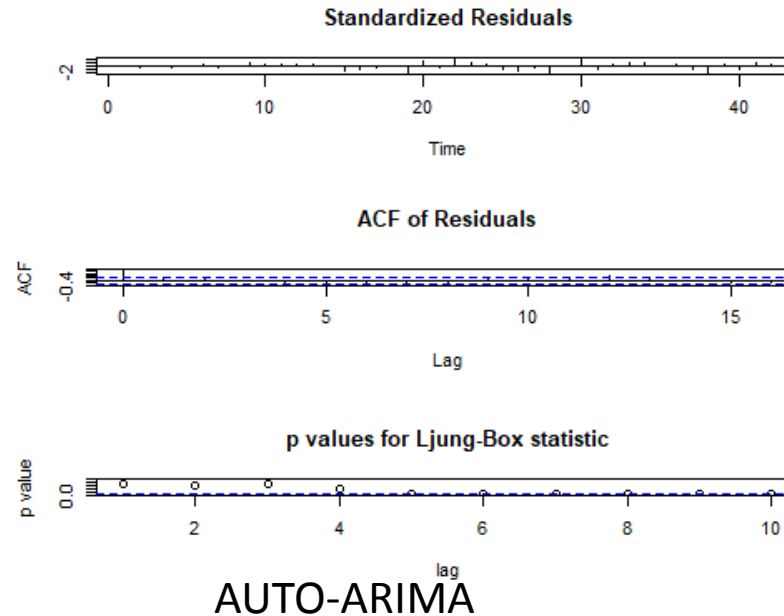
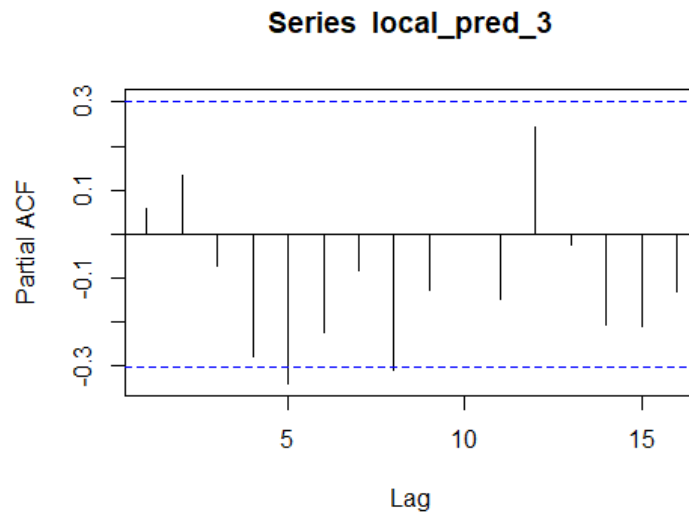
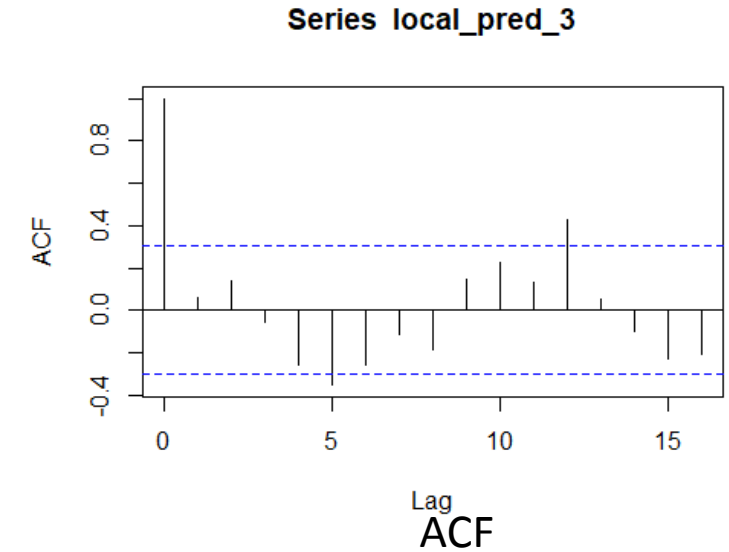
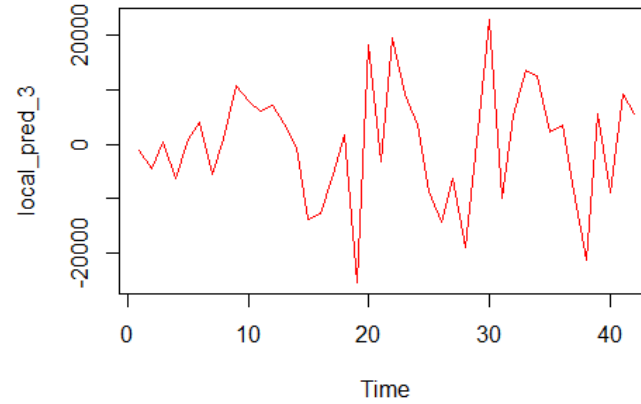
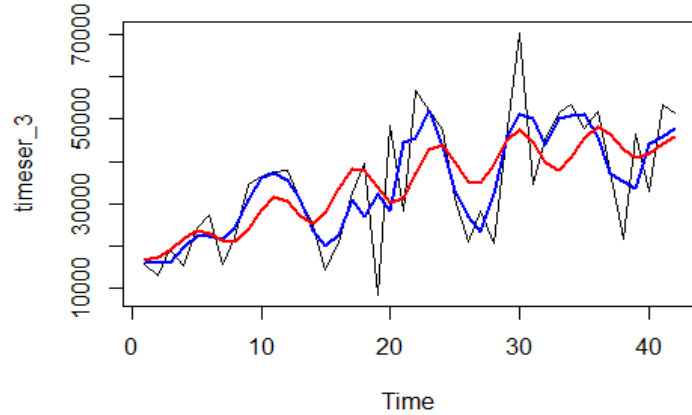


KPSS Level = 0.75633, Truncation
lag parameter = 3, p-value = 0.01

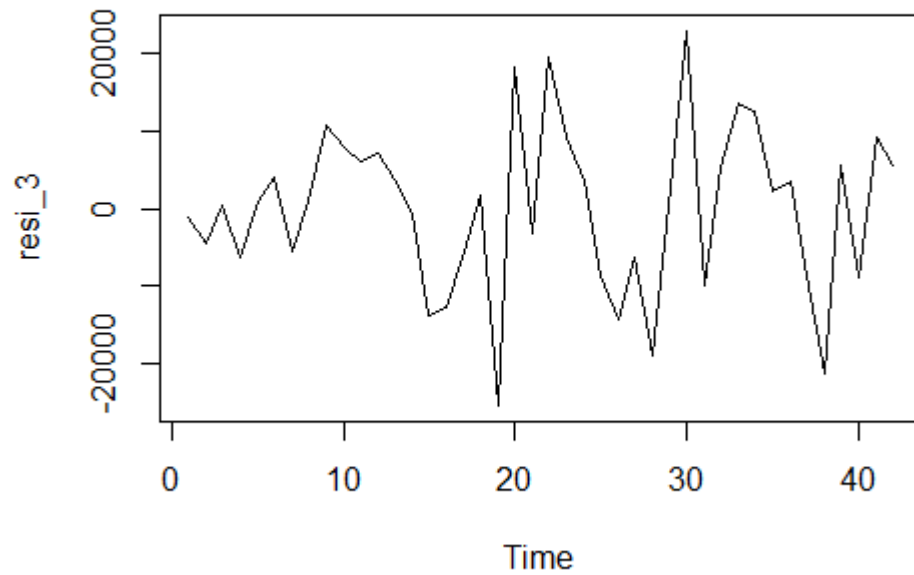
The original data series is Not Stationary, so it needs to be smoothed to remove seasonality and trend



The series After Smoothing



Residual test for the locally predicted part to check white noise

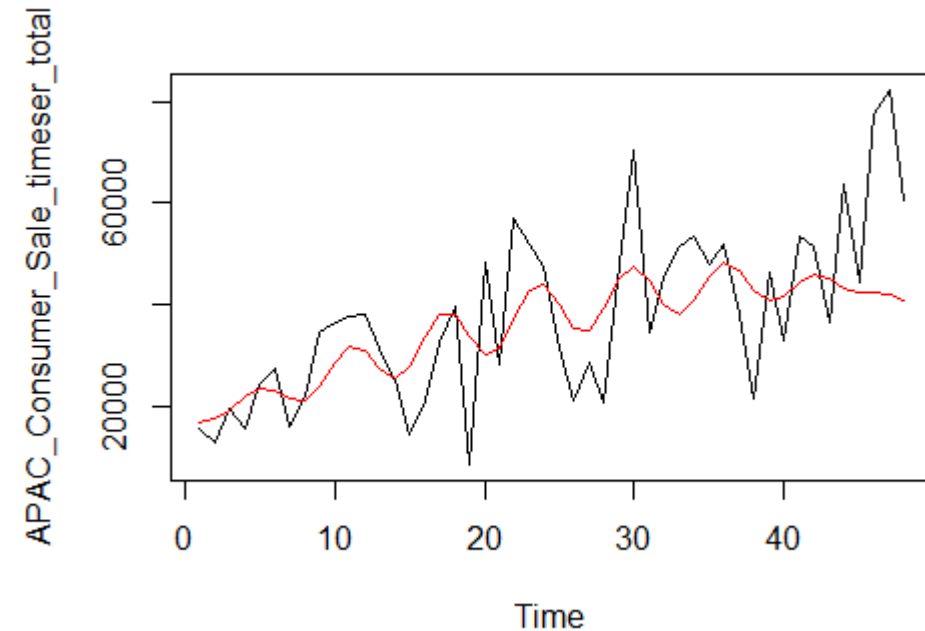


KPSS Level = 0.037483, Truncation lag parameter = 3, p-value = 0.1

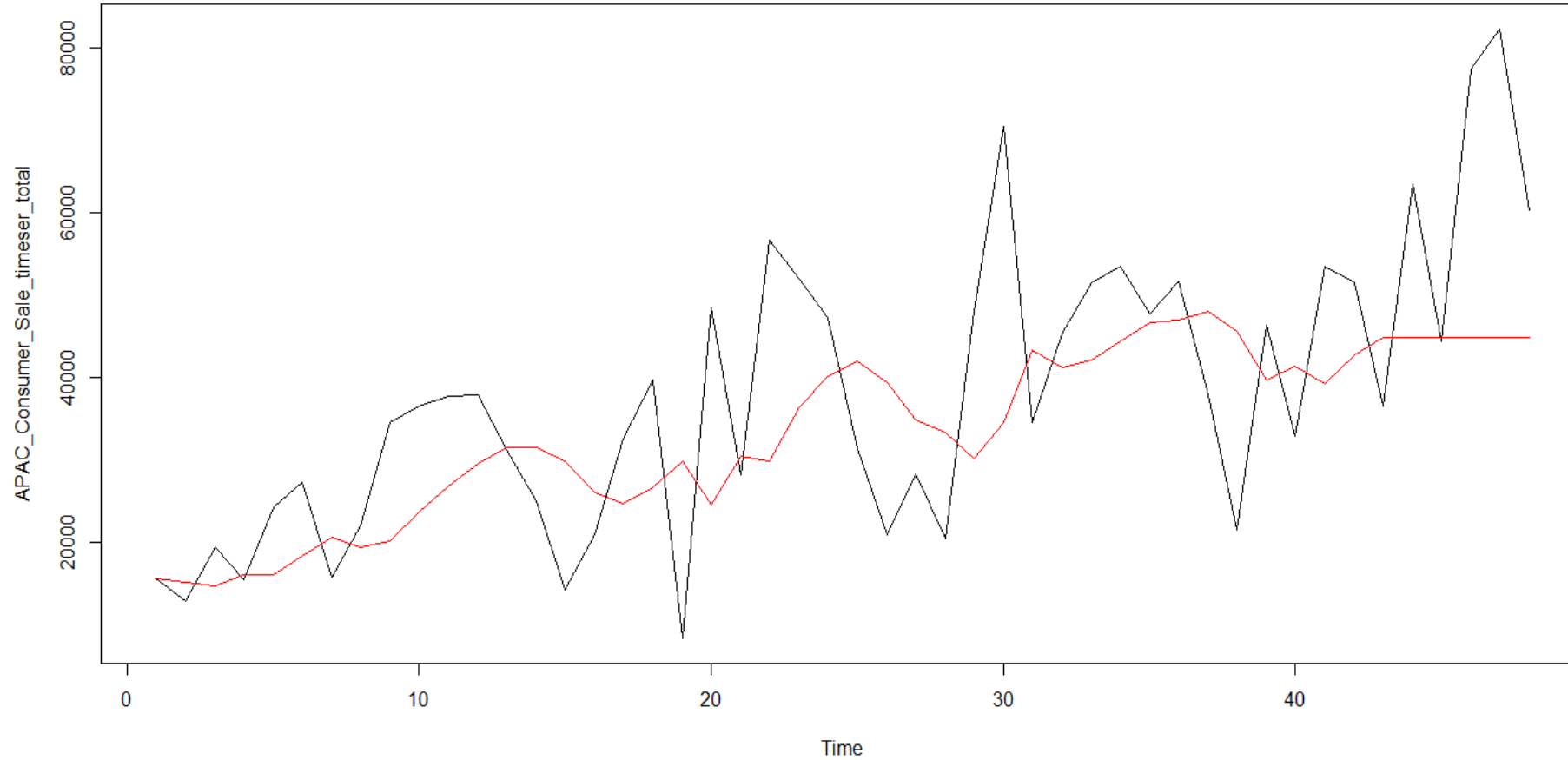
From both the tests it is clear that the series is now stationary

MAPE value of the globally predicted part

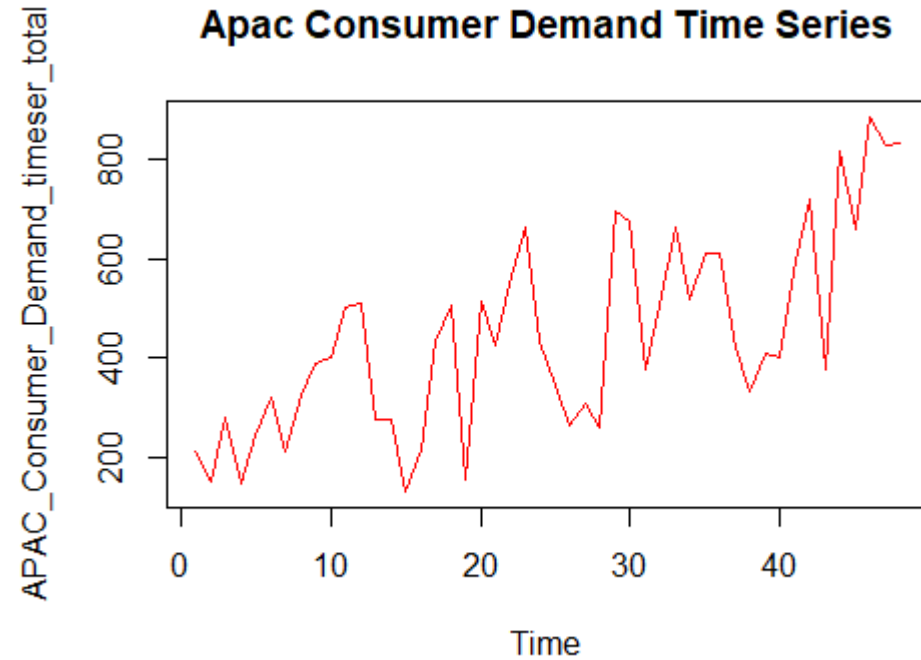
31.15703



Fitting the final data

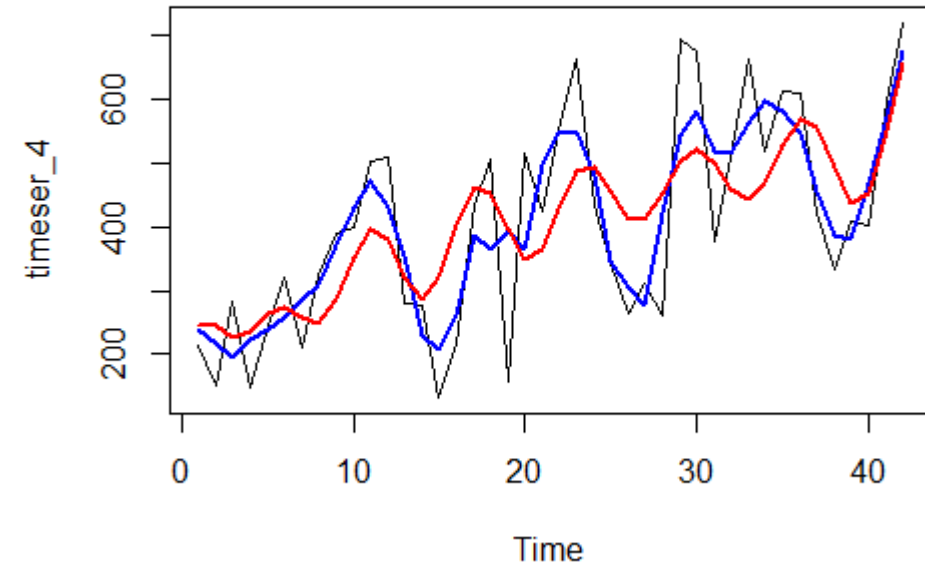


MAPE value->27.68952

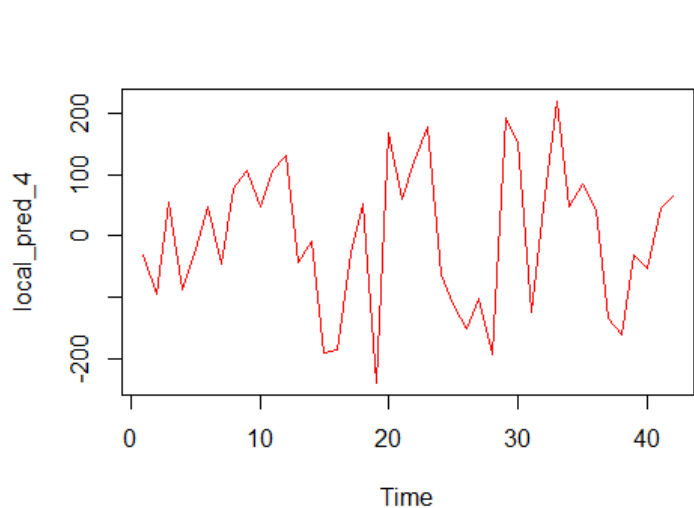


KPSS Level = 0.74074, Truncation lag
parameter = 3, p-value = 0.01

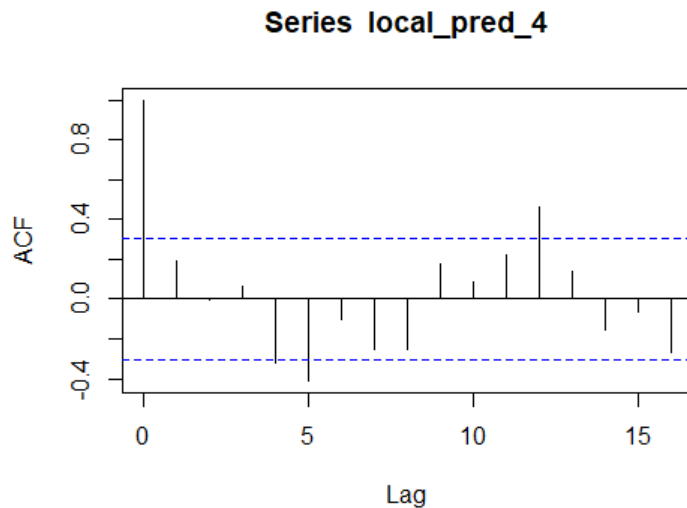
The original data series is Not Stationary, so it needs to be smoothed to remove seasonality and trend



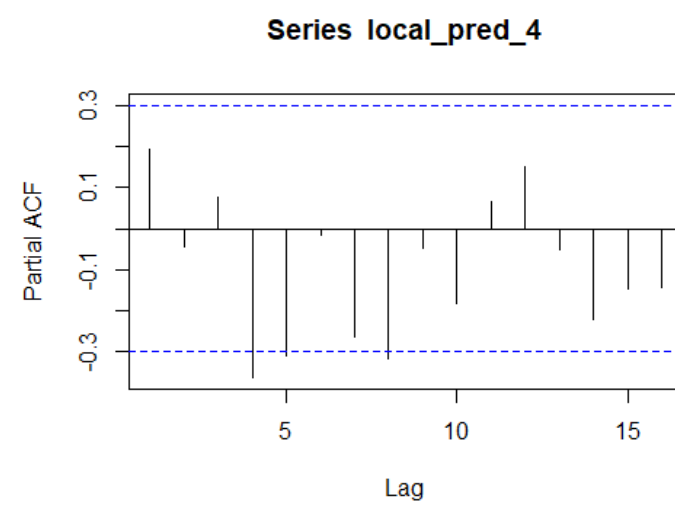
The series After Smoothing



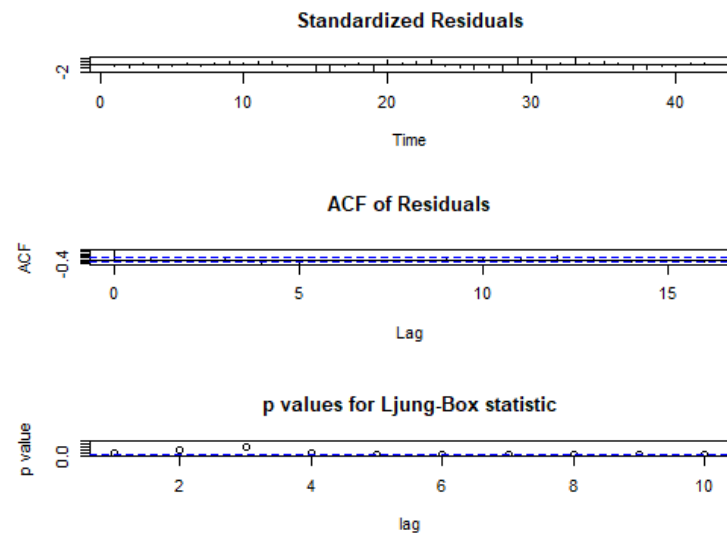
Global Prediction Plot – Red line



Locally Predicted Plot



ACF

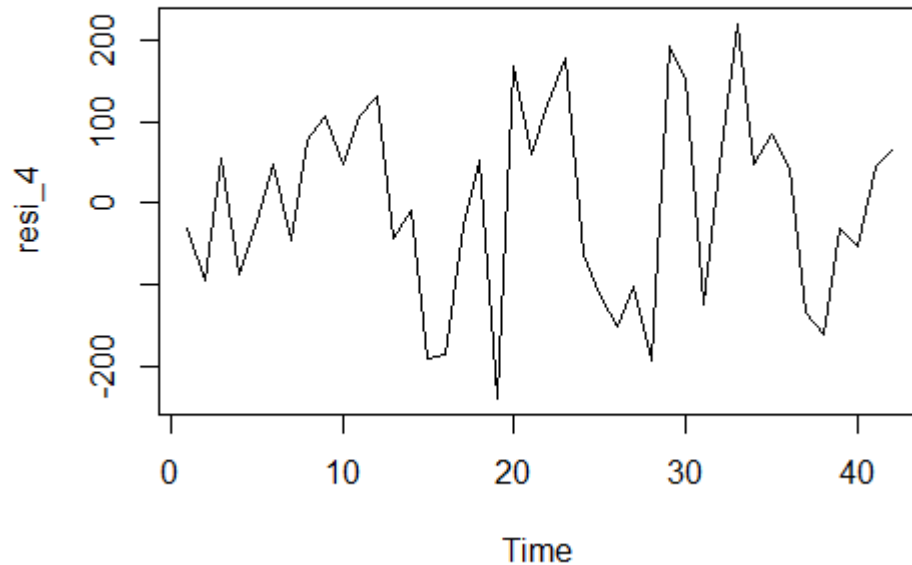


AUTO-ARIMA

MAPE value of the globally predicted part

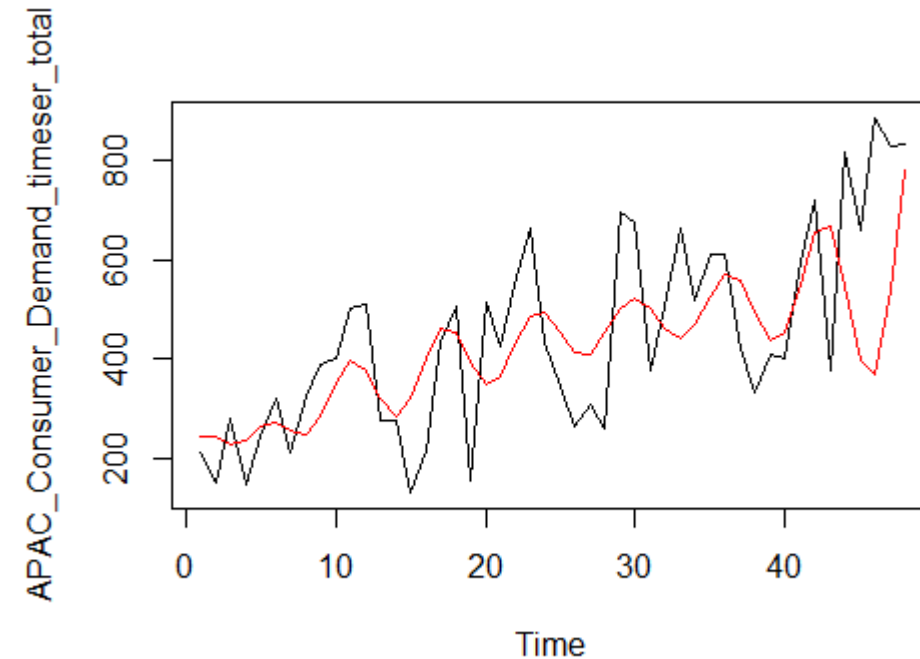
41.49113

Residual test for the locally predicted part to check white noise

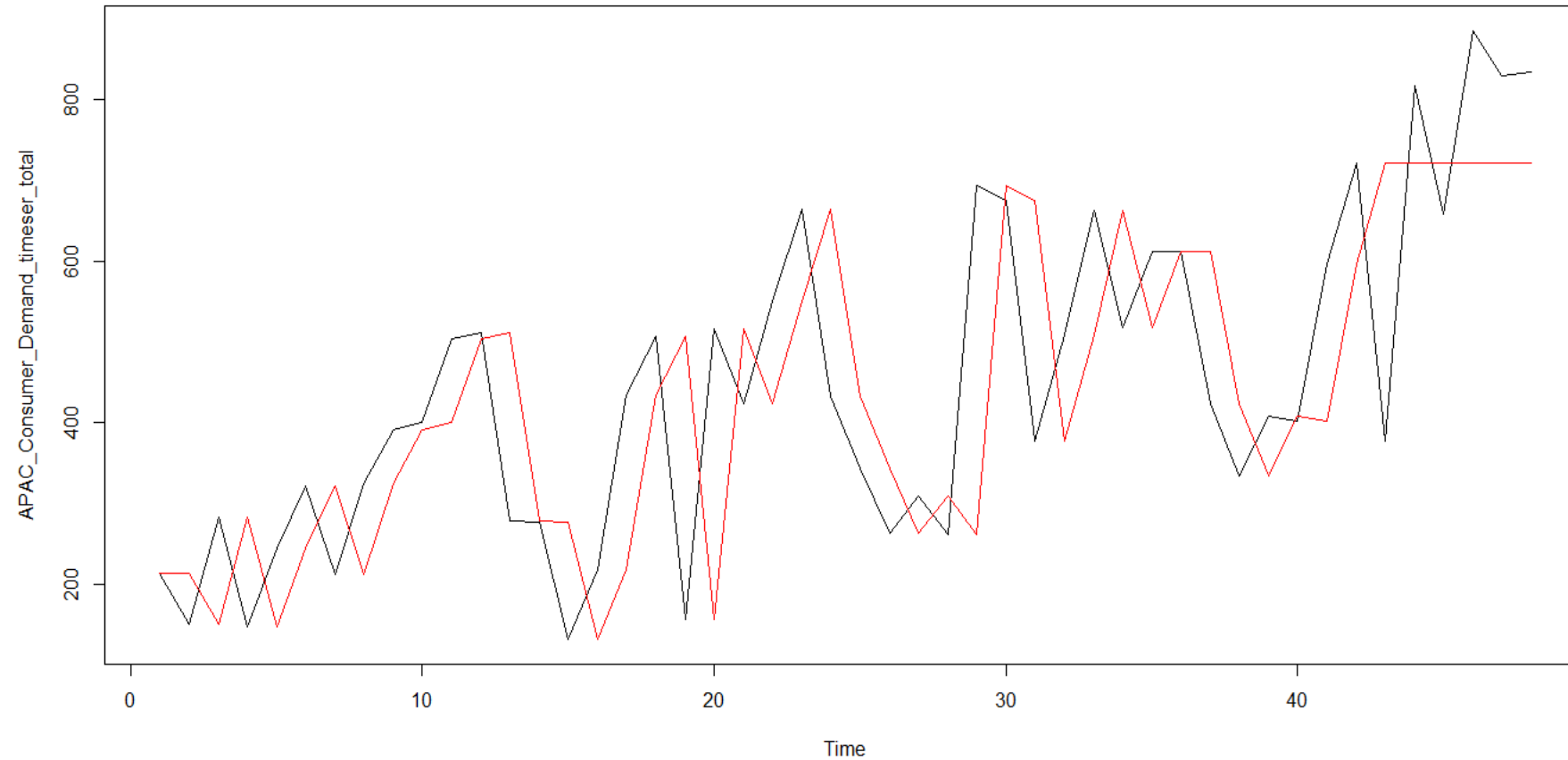


KPSS Level = 0.037207, Truncation lag parameter = 3, p-value = 0.1

From both the tests it is clear that the series is now stationary



Fitting the final data



MAPE value->26.24458

- The two most consistently profitable Market-Segments are
 1. APAC Consumer
 2. EU Consumer
- Sales across both Market-Segments show seasonal behaviour.
- Demand across above market segment have shown a linear increase in the Trend component.
- Auto ARIMA method has proven to be better than Classical Decomposition technique for APAC Consumer and EU Consumer for both Sales & Demand.
- So, it is recommended to invest more in EU and APAC Consumer Market-Segments as sales are expected to grow and sufficient inventory arrangement needs to be done