



RETAIL GIANT SALES FORECASTING CASE STUDY Group: DSWarriors

Group Members:

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



Retail Giant Sales Forecasting



- •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



Identification of the two most profitable segments

Problem Solving Methodology



Verification of Time Series Data Preparation for Time Series modeling output Forecasting **Business** Creation of time series data understanding Checking the stationary of the series ARIMA Model Classical Decomposition Method Data understanding Creation of time Creation of time series model series model using AutoARIMA function **Checking the Error value** Loading of data into R using the MAPE function Smoothening Cleaning of data & modification as required Visualizing analyzing the Forecasting Forecasting the global forecasted data and actual predictable part data Loading packages required for operations Forecasting the locally predictable part



Analysis – Data cleaning and preparation



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:

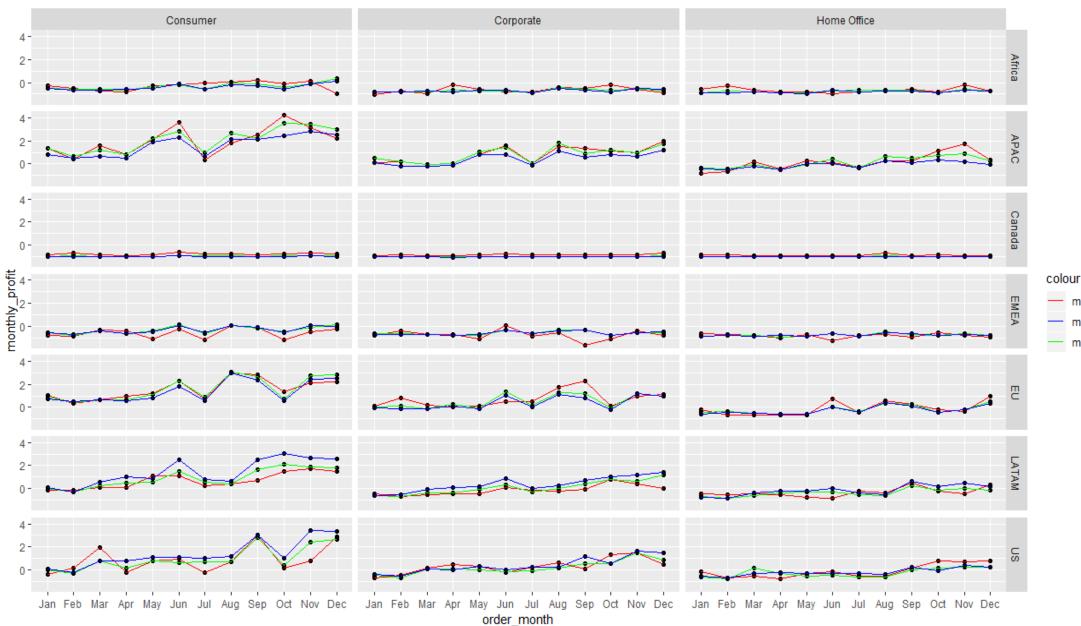
- 1. Extracted order month and year from the order date
- 2. Aggregating the data based in the profit per month per segment.
- 3. Identifying the top two segments based on the consistenancy in the profitability

Tools Used:

Rstudio is used for data cleaning and modelling



Exploratory Data Analysis for Profit, Sales, Qty Up Grad

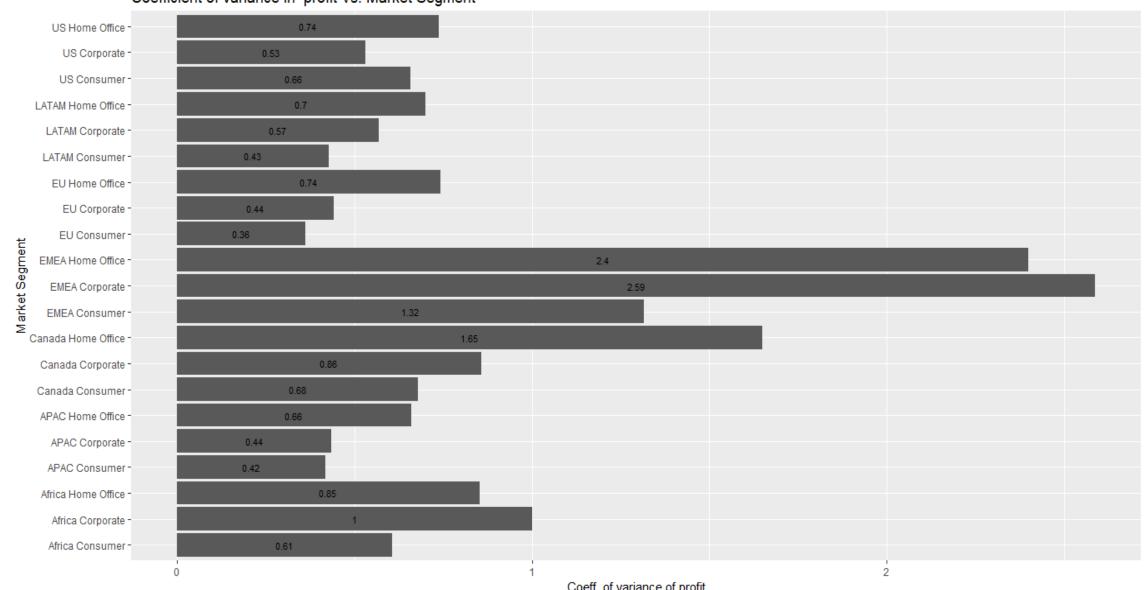




Determining Coefficient of Variation for Market Segment



Coefficient of variance in profit Vs. Market Segment

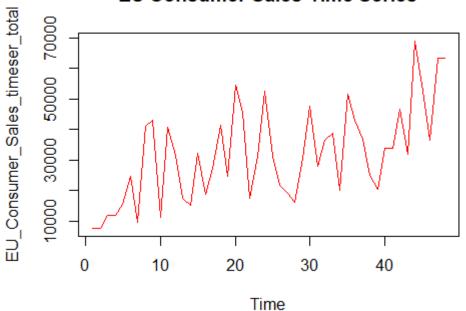




EU Consumer Sales

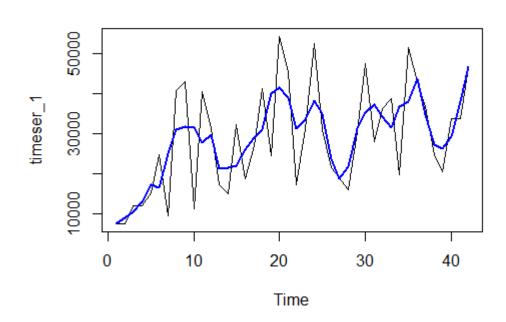


EU Consumer Sales Time Series



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

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

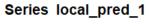


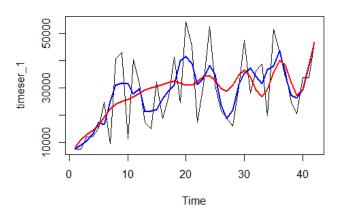
The series After Smoothening



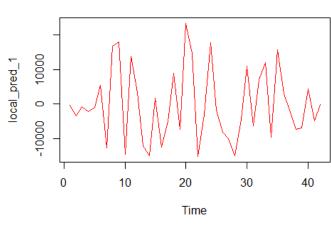
EU Consumer Sales



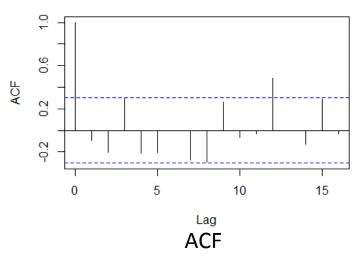


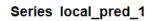


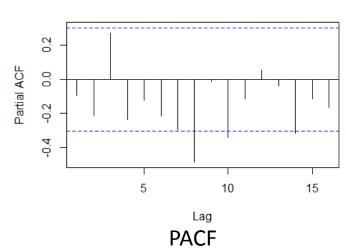
Global Prediction Plot – Red line

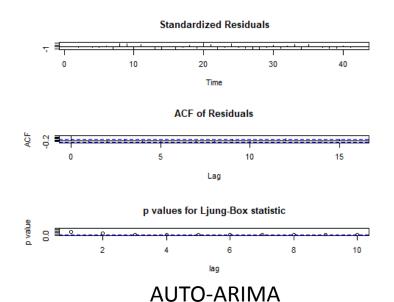


Locally Predicted Plot







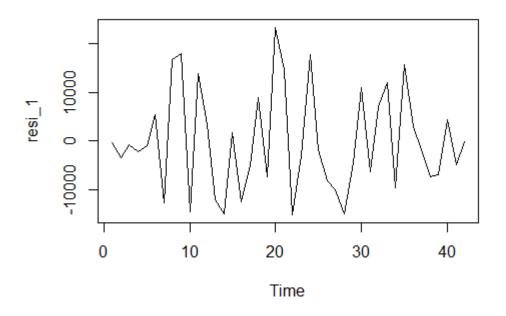




EU Consumer Sales



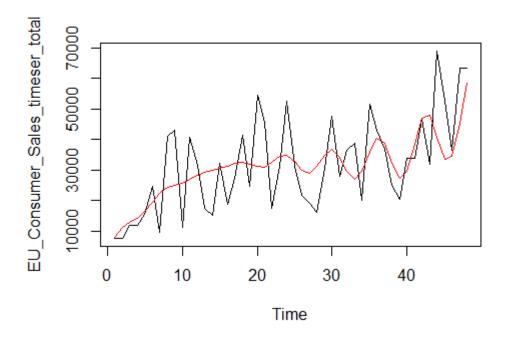
Residual test for the locally predicted part to check white noise



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

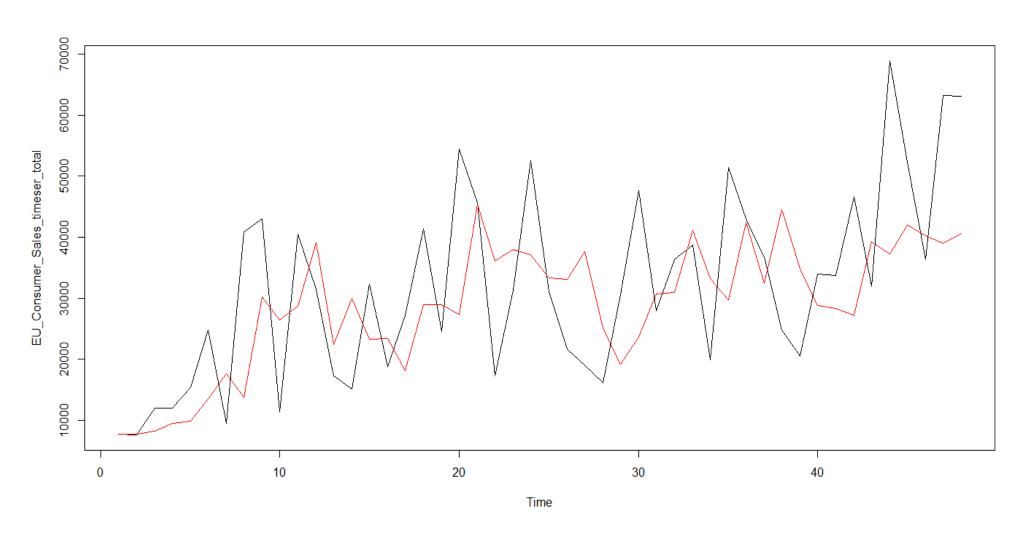


Fitting the final data



EU Consumer Sales-ARIMA Model



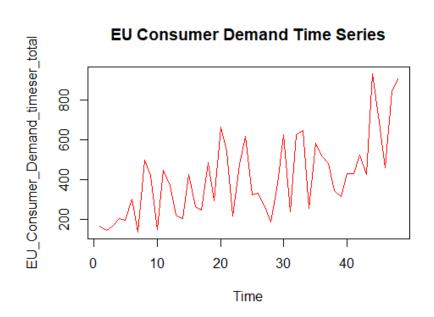


MAPE value->28.9226



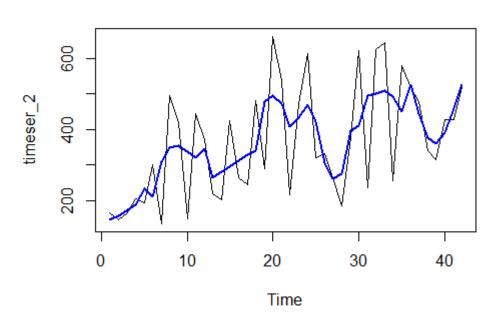
EU Consumer Demand





KPSS Test for Level Stationarity data: timeser_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

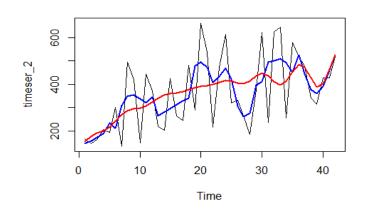


The series After Smoothening

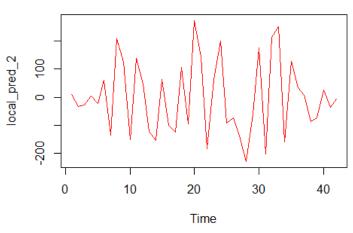


EU Consumer Demand

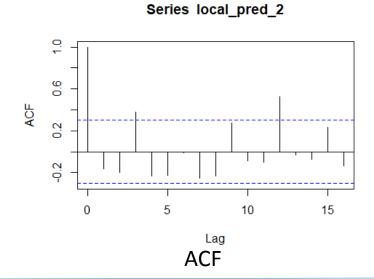


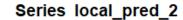


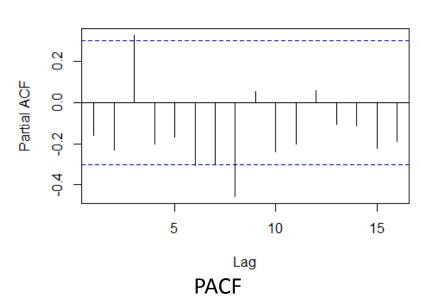
Global Prediction Plot - Red line



Locally Predicted Plot



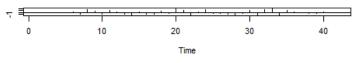




2 4 6 8 10

AUTO-ARIMA

Standardized Residuals



ACF of Residuals

p values for Ljung-Box statistic

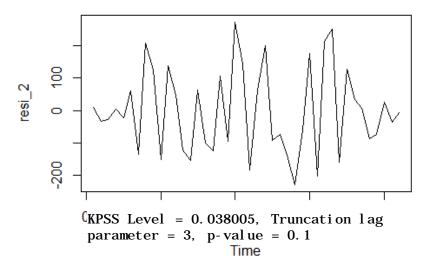




EU Consumer Demand

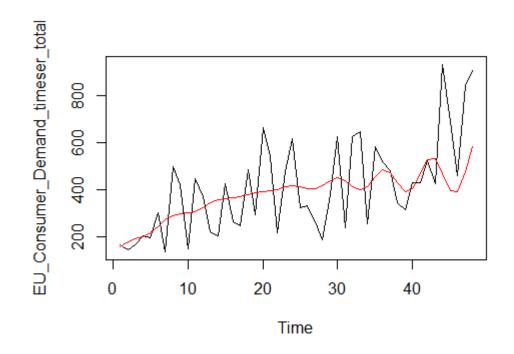


Residual test for the locally predicted part to check white noise



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

MAPE value of the globally predicted part 35. 31684

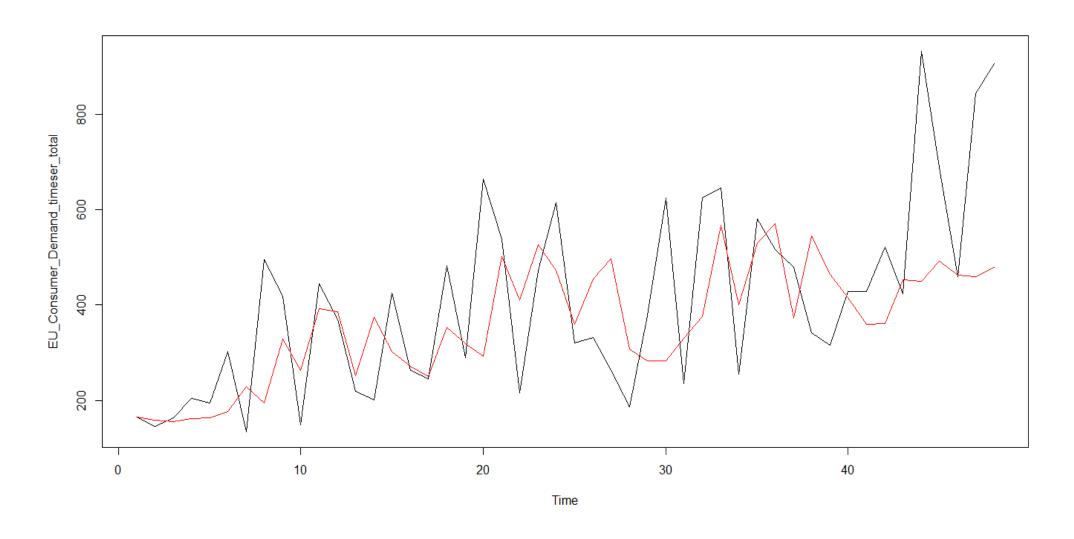


Fitting the final data



EU Consumer Demand-ARIMA Model



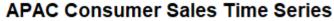


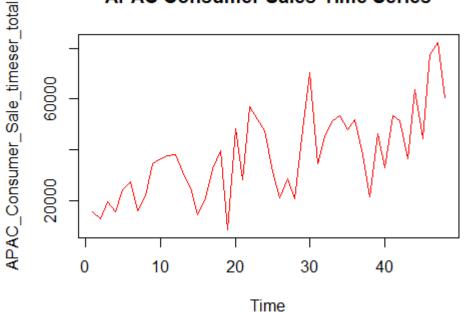
MAPE value->30.13319



APAC Consumer Sales

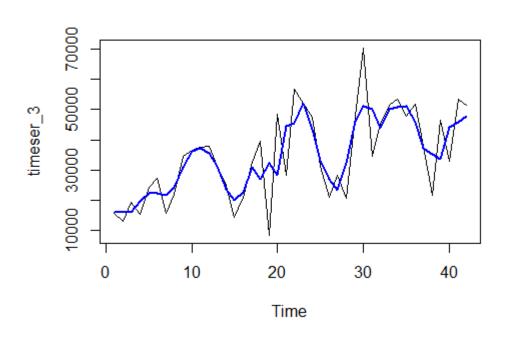






KPSS Level = 0.75633, Truncation lag parameter = 3, p-value = 0.01

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

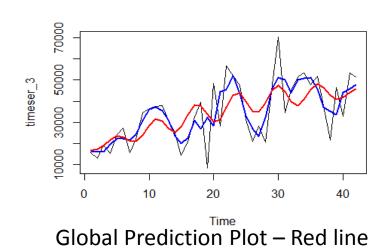


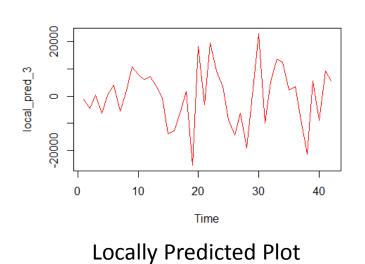
The series After Smoothening

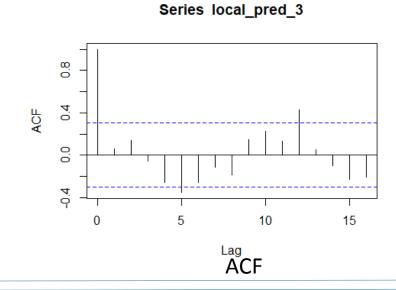


APAC Consumer Sales

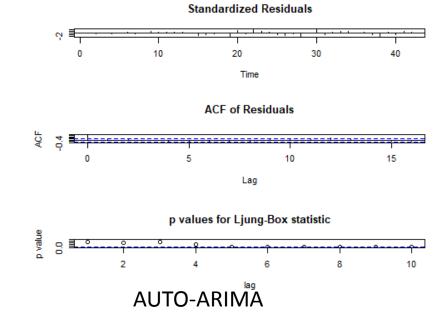








PACF

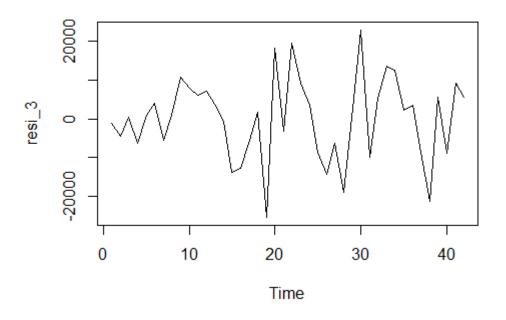




APAC Consumer Sales



Residual test for the locally predicted part to check white noise

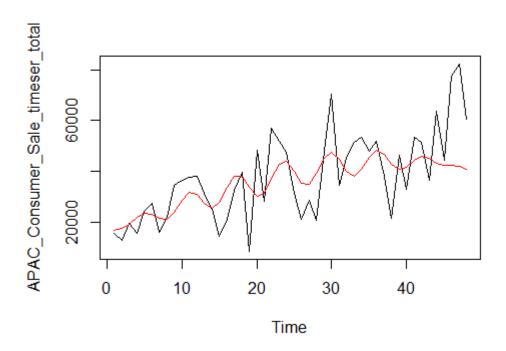


KPSS Level = 0.037483, Truncation lag parameter = 3, p-value = 0.1

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

MAPE value of the globally predicted part

31. 15703

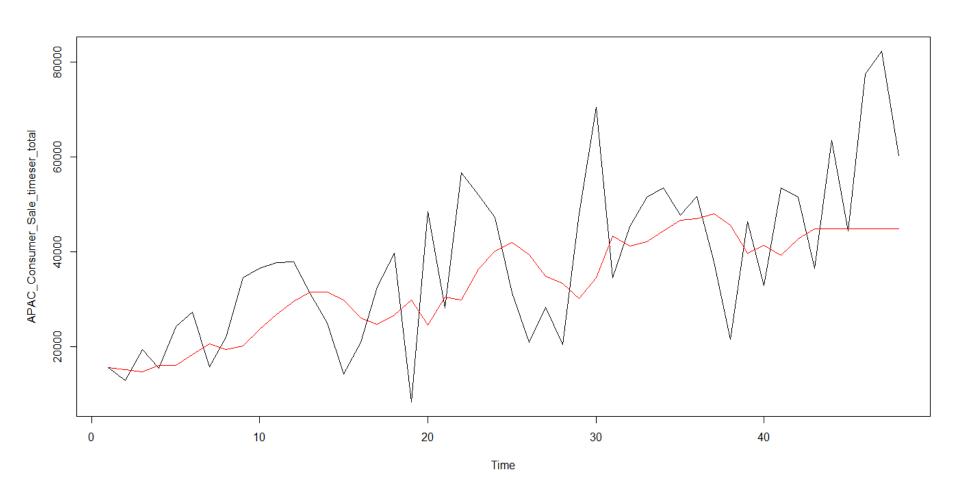


Fitting the final data



APAC Consumer Sales-ARIMA Model



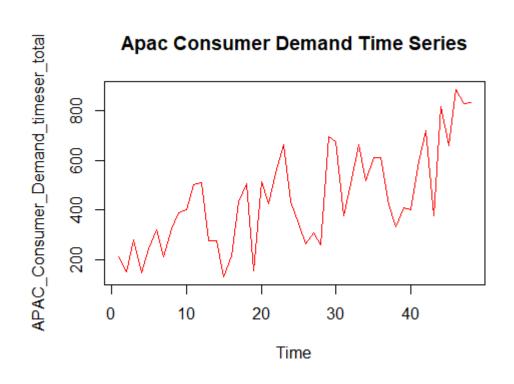


MAPE value->27.68952



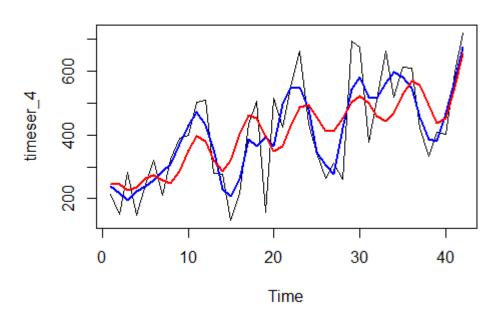
APAC Consumer Demand





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

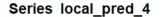


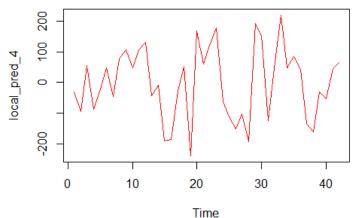
The series After Smoothening



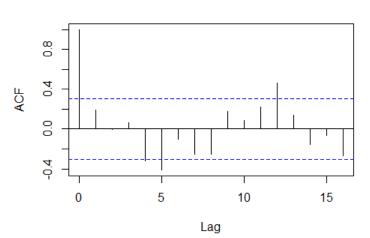
APAC Consumer Demand





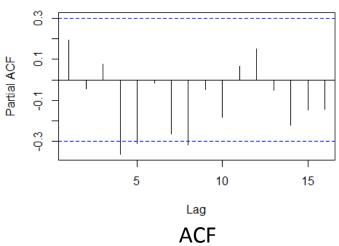


Global Prediction Plot – Red line

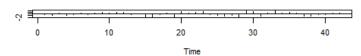


Locally Predicted Plot

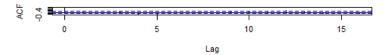
Series local_pred_4

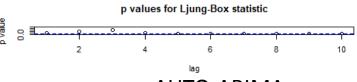


Standardized Residuals



ACF of Residuals





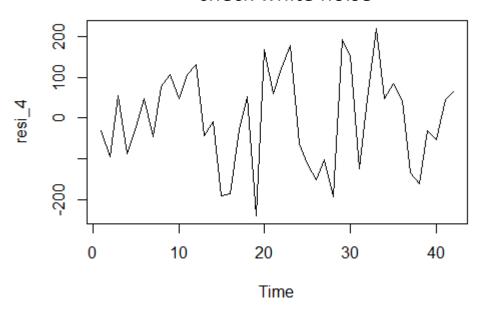
AUTO-ARIMA



APAC Consumer Demand



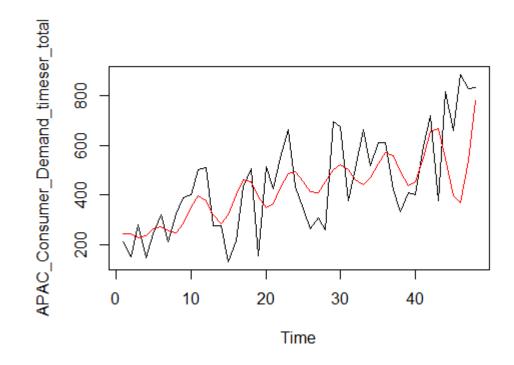
Residual test for the locally predicted part to check white noise



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

MAPE value of the globally predicted part 41. 49113

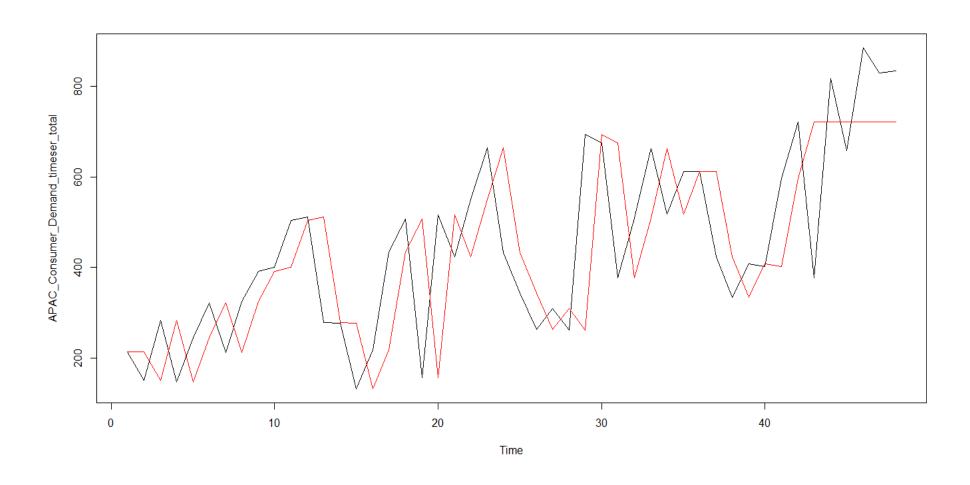


Fitting the final data



APAC Consumer Demand-ARIMA Model





MAPE value->26.24458



Conclusions



- The two most consistently profitable Market-Segments are
- 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