

# TIME SERIES – RETAIL GIANT SALES FORECASTING CASE STUDY

## SUBMISSION

Group Name:

1. Ashwin Rangarajan - DDA1730137
2. Prachi Prakash – DDA1730058
3. Shikha Chaturvedi – DDA1730288
4. Sumanth Sundaramurthy - DDA1730138

## Business Objective – Forecast The Sales & Demand For Next 6 Months To Manage Revenue & Inventory

“Global Mart” is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories. Store caters to 7 different market segments i.e. “US”, “APAC”, “EU”, “AFRICA”, “EMEA”, “LATAM” & “CANADA” and 3 major categories i.e. Consumer, Corporate & Home Office.

### **Business Strategy:**

The Sales/Operation manager now needs to forecast the sales and the demand for next 6 months however to forecast not all the 21 market buckets are important from the store’s point of view. For this Analysis one needs to find out the 2 most profitable market segment and forecast the sales and demand for these 2.

The strategy of the analysis is to perform:

- Using Coefficient of Variation Identify 2 profitable market segments. Coefficient of Variation often expressed as percentage is defined as the ratio of the Standard Deviation to the Mean.
- Once the required segments have been identified, using Classical Decomposition and auto ARIMA we would build the required model.
- With a satisfactory model, we will forecast the sales/demand data for next 6 months and using MAPE value check the accuracy of the forecast.

The approach taken here is the CRISP-DM framework. The Steps involved in CRISP-DM framework are –

## Business Understanding

01

## Data Understanding

02

## Data Preparation

03

## Data Modelling

04

## Model Evaluation

05

- “Global Mart” is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office.
- The Sales/Operation manager of Global Mart needs to forecast the sales and demand for next 6 months pertaining to the 2 most profitable Segments across markets.

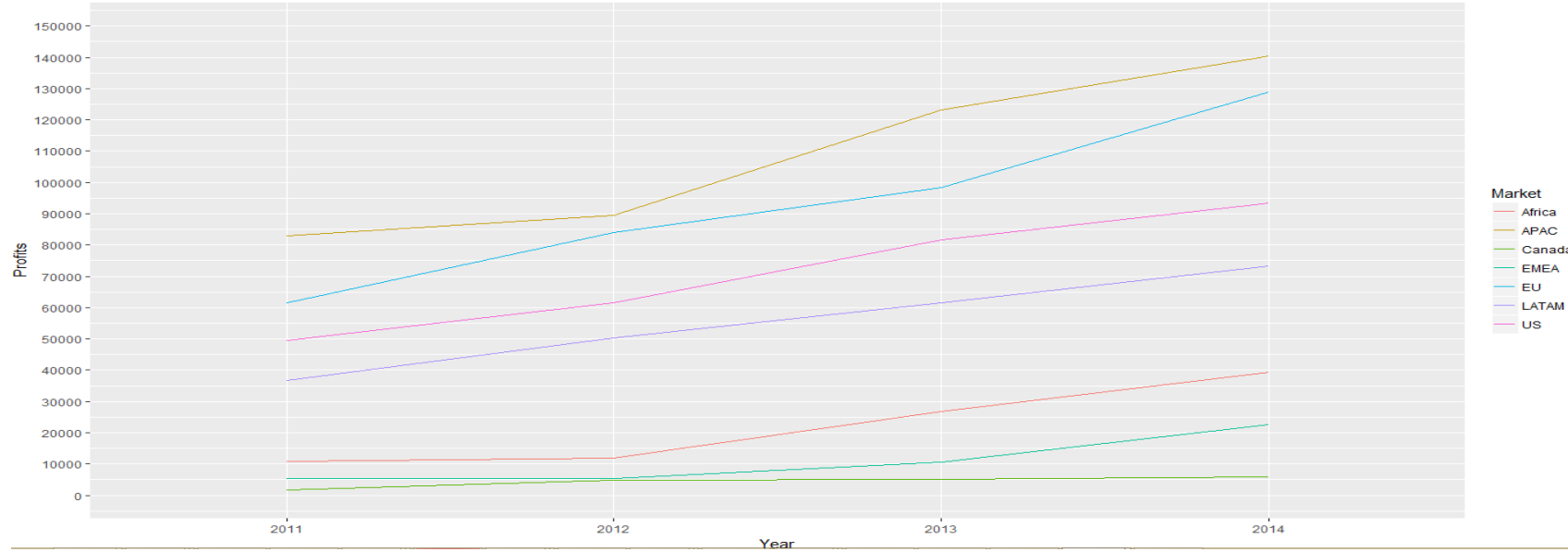
- Dataset provided represents a particular order made to the online store.
- Each record has 24 attributes related to each transaction and the transaction is available from year 2011 to 2014
- Market Attribute represents the geographical market sector the customer belongs to and Segment attribute gives which product categories the customer belongs to.
- Data Dictionary provided helps us understand the importance of each variable and their need for the analysis.

- Verified for deduplication of data and converted Order & Ship date to R Date format.
- Checked for NA across dataset and handled them appropriately either by imputing or removing them.
- Created a derived metrics Year-Month from Order Date.
- For identifying the 2 most profitable market segments, entire data is aggregated on 3 attributes – Sales, Quantity & Profit over the order date. Using the **coefficient of variation of the Profit** the 2 market segments identified were – **Consumer segment for APAC & EU markets.**

- Modelling is the heart of data analytics. A model which takes relevant data as input and gives an output one is interested in.
- Once the data set is divided to 2 most profitable market segment, we have used Classical Decomposition /h Moving Average & Exponential Smoothing and **auto ARIMA** to build the appropriate model.
- Also prior to modelling the data set was divided into InData and OutData (last 6 months values) and modeling was performed on the InData.

- Model Evaluation is a step where one does the litmus test.
- In this step, we used the final model to forecast the sales & demand on the OutData.
- To evaluate the accuracy of the model using MAPE as an indicator.

Profits Across Markets From 2011 to 2014

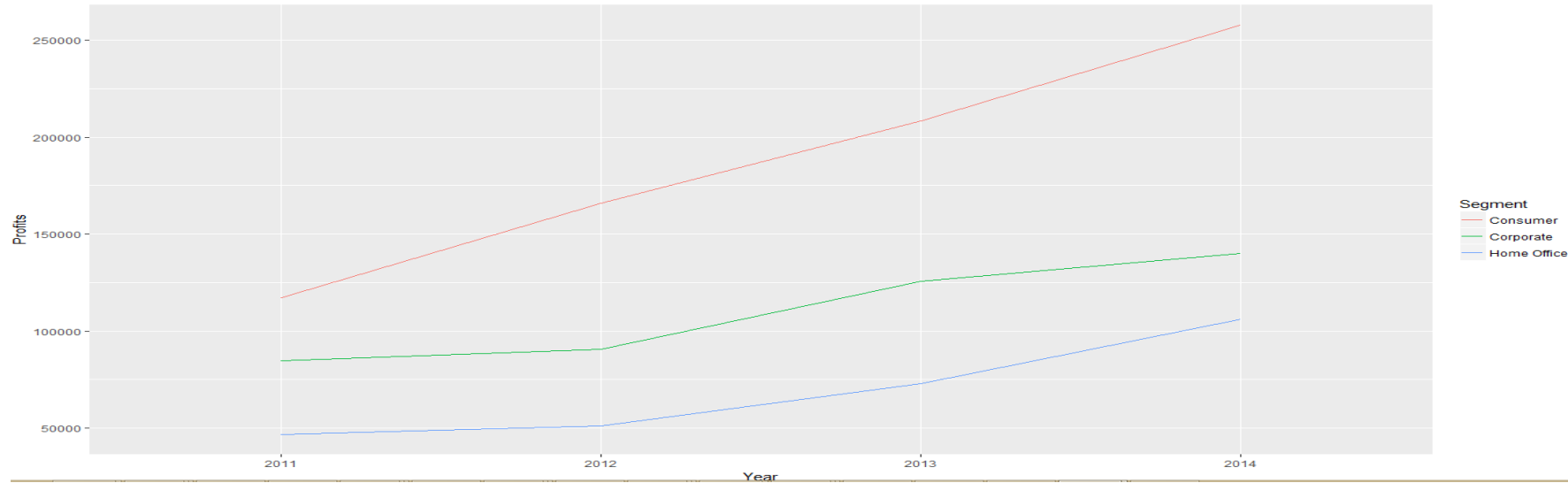


Observation:

Top 3 Profits Market Areas –

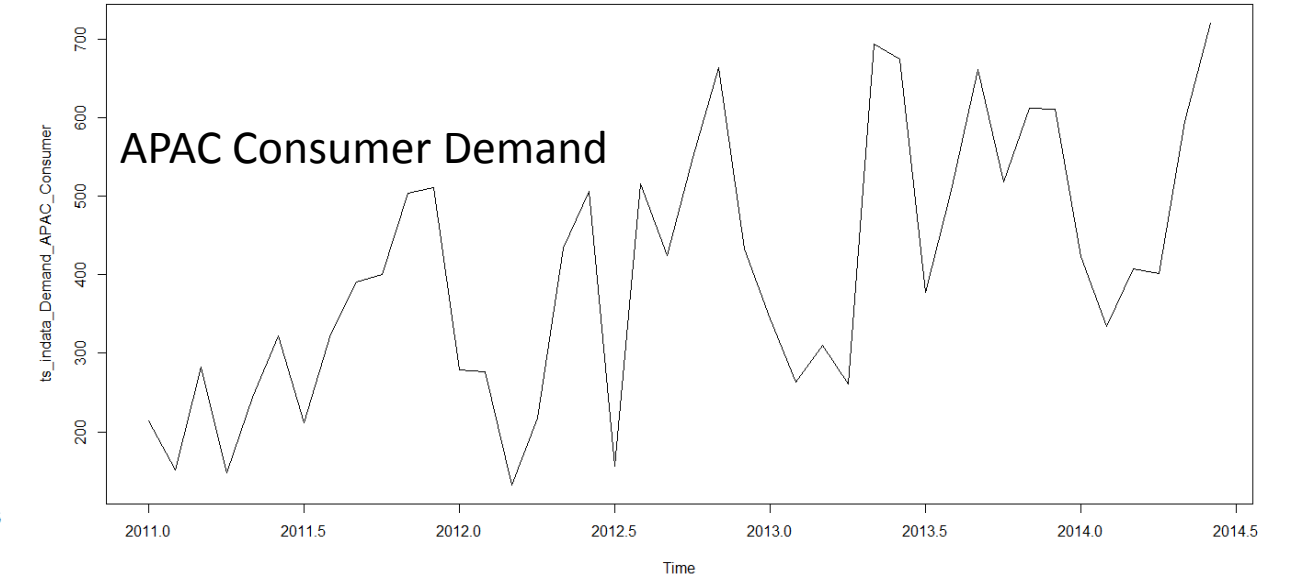
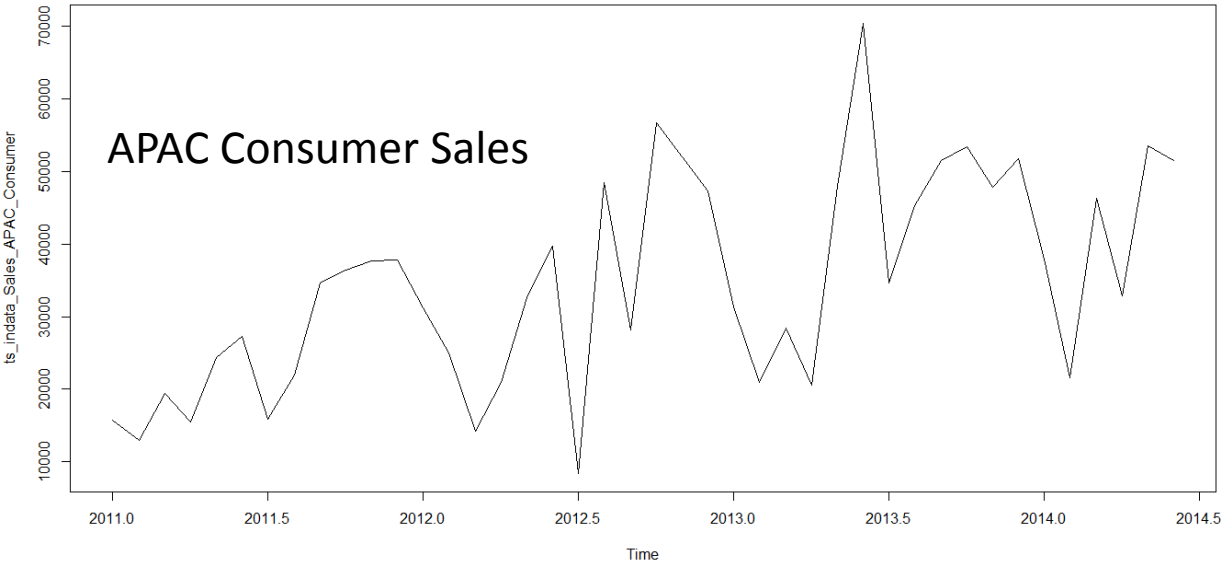
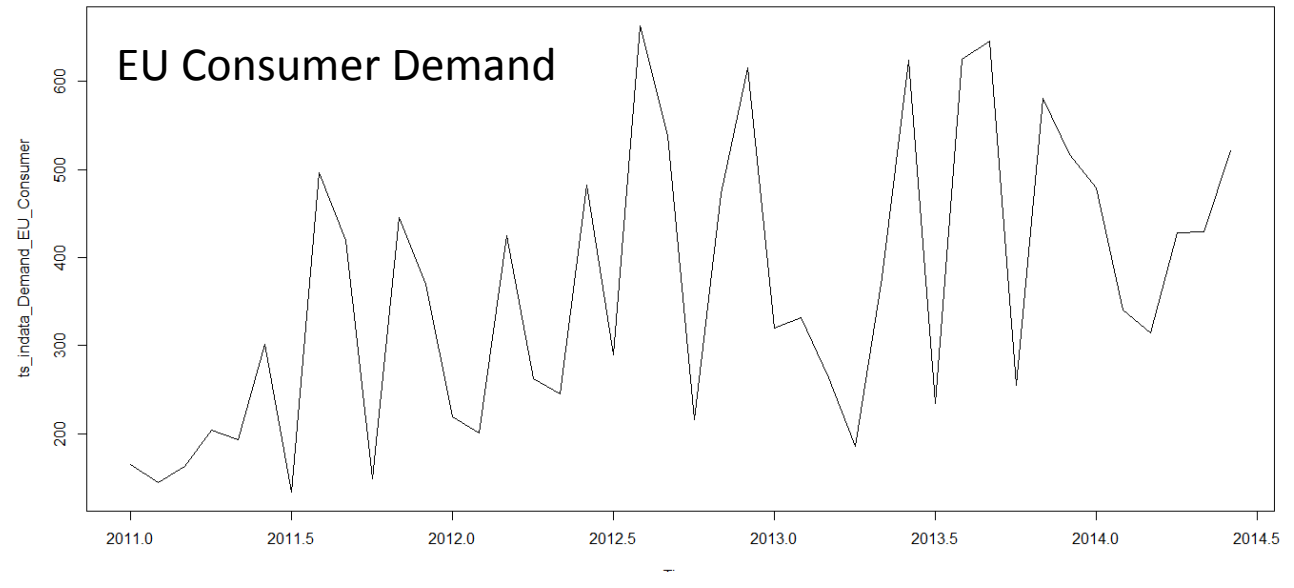
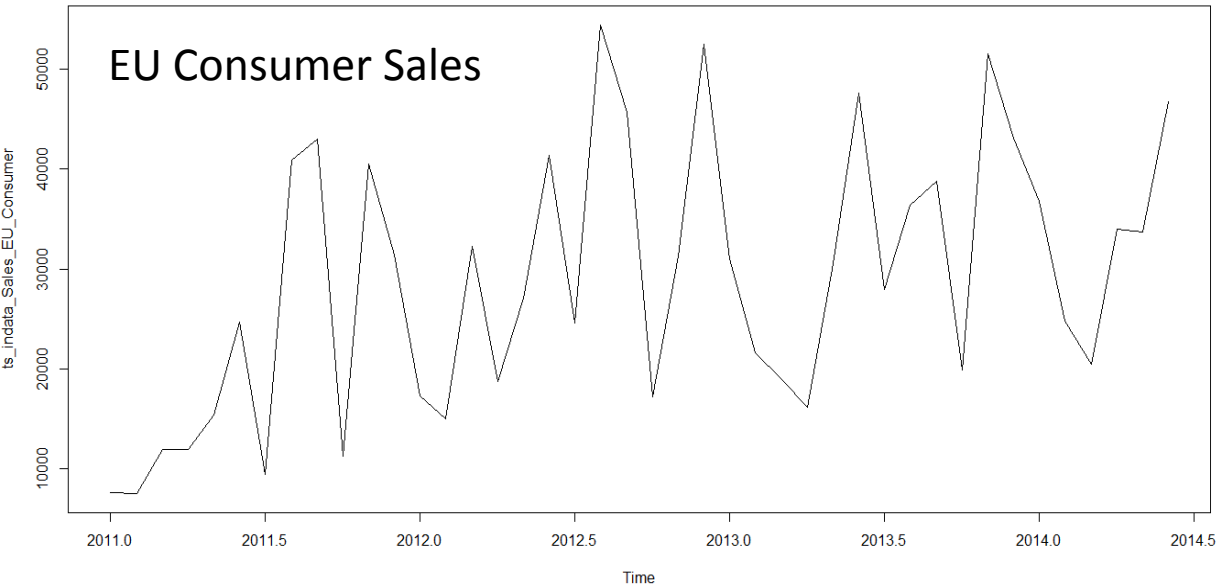
1. APAC
2. EU
3. US

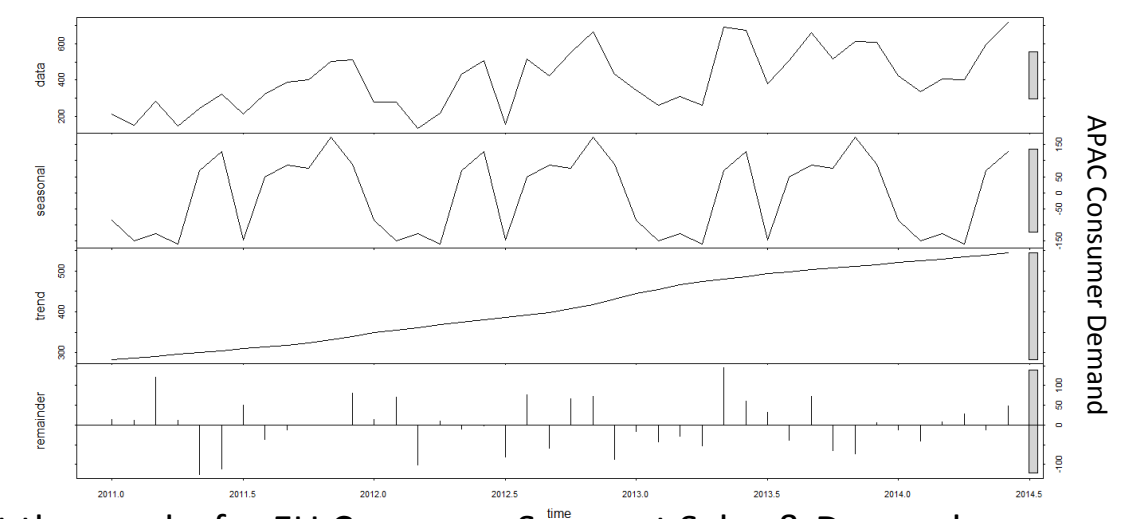
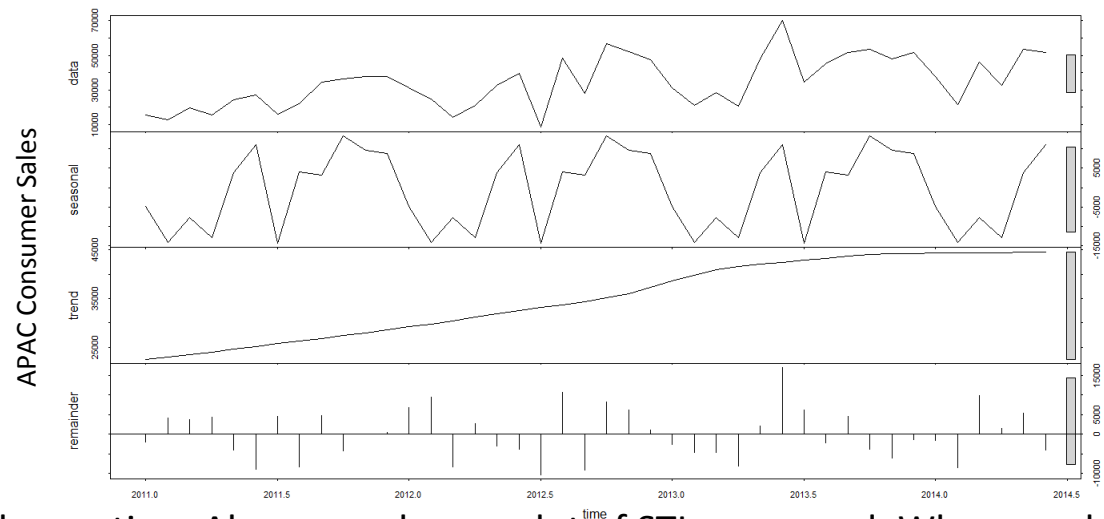
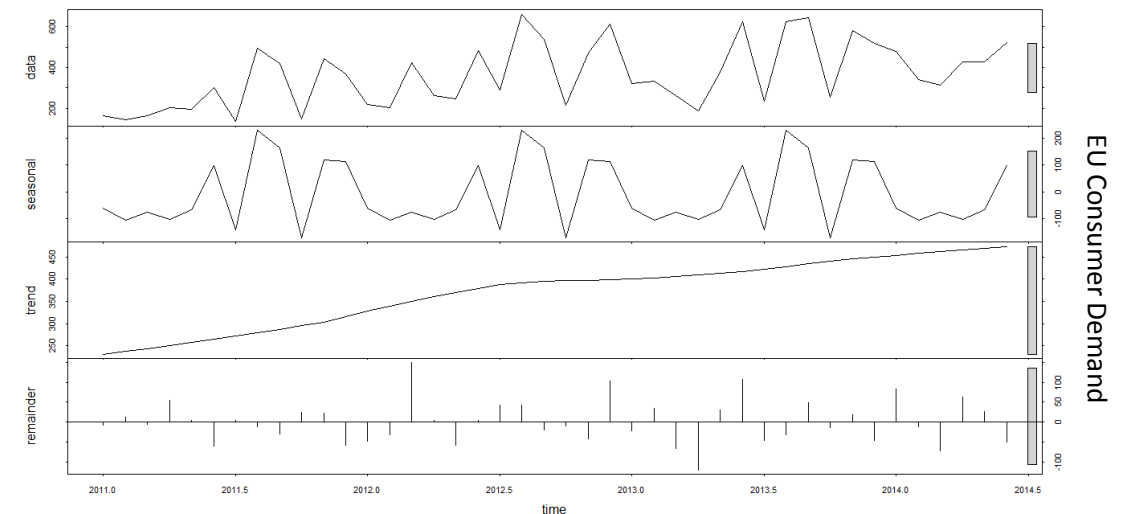
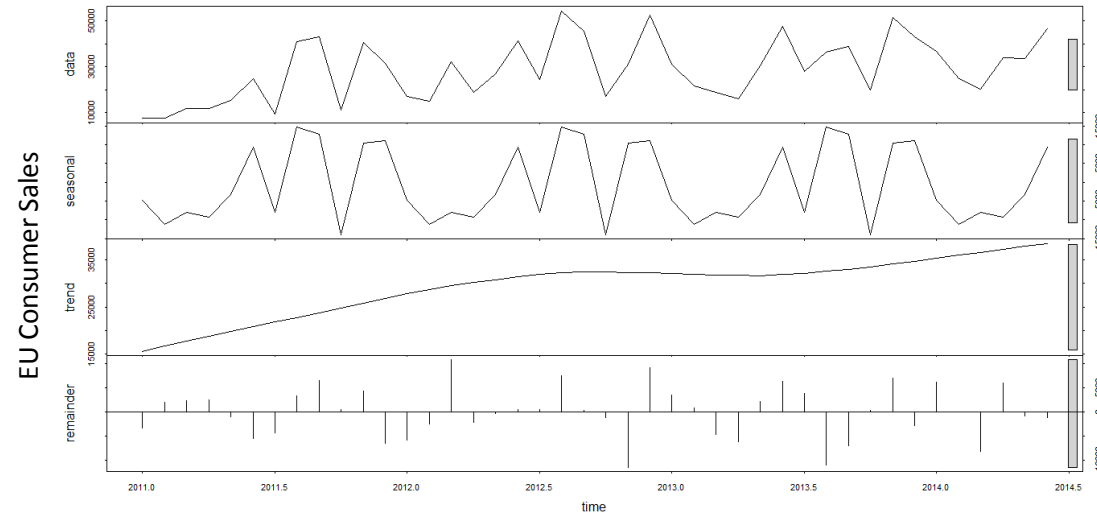
Profits Across Segments From 2011 to 2014



Observation:

Consumer is the most profitable segment and it has been consistently increasing over the years.



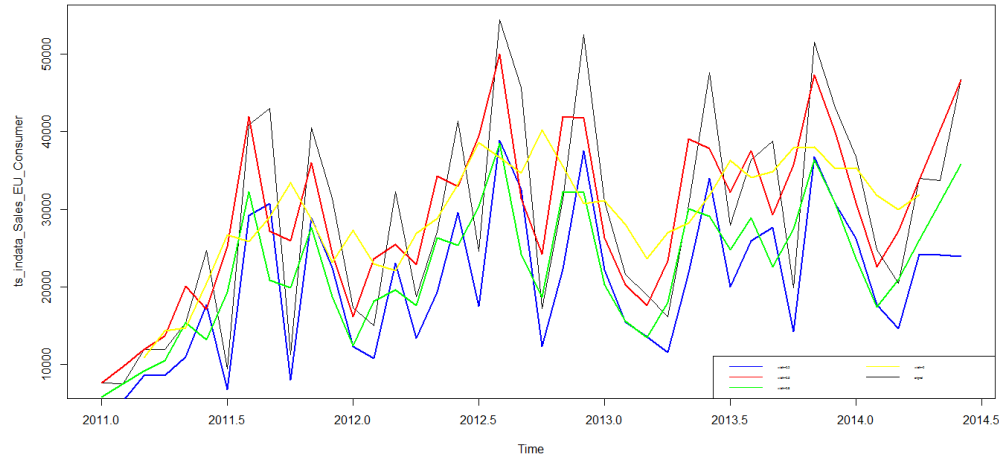


Observation: Above graphs are plot of STL command. When you look at the graphs for EU Consumer Segment Sales & Demand we can see that the **Additive model** fits well for the given time series.

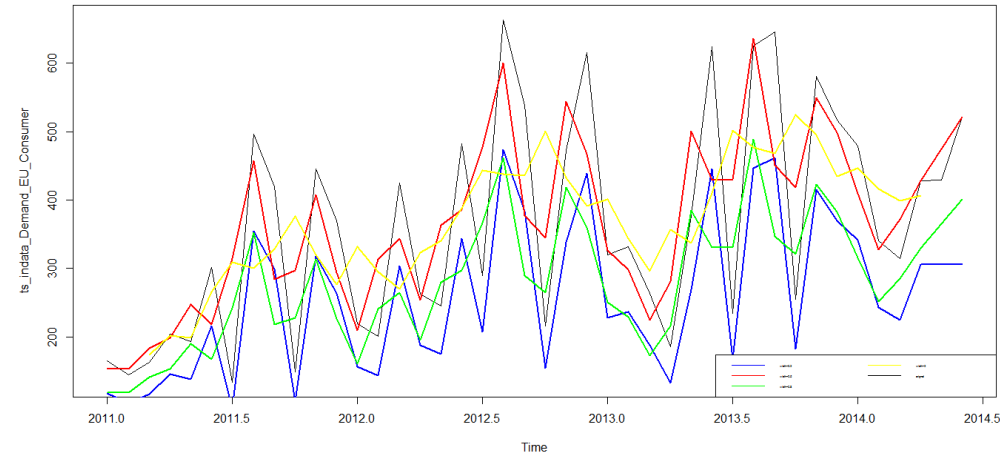
Various components of the time series Trend, Seasonality and noise is present in the time series. Similarly same conclusion can be drawn for APAC Consumer Sales & Demand Data.

## Model Building & Evaluation

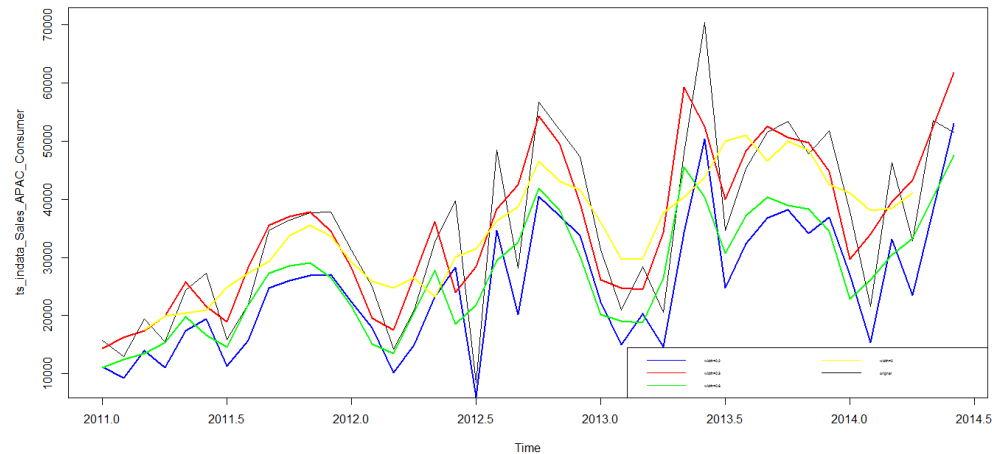
smooth time series : ts\_indata\_Sales\_EU\_Consumer



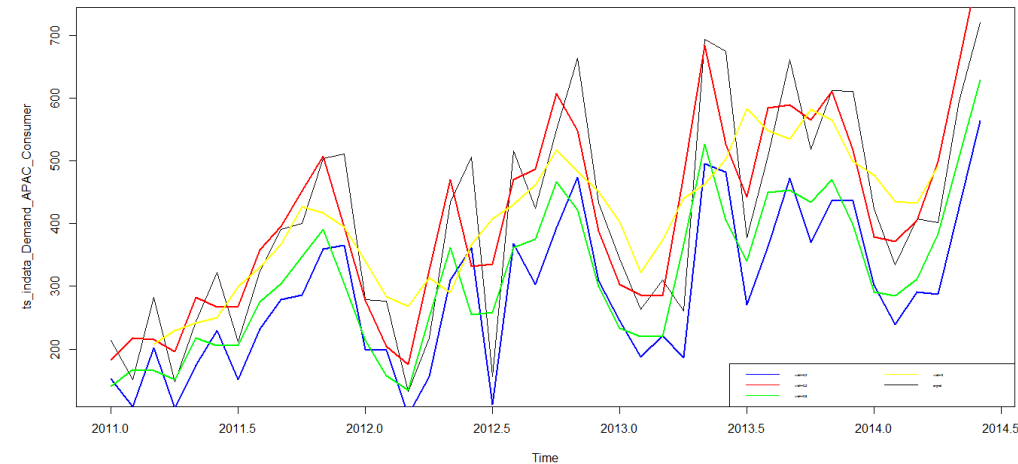
smooth time series : ts\_indata\_Demand\_EU\_Consumer



smooth time series : ts\_indata\_Sales\_APAC\_Consumer



smooth time series : ts\_indata\_Demand\_APAC\_Consumer



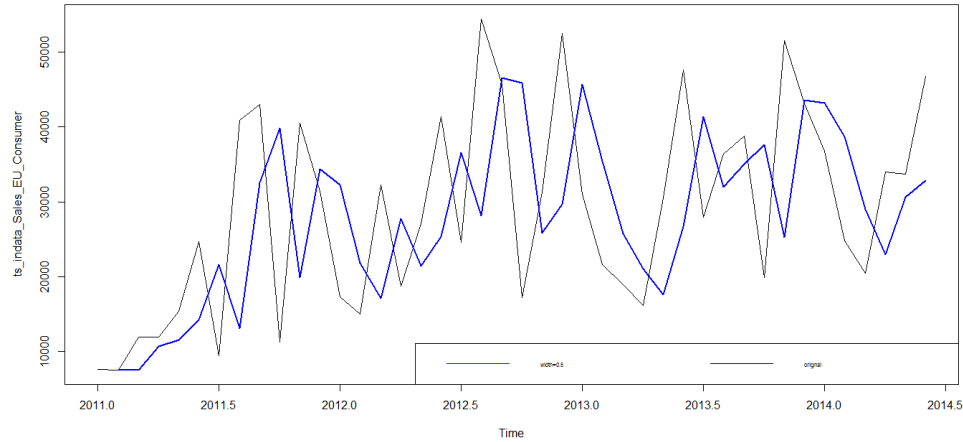
**Observation:**  
By plotting the EU Sales and Demand time series data across different values of Smoothing

- 0.2 (blue),
- 0.5 (red),
- 0.8 (green) and
- 2 (yellow);

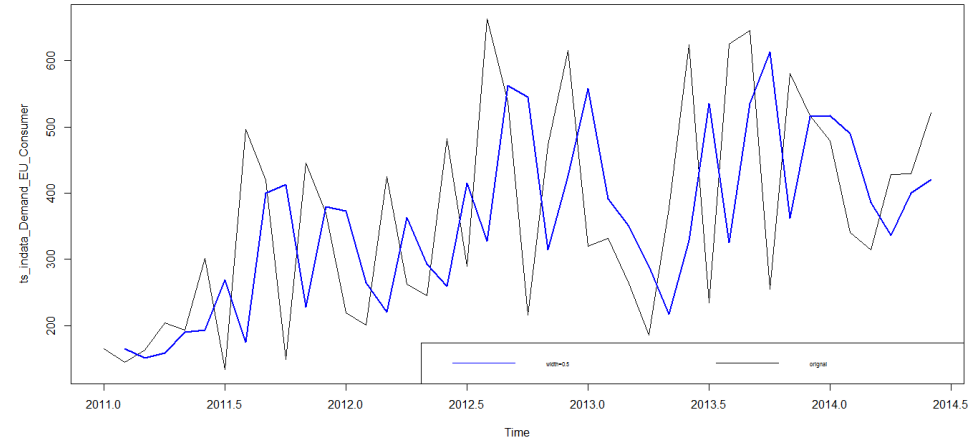
we can see that the 0.5 is ideal width to be used for smoothing as it still retains the trend & seasonality of the original data set.



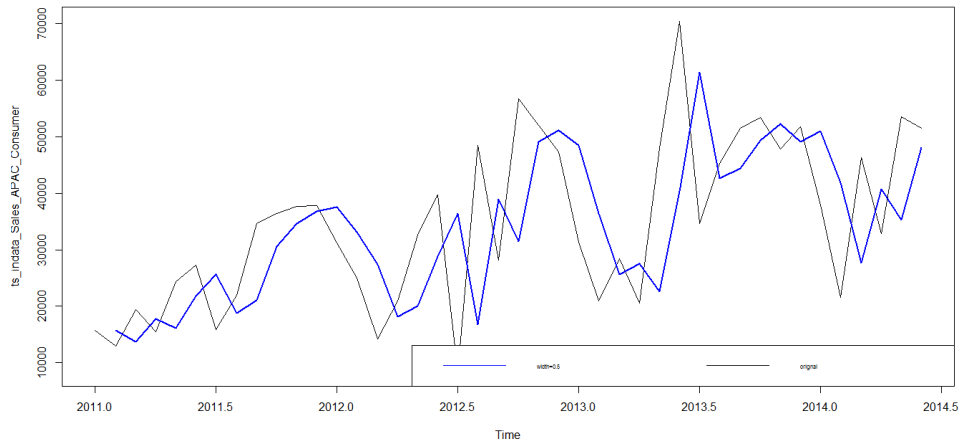
smooth time series : ts\_indata\_Sales\_EU\_Consumer



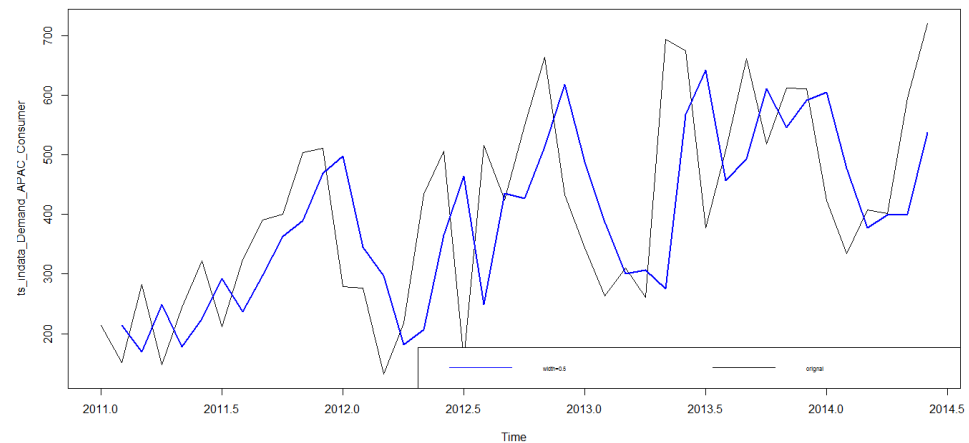
smooth time series : ts\_indata\_Demand\_EU\_Consumer



smooth time series : ts\_indata\_Sales\_APAC\_Consumer



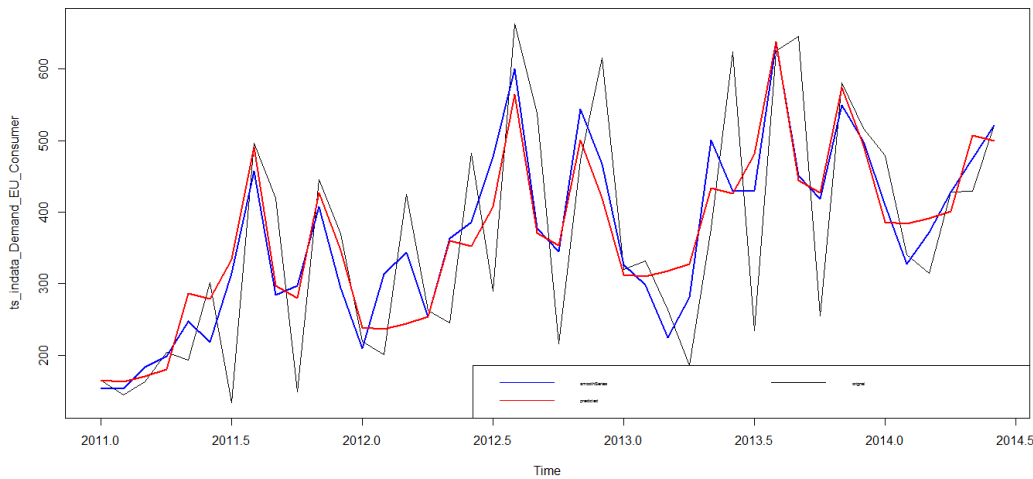
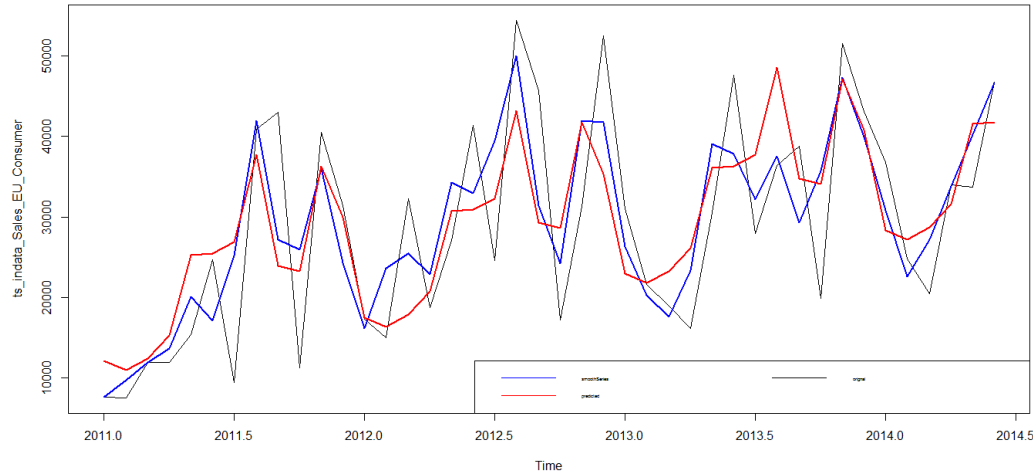
smooth time series : ts\_indata\_Demand\_APAC\_Consumer



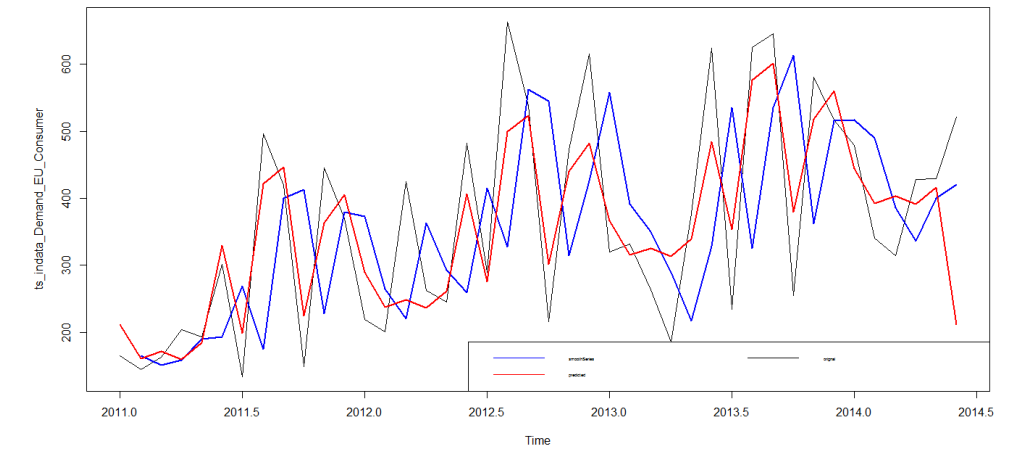
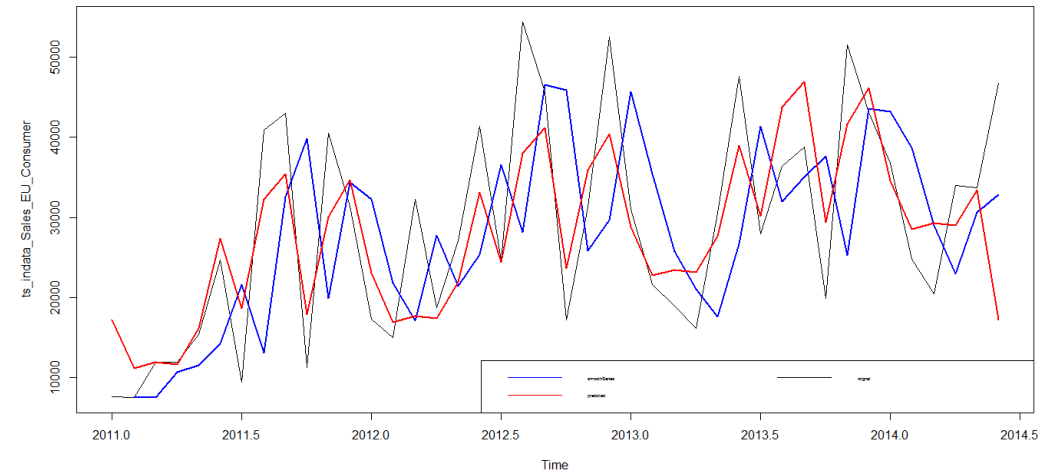
## Observation:

Since we know width of 0.5 is the ideal value, same value can be used for Exponential Smoothing as well. Clearly the smoothened plot is having the trend and seasonality of the data set.

## MOVING AVERAGE



## EXPONENTIAL SMOOTHING



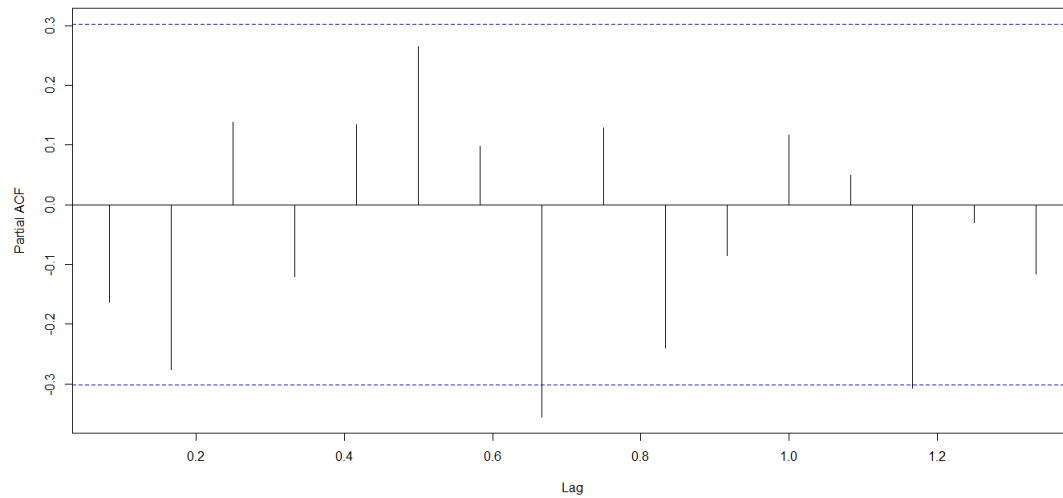
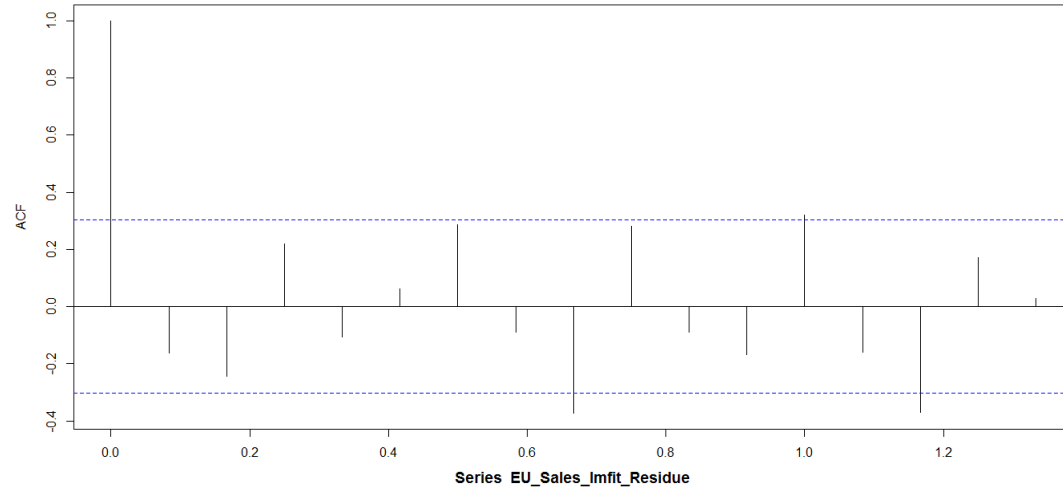
Observation: The plot across graphs shows that the generated model using classical decomposition w/h Moving Average & Exponential Smoothing fits closely to the original series.

EU Demand

EU Sales

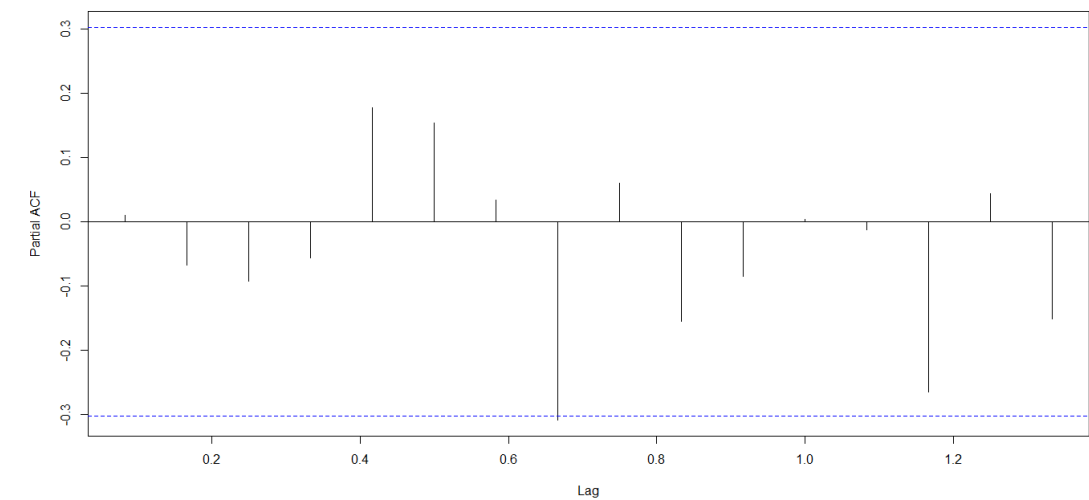
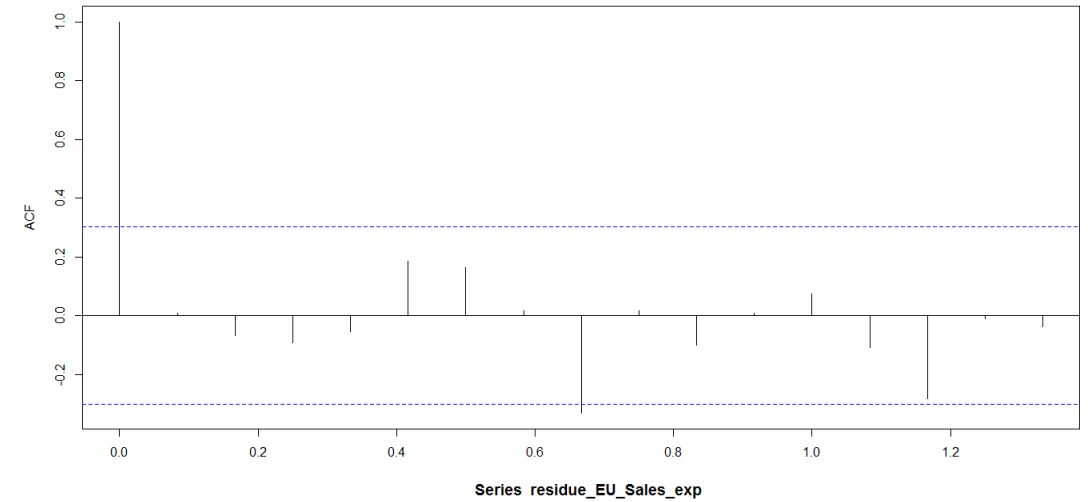
## MOVING AVERAGE

Series EU\_Sales\_lmfit\_Residue



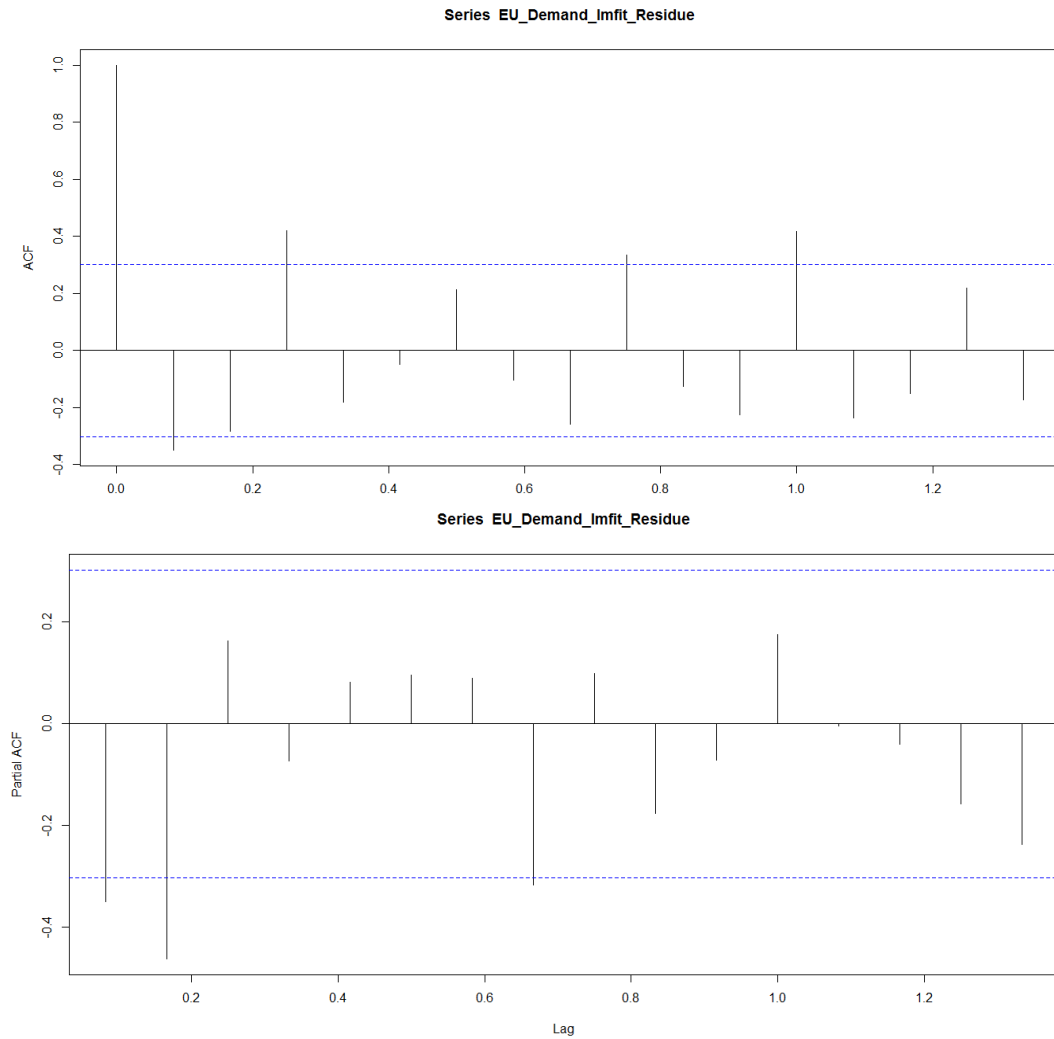
## EXPONENTIAL SMOOTHING

Series residue\_EU\_Sales\_exp

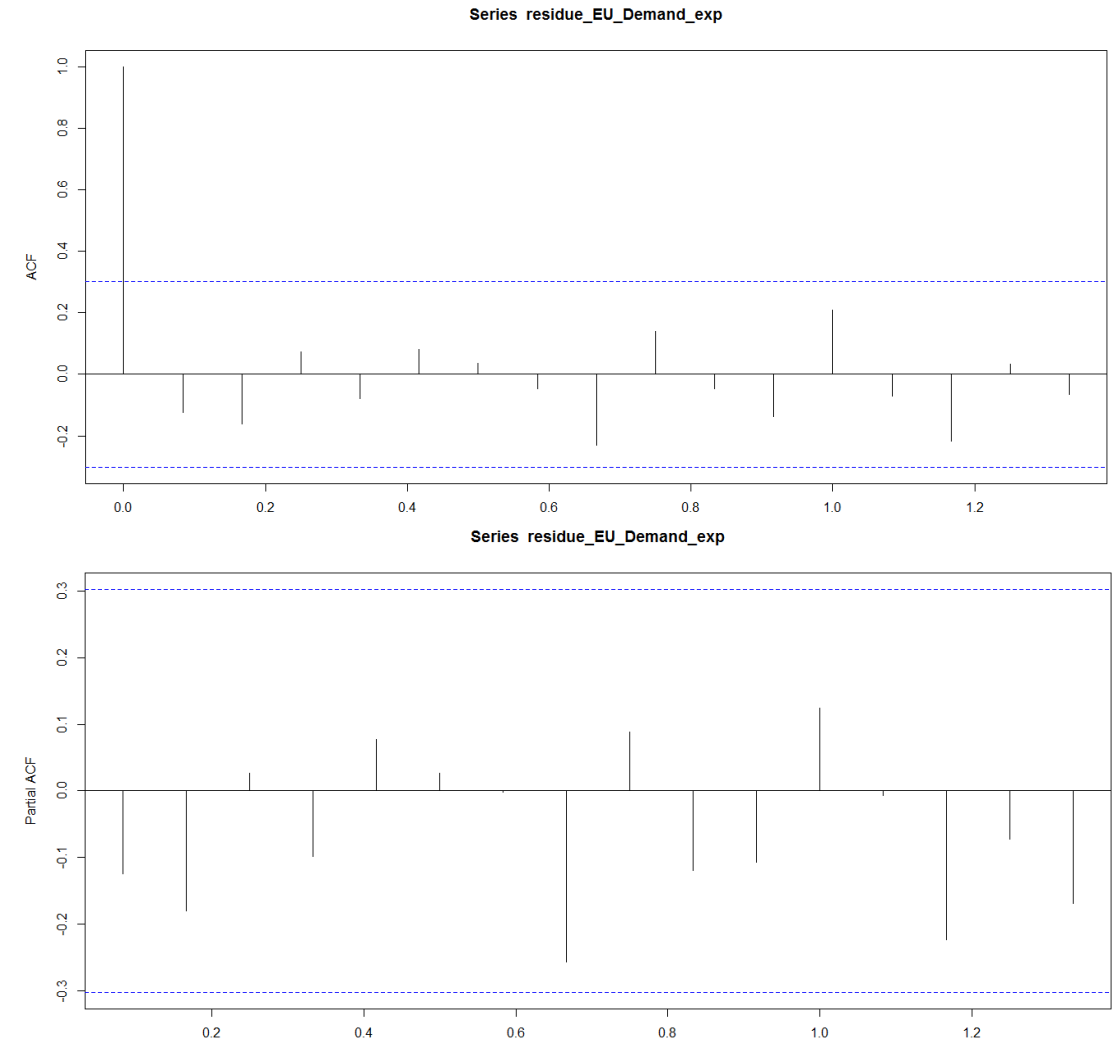


Observation: ACF and PACF plots are within the margins which indicate that the residual series is stationary.

## MOVING AVERAGE



## EXPONENTIAL SMOOTHING



Observation: ACF and PACF plots are within the margins which indicate that the residual series is stationary.



## EU Sales

## Forecasted Plots

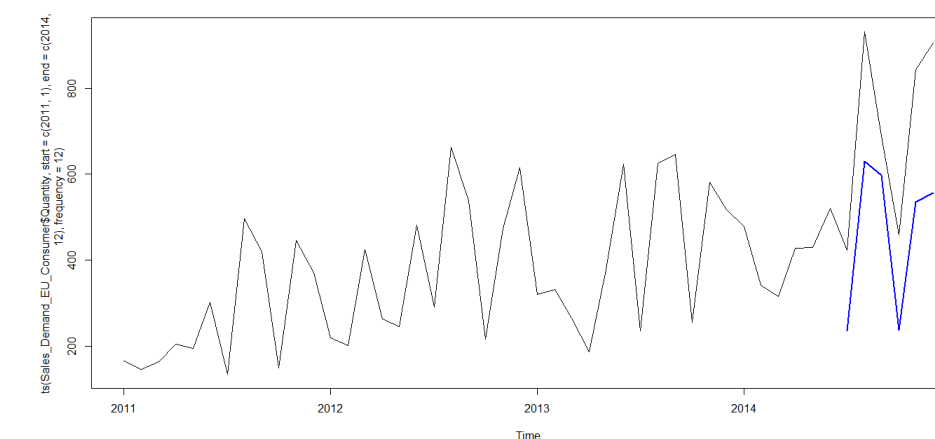
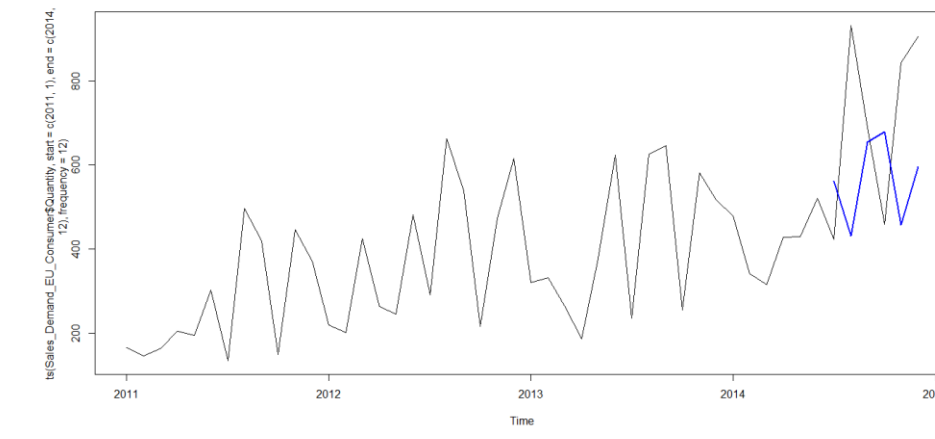
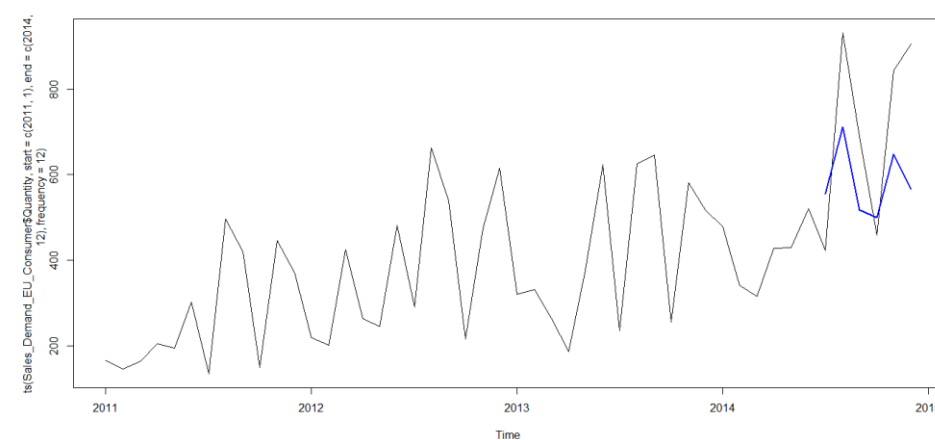
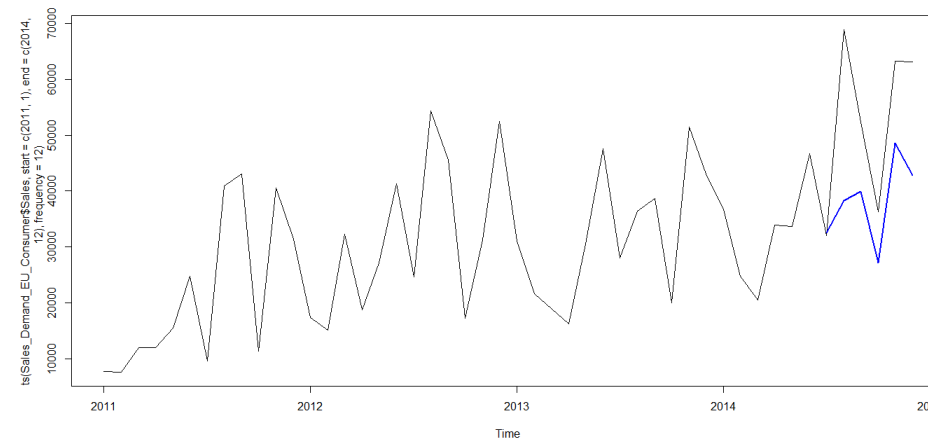
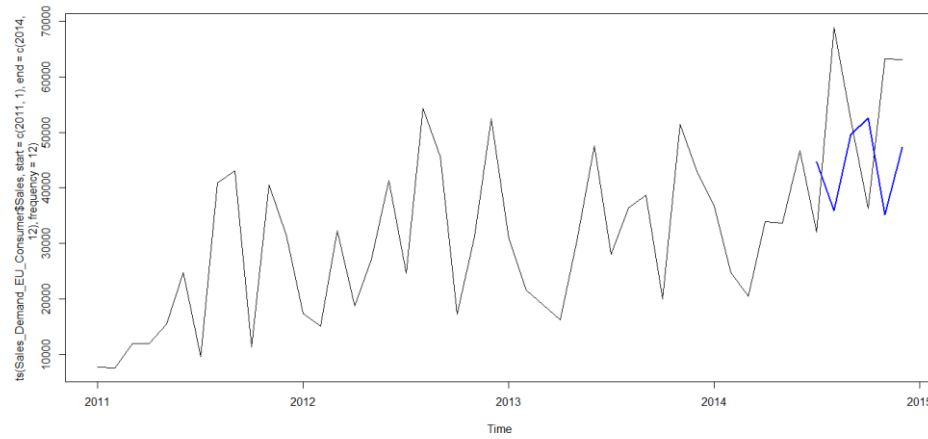
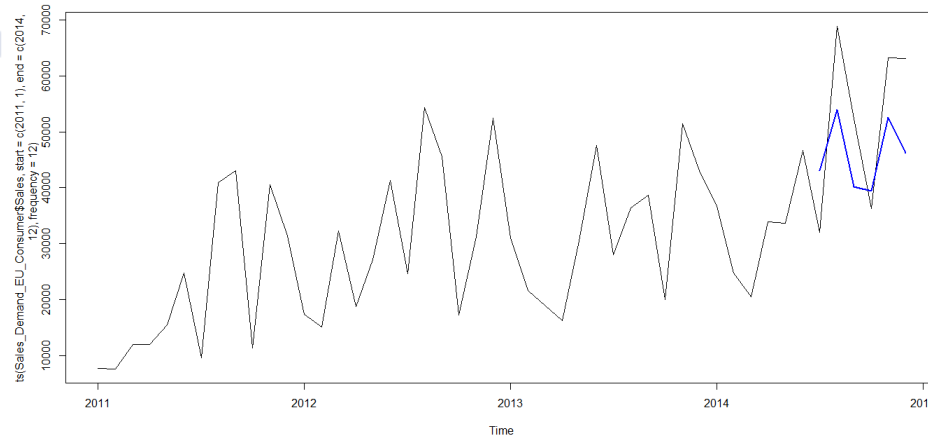
## EU Demand



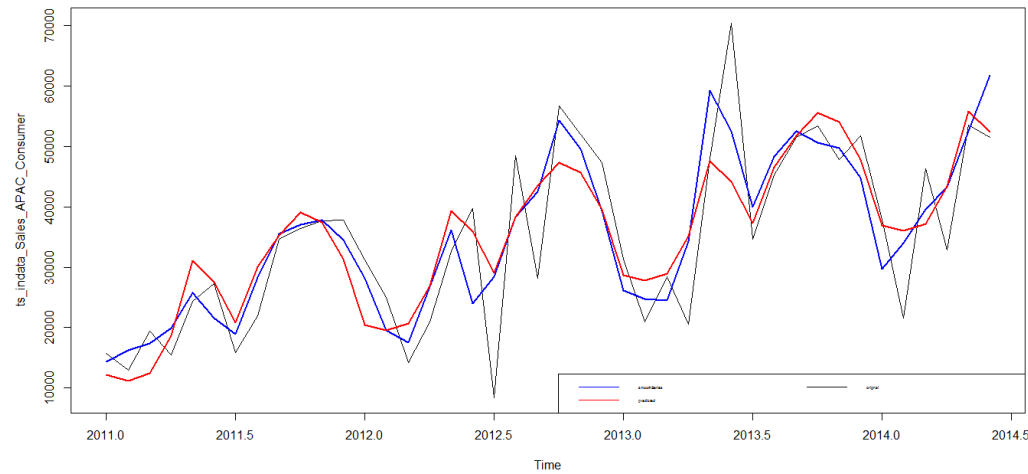
MOVING  
AVERAGE

EXPONENTIAL  
SMOOTHING

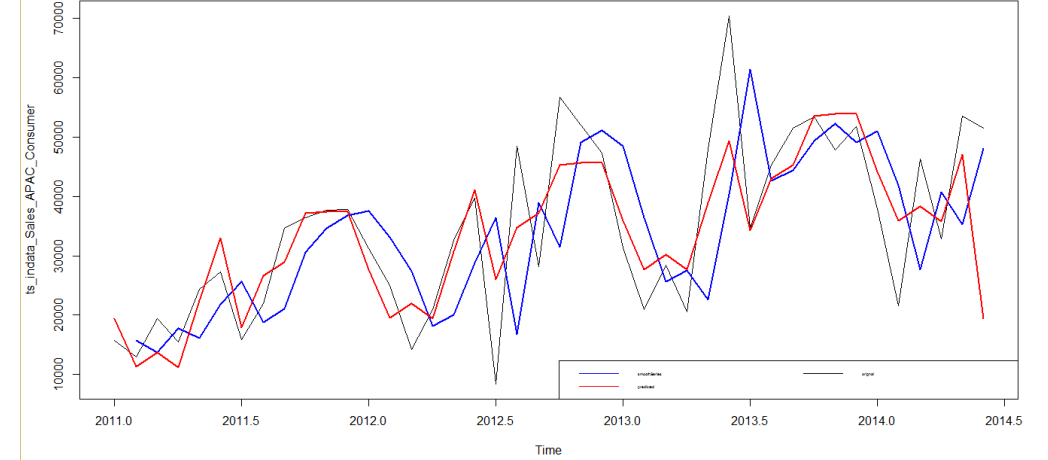
AUTO ARIMA



## MOVING AVERAGE

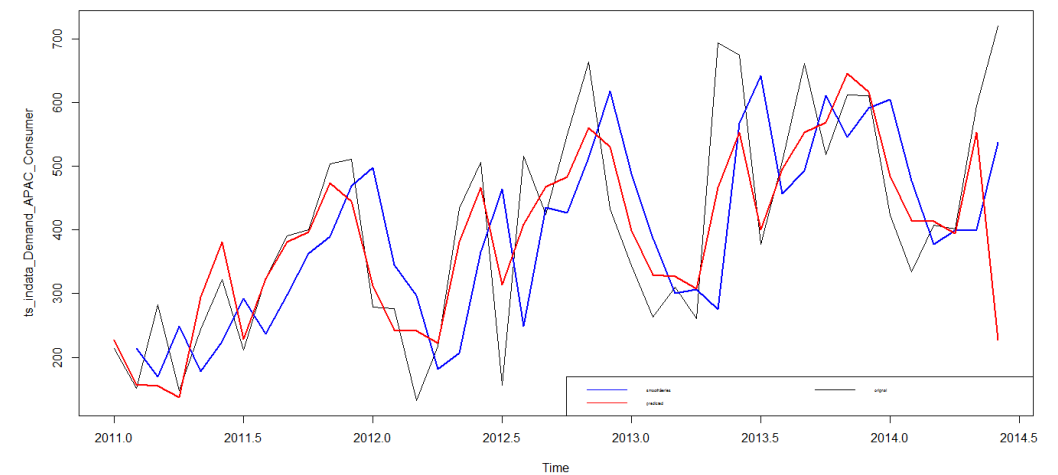
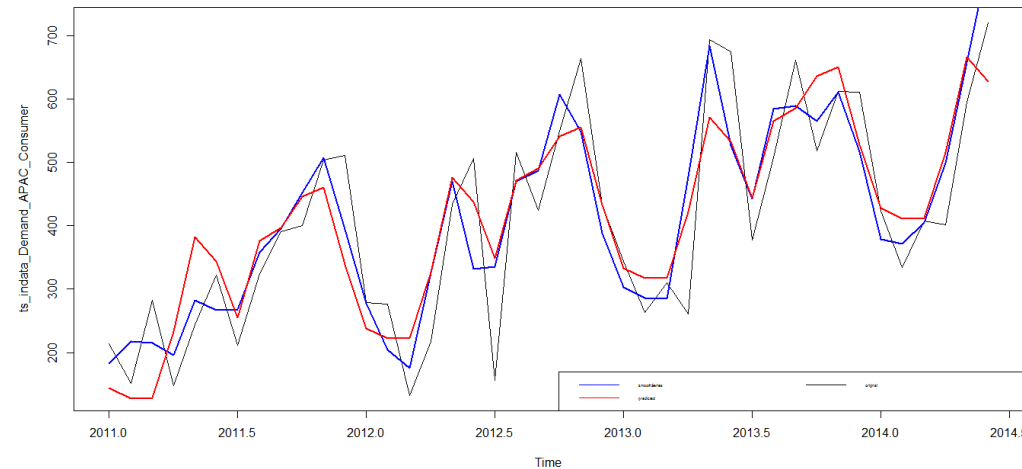


## EXPONENTIAL SMOOTHING



APAC Sales

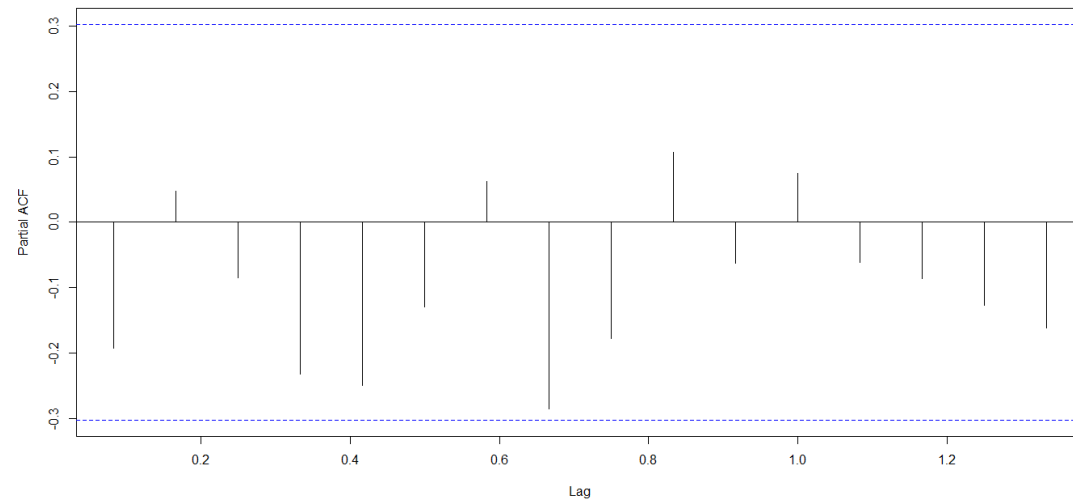
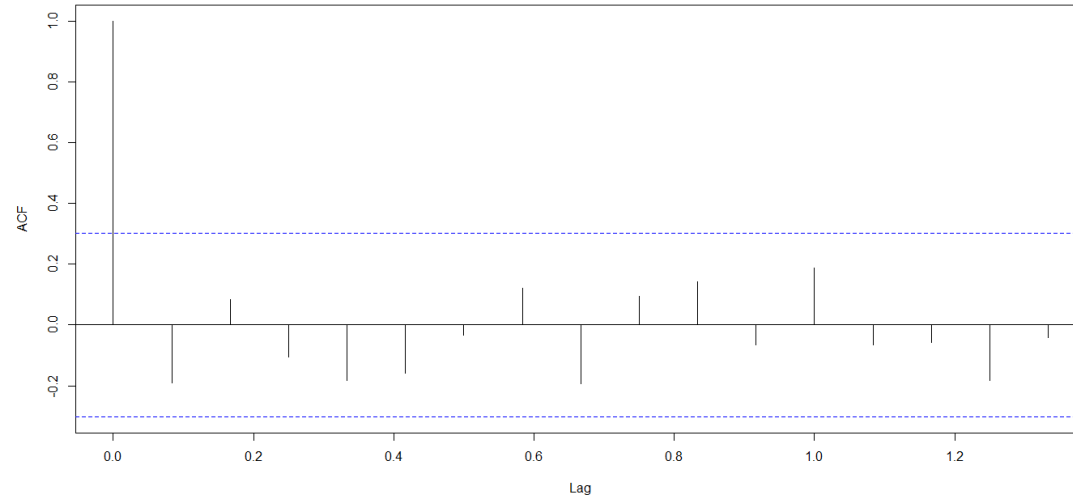
APAC Demand



Observation: The plot across graphs shows that the generated model using classical decomposition w/h Moving Average & Exponential Smoothing fits closely to the original series.

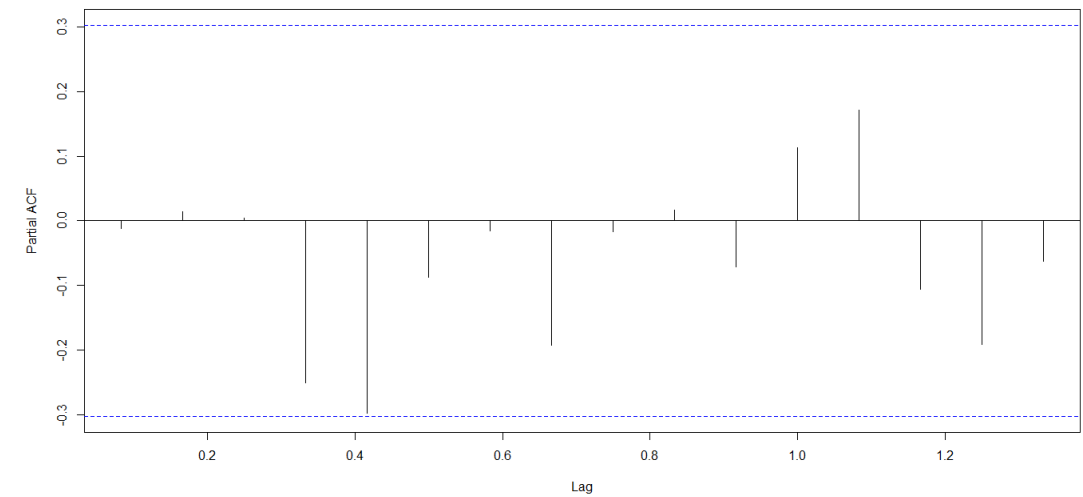
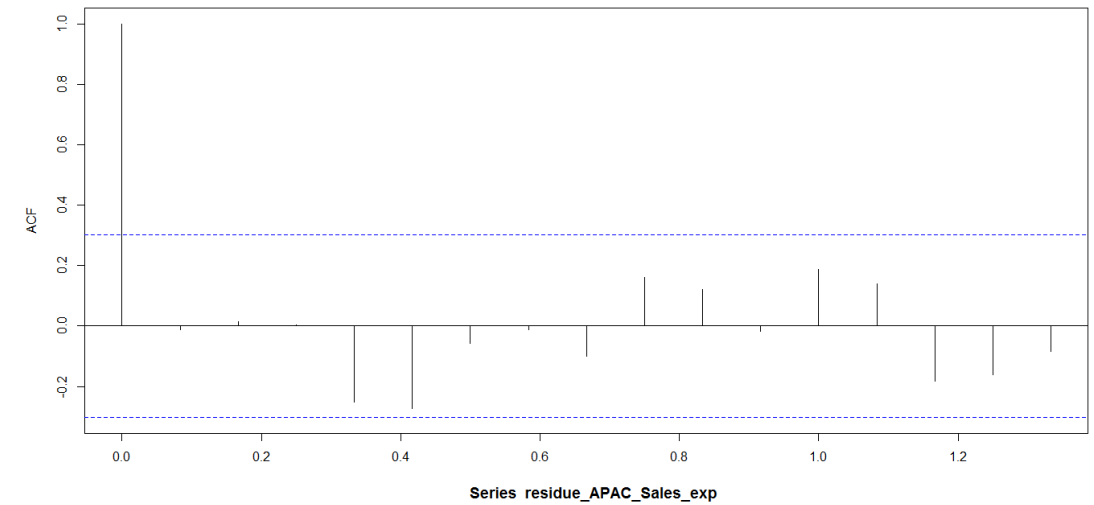
## MOVING AVERAGE

Series APAC\_Sales\_lmfit\_Residue



## EXPONENTIAL SMOOTHING

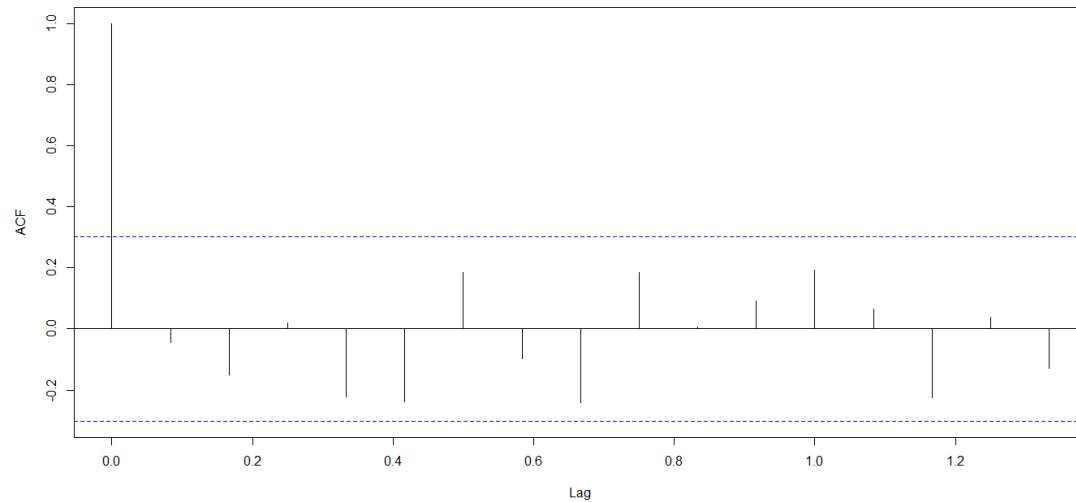
Series residue\_APAC\_Sales\_exp



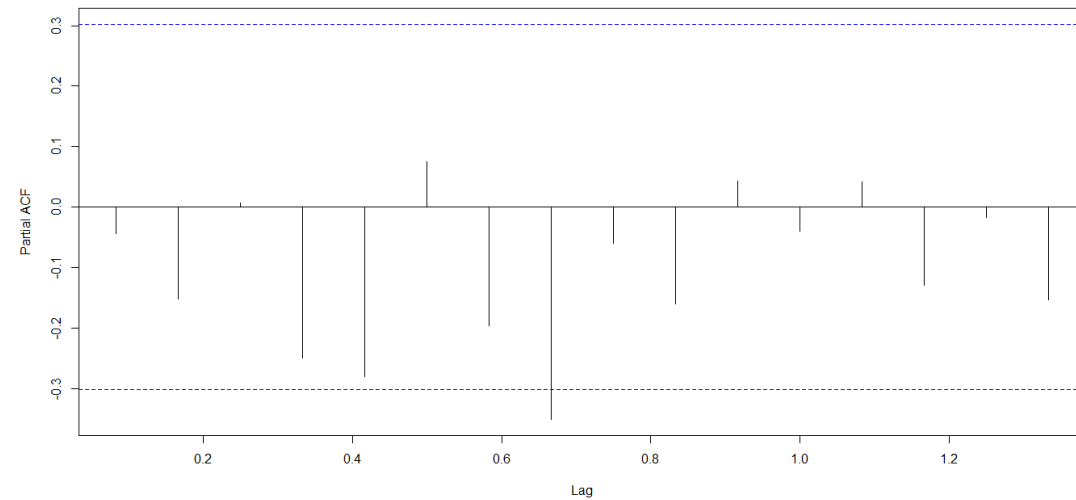
Observation: ACF and PACF plots are within the margins which indicate that the residual series is stationary.

## MOVING AVERAGE

Series APAC\_Demand\_lmfit\_Residue

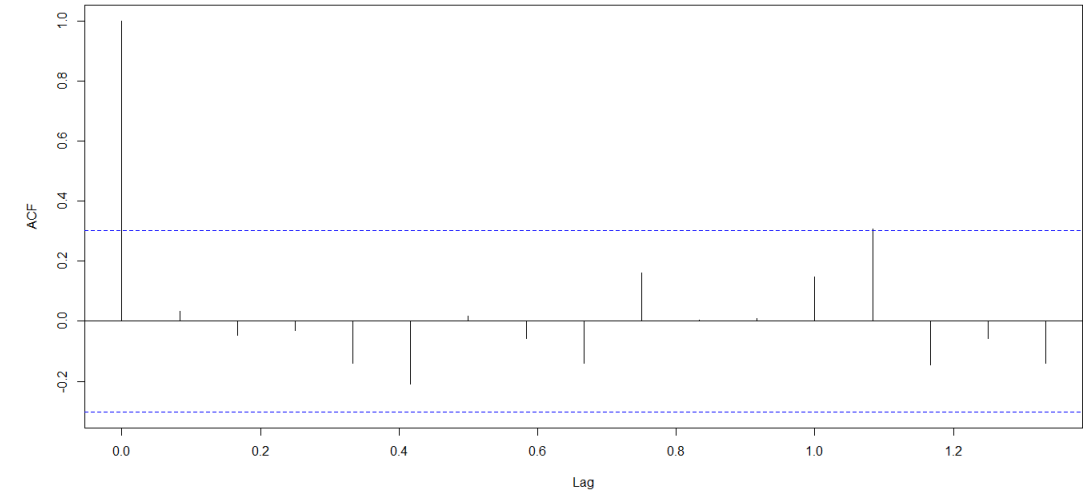


Series APAC\_Demand\_lmfit\_Residue

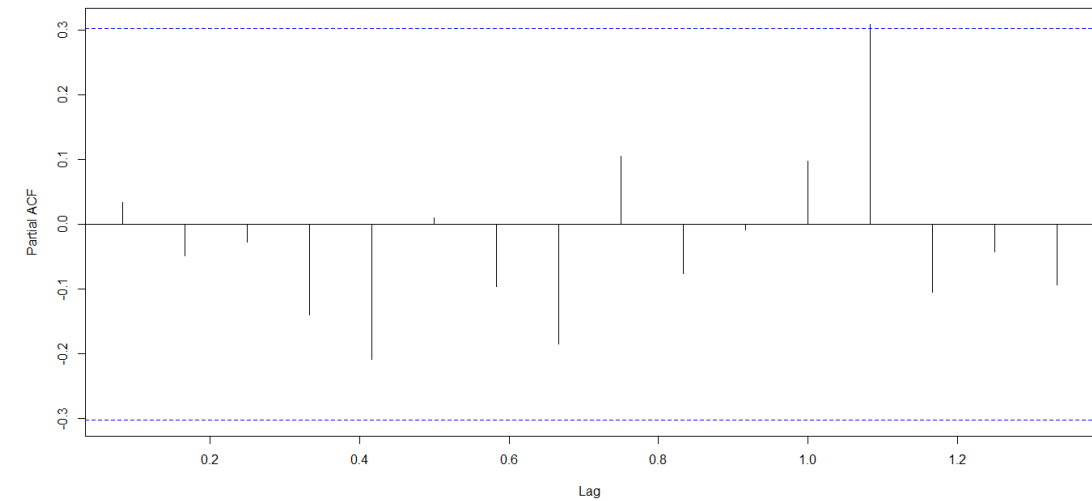


## EXPONENTIAL SMOOTHING

Series residue\_APAC\_Demand\_exp



Series residue\_APAC\_Demand\_exp



Observation: ACF and PACF plots are within the margins which indicate that the residual series is stationary.





# APAC Sales

# Forecasted Plots

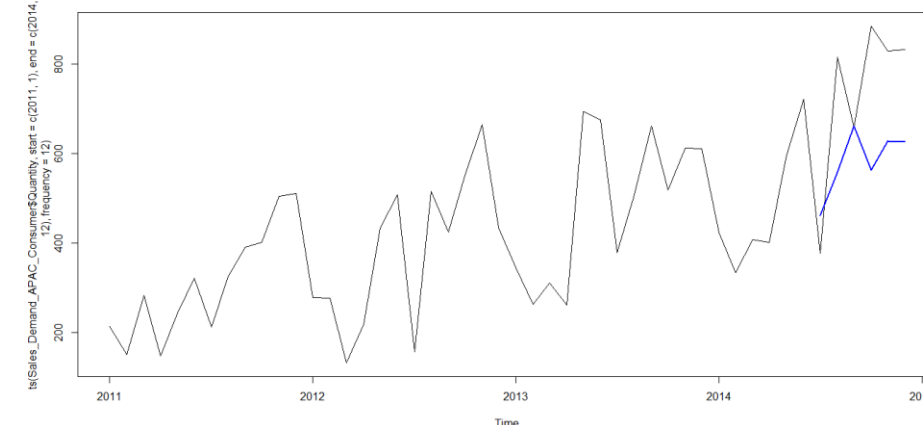
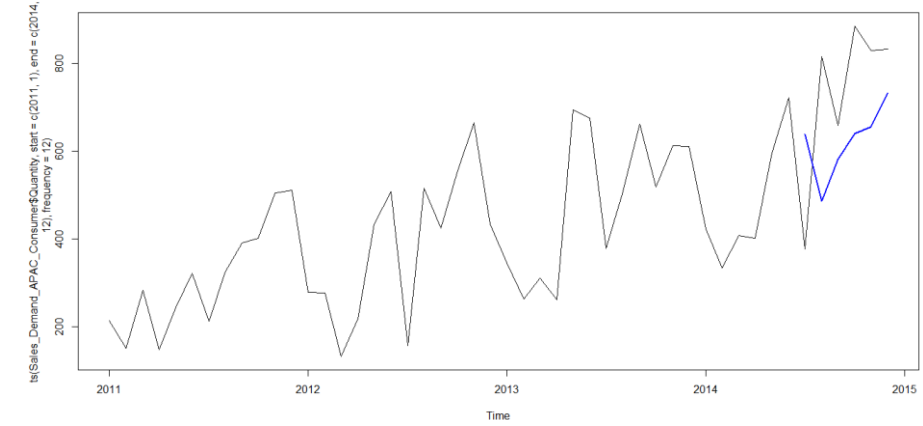
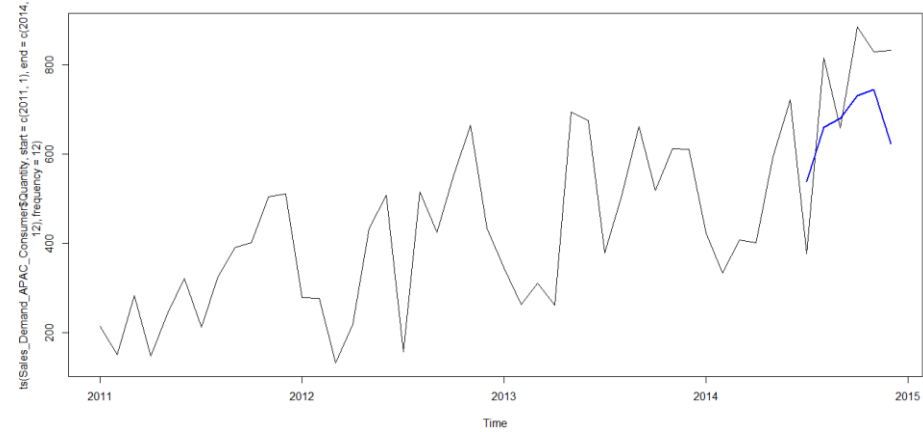
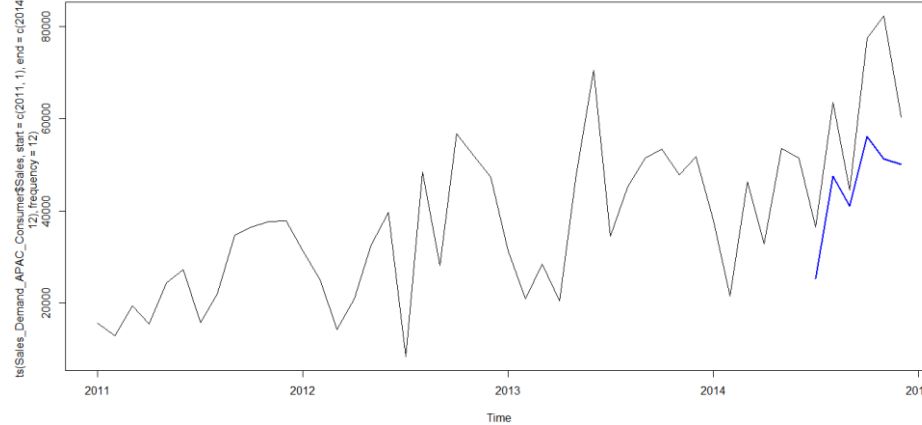
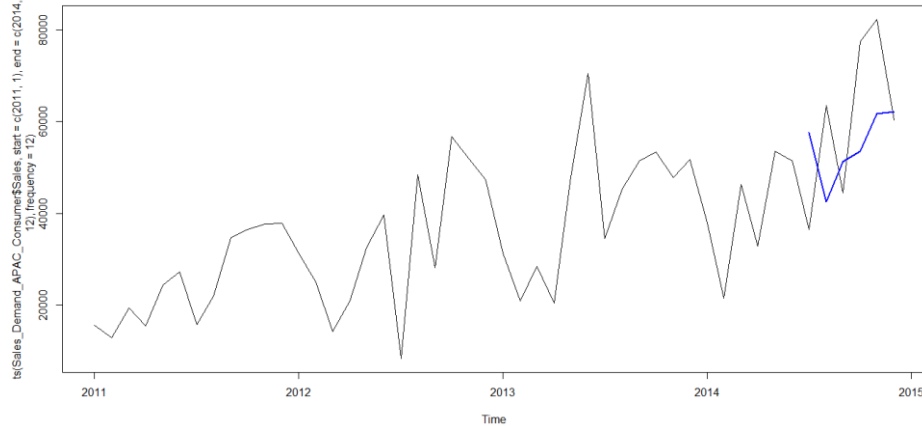
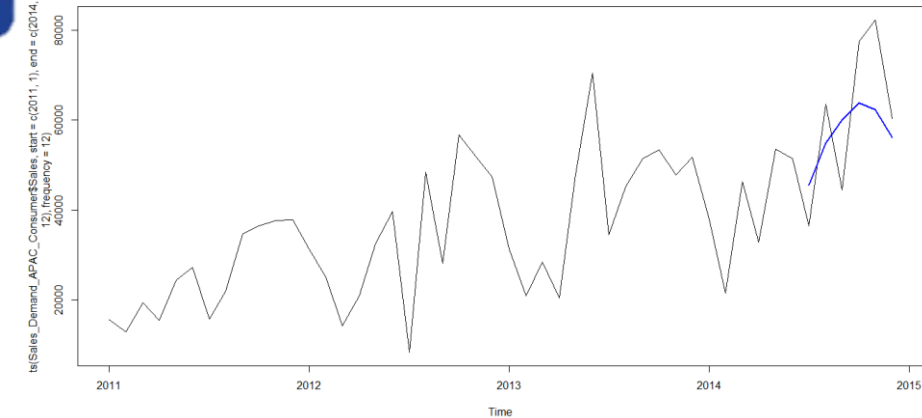
# APAC Demand



MOVING  
AVERAGE

EXPONENTIAL  
SMOOTHING

AUTO ARIMA



### – Sales EU Consumer –

1. For the Classical Decomposition Using Moving Average the MAPE Value # 22.029
2. For the Classical Decomposition Using Exponential Smoothing the MAPE Value # 34.528
3. For the Classical Decomposition Using Exponential Smoothing + Polynomial and Sinusoidal Model the MAPE Value # 77.83
4. For the Auto Arima Model the MAPE Value # **19.23**

So the best model based on the MAPE values seems to be Auto Arima, hence we will use Auto Arima to forecast the Sales.

### – Demand EU Consumer –

1. For the Classical Decomposition Using Moving Average the MAPE Value # **24.833**
2. For the Classical Decomposition Using Exponential Smoothing the MAPE Value # 36.571
3. For the Classical Decomposition Using Exponential Smoothing + Polynomial and Sinusoidal Model the MAPE Value # 99.635
4. For the Auto Arima Model the MAPE Value # 35.537

So the best model based on the MAPE values seems to be Moving Average, hence we will use Moving Average to forecast the Demand.

### – Sales APAC Consumer –

1. For the Classical Decomposition Using Moving Average the MAPE Value # **20.368**
2. For the Classical Decomposition Using Exponential Smoothing the MAPE Value # 27.454
3. For the Classical Decomposition Using Exponential Smoothing + Polynomial and Sinusoidal Model the MAPE Value # 69.1124
4. For the Auto Arima Model the MAPE Value # 24.300

So the best model based on the MAPE values seems to be Moving average, hence we will use Moving Average to forecast the Sales.

### – Demand APAC Consumer –

1. For the Classical Decomposition Using Moving Average the MAPE Value # **19.678**
2. For the Classical Decomposition Using Exponential Smoothing the MAPE Value # 30.400
3. For the Classical Decomposition Using Exponential Smoothing + Polynomial and Sinusoidal Model the MAPE Value # 99.619
4. For the Auto Arima Model the MAPE Value # 22.152

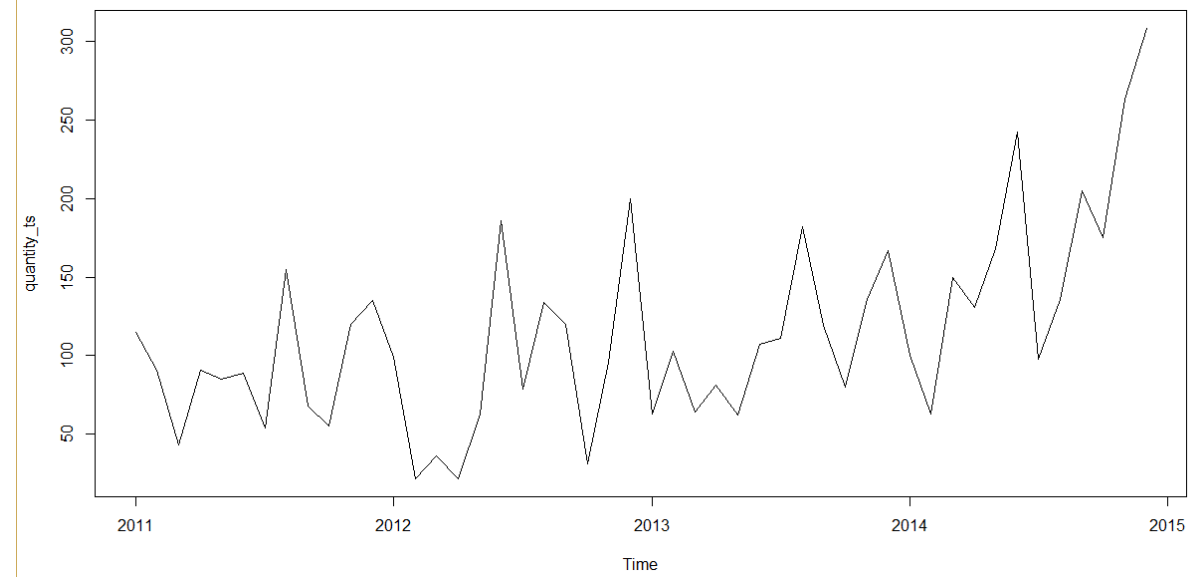
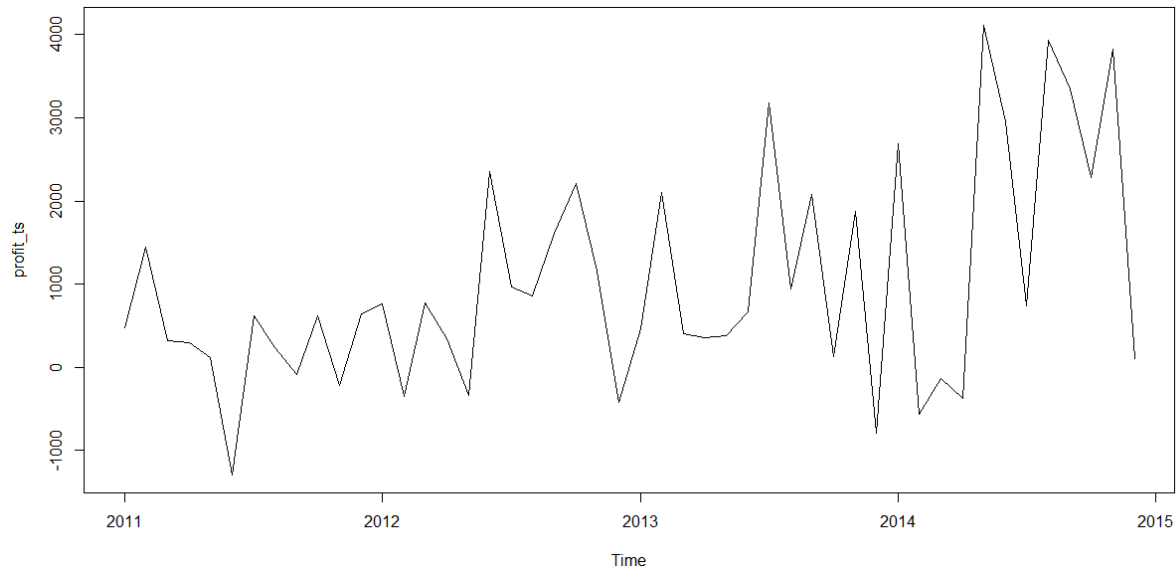
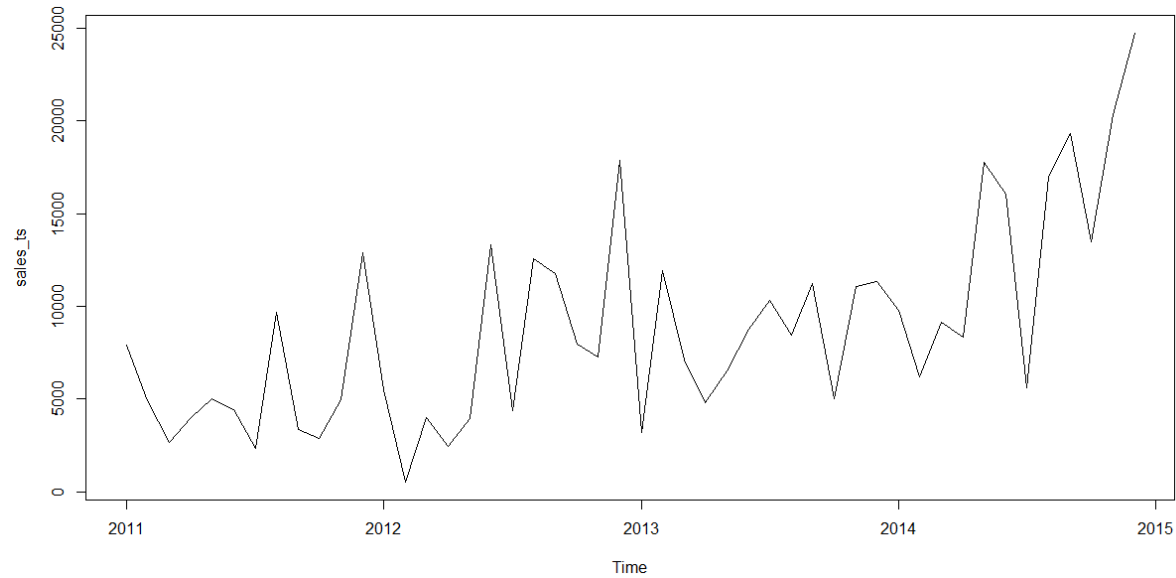
So the best model based on the MAPE values seems to be Moving Average, hence we will use Moving Average to forecast the Demand.

Based on the models chosen below is the approximate forecasted sales & demand for Consumer segment for EU & APAC markets.

EU - Consumer		
Month	Sales	Demand
Jul 2014	~ 32,580	~ 554
Aug 2014	~ 38,294	~ 711
Sep 2014	~ 39,908	~ 517
Oct 2014	~ 27,094	~ 500
Nov 2014	~ 48,598	~ 647
Dec 2014	~ 42,853	~ 566

APAC - Consumer		
Month	Sales	Demand
Jul 2014	~ 45,602	~ 538
Aug 2014	~ 54,851	~ 660
Sep 2014	~ 60,010	~ 680
Oct 2014	~ 63,874	~ 730
Nov 2014	~ 62,248	~ 745
Dec 2014	~ 56,086	~ 622

## APPENDIX

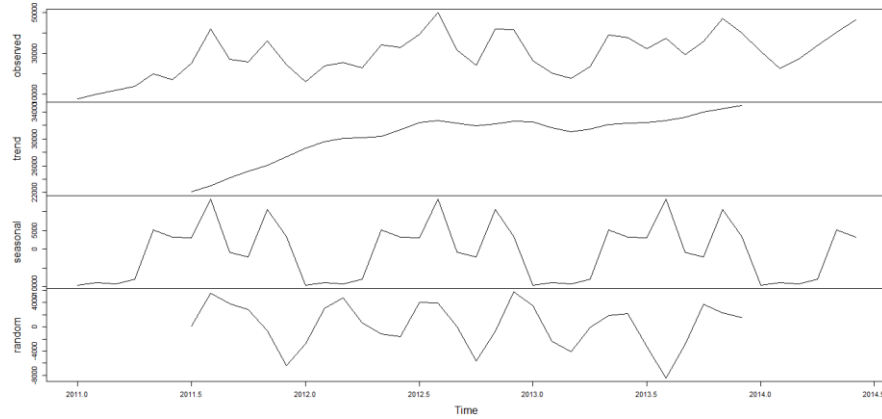


## Observation:

1. The time series plot shows that the data has good amount of trend and seasonality
2. Between year 2011 and 2012 the company profits were negative and same trend is visible in 2013 and 2014.

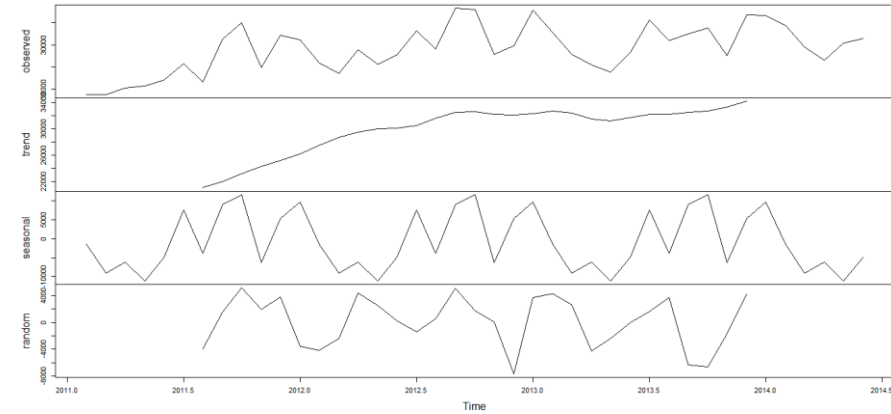
## MOVING AVERAGE

Decomposition of additive time series

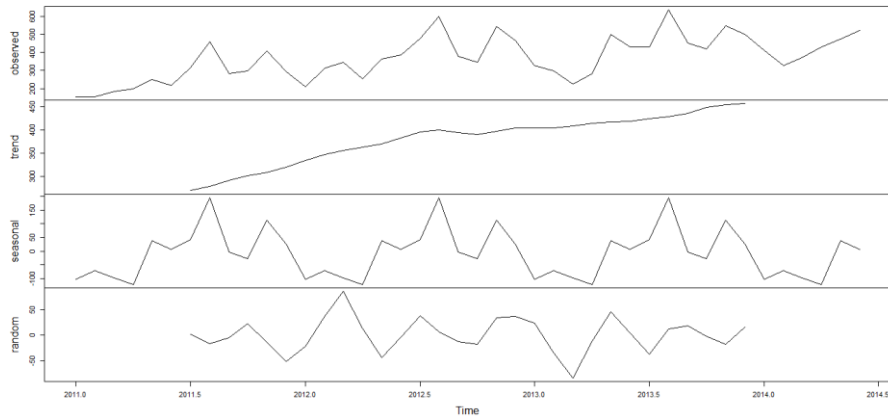


## EXPONENTIAL SMOOTHING

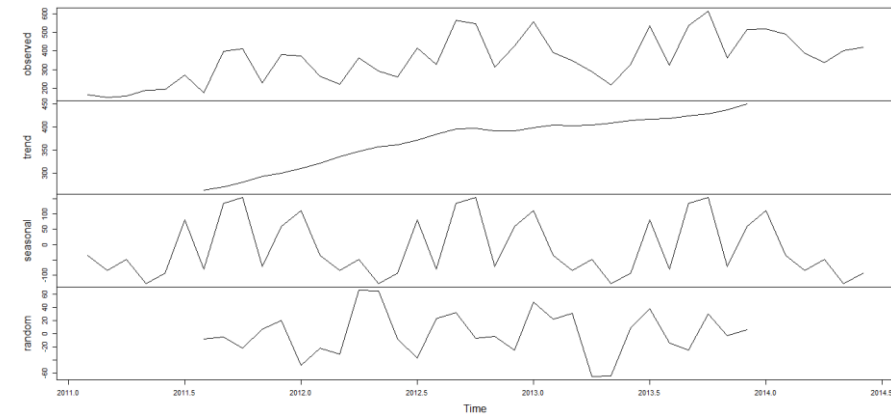
Decomposition of additive time series



Decomposition of additive time series



Decomposition of additive time series

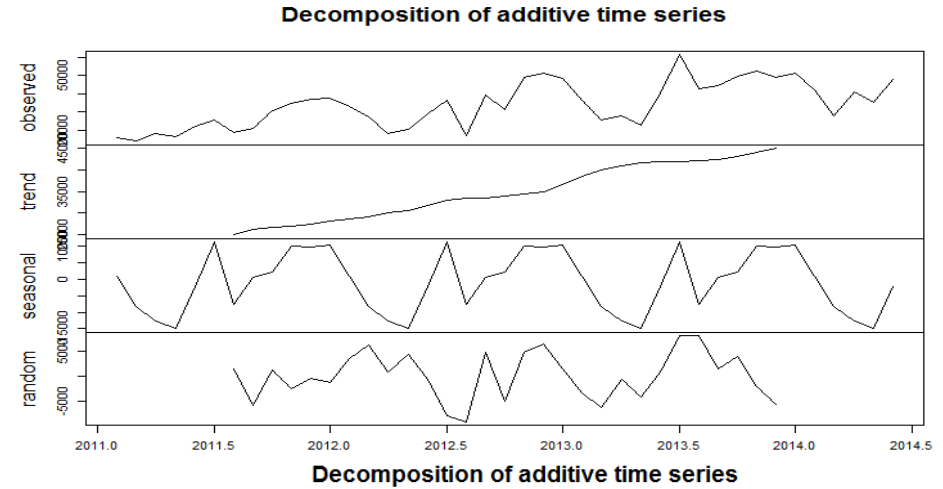
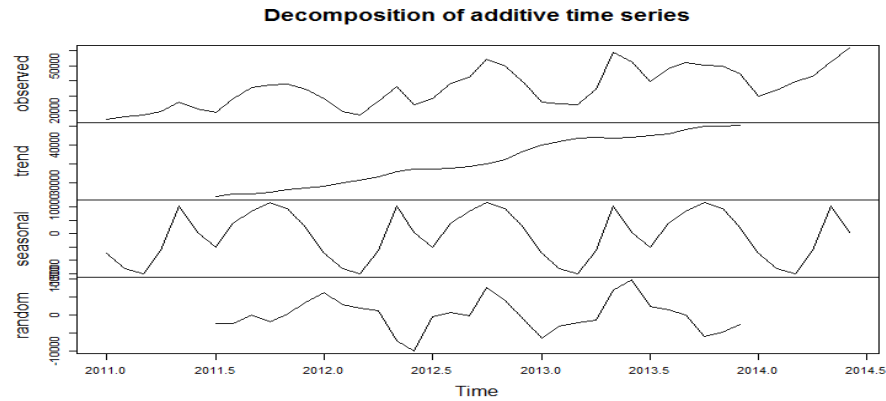


Observation: Clearly the plot shows the value of Trend, Seasonality & Noise across all time series.

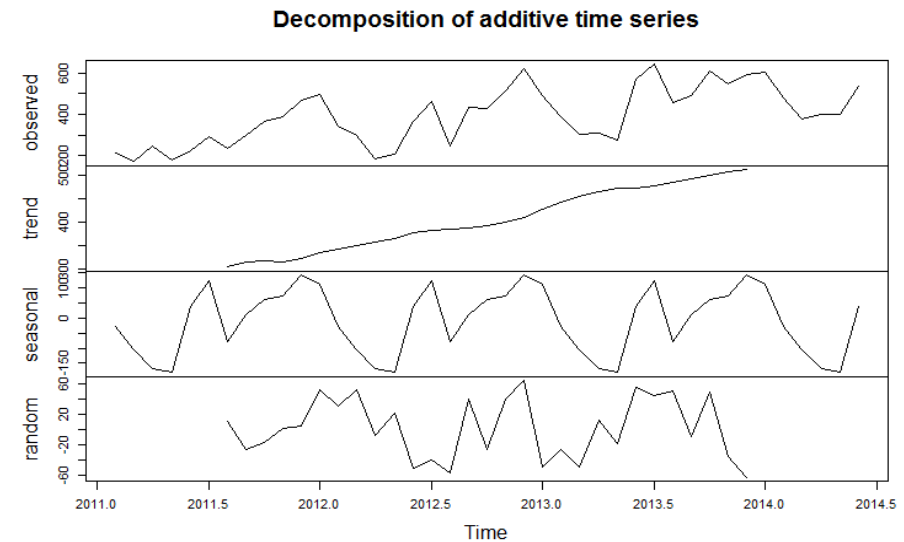
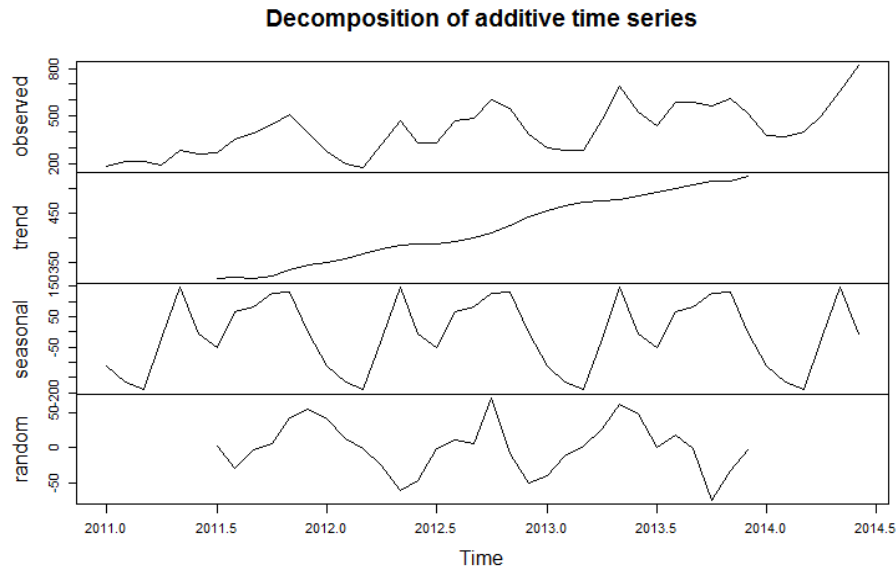
## MOVING AVERAGE

## EXPONENTIAL SMOOTHING

APAC Sales



APAC Demand



Observation: Clearly the plot shows the value of Trend, Seasonality & Noise across all time series.