Retail-Giant Sales Forecasting Case Study

SUBMISSION

Group Members:

1. Sudeep Upadhya
2. Rishabh
3. Vikash Prasad
4. Amit Mankikar

Case Study Objectives

- Find the two most profitable & consistent segment from 21 market & 3 Customer segments
- Forecast the sales and demand (Quantity) for the two most profitable segment for the next 6 months
- Recommend average inventory required for next 6 months

Problem solving methodology

Getting & Understanding the data

- Import the dataset
- Retain only required columns viz. *Market, Segment, Order Date, Sales, Quantity & Profit*

Data Cleaning & Preparation

- Identification & removal of missing values (is.na)
- Data segmentation into 21 subsets of 7 Market with 3 customer segments
- Using Coefficient of Variance (CV), concluded ***APAC.Consumer*** & ***EU.Consumer*** are the 2 most profitable *Market, Segment* combinations

Model Building

- Smoothen the Data (to separate Trend and Seasonality)
  - Identify & use of appropriate alpha parameter for Sales & Quantity
- Forecast 6 months using Classical Decomposition and auto ARIMA
  - Use of sinusoidal function to fit the model in Classical Decomposition method
- Modeling the Stationary component in the Time Series
  - Identifying Local and White noise
- Calculate the Mean Absolute Percentage Error (MAPE)
  - Compare which method is best suited for the 4 predictions

Final Forecasting

- Forecasting with most appropriate model (Classical Decomposition/ Auto ARIMA)

EDA: Understanding the data

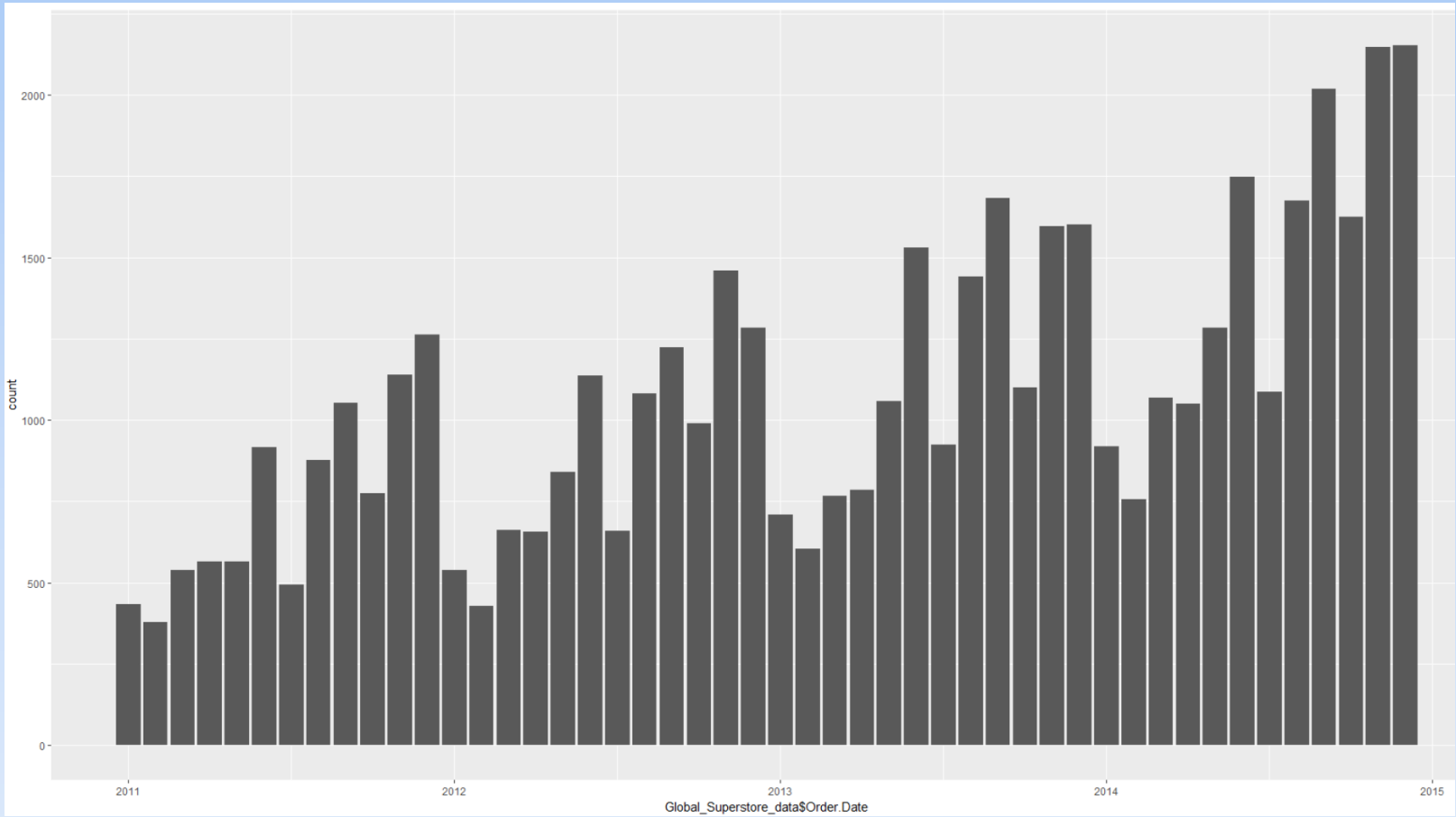
1. None of the retained columns have ‘na’ values

2. Time Series data is available from Jan-2011 to Dec-2014

3. During every holiday season - June, November and December of every year, the no. of sales is high

We can see "Seasonality". Also there is an upward "Trend"

4. Conversion of the Order.Date column from String to yearmon was done



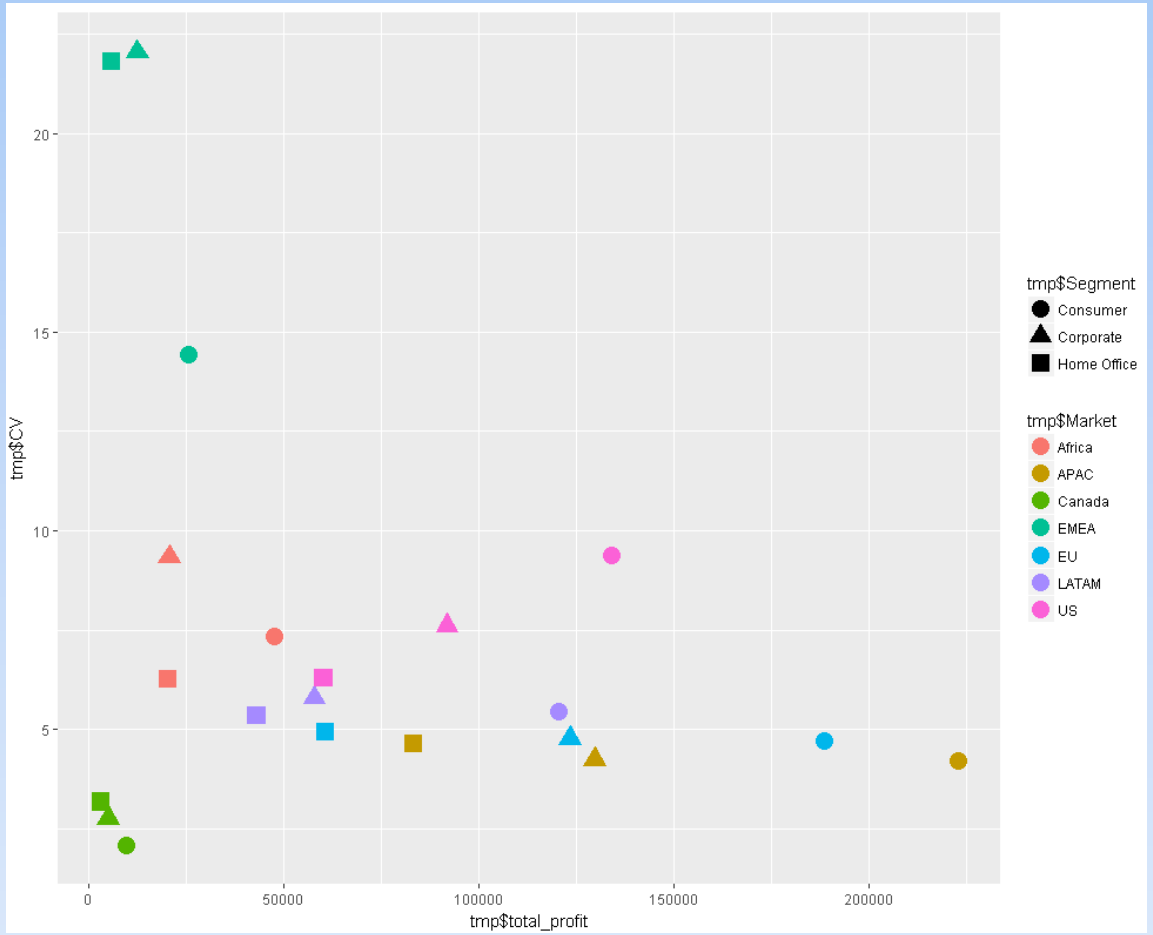
EDA: Understanding the data

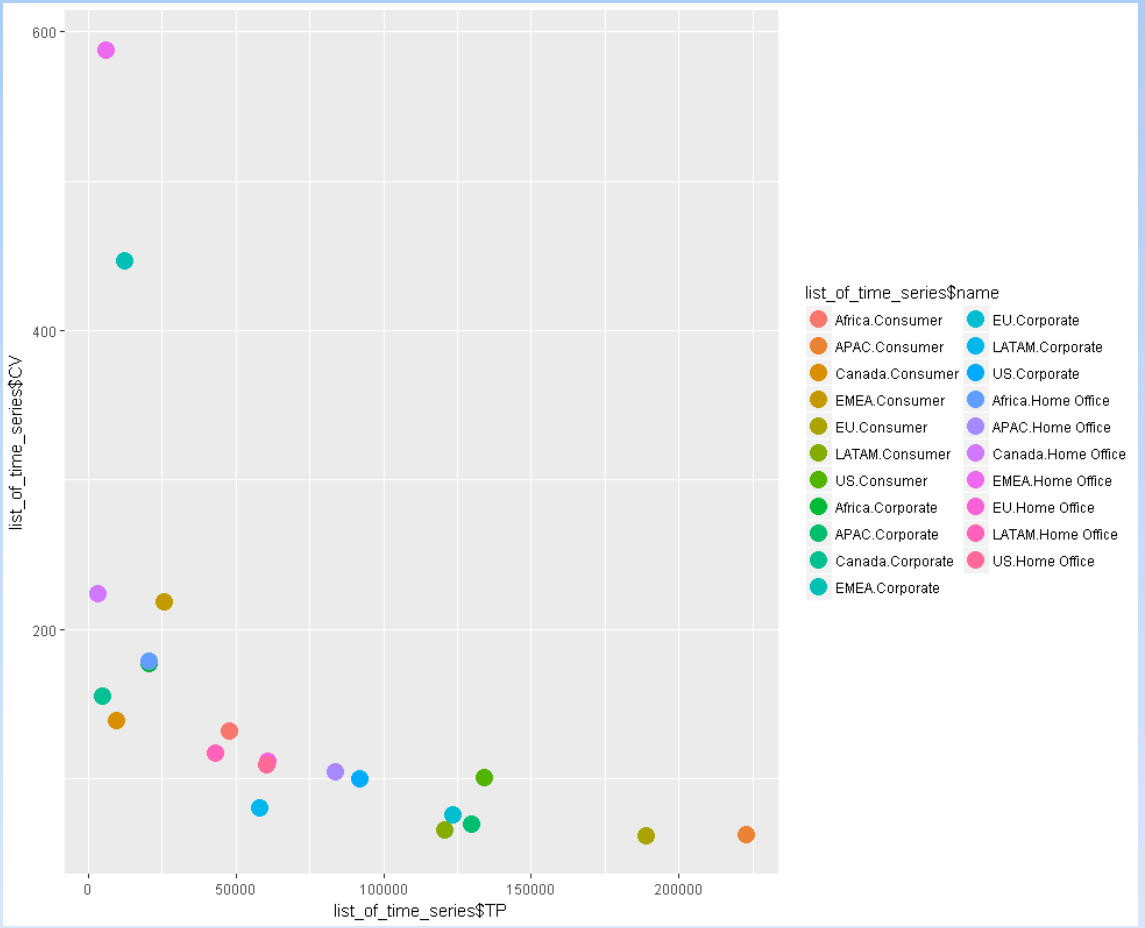
We extracted data of 7 Markets and 3 segments along with the CV and TP

EU Consumer & **APAC Consumer** have highest TP and lowest CV

	name	CV	TP
5	EU.Consumer	62.43052	188687.707
2	APAC.Consumer	63.21323	222817.560
6	LATAM.Consumer	66.14828	120632.932
9	APAC.Corporate	69.80869	129737.235
12	EU.Corporate	76.38072	123393.980
13	LATAM.Corporate	81.11217	57875.421
14	US.Corporate	100.24089	91979.134
7	US.Consumer	101.23900	134119.209
16	APAC.Home Office	104.59784	83445.254
21	US.Home Office	109.61473	60298.679
19	EU.Home Office	111.65073	60748.054
20	LATAM.Home Office	117.56978	43135.134
1	Africa.Consumer	131.95849	47772.099
3	Canada.Consumer	139.53122	9677.700
10	Canada.Corporate	155.27751	5036.460
8	Africa.Corporate	177.61054	20686.965
15	Africa.Home Office	178.99957	20412.567
4	EMEA.Consumer	218.82709	25532.574
17	Canada.Home Office	224.34607	3103.230
11	EMEA.Corporate	446.71024	12499.134
18	EMEA.Home Office	588.07467	5866.263

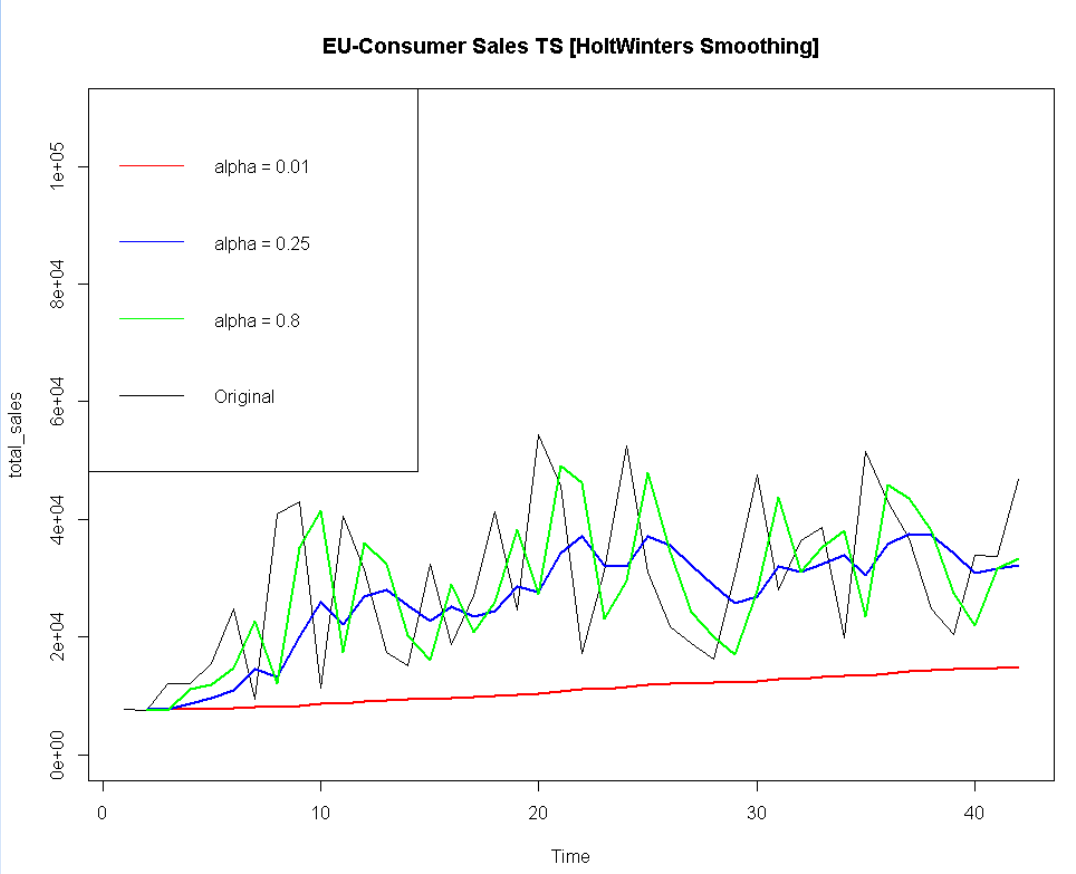
Identification of Best 2 Market/Segment combination

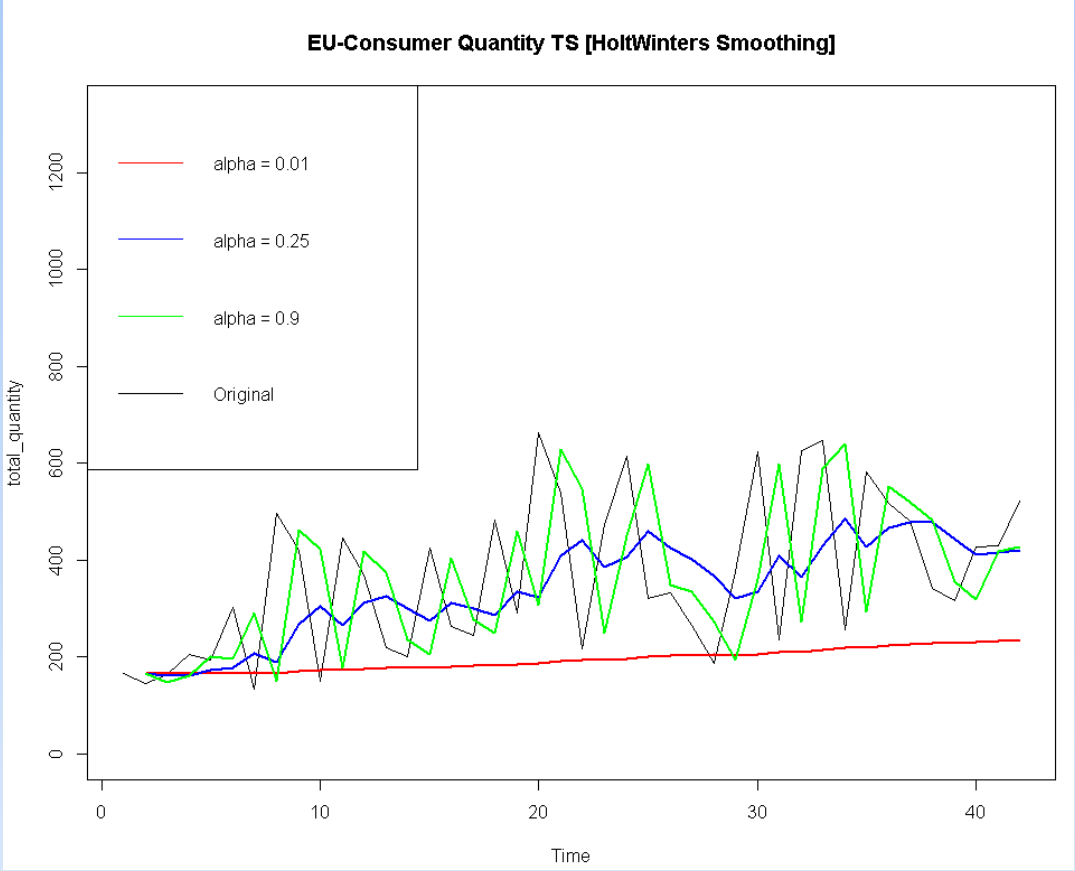




From these plots, it is evident that APAC.Consumer & EU.Consumer are best suited as the 2 most profitable Market, Segment combination.

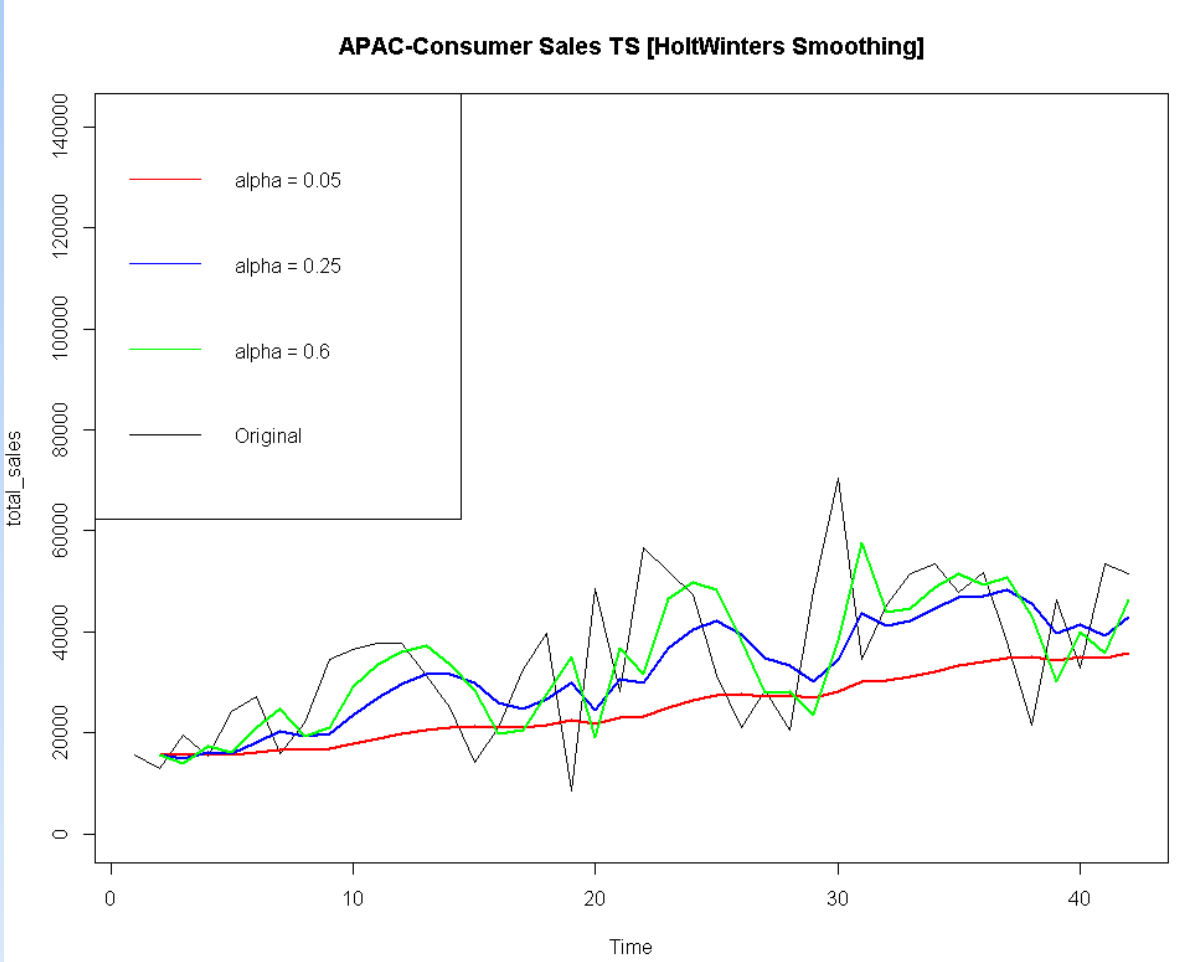
Model Building: Smoothening (EU-Consumer dataset)

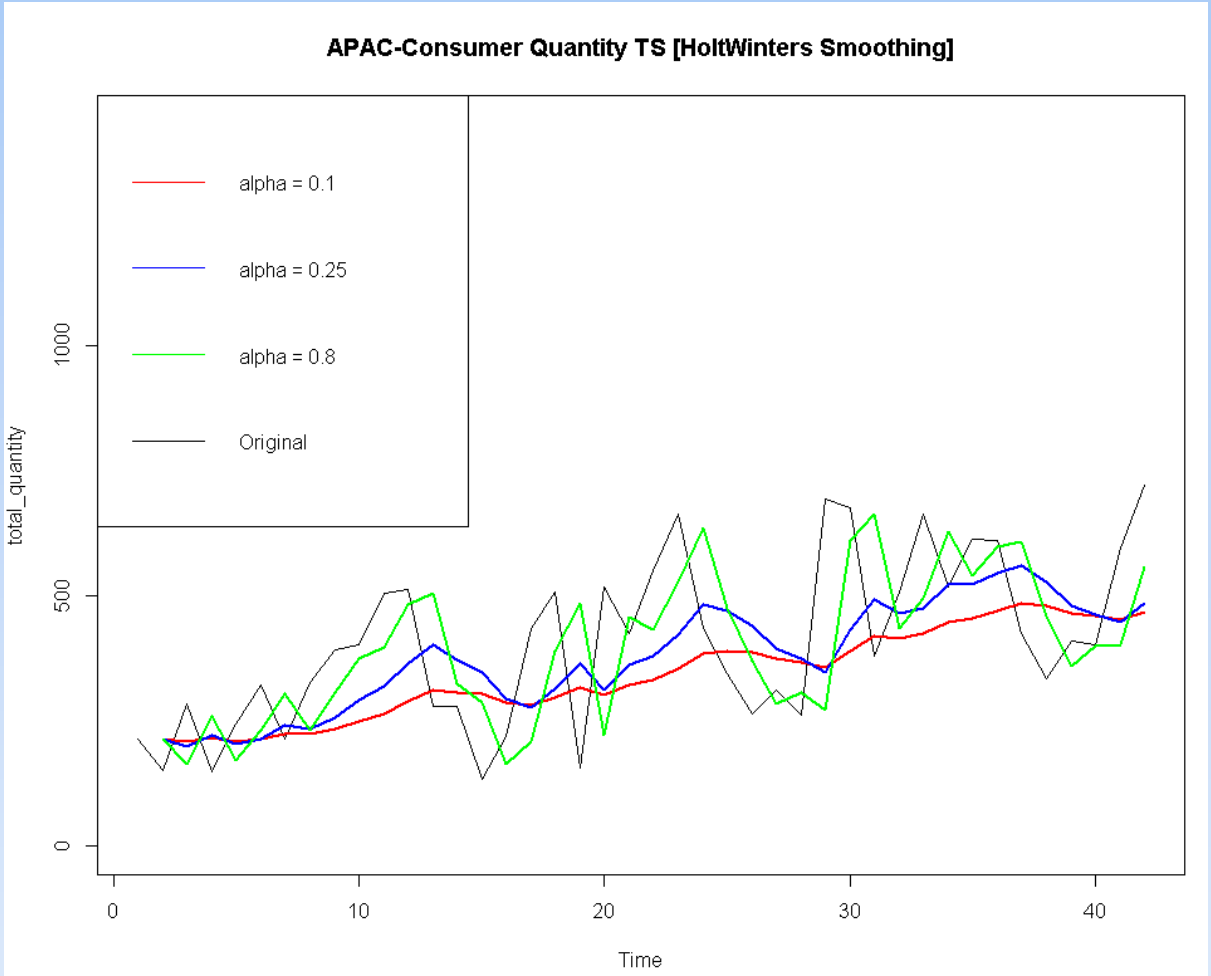




For both Sales and Quantity time series, an alpha value of 0.25 (blue line) seems to give the best & smoothest fit

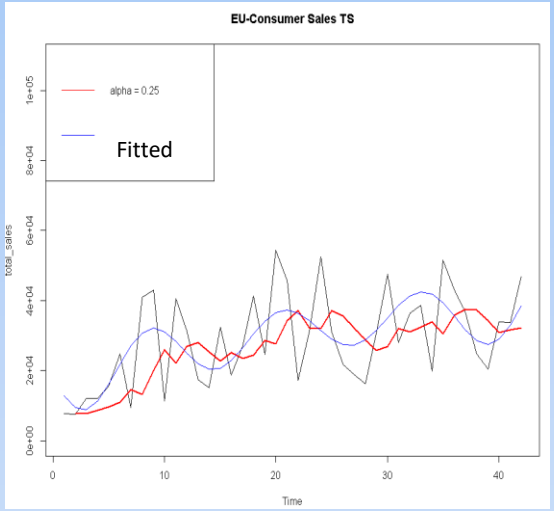
Model Building: Smoothening (APAC-Consumer dataset)



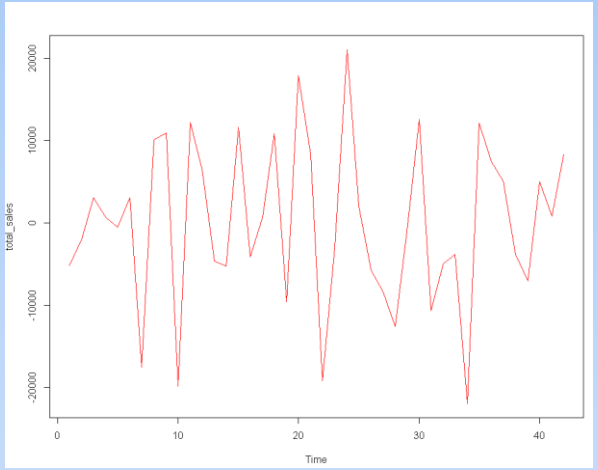


For both Sales and Quantity time series, an alpha value of 0.25 (blue line) seems to give the best & smoothest fit

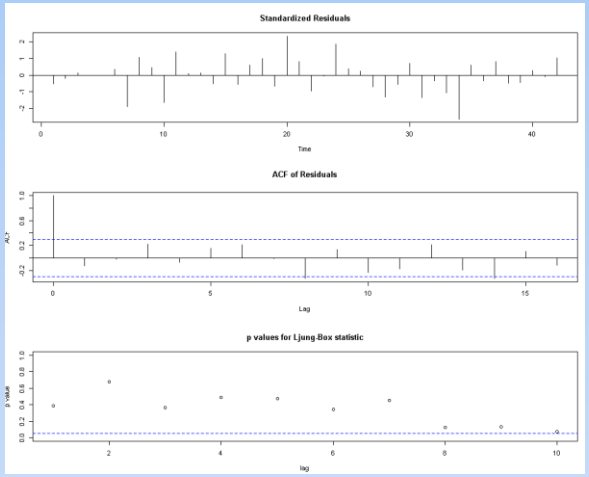
Model Building: Classical Decomposition (EU-Consumer, Sales)



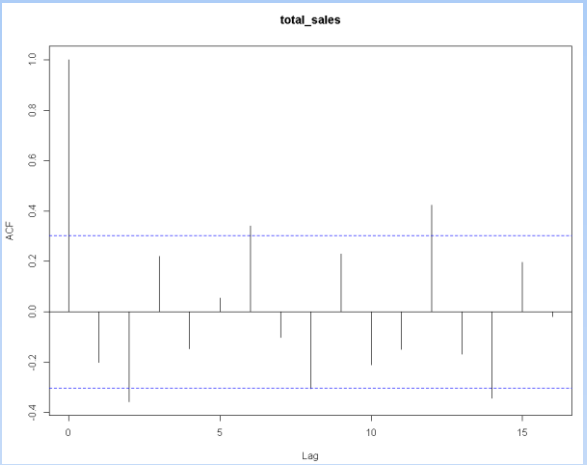
Step 1. Model the Trend and Seasonality



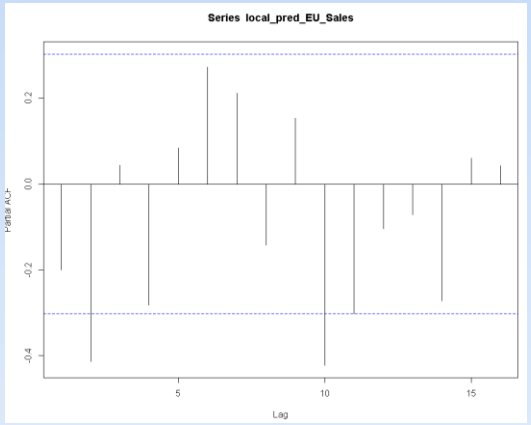
Step 2. Analyse the stationary component



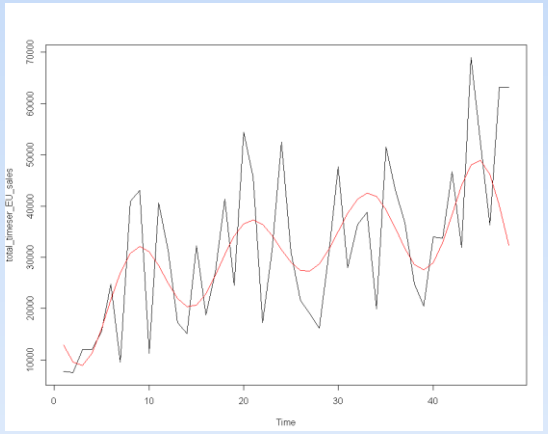
Step 3. Analyse the diagnostic plots for TS



Step 4. Analyse the ACF



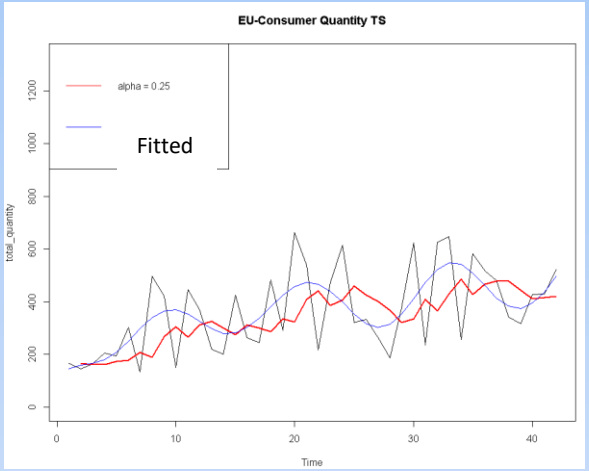
Step 5. Analyse the PACF



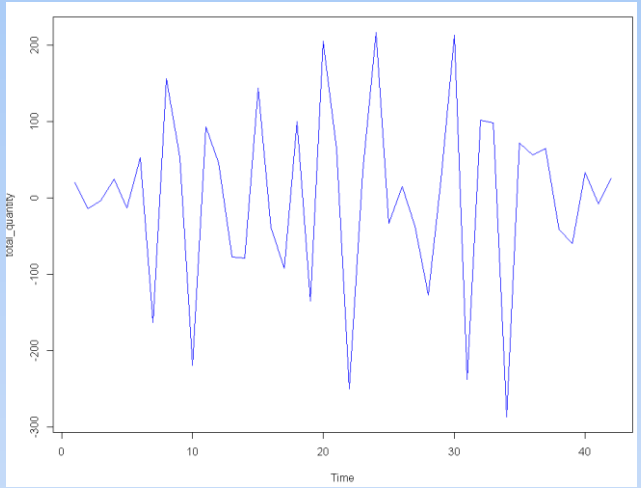
Step 6. Plot the predicted values

ARIMA(0,0,2)
Dickiey Fuller Test P-Value = 0.03 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 31.14

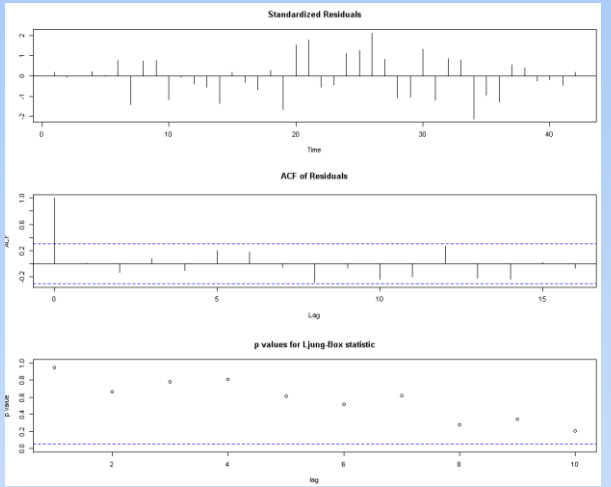
Model Building: Classical Decomposition (EU-Consumer, Quantity)



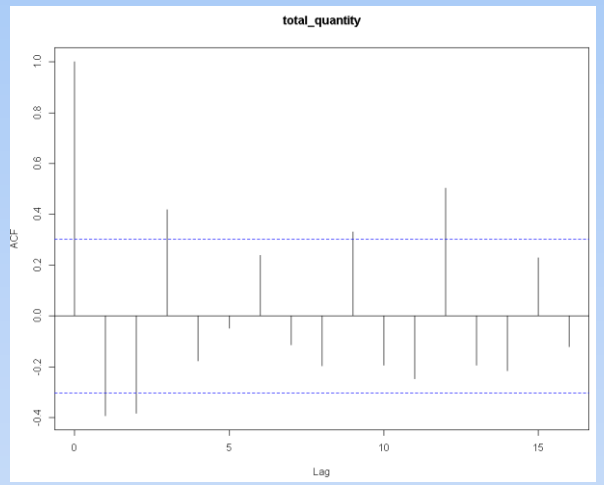
Step 1. Model the Trend and Seasonality



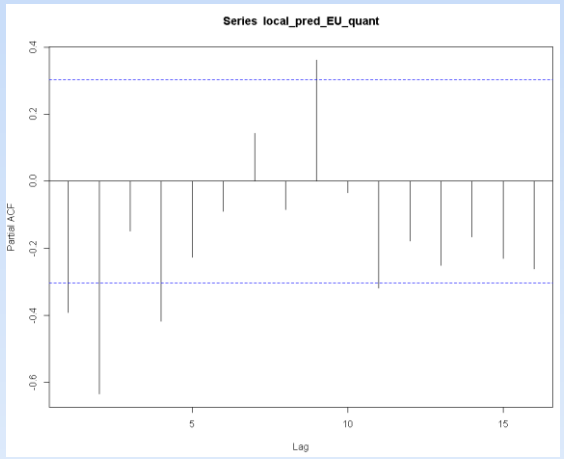
Step 2. Analyse the stationary component



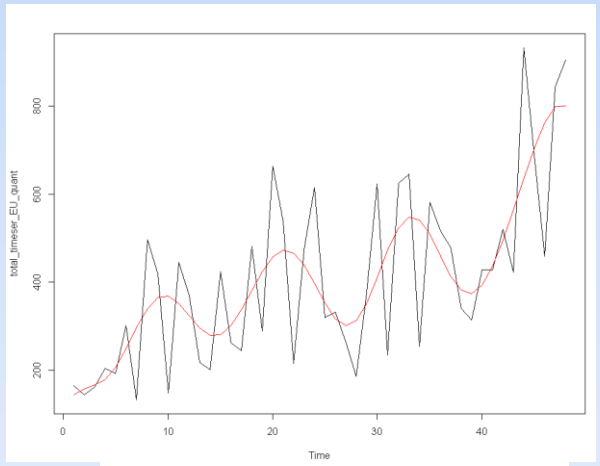
Step 3. Analyse the diagnostic plots for TS



Step 4. Analyse the ACF



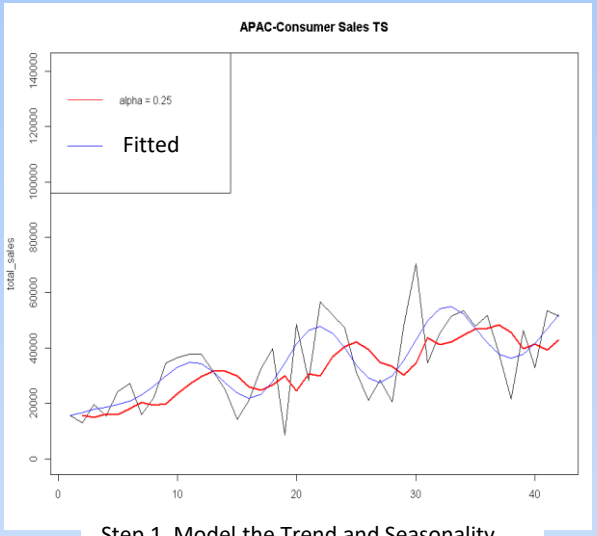
Step 5. Analyse the PACF



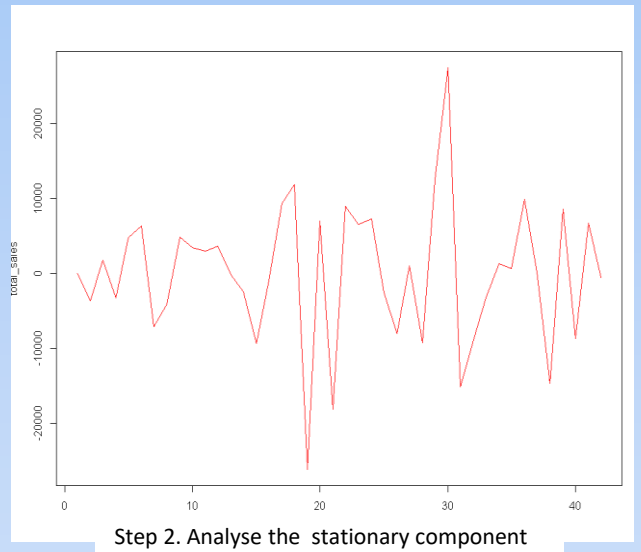
Step 6. Plot the predicted values

ARIMA(0,0,0)
Dickiey Fuller Test P-Value = 0.097 > 0.05
KPSS Test P-Value = 0.1 > 0.05
So DF test says TS is not stationary whereas KPSS indicates it is stationary
MAPE = 25.09

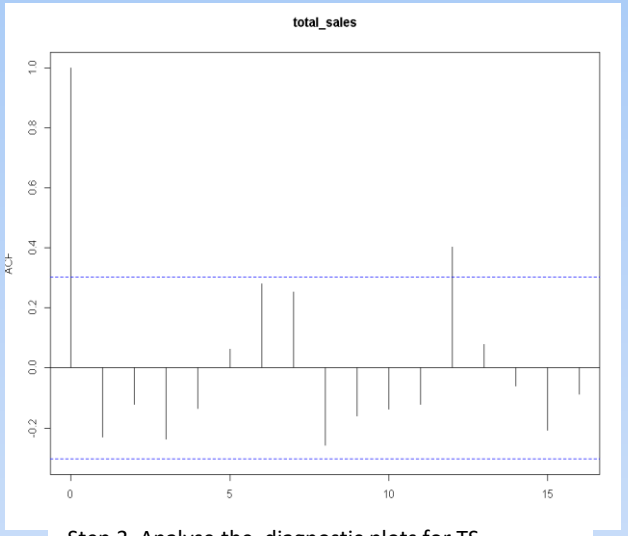
Model Building: Classical Decomposition (APAC-Consumer, Sales)



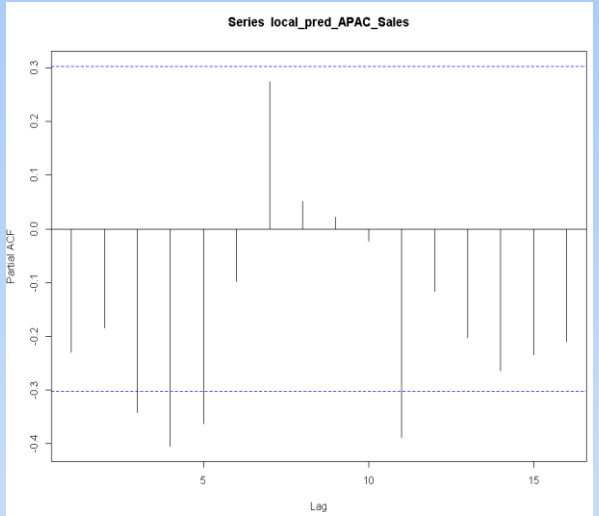
Step 1. Model the Trend and Seasonality



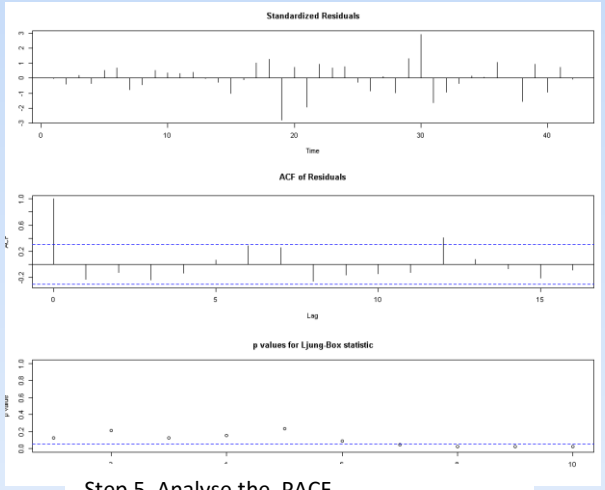
Step 2. Analyse the stationary component



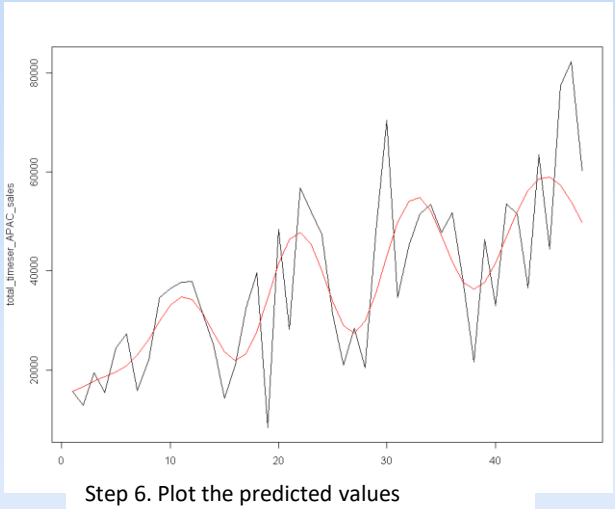
Step 3. Analyse the diagnostic plots for TS



Step 4. Analyse the ACF



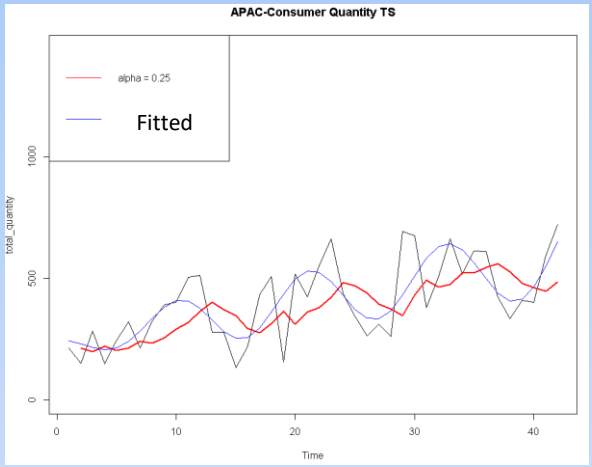
Step 5. Analyse the PACF



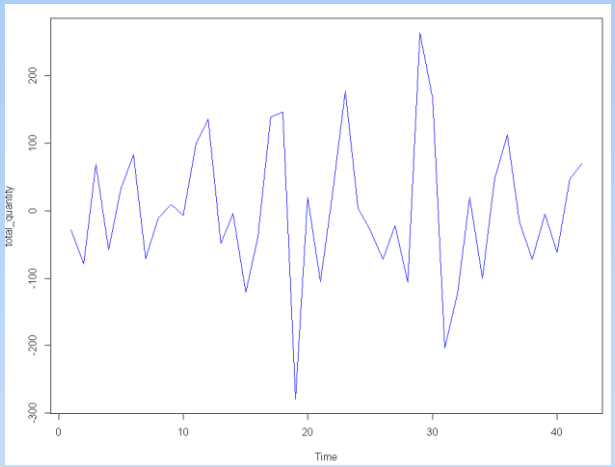
Step 6. Plot the predicted values

ARIMA(0,0,0)
Dickiey Fuller Test P-Value = 0.01 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 28.66

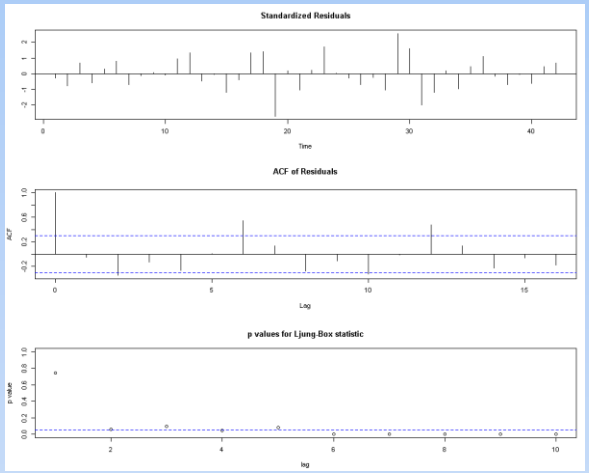
Model Building: Classical Decomposition (APAC-Consumer, Quantity)



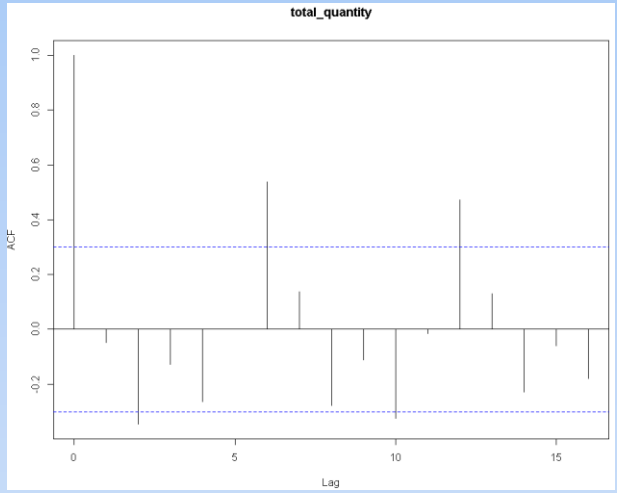
Step 1. Model the Trend and Seasonality



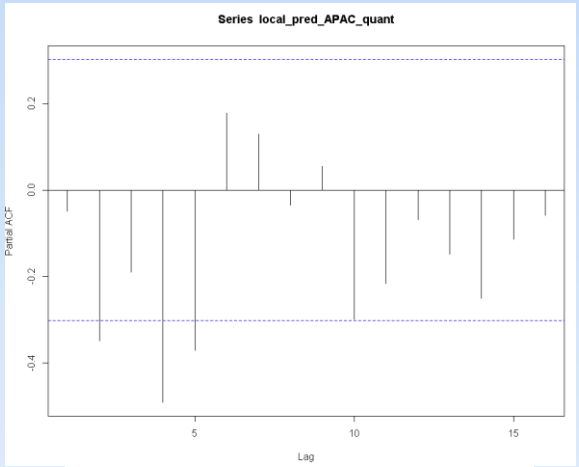
Step 2. Analyse the stationary component



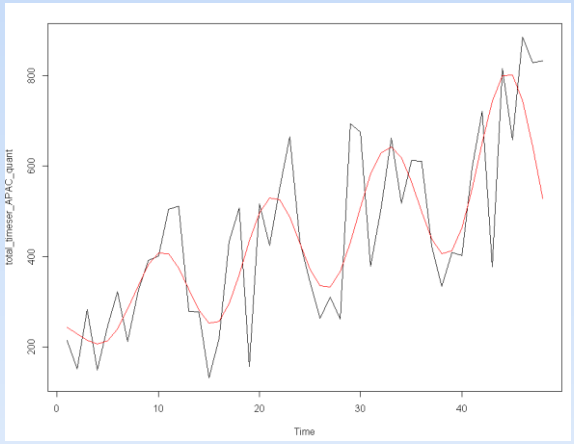
Step 3. Analyse the diagnostic plots for TS



Step 4. Analyse the ACF



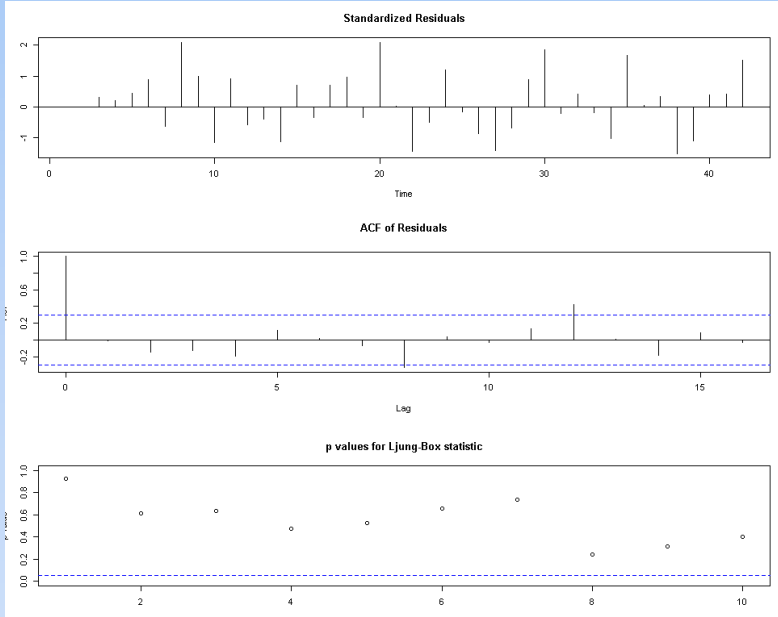
Step 5. Analyse the PACF



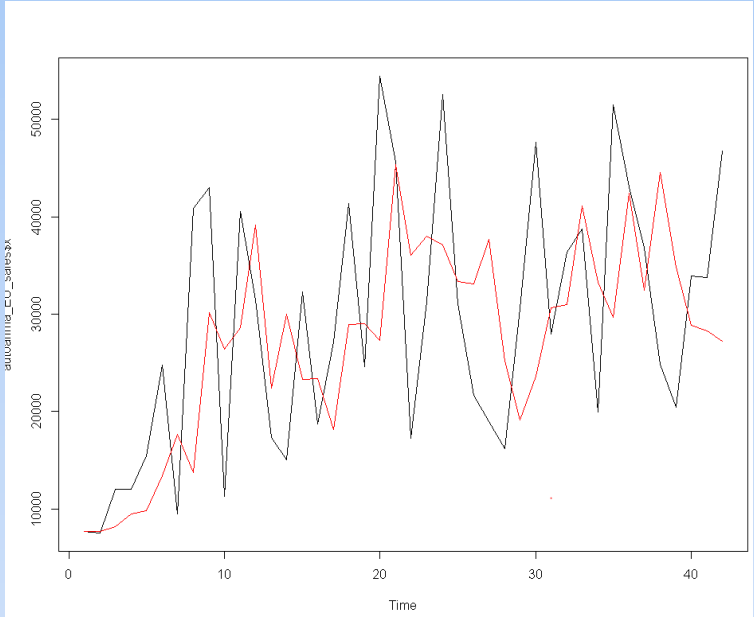
Step 6. Plot the predicted values

ARIMA (0,0,0)
Dickiey Fuller Test P-Value = 0.01 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 32.64

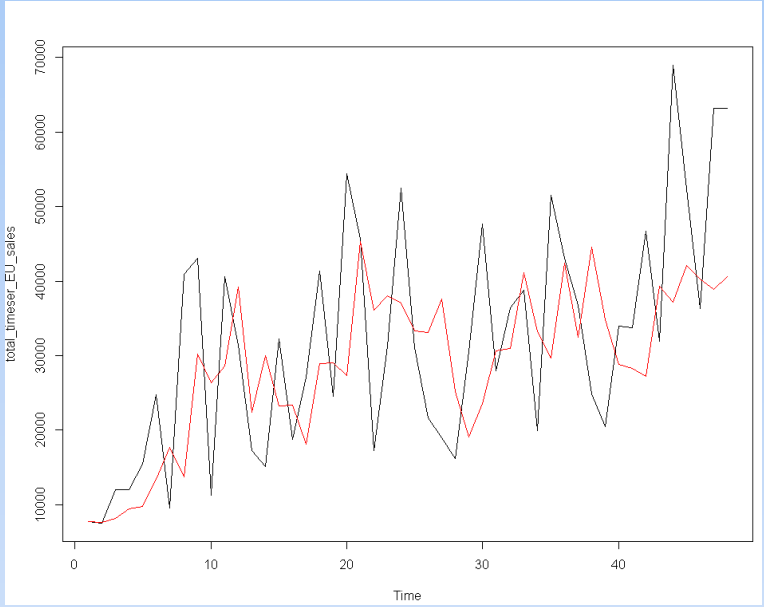
Model Building: Auto-ARIMA (EU-Consumer, Sales)



Step 1. Analyse the diagnostic plots for TS



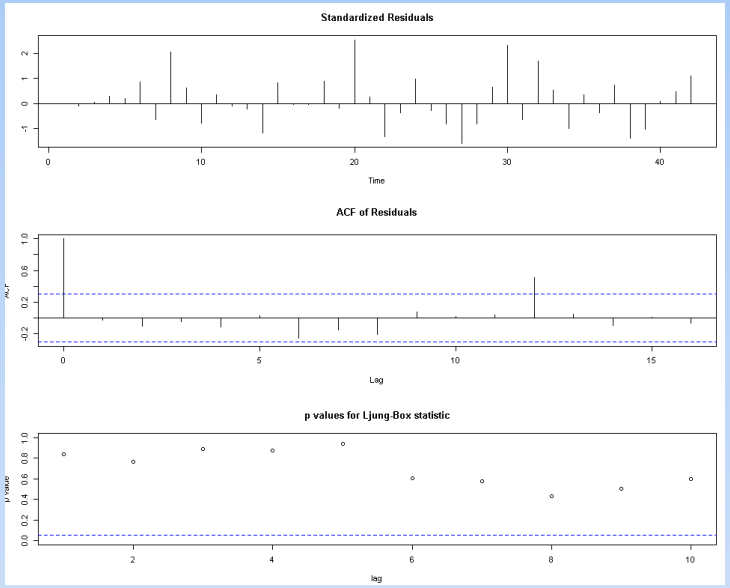
Step 2. Analyse the auto ARIMA fit



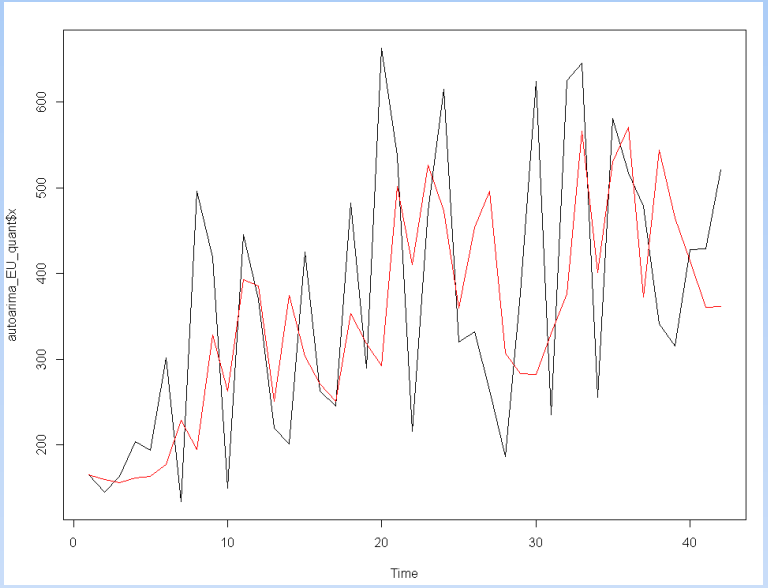
Step 3. Plot complete time series and analyze

ARIMA(2,1,0)
Dickiey Fuller Test P-Value = 0.01 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 28.92

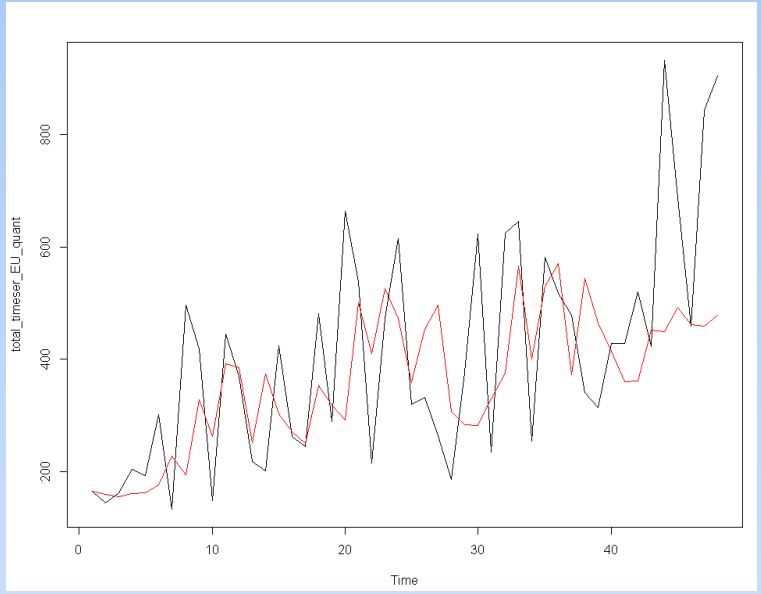
Model Building: Auto-ARIMA (EU-Consumer, Quantity)



Step 1. Analyse the diagnostic plots for TS



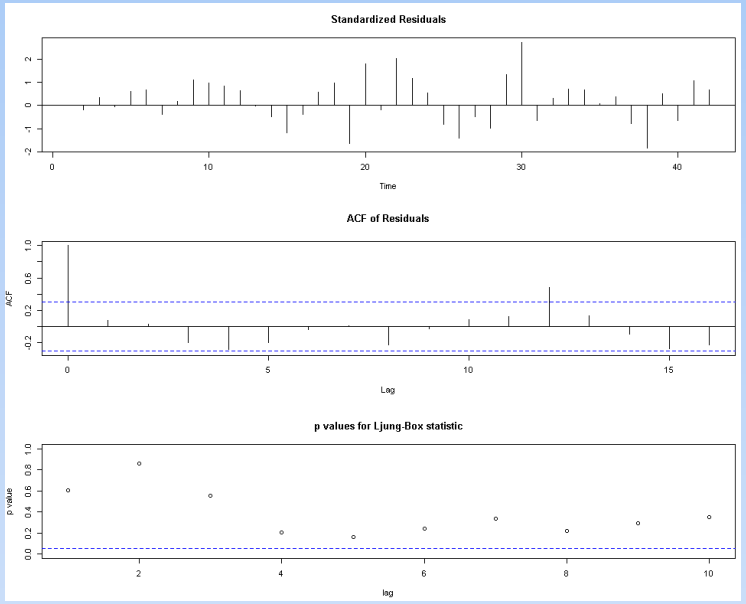
Step 2. Analyse the auto ARIMA fit



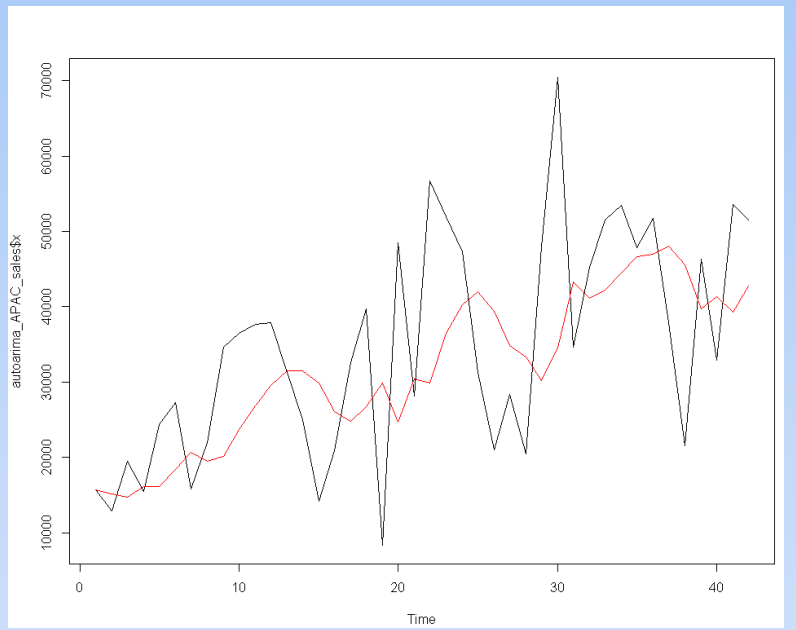
Step 3. Plot complete time series and analyse

ARIMA(2,1,0)
Dickiey Fuller Test P-Value = 0.04 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 30.13

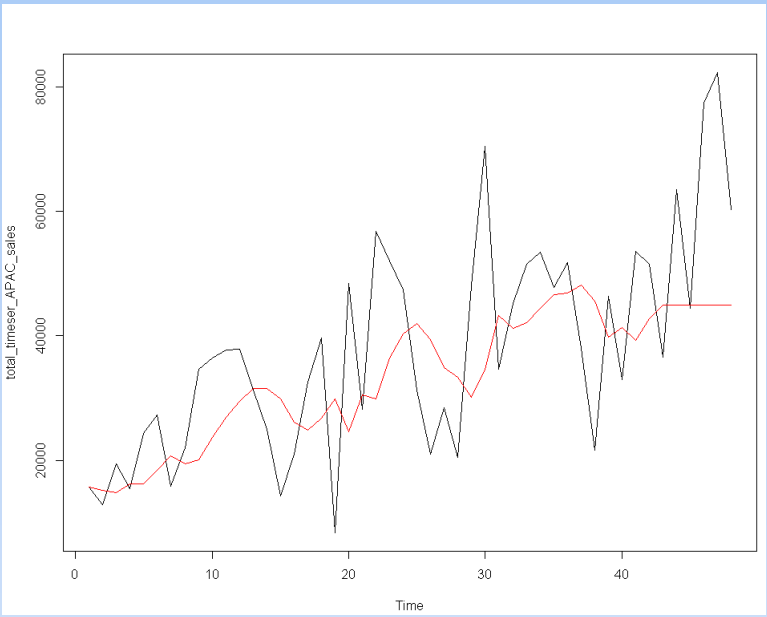
Model Building: Auto-ARIMA (APAC-Consumer, Sales)



Step 1. Analyse the diagnostic plots for TS



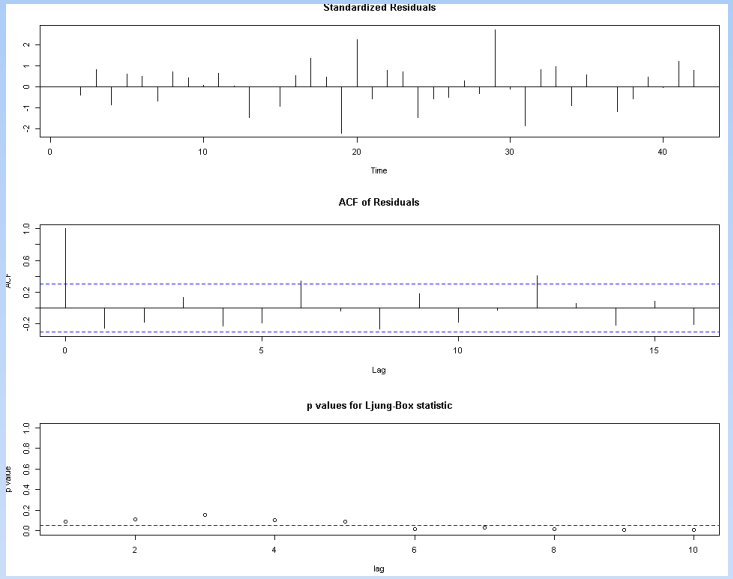
Step 2. Analyse the auto ARIMA fit



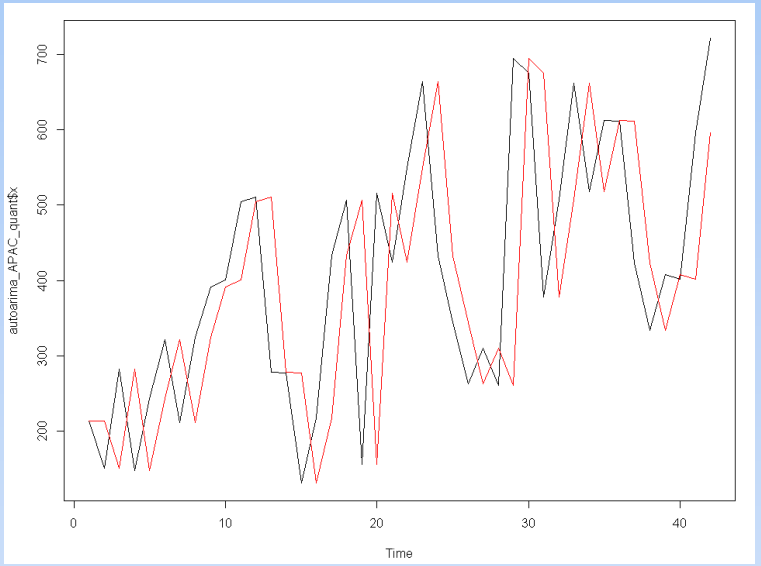
Step 3. Plot complete time series and analyse

ARIMA(0,1,1)
Dickiey Fuller Test P-Value = 0.01 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 27.67

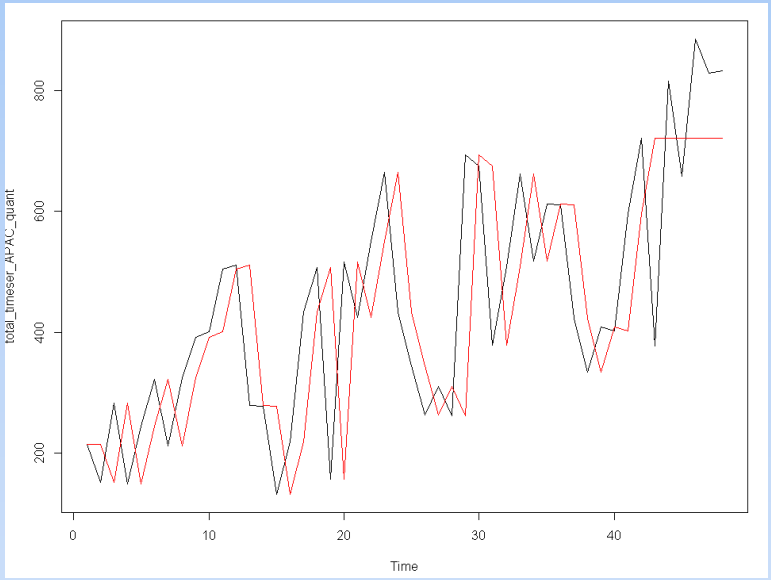
Model Building: Auto-ARIMA (APAC-Consumer, Quantity)



Step 1. Analyse the diagnostic plots for TS



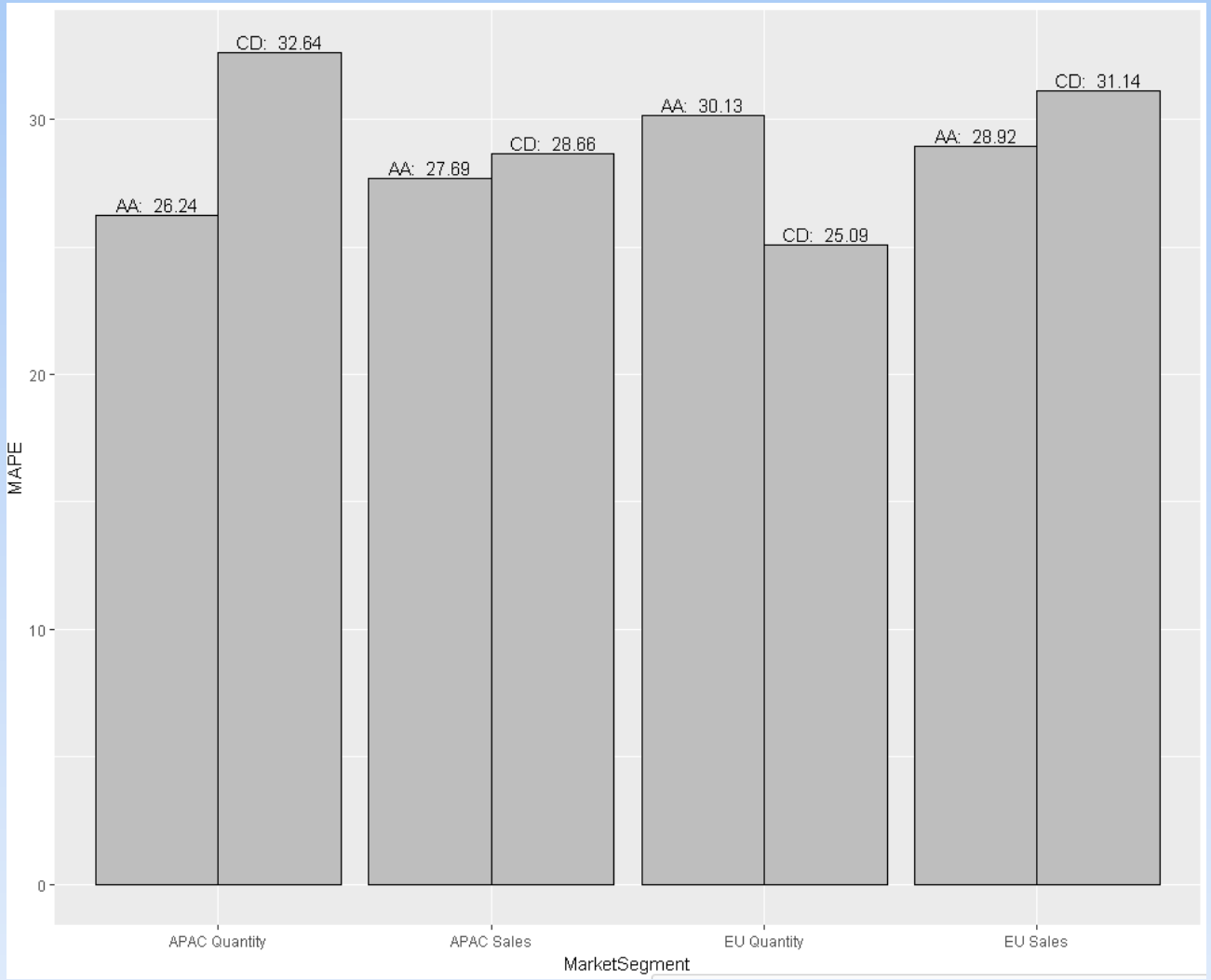
Step 2. Analyse the auto ARIMA fit



Step 3. Plot complete time series and analyse

ARIMA(0,1,0)
Dickiey Fuller Test P-Value = 0.01 < 0.05
KPSS Test P-Value = 0.1 > 0.05
So, we conclude that time series is stationary
MAPE = 26.24

Model Comparison

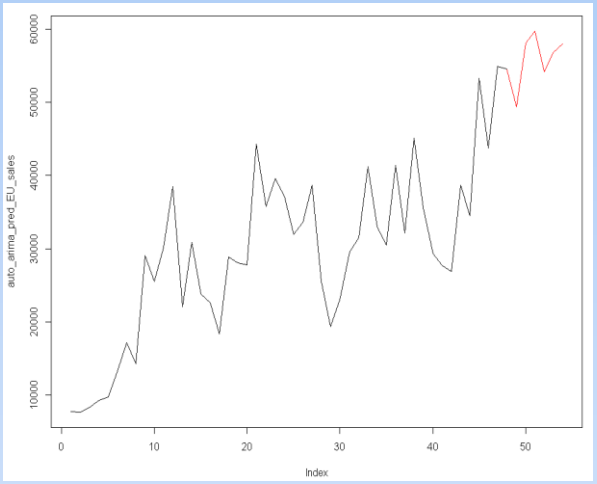


As we can see Auto ARIMA models have lower MAPE than Classical Decomposition in all time series analysis except in EU- Consumer – Quantity

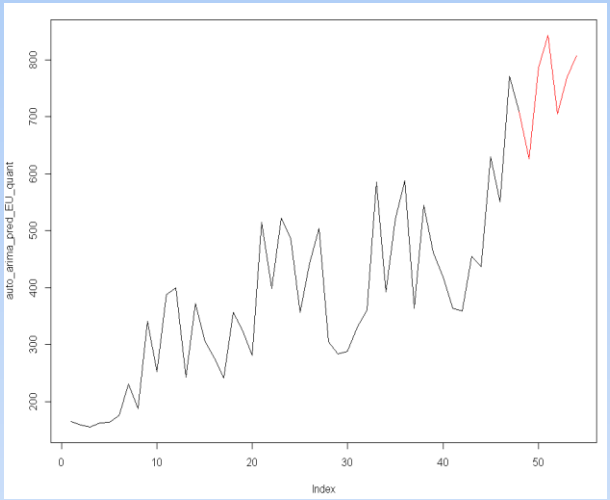
We will use Auto ARIMA models in all prediction except in EU-Consumer – Quantity where we will use the model built using Classical Decomposition

Forecast:

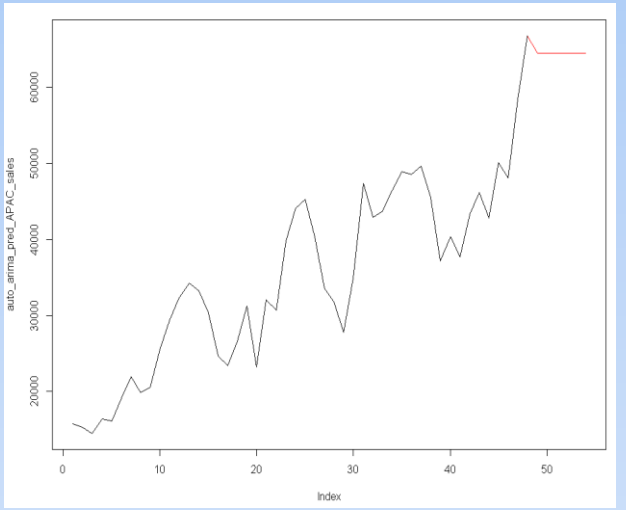
2015 Prediction of EU-Consumer Sales



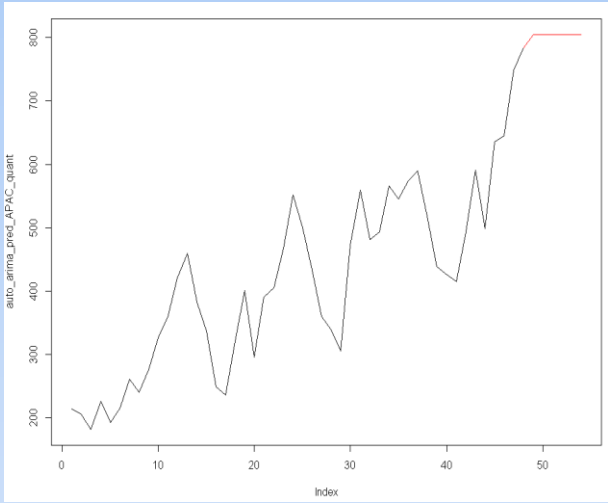
2015 Prediction of EU-Consumer Quantity



2015 Prediction of APAC-Consumer Sales



2015 Prediction of APAC-Consumer Quantity



EU Next 6 months:

Total Sales predicted in EU-Consumer = 336150
Average Sales predicted in EU-Consumer = 56025

Total Quantity predicted in EU-Consumer = 4536
Average Sales predicted in EU-Consumer = 756

APAC Next 6 months:

Total Sales predicted in EU-Consumer = 386969
Average Sales predicted in EU-Consumer = 64494

Total Quantity predicted in EU-Consumer = 4826
Average Sales predicted in EU-Consumer = 804

Recommendation

- Management is recommended to invest more in EU-Consumer and APAC-Consumer market segments
- Following the expected trend, Sales are expected to grow
- But, in the next 6 months, the seasonality will affect Sales
- Sales will be lower in Jan & Feb ‘15 than in Nov & Dec ‘14
- There will be average Sales until May 15 & Sales will jump again by June
- Average Quantity per month in EU will be 804 and in APAC will be 756. So sufficient inventory needs to be accommodated.