

Retail-Giant Sales Forecasting Case Study Presentation

SUBMISSION

1. Shivam Kakkar (Facilitator)
2. Ashwin Suresh
3. Manohar Shanmugasundaram
4. P Sai Prathyusha

Retail-Giant Sales Forecasting - Case Study Analysis

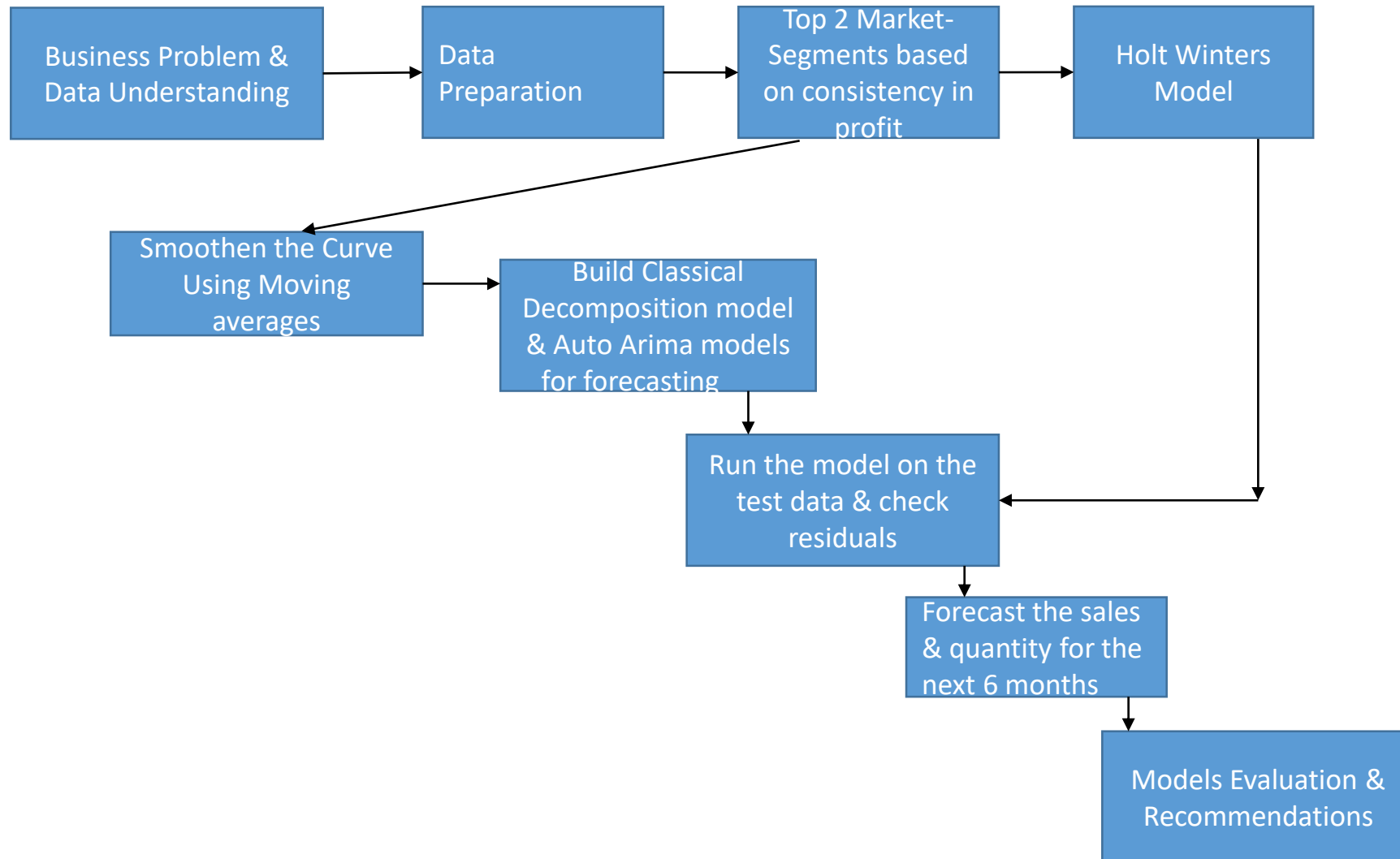
Background

“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.

Business Objectives

- To Forecast sales and demand for next six months using time series model .
- The model would be further helpful to manage the revenue and inventory according to the sales forecasted.
- Sales transaction data is available for four(4) years
- Each data point is a transaction
- Attributes of Interest:
 - Segment
 - Market
 - Sales
 - Quantity
 - Profits made on the sales
- We have multiple markets and segments so we would obtain top 2 using Coefficient of Variance

Problem Solving Approach Flow Chart



Step 1 - Data Preparation & Derived Metrics

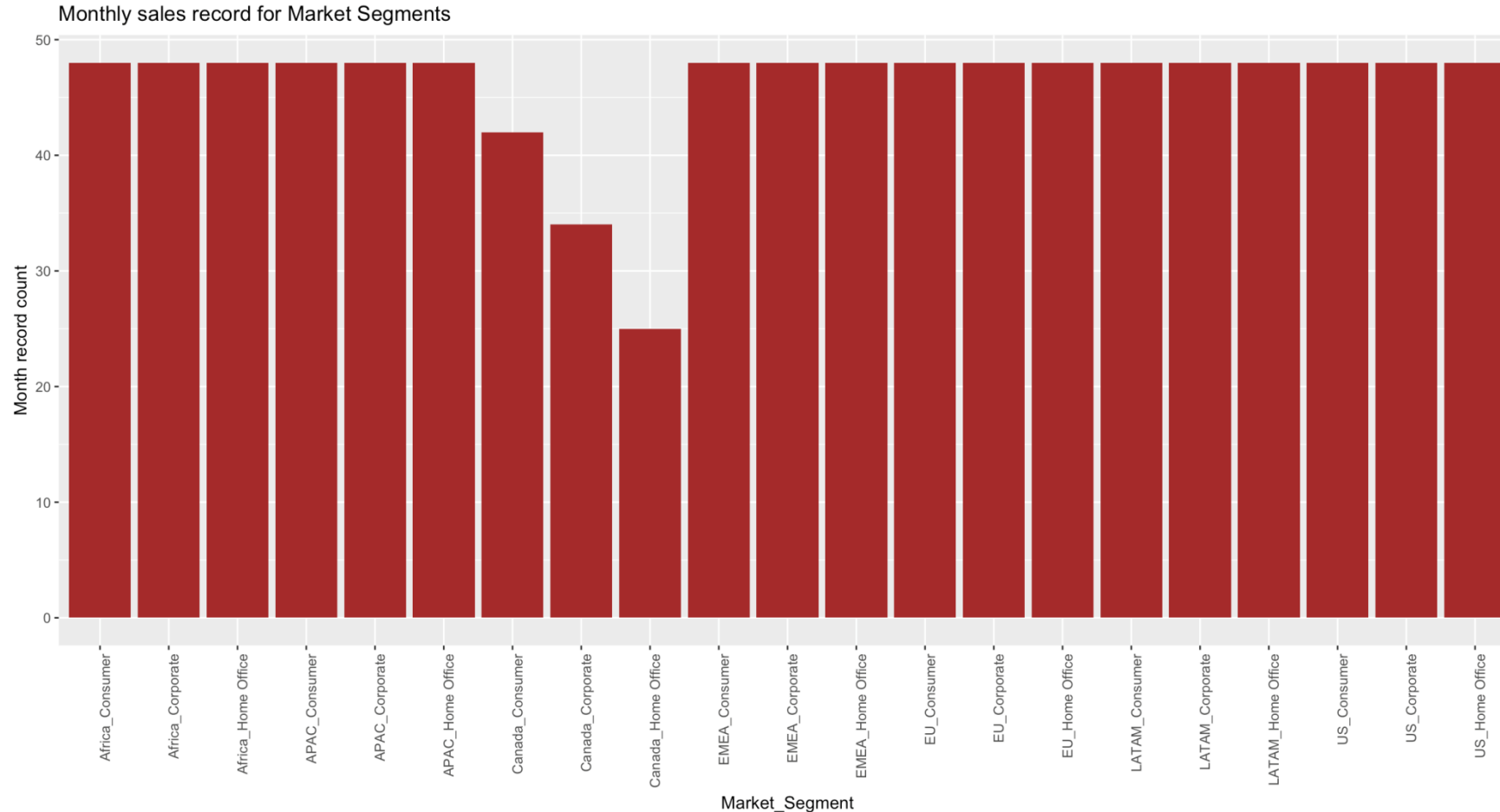
Preparation

- In the given Order Date and Sales are the main attributes required for the time series model.
- Year Month is retrieved from Order Date to achieve Chronological order.
- Data is converted to time series in-order to build the model

Metrics

- Data is divided into 21 buckets respective to every "Market and Segment combination"
- Aggregated "Profit" , "Sales" & "Quantity" at monthly level
- Coefficient of variation is calculated for every bucket

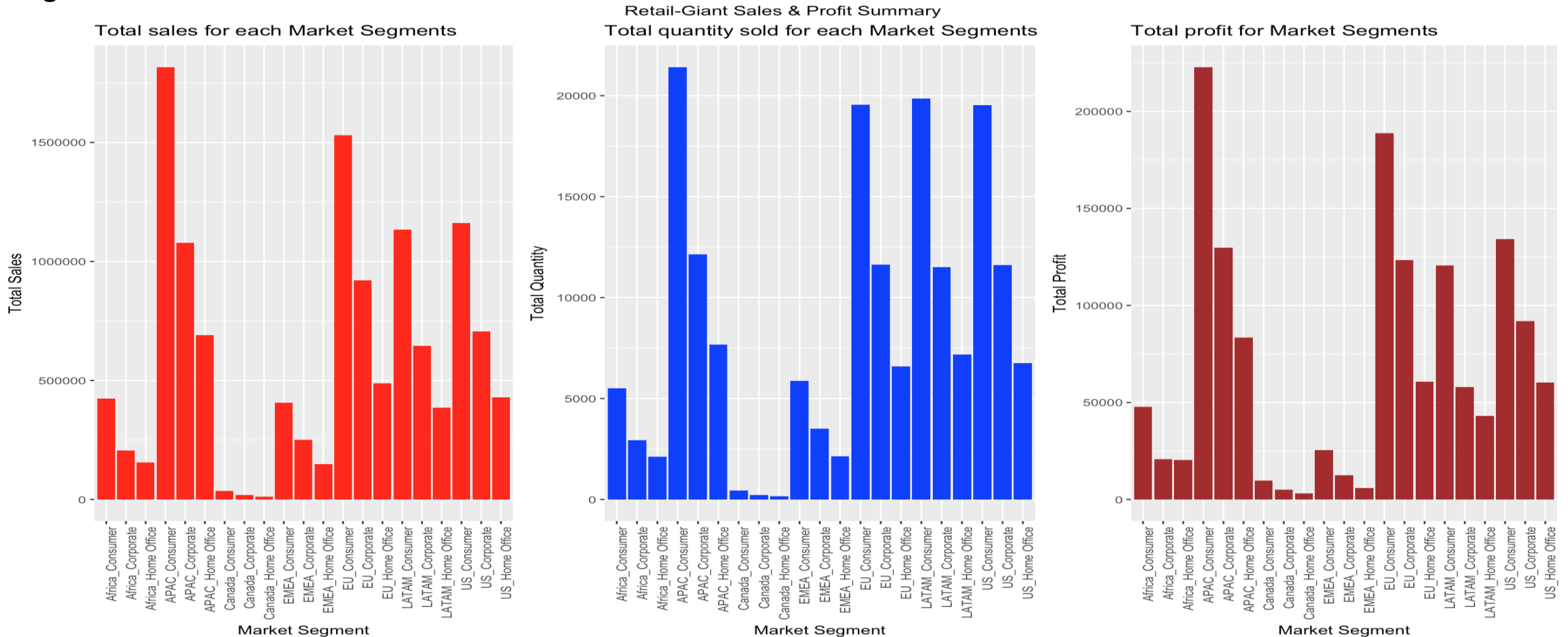
Univariate analysis on the aggregated data to check whether we have all market segments with 48 months of data.



INFERENCE:

Clearly we could see market segments “Canada Consumer”, “Canada Corporate” & “Canada Home Office” does not have records for all the months. We assumed it could be due to no sales or missing records.

Bivariate analysis on the aggregated data to insight on total sales, total quantity and total profit for each market segments.

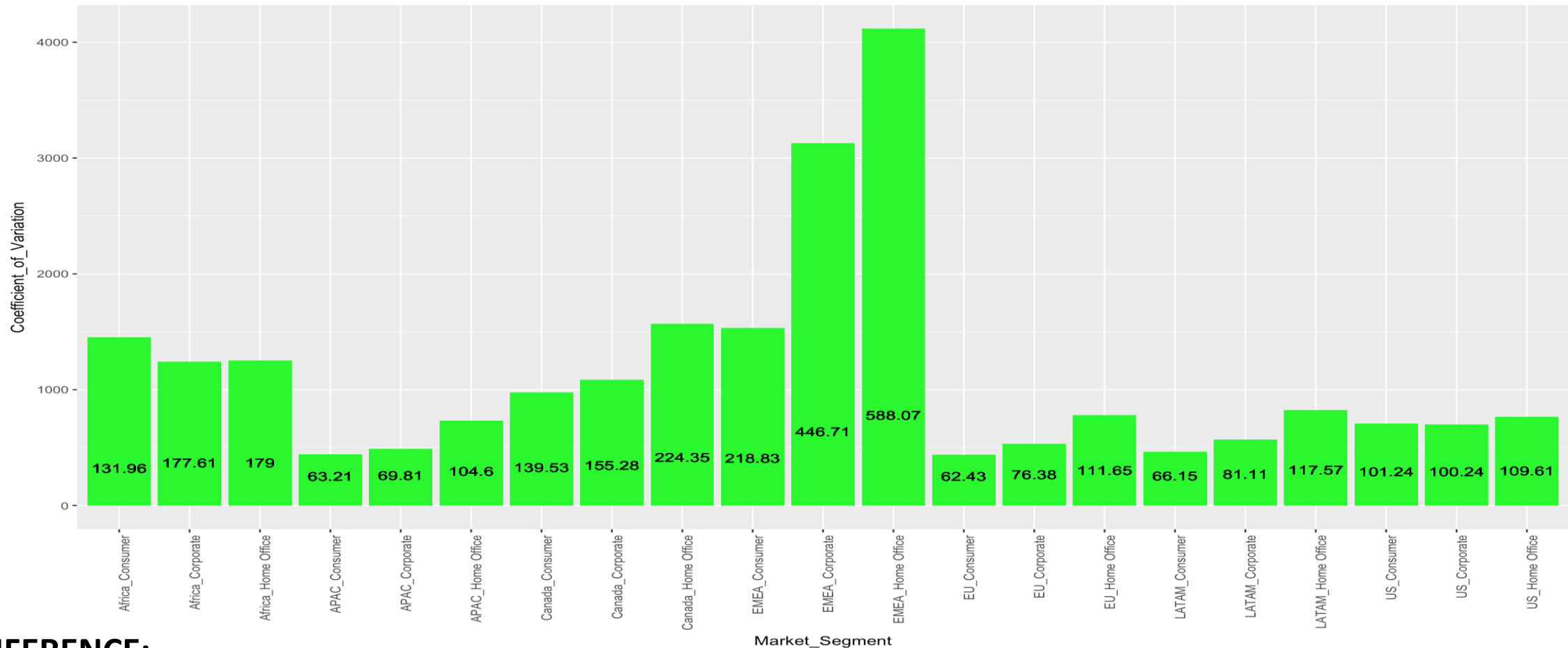


INFERENCE:

From the above plots, we could see that market segments **“APAC Consumer”** and **“EU Consumer”** are performing better than all other segments.

Coefficient of Variation

Coefficient of variation is used find the 2 most profitable and consistently profitable segments across all the market segments.



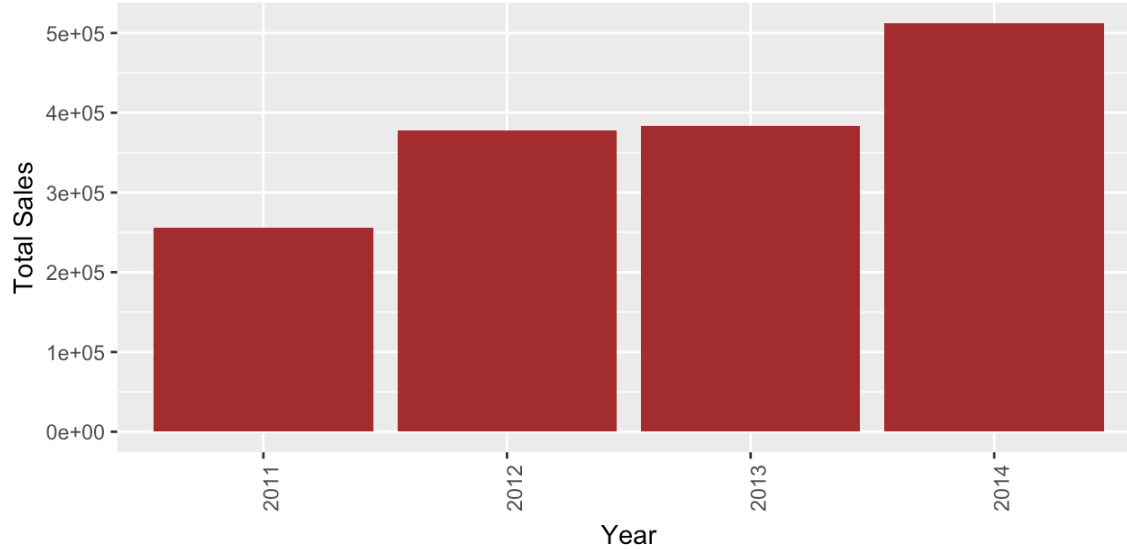
INFERENCE:

From the plot above we could see the market segments “**EU Consumer**” & “**APAC Consumer**” are the consistent profitable segments (based on the least COV values).

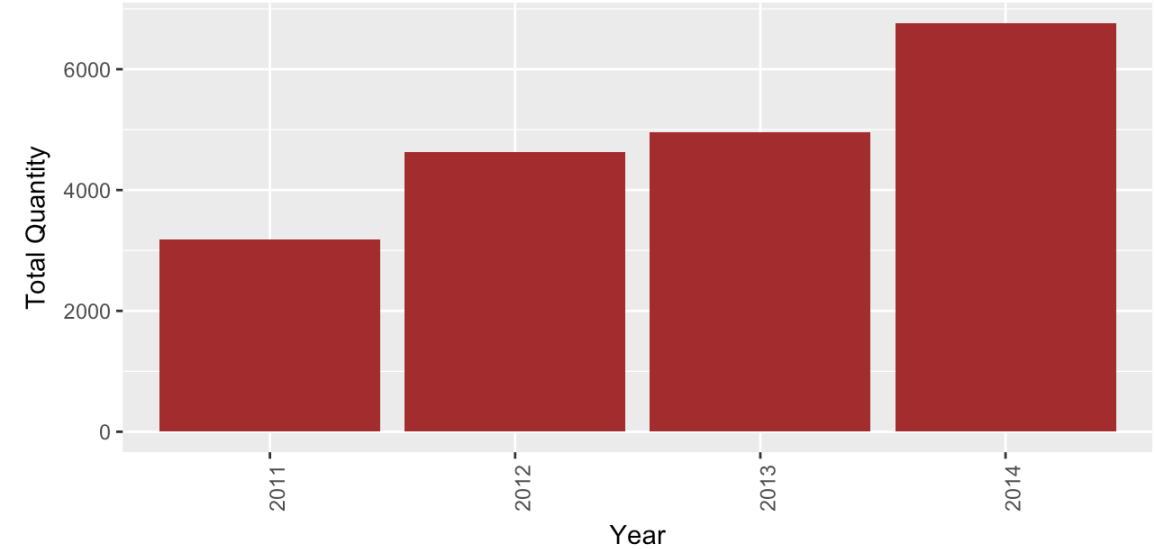
Profitable Market Segment Analysis

Sale and Quantity summary for EU and APAC Consumer segments

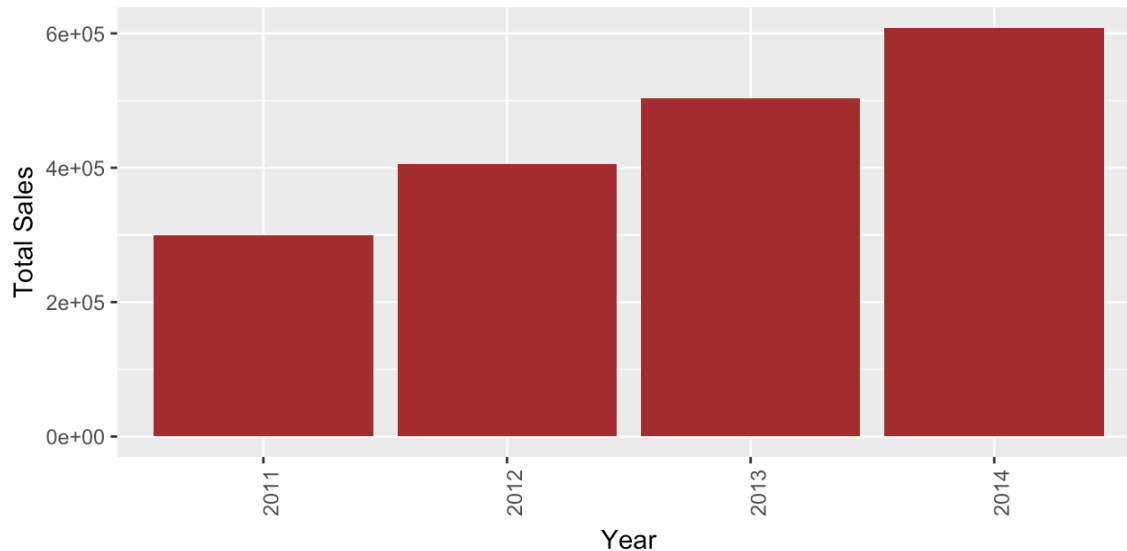
Total sales for EU Consumer per year



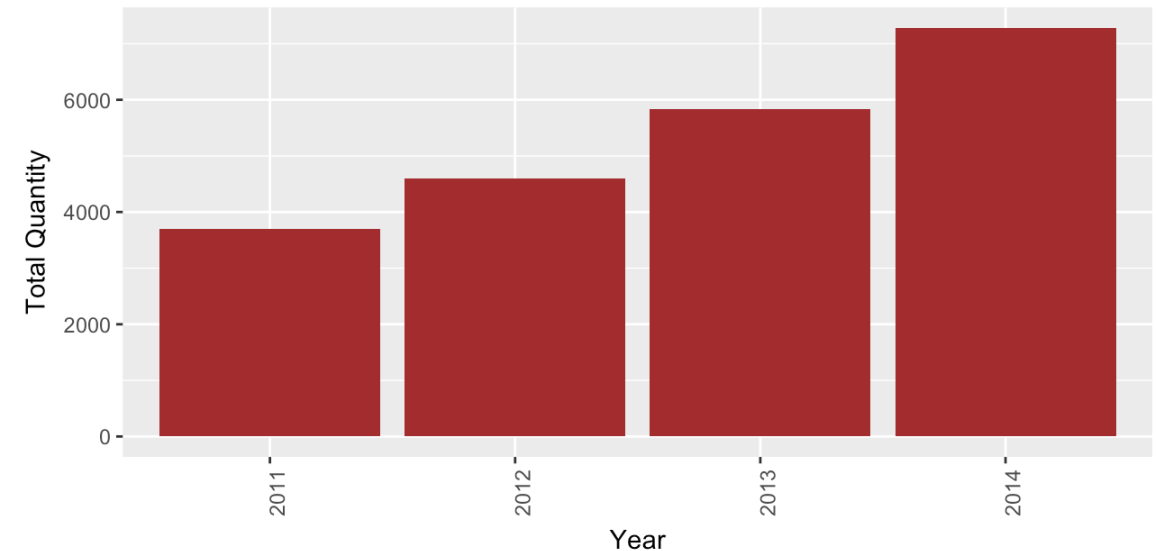
Total quantity sold for EU Consumer per year



Total sales for APAC Consumer per year



Total quantity sold for APAC Consumer per year

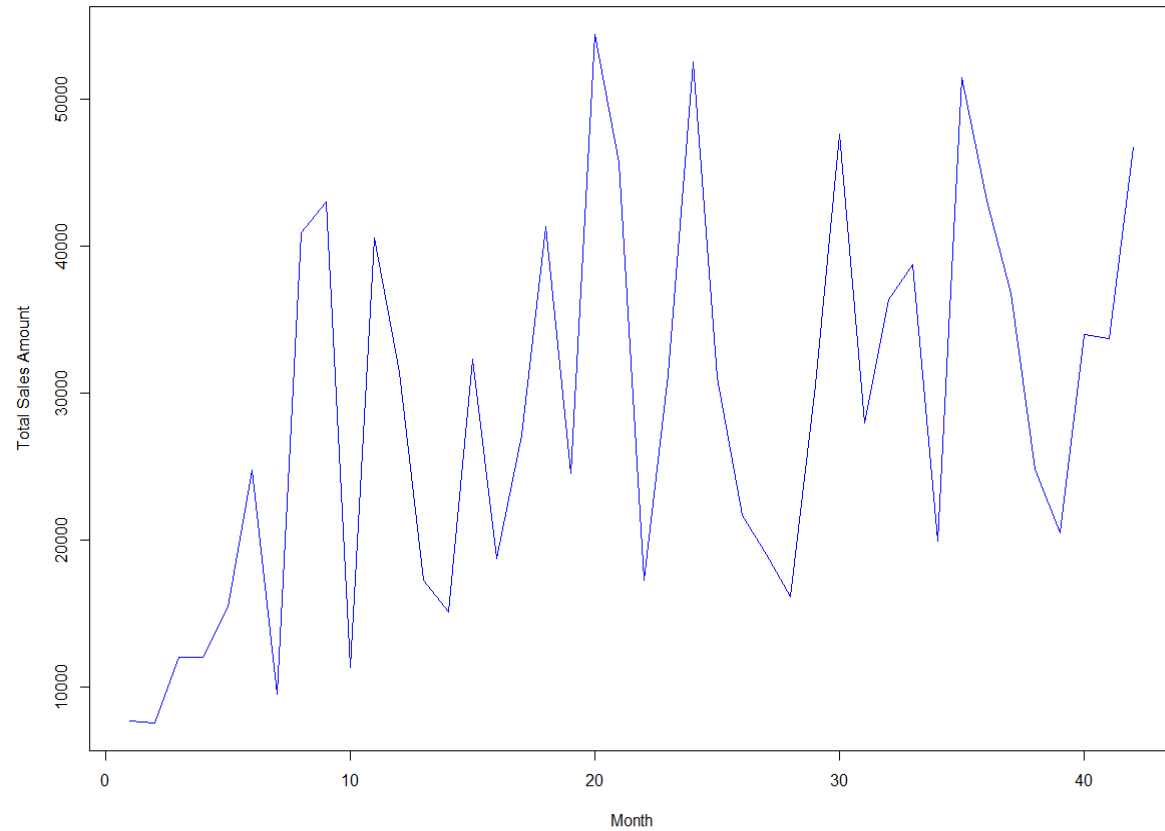


Model Building I

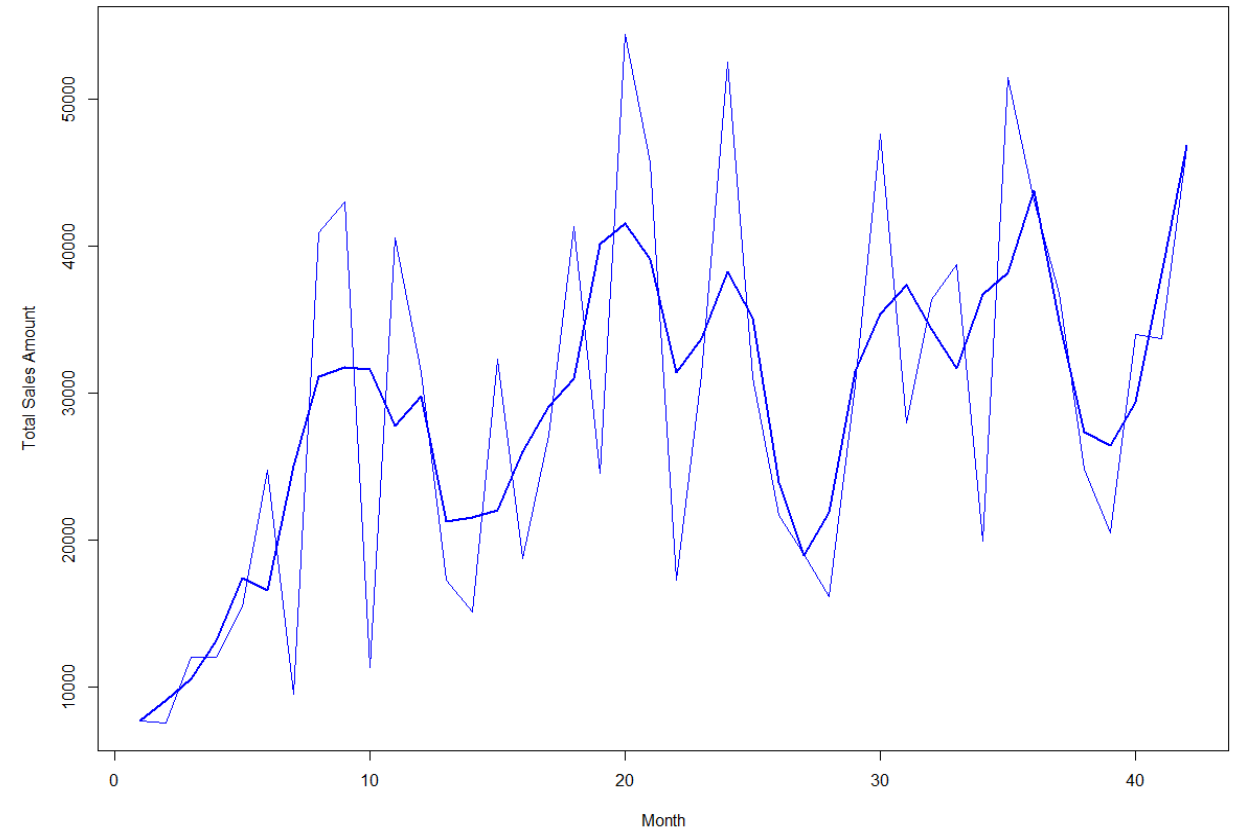
- As discussed in the previous slides we have narrowed down to 2 most profitable segments:
 - EU Consumer
 - APAC Consumer
- Hence, we have used “Classical Decomposition” & auto.arima methodologies in our “Model Building” process.
- Models are built for “EU Consumer Sales” , “EU Consumer Qty” , “APAC Consumer Sales” and “APAC Consumer Qty”
- Model Building process:
 - Plotting the original time series.
 - Smoothing the original time series using “simple moving average” window size of 3.
 - Build a linear model to predict the values globally. Sinusoidal function is used to address the seasonality.
 - Removing the predicted values from the time series to get a stationary time series.
 - Modelling the stationary series using auto.arima function.
 - Applied auto.arima on the original time series.
 - Compared the models (linear model) & model built with “auto.arima” on the original time series using MAPE metric.

Smoothing Time Series: EU Consumer Sales

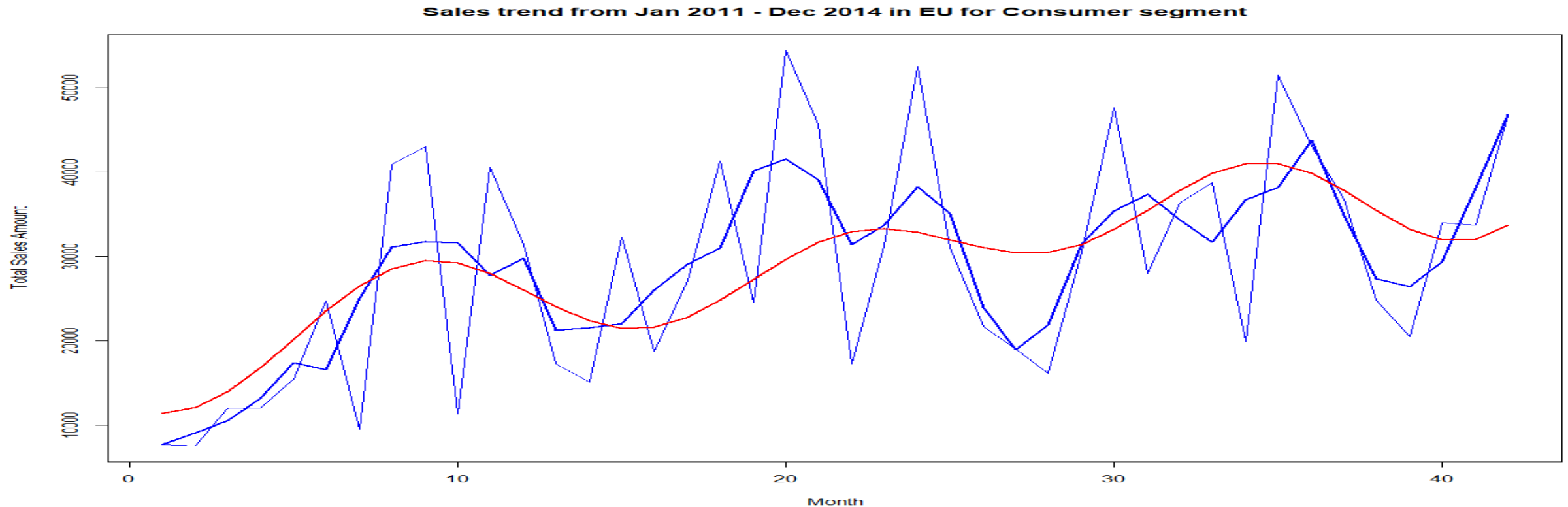
Sales trend from Jan 2011 - Dec 2014 in EU for Consumer segment



Sales trend from Jan 2011 - Dec 2014 in EU for Consumer segment



Modelling Trend & Seasonality: EU Consumer Sales

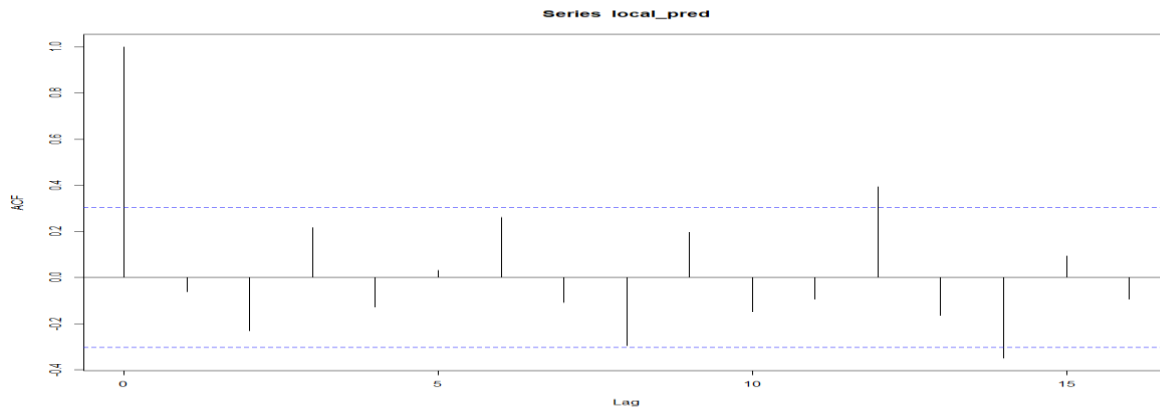
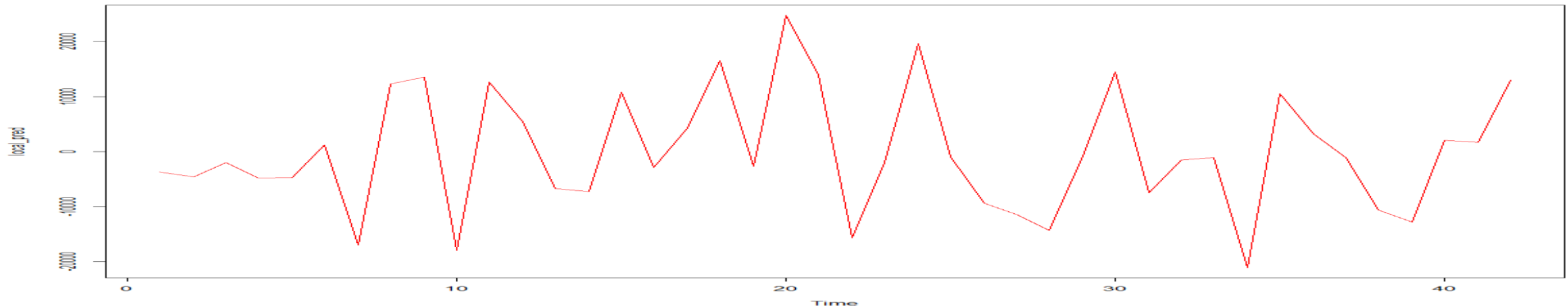


The red line in the above graph is trying to predict the trend and seasonality with the linear model :

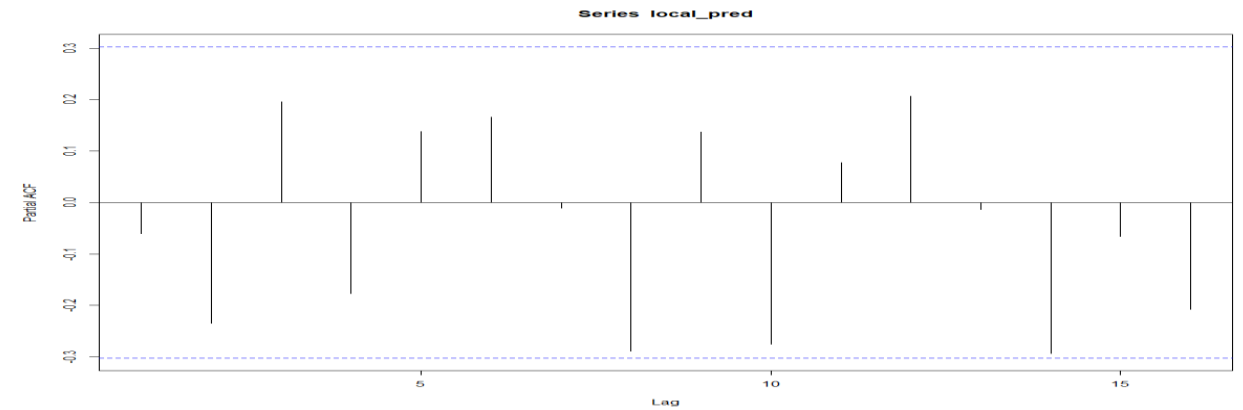
$$\text{Sales} \sim \sin(0.483 * \text{Month}) + \cos(0.4 * \text{Month}) * \text{poly}(\text{Month}, 1) + \text{Month}$$

Stationary Time Series-ACF, PACF Plots : EU Consumer Sales

Stationary Component



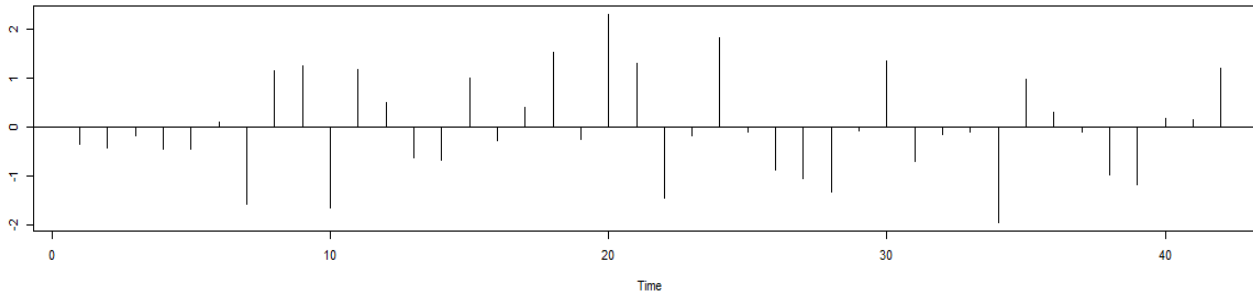
ACF Plot



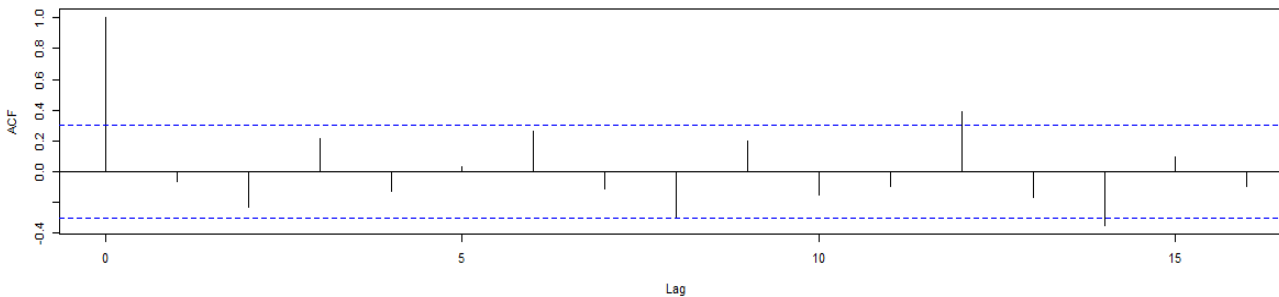
PACF Plot

Time Series Diagnostics : EU Consumer Sales

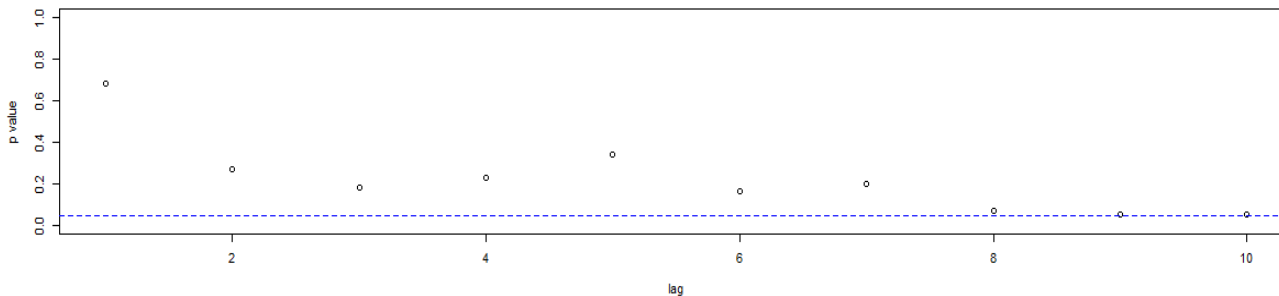
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



Series: local_pred
ARIMA(0,0,0) with zero mean

Verify if the residual series is stationary or not:

Dickey-Fuller test , p-value = 0.07

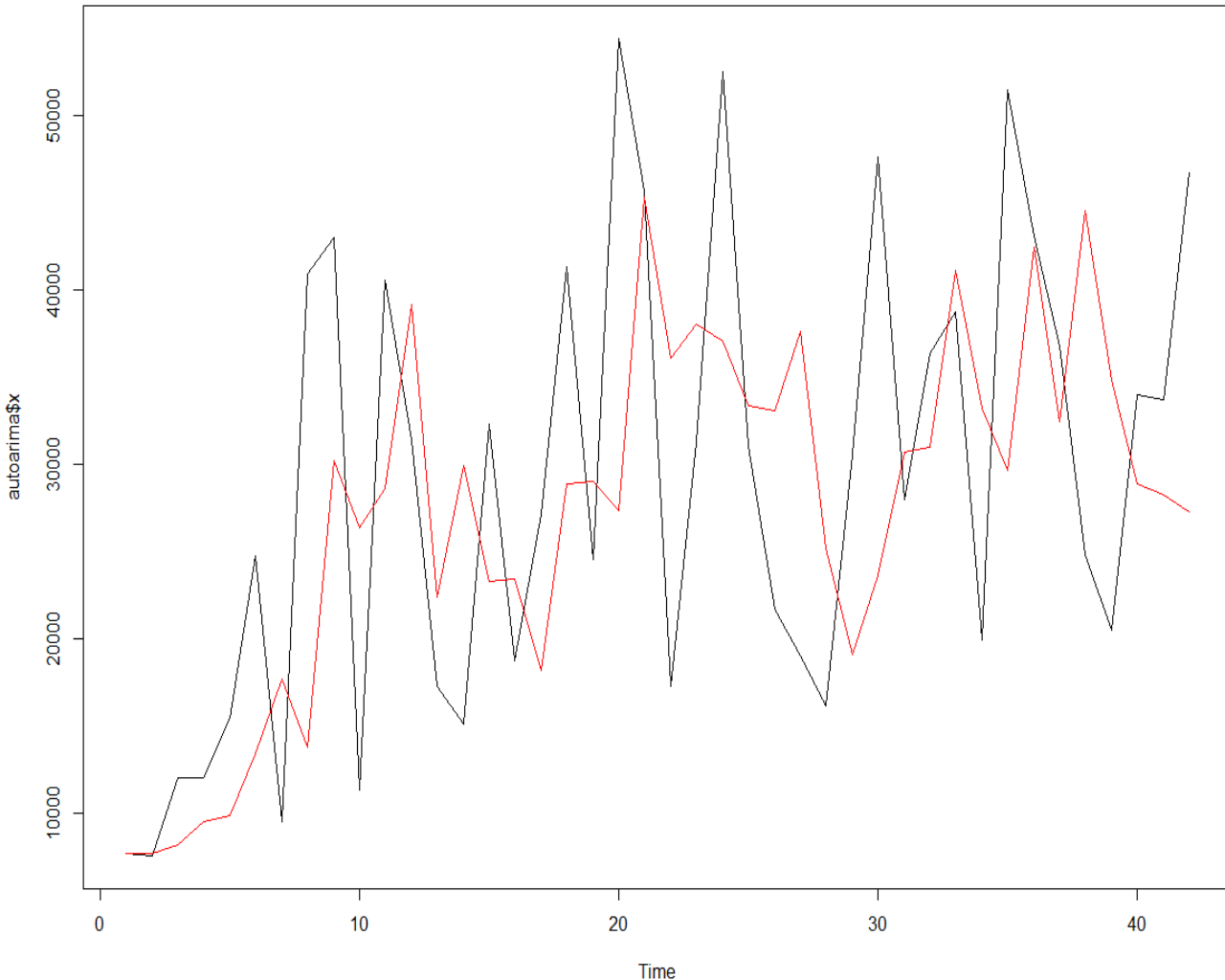
Kpss test , p-value = 0.1

By Dickey fuller test, series is non- stationary
But by Kpss test, **residual series is stationary**

Predicting the last 6 months sales using our manually built model. Checking its accuracy.

MAPE = 22.28

Model Building via auto.arima for : EU Consumer Sales



Series: local_pred

ARIMA(0,0,0) with zero mean

Verify if the residual series is stationary or not:

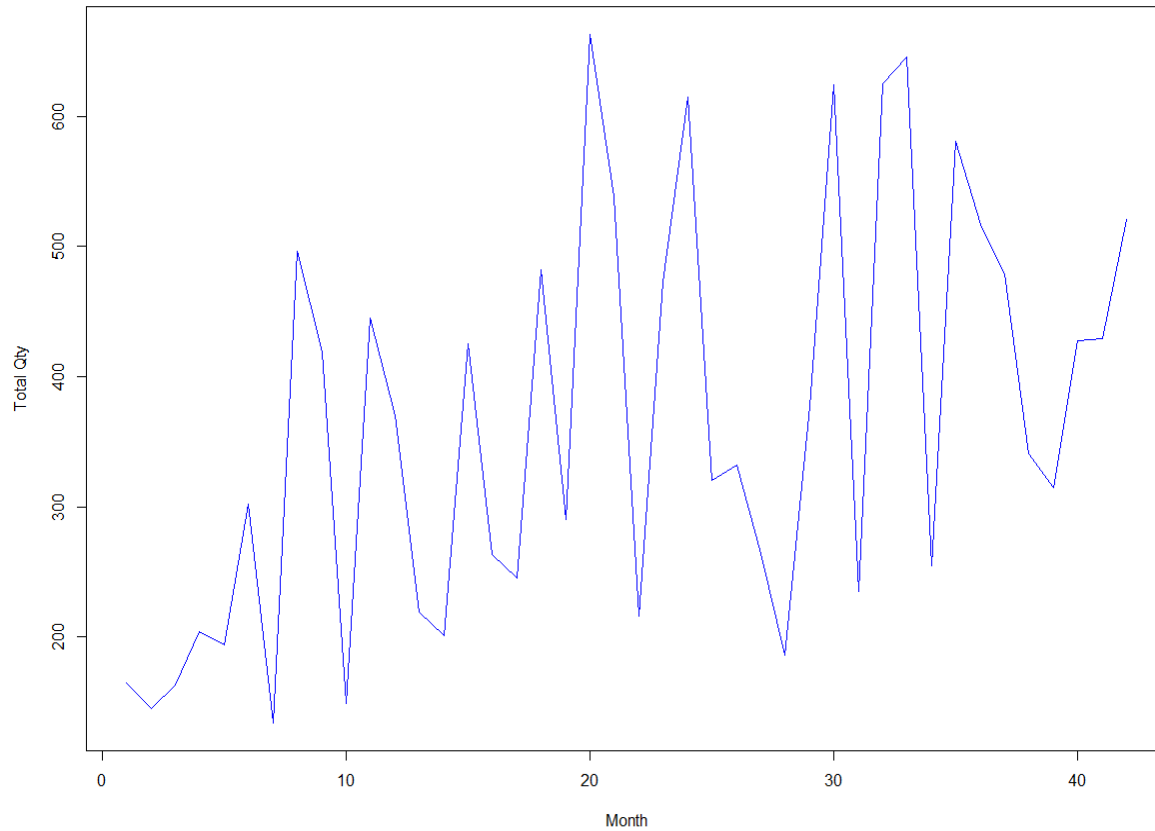
Dickey-Fuller test , p-value = 0.01

Kpss test , p-value = 0.1

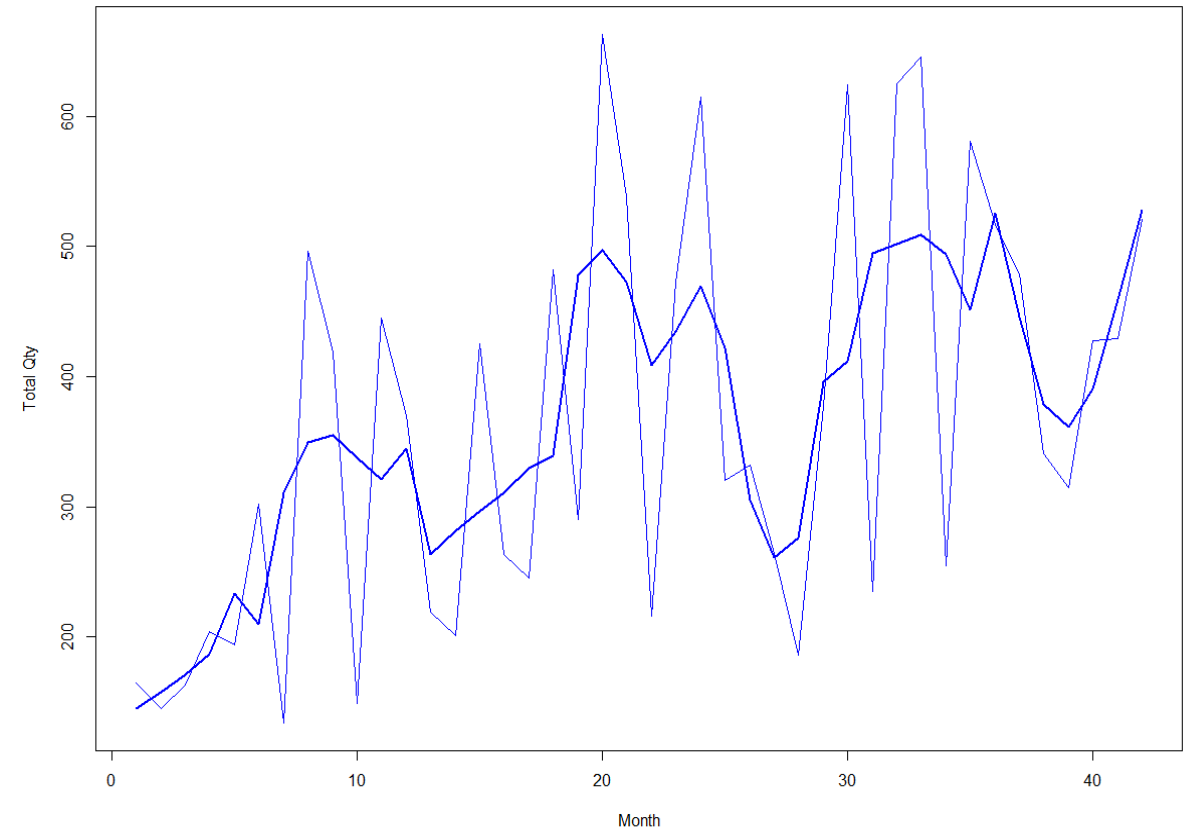
It confirms residual series is stationary

Smoothing Time Series: EU Consumer Quantity

Qty trend from Jan 2011 - Dec 2014 in EU for Consumer segment

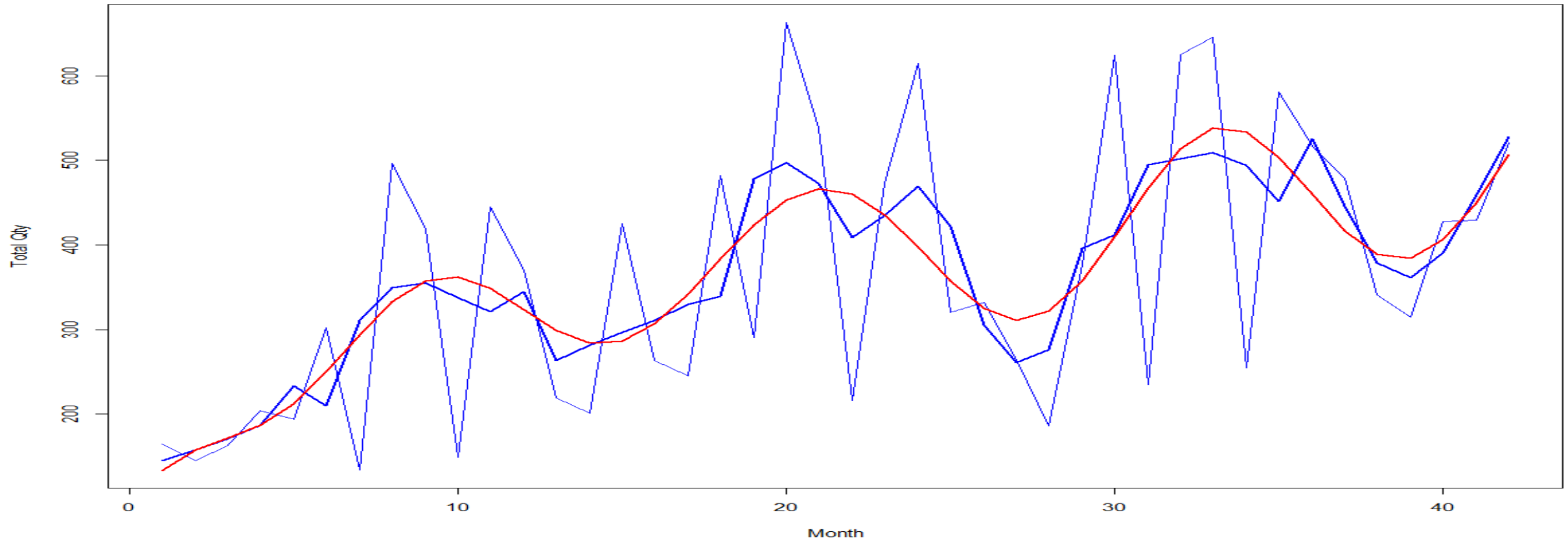


Qty trend from Jan 2011 - Dec 2014 in EU for Consumer segment



Modelling Trend & Seasonality: EU Consumer Quantity

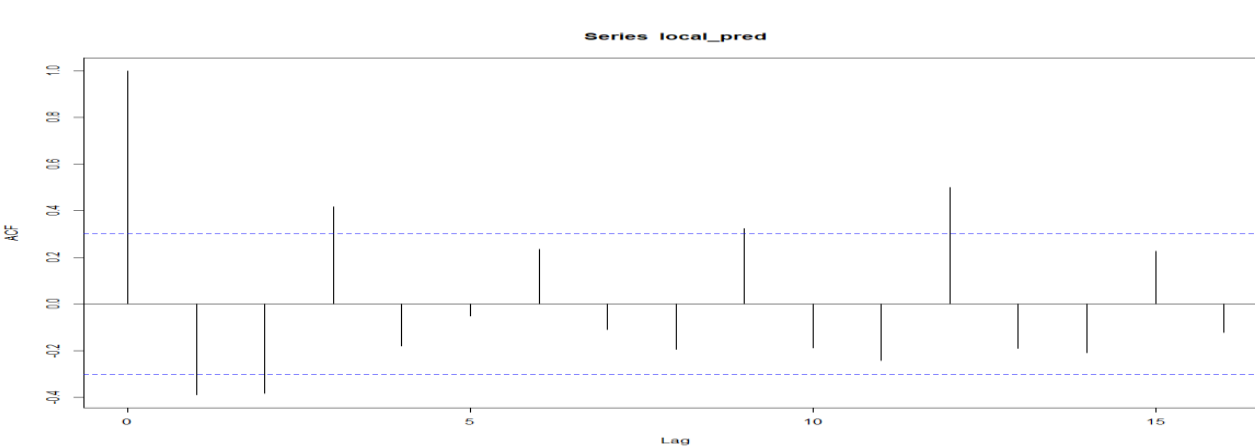
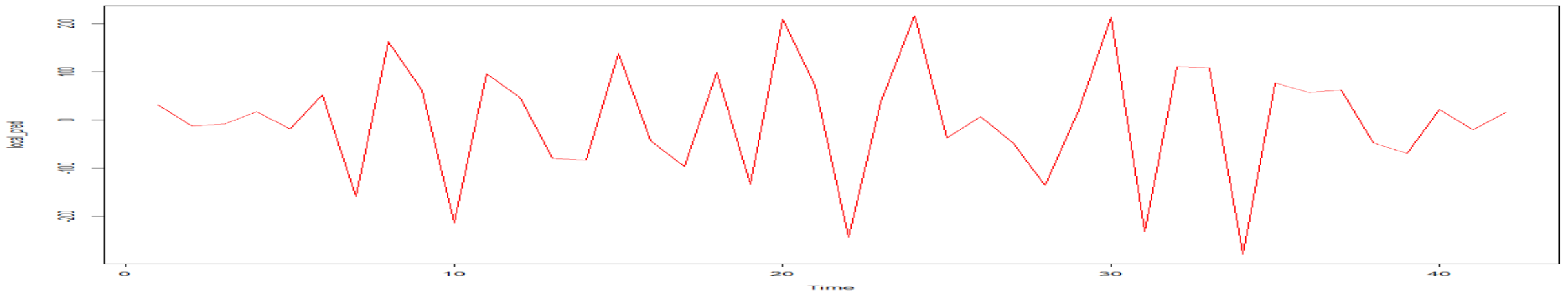
Qty trend from Jan 2011 - Dec 2014 in EU for Consumer segment



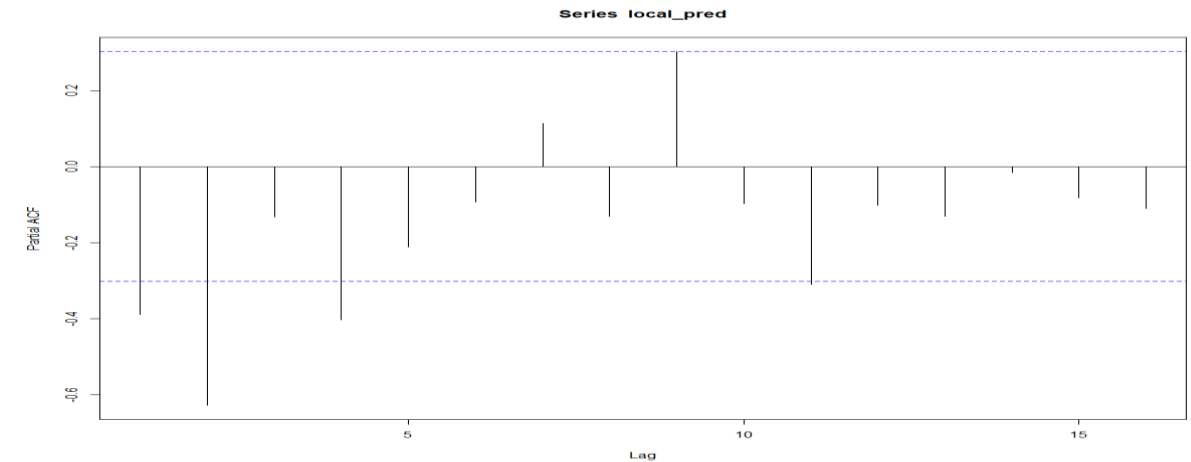
The **red** line in the above graph is trying to predict the trend and seasonality with the linear model :
 “Qty ~ sin(0.60*Month) * poly(Month,3) + cos(0.6*Month) * poly(Month,3) + Month”

Stationary Time Series-ACF, PACF Plots : EU Consumer Quantity

Stationary Component



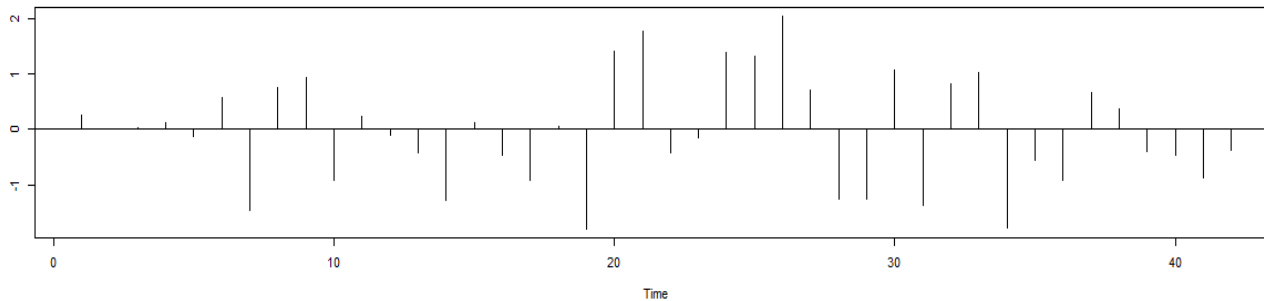
ACF Plot



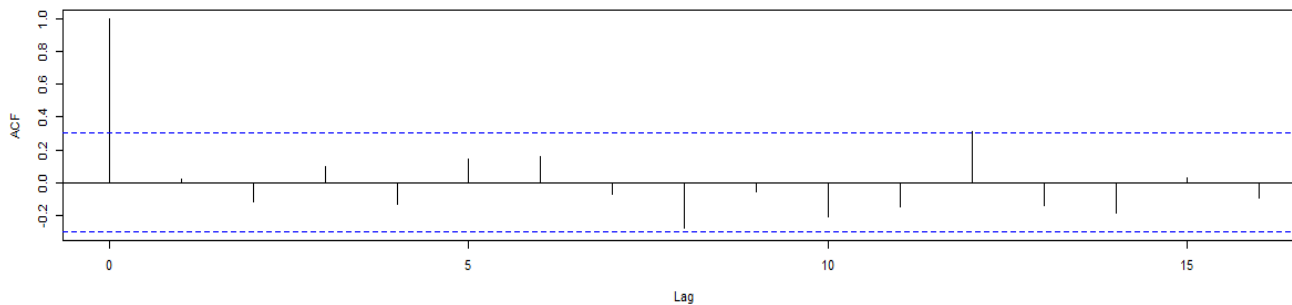
PACF Plot

Time Series Diagnostics : EU Consumer Quantity

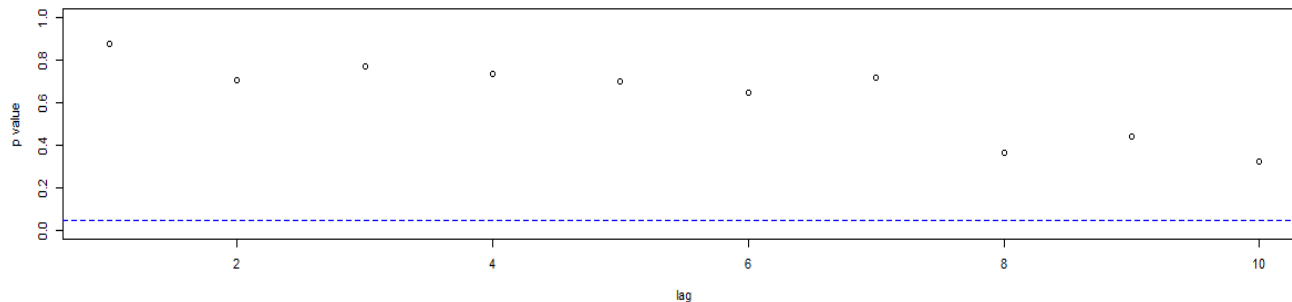
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



Series: local_pred
ARIMA(2,0,1) with zero mean

Verify if the residual series is stationary or not:

Dickey-Fuller test , p-value = 0.1019

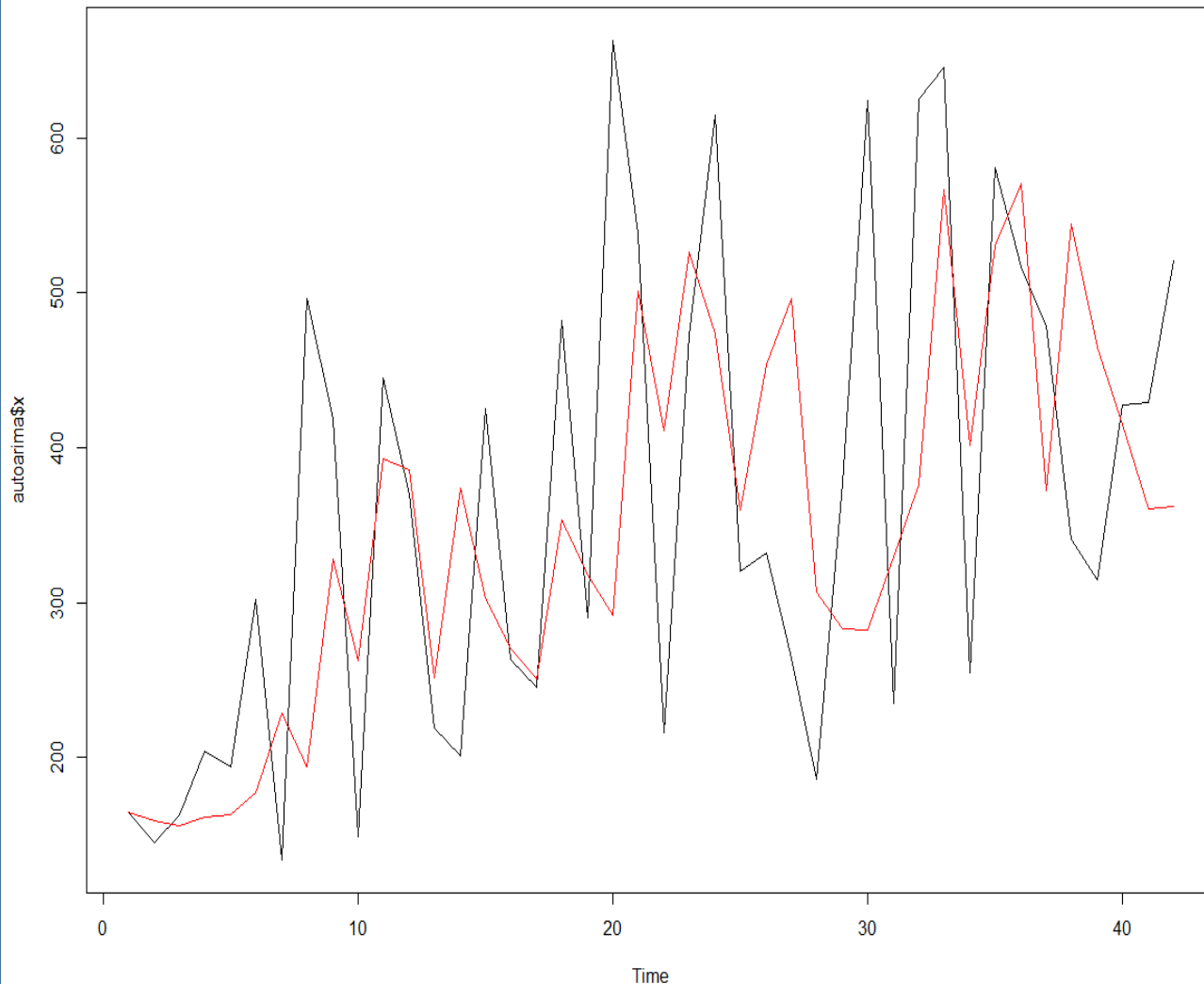
Kpss test , p-value = 0.1

It confirms that **residual series is stationary**

Predicting the last 6 months sales quantity using our manually built model. Checking its accuracy.

MAPE = 24.98

Model Building via auto.arima for : EU Consumer Qty



Series: local_pred

ARIMA(2,1,0) with zero mean

Verify if the residual series is stationary or not:

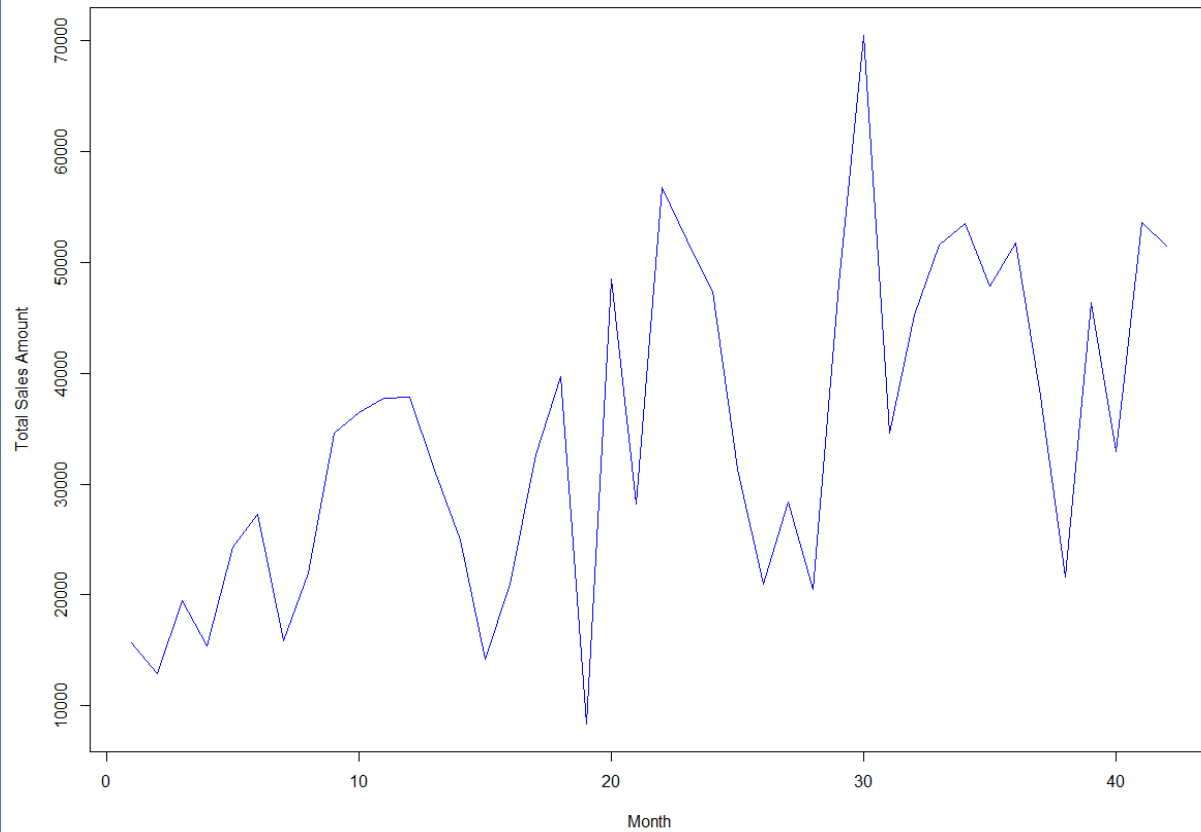
Dickey-Fuller test , p-value = 0.04

Kpss test , p-value = 0.1

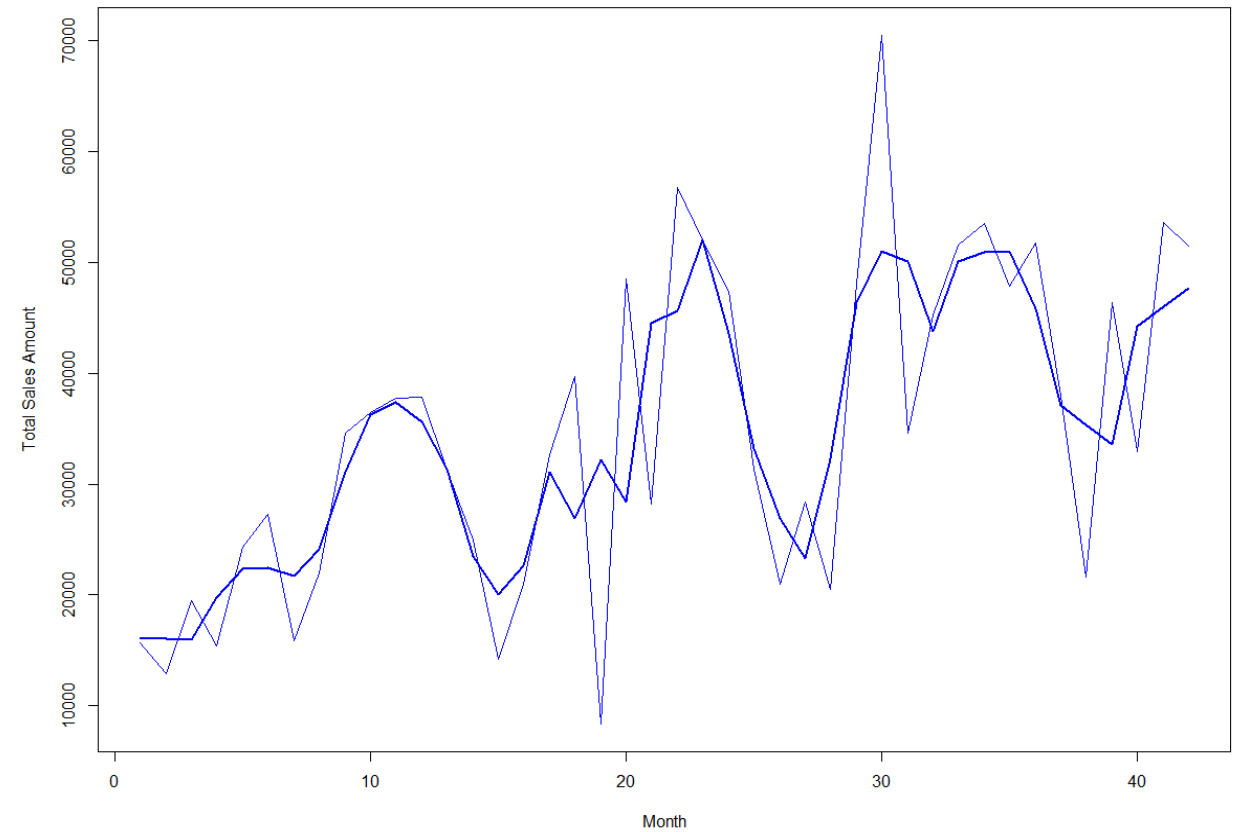
It confirms residual series is stationary

Smoothing Time Series: APAC Consumer Sales

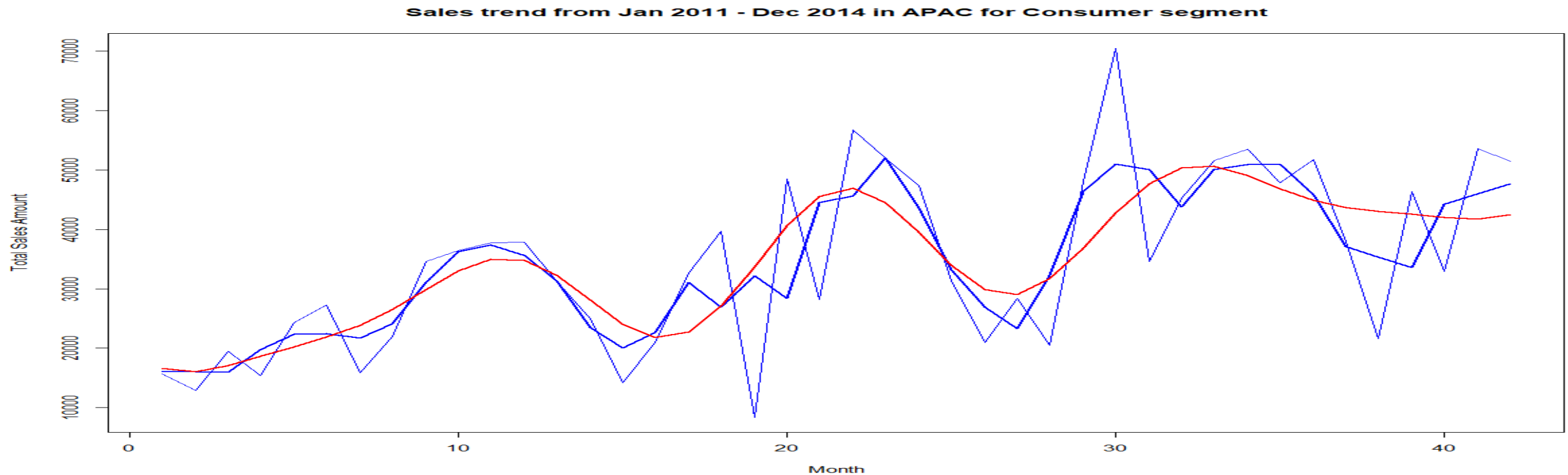
Sales trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



Sales trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



Modelling Trend & Seasonality: APAC Consumer Sales

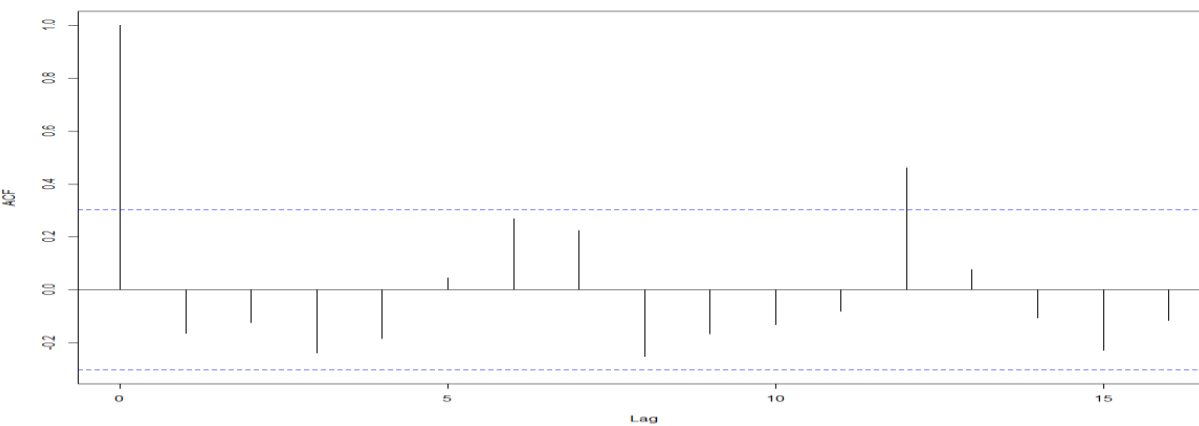
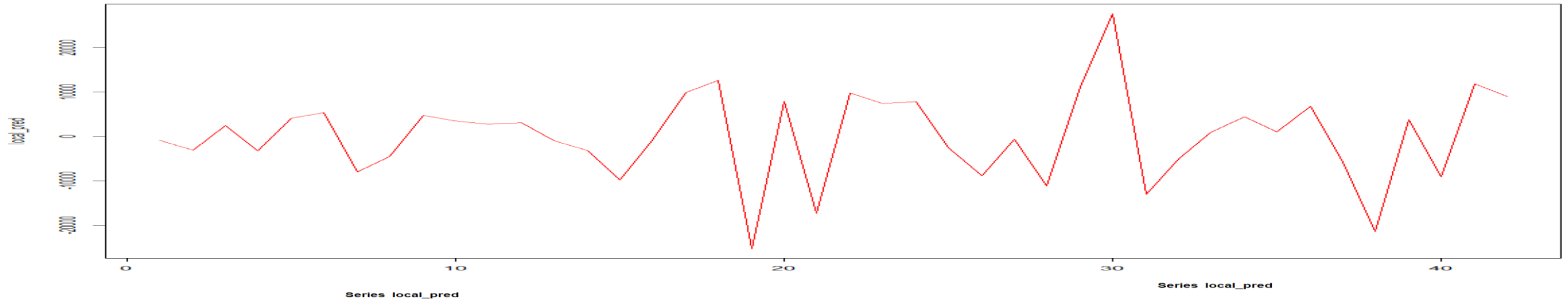


The red line in the above graph is trying to predict the trend and seasonality with the linear model :

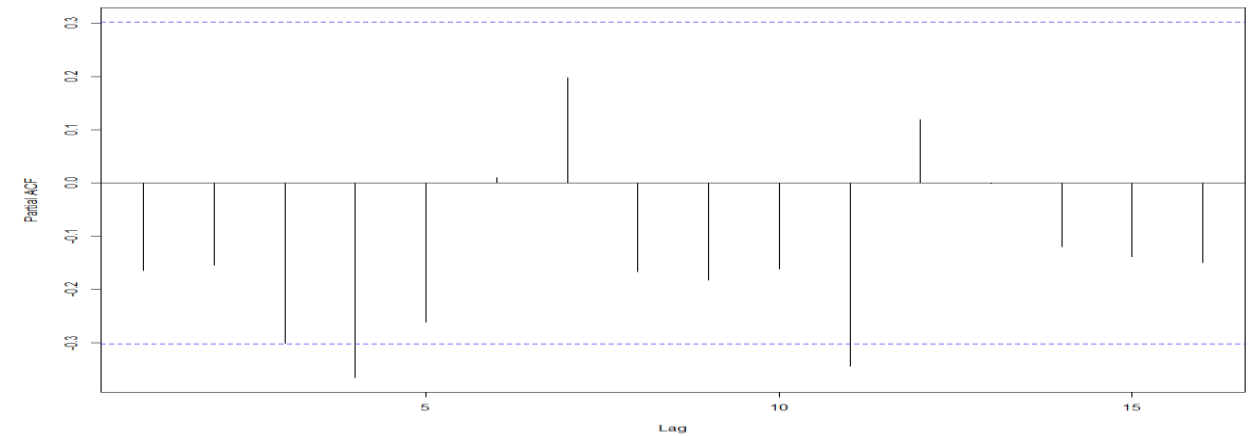
$$\text{Sales} \sim \sin(0.59 * \text{Month}) * \text{poly}(\text{Month}, 3) + \cos(0.55 * \text{Month}) * \text{poly}(\text{Month}, 2) + \text{Month}$$

Stationary Time Series-ACF, PACF Plots : APAC Consumer Sales

Stationary Component

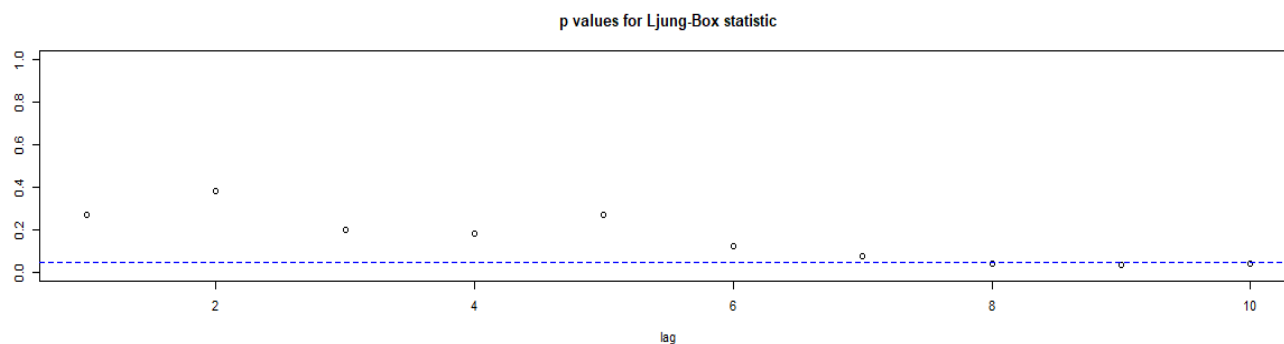
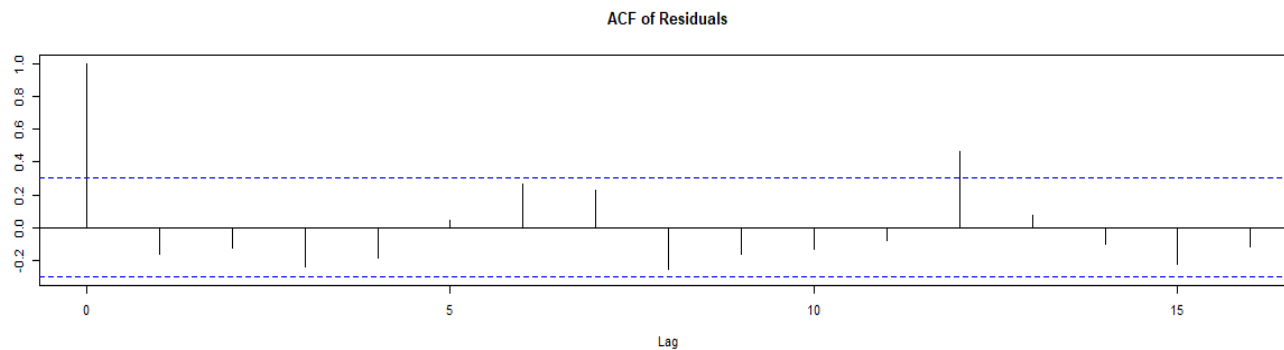
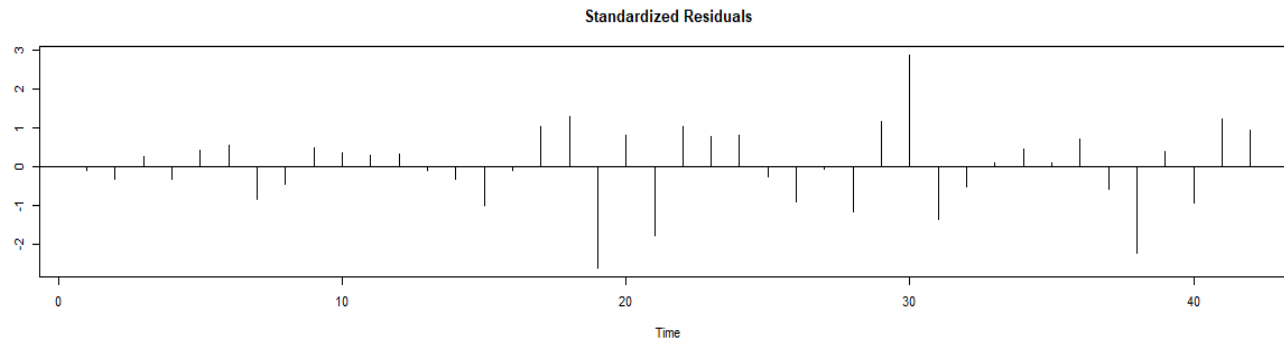


ACF Plot



PACF Plot

Time Series Diagnostics : APAC Consumer Sales



Series: local_pred
ARIMA(0,0,0) with zero mean

Verify if the residual series is stationary or not:

Dickey-Fuller test , p-value = 0.01

Kpss test , p-value = 0.1

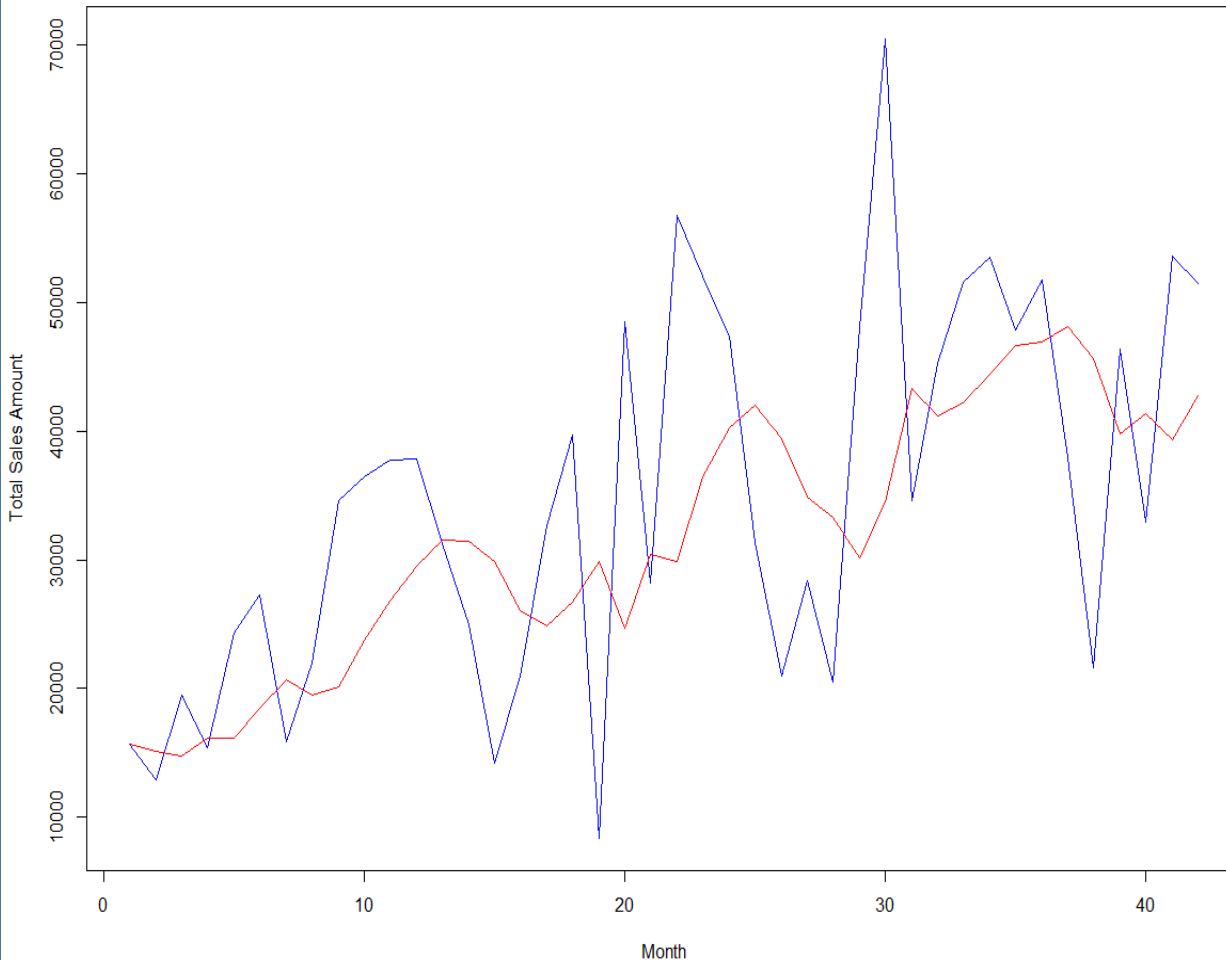
It confirms that **residual series is stationary**

Predicting the last 6 months sales using our manually built model. Checking its accuracy.

MAPE = 19.43

Model Building via auto.arima for : APAC Consumer Sales

Sales trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



Series: local_pred
ARIMA(0,1,1) with zero mean

Verify if the residual series is stationary or not:

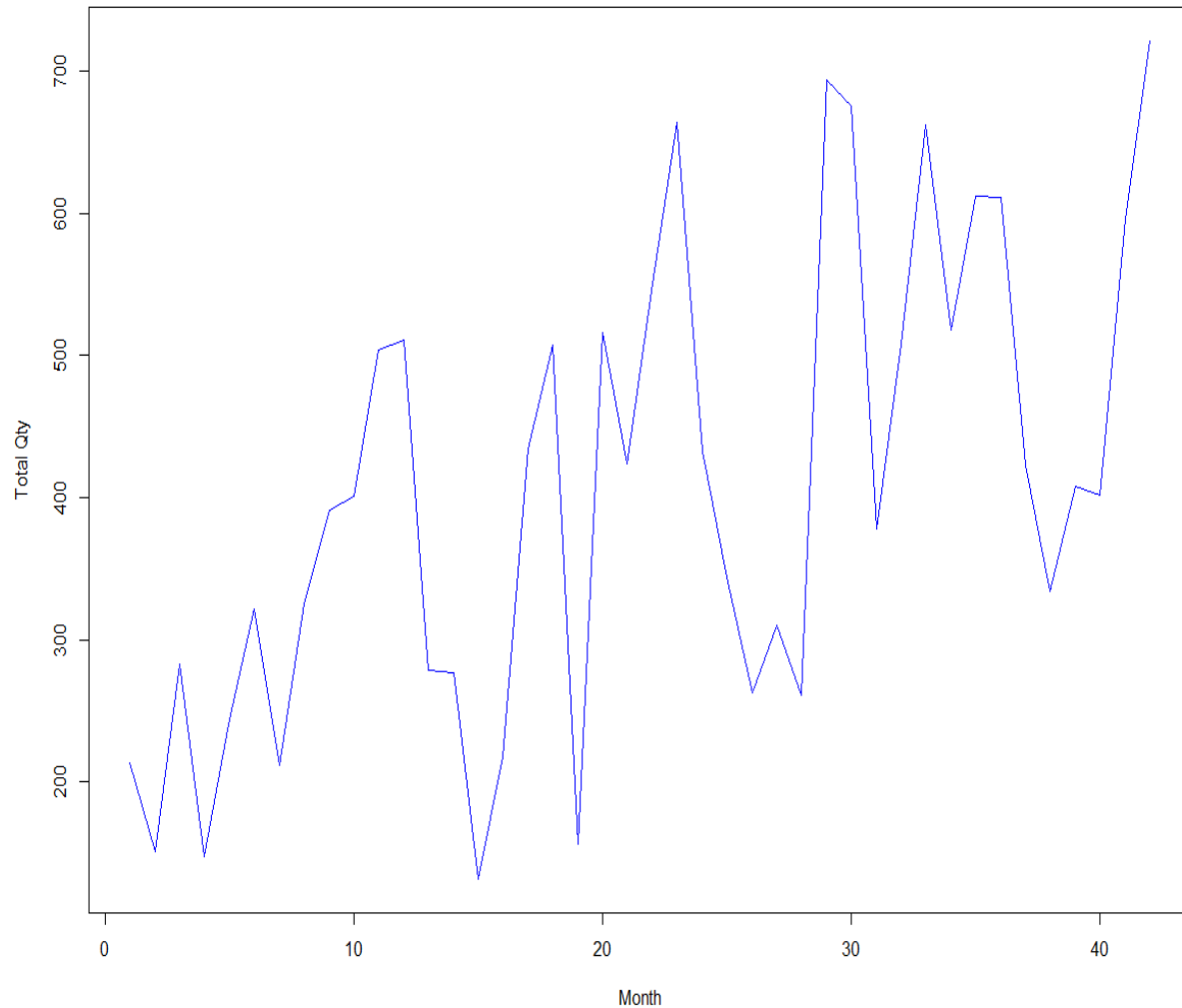
Dickey-Fuller test , p-value = 0.04

Kpss test , p-value = 0.1

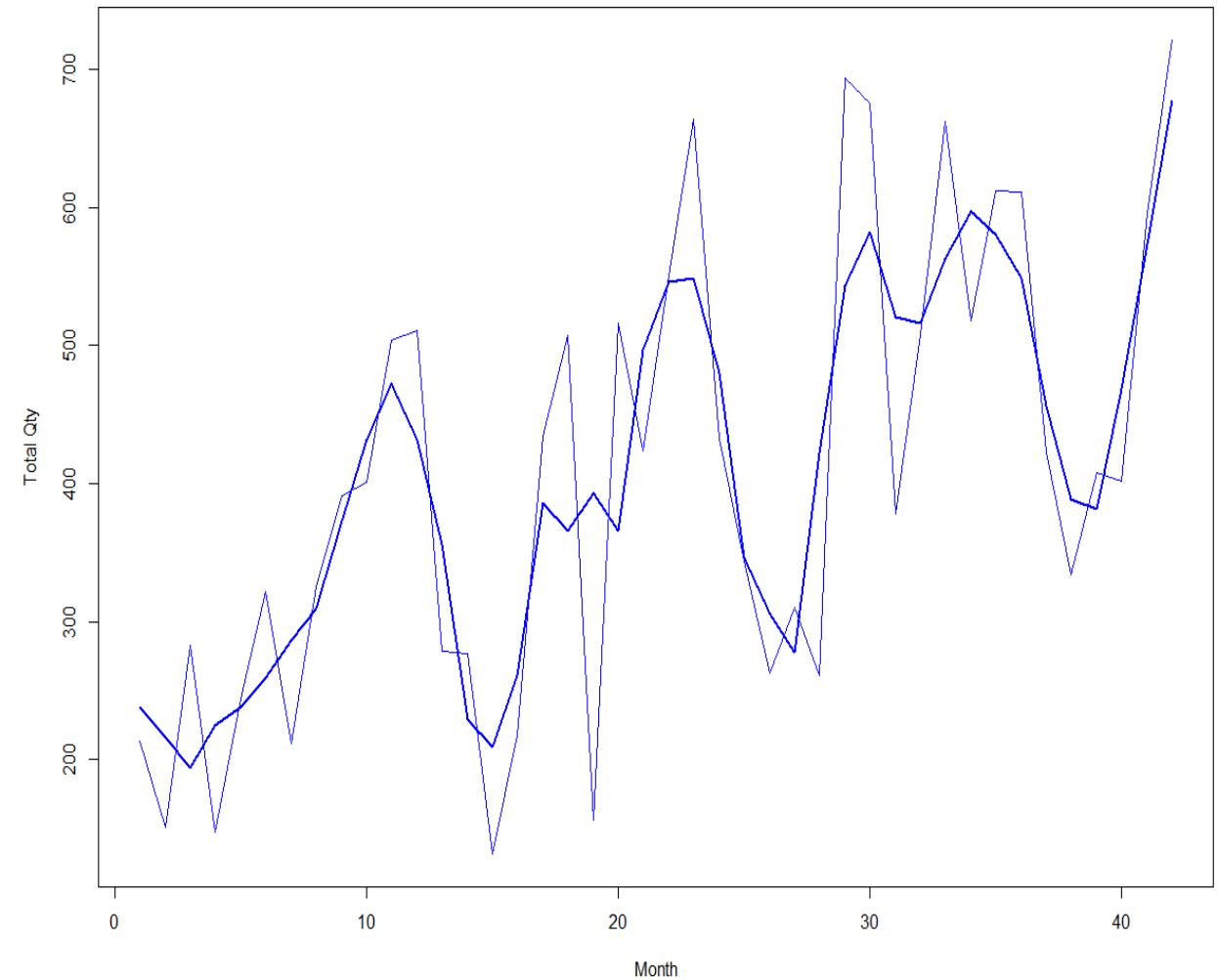
It confirms residual series is stationary

Smoothing Time Series: APAC Consumer Qty

Qty trend from Jan 2011 - Dec 2014 in APAC for Consumer segment

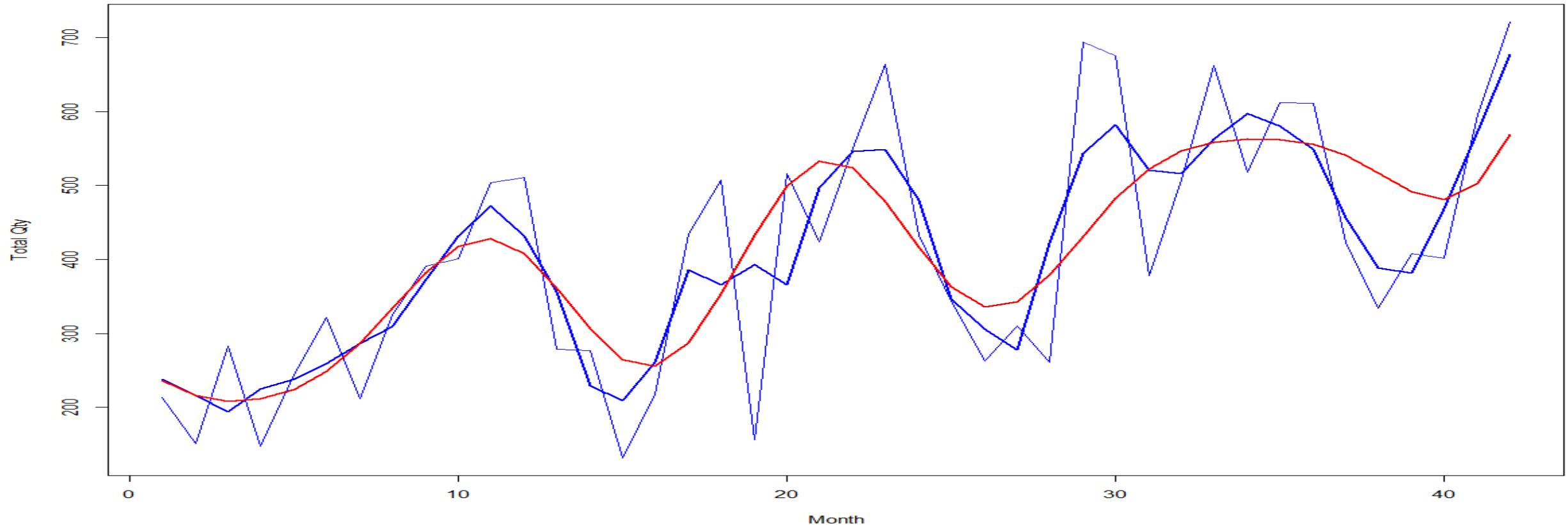


Qty trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



Modelling Trend & Seasonality: APAC Consumer Qty

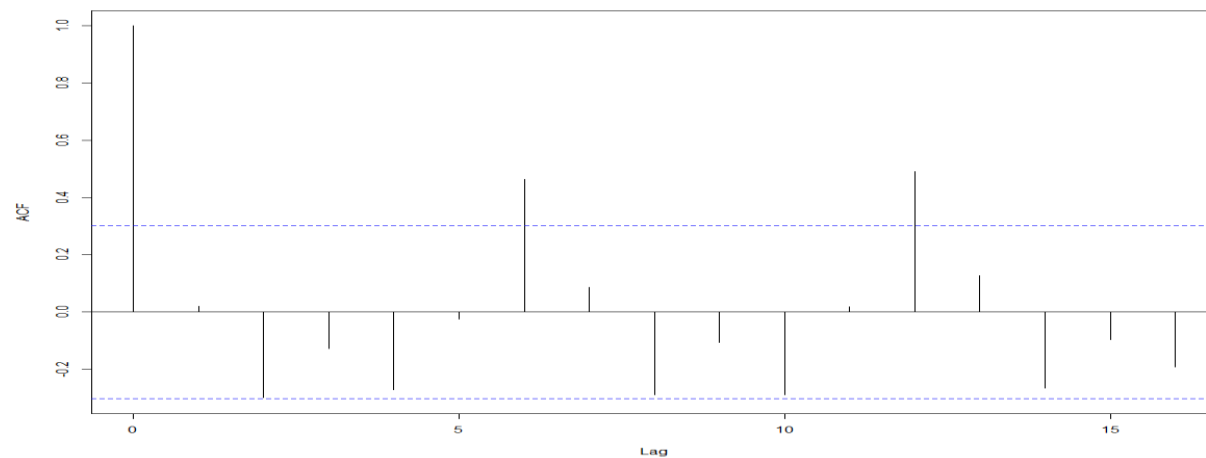
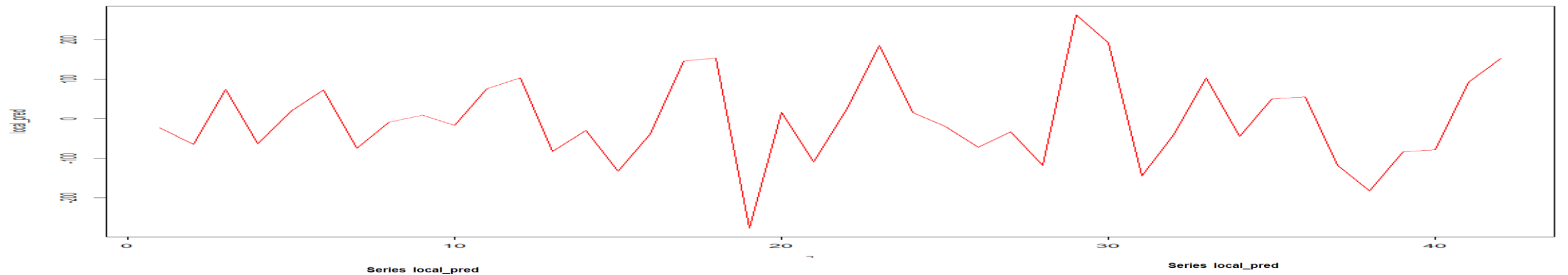
Qty trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



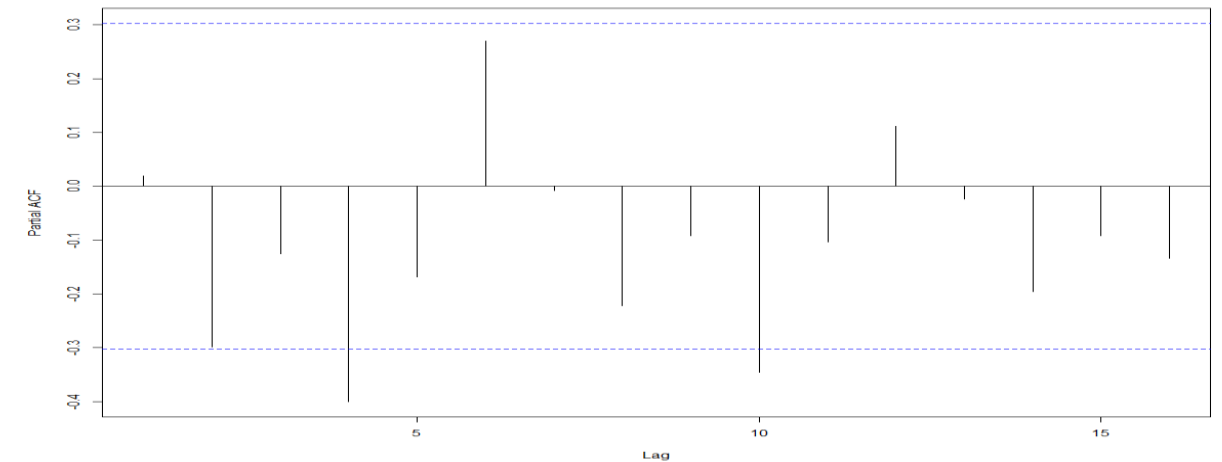
The red line in the above graph is trying to predict the trend and seasonality with the linear model :
 $Qty \sim \sin(0.59 * Month) * \text{poly}(Month, 3) + \cos(0.55 * Month) * \text{poly}(Month, 3) + Month$

Stationary Time Series-ACF, PACF Plots : APAC Consumer Qty

Stationary Component



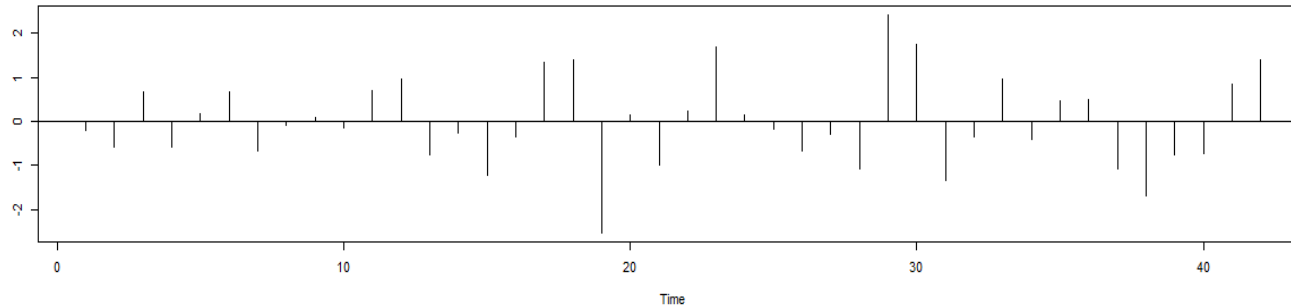
ACF Plot



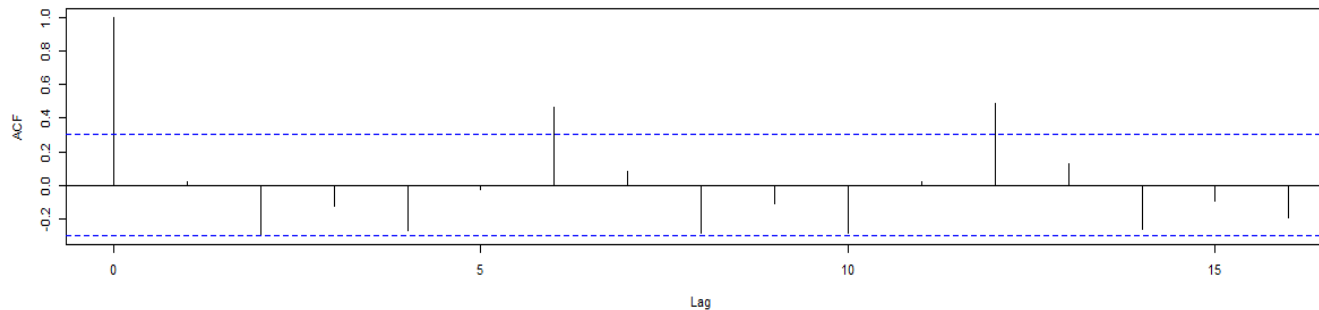
PACF Plot

Time Series Diagnostics : APAC Consumer Qty

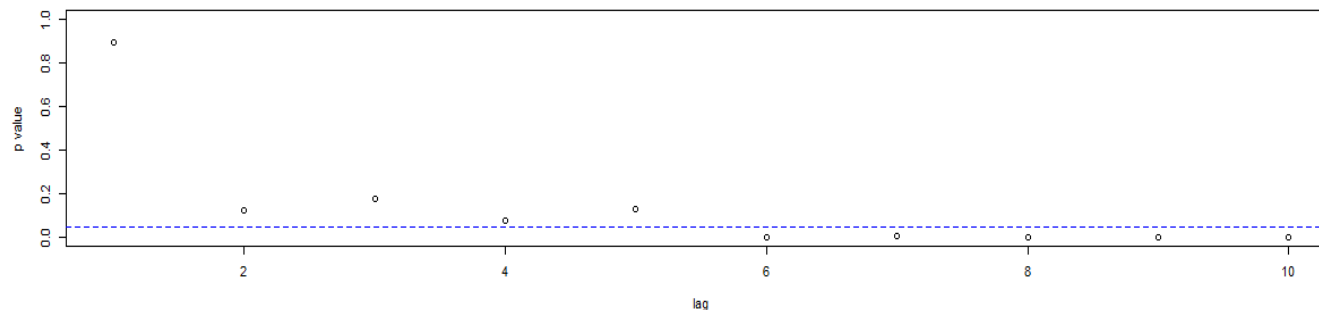
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



Series: local_pred
ARIMA(0,0,0) with zero mean

Verify if the residual series is stationary or not:

Dickey-Fuller test , p-value = 0.01

Kpss test , p-value = 0.1

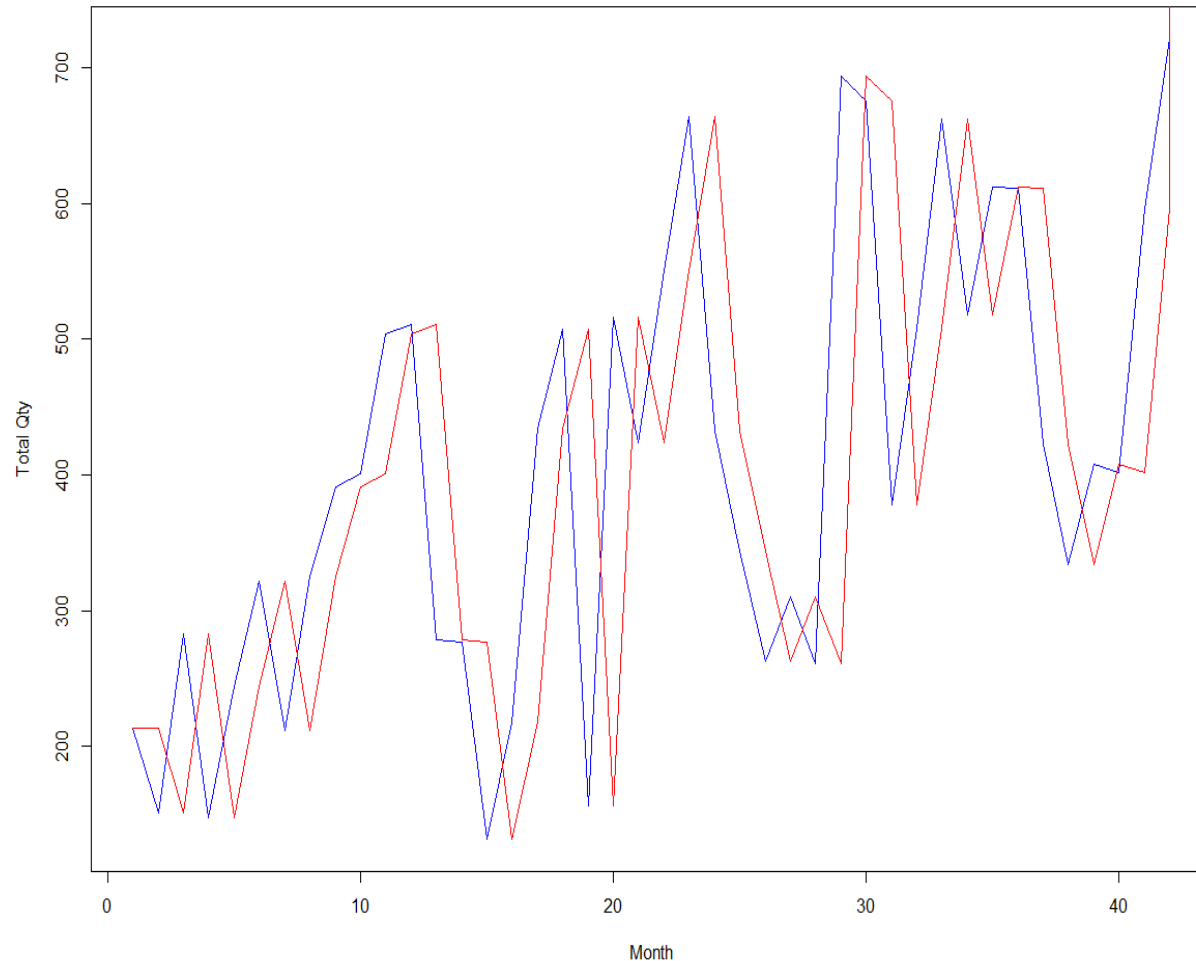
It confirms that **residual series is stationary**

Predicting the last 6 months sales quantity using our manually built model. Checking its accuracy.

MAPE = 22.81

Model Building via auto.arima for : APAC Consumer Qty

Qty trend from Jan 2011 - Dec 2014 in APAC for Consumer segment



Series: local_pred

ARIMA(0,1,0) with zero mean

Verify if the residual series is stationary or not:

Dickey-Fuller test , p-value = 0.04

Kpss test , p-value = 0.1

It confirms residual series is stationary

Model Evaluation & Recommendation I

- As we have manually build the models using linear model & using “auto.arima” also
- Comparison of the manually built models and the models built by auto.arima using the MAPE metric and the forecast for next 6 months. We have selected the model that has lower MAPE value.

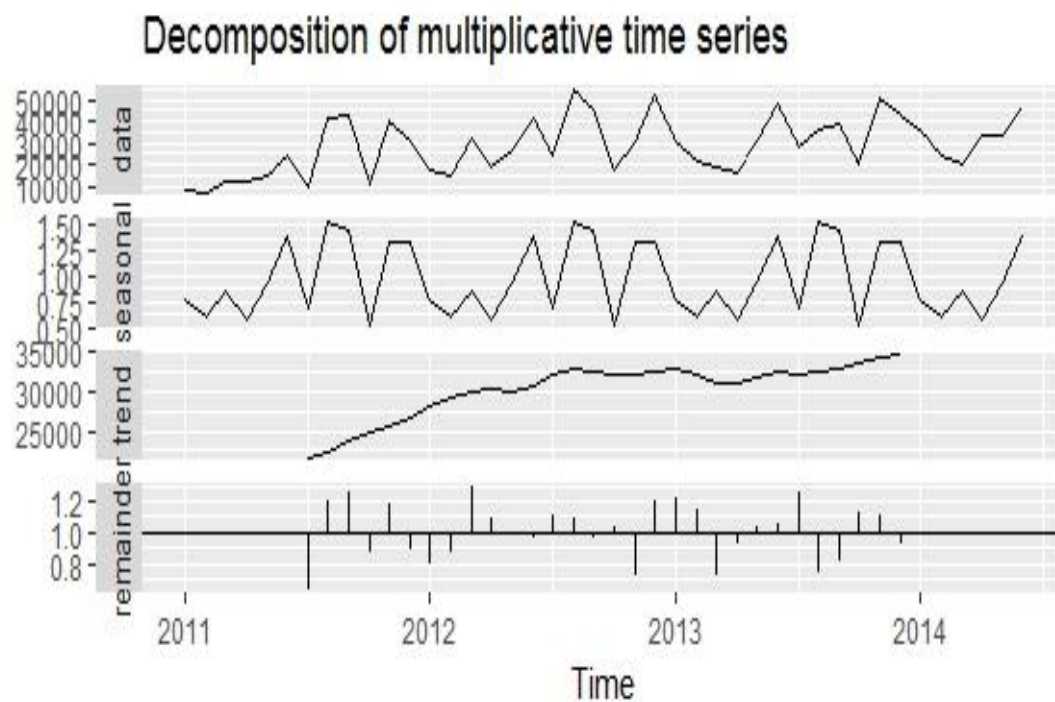
Model	Mape of Manually Built Model	Mape of auto.arima Model	Model Selected	Forecast for next 6 months
EU Consumer Sales	22.98	28.92	Manually Built	54656 , 51821 , 47451 , 42396 , 37678 , 34306
EU Consumer Qty	24.98	30.13	Manually Built	758 , 707 , 640 , 583 , 573 , 645
APAC Consumer Sales	19.43	27.68	Manually Built	55741 , 43470 , 31071 , 23581 , 25149 , 37091
APAC Consumer Qty	22.81	26.24	Manually Built	508 , 417 , 470 , 697 , 1060 1446

Model Evaluation & Recommendation I (Cont.)

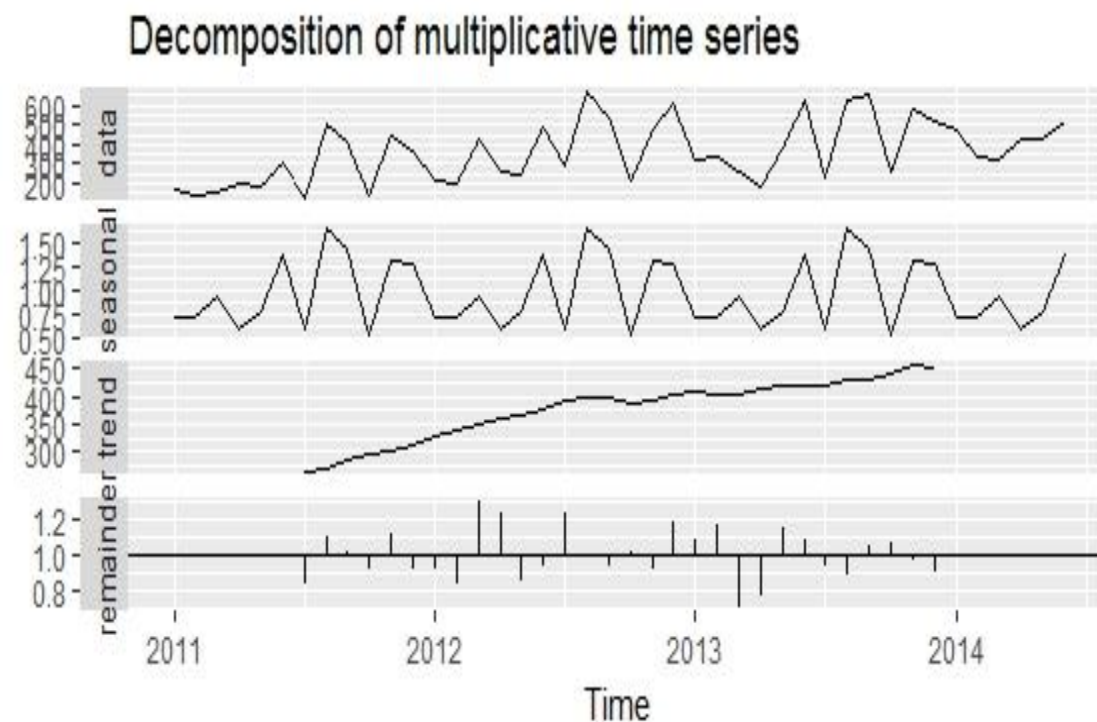
- For the EU Consumer segment the store needs to maintain an inventory of nearly 4000 units for next 6 months (Jan – June , 2015) : (forecast is of : 3906)
- From the EU Consumer segment the store can expect total sales of : 2,68,308 in the next 6 months.
- For APAC Consumer segment the store needs to maintain an inventory of nearly 4700 as the forecast is of 4598 units.
- From APAC Consumer segment the store can expect total sales of : 2,16,103

Holt-Winter's Model Decomposition Plots -1

EU_Consumer_Sales

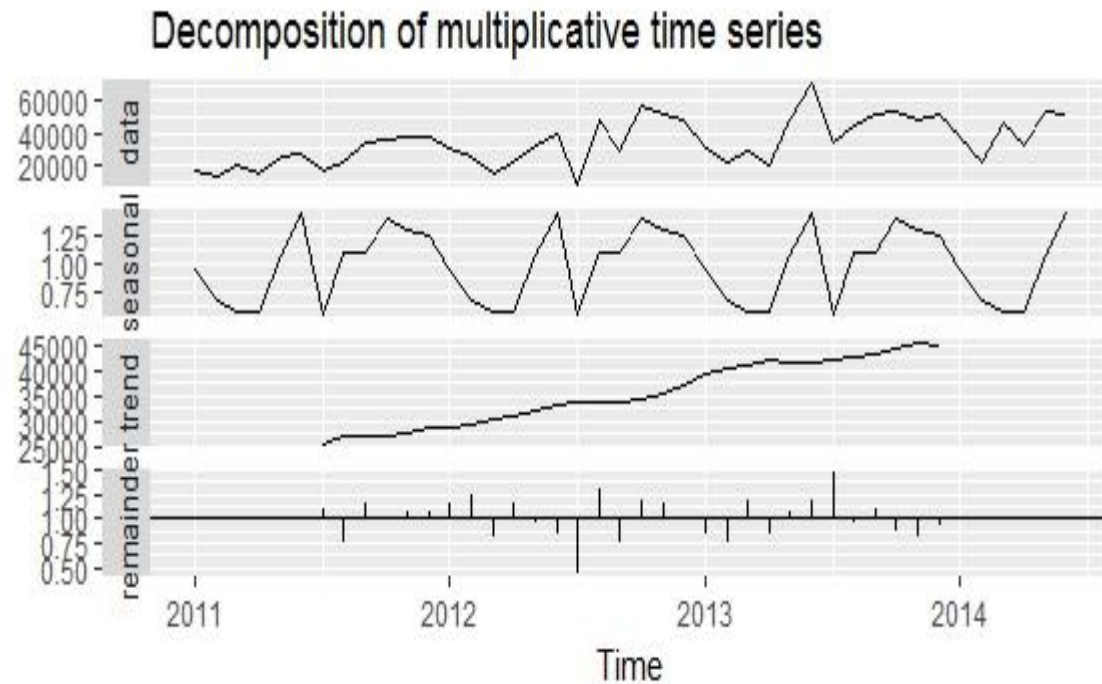


EU_Consumer_Quantity

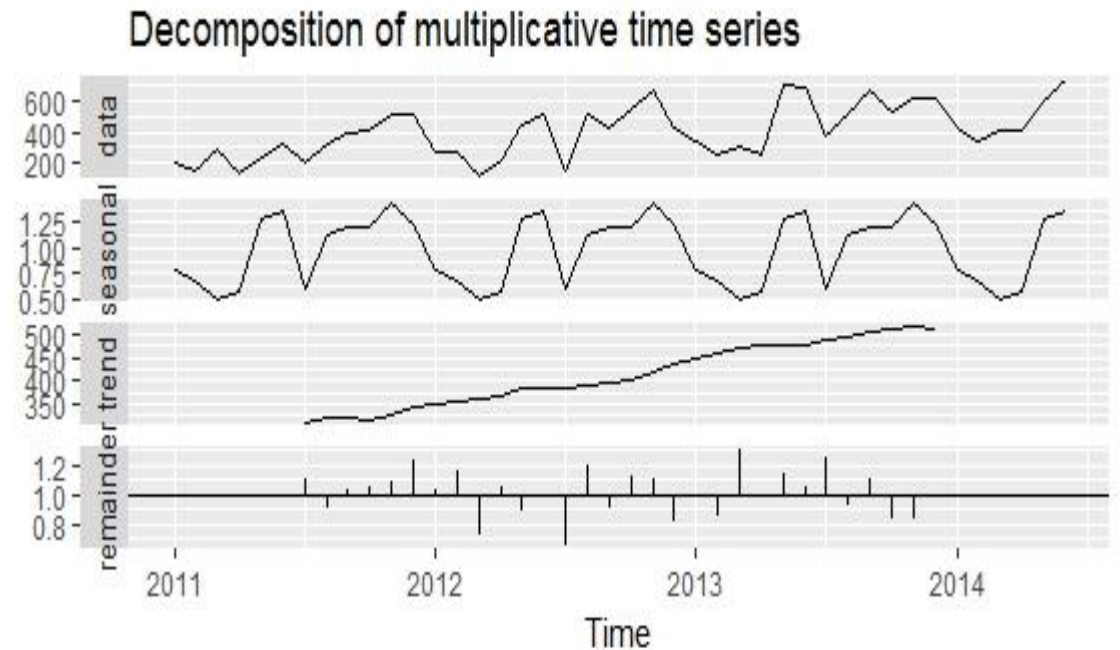


Holt-Winter's Model Decomposition Plots -2

APAC_Consumer_Sales



APAC_Consumer_Quantity



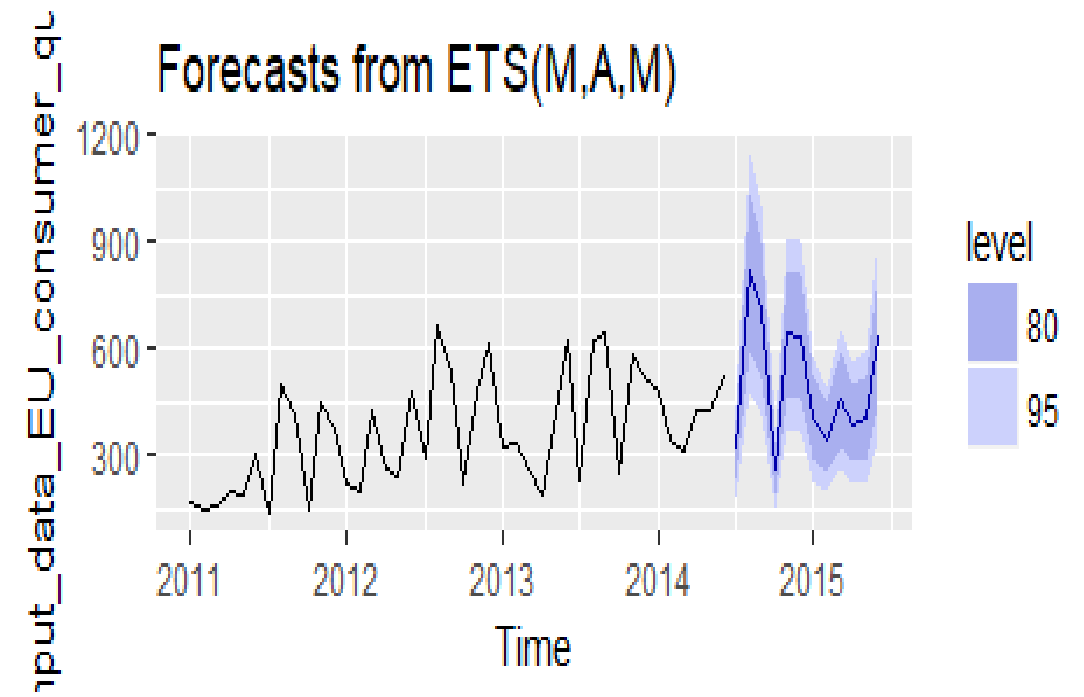
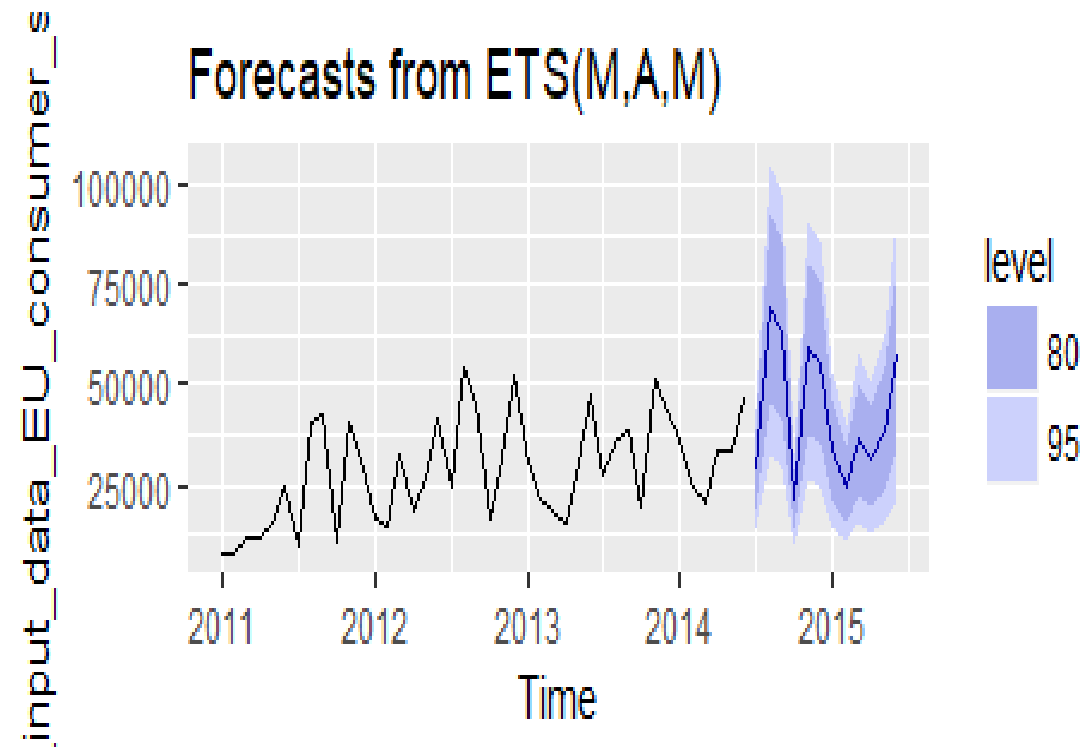
Holt-Winter's Forecast For EU

EU_Consumer_Sales

MASE = 0.7484215

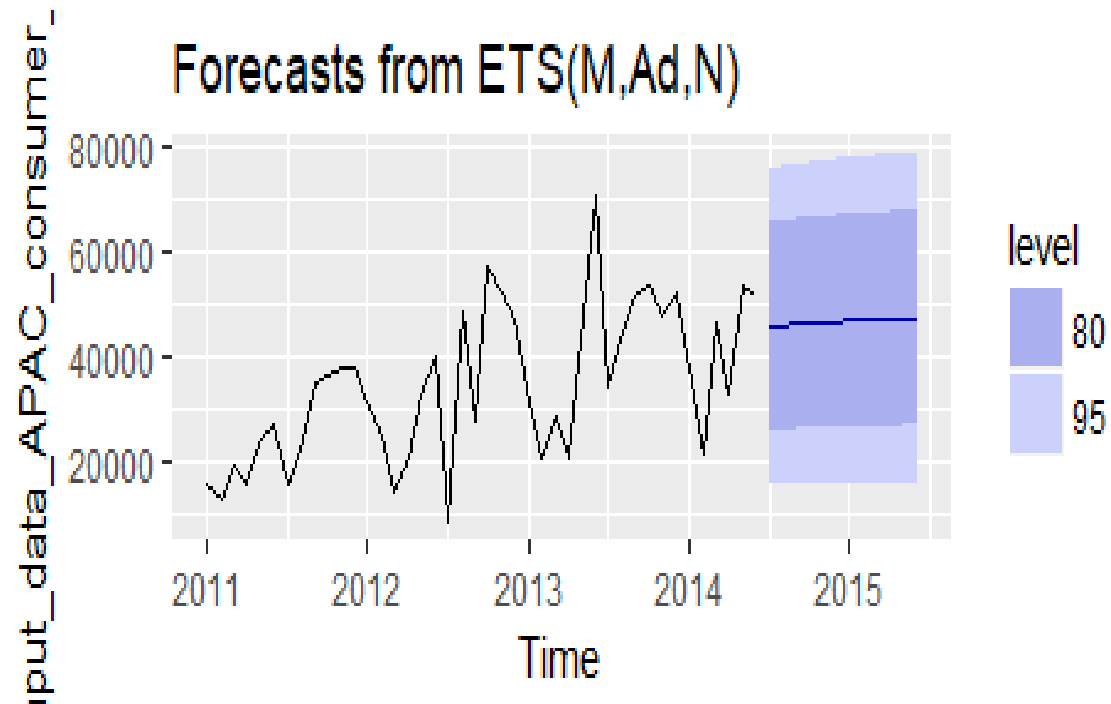
EU_Consumer_Quantity

MASE =1.4213443

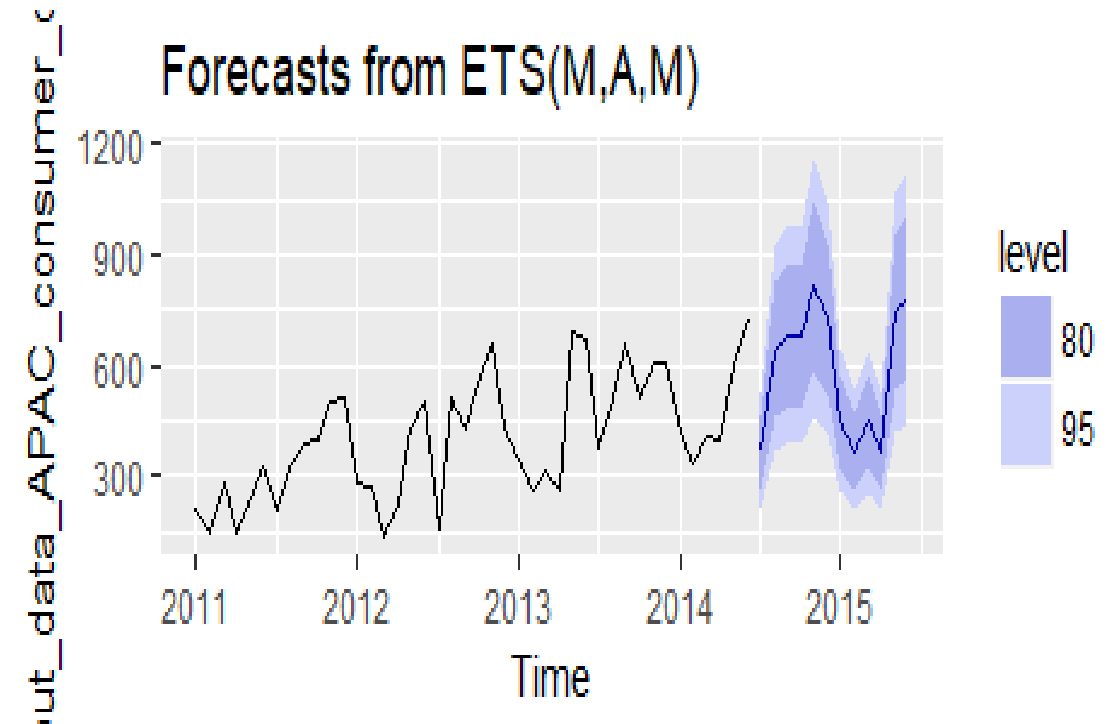


Holt-Winter's Forecast For APAC

APAC_Consumer_Sales
MASE =1.6311864



APAC_Consumer_Quantity
MASE =0.7731437



Holt-Winter's Residuals for EU

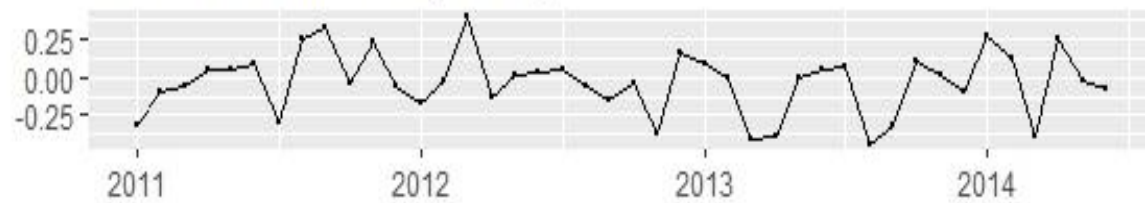
EU_Consumer_Sales

Ljung-Box test data: p-value = $1.86e-05$

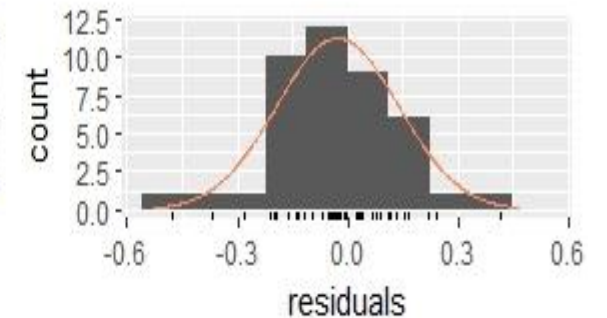
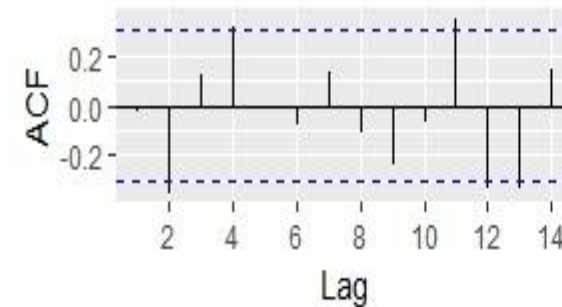
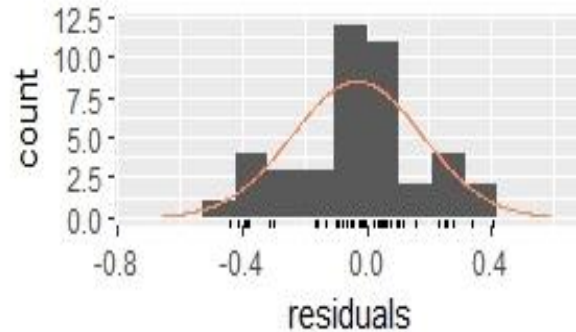
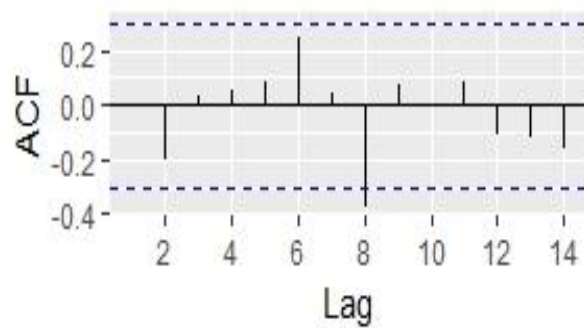
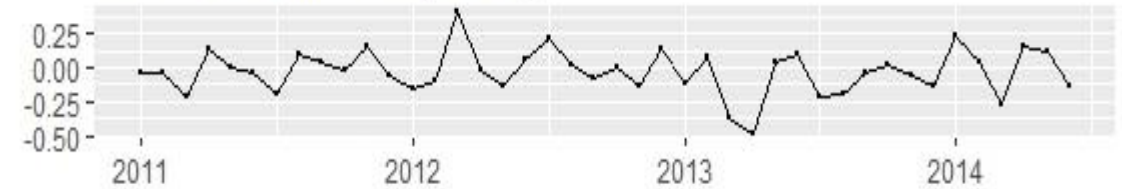
EU_Consumer_Quantity

Ljung-Box test data: p-value = $7.857e-09$

Residuals from ETS(M,A,M)



Residuals from ETS(M,A,M)



Holt-Winter's Residuals for APAC

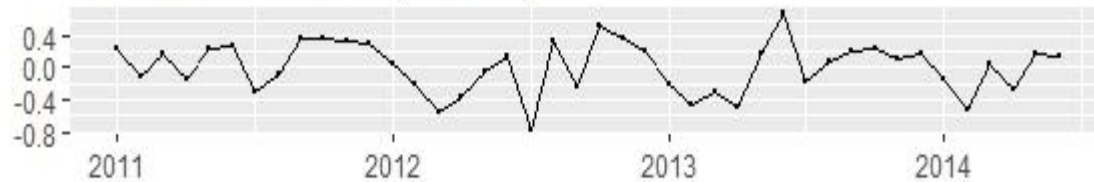
APAC_Consumer_Sales

Ljung-Box test data: p-value = 0.0007644

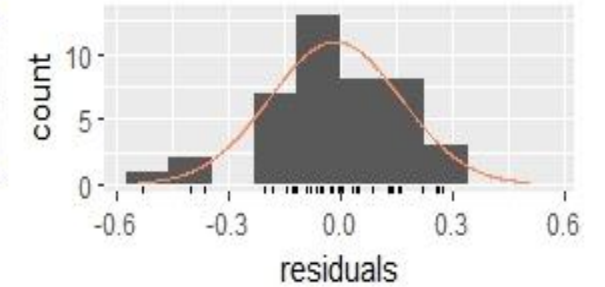
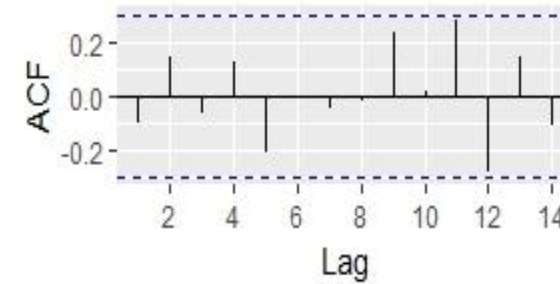
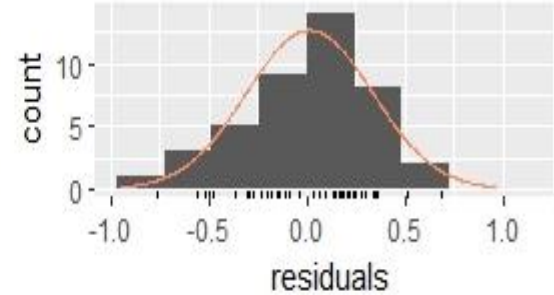
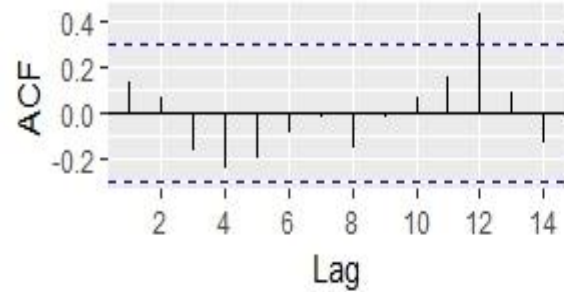
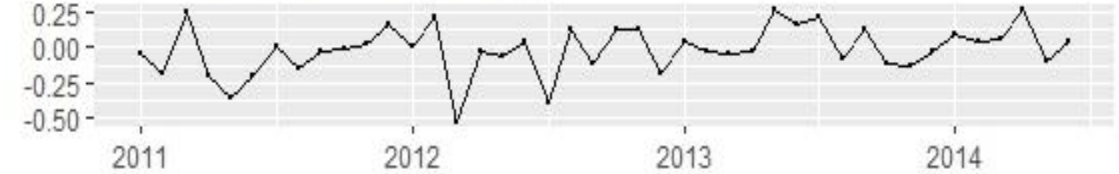
APAC_Consumer_Quantity

Ljung-Box test data: p-value = 3.582e-05

Residuals from ETS(M,Ad,N)



Residuals from ETS(M,A,M)



Model Evaluation & Recommendation - 2 (based on Holts Winters Model)

Customer Segment	MASE	Forecasted values
EU Consumer Sales	0.7484215	34036.33, 25084.59, 36869.57, 31894.55, 39060.28, 57245.09
EU Consumer Qty	1.4213443	410.5582, 344.8788, 458.0014, 390.2239, 405.2258, 633.5119
APAC Consumer Sales	1.6311864	46692.45, 46837.75, 46975.41, 47105.85, 47229.44, 47346.54
APAC Consumer Qty	0.7731437	451.9506, 368.9511, 447.1801, 362.8962, 746.0587, 784.5692

Model Evaluation & Recommendation II (Holt Winters) (Cont.)

- For the EU Consumer segment the store needs to maintain an inventory of nearly 2700 units for next 6 months (Jan – June , 2015) : (forecast is of : 2642).
- From the EU Consumer segment the store can expect total sales of : 2,24,190 in the next 6 months.
- For APAC Consumer segment the store needs to maintain an inventory of nearly 3200 units as the forecast is of 3161 units.
- From APAC Consumer segment the store can expect total sales of : 2,82,183 in next 6 months.