

Retail-Giant Sales Forecasting Case Study - Time Series

PRESENTATION

Group Members Name:

1. Thanigachalam TamizhSelvam
2. Swati Bansal
3. Rajanish Mirajkar

Problem Statement

“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

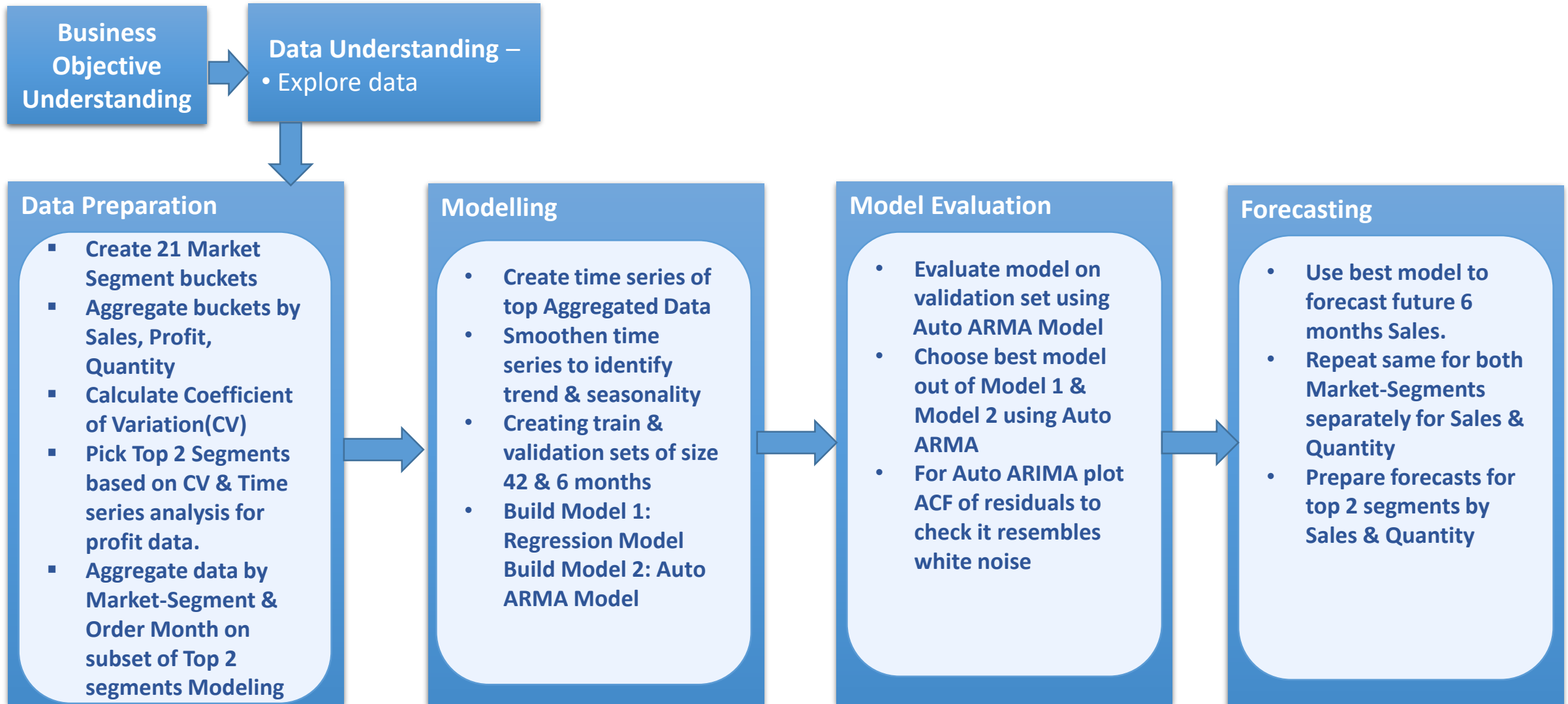
Now as a sales/operations manager, need to finalize the plan for the next 6 months. So, want to forecast the sales and the demand for the next 6 months, that would help to manage the revenue and inventory accordingly.

Business Objective

The “Global Mart” caters to 7 different market segments and in 3 major categories. We want to forecast at this granular level, into 21 ($7*3$) buckets.

Important from the store’s point of view is need to find out 2 most profitable (and consistent) segment from these 21 and forecast the sales and demand for these segments.

Identifying 2 most profitable (and consistent) segment from these 21 and forecast the sales and demand for these segments

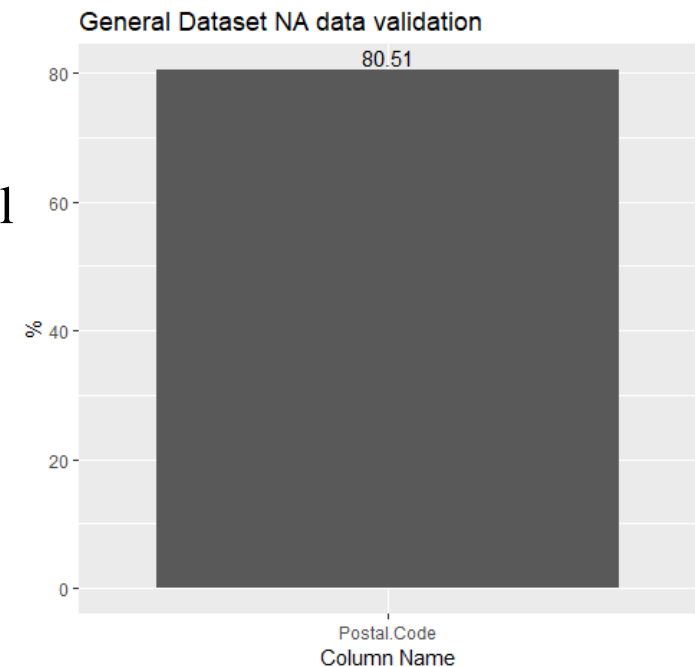


Data Clean Up Activity

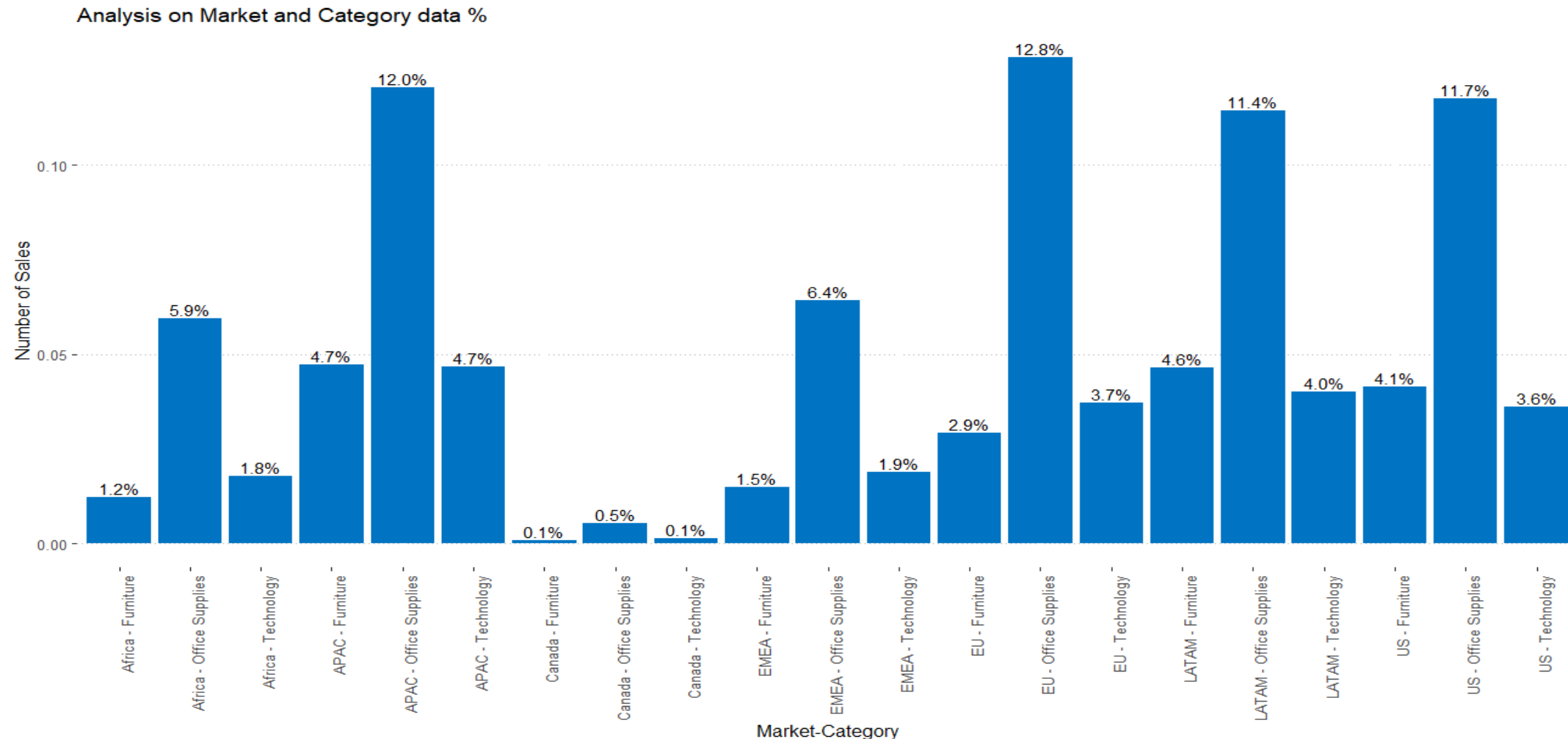
Performed the following data clean up :-

Source File : 'Global Superstore.csv'

1. Converted categorical columns has a vector type
2. Validated for NA values, and found 80 percentage of data is NA for Postal Code column, not considered this for analysis.
3. Formated date column to ensure all the column date values in same formate (Order.Date)
4. created derived columns (Market.Segment and Market.Category)

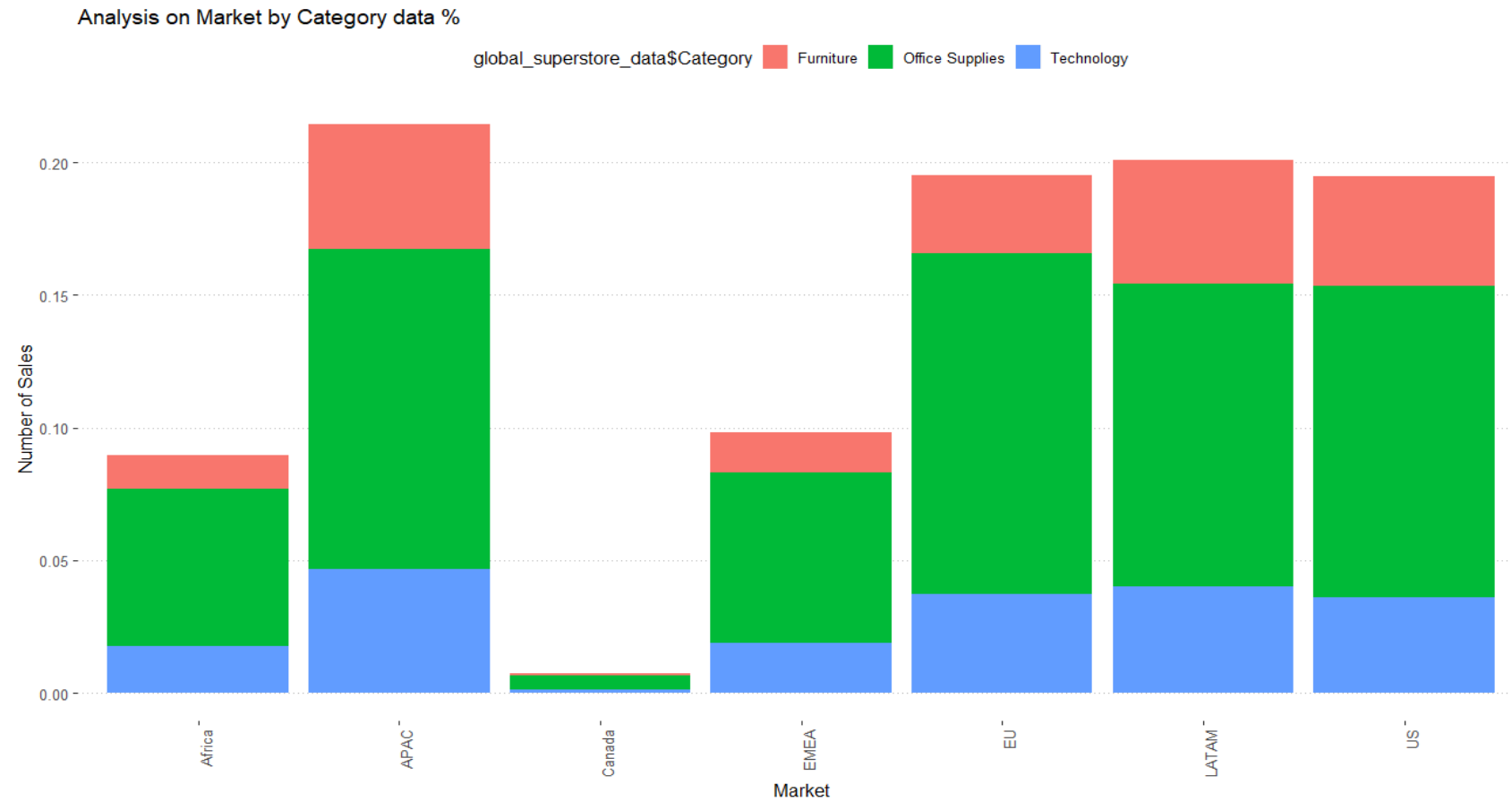


The below graph shows office supply is the major category across all market place



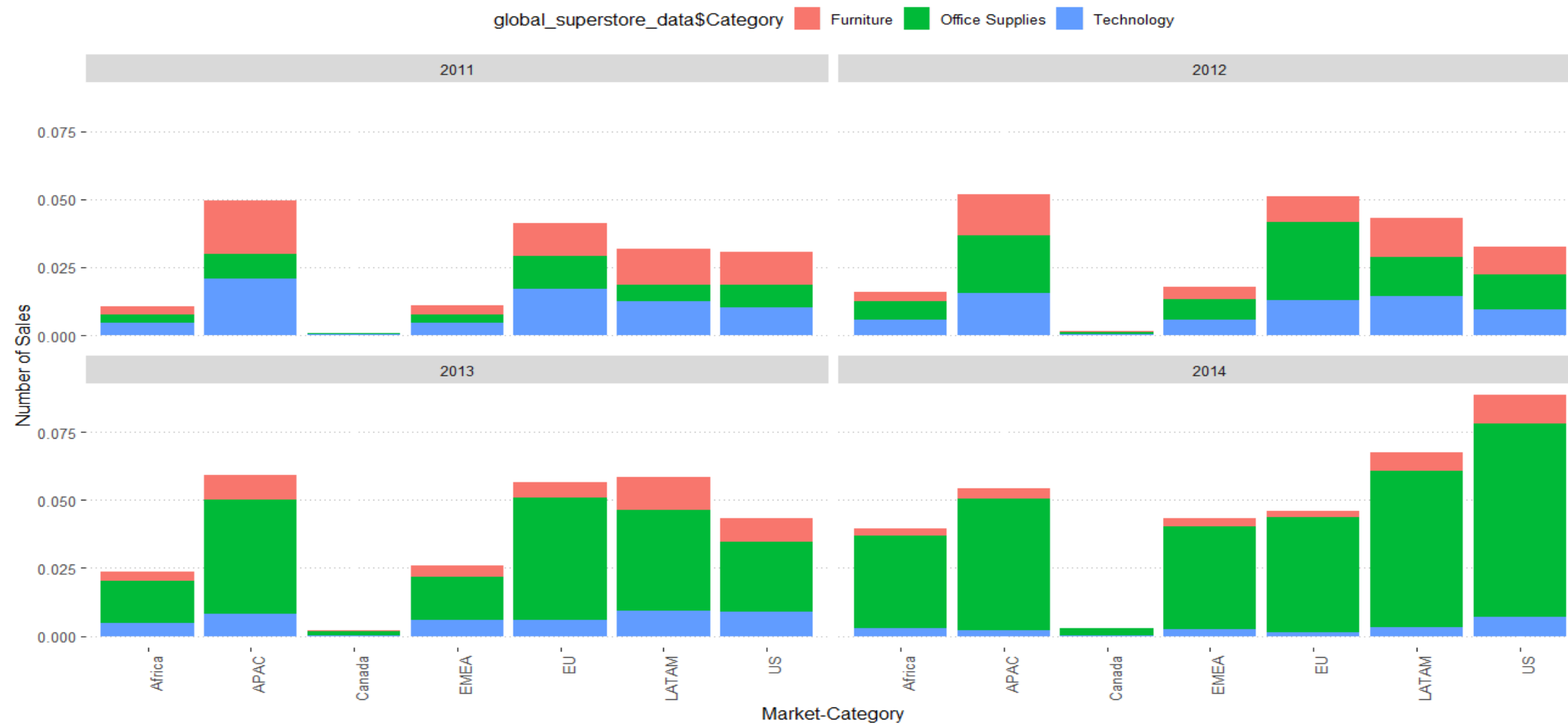
Analysis of All Market and all Category Data

With stack chart
it is observed that
office supply is
major category
across all market
place



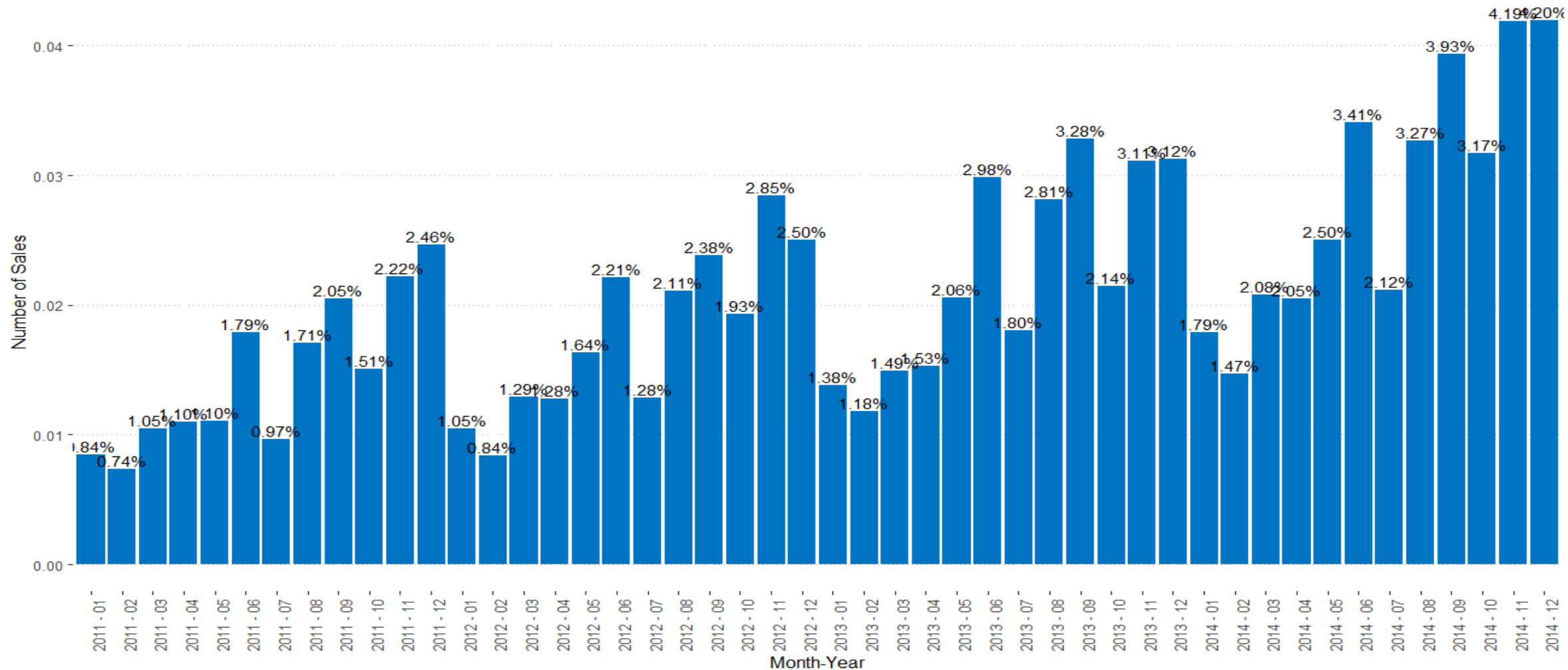
Analysis of All Market and all Category Data by Year

Analysis on Market and Category data by year %

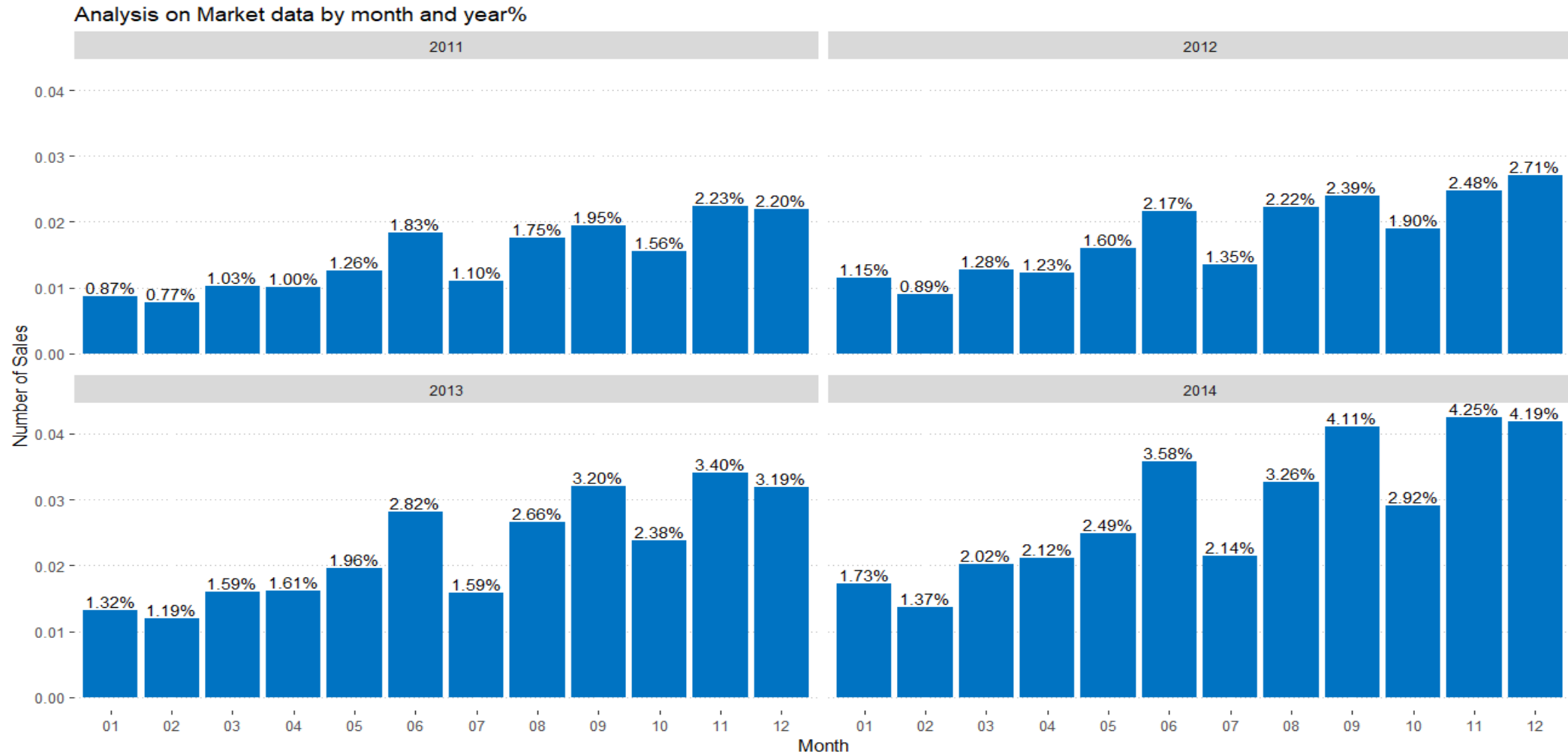


Analysis of Market Data by Month and Year

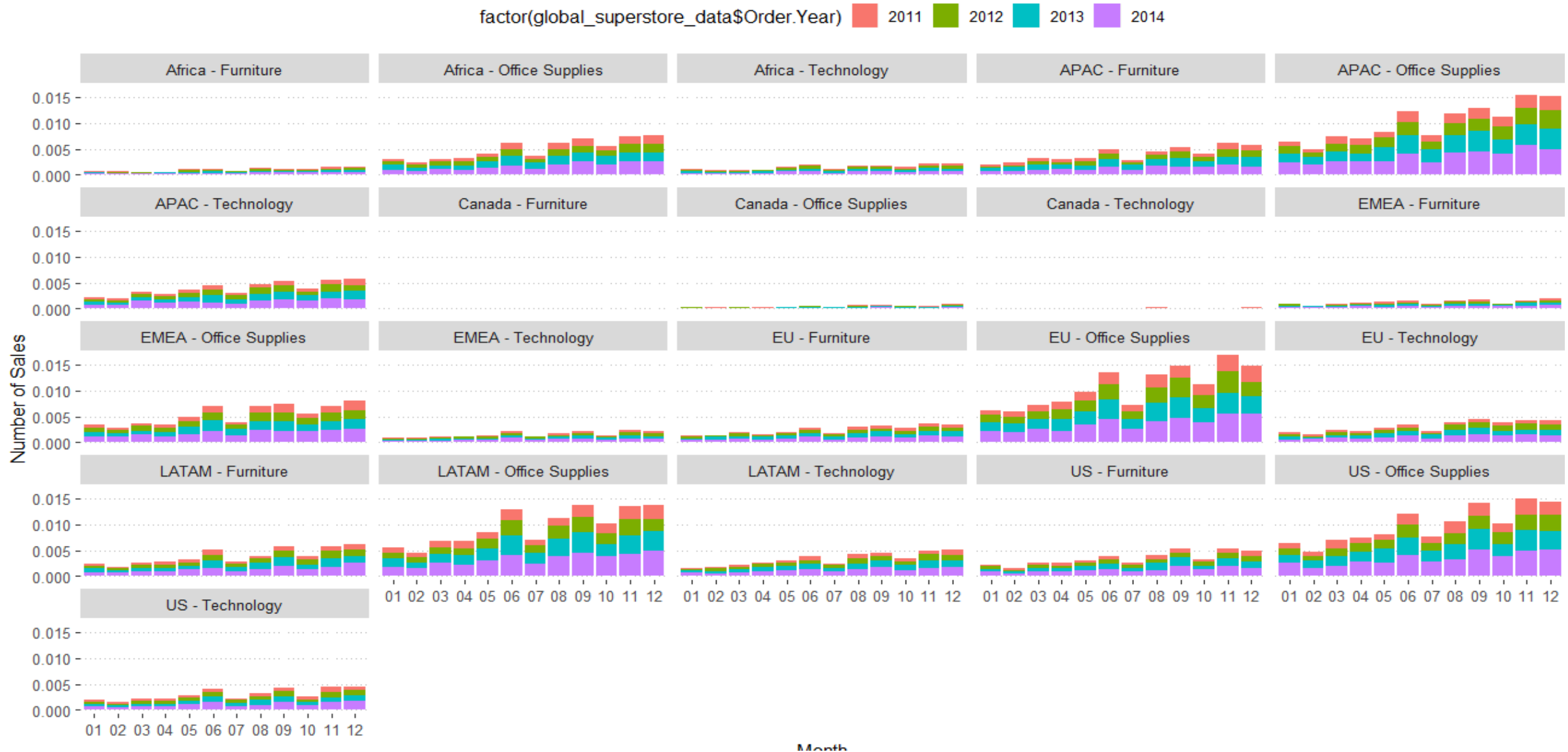
Analysis on Market data by month and year%



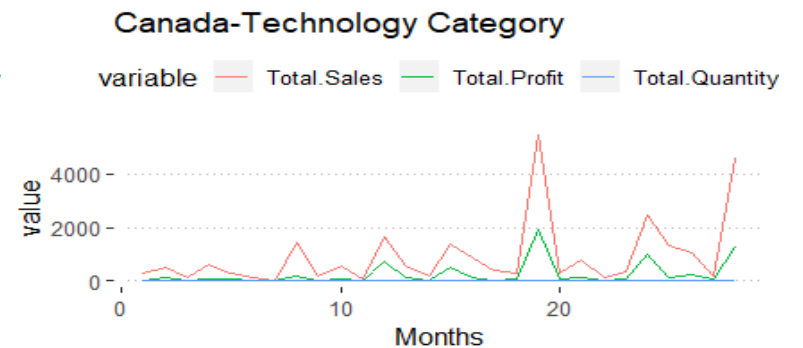
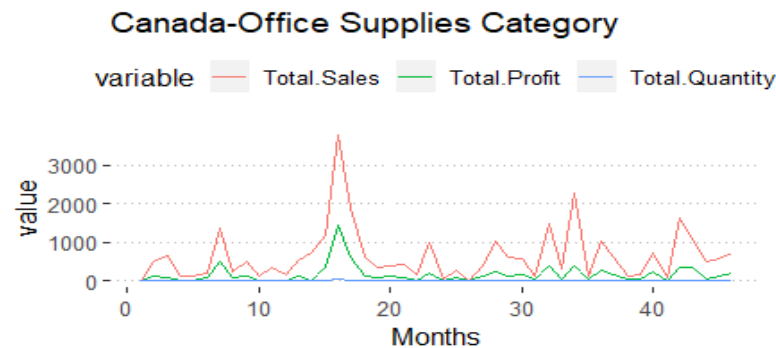
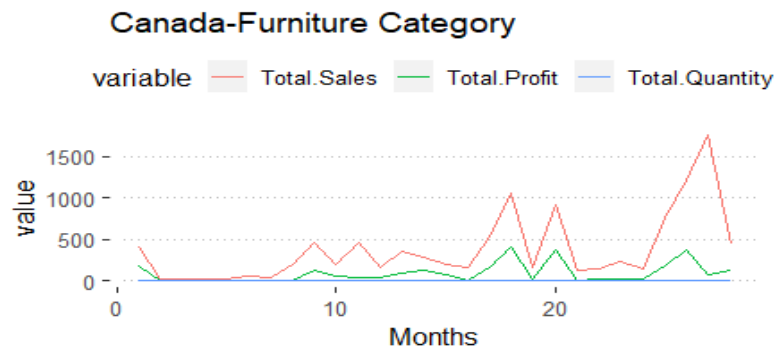
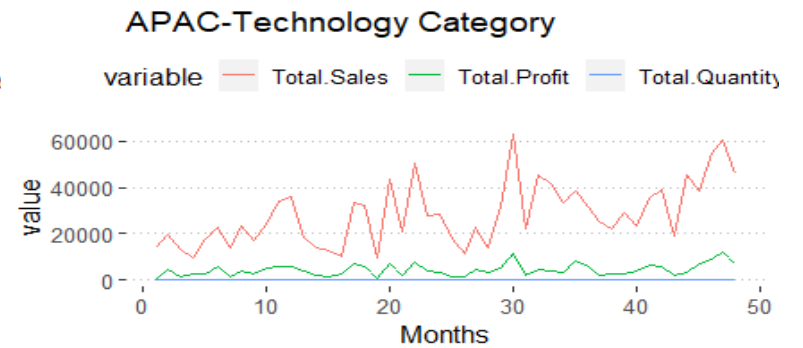
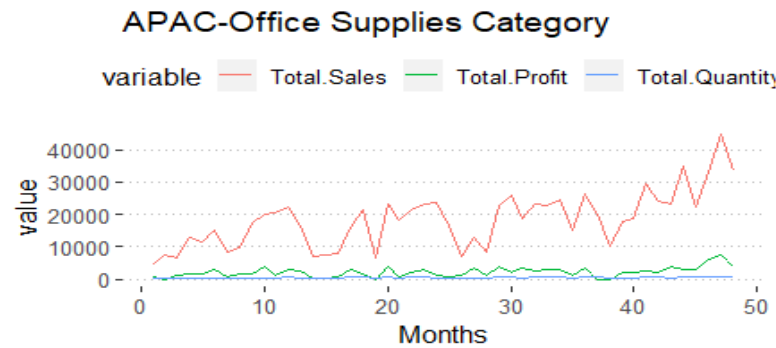
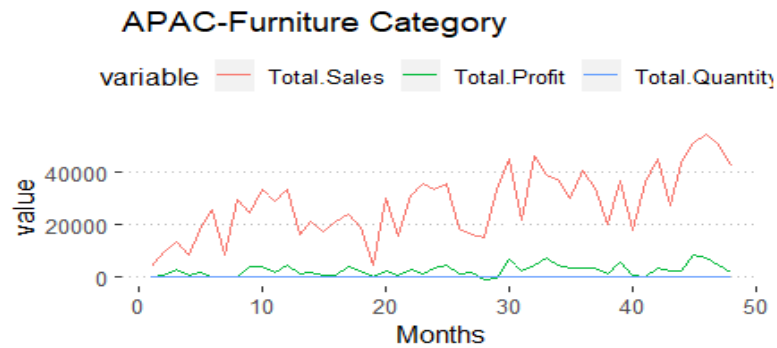
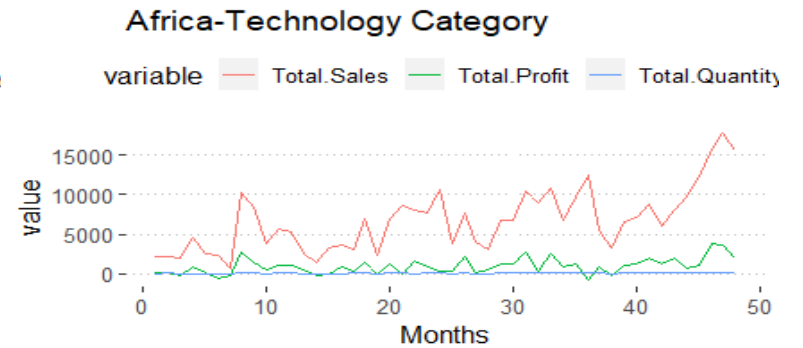
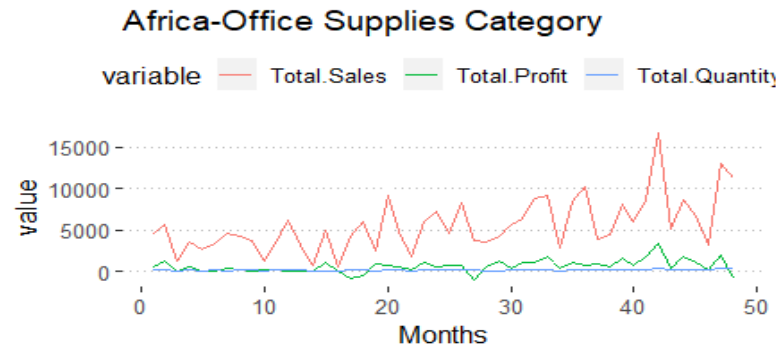
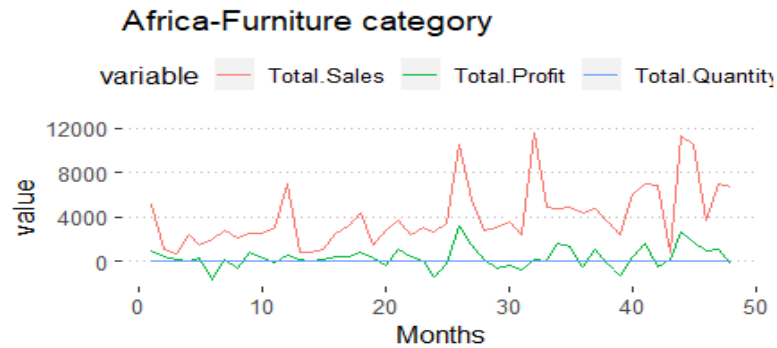
Analysis of Market Data by Month and Year (Continued)



Analysis on Market data by month and year%

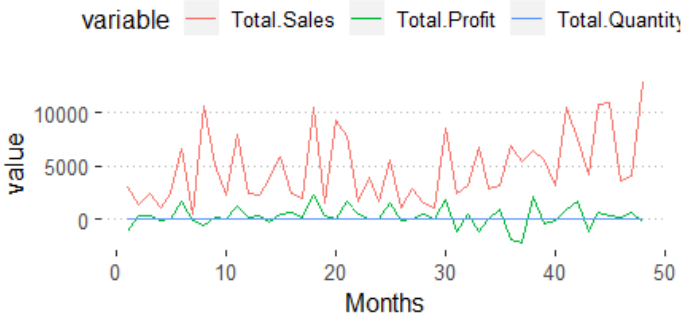


EDA - Africa, APAC and Canada for all category

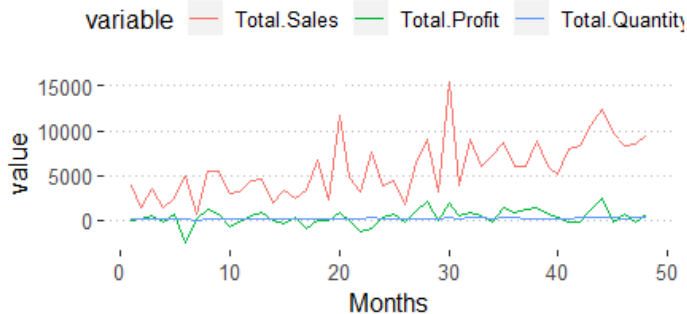


EDA - EMEA, EU and LATAM for all category

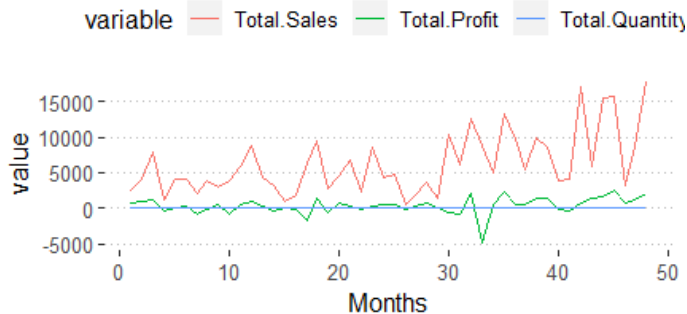
EMEA-Furniture Category



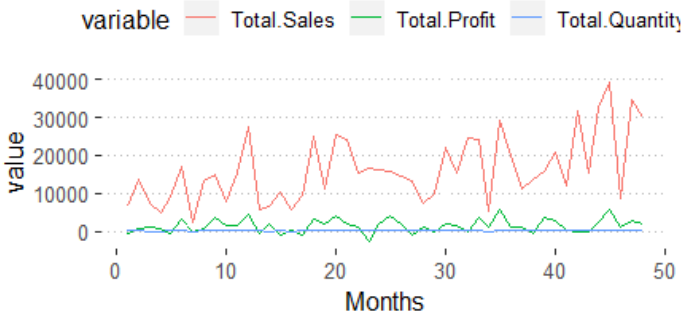
EMEA-Office Supplies Category



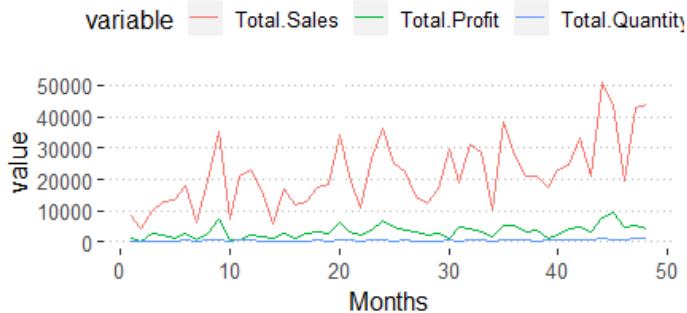
EMEA-Technology Category



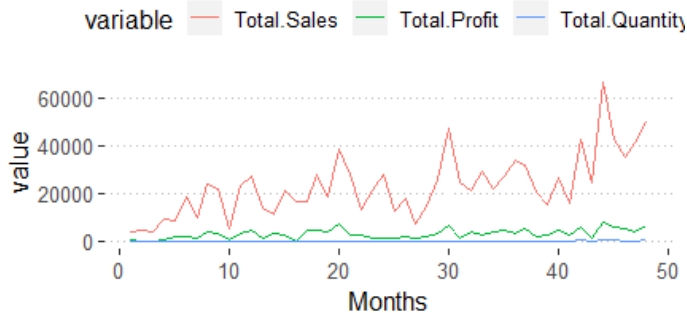
EU-Furniture Category



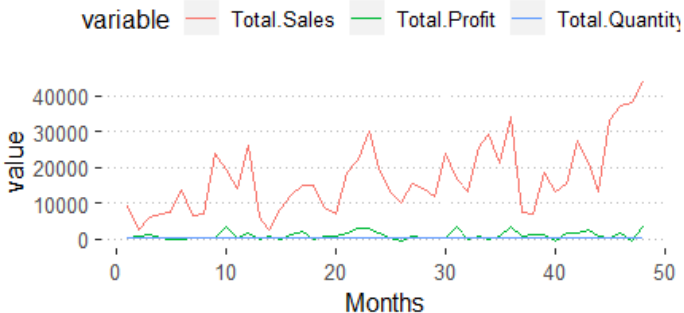
EU-Office Supplies Category



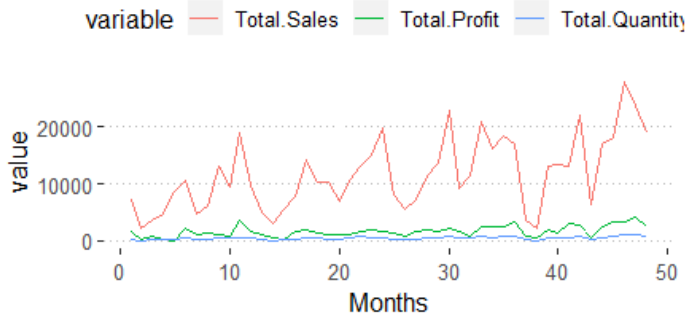
EU-Technology Category



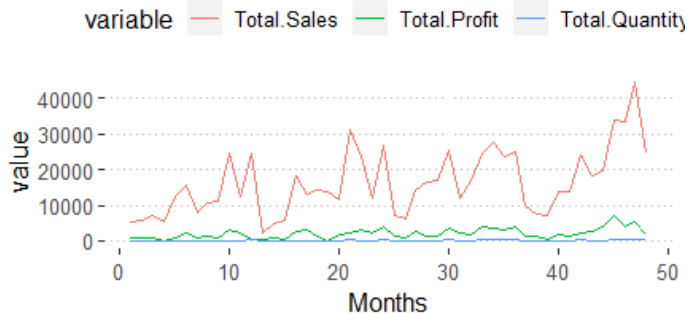
LATAM-Furniture Category



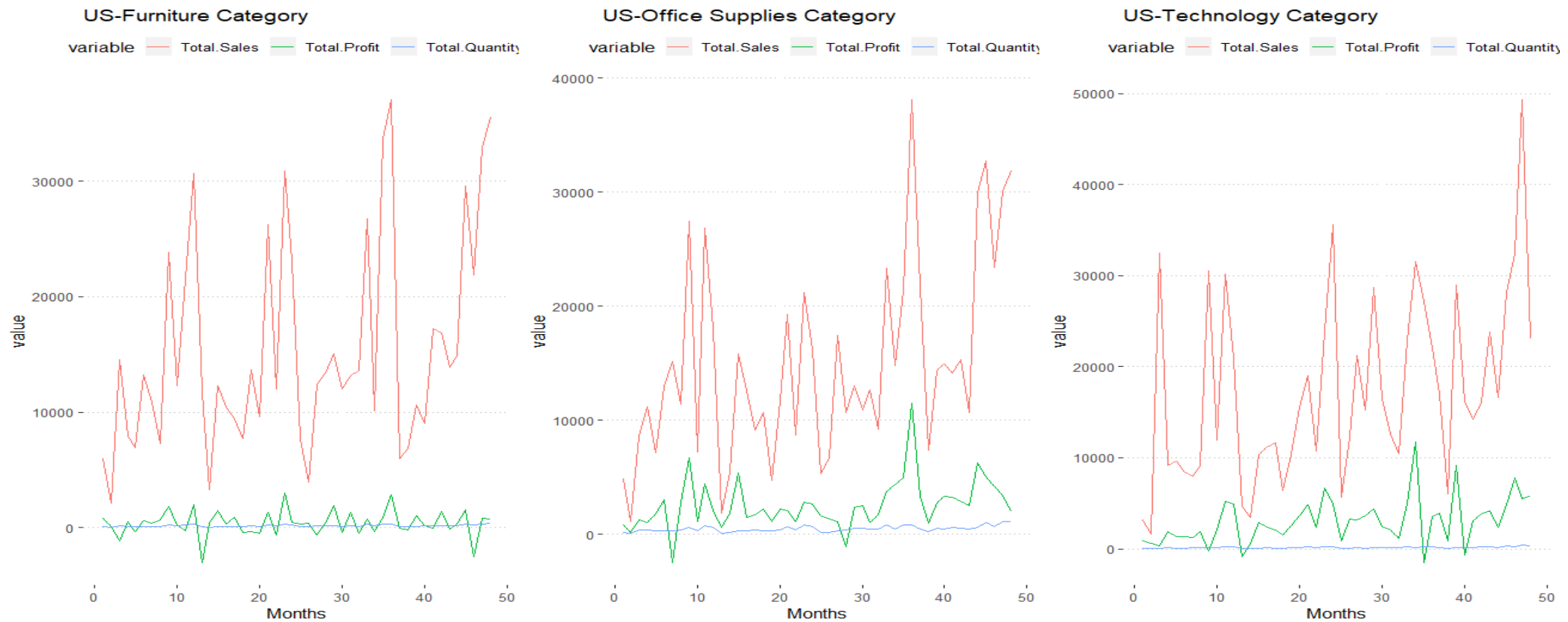
LATAM-Office Supplies Category



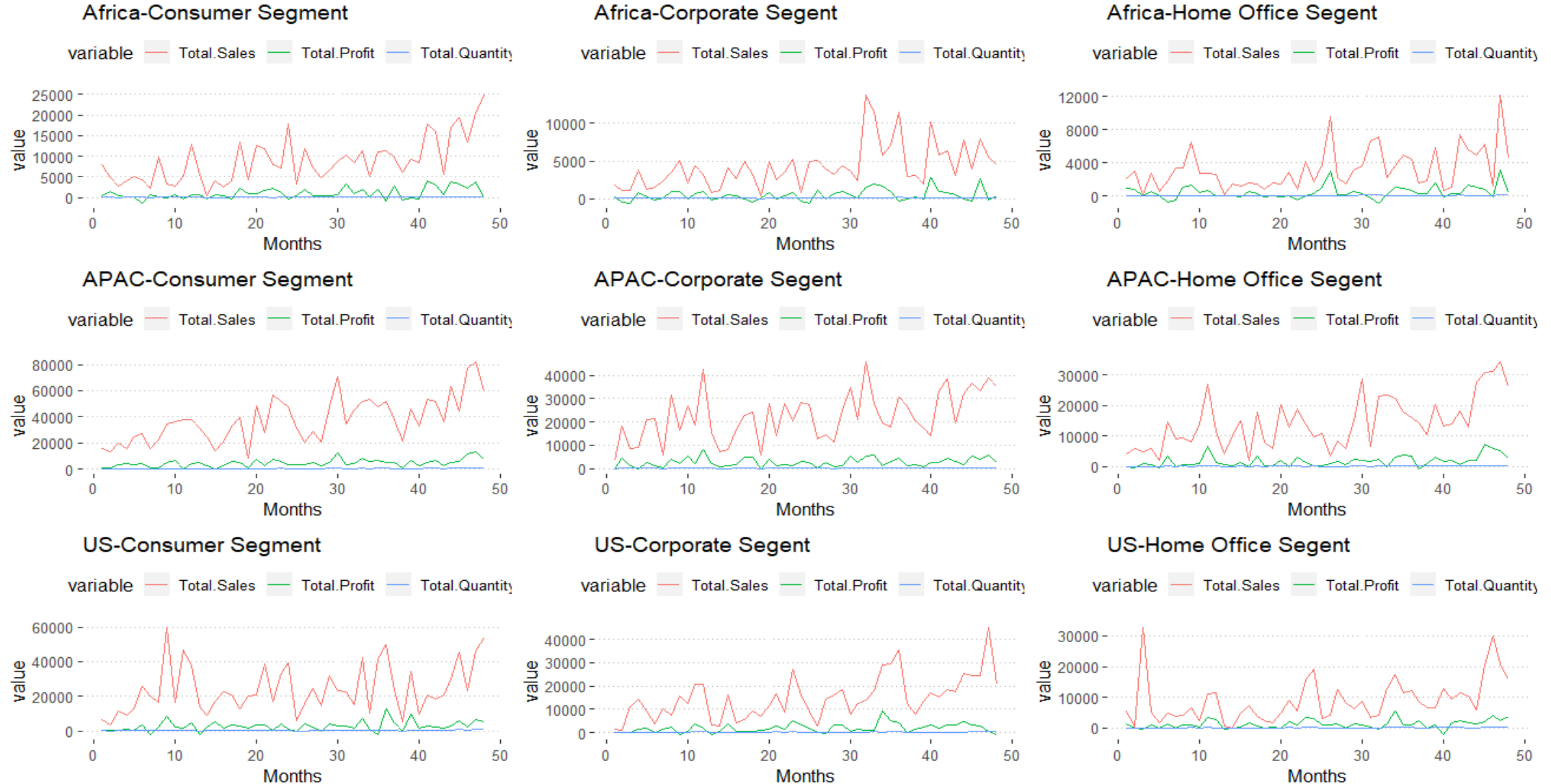
LATAM-Technology Category



EDA - US for all category

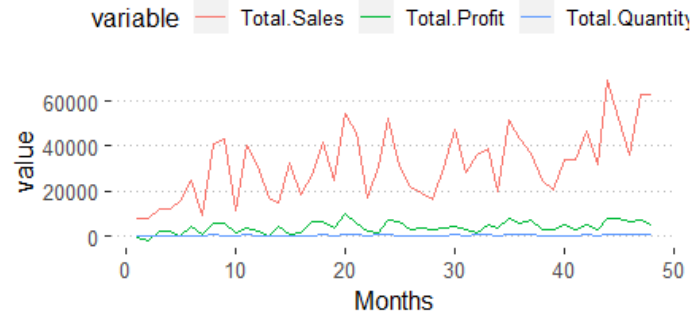


EDA - Africa, APAC and US for all Segment

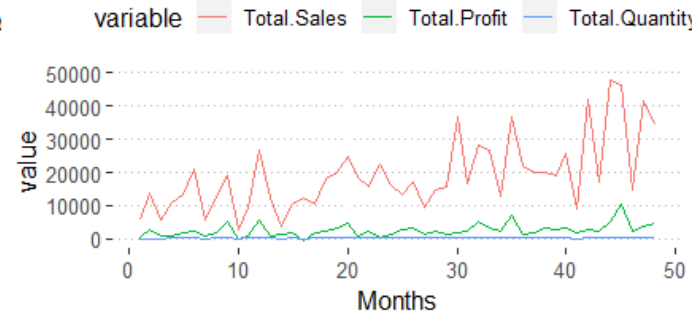


EDA - EU, EMEA and LATAM for all Segment

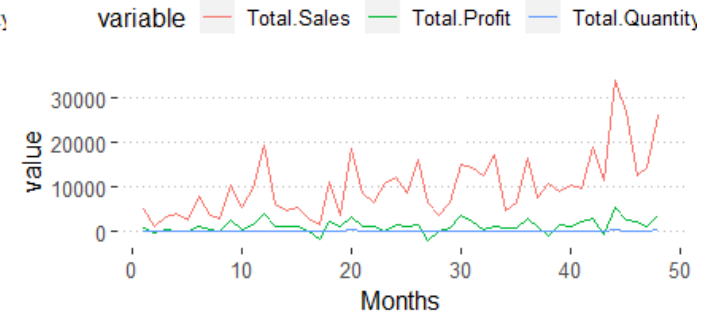
EU-Consumer Segment



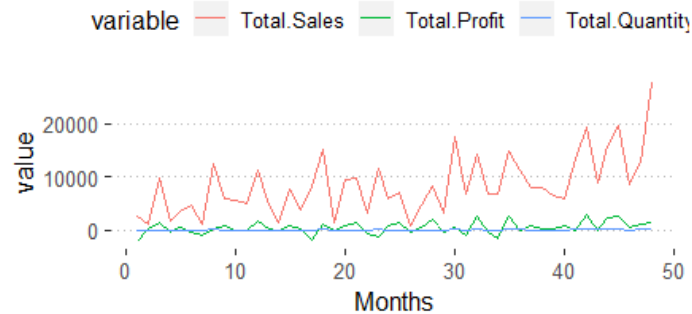
EU-Corporate Segent



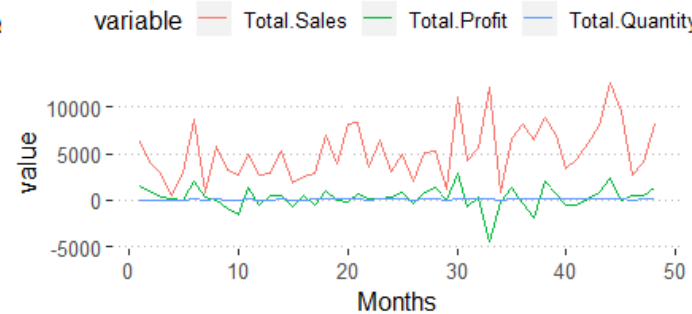
EU-Corporate Segent



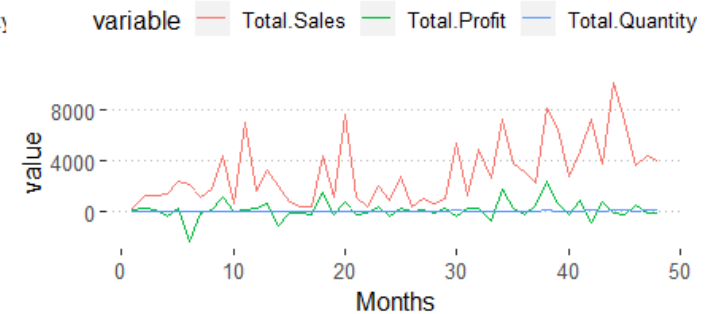
EMEA-Consumer Segent



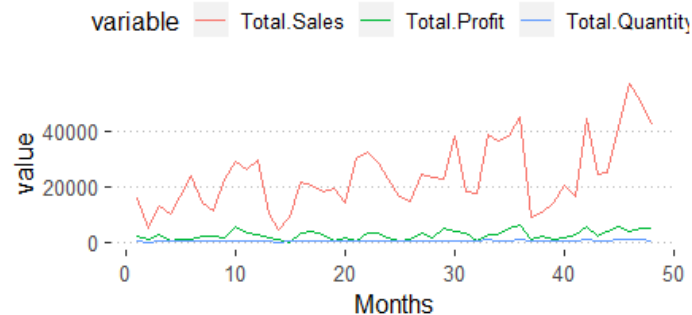
EMEA-Corporate Segent



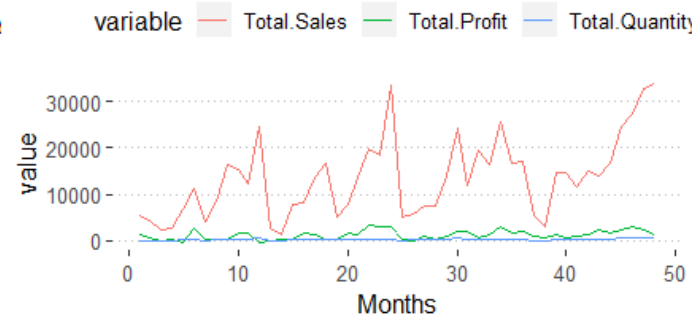
EMEA-Home Office Segent



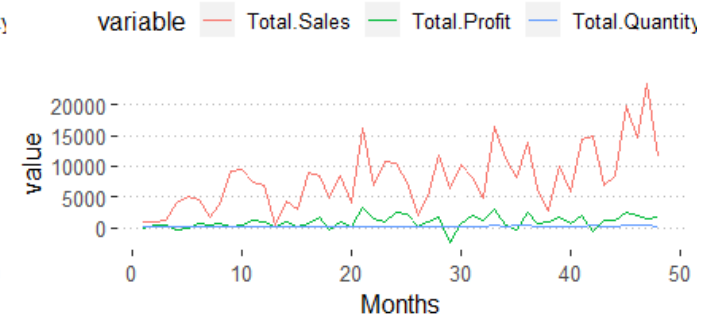
LATAM-Consumer Segent



LATAM-Corporate Segent

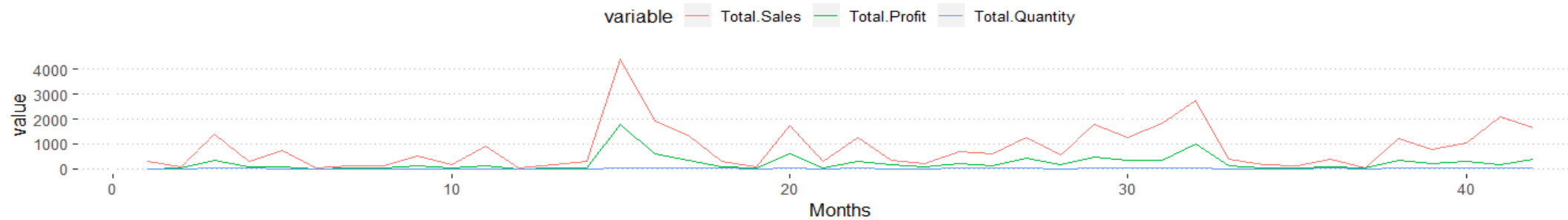


LATAM-Home Office Segent

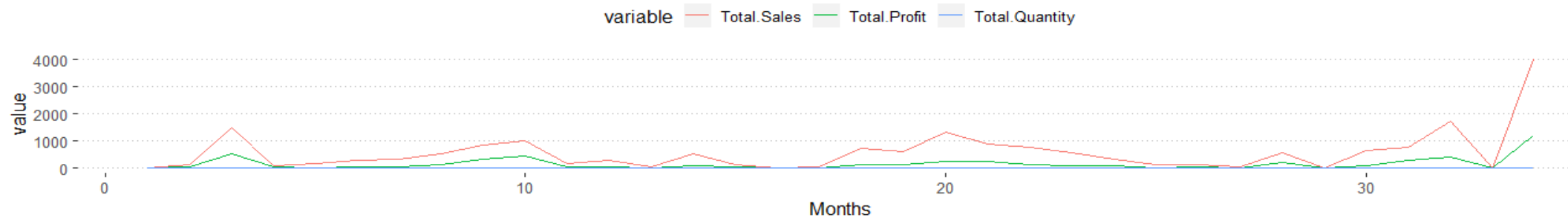


EDA - Canada for all Segment

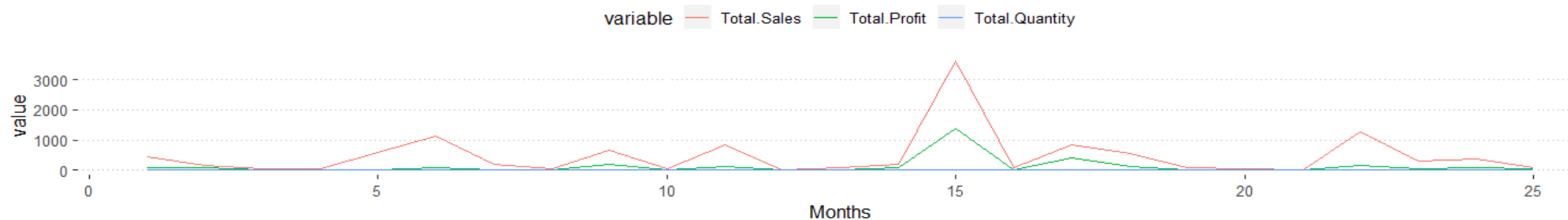
Canada-Consumer Segent



Canada-Corporate Segent

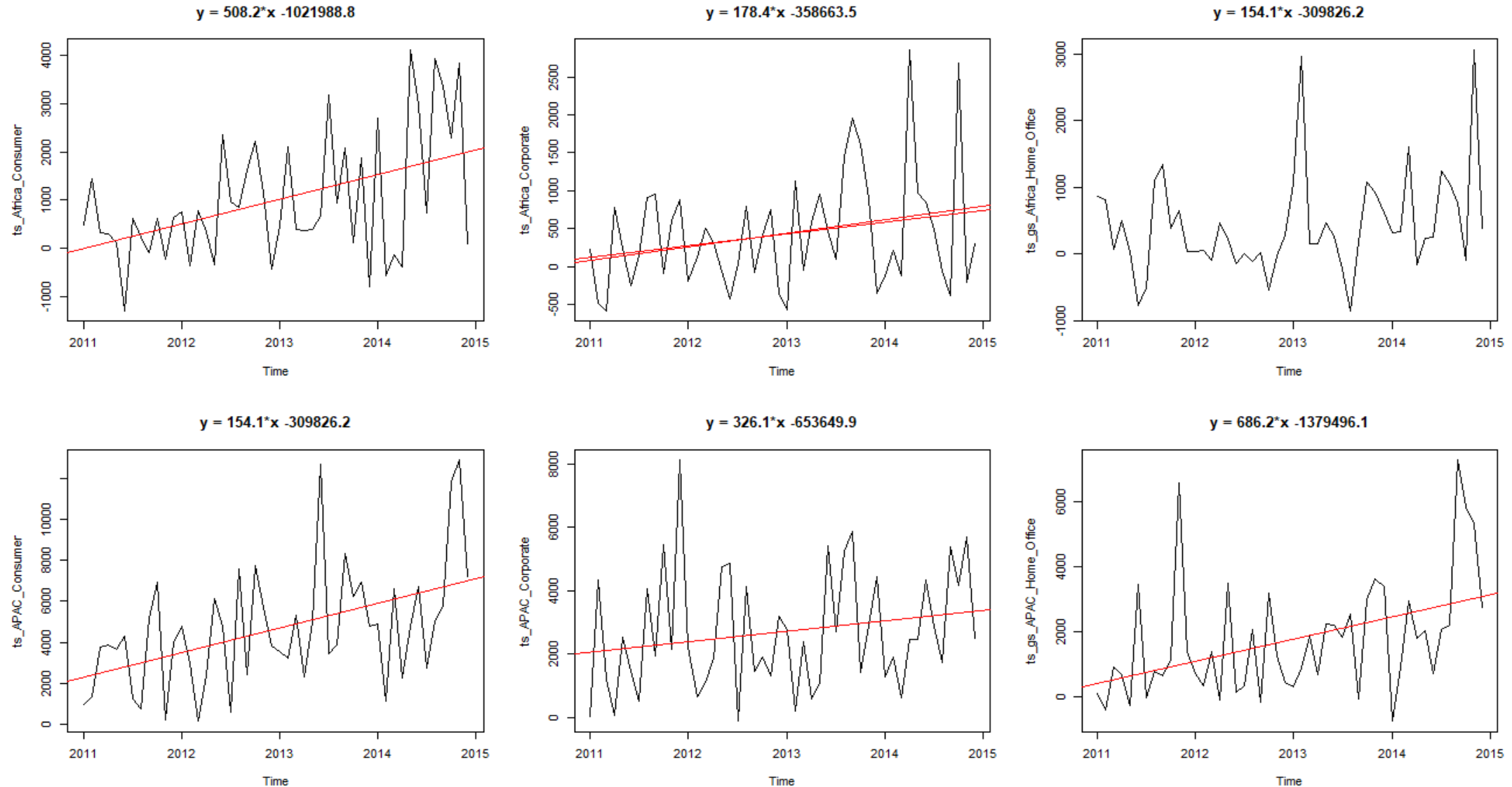


Canada-Home Office Segent

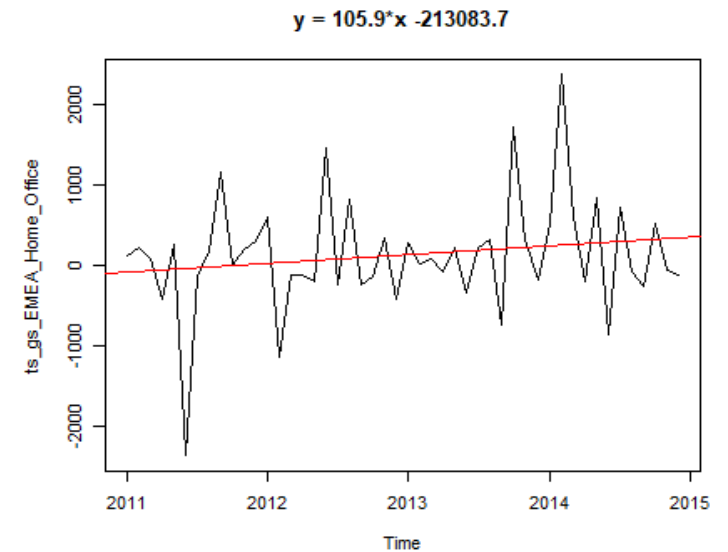
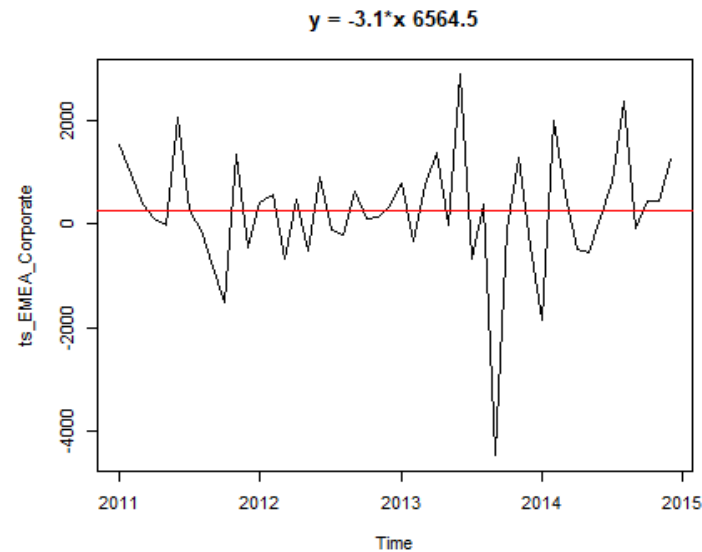
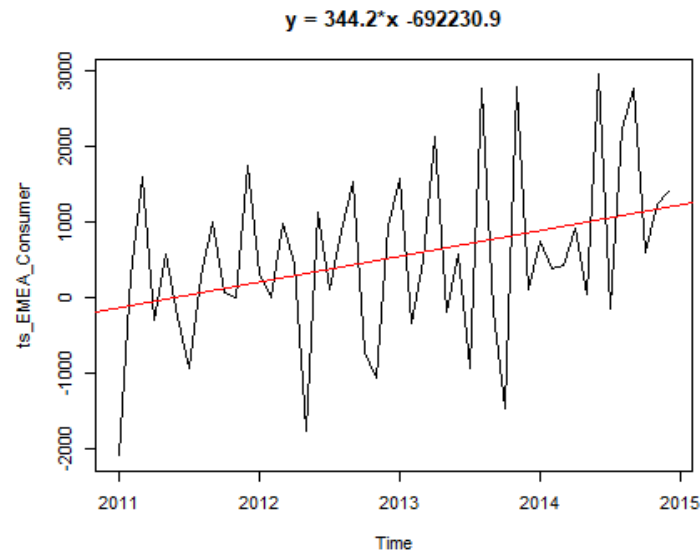
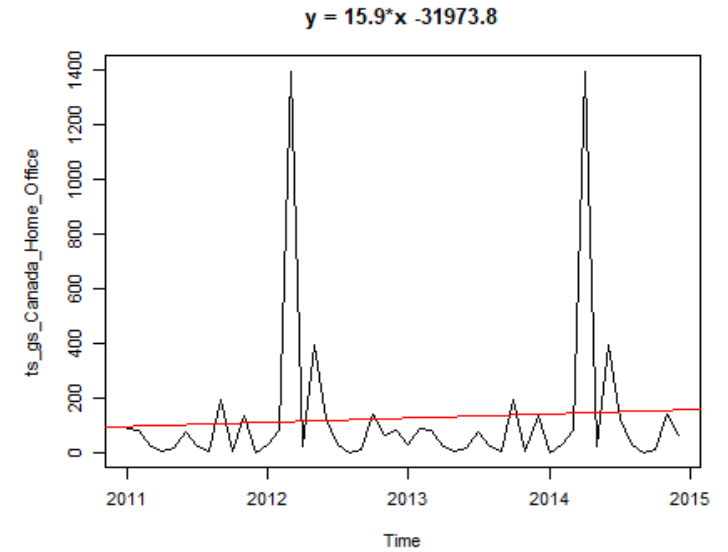
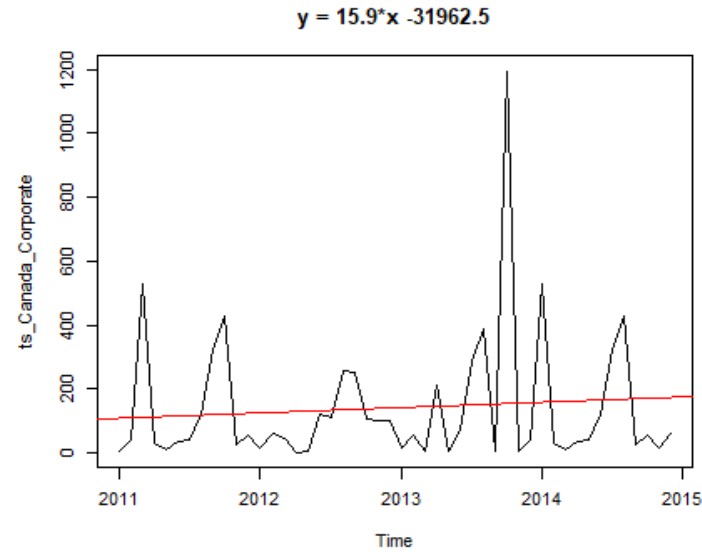
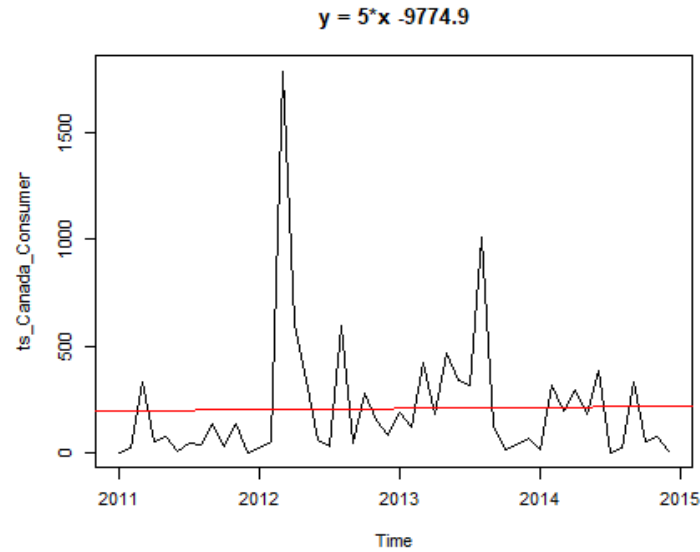


You would need to first segment the whole dataset into the 21 subsets based on the market and the customer segment level. Next, comes the most important data preparation step. That is to convert the transaction-level data into a time series. Thus, you would need to aggregate the 3 attributes - Sales, Quantity & Profit, over the Order Date to arrive at monthly values for these attributes. Once, you arrive at these 3 time series for each of the 21 segments, we need to find the 2 most profitable and consistently profitable segments. For this, the metric that you can use is the coefficient of variation of the Profit for all 21 market segments. If you wish to know more about the coefficient of variation

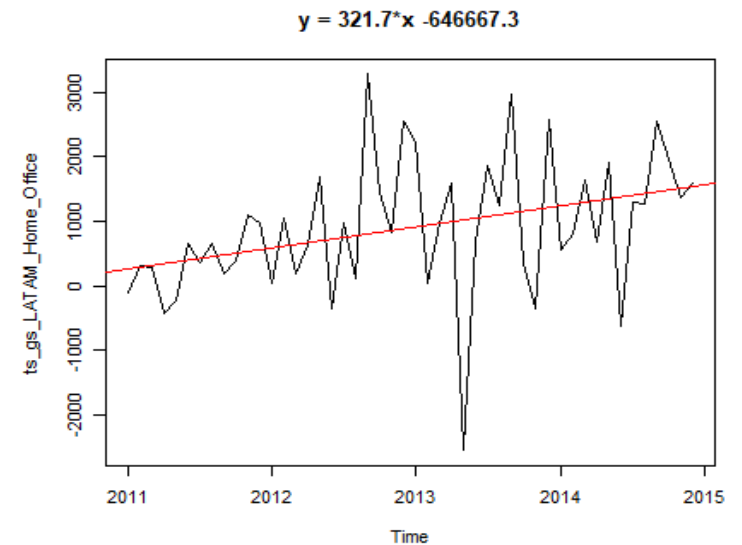
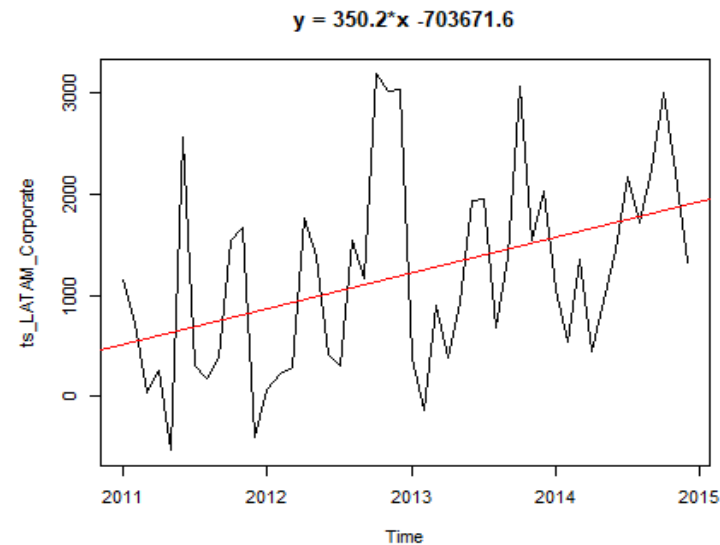
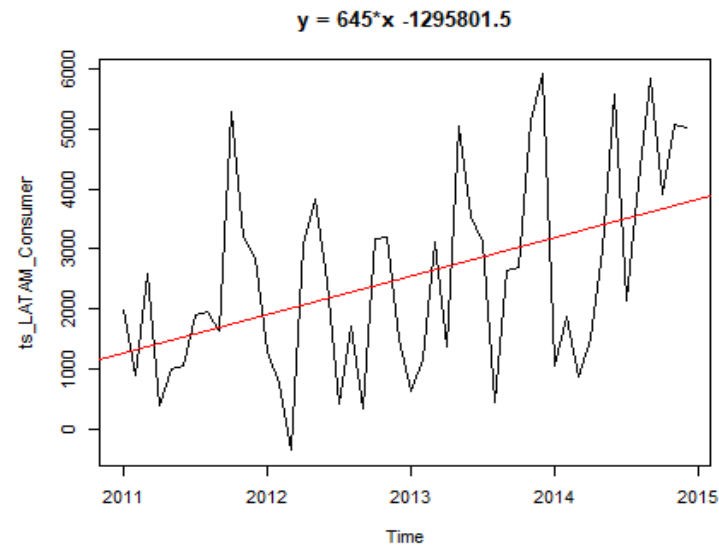
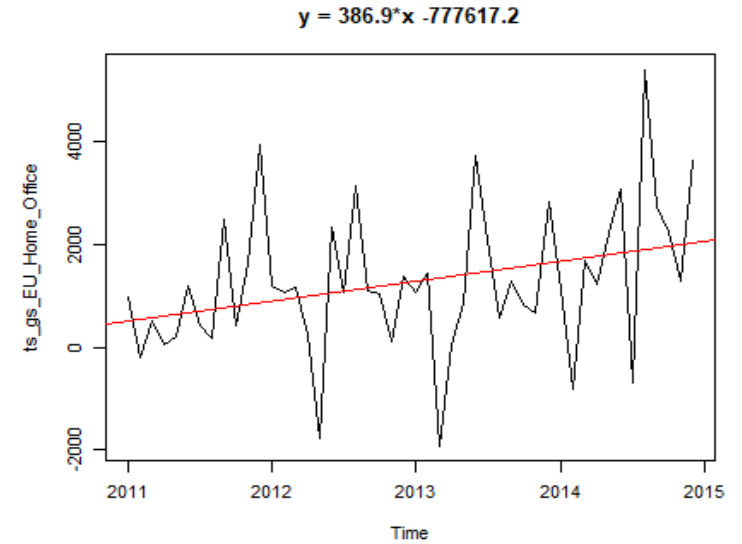
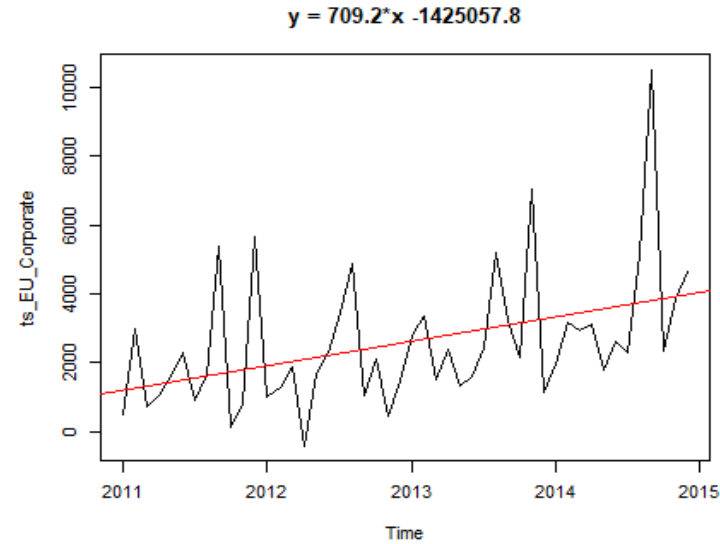
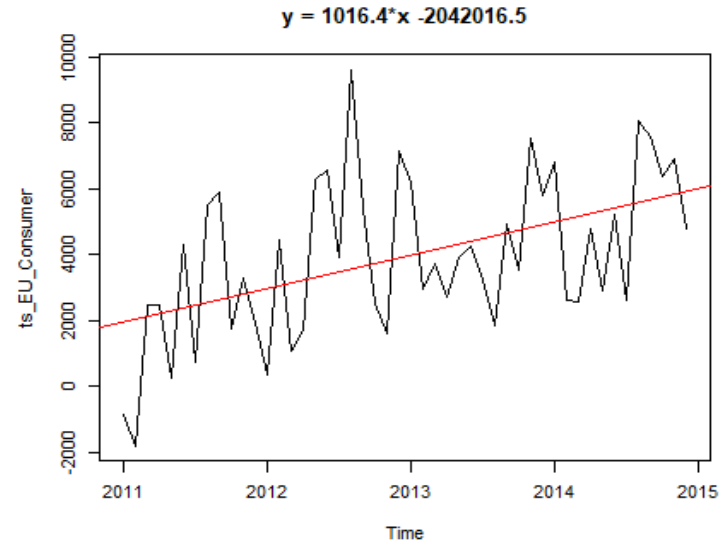
Time Series Plot for Africa and APAC consumer



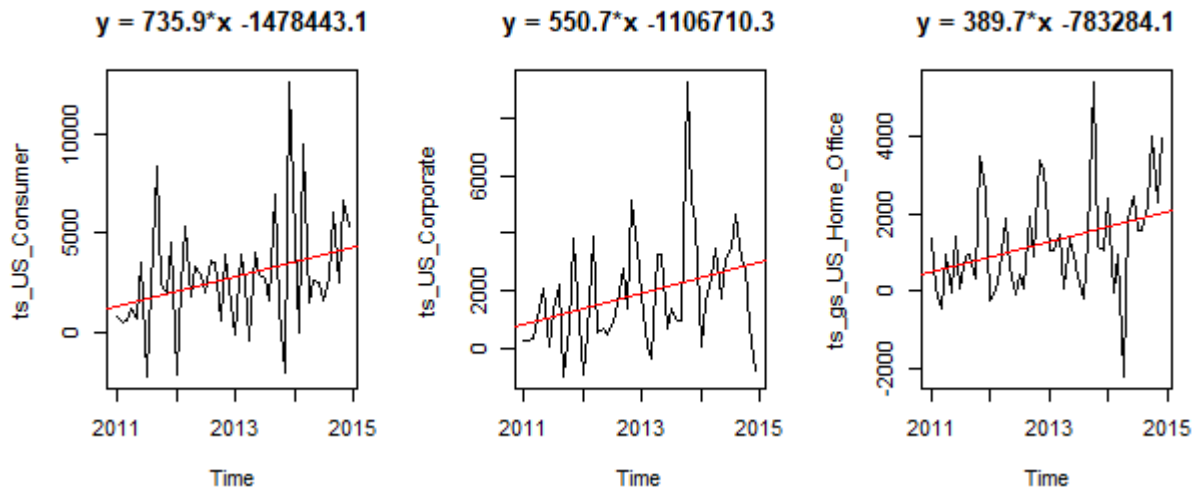
Time Series Plot for Canada and EMEA consumer



Time Series Plot for EU and LATAM consumer



Time Series Plot for US consumer



Below table shows the COV on Profit for All Market segment

Market.Segment	COV
EU-Consumer	0.624305
APAC-Consumer	0.632132
LATAM-Consumer	0.661483
APAC-Corporate	0.698087
EU-Corporate	0.763807
LATAM-Corporate	0.811122
US-Corporate	1.002409
US-Consumer	1.01239
APAC-Home Office	1.045978
US-Home Office	1.096147
EU-Home Office	1.116507
LATAM-Home Office	1.175698
Africa-Consumer	1.319585
Canada-Consumer	1.395312
Canada-Corporate	1.552775
Africa-Corporate	1.776105
Africa-Home Office	1.789996
EMEA-Consumer	2.188271
Canada-Home Office	2.243461
EMEA-Corporate	4.467102
EMEA-Home Office	5.880747

Below table shows Top 5 COV based on EDA and COV will be using
Top 2 i.e EU-Consumer and APAC-Consumer for our Model

Market.Segment	COV
EU-Consumer	0.624305
APAC-Consumer	0.632132
LATAM-Consumer	0.661483
APAC-Corporate	0.698087
EU-Corporate	0.763807

Model building:

Once you arrive at the 2 most profitable segments, the next challenge is to forecast the sales and quantity for the next 6 months. You are supposed to use classical decomposition and auto ARIMA for forecasting. Also, it is advised that you smoothen the data before you perform classical decomposition.

Model building: EU Consumer Sales

EU Consumer Quantity

APAC Consumer Sales

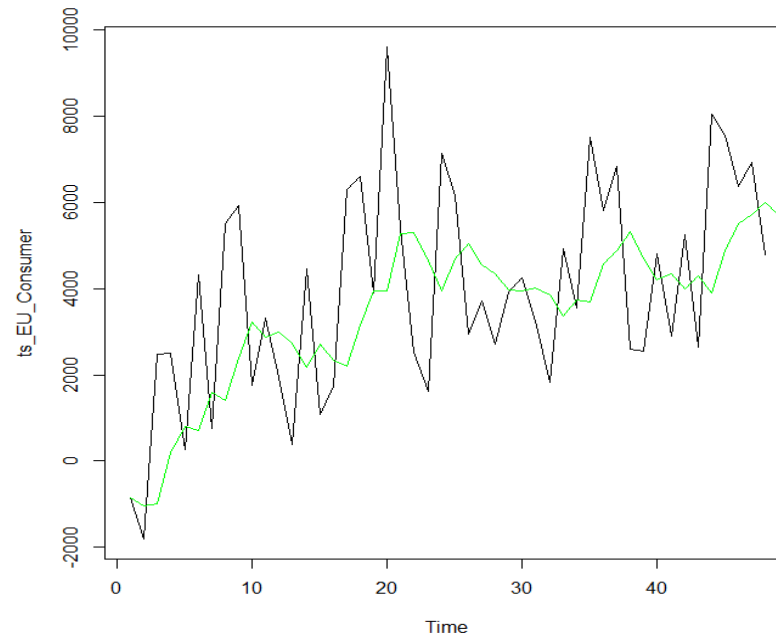
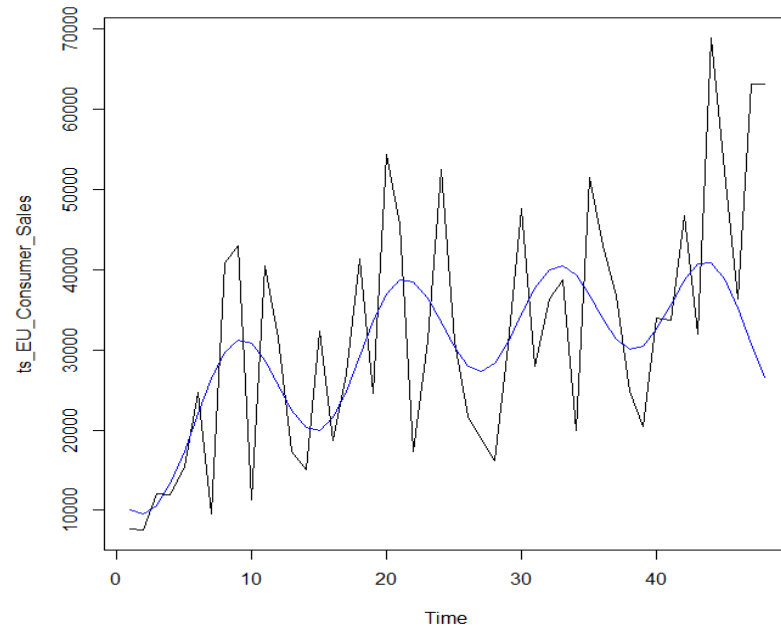
APAC Consumer Quantity

Model Evaluation & Conclusions

Model evaluation: EU Consumer Sales

After Smoothing and building model, Classical Decomposition and ARIMA has achieved MAPE of 34.35 approx

Plotting the predictions along with original values. The left graph represent classical decomposition and right graph shows ARIMA fit for the future 6 months



Next 6 months sales

10070.430

9500.265

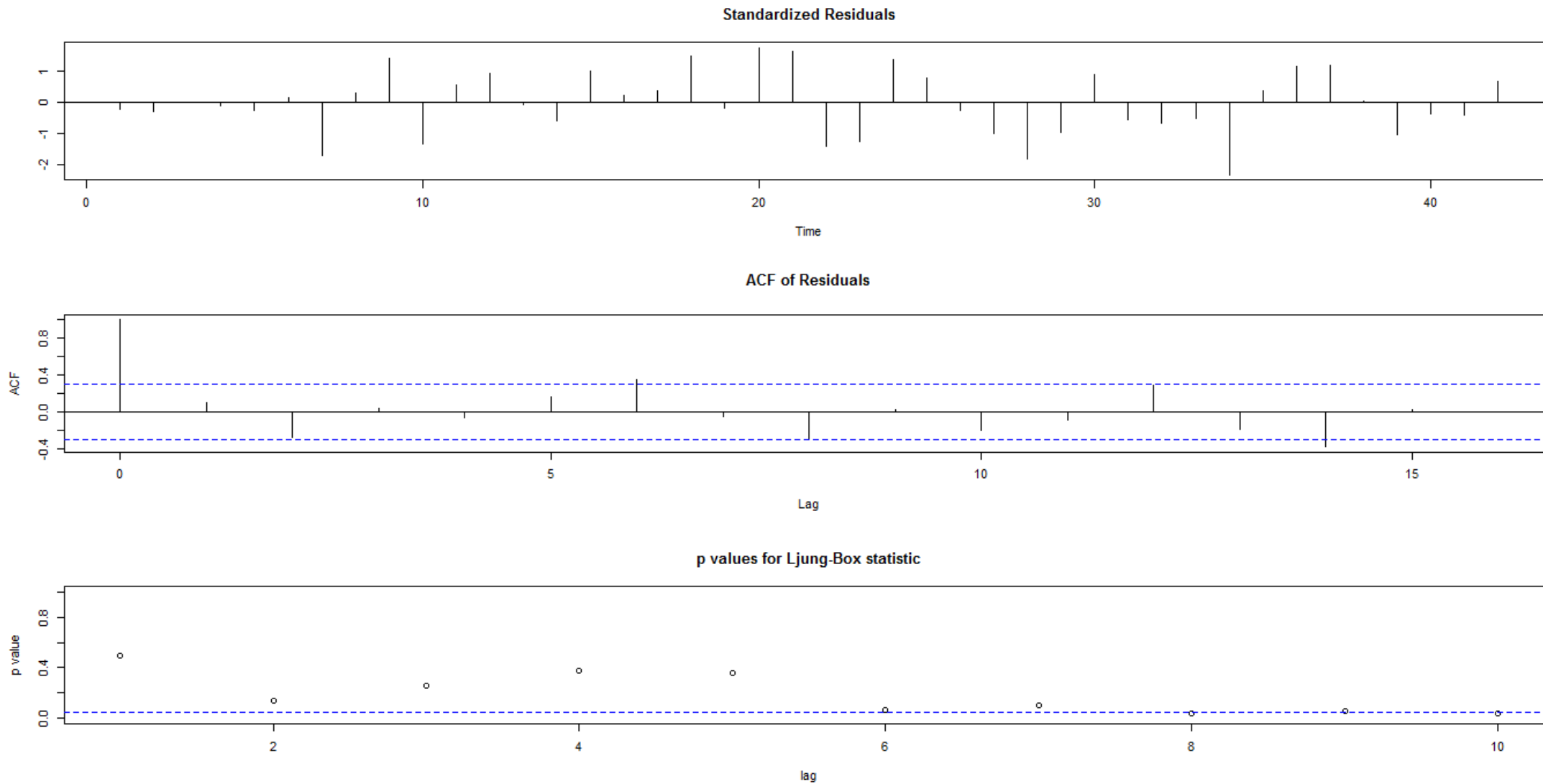
10535.857

13261.886

17334.104

22044.010

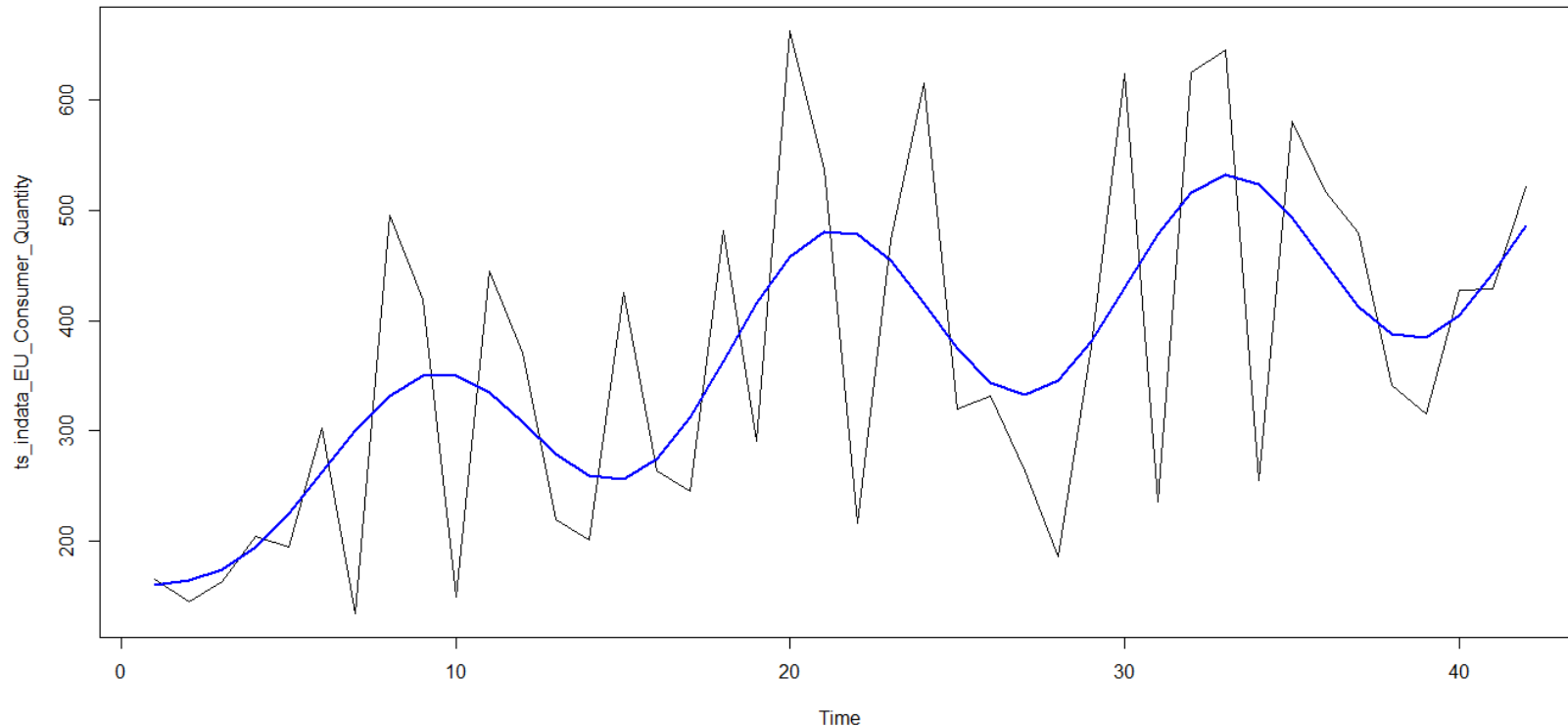
Diagnostic Plot For Time series



Model evaluation: EU Consumer Quantity

After Smoothing and building model, Classical Decomposition and ARIMA has achieved MAPE of 33.24 approx

The below Graph blue line shows the global trend

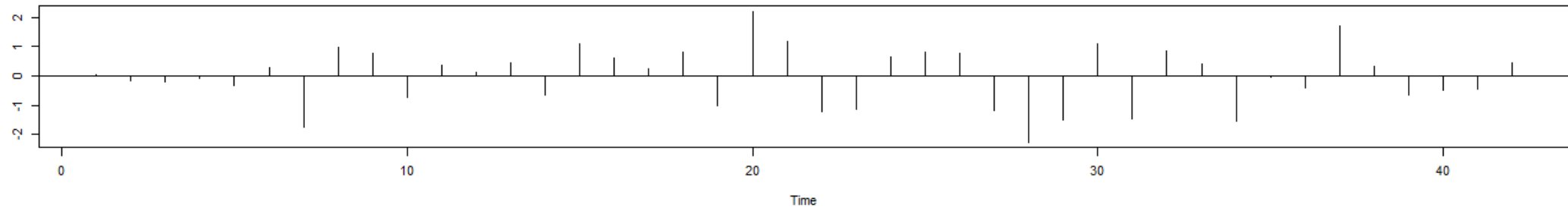


Next 6 months Quantity

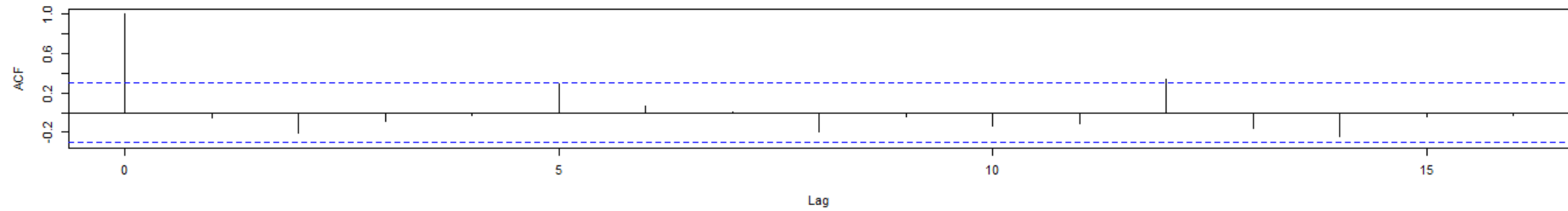
160.1148
 163.8136
 174.0631
 194.1950
 224.5899
 262.0422

Diagnostic Plot For Time series

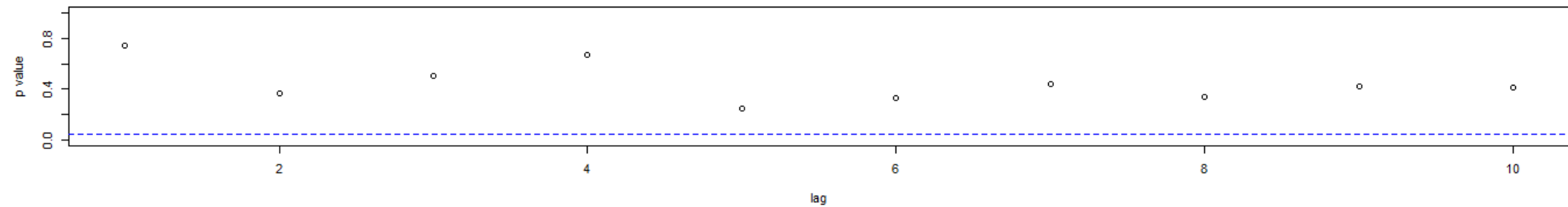
Standardized Residuals



ACF of Residuals



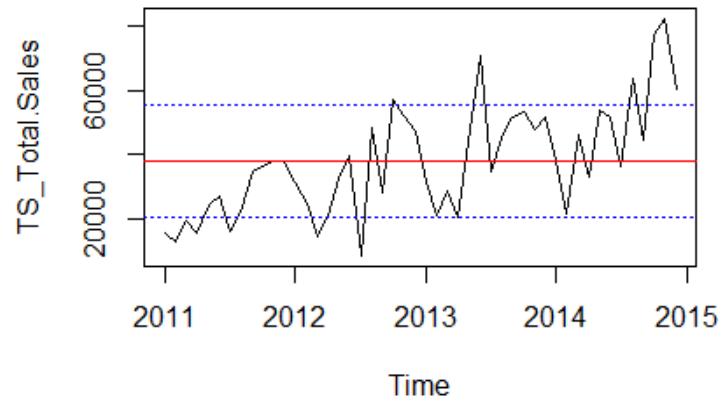
p values for Ljung-Box statistic



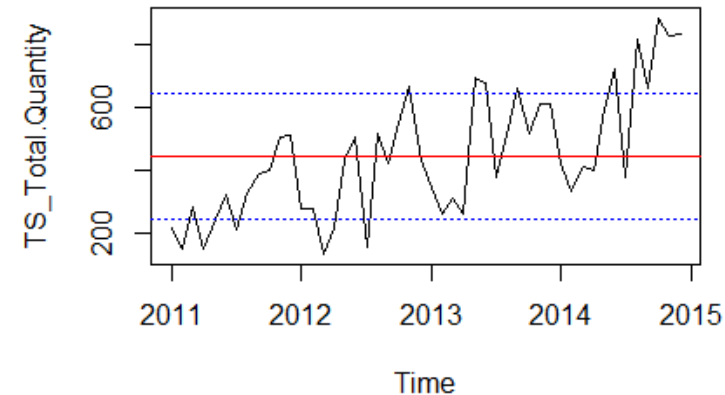
Stationarity:

In this we can see that the plot is stationary since the most of the data in Sales and Quantity for APAC Consumer is between standard deviation and mean values.

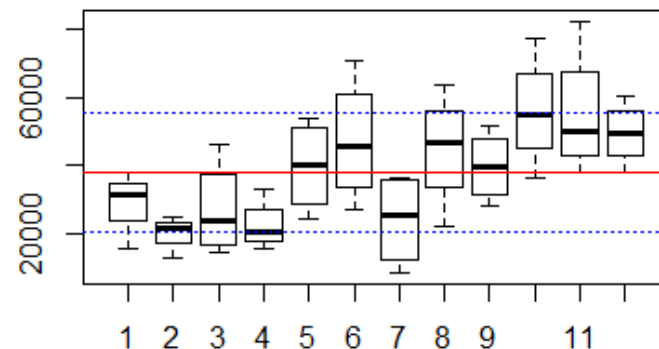
APAC Consumer Total Sales



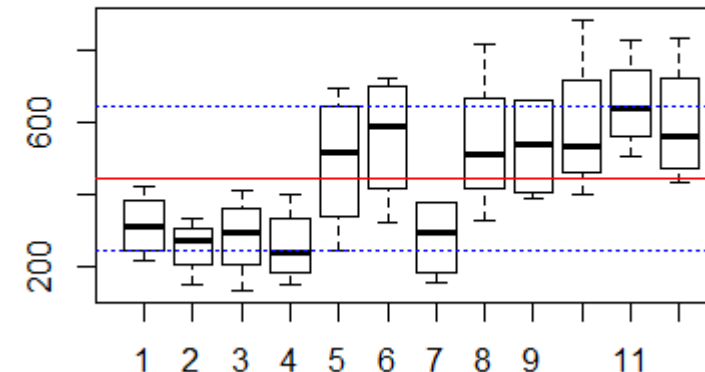
APAC Consumer Total Quantity



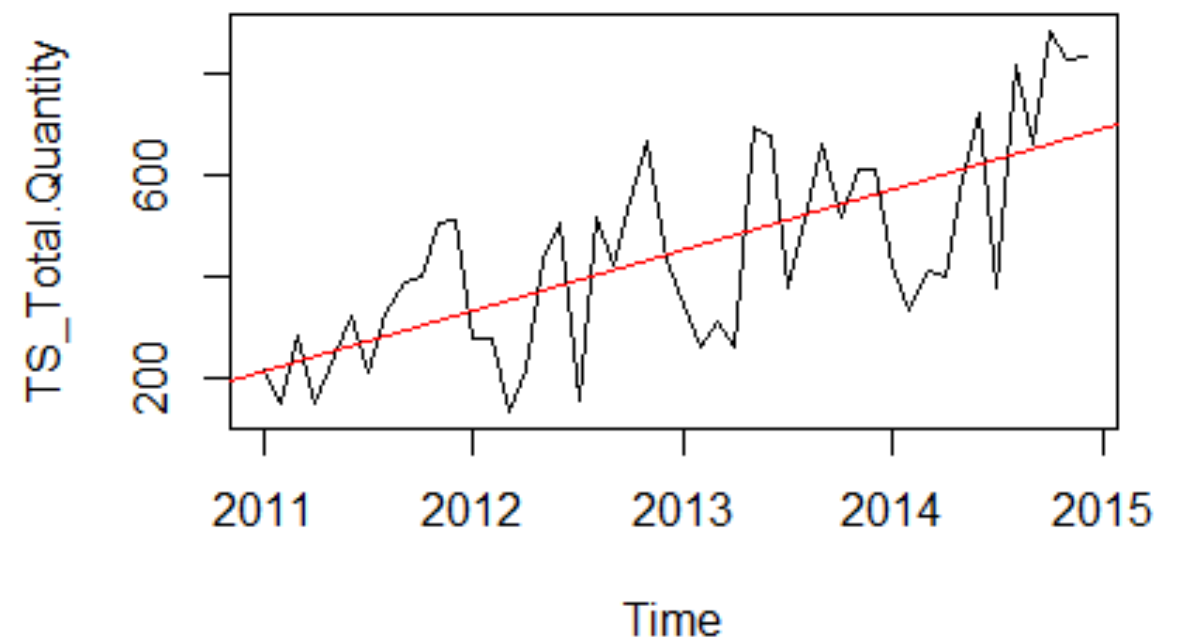
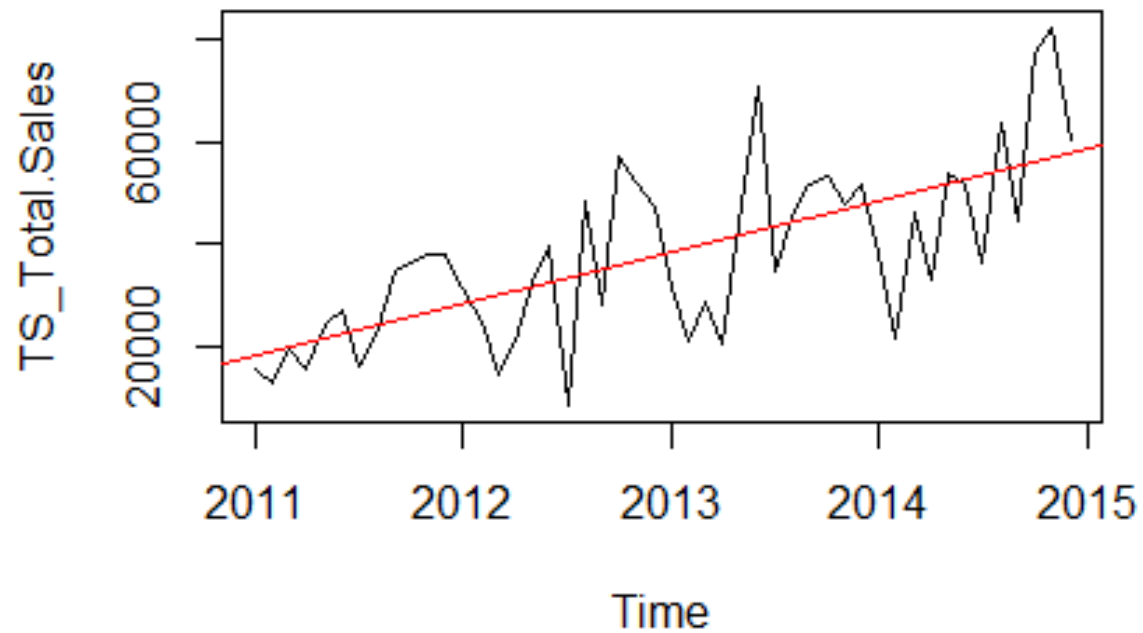
APAC Consumer Total Sales



APAC Consumer Total Quantity

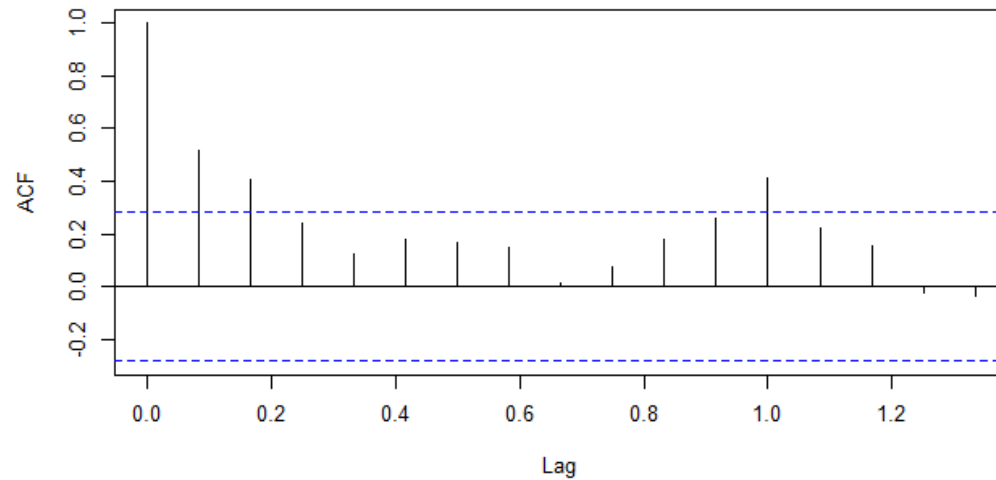


Finding linear trend for Sales and Quantity

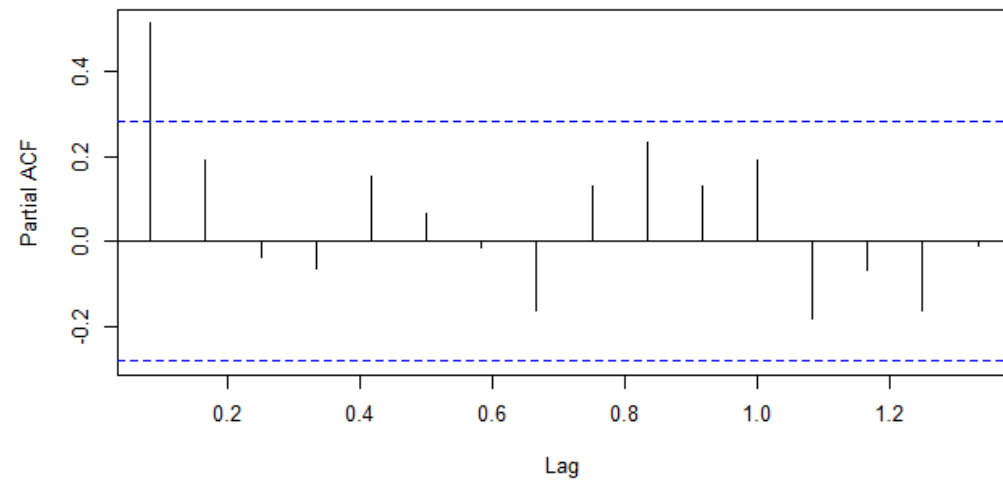


ACF and PACF plots to identify the cuts off values of the trend

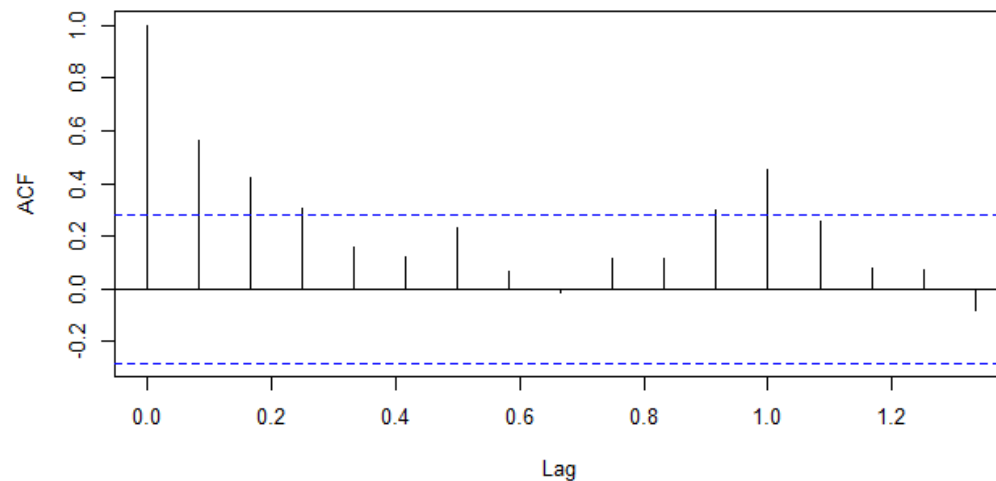
Series TS_Total.Sales



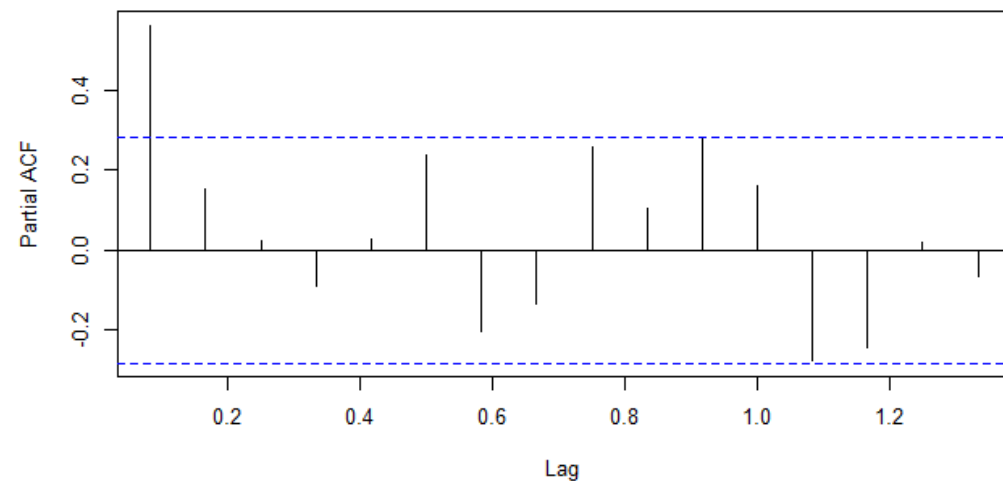
Series TS_Total.Sales



Series TS_Total.Quantity

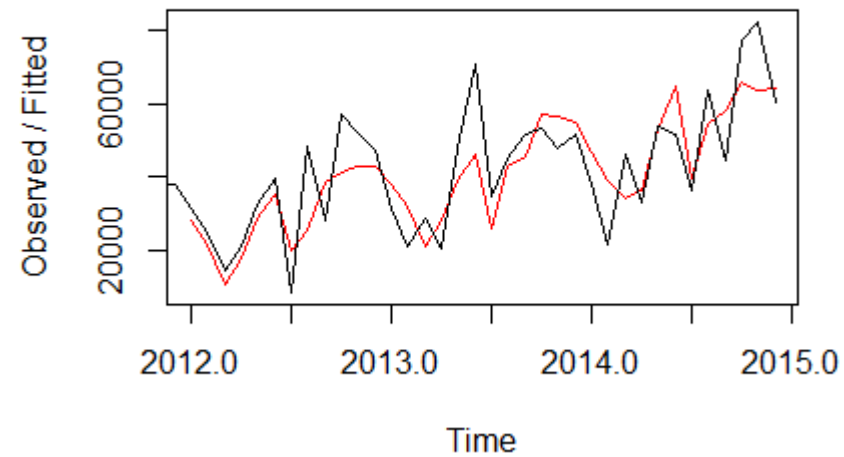


Series TS_Total.Quantity

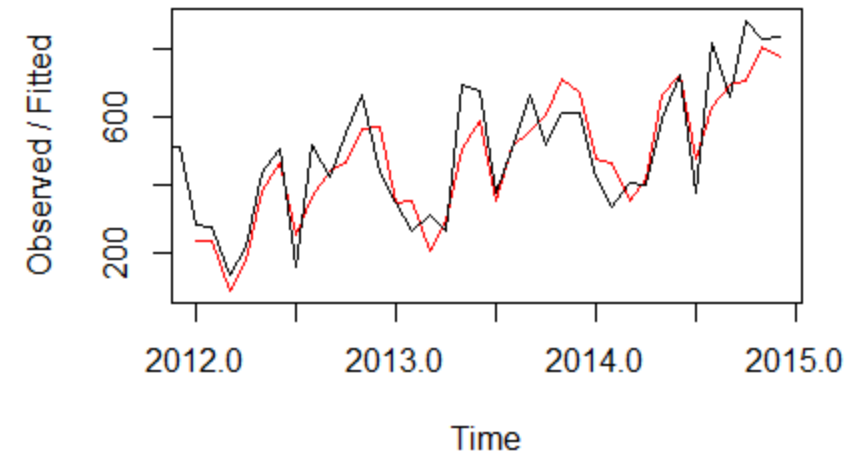


HoltWinters Smoothing : Total Sales and Quantity

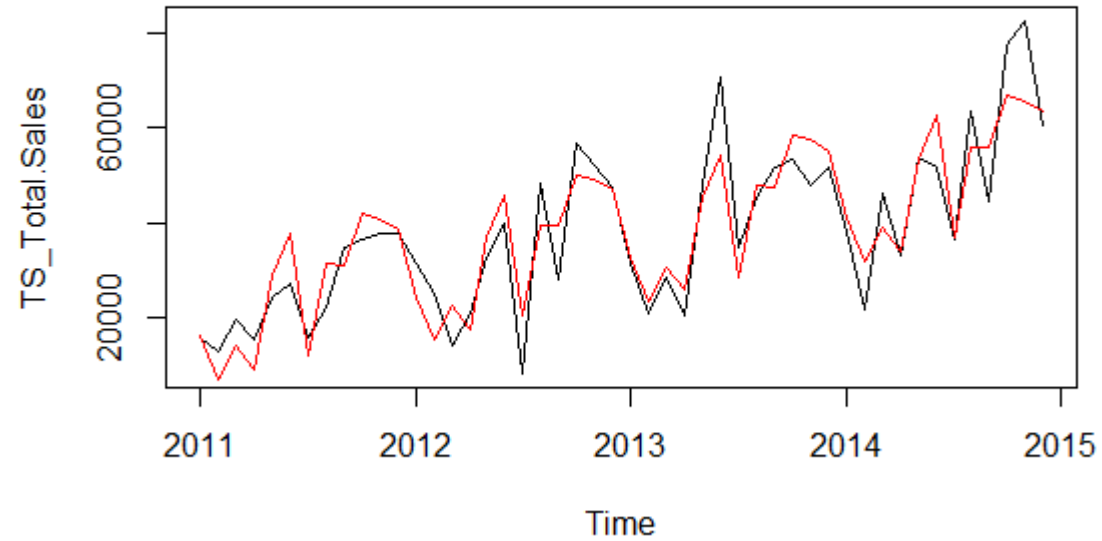
Holt-Winters filtering for Sales



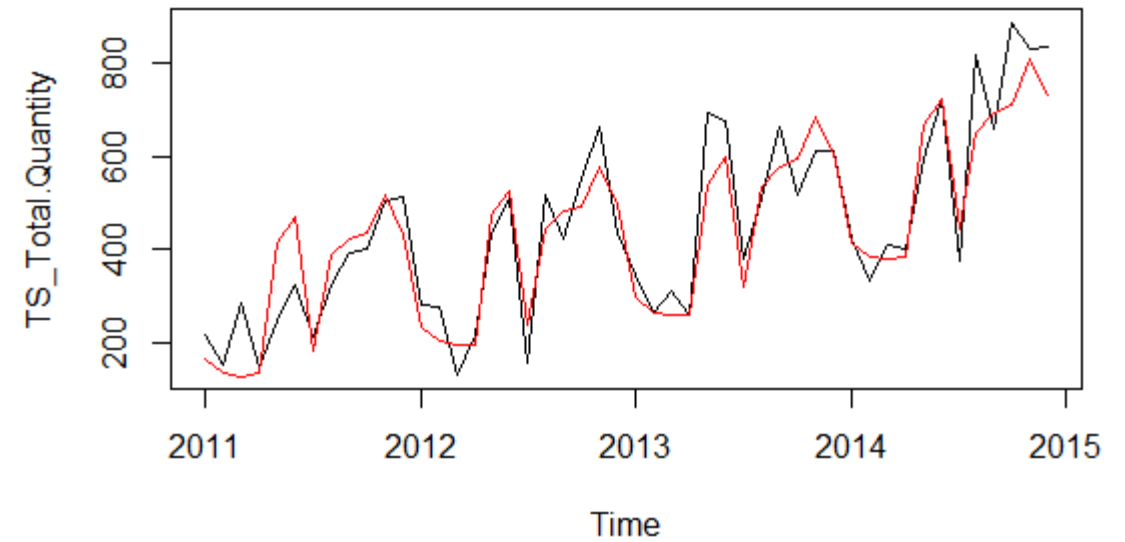
Holt-Winters filtering for Quantity



Exponential smoothing forecasts for Total Sales

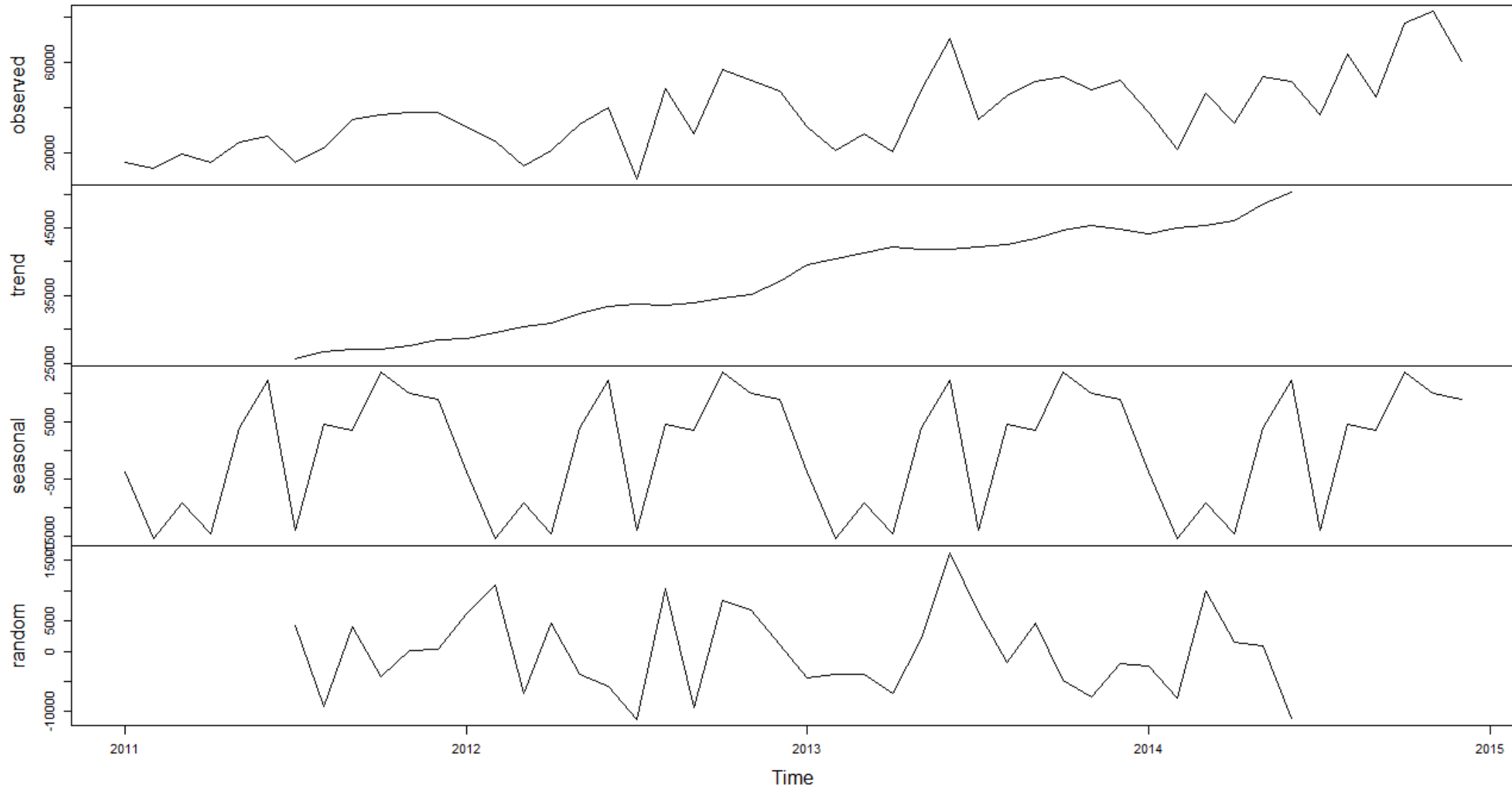


Exponential smoothing forecasts for Total Quantity



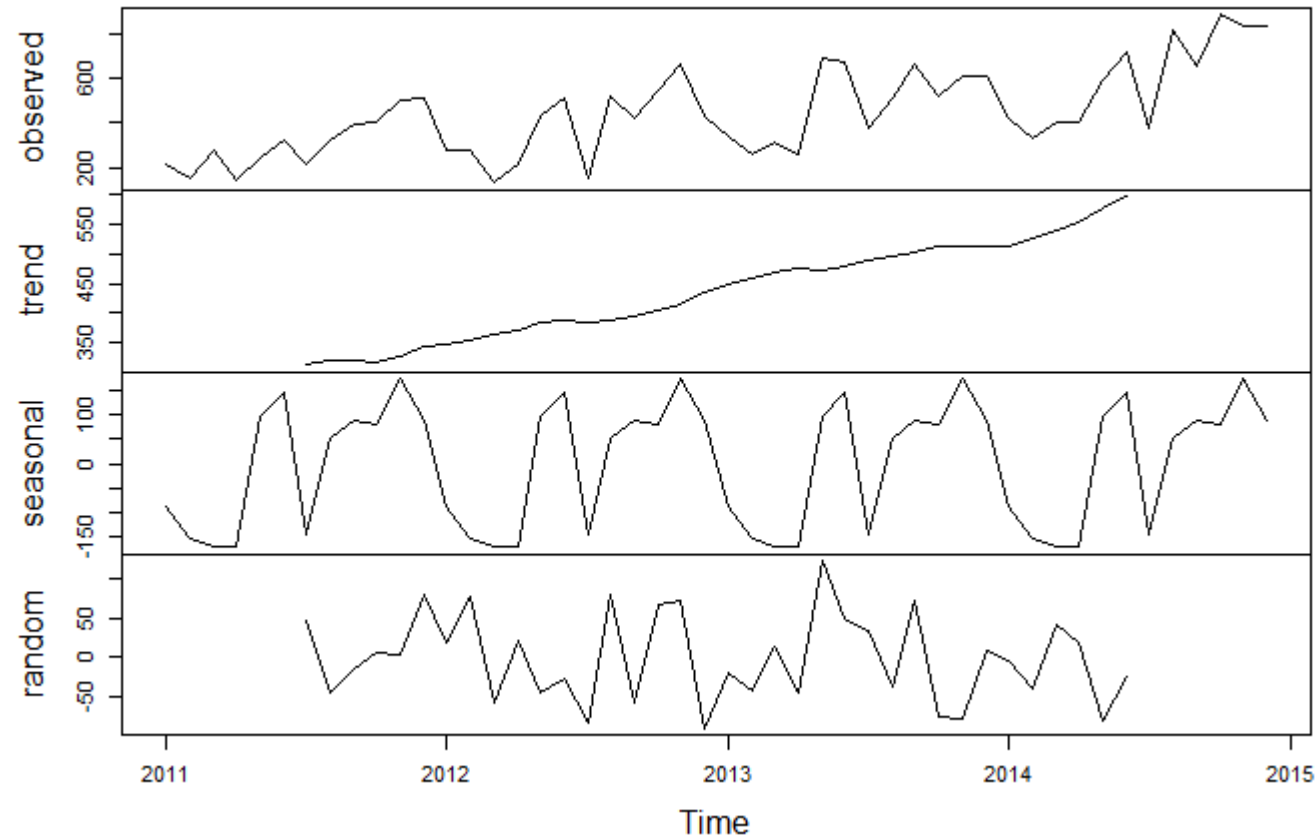
Decomposition : Applied decompose to find the moving average for Sales

Decomposition of additive time series



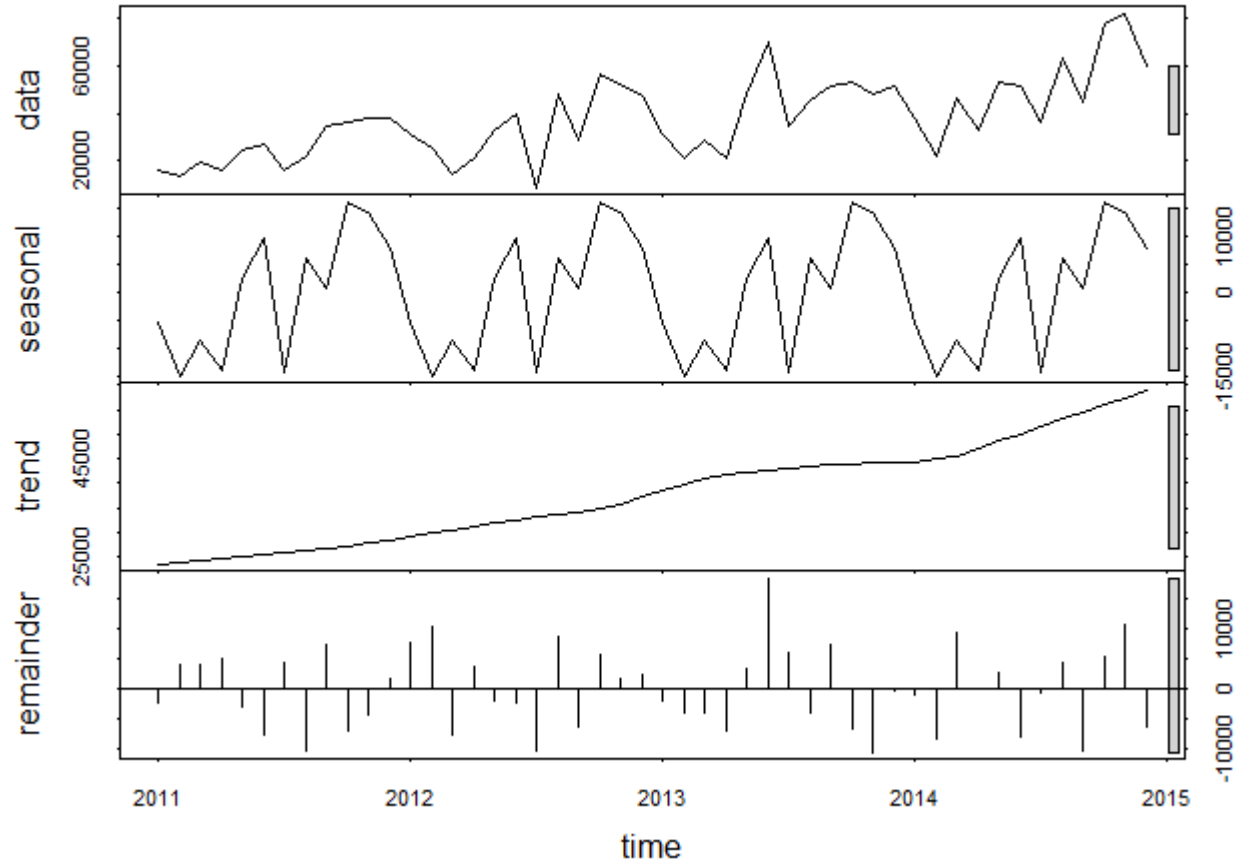
Decomposition : Applied decompose to find the moving average for Quantity

Decomposition of additive time series

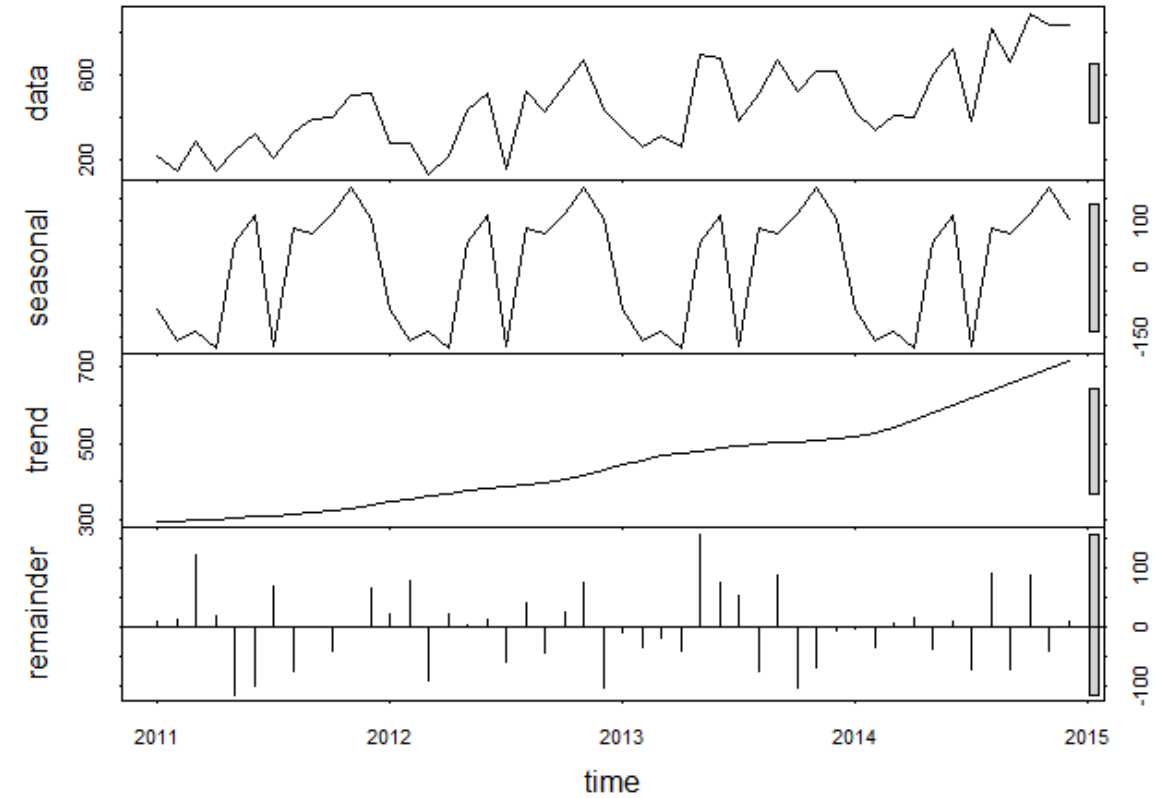


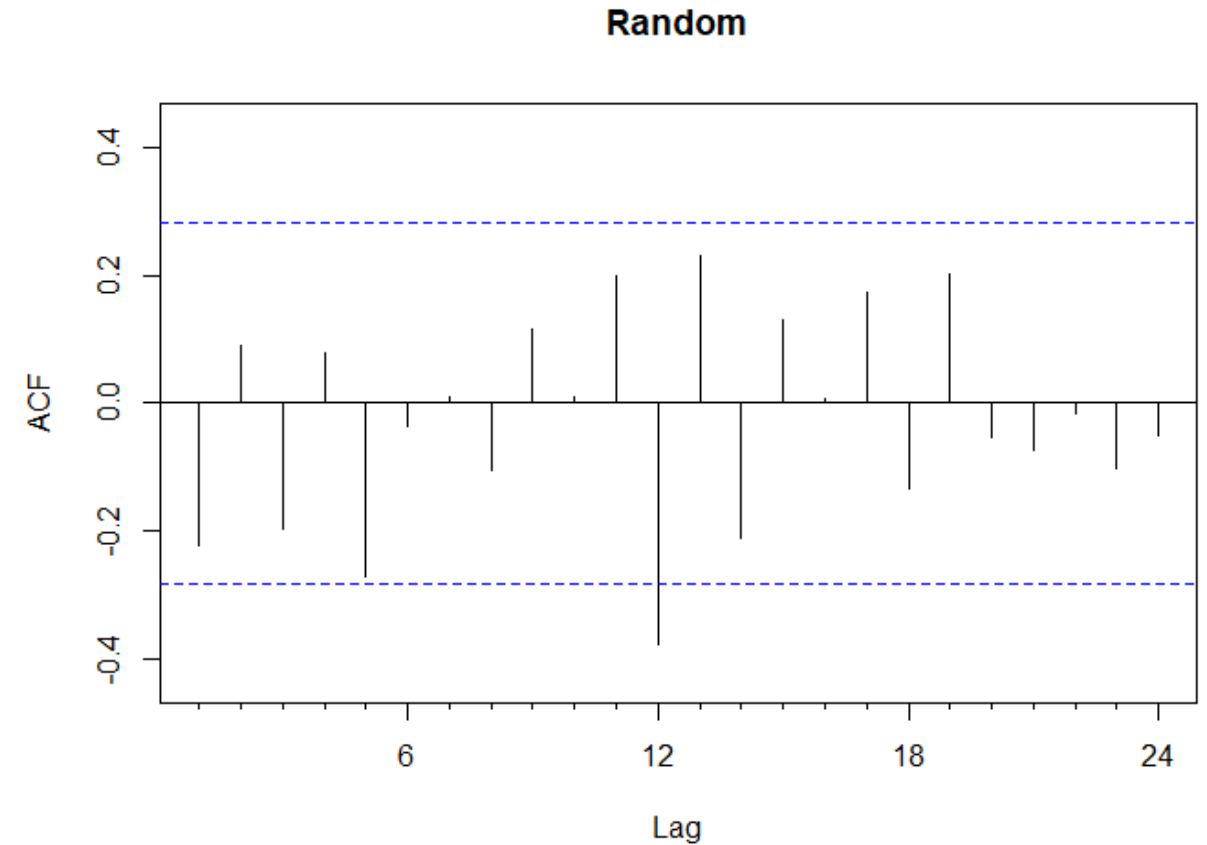
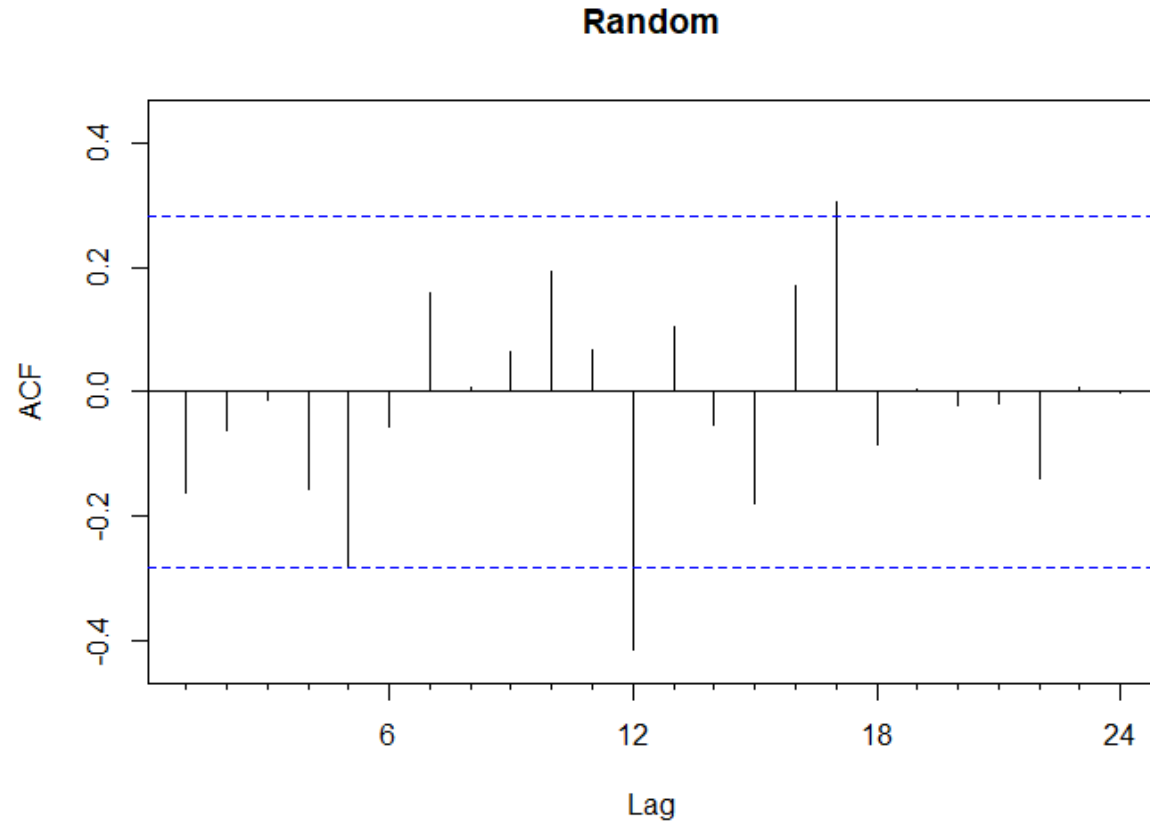
stl used the lowest smoothness

stl : lowest smooth for Sales



stl : lowest smooth for Quantity

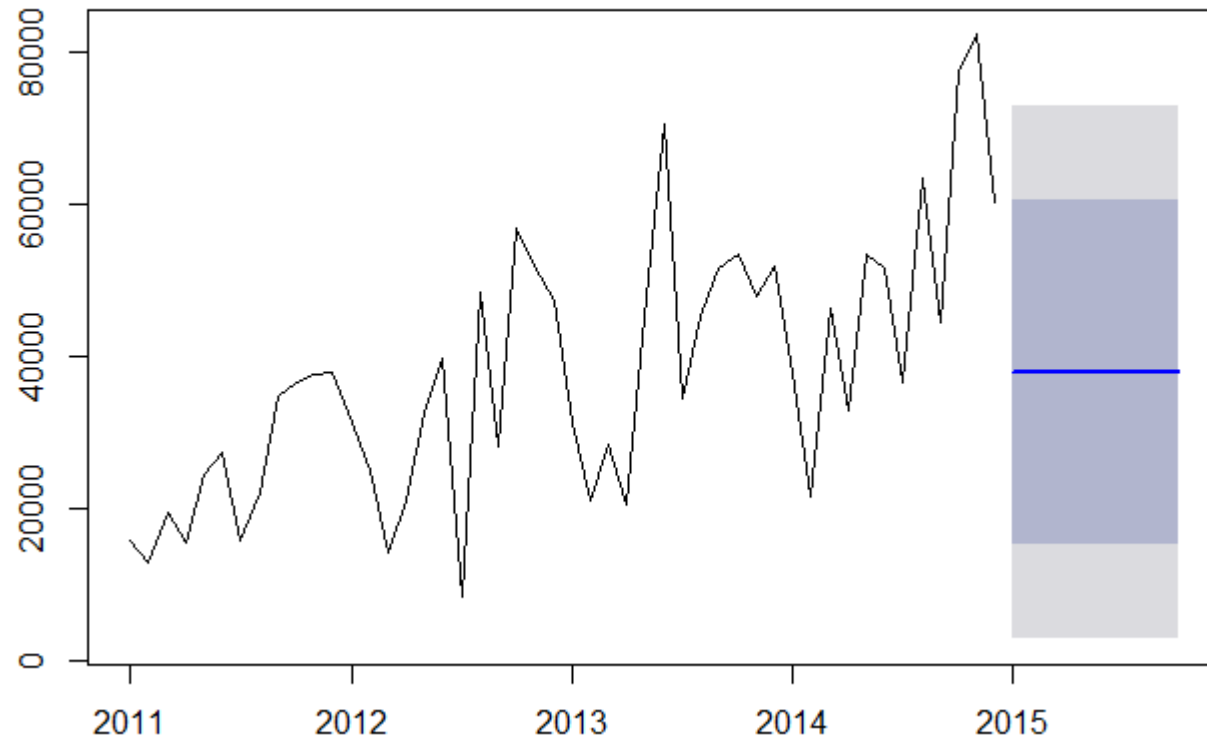




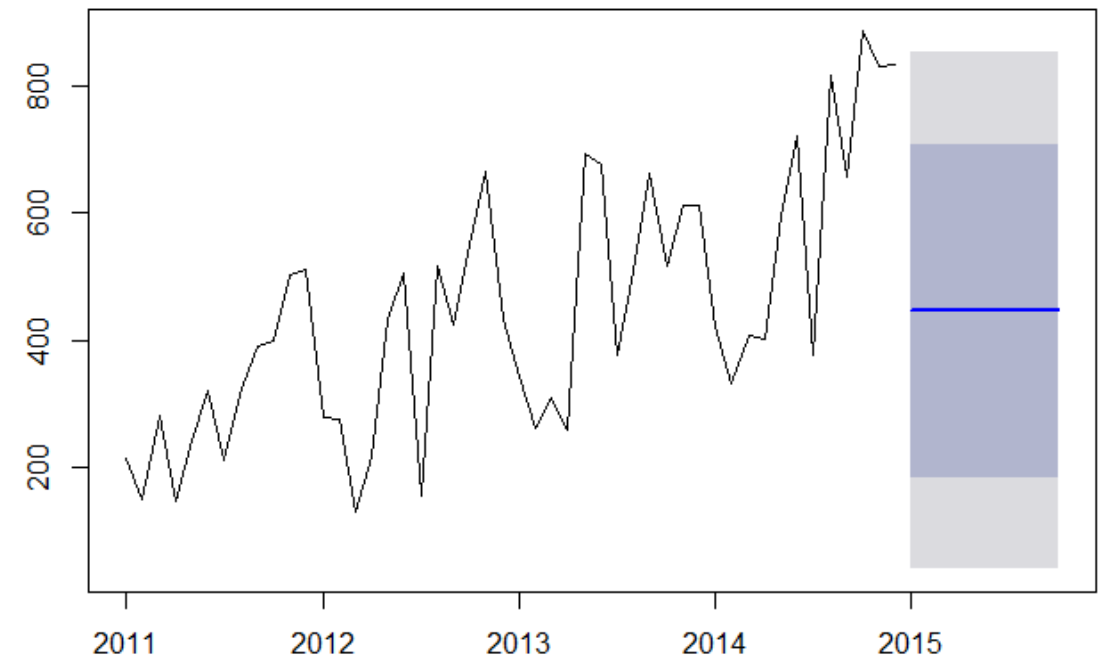
In this above chart dot line indicate significant correlation
below the blue line data observation are not correlated

Forecast : APAC-Consumer sales and quantity

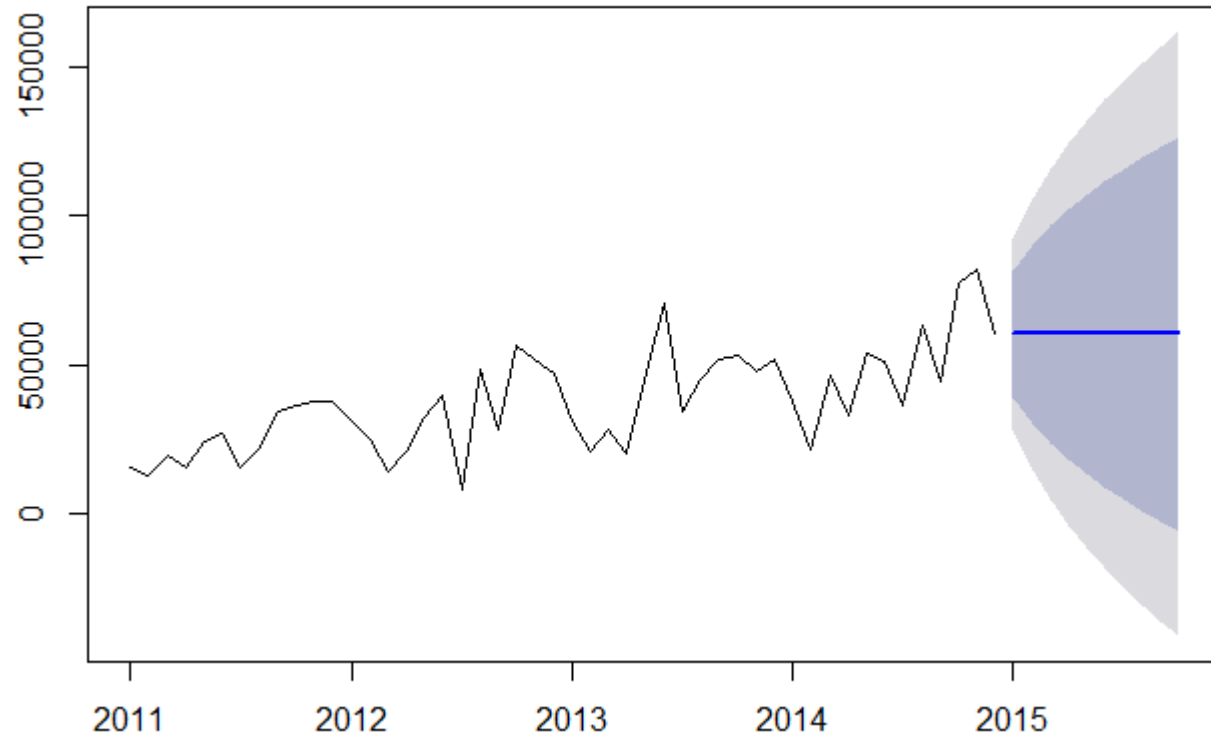
Forecasts from Mean - APAC_Consumer Sales



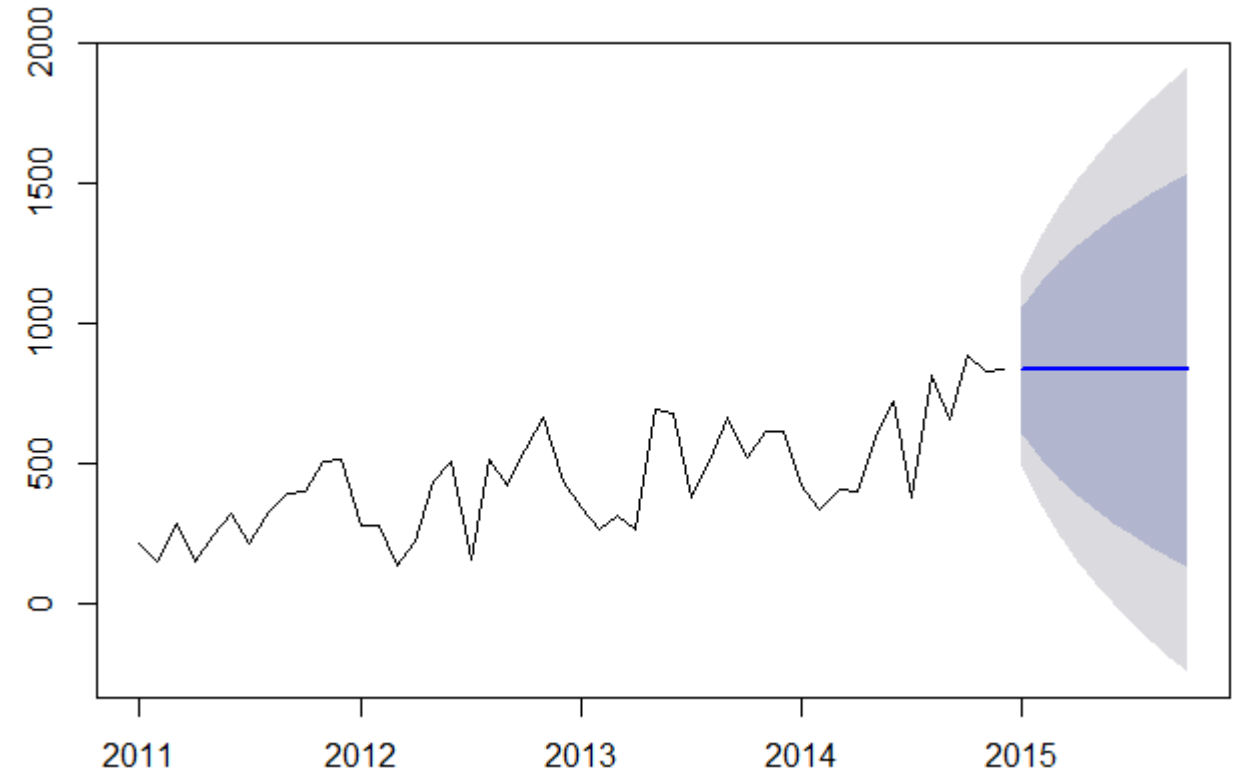
Forecasts from Mean - APAC_Consumer Quantity



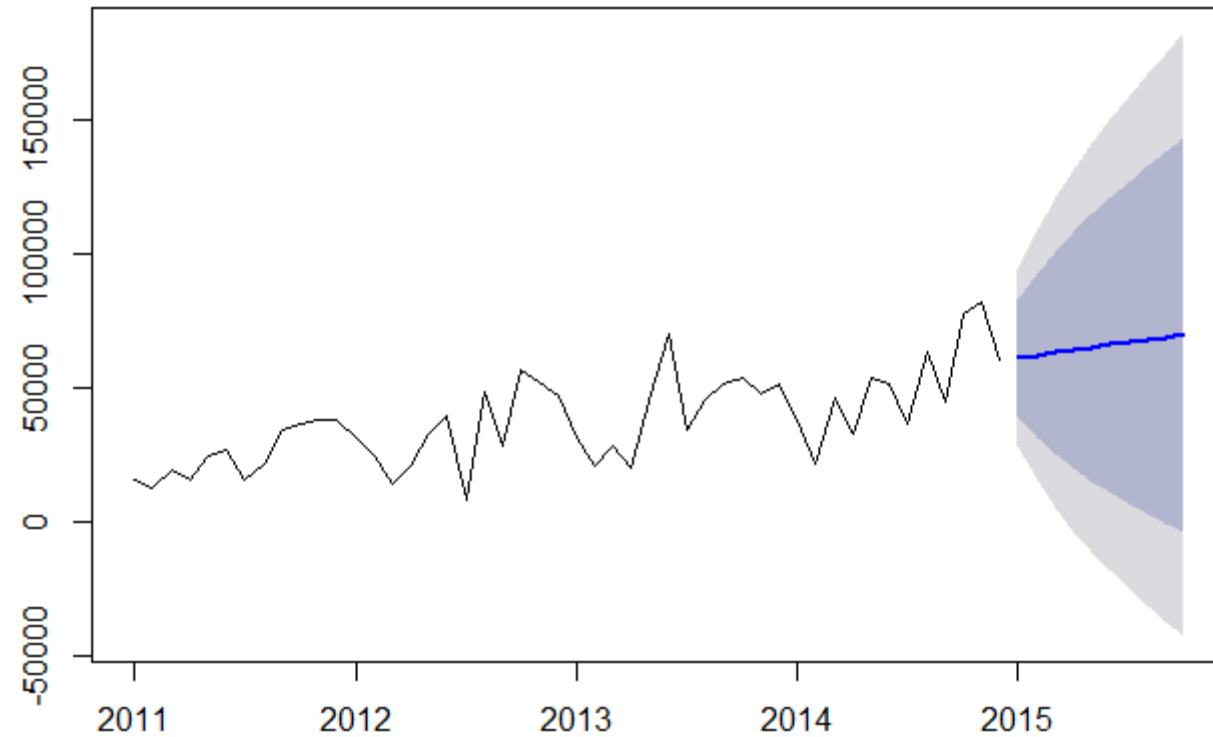
Forecasts from Navi - APAC_Consumer Sales



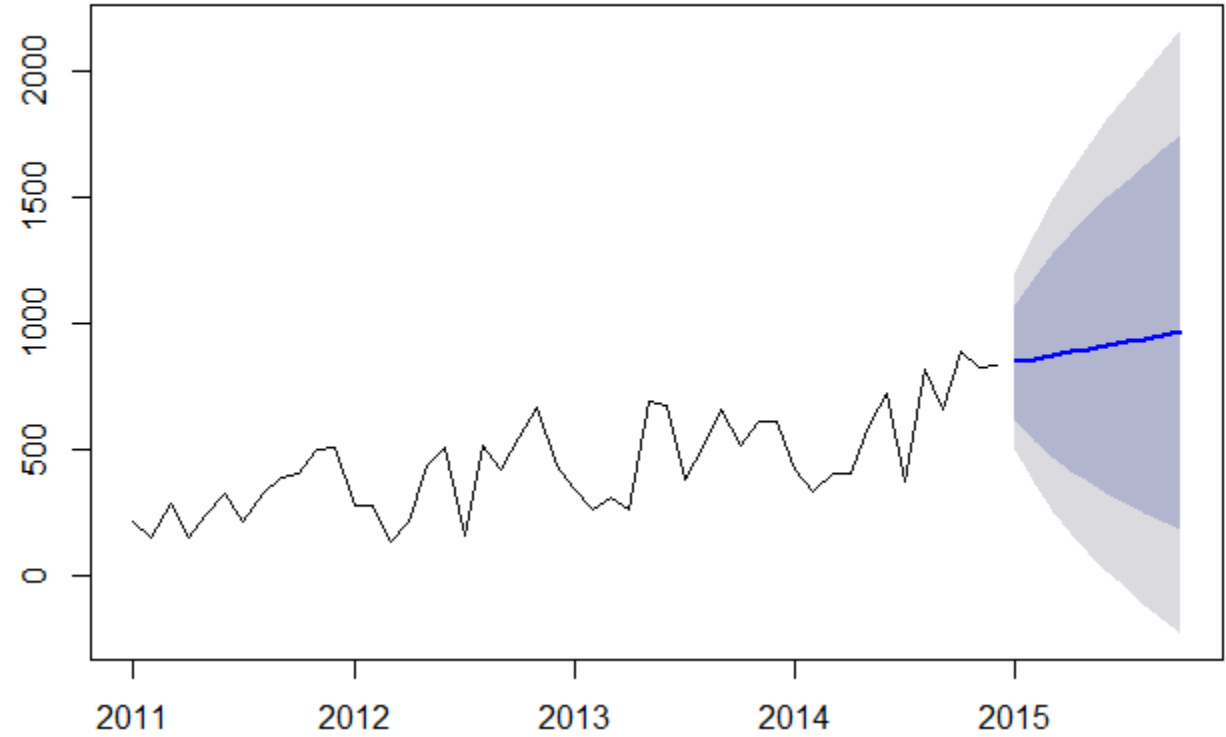
Forecasts from Navi - APAC_Consumer Quantity



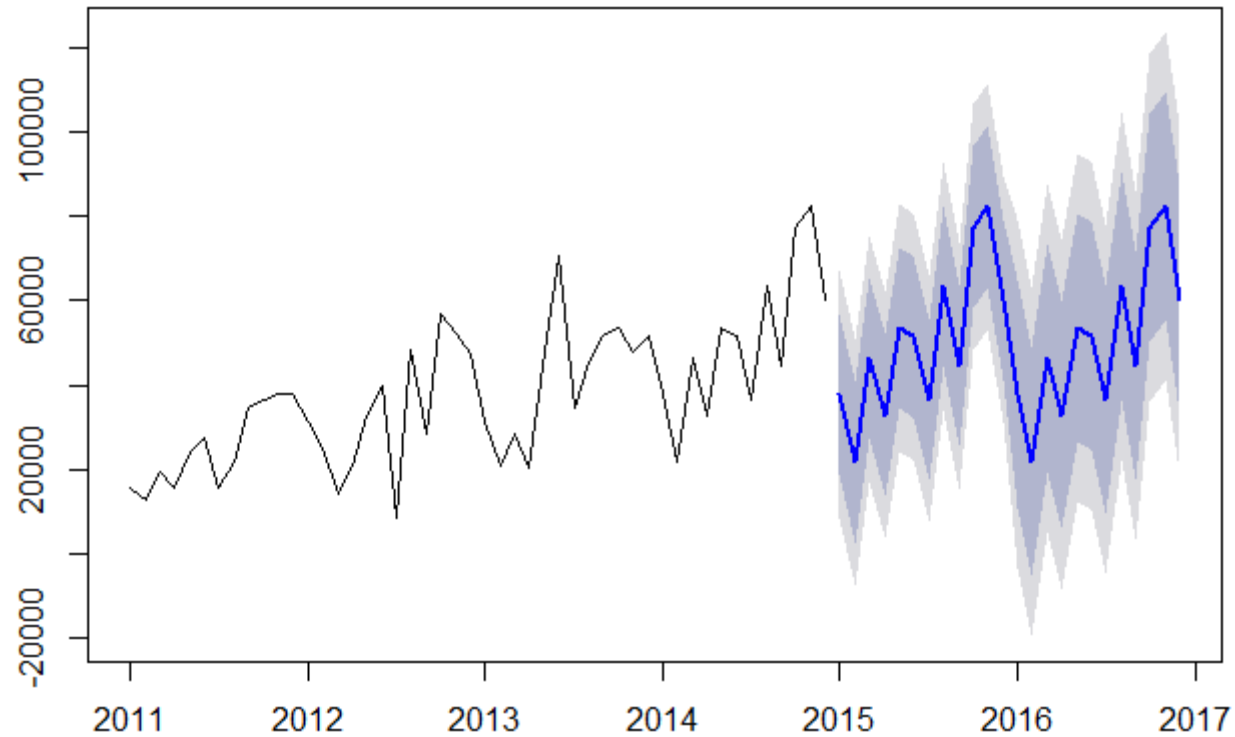
Forecasts from Random walk with drift - APAC_Consumer Sales



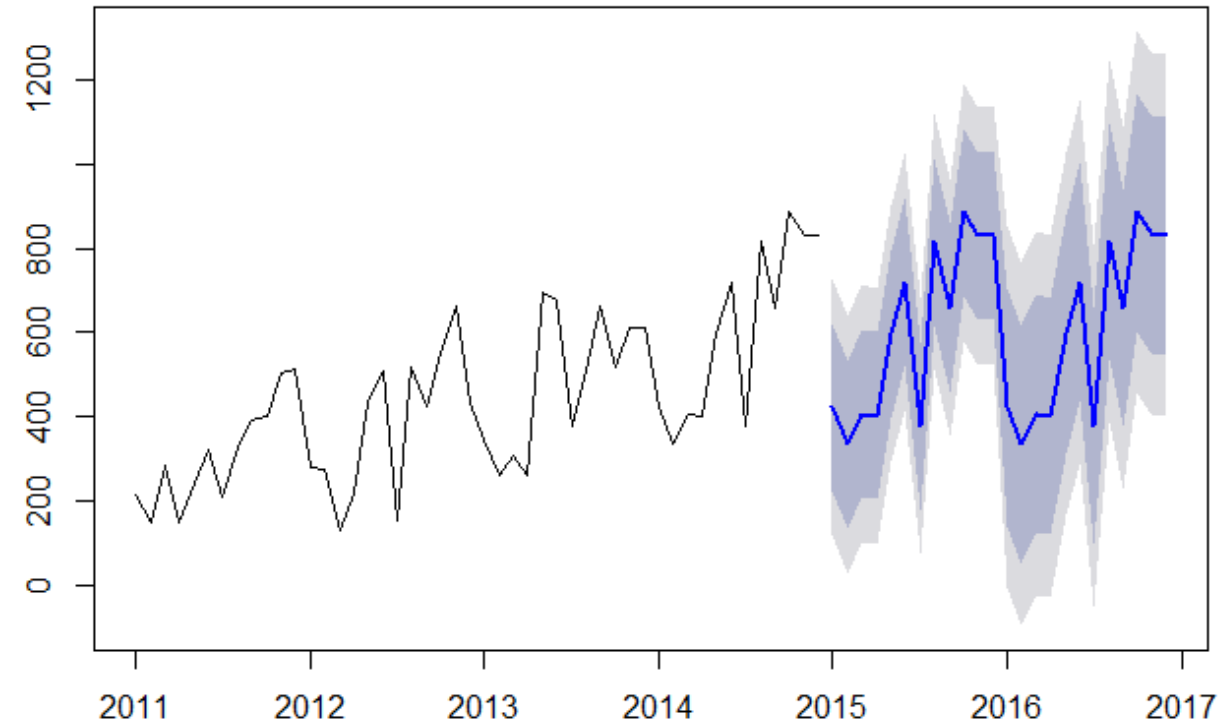
Forecasts from Random walk with drift - APAC_Consumer Quantity



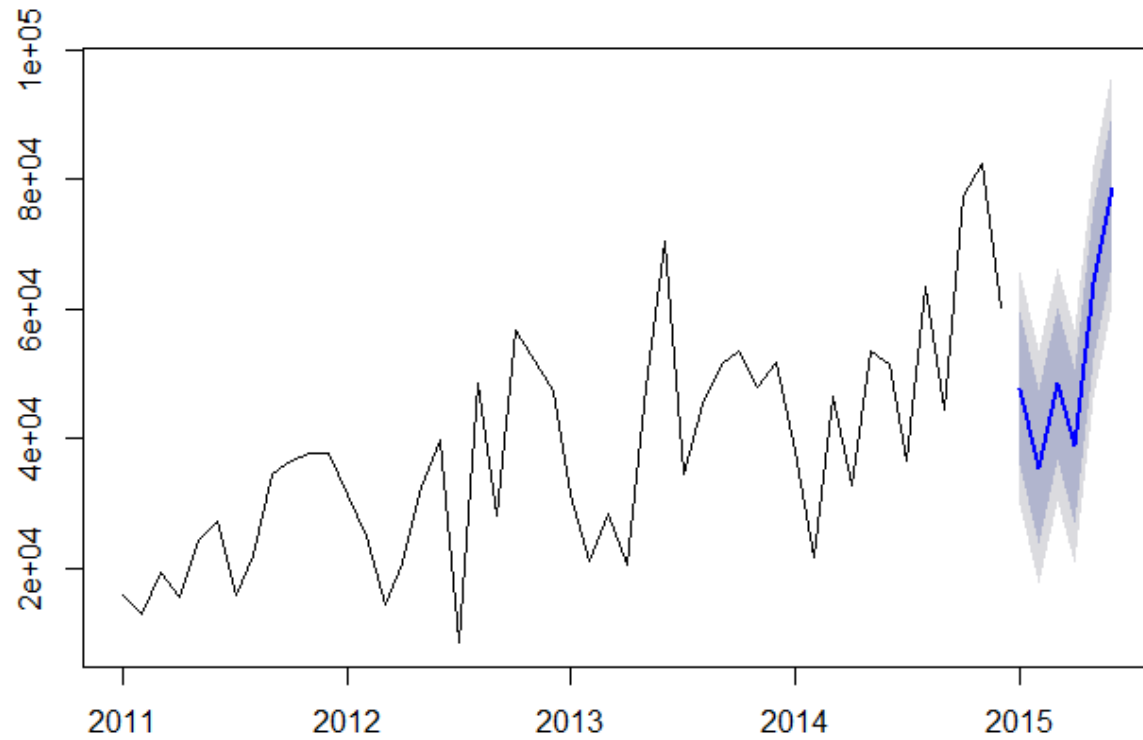
Forecasts from Seasonal naive method - APAC_Consumer Sales



Forecasts from Seasonal naive method - APAC_Consumer Quantity

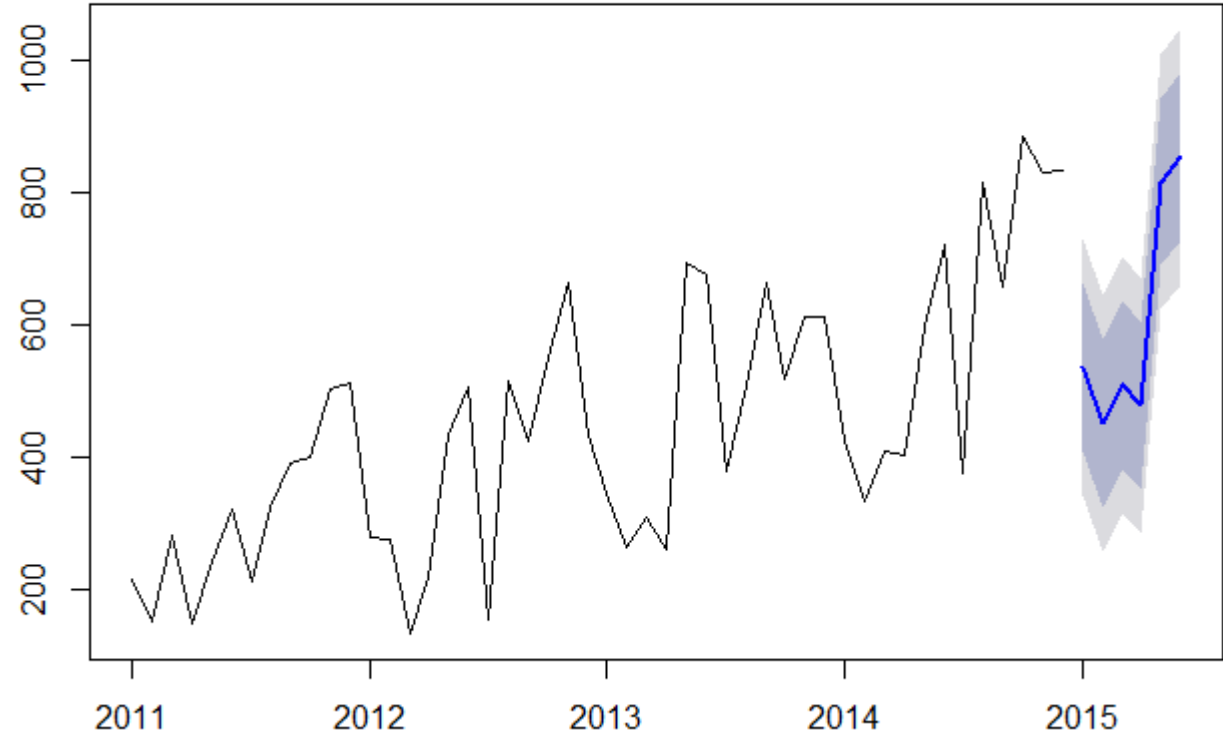


Forecasts from ARIMA(0,0,0)(1,1,0)[12] with drift



Sales

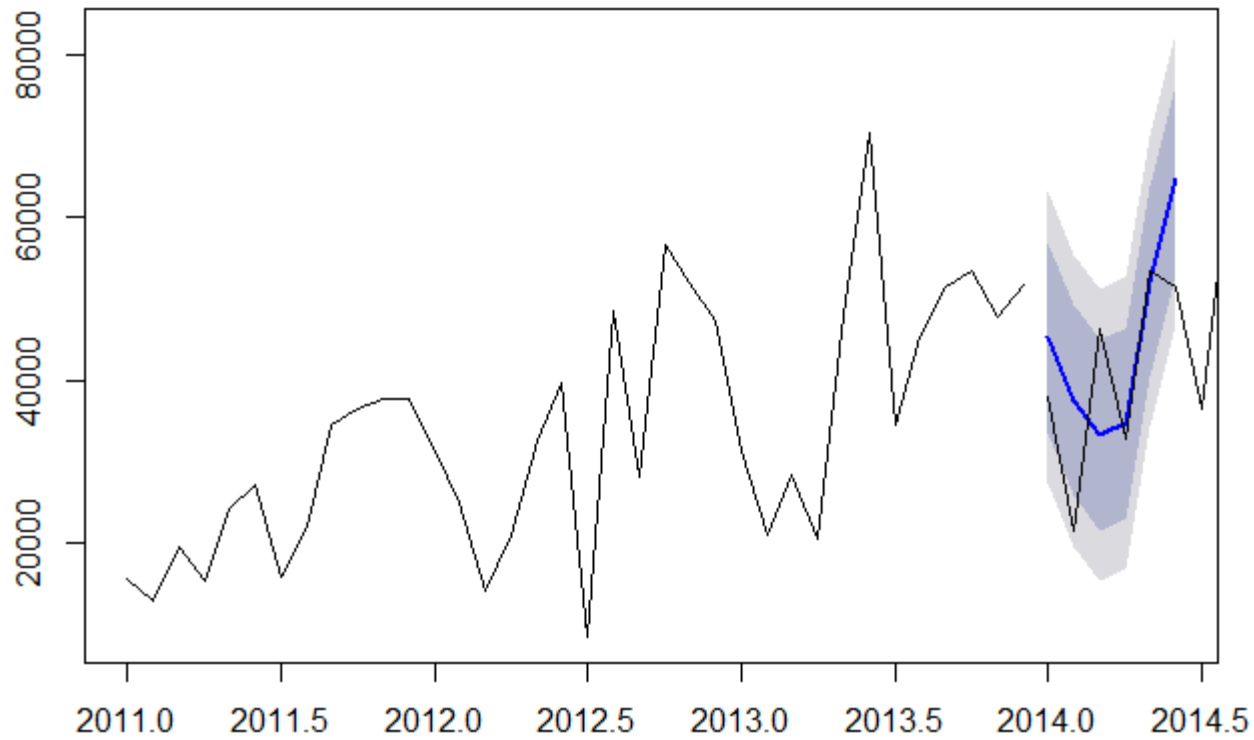
Forecasts from ARIMA(0,0,0)(1,1,0)[12] with drift



Quantity

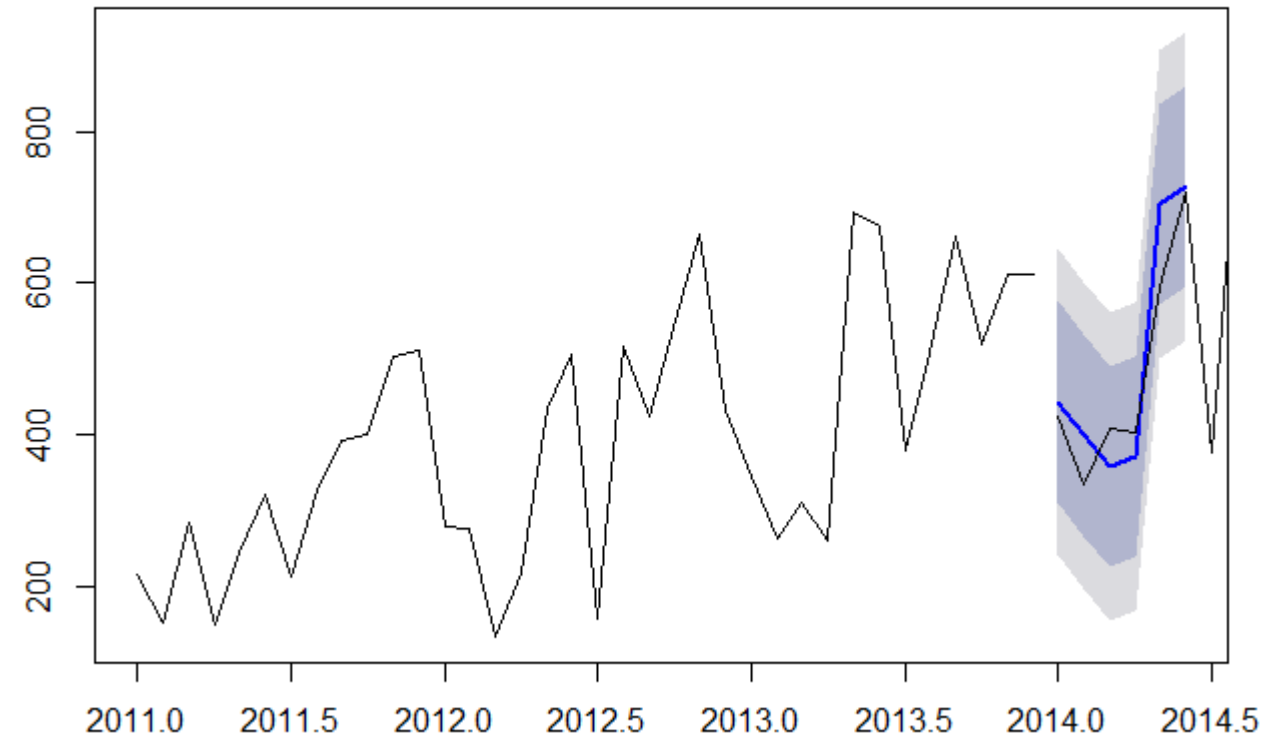
Forecasted Sales

Forecasts from $ARIMA(0,0,0)(1,1,0)[12]$ with drift



Forecasted Quantity

Forecasts from $ARIMA(0,0,0)(1,1,0)[12]$ with drift



Conclusion

1. Based on data provided we identified as APAC Consumer and EU Consumer are 2 most profitable market segments for “Global Mart” online store.
2. Below is summary of 4 key forecasts on test data(Jan – June 2015):
 - a. APAC Consumer Sales is likely to rise in next 6 months with small fluctuations.
 - b. APAC Consumer is also likely to rise steeply in coming 6 months.
 - c. EU Consumer Sales may show slow rise in coming months.
 - d. EU Consumer Quantity is likely to drop during initial 1 or 2 months & then rise rapidly in next 3 months, eventually reaching a plateau.