



Time Series Case Study

Group -

- ☐ Pritha Banerjee
- ☐ Rajarshi Ghoshal
- ☐ Priyanka Kapoor
- ☐ Piyush Baid

PGDDS C7
June'18
Group 2

The background of the slide is a blurred image of a person in a white shirt writing on a document with a pen. Overlaid on this are various business-related graphics: a line graph with an upward arrow in the top left, a bar chart in the center, and a tablet displaying charts and graphs in the bottom left. A black arrow points from the left edge towards the title.

Business Understanding

“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.

Objective-

Retail-Giant Sales Forecasting:

- ❖ To forecast the sales and the demand for the next 6 months.

Goals and Methodology

As sales/operations managers, we want to finalize the plan for the next 6 months, to forecast the sales and the demand for the next 6 months and to manage the revenue and inventory accordingly.

➤ Our Methodology–

- ❖ The store caters to 7 different market segments and in 3 major categories.
- ❖ To forecast at the granular level, we will subset the data into 21 ($7*3$) buckets
- ❖ Out of these 21 market buckets we need to find out 2 most profitable (and consistent) segment
- ❖ Forecast the sales and demand for these two segments -
 - ❖ Smoothing the data before performing classical decomposition.
 - ❖ Using Classical Decomposition and Auto ARIMA to build Time Series Models.
 - ❖ To test the Accuracy of prediction with the help of MAPE values for both the models.
 - ❖ With the selected model, forecasting the sales/demand for next 6 months.

Metadata and Data Understanding

- There are **51290 unique** transaction level data of **24 attributes**, where each row represents a particular order made on the online store.
- ❑ No Duplicate data nor any missing values were observed with the data set provided.
- ❑ NA was observed only in one column: Postal.Code. That column was deleted as it was of no use with respect to forecasting.
- ❑ **A Brief Description of the Data Attributes: Data Dictionary**

Attributes	Description
Order ID	Unique ID of the transaction
Order Date	Date on which the order was placed
Ship Date	Date on which the shipment was made
Ship Mode	The mode of shipment (category)
Customer ID	The unique ID of the customer
Customer Name	Name of the customer
Segment	The market segment to which the product belongs
City	City of the delivery address
State	State of the delivery address
Country	Country of the delivery address
Order Priority	Priority assigned to the order

Attributes	Description
Market	Market segment to which the customer belongs
Region	Geographical region of the customer
Product ID	Unique ID of the product
Category	Category of the product
Sub-Category	Sub-category of the product
Product Name	Name of the product
Sales	Total sales value of the transaction
Quantity	Quantity of the product ordered
Discount	Discount percentage offered on the product
Profit	Profit made on the transaction
Shipping Cost	Shipping cost incurred on the transaction

Assumptions

- ❑ All Sales and profit amounts are of same unit

Problem Solving Methodology -

Business Understanding

- Getting a clear picture of the business requirement



Extracting the Data

- Downloading / Importing the Data into R



Data Preparation

- Create 21 Market-Segment buckets
- Find Sales, Profit, Quantity for each bucket
- Calculate Total Profit, Coefficient of Variation(CV)
- Pick Top 2 Segments based on lowest CV & highest Profit
- Aggregate data by Order Month on subset of Top 2 segments



Model Building

- Create time series of top Aggregated Data
- Smoothen time series with Moving Average method to identify trend & seasonality
- Creating Time Series models with the **Classical Decomposition** and **Auto ARIMA** methods



Evaluation

- Create **plots for ACF of residuals** to check if the model falls under the acceptable band of correlation values.
- Perform **ADF, KPSS test** to check if the residual series is a white noise
- Evaluate model on validation set using **MAPE**
- Choose the best model



Forecasting

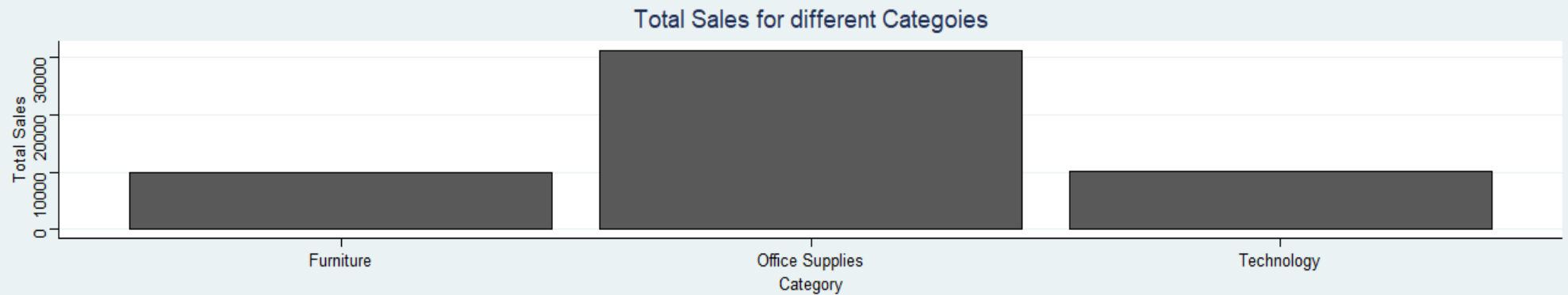
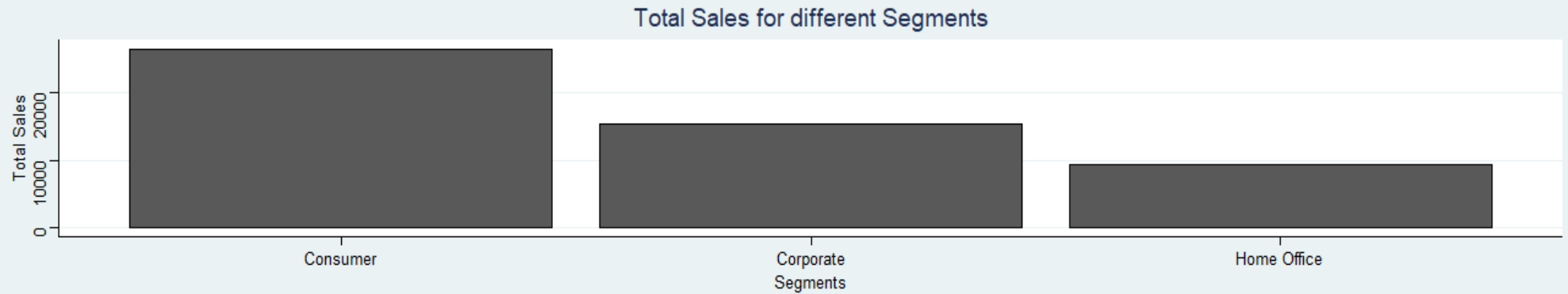
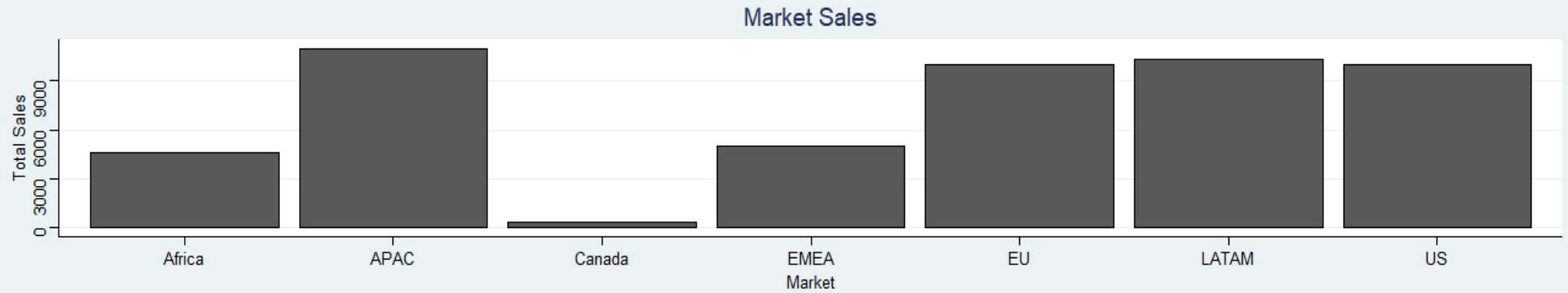
- Use best model to forecast future 6 months Sales.
- Repeat same for both Market-Segments separately for Sales & Quantity
- Prepare 4 forecasts for top 2 segments by Sales & Quantity

Data Preparation and EDA

- All the data sets were cleaned and prepared to be merged to form a core data file for analysis.
- A new column: Market_Segment was created by combining the two attributes
- 21 subsets were created grouping by market_segment
- Creating Timestamp for the dataset, for each month
- Calculated (monthly) Sum of Profit, Mean of Profit, SD of Profit and finally derived Coefficient of Variance, i.e. CV of Profit.
- (CV is ratio of Mean and SD.)
- Calculated Total Sales Quantity for each month
- Comparing the Total Profit (Maximum) and CV (Lowest) we derive at the conclusion that **APAC_Consumer** and **EU_Consumer** are two of the best segments
- Start with Time Series for these two segments

Visualization - Grid 1

- Finding Total Sales of Different Markets, Segments and Categories.



Visualization - Grid 2

- Finding Total Profit of Different Markets, Segments and Categories.

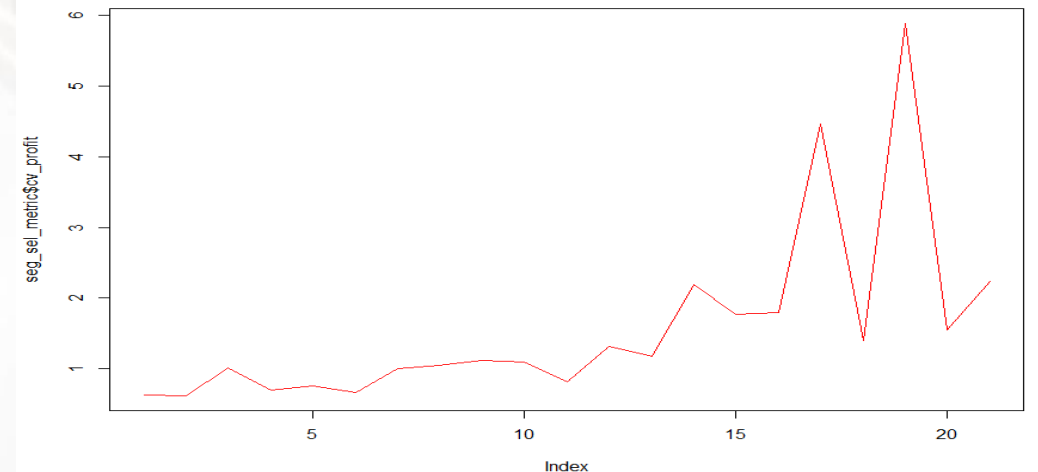
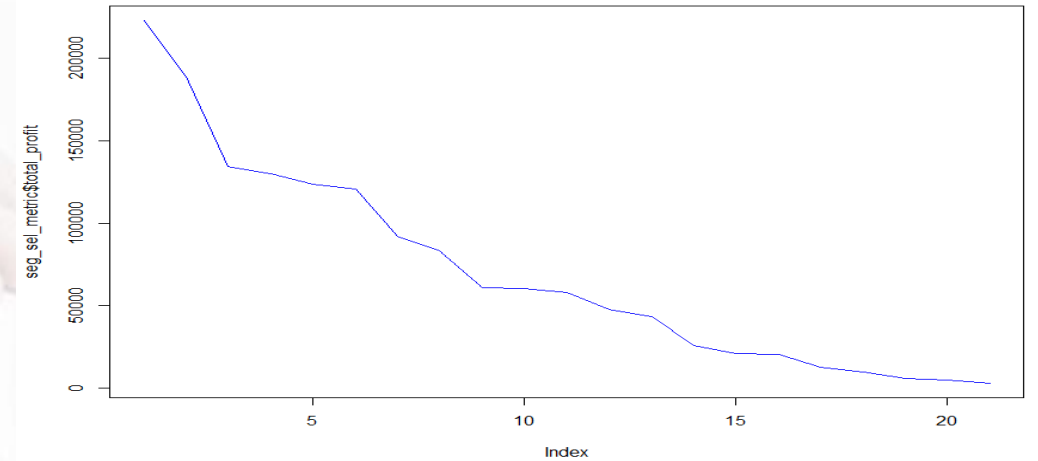


Visualization - Grid 3

- Finding the top Two market segments based on High Profit and Low CV

	market_seg	total_profit	cv_profit	total_Sales	total_quantity
4	APAC_Consumer	222817.560	0.6321323	1816753.70	21414
13	EU_Consumer	188687.707	0.6243052	1529716.24	19541
19	US_Consumer	134119.209	1.0123900	1161401.34	19521
5	APAC_Corporate	129737.235	0.6980869	1078466.31	12142
14	EU_Corporate	123393.980	0.7638072	920008.28	11635
16	LATAM_Consumer	120632.932	0.6614828	1133847.03	19853
20	US_Corporate	91979.134	1.0024089	706146.37	11608
6	APAC_Home Office	83445.254	1.0459784	690524.11	7670
15	EU_Home Office	60748.054	1.1165073	488364.54	6597
21	US_Home Office	60298.679	1.0961473	429653.15	6744
17	LATAM_Corporate	57875.421	0.8111217	645252.48	11499
1	Africa_Consumer	47772.099	1.3195849	423766.81	5503
18	LATAM_Home Office	43135.134	1.1756978	385505.65	7174
10	EMEA_Consumer	25532.574	2.1882709	406745.18	5871
2	Africa_Corporate	20686.965	1.7761054	204938.95	2943
3	Africa_Home Office	20412.567	1.7899957	155067.45	2118
11	EMEA_Corporate	12499.134	4.4671024	250571.39	3515
7	Canada_Consumer	9677.700	1.3953122	35719.11	454
12	EMEA_Home Office	5866.263	5.8807467	148844.73	2131
8	Canada_Corporate	5036.460	1.5527751	19313.73	223
9	Canada_Home Office	3103.230	2.2434607	11895.33	156

- The two plots below show, how CV decreases with Increase in Profit

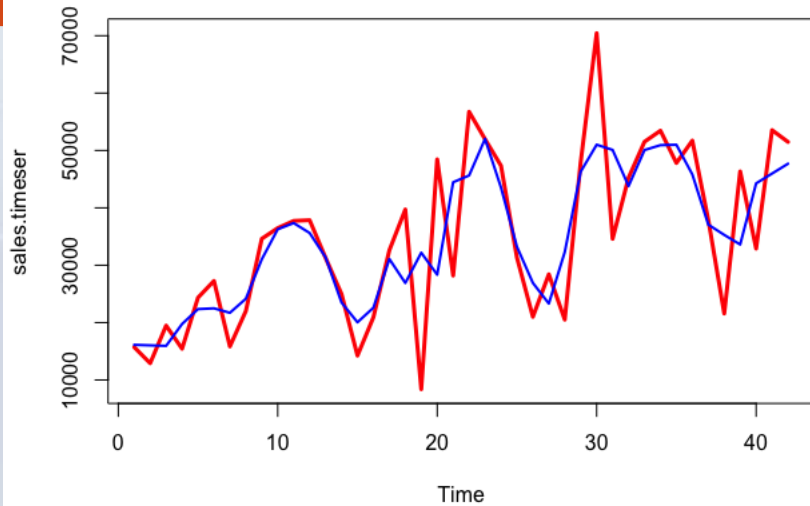


Model Building & Evaluation

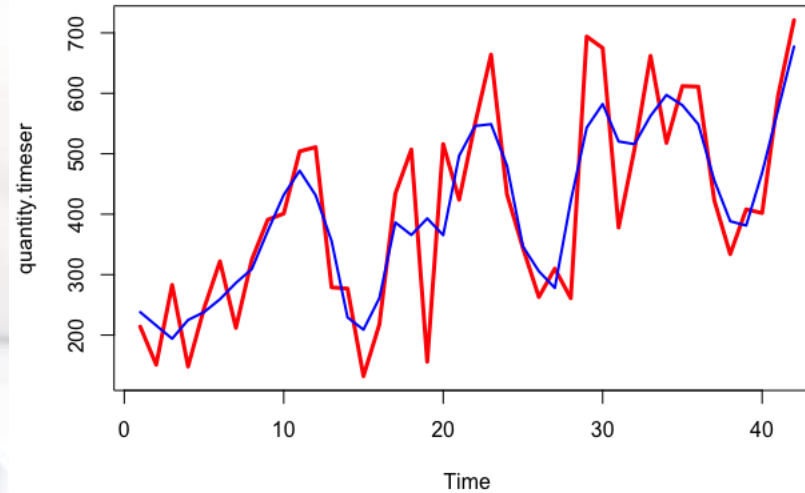
- Start with any one of the two dataset of **APAC Consumer** and **EU Consumer**.
- Create Sales Time Series for **42 timestamps** (out of 48 timestamps) – Training Model.
- Smoothened time series using **Moving Average method**.
- Created a smoothed dataframe.
- Use **Classical Decomposition** Method for modelling
- Create **plots for ACF of residuals** to check if the model falls under the acceptable band of correlation values.
- Perform **ADF, KPSS test** to check if the residual series is a white noise.
- Evaluate the Model using **MAPE**.
- Plotted the predicted value along with original values.
- Using **Auto-ARIMA** Method
- Repeat Checking for White Noise, Evaluation using MAPE and Plotting of predicted value along with original values.
- The same process shall be repeated for Quantity.
- All the Steps shall be repeated for the next dataset also.

Smoothened Time Series

APA Consumer - Sales

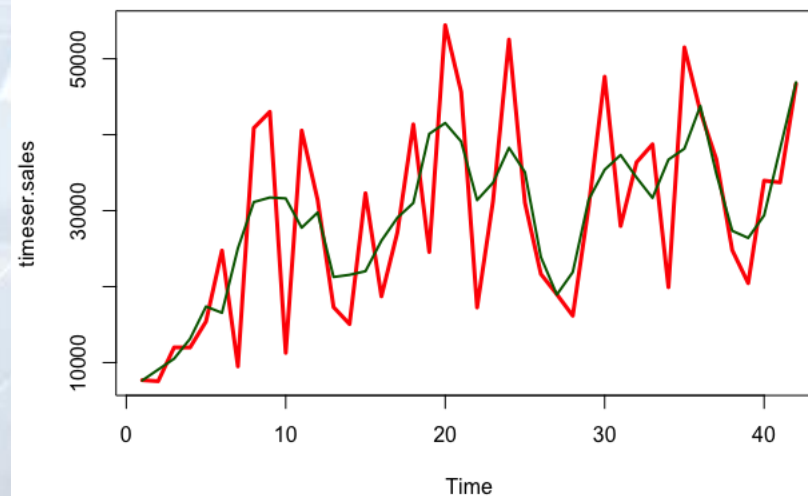


APAC Consumer - Quantity

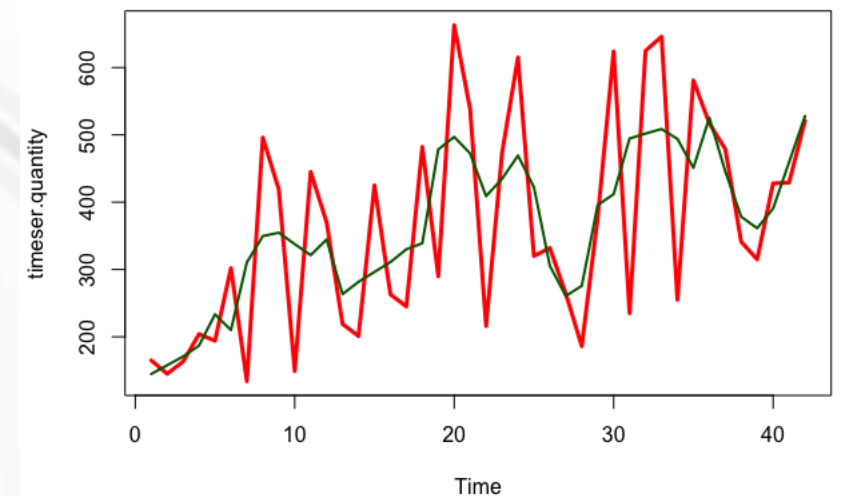


Time series were
smoothened with Moving
Average technique,
where **Width was taken
as 1** for all the models.

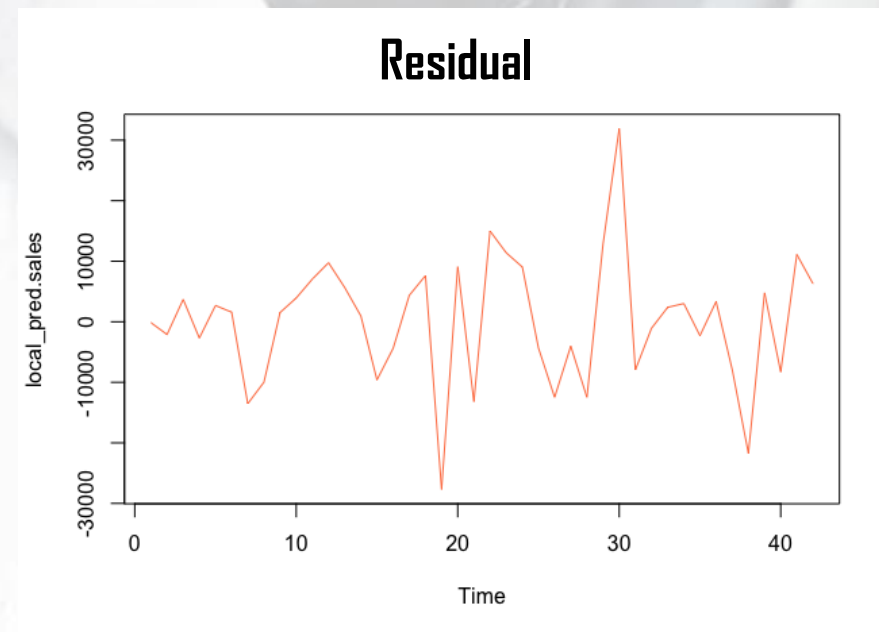
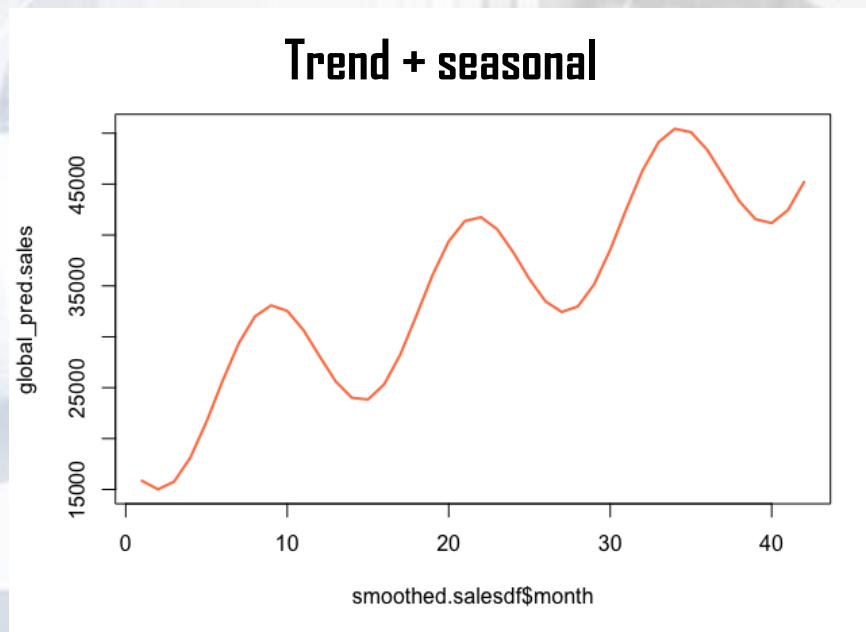
EU Consumer - Sales



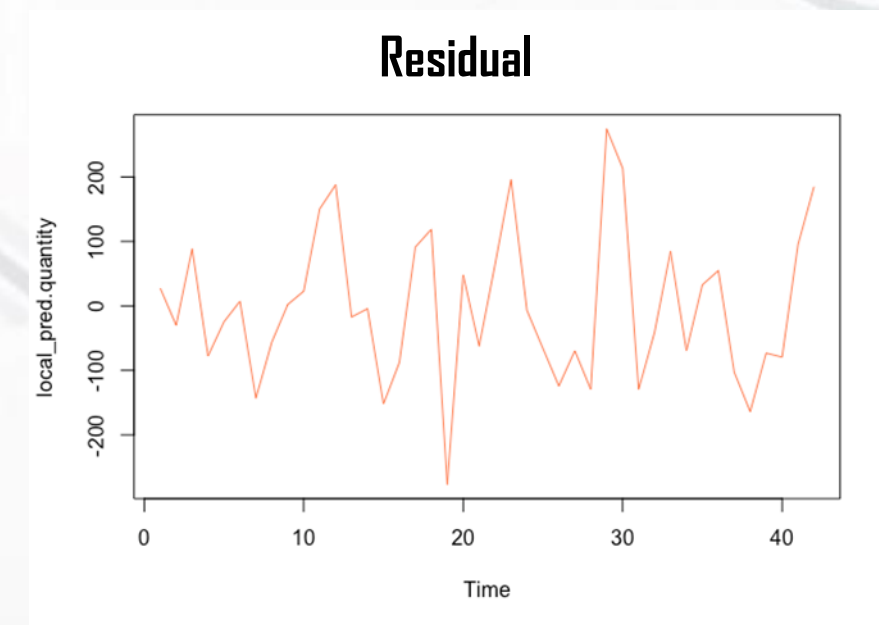
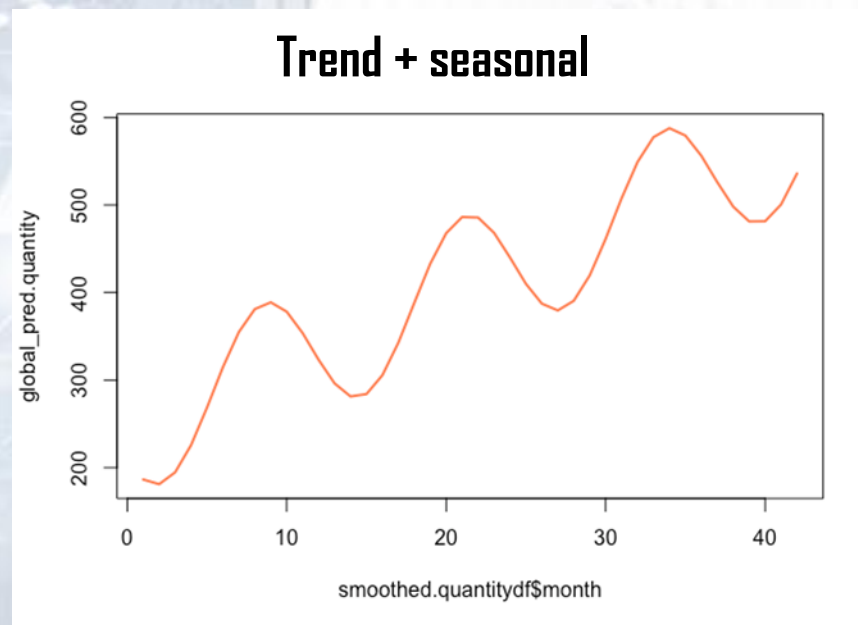
EU Consumer - Quantity



APAC Consumer Sales Decomposition Chart

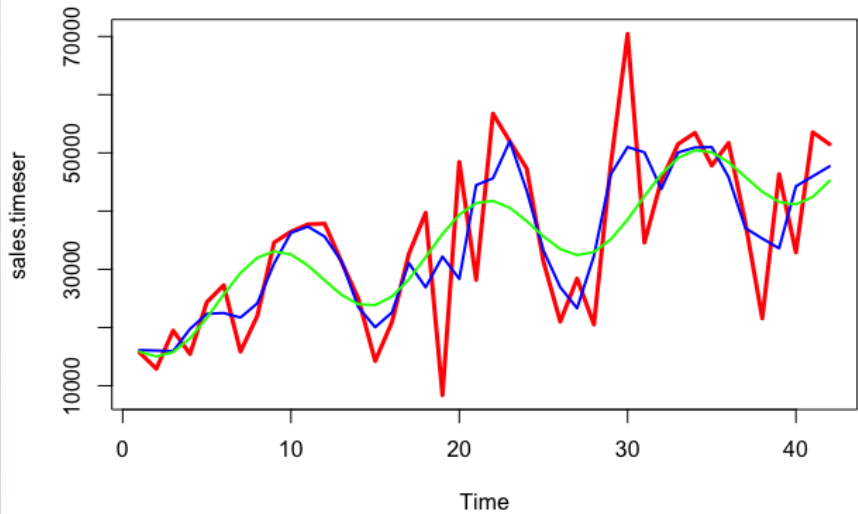


APAC Consumer Quantity Decomposition Chart

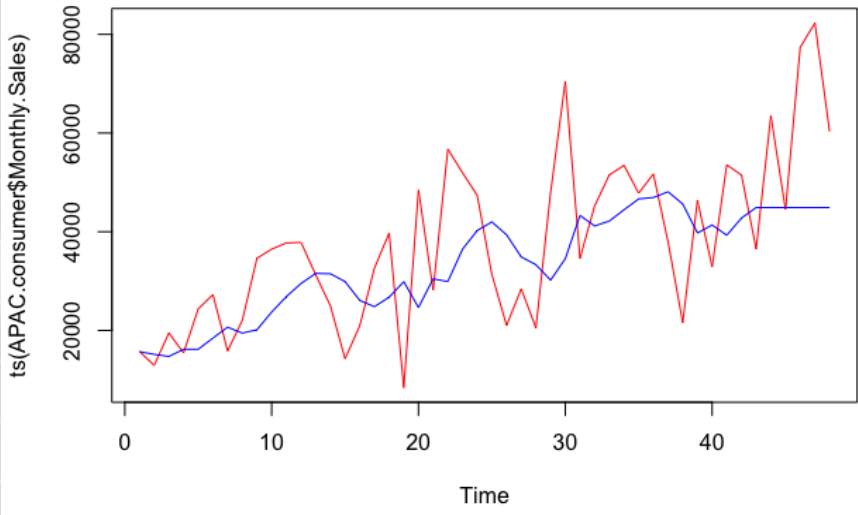


APAC Consumer Sales: Curve fitting & Residue Analysis -

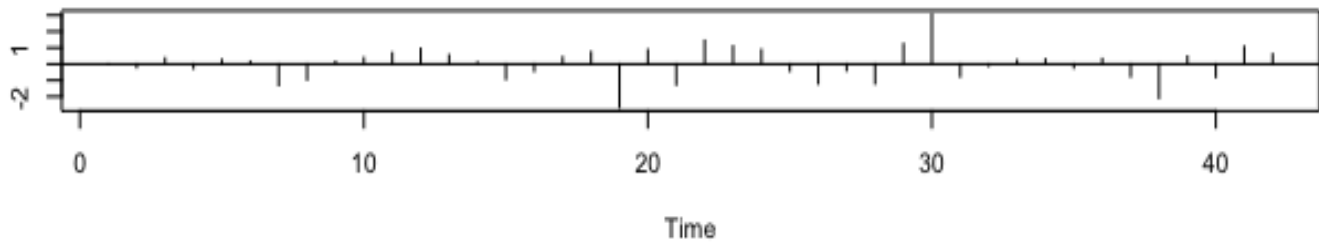
Classical Decomposition (MAPE - 22.94291)



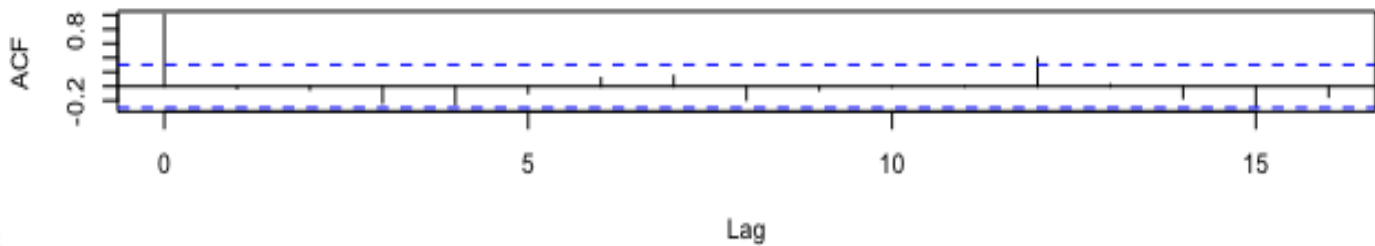
Auto Arima (MAPE - 27.68952)



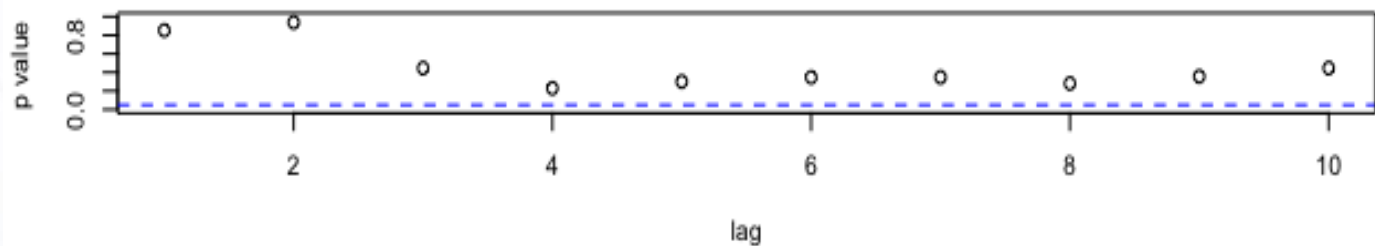
Standardized Residuals



ACF of Residuals

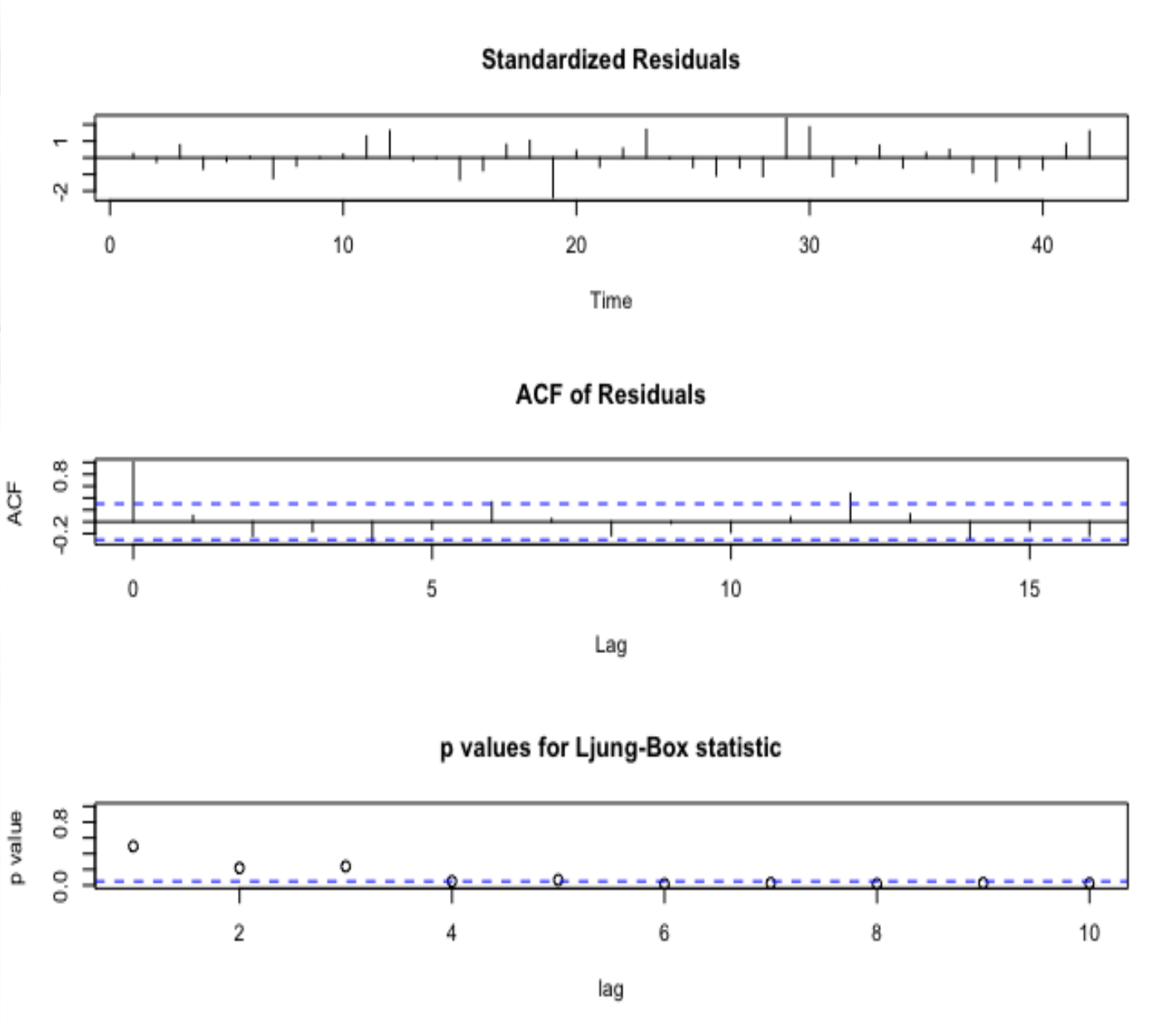
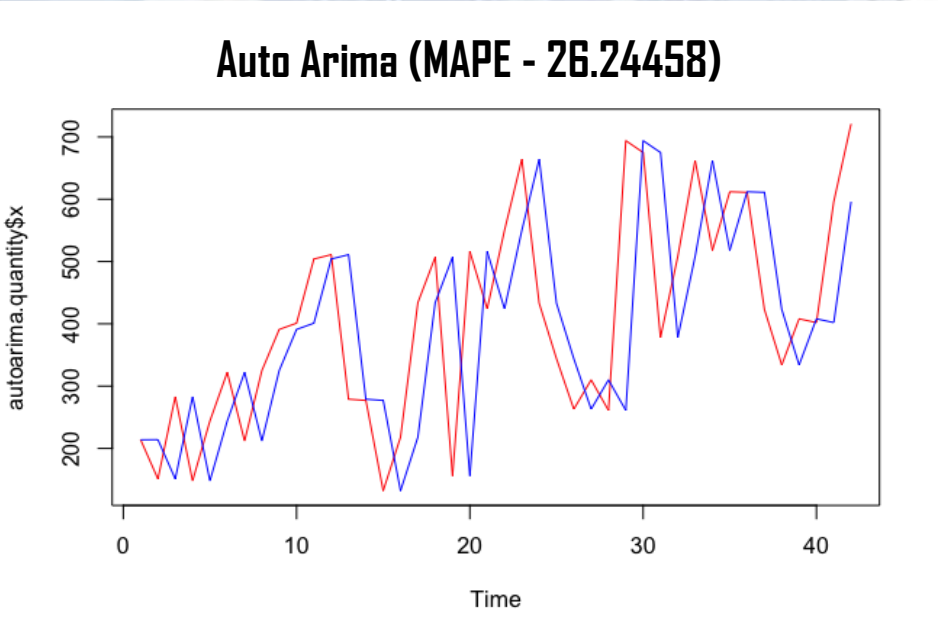
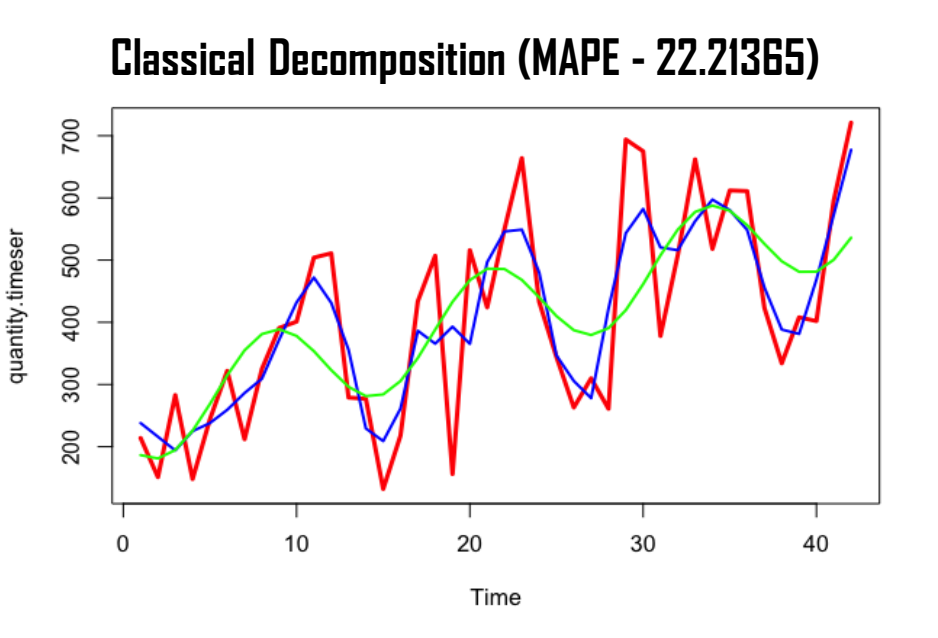


p values for Ljung-Box statistic



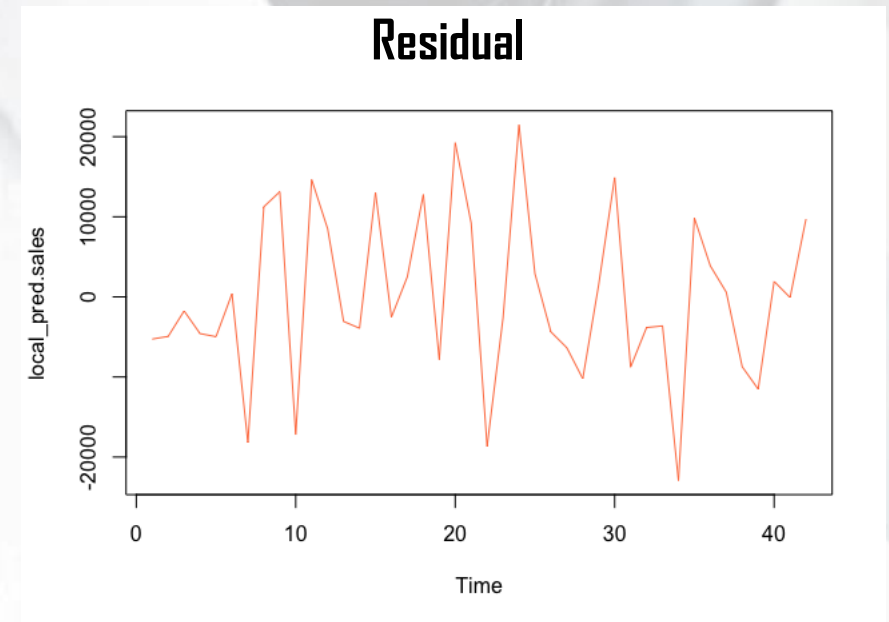
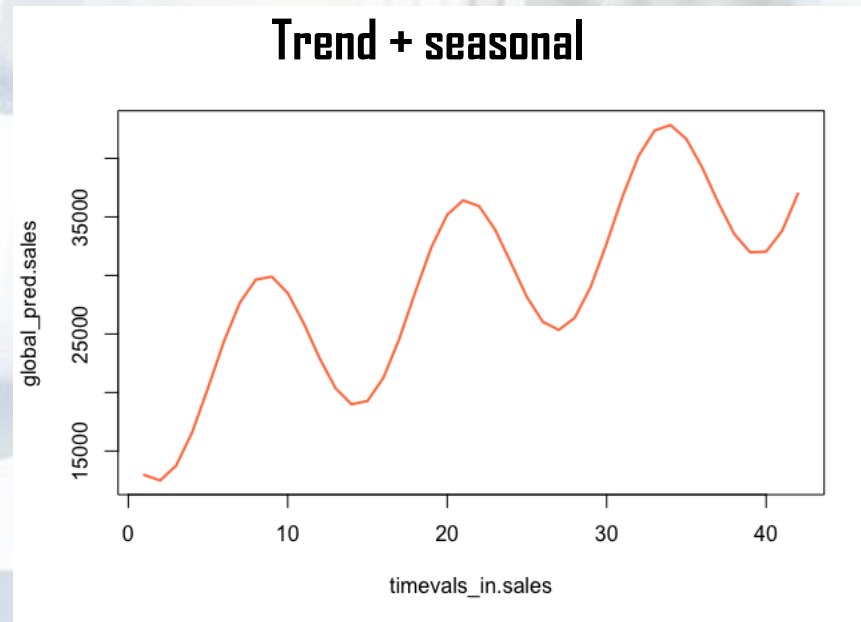
It is clearly indicating White-Noise

APAC Consumer Quantity: Curve fitting & Residue Analysis

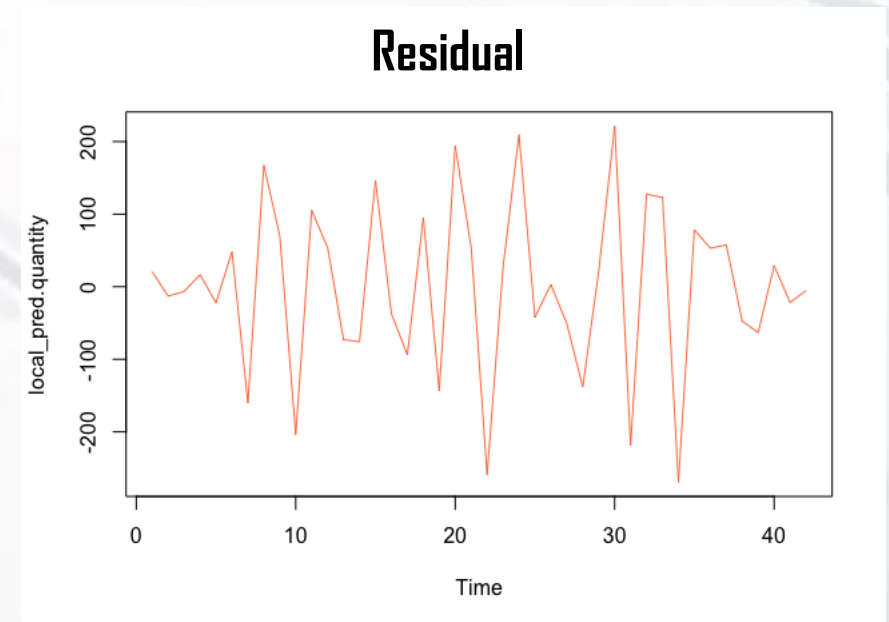
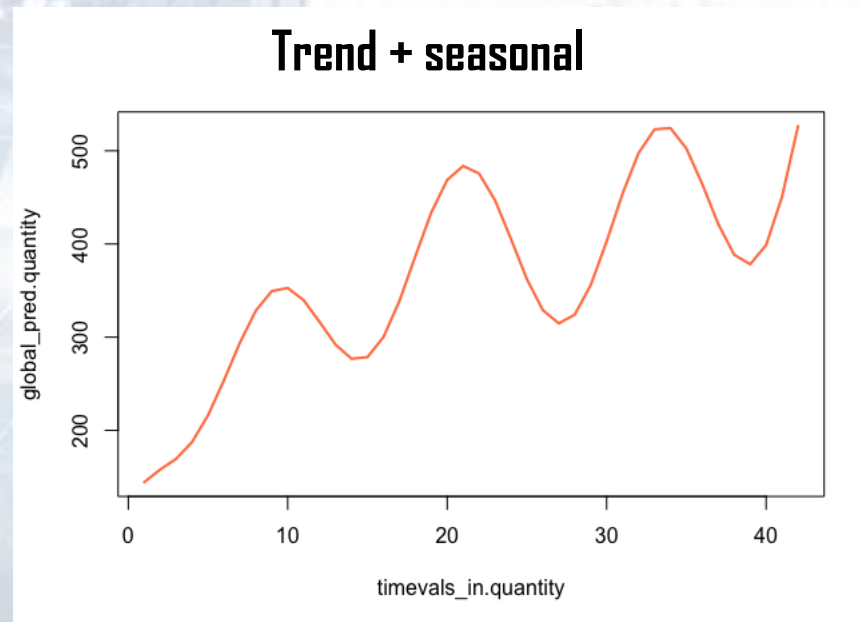


It is clearly indicating White-Noise

EU Consumer Sales Decomposition Chart

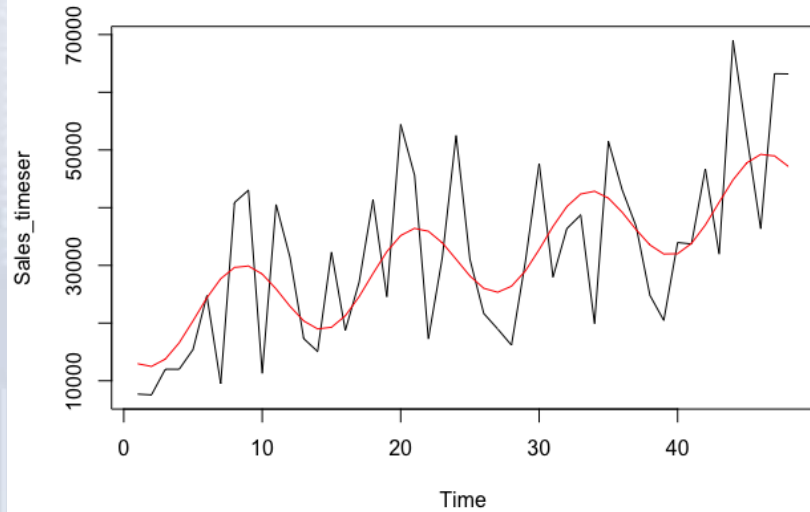


EU Consumer Quantity Decomposition Chart

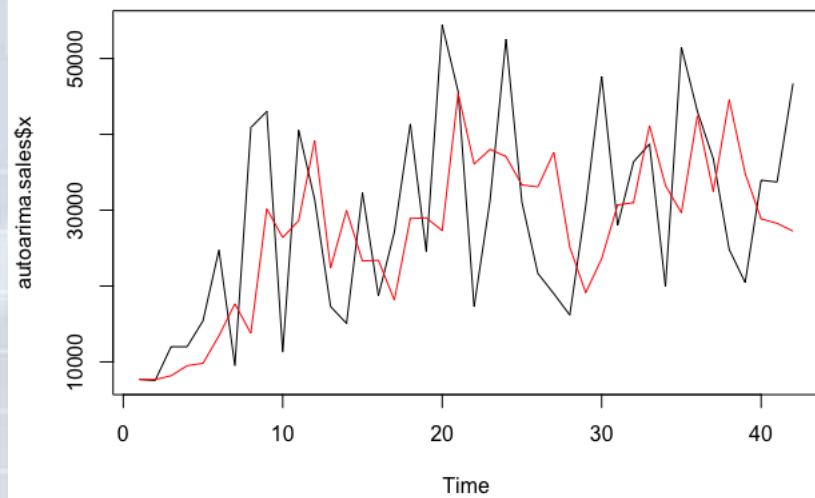


EU Consumer Sales: Curve fitting & Residue Analysis -

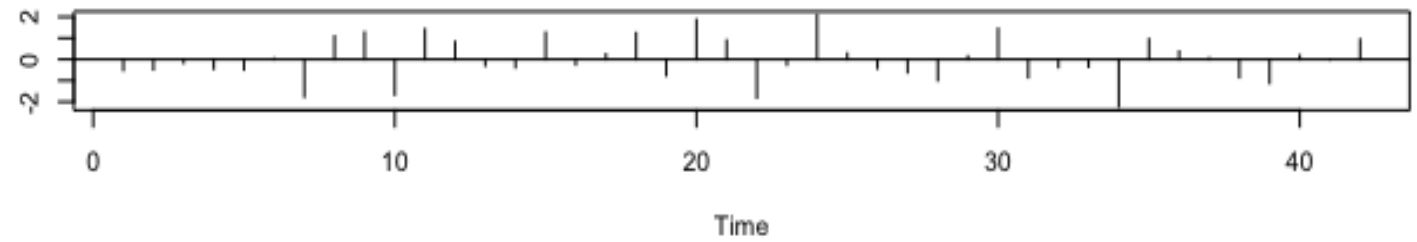
Classical Decomposition (MAPE - 25.85999)



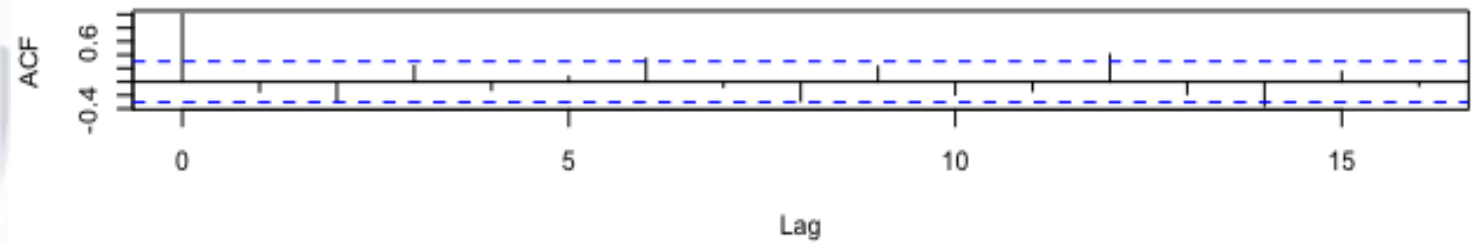
Auto Arima (MAPE - 28.9226)



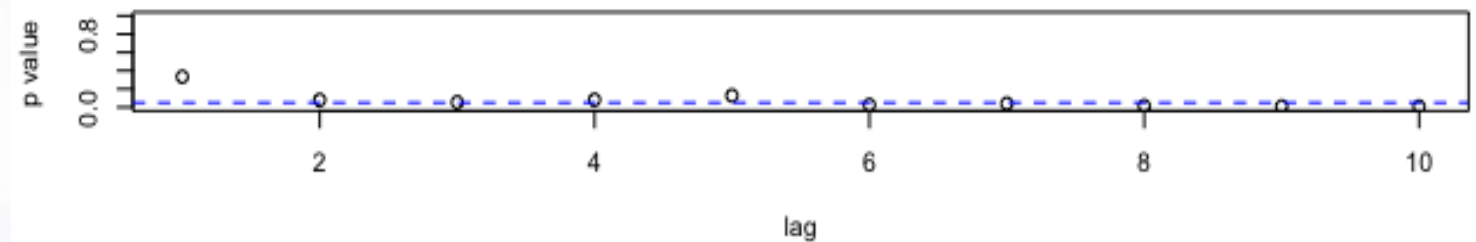
Standardized Residuals



ACF of Residuals



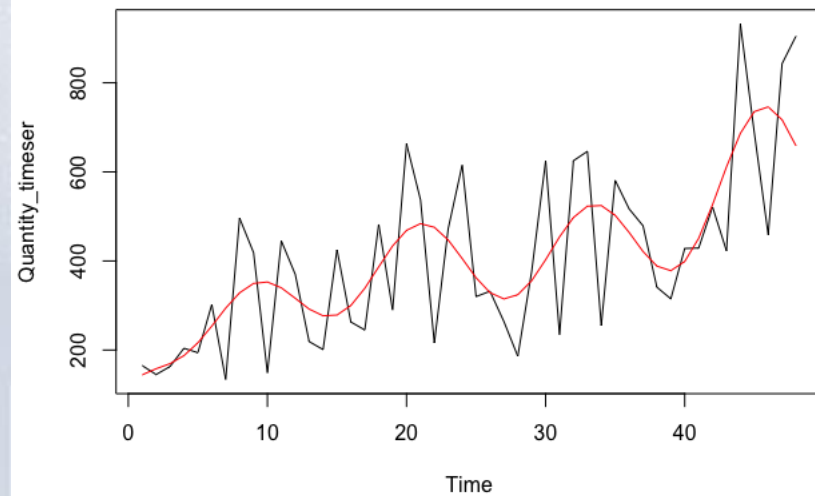
p values for Ljung-Box statistic



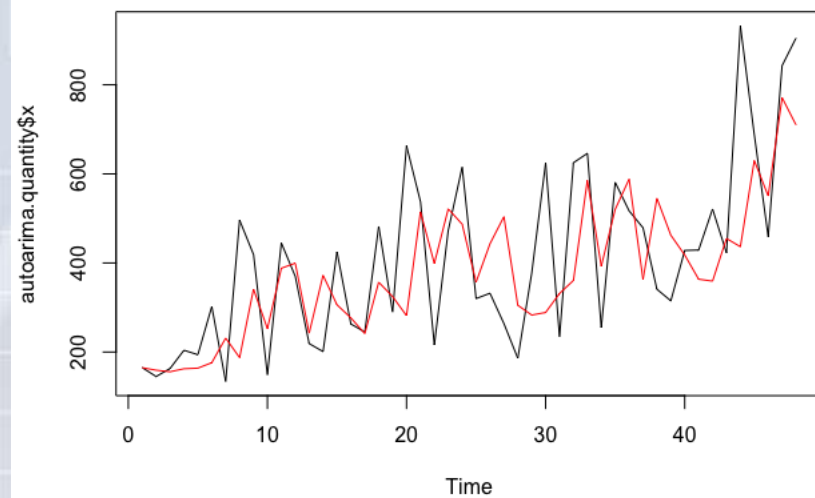
It is clearly indicating White-Noise

EU Consumer Quantity: Curve fitting & Residue Analysis -

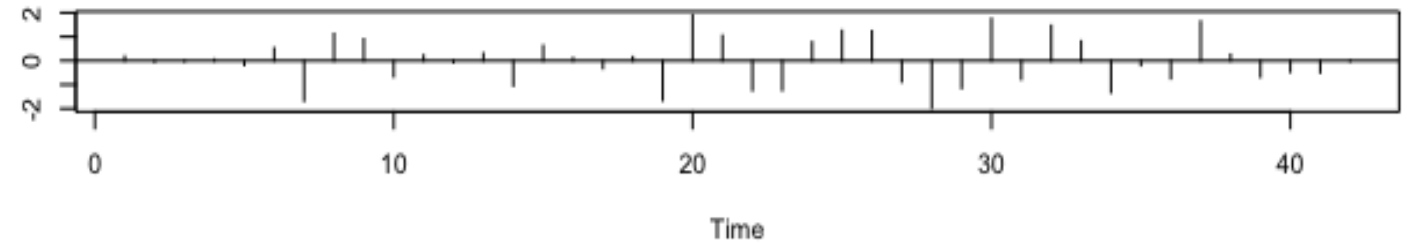
Classical Decomposition (MAPE - 30.39741)



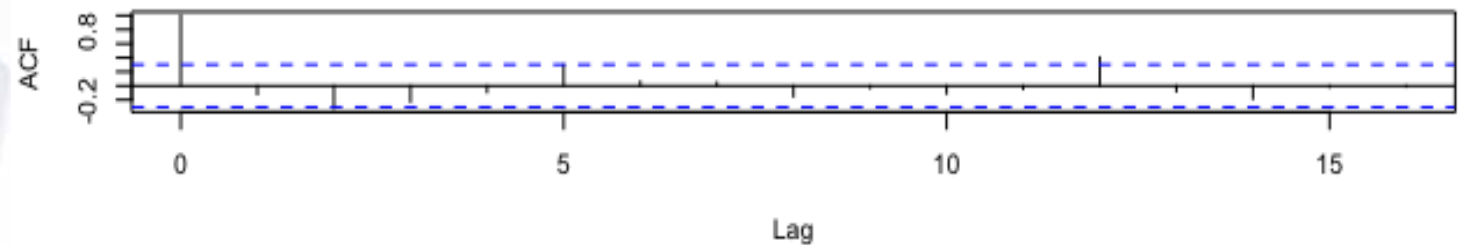
Auto Arima (MAPE - 26.54859)



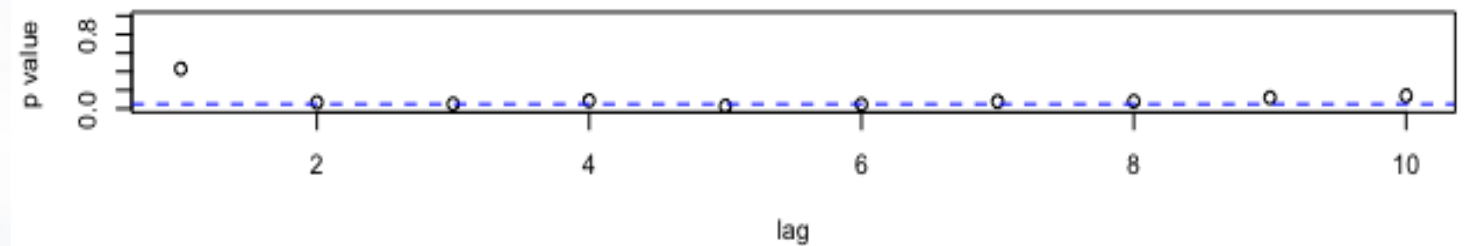
Standardized Residuals



ACF of Residuals

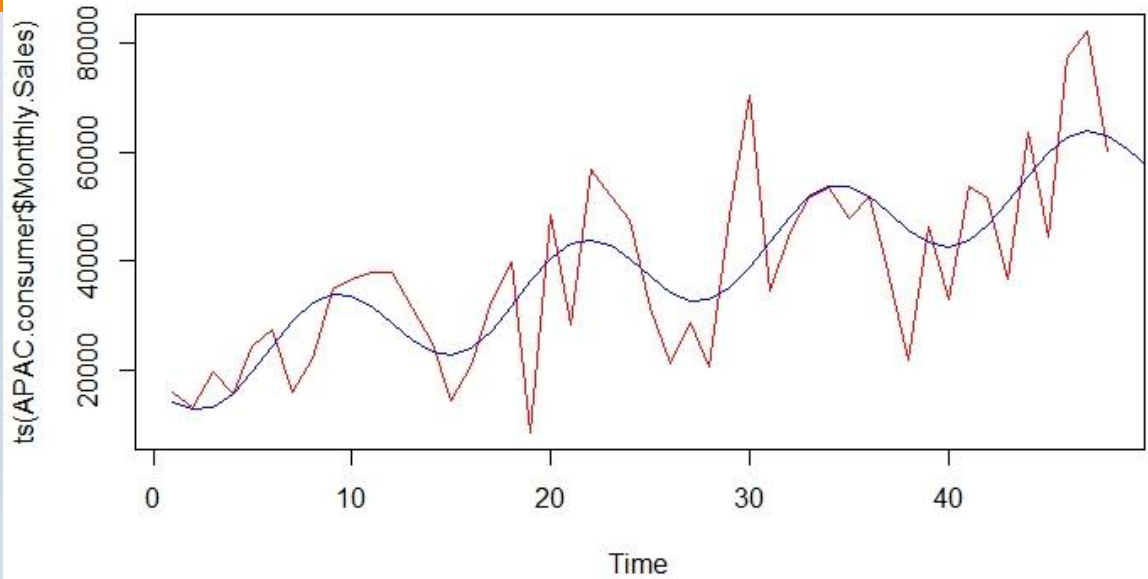


p values for Ljung-Box statistic



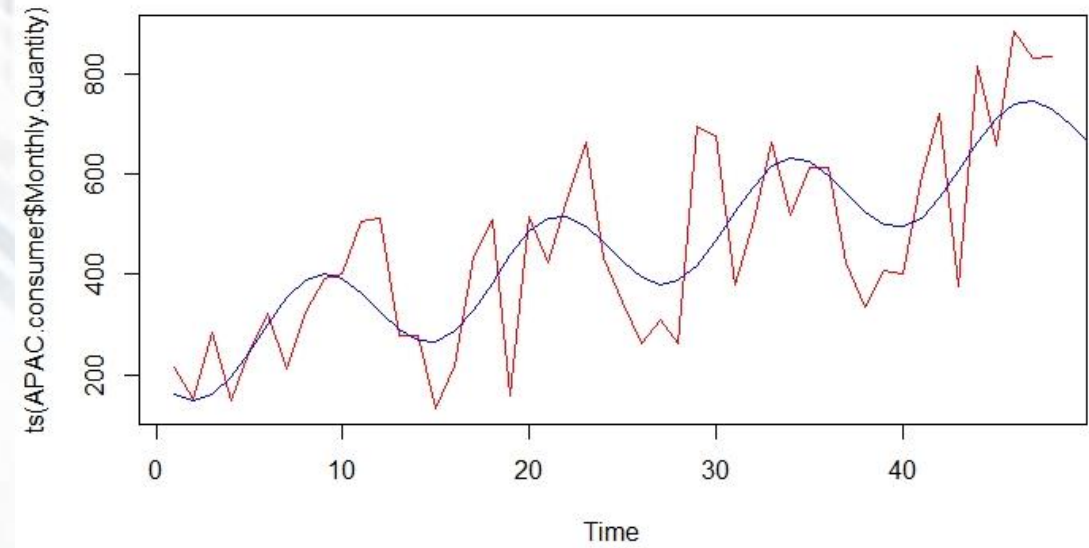
It is clearly indicating White-Noise

Forecast Plot : APAC Consumer

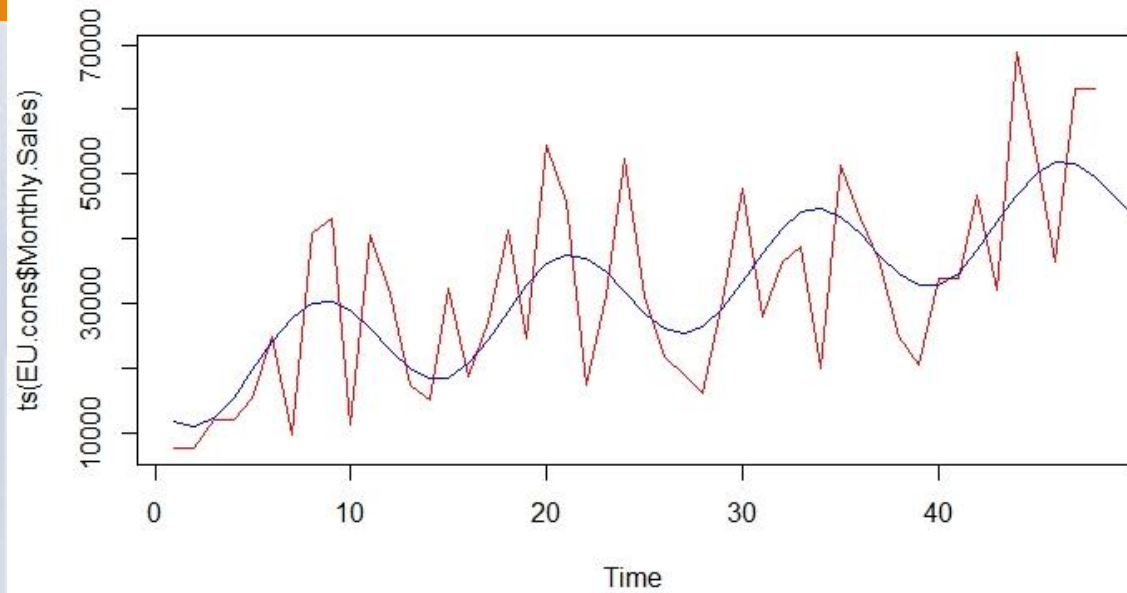


APAC Consumer Sales is likely to drop in coming 6 months

APAC Consumer Quantity is likely to drop in coming 6 months

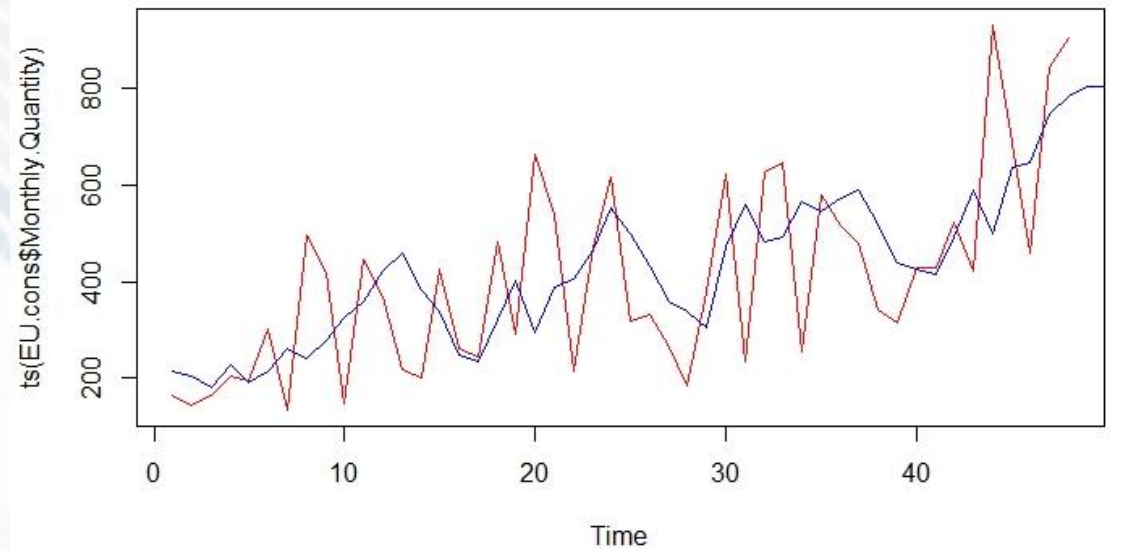


Forecast Plot : EU Consumer



EU Consumer Sales is likely to drop in coming 6 months

EU Consumer Quantity is likely to rise rapidly in coming 6 months.



Conclusions

According to the Final Forecasting, we arrive at the following observations –

- ❑ 2 most profitable market segments as APAC Consumer and EU Consumer.
- ❑ Sales across the Market Segments have seasonal behavior
- ❑ We created total 8 forecasting models for top 2 segments for forecasting future 6 months sales & quantity
- ❑ Below is Forecast Summary:
 - APAC Consumer Sales is likely to drop in next 6 months.
 - APAC Consumer Quantity is also likely to drop in coming 6 months.
 - EU Consumer Sales is also likely to drop in coming 6 months.
 - EU Consumer Quantity is likely to rise rapidly in coming 6 months.