Retail-Giant Sales Forecasting
CASE STUDY

1. Swati Kamat
2. Rounak Gurjar
3. Suraj Talreja
4. Saurabh Kumar Singh

Objective/Goals

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

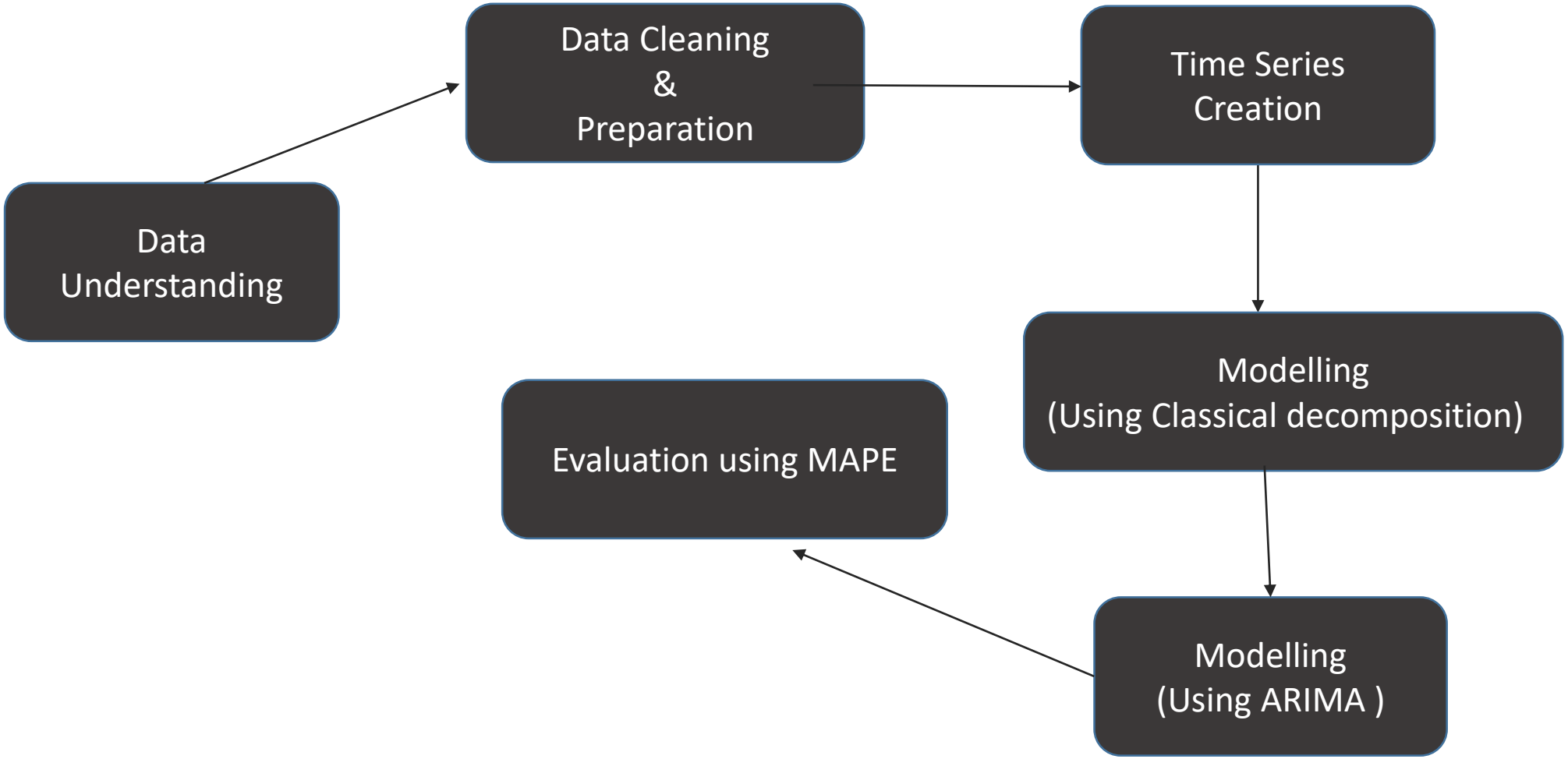
Objective

- Finalize the plan for the next 6 months that would help to manage the revenue and inventory accordingly.
- Find the 2 most profitable and consistent buckets of 21 buckets from 7 different market segments and in 3 major categories.
- After finding the 2 buckets we need to forecast the sales and quantity for next 6 months

Strategy

- Perform Data understanding, cleaning & Preparation
- Perform Time Series Model building
- Perform Model evaluation, Validate forecast using out of the sample figures.

Problem Solving Methodology



Data Cleaning & Preparation

- Missing values for Postal Code column only
  - As this is insignificant column, **not handling missing values**. Rather discarding this column.

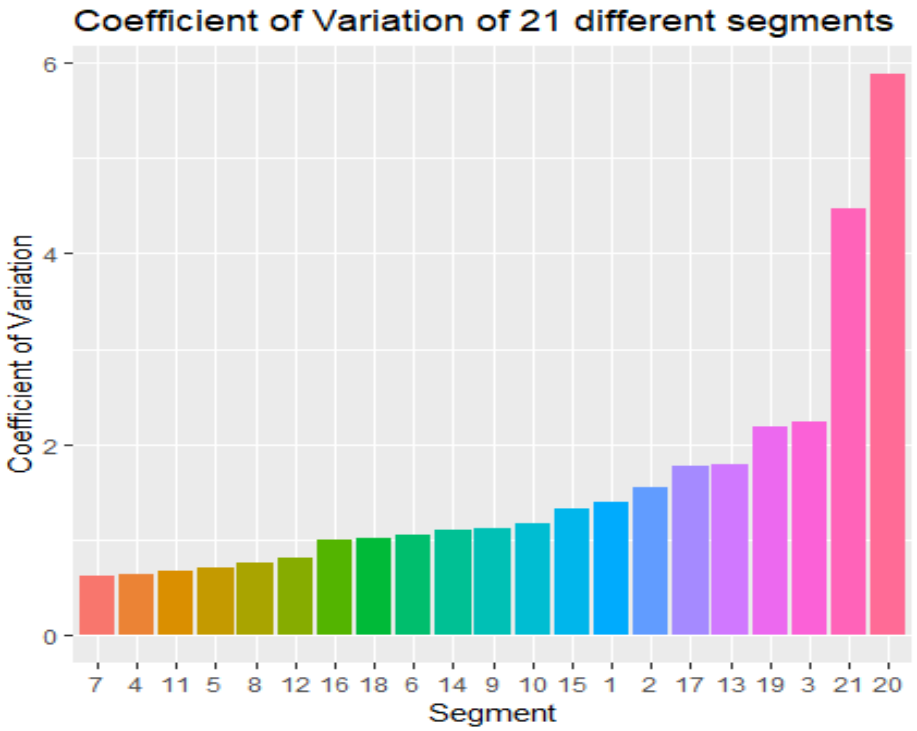
- There are 24 attributes in the dataset
  - Considering columns of interest, i.e. **Market, Order Date, Segment, Market, Sales, Quantity and Profit**

- There is 48 months data starting from 01-01-2011
  - Date formatting and converting to Months, so as to create Timeseries

- No Outlier treatment for Sales, Quantity and Profit
  - As we wish to forecast for Time Series, avoiding outlier treatment.

Data Understanding

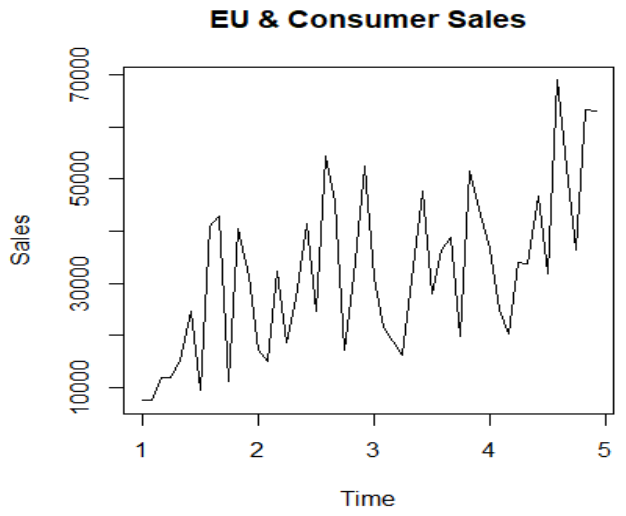
- Transaction level data for order details is present in the dataset
- There are 7 Markets and 3 Segments
  - Converted these into 21 market segments
- Calculated Total Profit, Average Profit and COV for Profit.
  - Based on these factors determined 2 most profitable and consistently profitable segments

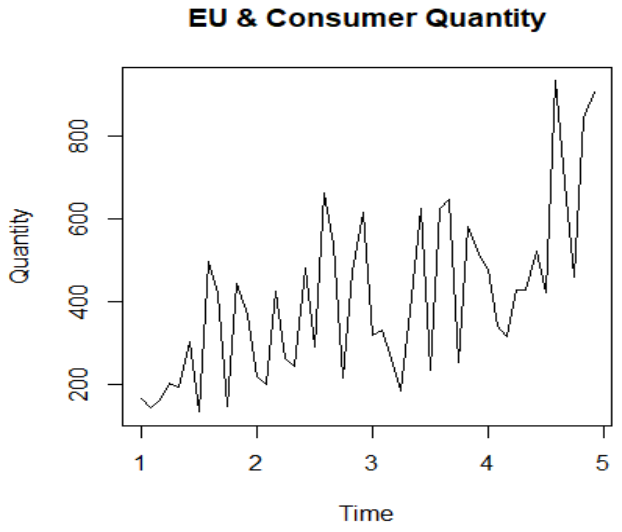


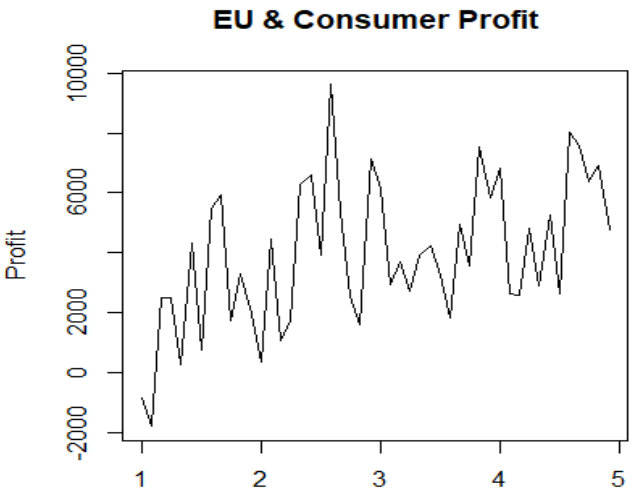
segment	cov	total_profit	average_profit	SegmentName
7	0.62	188687.71	3930.99	EU_Consumer
4	0.63	222817.56	4642.03	APAC_Consumer
11	0.66	120632.93	2513.19	EMEA_Corporate
5	0.70	129737.23	2702.86	APAC_Corporate
8	0.76	123393.98	2570.71	EU_Corporate
12	0.81	57875.42	1205.74	EMEA_HomeOffice
16	1.00	91979.13	1916.23	LATAM_Consumer
18	1.01	134119.21	2794.15	LATAM_HomeOffice
6	1.05	83445.25	1738.44	APAC_HomeOffice
14	1.10	60298.68	1256.22	Africa_Corporate
9	1.12	60748.05	1265.58	EU_HomeOffice
10	1.18	43135.13	898.65	EMEA_Consumer
15	1.32	47772.10	995.25	Africa_HomeOffice
1	1.40	9677.70	230.42	US_Consumer
2	1.55	5036.46	148.13	US_Corporate
17	1.78	20686.97	430.98	LATAM_Corporate
13	1.79	20412.57	425.26	Africa_Consumer
19	2.19	25532.57	531.93	Canada_Consumer
3	2.24	3103.23	124.13	US_HomeOffice
21	4.47	12499.13	260.40	Canada_HomeOffice
20	5.88	5866.26	122.21	Canada_Corporate

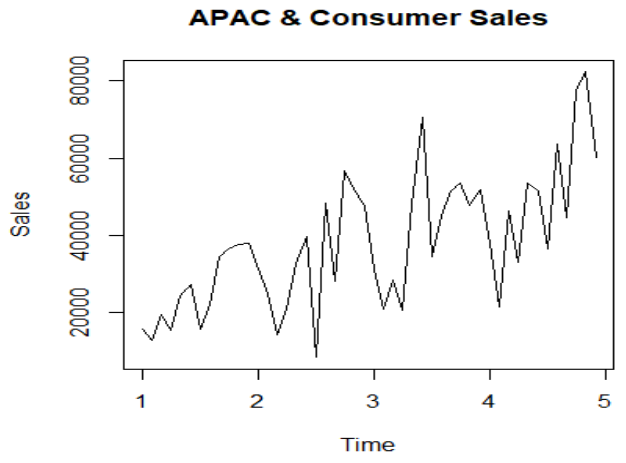
Time Series for EU and APAC regions

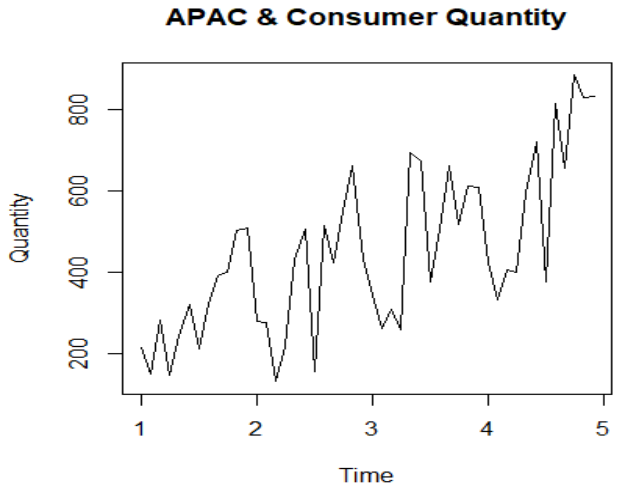
[Sales, Demand/Quantity and Profit]

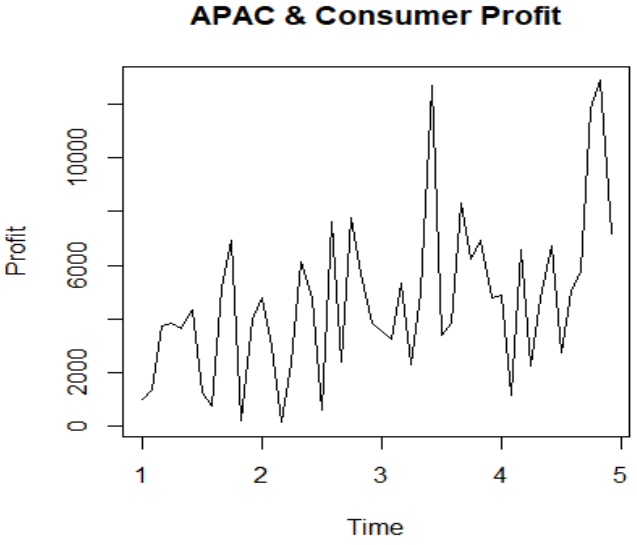






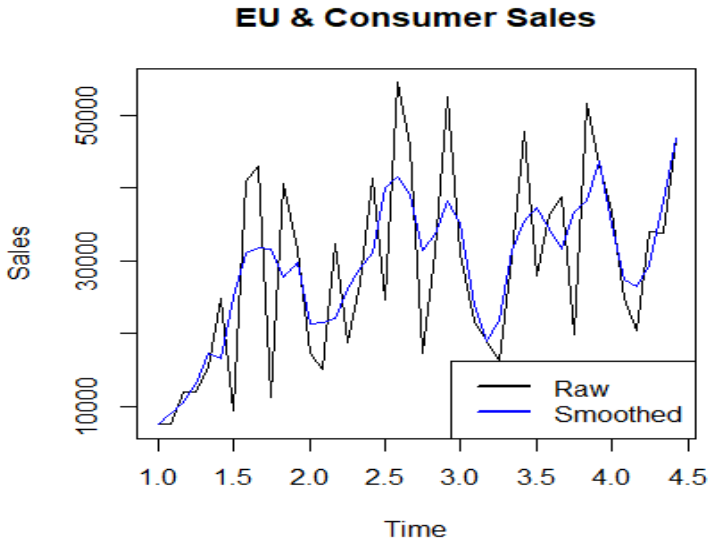


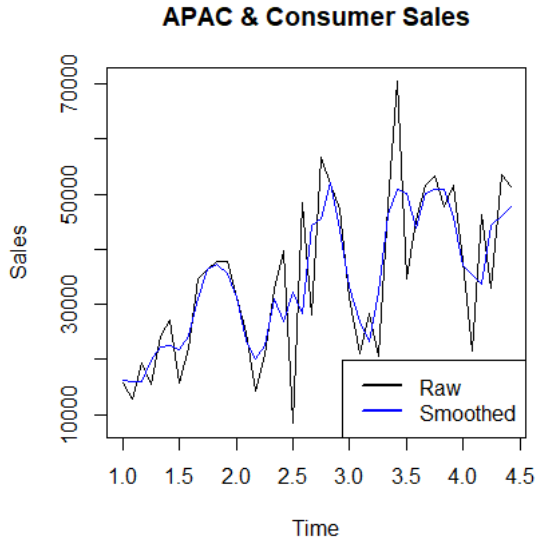


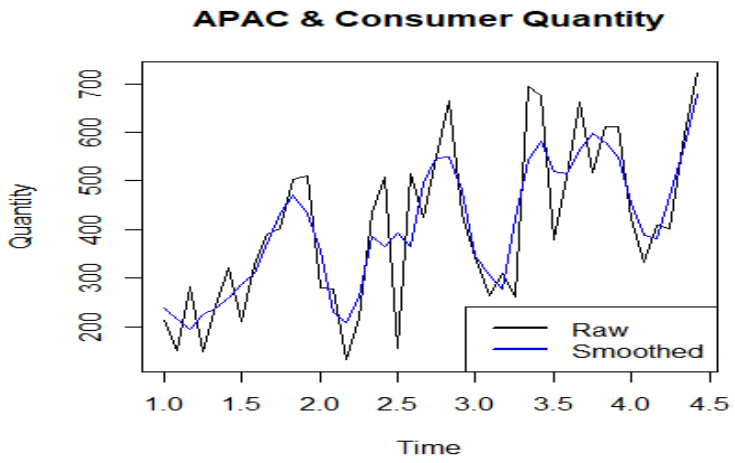


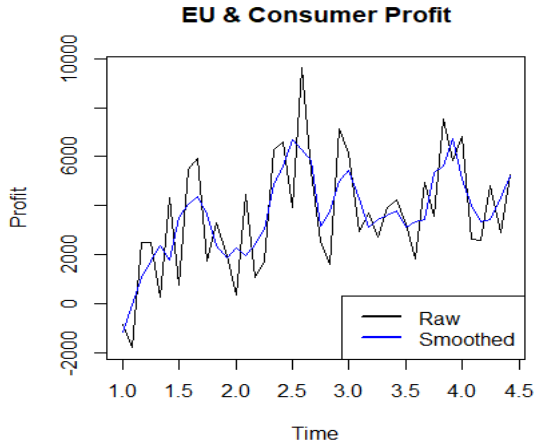
Model building

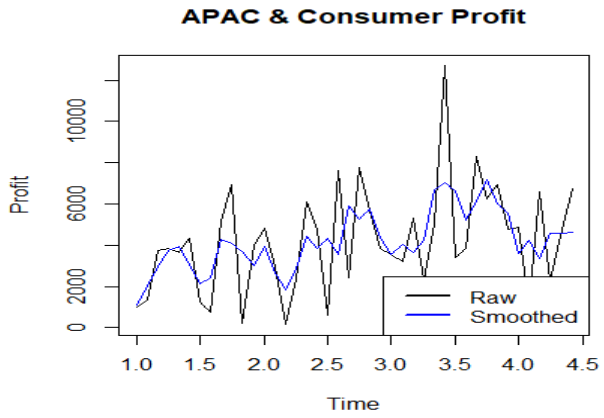
Considering 42 months data for training and the last 6 months for Out of Sample test.



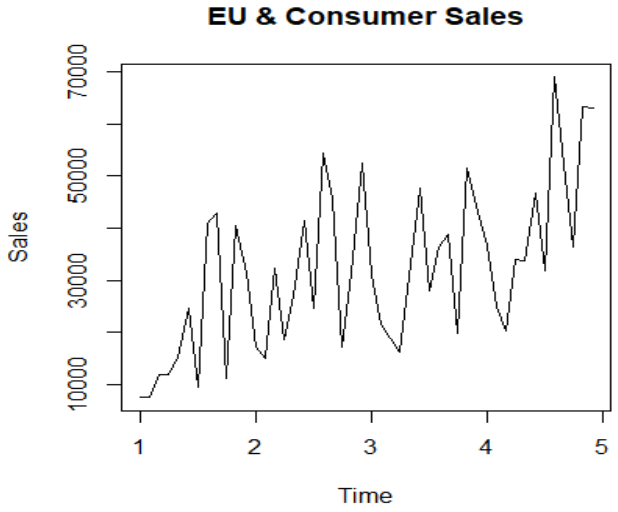


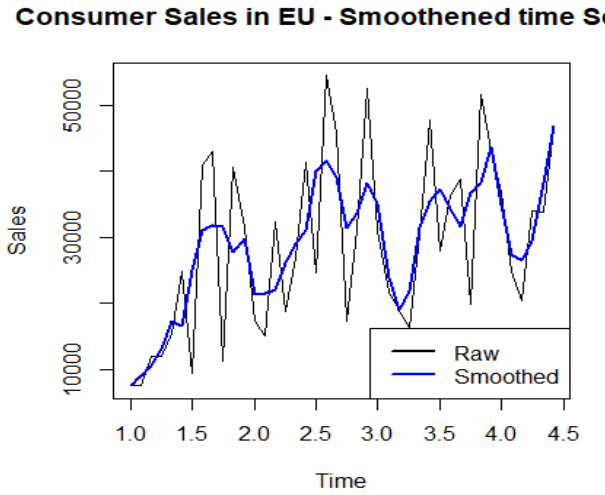


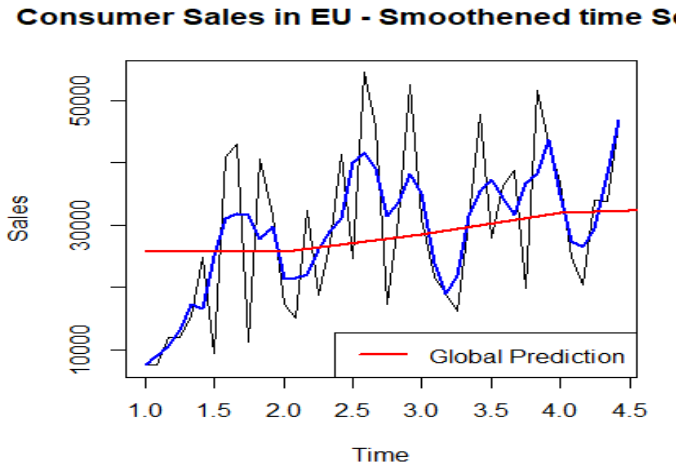


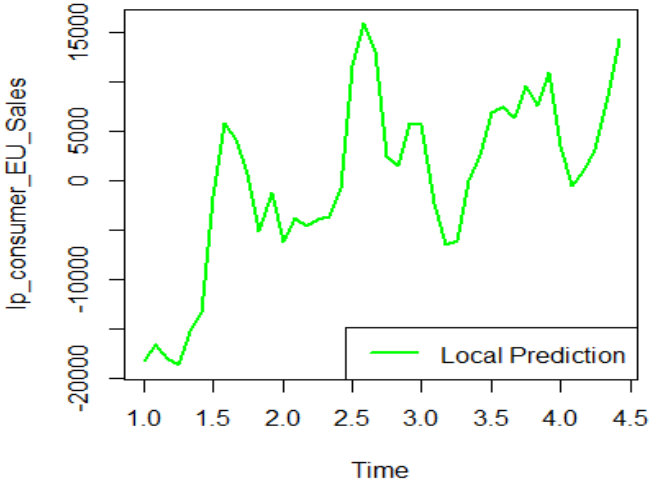


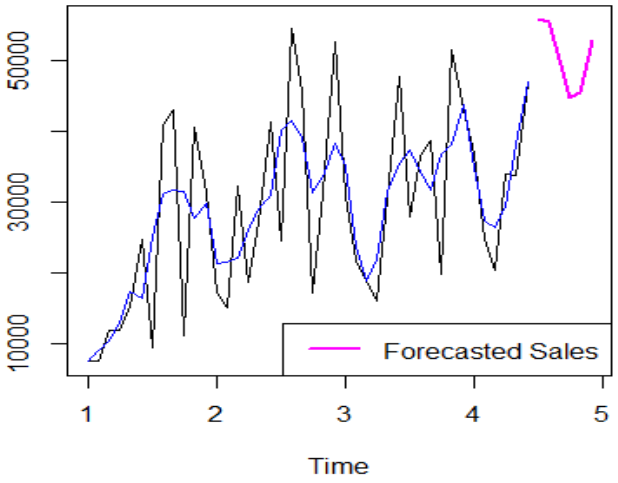
Consumer EU – Sales
[Classical Decomposition]

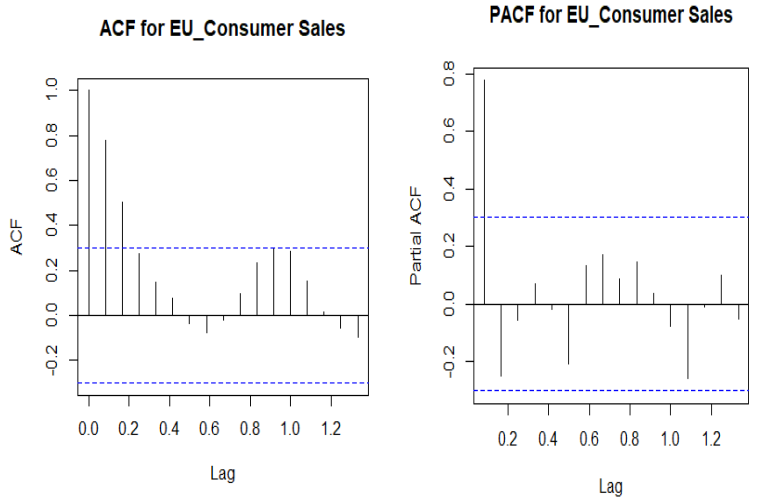






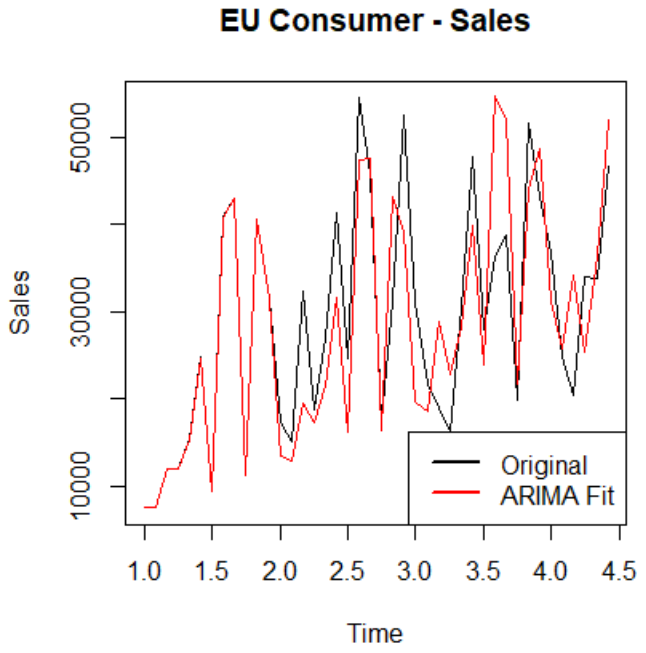


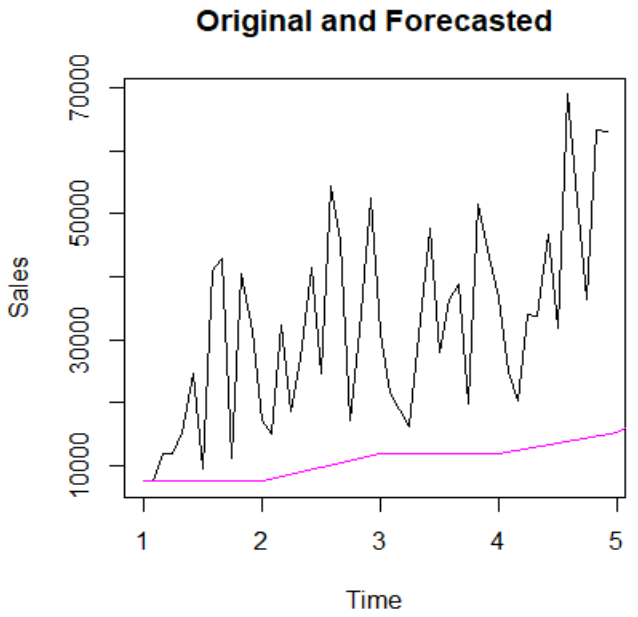




Residual Series
(No Autoregressive behavior in residual time series based on KPSS test and ADF test)

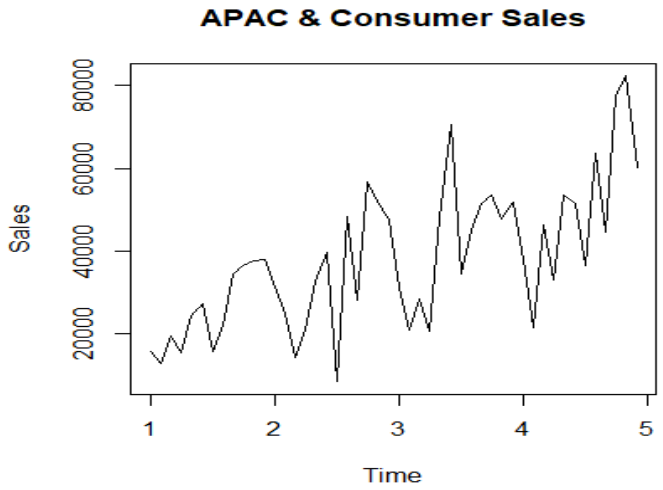
Consumer EU – Sales
[ARIMA method]

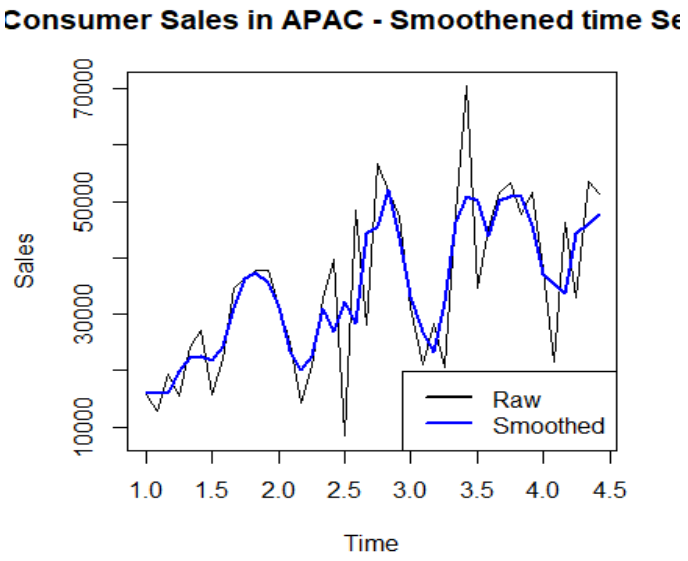


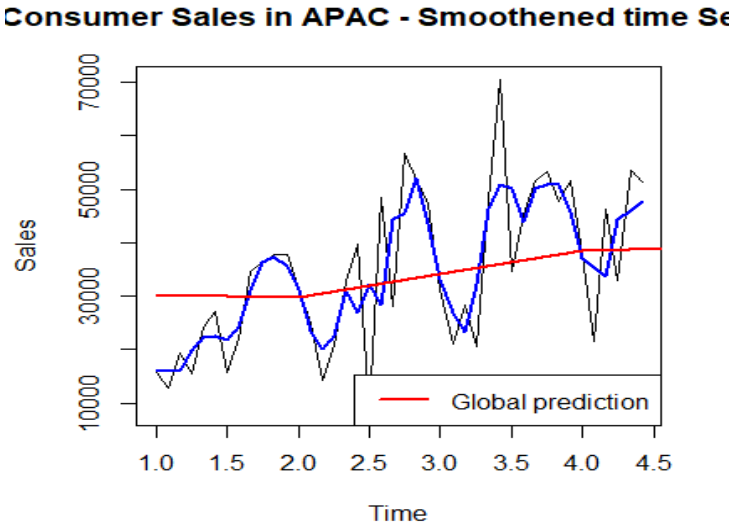


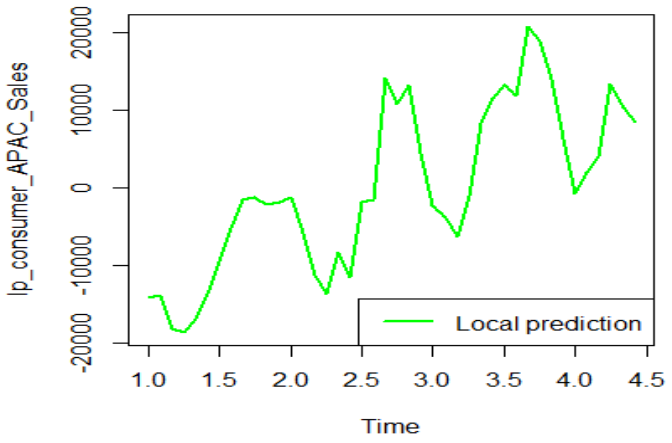
Forecasting Method	MAPE
Classical Decomposition	40.43501
ARIMA	15.7522

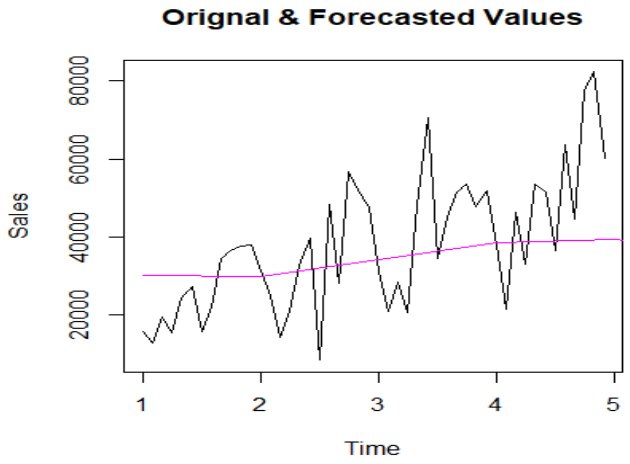
Consumer APAC – Sales
[Classical Decomposition]

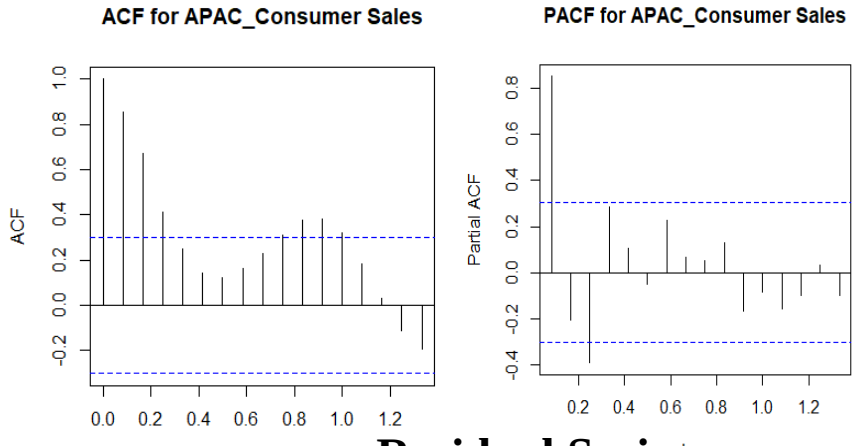






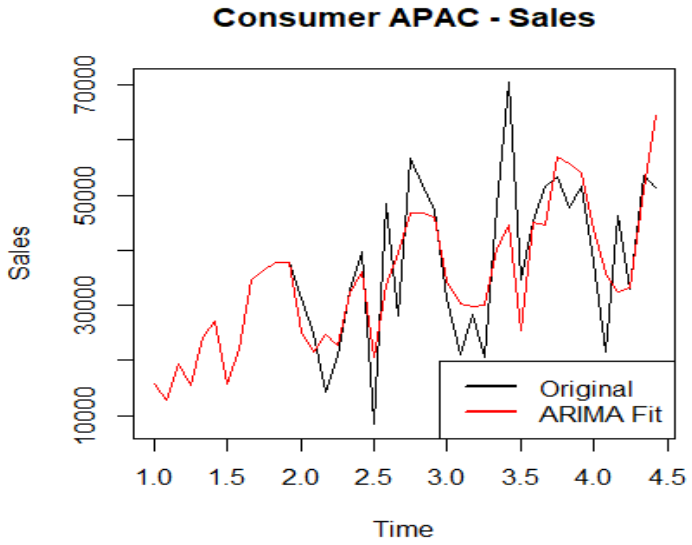


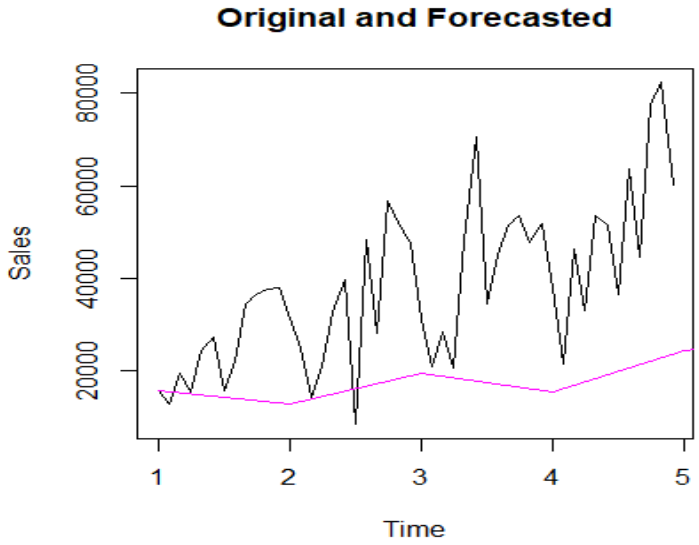




Residual Series
(No Autoregressive behavior in residual time series based on KPSS test and ADF test)

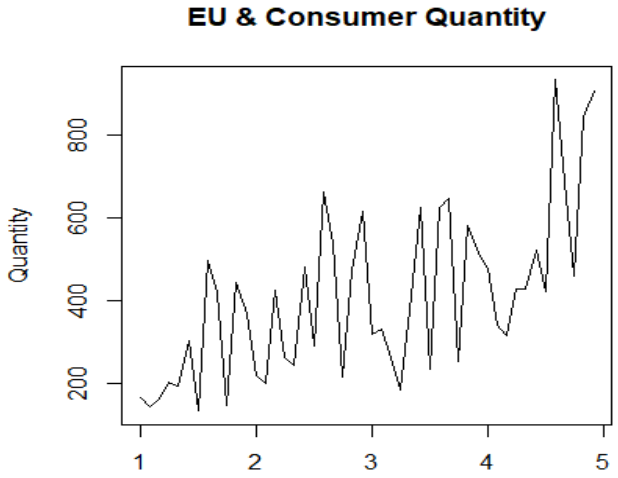
Consumer APAC – Sales
[ARIMA method]

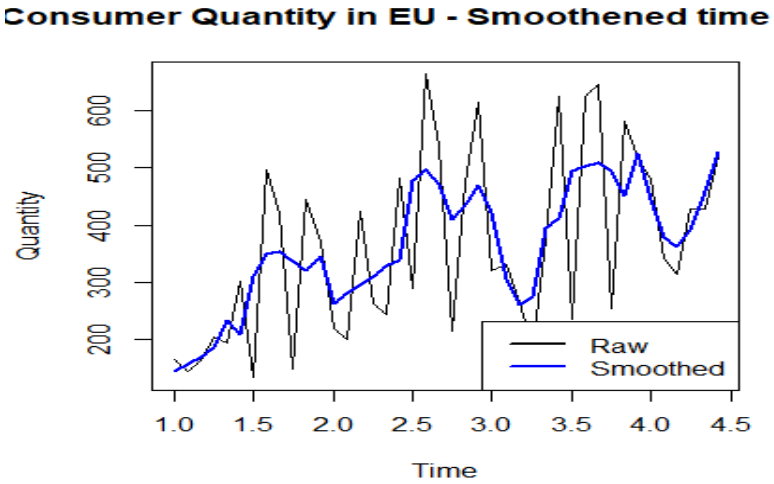


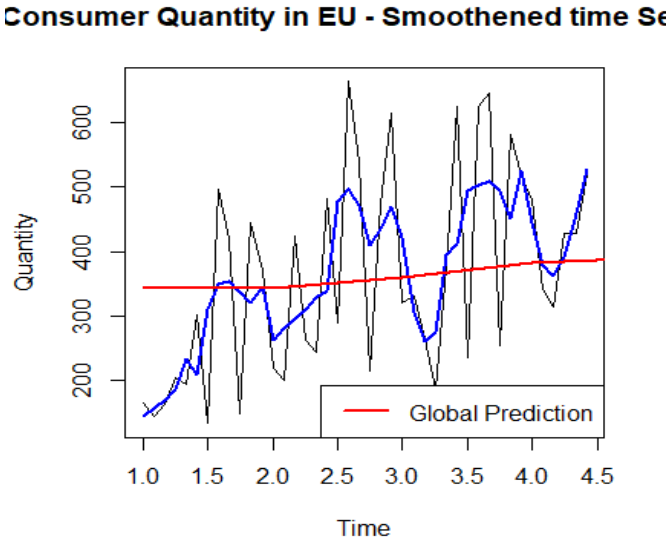


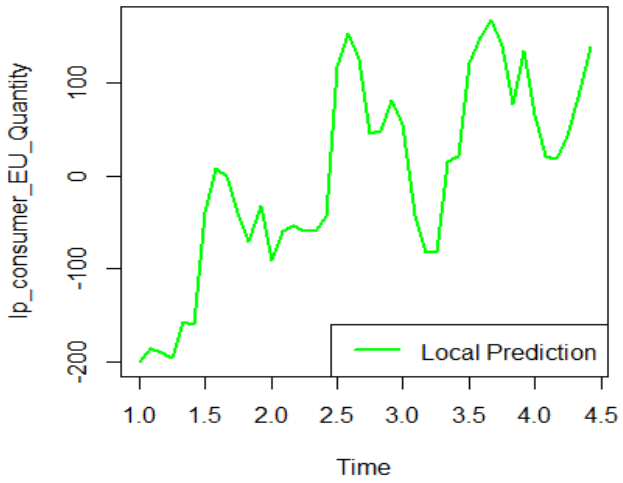
Forecasting Method	MAPE
Classical Decomposition	39.9417
ARIMA	11.776

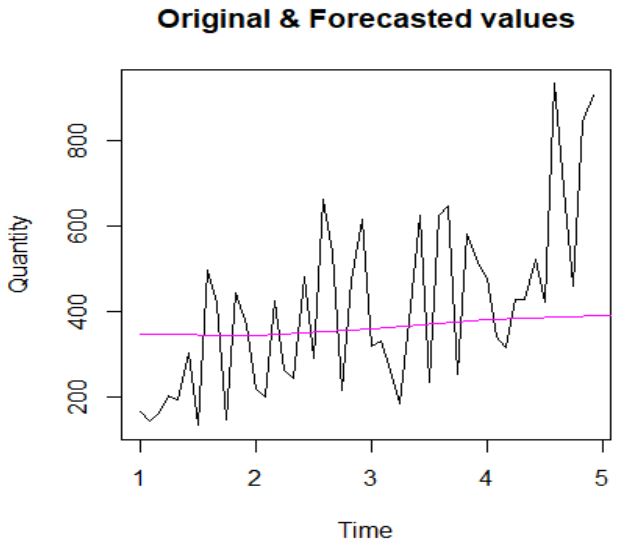
Consumer EU – Quantity
[Classical Decomposition]

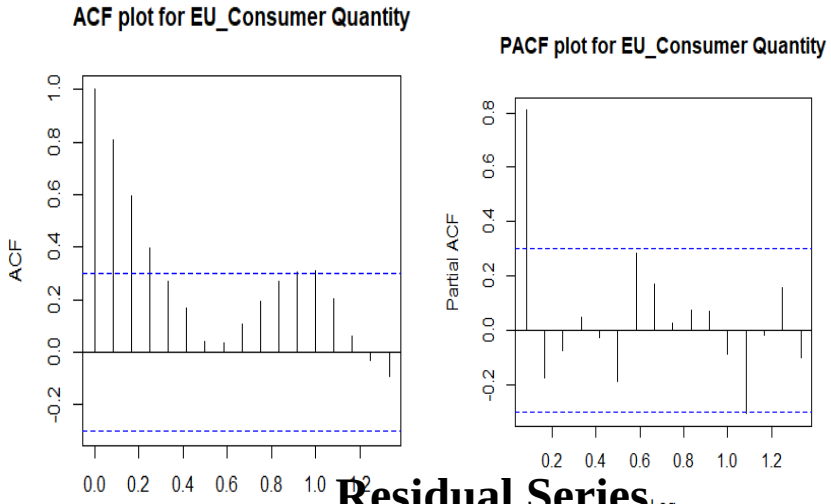






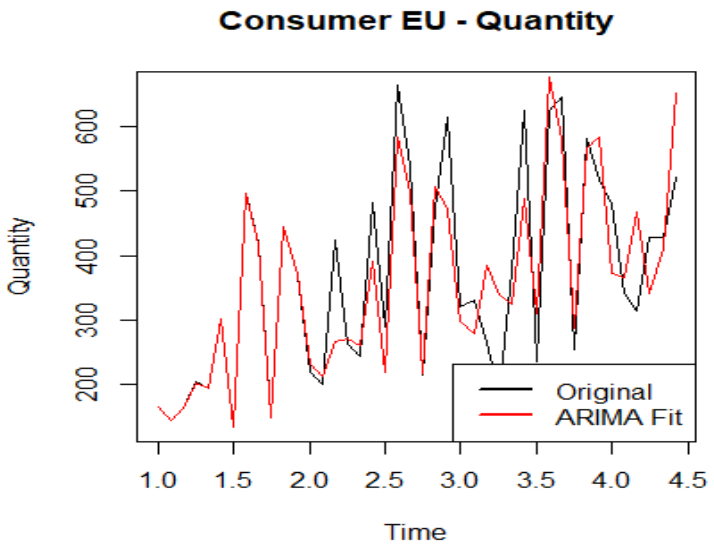


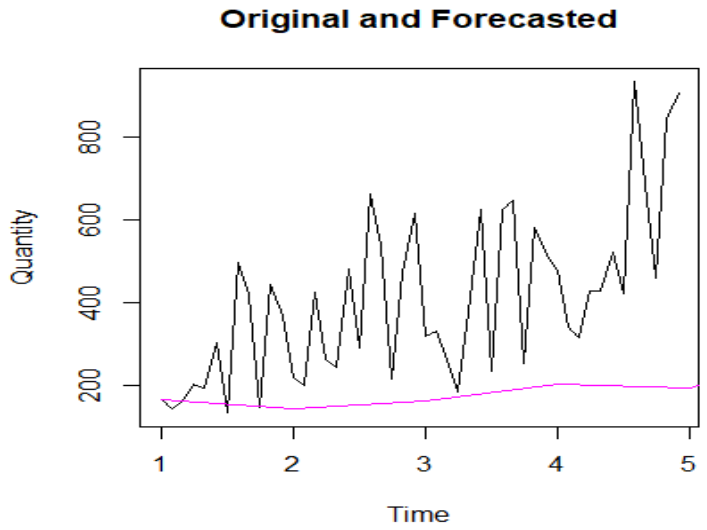




Residual Series
(No Autoregressive behavior in residual time series based on KPSS test and ADF test)

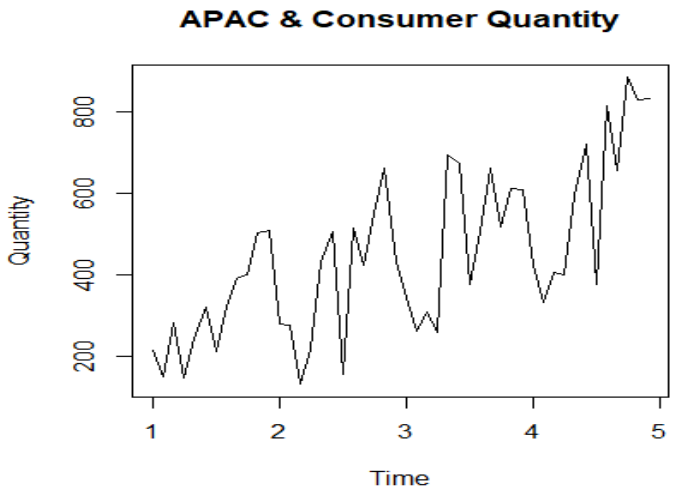
Consumer EU – Quantity
[ARIMA method]

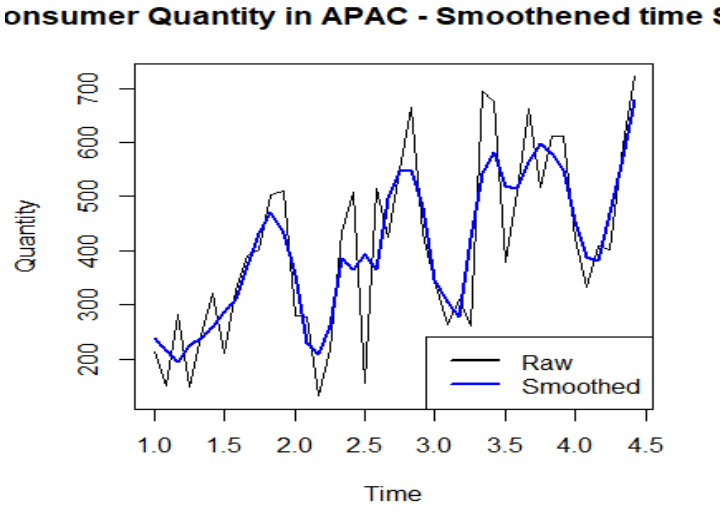


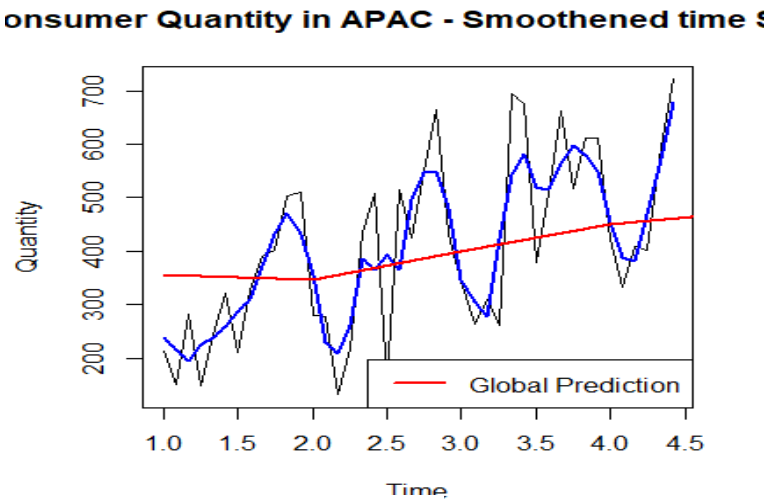


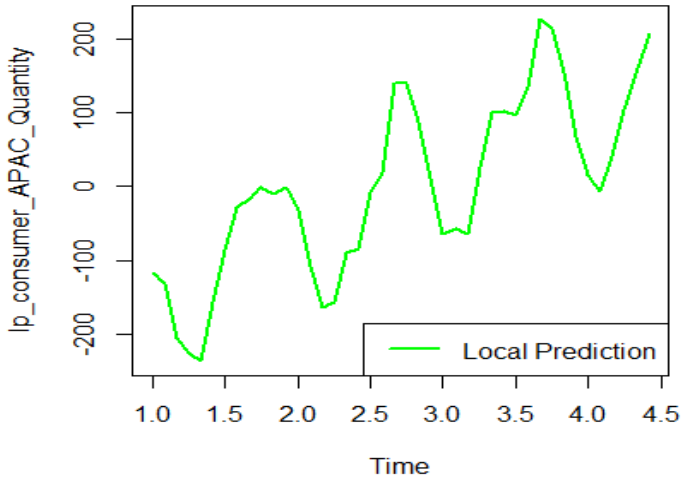
Forecasting Method	MAPE
Classical Decomposition	43.25895
ARIMA	17.61852

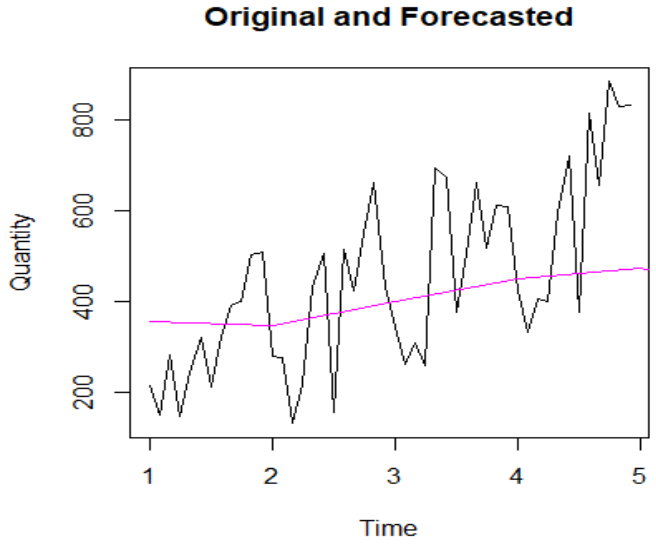
Consumer APAC - Quantity
[Classical Decomposition]

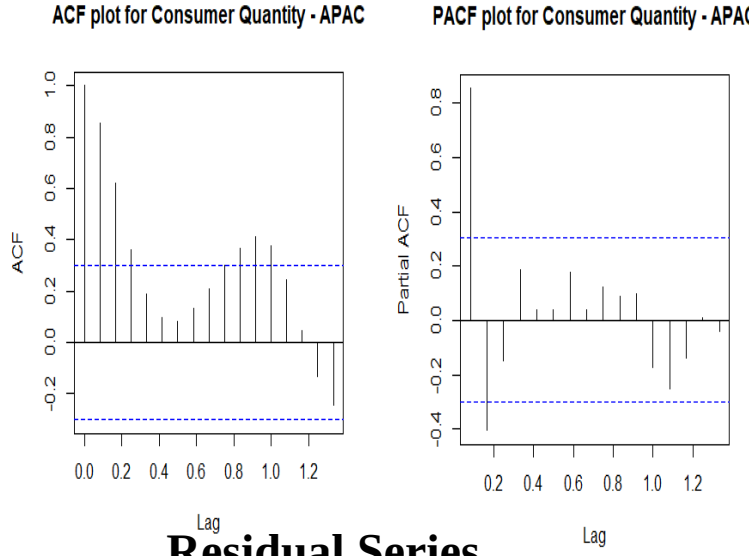






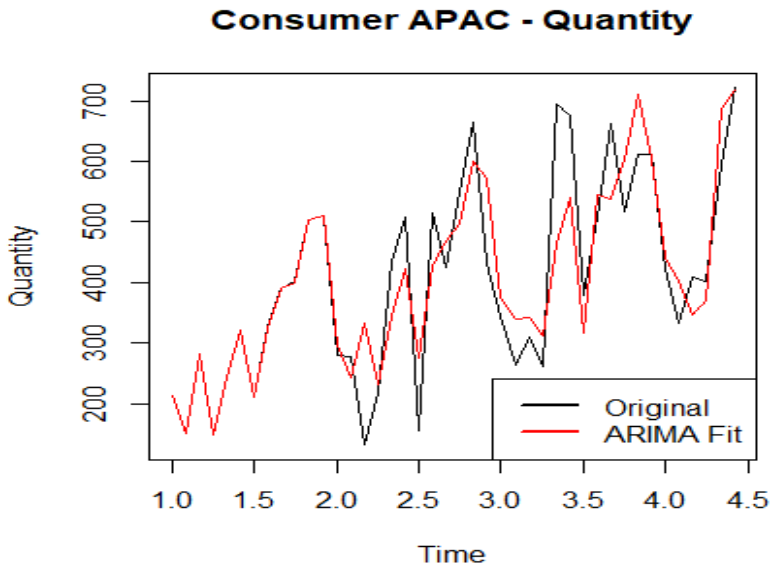


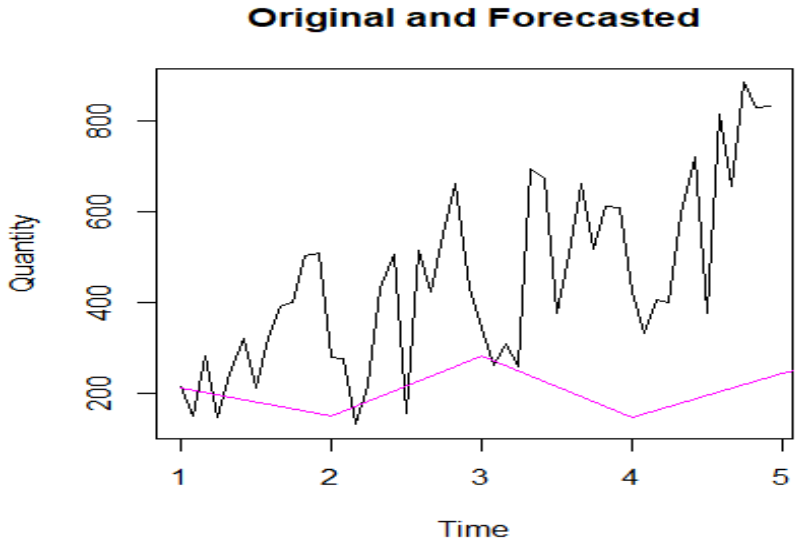




Residual Series
(No Autoregressive behavior in residual time series based on KPSS test and ADF test)

Consumer APAC - Quantity





Forecasting Method	MAPE
Classical Decomposition	46.02607
ARIMA	13.36779