

A Case Study
on
Global Mart Sales & Demand
using
Time Series Analysis & Forecasting



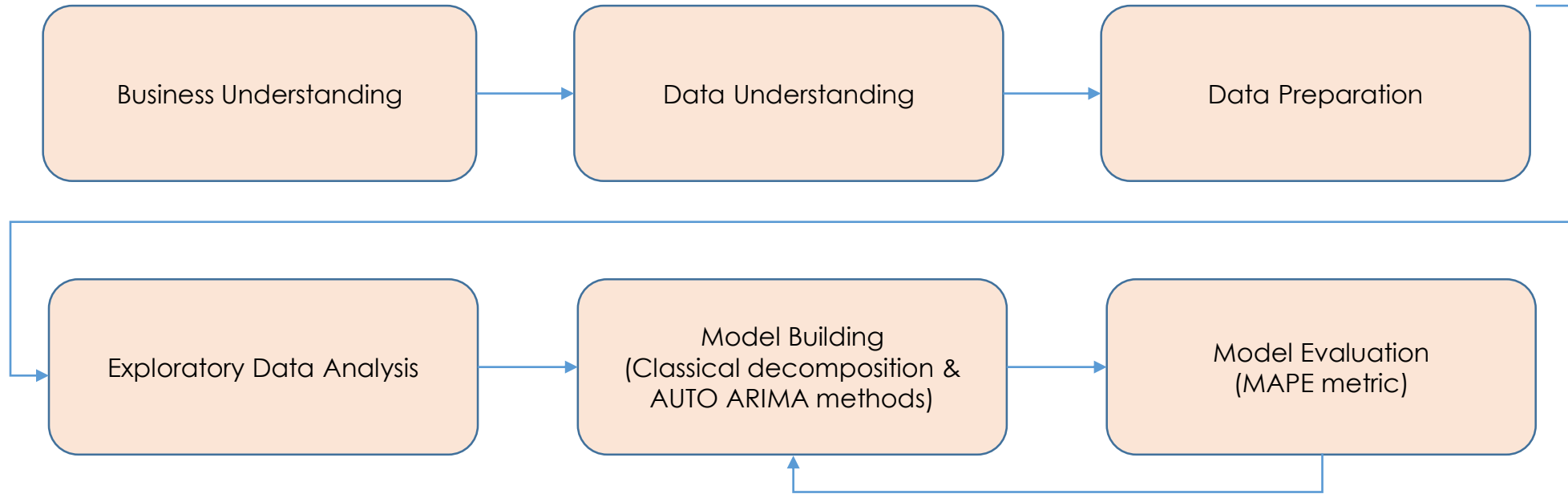
Business Understanding

Global Mart is an online store super giant having worldwide operations. The store process orders across the globe and deals with 3 major product segments - consumer, corporate & home office. The store caters to 7 different geographies (Markets).

Objective

Identify top 2 consistently profitable market segments of Global Mart and forecast Sales & Demand for next 6 months in those Market Segments using Time Series Analysis

Problem Solving Methodology



Classical decomposition method

- Visualize the time series by plotting
- Smoothen timeseries using a simple moving average of a window size of 3 & visualize by plotting
- Fit a model to the global Trend & Seasonality in the smoothened series using regression
- Remove Trend & Seasonality (the predicted values) from the series to get a stationary time series(Autoregressive + white noise)
- Check for Stationarity using ACF & PACF plots
- Model the stationary series using the auto.arima function
- Make Predictions using this model and find the residual series by subtracting the forecasted value from the actual observed value.
- Check if the residual series is pure white noise using ADF , KPSS & Ijung box statistic test.
- Evaluate the model using MAPE metric and compare it against the predictions of an AutoARIMA method (fitting a model using the original time series itself and then predictions)

Data understanding

- Contains **51290 observations & 24 attributes**
- Contains no Duplicates
- Contains 4 years' Order details from **Jan 2011 to Dec 2014**
- Attributes of interest for this case study
 - Order Date - Date on which the order was placed
 - Market - The Market segment to which the customer belongs
 - Segment - The market segment to which the product belongs
 - Sales - Total sales value of the transaction
 - Quantity - Quantity of the product ordered
 - Profit - Profit made on the transaction
- Markets include:

▪ Africa	▪ EU
▪ APAC	▪ LATAM
▪ Canada	▪ US
▪ EMEA	
- Segments include:
 - Consumer
 - Corporate
 - Home Office

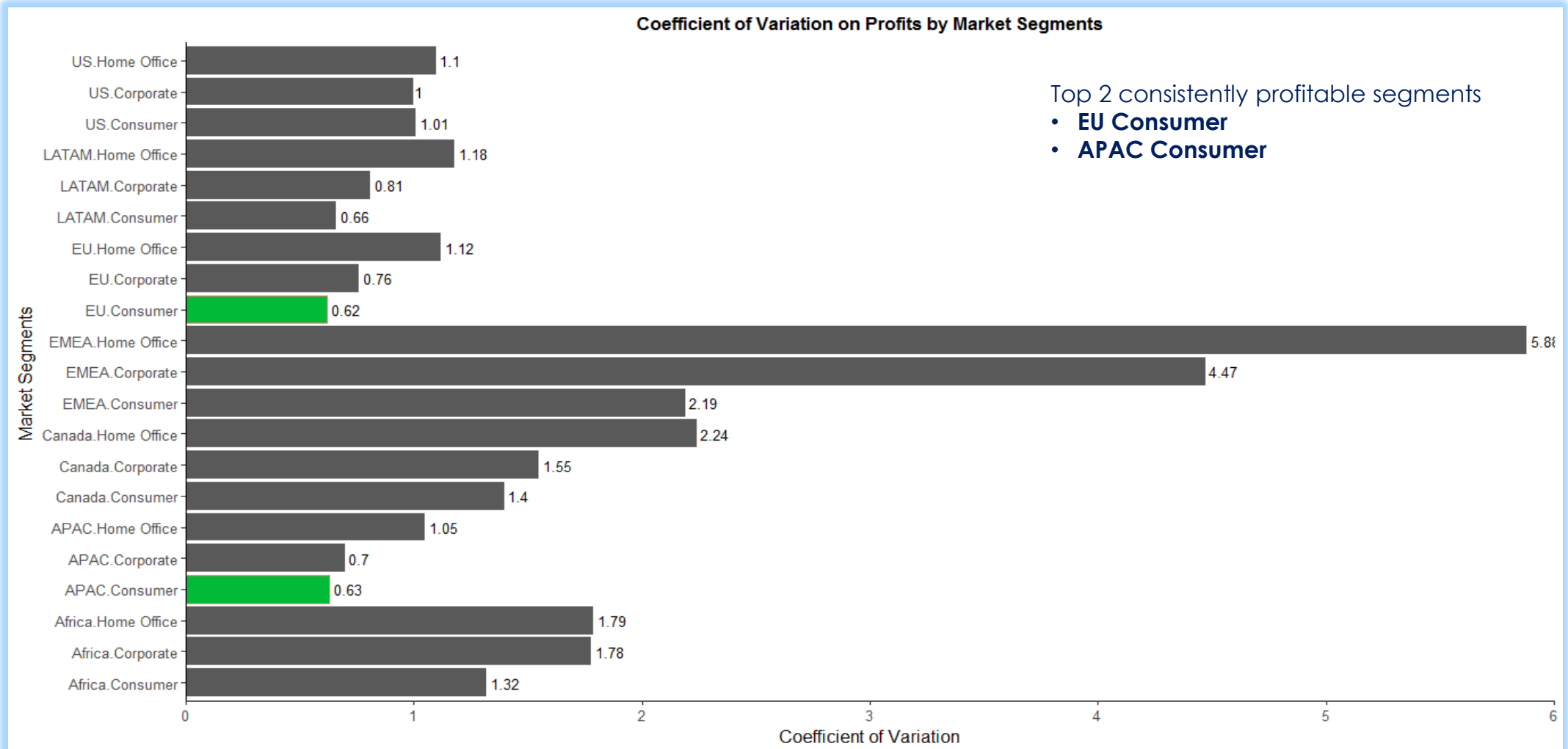
Data Preparation

- Convert order date into correct date format
- Add a new derived column which will have Year, Month values from Order date
- NA Values: Postal Code column alone has 41296 NA values. We will not treat these, since we are not interested in this column for this case study.

Identifying profitable market segments

- Aggregate sum of Sales, Profit, Quantity month on month
- Split the data into 21 Market segments (7 Markets * 3 Segments)
- Identify the top most 2 consistently profitable segments based on coefficient of variation metric on Profits.

Identifying profitable market segments

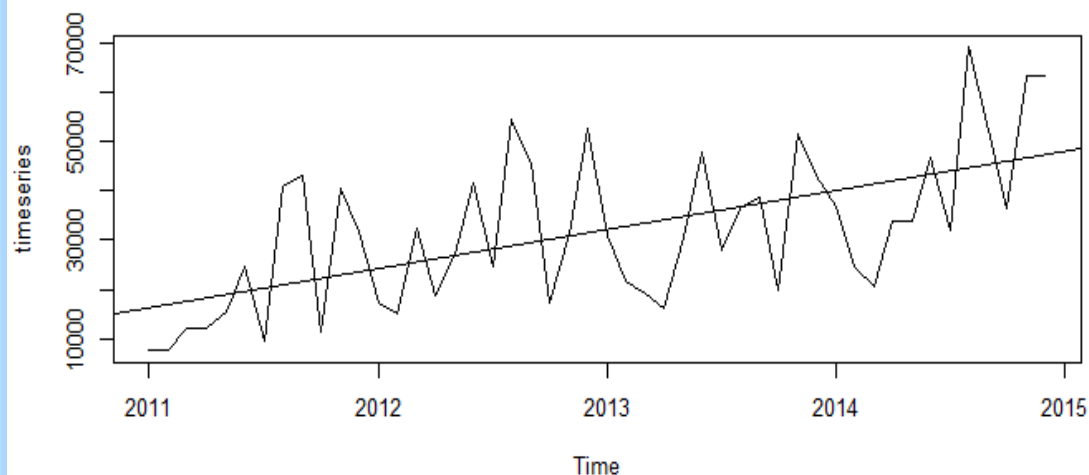


APAC consumer and EA Consumer market segments has the least Coefficient of Variance metric indicating that these are the consistently performing market segments in terms of profits.

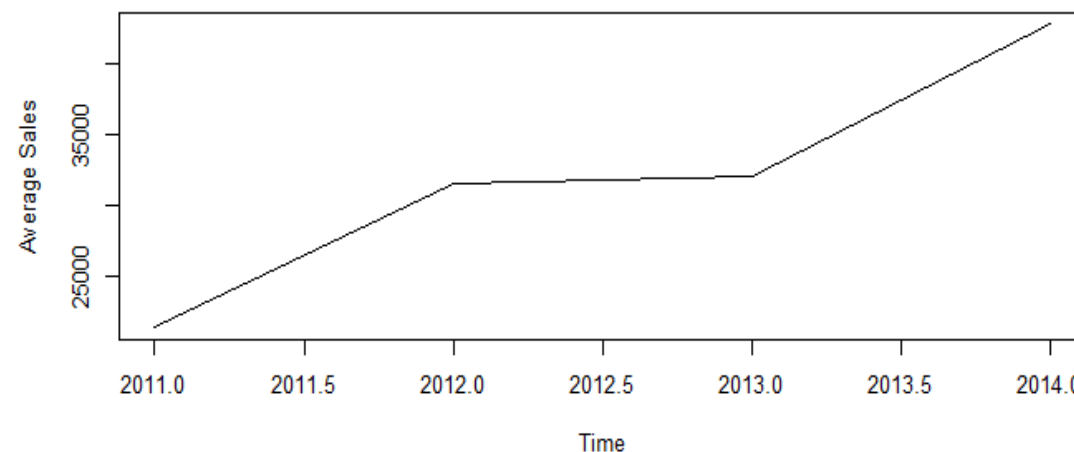
We will focus on these market segments during model building, evaluating and forecasting stages.

Exploratory data analysis on EU Consumer : Sales

Timeseries with trend line

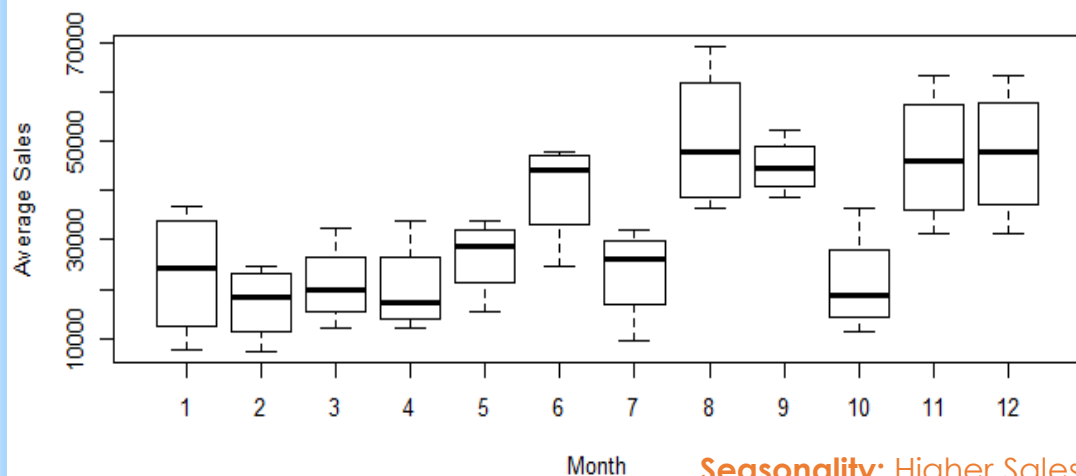


Year on Year trend based on average Sales

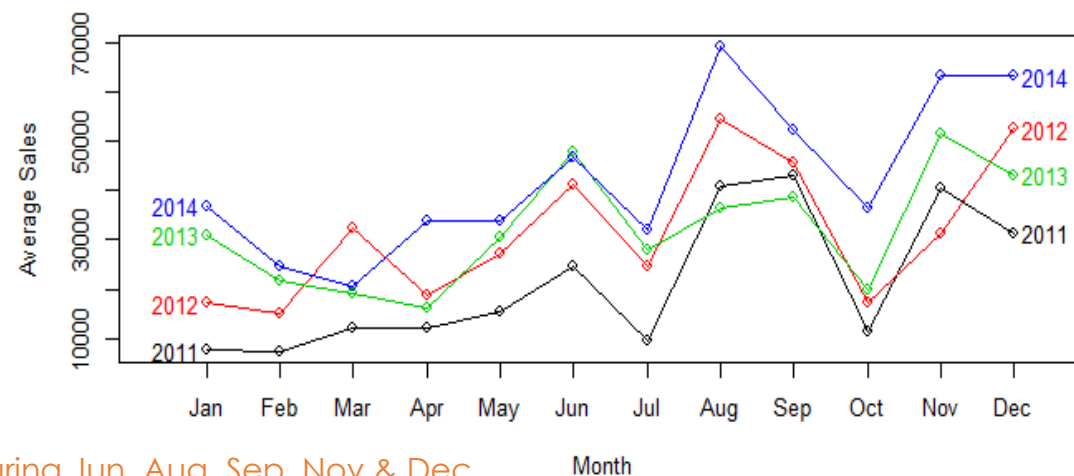


Trend: Sales sharply increased in 2011 and remained somewhat constant in 2012 and again started increasing from 2013 reaching peak values in 2014.

Box plot across months



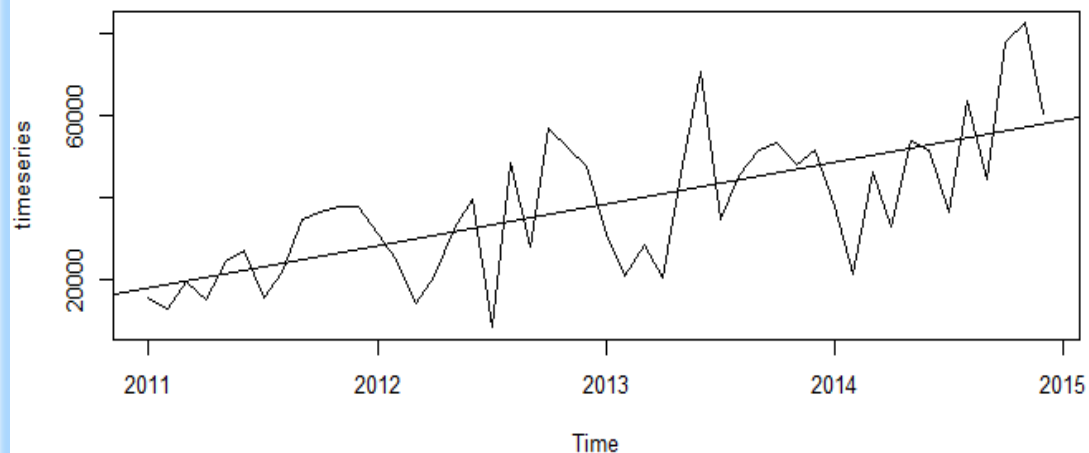
Seasonal plot: timeseries



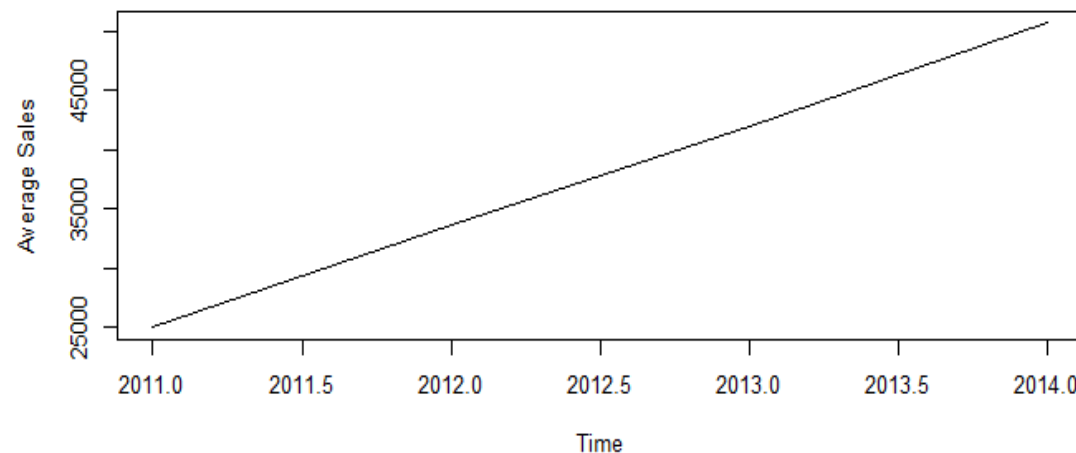
Seasonality: Higher Sales during Jun, Aug, Sep, Nov & Dec

Exploratory data analysis on APAC Consumer : Sales

Timeseries with trend line

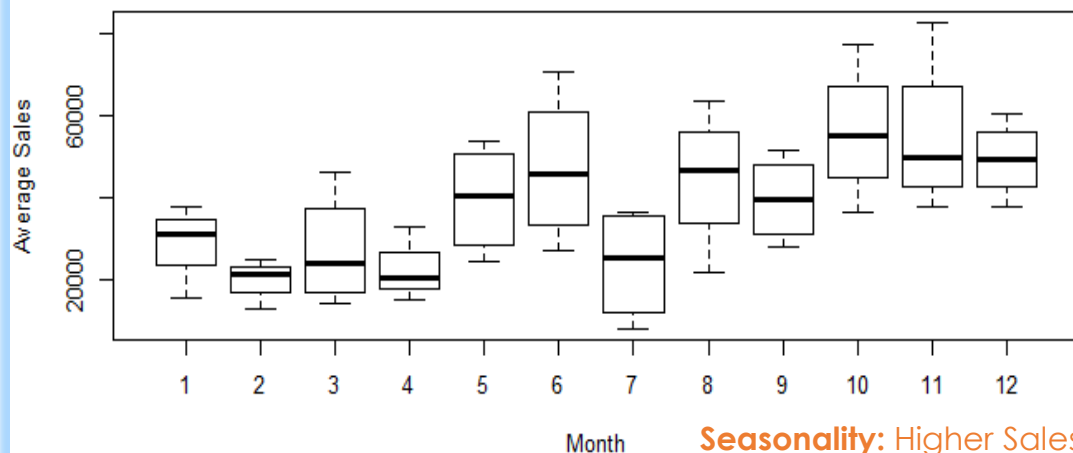


Year on Year trend based on average Sales

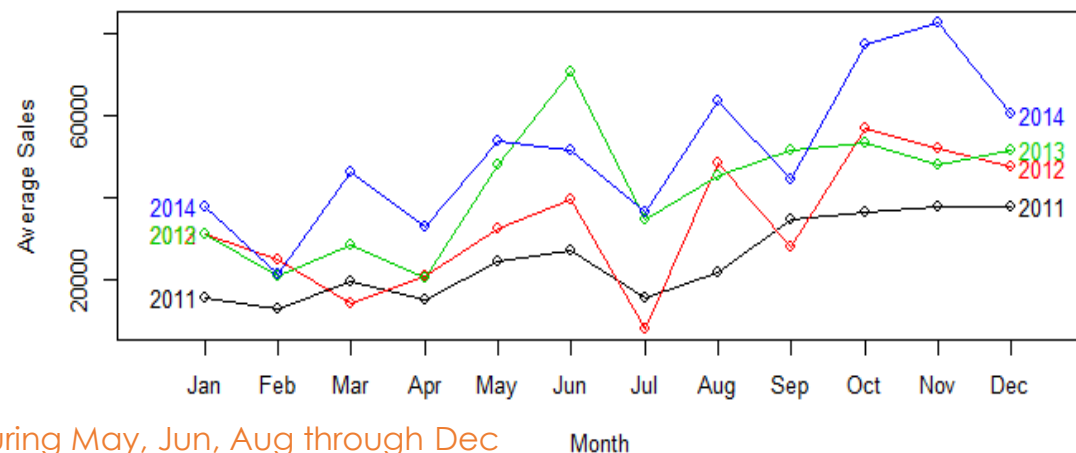


Trend: Sales has steadily increased year on year from 2011 to 2014.

Box plot across months



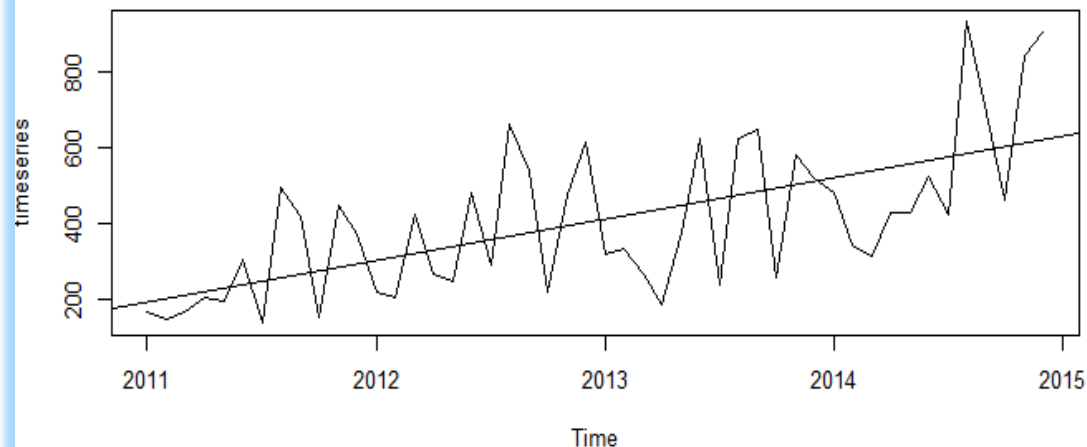
Seasonal plot: timeseries



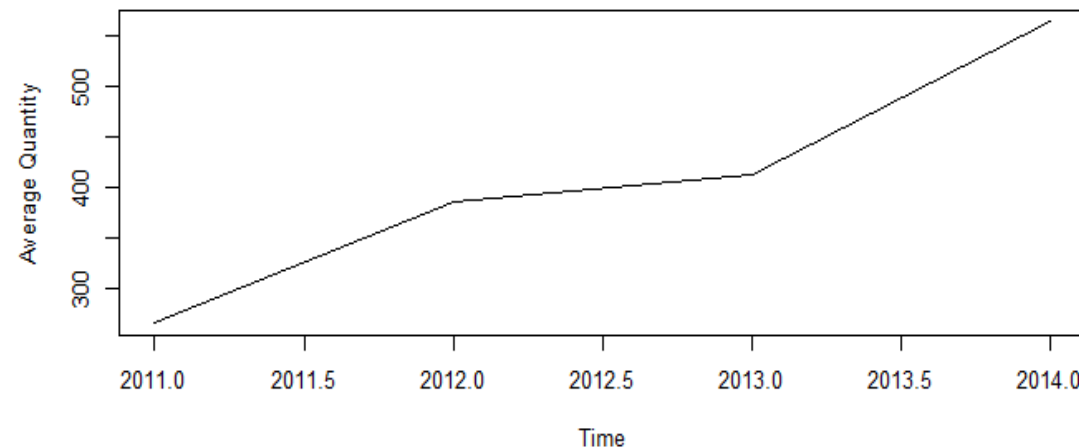
Seasonality: Higher Sales during May, Jun, Aug through Dec

Exploratory data analysis on EU Consumer : Quantity

Timeseries with trend line

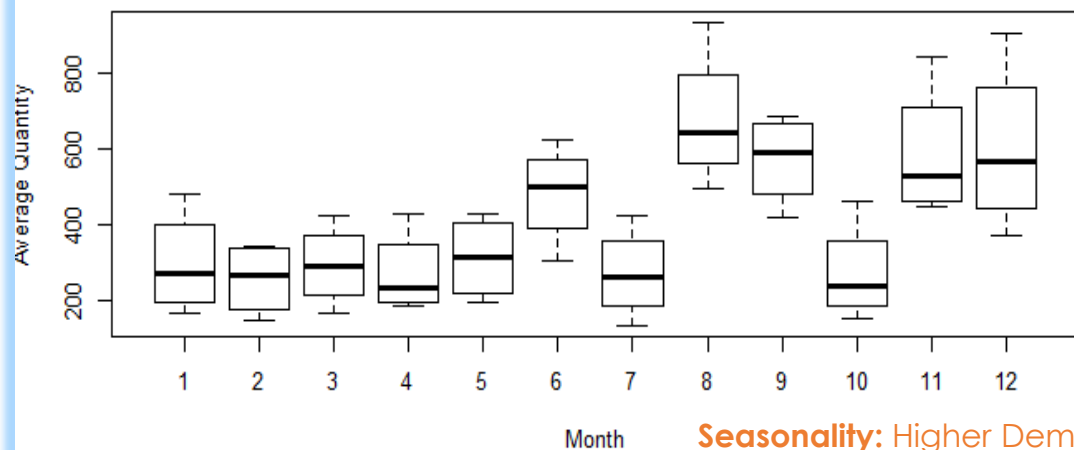


Year on Year trend based on average Sales

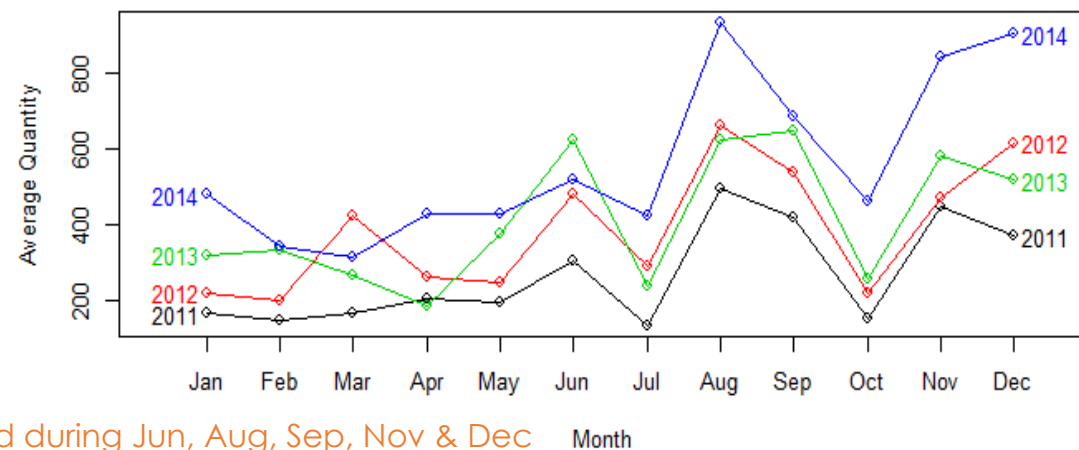


Trend: As expected like Sales, Demand sharply increased in 2011 and remained somewhat constant in 2012 and again started increasing from 2013 reaching peak values in 2014.

Box plot across months



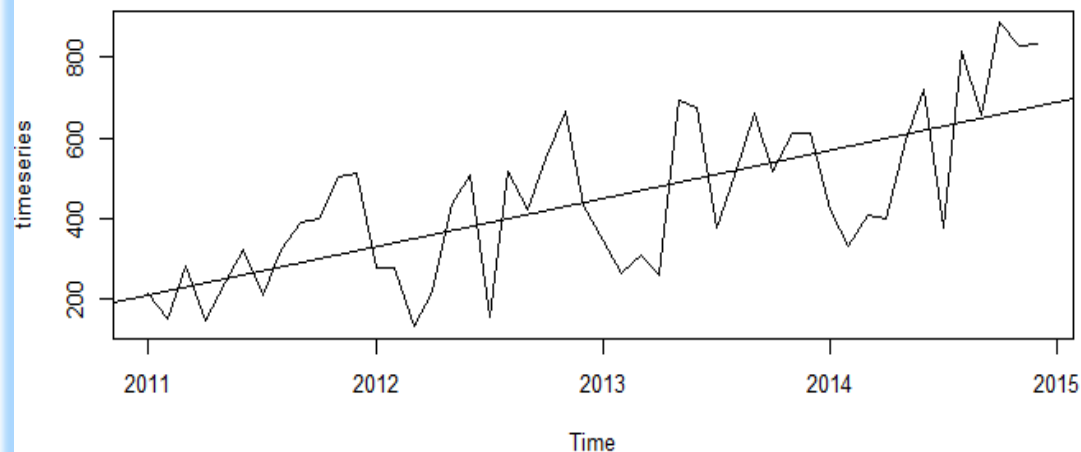
Seasonal plot: timeseries



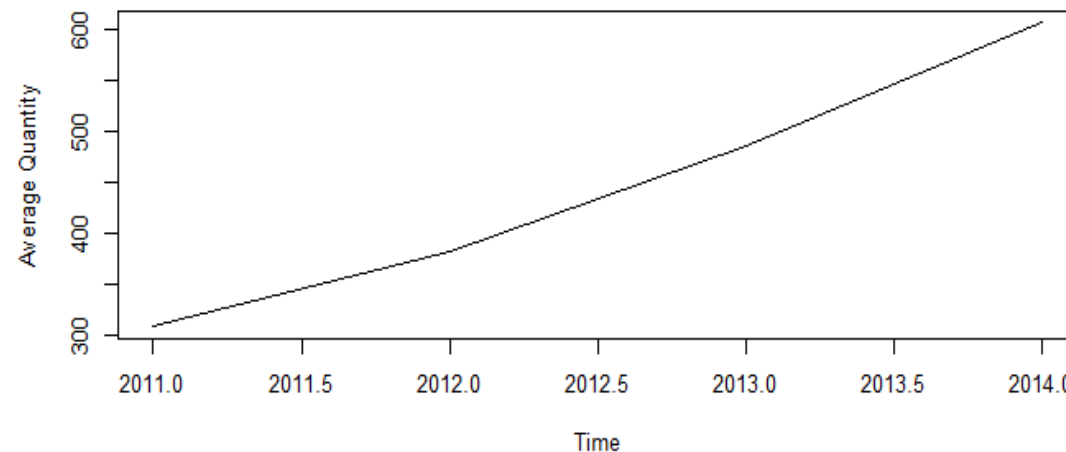
Seasonality: Higher Demand during Jun, Aug, Sep, Nov & Dec

Exploratory data analysis on APAC Consumer : Quantity

Timeseries with trend line

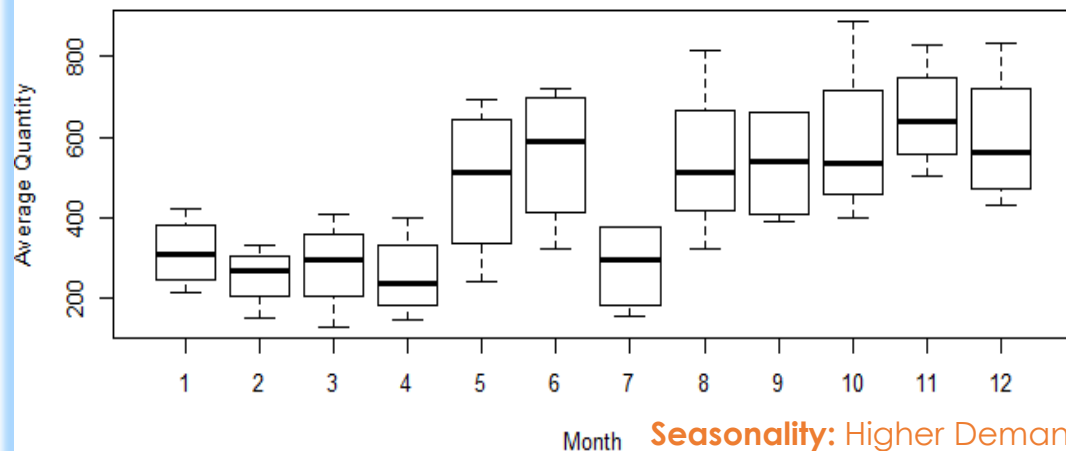


Year on Year trend based on average Sales

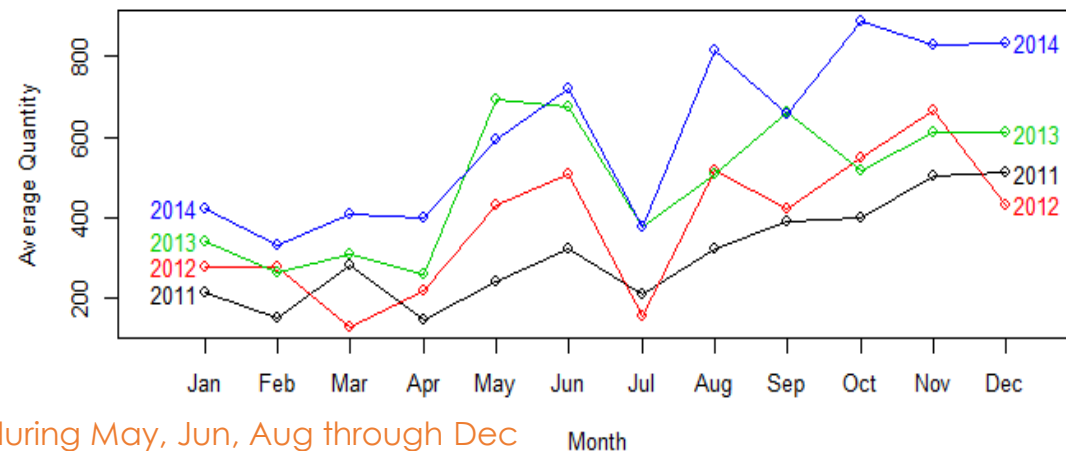


Trend: Demand steadily increased year on year from 2011 to 2014.

Box plot across months



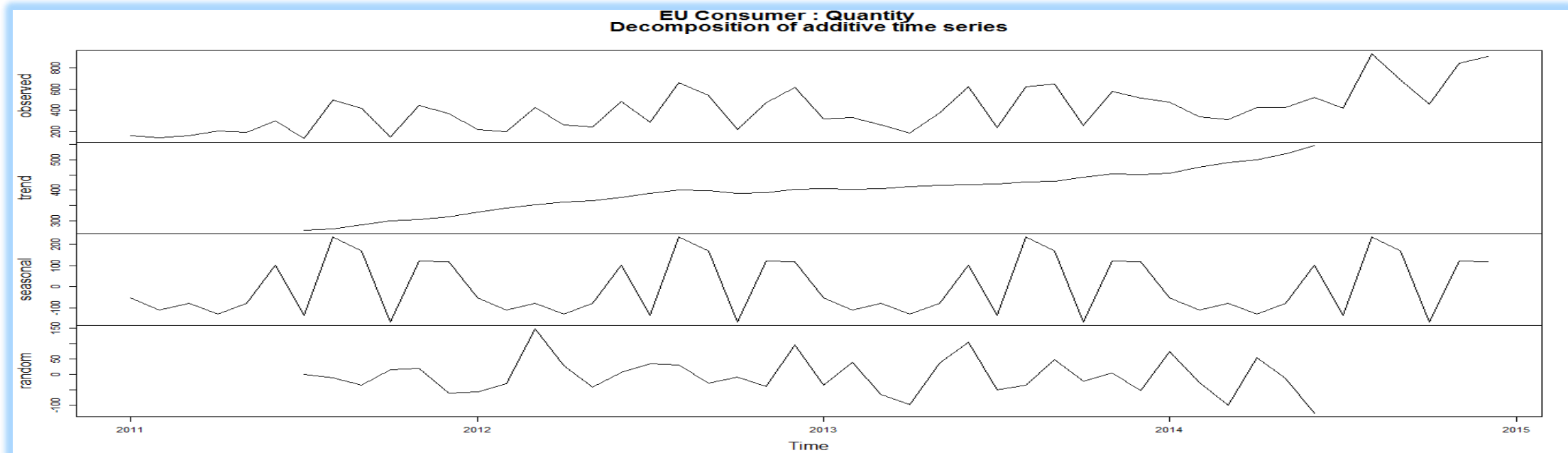
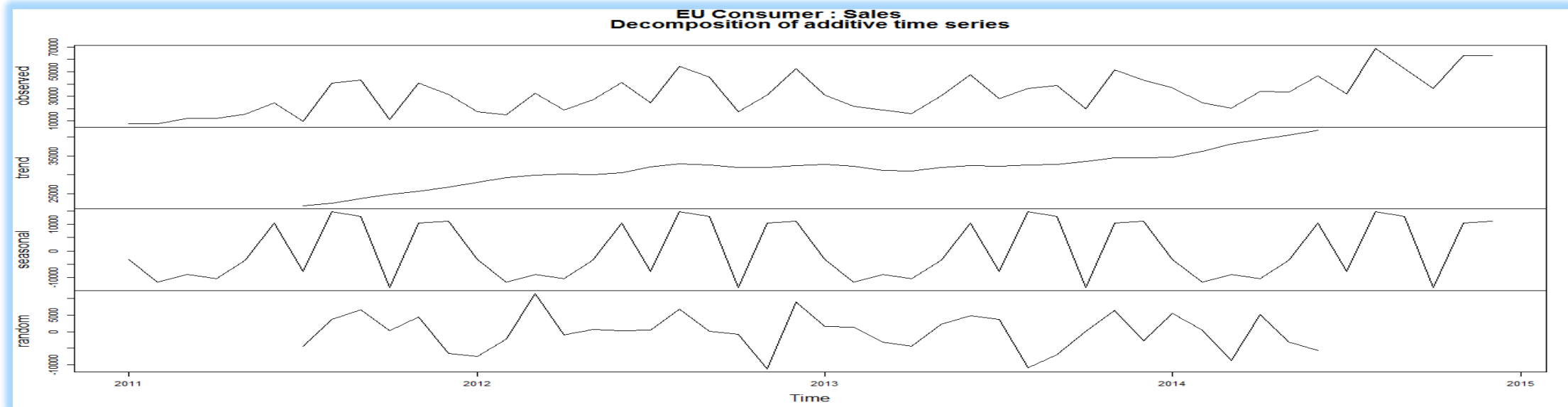
Seasonal plot: timeseries



Seasonality: Higher Demand during May, Jun, Aug through Dec

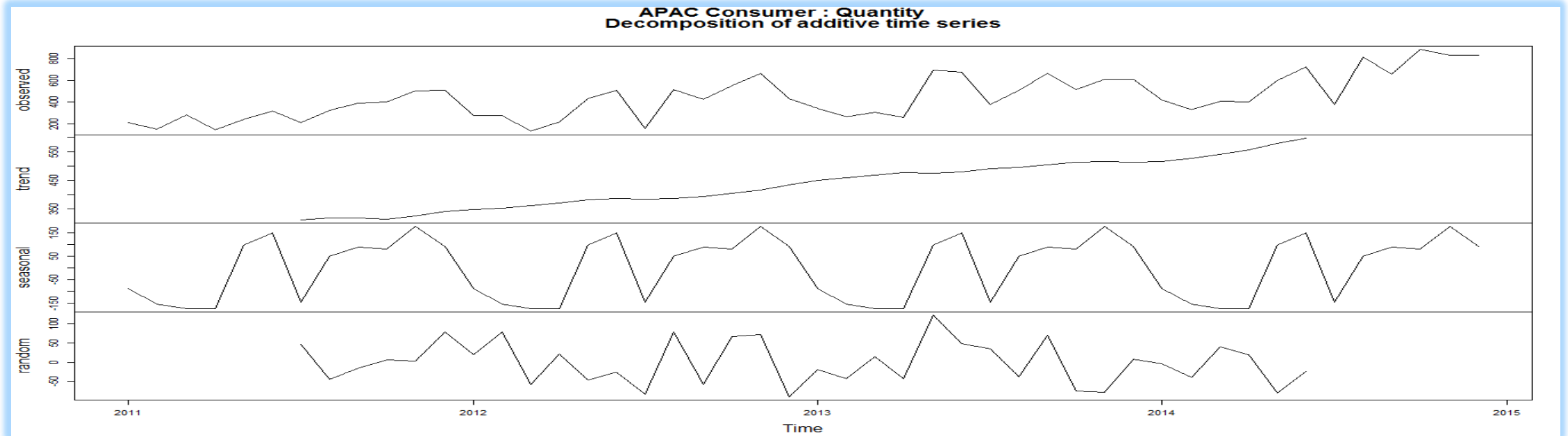
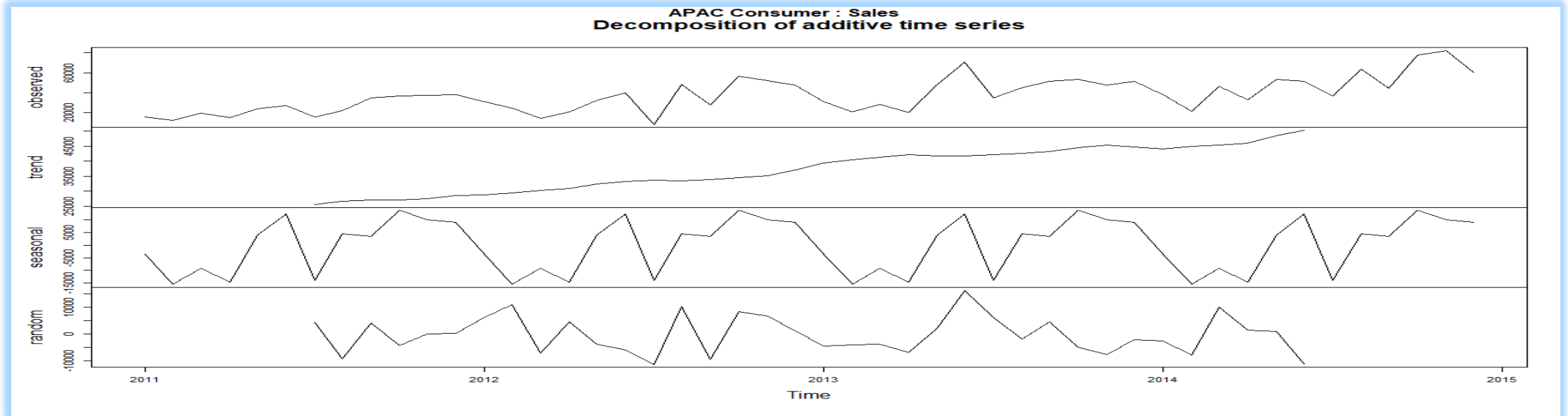
Exploratory data analysis

EU Consumer – Decomposing Time Series

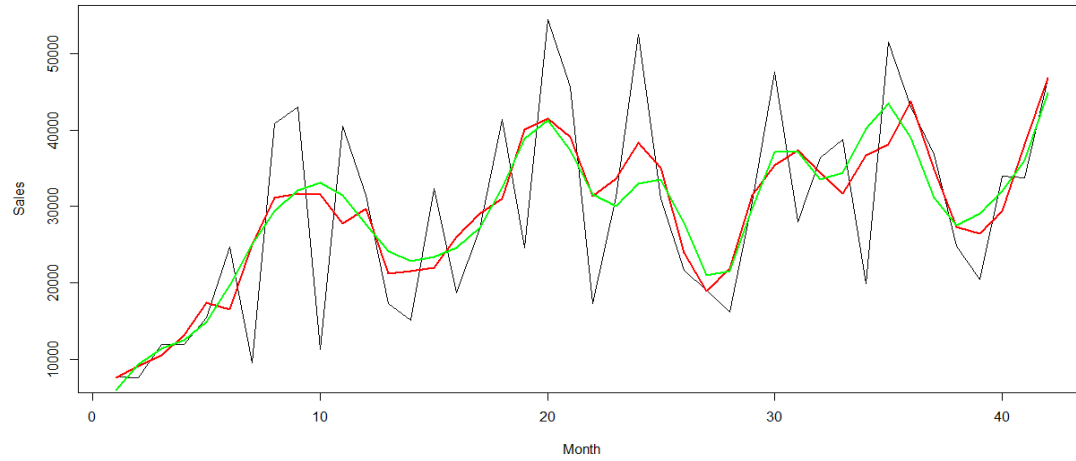


Exploratory data analysis

APAC Consumer – Decomposing Time Series

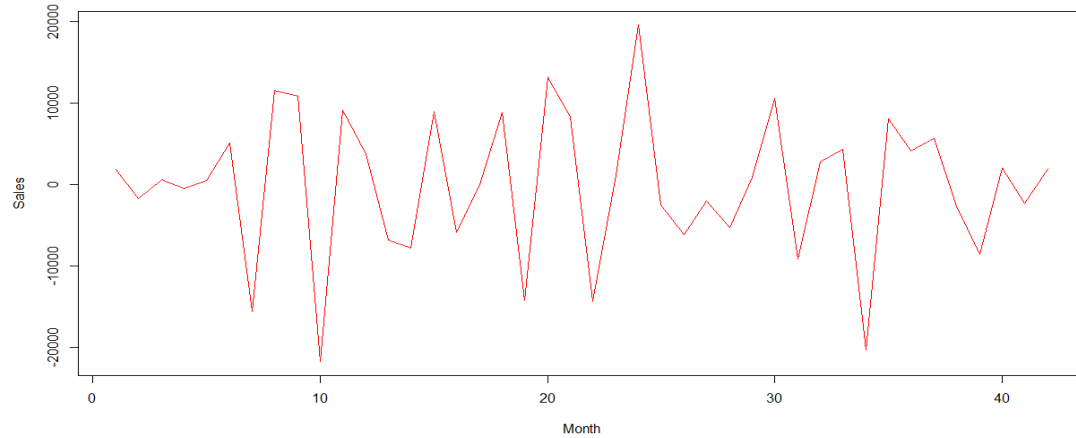


EU Consumer: Sales Fitted



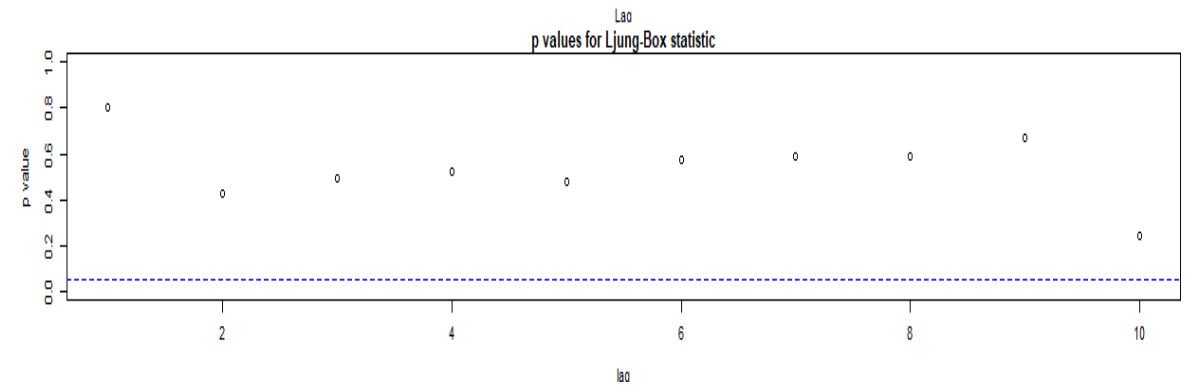
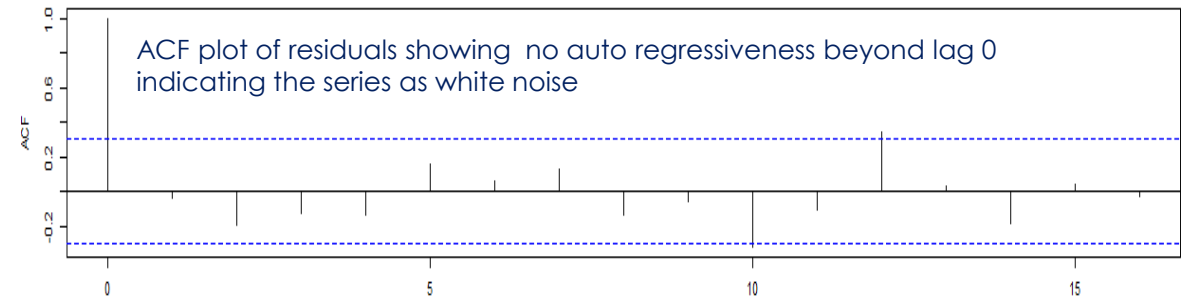
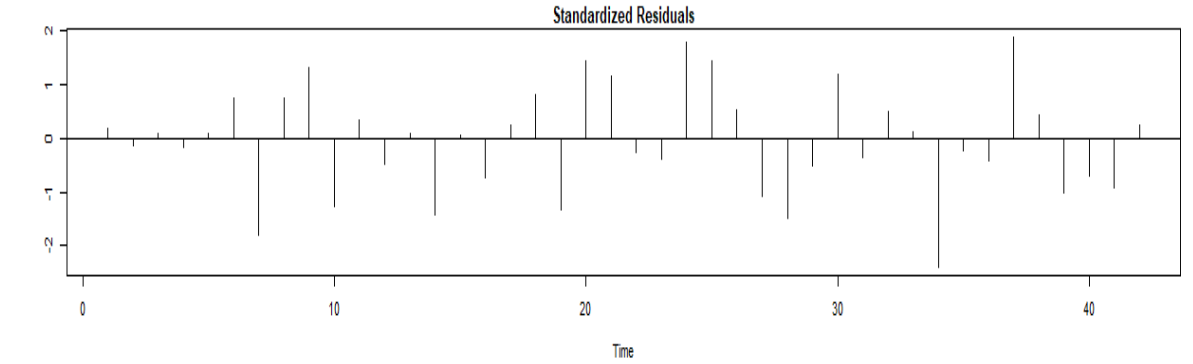
— Smoothened series — Regression fitted series

Local predictable Series



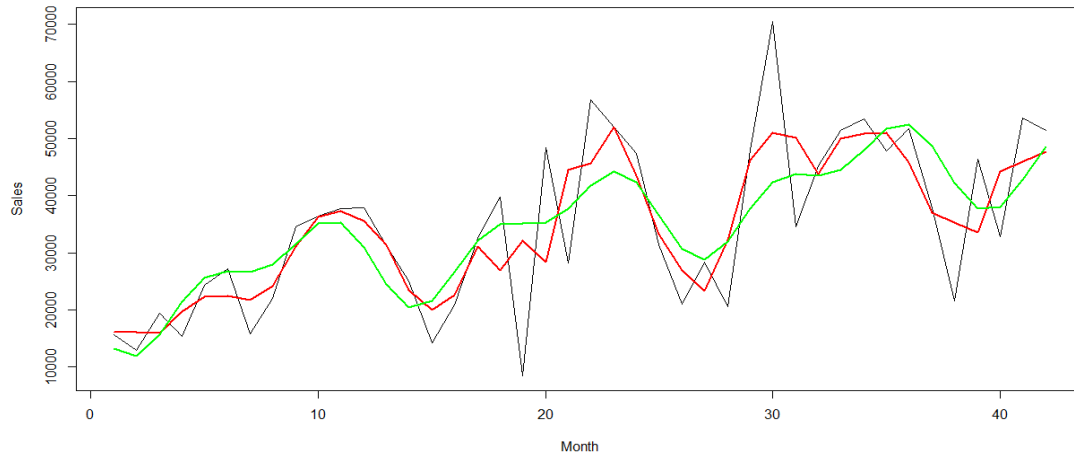
Local component after removing trend and Seasonality

Residuals after removing trend, seasonality and local components



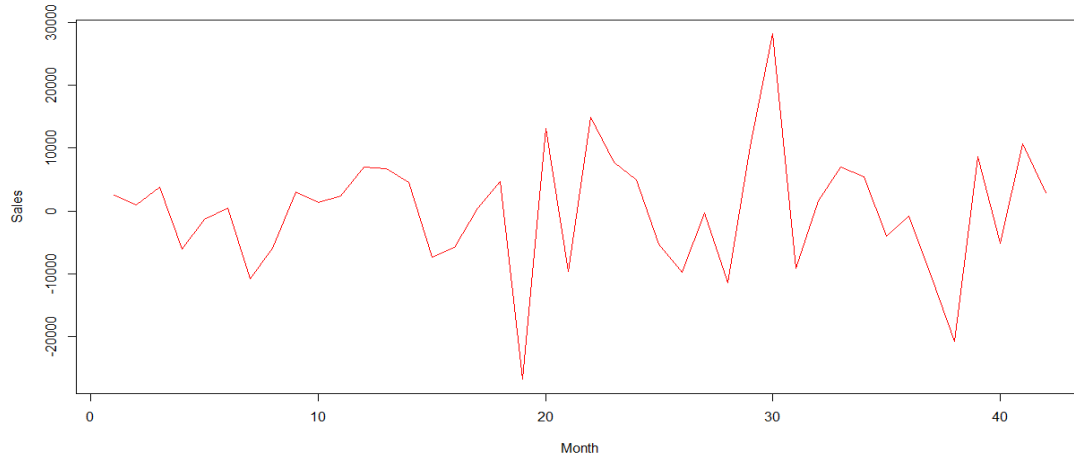
P values from Ljung box indicating no autocorrelation

APAC Consumer: Sales Fitted



— Smoothened series — Regression fitted series

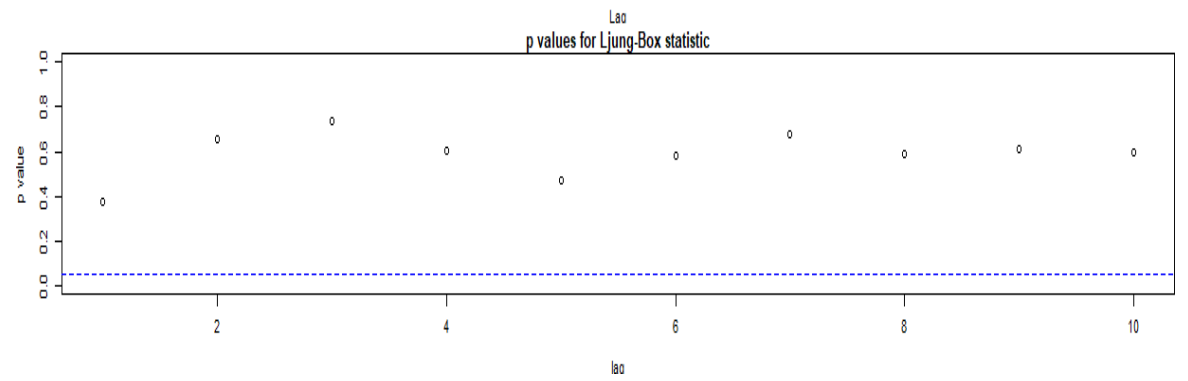
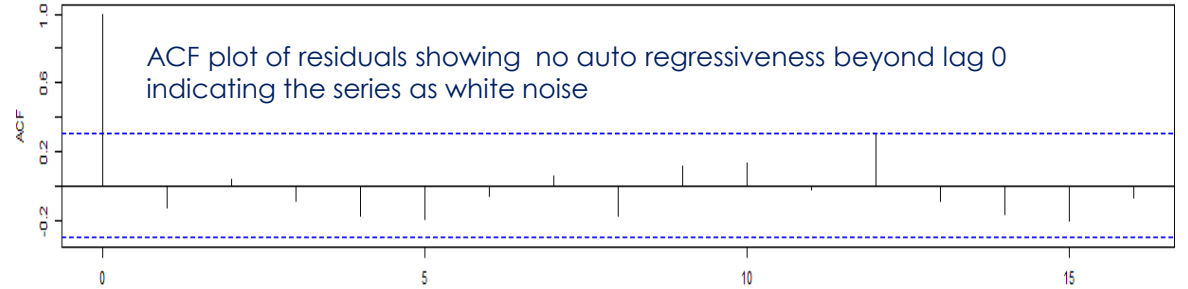
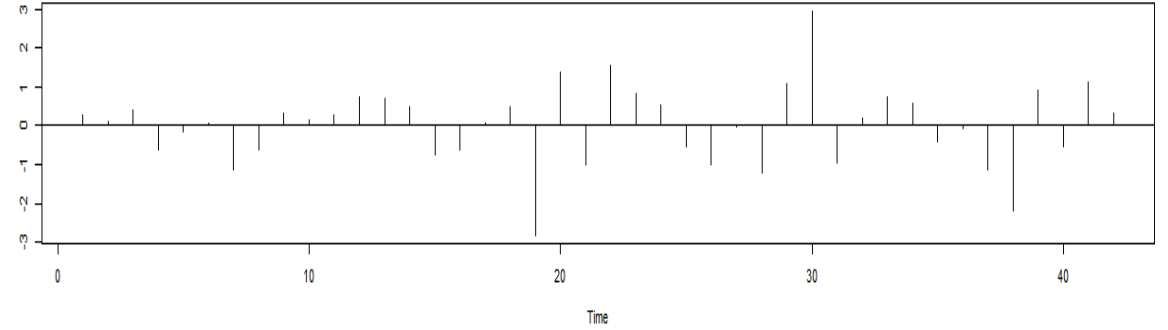
Local predictable Series



Local component after removing trend and Seasonality

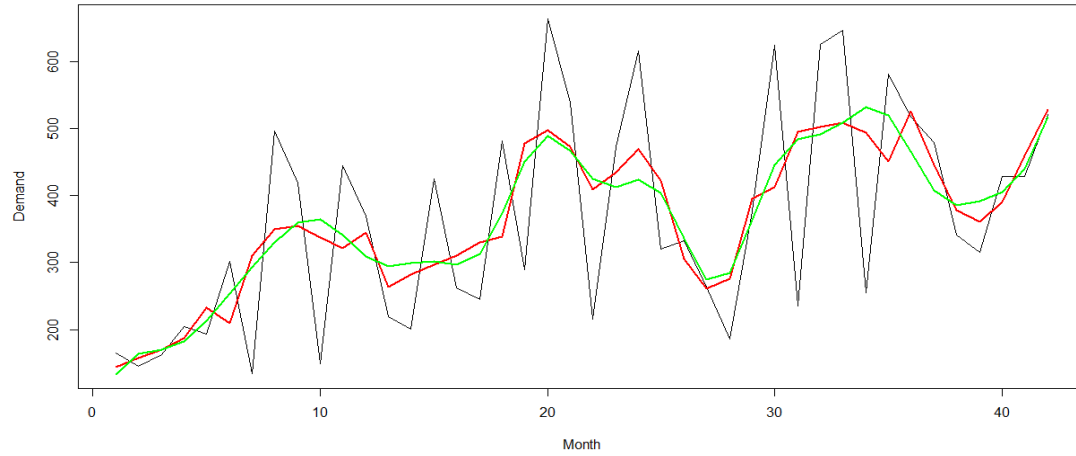
Residuals after removing trend, seasonality and local components

Standardized Residuals



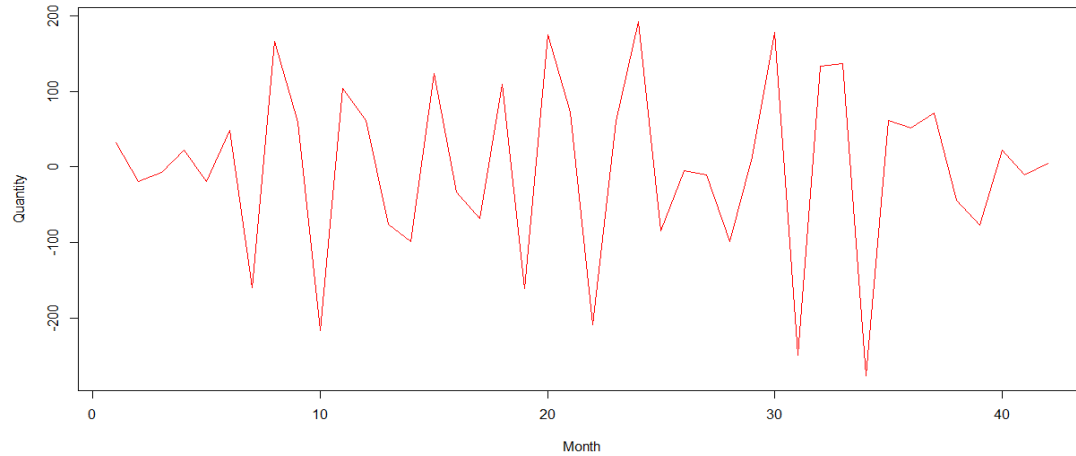
P values from Ljung box indicating no autocorrelation

EU Consumer: Demand Fitted



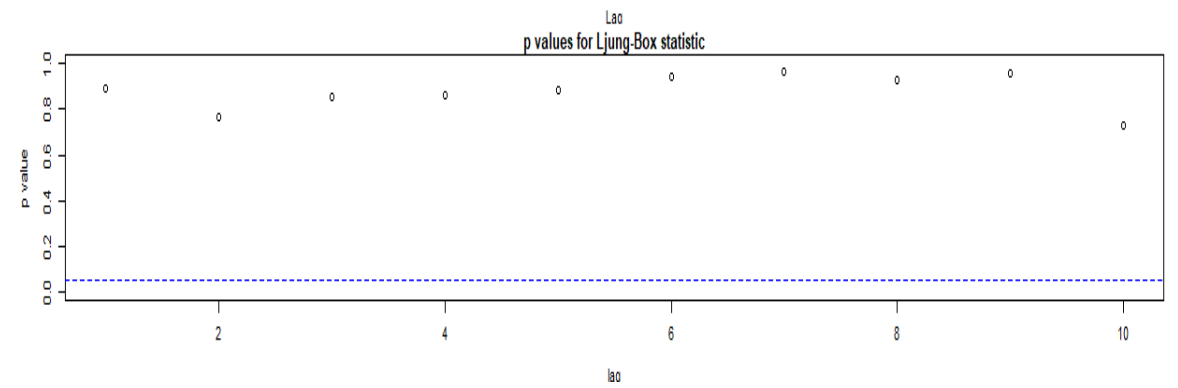
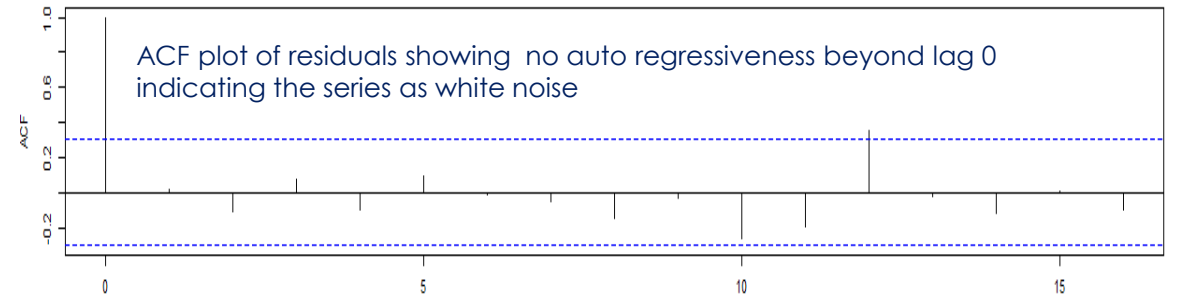
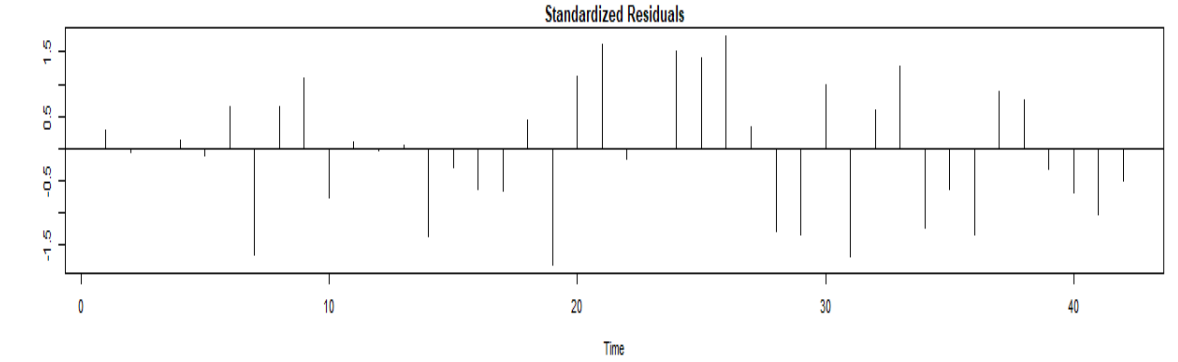
— Smoothened series — Regression fitted series

Local predictable Series



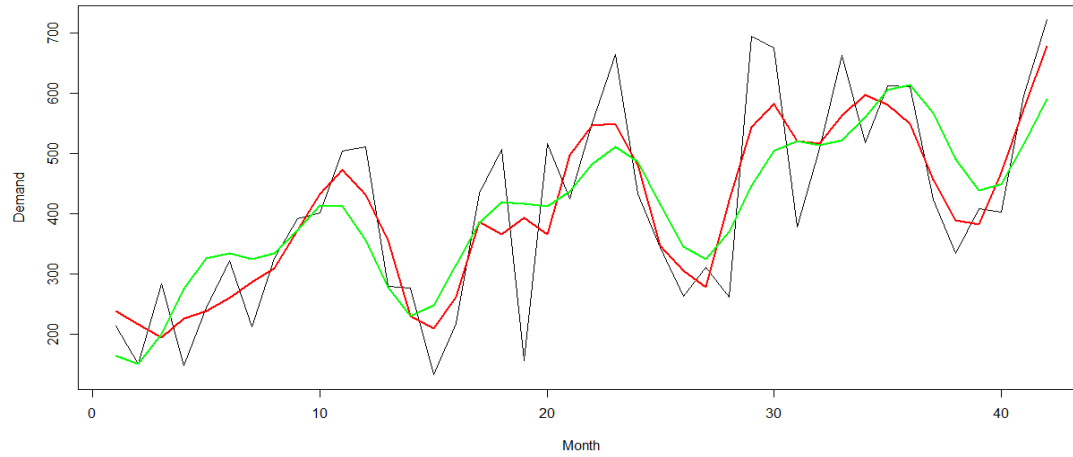
Local component after removing trend and Seasonality

Residuals after removing trend, seasonality and local components



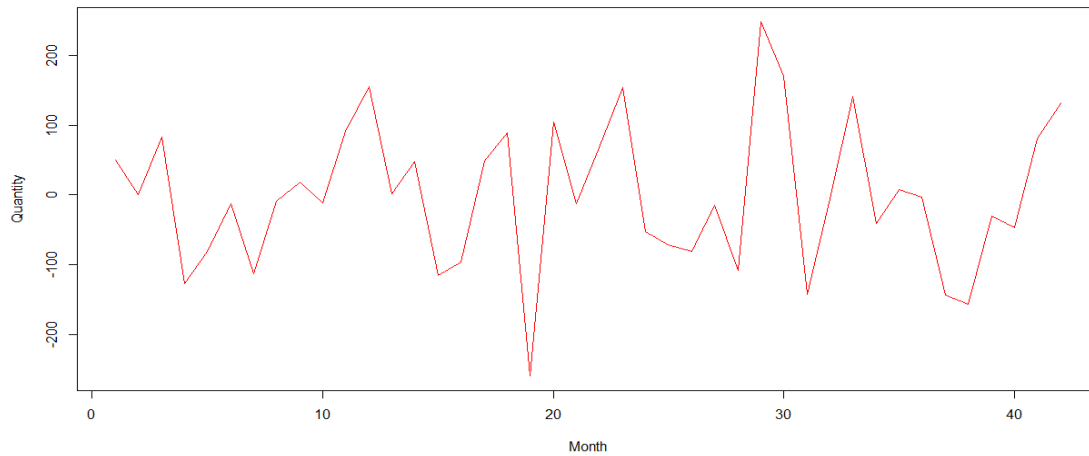
P values from Ljung box indicating no autocorrelation

APAC Consumer: Demand Fitted



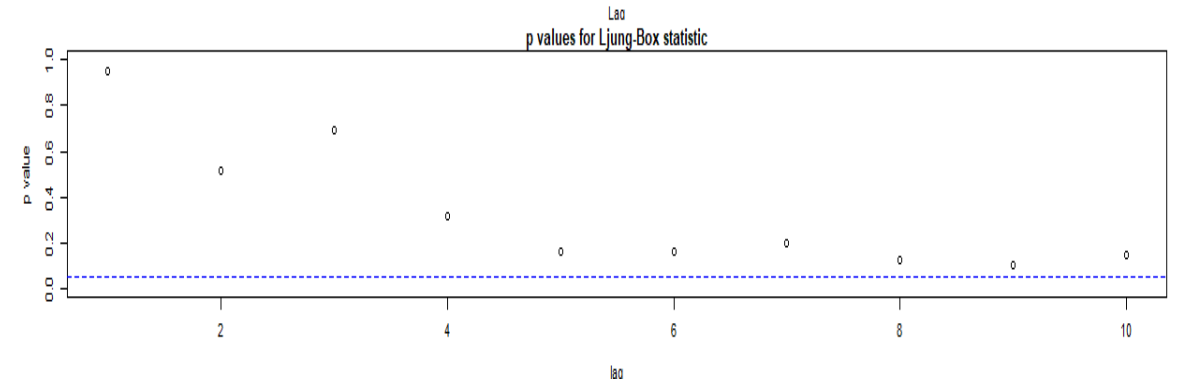
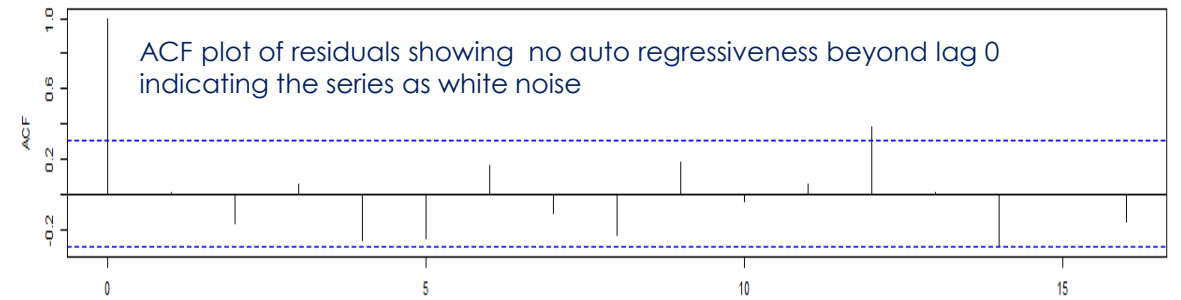
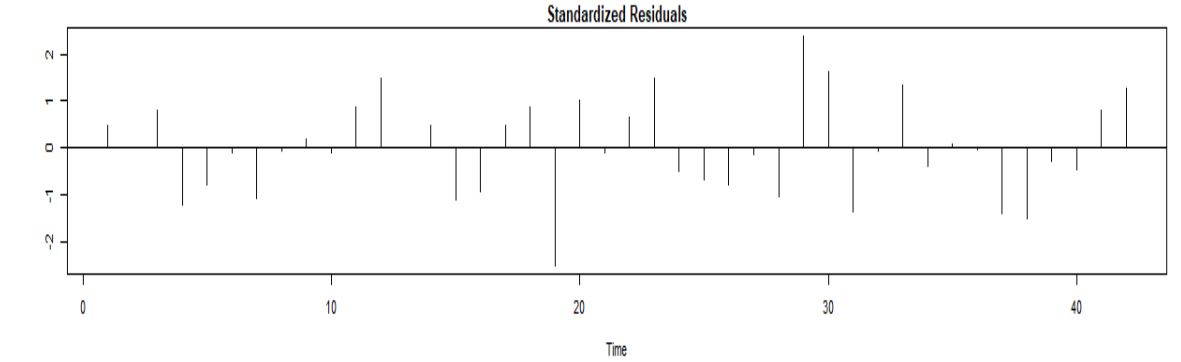
— Smoothed series — Regression fitted series

Local predictable Series



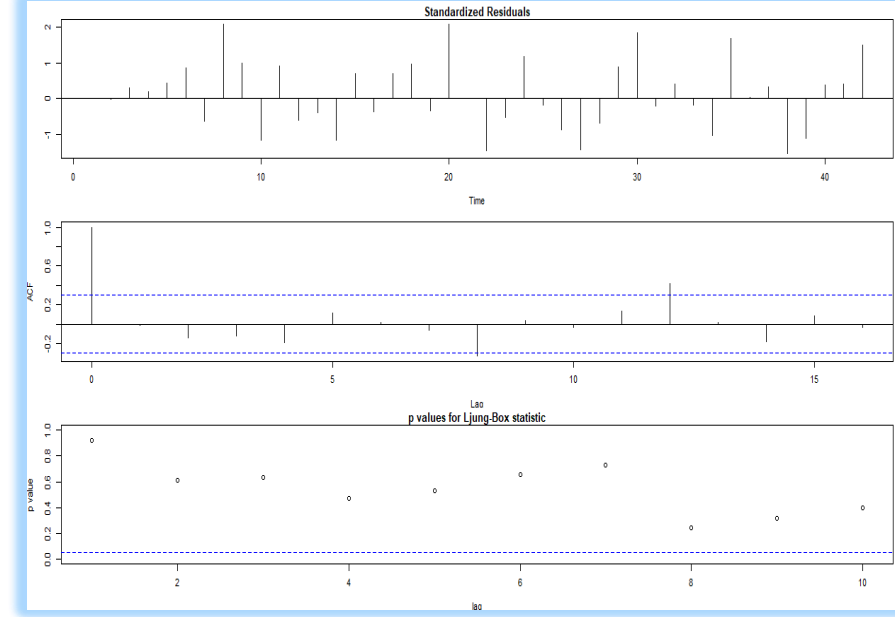
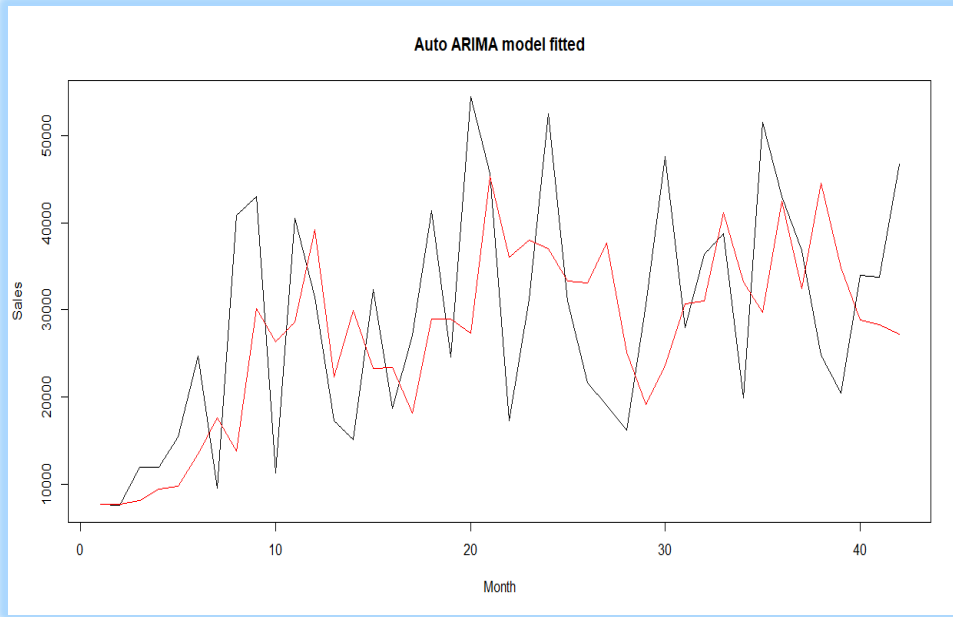
Local component after removing trend and Seasonality

Residuals after removing trend, seasonality and local components



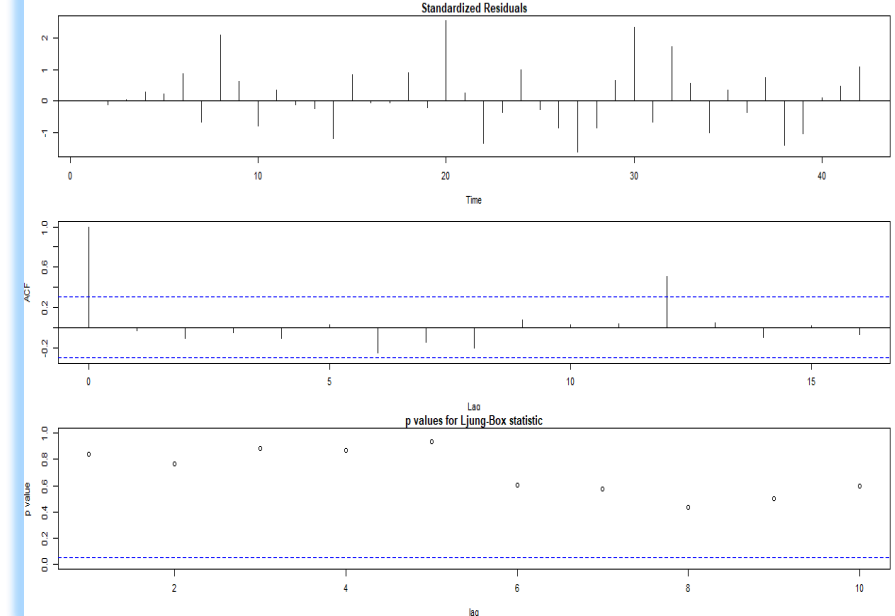
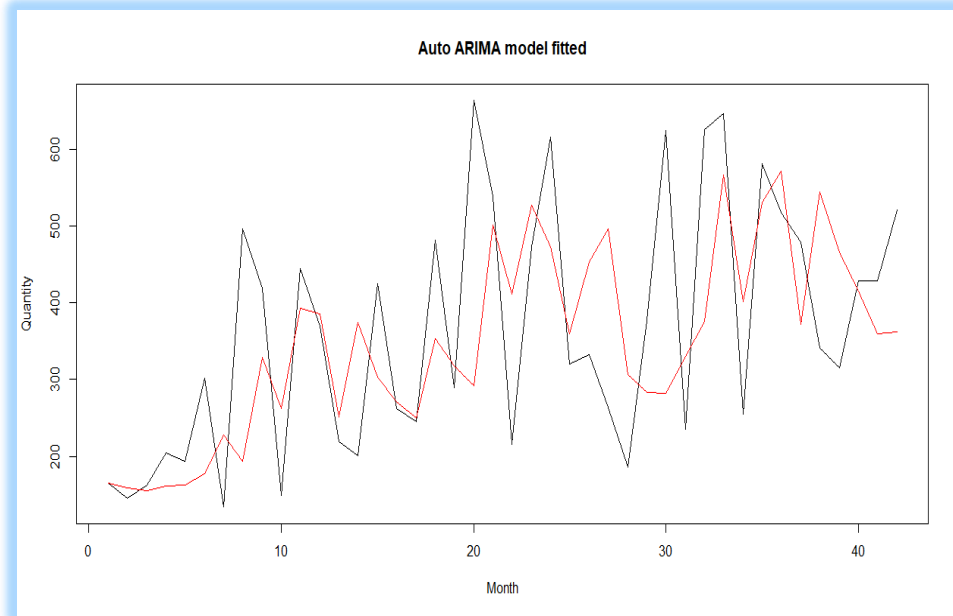
P values from Ljung box indicating no autocorrelation

Sales

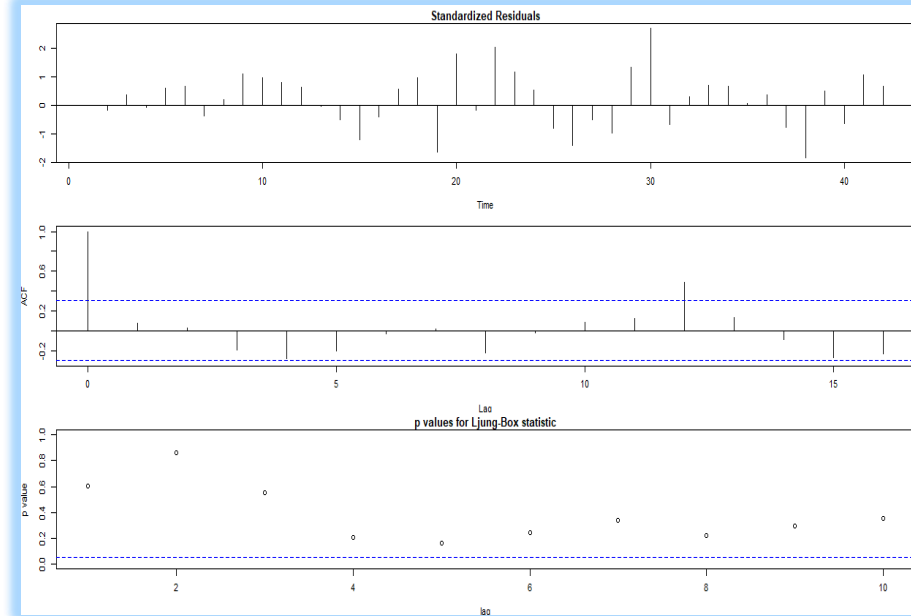
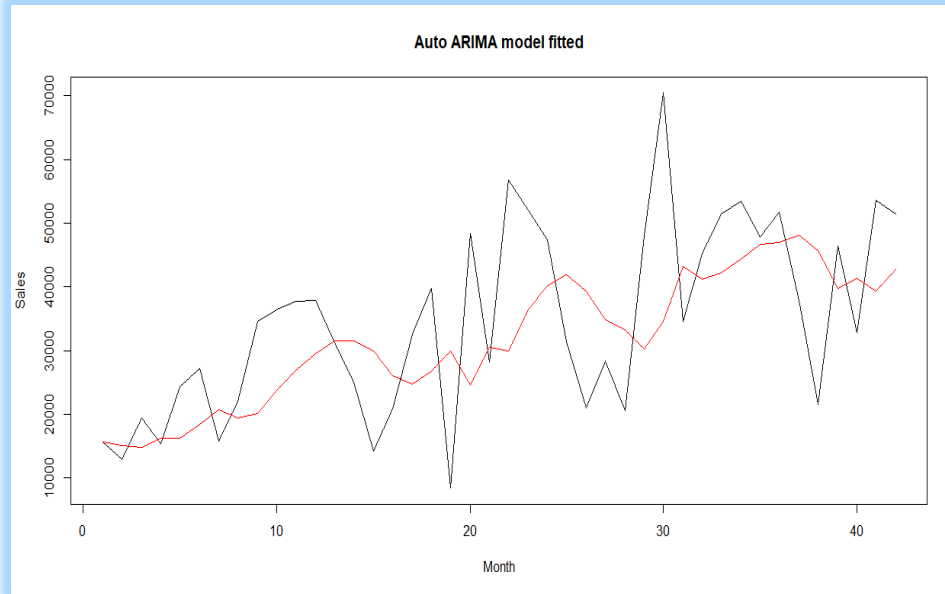


- A. Residuals after removing trend, seasonality and local components
- B. ACF plot of residuals showing no auto regressiveness beyond lag 0 indicating the series as white noise
- C. P values from Ljung box indicating no autocorrelation

Demand

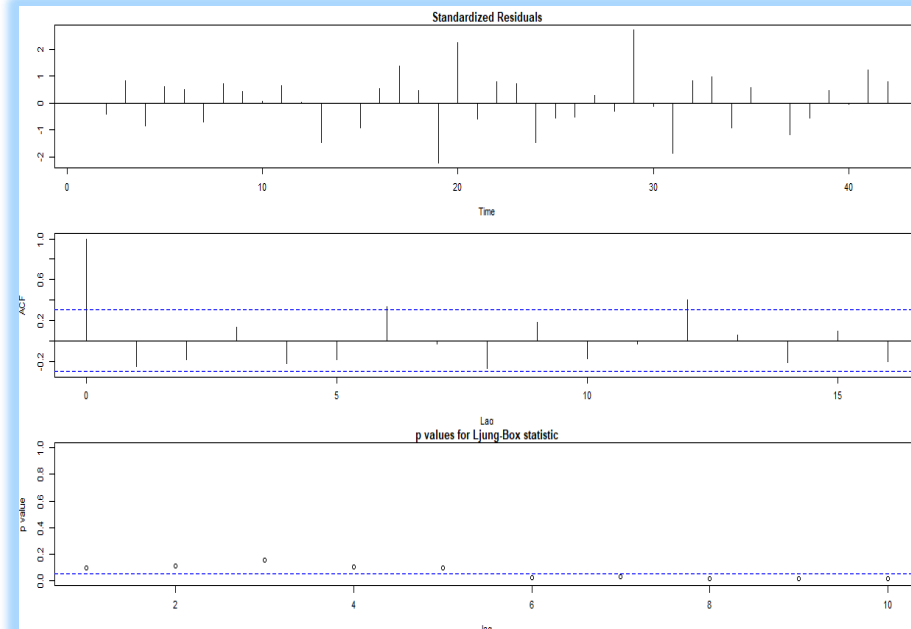
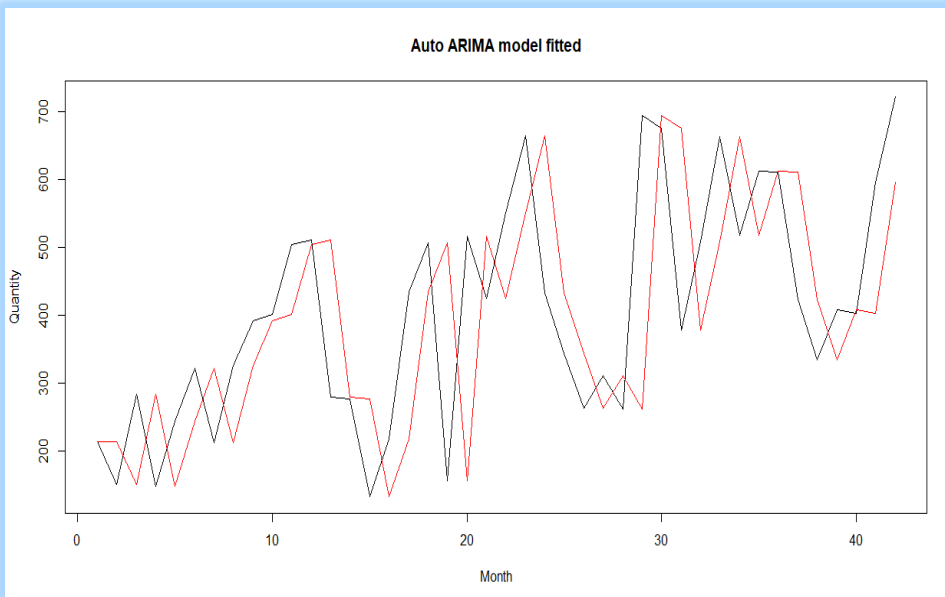


Sales



- A. Residuals after removing trend, seasonality and local components
- B. ACF plot of residuals showing no auto regressiveness beyond lag 0 indicating the series as white noise
- C. P values from Ljung box indicating some autocorrelation

Demand



Model Evaluation using MAPE

Market Segment	Attribute	Classical Decomp. MAPE	Auto ARIMA MAPE
EU Consumer	Sales	24.17	28.92
APAC Consumer	Sales	22.89	27.69
EU Consumer	Quantity	22.82	30.13
APAC Consumer	Quantity	24.52	26.24

The model using Classical decomposition method has significantly lower MAPE values indicating it to be better models than the AUTO ARIMA model.

Actual vs Forecast for last 6 Months in 2014:

Market Segment	Attribute	Actuals/Forecast	Jul.2014	Aug.2014	Sep.2014	Oct.2014	Nov.2014	Dec.2014
EU Consumer	Sales	Actuals	31967.69	68951.72	52328.68	36348.31	63218.71	63178.60
		Forecast	55965.73	59967.19	53033.26	46781.88	55095.88	72025.72
APAC Consumer	Sales	Actuals	423	932	688	459	843	905
		Forecast	605	635	598	594	710	872
EU Consumer	Quantity	Actuals	36524.30	63521.77	44477.27	77379.83	82286.36	60292.13
		Forecast	51775.86	52192.08	52202.86	54287.98	58243.78	61102.96
APAC Consumer	Quantity	Actuals	377	816	658	885	829	833
		Forecast	632	635	631	653	701	736