
TIME SERIES
CASE STUDY

Understanding Business & the requirements

"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 –

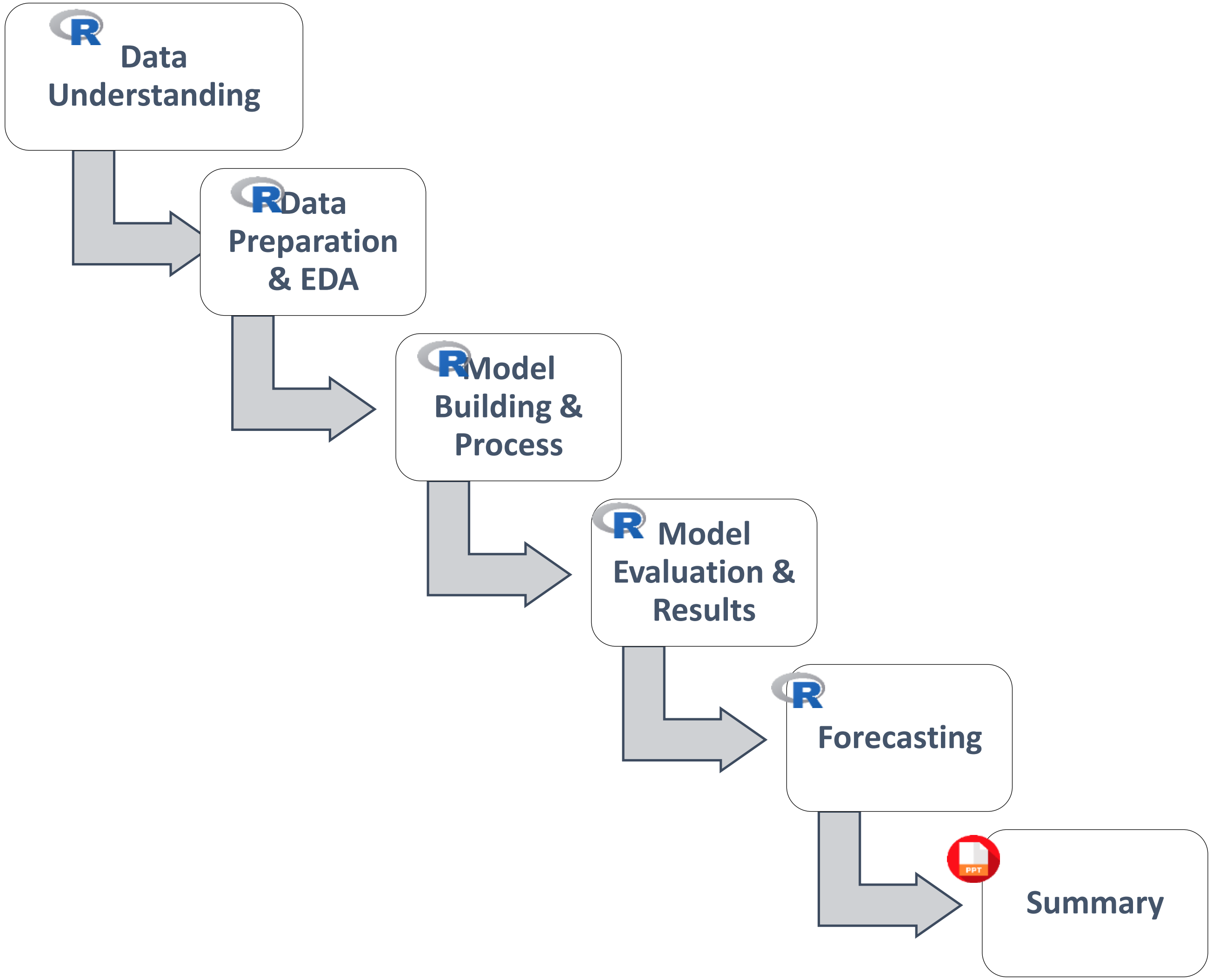
- Consumer
- Corporate
- Home Office

"Objective is to forecast the sales and the demand for the next 6 months, that would help manage the revenue and inventory accordingly."

"Global Mart" caters to 7 different market segments and in 3 major categories.

Project objective is to find out **2 most profitable (and consistent) segment** among all 21 Buckets [7*3 buckets] and **forecast the sales and demand for the next 6 months**.

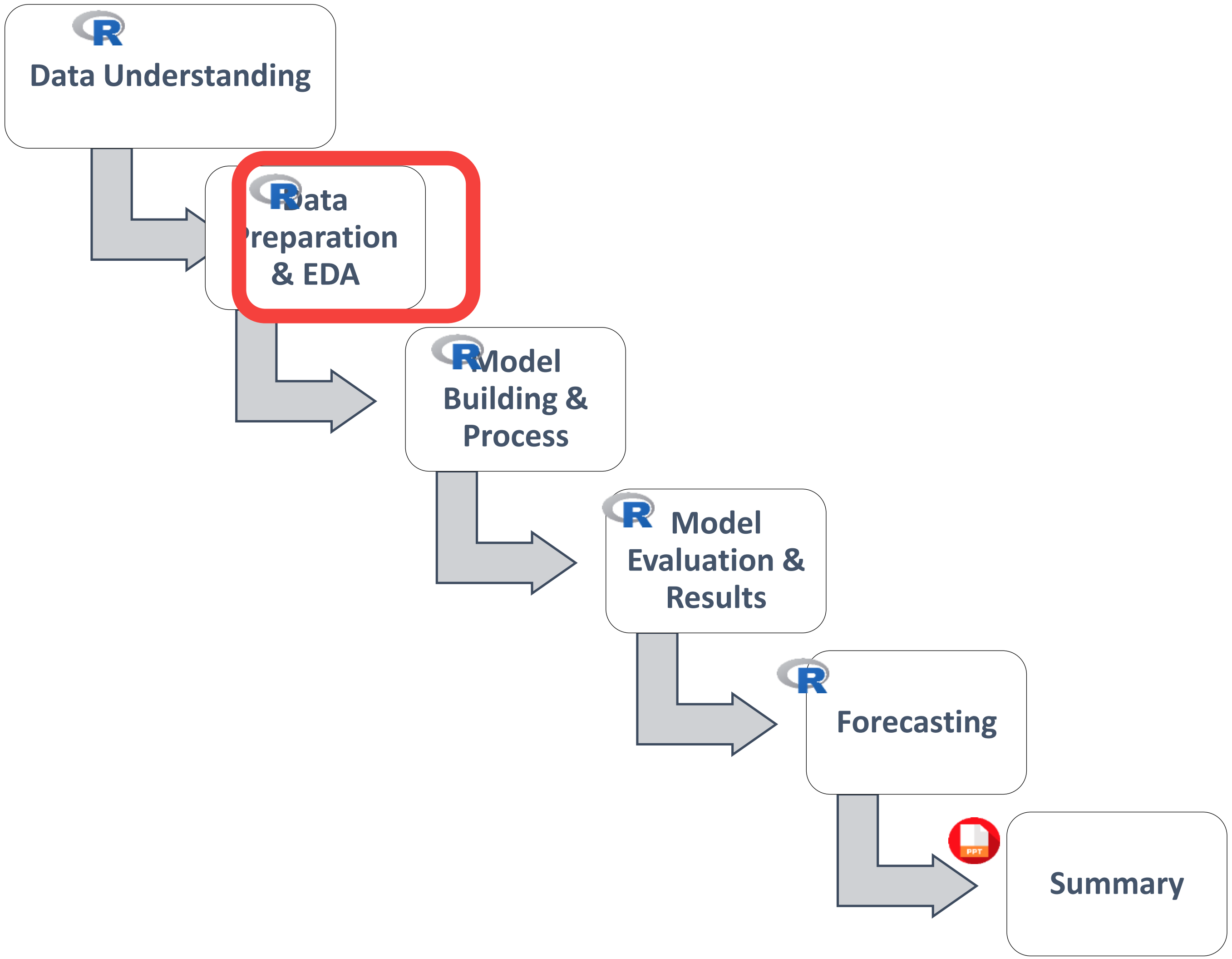
Analysis Process



Data Understanding

- The data currently has the transaction level data, where each row represents a particular order made on the online store. And there are 51,290 transaction records on the dataset.
- Data consists of 24 attributes with relates all these transactions. Two key attributes which are very essential for the model objectives
  - The “Market” attribute has 7-factor levels representing the geographical market sector that the customer belongs to.
  - The “Segment” attribute tells which of the 3 segments that customer belongs to.

Analysis Process



Data Preparation

Final 2 Data Set for Model Building

1. **APAC Consumer**
2. **EU Consumer**

Subset the dataset

Finding 2 most profitable and consistently profitable buckets through 2 ways -

- Monthly higher Average Profitable segments
- Low Coefficient of Variation among the buckets

Changing the Attribute Types

Changing all date columns (Order.Date & Shipping.Date) to "Date-Type" format

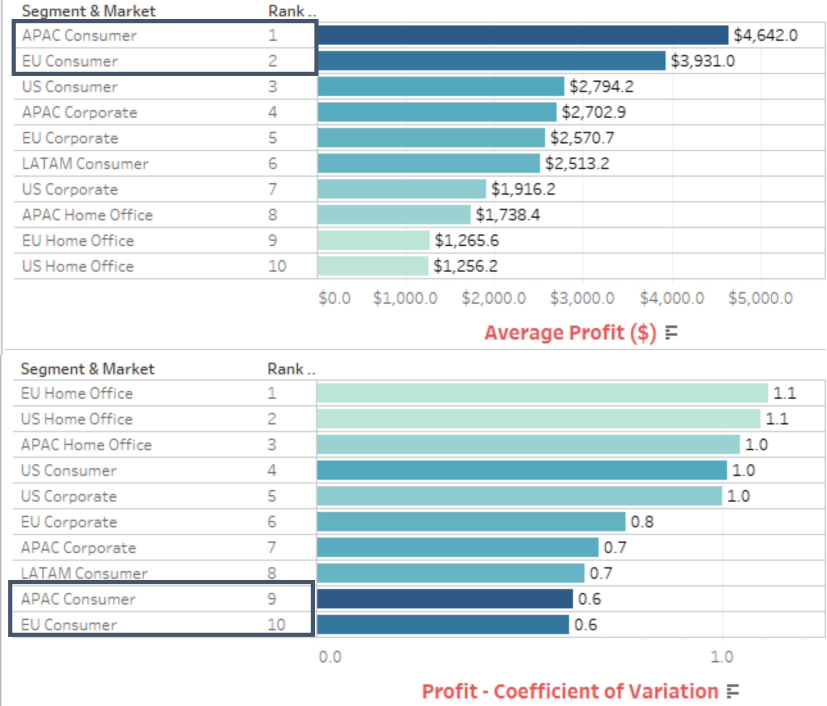
Finding the NAs

Dataset has 41,296 'NA' values which is only exists in "Postal-Code" attribute. No action has been taken here as this attribute has 'NOT' been considered for the Model Building

Data Uniqueness

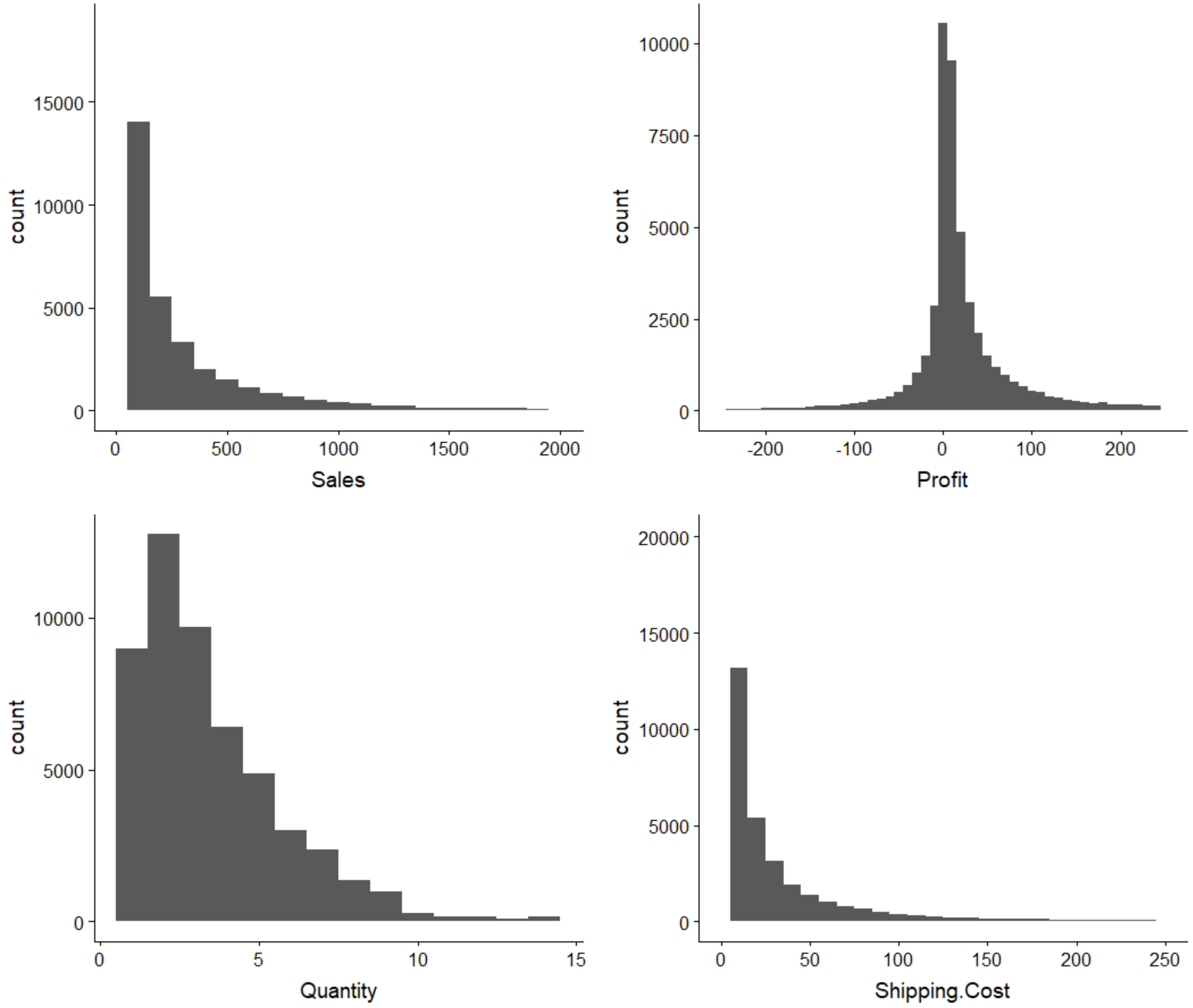
Dataset cover 51,290 records and there are NO duplicates.

Most Profitable Buckets



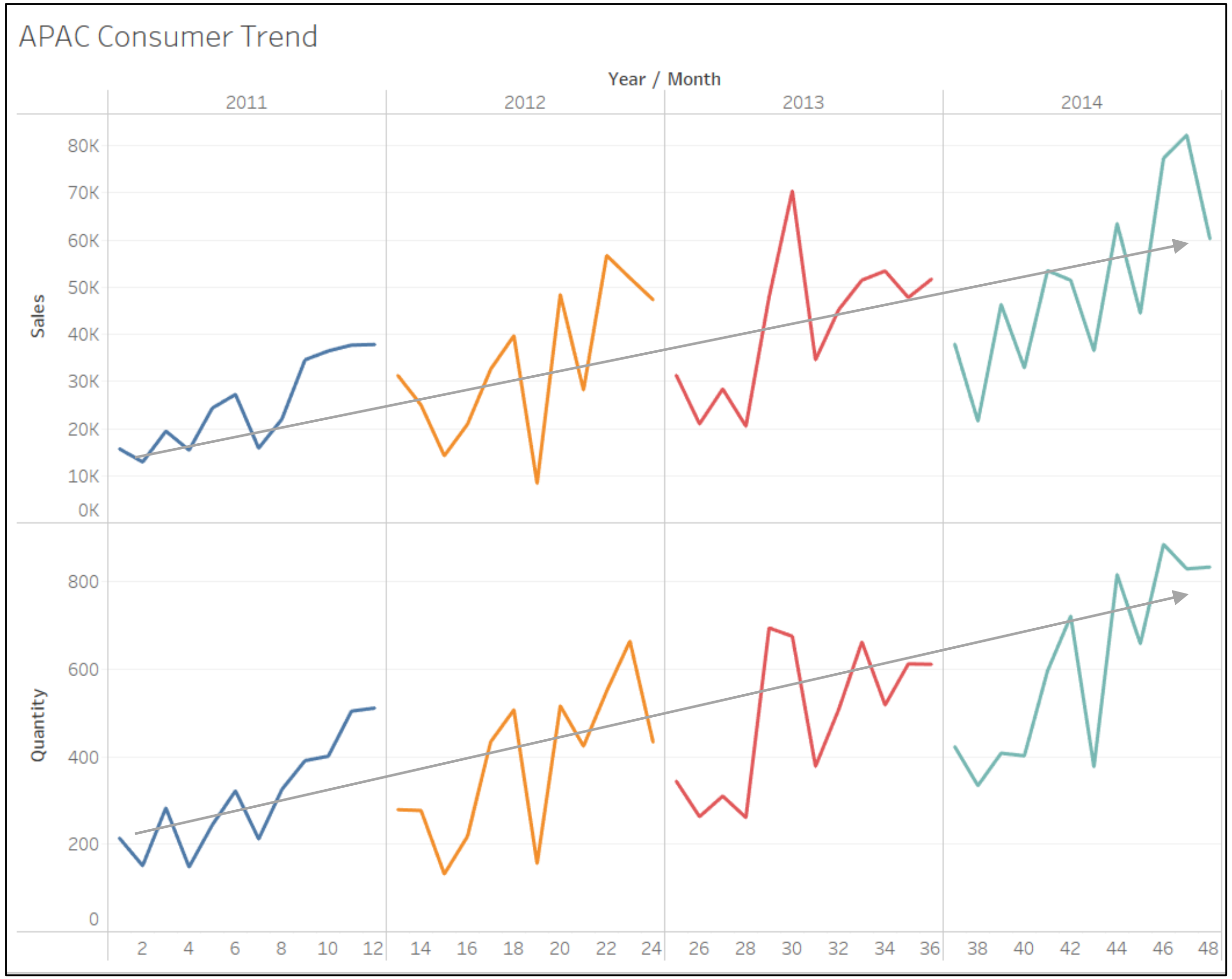
EDA- Overview of the data

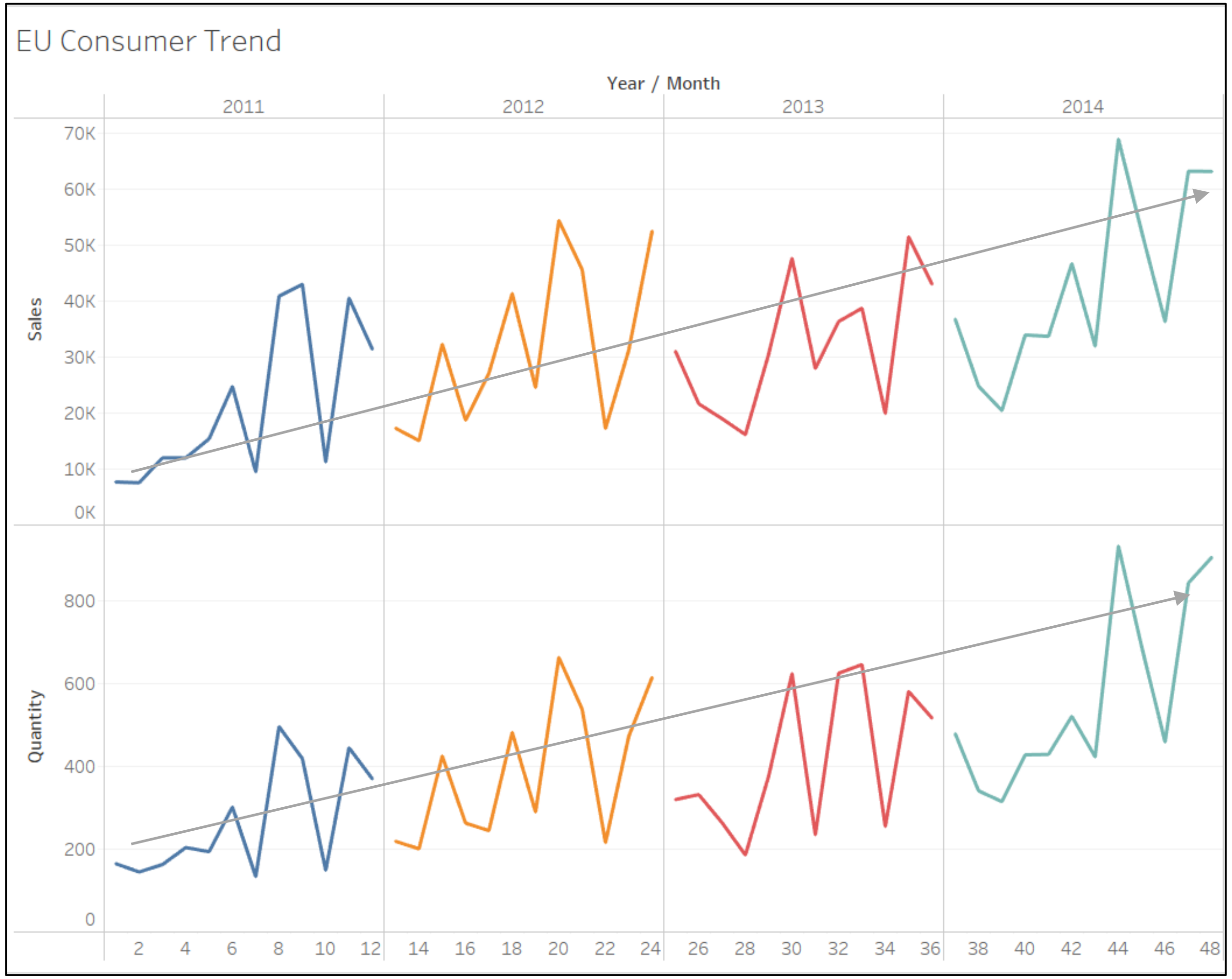
Except Profit, rest other attributes are positive variables



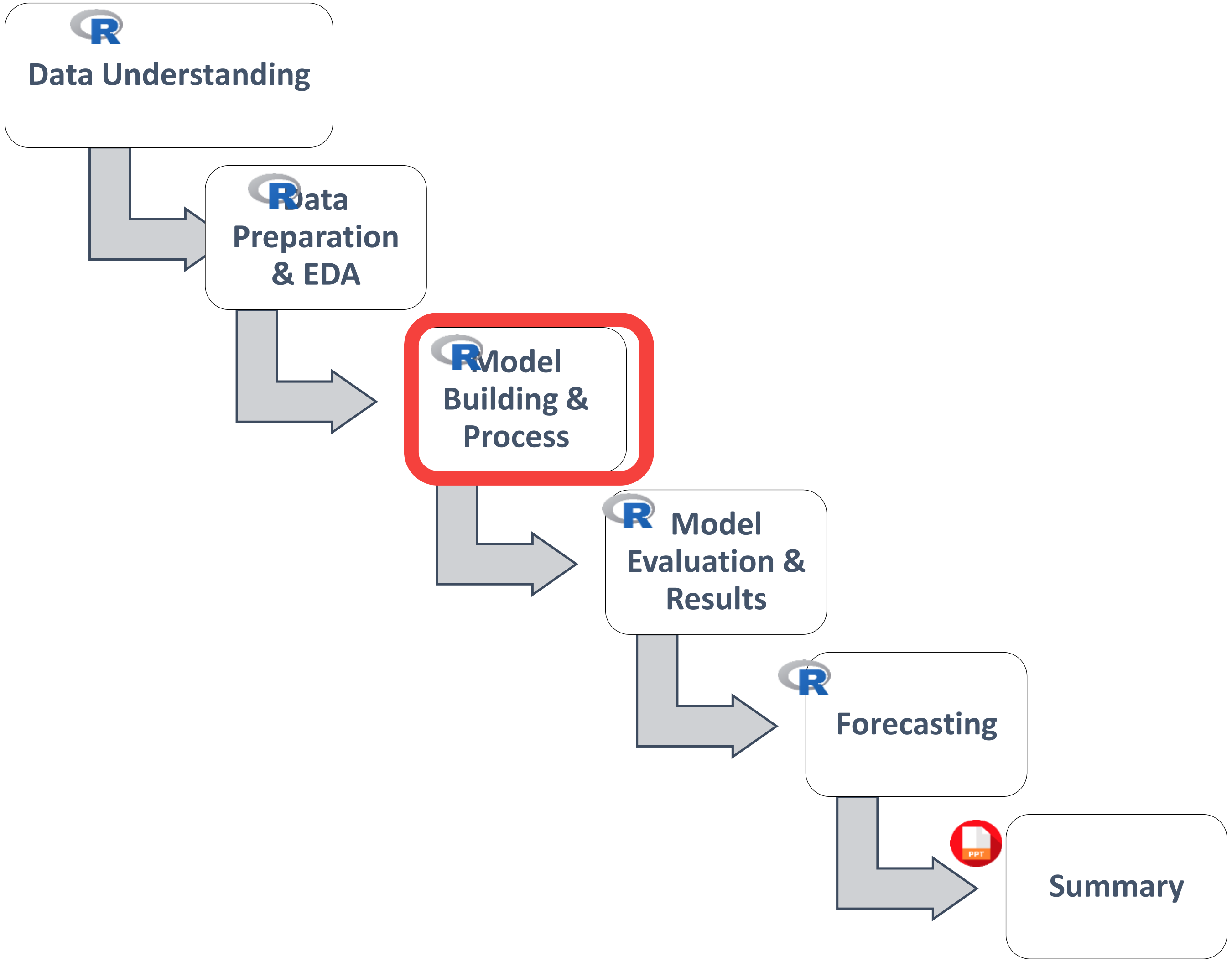
EDA- Visualise the actual data

- Data shows a linear increasing seasonal trend
- A clear drop in 'Sales' & 'Quantity' for both markets (APAC & EU) at the beginning of the each year; at same time peak occurs around the end of each year





Analysis Process



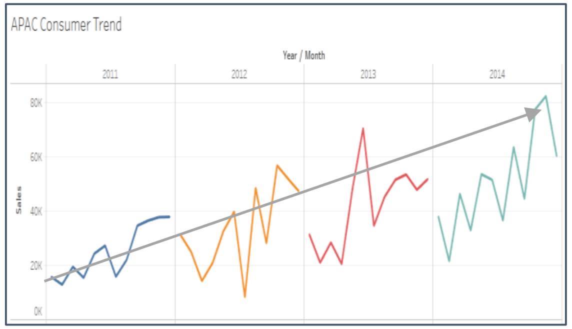
Model Building & Process

- Post Data preparations & EDA we got **“APAC and EU Consumer”** buckets are the most profitable. Forecasting will be done on these 2 buckets on ‘Sales’ & ‘Quantity’ are the attributes
- Being it’s a Times Series Forecasting Model, data has been split by
  - For Training - 1:42 Months
  - For Testing – 43:48 Months
  - For Forecasting – 49:54 Months
- Time Series Modeling has been done by 4 buckets
  - APAC Consumer Sales
  - APAC Consumer Quantity
  - EU Consumer Sales
  - EU Consumer Quantity
- Modeling & Evaluation has been done through
  - Classical decomposition Model
  - Auto-ARIMA Model

Classical decomposition Model

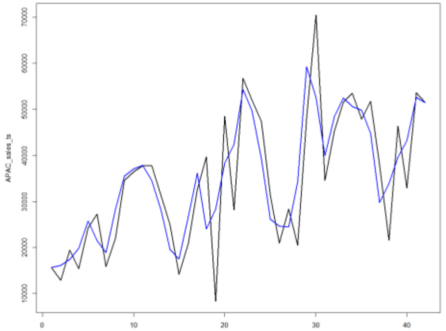
Model Process

Model has been shown here for APAC Consumer Sales as an example; and same process have been followed across all the Buckets



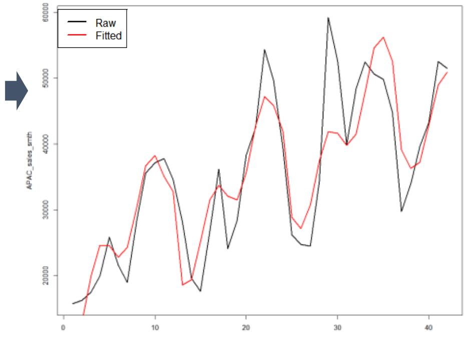
1. Visualize the Actual data *1-48 month*

- Data shows a linear increasing seasonal trend
- A clear drop in 'Sales' has been noticed at the beginning of the each year
- Hence we considered **MODULO FUNCTION** for a period of 12.
- Within YEAR seasonality has been modeled through **'SIN & COS FUNCTION'**



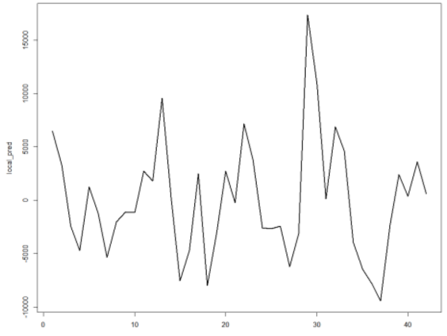
2. Smoothing the data

Train Data: 1-42 month



3. Global Prediction Seasonality & Trend fit

Train Data: 1-42 month



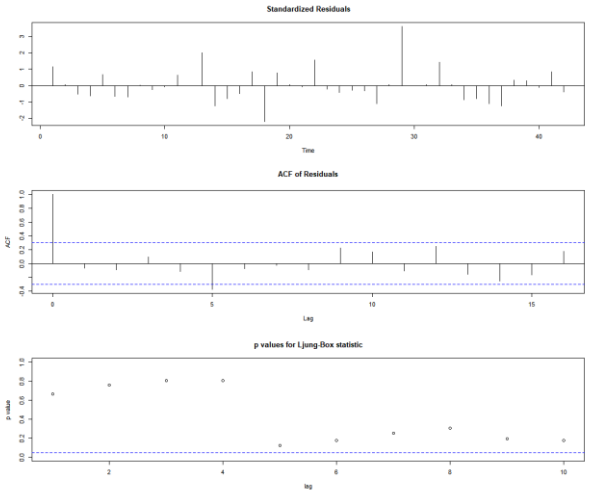
4. Local Prediction = Actual data less Global Prediction

Train Data: 1-42 month

5 – ACF & Residual

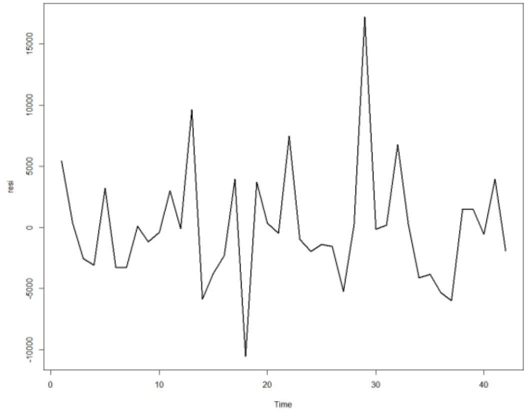
ARIMA(0,0,1) model or equivalently a MA(1) model

Train Data: 1-42 month



6. Residual plot to check randomness

Train Data: 1-42 month



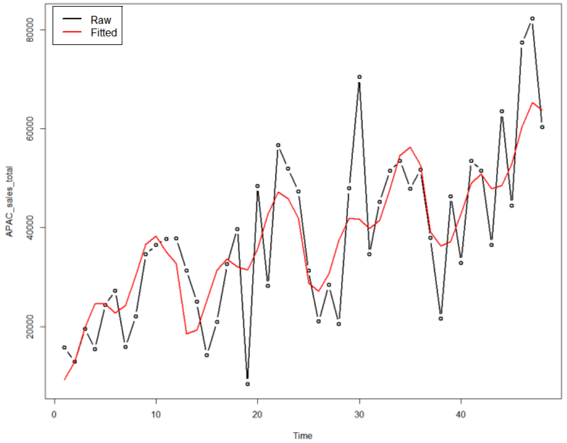
7. ADF and KPSS tests to check the Noise and result shows residuals is most likely white noise

Train Data: 1-42 month

```
> adf.test(local_pred,alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  local_pred
Dickey-Fuller = -4.0134, Lag order = 3, p-value = 0.01869
alternative hypothesis: stationary

> kpss.test(local_pred)

        KPSS Test for Level Stationarity

data:  local_pred
KPSS Level = 0.046819, Truncation lag parameter = 1, p-value = 0.1
```

8.Predict the Model on Test Data

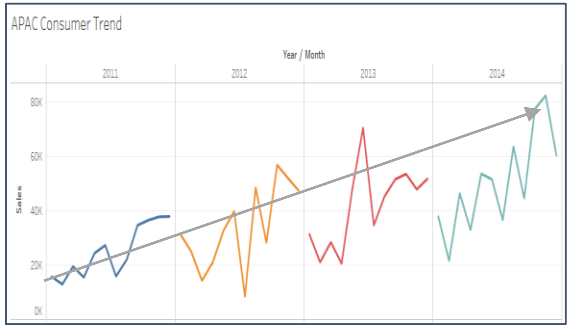
[43-48 month] **for evolution**



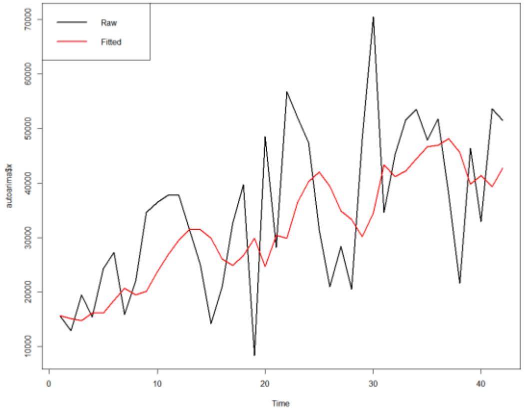
Auto-ARIMA Model

Model Process

Model has been shown here for APAC Consumer Sales as an example; and same process have been followed across all the Buckets



1. Visualize the Actual data *1-48 month*



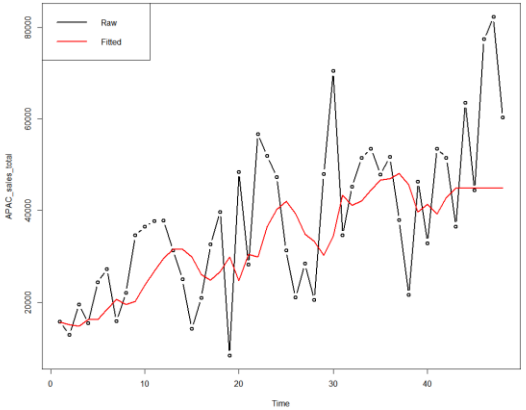
2. Model the dataset AUTO.ARIMA function

Train Data: 1-42 month

```
> resi_auto_arima <- APAC_sales_ts - fitted(autoarima)
> adf.test(resi_auto_arima,alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  resi_auto_arima
Dickey-Fuller = -4.2563, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(resi_auto_arima, alternative = "stationary") :
  p-value smaller than printed p-value
> kpss.test(resi_auto_arima)

        KPSS Test for Level Stationarity

data:  resi_auto_arima
KPSS Level = 0.042734, Truncation lag parameter = 1, p-value = 0.1

Warning message:
In kpss.test(resi_auto_arima) : p-value greater than printed p-value
```

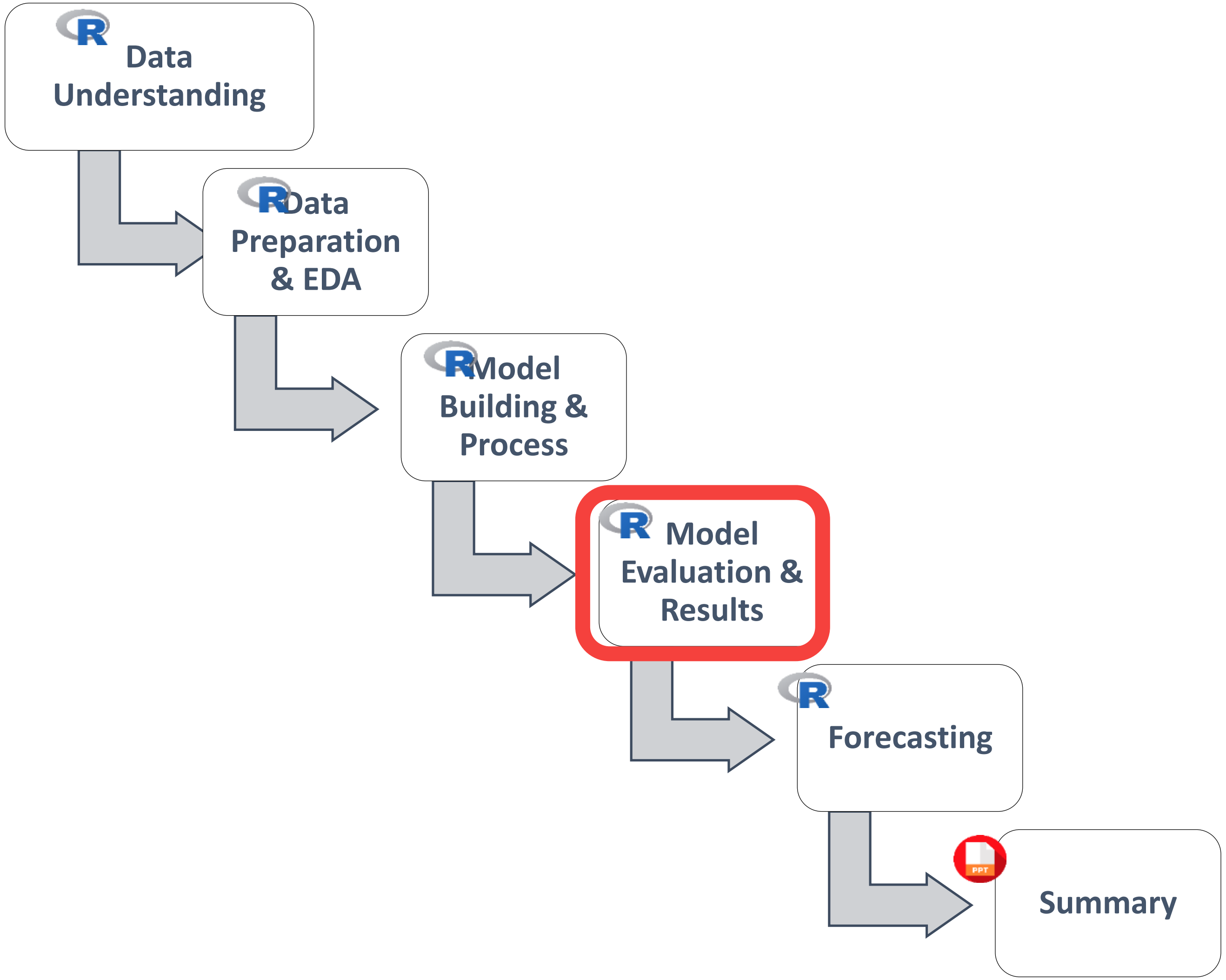
3. ADF and KPSS tests to check the Noise and result shows residuals is most likely white noise

Train Data: 1-42 month



4.Predict the Model on Test Data *[43-48th month]* **for evolution**

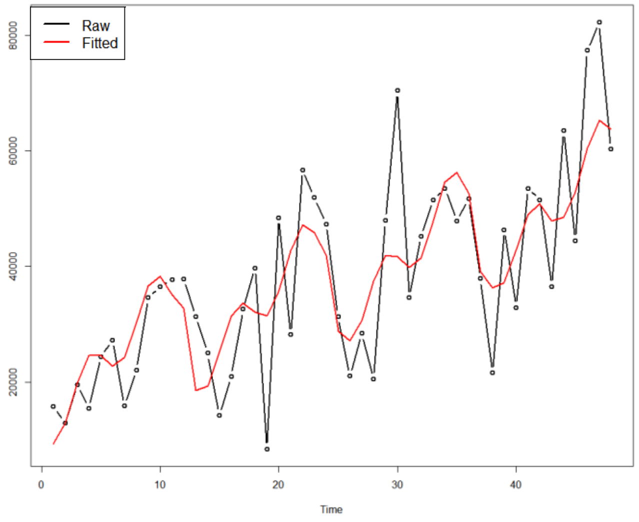
Analysis Process



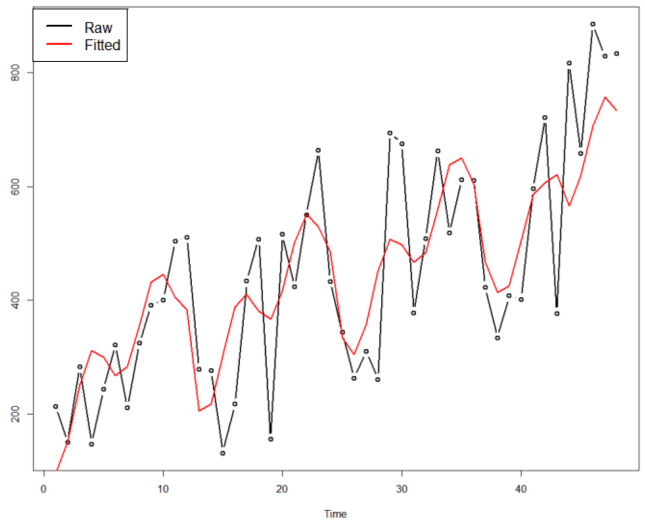
Classical decomposition Model

Model Evaluation & Results

APAC Consumer Sale



APAC Consumer Quantity



MAPE

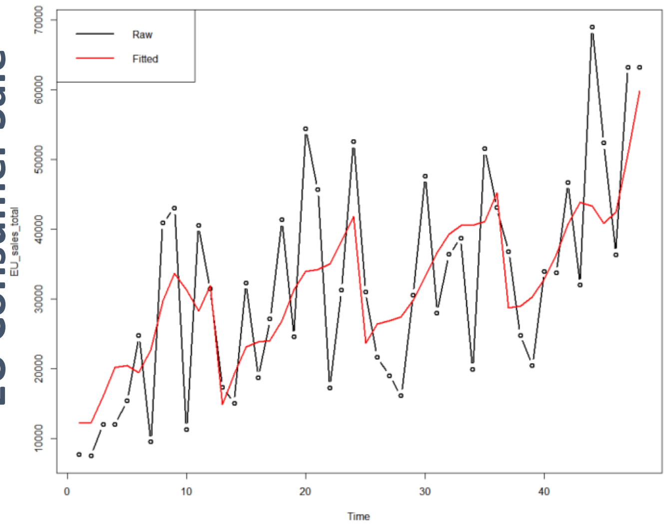
APAC Sales: ~20%

APAC Quantity: ~24%

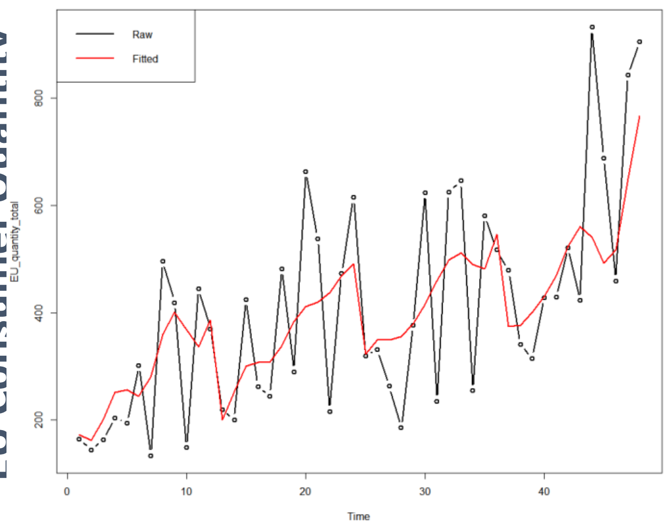
EU Sales: ~23%

EU Quantity: ~26%

EU Consumer Sale



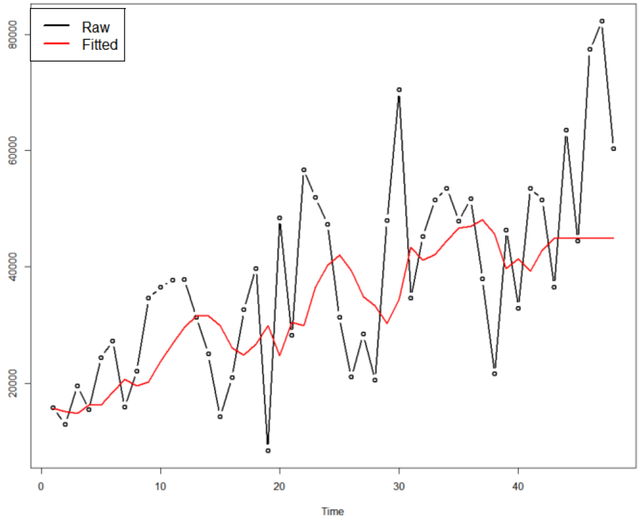
EU Consumer Quantity



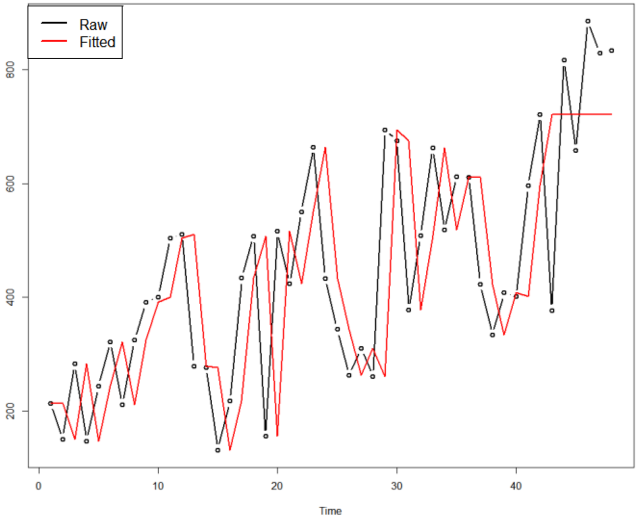
Auto-ARIMA position Model

Model Evaluation & Results

APAC Consumer Sale



APAC Consumer Quantity



MAPE

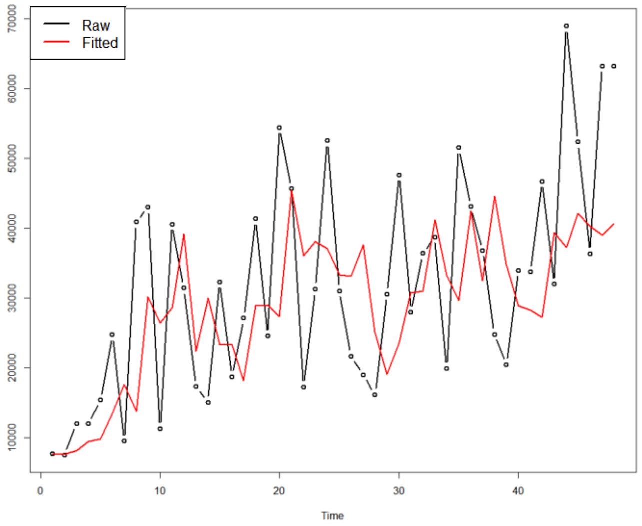
APAC Sales: ~28%

APAC Quantity: ~26%

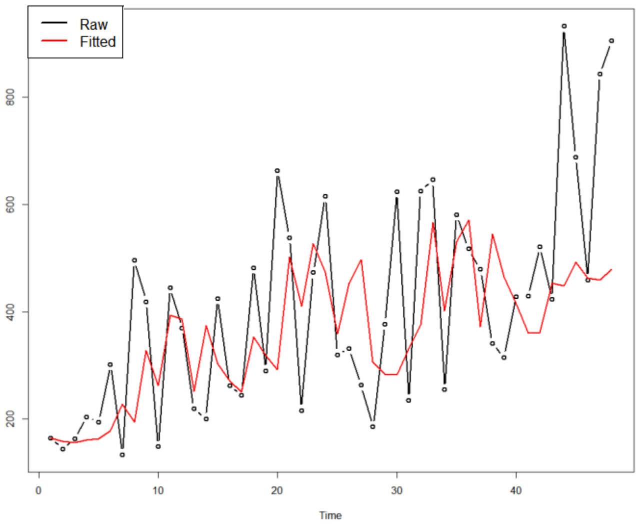
EU Sales: ~29%

EU Quantity: ~30%

EU Consumer Sale



EU Consumer Quantity

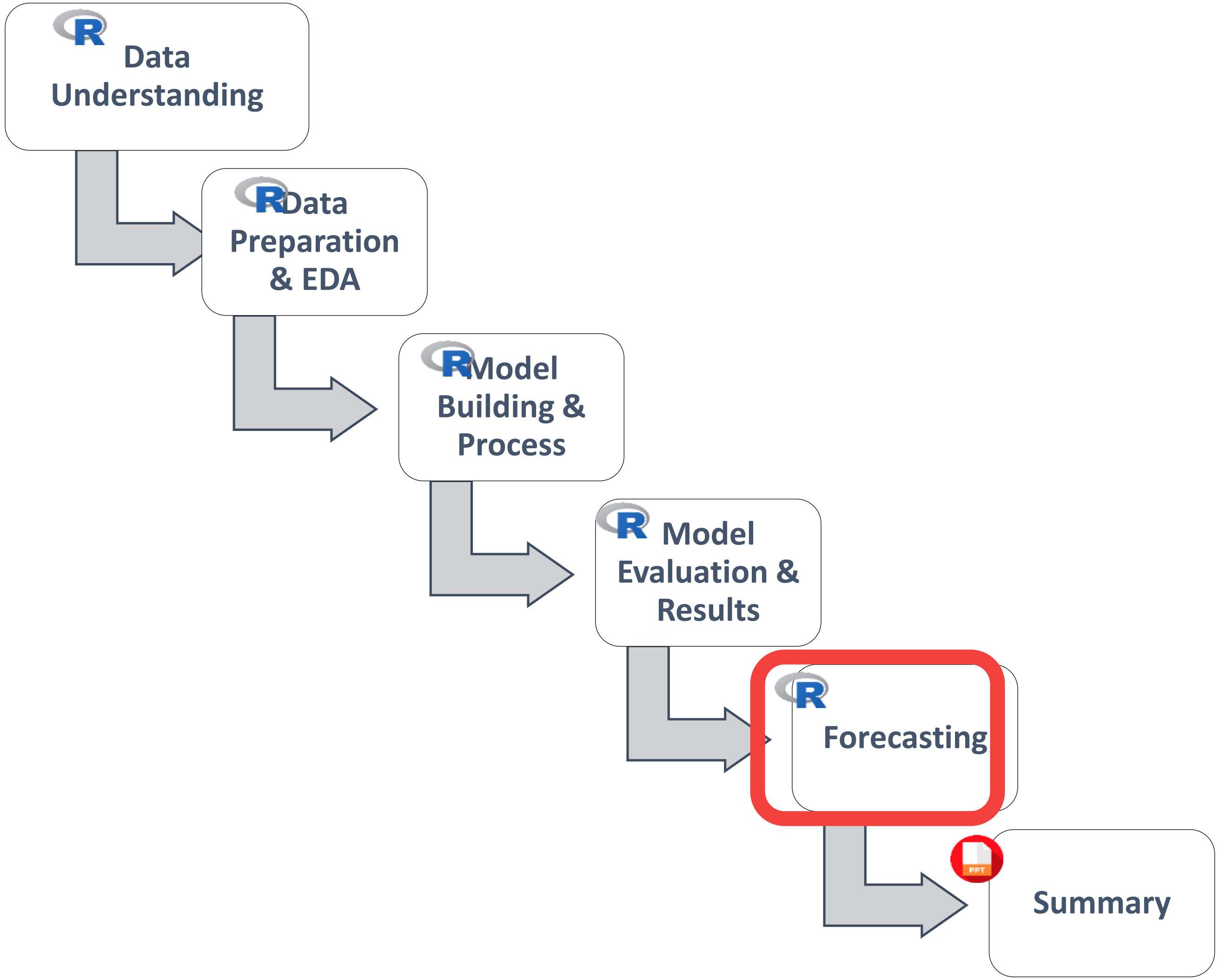


Model Evaluation & Results

Business Interpretation

- **Classical Decomposition** is giving a better MAPE value compared to the Auto-ARIMA model.
- It also seems that seasonality for the EU Market is slightly different compared to APAC.
- EU market is perhaps showing a slight quarterly seasonality on top of yearly seasonality, based on ACF plots and the Sales and Quantity vs Month plots.
- Modulo function along-with sin and cos can be tweaked to give better MAPE values for the EU market.

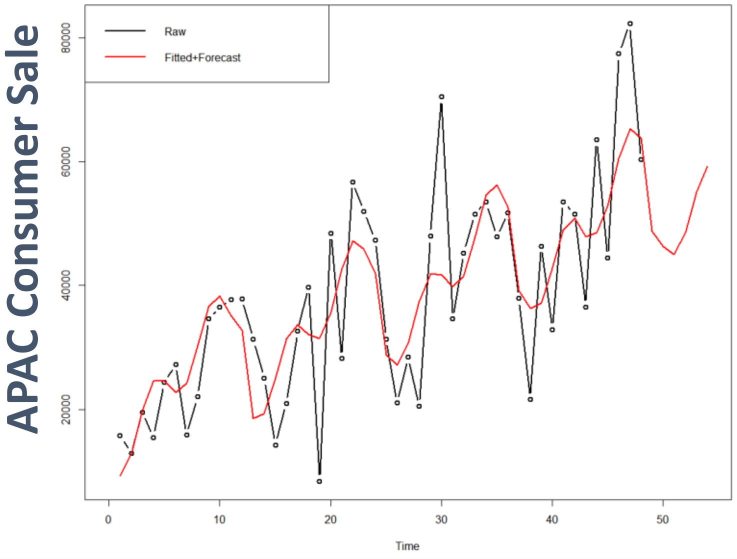
Analysis Process

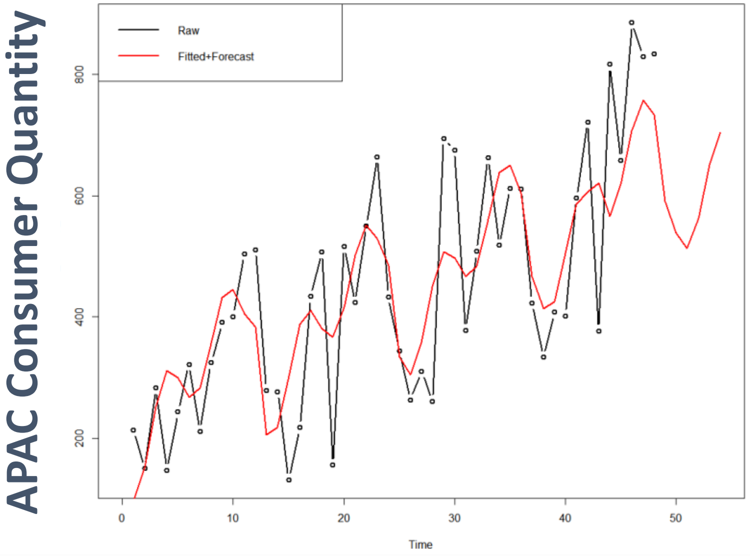


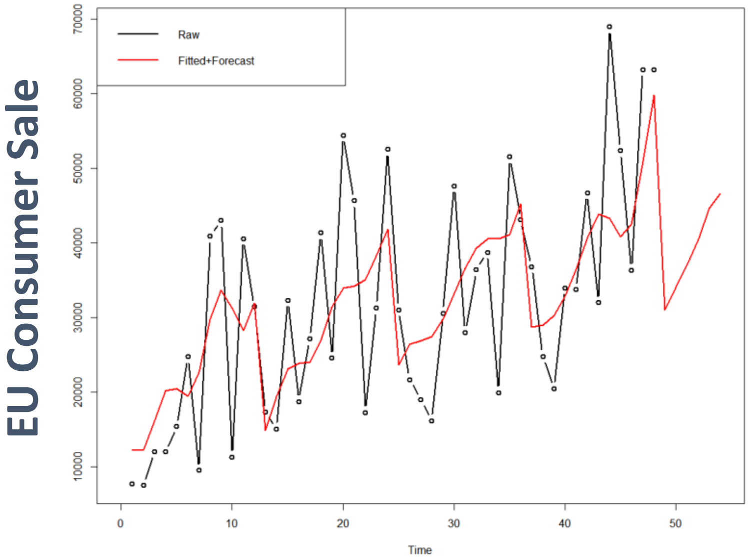
Forecasting for next 6 months [49:54 Month]

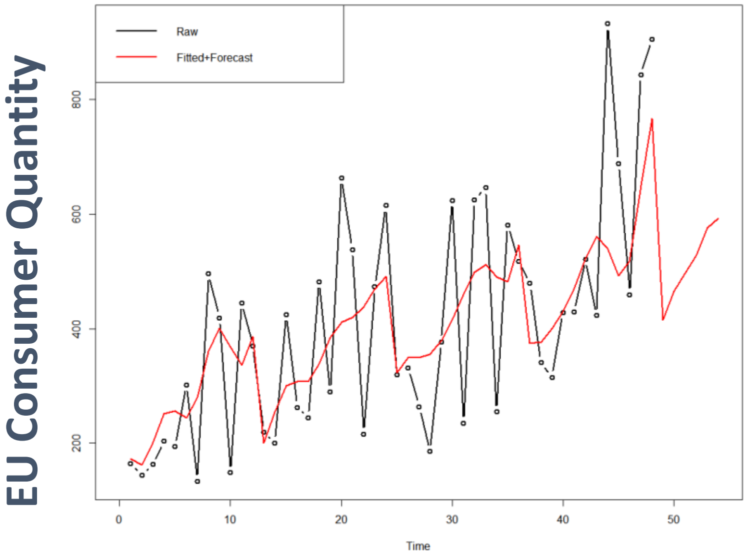
Classical Decomposition is showing better *MAPE* values during Training & Test Dataset, hence we have done the forecasting for next 6 months based on the same.

Forecasting shows, there will be a down trend for initial months but it will be recovered later part of the year.

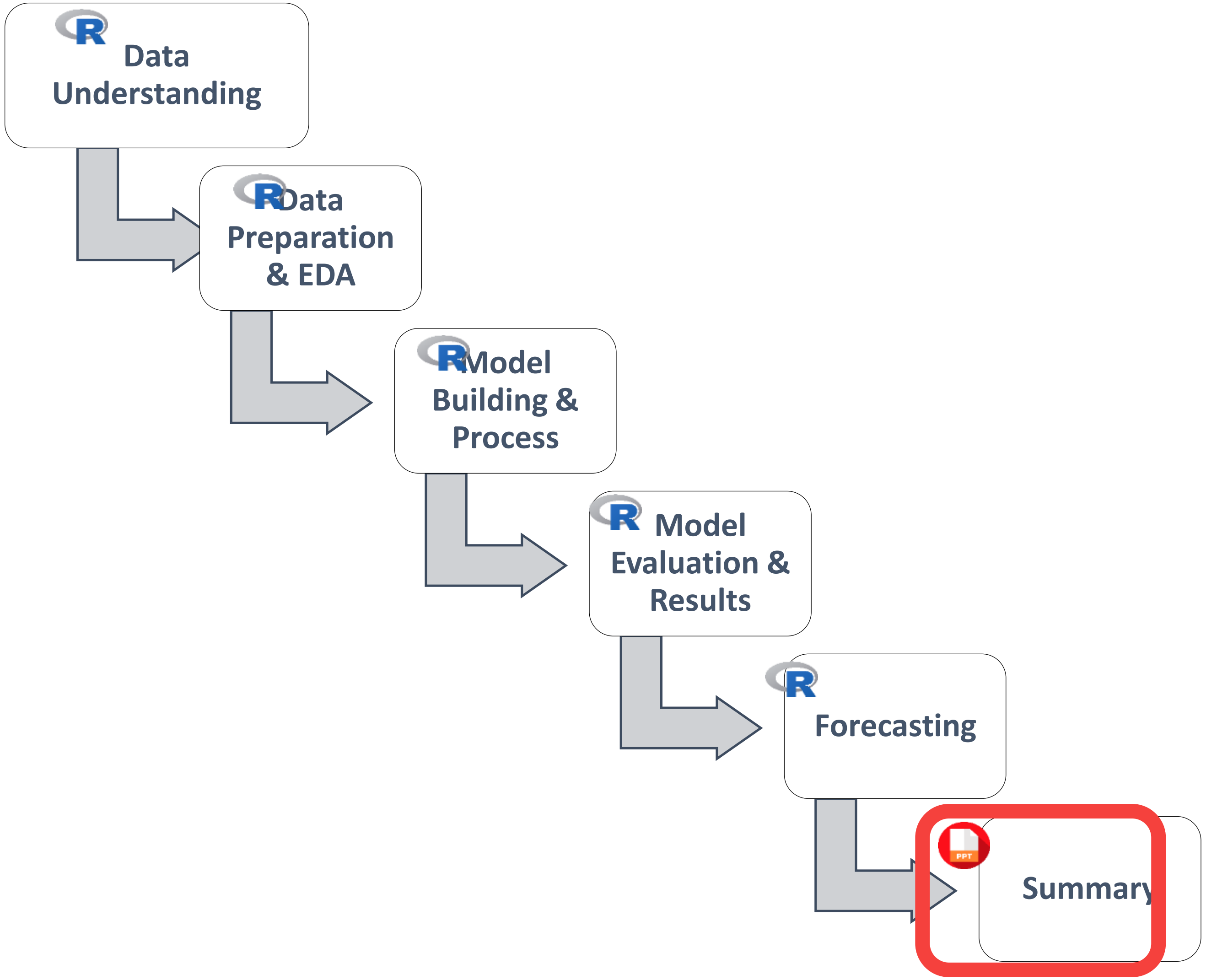








Analysis Process



Summary

Actual Data shows a linear increasing seasonal trend. A clear drop in 'Sales' & 'Quantity' for both markets (APAC & EU) at the beginning of the each year; at same time peak occurs around the end of each year.

Model Building & Evaluation

- Classical Decomposition is giving a better MAPE value compared to the Auto-ARIMA model.
- Seasonality for the EU Market is slightly different compared to APAC.
- EU market is perhaps showing a slight quarterly seasonality on top of yearly seasonality, based on ACF plots and the Sales and Quantity vs Month plots.
- Modulo function along-with sin and cos can be tweaked to give better MAPE values for the EU market.

Forecasting (49:54 Month)

- Based on Classical decomposition mode, Forecasting [48:54 month] shows a down trend for initial few month but it will be recovered later part of the year.