

## **Global Mart Sales Forecasting Case Study**

Group Name:Upgrad Pune- Bhubaneswar Cohorts

1. Rishabh Shrivastava
2. Vinod Jha
3. Saurav Kumar
4. Amitabha Banerjee

# Business Understanding & objectives from the Data analysis

## **Business Understanding –**

Global Mart” is an online super store giant having worldwide operations and with branches spread across globally. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office. Sales Manager wants to have the sales & demand forecast for coming 6 months so as to manage the revenue & stock inventory.

## **Business Objectives -**

Since the store caters to 7 different market segments and 3 major categories the objective of analysis is to find out among these 21 market buckets 2 most profitable and consistent segment and forecast the sales and demand for these 2 segments .

## **Business Constraints –**

Only 4 years of transaction data is available for analysis

## **Major deliverables from Data analysis-**

- ☐ Convert the transaction level data into time series
- ☐ Find the 2 most profitable & consistently profitable segments using the coefficient of variation of Profit
- ☐ Forecast the sales & quantity for next 6 months for the above 2 profitable segments
- ☐ Test the accuracy of the forecast

# Data Exploration & Explanation

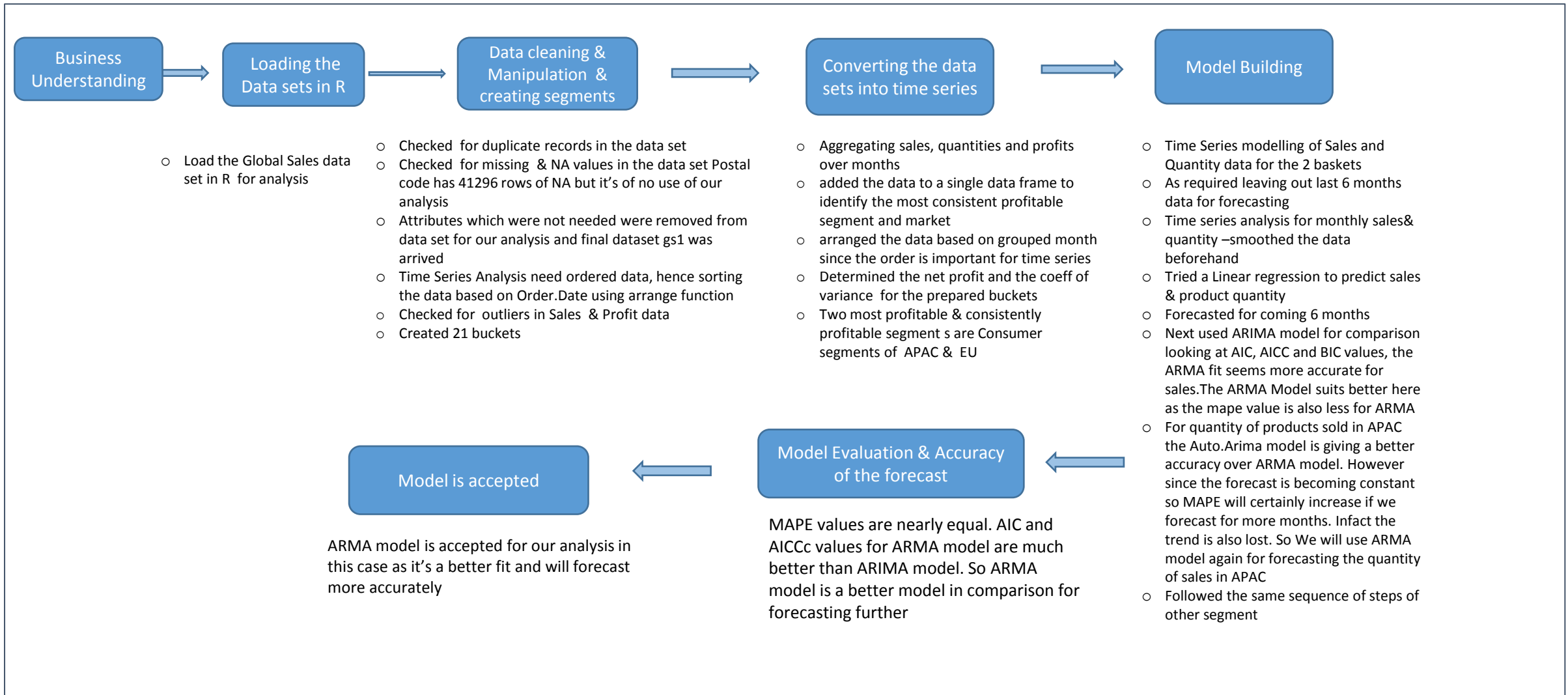
## Relevant variables for our Time series analysis from the Data Dictionary

Attributes	Description
Order Date	Date on which the order was placed
Segment	The market segment to which the product belongs
Market	Market segment to which the customer belongs
Sales	Total sales value of the transaction
Quantity	Quantity of the product ordered
Profit	Profit made on the transaction

### Observations –

- ☐ Order date is required to arrive at monthly aggregate value of sales , quantity & profit.
- ☐ Grouping by Segment and Markets as required hence they are the significant attributes
- ☐ Sales & quantity both are required for modelling and forecasting
- ☐ Our objective is to find 2 most profitable segment with consistent profit hence profit is a significant variable for data analysis

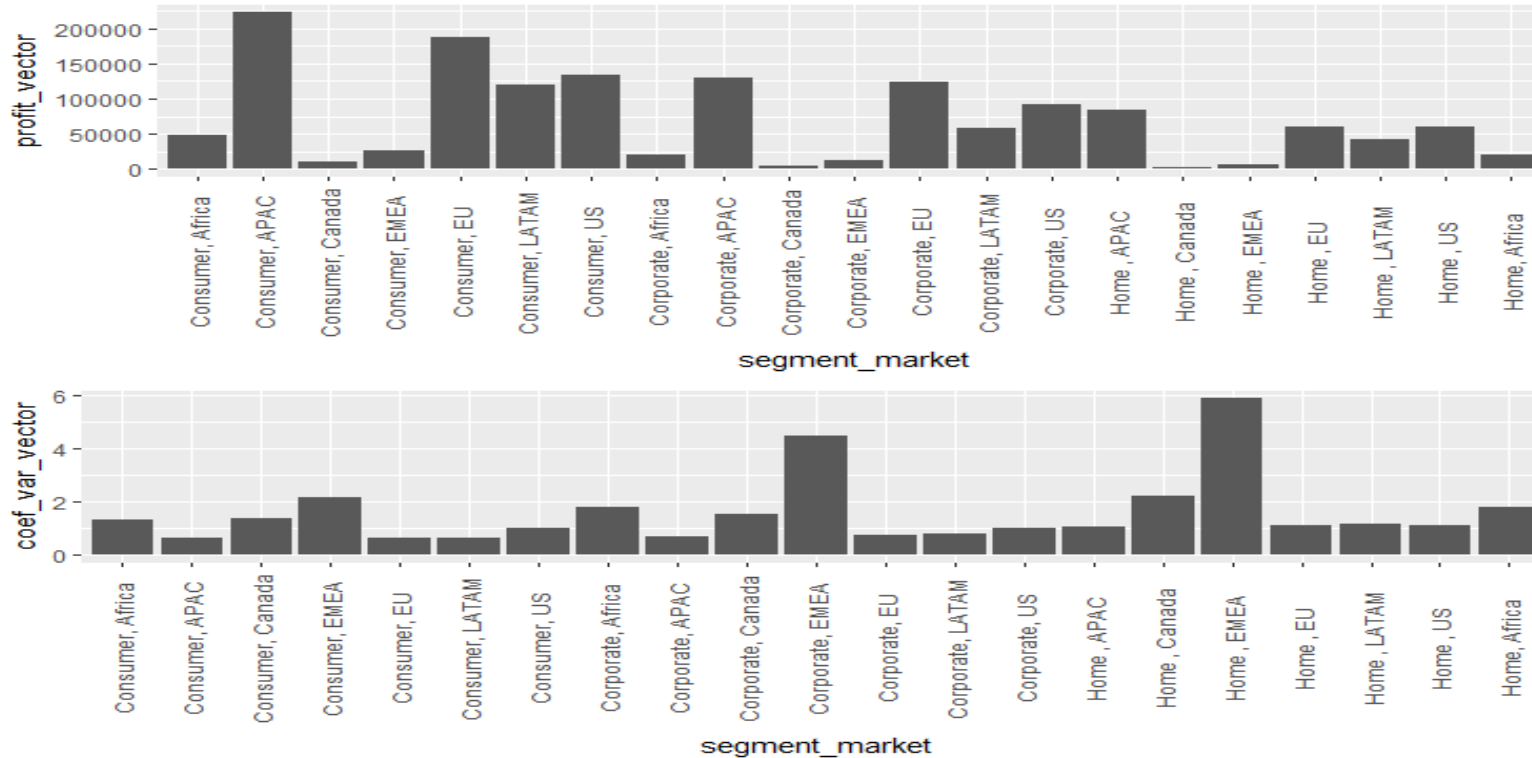
# Problem solving methodology- Process Flow



# Data Cleaning & Manipulation

- ❑ Checked for duplicate records found there are none
- ❑ Checked for missing & NA records, found that postal code has 41296 as NA. This data is not of much significance & we are not using it for our analysis hence ignoring it.
- ❑ Attributes which were not needed were removed from data set for our analysis and final dataset gs1 was arrived
- ❑ Time Series Analysis need ordered data, hence sorting the data based on Order.Date using arrange function
- ❑ Needed to aggregate on the basis of monthly profit etc, so we considered month and year of order date
- ❑ Checked for outliers in Sales & Profit data and verified that the data is correct and we can ignore the spikes
- ❑ Created 21 buckets including all the 7 market & segments
- ❑ Calculated the net profit and the coefficient of variance for all the above buckets using function sd for standard deviation and mean for average values
- ❑ Coefficient of variance is lowest for **Consumer - APAC** and **Consumer-EU** , which concludes that these two segments are most profitable and consistently profitable

# Top 4 Profitable Market Segments

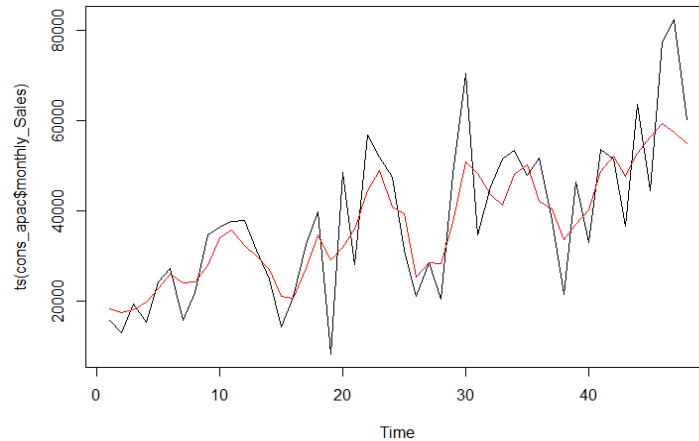


## Observations –

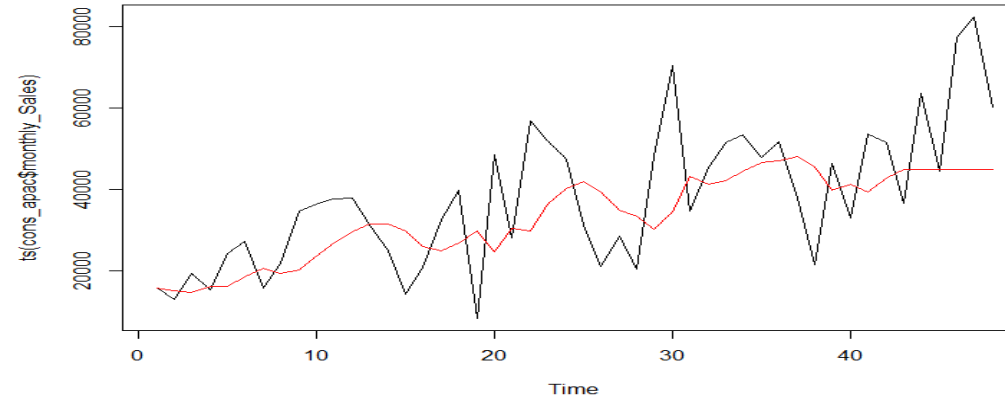
- Consumer APAC, Consumer EU, Consumer LATAM, Consumer US are the top 4 profitable market segments as obvious from the graphs
- The Coefficient of variance also suggests the same as the coefficients are less for these segments as compared to others
- Coefficient of variance is lowest for **Consumer - APAC** and **Consumer-EU** , which suggests that these two segments are the most profitable and consistently most profitable

## Time Series Analysis for monthly Sales- Consumer Segment APAC

ARMA model fit



ARIMA model fit



```
local_fit_arma
Series: local_fit
ARIMA(5,0,1) with zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1
      1.5413   -0.7321   -0.6232    0.9811   -0.4778   -0.6902
s.e.   0.1706    0.2342    0.2285    0.2372    0.1361    0.1460

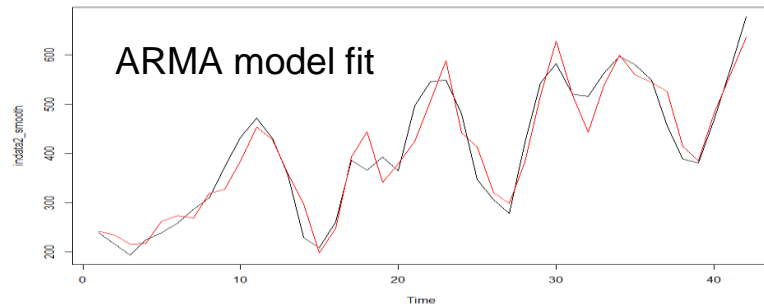
sigma^2 estimated as 16414923:  log likelihood=-406.82
AIC=827.65   AICc=830.94   BIC=839.81
>
cons_apac_sales_arma
# ARIMA(0,1,1)
#
# Coefficients:
#   ma1
#  -0.7559
# s.e.   0.1381
#
# sigma^2 estimated as 174361555:  log likelihood=-447.11
# AIC=898.23   AICc=898.55   BIC=901.66
|

Mape_val_arma1 #22.76
looking at AIC, AICc and BIC values, the ARMA fit seems more accurate
Mape_val_arma1 #27.69
```

### Observations –

- ☐ Looking at AIC, AICc and BIC values, the ARMA fit seems more accurate
- ☐ MAPE value ARMA- 22.76 & MAPE value ARIMA -27.69
- ☐ On the basis of MAPE values ARMA is a better model for forecasting sales in APAC segment

# Time series analysis for monthly quantity sold in APAC



```
local_fit_arma
Series: local_fit
ARIMA(5,0,1) with zero mean

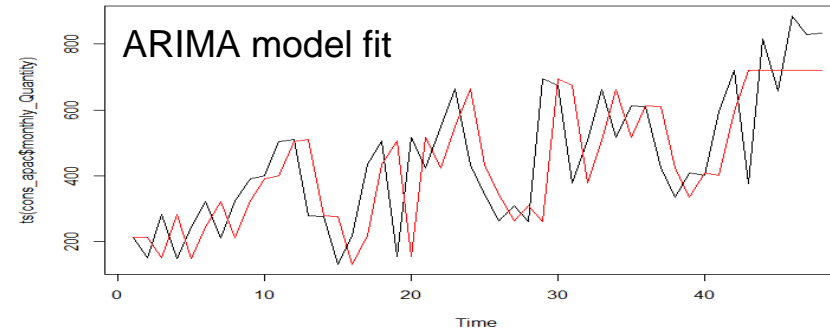
Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1
s.e.  1.5413  -0.7321  -0.6232  0.9011  -0.4778  -0.6902
      0.1706   0.2342   0.2285  0.2372   0.1361   0.1460

sigma^2 estimated as 16414923:  log likelihood=-406.82
AIC=827.65   AICc=830.94   BIC=839.81

cons_apac_quantity_arima
Series: indata2
ARIMA(0,1,0)

sigma^2 estimated as 25366:  log likelihood=-266.07
AIC=534.14   AICc=534.24   BIC=535.85

Mape_val_arma #36.12
Mape_val_arima #26.24
```

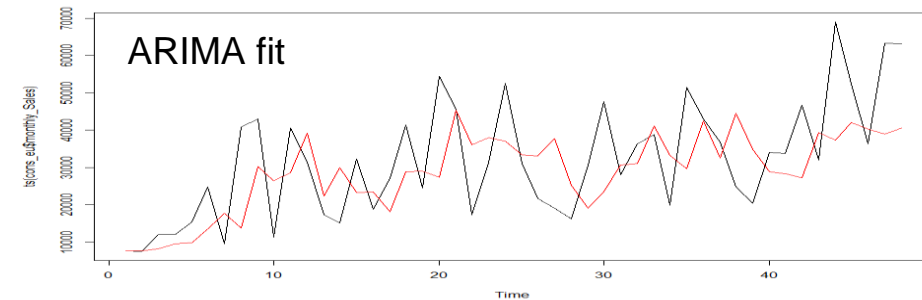
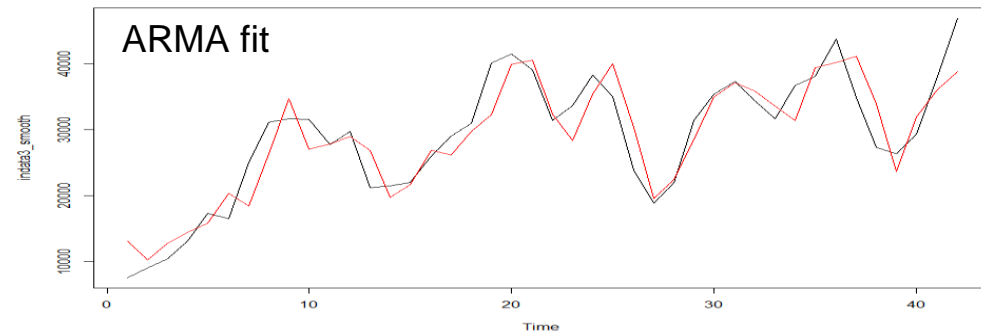


## Observations –

In this scenario the Auto.Arima model is giving a better accuracy over ARMA model. However since the forecast is becoming constant so MAPE will certainly increase if we forecast for more months. Infact the trend is also lost. So we will use ARMA model again for forecasting the quantity of sales in APAC



# Time Series Modelling for the sales of EU



```
local_fit_arma
Series: local_fit
ARIMA(5,0,1) with zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1
      1.5413  -0.7321  -0.6232  0.9011  -0.4778  -0.6902
s.e.    0.1706    0.2342    0.2285    0.2372    0.1361    0.1460

sigma^2 estimated as 16414923:  log likelihood=-406.82
AIC=827.65   AICc=830.94   BIC=839.81
```

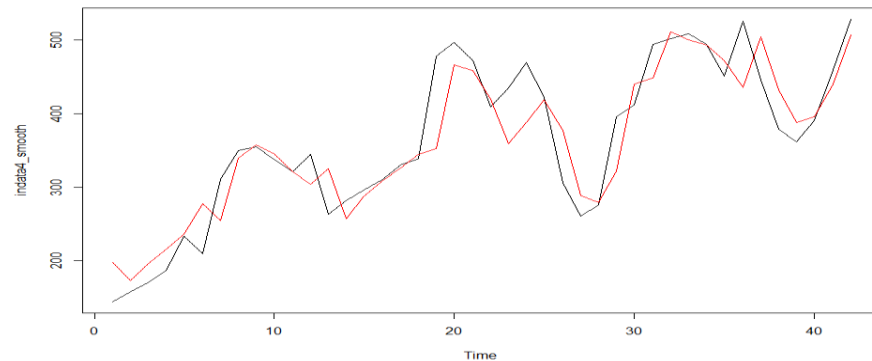
```
cons_eu_sale_arma
# ARIMA(2,1,0)
#
# Coefficients:
#   ar1      ar2
#  -0.5796  -0.4906
# s.e.    0.1346   0.1310
#
# sigma^2 estimated as 168564623:  log likelihood=-445.84
# AIC=897.67   AICc=898.32   BIC=902.81
|
Mape_val_arma #30.9
```

## Observations-

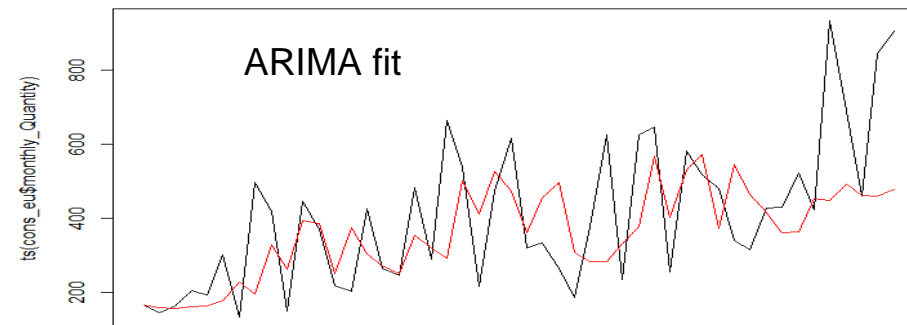
- ☐ Mape\_val\_arma #30.9
- ☐ Mape\_val\_arma #28.92
- ☐ The ARIMA model is giving a better prediction according to the MAPE value for forecasting the sales for EU market segment
- ☐ But AIC values is less for ARMA model .

# Modelling the time series for the Quantity of EU

ARMA fit



ARIMA fit



```
local_fit_arma
# ARIMA(2,0,1) with zero mean
#
# Coefficients:
#   ar1    ar2    ma1
# 1.5434 -0.7719 -0.7242
# s.e.   0.1209   0.1010   0.1451
#
# sigma^2 estimated as 2024:  log likelihood=-218.65
# AIC=445.3   AICc=446.39   BIC=452.26
```

```
-cons_eu_quantity_arima
# ARIMA(2,1,0)
#
# Coefficients:
#   ar1    ar2
# -0.7359 -0.5879
# s.e.    0.1224   0.1185
#
# sigma^2 estimated as 21185:  log likelihood=-261.9
# AIC=529.8   AICc=530.44   BIC=534.94
```

## Observations -

Mape values are nearly equal(Mape\_val\_arma #31.12 &Mape\_val\_arima #30.13) AIC and AICc values for ARMA model is much better than ARIMA model. So ARMA model is a better model for forecasting further

## Model Evaluation-MAPE

- ☐ For Sales & Quantity sold forecast in Consumer Segment of APAC, ARMA model is more suitable one as compared to ARIMA on the basis of lower MAPE values
- ☐ For Sales & Quantity sold forecast in Consumer segment of EU, ARIMA model is more suitable as compared to ARMA on the basis of lower MAPE values

## Conclusion-

Trend noticed in Sales for all the market segments

Consumer APAC, Consumer EU, Consumer LATAM, Consumer US are the top 4 most profitable market segments as per our analysis

On the basis of Coefficient of variance Consumer - APAC and Consumer-EU are the two most profitable and consistently profitable segments

For Sales & Quantity sold forecast in Consumer Segment of APAC, ARMA model is more suitable one as compared to ARIMA on the basis of lower MAPE values

For Sales & Quantity sold forecast in Consumer segment of EU, ARIMA model is more suitable as compared to ARMA on the basis of lower MAPE values

# Recommendations

- Since we have closed numbers for forecasted values, company can maintain the appropriate inventory considering the situation.
- The prediction do not consist the white noise, hence the actual numbers will be different, the prediction is just to understand the trend and help company to maintain the adequate inventory and resources and be ready for short term market.
- Prediction Number also tell about the approximate sales, so to increase sales company may introduce some offer for high sales.