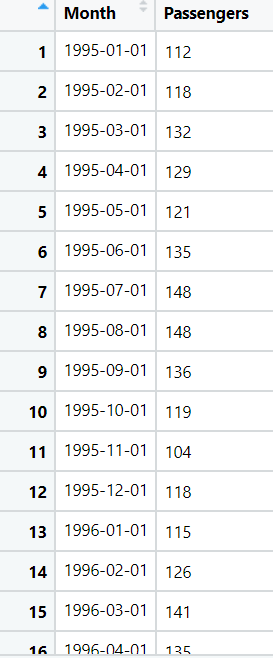
# Topic: Forecasting – Time Series

Forecast the Coca-Cola prices and Airlines Passengers data set. Prepare a document for each model explaining how many dummy variables you have created and RMSE value for each model. Finally which model you will use for Forecasting.

## 1.) Airlines.xlsx

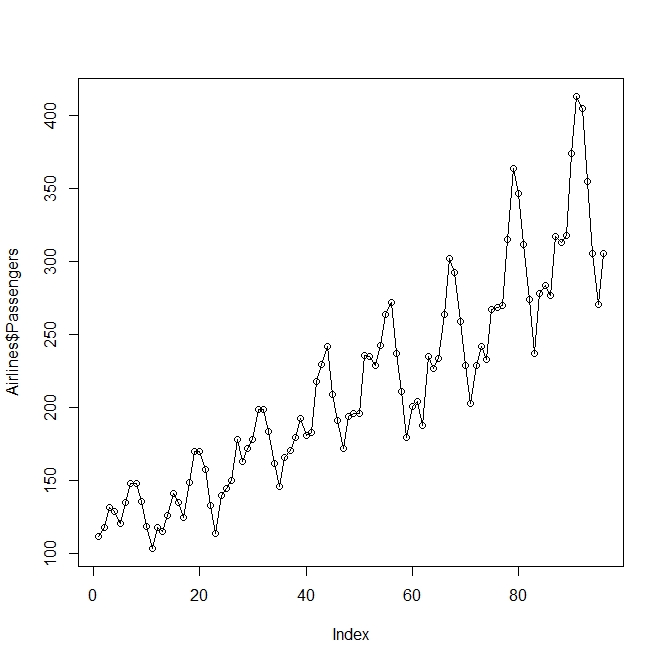


**Ans:**

* The dataset file is converted to csv as to remove the time stamp error in python.

**Data Preprocessing:**

* The dataset consists of 96 observations with 2 variables of Month-Year and passengers list.
* Below is the graph that represents passengers list for the given period of time



* Checking the NA values, as there are no NA values no further imputation is required.
* As the data is of 12 months, so the frequency is taken as 12 and created 12dummy variables
* Now the months names are assigned to the columns names of dummy variables and combined with the airline data.
* A time column “t” is assigned to the dataset and taking the log of passengers and square of t, so that the whole data is normalized

**Splitting the data:**

* Now the data is divided into training and test data with [1:84] and [85:96]
* Calculating the RMSE value using different models, as we cannot directly tell what is the exact trend followed by the data
* Only residual values are calculated for all the models to calculate the RMSE values

**Different Models to Calculate RMSE:**

**Linear Model:**

* RMSE is 53.19 and Adjusted R2 Value is 0.79

**Exponential Model:**

* RMSE is 46.05 and Adjusted R2 is 0.82, RMSE has reduced of the exponential model than linear.
* As predicted values are logged values, we do exponential of expo\_pred$fit to get actual values

**Quadratic Model:**

* RMSE is 48.05 and Adjusted R2 Value is 0.79

**Additive Seasonality Model:**

* RMSE is 132.8 and Adjusted R2 is 0.16. Hence, it may not be additive seasonality model.

**Additive Seasonality with Linear Model:**

* RMSE is 35.34 and Adjusted R2 is 0.94

**Additive Seasonality with Quadratic Model:**

* RMSE is 26.36 and Adjusted R2 is 0.95

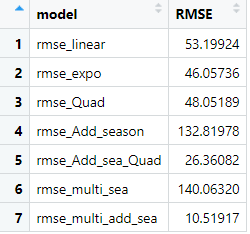
**Multiplicative Seasonality Model:**

* In multiplicative we multiply but we can't multiply directly hence we apply log
* RMSE is 140.06 & Adjusted R2 is 0.15

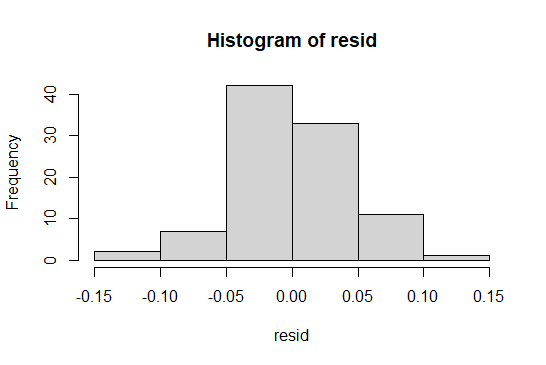
**Multiplicative Seasonality Linear trend:**

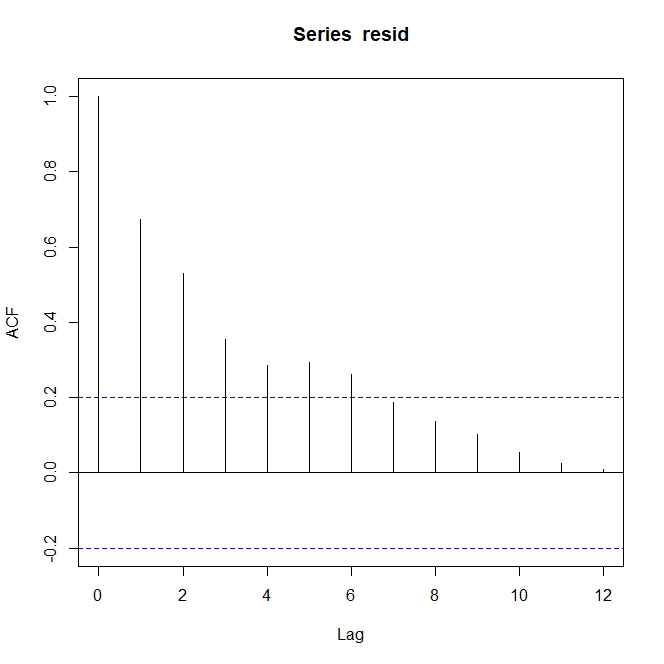
* RMSE is 10.51 and Adjusted R2 is 0.97
* This is the highest R2 & lowest RMSE

**Below are the over all RMSE values for all models**



* Here we find that Multiplicative Seasonality with Linear trend has least RMSE value of 10.51
* Now building the model with the whole dataset of airlines
* Histogram of residual values

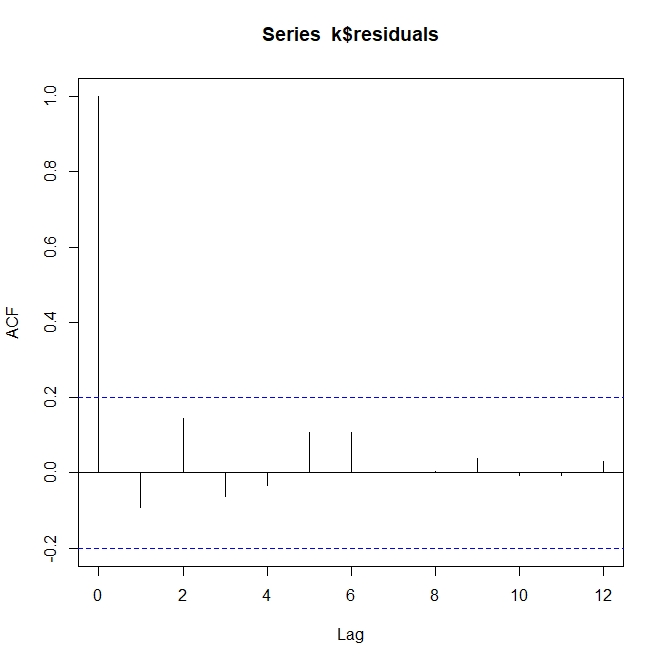




* From the above residual graph, we can say that the lag 1 to 6 are significant, so Arima can be built

**Building the Arima Model:**

* Auto regression is only used to forecast errors
* Performing auto regression with 2nd lag, p=2,d=0,q=0



* From the above graph, we can say that significance problem is removed & all are below threshold ACF values.

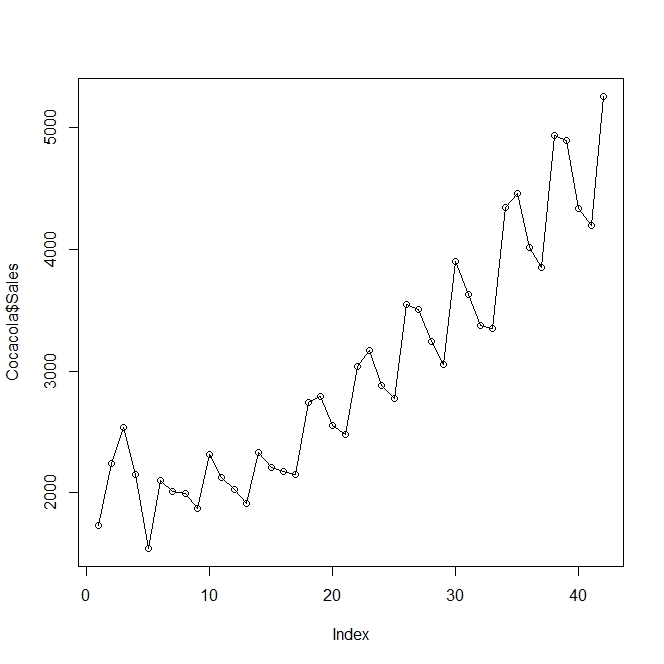
**2.) CocaCola\_Sales\_RawData.xlsx**



**Ans:**

**Data Preprocessing:**

* The dataset consists of 42 observations with 2 variables of Month-Year and Sales list.
* Below is the graph that represents sales for the given period of time



* Checking the NA values, as there are no NA values no further imputation is required.
* As the data is of 4 quarters, so the frequency is taken as 4 and created 4 dummy variables
* Now the quarters names are assigned to the columns names of dummy variables and combined with the data.
* A time column “t” is assigned to the dataset and taking the log of sales and square of t, so that the whole data is normalized.

**Splitting the data:**

* Now the data is divided into training and test data with [1:36] and [37:42]
* Calculating the RMSE value using different models, as we cannot directly tell what is the exact trend followed by the data
* Only residual values are calculated for all the models to calculate the RMSE values

**Different Models to Calculate RMSE:**

**Linear Model:**

* RMSE is 667.42 and Adjusted R2 Value is 0.7922

**Exponential Model:**

* RMSE is 526.76 and Adjusted R2 is 0.80, RMSE has reduced of the exponential model than linear.
* As predicted values are logged values, we do exponential of expo\_pred$fit to get actual values

**Quadratic Model:**

* RMSE is 485.14 and Adjusted R2 Value is 0.85

**Additive Seasonality Model:**

* RMSE is 1895.559 and Adjusted R2 is 0.11. Hence, it may not be additive seasonality model.

**Additive Seasonality with Linear Model:**

* RMSE is 555.34 and Adjusted R2 is 0.87

**Additive Seasonality with Quadratic Model:**

* RMSE is 283.06 and Adjusted R2 is 0.95

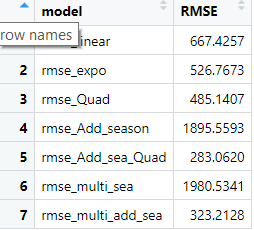
**Multiplicative Seasonality Model:**

* In multiplicative we multiply but we can't multiply directly hence we apply log
* RMSE is 1980.53 and Adjusted R2 is 0.05

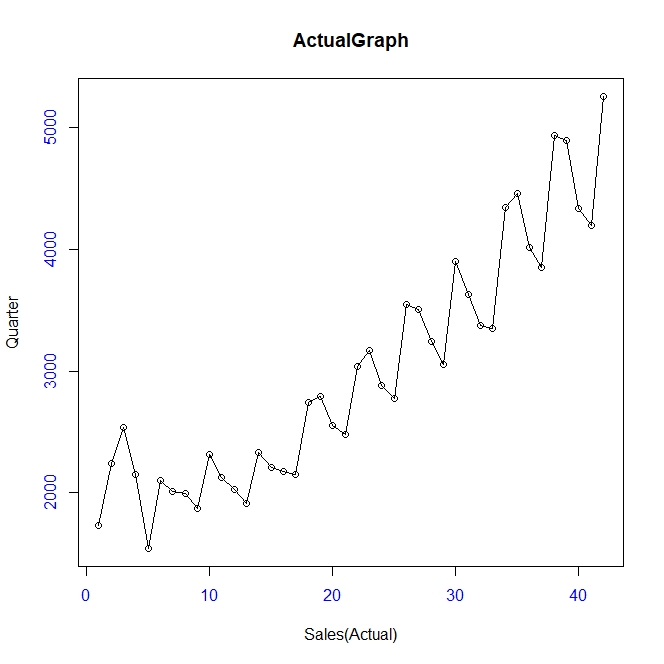
**Multiplicative Seasonality Linear trend:**

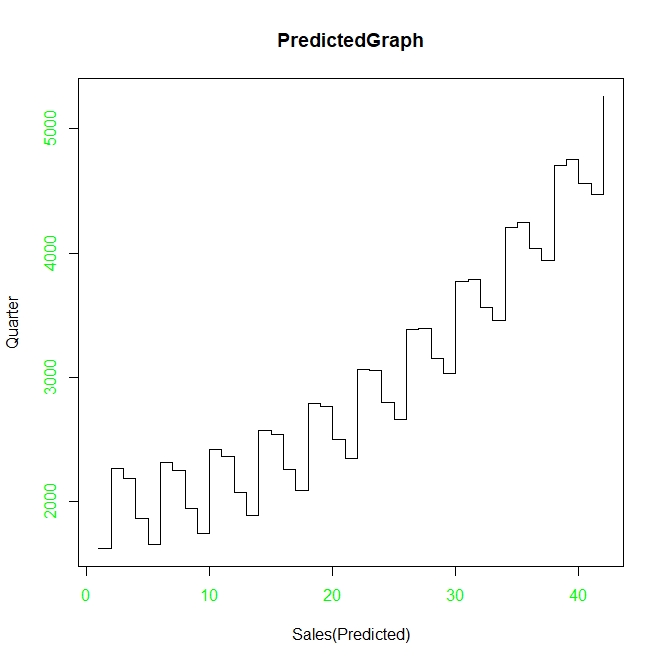
* RMSE is 323.21 and Adjusted R2 is 0.90
* This is the highest R2 & lowest RMSE

**Below are the overall RMSE values for all models**



* Here we find that Multiplicative Seasonality with Linear trend has least RMSE value of 10.51
* Now building the model with the whole dataset of airlines





# Hints:

1. Business Problem
   1. Objective
   2. Constraints (if any)
2. Data Pre-processing

2.1 Feature Engineering, EDA etc.

1. Model Building
   1. Partition the dataset
   2. Model(s) – Work with all the models (linear, exponential, quadratic etc.)
   3. Model(s) Improvement steps
   4. Model Evaluation
   5. Python and R codes
2. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

## Note:

1. For each assignment the solution should be submitted in the format
2. Research and Perform all possible steps for improving the model(s) accuracy & reduce the RMSE (also evaluate errors like MAPE, MAE etc.)
3. All the codes (executable programs) are running without errors
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here