# A picture containing shape, arrow Description automatically generatedTopic: Forecasting – TimeSeries

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name:** Nithin Dsouza

**Batch Id:** 05012021-10AM

**Topic: Forecasting – Time Series**

1. **Business Problem**
   1. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.**

**Using R and Python codes perform:**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

**3.2 Outlier Imputation if applicable**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Identify trend**
   3. **Identify seasonality**
2. **Model Building:**
   1. **Perform Forecasting on the given datasets Data Driven and Moving averages.**
   2. **Identify and apply Exponential smoothing, Model Based, ARIMA models on the dataset.**
   3. **Briefly explain the output in the documentation for each step as explained in the class.**
3. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

# A picture containing shape, arrow Description automatically generatedNote:

The assignment should be submitted in the following format:

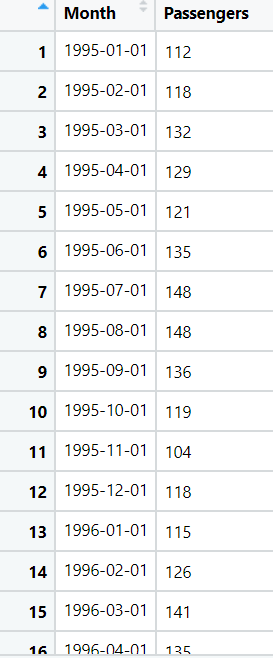
* R code
* Python code
* Code Modularization should be maintained
* Documentation of the model building (elaborating on steps mentioned above)

**Problem Statement:-**

1. The dataset consists of monthly totals of international airline passengers, from 1995 to 2002.Main aim is to predict number of passengers for next five years by using time series forecasting and prepare a document for each model explaining how many dummy variables you have created and RMSE value for each model.

## File:- Airlines.xlsx





1. **Business Problem**
   1. **Objective:** Maximize the accuracy in predicting future values of the time series variable.
   2. **Constraints (if any):** Problems in analyzing casual patterns are difficult.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| Month | month data for 5 years | Continuous, Ratio | Relevant, Provides useful information. |
| Passengers | Passengers data of each month | Discrete, count | Relevant, Provides useful information. |

**Problem Statement: -**

1. The dataset consists of quarterly sales of Coca-Cola data from 1986 to 1996.Main aim is to predict next two years of sales by using time series forecasting and prepare a document for each model explaining how many dummy variables you have created and RMSE value for each model.

**File:- CocaCola\_Sales\_RawData.xlsx**

1. **Business Problem**
   1. **Objective:** Maximize the accuracy in predicting future values of the time series variable.
   2. **Constraints (if any):** Problems in analyzing casual patterns are difficult.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| **Quarter** | Quarter data | Continuous, Ratio | Relevant, Provides useful information. |
| **Sales** | sales data of each Quarter | Discrete, count | Relevant, Provides useful information. |



**Problem Statement: -**

A Plastic manufacturing plant has recorded their sales data from 1949 to 1953 monthly. Perform Forecasting on the data and bring out insights from it, forecast the sale for the next year.

Plastic Sales.csv

1. **Business Problem**
   1. **Objective:** Maximize the accuracy in predicting future values of the time series variable.
   2. **Constraints (if any):** Problems in analyzing casual patterns are difficult.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| Month | month data | Continuous, Ratio | Relevant, Provides useful information. |
| Sales | sales data of each month | Discrete, count | Relevant, Provides useful information. |

A picture containing table

Description automatically generated

**Problem Statement: -**

Solar power consumption has been recorded by city counsels on a regular interval of times, the reason behind doing so is to understand how business are using solar power so that they can cut down on nonrenewable source of energy and shift towards renewable energies. Based on the data build a forecasting model and provide your insights on it.

Solarpower.csv

1. **Business Problem**
   1. **Objective:** Maximize the accuracy in predicting future values of the time series variable.
   2. **Constraints (if any):** Problems in analyzing casual patterns are difficult.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| date | date of each month | Continuous, Ratio | Relevant, Provides useful information. |
| cum\_power | cumulative power of each day | Continuous, interval | Relevant, Provides useful information. |

A picture containing table

Description automatically generated