TIME SERIES PROJECT

AND

BUSINESS REPORT

NAME- R INDUJA

PGP-DSBA ONLINE

26-06-2023

INTRODUCTION

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

In this module we are going to do the same and understand the given data by making predictions with TSF methods.

Problem 1- You are an analyst in the IJK shoe company and you are expected to forecast the sales of the pairs of shoes for the upcoming 12 months from where the data ends. The data for the pair of shoe sales have been given to you from January 1980 to July 1995.

Problem 2 - You are an analyst in the RST soft drink company and you are expected to forecast the sales of the production of the soft drink for the upcoming 12 months from where the data ends. The data for the production of soft drink has been given to you from January 1980 to July 1995.

Solution- For the above-mentioned problem we are going to perform EDA, splitting data into train and test, **build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data, to analyse the given reports.**

Solution 1 - First and the foremost thing we have to do is loading the important libraries in both the cases we are going to perform the same analysis

Important

Libraries are as follows:

# loading packages

# basic + dates

import numpy as np

import pandas as pd

from pandas import datetime

from datetime import datetime

# data visualization

import matplotlib.pyplot as plt

import seaborn as sns # advanced vizs

%matplotlib inline

from pylab import rcParams

from sklearn.metrics import mean\_squared\_error

# statistics

from statsmodels.distributions.empirical\_distribution import ECDF

# time series analysis

from statsmodels.tsa.seasonal import seasonal\_decompose

from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf

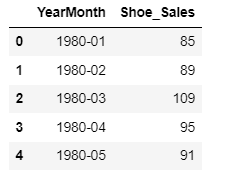
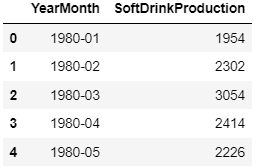
import warnings

warnings.filterwarnings("ignore")

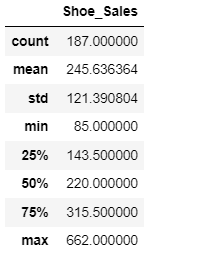
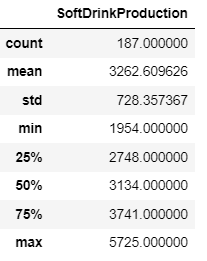
we have imported the libraries as per the requirement.

Now, let’s move towards performing and EDA **is applied to investigate the data and summarize the key insights. It will give you the basic understanding of your data, it's distribution, null values and much more.**

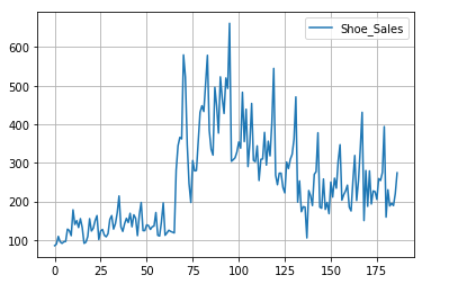
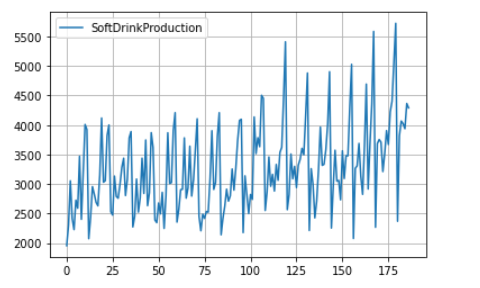
**Let’s see the results for the same first let’s see the data of Shoe sales and parallelly soft drink production data, by doing this we can see the first half of the data the columns and row we have in data, but this is not all we cannot see the entire rows and columns and thus we still need the summaries form of data. Let’s see further.**

Now we will see the summarized data of the modules, now it is clearly visible we can see the total count of the data, mean , median and standard deviation of the data.

Lets see the data in plot for both the module.

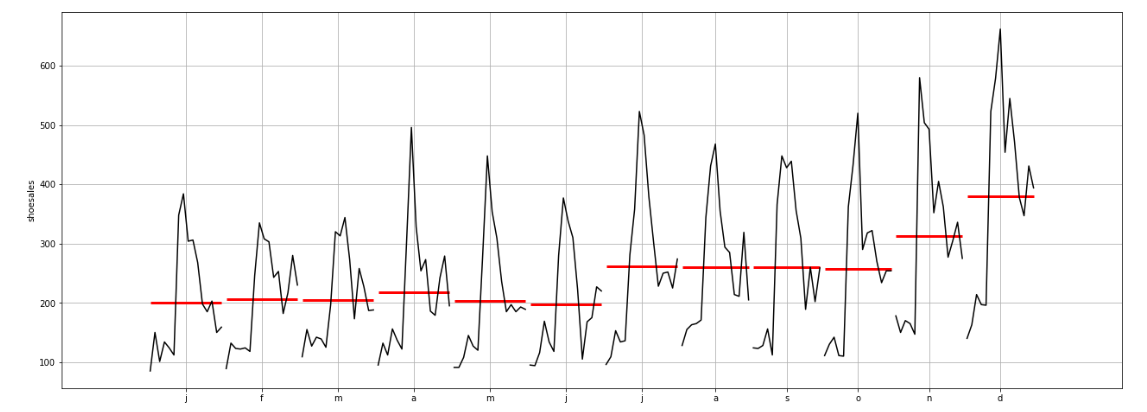
We can see the variation of data in the first graph it is explained graphically the variation of the shoe sales in each quarter. On the second hand we can see the graphical variation of soft drink production.

Shoe sales is in increasing wave in between 75-100 quadrant and on the other side we can see the increasing and decreasing variation of soft drink production

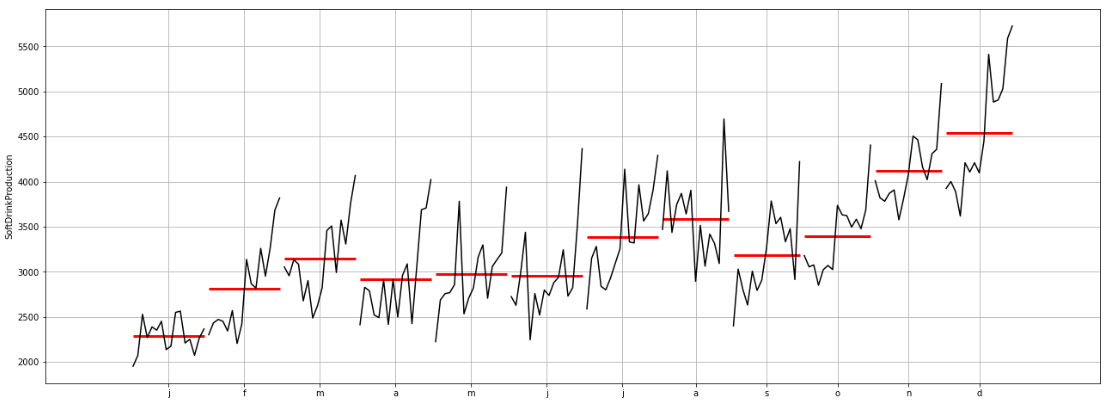
We cannot exactly imagine the production of shoes and soft drink in the plot mentioned above. Let’s move forward with the other insights.

Let’s plot the data monthly, to know the monthly production of shoes and soft drink production.

Monthly Shoe Sales Plot

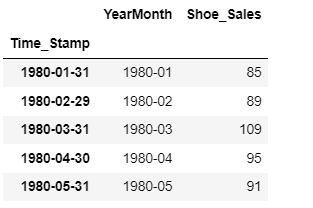
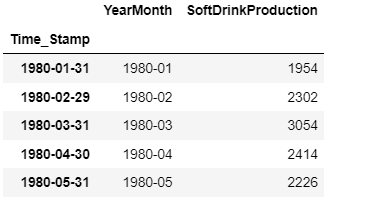


Monthly Soft drink production



On the above-mentioned plot, we cannot identify the monthly sales of the shoe sales and soft drink production.

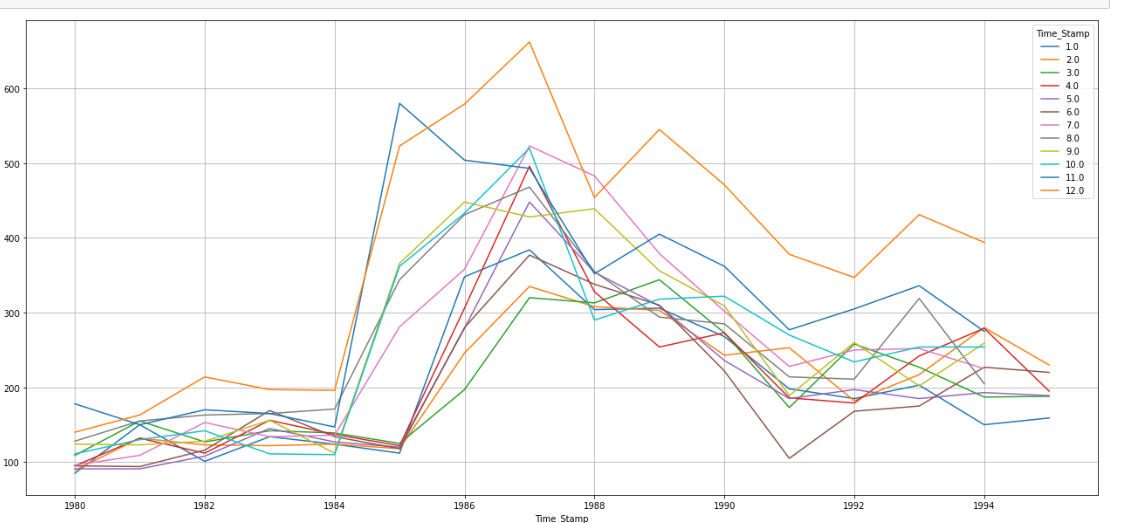
# Now plot the time series according to different months for different year to know the detailed production of shoes and soft drink month wise in each year.

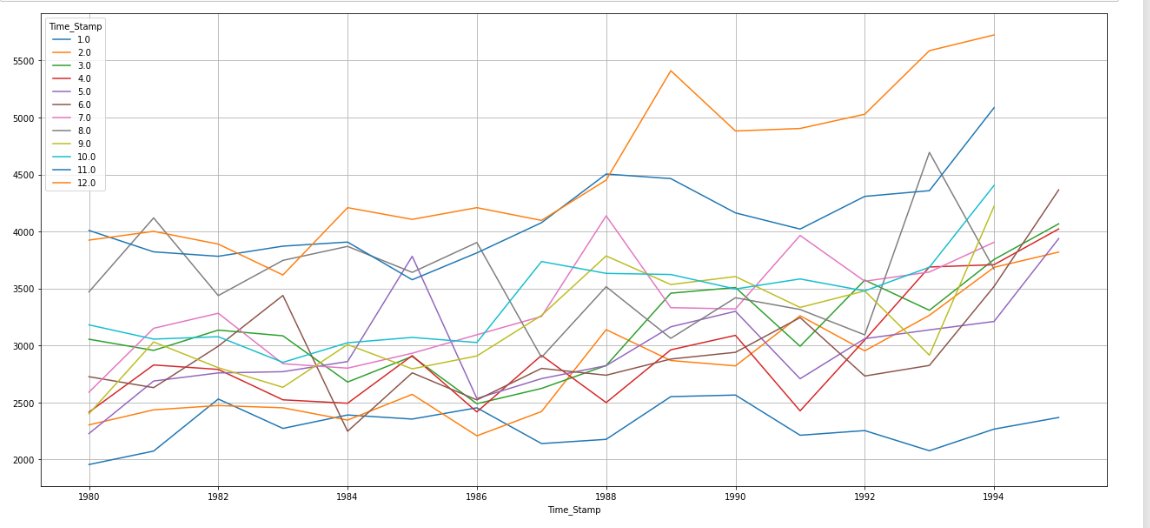
In this image we can see the monthly production of shoes and soft drink clearly in each year. Let’s

Move to see the clearer insights of the module, lets see the data into plot.

Shoe sales

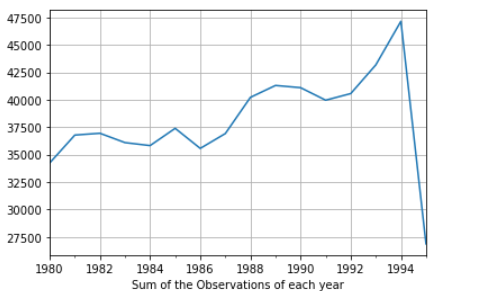
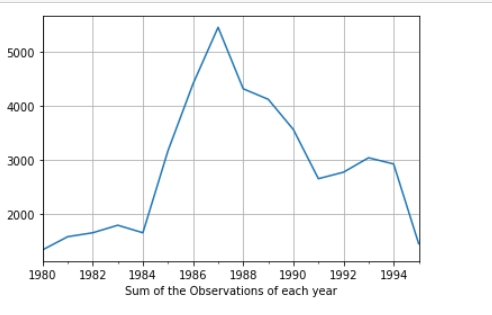


Soft drink production



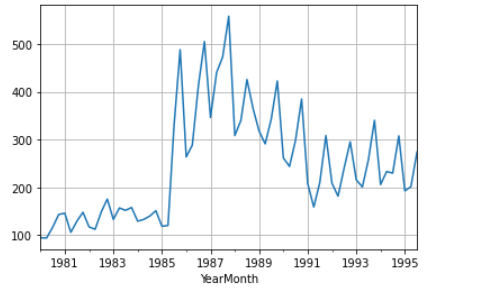
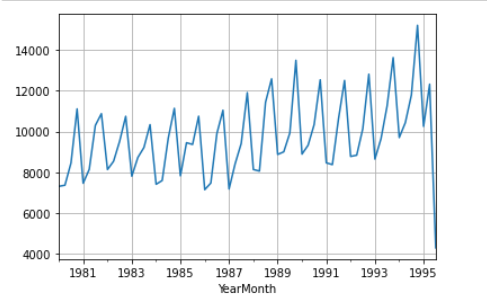
We do not have the clear insights on the above-mentioned plot so we move to see the data into yearly quarterly and decade wise data to analyse the sales.

Let’s first see the yearly insights of data for shoe sales and soft drink productions.

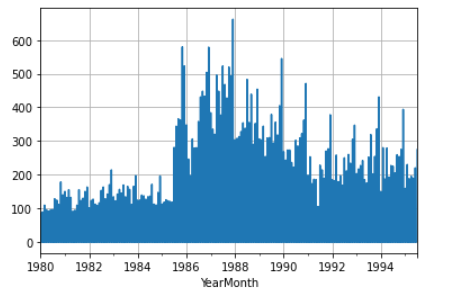
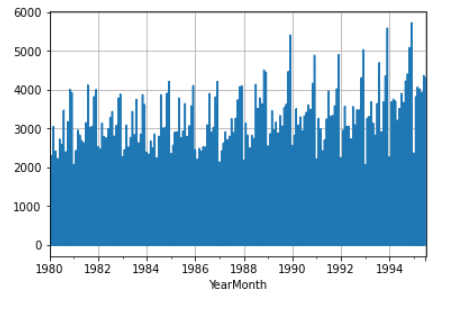
 

The first image we can see is the soft drink yearly production data and the second image we can see is the shoe sales yearly data. During the year 1994 the production of the soft drink is more as copared to other years and in 1987 the sale of shoe is more than in any other year.

Moving forward to see the quarterly performance of shoe sales and soft drink production.

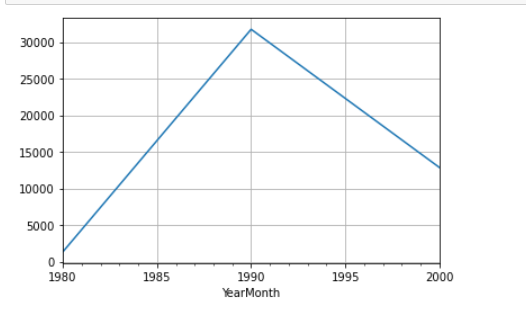
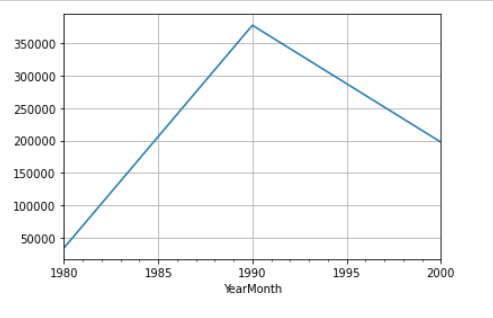
 

The first image is of shoe sales and the second image is of soft drink production, in the quarterly data we are seeing the variation of the sales more accurately. Now we will see the daily plot to know the sales.

On the above mentioned image we see the first daily plot is of shoe sales and the second is the daily plot of soft drink production.the insights are not clear as we cannot examine the sales of any year clearly, let’s move forward to see the decade plot to know the better insights.

Lets see the decade plots, the first plot will be the shoe sales decade plot and the second plot will be of soft drink production decate plot

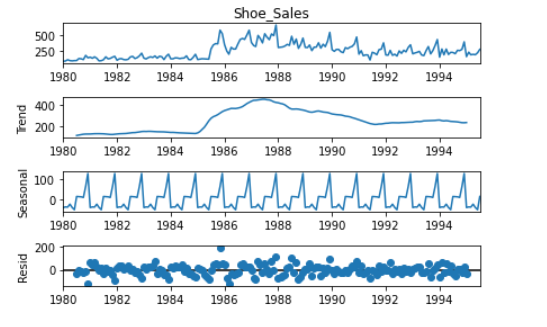
We have the clear insight on the above mentioned image for the sales.

Now we will learn to decompose the data the imortant of decomposition and the insights of data by decomposing

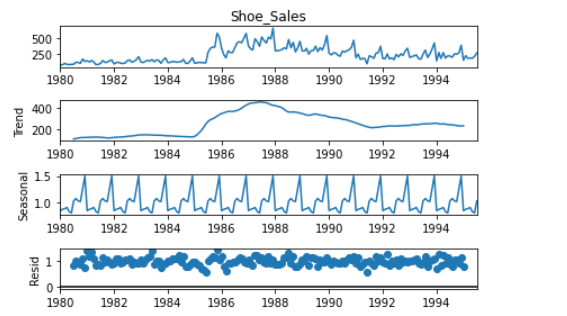
**Decomposition**- Time series decomposition helps us disentangle the time series into components that may be easier to understand and, yes, to forecast.

# **Decompose the Time Series**

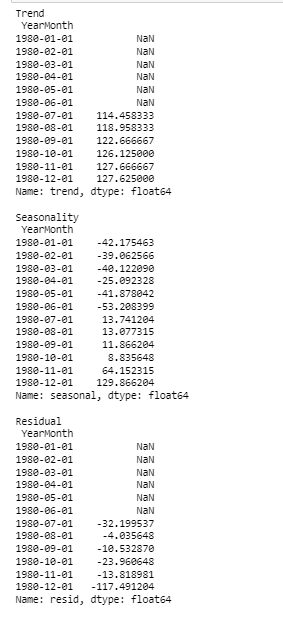
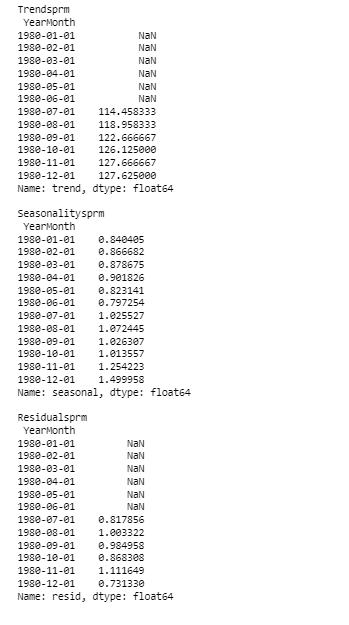
# Additive Model- As per the 'additive' decomposition, we see that there is a pronounced trend in the earlier years of the data. There is a seasonality as well.



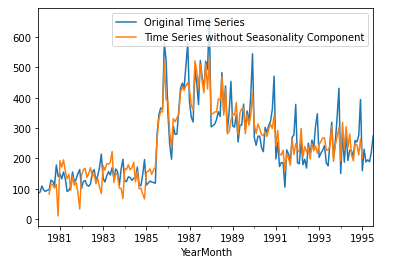
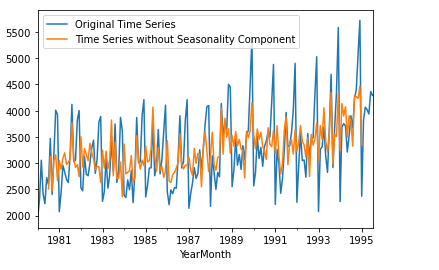
**Multiplicative Model**- The multiplicative decomposition argues that time series data is a function of the product of its components.



We can see the data spliting into trends sesional and resid form to know how much sale is done as per trend seasonal and residual.

We have the additive and multiplicative data the data is accoring to trend the sales and production, sales according to sesonal and the sales accoring to residual. Now lets observe the data in plot

Now we will look into the splitting data observation for both Shoe sales and soft drink production.

Splitting data is done to use values at the rear of the dataset for testing and everything else for training, we can split the data in different manner

Model on Training data and RMSE on test data

Model 1:Linear Regression

Model 2: Naive Approach

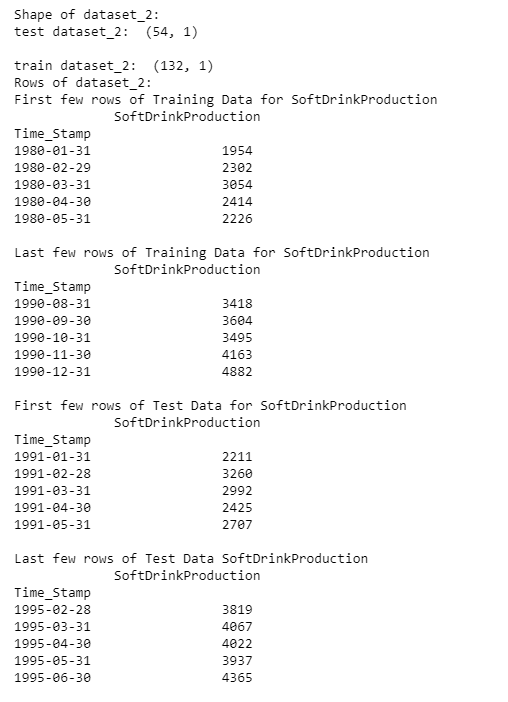
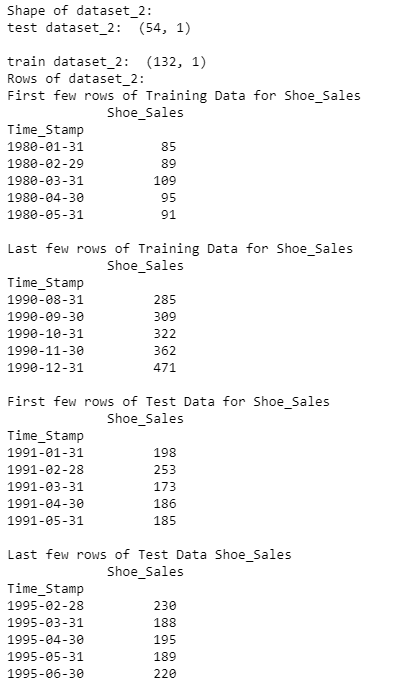
Model 3: Moving Average(MA)

Model 4: Simple Average

Model 5: Simple Exponential Smoothing

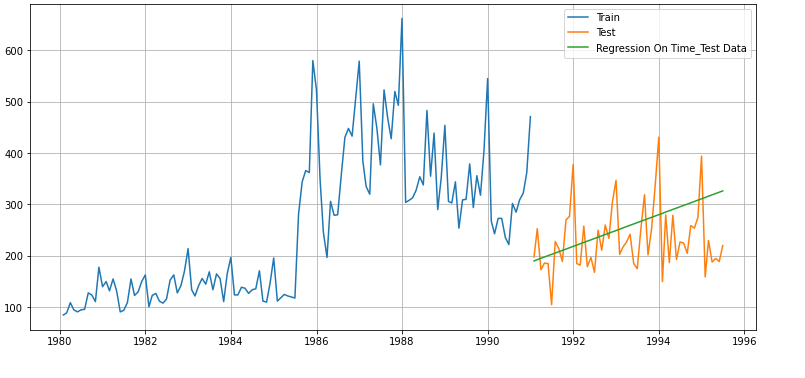
Model 6: Double Exponential Smoothing (Holt's Model)

Model 7: Triple Exponential Smoothing (Holt - Winter's Model)

**Linear Regression**- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

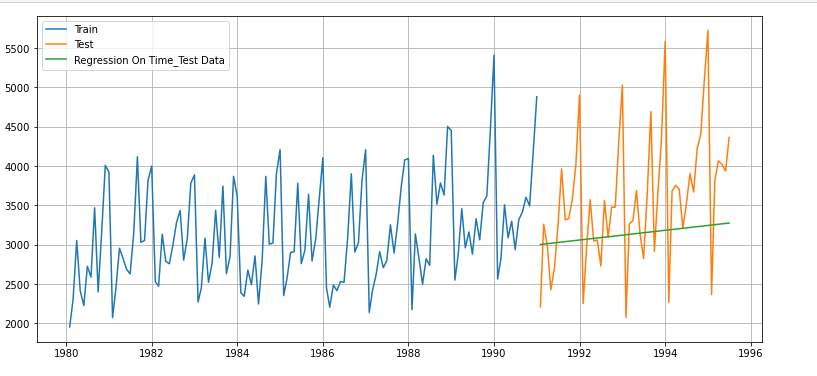
**Data after performing linear regression for shoe sales.**



**Result of regression data of shoe sales**

For RegressionOnTime forecast on the Test Data, RMSE is 73.404

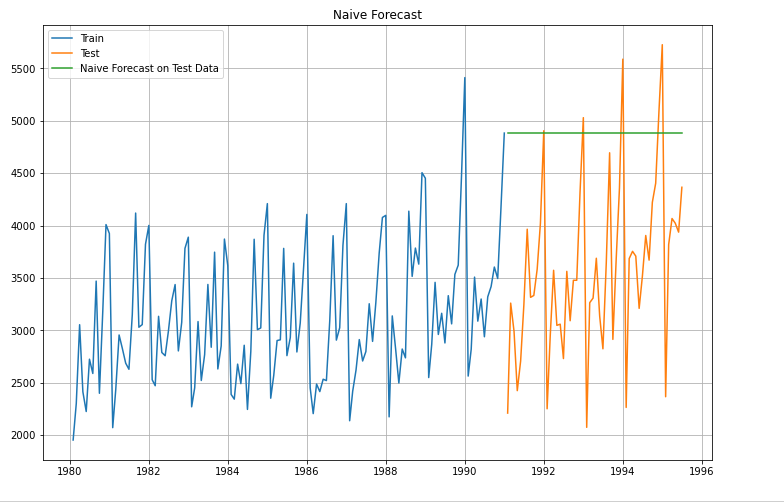
**Data after performing linear regression for Soft drink production.**

****

**Result of regression data of soft drink production**

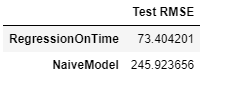
For RegressionOnTime forecast on the Test Data,RMSE is 895.992

# Naive Approach-The Naïve Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification. It is also part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category.



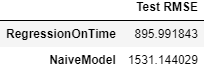
Result of Navies Bayes for Shoe sales

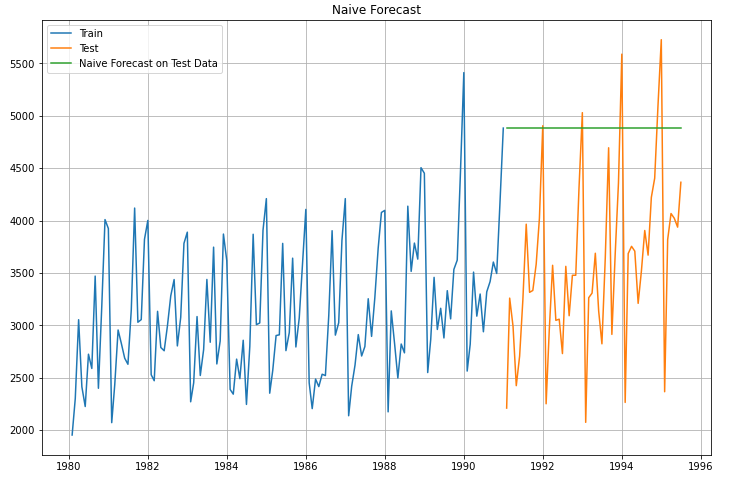
For RegressionOnTime forecast on the Test Data, RMSE is 245.924



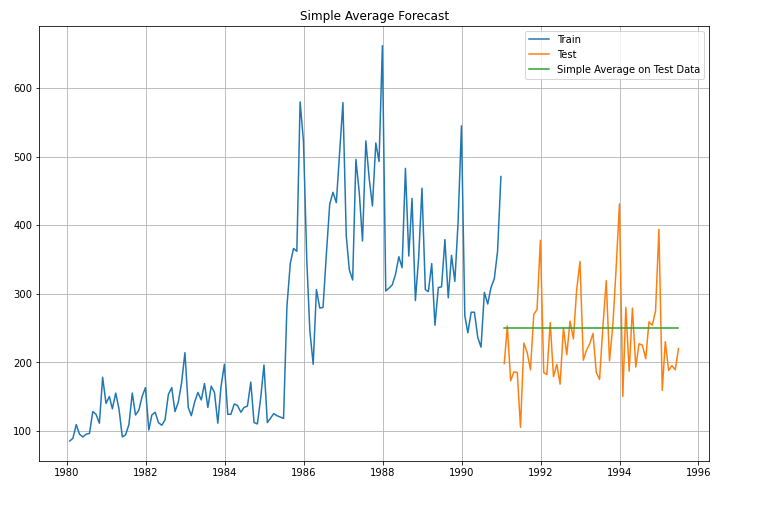
Result of Navies Bayes for Shoe sales

For RegressionOnTime forecast on the Test Data, RMSE is 1531.144



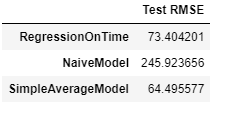


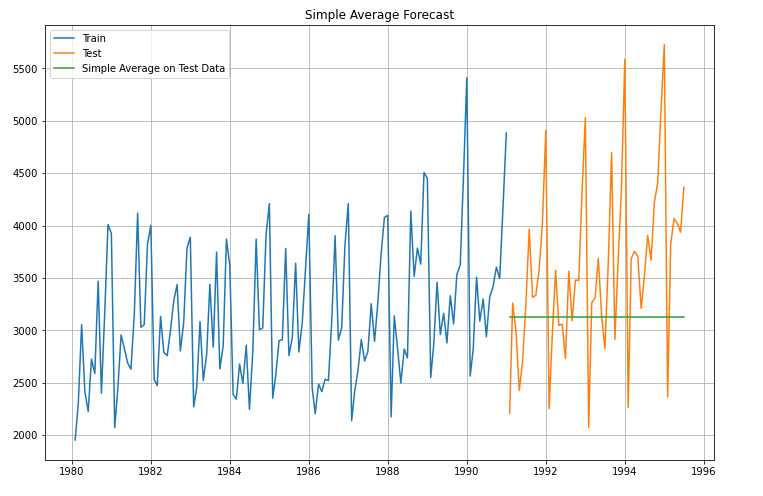
# **Simple Average**- It is a method for inventory valuation or delivery cost calculation, where even if accepting inventory goods with different unit cost, the average unit cost is calculated by multiplying the total of these unit costs simply by the number of receiving.



Result of simple average forecasting for shoe sales

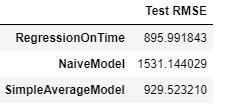
For Simple Average forecast on the Test Data, RMSE is 64.496





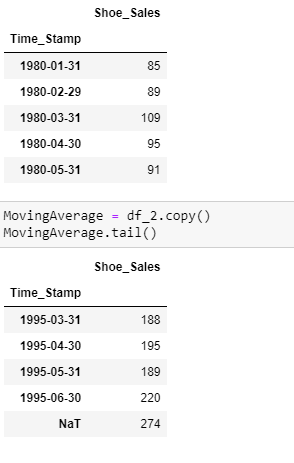
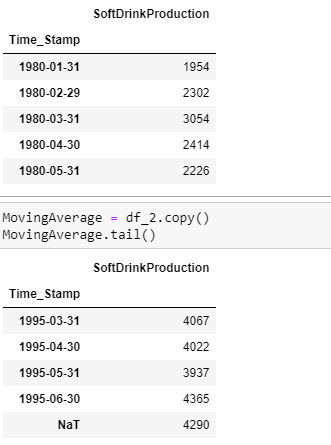
Result of simple average forecasting for soft drink production

For Simple Average forecast on the Test Data, RMSE is 929.523

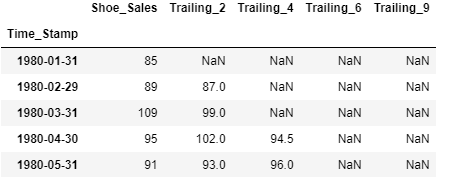
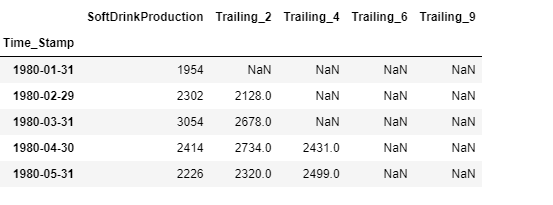


# **Moving Average**

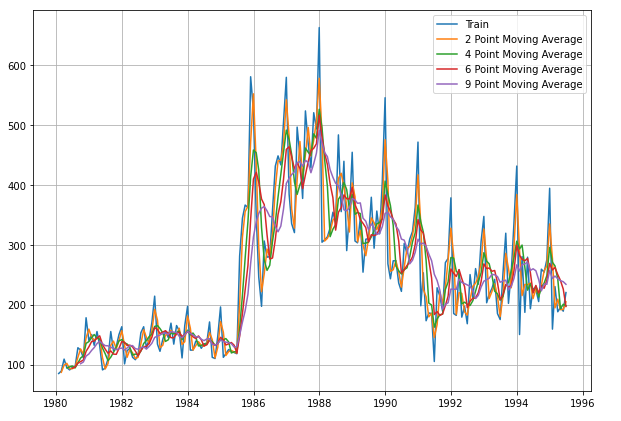
# For the moving average model, we are going to calculate rolling means (or moving averages) for different intervals. The best interval can be determined by the maximum accuracy (or the minimum error) over here. For Moving Average, we are going to average over the entire.

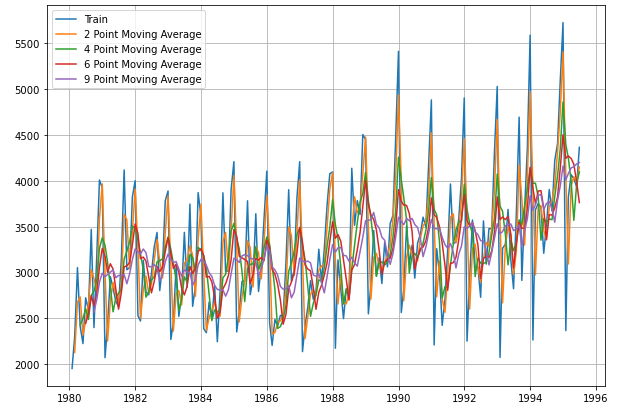
# Trailing moving averages

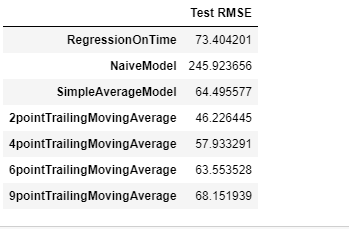
**SHOE SALES TRAINING DATA**



**SOFT DRINK PRODUCTION DATA**

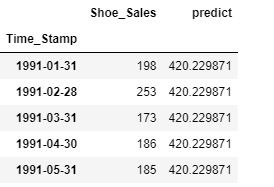


**RESULT OF TRAILING MOVING AVERAGE DATA OF SHOE SALES**

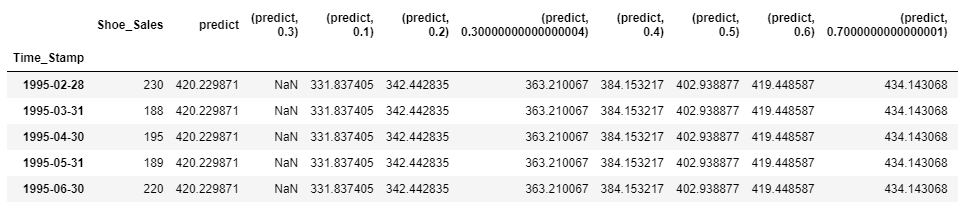


**Simple Exponential Smoothing :** Simple or single exponential smoothing (SES) is the method of time series forecasting used with univariate data with no trend and no seasonal pattern.

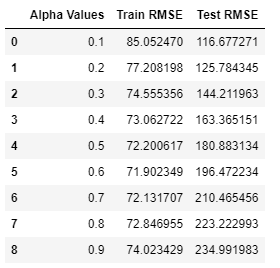
**RESULT OF SES TEST DATA**

****

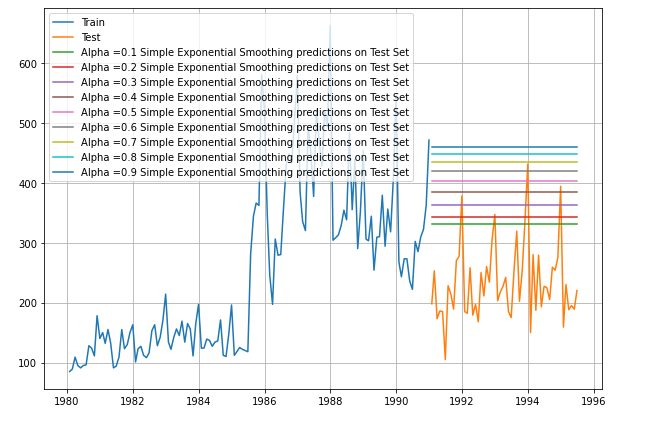
**RESULT OF SES RMSE TRAIN DATA**

****

**RESULT OF SES RMSE TRAIN AND TEST DATA**

****

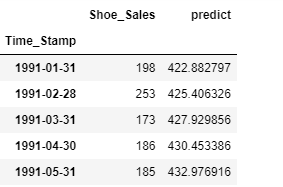
**PLOT OF THE ABOVE-MENTIONED TRAIN AND TEST DATA**

****

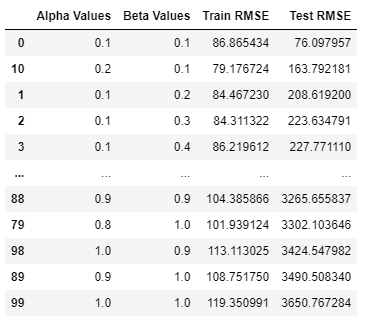
**Model 6: Double Exponential Smoothing (Holt's Model)**

Double exponential smoothing employs a level component and a trend component at each period. Double exponential smoothing uses two weights, (also called smoothing parameters), to update the components at each period.

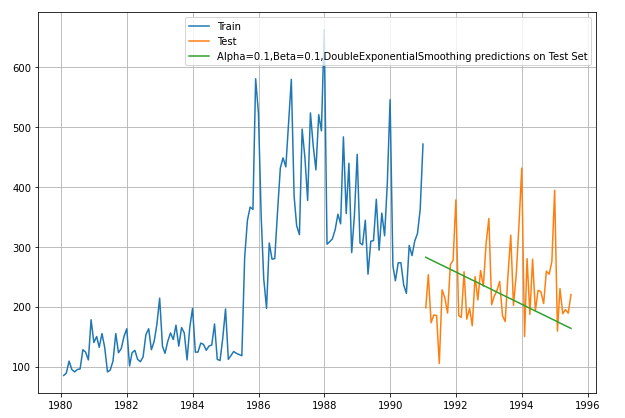
**Result of DES Data for shoe sales**

****

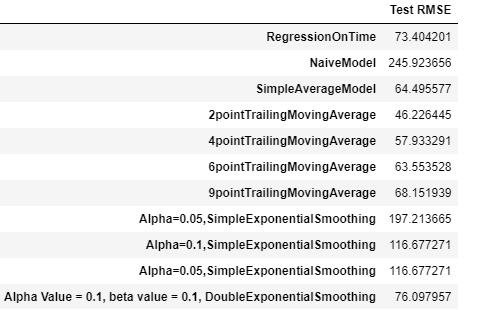
**Result of Train RMSE and Test RMSE data for shoe sales**

****

**Result of above mentioned data into plot.**

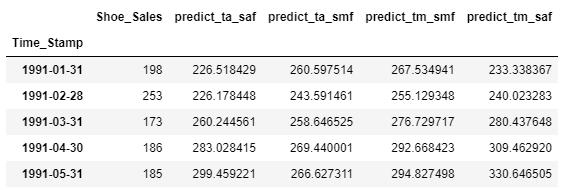
****

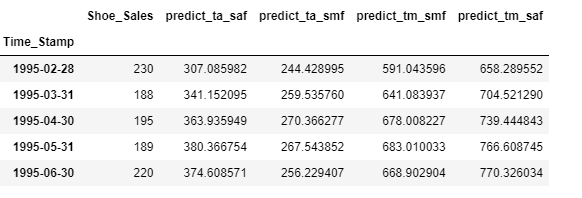
**Overall result of DES Data**

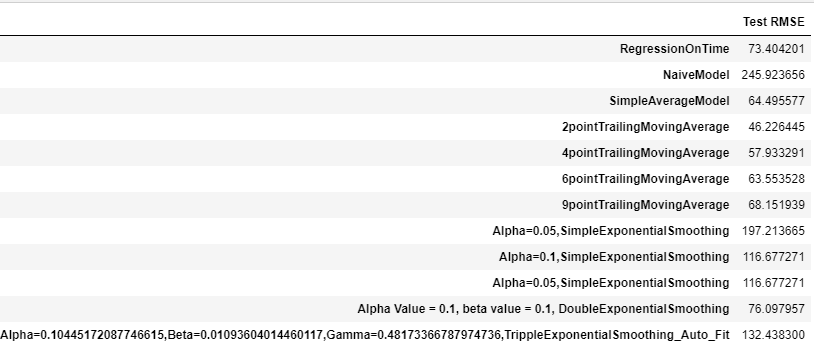
****

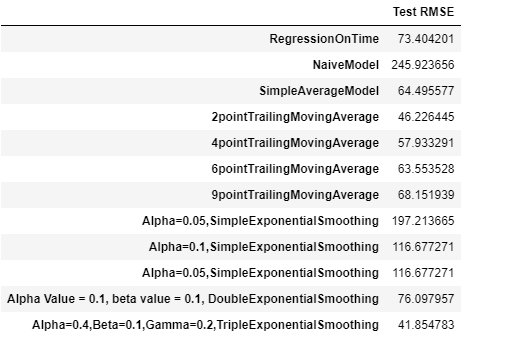
**Triple Exponential Smoothing-** Triple exponential smoothing is used to handle the time series data containing a seasonal component. This method is based on three smoothing equations: stationary component, trend, and seasonal. Both seasonal and trend can be additive or multiplicative.

**Result of TES Data**





**Overall result of TES Data**



Conclusion

Time series analysis is a must for every company to understand seasonality, cyclicality, trend and randomness in the sales and other attributes. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

Importance of Time Series  
  
It is helpful to compare the present trend with the past trend that has already happened so the future trend can be estimated and prepared. The cycle variations over a period using time series will allow us to understand the business cycle quite effectively.

Understanding after making time series

It occurs when you make scientific predictions based on historical time stamped data. It involves building models through historical analysis and using them to make observations and drive future strategic decision-making.

**THANK YOU!**