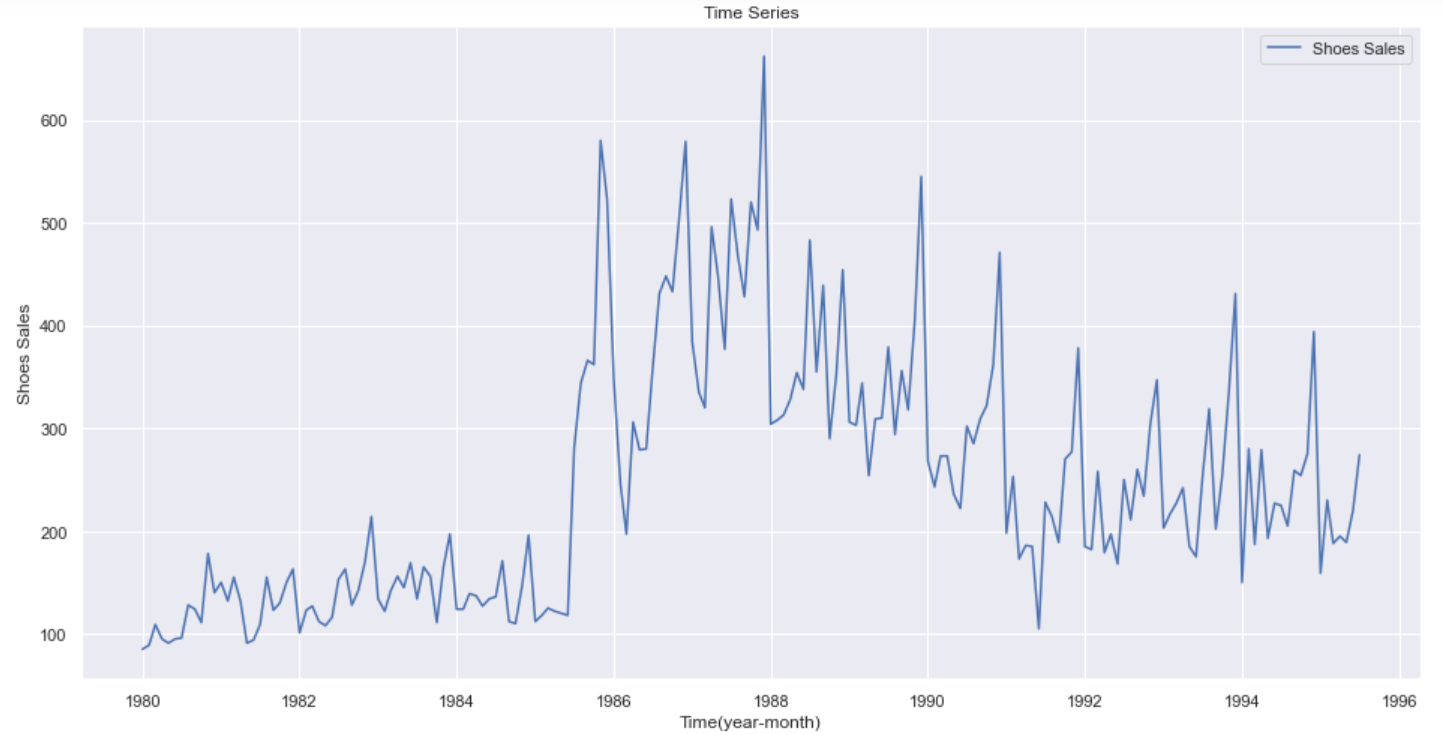
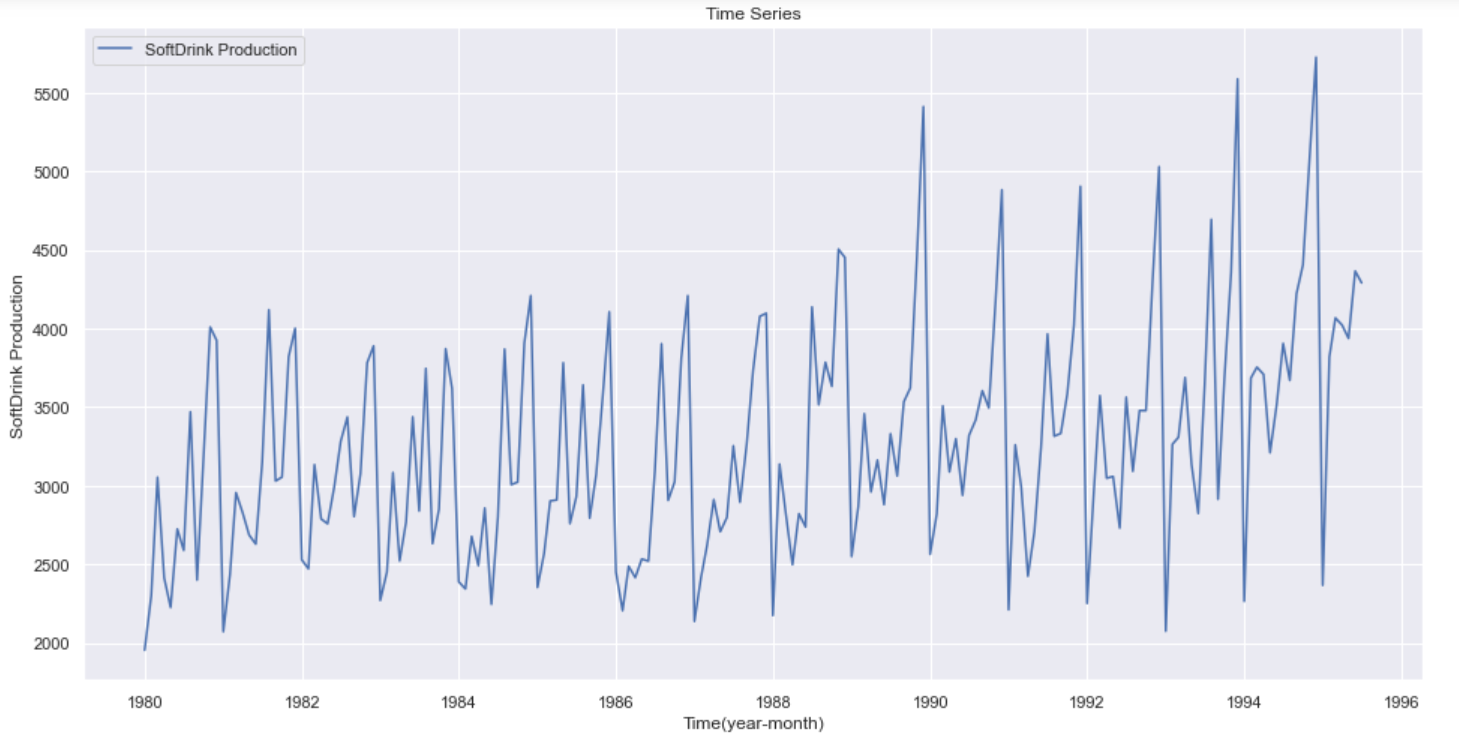
**Report**

There are 2 problems on which we have to work:

* Shoe Sales
* Soft drink Production

Let’s Start with plotting the dataset one by one



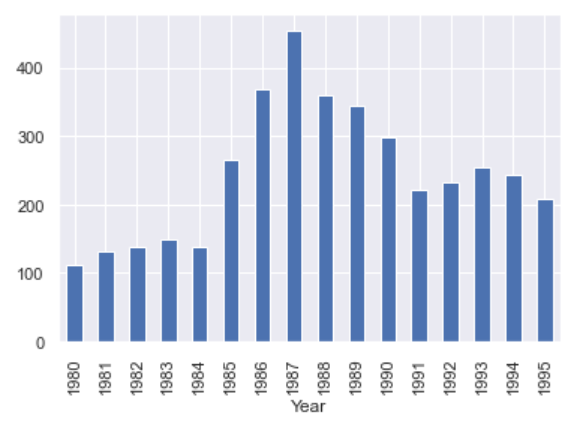


As we can see both are the problem of time series because we have data with time wise. Here we can see data is at monthly level. We have given with the monthly Sales of shoes and Soft Drink production and we have to predict what will be the sale or production in the next 12 months.

Now, Let’s perform some exploratory data analysis on both of the dataset.

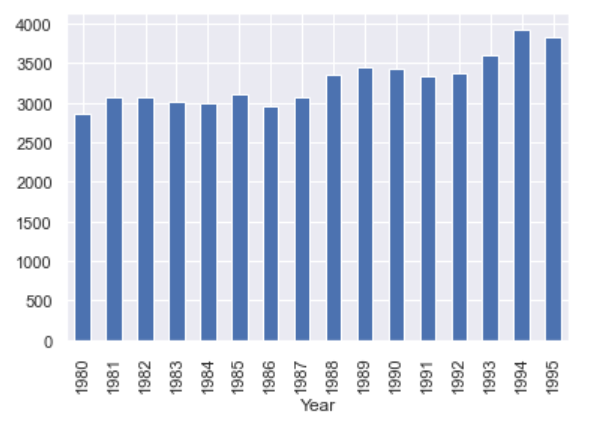
As we can see that **Shoe Sales between 1986 to 1988 got increased drastically** and then decreased again.

**Let’s check year wise sales and production**



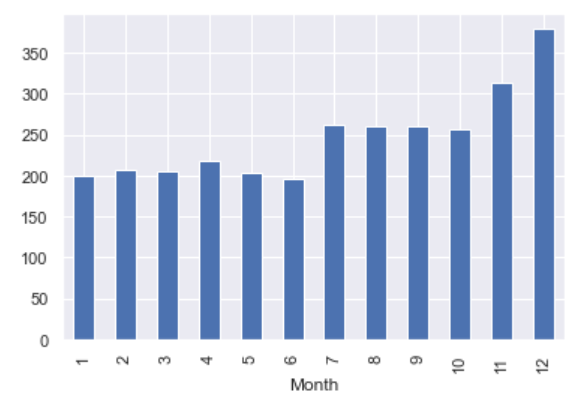
We can also verify my above statement by looking at this graph. The sales of shoes got increased after 1985 and decreased again after some years.

**Lets have a look on the Soft drink Production Data.**



We can see this that, Soft drink production is overall increasing.

**Let’s check on which month Shoe Sales are maximum**



We can see that the shoes sales in December is maximum and minimum in the June month because generally people wear shoes more in winters as compare to the summer.

**Let’s check Monthly Wise Soft drink production**



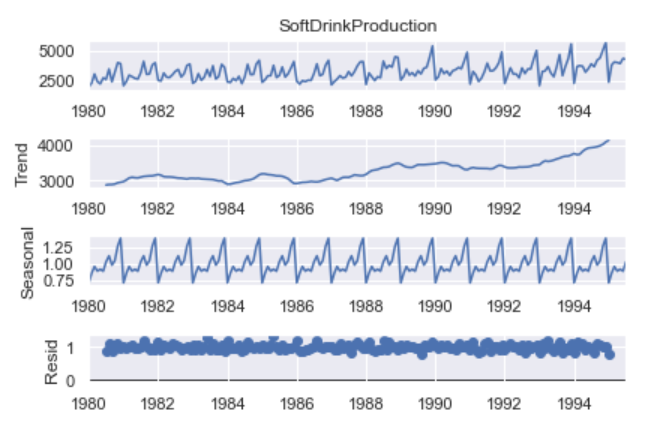
Production of soft drinks are maximum in December month and minimum in the jan month.

**Let’s Decompose the Shoe sales to see the seasonality, trend and noise in the dataset**



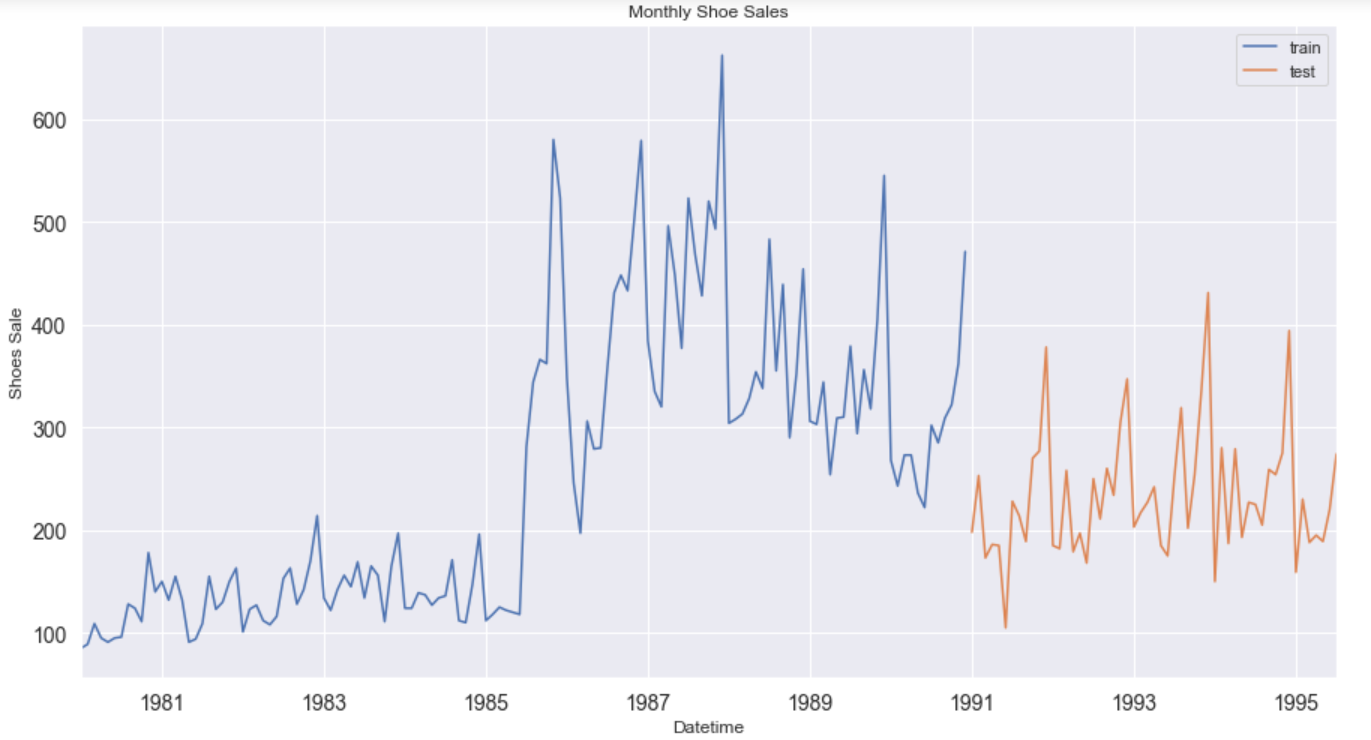
In this we can clearly see that shoes have some seasonality, trend and lots of noise as well.

**Let’s Check the Soft drink Data set Decomposition**

****

We can clearly see that the production is increasing and it also follows some pattern.

Before Applying any model, first we need to split the dataset into training set and test set to validate our model.

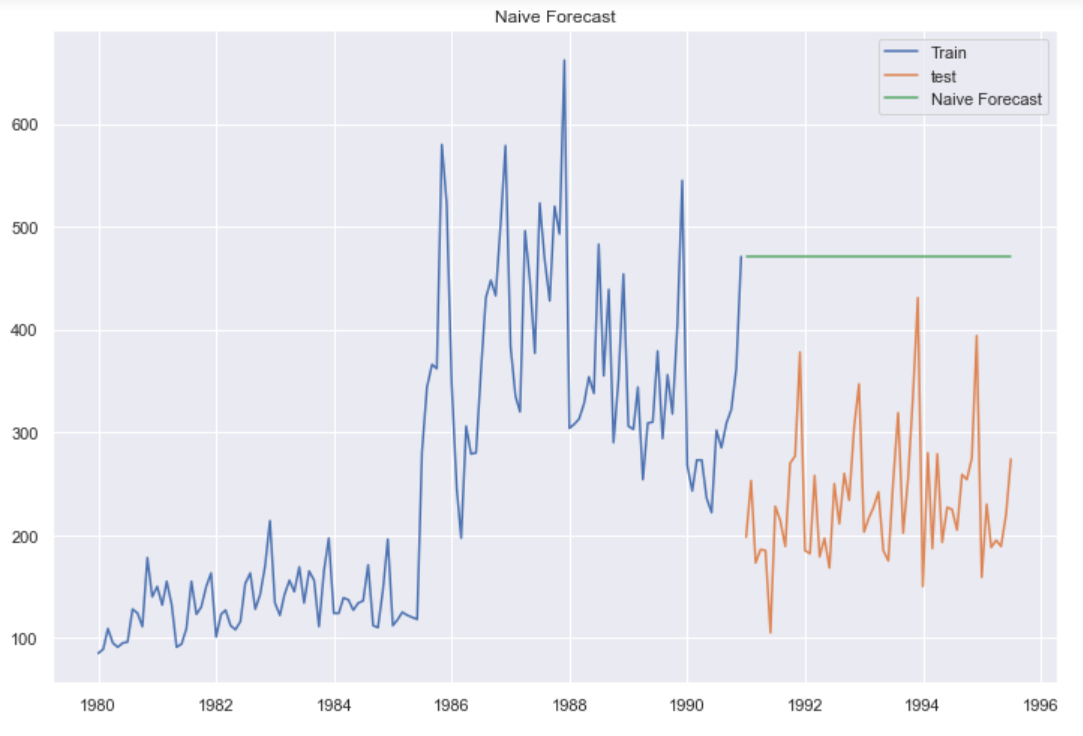


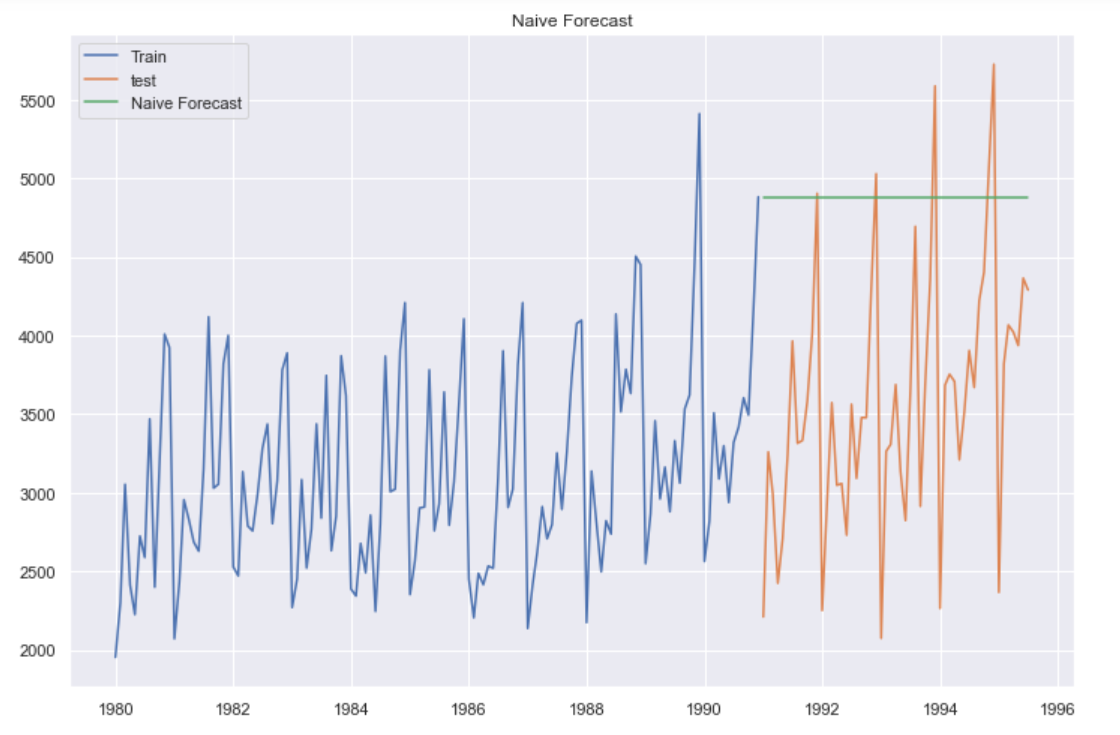


I have taken data from 1991 year as the test data and before 1991th year data is training data.

There are So many models which are available in the market lets try them one by one

1. **Naïve Forecasting Model**

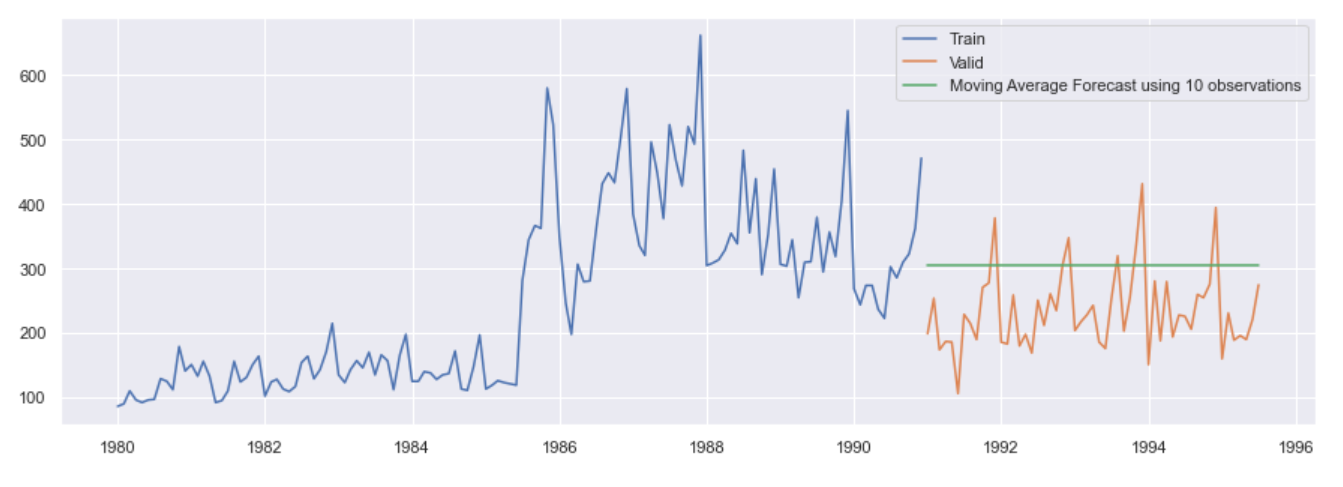
****

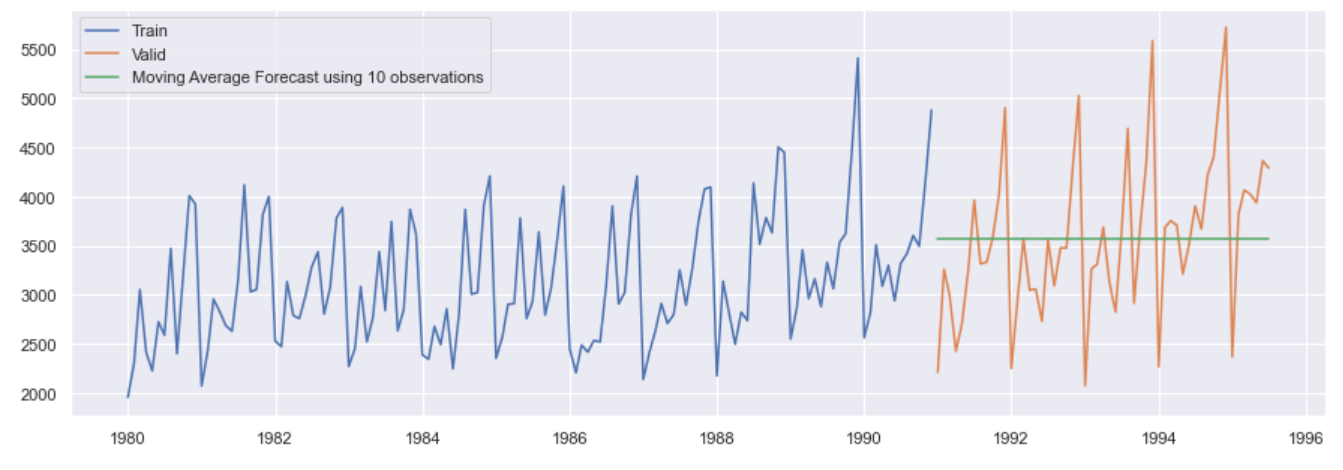


We can see the straight line after 1991 because moving average model says that put the last sale value to all the future outcomes.

1. **Moving Average Model**

Moving average model says that instead of assigning the last value, Assign the average of last n values, I took 10 as the value of n.

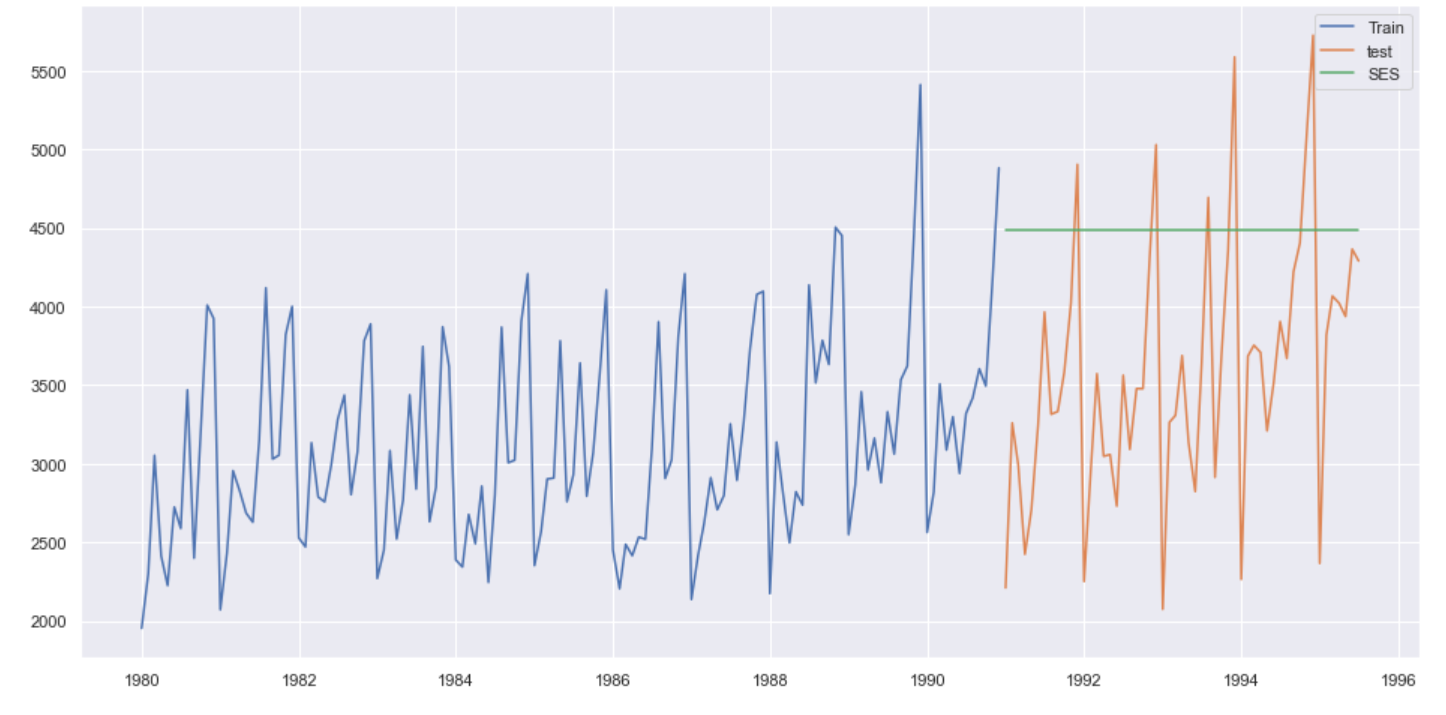




1. **Simple Exponential Smoothing**

Simple Exponential smoothing method give more weightage to the latest sales as compare to old one. So forecasting result will more rely on the latest monthly sales more.

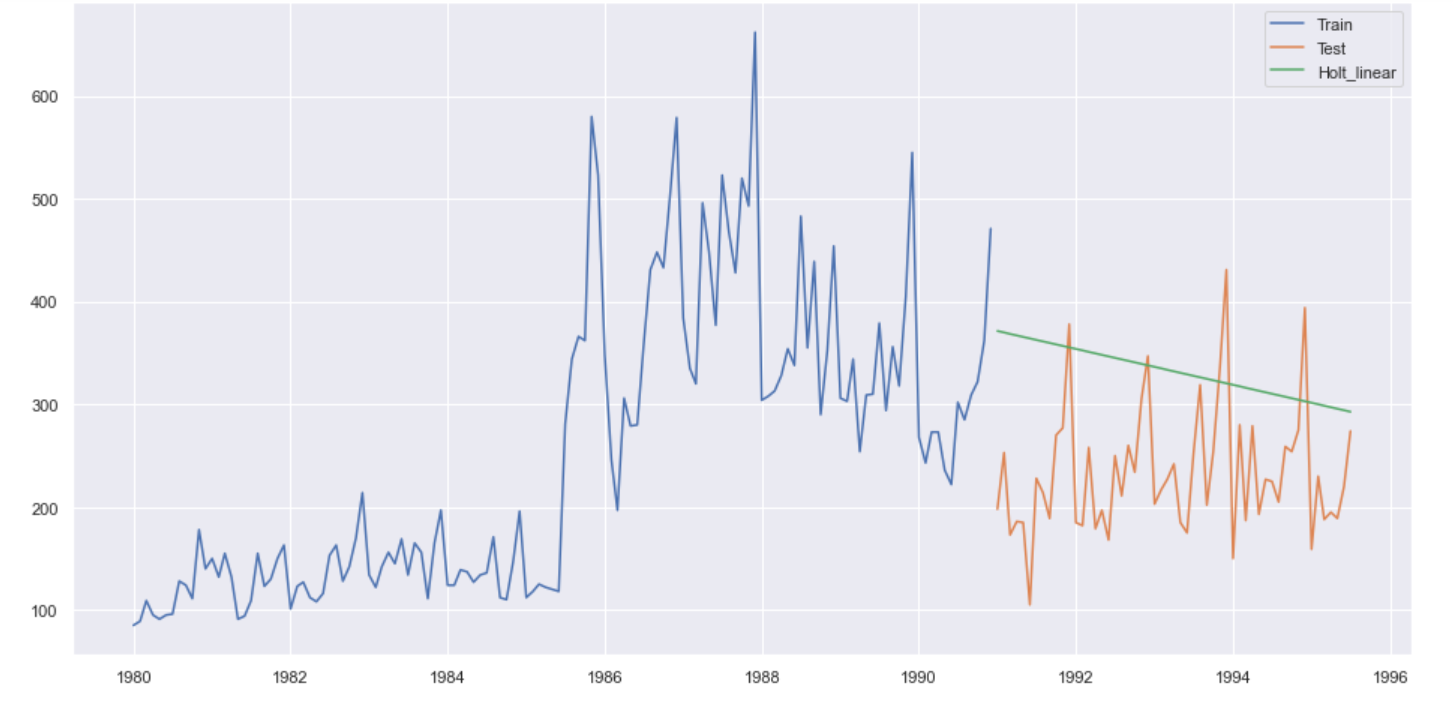


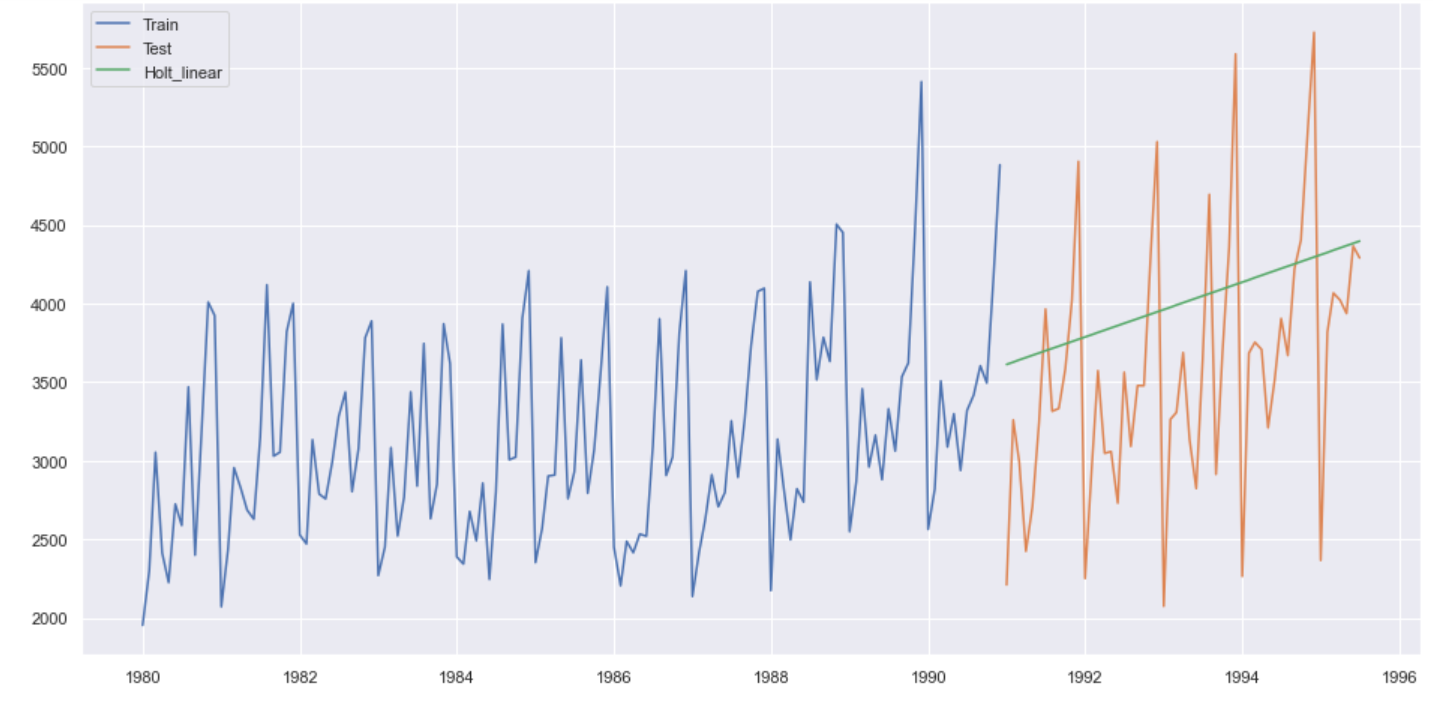


As we can see in the shoe sales the forecasting value of exponential model is pretty high because the latest monthly shoe sale was high and after that sale got decreased but this method doesn’t consider trend that’s why an another model came into the picture.

1. **Holt’s Linear Trend Model**

This method is like the exponential smoothing but with the trend functionality.

****

****

Now we can see that, Holt winter model also encounter the trend.

**Check for the stationarity**

First of all lets understand why we need to check this factor and what is the meaning of stationarity in the time series.

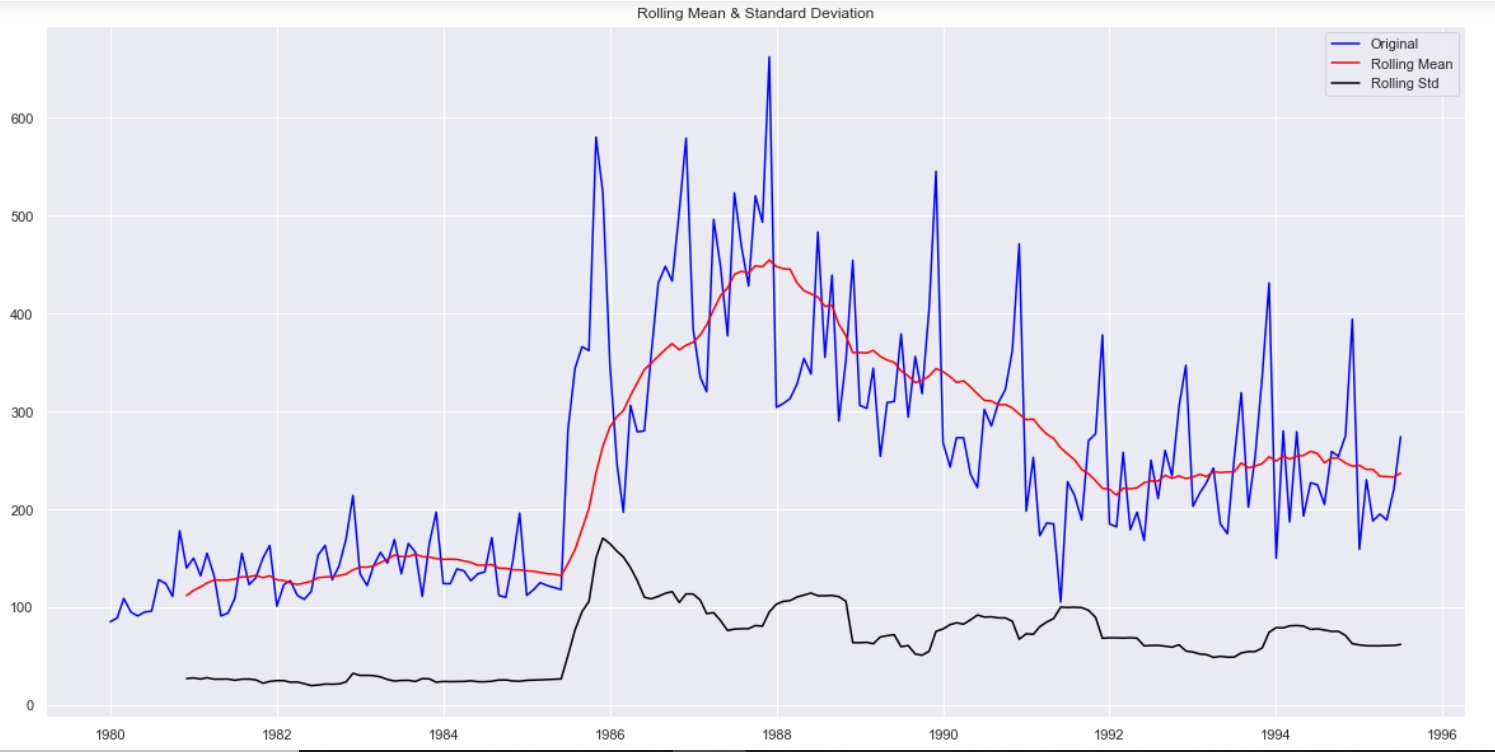
**There are three basic criterion for a series to be classified as stationary series :**

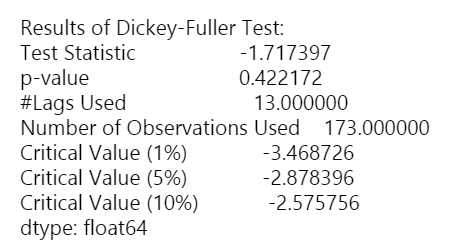
* The mean of the time series should not be a function of time. It should be constant.
* The variance of the time series should not be a function of time.
* The covariance of the ith term and the (i+m)th term should not be a function of time.

**Why do we have to make the time series stationary?**

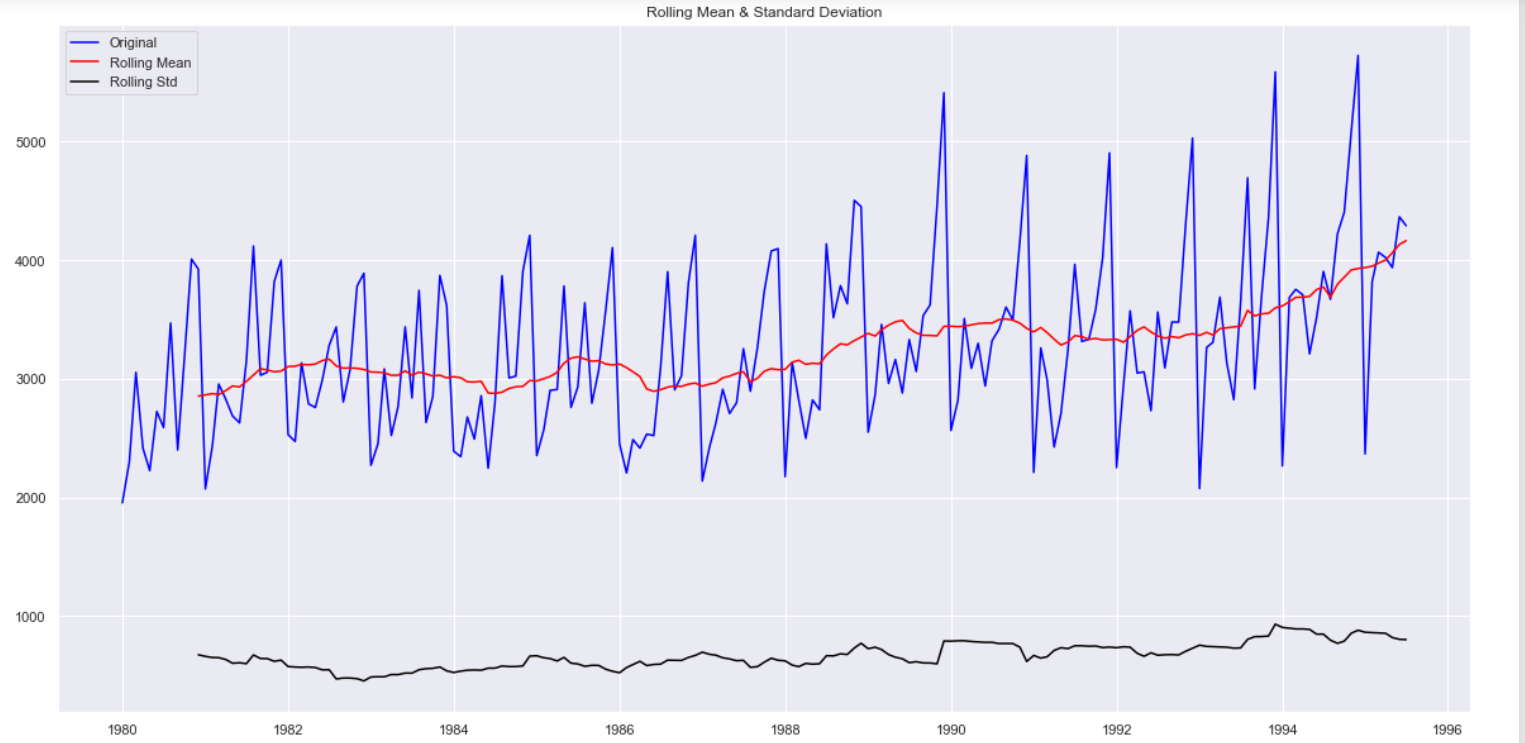
We make the series stationary to make the variables independent. Variables can be dependent in various ways, but can only be independent in one way. So, we will get more information when they are independent. Hence the time series must be stationary.

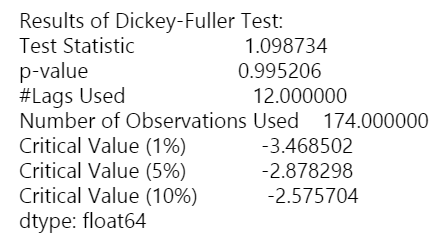
**Checking for stationarity**

****

****

We can clearly see the p-value. It is very high So, we can say that this time series is not stationary.



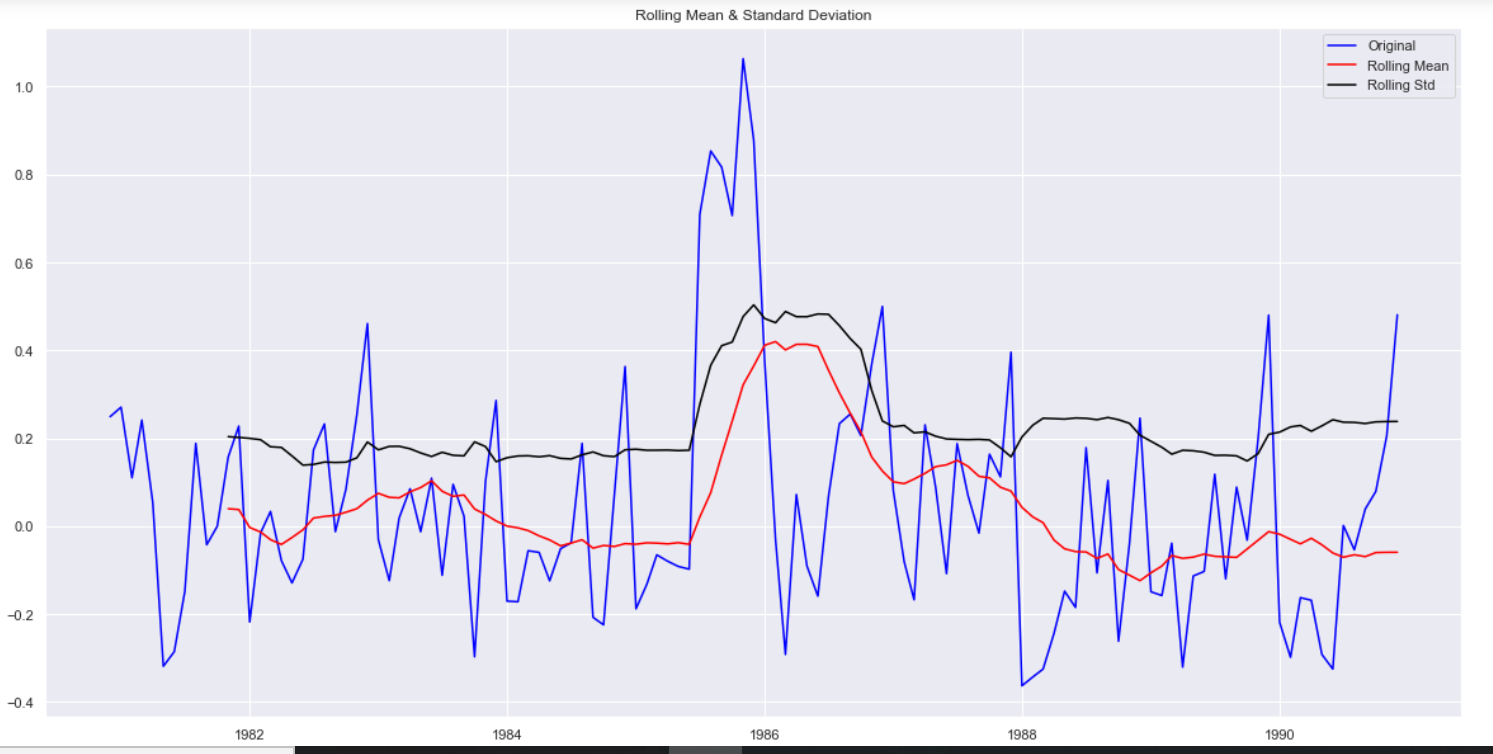


Having high p-value hence, it is non-stationary.

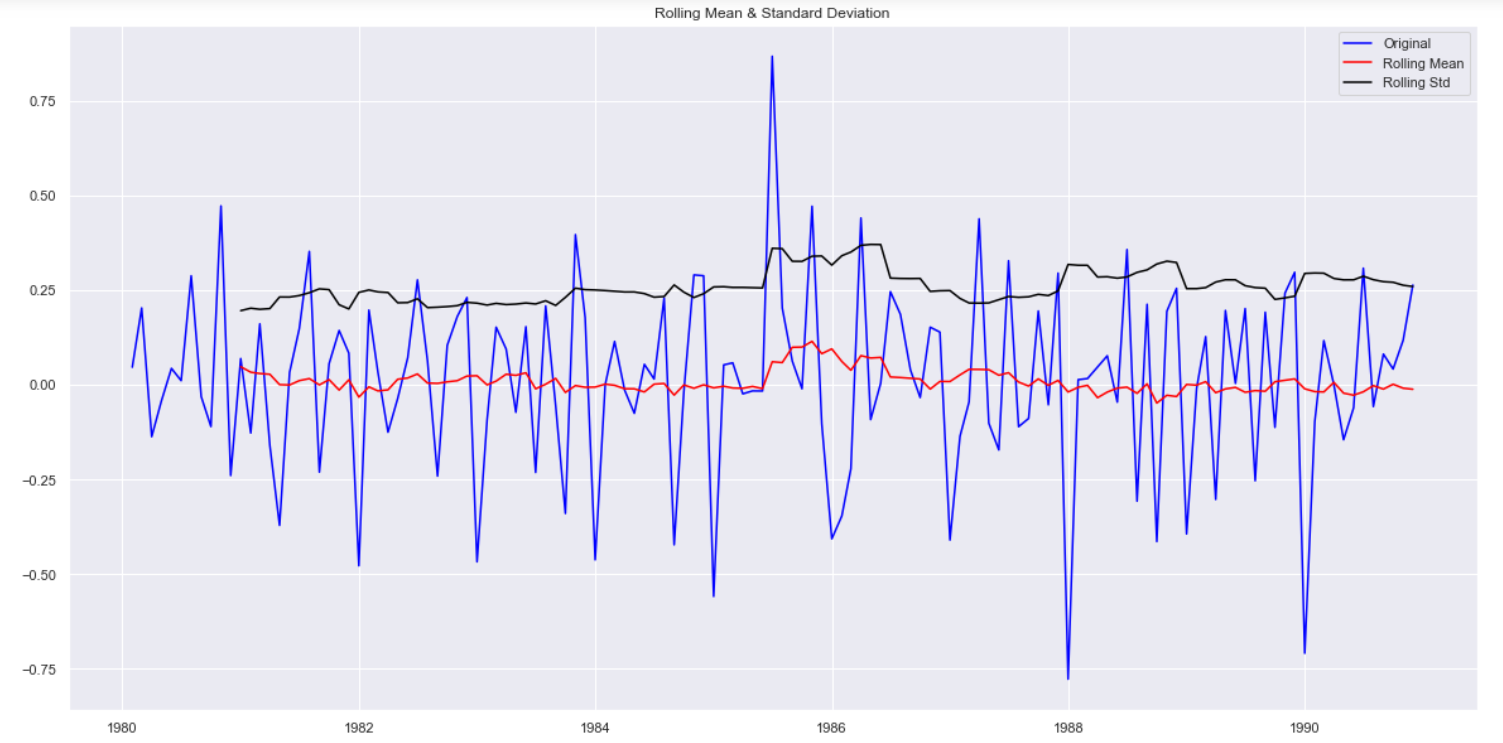
Both of the time series are non-stationary So, Let’s do the treatment of both of them.

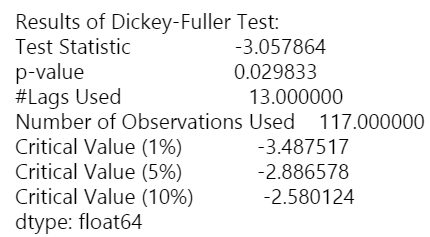
**Shoe Sales Time Series:**

* Lets remove the trend first by taking the difference moving average

****

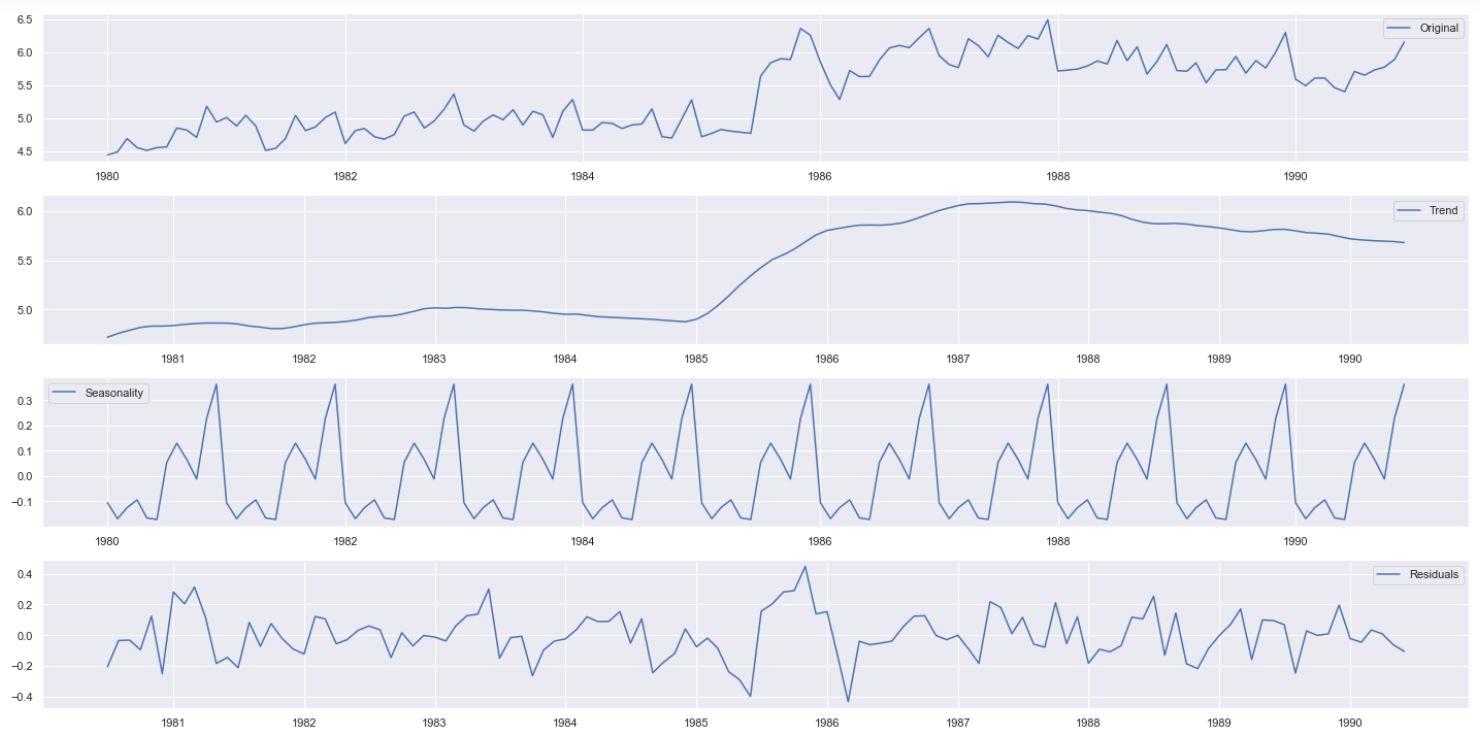
Take the first order difference and check the stationarity again

****



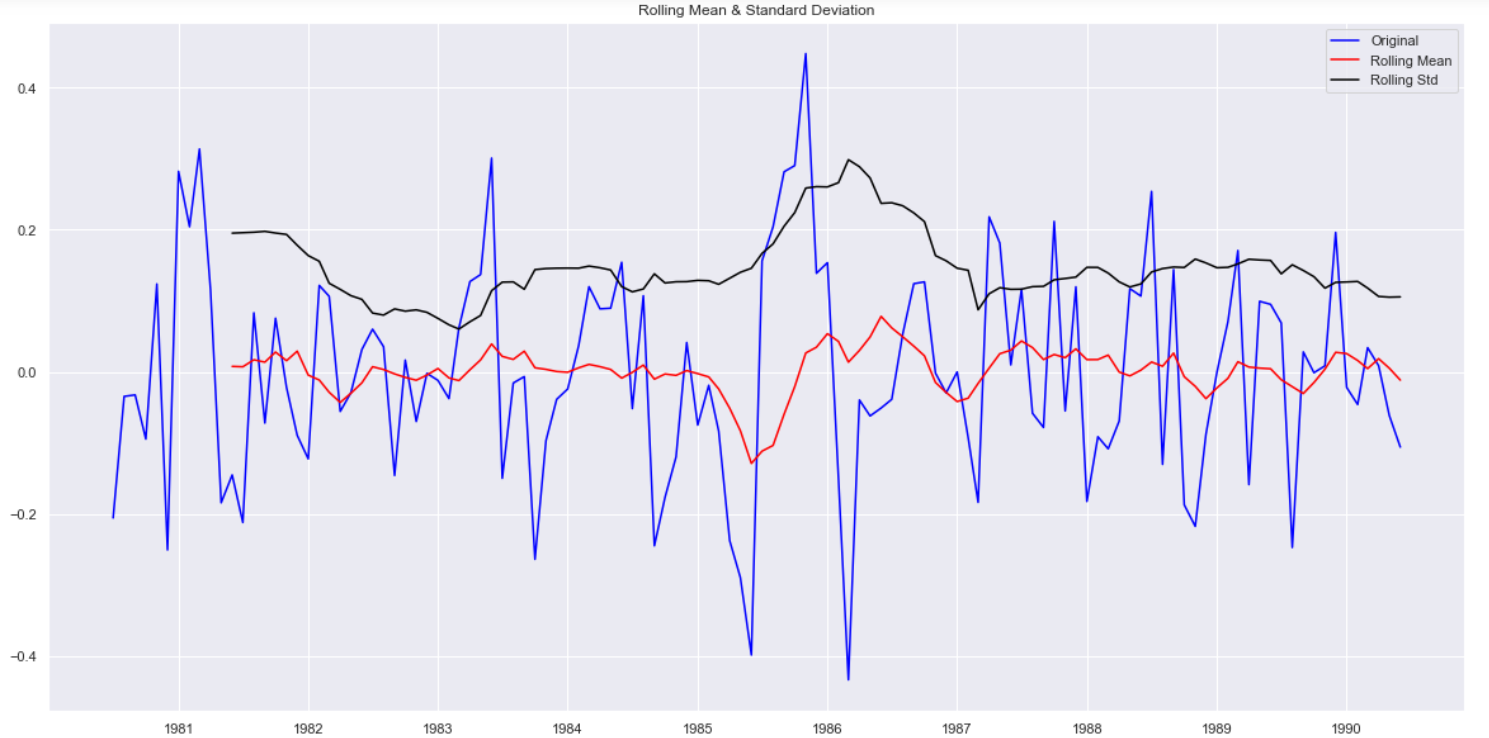
P-value has decreased, Hence It is becoming stationary.

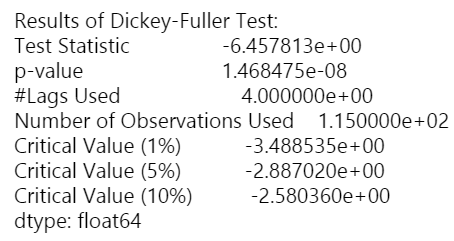
Lets check the seasonality pattern and residuals and decompose the data.



We can see the trend, residuals and the seasonality clearly in the above graph. Seasonality shows a constant trend in counter.

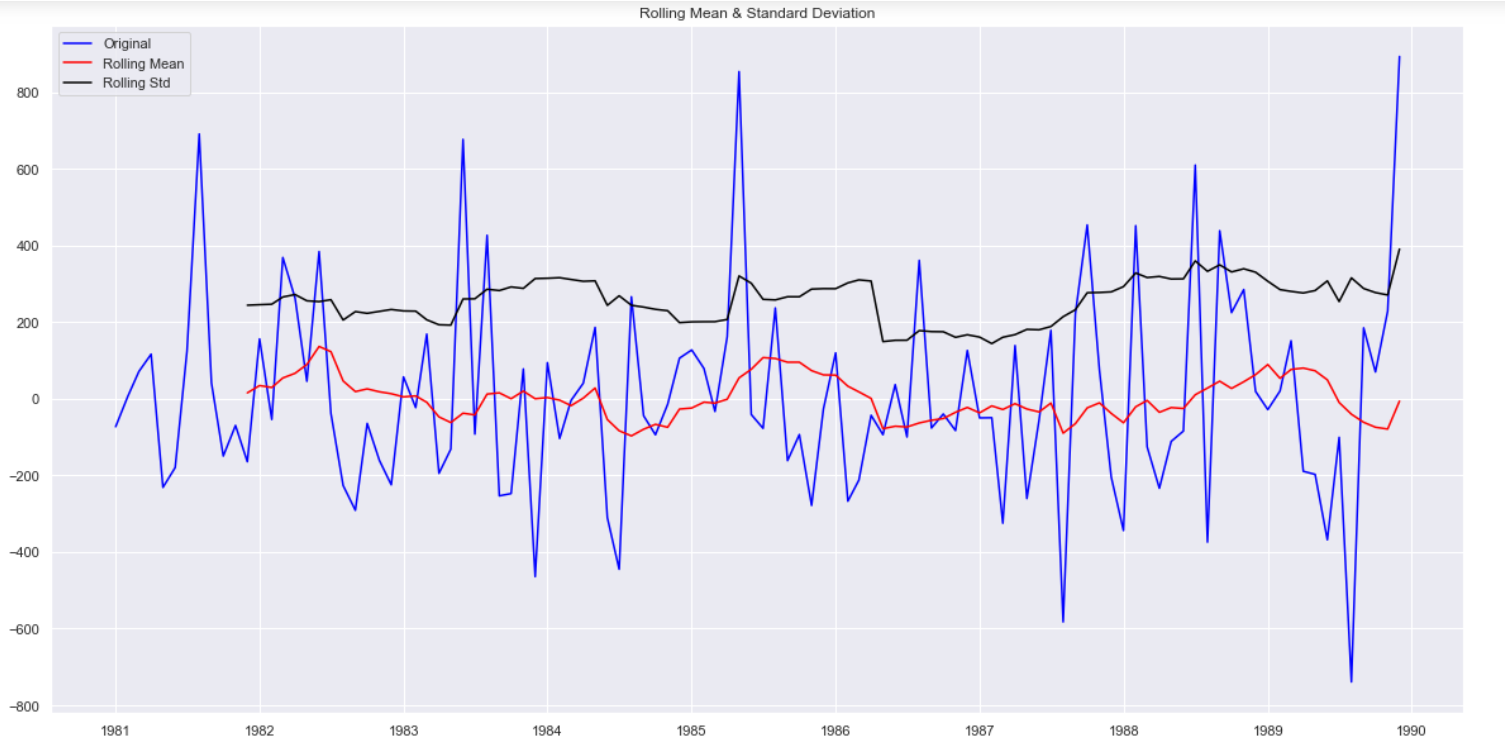
Lets see the residuals

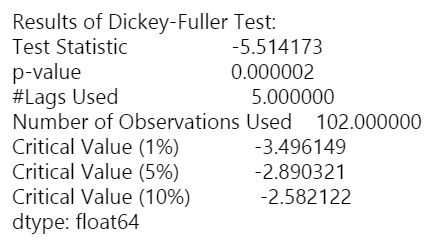




Now, this time series has become stationary and we can clearly verify that thing by looking at the p- value.

I have performed the same operation for the soft drink production dataset and I got the same result.





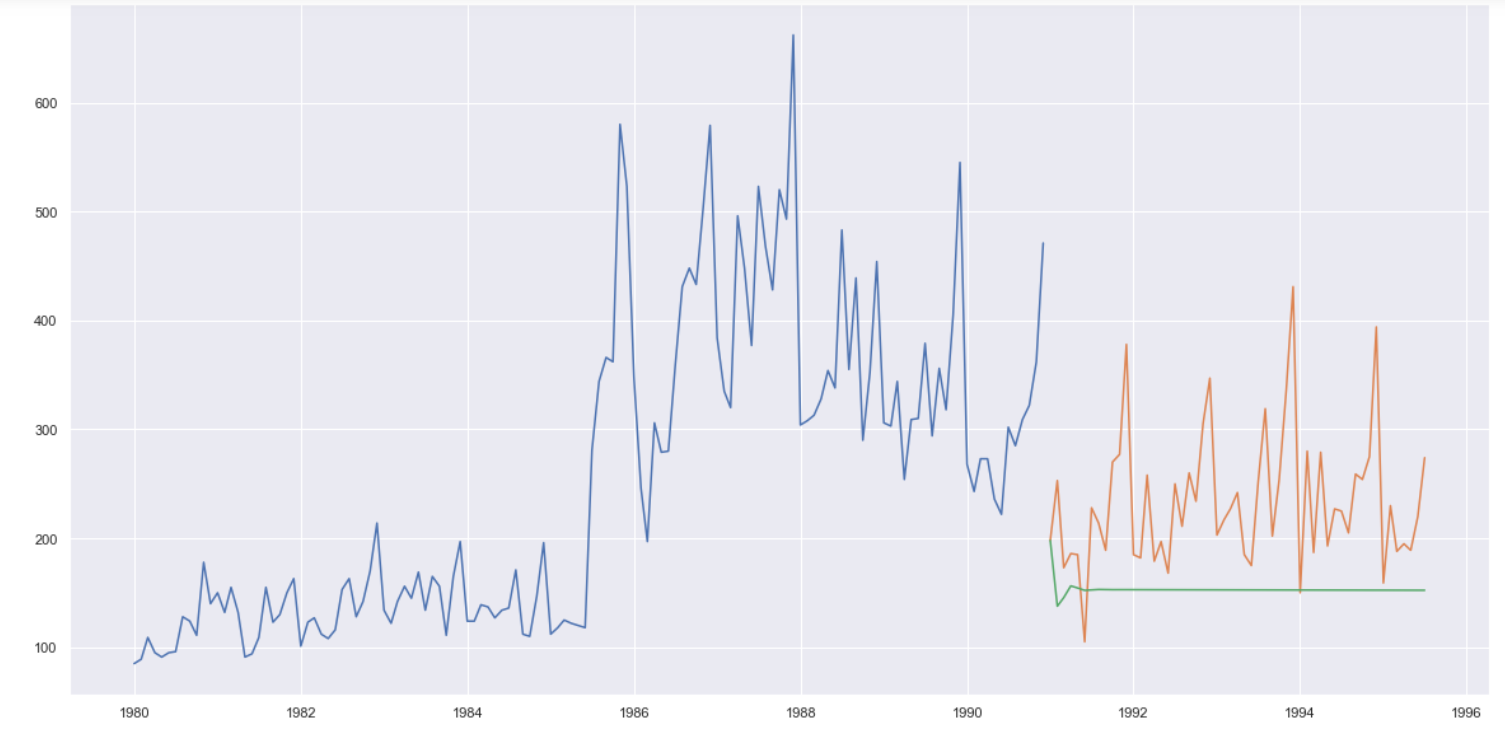
The p-value of the Soft Drink production time series is also very less So, we can reject the null hypothesis which assumes that our time series is non-stationary.

So, we can say that we have statistically proved that our both of the time series has become stationary.

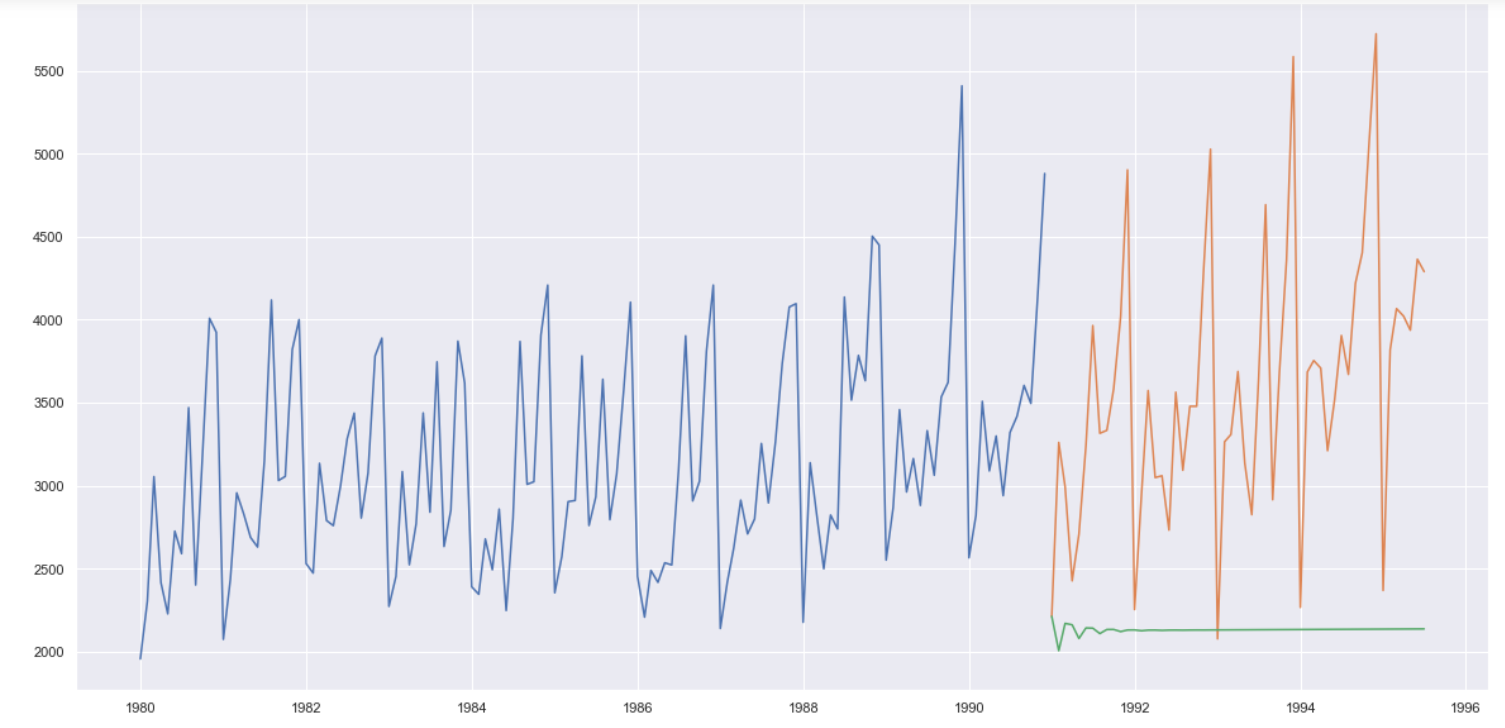
**Applied Auto Arima:**

In Auto arima, we don’t need to take care about the stationarity and parameters value because, it handles it by its own, It finds out the AIC value and on that basis it evaluates the model.

On Shoe data



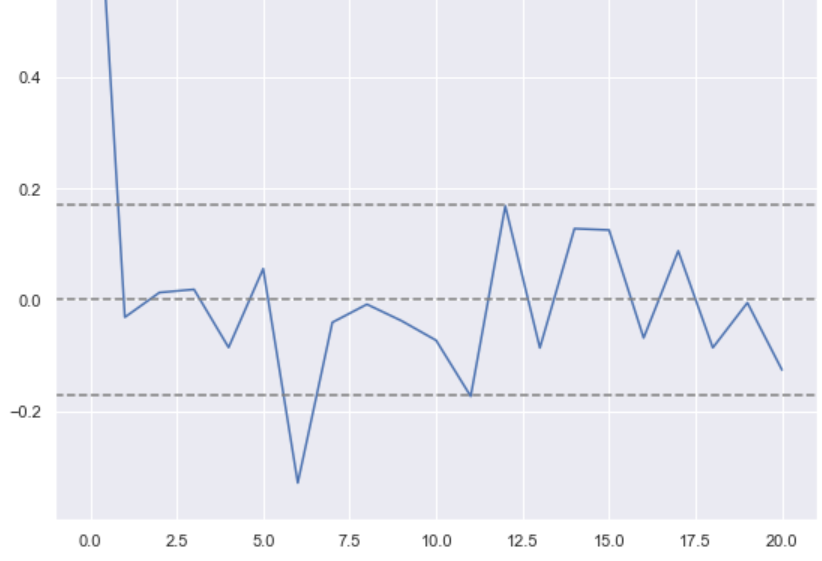
On softdrink data



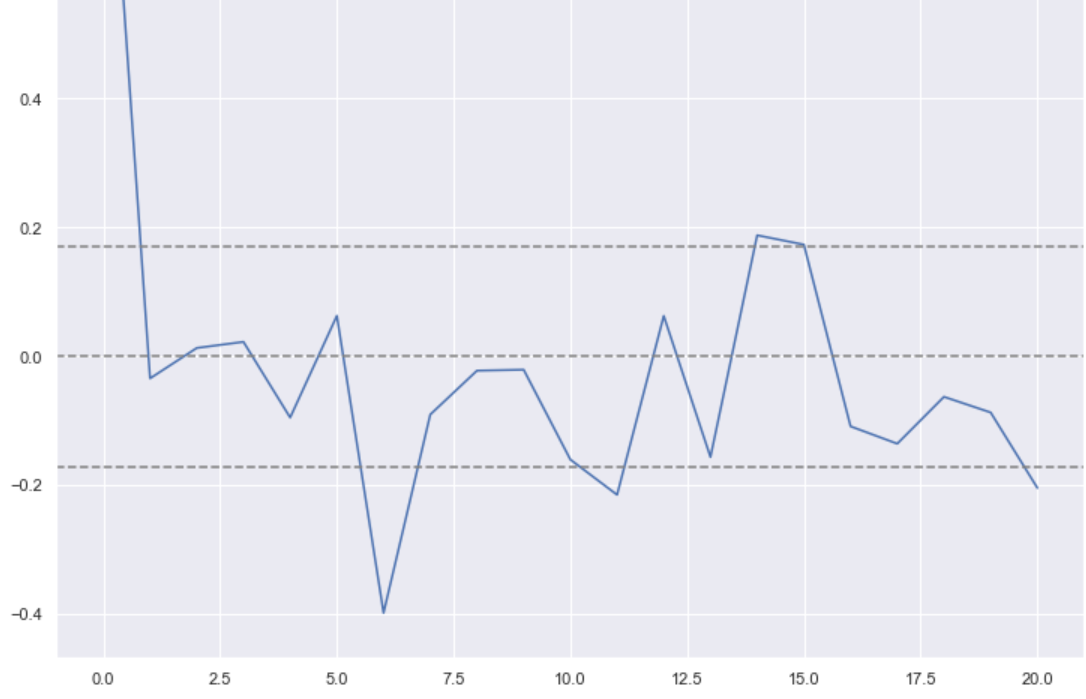
Now, lets try to find out PACF, ACF plots for finding the value of p, d, q for the Arima model.

**Soft Drink Data**

* **ACF**

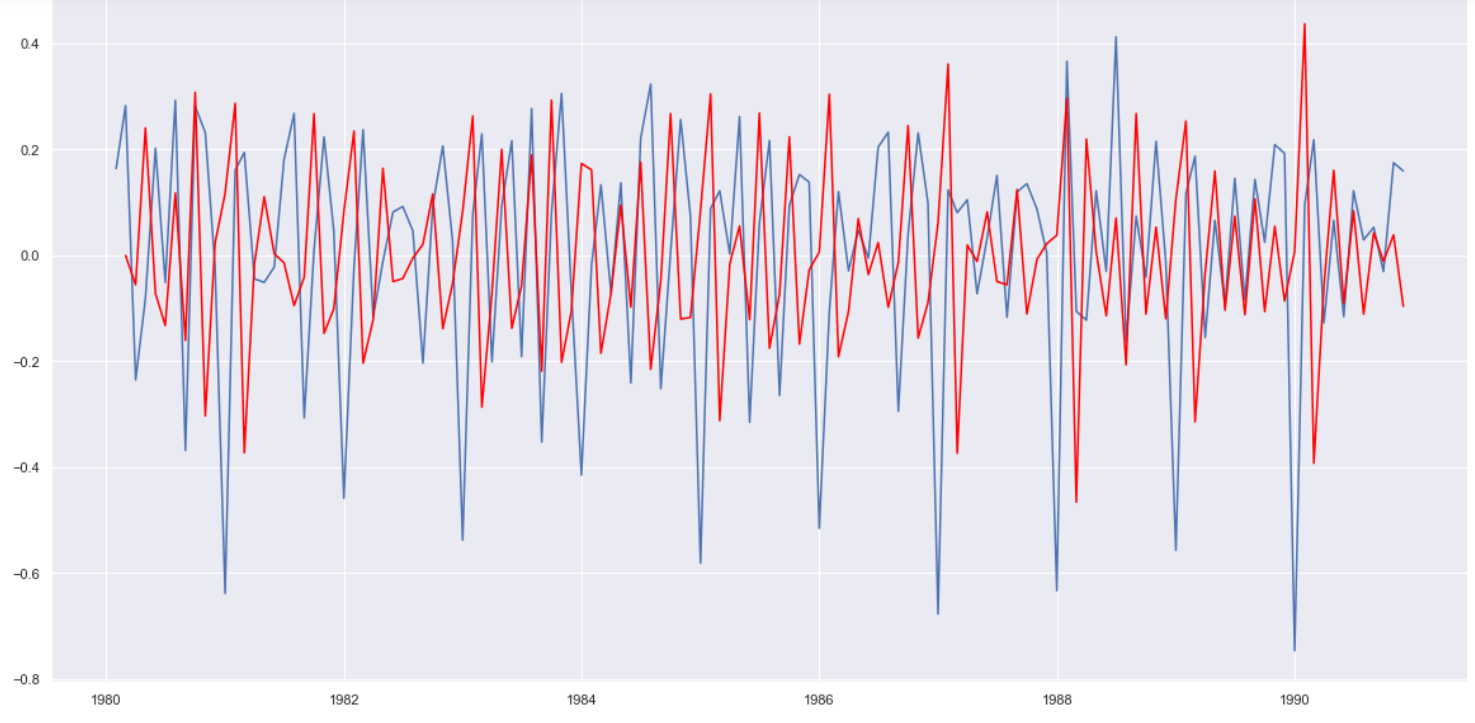
****

* **PACF**

****

**We can see the in ACF plot, it is not crossing the upper bound anytime So, the value of q will be 0. In PACF it is crossing one time So, p value will become 1.**

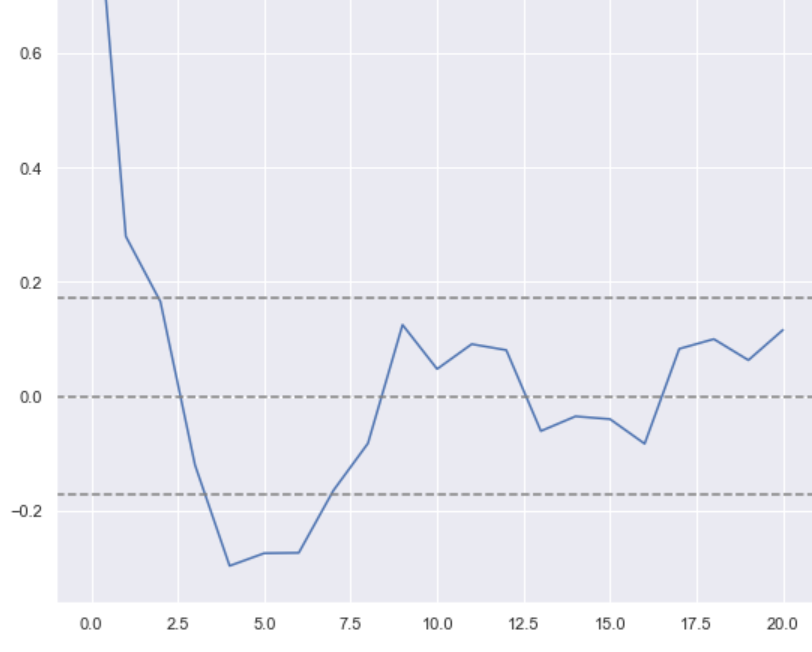
**Let’s model it (ARIMA (1,1,0))**

****

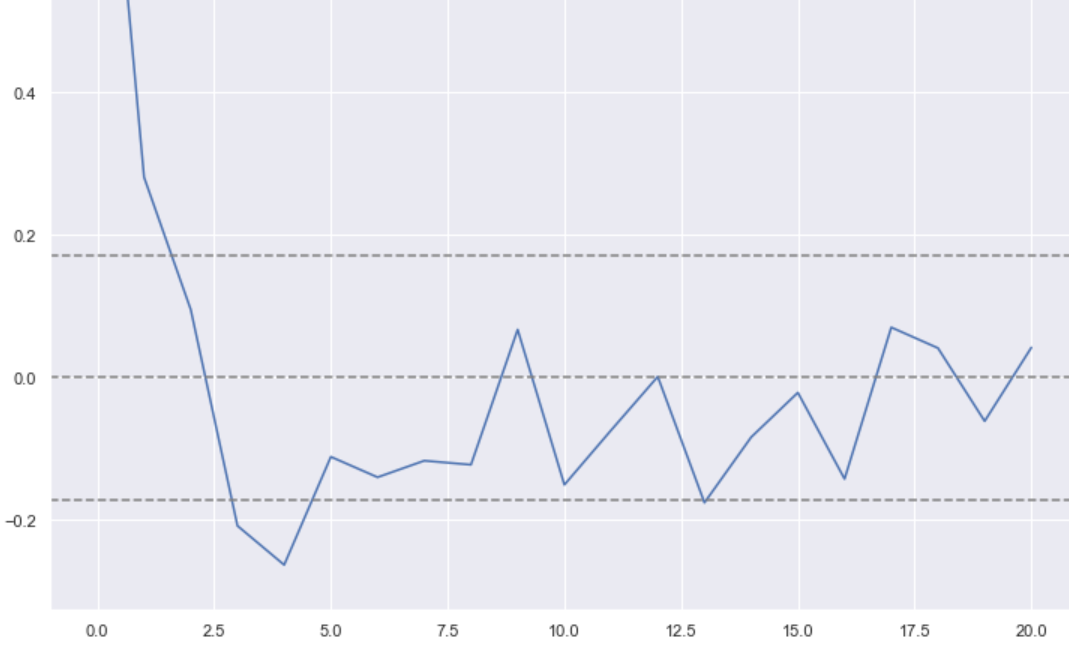
**We can the red color is predicted by arima model and blue one is training data.**

**Shoe Data**

* **ACF plot**

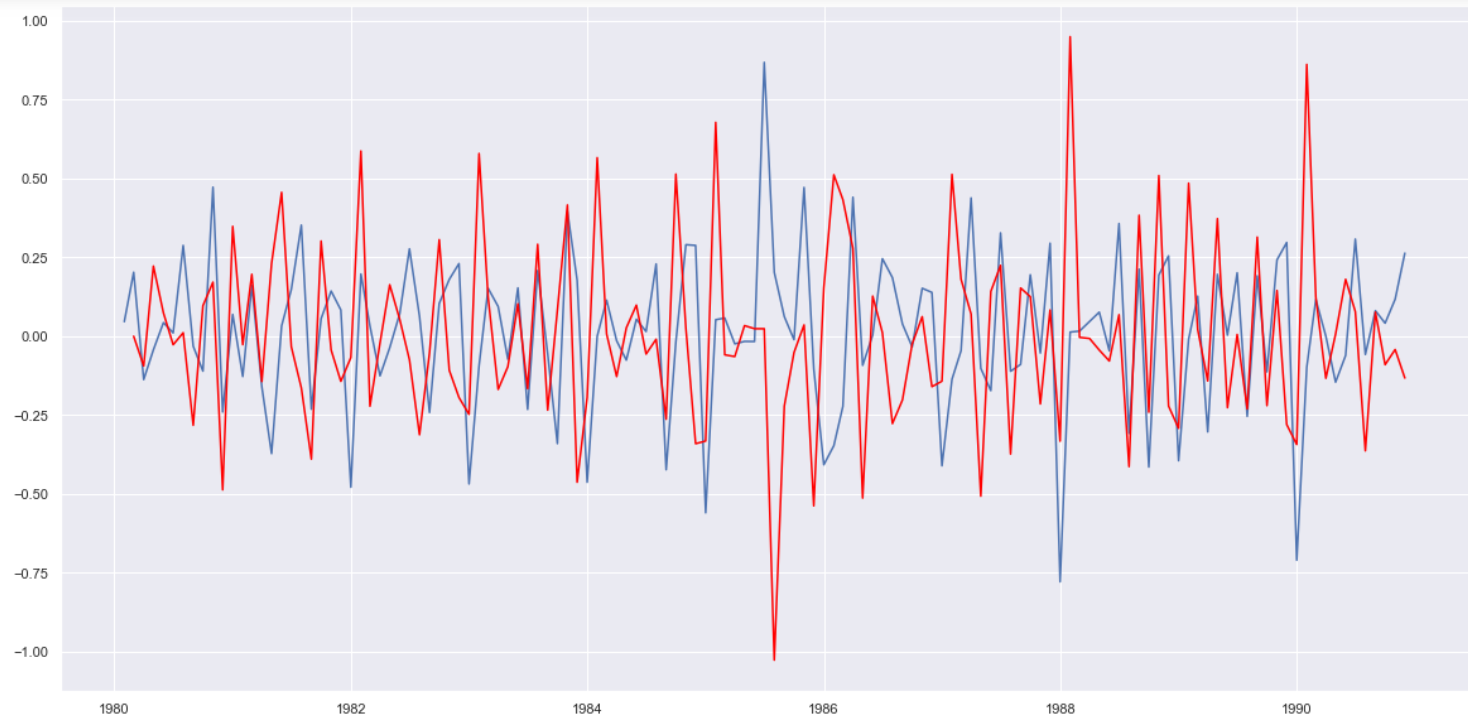
****

* **PACF**

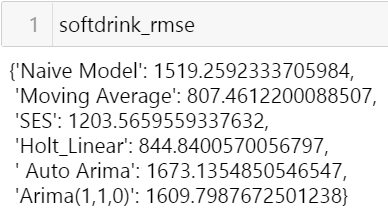
****

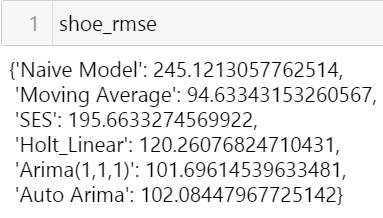
By this plot, we can clearly see the value of p= 0, q=1,d=1

Lets model it (ARIMA (0,1,1))

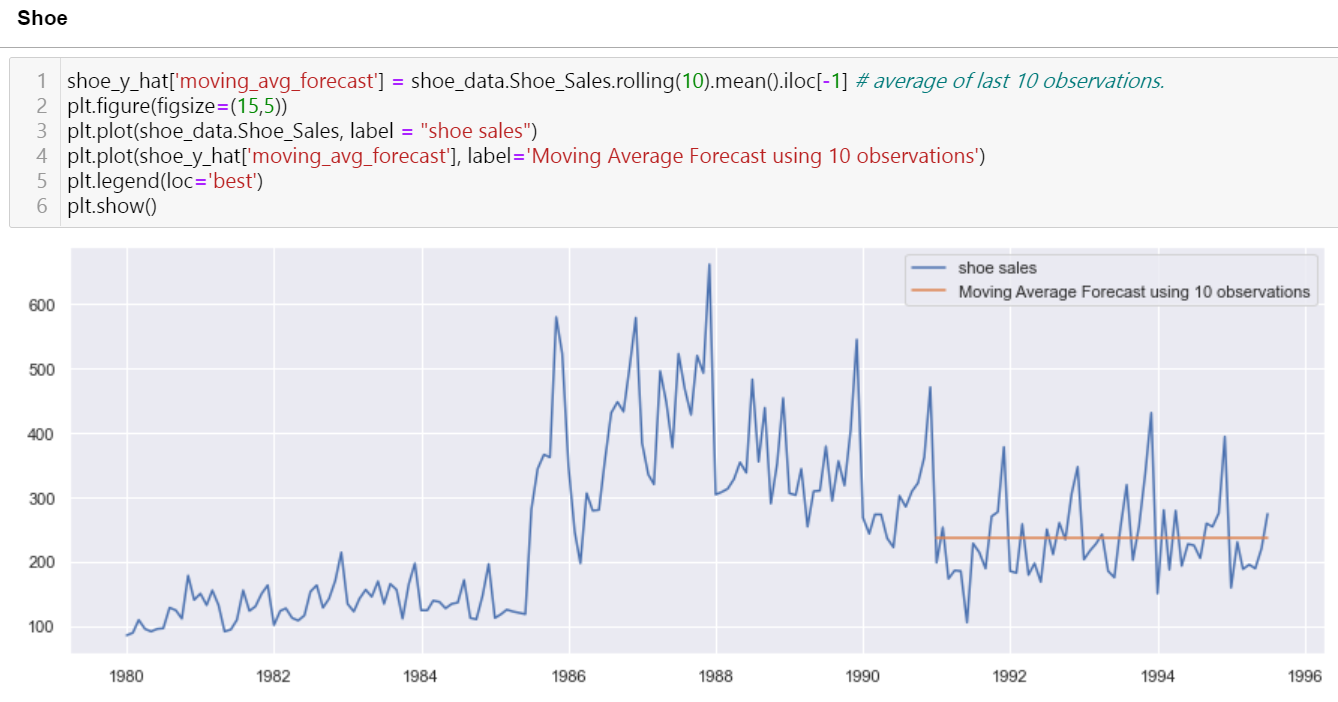


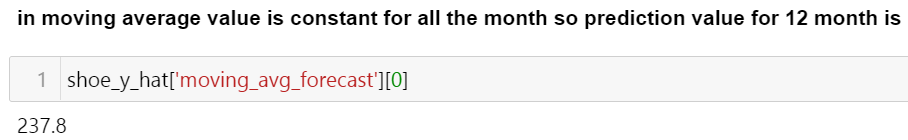
**Final Table with RMSE and parameter value**

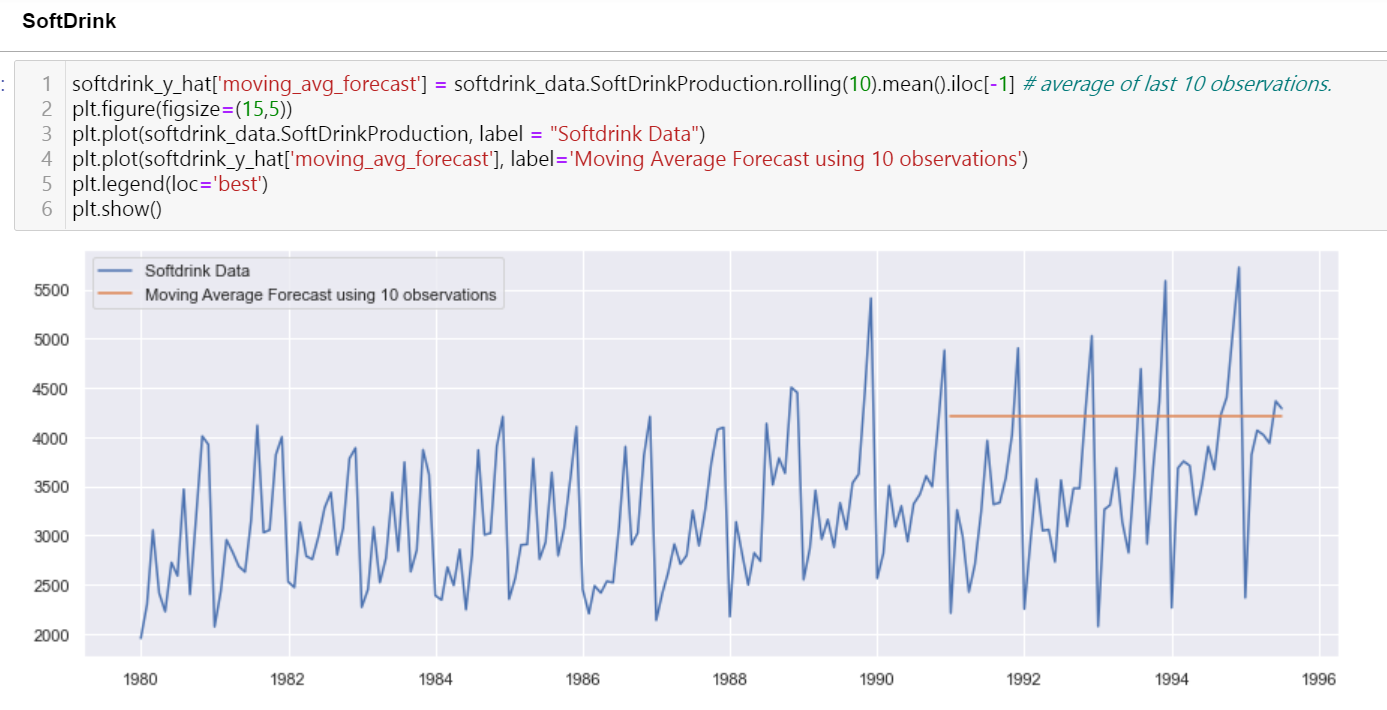
****

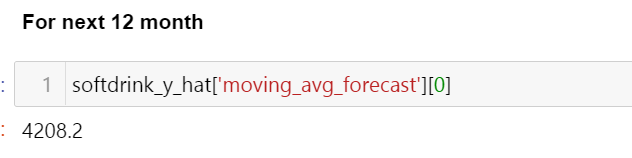
****

**The best model in my case is Moving Average So, for finalizing, I am going with the Moving average.**

****

****

****

****

**For Shoe Sales Company:- I would like to say that their revenue is decreasing because of less sales So they should focus on promoting the sales and They their sales increases in winters So, they should more do production because the winter come.**

**For Soft drink Production Company:- I would say they are doing well and their production of soft drink is increasing year by year but there is one thing on which they should focus and that is, generally people drink more in summers but they don’t produce too much in summer. They are producing more in December. So try to give more offers in summer.**