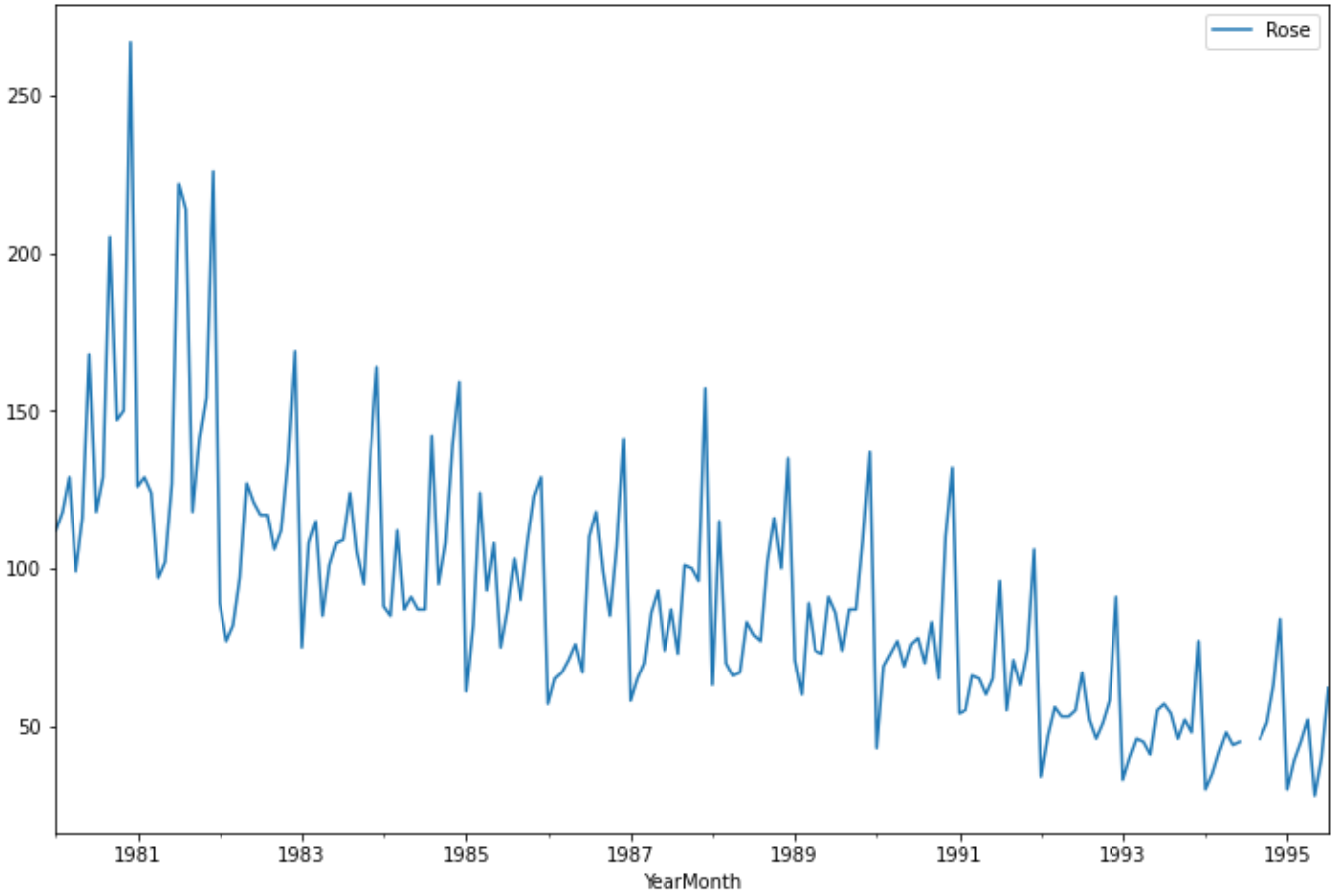
TSF – ROSE

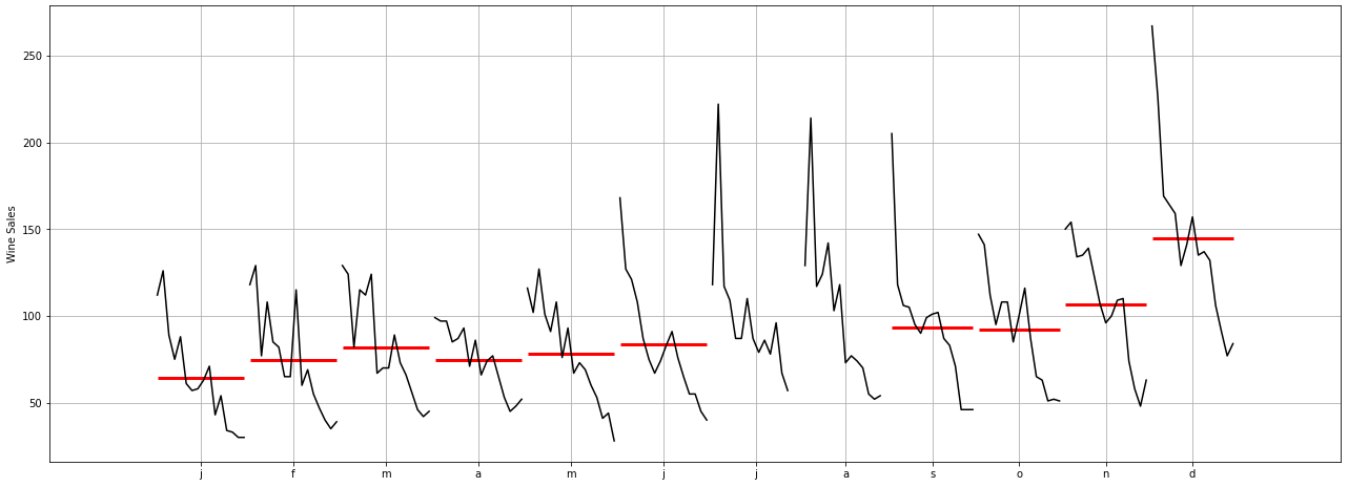
1. Data has been read as an appropriate time series and has been plotted.



We can see a decreasing trend and seasonality which is not constant in nature.

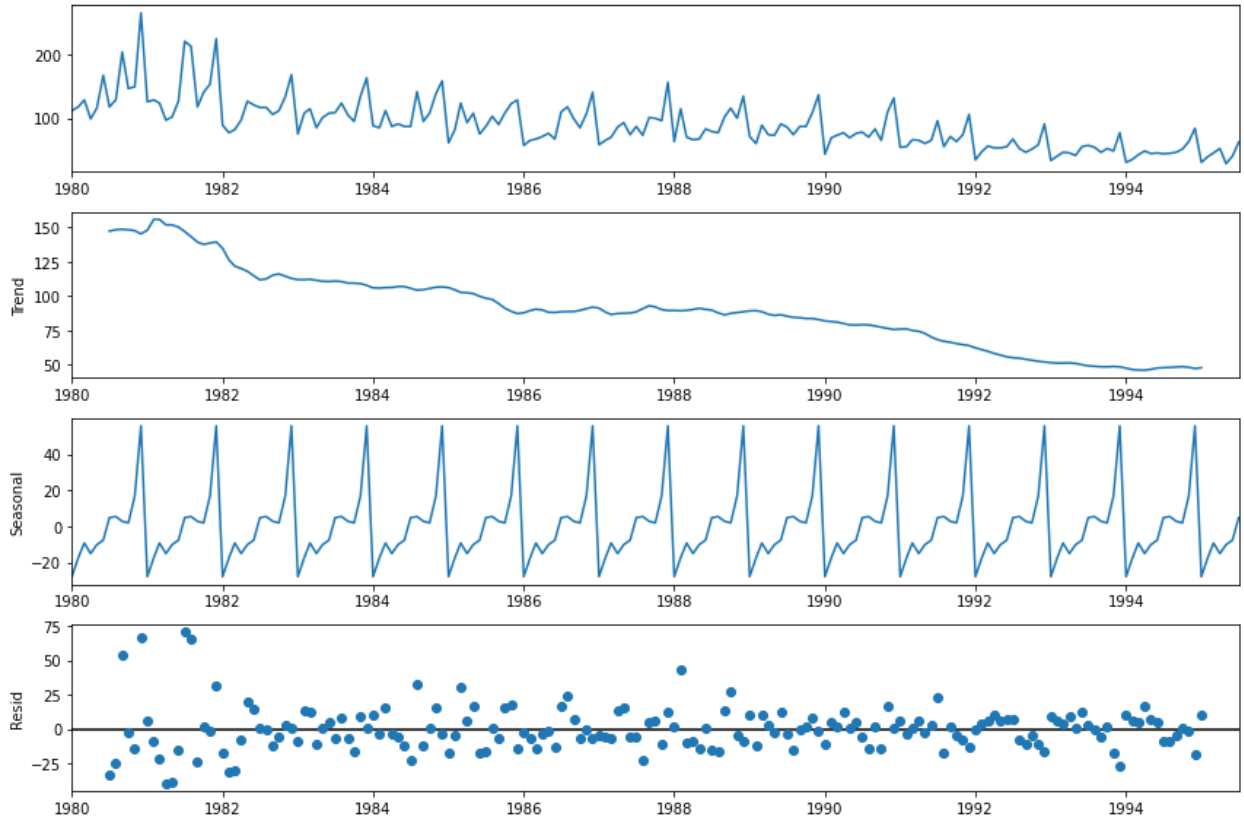
2. Basic EDA have been performed on the dataset. Please refer to the Jupyter notebook for the same.

The dataset had 2 Missing values which have been imputed accordingly.

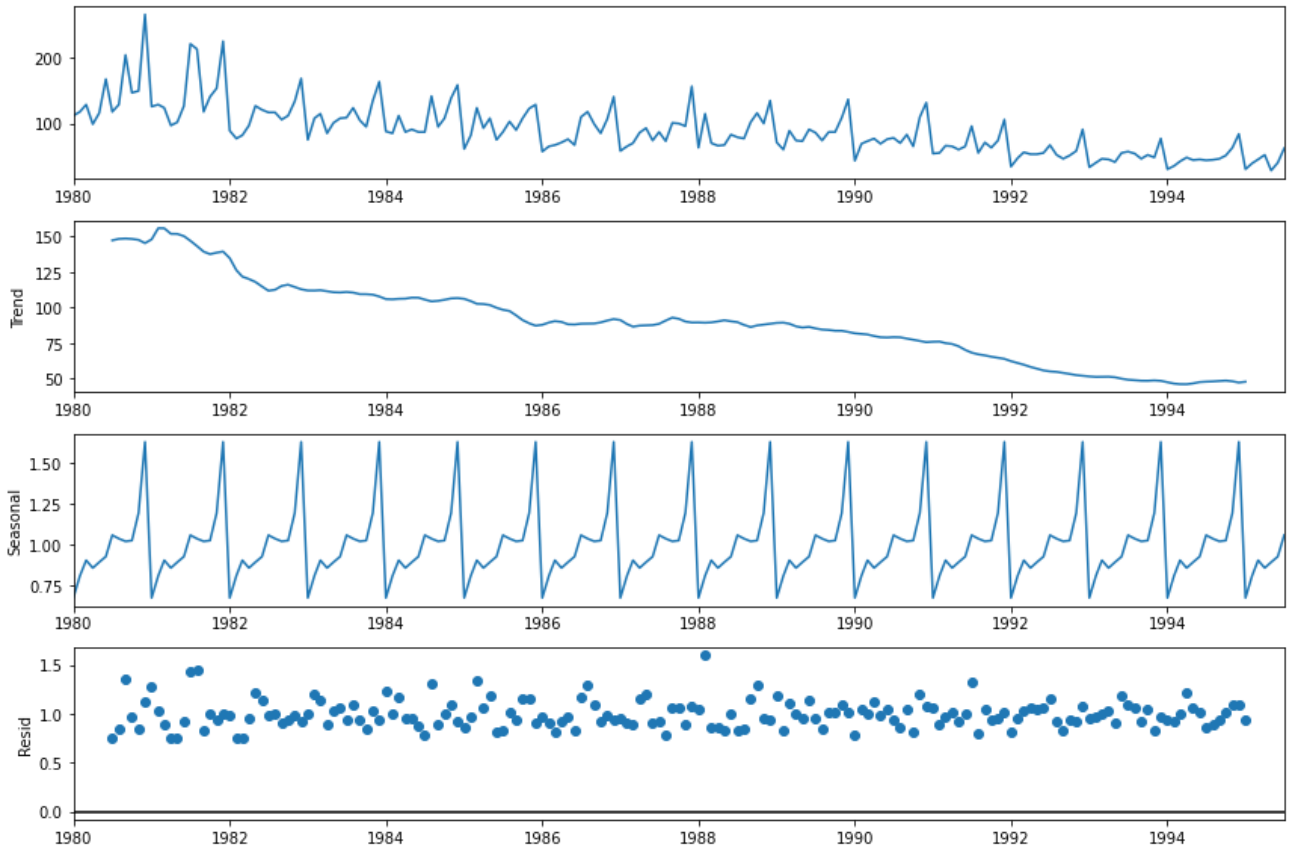


The above figure shows the month plot of the given time series.

Both Additive and Multiplicative decomposition have been performed.



Additive Decomposition



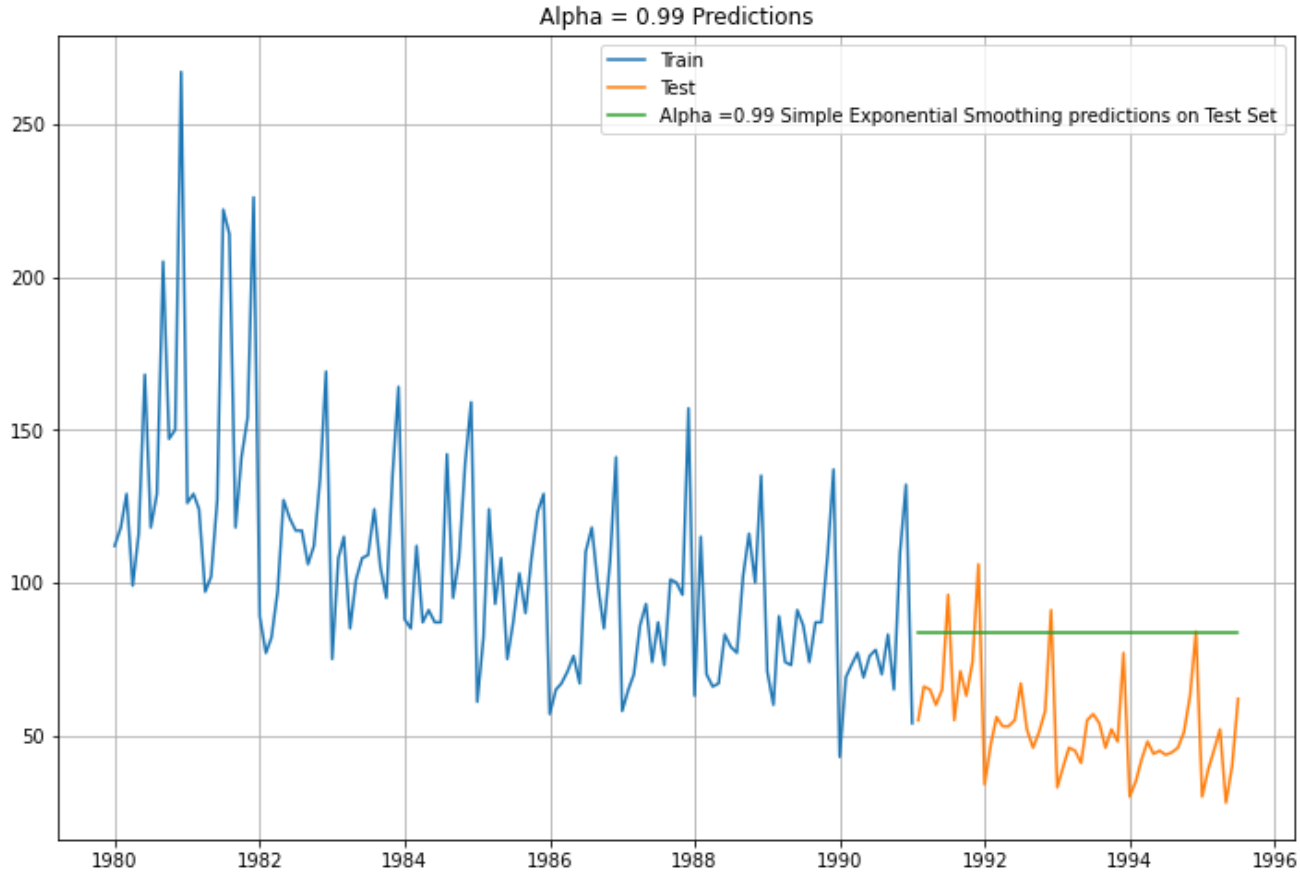
Multiplicative Decomposition

Comparing additive with multiplicative, we choose to go with multiplicative here.

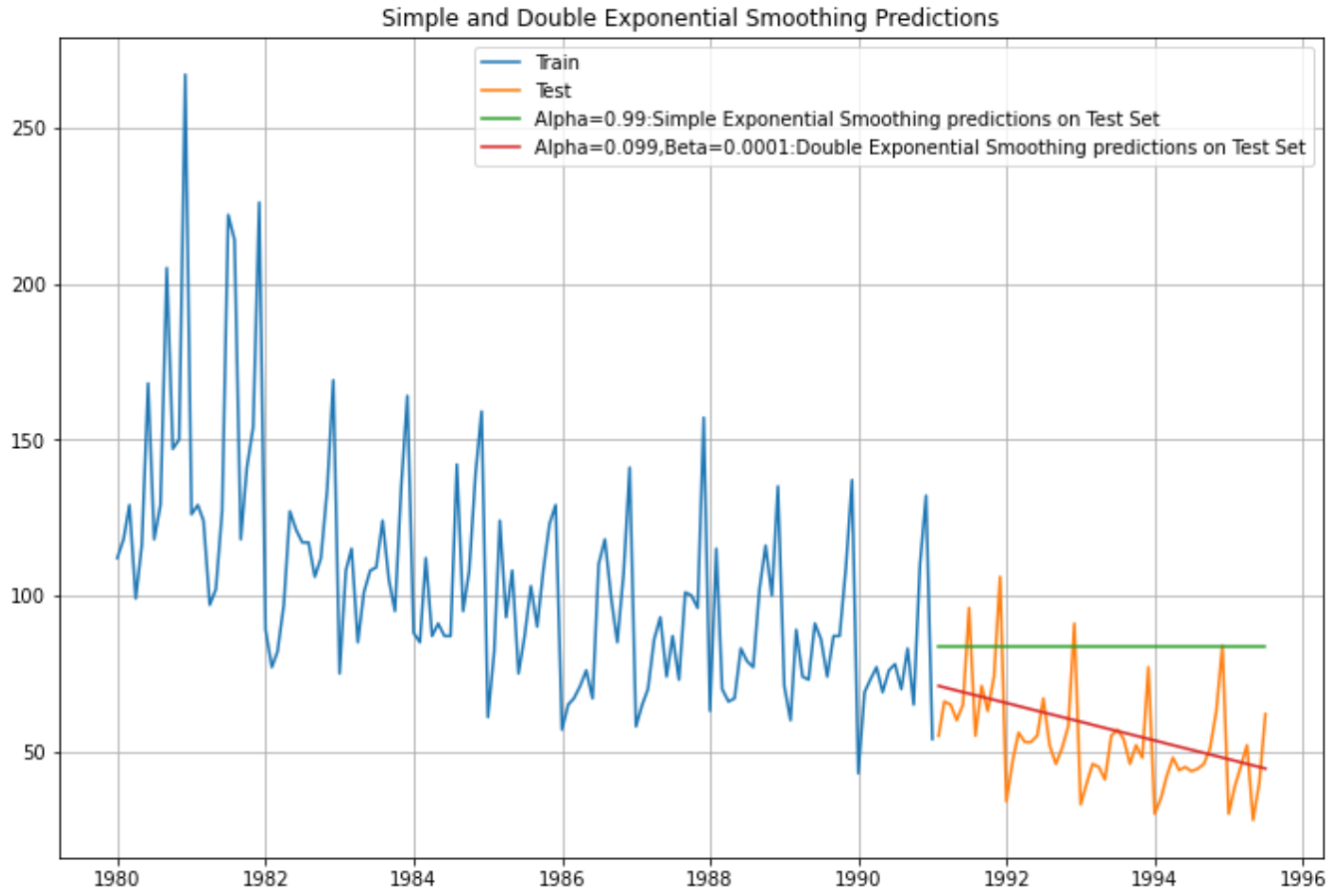
3. Train test split has been done satisfying the necessary requirements.

4. Various exponential smoothing models have been built and the performance have been measured on test data which have been summarized below.

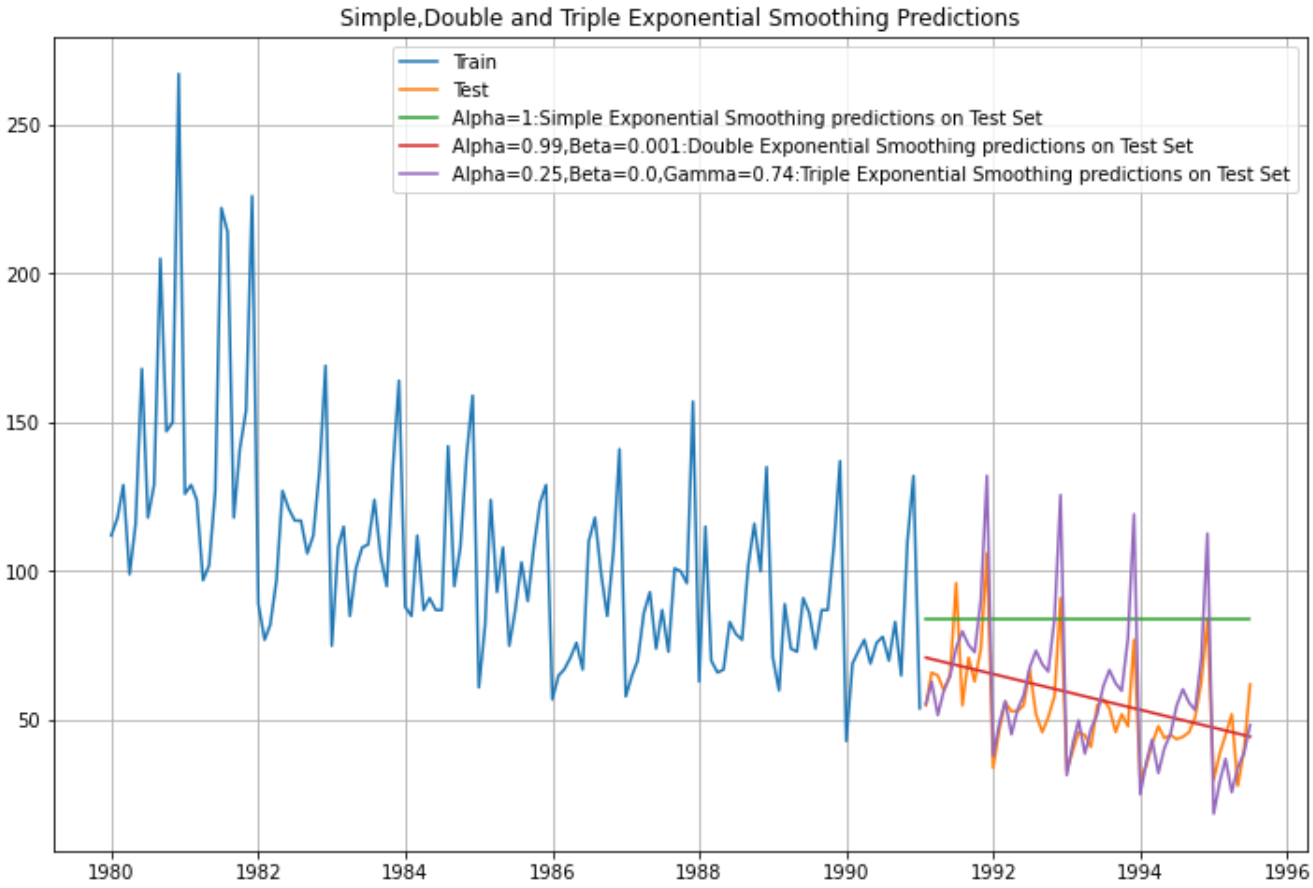
Simple Exponential Smoothing with additive errors



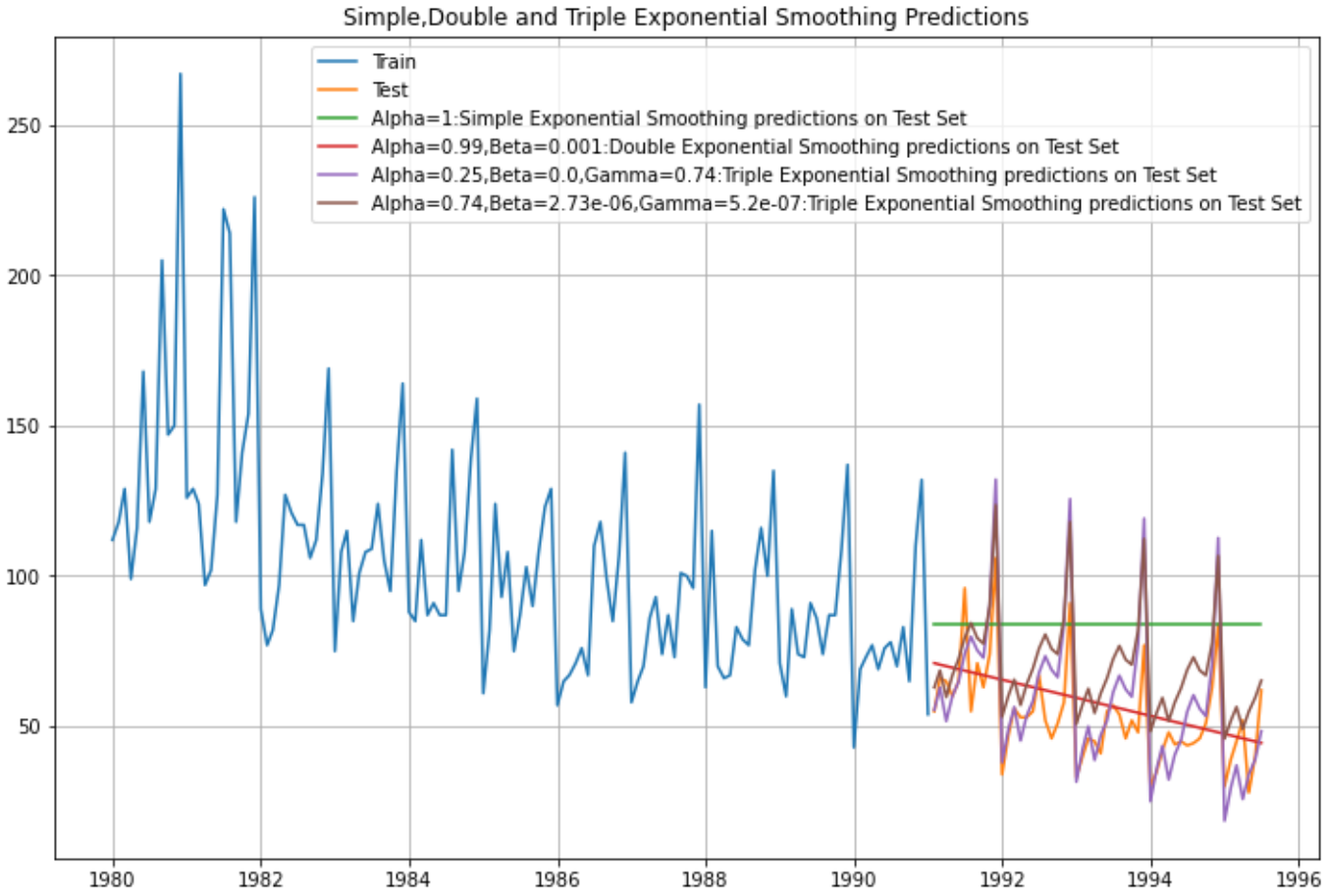
Holt's linear method with additive errors - Double Exponential Smoothing



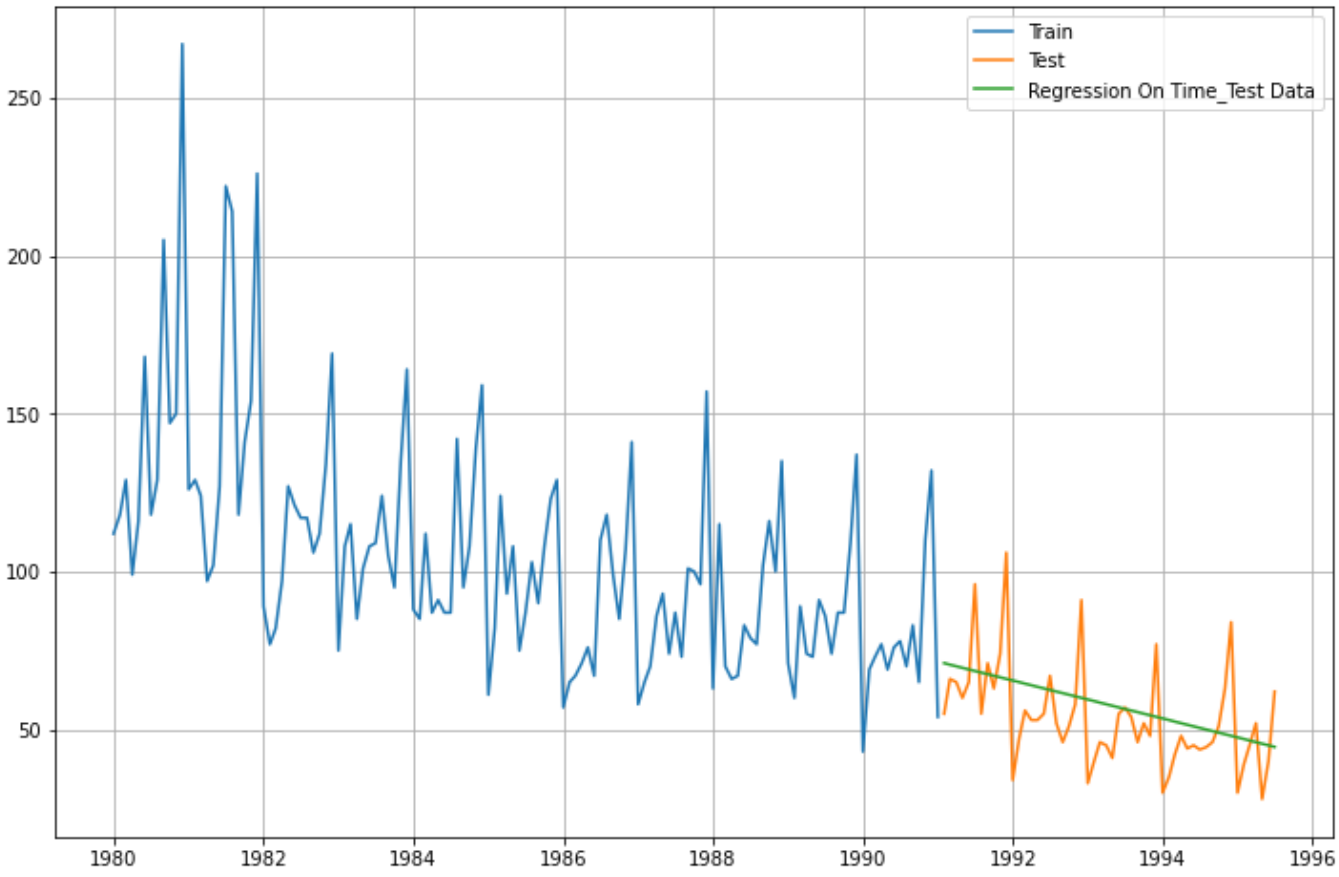
Holt Winter's linear method with additive errors



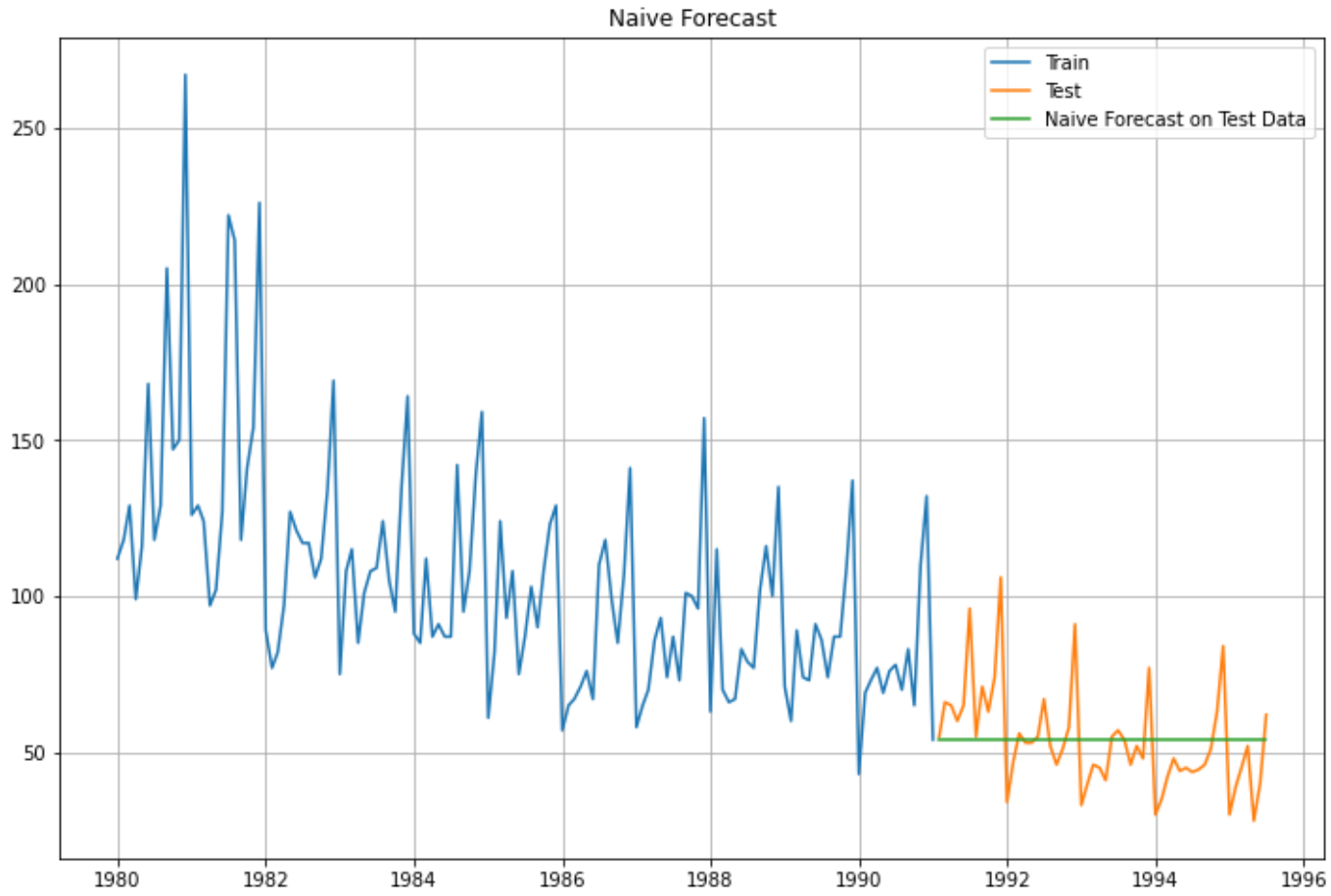
Holt Winter's linear method - ETS (A, A, M) model



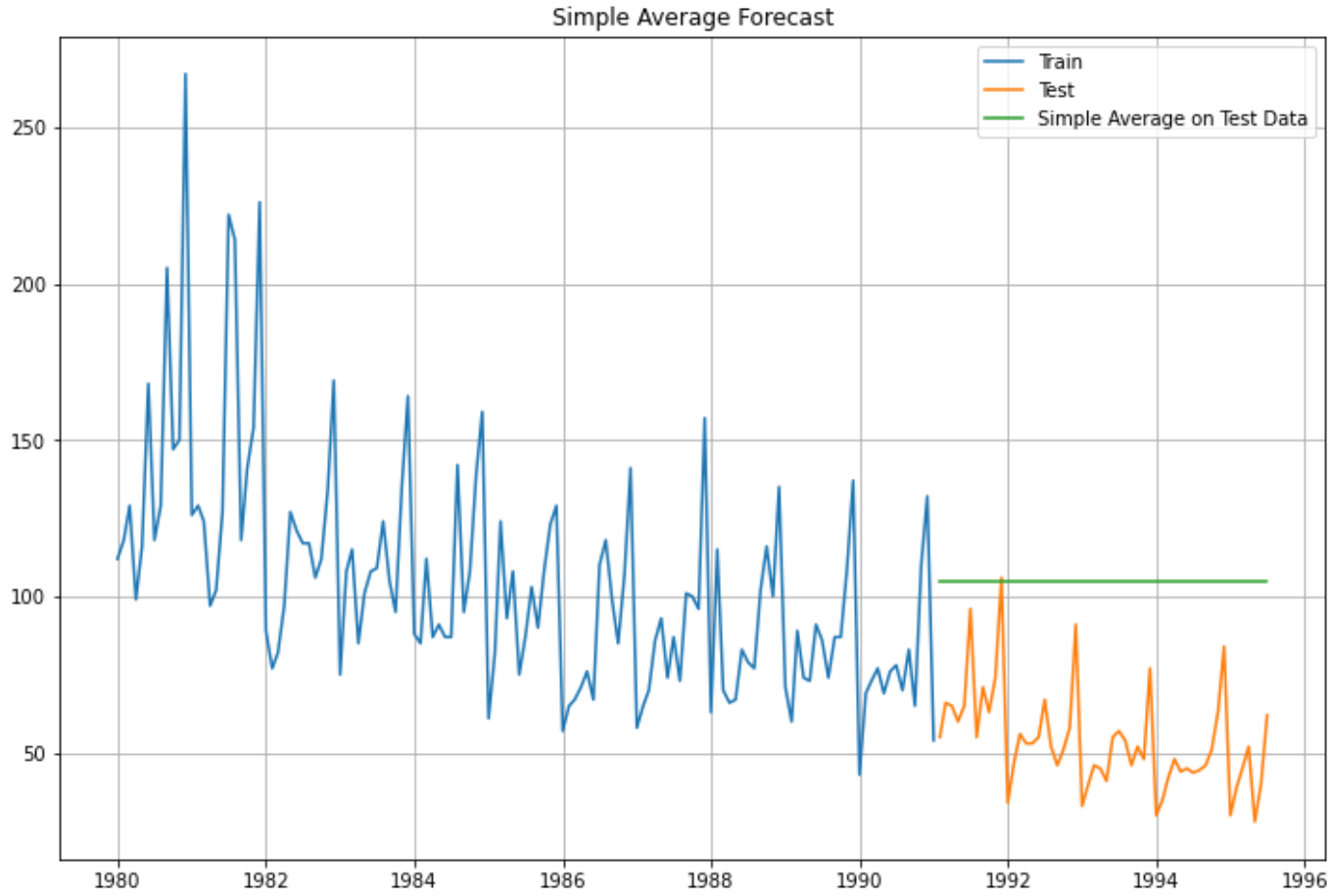
Linear Regression



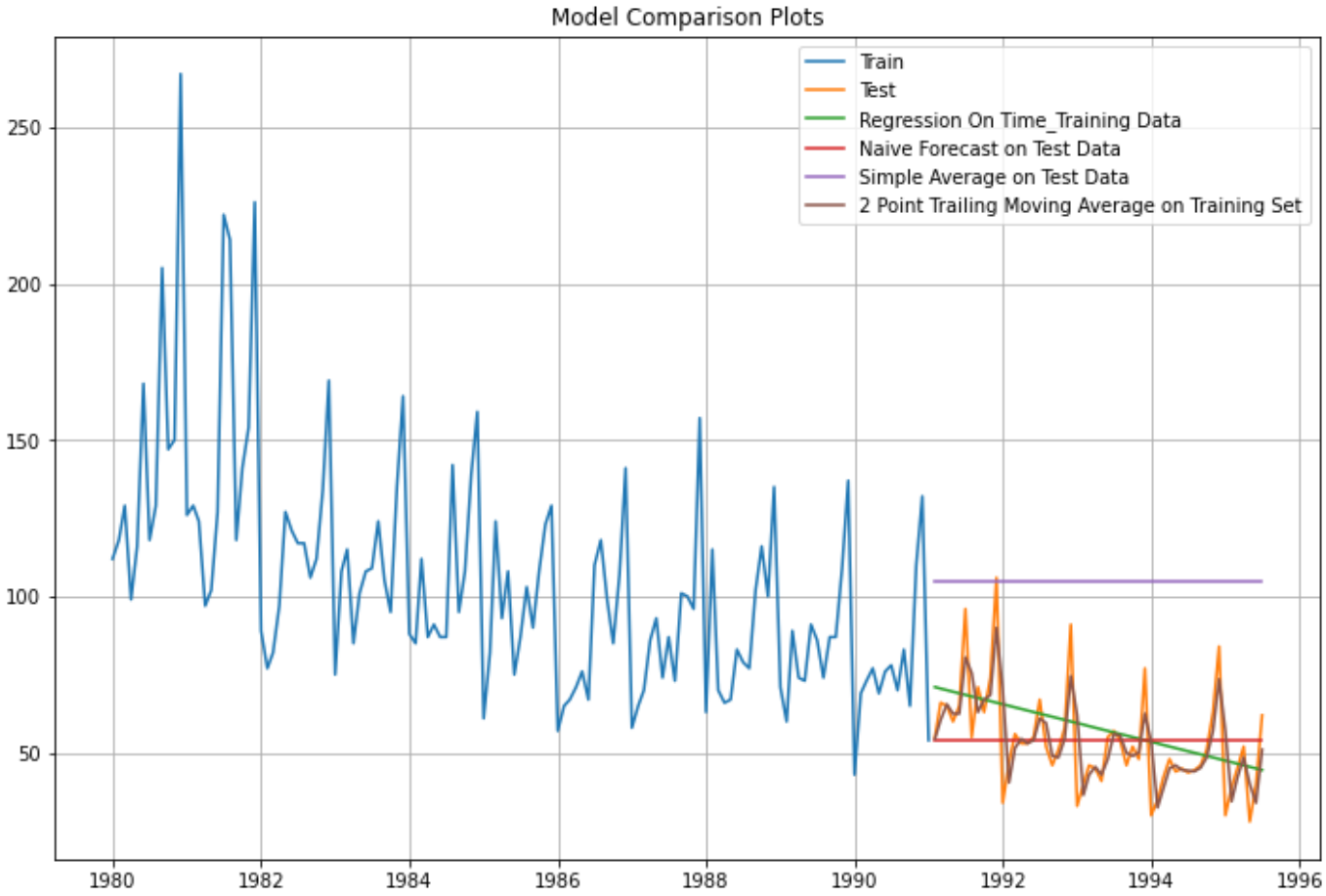
Naive Approach



Simple Average



Moving Average (MA)



RMSE for the models built:

	Test RMSE
Alpha=0.99, SES	34.012215
Alpha=1, Beta=0.0189: DES	15.040859
Alpha=0.25, Beta=0.0, Gamma=0.74: TES	14.670057
Alpha=0.74, Beta=2.73e-06, Gamma=5.2e-07, Gamma=0: TES	18.675110
RegressionOnTime	15.040846
NaiveModel	15.937414
SimpleAverageModel	53.201811
2pointTrailingMovingAverage	10.355691

4pointTrailingMovingAverage	13.736403
6pointTrailingMovingAverage	14.076433
9pointTrailingMovingAverage	14.383541

5. The Augmented Dickey-Fuller test is a unit root test which determines whether there is a unit root and subsequently whether the series is non-stationary.

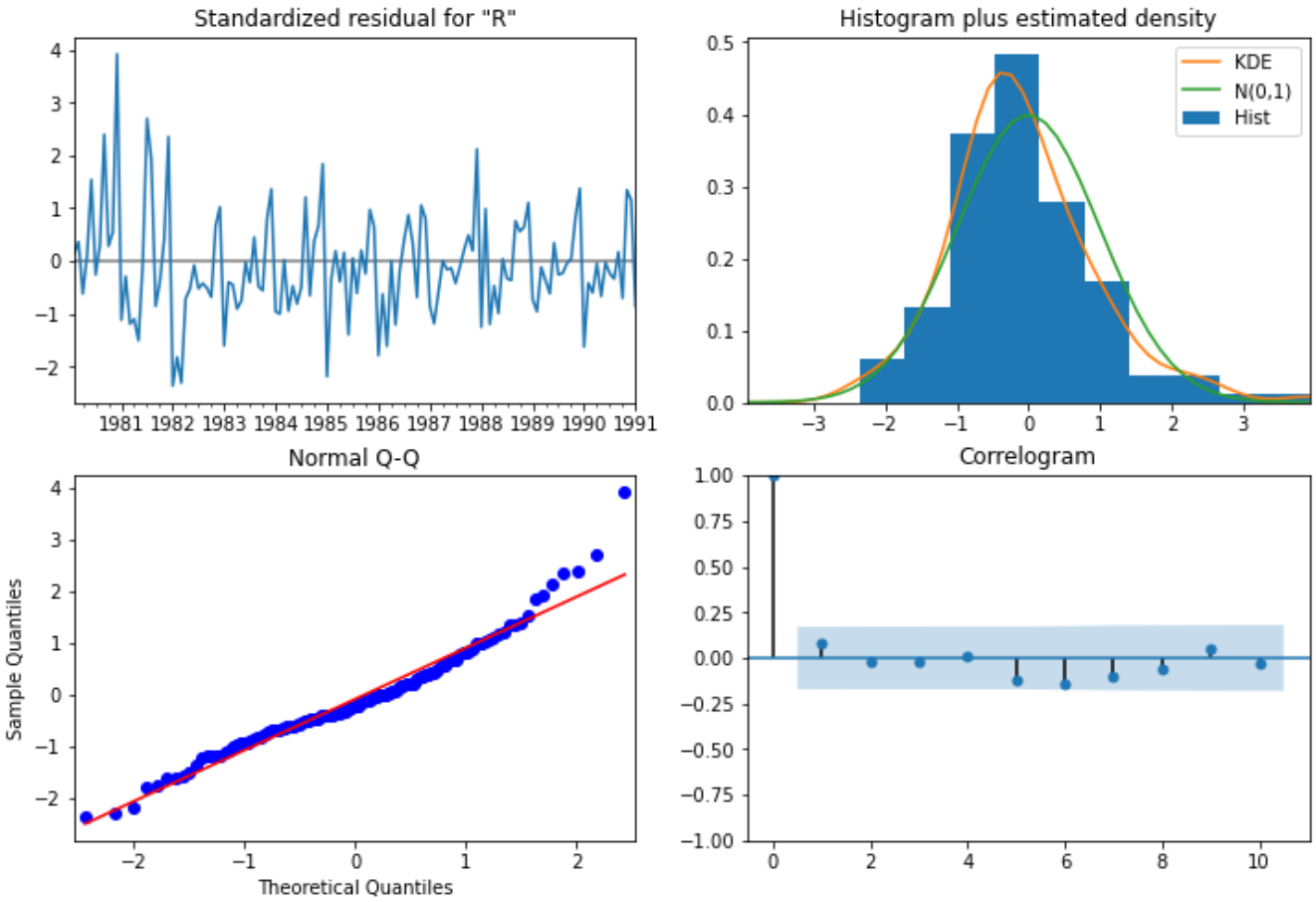
   The hypothesis in a simple form for the ADF test is:

- $H0$: The Time Series has a unit root and is thus non-stationary.
- $H1$: The Time Series does not have a unit root and is thus stationary.

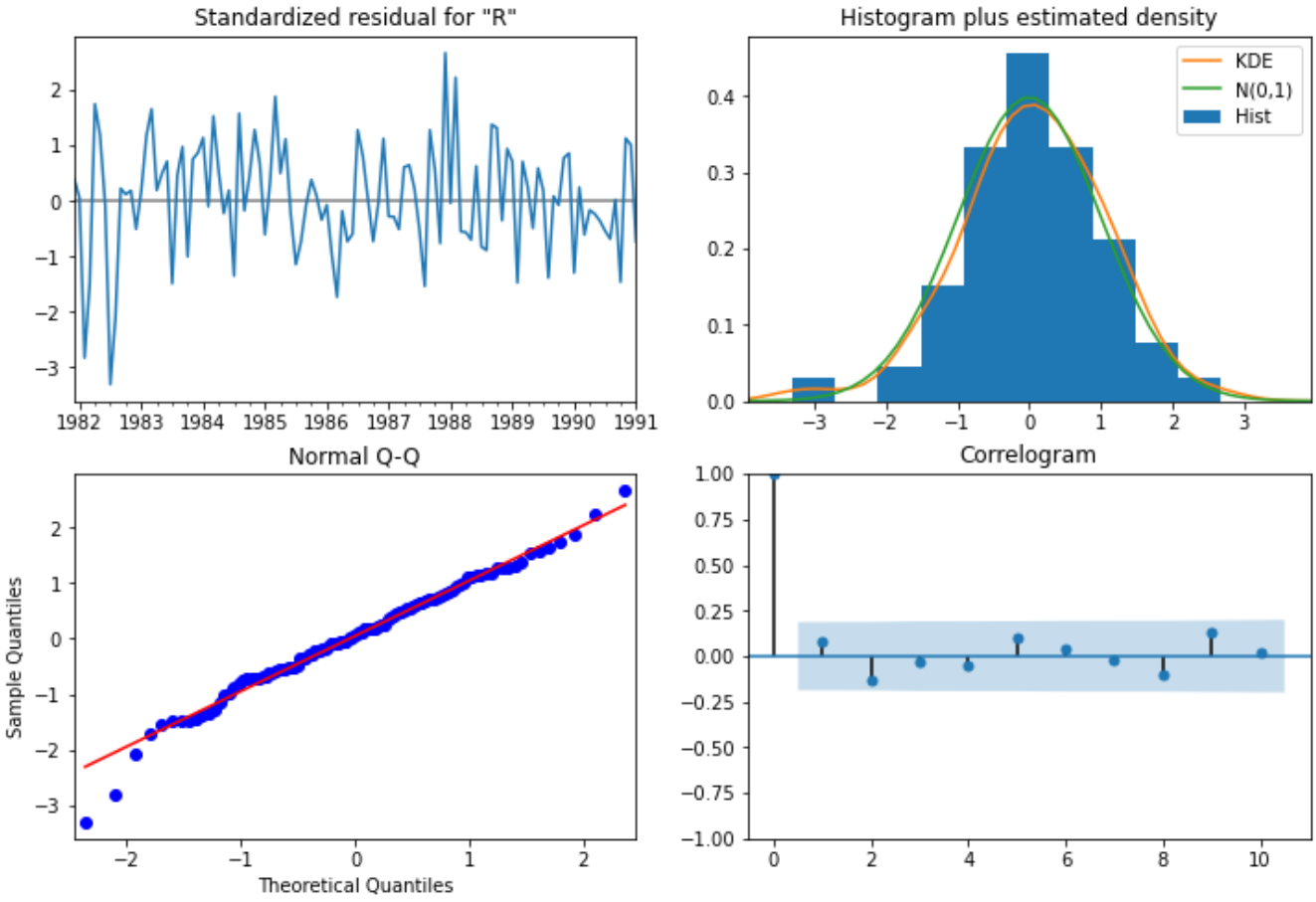
We would want the series to be stationary for building ARIMA models and thus we would want the p-value of this test to be less than the α value.

Data has been found to be non – stationary and necessary steps have been taken to make the data stationary. Stationarity is checked at alpha = 0.05.

6. Automated versions of both ARIMA and SARIMA models have been built in which the parameters were selected using the lowest Akaike Information Criteria (AIC) on the training data and has been evaluated on the test data using RMSE.



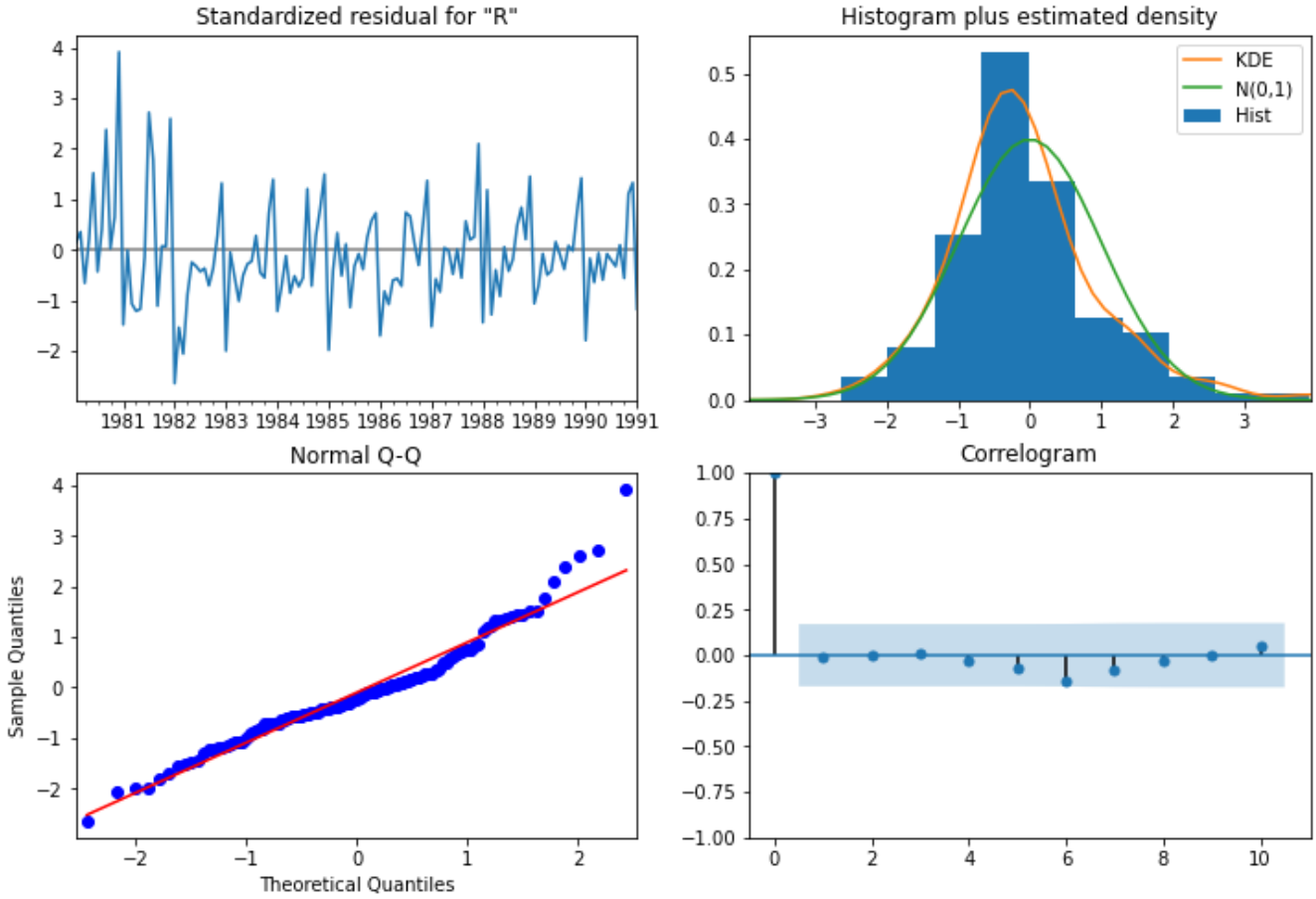
Automated ARMIA diagnostics



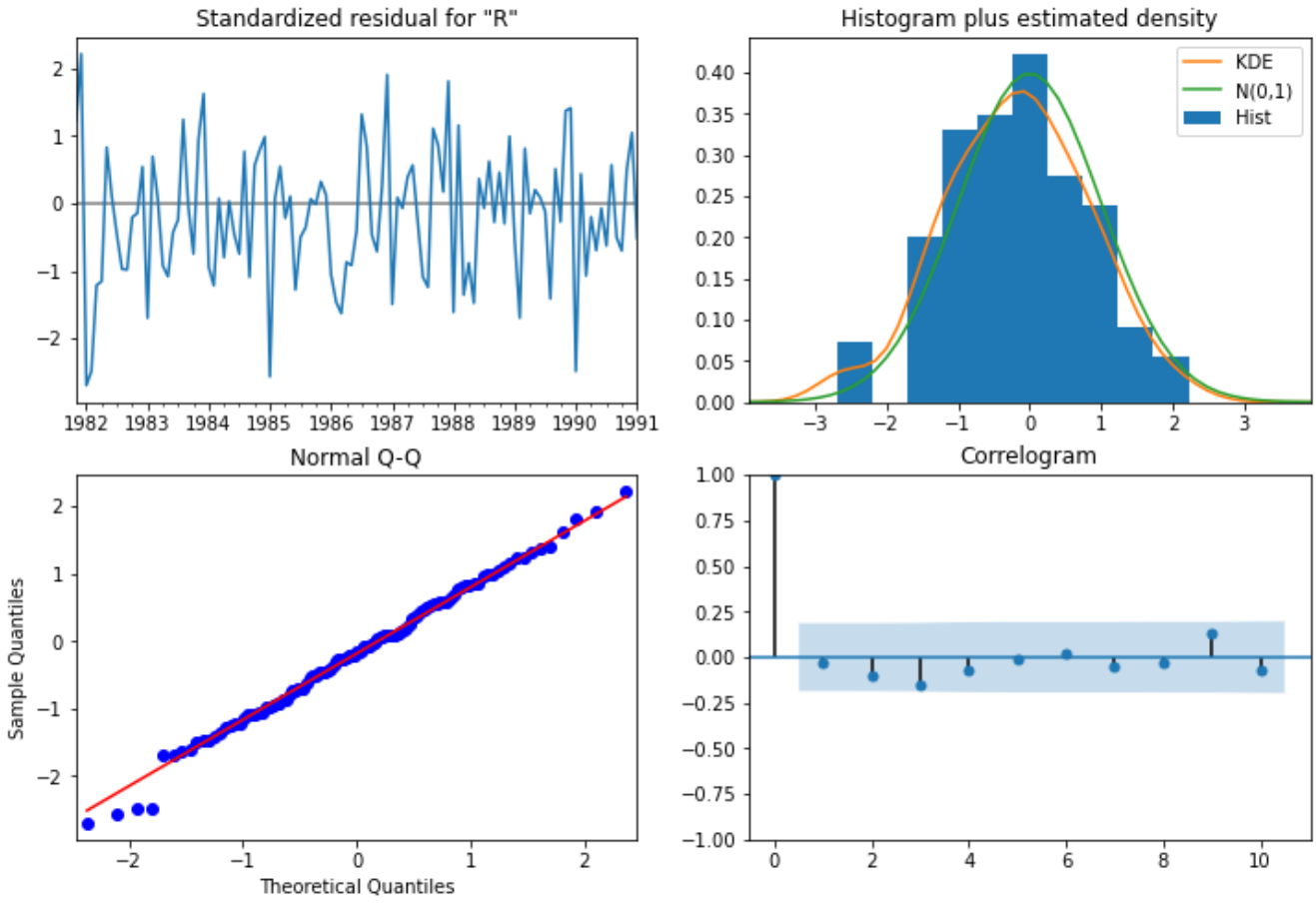
Automated SARIMA diagnostics

	RMSE
ARIMA (2,1,2)	33.955042
SARIMA (2,1,3) (2,0,3,6)	26.567358

7. Manual versions of both ARIMA and SARIMA models have been built in which the parameters were selected manually on the training data and has been evaluated on the test data using RMSE.



Manual ARIMA diagnostics



Manual SARIMA diagnostics

	RMSE
ARIMA (2,1,2)	33.532500
SARIMA (2,1,2) (0,0,3,6)	31.879425

8. Consolidated table of all different RMSEs for the different models built.

	Test RMSE
Alpha=0.99, SES	34.012215
Alpha=1, Beta=0.0189: DES	15.040859
Alpha=0.25, Beta=0.0, Gamma=0.74: TES	14.670057
Alpha=0.74, Beta=2.73e-06, Gamma=5.2e-07, Gamma=0: TES	18.675110
RegressionOnTime	15.040846
NaiveModel	15.937414
SimpleAverageModel	53.201811
2pointTrailingMovingAverage	10.355691
4pointTrailingMovingAverage	13.736403
6pointTrailingMovingAverage	14.076433
9pointTrailingMovingAverage	14.383541
ARIMA (2,1,2) *(Automated)*	33.955042
SARIMA (2,1,3) (2,0,3,6) *(Automated)*	26.567358
ARIMA (2,1,2) *(Manual)*	33.532500
SARIMA (2,1,2) (0,0,3,6) *(Manual)*	31.879425

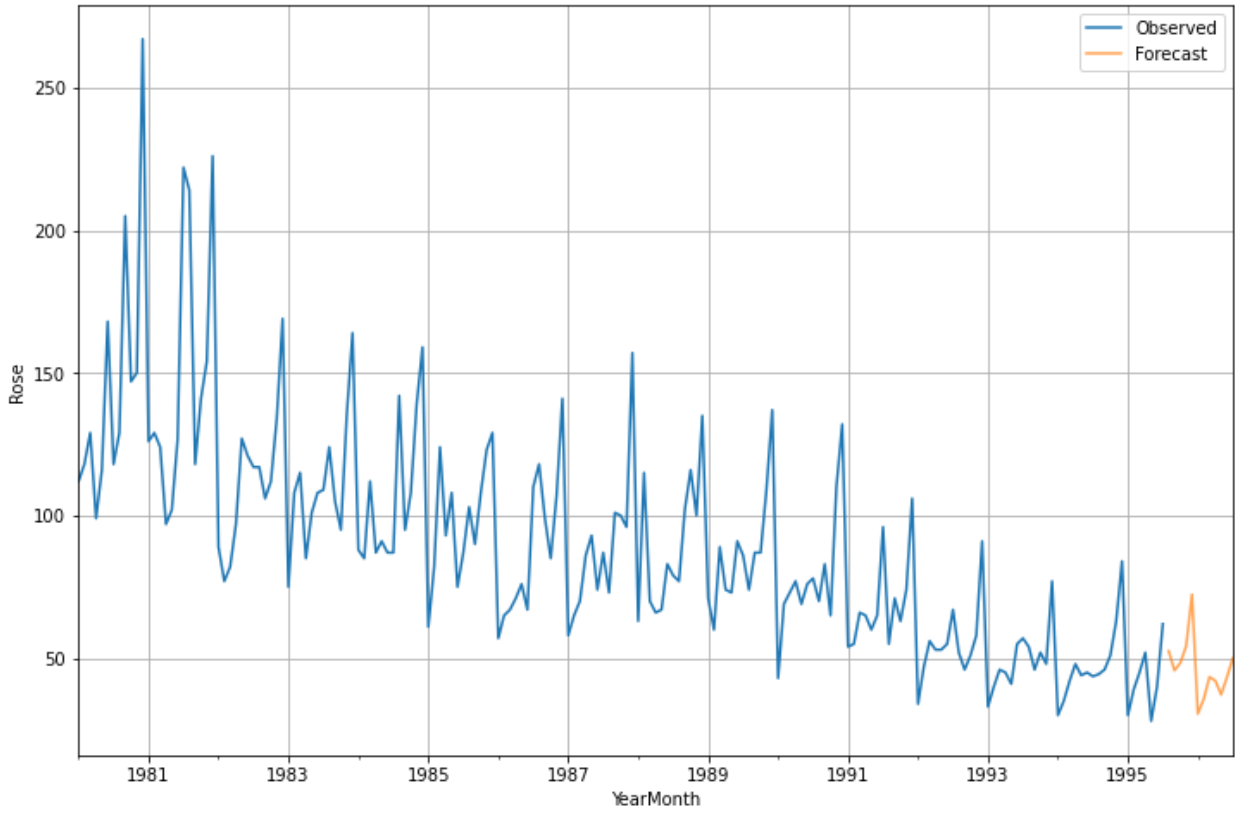
RMSE is a way to measure the error of a model predicting quantitative data. Since it is an error value, the model with the least RMSE value will give the most accurate predictions.

Of all the different models built on the train data and tested on the test data, for the full data forecast we go with the automated SARIMA model.

9. The most optimum model has been built on the complete data and the immediate 12 months into the future has been predicted.

RMSE of the total data comes to 25.51

10. Various models have been built on the train data, performance has been tested on the test data using the RMSE scores and the best model has been chosen to be built on the entire data for forecasting.



The forecasted values show that our forecast is in line with the actual sales the company has done in the past years proving that this forecast can be relied upon.

From the analysis, it can be seen that the overall demand for this wine product has gradually come down over the years both during the holiday and the non-holiday season.

People had a fair liking for this label in the beginning and it saw a boost in sales during the holiday season, however, gradually, the sales really went downhill over the period of time and if the trend suggests anything, it will continue to go down.

Either the company has to come up with a very innovative marketing strategy and development in the product to attract consumers towards this product or shut down the product before it starts failing to meet the necessary making costs and burdens the owners with the losses.