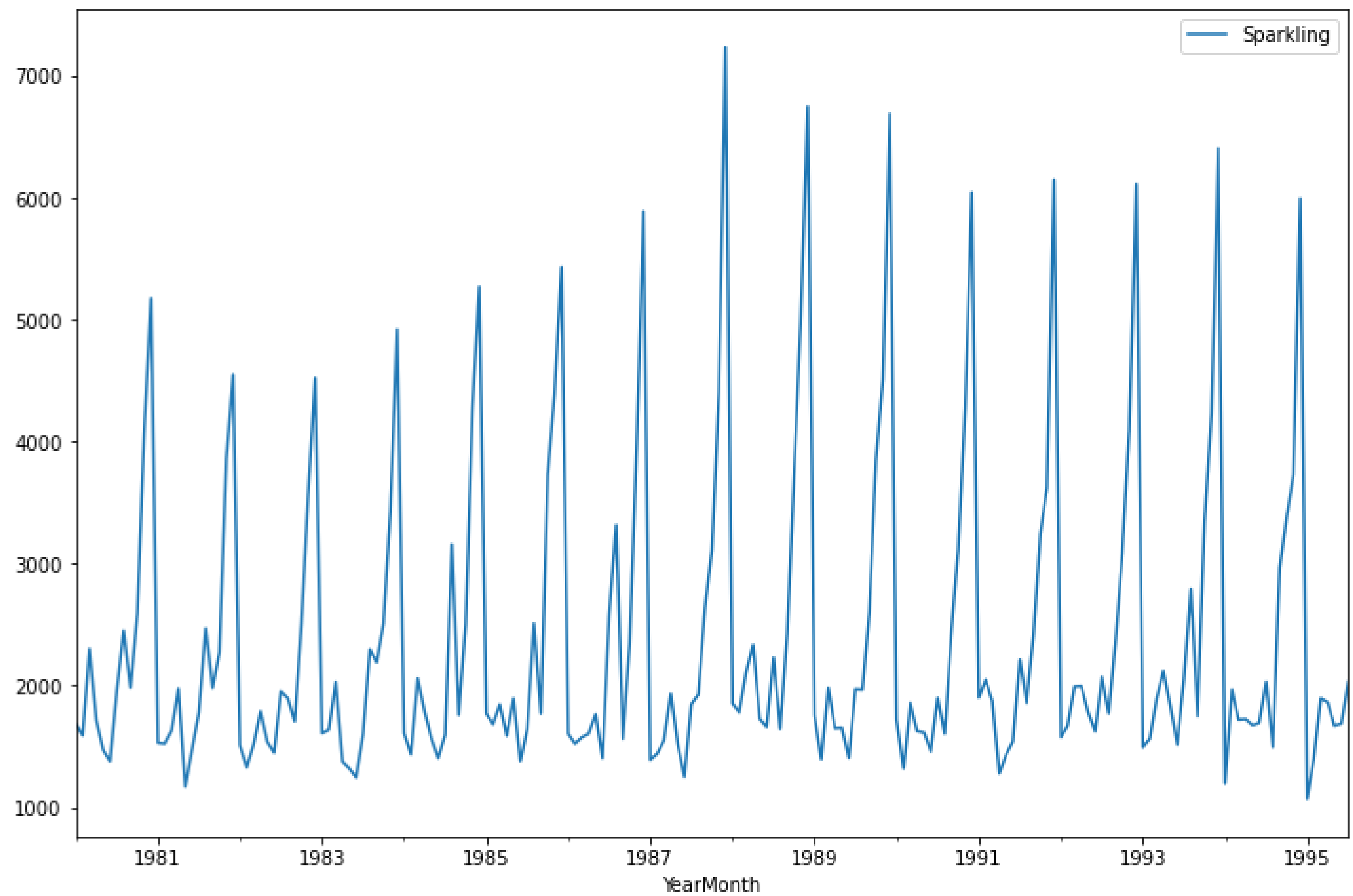
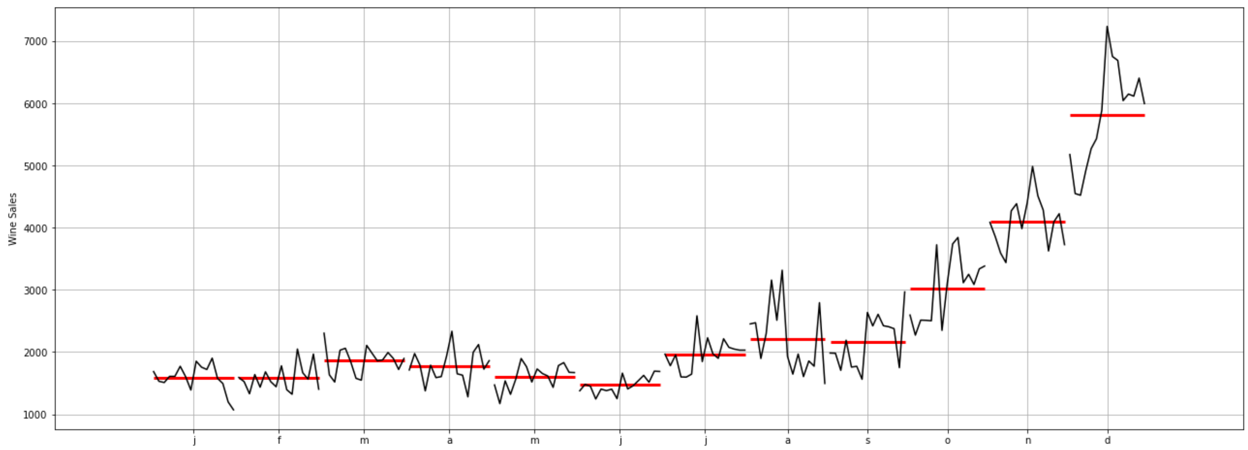
TSF – SPARKLING

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



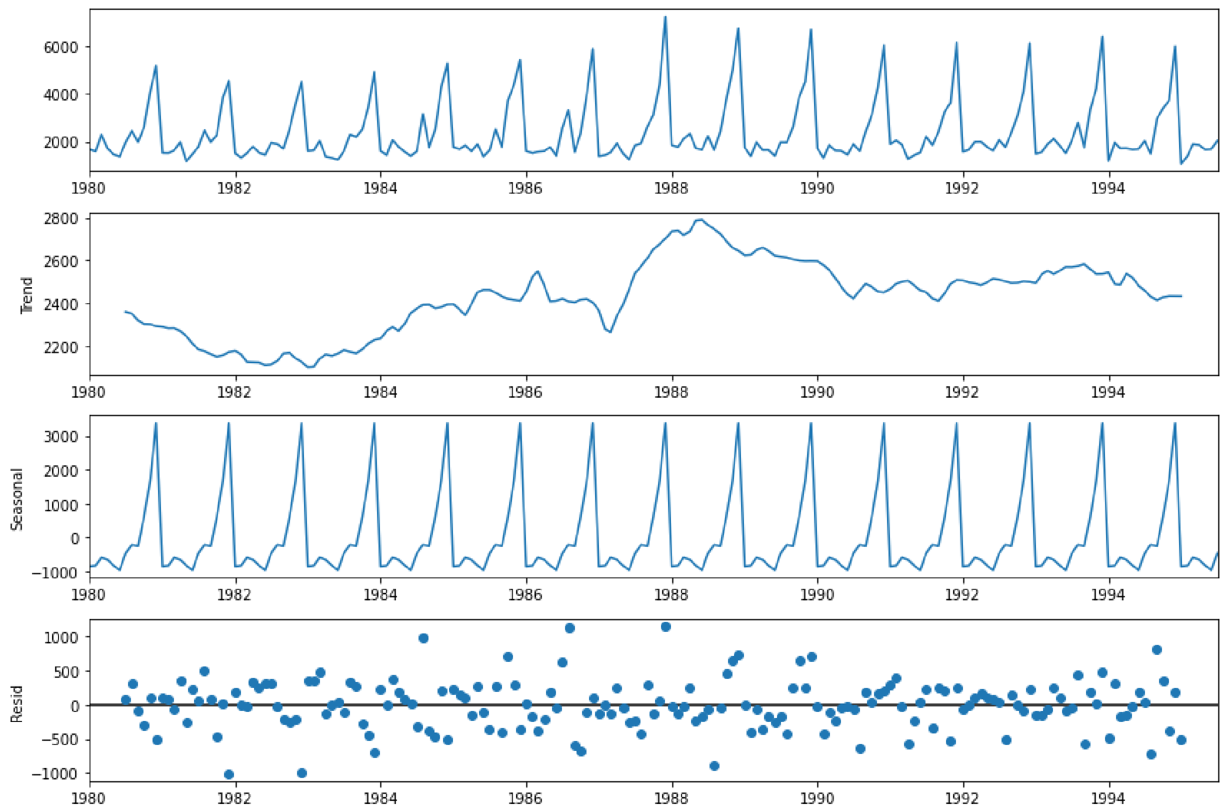
We can see an almost constant trend and a seasonality which is not constant.

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

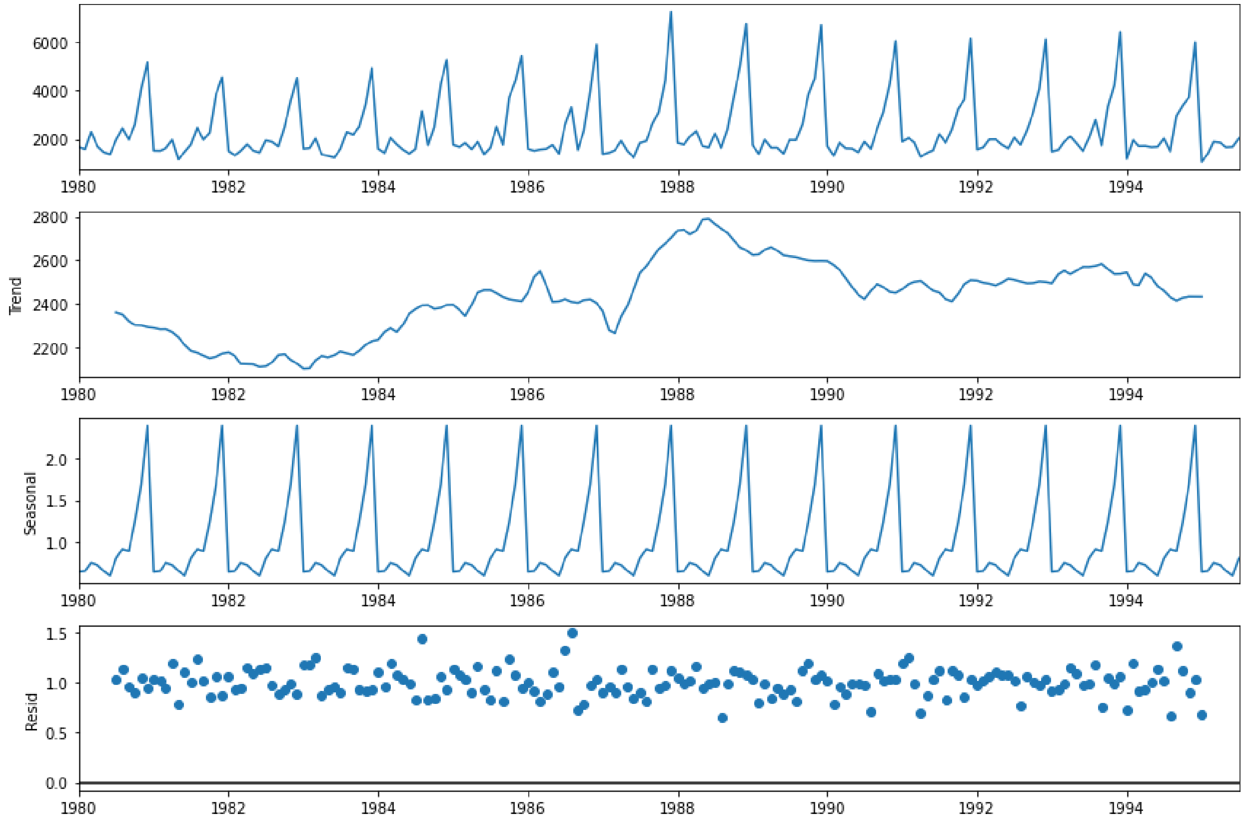


The above figure shows the month plot of the given time series.

Both Additive and Multiplicative decomposition have been performed.



Additive Decomposition



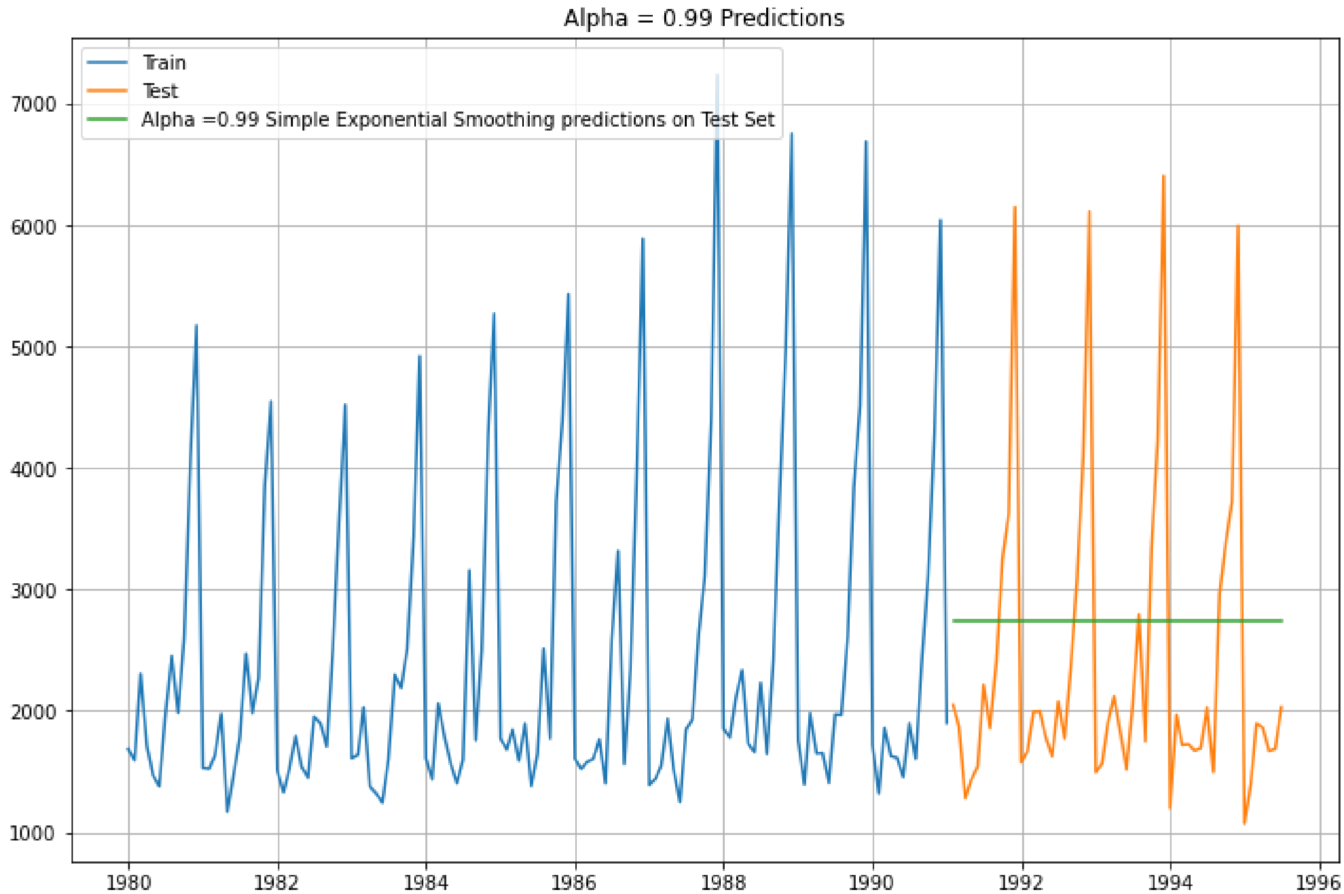
Multiplicative Decomposition

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

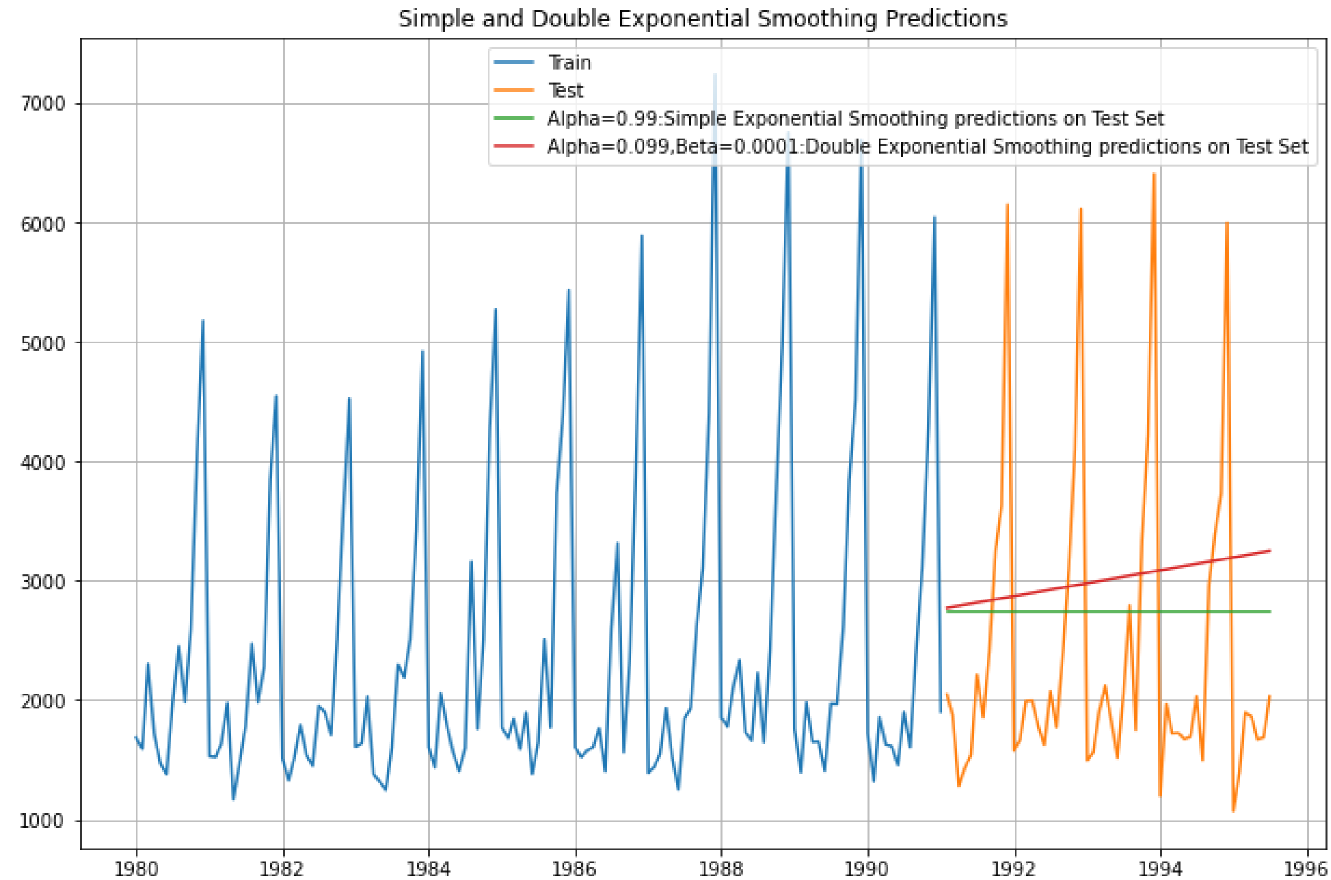
3. Train test split has been done satisfying the necessary requirements.

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

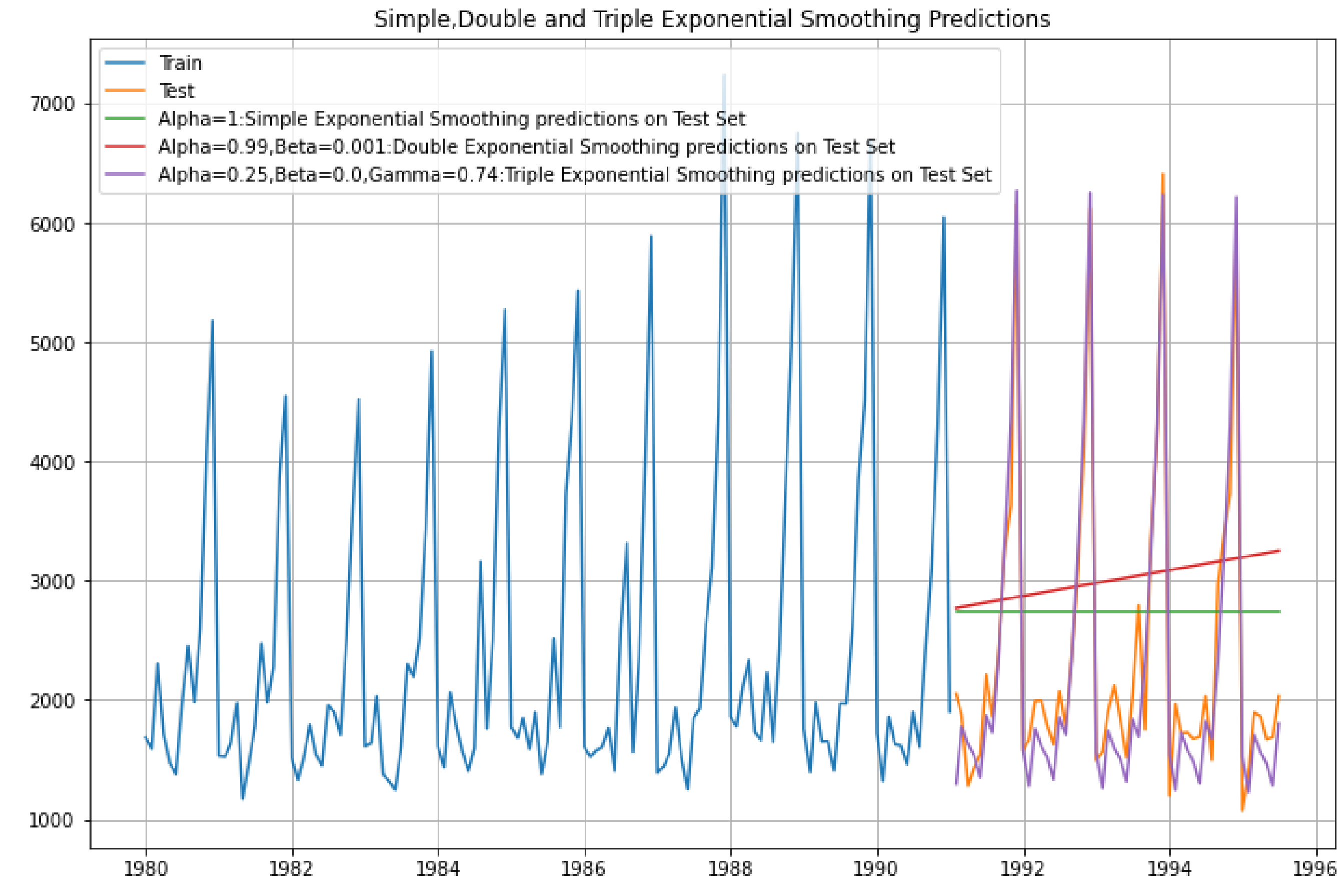
Simple Exponential Smoothing with additive errors



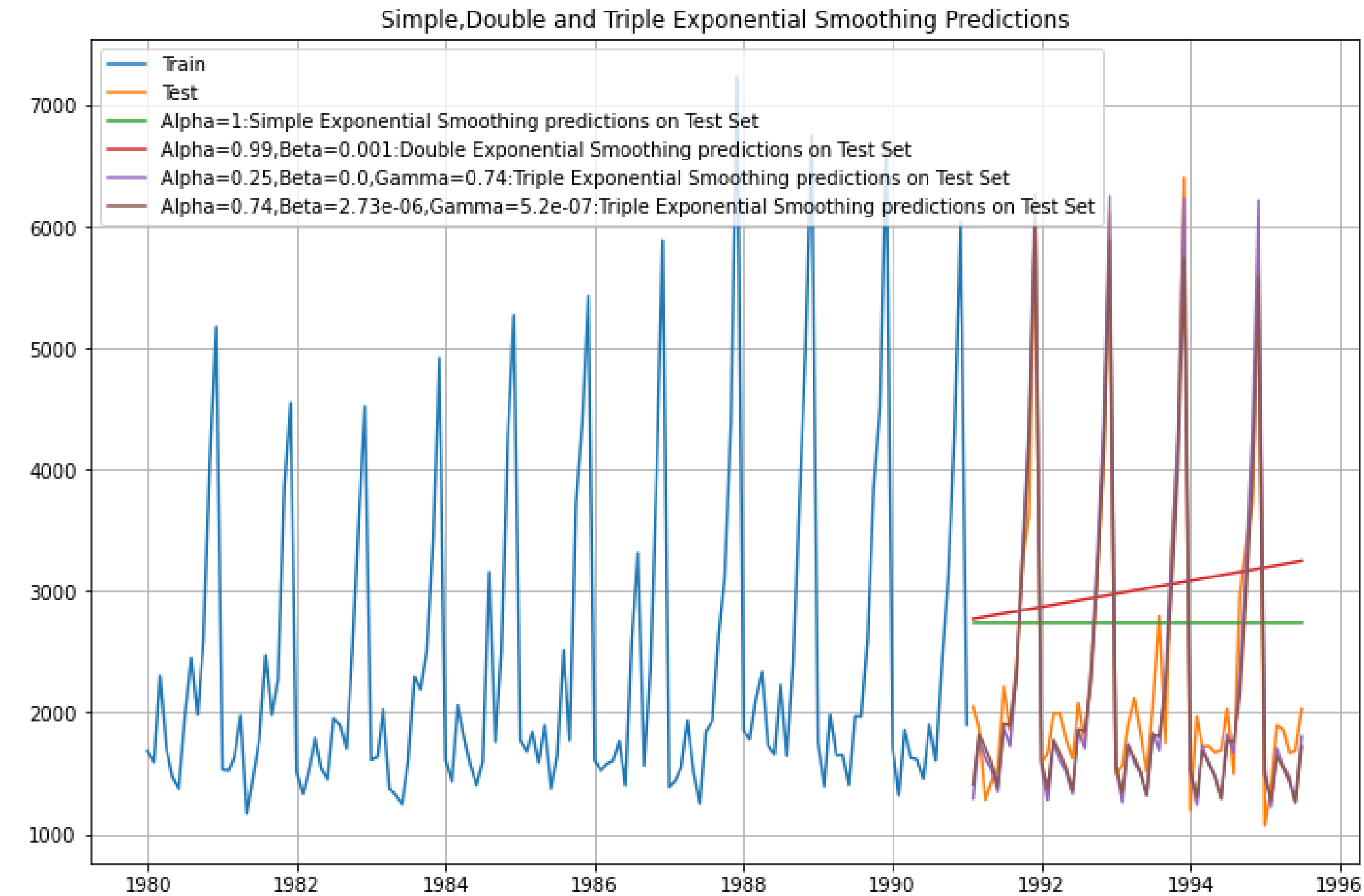
Holt's linear method with additive errors - Double Exponential Smoothing



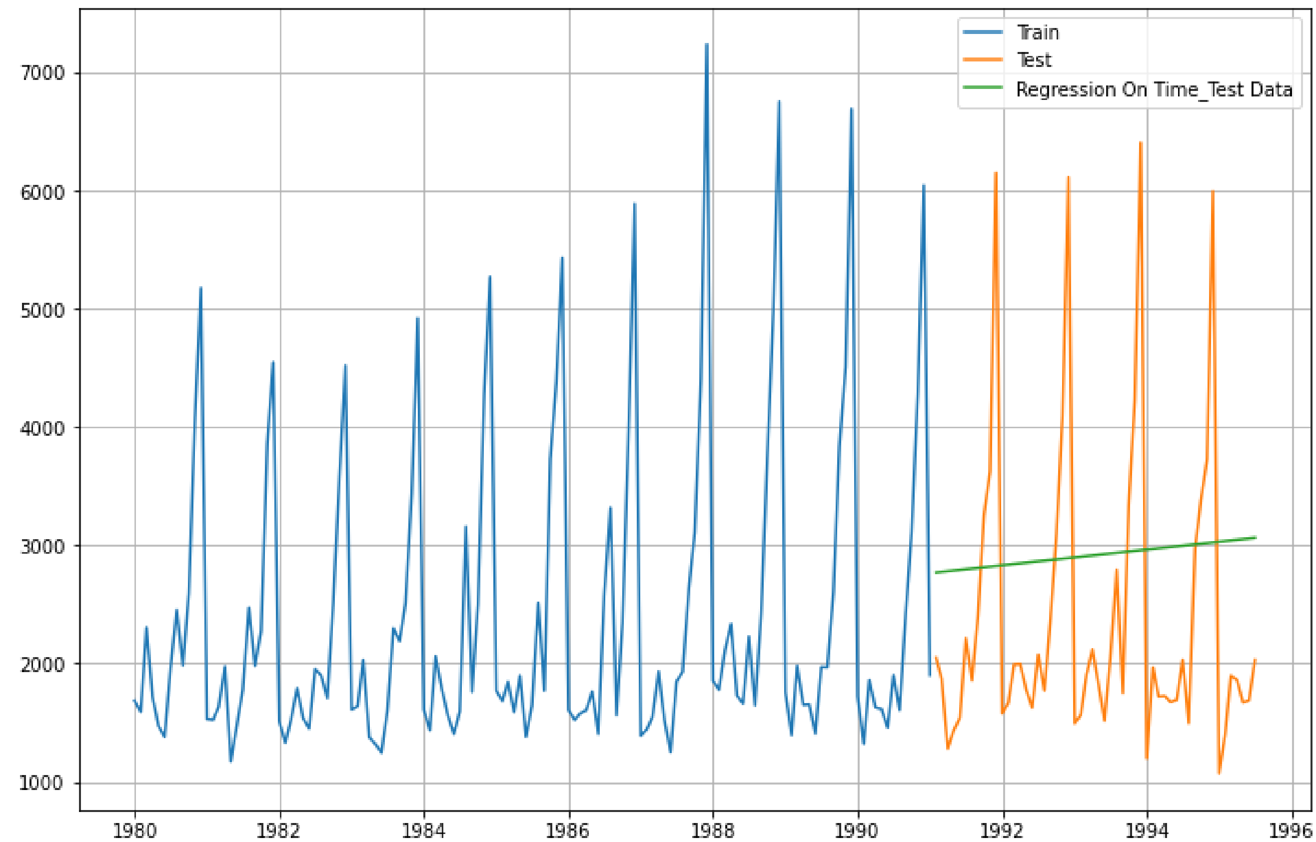
Holt Winter's linear method with additive errors



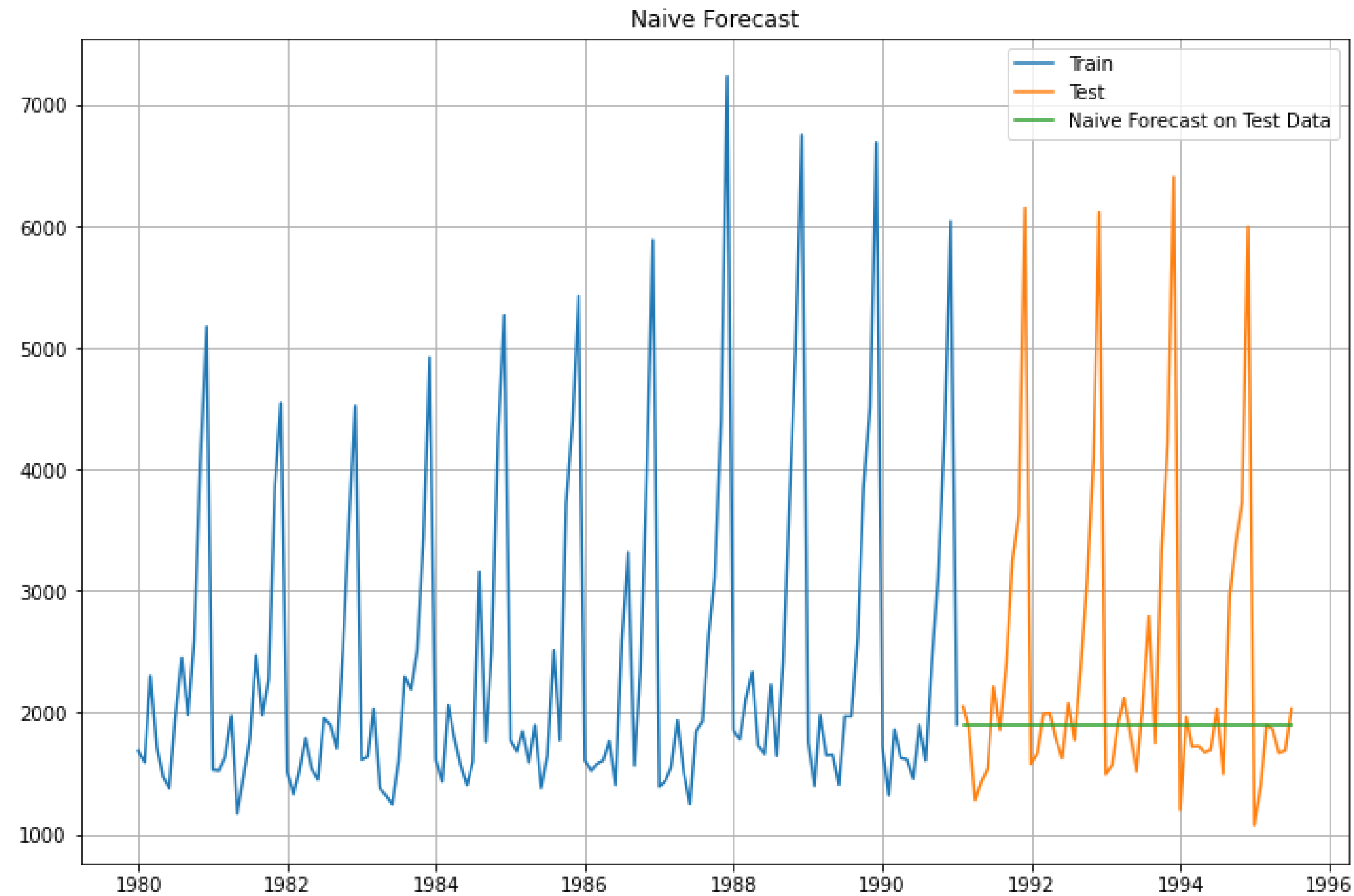
Holt Winter's linear method - ETS (A, A, M) model



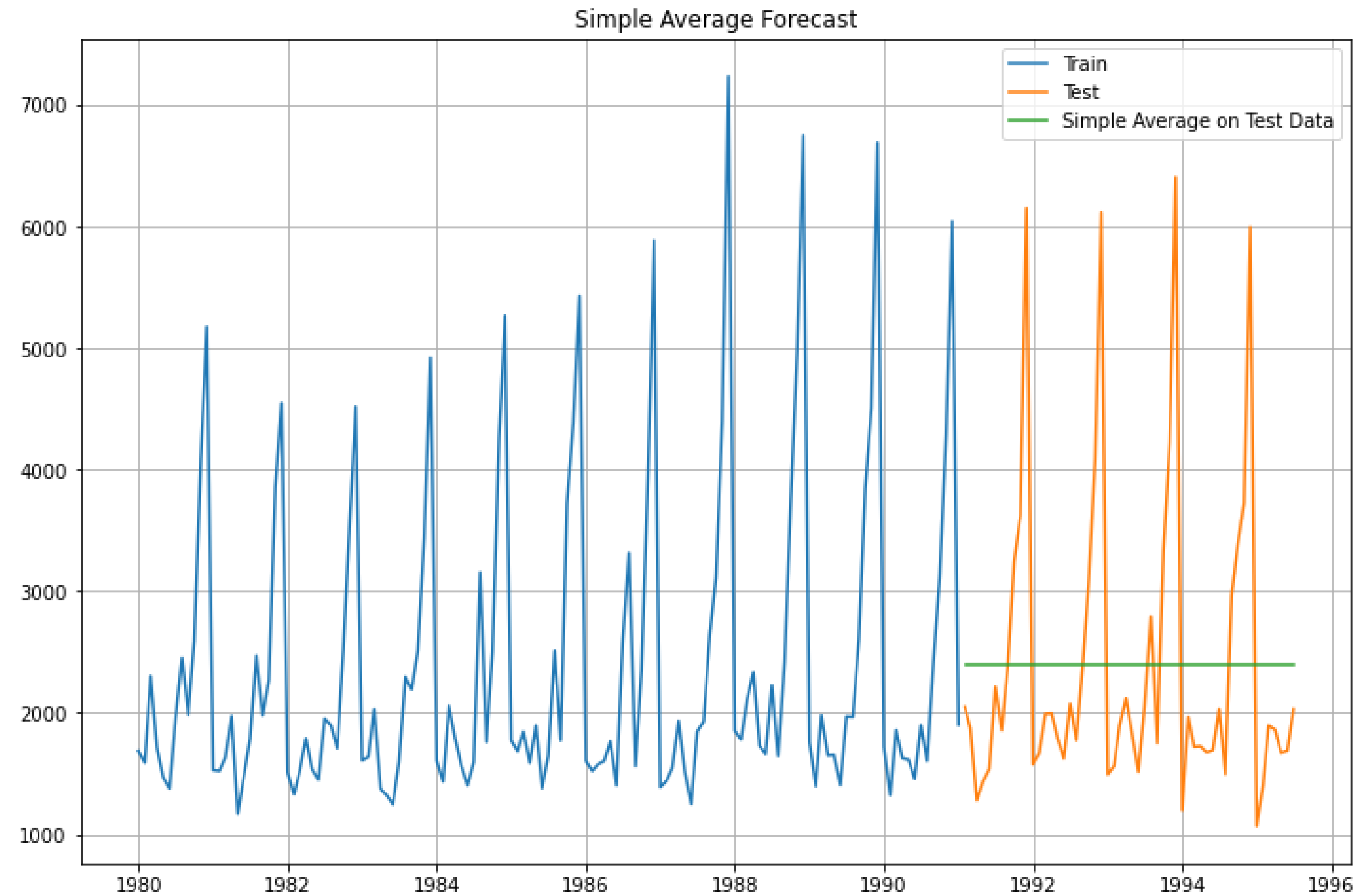
Linear Regression



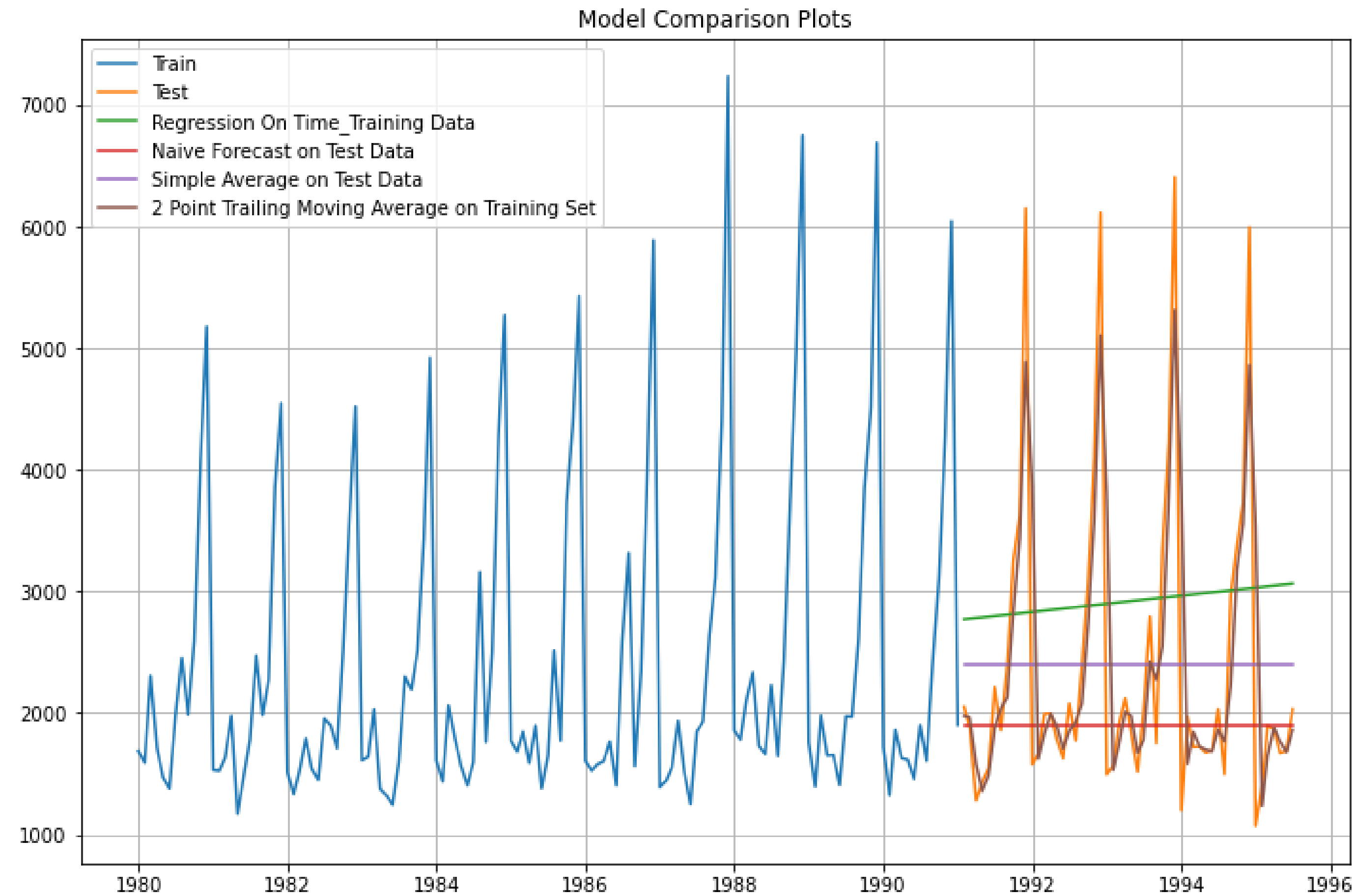
Naive Approach



Simple Average



Moving Average (MA)



RMSE for the models built:

	Test RMSE
Alpha=0.99, SES	1325.947872
Alpha=1, Beta=0.0189: DES	1423.848469
Alpha=0.25, Beta=0.0, Gamma=0.74: TES	351.577674
Alpha=0.74, Beta=2.73e-06, Gamma=5.2e-07, Gamma=0: TES	348.655594
RegressionOnTime	1383.347145
NaiveModel	1381.177135
SimpleAverageModel	1285.039964
2pointTrailingMovingAverage	770.928742
4pointTrailingMovingAverage	1137.137053

6pointTrailingMovingAverage	1283.096993
9pointTrailingMovingAverage	1354.277938

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.

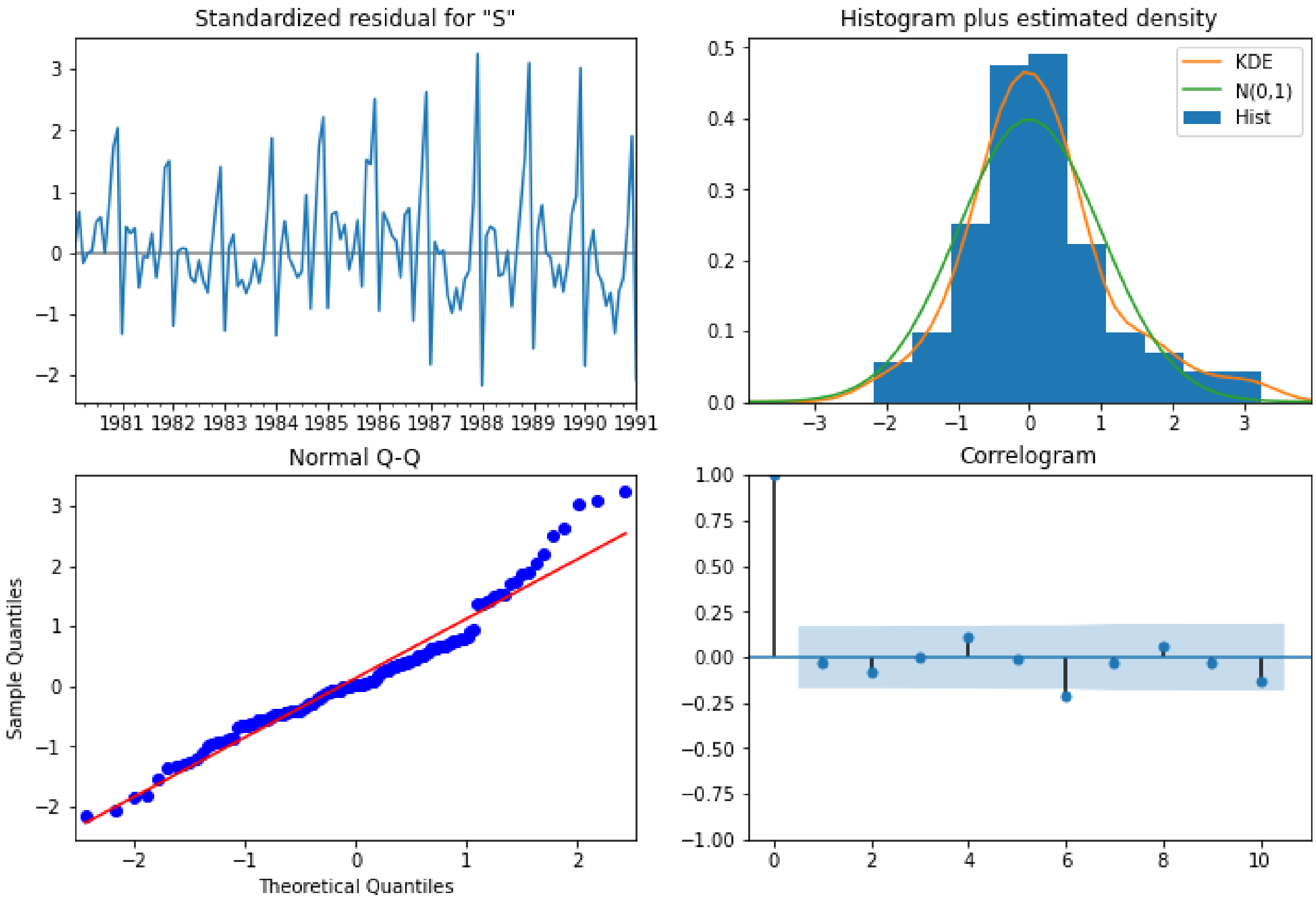
The hypothesis in a simple form for the ADF test is:

- *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.

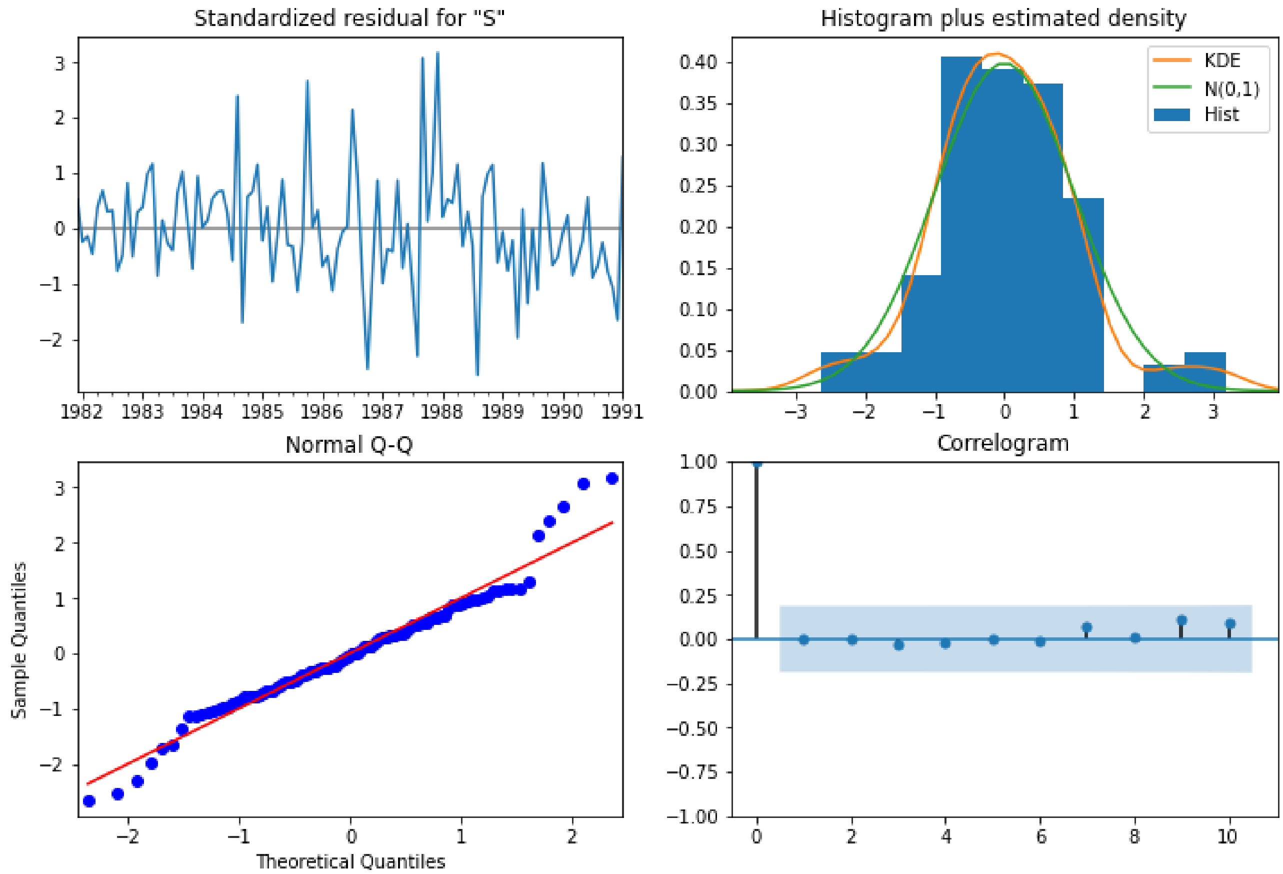
We would want the series to be stationary for building ARIMA models and thus we would want the p-value of this test to be less than the α value.

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.



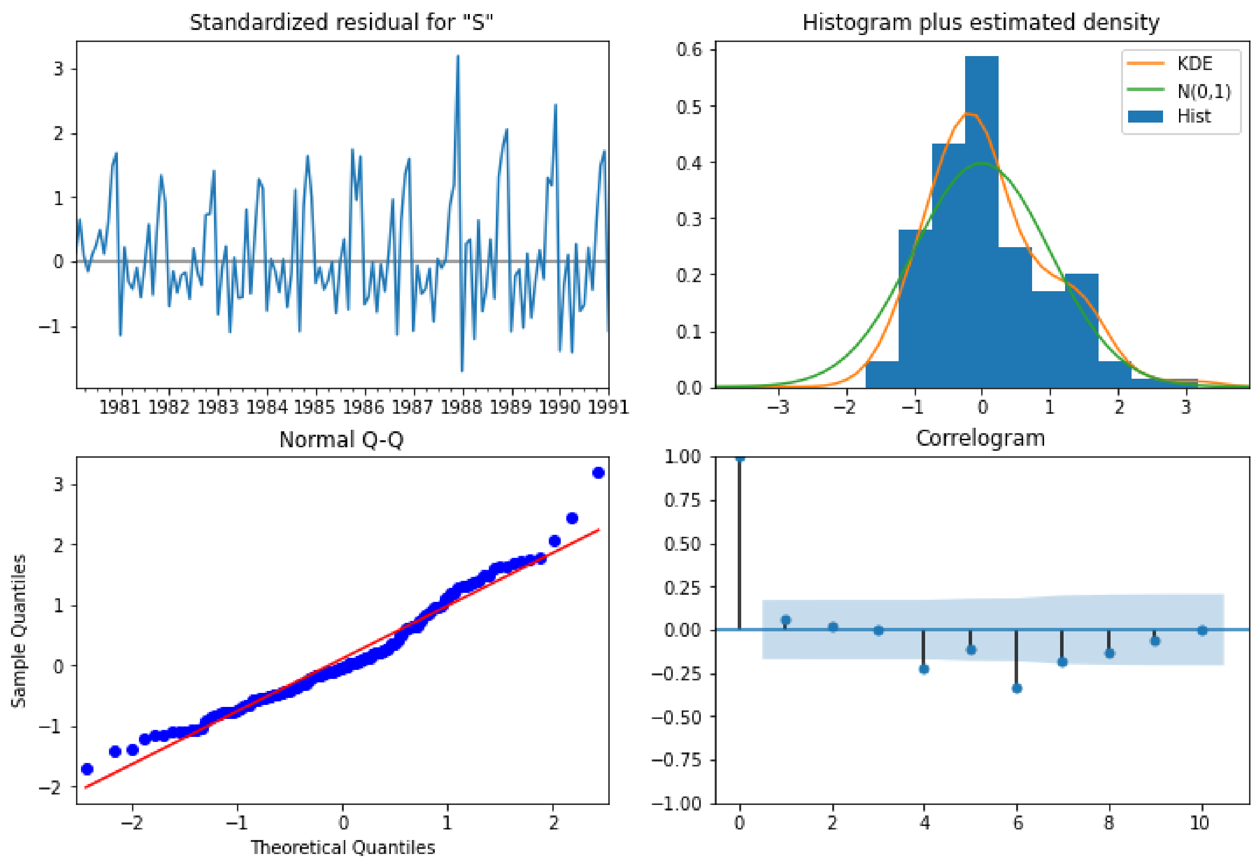
Automated ARIMA diagnostics



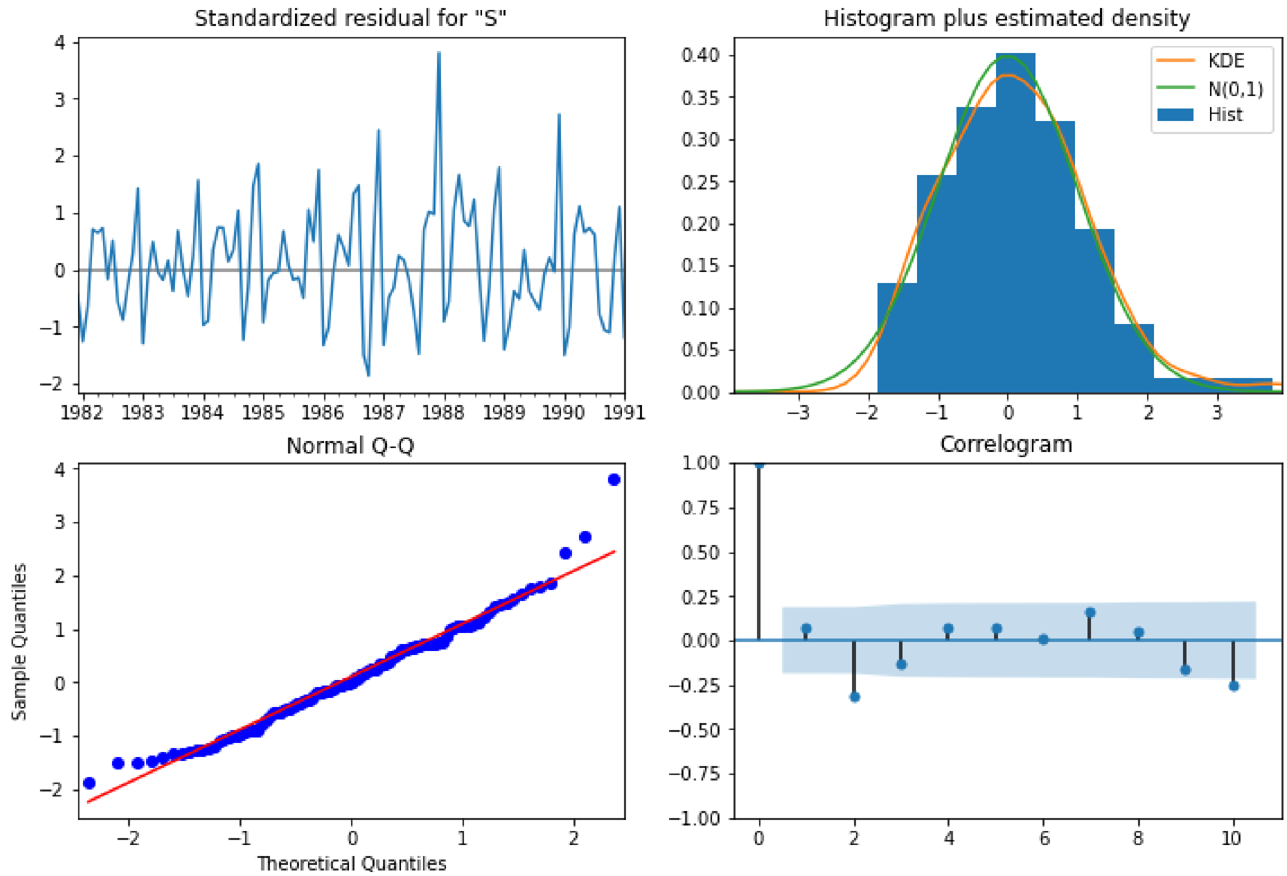
Automated SARIMA diagnostics

	RMSE
ARIMA (2,1,2)	1298.678641
SARIMA (2,1,3) (2,0,3,6)	652.178782

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 (2,1,3)	1295.659357
SARIMA (2,1,3) (0,0,3,6)	1080.007356

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

	Test RMSE
Alpha=0.99, SES	1325.947872

Alpha=1, Beta=0.0189: <u>DES</u>	1423.848469
Alpha=0.25, Beta=0.0, Gamma=0.74: <u>TES</u>	351.577674
Alpha=0.74, Beta=2.73e-06, Gamma=5.2e-07, Gamma=0: <u>TES</u>	348.655594
RegressionOnTime	1383.347145
NaiveModel	1381.177135
SimpleAverageModel	1285.039964
2pointTrailingMovingAverage	770.928742
4pointTrailingMovingAverage	1137.137053
6pointTrailingMovingAverage	1283.096993
9pointTrailingMovingAverage	1354.277938
ARIMA (2,1,2) *<u>(Automated)</u>*	1298.678641
SARIMA (2,1,3) (2,0,3,6) *<u>(Automated)</u>*	652.178782
ARIMA (2,1,3) *<u>(Manual)</u>*	1295.659357
SARIMA (2,1,3) (0,0,3,6) *<u>(Manual)</u>*	1080.007356

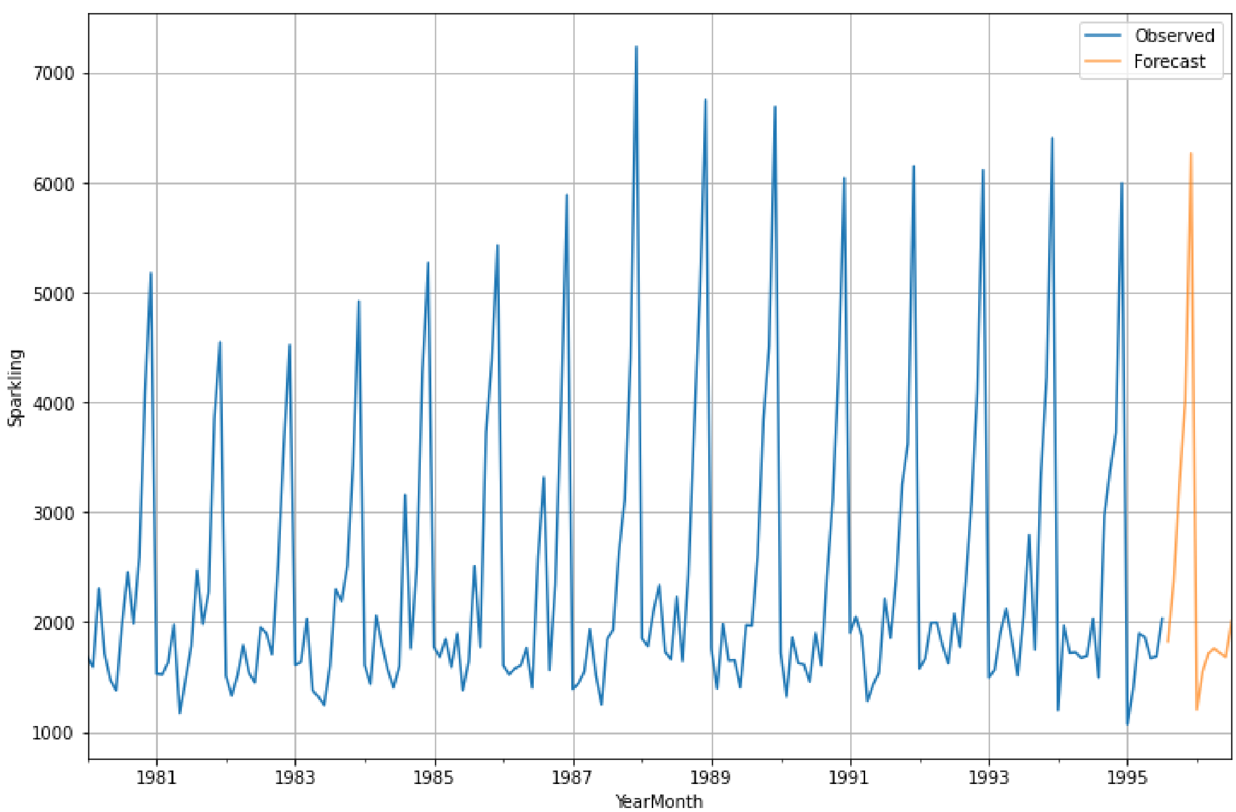
RMSE is a way to measure the error of a model predicting quantitative data. Since it is an error value, the model with the least RMSE value will give the most accurate predictions.

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 forecasted.

RMSE of the total data is 564.27

10. Various models have been built on the train data, performance has been tested on the test data using RMSE scores and the best one 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 generally, the sales have been at its peak during the months of November and December which makes sense since it’s the holiday season.

However, when the sales of the holiday season in the 1980s is compared to the recent year (1990-1995), it can be seen that there’s a massive uprise in sale during the recent times.

But there is no significant difference in sales in the other months of the year.

Hence the company must focus on being all equipped and ready to meet the demand during the holiday season than during the other months.

Manufacturing and having just the right amount of stock in the early months is sufficient while the company must be well prepared to meet even excess demand during its holiday season to seize the opportunity to sell more and make more profits than ever before.