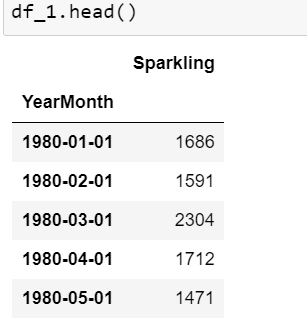
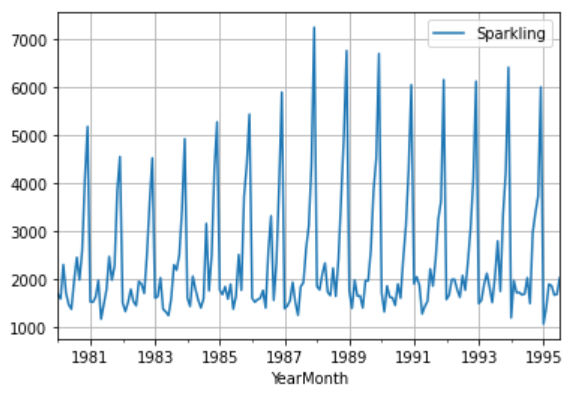
**Business Report**

**Time Series Forecasting**

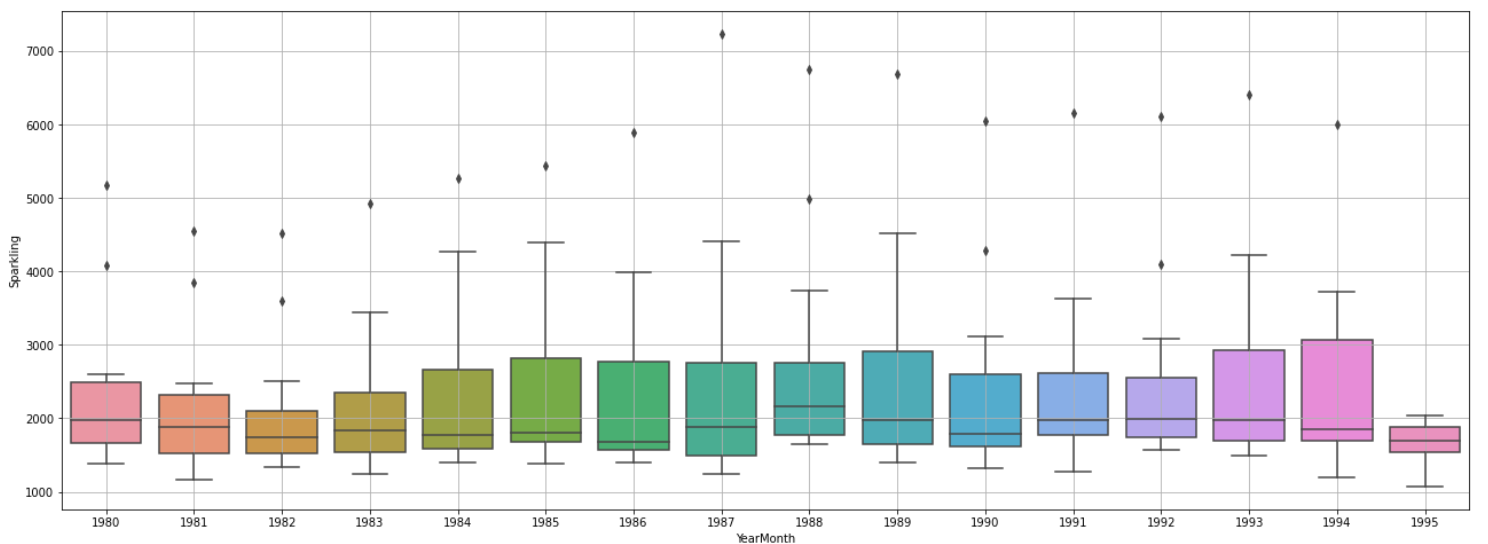
Dataset 1: Sparkling.csv

1. Read the data as an appropriate Time Series data and plot the data.



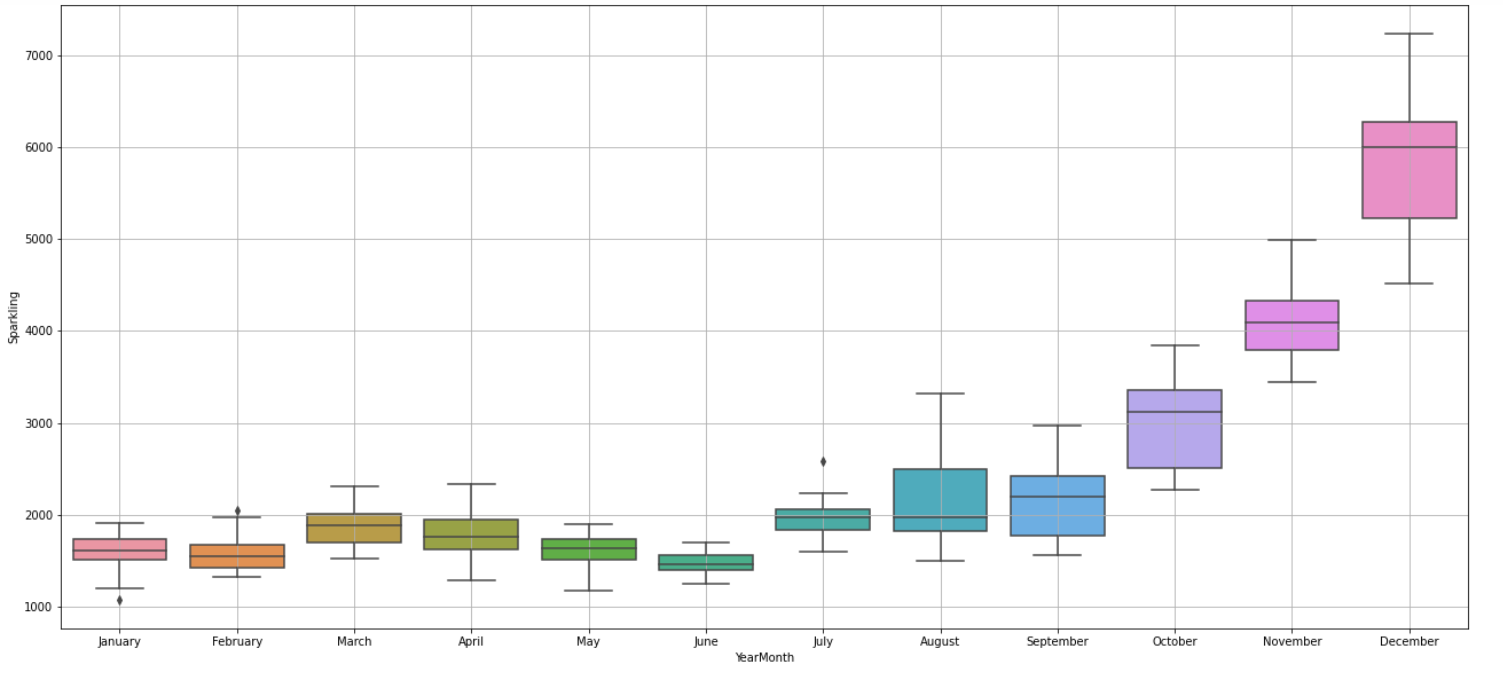


1. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.



The data shows trend.

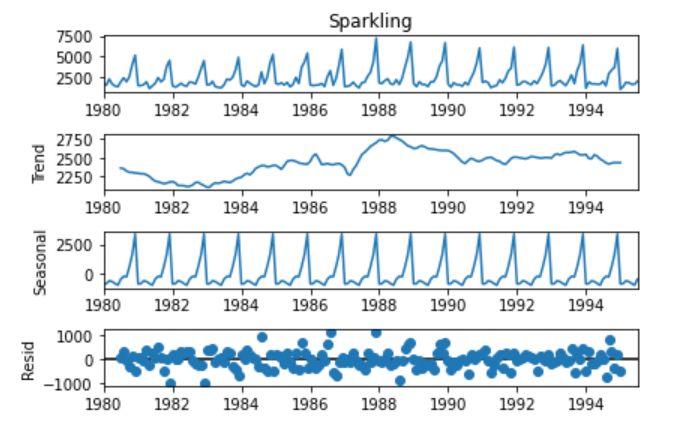
Also there are outliers in most of the years.



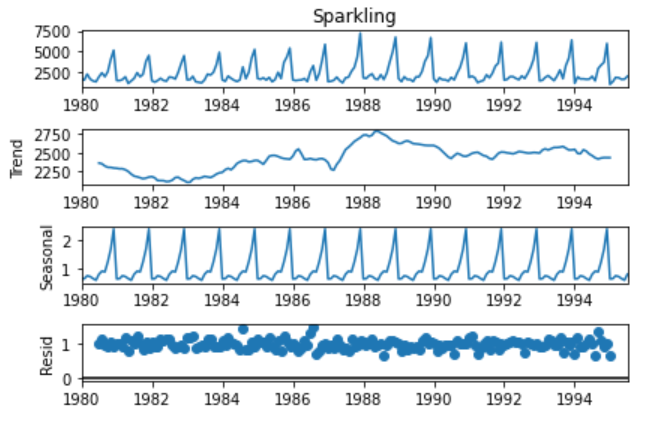
The data also shows Seasonality.

The wines sales are more in the month of October, November and December(Highest).

Additive Decomposition

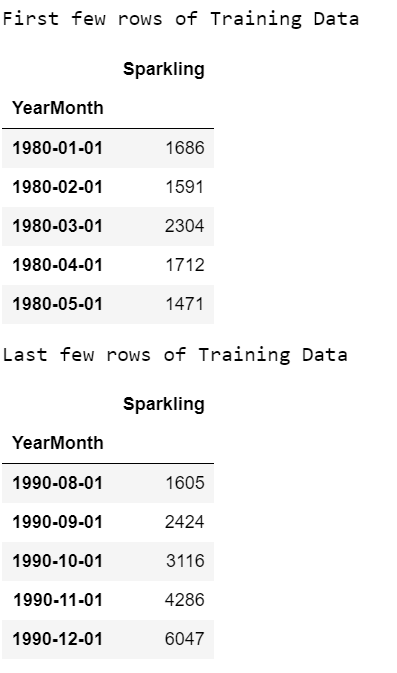


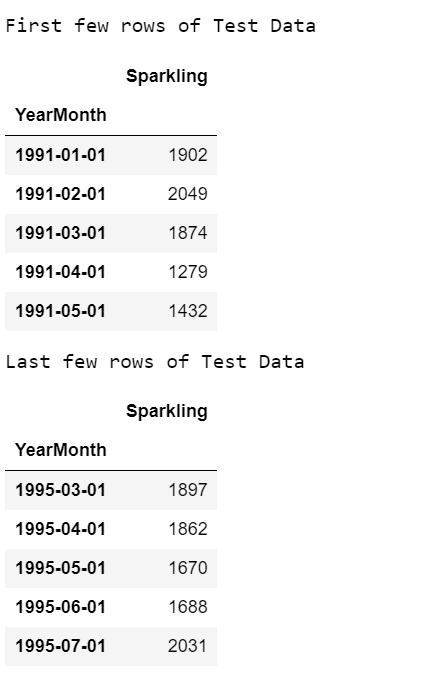
Multiplicative Decomposition



The data has multiplicative seasonality.

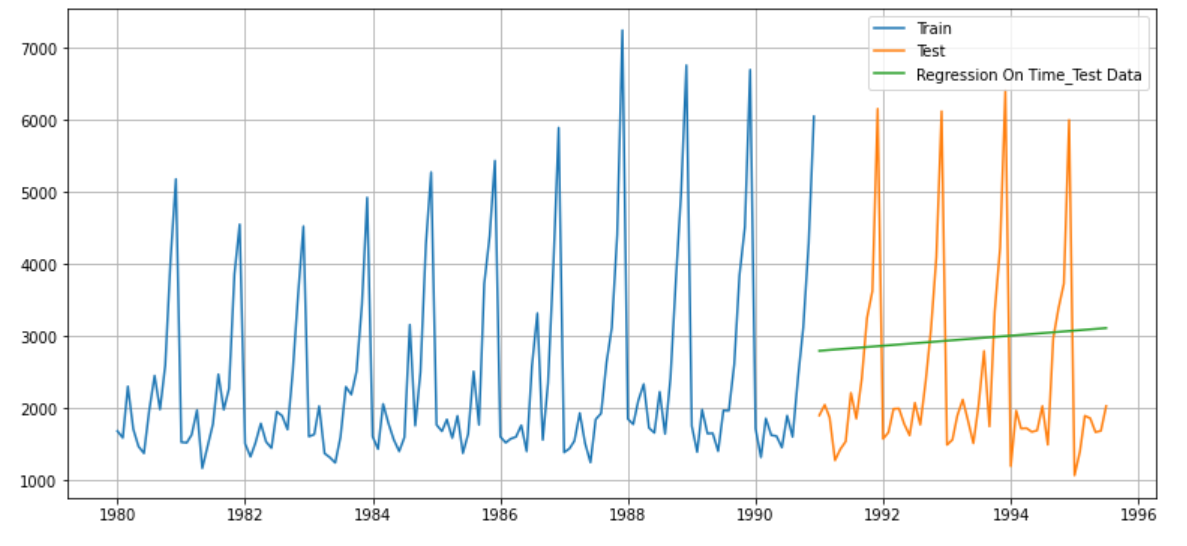
1. Split the data into training and test. The test data should start in 1991.





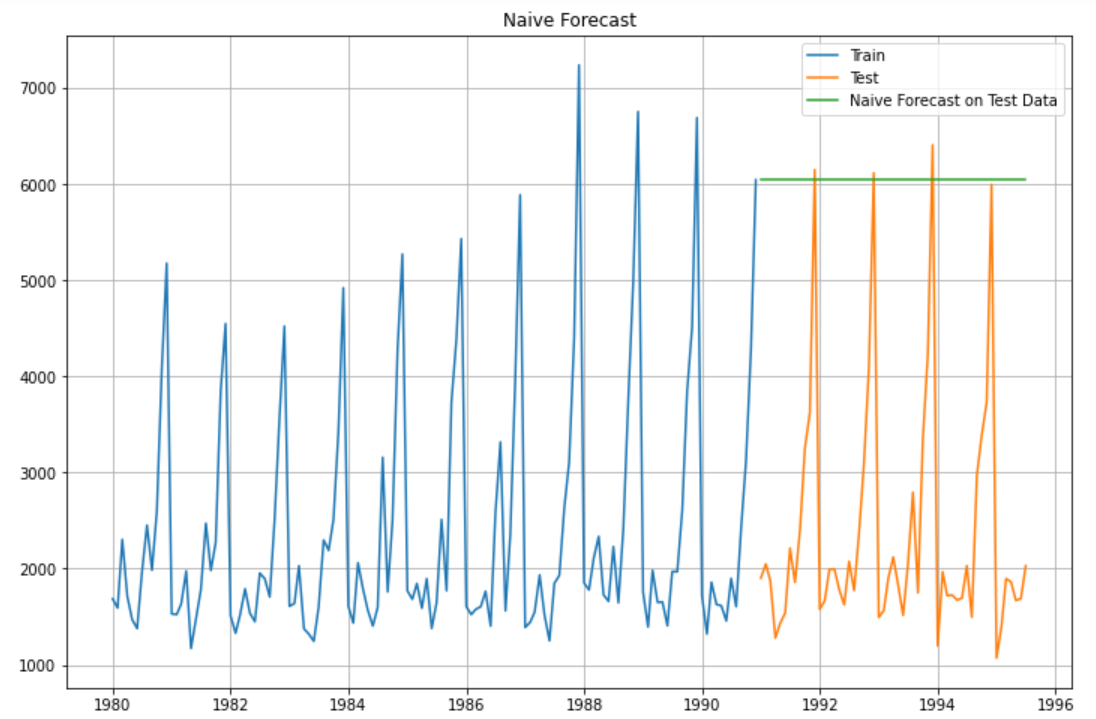
1. Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data.  
   Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

Graph by Linear regression



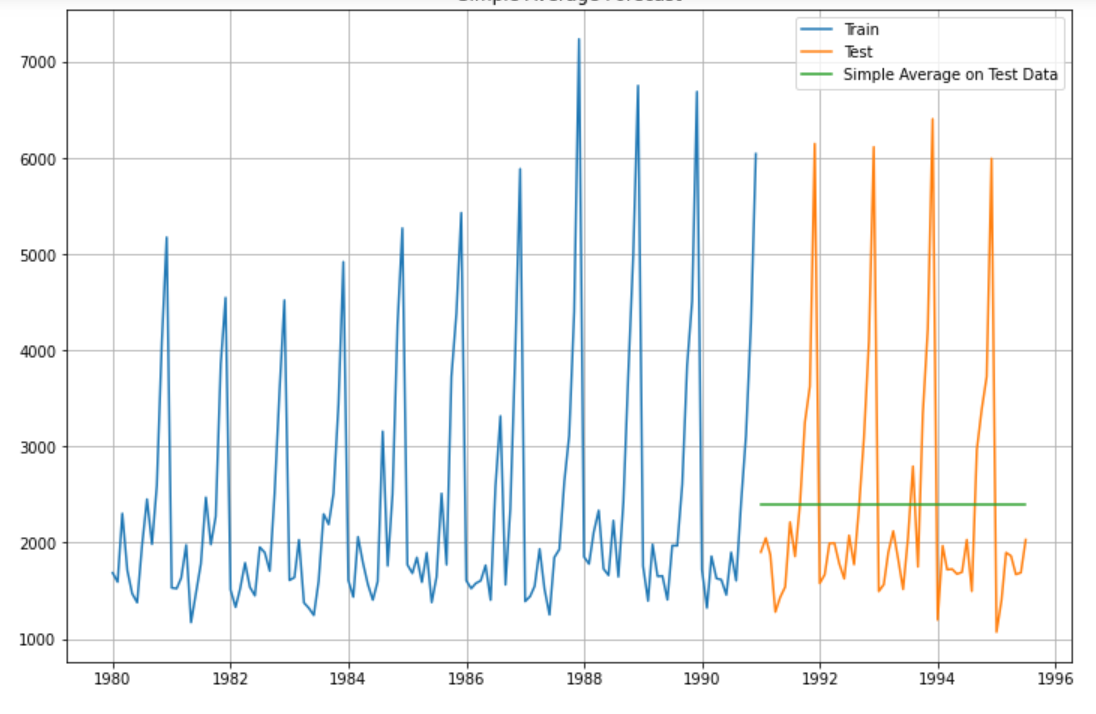
Linear regression fails trend and seasonality. It is a straight line

Naïve Approach



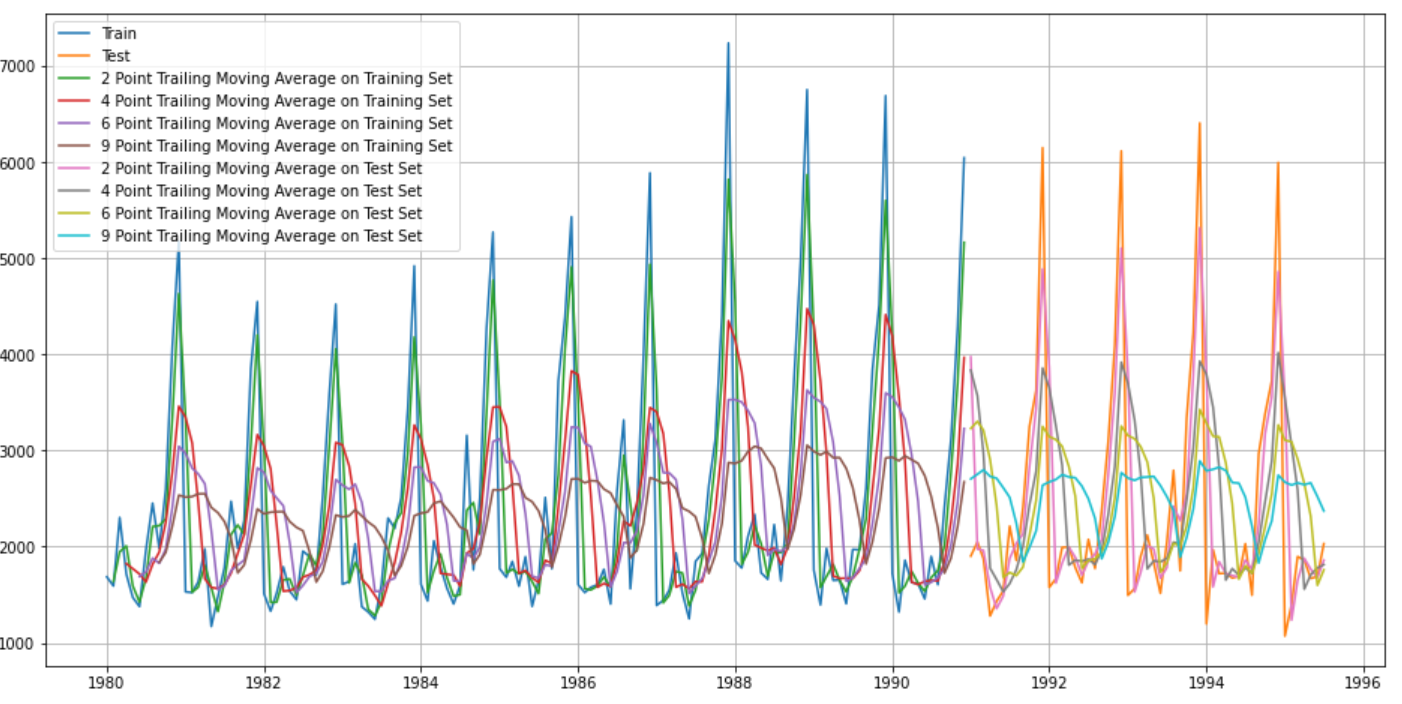
Naïve approach fails trend and seasonality. It is a straight line

Simple Average



Simple average also fails trend and seasonality. It is a straight line

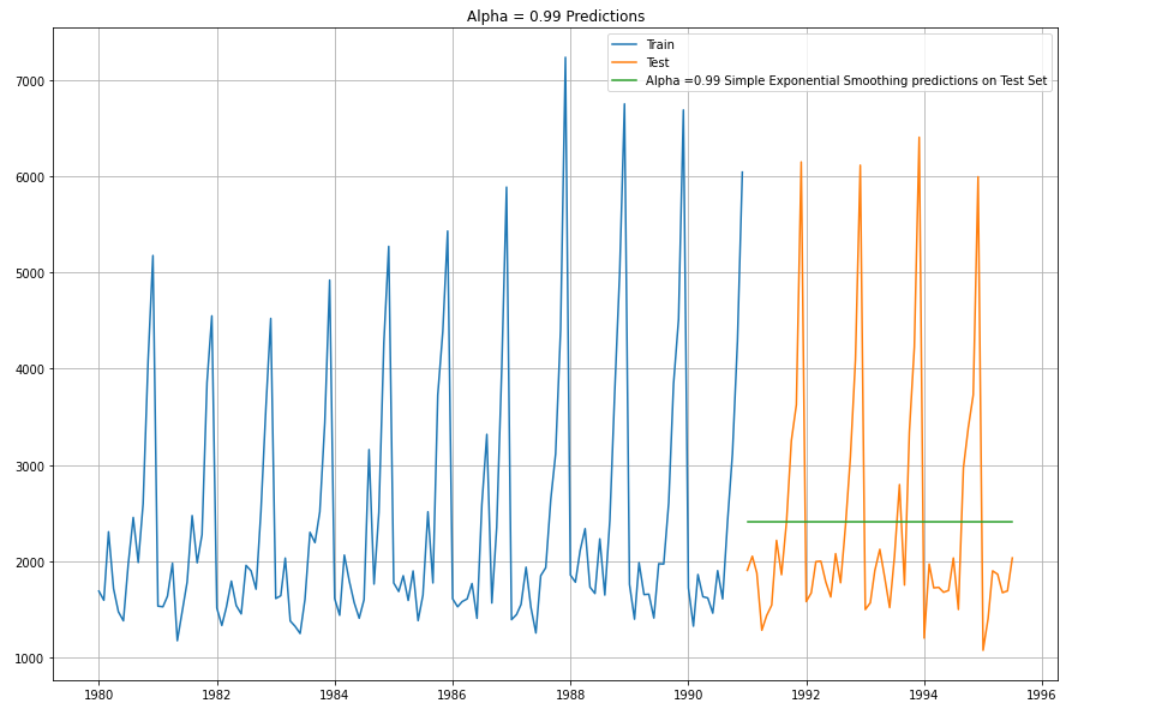
Moving Average



2 point moving average has matched trend and seasonality of test data to some extent.

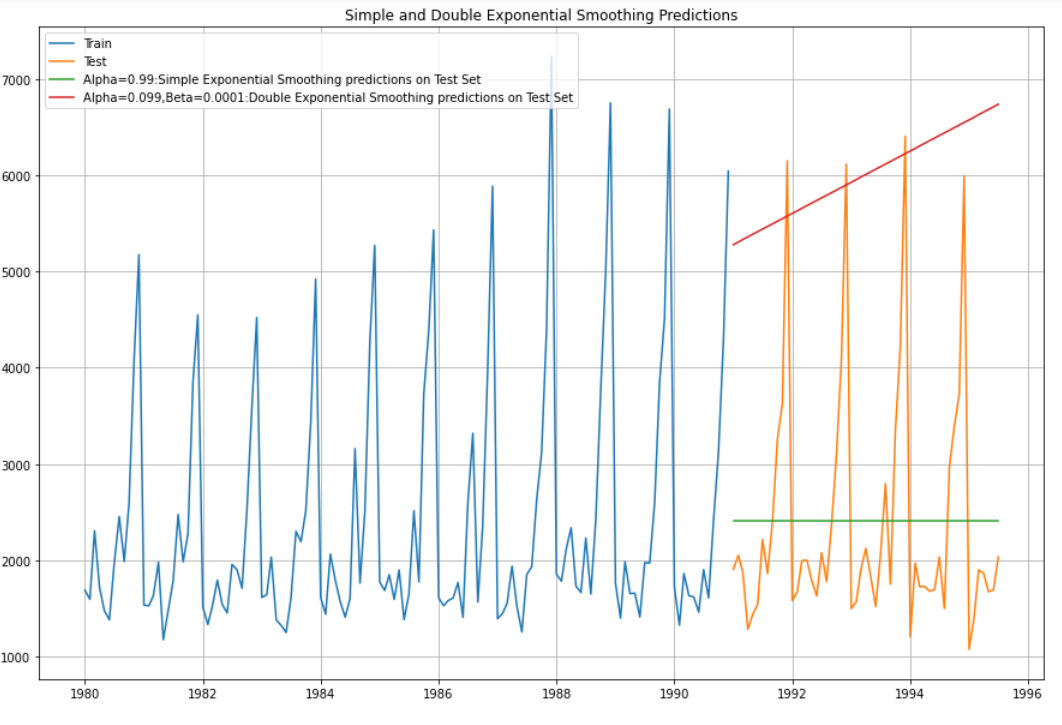
Simple Exponential Smoothing

SES also fails to capture trend and seasonality.



Double Exponential Smoothing

Holts method also fails.

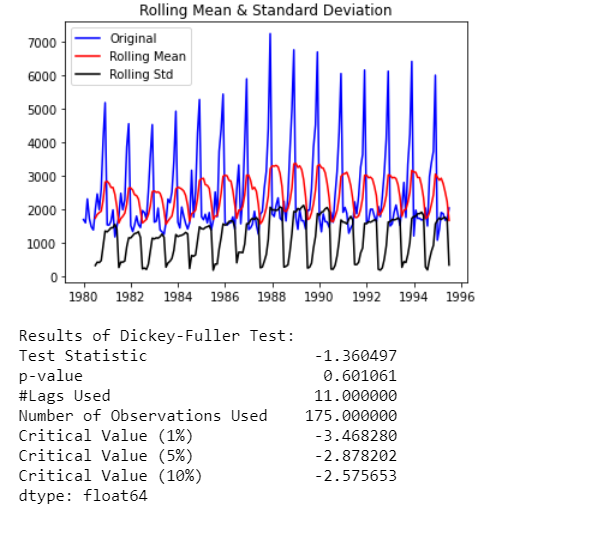


Triple Exponential Smoothing

Holts-Winter method captures both trend and seasonlity.

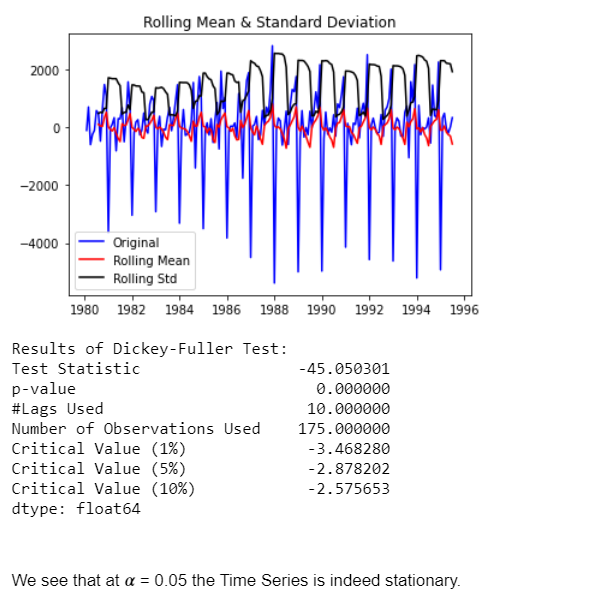


1. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment.  
   Note: Stationarity should be checked at alpha = 0.05.



At 5% significant level the Time Series is non-stationary.

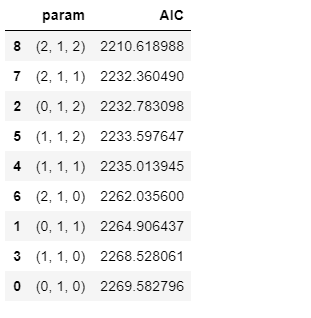
After 1st differencing it is Stationary

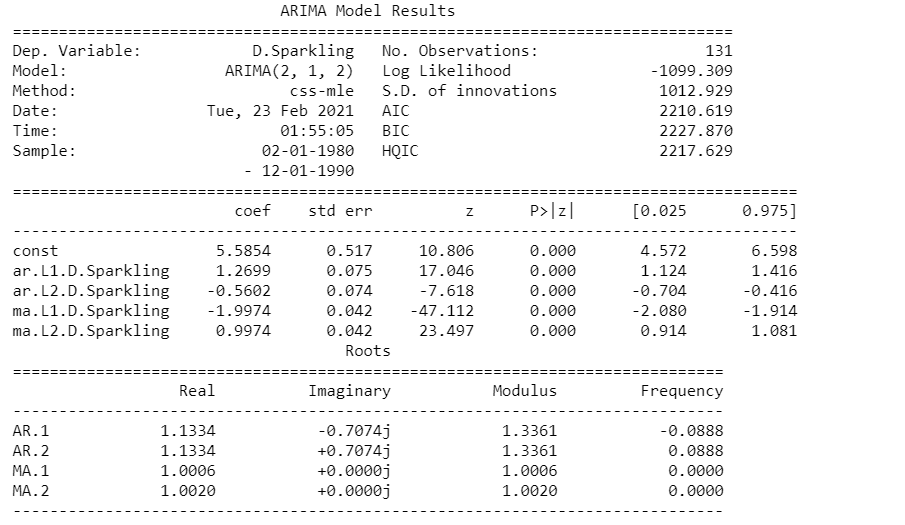


1. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

Arima model

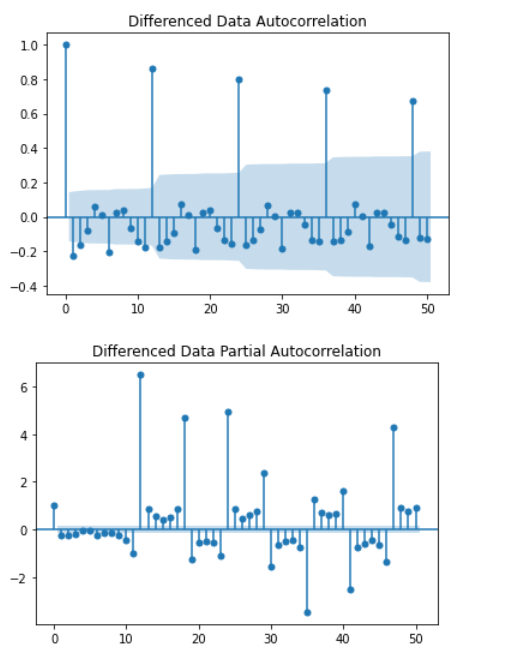
Lowest AIC





1. Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.

ARIMA model based on the cut-off points of ACF and PACF

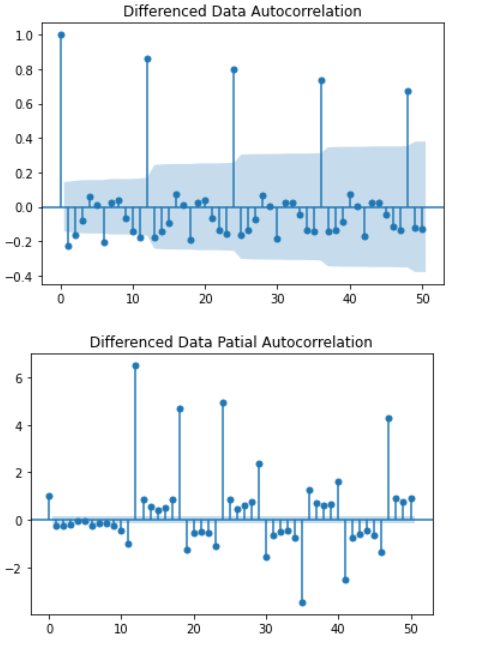


The Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 0.

The Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 0.

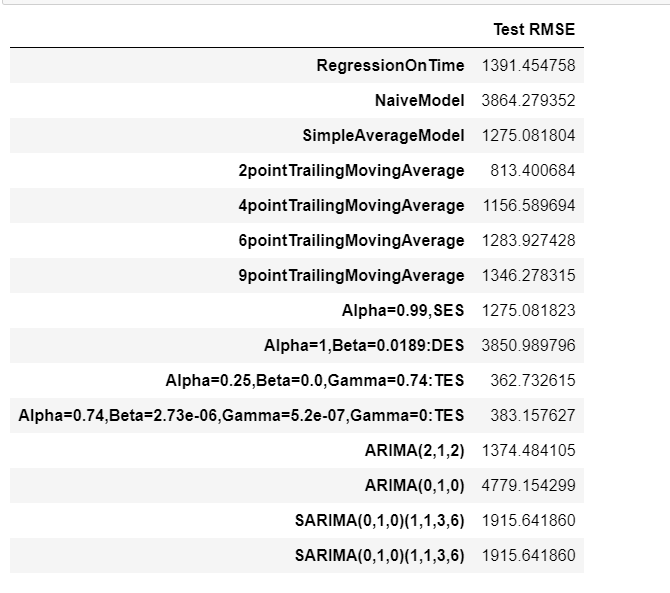
By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.

SARIMA model based on the cut-off points of ACF and PACF:-



By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.

1. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data.



1. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.



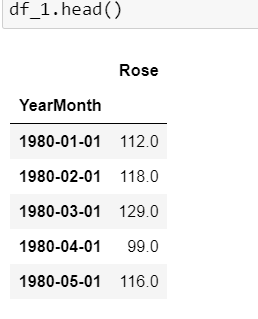
1. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

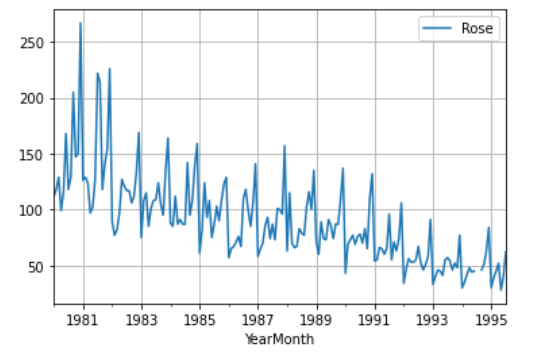
As Sparkling wines sales are more in the months of October , November and December more production of wine, offers to customers should be targeted during this period(Festive season)

Yearly sales of sparkling wine is more or less the same in this dataset.

Dataset 2: Rose.csv

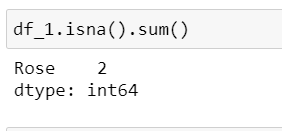
1. Read the data as an appropriate Time Series data and plot the data.



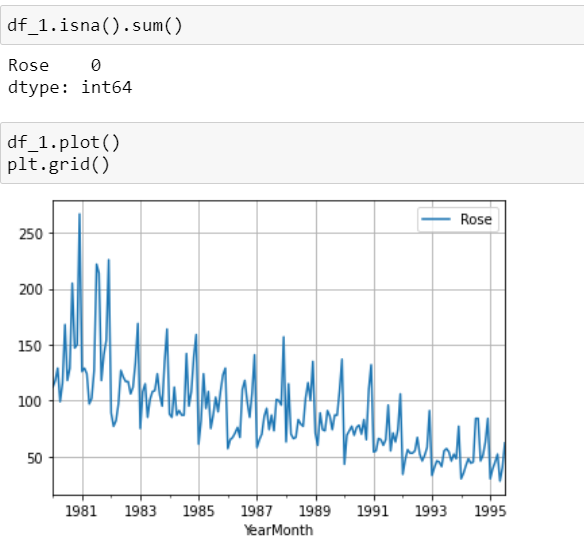


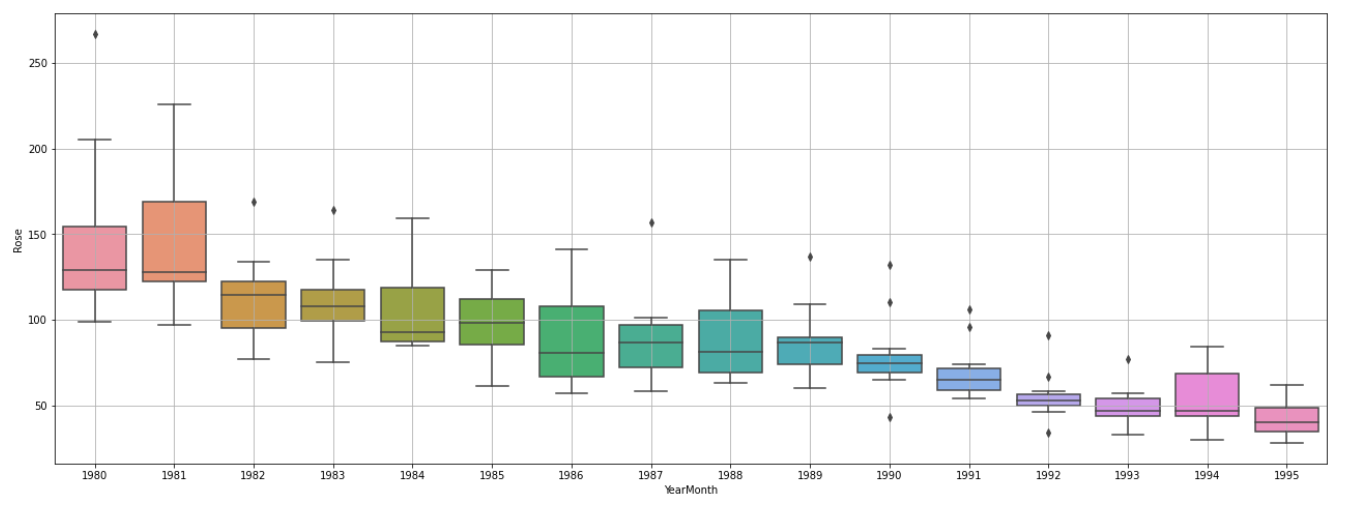
1. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

There were 2 missing values in the time series



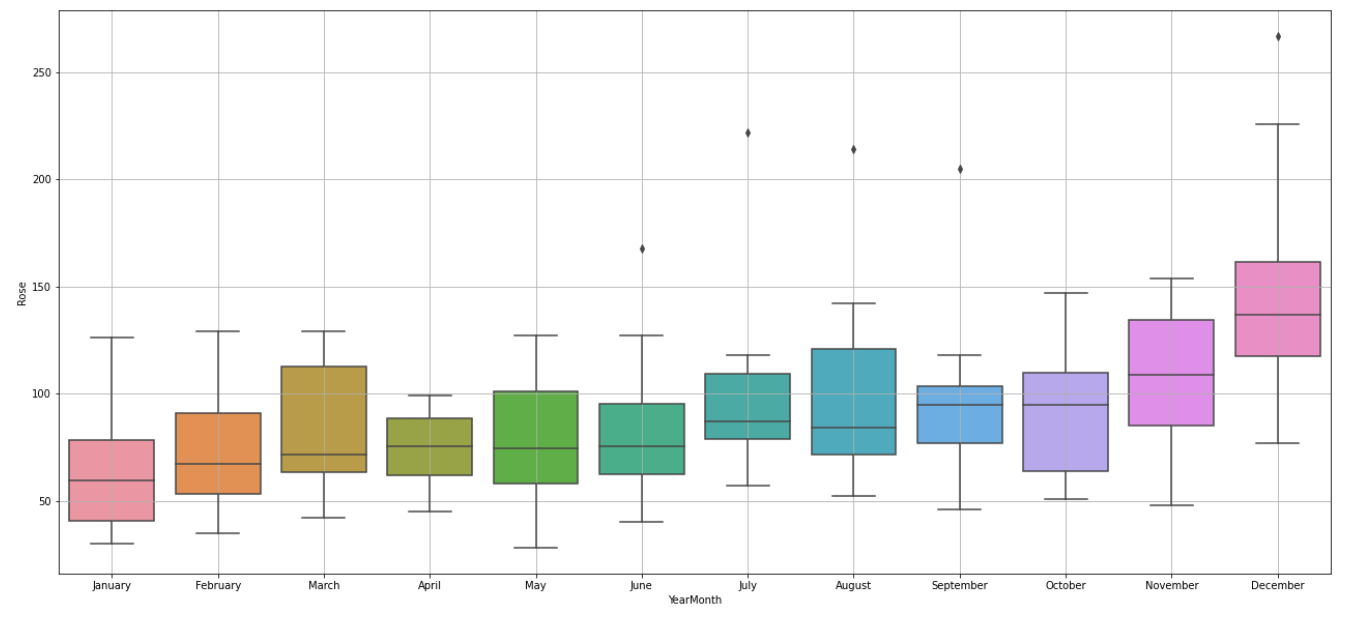
After imputing missing values





The TS shows declining trend

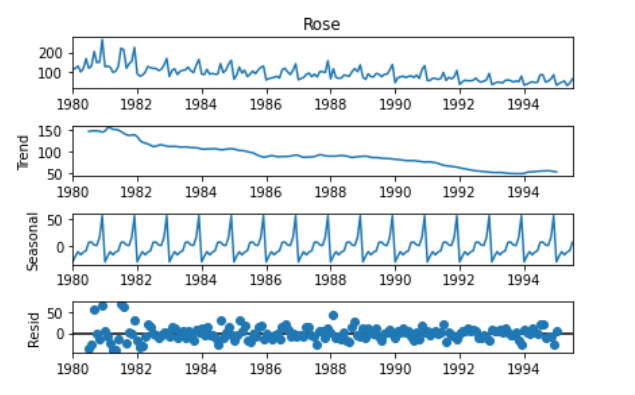
Also there are outliers



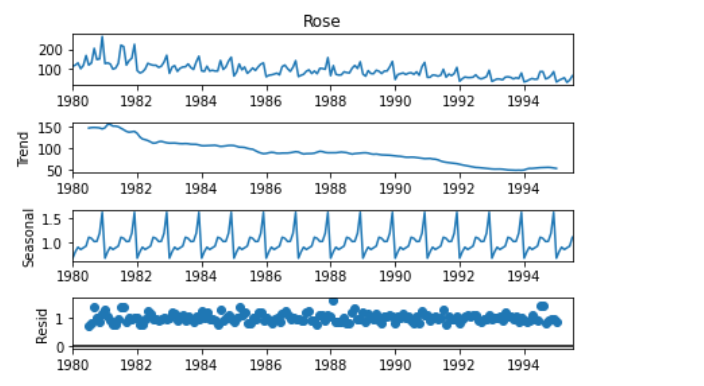
There is seasonality as well in the TS

December shows maximum wine sales

Additive Decomposition

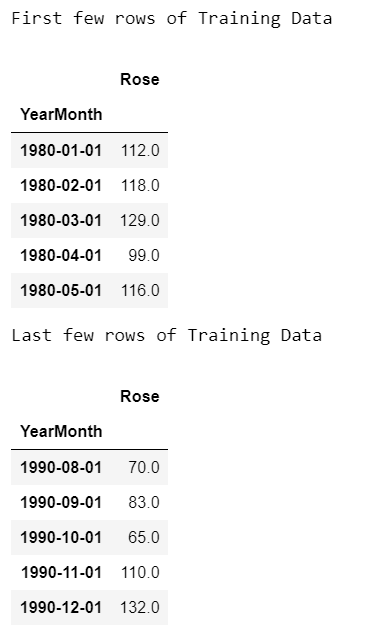


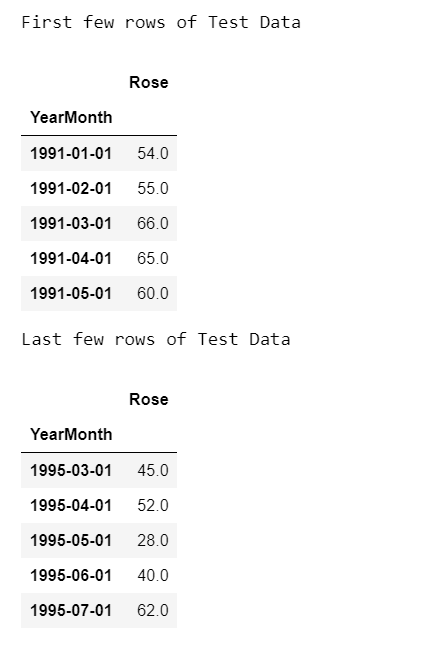
Multiplicative Decomposition



TS shows multiplicative seasonality

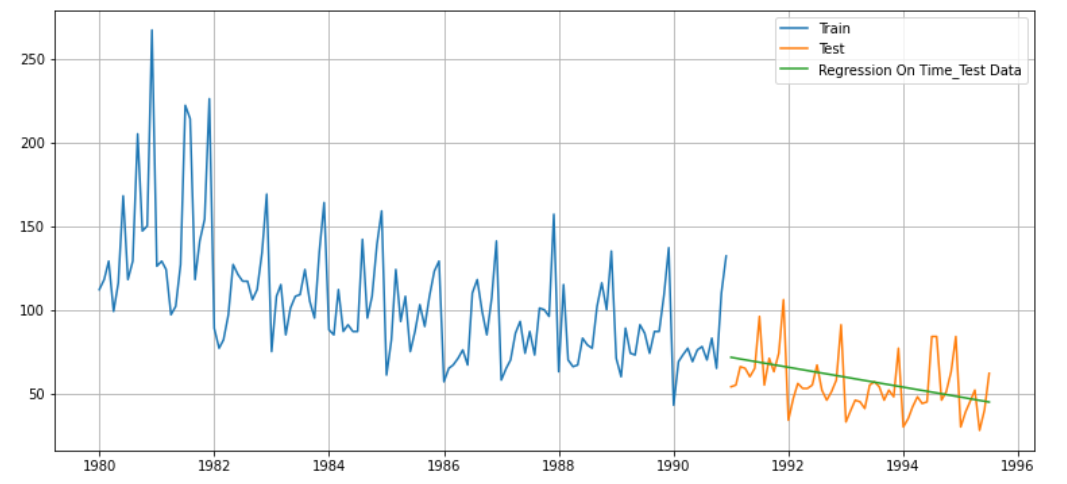
1. Split the data into training and test. The test data should start in 1991.

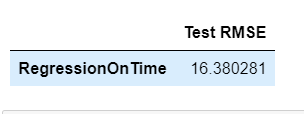




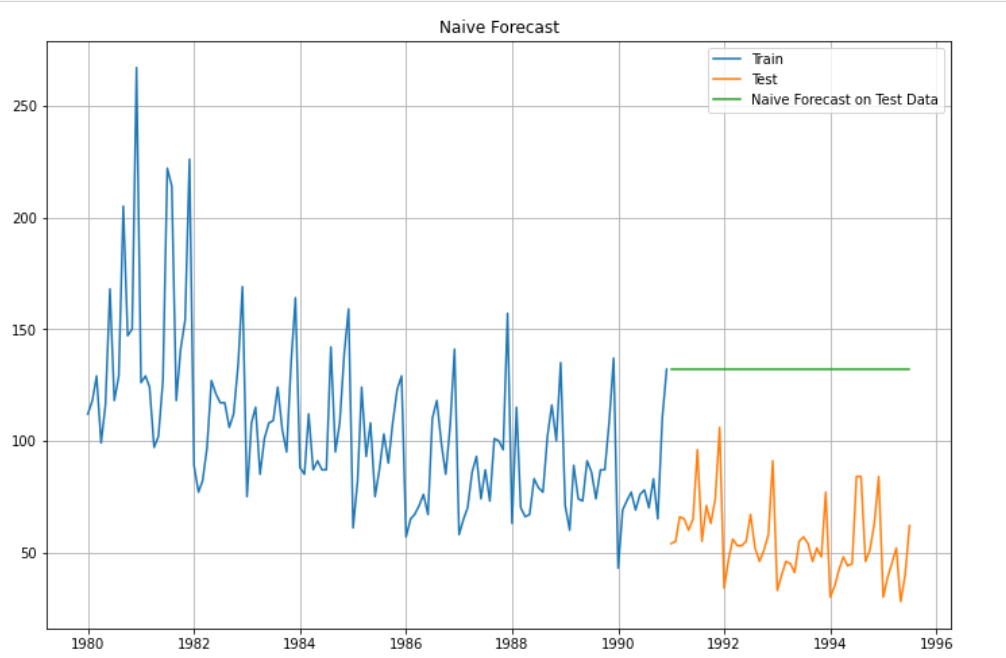
1. Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data.  
   Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

Linear regression capture trend but not seasonality



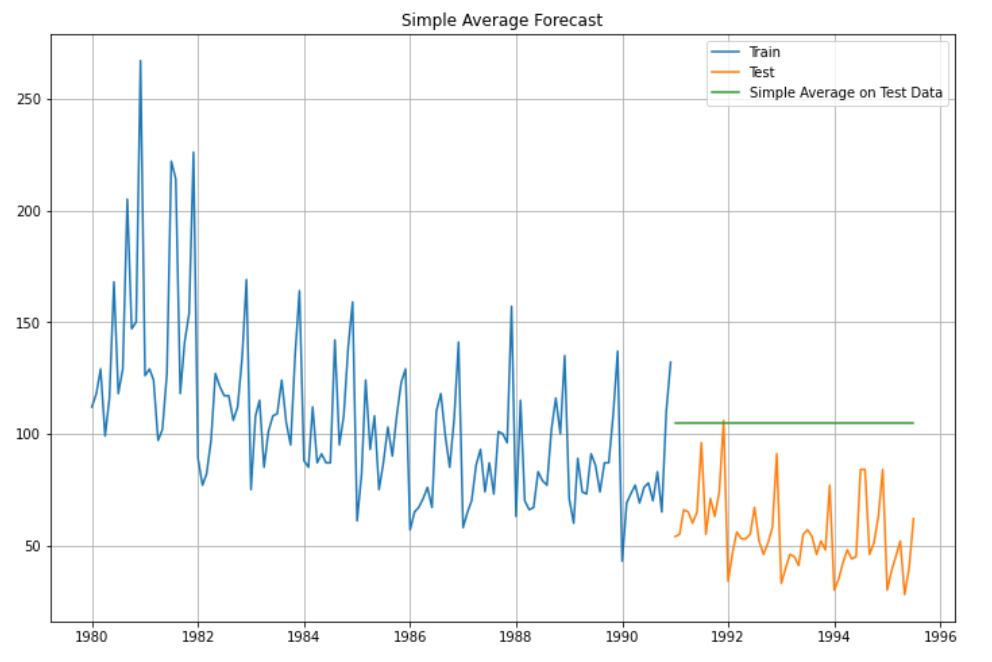


Naïve Approach fails to capture both trend and seasonality





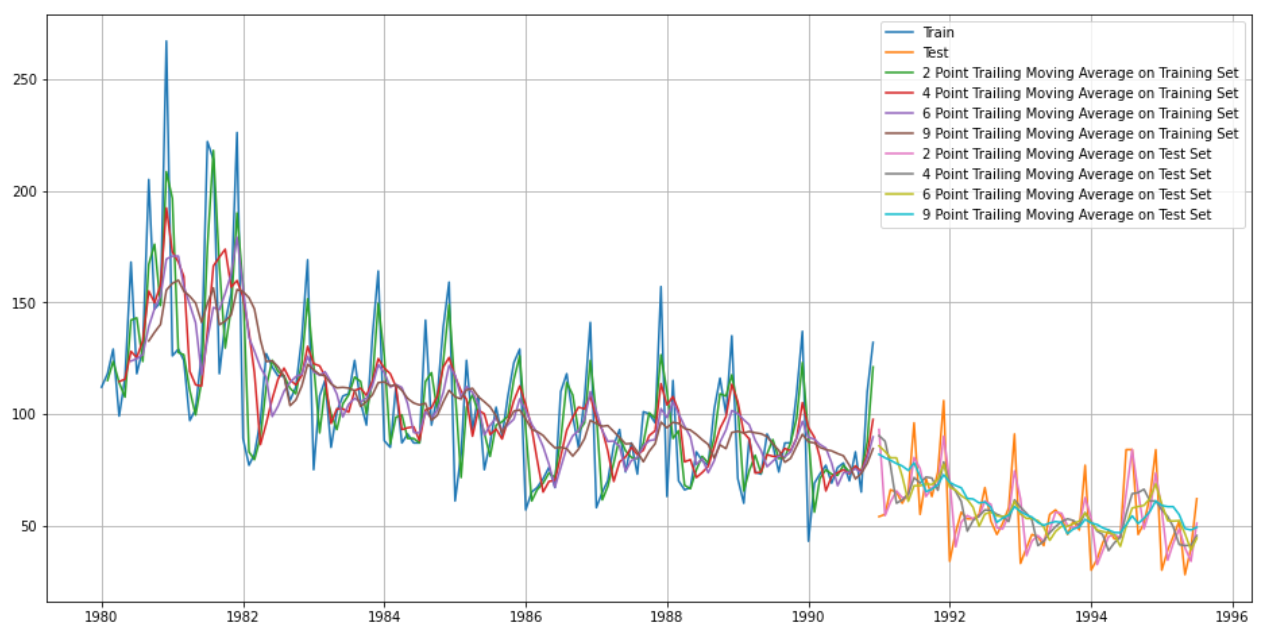
Simple Average fails to capture trend and seasonality





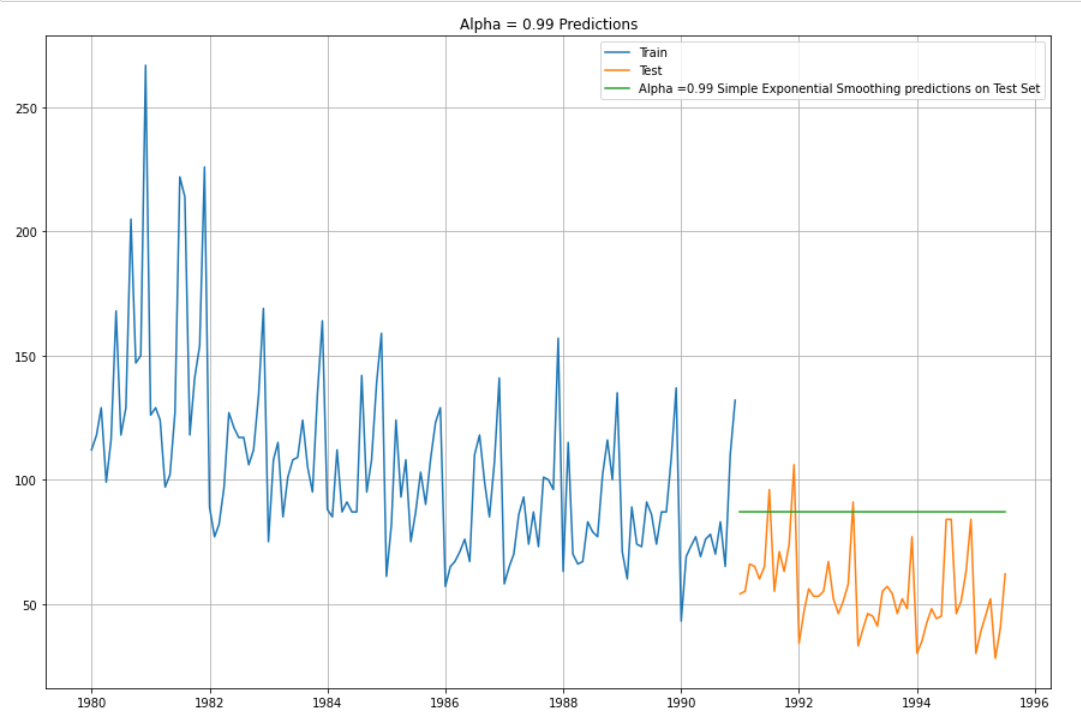
Moving Average

2 point MA capture trend and seasonality almost same as test data

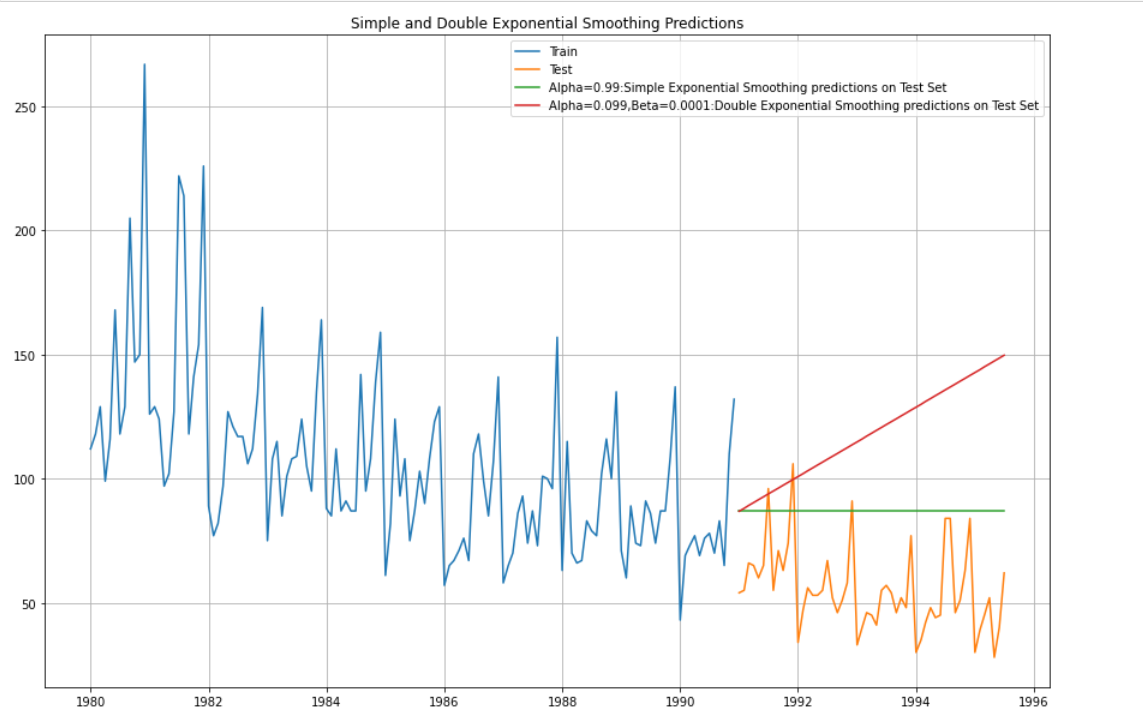




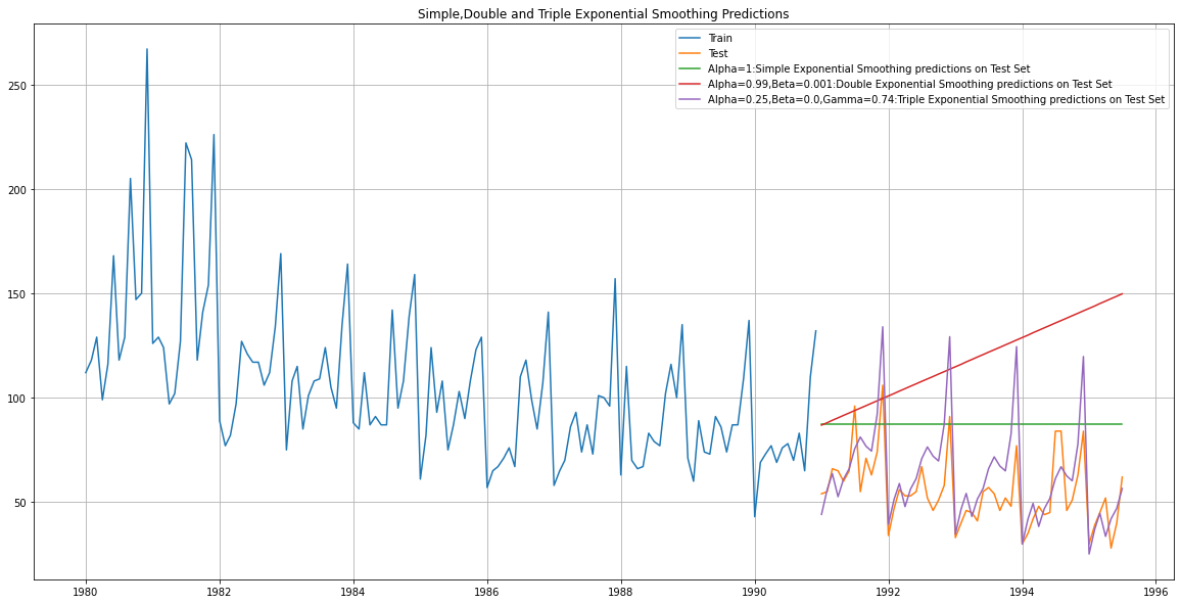
Simple Exponential Smoothing fails to capture trend and seasonality



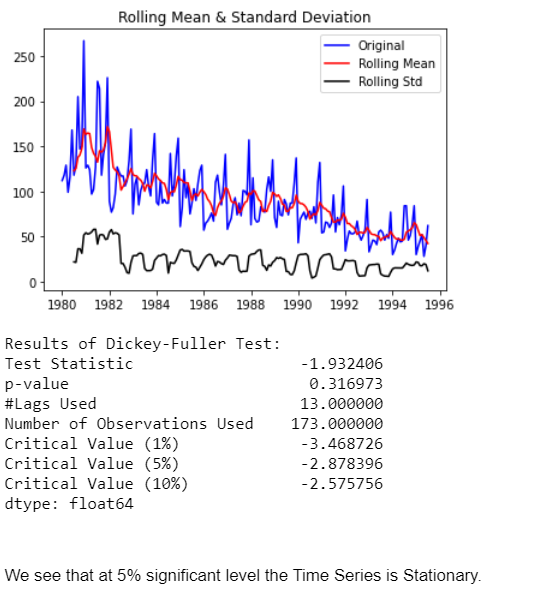
Double Exponential Smoothing fails to capture trend and seasonality



Triple Exponential Smoothing captures trend and seasonality. But it is slightly more than test data

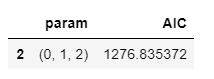


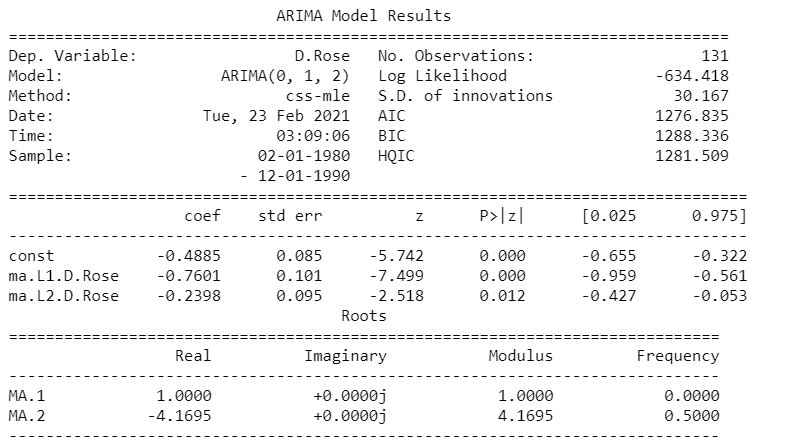
1. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment.  
   Note: Stationarity should be checked at alpha = 0.05.



1. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

Parameters with lowest AIC





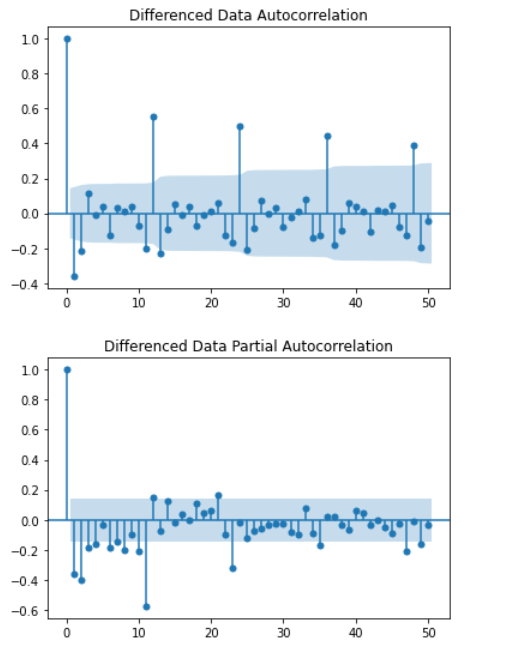
1. Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.

ARIMA model based on the cut-off points of ACF and PACF

The Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 0.

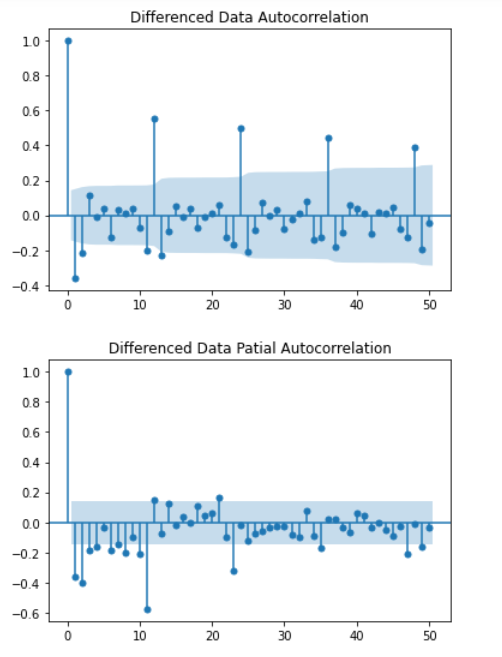
The Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 0.

By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.

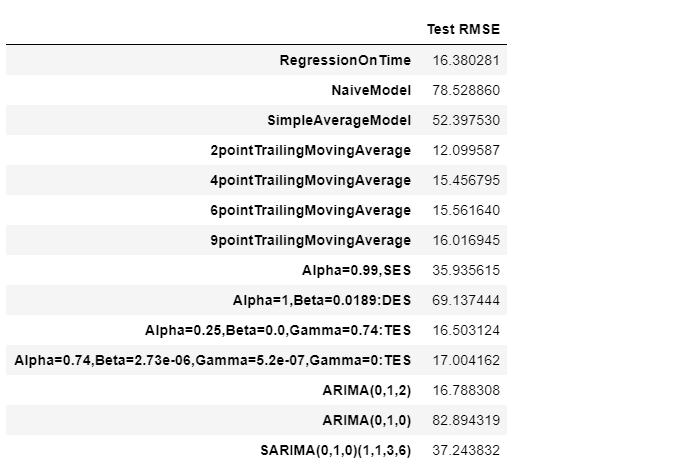


SARIMA model based on the cut-off points of ACF and PACF

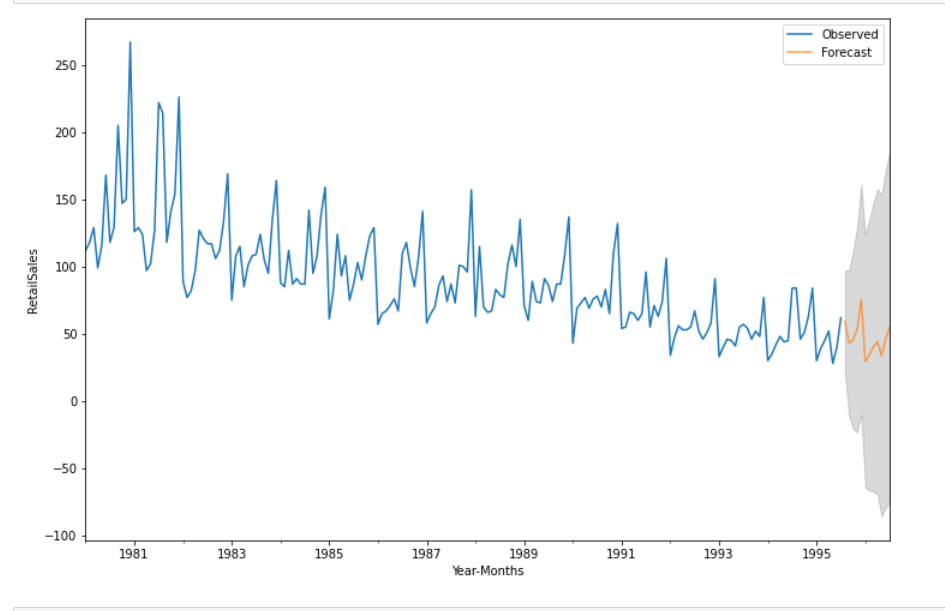
- By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.



1. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data.



1. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.



1. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

As the sales of Rose wine id declining therefore its production can be limited.

Since only December shows maximum sales, wine produce can be increased for December only as it is Christmas month.

For rest of the month it can be the same no of produce no need of extra wine.