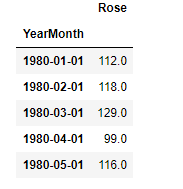
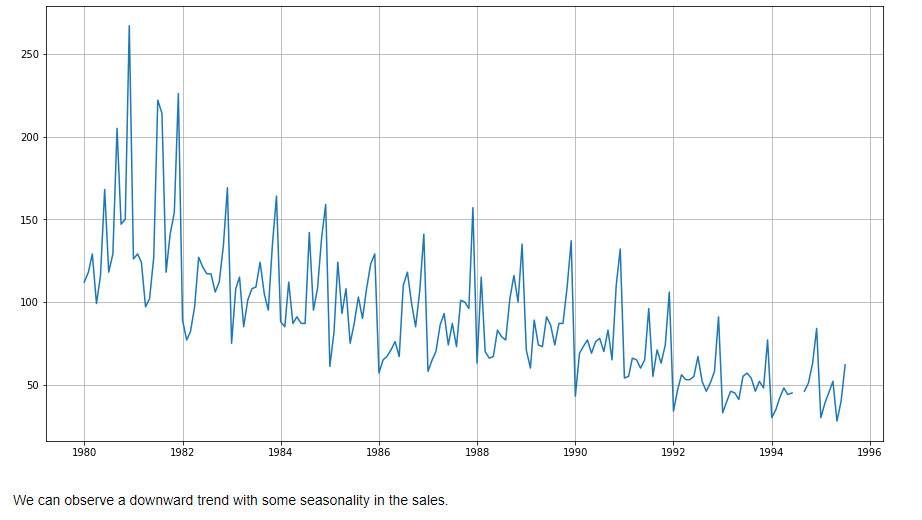
# ABC Estate Wines Report

## Rose Wine Analysis

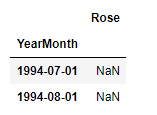
### Read the data as an appropriate Time Series data and plot the data.



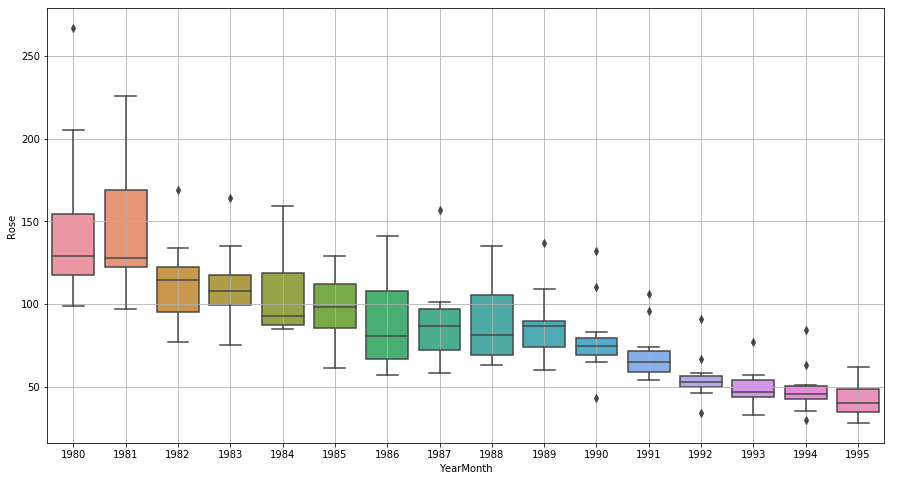


### Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

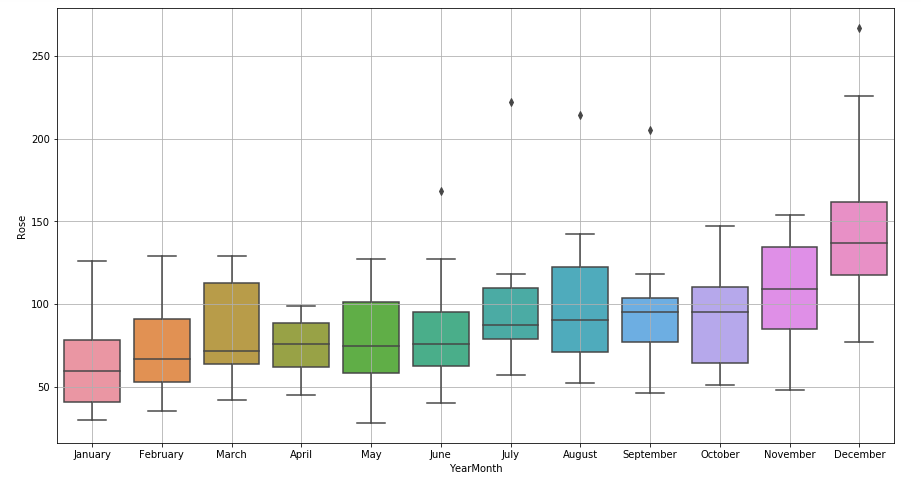
* There are two null values in the dataset



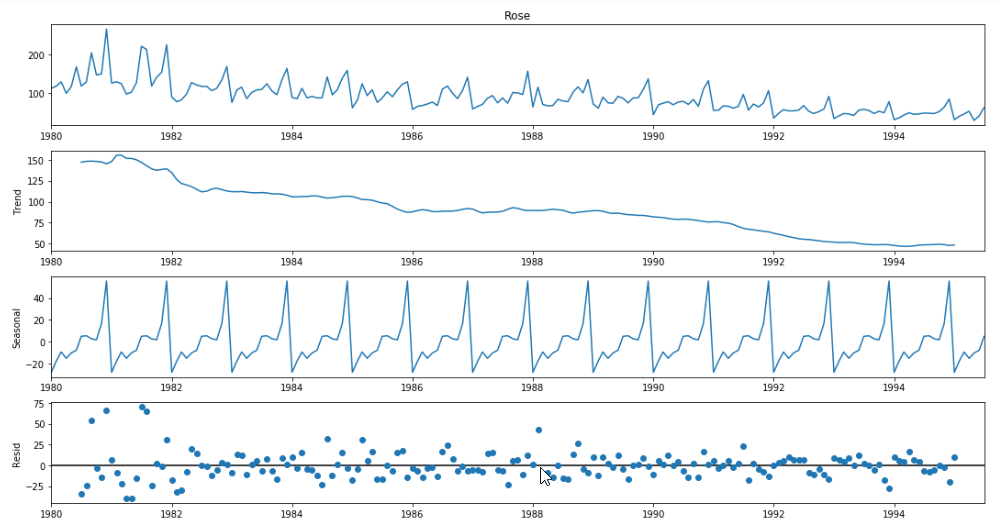
* From the graph we can also observe that there is a downward trend in the sales and seasonality is present.
* **Yearly Sales Analysis :** Maximum Sales were in the year 1981 and minimum in 1995



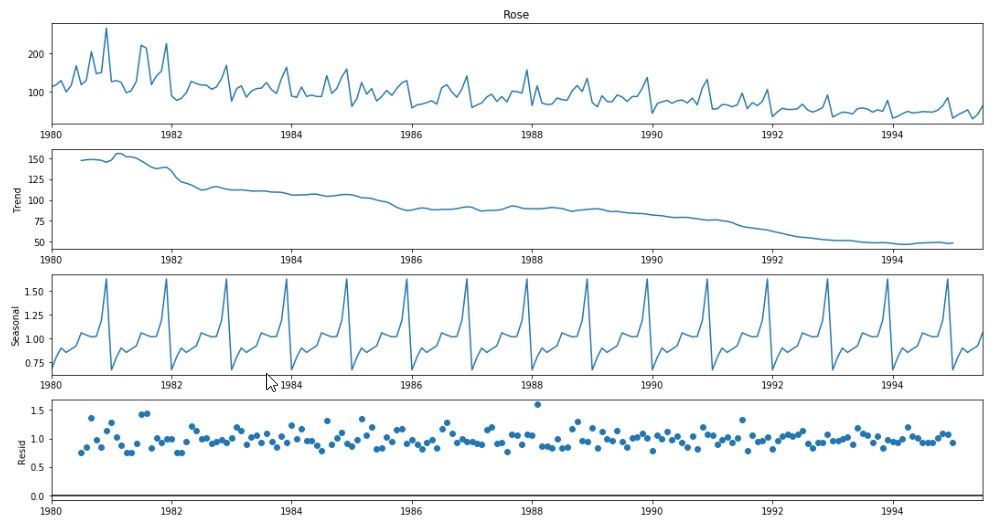
* **Monthly Sales Analysis:** Wine sales have a seasonal effect. Maximum sales are in December and minimum in January.

****

* **Additive Decomposition :**

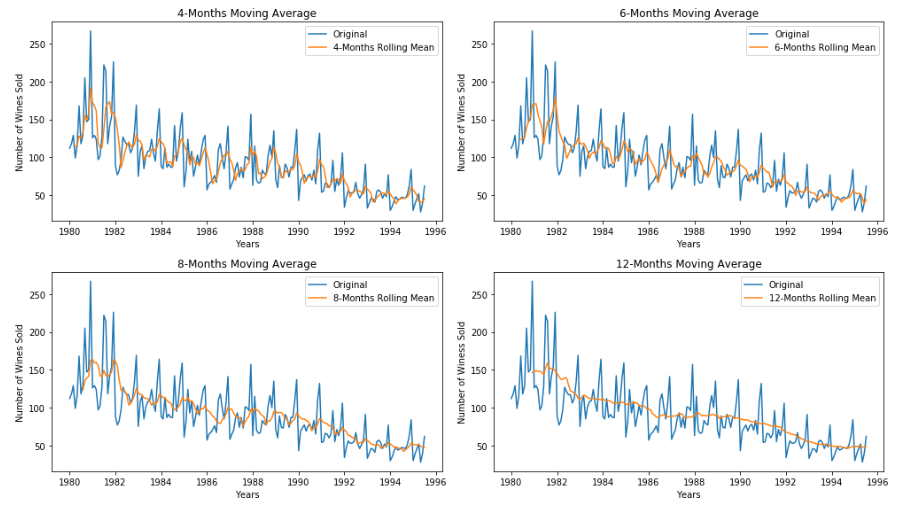


* **Multiplicative Decomposition:**

****

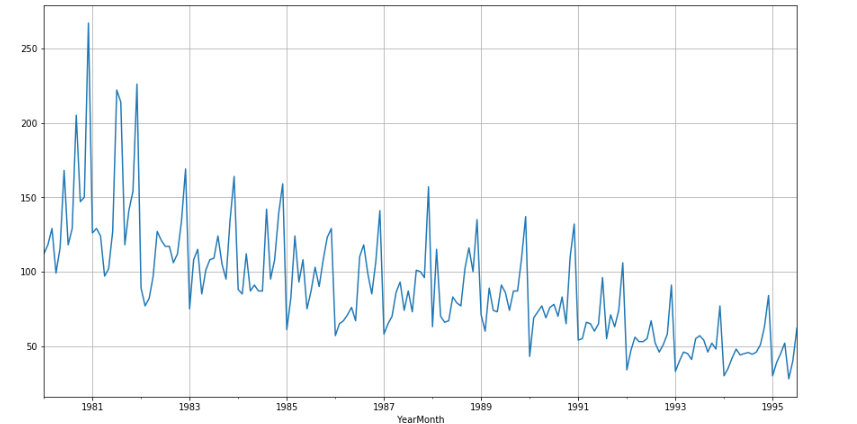
The trend is downward and there is a regular seasonality. As the seasonal effect is not changing with time, we can say that the decomposition is additive.

* Analysing Moving Averages



* Imputing Null Values using rolling Average of 4.

Plot After Imputing Null Value

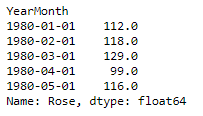


### Split the data into training and test. The test data should start in 1991.

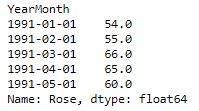
Number of rows in Training dataset: 132

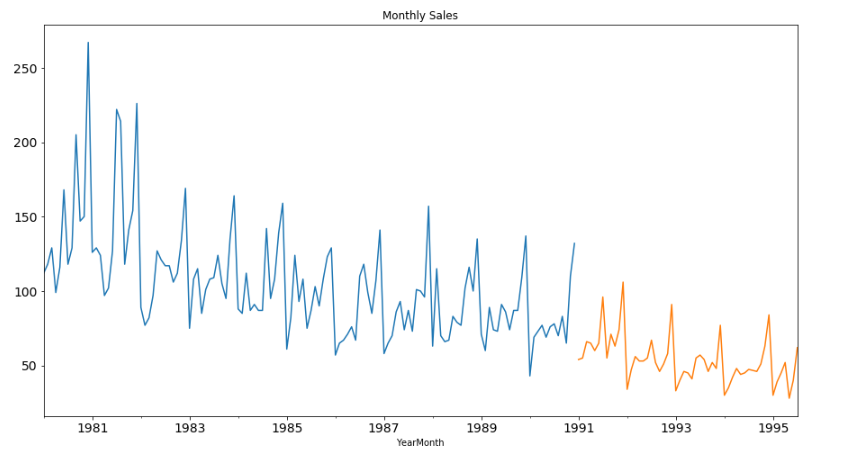
Number of rows in Test dataset: 55

Top rows of Training Dataset



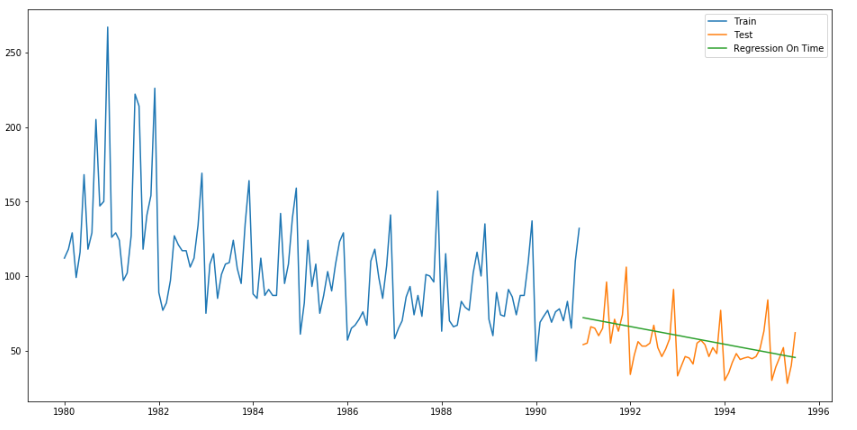
Top rows of Test Dataset





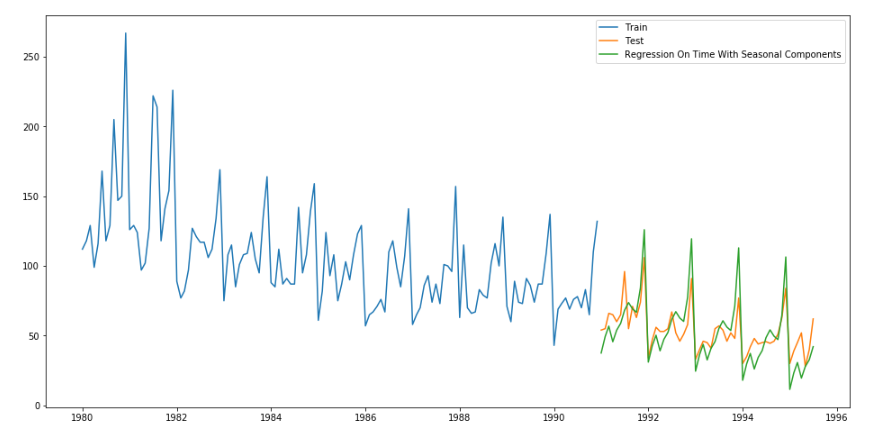
### Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE. - Please do try to build as many models as possible and as many iterations of models as possible with different parameters.

#### Method 1 - Regression On Time



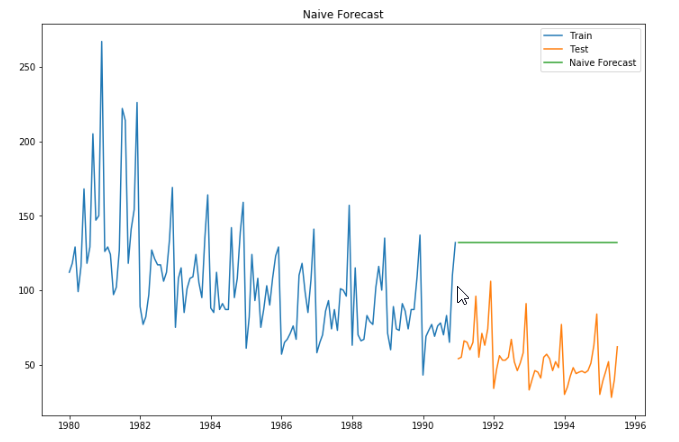
For RegressionOnTime, RMSE is 15.275 MAPE is 22.85

#### Method 2: Regression on Time With Seasonal Components



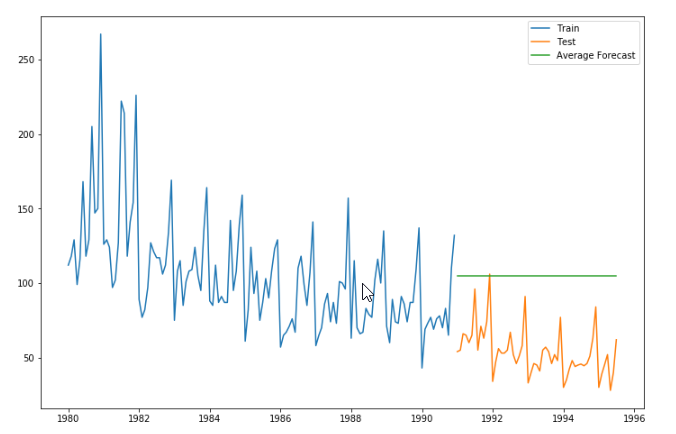
For RegOnTimeSeasonal, RMSE is 13.913 MAPE is 20.35

#### Method 3: Naive Approach: 𝑦̂ 𝑡+1=𝑦𝑡



For Naive model, RMSE is 79.735 MAPE is 145.17

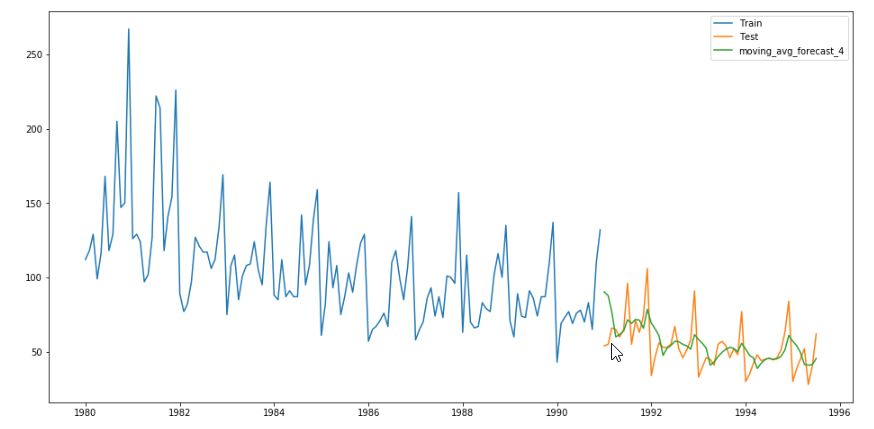
#### Method 4: Simple Average



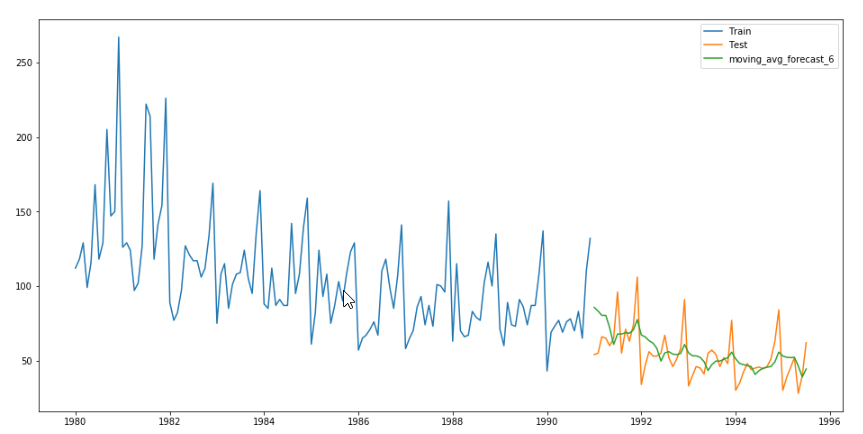
For Simple Average model, RMSE is 53.478 MAPE is 94.98

#### Method 5: Moving Average(MA)

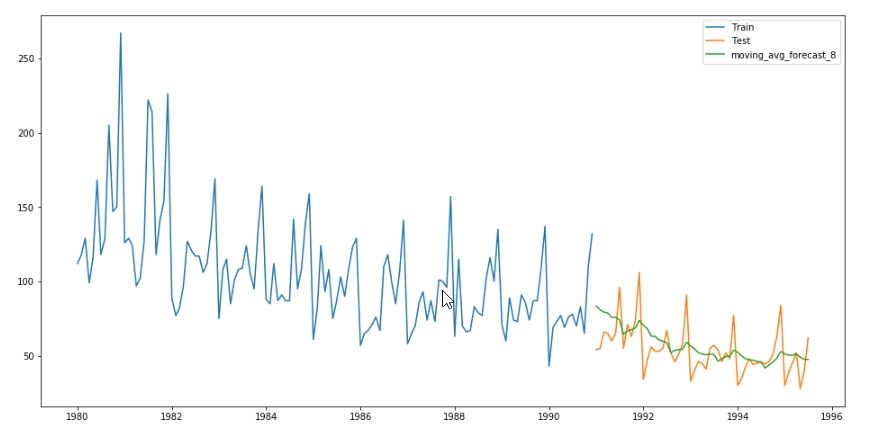
Moving Average(4)



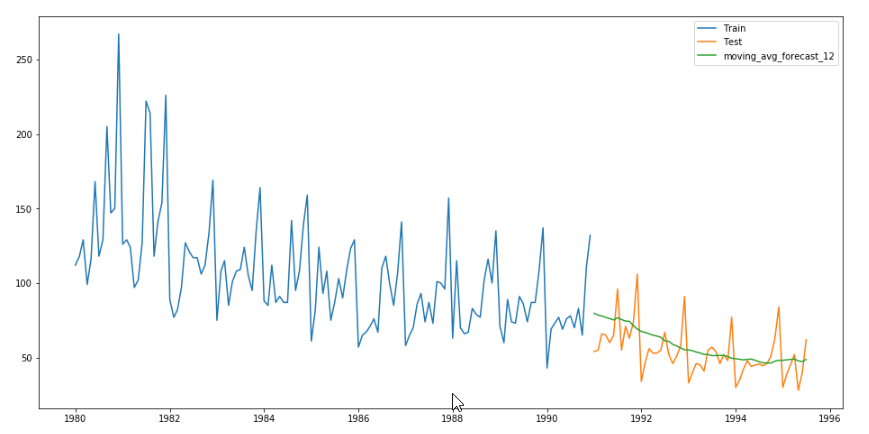
Moving Average(6)



Moving Average(8)



Moving Average(12)



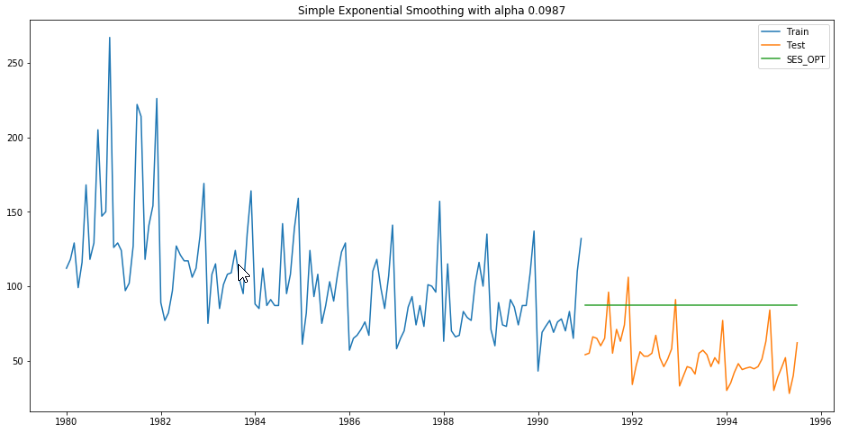
For Simple Average model, moving\_avg\_forecast\_4 RMSE is 14.457 MAPE is 19.50

For Simple Average model, moving\_avg\_forecast\_6 RMSE is 14.569 MAPE is 20.83

For Simple Average model, moving\_avg\_forecast\_8 RMSE is 14.802 MAPE is 21.02

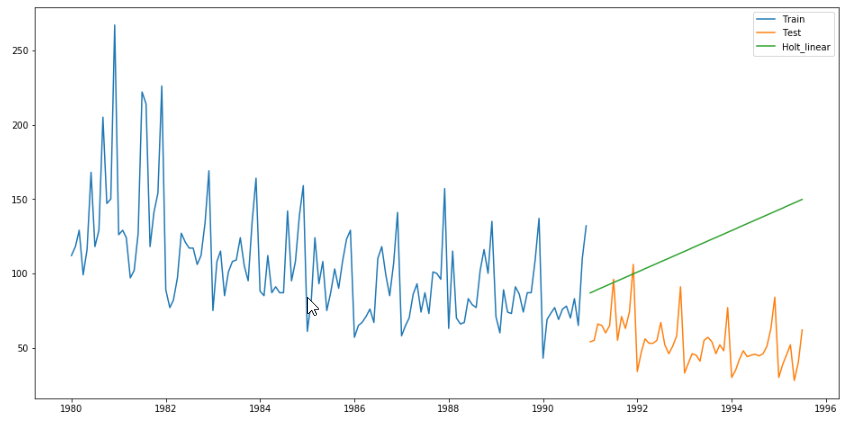
For Simple Average model, moving\_avg\_forecast\_12 RMSE is 15.239 MAPE is 22.10

### Method 6: Simple Exponential Smoothing



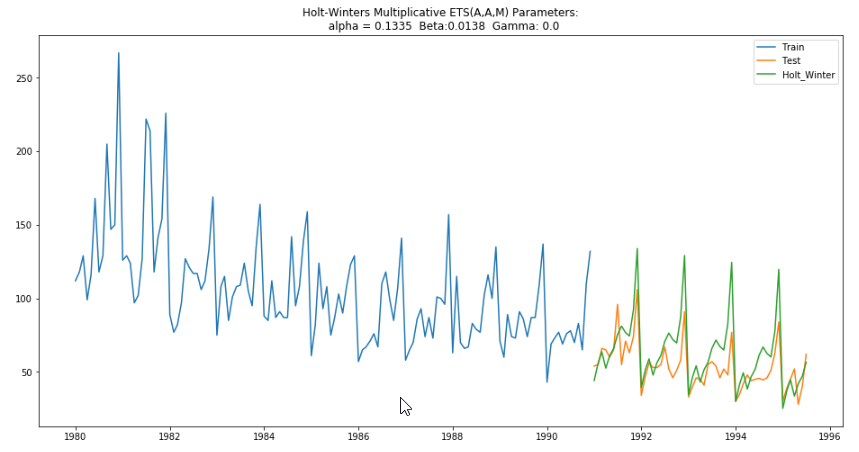
For alpha = 0.10, RMSE is 36.8136 MAPE is 63.93

### Method 7: Holt's Linear Trend Method (Double Exponential Smoothing)



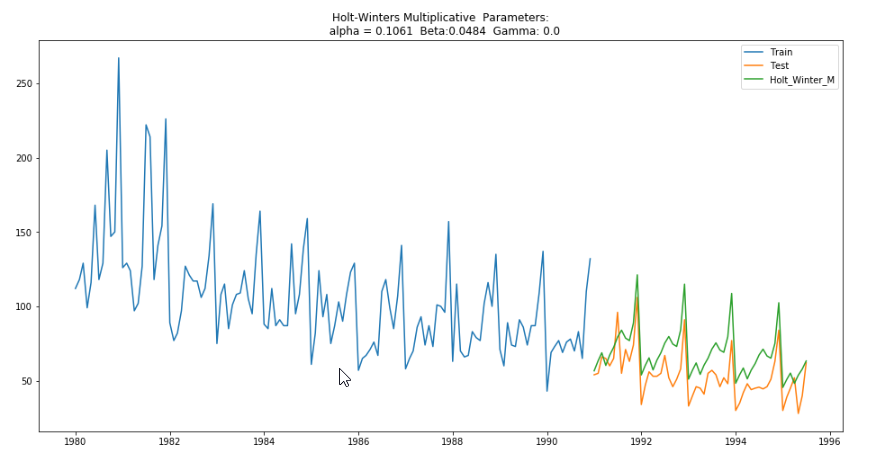
For alpha = 0.16, RMSE is 70.5923 MAPE is 120.31

### Method 8: Holt-Winters Method - Additive seasonality - Build this model and compare with the other models.



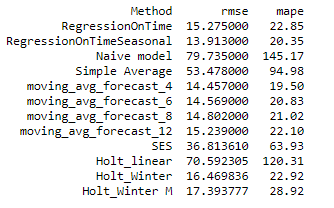
For alpha = 0.13, beta = 0.01, gamma = 0.00, RMSE is 16.4698 MAPE is 22.92

### Method 9: Holt-Winters Method - Multiplicative Model



For alpha = 0.11, beta = 0.05, gamma = 0.00, RMSE is 17.3938 MAPE is 28.92

### Comparing the Results of All Models



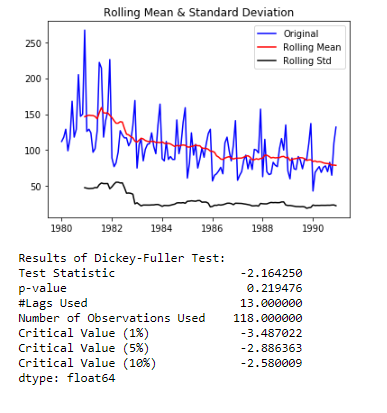
From the above comparison we can observe that so far **RegressionOnTimeSeasonal** seems to be a good fit for the data

### 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.

Augmented Dickey Fuller (ADF) Test for Stationarity

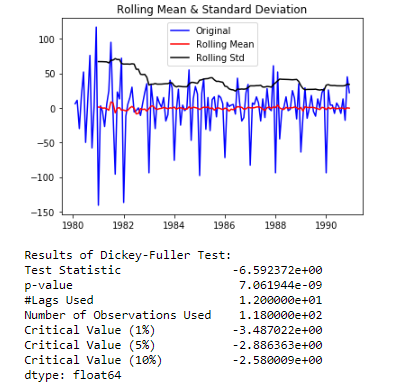
Null Hypothesis: Series is Non Stationary

Alternate Hypothesis: Series is Stationary



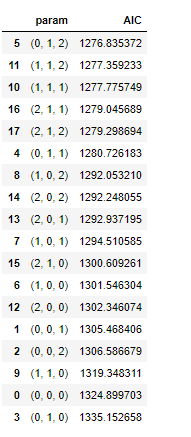
**At alpha 0.05, we cannot reject the null hypothesis. This indicates that there is some evidence that the series is non-stationary. Hence, differentiation is required.**

After Differentiation on the series we observe p-value < 0.05. Therefore we reject the null hypothesis and conclude that the series is stationary after differentiation.



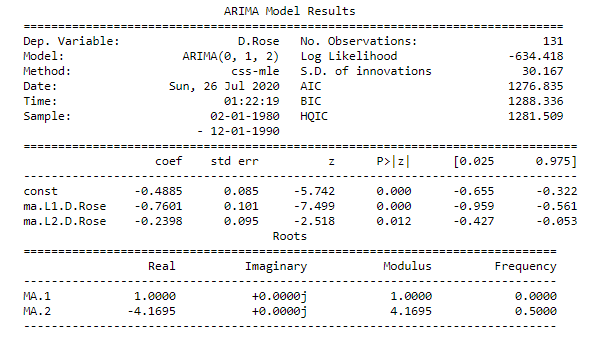
### Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

#### ARIMA



Lowest AIC: 1276.84

Parameter: (0, 1, 2)



**RMSE on Test Data: 15.62369038064324**

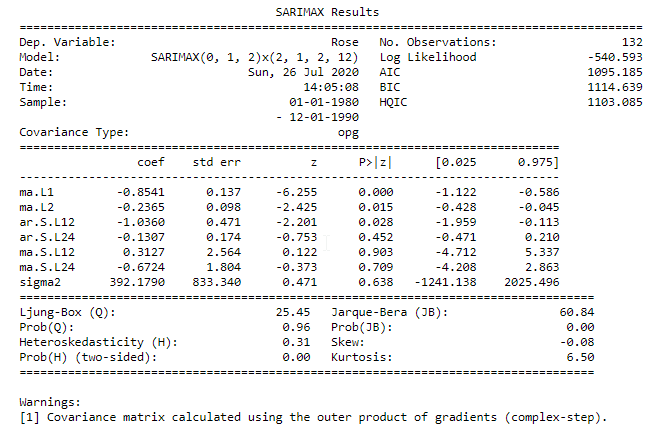
#### SARIMA



Lowest AIC: 774.97

Order: (0, 1, 2)

Seasonal Parameter: (2, 1, 2, 12)

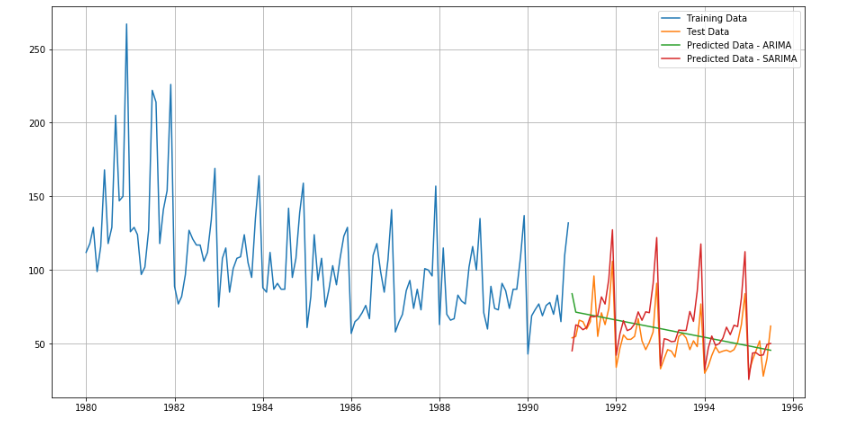


**Diagnostic Plot**



**RMSE on Test Data: 15.580771816054948**

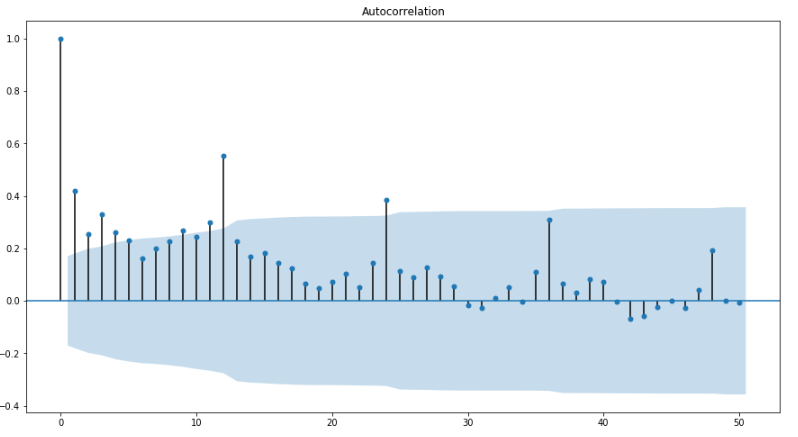
**Comparison**

****

Comparing the RMSE of ARIMA and SARIMA on test data we can conclude that SARIMA has performed better on the series.

#### 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.

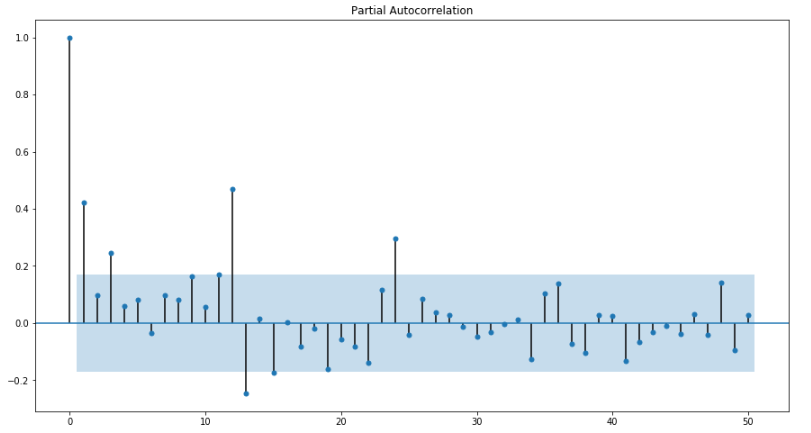
#### ACF Plot



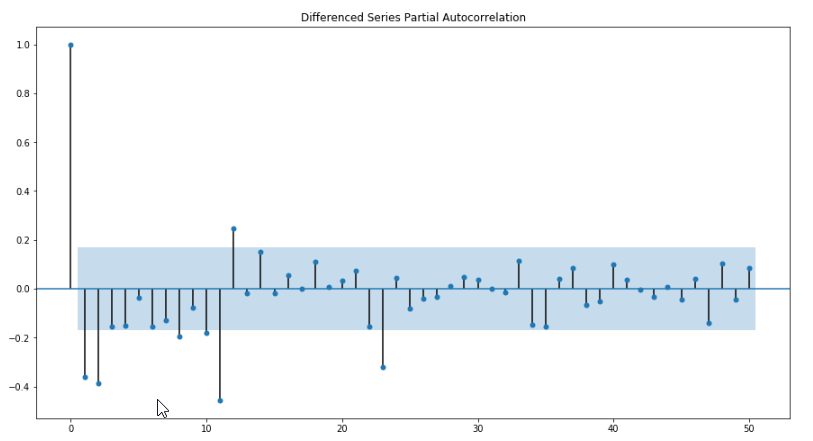
#### Differenced Series AutoCorrelation



#### PACF Plot

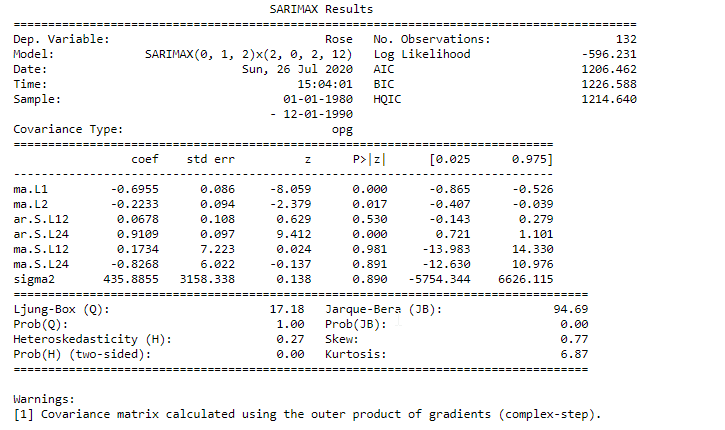


#### Differenced Series Partial Autocorrelation



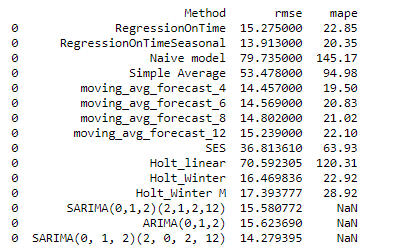
Order: (0, 1, 2)

Seasonal Parameter: (2, 0, 2, 12)



**RMSE: 14.27939540404599**

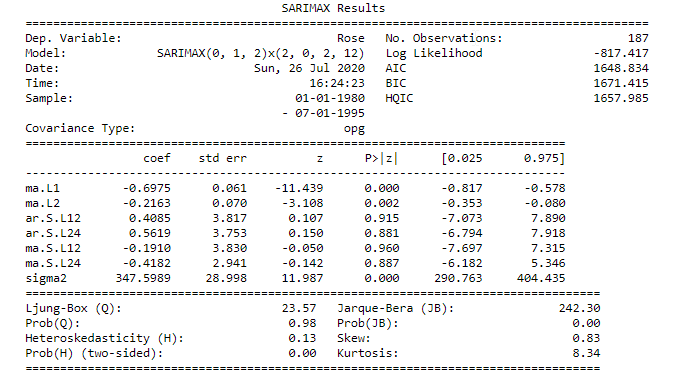
### Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.



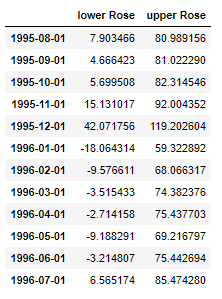
### Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

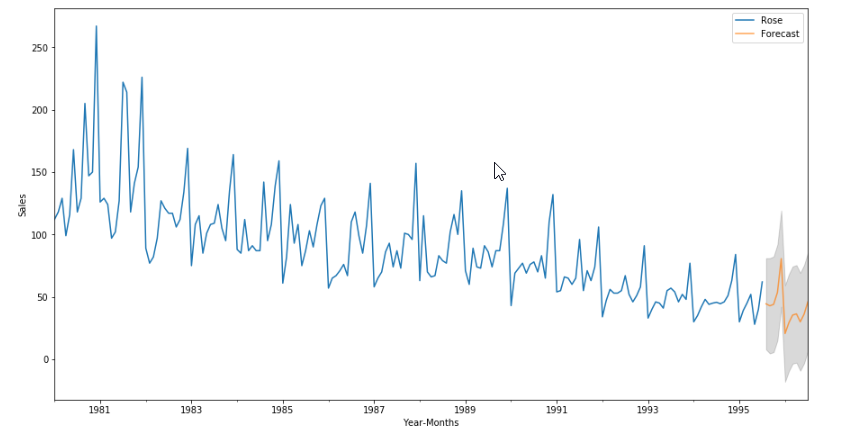
From the above table, we can say that RegressionOnTimeSeasonal, SARIMA(0,1,2)(2,0,2,12) and Moving Average Forecast (4,6 and 8) perform better than the others on the series.

**Building SARIMA(0,1,2)(2,0,2,12) on complete dataset**



Forecast for the next 12 months



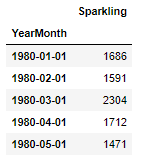


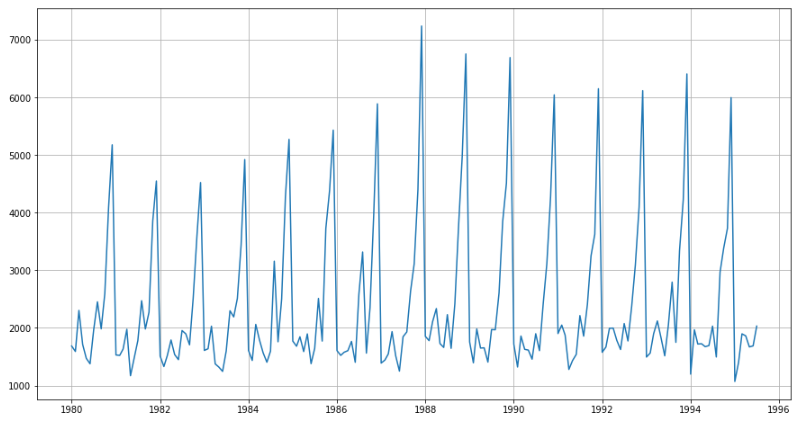
### Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

* We observe a downward trend in the sales of the wine over the years. Company needs to look at reasons of the decline. There may be new competitors offering better quality wine over similar price range leading to customer churn for the wine estate. Business should start analysing competition and reduce price or increase quality accordingly.
* We can see seasonality in the sales which indicates maximum sales are in the month of December and minimum in January. New marketing strategies like discounts/ promotional offers need to be planned to increase the sales in the months that currently have lower sales.
* Looking at the current value, sales will continue to decrease for the next ear and immediate action should be taken in order to prevent loss.

## Sparkling Wine Analysis

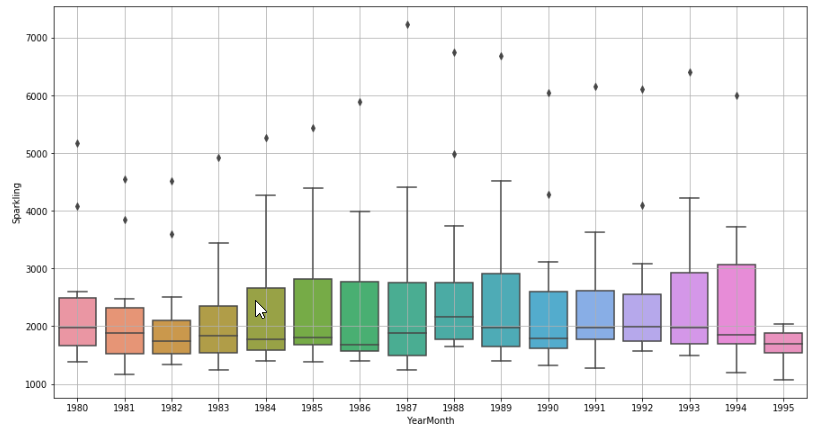
### Read the data as an appropriate Time Series data and plot the data.



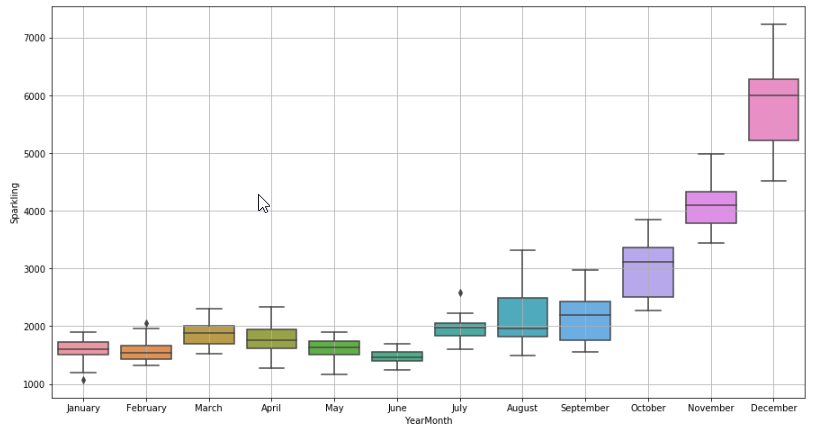


### Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

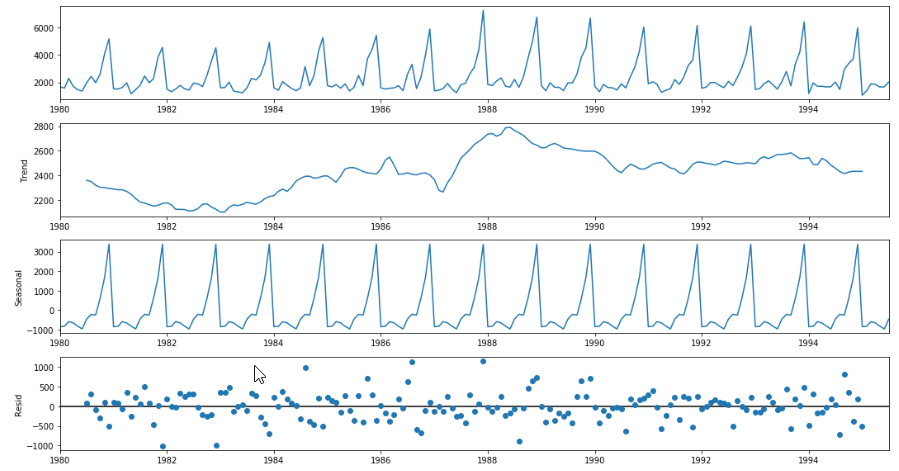
* There are no null values in the dataset
* From the graph we can also observe that there seasonality is the dataset.
* **Yearly Sales Analysis**



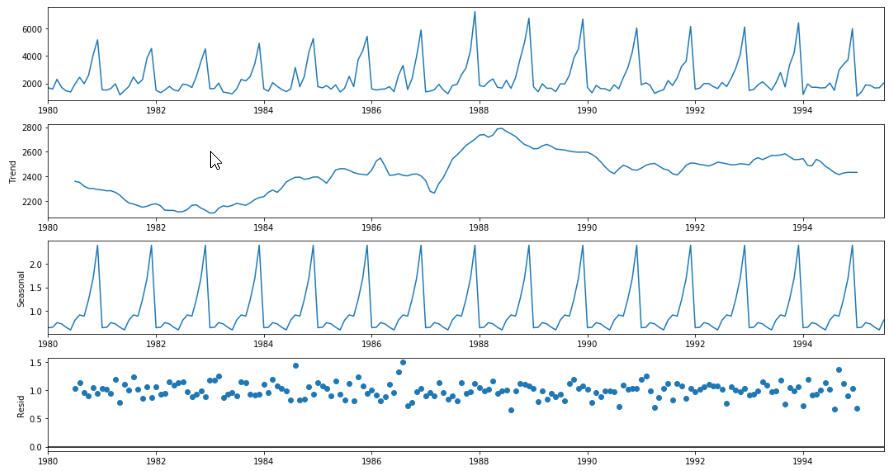
* **Monthly Sales Analysis:** Wine sales have a seasonal effect. Maximum sales are in December and minimum in June.

****

* **Additive Decomposition :**

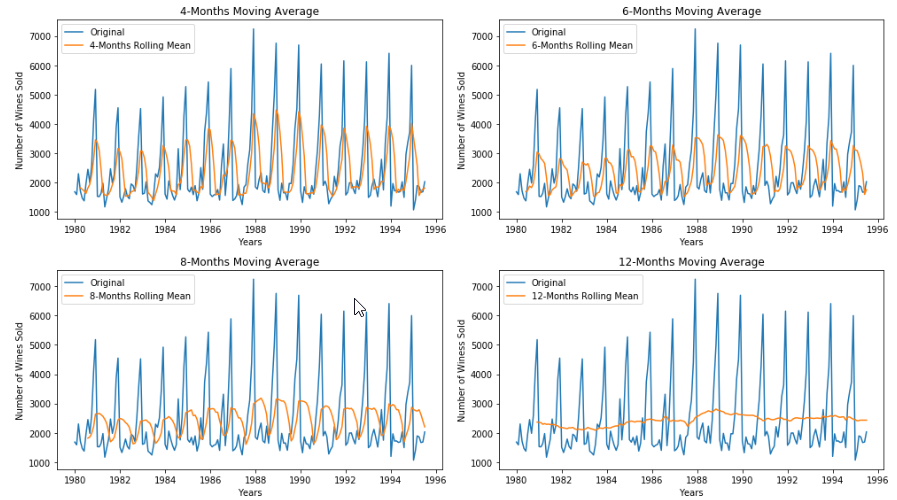


* **Multiplicative Decomposition:**

****

The trend is increasing till 1988 and there is a decline in sales till 1991 after which the trend is uniform. There is a regular seasonality in the sales. The series is Multiplicative.

* Analysing Moving Averages

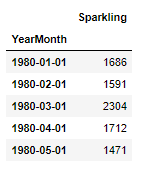


### Split the data into training and test. The test data should start in 1991.

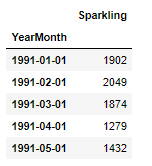
Number of rows in Training dataset: 132

Number of rows in Test dataset: 55

Top rows of Training Dataset



Top rows of Test Dataset



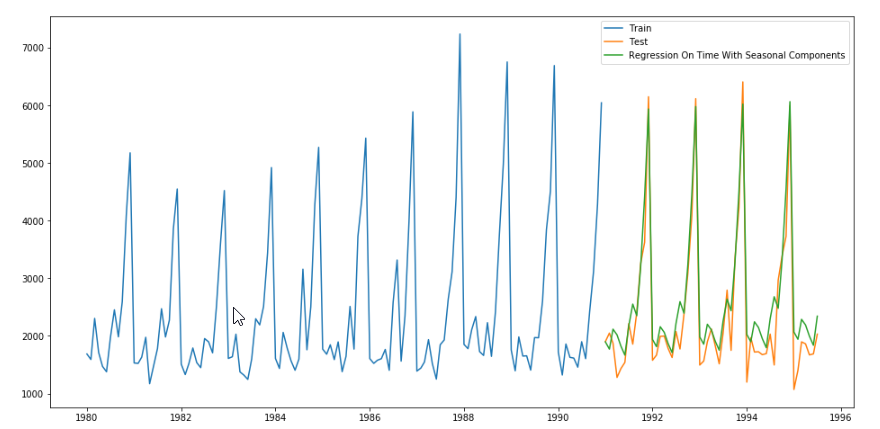
### Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE. - Please do try to build as many models as possible and as many iterations of models as possible with different parameters.

#### Method 1 - Regression On Time



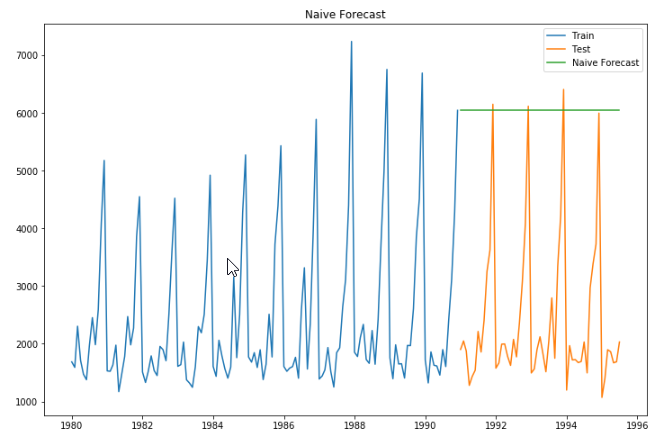
For RegressionOnTime, RMSE is 1389.135 MAPE is 50.15

#### Method 2: Regression on Time With Seasonal Components



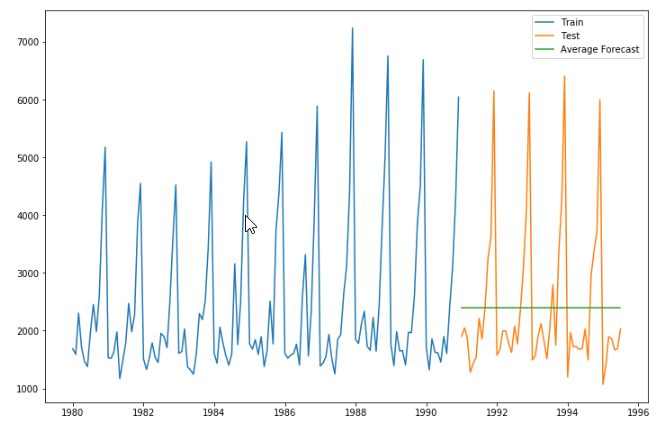
For RegOnTimeSeasonal, RMSE is 420.867 MAPE is 13.29

#### Method 3: Naive Approach: 𝑦̂ 𝑡+1=𝑦𝑡



For Naive model, 3864.279 MAPE is 152.87

#### Method 4: Simple Average

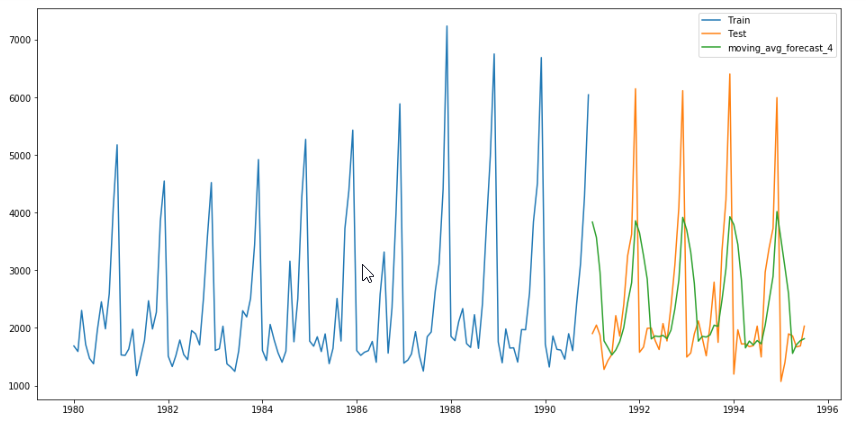


For Simple Average model, RMSE is 1275.082 MAPE is 38.90

#### Method 5: Moving Average(MA)

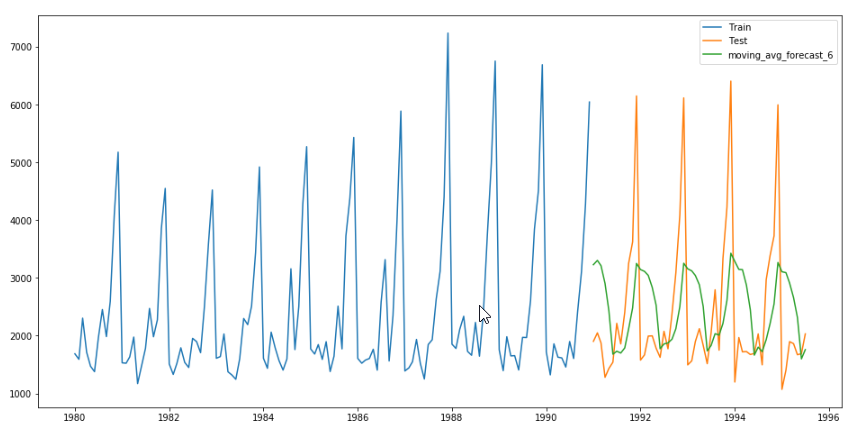
Moving Average(4)

For Simple Average model, moving\_avg\_forecast\_4 RMSE is 1156.590 MAPE is 35.96



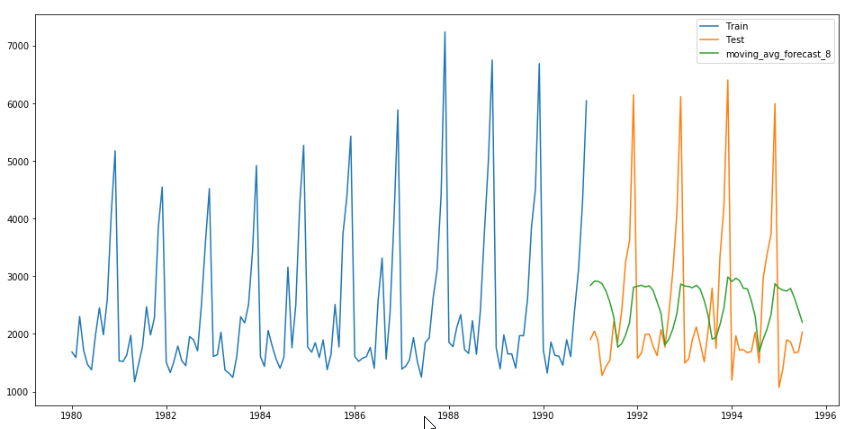
Moving Average(6)

For Simple Average model, moving\_avg\_forecast\_6 RMSE is 1283.927 MAPE is 43.86



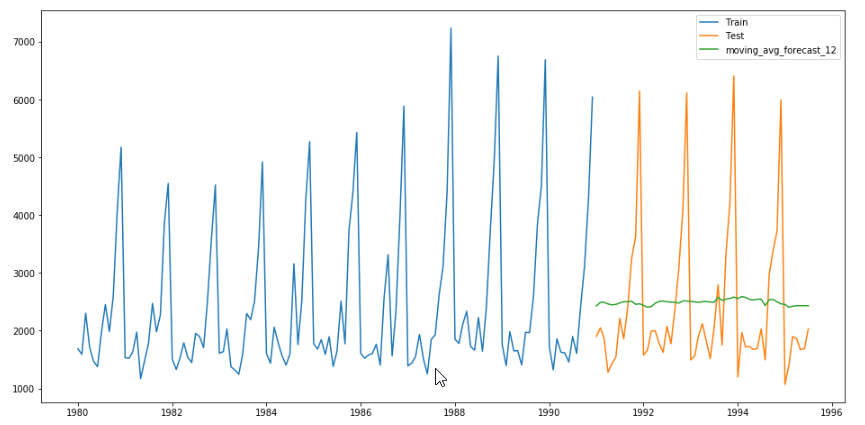
Moving Average(8)

For Simple Average model, moving\_avg\_forecast\_8 RMSE is 1342.568 MAPE is 46.46

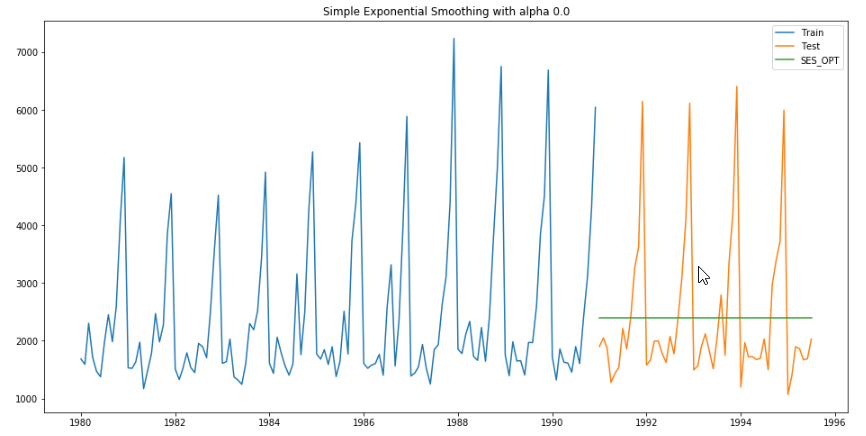


Moving Average(12)

For Simple Average model, moving\_avg\_forecast\_12 RMSE is 1267.925 MAPE is 40.19

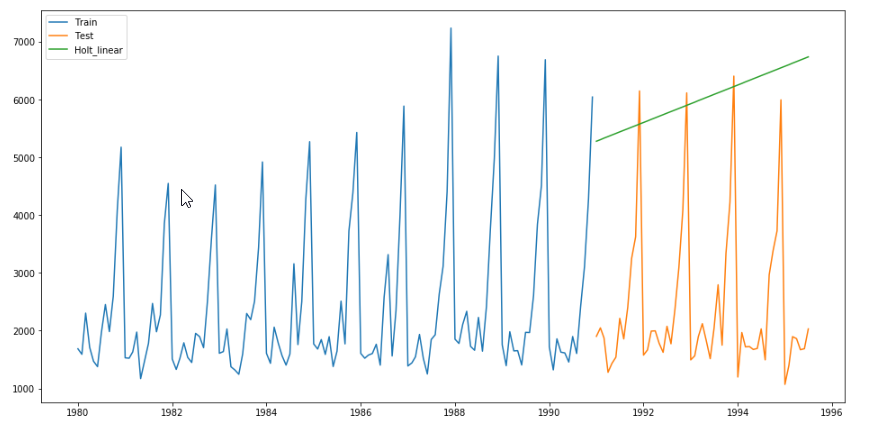


### Method 6: Simple Exponential Smoothing

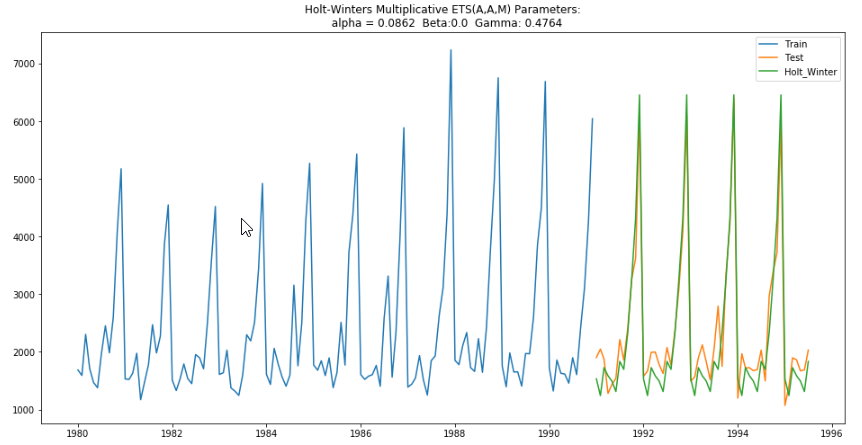


For alpha = 0.00, RMSE is 1275.0818 MAPE is 38.90

### Method 7: Holt's Linear Trend Method (Double Exponential Smoothing)

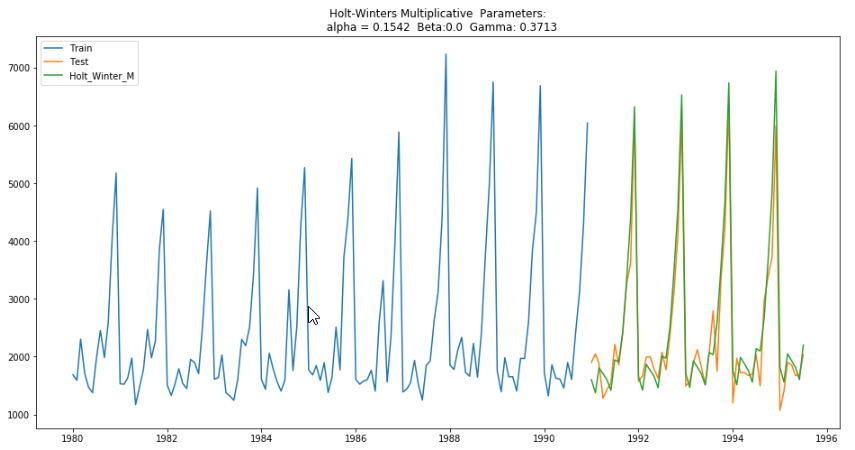
 For alpha = 0.65, RMSE is 3850.9898 MAPE is 152.06

### Method 8: Holt-Winters Method - Additive seasonality - Build this model and compare with the other models.



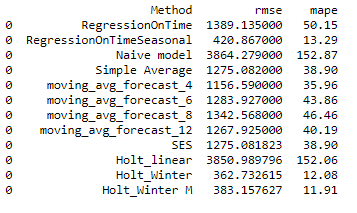
### For alpha = 0.09, beta = 0.00, gamma = 0.48, RMSE is 362.7326 MAPE is 12.08

### Method 9: Holt-Winters Method - Multiplicative Model



For alpha = 0.15, beta = 0.00, gamma = 0.37, RMSE is 383.1576 MAPE is 11.91

### Comparing the Results of All Models



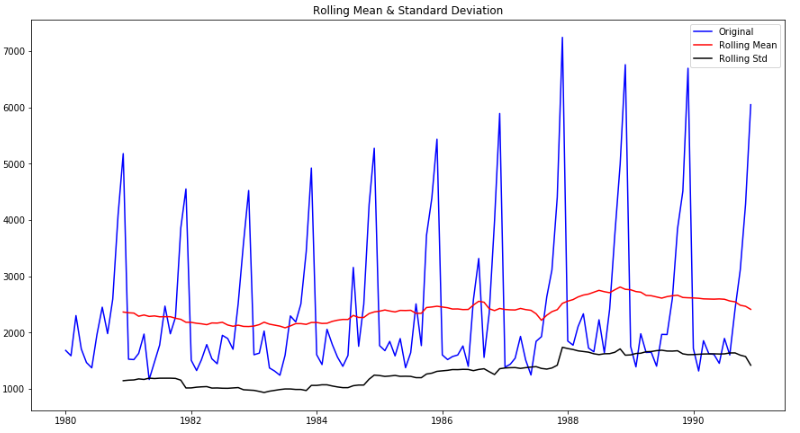
From the above comparison we can observe that so far Holt's Winter Multiplicative seems to be a good fit for the data

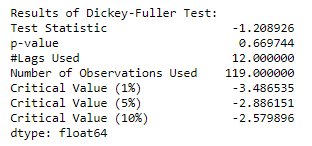
### Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and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

Augmented Dickey Fuller (ADF) Test for Stationarity

Null Hypothesis: Series is Non Stationary

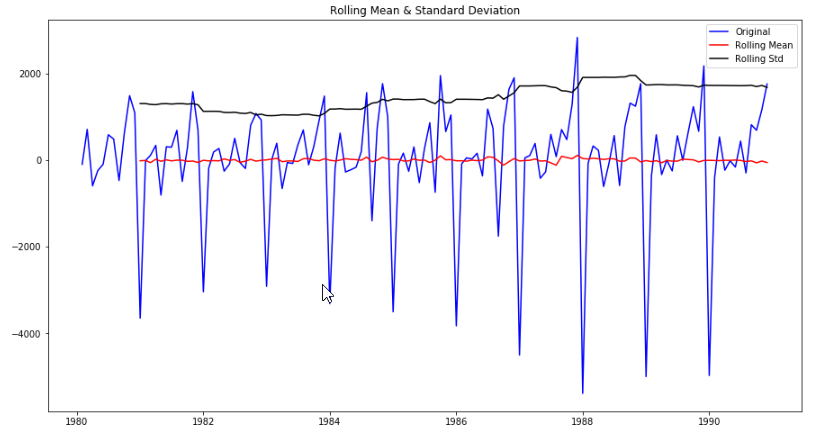
Alternate Hypothesis: Series is Stationary

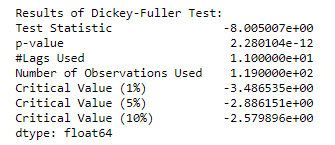




**At alpha 0.05, we cannot reject the null hypothesis. This indicates that there is some evidence that the series is non-stationary. Hence, differentiation is required.**

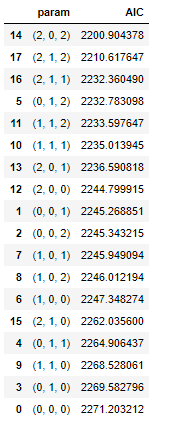
After Differentiation on the series we observe p-value < 0.05. Therefore we reject the null hypothesis and conclude that the series is stationary after differentiation.





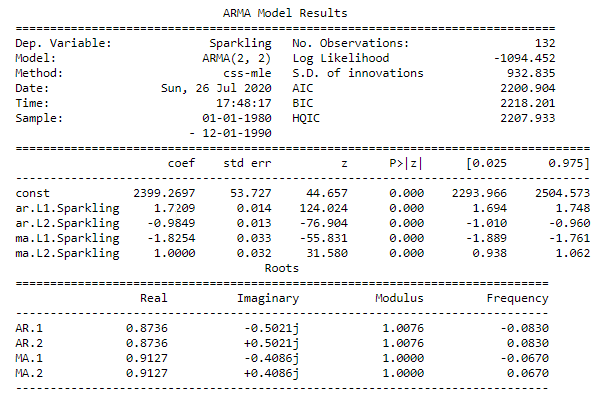
### Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

#### ARIMA



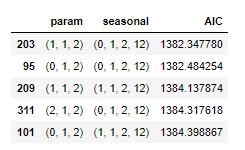
Lowest AIC: 2200.90

Parameter: (2, 0, 2)



**RMSE on Test Data: 1004.87**

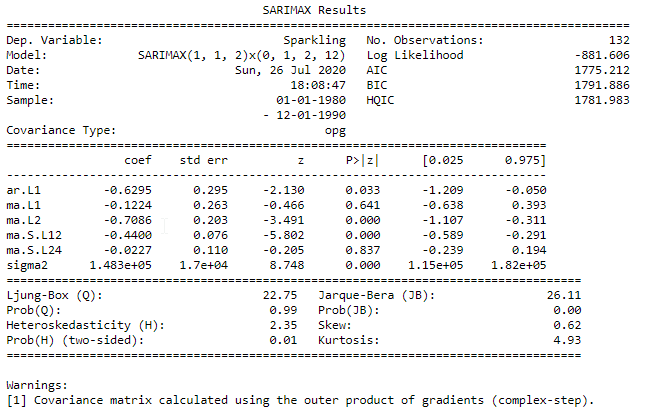
#### SARIMA



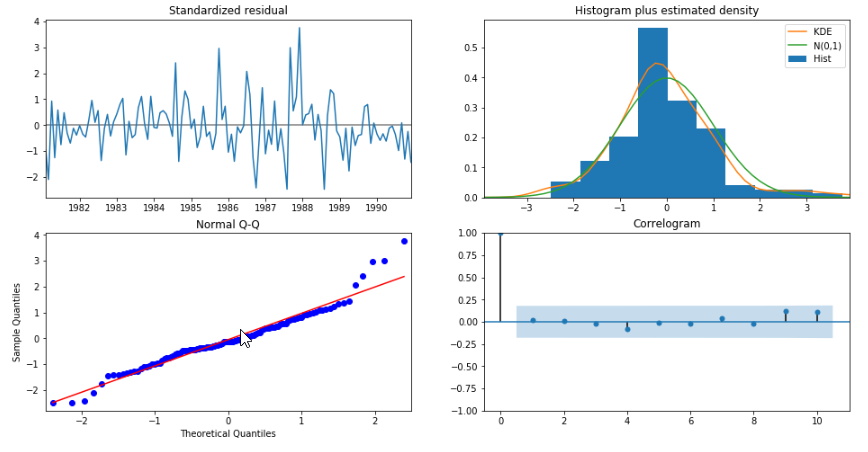
Lowest AIC: 1382.35

Order: (1, 1, 2)

Seasonal Parameter: (0, 1, 2, 12)



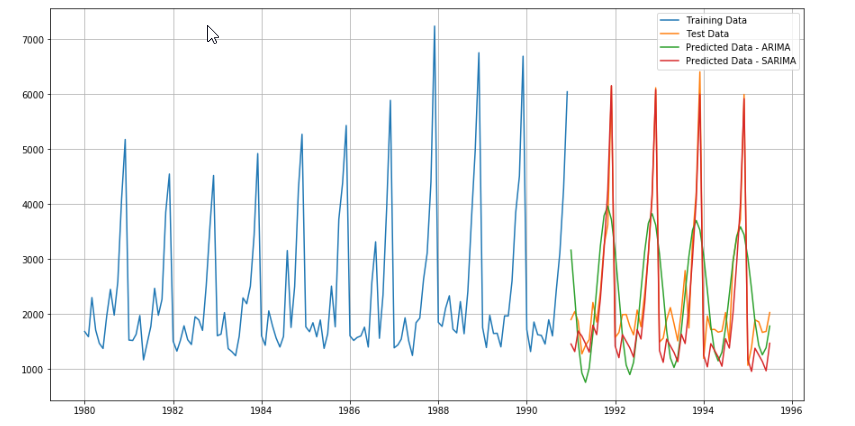
**Diagnostic Plot**



**RMSE on Test Data: 453.62129428741446**

**Comparison**

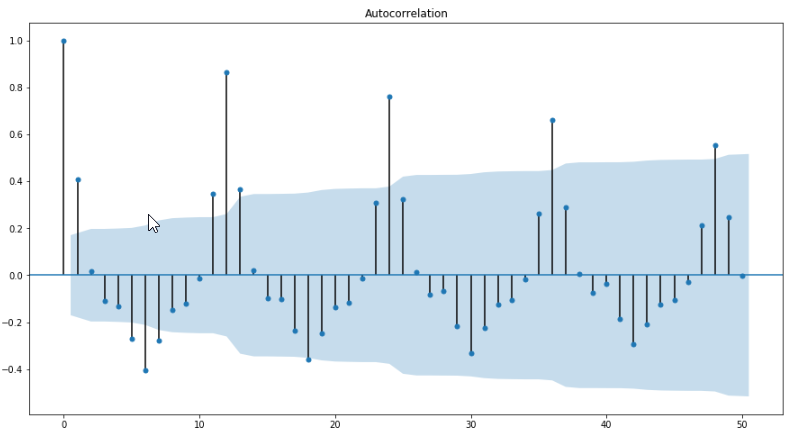
****

****

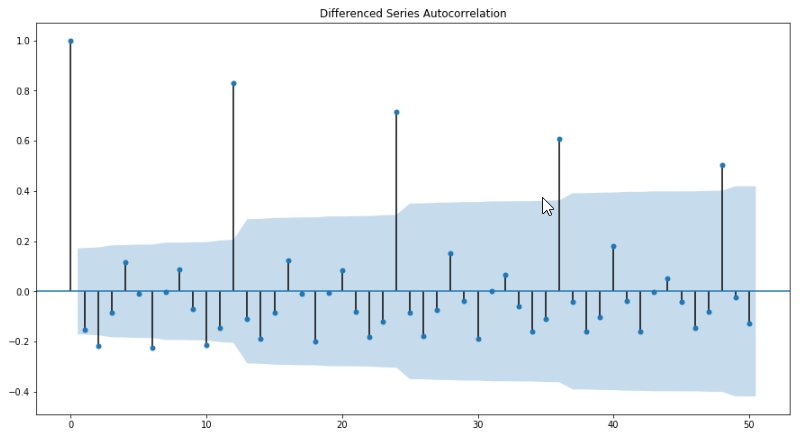
Comparing the RMSE of ARIMA and SARIMA on test data we can conclude that SARIMA has performed better on the series.

#### 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.

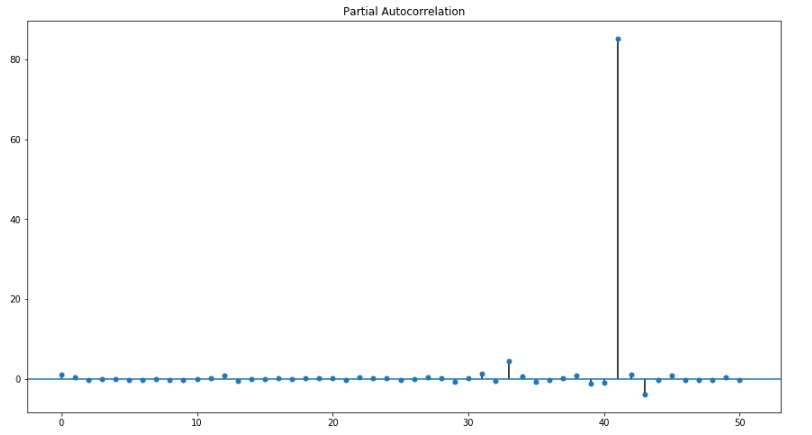
#### ACF Plot



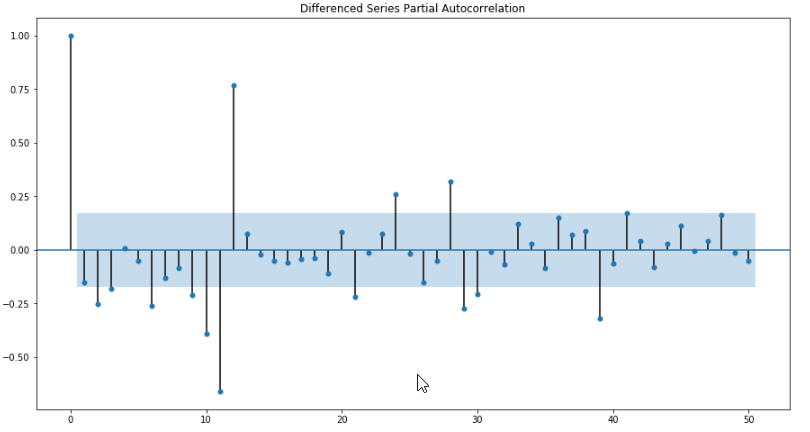
#### Differenced Series AutoCorrelation



#### PACF Plot

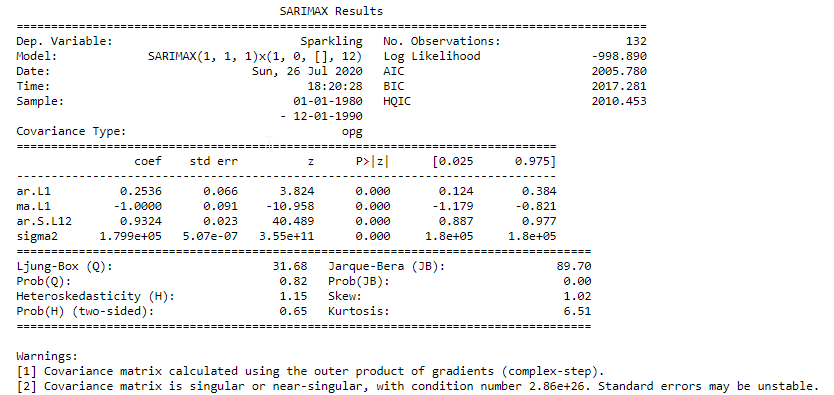


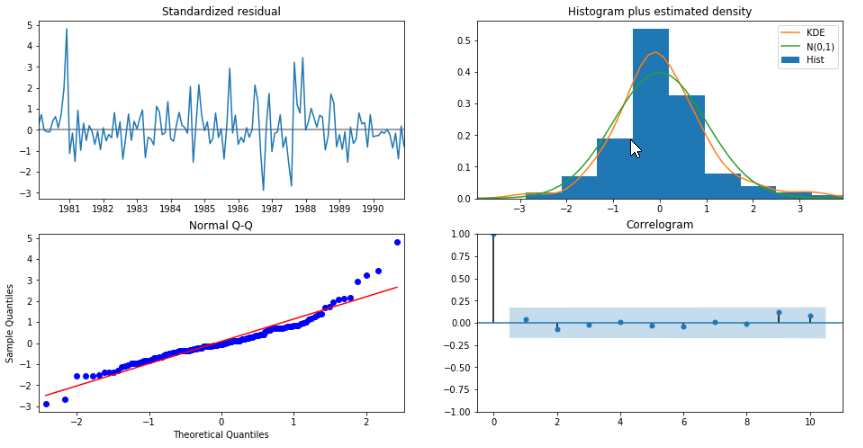
#### Differenced Series Partial Autocorrelation



Order: (1, 1, 1)

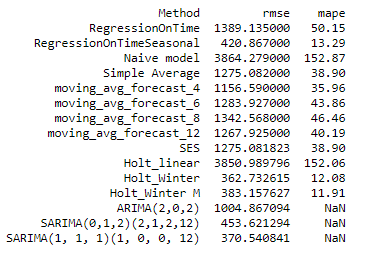
Seasonal Parameter: (1, 0, 0, 12)





**RMSE: 370.54084144211396**

### Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.



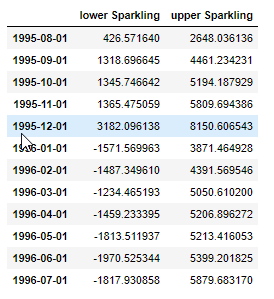
### Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

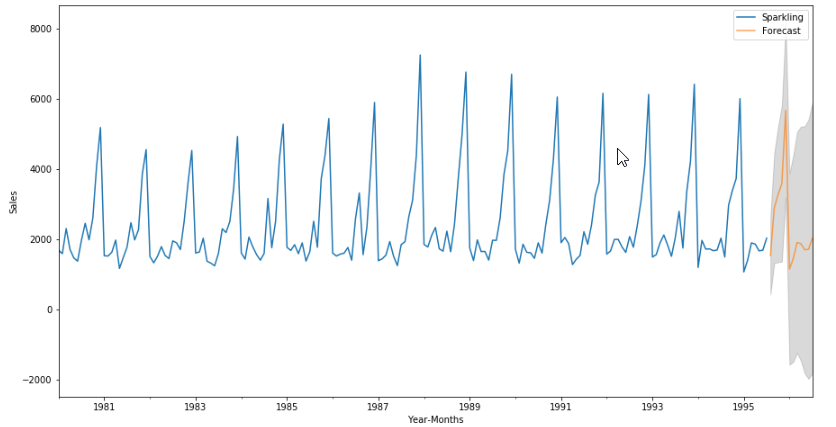
From the above table, we can say that Holt’s Winter and SARIMA(1,1,1)(1,0,0,12) perform better than the others on the series.

**Building SARIMA(1,1,1)(1,0,0,12) on complete dataset**



Forecast for the next 12 months





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

* We observe that the trend is increasing till 1988 and there is a decline in sales till 1991 after which the trend is uniform. As the sales are not dropping significantly, business can plan promotional strategies to target more customers to increase the sales rather than changing the product.
* We can see seasonality in the sales which indicates maximum sales are in the month of December and minimum in June. New marketing strategies like discounts/ promotional offers need to be planned to increase the sales in the months that currently have lower sales.