Yash Joshi
07-11-2021

Time Series - **Project Report**

Problem

1. Read the data as an appropriate Time Series data and plot the data.
2. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.
3. Split the data into training and test. The test data should start in 1991.
4. Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data.

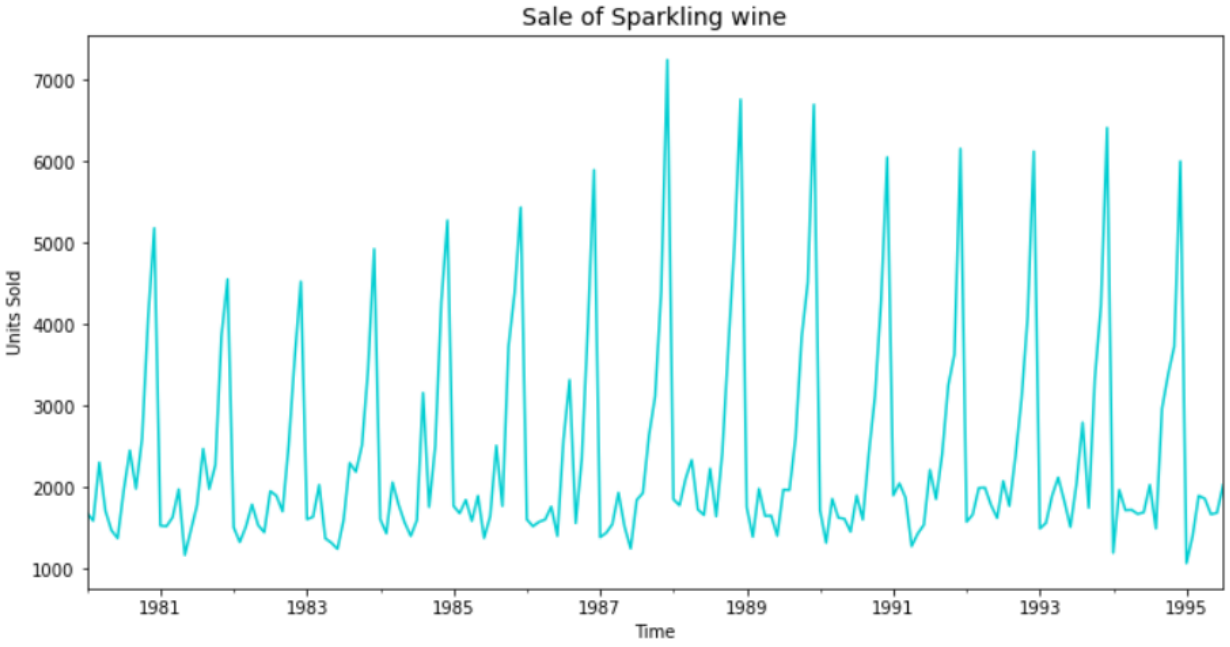
   Other models such as regression,naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.
5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also 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.

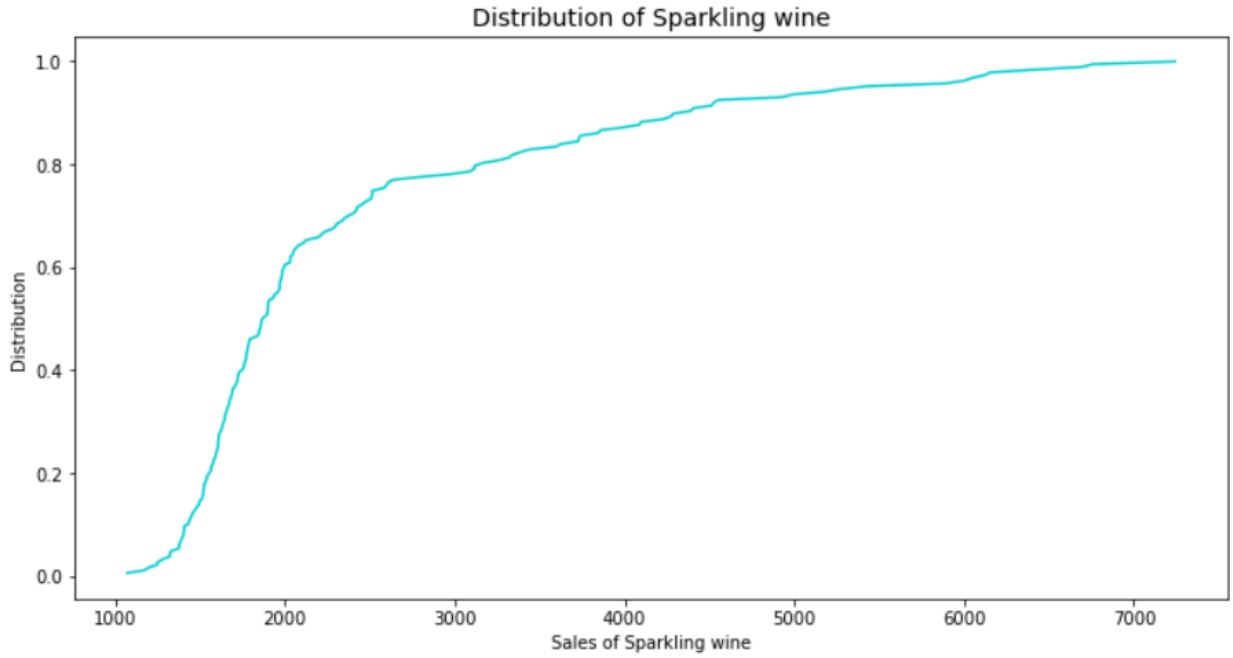
   Note: Stationarity should be checked at alpha = 0.05.
6. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.
7. Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.
8. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data.
9. 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.
10. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

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

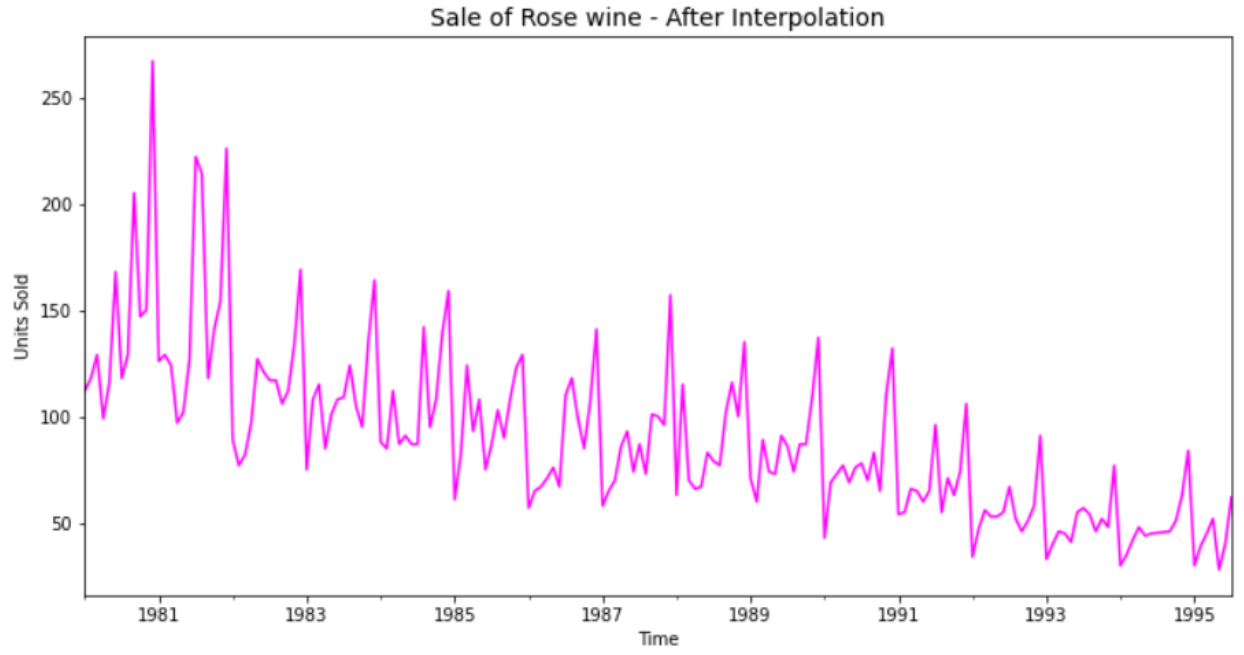
Ans 1.

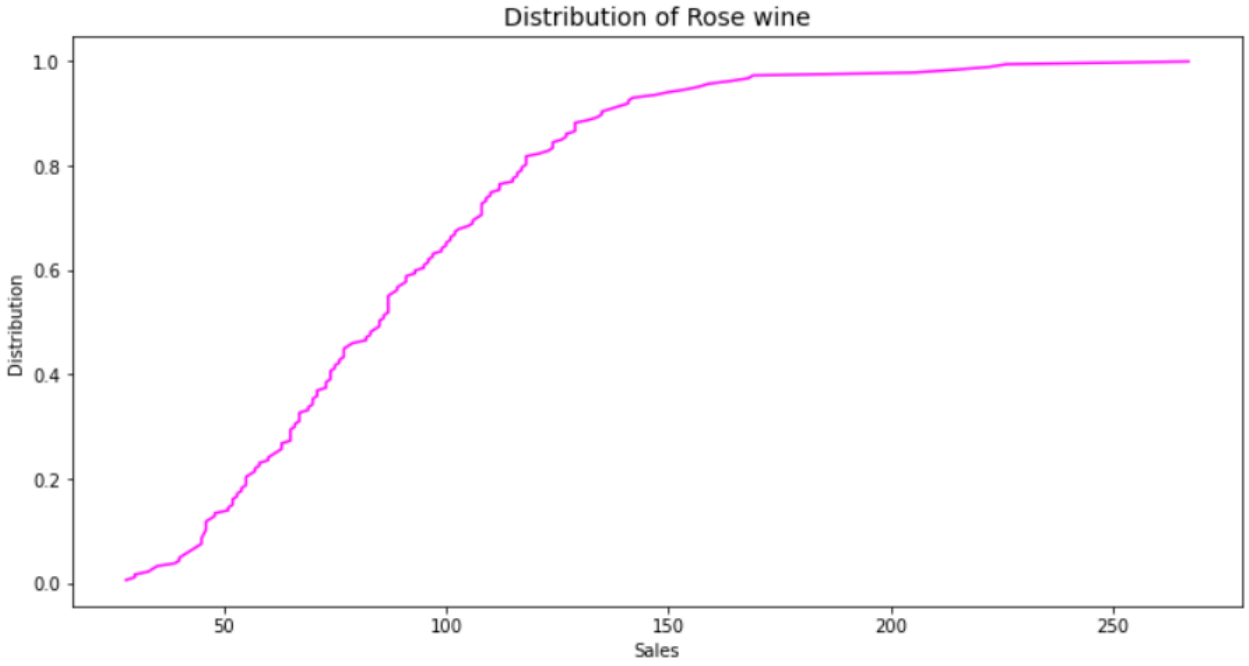
Sparkling Dataset:





Rose Dataset:

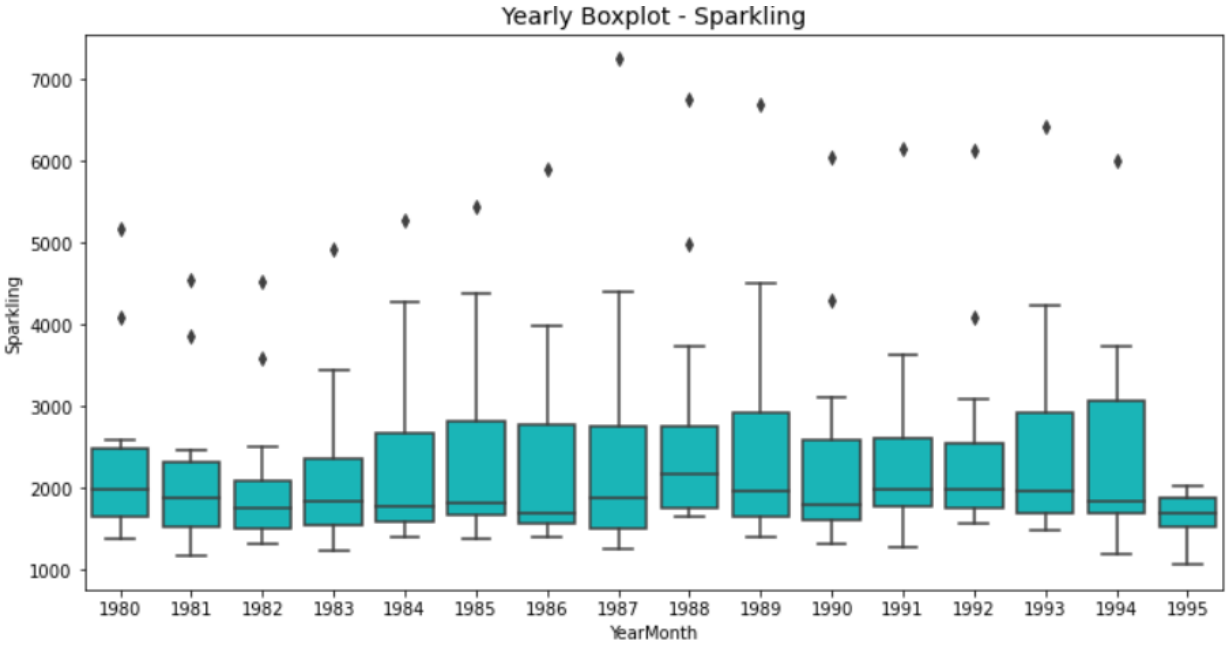


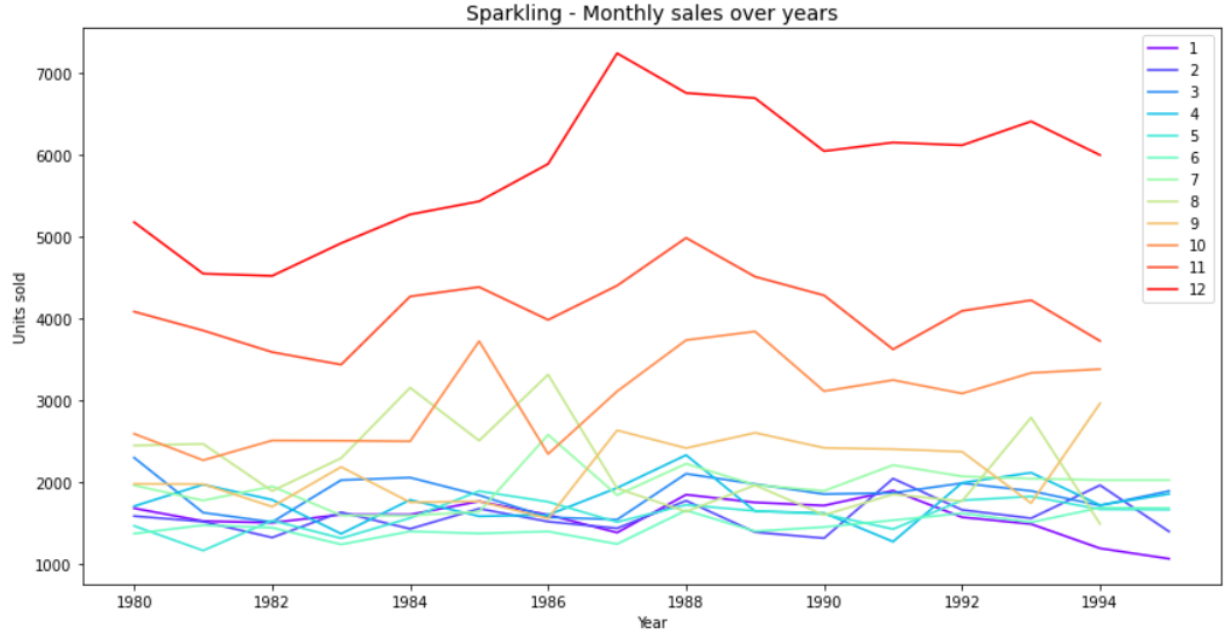


2. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

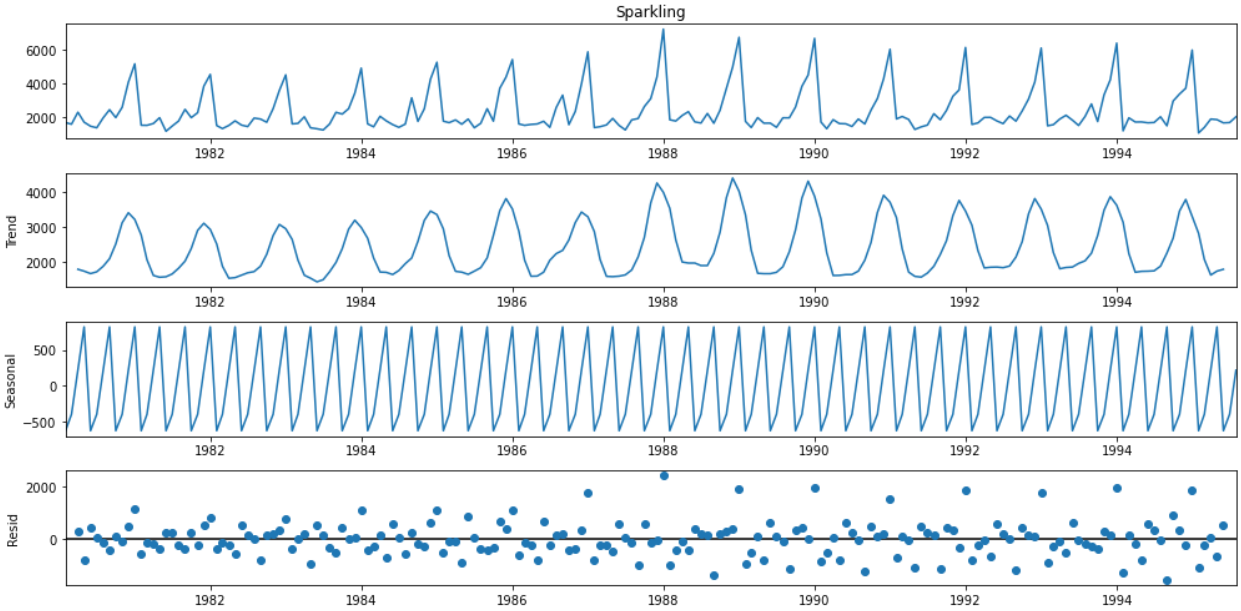
Ans 2.

Sparking -

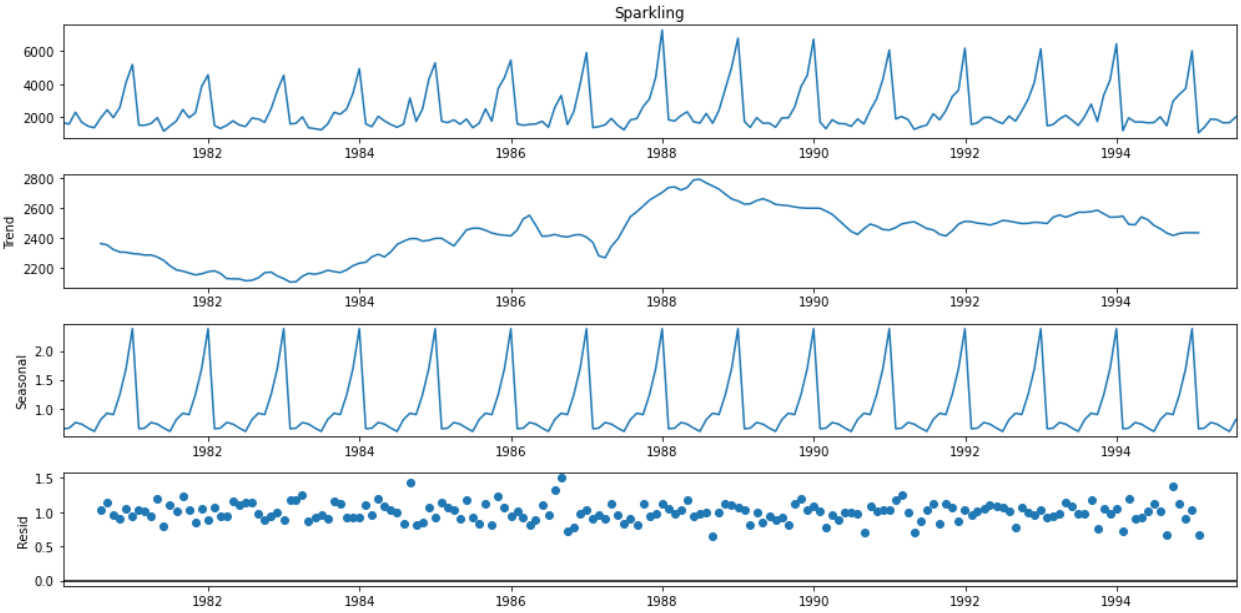




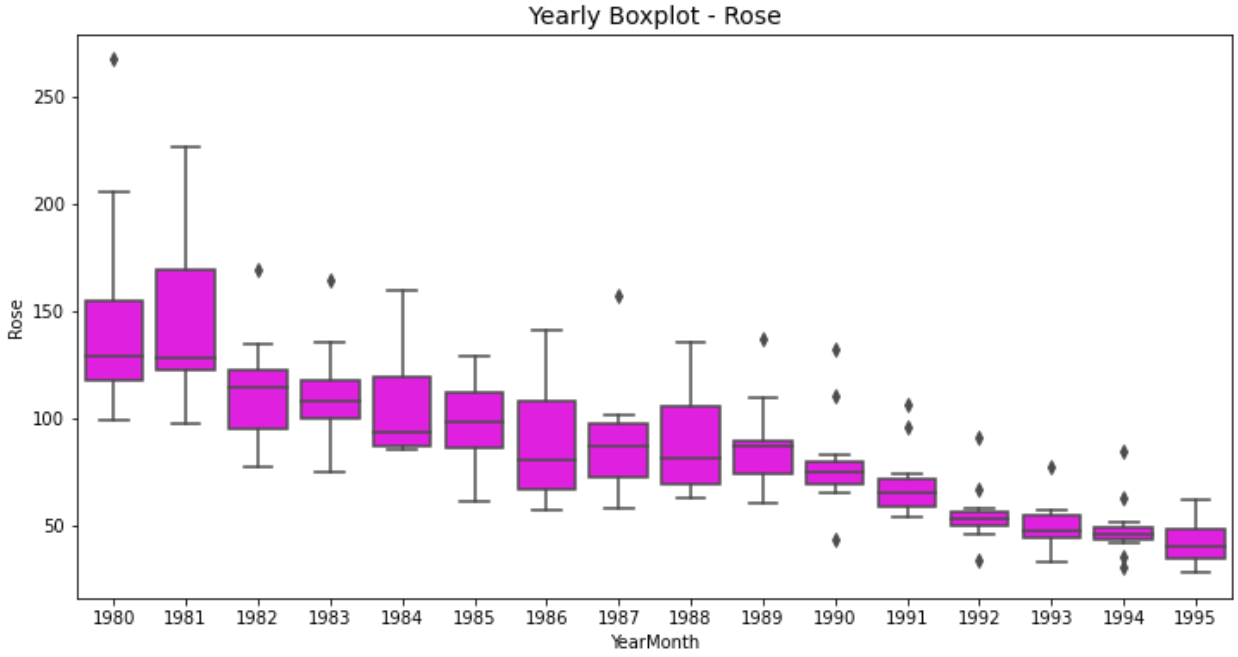
Additive Decomposition:

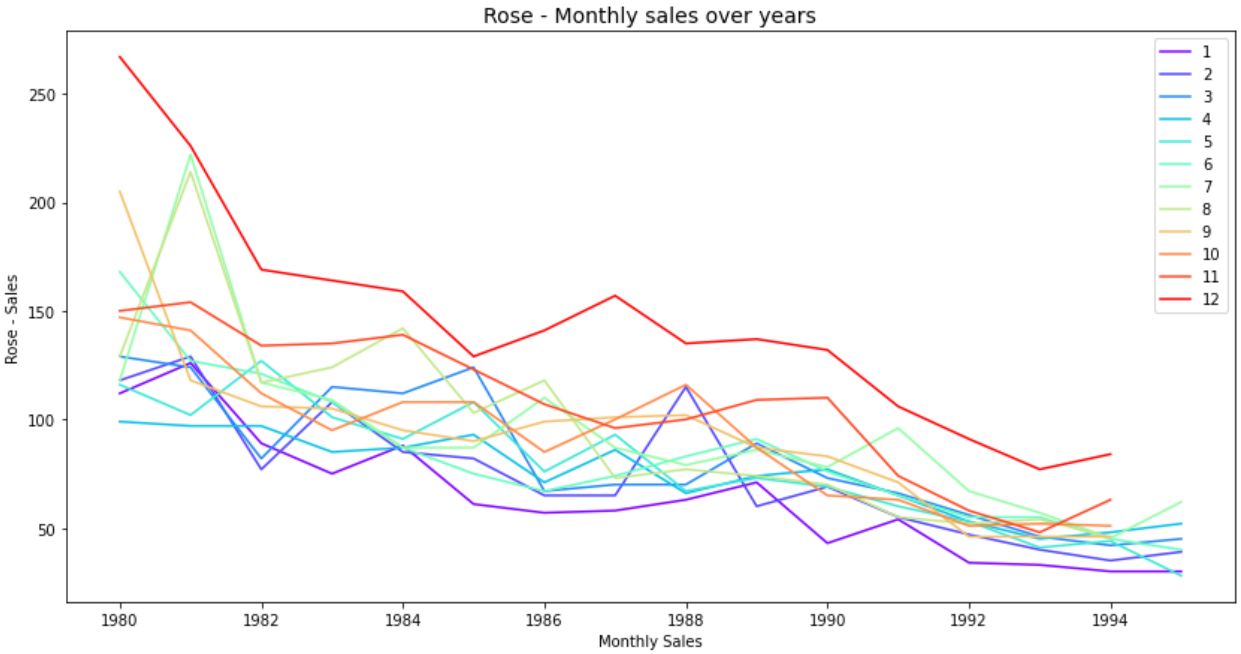


Multiplicative Decomposition:

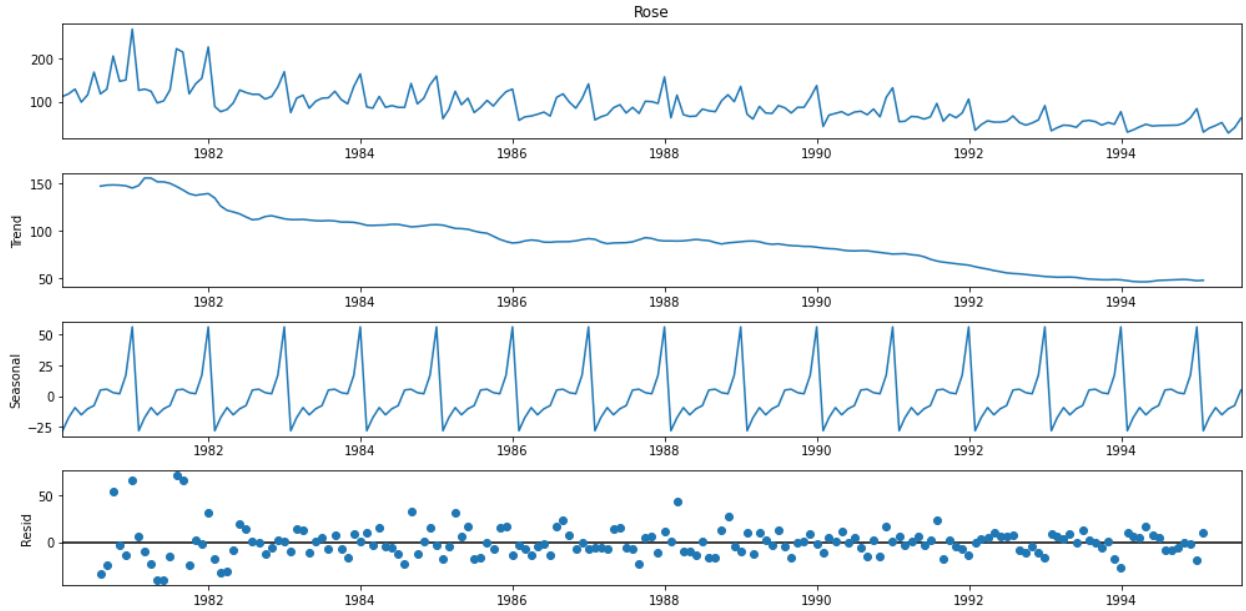


Rose -

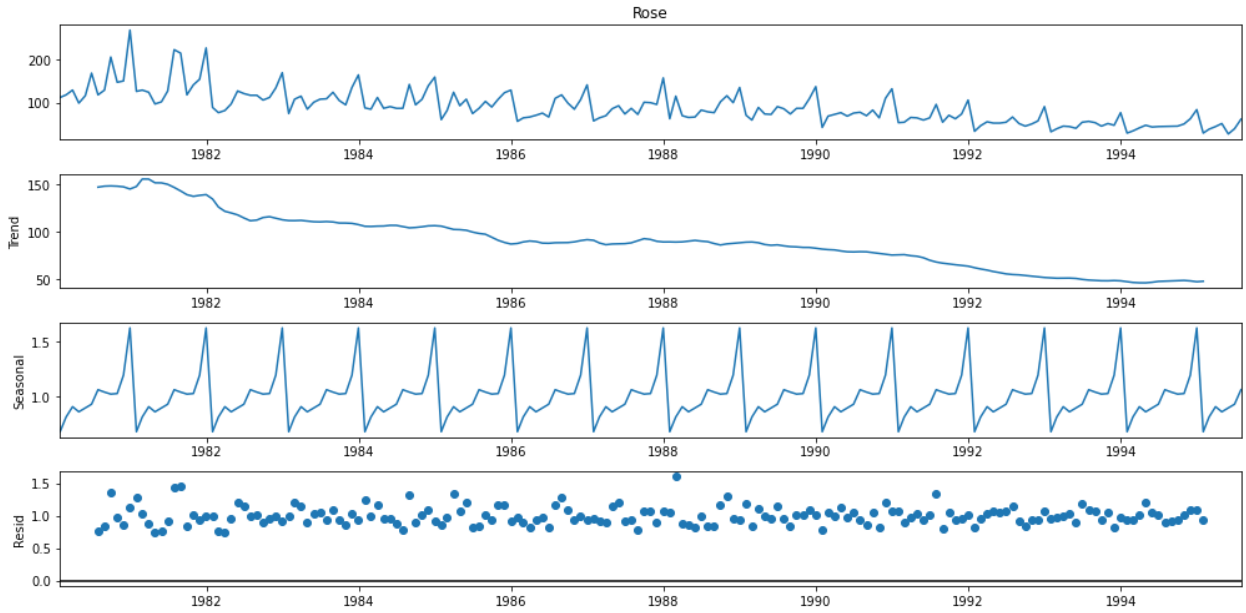

Yearly Boxplot - Rose
Rose
250
200
150
100
50
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
YearMonth


Rose - Monthly sales over years
Rose - Sales
250
200
150
100
50
1980
1982
1984
1986
1988
1990
1992
1994
Monthly Sales
1
2
3
4
5
6
7
8
9
10
11
12

Additive Decomposition:



Multiplicative Decomposition:



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

Ans 3.

The data is split using the comparison operator into train and test.

4. 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.

Ans 4.

Model Name	RMSE on Test Data - Sparkling	RMSE on Test Data - Rose
Linear Regression	1389.135	15.26
Naive Model	3864.27	79.7
Simple Average	1275.08	53.4
2 point TMA	813.4	11.52
Simple Exponential Smoothing	1316.41	36.82
Double Exponential Smoothing	1779.42	37.05
Triple Exponential Smoothing	314.8	9.1716

5. 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.

Ans 5.

Dickey-Fuller Test:

	Sparkling	Rose
Original series	Results of Dickey-Fuller Test: Test Statistic -1.360497 p-value 0.601061 #Lags Used 11.000000 Number of Observations Used 175.000000 Critical Value (1%) -3.468280 Critical Value (5%) -2.878202 Critical Value (10%) -2.575653 dtype: float64	Results of Dickey-Fuller Test: Test Statistic -1.876719 p-value 0.343091 #Lags Used 13.000000 Number of Observations Used 173.000000 Critical Value (1%) -3.468726 Critical Value (5%) -2.878396 Critical Value (10%) -2.575756
	We see that at 5% significant level the Time Series is non-stationary. But the seasonality is multiplicative as the Std deviation and mean varies according to the change in trend	**We see that at 5% significant level the Time Series is non-stationary.**
Difference of original series	Results of Dickey-Fuller Test: Test Statistic -4.460165 p-value 0.000232 #Lags Used 11.000000 Number of Observations Used 163.000000 Critical Value (1%) -3.471119 Critical Value (5%) -2.879441 Critical Value (10%) -2.576314	Results of Dickey-Fuller Test: Test Statistic -8.044395e+00 p-value 1.810868e-12 #Lags Used 1.200000e+01 Number of Observations Used 1.730000e+02 Critical Value (1%) -3.468726e+00 Critical Value (5%) -2.878396e+00 Critical Value (10%) -2.575756e+00
	We see that at α = 0.05 the Time Series is indeed stationary. But seasonality is multiplicative	**At difference of order 1, Rose Time Series is stationary with no trend**
Log of series	Results of Dickey-Fuller Test: Test Statistic -1.749630 p-value 0.405740 #Lags Used 11.000000 Number of Observations Used 175.000000 Critical Value (1%) -3.468280 Critical Value (5%) -2.878202 Critical Value (10%) -2.575653	Results of Dickey-Fuller Test: Test Statistic -4.605732 p-value 0.000126 #Lags Used 11.000000 Number of Observations Used 162.000000 Critical Value (1%) -3.471374 Critical Value (5%) -2.879552 Critical Value (10%) -2.576373

	Seasonality is now additive but non stationary	
Difference of log of series	Results of Dickey-Fuller Test: Test Statistic -5.183811 p-value 0.000009 #Lags Used 11.000000 Number of Observations Used 163.000000 Critical Value (1%) -3.471119 Critical Value (5%) -2.879441 Critical Value (10%) -2.576314	Results of Dickey-Fuller Test: Test Statistic -8.669696e+00 p-value 4.581847e-14 #Lags Used 1.100000e+01 Number of Observations Used 1.740000e+02 Critical Value (1%) -3.468502e+00 Critical Value (5%) -2.878298e+00 Critical Value (10%) -2.575704e+00

6. 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.

Ans 6.

Automated Version has been created as per the question

7. 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.

Ans 7.

ARIMA/SARIMA models have been created as per the question

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

Ans 8.

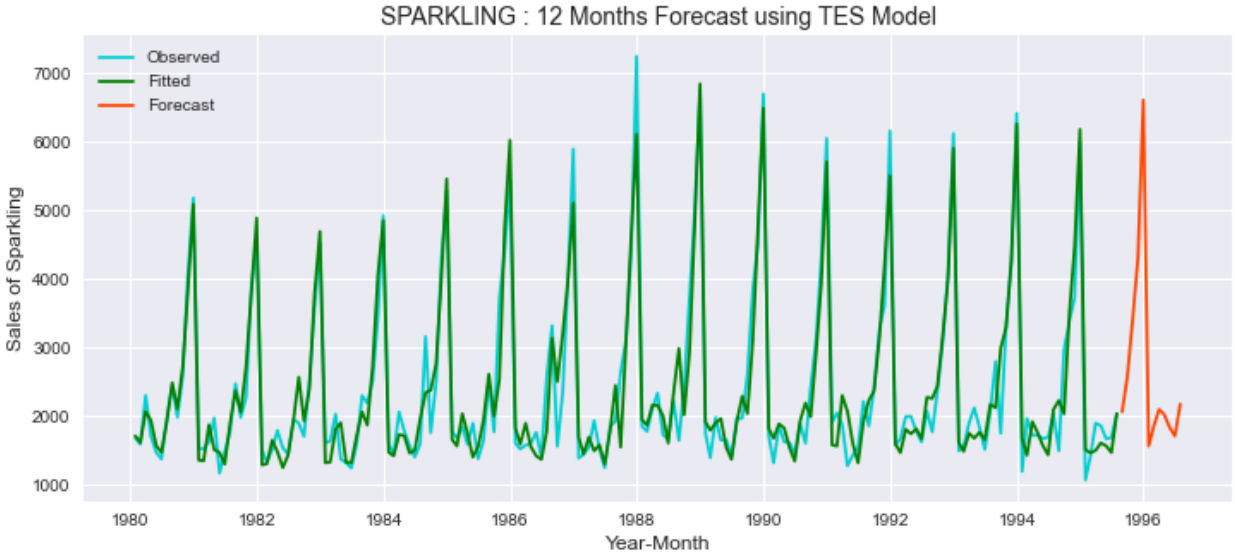
Model Name	RMSE on Test Data - Sparkling	RMSE on Test Data - Rose
Linear Regression	1389.135	15.26
Naive Model	3864.27	79.7
Simple Average	1275.08	53.4
2 point TMA	813.4	11.52
Simple Exponential Smoothing	1316.41	36.82
Double Exponential Smoothing	1779.42	37.05

Triple Exponential Smoothing	314.8	9.1716
Auto SARIMA	331.845160	16.823573

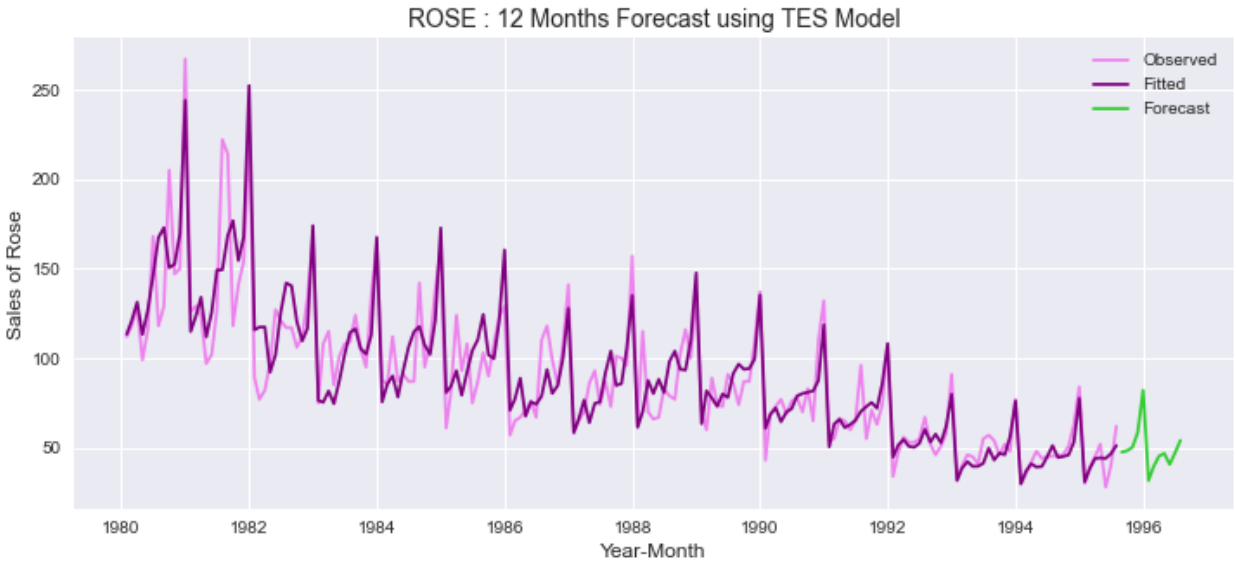
9. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

Ans 9.

Sparkling - Using Triple Exponential Smoothing Model -



Rose - Using Triple Exponential Smoothing Model -



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

Ans 10.

- The company can use the Triple Exponential Smoothing Model model for forecasting data as it gives the best results on Sparkling and Rose.
- There is missing data in the dataset which should be addressed immediately.
- To get more accurate forecasting results the company must provide more data and at a greater granularity.
- The later months of the year usually have more sales for both Sparkling and Rose. The company can focus on new advertisement ideas or different product verticals to capture this market.
- The overall sales for Sparkling have remained more or less similar but the sales for Rose have reduced. The company should advertise the rose Wine more.