DSBA

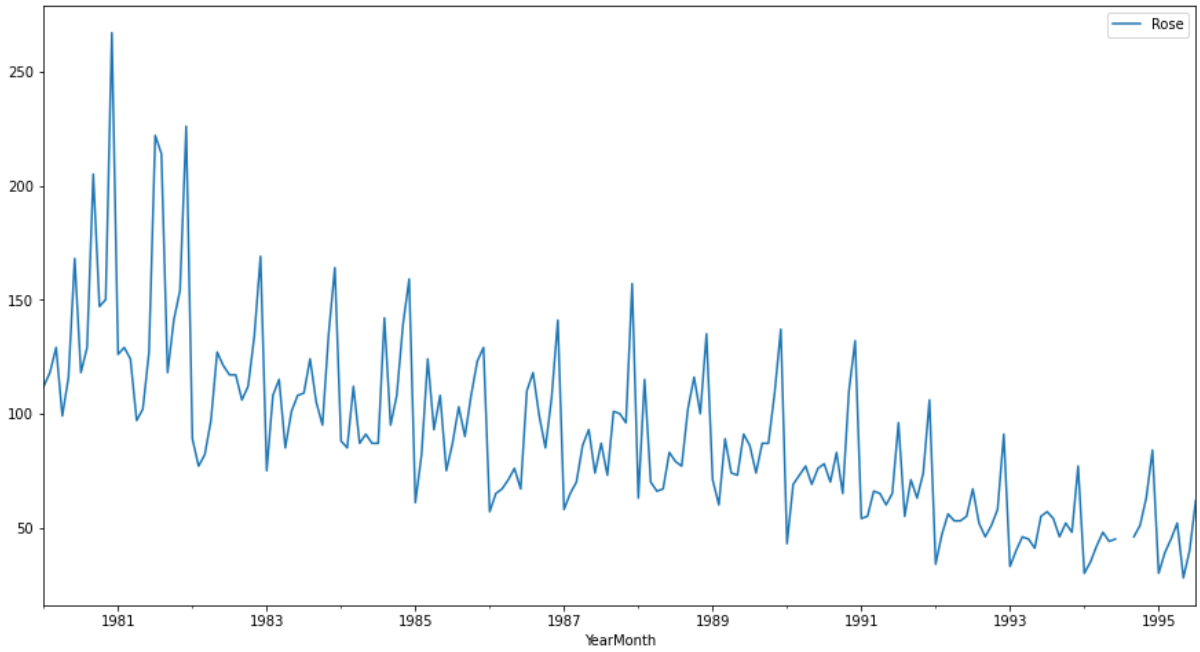
TIME SERIES FORCASTING PROJECT

Done By:

Hariharan Manickam

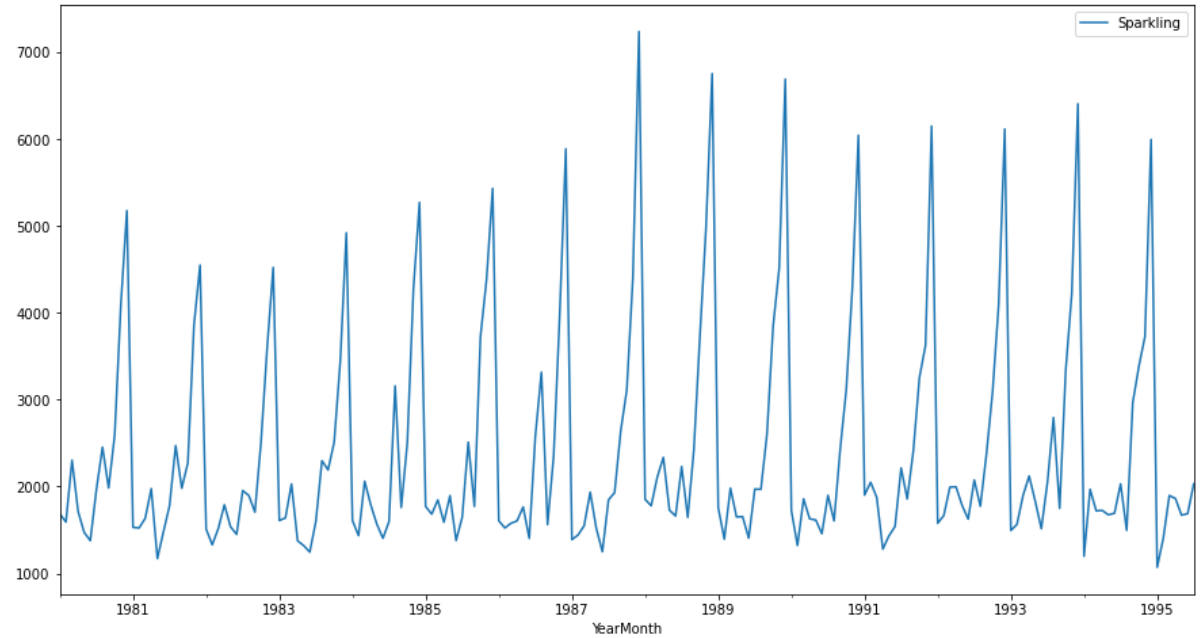
1. **Read the data as an appropriate Time Series data and plot the data.**

Rose Wine



We can observe that trend is downward with seasonality.

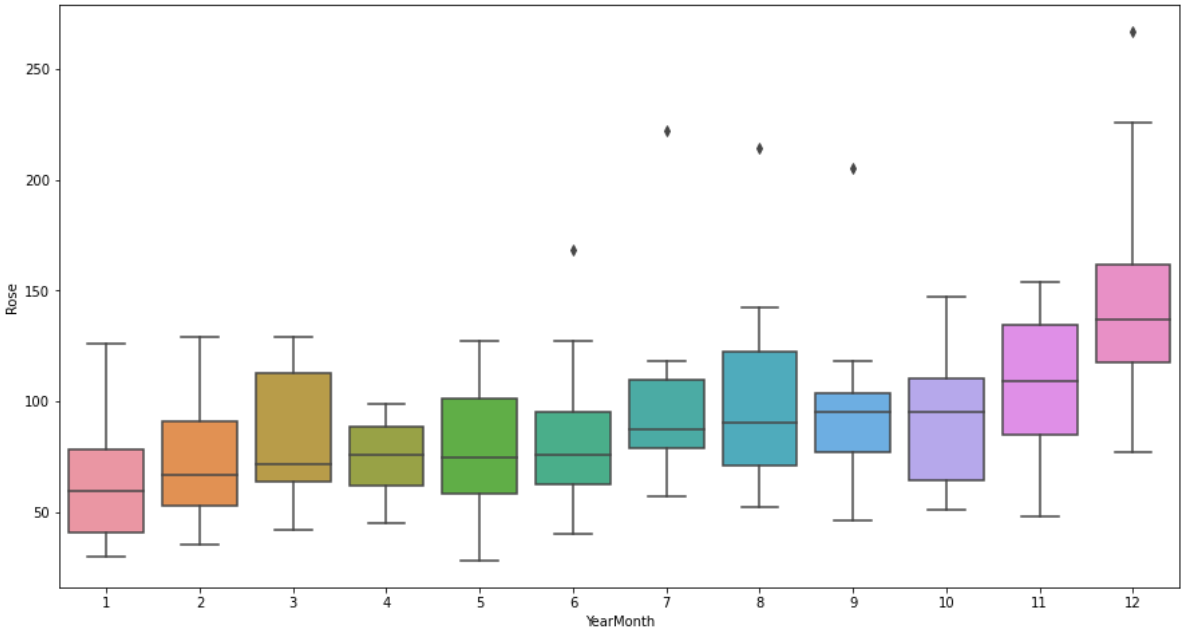
Sparkling Wine



We can observe there is no trend with a strong seasonality component.

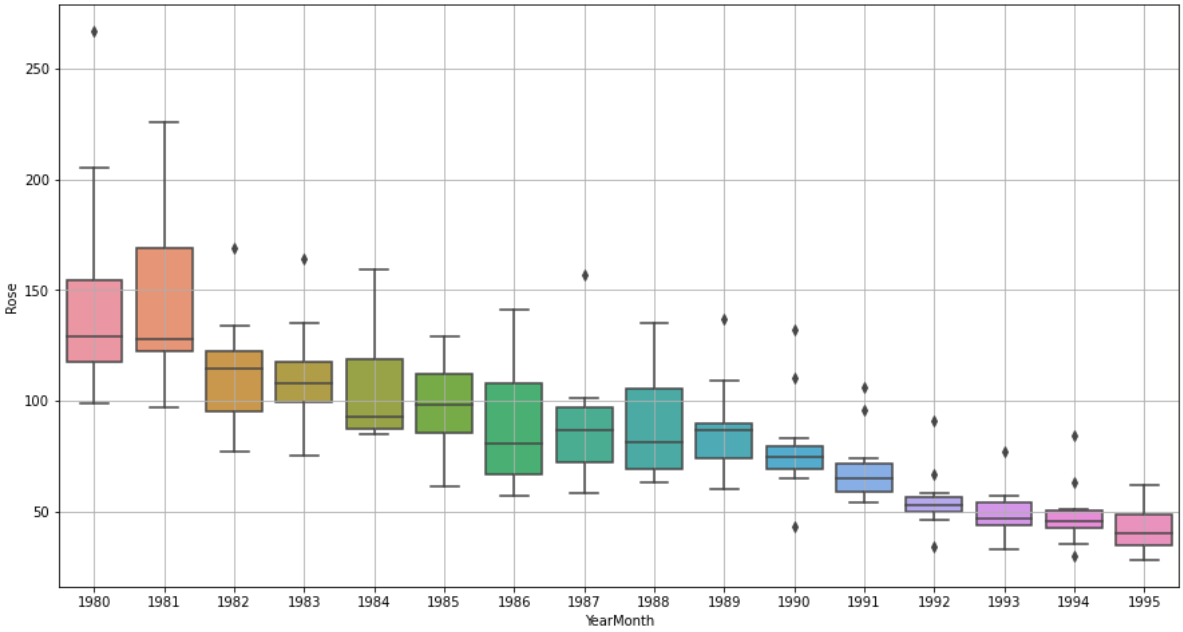
2. **Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.**

Rose wines (monthly)



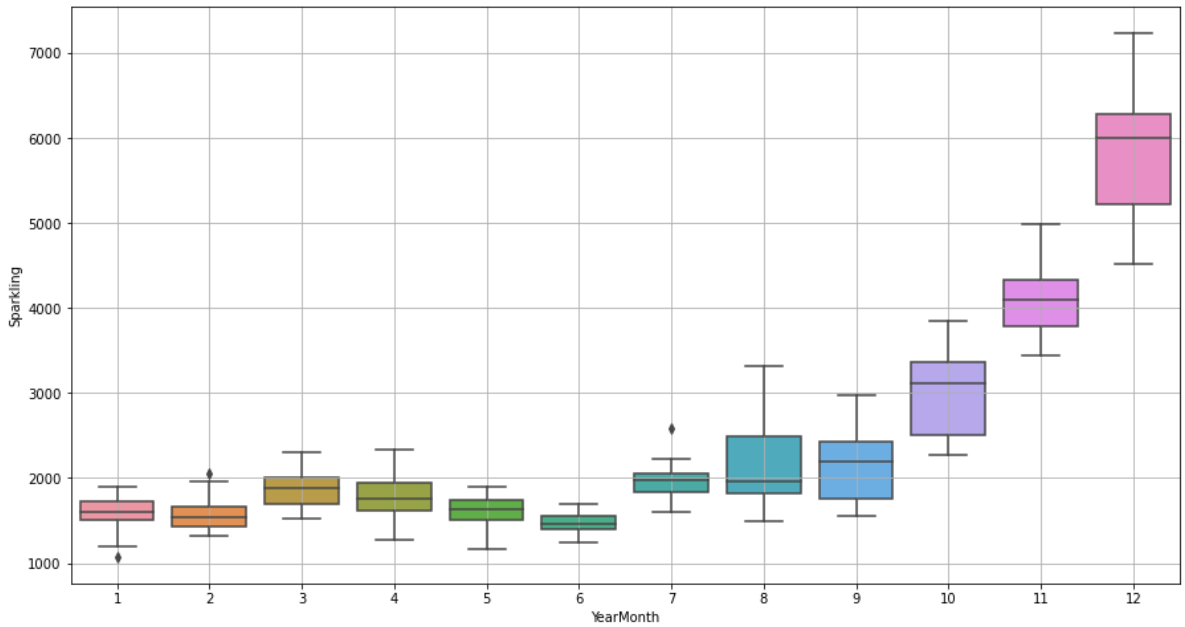
December has highest sales data. Average sales seems to be around 90.

Rose wines (yearly)



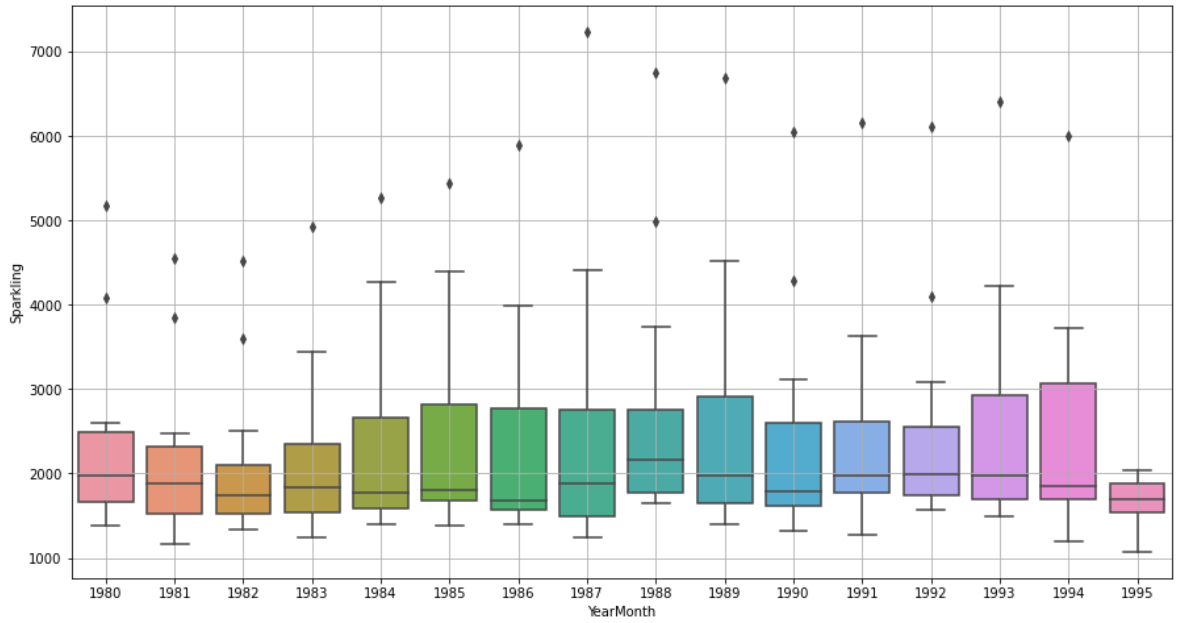
We can observe the downward trend again.

Sparkling wine (monthly)



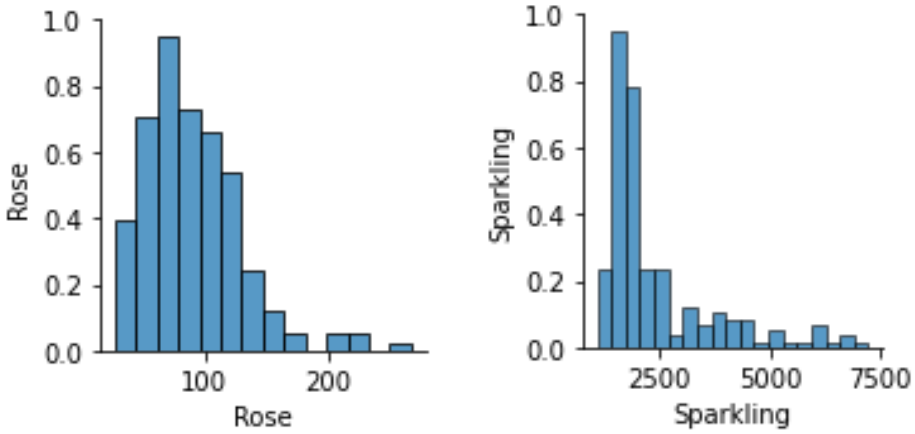
We can observe that December has the highest sales.

Sparkling wine (yearly)

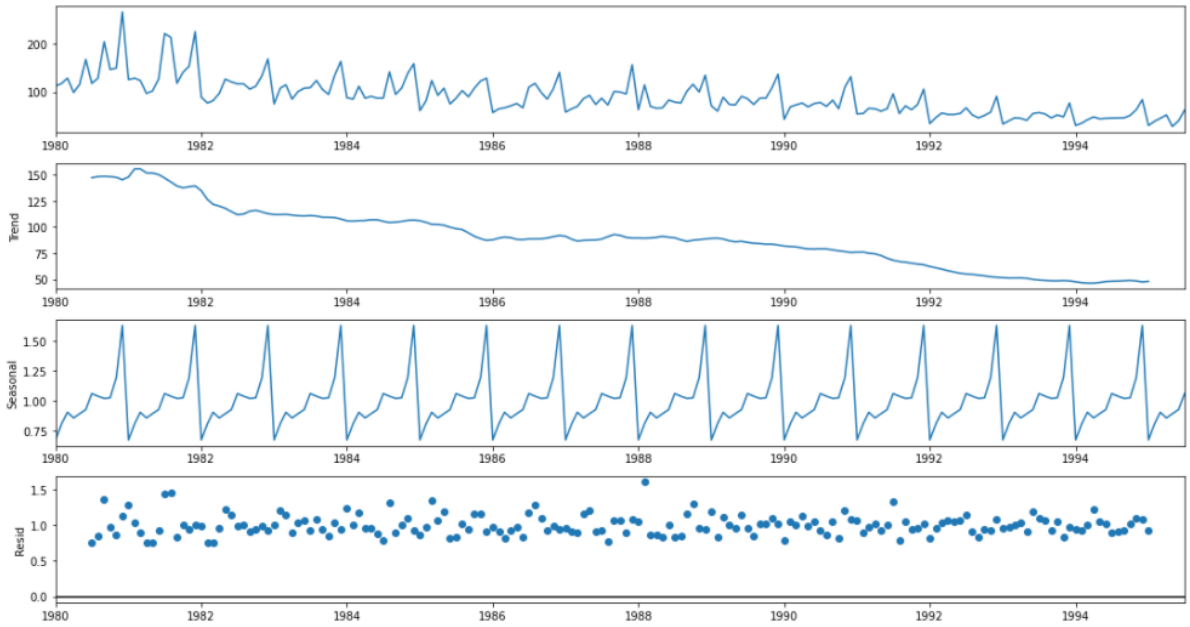


We can observe that average sales per year is around 2000.

Bivariate Analysis

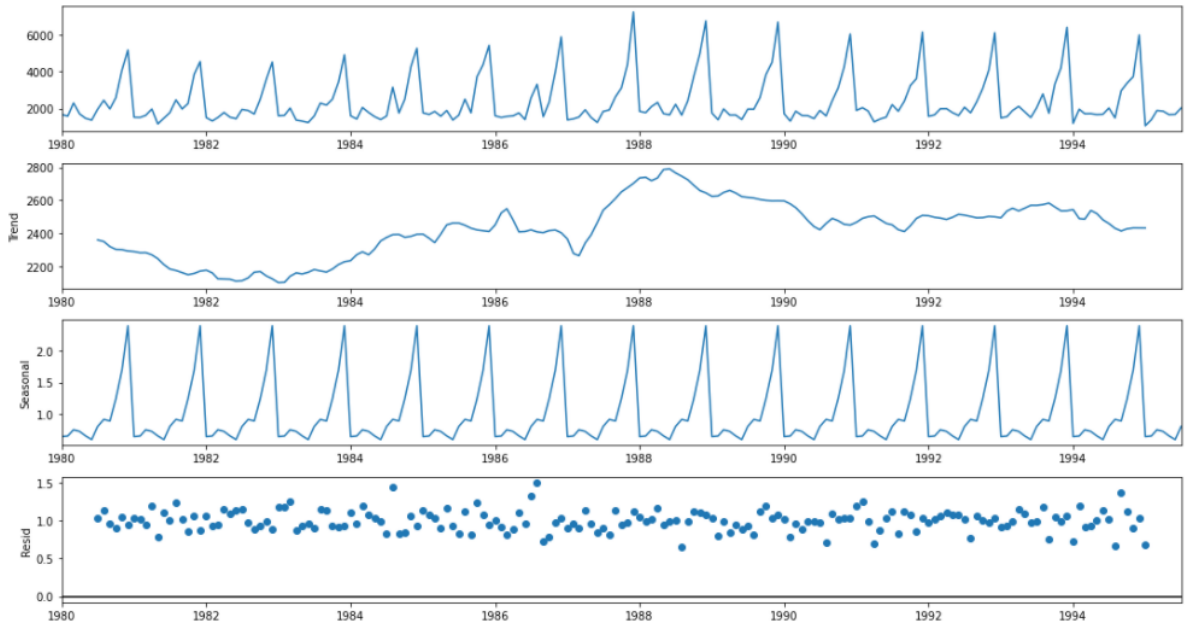


Rose Wine decomposed (multiplicative)



A downward trend can be observed in number of sales in Rose wine. Seasonality also observed in the dataset. Hence, Series has both trend and seasonality.

Sparkling wine (multiplicative)



An upward trend can be observed in number of sales in Sparkling wine. Seasonality also observed in the dataset. Hence, Series has both trend and seasonality.

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

First 2 columns are training data, next 2 are testing data.

YearMonth	Rose
1980-01-01	112.0
1980-02-01	118.0
1980-03-01	129.0
1980-04-01	99.0
1980-05-01	116.0

YearMonth	Sparkling
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471

YearMonth	Rose
1990-08-01	70.0
1990-09-01	83.0
1990-10-01	65.0
1990-11-01	110.0
1990-12-01	132.0

YearMonth	Sparkling
1990-08-01	1605
1990-09-01	2424
1990-10-01	3116
1990-11-01	4286
1990-12-01	6047

YearMonth	Rose
1991-01-01	54.0
1991-02-01	55.0
1991-03-01	66.0
1991-04-01	65.0
1991-05-01	60.0

YearMonth	Sparkling
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471

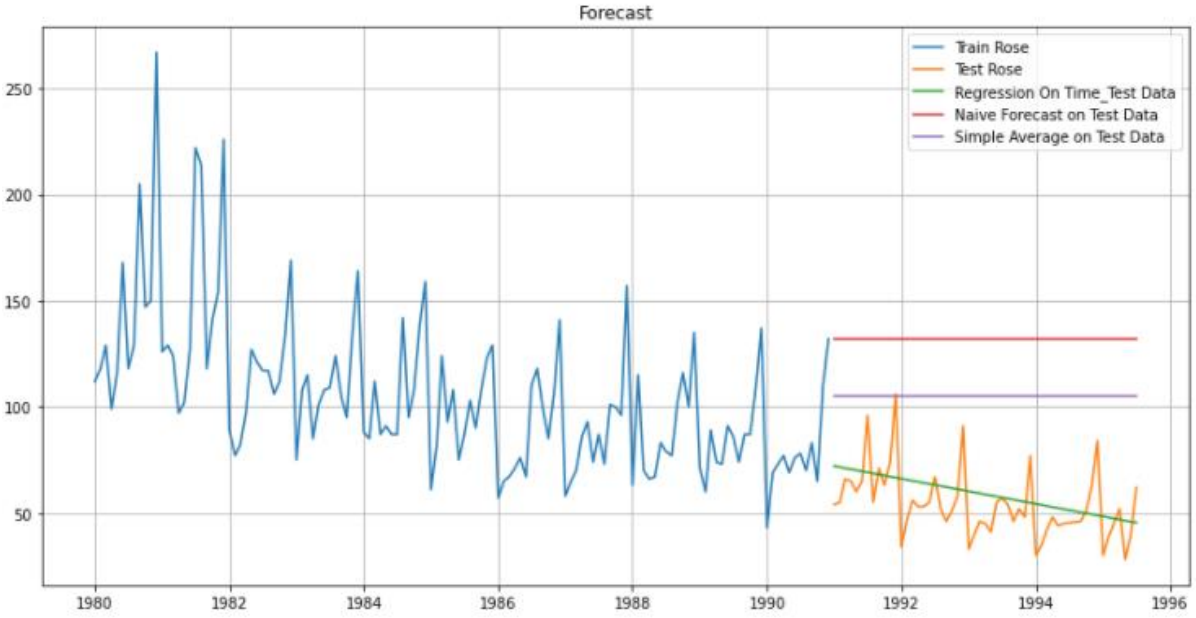
YearMonth	Rose
1995-03-01	45.0
1995-04-01	52.0
1995-05-01	28.0
1995-06-01	40.0
1995-07-01	62.0

YearMonth	Sparkling
1990-08-01	1605
1990-09-01	2424
1990-10-01	3116
1990-11-01	4286
1990-12-01	6047

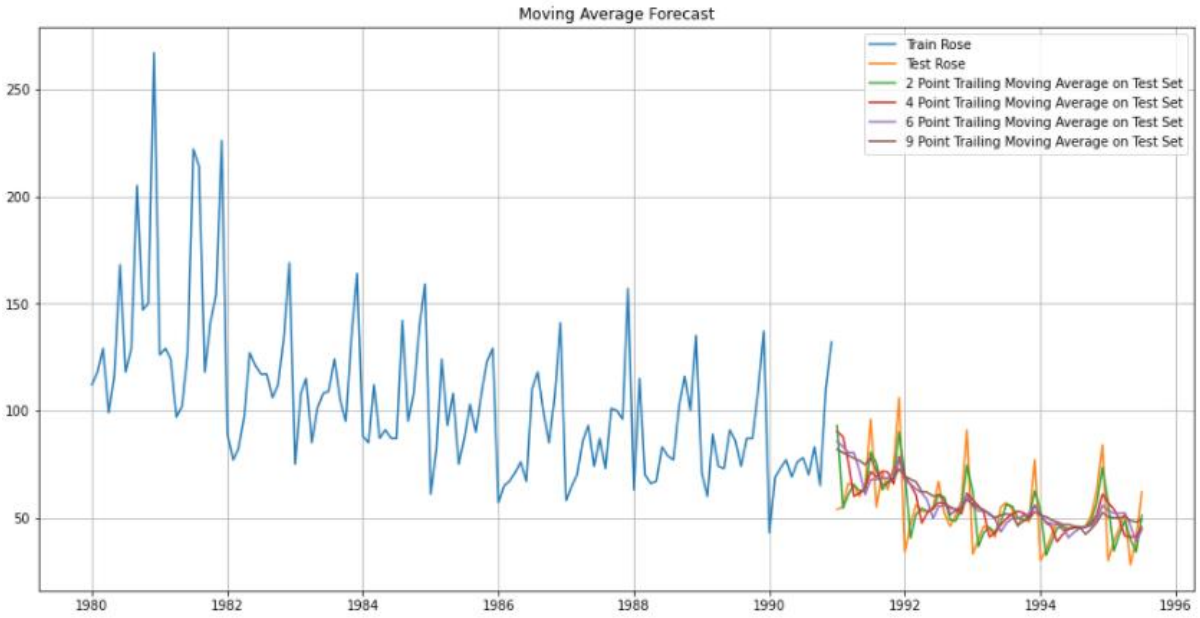
4. **Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data.**
   Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

<u>Rose Wine</u>

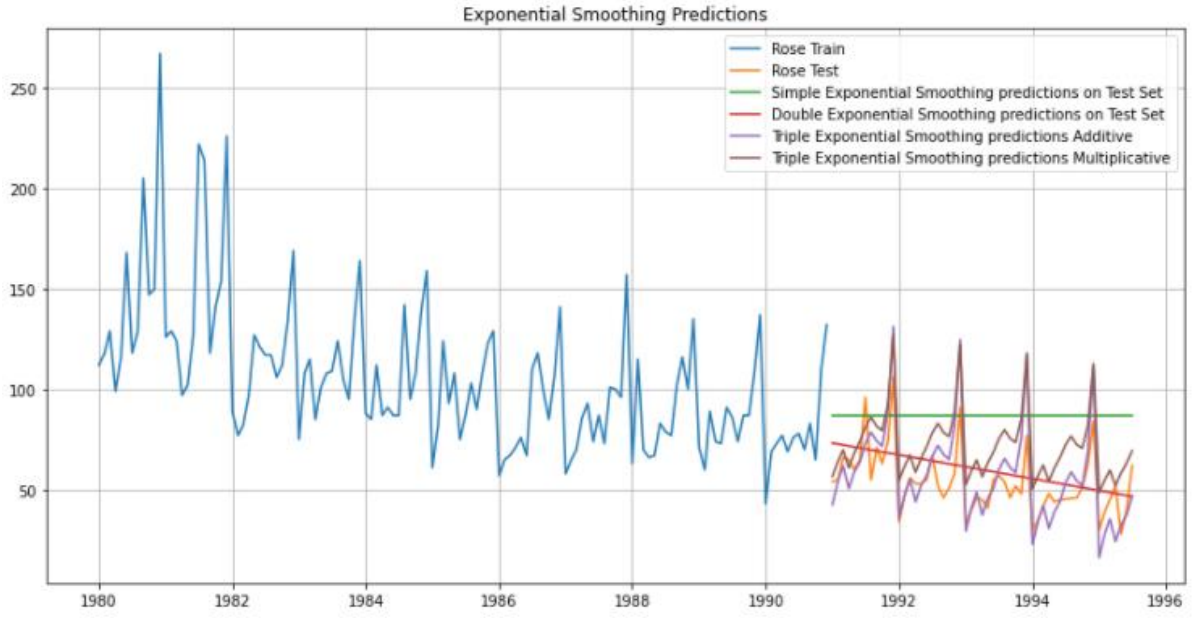
Regression, Naïve & Simple Average



Moving Average



Exponential Smoothing

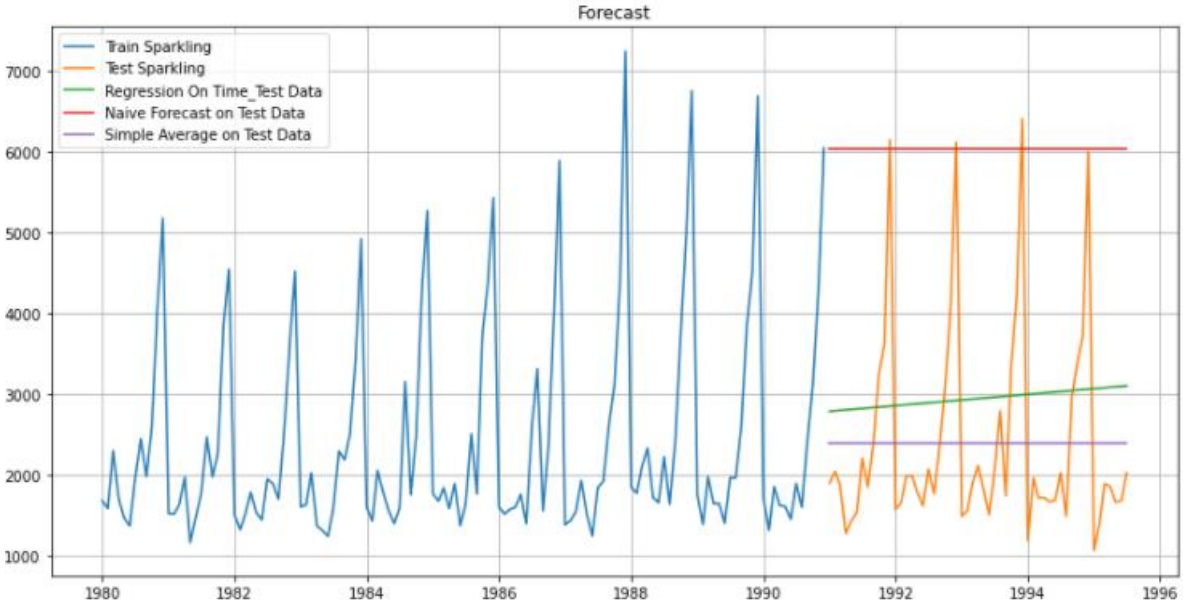


RMSE for Rose Wines

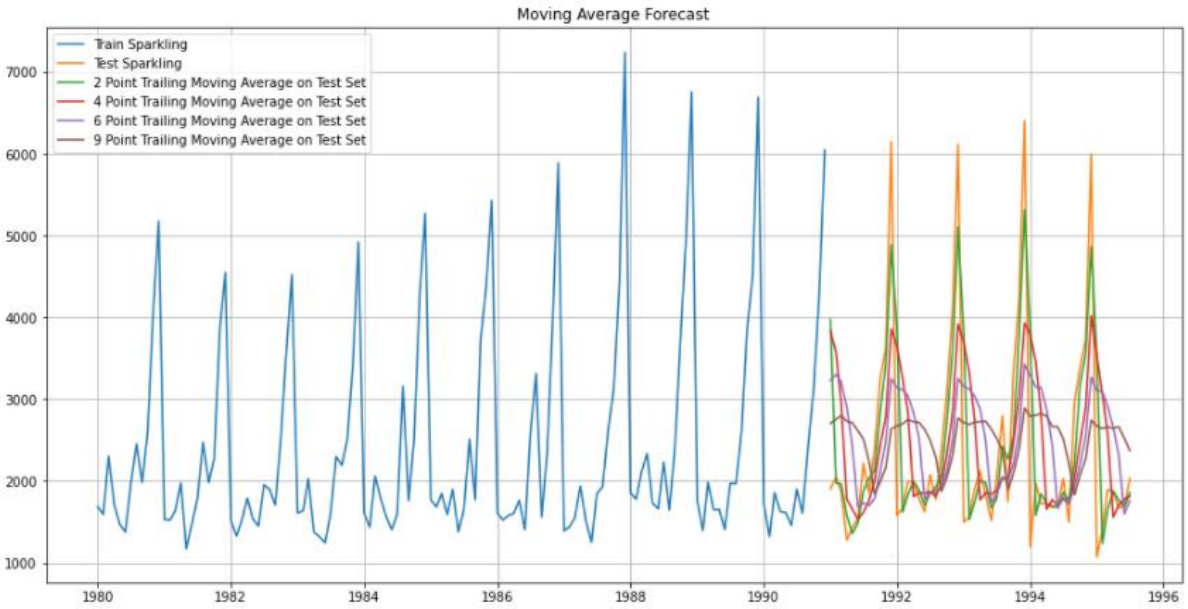
	RMSE
Linear Regression	15.268955
Naive Model	79.718773
Simple Average	53.460570
2 Point Moving Average	11.529278
4 Point Moving Average	14.451403
6 Point Moving Average	14.566327
9 Point Moving Average	14.727630
Simple Exponential, Alpha=0.05	36.796227
Double Exponential, Alpha=0.05	24.998640
Triple Exponential, Additive	14.257122
Triple Exponential, Multiplicative	21.019620

Sparkling Wine

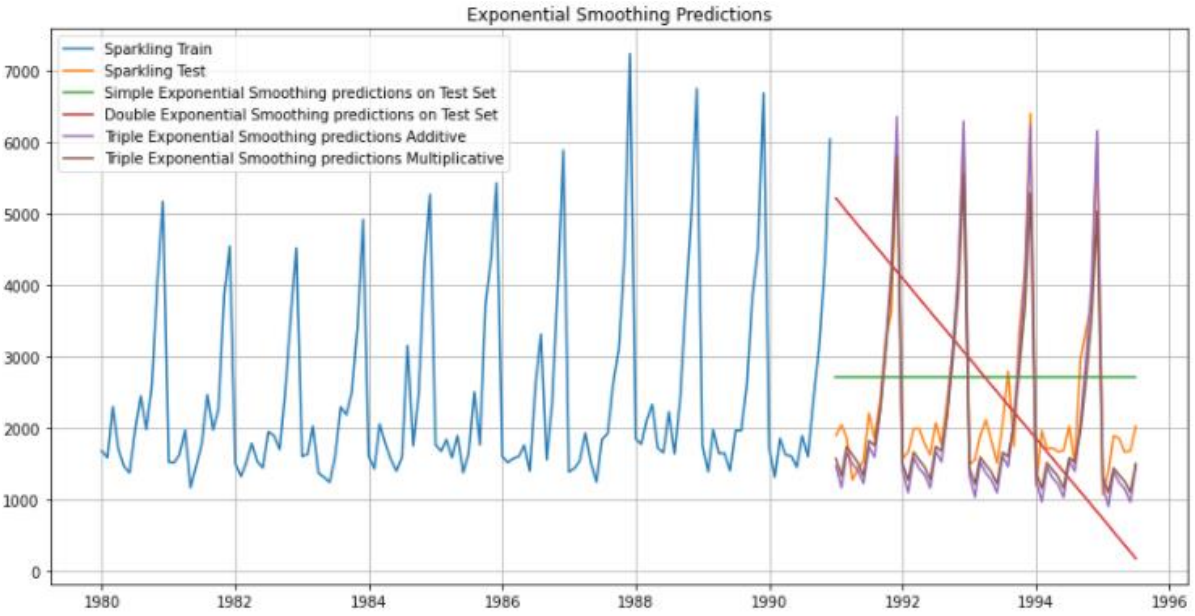
Regression, Naïve & Simple Average



Moving Average



Exponential Smoothing



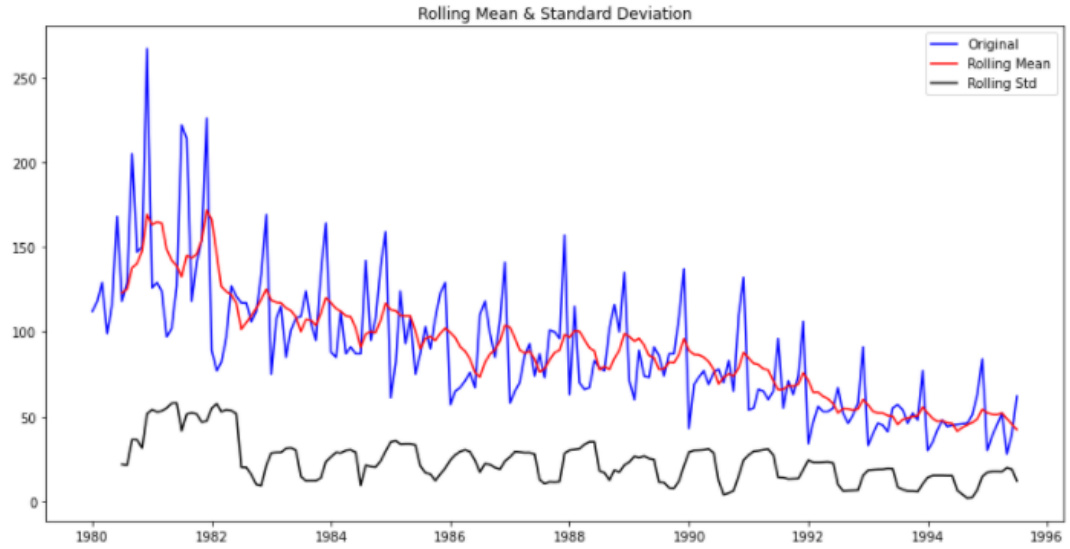
RMSE for Sparkling Wine

	RMSE
Linear Regression	1389.135175
Naive Model	1389.135175
Simple Average	1275.081804
2 Pt Moving Average	813.400684
4 Pt Moving Average	1156.589694
6 Pt Moving Average	1283.927428
9 Pt Moving Average	1346.278315
Simple Exponential, Alpha=0.05	1316.035487
Double Exponential, Alpha=0.05	3096.812583
Triple Exponential, Additive	473.152417
Triple Exponential, Multiplicative	469.767970

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.

ADF Hypothesis for Rose wines

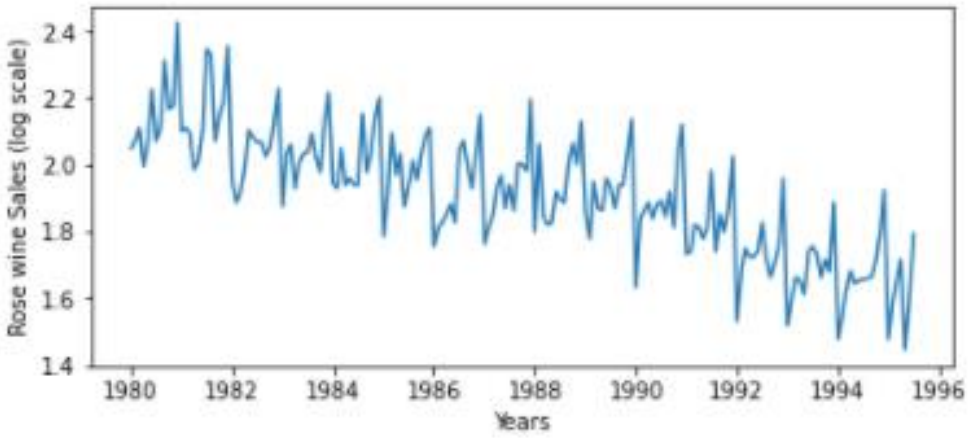
- H0 : The time series is not stationary
- H1 : The time series is Stationary

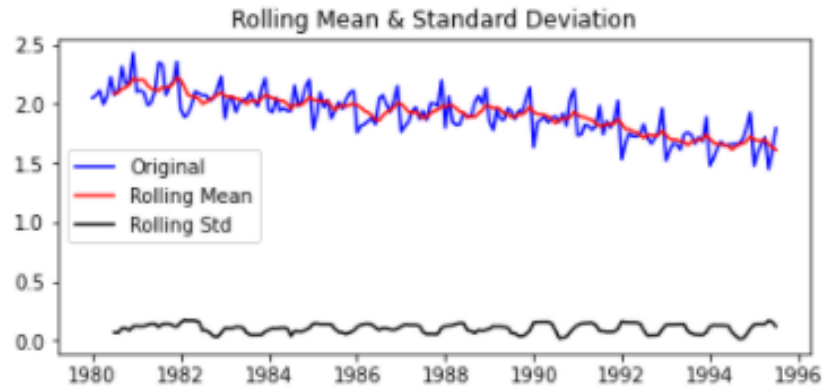



```
Results of Dickey-Fuller Test:
Test Statistic                  -1.727653
p-value                          0.416924
#Lags Used                      12.000000
Number of Observations Used    174.000000
Critical Value (1%)             -3.468502
Critical Value (5%)             -2.878298
Critical Value (10%)            -2.575704
dtype: float64
```

- P value is greater than 0.05 so cannot reject the null hypothesis

We repeat the test again after converting data to log scale

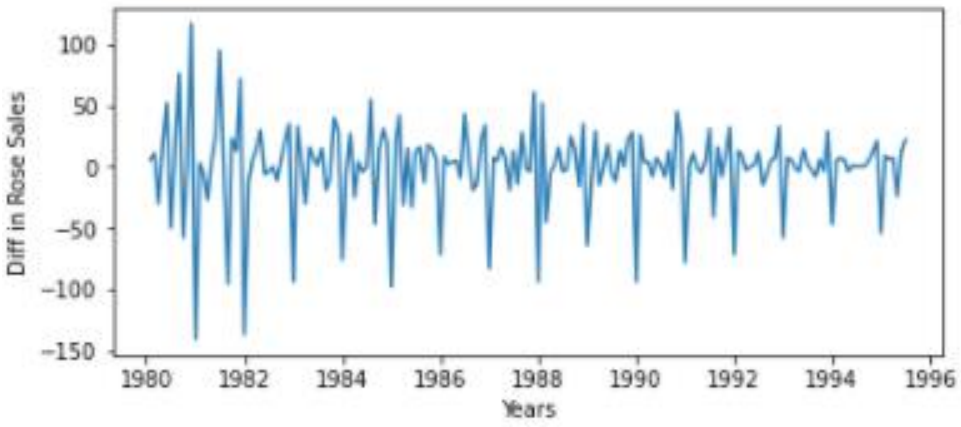


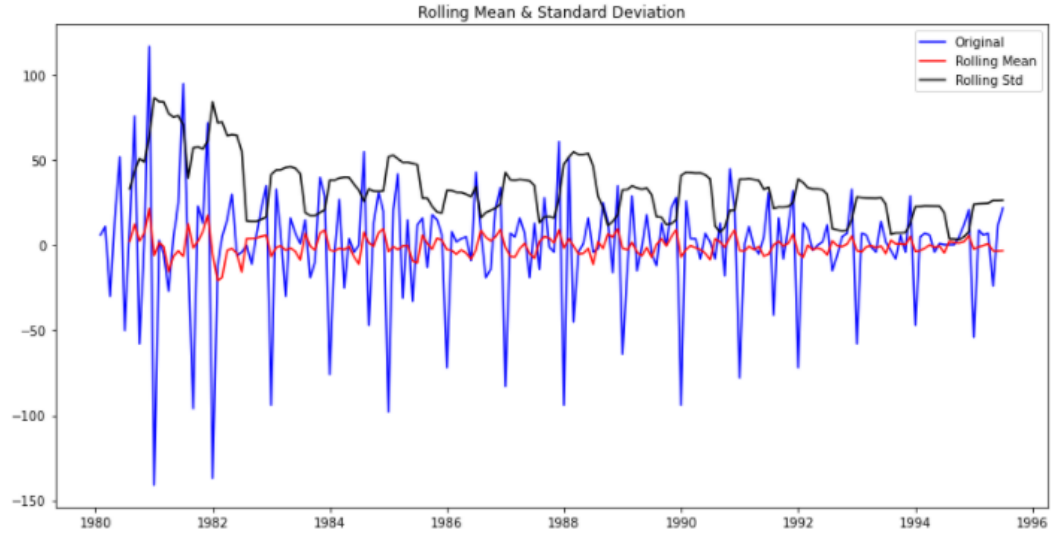



```
Results of Dickey-Fuller Test:
Test Statistic                  -0.422521
p-value                          0.906270
#Lags Used                      12.000000
Number of Observations Used    174.000000
Critical Value (1%)             -3.468502
Critical Value (5%)             -2.878298
Critical Value (10%)            -2.575704
dtype: float64
```

- p value is greater than 0.05 so we cannot reject the null hypothesis.

We now differentiate in the order of 1.



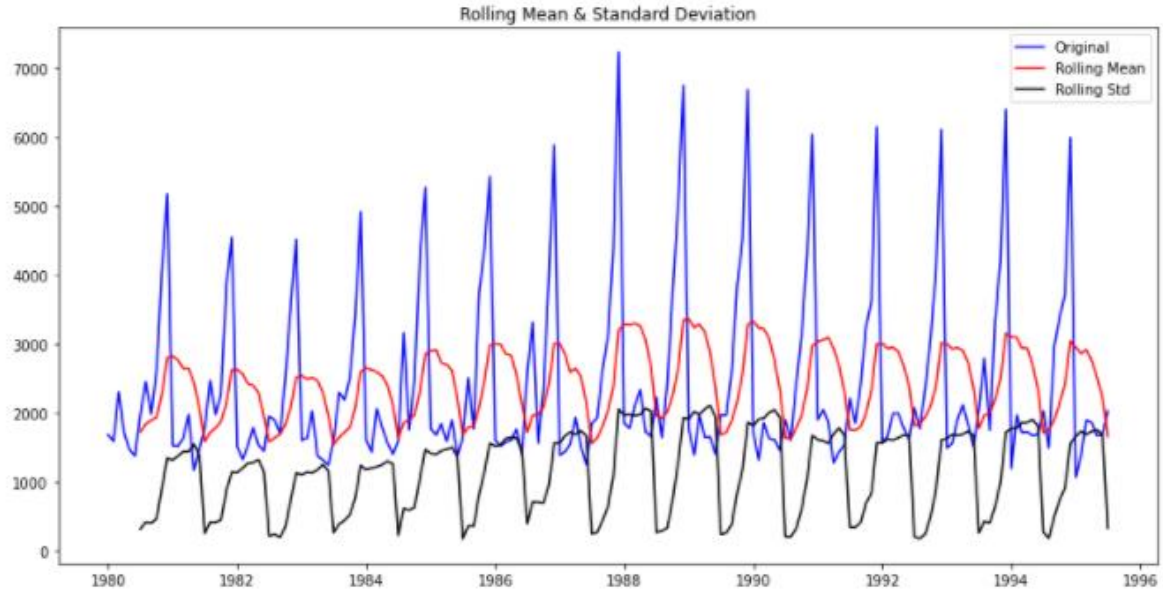



```
Results of Dickey-Fuller Test:
Test Statistic                -8.044392e+00
p-value                        1.810895e-12
#Lags Used                     1.200000e+01
Number of Observations Used    1.730000e+02
Critical Value (1%)           -3.468726e+00
Critical Value (5%)           -2.878396e+00
Critical Value (10%)          -2.575756e+00
dtype: float64
```

- P value is lower than 0.05, hence series has been converted to stationary

ADF Hypothesis for Sparkling wines

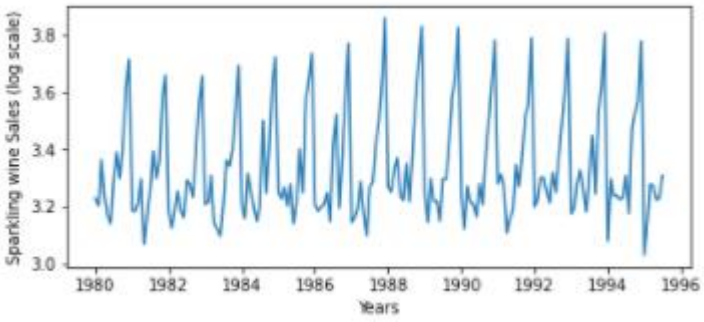
- H0 : The time series is not stationary
- H1 : The time series is Stationary

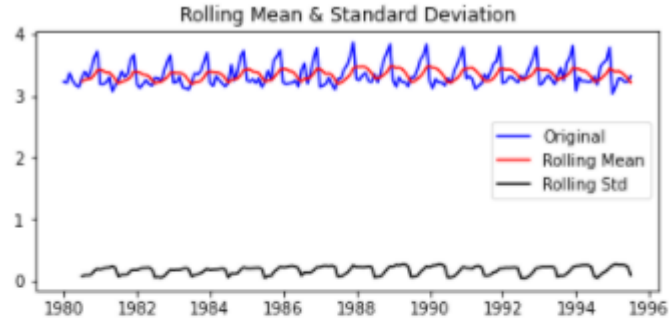



```
Results of Dickey-Fuller Test:
Test Statistic                  -1.497024
p-value                          0.534983
#Lags Used                      12.000000
Number of Observations Used    174.000000
Critical Value (1%)             -3.468502
Critical Value (5%)             -2.878298
Critical Value (10%)            -2.575704
dtype: float64
```

- P value is greater than 0.05 so cannot reject the null hypothesis

We repeat the test again after converting data to log scale

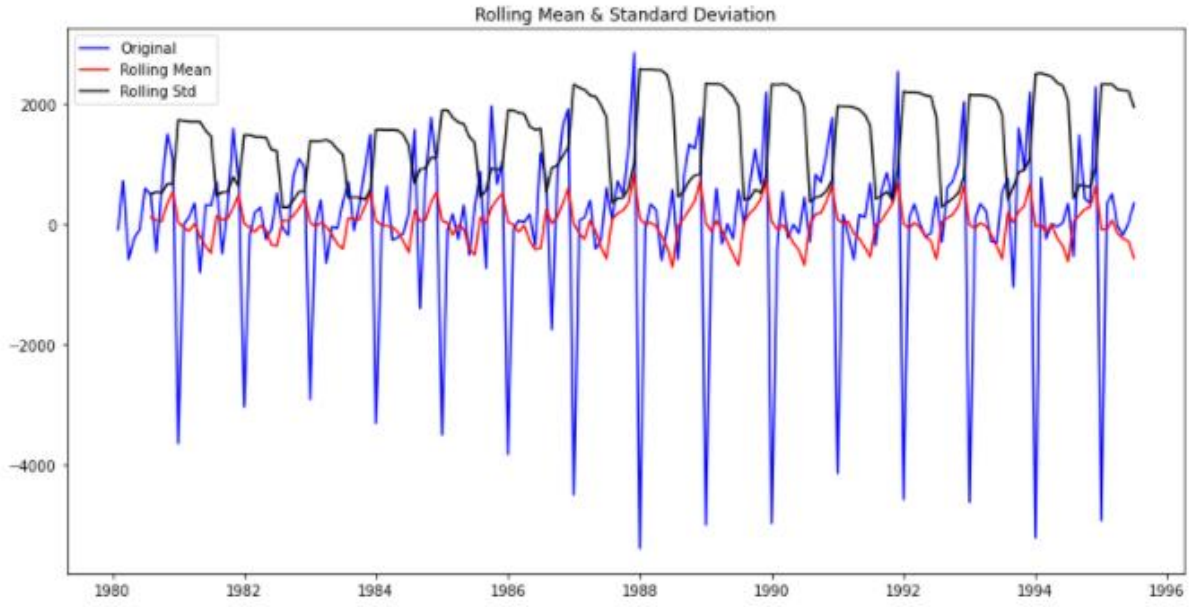





```
Results of Dickey-Fuller Test:
Test Statistic                  -1.859682
p-value                          0.351259
#Lags Used                      12.000000
Number of Observations Used    174.000000
Critical Value (1%)             -3.468502
Critical Value (5%)             -2.878298
Critical Value (10%)            -2.575704
dtype: float64
```

- p value is greater than 0.05 so we cannot reject the null hypothesis.

We now differentiate in the order of 1.




```
Results of Dickey-Fuller Test:
Test Statistic                -8.958891e+00
p-value                        8.328040e-15
#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
dtype: float64
```

- P value is lower than 0.05, hence series has been converted to stationary

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.**

Rose Wines

The following are the RMSE's for the ARIMA & SARIMA models with the lowest AIC scores.

	RMSE
ARIMA (0, 1, 2)	15.617834
SARIMA (1, 0, 0) (1, 0, 1, 12)	10.375552

Sparkling Wines

The following are the RMSE's for the ARIMA & SARIMA models with the lowest AIC scores.

	RMSE
ARIMA (2, 0, 2)	1005.083495
SARIMA(1,0,0)(1,0,1,12)	505.373773

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.**

Rose Wines

The following are the RMSE's for the ARIMA & SARIMA manual Cut Off models with the lowest AIC scores.

	RMSE
ARIMA (0, 1, 2)	15.617834
SARIMA (1, 0, 0) (1, 0, 1, 12)	10.375552
ARIMA (5, 1, 3) ACF/PACF Cutoff	15.633261
SARIMA (5, 1, 3) (1, 1, 0)12 ACF/PACF Cutoff	17.491317

Sparkling Wines

The following are the RMSE's for the ARIMA & SARIMA manual Cut Off models with the lowest AIC scores.

	RMSE
ARIMA (2, 0, 2)	1005.083495
SARIMA(1,0,0)(1,0,1,12)	505.373773
ARIMA (4, 1, 3) ACF/PACF Cutoff	1398.797957
SARIMA (1, 1, 3)(1, 1, 1)12 ACF/PACF Cutoff	648.039376

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

Rose Wines

	RMSE
Linear Regression	15.268955
Naive Model	79.718773
Simple Average	53.460570
2 Point Moving Average	11.529278
4 Point Moving Average	14.451403
6 Point Moving Average	14.566327
9 Point Moving Average	14.727630
Simple Exponential, Alpha=0.05	36.796227
Double Exponential, Alpha=0.05	24.998640
Triple Exponential, Additive	14.257122
Triple Exponential, Multiplicative	21.019620
ARIMA (0, 1, 2)	15.617834
SARIMA (1, 0, 0) (1, 0, 1, 12)	10.375552
ARIMA (5, 1, 3) ACF/PACF Cutoff	15.633261
SARIMA (5, 1, 3) (1, 1, 0)12 ACF/PACF Cutoff	17.491317

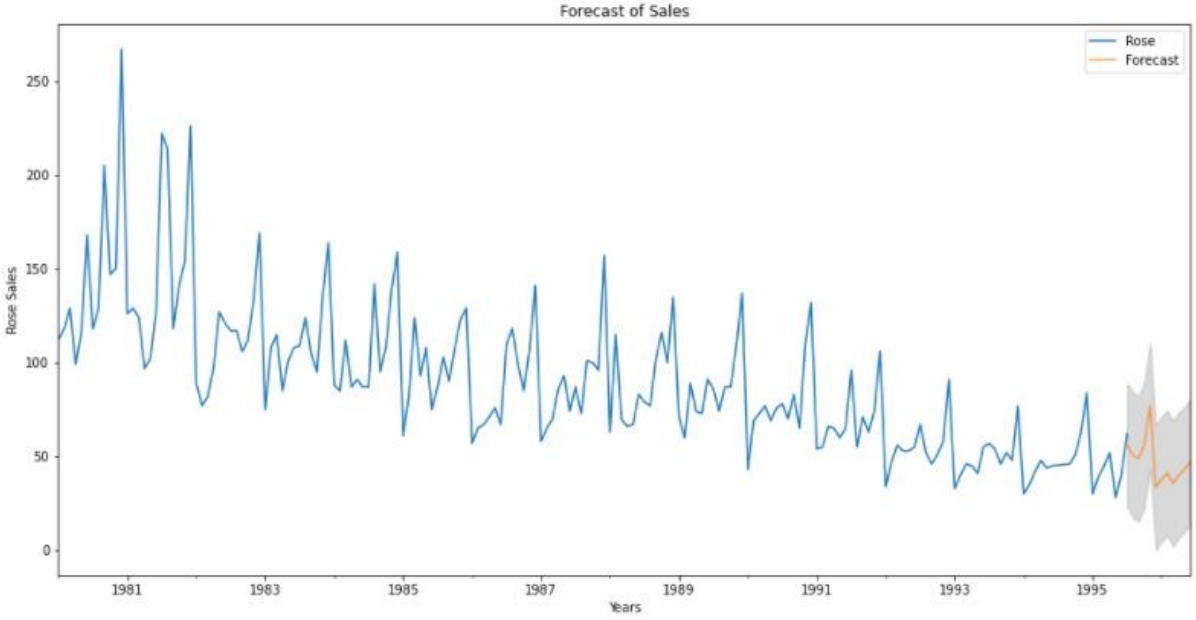
Sparkling Wines

	RMSE
Linear Regression	1389.135175
Naive Model	1389.135175
Simple Average	1275.081804
2 Pt Moving Average	813.400684
4 Pt Moving Average	1156.589694
6 Pt Moving Average	1283.927428
9 Pt Moving Average	1346.278315
Simple Exponential, Alpha=0.05	1316.035487
Double Exponential, Alpha=0.05	3096.812583
Triple Exponential, Additive	473.152417
Triple Exponential, Multiplicative	469.767970
ARIMA (2, 0, 2)	1005.083495
SARIMA(1,0,0)(1,0,1,12)	505.373773
ARIMA (4, 1, 3) ACF/PACF Cutoff	1398.797957
SARIMA (1, 1, 3)(1, 1, 1)12 ACF/PACF Cutoff	648.039376

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.**

Rose Wines

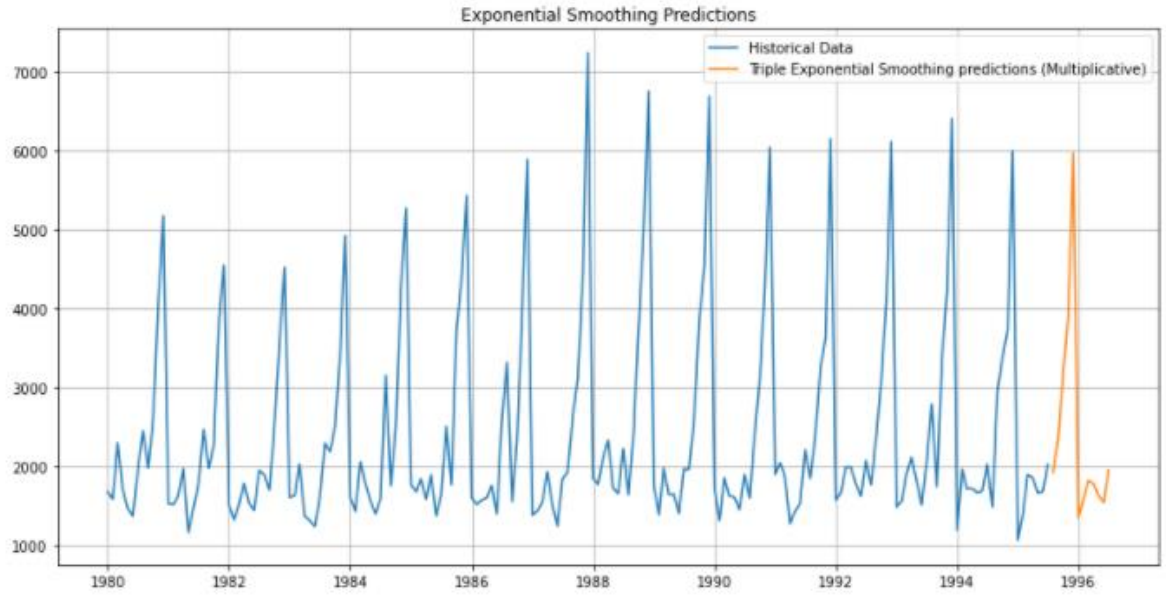
The SARIMA(1, 0, 0)(1, 0, 1,12) model has the lowest RMSE. We use it to predict the future sales.



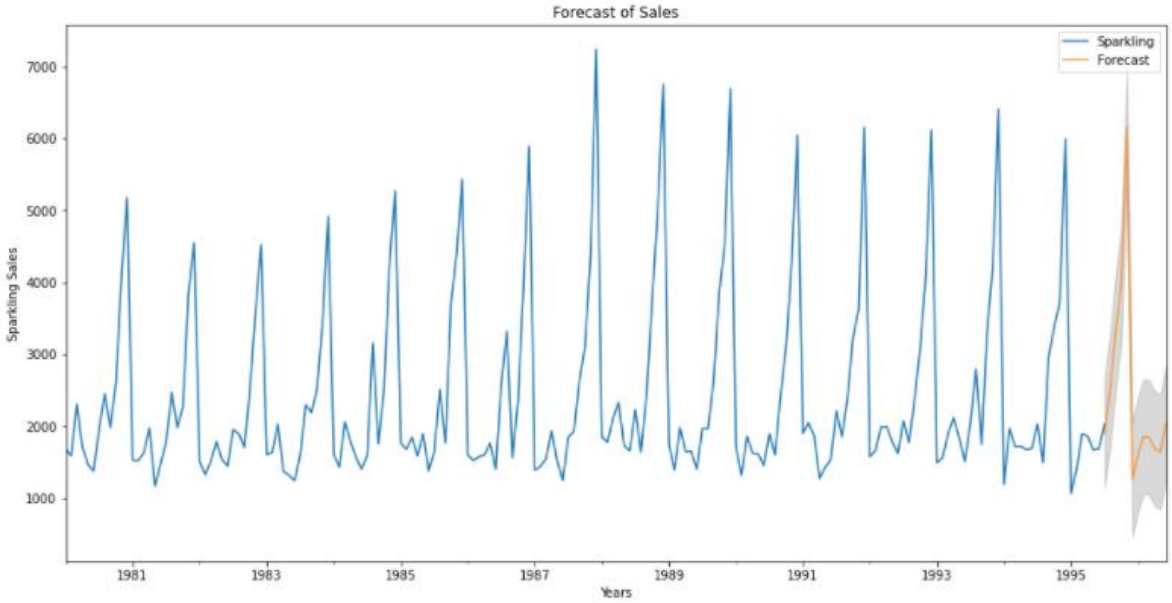
Sparkling Wine

Triple Exponential Smoothing Multiplicative & SARIMA (1, 0, 0)(1, 0, 1)12 are the 2 best models.

First we predict using Triple Exponential.



Now, we predict using SARIMAX (1, 0, 0)(1, 0, 1)12 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.**

Rose Wines

- We have predicted the future sales for 12 months for Rose Wines.
- Based on the model, we can observe that the macro trend for Sales of Rose Wines is declining.
- My suggestion would be to slowly ramp down production of Rose Wines.
- The company can perform market research to identify other products which have an increasing macro trend, and then explore selling other products.

Sparkling Wines

- We have predicted the future sales for 12 months for Sparkling Wines.
- My suggestion would be to manufacture Sparkling wine according to the seasonality.
- There seems to be no macro trend for the sales of Sparkling wines.
- The company would be able to maintain its current profitability, however, there is not much scope for growth.
- The company can perform market research to identify other products which have an increasing macro trend, and then explore selling other products.