
# Time Series Forecasting - Business Report

**By: Aaryani Kadiyala**

**PGP-Data Science and Business Analytics**

**(PGPDSBA.O.JAN24.A)**

# Table of Contents



# List of Tables:



# List of Figures:

## Problem Statment :

### Context

As an analyst at ABC Estate Wines, we are presented with historical data encompassing the sales of different types of wines throughout the 20th century. These datasets originate from the same company but represent sales figures for distinct wine varieties. Our objective is to delve into the data, analyze trends, patterns, and factors influencing wine sales over the course of the century. By leveraging data analytics and forecasting techniques, we aim to gain actionable insights that can inform strategic decision-making and optimize sales strategies for the future.

### Objective

The primary objective of this project is to analyze and forecast wine sales trends for the 20th century based on historical data provided by ABC Estate Wines. We aim to equip ABC Estate Wines with the necessary insights and foresight to enhance sales performance, capitalize on emerging market opportunities, and maintain a competitive edge in the wine industry.

## Solution :

## Rose Wine:

### Synopsis:

- Total no. of Entries = 187
- No. of Null Value = 2
- No Duplicates
- Train Set split till 1990 & Test set from 1991.
- Forecasting Models used
    - Linear Regression
    - Simple Average
    - 2 Point Moving Average
    - 4 Point Moving Average
    - 6 Point Moving Average
    - 9 Point Moving Average
    - Single Exponential Smoothing
    - Double Exponential Smoothing (Holt's Model)
    - Triple Exponential Smoothing (Holt-Winter Model) – Additive Model
    - Triple Exponential Smoothing (Holt-Winter Model) – Multiplicative Model
    - ARIMA / SARIMA (Auto fitted)
    - ARIMA / SARIMA (Manually fitted)

## Sparkling Wine:

**Synopsis:**

- Total no. of Entries = 187
- No Null values
- No Duplicates
- Train Set split till 1990 & Test set from 1991.
- Forecasting Models used
    - Linear Regression
    - Simple Average
    - 2 Point Moving Average
    - 4 Point Moving Average
    - 6 Point Moving Average
    - 9 Point Moving Average
    - Single Exponential Smoothing
    - Double Exponential Smoothing (Holt's Model)
    - Triple Exponential Smoothing (Holt-Winter Model) – Additive Model
    - Triple Exponential Smoothing (Holt-Winter Model) – Multiplicative Model
    - ARIMA / SARIMA (Auto fitted)
    - ARIMA / SARIMA (Manually fitted)

| | YearMonth | Rose |
|---|---|---|
| 0 | 1980-01 | 112.0 |
| 1 | 1980-02 | 118.0 |
| 2 | 1980-03 | 129.0 |
| 3 | 1980-04 | 99.0 |
| 4 | 1980-05 | 116.0 |

Figure 1: Head of Rose Wine

| | YearMonth | Sparkling |
|---|---|---|
| 0 | 1980-01 | 1686 |
| 1 | 1980-02 | 1591 |
| 2 | 1980-03 | 2304 |
| 3 | 1980-04 | 1712 |
| 4 | 1980-05 | 1471 |

Figure 2: Head of Sparkling Wine

Converted 'YearMonth' Column of both Rose & Sparkling wine into datetime.

| | YearMonth | Rose |
|---|---|---|
| 0 | 1980-01-31 | 112.0 |
| 1 | 1980-02-29 | 118.0 |
| 2 | 1980-03-31 | 129.0 |
| 3 | 1980-04-30 | 99.0 |
| 4 | 1980-05-31 | 116.0 |

Figure 3: Head of Rose Wine after conversion

| | YearMonth | Sparkling |
|---|---|---|
| 0 | 1980-01-31 | 1686 |
| 1 | 1980-02-29 | 1591 |
| 2 | 1980-03-31 | 2304 |
| 3 | 1980-04-30 | 1712 |
| 4 | 1980-05-31 | 1471 |

Figure 4: Head of Sparkling Wine after conversion

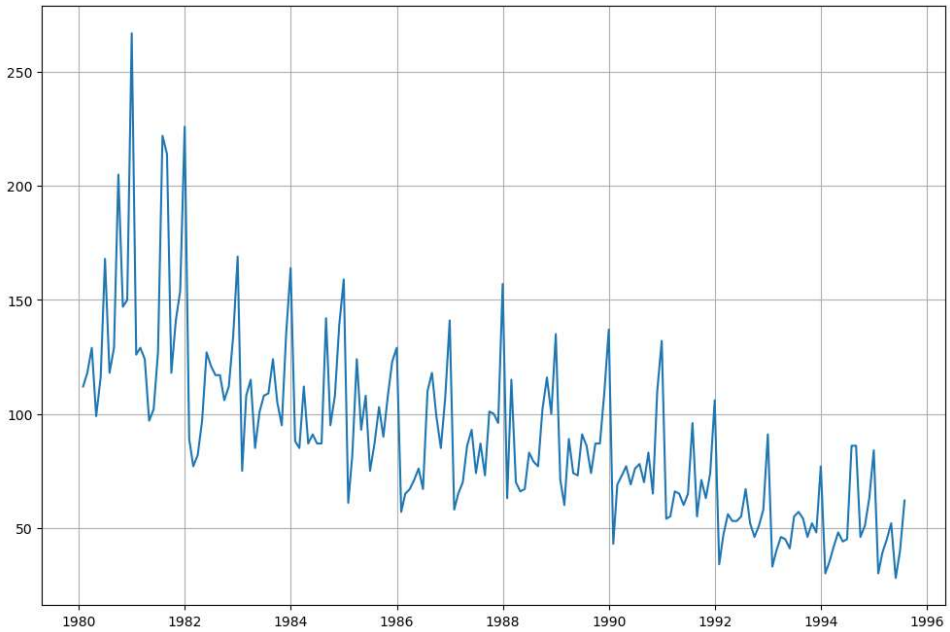



Figure 5 : Rose Wine Plot on Original Data

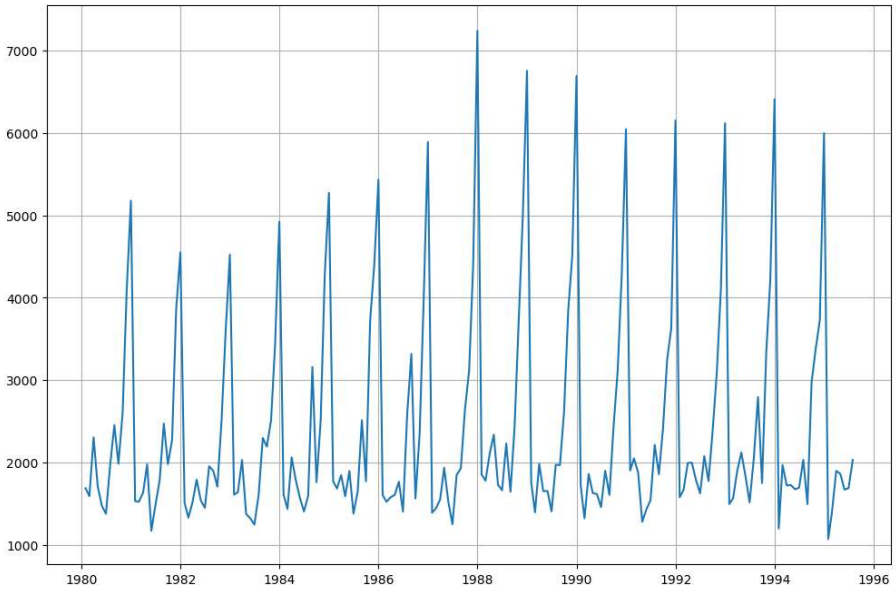



Figure 6: Sparkling Wine Plot on Original Data

## Exploratory Data Analysis:

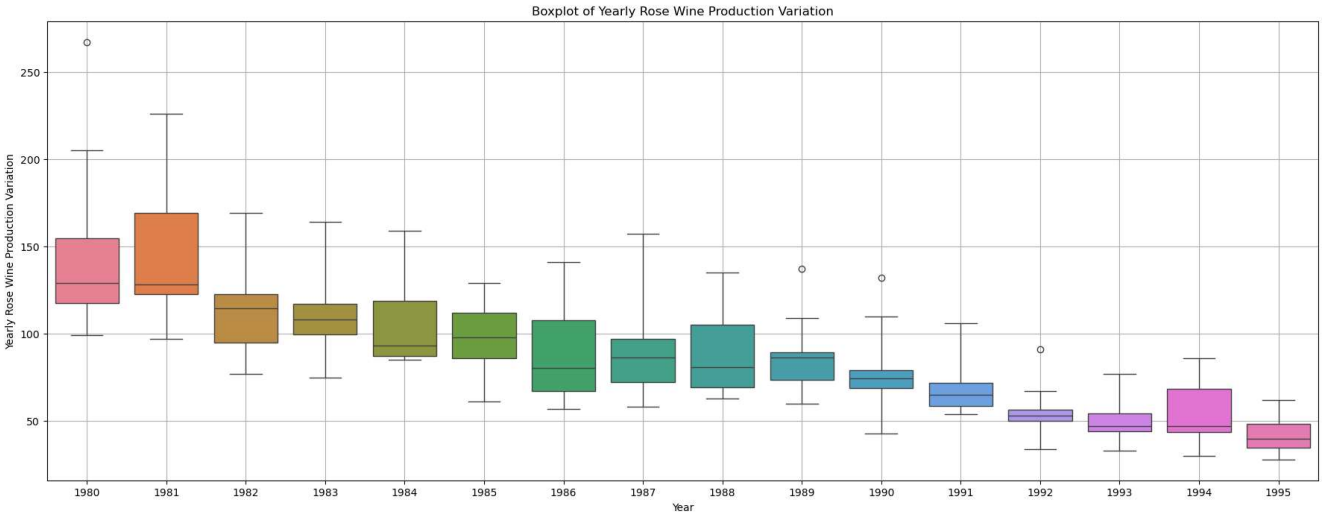



Figure 7: Yearly Production of Rose Wine

- Rose wine has mostly a downward sales trend.
- The highest sales for Rose wine can be observed in 1981 and the lowest sales in 1994.

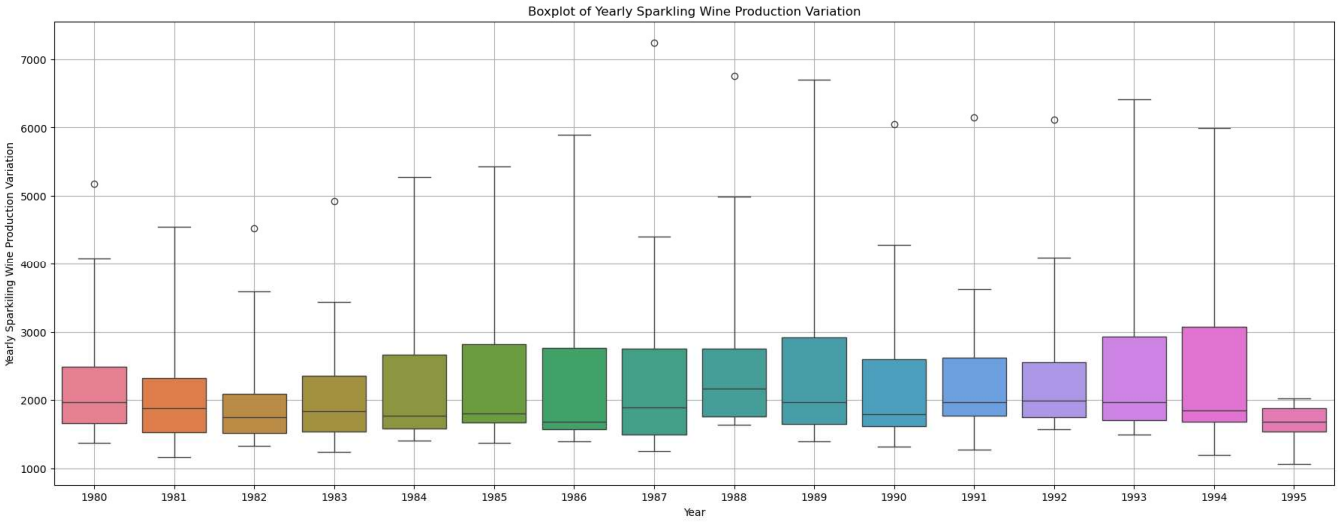



Figure 8: Yearly Production of Sparkling Wine

- Sparkling Wine sales have a variation each year.
- The years 1985 and 1986 seem to be the years with the least variation, so the 2 years show certain consistency in terms of sales.

- The highest sales for Sparkling Wine seem to happen in the year 1994 and the lowest in the year 1982.
- There is clear skewness that can be observed for Sparkling wine sales for all the years, except maybe in 1981.
- There are outliers in the yearly sales data, however as it is a Time Series, we can ignore the outlier data.
- The Sparkling wine sales appear to be going down from the year 1980 and have started increasing from the year 1983.
- The variation in Sparkling Wine sales seem to be increasing for the period 1983-1986, while the highest variation in Sparkling wine sales is in the year 1994.

| Month<br>Year | April | August | December | February | January | July | June | March | May | November | October | September |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1980** | 99.0 | 129.0 | 267.0 | 118.0 | 112.0 | 118.0 | 168.0 | 129.0 | 116.0 | 150.0 | 147.0 | 205.0 |
| **1981** | 97.0 | 214.0 | 226.0 | 129.0 | 126.0 | 222.0 | 127.0 | 124.0 | 102.0 | 154.0 | 141.0 | 118.0 |
| **1982** | 97.0 | 117.0 | 169.0 | 77.0 | 89.0 | 117.0 | 121.0 | 82.0 | 127.0 | 134.0 | 112.0 | 106.0 |
| **1983** | 85.0 | 124.0 | 164.0 | 108.0 | 75.0 | 109.0 | 108.0 | 115.0 | 101.0 | 135.0 | 95.0 | 105.0 |
| **1984** | 87.0 | 142.0 | 159.0 | 85.0 | 88.0 | 87.0 | 87.0 | 112.0 | 91.0 | 139.0 | 108.0 | 95.0 |
| **1985** | 93.0 | 103.0 | 129.0 | 82.0 | 61.0 | 87.0 | 75.0 | 124.0 | 108.0 | 123.0 | 108.0 | 90.0 |
| **1986** | 71.0 | 118.0 | 141.0 | 65.0 | 57.0 | 110.0 | 67.0 | 67.0 | 76.0 | 107.0 | 85.0 | 99.0 |
| **1987** | 86.0 | 73.0 | 157.0 | 65.0 | 58.0 | 87.0 | 74.0 | 70.0 | 93.0 | 96.0 | 100.0 | 101.0 |
| **1988** | 66.0 | 77.0 | 135.0 | 115.0 | 63.0 | 79.0 | 83.0 | 70.0 | 67.0 | 100.0 | 116.0 | 102.0 |
| **1989** | 74.0 | 74.0 | 137.0 | 60.0 | 71.0 | 86.0 | 91.0 | 89.0 | 73.0 | 109.0 | 87.0 | 87.0 |
| **1990** | 77.0 | 70.0 | 132.0 | 69.0 | 43.0 | 78.0 | 76.0 | 73.0 | 69.0 | 110.0 | 65.0 | 83.0 |
| **1991** | 65.0 | 55.0 | 106.0 | 55.0 | 54.0 | 96.0 | 65.0 | 66.0 | 60.0 | 74.0 | 63.0 | 71.0 |
| **1992** | 53.0 | 52.0 | 91.0 | 47.0 | 34.0 | 67.0 | 55.0 | 56.0 | 53.0 | 58.0 | 51.0 | 46.0 |
| **1993** | 45.0 | 54.0 | 77.0 | 40.0 | 33.0 | 57.0 | 55.0 | 46.0 | 41.0 | 48.0 | 52.0 | 46.0 |
| **1994** | 48.0 | 86.0 | 84.0 | 35.0 | 30.0 | 86.0 | 45.0 | 42.0 | 44.0 | 63.0 | 51.0 | 46.0 |
| **1995** | 52.0 | NaN | NaN | 39.0 | 30.0 | 62.0 | 40.0 | 45.0 | 28.0 | NaN | NaN | NaN |

Figure 9: Monthly Sales Across years for Rose Wine

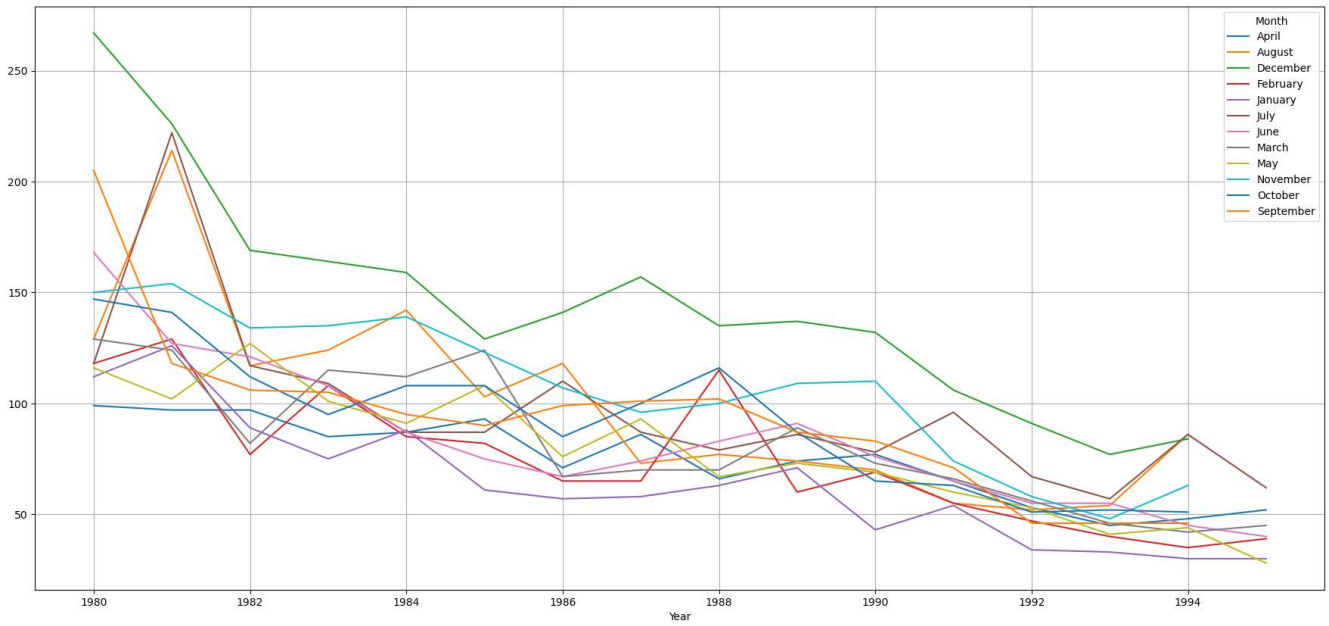



Figure 10: Monthly Production of Rose Wine

| Month<br>Year | April | August | December | February | January | July | June | March | May | November | October | September |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1980** | 1712.0 | 2453.0 | 5179.0 | 1591.0 | 1686.0 | 1966.0 | 1377.0 | 2304.0 | 1471.0 | 4087.0 | 2596.0 | 1984.0 |
| **1981** | 1976.0 | 2472.0 | 4551.0 | 1523.0 | 1530.0 | 1781.0 | 1480.0 | 1633.0 | 1170.0 | 3857.0 | 2273.0 | 1981.0 |
| **1982** | 1790.0 | 1897.0 | 4524.0 | 1329.0 | 1510.0 | 1954.0 | 1449.0 | 1518.0 | 1537.0 | 3593.0 | 2514.0 | 1706.0 |
| **1983** | 1375.0 | 2298.0 | 4923.0 | 1638.0 | 1609.0 | 1600.0 | 1245.0 | 2030.0 | 1320.0 | 3440.0 | 2511.0 | 2191.0 |
| **1984** | 1789.0 | 3159.0 | 5274.0 | 1435.0 | 1609.0 | 1597.0 | 1404.0 | 2061.0 | 1567.0 | 4273.0 | 2504.0 | 1759.0 |
| **1985** | 1589.0 | 2512.0 | 5434.0 | 1682.0 | 1771.0 | 1645.0 | 1379.0 | 1846.0 | 1896.0 | 4388.0 | 3727.0 | 1771.0 |
| **1986** | 1605.0 | 3318.0 | 5891.0 | 1523.0 | 1606.0 | 2584.0 | 1403.0 | 1577.0 | 1765.0 | 3987.0 | 2349.0 | 1562.0 |
| **1987** | 1935.0 | 1930.0 | 7242.0 | 1442.0 | 1389.0 | 1847.0 | 1250.0 | 1548.0 | 1518.0 | 4405.0 | 3114.0 | 2638.0 |
| **1988** | 2336.0 | 1645.0 | 6757.0 | 1779.0 | 1853.0 | 2230.0 | 1661.0 | 2108.0 | 1728.0 | 4988.0 | 3740.0 | 2421.0 |
| **1989** | 1650.0 | 1968.0 | 6694.0 | 1394.0 | 1757.0 | 1971.0 | 1406.0 | 1982.0 | 1654.0 | 4514.0 | 3845.0 | 2608.0 |
| **1990** | 1628.0 | 1605.0 | 6047.0 | 1321.0 | 1720.0 | 1899.0 | 1457.0 | 1859.0 | 1615.0 | 4286.0 | 3116.0 | 2424.0 |
| **1991** | 1279.0 | 1857.0 | 6153.0 | 2049.0 | 1902.0 | 2214.0 | 1540.0 | 1874.0 | 1432.0 | 3627.0 | 3252.0 | 2408.0 |
| **1992** | 1997.0 | 1773.0 | 6119.0 | 1667.0 | 1577.0 | 2076.0 | 1625.0 | 1993.0 | 1783.0 | 4096.0 | 3088.0 | 2377.0 |
| **1993** | 2121.0 | 2795.0 | 6410.0 | 1564.0 | 1494.0 | 2048.0 | 1515.0 | 1898.0 | 1831.0 | 4227.0 | 3339.0 | 1749.0 |
| **1994** | 1725.0 | 1495.0 | 5999.0 | 1968.0 | 1197.0 | 2031.0 | 1693.0 | 1720.0 | 1674.0 | 3729.0 | 3385.0 | 2968.0 |
| **1995** | 1862.0 | NaN | NaN | 1402.0 | 1070.0 | 2031.0 | 1688.0 | 1897.0 | 1670.0 | NaN | NaN | NaN |

Figure 11: Monthly Sales Across years for Sparkling Wine

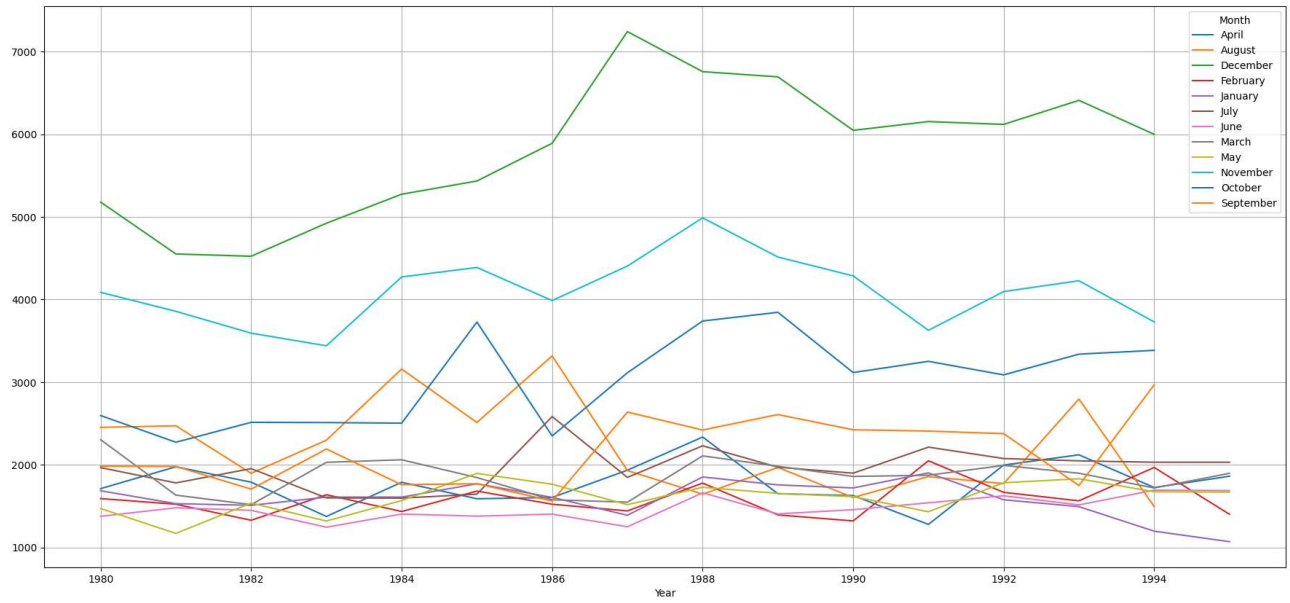



Figure 12: Monthly Production of Sparkling Wine

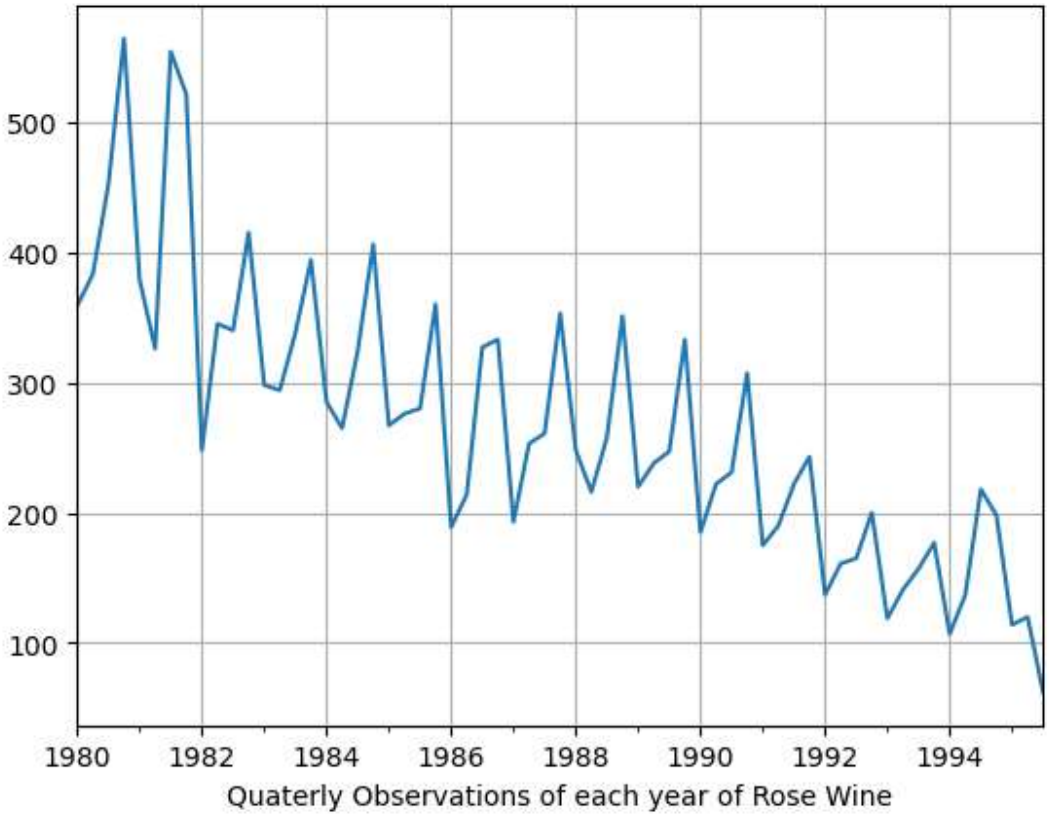



Figure 13: Quarterly Production of Rose Wine

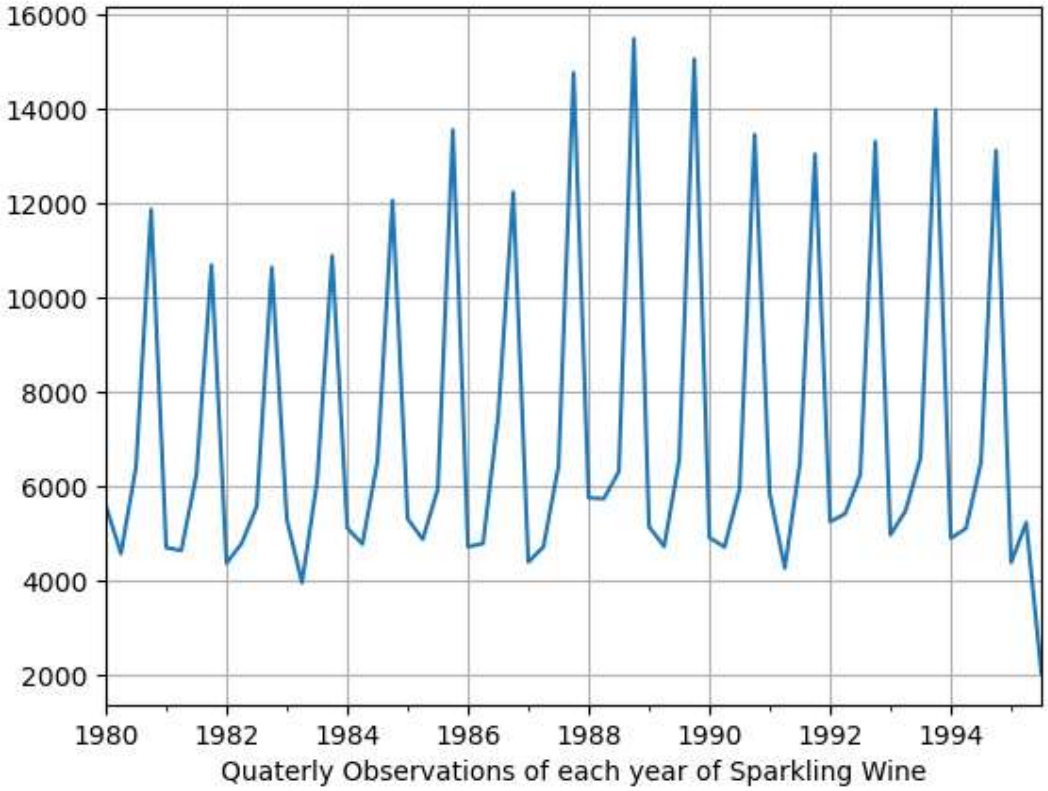



Figure 14: Quarterly Production of Sparkling Wine

- There is a seasonality element visible in both the Rose as well as Sparkling Wine time series datasets.
- The Sparkling wine seem to have a higher seasonality element as compared to Rose Wine.
- The sales have an increasing sales trend in the last quarter of the year, with Sparkling wine observing a steeper rise in sales during last quarter.
- The sales for Rose wine seem to pick up from January month and is more or less consistent till June, observes some stagnancy till September month and then starts to pick up again from October (i.e. last quarter); while for Sparkling wine, the sales is relatively low in first two quarters, slowly picks up pace during the third quarter and goes on a rise till the end of the year.
- Monthly sales data for both the type of wines shows skewness without much exceptions.

**Decomposition:**

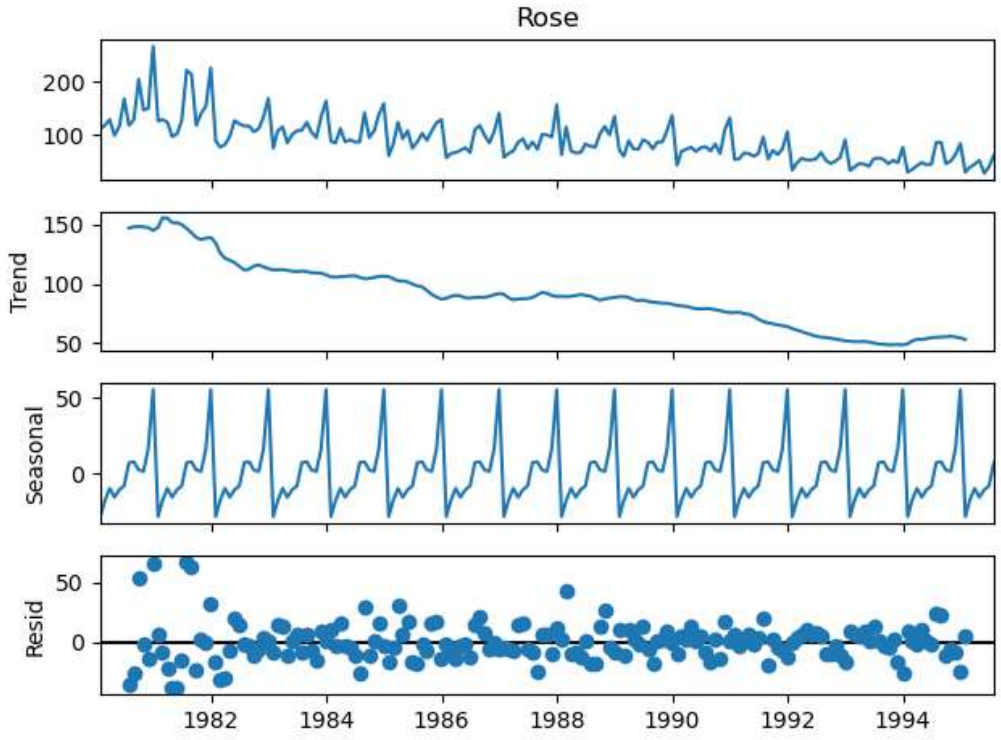



Figure 15: Additive Decomposition of Rose Wine

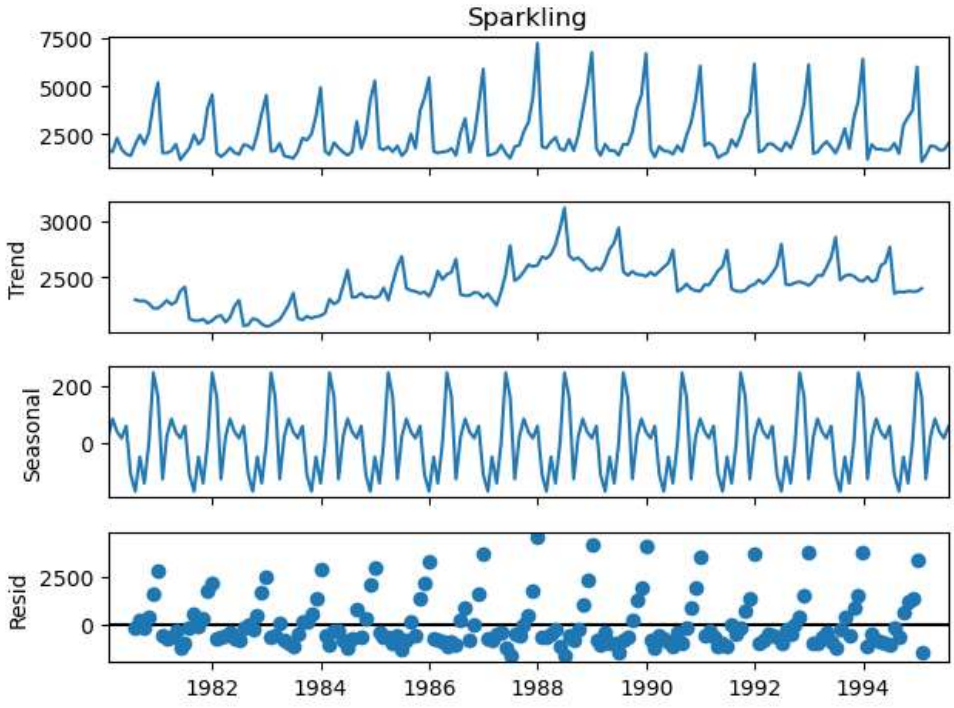



Figure 16: Additive Decomposition of Sparkling Wine

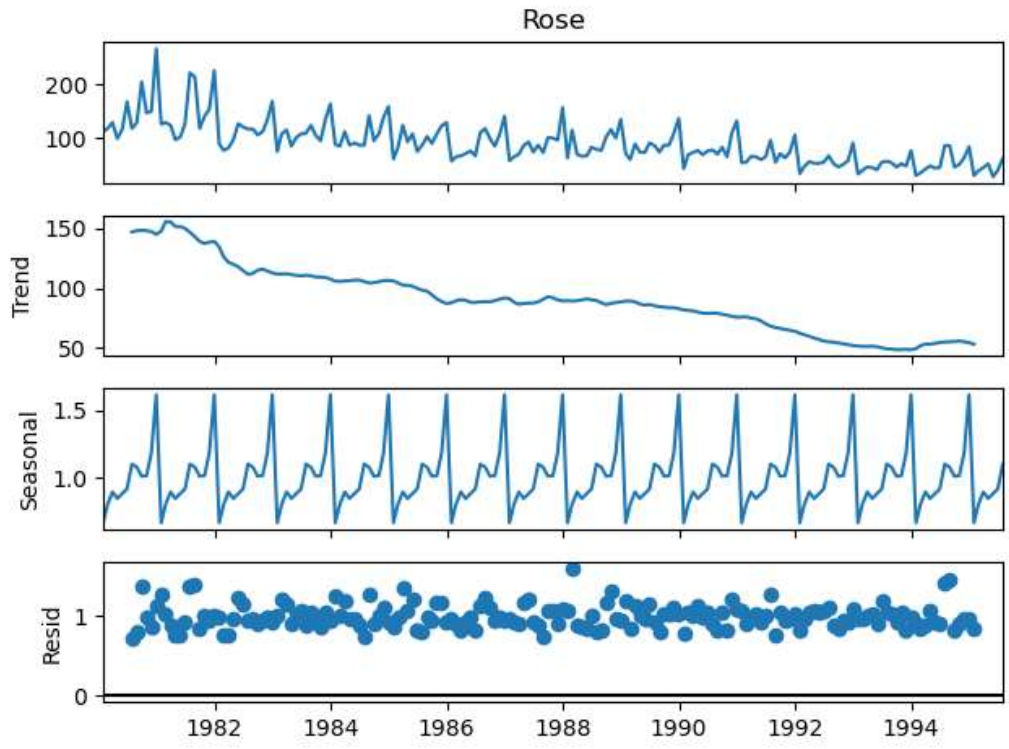



Figure 17: Multiplicative Decomposition of Rose Wine

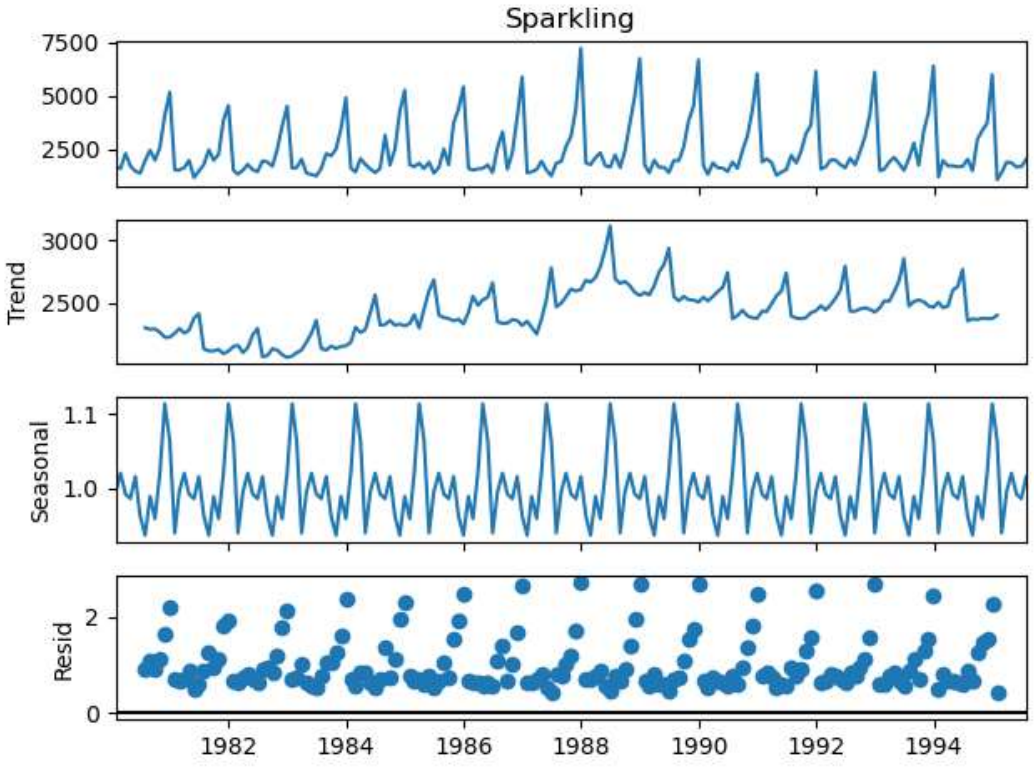


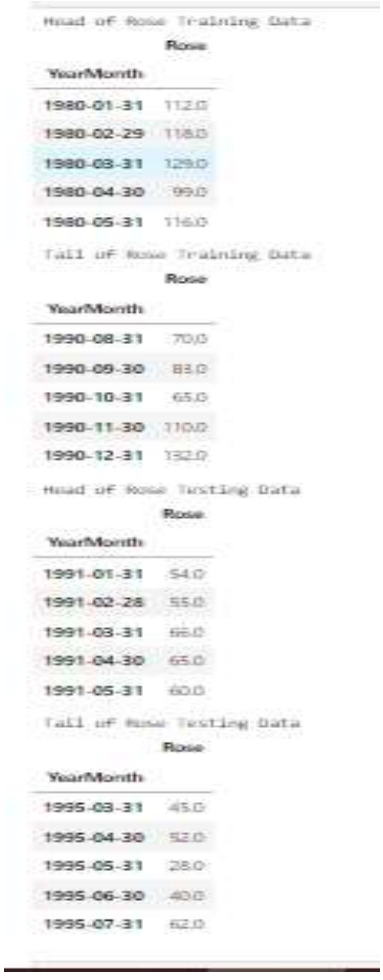

Figure 18: Multiplicative Decomposition of Sparkling Wine

- Here by just observing the Residual patterns of Additive and Multiplicative models of Rose and Sparkling datasets. It seems that
    - Rose is Multiplicative
    - Sparkling is Additive
- The plots above clearly indicate that the Wine sales are unstable and not uniform, and they have an apparent seasonality trend. Moreover, the seasonal variation seems to be more in the case of Sparkling wine as compared to the Rose wine; while the sales variation seems to be more in case of Rose wine as compared to Sparkling wine.

## Train & Test Data Split:

Rose & Sparkling wine data has been split into test and train datasets for further model evaluations.

- Train data for both rose & sparkling wine is till 1990
- Test data for both rose & sparkling wine is from 1991



Head of Rose Training Data

| YearMonth | Rose |
|---|---|
| 1980-01-31 | 112.0 |
| 1980-02-29 | 118.0 |
| 1980-03-31 | 129.0 |
| 1980-04-30 | 99.0 |
| 1980-05-31 | 116.0 |

Tail of Rose Training Data

| YearMonth | Rose |
|---|---|
| 1990-08-31 | 70.0 |
| 1990-09-30 | 83.0 |
| 1990-10-31 | 65.0 |
| 1990-11-30 | 110.0 |
| 1990-12-31 | 132.0 |

Head of Rose Testing Data

| YearMonth | Rose |
|---|---|
| 1991-01-31 | 54.0 |
| 1991-02-28 | 55.0 |
| 1991-03-31 | 66.0 |
| 1991-04-30 | 65.0 |
| 1991-05-31 | 60.0 |

Tail of Rose Testing Data

| YearMonth | Rose |
|---|---|
| 1995-03-31 | 45.0 |
| 1995-04-30 | 52.0 |
| 1995-05-31 | 28.0 |
| 1995-06-30 | 40.0 |
| 1995-07-31 | 62.0 |

Figure 19: Train & Test Split of Rose Wine

Head of Sparkling Training Data

| YearMonth | Sparkling |
|---|---|
| 1980-01-31 | 1686 |
| 1980-02-29 | 1591 |
| 1980-03-31 | 2304 |
| 1980-04-30 | 1712 |
| 1980-05-31 | 1471 |

Tail of Sparkling Training Data

| YearMonth | Sparkling |
|---|---|
| 1990-08-31 | 1605 |
| 1990-09-30 | 2424 |
| 1990-10-31 | 3116 |
| 1990-11-30 | 4286 |
| 1990-12-31 | 6047 |

Head of Sparkling Testing Data

| YearMonth | Sparkling |
|---|---|
| 1991-01-31 | 1902 |
| 1991-02-28 | 2049 |
| 1991-03-31 | 1874 |
| 1991-04-30 | 1279 |
| 1991-05-31 | 1432 |

Tail of Sparkling Testing Data

| YearMonth | Sparkling |
|---|---|
| 1995-03-31 | 1897 |
| 1995-04-30 | 1862 |
| 1995-05-31 | 1670 |
| 1995-06-30 | 1688 |
| 1995-07-31 | 2031 |

Figure 20: Train & Test Split of Sparkling Wine

- Shape of Train Rose & Sparkling wine dataset : (132, 1)
- Shape of Test Rose & Sparkling wine dataset : (55, 1)

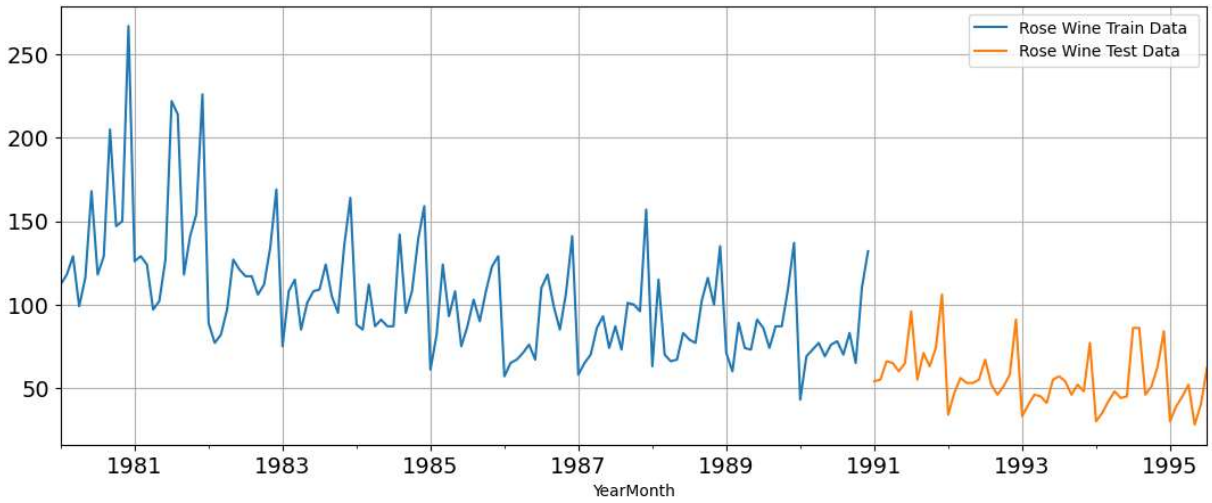



Figure 21: Train & Test Set of Rose Wine

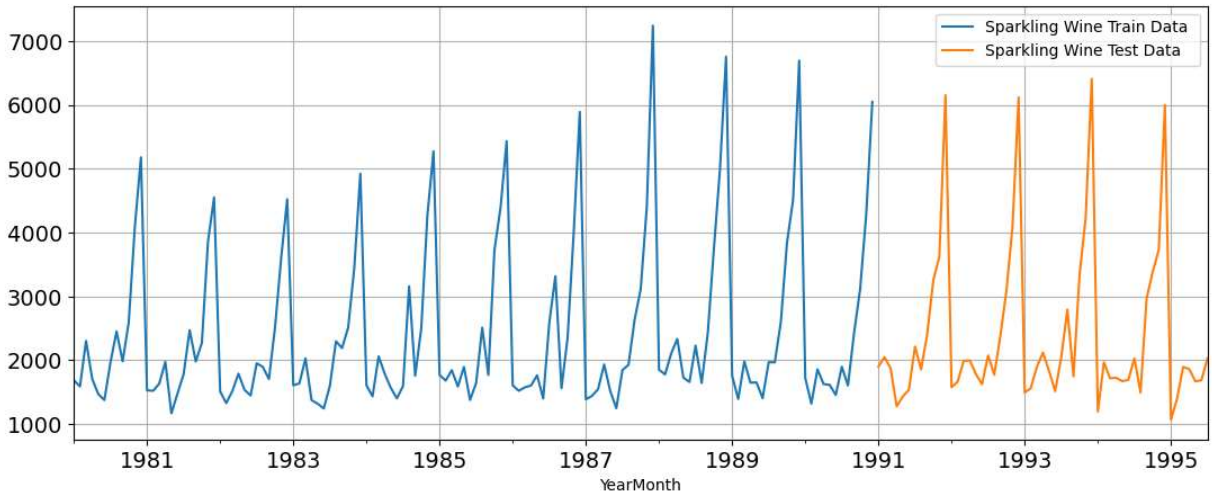



Figure 22: Train & Test Set of Sparkling Wine

## Model Building:

### Linear Regression:

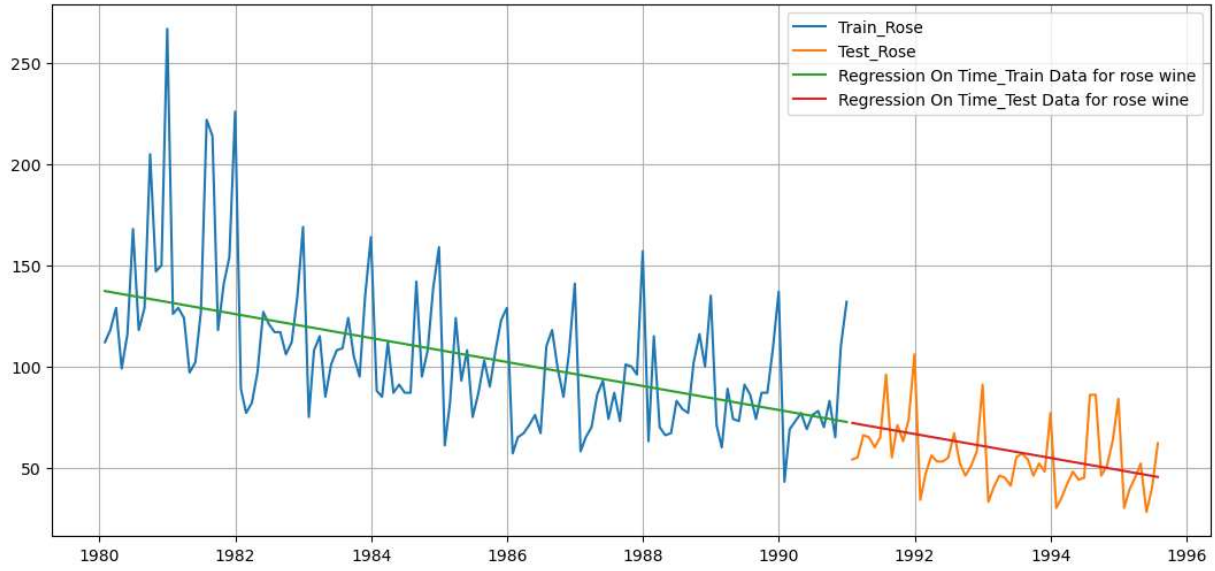



Figure 23: Linear Regression of Rose Wine

Linear Regression RSME on the Train Data of Rose Wine = 30.72
Linear Regression RSME on the Test Data of Rose Wine = 15.27

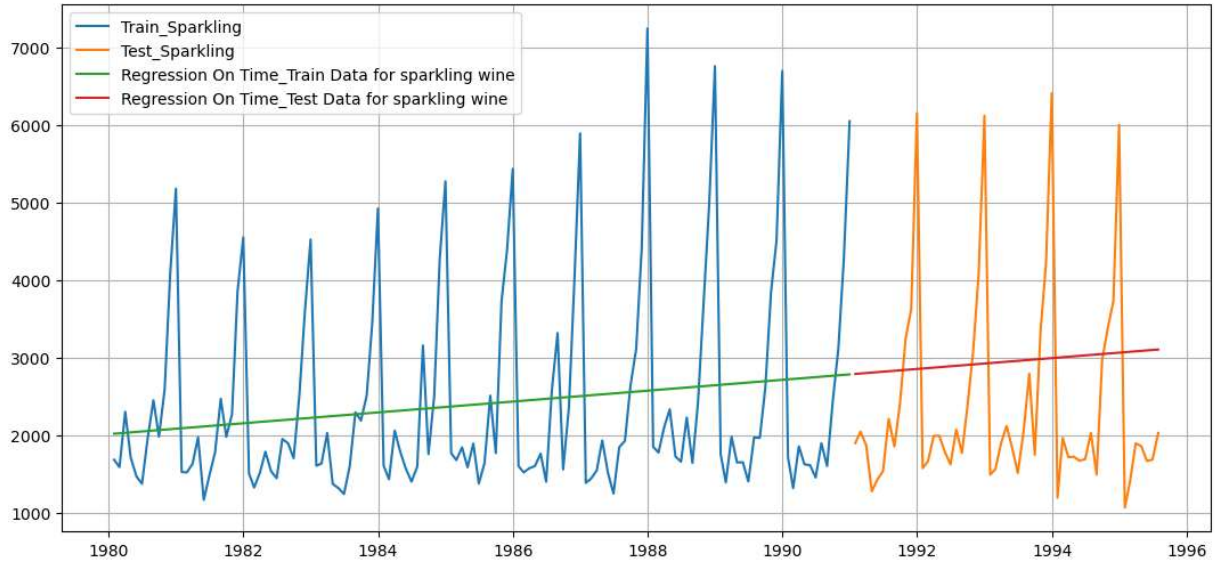



Figure 24: Linear Regression of Sparkling Wine

Linear Regression RSME on the Train Data of Sparkling Wine = 1279.32
Linear Regression RSME on the Test Data of Sparkling Wine = 1389.14

**Simple Average:**

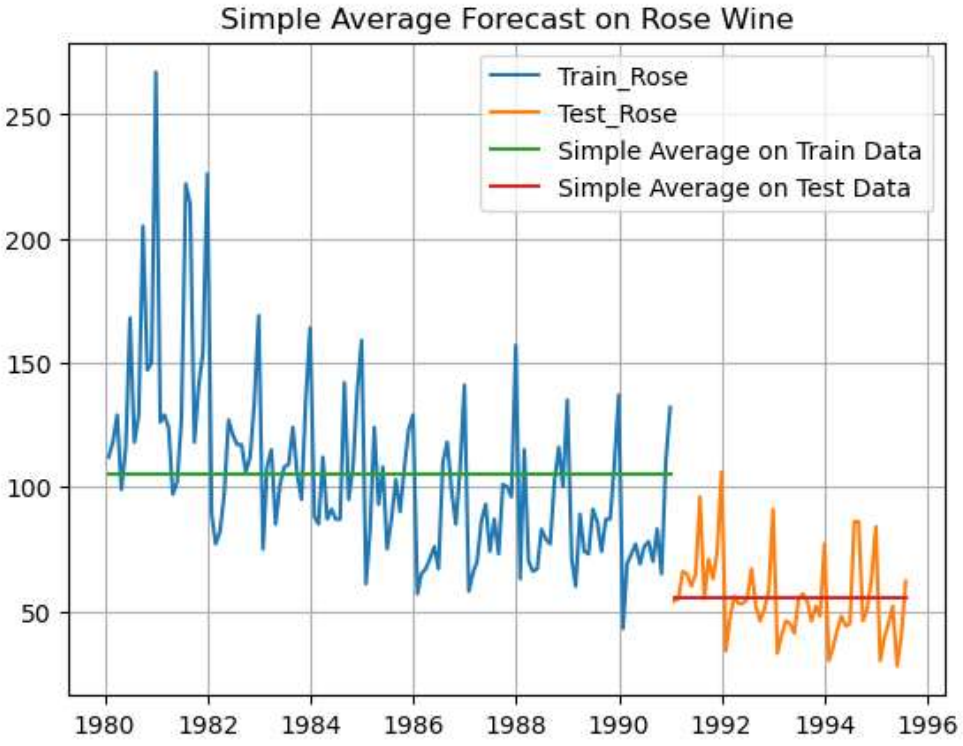



Figure 25: Simple Average of Rose Wine

RSME Simple Average forecast on the Train Data for Rose Wine = 36.034
RSME Simple Average forecast on the Test Data for Rose Wine = 15.760

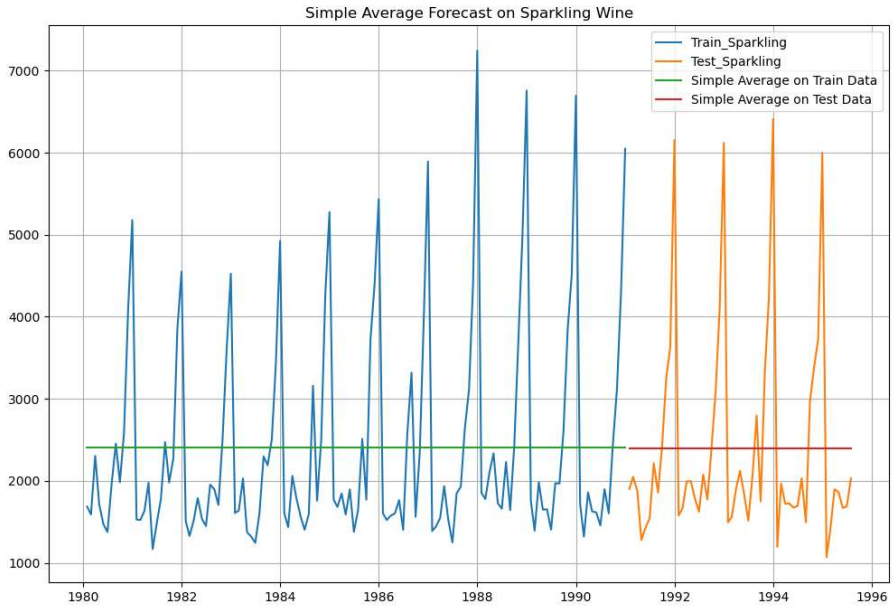



Figure 26: Simple Average of Sparkling Wine

RMSE Simple Average forecast on the Train Data for Sparkling Wine = 1298.484
RMSE Simple Average forecast on the Test Data for Sparkling Data = 1275.073

**Moving Average:**

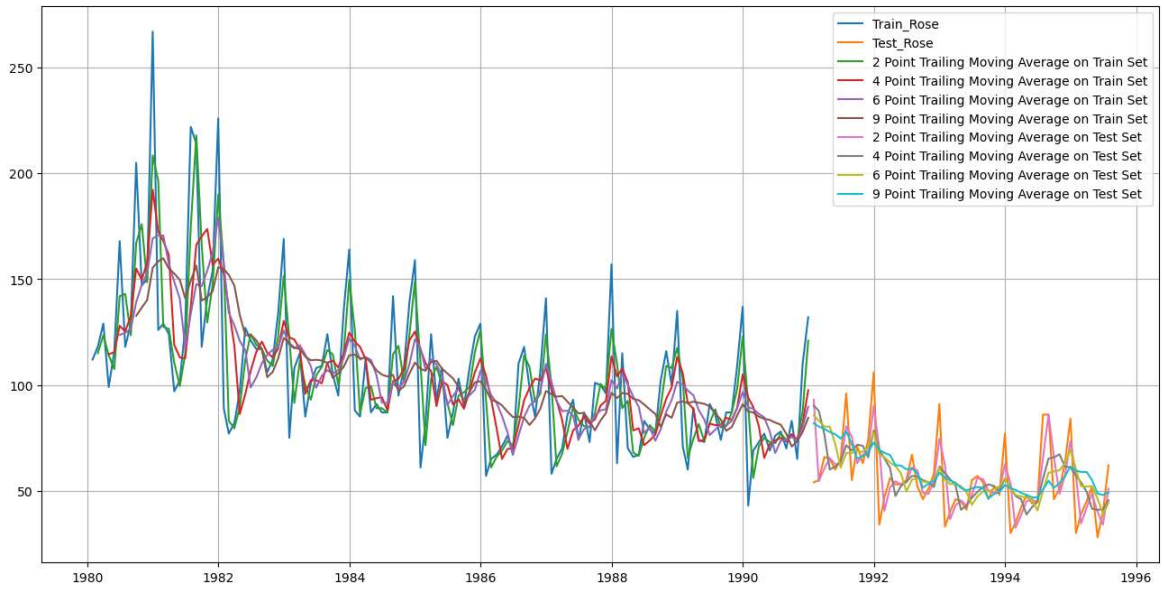



Figure 27: Moving Average of Rose Wine

2 point Moving Average Model forecast on the Test Data, RMSE is 11.529
4 point Moving Average Model forecast on the Test Data, RMSE is 14.451
6 point Moving Average Model forecast on the Test Data, RMSE is 14.567
9 point Moving Average Model forecast on the Test Data, RMSE is 14.728

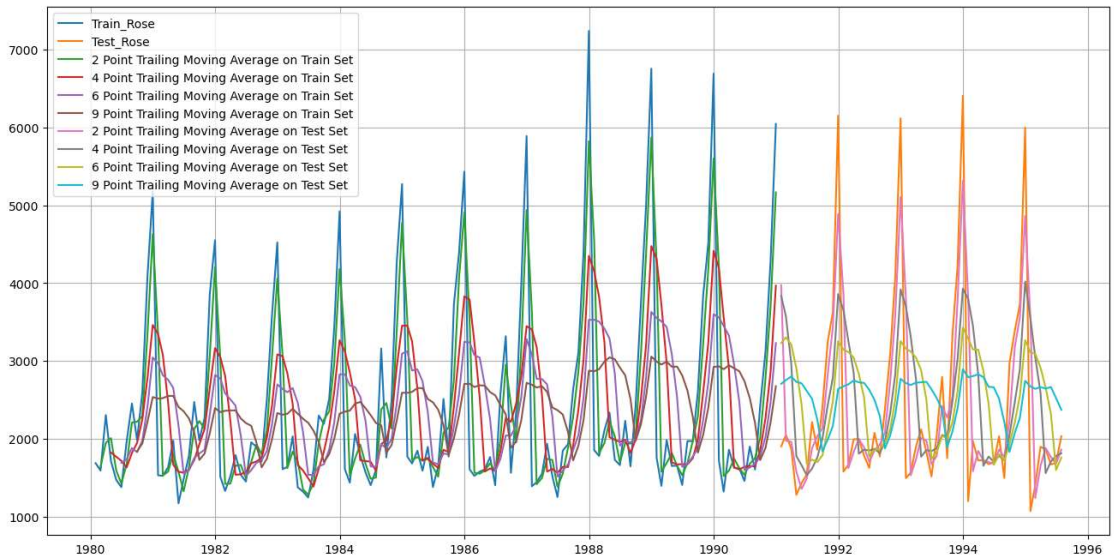



Figure 28: Moving Average of Sparkling Wine

For 2 point Moving Average Model forecast on the Test Data, RMSE is 813.401
For 4 point Moving Average Model forecast on the Test Data, RMSE is 1156.590
For 6 point Moving Average Model forecast on the Test Data, RMSE is 1283.928
For 9 point Moving Average Model forecast on the Test Data, RMSE is 1346.279

- We have built 4 models till now for both Rose and Sparkling Wine datasets
- We fitted various models to the Train split and tested it on Test split. Accuracy metrics used is R oot Mean Squared Error (RMSE) on Test data

- **Model 1 - Linear Regression**
  - We regressed variables 'Rose' and 'Sparkling' against their individual time instances
  - We modified the datasets and tagged individual sales to their time instances
  - TEST RMSE ROSE = 15.27 | TEST RMSE SPARKLING = 1389.14

- **Model 2 - Simple Average**
  - All future predictions are the same as the simple average of all data till today
  - TEST RMSE ROSE = 15.76 | TEST RMSE SPARKLING = 1275.073

- **Model 3 - Moving Average (MA)**
  - We calculate rolling means (Moving averages) over different intervals for the whole train data
  - 2 Point Moving Average:
    TEST RMSE ROSE = 11.53 | TEST RMSE SPARKLING = 813.401
  - 4 Point Moving Average:
    TEST RMSE ROSE = 14.45 | TEST RMSE SPARKLING = 1156.590
  - 6 Point Moving Average:
    TEST RMSE ROSE = 14.57 | TEST RMSE SPARKLING = 1283.928
  - 9 Point Moving Average:
    TEST RMSE ROSE = 14.73 | TEST RMSE SPARKLING = 1346.279

RSME of Test Data for Rose Wine

| | Rose Test RMSE |
|---|---|
| Linear Regression | 15.27 |
| Simple Average Model | 15.76 |
| 2pointTrailingMovingAverage | 11.53 |
| 4pointTrailingMovingAverage | 14.45 |
| 6pointTrailingMovingAverage | 14.57 |
| 9pointTrailingMovingAverage | 14.73 |

RSME of Test Data for Sparkling Wine

| | Sparkling Test RMSE |
|---|---|
| Linear Regression | 1389.14 |
| Simple Average Model | 1275.07 |
| 2pointTrailingMovingAverage | 813.40 |
| 4pointTrailingMovingAverage | 1156.59 |
| 6pointTrailingMovingAverage | 1283.93 |
| 9pointTrailingMovingAverage | 1346.28 |

Figure 29: Test RSME of all 3 models of Rose & Sparkling Wine

- Till now, Best Model which gives lowest RMSE score for both Rose and Sparkling is 2 Point Movi ng Average Model
- We'll continue to forecast using Exponential Smoothing Models for both datasets of Rose and Sp arkling Wine Sales
- We'll build following Exponential Smoothing Models –
  - Simple Exponential Smoothing
  - Double Exponential Smoothing (Holt's Model)
  - Triple Exponential Smoothing (Holt-Winter Model)

**Single Exponential Smoothing:**

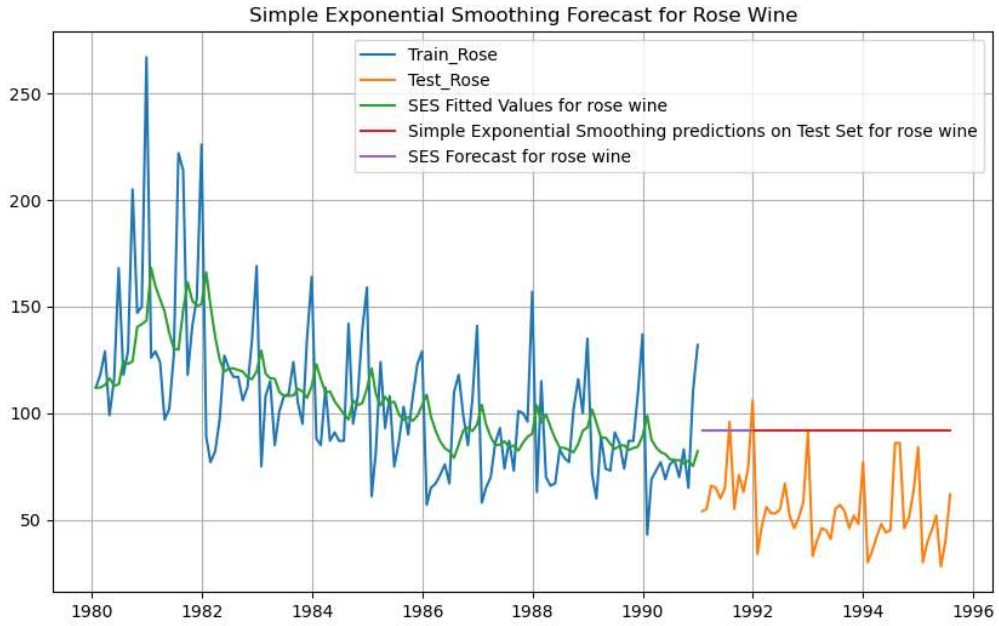



Figure 30: Simple Exponential Smoothing of Rose Wine

Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 41.362

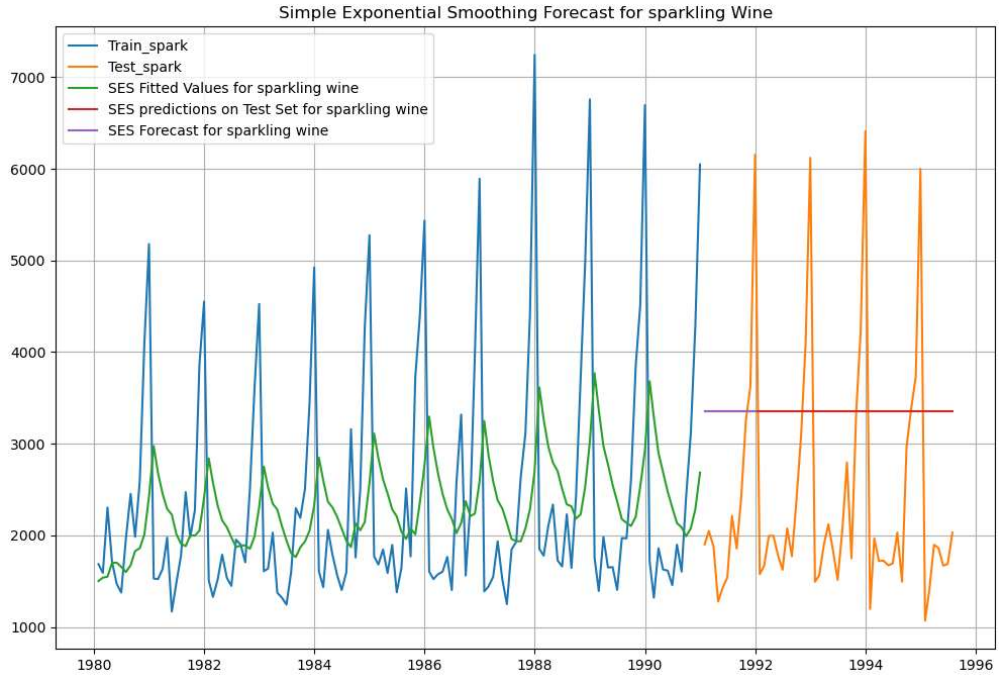



Figure 31: Simple Exponential Smoothing of Sparkling Wine

Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 1595.207

Best Model till now for Rose and Sparkling is 2 Point Moving Average Model

**Double Exponential Smoothing (Holt's Model):**

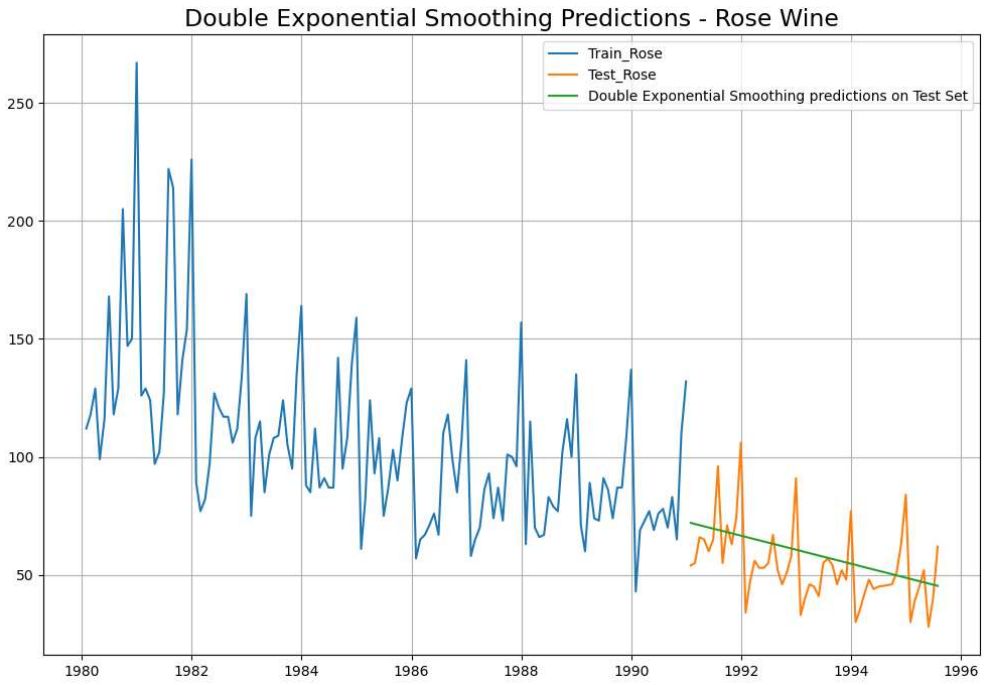



Figure 32: Double Exponential Smoothing of Rose Wine

Double Exponential Smoothing Model forecast on the Test Data, RMSE is 15.268

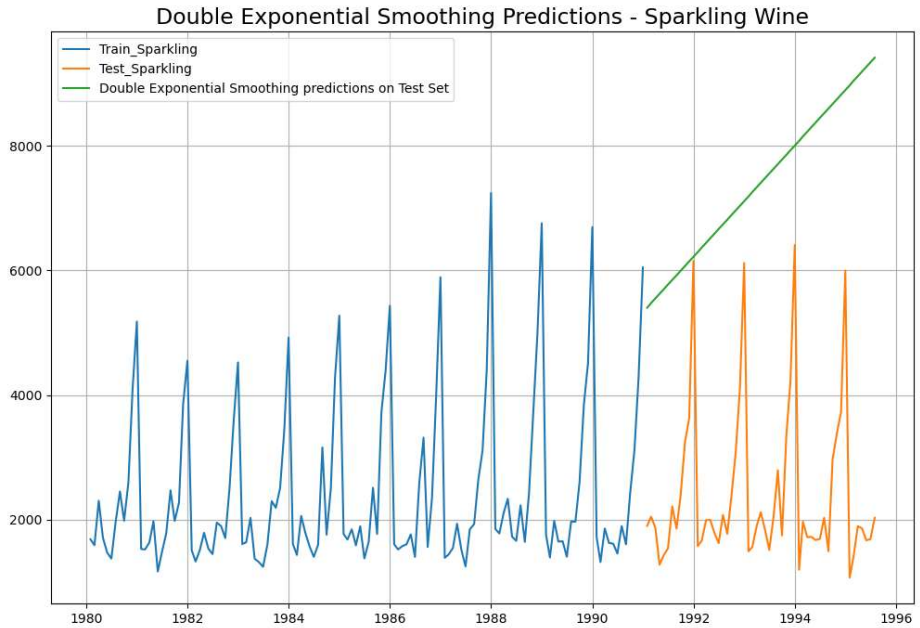



Figure 33: Double Exponential Smoothing of Sparkling Wine

Double Exponential Smoothing Model forecast on the Test Data, RMSE is 5291.87

- In Rose - DES has picked up the trend well. DES seems to perform better than SES here
- In Sparkling - DES shows a non-existent trend. DES does not perform well here
- Best Model till now for Rose and Sparkling is 2 Point Moving Average Model

**Triple Exponential Smoothing (Holt - Winter Model) – Additive Model:**

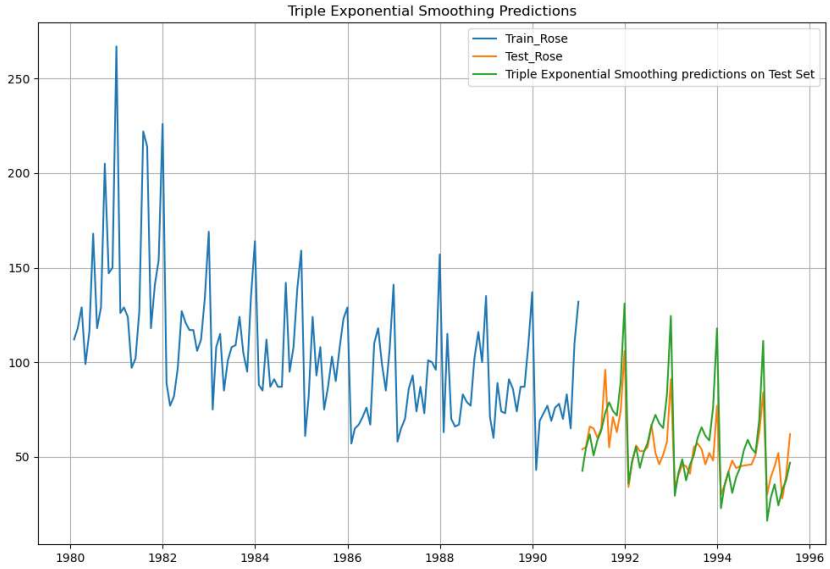



Figure 34: Triple Exponential Smoothing of Rose Wine - Additive

Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 14.26

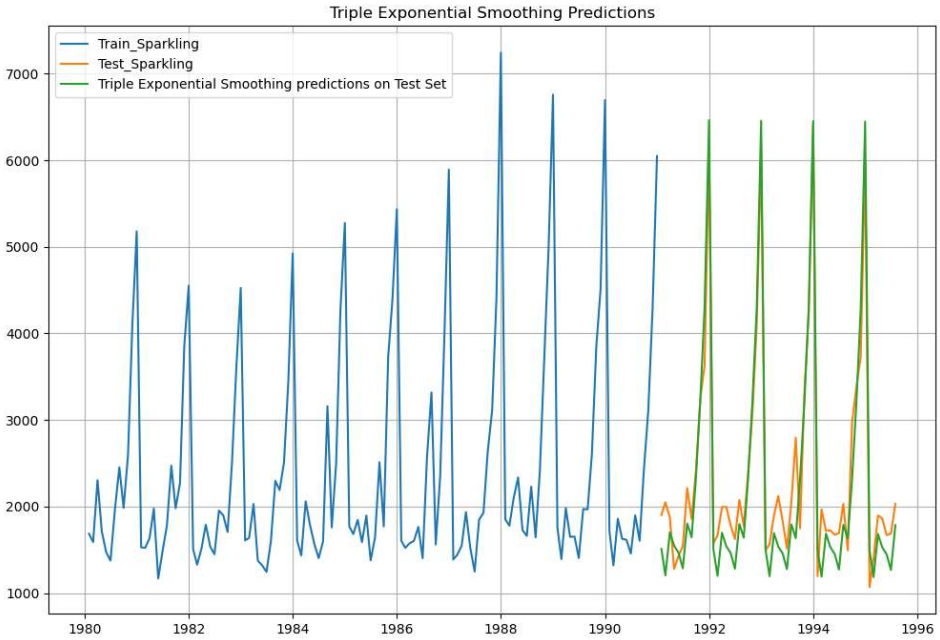



Figure 35: Triple Exponential Smoothing of Sparkling Wine - Additive

Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 380.09

**Triple Exponential Smoothing (Holt - Winter Model) – Multiplicative Model:**

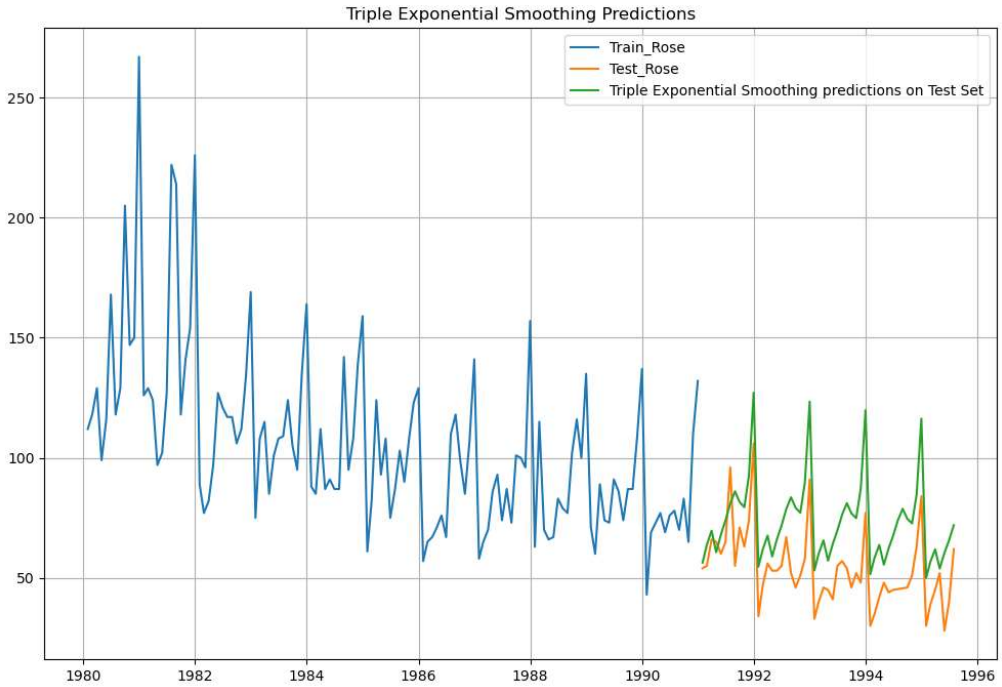



Figure 36: Triple Exponential Smoothing of Rose Wine – Multiplicative

RSME of Triple Exponential Smoothing Model forecast on the Test Data : 22.101

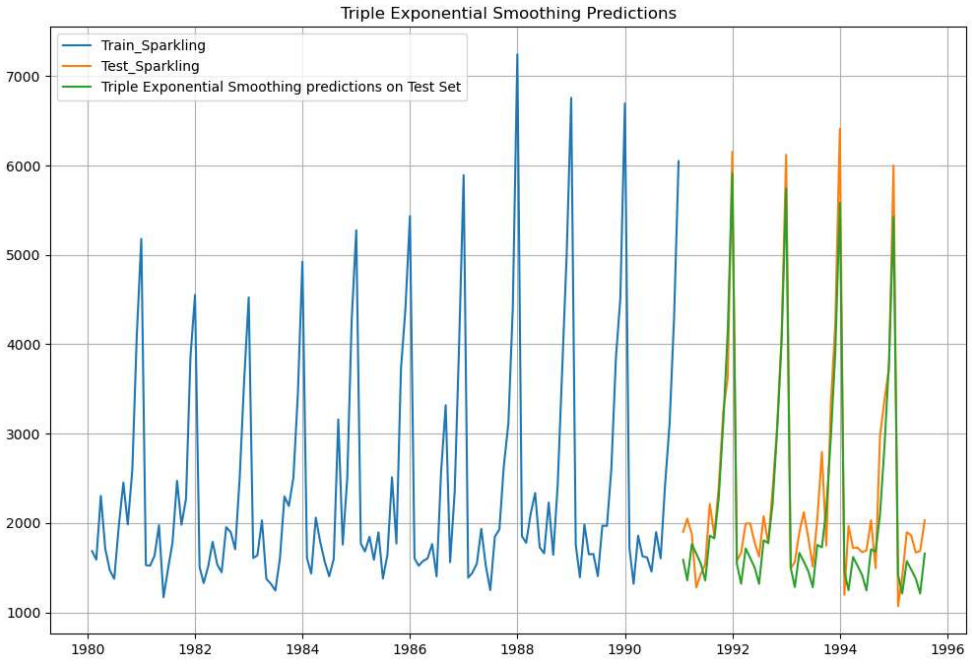



Figure 37: Triple Exponential Smoothing of Sparkling Wine – Multiplicative

RSME of Triple Exponential Smoothing Model forecast on the Test Data : 386.836

| | Rose Test RMSE |
|---|---|
| Linear Regression | 15.27 |
| Simple Average Model | 15.76 |
| 2pointTrailingMovingAverage | 11.53 |
| 4pointTrailingMovingAverage | 14.45 |
| 6pointTrailingMovingAverage | 14.57 |
| 9pointTrailingMovingAverage | 14.73 |
| Simple Exponential Smoothing | 41.36 |
| Double Exponential Smoothing | 15.27 |
| Triple Exponential Smoothing - Additive Model | 14.26 |
| Triple Exponential Smoothing - Multiplicative Model | 22.10 |

Figure 38: Test RSME of all models of Rose Wine

| | Sparkling Test RMSE |
|---|---|
| Linear Regression | 1389.14 |
| Simple Average Model | 1275.07 |
| 2pointTrailingMovingAverage | 813.40 |
| 4pointTrailingMovingAverage | 1156.59 |
| 6pointTrailingMovingAverage | 1283.93 |
| 9pointTrailingMovingAverage | 1346.28 |
| Simple Exponential Smoothing | 1595.21 |
| Double Exponential Smoothing | 5291.88 |
| Triple Exponential Smoothing - Additive Model | 380.09 |
| Triple Exponential Smoothing - Multiplicative Model | 386.84 |

Figure 39: Test RSME of all models of Sparkling Wine

Best Model till now,

- Rose Wine : 2 Point Moving Average Model
- Sparkling Wine : Triple Exponential Smoothing – Additive Model

**Stationary Test:**

Dickey-Fuller Test - Dicky Fuller Test on the timeseries is run to check for stationarity of data.

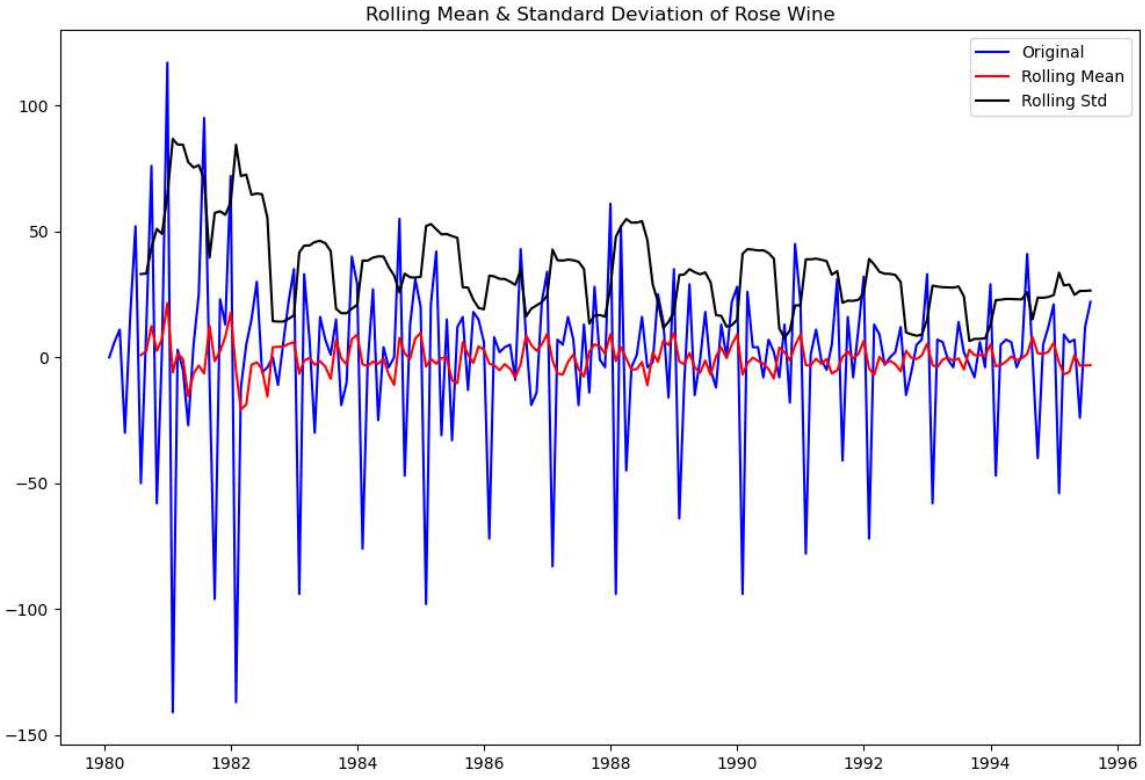
- Null Hypothesis H0: Time Series is non-stationary.

Alternate Hypothesis $\boldsymbol{Ha}$: Time Series is stationary.

- So, in ADF Test, if p-value < 0.05 - We reject the Null Hypothesis and hence conclude that given Time Series is Stationary
- So, in ADF Test, if p-value > 0.05 - We fail to reject the Null Hypothesis and hence conclude that given Time Series is Not Stationary
- If Time Series is not Stationary then we apply one level of differencing and check for Stationarity again.
- Again, if the Time Series is still not Stationary, we apply one more level of differencing and check for Stationarity again
- Once the Time Series is Stationary then we are ready to apply ARIMA / SARIMA models

**Stationarity of Rose Wine Dataset:**

- Augmented Dicky-Fuller Test was applied to the whole Rose dataset
  - We found, p-value = 0.343
  - Here, p-value > 0.05
  - We fail to reject the Null Hypothesis and hence conclude that Rose Wine Time Series is Not Stationary
  - We take 1 level of differencing and check again for Stationarity
  - Now, p-value = 1.656242e-12
  - Now, p-value < 0.05
  - Now, we reject the Null Hypothesis and conclude that Rose Time Series is Stationary with a lag of 1

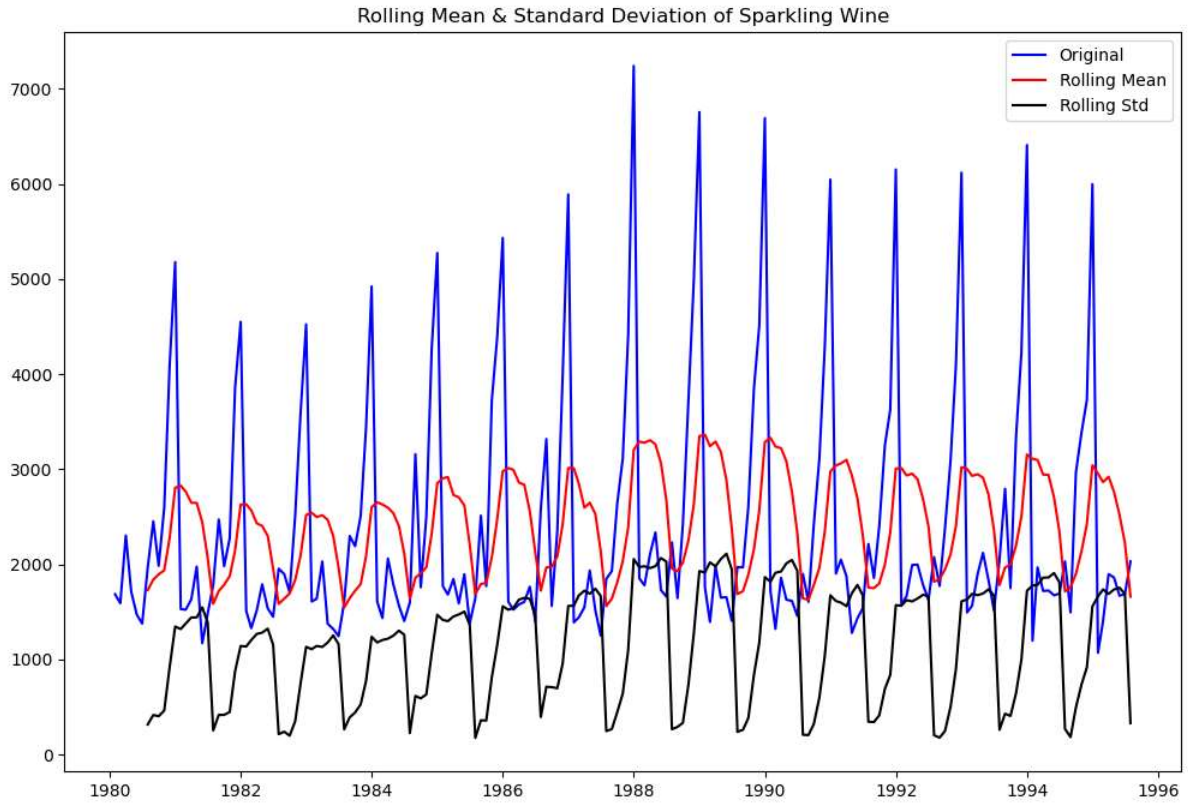


```
Results of Dickey-Fuller Test:
Test Statistic                -8.059643e+00
p-value                        1.656242e-12
#Lags Used                     1.200000e+01
Number of Observations Used    1.740000e+02
Critical Value (1%)           -3.468502e+00
Critical Value (5%)           -2.878298e+00
Critical Value (10%)          -2.575704e+00
dtype: float64
```

Figure 40: Stationary of Rose Wine

**Stationarity of Sparkling Wine Dataset:**

- Augmented Dicky-Fuller Test was applied to the whole Sparkling dataset
    - We found, p-value = 0.60106
    - Here, p-value > 0.05
    - We fail to reject the Null Hypothesis and hence conclude that Sparkling Wine Time Series is Not Stationary
    - We take 1 level of differencing and check again for Stationarity
    - Now, p-value = 0.000000
    - Now, p-value < 0.05
    - Now, we reject the Null Hypothesis and conclude that Sparkling Time Series is Stationary with a lag of 1



```
Results of Dickey-Fuller Test:
Test Statistic                -38.042867
p-value                         0.000000
#Lags Used                     10.000000
Number of Observations Used   176.000000
Critical Value (1%)            -3.468062
Critical Value (5%)            -2.878106
Critical Value (10%)           -2.575602
dtype: float64
```

Figure 41: Stationary of Sparkling Wine

## ACF & PACF:

**Auto-Correlation Function (ACF) –**

- Autocorrelation refers to how correlated a time series is with its past values.
- 'Auto' part of Autocorrelation refers to Correlation of any time instance with its previous time instance in the SAME Time Series
- ACF is the plot used to see the correlation between the points, up to and including the lag unit.
- ACF indicates the value of 'q' - which is the Moving Average parameter in ARIMA / SARIMA models

**Partial Auto-Correlation Function (PACF) –**

- Partial Autocorrelation refers to how correlated a time series is with its past lag values.
- PACF is the plot used to see the correlation between the lag points
- PACF indicates the value of 'p' - which is the Auto-Regressive parameter in ARIMA / SARIMA models

## Rose Wine:

- Observing the cut-offs in ACF and PACF plots for Rose dataset, we get –

  FOR ARIMA: p = 2, q = 2 and difference d = 1

  FOR SARIMA: p = 2, q = 2, d = 1 and P = 2, D = 1, Q = 2, Seasonality=12

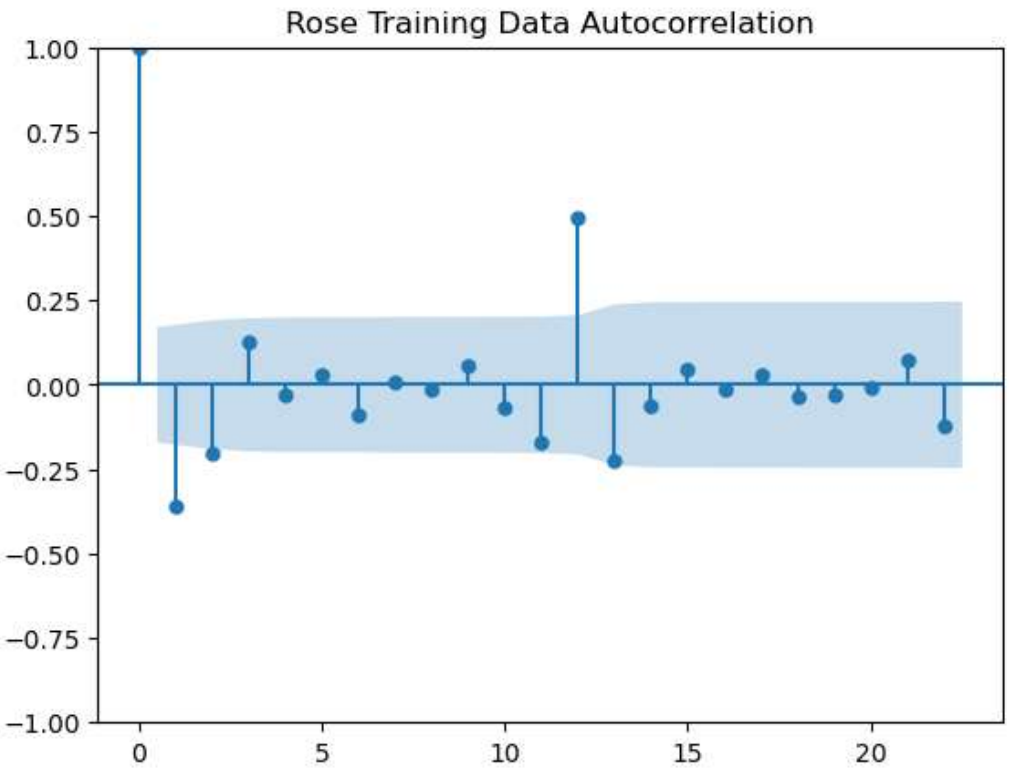



Figure 42: ACF of Auto ARIMA of Rose Wine

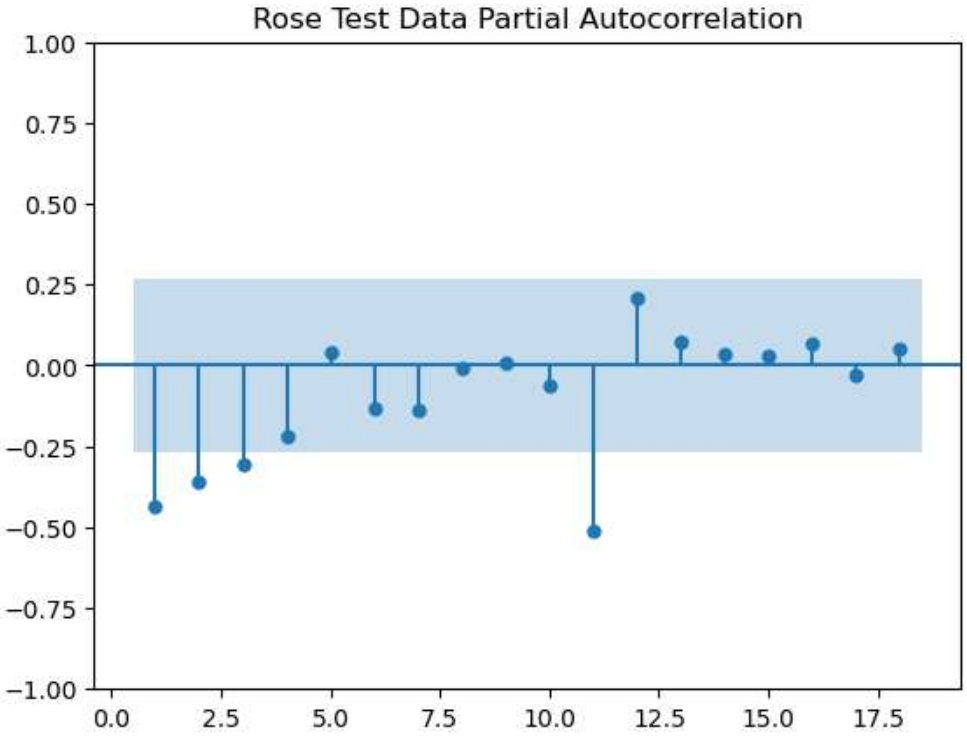



Figure 43: PACF of Auto ARIMA of Rose Wine

**ACF & PACF – Sparkling Wine:**

- Observing the cut-offs in ACF and PACF plots for Sparkling dataset, we get –
  FOR ARIMA: p = 0, q = 0 and difference d = 1
  FOR SARIMA: p = 0, q = 0, d = 1 and P = 0, 1, 2, 3 | D = 0, Q = 1, 2, 3

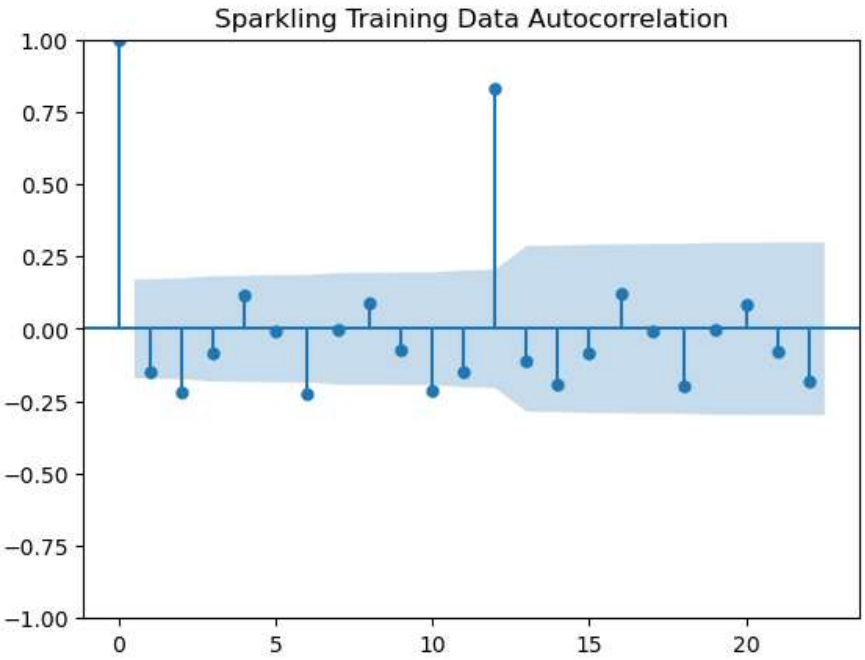



Figure 44: ACF of Auto ARIMA of Sparkling Wine

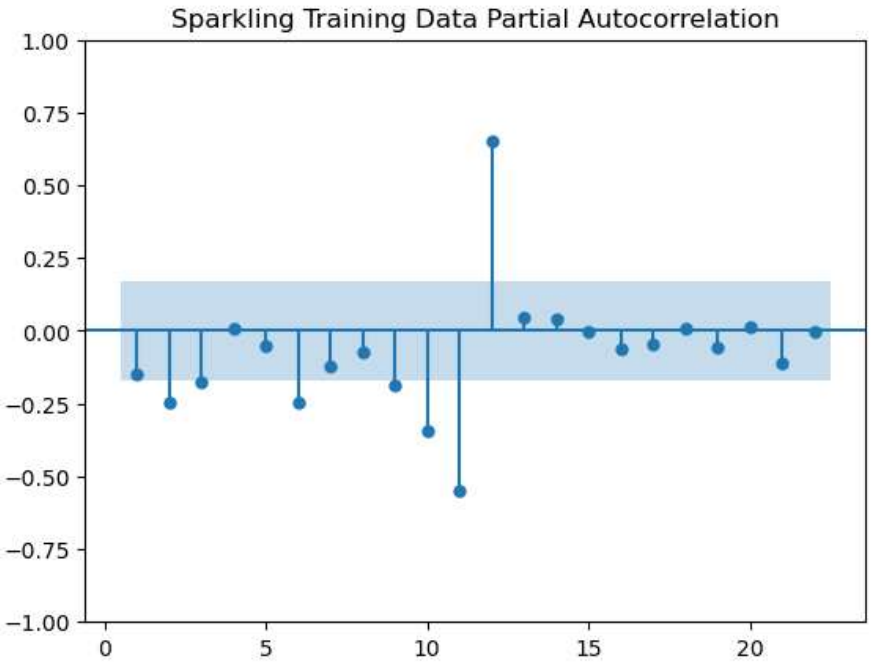



Figure 45: PACF of Auto ARIMA of Sparkling Wine

**Auto ARIMA:**

- We create a grid of all possible combinations of (p, d, q)
- Range of p = Range of q = 0 to 3, Constant d = 1

```
Some parameter combinations for the Model..
Model: (0, 1, 1)
Model: (0, 1, 2)
Model: (0, 1, 3)
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (1, 1, 3)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)
Model: (2, 1, 3)
Model: (3, 1, 0)
Model: (3, 1, 1)
Model: (3, 1, 2)
Model: (3, 1, 3)
```

Figure 46: Model Combinations of Rose & Sparkling Wine

- We fit ARIMA models to each of these combinations for both datasets
- We choose the combination with the least Akaike Information Criteria (AIC)
- We fit ARIMA to this combination of (p, d, q) to the Train set and forecast on the Test set
- Finally, we check the accuracy of this model by checking RMSE of Test set

For Rose, Best Combination with Least AIC is - (p, d, q) is (2, 1, 3)

RMSE of Rose Wine - Auto ARIMA : 36.807

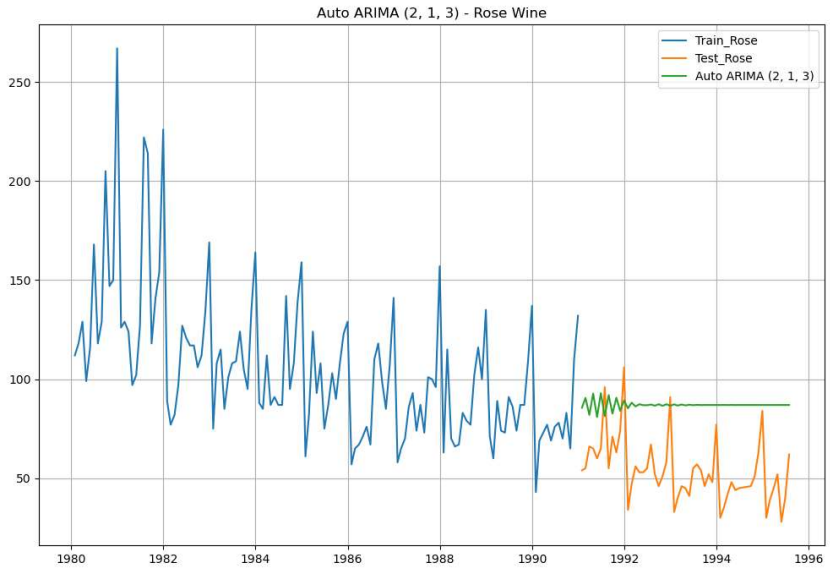



Figure 47: Auto ARIMA of Rose Wine

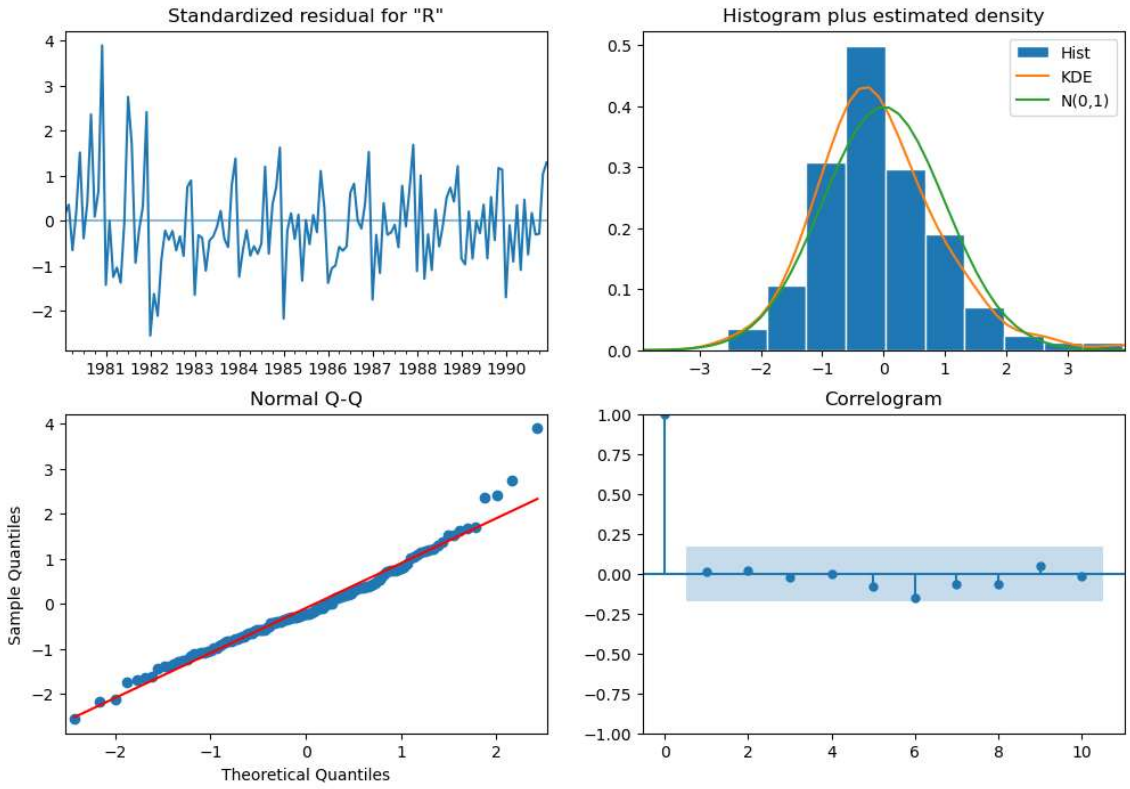



Figure 48: Diagnostic Plot of Auto ARIMA of Rose Wine

For Sparkling, Best Combination with Least AIC is - (p, d, q) is (2, 1, 2)

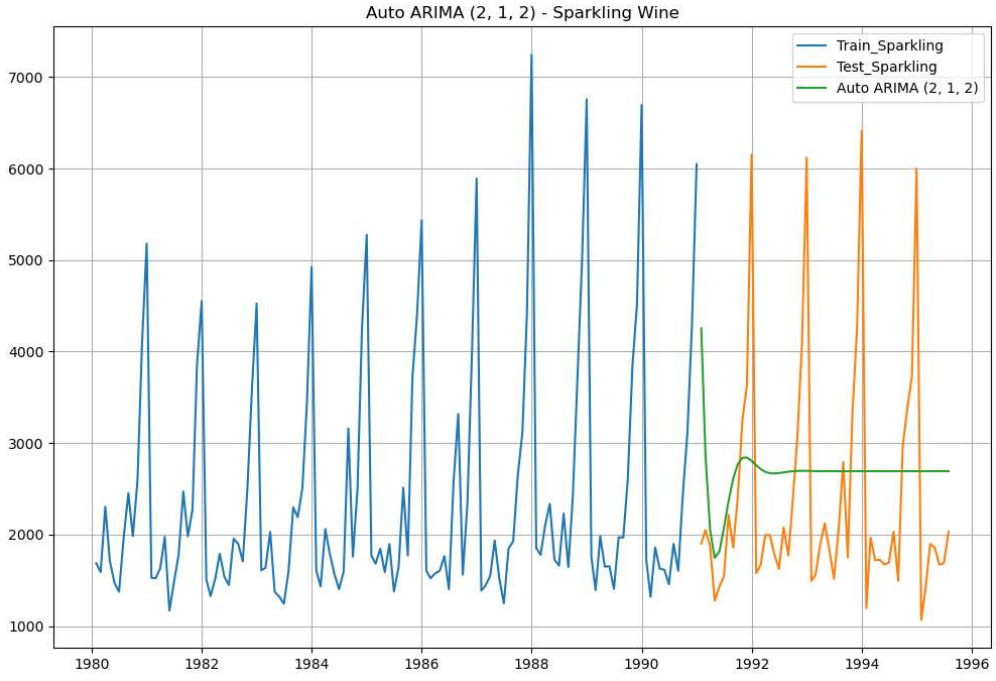



Figure 49: Auto ARIMA of Sparkling Wine

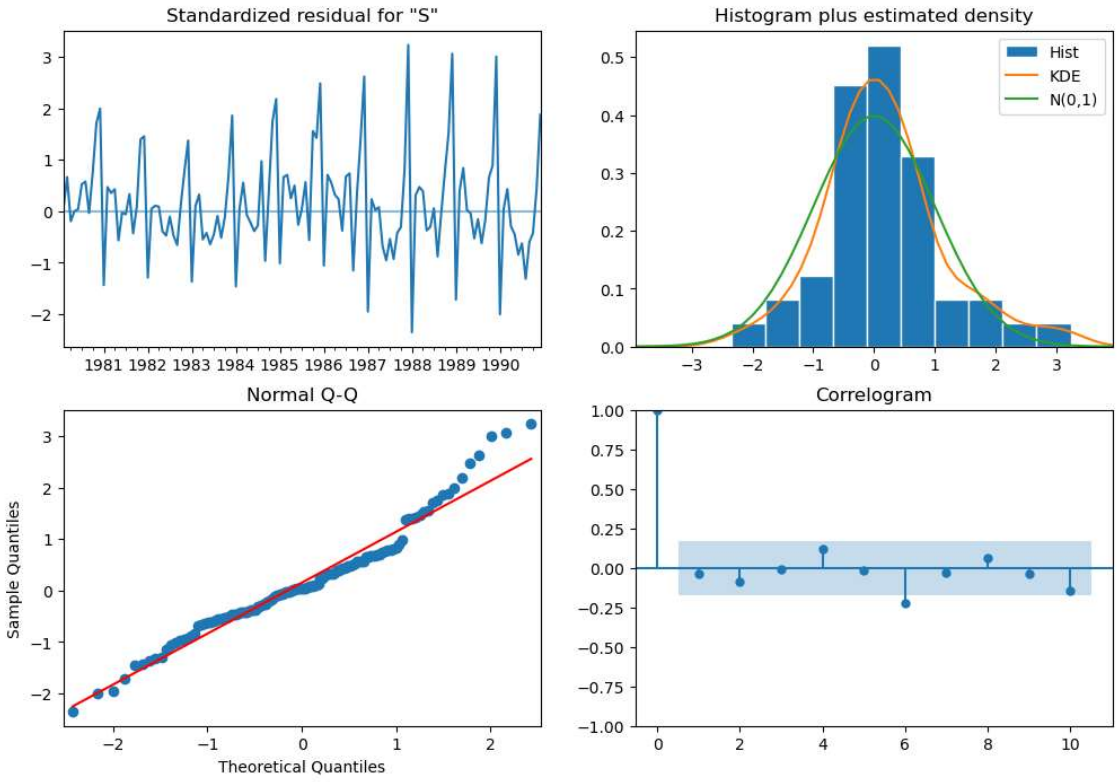



Figure 50: Diagnostic Plot of Auto ARIMA of Sparkling Wine

**Manual ARIMA:**

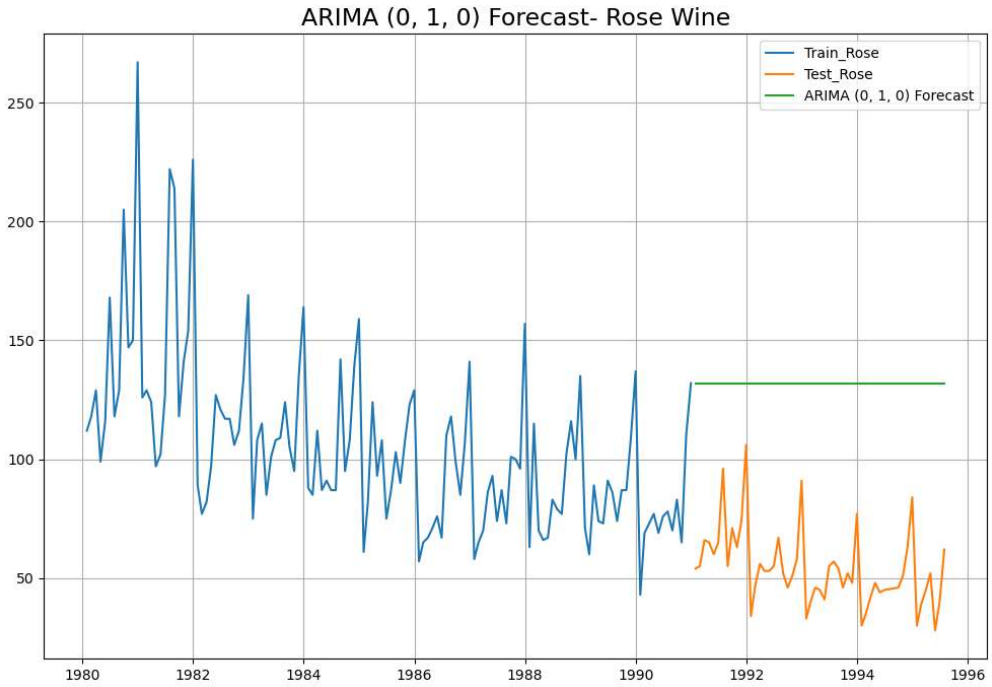



Figure 51: Manual ARIMA of Rose Wine

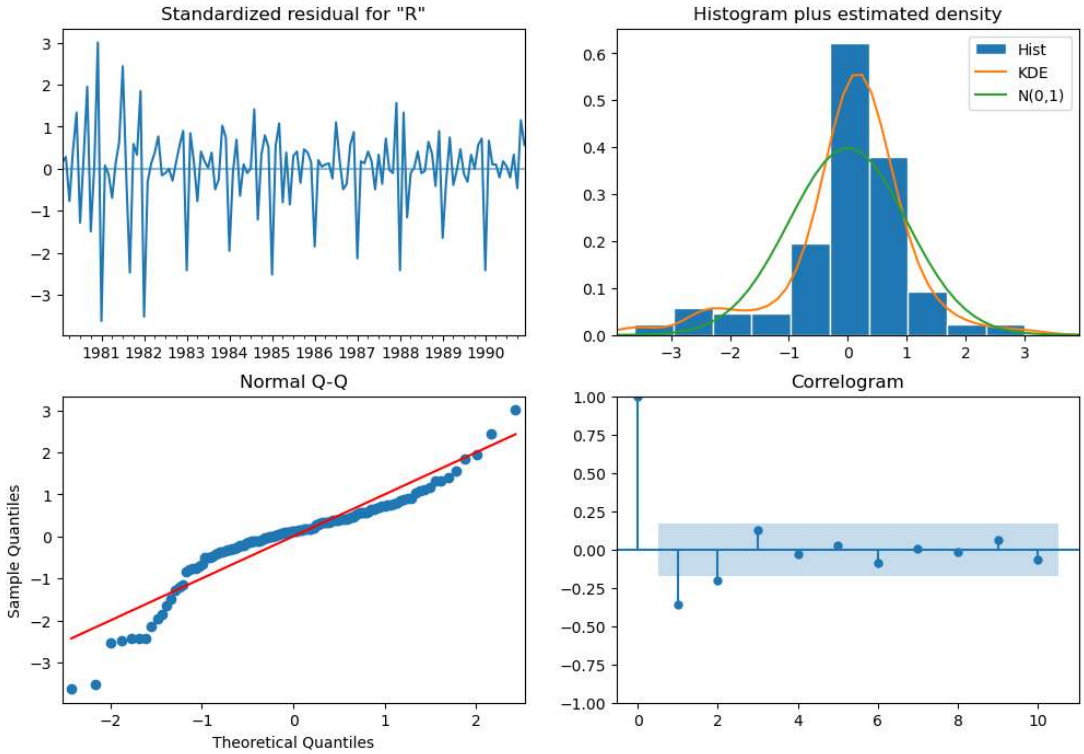



Figure 52: Diagnostic Plot of Manual ARIMA of Rose Wine

RMSE of Rose Wine - Manual ARIMA : 79.718

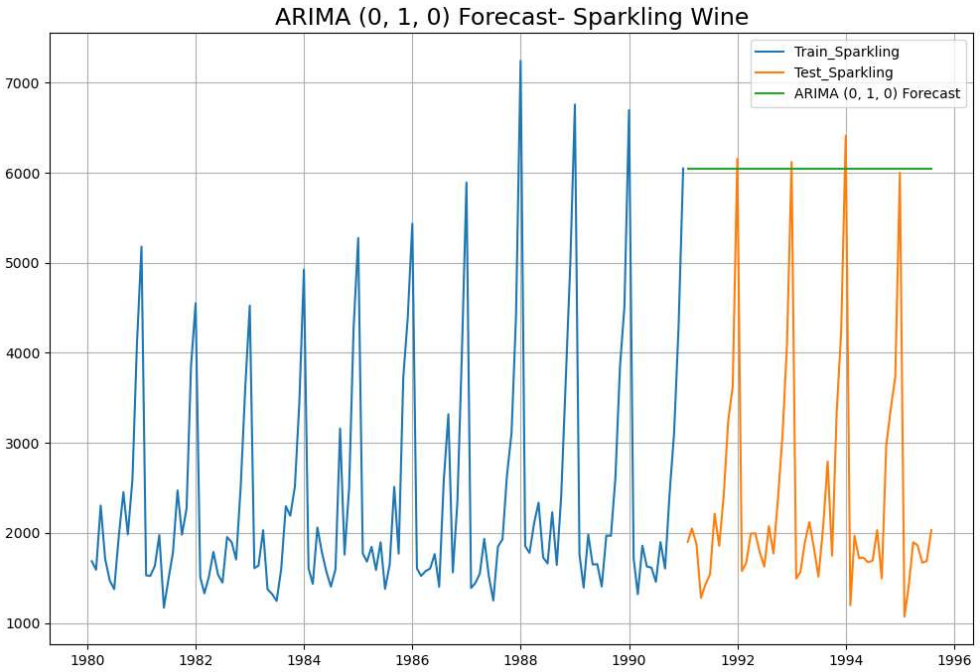



Figure 53: Manual ARIMA of Sparkling Wine

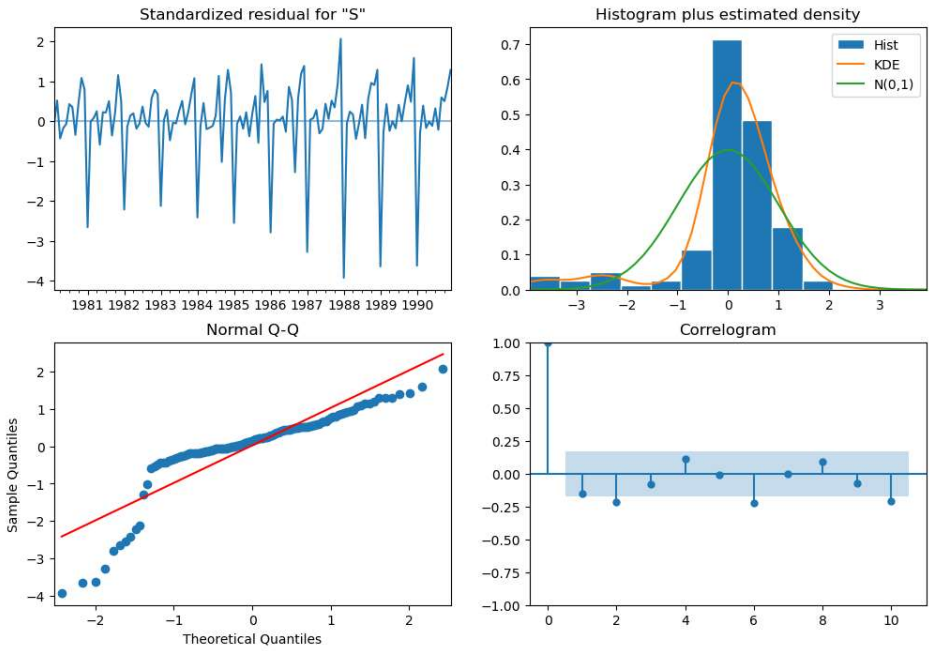



Figure 54: Diagnostic Plot of Manual ARIMA of Sparkling Wine

RMSE of Sparkling Wine - Manual ARIMA: 3864.279

**Auto SARIMA:**

- We create a grid of all possible combinations of (p, d, q) along with Seasonal (P, D, Q) & Seasonal ity of 12 (for both datasets)
- Range of p = Range of q = 0 to 3, Constant d = 1
- Range of Seasonal P = Range of Seasonal Q = 0 to 3, Constant D = 1, Seasonality m = 12

```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

Figure 55: Model Combinations of Rose & Sparkling Wine

- We fit SARIMA models to each of these combinations and select with least AIC
- We fit SARIMA to this best combination of (p, d, q) (P, D, Q, m) to the Train set and forecast on t he Test set. Then, we check accuracy using RMSE on Test set

**Auto SARIMA - Rose Wine – (1,1,1) (2,0,3,6):**

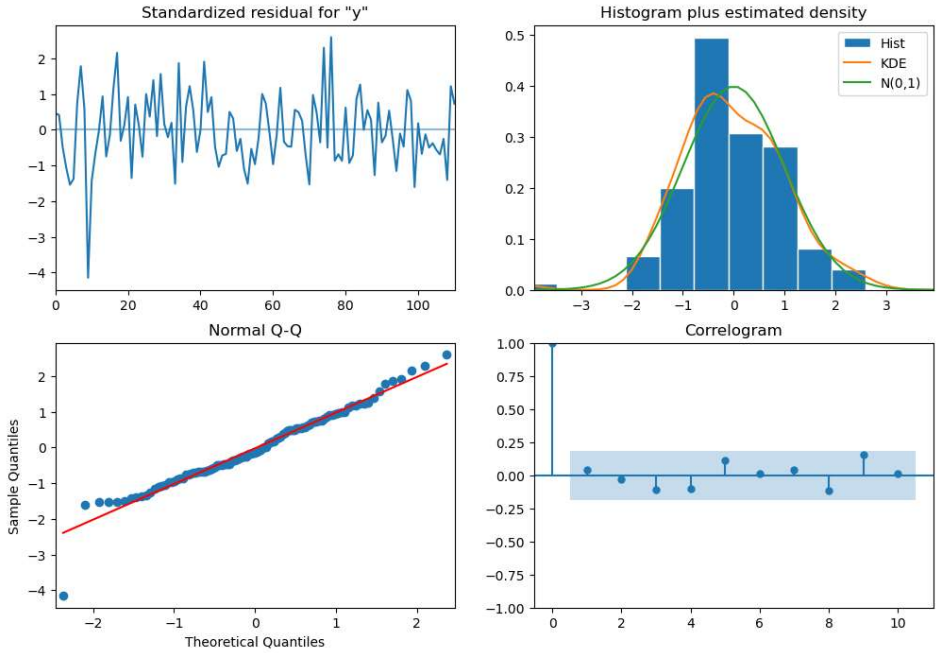



Figure 56: Diagnostic Plot of Auto SARIMA of Rose Wine – (1,1,1) (2,0,3,6)

RMSE of Test set for rose wine : 29.76

**Auto SARIMA - Rose Wine – (1,1,2) (2,0,2,12):**

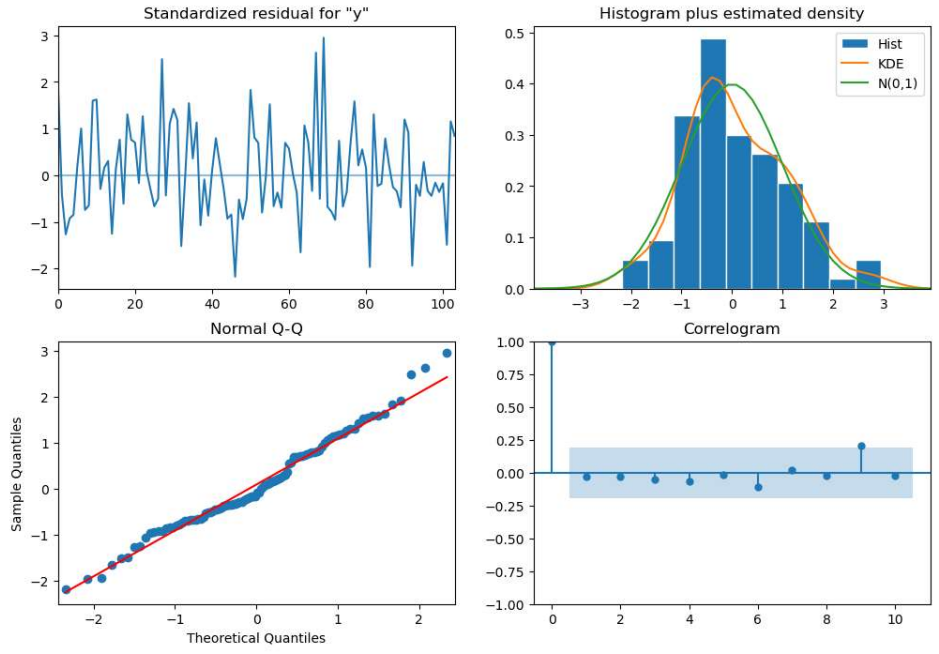



Figure 57: Diagnostic Plot of Auto SARIMA of Rose Wine – (1,1,2)) (2,0,2,12)

RMSE of Test set for rose wine : 26.928

**Auto SARIMA - Sparkling Wine – (0,1,1) (2,0,2,6):**

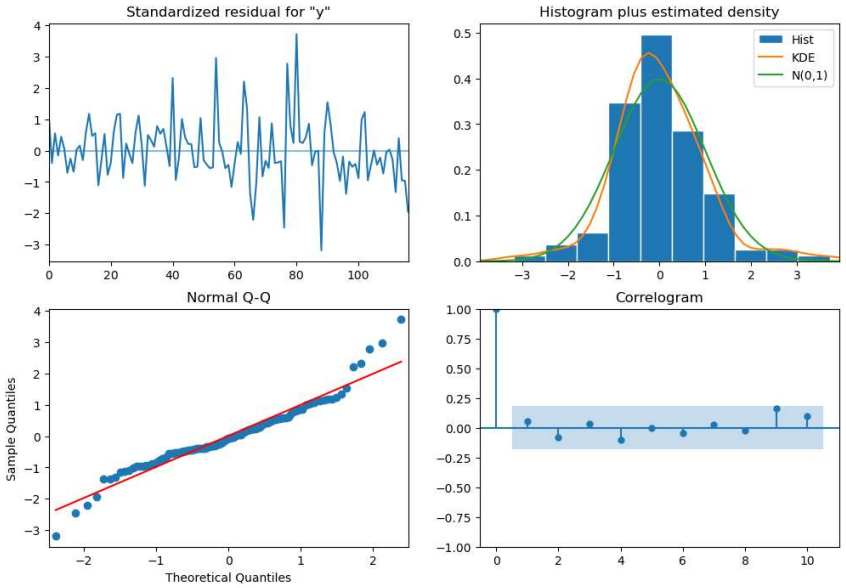



Figure 58: Diagnostic Plot of Auto SARIMA of Sparkling Wine – (0,1,1)) (2,0,2,6)

RMSE of Test set for Sparkling wine : 631.373

**Auto SARIMA - Sparkling Wine – (1,1,2) (2,0,2,12):**

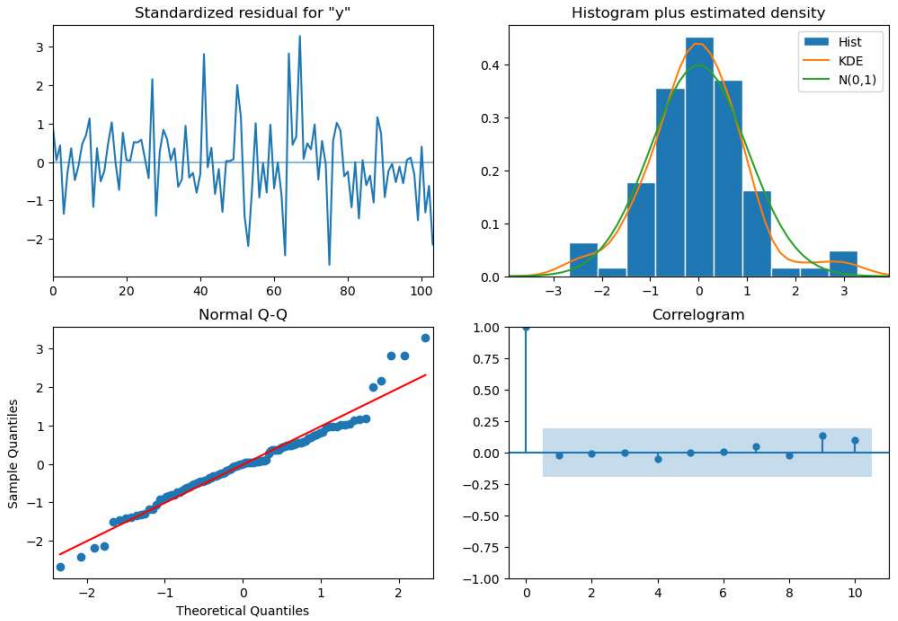



Figure 59: Diagnostic Plot of Auto SARIMA of Sparkling Wine – (1,1,2)) (2,0,2,12)

RMSE of Test set for Sparkling wine: 548.080

**Manual SARIMA - :**

**Rose Wine – (0,1,0) (1,1,3,6) :**

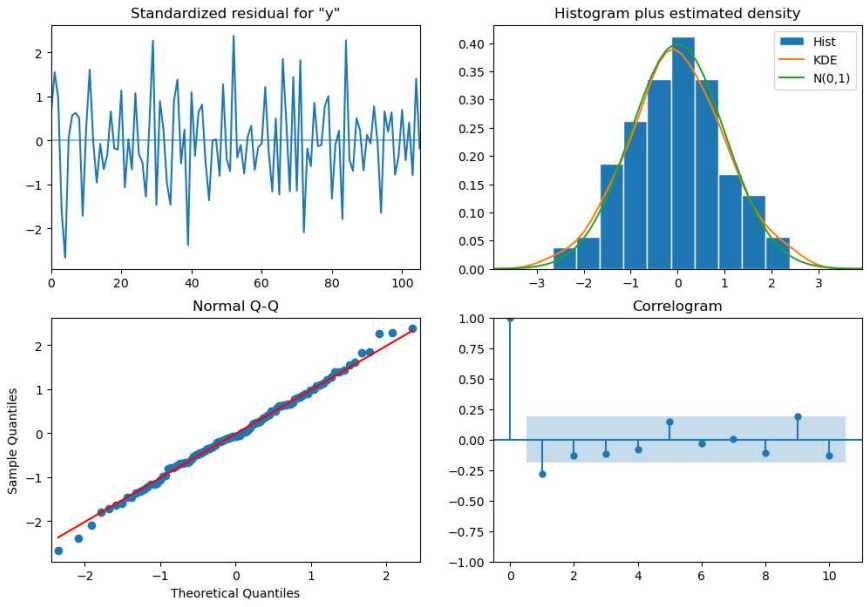



Figure 60: Diagnostic Plot of Manual SARIMA of Rose Wine – (0,1,0) (1,1,3,6)

RMSE of Rose Wine – Manual SARIMA: 37.874

**Rose Wine – (2,1,2) (2,1,2,12) :**

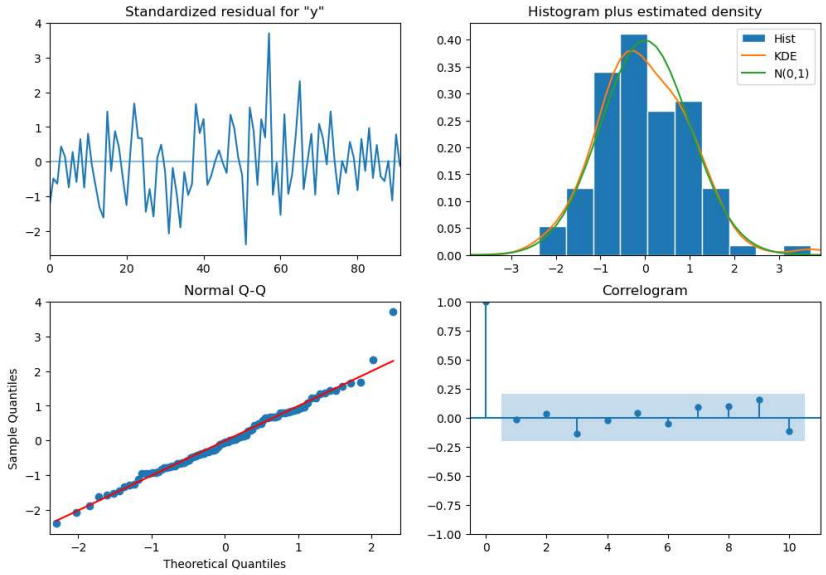



Figure 61: Diagnostic Plot of Manual SARIMA of Rose Wine – (2,1,2) (2,1,2,12)

RMSE of Rose Wine – Manual SARIMA: 37.874

**Sparkling Wine – (0,1,0) (1,1,3,6) :**

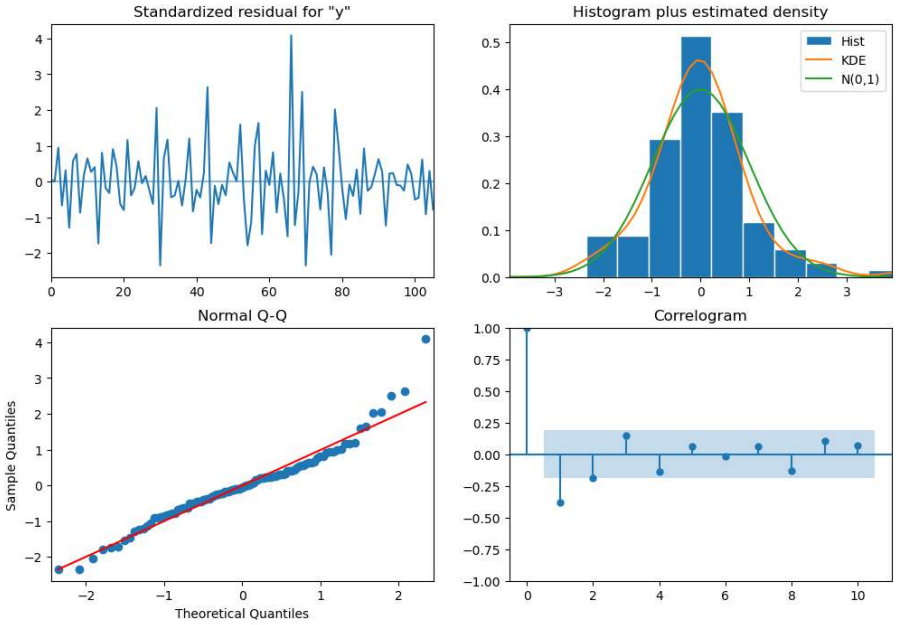



Figure 62: Diagnostic Plot of Manual SARIMA of Sparkling Wine – (0,1,0) (1,1,3,6)

RMSE of Sparkling Wine – Manual SARIMA: 1914.641

**Sparkling Wine – (2,1,2) (2,1,2,12) :**

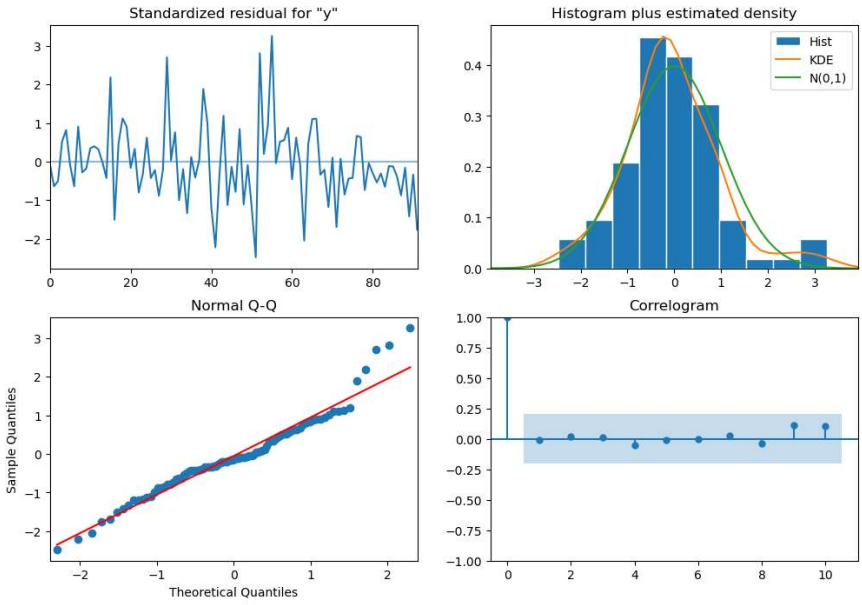



Figure 63: Diagnostic Plot of Manual SARIMA of Sparkling Wine – (2,1,2) (2,1,2,12)

RMSE of Sparkling Wine – Manual SARIMA: 1914.641

## Model Evaluation for next 12 Months:

### Rose Wine:

| | Rose Test RMSE |
|---|---|
| 2pointTrailingMovingAverage | 11.530000 |
| Triple Exponential Smoothing - Additive Model | 14.260000 |
| 4pointTrailingMovingAverage | 14.450000 |
| 6pointTrailingMovingAverage | 14.570000 |
| 9pointTrailingMovingAverage | 14.730000 |
| Linear Regression | 15.270000 |
| Double Exponential Smoothing | 15.270000 |
| Simple Average Model | 15.760000 |
| Triple Exponential Smoothing - Multiplicative Model | 22.100000 |
| Auto SARIMA (1,1,2)(2,0,2,12) | 26.928362 |
| Auto SARIMA (1,1,1)(2,0,3,6) | 29.760281 |
| Auto ARIMA(2,1,3) | 36.807150 |
| Mannual SARIMA (0,1,0)(1,1,3,6) | 37.874034 |
| Mannual SARIMA (2,1,2)(2,1,2,12) | 37.874034 |
| Simple Exponential Smoothing | 41.360000 |
| Manual ARIMA(0,1,0) | 79.718773 |

Figure 64: RSME Values of Rose Wine in Ascending Order

The Best Model so far is 2 point Trailing Moving Average.

### 2 Point Trailing Moving Average:

| YearMonth | Forecast |
|---|---|
| 31-08-1995 | 48.80556 |
| 30-09-1995 | 48.80556 |
| 31-10-1995 | 48.80556 |
| 30-11-1995 | 48.80556 |
| 31-12-1995 | 48.80556 |
| 31-01-1996 | 48.80556 |

| | |
|---|---|
| 29-02-1996 | 48.80556 |
| 31-03-1996 | 48.80556 |
| 30-04-1996 | 48.80556 |
| 31-05-1996 | 48.80556 |
| 30-06-1996 | 48.80556 |
| 31-07-1996 | 48.80556 |

Table 1: Rose Wine Forecast for next 12 months - 2 point Moving Average

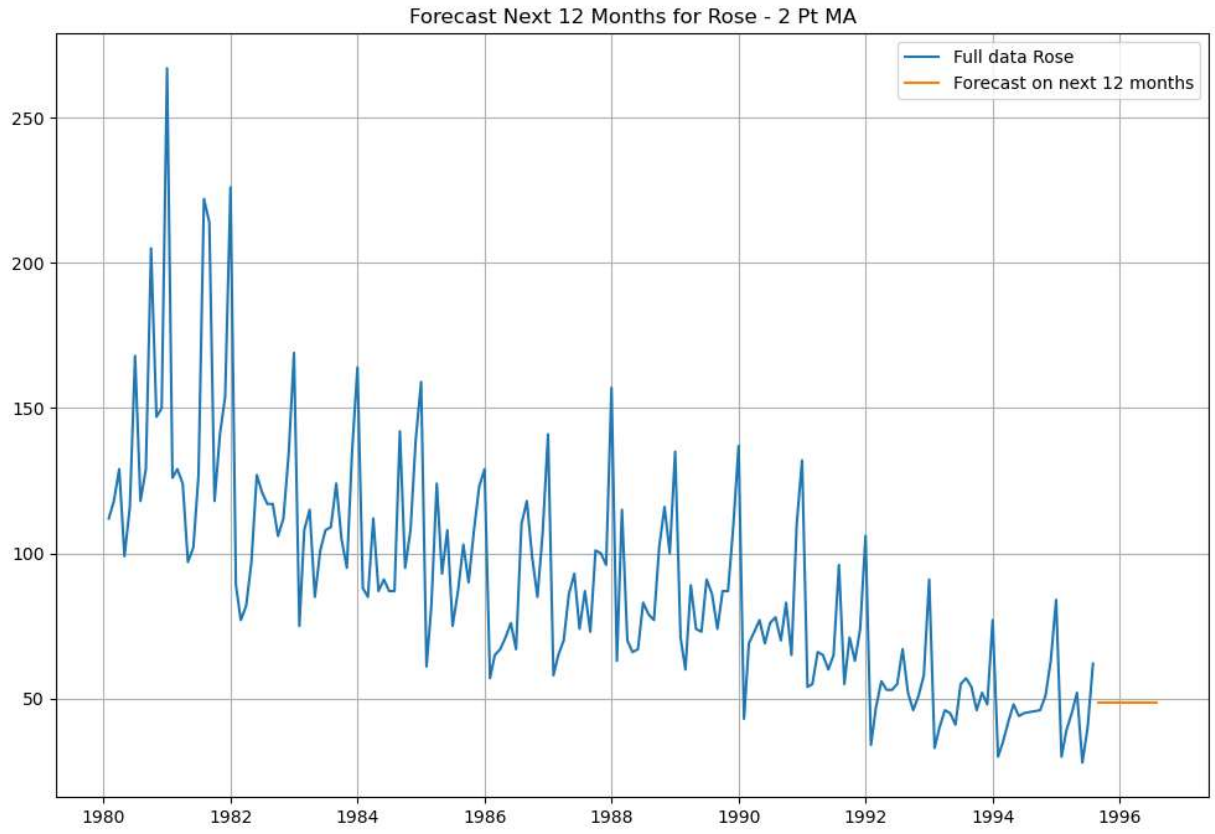



Figure 65: Rose Wine Forecast for next 12 months - 2 point Moving Average

Based on the above forecast, it can be said that the model is not predicting well. So, We forecast on the 2nd best model – Triple Exponential Smoothing – Additive Model

## Triple Exponential Smoothing – Additive Model:

Here, we have a scenario where our training data was stationary but our full data was not stationary. So, we will use the same parameters as our training data but with adding a level of differencing which is needed for the data to be stationary.

| YearMonth | Forecast |
|---|---|
| 31-08-1995 | 49.83258 |
| 30-09-1995 | 46.62195 |
| 31-10-1995 | 45.41716 |
| 30-11-1995 | 60.07118 |
| 31-12-1995 | 98.23646 |
| 31-01-1996 | 13.89411 |
| 29-02-1996 | 24.13636 |
| 31-03-1996 | 31.66632 |
| 30-04-1996 | 24.43448 |
| 31-05-1996 | 27.91752 |
| 30-06-1996 | 33.38828 |
| 31-07-1996 | 43.83555 |

Table 2: Rose Wine Forecast - next 12 months - Triple Exponential Smoothing – Additive

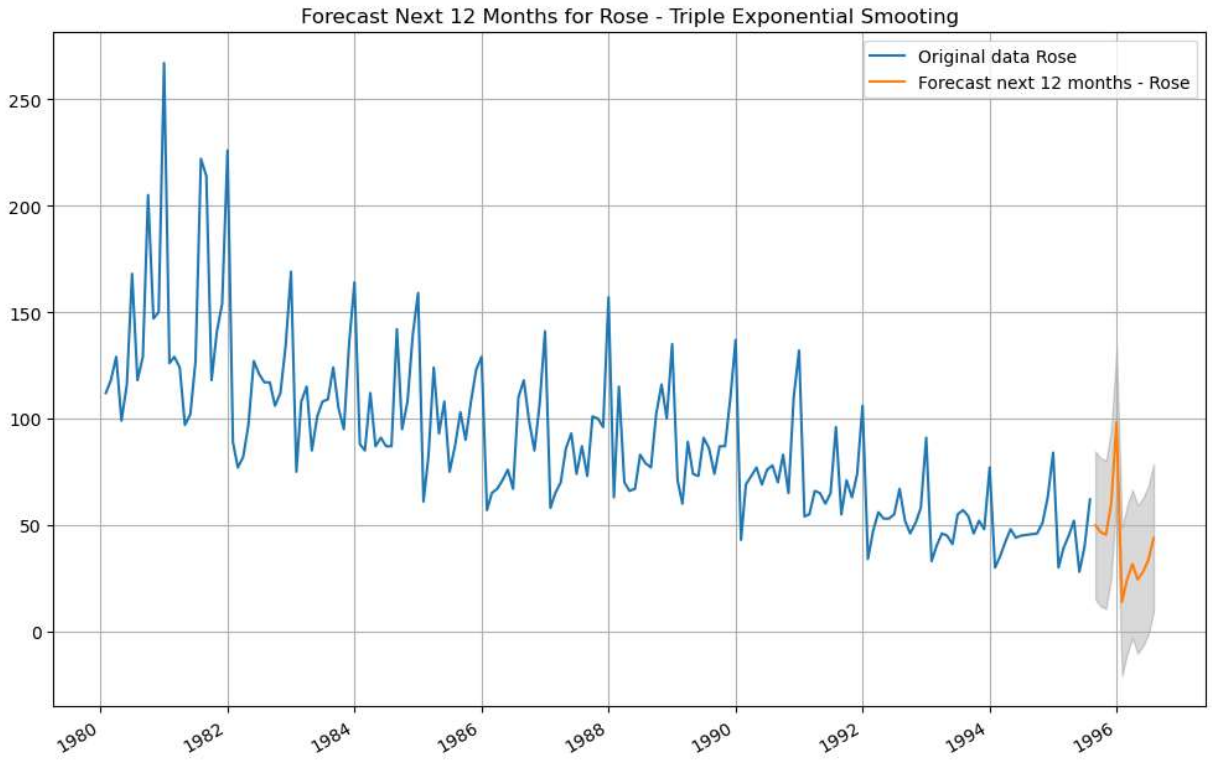



Figure 66: Rose Wine Forecast - next 12 months - Triple Exponential Smoothing – Additive

**Sparkling Wine:**

| | Sparkling Test RMSE |
|---|---|
| Triple Exponential Smoothing - Additive Model | 380.09 |
| Triple Exponential Smoothing - Multiplicative Model | 386.84 |
| Auto SARIMA (1,1,2)(2,0,2,12) | 548.08 |
| Auto SARIMA (0,1,1)(2,0,2,6) | 631.37 |
| 2pointTrailingMovingAverage | 813.40 |
| 4pointTrailingMovingAverage | 1156.59 |
| Simple Average Model | 1275.07 |
| 6pointTrailingMovingAverage | 1283.93 |
| Auto ARIMA(2,1,2) | 1299.98 |
| 9pointTrailingMovingAverage | 1346.28 |
| Linear Regression | 1389.14 |
| Simple Exponential Smoothing | 1595.21 |
| Mannual SARIMA (0,1,0)(1,1,3,6) | 1914.64 |
| Mannual SARIMA (2,1,2)(2,1,2,12) | 1914.64 |
| Manual ARIMA(0,1,0) | 3864.28 |
| Double Exponential Smoothing | 5291.88 |

Figure 67: RSME Values of Sparkling Wine in Ascending Order

The Best Model so far is Triple Exponential Smoothing – Additive Model.

**Triple Exponential Smoothing – Additive Model:**

| YearMonth | Forecast |
|---|---|
| 31-08-1995 | 1877.432 |
| 30-09-1995 | 2405.285 |
| 31-10-1995 | 3242.105 |
| 30-11-1995 | 3922.189 |
| 31-12-1995 | 6118.502 |
| 31-01-1996 | 1262.619 |
| 29-02-1996 | 1592.138 |
| 31-03-1996 | 1831.653 |
| 30-04-1996 | 1806.47 |

| | |
|---|---|
| 31-05-1996 | 1651.723 |
| 30-06-1996 | 1586.507 |
| 31-07-1996 | 1977.014 |

Table 3: Sparkling Wine Forecast - next 12 months - Triple Exponential Smoothing – Additive

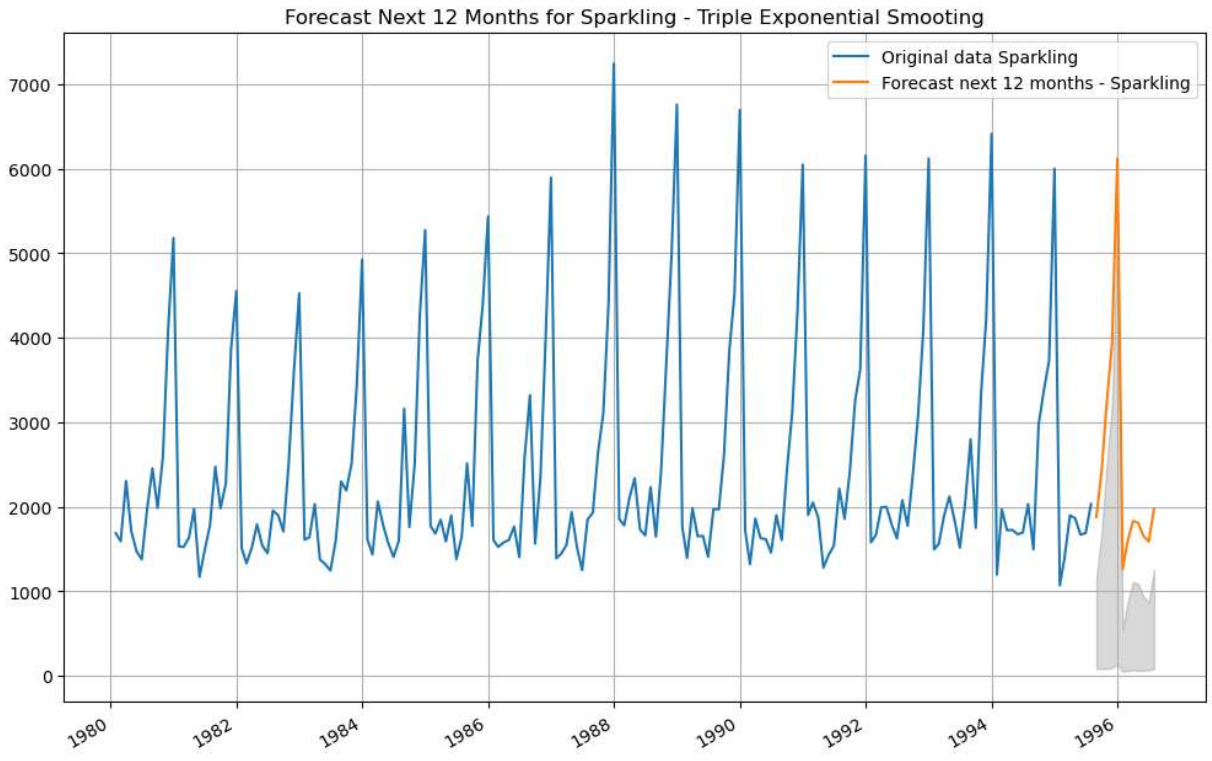



Figure 68: Sparkling Wine Forecast - next 12 months - Triple Exponential Smoothing – Additive

## Actionable Insights & Recommendations:

### Rose Wine - Comments:

- Sales of rose wine have steadily declined since 1980.
- This indicates a drop in popularity of this wine variety.
- Sales typically increase in the fourth quarter of the year (October-December).
    - This might be due to the Holiday season in this period.
    - Highest peak in sales is seen in Dec every year.
- Sales experience a sharp decline in the first quarter of each year beginning in January.

- This might be due to the after effect or hangover of Holidays.
- Sales slowly pick up only after May-June.

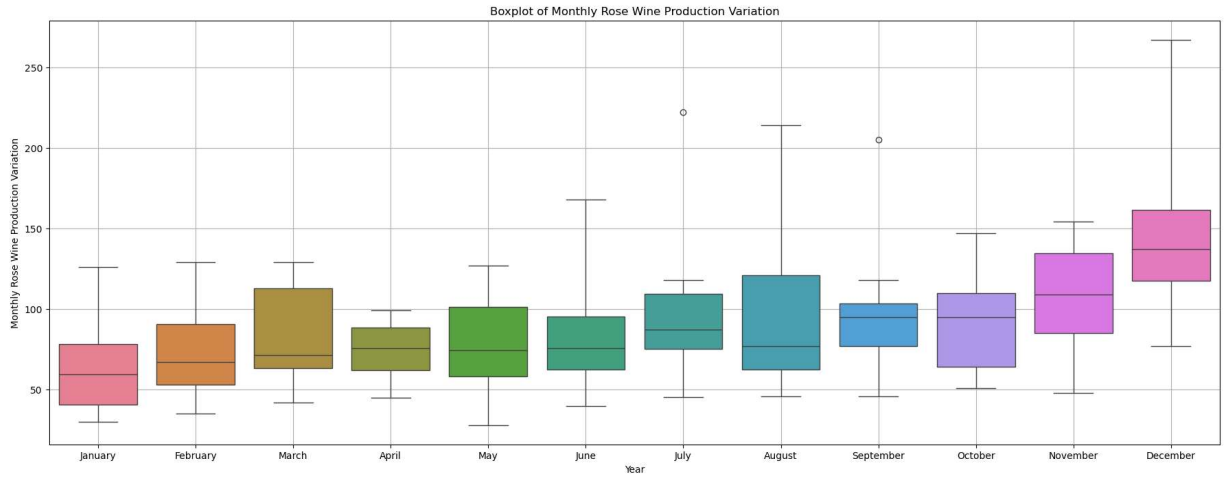



Figure 69: Monthly Sales of Rose Wine

**Rose Wine Sales - Forecast Models:**

- Top 2 best models as per lowest Test RMSE were found to be,
  - 2 Pt Moving Average
  - Triple Exponential Smoothing (Holt-Winters) - Additive Model
- When utilized for forecasting, the 2-point moving average approach does not appear to produce accurate forecasts. After a few repetitions, the forecast values level out.
- Triple Exponential Smoothing (Holt-Winters), appears to provide a forecast that is consistent with the data. Therefore, we have decided to use Holt-Winters for our final Rose Wine Sales prediction.

**Rose Wine Sales - Suggestions:**

- To start with, the holiday season is quickly approaching, and sales are expected to rise, peaking sharply in December. Thus, the company need to restock.
- However, a longer period of time and more data analysis should be done to look at declining rose wine sales.
- Company can rebrand its Rose variant along-with a new Wine-master.
- The business should launch aggressive deals and advertising efforts to capitalize on the impending rise from August to October.
- This will entice first time Wine drinkers and fence sitters (who don't have specific loyalties to any particular brand).

- Should sales not show a discernible increase trend by December, the company's options are as follows: either engage in research and development, or consider terminating current variant and developing a whole new product.

**Sparkling Wine - Comments:**

- Sparkling wine sales don’t show any upward or downward trend,
    - This shows flat sales over long term range.
- Moreover, a sharp increase in sales is observed in the final quarter of each year, from October to December.
    - This might be due to the Holiday season in this period.
    - Highest peak in sales is seen in Dec every year - Dec sales are almost 3 times of Sep sales
- Just like with rose wine, the first quarter of every year from January to March sees an instantaneous, crashing decline in sales, even in sparkling wine.
    - This might be due to the after effect or hangover of Holidays.
- Sales only gradually increase from July to August.

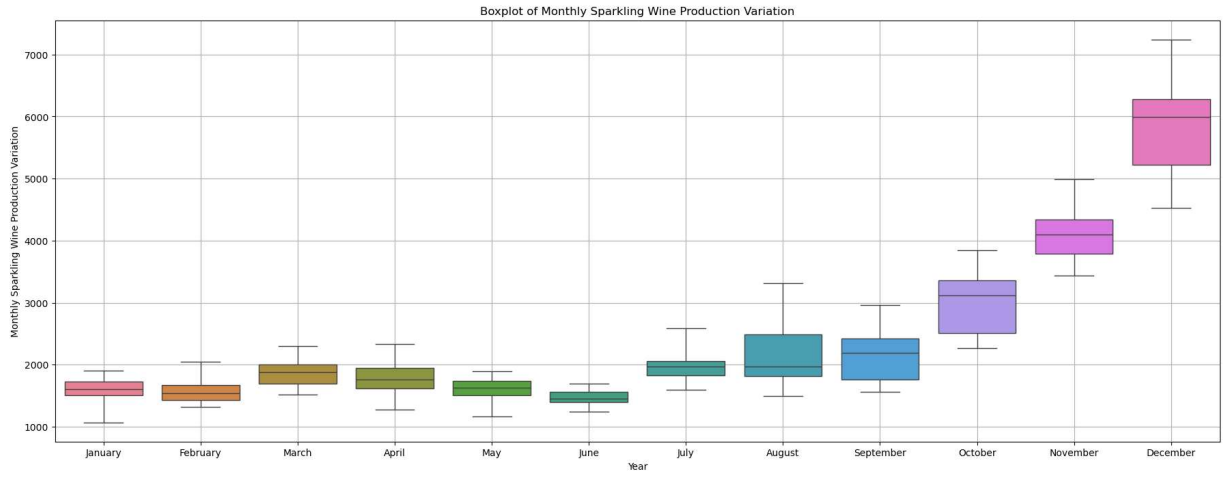



Figure 70: Monthly Sales of Sparkling Wine

**Sparkling Wine Sales - Forecast Models:**

- Triple Exponential Smoothing (Holt-Winters) Models perform the best on Sparkling dataset, consideri ng the least RMSE on Test data.

- Finally, for forecast of Sparkling Wine Sales - we choose Triple Exponential Smoothing (Holt-Winters) with Additive Seasonality and Additive Damped Trend.

**Sparkling Wine Sales - Suggestions:**

- Even for sparkling, the holiday season is quickly approaching, and sales are expected to rise, peaking sharply in December. Thus, the company should restock.
- Sparkling wine is very popular over the holidays, as seen by its strong sales.
- Ads are advised during the months of October through December, but there's no need to introduce any deals at this time. This will further increase sales.
- Sparkling wines are typically connected to festivities and are primarily meant to pop.
- For bursting, a unique designer bottle might be introduced at a lower cost. Profits will be maximized by this.
- There is no discernible rise or fall in sales year over year.
- Even though holiday sales surges are unusual, more research has to be done on year-over-year sales in general. This in-depth examination should be done in the early part of January.