TIME SERIES

Time series is a technique to look in to the future. Using Data, it studies its trend and seasonality and try to learn from it and forecast the future.

SAHIL SAXENA [Date] DATA SCIENCE AND BUSINESS ANALYST

CONTENT

TABLES.

FIGURES

PROBLEM:

For this particular assignment, the data of different types of wine sales in the 20th century is to be analysed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyse and forecast Wine Sales in the 20th century.

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

SOLUTION:

SPARKLING WINE DATA

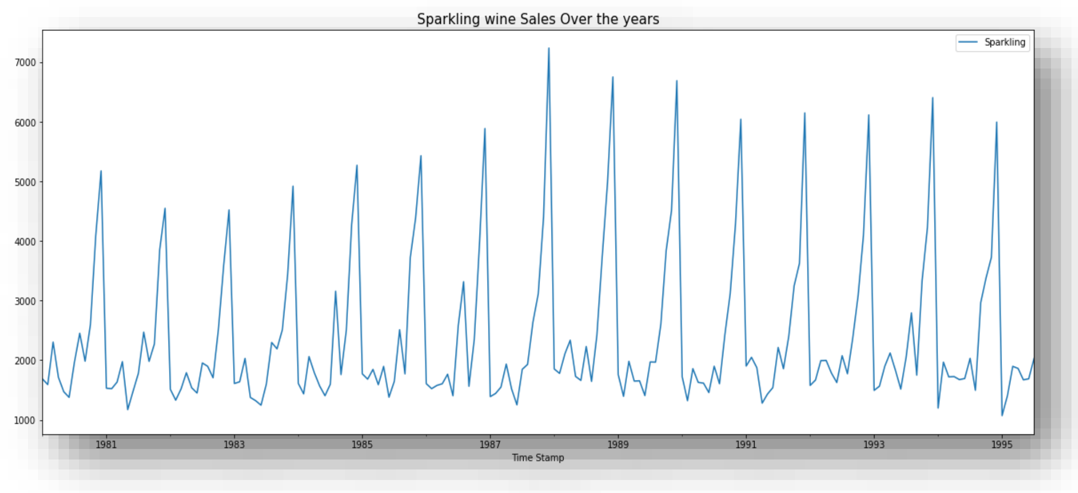
ORIGNAL DATA		
SR NO.	YearMonth	Sparkling
0	1980-01	1686
1	1980-02	1591
2	1980-03	2304
3	1980-04	1712
4	1980-05	1471
…	…	…
182	1995-03	1897
183	1995-04	1862
184	1995-05	1670
185	1995-06	1688
186	1995-07	2031

DATA AS TIME SERIES	
Time Stamp	Sparkling
31-01-1980	1686
29-02-1980	1591
31-03-1980	2304
30-04-1980	1712
31-05-1980	1471
…	…
31-03-1995	1897
30-04-1995	1862
31-05-1995	1670
30-06-1995	1688
31-07-1995	2031

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ------      --------------  -----
0   YearMonth   187 non-null    object
1   Sparkling   187 non-null    int64
2   Time Stamp  187 non-null    datetime64[ns]
dtypes: datetime64[ns](1), int64(1), object(
1)
memory usage: 4.5+ KB
```

TABLE 1. 1 – SPARKLING WINE ORIGNAL DATA AND DATA AS TIME SERIES

Original data for the sparkling wine had time column as object data type, so to read the data as time series we need to add one more column as time stamp and set it as index to make the data time series. Then there is no need to keep the Year Month column for further analysis, we will drop it from the data frame.



INFERENCE.

As we can see in the plot of sparkling wine that there is no trend in the data over the years but there is presence of seasonality is very much evident in the data.

Figure 1. 1 – SPARKLING WINE DATA VISUALIZATION

Sparkling wine data visualization for sales across the year in which Time Stamp represent the data collected regarding the sales of sparkling Wine from January 1980 to July 1995 and Sparkling column represents the sales of wine units in their respective months.

ROSE WINE

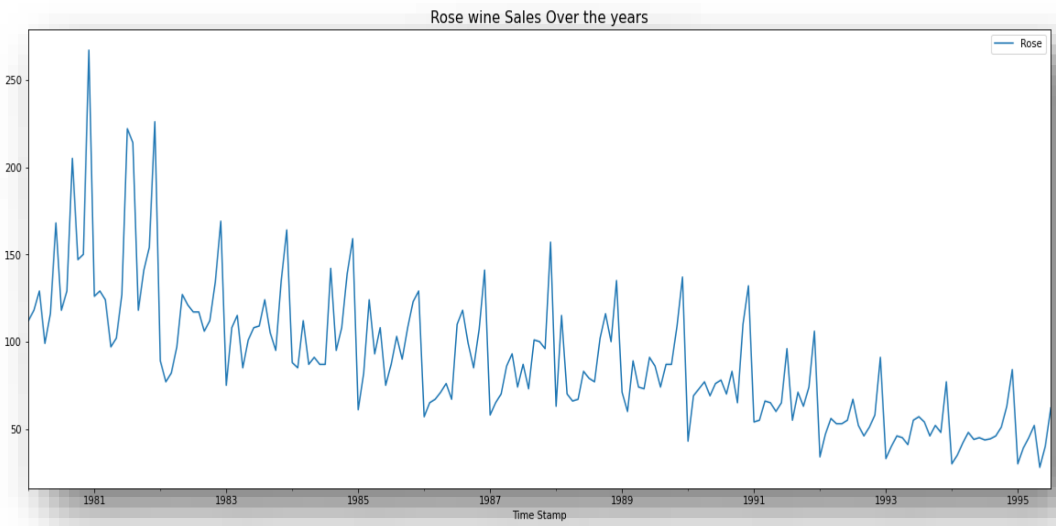
ORIGNAL DATA		
SR NO.	YearMonth	Rose
0	1980-01	112
1	1980-02	118
2	1980-03	129
3	1980-04	99
4	1980-05	116
...	...	...
182	1995-03	45
183	1995-04	52
184	1995-05	28
185	1995-06	40
186	1995-07	62

DATA AS TIME SERIES	
Time Stamp	Rose
31-01-1980	112
29-02-1980	118
31-03-1980	129
30-04-1980	99
31-05-1980	116
...	...
31-03-1995	45
30-04-1995	52
31-05-1995	28
30-06-1995	40
31-07-1995	62

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
0   YearMonth   187 non-null    object
1   Rose        185 non-null    float64
2   Time Stamp  187 non-null    datetime64[ns]
dtypes: datetime64[ns](1), float64(1), object(
1)
memory usage: 4.5+ KB
```

TABLE 1. 2 – ROSE WINE ORIGNAL AND DATA AS TIME SERIES

Original data for the Rose wine had time column named YearMonth as object data type, so to read the data as time series we need to add one more column as time stamp and set it as index to make the data time series. Then there is no need to keep the Year Month column for further analysis, we will drop it from the data frame.



INFERENCE.

As we can see in the plot of Rose wine that there is downward trend present in the data over the years and there is presence of seasonality is also very much evident.

Figure 1. 2 – ROSE WINE DATA VISUALIZATION

Rose wine data visualization for sales across the year in which Time Stamp represent the data collected regarding the sales of sparkling Wine from January 1980 to July 1995 and Rose column represents the sales of wine units in their respective months.

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

SOLUTION:

EXPLORATORY DATA ANALYSIS FOR SPARKLING WINE DATA.

DATA INFORMATION.

1. As we can see there is only one column Sparkling which is int64 datatype
2. There are total 187 rows in the data.
3. There are no null values in the data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 3 columns):
  #   Column      Non-Null  Count    Dtype
 ---  ------      --------------   -----
  0  Sparkling   non-null   187    int64
dtypes: datetime64[ns](1), int64(1), object(
1)
memory usage: 2.9 KB
```

DATA DESCRIPTION	
	Sparkling
count	187
mean	2402.417112
std	1295.11154
min	1070
25%	1605
50%	1874
75%	2549
max	7242

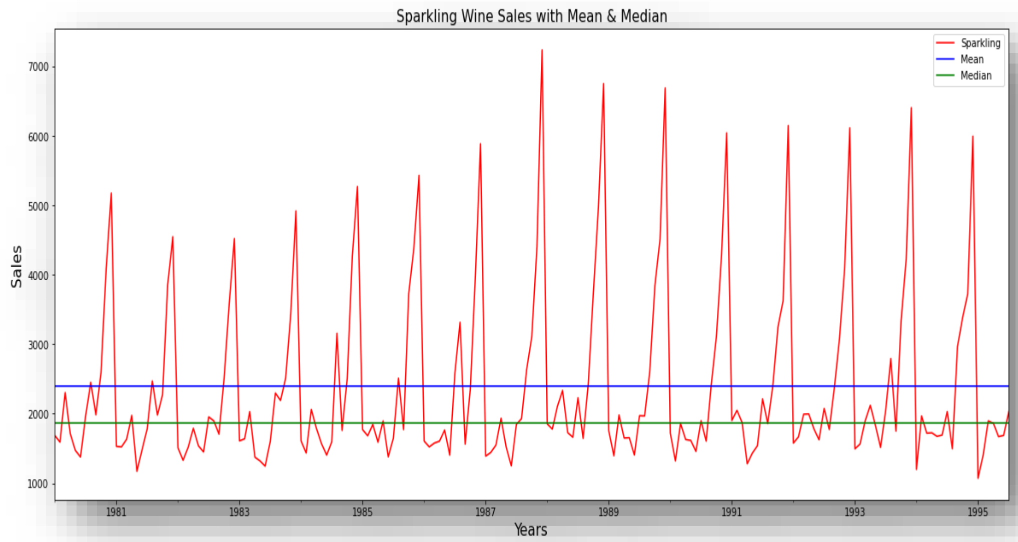
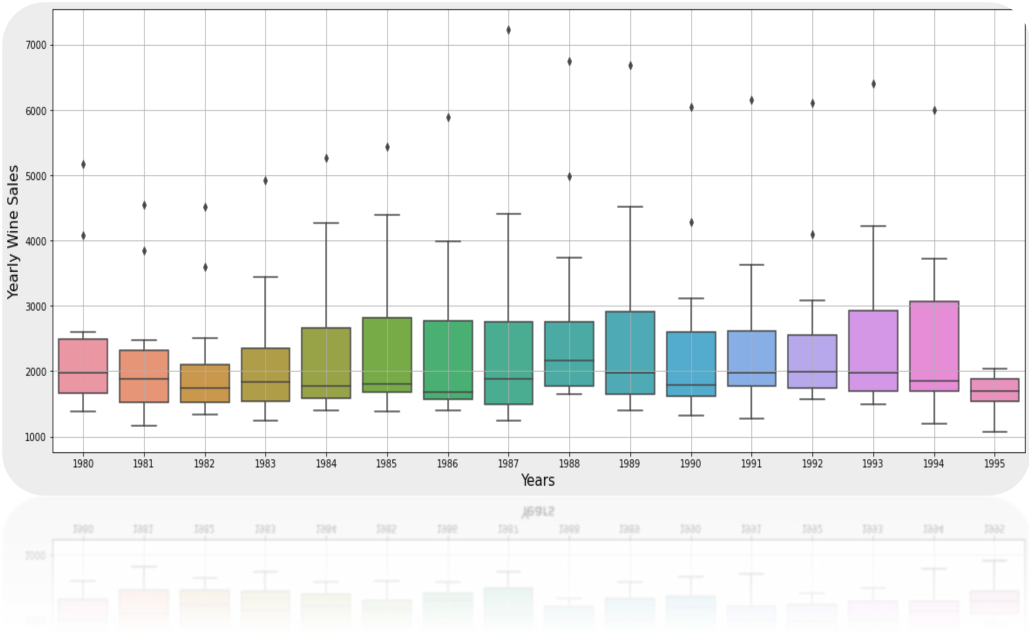


Figure 1. 3 – DATA DISCRIPTION VISUALIZATION

INSIGHTS:

1. Mean sales across 15 years is around 2402 units.
2. Maximum sale achieved by the company in 14 years is around 7242 units

❖ Let's check yearly sales of sparkling wine.



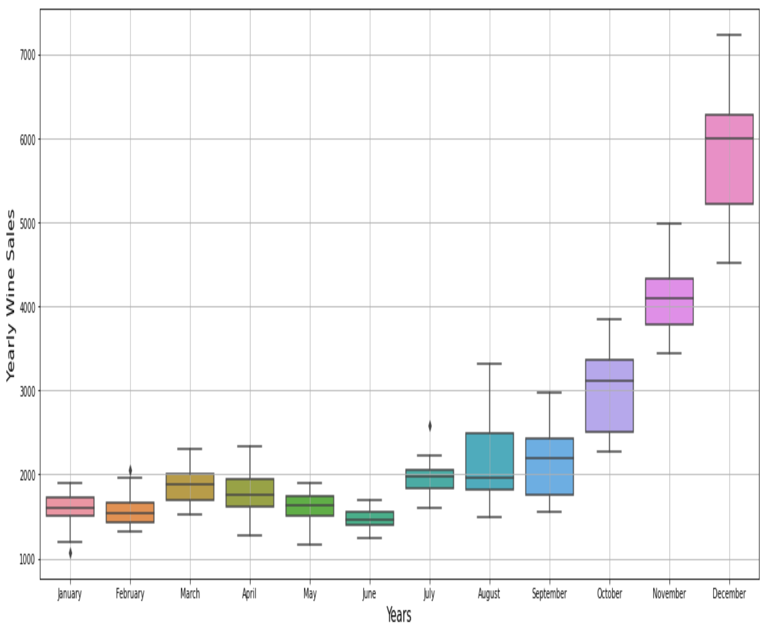
YEARLY TOTAL SALES	
Time Stamp	Yearly Sales
1988	33246
1989	31443
1993	30991
1987	30258
1992	30171
1985	29640
1991	29587
1994	29584
1986	29170
1990	28977
1984	28431
1980	28406
1981	26227
1983	26180
1982	25321
1995	11620

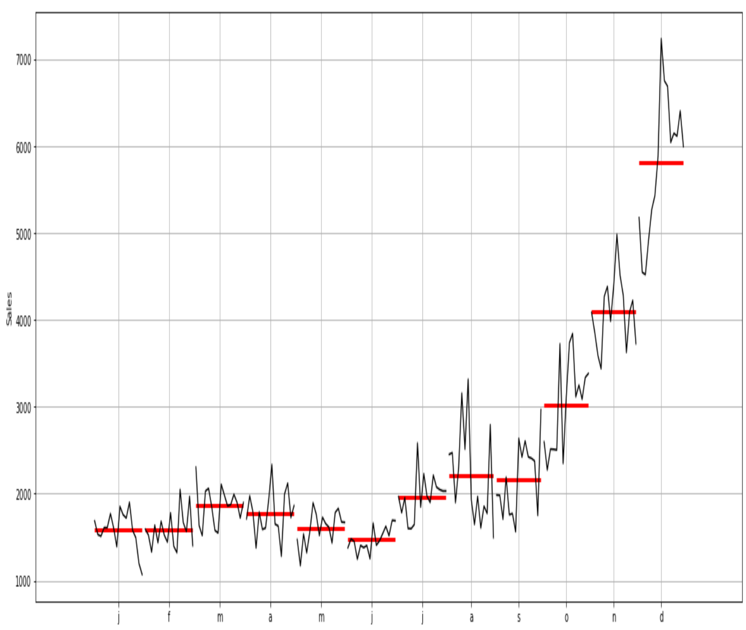
Figure 1. 4 – YEAR ON YEAR SALES OF SPARKLING WINE

INFERENCE

1. 1988 year has maximum sale more than 33000 units.
2. Every year has outlier sale
3. In 1995 only 7 months of sales is recorded, so 1982 has the lowest wine sales in the 14 year history.
4. This plot shows there is no increase in sales over the years which can lead to that, there is no trend present in data.

❖ Let's check Monthly Sales of sparkling wine.





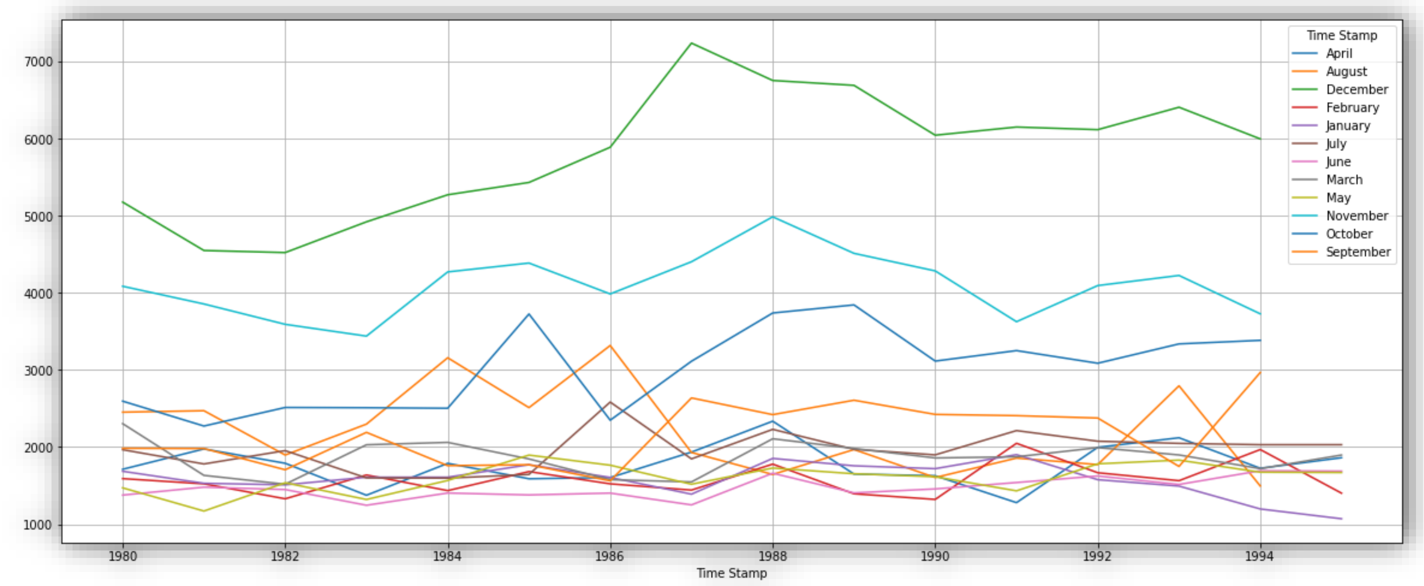


Figure 1. 5- MONTH ON MONTH SALES OF SPARKLING WINE

INFERENCE

From the above plots we can easily infer that year end months (last quarter) has the maximum number of sales recorded across the years and December has the highest unit sales of all the month. This is clear indication of presence of seasonality in the data.

SPARKLING WINE TIME SERIES DECOMPOSITION

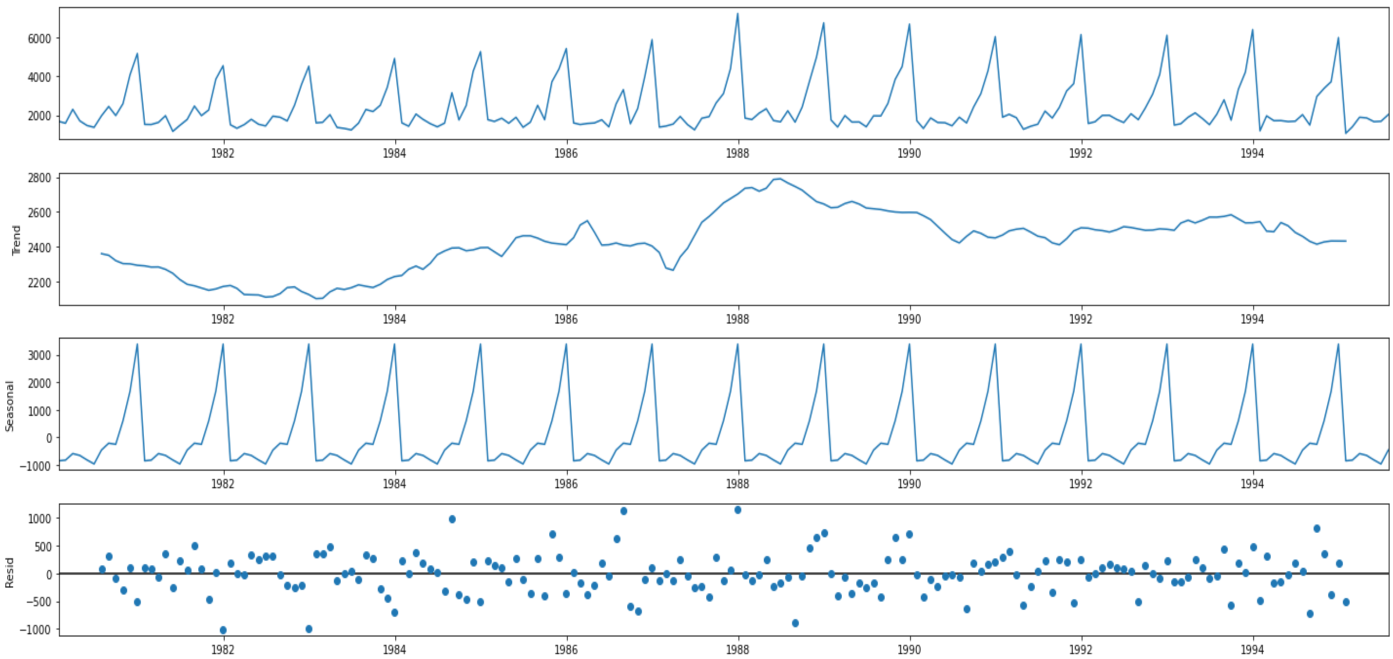


Figure 1. 6 – SPARKLING WINE DATA DECOMPOSITION

INFERENCE:

We can Cleary see from the decomposition plot that there is not much of a trend present in the data, there has been an increase in the sales around 1988 but then it again followed the previous pattern.

There is seasonality present in the data and the time series follows and additive type of time series.

EXPLORATORY DATA ANALYSIS FOR ROSE WINE DATA.

1. As we can see there is only one column Rose which is int64 datatype
2. There are total 187 rows in the data.
3. There are 02 null values present in the data i.e. on July 1984 and Aug 1984

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 19
95-07-31
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Rose    185 non-null    float64
dtypes: float64(1)
memory usage: 2.9 KB
```

Treating Missing Value

Time Stamp	Rose
31-07-1994	NaN
31-08-1994	NaN

Using Interpolate with method 'spline' and order '3' to fill the missing values in the time series data

Time Stamp	Rose
31-07-1994	43.693064
31-08-1994	44.326877

DATA DESCRIPTION	
	Rose
count	187
mean	89.898502
std	39.256767
min	28
25%	62.5
50%	85
75%	111
max	267

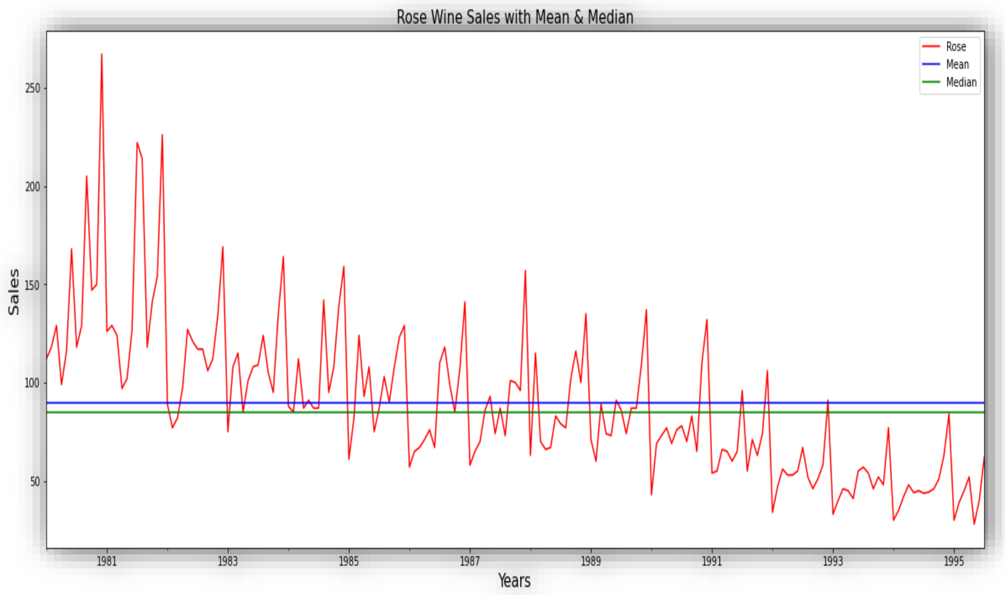


Figure 1. 7 – ROSE WINE DATA DESCRIPTION AND VISUALIZATION

INSIGHTS:

1. Mean sales across 15 years is around 90.
2. Maximum sale done by the company in 14 years is around 267.

❖ **Let's check yearly sales of Rose wine.**

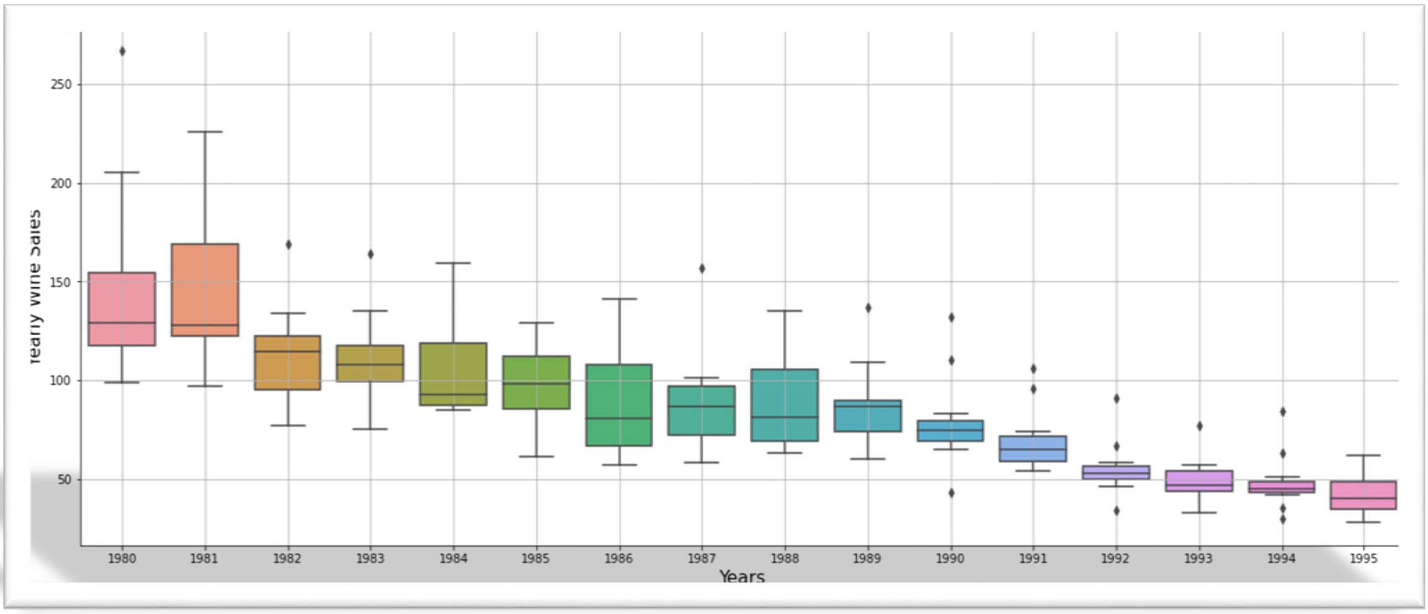
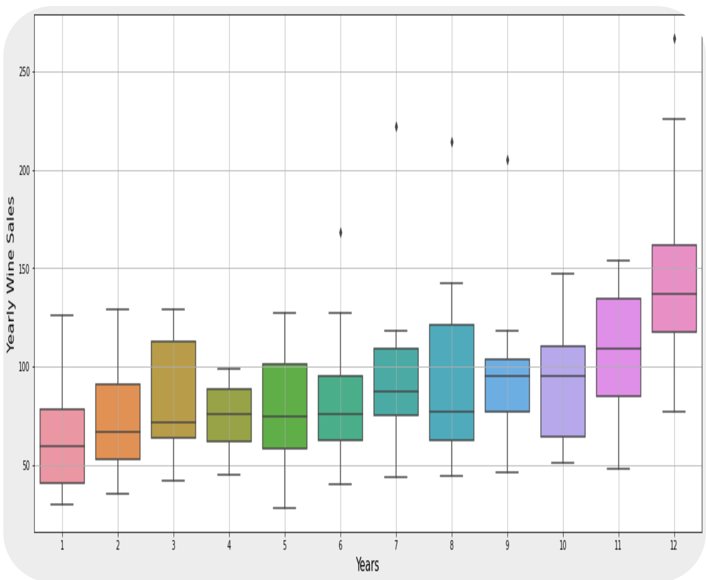


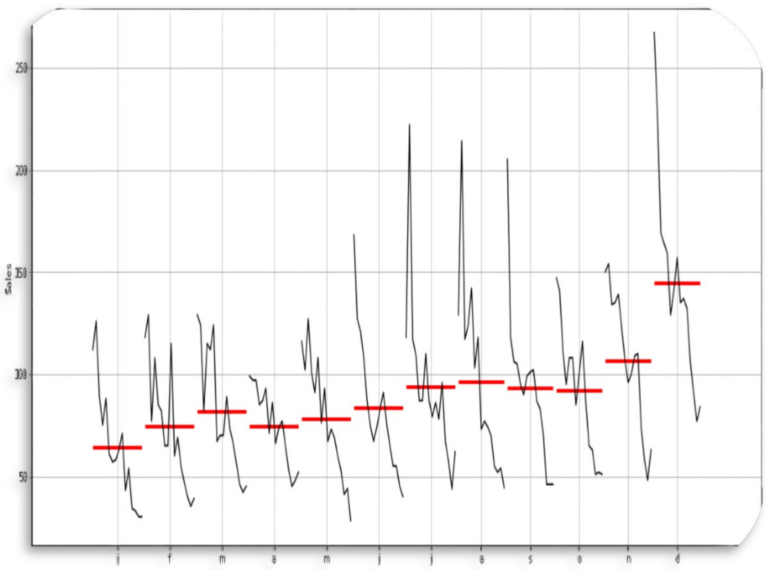
Figure 1. 8 – YEAR ON YEAR SALES OF ROSE WINE

INFERENCE

1. 1981 year has maximum sale more than units.
2. Almost every year has outlier sale
3. We can clearly interpret from the plot that there is significant decrease in the sales over the years causing a downward trend.

❖ **Let's check Monthly sales of Rose wine.**





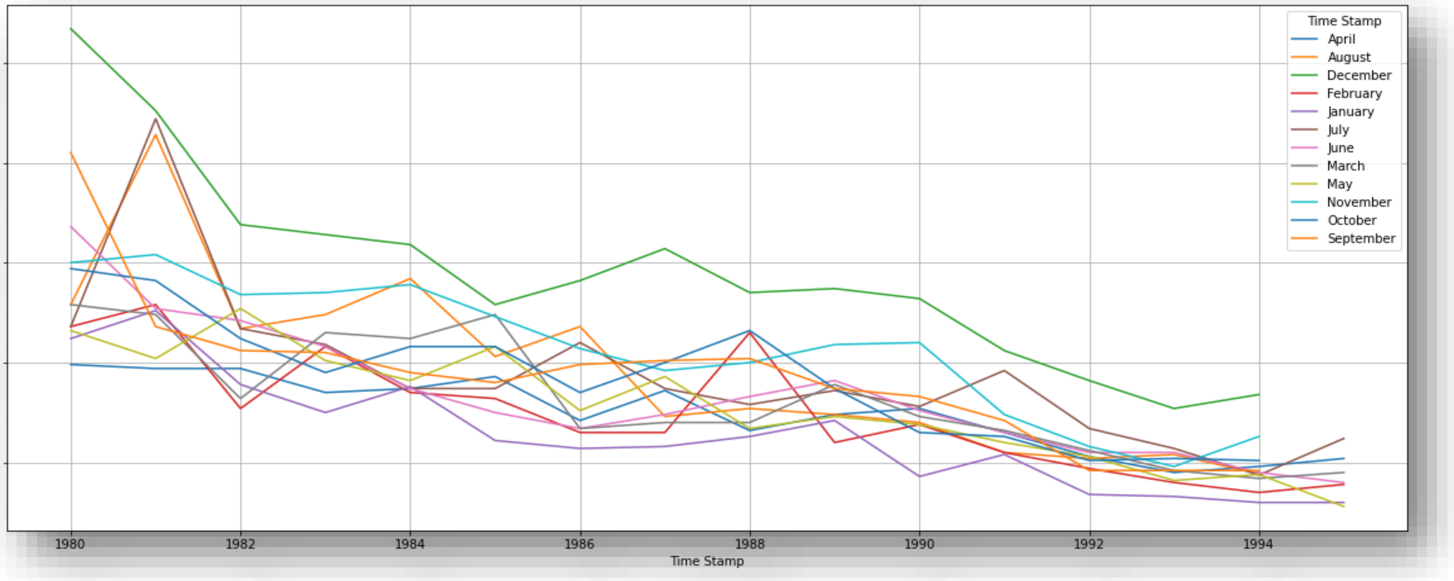


Figure 1. 9 – MONTH ON MONTH SALES OF ROSE WINE

INFERENCE

From the above plots we can easily infer that year end months (last 02 months) has the maximum number of sales recorded across the years and December has the highest unit sales of all the month followed by November than August. This is clear indication of presence of seasonality in the data.

<u>ROSE WINE TIME SERIES DECOMPOSITION</u>

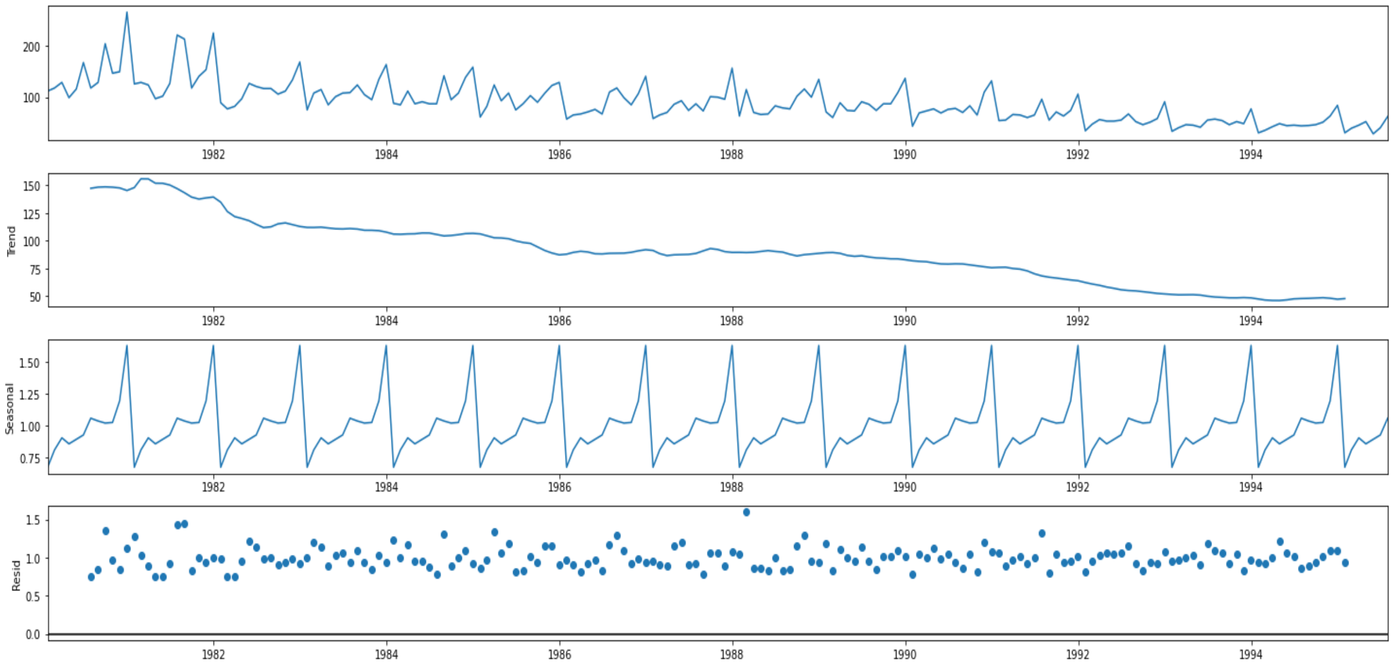


Figure 1. 10 – ROSE WINE DATA DECOMPOSITION

INFERENCE:

We can Cleary see from the decomposition plot that there is a downward trend present in the data, this plot shows significant decrease in the sales over the years. There is also monthly changes in the sales means seasonality is present in the data and the time series follows and multiplicative type of time series.

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

SOLUTION:

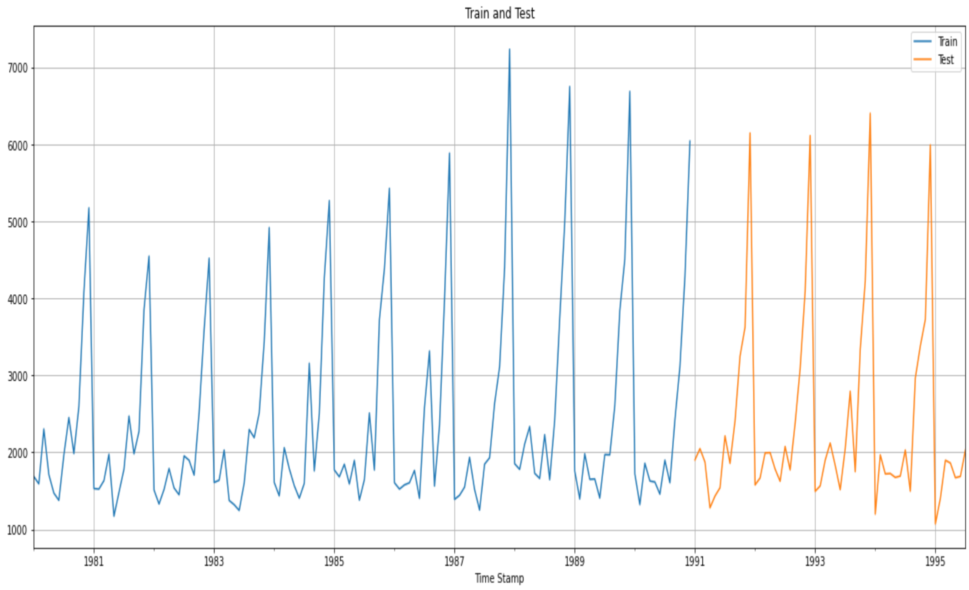
TRAIN AND TEST SPLIT FOR SPARKLING WINE

TRAIN TOP 5 ROWS

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

TRAIN LAST 5 ROWS

Time Stamp	Sparkling
31-08-1990	1605
30-09-1990	2424
31-10-1990	3116
30-11-1990	4286
31-12-1990	6047



TEST TOP 5 ROWS

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

TEST LAST 5 ROWS

Time Stamp	Sparkling
31-03-1995	1897
30-04-1995	1862
31-05-1995	1670
30-06-1995	1688
31-07-1995	2031

Figure 1. 11 – TRAIN AND TEST VISUALIZATION OF SPARKLING WINE

As we can for Sparkling wine Train data contains data from 1980 to 1990 and Test Data contains data from 1991 to 1995.

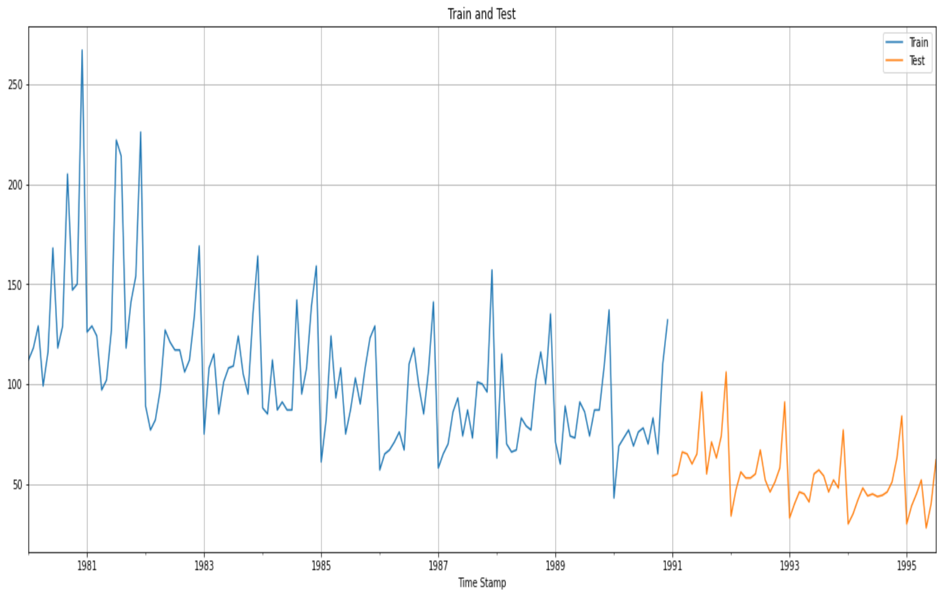
TRAIN AND TEST SPLIT FOR ROSE WINE

TRAIN TOP 5 ROWS

Time Stamp	Rose
31-01-1980	112
29-02-1980	118
31-03-1980	129
30-04-1980	99
31-05-1980	116

TRAIN LAST 5 ROWS

Time Stamp	Rose
31-08-1990	70
30-09-1990	83
31-10-1990	65
30-11-1990	110
31-12-1990	132



TEST TOP 5 ROWS

Time Stamp	Rose
31-01-1991	54
28-02-1991	55
31-03-1991	66
30-04-1991	65
31-05-1991	60

TEST LAST 5 ROWS

Time Stamp	Rose
31-03-1995	45
30-04-1995	52
31-05-1995	28
30-06-1995	40
31-07-1995	62

Figure 1. 12 - TRAIN AND TEST VISUALIZATION OF ROSE WINE

As we can for Rose wine also Train data contains data from 1980 to 1990 and Test Data contains data from 1991 to 1995.

Q-4. Build all the exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other additional models such as regression, naïve forecast models, simple average models, moving average models should also be built on the training data and check the performance on the test data using RMSE.

SOLUTION:

SPARKLING WINE

- **REGRESSION MODEL**

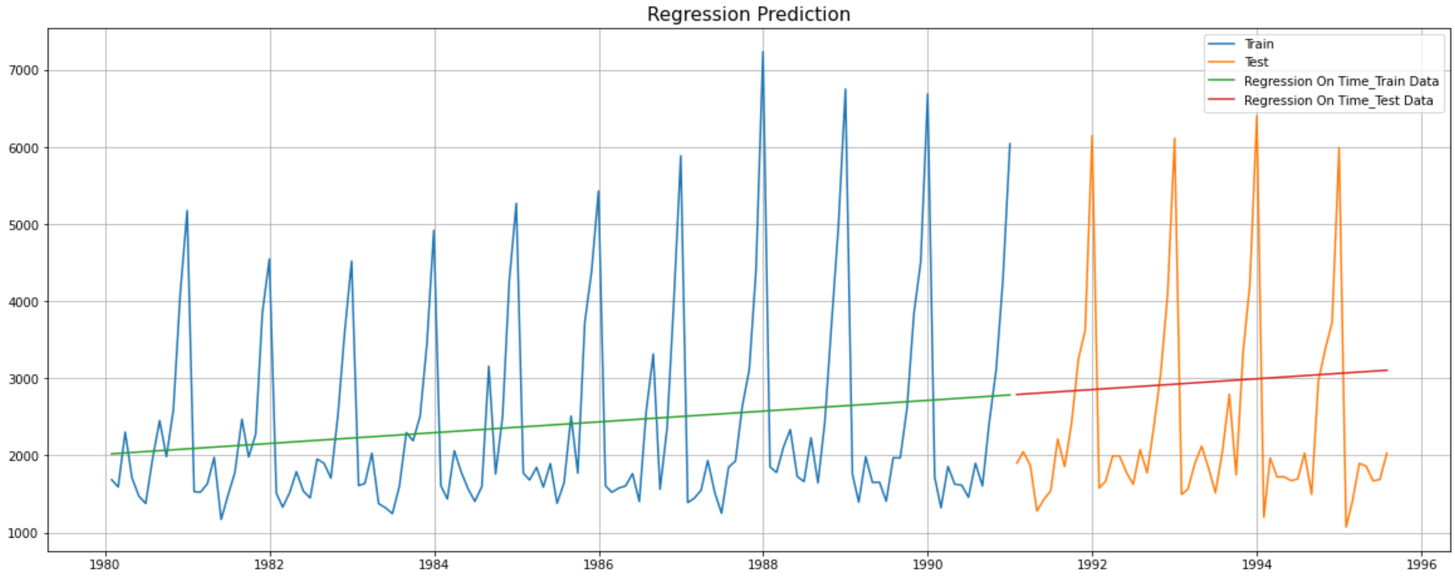


Figure 1. 13 – REGRESSION MODEL PREDICTION ON SPARKLING WINE

REGRESSION MODEL RESULT		
	TEST RMSE	TEST MAPE
Regression On Time	1389.135175	50.15

TABLE 1. 3 – REGRESSION MODEL PERFORMANCE

As the result shows MAPE value for the model is very high and the model only captured the trend part of the data not the seasonality.

- **NAÏVE FORECAST MODEL**

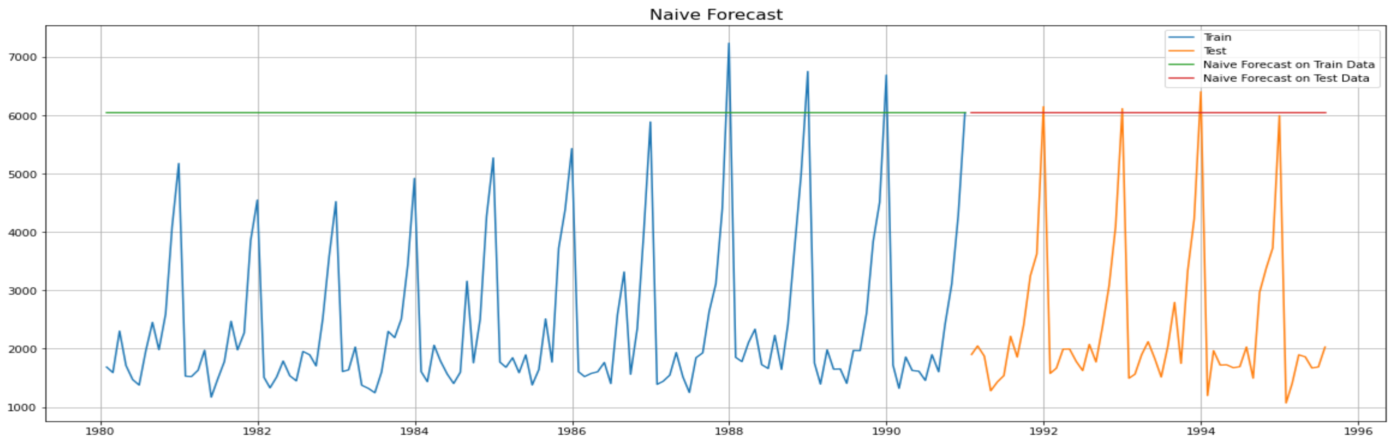


Figure 1. 14 – NAÏVE FORECAST MODEL PREDICTION ON SPARKLING WINE

NAÏVE FORECAST MODEL RESULT		
	TEST RMSE	TEST MAPE
Naïve Forecast	3864.279352	152.87

TABLE 1. 4 – NAÏVE FORECAST MODEL PERFORMANCE FOR SPARKLING WINE

As the result shows MAPE value for the model is very high. Model took the last value in train data and predicted the same value for the entire test data.

➢ SIMPLE AVERAGE

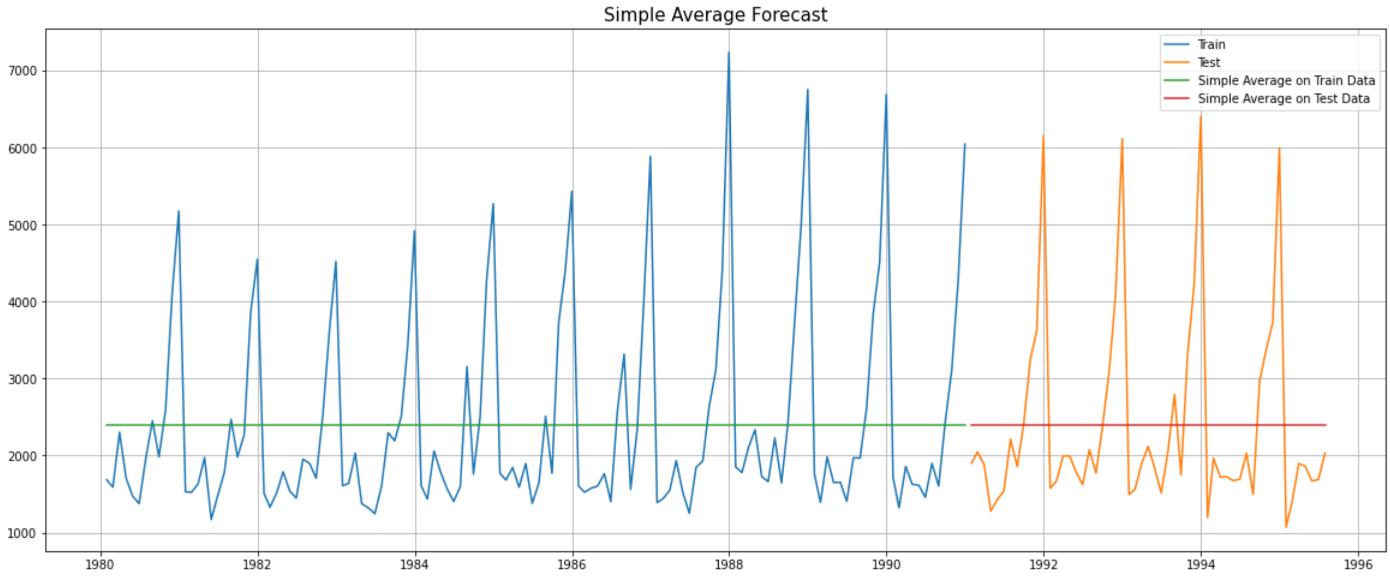


Figure 1. 15 – SIMPLE AVERAGE MODEL PREDICTION ON SPARKLING WINE

SIMPLE AVERAGE MODEL RESULT		
	TEST RMSE	TEST MAPE
Simple Average	1275.081804	38.9

TABLE 1. 5 – SIMPLE AVERAGE MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model is very high. Model took the average value of train data and predicted the same value for the entire test data.

➢ MOVING AVERAGE

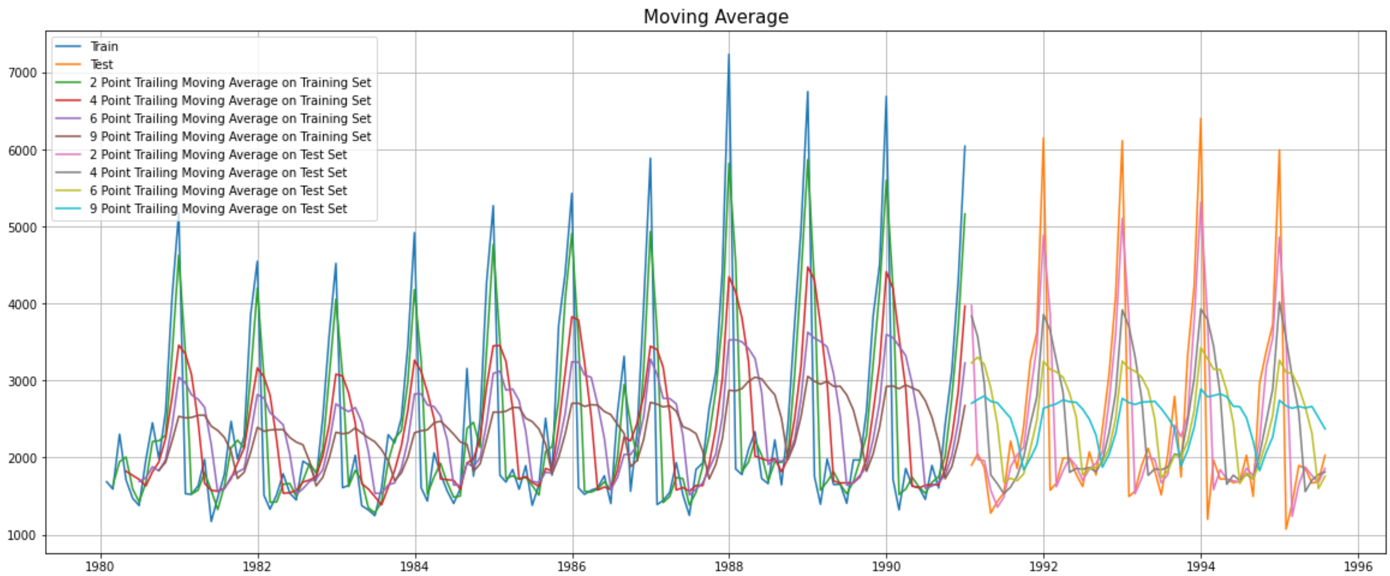


Figure 1. 16 – MOVING AVERAGE MODEL PREDICTION ON SPARKLING WINE

MOVING AVERAGE MODEL RESULT		
	TEST RMSE	TEST MAPE
2pointTrailingMovingAverage	813.400684	19.7
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1156.589694	43.86
9pointTrailingMovingAverage	1346.278315	46.86

TABLE 1. 6 – MOVING AVERAGE MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model is better for some models. Model took the Rolling average value of train data for different window (2, 4, 6, 9) and predicted the same value for the entire test data. This model is able to capture seasonality to some

➢ SIMPLE EXPONENTIAL SMOOTHING

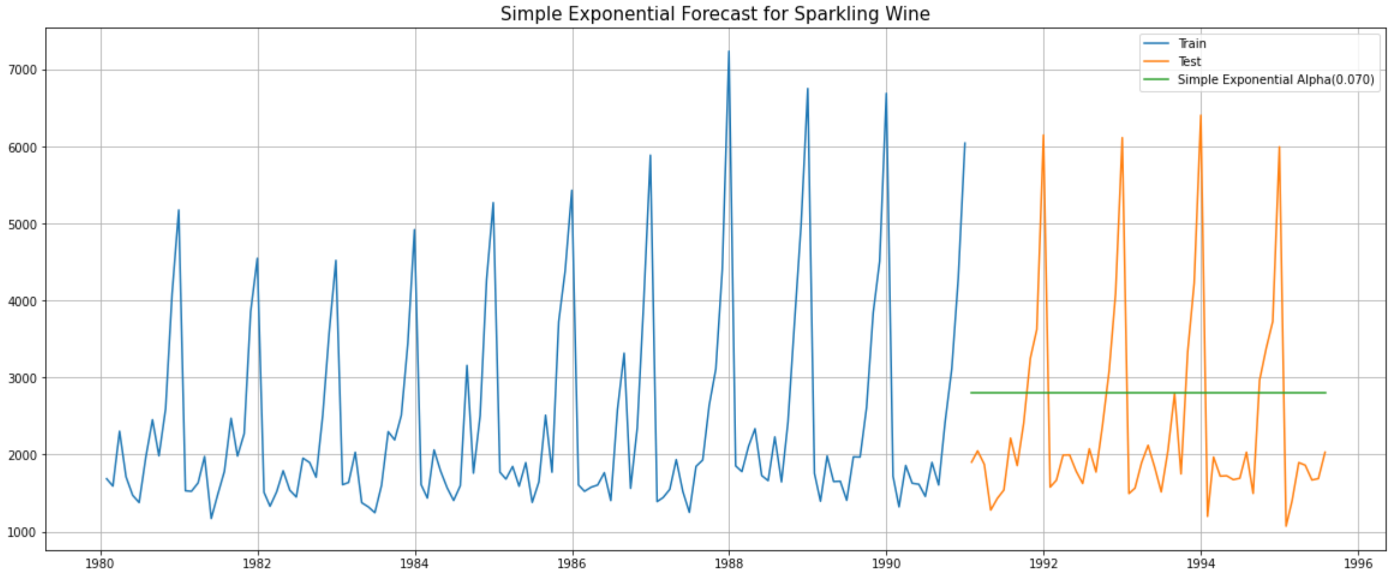


Figure 1. 17 – SIMPLE EXPONENTIAL SMOOTHENING MODEL PREDICTION ON SPARKLING WINE

SINGLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Simple Exponential Alpha(0.070)	1338	47.11

TABLE 1. 7 – SIMPLE EXPONENTIAL SMOOTHENING PERFORMANCE FOR SPARKLING WINE

As the result shows MAPE value for the model is high. It took only one parameter into the account i.e. LEVEL and forecasted the test data on that with value of alpha (0.07)

➢ SIMPLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA

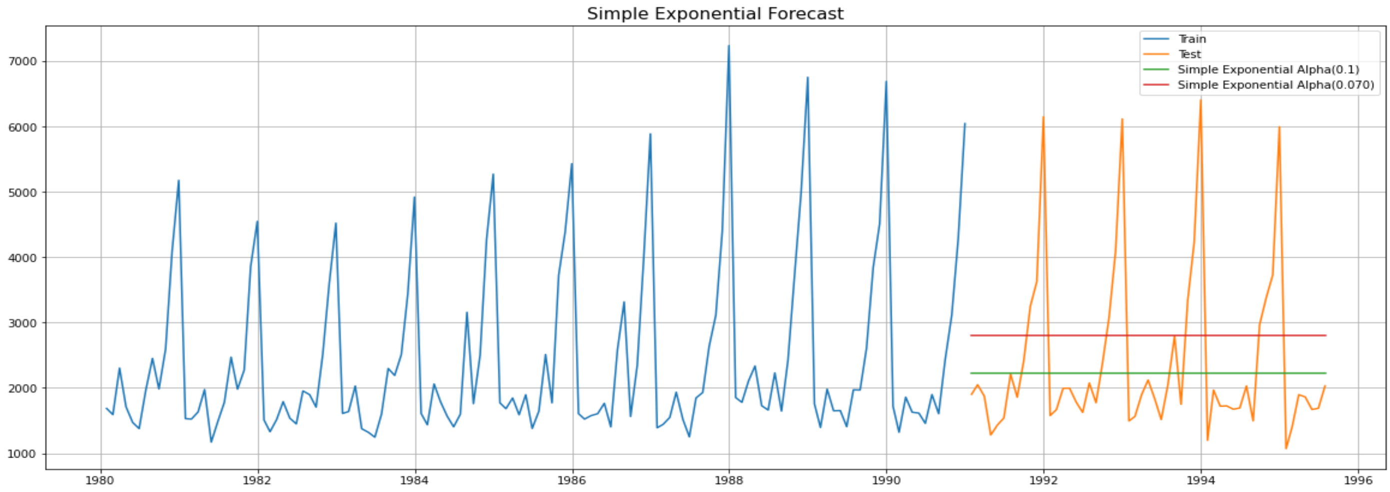


Figure 1. 18 – SIMPLE EXPONENTIAL SMOOTHENING ALPHA (0.01) MODEL PREDICTION ON SPARKLING WINE

SINGLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Single Exponential Smoothing Alpha(0.01)	1286.648	35.78

TABLE 1. 8 - SIMPLE EXPONENTIAL SMOOTHENING ALPHA (0.01) MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model is high but improved. It took only one parameter into the account i.e. LEVEL and forecasted the test data on that with value of alpha (0.01)

➢ DOUBLE EXPONENTIAL SMOOTHING

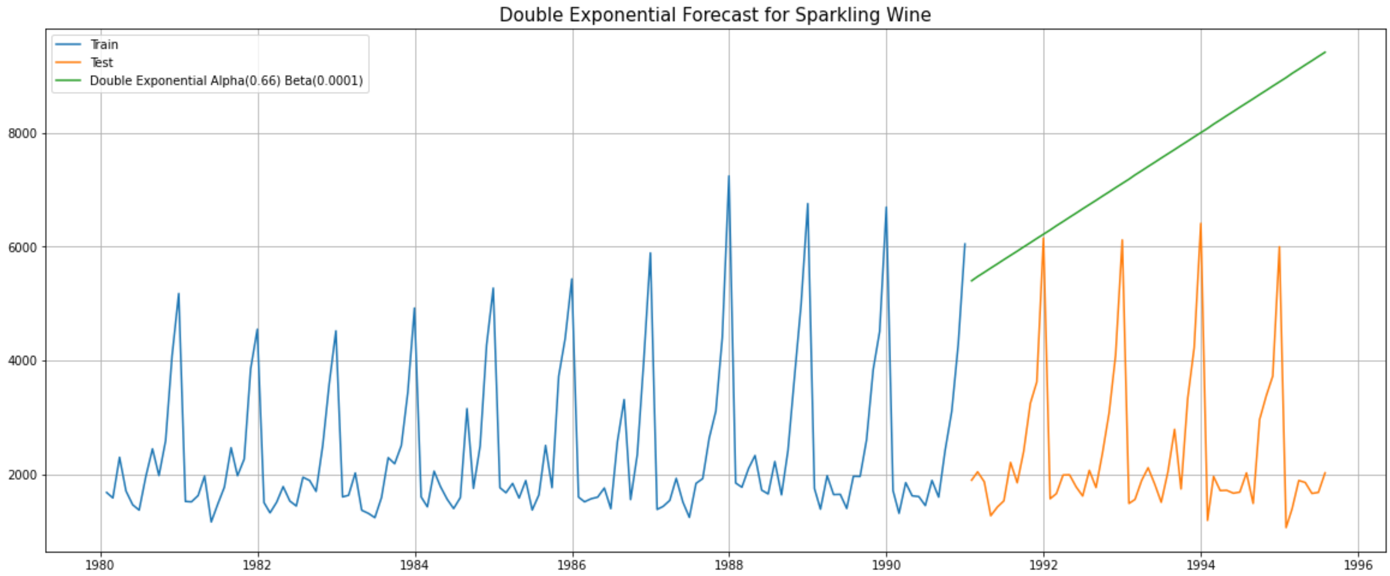


Figure 1. 19 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.66) BETA (0.001) MODEL PREDICTION ON SPARKLING WINE

DOUBLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Double Exponential Alpha(0.66) Beta(0.0001)	5291.88	208.74

TABLE 1. 9 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.66) BETA (0.001) MODEL PREDICTION ON SPARKLING WINE

As the result shows MAPE value for the model is Very high. It took two parameter into the account i.e. LEVEL and TREND forecasted the test data on that with value of alpha (0.01), Beta (0.0001) causing it to predict increasing trend.

➢ DOUBLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA AND BETA

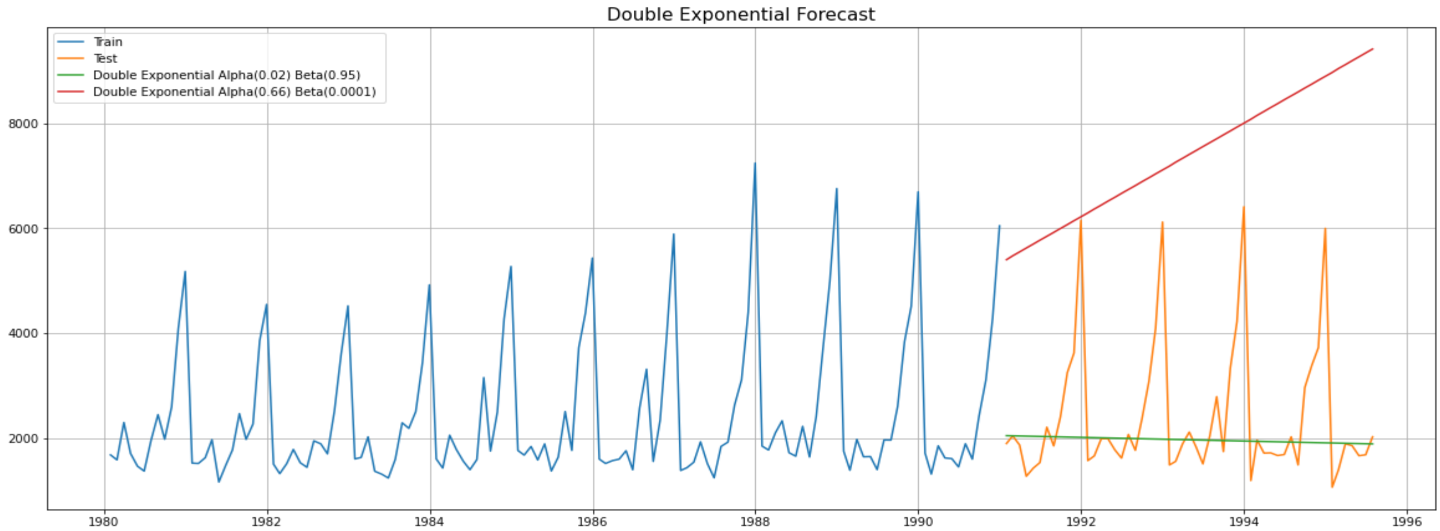


Figure 1. 20 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.02) BETA (0.95) MODEL PREDICTION ON SPARKLING WINE

DOUBLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Double Exponential Smoothing Alpha(0.02) Beta(0.95)	1498.73	55.31

TABLE 1. 10 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.02) BETA (0.95) MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model is high but improved before. It took two parameters into the account i.e. LEVEL and TREND forecasted the test data on that with value of alpha (0.02), Beta (0.95) causing it to predict increasing

- ## <u>TRIPLE EXPONENTIAL SMOOTHING</u>

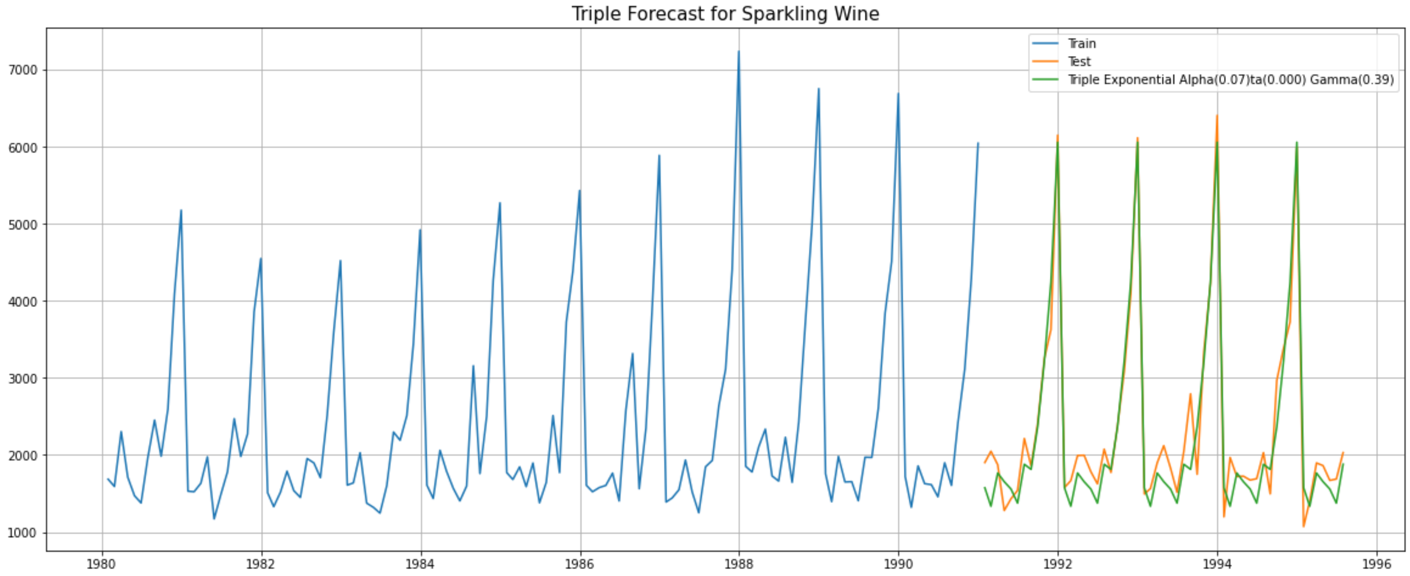


Figure 1. 21 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.07) BETA (0.00) GAMMA (0.39) MODEL PREDICTION ON SPARKLING WINE

TRIPLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Triple Exponential Alpha(0.07)ta(0.000) Gamma(0.39)	318.83	10.14

TABLE 1. 11 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.07) BETA (0.00) GAMMA (0.39) MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model has significantly improved. It took three parameters into the account i.e., LEVEL, TREND, SEASONALITY, then forecasted the test data on that with value of alpha (0.07), Beta (0.000) and Gamma (0.39).

- ## <u>TRIPLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA AND BETA AND GAMMA</u>

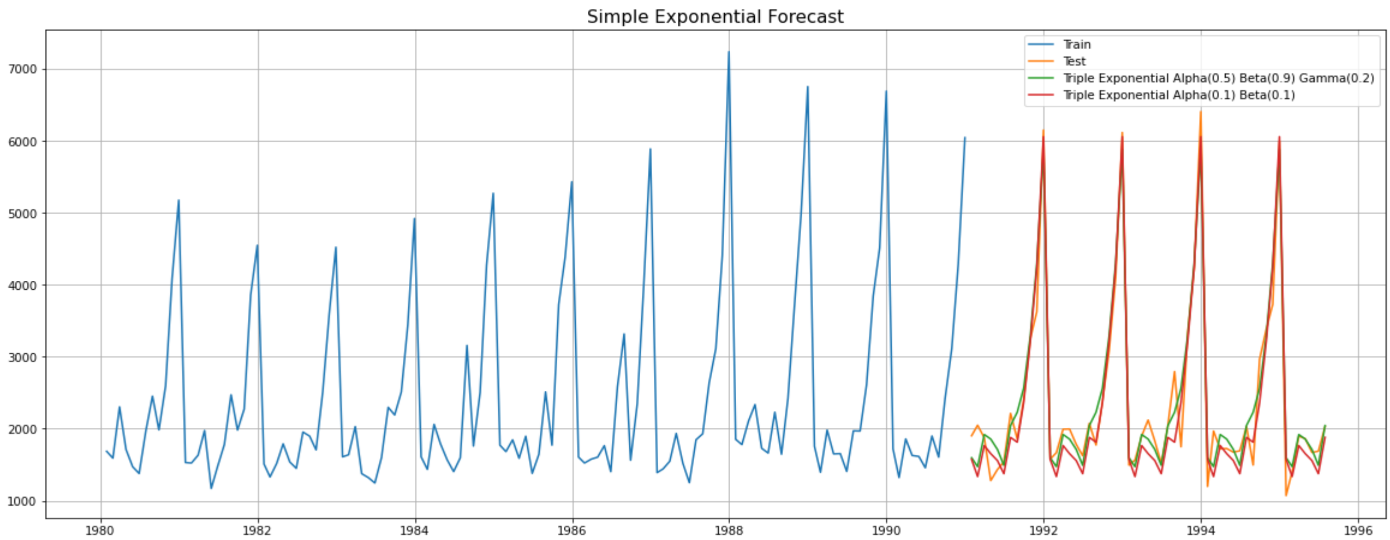


Figure 1. 22 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.4) BETA (0.9) GAMMA (0.2) MODEL PREDICTION ON SPARKLING WINE

TRIPLE EXPONENTIAL SMOOTHING MODEL RESULT		
	TEST RMSE	TEST MAPE
Triple Exponential Smoothing Alpha(0.5) Beta(0.9) Gamma(0.2)	309.86	9.39

TABLE 1. 12 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.4) BETA (0.9) GAMMA (0.2) MODEL PERFORMANCE ON SPARKLING WINE

As the result shows MAPE value for the model has drastically improved from the above models. It took three parameters into the account i.e., LEVEL, TREND, SEASONALITY, then forecasted the test data on that with value of alpha (0.5), Beta (0.9) and Gamma (0.2).

ROSE WINE

➢ REGRESSION MODEL

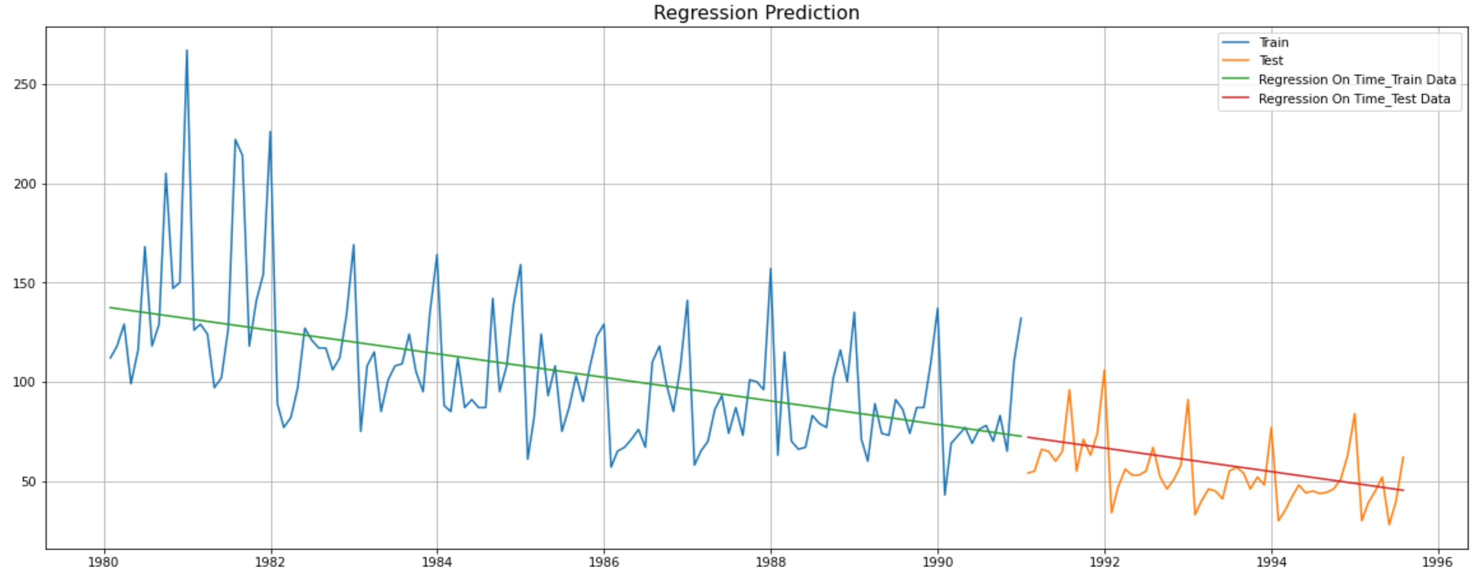


Figure 1. 23– REGRESSION MODEL PREDICTION ON ROSE WINE

REGRESSION MODEL RESULT		
	TEST RMSE	TEST MAPE
Regression On Time	15.29146	22.94

TABLE 1. 13 – REGRESSION MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model is very high and the model only captured the trend part of the data not the seasonality.

➢ NAÏVE FORECAST MODEL

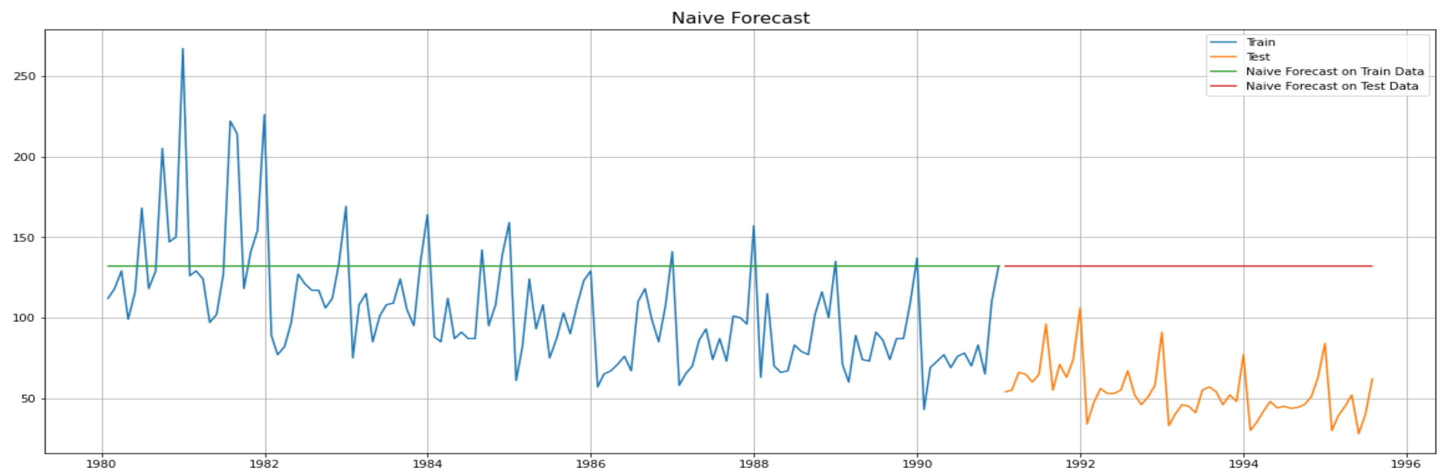


Figure 1. 24 – NAÏVE FORECAST MODEL PREDICTION ON ROSE WINE

NAIVE MODEL RESULT		
	TEST RMSE	TEST MAPE
NAÏVE MODEL	79.778066	145.35

TABLE 1. 14 – NAÏVE FORECAST MODEL PERFORMANCE FOR ROSE WINE

As the result shows MAPE value for the model is very high. Model took the last value in train data and predicted the same value for the entire test data.

- **SIMPLE AVERAGE**

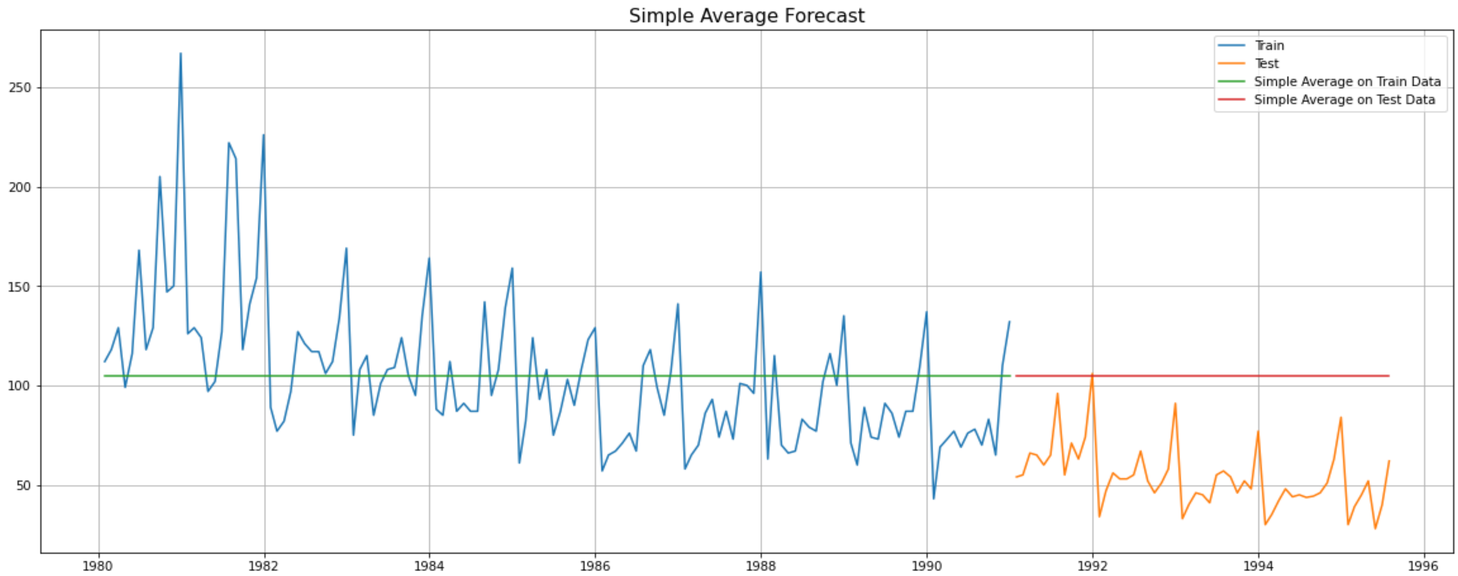


Figure 1. 25– SIMPLE AVERAGE MODEL PREDICTION ON ROSE WINE

SIMPLE AVERAGE MODEL RESULT		
	TEST RMSE	TEST MAPE
SIMPLE AVERAGE	53.521	95.13

TABLE 1. 15 – SIMPLE AVERAGE MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model is very high. Model took the average value of train data and predicted the same value for the entire test data.

- **MOVING AVERAGE**

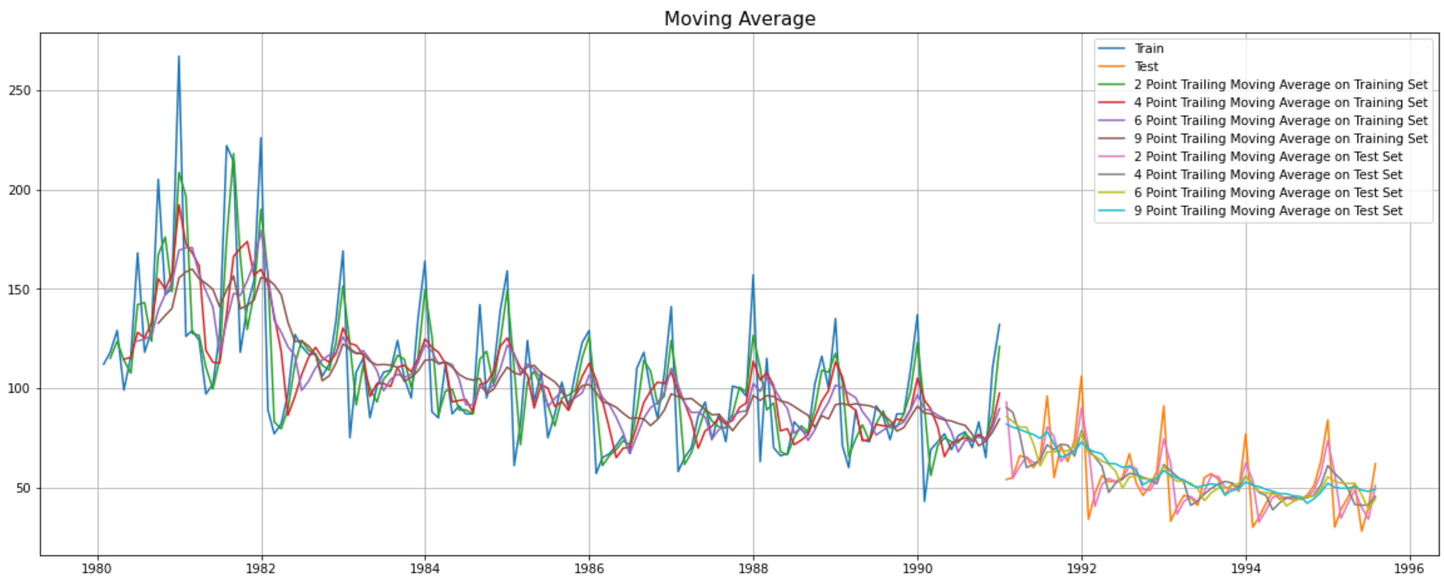


Figure 1. 26 – MOVING AVERAGE MODEL PREDICTION ON ROSE WINE

MOVING AVERAGE MODEL RESULT		
	TEST RMSE	TEST MAPE
2pointTrailingMovingAverage	11.53018	13.6
4pointTrailingMovingAverage	14.46233	19.59
6pointTrailingMovingAverage	14.46233	20.83
9pointTrailingMovingAverage	14.740112	21.13

As the result shows MAPE value for the model is better for some models. Model took the Rolling average value of train data for different window (2, 4, 6, 9) and predicted the same value for the entire test data. This model is able to capture seasonality to some

TABLE 1. 16 – MOVING AVERAGE MODEL PERFORMANCE ON ROSE WINE

➢ SIMPLE EXPONENTIAL SMOOTHING

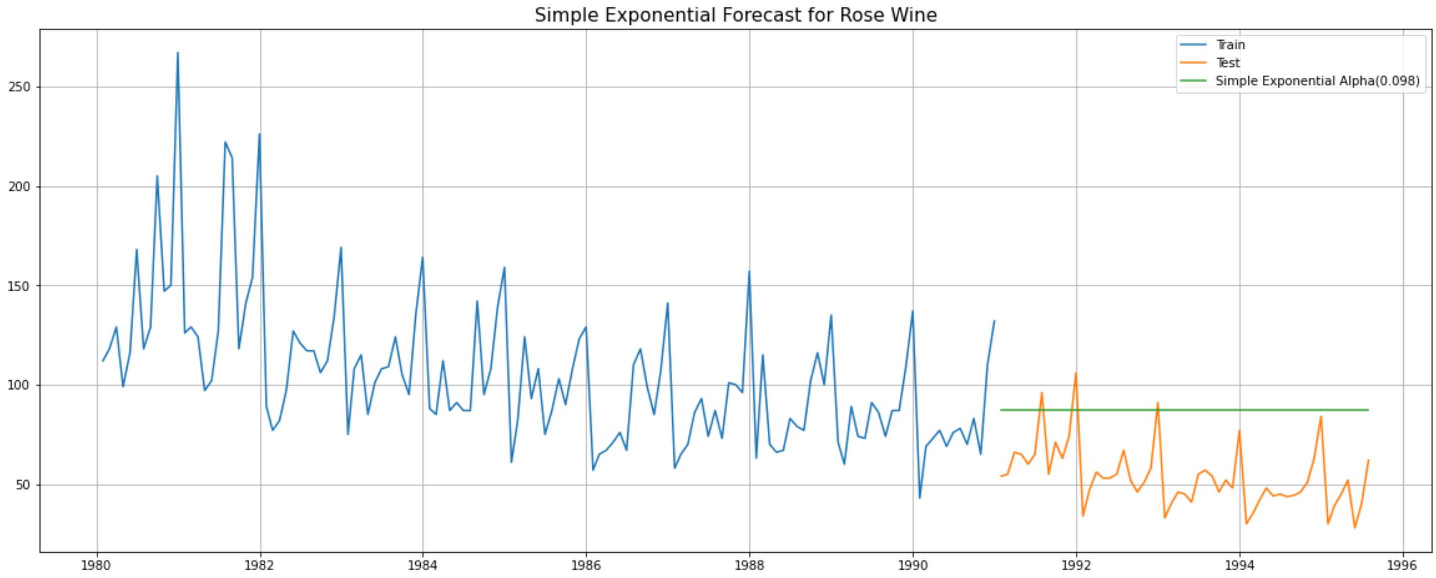


Figure 1. 27 – SIMPLE EXPONENTIAL SMOOTHENING MODEL PREDICTION ON ROSE WINE

SINGLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Simple Exponential Alpha(0.098)	36.858569	64.05

TABLE 1. 17 – SIMPLE EXPONENTIAL SMOOTHENING PERFORMANCE FOR ROSE WINE

As the result shows MAPE value for the model is high. It took only one parameter into the account i.e., LEVEL and forecasted the test data on that with value of alpha (0.098)

➢ SIMPLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA

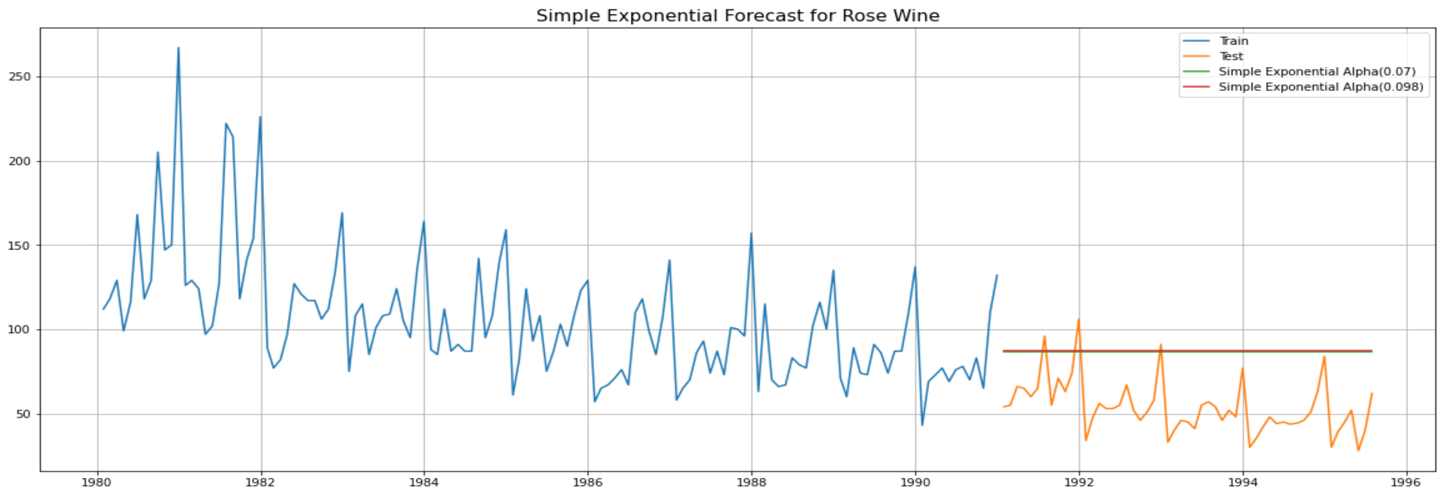


Figure 1. 28 – SIMPLE EXPONENTIAL SMOOTHENING ALPHA (0.01) MODEL PREDICTION ON SPARKLING WINE

SINGLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Single Exponential Alpha(0.07)	36.49737	63.38

TABLE 1. 18 - SIMPLE EXPONENTIAL SMOOTHENING ALPHA (0.07) MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model is high but improved. It took only one parameter into the account i.e., LEVEL and forecasted the test data on that with value of alpha (0.07)

➢ DOUBLE EXPONENTIAL SMOOTHING

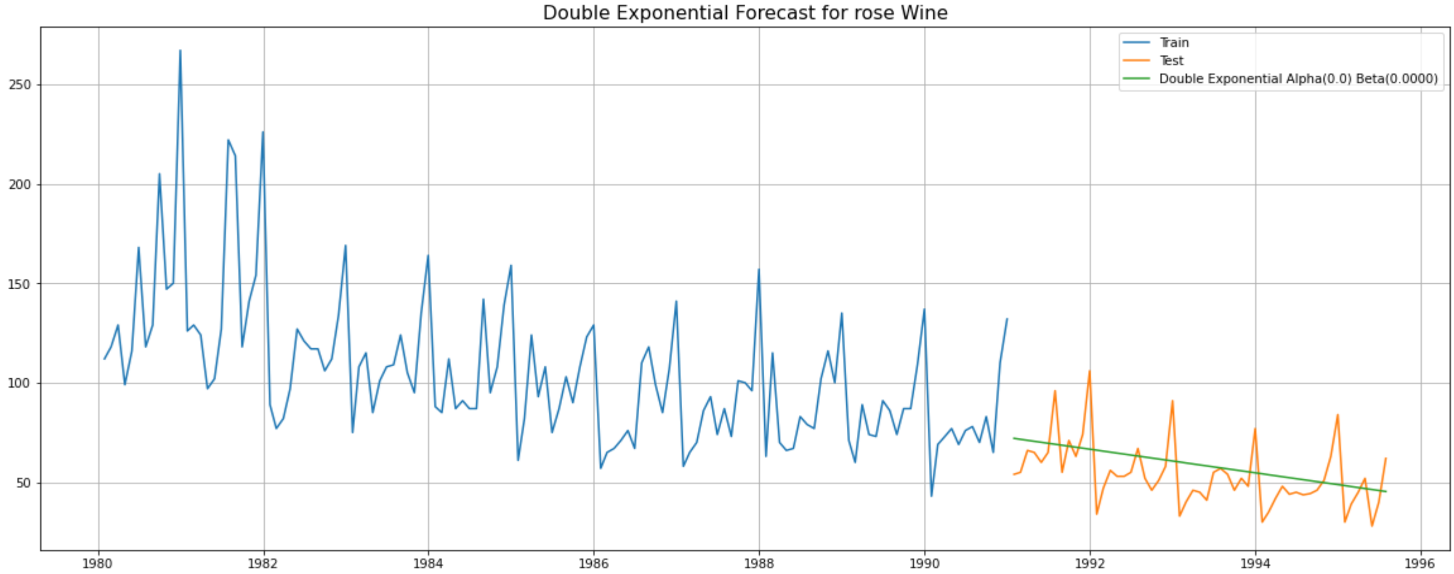


Figure 1. 29 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.00) BETA (0.000) MODEL PREDICTION ON ROSE WINE

As the result shows MAPE value for the model is Very high. It took two parameters into the account i.e., LEVEL and TREND forecasted the test data on that with value of alpha (0.00), Beta (0.0000) causing it to predict increasing trend.

DOUBLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Double Exponential Alpha(0.0) Beta(0.0000)	15.293494	22.95

TABLE 1. 19 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.00) BETA (0.0000) MODEL PREDICTION ON ROSE WINE

➢ DOUBLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA AND BETA

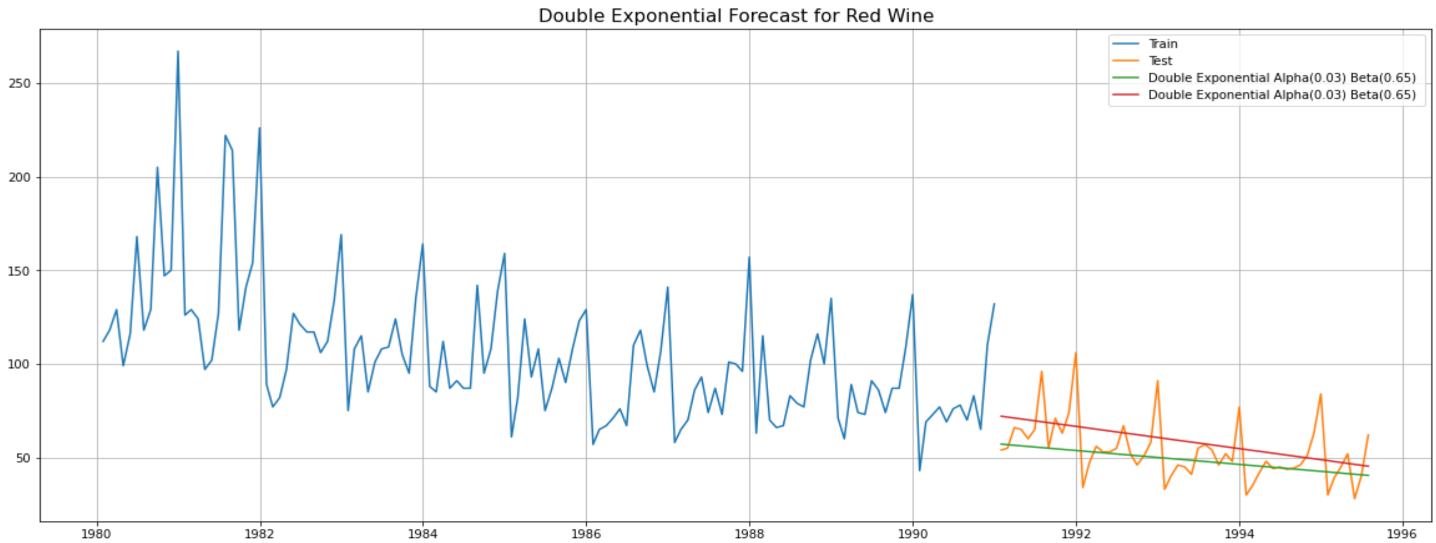


Figure 1. 30 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.03) BETA (0.65) MODEL PREDICTION ON ROSE WINE

DOUBLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Double Exponential Alpha(0.03) Beta(0.65)	15.331019	18.61

TABLE 1. 20 - DOUBLE EXPONENTIAL SMOOTHENING ALPHA (0.03) BETA (0.65) MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model is high but improved before. It took two parameters into the account i.e., LEVEL and TREND forecasted the test data on that with value of alpha (0.03), Beta (0.65) causing it to predict increasing

➢ TRIPLE EXPONENTIAL SMOOTHING

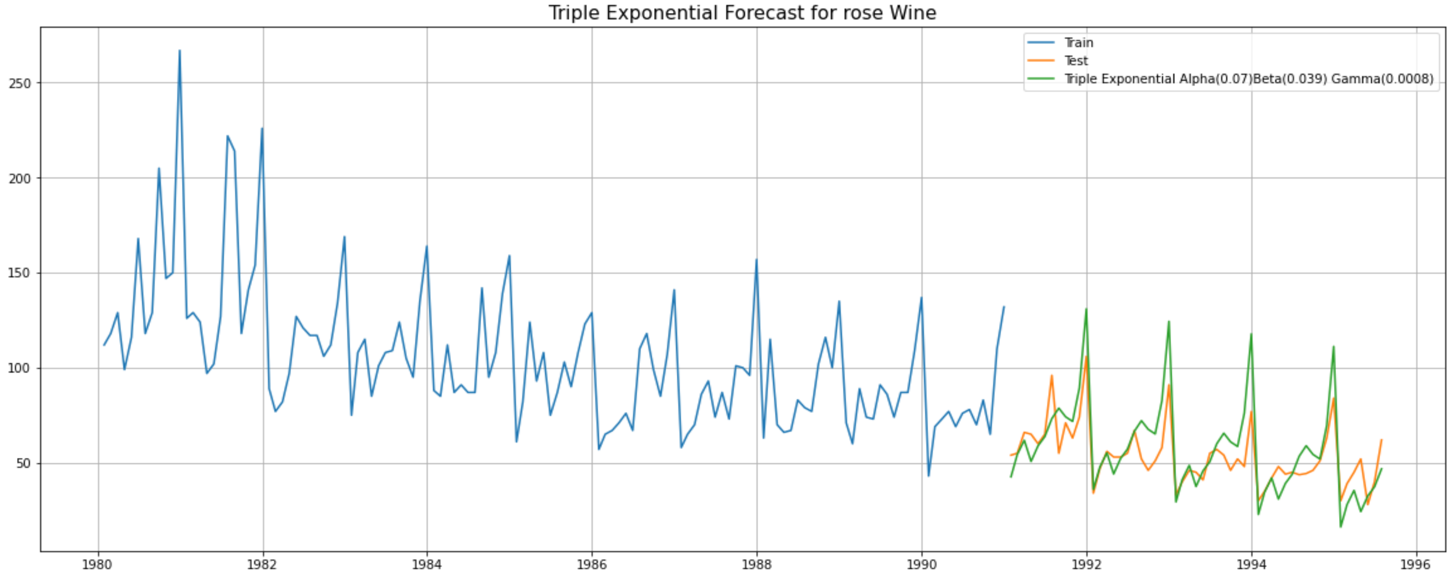


Figure 1. 31 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.08) BETA (5.52) GAMMA (0.0005) MODEL PREDICTION ON ROSE WINE

TRIPLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Triple Exponential Alpha(0.08) Beta(5.52) Gamma(0.0005)	14.28	19.37

TABLE 1. 21 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.08) BETA (5.52) GAMMA (0.0005) MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model has significantly improved. It took three parameters into the account i.e., LEVEL, TREND, SEASONALITY, then forecasted the test data on that with value of alpha (0.08), Beta (5.52) and Gamma (0.0005).

➢ TRIPLE EXPONENTIAL SMOOTHING WITH DIFFERENT PARAMETERS OF ALPHA AND BETA AND GAMMA

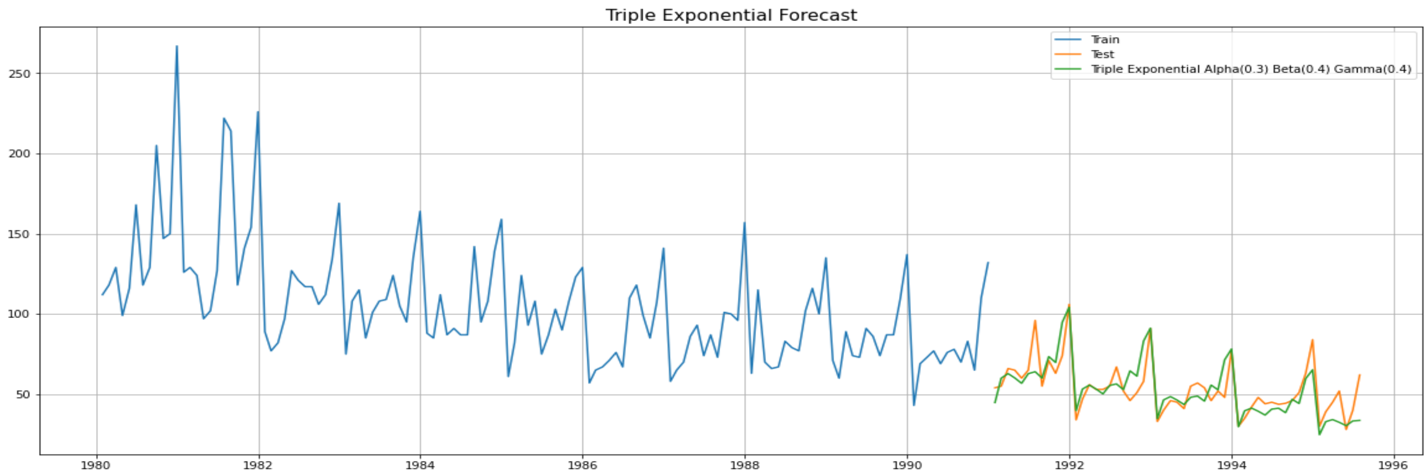


Figure 1. 32 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.3) BETA (0.4) GAMMA (0.4) MODEL PREDICTION ON ROSE WINE

TRIPLE EXPONENTIAL SMOOTHNING MODEL RESULT		
	TEST RMSE	TEST MAPE
Triple Exponential Alpha(0.3) Beta(0.4) Gamma(0.4)	10.34	13.3

TABLE 1. 22 - TRIPLE EXPONENTIAL SMOOTHENING ALPHA (0.3) BETA (0.4) GAMMA (0.4) MODEL PERFORMANCE ON ROSE WINE

As the result shows MAPE value for the model has drastically improved from the above models. It took three parameters into the account i.e., LEVEL, TREND, SEASONALITY, then forecasted the test data on that with value of alpha (0.3), Beta (0.4) and Gamma (0.4).

Q.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.

SOLUTION:

CHECKING STATIONARITY ON TRAIN DATA

The Augmented Dickey-Fuller test is a unit root test which determines whether there is a unit root and subsequently whether the series is non-stationary.

The hypothesis in a simple form for the ADF test is:

- H_0: The Time Series has a unit root and is thus non-stationary.
- H_1: The Time Series does not have a unit root and is thus stationary.

We would want the series to be stationary for building ARIMA models and thus we would want the p-value of this test to be less than the αα value.

FOR SPARKLING WINE

NO ORDER OF DIFFERENCING

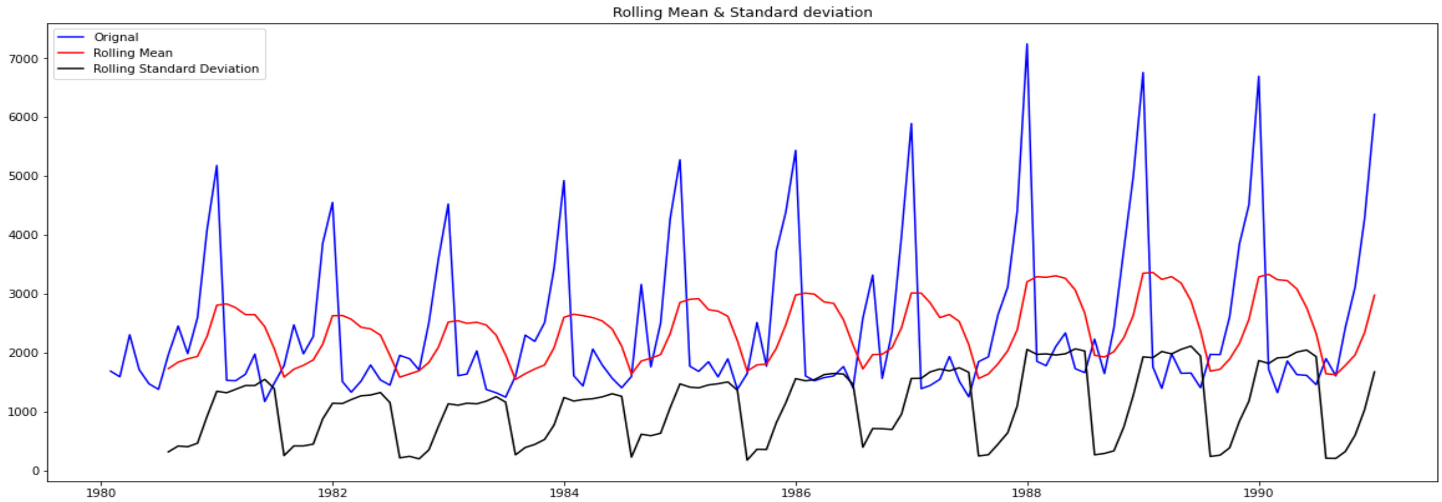


Figure 1. 33 – 0TH ODER DIFFERENCING PLOT FOR SPARKLING WINE

As we can see from the plot and ADF test that with there is some trend present in the data and p-value (0.669 > 0.05) which concludes that we fail to reject the null hypothesis that the time series is not stationary

```
Dicky Fuller test Result
T-stat value                -1.208926
p-value                      0.669744
Lags Used                   12.000000
Number of observations 119
```

1st ORDER OF DIFFERENCING

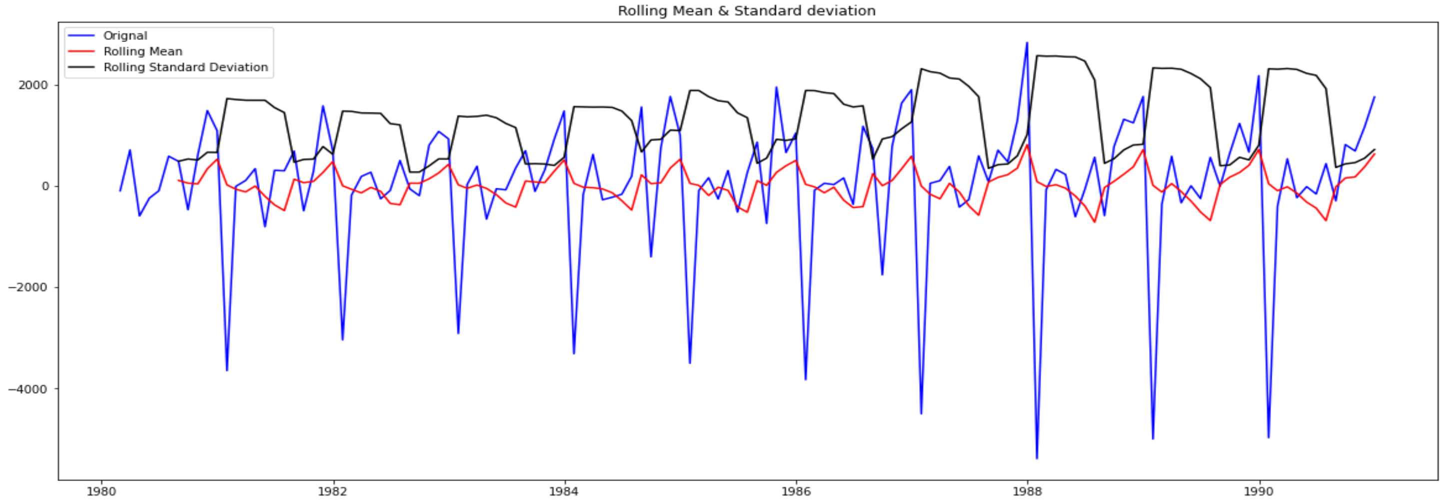


Figure 1. 34 – 1ST ODER DIFFERENCING PLOT FOR SPARKLING WINE

As we can see from the plot and ADF test that with there is no trend present in the data and p-value (2.280104e-12> 0.05) which concludes that we reject the null hypothesis, that the time series is stationary

```
Dicky Fuller test Result
T-stat value      -8.005007e+00
p-value            2.280104e-12
Lags Used          1.100000e+01
Number of observations 119
```

FOR ROSE WINE

NO ORDER OF DIFFERENCING

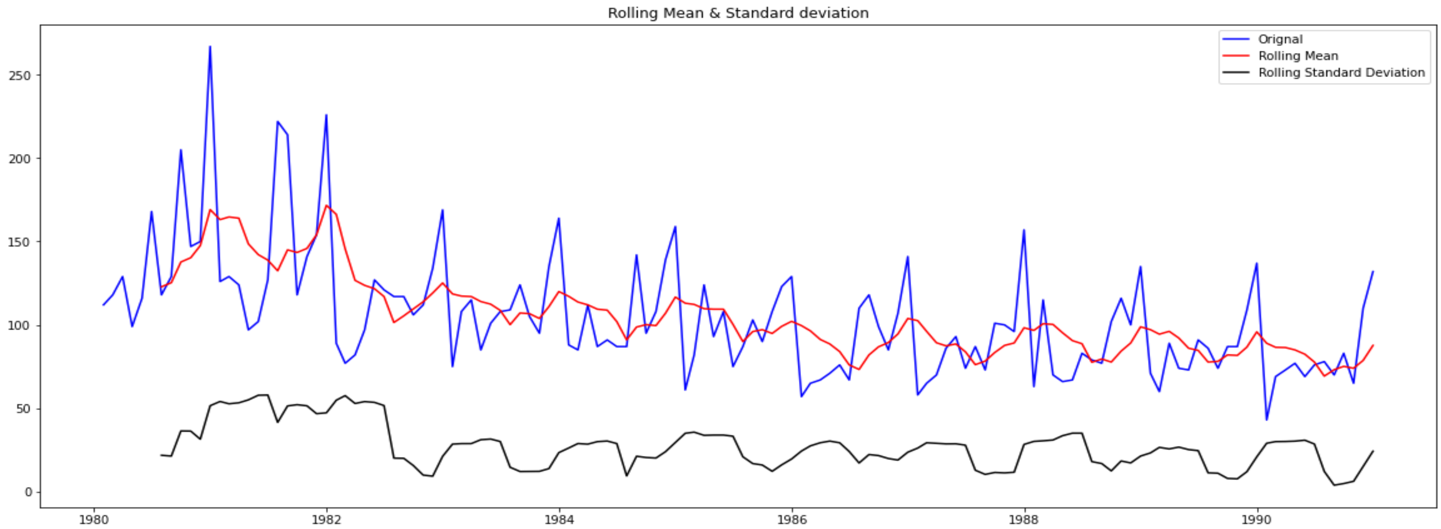


Figure 1. 35 - 0TH ODER DIFFERENCING PLOT FOR ROSE WINE

As we can see from the plot and ADF test that with there is some trend present in the data and p-value (0.219 > 0.05) which concludes that we fail to reject the null hypothesis that the time series is not stationary

```
Dicky Fuller test Result
T-stat value            -2.164250
p-value                  0.219476
Lags Used               13.000000
Number of observations 118
```

\

1st ORDER OF DIFFERENCING

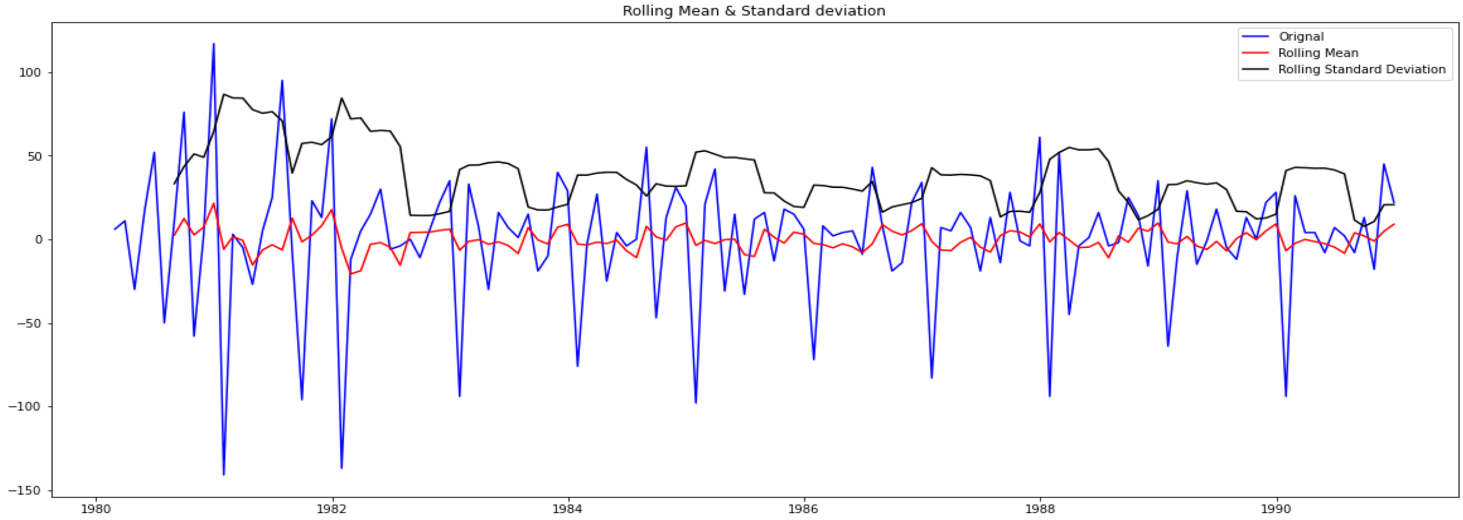


Figure 1. 36 – 1ST ODER DIFFERENCING PLOT FOR ROSE WINE

As we can see from the plot and ADF test that with there is no trend present in the data and p-value (7.061944e-09 > 0.05) which concludes that we reject the null hypothesis, that the time series is stationary

```
Dicky Fuller test Result
T-stat value          -6.592372e+00
p-value                7.061944e-09
Lags Used              1.200000e+01
Number of observations 118
```

Q-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.

SOLUTION:

FOR SPARKLING WINE

- ***AUTO ARIMA MODEL***

For Arima model we need to find the best combination of (p, d, q) to create a model and the choosing the best model which has the lowest Akaike Information Criteria value.

p and q value – range (0,4)

d value is 1 as 1st order differencing made time series stationary.

MODELS
Model : (0, 1, 0)
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)

	ARIMA MODEL	AIC VALUE
10	(2, 1, 2)	2213.509212
15	(3, 1, 3)	2221.459263
14	(3, 1, 2)	2230.759496
11	(2, 1, 3)	2232.890461
9	(2, 1, 1)	2233.777626
2	(0, 1, 2)	2233.994858
3	(0, 1, 3)	2234.408323
6	(1, 1, 2)	2234.5272
13	(3, 1, 1)	2235.498605
7	(1, 1, 3)	2235.607815
5	(1, 1, 1)	2235.755095
12	(3, 1, 0)	2257.723379
8	(2, 1, 0)	2260.365744
1	(0, 1, 1)	2263.060016
4	(1, 1, 0)	2266.608539
0	(0, 1, 0)	2267.663036

INFERENCE

Best model combination of (p, d, q) for ARIMA is (2,1,2) which has the lowest AIC value. using this model to predict test data and checking the performance measure by RMSE and MAPE.

ARIMA MODEL RESULT		
	TEST RMSE	TEST MAPE
ARIMA (2,1,2) MODEL RESULT	1299.9	43.2

- ***AUTO ARIMA MODEL DIAGNOSTICS.***

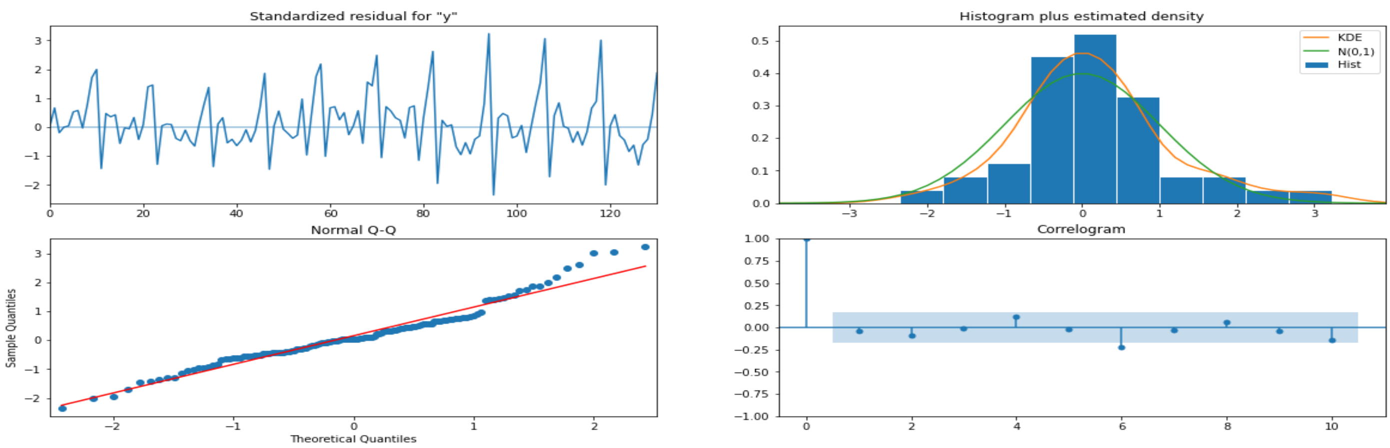


Figure 1. 37 – AUTO ARIMA DIAGNOSTIC PLOT FOR SPARKLING WINE

From the model diagnostics we can conclude that model performance can be improved.

➤ *AUTO ARIMA MODEL SUMMARY*

```
                              SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  132
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1101.755
Date:                Tue, 22 Feb 2022   AIC                           2213.509
Time:                        20:35:50   BIC                           2227.885
Sample:                             0   HQIC                          2219.351
                                - 132
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.3121      0.046     28.782      0.000       1.223       1.401
ar.L2         -0.5593      0.072     -7.741      0.000      -0.701      -0.418
ma.L1         -1.9917      0.109    -18.217      0.000      -2.206      -1.777
ma.L2          0.9999      0.110      9.109      0.000       0.785       1.215
sigma2      1.099e+06   1.99e-07   5.51e+12      0.000      1.1e+06     1.1e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.19   Jarque-Bera (JB):                14.46
Prob(Q):                              0.67   Prob(JB):                         0.00
Heteroskedasticity (H):               2.43   Skew:                             0.61
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.08
===================================================================================
```

Figure 1. 38 - AUTO ARIMA SUMMARY FOR SPARKLING WINE

➤ *AUTO SARIMA MODEL*

For Arima model we need to find the best combination of (p, d, q) (P, D, Q) to create a model and the choosing the best model which has the lowest Akaike Information Criteria value.

p, q, P and Q values – range (0,4)

d value is 1 as 1st order differencing made time series stationary.

MODELS
Model : (0, 1, 1)(0, 0, 1, 12)
Model : (0, 1, 2)(0, 0, 2, 12)
Model : (0, 1, 3)(0, 0, 3, 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 : (1, 1, 3)(1, 0, 3, 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)
Model : (2, 1, 3)(2, 0, 3, 12)
Model : (3, 1, 0)(3, 0, 0, 12)
Model : (3, 1, 1)(3, 0, 1, 12)
Model : (3, 1, 2)(3, 0, 2, 12)
Model : (3, 1, 3)(3, 0, 3, 12)

	SARIMA MODEL	SARIMA SEASONAL	AIC VALUE
220	(3, 1, 1)	(3, 0, 0, 12)	1387.8
252	(3, 1, 3)	(3, 0, 0, 12)	1387.9
237	(3, 1, 2)	(3, 0, 1, 12)	1388.6
221	(3, 1, 1)	(3, 0, 1, 12)	1388.7
222	(3, 1, 1)	(3, 0, 2, 12)	1389.2
...	...	...	...
107	(1, 1, 2)	(2, 0, 3, 12)	3922.2
39	(0, 1, 2)	(1, 0, 3, 12)	3935.7
99	(1, 1, 2)	(0, 0, 3, 12)	4140.7
103	(1, 1, 2)	(1, 0, 3, 12)	4180.4
131	(2, 1, 0)	(0, 0, 3, 12)	4268.9

INFERENCE

Best model combination of (p, d, q) and (P, D, Q) for SARIMA is (3,1,1) & (3,0,0,12) which has the lowest AIC value. Using this model to predict on test data and measuring the performance measure by RMSE and MAPE value

SARIMA MODEL PERFORMANCE		
	TEST RMSE	TEST MAPE
Auto Sarima Model (3,1,1) (3,0,0,12)	601.266	21.18

➢ *AUTO SARIMA MODEL DIAGNOSTICS.*

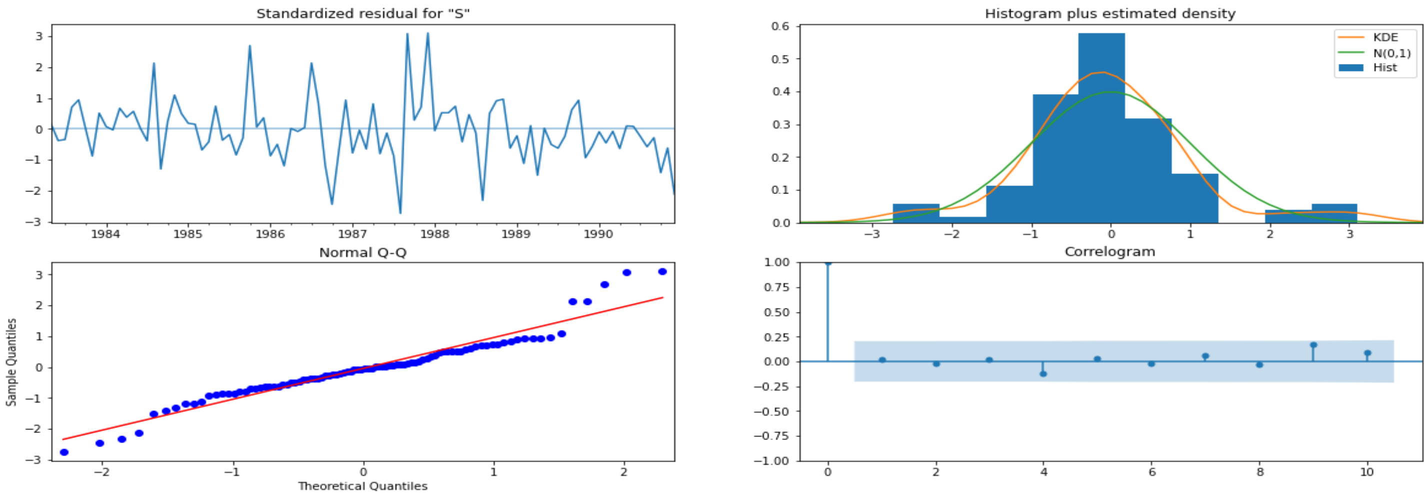


Figure 1. 39 - AUTO SARIMA DIAGNOSTIC PLOT FOR SPARKLING WINE

From the model diagnostics we can conclude that model performance better than Arima model but can be improved.

➢ *AUTO SARIMA MODEL SUMMARY*

```
                                      SARIMAX Results
======================================================================================
Dep. Variable:                        Sparkling   No. Observations:                132
Model:    SARIMAX(3, 1, 1)x(3, 0, [], 12)   Log Likelihood                -685.894
Date:                      Tue, 22 Feb 2022   AIC                           1387.788
Time:                              20:48:04   BIC                           1407.963
Sample:                          01-31-1980   HQIC                          1395.931
                               - 12-31-1990
Covariance Type:                        opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1615      0.150      1.075      0.282      -0.133       0.456
ar.L2         -0.0928      0.150     -0.617      0.537      -0.387       0.202
ar.L3          0.0917      0.136      0.676      0.499      -0.174       0.357
ma.L1         -0.9195      0.092    -10.032      0.000      -1.099      -0.740
ar.S.L12       0.5804      0.104      5.574      0.000       0.376       0.784
ar.S.L24       0.2559      0.119      2.159      0.031       0.024       0.488
ar.S.L36       0.2132      0.121      1.761      0.078      -0.024       0.451
sigma2      1.729e+05   2.18e+04      7.940      0.000      1.3e+05    2.16e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                18.77
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               1.08   Skew:                             0.47
Prob(H) (two-sided):                  0.84   Kurtosis:                         5.00
===================================================================================
```

Figure 1. 40 - AUTO SARIMA SUMMARY FOR SPARKLING WINE

FOR ROSE WINE

➢ ***AUTO ARIMA MODEL***

For Arima model we need to find the best combination of (p, d, q) to create a model and the choosing the best model which has the lowest Akaike Information Criteria value.

p and q value – range (0,4)

d value is 1 as 1st order differencing made time series stationary.

MODELS
Model : (0, 1, 0)
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)

	ARIMA MODEL	AIC VALUE
11	(2, 1, 3)	1274.695
15	(3, 1, 3)	1278.65653
2	(0, 1, 2)	1279.67153
6	(1, 1, 2)	1279.87072
3	(0, 1, 3)	1280.54538
5	(1, 1, 1)	1280.57423
9	(2, 1, 1)	1281.50786
10	(2, 1, 2)	1281.87072
7	(1, 1, 3)	1281.87072
1	(0, 1, 1)	1282.30983
13	(3, 1, 1)	1282.41928
14	(3, 1, 2)	1283.72074
12	(3, 1, 0)	1297.48109
8	(2, 1, 0)	1298.61103
4	(1, 1, 0)	1317.35031
0	(0, 1, 0)	1333.15467

INFERENCE

Best model combination of (p, d, q) for ARIMA is (2,1,3) which has the lowest AIC value. using this model to predict test data and checking the performance measure by RMSE and MAPE.

ARIMA MODEL RESULT		
	TEST RMSE	TEST MAPE
ARIMA (2,1,3) MODEL RESULT	36.87	64.06

➢ ***AUTO ARIMA MODEL DIAGNOSTICS.***

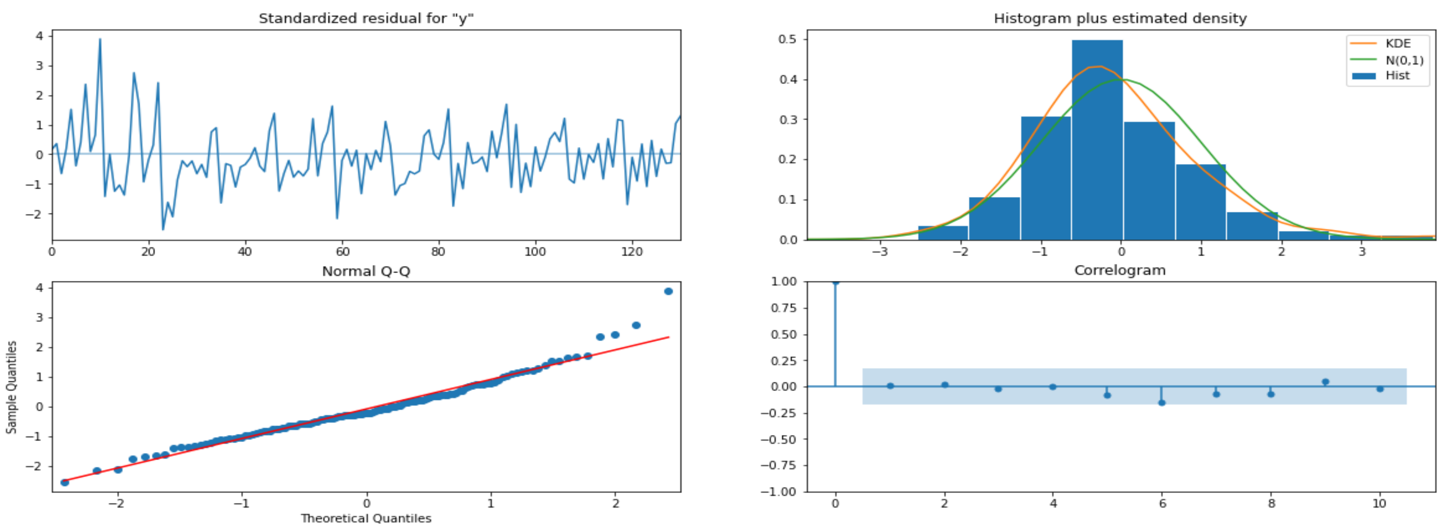


Figure 1. 41 - AUTO ARIMA DIAGNOSTIC PLOT FOR ROSE WINE

From the model diagnostics we can conclude that model performance can be improved.

➢ *AUTO ARIMA MODEL SUMMARY*

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  132
Model:               ARIMA(2, 1, 3)   Log Likelihood                -631.347
Date:              Tue, 22 Feb 2022   AIC                           1274.695
Time:                      20:48:08   BIC                           1291.946
Sample:                           0   HQIC                          1281.705
                              - 132
Covariance Type:                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6775      0.084    -20.007      0.000      -1.842      -1.513
ar.L2         -0.7283      0.084     -8.685      0.000      -0.893      -0.564
ma.L1          1.0448      0.662      1.579      0.114      -0.252       2.341
ma.L2         -0.7719      0.135     -5.713      0.000      -1.037      -0.507
ma.L3         -0.9047      0.600     -1.507      0.132      -2.082       0.272
sigma2       860.0026    557.970      1.541      0.123    -233.598    1953.603
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                24.40
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.40   Skew:                             0.71
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.57
===================================================================================
```

Figure 1. 42 - AUTO ARIMA SUMMARY FOR ROSE WINE

➢ *AUTO SARIMA MODEL*

For Arima model we need to find the best combination of (p, d, q) (P, D, Q) to create a model and the choosing the best model which has the lowest Akaike Information Criteria value.

p, q, P and Q values – range (0,4)

d value is 1 as 1st order differencing made time series stationary.

MODELS
Model : (0, 1, 1)(0, 0, 1, 12)
Model : (0, 1, 2)(0, 0, 2, 12)
Model : (0, 1, 3)(0, 0, 3, 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 : (1, 1, 3)(1, 0, 3, 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)
Model : (2, 1, 3)(2, 0, 3, 12)
Model : (3, 1, 0)(3, 0, 0, 12)
Model : (3, 1, 1)(3, 0, 1, 12)
Model : (3, 1, 2)(3, 0, 2, 12)
Model : (3, 1, 3)(3, 0, 3, 12)

	SARIMA PARAM	SARIMA SEASONAL	AIC VALUE
222	(3, 1, 1)	(3, 0, 2, 12)	774.4
238	(3, 1, 2)	(3, 0, 2, 12)	774.881
220	(3, 1, 1)	(3, 0, 0, 12)	775.427
221	(3, 1, 1)	(3, 0, 1, 12)	775.495
252	(3, 1, 3)	(3, 0, 0, 12)	775.561
...	...	...	...
135	(2, 1, 0)	(1, 0, 3, 12)	4053.6
199	(3, 1, 0)	(1, 0, 3, 12)	4055.6
131	(2, 1, 0)	(0, 0, 3, 12)	4069.19
195	(3, 1, 0)	(0, 0, 3, 12)	4071.19
7	(0, 1, 0)	(1, 0, 3, 12)	4554.89

INFERENCE

Best model combination of (p, d, q) and (P, D, Q) for SARIMA is (3,1,1) & (3,0,2,12) which has the lowest AIC value. Using this model to predict on test data and measuring the performance measure by RMSE and MAPE value

SARIMA MODEL PERFORMANCE		
	TEST RMSE	TEST MAPE
Auto Sarima Model (3,1,1) (3,0,2,12)	18.94	32.15

➢ *AUTO SARIMA MODEL DIAGNOSTICS.*

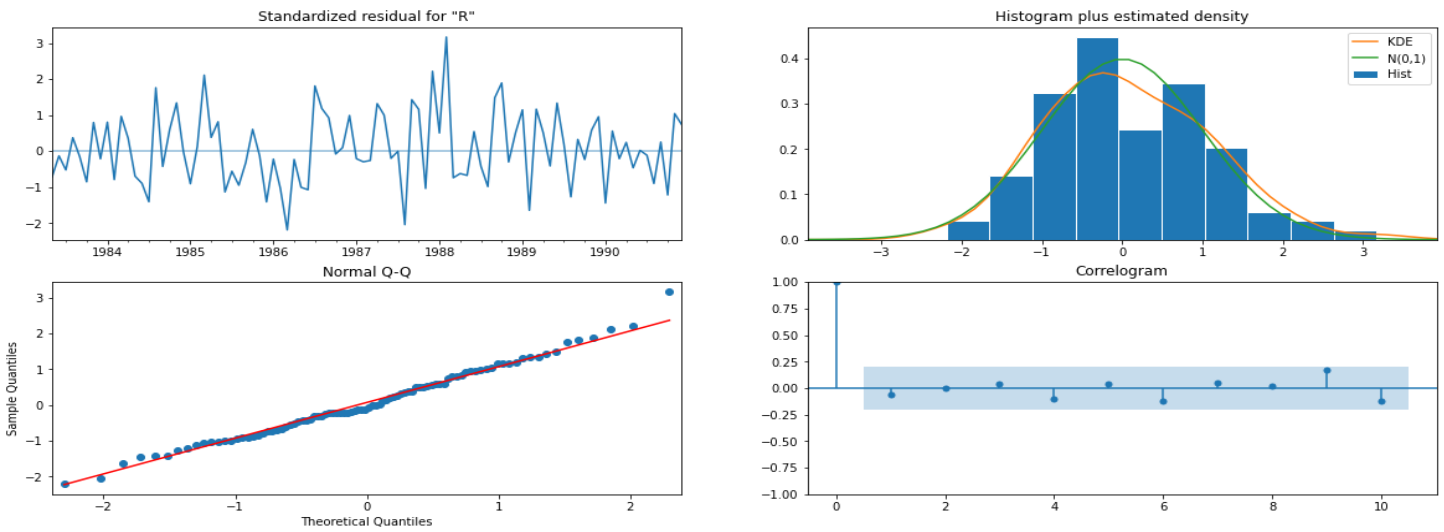


Figure 1. 43 - AUTO SARIMA DIAGNOSTIC PLOT FOR ROSE WINE

From the model diagnostics we can conclude that model performance is better than Arima model but can be improved.

➢ *AUTO SARIMA MODEL SUMMARY*

```
                                    SARIMAX Results
==========================================================================================
===========
Dep. Variable:                               Rose   No. Observations:                  132
Model:    SARIMAX(3, 1, 1)x(3, 0, [1, 2], 12)   Log Likelihood                -377.200
Date:                            Tue, 22 Feb 2022   AIC                            774.400
Time:                                    20:55:29   BIC                            799.618
Sample:                                01-31-1980   HQIC                           784.578
                                     - 12-31-1990
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0464      0.126      0.367      0.714      -0.202       0.294
ar.L2         -0.0060      0.120     -0.050      0.960      -0.241       0.229
ar.L3         -0.1808      0.098     -1.837      0.066      -0.374       0.012
ma.L1         -0.9370      0.067    -13.903      0.000      -1.069      -0.805
ar.S.L12       0.7639      0.165      4.639      0.000       0.441       1.087
ar.S.L24       0.0840      0.159      0.527      0.598      -0.229       0.397
ar.S.L36       0.0727      0.095      0.764      0.445      -0.114       0.259
ma.S.L12      -0.4968      0.250     -1.988      0.047      -0.987      -0.007
ma.S.L24      -0.2190      0.210     -1.044      0.297      -0.630       0.192
sigma2       192.1721     39.633      4.849      0.000     114.492     269.852
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                 1.64
Prob(Q):                              0.58   Prob(JB):                         0.44
Heteroskedasticity (H):               1.11   Skew:                             0.33
Prob(H) (two-sided):                  0.77   Kurtosis:                         3.03
===================================================================================
```

Figure 1. 44 - AUTO SARIMA SUMMARY FOR ROSE WINE

Q.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.

SOLUTION:

FOR SPARKLING WINE

- ***AUTO CORRELATION PLOT***

1st order ACF plot for the time Series

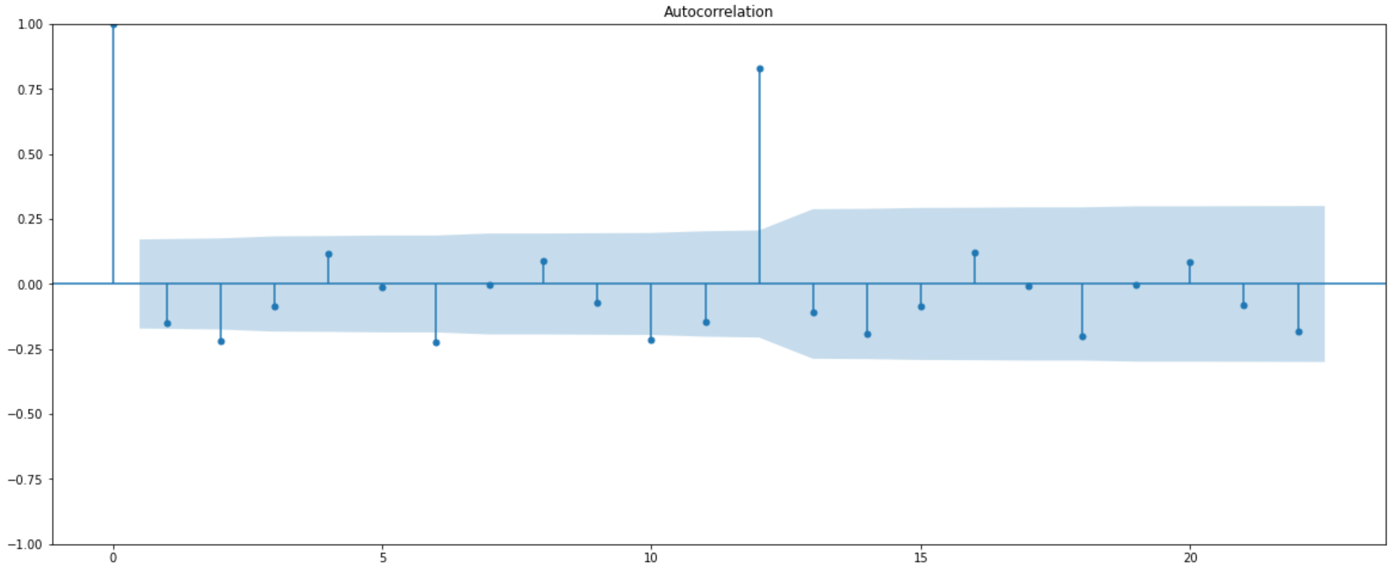


Figure 1. 45 – 1ST ACF PLOT FOR SPARKLING WINE

- ***PARTIAL AUTO CORRELATION PLOT***

1st order PACF plot for the time Series

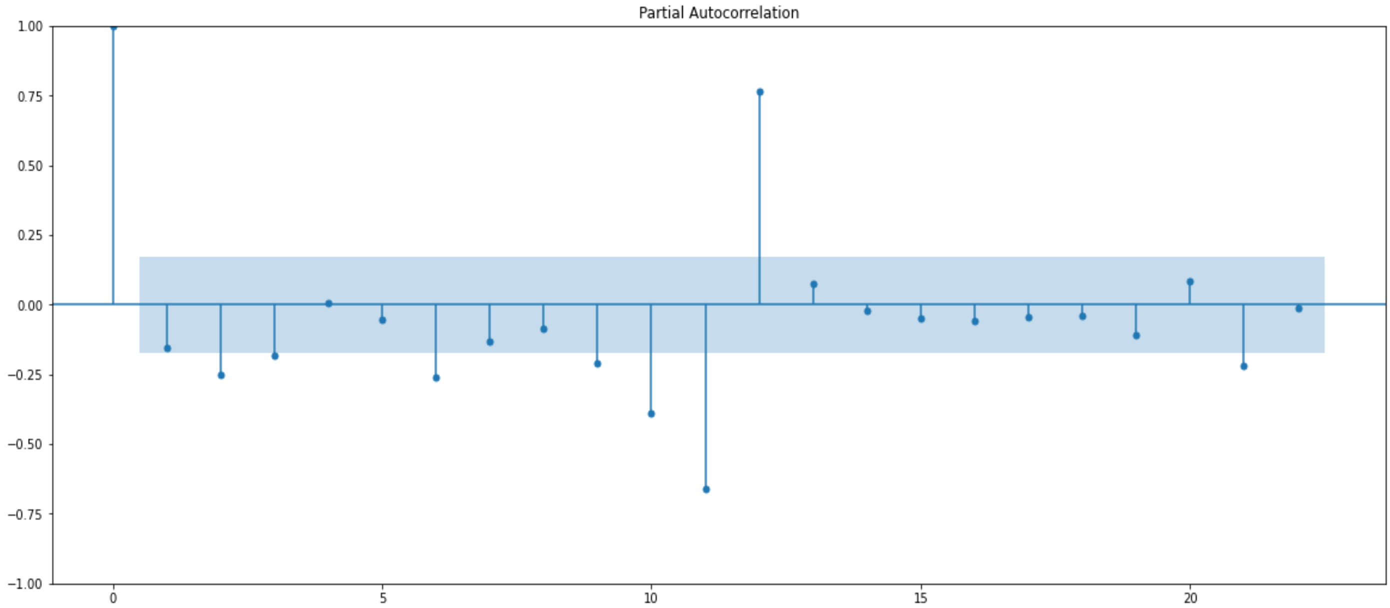


Figure 1. 46 - 1ST PACF PLOT FOR SPARKLING WINE

➢ *MANUAL ARIMA MODEL*

For Arima model we need to find the best combination of (p, d, q) using ACF and PACF plot to create a model.

1. p value – 0
2. q – value - 0
3. d value is 1 as 1st order differencing made time series stationary.

AUTO ARIMA MODEL DIAGNOSTICS.

ARIMA MODEL RESULT		
	TEST RMSE	TEST MAPE
ARIMA (0,1,0) MODEL RESULT	3864.27	152.87

INFERENCE

Best model combination of (p, d, q) for ARIMA is (0,1,0) which was selected using ACF and PACF plots and will use this model to predict test data and checking the performance measure by RMSE and MAPE.

➢ *MANUAL ARIMA DIAGNOSTIC PLOT*

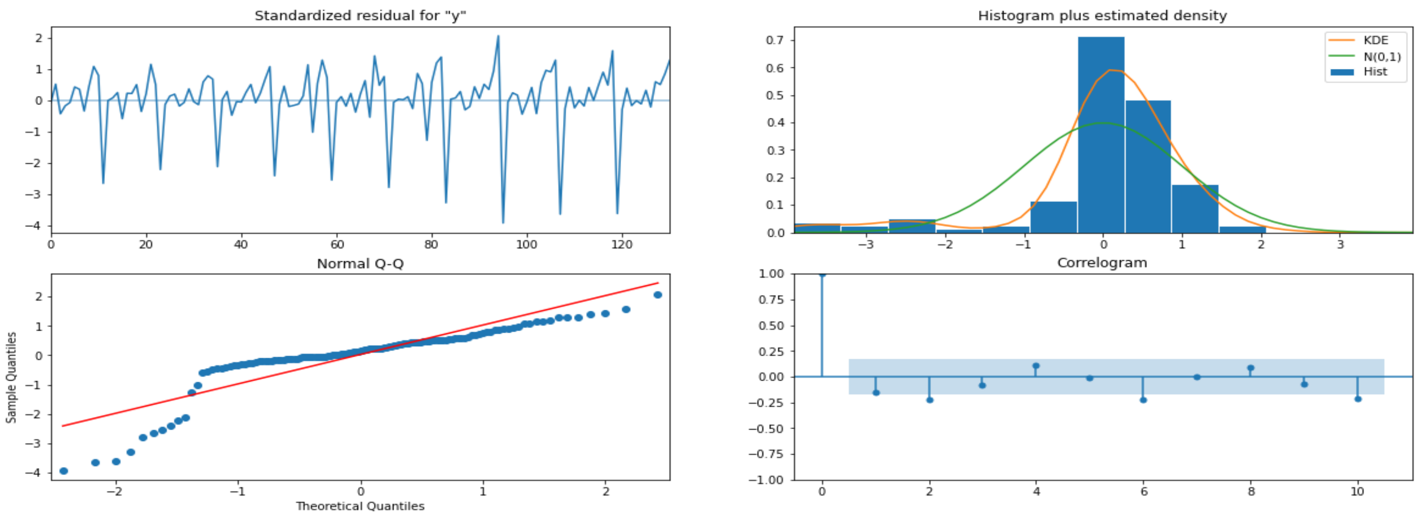


Figure 1. 47 – MANUAL ARIMA DIAGNOSTIC PLOT FOR SPARKLING WINE

From the model diagnostics we can conclude that model performance can be improved as Mape value are very high.

➢ *MANUAL ARIMA MODEL SUMMARY*

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  132
Model:                 ARIMA(0, 1, 0)   Log Likelihood               -1132.832
Date:                Tue, 22 Feb 2022   AIC                           2267.663
Time:                        21:14:46   BIC                           2270.538
Sample:                             0   HQIC                          2268.831
                                - 132
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      1.885e+06   1.29e+05     14.658      0.000    1.63e+06    2.14e+06
===================================================================================
Ljung-Box (L1) (Q):                   3.07   Jarque-Bera (JB):               198.83
Prob(Q):                              0.08   Prob(JB):                         0.00
Heteroskedasticity (H):               2.46   Skew:                            -1.92
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.65
===================================================================================
```

Figure 1. 48 - MANUAL ARIMA SUMMARY FOR SPARKLING WINE

➢ *MANUAL SARIMA MODEL*

For Arima model we need to find the best combination of (p, d, q) (P, D, Q) to create a model using ACF and PACF plots of the time series.

1. p – value - 0
2. q – value - 0
3. P – value - 4
4. Q - values – 0
5. d value is 1 as 1st order differencing made time series stationary.

SARIMA MODEL PERFORMANCE		
	TEST RMSE	TEST MAPE
Manual Sarima Model (0,1,0) (0,0,4,12)	1991.42	71.68

INFERENCE

Best model combination of (p, d, q) and (P, D, Q) for SARIMA is (0,1,0) & (0,0,4,12) which has been selecting using the ACF and PACF plots. Using this model to predict on test data and measuring the performance measure by RMSE and MAPE value

➢ *MANUAL SARIMA MODEL DIAGNOSTICS.*

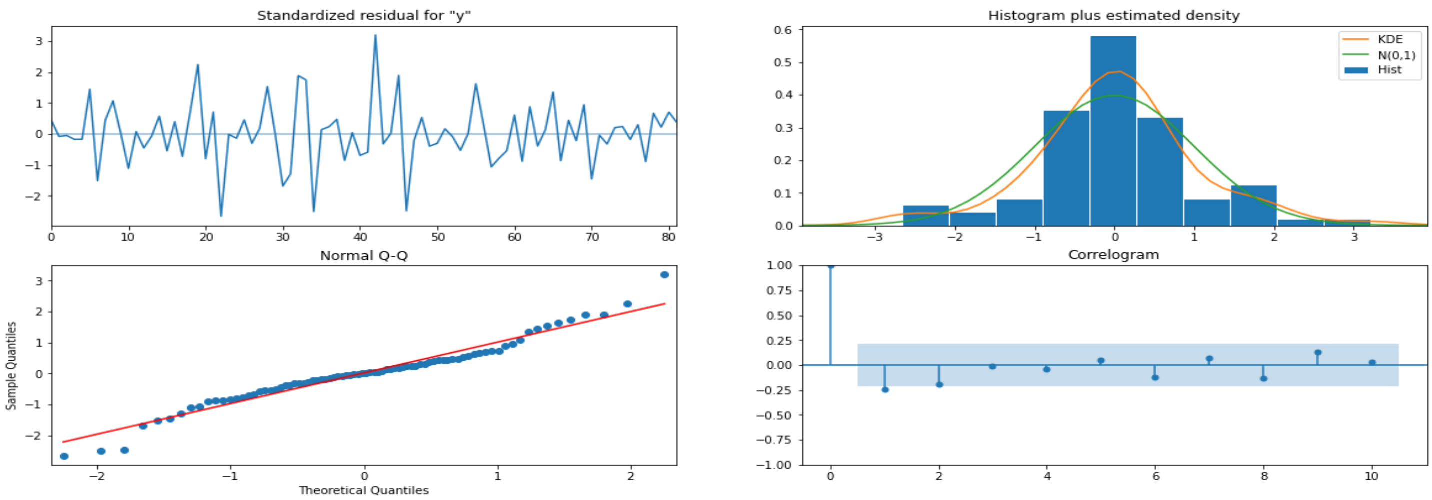


Figure 1. 49 - MANUAL SARIMA DIAGNOSTIC PLOT FOR SPARKLING WINE

From the model diagnostics we can conclude that model performance better than Arima model but can be improved as performance measure are still very high.

➢ *MANUAL SARIMA MODEL SUMMARY*

```
                                  SARIMAX Results
=====================================================================================
Dep. Variable:                               y   No. Observations:              132
Model:   SARIMAX(0, 1, 0)x(0, 0, [1, 2, 3, 4], 12)   Log Likelihood          -648.636
Date:                         Tue, 22 Feb 2022   AIC                       1307.272
Time:                                 21:16:34   BIC                       1319.305
Sample:                                      0   HQIC                      1312.103
                                         - 132
Covariance Type:                           opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.S.L12       1.3205      0.618      2.136      0.033       0.109       2.532
ma.S.L24       1.3428      0.625      2.149      0.032       0.118       2.567
ma.S.L36       1.1659      0.639      1.825      0.068      -0.086       2.418
ma.S.L48       1.3065      0.727      1.797      0.072      -0.118       2.731
sigma2      1.853e+05   1.52e+05      1.220      0.222   -1.12e+05    4.83e+05
===================================================================================
Ljung-Box (L1) (Q):                   5.04   Jarque-Bera (JB):                 6.46
Prob(Q):                              0.02   Prob(JB):                         0.04
Heteroskedasticity (H):               0.66   Skew:                             0.08
Prob(H) (two-sided):                  0.28   Kurtosis:                         4.37
===================================================================================
```

Figure 1. 50 - MANUAL SARIMA SUMMARY FOR SPARKLING WINE

FOR ROSE WINE

- ***AUTO CORRELATION PLOT***

1st order ACF plot for the time Series

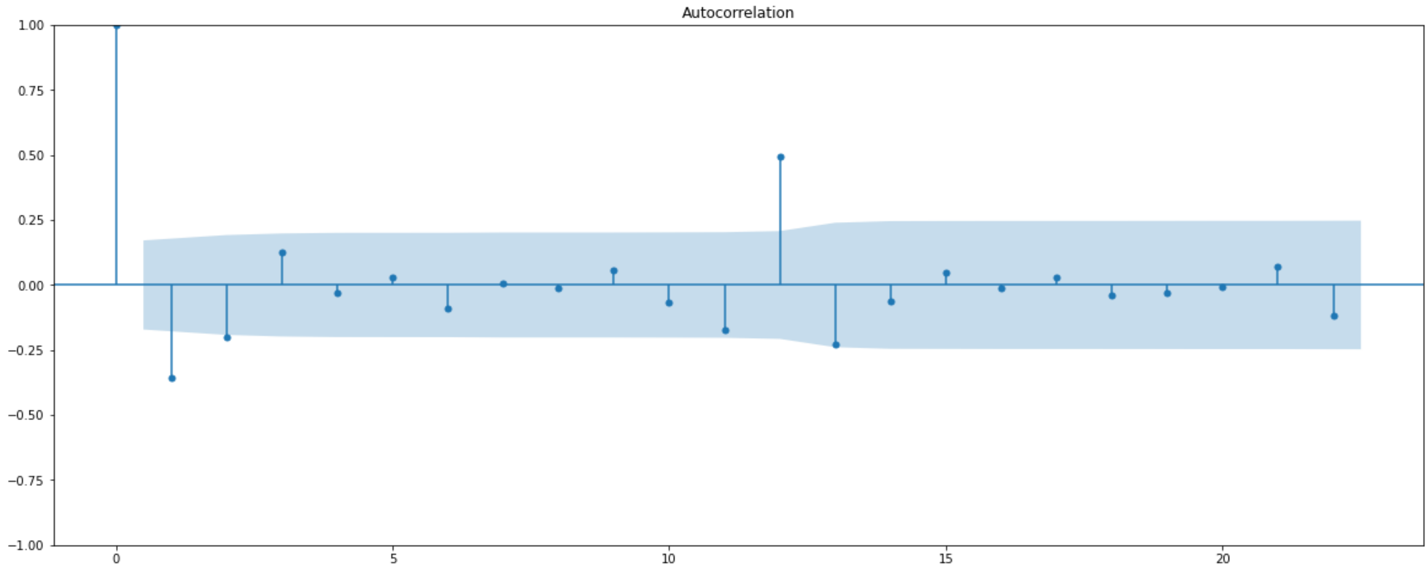


Figure 1. 51 – 1ST ORDER ACF PLOT FOR ROSE WINE

- ***PARTIAL AUTO CORRELATION PLOT***

1st order PACF plot for the time Series

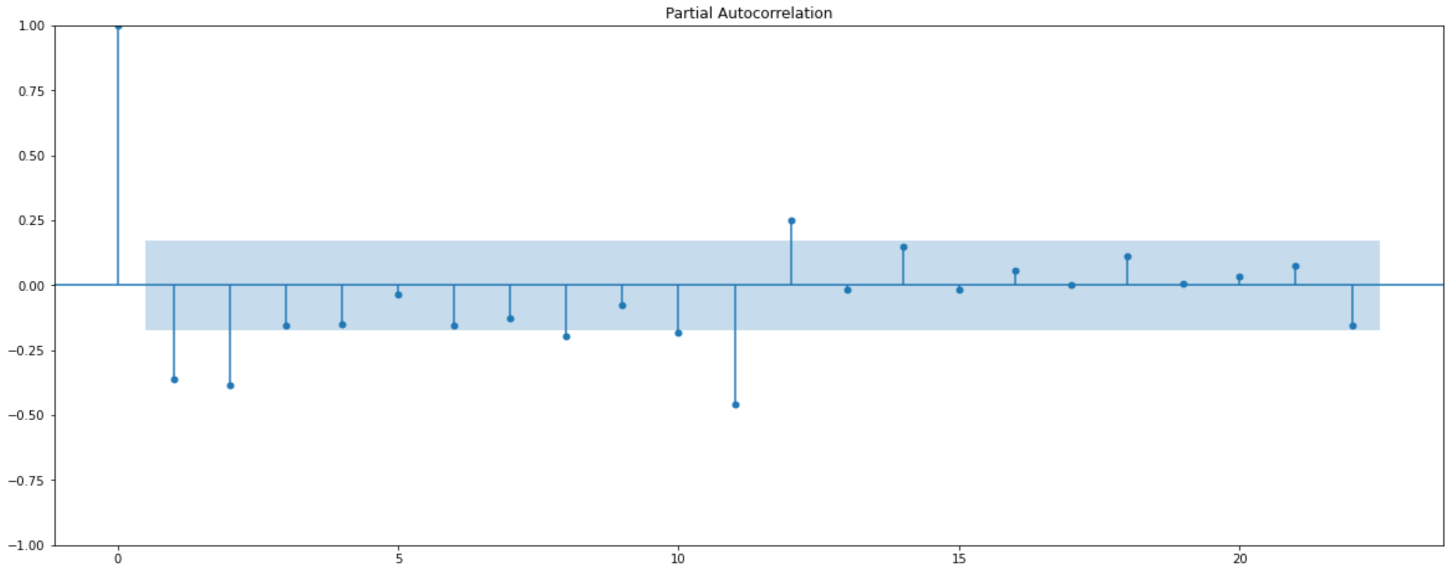


Figure 1. 52 - 1ST ORDER PACF PLOT FOR ROSE WINE

- ***MANUAL ARIMA MODEL***

For Arima model we need to find the best combination of (p, d, q) using ACF and PACF plot to create a model.

1. p value – 2
2. q – value - 3
3. d value is 1 as 1st order differencing made time series stationary.

ARIMA MODEL RESULT		
	TEST RMSE	TEST MAPE
ARIMA (2,1,3) MODEL RESULT	36.87	64.06

INFERENCE

Best model combination of (p, d, q) for ARIMA is (2,1,3) which was selected using ACF and PACF plots and will use this model to predict test data and checking the performance measure by RMSE and MAPE.

- ***MANUAL ARIMA DIAGNOSTIC PLOT***

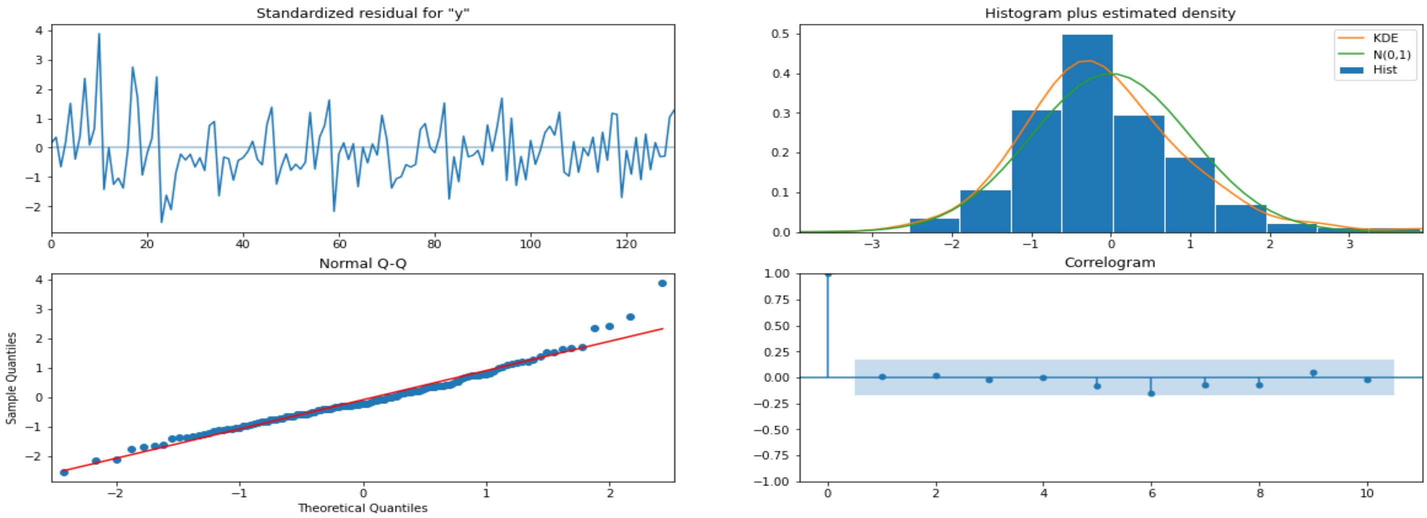


Figure 1. 53 – MAUAL ARIMA DIAGNOSTIC PLOT FOR ROSE WINE

From the model diagnostics we can conclude that model performance can be improved as Mape value are very high.

- ***MANUAL ARIMA MODEL SUMMARY***

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  132
Model:                 ARIMA(2, 1, 3)   Log Likelihood                -631.347
Date:                Tue, 22 Feb 2022   AIC                           1274.695
Time:                        21:18:12   BIC                           1291.946
Sample:                             0   HQIC                          1281.705
                                - 132
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6775      0.084    -20.007      0.000      -1.842      -1.513
ar.L2         -0.7283      0.084     -8.685      0.000      -0.893      -0.564
ma.L1          1.0448      0.662      1.579      0.114      -0.252       2.341
ma.L2         -0.7719      0.135     -5.713      0.000      -1.037      -0.507
ma.L3         -0.9047      0.600     -1.507      0.132      -2.082       0.272
sigma2       860.0026    557.970      1.541      0.123    -233.598    1953.603
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                24.40
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.40   Skew:                             0.71
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.57
===================================================================================
```

Figure 1. 54 - MAUAL ARIMA SUMMARY FOR ROSE WINE

- ***MANUAL SARIMA MODEL***

For Arima model we need to find the best combination of (p, d, q) (P, D, Q) to create a model using ACF and PACF plots of the time series.

1. p – value - 2
2. q – value - 3
3. P – value - 0
4. Q - values – 0
5. d value is 1 as 1st order differencing made time series stationary.

SARIMA MODEL PERFORMANCE		
	TEST RMSE	TEST MAPE
Auto Sarima Model (2,1,3) (0,0,0,12)	36.25	62.88

INFERENCE

Best model combination of (p, d, q) and (P, D, Q) for SARIMA is (2,1,3) & (0,0,0,12) which has been selecting using the ACF and PACF plots. Using this model to predict on test data and measuring the performance measure by RMSE and MAPE value

➢ *MANUAL SARIMA MODEL DIAGNOSTICS.*

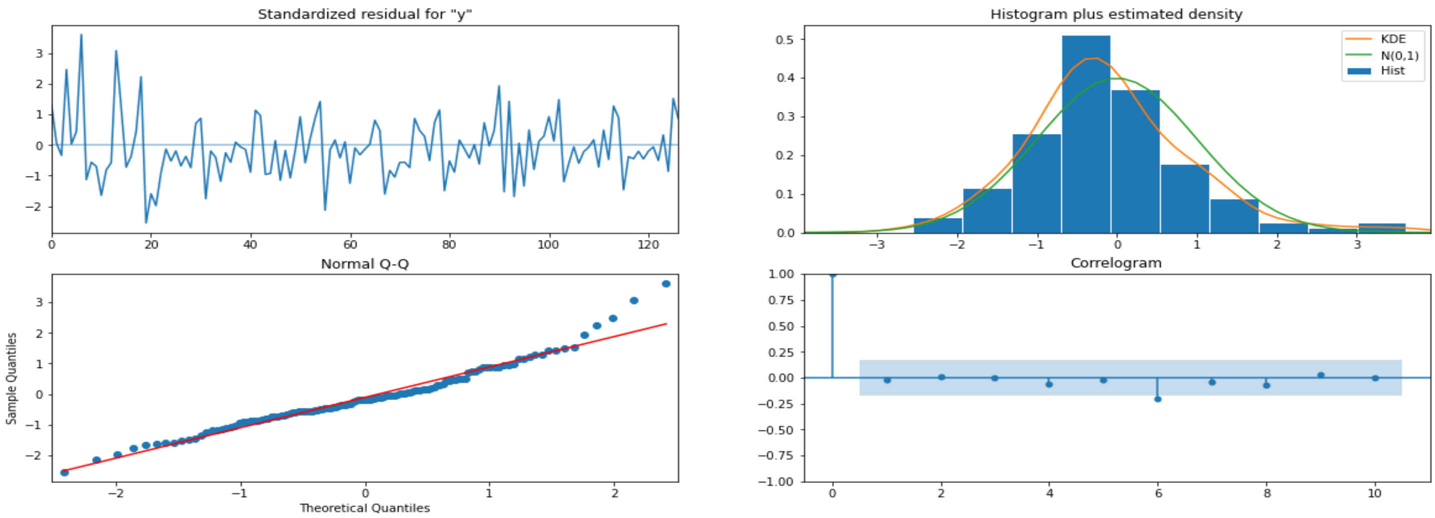


Figure 1. 55 - MAUAL SARIMA DIAGNOSTIC PLOT FOR ROSE WINE

From the model diagnostics we can conclude that model performance better than Arima model but can be improved as performance measure are still very high.

➢ *MANUAL SARIMA MODEL SUMMARY*

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  132
Model:               SARIMAX(2, 1, 3)   Log Likelihood                -612.935
Date:                Tue, 22 Feb 2022   AIC                           1237.870
Time:                        21:18:48   BIC                           1254.935
Sample:                             0   HQIC                          1244.804
                                - 132
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.5766      0.109    -14.499      0.000      -1.790      -1.363
ar.L2         -0.5869      0.110     -5.338      0.000      -0.802      -0.371
ma.L1          0.9786      0.100      9.766      0.000       0.782       1.175
ma.L2         -0.8199      0.103     -7.976      0.000      -1.021      -0.618
ma.L3         -0.8200      0.084     -9.820      0.000      -0.984      -0.656
sigma2       869.8914     93.757      9.278      0.000     686.131    1053.651
===================================================================================
Ljung-Box (L1) (Q):                   0.08   Jarque-Bera (JB):                23.60
Prob(Q):                              0.77   Prob(JB):                         0.00
Heteroskedasticity (H):               0.45   Skew:                             0.73
Prob(H) (two-sided):                  0.01   Kurtosis:                         4.53
```

Figure 1. 56 - MAUAL SARIMA SUMMARY FOR ROSE WINE

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

SOLUTION:

FOR SPARKLING WINE

MODELS	TEST RMSE	TEST MAPE
Triple Exponential Alpha(0.4) Beta(0.1) Gamma(0.2)	317.434302	9.64
Triple Exponential Alpha(0.07)ta(0.000) Gamma(0.39)	403.706228	13.91
Auto Sarima Model (3,1,1) (3,0,0,12)	601.26622	21.18
2pointTrailingMovingAverage	813.400684	19.7
Manual Sarima Model (1,1,3) (0,0,4,12)	911.986049	27.87
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1156.589694	43.86
Simple Average	1275.081804	38.9
Single Exponential Smoothing Alpha(0.01)	1286.648058	35.78
Auto Arima (2,1,2)	1299.980084	43.2
Manual Arima Model (2,1,1)	1308.89794	82.73
Simple Exponential Alpha(0.070)	1338.004623	47.11
9pointTrailingMovingAverage	1346.278315	46.86
Regression On Time	1389.135175	50.15
Double Exponential Smoothing Alpha(0.02) Beta(0.95)	1498.738766	55.31
Manual Sarima Model (1,1,3) (0,0,4,12)	1991.42288	71.68
Naive Forecast	3864.279352	152.87
Manual Arima Model (0,1,0)	3864.279352	152.87
Double Exponential Alpha(0.66) Beta(0.0001)	5291.879833	208.74

TABLE 1. 23 – SPARKLING WINE ALL MODELS PERFORMANCE RESULT

FOR ROSE WINE

MODELS	TEST RMSE	TEST MAPE
Triple Exponential Smoothing Alpha(0.3) Beta(0.4) Gamma(0.4)	10.347078	13.3
2pointTrailingMovingAverage	11.53018	13.6
Double Exponential Smoothing Alpha(0.03) Beta(0.65)	15.331019	18.61
Triple Exponential Alpha(0.08)Beta(5.52) Gamma(0.0005)	14.285684	19.37
4pointTrailingMovingAverage	14.46233	19.59
6pointTrailingMovingAverage	14.46233	20.83
9pointTrailingMovingAverage	14.740112	21.13
Regression On Time	15.29146	22.94
Double Exponential Alpha(0.0) Beta(0.0000)	15.293494	22.95
Auto Sarima Model (3,1,1) (3,0,2,12)	18.946998	32.15
Manual Sarima Model (2,1,5) (0,0,6,12)	28.87605	48.31
Manual Sarima Model (2,1,5) (0,0,2,12)	30.939303	51.75
Manual Sarima Model (2,1,5) (0,0,2,12)	36.253116	62.88
Single Exponential Smoothing Alpha(0.07)	36.49737	63.38
Simple Exponential Alpha(0.098)	36.858569	64.05
Manual Arima Model (2,1,3)	36.87608	64.06
Auto Arima (2,1,3)	36.87608	64.06
Manual Arima Model (2,1,5)	37.26752	64.84
Simple Average	53.521557	95.13
Naive Forecast	79.778066	145.35

TABLE 1. 24 - ROSE WINE ALL MODELS PERFORMANCE RESULT

INFERENCE

SPARKLING WINE.

From the table there are multiple models that were created for this time series in which **"Triple Exponential Smoothening Alpha (0.4) Beta(0.1) Gamma(0.2)"** with given parameters performed best with the lowest RMSE (317.43) and MAPE (9.64) scores. So, we are going to choose this model to future forecast for the time Series.

ROSE WINE.

From the table there are multiple models that were created for this time series in which **"Triple Exponential Smoothening Alpha (0.3) Beta (0.4) Gamma (0.4)"** with given parameters performed best with the lowest RMSE (10.34) and MAPE (13.3) scores. So, we are going to choose this model to future forecast for the time Series.

Q.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.

SOLUTION:

FOR SPARKLING WINE

After choosing the most optimum model with best performance measure i.e., Triple Exponential smoothening Model with parameters Alpha (0.4), Beta (0.1) and Gamma (0.2), We built the model on the whole data and forecasted the next 12 months sales for the sparkling wine.

Sparkling wine forecast with confidence interval for 12 months (Aug-1995 to July 1996)

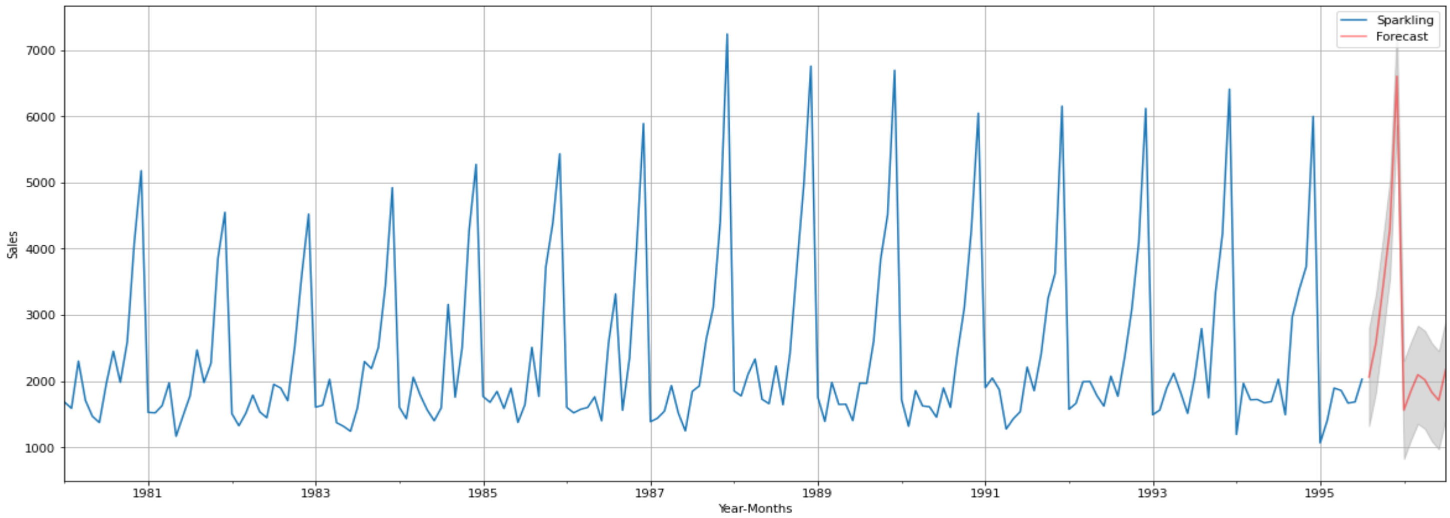


Figure 1. 57 – SPARKLING WINE FUTURE SALES FORECAST FOR NEXT 12 MONTHS VISUAL

FOR ROSE WINE

After choosing the most optimum model with best performance measure i.e., Triple Exponential smoothening Model with parameters Alpha (0.3), Beta (0.4) and Gamma (0.4), We built the model on the whole data and forecasted the next 12 months sales for the rose wine.

Rose wine forecast with confidence interval for 12 months (Aug-1995 to July 1996)

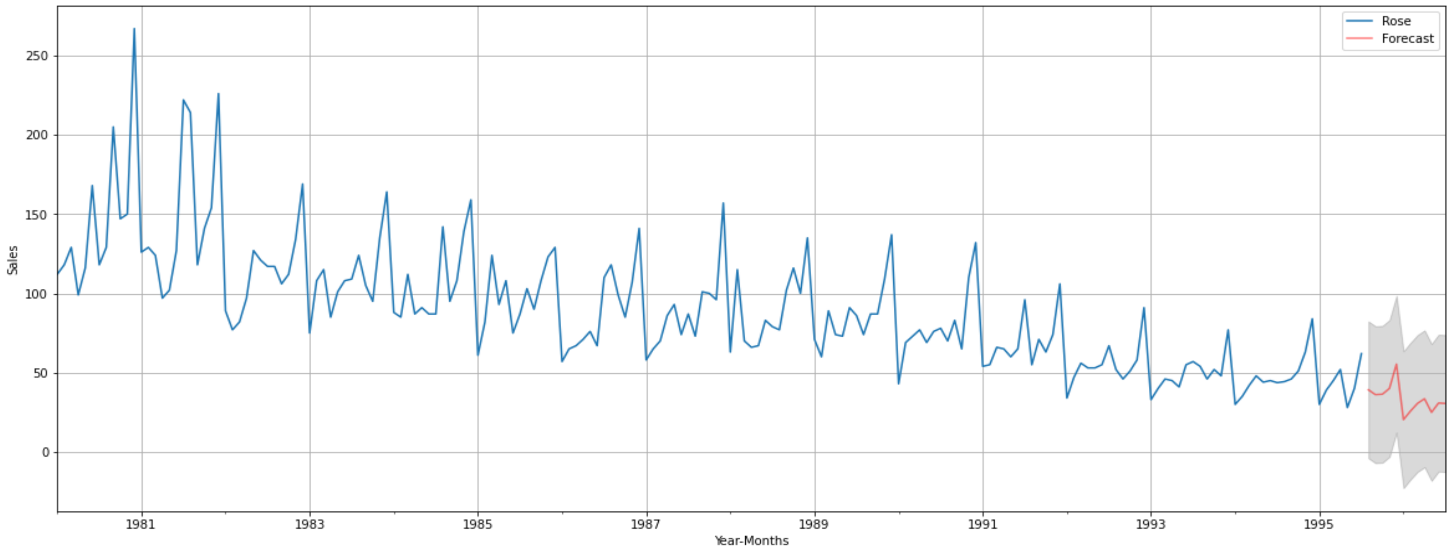


Figure 1. 58 - ROSE WINE FUTURE SALES FORECAST FOR NEXT 12 MONTHS VISUAL

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

SOLUTION:

FOR SPARKLING WINE

SPARKLING WINE FORECAST	
MONTHS	SALES
31-08-1995	2063.449006
30-09-1995	2579.407584
31-10-1995	3416.654489
30-11-1995	4304.477396
31-12-1995	6604.876921
31-01-1996	1564.539905
29-02-1996	1849.760079
31-03-1996	2098.878914
30-04-1996	2022.428902
31-05-1996	1834.540669
30-06-1996	1712.408912
31-07-1996	2176.425517

TABLE 1. 25 – SPARKLING WINE NEXT 12 MONTH FORECAST

COMMENTS

The model forecast is well followed the existing pattern and forecasted good sales values for the next 12 months. This forecast sales values can really help company to be prepared for the future and take some measures to improve sales.

MEASURES

1. December 1995 has been forecasted for the maximum Wine sales so company can prepare the wine stocks accordingly.
2. Company can precure raw materials according to their forecast sales so that company don't have to face problems like over or under purchase of the material.
3. Company can check months like January and February have less sales forecast and can prepare some offers and discounts accordingly.

FOR ROSE WINE

ROSE WINE FORECAST	
MONTHS	**SALES**
31-08-1995	39.201883
30-09-1995	36.138936
31-10-1995	36.415437
30-11-1995	40.082038
31-12-1995	55.368083
31-01-1996	20.368932
29-02-1996	25.785196
31-03-1996	30.609484
30-04-1996	33.559613
31-05-1996	24.98005
30-06-1996	30.832509
31-07-1996	30.590528

TABLE 1. 26 - ROSE WINE NEXT 12 MONTH FORECAST

COMMENTS

The model forecast is well followed the existing pattern and forecasted good sales values for the next 12 months. As the pattern suggested that winter months have the most wine sales. But as the trend suggested rose wine sales have decreased significantly over the years.

MEASURES

1. December 1995 has been forecasted for the maximum Wine sales so company can prepare the wine stocks accordingly, as the sales decreased so that company won't have to bear too much loss.
2. As the trend suggest decreased sales for the wine, company should go for some opt for some good marketing strategies or should think of removing the product from the market.
3. Company can opt for the production of wine only for those months which has peak sales.
4. Wine production for only few months can also be used as marketing strategy to increase demand.