TIME SERIES FORECASTING
PROJECT

ABC ESTATE WINES
CASE - ROSE WINE

SABITA NAIR PANCHAL

BATCH - DSBA (FEBRUARY, 2021)

SUBMISSION DATE - 10/10/2021

TABLE OF CONTENTS

LIST OF TABLES		
Table no.:	**Description:**	**Page no.:**
1.1	First five rows of dataset	4
1.2.1	Statistical summary	7
1.2.2	Pivot table (years v/s months)	9
1.3	Train and test data	18
1.4.6	Various Moving average predictions on test data	29
1.6	AIC values for various SARIMA models	35
1.8	Model-wise RMSE	41

Executive Summary:

ABC Estate Wines have asked for an analysis into the sales pattern of their brand 'Rose' wine. A data of monthly Rose wine sales from January, 1980 to July 1995 has been shared. Basis on this, the company is also expecting a future forecast of the sales of this brand of wine into the next 12 months.

1.1 Reading the Time series:

The data of the 'Rose' wine sales is a classic example of a univariate time series with monthly frequency. The first five rows of the dataset is given in the table below:

Table: 1.1

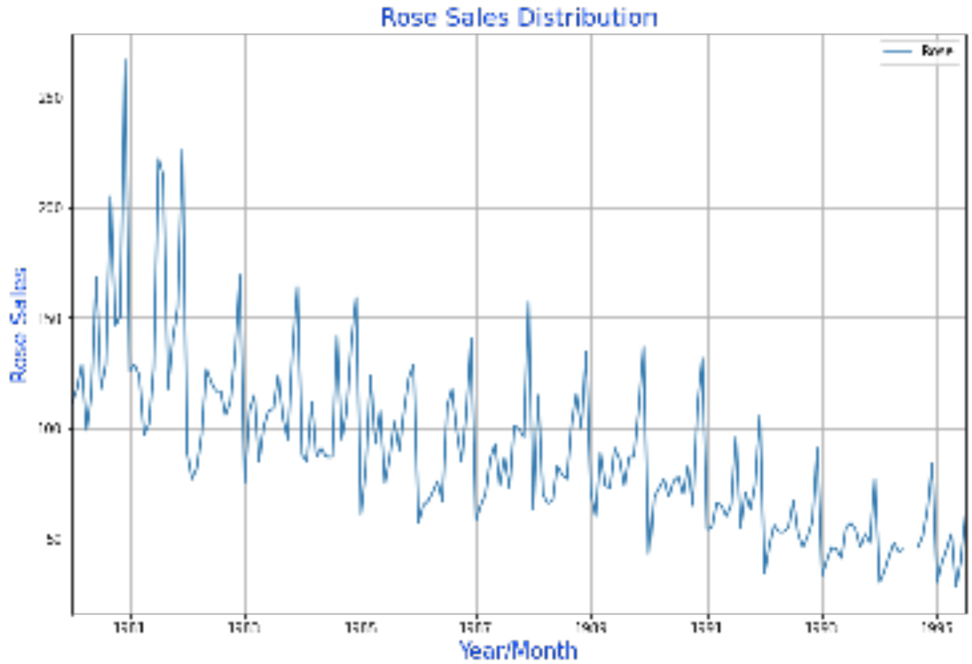
YearMonth	Rose
1980-01-01	112
1980-02-01	118
1980-03-01	129
1980-04-01	99
1980-05-01	116

There are 187 rows in the dataset, and two null/missing values. The data description is given below:

YearMonth	represents the month and year (Parsing was done to convert this into DateTimeIndex, essential for Time series analysis)
Rose	represents the monthly sales of the 'Rose' brand of wine

A time series plot was used to graphically represent the distribution of 'Rose' wine sales over the years (as shown in figure below).

Figure: 1.1



Inferences:

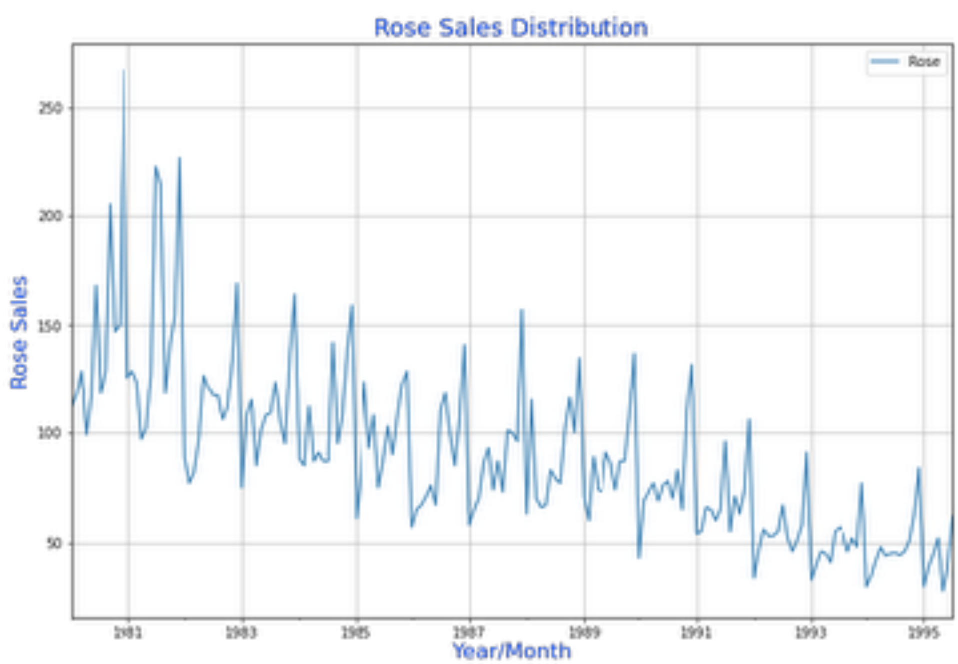
- The data is contiguous, without any change in time sequence.
- There are two apparent missing values.
- There is a very evident downward trend. After a brief period of increase in 1980, sales seemed to drop from 1981 till 1986. Then, there was a brief upward movement in sales till 1988. Thereafter, there has been a steady decrease to 1995.
- The seasonality factor is evident here - most years have seen a sharp spike in sales around the same time - this can only be attributed to a seasonal component.

1.2 Exploratory data analysis:

—> Missing value imputation:

- The dataset was checked for null and missing values, using the isnull() function. It revealed that there were two missing values.
- In a time series, each observation is time-dependent, missing values lead to a definite loss of information, leading to inefficiencies in the forecasting.
- Hence, missing values were interpolated using the ‘polynomial’ method of order 2. Post interpolation plot is given below:

Figure: 1.2.1



—> **Statistical summary of the data**:

Statistical summary of the dataset was checked using the describe() function, the results are given below:

Table: 1.2.1

	Rose
count	187.00
mean	89.91
std	39.25
min	28.00
25%	62.50
50%	85.00
75%	111.00
max	267.00

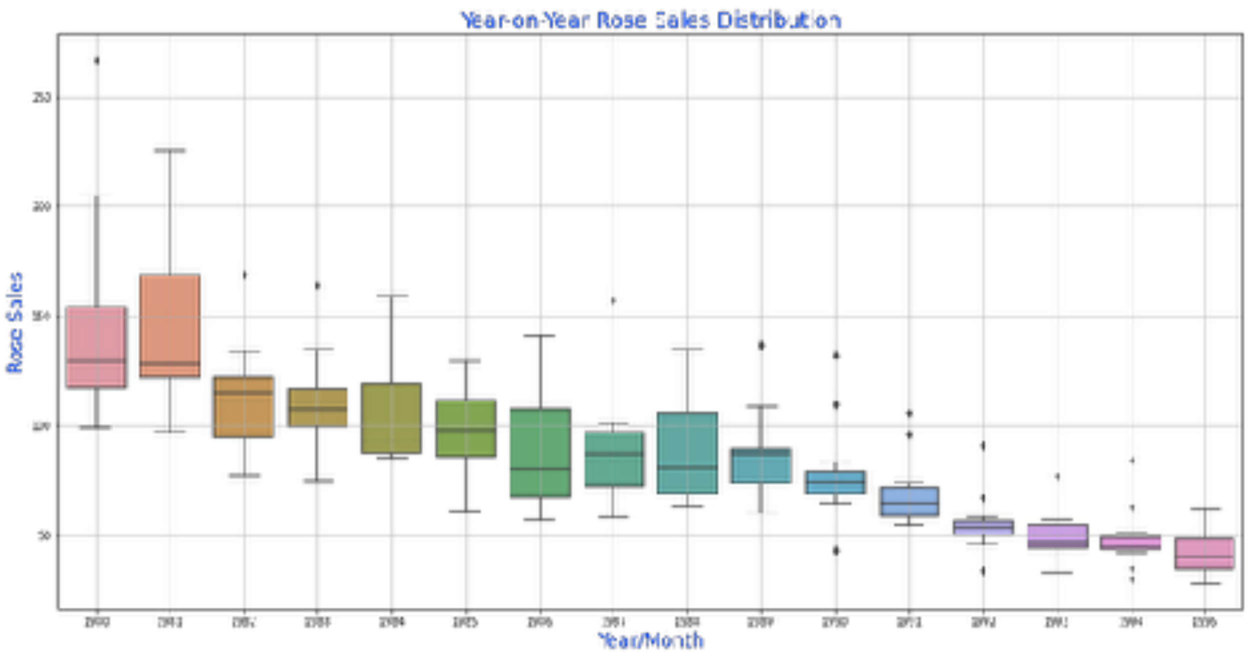
Inferences: The following can be inferred from the above statistics:

- The average monthly sales of ‘Rose’ wine is approx. 90 units.
- The minimum sales registered in a month is 28, while the maximum sales registered is 267.
- The mean of the observations (89.91) is not very far off from the median (85) - proving that the wine sales figures follow a somewhat normal distribution, i.e., they are fairly centered around the mean value.
- Also, a standard deviation of 39.25 shows that there is some variation in the data.

—-> **Univariate analysis**:

Year-on-year spread of ‘Rose’ sales:

Figure: 1.2.2



Inferences: This plot reveals the distribution of sales of ‘Rose’ wine per year. We can observe the following:

- The largest sales distribution is evident in the years 1981 followed by 1980. The least distribution was in years 1994, 1992, 1990 and 1993.
- There is evidently a repeating rise-and-fall pattern in sales between 1982 and 1989, after which a decreasing trend is clearly seen.
- The year 1995 will not be considered as we only have data for the first seven months of the year.
- The least sales recorded for any year was in 1994 (approx. 25), and the highest sales was in 1980 (over 260 - albeit an outlier).

- After the year 1981, not a single year has recorded sales greater than 200.

—> Pivot table of month-wise and year-wise sales:

Table: 1.2.2

Year Month	April	Aug	Dec	Feb	Jan	July	June	March	May	Nov	Oct	Sept
1980	99	129	267	118	112	118	168	129	116	150	147	205
1981	97	214	226	129	126	222	127	124	102	154	141	118
1982	97	117	169	77	89	117	121	82	127	134	112	106
1983	85	124	164	108	75	109	108	115	101	135	95	105
1984	87	142	159	85	88	87	87	112	91	139	108	95
1985	93	103	129	82	61	87	75	124	108	123	108	90
1986	71	118	141	65	57	110	67	67	76	107	85	99
1987	86	73	157	65	58	87	74	70	93	96	100	101
1988	66	77	135	115	63	79	83	70	67	100	116	102
1989	74	74	137	60	71	86	91	89	73	109	87	87
1990	77	70	132	69	43	78	76	73	69	110	65	83
1991	65	55	106	55	54	96	65	66	60	74	63	71
1992	53	52	91	47	34	67	55	56	53	58	51	46
1993	45	54	77	40	33	57	55	46	41	48	52	46
1994	48	44	84	35	30	45	45	42	44	63	51	46
1995	52	NaN	NaN	39	30	62	40	45	28	NaN	NaN	NaN

Inferences: The pivot-table reveals the following:

- The year with the highest total sales is 1981 (1,780) selling approx. 148.33 units of wine every month

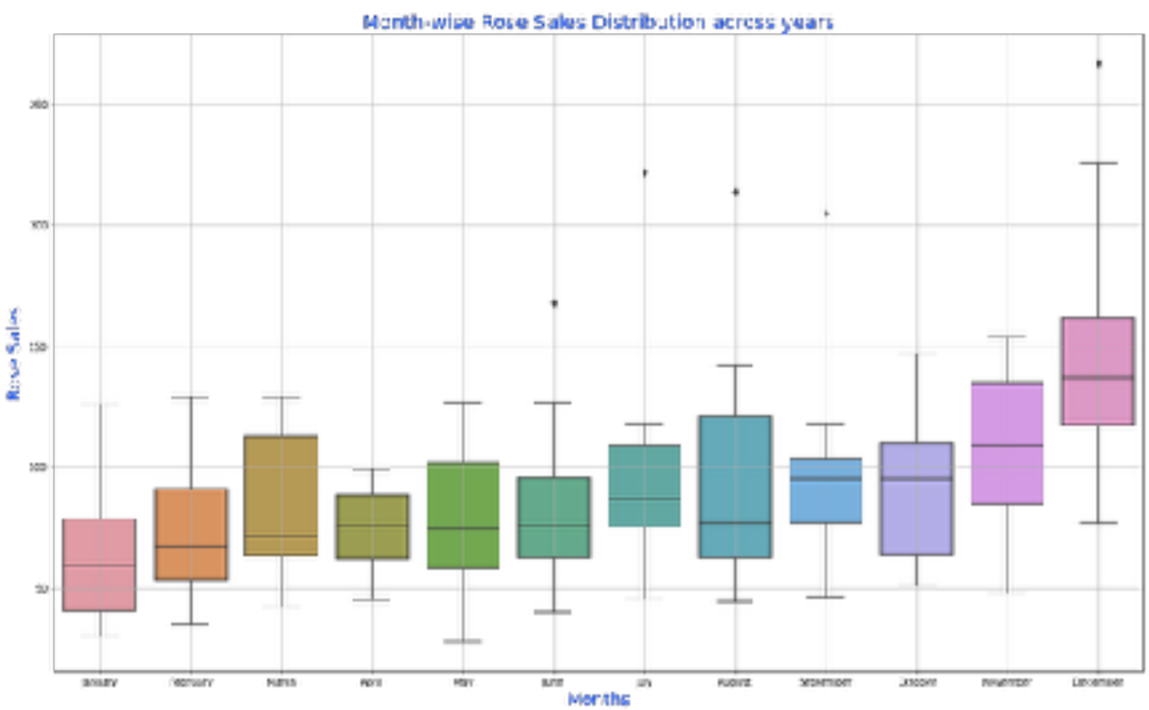
- The least total sales was recorded in 1994 (578), selling approx. 48.14 units every month. (we have discounted 1995, since we only have data for the first seven months of the year).

- The month with highest sum of sales across the years is December, selling over 2,174 units of wine, averaging 144.93 per year.

- The month with least total sales across years is January (1,024), averaging about 64 units per year.

—> <u>Monthly sales distribution across years:</u>

<u>Figure: 1.2.3</u>

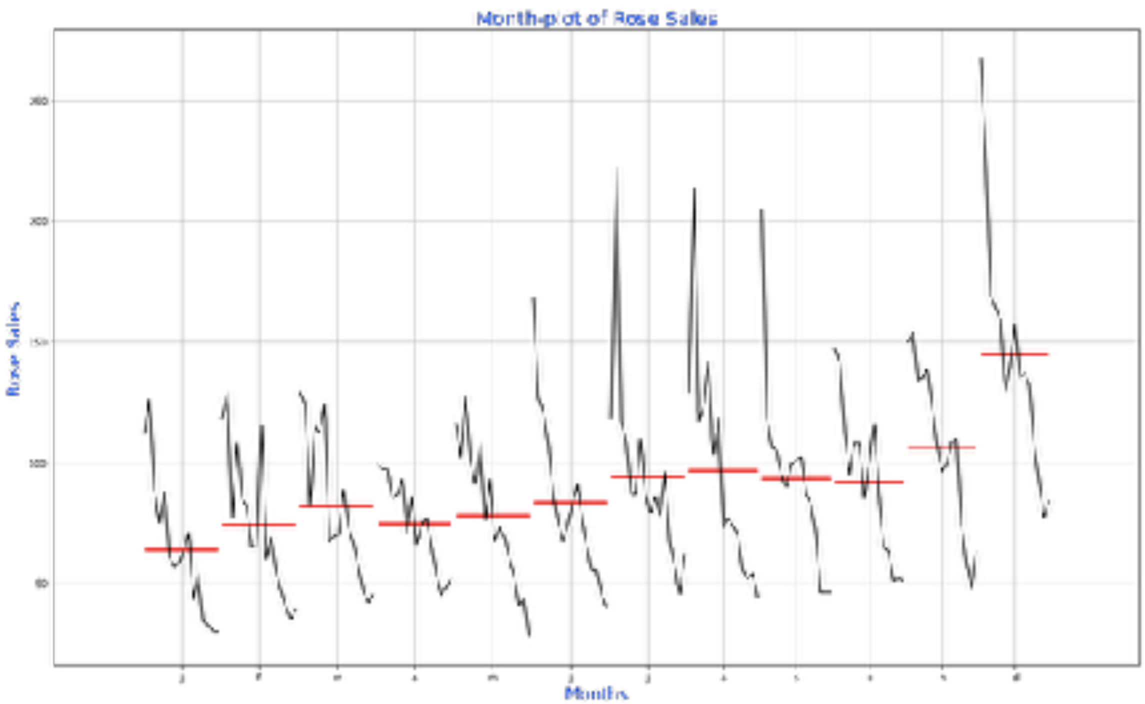


Inferences: This box-plot corroborates the observations of the pivot-table above, viz:

- December and January are the months with the highest and lowest distribution of Rose wine sales respectively.

—> Yearly sales distribution across months:

Figure: 1.2.4

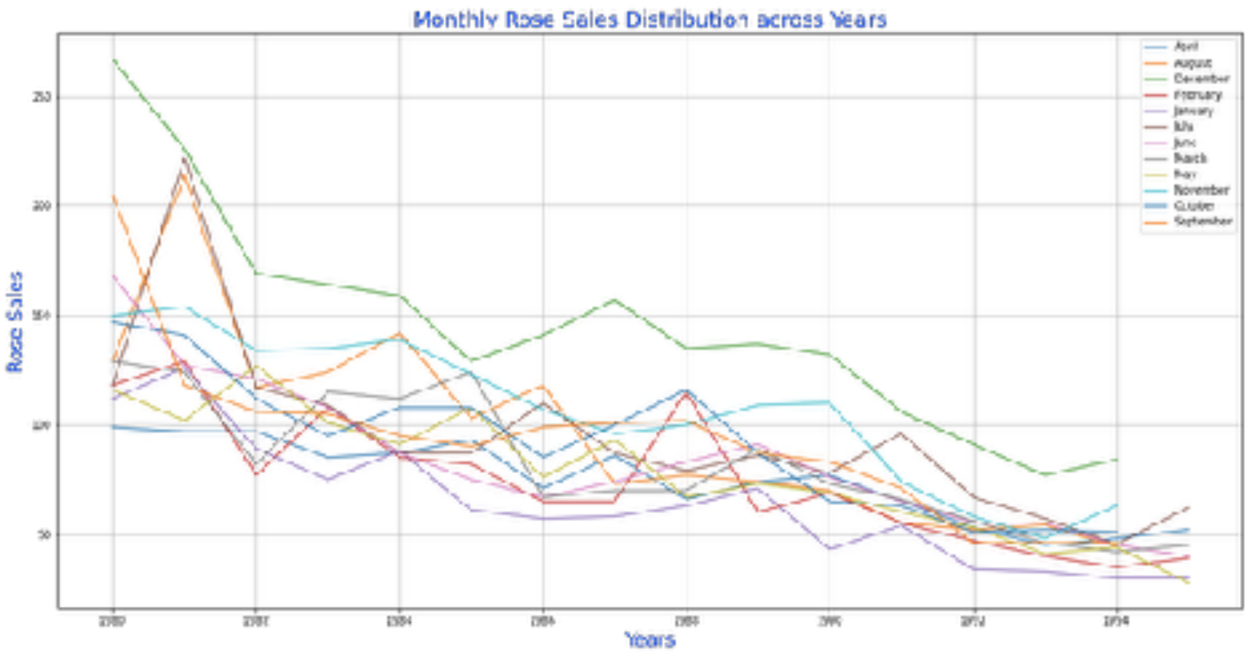


Inferences: This month-plot shows the yearly distribution of wine sales for each month. It reinforces our inferences from the pivot-table, as given below:

- December is the month that has recorded the highest sales across years.
- January has recorded the least mean sales across the years.
- 1995 is the year that gave the lowest sales figure.
- Across most months, there seems to be a drop in sales towards 1995.

—> <u>Monthly sales distribution across years:</u>

<u>Figure: 1.2.5</u>

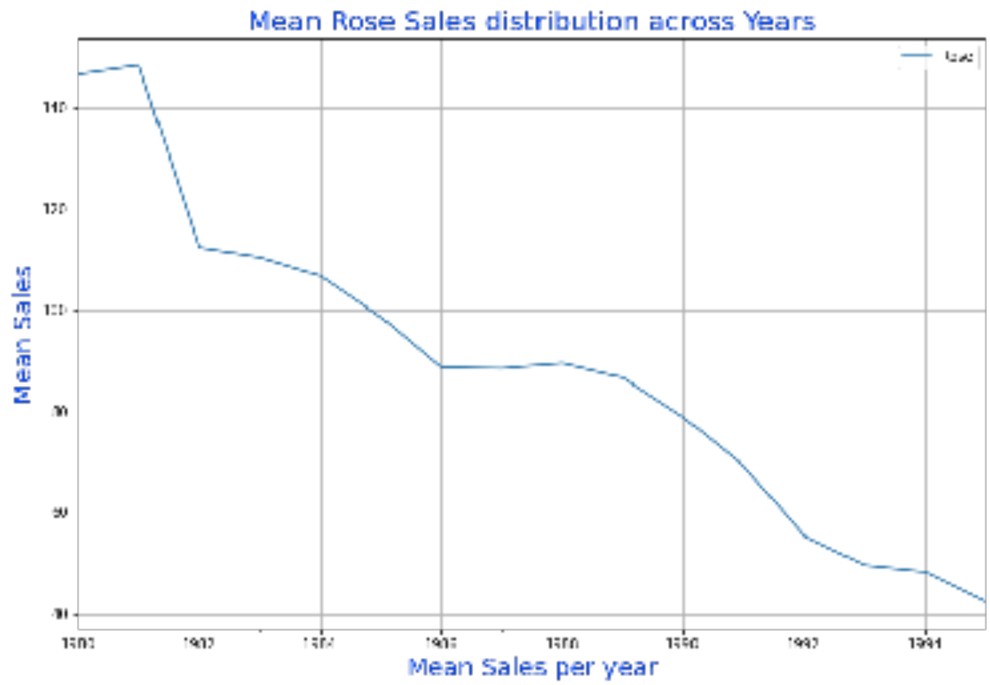


Inferences: the following data can be inferred from the above line-plot:

- The months of December, followed by November, have contributed to the highest sales figures of ‘Rose’ wine.
- The months with least sales seem to be January and April.
- Post 1988, there is an evident drop in sales across months.
- Between 1994 to1995 the sales seem to be steady at least for some of the months.

—> <u>Mean 'Rose' sales across years:</u>

<u>Figure: 1.2.6</u>

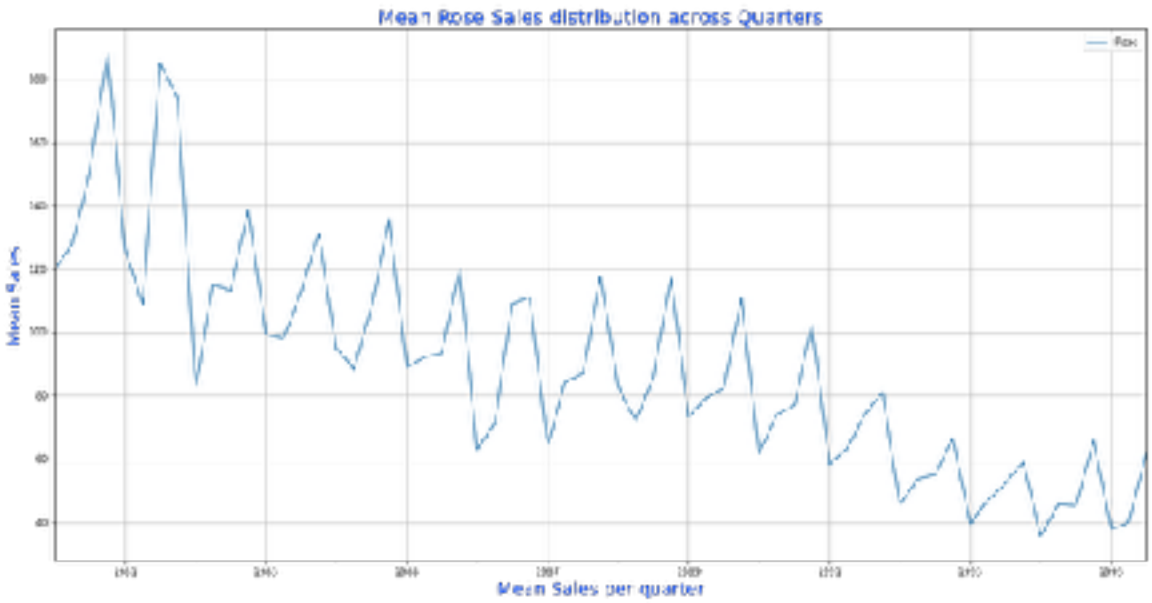


Inferences: This plot corroborates what has been inferred from the previous plots and the pivot-table:

- Barring a slight rise in mean sales 1981 and then in 1988, other years have shown consistent decreasing trend in terms of average sales per year.
- Although there is a rising and falling tendency, we can clearly see that average sales has been decreasing over the years.

—> Mean ‘Rose’ sales across quarters:

Figure: 1.2.7



Inferences:

- The quarterly sales reveal a certain seasonality in sales distribution. We can see that every year begins with low sales, then peaks to the maximum during the 4th quarter (November and December).
- The seasonality has a somewhat steady pattern, but is not similar across the years.
- The trend component, as we saw in the previous plot, shows a steady decline.

Outlier check:

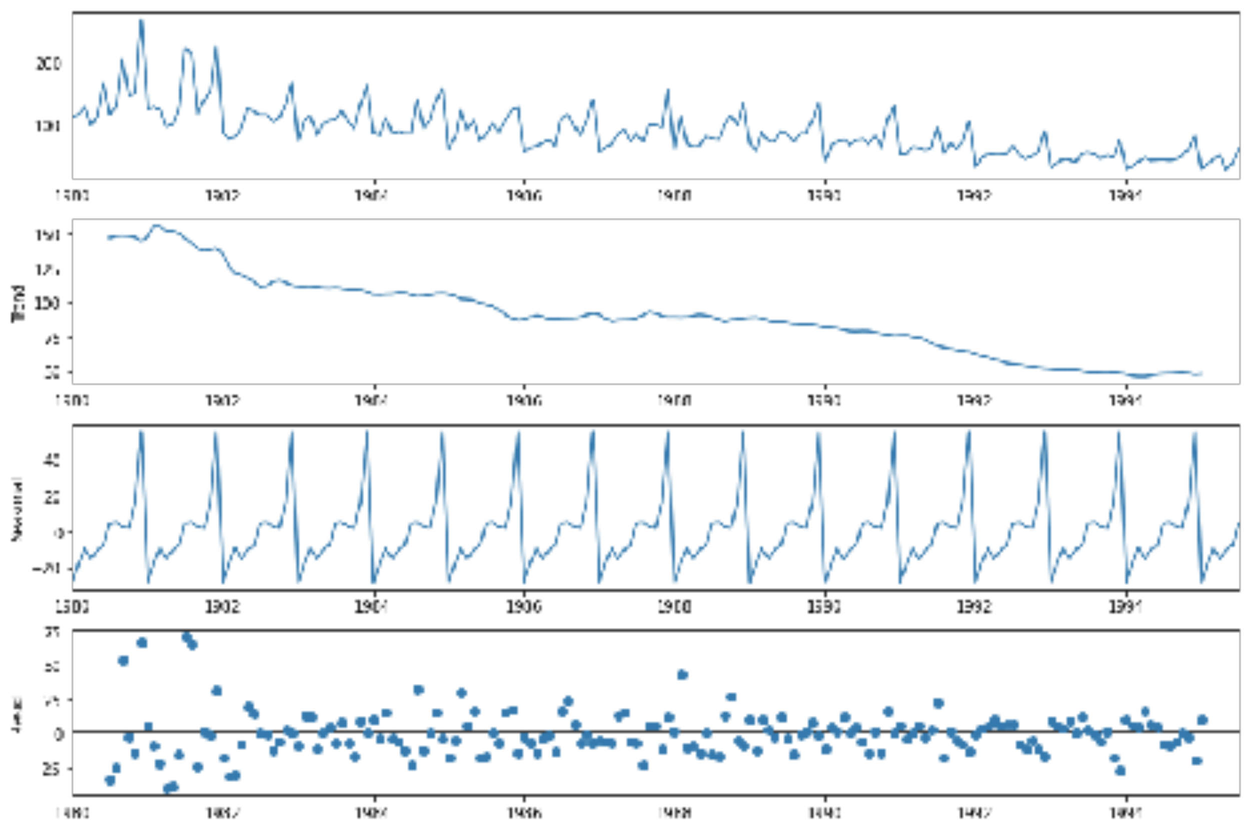
As seen in Figure: 1.2.2, there are a few outliers in the data, but they are not substantial or numerous enough to warrant an outlier treatment. Hence, outliers have not been treated here.

<u>Time Series decomposition</u>:

Decomposition aids to segregate the trend, seasonality and the residuals/ error elements in the data. Both additive and multiplicative decomposition was performed on the given data using the seasonal_decompose() function, as given below:

—> <u>Additive decomposition:</u>

<u>Figure: 1.2.8</u>



Inferences:

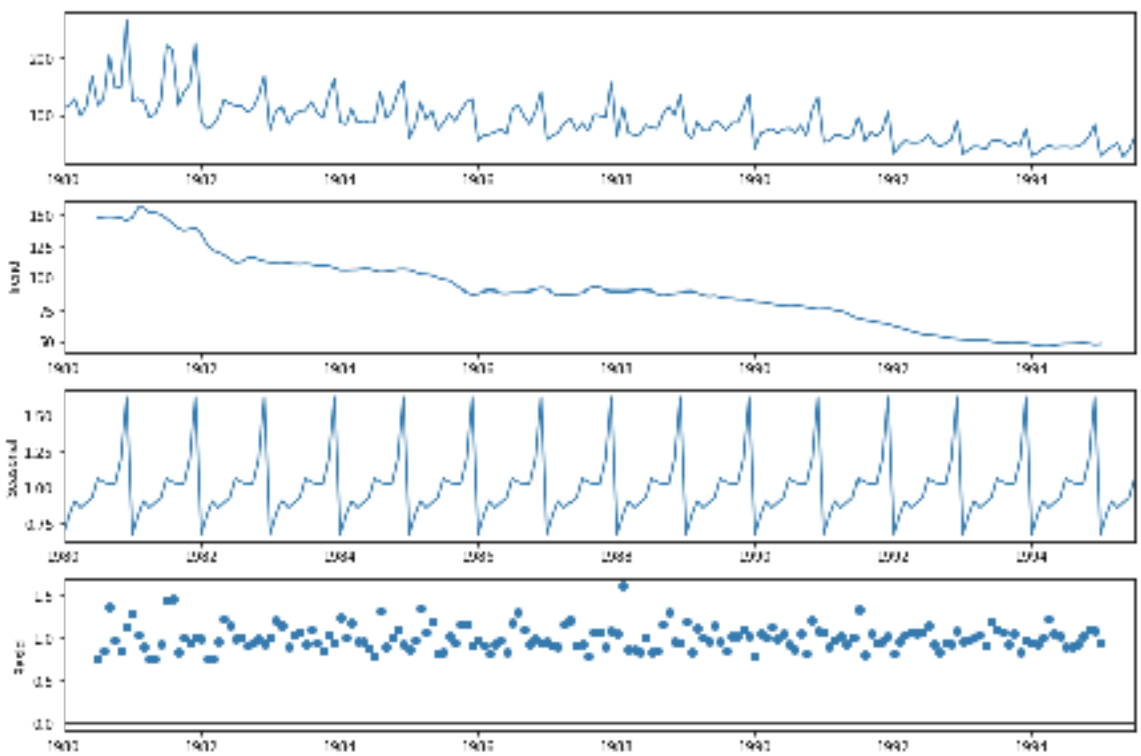
- We can see that there is a very clear declining pattern in the trend component. However, seasonality is also evident in the data.
- The scale of residuals varies from -25 to +75. Based on the plotting of the residuals, we can infer that error component is not randomly

distributed - it still carries some information/pattern in it, especially in the early years.

- Thus, we will not choose the additive decomposition to analyze this time series.

—> Multiplicative decomposition:

Figure: 1.2.9



Inferences:

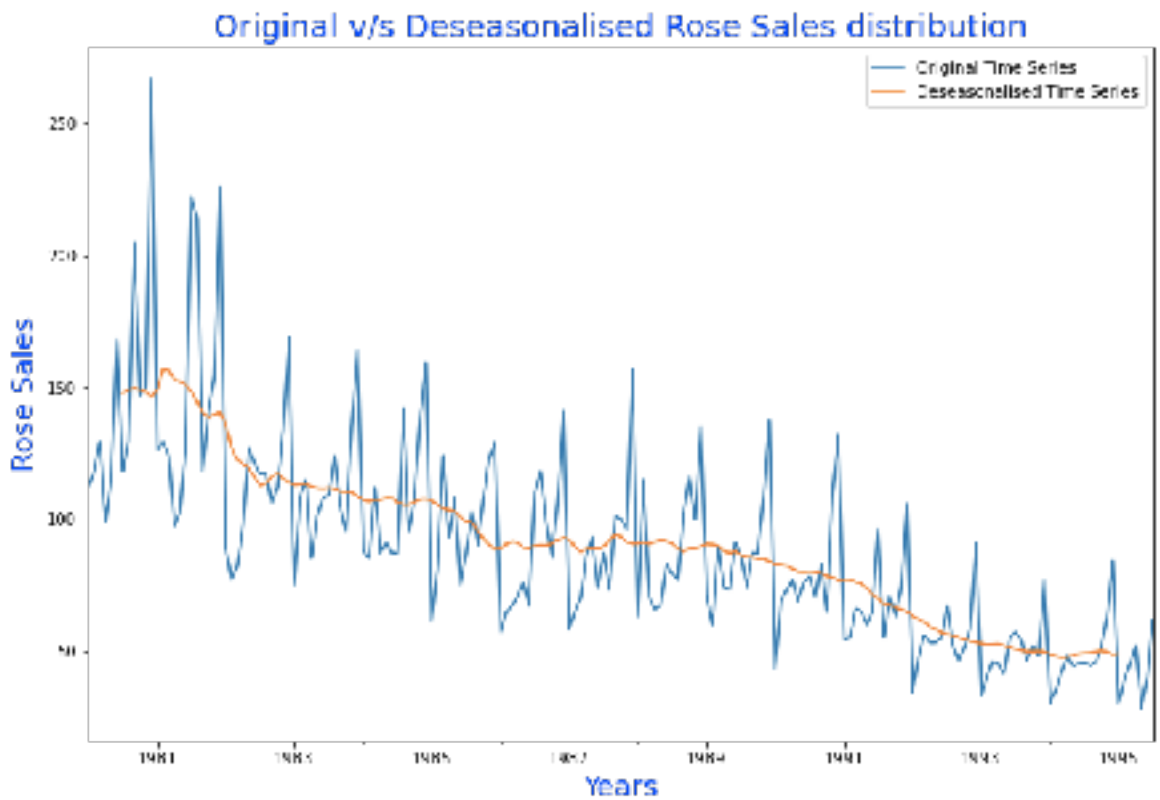
- Here, the residuals are centered around the point 1.0. Thus we can effectively say that error component is purely random, and does not conceal any information in it.

- Thus we will select Multiplicative decomposition to proceed with the time series analysis and forecasting.

—> <u>Plotting of de-seasonalised time series:</u>

To verify the choice of decomposition, a plotting of the deseasonalised data time series was done (with only trend and residual elements), as shown in graph below:

<u>Figure: 1.2.10</u>



Inferences:

- This plot reveals beyond doubt that once seasonality is removed, data is clearly represented in a wavy, declining line.
- Thus, we can affirm that seasonality is a major component of this data.
- A decreasing trend is also evident here.

1.3 Splitting data into train and test set:

The dataset was split using the following criteria:

Set:	Criteria:	Observations:
Train	Data uptil 1991	132
Test	Data from 1991 onwards	55

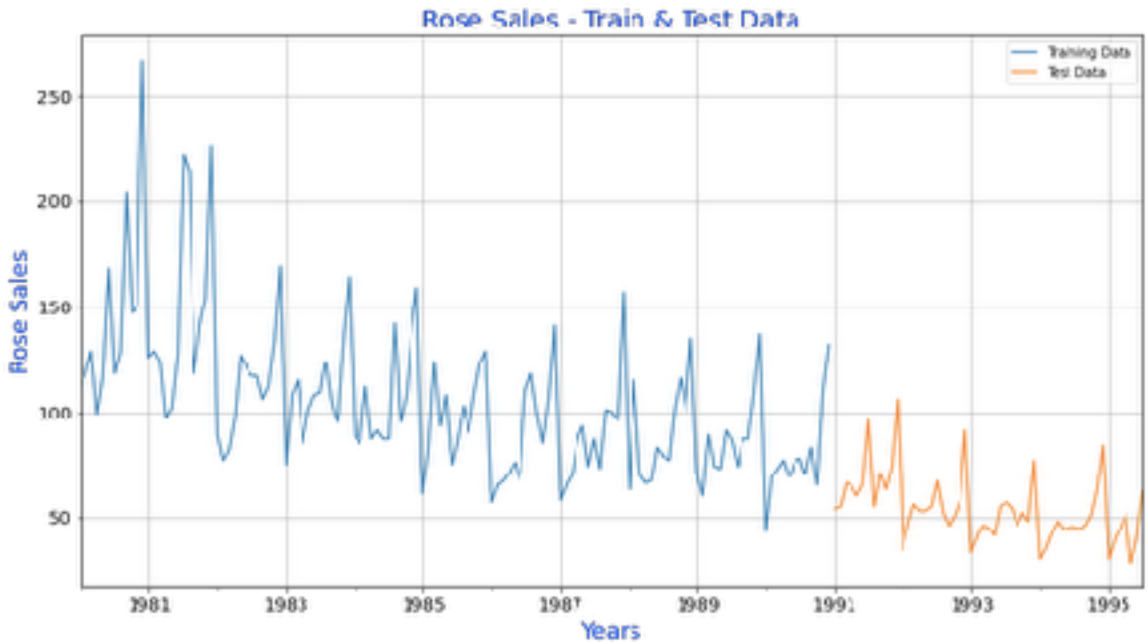
The first five rows of the training and test set are given below:

Table: 1.3

Training data		Test data	
YearMonth	Rose	YearMonth	Rose
1980-01-01	112	1991-01-01	54
1980-02-01	118	1991-02-01	55
1980-03-01	129	1991-03-01	66
1980-04-01	99	1991-04-01	65
1980-05-01	116	1991-05-01	60

Post splitting, the train and test data was plotted to cross-check the split, as given in figure overleaf:

Figure: 1.3


Rose Sales - Train & Test Data
Training Data
Test Data
Rose Sales
250
200
150
100
50
1981
1983
1985
1987
1989
1991
1993
1995
Years

1.4 Exponential Smoothing:

- Exponential smoothing is a forecasting technique which is an extension of the weighted moving average method (wherein higher weights are given to more recent observations).

- There are various exponential smoothing methods, some of which are:

    1. *Single exponential smoothing (SES)* - where trend and seasonality are absent
    2. *Double exponential smoothing (DES)* - where only trend is present
    3. *Triple exponential smoothing (TES)* - where both trend and seasonality are present

- In this case, since there is the presence of very strong trend and seasonality, we will utilize the TES method first.

—> Model evaluation metric:

- Throughout this case, the evaluation criteria used is the RMSE (root mean squared error). All models have been compared for effectiveness using their RMSEs.

- RMSE is a robust model evaluation metric. It is the square root of mean of squares of all error terms. It is derived as follows:

$$\text{rmse} = \sqrt{\frac{\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2}{n}}$$

where: n = no. of observations

t = time period

Y_t = actual value of observation

$Yhat_t$ = forecasted value of observation

1.4.1 Triple Exponential Smoothing (TES) using Holt's Winter method:

Using the ExponentialSmoothing() function from statsmodels.tsa.api, the TES model was built for the time series.

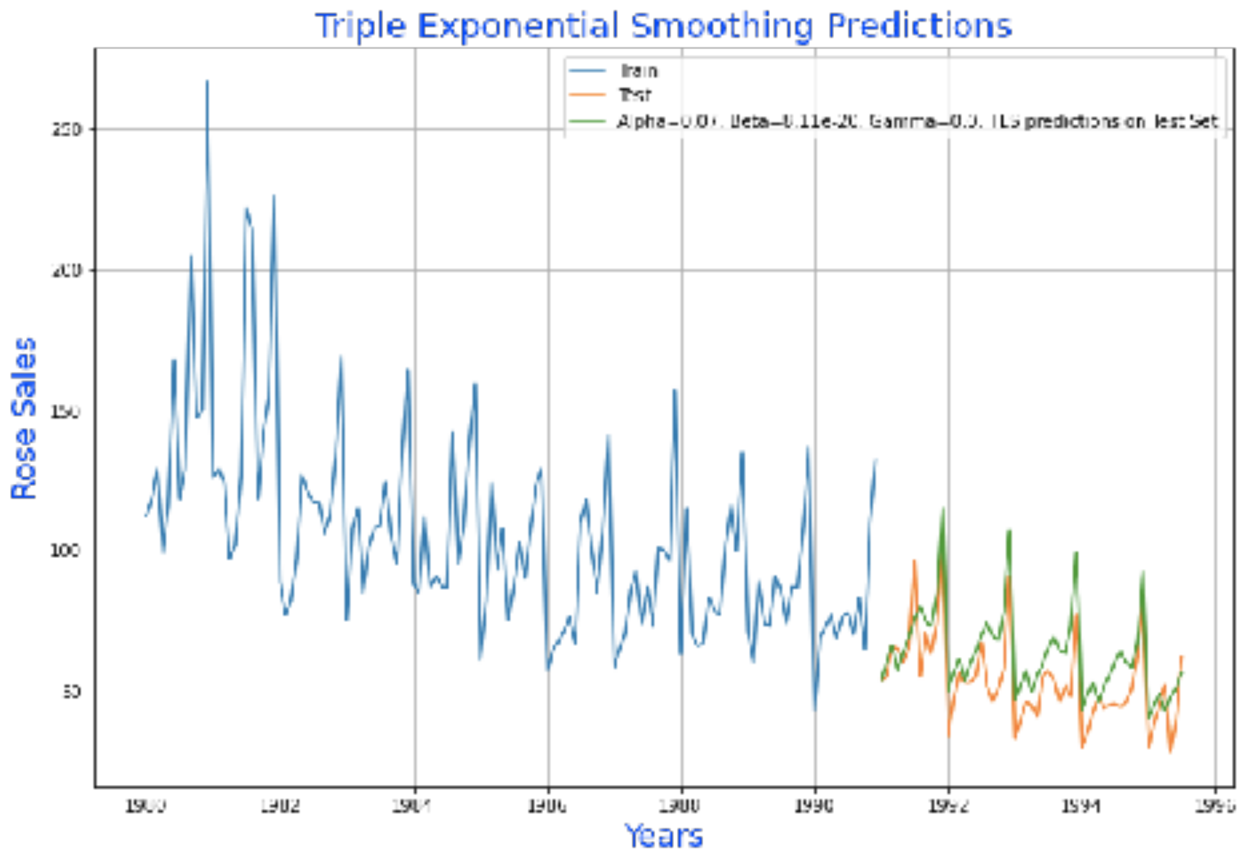
The best parameters for the automated TES model are given below:

Parameter:	Description:	Value:
alpha	smoothing level	0.0699
beta	smoothing slope	8.106E-20
gamma	smoothing seasonal	0.0

Test predictions:

The TES model was built to predict sales on the test data, which is plotted in the graph overleaf:

Figure: 1.4.1



Observations:

The forecast on test data (green line) seems to be somewhat similar, it tries to roughly model the test data (orange line).

Model evaluation:

RMSE for TES was computed, and shared in a DataFrame, to be used for later comparison with RMSEs of other models, as given below:

Model:	Test RMSE:
TES: Alpha=0.07, Beta=8.11E-20, Gamma=0.0	12.8311

1.4.2 Double Exponential Smoothing (DES) using Holt's method:

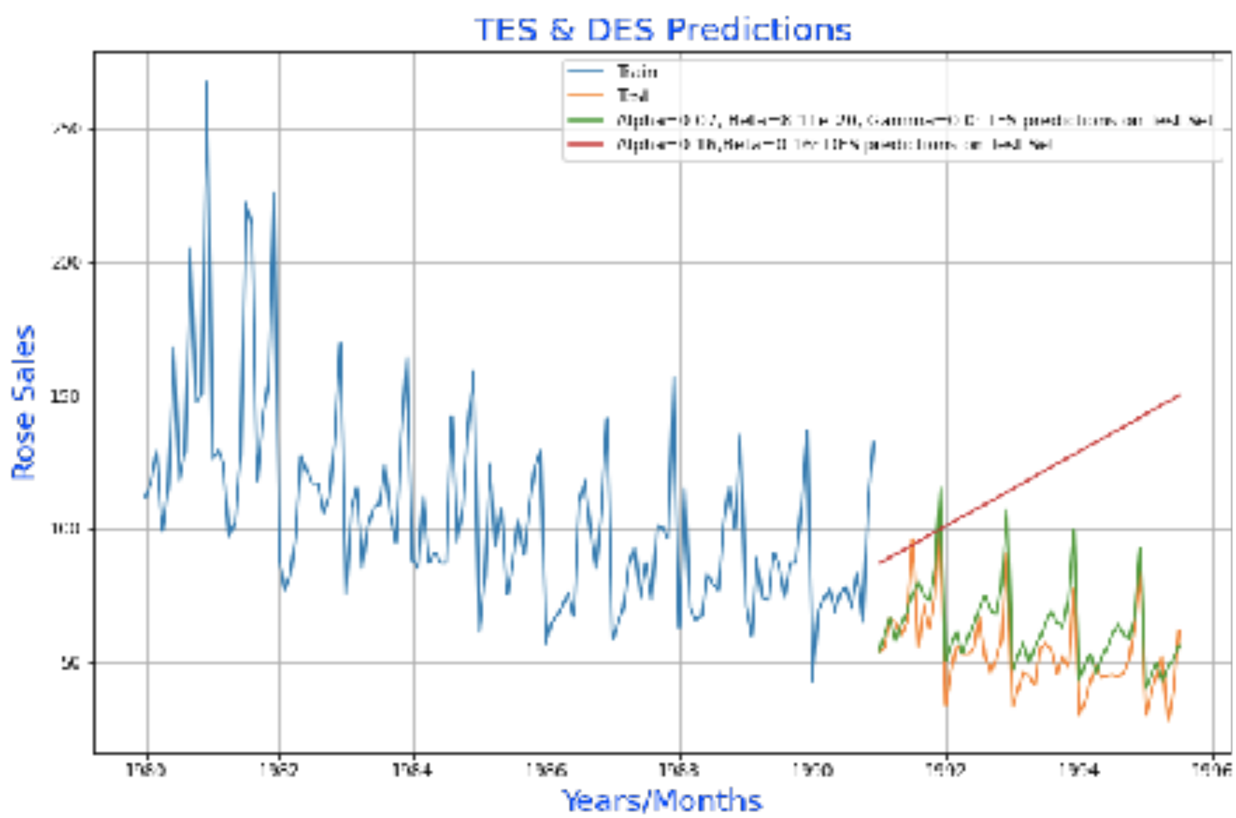
Despite the presence of seasonality in the data, the DES model was built using the Holt() function from statsmodels.tsa.api (this exercise was done for sheer comparison sake), which gave the following best parameters:

Parameter:	Description:	Value:
alpha	smoothing level	0.1579
beta	smoothing slope	0.1579

Test predictions:

The DES model test predictions were plotted in a comparison plot along with the test predictions of previous models, as given below:

Figure: 1.4.2



Observations:

The DES forecasts on the test data is a straight-line prediction (as seen in red line). This is very unrealistic prediction, showing an increasing sales trend, proving that the DES model is not apt in this case.

Model evaluation:

RMSE for DES is as given below:

Model:	Test RMSE:
DES: Alpha=0.16,Beta=0.16	70.6046

1.4.3 Linear Regression model:

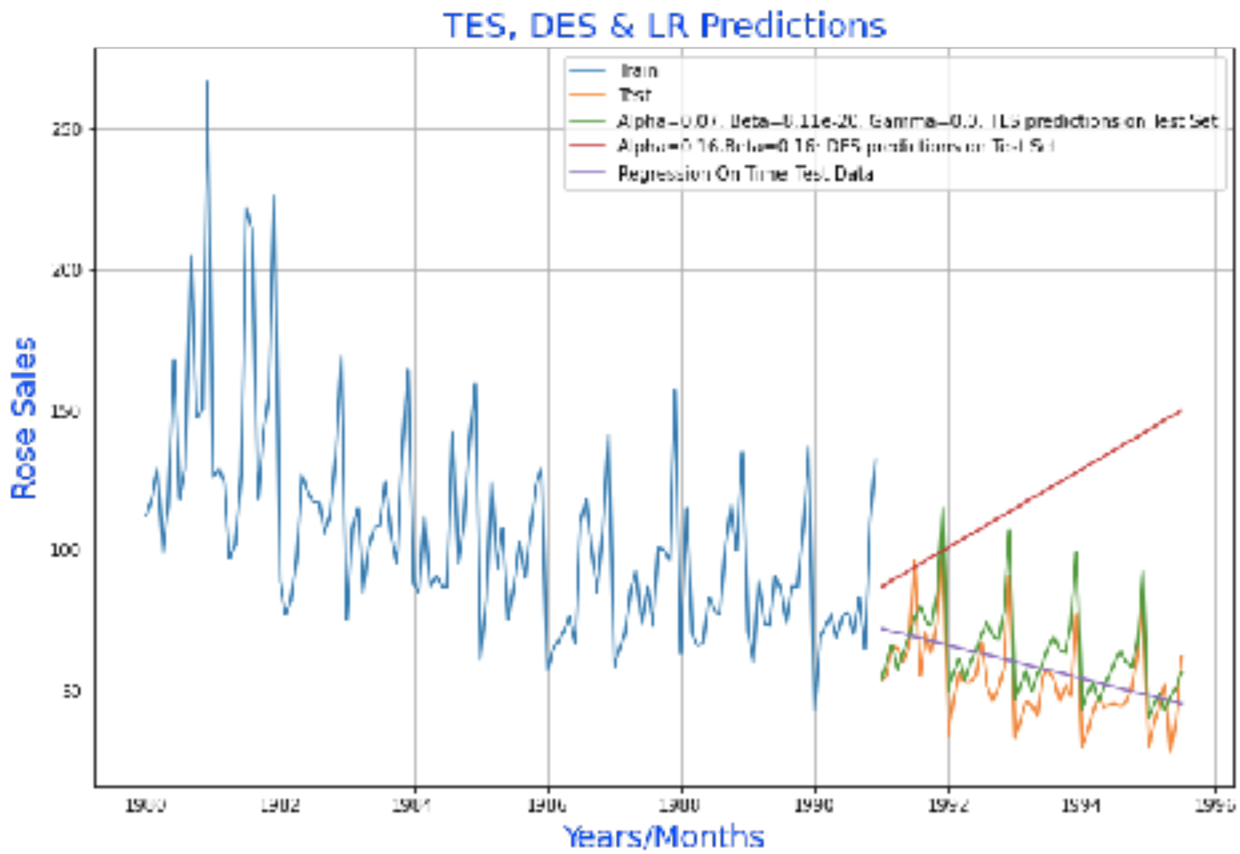
- The Linear regression (LR) model was used for forecasting test data using the *LinearRegression()* function from *sklearn.linear_model*.

- For running the LR model, a time instance (cardinal numbering indexing) had to be given to the train and test set, since the LR model cannot run with the DateTimeIndex.

Test predictions:

The LR model test predictions were plotted in a comparison plot along with the test predictions of previous models, as given overleaf:

Figure: 1.4.3



Observations:

The LR forecasts on the test data is a straight-line prediction (as seen in purple line). Although it captures the downward trend of data, it has failed to capture the seasonality, proving that the LR model is not apt in this case.

Model evaluation:

RMSE for LR model is as given below:

Model:	Test RMSE:
Linear regression	15.2784

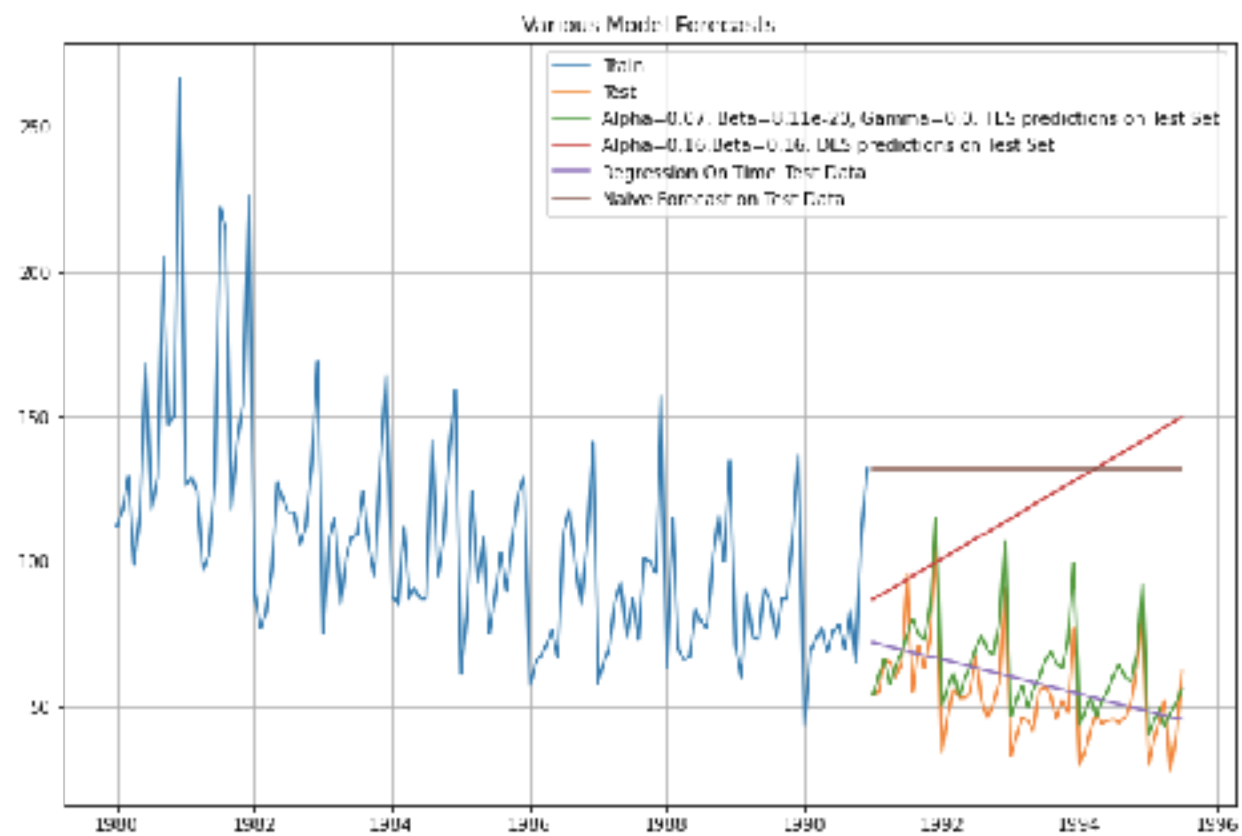
1.4.4 Naive approach model:

The Naive approach was used to make predictions on the test data, which essentially uses the previous/final observed value as predictions.

Test predictions:

The Naive approach test predictions are plotted against other model predictions in the plot below:

Figure: 1.4.4



Observations:

The Naive approach forecasts on the test data is again a straight-line prediction (as shown by brown line). This is a highly unrealistic prediction, as it captures neither the trend nor the seasonality elements.

Model evaluation:

RMSE for Naive model is given below:

Model:	Test RMSE:
Naive approach	79.7457

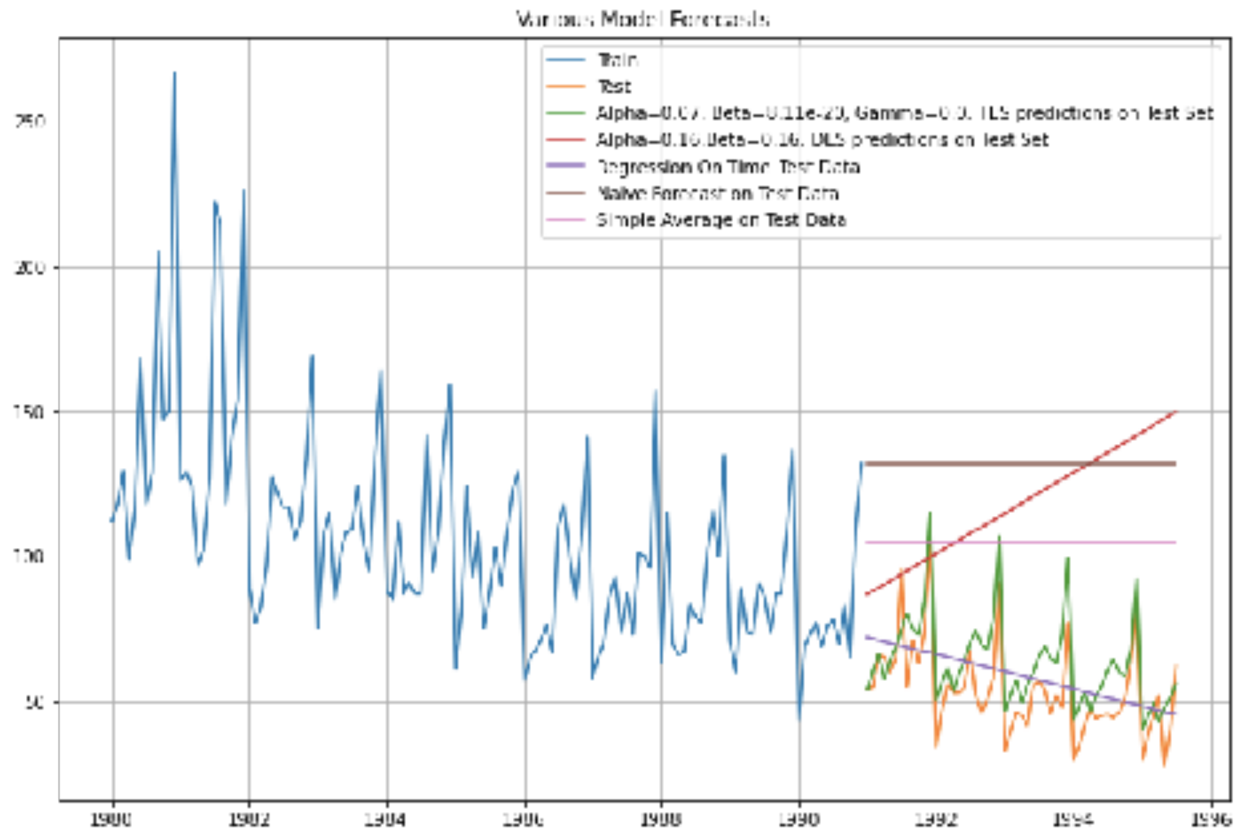
1.4.5 Simple Average method:

The Simple average (SA) method was used to make predictions on the test data, which essentially uses the mean of the train data to make predictions.

Test predictions:

The SA test predictions are plotted against other model predictions in the plot overleaf:

Figure: 1.4.5



Observations:

The SA predictions on the test data is again an absolute straight-line prediction (as shown by pink line), that does not even capture the trend or seasonality elements.

Model evaluation:

RMSE for SA model is given below:

Model:	Test RMSE:
Simple Average	53.4882

1.4.6 Moving Average method:

The Moving average (MA) method uses the mean of a specified number of sliding window / trailing observations to make predictions on the test data.

Here, we have taken four different trailing windows for mean computation - 2, 4, 6 and 9, in order to test out their effectiveness.

Test predictions:

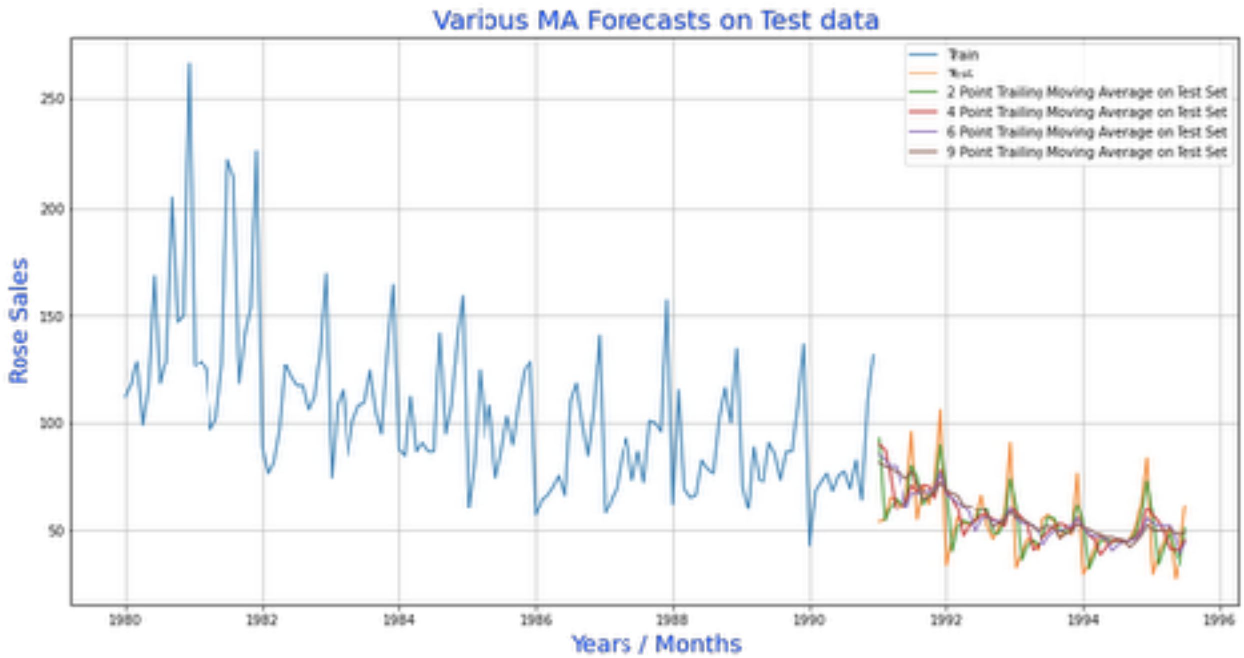
The below table shows the first ten MA predictions for the various trailing windows:

Table: 1.4.6

YearMonth	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1980-01-01	112.00	NaN	NaN	NaN	NaN
1980-02-01	118.00	115.00	NaN	NaN	NaN
1980-03-01	129.00	123.50	NaN	NaN	NaN
1980-04-01	99.00	114.00	114.50	NaN	NaN
1980-05-01	116.00	107.50	115.50	NaN	NaN
1980-06-01	168.00	142.00	128.00	123.67	NaN
1980-07-01	118.00	143.00	125.25	124.67	NaN
1980-08-01	129.00	123.50	132.75	126.50	NaN
1980-09-01	205.00	167.00	155.00	139.17	132.67
1980-10-01	147.00	176.00	149.75	147.17	136.56

The MA test predictions are plotted against other model predictions in the plot overleaf:

Figure: 1.4.6



Observations:

Out of the MA predictions on the test data, the 2-point trailing average (demonstrated by green line) seems to give a prediction closest to the test data as compared to the other three trailing windows.

Model evaluation:

RMSE for the four MA models are given below:

Model:	Test RMSE:
2-point MA	11.5301
4-point MA	14.4584
6-point MA	14.5730
9-point MA	14.7329

1.5 Checking stationarity of the data:

A stationary time series is one where the mean and the variance of the series is constant over a period of time, and the correlation between two observations depends only on the distance/lag between them.

Stationarity check on train data:

- Checking the stationarity of the train data is an essential step before proceeding to build more complex forecast models like ARIMA and SARIMA.
- Stationarity need not be checked for the test data, since models will only be trained on the train data. Therefore, test data features will not affect the predictions.
- To check for stationarity, the Augmented Dickey-Fuller (ADF) test was used on the train set, using the adfuller() function from statsmodels.tsa.stattools.
- With the derived ADF test statistics, hypothesis testing was done to come to a conclusion. The hypothesis were:

  H_0 = Series is non-stationary

  H_1 = Series is stationary

- The ADF test gave the following results:

```
Test statistic is -2.242
Test p-value is 0.4663
Number of lags used 13
```

Conclusion:

At 95% confidence interval: **p-value (0.47) > alpha (0.05)**

Therefore, the time series is **non-stationary**.

Differencing the series:

Level-1 differencing was done on train data to make the non-stationary time series stationary, using the diff() function.

Post level-1 differencing, the following test stats were derived:

```
Test statistic is -8.161
Test p-value is 3.038E-11
Number of lags used 12
```

Conclusion:

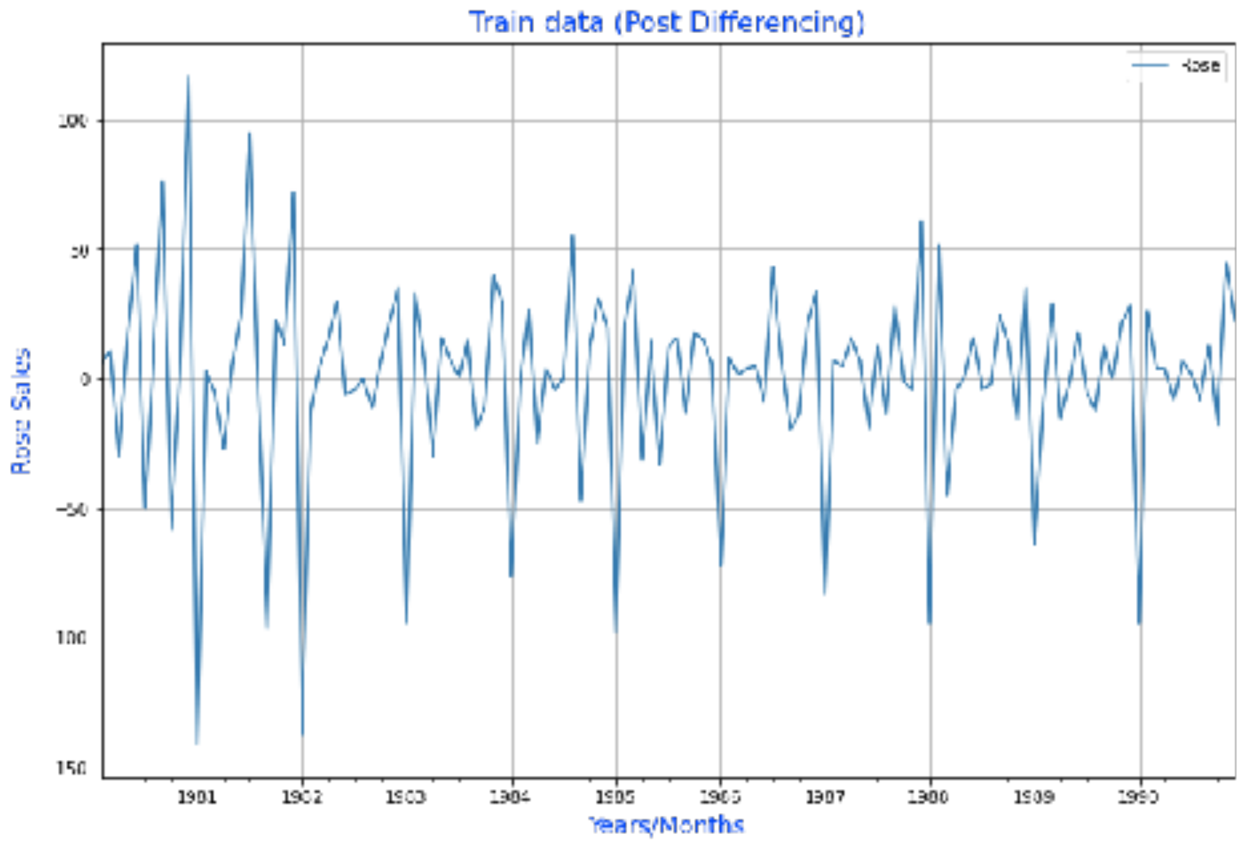
At 95% confidence interval: **p-value (3.04E-11) < alpha (0.05)**

Therefore, the time series is now **stationary.**

Plotting the differenced time series:

The time series train data, post differencing, which is stationary in nature, is as shown overleaf:

Figure: 1.5



1.6 Building automated ARIMA/SARIMA model:

- Auto-regressive Integrated Moving Average (ARIMA) models are powerful models that can make future forecasts on time series data.
- However, in this case, wherein the series demonstrates a very strong seasonality, a simple ARIMA will not suffice.
- Here we will employ the SARIMA (Seasonal Auto-regressive Integrated Moving Average) model, so that the seasonal adjustments are fully accounted for in the future predictions.
- The SARIMA can be defined as a function of the ARIMA, being represented as **SARIMA(p,d,q)(P,D,Q)[m]**, which is equal to:

$$\boxed{\text{ARIMA (p,d,q)} \;*\; \text{ARIMA (P,D,Q)[m]}}$$

where: p = no. of auto-regressive components

d = no. of differencing done

q = no. of moving average components

P = seasonal auto-regressive component

D = seasonal differencing

Q = seasonal moving average components

m = frequency of observations

Parameter generation:

- Using the 'itertools' function, random parameter combinations for the SARIMA model were automatically generated.

- With the sm.tsa.statespace.SARIMAX() function from statsmodels.api, various combinations of the random parameters were tested to a maximum iteration of 1000, to find the best parameters to build the automated model.

Scoring and selecting the best SARIMA model:

- The *Akaike Information Criterion (AIC)* metric was used to select the best parameter. The parameters with least AIC values was chosen to build the model with.

- Post running the iteration with random parameters, a data frame was generated with the five models with least-AIC parameters, as shown below:

Table: 1.6

Parameter	Seasonal parameter	AIC
(3, 1, 1)	(3, 0, 2, 12)	774.4003
(3, 1, 2)	(3, 0, 2, 12)	774.8809
(3, 1, 1)	(3, 0, 0, 12)	775.4267
(3, 1, 1)	(3, 0, 1, 12)	775.4953
(3, 1, 3)	(3, 0, 0, 12)	775.5610

- Based on this information, SARIMA(3,1,1)(3,0,2,12) was chosen as the best parameter for automated SARIMA generation.

Generating the auto-SARIMA model:

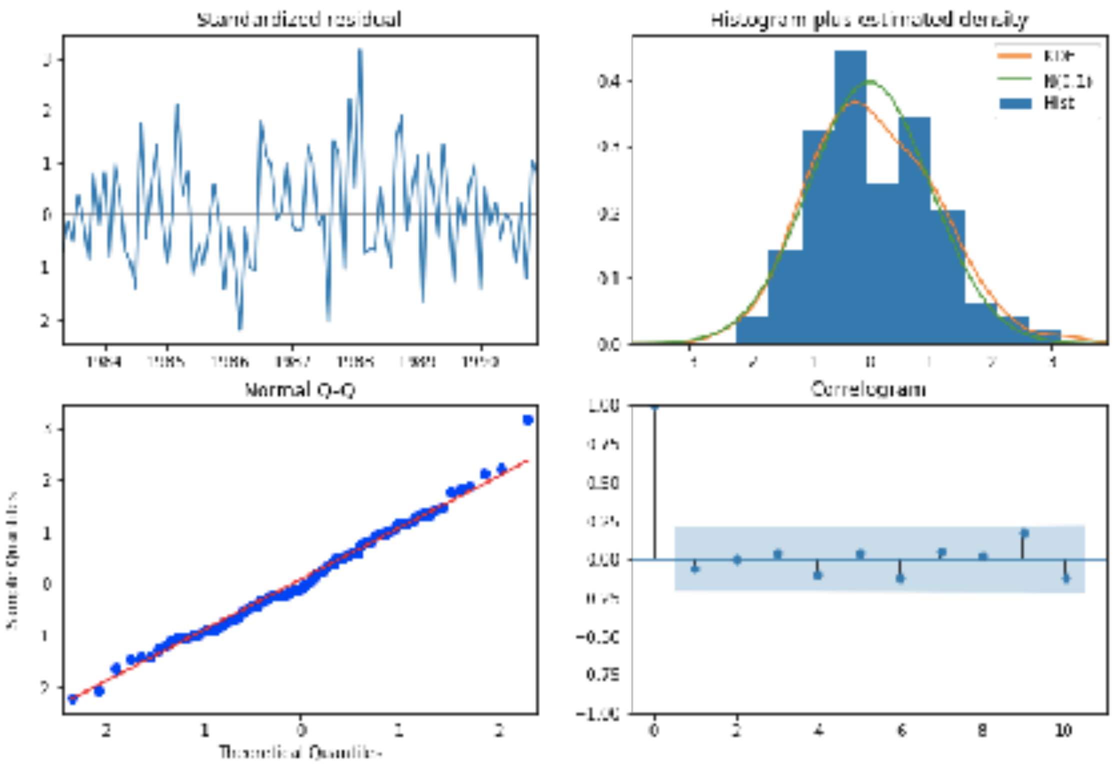
The auto-SARIMA model was built and the following model diagnostics was generated.

Figure: 1.6.1

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 Rose   No. Observations:                132
Model:             SARIMAX(3, 1, 1)x(3, 0, [1, 2], 12)   Log Likelihood                -377.200
Date:                              Sat, 09 Oct 2021   AIC                            774.400
Time:                                      09:28:34   BIC                            799.618
Sample:                                  01-01-1980   HQIC                           784.578
                                       - 12-01-1990
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0464      0.127      0.366      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    -13.902      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      -0.4967      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       192.1979     39.638      4.849      0.000     114.508     269.888
===================================================================================
Ljung-Box (Q):                       34.22   Jarque-Bera (JB):                 1.64
Prob(Q):                              0.73   Prob(JB):                         0.44
Heteroskedasticity (H):               1.11   Skew:                             0.33
Prob(H) (two-sided):                  0.78   Kurtosis:                         3.03
===================================================================================
```

A diagnostic plot was drawn to check the nature of the residuals in the model, as shown overleaf:

Figure: 1.6.2



Evaluating the auto-SARIMA model:

For the model evaluation, two metrics were used:

- *RMSE - root mean square error* (square root of mean of squares of all error terms)
- *MAPE - mean absolute percentage error* (mean of absolute difference between error terms, represented as a percentage)

The evaluation results for the auto-SARIMA is given below:

	Test RMSE	MAPE
SARIMA(3,1,1)(3,0,2,12)	18.912	36.46

1.7 <u>Building manual SARIMA model:</u>

In order to build manual SARIMA models, the parameters p,d,q and seasonal parameters P,D,Q will have to be manually selected. For this purpose, the following were plotted:

- ***ACF (auto-correlation function)*** - measures correlation of current observations with past observations
- ***PACF (partial auto-correlation function)*** - measures correlation between current and k-lagged series by removing intermediate observations.

ACF and PACF plots:

<u>Figure: 1.7.1</u>

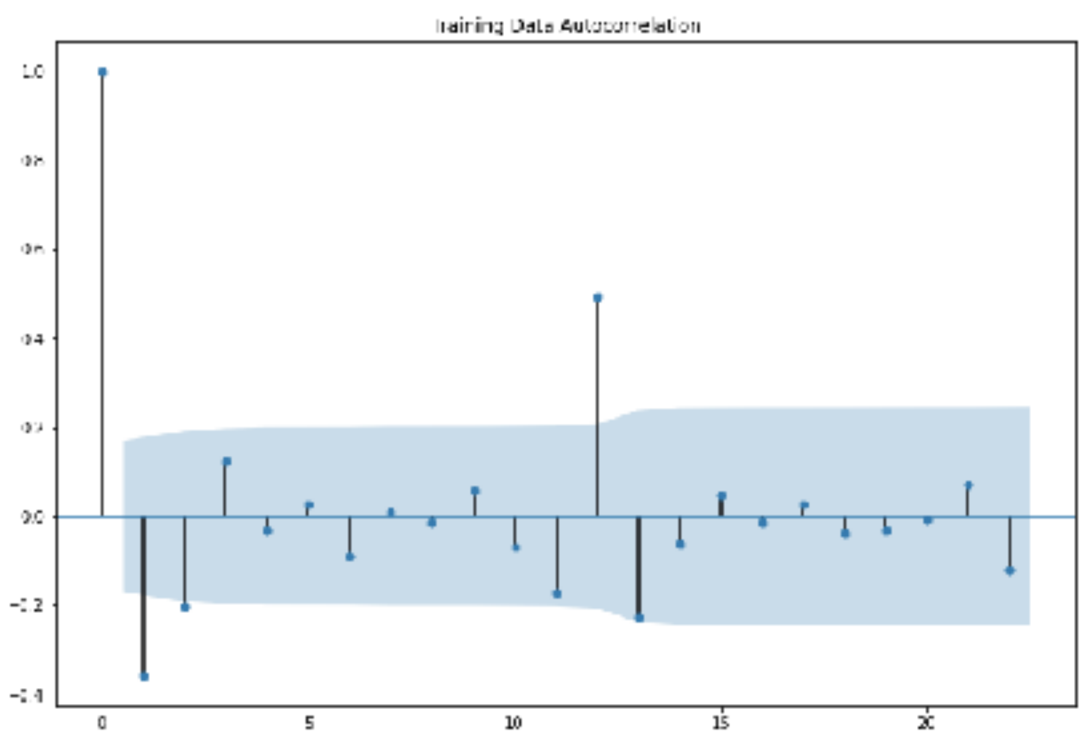
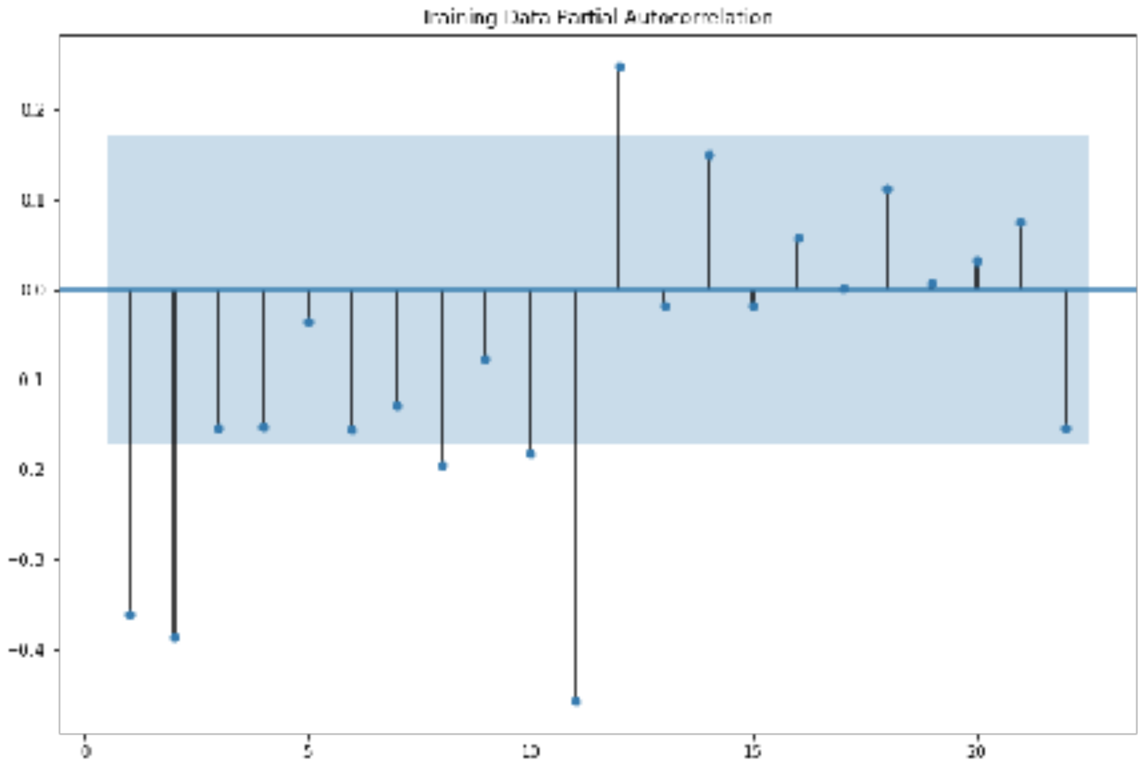


Figure: 1.7.2



As per the ACF and PACF plots, we can derive the following values:

Parameters	p=2, d=1, q=2
Seasonal parameters	P=1, D=1, Q=1, m=12

The manual SARIMA model was built and test data forecasted using the above selected parameters.

The evaluation results for the manual-SARIMA is given below:

	Test RMSE	MAPE
SARIMA(2,1,2)(1,1,1,12)	13.279	17.134

Additional SARIMA models with random parameters:

Four additional models were built using the following:

Parameters	Seasonal parameters
(1,1,1)	(0,0,0,12)
(3,1,0)	(1,0,1,12)
(2,1,1)	(1,1,0,12)
(3,1,1)	(1,0,1,6)

1.8 Model evaluation:

The RMSE values were compared to decide on the best model to use for making the future sales predictions. The least RMSE suggests the most effective model. Tabular comparison of model-wise RMSE values is given below:

Table: 1.8

Model	Test RMSE
TES: Alpha=0.07, Beta=8.11e-20, Gamma=0.0	12.8311
DES: Alpha=0.16,Beta=0.16	70.6046
LR RMSE	15.2784
Naive RMSE	79.7457
SA RMSE	53.4882
2-point MA	11.5301
4-point MA	14.4584
6-point MA	14.5730
9-point MA	14.7329
SARIMA(3,1,1)(3,0,2,12)	18.9121
SARIMA(2,1,2)(1,1,1,12)	13.2787
SARIMA(1,1,1)(0,0,0,12)	37.4466
SARIMA(3,1,0)(1,0,1,12)	32.6546
SARIMA(2,1,1)(1,1,0,12)	16.4633
SARIMA(3,1,1)(1,0,1,6)	31.1900

Conclusion:

Based on the RMSE values, the best models are:

- 2-point Moving Average
- TES (Alpha=0.07, Beta=8.11e-20, Gamma=0.0)
- SARIMA (2,1,2) (1,1,1,12)

1.9 Forecasting future sales:

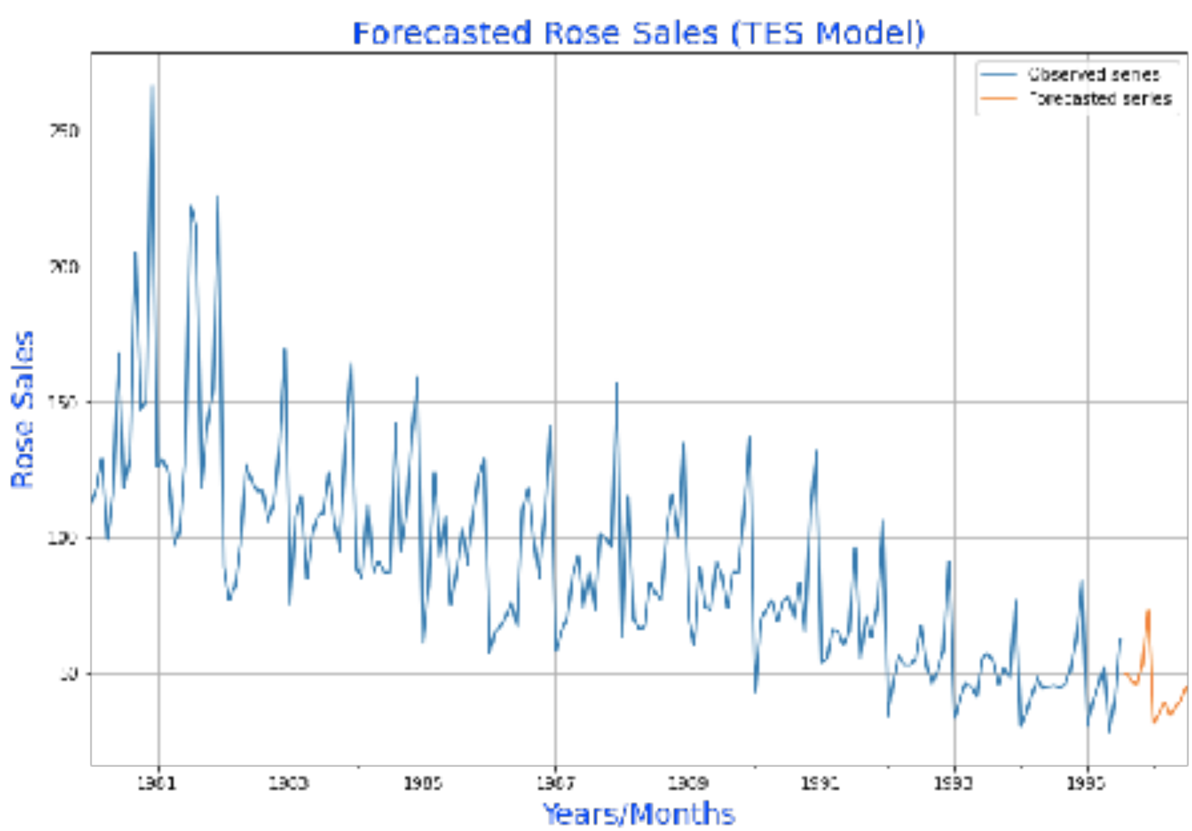
- Since the Moving Average method is good for predictions on the test data, it is not really ideal in making future forecasts. Hence we will only use the TES and SARIMA model for forecasting future sales.

- The time frame for which forecast is to be done is 12 months, i.e. from 01-08-1995 to 01-07-1996.

—> **Forecasts with TES**:

The TES predictions for the future sale of Rose wine for the next 12 months is as plotted below:

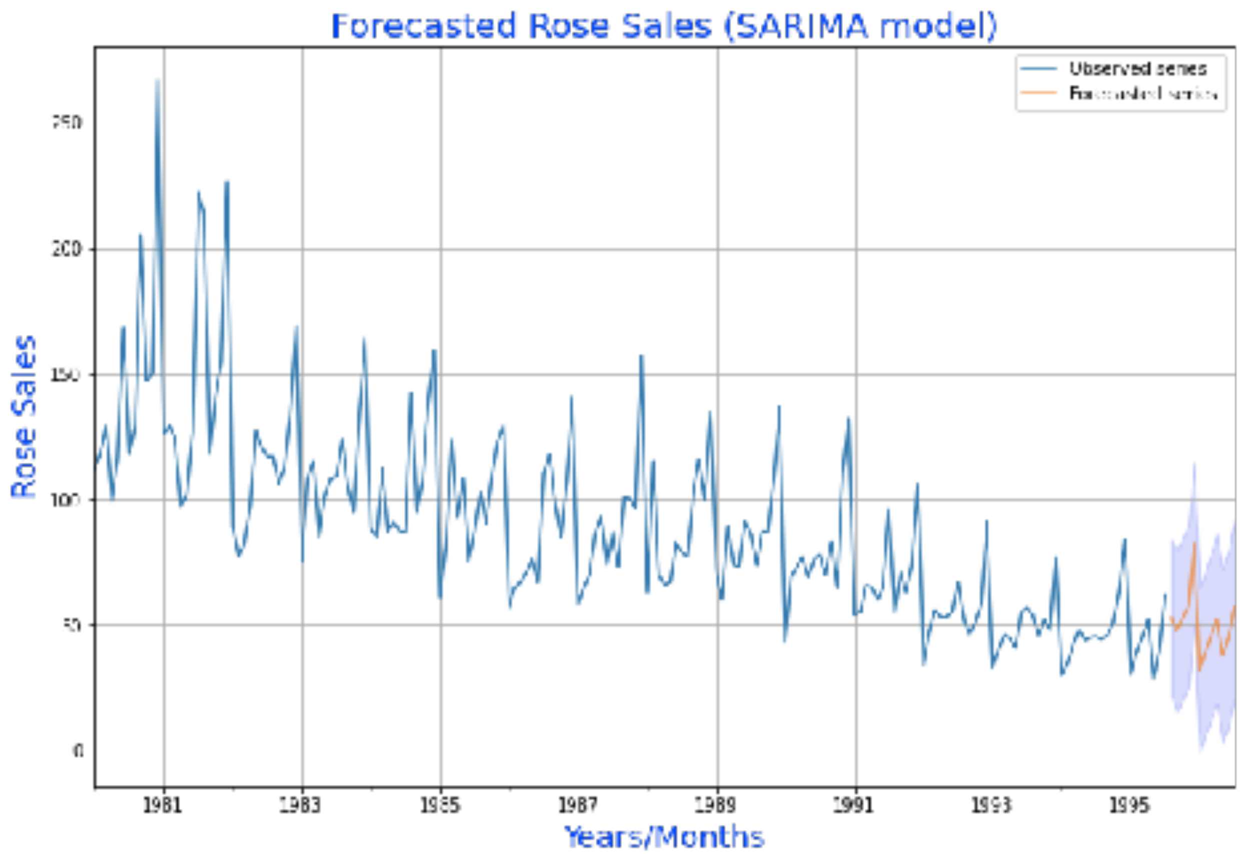
Figure: 1.9.1



—> **Forecasts with SARIMA (2,1,2)(1,1,1,12)**:

The SARIMA predictions for the future sale of Rose wine for the next 12 months (with confidence interval marked in light blue) is as plotted below:

Figure: 1.9.2



1.10 Comments and business insights:

The following can be observed from the two different forecasts made:

—> **Observations on TES forecasts**:

The 12-month future forecast exhibits a somewhat downward trend, while still maintaining the seasonality of the observed time series.

—> **Observations on SARIMA forecasts**:

The 12-month forecasted series demonstrates a somewhat steady trend, neither promising a wide increment or decline in sales of ‘Rose’ wine. It has, however, maintained the seasonal component of the observed data series.

—> Business insights:

- Data has shown a very steady decrease in the sale of ‘Rose’ wine. Thus, it is an alarm for ABC Estate wines. Also, the future forecasts do to seem to be very optimistic.
- The company needs to analyze the reasons behind this decline in sales, and remedy the situation accordingly.
- They can opt for a customer survey to gauge the exact reasons for the dwindling sales.

- A reassessment of the target audience could be carried out - to ensure that the right people are being targeted with the right kind of promotional activities.
- Newer and more effective strategies for advertising and promotion may be used.
- The pricing policy may need to be changed, making the product a bit more accessible in terms of price.
- A repackaging strategy could be worked out to make the 'Rose' wine more attractive and to entice customers to buy it.
- Data shows a steep seasonal spike during the final quarter of the year, perhaps due to festivities like Christmas and New Year, when people indulge in wine.
- The company can promote sales by offering many schemes and slash prices during the festive season.
- For the periods when sales are very slow, especially the first and second quarter, certain discounts and other offers can be given.

========================*The End*========================