Business Report

Time Series Forecasting Coded Project

PGPDSBA

Chithira Raj

Table of Contents

List of Tables

List of Figures

1. Context

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

2. Objective

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

3. Sparkling Wine Dataset

3.1. Data Overview

3.1.1. Import libraries and load the data

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

Figure 1: Sparkling Wine Data Overview

3.1.2. Check the structure of data

Shape of the dataset: 187 rows and 2 columns

3.1.3. Check the types of the data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   YearMonth  187 non-null    object
 1   Sparkling  187 non-null    int64
dtypes: int64(1), object(1)
memory usage: 3.0+ KB
```

Figure 2: Datatypes

3.1.4. Check for and treat (if needed) missing values

	0
YearMonth	0
Sparkling	0

Figure 3: Missing values check

3.1.5. Data Duplicates

There are no duplicate rows.

3.1.6. Statistical Summary

	count	mean	std	min	25%	50%	75%	max
Sparkling	187.0	2402.417112	1295.11154	1070.0	1605.0	1874.0	2549.0	7242.0

Figure 4: Statistical Summary - Numeric

3.1.7. Insights

- The standard deviation (1295.11) is more than 50% of the mean, indicating significant variability in sales.
- The mean (2402.42) is higher than the median (1874.0), suggesting the data is positively skewed, possibly due to some very high sales values (e.g., the maximum of 7242.0).
- Consistently high values in the 75th percentile and maximum range may reflect periods of heightened demand, which can guide targeted marketing strategies during similar times.

3.2. Exploratory Data Analysis

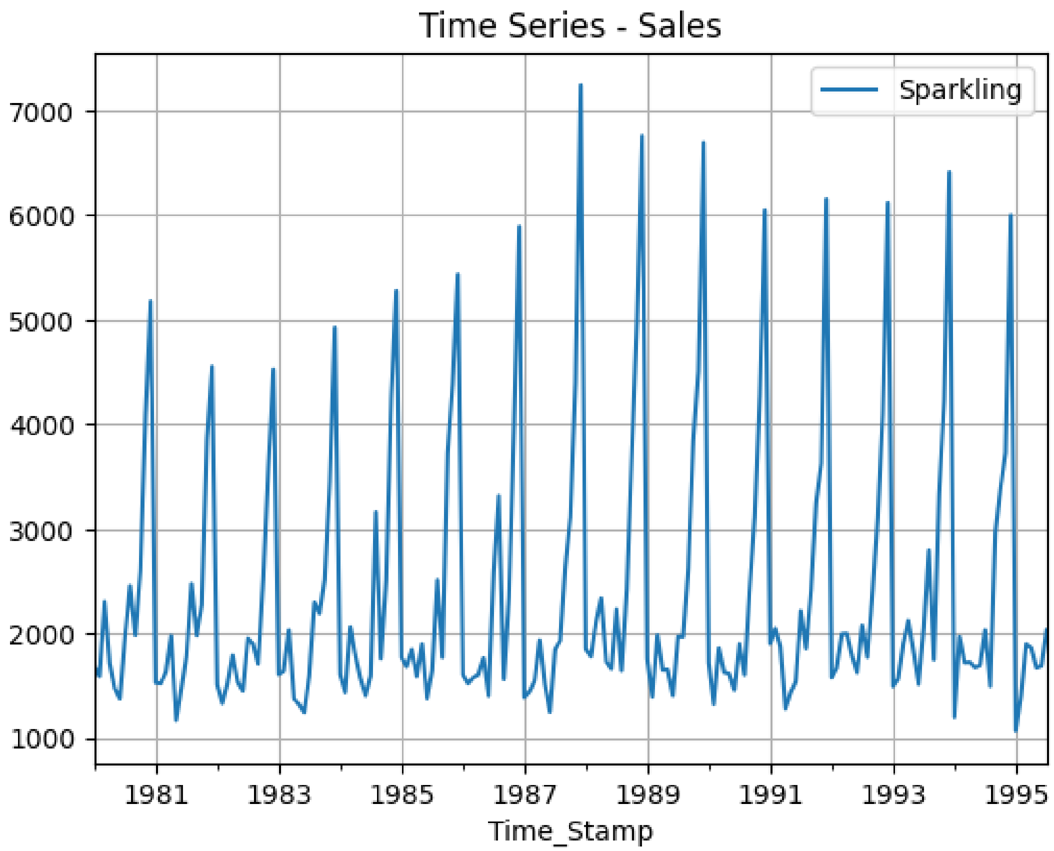


Figure 5: Time Series Plot

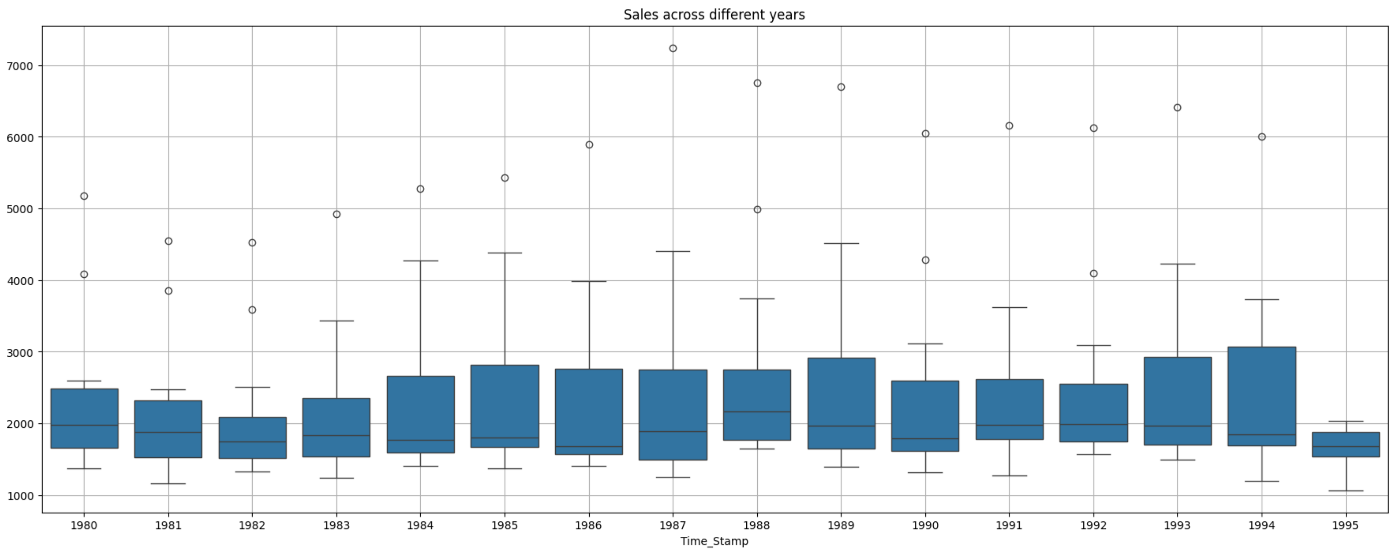


Figure 6: Sales across different years

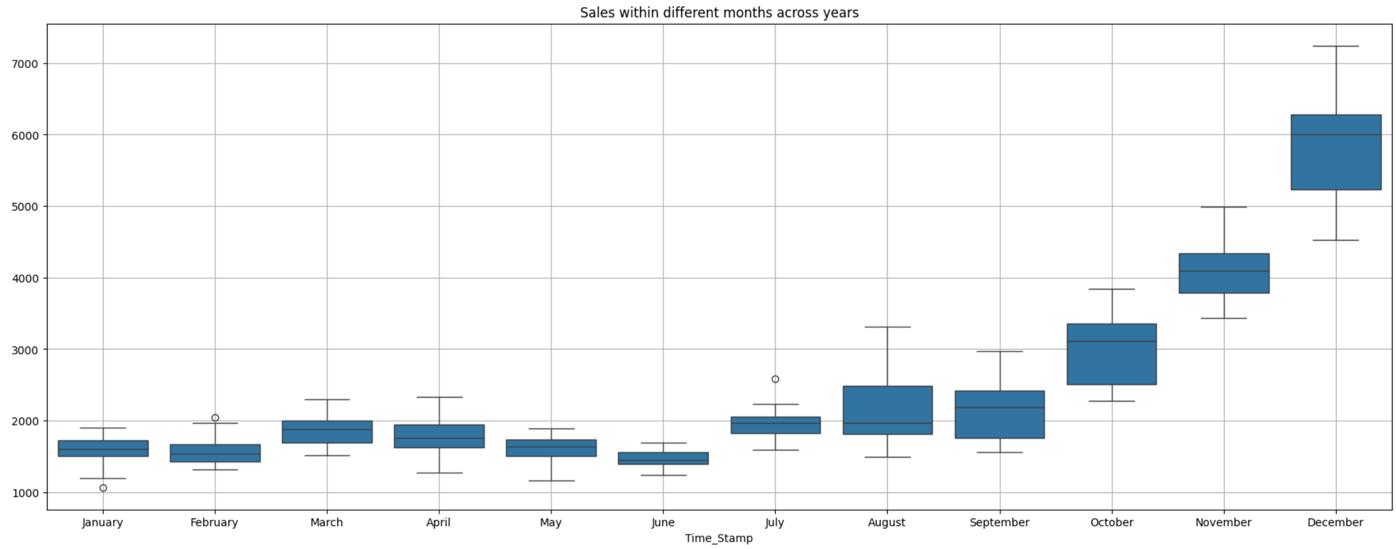


Figure 7: Sales within different months across years

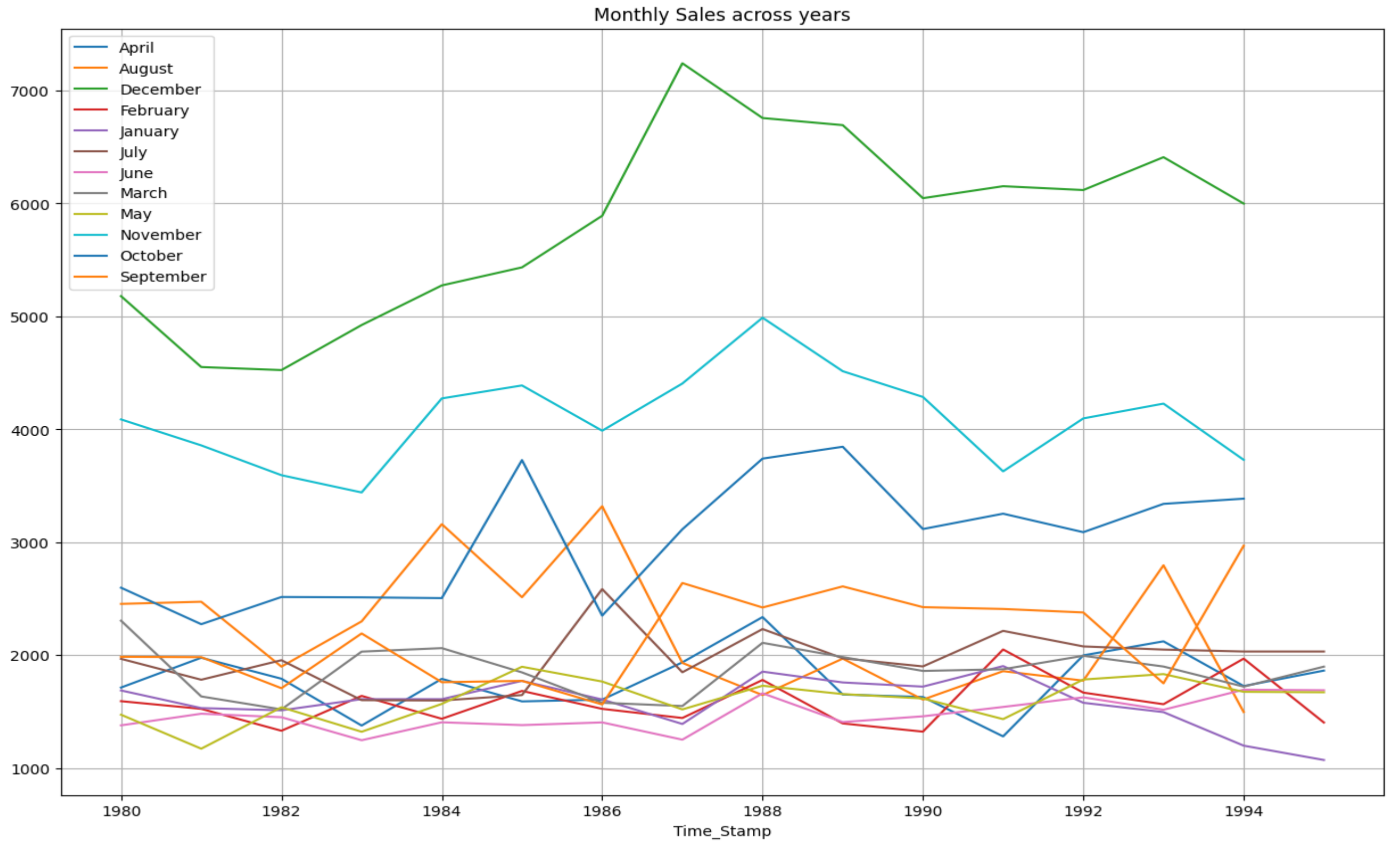


Figure 8: Monthly Sales

Insights

- December is the standout peak for sales across years, reflecting strong holiday demand.
- Sales rise from 1980 to the mid-1980s, particularly in December and mid-year months, with 1987 and 1988 showing peak sales in December.
- Sales stabilize in the early 1990s, though there is some variation across months.

3.2.1. Time series Decomposition

Additive

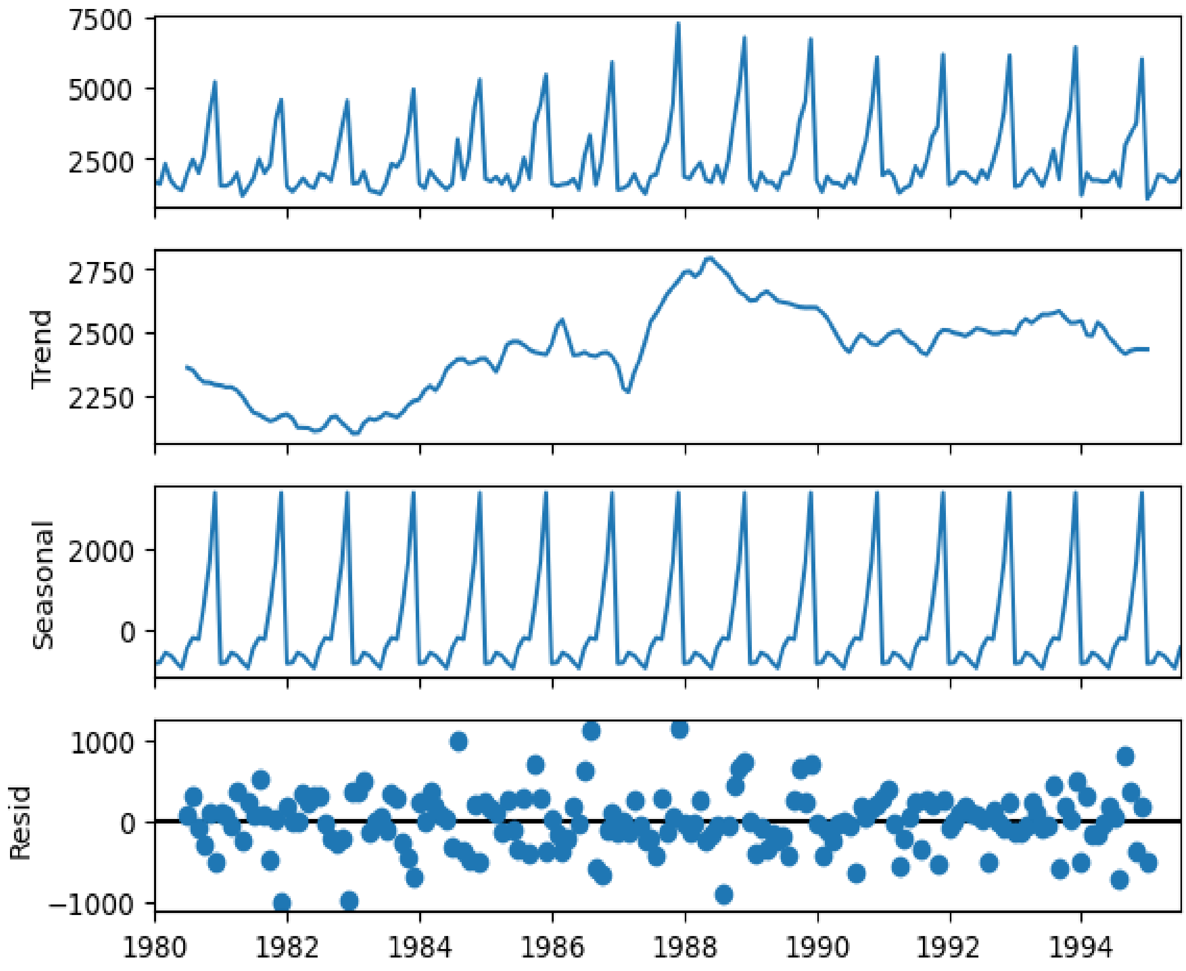


Figure 9: Additive

```
Trend
 Time_Stamp
1980-01-01          NaN
1980-02-01          NaN
1980-03-01          NaN
1980-04-01          NaN
1980-05-01          NaN
1980-06-01          NaN
1980-07-01      2360.67
1980-08-01      2351.33
1980-09-01      2320.54
1980-10-01      2303.58
1980-11-01      2302.04
1980-12-01      2293.79
Name: trend, dtype: float64

Seasonality
 Time_Stamp
1980-01-01     -854.26
1980-02-01     -830.35
1980-03-01     -592.36
1980-04-01     -658.49
1980-05-01     -824.42
1980-06-01     -967.43
1980-07-01     -465.50
1980-08-01     -214.33
1980-09-01     -254.68
1980-10-01      599.77
1980-11-01     1675.07
1980-12-01     3386.98
Name: seasonal, dtype: float64

Residual
 Time_Stamp
1980-01-01        NaN
1980-02-01        NaN
1980-03-01        NaN
1980-04-01        NaN
1980-05-01        NaN
1980-06-01        NaN
1980-07-01      70.84
1980-08-01     316.00
1980-09-01     -81.86
1980-10-01    -307.35
1980-11-01     109.89
1980-12-01    -501.78
Name: resid, dtype: float64
```

Figure 10: Additive Components

Multiplicative

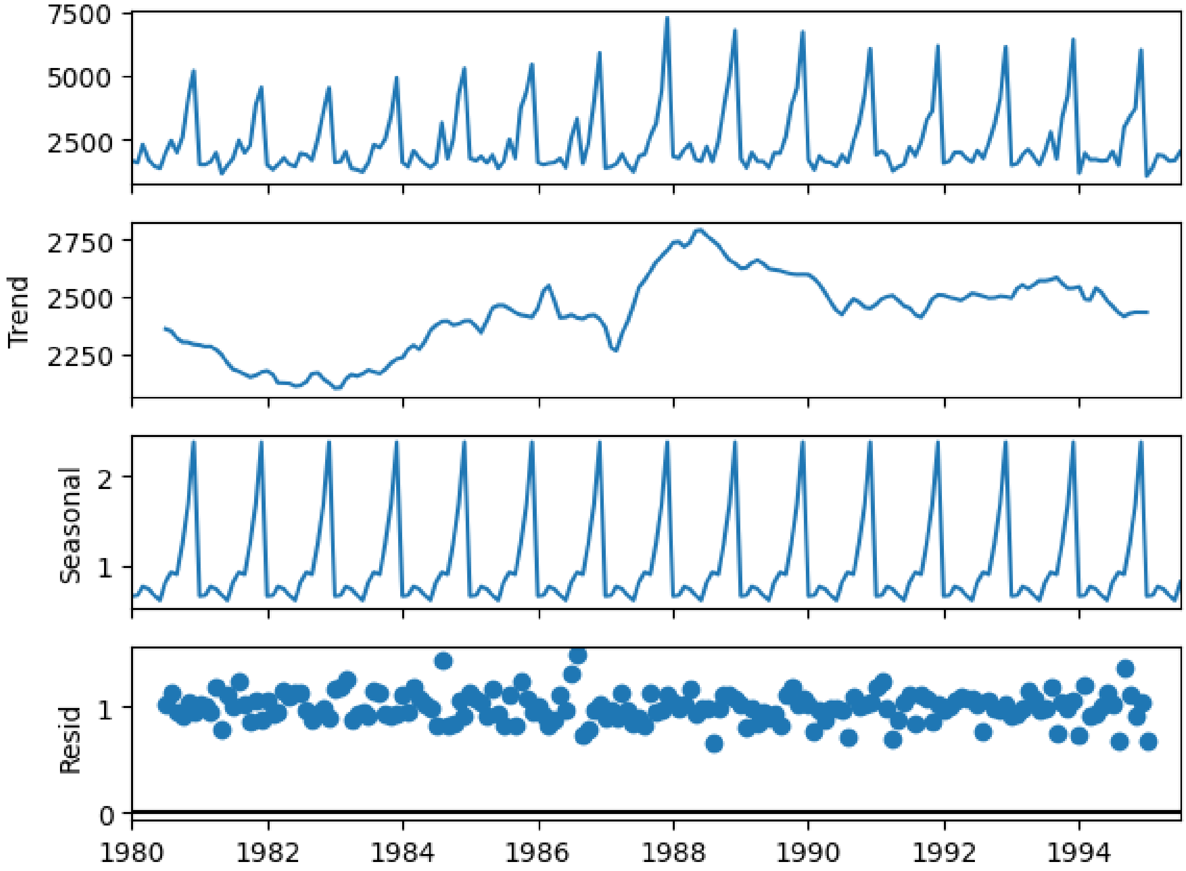


Figure 11: Multiplicative

```
Trend
 Time_Stamp
1980-01-01          NaN
1980-02-01          NaN
1980-03-01          NaN
1980-04-01          NaN
1980-05-01          NaN
1980-06-01          NaN
1980-07-01      2360.67
1980-08-01      2351.33
1980-09-01      2320.54
1980-10-01      2303.58
1980-11-01      2302.04
1980-12-01      2293.79
Name: trend, dtype: float64

Seasonality
 Time_Stamp
1980-01-01     0.65
1980-02-01     0.66
1980-03-01     0.76
1980-04-01     0.73
1980-05-01     0.66
1980-06-01     0.60
1980-07-01     0.81
1980-08-01     0.92
1980-09-01     0.89
1980-10-01     1.24
1980-11-01     1.69
1980-12-01     2.38
Name: seasonal, dtype: float64

Residual
 Time_Stamp
1980-01-01      NaN
1980-02-01      NaN
1980-03-01      NaN
1980-04-01      NaN
1980-05-01      NaN
1980-06-01      NaN
1980-07-01     1.03
1980-08-01     1.14
1980-09-01     0.96
1980-10-01     0.91
1980-11-01     1.05
1980-12-01     0.95
Name: resid, dtype: float64
```

Figure 12: Multiplicative Components

Insights

- It appears that the multiplicative decomposition may be a better fit for this data. The seasonal fluctuations seem to have a stronger relationship with the level of the series, which is characteristic of a multiplicative model.
- In the multiplicative model, the seasonal variations grew/shrank proportionally with the trend, while in this additive model, the seasonal variations are constant over time.
- If the seasonal variation in the data changes with the level of the trend, the multiplicative model is better suited.

3.3. Data Preprocessing

3.3.1. Missing Value treatment

There are no missing values.

3.3.2. Duplicate value check

There are no duplicate rows.

3.3.3. Train – Test Split

Split on 70:30 ratio.

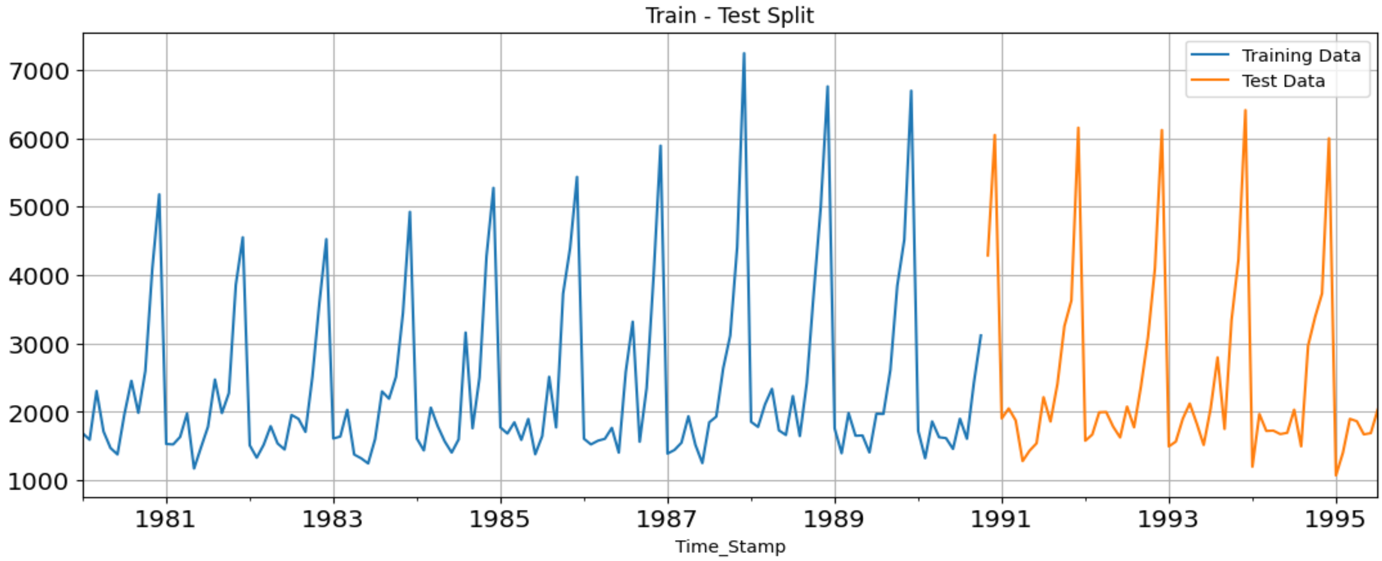


Figure 13: Train Test Split

3.4. Model Building- Original Data

3.4.1. Linear Regression

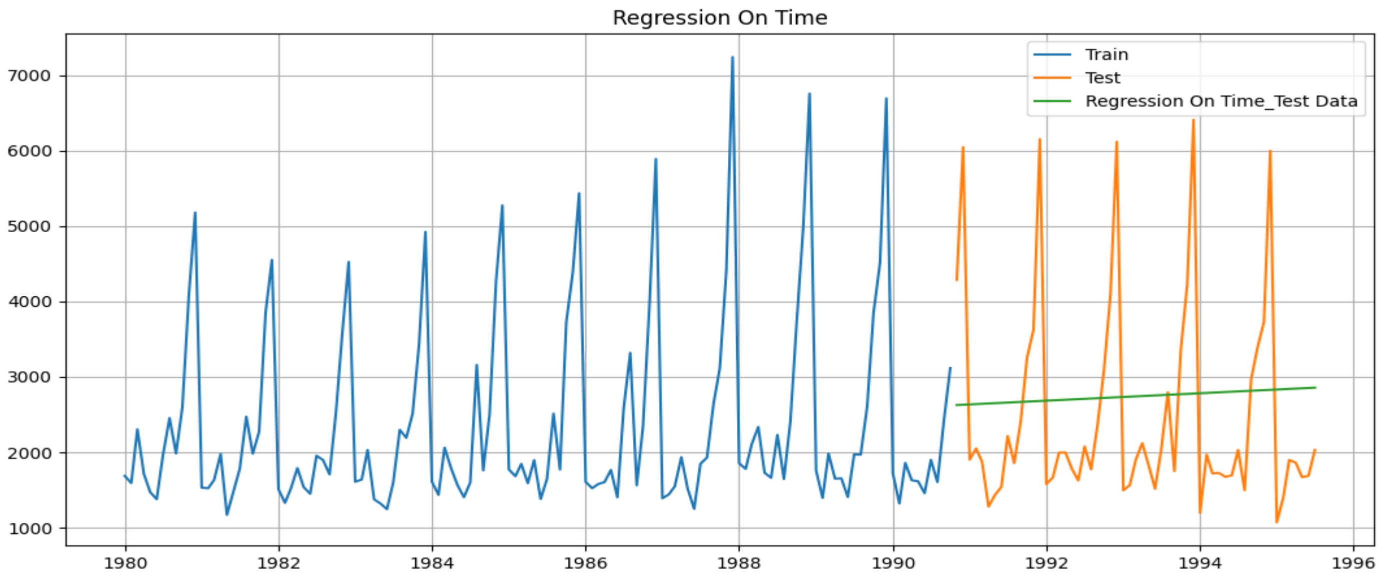


Figure 14: Linear Regression

For Linear regression forecast on the Test Data, RMSE is 1391.71

3.4.2. Simple Average

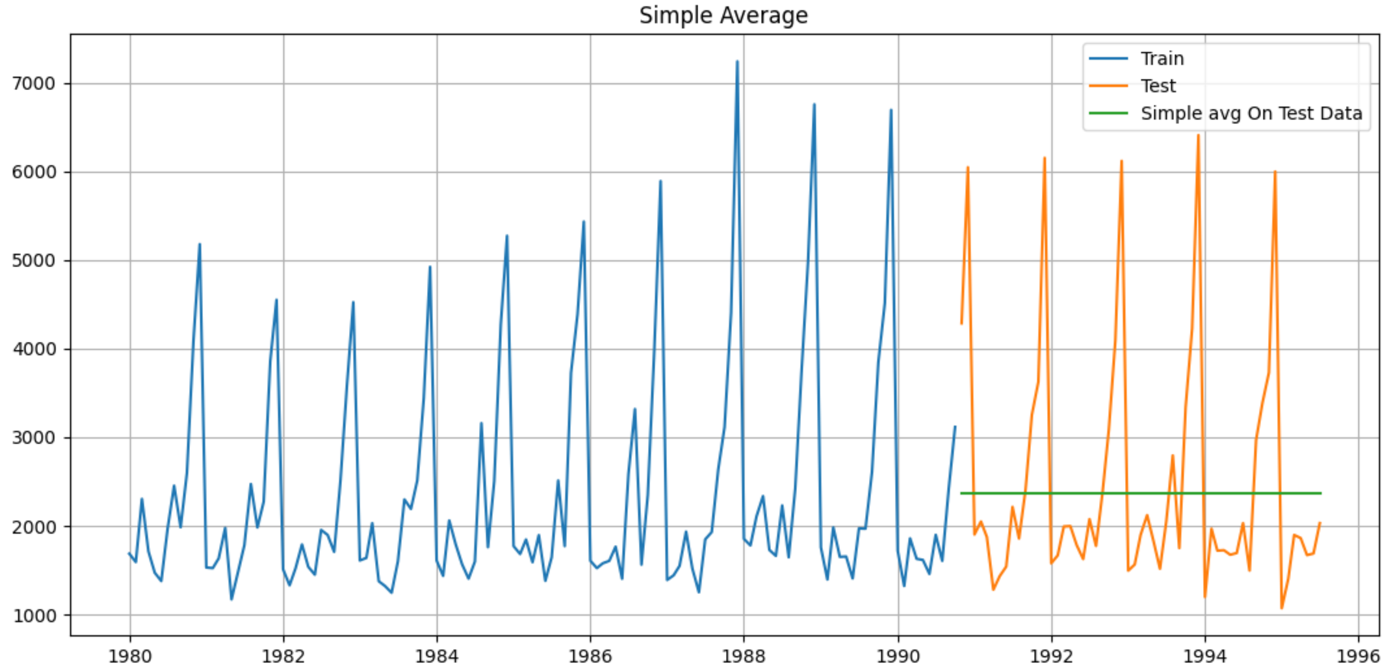


Figure 15: Simple Average

For Simple Average forecast on the Test Data, RMSE is 1368.77

3.4.3. Moving Average

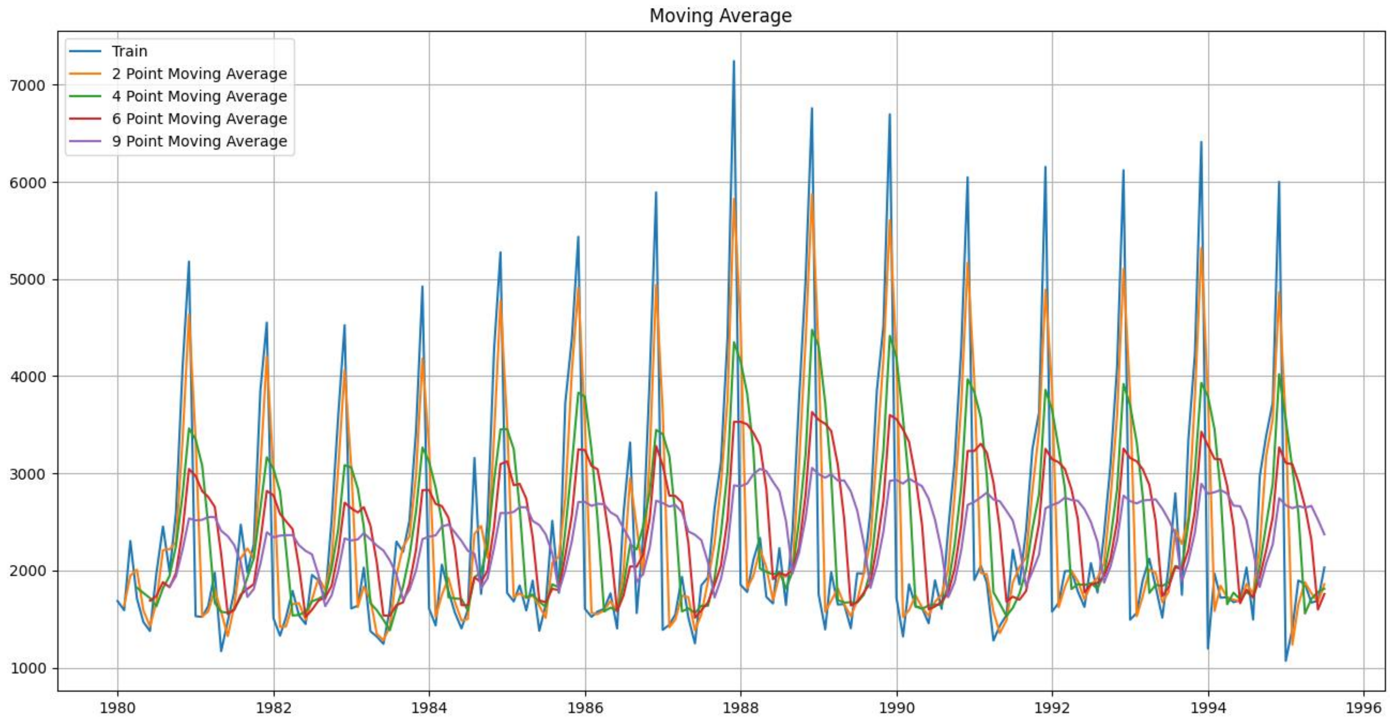


Figure 16: Moving Average

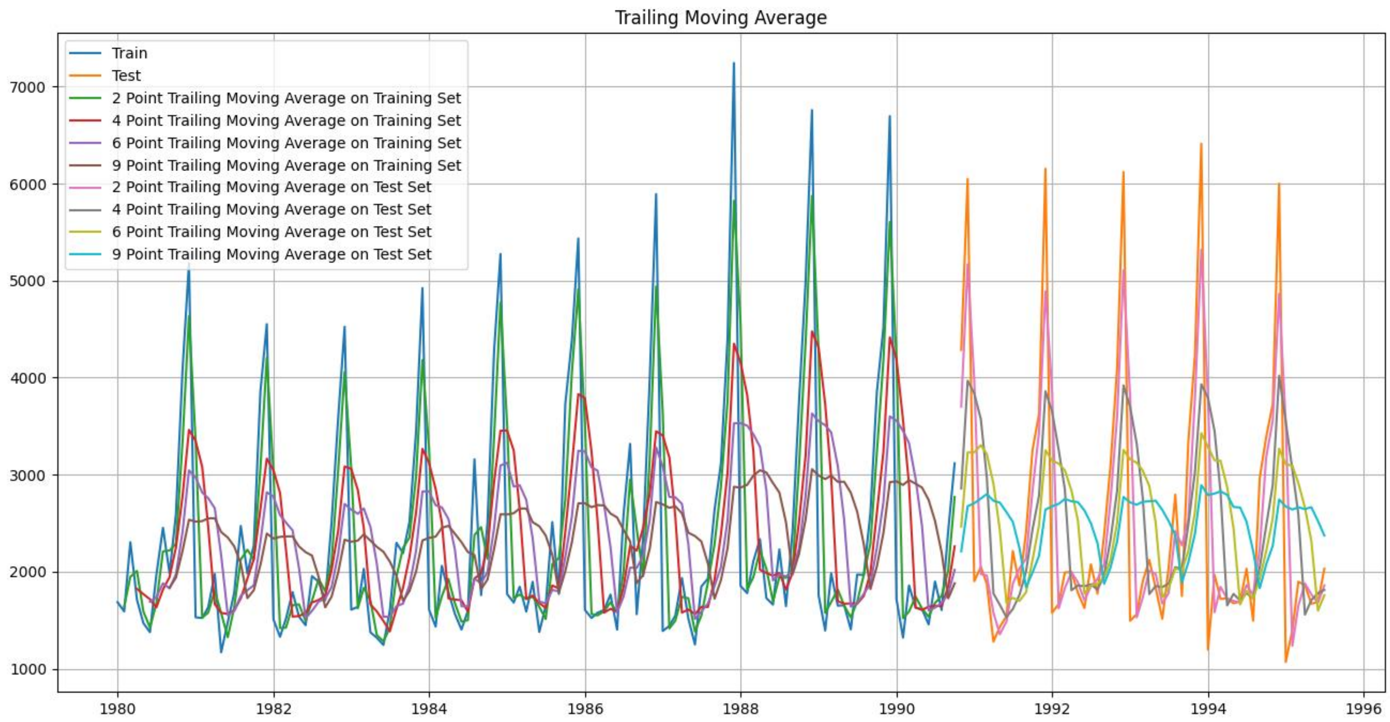


Figure 17: Trailing Moving Average

For 2 point Moving Average Model forecast on the Training Data, RMSE is 811.179

For 4 point Moving Average Model forecast on the Training Data, RMSE is 1184.213

For 6 point Moving Average Model forecast on the Training Data, RMSE is 1337.201

For 9 point Moving Average Model forecast on the Training Data, RMSE is 1422.653

3.4.4. Simple Exponential Smoothing

```
{'smoothing_level': 0.037534299016257884,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 1686.0,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 18: SES Model Params

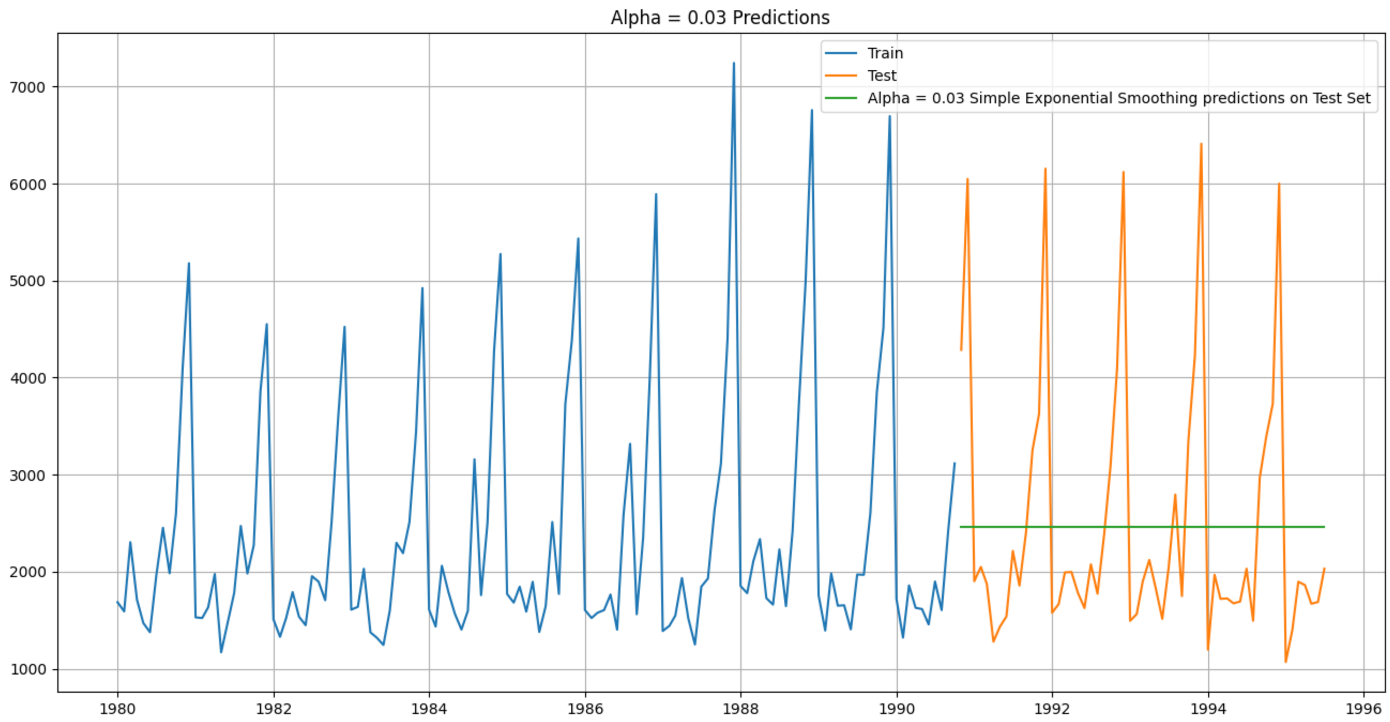


Figure 19: SES Model

For Alpha = 0.03 Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 1362.429

Different Alpha values

Alpha Values	Train RMSE	Test RMSE
0.3	1331.102204	1372.323705
0.4	1329.814823	1363.037803
0.5	1326.403864	1364.863549
0.6	1325.588422	1379.988733
0.7	1329.257530	1404.659104
0.8	1337.879425	1434.578214
0.9	1351.645478	1466.179706

Figure 20: RMSE for different alpha values

Chose Alpha = 0.4 based on Test RMSE

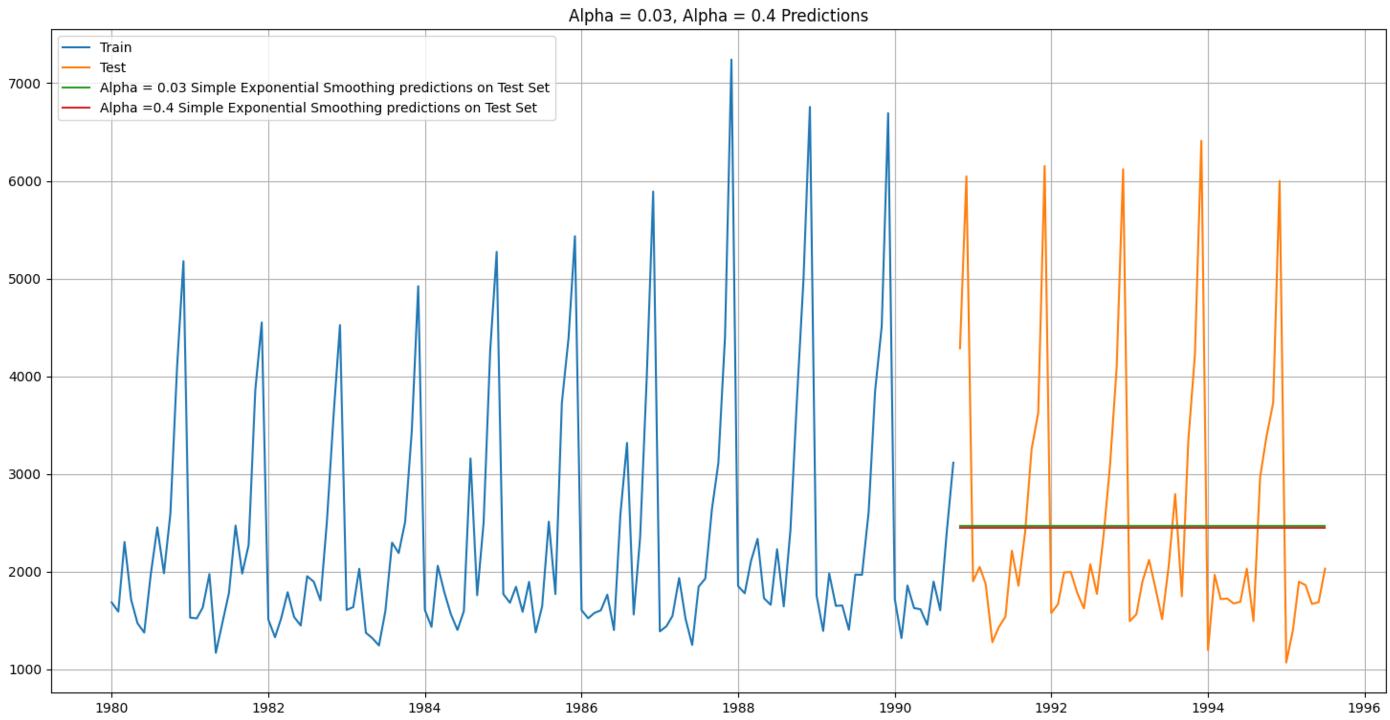


Figure 21: SES Predictions for different alpha values

3.4.5. Double Exponential Smoothing – Holt's Model

Computed RMSE for different Alpha and Beta values.

Alpha Values	Beta Values	Train RMSE	Test RMSE
0.3	0.3	1567.524066	1597.853999
0.3	0.4	1662.549225	4023.672164
0.3	0.5	1758.543876	8879.172380
0.3	0.6	1843.560670	15645.080035
0.3	0.7	1902.735965	23205.442323
...	...	...	...
1.0	0.6	1764.658812	20558.025827
1.0	0.7	1837.425218	22155.074151
1.0	0.8	1915.148280	23241.839479
1.0	0.9	1999.362743	23787.747852
1.0	1.0	2092.531564	23712.944127

Figure 22: DES Alpha Beta RMSE

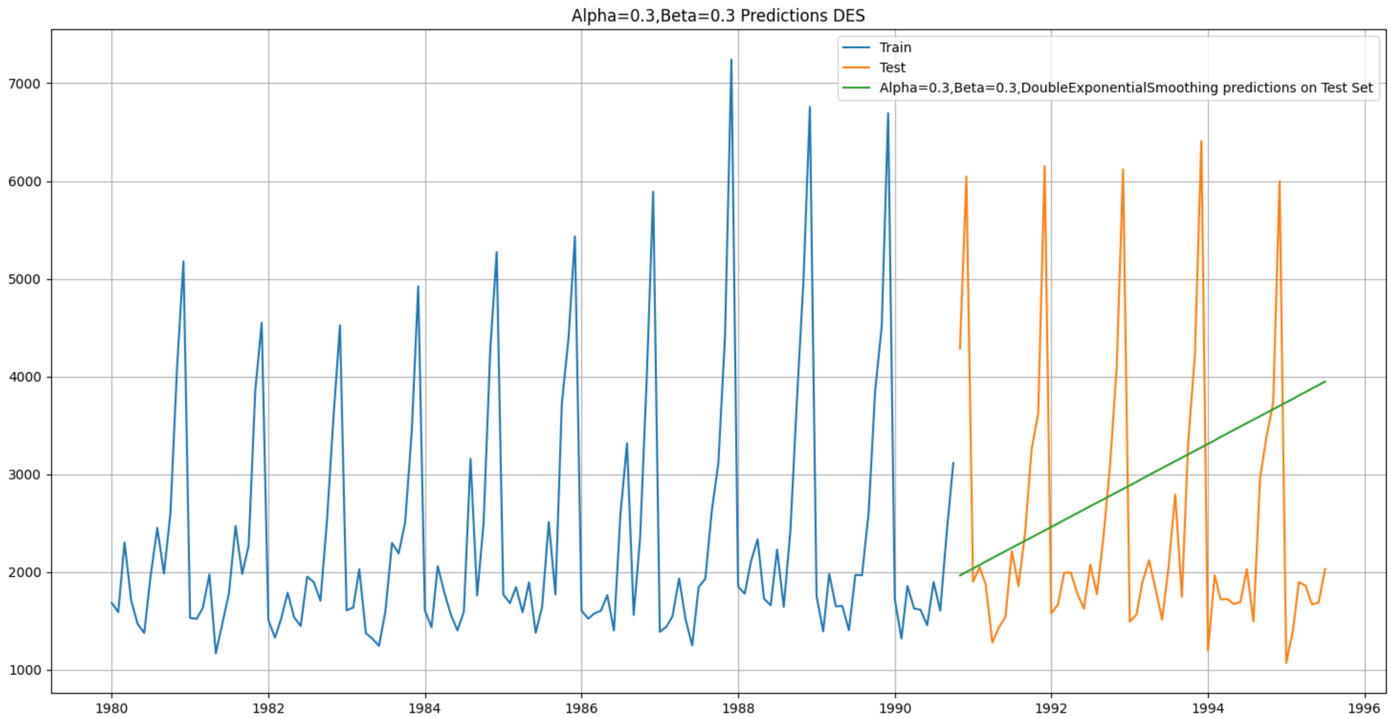


Figure 23: DES Predictions

3.4.6. Triple Exponential Smoothing – Holt's Winter

```
{'smoothing_level': 0.07571445210103464,
 'smoothing_trend': 0.06489808813237438,
 'smoothing_seasonal': 0.3765608370780376,
 'damping_trend': nan,
 'initial_level': 2356.54174944041,
 'initial_trend': -9.180926180482402,
 'initial_seasons': array([0.71186629, 0.67768289, 0.89647955, 0.79722705, 0.64099767,
        0.64026213, 0.86701095, 1.11336214, 0.89797444, 1.18549449,
        1.8343214 , 2.32723166]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 24: TES params

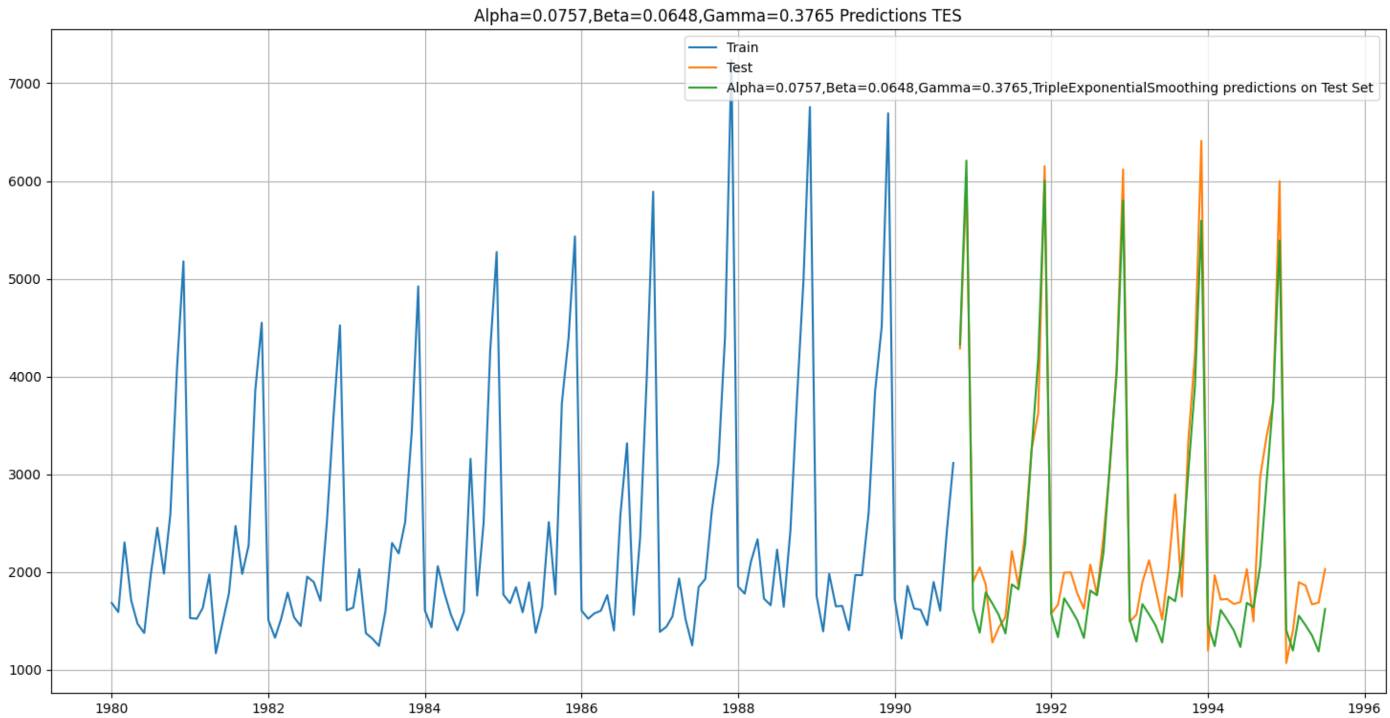


Figure 25: TES Predictions

For Alpha=0.0757,Beta=0.0648,Gamma=0.3765, Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 381.657

Computed RMSE for different Alpha, Beta and Gamma values.

Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
0.7	0.4	0.3	512.023844	422.908833
0.5	0.5	0.3	472.088500	451.601686
0.5	0.8	0.4	625.557444	481.151676
0.6	0.4	0.3	479.344459	498.796626
0.8	0.4	0.3	544.126424	502.371290

Figure 26: RMSE Alpha Beta Gamma Values

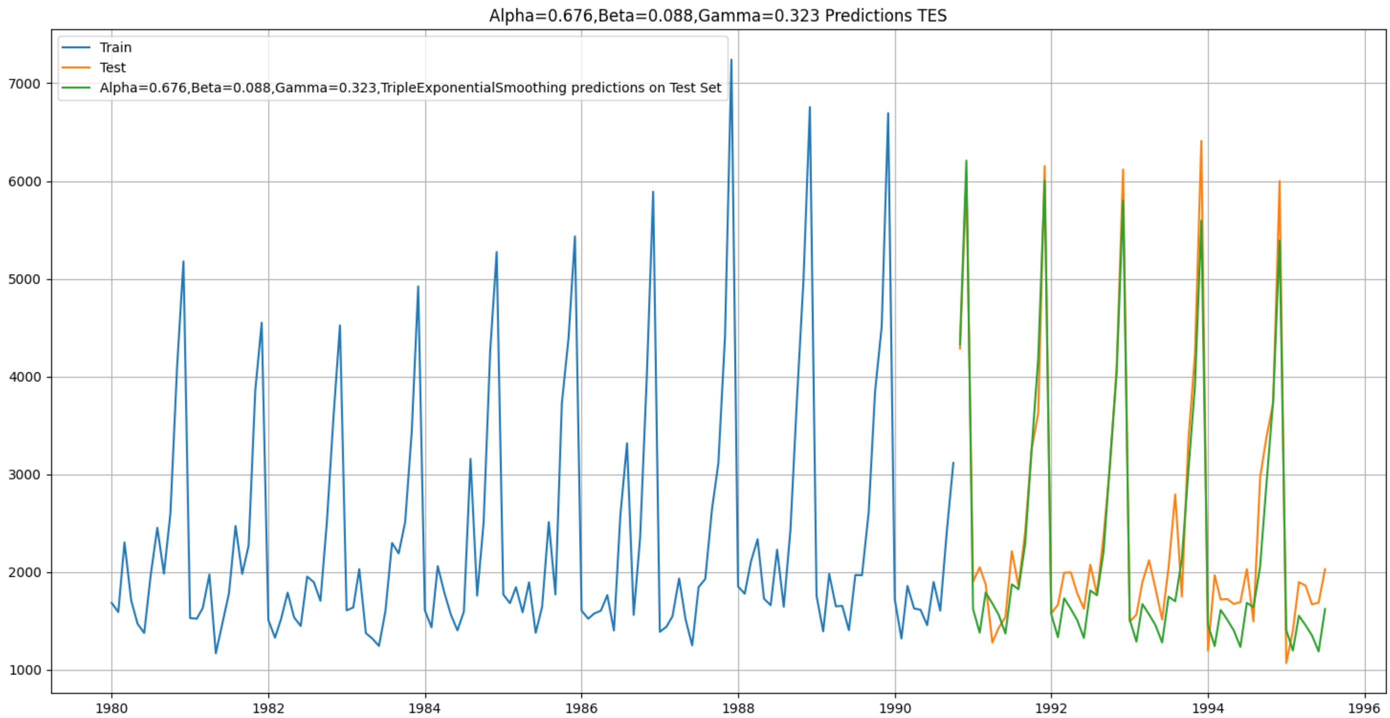
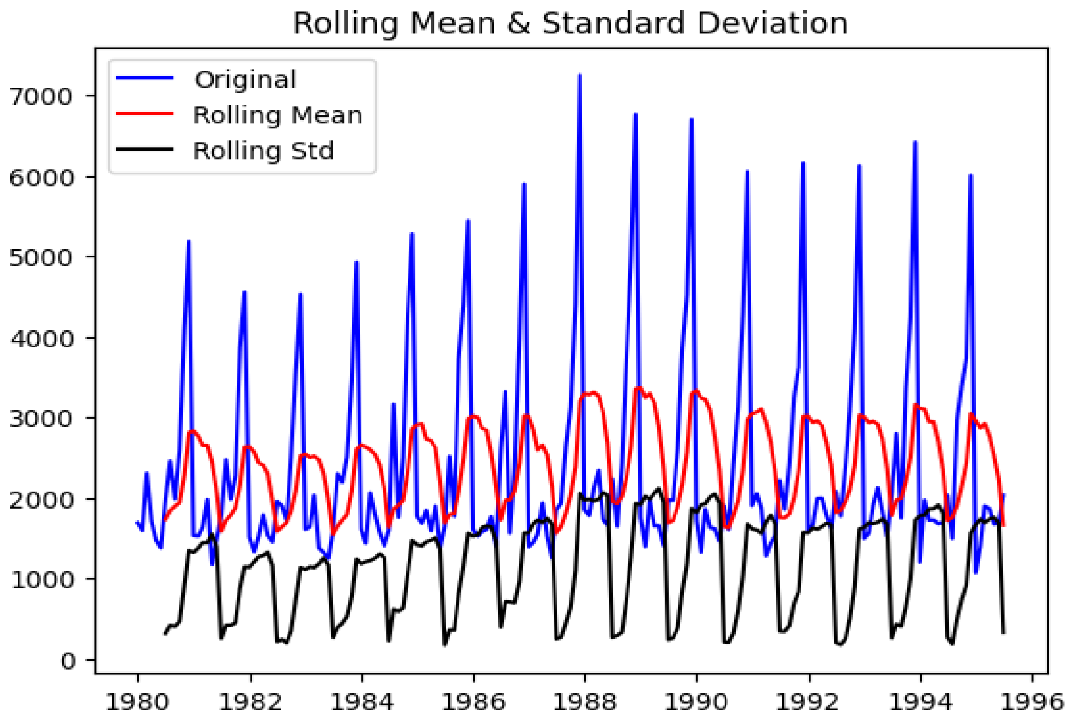


Figure 27: TES Prediction for different params

Insights

- Model: Triple Exponential Smoothing (Alpha=0.0757, Beta=0.0648, Gamma=0.3765)
- RMSE: 381.657232
- This model has the lowest RMSE, indicating it provides the most accurate forecasts among all models evaluated. It is well-suited for data with seasonality, which aligns with the observed patterns in Rose wine sales.
- The 2-point trailing moving average performs better (RMSE = 811.178937) than the 4-point, 6-point, or 9-point trailing moving averages. However, all moving average models have significantly higher RMSE values than the triple exponential smoothing models.
- Regression on time (RMSE = 1391.708631) and simple average (RMSE = 1368.774051) are among the least accurate models.
- Use Triple Exponential Smoothing (Alpha=0.0757, Beta=0.0648, Gamma=0.3765) for forecasting. It has the lowest RMSE and is designed to capture seasonality, trend, and level changes effectively.
- If further validation is required, consider the second triple exponential smoothing model (Alpha=0.7, Beta=0.4, Gamma=0.3), which also performs well but is slightly less accurate.

3.5. Check for stationarity

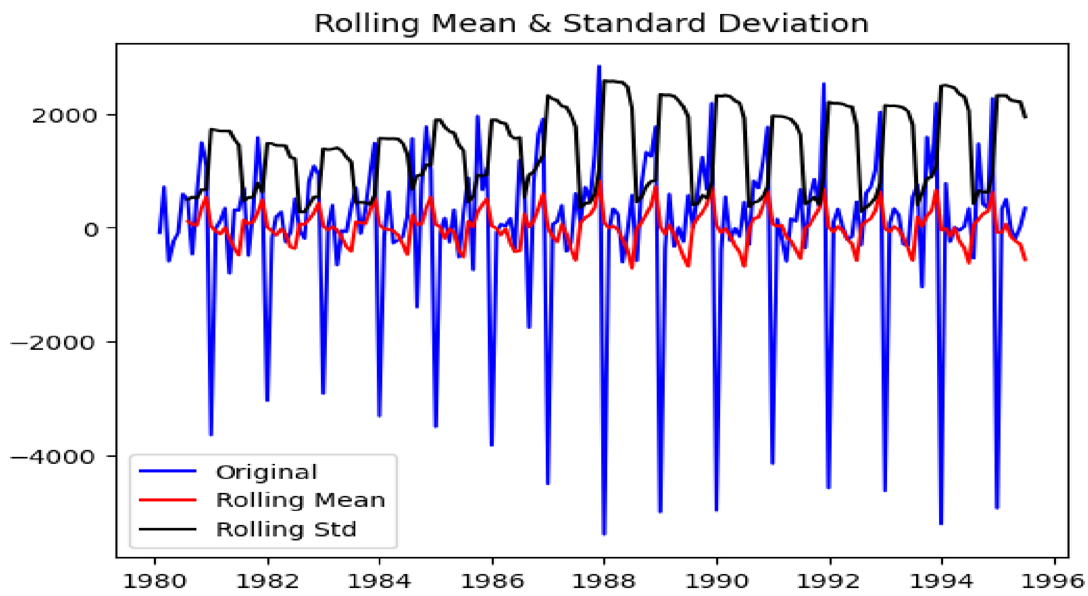



```
Results of Dickey-Fuller Test:
Test Statistic                  -1.360497
p-value                          0.601061
#Lags Used                      11.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
dtype: float64
```

Figure 28: Dickey Fuller Test

The results of the Dickey-Fuller Test indicate that the time series is not stationary. Here's the interpretation:
Test Statistic = -1.360497:
This value is not smaller than the critical values at 1%, 5%, or 10% significance levels, meaning we fail to reject the null hypothesis.
p-value = 0.601061:
Since the p-value is greater than 0.05 (or 0.01 for stricter significance), we do not have sufficient evidence to reject the null hypothesis that the series has a unit root (i.e., it is non-stationary).
Critical Values:
1% level = -3.468280
5% level = -2.878202
10% level = -2.575653
Since the test statistic (-1.360497) is greater than all three critical values, we fail to reject the null hypothesis.

Differencing the Series




```
Results of Dickey-Fuller Test:
Test Statistic                 -45.050301
p-value                          0.000000
#Lags Used                      10.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
dtype: float64
```

Figure 29: Dickey Fuller Test After differencing

The results of the Dickey-Fuller Test indicate that the time series is stationary. Here's the interpretation:

Test Statistic = -45.050301
This value is much smaller than all the critical values at 1%, 5%, and 10% significance levels. This suggests strong evidence to reject the null hypothesis that the series has a unit root.

p-value = 0.000000
The p-value is significantly less than 0.05 (or even 0.01), providing strong evidence to reject the null hypothesis that the series is non-stationary.

Critical Values:

1% level = -3.468280
5% level = -2.878202
10% level = -2.575653
The test statistic (-45.050301) is far smaller than all the critical values, further supporting the conclusion that the series is stationary.

- The original series was non-stationary, but after applying differencing, the series is now stationary.
- Since stationarity is a key assumption for ARIMA models, we can proceed with model fitting.
- If applied first-order differencing (d=1), we can now set d=1 in the ARIMA model.

3.6. Model Building – Stationary Data

3.6.1. ACF Plots

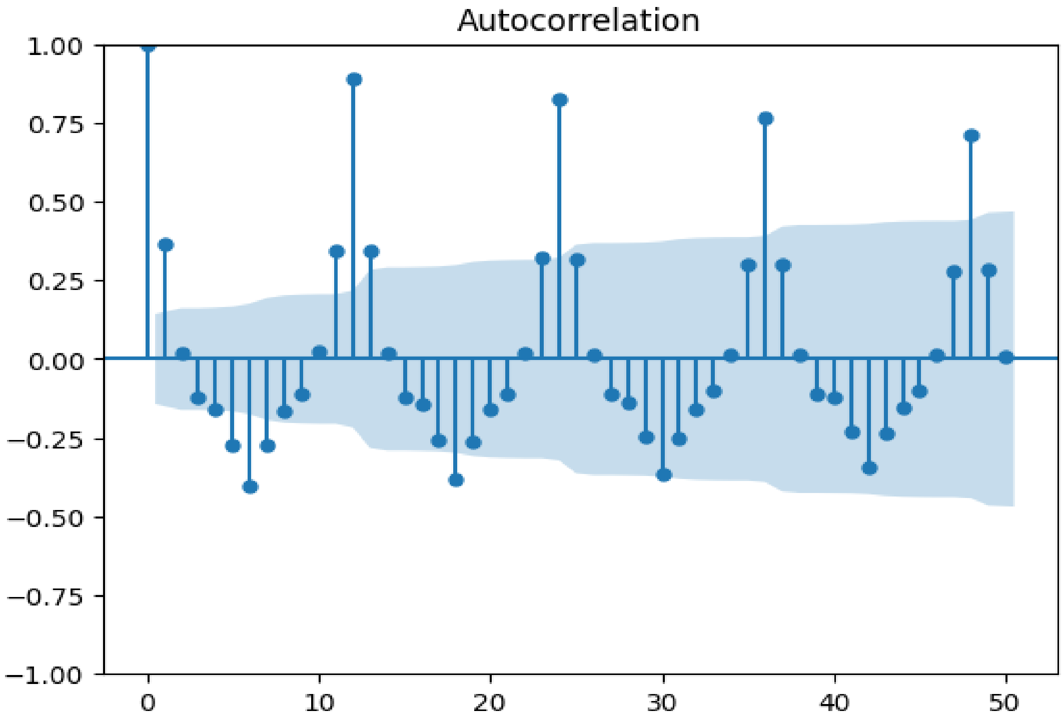


Figure 30: ACF

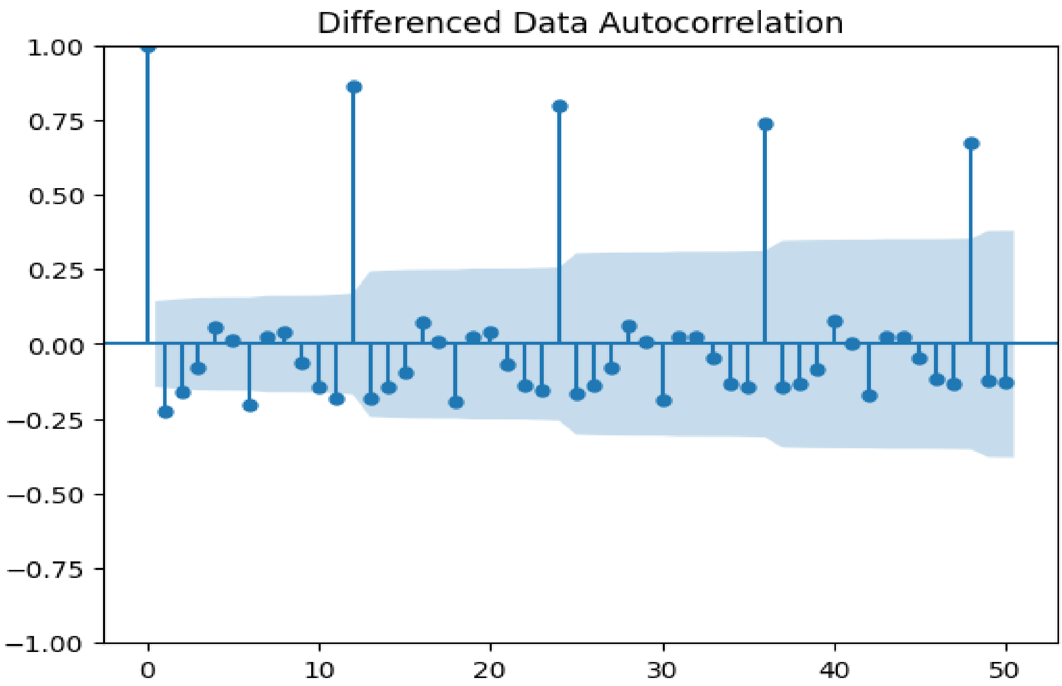


Figure 31: ACF Differenced

Insights

- The original data exhibits some cyclical or periodic behavior, as evidenced by the fluctuating autocorrelation values.
- Differencing the data has helped remove any non-stationarity, resulting in a more stable autocorrelation structure.

3.6.2. PACF Plots

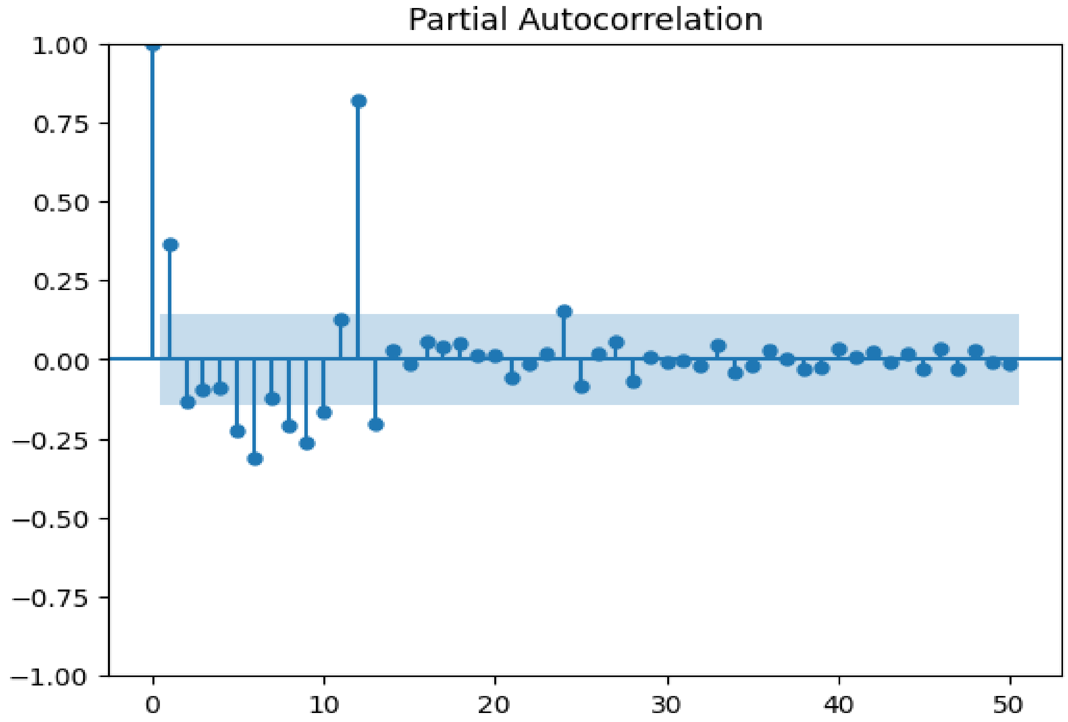


Figure 32: PACF

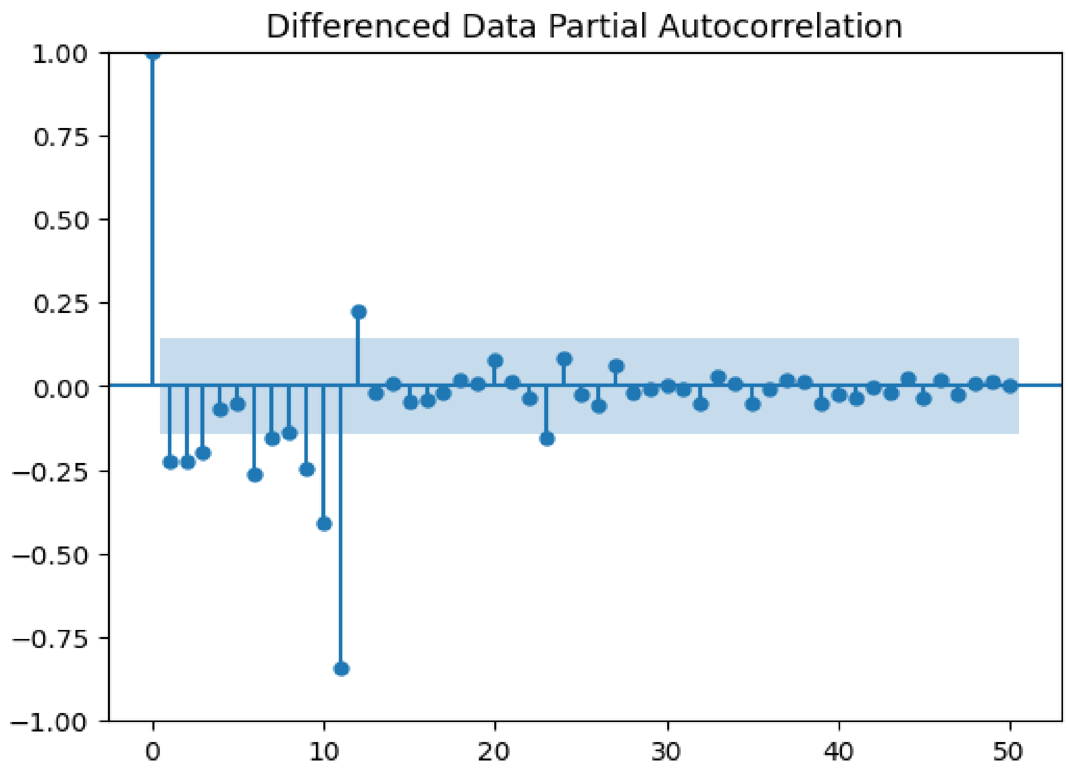


Figure 33: PACF Differenced

Insights

- The original data has an autoregressive structure, as indicated by the significant partial autocorrelation at the early lags.
- Differencing the data has helped remove any non-stationarity, resulting in a more stable PACF structure.

3.6.3. Auto ARIMA

```
Best model:  ARIMA(0,0,1)(0,0,0)[0] intercept
Total fit time: 1.708 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  130
Model:               SARIMAX(0, 0, 1)   Log Likelihood               -1099.723
Date:                Fri, 06 Dec 2024   AIC                           2205.446
Time:                        22:51:32   BIC                           2214.048
Sample:                    01-01-1980   HQIC                          2208.941
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept   2265.1074    284.849      7.952      0.000    1706.813    2823.402
ma.L1          0.4317      0.112      3.848      0.000       0.212       0.652
sigma2      1.357e+06   1.56e+05      8.677      0.000    1.05e+06    1.66e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.20   Jarque-Bera (JB):                39.99
Prob(Q):                              0.66   Prob(JB):                         0.00
Heteroskedasticity (H):               2.49   Skew:                             1.05
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.72
===================================================================================
```

Figure 34: Auto ARIMA Summary

For auto_ARIMA forecast on the Test Data, RMSE is 1374.66

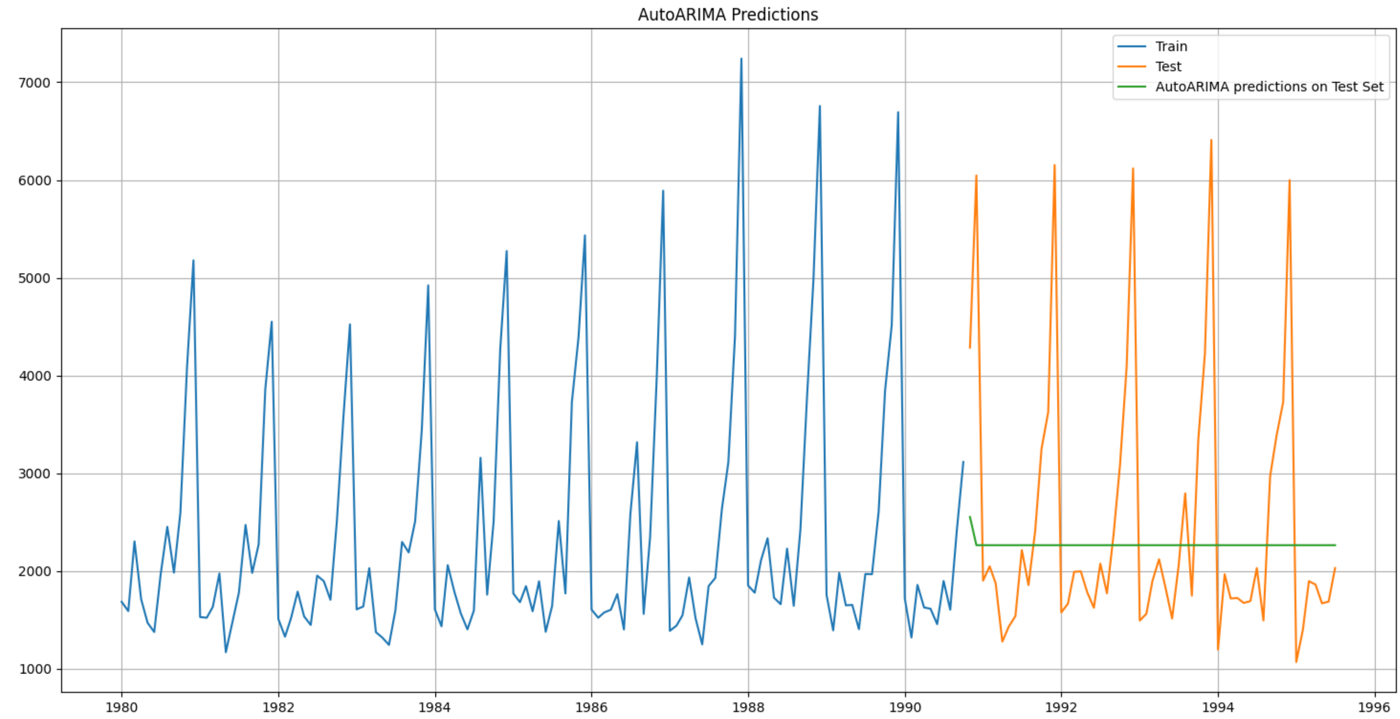


Figure 35: Auto ARIMA Predictions

3.6.4. ARIMA

param	AIC
(2, 1, 2)	2178.109742
(2, 1, 1)	2193.974962
(0, 1, 2)	2194.034361
(1, 1, 2)	2194.959653
(1, 1, 1)	2196.050086
(0, 1, 1)	2217.939217
(2, 1, 0)	2223.899470
(1, 1, 0)	2231.137663
(0, 1, 0)	2232.719438

Figure 36: AIC

ARIMA(2,1,2) has the lowest AIC.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1084.055
Date:                Sat, 07 Dec 2024   AIC                           2178.110
Time:                        21:25:46   BIC                           2192.409
Sample:                    01-01-1980   HQIC                          2183.920
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.3020      0.046     28.556      0.000       1.213       1.391
ar.L2         -0.5360      0.079     -6.744      0.000      -0.692      -0.380
ma.L1         -1.9914      0.110    -18.144      0.000      -2.207      -1.776
ma.L2          0.9997      0.110      9.069      0.000       0.784       1.216
sigma2      1.085e+06   2.04e-07   5.31e+12      0.000    1.08e+06    1.08e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):                19.53
Prob(Q):                              0.75   Prob(JB):                         0.00
Heteroskedasticity (H):               2.30   Skew:                             0.71
Prob(H) (two-sided):                  0.01   Kurtosis:                         4.27
===================================================================================
```

Figure 37: ARIMA Summary

For ARIMA forecast on the Test Data, RMSE is 1325.17

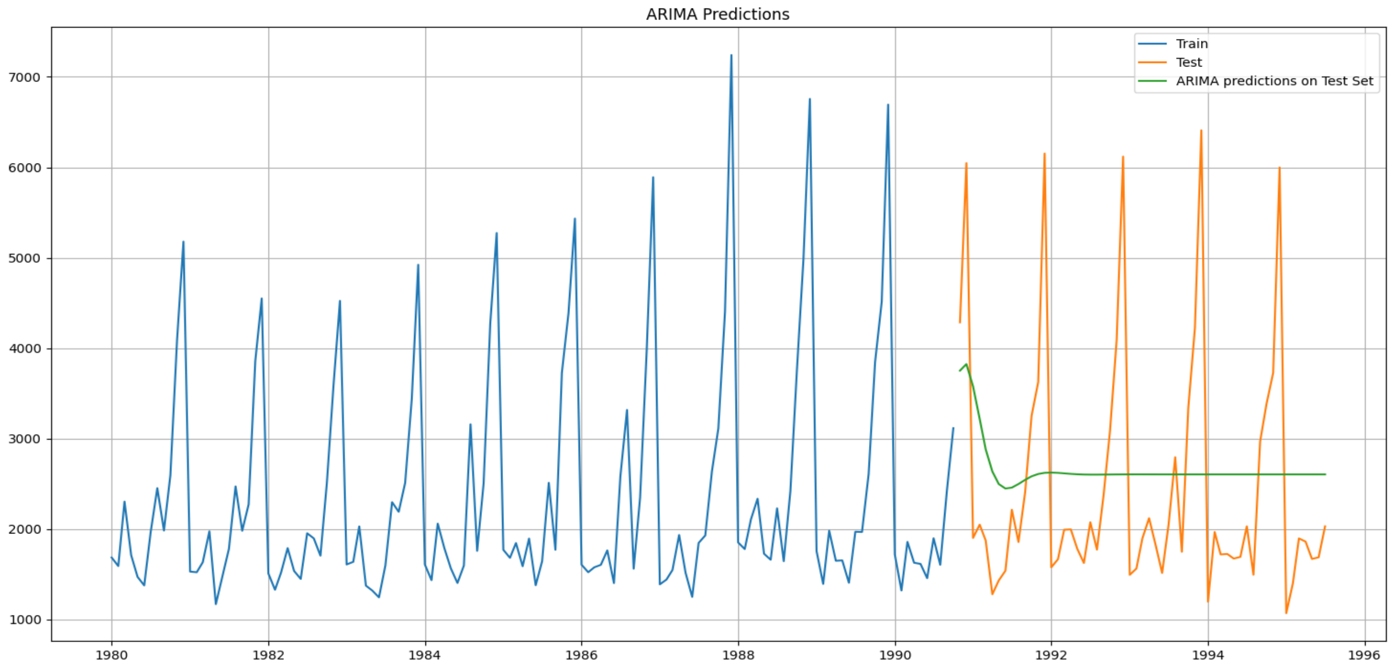


Figure 38: ARIMA Predictions

3.6.5. Auto SARIMA

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(0, 0, 1)x(0, 1, 1, 12)   Log Likelihood                -869.820
Date:                            Sat, 07 Dec 2024   AIC                           1747.640
Time:                                    21:26:08   BIC                           1758.723
Sample:                                01-01-1980   HQIC                          1752.140
                                     - 10-01-1990
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     42.8247     26.717      1.603      0.109      -9.540      95.190
ma.L1          0.1764      0.090      1.953      0.051      -0.001       0.353
ma.S.L12      -0.4656      0.072     -6.491      0.000      -0.606      -0.325
sigma2      1.472e+05   1.39e+04     10.621      0.000     1.2e+05    1.74e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                48.60
Prob(Q):                              0.95   Prob(JB):                         0.00
Heteroskedasticity (H):               3.37   Skew:                             0.84
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.66
===================================================================================
```

Figure 39: Auto SARIMA Summary

For auto SARIMA forecast on the Test Data, RMSE is 426.96

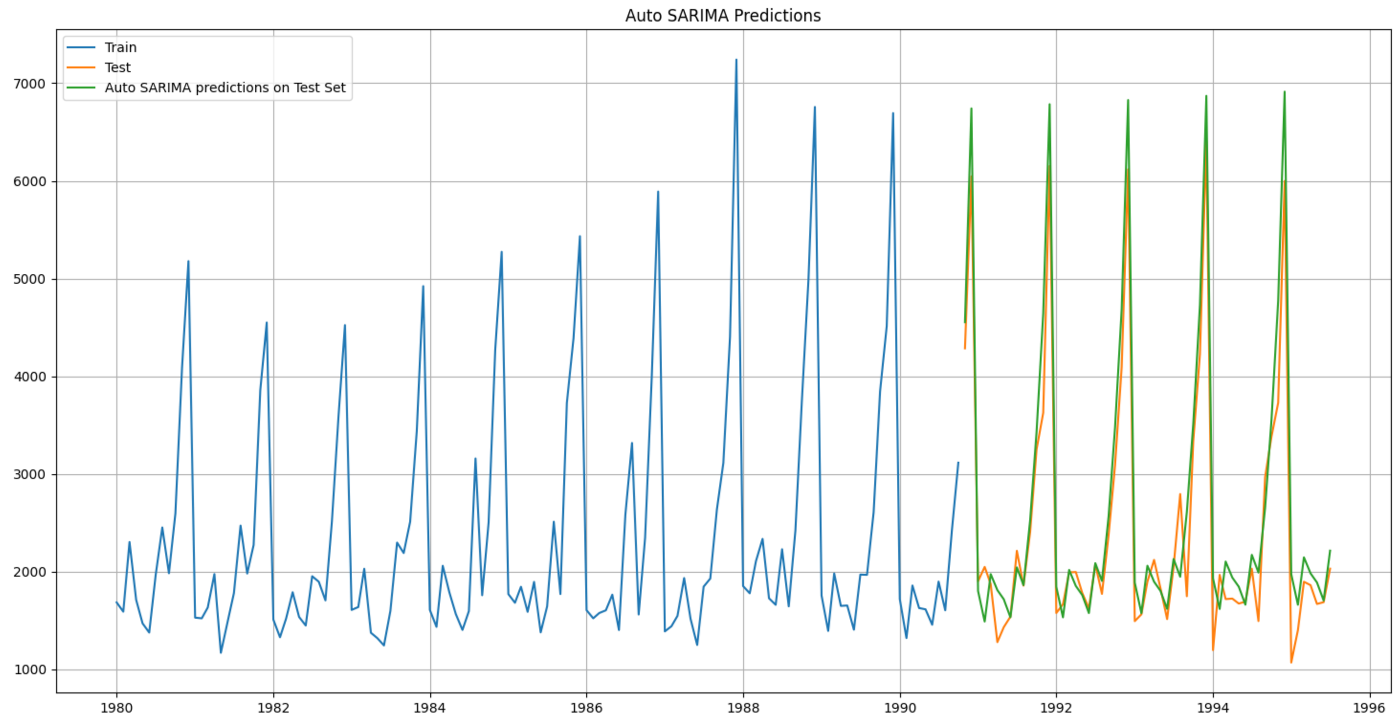


Figure 40: Auto SARIMA Predictions

3.6.6. SARIMA

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                            Sparkling   No. Observations:                130
Model:             SARIMAX(0, 1, 2)x(0, 1, [1], 12)   Log Likelihood                -867.096
Date:                              Sat, 07 Dec 2024   AIC                           1742.193
Time:                                      21:29:20   BIC                           1753.241
Sample:                                  01-01-1980   HQIC                          1746.678
                                       - 10-01-1990
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7754      2.413     -0.321      0.748      -5.504       3.953
ma.L2         -0.2243      0.583     -0.385      0.700      -1.366       0.918
ma.S.L12      -0.4290      0.072     -5.959      0.000      -0.570      -0.288
sigma2      1.477e+05   3.57e+05      0.414      0.679   -5.51e+05    8.47e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB):                37.64
Prob(Q):                              0.84   Prob(JB):                         0.00
Heteroskedasticity (H):               2.89   Skew:                             0.74
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.35
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 41: SARIMA Summary

For SARIMA forecast on the Test Data, RMSE is 412.82

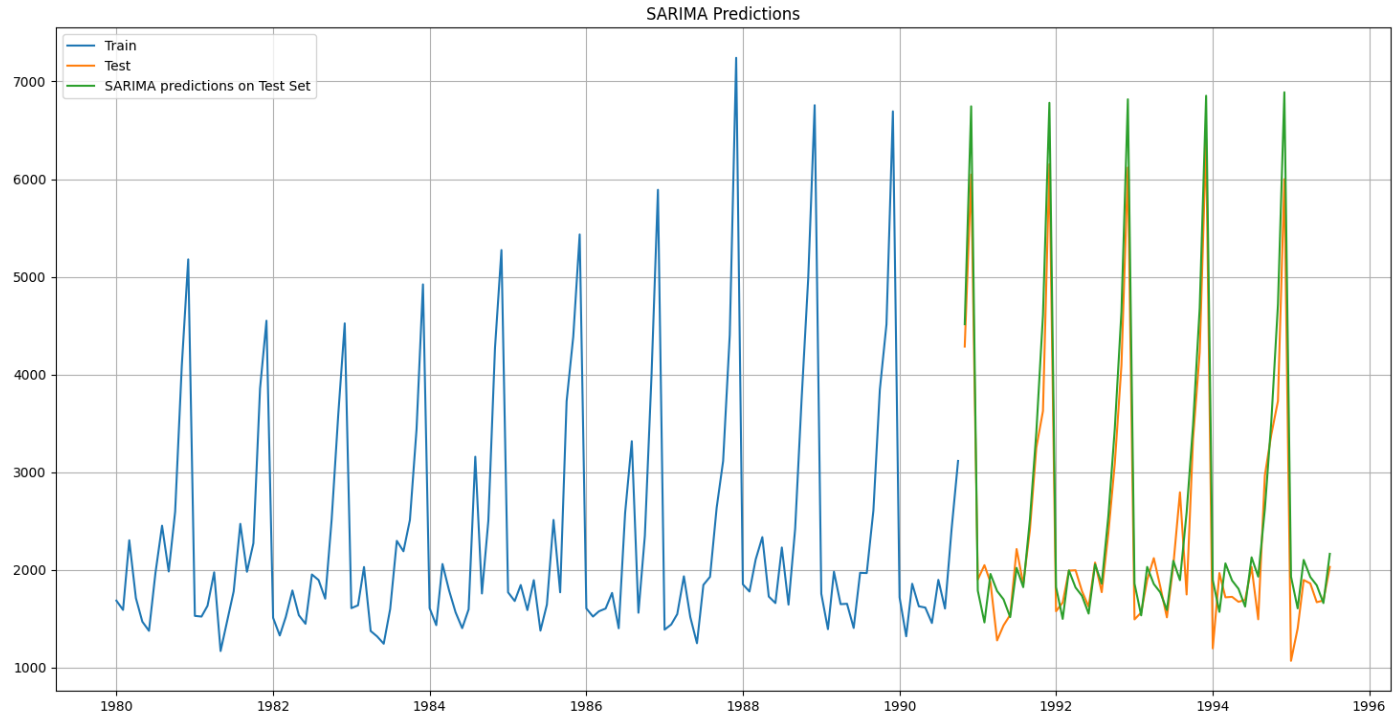


Figure 42: SARIMA Predictions

Insights

- Model: SARIMA
- RMSE: 412.824322
- The SARIMA model has the lowest RMSE, making it the most accurate among all models tested. SARIMA effectively handles seasonality, trends, and noise, which are prominent in Rose wine sales data. This model is the best choice for forecasting in this scenario.
- Both ARIMA models have significantly higher RMSE compared to SARIMA models. These models lack the capability to handle seasonality effectively, which is likely a major factor in their underperformance.
- Seasonality is Crucial: SARIMA models outperform ARIMA models due to their ability to account for seasonality in the data.
- Auto vs. Manual: The manually tuned SARIMA model has a slight edge over Auto SARIMA, suggesting that manual tuning based on domain knowledge and diagnostics (e.g., ACF/PACF) can improve performance.

3.7. Compare the performance of the models

	Test RMSE
RegressionOnTime	1391.708631
SimpleAverage	1368.774051
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.03,SimpleExponentialSmoothing	1362.428949
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999
Alpha=0.0757,Beta=0.0648,Gamma=0.3765,TripleExponentialSmoothing	381.657232
Alpha=0.7,Beta=0.4,Gamma=0.3,TripleExponentialSmoothing	422.908833
AutoARIMA	1374.664628
ARIMA	1325.166743
Auto SARIMA	426.961615
SARIMA	412.824322

Figure 43: Model RMSE

Model	RMSE	Performance
Triple Exponential Smoothing (Alpha=0.0757, Beta=0.0648, Gamma=0.3765)	381.657232	Best model overall for non-stationary data. Captures trends and seasonality well.
SARIMA	412.824322	Best for stationary data. Slightly less accurate but robust for stationary time series.
Auto SARIMA	426.961615	Close to SARIMA; good for automated setups.
Triple Exponential Smoothing (Alpha=0.7, Beta=0.4, Gamma=0.3)	422.908833	Strong performance for non-stationary data, slightly worse than the best triple exponential.
2-point Trailing Moving Average	811.178937	Better than other moving averages but lacks seasonality modeling.
ARIMA	1325.166743	Outperformed by SARIMA models due to seasonality issues.
Auto ARIMA	1374.664628	Poor performance due to lack of seasonality handling.
Other Simple Models (Regression, Simple Average)	>1300 RMSE	Simple models fail to account for trends and seasonality.

Table 1: Model Performance – Sparkling Wine Dataset

3.7.1. Best Model based on RMSE- Triple Exponential Smoothing (Holt's Winter)

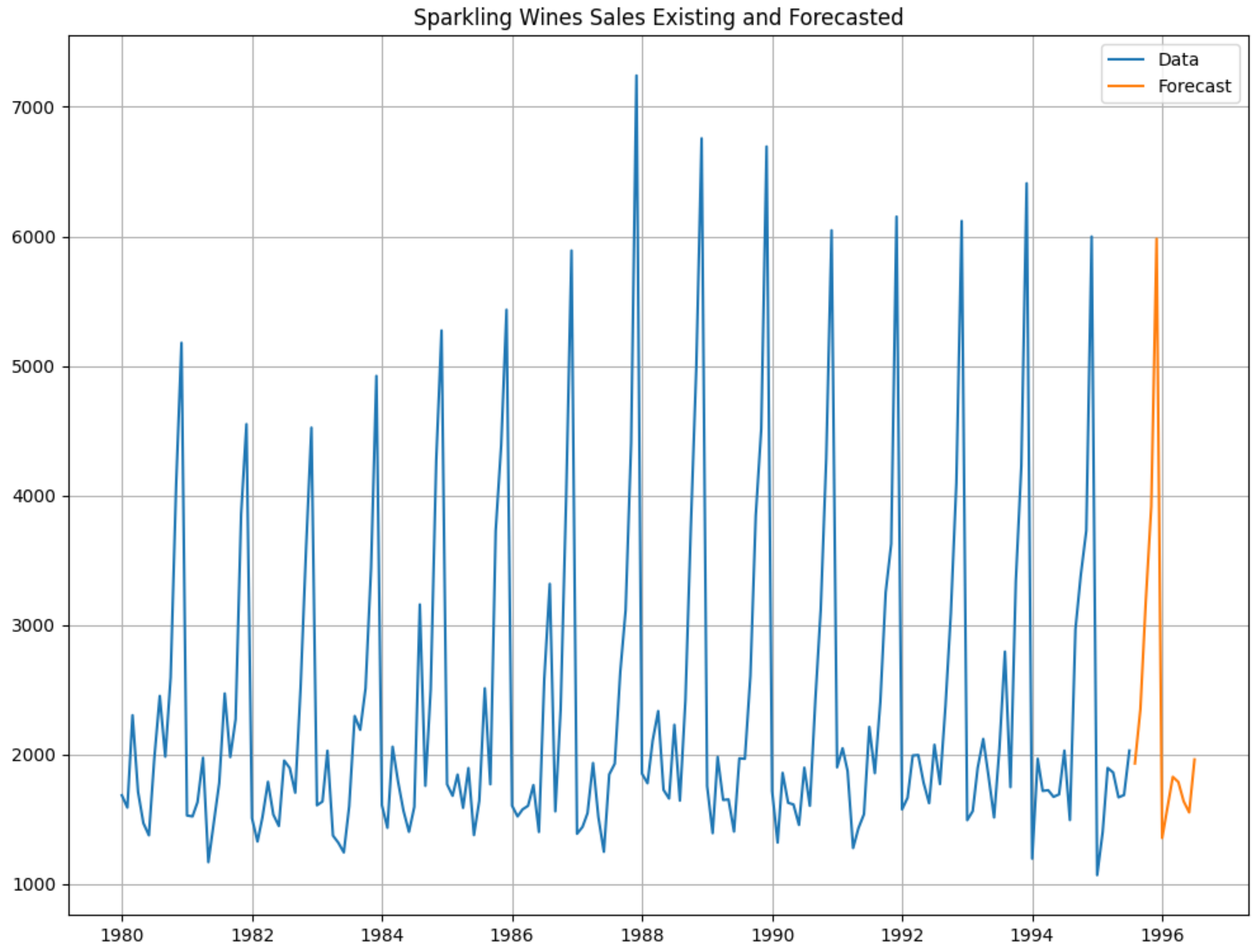


Figure 44: Holt's WInter Future Forecast

3.8. Actionable Insights and Recommendations

<u>Benefits using the Holt's Winter Model</u>

- Captures Seasonality and Trends Effectively: The triple exponential smoothing model incorporates both seasonal patterns and long-term trends, making it highly effective in forecasting wine sales with predictable peaks and dips. This allows the business to align marketing, inventory, and operations with demand fluctuations.
- Improved Forecast Accuracy: With the lowest RMSE compared to other models, this model minimizes forecast errors, ensuring more reliable and precise predictions. Accurate forecasting helps in optimizing supply chain management and reducing inventory costs.
- Data-Driven Decision-Making: By accurately predicting high-demand periods (e.g., holiday season) and low-demand months, ABC Estate Wines can make proactive business decisions, such as launching seasonal promotions, managing production schedules, and adjusting staffing requirements.
- Resource Optimization: The model's ability to predict sales dips during the early months (January–February) allows for better resource allocation, such as reducing production during low-demand periods and focusing efforts on other strategic activities.

- Strategic Marketing Insights: Insights derived from this model can guide personalized and time-sensitive campaigns. For example, leveraging predicted peaks in November–December for targeted holiday promotions can maximize ROI on marketing investments.
- Supports Long-Term Planning: The model helps in identifying long-term growth trends, which is crucial for strategic planning in production expansion, market diversification, or new product launches.
- Better Customer Experience: With reliable sales forecasts, ABC Estate Wines can ensure timely availability of popular products, thereby improving customer satisfaction and loyalty.
- Cost Savings Through Inventory Management: By predicting demand accurately, the business can reduce costs associated with overstocking or stockouts, such as storage fees, spoilage (for perishable products), or lost sales opportunities.
- Adapts to Changing Dynamics: The triple exponential smoothing model's inclusion of a smoothing parameter for seasonality (gamma) allows the model to adapt dynamically to changing customer preferences or market conditions, maintaining forecast accuracy over time.
- Enhances Competitive Edge: With a data-driven approach to forecasting, ABC Estate Wines can stay ahead of competitors by responding swiftly to demand changes, introducing innovative campaigns, and consistently meeting customer expectations.

Insights

- The forecast highlights significant seasonality in wine sales, with peaks around the holiday season (November–December) and dips in the early months of the year (January–February).
- The early months of the year show a considerable decline in sales, which can impact overall revenue.
- Seasonal demand may also indicate varying levels of customer engagement throughout the year.
- The triple exponential smoothing model achieved a lower RMSE, indicating it captures seasonality and trends effectively. However, further improvements are possible.
- Trends may be influenced by external factors like climate, market preferences, or economic conditions.

Recommendations

- Focus on targeted marketing campaigns and promotions during the holiday season to capitalize on increased demand.
- Ensure adequate inventory levels to meet the peak season demand without shortages or overstocking.
- Introduce attractive discounts or bundle offers in January and February to stimulate sales.
- Launch a rewards program to retain high-value customers and encourage repeat purchases.

4. Rose Wine Dataset

4.1. Data Overview

4.1.1. Import libraries and load the data

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

Figure 45: Rose Wine Data Overview

4.1.2. Check the structure of data

Shape of the dataset: 187 rows and 2 columns

4.1.3. Check the types of the data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   YearMonth  187 non-null    object
 1   Rose       185 non-null    float64
dtypes: float64(1), object(1)
memory usage: 3.0+ KB
```

Figure 46: Datatypes

4.1.4. Check for and treat (if needed) missing values

	0
YearMonth	0
Rose	2

Figure 47: Missing values check

Interpolate the missing values with linear values. Smoothing out missing values in a logical, linear manner

4.1.5. Data Duplicates

There are no duplicate rows.

4.1.6. Statistical Summary

	count	mean	std	min	25%	50%	75%	max
Rose	187.0	89.914439	39.238325	28.0	62.5	85.0	111.0	267.0

Figure 48: Statistical Summary - Numeric

4.1.7. Insights

- The standard deviation (39.24) is approximately 43.6% of the mean (89.91), indicating moderate variability in sales.
- The mean (89.91) is slightly higher than the median (85.0), suggesting a slight positive skew in the data. This indicates the presence of some higher sales values, though the skewness is not extreme.
- The 75th percentile (111.0) and maximum value (267.0) highlight periods of elevated sales, which could reflect increased demand during certain times. These insights could help in planning targeted marketing strategies or promotions during similar periods.

4.2. Exploratory Data Analysis

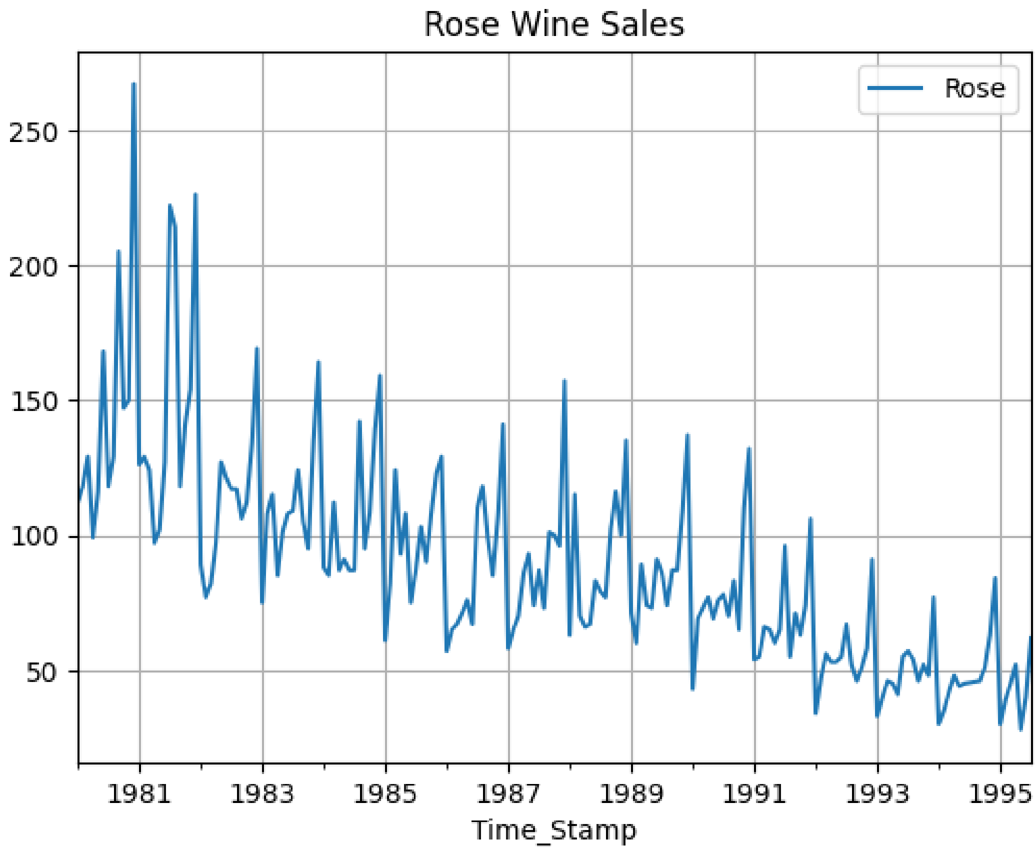


Figure 49: Time Series Plot

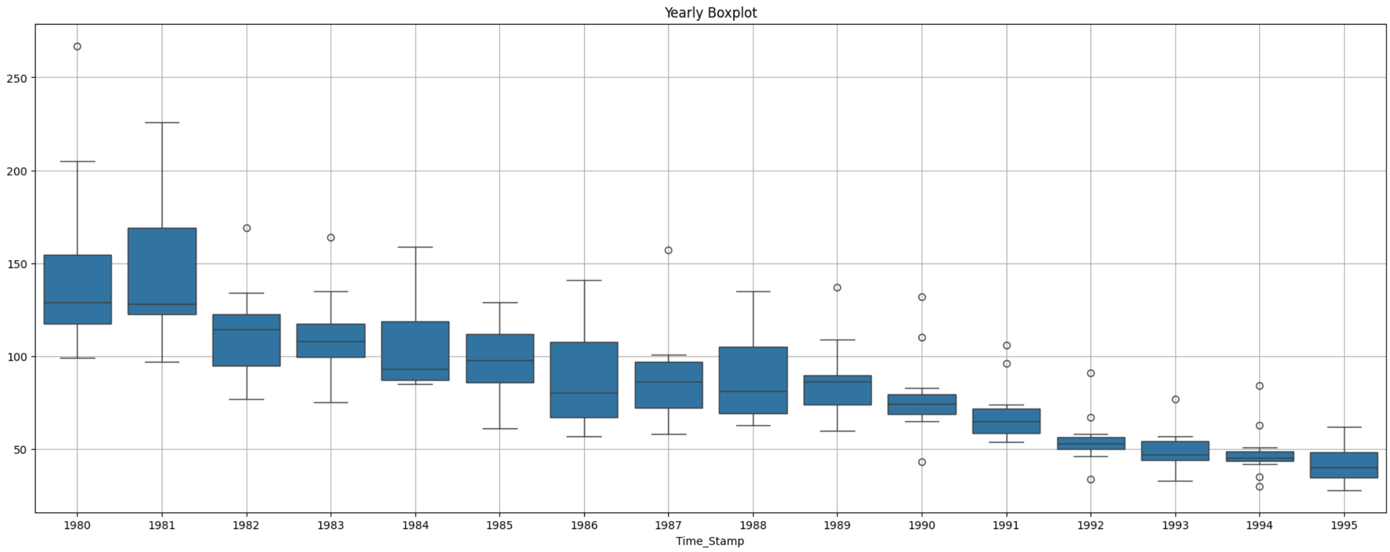


Figure 50: Sales across different years

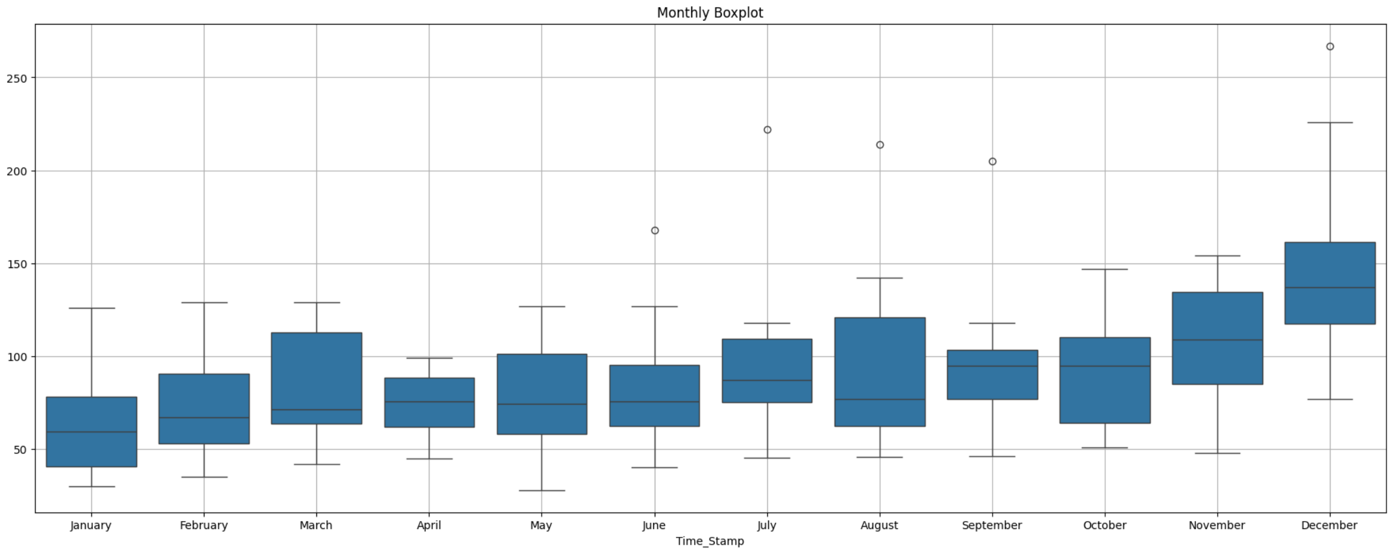


Figure 51: Sales within different months across years

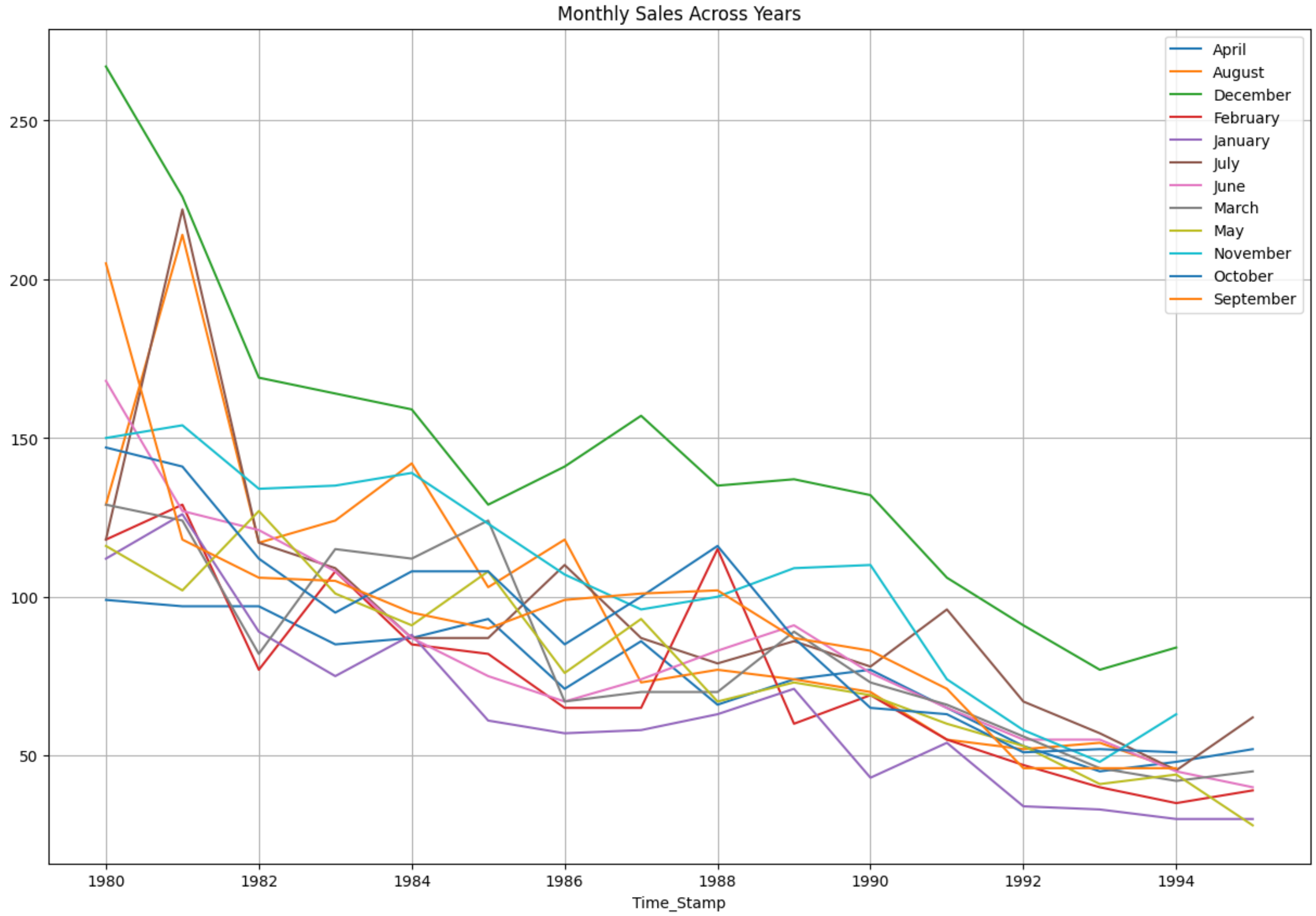


Figure 52: Monthly Sales

Insights

- 1980–1988: Sales show high variability, with peaks and troughs across months.
- 1989–1995: A steady decline is observed, especially in later years, indicating a potential reduction in demand for rose wine during this period.
- December consistently has the highest sales values across the years, likely indicating seasonal demand during the holiday period. For example, sales in 1980 (267) and 1981 (226) are significantly higher than other months.
- February, January, and March also exhibit relatively higher sales compared to mid-year months, hinting at strong performance during winter and early spring.
- December sales drop from 106 (1991) to 84 (1994) and August sales fall sharply, from 55 (1991) to 45.67 (1994).

4.2.1. Time series Decomposition

Additive

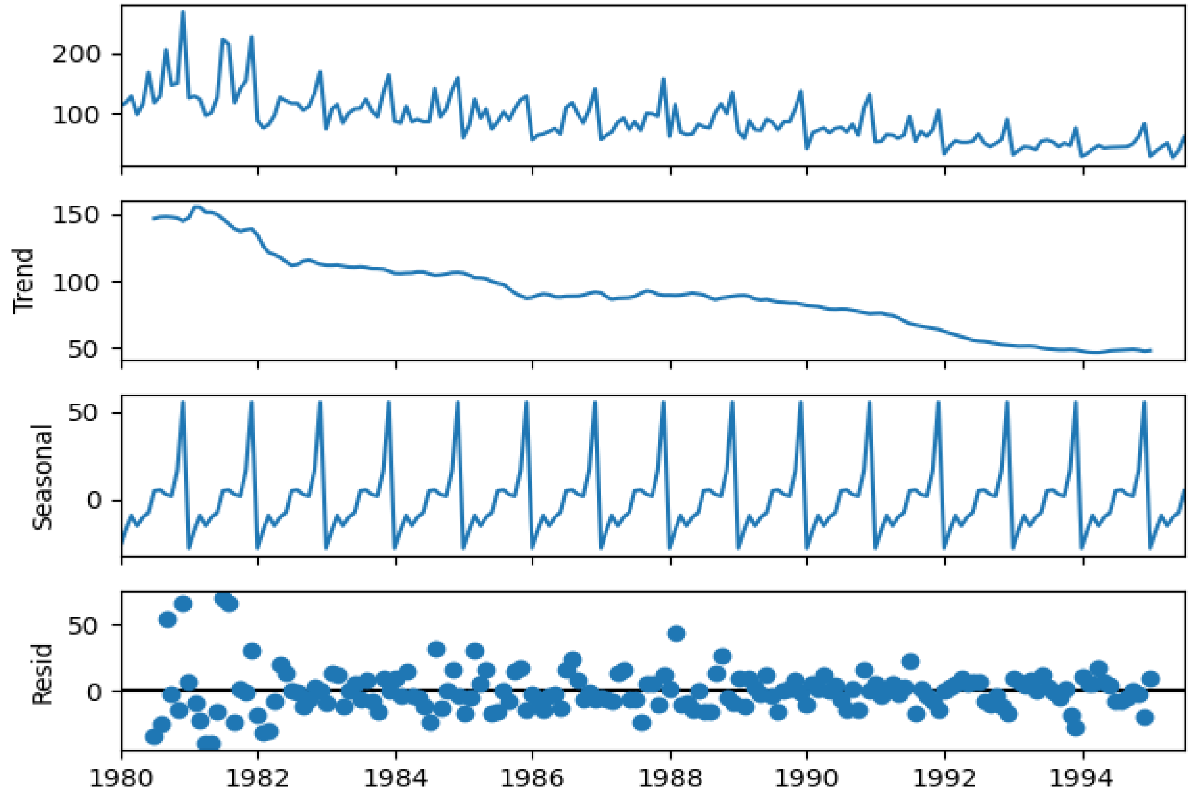


Figure 53: Additive

```
Trend
 Time_Stamp
1980-01-01       NaN
1980-02-01       NaN
1980-03-01       NaN
1980-04-01       NaN
1980-05-01       NaN
1980-06-01       NaN
1980-07-01    147.08
1980-08-01    148.12
1980-09-01    148.37
1980-10-01    148.08
1980-11-01    147.42
1980-12-01    145.12
Name: trend, dtype: float64

Seasonality
 Time_Stamp
1980-01-01   -27.91
1980-02-01   -17.44
1980-03-01    -9.29
1980-04-01   -15.10
1980-05-01   -10.20
1980-06-01    -7.68
1980-07-01     4.90
1980-08-01     5.50
1980-09-01     2.77
1980-10-01     1.87
1980-11-01    16.85
1980-12-01    55.71
Name: seasonal, dtype: float64

Residual
 Time_Stamp
1980-01-01      NaN
1980-02-01      NaN
1980-03-01      NaN
1980-04-01      NaN
1980-05-01      NaN
1980-06-01      NaN
1980-07-01   -33.98
1980-08-01   -24.62
1980-09-01    53.85
1980-10-01    -2.96
1980-11-01   -14.26
1980-12-01    66.16
Name: resid, dtype: float64
```

Figure 54: Additive Components

Multiplicative

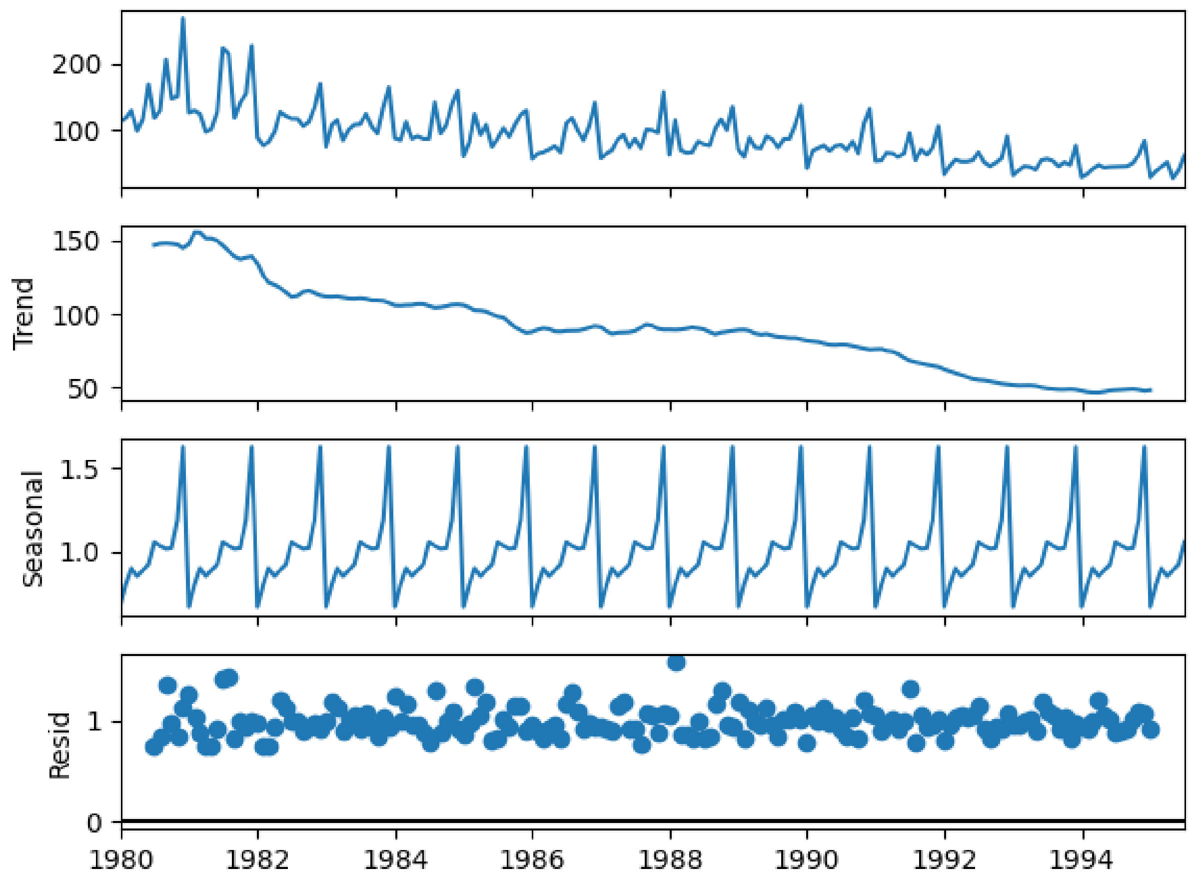


Figure 55: Multiplicative

```
Trend
 Time_Stamp
1980-01-01        NaN
1980-02-01        NaN
1980-03-01        NaN
1980-04-01        NaN
1980-05-01        NaN
1980-06-01        NaN
1980-07-01     147.08
1980-08-01     148.12
1980-09-01     148.37
1980-10-01     148.08
1980-11-01     147.42
1980-12-01     145.12
Name: trend, dtype: float64

Seasonality
 Time_Stamp
1980-01-01     0.67
1980-02-01     0.81
1980-03-01     0.90
1980-04-01     0.85
1980-05-01     0.89
1980-06-01     0.92
1980-07-01     1.06
1980-08-01     1.04
1980-09-01     1.02
1980-10-01     1.02
1980-11-01     1.19
1980-12-01     1.63
Name: seasonal, dtype: float64

Residual
 Time_Stamp
1980-01-01      NaN
1980-02-01      NaN
1980-03-01      NaN
1980-04-01      NaN
1980-05-01      NaN
1980-06-01      NaN
1980-07-01     0.76
1980-08-01     0.84
1980-09-01     1.36
1980-10-01     0.97
1980-11-01     0.85
1980-12-01     1.13
Name: resid, dtype: float64
```

Figure 56: Multiplicative Components

Insights

- It appears that the multiplicative decomposition may be a better fit for this data. The seasonal fluctuations seem to have a stronger relationship with the level of the series, which is characteristic of a multiplicative model.
- In the multiplicative model, the seasonal variations grew/shrank proportionally with the trend, while in this additive model, the seasonal variations are constant over time.
- The additive model assumes constant seasonal effects and residuals in absolute terms, which does not align with the data.

4.3. Data Preprocessing

4.3.1. Missing Value treatment

Interpolate the missing values with linear values. Smoothing out missing values in a logical, linear manner

4.3.2. Duplicate value check

There are no duplicate rows.

4.3.3. Train – Test Split

Split on 70:30 ratio.

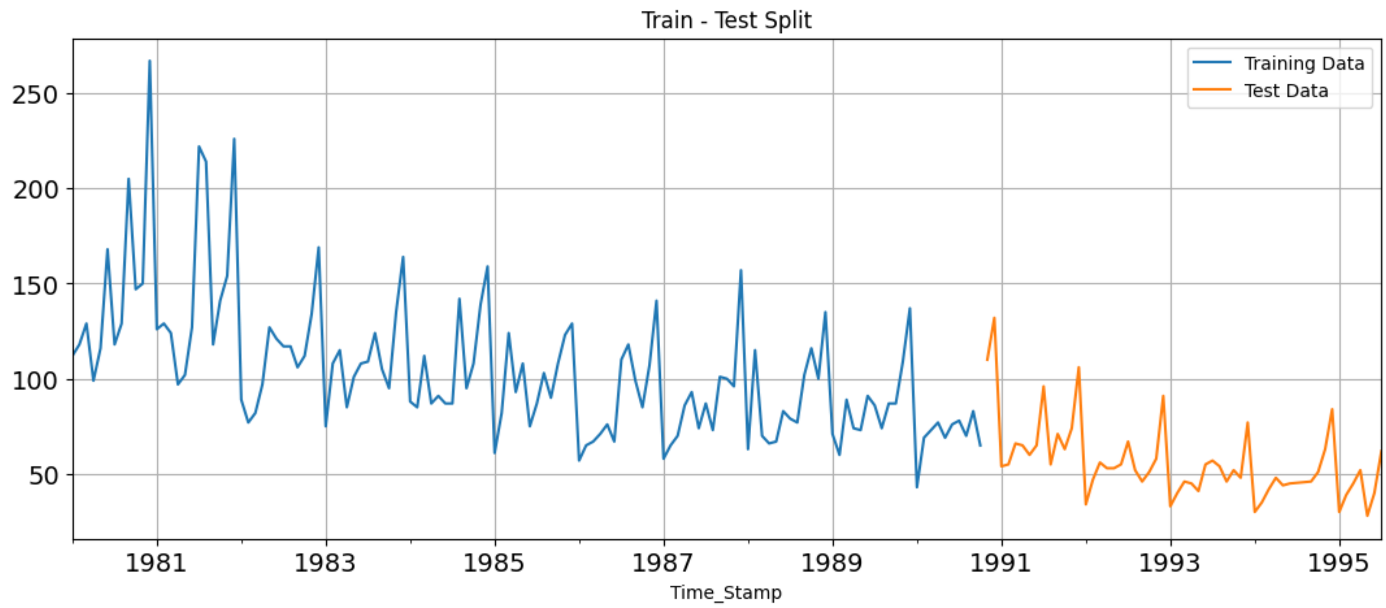


Figure 57: Train Test Split

4.4. Model Building- Original Data

4.4.1. Linear Regression

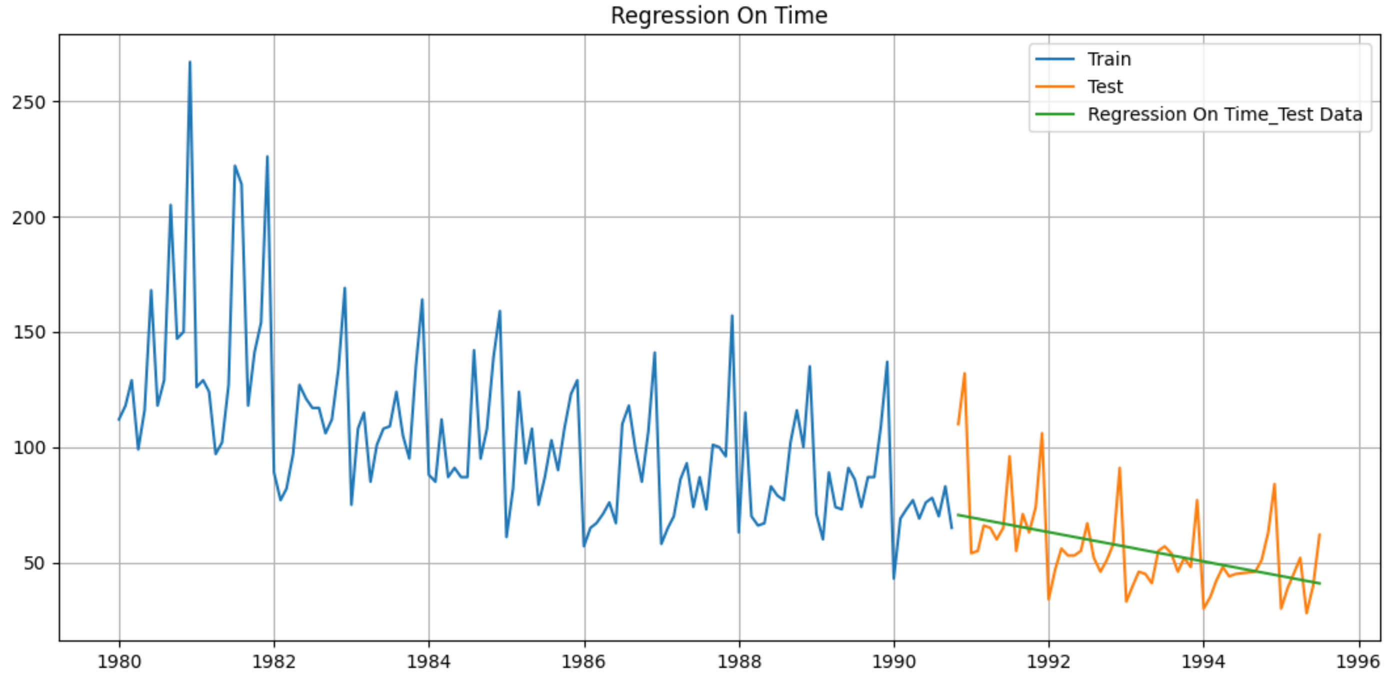


Figure 58: Linear Regression

For Linear regression forecast on the Test Data, RMSE is 17.33

4.4.2. Simple Average

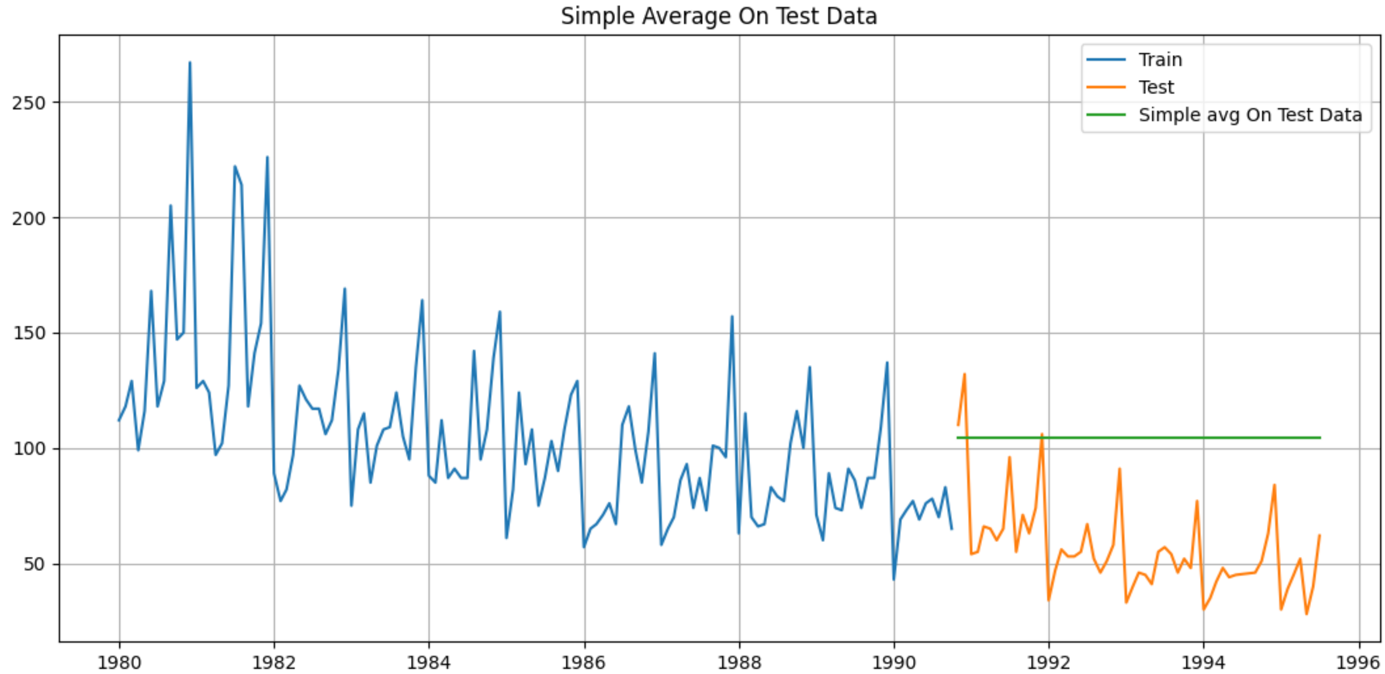


Figure 59: Simple Average

For Simple Average forecast on the Test Data, RMSE is 52.41

4.4.3. Moving Average

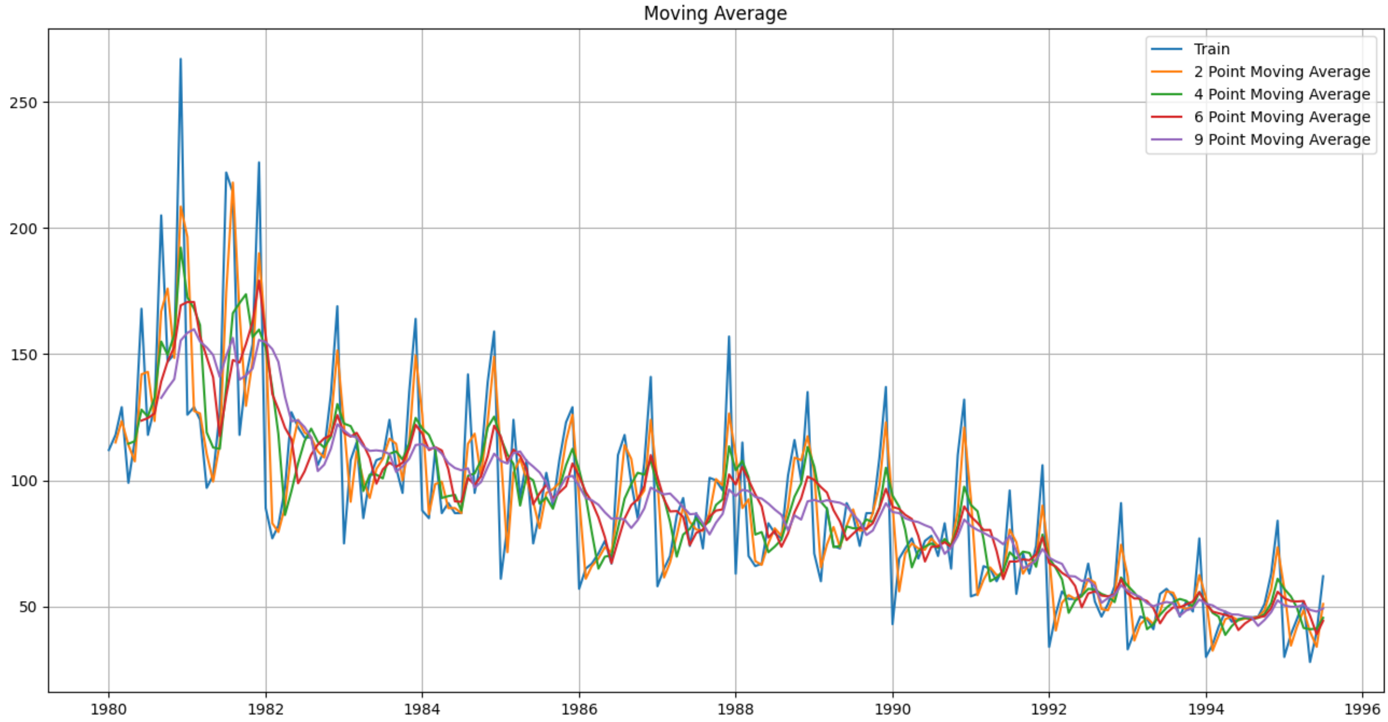


Figure 60: Moving Average

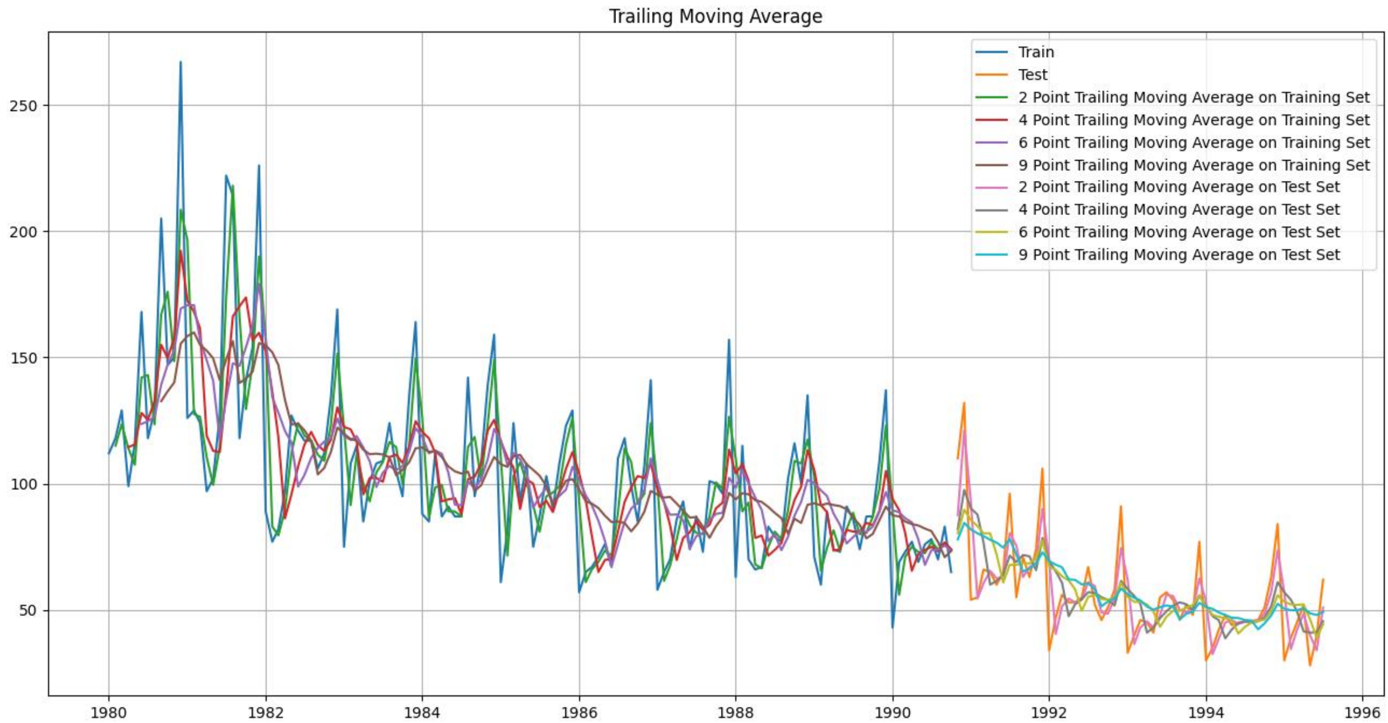


Figure 61: Trailing Moving Average

For 2 point Moving Average Model forecast on the Training Data, RMSE is 11.801

For 4 point Moving Average Model forecast on the Training Data, RMSE is 15.367

For 6 point Moving Average Model forecast on the Training Data, RMSE is 15.862

For 9 point Moving Average Model forecast on the Training Data, RMSE is 16.342

4.4.4. Simple Exponential Smoothing

```
{'smoothing_level': 0.1277774057492626,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 112.0,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 62: SES Model Params

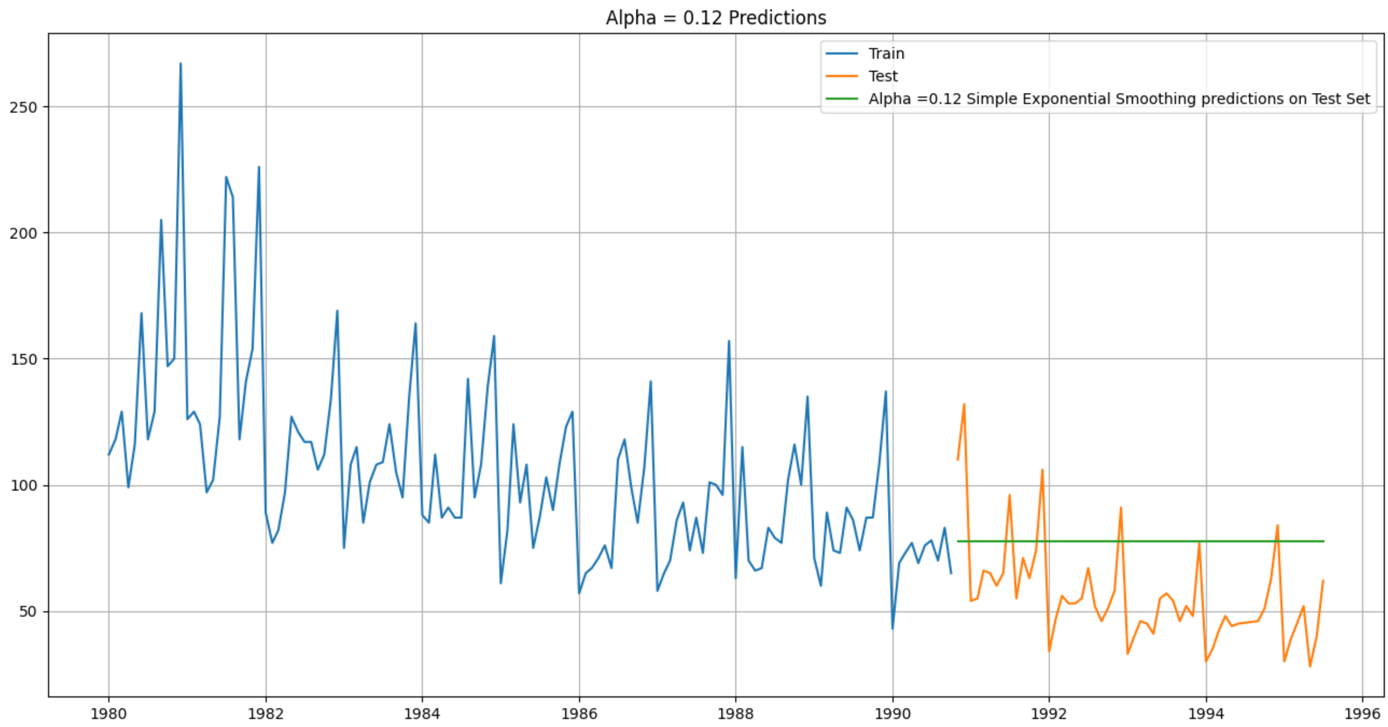


Figure 63: SES Model

For Alpha = 0.12 Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 29.224

Different Alpha values

Alpha Values	Train RMSE	Test RMSE
0.3	32.292266	26.310348
0.4	32.893017	25.657764
0.5	33.578304	25.109604
0.6	34.372651	24.529811
0.7	35.288467	23.894929
0.8	36.330954	23.212777
0.9	37.507371	22.496819

Figure 64: RMSE for different alpha values

Chose Alpha = 0.9 based on Test RMSE

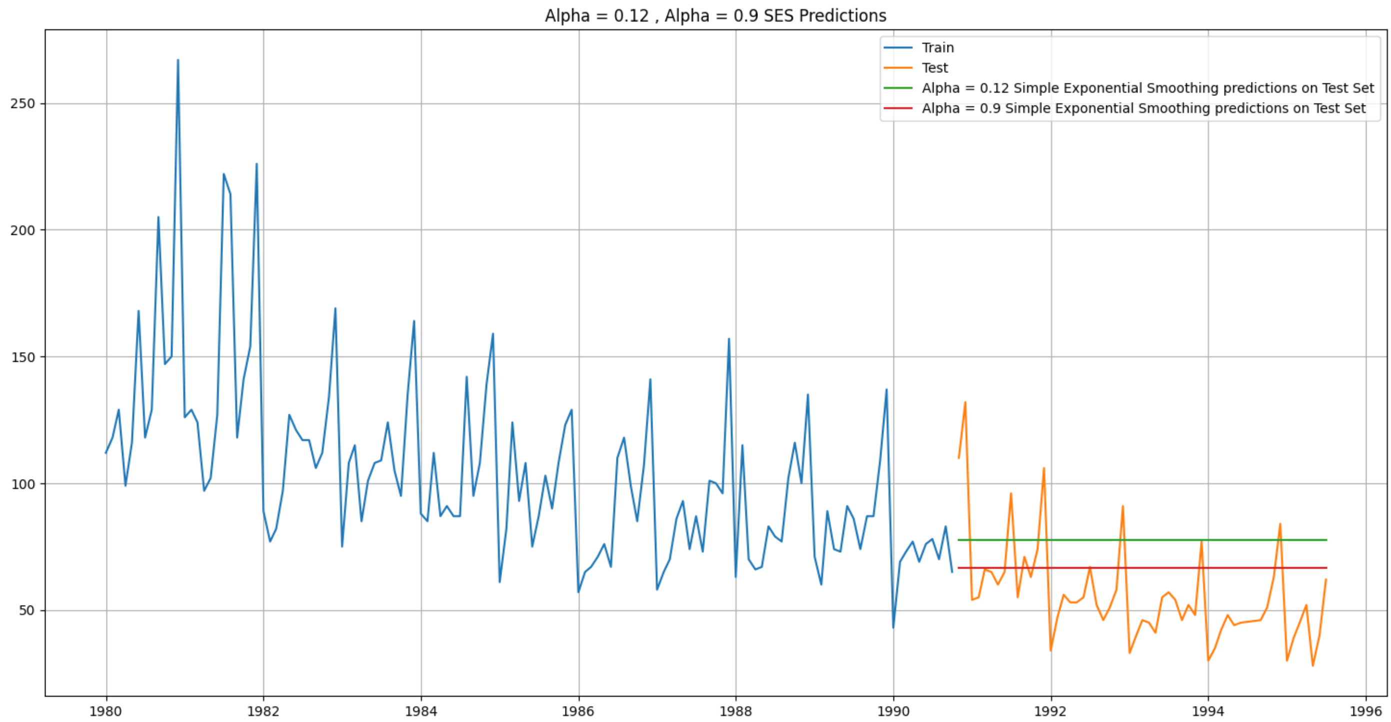


Figure 65: SES Predictions for different alpha values

4.4.5. Double Exponential Smoothing – Holt's Model

Computed RMSE for different Alpha and Beta values.

Alpha Values	Beta Values	Train RMSE	Test RMSE
0.3	0.4	37.287813	18.343250
0.4	0.7	40.744796	18.975318
0.4	0.4	37.990913	19.133156
0.5	0.4	38.598226	19.197151
0.4	0.3	36.682435	19.769770

Figure 66: DES Alpha Beta RMSE

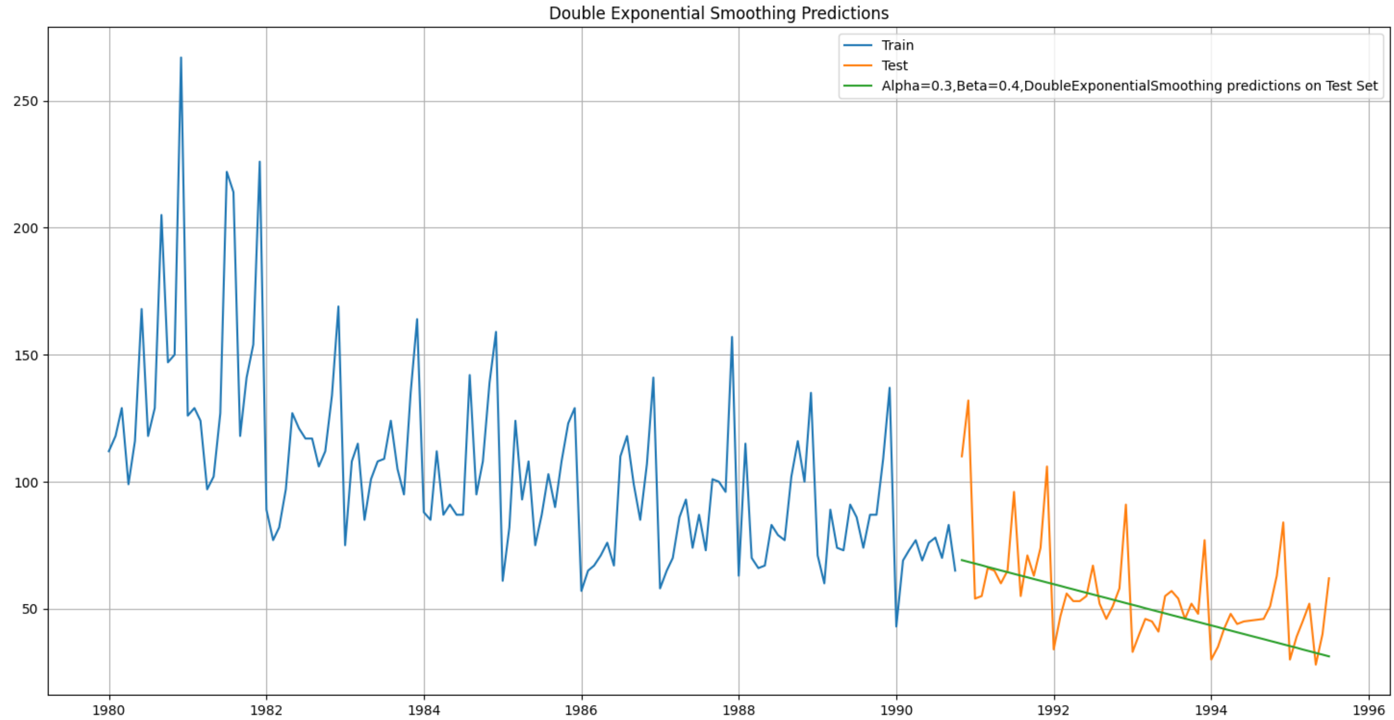


Figure 67: DES Predictions

4.4.6. Triple Exponential Smoothing – Holt's Winter

```
{'smoothing_level': 0.0999080139189177,
 'smoothing_trend': 1.9932826568022853e-06,
 'smoothing_seasonal': 0.00017683239767298466,
 'damping_trend': nan,
 'initial_level': 109.16836143052193,
 'initial_trend': -0.44137924420686336,
 'initial_seasons': array([1.0049411 , 1.13565754, 1.2416344 , 1.08896356, 1.2223928 ,
        1.31686195, 1.44959601, 1.55043078, 1.45169973, 1.42782318,
        1.64159637, 2.26353792]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 68: TES params

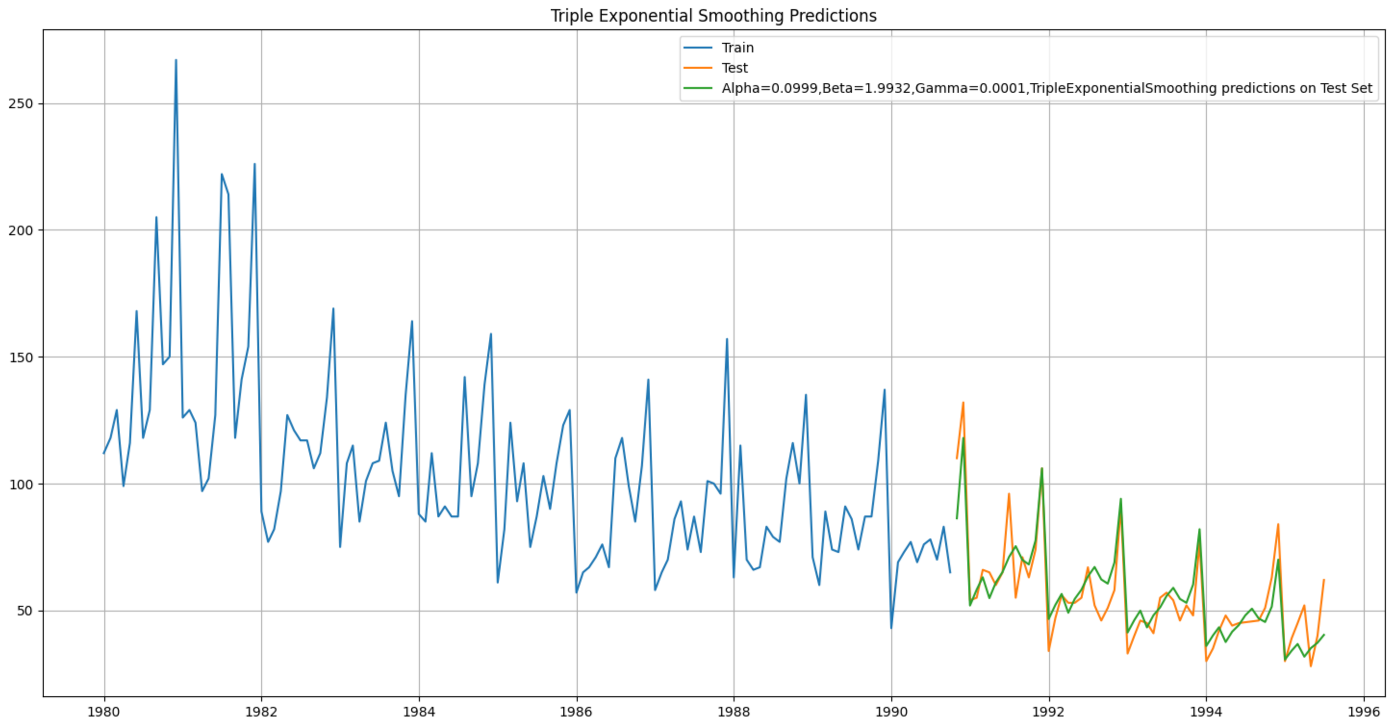


Figure 69: TES Predictions

For Alpha=0.0999,Beta=1.9932,Gamma=0.0001, Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 9.329

Computed RMSE for different Alpha, Beta and Gamma values.

Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
0.3	0.7	0.4	29.968505	28.362899
0.5	0.9	0.4	41.232290	28.802668
0.3	0.6	0.4	27.743621	39.656865
0.4	0.4	0.9	43.001123	51.375209
0.5	0.3	1.0	47.353331	62.161097

Figure 70: RMSE Alpha Beta Gamma Values

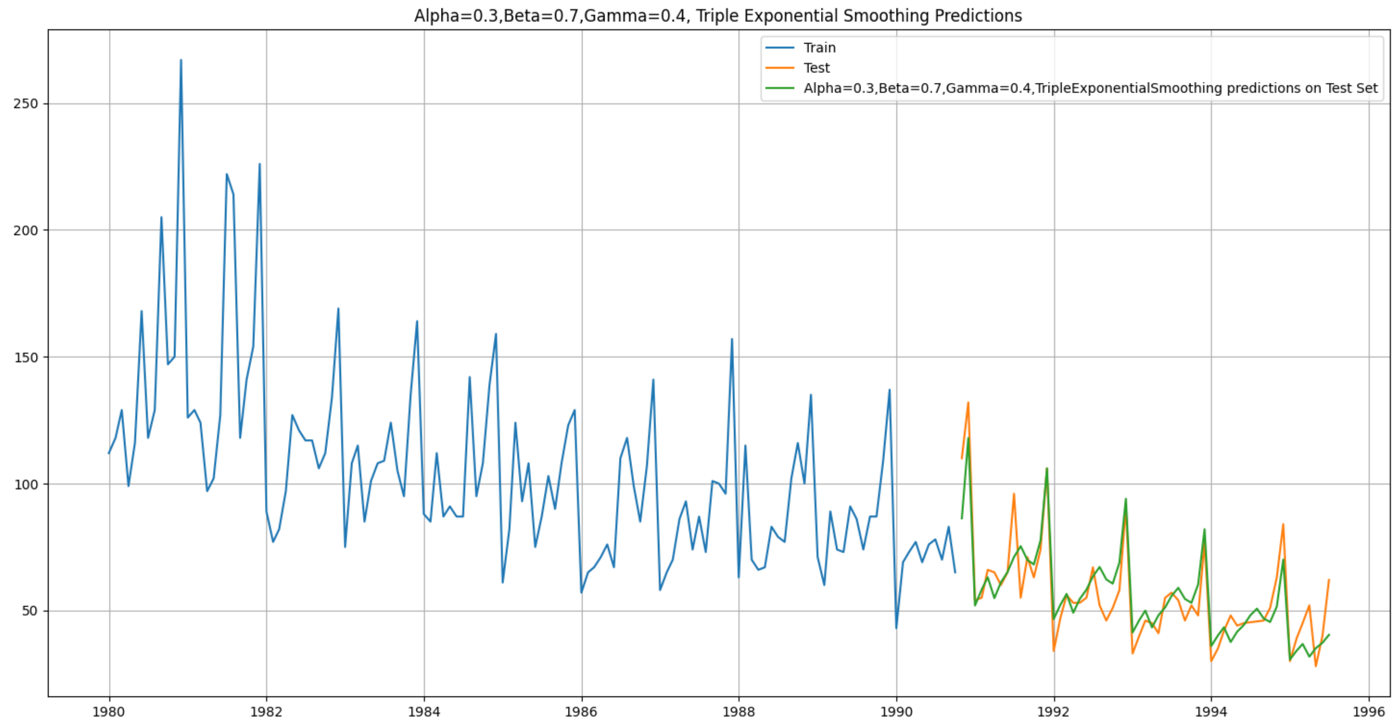
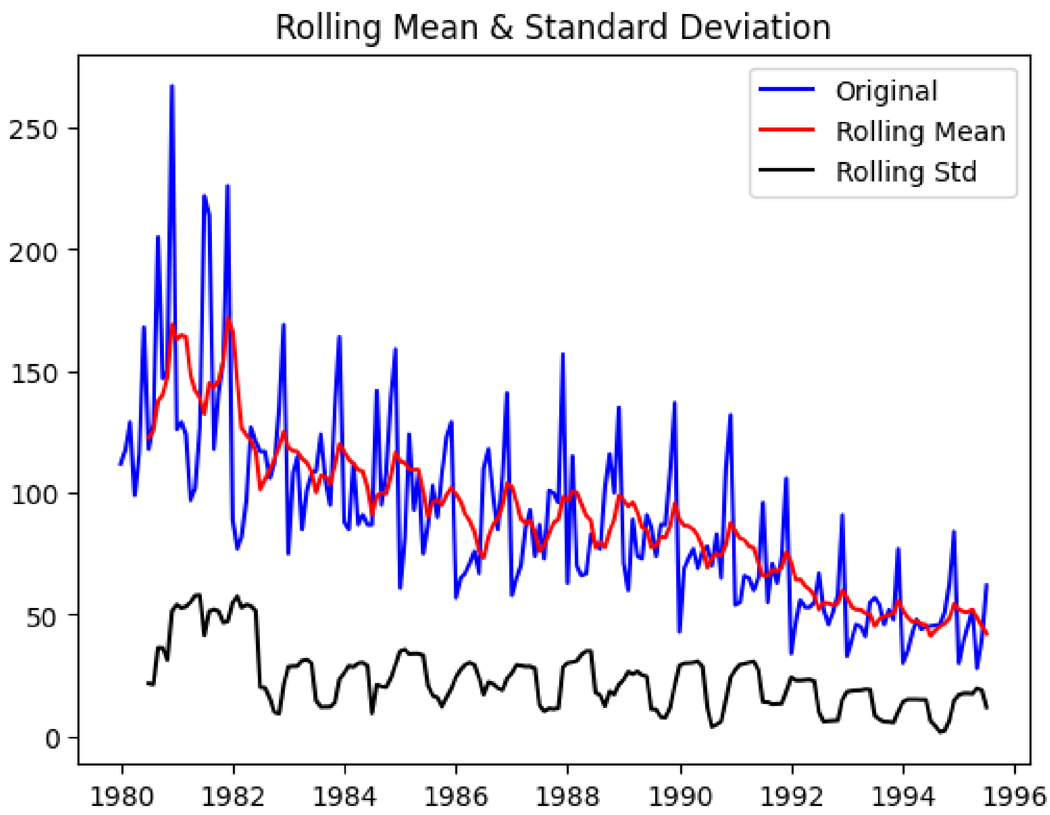


Figure 71: TES Prediction for different params

Insights

- The Triple Exponential Smoothing model with parameters Alpha=0.0999, Beta=1.9932, Gamma=0.0001 achieved the lowest RMSE of 9.328733, indicating it is the most accurate model for forecasting rose wine sales. This model effectively captures seasonality and trends, making it the best choice for planning and decision-making.
- The 2-point trailing moving average model performed better than the 4-point, 6-point, and 9-point moving averages, achieving an RMSE of 11.801043. While it provides a straightforward forecasting method, it lacks the ability to adapt to seasonal patterns, making it less suitable compared to triple exponential smoothing.
- The Double Exponential Smoothing model (RMSE: 18.343250) performed better than both Simple Exponential Smoothing models (RMSE: 22.496819 and 29.223677). However, it is outperformed by models that account for seasonality, such as triple exponential smoothing.
- Models with higher RMSE values, such as Simple Average and the less optimized Triple Exponential Smoothing (Alpha=0.3, Beta=0.7, Gamma=0.4), fail to capture trends and seasonality effectively, leading to less accurate forecasts.

4.5. Check for stationarity




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.876699
p-value                          0.343101
#Lags Used                      13.000000
Number of Observations Used    173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
dtype: float64
```

Figure 72: Dickey Fuller Test

The results of the Dickey-Fuller Test indicate that the time series is still non-stationary. Here's the detailed interpretation:

Key Results

Test Statistic = -1.876699

This value is not smaller than the critical values at the 1%, 5%, or 10% significance levels, meaning we fail to reject the null hypothesis that the series has a unit root (i.e., it is non-stationary).

p-value = 0.343101

Since the p-value is much greater than 0.05 (or 0.01 for stricter tests), we do not have enough evidence to reject the null hypothesis.

Critical Values:

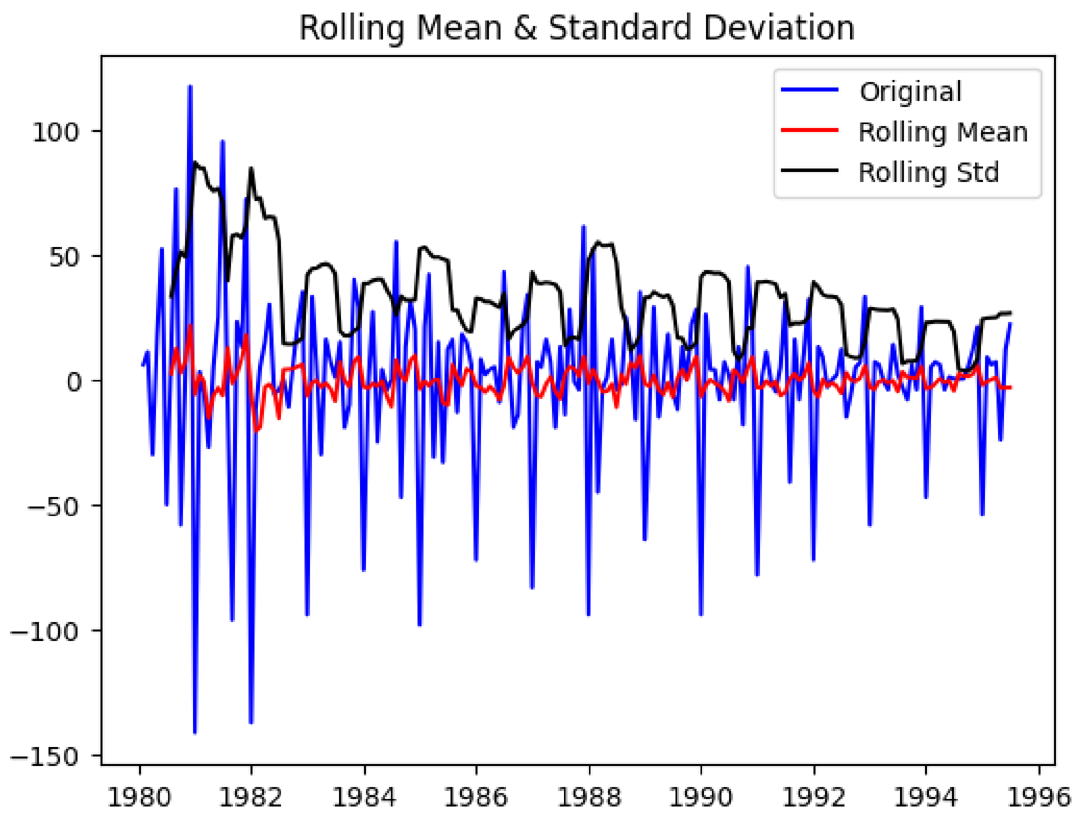
1% level = -3.468726

5% level = -2.878396

10% level = -2.575756

Since the test statistic (-1.876699) is greater than all three critical values, it reinforces that the series is non-stationary.

Differencing the Series




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

Figure 73: Dickey Fuller Test After differencing

The results of the Dickey-Fuller Test indicate that the time series is now stationary after differencing. Here's the detailed interpretation:

Key Results

Test Statistic = -8.044392

This value is much smaller than the critical values at 1%, 5%, and 10% significance levels. This suggests that we can reject the null hypothesis that the series has a unit root (i.e., it is stationary).

p-value = 1.81e-12 (essentially 0)

Since the p-value is far below 0.05 (or even 0.01 for a stricter test), there is strong evidence to reject the null hypothesis.

Critical Values:

1% level = -3.468726

5% level = -2.878396

10% level = -2.575756

The test statistic (-8.044392) is much smaller than all three critical values, further supporting the conclusion that the series is now stationary.

4.6. Model Building – Stationary Data

4.6.1. ACF Plots

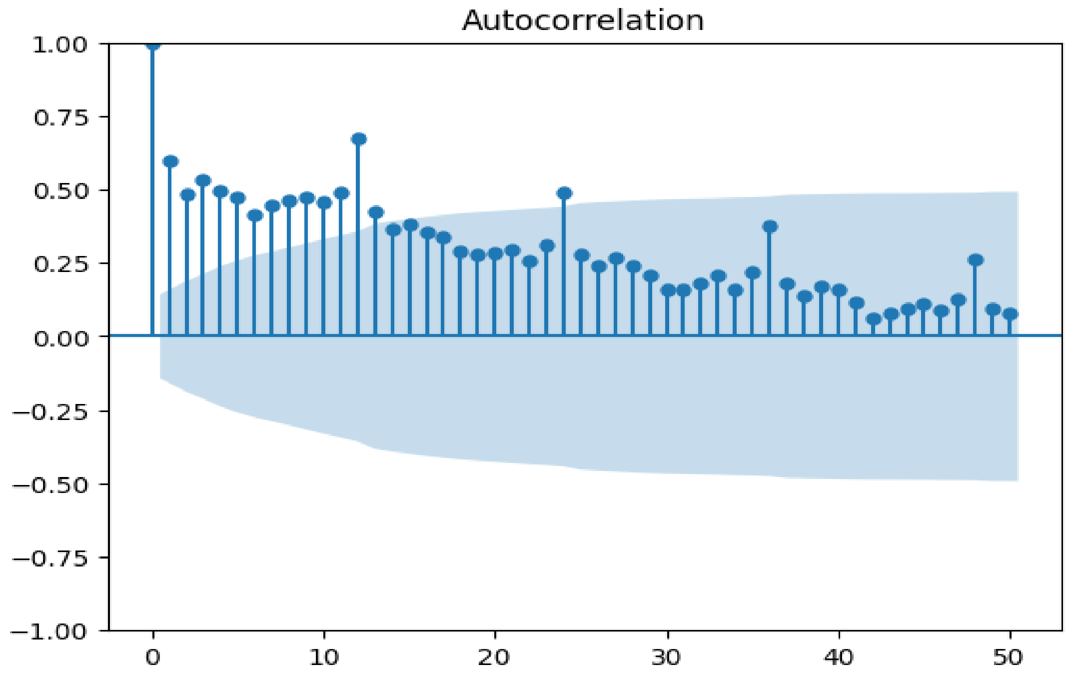


Figure 74: ACF

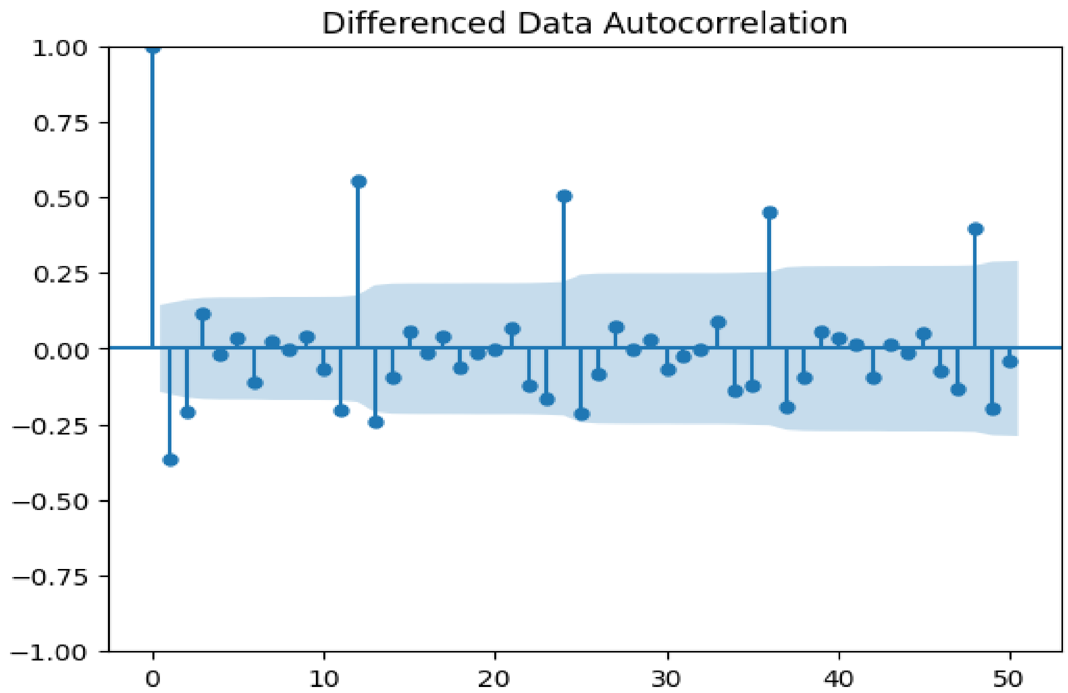


Figure 75: ACF Differenced

Insights

- The original data appears to have some cyclical or periodic patterns, as indicated by the fluctuating autocorrelation values.
- Differencing the data has helped remove any underlying trends or non-stationarity, resulting in a more stable autocorrelation structure.

4.6.2. PACF Plots

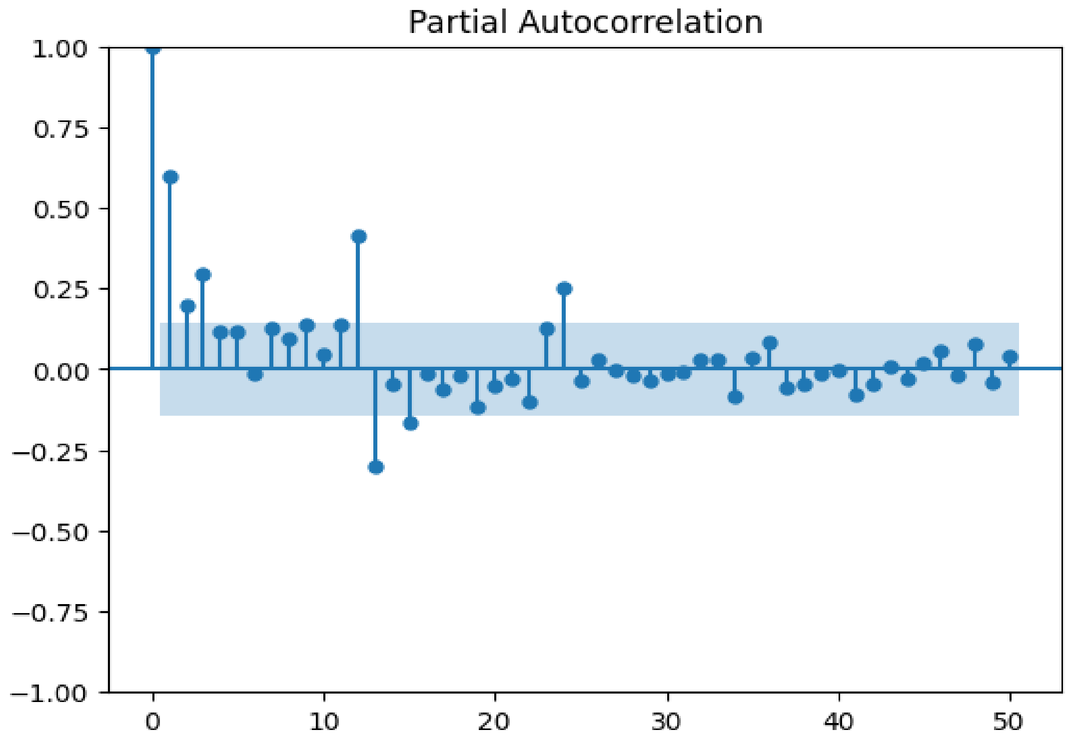


Figure 76: PACF

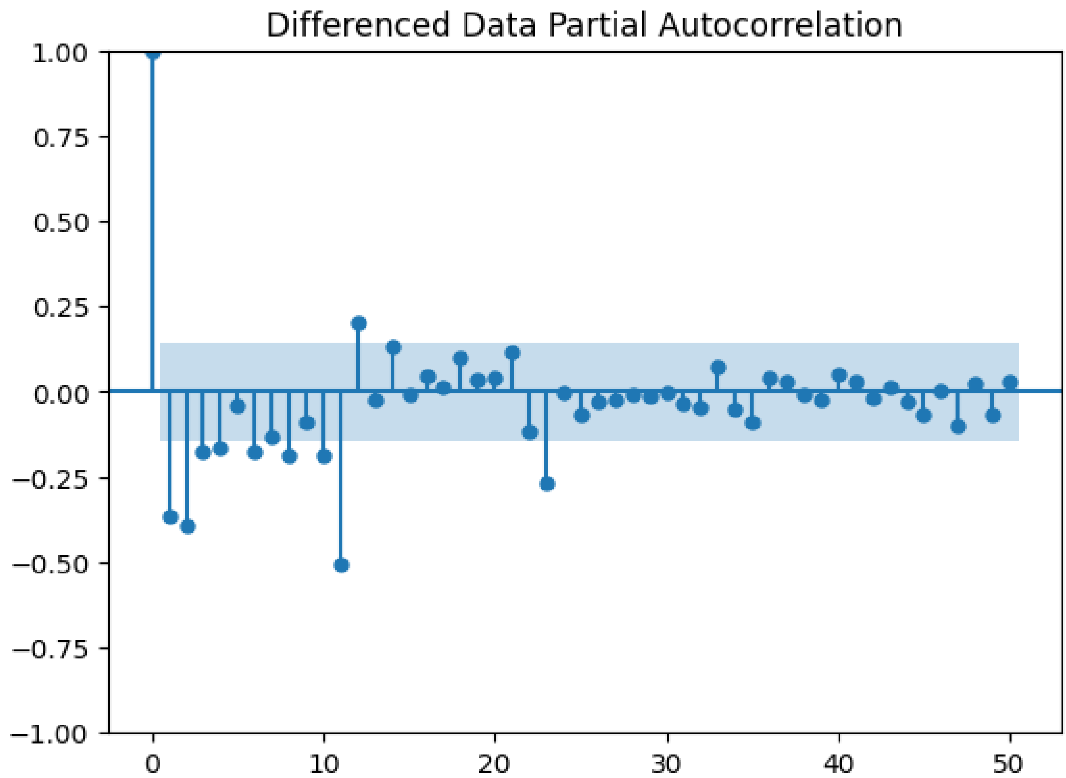


Figure 77: PACF Differenced

Insights

- The original data has an autoregressive structure, as indicated by the significant partial autocorrelation at the early lags.
- Differencing the data has helped remove any non-stationarity, resulting in a more stable PACF structure.

4.6.3. Auto ARIMA

```
Best model:  ARIMA(0,1,2)(0,0,0)[0]
Total fit time: 8.717 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  130
Model:               SARIMAX(0, 1, 2)   Log Likelihood                -626.624
Date:                Sat, 07 Dec 2024   AIC                           1259.248
Time:                        21:29:55   BIC                           1267.827
Sample:                    01-01-1980   HQIC                          1262.734
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7059      0.072     -9.851      0.000      -0.846      -0.565
ma.L2         -0.1915      0.074     -2.574      0.010      -0.337      -0.046
sigma2       958.5998     86.875     11.034      0.000     788.328    1128.872
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):                45.85
Prob(Q):                              0.70   Prob(JB):                         0.00
Heteroskedasticity (H):               0.32   Skew:                             0.88
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.34
===================================================================================
```

Figure 78: Auto ARIMA Summary

For auto_ARIMA forecast on the Test Data, RMSE is 30.90

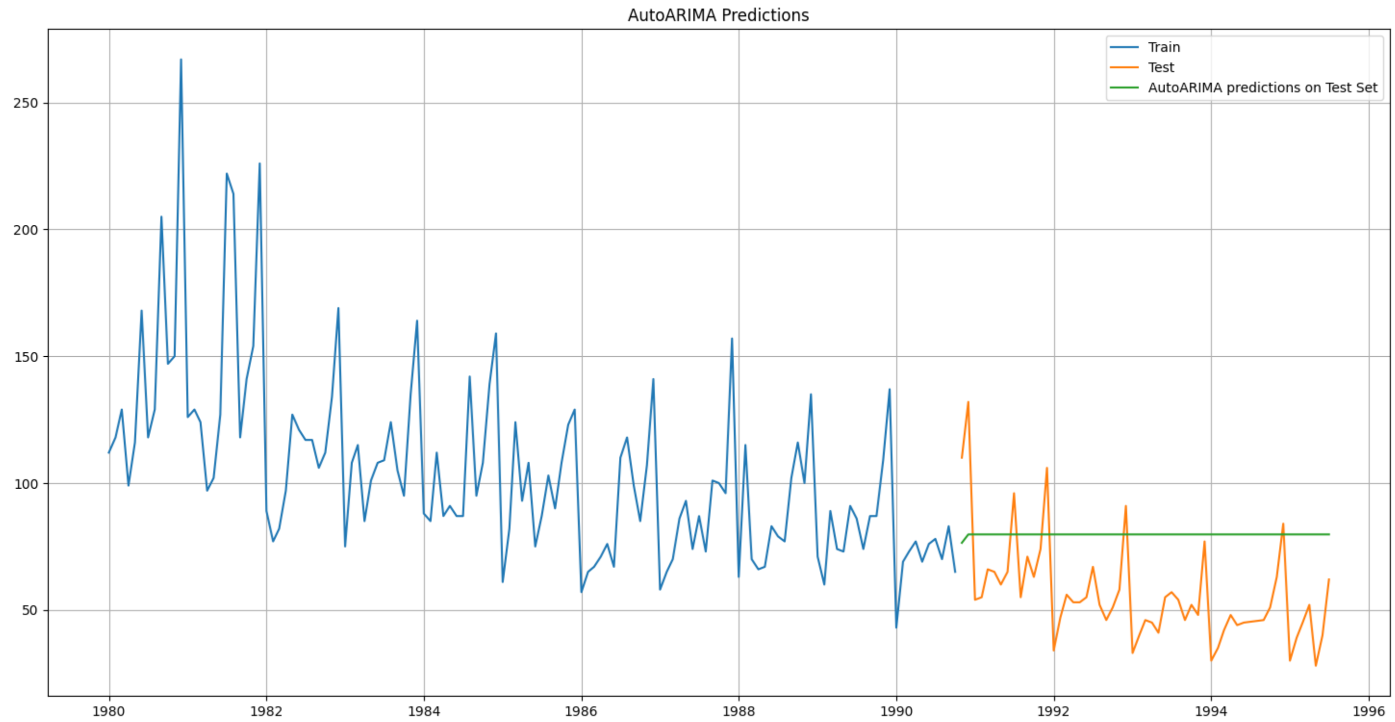


Figure 79: Auto ARIMA Predictions

4.6.4. ARIMA

param	AIC
(0, 1, 2)	1259.247780
(1, 1, 2)	1259.473205
(1, 1, 1)	1260.036763
(2, 1, 1)	1261.014076
(0, 1, 1)	1261.327444
(2, 1, 2)	1261.472001
(2, 1, 0)	1278.135281
(1, 1, 0)	1297.077294
(0, 1, 0)	1313.175861

Figure 80: AIC

ARIMA(0,1,2) has the lowest AIC.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  130
Model:                 ARIMA(0, 1, 2)   Log Likelihood                -626.624
Date:                Sat, 07 Dec 2024   AIC                           1259.248
Time:                        21:29:57   BIC                           1267.827
Sample:                    01-01-1980   HQIC                          1262.734
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7059      0.072     -9.851      0.000      -0.846      -0.565
ma.L2         -0.1915      0.074     -2.574      0.010      -0.337      -0.046
sigma2       958.5998     86.875     11.034      0.000     788.328    1128.872
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):                45.85
Prob(Q):                              0.70   Prob(JB):                         0.00
Heteroskedasticity (H):               0.32   Skew:                             0.88
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.34
===================================================================================
```

Figure 81: ARIMA Summary

For ARIMA forecast on the Test Data, RMSE is 30.90

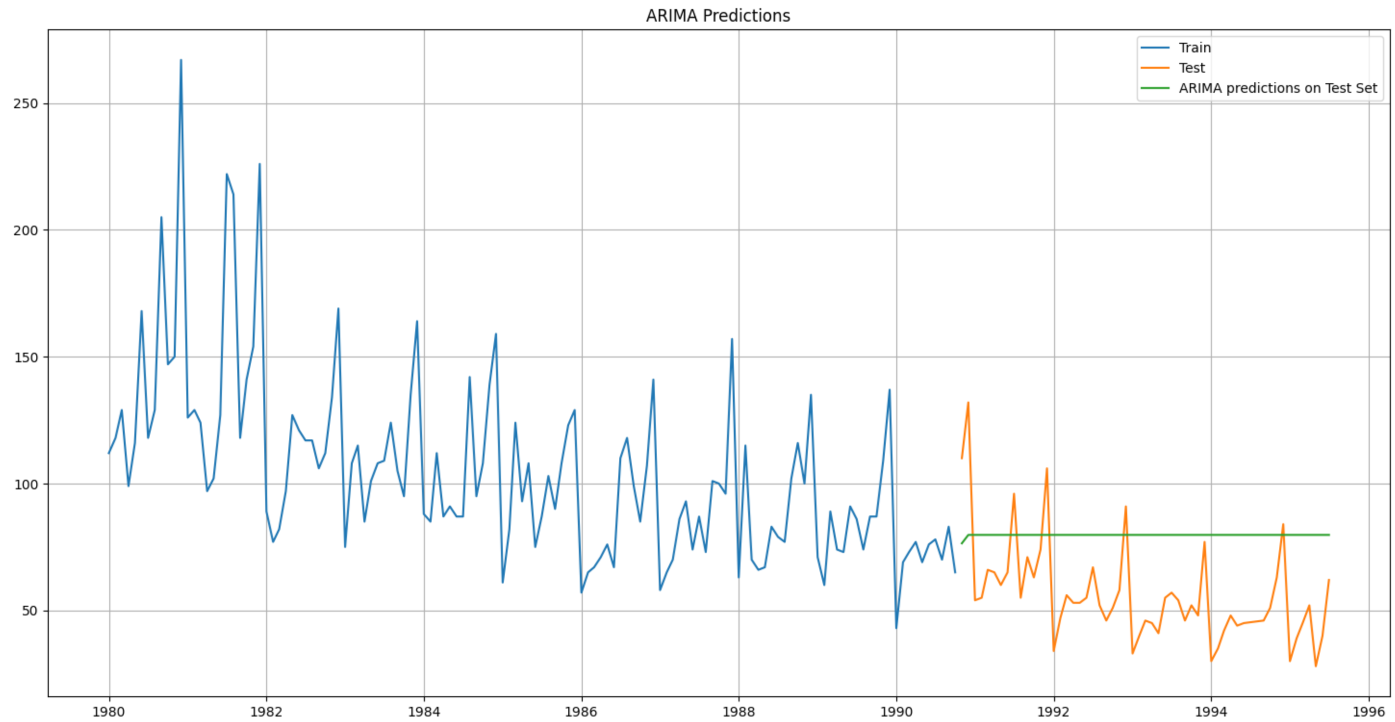


Figure 82: ARIMA Predictions

4.6.5. Auto SARIMA

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(5, 1, 1)x(1, 0, 1, 12)   Log Likelihood                -585.456
Date:                            Sat, 07 Dec 2024   AIC                           1190.913
Time:                                    21:31:46   BIC                           1219.511
Sample:                                01-01-1980   HQIC                          1202.533
                                     - 10-01-1990
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     -0.0070      0.015     -0.461      0.645      -0.037       0.023
ar.L1          0.2135      0.104      2.055      0.040       0.010       0.417
ar.L2         -0.1959      0.102     -1.913      0.056      -0.397       0.005
ar.L3          0.1336      0.111      1.204      0.229      -0.084       0.351
ar.L4         -0.0916      0.122     -0.753      0.451      -0.330       0.147
ar.L5          0.0738      0.106      0.694      0.488      -0.135       0.282
ma.L1         -0.9356      0.060    -15.520      0.000      -1.054      -0.817
ar.S.L12       0.9860      0.022     45.194      0.000       0.943       1.029
ma.S.L12      -0.7993      0.150     -5.314      0.000      -1.094      -0.504
sigma2       436.6381     62.403      6.997      0.000     314.331     558.945
===================================================================================
Ljung-Box (L1) (Q):                   0.06   Jarque-Bera (JB):                73.02
Prob(Q):                              0.81   Prob(JB):                         0.00
Heteroskedasticity (H):               0.34   Skew:                             0.95
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.16
===================================================================================
```

Figure 83: Auto SARIMA Summary

For auto SARIMA forecast on the Test Data, RMSE is 11.64

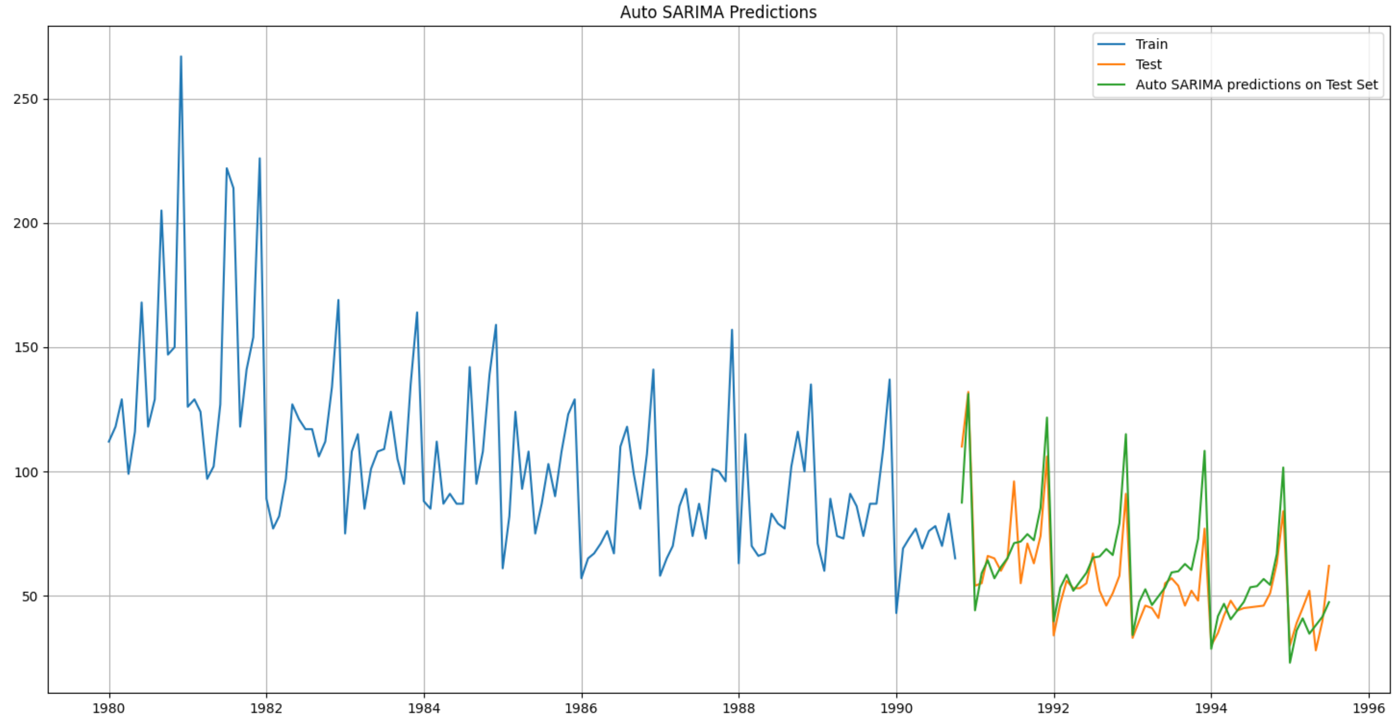


Figure 84: Auto SARIMA Predictions

4.6.6. SARIMA

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                Rose   No. Observations:                  130
Model:             SARIMAX(1, 1, 2)x(0, 1, [1], 12)   Log Likelihood                -530.109
Date:                             Sat, 07 Dec 2024   AIC                           1070.218
Time:                                     21:34:43   BIC                           1084.029
Sample:                                 01-01-1980   HQIC                          1075.825
                                      - 10-01-1990
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5383      0.200     -2.694      0.007      -0.930      -0.147
ma.L1         -0.1374      0.171     -0.803      0.422      -0.473       0.198
ma.L2         -0.7208      0.148     -4.870      0.000      -1.011      -0.431
ma.S.L12      -0.7736      0.136     -5.683      0.000      -1.040      -0.507
sigma2       450.0689     60.467      7.443      0.000     331.555     568.583
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                38.18
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               0.35   Skew:                             0.22
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.76
===================================================================================
```

Figure 85: SARIMA Summary

For SARIMA forecast on the Test Data, RMSE is 13.15

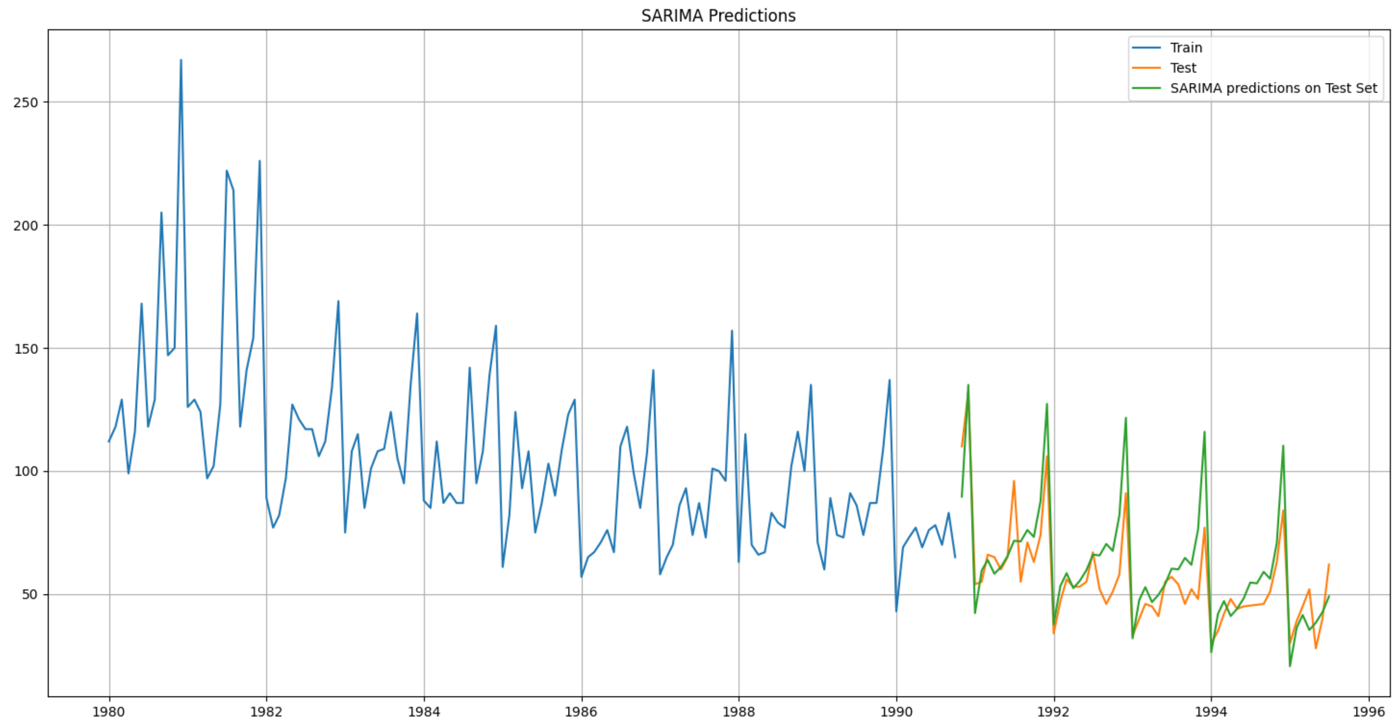


Figure 86: SARIMA Predictions

Insights

- The Auto SARIMA model achieved the lowest RMSE of 11.637070, indicating that it provides the most accurate forecasts among the models built with stationary data. This model effectively captures both seasonality and stationary trends in the rose wine sales data.
- The manually tuned SARIMA model achieved an RMSE of 13.150083, which is close to the Auto SARIMA model but slightly less accurate. This suggests that while SARIMA is a strong candidate, automatic parameter tuning (Auto SARIMA) provides better results for this dataset.
- Both ARIMA and Auto ARIMA models produced the same RMSE of 30.903804, significantly higher than Auto SARIMA and SARIMA. This indicates that these models are less effective due to their inability to account for seasonal components, which are crucial in rose wine sales forecasting.
- The strong performance of Auto SARIMA and SARIMA highlights the importance of incorporating seasonal adjustments when modeling stationary data. Non-seasonal models like ARIMA fall short in capturing these critical patterns.

4.7. Compare the performance of the models

	Test RMSE
Alpha=0.0999,Beta=1.9932,Gamma=0.0001,TripleExponentialSmoothing	9.328733
Auto SARIMA	11.637070
2pointTrailingMovingAverage	11.801043
SARIMA	13.150083
4pointTrailingMovingAverage	15.367212
6pointTrailingMovingAverage	15.862350
9pointTrailingMovingAverage	16.341919
RegressionOnTime	17.333437
Alpha=0.3,Beta=0.4,DoubleExponentialSmoothing	18.343250
Alpha=0.9,SimpleExponentialSmoothing	22.496819
Alpha=0.3,Beta=0.7,Gamma=0.4,TripleExponentialSmoothing	28.362899
Alpha=0.12,SimpleExponentialSmoothing	29.223677
AutoARIMA	30.903804
ARIMA	30.903804
SimpleAverage	52.412093

Figure 87: Model RMSE

Model	RMSE	Performance
Triple Exponential Smoothing (Alpha=0.0999, Beta=1.9932, Gamma=0.0001)	9.328733	Best model overall for non-stationary data. Captures trends and seasonality well.
Auto SARIMA	11.637070	Strong performance for stationary data. Robust for handling seasonality in a stationary time series.
2-Point Trailing Moving Average	11.801043	Close to Auto SARIMA but lacks explicit seasonality modeling. Suitable for short-term smoothing.
SARIMA	13.150083	Performs slightly worse than Auto SARIMA but still a reliable option for stationary data.
4-Point Trailing Moving Average	15.367212	Useful for medium-term trend smoothing but cannot handle seasonality.
6-Point Trailing Moving Average	15.86235	Highlights longer-term trends but sacrifices responsiveness to short-term fluctuations.

9-Point Trailing Moving Average	16.341919	Emphasizes long-term trends but ignores seasonality and short-term variations.
Regression on Time	17.333437	Simple to implement, but fails to capture seasonality, making it unsuitable for seasonal data like rose wine sales.
Double Exponential Smoothing (Alpha=0.3, Beta=0.4)	18.34325	Captures trends effectively but cannot model seasonality.
Simple Exponential Smoothing (Alpha=0.9)	22.496819	Focuses on short-term changes but performs poorly on seasonal data.
Triple Exponential Smoothing (Alpha=0.3, Beta=0.7, Gamma=0.4)	28.362899	Captures seasonality but performs poorly due to suboptimal parameter selection.
Simple Exponential Smoothing (Alpha=0.12)	29.223677	Smoothing method suited for volatile data without trends or seasonality.
Auto ARIMA	30.903804	Poor performance for seasonal data due to lack of explicit seasonality handling.
ARIMA	30.903804	Similar to Auto ARIMA but requires manual parameter tuning. Struggles to handle seasonality effectively.
Simple Average	52.412093	Baseline model with no ability to capture trends or seasonality, performing the worst.

Table 2: Model Performance – Rose Wine Dataset

4.7.1. Best Model based on RMSE- Triple Exponential Smoothing (Holt's Winter)

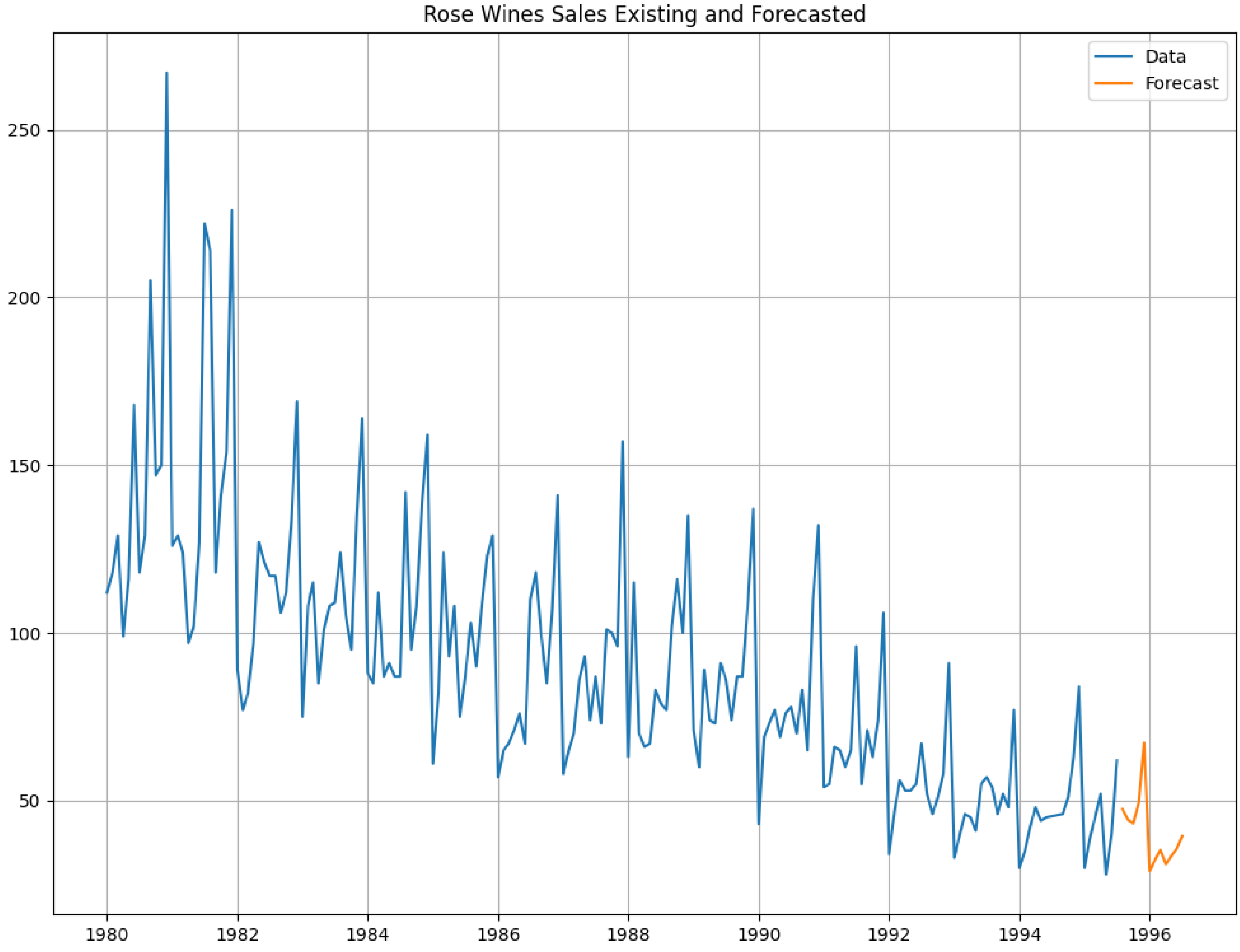


Figure 88: Holt's WInter Future Forecast

4.8. Actionable Insights and Recommendations

Benefits using the Holt's Winter Model

- Accurate Demand Forecasting: Predicts seasonal demand fluctuations, helping to optimize stock availability and prevent overstocking or stockouts.
- Inventory Optimization: Ensures optimal inventory levels, reducing storage costs and avoiding lost sales due to insufficient stock.
- Improved Sales & Marketing Strategies: Identifies peak and low-demand periods, enabling targeted marketing campaigns and promotions.
- Efficient Production Planning: Helps production teams align schedules with demand forecasts, avoiding bottlenecks and production delays.
- Data-Driven Decision-Making: Provides actionable insights for better decision-making on production, marketing, and distribution strategies.
- Cost Reduction & Operational Efficiency: Minimizes storage, labour, and production costs by aligning operations with demand forecasts.

- Scenario Planning & Risk Mitigation: Prepares the business for demand surges or declines, ensuring better responsiveness to unexpected changes.
- Competitive Advantage: Enhances customer satisfaction and brand loyalty by ensuring timely availability of products during peak demand periods.

Insights

- Rose wine sales show a fluctuating trend, with peaks observed in December and November (holiday season) and lower sales in the early months of the year (January–February).
- There appears to be a clear seasonal pattern, with sales picking up towards the end of the year, particularly around the holiday season.
- The forecasted values, show a rising trend towards the end of 1995 and into 1996. However, the forecast doesn't perfectly align with the actual data, indicating that while the trend is correct, the magnitude of sales might need further refinement.

Recommendations

- Focus on marketing campaigns that target the high sales season (November–December), such as holiday discounts or bundle offers with other wine varieties.
- Winter Promotions: Introduce special offers or discounts during the low-demand months (January–March) to incentivize purchases and maintain steady sales.
- Engage with Customers Year-Round: Develop year-round customer engagement strategies such as loyalty programs, wine-tasting events, and educational content to keep customers interested throughout the year.