TSF Project - Coded

Anirudh Sardiwal
7 July 2024

Table of Contents

1.0 Problem Definition & EDA

1.1 Problem Definition

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 - Rose and Sparkling. 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.

1.2 Basic EDA of Both Varieties

Date	Sales
1980-01-01	112.0
1980-02-01	118.0
1980-03-01	129.0
1980-04-01	99.0
1980-05-01	116.0

1.1 - Rose data head

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

1.2 Sparkling data head

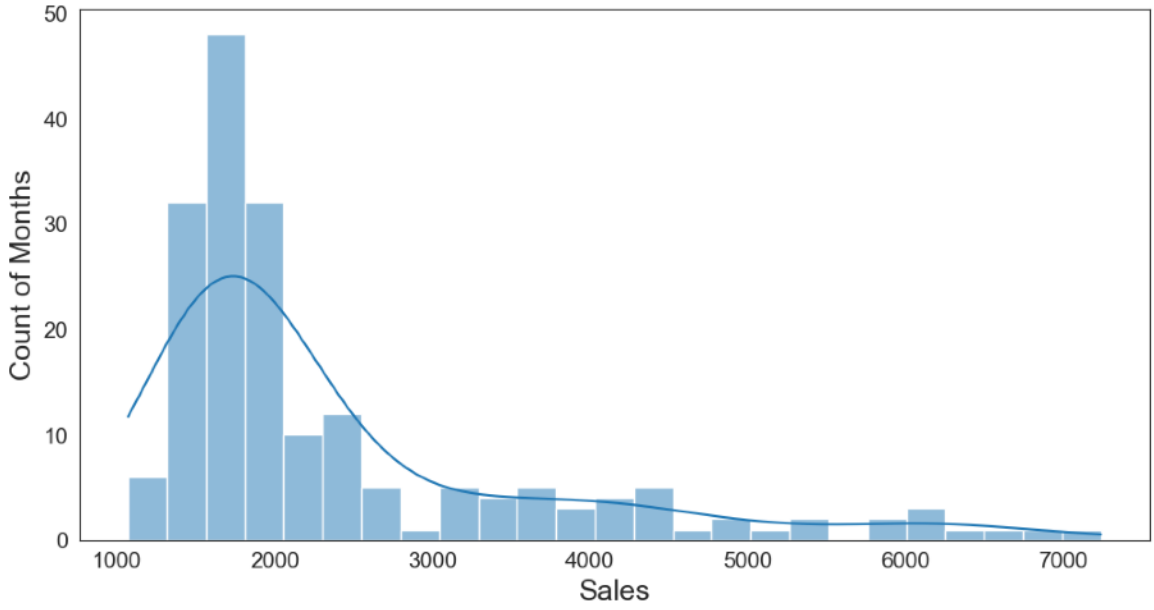
The data for both varieties is monthly and ranges from 1980-Jan to 1995-July, thus producing 187 rows.

	count	mean	std	min	25%	50%	75%	max
Sales	187.0	2402.0	1295.0	1070.0	1605.0	1874.0	2549.0	7242.0

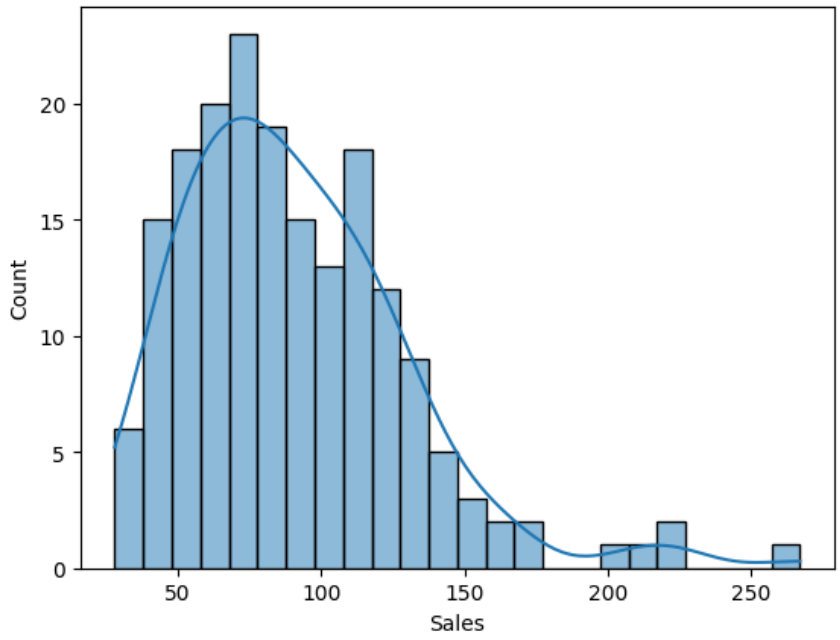
1.3 - Descriptive stats for Sparkling

	count	mean	std	min	25%	50%	75%	max
Sales	187.0	90.0	39.0	28.0	62.0	85.0	111.0	267.0

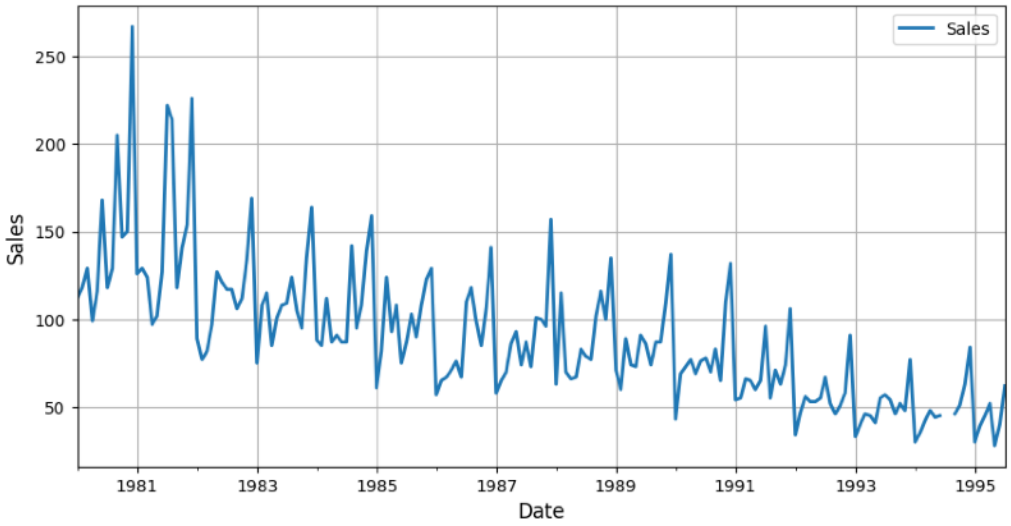
1.4 - Descriptive stats for Rose



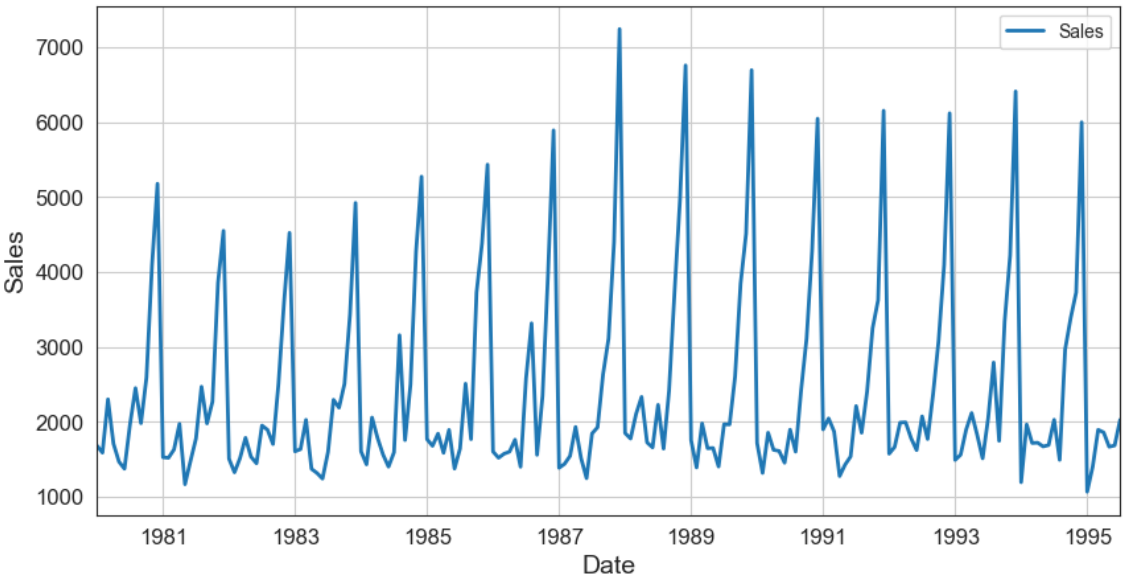
1.5 Distribution plot for Sparkling



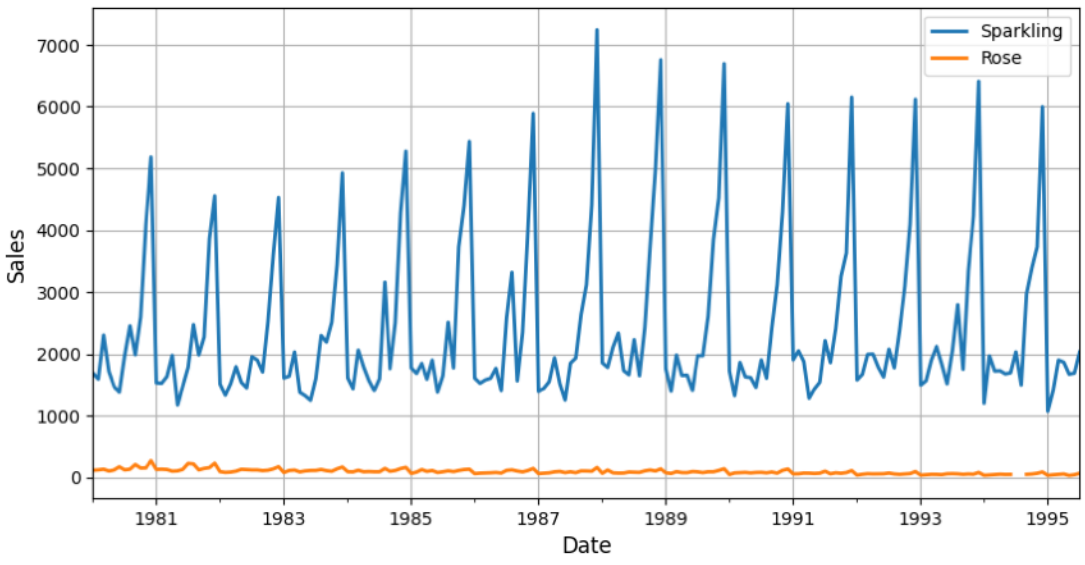
1.6 Distribution plot for Rose



1.7 - Series plot of Rose



1.8 - Series plot of Sparkling



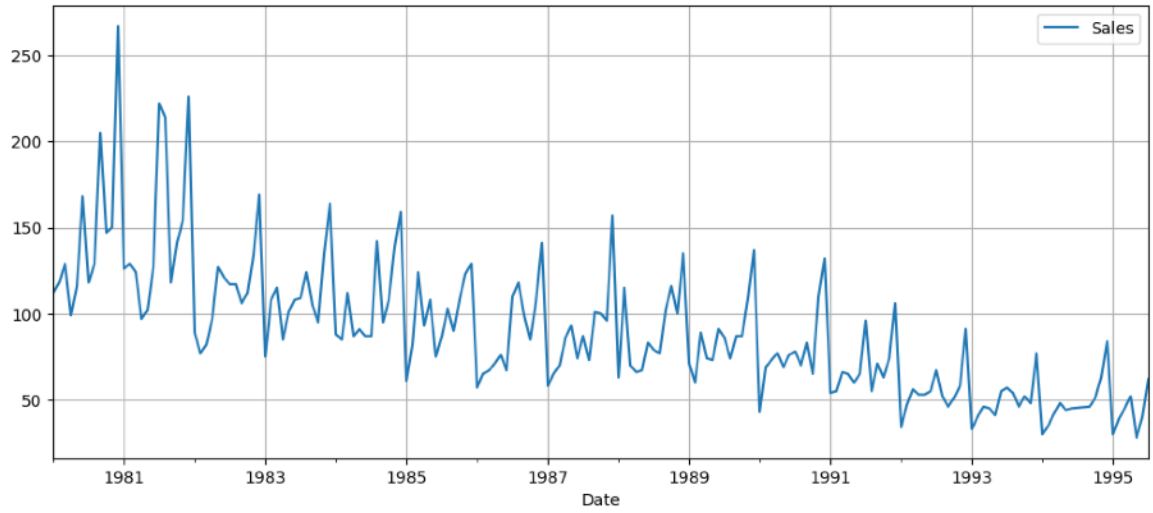
1.9 - Series plot of both varieties together

2.0 Rose

2.1 EDA contd.

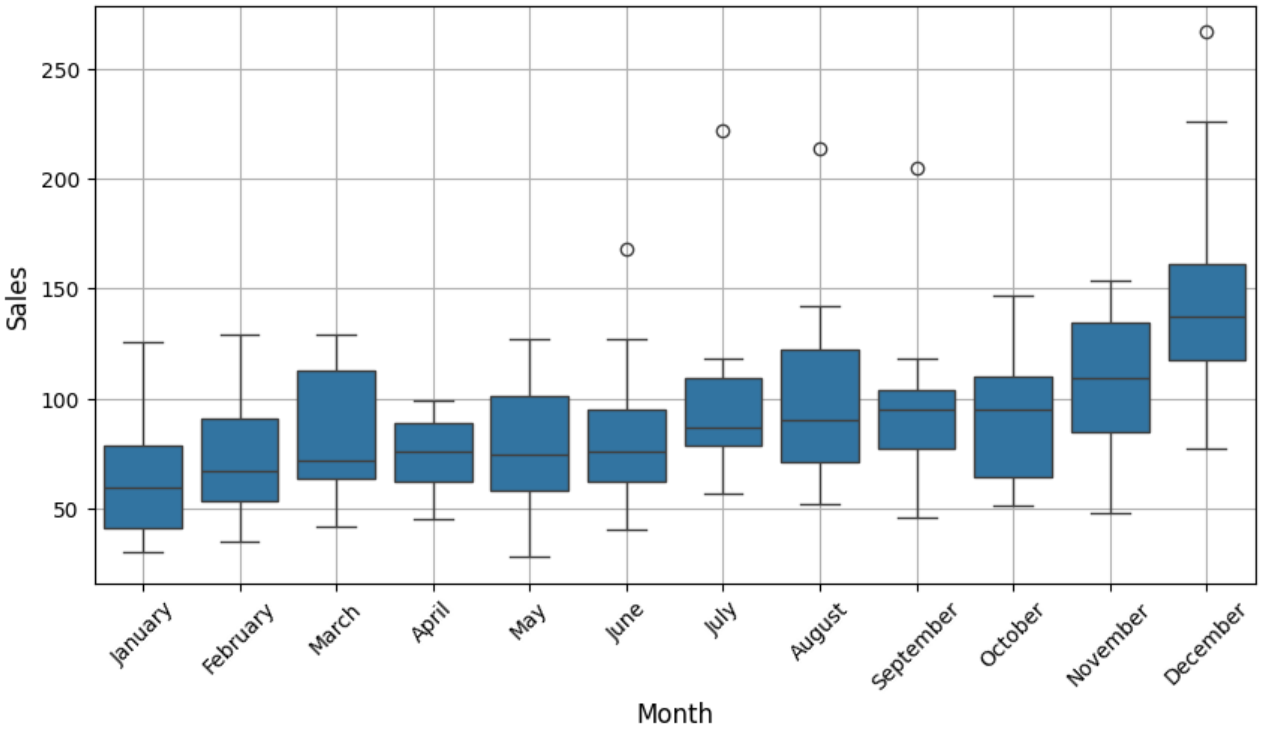
2.1.1 Imputing Missing Values

Missing values were imputed using Linear Interpolation.

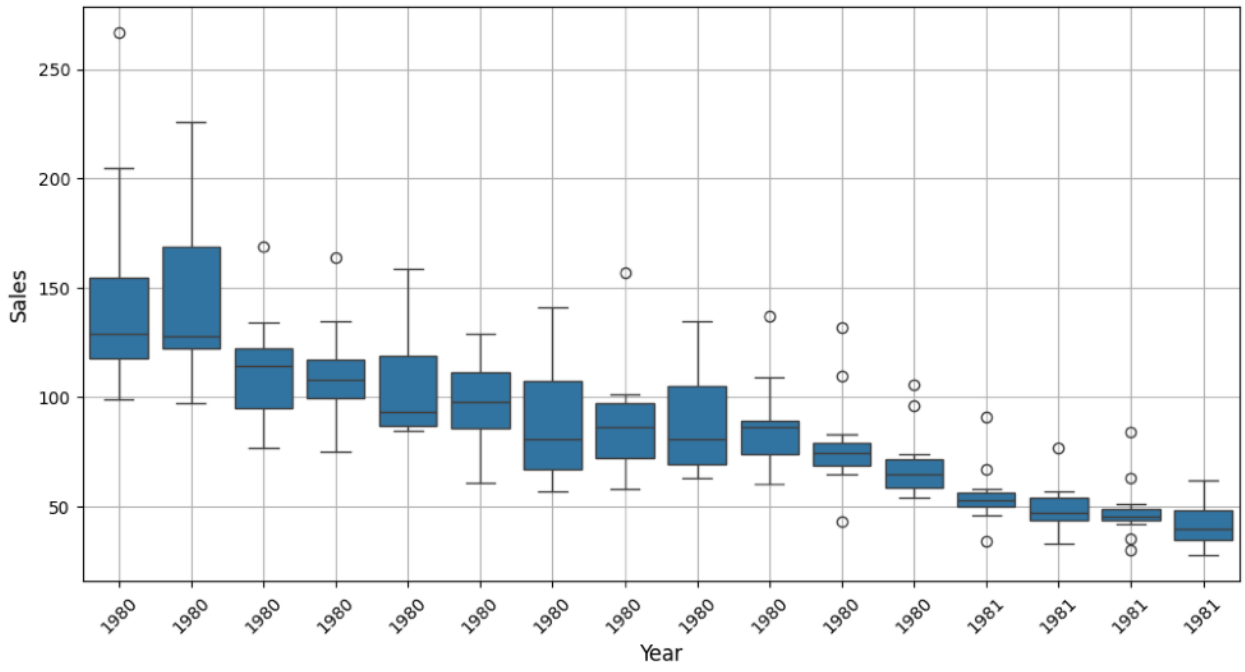


2.1 - Series plot of Rose after imputing missing values

2.1.2 Boxplot Visualization



2.2 - Boxplots by Month

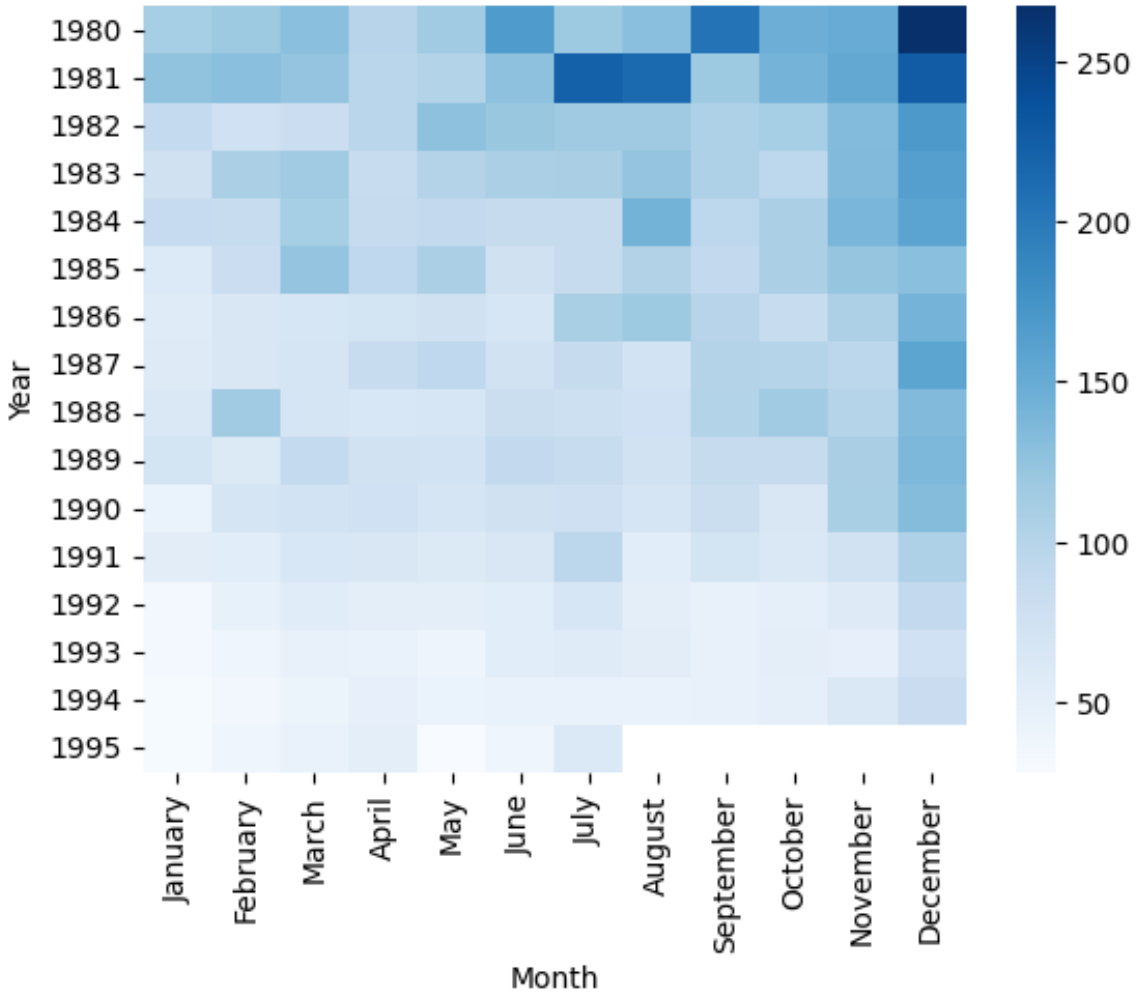


2.3 - Boxplots by Year

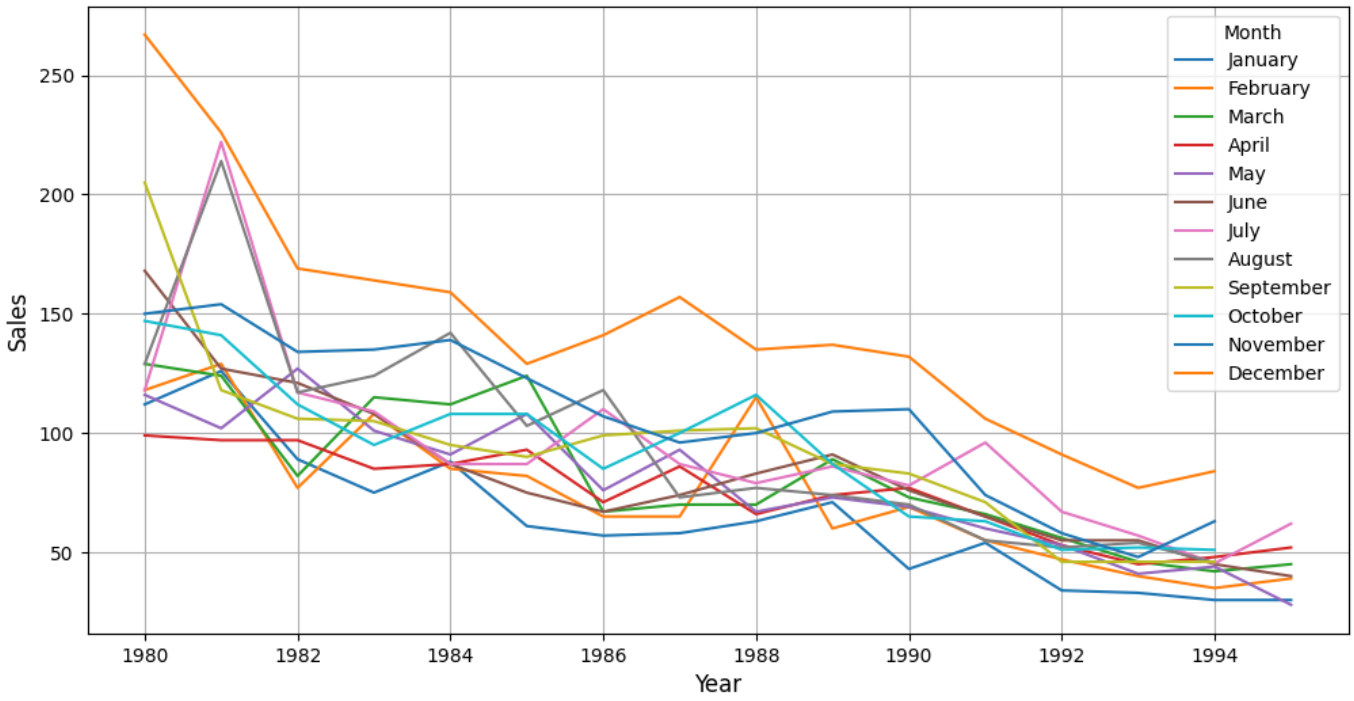
2.1.3 Pivot Table Analysis

Month Year	January	February	March	April	May	June	July	August	September	October	November	December
1980	112.0	118.0	129.0	99.0	116.0	168.0	118.0	129.0	205.0	147.0	150.0	267.0
1981	126.0	129.0	124.0	97.0	102.0	127.0	222.0	214.0	118.0	141.0	154.0	226.0
1982	89.0	77.0	82.0	97.0	127.0	121.0	117.0	117.0	106.0	112.0	134.0	169.0
1983	75.0	108.0	115.0	85.0	101.0	108.0	109.0	124.0	105.0	95.0	135.0	164.0
1984	88.0	85.0	112.0	87.0	91.0	87.0	87.0	142.0	95.0	108.0	139.0	159.0
1985	61.0	82.0	124.0	93.0	108.0	75.0	87.0	103.0	90.0	108.0	123.0	129.0
1986	57.0	65.0	67.0	71.0	76.0	67.0	110.0	118.0	99.0	85.0	107.0	141.0
1987	58.0	65.0	70.0	86.0	93.0	74.0	87.0	73.0	101.0	100.0	96.0	157.0
1988	63.0	115.0	70.0	66.0	67.0	83.0	79.0	77.0	102.0	116.0	100.0	135.0
1989	71.0	60.0	89.0	74.0	73.0	91.0	86.0	74.0	87.0	87.0	109.0	137.0
1990	43.0	69.0	73.0	77.0	69.0	76.0	78.0	70.0	83.0	65.0	110.0	132.0
1991	54.0	55.0	66.0	65.0	60.0	65.0	96.0	55.0	71.0	63.0	74.0	106.0
1992	34.0	47.0	56.0	53.0	53.0	55.0	67.0	52.0	46.0	51.0	58.0	91.0
1993	33.0	40.0	46.0	45.0	41.0	55.0	57.0	54.0	46.0	52.0	48.0	77.0
1994	30.0	35.0	42.0	48.0	44.0	45.0	0.0	0.0	46.0	51.0	63.0	84.0
1995	30.0	39.0	45.0	52.0	28.0	40.0	62.0	NaN	NaN	NaN	NaN	NaN

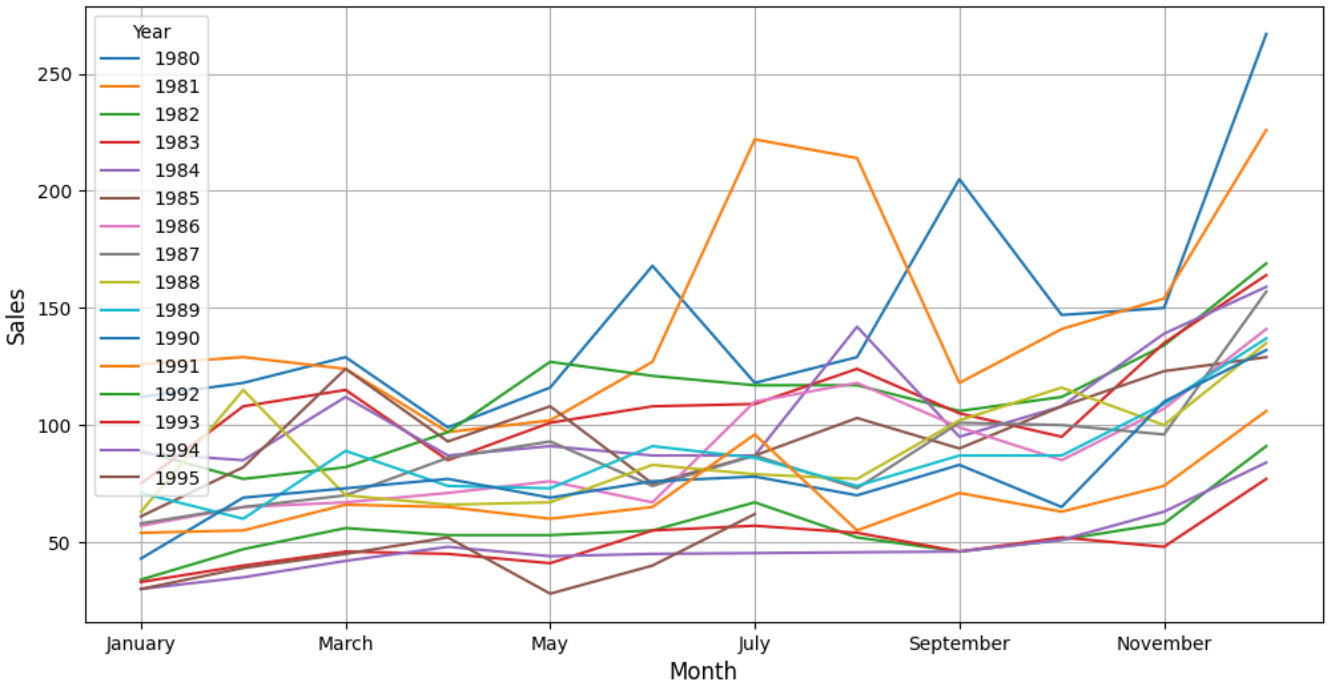
2.4 - Pivot Table - Month vs Year



2.5 - Heatmap of Pivot Table

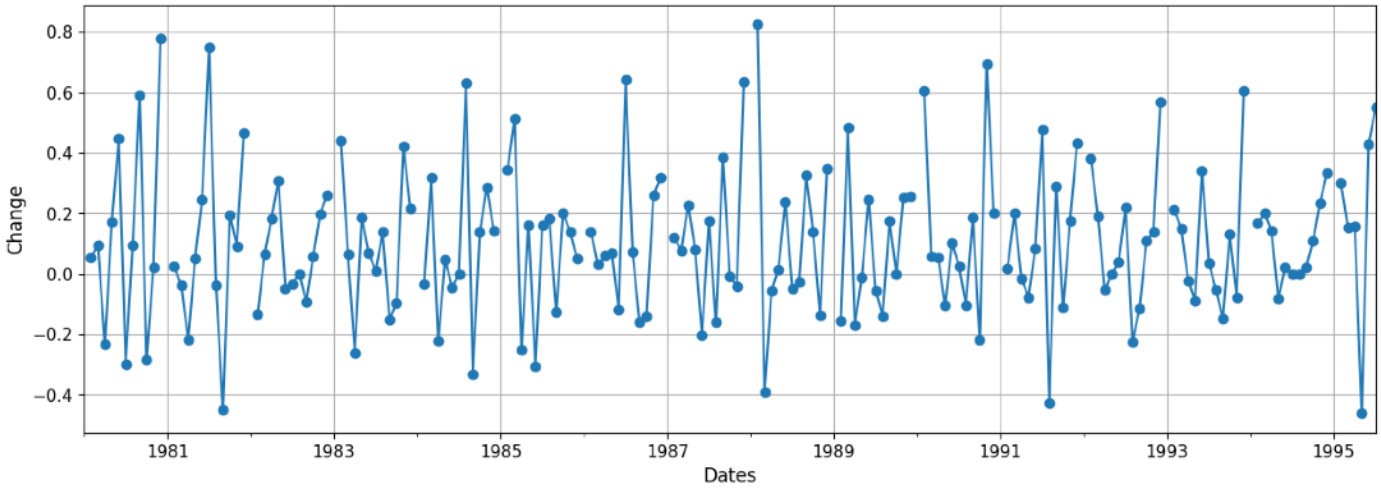


2.6 - Lineplots of Monthly Sales by Year

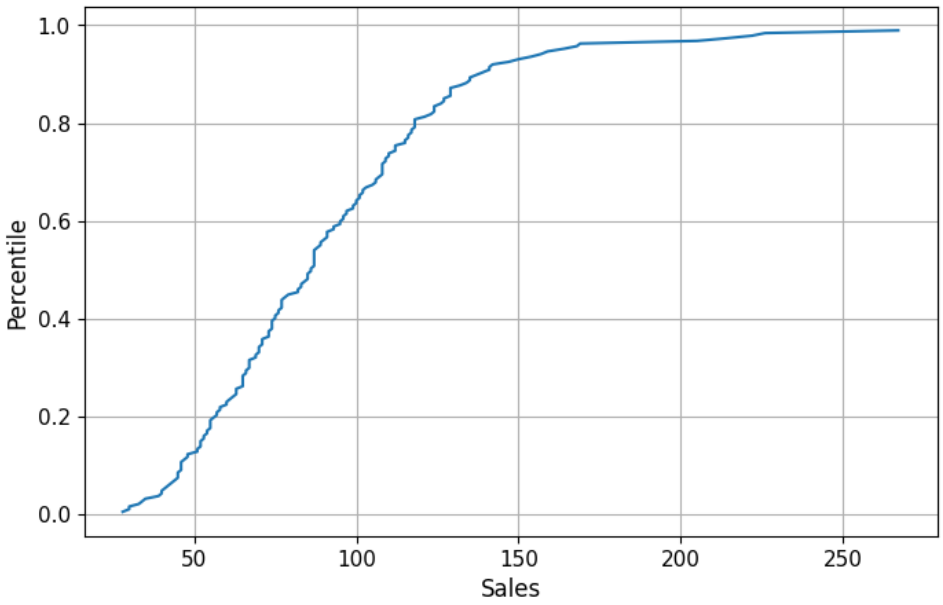


2.7 - Lineplots of Yearly Sales by Month

2.1.4 Other Plots

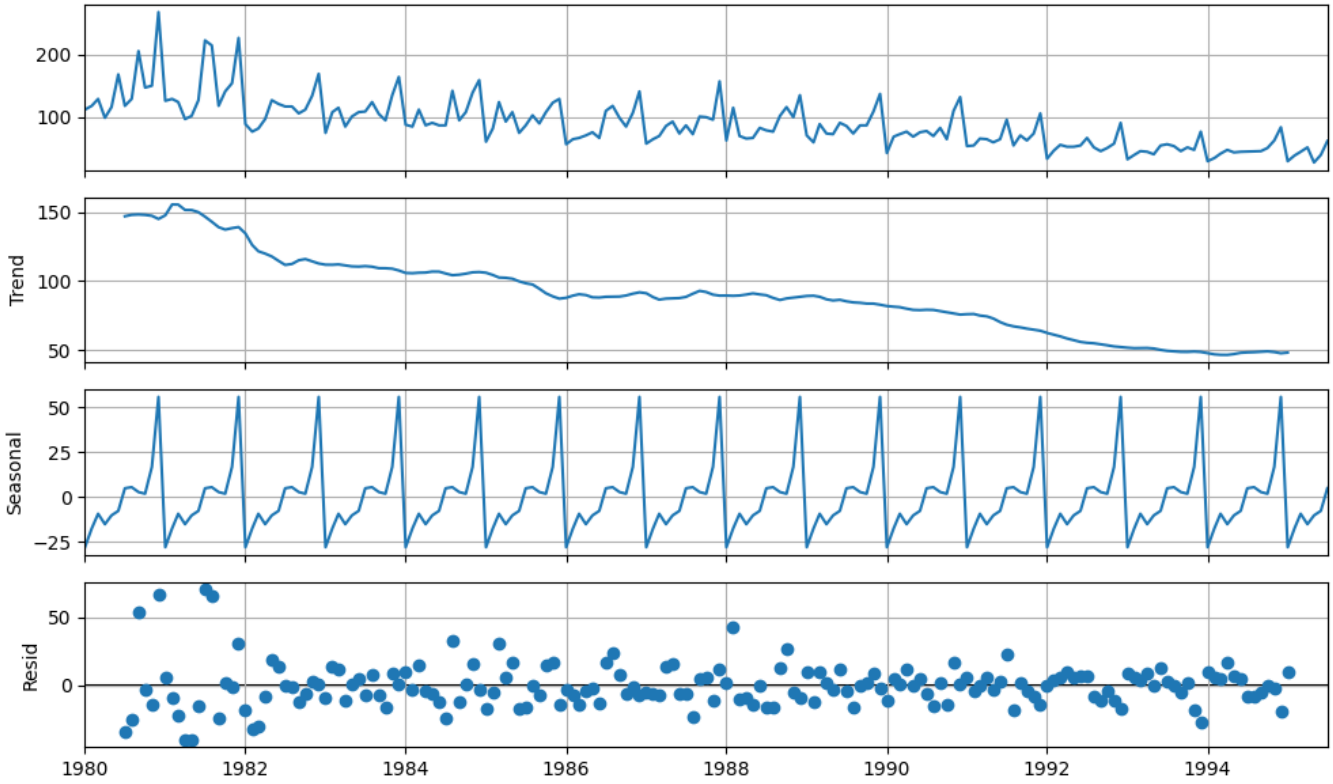


2.8 - MoM Percentage Change in Sales

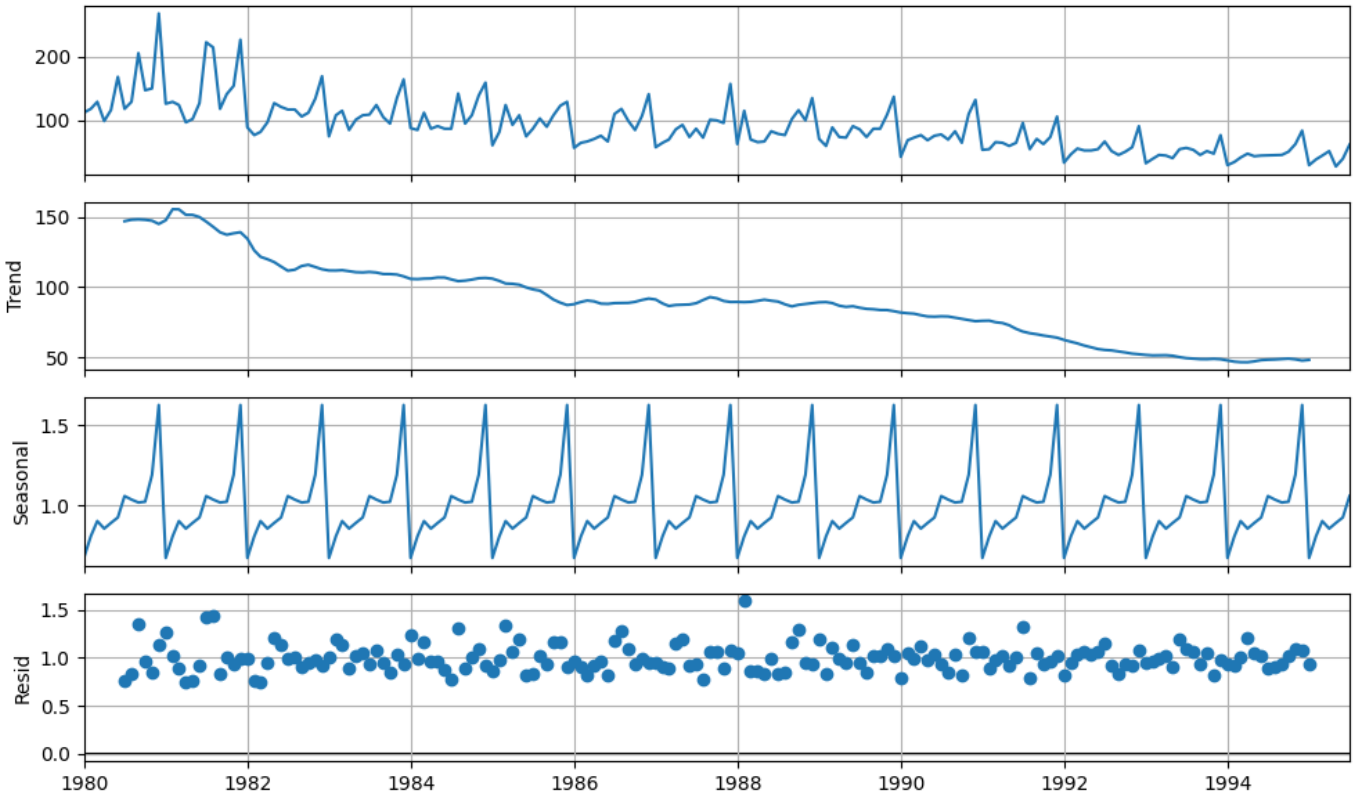


2.9 - ECDF Plot

2.1.5 Decomposition



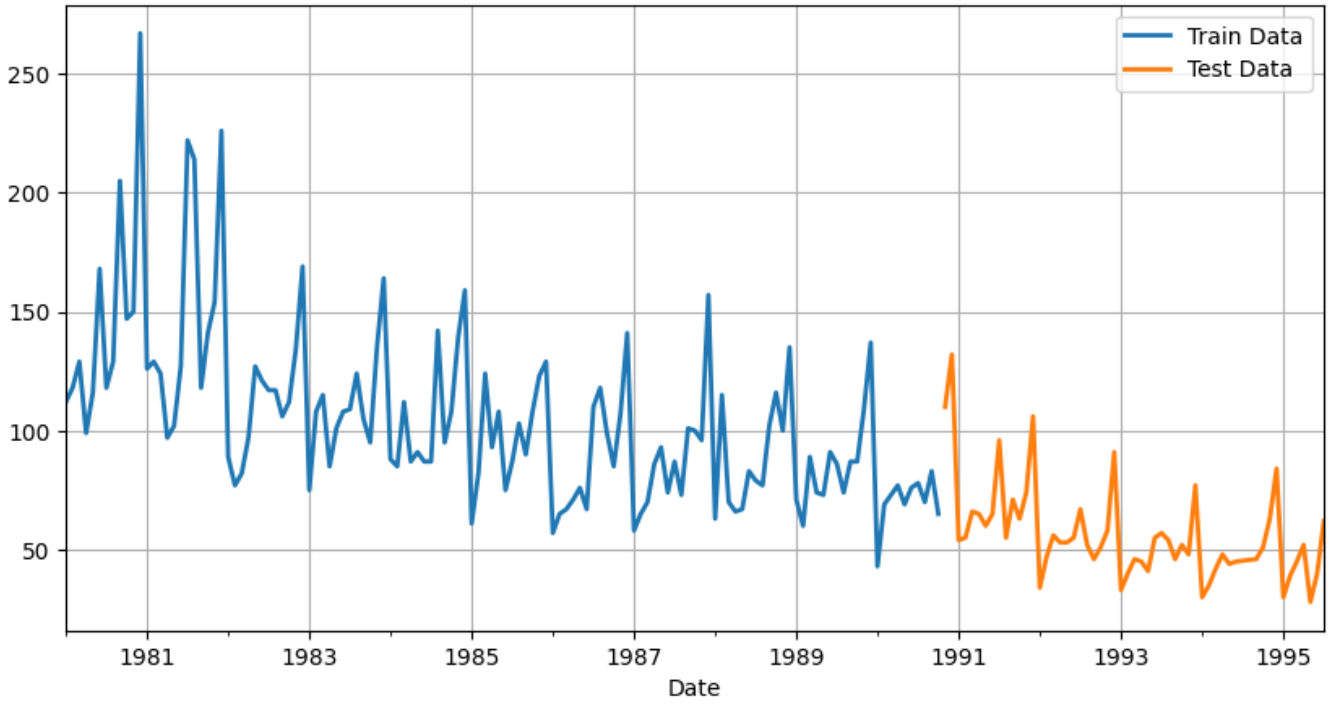
2.11 - Additive Decomposition



2.12 - Multiplicative Decomposition

2.2 Model Building - Original Data

2.2.1 Train Test Split



2.2.1 - Train Test Split

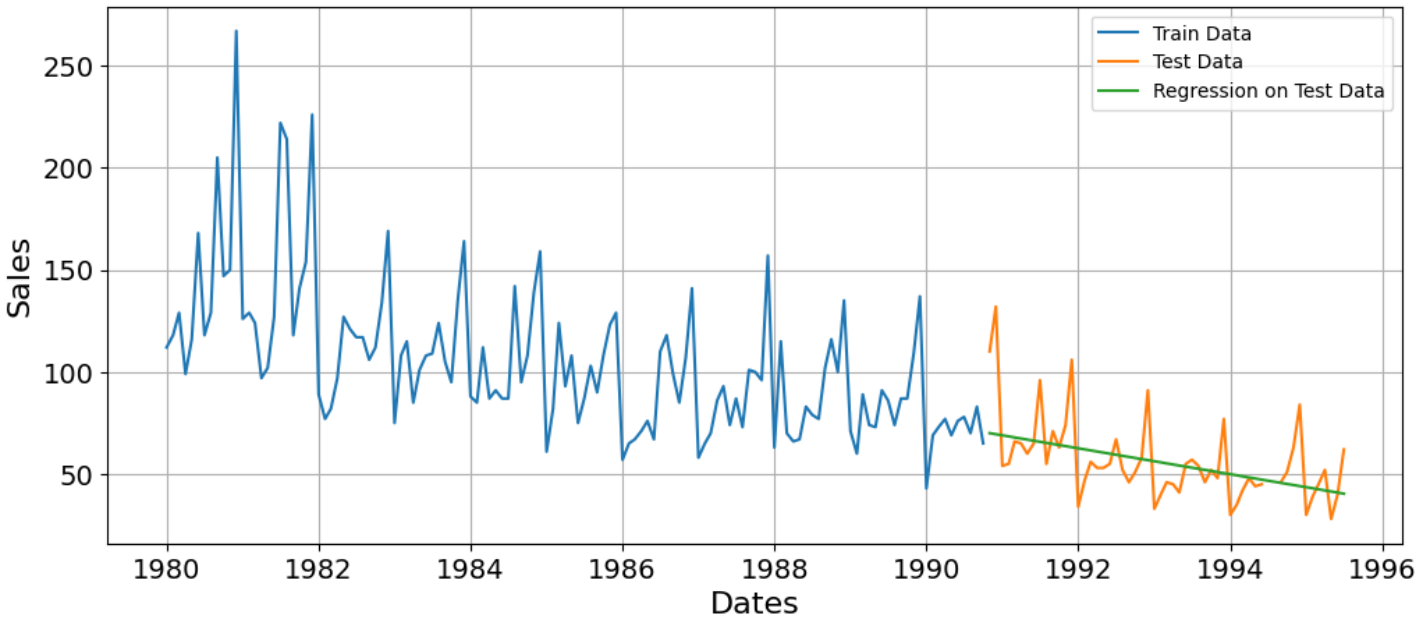
2.2.2 Linear Regression

Date	Sales	time
1980-01-01	112.0	1
1980-02-01	118.0	2
1980-03-01	129.0	3
1980-04-01	99.0	4
1980-05-01	116.0	5

2.2.2 - Head of Train Data with Time numbers

Date	Sales	time
1990-11-01	110.0	131
1990-12-01	132.0	132
1991-01-01	54.0	133
1991-02-01	55.0	134
1991-03-01	66.0	135

2.2.3 - Head of Test Data with Time numbers



2.2.4 - Linear Regression plot on Test data

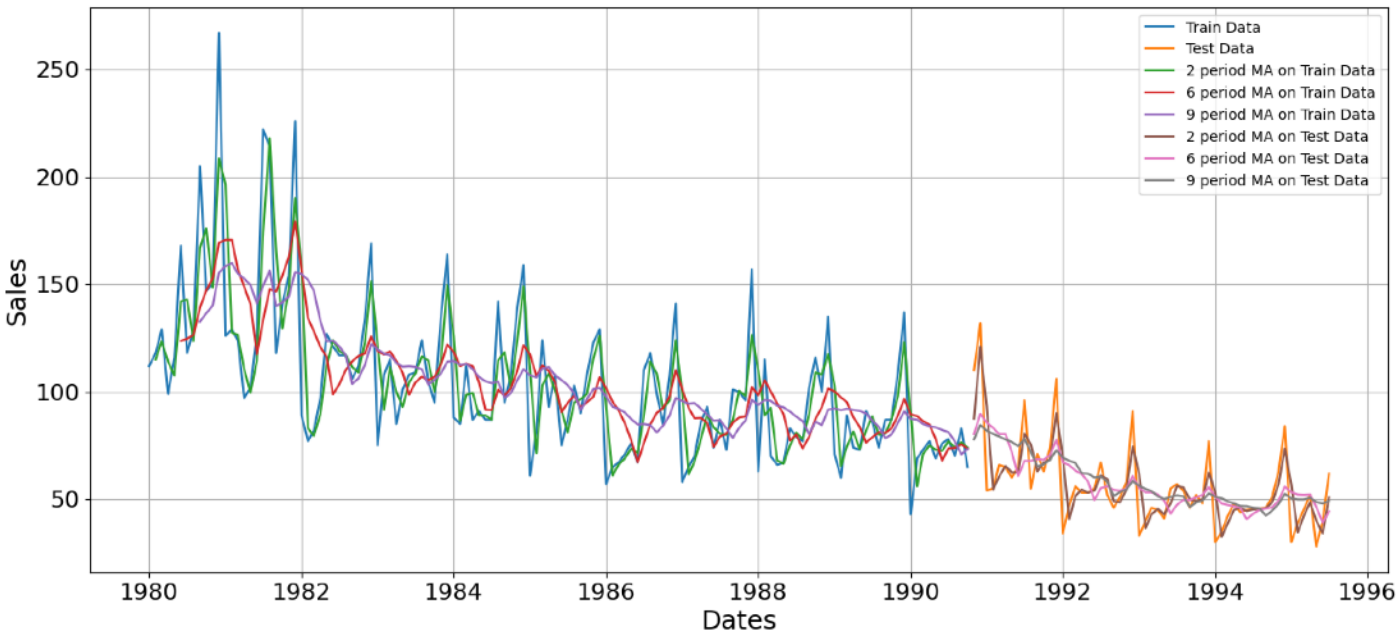
For Linear Regression on Test Data the RMSE is 17.36.

2.2.3 Moving Averages

For 2 point MA the RMSE is 11.80

For 6 point MA the RMSE is 15.86

For 9 point MA the RMSE is 16.34



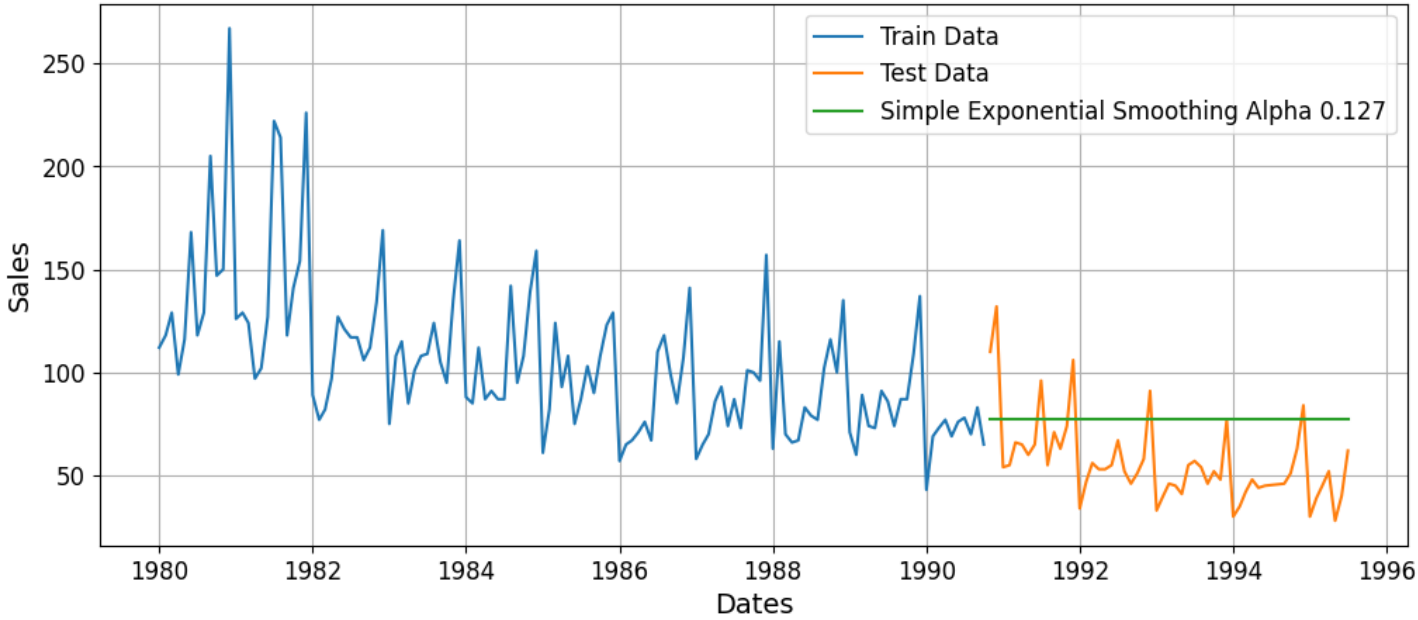
2.2.6 - Various Moving Averages plotted on Train and Test data

2.2.4 Simple Exponential Smoothing

For Simple Exponential Smoothing Alpha 0.127 the RMSE is 29.22

```
{'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}
```

2.2.8 - SES Autofit Parameters

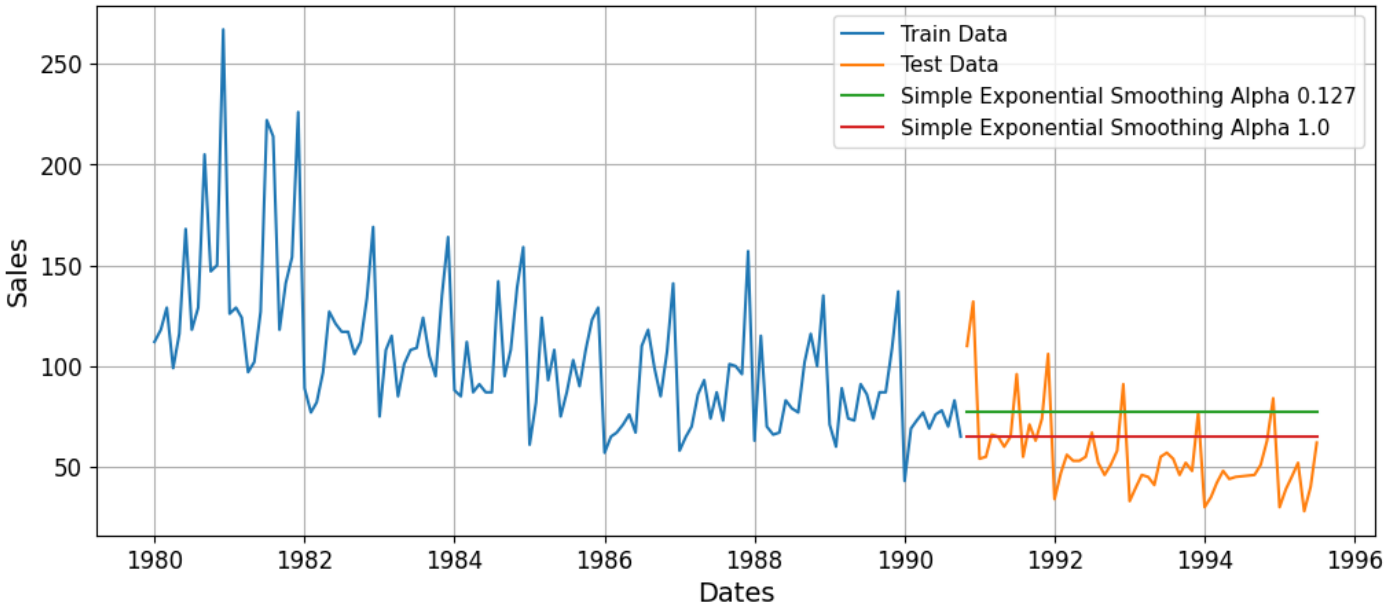


2.2.9 - SES Results on Test data

2.2.4.1 Refining Simple Exponential Smoothing

	Alpha	Train RMSE	Test RMSE
9	1.0	38.833273	21.766930
8	0.9	37.507371	22.496819
7	0.8	36.330954	23.212777
6	0.7	35.288467	23.894929
5	0.6	34.372651	24.529811

2.2.10 - Experimenting to find the best Alpha



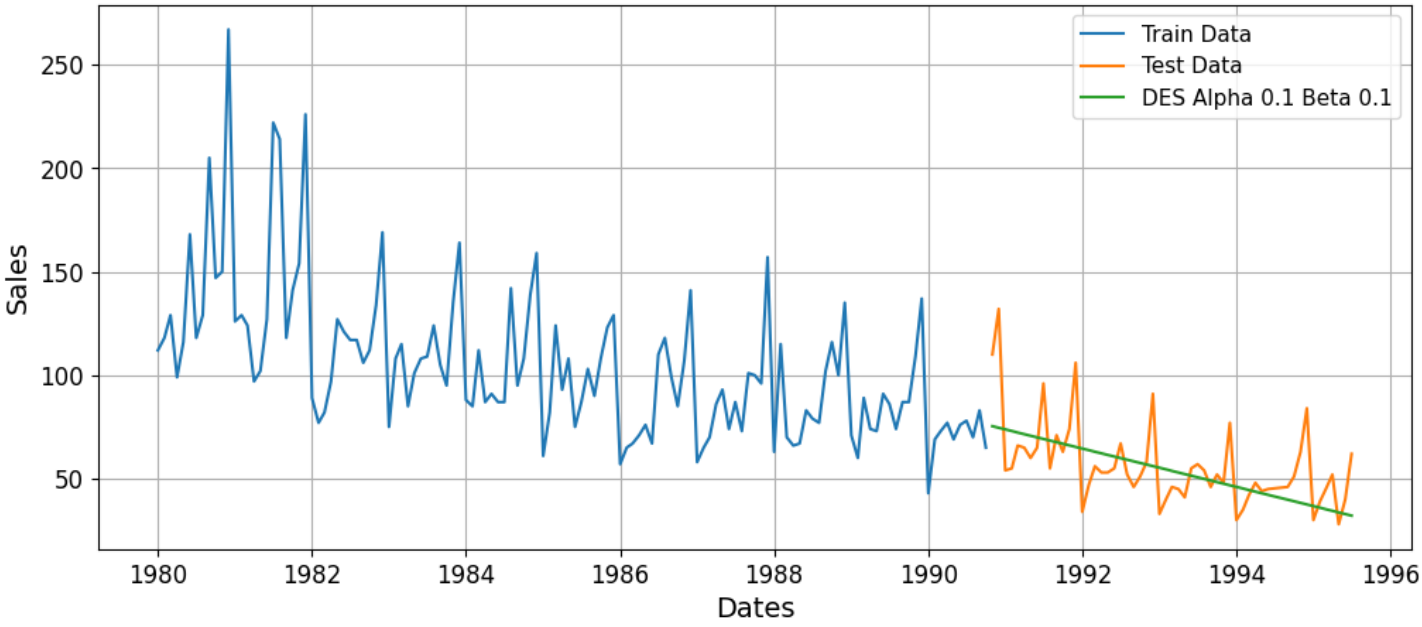
2.2.11 - Both SES models on Test data

2.2.5 Double Exponential Smoothing

	Alpha	Beta	Train RMSE	Test RMSE
0	0.1	0.1	34.248544	17.685906
16	0.2	0.7	40.347897	17.732195
23	0.3	0.4	37.287813	18.343250
36	0.4	0.7	40.744796	18.975318
33	0.4	0.4	37.990913	19.133156
...	...	...	...	...
88	0.9	0.9	55.814369	369.989590
97	1.0	0.8	57.575818	406.506557
89	0.9	1.0	58.406268	436.031598
98	1.0	0.9	60.691028	491.247514
99	1.0	1.0	64.093561	587.897332

100 rows × 4 columns

2.2.12 - Experimenting to find the best combination of Alpha and Beta



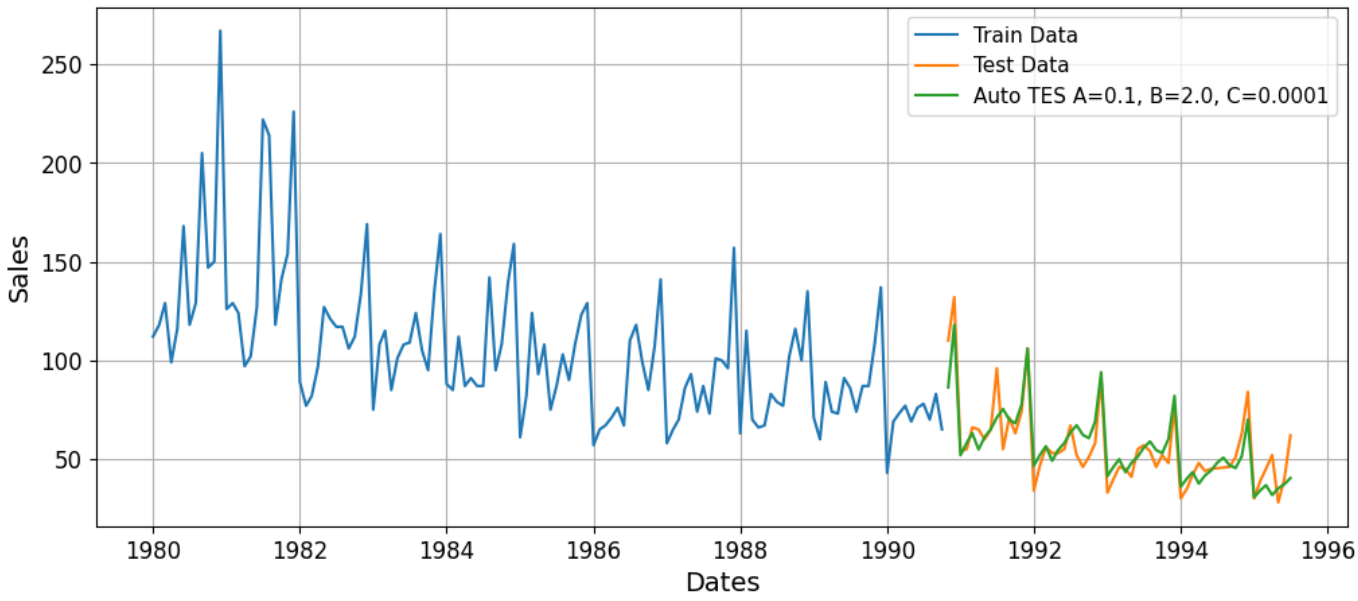
2.2.13 - Best DES model on Test data

2.2.6 Triple Exponential Smoothing

```
{'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}
```

2.2.14 - Auto TES parameters

For Auto TES the RMSE is 9.33

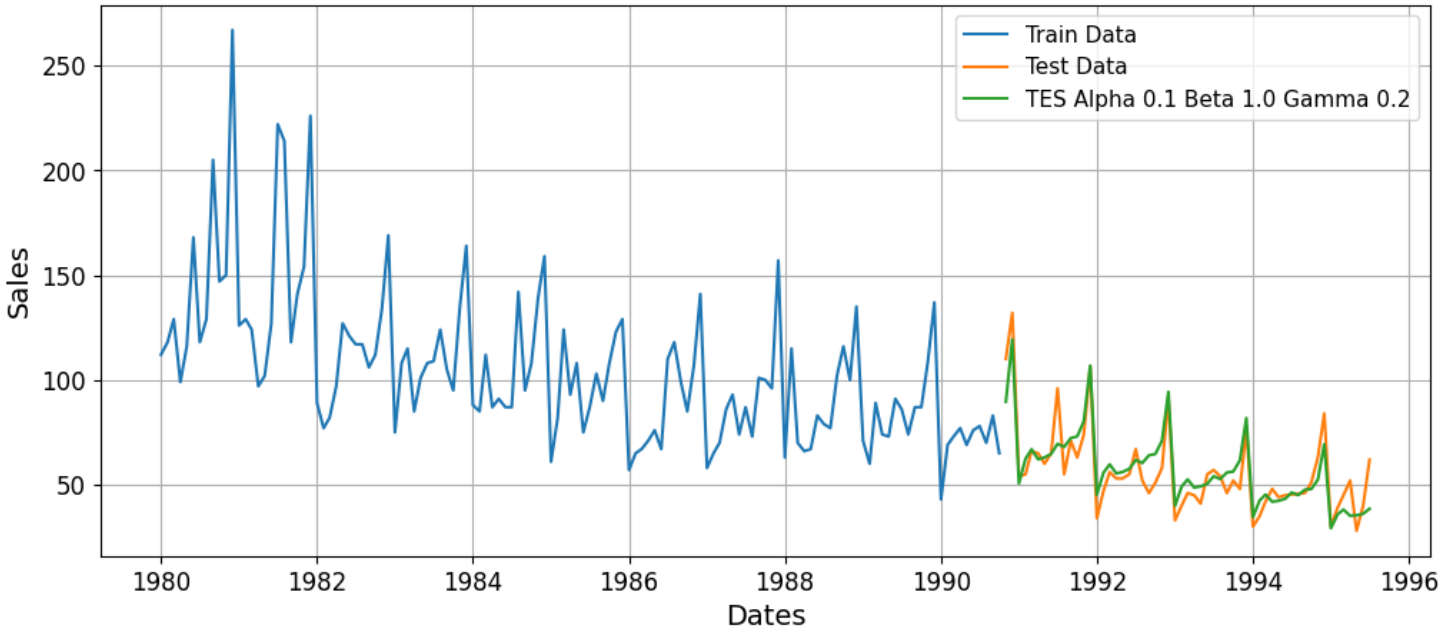


2.2.15 - Auto TES plot on Test data

	Alpha	Beta	Gamma	Train RMSE	Test RMSE
91	0.1	1.0	0.2	23.140673	9.129075
4	0.1	0.1	0.5	22.415877	9.212838
5	0.1	0.1	0.6	23.396696	9.282132
3	0.1	0.1	0.4	21.552201	9.288651
106	0.2	0.1	0.7	25.623069	9.331747
...	...	...	...	...	...
839	0.9	0.4	1.0	492.492945	325300.586086
488	0.5	0.9	0.9	816.234770	327275.368132
195	0.2	1.0	0.6	294.231444	380649.281071
982	1.0	0.9	0.3	3843.552854	478966.918432
587	0.6	0.9	0.8	2700.592196	630046.005970

1000 rows × 5 columns

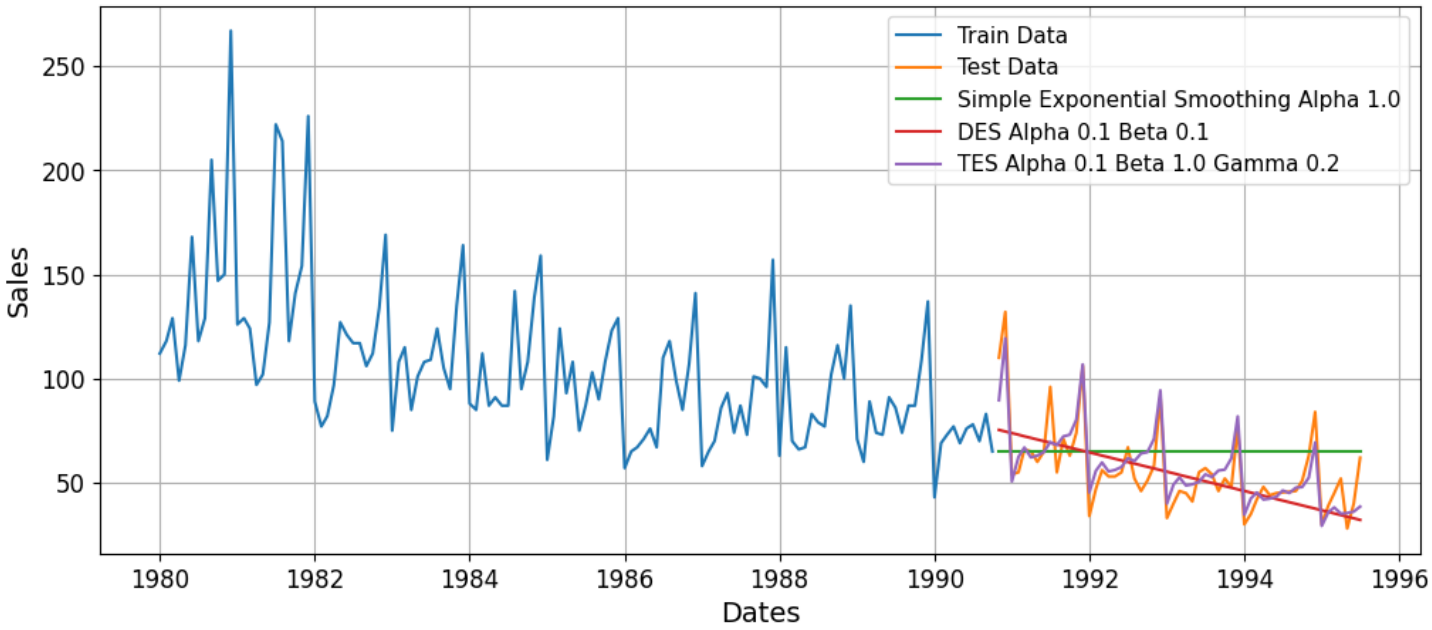
2.2.16 - Experimenting to Refine TES



2.2.17 - Plot of TES model chosen by least RMSE

For Refined TES the RMSE is 9.13

2.2.7 Plotting all 3 Exponential Smoothing Together



2.2.18 - All 3 forms of Exponential Smoothing plotted togethere on Test data

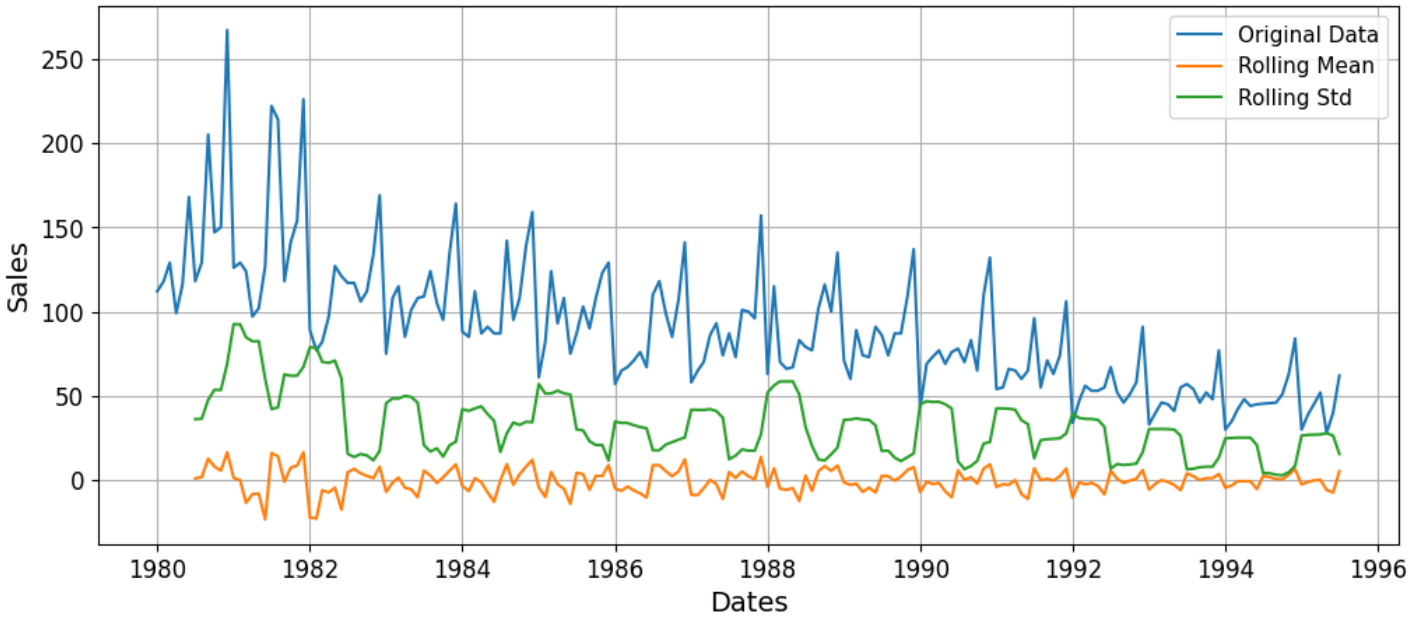
2.3 Model Building - Stationary Data

2.3.1 Check for Stationarity

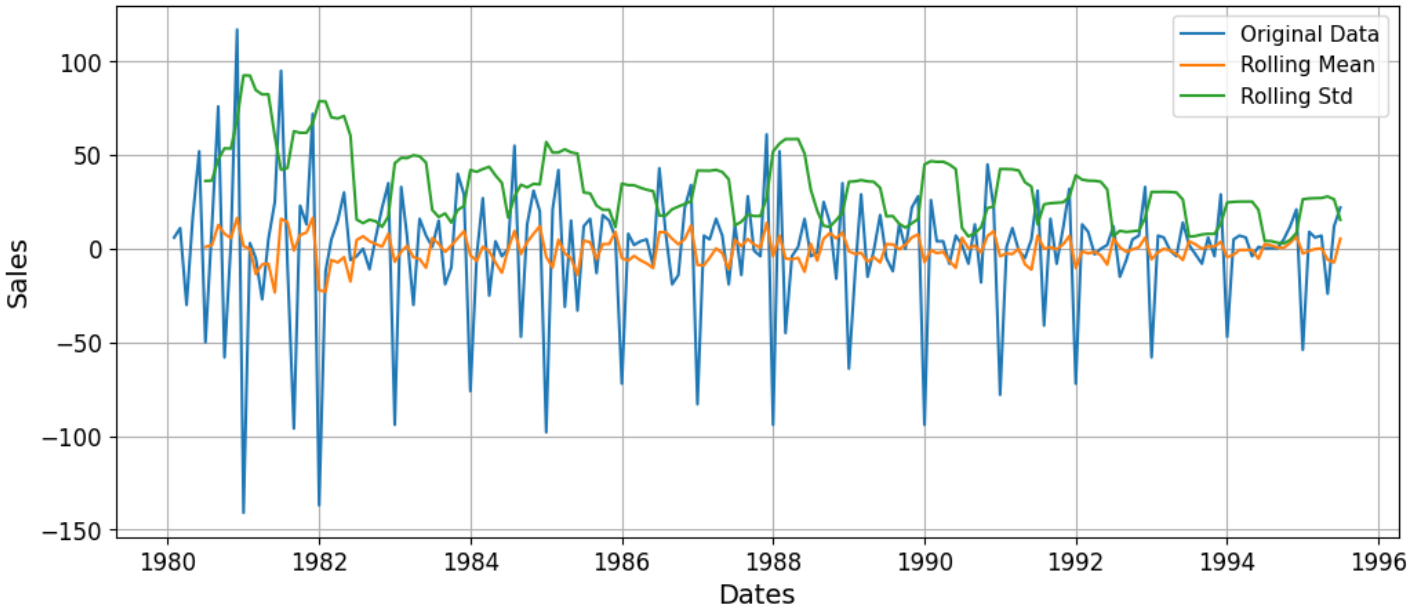
P-value is 0.34. Series is not Stationary.

```
(-1.876699107990828,
 0.3431007142833735,
 13,
 173,
 {'1%': -3.4687256239864017,
  '5%': -2.8783961376954363,
  '10%': -2.57575634100705},
 1516.2583629826856)
```

2.3.1 - Results of ADF test

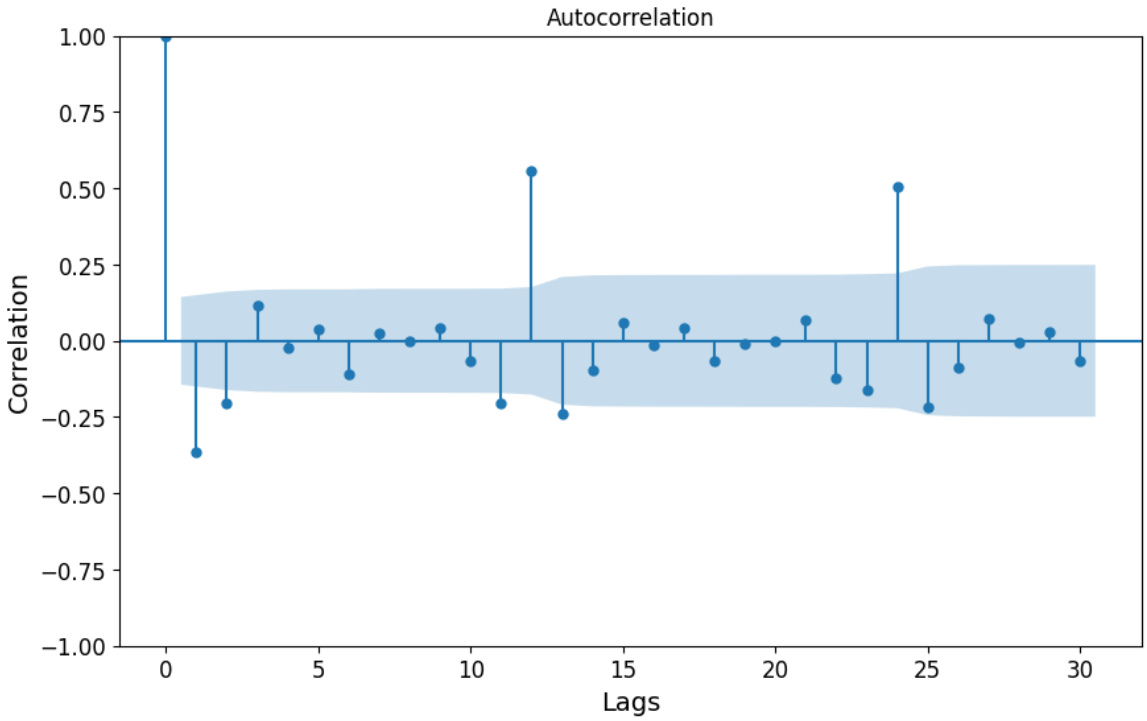


2.3.2 - Data before differencing

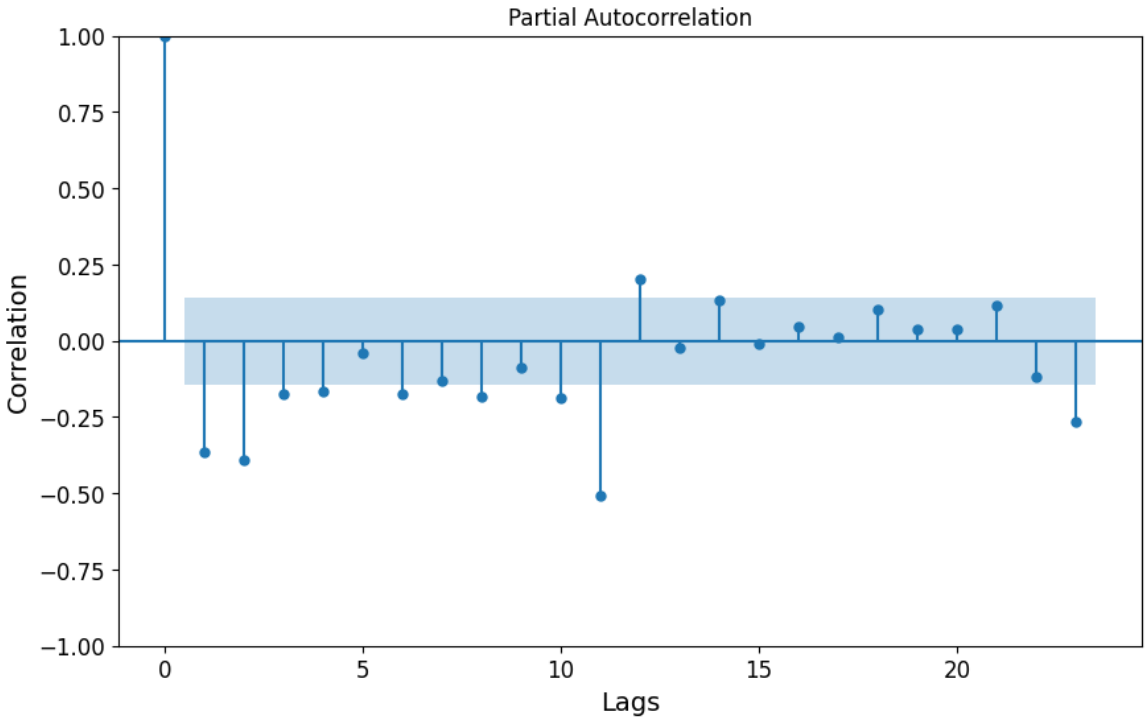


2.3.3 - Data after differencing

2.3.2 ACF & PACF Plots



2.3.4 - Autocorrelation Plot of differenced series



2.3.4 - Partial Autocorrelation Plot of differenced series

2.3.3 ARIMA Model

	param	AIC
2	(0, 1, 2)	1259.247780
5	(1, 1, 2)	1259.473205
4	(1, 1, 1)	1260.036763
7	(2, 1, 1)	1261.014076
1	(0, 1, 1)	1261.327444

2.3.5 - Selecting parameters based on AIC

For ARIMA(0,1,2) the RMSE is 30.90

SARIMAX Results

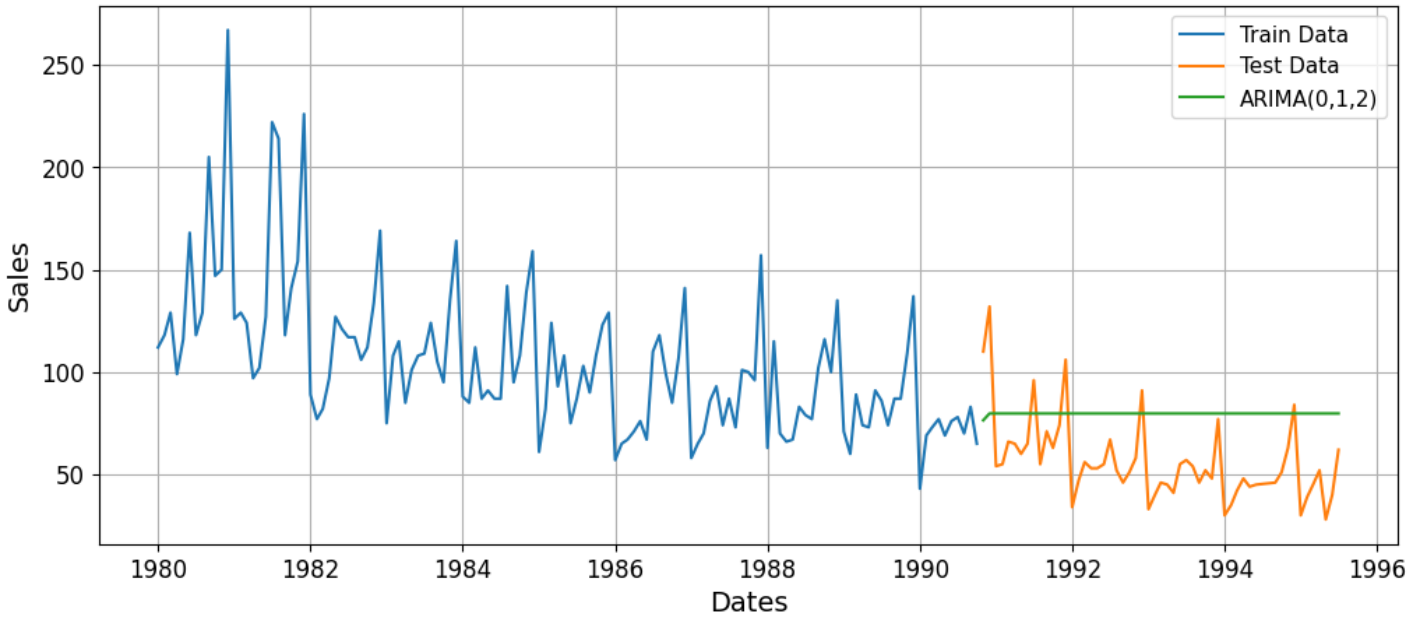
Dep. Variable:	Sales	No. Observations:	130
Model:	ARIMA(0, 1, 2)	Log Likelihood	-626.624
Date:	Tue, 09 Jul 2024	AIC	1259.248
Time:	05:43:09	BIC	1267.827
Sample:	01-01-1980	HQIC	1262.734
	- 10-01-1990		
Covariance Type:	opg		

	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

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

2.3.6 - ARIMA Model results



2.3.7 - Plot of ARIMA Model on Test data

2.3.4 SARIMA Model

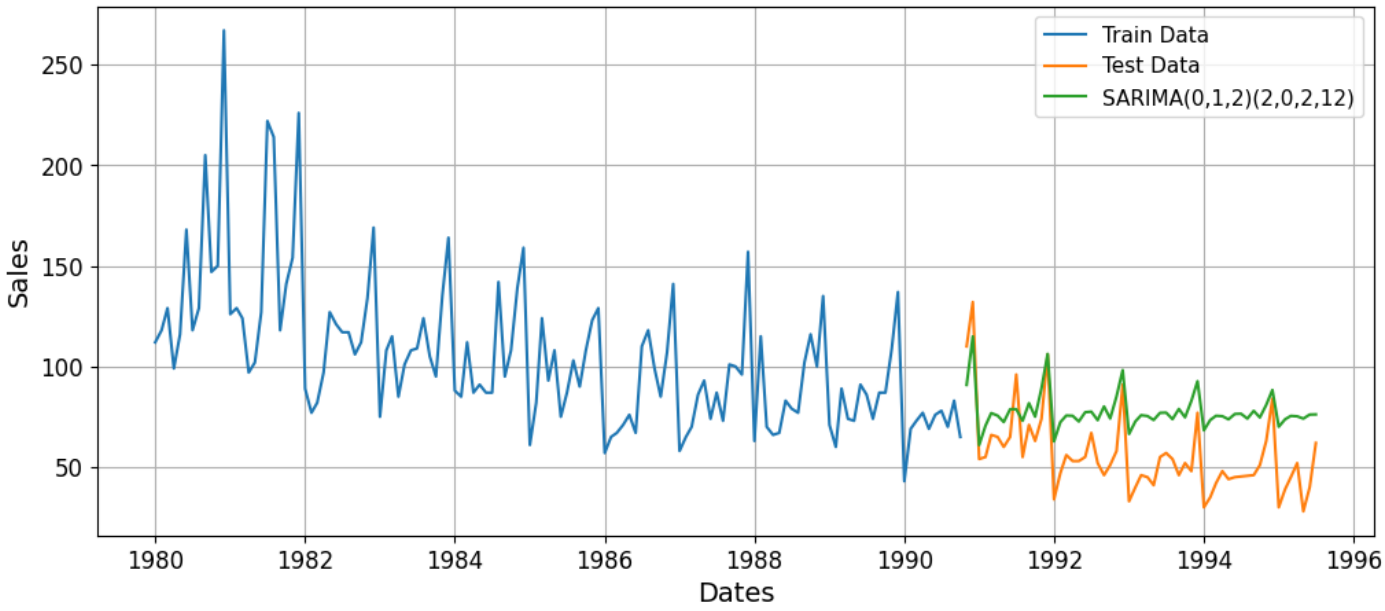
	param	param_seasonal	AIC
53	(0, 1, 2)	(2, 0, 2, 12)	871.075238
107	(1, 1, 2)	(2, 0, 2, 12)	873.003875
161	(2, 1, 2)	(2, 0, 2, 12)	874.213961
150	(2, 1, 1)	(2, 0, 0, 12)	879.792363
159	(2, 1, 2)	(2, 0, 0, 12)	880.763857

2.3.8 - Selecting parameters based on AIC

```
                                SARIMAX Results
==========================================================================================
Dep. Variable:                             Sales   No. Observations:                  130
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -428.538
Date:                            Tue, 09 Jul 2024   AIC                             871.075
Time:                                    05:51:31   BIC                             889.450
Sample:                                01-01-1980   HQIC                            878.516
                                     - 10-01-1990
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.8367    239.070     -0.003      0.997    -469.404     467.731
ma.L2         -0.1633     39.021     -0.004      0.997     -76.642      76.315
ar.S.L12       0.3494      0.079      4.408      0.000       0.194       0.505
ar.S.L24       0.3067      0.075      4.103      0.000       0.160       0.453
ma.S.L12       0.0454      0.134      0.338      0.735      -0.218       0.309
ma.S.L24      -0.0912      0.145     -0.628      0.530      -0.376       0.193
sigma2       250.7786      6e+04      0.004      0.997   -1.17e+05    1.18e+05
==========================================================================================
Ljung-Box (L1) (Q):                   0.09   Jarque-Bera (JB):                 3.10
Prob(Q):                              0.76   Prob(JB):                         0.21
Heteroskedasticity (H):               0.88   Skew:                             0.43
Prob(H) (two-sided):                  0.71   Kurtosis:                         3.05
==========================================================================================

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

2.3.9 - SARIMA Model results

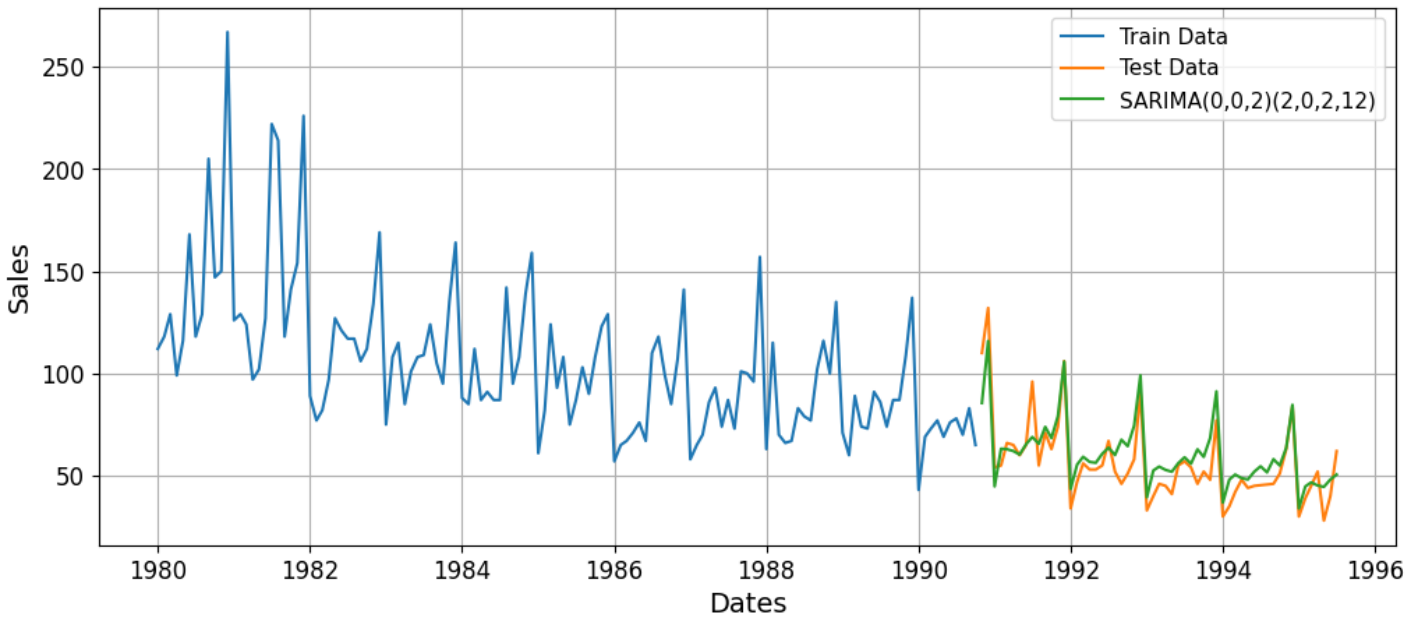


2.3.10 - Plot of SARIMA Model on Test data

RMSE for SARIMA(0,1,2)(2,0,2,12) is 25.34

2.3.5 Making d=0 and Generating Another SARIMA Model

RMSE for SARIMA(0,0,2)(2,0,2,12) is 10.13



2.3.11 - Plot of alternate SARIMA Model

2.4 Final Model Selection & Forecasting

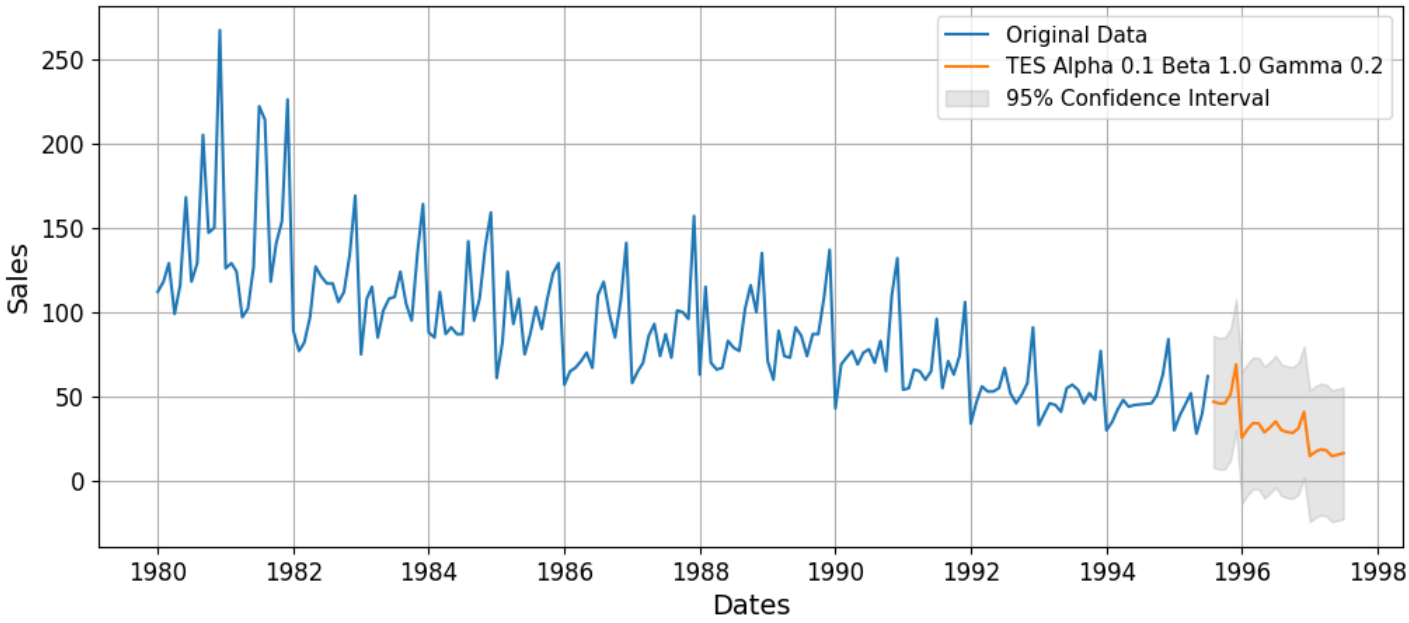
2.4.1 Model Selection

	Model	Test RMSE
8	TES Refined Alpha 0.1 Beta 1.0 Gamma 0.2	9.129075
7	TES Auto 0.1,2.0,0.0001	9.328733
11	SARIMA(0,0,2)(2,0,2,12)	10.128574
1	2 point MA	11.801043
2	6 point MA	15.862350
3	9 point MA	16.341919
0	Linear Regression	17.355796
6	DES Refined Alpha 0.1 Beta 0.1	17.685906
5	SES Refined Alpha 1.0	21.766930
10	SARIMA(0,1,2)(2,0,2,12)	25.343324
4	Auto SES Alpha 0.127	29.223677
9	ARIMA(0,1,2)	30.903804

2.4.1 - Comparison of all models by RMSE

Final model selected (based on least RMSE) is Triple Exponential Smoothing (Alpha 0.1, Beta 1.0, Gamma 0.2) with RMSE of 9.13

2.4.2 Forecasting

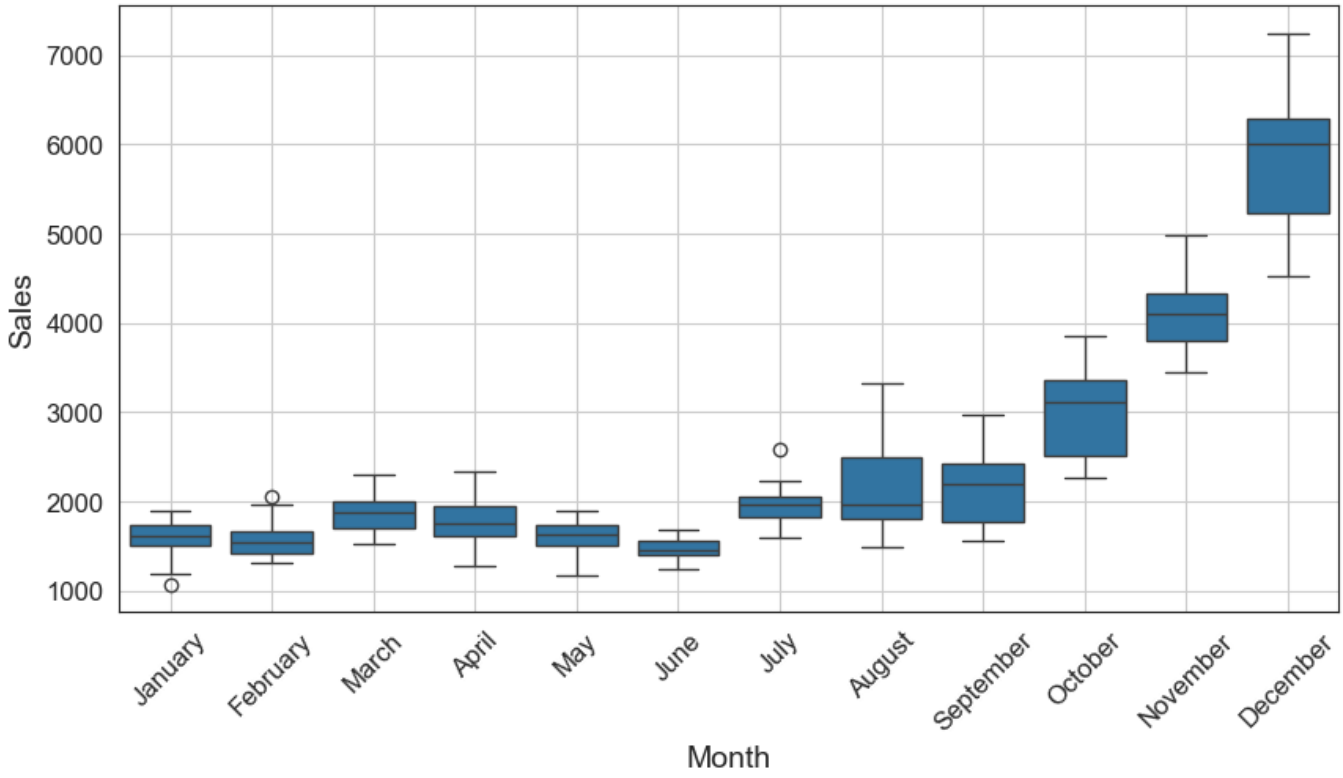


2.4.2 - 24 month Forecast with Final Model

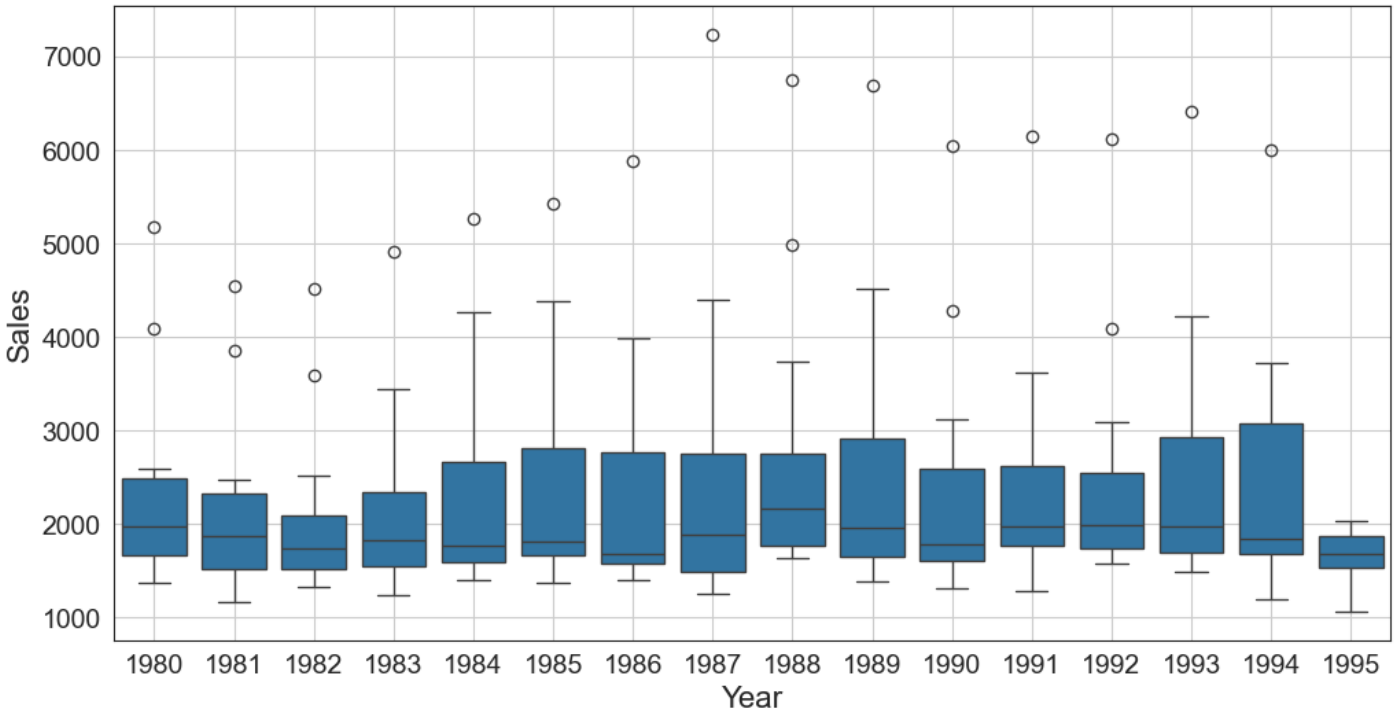
3.0 Sparkling

3.1 EDA Contd.

3.1.1 Boxplot Visualization



3.1.1 - Boxplots by Month

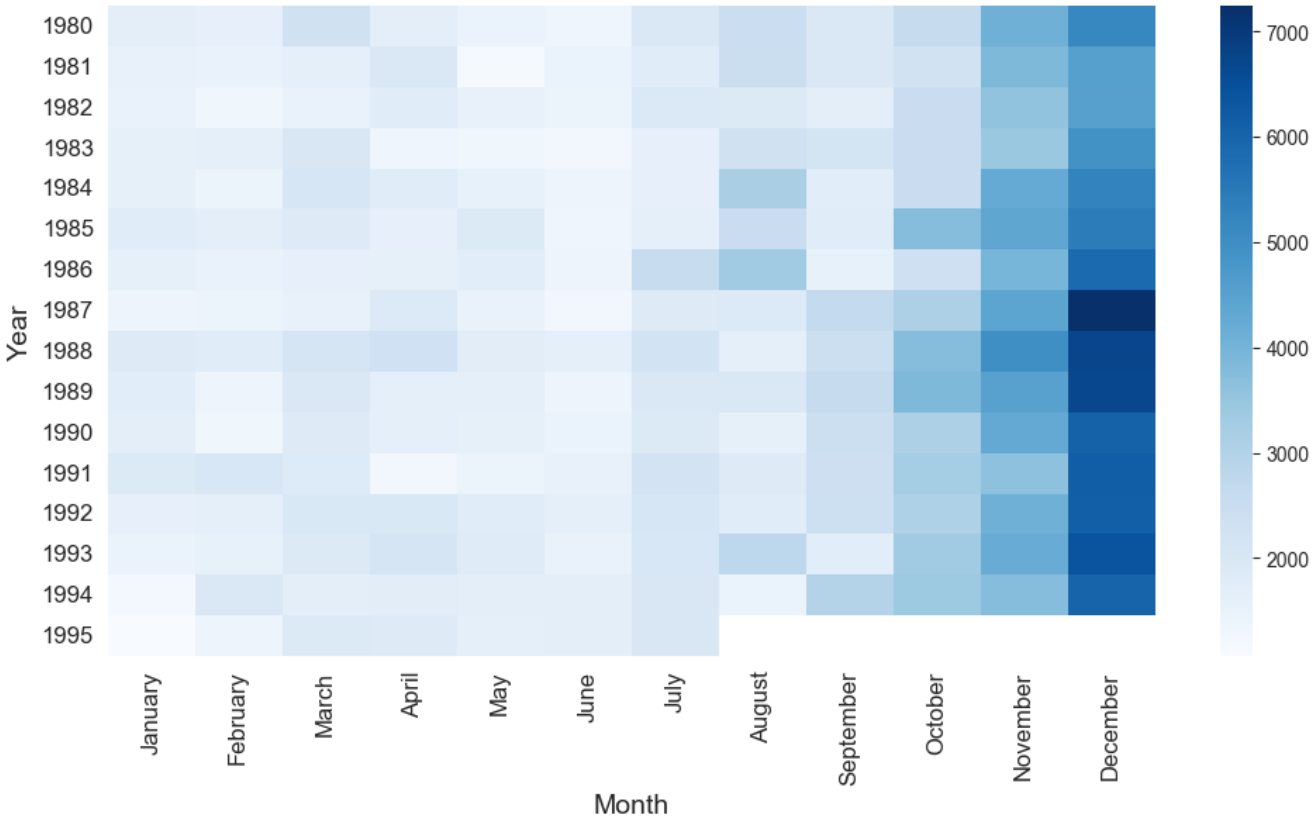


3.1.2 - Boxplots by Year

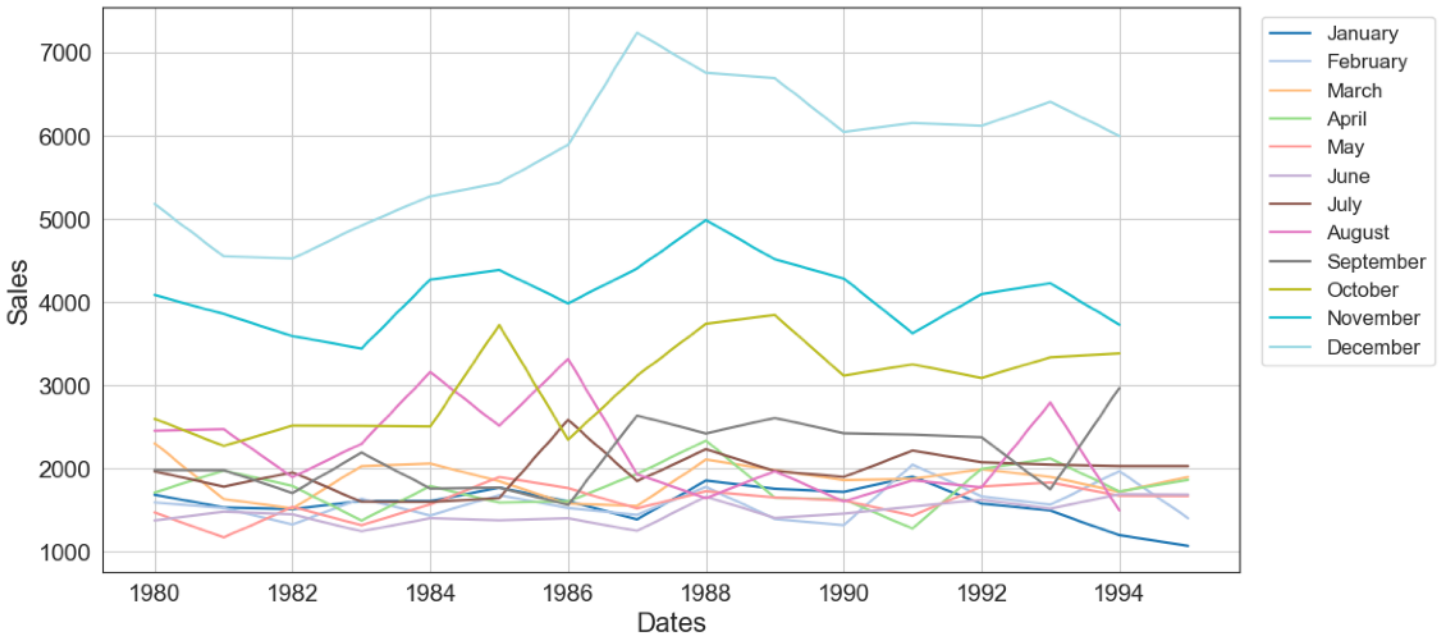
3.1.2 Pivot Table Analysis

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

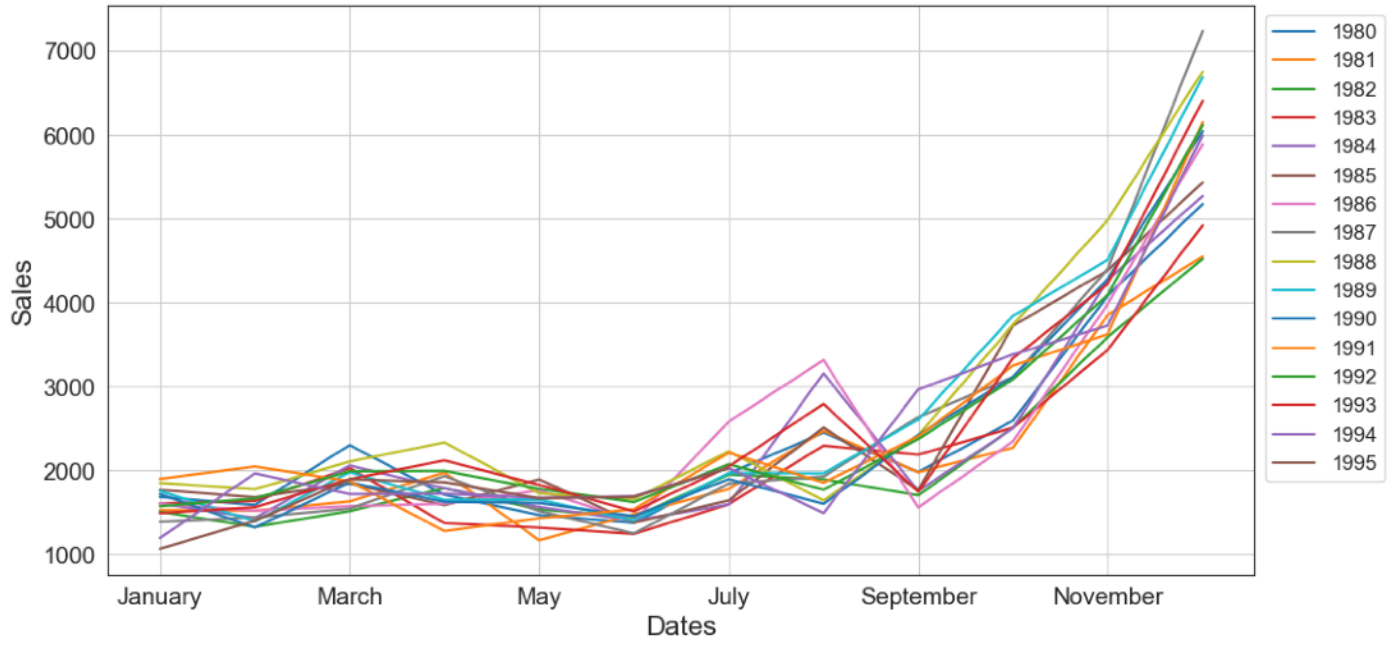
3.1.3 - Pivot Table - Month vs Year



3.1.4 - Heatmap of Pivot Table

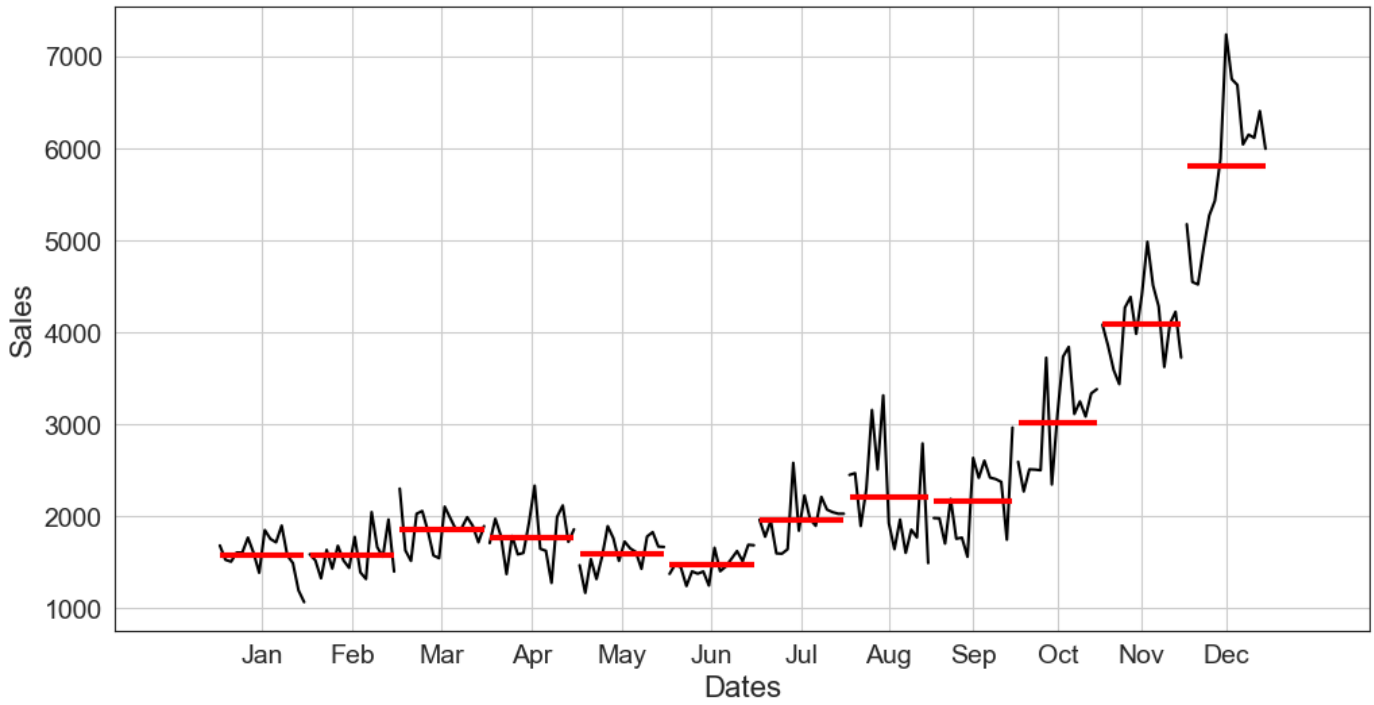


3.1.5 - Monthly Sales by Year

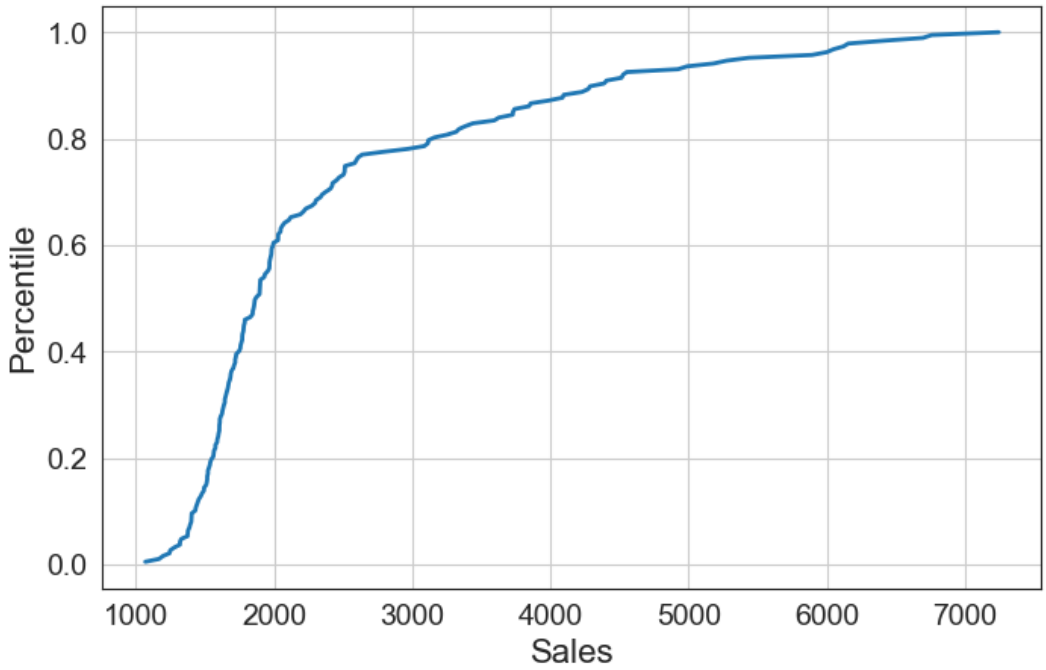


3.1.6 - Yearly Sales by Month

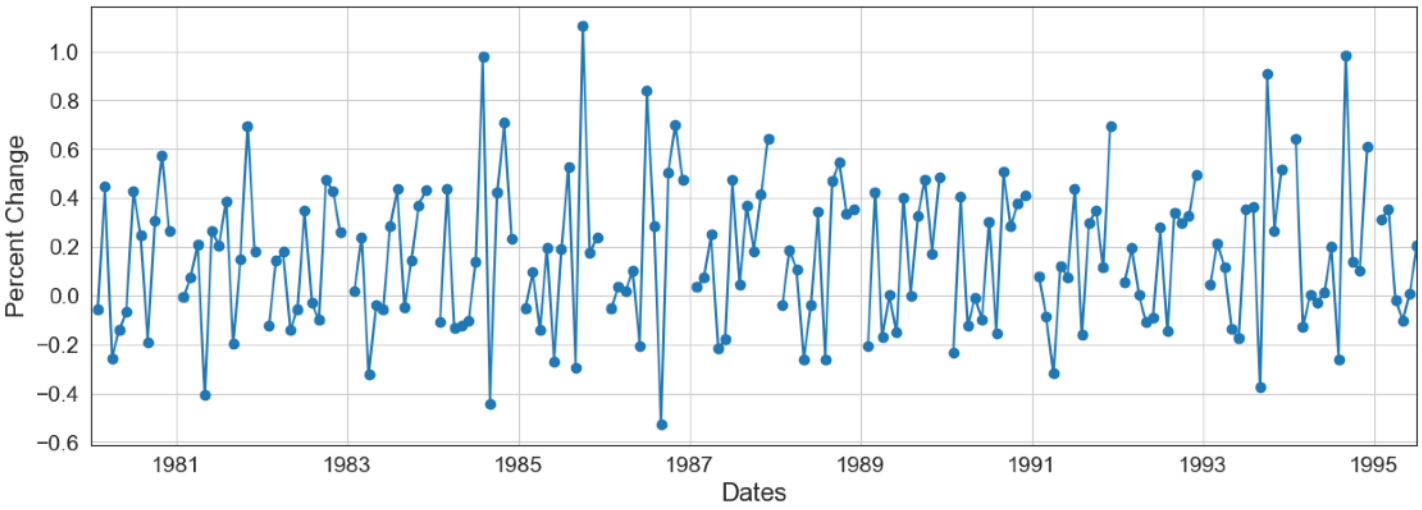
3.1.3 Other Plots



3.1.7 - Month Plot

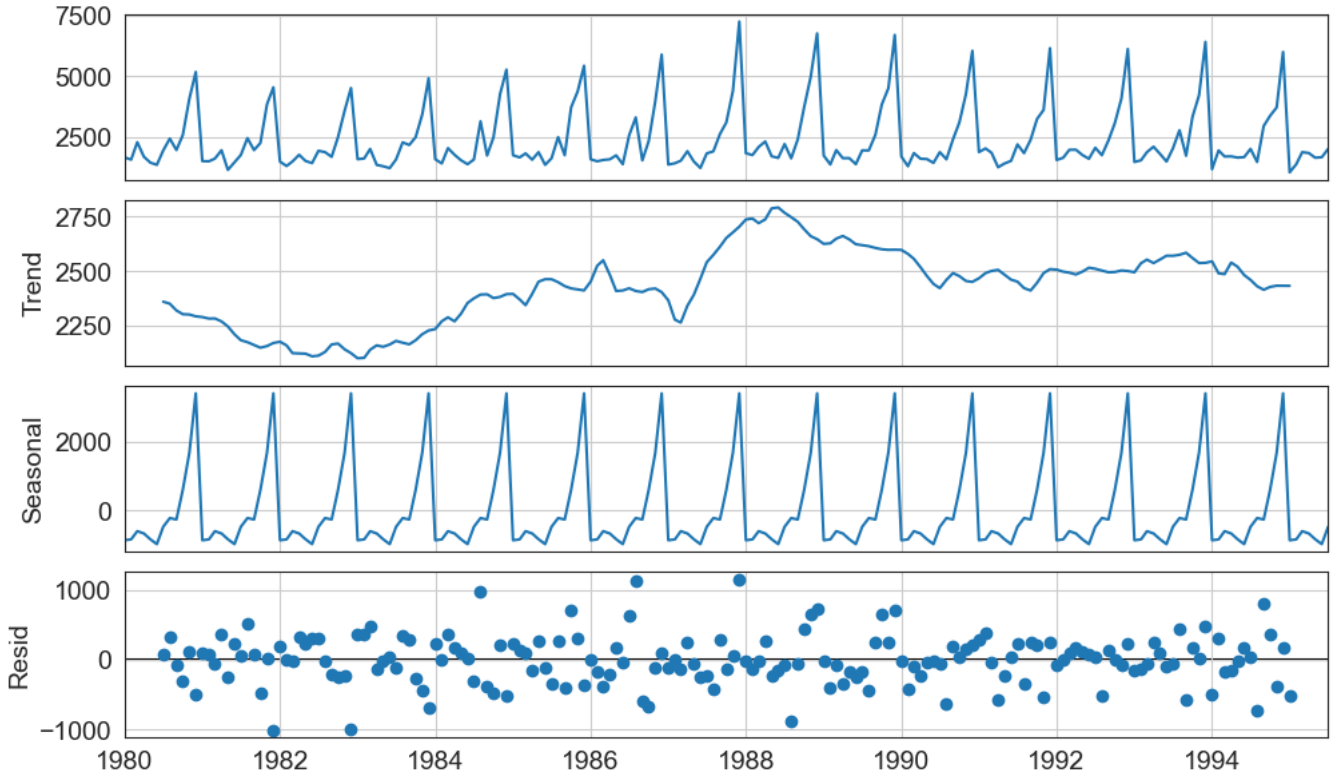


3.1.8 - ECDF Plot

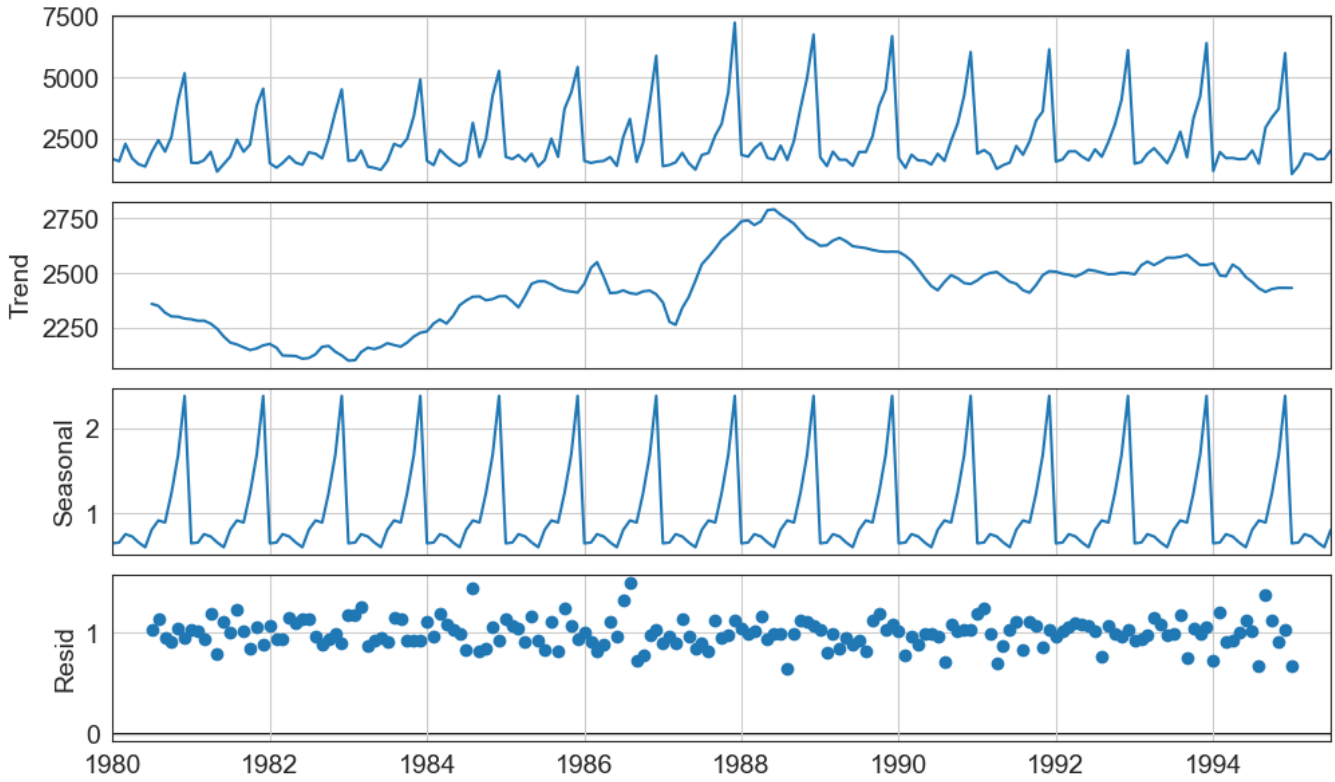


3.1.9 - Percentage Change Plot

3.1.4 Decomposition



3.1.10 - Decomposition Additive

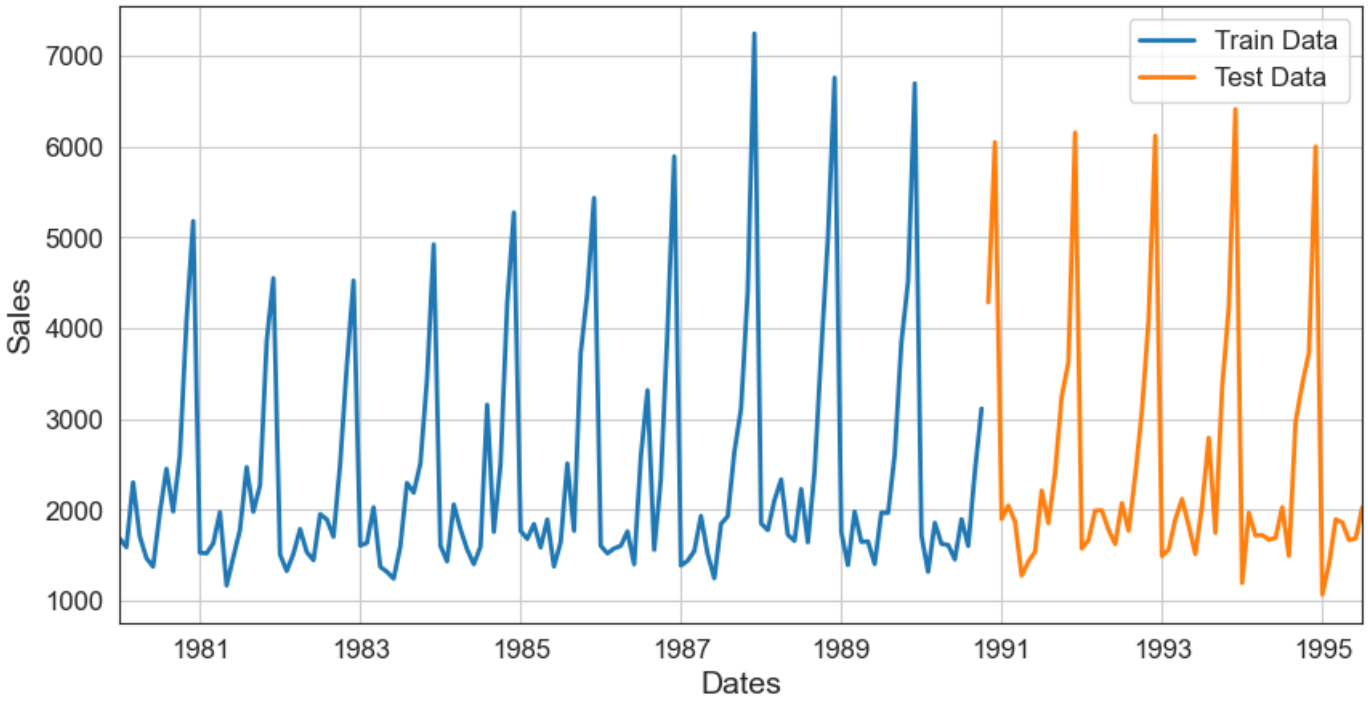


3.1.11 - Decomposition Multiplicative

3.2 Model Building - Original Data

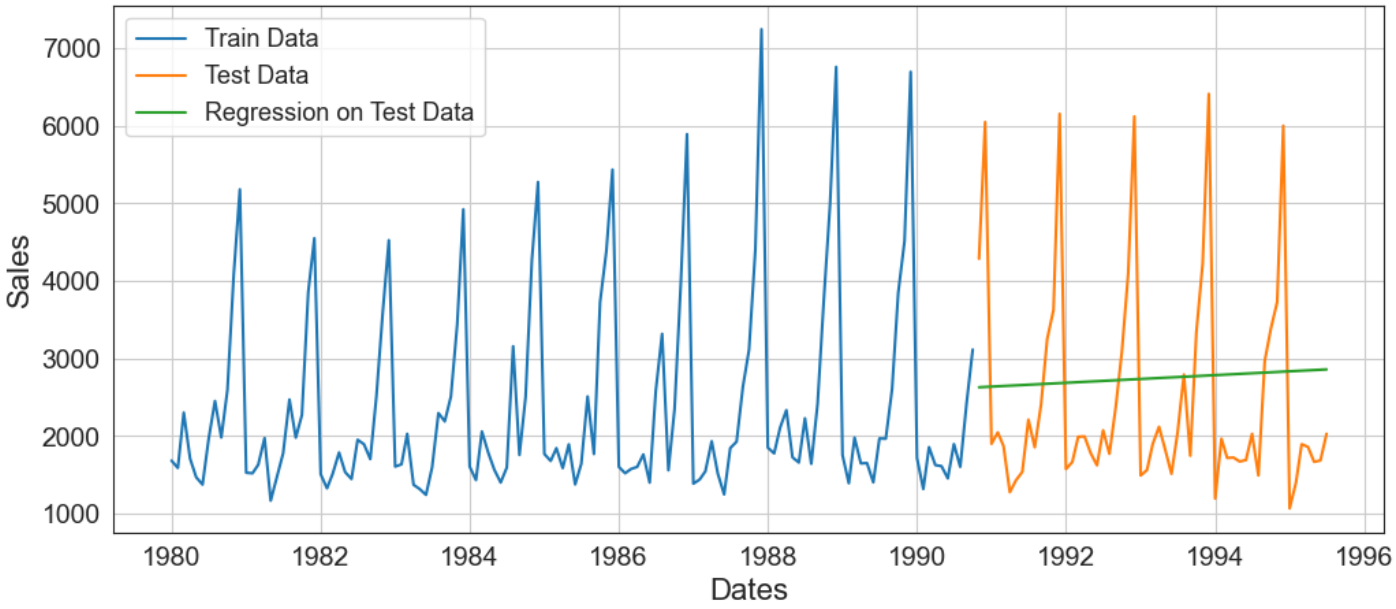
3.2.1 Train Test Split

The data was split in 70-30 ratio.



3.2.1 - Train Test Split

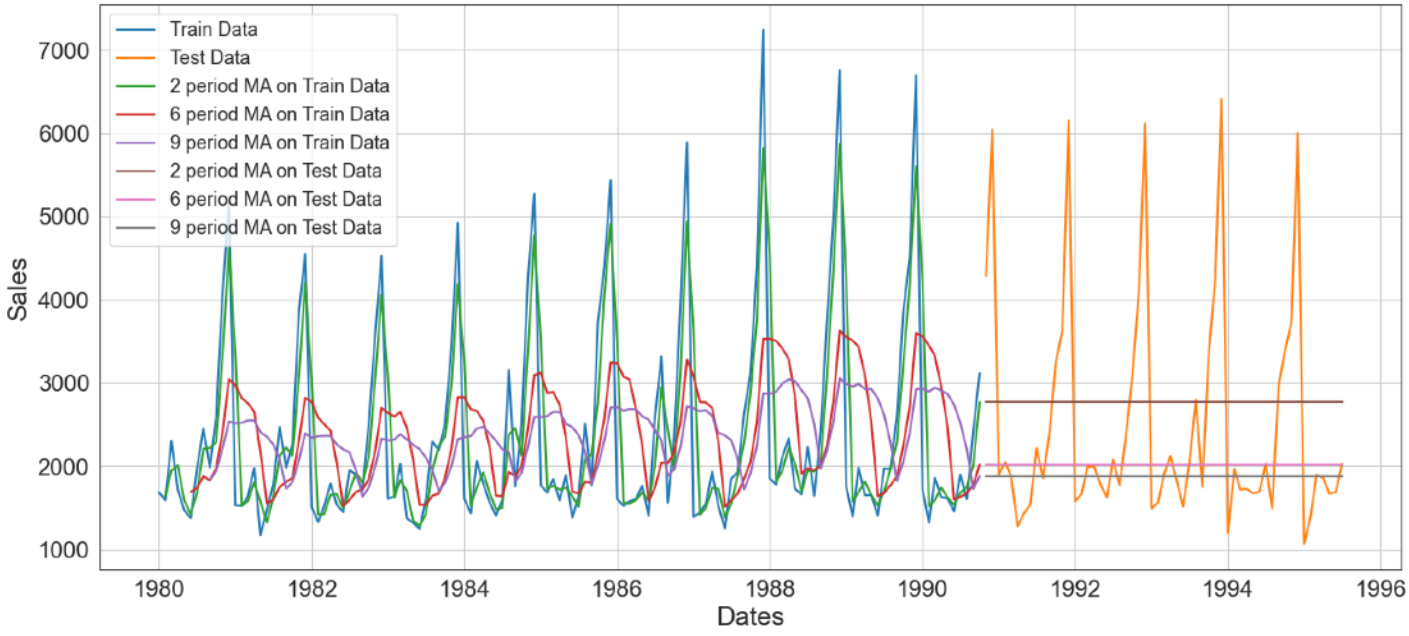
3.2.2 Linear Regression



3.2.2 - Linear Regression on Test data

For Linear Regression on Test Data the RMSE is 1392.44

3.2.3 Moving Averages



3.2.3 - Moving Averages plot on Test data

For 2 point MA the RMSE is 1389.31

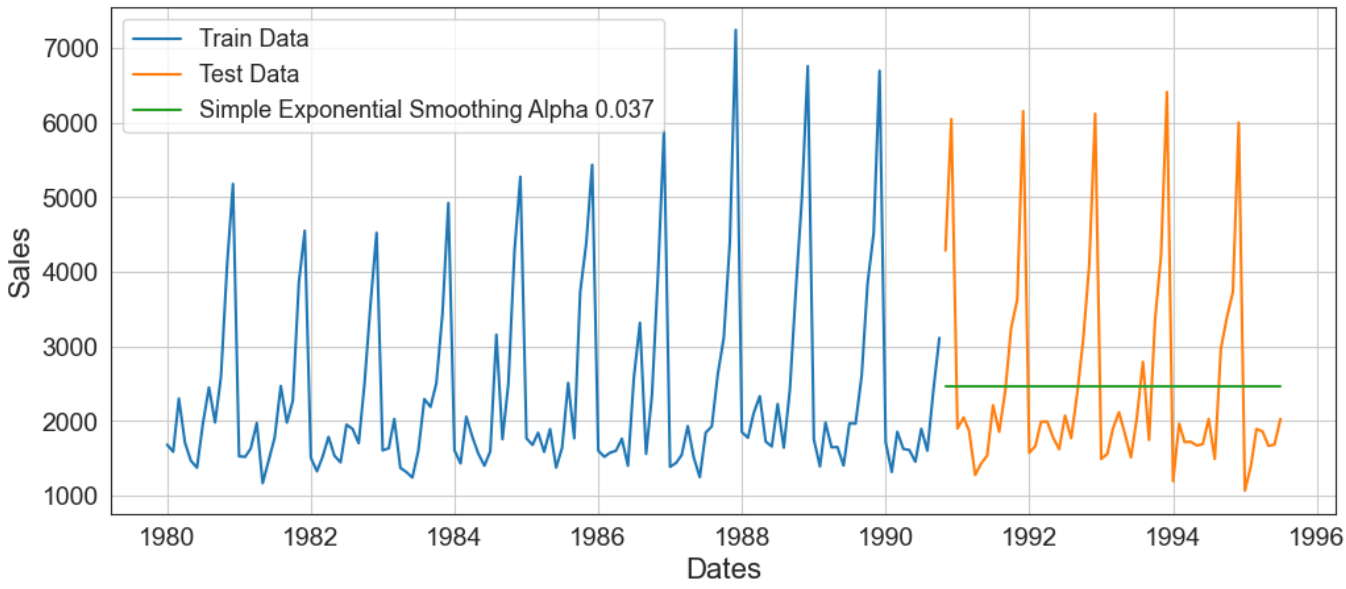
For 6 point MA the RMSE is 1443.15

For 9 point MA the RMSE is 1494.81

3.2.4 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}
```

3.2.4 - SES Autofit Parameters



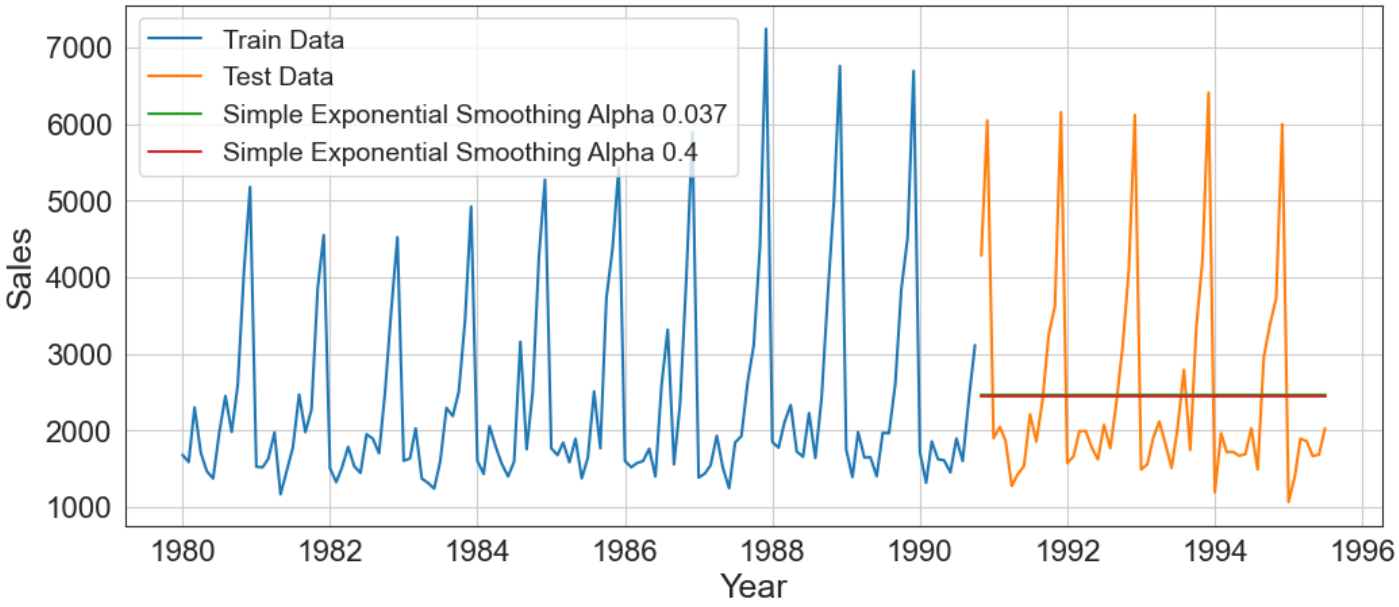
3.2.5 - SES Plot on Test data

For Simple Exponential Smoothing Alpha 0.037 the RMSE is 1362.43

3.2.4.1 SES Refined

	Alpha	Train RMSE	Test RMSE
3	0.4	1329.814823	1363.037803
4	0.5	1326.403864	1364.863549
0	0.1	1298.211536	1367.395642
2	0.3	1331.102204	1372.323705
1	0.2	1322.658289	1378.320562
5	0.6	1325.588422	1379.988733
6	0.7	1329.257530	1404.659104
7	0.8	1337.879425	1434.578214
8	0.9	1351.645478	1466.179706
9	1.0	1371.122286	1496.444629

3.2.6 - Selecting Alpha based on RMSE



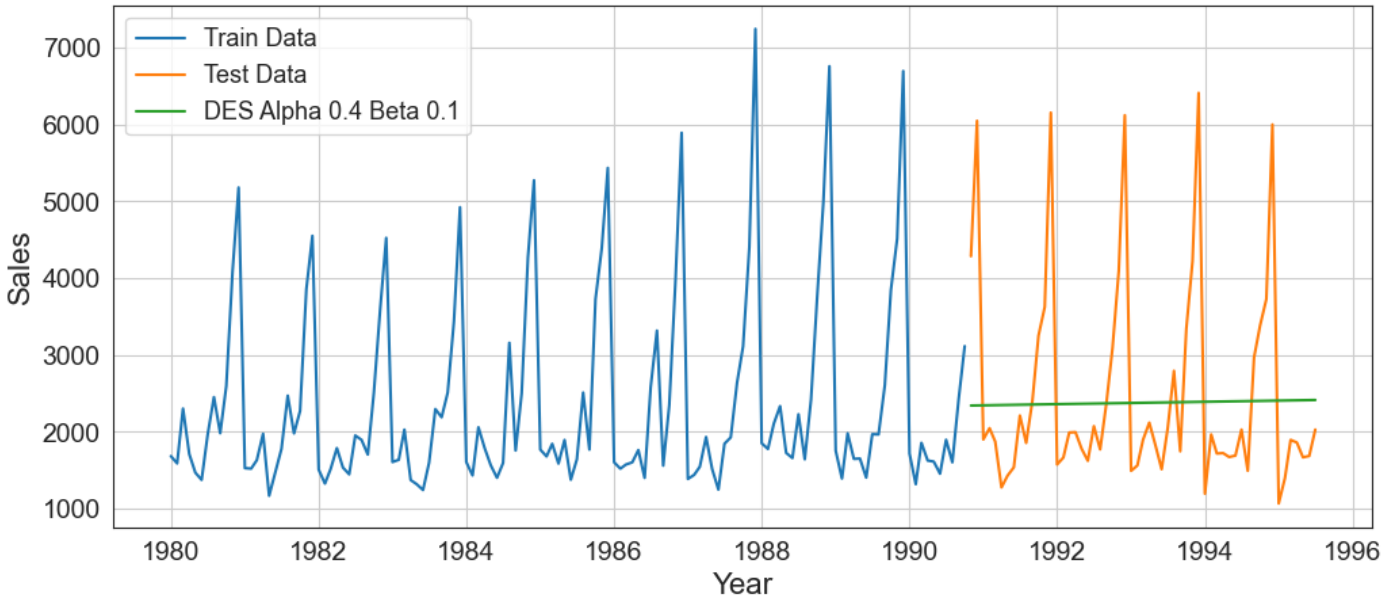
3.2.7 - Plotting both SES models

3.2.5 Double Exponential Smoothing

	Alpha	Beta	Train RMSE	Test RMSE
30	0.4	0.1	1402.987160	1369.139116
21	0.3	0.2	1479.936342	1516.785129
15	0.2	0.6	1798.093767	1540.320476
40	0.5	0.1	1396.300778	1585.868237
22	0.3	0.3	1567.524066	1597.853999
...	...	...	...	...
69	0.7	1.0	1829.175506	26841.074837
79	0.8	1.0	1885.669827	27176.057077
27	0.3	0.8	1925.999079	29603.277989
29	0.3	1.0	1883.511575	33015.522624
28	0.3	0.9	1915.332971	33043.719889

100 rows × 4 columns

3.2.8 - Selecting Alpha and Beta based on RMSE

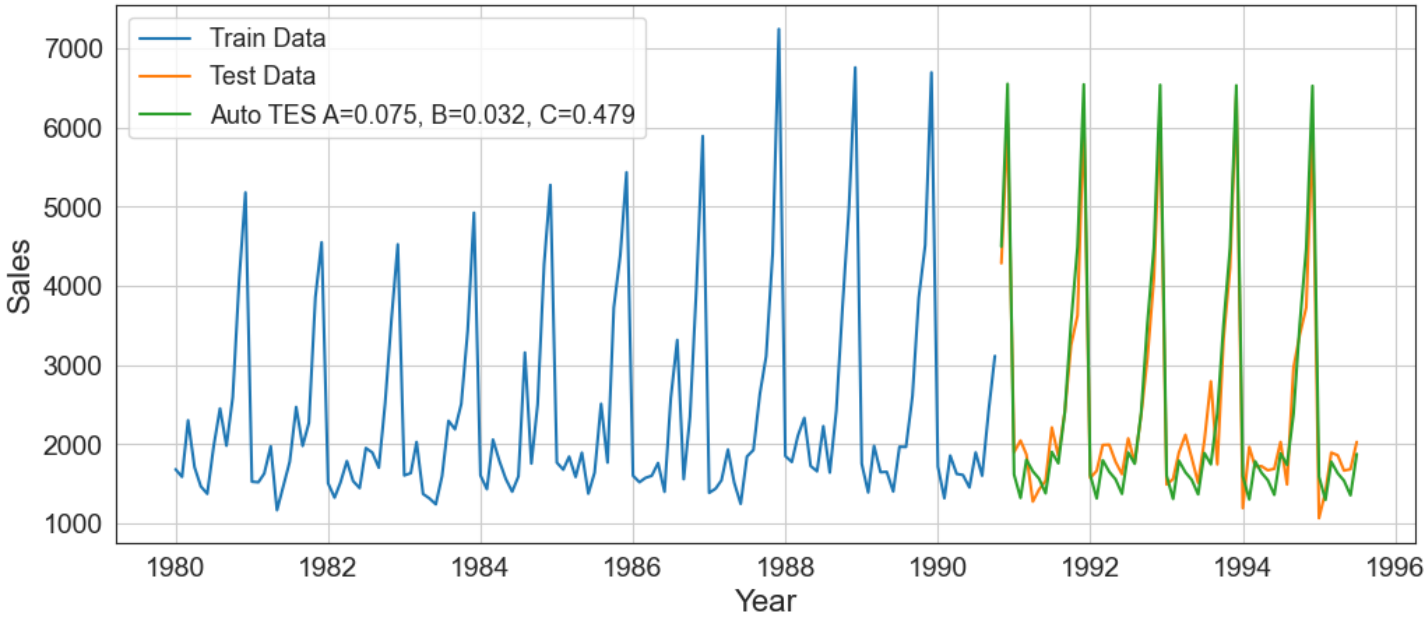


3.2.9 - DES Plot on Test data

3.2.6 Triple Exponential Smoothing

```
{'smoothing_level': 0.07569306567652224,
 'smoothing_trend': 0.03243079353506148,
 'smoothing_seasonal': 0.4791359397032058,
 'damping_trend': nan,
 'initial_level': 2356.526392240251,
 'initial_trend': -0.7790750851036652,
 'initial_seasons': array([-636.24268489, -722.99133043, -398.63354546, -473.44405847,
        -808.44186274, -815.36460197, -384.23754735,   72.99614946,
        -237.45739424,  272.31927729, 1541.39485453, 2590.08852084]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

3.2.10 - TES Autofit Parameters



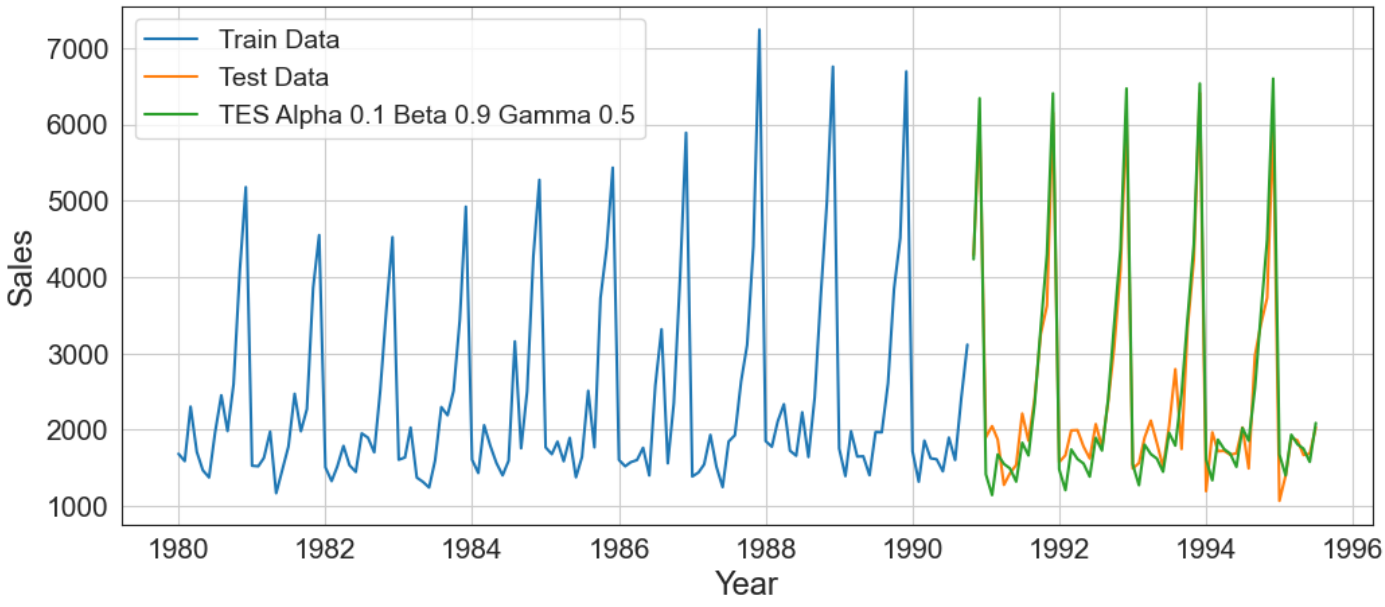
3.2.11 - TES Autofit plot on Test data

3.2.6.1 Refining TES

	Alpha	Beta	Gamma	Train RMSE	Test RMSE
84	0.1	0.9	0.5	437.48	353.28
93	0.1	1.0	0.4	437.15	378.55
600	0.7	0.1	0.1	494.13	381.17
801	0.9	0.1	0.2	501.50	399.94
410	0.5	0.2	0.1	493.54	405.09
...	...	...	...	...	...
988	1.0	0.9	0.9	355660.95	7597638.87
979	1.0	0.8	1.0	404701.27	8322372.88
989	1.0	0.9	1.0	1136914.67	26150971.98
998	1.0	1.0	0.9	1085896.81	26642032.99
999	1.0	1.0	1.0	3404261.19	87098191.40

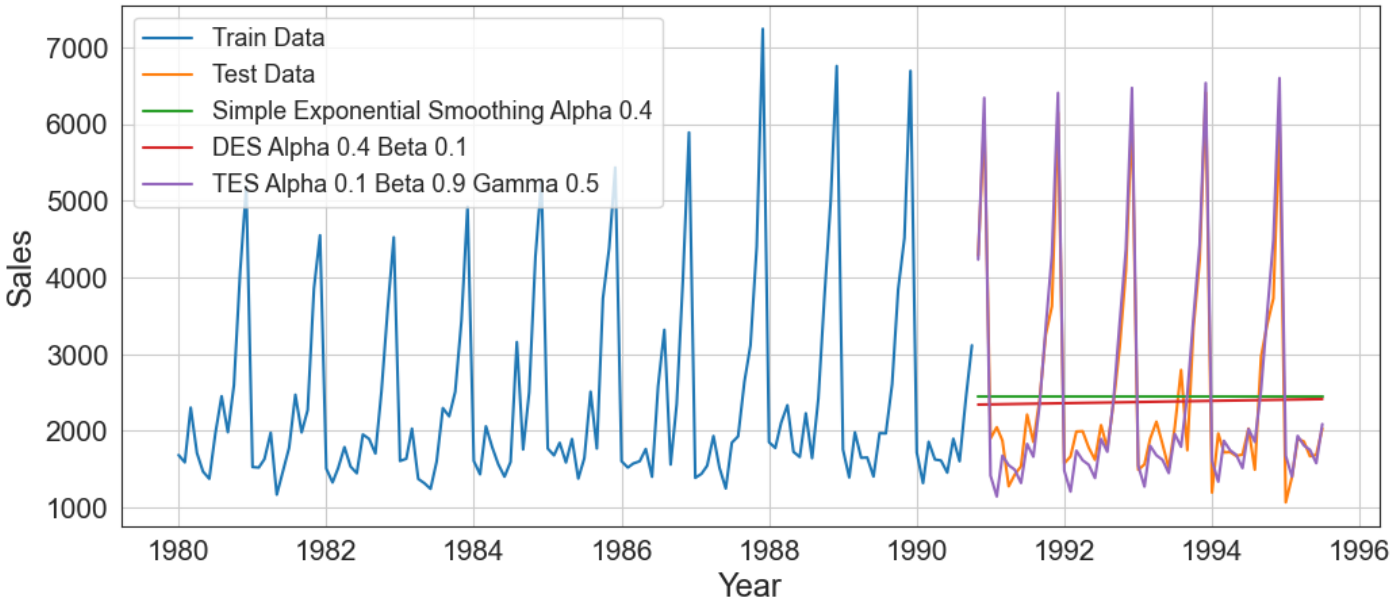
1000 rows × 5 columns

3.2.12 - Selecting Parameters based on RMSE



3.2.13 - Refined TES plot on Test data

3.2.7 Plotting all 3 Together



3.2.14 - Plotting all 3 Exponential Smoothing models

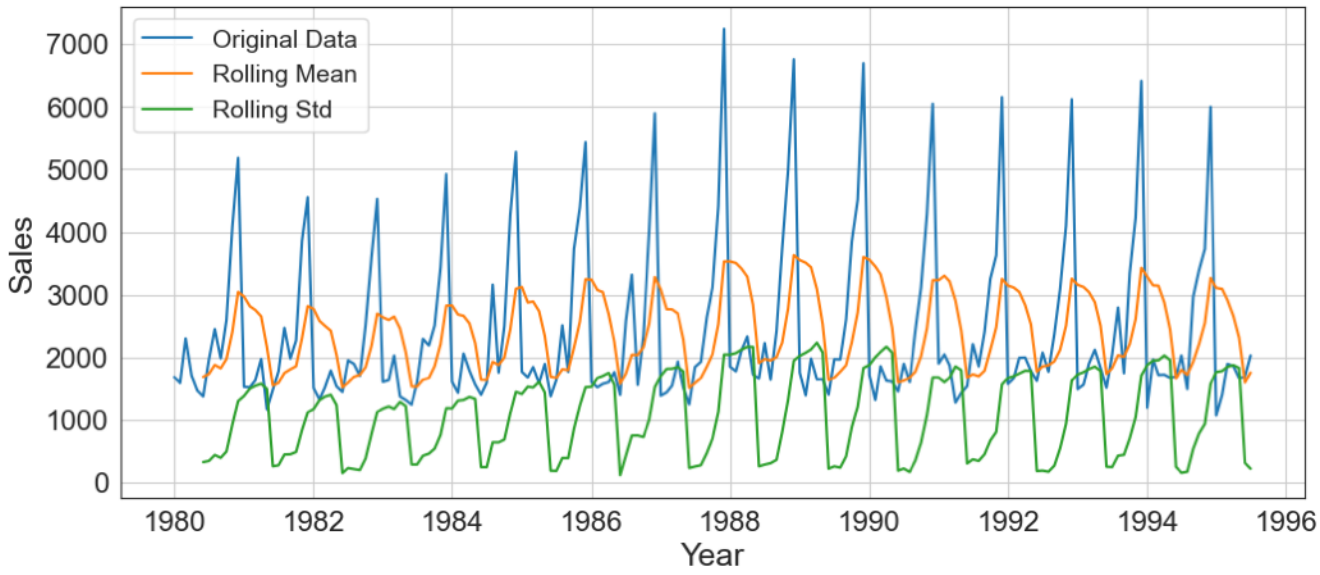
3.3 Model Building - Stationary Data

3.3.1 Check for Stationarity

```
(-1.3604974548123367,
 0.6010608871634855,
 11,
 175,
 {'1%': -3.4682803641749267,
  '5%': -2.8782017240816327,
  '10%': -2.5756525795918366},
 2573.1222090270685)
```

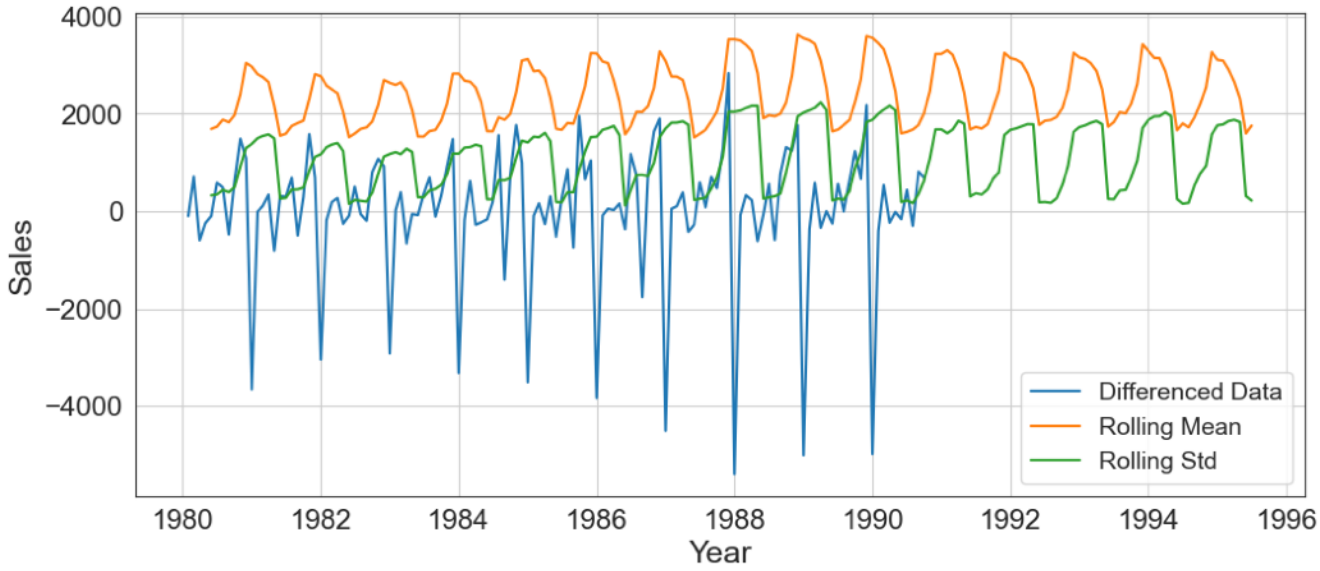
3.3.1 - ADF Test results

The p-value is 0.60. Data is not Stationary.



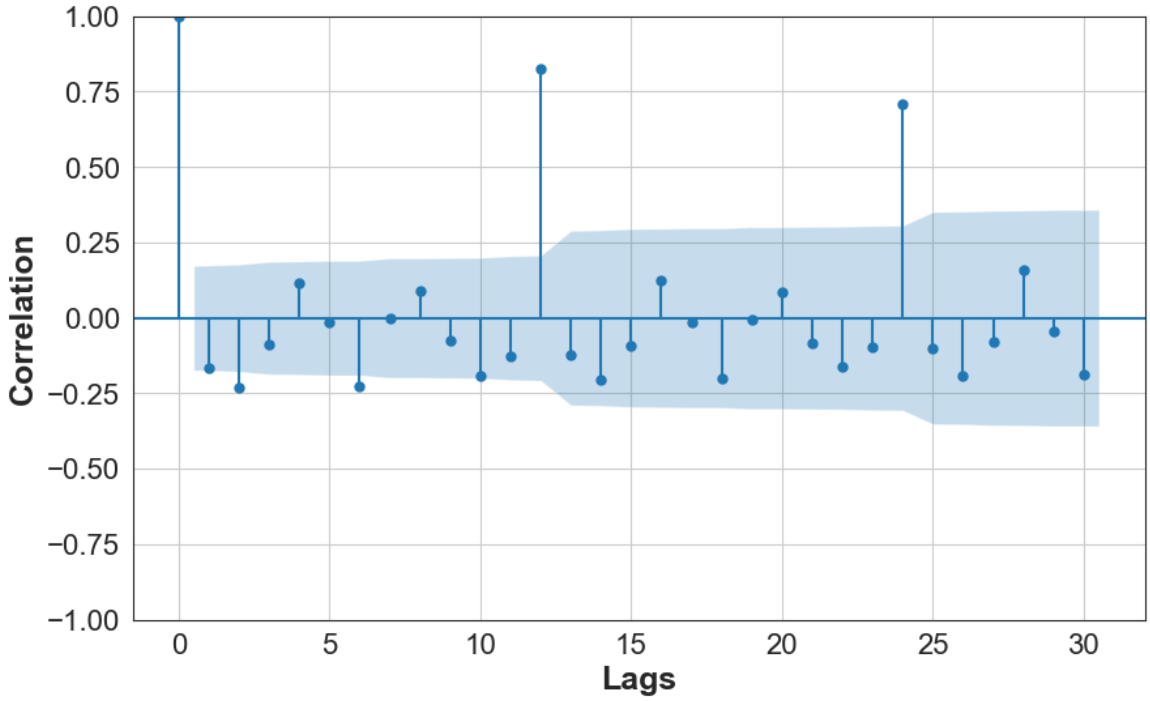
3.3.2 - Rolling Mean and Rolling Std. of entire data before differencing

After differencing the p-value is 0.00. Data becomes Stationary.

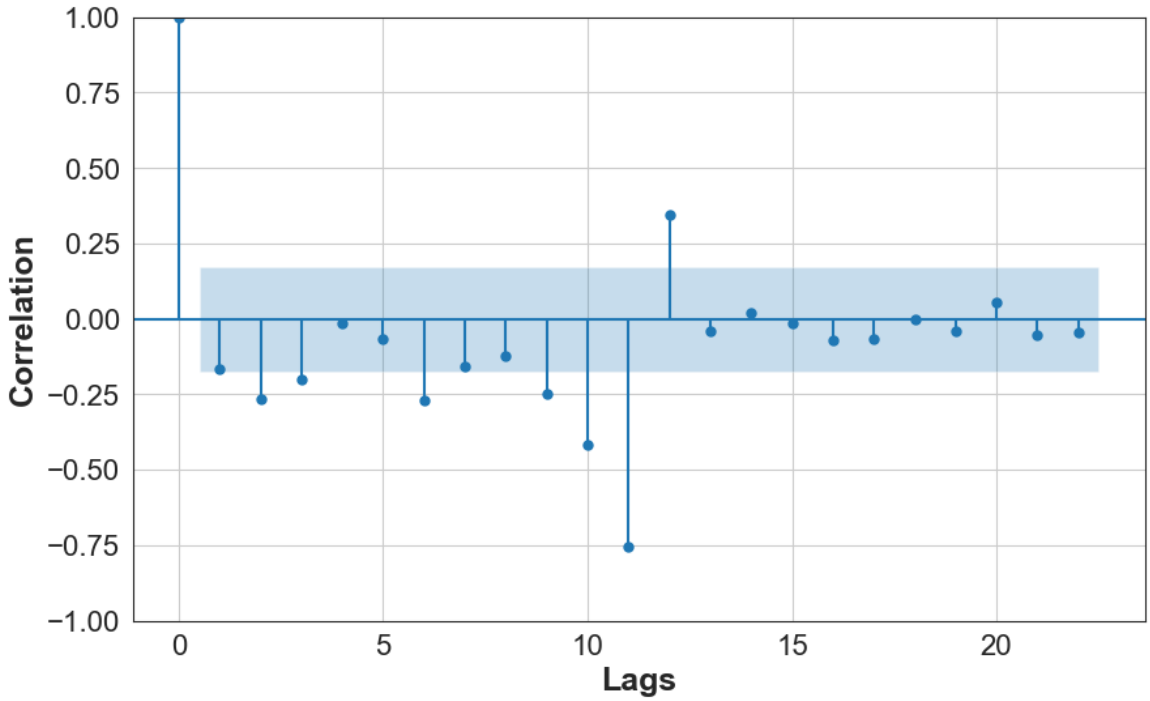


3.3.3 - Rolling Mean and Rolling Std. of entire data after differencing

3.3.2 ACF & PACF Plots



3.3.4 - ACF Plot of differenced data



3.3.5 - PACF Plot of differenced data

3.3.3 ARIMA Model

	param	AIC
14	(2, 0, 2)	2178.848194
17	(2, 1, 2)	2187.221700
11	(1, 1, 2)	2190.728278
13	(2, 0, 1)	2193.868655
2	(0, 0, 2)	2193.882439

3.3.6 - Selecting pdq
based on AIC values

SARIMAX Results

Dep. Variable:	Sales	No. Observations:	129
Model:	ARIMA(2, 1, 2)	Log Likelihood	-1088.611
Date:	Tue, 09 Jul 2024	AIC	2187.222
Time:	19:35:21	BIC	2201.482
Sample:	02-01-1980	HQIC	2193.016
	- 10-01-1990		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4934	0.107	4.604	0.000	0.283	0.703
ar.L2	-0.1783	0.200	-0.894	0.372	-0.570	0.213
ma.L1	-1.9930	0.112	-17.727	0.000	-2.213	-1.773
ma.L2	0.9949	0.112	8.886	0.000	0.775	1.214
sigma2	1.316e+06	1.84e-07	7.14e+12	0.000	1.32e+06	1.32e+06

Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	15.15
Prob(Q):	0.79	Prob(JB):	0.00
Heteroskedasticity (H):	2.49	Skew:	0.51
Prob(H) (two-sided):	0.00	Kurtosis:	4.34

3.3.7 - Selected model results

For ARIMA(2,0,2) the RMSE is 1340.63

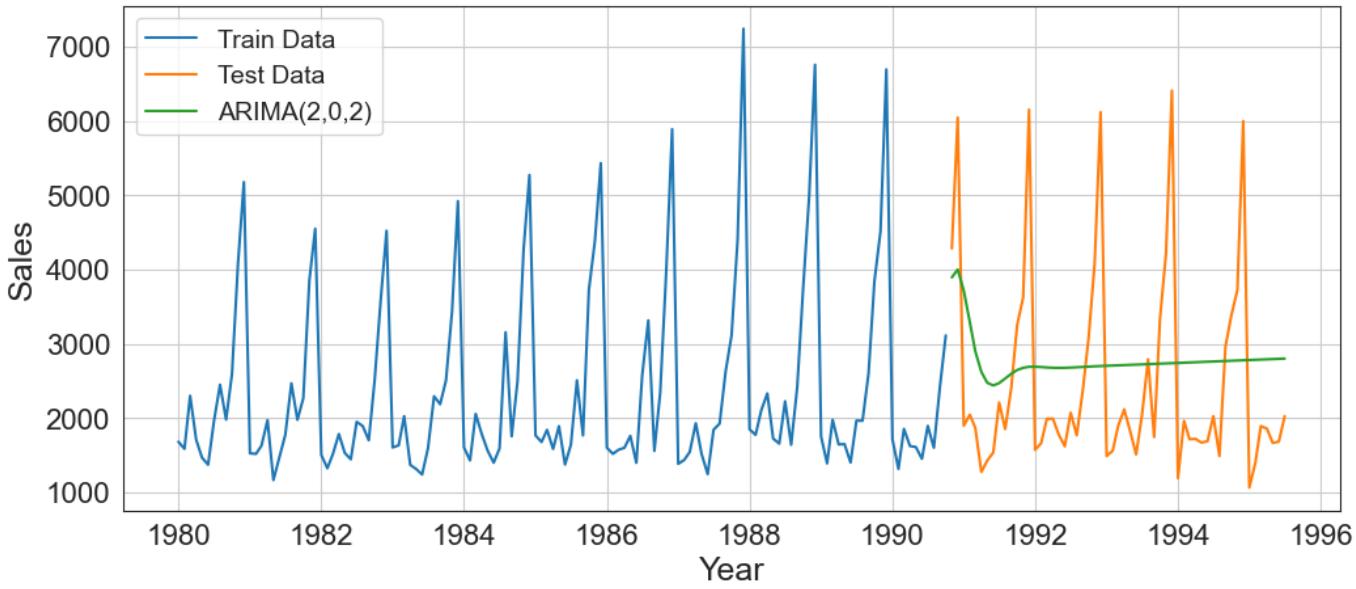
```
1990-11-01    779.899158
1990-12-01    106.939963
1991-01-01   -283.187034
1991-02-01   -423.128626
1991-03-01   -392.806438
Freq: MS, Name: predicted_mean

1995-03-01    3.245939
1995-04-01    3.245893
1995-05-01    3.245887
1995-06-01    3.245904
1995-07-01    3.245929
Freq: MS, Name: predicted_mean
```

3.3.8 - Forecast values on differenced data

```
1990-11-01    779.899158
1990-12-01    886.839122
1991-01-01    603.652088
1991-02-01    180.523462
1991-03-01   -212.282976
1991-04-01   -490.133913
1991-05-01   -635.811917
1991-06-01   -672.881018
1991-07-01   -641.120728
1991-08-01   -579.191993
Freq: MS, Name: predicted_mean
```

3.3.9 - Converting back to original values



3.3.10 - Plot of ARIMA model

3.3.4 SARIMA Model

	param	param_seasonal	AIC
50	(0, 1, 2)	(1, 0, 2, 12)	1518.30
53	(0, 1, 2)	(2, 0, 2, 12)	1518.54
158	(2, 1, 2)	(1, 0, 2, 12)	1520.03
104	(1, 1, 2)	(1, 0, 2, 12)	1520.21
107	(1, 1, 2)	(2, 0, 2, 12)	1520.31
...	...	...	...
54	(1, 0, 0)	(0, 0, 0, 12)	2214.85
0	(0, 0, 0)	(0, 0, 0, 12)	2216.42
135	(2, 1, 0)	(0, 0, 0, 12)	2240.11
81	(1, 1, 0)	(0, 0, 0, 12)	2277.31
27	(0, 1, 0)	(0, 0, 0, 12)	2307.36

162 rows × 3 columns

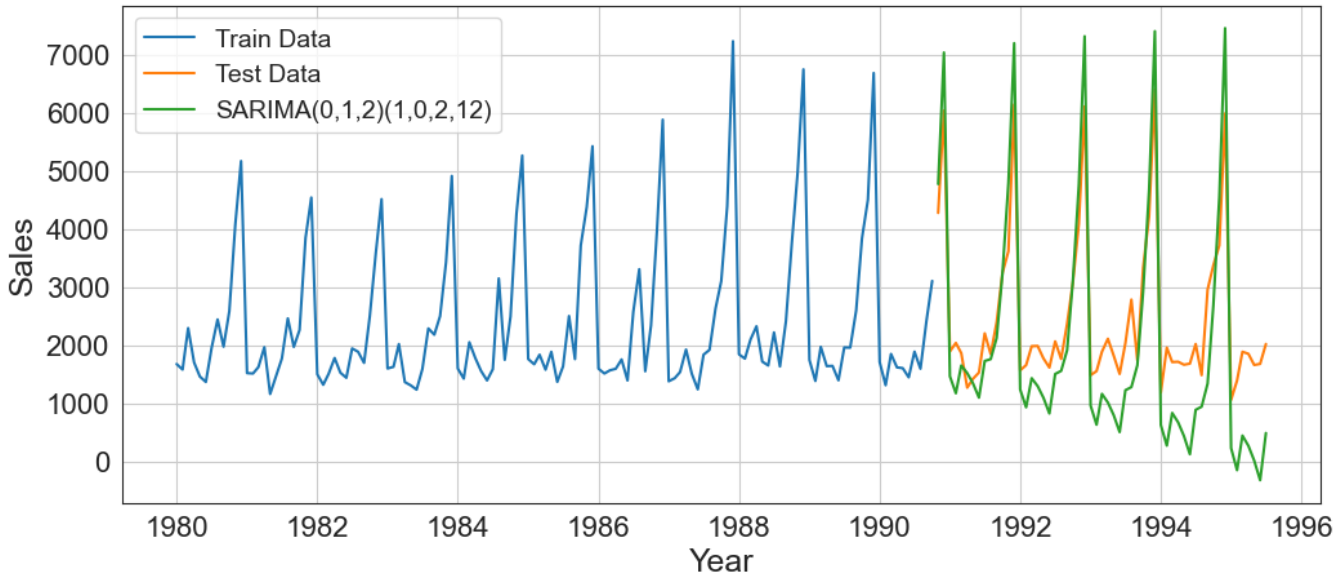
3.3.11 - Experimenting to get the best parameters based on AIC

Parameters (0,1,2)(1,0,2,12) are chosen.

RMSE for SARIMA(0,1,2)(2,0,2,12) is 961.27

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                Sales   No. Observations:                  129
Model:             SARIMAX(0, 1, 2)x(1, 0, 2, 12)   Log Likelihood                -753.187
Date:                            Tue, 09 Jul 2024   AIC                           1518.374
Time:                                    19:38:40   BIC                           1534.065
Sample:                                01-01-1980   HQIC                          1524.726
                                     - 10-01-1990
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -1.8932      0.054    -35.172      0.000      -1.999      -1.788
ma.L2          0.8985      0.050     18.129      0.000       0.801       0.996
ar.S.L12       1.0658      0.016     68.346      0.000       1.035       1.096
ma.S.L12      -0.6709      0.085     -7.856      0.000      -0.838      -0.504
ma.S.L24      -0.0448      0.111     -0.402      0.688      -0.263       0.173
sigma2      1.56e+05   2.06e+04      7.562      0.000    1.16e+05    1.96e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.13   Jarque-Bera (JB):                 14.95
Prob(Q):                              0.72   Prob(JB):                          0.00
Heteroskedasticity (H):               1.58   Skew:                              0.28
Prob(H) (two-sided):                  0.19   Kurtosis:                          4.80
===================================================================================
```

3.3.12 - Results of chosen model



3.3.13 - Plot of SARIMA model (0,1,2)(1,0,2,12)

3.3.4.1 Making d=0 and generating another Model

Making d=0 because there's no trend in differenced data.

RMSE for SARIMA(0,0,2)(1,0,2,12) is 708.86

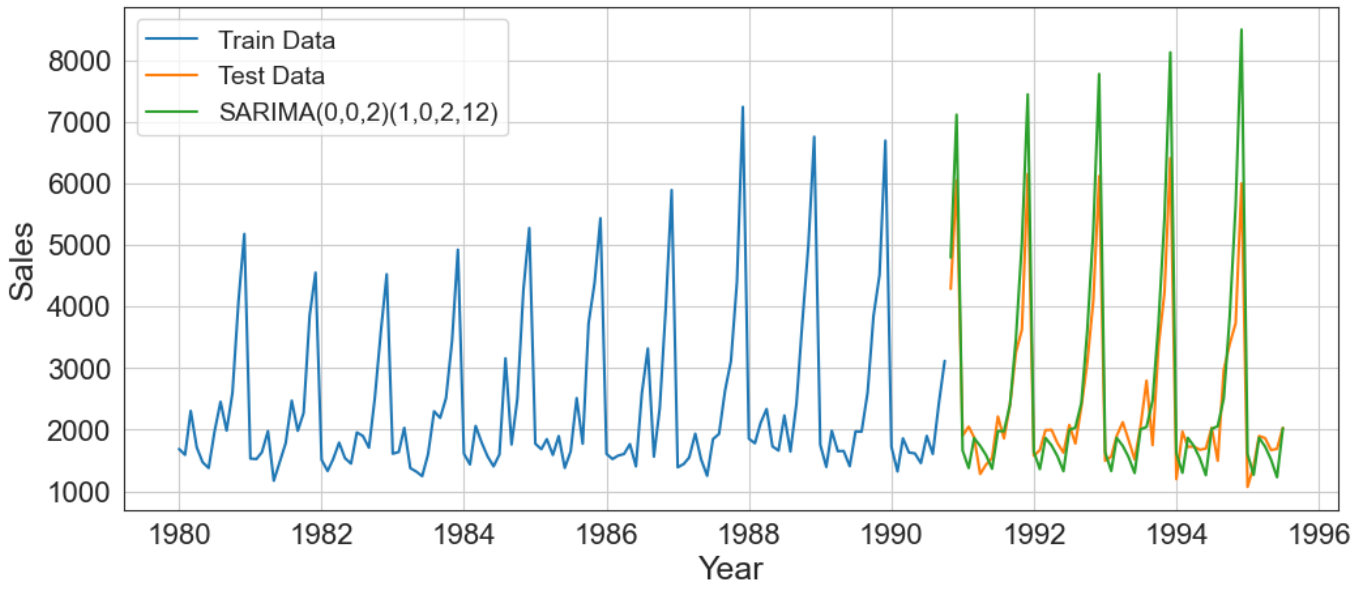
SARIMAX Results

Dep. Variable:	Sales	No. Observations:	129
Model:	SARIMAX(0, 0, 2)x(1, 0, 2, 12)	Log Likelihood	-756.361
Date:	Wed, 10 Jul 2024	AIC	1524.721
Time:	13:50:38	BIC	1540.471
Sample:	02-01-1980	HQIC	1531.099
	- 10-01-1990		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.8378	0.099	-8.451	0.000	-1.032	-0.644
ma.L2	-0.1105	0.116	-0.955	0.340	-0.337	0.116
ar.S.L12	1.0591	0.014	73.522	0.000	1.031	1.087
ma.S.L12	-1.2715	0.395	-3.220	0.001	-2.045	-0.498
ma.S.L24	-0.1663	0.160	-1.038	0.299	-0.480	0.148
sigma2	7.735e+04	2.96e+04	2.614	0.009	1.94e+04	1.35e+05

Ljung-Box (L1) (Q):	0.03	Jarque-Bera (JB):	15.08
Prob(Q):	0.87	Prob(JB):	0.00
Heteroskedasticity (H):	1.35	Skew:	0.33
Prob(H) (two-sided):	0.38	Kurtosis:	4.76

3.3.14 - Results of model with d=0



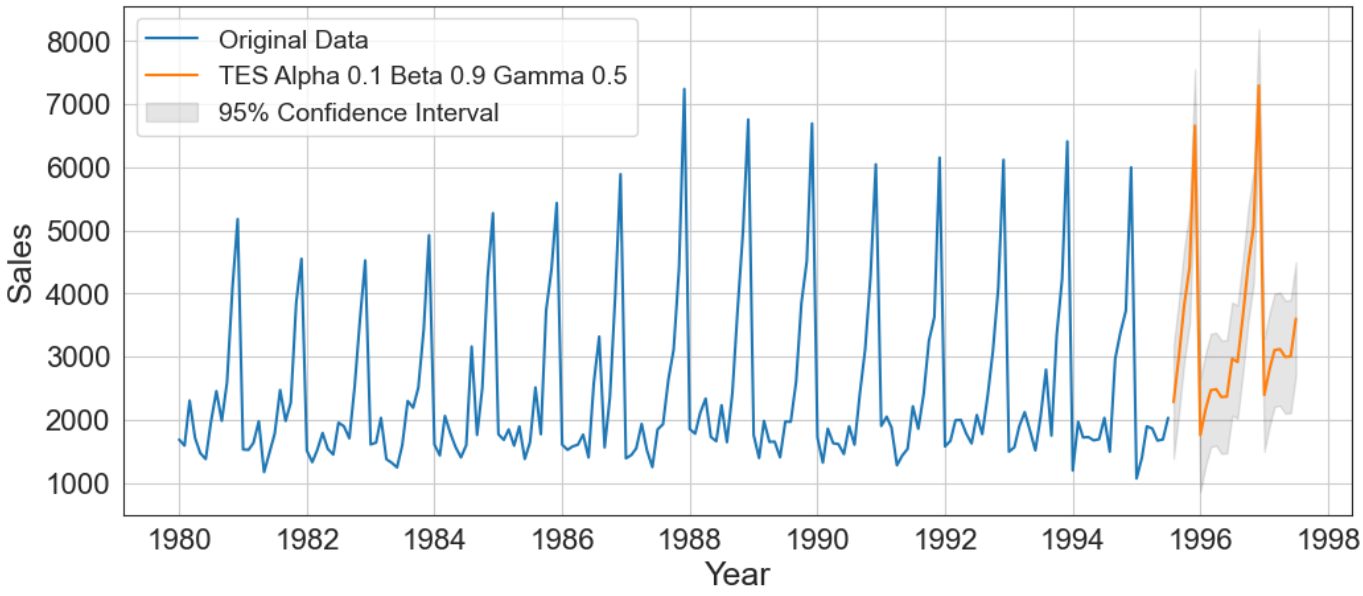
3.3.15 - Plot of alternative SARIMA model

3.4 Final Model Selection & Forecasting

Chosen Model is Triple Exponential Smoothing (Alpha 0.1, Beta 0.9, Gamma 0.5)

	Model	Test RMSE
8	TES Refined Alpha 0.1 Beta 0.9 Gamma 0.5	353.3
7	TES Auto A=0.075, B=0.032, C=0.479	366.9
11	SARIMA(0,0,2)(1,0,2,12)	708.9
10	SARIMA(0,1,2)(1,0,2,12)	961.3
9	ARIMA(2,0,2)	1340.6
4	Auto SES Alpha 0.037	1362.4
5	SES Refined Alpha 0.4	1363.0
6	DES Refined Alpha 0.4 Beta 0.1	1369.1
1	2 point MA	1389.3
0	Linear Regression	1392.4
2	6 point MA	1443.2
3	9 point MA	1494.8

3.3.16 - List of all models arranged by RMSE



3.3.17 - Forecast using chosen model

4.0 Actionable Insights & Recommendations

- Sales of Rose are really low and are projected to go lower, therefore it's recommended that it be stopped altogether and the resources used for other purposes.
- Sales of Sparkling are going strong and are projected to do either same or better. Since the sales are strong and stable, therefore efforts can be made to expand this portion of the business.