TIME SERIES FORECASTING

Aniket Ganguly

PGPDSBA – JULY 2023

SOLUTION

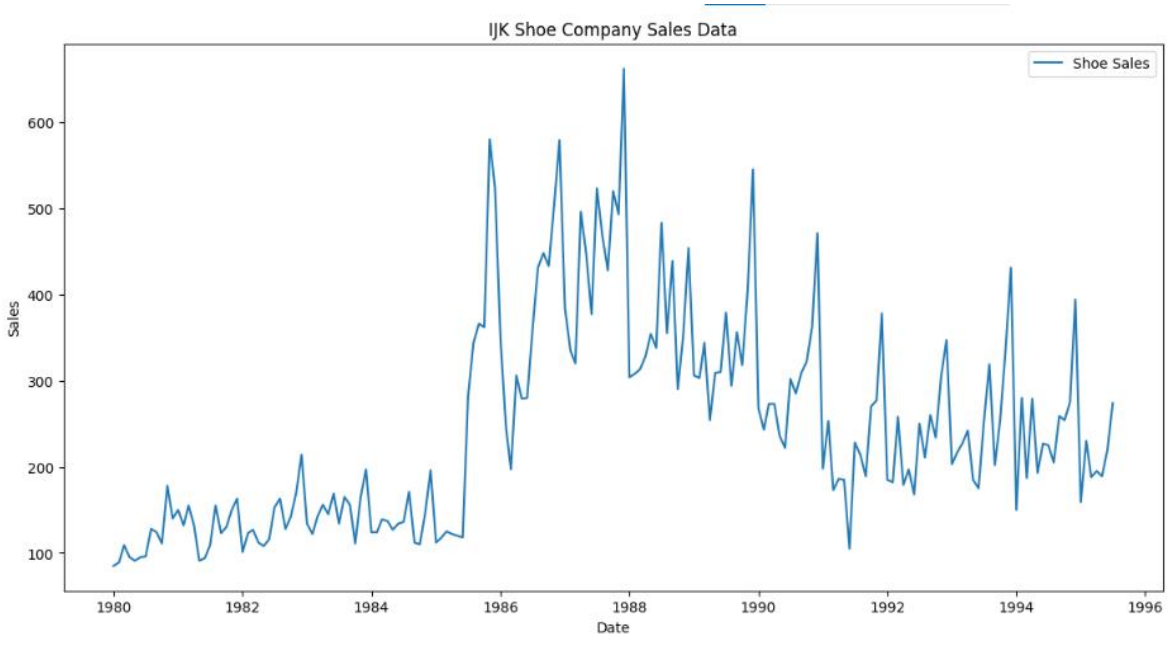
1. **Define the problem and perform Exploratory Data Analysis**

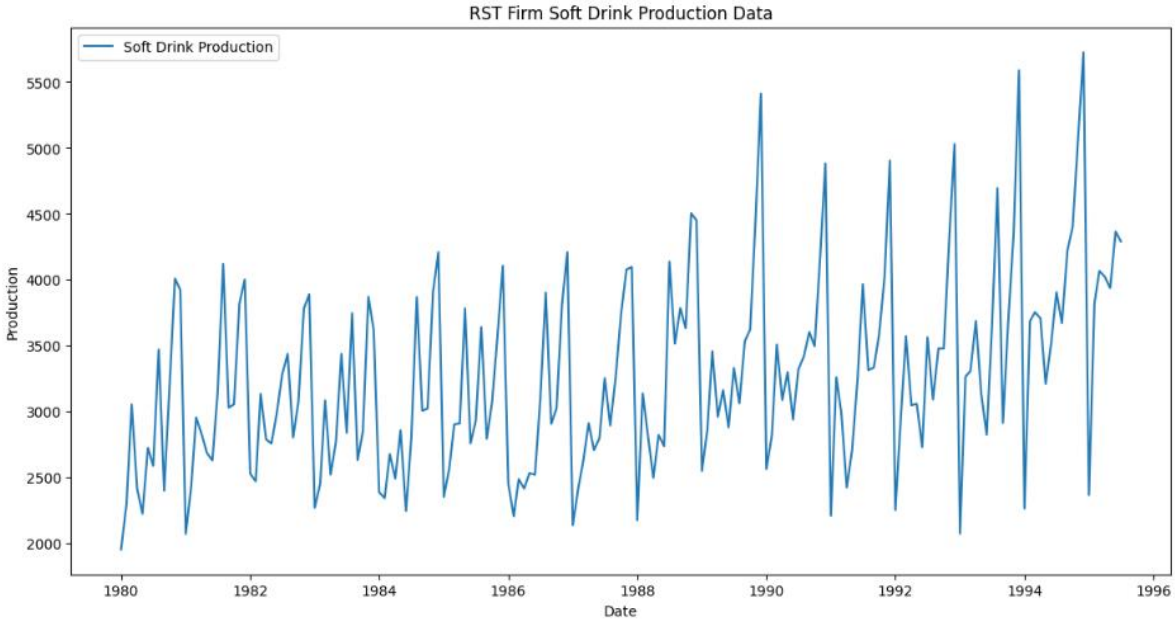
   - Read the data as an appropriate time series data - Plot the data - Perform EDA - Perform Decomposition

   - Top 5 rows of Shoe Sales data and Soft Drinks data

```
(            Shoe_Sales
YearMonth
1980-01-01          85
1980-02-01          89
1980-03-01         109
1980-04-01          95
1980-05-01          91,
             SoftDrinkProduction
YearMonth
1980-01-01                  1954
1980-02-01                  2302
1980-03-01                  3054
1980-04-01                  2414
1980-05-01                  2226)
```

   - Plotted the Data



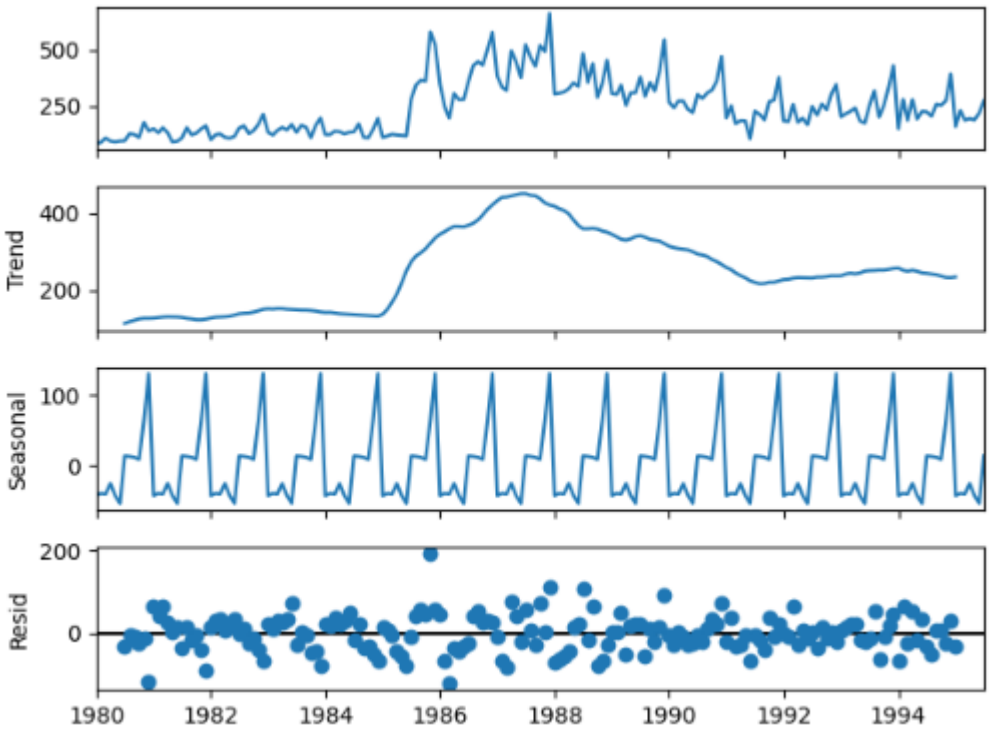


- Performed EDA

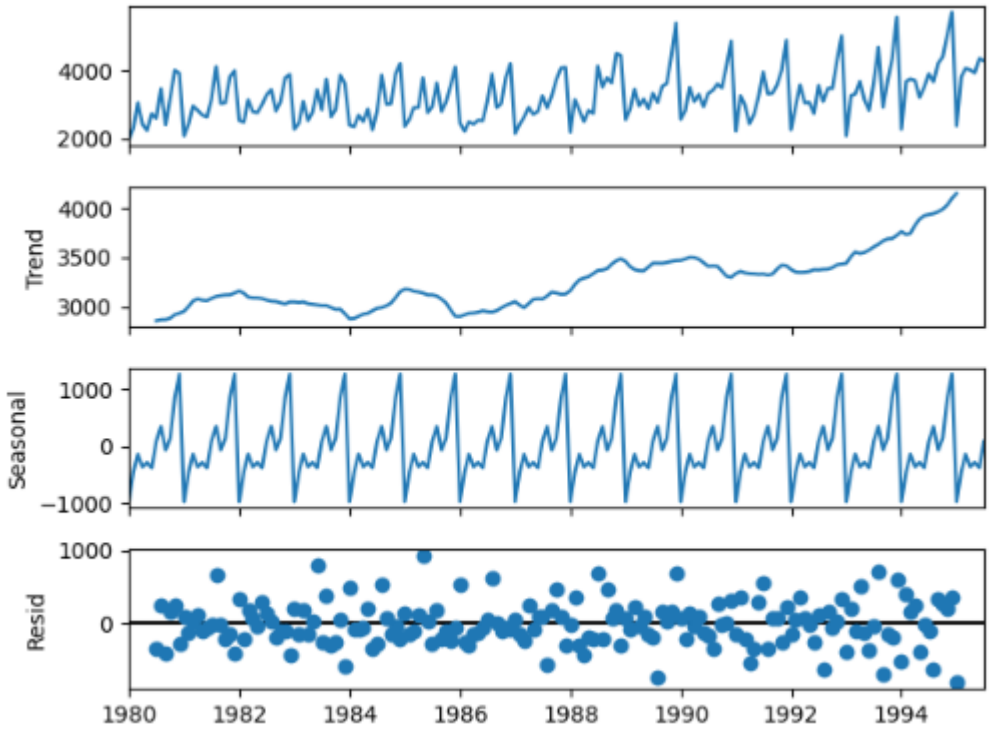
```
(Shoe_Sales    0
 dtype: int64,
 SoftDrinkProduction    0
 dtype: int64,
        Shoe_Sales
 count  187.000000
 mean   245.636364
 std    121.390804
 min     85.000000
 25%    143.500000
 50%    220.000000
 75%    315.500000
 max    662.000000,
        SoftDrinkProduction
 count           187.000000
 mean           3262.609626
 std             728.357367
 min            1954.000000
 25%            2748.000000
 50%            3134.000000
 75%            3741.000000
 max            5725.000000)
```

- Performed Decomposition

Shoe Sales data



Soft Drinks Data



2. **Data Pre-processing**
   - Missing value treatment - Visualize the processed data - Train-test split

   - Missing Values

```
Shoe_Sales    0
dtype: int64
```

```
SoftDrinkProduction    0
dtype: int64
```

   - Train-test split

```
Shoe sales training set length: 149
Shoe sales testing set length: 38
Soft drink production training set length: 149
Soft drink production testing set length: 38
```

3. **Model Building - Original Data**
   - Build forecasting models - Linear regression - Simple Average - Moving Average - Exponential Models (Single, Double, Triple) - Check the performance of the models built

```
Mean Squared Error for Linear Regression (Shoe Sales): 29324.28073336573
Mean Squared Error for Linear Regression (Soft Drink Production): 699768.9626079414
```

```
Mean Squared Error for Simple Average (Shoe Sales): 3809.0410922694327
Mean Squared Error for Simple Average (Soft Drink Production): 1051067.673177358
```

```
Mean Squared Error for Moving Average (Shoe Sales): 4738.225146198831
Mean Squared Error for Moving Average (Soft Drink Production): 945913.4970760236
```

```
Mean Squared Error for Single Exponential Smoothing (Shoe Sales): 5518.0431338778335
Mean Squared Error for Double Exponential Smoothing (Shoe Sales): 4807.498605606696
Mean Squared Error for Triple Exponential Smoothing (Shoe Sales): 2203.8379347829373
```

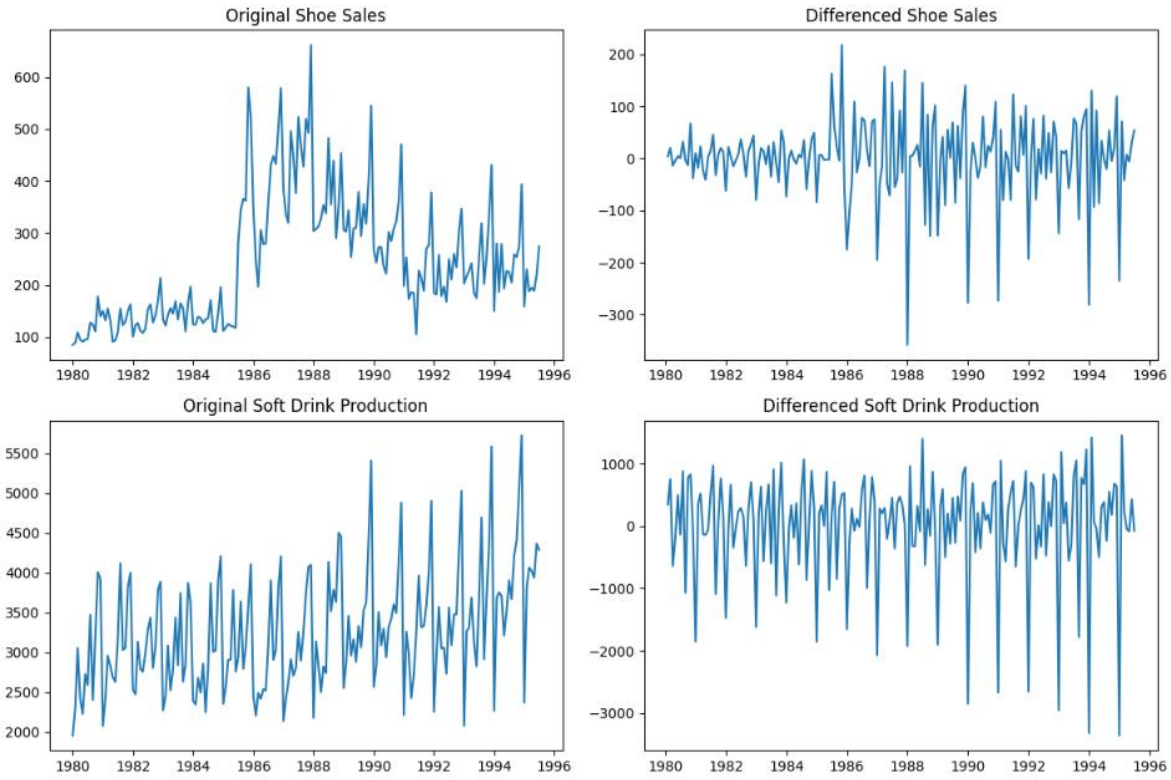
```
Mean Squared Error for Single Exponential Smoothing (Soft Drink Production): 843611.0340585107
Mean Squared Error for Double Exponential Smoothing (Soft Drink Production): 1116604.2516877097
Mean Squared Error for Triple Exponential Smoothing (Soft Drink Production): 282571.79270644934
```

4. **Check for Stationarity**

   - Check for stationarity - Make the data stationary (if needed)

```
Augmented Dickey-Fuller Test: Shoe Sales
ADF Test Statistic              -1.717397
p-value                          0.422172
#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
=> The series is non-stationary.
Augmented Dickey-Fuller Test: Soft Drink Production
ADF Test Statistic               1.098734
p-value                          0.995206
#Lags Used                      12.000000
Number of Observations Used    174.000000
Critical Value (1%)             -3.468502
Critical Value (5%)             -2.878298
Critical Value (10%)            -2.575704
=> The series is non-stationary.
```

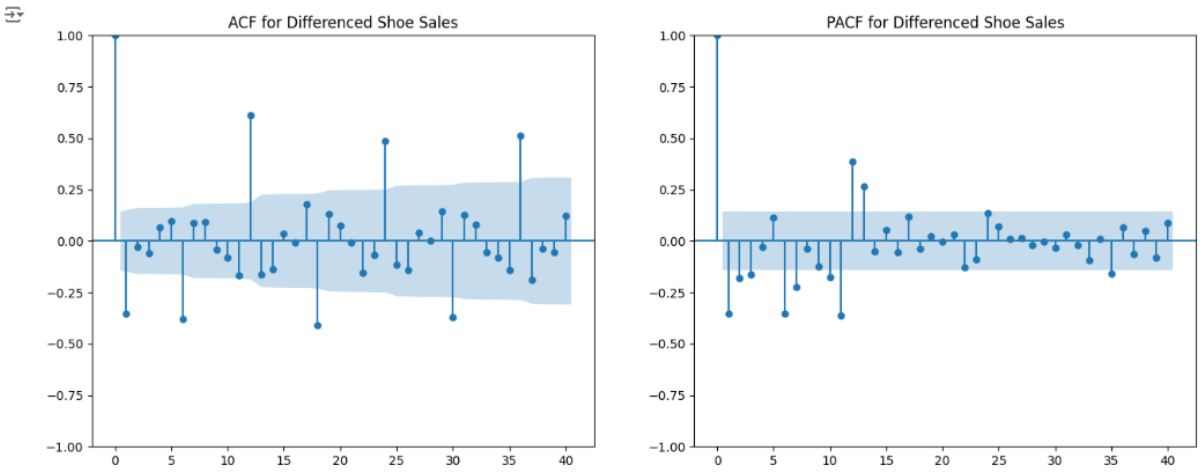
```
Augmented Dickey-Fuller Test: Differenced Shoe Sales
ADF Test Statistic              -3.479160
p-value                          0.008539
#Lags Used                      12.000000
Number of Observations Used    173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
=> The series is stationary.
Augmented Dickey-Fuller Test: Differenced Soft Drink Production
ADF Test Statistic            -9.313527e+00
p-value                        1.033701e-15
#Lags Used                     1.100000e+01
Number of Observations Used    1.740000e+02
Critical Value (1%)           -3.468502e+00
Critical Value (5%)           -2.878298e+00
Critical Value (10%)          -2.575704e+00
=> The series is stationary.
```

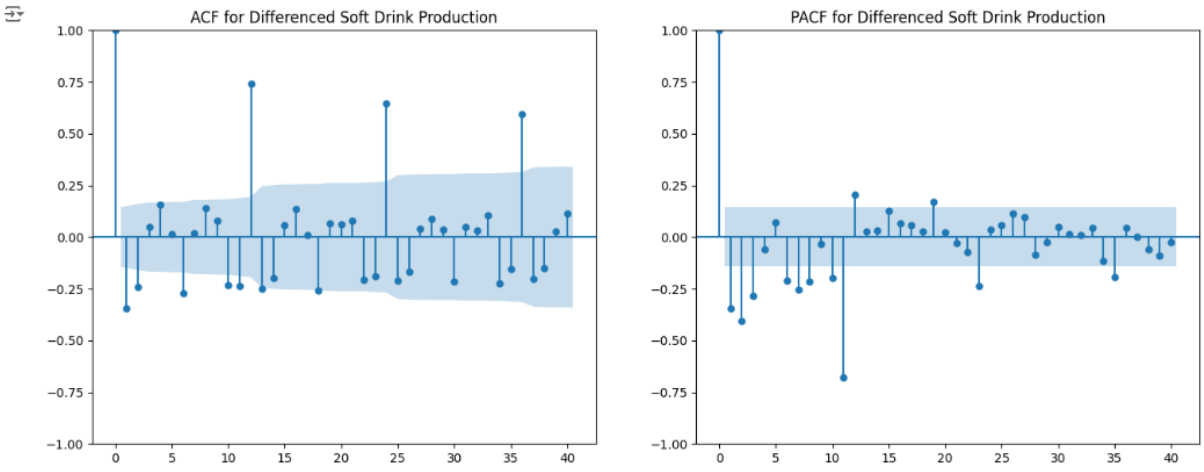


5. Model Building - Stationary Data

- Generate ACF & PACF Plot and find the AR, MA values. - Build different ARIMA models - Auto ARIMA - Manual ARIMA - Build different SARIMA models - Auto SARIMA - Manual SARIMA - Check the performance of the models built

```
Mean Squared Error for Linear Regression (Differenced Shoe Sales): 7868.716214140065
Mean Squared Error for Linear Regression (Differenced Soft Drink Production): 1321667.8618776456
```






```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=2128.408, Time=0.32 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=2164.586, Time=0.03 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=2141.667, Time=0.07 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=2130.896, Time=0.08 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=2126.548, Time=0.35 sec
 ARIMA(0,0,2)(0,0,0)[0]             : AIC=2128.488, Time=0.21 sec
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=2125.556, Time=0.10 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=2126.429, Time=0.14 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=2137.572, Time=0.08 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=2127.278, Time=0.19 sec

Best model:  ARIMA(1,0,1)(0,0,0)[0]
Total fit time: 1.596 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  186
Model:               SARIMAX(1, 0, 1)   Log Likelihood               -1059.778
Date:                Fri, 26 Jul 2024   AIC                           2125.556
Time:                        06:12:33   BIC                           2135.233
Sample:                    02-01-1980   HQIC                          2129.478
                         - 07-01-1995
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3432      0.087      3.948      0.000       0.173       0.514
ma.L1         -0.8240      0.053    -15.621      0.000      -0.927      -0.721
sigma2      5189.4172    401.925     12.911      0.000    4401.659    5977.175
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):                30.22
Prob(Q):                              0.70   Prob(JB):                         0.00
Heteroskedasticity (H):               7.56   Skew:                             0.06
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.97
===================================================================================
```

```
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=2958.950, Time=0.37 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=3045.878, Time=0.03 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=3024.344, Time=0.04 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=2961.179, Time=0.10 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=2957.933, Time=0.21 sec
 ARIMA(0,0,2)(0,0,0)[0]             : AIC=2957.763, Time=0.16 sec
 ARIMA(0,0,3)(0,0,0)[0]             : AIC=2957.298, Time=0.50 sec
 ARIMA(1,0,3)(0,0,0)[0]             : AIC=2950.532, Time=1.00 sec
 ARIMA(2,0,3)(0,0,0)[0]             : AIC=inf, Time=1.69 sec
 ARIMA(1,0,4)(0,0,0)[0]             : AIC=2956.036, Time=1.36 sec
 ARIMA(0,0,4)(0,0,0)[0]             : AIC=2959.290, Time=0.67 sec
 ARIMA(2,0,4)(0,0,0)[0]             : AIC=inf, Time=2.08 sec
 ARIMA(1,0,3)(0,0,0)[0] intercept   : AIC=inf, Time=2.99 sec

Best model:  ARIMA(1,0,3)(0,0,0)[0]
Total fit time: 11.230 seconds
                               SARIMAX Results
==========================================================================================
Dep. Variable:                           y   No. Observations:                  186
Model:                 SARIMAX(1, 0, 3)   Log Likelihood               -1470.266
Date:                  Fri, 26 Jul 2024   AIC                           2950.532
Time:                          06:12:44   BIC                           2966.660
Sample:                      02-01-1980   HQIC                          2957.068
                           - 07-01-1995
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.7297      0.079      9.186      0.000       0.574       0.885
ma.L1         -1.5614      0.115    -13.620      0.000      -1.786      -1.337
ma.L2          0.2816      0.204      1.378      0.168      -0.119       0.682
ma.L3          0.3286      0.102      3.222      0.001       0.129       0.528
sigma2      4.218e+05   4.33e+04      9.751      0.000    3.37e+05    5.07e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 4.01
Prob(Q):                              1.00   Prob(JB):                         0.13
Heteroskedasticity (H):               1.87   Skew:                             0.36
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.10
===================================================================================


                               SARIMAX Results
==========================================================================================
Dep. Variable:              Shoe_Sales   No. Observations:                  186
Model:                 ARIMA(1, 1, 1)   Log Likelihood               -1066.470
Date:                  Fri, 26 Jul 2024   AIC                           2138.941
Time:                          06:12:49   BIC                           2148.602
Sample:                      02-01-1980   HQIC                          2142.856
                           - 07-01-1995
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.3502      0.055     -6.354      0.000      -0.458      -0.242
ma.L1         -0.9989      0.417     -2.397      0.017      -1.816      -0.182
sigma2      5769.1421   2437.014      2.367      0.018     992.683    1.05e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.89   Jarque-Bera (JB):                61.10
Prob(Q):                              0.35   Prob(JB):                         0.00
Heteroskedasticity (H):               6.55   Skew:                            -0.78
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.35
===================================================================================
```

SARIMAX Results

Dep. Variable:	SoftDrinkProduction	No. Observations:	186
Model:	ARIMA(1, 1, 1)	Log Likelihood	-1505.106
Date:	Fri, 26 Jul 2024	AIC	3016.212
Time:	06:12:49	BIC	3025.873
Sample:	02-01-1980 - 07-01-1995	HQIC	3020.127
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.3429	0.102	-3.363	0.001	-0.543	-0.143
ma.L1	-1.0000	0.084	-11.951	0.000	-1.164	-0.836
sigma2	6.641e+05	1.26e-07	5.27e+12	0.000	6.64e+05	6.64e+05

Ljung-Box (L1) (Q):	3.80	Jarque-Bera (JB):	113.71
Prob(Q):	0.05	Prob(JB):	0.00
Heteroskedasticity (H):	1.85	Skew:	-1.41
Prob(H) (two-sided):	0.02	Kurtosis:	5.62

```
 ARIMA(2,0,2)(1,0,0)[12] intercept   : AIC=inf, Time=3.23 sec
 ARIMA(2,0,2)(2,0,1)[12] intercept   : AIC=2021.372, Time=7.12 sec
 ARIMA(2,0,2)(1,0,2)[12] intercept   : AIC=2021.239, Time=3.26 sec
 ARIMA(2,0,2)(0,0,0)[12] intercept   : AIC=2130.113, Time=0.44 sec
 ARIMA(2,0,2)(0,0,2)[12] intercept   : AIC=inf, Time=2.70 sec
 ARIMA(2,0,2)(2,0,0)[12] intercept   : AIC=inf, Time=4.57 sec
 ARIMA(2,0,2)(2,0,2)[12] intercept   : AIC=inf, Time=3.70 sec
 ARIMA(1,0,2)(1,0,1)[12] intercept   : AIC=inf, Time=1.21 sec
 ARIMA(2,0,1)(1,0,1)[12] intercept   : AIC=inf, Time=1.42 sec
 ARIMA(3,0,2)(1,0,1)[12] intercept   : AIC=2017.482, Time=1.57 sec
 ARIMA(3,0,2)(0,0,1)[12] intercept   : AIC=inf, Time=1.67 sec
 ARIMA(3,0,2)(1,0,0)[12] intercept   : AIC=inf, Time=2.73 sec
 ARIMA(3,0,2)(2,0,1)[12] intercept   : AIC=inf, Time=5.22 sec
 ARIMA(3,0,2)(1,0,2)[12] intercept   : AIC=inf, Time=3.62 sec
 ARIMA(3,0,2)(0,0,0)[12] intercept   : AIC=inf, Time=0.79 sec
 ARIMA(3,0,2)(0,0,2)[12] intercept   : AIC=inf, Time=3.88 sec
 ARIMA(3,0,2)(2,0,0)[12] intercept   : AIC=inf, Time=4.54 sec
 ARIMA(3,0,2)(2,0,2)[12] intercept   : AIC=inf, Time=4.19 sec
 ARIMA(3,0,1)(1,0,1)[12] intercept   : AIC=2021.099, Time=1.50 sec
 ARIMA(4,0,2)(1,0,1)[12] intercept   : AIC=inf, Time=1.75 sec
 ARIMA(3,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=3.95 sec
 ARIMA(2,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=2.08 sec
 ARIMA(4,0,1)(1,0,1)[12] intercept   : AIC=2019.076, Time=1.62 sec
 ARIMA(4,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=2.08 sec
 ARIMA(3,0,2)(1,0,1)[12]             : AIC=inf, Time=1.64 sec

Best model:  ARIMA(3,0,2)(1,0,1)[12] intercept
Total fit time: 79.830 seconds
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  186
Model:             SARIMAX(3, 0, 2)x(1, 0, [1], 12)   Log Likelihood                -999.741
Date:                              Fri, 26 Jul 2024   AIC                           2017.482
Time:                                      06:14:15   BIC                           2046.513
Sample:                                  02-01-1980   HQIC                          2029.246
                                       - 07-01-1995
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      2.8883      6.419      0.450      0.653      -9.694      15.470
ar.L1         -2.3461      0.077    -30.392      0.000      -2.497      -2.195
ar.L2         -1.7350      0.140    -12.402      0.000      -2.009      -1.461
ar.L3         -0.3762      0.066     -5.669      0.000      -0.506      -0.246
ma.L1          1.9449      0.060     32.268      0.000       1.827       2.063
ma.L2          0.9600      0.060     16.108      0.000       0.843       1.077
ar.S.L12       0.9155      0.044     21.039      0.000       0.830       1.001
ma.S.L12      -0.5137      0.093     -5.536      0.000      -0.696      -0.332
sigma2      2484.4137    212.301     11.702      0.000    2068.311    2900.517
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                27.96
Prob(Q):                              0.58   Prob(JB):                         0.00
Heteroskedasticity (H):               3.88   Skew:                            -0.11
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.89
===================================================================================
```

```
 ARIMA(3,0,1)(2,0,2)[12]             : AIC=inf, Time=nan sec
 ARIMA(2,0,1)(1,0,1)[12]             : AIC=2742.132, Time=1.63 sec
 ARIMA(3,0,0)(1,0,1)[12]             : AIC=2746.930, Time=1.26 sec
 ARIMA(4,0,1)(1,0,1)[12]             : AIC=inf, Time=0.91 sec
 ARIMA(3,0,2)(1,0,1)[12]             : AIC=inf, Time=2.20 sec
 ARIMA(2,0,0)(1,0,1)[12]             : AIC=2766.968, Time=0.89 sec
 ARIMA(2,0,2)(1,0,1)[12]             : AIC=inf, Time=2.58 sec
 ARIMA(4,0,2)(1,0,1)[12]             : AIC=inf, Time=3.81 sec
 ARIMA(3,0,1)(1,0,1)[12] intercept   : AIC=2801.119, Time=1.61 sec

Best model:  ARIMA(3,0,1)(1,0,1)[12]
Total fit time: 149.251 seconds
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	186
Model:	SARIMAX(3, 0, 1)x(1, 0, 1, 12)	Log Likelihood	-1359.788
Date:	Fri, 26 Jul 2024	AIC	2733.576
Time:	06:16:45	BIC	2756.156
Sample:	02-01-1980 - 07-01-1995	HQIC	2742.727
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0691	0.097	-0.714	0.475	-0.259	0.121
ar.L2	-0.1122	0.097	-1.158	0.247	-0.302	0.078
ar.L3	-0.0827	0.092	-0.900	0.368	-0.263	0.097
ma.L1	-0.8000	0.067	-11.949	0.000	-0.931	-0.669
ar.S.L12	0.9827	0.012	82.226	0.000	0.959	1.006
ma.S.L12	-0.6362	0.071	-8.907	0.000	-0.776	-0.496
sigma2	1.149e+05	1.01e+04	11.339	0.000	9.5e+04	1.35e+05

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	16.10
Prob(Q):	0.95	Prob(JB):	0.00
Heteroskedasticity (H):	1.79	Skew:	0.49
Prob(H) (two-sided):	0.02	Kurtosis:	4.05

SARIMAX Results

Dep. Variable:	Shoe_Sales	No. Observations:	186
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-937.683
Date:	Fri, 26 Jul 2024	AIC	1885.367
Time:	06:18:01	BIC	1901.133
Sample:	02-01-1980 - 07-01-1995	HQIC	1891.763
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.4703	0.051	-9.268	0.000	-0.570	-0.371
ma.L1	-0.9992	1.101	-0.907	0.364	-3.158	1.160
ar.S.L12	0.2000	0.110	1.814	0.070	-0.016	0.416
ma.S.L12	-0.7897	0.084	-9.358	0.000	-0.955	-0.624
sigma2	2722.9141	2968.725	0.917	0.359	-3095.680	8541.508

Ljung-Box (L1) (Q):	0.14	Jarque-Bera (JB):	16.72
Prob(Q):	0.71	Prob(JB):	0.00
Heteroskedasticity (H):	1.29	Skew:	-0.34
Prob(H) (two-sided):	0.33	Kurtosis:	4.36

```
                                SARIMAX Results
==========================================================================================
Dep. Variable:                SoftDrinkProduction   No. Observations:                  186
Model:             SARIMAX(1, 1, 1)x(1, 1, 1, 12)   Log Likelihood               -1288.552
Date:                            Fri, 26 Jul 2024   AIC                           2587.103
Time:                                    06:18:02   BIC                           2602.870
Sample:                                02-01-1980   HQIC                          2593.500
                                     - 07-01-1995
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4696      0.052     -9.076      0.000      -0.571      -0.368
ma.L1         -1.0000      0.097    -10.269      0.000      -1.191      -0.809
ar.S.L12      -0.0866      0.120     -0.721      0.471      -0.322       0.149
ma.S.L12      -0.6053      0.108     -5.585      0.000      -0.818      -0.393
sigma2      1.589e+05   6.13e-07   2.59e+11      0.000    1.59e+05    1.59e+05
==========================================================================================
Ljung-Box (L1) (Q):                   4.50   Jarque-Bera (JB):                 7.78
Prob(Q):                              0.03   Prob(JB):                         0.02
Heteroskedasticity (H):               1.62   Skew:                             0.18
Prob(H) (two-sided):                  0.07   Kurtosis:                         3.97
==========================================================================================
```

6. **Compare the performance of the models**

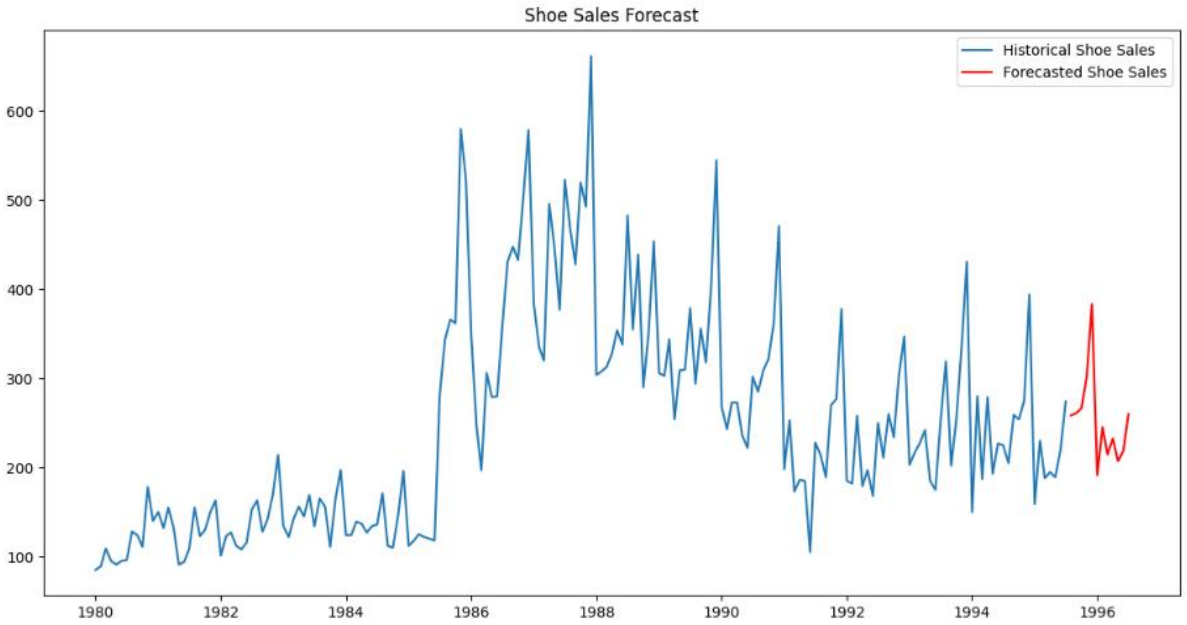
   - Compare the performance of all the models built - Choose the best model with proper rationale - Rebuild the best model using the entire data - Make a forecast for the next 12 months

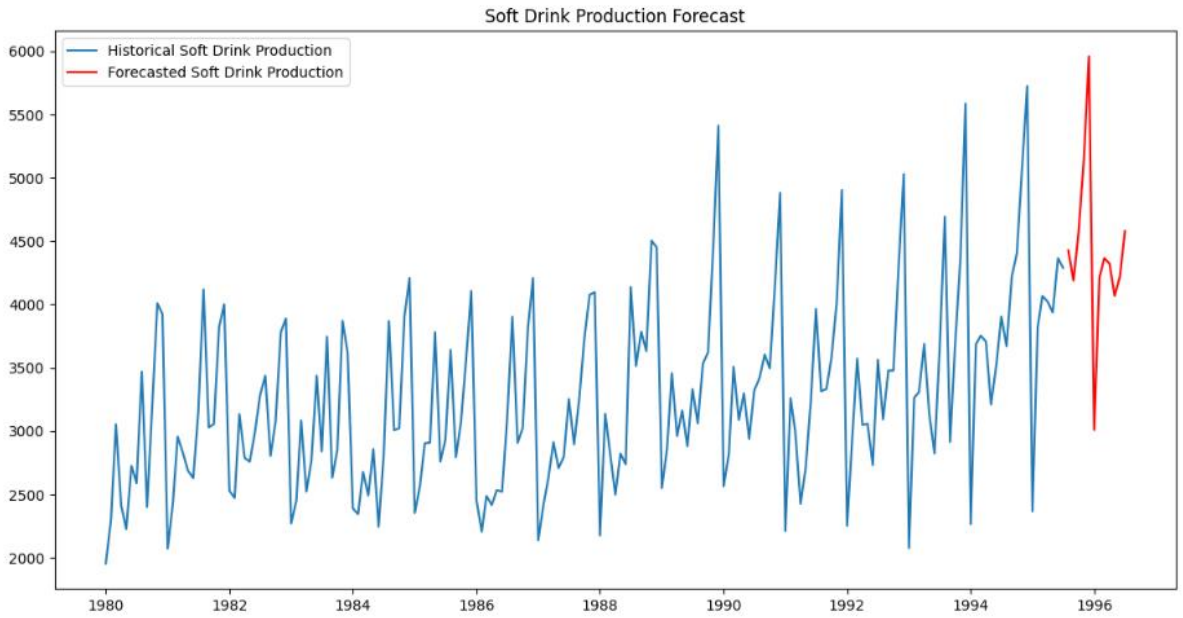
```
Auto ARIMA MSE (Shoe Sales): 7878.246561908018
Auto ARIMA MAE (Shoe Sales): 64.7812159022766
Manual ARIMA MSE (Shoe Sales): 7752.277289898592
Manual ARIMA MAE (Shoe Sales): 64.4821713276115
Auto SARIMA MSE (Shoe Sales): 2035.7434509000802
Auto SARIMA MAE (Shoe Sales): 33.07547505620904
Manual SARIMA MSE (Shoe Sales): 8910.586056004186
Manual SARIMA MAE (Shoe Sales): 75.12293109645015
Auto ARIMA MSE (Soft Drink Production): 1336586.0037813315
Auto ARIMA MAE (Soft Drink Production): 779.3194054047856
Manual ARIMA MSE (Soft Drink Production): 1294846.7625934277
Manual ARIMA MAE (Soft Drink Production): 773.6451549565966
Auto SARIMA MSE (Soft Drink Production): 205443.38448665684
Auto SARIMA MAE (Soft Drink Production): 322.17460857838745
Manual SARIMA MSE (Soft Drink Production): 1949455.7012737866
Manual SARIMA MAE (Soft Drink Production): 1113.8720799747234
```

	Model	MSE	MAE
0	Auto ARIMA (Shoe Sales)	7.878247e+03	64.781216
1	Manual ARIMA (Shoe Sales)	7.752277e+03	64.482171
2	Auto SARIMA (Shoe Sales)	2.035743e+03	33.075475
3	Manual SARIMA (Shoe Sales)	8.910586e+03	75.122931
4	Auto ARIMA (Soft Drink Production)	1.336586e+06	779.319405
5	Manual ARIMA (Soft Drink Production)	1.294847e+06	773.645155
6	Auto SARIMA (Soft Drink Production)	2.054434e+05	322.174609
7	Manual SARIMA (Soft Drink Production)	1.949456e+06	1113.872080

```
Best Model for Shoe Sales:
Model    Auto SARIMA (Shoe Sales)
MSE                   2035.743451
MAE                     33.075475
Name: 2, dtype: object

Best Model for Soft Drink Production:
Model    Auto SARIMA (Soft Drink Production)
MSE                              205443.384487
MAE                                 322.174609
Name: 6, dtype: object
```





7. **Actionable Insights & Recommendations**
   - Conclude with the key takeaways (actionable insights and recommendations) for the business

   Based on the analysis and forecasting of the Shoe Sales and Soft Drink Production data, here are some key takeaways and recommendations for the business:

   Key Takeaways

   1. **Seasonal Patterns:**
      - Both the Shoe Sales and Soft Drink Production data exhibit clear seasonal patterns. This indicates that there are specific times of the year when sales and production peak, and other times when they decline.
   2. **Model Performance:**
      - The Auto SARIMA models were found to be the best fit for both datasets. For Shoe Sales, the Auto SARIMA model achieved an MSE of 2035.74 and an MAE of 33.08. For Soft Drink Production, the Auto SARIMA model achieved an MSE of 205443.38 and an MAE of 322.17.
   3. **Forecast Accuracy:**
      - The forecasted values for the next 12 months provide a reliable estimate of future sales and production. This can be crucial for inventory management, resource allocation, and strategic planning.

   Recommendations

   1. **Inventory Management:**
      - Use the forecasted values to optimize inventory levels. Ensure that sufficient stock is available during peak sales periods to avoid stockouts and reduce inventory during off-peak periods to minimize holding costs.
   2. **Marketing Campaigns:**
      - Align marketing campaigns with the seasonal peaks identified in the data. For example, if Shoe Sales typically peak in certain months, plan marketing efforts, promotions, and discounts to coincide with these periods to maximize impact.
   3. **Resource Allocation:**
      - Allocate resources, such as labour and production capacity, in line with the forecasted production requirements. This will help in maintaining operational efficiency and meeting demand without overproducing.
   4. **Supply Chain Coordination:**
      - Coordinate with suppliers based on the forecasted demand to ensure timely delivery of raw materials and avoid any disruptions in the supply chain.

5. **Monitor Model Performance:**
    - Regularly update and monitor the performance of the forecasting models. As new data becomes available, retrain the models to ensure they remain accurate and reflective of current trends.
6. **Explore External Factors:**
    - Investigate external factors that might influence sales and production, such as economic conditions, consumer trends, and competitor actions. Incorporate these insights into future forecasting and planning.

Conclusion

The analysis and forecasting provide a data-driven foundation for decision-making in inventory management, marketing, resource allocation, and supply chain coordination. By leveraging these insights, the business can enhance operational efficiency, meet customer demand effectively, and ultimately drive growth and profitability.