FORECASTING NIFTY PHARMA STOCKS USING TIME SERIES ANALYSIS

SUMMER PROJECT

TIME SERIES

SHIV YADAV 22N0064 | MSC ASI

UNDER THE GUIDANCE OF PROF. SANJEEV V. SABNIS (IIT- BOMBAY)

Time Series Analysis and Forecasting of Nifty Pharma Stock

1. **Objective:**
   The objective of the Time Series Analysis and Forecasting of Nifty Pharma Stock is to gain insights into the historical price movements and patterns of the Nifty Pharma Index, which represents the pharmaceutical sector's performance on the National Stock Exchange of India (NSE). This analysis aims to provide a comprehensive understanding of the underlying dynamics of the Nifty Pharma Stock and develop accurate forecasting models to predict its future price trends.
2. **Data Description:**
   The data for the Time Series Analysis and Forecasting of Nifty Pharma Stock is obtained from the official website of the National Stock Exchange of India (NSE).
   https://www.niftyindices.com/reports/historical-data
   The NSE is a premier stock exchange in India, and it provides reliable and accurate historical price data for various financial instruments, including the Nifty Pharma Index.

   Stock Symbol: Symbol used to identify the Nifty Pharma Index.

   Date: The timestamp of the data point.

   Open Price: The opening price of the Nifty Pharma Index on that particular date.

   High Price: The highest price recorded during the trading session.

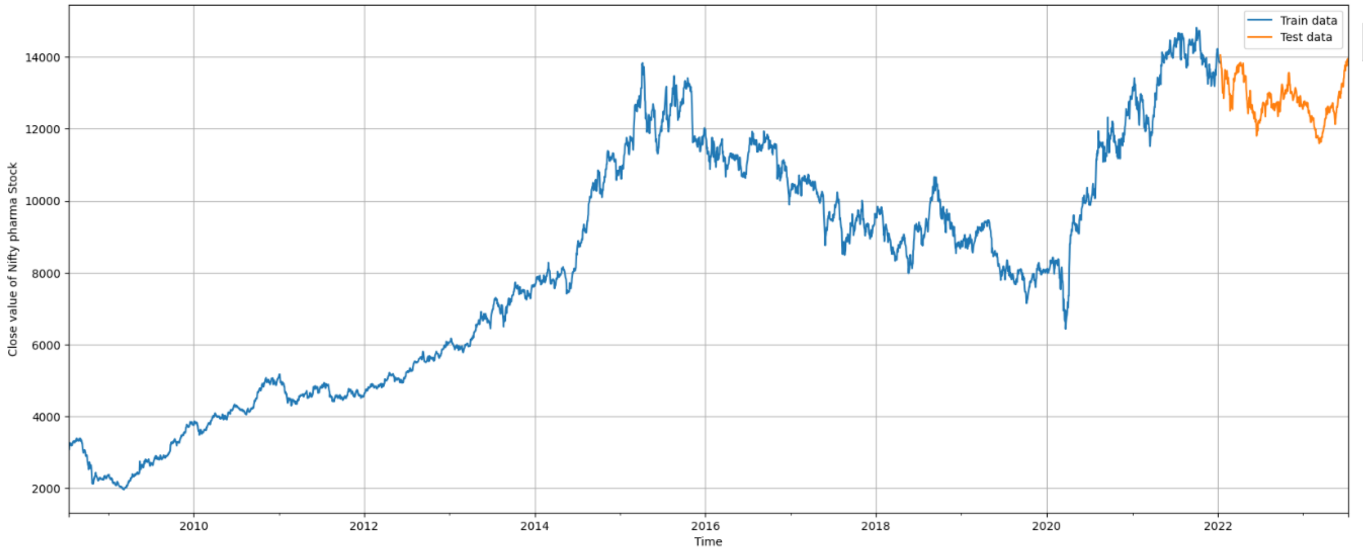
   Low Price: The lowest price recorded during the trading session.

   Close Price: The closing price of the Nifty Pharma Index on that particular date.
3. **Glimpse of Nifty Pharma Data set:**

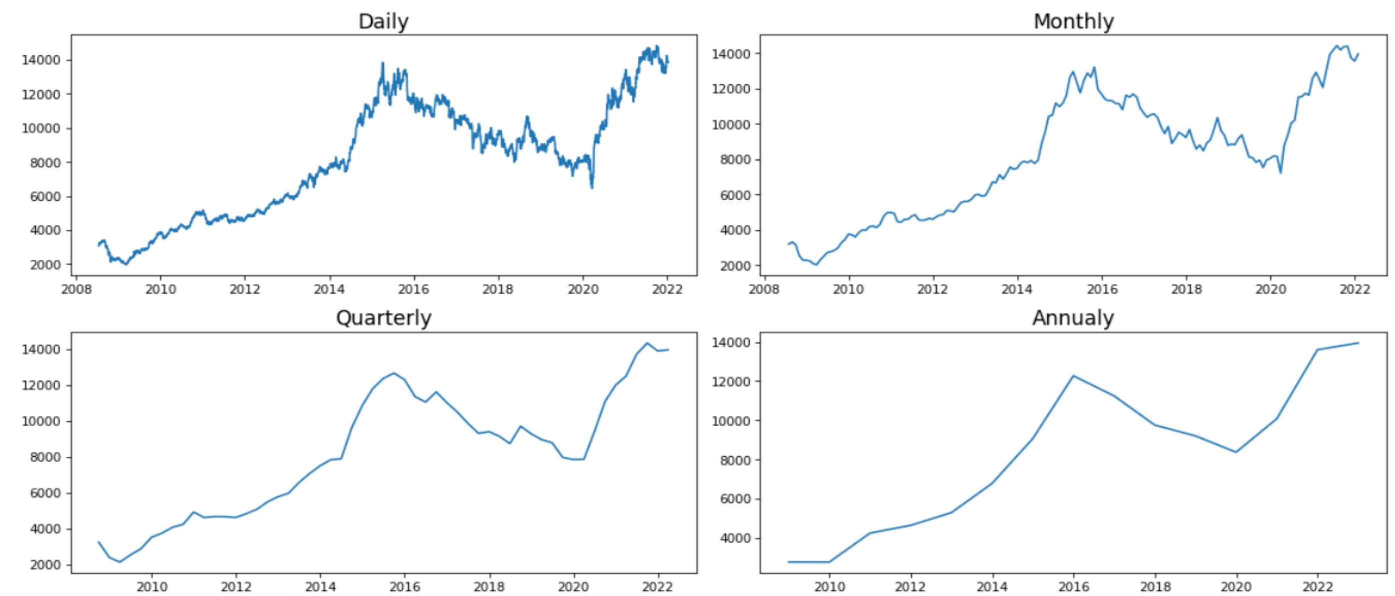
	Date	Open	High	Low	Close
0	14 Jul 2023	13931.75	13961.20	13852.90	13945.65
1	13 Jul 2023	13979.40	13980.05	13841.45	13876.55
2	12 Jul 2023	13918.20	13970.30	13873.75	13936.70
3	11 Jul 2023	13795.55	13915.50	13780.45	13895.40
4	10 Jul 2023	13811.15	13821.00	13705.35	13751.80
5	07 Jul 2023	13800.40	13917.55	13730.35	13769.00
6	06 Jul 2023	13769.40	13904.65	13759.75	13871.60
7	05 Jul 2023	13726.25	13788.65	13682.85	13772.00
8	04 Jul 2023	13683.15	13713.95	13578.35	13675.45
9	03 Jul 2023	13796.55	13811.70	13597.40	13614.00

4. **Plotting the Time Series Data:**
   Firstly, to get some insights our time series data, we will take the help of graphical method, i.e., we will plot our time series data over whole time period of study to know that how our data varies over the time.

• **Plotting the Nifty Pharma Time series data (July/2008 - July/2023)**



➢ **Comment**: From the above plot of Nifty Pharma over the whole time period of study we can see down trend in the stock prices from (2016 till 2020) and after 2020 onwards the stock price of Nifty pharma started increasing (upward trend)



➢ Decomposition of the time series Data

Decomposition:

Decomposing a time series data refers to the process of breaking down the time series into its constituent components, namely trend, seasonality, and Residual (also called noise). Time series decomposition is a common technique used in analysing and understanding time series data.

- **Naive Decomposition (Classical Decomposition) "The Additive Model"**
  The classical decomposition of a time series refers to a widely used method for decomposing a time series into its trend, seasonality, and residual components. It is also known as the additive

decomposition method. The classical decomposition assumes that the time series can be expressed as the sum of following three components:

1. $\boldsymbol{Trend\ T}$
2. $\boldsymbol{Seasonality\ S}$
3. $\boldsymbol{Residuals\ R}$

Thus, the Additive Model is given by:

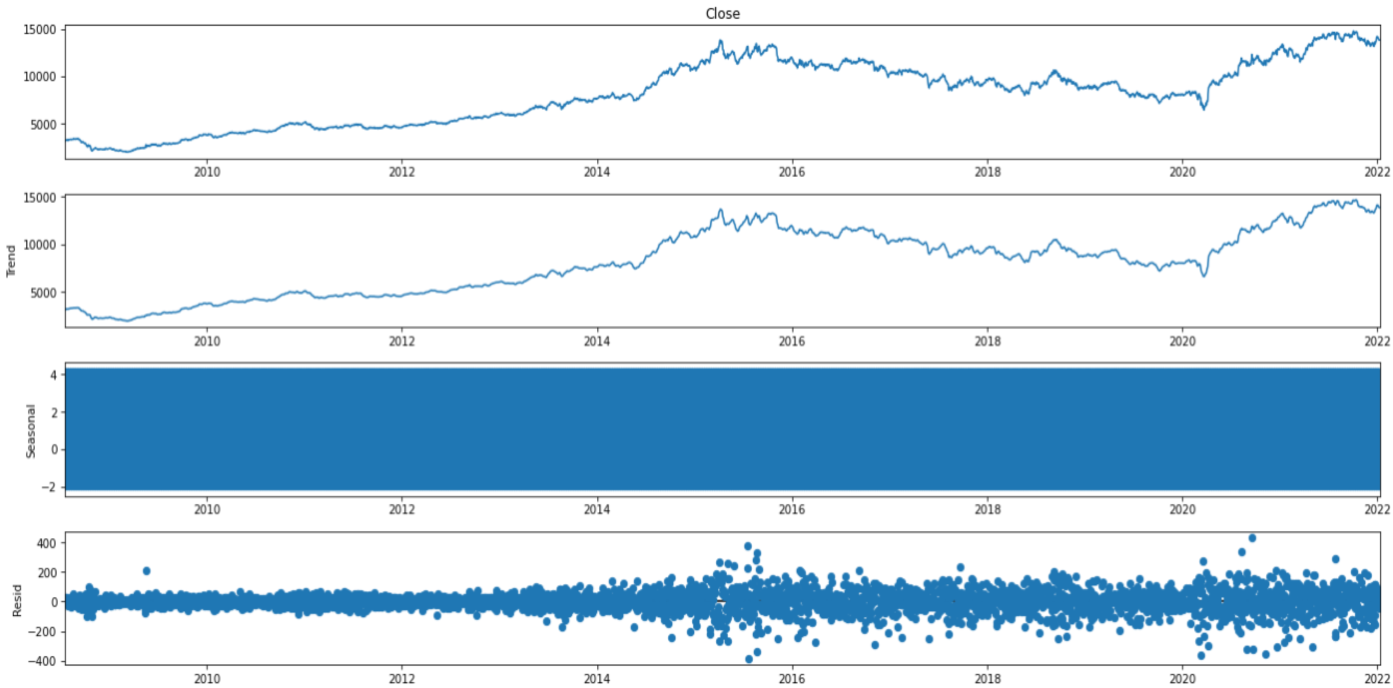
$$\boldsymbol{Xt = T + S + R}$$

- **Multiplicative Decomposition of the Time Series Data**:
  This method of decomposition assumes that the time series can be expressed as the product of following three components:
  1. $\boldsymbol{Trend\ T}$
  2. $\boldsymbol{Seasonality\ S}$
  3. $\boldsymbol{Residuals\ R}$

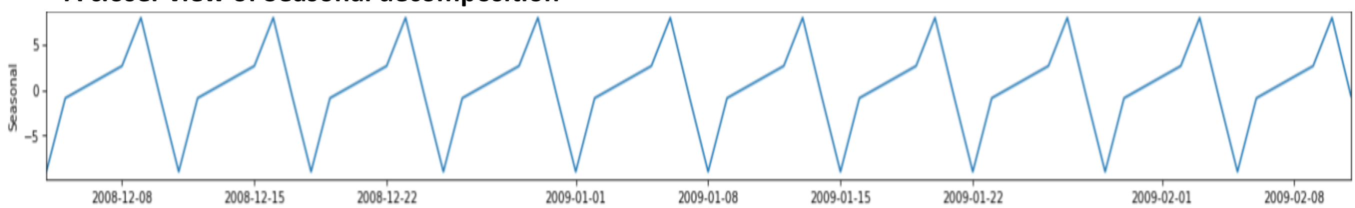
  And the Multiplicative Model is given by:

$$\boldsymbol{Xt = T * S * R}$$

(a) Additive Decomposition for Nifty Pharma Stock:

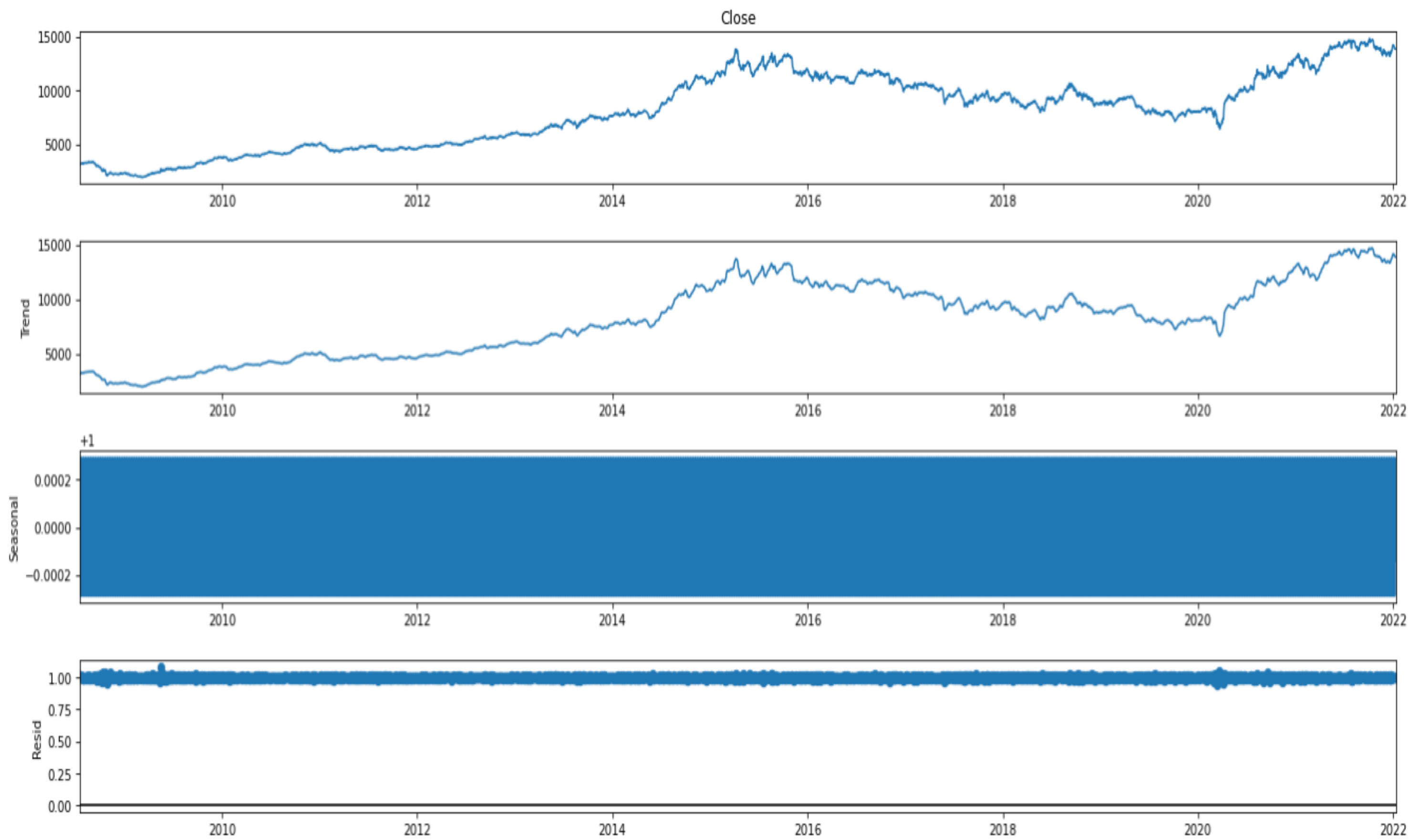


A closer view of Seasonal decomposition



➢ **Comment**: From the above plots it can be concluded that there is an upward trend and a seasonal component with very small period (maybe 7 days) present in the Nifty pharma stock data also from the residual plot it can be interpreted that before July 2016 residuals are evenly distributed around the Zero line but after that residuals have more variation around the Zero line

b.) Multiplicative Decomposition for UPI Volume Data:



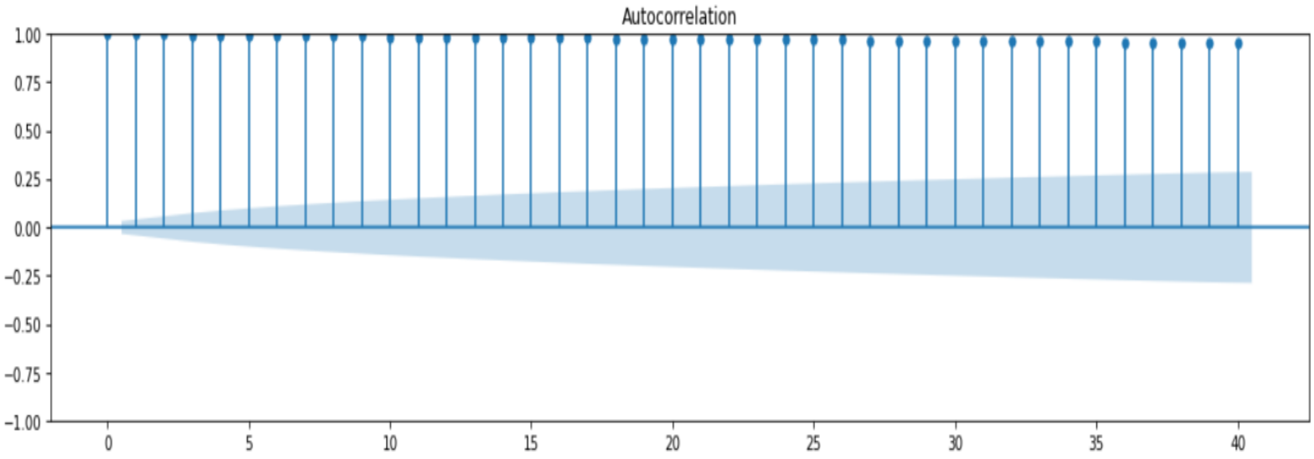
- **Comment:** From the above multiplicative decomposition, the same can be concluded, which we have concluded from the Additive decomposition except the residuals terms, in the additive decomposition the residuals are centred around the Zero line(which can so, because in Additive decomposition, $\boldsymbol{Residual} = \boldsymbol{Actual\ value} - \boldsymbol{Predicted\ value}$) while in the multiplicative decomposition the residuals are centred around the $y = 1$line again which can so, because in multiplicative decomposition $\boldsymbol{Residual} = \boldsymbol{Actual\ Value\ Predicted\ Value}$
  In Multiplicative model the residuals have lesser fluctuation around the $\boldsymbol{y} = \mathbf{1}$ line, this indicates that Multiplicative model is quite preferable over the additive model for the Nifty Pharma Stock data

Modelling to the Nifty Pharma Stock Data

The plot reveals that the data is non-stationary, as the level of the process is not constant. We difference the data in order to convert the time series to stationary.

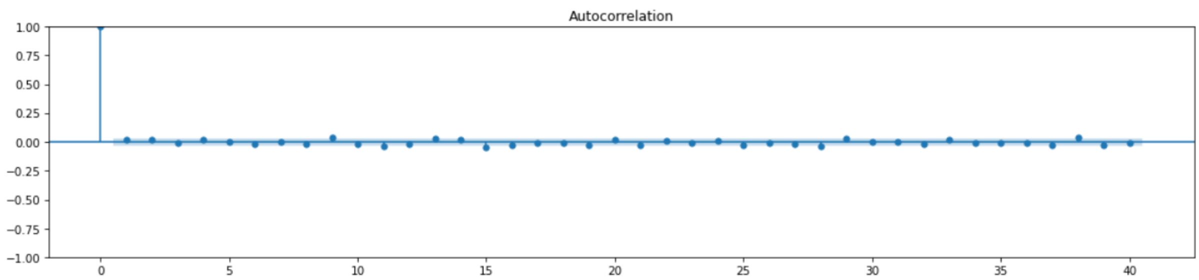
Differencing the data

We first plot the ACF (auto- correlation function) plot of the original time series data.



The estimated autocorrelation function does not decay quickly, suggesting the need for differencing. We conduct **first order and second order differencing**.

ACF plot of first order differenced Data



The ACF plot die out quickly, which suggests that we can perform either first or second order differencing.

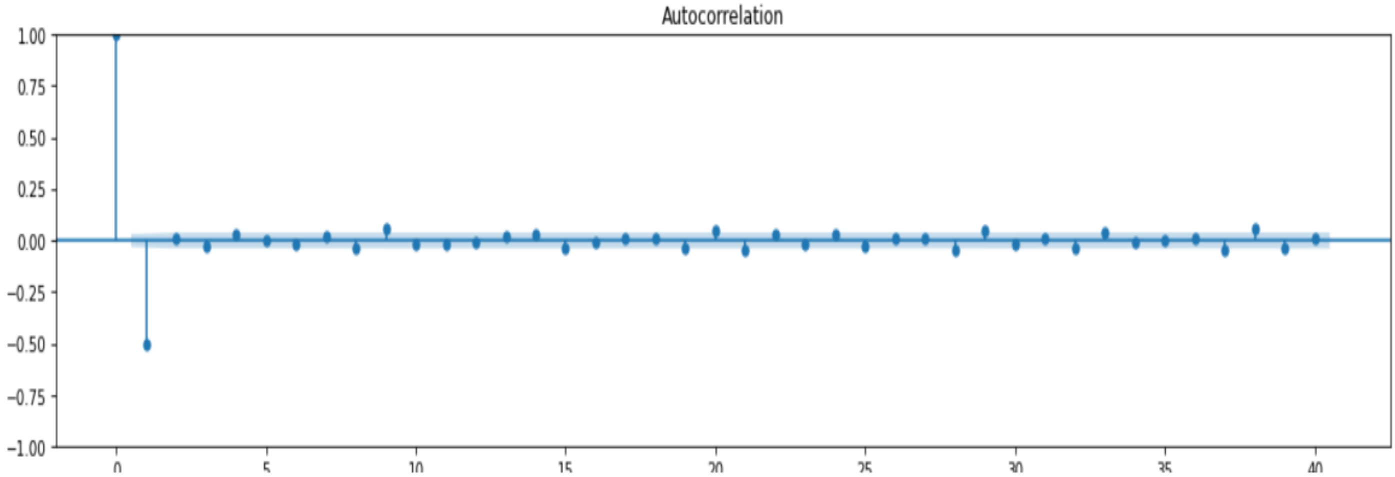
We use the augmented Dickey- Fuller test (ADF Test) to determine whether stationarity has been achieved.

```
ADF Statistic: -16.387397
p-value: 0.000000
Critical Values:
     1%: -3.432
     5%: -2.862
     10%: -2.567
```

As the p- value for the ADF test, after differencing the data once < alpha, we conclude that stationarity has been achieved by 1st order differencing.

Let's see what happen if take another differencing.

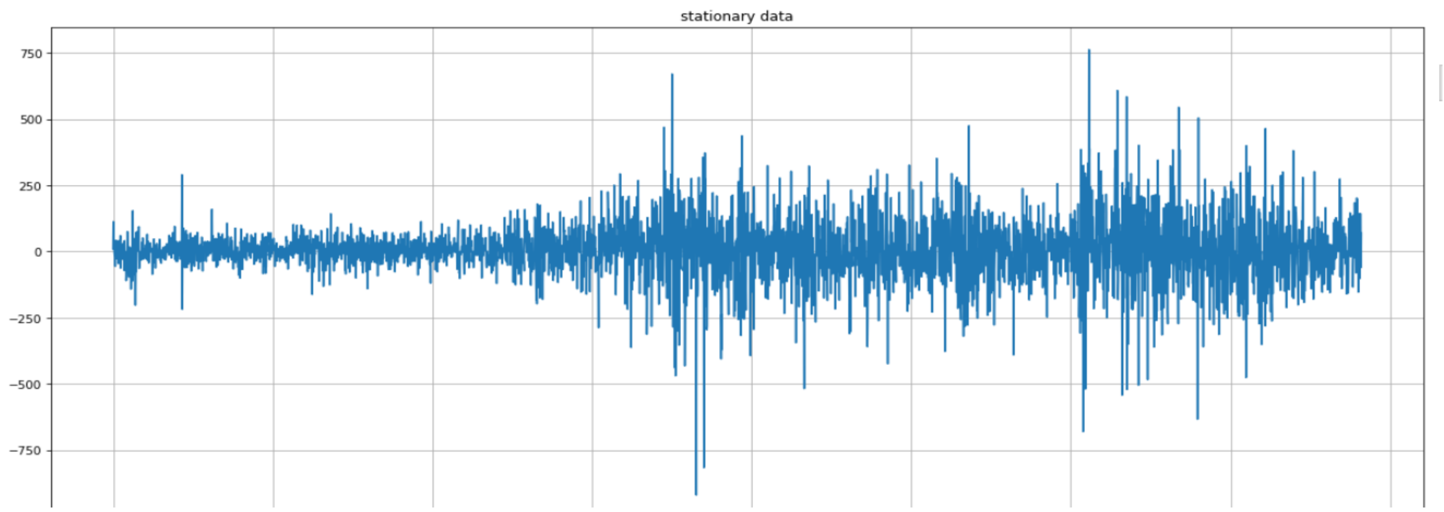
ACF plot of second order differenced Data:



We use the augmented Dickey- Fuller test (ADF Test) to determine whether stationarity has been achieved.

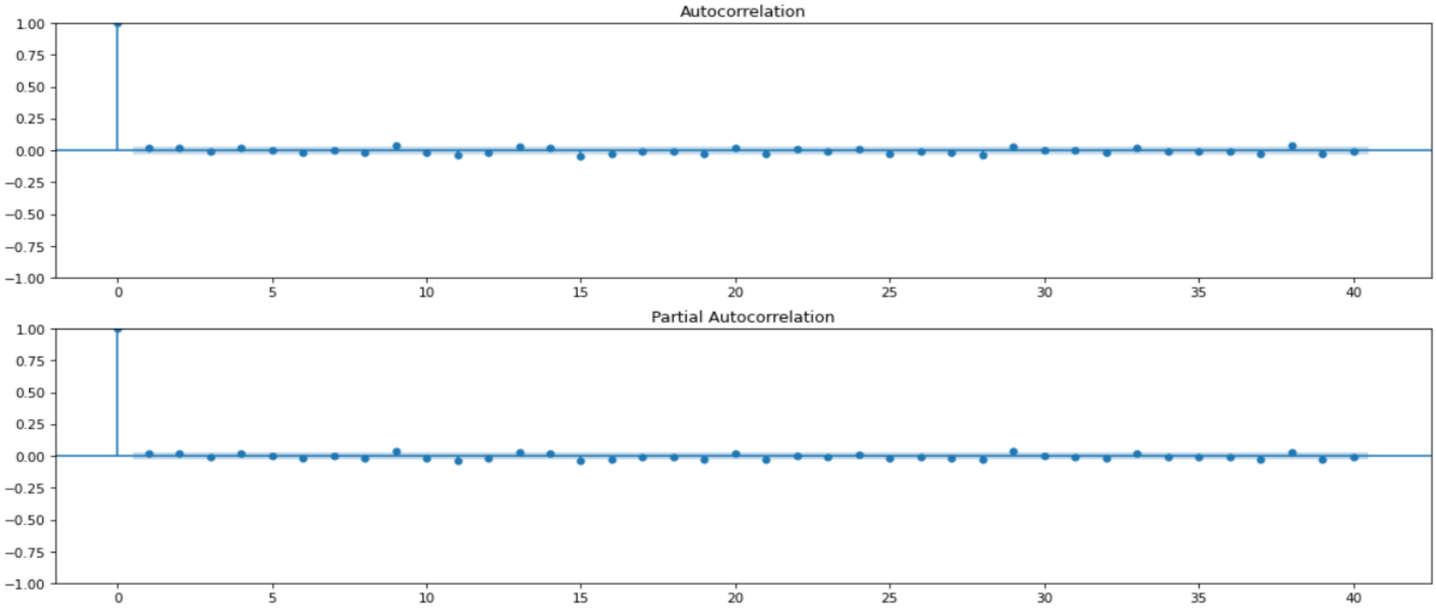
```
ADF Statistic: -0.700273
p-value: 0.846679
Critical Values:
     1%: -3.432
     5%: -2.862
     10%: -2.567
```

As the p- value for the ADF test, after differencing the data twice > alpha, we conclude that we have over differenced.



So, d=1 is preferred.

- **Now Analysis of acf and pacf plots of the differenced data (d=1)**



We can see that the acf as well as pacf plot cut off after lag 0. From this plot we can say p=0 and q=0 was maybe a good initial choice

So a potential model can be **ARIMA(0,1,0).**

Model Building

From the above discussions and results we came to know that ARIMA (0,1,0) be a good choice:

Model Summary:

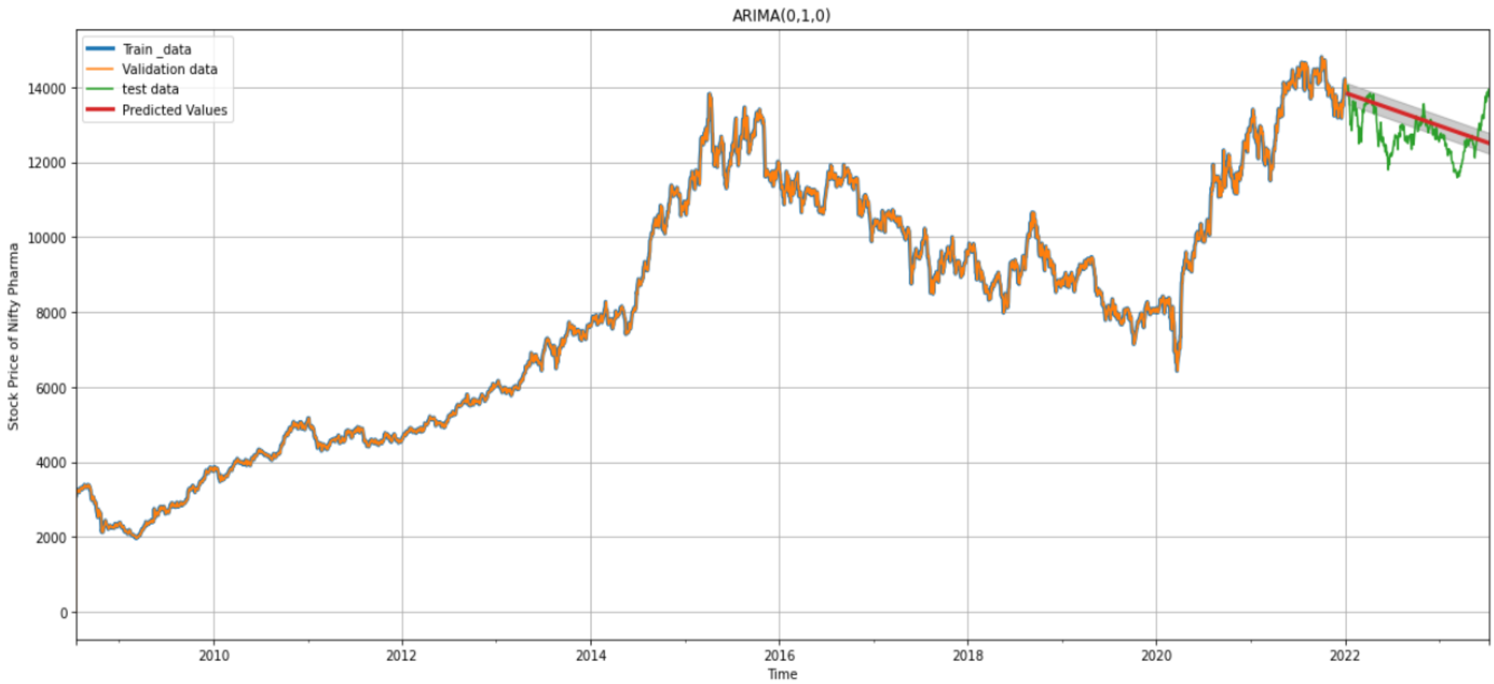
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                 3520
Model:                 ARIMA(0, 1, 0)   Log Likelihood              -21439.010
Date:                Mon, 31 Jul 2023   AIC                          42880.020
Time:                        11:23:45   BIC                          42886.186
Sample:                    07-17-2008   HQIC                         42882.220
                         - 01-12-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      1.146e+04    126.012     90.938      0.000    1.12e+04    1.17e+04
===================================================================================
Ljung-Box (L1) (Q):                   2.03   Jarque-Bera (JB):              8169.94
Prob(Q):                              0.15   Prob(JB):                         0.00
Heteroskedasticity (H):              10.96   Skew:                            -0.30
Prob(H) (two-sided):                  0.00   Kurtosis:                        10.44
===================================================================================
```

Result:

- Margin of Error (ME): ME = 1.2853275 #CHange
- Maximum 95% CI length: 2.5706550

Model Performance:

Now to see how our model fitted the actual series, we can plot the fitted values and the true values to get the idea about the performance of our model as follows:



- **AUTO-ARIMA Method:**

  In this method, I am going to apply a Grid-Search method to get the best optimum values of $p, d, and\ q$ based on the **AIC and BIC criteria**. And the results of the Grid-Search method is as follows:

```
best_order1 = None
best_order2 = None

for param in pdq:
    try:
        # Fit ARIMA model
        model = sm.tsa.ARIMA(train1['Close'], order=param)
        results = model.fit()

        #forecast=result.forecast(591)
        # Calculate AIC and BIC
        aic = results.aic
        bic = results.bic
        # Check if the current model has lower AIC or BIC than the best model so far
        if aic < best_aic:
            best_aic = aic
            best_order1 = param
        if bic < best_bic:
            best_bic=bic
            best_order2= param


    except Exception as e:
        continue

# Print the best model's parameters, AIC, and BIC
print("Best AIC:", best_aic)
print("Best ARIMA wrt AIC :", best_order1)
print("Best BIC:", best_bic)
print("Best ARIMA wrt BIC :",best_order2)
```

```
Best AIC: 42871.930089398535
Best ARIMA wrt AIC : (4, 1, 3)
Best BIC: 42886.18570621029
Best ARIMA wrt BIC : (0, 1, 0)
```

Thus, from above results, it can be concluded that the best ARIMA model for our time series data is:
ARIMA (4,1,3) and ARIMA (0,1,0)

Since we have already built our model on ARIMA (0,1,0) in our first model.

▪ Fitting the ARIMA (4, 1, 3) model:

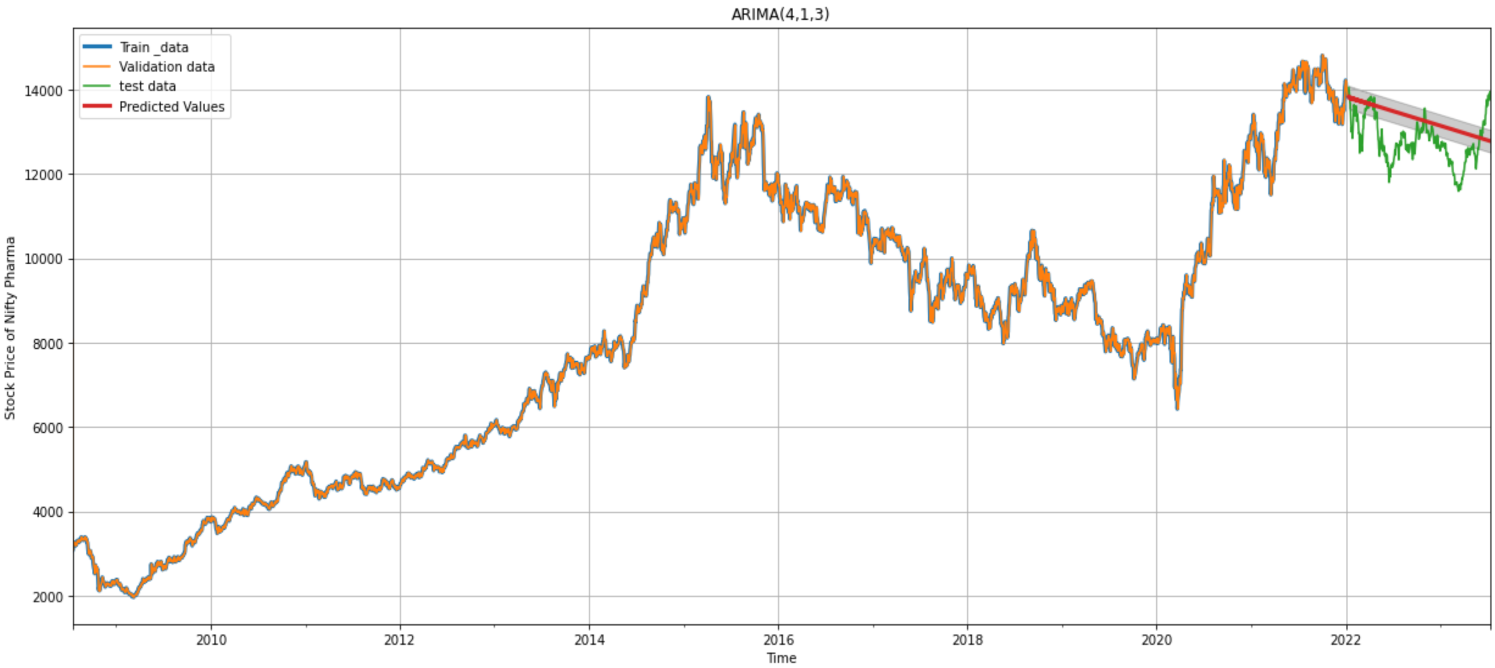
Model Summary:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                 3520
Model:                 ARIMA(4, 1, 3)   Log Likelihood              -21427.965
Date:                Mon, 31 Jul 2023   AIC                          42871.930
Time:                        18:49:15   BIC                          42921.258
Sample:                    07-17-2008   HQIC                         42889.530
                         - 01-12-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1283      0.276     -0.465      0.642      -0.669       0.413
ar.L2         -0.5546      0.158     -3.517      0.000      -0.864      -0.245
ar.L3         -0.6990      0.273     -2.557      0.011      -1.235      -0.163
ar.L4          0.0390      0.012      3.341      0.001       0.016       0.062
ma.L1          0.1472      0.275      0.534      0.593      -0.393       0.687
ma.L2          0.5664      0.159      3.569      0.000       0.255       0.877
ma.L3          0.7103      0.272      2.613      0.009       0.178       1.243
sigma2       1.14e+04    131.787     86.535      0.000    1.11e+04    1.17e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):              7863.54
Prob(Q):                              0.86   Prob(JB):                         0.00
Heteroskedasticity (H):              10.88   Skew:                            -0.26
Prob(H) (two-sided):                  0.00   Kurtosis:                        10.30
===================================================================================
```

Interpretation: Looking to the summary of ARIMA (4,1,3) model, one can easily see that, the 1st and 3rd coefficient of AR i.e., ar. L1 and ar.L3, 1 st coefficient of MA i.e., ma. L1 are statistically insignificant as the corresponding p-values are greater than 0.05, while rest of the coefficients are statistically significant.

Model Performance: ARIMA (4,1,3)

Now to see how our model fitted the actual series, we can plot the fitted values and the true values to get the idea about the performance of our model as follows:



• Seasonal ARIMA Method (SARIMA):

1. SARIMA fitting

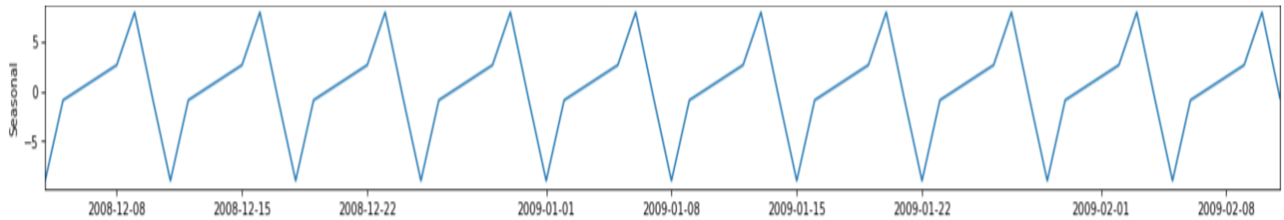
SARIMA (Seasonal Autoregressive Integrated Moving Average) modeling is a time series forecasting technique that extends the ARIMA model to account for seasonality in the data. It is useful when working with data that exhibits a repeating pattern or seasonal variations. The SARIMA model is typically denoted as SARIMA (p, d, q)(P, D, Q, s), where:

- **p: The order of the autoregressive (AR) component.**
- **d: The degree of differencing required to make the time series stationary.**
- **q: The order of the moving average (MA) component.**
- **P: The order of the seasonal autoregressive (SAR) component.**
- **D: The degree of seasonal differencing.**
- **Q: The order of the seasonal moving average (SMA) component.**
- **s: The length of the seasonal cycle.**

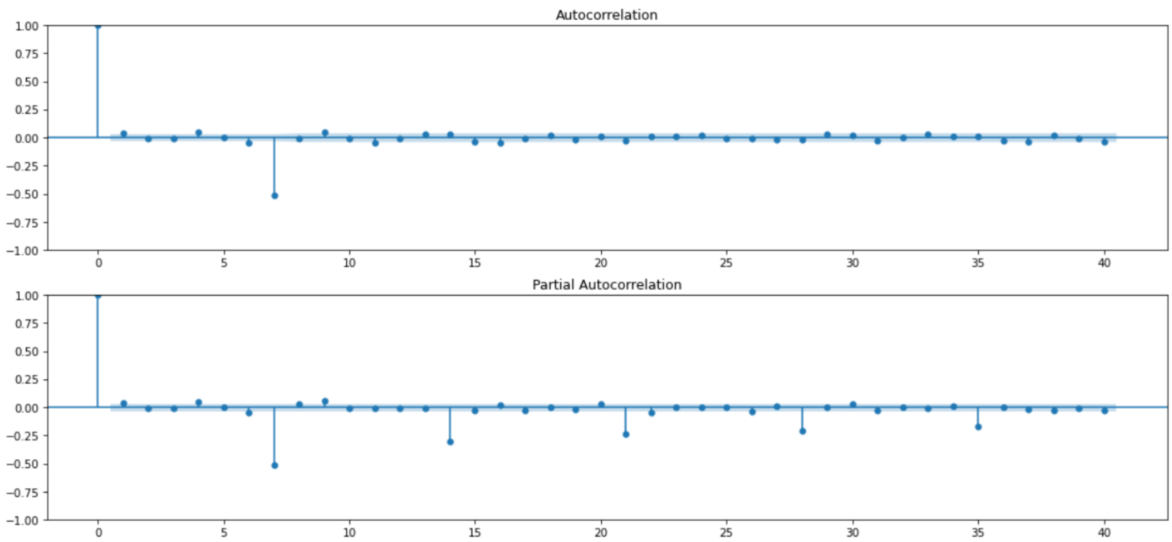
To build a SARIMA model, we typically start by analysing the time series data, identifying the seasonal pattern, and checking for stationarity. Then, we use methods like autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine the appropriate values for the model parameters. Finally, we can estimate the parameters using techniques like maximum likelihood estimation and perform model diagnostics to assess the model's fit and make prediction.

Finding parameters of ARIMA model manually:

Now, from the decomposition of my time series, I have seen that my time series have some seasonality patten and the period of the seasonality is around 7 days, therefore, I am going to fit the Seasonal ARIMA model



- **Now Analysis of ACF and PACF plots of the differenced data (d=1) at 7th lag**



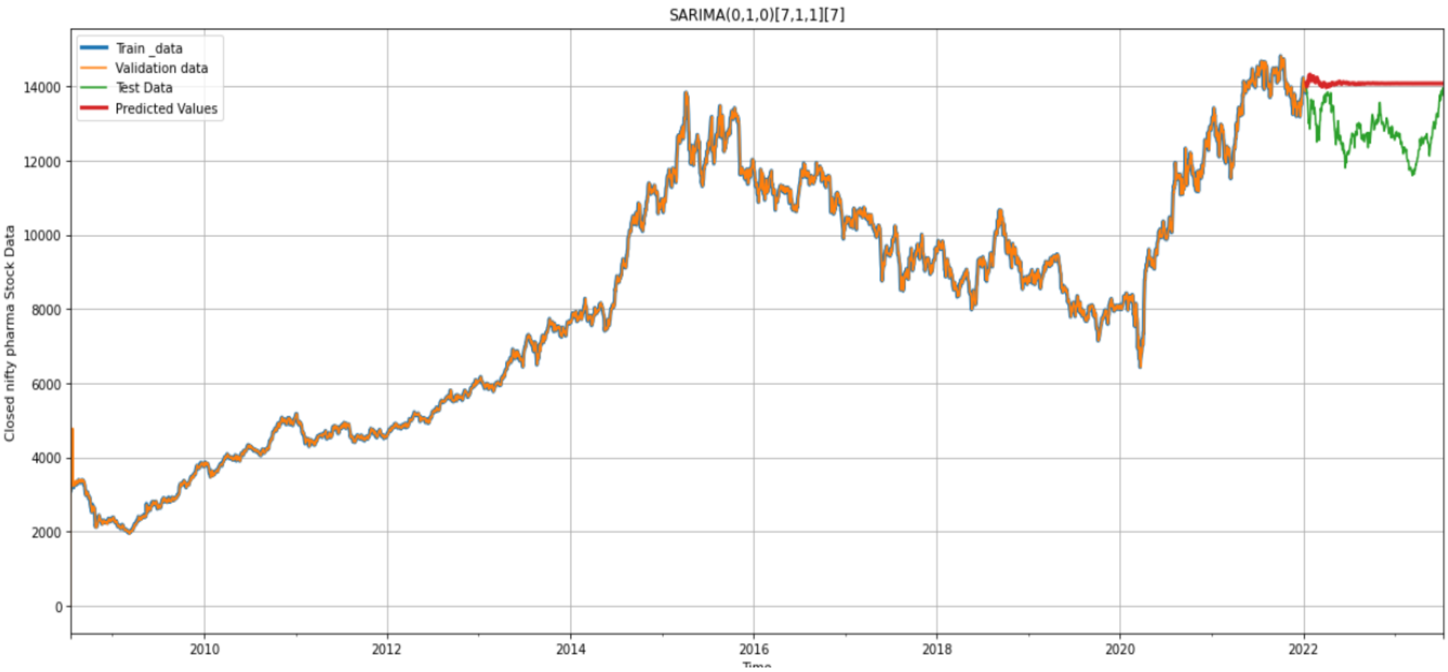
We can see that the acf as well as pacf plot cut off after lag 0. From this plot we can say p=0 and q=0 was maybe a good initial choice. I have seen that my time series have some seasonality patten and the period of the seasonality is around 7 days, and to determine the seasonal AR(P) is 7 and seasonal MA(q) is 1 as it shut off after 1st seasonal lag. The SARIMA model to my time series data and the summary of the SARIMA model is as follows: SARIMA (0,1,0) [7,1,1,7]

Model Summary:

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 Close   No. Observations:             3520
Model:             SARIMAX(0, 1, 0)x(7, 1, [1], 7)   Log Likelihood          -21413.284
Date:                               Tue, 01 Aug 2023   AIC                      42844.568
Time:                                       08:49:43   BIC                      42900.044
Sample:                                   07-17-2008   HQIC                     42864.363
                                        - 01-12-2022
Covariance Type:                                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.S.L7        0.0020      0.012      0.166      0.868      -0.022       0.026
ar.S.L14       0.0299      0.013      2.288      0.022       0.004       0.056
ar.S.L21      -0.0312      0.013     -2.460      0.014      -0.056      -0.006
ar.S.L28      -0.0353      0.014     -2.592      0.010      -0.062      -0.009
ar.S.L35      -0.0109      0.014     -0.776      0.438      -0.038       0.017
ar.S.L42       0.0036      0.014      0.253      0.801      -0.024       0.031
ar.S.L49       0.0079      0.014      0.588      0.556      -0.019       0.034
ma.S.L7       -1.0000      0.365     -2.739      0.006      -1.716      -0.284
sigma2      1.143e+04   4180.964      2.734      0.006    3234.833    1.96e+04
===================================================================================
Ljung-Box (L1) (Q):                   2.48   Jarque-Bera (JB):              8345.61
Prob(Q):                              0.12   Prob(JB):                         0.00
Heteroskedasticity (H):              11.05   Skew:                            -0.31
Prob(H) (two-sided):                  0.00   Kurtosis:                        10.53
===================================================================================
```

Model Performance:

Now to see how our model fitted the actual series, we can plot the fitted values and the true values to get the idea about the performance of our model as follows



<u>AUTO-SARIMA Method</u>:

In this method, I am going to apply a Grid-Search method to get the best optimum values of p, d, *and* q and Seasonal **P, D, Q** based on the **AIC and BIC criteria**. And the results of the Grid-Search method is as follows:

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,1,0)[7]             : AIC=45226.226, Time=0.24 sec
 ARIMA(1,1,0)(1,1,0)[7]             : AIC=44153.118, Time=0.98 sec
 ARIMA(0,1,1)(0,1,1)[7]             : AIC=inf, Time=1.72 sec
 ARIMA(1,1,0)(0,1,0)[7]             : AIC=45223.012, Time=0.12 sec
 ARIMA(1,1,0)(2,1,0)[7]             : AIC=43831.203, Time=1.77 sec
 ARIMA(1,1,0)(3,1,0)[7]             : AIC=43646.584, Time=2.55 sec
 ARIMA(1,1,0)(4,1,0)[7]             : AIC=43497.029, Time=5.84 sec
 ARIMA(1,1,0)(5,1,0)[7]             : AIC=43379.593, Time=6.11 sec
 ARIMA(1,1,0)(5,1,1)[7]             : AIC=inf, Time=29.20 sec
 ARIMA(1,1,0)(4,1,1)[7]             : AIC=inf, Time=17.17 sec
 ARIMA(0,1,0)(5,1,0)[7]             : AIC=43379.859, Time=2.30 sec
 ARIMA(2,1,0)(5,1,0)[7]             : AIC=43379.575, Time=8.94 sec
 ARIMA(2,1,0)(4,1,0)[7]             : AIC=43496.451, Time=5.39 sec
 ARIMA(2,1,0)(5,1,1)[7]             : AIC=inf, Time=35.66 sec
 ARIMA(2,1,0)(4,1,1)[7]             : AIC=inf, Time=23.81 sec
 ARIMA(3,1,0)(5,1,0)[7]             : AIC=43381.556, Time=10.74 sec
 ARIMA(2,1,1)(5,1,0)[7]             : AIC=43380.769, Time=17.78 sec
 ARIMA(1,1,1)(5,1,0)[7]             : AIC=inf, Time=38.66 sec
```

Thus, from above results, it can be concluded that the best SARIMA model based on lowest AIC value for our time series data is: SARIMA (1,1,0)(5,1,0)[7]

<u>Fitting the SARIMA (1,1,0)(5,1,0)[7] model:</u>

Model Summary:

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                               Close   No. Observations:                 3520
Model:             SARIMAX(1, 1, 0)x(5, 1, 0, 7)   Log Likelihood              -21682.796
Date:                             Tue, 01 Aug 2023   AIC                          43379.593
Time:                                     17:49:23   BIC                          43422.740
Sample:                                 07-17-2008   HQIC                         43394.989
                                      - 01-12-2022
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0247      0.012      2.056      0.040       0.001       0.048
ar.S.L7       -0.8179      0.012    -69.831      0.000      -0.841      -0.795
ar.S.L14      -0.6106      0.016    -38.708      0.000      -0.642      -0.580
ar.S.L21      -0.4774      0.016    -29.381      0.000      -0.509      -0.446
ar.S.L28      -0.3488      0.016    -21.811      0.000      -0.380      -0.317
ar.S.L35      -0.1835      0.013    -13.953      0.000      -0.209      -0.158
sigma2      1.347e+04    157.131     85.699      0.000    1.32e+04    1.38e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              6346.33
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):              11.33   Skew:                            -0.29
Prob(H) (two-sided):                  0.00   Kurtosis:                         9.56
===================================================================================
```

Model Performance:

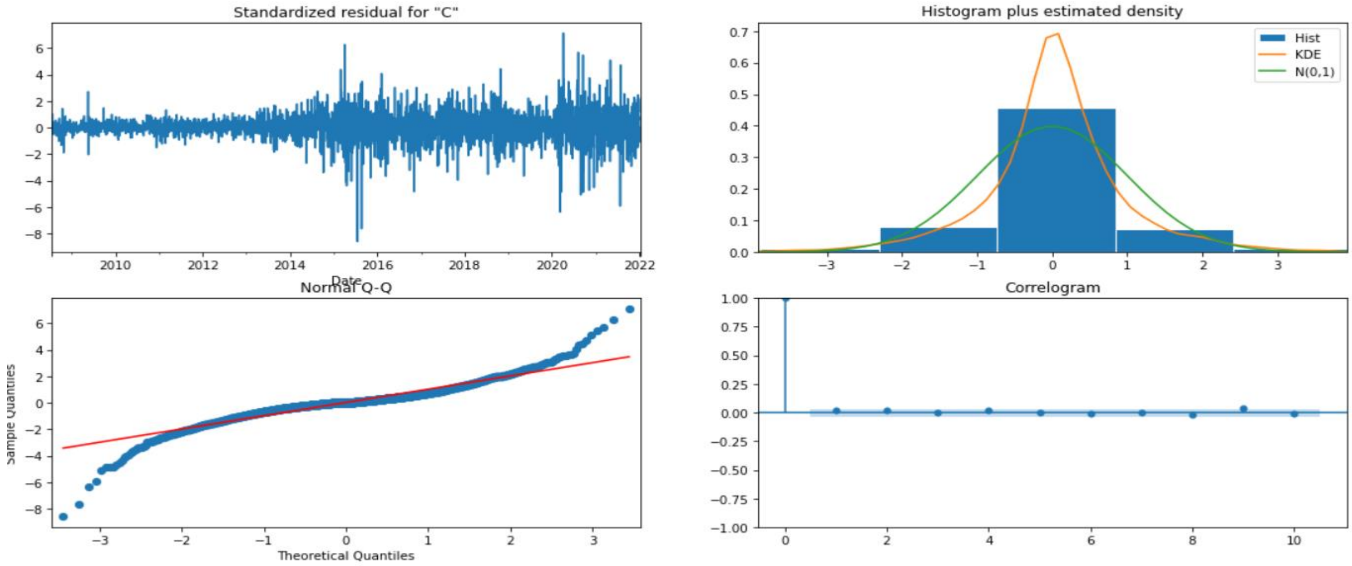
Now to see how our model fitted the actual series, we can plot the fitted values and the true values to get the idea about the performance of our model as follows

Model Diagnostics

We conduct diagnostic checks for each of the models based on the residuals and their auto-correlation function:

1. ARIMA (0,1,0)
2. ARIMA (4,1,3)
3. SARIMA (0,1,0) (7,1,1) [7]
4. SARIMA (0,1,0) (5,1,0) [7]

- **Model Diagnostics for the ARIMA (0,1,0) Model:**



Interpretation: • From the above 4 different plots of the residual of the ARIMA model, we have the following

Plot 1: From the plot 1, which is representing variation of the standardized residuals over the time, and from the plot one can easily see that the residuals are fluctuating around the ZERO line of residuals and almost having the constant variance if we can ignore the some of the high spikes of the residuals.

Plot 2: the histogram, KDE (Kernal Density Estimation) of the residuals and the density function of the standard normal Distribution, and seeing/ Comparing KDE with the standard Normal Density Curve one can easily interpret that the distribution of the residuals is approximately normal.

Plot 3: The plot of Sample quantiles of the residual's vs the Theoretical Quantiles of the residuals and from the plot one can easily see that most of the points are lies on the straight line, depicting the fact that distribution of the residuals is approximately Normal.

Plot 4: The correlogram (The plot the Auto-Correlation Function Vs The lags k is called the Correlogram) and from the plot, one can easily see that none of the autocorrelation is statistically significant as they lie in the blue zone, thus this shows that the residuals are from the White Noise process.

Performing the Ljung-Box test on the residuals of the ARIMA (0,1,0) model:

	lb_stat	lb_pvalue
1	0.141995	0.706306
2	0.635601	0.727748
3	0.638373	0.887594
4	0.912477	0.922764
5	1.896064	0.863332
6	4.258770	0.641703
7	4.277073	0.747368
8	5.850191	0.664008
9	6.294437	0.710126
10	9.225479	0.510847
11	9.273004	0.596709
12	10.444051	0.577065
13	13.303755	0.424632
14	13.567762	0.482384
15	14.912486	0.457739
16	16.735332	0.402920
17	18.329043	0.368373
18	19.494274	0.361999

Here Our Hypothesis for Ljung Box test is: -

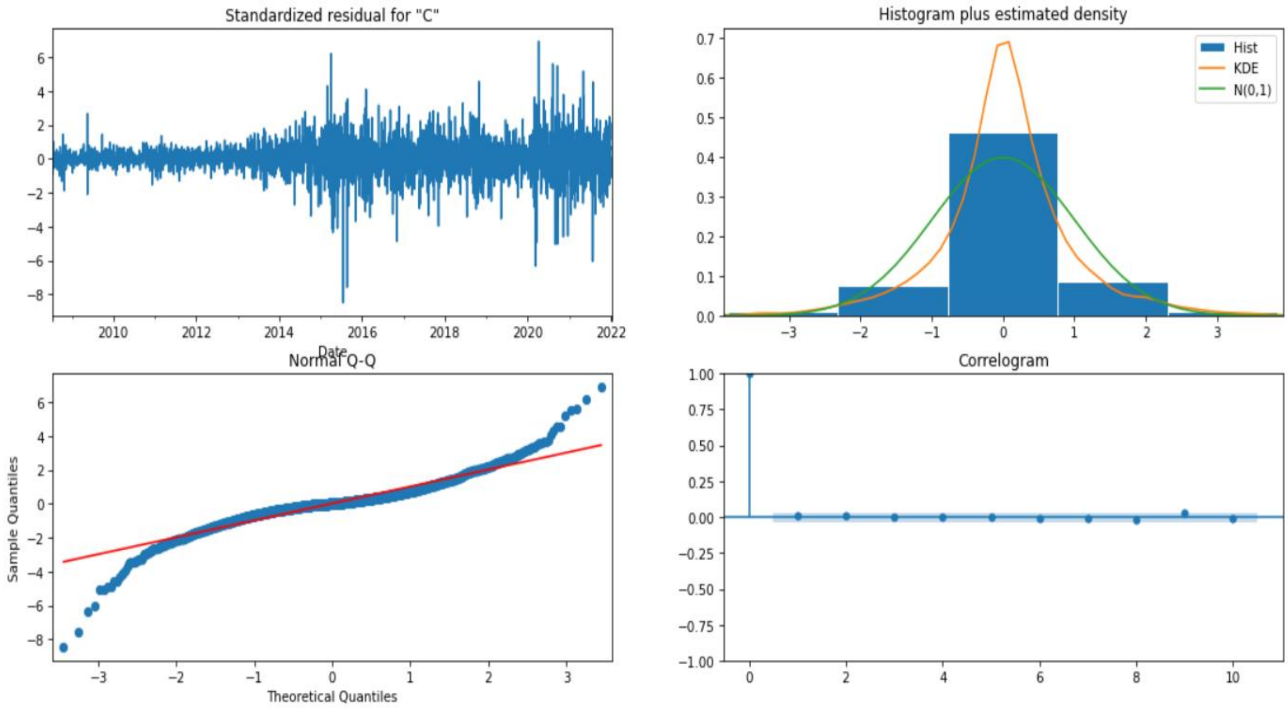
▪ $\boldsymbol{Null\ Hypothesis\ (H0\ ):}$ **The residuals are independently distributed**

vs

▪ $\boldsymbol{Alternative\ Hypothesis\ (H1\ ):}$ **The residuals are not independently distributed; they exhibit serial correlation.**

In summary, based on the provided results, we do not have significant evidence to reject the null hypothesis of no autocorrelation in the data for any of the lags tested

• **Model Diagnostics for the ARIMA (4,1,3) Model:**



Interpretation: • From the above 4 different plots of the residual of the ARIMA model, we have the following

Plot 1: From the plot 1, which is representing variation of the standardized residuals over the time, and from the plot one can easily see that the residuals are fluctuating around the ZERO line of residuals and almost having the constant variance if we can ignore the some of the high spikes of the residuals.

Plot 2: the histogram, KDE (Kernal Density Estimation) of the residuals and the density function of the standard normal Distribution, and seeing/ Comparing KDE with the standard Normal Density Curve one can easily interpret that the distribution of the residuals is approximately normal.

Plot 3: The plot of Sample quantiles of the residual's vs the Theoretical Quantiles of the residuals and from the plot one can easily see that most of the points are lies on the straight line, depicting the fact that distribution of the residuals is approximately Normal.

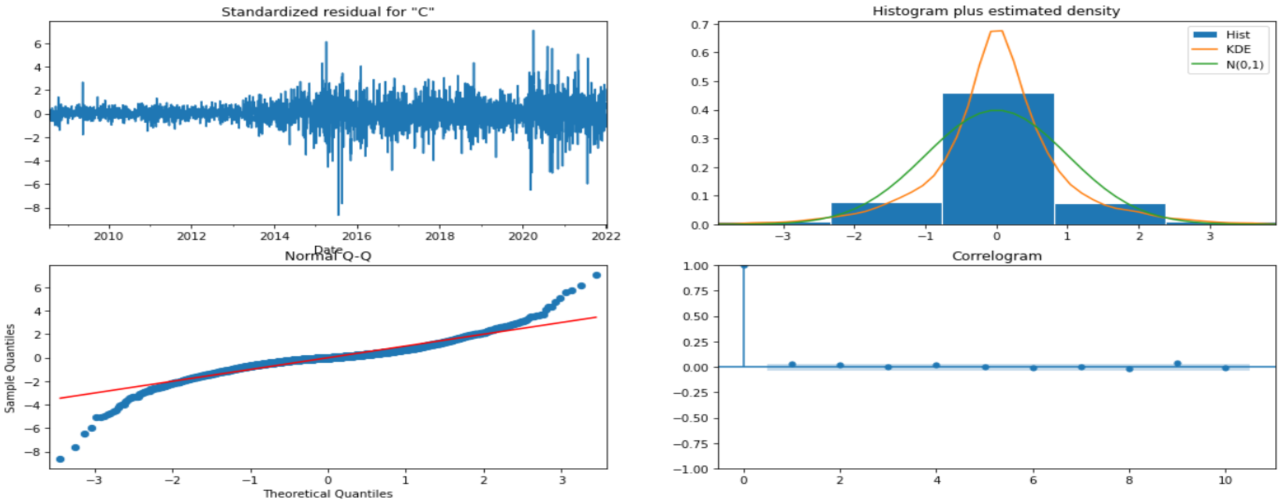
Plot 4: The correlogram (The plot the Auto-Correlation Function Vs The lags k is called the Correlogram) and from the plot, one can easily see that none of the autocorrelation is statistically significant as they lie in the blue zone, thus this shows that the residuals are from the White Noise process.

Performing the Ljung-Box test on the residuals of the ARIMA (4,1,3) model:

	lb_stat	lb_pvalue
1	0.160437	0.688754
2	0.221037	0.895370
3	0.226990	0.973119
4	0.407983	0.981818
5	2.014878	0.847083
6	4.235596	0.644828
7	4.444380	0.727404
8	5.264530	0.728965
9	5.271653	0.810014
10	7.075978	0.718253
11	7.485828	0.758487
12	10.027065	0.613586
13	13.229812	0.430221
14	13.229994	0.508499
15	13.996567	0.525789
16	16.422362	0.423892

In summary, based on the provided results, we do not have significant evidence to reject the null hypothesis of no autocorrelation in the data for any of the lags tested

- **Model Diagnostics for the SARIMA (0,1,0) (7,1,1) [7]Model:**



Interpretation: • From the above 4 different plots of the residual of the ARIMA model, we have the following

Plot 1: From the plot 1, which is representing variation of the standardized residuals over the time, and from the plot one can easily see that the residuals are fluctuating around the ZERO line of residuals and almost having the constant variance if we can ignore the some of the high spikes of the residuals.

Plot 2: the histogram, KDE (Kernal Density Estimation) of the residuals and the density function of the standard normal Distribution, and seeing/ Comparing KDE with the standard Normal Density Curve one can easily interpret that the distribution of the residuals is approximately normal.

Plot 3: The plot of Sample quantiles of the residual's vs the Theoretical Quantiles of the residuals and from the plot one can easily see that most of the points are lies on the straight line, depicting the fact that distribution of the residuals is approximately Normal.

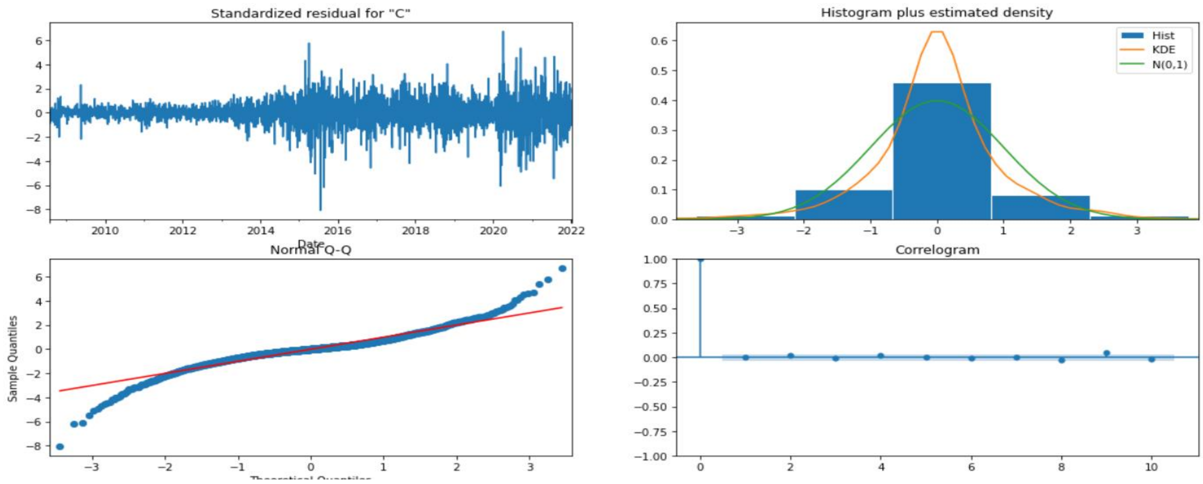
Plot 4: The correlogram (The plot the Auto-Correlation Function Vs The lags k is called the Correlogram) and from the plot, one can easily see that none of the autocorrelation is statistically significant as they lie in the blue zone, thus this shows that the residuals are from the White Noise process.

Performing the Ljung-Box test on the residuals of the SARIMA (0,1,0) (7,1,1) [7]Model:

	lb_stat	lb_pvalue
1	0.004143	0.948681
2	1.632696	0.442043
3	1.824689	0.609577
4	1.826546	0.767622
5	3.917069	0.561417
6	4.284604	0.638222
7	4.391502	0.733740
8	4.397038	0.819644
9	4.824297	0.849348
10	6.968117	0.728451
11	7.958298	0.717034
12	7.989686	0.785936
13	8.396980	0.816811
14	8.431941	0.865632
15	14.388194	0.496315
16	14.403285	0.568697
17	15.810473	0.537308

In summary, based on the provided results, we do not have significant evidence to reject the null hypothesis of no autocorrelation in the data for any of the lags tested

Model Diagnostics for the SARIMA (0,1,0) (5,1,0) [7] Model:



Interpretation: • From the above 4 different plots of the residual of the ARIMA model, we have the following

Plot 1: From the plot 1, which is representing variation of the standardized residuals over the time, and from the plot one can easily see that the residuals are fluctuating around the ZERO line of residuals and almost having the constant variance if we can ignore the some of the high spikes of the residuals.

Plot 2: the histogram, KDE (Kernal Density Estimation) of the residuals and the density function of the standard normal Distribution, and seeing/ Comparing KDE with the standard Normal Density Curve one can easily interpret that the distribution of the residuals is approximately normal.

Plot 3: The plot of Sample quantiles of the residual's vs the Theoretical Quantiles of the residuals and from the plot one can easily see that most of the points are lies on the straight line, depicting the fact that distribution of the residuals is approximately Normal.

Plot 4: The correlogram (The plot the Auto-Correlation Function Vs The lags k is called the Correlogram) and from the plot, one can easily see that none of the autocorrelation is statistically significant as they lie in the blue zone, thus this shows that the residuals are from the White Noise process.

Performing the Ljung-Box test on the residuals of the SARIMA (0,1,0) (7,1,1) [7]Model:

	lb_stat	lb_pvalue
1	0.440986	0.506647
2	2.552051	0.279145
3	2.696229	0.440869
4	2.771213	0.596812
5	4.469695	0.483948
6	7.788951	0.253977
7	7.848751	0.346120
8	7.937769	0.439573
9	8.526157	0.482111
10	12.296432	0.265707
11	12.410436	0.333596
12	14.400814	0.275848
13	16.653039	0.215654
14	16.671956	0.274081
15	19.490892	0.192344
16	20.030623	0.218845
17	21.369401	0.210201
18	21.495303	0.255164

In summary, based on the provided results, we do not have significant evidence to reject the null hypothesis of no autocorrelation in the data for any of the lags tested

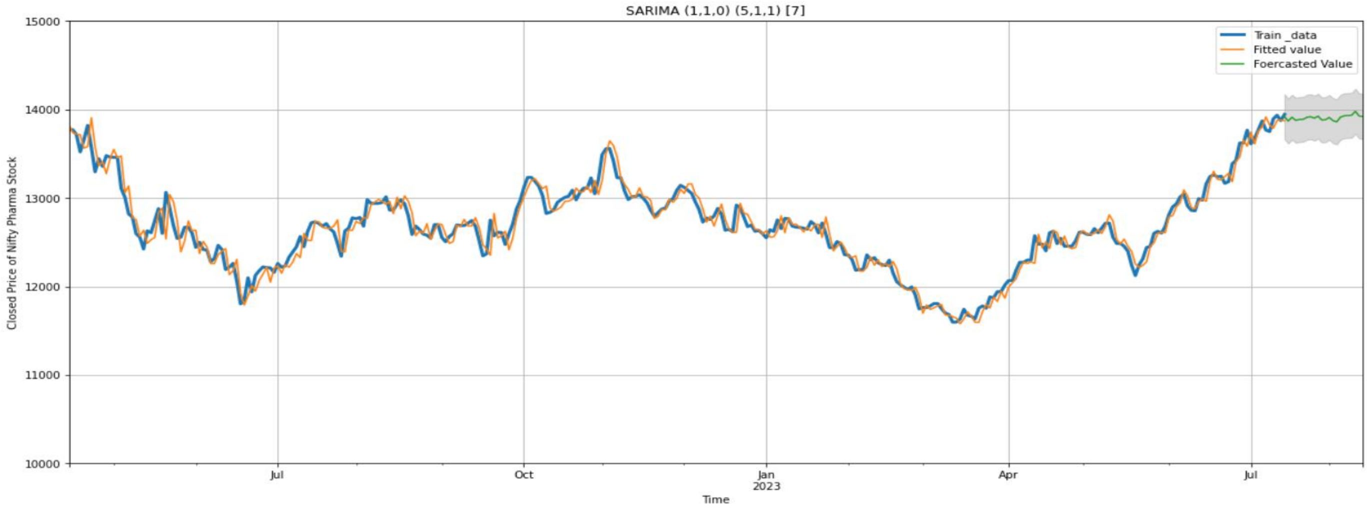
As we can see all the Models are satisfying the model assumption. Now we will choose the model based on MAPE (accuracy):

	Method	RMSE	MAPE	AIC	BIC
0	ARIMA(0,1,0)	656.34	4.21	862.310752	882.017645
0	ARIMA(4,1,3)	718.70	4.83	781.424893	798.179650
0	SARIMA(0,1,0)[7,1,1][7]	1294.92	9.35	862.310752	882.017645
0	SARIMA(1,1,0)[5,1,1][7]	588.88	3.78	571.188729	583.678057

As we can see from the table that SARIMA (1,1,0) (5,1,1)[7] has the lowest MAPE, Hence I'll continue with SARIMA (1,1,0) (5,1,1) [7]

SARIMA (1,1,0) (5,1,1) [7]

Here's the Forecasting value of Nifty Pharma Stock for next 30 days,



Conclusion: From the above plot showing the forecasted values for the next 1 months, one can see that our model is quite well as the forecasted values are approximately in the same pattern as in the past.

****