Time Series Analysis TATA STEEL Stock Price Prediction

Hindi Vidya Prachar Samiti's

R. J. COLLEGE

of Arts, Science & Commerce

(Autonomous College)

NAAC Re-Accredited 'A' Grade (CGPA 3.50)

DEPARTMENT OF STATISTICS

GROUP PROJECT

(2021 – 2022)

CLASS: M.Sc. - I **GROUP NUMBER:1**

TITLE: Time Series Analysis

Project Group Members		
UID	Name	Email ID
24	Shubham Suratran	shubhamsuratran2@gmail.com
25	DikshitaKandari	dikshitakandari11@gmail.com
12	AnaghaSuryawanshi	anaghavs020@gmail.com
35	Shweta Prajapati	shwetaprajapati53@gmail.com
49	Faisal Shaikh	shk.faisal020@gmail.com

_____ _____

PROF. MAYUR MORE **PROF. CHHAYA PINGE**

PROFESSOR IN-CHARGE **HEAD OF STATISTICS DEPARTMENT**

I N D E X

Dataset information in detail:

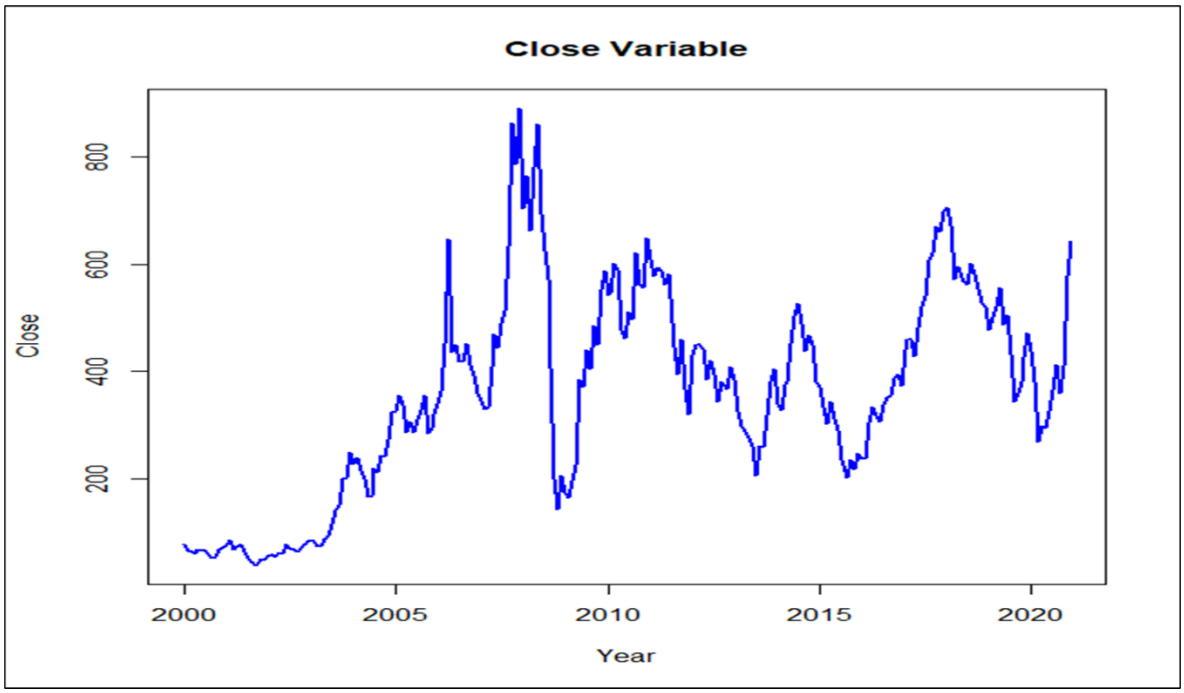
We have taken the monthly data of stock price for Tata Steel from year 2000 January to 2020 December, where we had various components depending on the stock market terms.

Our dataset contains 6 columns of indicators and date columns which will be then set as index.

Data Source:

link of data

Data Visualization



Objective:

The scope of this project is to predict future values of TataSteel Company closing stock price using Time-Series Forecasting methods in order to achieve the highest possible accuracy. This can be broken down into the following milestones:

1. Data Exploration and evaluation of Stationarity.
2. Modelling and selection of best model.
3. Prediction of future values for next 12 months.

Methodology:

All the analysis part was carried in R software. The libraries of R used for the time series analysis are

1)forecast

2)tseries

3)timeSeries

4)TTR

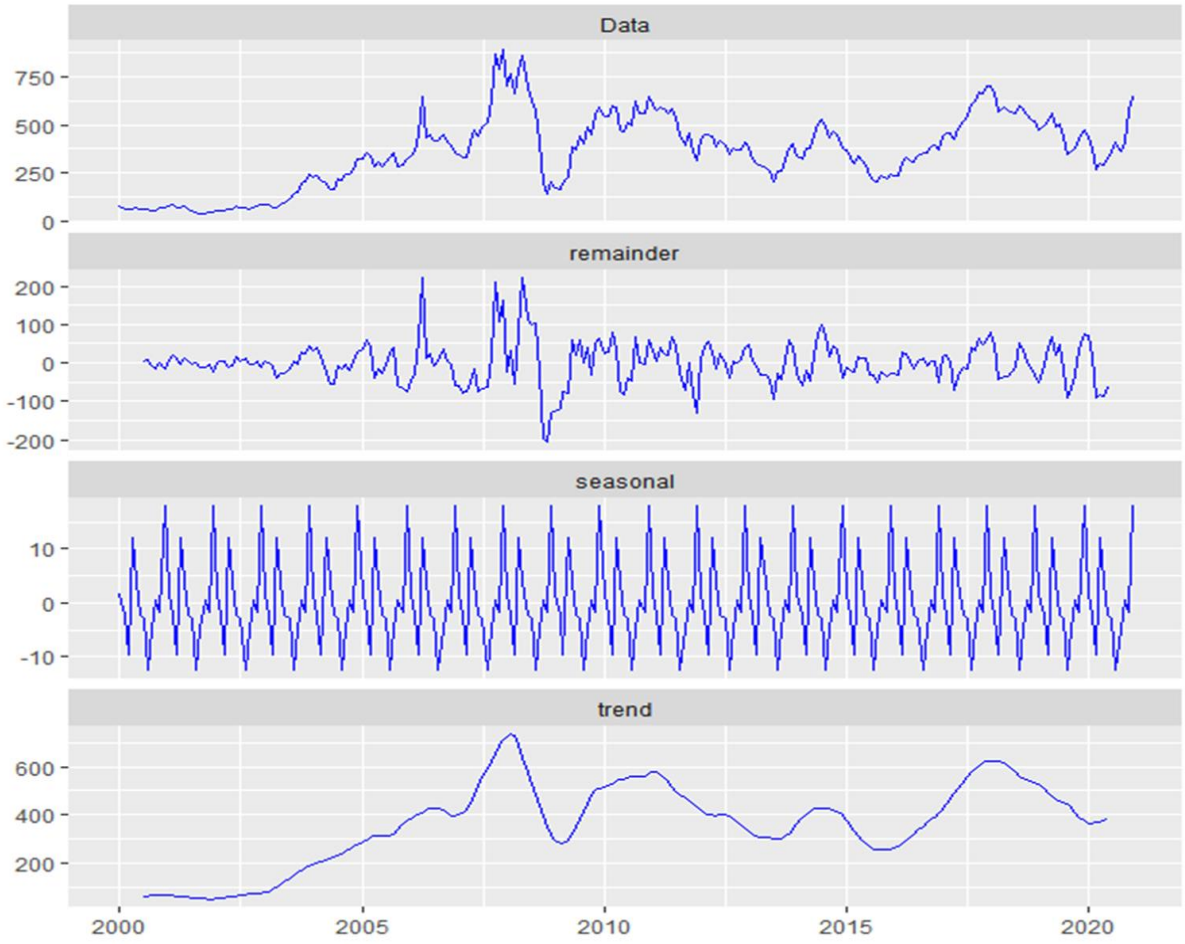
5)ggplot

6)ggfortify

CODES: R Script of this project can be accessed through the given link below. Link to code :R codes

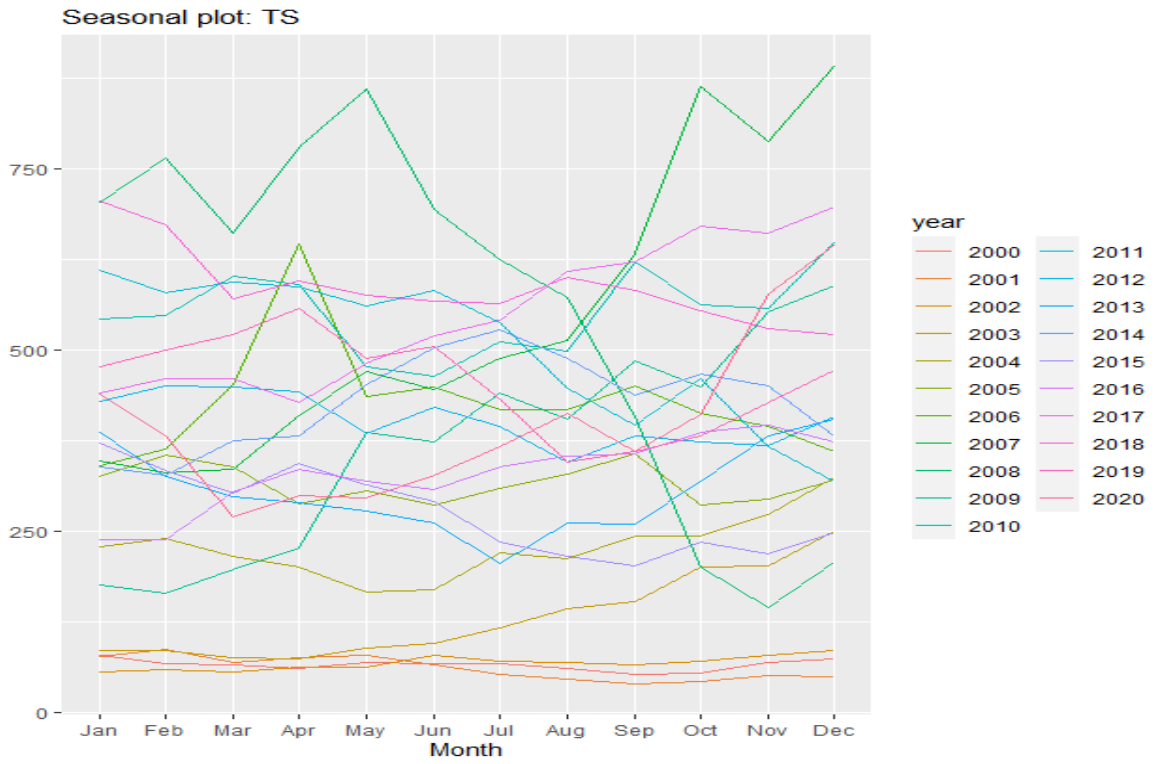
Analysis:

Decomposition:



Time series decomposition involves thinking of a series as a combination of level, trend, seasonality.From the above graph we can conclude that our data has trend component but for the clear picture of seasonalitywe have plot seasonal plot.

Checking seasonality by seasonal plot:



In time series data, seasonality refers to the presence of variations which occur at certain regular intervals either on a weekly basis, monthly basis, or even quarterly (but never up to a year) but from the above graph we can conclude that in our data variations is absent. Hence seasonality is not present.

Testing for Stationarity

Augmented Dickey-Fuller test:

H0: Data is non-Stationary v/s H1: Data is Stationary

p-value: 0.07667

Conclusion: Since p-value > 0.05 therefore data is non-Stationary

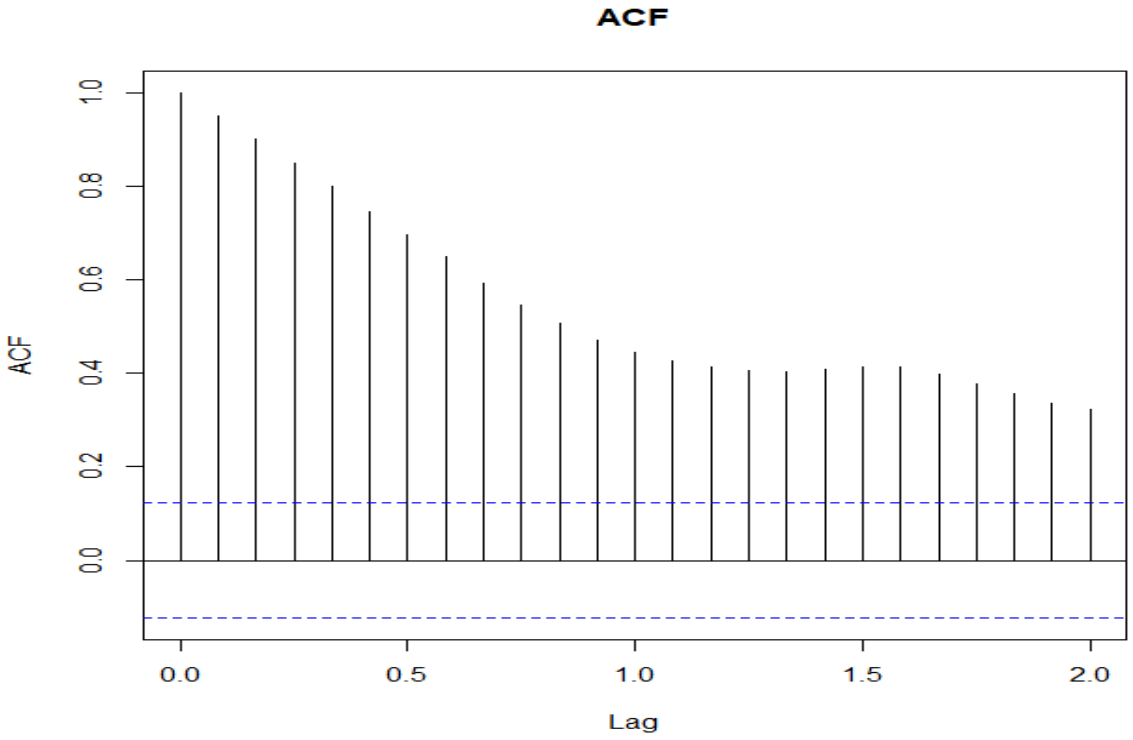
Ljung-Box test:

H0: Data is Stationary v/s H1: Data is non-Stationary

p-value: 2.2e-16

Conclusion: Since p-value <0.05 therefore data is non-Stationary

Autocorrelation plot



Conclusion:The ACF of datadrop to zero relativelyslowlyso it is non-Stationary.

To convert the data to Stationary,we use difference function.

So after using difference function, we again perform the ADF and Ljung-Box test.

Augmented Dickey-Fuller test:

H0: Data is non-Stationary v/s H1: Data is Stationary

p-value: 0.01

Conclusion: Since p-value < 0.05 therefore data is Stationary

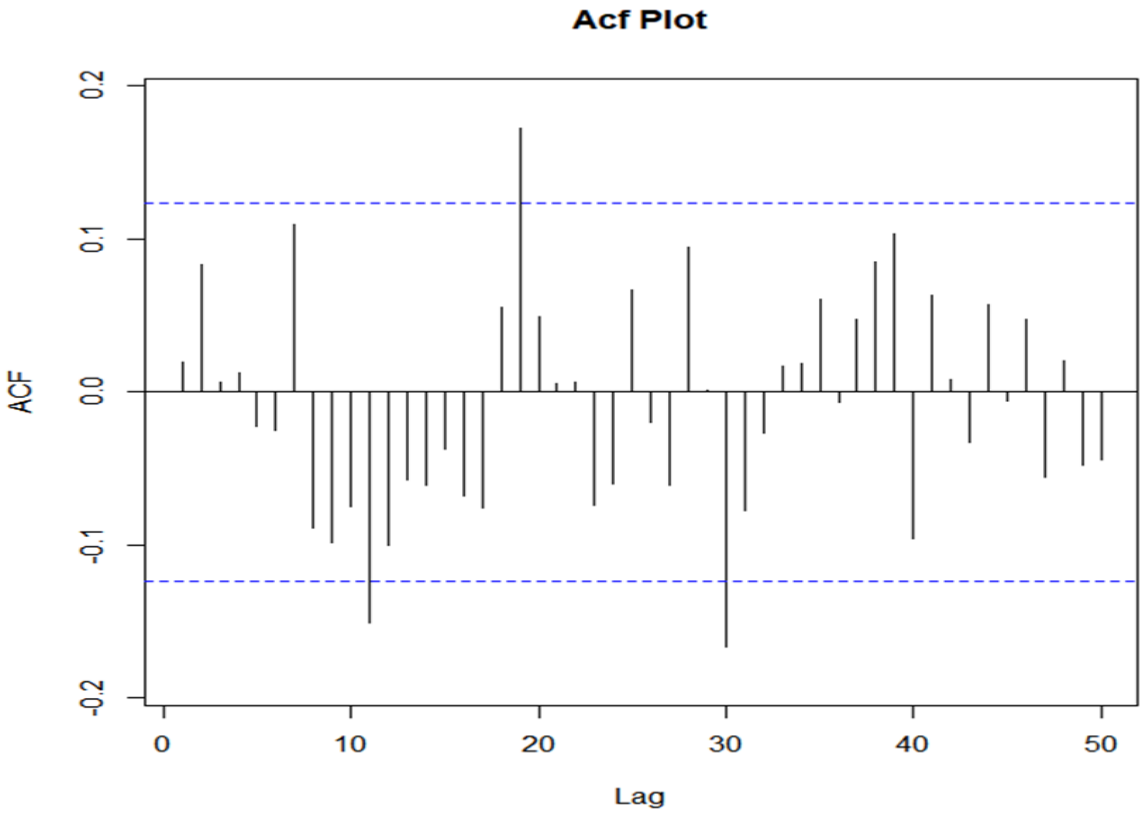
Ljung-Box test:

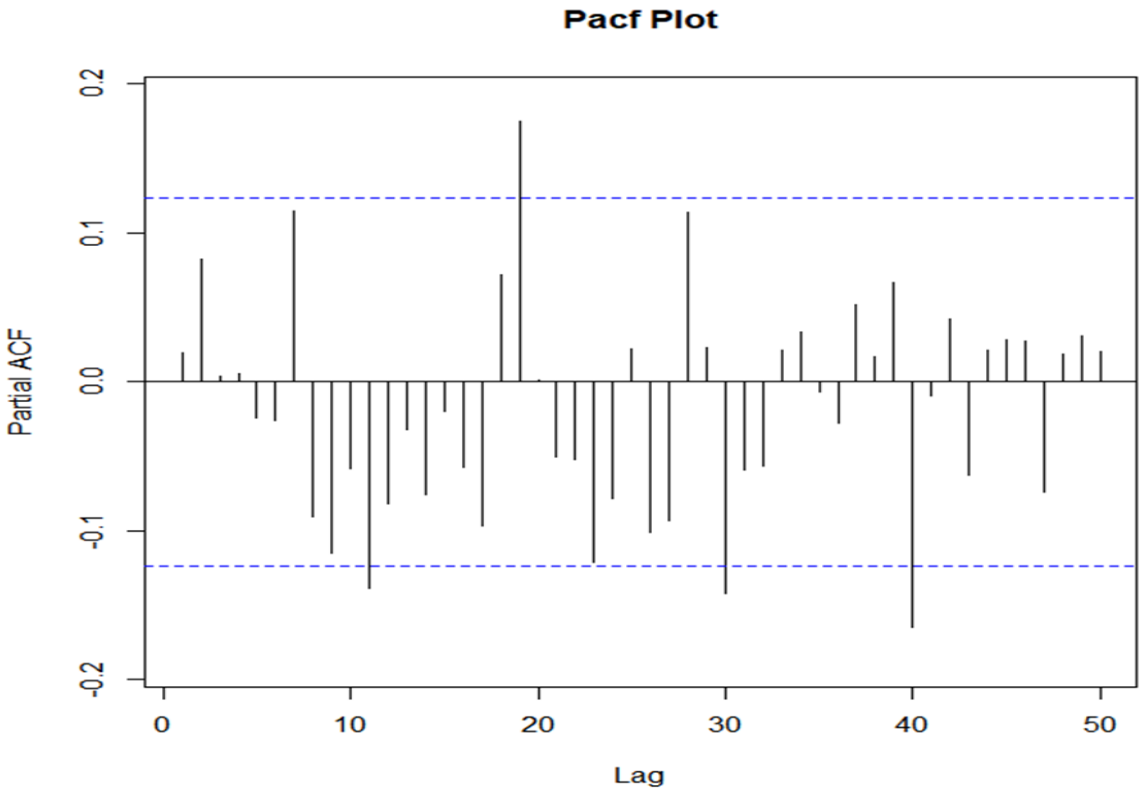
H0: Data is Stationary v/s H1: Data is non-Stationary

p-value: 0.9013

Conclusion: Since p-value > 0.05 therefore data is Stationary

Autocorrelation plot and Partial-autocorrelation plot





In time series analysis, Autocorrelation Function (ACF) and the partial autocorrelation function (PACF) plots are essential in providing the model's orders such as q for MA and p for AR respectively. From above we get p=4 & q=3.

Auto.arima Function:

From Auto.arima Function we get to know the best model:

And from that our best model is (5,1,0)

Train and Test:

To prevent the model from overfitting and to accurately evaluate the model

We use train data to fit the model and test accuracy on the test data

Train Data: Used to estimate any Parameters of forecasting methods.

We have selected first 200 observations for train data set. (i.e80%)

Test Data: Used to evaluate its accuracy.

The test set should be the most recent part of data.

51 observations left for test data set. (i.e20%)

We check whether our predicted value based on train dataset match the actual values of test dataset. If it is same which means our forecast model is a good fit.

From above we conclude the MAPE value

MAPE=4.161563

With a MAPE of around 4.2%, the model is 95.8% accurate in predicting the next 12 months observations.

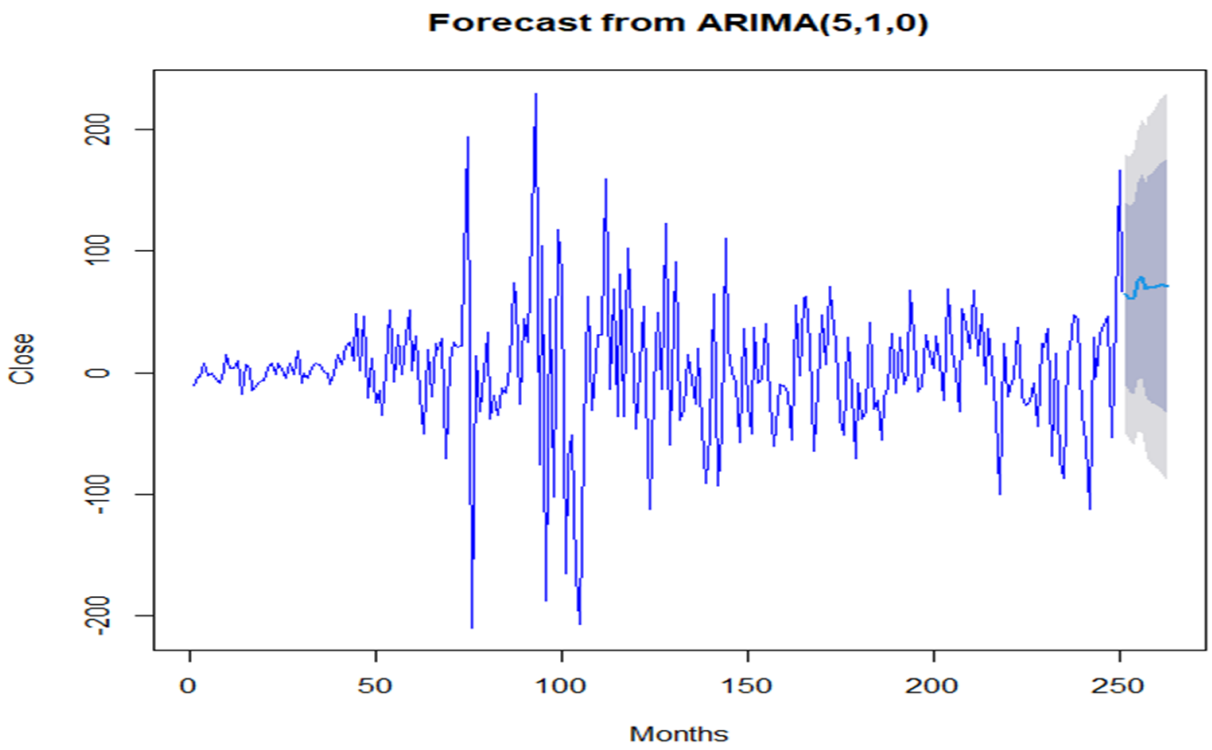
So from this we can say our model is good fit.

Forecasting:

As we had seen that our model is the best fit one so we can further forcast it .We have forecasted the values for next 12 months.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
252	64.70374	-10.332999	139.7405	-50.05505	179.4625
253	60.79819	-15.407507	137.0039	-55.74837	177.3448
254	61.35096	-17.669775	140.3717	-59.50082	182.2027
255	75.10137	-6.110059	156.3128	-49.10079	199.3035
256	79.33740	-4.722202	163.3970	-49.22067	207.8955
257	68.98212	-18.364516	156.3288	-64.60303	202.5673
258	69.68787	-21.939514	161.3153	-70.44412	209.8199
259	69.49307	-24.637265	163.6234	-74.46685	213.4530
260	70.55191	-26.339721	167.4435	-77.63105	218.7349
261	71.71963	-27.800341	171.2396	-80.48303	223.9223
262	71.46168	-30.756407	173.6798	-84.86739	227.7907
263	70.51385	-34.370103	175.3978	-89.89231	230.9200

Forecasting plot:



Double Exponential Smoothing:

Double exponential Smoothing it is also known as Holts model or second-order exponential smoothing.We choose this model because our data follows trend but not seasonality.Double exponential smoothing uses the level and trend components to generate forecasts. It has 2 smoothing parameter alpha and beta where alpha is smoothing factory for level and beta is smoothing factor for trend.Trend shows general tendency of data.

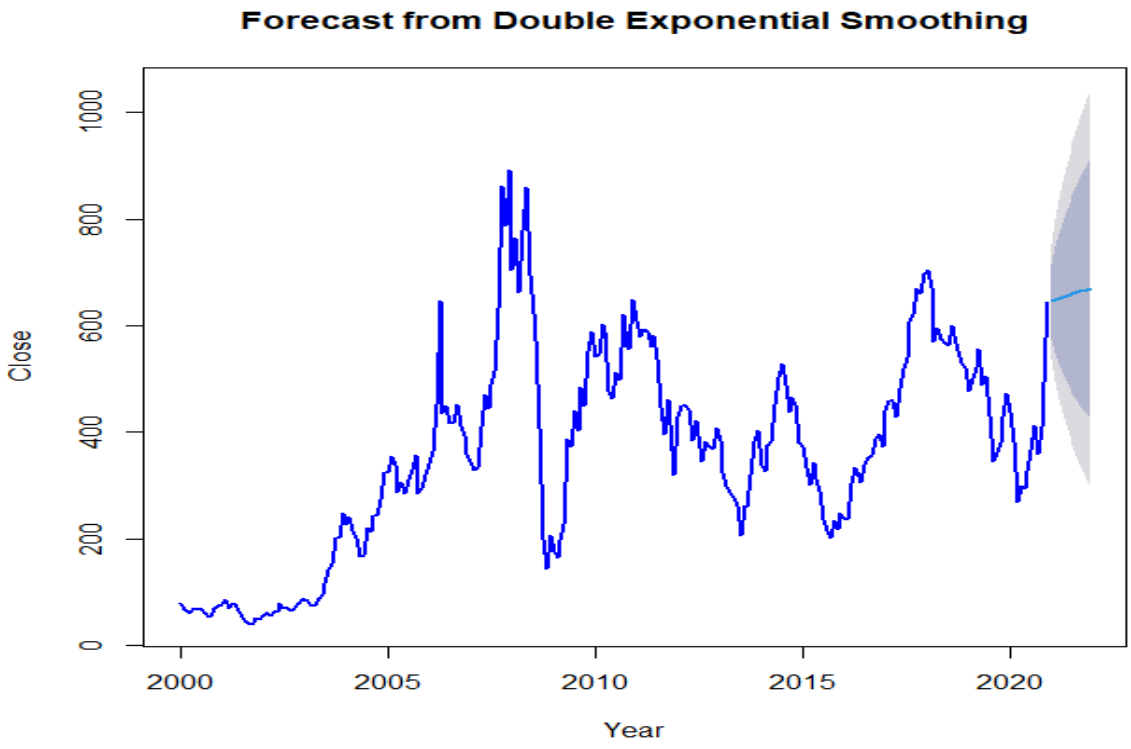
Forecasted values:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2021	645.8885	575.5162	716.2609	538.2633	753.5138
Feb 2021	648.1334	548.6119	747.6550	495.9284	800.3385
Mar 2021	650.3784	528.4858	772.2709	463.9598	836.7970
Apr 2021	652.6233	511.8680	793.3786	437.3566	867.8899
May 2021	654.8682	497.4919	812.2445	414.1820	895.5544
Jun 2021	657.1131	484.7080	829.5182	393.4422	920.7840
Jul 2021	659.3580	473.1304	845.5857	374.5474	944.1687
Aug 2021	661.6029	462.5076	860.6983	357.1129	966.0930
Sep 2021	663.8479	452.6651	875.0307	340.8717	986.8241
Oct 2021	666.0928	443.4757	888.7099	325.6293	1006.5562
Nov 2021	668.3377	434.8435	901.8319	311.2391	1025.4363
Dec 2021	670.5826	426.6938	914.4714	297.5869	1043.5783

Above is the forecasted values for 12 Months.

From this model we get our alpha value as 0.999 and beta as $1e^{-04}$

Forecasted plot:



This is the graphical representation of the forecasted values in which the light blue line indicates the forecasted values for next 12 months and as you can see the line is slightly tilted upward it means trend is present.

Concluding remarks:

Time Series forecasting is really useful when we have to take future decisions or we have to do analysis, we can quickly do that using ARIMA, there are lots of other Models from we can do the time series forecasting but ARIMA is really easy to understand.

Reference used:

For stationarity :1)https://rpubs.com/richkt/269797

2)https://stats.stackexchange.com

Analysis :https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html

Importing stock data:

https://analyticswithfubu.wordpress.com/2017/10/28/importing-and-visualising-stock-market-data/