

Stock Assessment of Tatasteel using Time Series Analysis

Dishi Patangiya¹, Bhavya Sharma¹, Dr. Vikas Khare²
¹MBA Tech III Year Students, STME, NMIMS, Indore, INDIA
²Associate Professor, STME, NMIMS, Indore, INDIA

Abstract: Time series is a collection of continuous data points that have been ordered by date and time. The analysis of the data through this method can be utilised to understand the stock's behaviour and quantify the risk associated with it. The methods used to predict the future price of TATASTEEL stock are ARIMA (Auto regressive integrated moving average) model, python, and power BI. There were various libraries of python used, which are numpy, matplotlib, pandas, scikit learn, pmdarima. Power BI is an interactive data visualisation software aimed largely at business intelligence. . The Power BI dashboard is a story-telling one-page display. Reports are used to create dashboard displays, and each report is based on a dataset. The dataset of TATASTEEL used here is of 11 years.

Key words – Time series Analysis, Python, Power BI, Stock market

1. INTRODUCTION

A time series is a collection of discrete data points that have been arranged chronologically.

A continuous sequence of temporal data connects the data. Time series data analysis is used to derive relevant statistics and other data characteristics [1]. Before making any investments, statistical time series data calculations and data analysis can be used to gain a better understanding of the stock's behaviour and estimate the risk. Time series forecasting is a step forward in the process of learning more about what will happen in the future. It refers to the application of mathematical models to forecast future values based on past data. [3]. The Auto-Regressive Integrated Moving Average (ARIMA) model, as well as the Augmented Dickey-Fuller Test is used to determine the stationarity of time series data and to estimate future stock prices for a specified period of time.

The stock market is a market that allows people to buy and sell business equity. The Stock Index has its unique value on each Stock Exchange. [2]. The index is the average value generated by combining the prices of several stocks. This makes it simpler to see the entire stock market as well as market forecasts over time. Individuals and the economy as a whole are heavily influenced by the stock market. As a result, correctly predicting market trends can lower the risk of losing money while increasing profits.

The Indian Stock Exchange, often known as the National Stock Exchange of India (NSE), is a private company based in India. The country's first demutualized electronic exchange, this market, which is located in India's economic metropolis, was formed in 1992. The National Stock Exchange was India's first exchange to offer a contemporary, fully automated screen-based electronic trading system, allowing investors from throughout the country to trade with ease. Nifty is frequently utilised as a barometer of the Indian capital market by investors in India and throughout the world.

Kwon and Shin (1999), **Christiansen et al.** (2012), **Engle et al.** (2013), and **Bekrios et al.** (2013) all underline the relevance of economic factors, particularly economic growth, on stock market return or volatility (2016). Several research, such as **Erb, Harvey, and Viskanta** (1995) and **Hassan et al.** (2003), have looked into the impact of country risk on the stock market. Erb et al. (1995) indicate that country risk indicators, such as political, economic, and financial hazards, are essential for predicting predicted stock returns using a panel based model. Hassan et al. (2003), on the other hand, want to look at the impact of country risk on stock market volatility in the Middle East and Africa from 1984 to 1999. According to their findings, country risk characteristics are significant determinants of stock market return volatility.

2. METHODOLOGY

The specific procedure or technique used in this paper to identify, select, process, and analyse information about the stock assessment using time series are ARIMA model, various libraries of python, and Power BI. The methodologies used in this research paper are represented through figure (1).

2.1 ARIMA MODEL

Before working with non-stationary data, the Autoregressive Integrated Moving Average (ARIMA) Model converts it to stationary data. One of the most widely used models for predicting linear time series data is this one. The ARIMA model has been widely utilised in banking and economics since it was shown to be dependable, efficient, and capable of anticipating short-term share market fluctuations. The abbreviation for "autoregressive integrated moving average" (ARIMA).

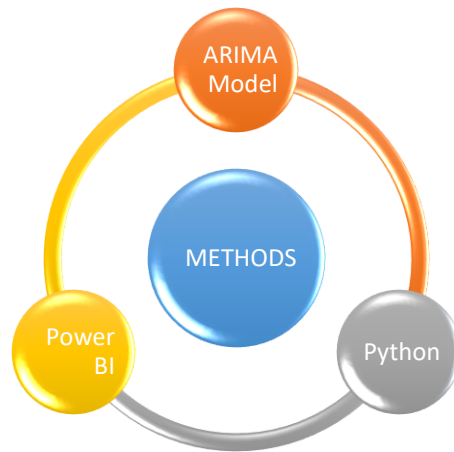


Figure 1: Methodologies

It's a time series model that's used to track events across time in statistics and econometrics. The model is used to interpret past data or predict future data in a series.

When a metric is measured at regular intervals, such as fractions of a second, daily, weekly, or monthly, it's called a periodic metric. ARIMA is a model based on the Box-Jenkins approach. Consider the case when you have a certain value A that is influenced by another value B. The link between data points A and B must then be determined in order to perform linear regression. A and B (A's previous value) have now become linked, and A's present value is now dependant on A's past value. As a result, the current value of A will decide any future value for it.

2.2 PYTHON LIBRARIES

Pandas: Pandas is an open-source library that makes working with relational or labelled data simple and intuitive. It comes with a number of data structures and methods for working with numerical and time series data. The NumPy Python library provides the foundation for this library. Pandas is fast and provides its users with a high level of performance and productivity. It includes data analysis, cleansing, exploration, and manipulation tools. Pandas allows you to analyse enormous volumes of data and come up with findings based on statistical theory. It can clean up data sets and make them readable and valuable. Relevant data is crucial in data science. Figure (2) shows all of the python libraries that were utilised in the analysis.

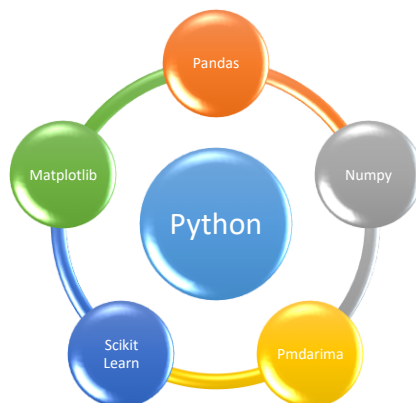


Figure (2): Python Libraries

NumPy: t's a Python library that includes a multidimensional array object, derived objects (such as masked arrays and matrices), and a number of routines for performing fast array operations such as mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation, and more.

Matplotlib: For 2D array charts, it's a useful Python visualisation package. It's a multi-platform data visualisation package built on NumPy arrays that's meant to work with the entire SciPy stack. One of the most important benefits of visualisation is that it allows us to see large amounts of data in easily understood images. There are many plots available, including line, bar, scatter, histogram, and so on. **Scikit learn:** Python's most useful and robust machine learning library is Scikitlearn (Sklearn).

Through Python's integrity interface, it provides a set of efficient machine learning and statistical modelling methods such as classification, regression, clustering, and dimensionality reduction.

This library is mostly written in Python and is based on NumPy, SciPy, and Matplotlib.

Pmdarima: Pmdarima is a statistical library that connects Python's time series analysis capabilities with other statistical libraries. A collection of statistical tests for stationarity and seasonality, B., for example, is a time series utility. Differentiation and inverse differentiation are two terms that are used interchangeably. BoxCox and Fourier transforms are just two examples of intrinsic and extrinsic transformers and functionalizers. Decomposition of seasonal time series utility for mutual verification for experimentation and examples, there is a large number of built-in time series datasets. To integrate quotes and drive production, use the Scikitlearnesque pipeline.

2.3 POWER BI

Power BI is a business intelligence-focused interactive data visualisation product from Microsoft. Data preparation, data detection, and interactive dashboards are among the data warehousing tools available. Microsoft Power BI is a tool for creating reports and gaining insights from your company's data. It links too many datasets and "cleans up" the data so that it can be processed and understood more easily. The Power BI architecture is an Azure service that allows you to connect to a variety of data sources. You can build dataset reports and data visualisations using Power BI Desktop. To access continuous data for reporting and analysis, the Power BI gateway connects to your on-premises data source. A Power BI service is a cloud-based service that enables the publication of Power BI reports and data visualisations. With the Power BI mobile app, you can view your data from anywhere. The Power BI app is available for Windows, iOS, and Android. The Power BI dashboard is a one-page presentation that tells a story. Reports are used to create dashboard displays, and each report is based on a dataset. The canvas is the name for the one-page dashboard. The visualisations that display on the dashboard are known as tiles, and the report creator pins them to the dashboard. Power Query, Power Pivot, Power View, Power Map, Power Q&A, and Power BI Desktop are the components of Power BI (Development Tool). Stream Analytics, multiple data sources, and custom visualization are some of the Power BI's features.

3. DATA

Tata Steel Limited is an Indian multinational steel-making firm based in Jamshedpur, Jharkhand. It is headquartered in Mumbai, Maharashtra, India. The corporation is owned by the Tata Group. With an annual crude steel capacity of 34 million tonnes, Tata Steel, formerly known as Tata Iron and Steel Company Limited (TISCO), is one of the world's major steel makers. It is one of the world's most geographically diverse steel producers, with operations and commercial presence all over the globe. The group generated a consolidated turnover of US\$19.7 billion in the financial year ending March 31, 2020 (excluding SEA activities). Tata Steel employs more than 80,500 people across 26 countries, with the majority of its operations in India, the Netherlands, and the United Kingdom. The company's largest factory is located in Jamshedpur, Jharkhand (10 MTPA capacity). In 2007, Tata Steel purchased Corus, a steel manufacturer based in the United Kingdom. It was ranked 486th on the Fortune Global 500 list of the world's largest companies in 2014. TATASTEEL LTD's dataset is being used for an 11-year period, from January 1, 2011 to January 1, 2022. After Steel Authority of India Ltd, it is India's second largest steel company (measured by domestic production) with an annual capacity of 13 million tonnes (SAIL). Table (1) shows the data for the stock that was used in the analysis, which includes attributes such as Date, Open, Close, High, Low, Adj. Close, and Volume. The table also shows the first 10 tuples of the dataset.

Table (1): Dataset of TATSTEEL stock

Date	Open	High	Low	Close	Adj Close	Volume
01-01-2011	652.815125	680.158691	625.138123	629.997131	485.806061	30599552
08-01-2011	628.806152	634.236755	588.31488	593.030945	457.300568	35957677
15-01-2011	587.981445	614.038757	584.218079	599.795349	462.516724	28110844
22-01-2011	599.176086	633.093506	595.460388	607.32196	468.320618	23250152
29-01-2011	591.649475	621.184265	586.695251	605.845215	467.181946	43123736
05-02-2011	609.465637	614.991516	547.918823	567.164124	437.354004	39067026
12-02-2011	578.596985	630.140015	571.165588	608.179443	468.981903	34980617
19-02-2011	611.085266	702.166931	563.257874	578.168213	445.839508	34980617
26-02-2011	584.980286	605.749939	570.975098	588.981812	454.178131	22202239
05-03-2011	584.456299	591.220703	549.871948	553.825806	427.068512	27059203

4. Analysis

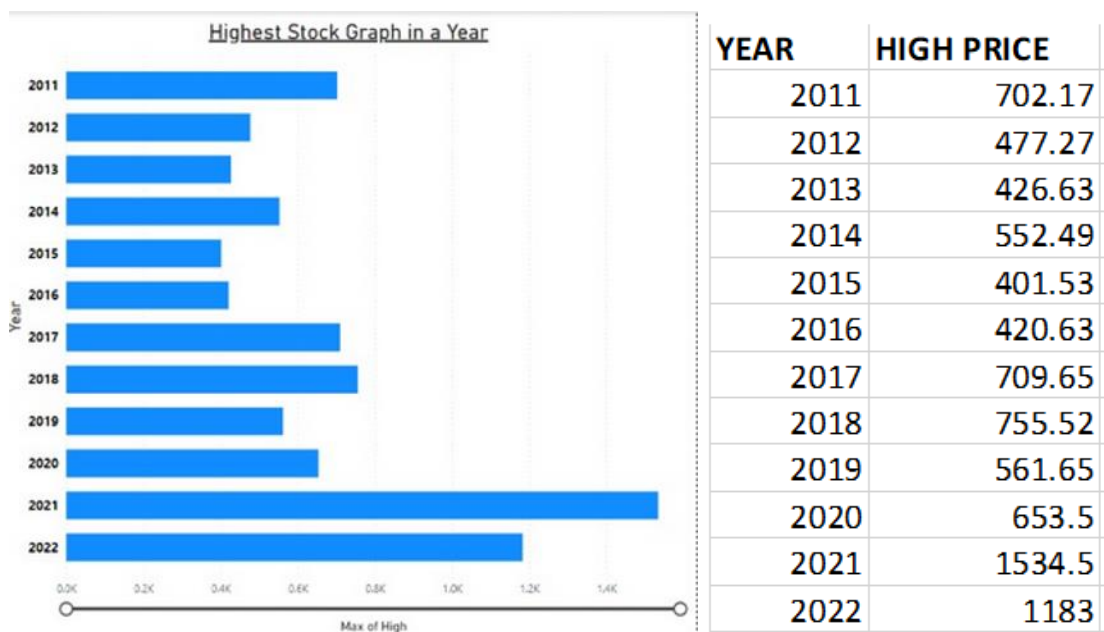


Figure (3): Highest Stock Price Graph



Figure (4): Visualization of stock's daily closing price

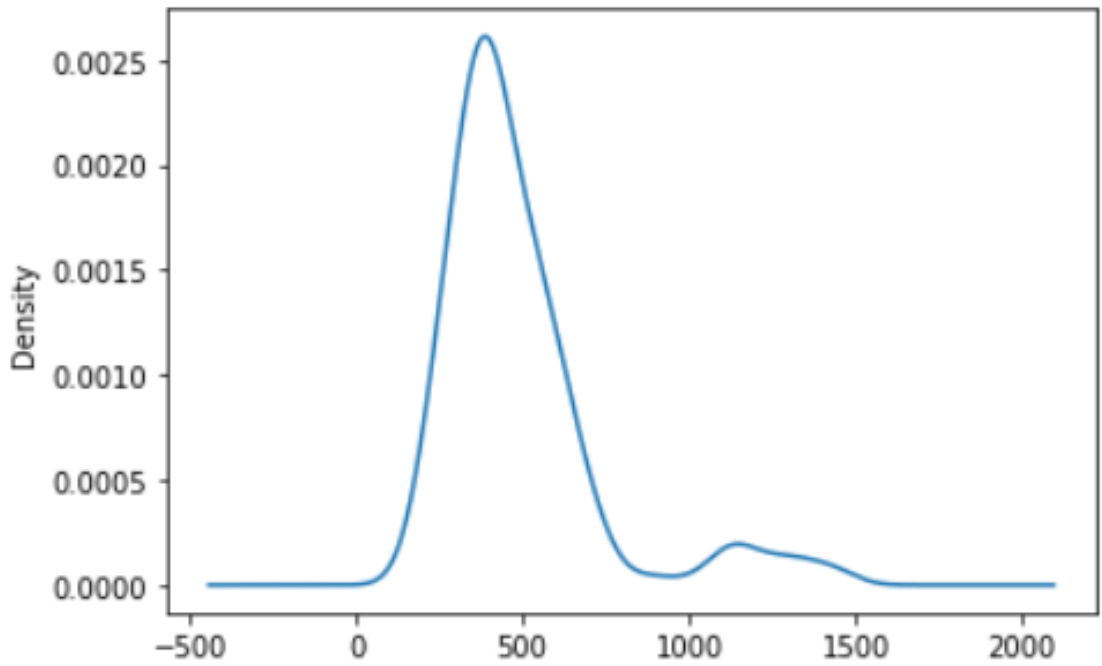
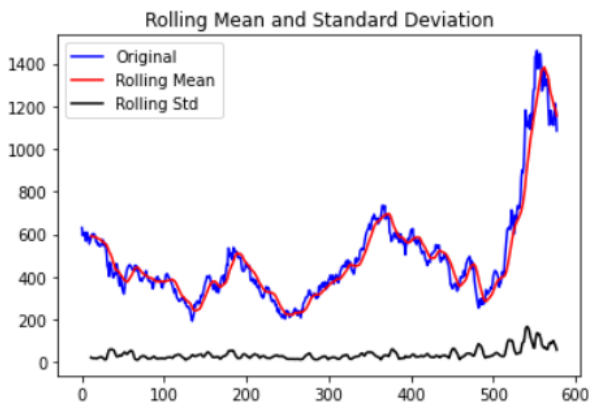


Figure (5): Probability distribution of closing price of stock



```
Results of dickey fuller test
Test Statistics          -1.746701
p-value                 0.407226
No. of lags used       15.000000
Number of observations used 562.000000
critical value (1%)    -3.442039
critical value (5%)    -2.866697
critical value (10%)   -2.569516
dtype: float64
```

Figure (6)

They applied the ADF (Augmented Dickey- Fuller) Test, which is the most extensively used statistical test, in figure (6). It's used to see if a series is stationary or has a unit root.

The following are the null and alternate hypotheses:

The series has a unit root, according to the null hypothesis.

Alternative Hypothesis: There is no unit root in the series.

When the null hypothesis is rejected, the series becomes stationary, and the mean and standard deviation become flat lines.

This graph's conclusion is that it is non-stationary.

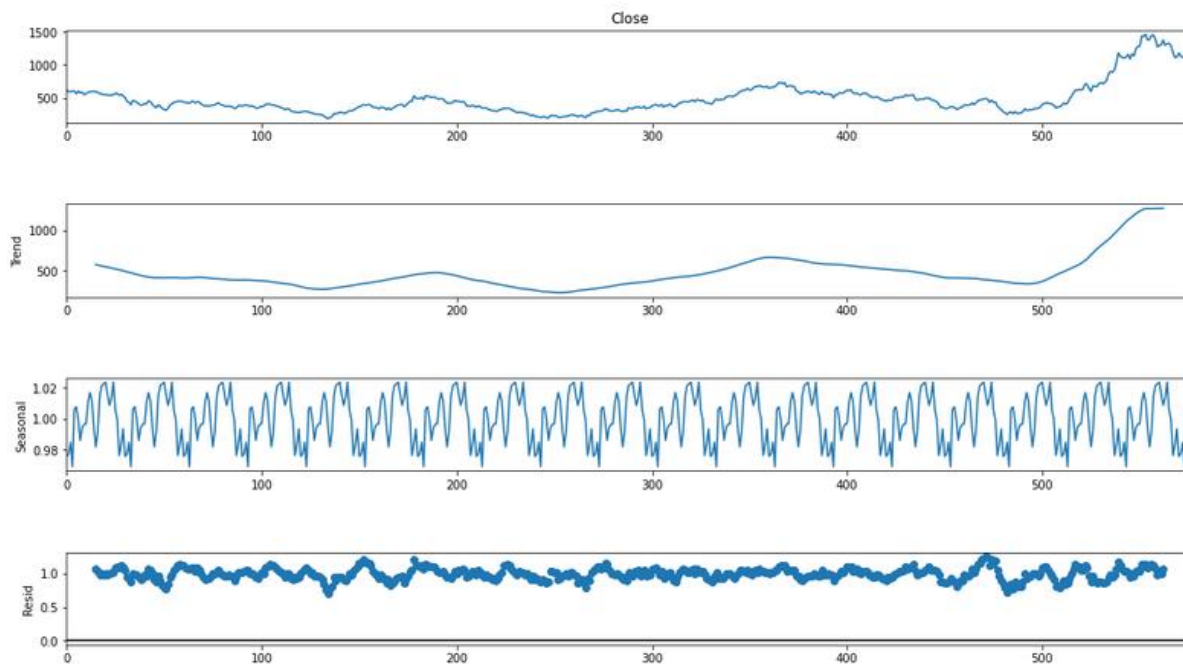


Figure (7): Separation of Trend and Seasonality

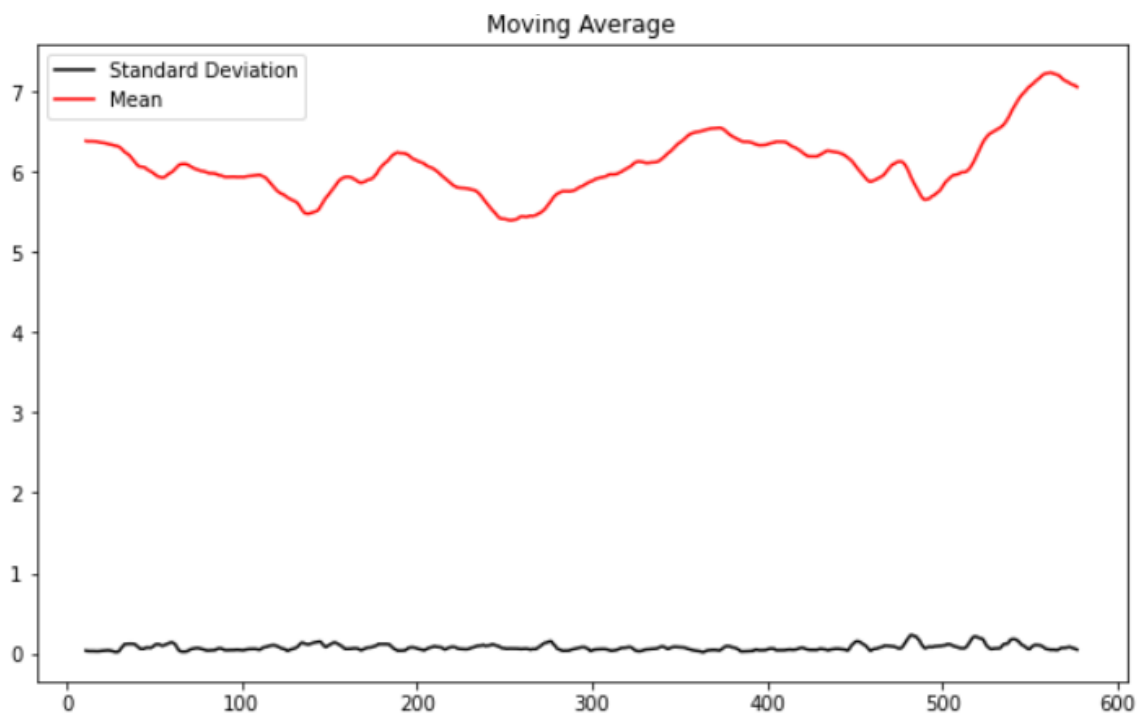


Figure (8): Log of the series and calculating rolling average

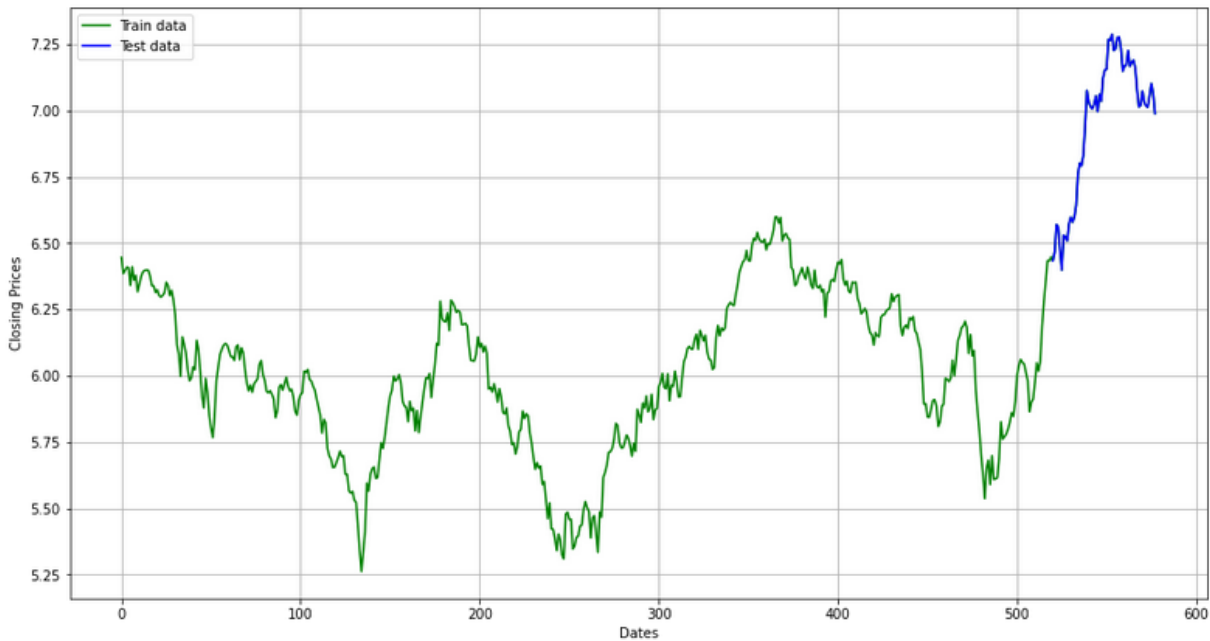


Figure (9): ARIMA model to train the data

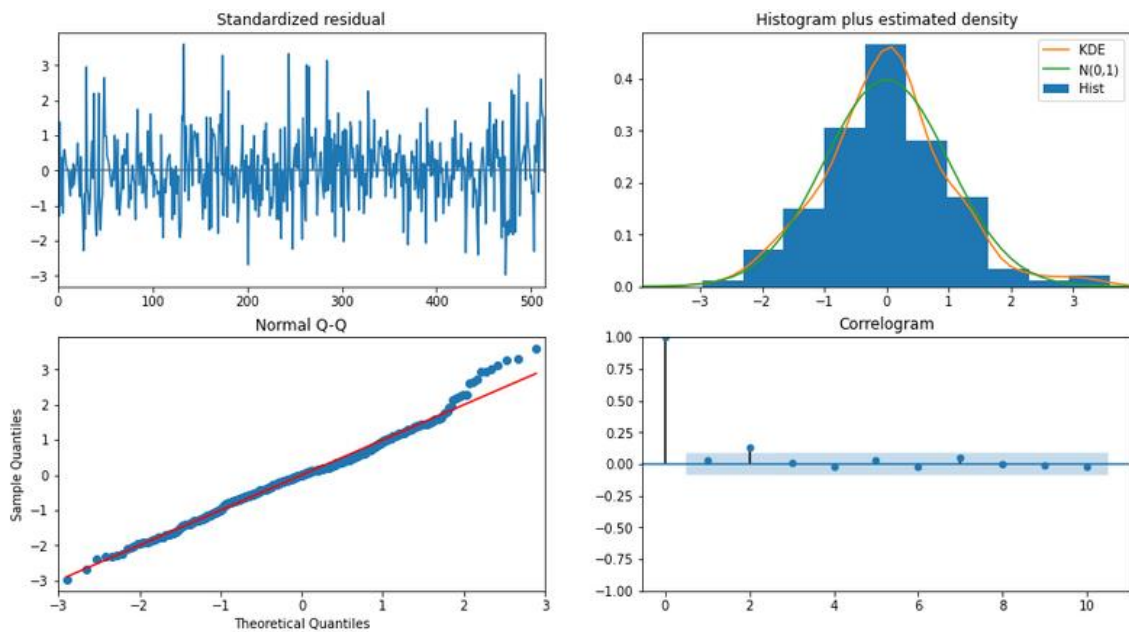


Figure (10)

The researchers choose the parameters for the ARIMA model, which are p , q , and d , in Fig. (10). This time, they used Auto ARIMA to select the parameters (Automatically discover the optimal order for an ARIMA model). The auto arima model returns the most fitting ARIMA model after calculating the optimal parameters.

The residual error appears to have a uniform variance and fluctuate around a mean of zero in the **top-left** graph in the preceding figure.

The density plot on the **top-right** graph shows a normal distribution with a mean of zero. Because the red line is not completely aligned with the dots in the **bottom – left** graph. This illustrates the data's skewness. The residual errors are not auto-correlated, as shown in the **bottom-right** graph.

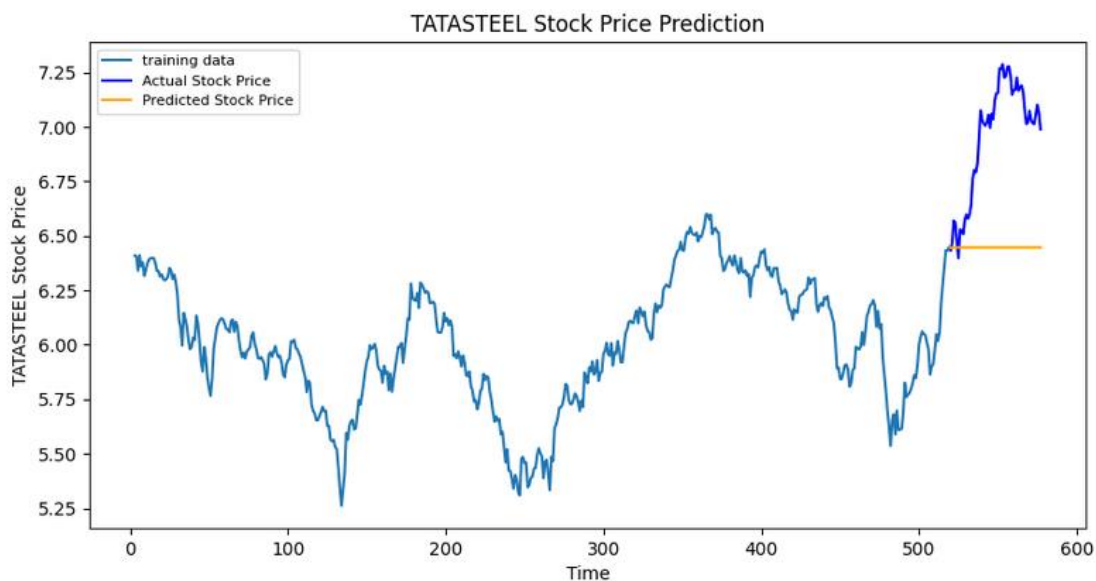


Figure (11): Forecasting with 95% confidence level

Average of Open, Average of Close, Average of High, Average of Low, Average of High and Average of Low by Year

Stock Price ● Average of High ● Average of Low

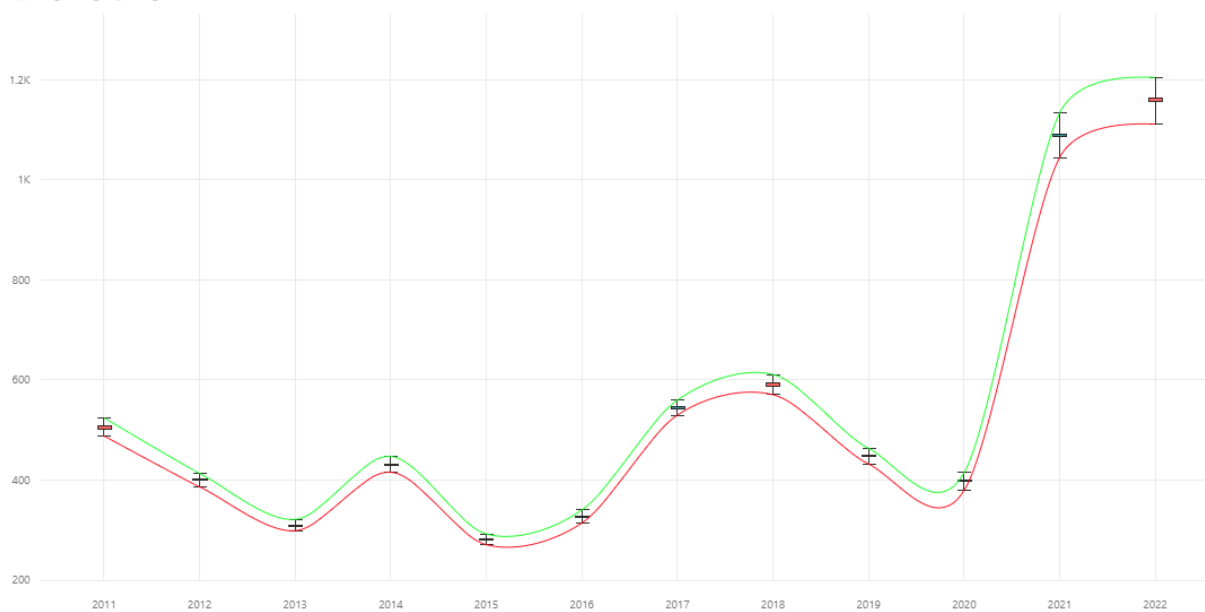


Figure (12)(a): Candlestick Graph



Figure (12) (b): Candlestick Graph

In Figure (12)(b), the researchers examined the stock price on a monthly basis using a line chart and a candlestick graph for the years 2019 and 2020.

The world was hit by COVID -19 during this time, and stock values plummeted, with the lowest price of Rs.250.85 and the highest price of Rs.653.50.

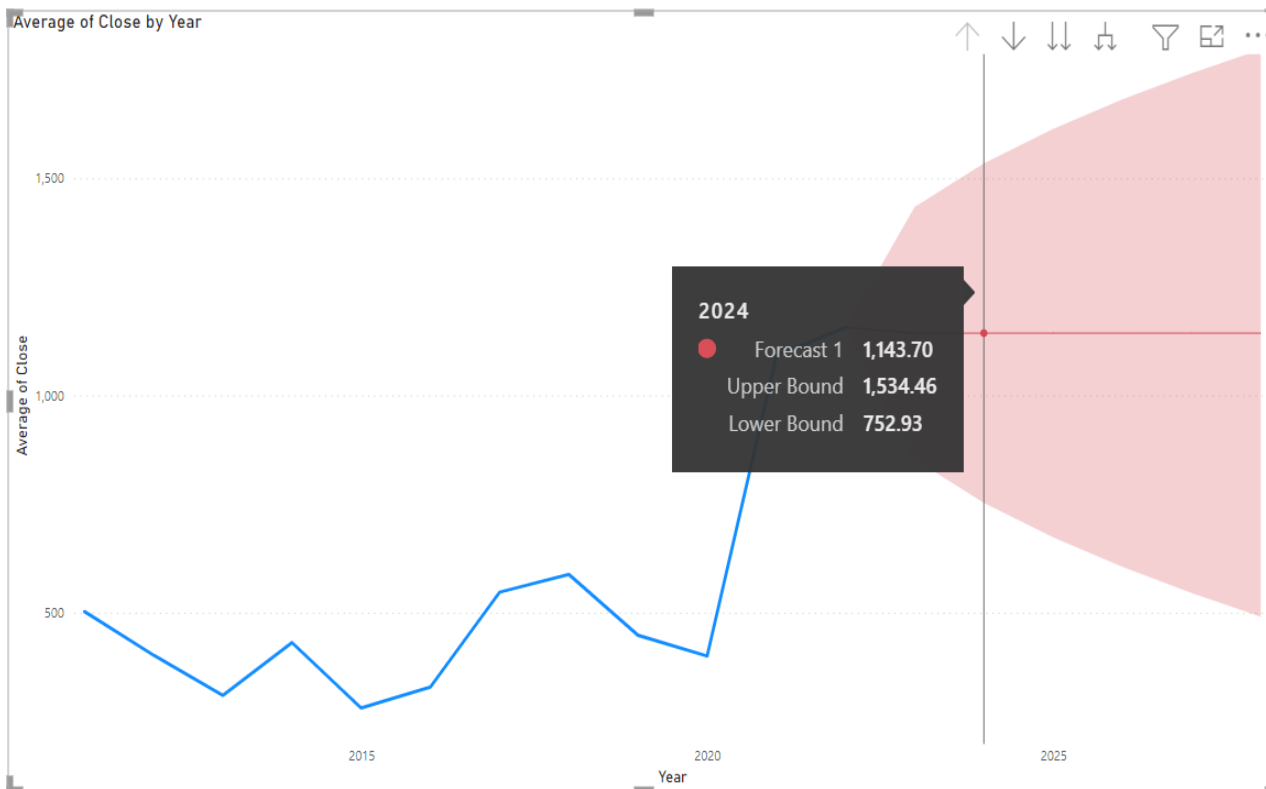


Figure (13) Prediction of stock price for next 6 years

CONCLUSION:

From the analysis, the researchers conclude that the stock has hit the lowest price during the starting of the COVID -19 pandemic and highest (until January, 2022) in the year 2021. By using power BI and time series analysis they also predicted the price of this particular stock for the next 6 years. We compared tata steel stocks for the past ten years and concluded that the stock's open price, close price. By using a power BI tool i.e., forecasting we have predicted the price for 6 years. The upper bound for 6 years are 1434.15, 1534.45, 1613.83, 1681.62, 1741.77, and 1796.41 and the lower bound is 853.24, 752.93, 673.56, 605.77, 545.62, and 490.99.

REFERENCES:

- [1] Kirikkaleli D. , 2020, The effect of domestic and foreign risks on an emerging stock market: A time series analysis, North American Journal of Economics and Finance(51)
- [2] Christy Jackson J. a, Prassanna J. a , Abdul Quadir Md. a , Sivakumar V. , 2020, Stock market analysis and prediction using time series analysis, materials today: Proceedings.
- [3] Jia Zhu, Daijun Wei, 2021, Analysis of stock market based on visibility graph and structure entropy, Physica A: Statistical Mechanics and its Applications. Volume 576