

# BSE 500 Stocks Fluctuations over 20 years

## ***Abstract***

Stock market is highly volatile and thus it is beneficial for a trader to analyse the past and current performance of a stock before making an investment. It assists in decision-making and offers a better comprehension of the stock market. In order to analyze daily price data of a stock over the previous 20 years, we plan to visualize a stock's price trend. The visualization shows a variety of patterns, each of which is a predictor of price movement and may represent the steadiness between the buying or selling pressure. In addition to the price movement, risk analysis is of utmost importance, so we also intend to visualize the risk associated with investing in a specific stock. Lastly, the stability of a stock is an important factor which influences the investment decision, so we have visualized the daily percentage change of closing price.

## ***Dataset***

We used the S&P Bombay Stock Exchange 500 Companies' stock data from the previous 20 years. First, we obtained the Security Code, Security ID, and Security Name for each BSE 500 company from [S&P BSE 500 Index - Stocks Screener | Value Research \(valueresearchonline.com\)](https://www.valueresearchonline.com). The daily stock data for each of the companies during the previous 20 years were then obtained from Alpha Vantage (<https://www.alphavantage.co>) API using this information. We obtained the historical data using this website's daily adjusted data API, which offers access to real-time financial market data.

Our stock dataset has mainly two aspects of big data. We are dealing with a real time dataset and the daily stock data is getting updated on the website for every company, thus our data has *velocity*. After combining the data for every stock over the last 20 years, our dataset has 1657280 rows which indicates *voluminous* property of data.

The dataset consists of the attributes such as the daily opening and closing prices of stock on that date, the highest and lowest price along with adjusted close, dividend amount and split coefficient. These attributes are (quantitative) ratio measurement attributes.

## ***Data Exploration, Processing, Cleaning and/or Integration***

### **Data Processing and Cleaning**

To obtain the data for every stock listed in the BSE 500, we had to execute separate API calls. So, 500 API calls in total were needed for this. Additionally, in order to obtain the data for some stocks, we had to include their unique Security Code in the API call, while Security ID was needed for the remaining stocks. We handled this by conditional statements to get the data by one of the ways since the documentation didn't specify when to send the security code and when the security id. Lastly, the API had a limit of 5 calls per minute, so we added a delay of 60 secs after every 5 API hits in the execution of code.

The data returned as a response consisted of all the attributes mentioned in the above section out of which we only stored the below attributes,

- *Opening Price* of the stock on a day
- *Closing Price* of the stock on a day
- *Highest Price* of the day
- *Lowest Price* of the day
- *Daily Volume* of the stock
- *Adjusted Closing Price* of the stock

An additional column of the *Stock Name* was appended and the individual stock data was stored in separate csv files. We loaded this data to MySQL workbench and combined the csv files of all the stocks for easy maintenance, handling and real time data updation.

Stock Name	timestamp	open	high	low	close	adjusted_close	volume
3m india ltd	2022-12-02	23399.95	23399.95	23035.2	23173.25	23173.25	269
3m india ltd	2022-12-01	23420	23420.65	23307.7	23393.9	23393.9	71
3m india ltd	2022-11-30	23347.05	23466.45	23200	23347.55	23347.55	237
3m india ltd	2022-11-29	23271.85	23460	23096.15	23381.3	23381.3	138
3m india ltd	2022-11-28	22855	23299	22800	23201.7	23201.7	200
3m india ltd	2022-11-25	22944	23042.3008	22770.4492	22884.1992	22884.1992	249
3m india ltd	2022-11-24	22836.8008	23172.6504	22779.3496	22946.0996	22946.0996	191
3m india ltd	2022-11-23	23171.3496	23212.9492	22685	22836.75	22836.75	556
3m india ltd	2022-11-22	23515.0508	23640.75	23051.0996	23171.3008	23171.3008	448
3m india ltd	2022-11-21	23905.0508	23920.0508	23500	23756.5508	23756.5508	387

## **Data Exploration**

After the completion of preprocessing, cleaning and integration we explored the data by understanding the significance of each column available in our dataset. We also did some research on the indicators of the stock market based on which the traders can interpret the market trend and make investment decisions.

While exploring the dataset we found that out of the 500 stocks listed in the BSE 500 companies, different stocks entered the stock market at different time frames and thus the data available for each stock varied between 1 to 20 years, for instance *CSB Bank LTD* had the daily stock data from January 2020 while *Gillette India LTD* had data from January 2005 and so on. Initially we were going to consider data for a couple of years however, this reduced the size of data and left us with fewer data points to visualize and thus we decided to consider the entire dataset for all 500 BSE stocks.

We decided to use the daily opening, closing and high and low prices and the adjusted closing price. We used these attributes as we had to show the daily fluctuations, trends in price of a specific stock as well as the risk and returns associated with various stocks. We plotted a few line graphs initially to understand how open and close prices are varying over time period for different stocks. The below graph was only used for understanding the patterns and is not part of the final visualization. We realised that line graphs can only highlight one feature of the stock.

Opening Prices



## Visualisation

### First Visualisation

In the first visualization we have used the traditional **Volume Candlestick** chart to show the price movement of a stock over a time period and to understand the daily volume changes according to the price movement.

#### LIFE INSURANCE CORPORATION OF INDIA



### Choice of Chart Type

The attributes open, close, high, low and volume of stock all fall under the *Ratio* type of measurement and the data we are dealing with is a *Time Series* dataset.

Usually, a line graph is an ideal choice for visualizing the changing data over a period of time, however the line graph displays only one data point (either the opening or closing price) and we wanted to display the entire day's data for a specific stock rather than plotting only one data point, so we used the candlestick instead. The Candlestick chart is extremely useful for interpreting the current and past trends where every candle provides us with the opening and closing, highest and lowest price on a day as well as the overall chart demonstrates the direction of price movement. Instead of using the traditional candlestick we used the Volume Candlestick to add the 5th parameter where the user can also understand how many shares were traded on a particular day.

## Design Choices

We referred to the following tutorial for plotting the graph —

- We used Python to plot this graph since it provided an inbuilt *Plotly* library as well as the candlestick function and had an interactive range slider which allows users to see the data only for the required time period (day, month and year). We provided an explicit dropdown option from where the user has to select the required stock for which he wants to visualize the current and past trends of that stock. This provides flexibility to the user both over the time range and the stock selection.
- We used the standard candlestick layout where the Green candle indicates the price moved up on the day and Red indicates the price went down by the end of the day. These standard colors justify the price movement very well. The volume bars have the —
- We have given the title in bold highlighting the stock for which the data is displayed and we have provided two y-axes one for price and another for volume which provides better readability.

## Second Visualisation

In the second visualization, we are trying to analyse the risk associated with the top 30 BSE stocks using Scatter Plot. We have used only the 2022 stock data since it is the most recent data which can provide a better estimate of the Risks.



### Choice of Chart Type

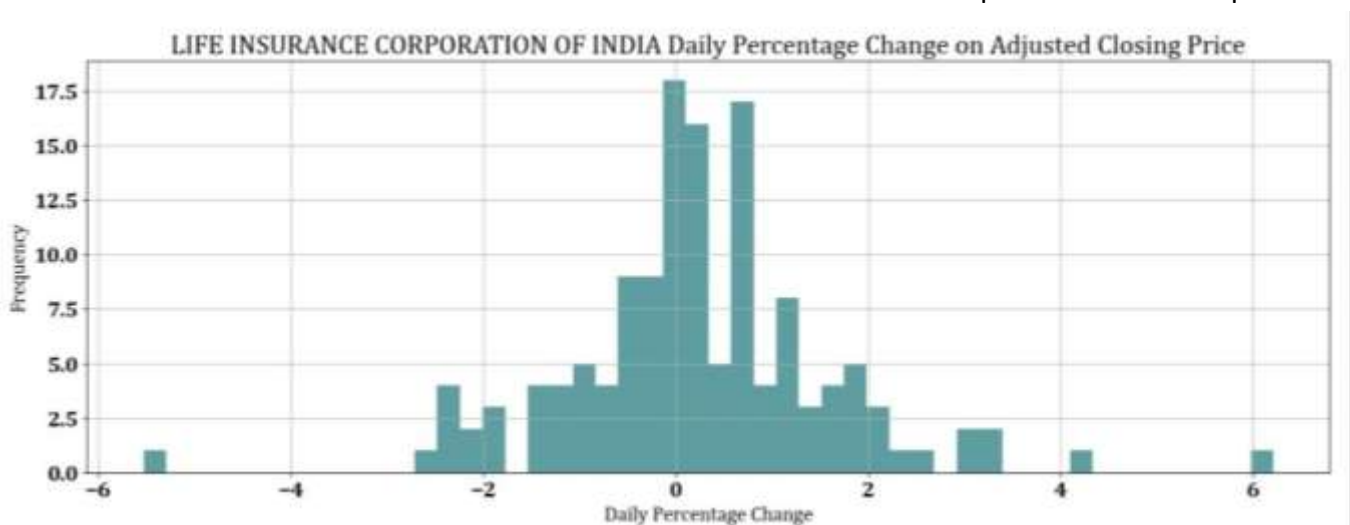
We first calculated the correlation coefficient between daily returns of different stocks and then used the mean and standard deviation of the daily returns to plot the relationship between Expected Returns and Risk. We choose the Scatter plot because it shows the relationship between two variables, also the data we are dealing with is numeric which makes the scatter plot an ideal choice.

### Design Choices

- We have used the Scatter plot from the Plotly library of python to visualize the scatter plot.
- The user can hover over any data point to understand the Risk and Expected Return of a Stock.
- Since there are only 30 data points the scatter plot is readable and allows users to analyse the Risk associated with the top 30 BSE stocks.

### Third Visualisation

In the third visualization we have used the daily adjusted closing price values to calculate the daily percentage change on the adjusted closing prices for every stock and visualized the same using Histogram, so that the user can understand how stable the stock is by looking at the distribution. The user can select the stock for which he wants to view the distribution from the list of stocks provided in the dropdown.



### Choice of Chart Type

We wanted to analyse the stability of a stock by plotting the daily percentage change over the years and we could not find a better chart than histogram which is used for displaying the distribution of a continuous variable; the adjusted closing price in this case.

### Design Choices

We have used the Matplotlib.Pyplot library of Python to plot the histogram. The Title is a combination of the stock selected by the user and the description of the graph.

## ***Conclusion***

Stock analysis dashboards have been an effective way for keen investors to study the market before making any investments. Our visualisations portray the most significant aspects that any investor would want to analyse. In the first visualisation we have mapped the stock prices with the volume traded since the stock has been listed. The range slider feature provides the investor to see the trend in the selected stock for any given date range. It would be appropriate to say that the first visualization provides a detailed trend analysis for the big data.

The second visualisation calculates the Pearson's coefficient for the BSE SENSEX (Top 30 ) companies to study the Risk associated with the stock and the returns expected from them. By looking at this chart the investor can analyse how risky the stock is and what kind of returns he can expect. The third visualisation can be more beneficial for finance experts to study the total distribution of the changes taking place on a daily basis with respect to the adjusted closing price.

Each visualisation that we have created aims at providing various angles to the selected stock for the investor to make a calculated decision.

There were few aspects that could be improved in our visualizations. For instance, in the first graph we wanted to change the color of the volume bar chart to a contrasting color for better differentiation. Also, we could have provided a better alignment to the titles of both y-axes (Volume and Stock) for making it more readable. In the third graph, we could have focused on the normal distribution and skewness to provide additional insights to the user. We were initially aiming to create a word cloud by considering the traded volume of the stocks, however we could not find any relevant sources which described how categorical data could be pictured in the word cloud. Another challenge we faced was there was a callout limit of 5 per minute to the API due to which it took us a large amount of time in data collection and integration.

## ***References***

- [1] List of BSE 500 S & P Stocks: [BSE Top 500](#)
- [2] List of BSE SENSEX Top 30 Stocks: [BSE Top 30](#)
- [3] Alphavantage API : [API](#)
- [4] Making of Candle Stick Graph : [Candlestick Chart in Python](#)
- [5] Pearson's Coefficient Scatter Plot Reference: [Scatter Plot in Python](#)
- [6] Histogram for Continuous Data: [Histogram in Python](#)