Section 1: Idea/Goal

One of the most common areas of study when it comes to building any sort of predictive model are the metrics surrounding money and the value of a product. In this case I would like to look the possibility of creating a model that provide some insight in the closing value of a stock. There is some difficulty in attaining this type of data from the American stock market in an easily accessible format. The Indian stock market, on the other hand, has readily available datasets surrounding the opening and closing prices (among other metrics) for a large portion of the companies listed on the Nifty 500. While this model is not necessarily possible and or accurate I would like to see if there is even a remote possibly of predicting volatility between the day to day value of a stock. If there is validity to this model I think it would provide an interesting look at whether or not there is any data based rhyme or reason for the high volatility in some stocks or the lack there of in others. On the other side of the coin, I think that a lack of evidence to support the validity of the model may indicate that stocks are simply beholden to the individual traders personal opinion of the value more so than the actual tangible value of the stock, which could also provide some insight in to the volatility of said stocks. Overall I am curious to find these results either way.

I believe that this problem is a focal point for many investors in not only foreign markets but also the US market. Predicting volatility to some degree would be an extremely helpful tool for any major stakeholder in the market. This goes double for stakeholders interested in foreign markets, as they are traditionally a bit more unstable. A pitch for this idea would likely try to reinforce the idea that any sort of advantage gained from a model like this would provide quite a boost over the competition. Even if share prices are not moving dramatically being able to predict some level of changes would have a noticeable affect at high volumes. This type of volume would likely appeal to the type of stakeholder is interested in buying into a foreign market.

This data was obtained via Kaggle, I am unsure of how the user in question obtained this data but it is likely either obtained from the market itself or from some kind of third party tracker. The third party tool is the most likely choice as there are a number of different websites that provide live tracking and tickers for the Nifty 500 much like the markets here in the US. I am unsure if there is an official site provided by the market, but the sites providing these tickers in English seem to all be third party.

Section 2: Analysis/ Process

In the case of our study here, the most likely stakeholders would be American based investors looking to start purchasing stocks in a foreign market, in this case India. The Nifty 500 is one of the largest stock exchanges in the country and India’s growing economy could open opportunities for foreign investments to make large returns. In this study one of the main barriers to entry for understanding what we are viewing is the large discrepancy between the value of the Indian Rupee and the USD. It is important to convert these prices into USD in order to best convey to our stakeholders what the value of a stock/company is in terms they are more likely to understand. This is where the use of different visualization tools such as graphs comes in handy. Simply converting prices is not enough, providing visuals to accompany these changes and the present data that has not be altered is key. I accomplished these visualizations through the use of the Matplotlib library by comparing the different columns of our pandas dataframe containing our selected columns.

These graphs give a bit of insight into some of the valuations of these companies and how they are fluctuating in the market. For the most part the seems to be very few significant changes in the overall values of the stocks aside from a few outliers. These findings segway into the findings of our model.

Chart, scatter chart

Description automatically generated

Much of the data provided in the initial set was not overly relevant or acitvly hindered out ability to work with and process the information. Multiple null values or values that did not pair well with our model were present in a number of columns. This highlighted the need for a bit of necessary data wrangling and cleansing. While this may seem like a time consuming and unnecessary task but outside stakeholders, making sure that our data is useable is key. Without these changes we would likely not be able to see the full scope of the data or it could even provide information that is inaccurate or otherwise unhelpful.

Section 3: Model

With this type of information choosing a model can be a bit difficult. Many of the more commonly used models do not work particularly well with this type of data or do not provide information that is actionable. In this case I chose to use a linear regression model as it seemed the most appropriate for the dataset. Almost of the bat the information provided by the model, after slotting in our training and test sets, was extremely accurate. This type of accuracy (over 99%) is very misleading. At first glance, most stakeholders would be very excited to see this level of precision but it actually indicates that the information provided to the model is not very helpful. As pointed out to me previously, the level of change between opening and closing prices is often minimal. While in the context of the stock market any of these changes could indicate a sign of volatility and provide large profits at huge volumes, it doesn’t really reflect the spirit of these tests. Outside of significant changes within the pricing of the stock, trying to predict the closing price of a stock will often just reflect the opening price. This indicates that our predictors are not providing a group of data that will allows to make actionable decisions.

This would indicate to me that model of this type would not be particularly useful to our stakeholders. I am unsure if there is a model that we could use on this data alone to attain the desired outcome. This would mean that there would not be a point in time where this model could be considered to be in a state that is ready for deployment on live data. It would require the use of data outside of our selected dataset in order to get anywhere near a deployment stage let alone provide actionable information to our stakeholders.s

Section 4: Conclusions

Based on the finding from the model I believe it is safe to say that within the context of pure market data, it would be very difficult to make decision based on that data alone. Pure market data by its nature is unable to give us the full story. There are simply too many business decisions and irrational behaviors surrounding the market to work with this data alone. Much of the actions that do create dramatic changes in market behavior are often not quantifiable and even if they were it is very unlikely that a group would have access to enough of that data to build models that could predict these changes. In that same vein of thinking large market changes are often the affects of global changes and conflicts. Even within the context of the American stock market, there are too many outside factors beyond the scope of those participating in the market to take any actionable data with a high degree of confidence.

Even if all of the previously mentioned issues were addressed and solved I think this type of project could raise some ethical concerns as well. Having a model that could solve the market, so to speak, would give anyone with the model an enormous advantage of every other competitor. This would likely lead to a massive change in the dynamics of the free market or if the model was provided to anyone outside of that group it would grind the market itself to a halt. It is for these reasons that I would recommend to any stakeholders involved that the cost and possibility of success are not viable as a strategy. It would likely be unsuccessful to pursue this goal and cost an astronomical amount to solve in the first place.