**Project Proposal**

**Motivation**

Predicting stock prices is one of the popular old business problems which still does not have a definite solution. While several methodologies were developed using technology with time, which are used by various industries to invest in stocks, it does not always work well. To understand the trends in trading stocks in a better manner, apart from performing traditional time-series analysis, we also want to utilize this data to the maximum as it contains an extensive range of features to play around and make meaningful interpretations and how exactly they influence U.S. stock market.

**Introduction**

The data utilized for the project is “Stocks IPO information & results” which contains information about the U.S. stocks of various companies that were publicly traded on 01/01/2018. The data mainly revolves around how the companies appeared before going public. Data will be utilized to study the stock market trends for various companies by performing time-series analysis and then utilizing the results obtained to identify the features that contribute to the rise and fall in stock prices in the U.S. market.

**Project scope**

We will perform the analysis in the scope of time series prediction combining the historical data and the related marketing information to understand the behavior of the IPO stock market. The selected stock market indicators will be generated based on past trading information, and then are used to forecast the short-term trend with other features.

Python will be the primary programming language we utilize for this project. We will use Tableau for interactive visualization.

**Data source**

The dataset is obtained from Kaggle.

<https://www.kaggle.com/proselotis/financial-ipo-data>

**Description of the data set**

The data contains the historical prices and the related information about the U.S. stocks of 3762 companies. The historical prices consist of the open value, close value, lowest value, highest value, and volume for 261 days, and the related information includes the stock symbol, industry, revenue, and so on. Some of the essential features are listed as follows.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Symbol | Stock symbol |
| DayBetterThanSP | Days the stock had a higher percentage change than the S&P500 |
| daysProfit | Days the stock was positive |
| daysProfitGrouped | A grouped form of positive days |
| Year | Year the stock went public |
| Month | Month the stock went public |
| Day | Day the stock went public |
| closeDay (X) | The value of the stock on close of the Xth day |
| highDay(X) | The highest value of the stock on the Xth day |
| openDay(X) | The value of the stock on the opening of the Xth day |
| lowDay(X) | The lowest value of the stock on the Xth day |
| volumnDay | The total amount of trades on the Xth day |
| Sector | The sector of the stock |
| Industry | The industry of the company |
| CEONname | The name of the company’s CEO |
| CEOAge | The age of the company’s CEO |
| Revenue | Revenue the company obtains |
| Employees | The number of the employees of the company |
| Name | The name of the company |

Due to the massive amount of the variables, we do not list all the variables; however, the list should provide a general view of the structure and the nature of the dataset.