Method for Data Gathering:

There are lots of APIs online, for example, Tushare, is a company that provides API for daily stock market prices. It has provided databases on the stock market around the world. I have already written codes to incorporate the dataset into my code. Once I have provided it with the token number of the stock, it will generate information such as open price, highest price, lowest price, low price, the volume of transaction, etc. Another platform called JoinQuant is a Chinese company that provides APIs for companies and individuals to design their own APIs for quantitive tradings, it also provides crash courses and open source codes.

Method for Data Processing:

Most of the APIs have provided ways to generate data. They are usually pre-written functions in the packages. All I need to do is to incorporate it into my code and perform data cleaning. Finding a seasonal frequency can be a hard job in time series analysis.

Methods for Analytics and Modeling:

1. Differencing: In the differencing approach, I used first-order differencing because there is no obvious seasonality (have not thoroughly analyzed the seasonality yet). Also, to mitigate the increasing variance, the first-order difference is taken after applying a log VST. This model can be expressed in Equation as $Yt = Yt-1 + Xt$, where Xt is the additive noise term.
2. Smoothing with ARIMA: ARIMA stands for autoregressive integrated moving average model, which is a generalization of autoregressive moving average (ARMA) model that can be applied in the differencing approach. ARIMA was good with no non-stationarity data, and an initial differencing step can be applied multiple times for modeling the upward or downward trends.

Is there an obvious outcome/classification/model output that you can describe:

The obvious outcome can be the forecasted values for stock price for the next 10 trading days. I can create training data on certain time periods and validation data in the next few days, or I can choose a longer range depending on the seasonality that will be analyzed. The quantitive

evaluations can be the loss, distance between the predicted value and actual data; which can be modeled as l1 or l2 norm.

How will Methods be documented:

The method will be documented through coding. I will use both R and python in data analysis. I have asked permission to audit stat 153 (times series) this semester, and they use R to perform time series analysis, such as stock prices and COVID cases changes. I will also use python, since in the real world, most projects are done by python, and lots of APIs are designed for Python programmers. I will also generate R markdowns for documenting the codes and methods I have employed and written.

References to Literature in techniques or methods:

I am reading stat 153's textbook, Time Series Analysis, and Its Applications: With R Examples by professor Robert H.Shymway and professor David S.Stoffer. I am also watching some algorithmic trading crash courses.

What is new about your research:

The most important new feature in my research is that I am considering companies that are listed in different markets. Specifically, I am analyzing the company Tencent first. In addition, I will look at the seasonality of those companies using both time series analysis, and machine learning modelings.

What is the contribution - to the field, to policy, to research:

The contribution is for the private equity companies and investment funds companies to better understand the market. It also helps public policy to see the correlation between different country's domestic policy and foreign policy's effect on each other countries' stock market.

Are there biases or blind spots:

There are lots of blind spots. Since companies are different if they have they are registered in different areas, we cannot assume both branches are the same. In addition, there are lots of companies that are duel-listed; I need to traverse lots of companies to give a statistically significant conclusion. What's more, analyzing markets only in China, Hong Kong, and the United States can be a biased sample, I need to be very careful on drawing the association, correlation, causal inference, and conclusion.