**Project Outline**

# Name of Project

Event-Driven Stock Trend Analysis Under COVID influence

# Team

**Team Name: Data Secrets**

Yangyin Ke: [yangyin\_ke@brown.edu](mailto:yangyin_ke@brown.edu)

Huaqi Nie: [huaqi\_nie@brown.edu](mailto:huaqi_nie@brown.edu)

Enmin Zhou: enmin\_zhou@brown.edu

# Location

https://github.com/Enmin/final-project.git

# Summary

The “big idea” is to analyze the potential relationships that drive the change in the stock market under COVID influences.

The project is constituted of 4 parts:

* Collecting stock price information from open source websites
* Collecting COVID data from open source websites
* Collecting events/news from social media (COVID related events especially)
* Make data analysis on the relationship between the content of events and stock price

We might be able to find the relationship between the change of stock prices and the events. At the end of the project, we will add predictions on the stock price based on COVID data.

# Data

DataSets:

1. 8-months Stock Market Information.

2. 8-months COVID data.

3. 8-months COVID events/news.

Collecting Method:

The initial dataset will be collected from the Stock Market historical data and COVID historical data.

The incremental updates will be collected from open-source websites from which we collect historical data.

# Previous Work

# [Afees A. Salisu](https://www.ncbi.nlm.nih.gov/pubmed/?term=Salisu%20AA%5BAuthor%5D) and [Xuan Vinh Vo](https://www.ncbi.nlm.nih.gov/pubmed/?term=Vo%20XV%5BAuthor%5D), Predicting stock returns in the presence of COVID-19 pandemic: The role of health news , **2020 Oct**, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7322492/>

#### [Andrea Ialenti](https://medium.com/@andrea.ialenti?source=post_page-----b7a4739406e8--------------------------------), The Coronavirus Effect on the Stock Market, **Mar 2020**, https://towardsdatascience.com/the-coronavirus-effect-on-the-stock-market-b7a4739406e8

# [Asim Kumer Dey](https://arxiv.org/search/q-fin?searchtype=author&query=Dey%2C+A+K), [Toufiqul Haq](https://arxiv.org/search/q-fin?searchtype=author&query=Haq%2C+T), [Kumer Das](https://arxiv.org/search/q-fin?searchtype=author&query=Das%2C+K), [Irina Panovska](https://arxiv.org/search/q-fin?searchtype=author&query=Panovska%2C+I), Quantifying the impact of COVID-19 on the US stock market: An analysis from multi-source information, **Aug 2020**, https://arxiv.org/abs/2008.10885

# Methodology

We planned to process the stock data so that only stock data that have been largely impacted by COVID are saved in our database. The text content of COVID events/news will be analyzed through some NLP packages such as “word2vec” or “bert”.

Machining learning algorithms are required to analyze our datasets. (we plan to use sklearn)

The stock price and COVID data will be visualized through charts; while hot-spotted events or news will be ranked in a sidebar on the website.

The base-line model will be an accurate correlated representation of COVID data and stock price data in the past 8 months.

# Visualization idea

The New York Oil and Gas model will be a good template for our final project (https://dash-gallery.plotly.host/dash-oil-and-gas/)

# Enhancement idea

1. We will test our predicting results by separating our 8-months datasets into different subsets by time.
2. The incremental updates will be used to update the model.

# Next Steps

1. Collect data: Each member will be responsible for collecting one of the target datasets (“stock”, “COVID”, “news”) and designing one or more tables in MongoDB to store the collected data.
2. Each of us will be assigned several machine learning algorithms to improve the model performance
3. Huaqi will research plot.ly and Dash for use with Python. Things about Dash and the final website will be discussed before making a final decision.
4. Enmin will research the backend system including MongoDB administration and AWS management.
5. Yangyin will research Dash and look for interesting examples to pattern final website