Project Title: COVID 19 Pandemic Analysis – ETL project

Team Members: John Costa, Michael Cary, Henry Randall

Project Description/Outline: Combine COVID-19 case/death data with data indicating time of implementation and expiration of stay-at-home data

ETL process:

1. Extract – take Hopkins data and FINRA data; scrape FINRA data
2. Transform – linking data, cleaning data, matching based states
3. Load - into SQL database (daily case count per state, daily death count per state, stay-at-home implementation and expiration dates)

Datasets to Be Used:

1) Johns Hopkins COVID-19 github repo - https://github.com/CSSEGISandData/COVID-19/blob/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_confirmed\_US.csv

2) Financial Industry Regulatory Authority (FINRA)

https://www.finra.org/rules-guidance/key-topics/covid-19/shelter-in-place

Rough Breakdown of Tasks:

* Extract (Henry)
  + Loading Hopkins data (in python)
  + Clean Hopkins data (in python)
  + Scrape FINRA data (in python)
  + Clean FINRA data (in python)
* Transform (Michael)
  + Export all tables to SQL (in python using SQLAlchemy)
* Load (John)
  + Create schema in SQL – (in postgres)
  + Develop diagram showing relationship between data – (in postgres)
  + Develop code to display tables – (in postgres)