

Data Mining Project Proposal

1 - Group: Ryan Rigby, Scott Gale

2 - Resources: We are going to use historical data from the US stock market. We have multiple sources for the data. There is a Yahoo finance API available for Python where we can get current and historical data for stocks . There is a comprehensive dataset on Kaggle that contains historical data from stocks traded on the NYSE. Lastly, we have access to TradeBuild, an interactive tool that provides data, charts, and real-time stock market profit and loss information via an API provided by the online brokerage Interactive Brokers. Each dataset provides data aggregated in different time intervals, ranging from time scales of 1 second to 1 day.

Python API: <https://pypi.org/project/yahoo-finance/>

Kaggle: <https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs/data#>

TradeBuild: <https://github.com/rlktradewright/tradebuild-platform>

3 - What: We want to analyze how stocks perform within a given market sector and market capital size within a presidential election year. Do certain sectors perform better / worse in an election year? Do companies with larger or smaller market capitalizations perform better / worse in an election year? Is there any correlation between the winning party in a presidential election and US stock market performance? Does market performance correlate with whether the current president is seeking re-election.

4 - Why: The state of the US economy affects US citizens and is a major factor in US politics. Many US citizens have money invested in the US stock market (investments and retirement) and might be swayed in who they vote for for president depending on how their investments are doing (high returns or otherwise). Also, more detailed studies such as the one we are proposing are generally part of proprietary trading models created by investment banks, brokerages, and private trading groups, and are generally not accessible to the public.

5 - Learning: Although the economy is a major topic in US presidential elections, we have never seen an analysis of market segments and market capital size studied within a presidential election year and correlated to the party of the elected candidate, so in that context we may obtain novel data. The results of this study may provide useful insights that we can use personally in our investment decisions. Market data is highly structured and labeled in several ways, so it will be readily adaptable to the techniques we will be studying in class.