

CMSC12200 Project Proposal

2 February 2022

Group Name: CapitolHillBets

Group Members: Zach Mariani, Gabriel Moos, Aarsh Kak, and Marcus Gerbich Pais

Introduction: There has recently been much controversy surrounding the personal trading activities of individual members of congress and their immediate partners. Our group aims to uncover whether or not individual members of congress outperform major indices, and how their performance compares between chambers, parties, and congressional districts. We anticipate that members of congress will significantly outperform the market.

Data Sources: We will need congressional trading data, stock price movement data, and GeoPandas geographical boundary data. We could attempt to locate and scrape individual filings, however, housestockwatcher.com has already located the past 18 months of data and made it readily available for scraping on its website [here](#). We will take advantage of the fact that the data source updates these filing so that our program will dynamically run the analysis on the most recent congress people's holdings. For stock price movement data, we could use the GoogleFinance or YahooFinance publicly available APIs, or scrape data directly from the YahooFinance website. For GeoPandas geographical boundary data, we will use the publicly available geographic data for congressional districts found [here](#).

Workflow:

- We will begin by gathering and cleaning all relevant data sets
- From here, we will attempt to generate individual index values for each congress person, i.e. a PnL over time for their individual trading
 - For each purchase of a stock, we will associate it with the congressperson. If they later close their position, we will record the profit/loss from the trade. Otherwise, we can mark their position to the current market value.
- From here, we will plot on a congressional map how each individual does compared to a given index

- Finally, we will group congress people into their own indices so that we might have RSenate, RHouse, DSenate, DHouse indices and graph their performance against each other and the major indices

Anticipated Challenges: Congressional trading filings only present ranges for amounts traded and they do not indicate initial holdings at the start of the period. In this way, we will have to make quite substantial assumptions about the actual amount that members are trading and their PnL over time given that we don't always know the cost basis of their holdings. The HouseStockWatcher website also only has 18 months of data and should it be of interest to us to include more than this, we will have to locate and scrape individual PDF filings of trades. This will be time intensive and subject to error.

Timeline for Completion:

- Fri Feb 11: Retrieve and clean all relevant data
- Wed Feb 23: Finish program that generates and stores PnL for each congressperson
- Fri Mar 4: Create code to generate output tables and visualizations that compare the congress people to major indices, how their performance compares between chambers, parties, and congressional districts
- Fri Mar 11 4:30 pm: Clean up, add description on how to run software, and submit