**Comparative Analysis on U.S Senator’s trade and S&P 500**

**#Team 40 | Summer 2022**

**#Team members: Yulia Kotelnikova; Muskan Yadav; Wenxin Li; Xingyu Lan; Yunhui Zhao;** [**Cici Yin**](https://app.slack.com/team/U03LK6NRGP4)

* **Problem**

In 2012, the STOCK Act was introduced to prevent the U.S. Congressmen from using non-public information to gain profit in the stock market. While all Congress members are required to report the transactions, previous data showed that they realized cumulative profits were much higher than average. Moreover, during the 2020 congressional insider trading scandal, four senators were investigated due to suspicious trading behaviors related to public health and the pandemic. This leads to the question about whether the high profits of congress members came from insider information. Through the studies on trading history in recent years, we intend to analyze their portfolios and determine whether they could be replicated or followed to achieve excess returns.

* **Proposal**

This project intends to investigate the trading behaviors of congress members in terms of profits against benchmark, primary industry of investment, preferences of investment among groups of congress members, and the influence of the pandemic.

Firstly, the cumulative profits of congress members are benchmarked against stock index S&P 500 as reference to average return. Secondly, tradings among different congress members are grouped to find possible shared investment interests or patterns. Thirdly, data pre and post pandemic are compared to investigate the performances during black swan effect. Lastly, trading will be replicated after disclosure to see whether common investors could follow their investment choices to gain excess profits. From analyzing the four aspects, this project intends to explain whether unexpected high profits for congress members came from insider information.

* **Feasibility**

The team uses publicly available data on the official United States financial disclosures website and the Senator trading. Python is relied upon to achieve the proposal. The team members also have backgrounds in mathematics, finance, business analytics and engineering to facilitate the project further through their knowledge.

* **Impact**

It is seen that if new data would be positive for stocks, patterns of stocks were observed to be rising before that information had become publicly available. Additionally at least 57 lawmakers [have failed to disclose financial transactions](https://www.businessinsider.com/congress-stock-act-violations-senate-house-trading-2021-9) since 2012 as required by law. ​​A 2020 study showed that [only about 15% of insider trading](https://dx.doi.org/10.2139/ssrn.3764192) in the U.S. was detected and prosecuted. An attempt is made to utilize the strict procedures governing the release of economic data and try to interpret whether congressmen use insider information to gain profit in the stock market. Hopefully this will help to alleviate the effects of insider trading on market liquidity, companies issuing stocks and bonds and the threatened [public confidence](http://dx.doi.org/10.2139/ssrn.3645579) in financial markets.

* **Usability**

The final product is delivered in the form of a presentation as well as a data dashboard. This will describe our methods and present our results. The results will be accessible to the general audience, and the methods will hopefully be appreciated by those familiar with basic statistics.

The results are nothing more than statistical findings and can not be used to indict any particular person.

* **Data**

**List of Included Datasets:**

* + Senator trading data in csv format [here](https://senatestockwatcher.com/) and from official United States financial disclosures website [here](https://efdsearch.senate.gov/search/)
  + S&P 500 (Yahoo finance) [here](https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC)
  + Data size: from Jan. 2017 to Jul. 2022

**Data Cleaning Methods:**

* + Remove duplicates
  + Remove irrelevant data
  + Fix errors
  + Restoring missing values
  + Handling missing values

**Advantages and Disadvantages of the dataset:**

* + There are 6367 senator’s stock transactions in the datasets from Jan. 2017 to Jul. 2022. The datasets consist of 11 features including Transaction\_Date, Owner (Spouse/Joint/Self), Ticker, Asset\_description (Stock; Securities; Options), Asset\_type, Type (Sale or Purchase), Amount (how much money they spent on investments); Senator Name; Disclosure Date.
  + The advantage of this dataset is that abundant information is provided to understand the senator’s transaction behavior. The disadvantage is that the data sample is relatively small and contains missing values.
* **Methods**
  + Visualization: Python: matplotlib, seaborn
  + Libraries: Python: pandas, numpy, scipy, datetime
* **Timeline**

| **Date** | **Deliverable** | **Details** |
| --- | --- | --- |
| **Week1 (Jul 11- Jul 16)** | **Project Scoping** | Clarifying the problem and writing the proposal. |
| **Week2 (Jul 18 - Jul 23)** | **Data Cleaning;**  **Data Wrangling; EDA** | Datafolio draft due |
| **Week3 (Jul 25 - Jul 30)** | **Datafolio due; Dashboards Building** | Presentation due |
| **Week4 (Aug 1 - Aug 6)** | **Project Symposium** | Delivering presentation |

* **Concerns**

Our primary concerns with the project will be the method to determine whether congressmen use insider information to gain illegal profit. With the fact that July, 17th, 2022, the US inflation reached 9.1% which was the highest rate since 1970. The wealthy people should earn more due to the big volatility in the market. To address this, we may narrow down the effect of the macro factors.

* **References**

1. What's insider trading and why it’s a big problem. (2022). Retrieved 16 July 2022, from <https://theconversation.com/whats-insider-trading-and-why-its-a-big-problem-176940>
2. The project was inspired by [this reddit post](https://www.reddit.com/r/wallstreetbets/comments/nanynr/i_analyzed_9000_trades_made_%5B%E2%80%A6%5Dcongress/?utm_source=share&utm_medium=ios_app&utm_name=iossmf).