Topic and purpose:

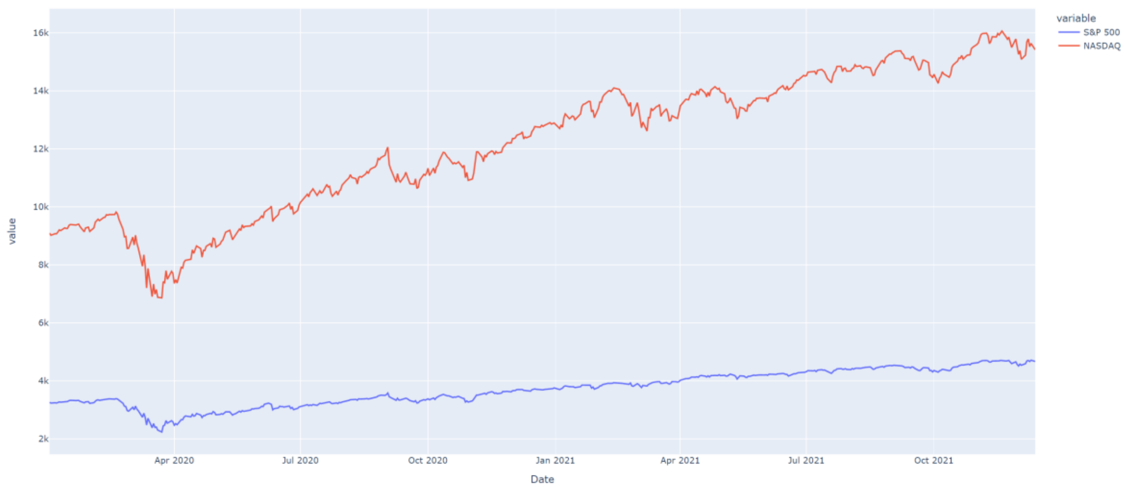
The topic of the project is Congress Financial Disclosures Sentiment Analysis. Inspired by the new that “Four senators sold stocks before coronavirus threat crashed market” (https://thehill.com/homenews/senate/488593-four-senators-sold-stocks-before-coronavirus-threat-crashed-market), I would like to develop a tool to extract useful information from trades made by senators and representatives, in order to improve the data transparency into Congress and make insider trading less likely. It’s crucially important because it helps the public to understand if and how politicians are benefited from the information they know before the public, which is the purpose of financial disclosures.

Process and methods:

I took advantages of the senatestockwatcher (https://senatestockwatcher.com/api) and housestockwatcher(https://housestockwatcher.com/api) APIs designed by Tim Carambat (https://ko-fi.com/rambat), got the disclosed trading data from the APIs and pre-processed the data into tabular forms, make data visualization about their buy or sale actions in different industries or the whole market
Also, I got the S&P 500 and Nasdaq index data from Yahoo Finance API (https://www.yahoofinanceapi.com/), and correlates them by various similarity matrices. Due to the limits of the disclosed information (no specific trading price, volumes, and trading time, but only trading volume ranges and dates), I do not expect to develop trading algorithms based on disclosed information.
I cleaned the data by dropping rows containing N/A values. In addition, because senateTransactions.csv and houseTransactions.csv have different words for “type”, I replace the data in senateTransactions.csv ['Purchase','Sale (Full)','Sale (Partial)','Exchange'], with ['purchase','sale_full','sale_partial', 'exchange'] in order to make the data in two dataframes consistent.

I planned to utilize one of the techniques in lecture “Contextual Text Mining: Mining Causal Topics with Time Series Supervision” -- the Granger Causality Test -- to make conclusion about the casual relationships between senate/house transactions data and S&P500/NASDAQ time series data.

Result:




```
(492, 2)
                S&P 500       NASDAQ
Date
2020-01-02  3257.850098  9092.190430
2020-01-03  3234.850098  9020.769531
2020-01-06  3246.280029  9071.469727
2020-01-07  3237.179932  9068.580078
2020-01-08  3253.050049  9129.240234
2020-01-09  3274.699951  9203.429688
2020-01-10  3265.350098  9178.860352
2020-01-13  3288.129883  9273.929688
2020-01-14  3283.149902  9251.330078
2020-01-15  3289.290039  9258.700195
```

```
(11628, 2)
                         type             amount
transaction_date
2021-09-27           purchase   $1,001 - $15,000
2021-09-13           purchase   $1,001 - $15,000
2021-09-10           purchase  $15,001 - $50,000
2021-09-28           purchase  $15,001 - $50,000
2021-09-17       sale_partial   $1,001 - $15,000
```

```
(2032, 2)
                      type                   amount
transaction_date
2021-11-01       sale_full    $250,001 - $500,000
2021-11-01       sale_full  $500,001 - $1,000,000
2021-11-01       sale_full    $250,001 - $500,000
2021-10-27       sale_full    $250,001 - $500,000
2021-10-29       sale_full    $250,001 - $500,000
```

I successfully visualized the S&P500/NASDAQ time series data and cleaned the data; however, I realized that it the senate/house transactions are categorical, but Granger Causality in Time Series requires to compare two (or more) numerical data, and the data frame need to be the same length.
With that been said, I would need to figure out how to convert the senate/house transactions data is to numerical data, and I would need to concatenate/reshape the data to have the same length. (the length of 492 is the sum of trading days from 1/2/2020 to 12/10/2021). That would be the steps for future works.