



PROBLEM STATEMENT

MScFE Capstone Project

Heng Lyu - aaronlvii@gmail.com

Lucas Rodrigues Abbruzzini - lucas.abbruzzini@proton.com

December 15th 2025

Project Track

The track selected was OTHER as this is a new research proposal. The project researches whether members of Congress (and their spouses/immediate family) continue to outperform the market with their personal stock and options trades even after the 2012 STOCK Act imposed mandatory public disclosure within 45 days. Earlier academic studies found large pre-2012 alphas ($\approx 6\text{--}12\%$ per year). Post-2012 results have been mixed and mostly limited to data ending 2020–2022. The project extends the analysis through the end of 2025, a period that includes the COVID market shock, the 2022 Inflation Reduction and CHIPS Acts, the AI boom, and the 2024–2025 Trump administration transition — all events with major legislative involvement.

Problem statement

Problems

1. Using disclosure-date timing (the only truly exploitable timing for outsiders), do congressional trades still generate statistically significant positive cumulative abnormal returns (CARs) in the days and weeks after public filing?
2. Which subgroups (party affiliation, committee membership, trade size, options vs. stock, sector) show the strongest (or weakest) performance?
3. Can a realistic “copy-trading” (mimetic) portfolio that buys/sells on the first trading day after public disclosure outperform standard factor benchmarks net of transaction costs through 2025?

Code Example and Plan

Step 1: Gather data

- Politicians transaction data: Quiver Strategies
- Stock candlestick data: [Polygon.io](https://polygon.io)

```
client = RESTClient(api_key=polygon_key[0])
✓ 0.0s
```

```
aggs = []
for a in client.list_aggs(
    "AAPL",
    1,
    "minute",
    "2024-01-01",
    "2024-02-03",
    limit=50000,
):
    aggs.append(a)
✓ 0.3s
```

```
df = pd.DataFrame(aggs)
✓ 0.1s
```

```
df.head()
✓ 0.0s
```

	open	high	low	close	volume	vwap	timestamp	transactions	otc
0	191.52	191.52	190.00	190.20	18933.0	190.2570	1704186000000	612	None
1	190.11	190.13	190.00	190.00	8562.0	190.0787	1704186060000	425	None
2	189.90	189.90	189.82	189.82	5355.0	189.8463	1704186120000	129	None
3	189.87	189.87	189.80	189.87	4405.0	189.8634	1704186180000	83	None
4	189.90	189.95	189.90	189.90	2225.0	189.9215	1704186240000	87	None

- Stock new and sentiment data: Alpha Vantage

```
ticker = "AAPL"
start = "20220410T0130"
end = "20220417T0130"
url = f"https://www.alphavantage.co/query?function=NEWS_SENTIMENT&tickers={ticker}&apikey={alphavantage_key[0]}&time_from={start}&time_to={end}"
r = requests.get(url)
data = r.json()

✓ 0.3s

{'items': '6', 'sentiment_score_definition': 'x <= -0.35: Bearish; -0.35 < x <= -0.15: Somewhat-Bearish; -0.15 < x < 0.15: Neutral; 0.15 <= x < 0.35: ...'}

news = pd.DataFrame(data)

✓ 0.0s

news.feed[0]

✓ 0.0s

{'title': 'Apple helps suppliers rapidly accelerate renewable energy use around the world',
'url': 'https://www.apple.com/newsroom/2022/04/apple-helps-suppliers-rapidly-accelerate-renewable-energy-use-around-the-world/',
'time_published': '20220414T182544',
'authors': [],
'summary': 'Apple announced that its suppliers more than doubled their use of clean power over the last year, with over 10 gigawatts operational out ...',
'source': 'Apple',
'category_within_source': 'General',
'source_domain': 'Apple',
'topics': [{'topic': 'technology', 'relevance_score': '0.826507'},
{'topic': 'energy transportation', 'relevance_score': '0.726096'},
{'topic': 'manufacturing', 'relevance_score': '0.601629'}],
'overall_sentiment_score': 0.285539,
'overall_sentiment_label': 'Somewhat-Bullish',
'ticker_sentiment': [{'ticker': 'RDN',
'relevance_score': '0.329672',
'ticker_sentiment_score': '0.041449',
'ticker_sentiment_label': 'Neutral'},
{'ticker': 'AAPL',
'relevance_score': '0.974955',
'ticker_sentiment_score': '0.406813',
'ticker_sentiment_label': 'Bullish'},
{'ticker': 'DO',
'relevance_score': '0.636420',
'ticker_sentiment_score': '0.319873',
'ticker_sentiment_label': 'Somewhat-Bullish'},
{'ticker': 'GLW',
'relevance_score': '0.626677',
'ticker_sentiment_score': '0.340710',
'ticker_sentiment_label': 'Somewhat-Bullish'}]}
```

- Stock financial statement and income statement data(10-K) data: Financial Model Prep

```
def get_financials_fmp(ticker, api_key):
    # FMP Endpoint for Income Statement (limit=5 gets last 5 years)
    url = f"https://financialmodelingprep.com/stable/financial-reports-json?symbol={ticker}&year=2022&period=FY&apikey={api_key}"

    try:
        response = requests.get(url)
        data = response.json()
        return data

    except Exception as e:
        print(f"Error fetching data: {e}")
        return pd.DataFrame()

statement_aapl = get_financials_fmp("AAPL", fmp_key[0])

statement_aapl

{'symbol': 'AAPL',
 'period': 'FY',
 'year': '2022',
 'Cover Page': [{'Cover Page - USD ($) shares in Thousands, $ in Millions': ['12 Months Ended']}],
 'Items': ['Sep. 24, 2022', 'Oct. 14, 2022', 'Mar. 25, 2022']],
 'Entity Information [Line Items]': ['\xa0', '\xa0', '\xa0']],
 'Document Type': ['10-K', '\xa0', '\xa0']],
 'Document Annual Report': ['true', '\xa0', '\xa0']],
 'Current Fiscal Year End Date': ['--09-24', '\xa0', '\xa0']],
 'Document Period End Date': ['Sep. 24, 2022', '\xa0', '\xa0']],
 'Document Transition Report': ['false', '\xa0', '\xa0']],
 'Entity File Number': ['001-36743', '\xa0', '\xa0']],
 'Entity Registrant Name': ['Apple Inc.', '\xa0', '\xa0']],
 'Entity Incorporation, State or Country Code': ['CA', '\xa0', '\xa0']],
 'Entity Tax Identification Number': ['94-2404110', '\xa0', '\xa0']],
 'Entity Address, Address Line One': ['One Apple Park Way', '\xa0', '\xa0']],
 'Entity Address, City or Town': ['Cupertino', '\xa0', '\xa0']],
 'Entity Address, State or Province': ['CA', '\xa0', '\xa0']],
 'Entity Address, Postal Zip Code': ['95014', '\xa0', '\xa0']],
 'City Area Code': ['408', '\xa0', '\xa0']],
 'Local Phone Number': ['996-1010', '\xa0', '\xa0']],
 'Entity Well-known Seasoned Issuer': ['Yes', '\xa0', '\xa0']}]
```

Step 2: EDA

- Find the top-performing member of the Congress trading portfolio.
- Filter all the stocks' daily candlestick data in the portfolios.
- Use plotly to plot the candlestick and figure out the spikes from 2022-2025.

```
# 1. Fetch Data
ticker = "NVDA"
df = yf.download(ticker, period="1y", interval="1d")
df = df.stack(level='Ticker').reset_index()

# 2. Create Subplots (2 Rows: Price on top, Volume on bottom)
# row_heights=[0.7, 0.3] gives 70% space to Price, 30% to Volume
fig = make_subplots(
    rows=2, cols=1,
    shared_xaxes=True,
    vertical_spacing=0.03,
    subplot_titles=(f'{ticker} Price', 'Volume'),
    row_heights=[0.7, 0.3]
)

# 3. Add Candlestick Trace (Row 1)
fig.add_trace(go.Candlestick(
    x=df.Date,
    open=df['Open'],
    high=df['High'],
    low=df['Low'],
    close=df['Close'],
    name='Price'
), row=1, col=1)

# 4. Add Volume Trace (Row 2)
# We color the volume bars based on price change (Green for up, Red for down)
colors = ['green' if row['Open'] - row['Close'] >= 0
          | else 'red' for index, row in df.iterrows()]

fig.add_trace(go.Bar(
    x=df.Date,
    y=df['Volume'],
    marker_color=colors,
    name='Volume'
), row=2, col=1)

# 5. Add Moving Average (Optional, on Row 1)
df['MA20'] = df['Close'].rolling(window=20).mean()
fig.add_trace(go.Scatter(
    x=df.Date,
    y=df['MA20'],
    line=dict(color='orange', width=2),
    name='20-Day MA'
), row=1, col=1)

# 6. Final Layout Adjustments
fig.update_layout(
    title=f'{ticker} Candlestick Chart with Volume',
    template='plotly_dark',
    xaxis_rangeslider_visible=False, # Hides the slider (redundant with zoom)
    height=800 # Taller figure to accommodate both plots
)

# Show the Plot
fig.show()
```



- Check how many percentages of spikes these politicians have captured.

Step 3: Categorical EDA

Comparative Distribution Analysis is required to provide answers to the question of which Party, Committee or Sector is the strongest.

1. Party & Committee Analysis

Use Box-and-Whisker plot. Use Political Party (Democrat vs Republican) or Committee name as X-axis. And use returns as Y-axis. Maybe the overall return or one month return.

Future thoughts: May divide members of the Congress into different subgroups. Maybe divided by the transaction frequency, invested amount, etc.. Also may divide groups through clustering methods like K-means.

2. The “Insider” Heatmap

May solve question like “Are Energy Committee members better at trading Energy stocks than Tech stocks?”

Matrix:

- *Rows:* Committees (Energy, Armed Services, Finance).
- *Columns:* Stock Sectors (Energy, Defense, Banks).
- *Color:* Returns.

Step 4: Optimization portfolio

The typical copy-cat portfolio can concentrate on a specific sector or stocks. For example, everybody loves Nvidia. But this situation can be corrected with optimization.

The Hypothesis: The unmanaged congress portfolio may be inefficient (high volatility, low diversification). We may come up with some smart portfolio or better asset allocation using modern portfolio theory. And then compare the performance of each portfolio within the test set.

The techniques:

- Mean-variance optimization
- Black-Litterman Model(PyPortfolioOpt)

Step 5: Back-testing

Back-test the portfolio performance on the testing dataset. There is a time gap between the transaction date and the disclosure date for the members of Congress. So the individual investor may see the transaction history 45 days after the actual transaction date. At the disclosure date or filing date, the situation or stock price is totally different. If an investor just copy-cat the transactions from the politicians, they may suffer incredible financial loss.

Goals and Objectives

Primary Objective

To perform an intensive Exploratory Data Analysis (EDA) that measures the number of cumulative abnormal returns (CARs) accessible to the market post-disclosure by congress, and to determine certain congressional subgroups that perform consistently better than the S&P 500.

Specific objectives

- Temporal Return Analysis: Compute and plot the Cumulative Abnormal Return (CAR) of all trades of the congress between a constant timeframe with a 95% confidence interval to find whether the mean of the returns is statistically significant or not.
- Subgroup Performance Profiling: Compare the returns of Democrats and Republicans and House and Senate on plots to identify the difference in the median performance and the number of outliers (risk).
- Optimized portfolio for individuals: Using modern techniques and machine learning to develop a smart portfolio for individuals.