

Dataset Context and Statistics

Context of the Dataset

The dataset, *legislative_trades.csv*, comprises financial transaction disclosures mandated by the Stop Trading on Congressional Knowledge (STOCK) Act. It captures stock and asset trades made by members of the United States Congress (both the Senate and the House of Representatives), their spouses, and dependent children. The demographic is entirely composed of federal lawmakers and their immediate families. The dataset covers transactions over several recent years, capturing major market events such as the early 2020 COVID-19 market crash and the subsequent tech bull run. Depending on the extraction timeframe, the sample size consists of tens of thousands of individual trade records spanning the 535 members of Congress. The primary purpose of collecting this dataset is to monitor ethical compliance, ensure transparency, and investigate potential insider trading by comparing congressional trading behavior against broader market indices (e.g., the S&P 500).

Key Variables

To investigate whether members of Congress achieve abnormal market returns or exhibit prescient market timing, several key variables are utilized:

- **transaction_date** and **disclosure_date**: Essential for establishing the timeline of the trade relative to market-moving events and evaluating compliance with reporting windows.
- **ticker** and **asset_description**: Allow for joining the dataset with historical financial market data to calculate post-transaction returns.
- **transaction_type** (e.g., Purchase, Sale): Determines the directional bet of the lawmaker.
- **amount_min**, **amount_max**, and **amount_avg**: Used to estimate the financial volume and weight of the trade.
- **legislator_name**, **chamber**, **party**, and **office**: Enable demographic and grouping analyses to see if specific affiliations correlate with outsized returns.

Strengths and Limitations

Strengths: The primary strength of this dataset is its authoritative nature; it is scraped directly from official US government disclosure portals (efdsearch.senate.gov and disclosures-clerk.house.gov). Legally mandated reporting provides a comprehensive overview of financial activity that would otherwise be private, offering a rare empirical lens into the intersection of legislative knowledge and personal financial gain.

Limitations: The dataset possesses several inherent limitations that complicate exact financial analysis. First, trade volumes are reported in broad value brackets (e.g., \$1,001 – \$15,000) rather than exact dollar amounts, meaning that the true financial gain or loss can only be approximated using the **amount_avg**. Second, there is a legal reporting delay of up to 45 days, and sometimes lawmakers file late, causing friction in real-time tracking. Third, the data includes trades made by spouses and trusts, muddying the waters on whether the legislator actively directed the trade based on privileged information. Finally, while Senate data is readily digitized, House disclosures often rely on PDF scans, which necessitates OCR processing and can introduce parsing errors or missing data.