

Insider Trading Analysis

Nihal Desai · Ramya Kondrakunta · Santosh Saranyan · Srinath Iyer

Summary: The project aims to deal with analyzing Insider trading data. This project will do an exploratory analysis of the insider filling and stock price data, based on the Form 4 filings from the U.S. Securities and Exchange Commission (SEC)[1] and Yahoo Finance Data[2]. In this project, we aim to answer questions such as how companies insider trade? How trading behavior of insiders changes over time, or by the position of an insider, so on. Also, we try to check if there is a correlation between insider trading and stock prices of the company?

Insider Trading is the trading of public company stocks or securities, based on nonpublic information about the company. An example can be a CEO buying shares right before making a big announcement, or a CFO selling stocks based on the quarterly sales information before they are made public. To avoid this the SEC requires Chief Officers, Directors, so on to file Form 4 whenever they trade their company's share[3][4].

For the project, we are generating the Insider trading dataset by scraping the SEC website. We will be using APIs from Yahoo Finance to get a company's stock prices. We will be using methods such as statistical inferences and data visualization to analyze and present the findings, and web scraping to collect data. Finally, we will be using tools such as R, Python, SQL, and possibly PowerBI.

Proposed Plan: To generate the data we had to use Web Scraping. As we were unable to find a dataset or API that was both free and had all the data that we require for the project. We used Python to do the web scraping directly from the SEC website. Once the data is generated, we will be using SQL, to tidy the data set. After which we will be using R, to perform ETL on the data, using statistical inferences and data visualization to do the analysis and generating visualization for the project presentation.

The challenges we faced so far were to find the data. We were unable to find any ready dataset, so we decided to scrape data from the SEC website. However, SEC only allows 10 requests per second because of which the data collection is taking a long time. A single Quarter has up to 80,000 filings, which took us 6 hours per quarter to scrape. We were able to collect data for one year. Another challenge will be to clean the data. We will be using SQL to store the data set in the Database and creating a schema and break the data into tables.

Preliminary Analysis: The final dataset contains 1,80,233 rows and 11 columns. The dataset description is as shown below:

| SNo. | Attribute Name | Description | Datatype |
|------|---------------------------------|---|-----------|
| 1 | fid | Unique ID of the Form | String |
| 2 | cik | Unique ID of the Company | String |
| 3 | rptownercik | Unique ID of the Share Owner | String |
| 4 | rptownername | Name of the Share Owner | String |
| 5 | relationship | Relationship with the company | List |
| 6 | date | Date | Date |
| 7 | transactioncode | Type of transaction | Character |
| 8 | transactionshares | Number or shares | Numeric |
| 9 | transactionpricepershare | Price per shares at the time of transaction | Numeric |
| 10 | sharesownedfollowingtransaction | Shares owned by Owner following transaction | Numeric |

We observed that there were duplicate and missing values in the dataset. After cleaning the data, we performed preliminary data analysis and plotted the following graphs. The first analysis was to see what type of transactions were filed the most. From Fig 1, we can see that Acquisition (A) and Sales (S) are the most common type of filings. The second analysis shows the selling and purchasing transactions during the year 2019. From Fig 2, we can see that there are more insider transactions for selling company stocks rather than purchasing stocks. This makes sense as the individuals that file insider forms are mostly paid through stocks.

References:

- [1] "EDGAR Application Programming Interfaces." <https://www.sec.gov/edgar/sec-api-documentation>.
- [2] "Yahoo Finance." <https://www.sec.gov/edgar/sec-api-documentation>.
- [3] A. Tamersoy, B. Xie, S. L. Lenkey, B. R. Routledge, D. H. Chau, and S. B. Navathe, "Inside insider trading: Patterns & discoveries from a large scale exploratory analysis," *Proc. 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2013*, pp. 797–804, 2013, doi: 10.1145/2492517.2500288.
- [4] A. Tamersoy *et al.*, "Large-scale insider trading analysis: patterns and discoveries," *Soc. Netw. Anal. Min.*, vol. 4, no. 1, pp. 1–17, 2014, doi: 10.1007/s13278-014-0201-9.

Supplementary Material:

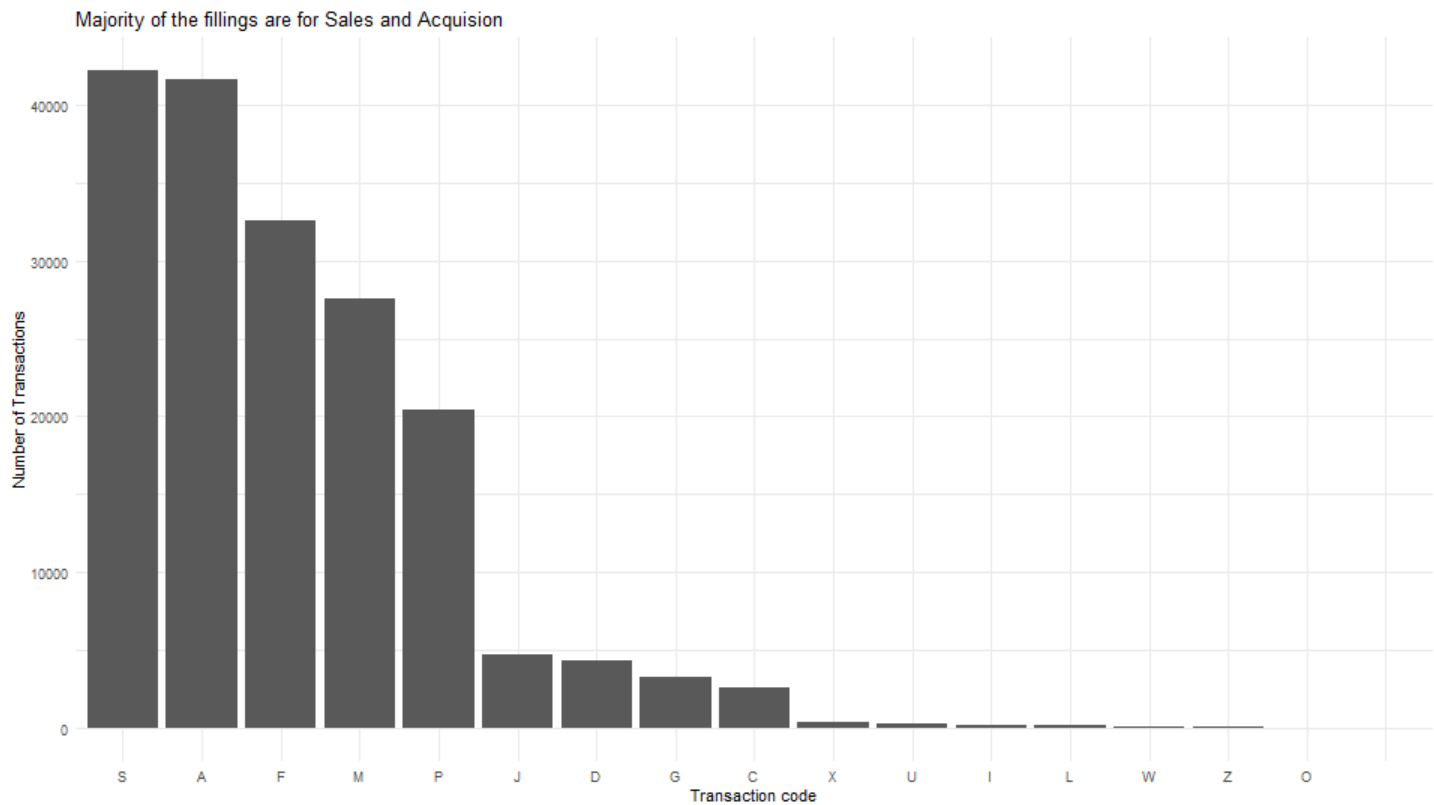


Figure 1

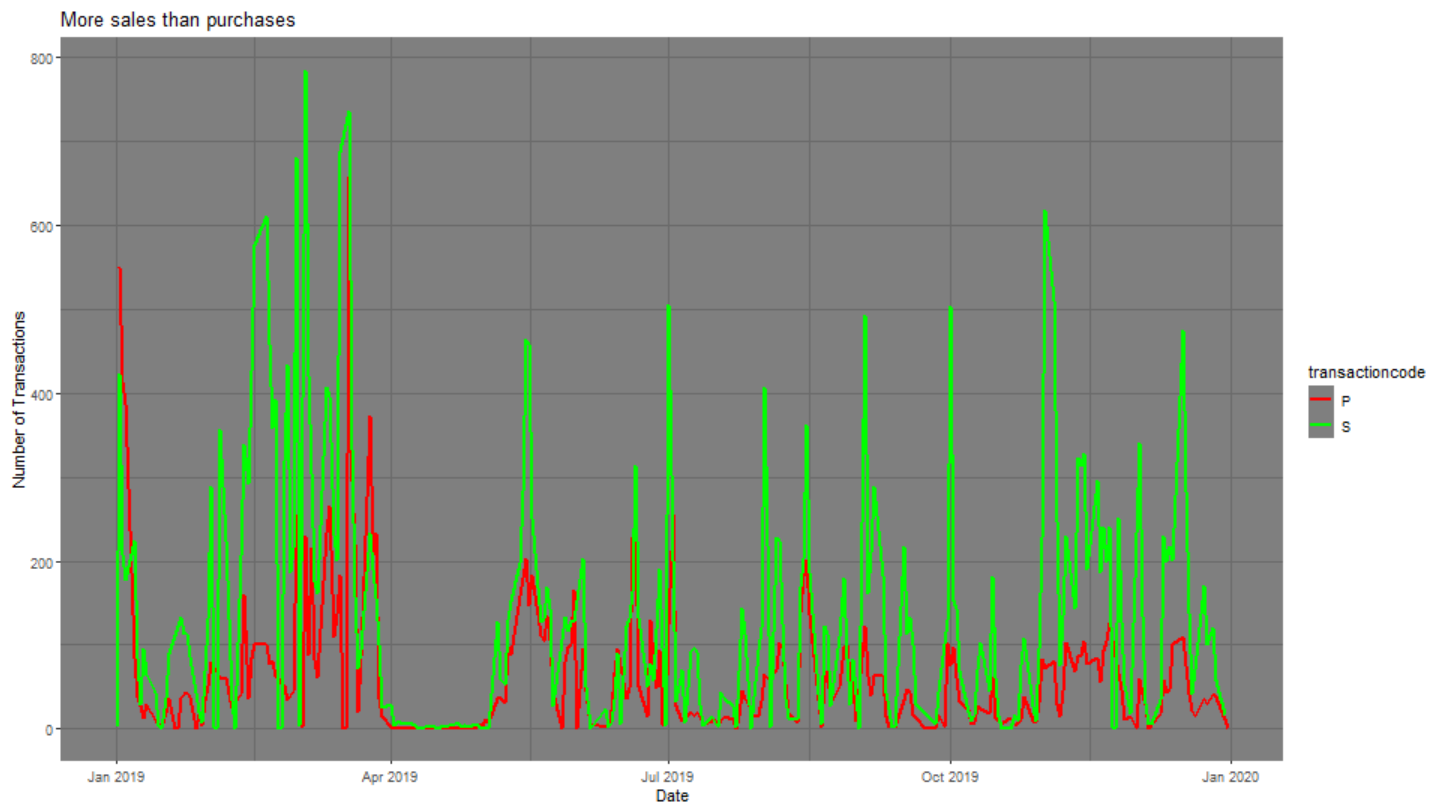


Figure 2