

MIS 6382 - Object Oriented Programming in Python – Data Visualization Project

Group 2

-
- *SXC220023 - Sai Thushara Chowdary Chundru Venkateswara Prasad*
 - *RXD230004 - Radhika Dubey*
 - *AXY210025 - Aditya Praveen Yewankar*
 - *JXA220069 - Jagannath Ajaya Kumar*
 - *SXR210051 - Sadaf Riaz*
-

Introduction

The data set contains information about different companies' financial fluctuations over a while by noticing the change in the price of their stocks followed by the change in price. To begin with, the dataset has 10000 rows of data with the following columns:

- **uniqueID** -> It can be identified as the primary key in the dataset to uniquely identify each record in the table.
- **date** -> The date column ranges from January 2022 to January 2023 helping us identify stock prices and fluctuations of certain companies on different dates.
- **index_name** -> The name of the stock index.
- **company** -> The different companies that were taken into consideration in the dataset to which the stock indexes belonged.
- **type** -> This tells us the type of the stock market index eg: large cap or nonlarge cap.
- **rating** -> It ranges from 1-5 which speaks about the rating of the company.
- **price** -> the stock price of the company or the index.
- **fluct_rate** -> The fluctuation rate regarding the change of price of the stock over the given dates.

Data types:

Column	Data Type
uniqueID	int 64
date	object
company	object
index_name	object
type	object
rating	float 64
price	float 64
fluct_rate	float 64

In a nutshell, the data types are a mix of numerical (int, float) and categorical/text (object) values.

Handling Missing Data

- An initial exploration showed there were some missing values in the date, rating, price, and fluct_rate columns.
- In terms of handling these missing values, all the numerical values were replaced with the median of the non-null values. The median was considered instead of the mean to make sure the replaced values were not affected by the outliers.
- For the date column, the statistical value mode was taken by replacing the null values with the most frequent date in the entire dataset.

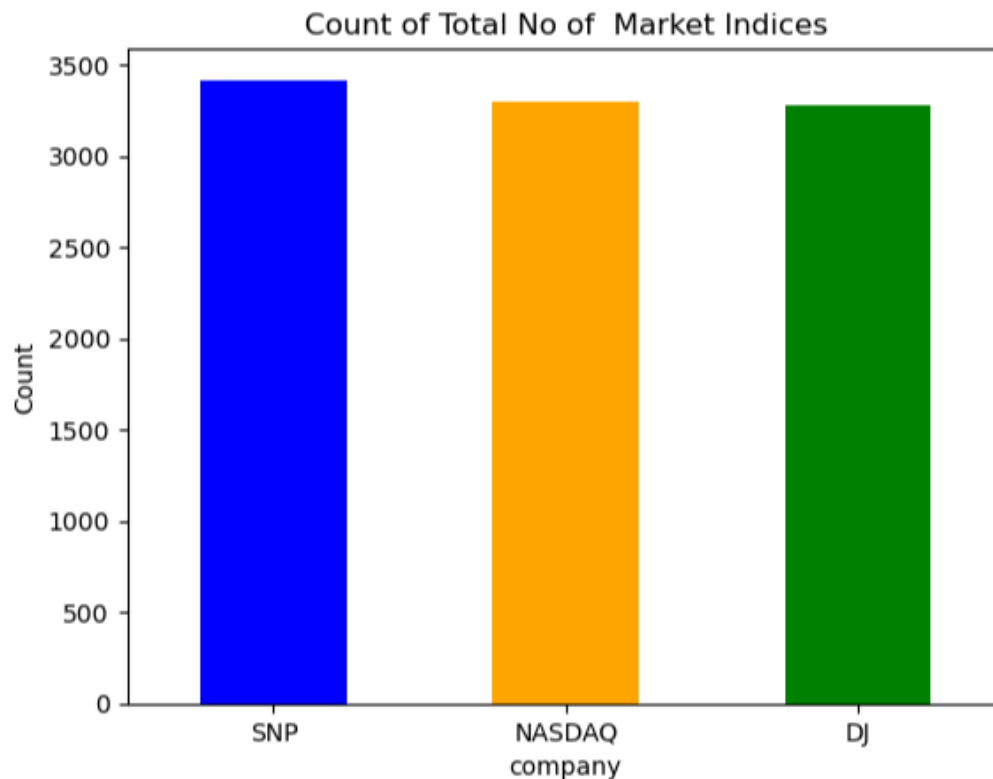
Visualizations and Trends

1. **Line chart showing price overtime:** Prices tend to increase over the year 2022 but with some periodic drops. There is noticeable volatility.
2. **Scatter plot of rating vs price:** There is a weak positive correlation between higher prices and increased rating.
3. **Bar chart of the total number of companies:** There is a higher count of non-large cap companies than the large-cap companies.
4. **Pie chart of company counts:** NASDAQ indexes make up the highest percentage followed by Dow Jones and S&P.
5. **Box plot of prices by company type:** Large-cap indexes and non-large indexes have almost the same median and equal spread of prices.
6. **Histogram of fluctuation rate:** Markets fluctuate fairly frequently. The average fluctuation is about 15% during a year.
7. **Heatmap of price changes:** For all three indices, ratings 2, 3 and 4 are more common than ratings 1 and 5, with 3 being the most common rating for all three indices.

Plot: Bar Graph

Objective: Visualize the total number of companies in terms of count and in terms of their type

Inference:

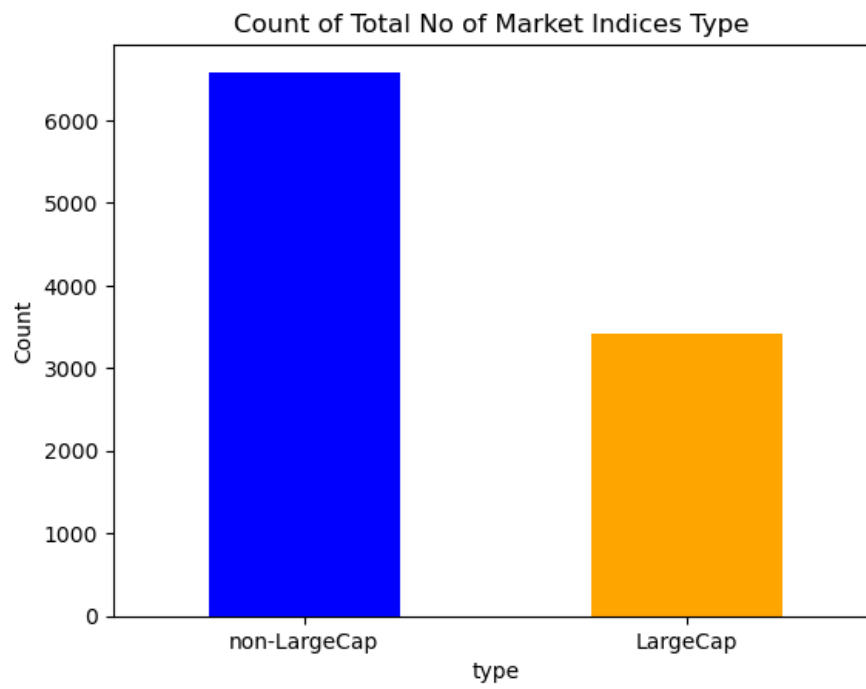


We can infer from the graph that there are a total of 3 companies with almost an equal number present in our dataset. The three companies are SNP, NASDAQ, DJ followed with a count between the range of 1500-1750.

This shows that almost an equal number of instances of 3 different companies have been considered with regards to their stock price and fluctuation rate.

However, while taking into consideration the types(Large cap and Non large cap) of the company, there seems to be a dominant presence of more non-large companies than the large-cap type.

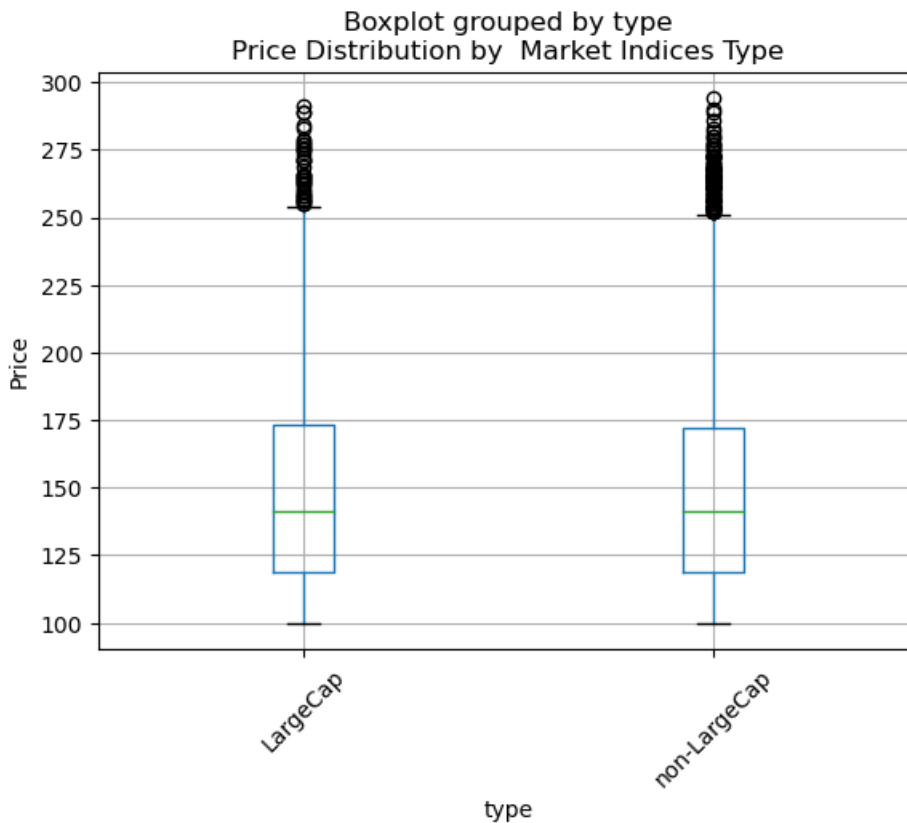
Group2



Plot: Box Plot

Objective: Visualize the distribution of price of stocks concerning company type

Inference:



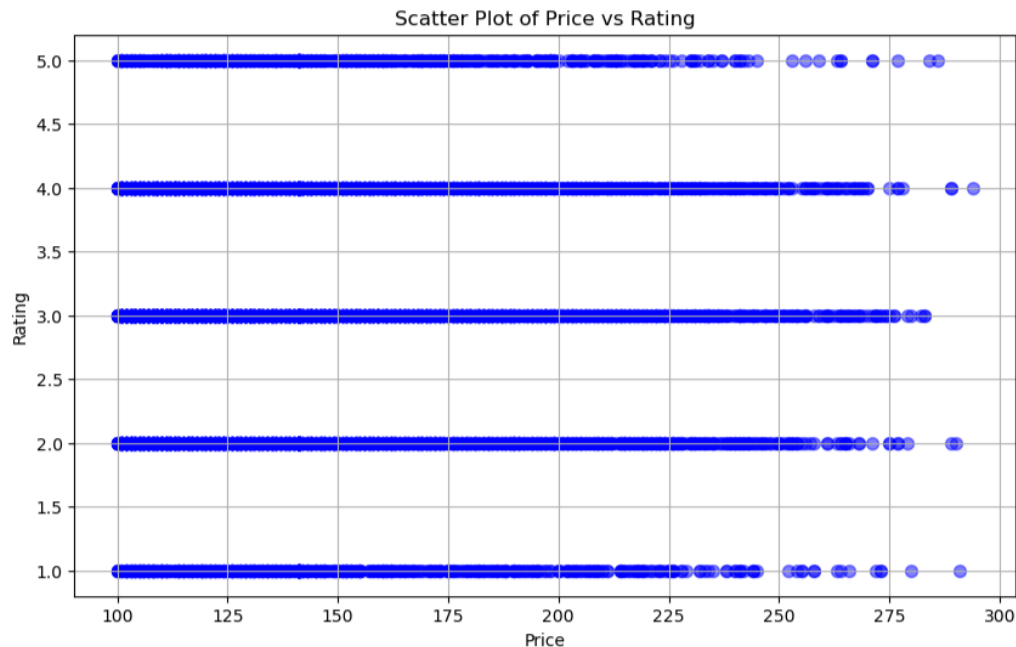
We can infer from the graph that the price of the stock prices is distributed within a range of 100 – 250 which shows the maximum increase in the prices of the stock is around 250 dollars and there is a rare occurrence beyond 250 dollars and further considered as outliers. The median price of the stock is approximately 140.

The type of company has no causal effect on the price of their shares and almost remains the same.

Plot: Scatter Plot of Index Price vs Rating

Objective: Visualize the relationship between index price and rating

Inference:



Data Points:

Each point represents an index observation.

Point color indicates index type.

Point size indicates fluctuation rate.

Trend:

There is a weak positive correlation between price and rating, where higher priced indexes tend to have slightly higher ratings.

Insights:

Large cap indexes dominate the higher price tiers

Non large cap indexes show inconsistent ratings despite lower prices

Inference

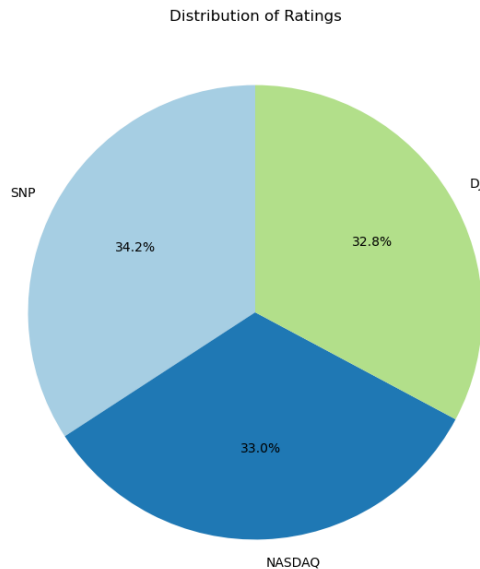
The scatter plot shows a minor relationship between increased index price and higher ratings. This suggests investors may associate very high prices with quality when evaluating indexes. However, lower rated large caps are rare so the price-rating correlation is not extremely strong.

Plot: Pie Chart Distribution of Category

Objective: Visualize the relationship between companies' performance levels

Data Points:

- Count the occurrences of each category.
- Percentage represents the distribution of rating.
- Rating status



Trend:

The companies are categorized by performance levels. A slice represents each company, and the size of each slide corresponds to the percentage of the whole.

Inference

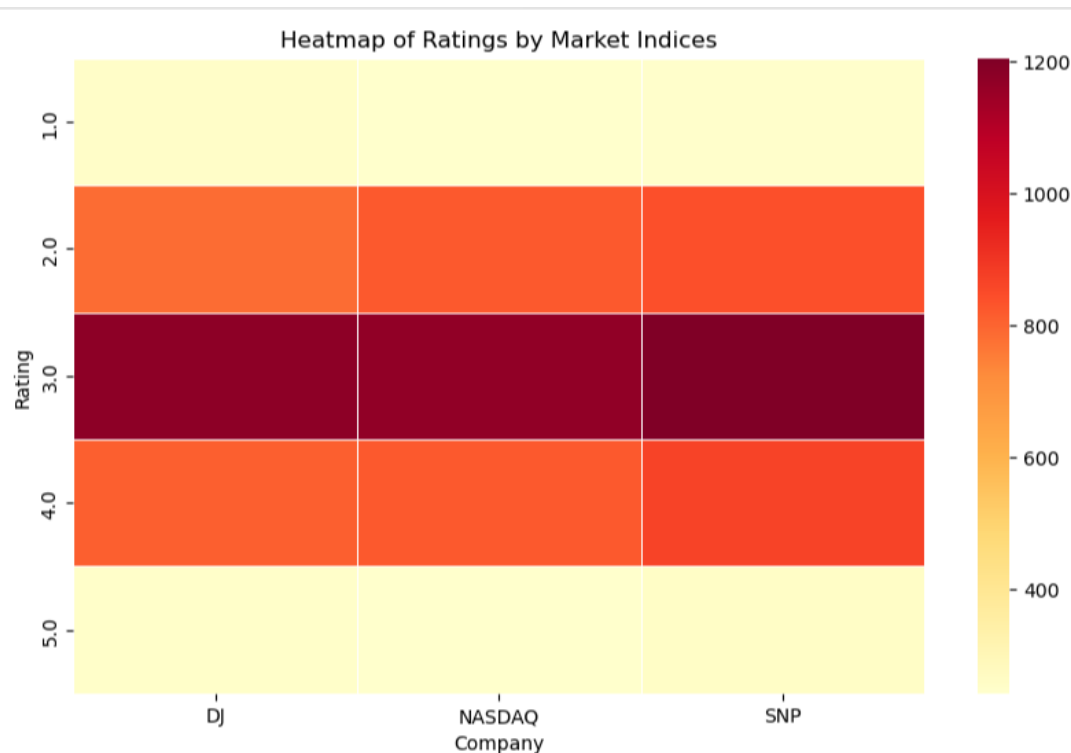
This Pie chart distributes the companies by their category and their performance. Each category represents a specific aspect of criteria by which companies are classified. The performance of each company is depicted by the percentage of the pie slice. The higher percentage of the slice of the pie chart suggests that a significant proportion of companies fall into the high-performance category. The lower percentage slice of the pie chart indicates that fewer companies perform well in that category.

Heatmap: Rating Frequencies by Indices

Objective: Based on the given statistics, the heatmap shows a study of the rating frequencies by index. Data on different financial indices, firms, kinds, ratings, costs, and fluctuation rates are all included in the dataset. Visually examining how ratings are distributed among the indices is the goal.

Inference:

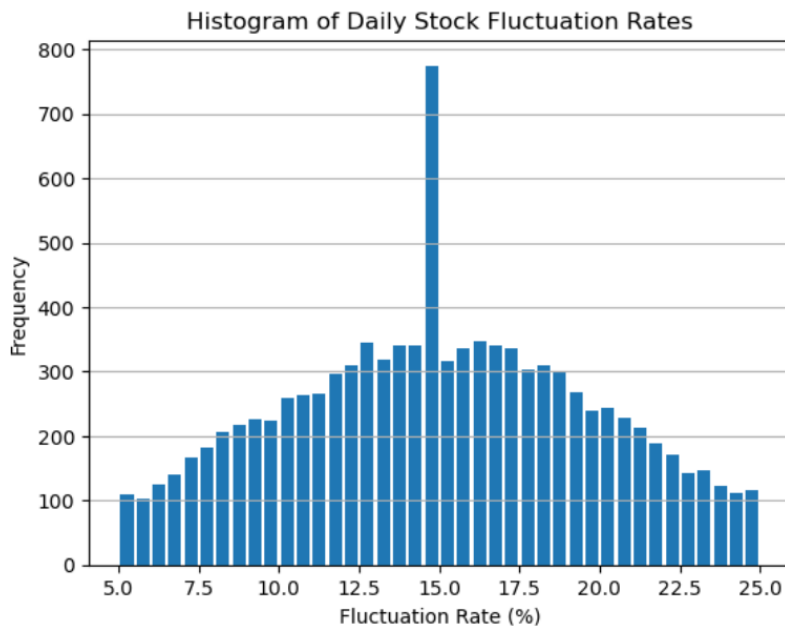
- I. Distribution of Ratings:
 - Compared to rating 1, ratings 2, 3, and 4 are more prevalent.
 - The rating that is most given on DJ, NASDAQ, and SNP is 3.
 -
- II. Business-Rating Relationship:
 - In every case, DJ, NASDAQ, and SNP get comparable ratings with SNP having a higher frequency of ratings 2,3 and 4 than DJ and NASDAQ.
 - For all three of the companies, the most common rating is 3.
 - From this we can say that investments made in all 3 indices can be considered as having the same level of risk associated with it as the most common rating associated with all of them is 3.



Histogram- Visualizing fluctuation rate.

Trend- Markets fluctuate fairly frequently. As can be seen from the graph that the fluctuation rate is centered at 15%. This is something that can be observed in stock markets in general.

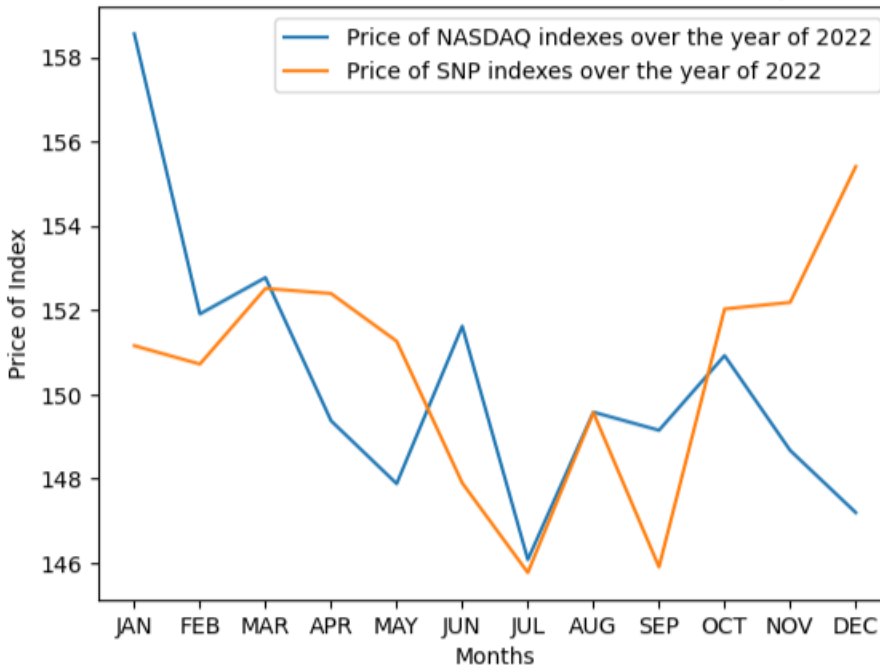
Inference- The stock market fluctuates for several different reasons, but the biggest overall factor is supply and demand. Prices of stocks rise when the supply of shares for sale is not enough to meet investors' demands. When investors' demand for shares falls, so does the price of the shares. This causes volatility.



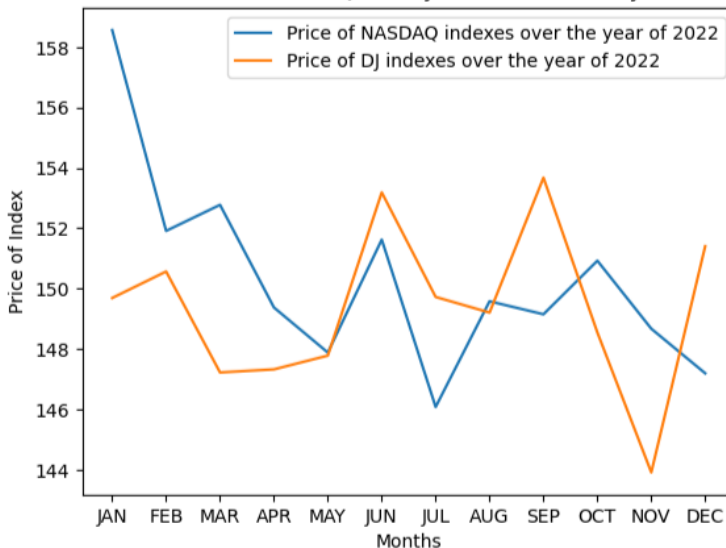
Line chart: Comparing price trends of different indices over the year 2022.

Introduction: The line charts show the price trend of the average price of all the companies within the indices NASDAQ, SNP, and DJ. Every point on the x-axis represents a month in the year 2022. The first chart compares a large cap index with a non-large index SNP and NASDAQ, respectively. While the second chart compares two non-large indices NASDAQ and DJ.

Price fluctuation of NASDAQ and SNP indexes over the year of 2022



Price fluctuation of NASDAQ and DJ indexes over the year of 2022



Slicing the data-

- The data is first split into months.
- Indices are then segregated based on their name like NASDAQ, SNP and DJ.
- Mean for these indices for each month is then calculated and plotted over each month comparing them for different indices.

Trend-

- Prices tend to increase over the years but with some periodic drops.
- There is noticeable volatility.

Inference-

- Both the non-large cap indexes tend to move in similar trends, they rise and dip in the same months.
- Having a large cap doesn't necessarily mean the company will outperform a company with a non-large cap.

Conclusion of the project:

The data concludes that both the large-cap and non-large-cap stock market indexes perform at an equal rate and they don't seem to outperform the other. We could safely assume all the given stock market index moves are influenced by the market performance and economy. This could be further backed up by our findings from the line graph, heat map, and box plot that the prices almost equally increase or decrease at the same time of the year and almost an equal rating is given to all the stock market indexes. Even with the box plot, we could say that the median of the stock price remains the same for both the large cap and non-large cap index. Overall, the market has been growing at a healthy pace, likely contributing to the higher ratings despite some volatility concerns. There do not appear to be any severe anomalies or crashes present in this slice of data. Expanding the analysis to a longer time horizon could reveal more extreme events and test the strength of some of the hypothesized trends.