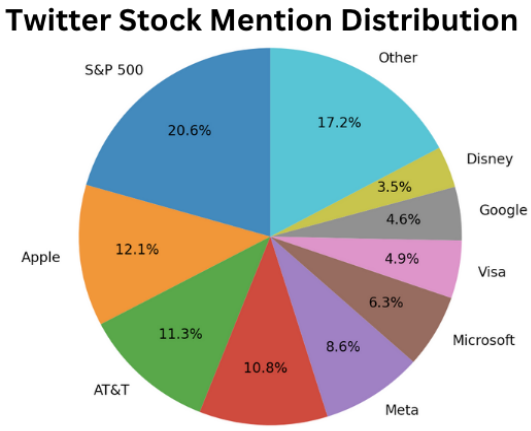
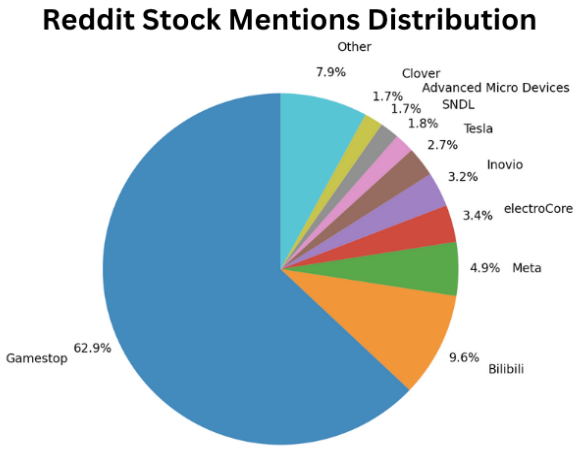
CSCI 1951A Final Project Visualizations

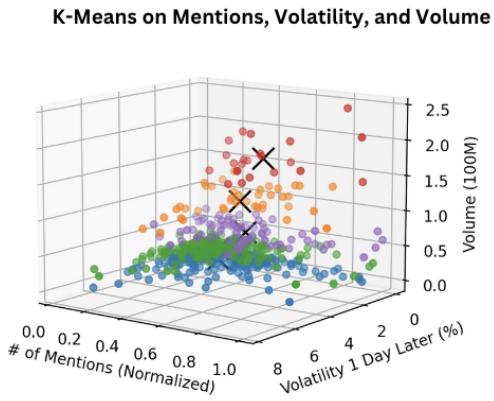
Social Media Shares Pie Chart

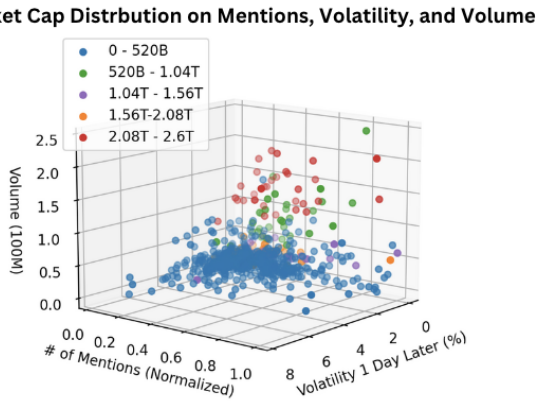




- Since the intent of this visualization is to display the shares of social media mentions across Twitter and Reddit we thought a pie chart would be most effective. This would allow viewers to observe the number of social media mentions for each company relative in size to others.
- A bar graph might have been used instead; however, it would not have had the same comparative element that the pie chart offers.
- Most of the issues visualizing these results were common pie chart alignment tasks. Being able to space the percentages and stock labels adequately took some time, but mainly required a bit of testing.
- When paired with their respective titles, these pie charts are pretty intuitive. Most viewers would understand the data at first glance. Text might benefit viewers who are unfamiliar with some of the stock tickers and might be confused as to their meaning.

K-Means Scatter Plots





- When plotting the data distribution grouped by K Means and grouping by market capitalization, we knew a scatter plot would be of best use considering the involvement of distributions. Since our data involved three variables; volume, volatility, and number of social media mentions, we opted for a three dimensional scatter plot.
- Another graph showing distribution density, such as a Hexbin graph, could have been used. However, we thought the scatter plot would be the most straightforward for viewers and highlight the distinction between each data point the best.
- We had challenges matching up the coloring between the k-means visualization and the market cap visualization. Our k-means coloring would be random throughout the groups, and we wanted to match these colors for the market cap visualization as to better depict the relationship. We also had challenges in getting the axis depiction to be the same for the k-means visualization and market cap visualization. For some reason one of the axes was inverted, and we had to flip it back in order to get the same view on both visualizations.
- Our visualizations do require text for context, because we wanted to explain to the reader that we were comparing the groupings on the two graphs, which isn't necessarily obvious upon looking at them even though they are formatted the same.

Linear Regression Scatter Plot

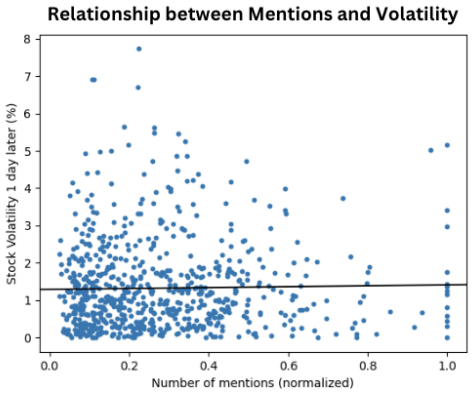


Figure 3: Linear Regression on Twitter Data

- The intent of this visualization is to observe the results of the linear regression we ran as part of our analysis. We wanted to display the data alongside the linear regression model and decided a scatter plot with the regression line would do so best.
- When it comes to observing the direction of our regression model and whether or not it's valid, there really is no other graph that would have been as effective. A box plot might have been used to observe the distribution of the data along the x-axis which might have helped highlight outlier data; however, this was not the purpose of this graph, so it would not have been as effective.
- We found challenges in graphing the linear regression because we weren't sure how to graph the regression line. After experimenting with graphing it in seaborn, we eventually found the statsmodel abline_plot() method which worked to plot it.
- Paired with the title the graph is pretty straight forward. Considering the two-dimensionality of the linear regression, it does not require too much explanation in order for viewers to understand its meaning. However, text will provide context regarding the regression approach and our findings that are pertinent to the overall visual.