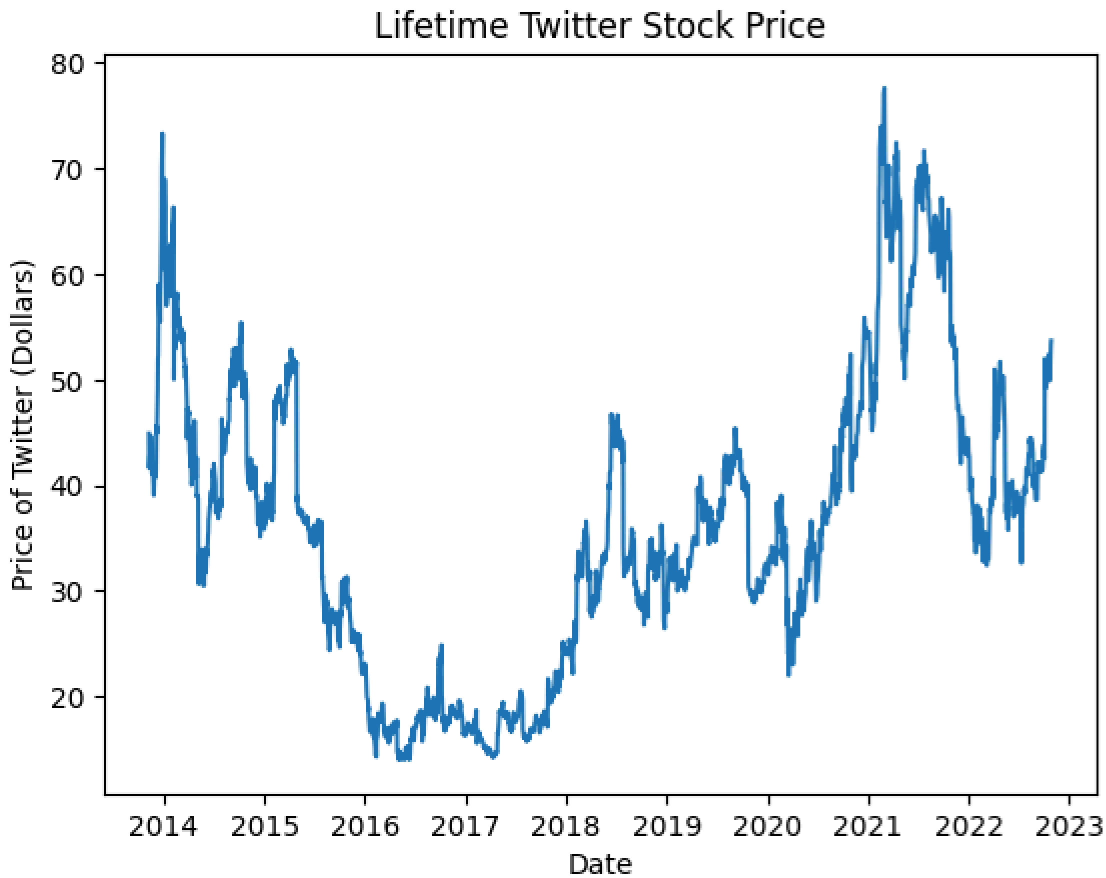
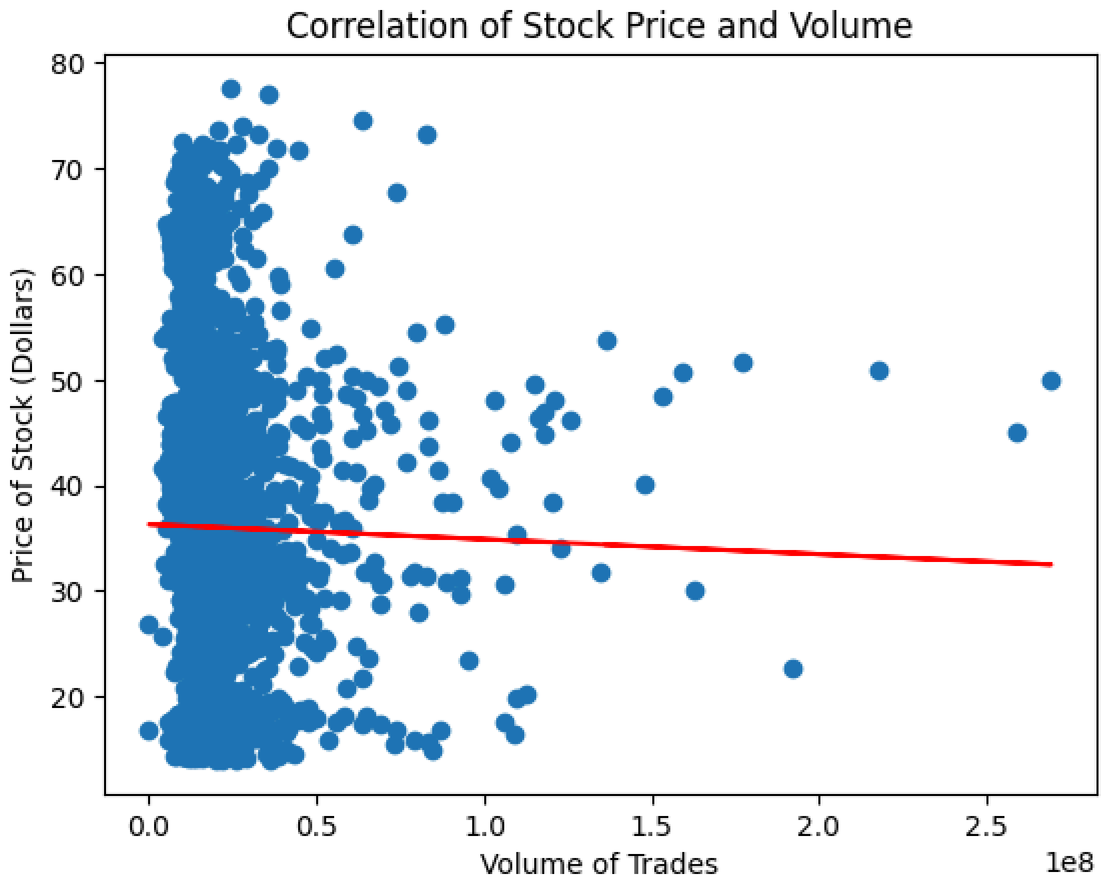
For my project the topic I chose was Twitter/X, specifically the effect that Twitter transitioning into X had on the platform itself. Other social media platforms have rebranded themselves but none were as prolific or controversial as Elon Musk purchasing Twitter and rebranding it into X. I believe that this makes Twitter/X a good platform to research because the controversial rebranding could have a significant impact on the platform. Social media itself is important to look at because it has become so ingrained in our everyday lives. For some people social media might be where they get the majority of their information, especially on a platform like Twitter/X known for its accessibility to news coverage. Another reason that I chose this topic was because the companies that run social media companies are tech giants. Social media in general is a field that has a lot of opportunities for computer science related jobs. By researching this topic I am better informed about how these social media platforms work. This will hopefully give me an advantage if I ever decide to seek employment from any of these companies.

One of my first thoughts was to find a dataset about active twitter users, so that I could compare the user base before and after the rebranding. This is where I ran into the first of my issues. Shortly after the rebranding there were several API changes that made it much more difficult to collect data. This meant that there wasn't a lot of quality data after the rebranding took place. Putting this idea on the backburner I began to look at Twitter's stock price. I figured this could be a good metric to measure the public's opinion on the rebranding and it was easily available data. This is where I ran into my next problem, when Twitter rebranded to X the company actually went private as well. Luckily for me I found out that the company actually didn't go private until several months after the rebranding happened, so I had enough data to compare it. In this dataset the key values I looked at were the closing price of the stock and the volume. The closing price of the stock is the easy choice because it best represents the value of the stock from that day. The reason I chose to look at the volume too is it showcases the 'popularity' of the stock. For this dataset most of the processing I had to do was dropping the columns that were unnecessary to me. I also ensured that the date was recognized as a date by pandas so that I could use it for the DataFrame's index.



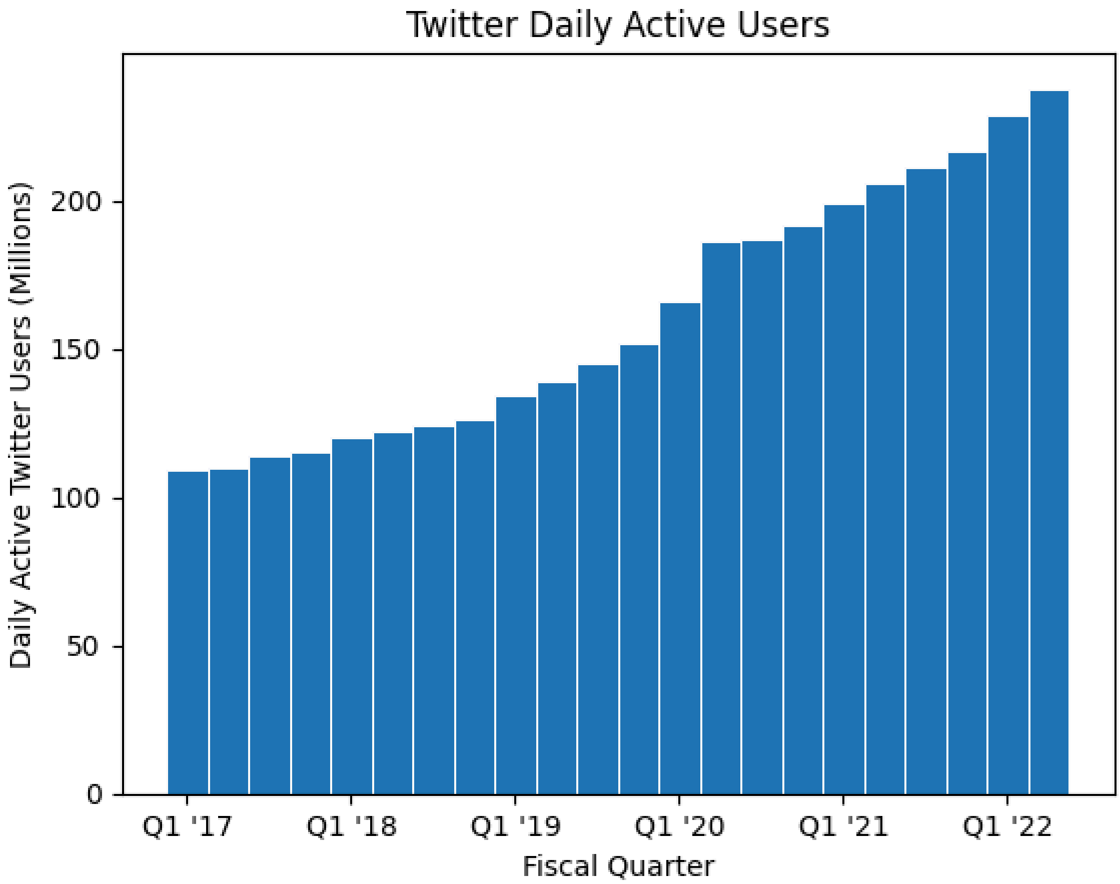
This is the first graph I created with this dataset. It is a basic Line plot that showcases the stock price of Twitter from when it first went public to when it went private. This information is vital for showing us the confidence people have in the stock. With the knowledge that Elon Musk started the process of purchasing Twitter on April 14, 2022 we can see that the stock continued to decrease around that time. This would not be fair to contribute to Elon Musk alone as we can see that the stock price has been on a downward trend since 2021. Then we can see the stock price begin to rise again until late 2022 around the time Elon Musk officially purchased Twitter on October 27th, 2022. Then the stock drops before rising again until the company goes private. Unfortunately because we are lacking data after the company goes private it making it difficult to draw a strong conclusion because we do not know the long term effects of the stock. What we can conclude is that the change helped to reverse the overall decline of the stock from the previous year.



The next graph I decided to do was a scatter plot showcasing the correlation between stock price and volume. This was something I had experience doing from my previous data visualization assignment so I felt comfortable with being able to interpret the results of this graph. I also wanted something that would showcase the effect of the volume because when we just look at the stock price this vital statistic is completely hidden. The line of correlation shows us that the price of the stock generally decreases as the volume of the trades increases. This type of relationship usually represents a strong selling pressure meaning that people are looking to sell off the stock. This represents a decline for the company because if everyone is trying to sell off their stocks the company must not be doing too well. This is consistent with the information we saw on the previous graph showcasing that Twitter had not been doing well in 2021 and 2022. This graph alone does not tell us much about the impact of the rebranding, but if we consider the previous graph it showcases that reversing the downward trend may have been more difficult than previously thought.

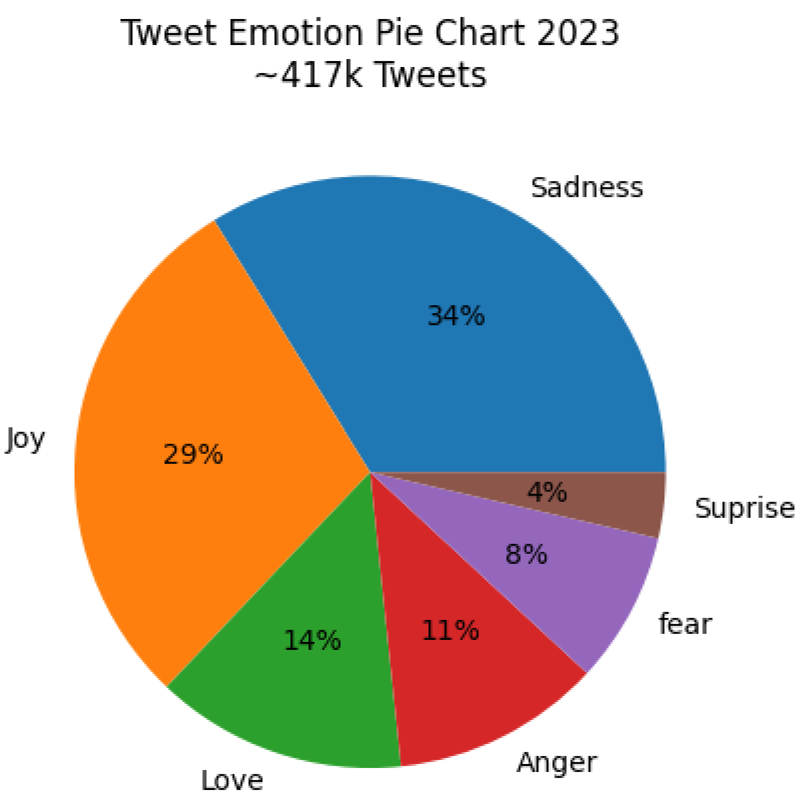
After creating the graphs for the stock market dataset I once again came back to the idea of finding a dataset for active twitter users. I began to expand my search outside of kaggle and was able to find a dataset that looked promising. This is when I learned the dataset was locked behind a paywall. On the same website I was able to find a similar dataset that was free but lacked the quantity the initial dataset had. I still wanted to utilize this dataset even though it wasn't the greatest because it is an extremely important statistic for what I wanted to research.

For this dataset there wasn't much processing I had to do because it just involved the fiscal quarter and the daily active user count of twitter. I did have to drop an irrelevant column because the dataset had an extra index column. For this dataset I decided that a bar graph would best represent it. This is because there aren't as many data values so a line graph wouldn’t show the change overtime as well. The bar graph lets us see the overall trend of the daily active users even without as many data points. One of the issues I ran into with this graph was the date being shown under every single bar making it impossible to read. I was able to figure out how to limit the max ticks making it so matplotlib would calculate the best dates to showcase the data. This resulted in only the first quarter of each year being shown.



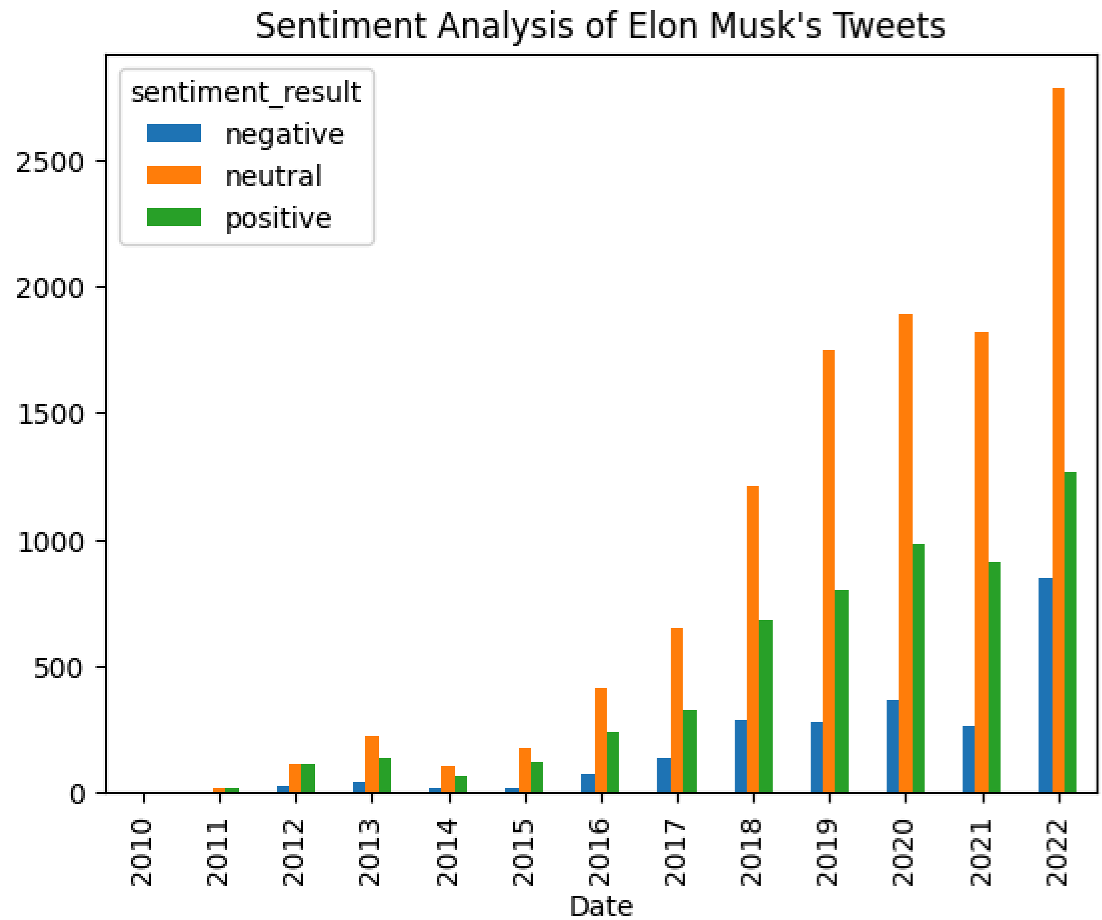
This graph shows us the daily active twitter users for each financial quarter. We can see in this graph that the daily active users are steadily on the rise. This is why I wanted to find a dataset for active users because this is in direct contrast to what we saw on the stock market graphs. This shows that while investors may have not been happy with the company the daily user base was still satisfied and growing. If we look around the time that Elon Musk purchases the platform we can see that there is no strong effect on the change in active daily users. It continues to increase at about the same rate it did before. In this case the rebranding did not have a strong effect on this statistic. Unfortunately the statistic does not extend further because we are unable to see if further changes affected the daily active user count.

The next dataset I investigated was a sentimental analysis that had been done on a sample of around 417k tweets during 2023. This analysis broke the content of the tweet down into one of six emotions: sadness, joy, love, anger, fear, and surprise. Since this analysis was done after the rebranding it can give us a look into the user base and the emotions that they are feeling. In an ideal world I would have liked to have a similar study done before the rebranding I could have compared the values to. Unfortunately I had to work with what data was available. This dataset had a little more processing that had to be done compared to the previous ones. The first thing I did was drop an extra index that had been included in the dataset and the actual text of the tweets that had been used. For the next step I had realized that I wanted to use a pie chart to represent this data, but in the dataset each of the emotions were linked to a numerical value. To fix this I created a dictionary where the key was the numerical value and the value was a string of which emotion it represented. Then using this dictionary I added a new column to the DataFrame by mapping the dictionary with the column of numerical values in the DataFrame. This allowed me to have a column of strings for the emotions making it easier to turn into a pie chart. I then learned that I needed to count each category before passing it to the pie chart function to ensure it worked correctly. After that I created the labels for the pie chart and made sure to include percentages on the pie chart.

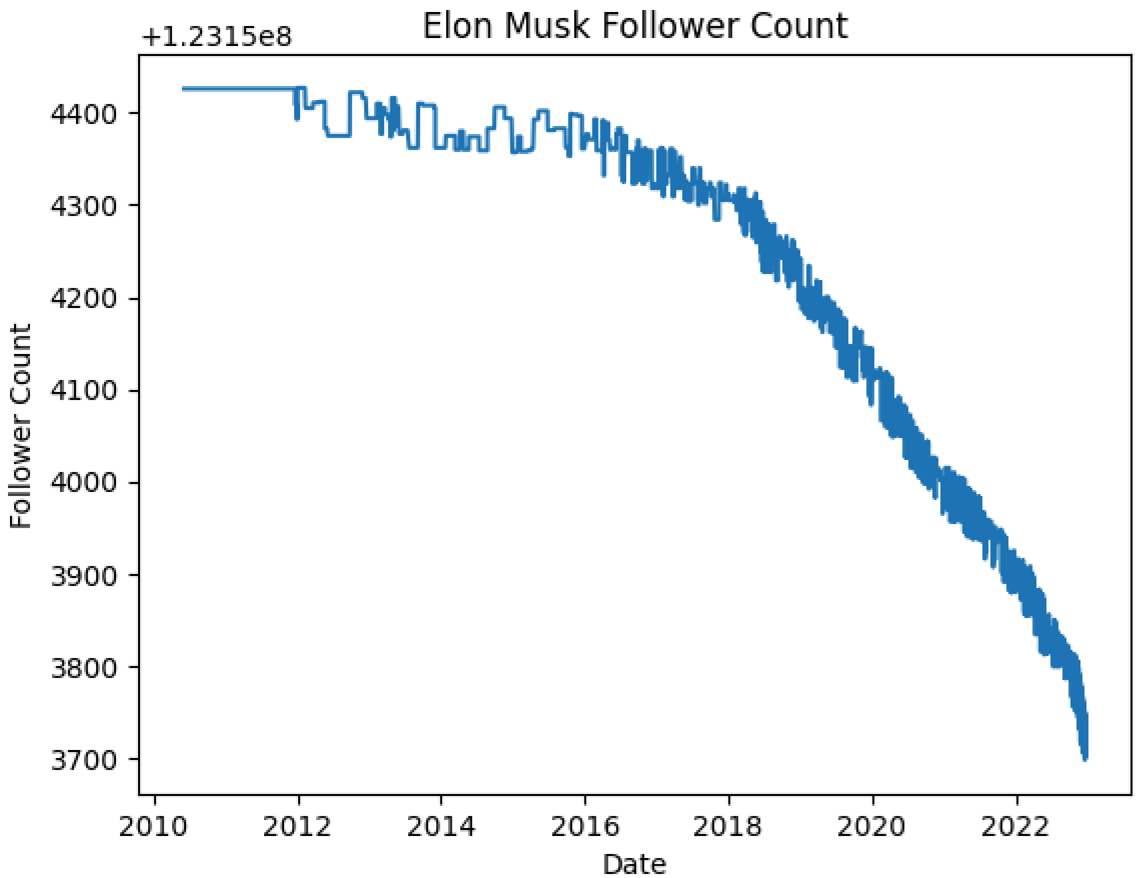


This pie chart shows us that the most common emotion to feel in 2023 was sadness. This emotion is closely followed with joy. The rest of the emotions make a much smaller portion when compared to sadness and joy. It is important to keep in mind that this data focuses on all tweets and not just people's feelings about the rebranding. This means that the emotions showcased here could just represent what happened in the world in 2023. Still I believe this dataset is useful because the rebranding was a major event that people would have been talking about. It is also important to consider what kind of emotions the platform even attracts in the first place. If sadness was clearly in first place it would be easy to mark this as a negative for the rebranding, but since joy is so closely behind it makes it hard to conclude this. Instead, this pie chart reinforces how the change was viewed to be controversial with many people supporting it and many others condemning it.

The next dataset I investigated was a sentiment analysis conducted on every tweet Elon Musk has ever posted. I thought this dataset could be an interesting insight into the platform because Elon has made himself somewhat of a public figurehead of the company. This dataset by far had the most processing required to make it usable. First off it had several completely unnecessary data points included, most likely because it just included all the metadata included in a tweet. So I went through and dropped all of these columns with one being very annoying to drop because it had a space in front of the name. The next thing I wanted to do was separate out the sentinement column because it has a sentiment result and a numerical value representing it. I tried several methods of separating out a list inside of a DataFrame, but none of them were working. This is when I realized that it was actually a string stopping any of these methods from working. After I realized this I was able to utilize RegEx to extract the sentiment result from the column and create a new column in the DataFrame containing only that. After that I encountered my next hurdle when trying to figure out how to utilize the dataset in a grouped bar graph. I first converted the date to a date recognized by pandas. Then I created a new DataFrame consisting only of the date and the new sentiment result column. After doing some research I was able to learn how to group the DataFrame by the year and then count the amount of each result for that year. Finally by utilizing pandas instead of matplotlib I was able to unstack the DataFrame and create a grouped bar graph.

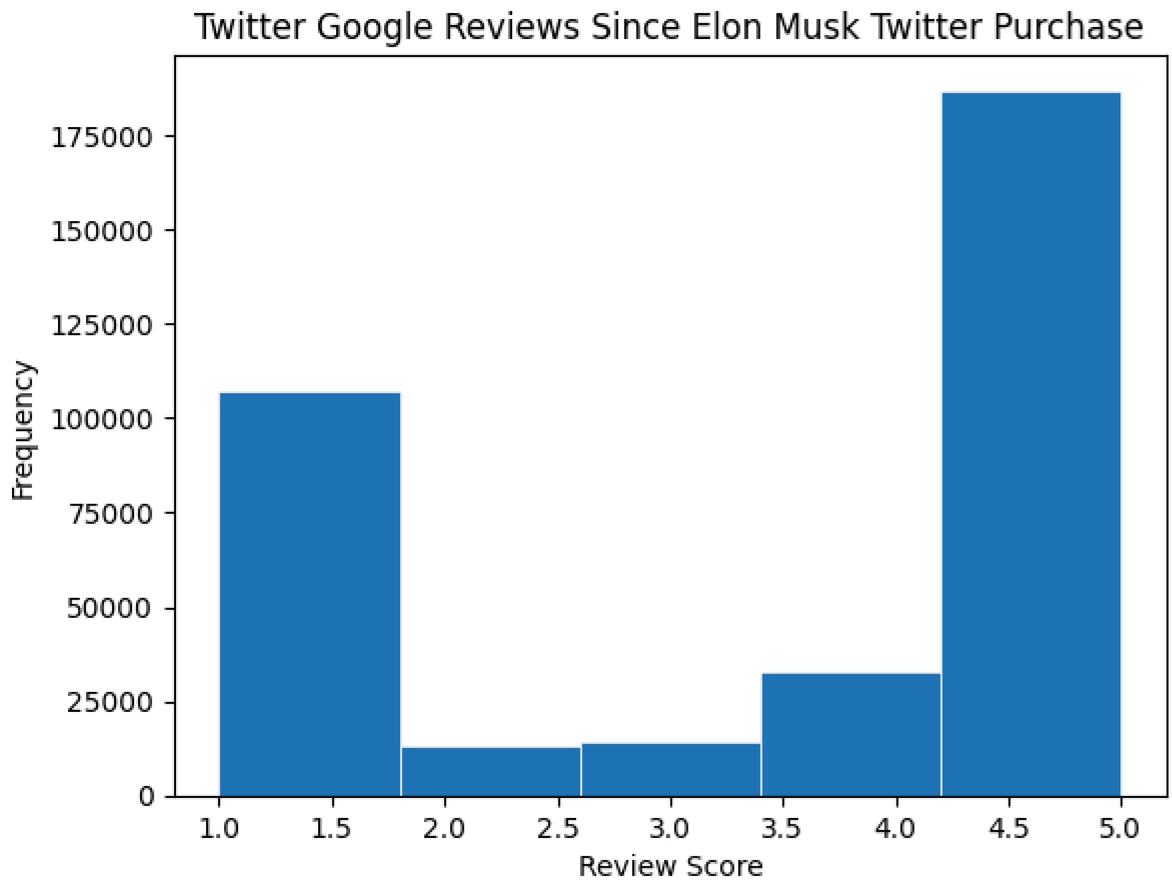


This graph shows us how many times Elon Musk tweeted every year and what the overall impression of the public was for each tweet. The key years I want to focus on are 2021 and 2022. When comparing the two we can see that the amount of negative sentiments in 2022 was over double what it was in 2021. You might expect to see each sentiment change by the same amount if nothing changed except for the amount of tweets. Instead we can see that the positive sentiment increased by a little and the neutral sentiment increased by a medium amount. This shows us that the public opinion of Elon Musk soured greatly in 2022, which is the same year that he bought twitter. This can act as evidence for the fact that the rebranding was controversial and that Elon Musk himself is a controversial figure.



Because there were so many leftover values in this dataset that I didn't utilize I thought it would also be interesting to look at Elon Musk's follower amount and how it changes over time. At first this graph seems like it is a huge downward trend where he is losing many followers, but if you take a closer look this isn't the case. It actually shows that over these years he has only lost around 700 followers out of his millions of followers. Of course this also showcases that his follower count is relatively stagnant since there is no growth and only a slight decline.

The final dataset I used took me a while to find because I had already looked through many of the most popular datasets. After exploring a couple of datasets that didn't pan out I was finally able to find a dataset that compiled 2 million google play store reviews for the twitter app. These reviews were from twitter was first listed on the google play store to late 2023. This dataset is very useful in a way the previous ones haven't been because it consists of the direct users' opinions on the platform. In contrast to many of the previous datasets I decided that the dataset had too many entries because it was taking a long time to load anytime I ran the program. This resulted in me deciding to narrow down the reviews to 2022 and 2023 only so that we could look at the reviews under Elon Musk's ownership of Twitter. This dataset was much easier to process, I first dropped all of the data value that were unnecessary to the graph. Then I made sure that the date was recognized in the date format by pandas. Finally I had to filter the dataset to only include values from 2022 and beyond. Luckily I already knew how to do this because I did something similar on the data visualization project. I decided that a histogram would be the best way to represent the data. This allows for us to easily see the distribution of different review scores.



This graph seems to follow the trend of most of the data we have looked at where it is either very positive or very negative. In this case we can see that the positive reviews have the largest frequency, but this is not enough to come to a positive conclusion. This is because there is quite a high frequency of negative reviews with a very low frequency in between. This showcases that people either rate the app very highly or very lowly and not anything in between. Once again we are lead to the conclusion that the rebranding is very controversial with many people liking it and many people hating it.

In conclusion, the rebranding of twitter is very controversial with no opinion being able to claim a strong majority. I would be interested in revisiting this project in the future if new datasets ever come out. This is because one of my biggest regrets with this project is that the datasets don't extend very far past the point of the rebranding. This greatly hurts the conclusion because many of the changes of the platform hadn’t been implemented yet. Also, in this time many competitors to the platform have been released which could further affect the conclusion. The fact that the platform has made this information so much harder to gather could be indicative of a certain conclusion as well. Overall, I learned a lot from this project about social media and how these companies operate. I also learned a valuable lesson about data analysis which is the fact that the biggest limitation is the data available to you.