Comparing User Sentiment Before and After Events

Problem Statement

Companies around the world are focused on building a brand and earning the favor of the public. What opportunities exist for companies to use real-time customer feedback to analyze the effects of changes in the company or the product or service that it provides? Can tweets be used to analyze changes in user sentiment before and after certain changes are made?

Context

In an environment where people can express their opinions on any topic to the entire world in a matter of seconds, companies have adapted and are interacting with the public on social media. Many individual users react to company news such as CEO changes, new advertisements, stock prices, and just about anything else related to the company. Specifically on Twitter, users often reference companies when complaining about products, praising good service, or speculating about investment. For companies who want to analyze the effects of key company events this is a great opportunity to look at user sentiment before and after the event to see if a significant change has occurred. This would enable companies to monitor the attitude of the general public towards them and could potentially help inform their decisions in the future.

For this project, user sentiment before and after Tesla announced the 2021 quarter four (Q4) earnings will be analyzed. The announcement was made on Wednesday January 26, 2022 after market close.

Data Wrangling

Two sets of data were collected using the Twitter API v2. Each dataset started with approximately 40,000 tweets referencing Tesla. The first dataset includes tweets from January 21, 2022 and the second dataset includes tweets from January 27, 2022.

Due to the messy nature of text data, each dataset required a lot of cleaning. A quick, albeit not all-inclusive, summary of the cleaning process is:

- Removing usernames (e.g., @user1234)
- Removing symbols (e.g., #$%^ etc.)
- Removing leading, trailing, and extra spaces
- Removing emojis
- Removing duplicate tweets

Data processing was also performed as part of data wrangling. Again, a quick summary of the data processing performed is:

- Removing stopwords (e.g., in, a, the, by, etc.)
- Adding a ‘Subjectivity’ feature
- Adding a ‘Polarity’ feature

Exploratory Data Analysis

In order to understand the tweets in each dataset, basic numerical descriptors as well as visual aids were used. In each dataset the distribution for tweet subjectivity and tweet polarity was very similar (see Figures 1 and 2).

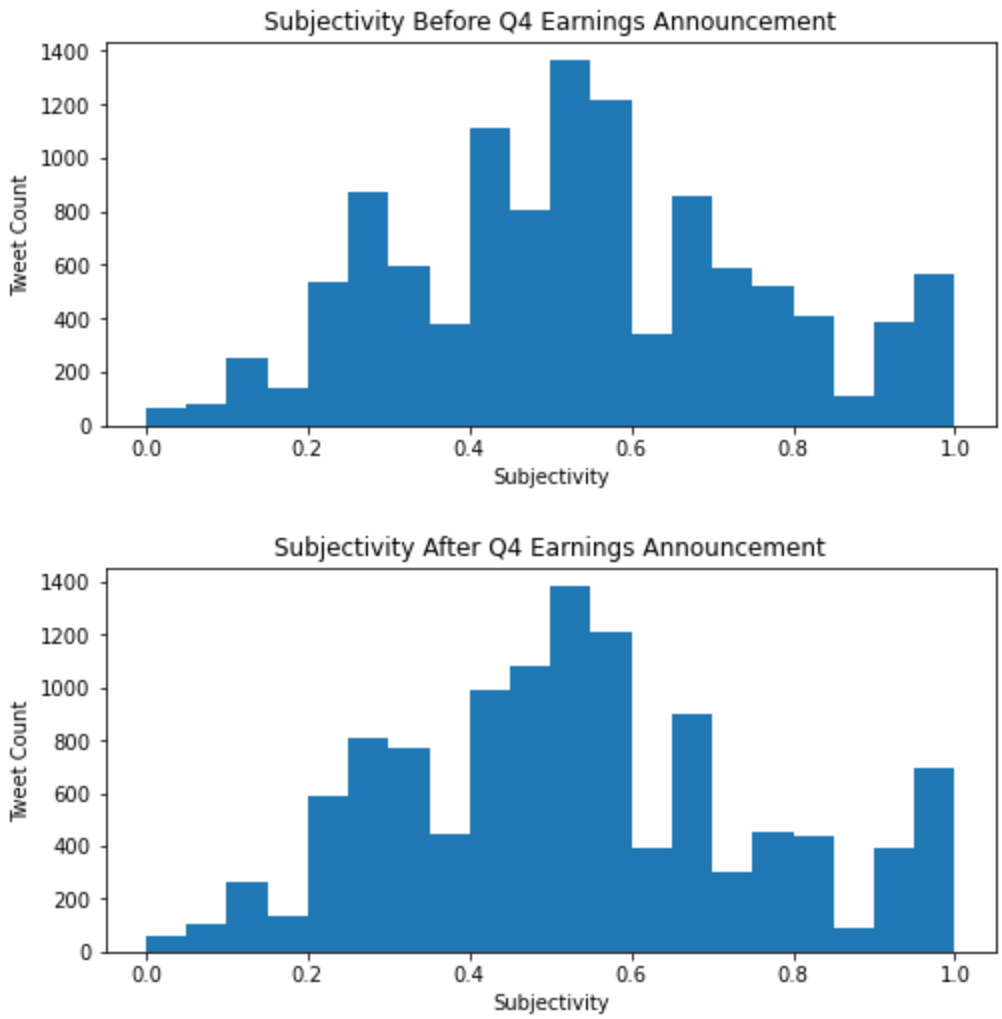

Figure 1: Distribution of tweet subjectivity before and after Tesla announced the 2021 Q4 earnings.

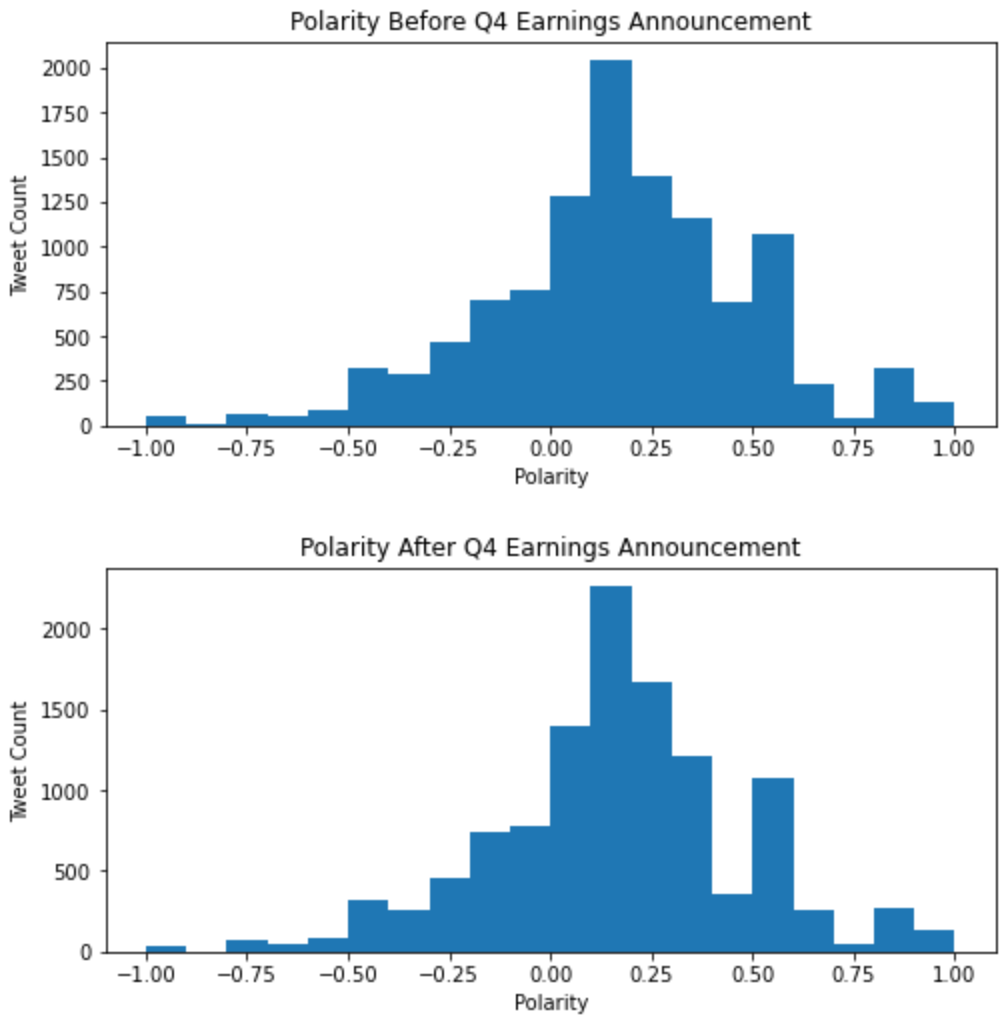

Figure 2: Distribution of tweet polarity before and after Tesla announced the 2021 Q4 earnings.

Both the subjectivity and polarity were slightly skewed positive which seems to indicate that there were more positive tweets about Tesla than negative tweets. However, the distribution is approximately normal for each variable in both datasets.

To visually represent the tweets, word clouds were made for the two datasets. Each word cloud shows common words in the tweets, with larger words being more common than smaller words (see Figures 3 and 4).

Figure 3: Word cloud for tweets collected before Tesla announced the 2021 Q4 earnings.

Figure 4: Word cloud for tweets collected before Tesla announced the 2021 Q4 earnings.

Analysis

To determine whether or not the 2021 Q4 earnings announcement had an effect on user sentiment, a two sample t-test was performed. A two sample t-test is a method used to test whether the unknown population means of two groups are equal or not. In this case, the two groups are users tweeting about Tesla before and after the 2021 Q4 earnings announcement.

Null hypothesis: The 2021 Q4 earnings announcement had no effect on user sentiment (population mean before is not different than population mean after).

Alternative hypothesis: The 2021 Q4 earnings announcement had an effect on user sentiment (population mean before is different than population mean after).

Summary Statistics

Subjectivity	Before	After
Mean	0.532215	0.526675
Standard Deviation	0.223828	0.226634

Polarity	Before	After
Mean	0.178973	0.171599
Standard Deviation	0.322279	0.310142

Since user sentiment is a combination of both subjectivity and polarity, the t-test was performed with each variable using the process shown below. It is assumed that the datasets are representative of the overall population due to the high number of samples in each one.

Difference = u_1 - u_2

$$s_p = \sqrt{\frac{((n_1 - 1){s_1}^2) + ((n_2 - 1){s_2}^2)}{n_1 + n_2 - 2}}$$

$$t = \frac{Difference}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Using the equations above for each of the two variables, the two t-values were calculated to be 1.85 and 1.76 respectively. The critical t-value for $\alpha=0.05$ is $t=1.96$.

Conclusion

For both test variables, the conclusion is that the null hypothesis could not be rejected. To put it in terms of the null hypothesis, there is not significant evidence that the population mean for either subjectivity or polarity changed as a result of Tesla's 2021 Q4 earnings announcement.

This project served to show that meaningful analysis can be done using real time data from Twitter. This type of data could easily be monitored so companies could keep an eye on how the public was reacting to changes. For example, a company with a new product and marketing campaign might wish to see if people are tweeting more positive things about the company or to see if tweets mentioning that specific product tend to be negative or positive. Using the process described in the previous sections, the company could determine whether or not a significant change had occurred in user sentiment.