Sentiment Mining Project

MGIS:650

Prof: Jing Tang

Team-2

Akash Taploo, Sameer Bhattarai, Raj Jasani and Sneha Mondal

Contents

1. Introduction

Twitter is a microblogging platform where users may share and express their opinions on various subjects or post messages. The Sentiment Analysis of Twitter data has received a lot of attention in recent times. This project involves gathering data from Twitter API, text mining and sentiment analysis using Tableau and R studio.

Objective:

The main objective behind this project is to find sentiments of people towards Tesla using its twitter data. The company is famous for its Automobile designs, Battery research and growing autonomous driving technology. we did our analysis on these things as these are the major revenue streams for the company.

About the company:

Tesla, Inc. is an American electric vehicle and clean energy company founded in 2003. The company is named after the renowned physicist and inventor Nikola Tesla.

Tesla designs, manufactures, and sells electric cars, energy storage systems, and solar products. Its vehicles include the Model S, Model X, Model 3, Model Y, and the Cybertruck, all of which have gained popularity for their sleek design, high performance, and environmentally friendly features. Tesla is also involved in battery research and has built one of the world's largest battery factories. In addition to producing electric vehicles, Tesla is also focused on developing autonomous driving technology, which has the potential to revolutionize the automotive industry.

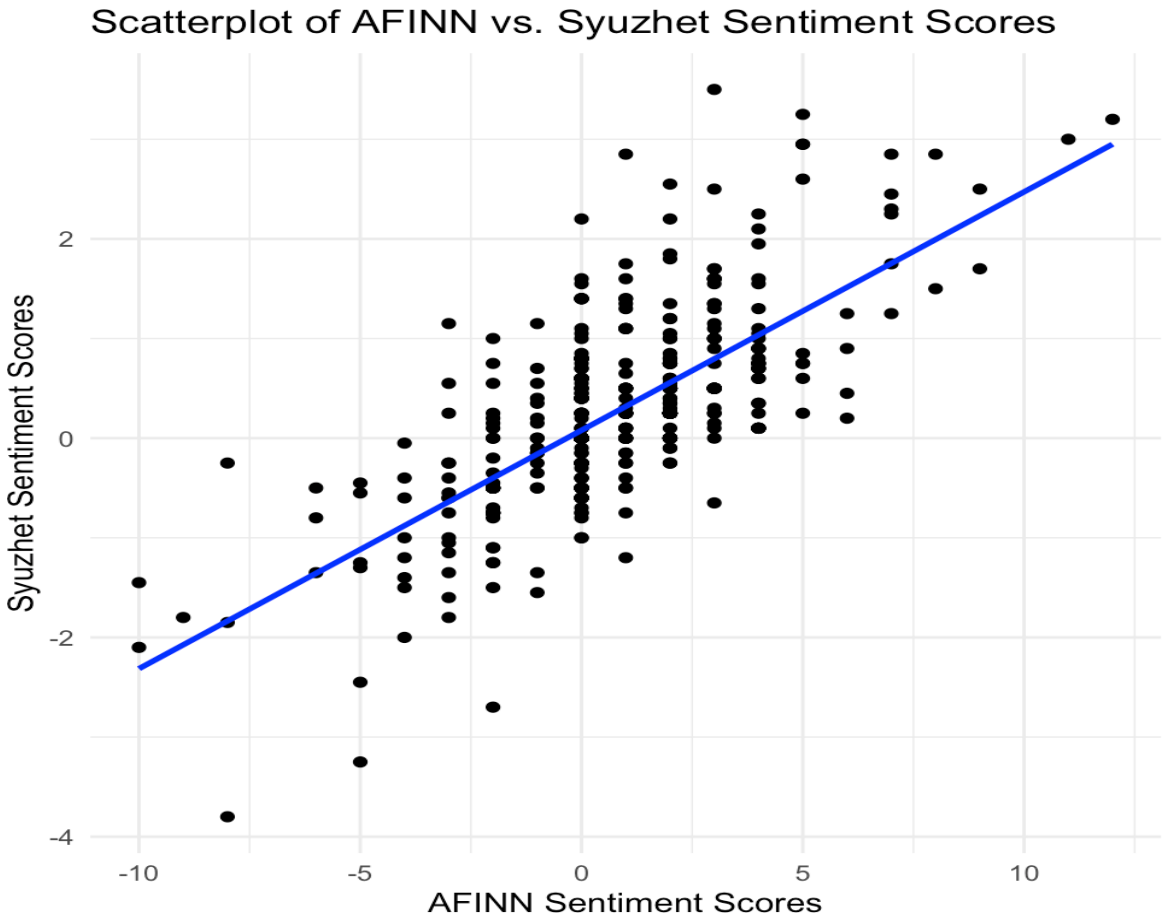
Keywords: Sentiment Analysis, Twitter API, Tableau, R Studio

2. Executive Summary

In this project we aim to showcase the sentiments of the general public towards our company “Tesla” using data extracted by Twitter API. Text mining and Sentiment analysis are the modern methods and tools to measure public sentiment from data. Information received through the Twitter API, analyzed in R, and displayed in Tableau software served as the basis for the analysis. The first step in this project is collecting data from Twitter API using R. To do this we collected 1,000 tweets from any source containing the word “Tesla” and then get 1,000 tweets from the official handle of Tesla which is “@Tesla”. we do not include retweets as a cleaning tool, restrict the data to English language only and remove the duplicates. Also, we used 4 key words related to Tesla to understand the data in an efficient manner. Then we used R to calculate the sentiment score of each tweet. If the sentiment score is positive, then the tweet is good and if the sentiment score is negative then the tweet is bad for the company’s image. The better the sentiment score, the more positive will be score. We used three different packages (Syuzhet, Afinn and NRC) to get the sentiment scores and emotions for the data and used different packages to combine, filter, clean and analyze the data in terms of sentiments and scores. At last, we used Tableau and R to visualize the data in the form of different scatter plots, bar graphs and packed bubbles in order to present the data in an effective manner. The word cloud has been prepared using tableau as well as R and we filter the word cloud using various codes and functions in both the tools.

3. Correlation Analysis

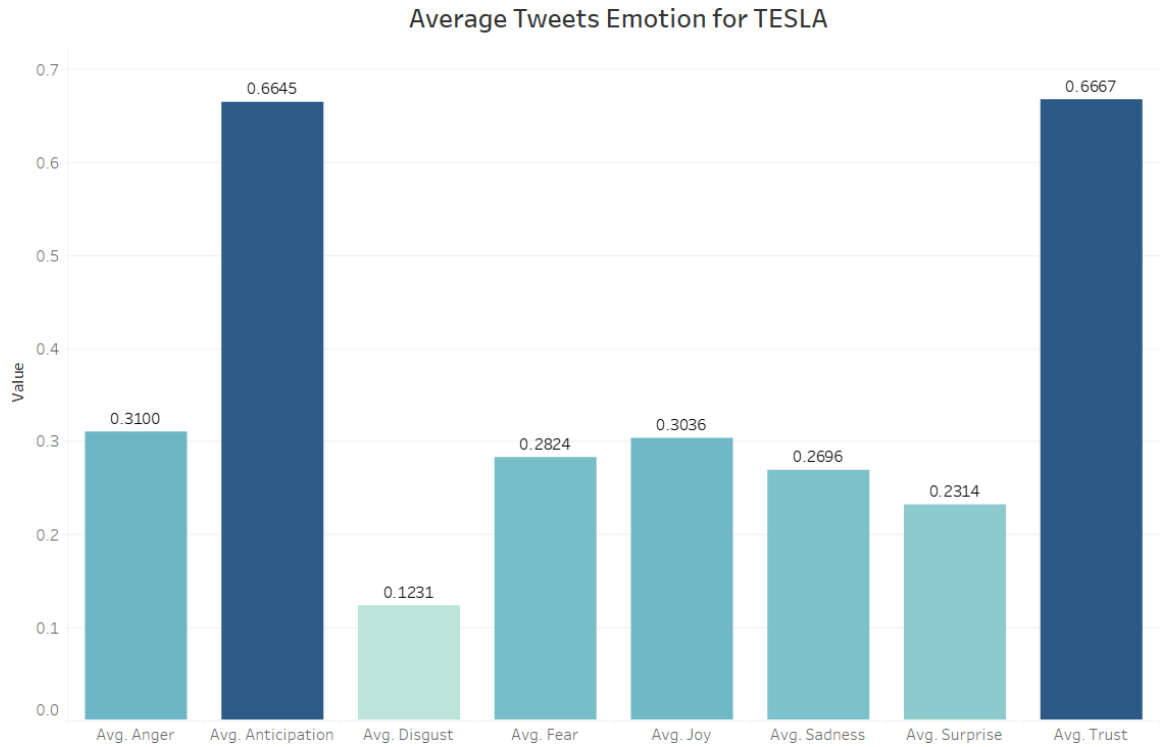
We used Afinn and Syuzhet, the two most well-liked sentiment analysis techniques in R, to determine the sentiment score for each tweet. The word's meaning and context determine whether the sentiment score is positive (1, 2, 3) or negative (-1, -2,-3) for that word. "Afinn" is the dependent variable and "Syuzhet" is the independent variable, we can detect a positive correlation between the two variables. The trend line displays the ideal relationship between two variables, while the scatter plots display the various data spread over the graph. Additionally, the fact that the p value is smaller than 0.009 indicates a significant correlation between the two variables. We can see numerous outliers in the plot that reflect both excessively positive and excessively negative public attitude.



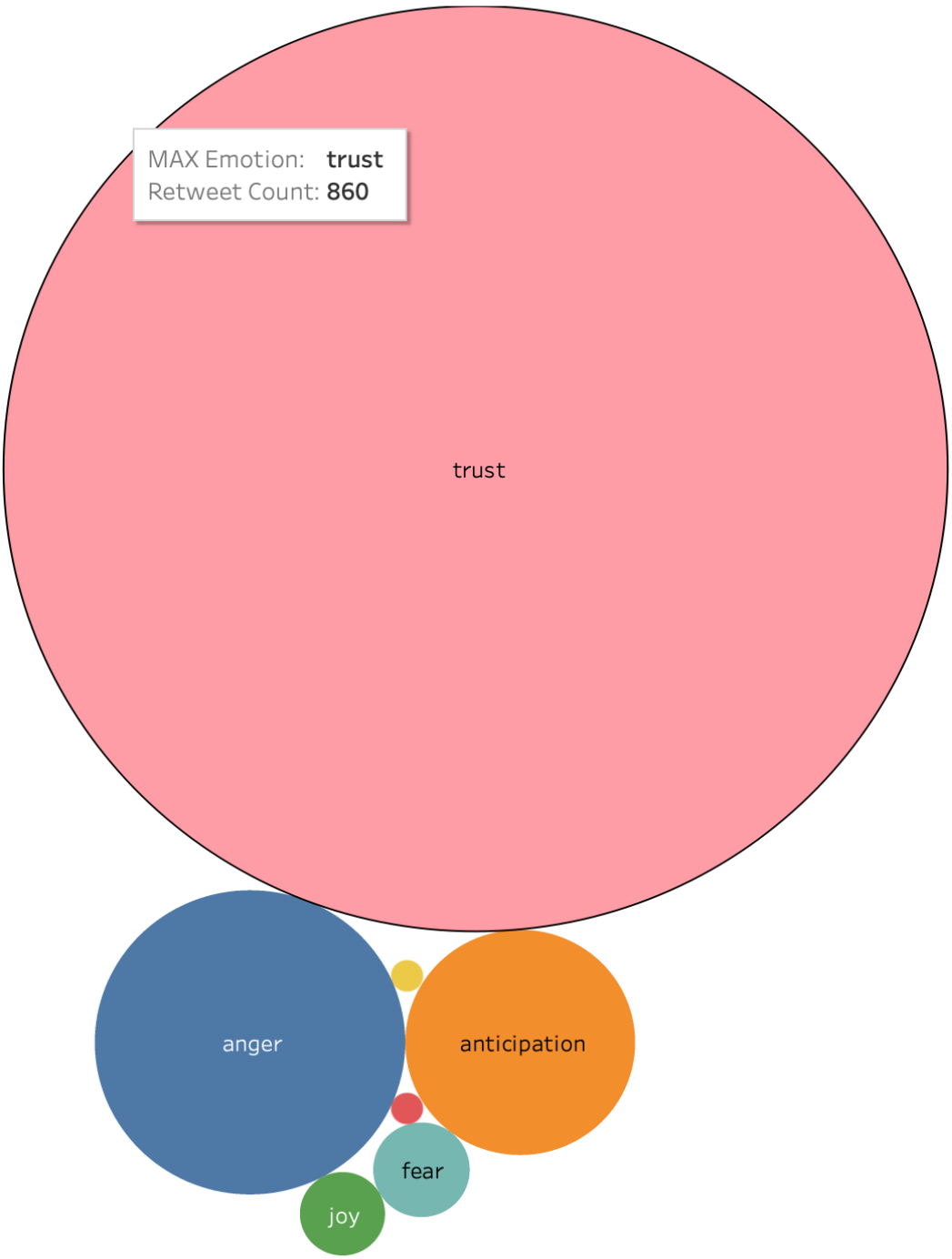
4. Emotions

(A) Emotions associated with Tesla Tweets

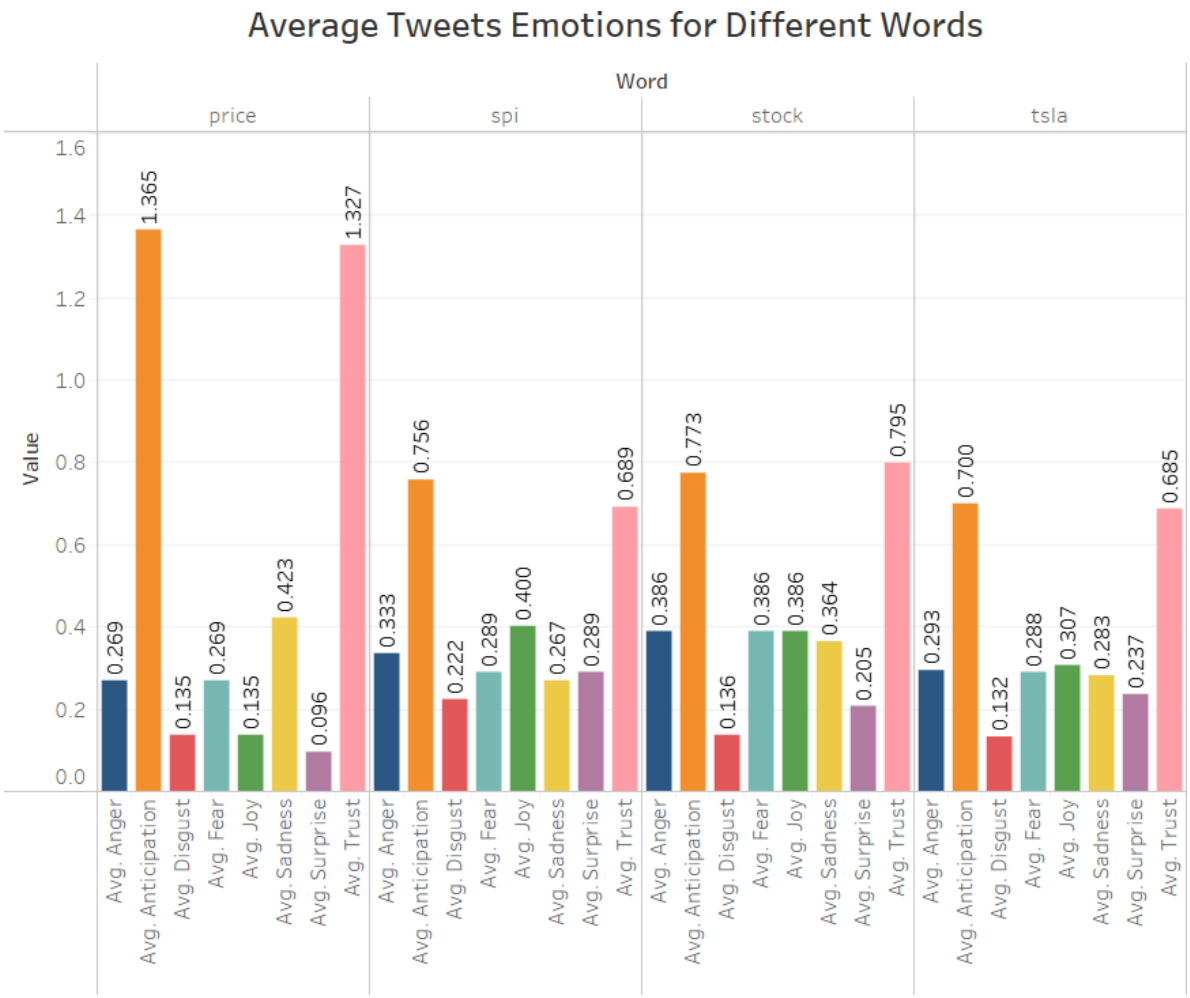
Using a predetermined set of emotions—Trust, Joy, Anticipation, Surprise, Sadness, Fear, Anger, and Disgust—R & its packages categorized the company's tweets. This graph displays the many emotions connected to the company's tweets. The average of positive emotions is clearly significantly higher than the average of negative emotions in this instance. The trust score has the highest average of the emotions (0.667), indicating that customers feel a sense of loyalty & trustworthiness to the business, while the disgust score has the lowest average (0.1231), indicating that some customers were dissatisfied with their order or delivery. The most prevalent negative emotion in tweets is Anger (0.3100), which is roughly half of the most prevalent positive feeling (Trust). This indicates that the positive emotions are considerably more prevalent in tweets than the negative ones, which builds brand recognition and goodwill for the business.



Below we have attached the packed bubble visualization for the emotions associated with Tesla:



(B) Emotions associated with different words

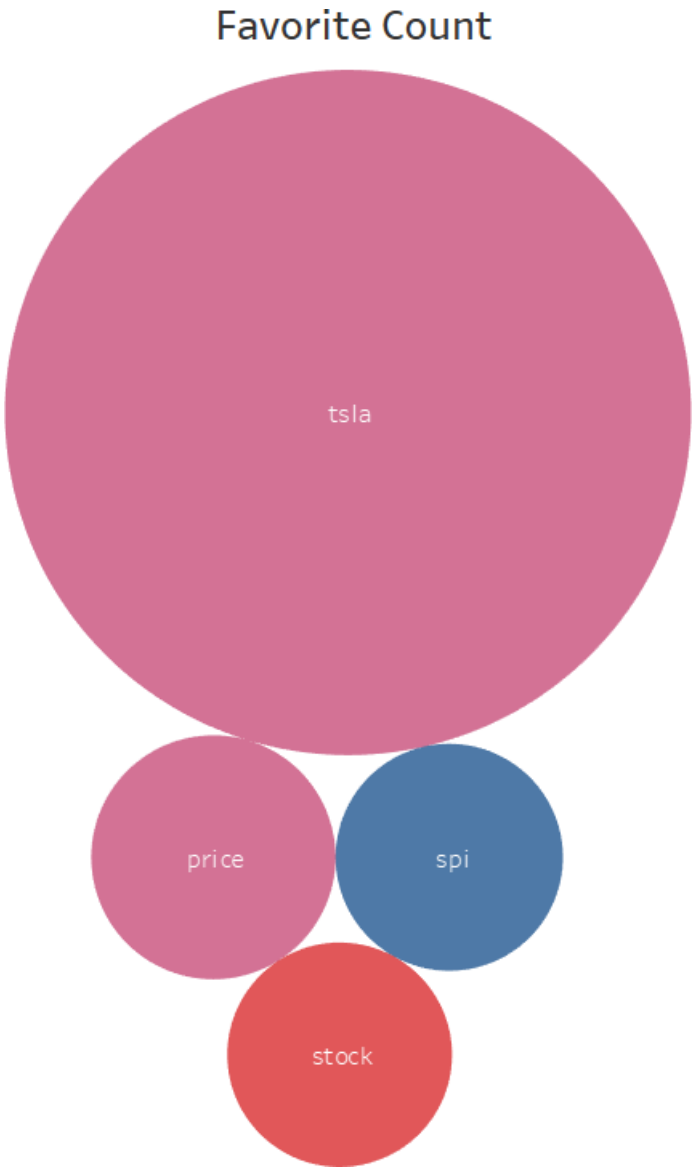


Price, Spi, Stock, and Tesla, are the four words that we use in R to connect to the company in all aspects, and they represent what many people tweet about and feel. Next, we examine the feelings evoked by these words. The highest average for all four terms is either "Anticipation" or "Trust," with "Disgust" or "Surprise" having the lowest average. However, if you compare the averages of the two types of emotions (positive &negative), you will notice that the positive ones are considerably more common than the negative ones. Additionally, the high averages of trust, joy, and anticipation demonstrate a favorable effect

of the business on its clients' minds. Additionally, the word with the greatest average across all words is "Price" (Avg Anticipation: 1.365), which shows that customers are quite excited with their Tesla purchases and that this is reflected in their tweets. The phrase with the lowest average across all words is also "Price" (Avg surprise:0.096). Sadness for price (0.423) is the negative feeling with the highest average, indicating that consumers are substantially more saddened by the company's prices than by any other factor we looked at. So, this might be a place where the business can look at and take the necessary action to improve its image.

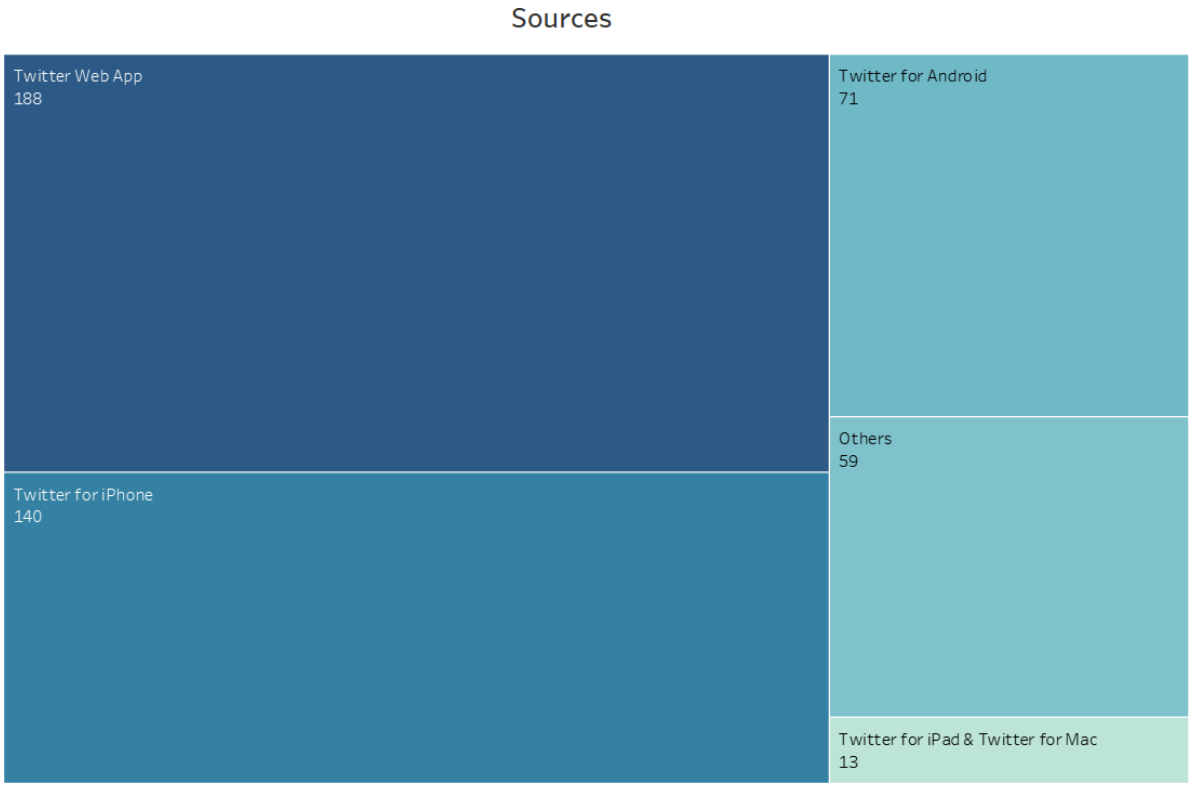
5. Favorites Count

We analyzed the frequency of favorite words used in tweets, filtering out those with a count of less than 50 to improve the clarity of the visualization. As a result, we excluded the words "tsla," "price," "spi," and "stock," with "tsla" being the most commonly used word with a count of 410, and "stock" being the least commonly used with a count of 44.



6. Sources of tweets

This matrix displays the company's tweets' many sources. There are more than 400 sources of tweets for the company, but only a select number are significant ones, with the others falling under the "Other" category. Twitter web App and Twitter for iPhone are the two most significant sources, with a count of 188 and 140; Twitter for iPad and Mac has the lowest count of 13. Although the "Other" category has some sources that may have a count lower than 13, we have grouped them together for presentational reasons. The most popular source for any business is the Twitter web app.



7. Word Cloud

The Tesla word cloud has been made using Tableau. Only words which have a count of more than 10 can be seen in the cloud and irrelevant words and numbers have been removed to make it look more presentable. The words with the highest count are “Tsla, price week and top”.

Using R and its packages, we created a positive and negative word cloud that displays positive words in blue and negative phrases in grey. The most positive terms used in tweets are "top, free, good, like" whereas the most negative ones are "loss, hard, miss, break." We employ several tools to filter out the data.

End of the report