An Analysis of the Associations between Twitter Sentiment of Autonomous Vehicles and Tesla Model Releases

By: Darren Jiang, Yining Yang, Yueqi Liao, and Zibin Gao

Question

This paper aims to find whether or not a potential association exists between the sentiment of self-driving cars on social media websites, like Twitter, and the Tesla model releases. Our hypothesis is that there would be a positive association, meaning that the sentiment scores of autonomous vehicles are expected to be positive, or more positive around the months where a new Tesla model is released compared to the months when there isn't aren't any.

Justification

The basis of our hypothesis rests on the idea that Tesla's branding centers around developing fully autonomous cars. It's possible, then, that advances of Tesla vehicle models could more broadly affect the sentiment of self-driving cars, and releases of improved self-driving capabilities may positively drive public sentiment. The positive hype of newly-released Tesla cars may spill over to have a positive attitude associated with self-driving cars in general.

Background Information

There have been many previous studies that look at the general public's opinion of self-driving cars. The 2018 Ipsos global poll found generally positive attitudes towards self-driving cars worldwide, the 2019 Microsoft questionnaire confirmed broadly positive views of AVs, and a 2019 AAA survey, which indicated that 71% of Americans claim to be afraid to ride in a self-driving car, are just a few examples. Yet, these studies tend to only report on what the general public's attitudes toward self-driving cars are, as opposed to whether those attitudes were affected by the releases or announcements of advancements in self-driving cars. There are few studies that look at possible relationships between changes in the sentiment of self-driving cars and breakthroughs in commercial self-driving car technology.

In general, we can say that public sentiment as reported by surveys and polls demonstrate broadly positive sentiments towards self-driving cars, though it's important to note that such reports can be contradictory and greatly dependent on the demographics of the people surveyed.

Data

Ideally, we would have a dataset that contains information on tweets about self-driving cars within a certain time period, likely 2012-2021. The dataset would track the text of the tweet and the date that the tweet was made. Sentiment analysis would also be performed on a clean version of each tweet to give us a polarity score, telling us how positive or negative each tweet was. To get observations that fall within the time period, we'll sample an arbitrary number of tweets each day, for an arbitrary amount of days per month, for each year in our time period. The exact number of tweets will vary based on what limits we choose and how many tweets were actually made during the time period that we're sampling from, but we expect to get roughly 57000-60000 observations.

Additionally, we need a dataset that contains dates for Tesla model releases, in order to see if there is an association between change in polarity score and those model releases. These were obtained via several Google searches.

The variables collected would be quantitative variables, those being the date of each tweet and the polarity score of each tweet. These variables allow us to track potential trends or changes in polarity score as time goes on, and let us see if there are certain spikes/dips in sentiment that correlate with Tesla model releases. If there is an association between Tesla model releases and the sentiment of self-driving cars, the polarity scores of the tweets should be statistically different compared to the polarity scores of times where there aren't Tesla model releases.

While there are databases that contain tweets about self-driving cars (like this one), they tend to be limited to a specific date. Therefore, we will have to collect the data for this project ourselves. Due to the number of tweets we were trying to get and the rate limitation that the Twitter API imposes, we decided to web scrape Twitter to get the tweets. We wrote a python script to do this (web_scrapping.py). The program gave us a table of tweets per year, and we just concatenated every yearly table together to form the dataset (tweets.csv). After cleaning the text of each tweet, we ran sentiment analysis using the NLTK library, which gave us a polarity score for each tweet.

The result was a dataset of 57215 tweets, which matched the ideal dataset that we wanted. Due to time and memory constraints, we only drew 21 tweets for each day. We also took values from the first 28 days of every month to avoid complications with months of differing dates.

Ethical Considerations

1. Data Collection:

The ethical consideration that must be taken care of during the data collection process is that we have to have the right to use the websites, articles, and sources we found for our analysis. For instance, the websites we are looking at are public resources that contain information on released Tesla models and tweets about them. Since the tweets are posted onto a public platform, we do not necessarily need to ask for the users' permission before using their tweets to analyze the sentiments. It is also important to avoid collection bias in the data collection process. We should include impersonal tweets that express the users' opinions on self-driving cars and try not to use misleading tweets that are not objective. Lastly, we must limit the exposure of the users' personally identifiable information and ensure their privacy. We are only analyzing the content of the tweets, thus the user names and any related personal information that might be listed on their main pages such as their locations, ages, genders must be well-protected. When collecting data, information such as usernames, location, age, etc. will not be gathered. The only data collected is the content of the tweet and the datetime the tweet was made.

2. Data Storage:

There are also many things that need to be considered during data storage. All users whose tweets are used will have the right to have access to his or her own personal information. Besides, data retention requires us to determine if the data needs to be deleted after being used. Since we collect data by ourselves and make our own dataset, we know that there isn't any sensitive information about specific, identifiable people. Therefore, we have the right to delete the data after we are done with our analysis, and data about the tweets will no longer be used. To preserve privacy, the raw data will be deleted after the analysis is completed. Additionally, the cleaned version of the content of the tweet will not have any mentions of other users, thus preserving privacy.

3. Analysis:

During analysis, our final conclusion must be that there is an association between the releases of Tesla models and the sentiments expressed on Twitter. In order to address the blind spots in data analysis, we must include all kinds of tweets from different genders, countries, and ages to enrich the variety of our data. We also need to check if the data contains any sources of bias. For example, some negative judgments about self-driving cars may come from their rivals. Thus, we must be careful when looking at the tweets to pick out the most helpful ones and mitigate the influence of bias on analysis. Secondly, we should check if the visualizations and representations honestly reflect the analysis. The representations should reflect the result of the analysis as accurately as possible. This can be achieved by being careful when making the visualizations and spending time checking every detail listed. Thirdly, there should not be any unnecessary procedures. The protection of PII should be strictly guaranteed and re-identifying the original user that made a specific tweet should be made impossible with the clean versions of data. Lastly, auditability refers to the process of producing analysis that is well documented and reproducible, that is, other groups must be able to draw the same conclusions using the data we generate.

4. Modeling:

For modeling, we shall ensure that our model is not relied upon or related to discriminatory variables and proxies of any kind. Possible discriminatory proxies for our model include race, gender, age, etc. The data collected should fairly represent different groups of people. Data should be displayed in a way that shows accuracy and is easy for people to understand from an unbiased point of view. Since the only criteria for collecting the data is whether or not the tweet mentions self-driving cars, we shouldn't have any issues disproportionately sampling a particular race, gender, etc. beyond the demographics of a typical Twitter user. How we represent the data on self-driving cars should not be biased towards a more positive or negative perspective, meaning that figures should have proportionate scaling and sizes, and shouldn't be visually misleading. The ends of the scales when measuring polarity scores should be set to -1 and 1, so that the data doesn't look more positive or negative than it really is and small differences don't get magnified. Moreover, we shall take into consideration whether the effects of the communication biases are properly addressed. Since we do not have any stakeholders in our research, we have to make sure that potential readers should be well informed about the biases and limitations of our model. Something like a disclaimer, or a notice in the discussion section of the project report would need to be made.

5. Deployment

Finally, in the deployment process, we have to first check our results are unbiased and precise before uploading. Moreover, if people discover any errors, we would fix and renew our conclusion immediately. We also should re-test our model with new information and new parameters to ensure the model remains fair over time, such as when Tesla releases the new model, and people post anything about their new feeling on Twitter. It is important to include redress where there is an organization's plan for responses if the user is harmed by the results. Furthermore, we should be able to withdraw results at any time if needed as well as have our data removed from the dataset in emergency circumstances. Finally, we should minimize unintended use and abuse of the model. In order to avoid people from drawing wrong conclusions from our model or using the results in an unethical way, we can set limits of authority on our model and make it only accessible to people under permission.

Analysis Proposal

1. **Data Collection (web scraping, APIs, etc.):**

To begin our analysis, we need to **collect data** on tweets about self-driving cars, and the release dates of all Tesla models. While the data about Tesla model release dates can be found fairly easily with a few Google searches, data on tweets are more difficult. Essentially, what we would need was a dataset of tweets about the subject of self-driving cars that go back from 2012 to 2021. To obtain this data, we web-scraped Twitter using a python library called "snscrape". By using snscrape, we can essentially piggyback off of Twitter's own search function and systematically search for any tweets that match a query in a specified time frame. This means that we can search for tweets about self-driving cars that were made on a specific date, and record them onto a list or table. Using this function, we grabbed 21 of the latest tweets of each day, for the first 28 days of each month, for every year from 2012 to 2021. There were some issues with the amount of tweets we were pulling, but they were sidestepped by separating each year into its own dataset and concatenating them back together once the data collection process was complete.

2. **Data Wrangling:**

The next step in our analysis would be to **wrangle the data** to make our analysis easier. We first combined all the tables into a single dataset so that any other operations could be done with all of the data. While query year and month were useful for checking over our data and making sure the web scraping algorithm functioned properly, they were no longer necessary, and could therefore be removed. The time that the tweet was made was also unnecessary (and removed). Since our analysis would be looking at dates, we converted the column datetime to date. The text of the tweet should be cleaned of hashtags, names, and links, and the original raw text should be removed from the dataset. This removes personal identifiers from the data, which further anonymizes the people the authors of the tweet, and prepares the data for sentiment analysis.

date	clean_text	polarity_score
2011-12-31	"seriously, gm it, will! self driving cars... i'm it!"	-0.3147
2011-12-31	bring self driving cars : driving lost cool young americans	0.0

2011-12-31	hey car makers- get self driving cars? want read/take nap car drive itself. let's get this. pls & ty	0.5065
2011-12-31	countdown: ideas change world via self-driving cars intesresting	0.0
2011-12-31	"sebastian thrun: self-driving cars save lives, parking spaces:"	0.4939

Table 1. First 5 entries of the table containing cleaned tweets, the date the tweets were made, and the polarity score of those tweets. *Data were obtained by web scraping Twitter*

3. Text Analysis (Sentiment Analysis, TF-IDF, etc.)

To determine the sentiment of each tweet, we used an NLTK open-sourced package called Vader (Valence Aware Dictionary for sEntiment Reasoning). Vader is an NLP algorithm that utilizes a mix of a sentiment lexicon approach and grammatical rules and syntactic conventions to calculate sentiment polarity. How this works is that Vader has a sentiment lexicon, which is essentially a dictionary that contains a list of sentiment features. This sentiment lexicon contains words, phrases, emoticons, and sentiment-laden acronyms (such as “WTF”) and each feature is rated with a polarity and intensity, which range from “-4 Extremely Negative” to “4 Extremely Positive”. These ratings were given by 10 independent human raters. Any words that aren’t a part of the lexicon are labelled as “0 neutral”. Beyond the sentimental lexicon, Vader also has several heuristic rules that check for punctuation, capitalization, adverbs, and contrastive conjunctions. These are all meant to detect structures that are inherently neutral, but affect the polarity or intensity of the text.

Input	neg	neu	pos	compound
"This computer is a good deal."	0	0.58	0.42	0.44
"This computer is a very good deal."	0	0.61	0.39	0.49
"This computer is a very good deal!!"	0	0.57	0.43	0.58
This computer is a very good deal!! :-)"	0	0.44	0.56	0.74
This computer is a VERY good deal!! :-)"	0	0.393	0.61	0.82

Table 2. Example of how Vader interprets different structures that modifies the sentiment of the text*. Neg indicates the percentage of the statement that has a negative polarity score, neu indicates the percentage of the statement that has a neutral polarity score, and pos indicates the percentage of the statement that has a positive polarity score. Compound is the text’s overall polarity score.*
Source: https://medium.com/ro-codes/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b

In practice, what all this means is that when given some text, Vader breaks up that text into its individual words and punctuation. Each sentimental feature -- words, phrases, abbreviations, or emoticons -- that Vader can recognize are given a polarity score. Those polarity scores are then modified by whatever syntactical structure that Vader can detect, and then added together to get some sum that we will call X. X is then normalized to the final compound polarity score, where the most negative polarity is -1 and the most positive polarity is 1, using the following formula:

$$\frac{x}{\sqrt{x^2 + \alpha}}$$

In Vader, alpha is the maximum expected value of x, and has a default value of 15.

4. Descriptive & Exploratory Data Analysis:

Once data wrangling is finished, we need to describe and explore the dataset that we have. The essential part of this step is to obtain basic statistical information such as min-value, max-value, mean, standard deviation, and check for possible correlations of data variables. Visualizations would also help us describe and understand the data that we have just collected

We first made a table with basic statistical information about the polarity scores of all collected tweets:

	Polarity score of all tweets
Count	57215
Mean	0.074
STD	0.343
Min	-0.970
25%	0.000
50%	0.000
75%	0.340
Max	0.976

Table 3. Basic descriptive statistics of the polarity scores of all tweets in our dataset. *Values are rounded to three decimal places*

	Annual average polarity score (new Tesla model releasing)	Annual average polarity score (No new Tesla model releasing)
Min	0.047633	0.057466
Mean	0.077682	0.081920
Max	0.111632	0.137140
S.D	0.030531	0.026732

Table 4. Basic descriptive statistics of average annual polarity scores of years with Tesla model releases and years without Tesla model releases. *Values are rounded to six decimal places*

Much of our data seemed to be concentrated at 0.000, implying that most of the tweets were completely neutral. To see if this was the case, we plotted a histogram of the polarity scores:

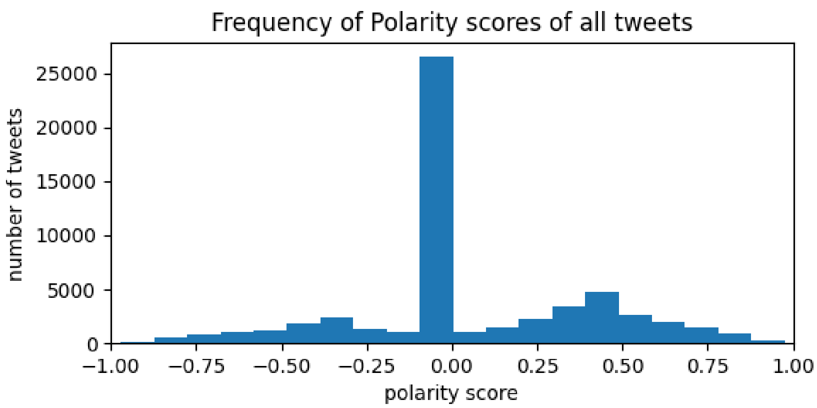

Figure 1. Frequency of Polarity scores of all tweets.
Tweets were grouped by polarity score and placed into bins of 0.1. Values that were the same as the separator (-0.2, 0, 0.5, etc.) were placed in the bin with smaller values.

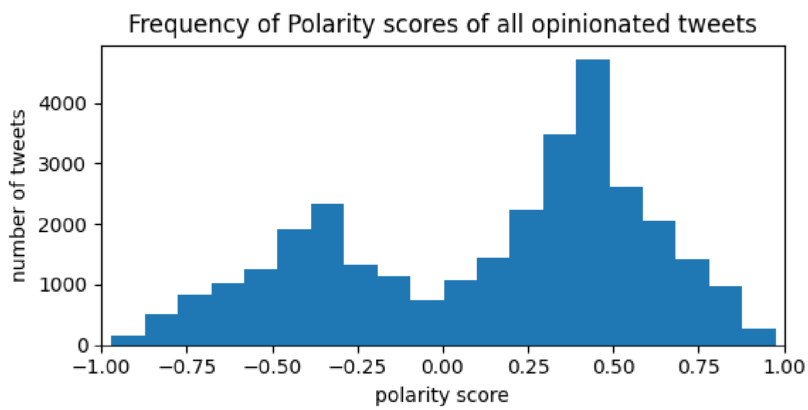

Figure 2. Frequency of Polarity scores of all opinionated tweets.
This histogram was the exact same as Figure 1, however, all tweets with a polarity score of 0 were removed.

From figure 1, we can see that the vast majority of tweets fell within the “-0.1 -- 0.0” bin, and figure 2 confirmed that the cause for why that bin was so large in figure 1 was because we had an enormous amount of tweets with a polarity score of 0.000. This was an outlier, and we initially thought that the large number of neutral tweets were because of references to articles about self-driving cars, or academic-related tweets. While this may be the case for some neutral tweets, after examining the dataset, we found that a significant portion of neutral tweets still appeared to be tweets made by the general public about the topic of self-driving cars. Therefore, we decided against removing completely neutral tweets in order to avoid biasing our data to more extreme results.

5. Data Visualization:

Once that potential outlier was out of the way, we focused on seeing if there was an association between the polarity score of the tweets and the release dates of Tesla vehicles. We first **visualized the data** with a line plot, plotting the date of the tweet against the tweet’s polarity score. This would allow us to see any trends that occur over time, creating easily identifiable spikes or valleys if an association between polarity score and release dates really did exist.

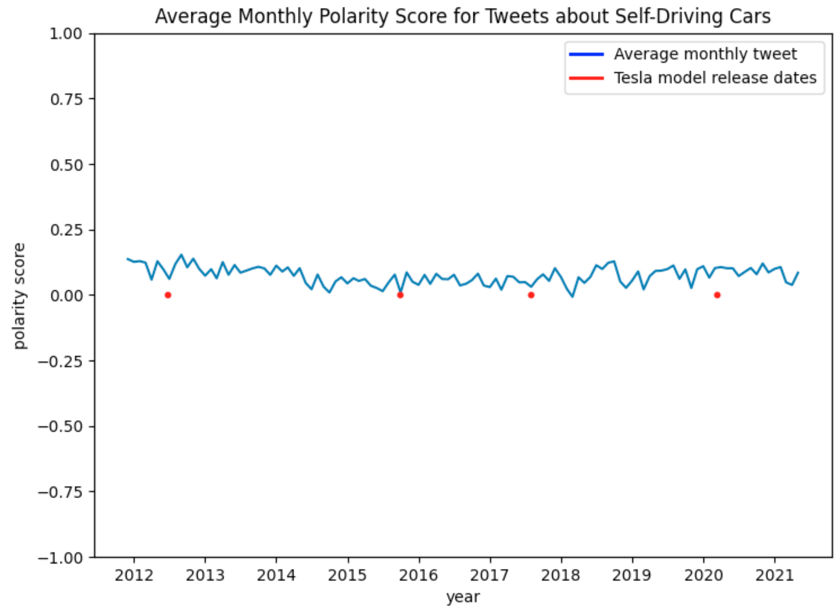

Figure 3. Average Monthly Polarity Score for Tweets about Self-Driving Cars. *The polarity scores of tweets relating to self-driving cars were averaged out per month and plotted against date. The uniformity of the trend line indicates that no significant increase/decrease in the public sentiment of self-driving cars occurred after the release of a new Tesla model, suggesting that no such relationship exists between the sentiment of self-driving cars and Tesla model releases*

Once we reached the conclusion that no relationship exists between the sentiment of self-driving cars and new Tesla car releases, we made an explanatory version of figure 3. We decided to keep the format as a line plot since it quite clearly showed the lack of any distinct trend or noticeable change in the sentiment of self-driving cars near new model release dates.

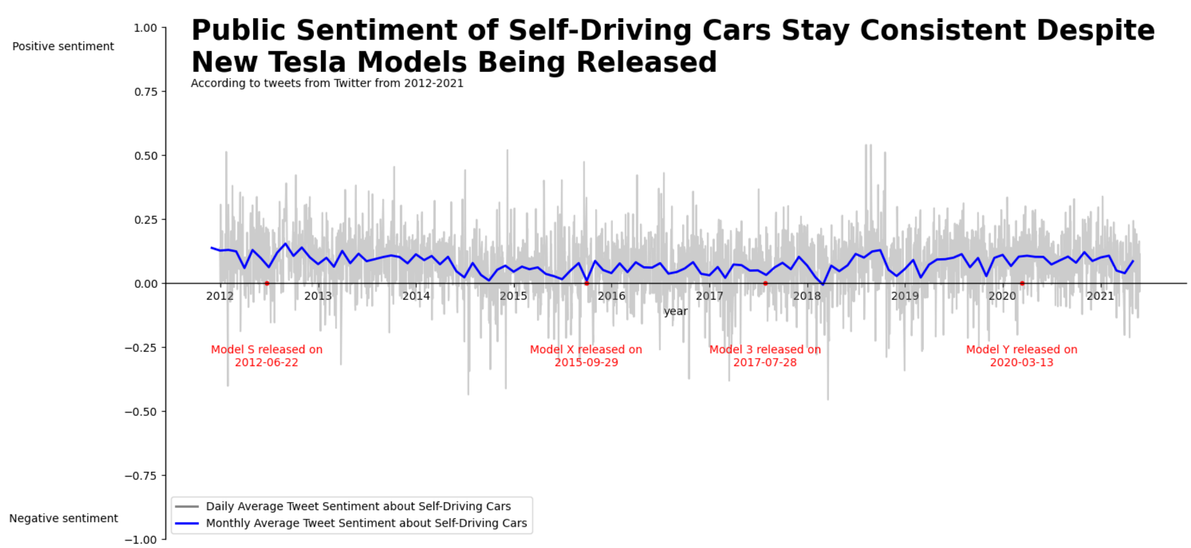

Figure 4. Explanatory Line Plot of sentiment analysis trends of tweets relating to self-driving cars. *Axes titles were modified from figure 3 to be more descriptive. Legend was modified to be more clear. The average daily tweet's polarity score was added in gray to show daily sentiment variance as time moves on. Underlying data is consistent for both figure 3 and figure 4.*

All the data collected, figures, and scripts used in this project were collected in this drive folder, which can be viewed by a UCSD Google account.

Discussion

The analysis would look to answer whether or not there was an association between public sentiment towards self-driving cars on Twitter and the release of new Tesla models. If the association does exist, we expect to see the polarity scores of tweets become more positive or negative around release dates. Based on figure 4, the trends in sentiment did not support a correlation between the two variables: public sentiments towards self-driving cars on Twitter and the release of new Tesla models. The line plot in figure 4 indicates a consistent trend in which the polarity scores of public sentiments towards autonomous cars on Twitter stay around 0.00 - 0.25. Hence, we conclude that the release of new Tesla models does not have an observable impact on the public sentiments towards self-driving cars.

The first limitation would be the age. People who use Twitter tend to be younger people, thus the data we collected from the tweets may have little information about how older people think about self-driving cars. Secondly, The NLP library that we used might be inaccurate in terms of processing the tweets and giving a polarity score. How the NLP library works is that it can break up a sentence into words and calculate the polarity score of each word. This means that phrases such as “not good” would be evaluated as two terms -- “not” and “good” -- and be individually assigned a sentiment score that doesn’t match the true sentiment of the tweet. Additionally, there may be sarcastic tweets that people use positive words to express their negative feelings about self-driving cars, or vice versa. Therefore, the NLP library might inadequately give tweets wrong polarity scores. Thirdly, people who use Twitter to express their sentiments towards autonomous vehicles may have stronger feelings towards self-driving cars. For instance, they may be more passionate about or have bad experiences with the subject; therefore, the tweets that we scraped might be biased and cannot properly represent the public sentiments towards self-driving cars. Moreover, since Twitter is a public communication platform, the rivals of self-driving cars are also able to post some negative judgments to Twitter, which is difficult to find and correct for us.

To address bias in our data, we would have to find the age of each user in our dataset and oversample from the age groups that aren’t fully represented so that the dataset we generate contains people’s opinions from different ages. Like all other personally identifiable information, this would be deleted after the initial confirmation that our data is more representative of the general US population. Additionally, we should pull from other sources of social media, such as Facebook or Reddit, to get a more accurate assessment of the general public. To get more accurate polarity scores, the NLP library could be used in tandem with another dataset built specifically for identifying positive or negative words on Twitter. This article demonstrates a possible way of using machine learning to better identify the sentiment of contextual phrases, and a similar method could be used to more accurately determine the polarity scores of tweets. Unfortunately, there’s not much we can do to combat more passionate people discussing self-driving cars, or find out whether commercial rivals of self-driving car companies are deliberately tainting the opinions of self-driving cars. People that are indifferent to self-driving cars likely won’t tweet about them.

It's important to note that our results are only meant to give a rough overview of the sentiment trends of Twitter users on self-driving cars. Because of issues of bias and the potential inaccuracy of the sentiment analysis, our results should not be taken as definite. The results should not be generalized across the entire global population or used as evidence for the global opinions of self-driving cars, as Twitter users tend to be younger and geographically located in wealthier countries. Additionally, we still need to make sure our own opinions of what the sentiment of self-driving cars should be do not act as a factor in our analysis and try not to mix up causation with correlation. Our study does not look for causal relationships. If we did find a statistically significant association, that doesn't immediately mean that a correlation exists between the release of new Tesla models and the sentiment of self-driving cars.

Group Participation

We decided to change our proposal since what we had for assignment 1 may not be suitable for data collection and later analysis. We came up with a new question, the hypothesis, and the justification together during our first meeting. **Zibin** looked for Tesla related articles for background information and worked on the ethical consideration portion. **Yining** did the initial version of Descriptive & Exploratory Data Analysis, and edited the rest part of ethical consideration. **Darren** wrote the data section of the proposal and the Data Collection, Data Wrangling, Text Analysis, and the Data Visualization sections for the Data Analysis. He also wrote and edited the final version of the Question, Justification, Background iInformation, and Descriptive & Exploratory Data Analysis, collected and wrangled the data, generated the figures, and created the tables and final visualizations. **Yueqi** made final edits to the entire project and contributed to the conclusions of what the analysis means in the discussion section. The rest of the discussion section was done collaboratively by everyone in the group.