I. Introduction

A. Problem Statement

In the past couple of years, Uber has found itself in the social spotlight quite a few times; unfortunately, this social spotlight has not always been a positive place for Uber to land. The goal of this project is to utilize social media and natural language processing to analyze the social sentiment of Uber. We will also analyze the sentiment surrounding Uber's direct competitor, Lyft, in an attempt to find answers to any struggles Uber may face, that Lyft has solved.

B. Background

In the recent months, Uber has become a social focus as their prices skyrocket following the COVID-19 pandemic. The increased pricing, as well as increased wait times, have left riders questioning if there is truly any convenience left when it comes to using Uber's services. These increased prices also seem to not take into account fair payouts for drivers, which was publicly voiced as Uber drivers led a strike for better pay this summer. In 2019, Uber fell short of their expected market valuation following their IPO closing and has struggled to reach that expected market valuation mark ever since. We are aiming to improve the social sentiment surrounding Uber in hopes of seeing a more positive financial outcome for the company as whole.

C. Goal

This project aims to determine any reason for negative sentiment that surrounds Uber as a company. Determining the source of any negative sentiment will allow us to propose practical solutions to help boost sentiment, allowing the public to view the company in a positive light and hopefully encourage market valuation to move forward.

II. Datasets

Social media sentiment was determined using 30,373 tweets scraped from Twitter. The Twitter API restrictions make it difficult to build a large dataset, to circumvent this issue, we downloaded a package called snscrape and utilized it's 'TwitterSearchScraper' function. We pulled all tweets containing the keyword 'Uber' from a date range of 02-01-21 to the date of scraping, 07-28-21. We also pulled tweets containing the keyword 'Lyft', but with a date range of 01-01-21 to the date of scraping, 07-28-21. The additional month in the Lyft date range was an attempt to balance out the number of mentions for either keyword; in the end, there was still a lack of balance. This lack of balance was seemingly due to Lyft running a much smaller operation than Uber's global business -- Lyft only operates within the US and Canada, while Uber operates in 69 countries around the globe. There were about 4x as many tweets mentioning Uber than there were mentioning Lyft. Columns of our final dataset included the Datetime for the tweet posted ('Datetime') , Twitter's tweet identification tag ('Tweet ID') , the actual text of the scraped tweet (‘Text’) , the username posting the tweet (‘Username’), the language the tweet was posted in (‘Language’), and eventually a sentiment label (‘Sentiment’).

III. Data Wrangling and Cleaning

Data Wrangling was performed using snscrape’s ‘TwitterSearchScraper’ function. Following the initial scraping, we began exploring the data to determine how best to clean it. The first step to cleaning the data was an attempt to identify tweets that did not contribute to the social sentiment of either company. Some of these tweets were identified by gathering the value counts of each username’s occurrence within the dataset. The idea behind this method was that most users reporting an opinion about something via social media are not likely to post sentiment based opinions more than a couple of times. To account for this, we read through tweets of users that accounted for more than 100 tweets within the dataset and assessed what their contribution to the project might be. From this group of users, we deleted all tweets that seemed to be strictly promotional or those who were tweeting about various unrelated topics but choosing to mention larger companies, such as Uber or Lyft, in hope of more online interaction. There were thirteen usernames that contributed over 100 tweets to the dataset; of these thirteen, seven usernames were dropped from the dataset -- resulting in just over 6,300 tweets being removed.

Our goal was to find a model that handled common social media styled text well enough to not require much text processing. To test this, we created two different datasets -- one was saved as an unprocessed dataset, and another saved following common text processing techniques. In the processed dataset: all html character entities, https links, and ‘#’ and ‘@’ symbols were removed -- all text was tokenized within the dataset as well. Stemming and removal of stop words was decided against due to risk of over stemming and loss of ability to determine sentiment. Each model was run using both datasets and it was ultimately determined that the text did not require processing to achieve results.

IV. Data Labeling

A. VADER Analysis

Initial attempt for data labeling was done using VADER (Valence Aware Dictionary and Sentiment Reasoner) Analysis. VADER works by scoring text on a scale from -1 to +1 and utilizing a user-set threshold to determine the label output. For this project, we set a threshold that rated all text scored from -1 through -0.05 as ‘Negative’, -0.05 through +0.05 as ‘Neutral’, and +0.05 through +1 as ‘Positive’. We ran VADER through each tweet within the dataset and set up a column for the text’s rating as well as another column that showed the sentiment label for the tweet. Following labeling with VADER, we noticed that there were quite a few tweets incorrectly labeled as ‘Positive’ and the model seemed to have a hard time accurately scoring tweets that were clearly negative. Misclassification was expected to a degree but we decided to attempt the same process using another model in hopes of better results.

B. Twitter-roBERTa-base for Sentiment Analysis

Ultimately, the decided upon model for sentiment analysis was found utilizing an AI community known as Hugging Face. Hugging Face is a domain where users can build, train and deploy state of the art models powered by the reference open source in natural language processing. The model we chose to use is a roBERTa-base model that was trained on ~58M tweets and fine tuned for sentiment analysis with the TweetEval benchmark.

After downloading and defining the model in our notebook, it was applied to the column of text within our dataframe -- creating a new column for sentiment labels of ‘positive’, ‘neutral’, or ‘negative’. Following the process of sentiment analysis for each tweet, we noticed that there were significantly fewer tweets classified as ‘positive’ when using this model in comparison to the VADER model.

Model	Positive Sentiment (%)
VADER	46.4
roBERTa-base	25.9

Table 1. Table of the percentage of positive sentiment labels determined by each model.

This lower rate of positively labeled tweets is what we had previously expected since users are more likely to speak out on social media when something has negatively affected them. We chose to move forward using the roBERTa-base model.

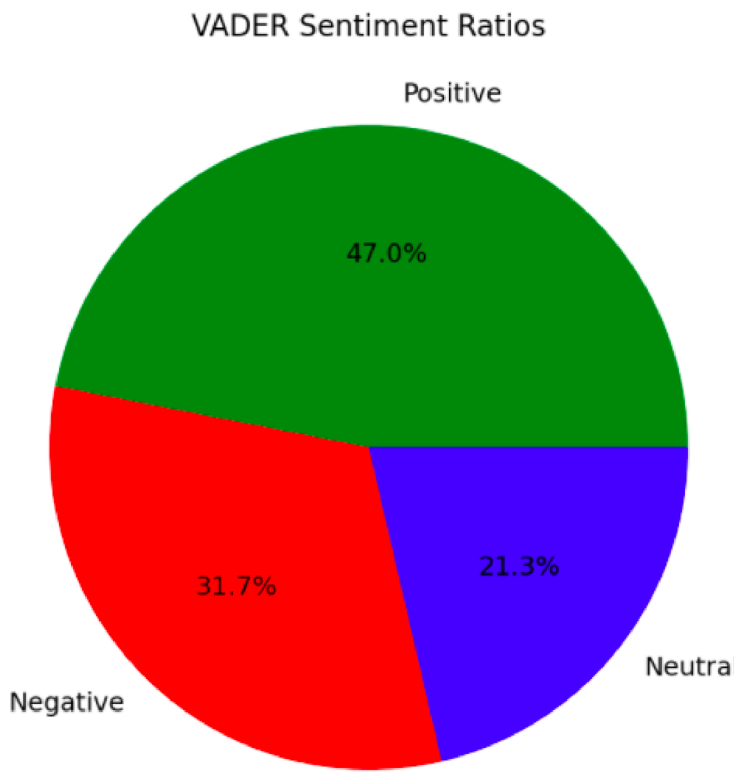

Figure 1. Pie chart showing VADER Sentiment Analysis ratio results.

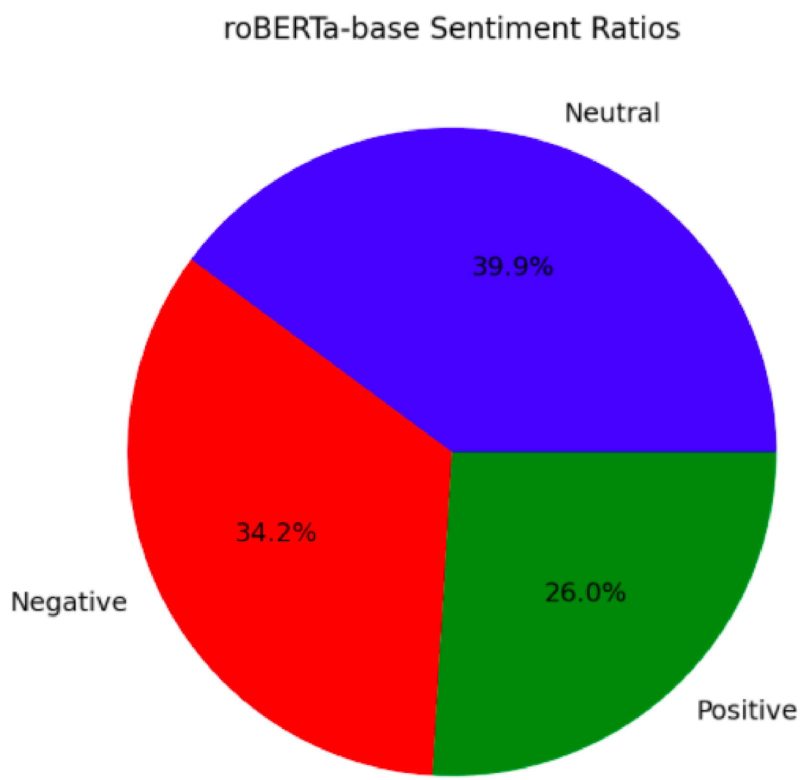

Figure 2. Pie chart showing roBERTa-base Sentiment Analysis ratio results.

V. Exploring Our Data

After creating a column of sentiment labels, it was time to explore our data and gain insight to the most realistic ways to improve company sentiment. We organized our data into four different datasets -- Positive tweets mentioning Uber (‘Uber Positive’), Negative tweets mentioning Uber (‘Uber Negative’), Positive tweets mentioning Lyft (‘Lyft Positive’), and Negative tweets mentioning Lyft (‘Lyft Negative’). Once these datasets were created, we were able to search for the most frequently occurring words within each dataset through the use of word clouds. These notable words allowed us to take a closer look at tweets determining the company’s social sentiment.

We made the assumption that these most frequently appearing words would lead to insight of what a company is doing correctly (found within the positive sentiment dataframes) or what a company could improve on (found within the negative sentiment dataframes). Once we took note of the most frequently occurring words that seemed like they would provide the most insight, we searched through the text that contained these keywords or phrases -- this is ultimately what led us to determine an approach to realistic solutions for improving sentiment surrounding Uber. For example, the word “driver” was a frequently occurring word in each of the four subsets of data. We filtered our data to only show tweets containing the word “driver” and read through the tweets until we had a better understanding of what types of issues or nonissues the public was having with Uber or Lyft. When going through the Lyft tweets, we went in with a focus on issues Uber might be dealing with, that Lyft might have already found an answer to.

Figure 3. Uber Positive WordCloud

Uber Positive notable words: ‘driver’, ‘ride’, ‘app’, ‘new’, ‘good’.

Uber Negative Notable words: ‘driver’, ‘ride’, ‘UberEat’, ‘cancel’, ‘time’, ‘Uber_Support’.

Figure 4. Uber Negative WordCloud

Figure 5. Lyft Positive WordCloud

Lyft Positive Notable words: ‘giftcard’, ‘driver’, ‘ride’, ‘car’, ‘win’.

Lyft Negative Notable words: ‘driver’, ‘AskLyft’, ‘price’, ‘ride’, ‘mask’, ‘passenger’.

Figure 6. Lyft Negative WordCloud

VI. Drawing Conclusions

After analyzing the social sentiment data for Uber, there are a few solutions we propose looking further into.

Proposal #1: A noticeably common mention within the 'negative' sentiment group is the lack of online presence when it comes to Uber Support. There seems to be a twitter handle set up to handle such tasks, @Uber_Support, but both riders and drivers are frequently mentioning that they are not receiving responses from this account. In cases where there is a response received, the response tends to be generic and unhelpful. In the data's 6 month date range, there are only 1201 mentions of Uber Support -- a potential solution could be a shift in the current support team to include a social media team that is tasked with responding to support requests of this type.

Proposal #2: The concern of driver pay being too low is not a new issue, but something that still requires addressal. Prices of rides are rumored to be increasing due to the lack of drivers available, but the price increase does not seem to include a wage increase for drivers. Drivers are also concerned that the charge increase may be a reason for noticeably smaller tips from riders. Our proposal is to adopt an incentive program similar to that of which is already in effect for Lyft drivers. Aim to offer weekly challenges for drivers and once these challenges are completed, a weekly bonus is to be distributed.

Challenge example: Complete 30 rides this week for a $X bonus

Sub proposals: these proposals are solutions for less frequently occurring concerns raised within the 'negative' sentiment group of tweets.

> sub proposal #1: There is concern about fraudulent damage reports -- these can hurt a rider who is not actually at fault, but can also hurt a driver who is falsely accused of submitting a fraudulent damage report. We propose a time restriction on when drivers can submit a report after drop off has occurred. For example, a driver would need to report damage created by a specific rider within 15 minutes of dropping them off. This should be a simple process to ensure the time restriction is a reasonable period for drivers to submit the report.
>
> sub proposal #2: Riders had reported issues about drivers making inappropriate or rude comments. Our proposal would be to implement a behavioral course that drivers need to complete prior to starting their position with Uber. The course should be no more than an hour and go through common topics such as sexual and racial harassment. All currently employed drivers will have 7 days to complete the training before a pause is placed on their account.
>
> sub proposal #3: All mentions of scheduled rides are about no-show or cancelled rides without refunds. Our proposal is to offer an incentive for drivers to follow

through on scheduled rides or a larger punishment for cancelling a scheduled ride after already accepting it.

sub proposal #4: There were a few mentions of drivers cancelling rides after seeing how far the drop location was. Our proposal is implementing the ability for drivers to set ranges for their rides. The larger the range, the higher the wage incentive for the driver.

Uber began as a competitor for outdated taxi services, but it is beginning to lose its upper hand in that competition. To continue to evolve beyond our current systems, we ultimately believe that a focus on customer support/service will set Uber apart from the rest.