Twitter Sentiment Mining

•••

Comparing Uber vs. Lyft Using rtweet and Tableau

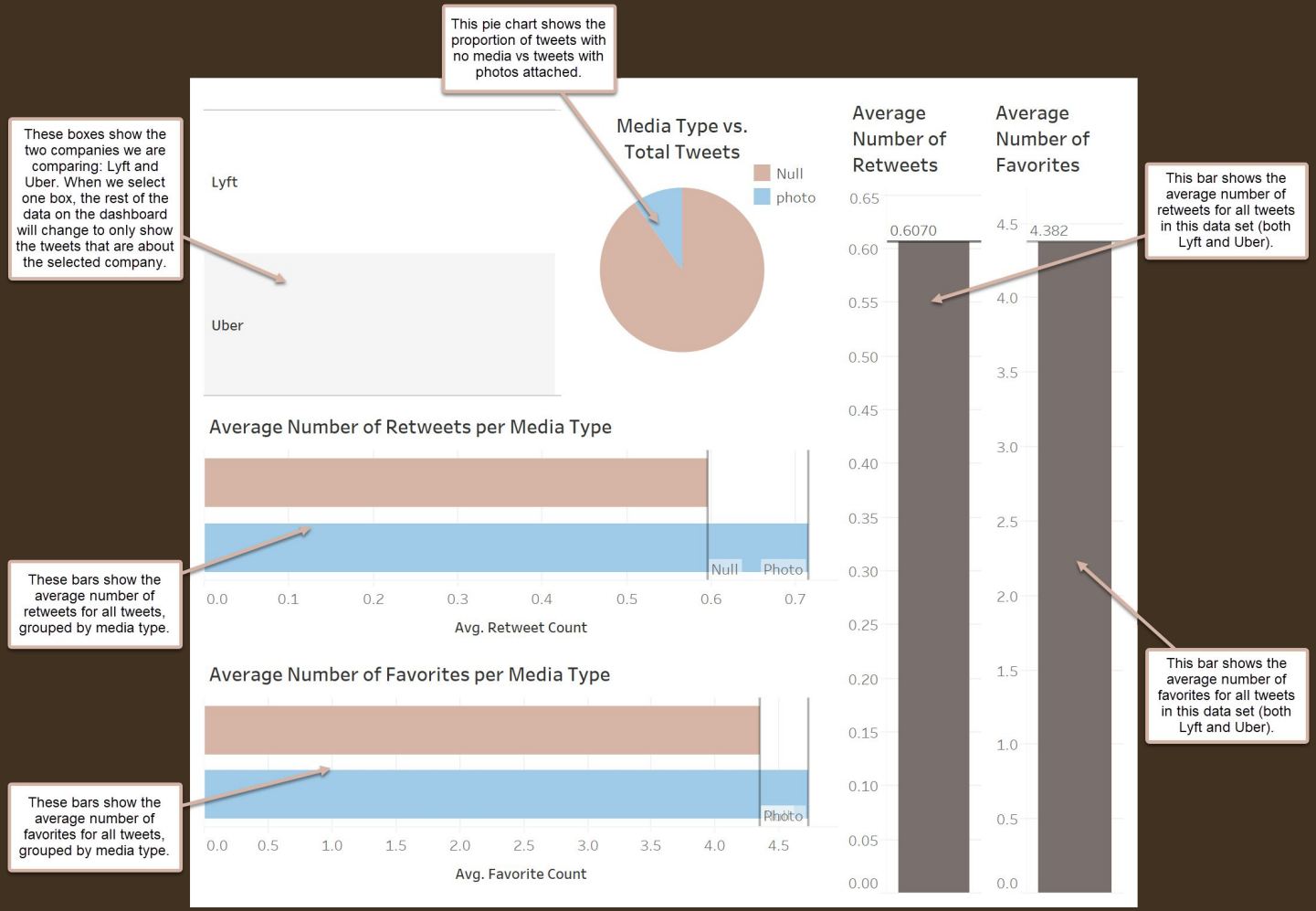
Dana DiVincenzo - MGIS.650.04

Scope

- Using R and the rtweet package, I downloaded 1000 tweets that mentioned the word “Lyft,” and 1000 tweets that mentioned the word “Uber”
- Rtweet outputted a data frame about each set of tweets, filled with information about the tweets themselves and about the account holders of the tweets
- A sentiment value was assigned to each tweet - tweets with positive words were given a positive value and tweets with negative words were given a negative value
- Using this data, I’ve created a Tableau dashboard which compares the Twitter presence of Lyft and Uber, and the visualizations can be used for further analysis of how the brands are discussed on Twitter

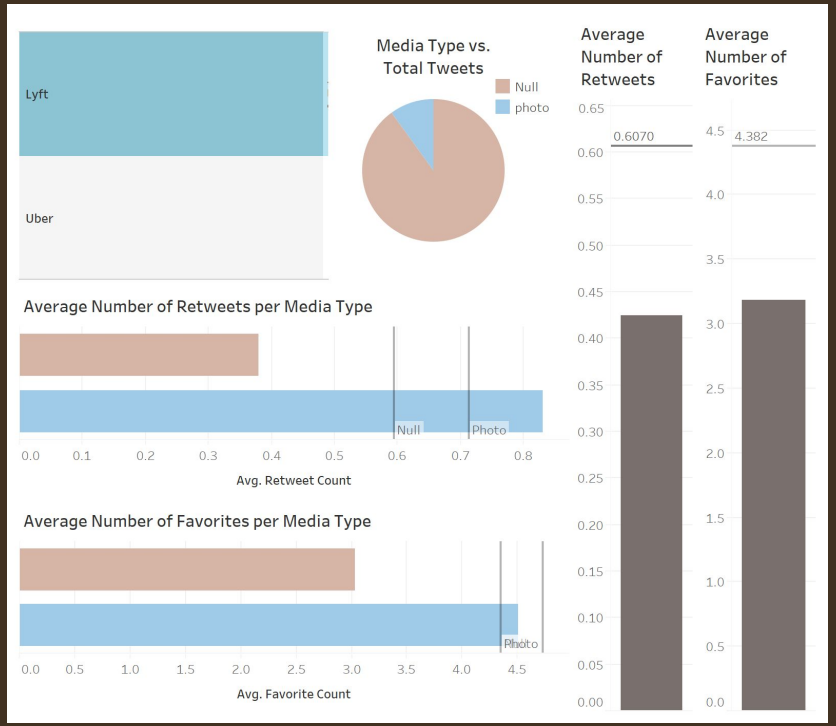
General Dashboard

- The dashboard on the next slide shows some visualizations which convey general information about the downloaded tweets
- A second dashboard which discusses sentiment values specifically will be discussed later on


This pie chart shows the proportion of tweets with no media vs tweets with photos attached.
These boxes show the two companies we are comparing: Lyft and Uber. When we select one box, the rest of the data on the dashboard will change to only show the tweets that are about the selected company.
Lyft
Uber
Media Type vs. Total Tweets
Null
photo
Average Number of Retweets
Average Number of Favorites
This bar shows the average number of retweets for all tweets in this data set (both Lyft and Uber).
0.6070
4.382
Average Number of Retweets per Media Type
These bars show the average number of retweets for all tweets, grouped by media type.
Null
Photo
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Avg. Retweet Count
Average Number of Favorites per Media Type
These bars show the average number of favorites for all tweets, grouped by media type.
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
Avg. Favorite Count
This bar shows the average number of favorites for all tweets in this data set (both Lyft and Uber).

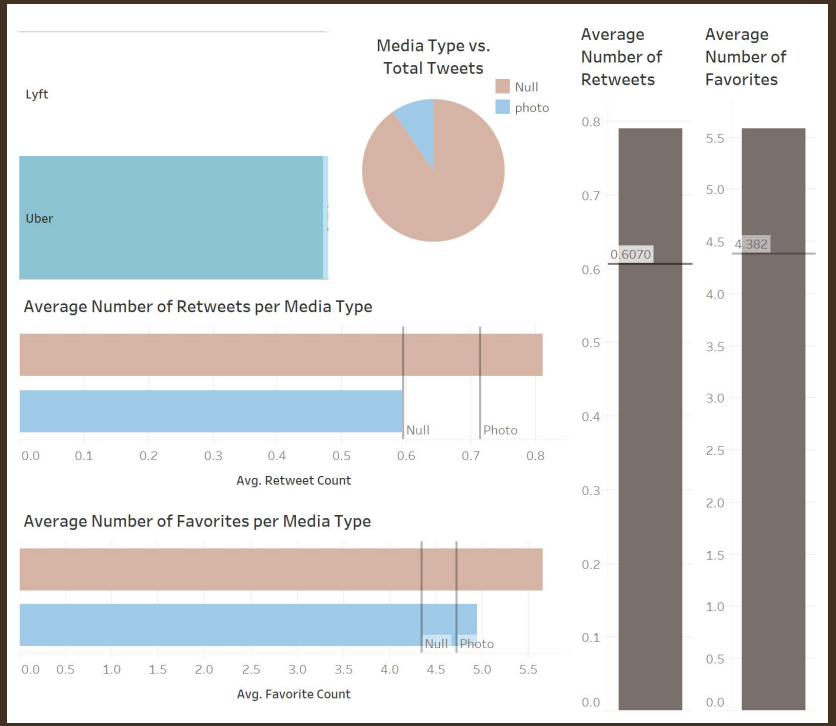
General Dashboard - Lyft

- In the Lyft dataset, the average number of retweets and average number of favorites are significantly lower than those averages of the entire dataset
- For the Lyft dataset, the proportion of photo tweets to non-photo tweets is almost exactly the same as the proportion for the entire dataset
- For the Lyft dataset, the average number of retweets for photo tweets is significantly higher than that average of the entire dataset
- For the Lyft dataset, the average number of favorites for photo tweets is very similar to that average of the entire dataset
- In the Lyft dataset, the average number of retweets and favorites for non-photo tweets are significantly lower than those averages of the entire dataset



General Dashboard - Uber

- In the Uber dataset, the average number of retweets and average number of favorites are significantly higher than those averages of the entire dataset
- In the Uber dataset, the proportion of photo tweets to non-photo tweets is almost exactly the same as the proportion for the entire dataset
- In the Uber dataset, the average number of retweets and favorites for non-photo tweets are significantly higher than those averages of the entire dataset
- In the Uber dataset, the average number of retweets for photo tweets is very similar to that average of the entire dataset
- In the Uber dataset, the average number of favorites for non-photo tweets is significantly higher than that average of the entire dataset

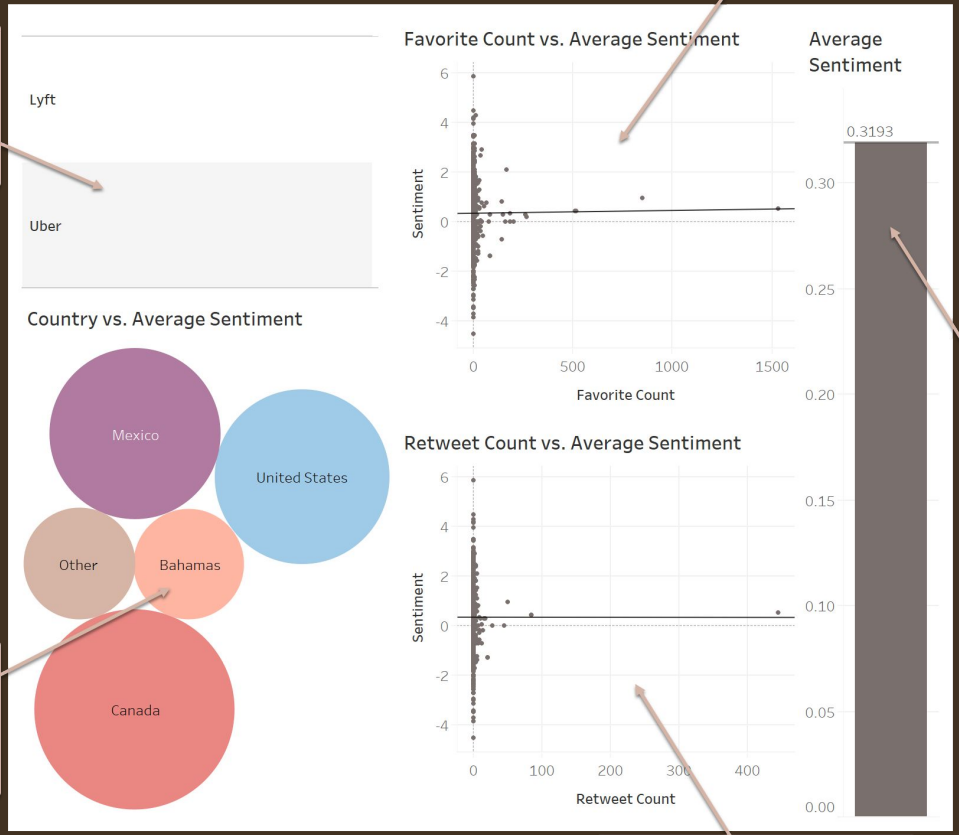


General - Uber vs. Lyft

- Tweets in the Uber dataset had significantly higher averages of both retweets and favorites than tweets in the Lyft dataset
  - This does not tell us if the tweets reflect a positive sentiment on Uber, just that the tweets about Uber are more popular than the tweets about Lyft
- Tweets in the Lyft dataset which contained photos had higher averages of both retweets and favorites when compared to the Uber dataset
  - This means that Lyft, or a person tweeting about Lyft, may want to include a photo in their tweet if they would like it to be more popular
- Tweets in the Uber dataset which did not contain photos had higher averages of both retweets and favorites when compared to the Lyft dataset
  - This means that Uber, or a person tweeting about Uber, may not want to include a photo in their tweet if they would like it to be more popular

Sentiment Dashboard

- The dashboard on the next slide shows some visualizations which convey information about the sentiment of the downloaded tweets



These boxes show the two companies we are comparing: Lyft and Uber. When we select one box, the rest of the data on the dashboard will change to only show the tweets that are about the selected company.

These circles show the average sentiment of all tweets based on the country of the account holder. Larger circles show a more positive average sentiment of the tweets in that country.

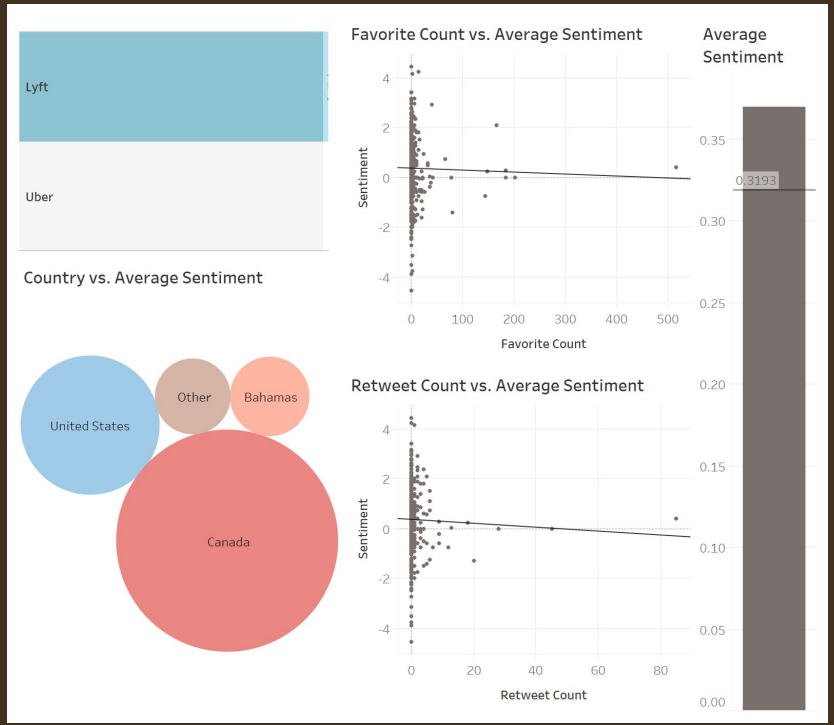
This graph shows the relationship between average number of favorites and average sentiment. A positive slope implies that more popular tweets have a more positive sentiment.

This bar shows the average sentiment of all tweets (Uber and Lyft). Positive sentiment values imply positive tweets, and negative sentiment values imply negative tweets. As the absolute value of the sentiment becomes greater, the tweets become more positive and more negative.

This graph shows the relationship between average number of retweets and average sentiment. A positive slope implies that more popular tweets have a more positive sentiment.

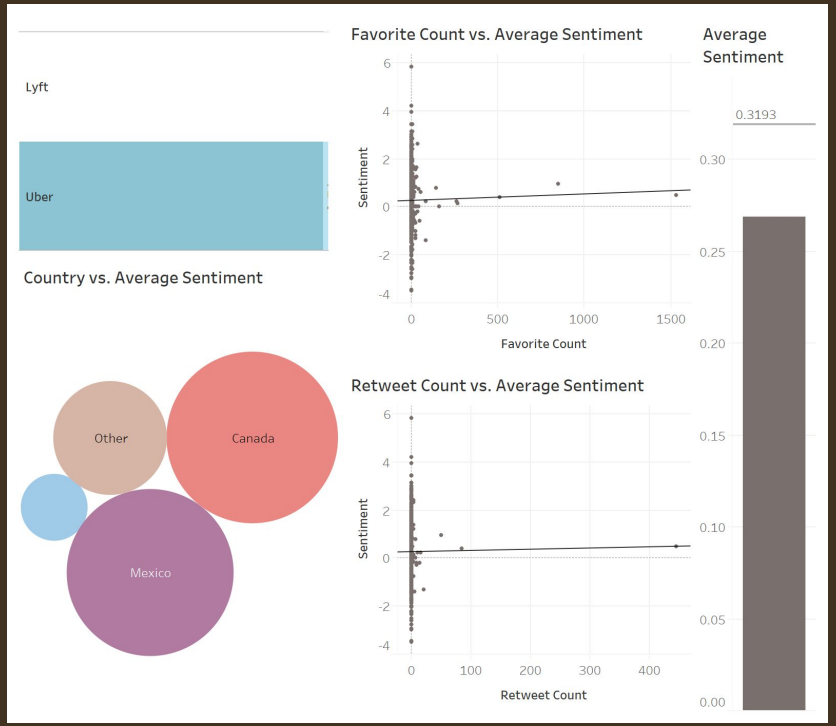
Sentiment Dashboard - Lyft

- In the Lyft dataset, tweets from Canada had the most positive average sentiment
  - Lyft does not operate in Mexico, so it makes sense that Mexico does not appear on this graph
- In the Lyft dataset, the slopes of both “favorite count vs. average sentiment” and “retweet count vs. average sentiment” are negative
  - This tells us that the more popular tweets involving Lyft have a more negative sentiment
- In the Lyft dataset, the average sentiment is significantly higher than the average sentiment of tweets from the entire dataset



Sentiment Dashboard - Uber

- In the Uber dataset, tweets from Canada had the most positive average sentiment, followed closely by Mexico
  - Uber does not operate in the Bahamas, so it makes sense that the Bahamas does not appear on this graph
- In the Uber dataset, the slopes of both “favorite count vs. average sentiment” and “retweet count vs. average sentiment” are positive
  - This tells us that the more popular tweets involving Uber have a more positive sentiment
- In the Uber dataset, the average sentiment is significantly lower than the average sentiment of tweets from the entire dataset



Sentiment - Uber vs. Lyft

- On average, tweets about Lyft have a significantly higher sentiment value than tweets about Uber
- However, on average, more popular tweets about Uber have a higher sentiment value, while more popular tweets about Lyft have a lower sentiment value
- On average, tweets from Canada have the highest sentiment value for tweets about both Uber and Lyft
    - This could mean that Canadians like Uber and Lyft more than other people do, or Canadians tweet with more positive language in general
- Tweets from the United States have a significantly lower sentiment value when mentioning Uber compared to mentioning Lyft

Conclusions

- Tweets about Uber are more popular on Twitter than tweets about Lyft
- However, tweets about Lyft are more positive, on average, than tweets about Uber
- Which is better for a company - to be talked about more online, or to be talked about positively online?
    - More analysis needs to be done to see which situation leads to the most revenue
- Because tweets with photos were much more popular in the Lyft dataset, if Lyft wanted to increase the visibility of a tweet, I would recommend including a photo in the tweet
- Because tweets without photos were more popular in the Uber dataset, if Uber wanted to increase the visibility of a tweet, I would recommend not including a photo in the tweet