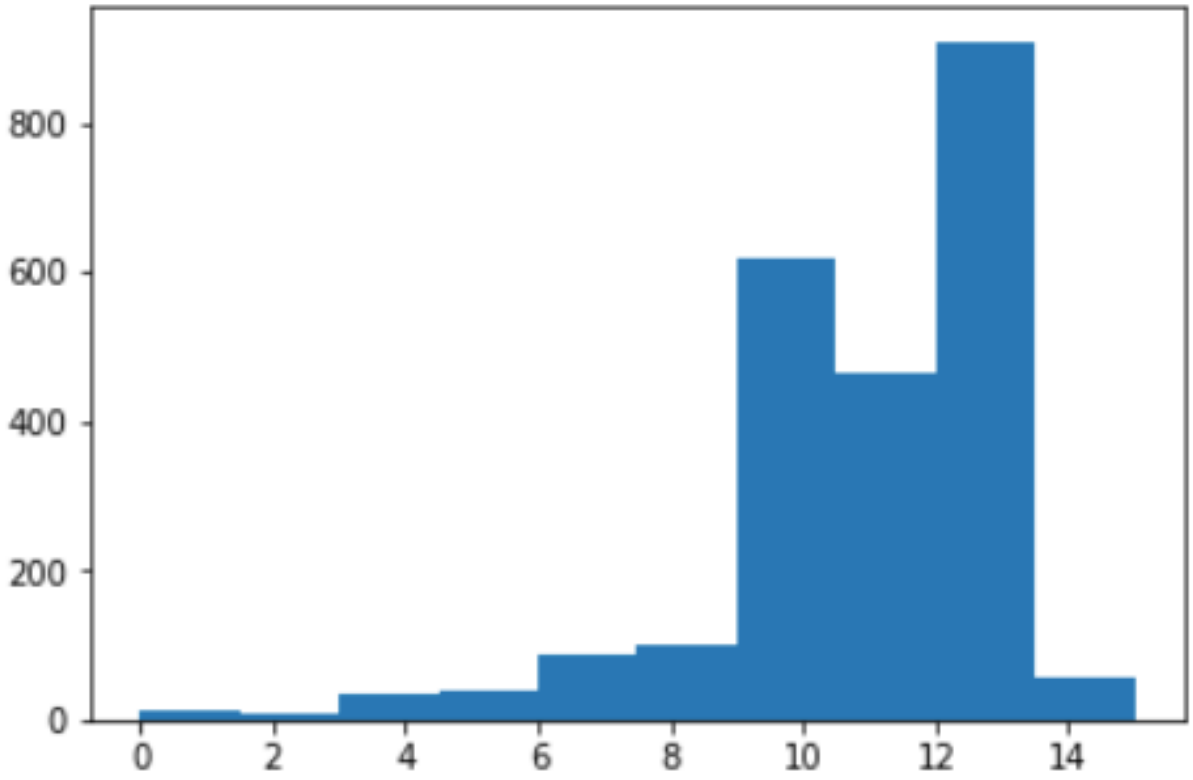
Introduction

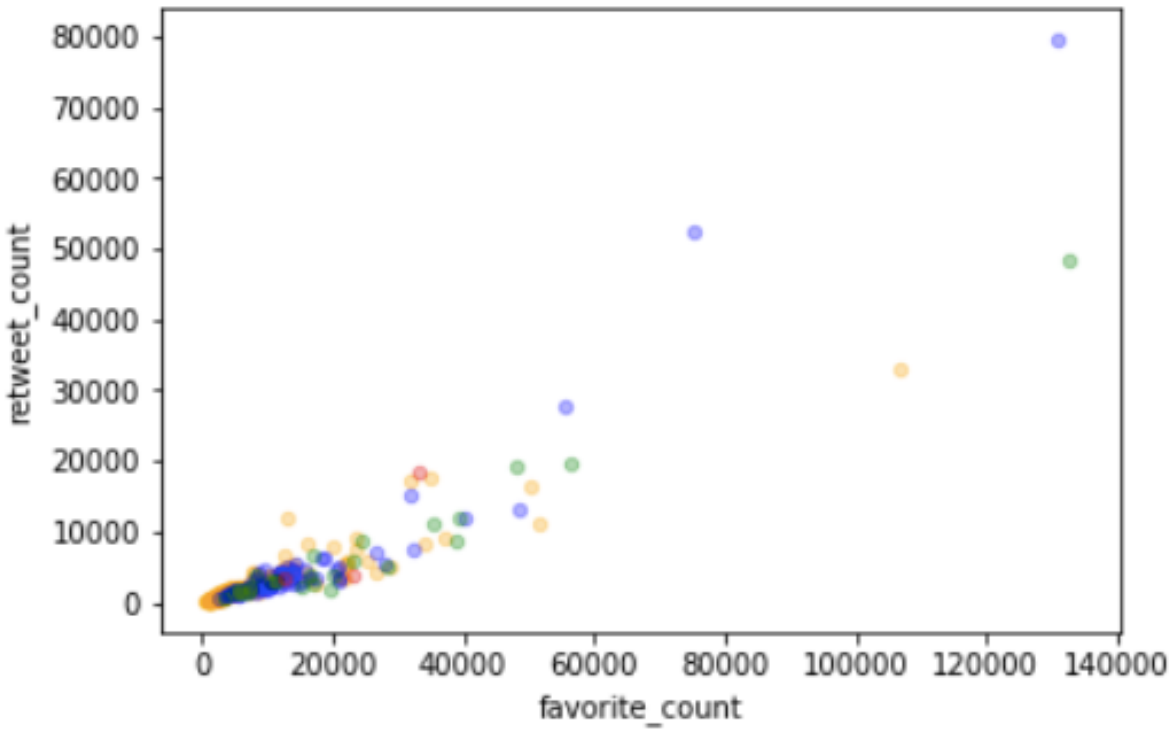
This project is part of Udaciy Data Wrangling and Visulization project. I use Python and its libraries to gather data from a variety of sources to analyze and visulize the tweet archive of Twitter user @dog_rates, also known as WeRateDogs, which is a Twitter account that rates people's dogs with a humorous comment about the dog. The data comes in a variety of formats, I assess its quality and tidiness, then clean it to analyze and visulize the key findings.

Analyze and Visualize



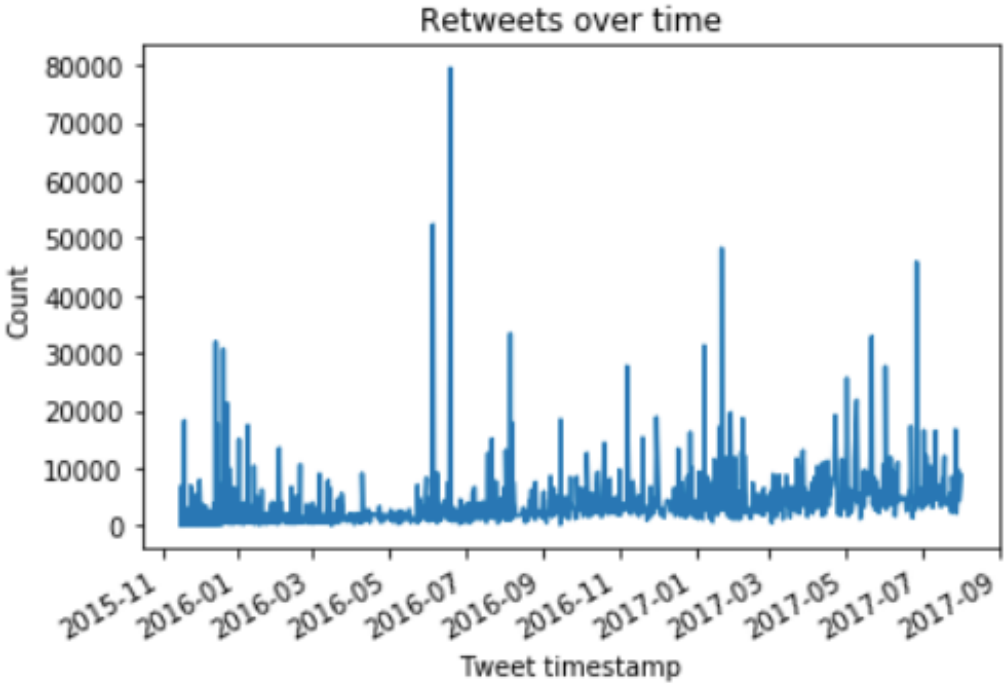
Key Findings:

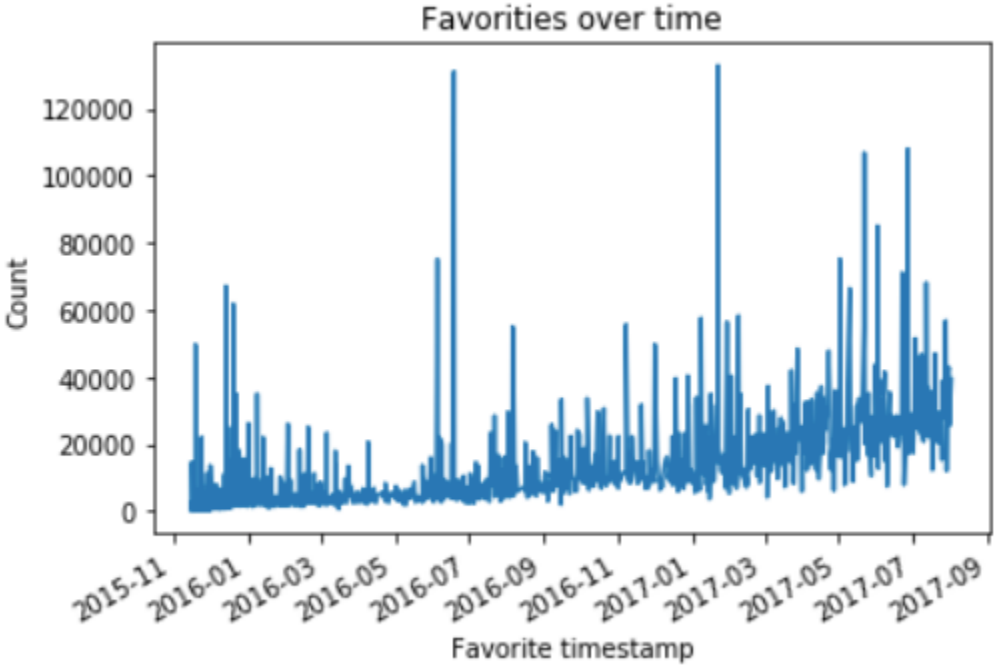
There are 26 ratings greater than 15. Majority of the ratting are range between 10 to 13, with 13 received the most ratings

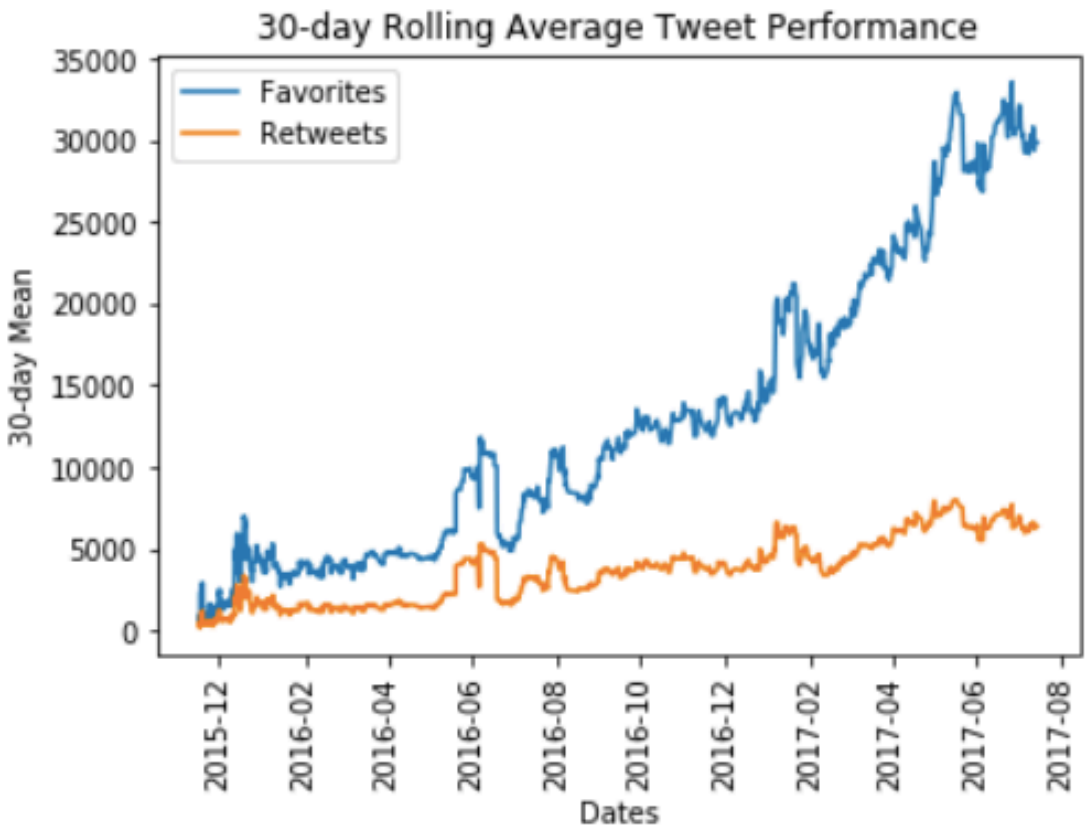


Key Findings

The chart shows there are direct relationship between retweet_count and favorite_count. The more favorite the dog received, the more retweet will happen. Most of the images have less than 20000 favorites.


Retweets over time
80000
70000
60000
50000
40000
30000
20000
10000
0
Count
2015-11
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
2017-07
2017-09
Tweet timestamp


Favorities over time
120000
100000
80000
60000
40000
20000
0
Count
2015-11
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
2017-07
2017-09
Favorite timestamp



Key Findings

There is a trend in the favorites and retweets, and both Favorites and Retweets saw increase over time. This trend increased as the account became more popular. In the chart above, we see a spike in both favorites and retweet in June 2016. The favorites count increase rate is bigger than retweets and it is more noticeable when compared to several large outliers in favorites for extremely popular tweets