

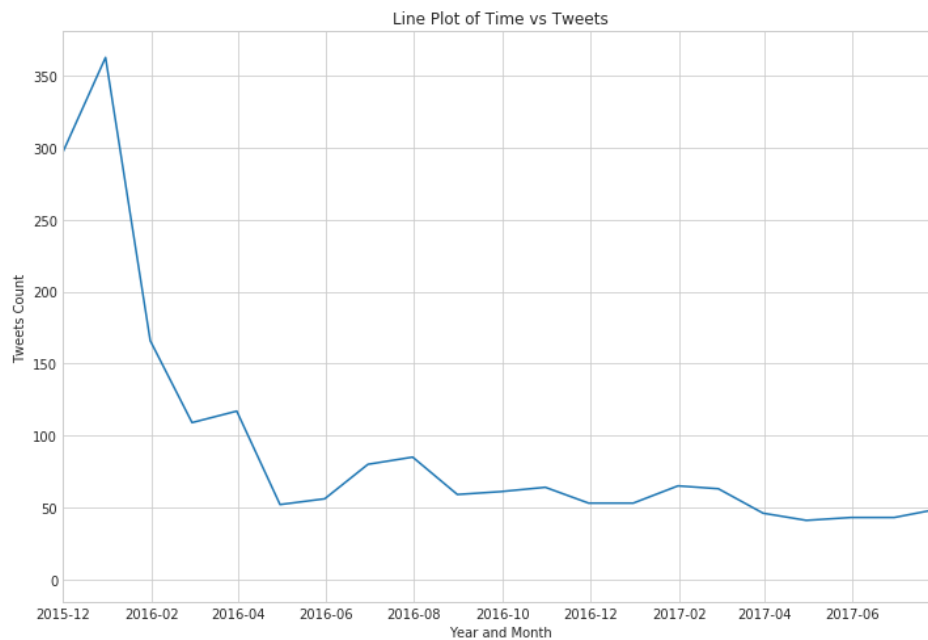
Act Report

By Nick Webb

8/5/2019

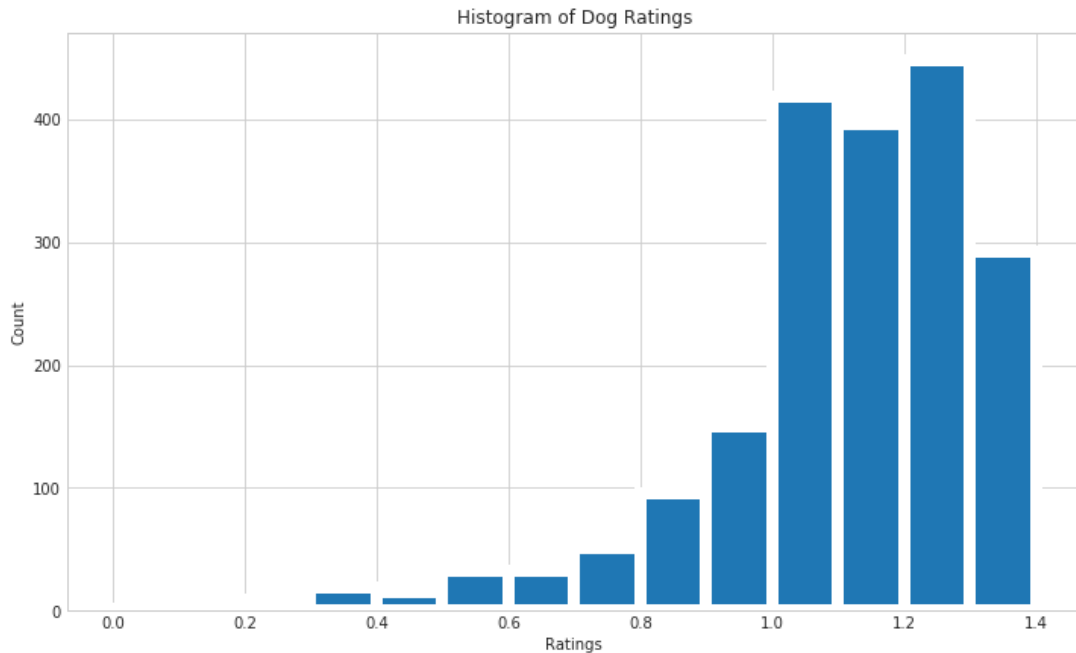
Tweets Over Time

To understand the frequency of the tweets and how they were established over time it felt fitting to do a basic time series analysis. The graph below shows the frequency of the original tweets from the account. From this, we can understand that the accounts activity is dwindling and the most popular posts might have been less popular if the rate of new material was going up.



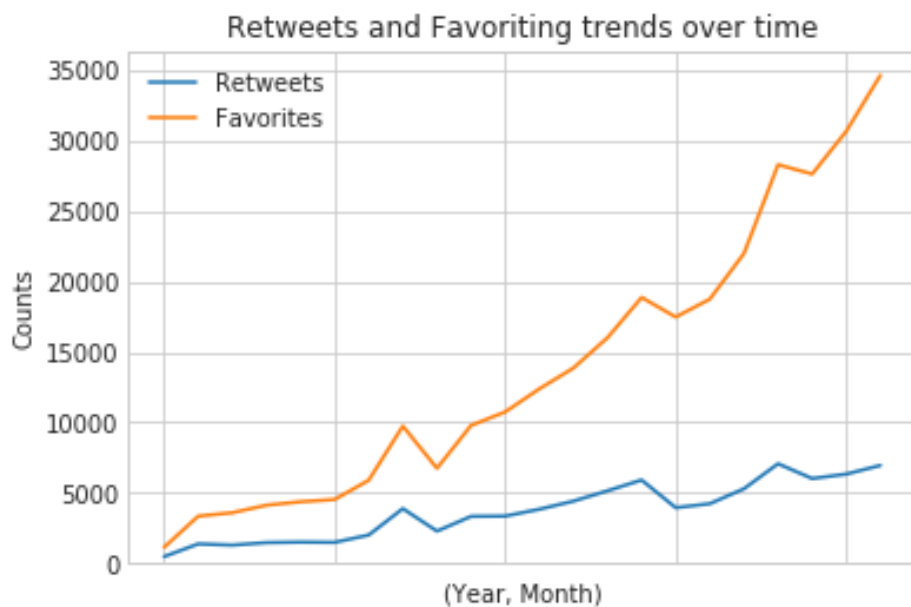
Dog Ratings Histogram

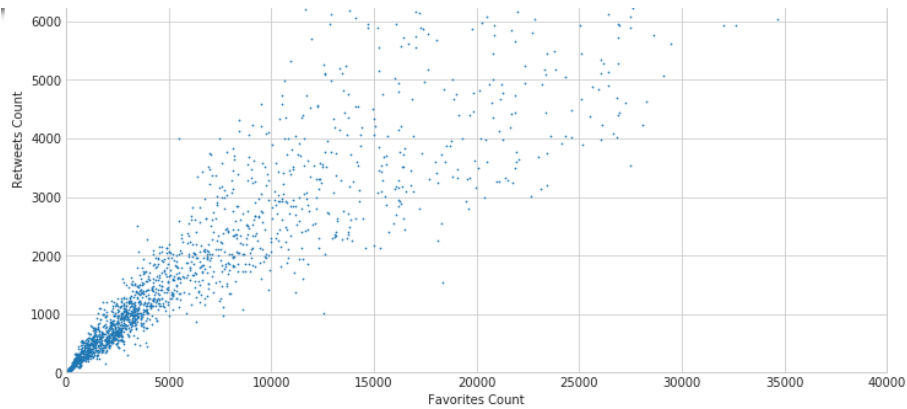
While consolidating, a new column was created called “rating” which was appropriately defined by the fraction ($\text{rating_numerator} / \text{rating_denominator}$). While holding the denominator to values of 10, we could see that the majority of the numerator values are indeed "Almost always greater than 10". The histogram is skewed to the left by the majority of ratings be equal or greater than 1.



Retweets Vs Favorites, Line and Scatter Plots

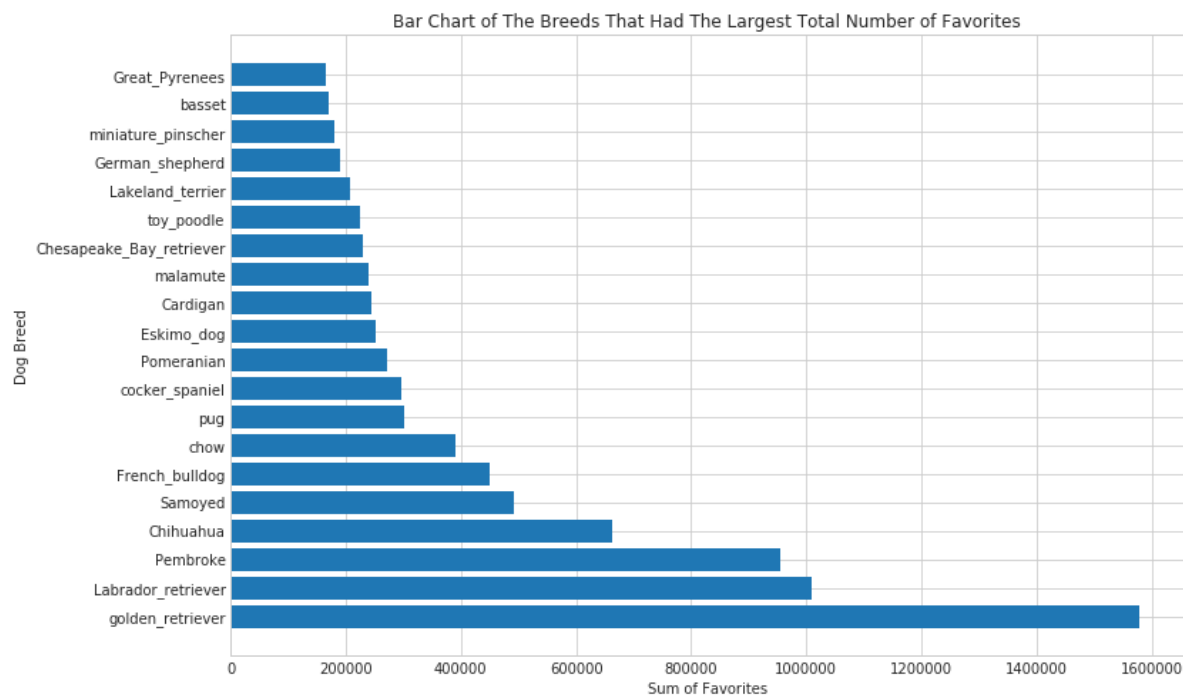
It seemed reasonable to investigate the relationship between “Retweets” and “Favorites”. One would assume a positive correlation between both variables and as one would expect from looking at the line chart of retweets and favorites over time (both have clearly have similar trends), using a scatterplot reassured that assumption. I was going to investigate this further and analyze the relationship between both counts and the rating scores, but due to being very ill lately, I have not had a ton of time to work on this and became behind.





Determining Breed Popularity with a Bar Chart

Using the image prediction results to establish which breed was associated to each text, the Golden Retriever was clearly the most popular breed on this account. I suspect that if there were more new tweets from this account that we would see more pitbulls (staffordshire terrier) posted and favorited as this breed has become very popular in the past few years. It would be interesting to look at each breeds post over time to see their frequency.



This project was very challenging. This project confirmed my feeling that gathering the data and connecting to different data sources is the most challenging part. All in all, it was pretty interesting.