WeRateDogs Project
Analysis of data wrangled from Twitter

By Aakash Behl
Date: April 10, 2018

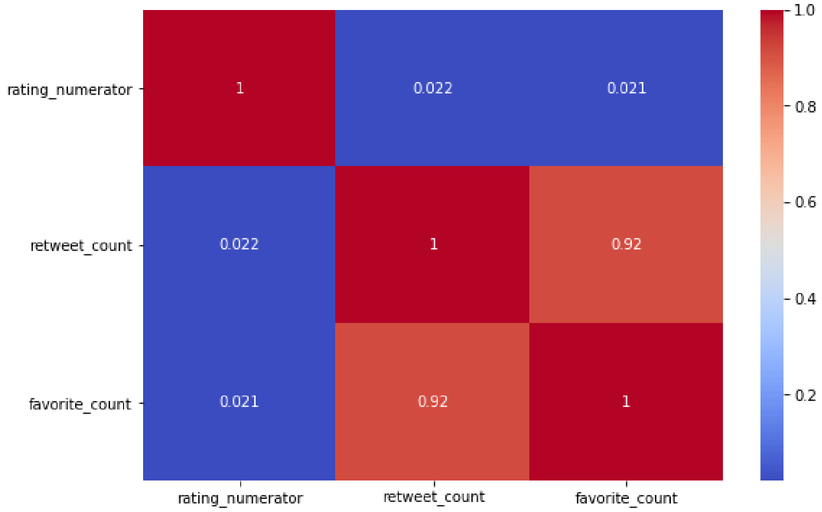
Twitter today has almost 200 million users worldwide.
Approximately 460,000 new Twitter accounts are opened daily. More than 140 million tweets are sent daily. That's one billion weekly tweets. The goal here was to capitalize on Twitter's vast amounts of tweet data.

WeRateDogs is a popular Twitter account with over 4 million followers and has received international media coverage. WeRateDogs got popular by rating people's dogs with a comment about the dog. The rating system is fraction based, with the denominator at 10 and the numerator is always a number greater than 10 (for the most part), the only exception to that being if it is a case of plagmarism or it is not a photo of a dog.

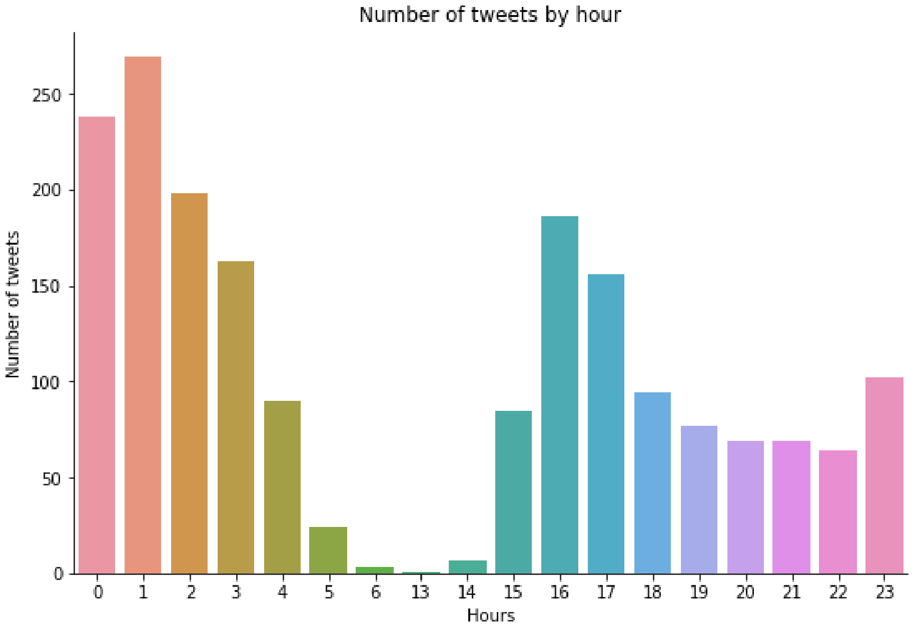
For this analysis I have gathered data from three different sources. WeRateDogs gave Udacity access to their Twitter archive for this project in the form of a csv file. This archive contains basic tweet data. Each tweet image was run through a convolutional neural network with the purpose of correctly identify the dog breeds. The predictions were programmatically downloaded using the Requests Python library as a tsv file. Then, using the tweet IDs from the WeRateDogs dataset I queried the Twitter API for each tweet's JSON data using the Python's Tweepy library I stored each tweet's entire set of JSON data, which I later used to analyze the tweet's retweet and favorite counts.

The highest value in rating_ numerator is a dog named as Atticus. The tweet was sent on July 4, 2016, "This is Atticus. He's quite simply America…" his picture is below.

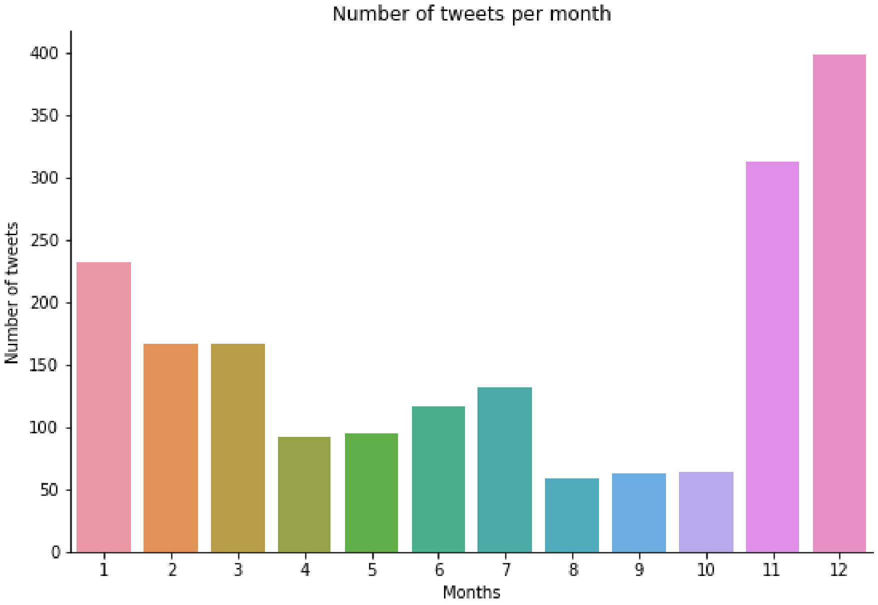
I tried to find a relation between the number of retweets and favourite counts and there was a very strong correlation of r = .92.



I also worked on finding the most popular time for a tweet, the analysis of which is as follows



The highest number of tweets are there in the night hours, but there are no tweets in the timeframe of 6:00 AM to 2:00 PM which is really strange



Looks like December is the most popular month for the tweets.

In summary, the data analysis of WeRateDogs, the favorite and retweet data is strongly correlated and the most active month is December.