MOHIT SINDHANI

Introduction

The dataset is a tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10.

Datasets:

1. The WeRateDogs Twitter archive: twitter_archive_enhanced.csv
2. The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. (image_predictions.tsv)
3. Each tweet's retweet count and favorite ("like") count at minimum, and any additional data you find interesting. Using the tweet IDs in the WeRateDogs Twitter archive, we have queried the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet_json.txt file

Insights

1. Top rated and lowest rated dogs

Tweet id 740214038584557568 is the lowest rated

Tweet id 702217446468493312 is the highest rated

2. Most retweeted photo

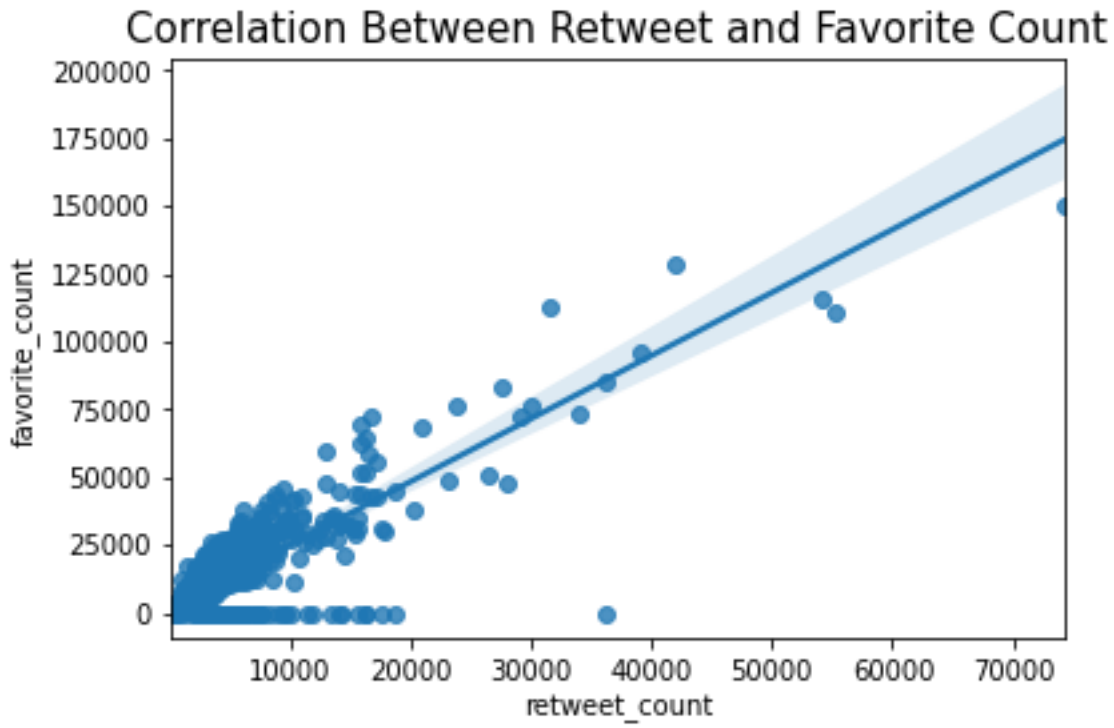
Insight: Tweet_id 744234799360020481 is the most retweeted tweet with 83727 retweets.

3. Tweet source

98% of users are using an iPhone to view tweets, 1.4% are using a desktop, and 0.5% are using Tweetdeck.

Visualization

Correlation between retweet and favorite count¶



I selected this visual image as a result of I detected that whereas assessing for three insights, the most favorited and most retweeted tweet ar identical. It got Maine thinking what is the correlation betweetn retweet_count and favorite_count variables? By viewing the graph on top of and doing a visible assessment, we will infer that a greater variety of retweets can increase the chance that the tweet will be favorited. there's a steep, correlation supported the most effective match line of the graph. Every increase in retweet count can get multiple favorite hits.