Analysis Report

Introduction

For this analysis I gathered data from three different sources. WeRateDogs gave Udacity exclusive access to their Twitter archive for this project in the form of a csv file. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. Each tweet image was run through a convolutional neural network with the purpose of analyzing the images to correctly identify the dog breeds. The convolutional neural network predictions were programmatically downloaded using the Requests Python library as a tsv file. And finally, using the tweet IDs from the WeRateDogs archive 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 would later use to analyze the tweet's retweet and favorite (i.e. "like") counts.

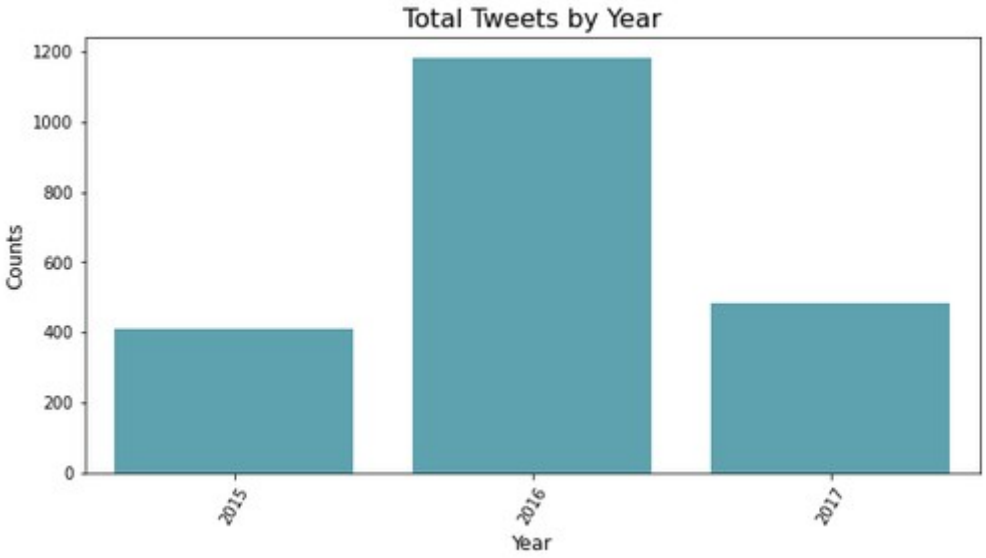
Analysis

The question intended to be answered afte the wrangling process were completed are:

1. What is the average rating for dogs?
2. What is the average rating for each dog stage?
3. what are the top five dog breed with the highest average ratings?
4. what are the total number of tweet by year? visualised with bar chart.

The insight drawn from the above questions are as follows

1. The average rating for dogs is `13.7`
2. `Doggopuppo` stage has the average rating of `13` while the `Pupper` has the average rating of `10.9`
3. Top five breeds with highest average ratings are `Banana` `Black-footed_ferret` `Rottweiler` `Shower_curtain` `Redbone`
4. `2016` has the number of tweets among other years.



Conclusion

This write-up offers a straightforward look at the data analysis proces

There is so much more that can be done
with this data set