Report: Act report

Introduction

Real-world data rarely comes clean. It needs to be wrangled and cleaned before any analysis and visualization
The management of three Datasets (wrangling and cleaning) used in this project is explained in details in the attached "wrangle_report.ipynb".

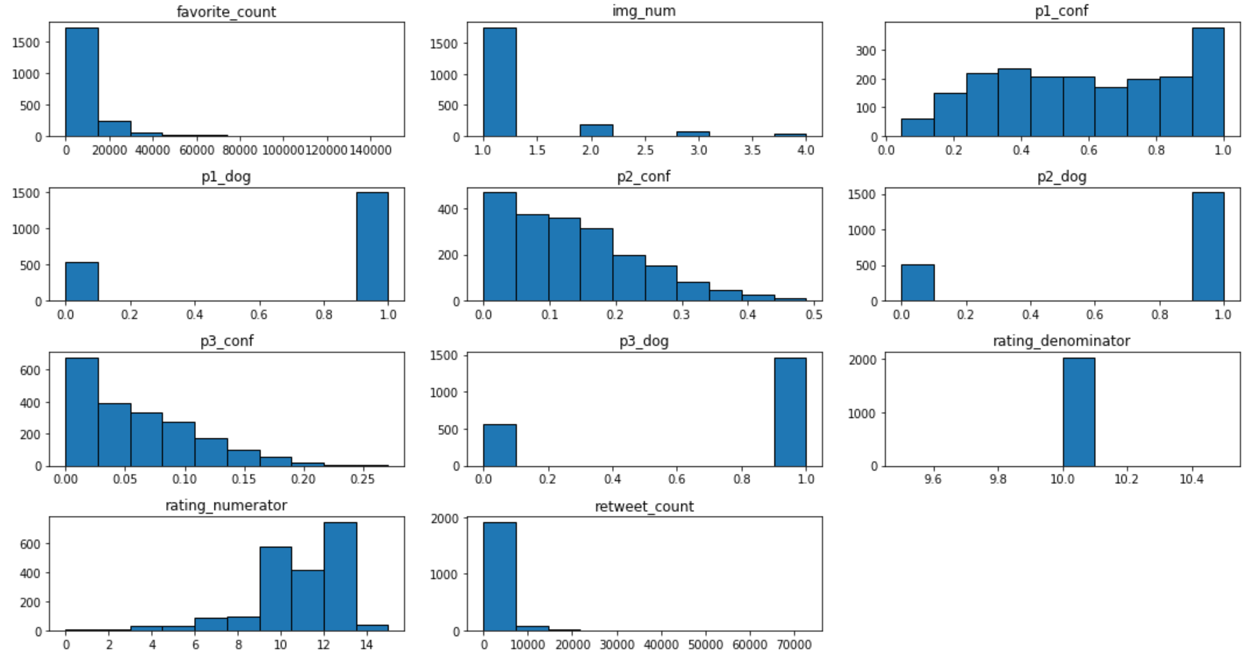
This report will includes some interested insights of the wrangled and cleaned data.
The analyzed data is the tweet archive of twitter user @dog_rates which is also known as WeRateDogs. WeRateDogs is a twitter account that rates people's dogs with a humorous comment about the dog. These rating almost always have a denonminator of 10; but the numerators almost always greater than 10 (11/10, 12/10, 13/10, etc.). Because "they're good dogs Brent" as per the founder of this page "Matt Nelson". WeRateDogs has over 9 million followers now on twitter and has received international media coverage.

Analysis of the Data

I will focus my analysis on the following:

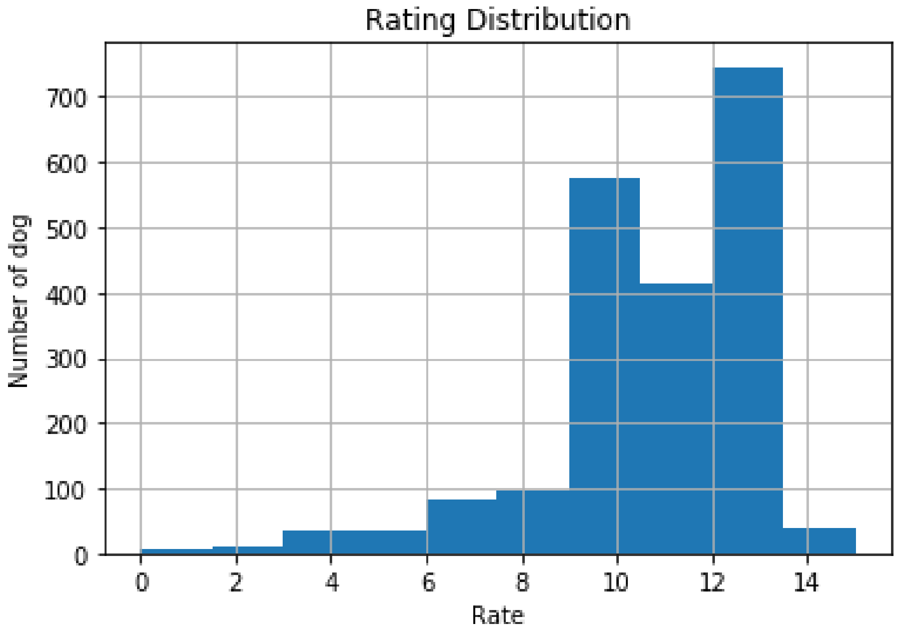
a. What is the most used name for dogs?
b. What is the highly rated dog stage?
c. How is the correlation between favorite count and retweet count?
d. What is the most truly identified dog breed by the neural network according to the first prediction of image's dog?

1. Overview of all variables in the dataset



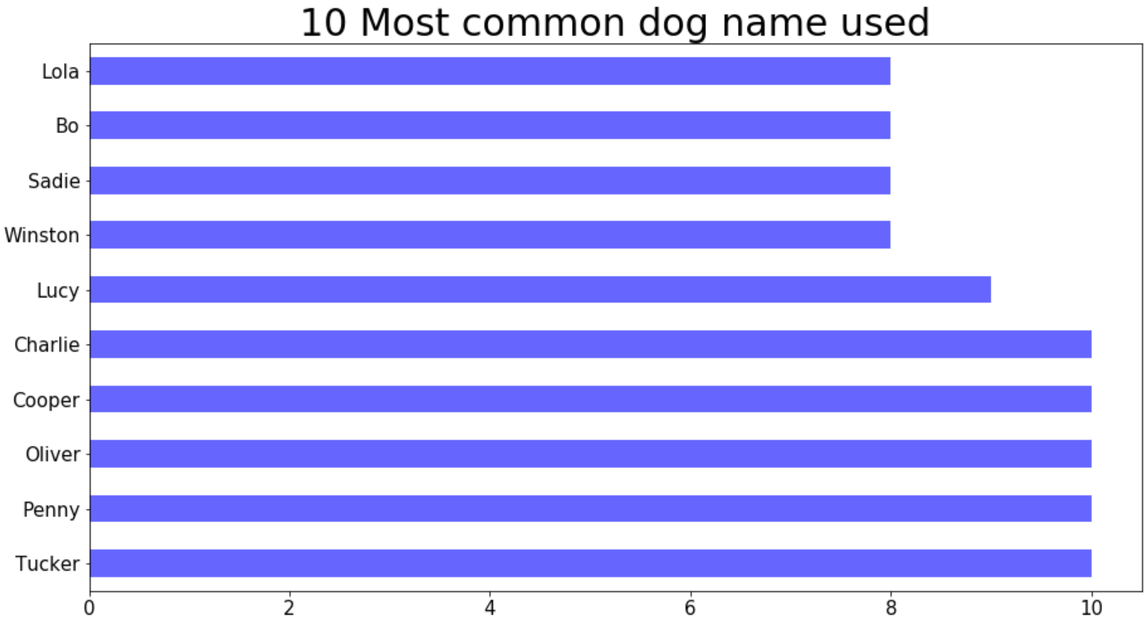
The most of tweet had a favorite and retweet counts less than 20000 and include only one image of the dog. The first algorithm seems to be the most confident one comparing to the second and the third algorithm. Most of rating for dogs varies between 10 and 14.

2. Dogs Rating Distribution



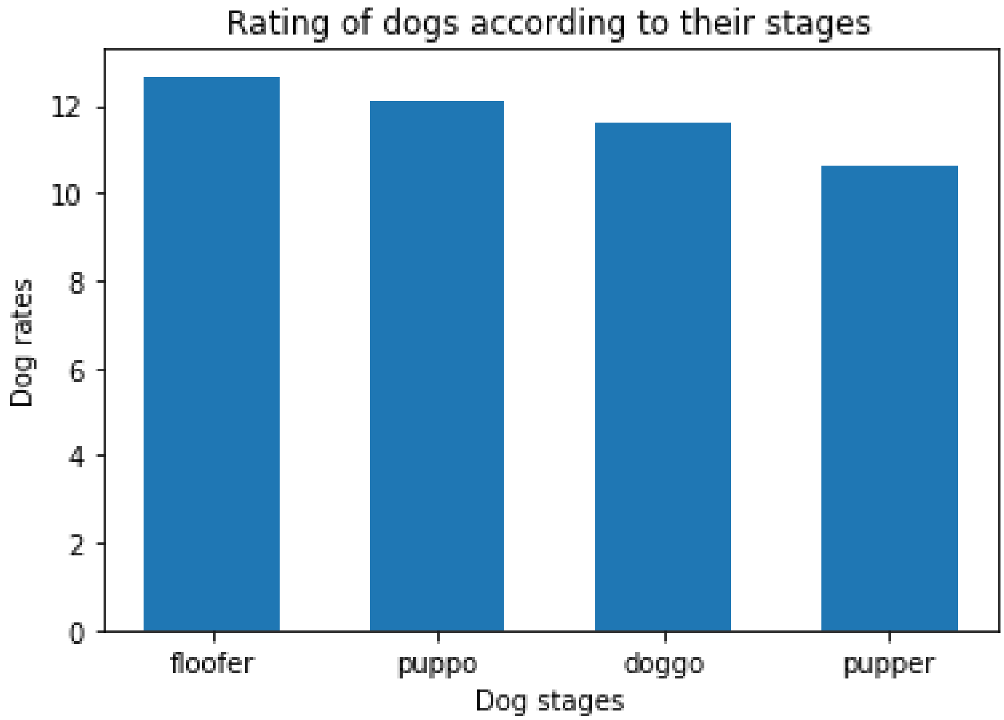
The mean rate of the dogs is 10.587/10
As per rating distribution, data show that most dogs are rated higher than 10/10 where the most rating varies from 9 to 13 over 10.

3. Most 10 common names between dogs



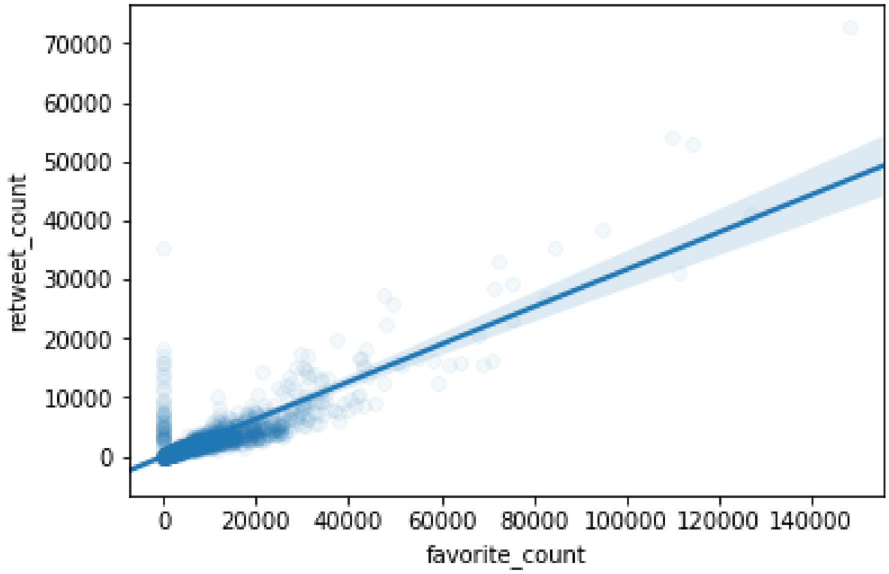
Most dogs are named: Tucker, Penny, Oliver, Cooper and Charlie.

4. Highly rated dog stages



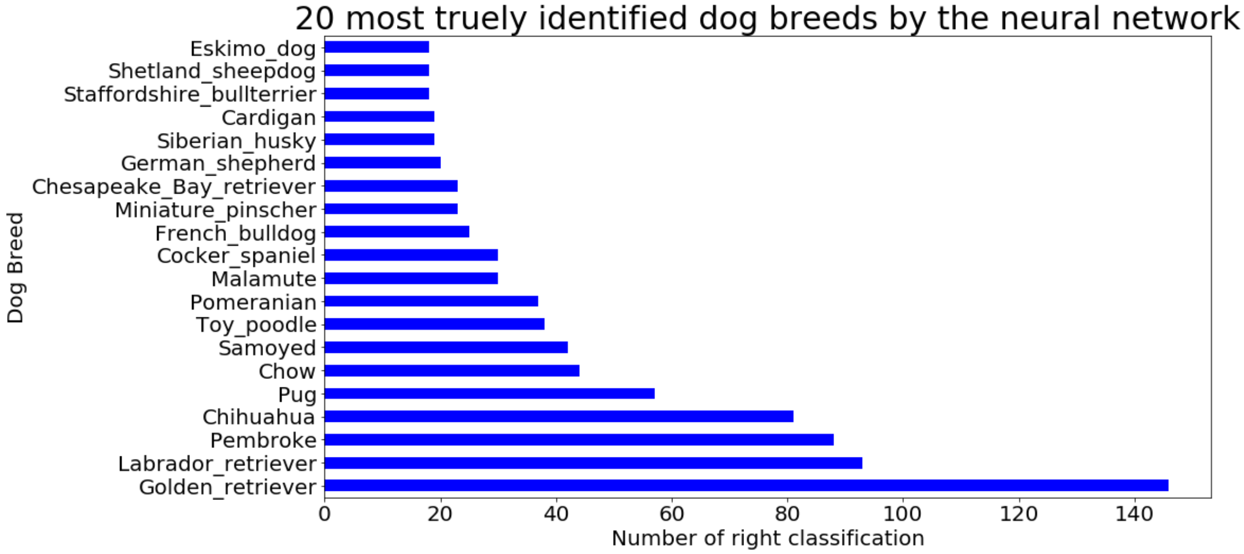
It seems that when the dog is in the floofer stage is become more rated than other stages like: "puppo", "doggo", and "pupper".

5. Favorite count and retweet count



The correlation between the favorite count and the retweet count is a positive strong correlation, meaning that tweets which are the favorite ones are more likely to be retweeted.

6. **20 most breed dog truly predicted by the algorithm**



Golden_retriever is the highest dog breed truly recognized by the neural network according to the first algorithm.