Data Analyst Nanodegree

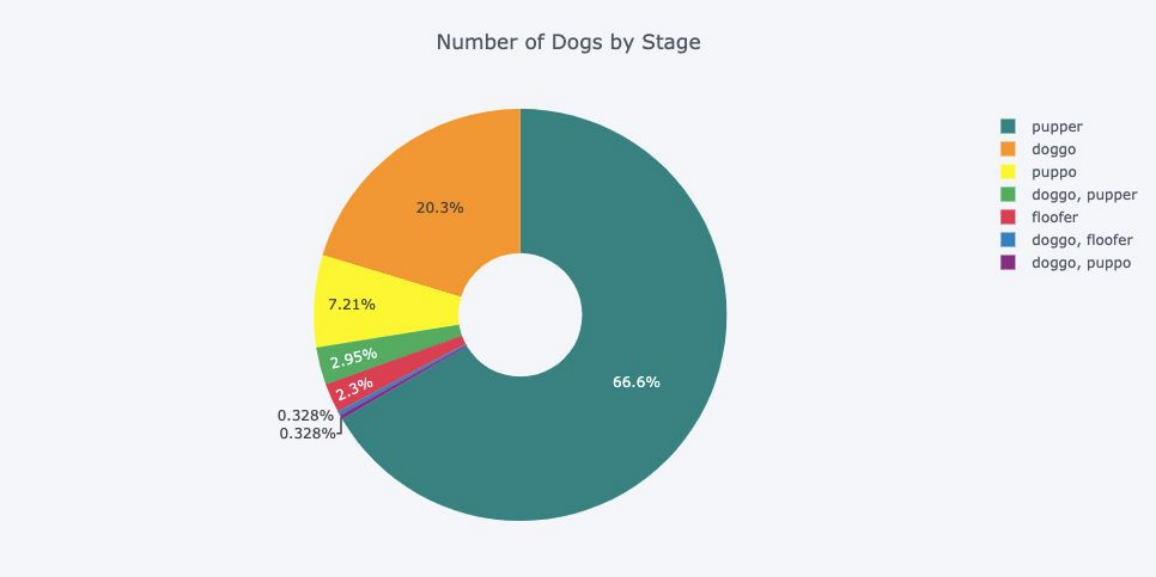
Project Report: Wrangle, Analyze and Visualize *WeRateDogs*

Introduction

WeRateDogs, the tweet archive of Twitter user @dog_rates, is a Twitter account that rates people's dogs with a humorous comment about the dog. I decided to use this archive of tweets as a dataset to practice my data wrangling skills recently acquired from Udacity's Data Analyst Nanodegree. Here I showcase key insights I found from analyzing this dataset.

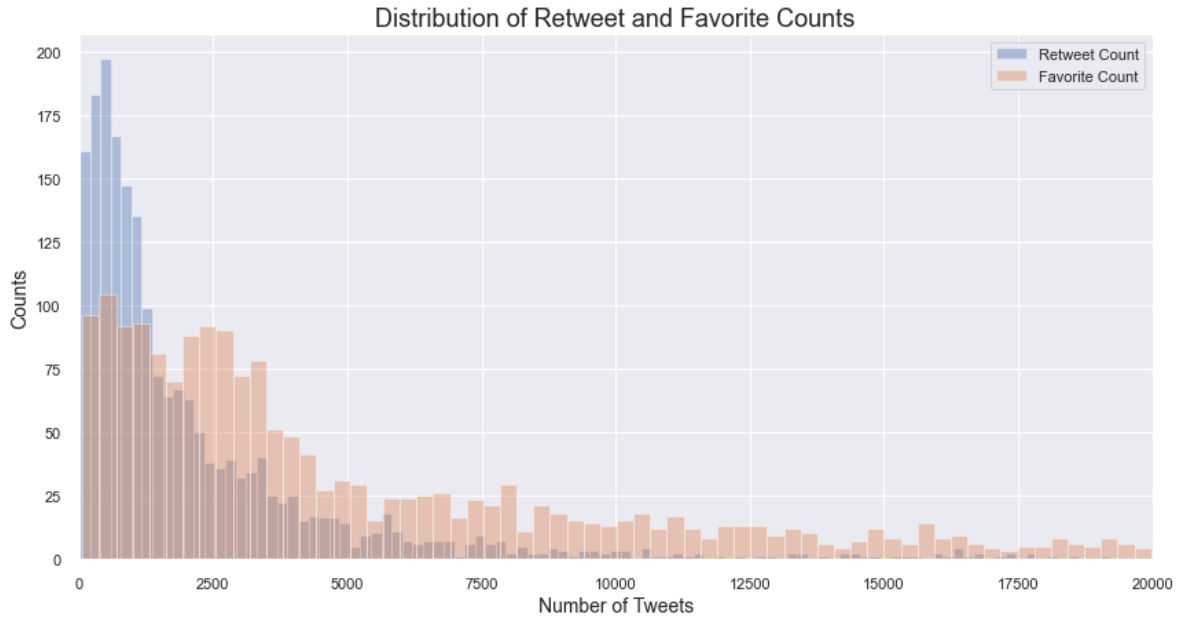
Dog Stages

Apparently, dogs can be classified into these four stages: pupper, doggo, puppo and floofer. Yep, I definitely didn't know about this before taking this course. Thanks Udacity. Anyway, I made a pie chart to see which stage is most dogs in, and pupper is the winner.



Distribution of Retweets and Favorites Counts

Next, I proceed to plot the histogram of the column *retweet_count* and *favorite_count* just to see their distribution. I found that both distributions are highly skewed, which tells me that the main bulk of the tweets have relatively low retweet counts and favorite counts.

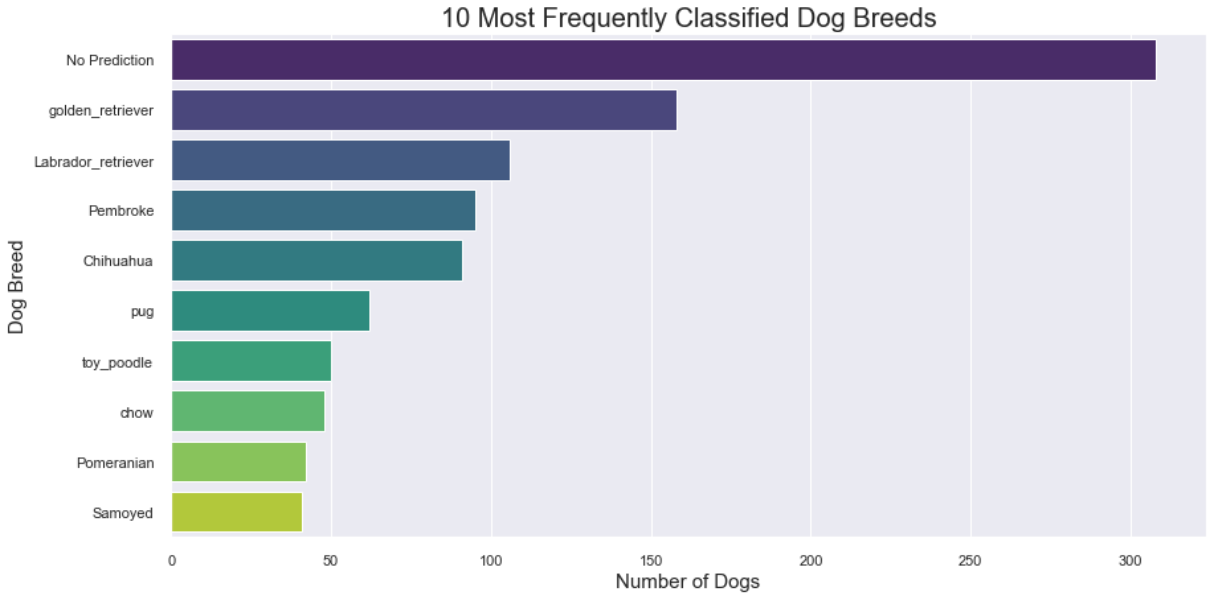


The Lucky Doggy

Curious about which dog has the highest retweet and favorite count? Well, here you are.

It totally makes sense why he/she is the winner.

The 10 most frequently classified dog breed



Shown here is a horizontal barchart of the 10 most frequently classified dog breeds. The predictions are done by a convolutional neural network (CNN)! It appears that besides the "No Predictions", the most frequently predicted breed is golden retriever.

The Timeline of WeRateDogs

Lastly, I created a line plot showing a timeline of the WeRateDogs movement. It appears that the popularity of the movement peaked in Jan 2016, and soon plunged, though having 50-100 monthly tweets after the peak and lasted for almost two years.