

Wrangle and Analyze Data

Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs has over 4 million followers and has received international media coverage. In this project, data wrangling and analyzing will be performed on three datasets, which includes the tweet archive of Twitter user @dog_rates (WeRateDogs), image predictions generated from neural network to classify the breed of dogs and additional WeRateDogs data captured via the Twitter API. Objective is to practise data gathering, accessing, cleaning and uncovering meaning findings in WeRateDog posts through data visualization.

Insights and Data Visualization

From the data wrangling, there are some interesting findings explored from data analyzing and visualization.

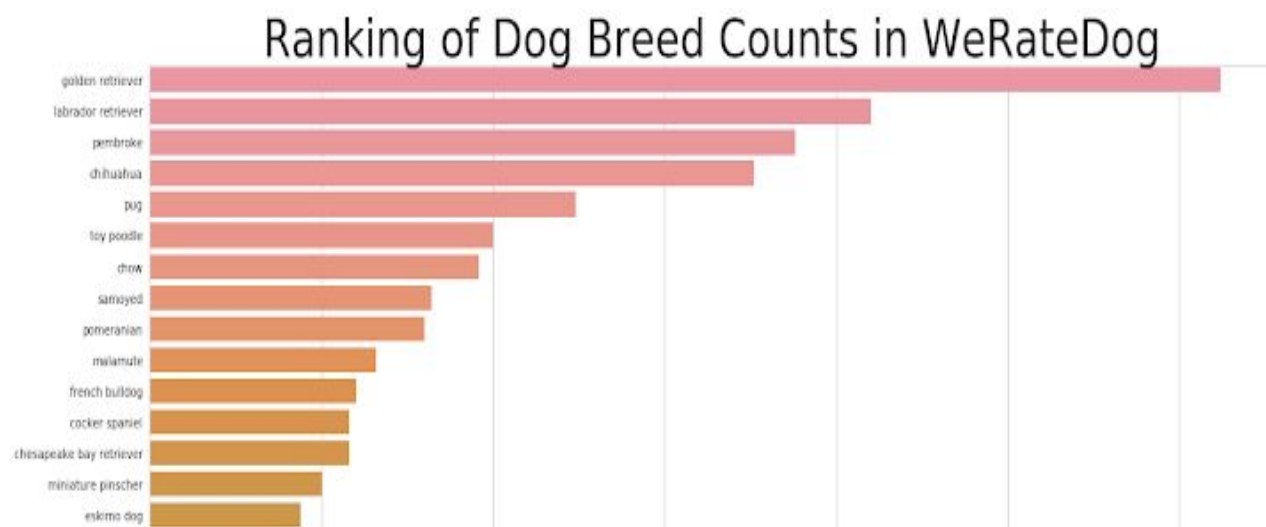
Insight 1: The most frequently posted dog is Golden Retriever!



Golden retriever, labrador retriever and pembroke (Pembroke Welsh Corgi) turned out to be the most popular dog breed in WeRateDog, they are on the top 3 in terms of number of dog breed counts. And those dog breeds match the most popular dog breed in 2017.

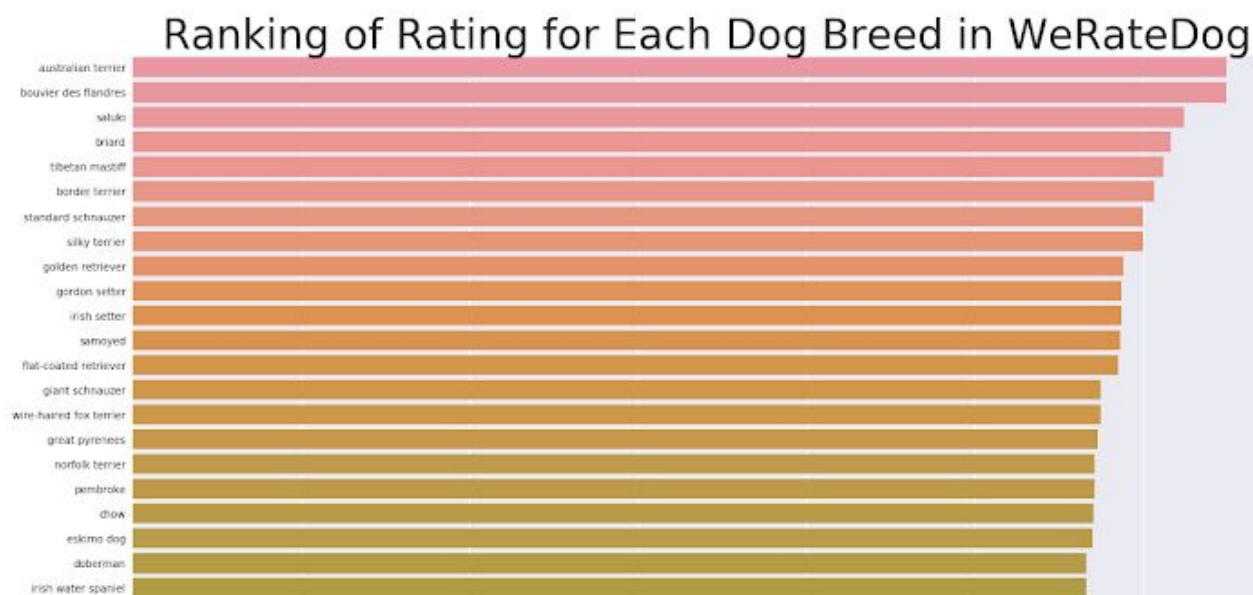
The top rankings for the most popular dog breeds of 2017 stand as:

1. Labrador Retriever
2. German Shepherd
3. Golden Retriever
4. French Bulldog
5. Bulldog



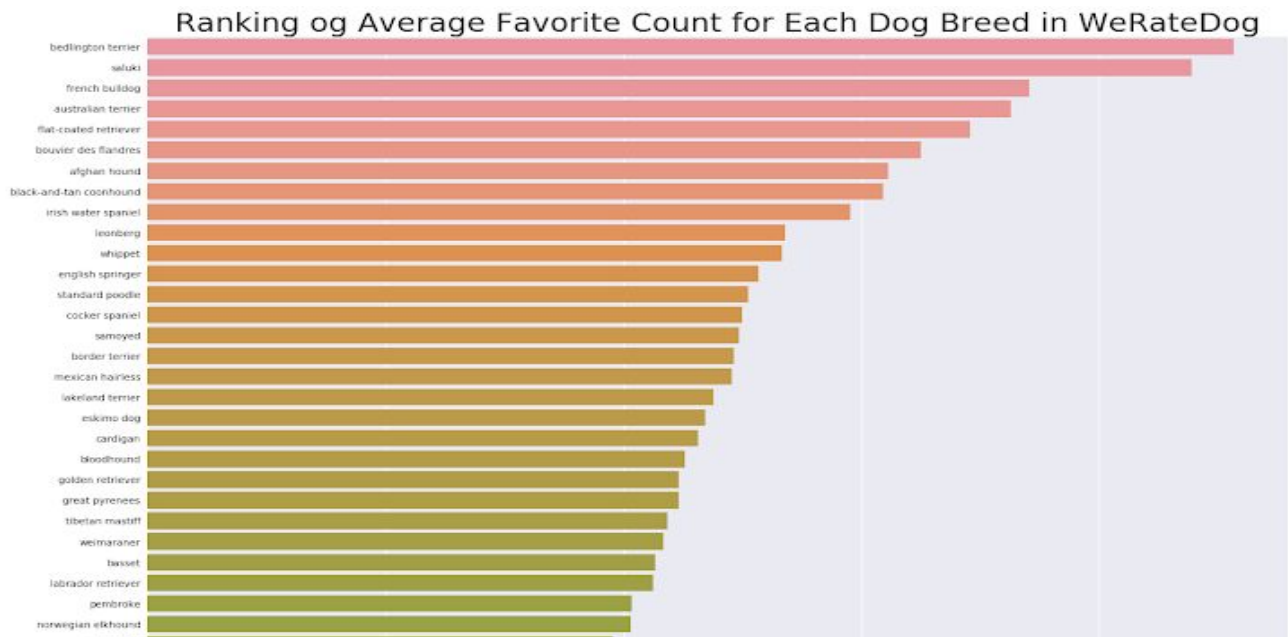
(Graph is cropped to show the top ranking only)

Insight 2: Australian terrier and bouvier des terrier have the highest rating score in WeRateDog with 13/10, followed by saluki with rating of 12.5/10 and briard, 12.33/10. But noticed the frequently

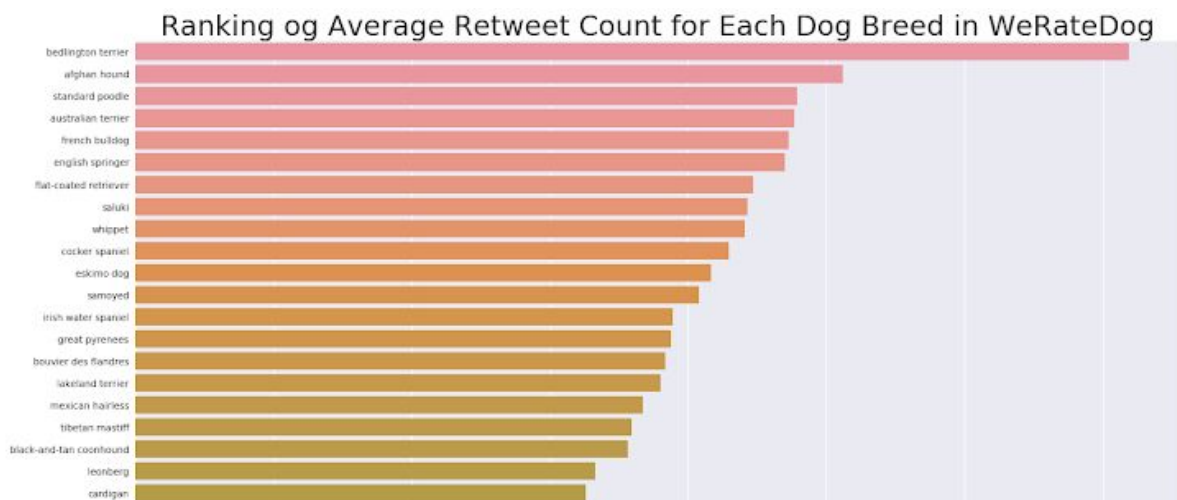


posted dog breed like Golden Retriever and Labrador Retriever are not in the top ranking in terms of rating from WeRateDog, this may due to some of the dog breeds are competition level and not common in pet dog breed, it it easier to obtain a higher rating.

Insight 3: Bedlington terrier, saluki and french bulldog received the most favorite counts over all other dog breeds.



Insight 4: Bedlington terrier, afghan hound and standard poodle ranked top 3 in terms of number of retweets.



Insight 5: Based on the dog breed ranking in favorite count and retweet count, noticed some of the dog breed appears repeated in the top list, and the joinplot between favorite count and retweet count shows some positive interaction, which matches real world observation, favorite and retweet are correlated.

