

Act Report – WeRateDogs Data

Gabriela Trindade de Almeida

July 28th, 2021

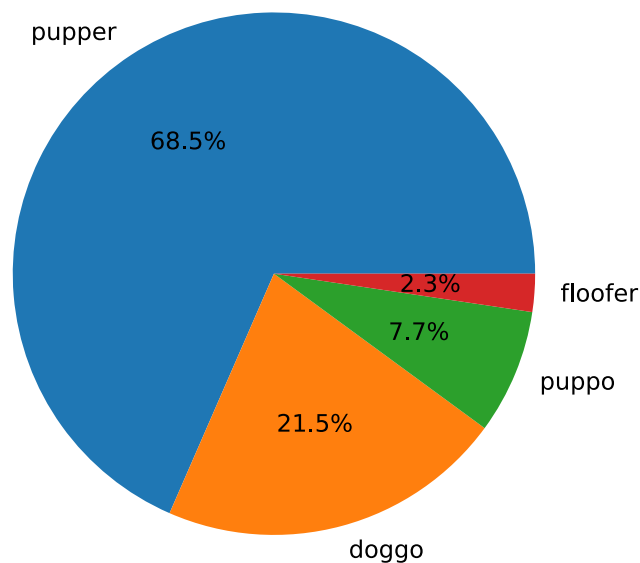
This project was written for Udacity's Data Analyst nanodegree program. The primary goal is to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. This is tweet archive's (November 2015 - August 2017) dataset for Tweeter user @dog_rates, also known as WeRateDogs, which is an account that rates people's dogs with a fun comment about each pet.

For this purpose, it was necessary to gather data from a variety of sources and file formats. After assessing and cleaning these datasets for data quality issues and lack of tidiness, the following insights and visualizations were produced.

1) Dog stages

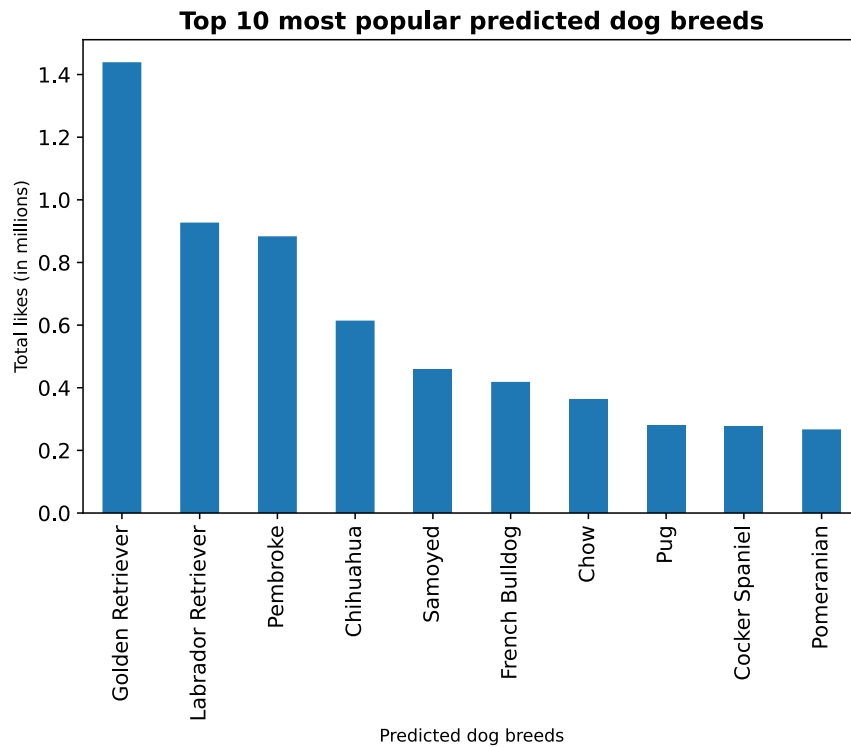
This pie chart represents the percentage of each dog's stage in the dataframe, considering only the stages different from "None". It is possible to notice that the majority of the classified dogs in this dataframe are puppies, with 68.5%, followed by doggo, with 21.5%.

Percentage of different dog stages



2) Popular dog breeds

This bar chart represents the predicted dog breeds with the top ten highest total of "likes". It is possible to notice that the dogs predicted as Golden Retriever received the highest amount of "likes", with 1.4 million total, followed by Labrador Retriever, with 0.9 million, and Pembroke, with 0.8 million.



3) Retweeted dog breeds

This bar chart represents the predicted dog breeds with the top ten highest total of retweets. It is possible to notice that the dogs predicted as Golden Retriever were the most retweeted, with 398 thousands of retweets, also followed by Labrador Retriever, 287 thousands, and Pembroke, 233 thousands.

