WeRateDogs Twitter Archive - Act Report

By Tarek Abdel Rahman

Introduction:

WeRateDogs is a famous twitter handle which rates dog in an unusual and humorous manner(They even rated Snoop Dogg, 420/10)



It became famous after a meme went viral.



And they appreciate creativity, creatively, like rating 666 to dog dressed as devil or rating 1776 on 4th July



Over the time, amount of tweets per month decreased and popularity (retweets and favourites) increased. Currently it has more than 4.3 million followers.

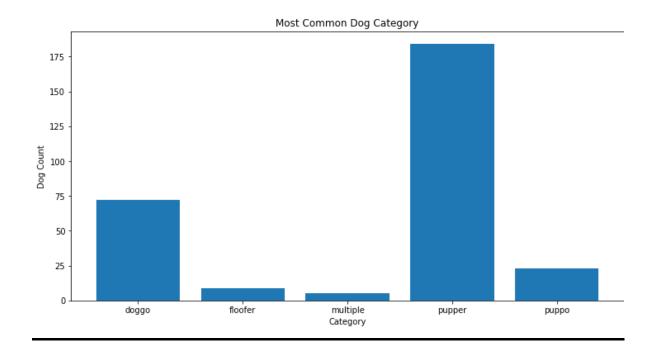
WeRateDogs Data:

The WeRateDogs Enhanced Twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account, posted between November 15, 2015 and August 1, 217. This data is comprised of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present. The retweet count and favourite count for each tweet were not included in the enhanced archive, and so I had to download this additional data from the twitter account using the tweet ID from the archive. Along with the Twitter data, I also downloaded an image predictions file from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets (there can be up to 4).

Wrangling Data:

Before I could begin the analysis, the data had to be wrangled into shape to make it easier. I assessed the data both visually and programmatically for quality and tidiness; the quality of data is determined mainly by looking at several aspects or dimensions to ensure that it is complete, valid, accurate and consistent. After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data.

Analysis:



Insights:

- The most popular dog type is a "pupper"
- The most popular dog, based on image predictions, is a Golden Retriever
- The median and mean ratings out of ten are 11 and 12.7 respectively
- The median and mean retweet count is 1337 and 2837 respectively
- The median and mean favorite count is 3744 and 8490 respectively