WeRateDogs Twitter Archive - Act Report

WeRateDogs Dataset

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The WeRateDogs Enhanced Twitter archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

Data Wrangling

My goal was to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Therefore, additional gathering, then assessing and cleaning was done for "*Wow!*"-worthy analyses and visualizations.
I gathered from Twitter's API retweet count and favorite count which are two of the notable column omissions. I also downloaded programmatically using the Requests library an Image Predictions File that can classify breeds of dogs.
After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data that was finally ready to investigate and analyze

Insights

1. **Top rated dog/s**

```
dog_stage
doggo, puppo      13.000000
floofer           12.000000
doggo             11.888889
doggo, floofer    11.000000
doggo, pupper     11.000000
Name: rating, dtype: float64
```
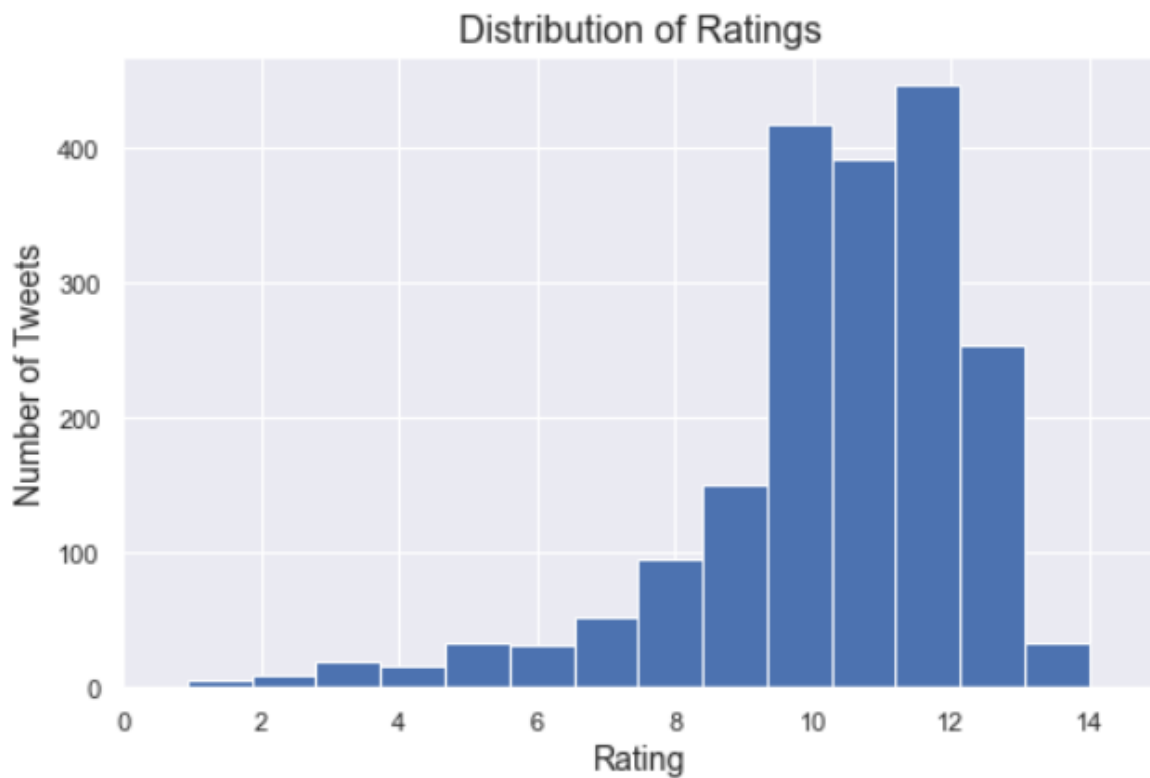
→ doggo & puppo are the most rated dog stages. Can't blame them, they are the cutest.

2. **Common dog breeds**

From the wordcloud it is understood that **golden_retriever**, **Labrador_retriever**, **Pembroke**, and **Chihuahua** are the most common dog breeds.

3. Distribution of Ratings

The distribution of ratings is very skewed to the left. From the descriptive statistics above we see that 75% of all ratings are between 10 and 14 inclusive (the IQR is from 10 to 12).

As the twitter account replied to one user, "They're good dogs (Brent)!"