

WeRateDogs blog post for the DAND by Udacity

by Japheth Rutoh

Wikipedia describes the WeRateDogs Twitter page as an account that rates people's dogs with humorous content about the dog.

The data for this project was provided by Udacity. This was part of the Data Analyst Nanodegree.

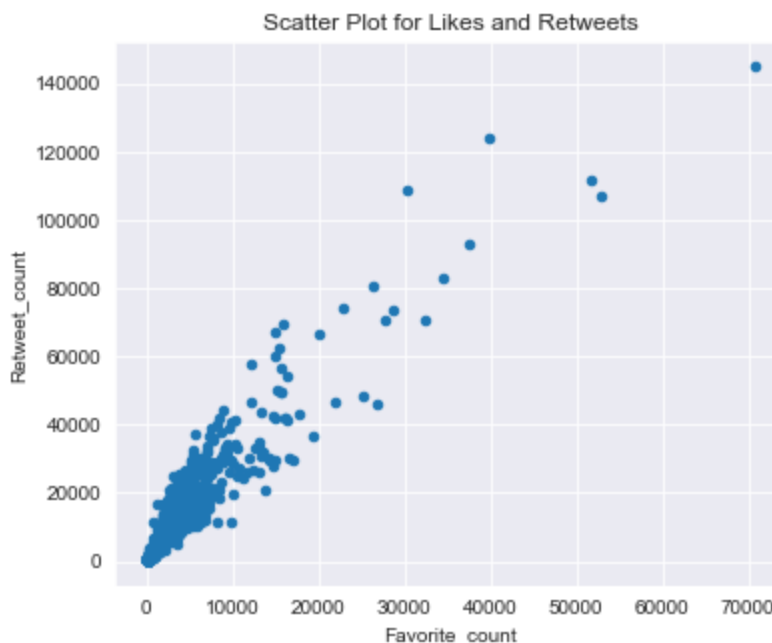
The dataset was made up of tweets by the WeRateDogs Twitter, amounting upto 5000+ tweets. The final analysed and cleaned dataset has 1951 entries and this is the one that is used to come up with insights for this blog.

The following are some insights I found after cleaning the data:

- The most common dog names are Cooper, Oliver, Charlie, Lucy and Tucker. It seems that many dogs have similar names with Cooper being the most common.
- Twitter for iPhone is the most common tweet source. Twitter for iPhone is the most common tweet source by a large mile.
- 38 dogs receive the highest rating of 14 out of 10; weird system rating :-) - but understandable.
- The pupper stage is the most common in the dataset while 1731 cases have unknown dog stages. Puppies are generally tiny dogs, and it seems people like them tiny and cute.

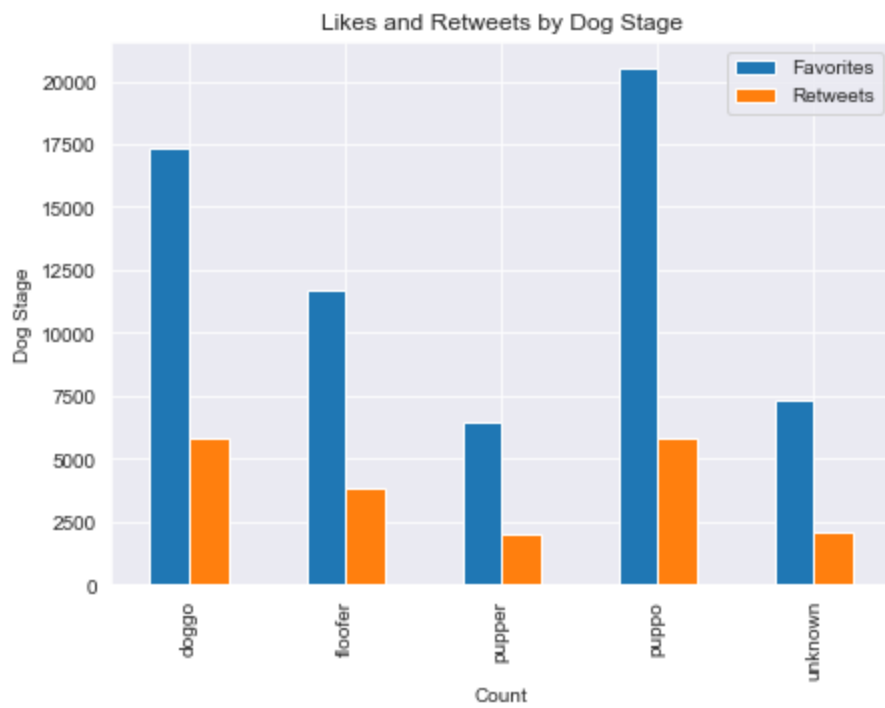
Visual Insights

1. What is the relationship between number of likes and number of retweets?



For this I plotted the number of retweets on the y-axis and the like count on the x-axis, and the picture above was the result. This plot shows that there is a positive relationship between number of likes a tweet gets and number of retweets. The number of likes generally increase with increasing retweets.

2. Which dog stage receives more likes and retweets?



For this I needed to see the how many likes and retweets a tweet for a dog at a certain stage gets; therefore I plotted a bargraph with likes and retweets for each side by side. The plot above showed that the Puppo stage had more likes followed by the Doggo stage with Floofer and Pupper stage following respectively. For retweets though, Doggo and Puppo stages were very closely matched.