Analysis of @WeRateDogs

Marc Hilty

07.01.2018

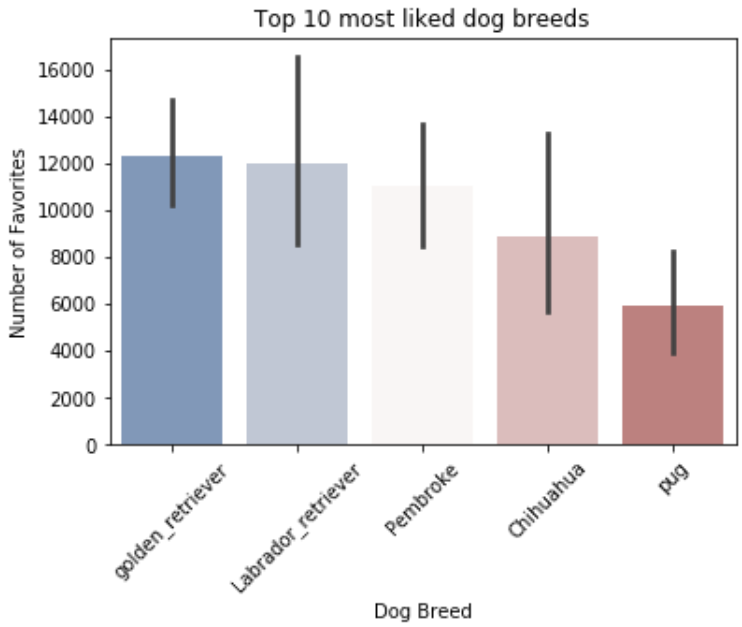
General Information

The data was provided by the twitter user @WeRateDogs, Udacity and additional information was gathered from the twitter API. This project is part of Udacity's Data Analyst Nanodegree program.

Analyze

This dataset provides a lot of material to work with. I defined relevant question before analyzing them in detail. The jupyter notebook contains a lot more questions I asked myself, but didn't answer since it would burst the limits of this project.

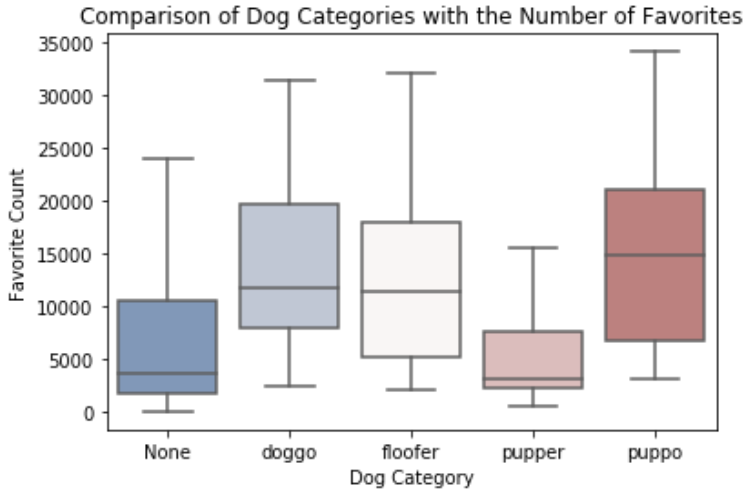
1. What is the most liked dog breed on @WeRateDogs



This graph is generated using only the top 5 most tweeted dogs from @WeRateDogs. It programmatically looks for the most posted dogs and uses them to plot the graph. I chose to include only the top 5 of most posted dogs instead of all dogs, because some dog breeds are only posted once or twice while the most posted are in the hundreds and the significance level of the graph therefore is much higher. I used the number of favorites for each tweet as an independent measurement of general appeal of the dog breed.
Because I plotted with seaborn it automatically bootstrapped the data in the barplot and shows the 95% confidence interval as a black bar directly in the graph.

We can say that the breed golden_retriever usually results in the highest number of favorites.

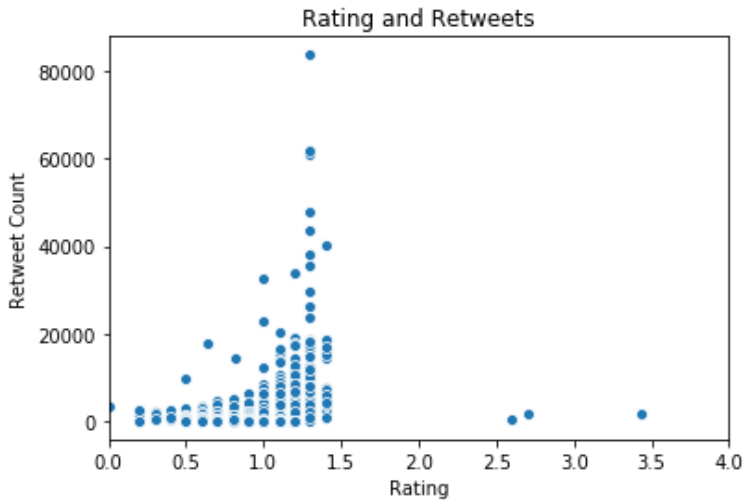
2. Which of the dog categories @WeRateDogs uses is the most liked?



This boxplot graph allows us to compare the dog_categories @WeRateDogs. Again I used the number of favorites as a measurement.

Doggos and puppos generally are favorited most often, followed by floofers. Interestingly puppers aren't favorited as often as the other categories. When a dog is not categorized it generally is favorited less often than doggos, floofers or puppos, but not puppers.

3. Does @WeRateDogs rating affect the number of retweets?



I calculated the individual ratings using the rating_numerator and rating_denominator. The retweet_count subsitutes the favorite count in this plot since during another analyse, which is not present in this file but in the wrangle_act.ipynb. It shows that the retweet_count and favorite_count strongly correlate with one another.

We can say that there is at least a weak positive correlation between the rating @WeRateDogs gives us and the number of retweets.