Data Analysis Nanodegree-2020

WeRateDogs Twitter Archive - Analysis and Visualization

Sarah Alhomoud

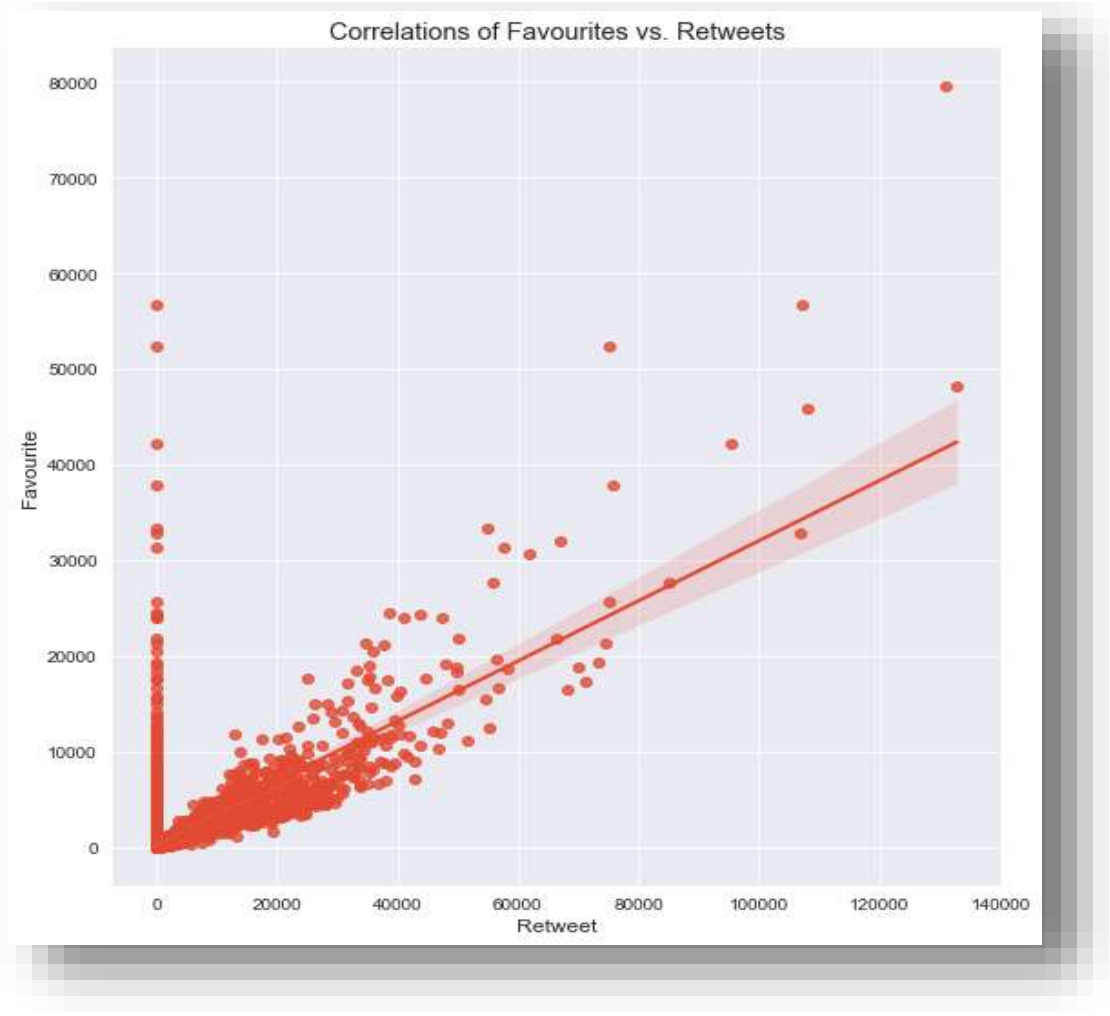
Data Analysis Nanodegree-2020

Introduction:

The dataset will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

- **Is there Correlation between Favorite and Retweet ?**

We want to determine if there is correlations in our data to find out any interesting relationships. Since the retweet and favorite are ways to express opinion in twitter and types of measurement, we assume there is correlation between them; lets test the theory?



Finding:

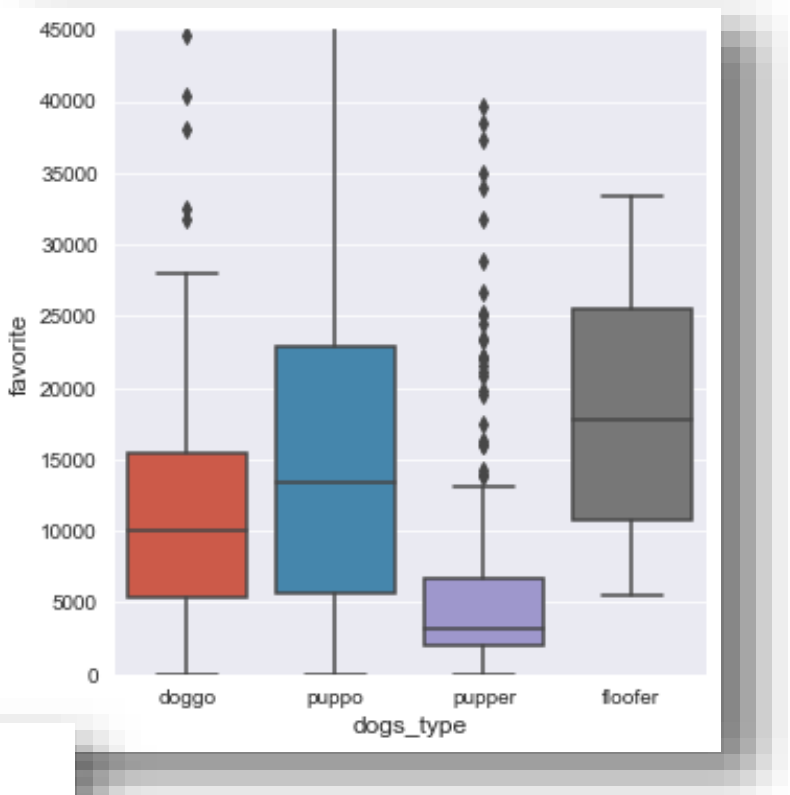
- We can notice there is a strong , positive liner relationship b\w Favorites and retweets as we assume.

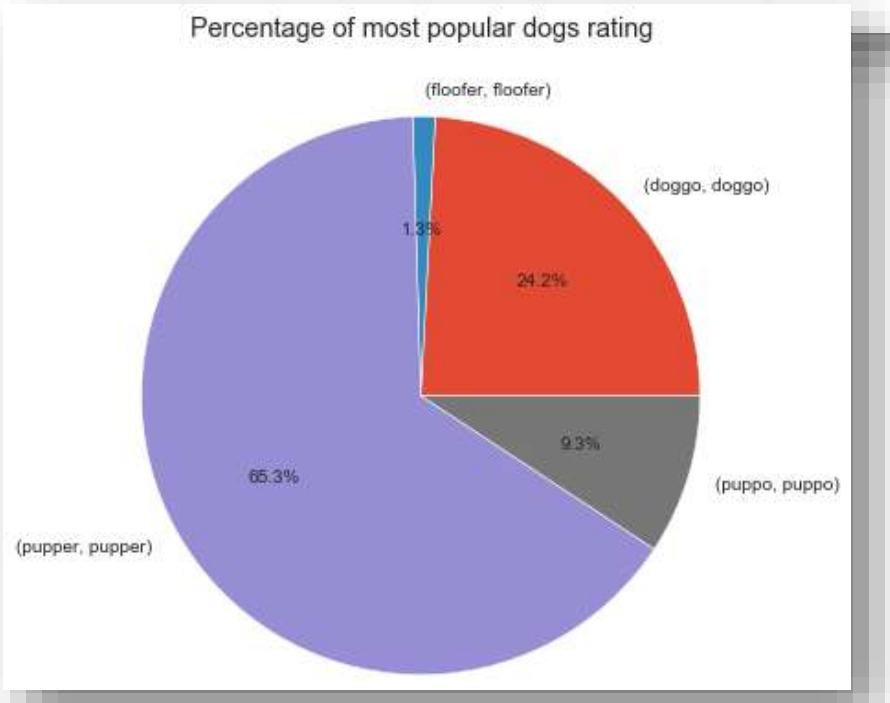
- **Which type of dogs is most popular, and more favorable by people ?**

There are four types of dogs in the dataset [doggo, pupper, fluffer, puppo]) , and I need to determine which type is more popular based on rating . And is the 1st popular dog type is also gets favorite than others?

Findings:

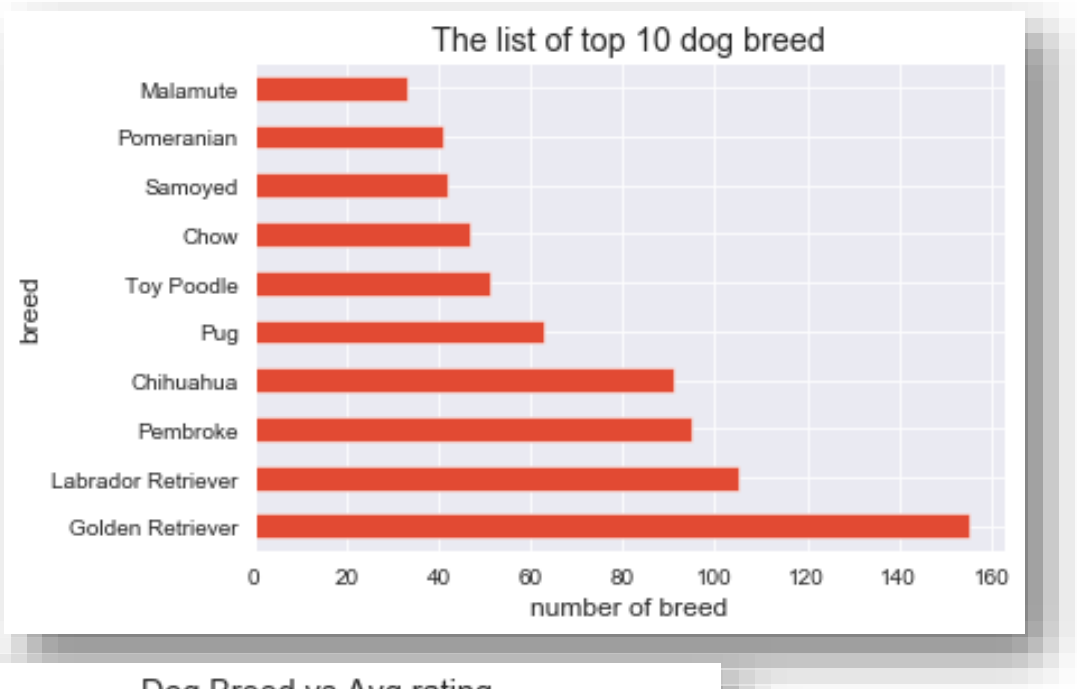
- As a result , the **Puppos** are the most popular type on average rating among the others by 65.7 % which is high percentage , after that doggo type by 24.2%.
- And based on the plotted points , pupper less favorable among the other types and has greater outlier than other. Where floffer then puppo are the most favorite one.

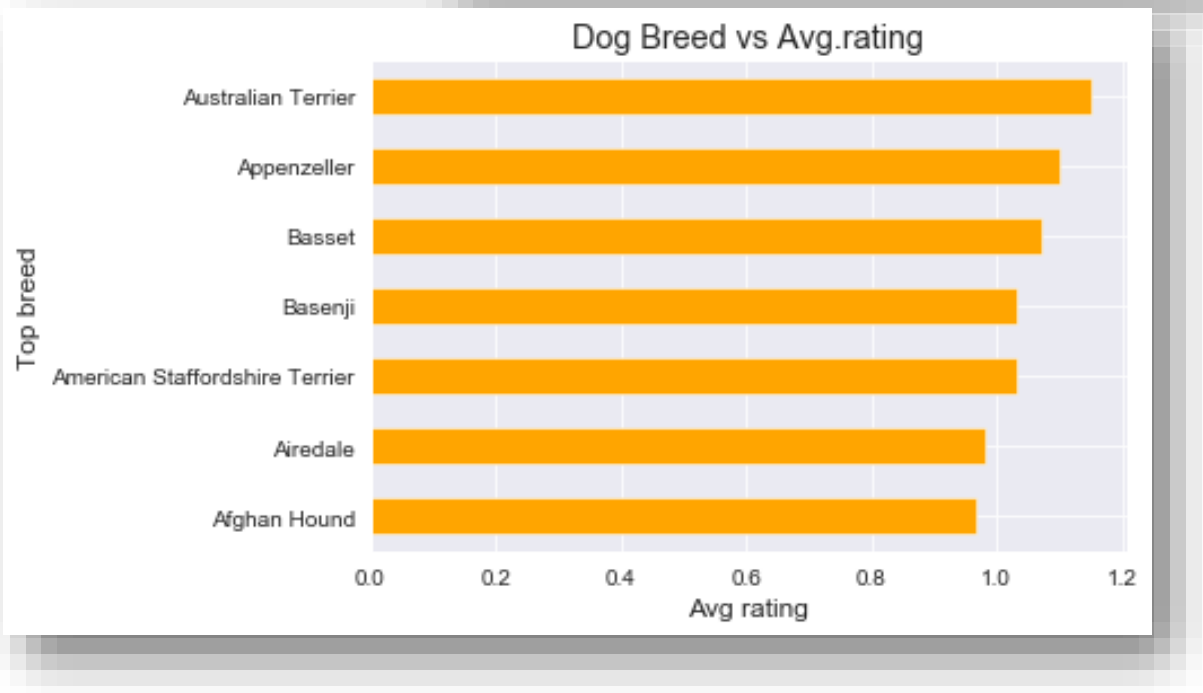




- **Which dog prediction is popular? And who gets high rating ?**

There are three predictions in the dataset, so in order to make it easy to compare I combine them to one column
called breed.





Findings:

- Golden Retriever is the most common breed with almost 160 count, Labrador Retriever comes in the second level of popularity with almost 110 count.
- Based on the rating , Australian Terrier gets the highest rating with average ~ 1.15. then Appenzeller with 1.10 rating.