Mathieu Rella

Analysis Act Report

We Rate Dogs

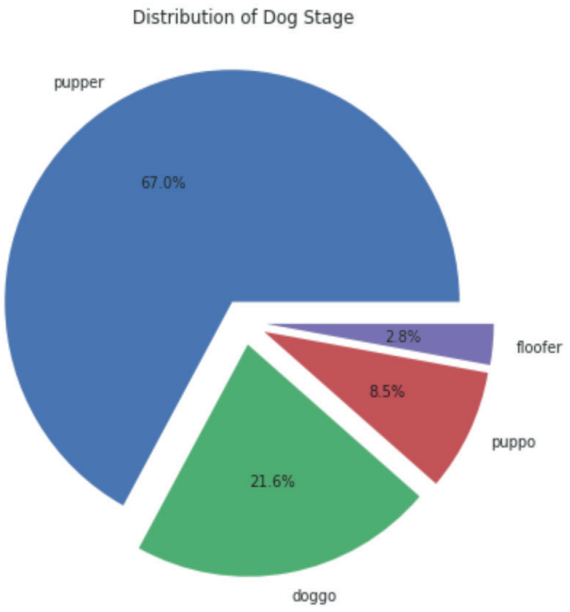
UDACITY | **Data Analyst Nanodegree**

This analysis report comes from a dataframe created and cleaned from a preliminary work of data wrangling.

This report will provide you with 4 insight and visualization allowing you to understand and establish correlations of this dataframe.

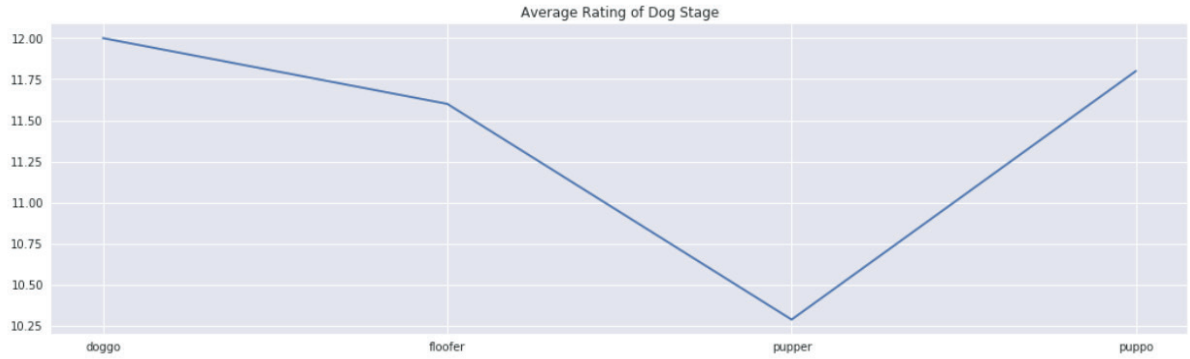
Data Analysis

I. Distribution of the stage

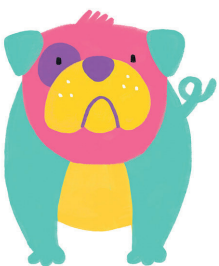


This pie chart show us the stage distribution of the dataset with over 67% of pupper tweet.

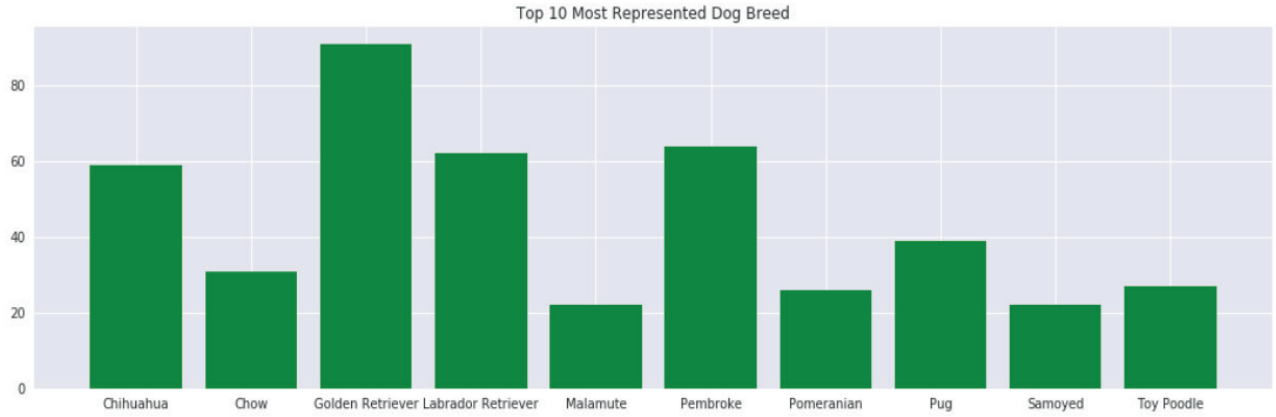
II. Average Stage Rating



The delta in the rating is only of 1.75, which is surprising because the data-set is populated for 67% of pupper and only 22% of doggo.
At first sight I would admit a bigger difference in ratings

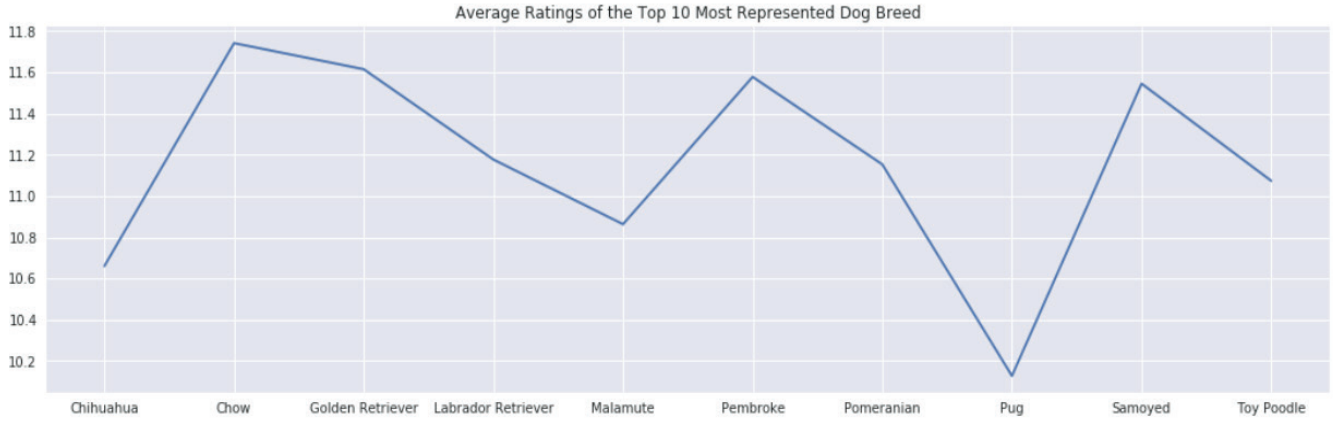


III. Top 10 Dogs Breeds

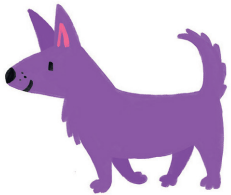


In this plot we see the 10 most breed represented in our dataset, The golden retriever(>90) is the most represented.

IV. Average Ratings of Top 10 Dogs Breeds

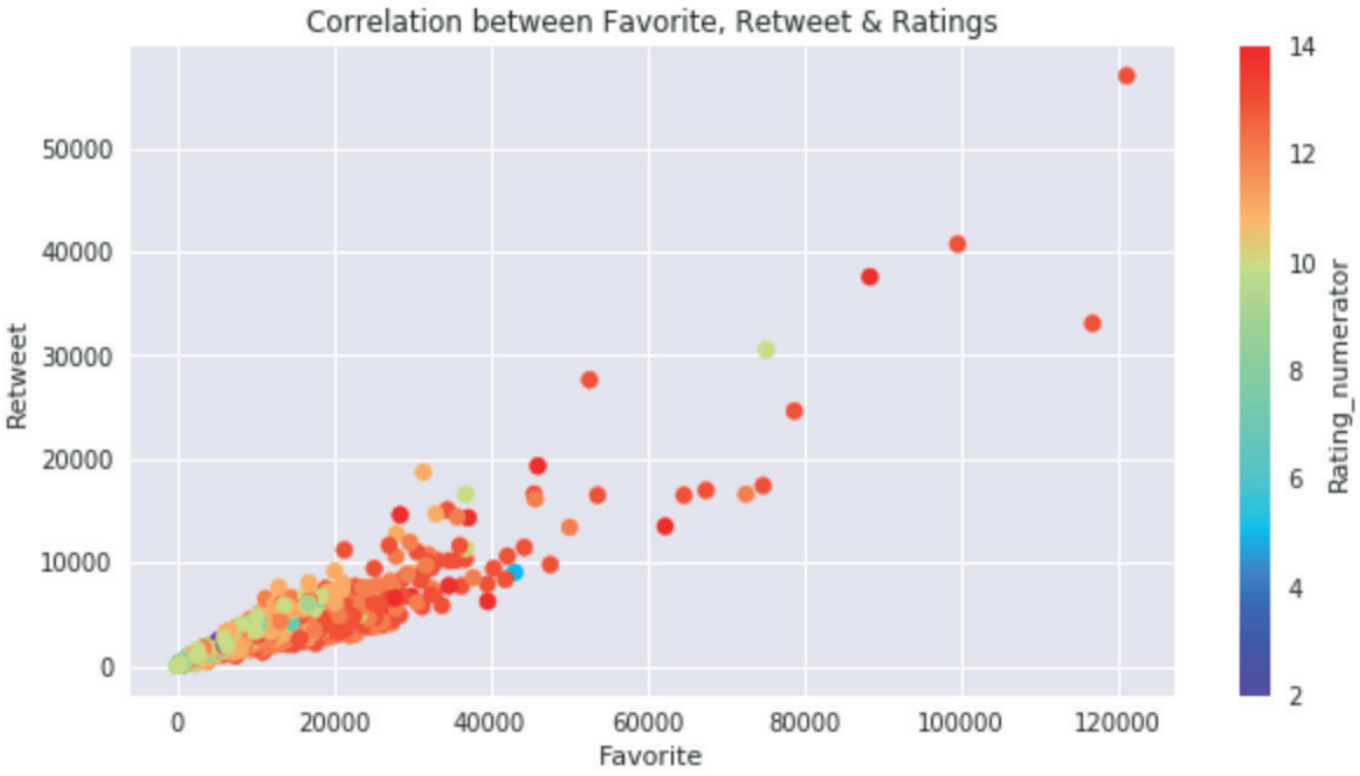


As seen with the stage average rating,The delta of the breed average rating is approximately the same (1.7), which is less astonishing since the density of dog breeds is more extensive than the dog stage.



V.Correlation between Favorite, Retweet & ratings

before establishing this graph I looked for the 5 tweet with the most favorites and retweet except 1 that did not correspond, the rest were the same, the following graph is the result of this observation to verify this correlation with the rating index.



This plot show us the positive correlation between retweet and correlation.

So the more likely a person who likes a tweet the more likely they are to retweet it.

We roughly see that's the more a tweet have rt and fav the more is rating is good (red)

On the 1341 tweet more than the half of them oscillates between 0 & 40000 favorites.

Finnaly on this plot we clearly see that retweet have a higher value for the user, their distribution is only half as important as the favorites.