Udacity Data Analyst Nanodegree Program

wrangle and analyze data

Introduction:

This document contains visualizations and insights into the data that was gathering, assessing and cleaning activities done in the "Wrangle_report" document.

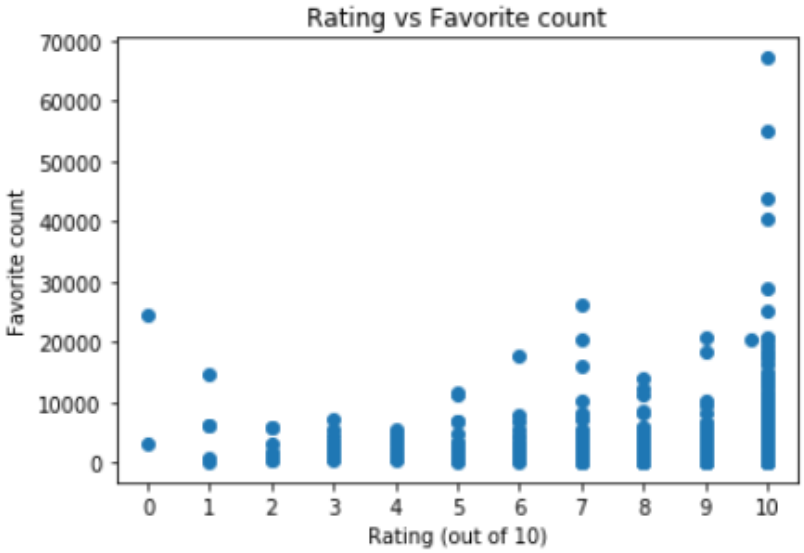
The data which was collected is a collection of tweets gathered from the user @dog_rates form twitter. Every tweet has a rating out of 10 (which can go over 10), name of the dog (if given), stage the dog (if given), a prediction of the breed (if given), the number of retweets, number of Favorite and where did the tweet come from.

In this document we will answer the following questions:

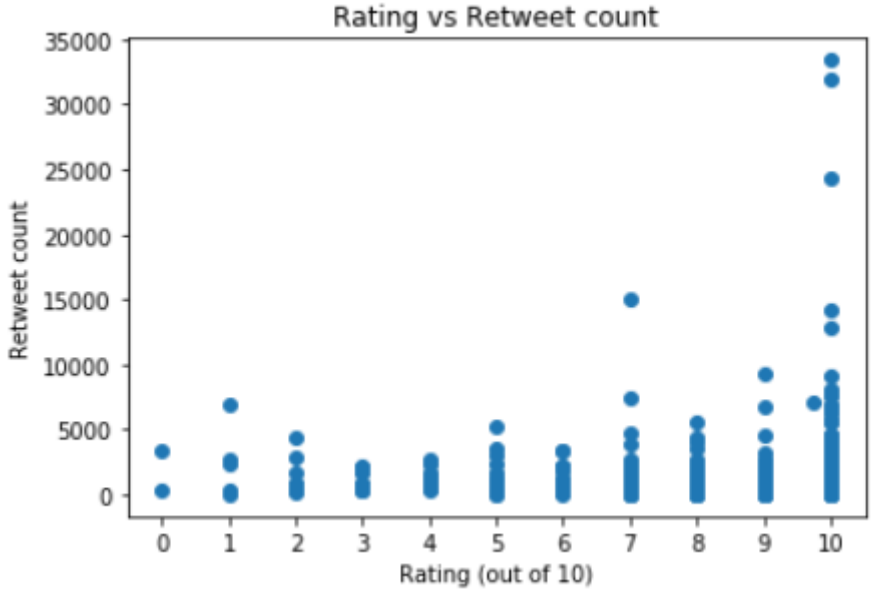
1. Where the most retweeted tweet came from?
2. Where the most favorited tweet came from?
3. Which is the most popular dog stage?
4. What is the most popular dog breed?
5. What is the most popular name?
6. Is there a relationship between the number of retweets and the dog rating?
7. Is there a relationship between the number of favorites and the dogs rating?

Results:

1. The most retweeted tweet came for an iPhone.
2. The most favorited tweet also came for an iPhone.
3. The most popular stage of a dog is pupper.
4. The most popular breed is the golden retriever.
5. The most popular name is both Charlie and Oliver.



6. The relation between the rating and the favorite count is a positive exponential relationship. (the relationship seems to be stronger than the relationship between the rating and the retweet count.)



7. The relation between the rating and the retweet count is a positive exponential relationship.