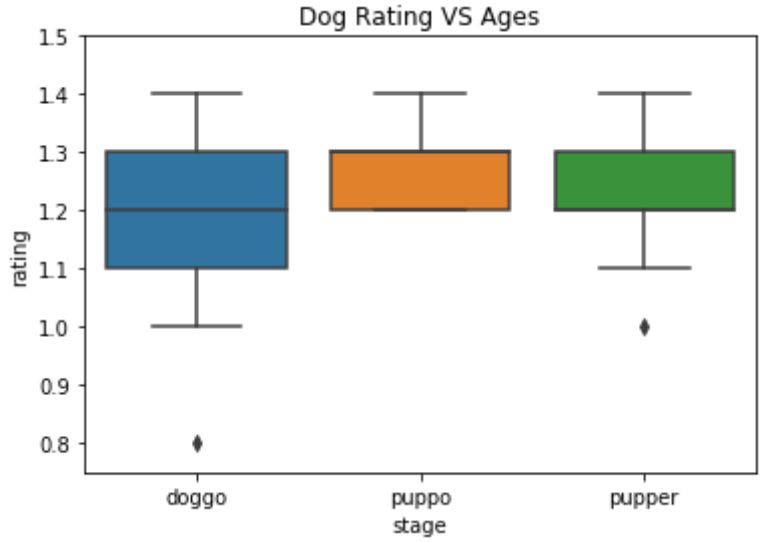
Wrangle and Analyze Data Report

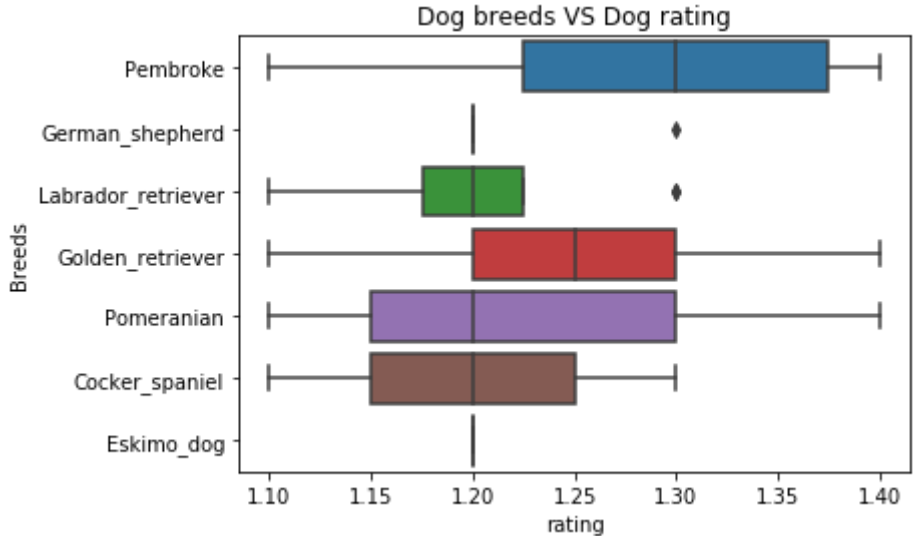
In this project, I was trying to find if there's some correlation between the dog age and the dog rating for these twitters; if there's some correlation between the dog breed and the dog rating for these twitters and if there's some correlation between the favorite counts and the dog ratings for these twitters. After analyzing and visualizing the datasets, I got something to share with you.

Insight 1:
About this insight, I was trying to find if there's some correlation between the dog age and the dog rating for these twitters.



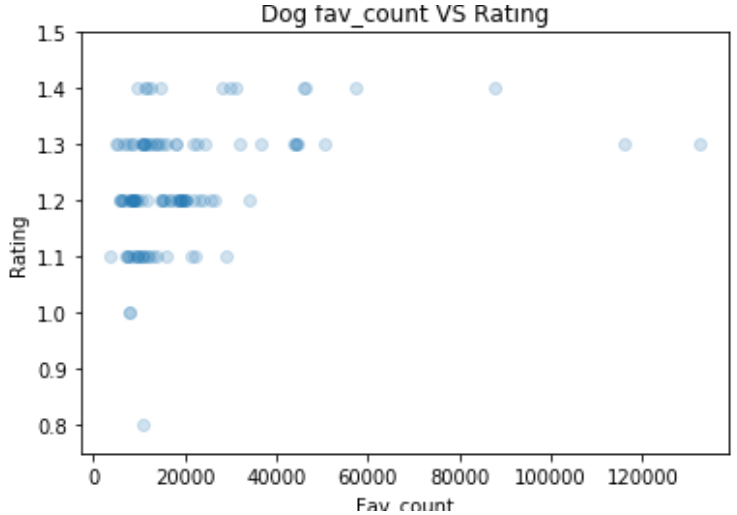
In the three stages, the ratings of the poppo stage(between old and young) are intensively distributed than the doggo and pupper; the pupper stage(the young) gets a same box range but the puppo has no bottom whisker which means it has less lower ratings. The doggo stage(the adult) gets a relatively lower ratings than pupper and poppo, and the dirtribution is much more scattered and has very low outliers, like 0.8.

Insight 2:
About this insight, I was trying to find if there's some correlation between the dog breed and the dog rating for these twitters.



In this plot, there're seven breeds showed from top to bottom using different colors. We can see there're three breeds got their whiskers all from 1.1 to 1.4, but the medians are different.Pembroke got a highest meadian and the ratings of Q1-Q3 are releatively higher than other six breeds. Pomeranian and Cocker_spaniel have the same median, but the median to Q3 of Pomeranian has a bigger range which means more higher ratings.

Insight 3:
About this insight, I was trying to find if there's some correlation between the favorite counts and the dog ratings for these twitters.



In this plot, we can see from 1.1 to 1.3 scores, when the ratings increased, the fav_count will also increase accordingly. But after 1.3 score the relationship dosen't exsit and also dogs who got above 1.3 scores are much less.

The results we got from these three analyzations may effected by the sample data we got. For more accurate results, we need to analyze if the data we collected is normally distributed.