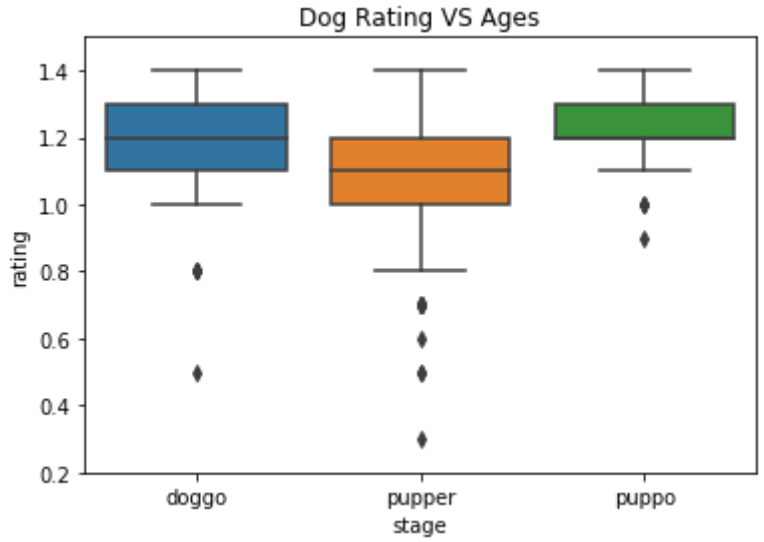
Wrangle and Analyze Data Report

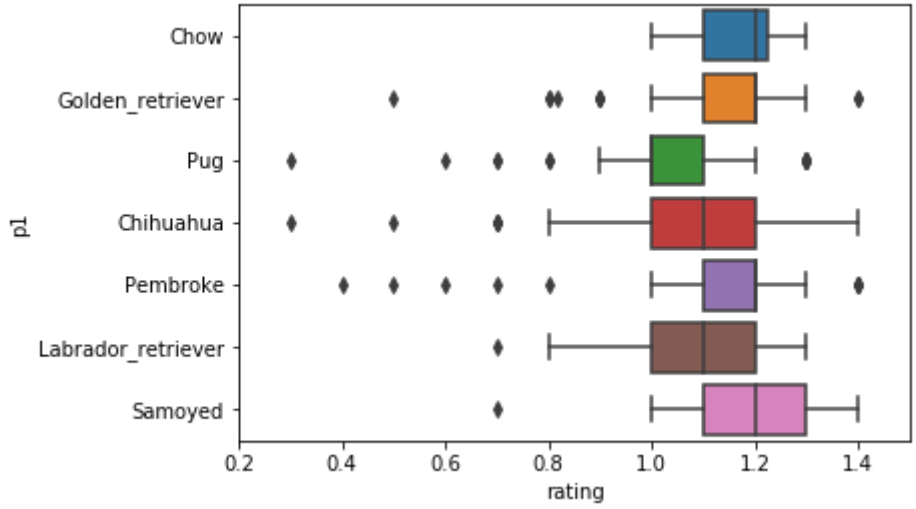
After analyzing the datasets about the dog twitters, I got three insights to share.

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.



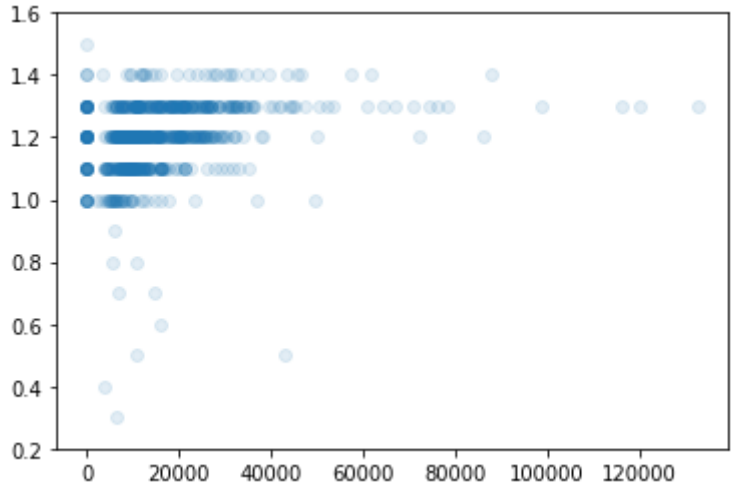
In the three stages, the poppo stage(between old and young) get the highest meadian, and the ratings are intensively distributed; the pupper stage(the young) gets a lowest median and the ratings are relatively scattered distributed, also gets some low rating outliers. And the doggo stage(the adult) gets a relatively higher median than pupper and lower than poppo, and the dirtribution is also the same. Before this analyzation, I thought people prefer to rate highly to the young puppies.

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 two breeds which get the highest median and least outliers and intensively distributed. They are "Chow" and "Samoyed" which are all floofers. This finding says that people prefer fluffy dogs.

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 0.2 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.