ACT REPORT

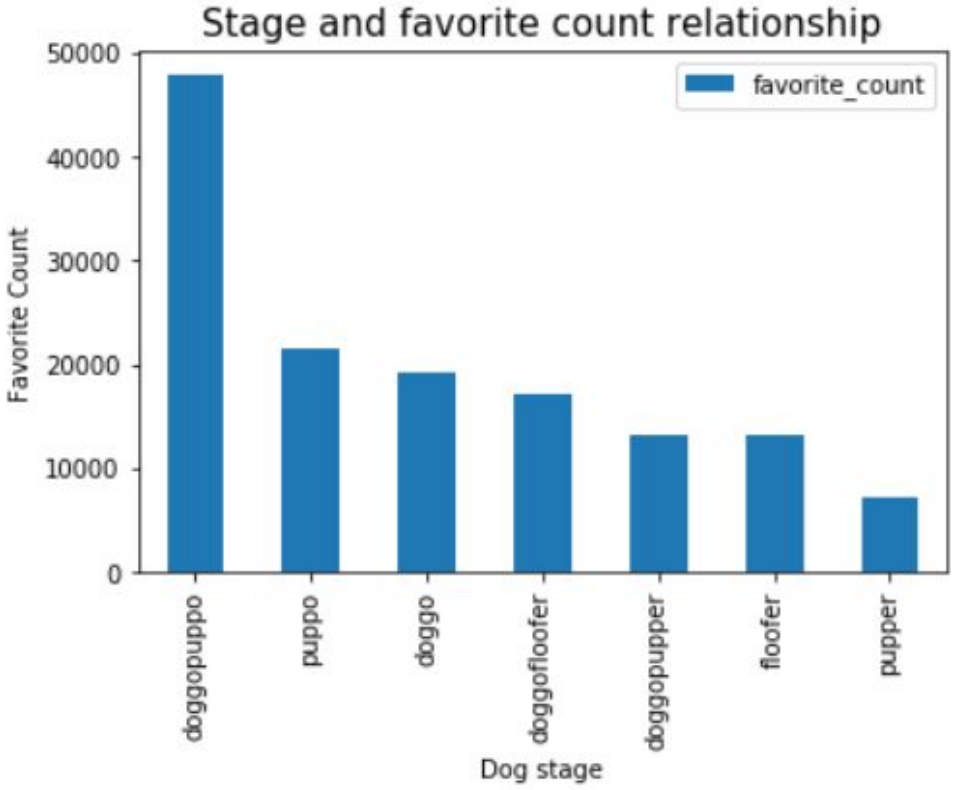
As introduced in the introduction, I will explore four different questions:

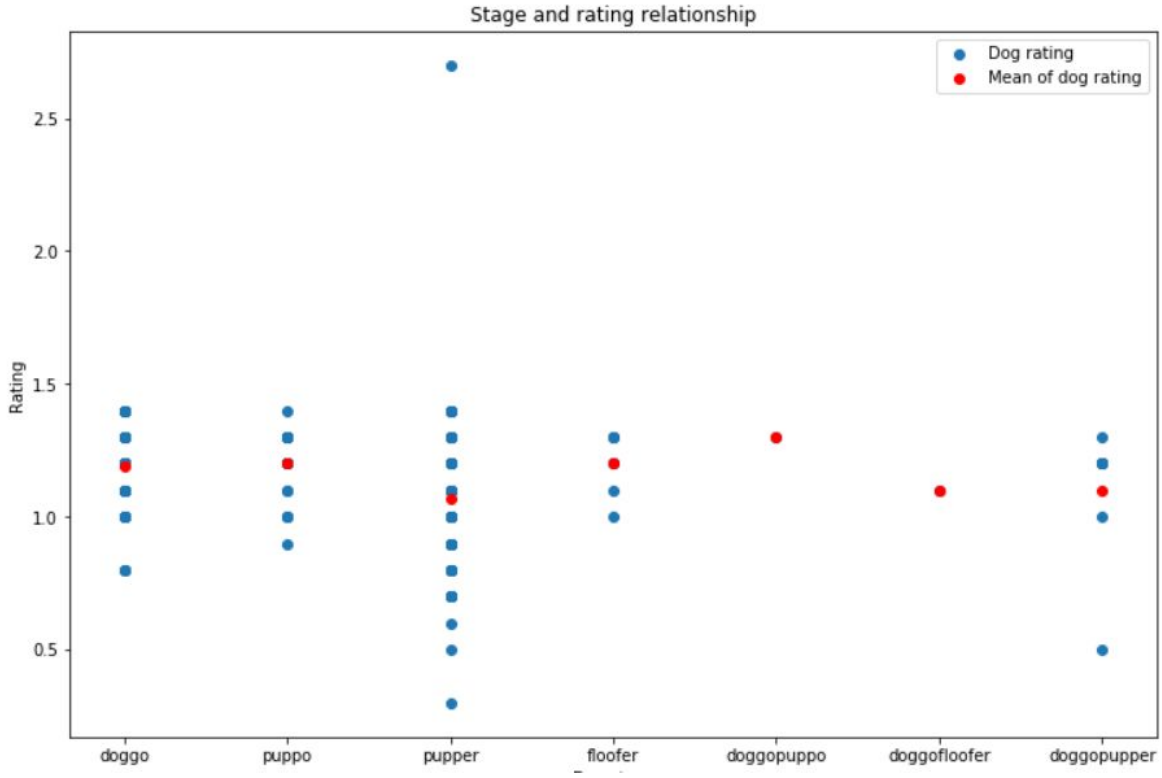
- Which dog stages receive the highest rating
- What is the relationship between breeds and rating
- What are the most popular dog names
- What dog stages have the highest favorite count

To answer these questions, the data cleaning and data restructure will aim to clean the best of relevant columns. The rest of the dataset will be omitted/ ignored as they do not contribute to give insights for the questions above (e.g: timestamp of the tweets). Some of the columns will be addressed to qualify the prompt (e.g: change the timestamp from object format to timestamp format), even though I will delete that column anyway.

The insights can be summarized below:

1. I define the question to analyze before performing the data cleaning process. Thus, I save time by not cleaning the unnecessary columns that I will delete anyway (e.g: source, expanded_URL).
2. The rating and dog breeds require some of my intuition in defining the final value of the dataset. I am open to discussion if someone with a better understanding of the rating algorithm wants to change the formula to calculate the dog breeds and rating.
3. I decided to keep some Nan rows at the final dataset because if I remove all rows with Nan, I will have such a small dataset that is not enough for further analysis.
4. Doggopuppo receives the smallest range of ratings, but ends up staying at the highest rate. Surprisingly, it also has the highest favorite count in the dataset. In the beginning, I thought the highest ratings come from the small sample size, but the favorite count says that it is not a coincidence (pictures below). That is, it has both the highest rates and the highest favorite count too. I guess I have to look at the URL of some of this dog stage.


Stage and favorite count relationship
favorite_count
Favorite Count
50000
40000
30000
20000
10000
0
doggopuppo
puppo
doggo
doggofloofer
doggopupper
floofer
pupper
Dog stage


Stage and rating relationship
Dog rating
Mean of dog rating
Rating
2.5
2.0
1.5
1.0
0.5
doggo
puppo
pupper
floofer
doggopuppo
doggofloofer
doggopupper