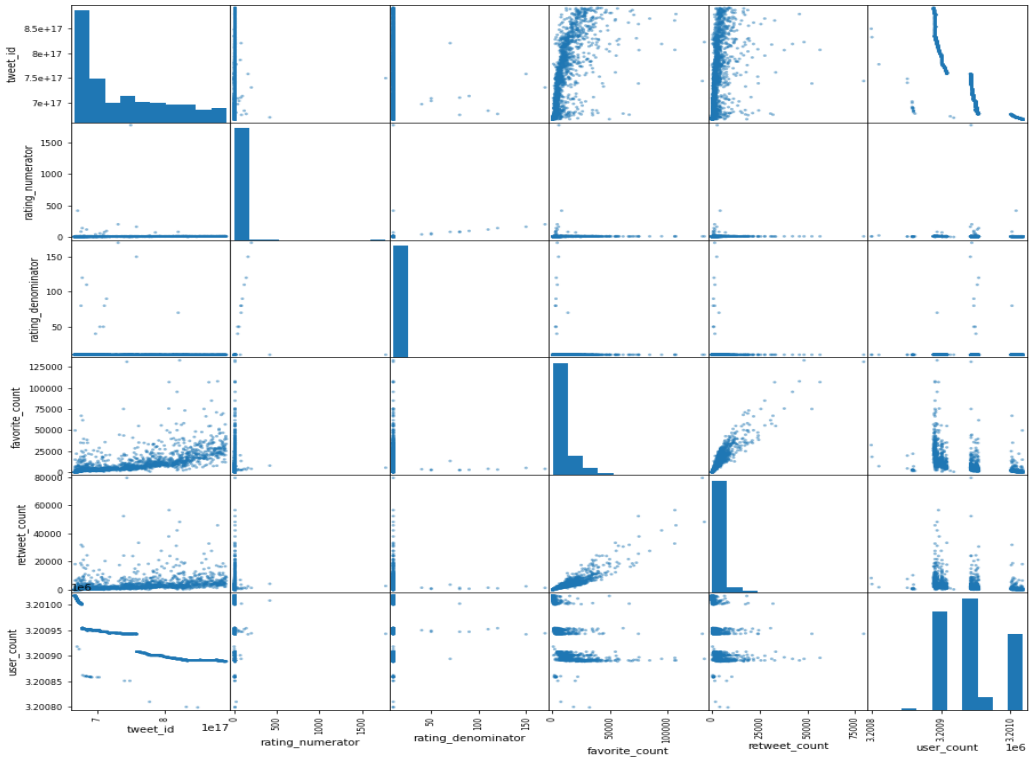
WeRateDogs Data Analysis Report

documentation of analysis and insights into final data

Analyze and visualize your wrangled data in your wrangle_act.ipynb Jupyter Notebook. At least three (3) insights and one (1) visualization must be produced.

A. load DataFrame and fix category and Datetime format
B. explore DataFrame by info() describe()
C. ask question:
    1) which dog breed get most interactions and ratings
    2) which dog stage gets most interactions and ratings
    3) how is images with no dogs affect rating.
    4) How does rating affect interaction.
    5) type of relation between follower count and retweet and favorite count.

D. Plot scatter matrix

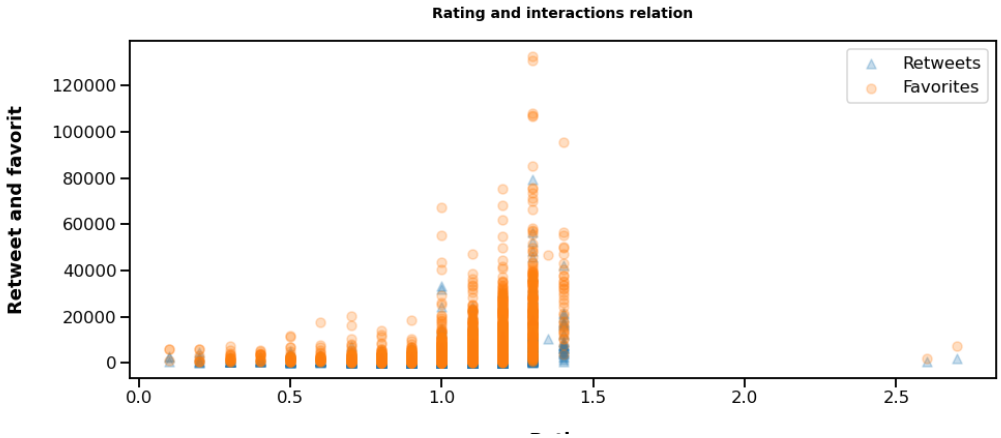


E. New columns needed for analysis:
    1) Calculate rating ratio
        - rating_numerator divided by archive_master.rating_denominator
    2) Tweet contains dog name
        - create new columns to with Boolean True and False to study relation between dog name and rating and interactions.
    3) All tweet Interactions
        - create new interaction column with sum favorite and retweet

4) Interaction in relation to user count
    - create new column relative_interaction to study the relation between user count and interaction

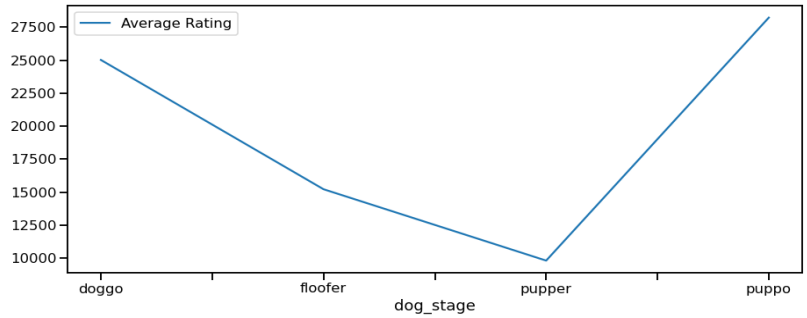
F. More data cleaning needed:
1) 1 tweet with rating = 0 , remove row
2) Identify rating ratio outliers and delete index rows.

Q: Relation between tweet rating and interactions:



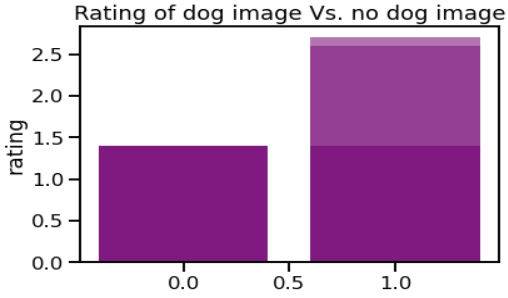
➢ tweets with high ratings gets more interactions

Q: which dog stage gets most interactions and ratings



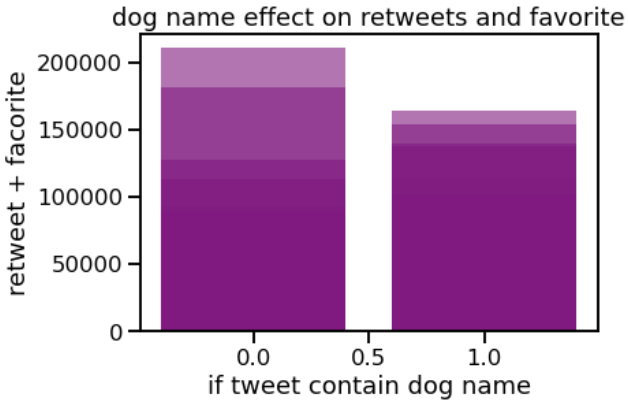
➢ dogs in puppo stage gets more interaction

Q: compare rating between Dog images with no images with no dog



➢ Images with no dog get higher ratings.

Q: does tweets with dog name get higher interactions?



- tweets with dog name have higher retweet count and favorite count

Analysis insights:

- tweets with high ratings gets more interactions
- images with no dog have higher retweet and favorite count
- favorite and retweet count are highly affected by follower count
- images with puppo dogs in it have higher ratings
- tweets with dog name have higher retweet count and favorite count

record insights
save modifies DataFrame.