Act report

Udacity DAND –

Wrangle and Analyse Data

Project Details

Gathering Data

The data, the most important step is to gather the data. First to gather the data from file than, download the image predictions. Lastly from the udacity's classroom resources

Assessing Data

After collecting the data and gathering it I used multiple function from Pandas library to explore the data to understand it better,

Quality

- **Twitter archive data:**

1. Wrong names in name table
2. The tweet-id is not str
3. Deleting unnecessary tables as example
4. The table “Timetemp” wasn’t datetype
5. In “Rating_Denomintor” table the rating is not fixed one
6. I merged it two tables together
7. Merged it 4 tables as stage instead of being individual tables

- **Image prediction data:**

1. Changing the Tweet-ID types into str

Tweet-Json

2. Retweets table convert into int
3. Favourite table converts into int

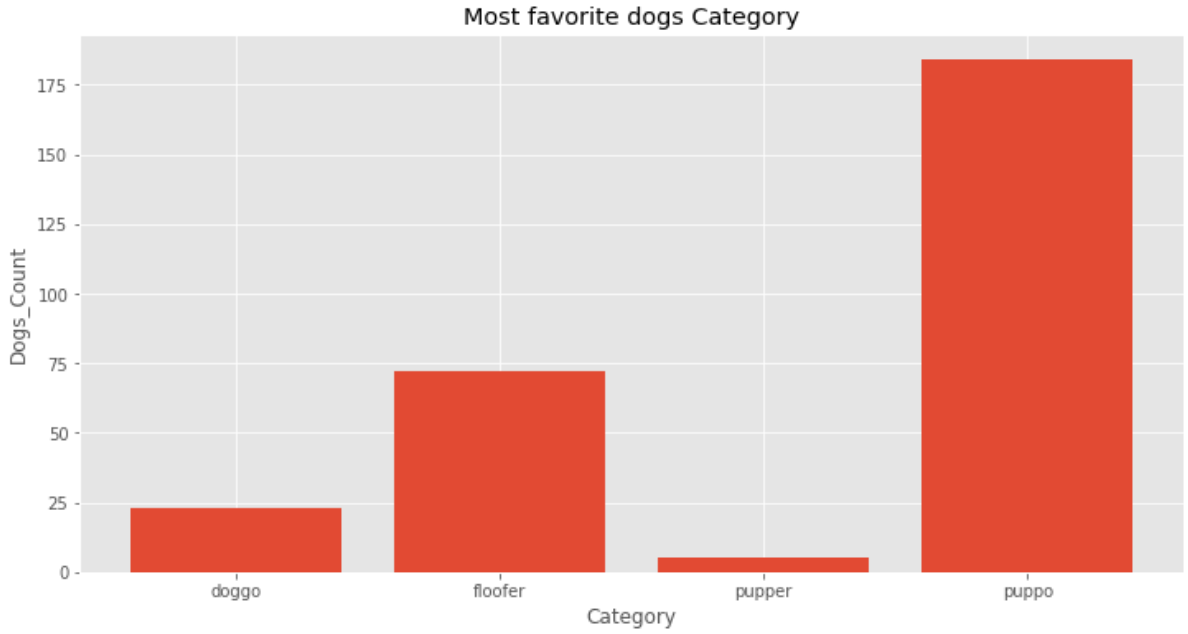
Tidyness

I used pandas libraries to figure out from stage which of the dogs are the favourite with the most high rating

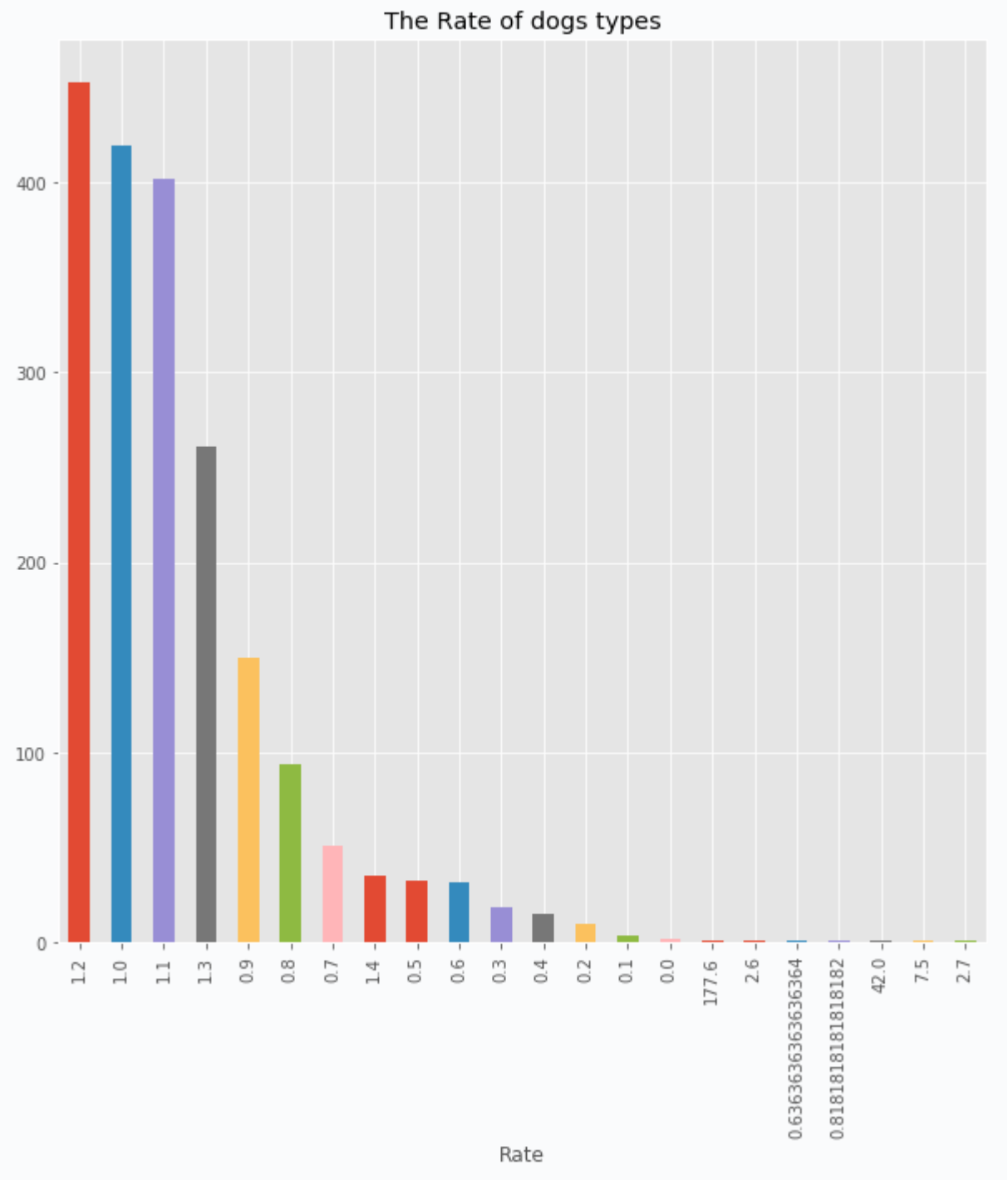
Visualization:

After wrapping up the wrangling portion which is include gathering, surveying, and cleaning the given data.

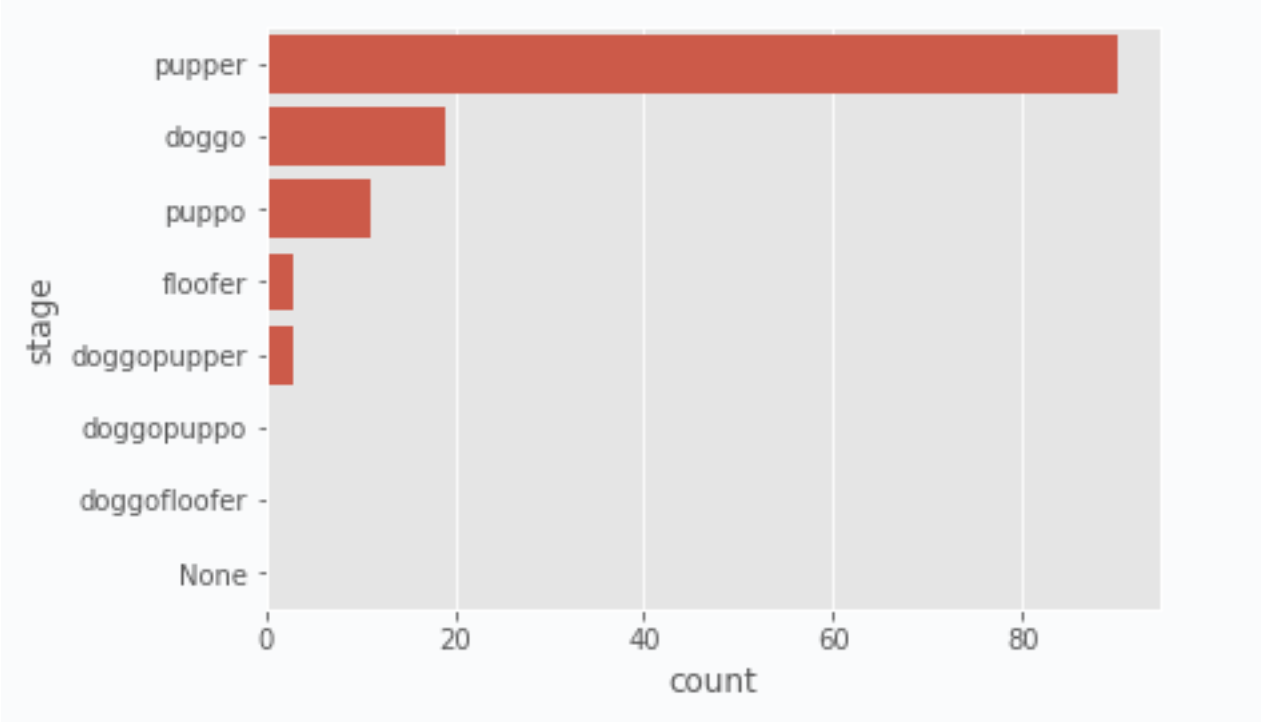
I used pandas libraries to figure out from dogs_type which of the dogs are the favourite with the most high rating among all the group. As shown in the graph below



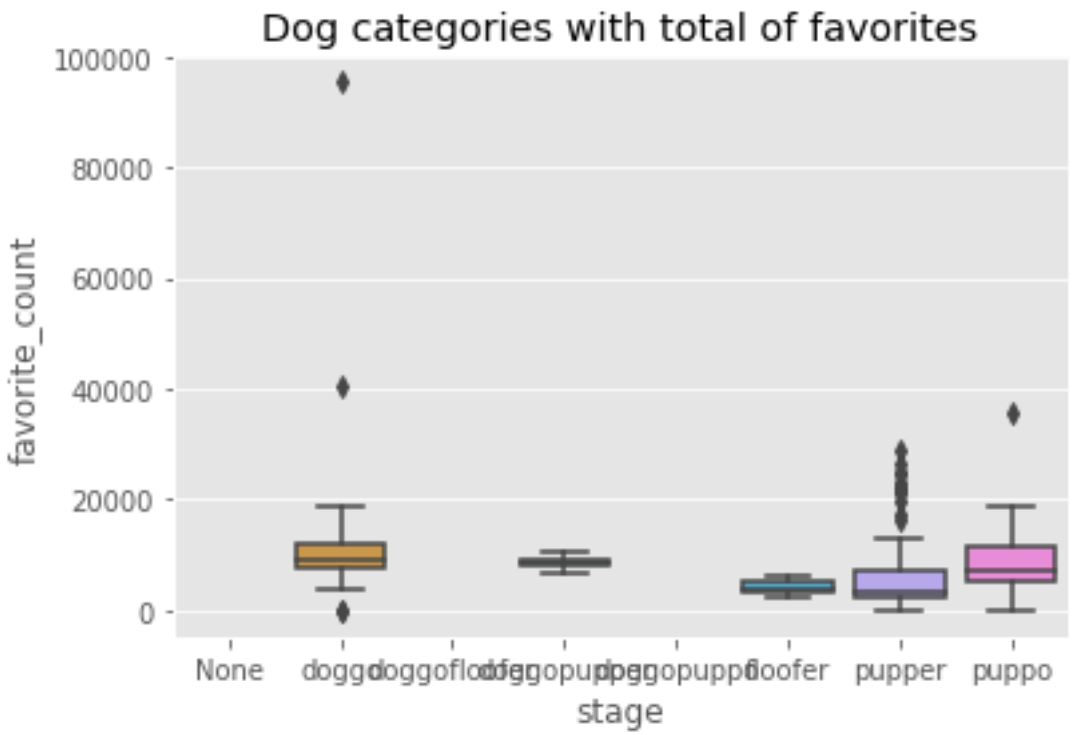
Another visual the rating_denominator / rating_numerator as shown in the graph below



The ratio between dogs type



As it appears above in the graph The most common dog is Pupper. Here I want to visualize the favourites dogs in dogs stage


Dog categories with total of favorites
favorite_count
100000
80000
60000
40000
20000
0
None
doggo
floofer
pupper
puppo
stage