Report: act_report

This report of project Data Wraling that communicates the insights and displays the visualization(s) produced from my wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

1. Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

These ratings almost always have a denominator of 10.

The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc.

Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

2. Data wrangling efforts

df_clean_merge is last table that we merge from 3 dataset df_clean_twitter, df_clean_image and df_clean_web_tweet

```
df_clean_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 39 columns):
tweet_id              2356 non-null object
timestamp             2175 non-null datetime64[ns]
source_x              2175 non-null object
text                  2175 non-null object
expanded_urls         2117 non-null object
rating_numerator      2175 non-null float64
rating_denominator    2175 non-null float64
name                  2175 non-null object
doggo                 2175 non-null object
floof                 2175 non-null object
pupper                2175 non-null object
puppo                 2175 non-null object
dog_stage             2175 non-null object
jpg_url               2075 non-null object
img_num               2075 non-null float64
p1                    2075 non-null object
p1_conf               2075 non-null float64
p1_dog                2075 non-null object
p2                    2075 non-null object
p2_conf               2075 non-null float64
p2_dog                2075 non-null object
p3                    2075 non-null object
p3_conf               2075 non-null float64
p3_dog                2075 non-null object
created_at            2354 non-null datetime64[ns]
display_text_range    2354 non-null object
entities              2354 non-null object
extended_entities     2073 non-null object
favorite_count        2354 non-null float64
favorited             2354 non-null object
full_text             2354 non-null object
id_str                2354 non-null float64
is_quote_status       2354 non-null object
lang                  2354 non-null object
retweet_count         2354 non-null float64
retweeted             2354 non-null object
source_y              2354 non-null object
truncated             2354 non-null object
user                  2354 non-null object
dtypes: datetime64[ns](2), float64(9), object(28)
memory usage: 736.2+ KB
```

3. Data analyses and visualizations

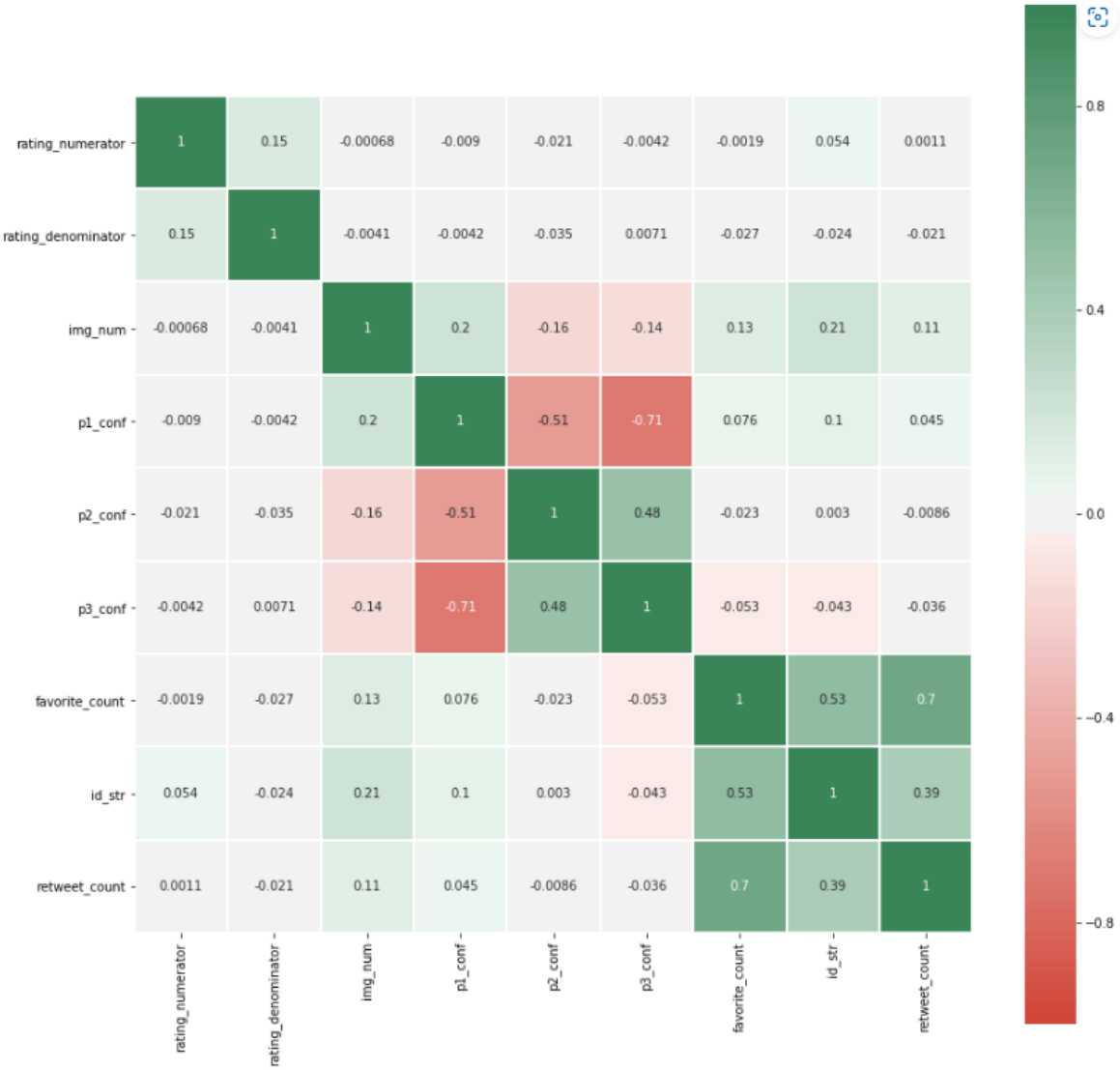
a. Describe table

	rating_numerator	rating_denominator	img_num	p1_conf	p2_conf	p3_conf	favorite_count	id_str	retweet_count
count	2175.000	2175.000	2075.000	2075.000	2075.000	2075.000	2354.000	2.354000e+03	2354.000
mean	13.215	10.493	1.204	0.595	0.135	0.060	8080.969	7.426978e+17	3164.797
std	47.726	7.019	0.562	0.271	0.101	0.051	11814.771	6.852812e+16	5284.770
min	0.000	0.000	1.000	0.044	0.000	0.000	0.000	6.660209e+17	0.000
25%	10.000	10.000	1.000	0.364	0.054	0.016	1415.000	6.783975e+17	624.500
50%	11.000	10.000	1.000	0.588	0.118	0.049	3603.500	7.194596e+17	1473.500
75%	12.000	10.000	1.000	0.844	0.196	0.092	10122.250	7.993058e+17	3652.000
max	1776.000	170.000	4.000	1.000	0.488	0.273	132810.000	8.924206e+17	79515.000

So we can see:

1. rating_numerator have 25%, 50%, 75% ~ 11 but max is 1776
2. rating_denominator have 25%, 50%, 75% = 10 but max is 170
3. Mean fo p1_conf > mean of p2_conf> mean of p3_conf so p1_conf is the most popular

b. Correlection data

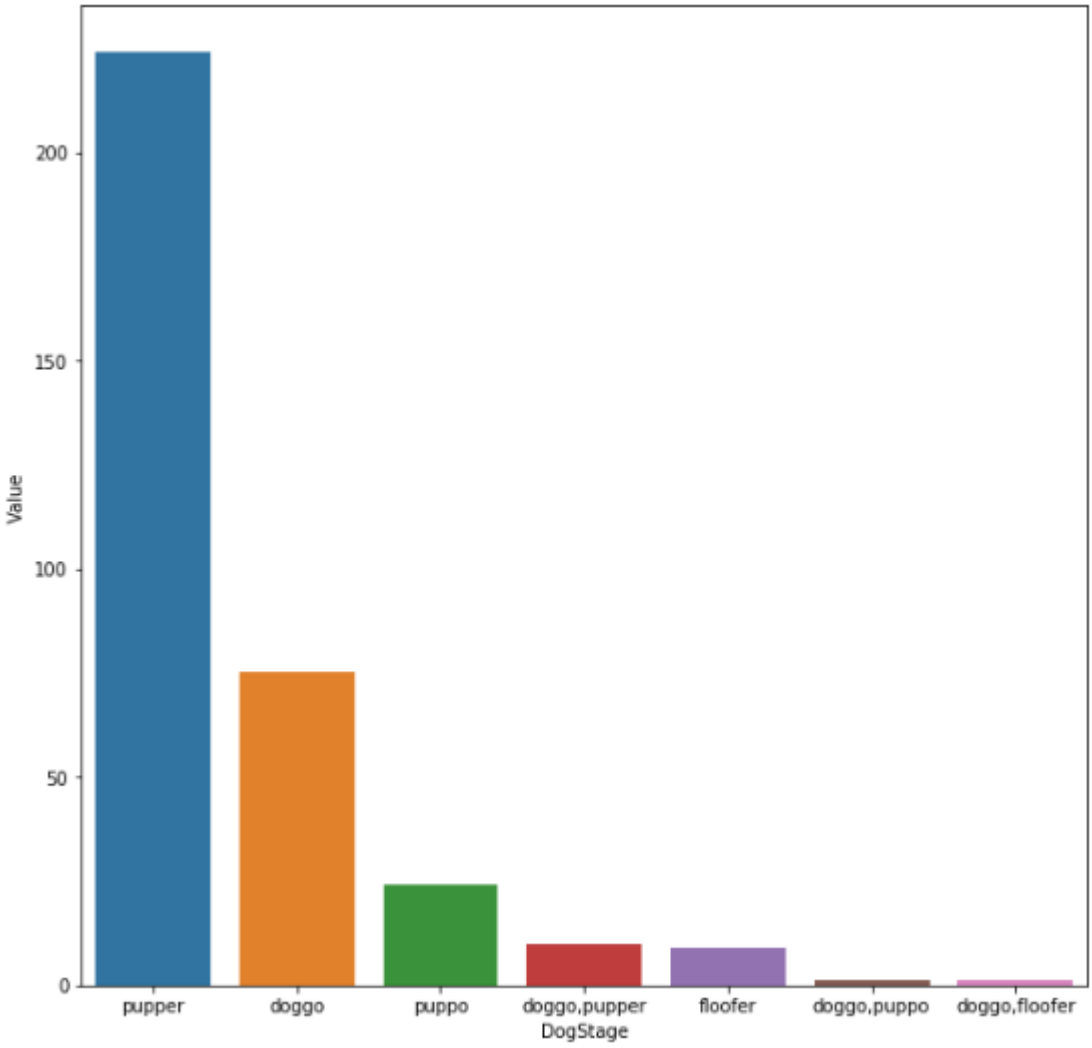


So we can see:

The same with p1_conf,p2_conf and p3_conf, p1_conf is positive

If p1_conf is correct, p2_conf and p3_conf will not correct

c. Compare the total number of 4 type of dog

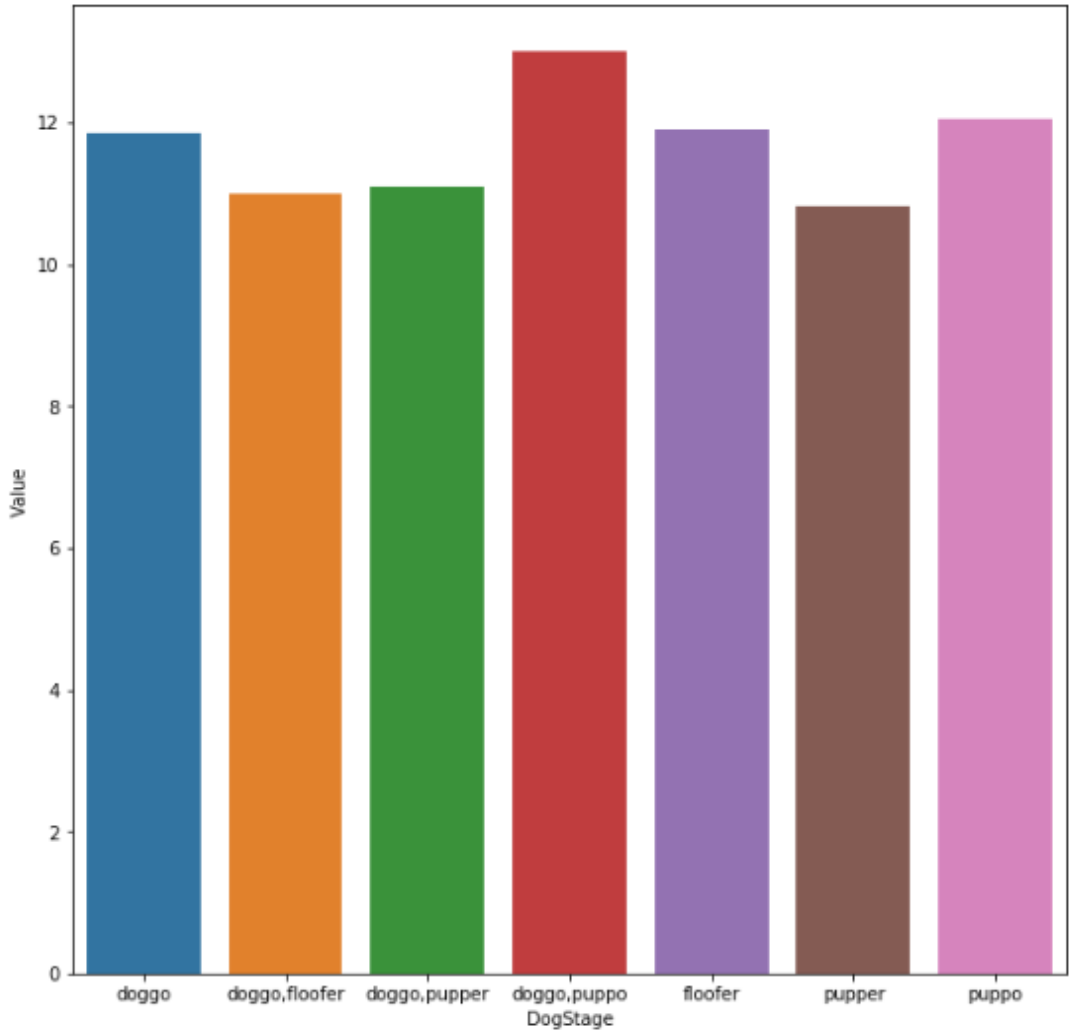


So we can see:

pupper is the most popular dog then doggo dog and floof is the lowest

pupper accounted for half of the remaining

d. Compare average rating of 4 type of dogs



So we can see:

puppo is the hightest rating

floof and doggo have the same rating

although pupper is the most popular dog but have the lowest rating