Dog Ratings- Data Analysis

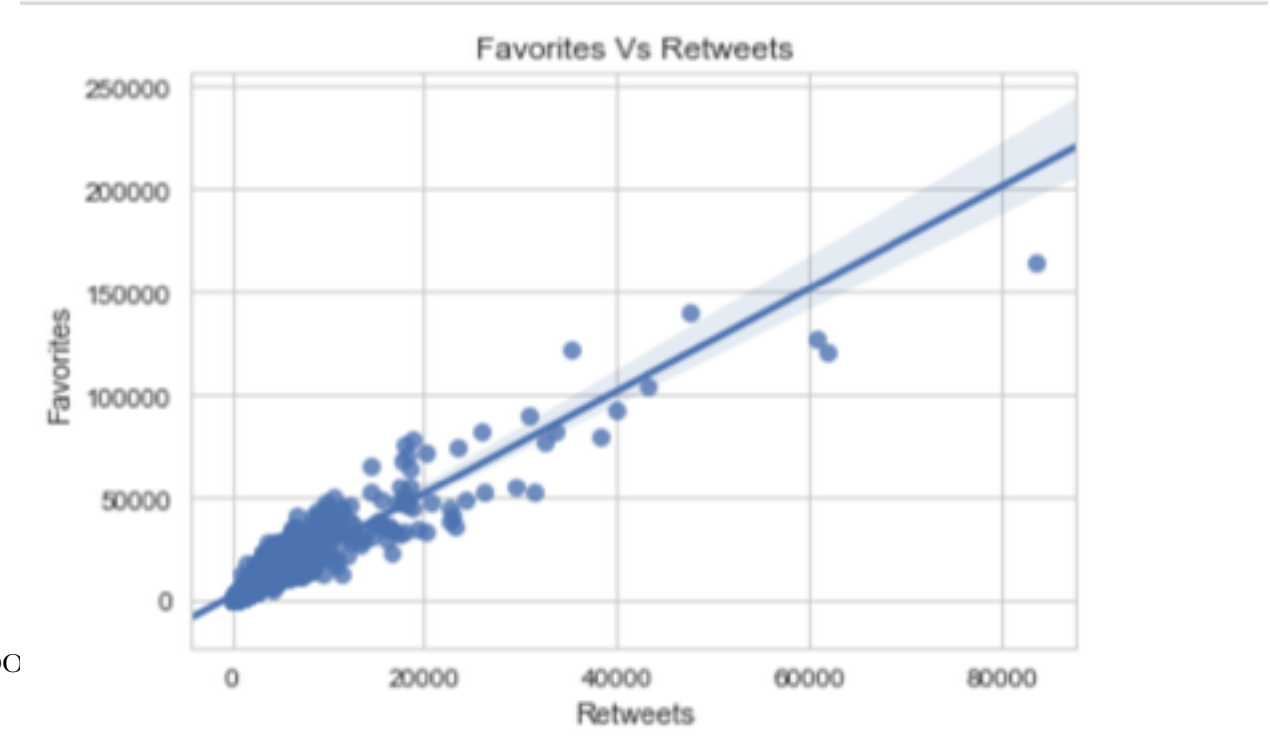
This data analysis focuses on analyzing the master dataset prepared after data wrangling various data sets for dogratings. This analysis focuses on the trends observed in the data with respect to retweets and favorites.

Data Analysis :

Generating correlation matrix on the data sets gives the following result.

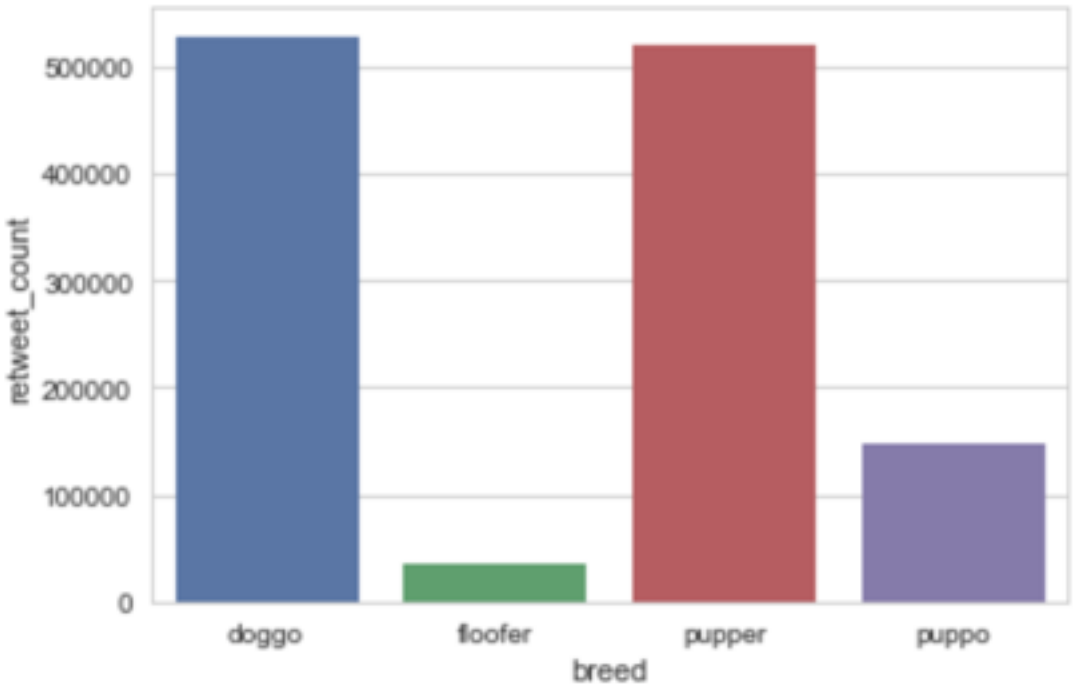
	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	confidence_level
tweet_id	1.000000	0.511789	NaN	0.382843	0.612506	0.104896
rating_numerator	0.511789	1.000000	NaN	0.289842	0.376723	0.100558
rating_denominator	NaN	NaN	NaN	NaN	NaN	NaN
retweet_count	0.382843	0.289842	NaN	1.000000	0.926834	0.052219
favorite_count	0.612506	0.376723	NaN	0.926834	1.000000	0.075380
confidence_level	0.104896	0.100558	NaN	0.052219	0.075380	1.000000

Retweets and favorites are highly correlated.The following scatter plots shows same correlation.

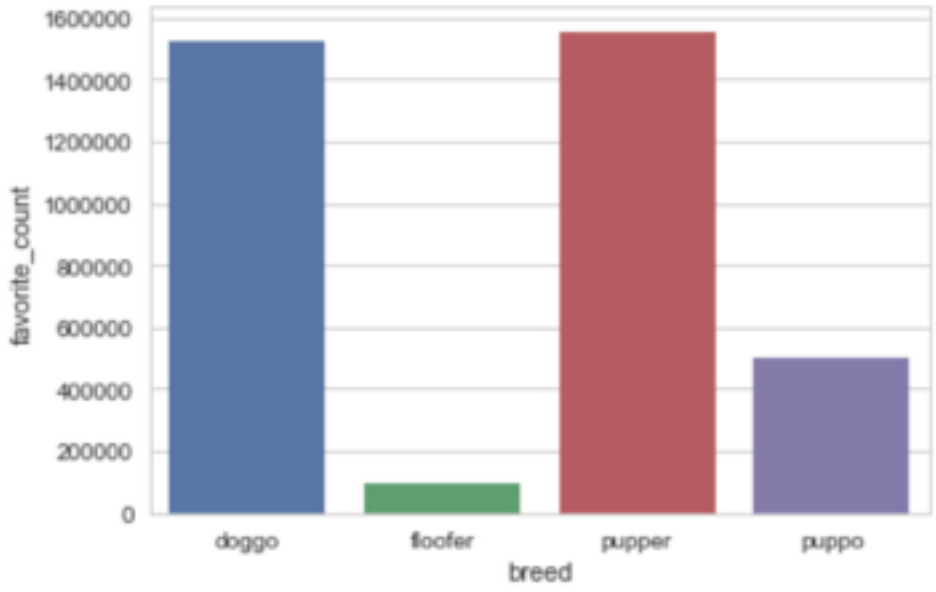


DC

Now lets look at the number of retweets grouped by Breed. " Pupper" has the most number of retweets.



Now lets look at the number of favorite grouped by Breed. " Pupper" has the most number of retweets.



Lets take a look at the max and min ratings (rating_numerator devided by ratings_denominator) grouped by breed .

	breed	ratings
0	doggo	1.4
1	floofer	1.3
2	pupper	2.7
3	puppo	1.4

	breed	ratings
0	doggo	0.5
1	floofer	1.0
2	pupper	0.3
3	puppo	0.9

Pupper has the maximum rating of 2.7 and also a minimum rating of 0.3

Here is the most rated Pupper