Dog Ratings- Data Analysis

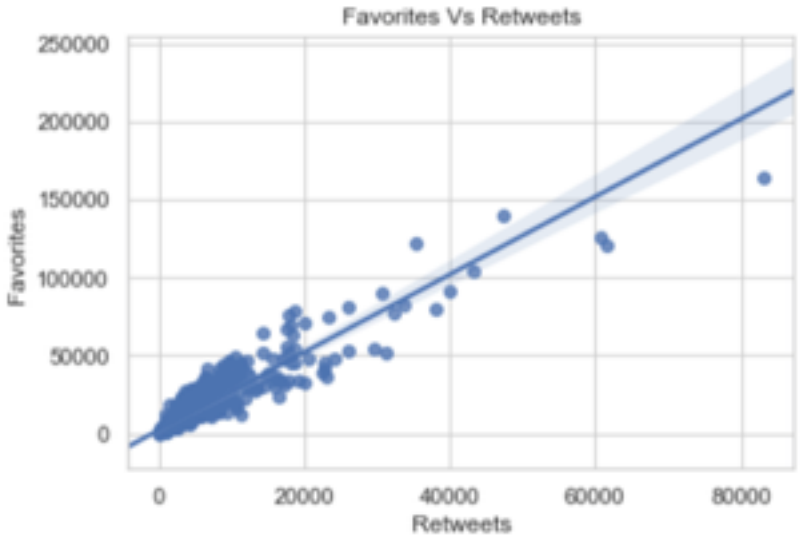
This Data analysis focuses on analyzing the master dataset prepared after data wrangling of various data sets that consists of dog ratings. This analysis focuses on the trends observed in the data with respect to retweets and favorites.

Data Analysis:

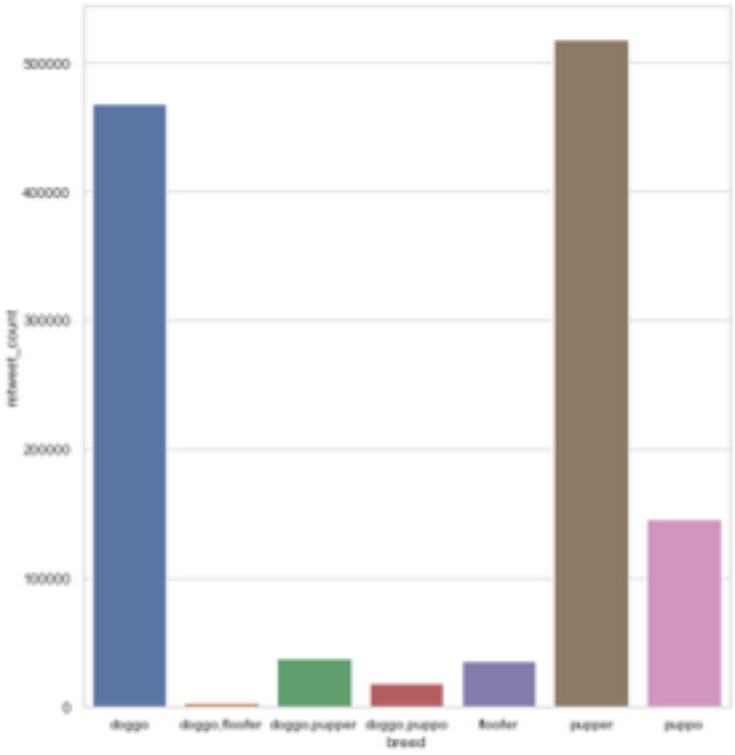
Generating correlation matrix on the data sets gives the following result.

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	confidence_level
tweet_id	1.000000	0.511789	NaN	0.382832	0.612645	0.104896
rating_numerator	0.511789	1.000000	NaN	0.289848	0.376777	0.100558
rating_denominator	NaN	NaN	NaN	NaN	NaN	NaN
retweet_count	0.382832	0.289848	NaN	1.000000	0.926728	0.052205
favorite_count	0.612645	0.376777	NaN	0.926728	1.000000	0.075395
confidence_level	0.104896	0.100558	NaN	0.052205	0.075395	1.000000

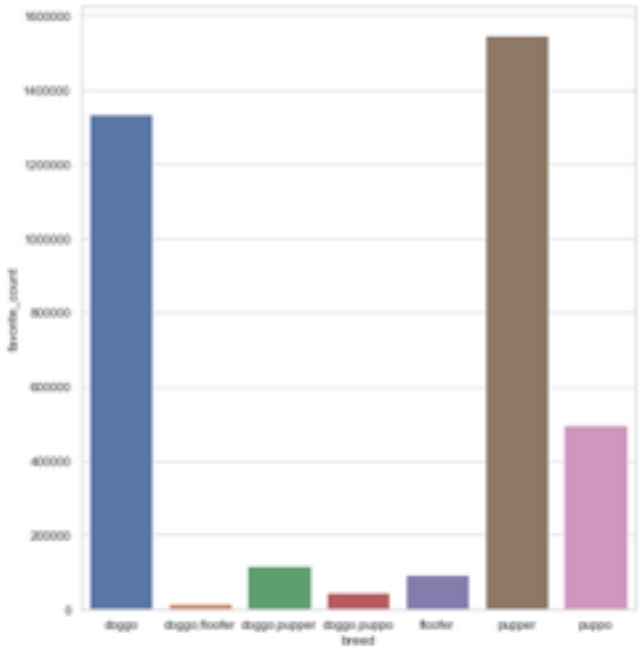
Retweets and favorites are highly correlated.The following scatter plots shows same correlation.



Number of retweets grouped by Breed. “ Pupper” has the most number of retweets.



Number of favorite grouped by Breed. “ Pupper” has the most number of favorites too.



If we just look at the number of retweets and favorites to determine popularity of a breed then "pupper" will qualify to be the most popular dog breed.

To also include dog ratings to determine popularity we need to calculate ratings using the data in rating_numerator and rating_denominator.

Here is the chart of Max and Min ratings grouped by breed after calculating the individual ratings.

	breed	ratings
0	doggo	1.4
1	doggo,floofer	1.1
2	doggo,pupper	1.3
3	doggo,puppo	1.3
4	floofer	1.3
5	pupper	1.4
6	puppo	1.4

	breed	ratings
0	doggo	0.8
1	doggo,floofer	1.1
2	doggo,pupper	0.5
3	doggo,puppo	1.3
4	floofer	1.0
5	pupper	0.3
6	puppo	0.9

Pupper also has the min rating of 0.3.
Highest rating is based on "Dog ratings" is given to breed types "doggo", "doggo, floofer", "pupper", and "puppo"Here is the most popular dog that highest rating of 1.4, highest retweets of 39854 and highest favorites count of 91771.