Analysis and Insights

After cleaning and splitting the data into two separate datasets (twitter_df.csv & dog_df.csv) some additional columns were added to these datasets to be able to derive some insights and visualizations.

These columns were:

- 'engage_value' and 'year' columns in the tweet_df.csv dataset
- 'dog_breed' and 'dog_breed_conf' columns in the dog_df.csv dataset

The engage_value column indicates the amount of social interaction (retweets and favorites count) a tweet have.

After that the 5-number summary, mean & std was calculated for engage value for each year and after some investigations it was found out that:

- tweets with very high/low value will not be taken into analysis as they skew the data very much instead, they would be treated separately
- WeRateDogs average engagement's value in 2016 was almost 3 times that of 2015 and 2017 was more than 7 times that of 2015
- the worst 50 tweets (in terms of engagement value) were in 2016 and 2015 only

Engage value stats per year

year	count	mean	std	min	25%	50%	75%	max
2015	606.0	5554.889533	13971.531429	107.911832	1214.828124	2470.601304	4828.696610	181914.208703
2016	1023.0	15714.317037	24890.786037	313.435981	5805.043777	9235.336492	17071.589676	411530.781709
2017	368.0	39222.372340	32675.148206	3608.825940	21058.824279	31161.639944	44305.269889	275796.842002

Also, it after investigating the relationship between dog score given by WeRateDogs and the amount of engagement value it was found out that:

- there is a positive correlation between dog score and engagement value in general between 0.1-0.3
- the year in which dog score affected engagement value the most was 2016
- the year with the highest number of tweets (more than 50% of the total) is 2016
- In 2017 WeRateDogs almost exclusively gave scores higher than 1 i.e. (10/10) while before ratings could either be higher or lower than 1
- 2015 has the lowest number of 'hit' tweets (9 tweets) while 2017 has the highest number of 'hit' tweets (88 tweets)

- hit tweets doesn't seem to be affected much by the score
- Tweets in 2017 are far better (in terms of engagement value per tweet) than other years as it has the least number of tweets yet the highest amount of hit tweets, nearly one of every four tweets is considered a hit compared to previous years
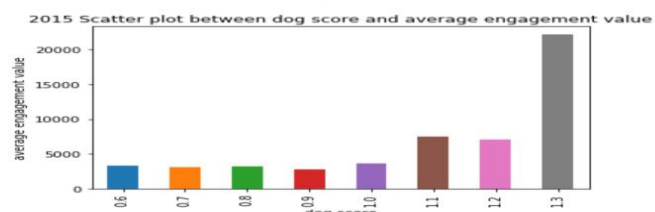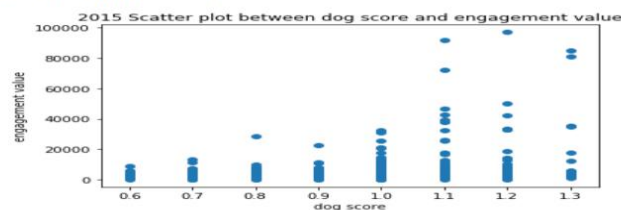
Year: 2017
corelation between dog score and engagement value: 0.2266482420008927
Tweet counts: 361
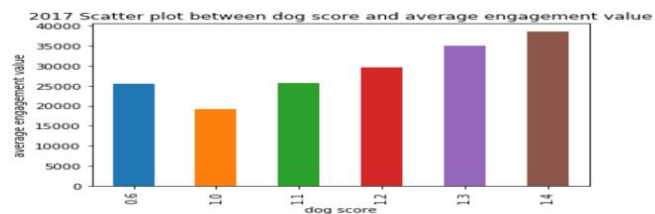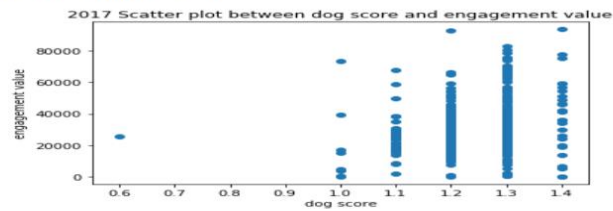
Year: 2016
corelation between dog score and engagement value: 0.3192284510495404
Tweet counts: 1016
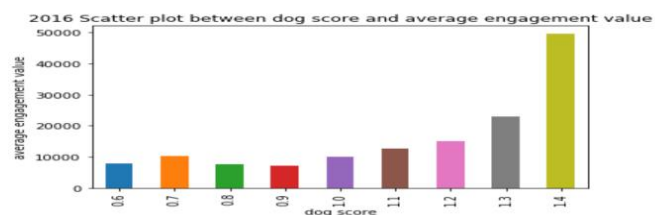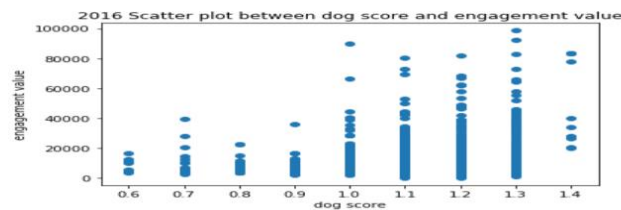
Year: 2015
corelation between dog score and engagement value: 0.21775723854257448
Tweet counts: 605

After filtering the dog_df.csv for dog breeds in which the model is at least have 50% confidence in its prediction, the number of dogs for each category and their average score was plotted.
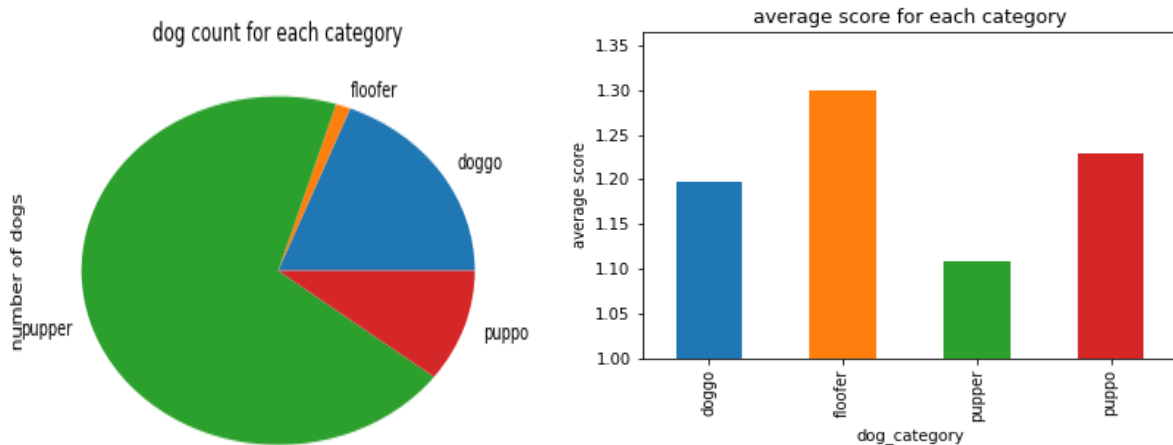
Also, the number of the top 10 popular dog breeds and their average score was plotted.
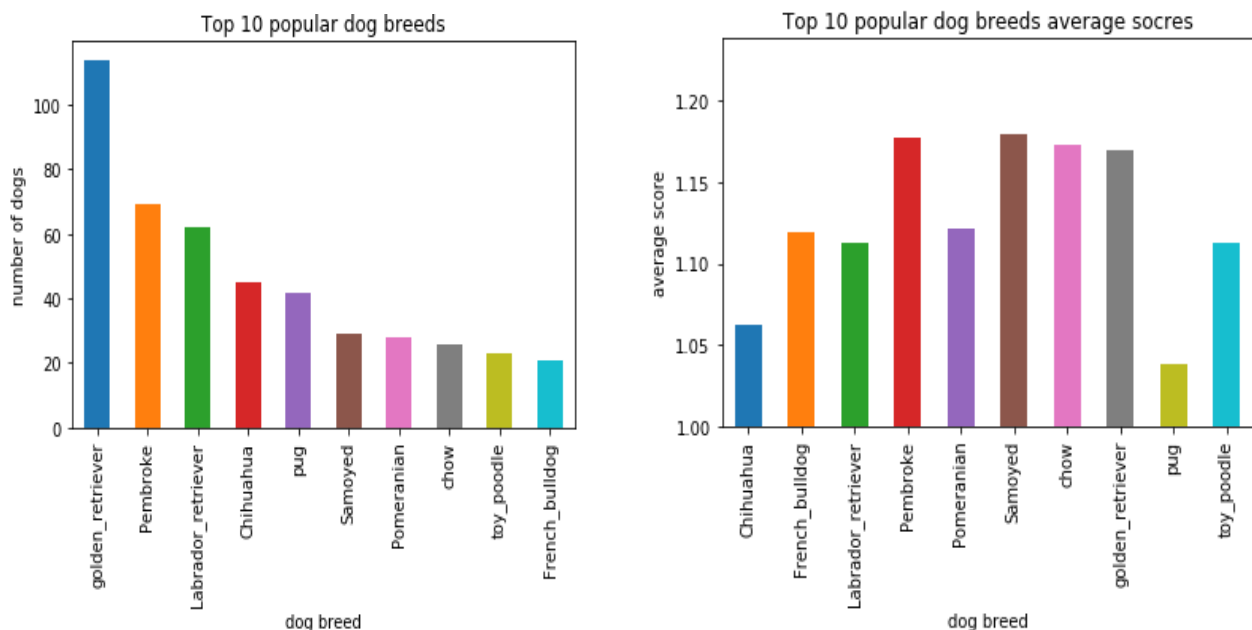
Insights were gained from these graphs and after some further investigations it was found that:

- most dogs that are classified are classified as puppers (113) and almost no dogs are classified as floofer (only 2!)
- most common dog type in WeRateDogs is the golden retreiver
- pupper although the most common type has the least average score (1.1) while doggo and puppo seems to be equal (around 1.2)
- some dog breeds tends to have a higher score than other breeds like golden retriever and chow, while others tend to be lower like pug and chihuahua