INTRODUCTION

The dataset for this project was from WeRateDogs, a twitter account that rates people's dogs with a humorous comment about the dog. The rating system is quite unique; the rating denominators are almost always 10, while the rating numerators are almost always greater than 10. This report presents some insights from the 'WeRateDogs' master dataset. The master dataset was the product of wrangling separate datasets with necessary information on WeRateDogs from 2015 to 2017.

The following table presents an overview of the final dataset used for this analysis:

SN	Variable	Description
1	tweet_id	tweet unique identification number
2	breed	dog breed
3	stage	dog stage with four levels: doggo, floofer, pupper, and puppo
4	rating_numerator	numerator of the rating score, which is almost always greater than the denominator
5	rating_denominator	denominator of the rating score, which is almost always equal to 10
6	favorite_count	total number of tweets that marked a specific as 'favorite'
7	retweet_count	total number of retweets for a specific dog
8	date	dog breed
9	time	tweet unique identification number
10	day	day of the week, which was extracted from the 'time' variable

RESEARCH QUESTIONS

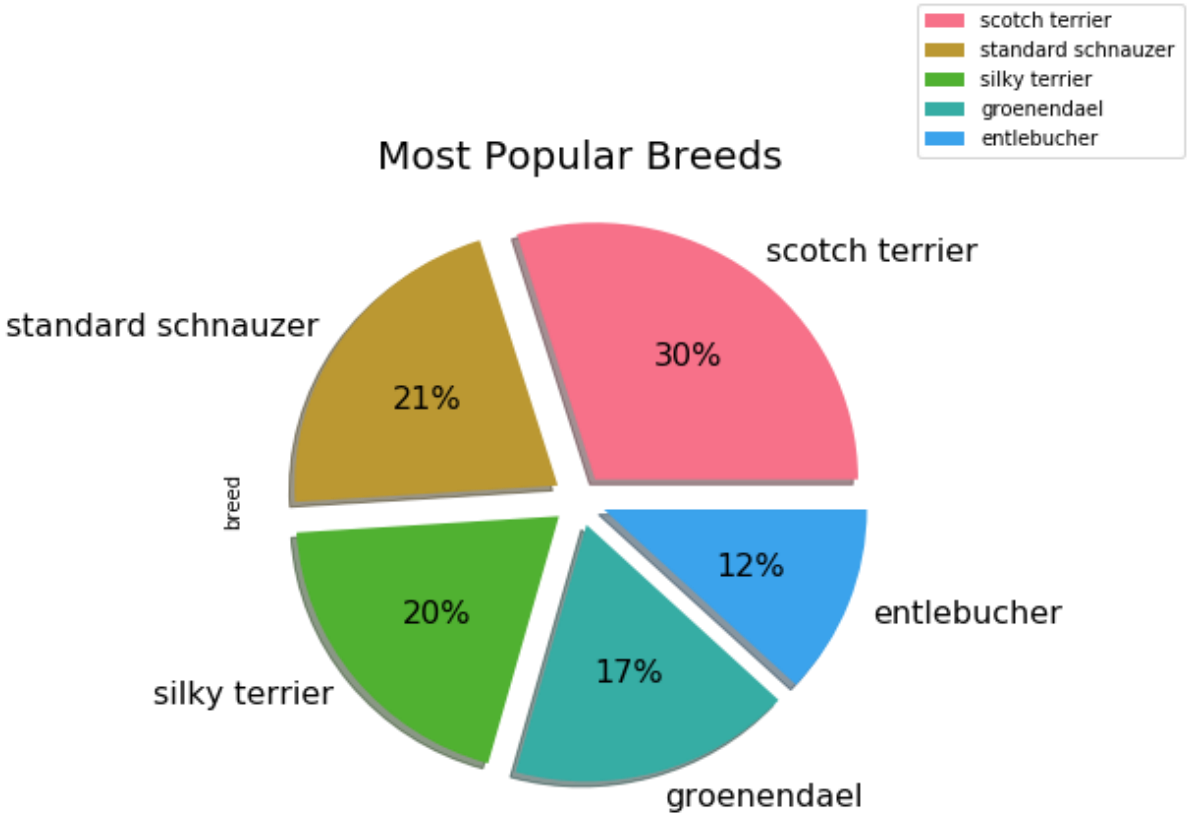
The analysis tried to answer the following questions:

1. What are the most popular dog breeds on WeRateDogs?
2. What are the least popular dog breeds?
3. What are the top five breeds with the highest rating scores?
4. What dog breeds have the lowest rating scores?
5. Is there any correlation between rating score, favorite count and retweet count?
6. Does favorite count have an effect on rating score?
7. What days have the highest retweet and favorite counts?

INSIGHTS AND VISUALIZATIONS

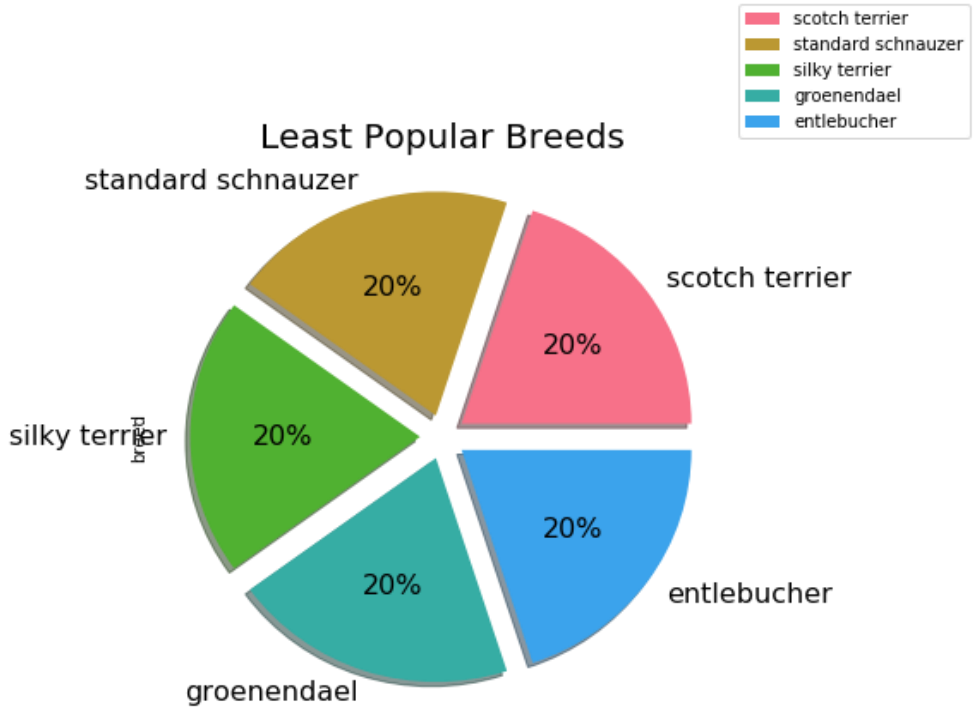
The most popular dog breeds on WeRateDogs are as follows:

a. Golden Retriever – 132 dogs
b. Labrador Retriever – 94 dogs
c. Pembroke – 87 dogs
d. Chihuahua – 77 dogs
e. Pug – 53 dogs



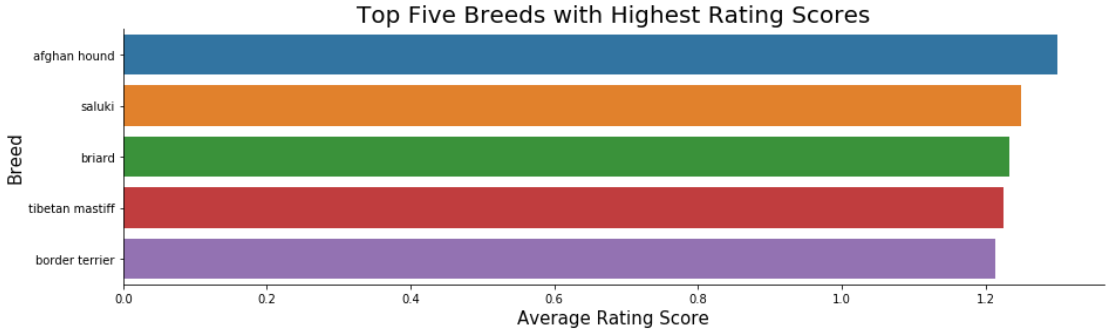
The least popular dog breeds on WeRateDogs are as follows:

a. Groenendael – 1 dog
b. Japanese Spaniel – 1 dog
c. Silky Terrier – 1 dog
d. Clumber – 1 dog
e. Standard Schnauzer – 1 dog



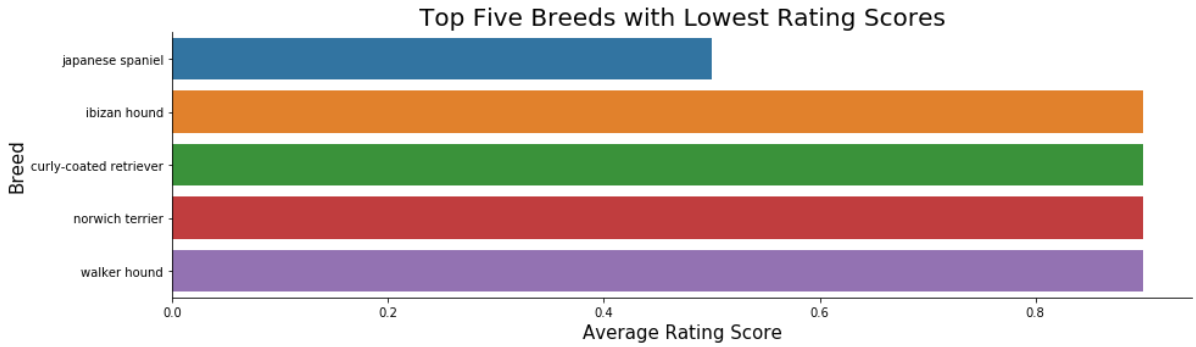
It was quite surprising that the top five breeds with the highest rating scores were not amongst the breed with the highest rating scores. The dog breeds with the highest rating scores are as follows:

a. Afghan hound – 1.30
b. Saluki – 1.25
c. Briard – 1.23
d. Tibetan Mastiff – 1.23
e. Border Terrier – 1.21

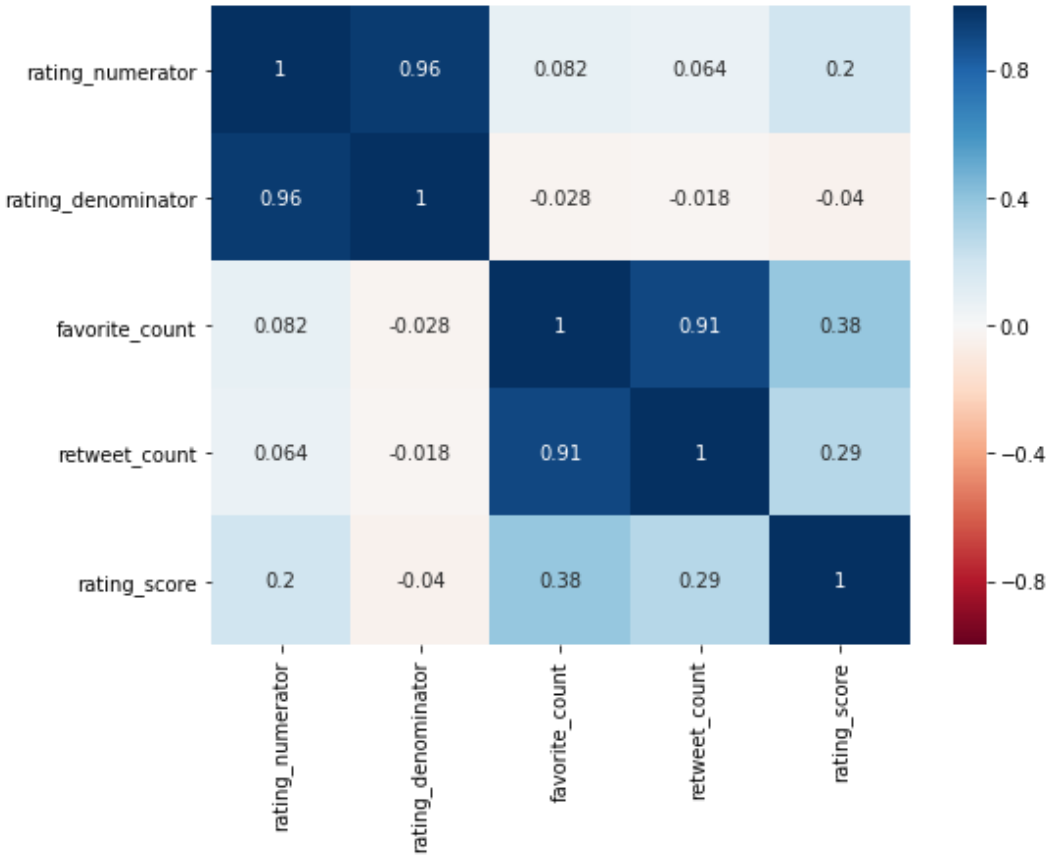


On the other hand, the top five breeds with the lowest rating scores were:

a. Japanese Spaniel – 0.5
b. Ibizan Hound – 0.9
c. Curly-Coated Retriever – 0.9
d. Norwich Retriever – 0.9
e. Walker Hound – 0.9

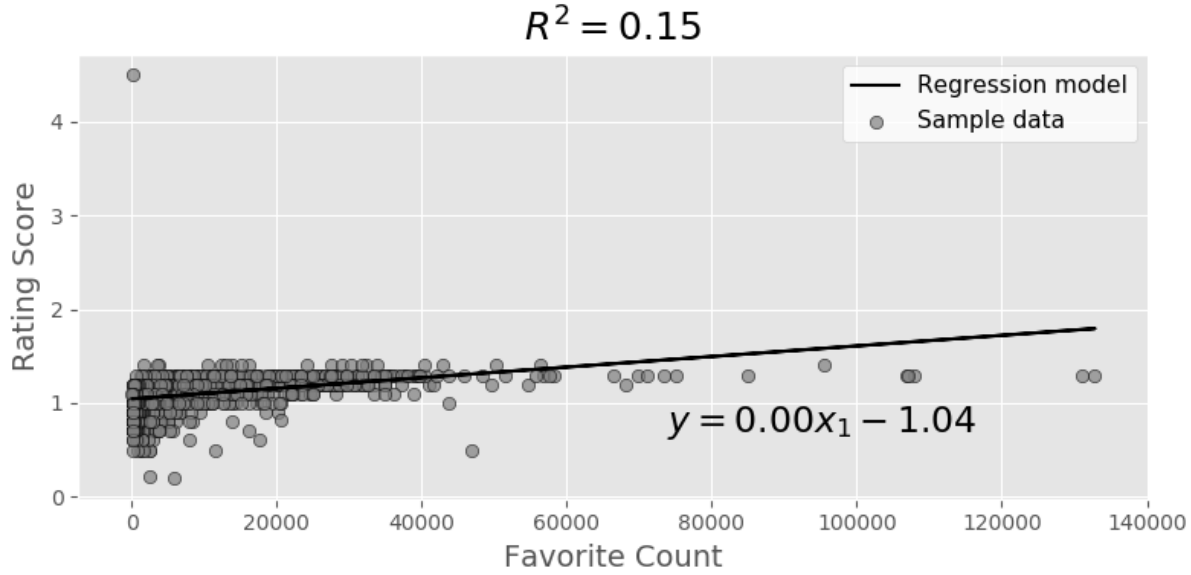


Relationship analysis like correlations and regression give important insights that could aid dog owners in their activities on WeRateDogs. Therefore, this analysis also considered the relationship between favorite count, retweet count and rating score.



The heatmap revealed a strong positive correlation between favorite count and retweet count (r=0.91), a weak positive correlation between rating score and retweet count (r=0.29), and a moderate positive correlation between rating score and favorite count (r=0.38). On the basis of the these results, this analysis tried to determine the effect of favorite count on rating scores. The table below shows the output of the regression model, and the scatterplot illustrates the regression model.

	coef	Std err	t	P>|t|
constant	1.0438	0.006	186.144	0.00
favorite_count	0.000005649	0.000000357	15.810	0.00



Favorite count has a significant effect on rating score (p-value = 0.00), but the magnitude of this effect is 0 (coef = 0.000005649). This implies that an increase in rating score is not as a result of an increase in the favorite count. Additionally, the validity of this model is only 15% (R2 = 0.15), which implies that the favorite count accounts for only 15% of the variations in rating scores. These insights buttress the fact that the WeRateDogs team is solely responsible for the rating scores of the dogs.

Finally, the analysis tried to give users an insight on the best days to post a dog with respect to favorite and retweet counts. The table below gives an overview of the mean favorite and retweet counts per day.

Day	Favorite Count	Retweet Count
Friday	9052.274038	2672.850962
Monday	8787.597403	2597.510823
Saturday	**9828.541436**	**3239.359116**
Sunday	8752.558974	2606.179487
Thursday	8854.911628	2553.697674
Tuesday	9443.229665	2784.602871
Wednesday	**10604.645631**	**3175.252427**

Saturday and Wednesday have the highest favorite and retweet counts. This implies that Saturdays and Wednesdays might be one of the best days to post a dog with respect to number of viewers and reposts.