Insights Report

Sean Kan

In this section, we will analyze post-clean-up data from WeRateDogs (@dog_rates) and address questions such as:

- Favorite dog breeds
- Most popular dog names
- Dog stage with the highest rating
- Relationships between Twitter likes and other variables.

Favorite dog breeds

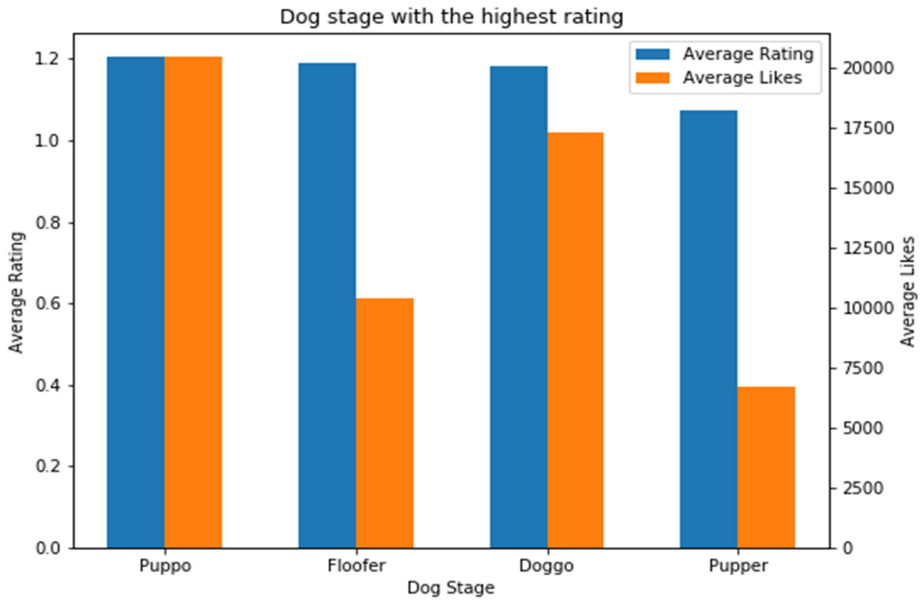
In our findings, *Saluki, Black and Tan Coonhound,* and *French bulldog* are the 3 most favorite dog breeds based on Twitter likes. Out of the three, *Saluki* is the clear favorite with an average of approximately 22,740 likes.

Most popular dog names

Cooper, Oliver, Charlie, and *Lucy* are the most popular dog names. There are 13 *Coopers*, 12 *Olivers*, and 11 *Charlies* and *Lucys*, respectively.

Dog stage with the highest rating

Puppo, the transitional stage between pupper and doggo, is ranked first based on ratings conducted by @dog_rates. Interestingly enough, it also has the most likes on average.



Relationship between Twitter likes (favorite_count) and other variables:

Correlation matrix

	tweet_id	true_rating	retweet_count	favorite_count
tweet_id	1.000000	0.510658	0.360955	0.583465
true_rating	0.510658	1.000000	0.279152	0.362346
retweet_count	0.360955	0.279152	1.000000	0.926252
favorite_count	0.583465	0.362346	0.926252	1.000000

Regression results for Twitter likes and @dog_rate ratings

Dep. Variable:	favorite_count	**R-squared:**	0.131
Model:	OLS	**Adj. R-squared:**	0.131
Method:	Least Squares	**F-statistic:**	326.3
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	4.87e-68
Time:	08:35:45	**Log-Likelihood:**	-23249.
No. Observations:	2161	**AIC:**	4.650e+04
Df Residuals:	2159	**BIC:**	4.651e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-1.31e+04	1204.698	-10.873	0.000	-1.55e+04	-1.07e+04
true_rating	2.007e+04	1111.248	18.064	0.000	1.79e+04	2.23e+04

Per the correlation matrix, the correlation coefficient for this pair is 0.36, which indicates there is a weak relationship between Twitters likes from the general public and official ratings from WeRateDogs. Moreover, the R-squared of this linear regression model is 0.131. There is no indication that ratings by @dog_rates have an impact on Twitter likes as only 13.1% of the variability in likes can be explained by the former variable.

Regression results for Twitter likes and retweets

Dep. Variable:	favorite_count	**R-squared:**	0.858
Model:	OLS	**Adj. R-squared:**	0.858
Method:	Least Squares	**F-statistic:**	1.304e+04
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	0.00
Time:	08:35:46	**Log-Likelihood:**	-21293.
No. Observations:	2161	**AIC:**	4.259e+04
Df Residuals:	2159	**BIC:**	4.260e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	1908.4261	113.373	16.833	0.000	1686.095	2130.757
retweet_count	2.5380	0.022	114.189	0.000	2.494	2.582

Per the correlation matrix, the correlation coefficient for this pair is 0.93, which indicates there is a strong relationship between Twitter likes and retweet counts. This relationship can also be verified by the linear regression model above, which has an extremely low p-value indicating that an increase of 1 retweet is associated with an increase of approximately 2.5 likes. Moreover, an R-squared of 0.858 means that retweet counts can explain almost 86% of the variance in Twitter likes, further cementing the relationship between these 2 variables.