Act Report

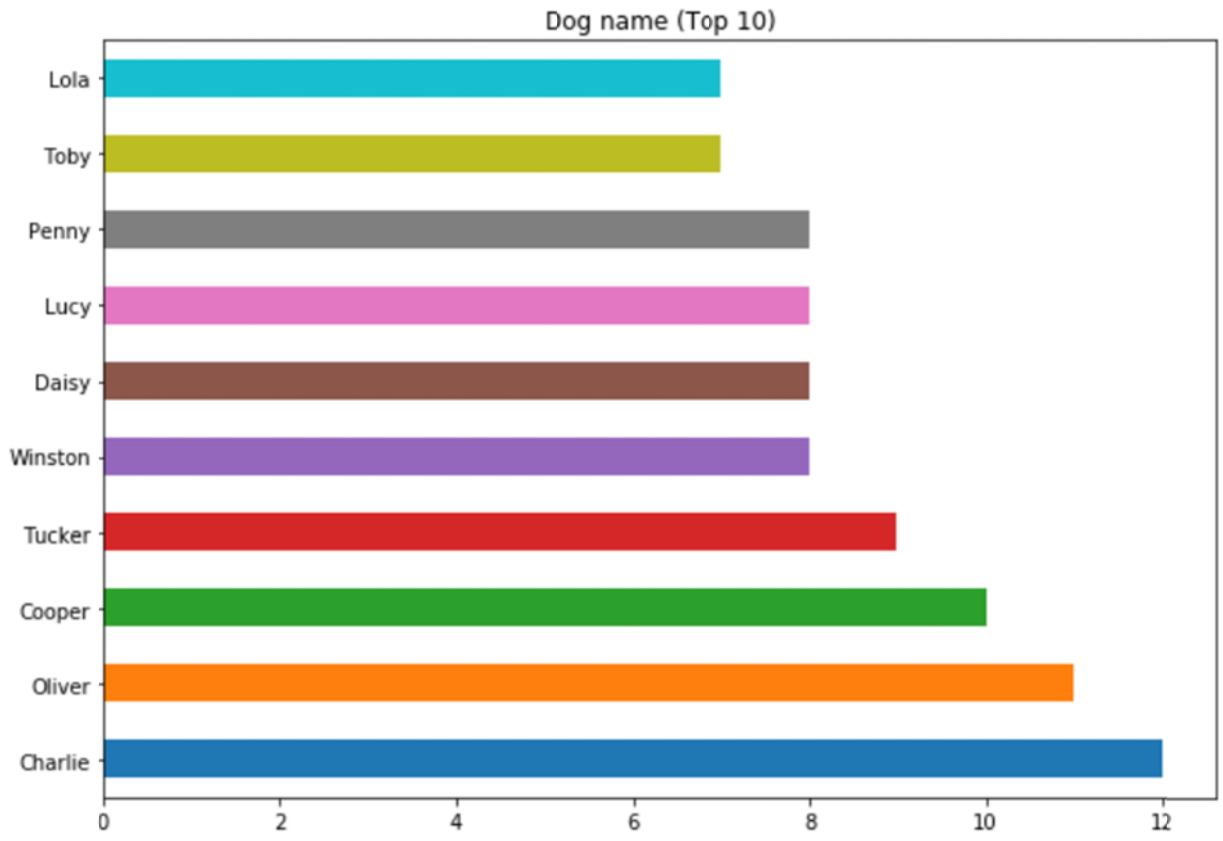
August 2, 2021

We were interested in the tweet account – WeRateDogs, where the tweeters primarily rate their beloved dogs per the unique rating method and their vivid images. We finally sorted out a clean text of 1,373 tweets and accessed those to find meaningful insights. As the basis of the original tweet text, we have gathered data that contains dog rate, stage, favorite count, and re-tweet count. From the data analysis, we have figured out five insights and made them visualized along with the question and answer as below.

Analyze and visualize #1

Question from insight

What dog names are most favored and commonly picked up from the Twitter users?



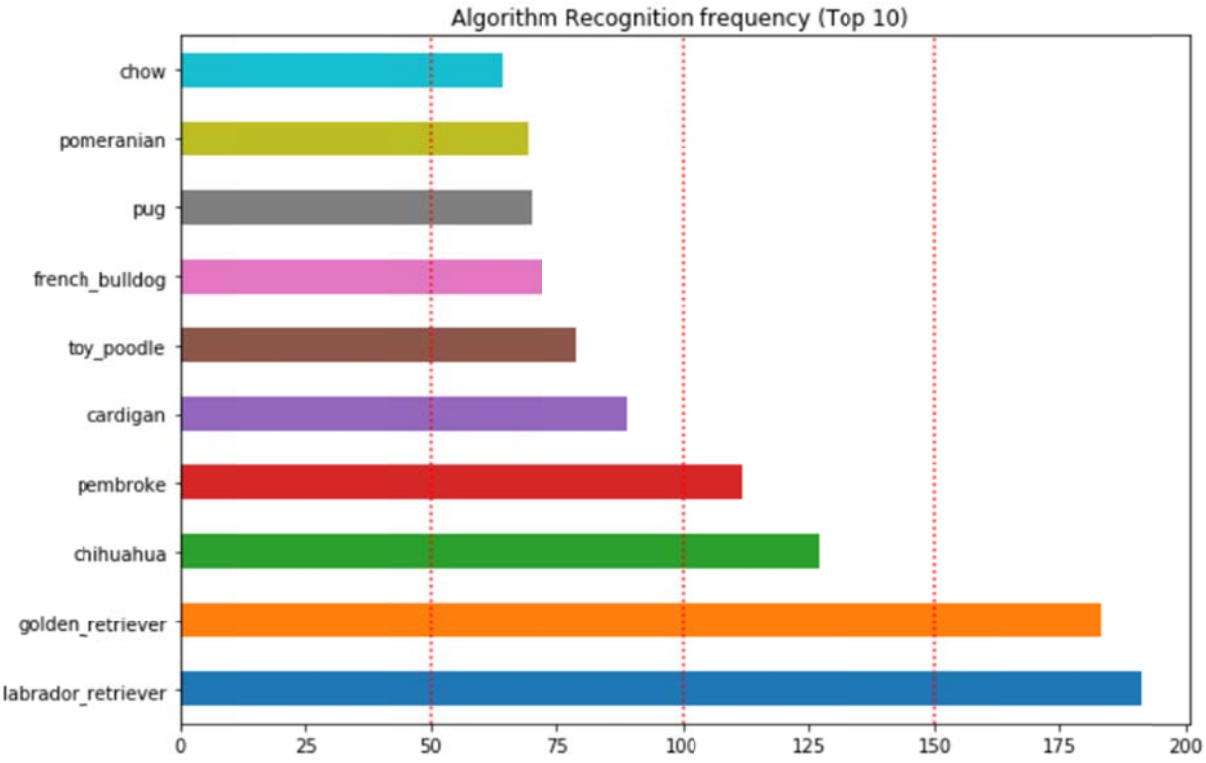
Answer from data analysis

Charlie, Oliver, and Cooper. These three names are on the top 3 from the result of the analysis.

Analyze and visualize #2

Question from insight

According to the neural network algorithm, what kind of breed of dogs are most frequently recognized by the algorithm?



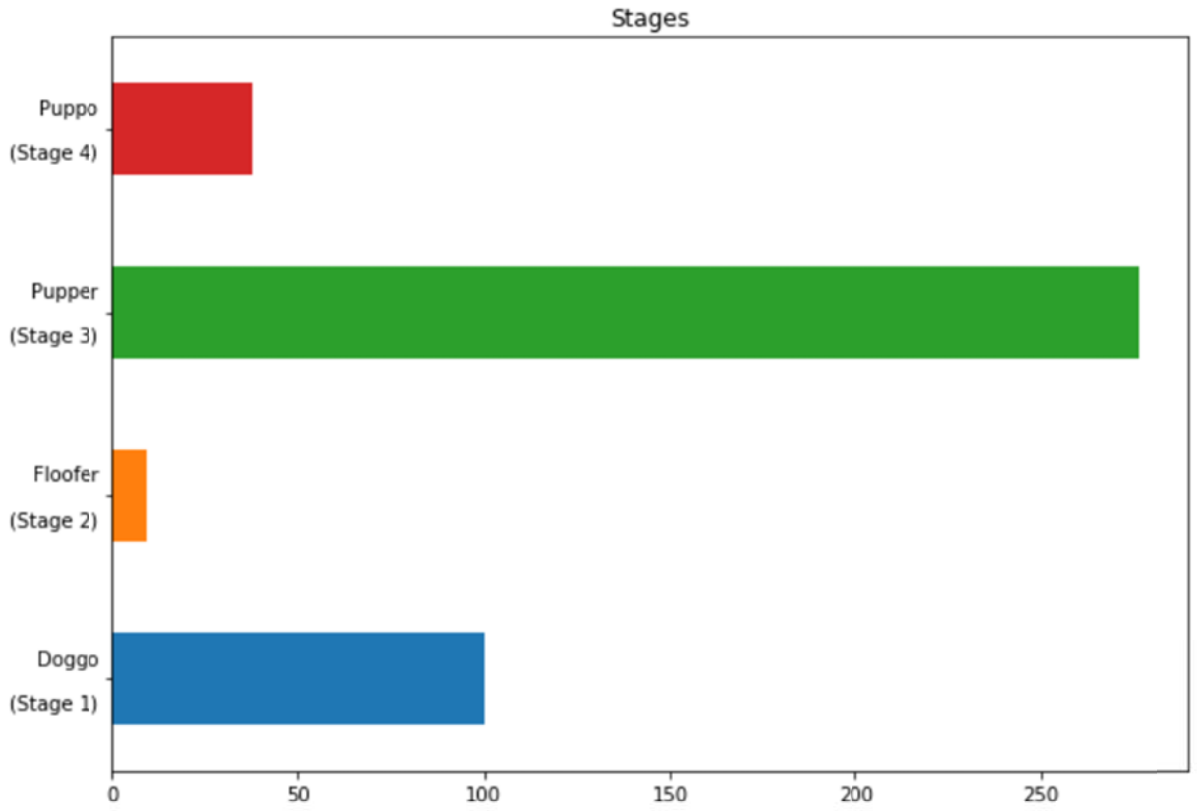
Answer from data analysis

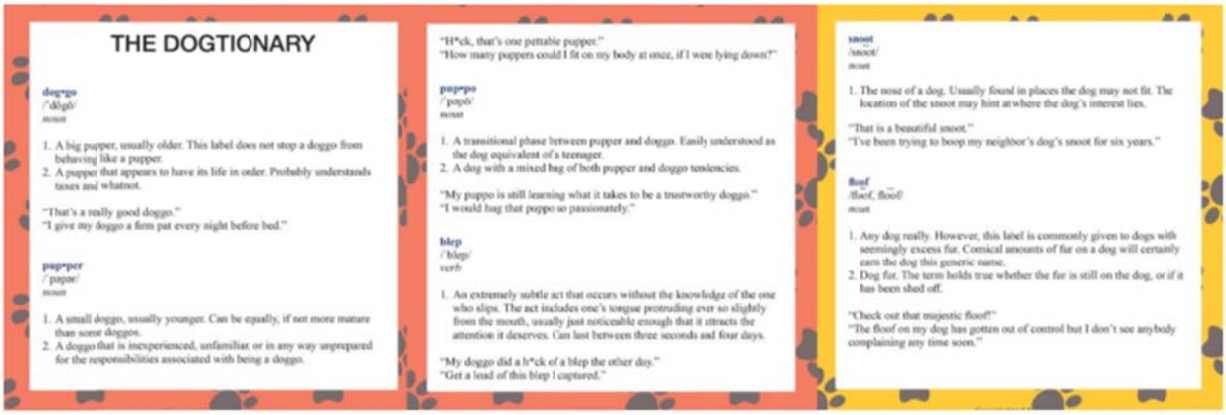
Labrador retriever, Golden retriever, and Chihuahua. These three names are on the top 3 from the result of the analysis.

Analyze and visualize #3

Question from insight

What dog stage do tweets belong to the most?





Answer from data analysis

Pupper which is stage 3, is the most common dog stage

Analyze and visualize #4

Question from insight

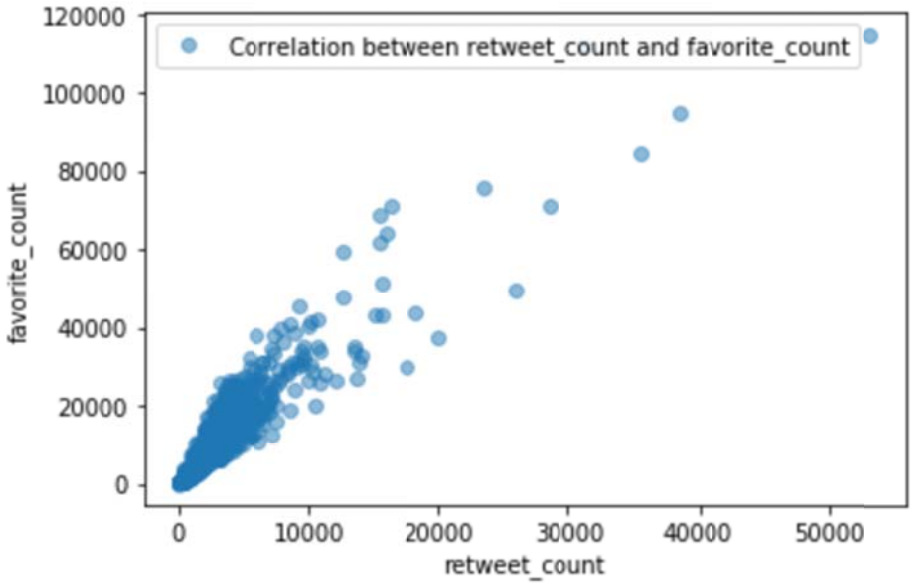
Can we find any evidence that there is a correlation between re-tweet count and favorite count?

OLS Regression Results

Dep. Variable:	favorite_count	**R-squared:**	0.853
Model:	OLS	**Adj. R-squared:**	0.853
Method:	Least Squares	**F-statistic:**	7959.
Date:	Mon, 02 Aug 2021	**Prob (F-statistic):**	0.00
Time:	16:23:58	**Log-Likelihood:**	-13372.
No. Observations:	1373	**AIC:**	2.675e+04
Df Residuals:	1371	**BIC:**	2.676e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	1648.9923	132.859	12.412	0.000	1388.364	1909.620
retweet_count	2.9018	0.033	89.211	0.000	2.838	2.966

Omnibus:	309.463	**Durbin-Watson:**	0.782
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	11740.678
Skew:	0.172	**Prob(JB):**	0.00
Kurtosis:	17.322	**Cond. No.**	4.89e+03



Answer from data analysis

We used OLS(ordinary least square) model to determine the statistical correlation between the re-tweet counts and favorite counts. With P-Value of 0.000, we are able to find significant evidence that those two variables are closely correlated. Scatter plot chart also shows as such.

Analyze and visualize #5

Question from insight

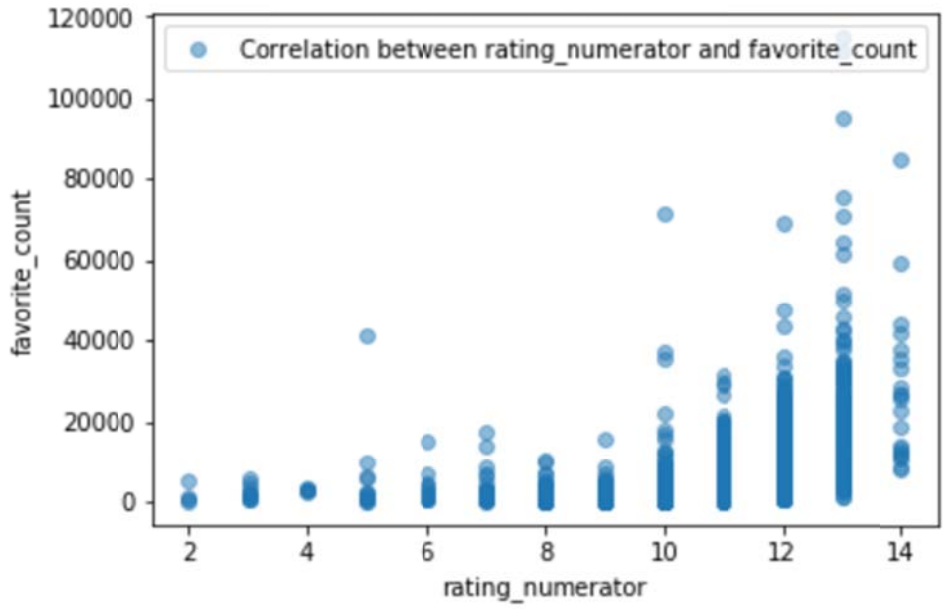
Can we find any evidence that there is a correlation between rating numerator and favorite count?

OLS Regression Results

Dep. Variable:	favorite_count	**R-squared:**	0.197
Model:	OLS	**Adj. R-squared:**	0.197
Method:	Least Squares	**F-statistic:**	337.3
Date:	Mon, 02 Aug 2021	**Prob (F-statistic):**	1.59e-67
Time:	16:23:58	**Log-Likelihood:**	-14537.
No. Observations:	1373	**AIC:**	2.908e+04
Df Residuals:	1371	**BIC:**	2.909e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-1.793e+04	1445.266	-12.408	0.000	-2.08e+04	-1.51e+04
rating_numerator	2446.1974	133.193	18.366	0.000	2184.914	2707.481

Omnibus:	1211.040	**Durbin-Watson:**	1.435
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	47218.223
Skew:	3.974	**Prob(JB):**	0.00
Kurtosis:	30.608	**Cond. No.**	61.0



Answer from data analysis

We used OLS(ordinary least square) model to determine the statistical correlation between the rating_numerator and favorite counts. R-squared value shows its fitness is looser than analysis #4. However, by referring to the P-Value and scatter plot, we are able to find evidence that those two variables are closely correlated.