WeRateDogs Project

Todd Allen

The dataset that I used to analyze and visualize is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

When performing the data wrangling activities for this analysis, I started thinking of questions that may be of interest. What day of week, month and hour of the day do most tweets occur? What is the most common dog name? Which is Oliver followed by Winston and Tucker.

After the data wrangling was complete and I viewed some descriptive statistics a few more questions came to mind. I noticed a very high favorite count and retweet count. I also noticed I high rating numerator and rating denominator. I was curious which tweet was the favorite and if that tweet was also the most retweeted tweet (that is a tongue twister). I also wanted to see which dog received the highest numerator rating and I assume it would be group of dogs for the high denominator rating.

	tweet_id	favorite_count	retweet_count	in_reply_to_status_id	in_reply_to_user_id	rating_numerator	rating_denominator	hour	img_num
count	1.300000e+03	1300.000000	1300.000000	1.500000e+01	1.500000e+01	1300.000000	1300.000000	1300.000000	1300.000000
mean	7.327177e+17	8346.620000	2558.210000	7.002882e+17	4.196984e+09	12.843077	10.545385	9.480769	1.186923
std	6.630500e+16	11505.413421	4080.831996	5.013317e+16	0.000000e+00	51.127955	7.871481	8.532622	0.540562
min	6.660293e+17	80.000000	13.000000	6.671522e+17	4.196984e+09	1.000000	2.000000	0.000000	1.000000
25%	6.749512e+17	1734.250000	592.750000	6.732617e+17	4.196984e+09	10.000000	10.000000	1.000000	1.000000
50%	7.062785e+17	3870.000000	1284.000000	6.757073e+17	4.196984e+09	11.000000	10.000000	4.000000	1.000000
75%	7.790874e+17	10358.000000	3040.000000	7.031489e+17	4.196984e+09	12.000000	10.000000	17.000000	1.000000
max	8.918152e+17	123746.000000	61805.000000	8.558181e+17	4.196984e+09	1776.000000	170.000000	23.000000	4.000000

It turned out that the tweet with the favorite count also had the highest retweet count. His name is Stephan and it seems he just like to help out with household choirs. For those interested in seeing Stephan in action, you can clik on this link https://t.co/DkBYaCAg2d.

This is Stephan. He just wants to help. 13/10
such a good boy

As far as the dog(s) with the highest denominator, it was a group of dogs that look to be well behaved. Seventeen dogs in all as one would expect with a denominator of 170.


For the dog with the highest numerator, well, this picture speaks for itself. This rating is well deserved for Atticus.

The other question I was curious about was when most of these tweets occurred. It seems December and November or the most popular months to tweet, maybe because it is too cold to go outside? The

tweets also drop as the weather gets nicer. Owner's are too busy with their dogs and the nice weather to tweet, maybe? The most popular hour for this set of tweets was surprising as it was 1 AM.

Number of Tweets by Hour

Number of Tweets by Month
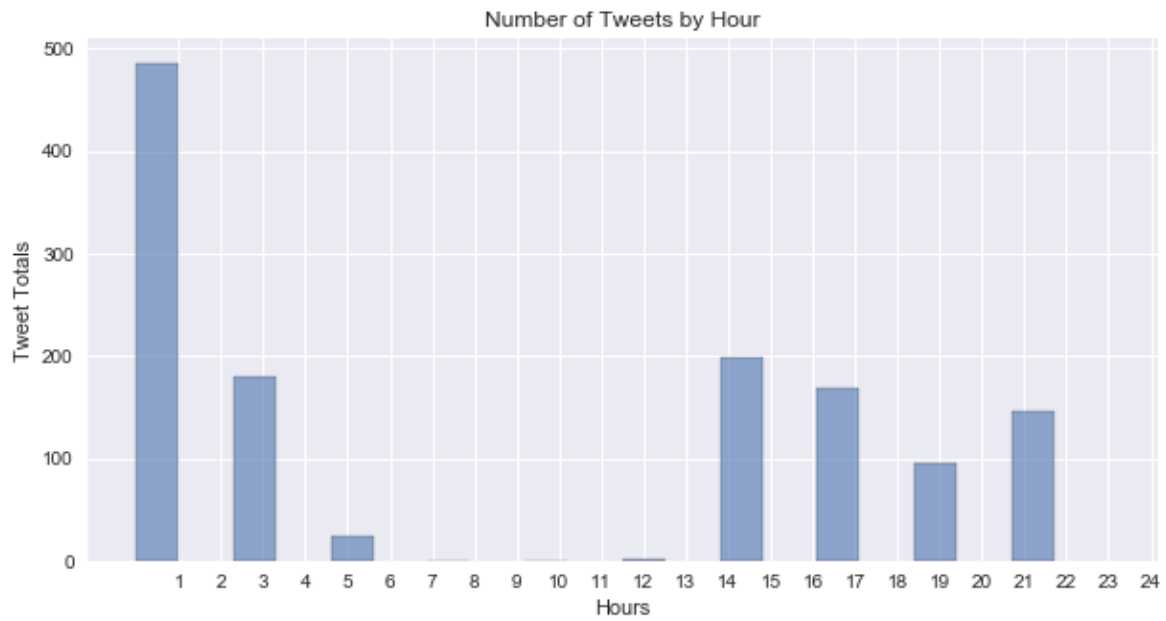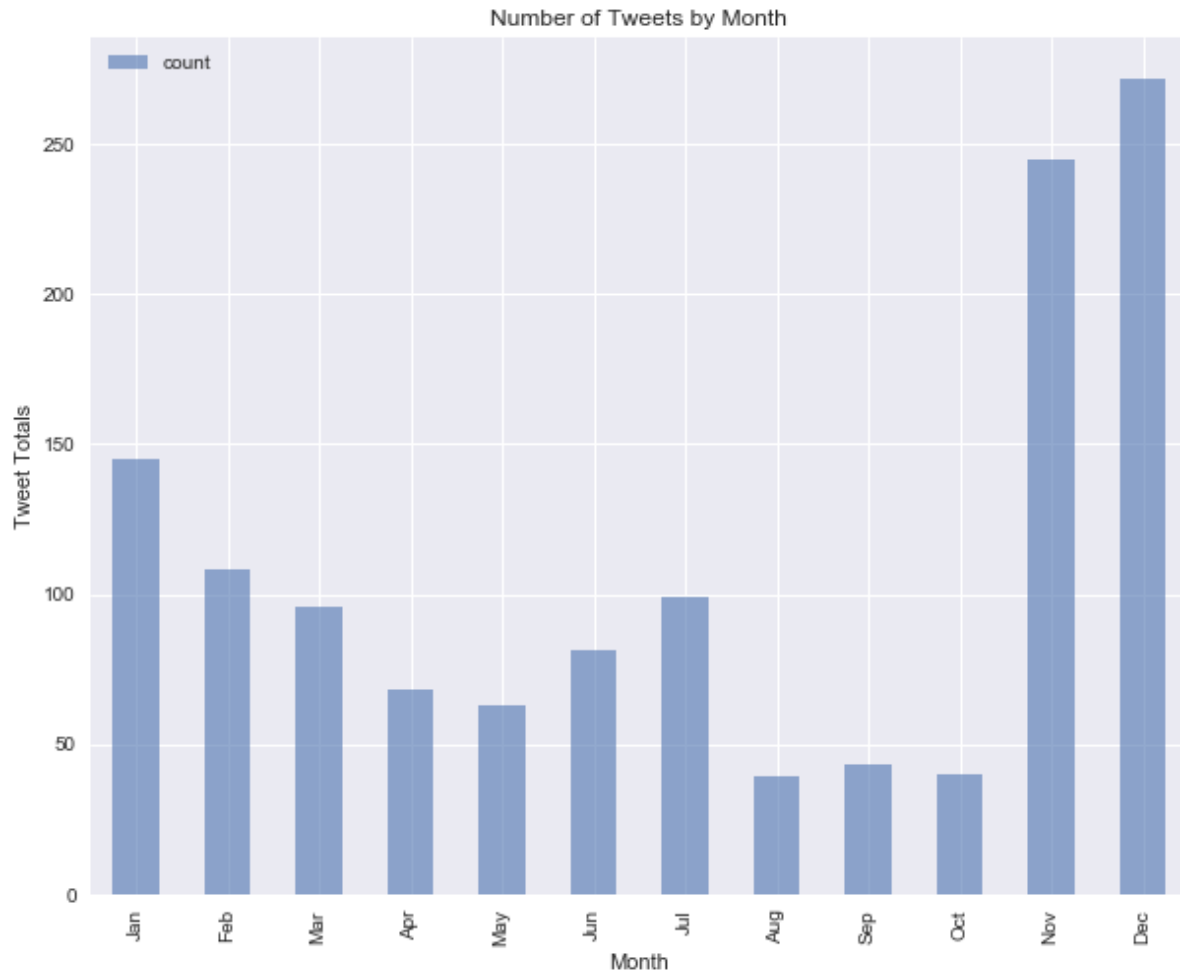
The last insight I wanted to review was the correlation between favorite count and retweet count. Just by intuition I would think that if a tweet was a favorite it would also be retweeted. Let's look to see if this is correct.

First, I looked at the correlation coefficient between the two categories which came back with .92 which suggest a strong positive correlation.

Next, I graphed a scatter plot with regression line and it to suggest a relationship between the two categories. The outlier is Stephan the floor mopping dog.
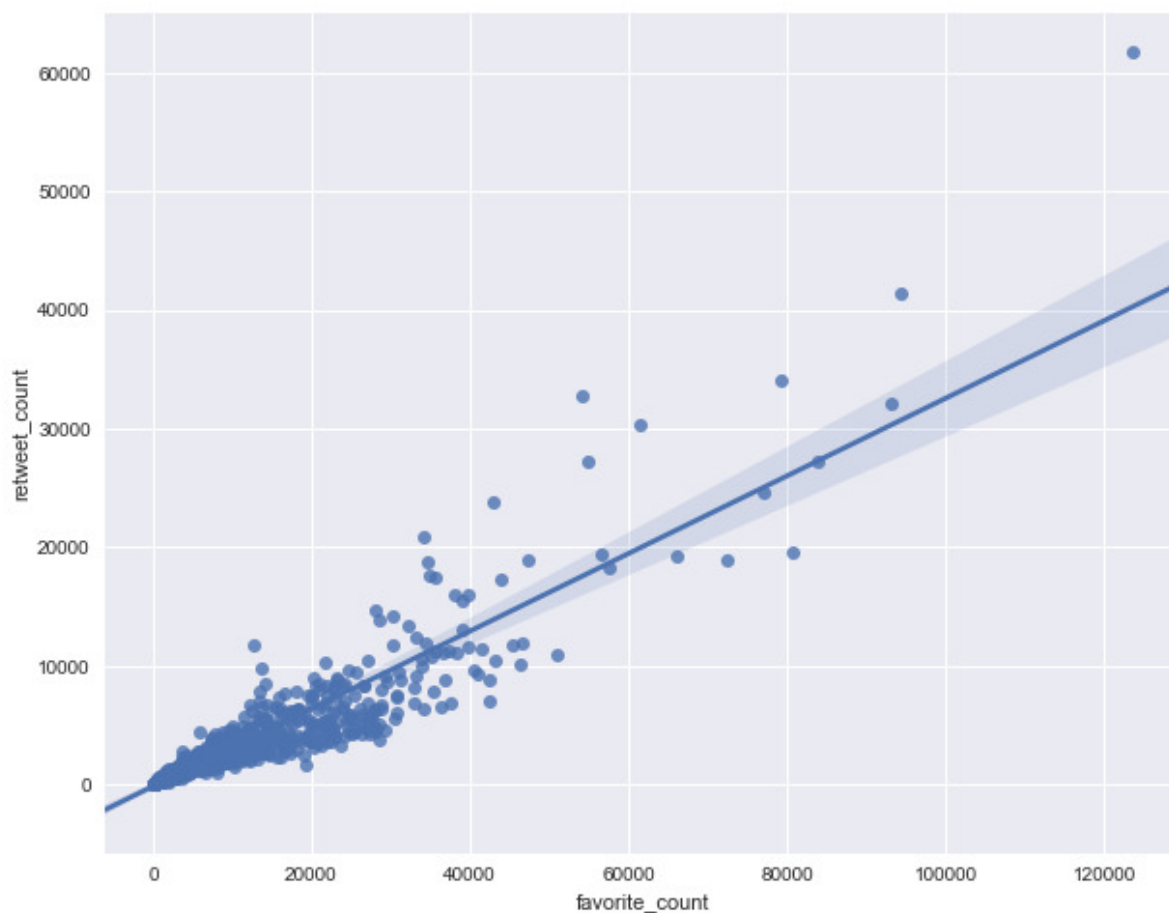
OLS Regression Results

Dep. Variable:	favorite_count	R-squared:	0.850
Model:	OLS	Adj. R-squared:	0.850
Method:	Least Squares	F-statistic:	7357.
Date:	Fri, 02 Mar 2018	Prob (F-statistic):	0.00
Time:	16:29:49	Log-Likelihood:	-12767.
No. Observations:	1300	AIC:	2.554e+04
Df Residuals:	1298	BIC:	2.555e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	1696.8484	145.922	11.628	0.000	1410.580	1983.117
retweet_count	2.5994	0.030	85.773	0.000	2.540	2.659

Omnibus:	313.649	Durbin-Watson:	0.760
Prob(Omnibus):	0.000	Jarque-Bera (JB):	8011.387
Skew:	0.500	Prob(JB):	0.00
Kurtosis:	15.120	Cond. No.	5.68e+03

Just to be sure, I ran a linear regression model which returned an R-squared factor of .85, which again suggest a very strong correlation between favorite count and retweet count. Which seems logical.

In summary, We found out that the most popular names are Oliver, Winston and Tucker. That most dop owners do not sleep or like to tweet during the 1 AM hour and during colder months of December and November. We also determined the highest rated dog (Atticus) and the most retweeted dog (Stephan) as he was mopping the floor. We looked at the correlation between retweet count and favorite count which was a strong positive relationship which makes sense.