Analysis and Visualization

What is included in the analysis?

1. To explore the relationship of tweet source and rating

2. To explore the relationship of retweet counts and ratings, favorite counts and ratings

3. To explore if there is any trend/change of retweet counts and favorites over time across 2015 to 2017.

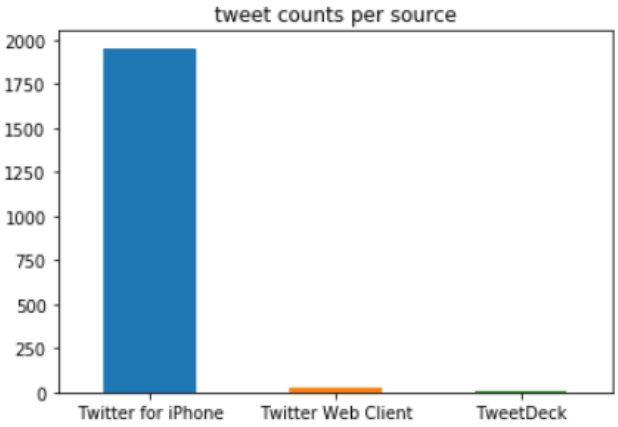
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. We have downloaded and combined data from 3 different sources, in order to get some insights from the data.

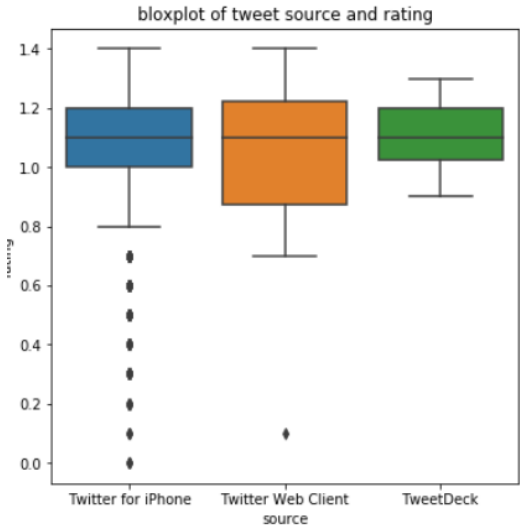
1. **Some observation and limitation of ratings from WeRateDogs**

For majority of the postings, the scores are within a well-accepted range, however, there are some outliers. When I created the rating column (rating numerator divided by denominator), below 2 tweets stood out. When I searched for the tweets, it looks like the user actually used the score to show some sentimental attachment to either the year of US independence, or a song performed by Snoop Dogg. This actually made the tweet itself more attractive and meaningful, and got a lot retweets, however, from a data analysis perspective, it's not very helpful. So I had to drop off these 2 records. Since the rating is kind of objective, we can't really draw conclusive conclusion based on the ratings.

2. **To explore if there is any relationship of source and rating**

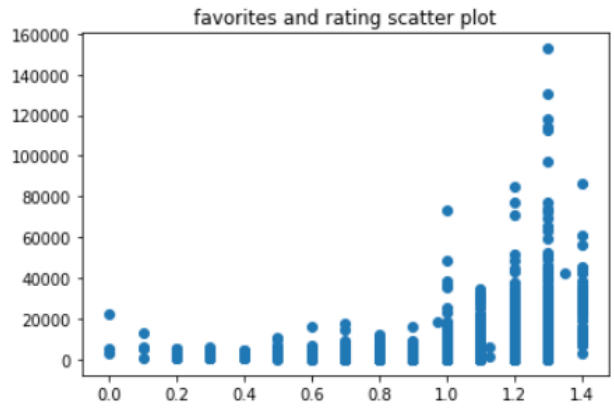
Based on the bar chart created to count the tweets by source, it looks like most tweets were tweeted from iPhone, which makes sense. Because most users would take photos of their dogs at any given time and at any place, so cell phone is very convenient to do so. But the score of the tweets posted by iPhone has a lot of lower range ratings, compared to web and Tweet Deck. But regardless of the source, the rating median for each source is pretty close, which I assume the source doesn't have much impact on the rating.

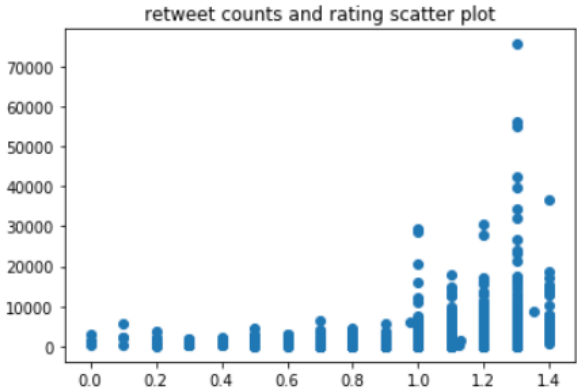


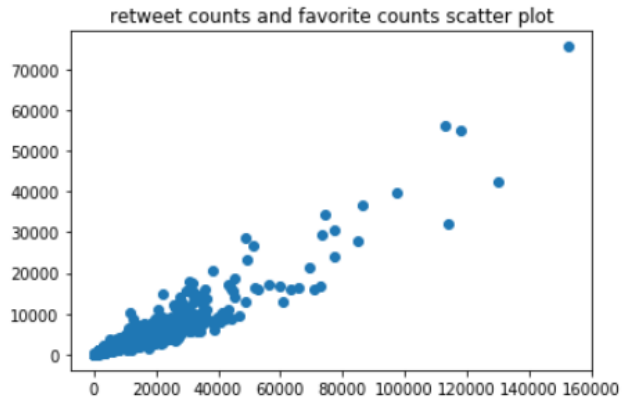


3. To explore the relationship of retweet counts and ratings, favorite counts and ratings

First 2 scatter chart were created to explore whether there is any correlation between retweet counts with ratings and favorite counts with ratings. It looks like both retweet counts and favorite counts have strong liner correlation with ratings. The higher the rating, more retweets and favorite counts it gets. However based on the statistics of a liner regression model that I ran between favorites and retweet counts, the r-square is 0.862, which means the favorite and retweets has very strong relationship. Therefore, the relationship of favorites and ratings resembles the relationship of retweet counts and ratings.







OLS Regression Results

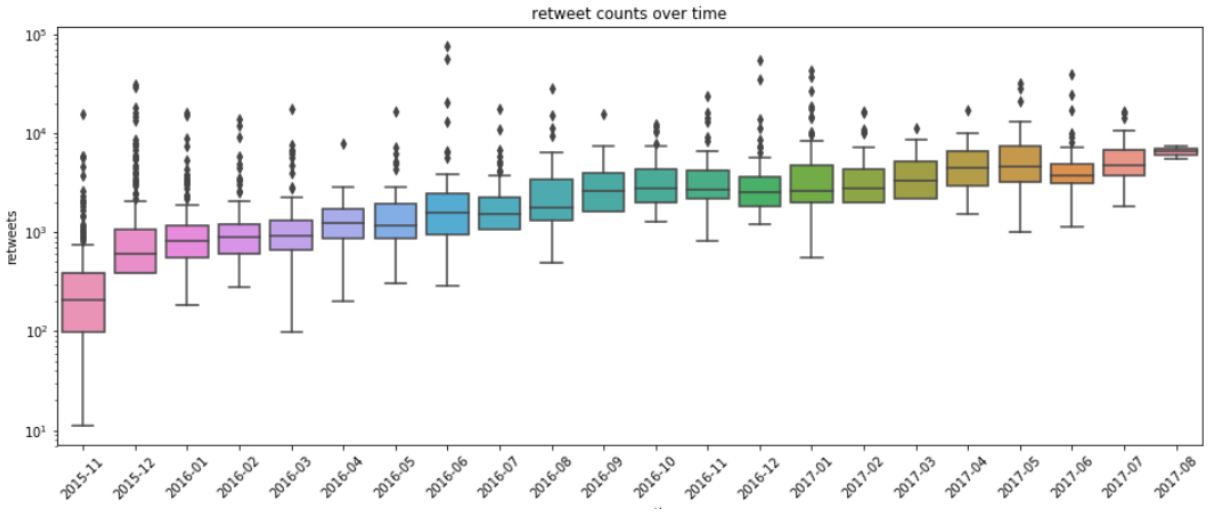
Dep. Variable:	favorites	**R-squared:**	0.862
Model:	OLS	**Adj. R-squared:**	0.862
Method:	Least Squares	**F-statistic:**	1.248e+04
Date:	Sat, 21 Nov 2020	**Prob (F-statistic):**	0.00
Time:	04:22:14	**Log-Likelihood:**	-19547.
No. Observations:	1992	**AIC:**	3.910e+04
Df Residuals:	1990	**BIC:**	3.911e+04
Df Model:	1		
Covariance Type:	nonrobust		

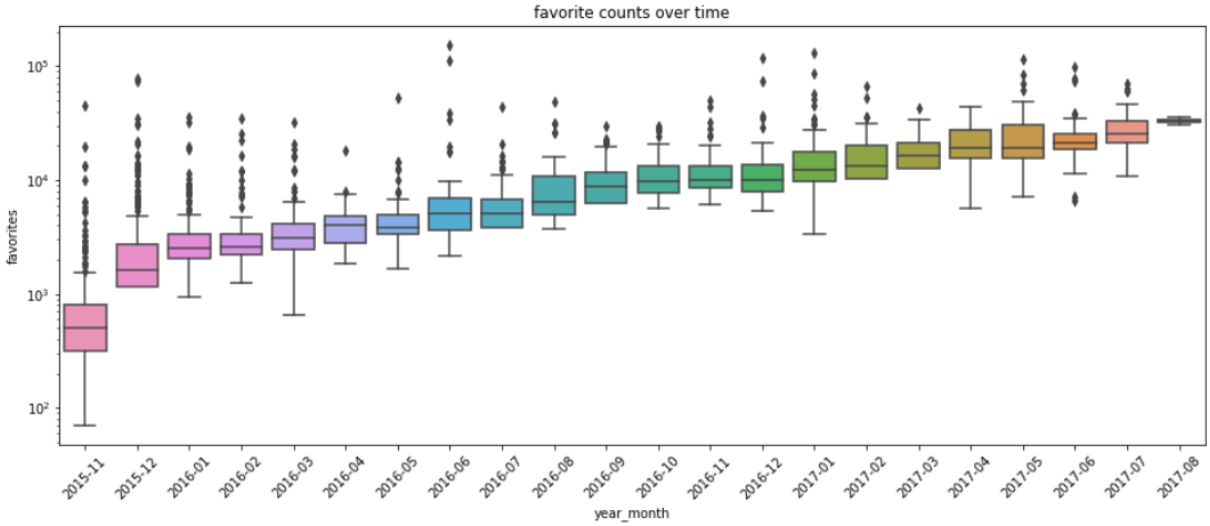
	coef	std err	t	P>|t|	[0.025	0.975]
intercept	1899.9362	113.430	16.750	0.000	1677.482	2122.390
retweets	2.5929	0.023	111.722	0.000	2.547	2.638

Omnibus:	530.580	**Durbin-Watson:**	0.880
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	17450.066
Skew:	0.576	**Prob(JB):**	0.00
Kurtosis:	17.454	**Cond. No.**	5.59e+03

4. **To explore if there is any trend/change of retweet counts and favorites over time? Since the data is across 2015 to 2017.**

Since the data was across year of 2015 to 2017, I was wondering if the retweets and favorite counts increase or decrease during the years, since internet and media developed so fast. The good news is, based on the boxplot done for retweets across 3 years, it's pretty clear that the retweet counts increased steadily. There might be a lot reasons such as Weratedog account got more clicks and impressions over the years, or more and more people started to use twitter, or because with the development of the world, iPhone are popularized, and it's easier to access to twitter which helps to increase the count of retweet and favorite. And since retweet and favorites are highly correlated, the trend of below 2 charts is very similar.




favorite counts over time
favorites
10^5
10^4
10^3
10^2
2015-11
2015-12
2016-01
2016-02
2016-03
2016-04
2016-05
2016-06
2016-07
2016-08
2016-09
2016-10
2016-11
2016-12
2017-01
2017-02
2017-03
2017-04
2017-05
2017-06
2017-07
2017-08
year_month