Act Report

By: Bassam Mohammed

goal: wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning is required for -worthy analyses and visualizations.

The Data

Enhanced Twitter Archive: The WeRateDogs Twitter archive contains basic tweet data for all 5000+ of their tweets, but not everything. One column the archive does contain though: each tweet's text, which I used to extract rating, dog name, and dog "stage" (i.e. doggo, floofer, pupper, and puppo) to make this Twitter archive "enhanced." Of the 5000+ tweets, I have filtered for tweets with ratings only (there are 2356).

Additional Data via the Twitter API: Back to the basic-ness of Twitter archives: retweet count and favorite count are two of the notable column omissions. Fortunately, this additional data can be gathered by anyone from Twitter's API. Well, "anyone" who has access to data for the 3000 most recent tweets, at least. But you, because you have the WeRateDogs Twitter archive and specifically the tweet IDs within it, can gather this data for all 5000+. And guess what? You're going to query Twitter's API to gather this valuable data.

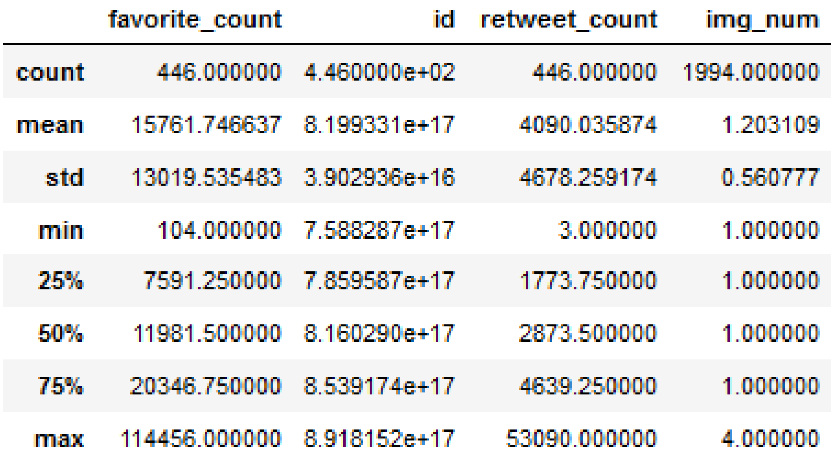
Image Predictions File One more cool thing: I ran every image in the WeRateDogs Twitter archive through a neural network that can classify breeds of dogs*. The results: a table full of image predictions (the top three only) alongside each tweetID,image URL, and the image number that corresponded to the most confident prediction (numbered 1 to 4 since tweets can have up to four images).

Analysis

I clean and analyze data from this account by show compare the favorite counts & retweet counts ,what the most source for tweet,what is the 6 frequent breed,what the most stage of doges,and number of tweets per month?, I used different types of graphs for display and analysis of matplotlib libray such as countplot,pie chart,bar chart,scatter,and plot.line

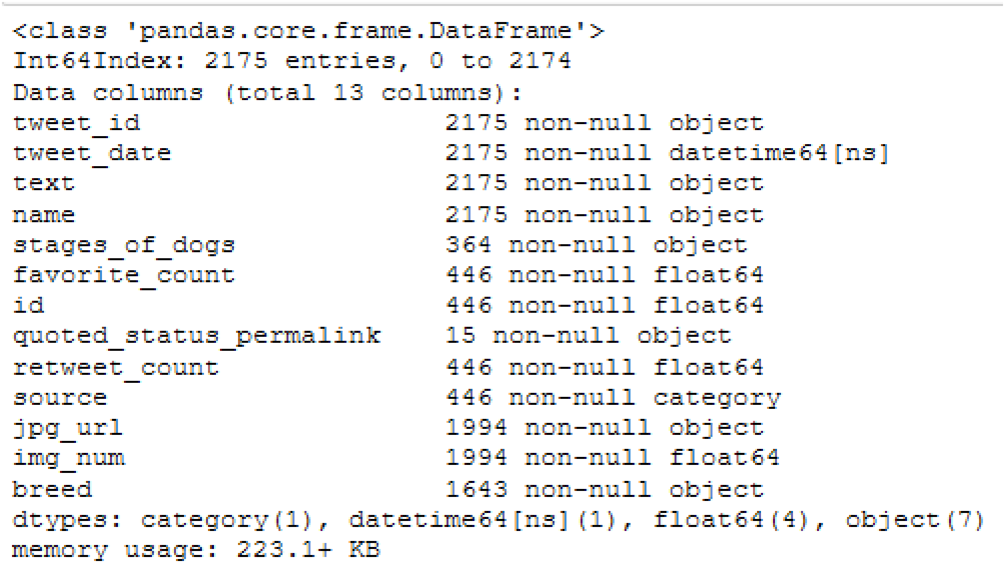
Finding

As shown in the table above that the mean (retweet:2971.322, favorite:7752.137) for retweet and favorite ,largest number of favorite equal to 107015



	favorite_count	id	retweet_count	img_num
count	446.000000	4.460000e+02	446.000000	1994.000000
mean	15761.746637	8.199331e+17	4090.035874	1.203109
std	13019.535483	3.902936e+16	4678.259174	0.560777
min	104.000000	7.588287e+17	3.000000	1.000000
25%	7591.250000	7.859587e+17	1773.750000	1.000000
50%	11981.500000	8.160290e+17	2873.500000	1.000000
75%	20346.750000	8.539174e+17	4639.250000	1.000000
max	114456.000000	8.918152e+17	53090.000000	4.000000

Merged Dataset info :

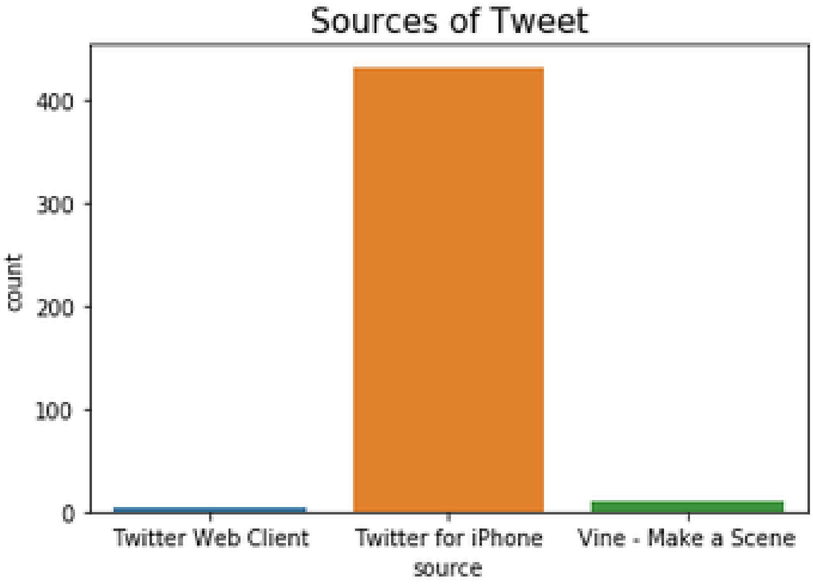


```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2174
Data columns (total 13 columns):
tweet_id                    2175 non-null object
tweet_date                  2175 non-null datetime64[ns]
text                        2175 non-null object
name                        2175 non-null object
stages_of_dogs              364 non-null object
favorite_count              446 non-null float64
id                          446 non-null float64
quoted_status_permalink     15 non-null object
retweet_count               446 non-null float64
source                      446 non-null category
jpg_url                     1994 non-null object
img_num                     1994 non-null float64
breed                       1643 non-null object
dtypes: category(1), datetime64[ns](1), float64(4), object(7)
memory usage: 223.1+ KB
```

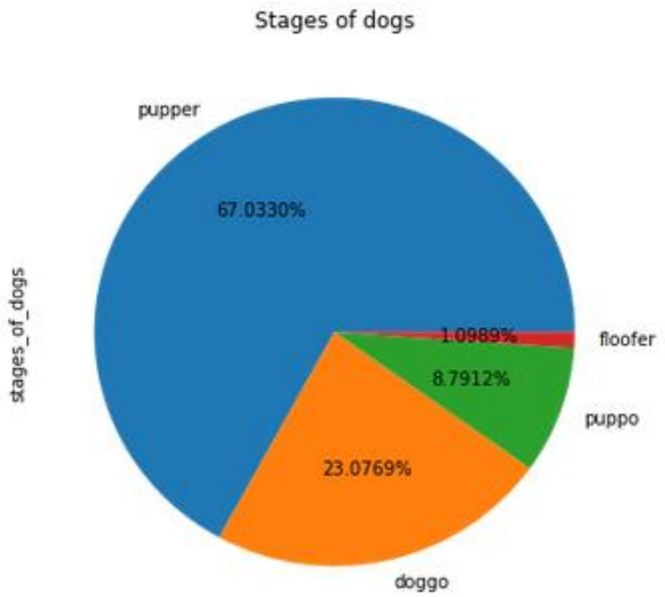
The dataset have 2356 observations,
12 columns and with no null values. The data types of the variables are divided in 4
float, 1 datetime , and 7 object.

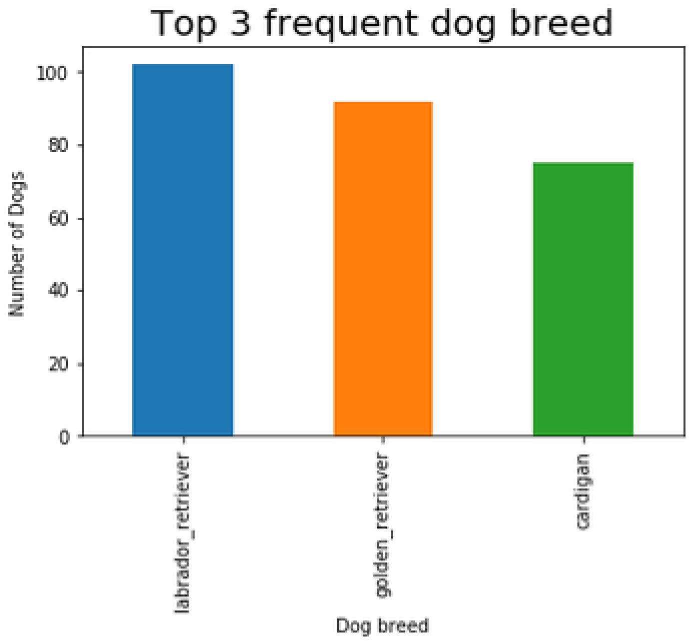
Analyzing and Visualizing Data

1. What is the most used source? Iphone is most used source for tweet

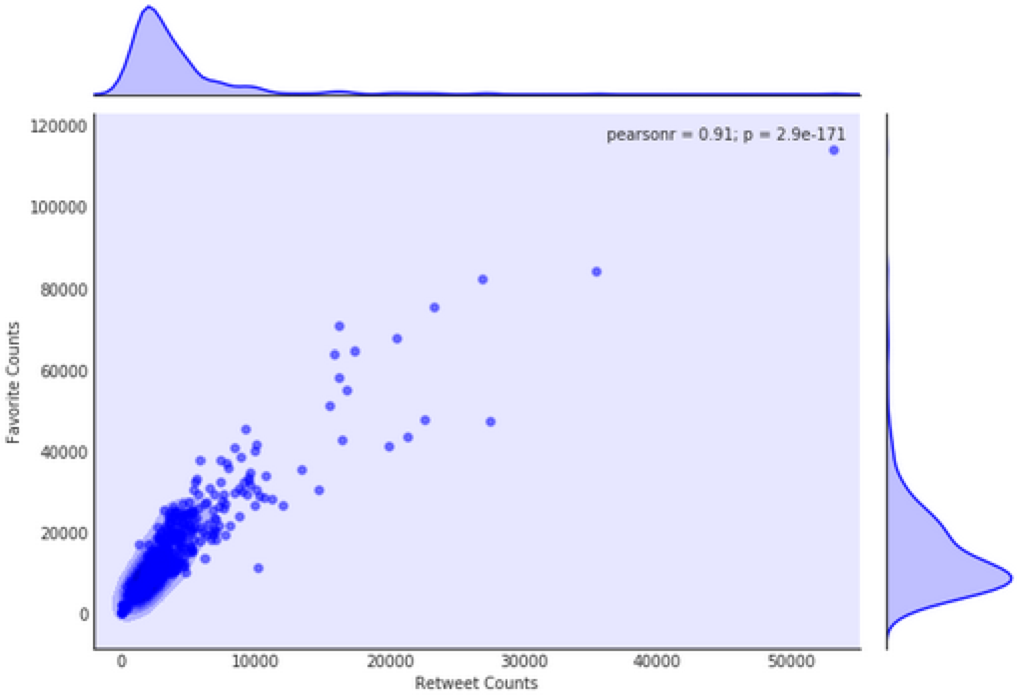


2. What is the most stage of doges?pupper the most stage of doges , doggo ,puppo then floofer

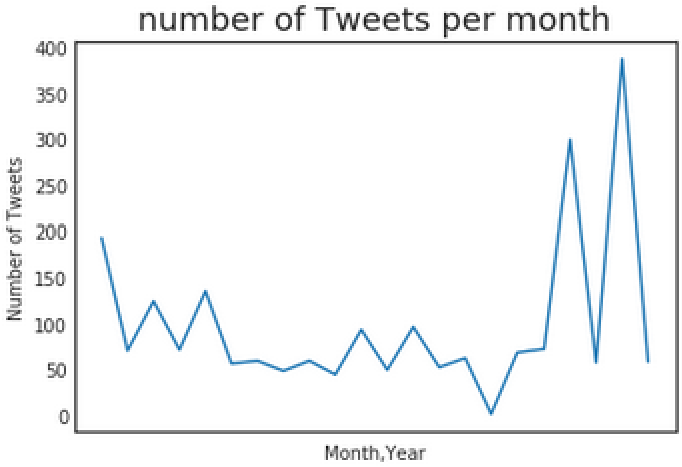




Visualization compare the favorite counts & retweet counts



Number of Tweets per month?
Notice the most increase in the number ion 12 month ,2015 year


number of Tweets per month
Number of Tweets
Month,Year