Wrangle and analyze twitter data
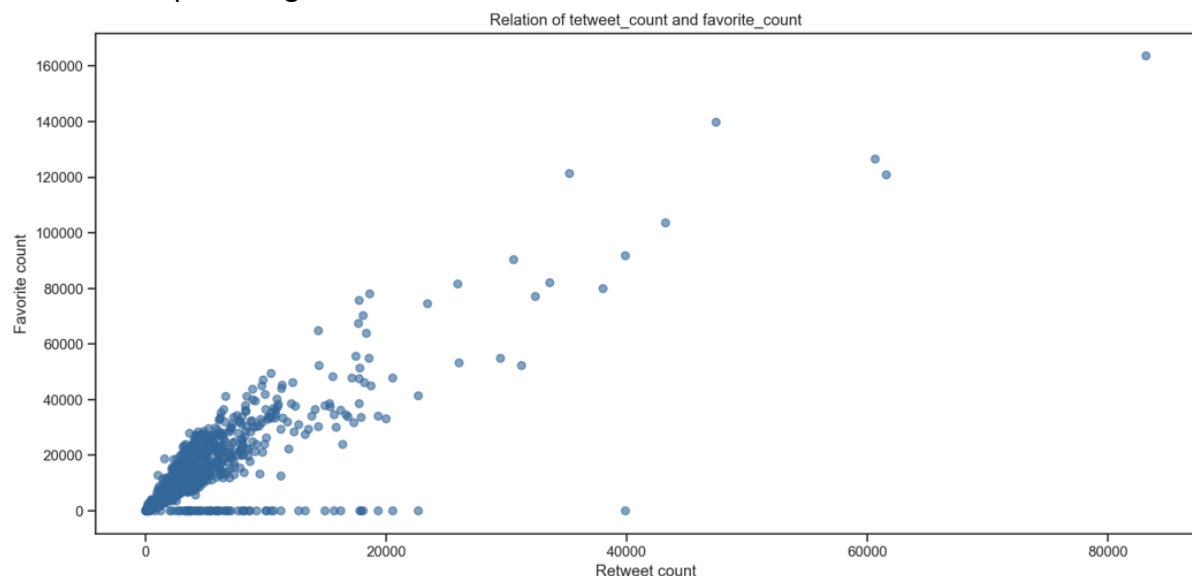
By Hana AL-alawi, 25 February 2019

@weRateDogs is the twitter account to rate people's dogs by special rating system. @weRateDogs has 7.78 million followers it is popular account.

At the first when I start working on this project I had some wondering questions like, *What is the most liked dog type?, Is the relation between retweet and likes strong or weak?, What is hashtags used with dog's images?, What is the dog type having the highest rate in years? And I start to answer it.*

Retweet and Favorite:

Each tweet has retweet count and favorite count. people express their admiration about tweets by retweet it or like it. The high retweet and favorite tweets are extremely popular. I used Scatter plot to figure out the relation between Retweets and likes.

This visualization shows how the relation between favorite and retweet count is strong, and the correlation coefficient is 0.85 with Pearson correlation coefficient, this correlation is strong positive make sense because when the tweet has more retweet will liked more too.
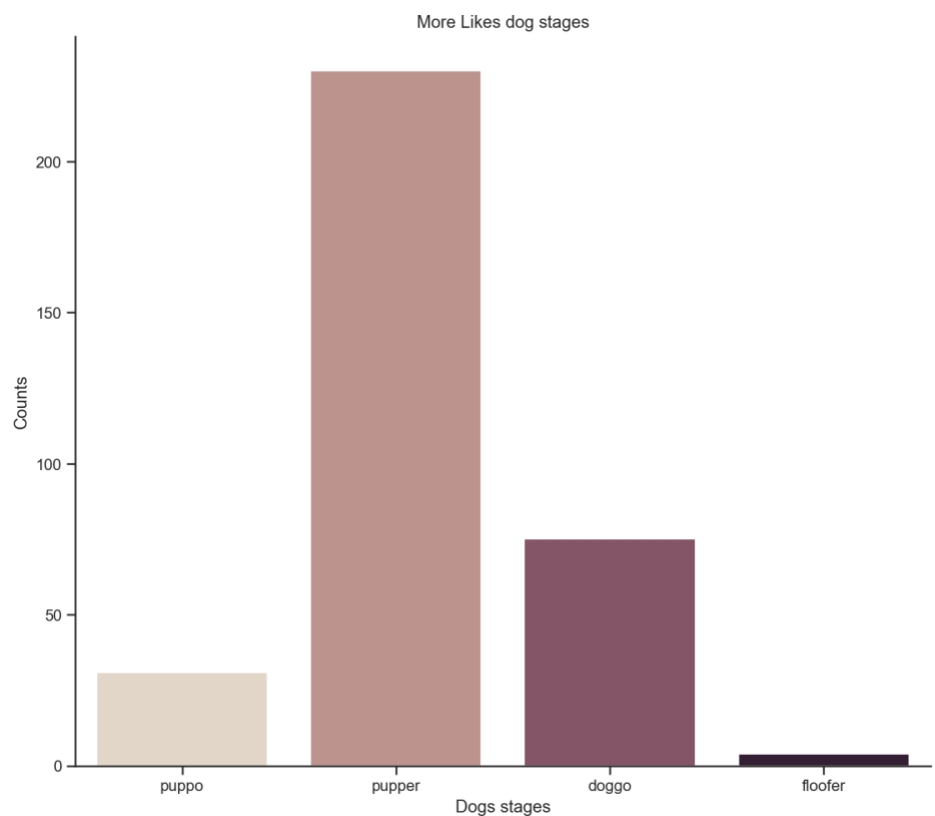
Most dog stages like:

From people's tweets (text column in the data frame) I searched for dog types. The dog types are Puppo, Floofer, Doggo and Pupper. I used Barplot to visualize and analyze.
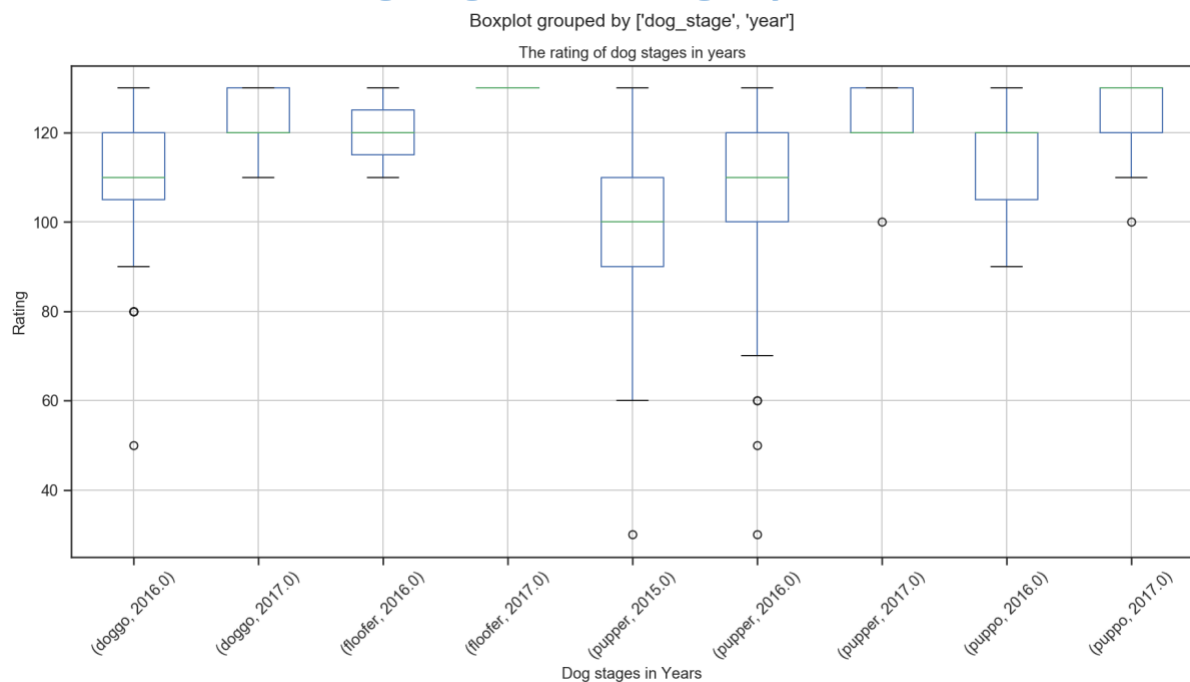
From the visualization the most liked dog stage is Pupper.
Pupper comes on tweets more than 250 times and the second is Doggo comes more than 50 times on tweets.

More Likes dog stages

	count	mean	std	min	25%	50%	75%	max
dog_stage								
doggo	75.0	113.866667	14.967833	50.0	110.0	110.0	125.0	130.0
floofer	4.0	122.500000	9.574271	110.0	117.5	125.0	130.0	130.0
pupper	230.0	105.478261	17.320179	30.0	100.0	110.0	120.0	130.0
puppo	31.0	120.000000	11.832160	90.0	115.0	120.0	130.0	130.0

Dog stages and rating in years:

Boxplot grouped by ['dog_stage', 'year']
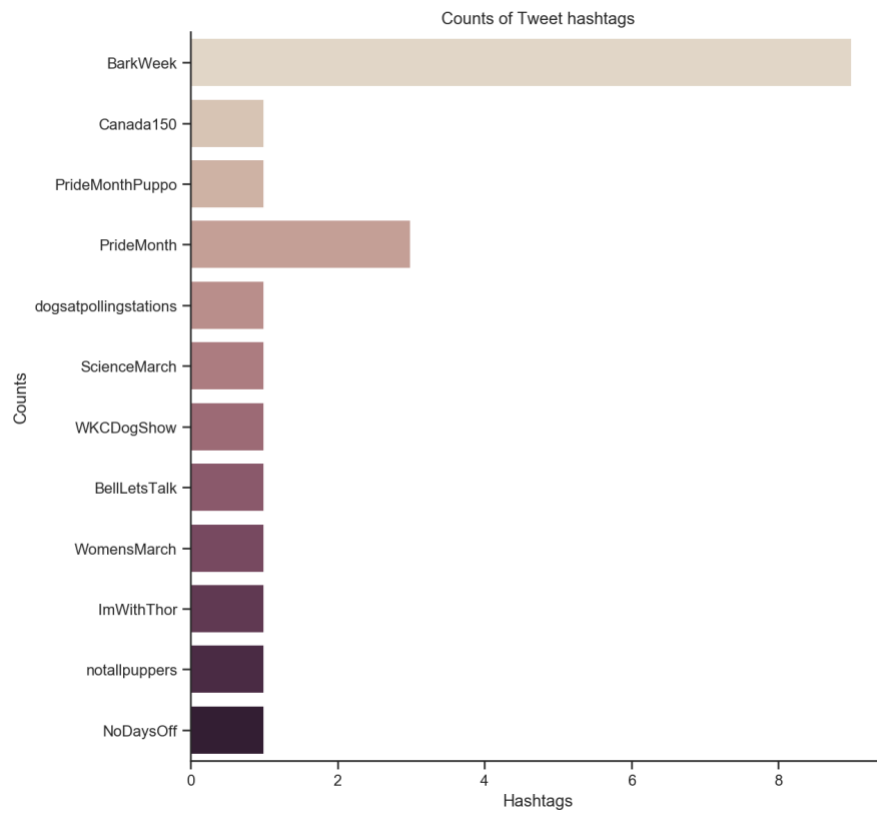
The rating of dog stages in years

This visualization shows each dog stage and its rating in years.

Tweet hashtags:

People using hashtags with their tweets to give the tweet keywords to categorize it, from tweets (text column) in data frame I extracted hashtags. Hashtags are the word after # symbol.

Counts of Tweet hashtags

The visualization shows the tweet hashtags, the Most hashtag repeat are #Barkweek and #PrideMonth.