Wrangling project analysis report

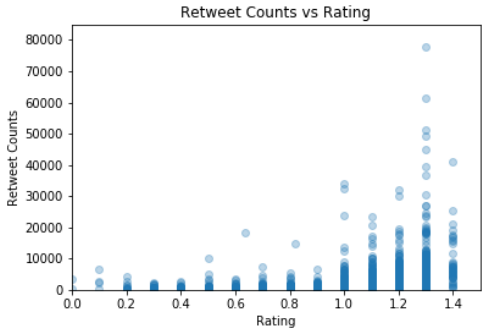
Prior to analyzation and visualization, gathering, assessing and cleaning were performed on dataset.

To analyze data, I have created rating column which is rating numerator divided by rating denominator. Through this rating column, I was able to know that dog pictures get average rating of 1.16. Some of them got 0 out of 10 because of plagiarism and no dog on the picture. Some pictures that are not dog also received 2/10 or 3/10 as well.

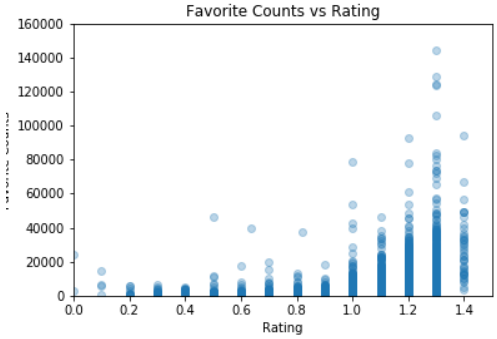
Another analysis was done to find out which source was most used to tweet. Majority was Ipone followed by vine, twitter website, and twitter deck.

It was also easy to find which name is most popular for dogs. Lucy and Charlie was the most common names followed by Cooper and Oliver.

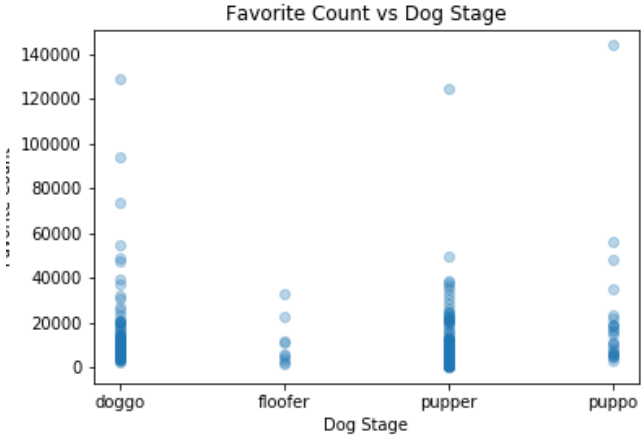
To check relationship between rating and retweets, I plotted scatter plot. Scatter plot showed that around rating of 1.3 got most retweets. However higher than 1.3 did not get more retweets which did not meet my assumption of higher rating getter more retweets.



I also checked relationship between rating and favorite counts. Again, I assumed higher ratings getting more favorite counts. It was somewhat true since it gradually increased until rating 1.3, but higher than 1.3 did not get as many favorite counts.



Last thing I checked was which dog_stage is the most popular. My assumption was puppo since I like baby dogs. The result actually showed that pupper is the most popular dog_stage. However, post that got most favorite counts was at puppo stage.


Favorite Count vs Dog Stage
140000
120000
100000
80000
60000
40000
20000
0
doggo
floofer
pupper
puppo
Dog Stage