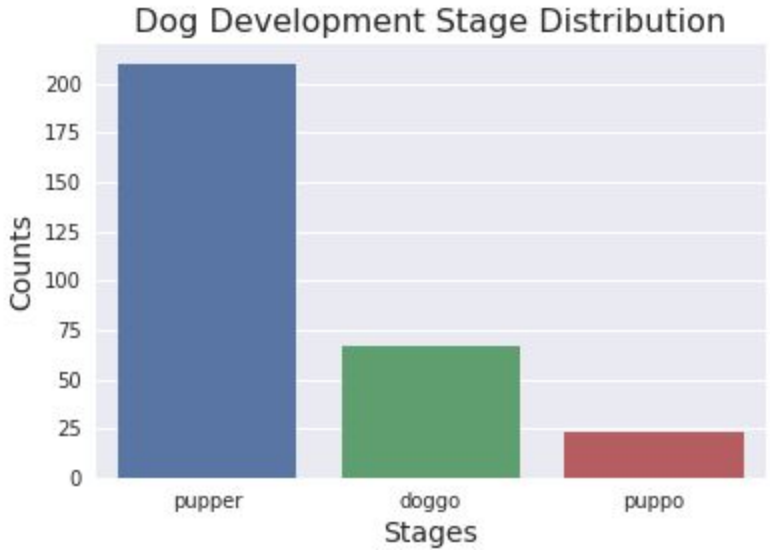
Analysis insights and visualization

Introduction

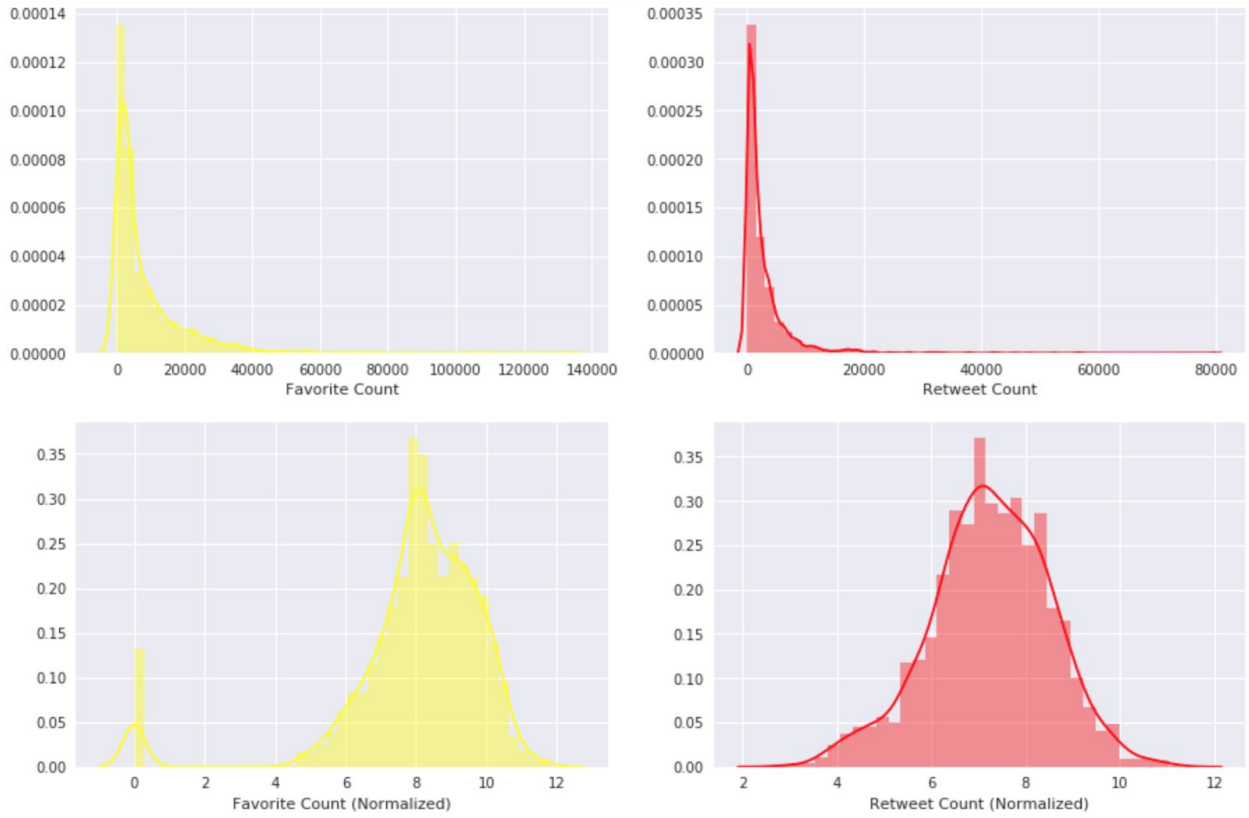
This report communicates the insights and displays the visualizations produced from the wrangled WeRateDogs Twitter archive data.

Dog Development Stage Distribution



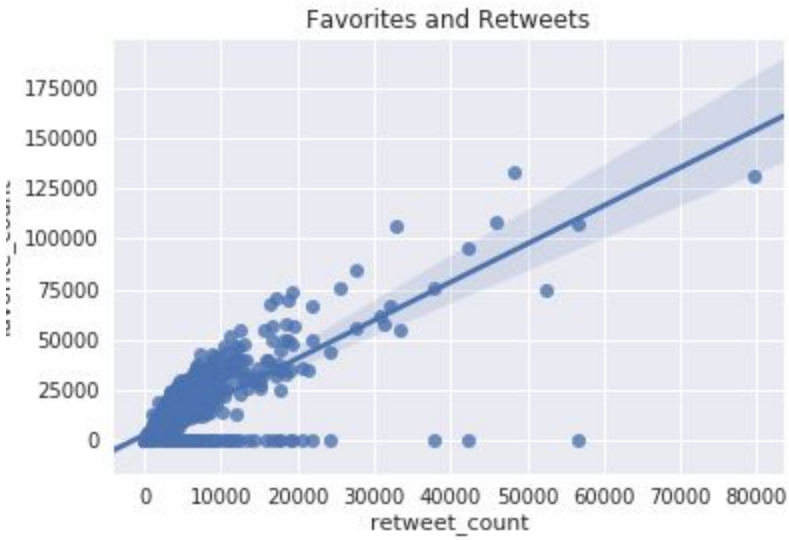
The distribution of dog stages shows that the ‘pupper’ dog development stage (a small doggo) is the most prevalent dog stage. It is followed by the ‘doggo’ and ‘puppo’ dog development stages. This can be explained by the fact the unmature dogs are cuter than their adult counterparts. A small note however, there was huge amount missing data in dog stages, thus the distribution can be slightly skewed and not reflect the truth at all… And this chart only represents the top 3 dog stages. Stages [nan, 'floofer', ('doggo, puppo'), ('doggo, floofer'), ('doggo, pupper')] are missing from the chart because their counts are relatively low to the top 3.

Distribution of retweet_count and favorite_count



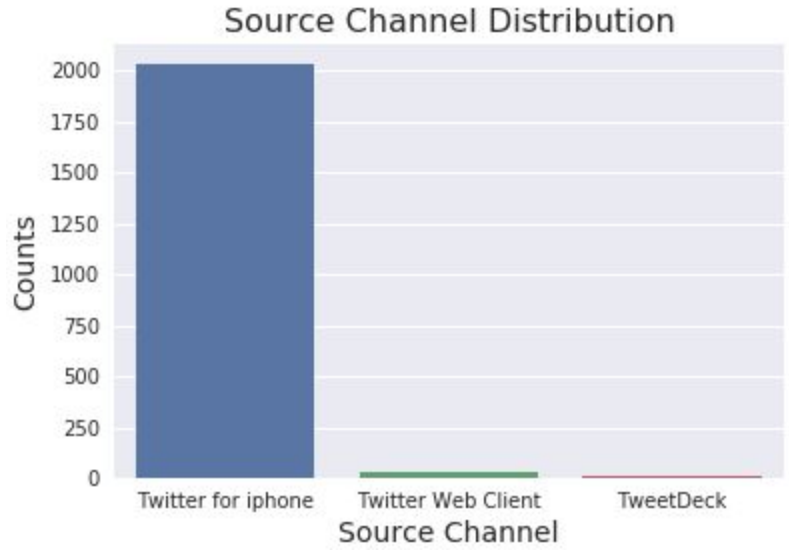
The non-normalized distributions for both favorites count and retweet count have long positive tails. This suggests that top end popular tweets are super rare. The normalized graphs also show similar distributions. The similarities between the two variables might give insight that they are both correlated… See next graph.

Correlation between retweet_count and favorite_count



Just like I suspected there seems to be a strong positive relationship (r-squared is 0.790481) between ‘retweet_count’ and ‘favorite_count’. My guess is that most the popular tweets get retweeted and favorited because of its popularity.

Source Channel Distribution



This clearly shows that Twitter for iPhone is the dominant channel source for WeRateDogs tweets, retweets, posts, and others. Hardly no one uses either the twitter web client or tweetDeck for the WeRateDogs channel...