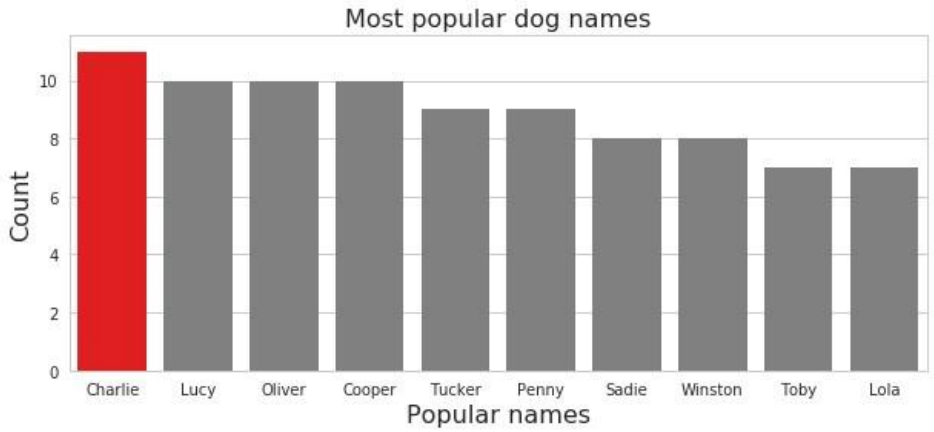
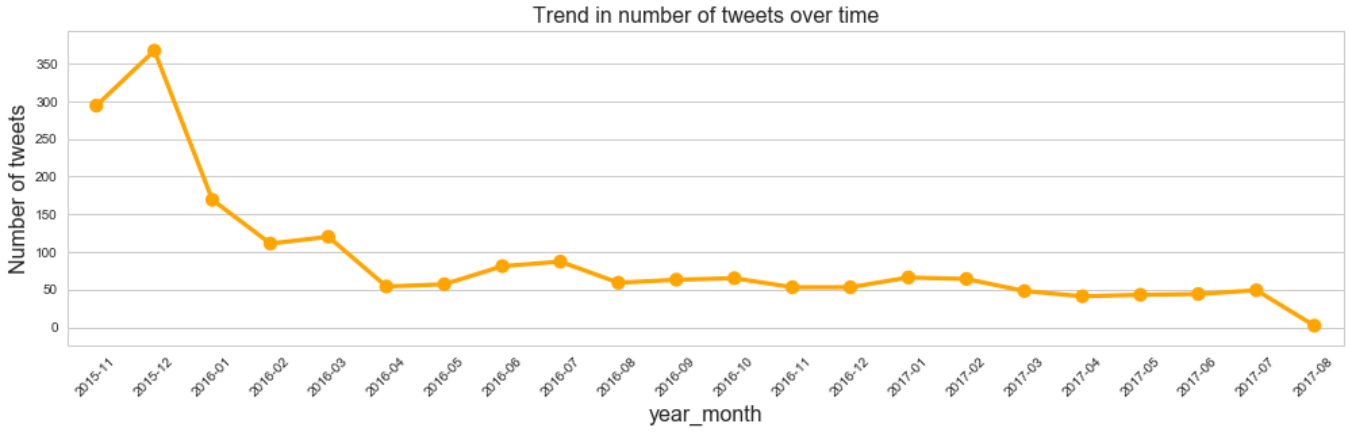
WeRateDogs Twitter - Data Analysis Insights and Visualization

1. Most popular dog names



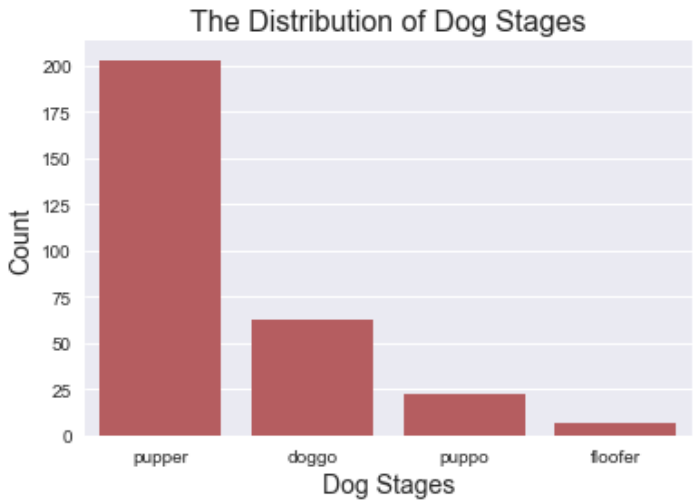
Charlie seems to be the most popular dog name in the dataset.

2. Analyzing trend in number of tweets posted over time



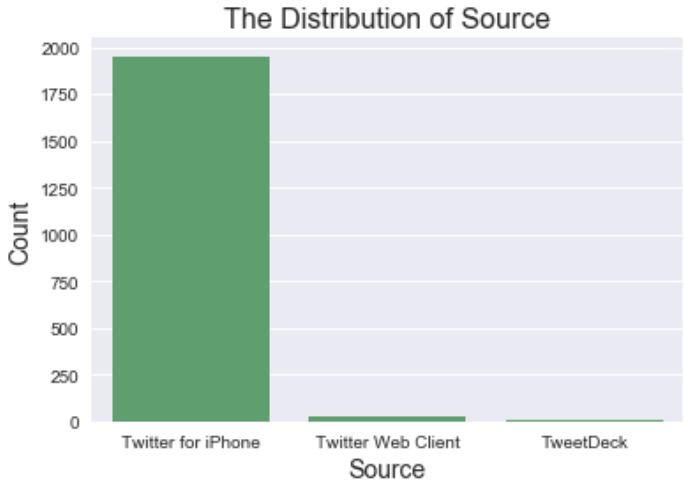
There is a decreasing trend in the number of tweets posted between Nov ’15 and April ‘ 16. After Apr ’16 the numbers are almost similar till Jul ‘ 17 after which there is drop again.

3. The Distribution of Dog Stages



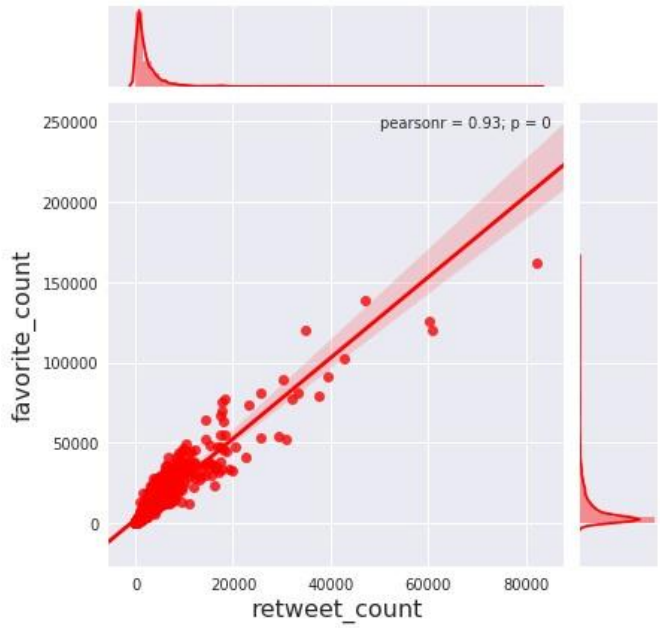
There is a lot of missing data in dog stage. But if we do not include missing data, then the distribution of dog stages shows that ‘pupper’ (a small doggo, usually younger) is the most popular dog stage, followed by ‘doggo’, ‘puppo’ and 'floofer'. It could be due to the young and unmatured dog is usually cuter than the adult dog.

3. Distribution of source



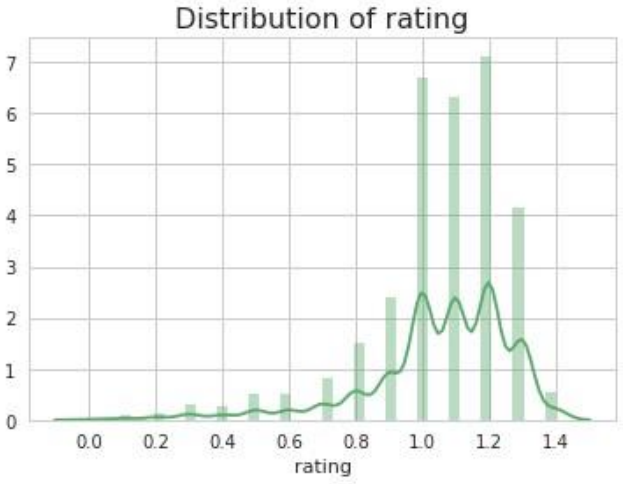
The dominate source of tweets is from iPhone twitter app, which is 94% in the total. That means the twitter app is the main channel for people using to tweet, retweet, post, and others, while the TweetDeck and Twitter pretty rare (less than 1%).

4. Relationship between retweet_count and favorite_count



It is seen that most popular tweets usually get a large number of retweets and favorite counts. As shown above correlation between 'retweet_count' and 'favorite_count' is 0.929101, that is a high value showing a strong positive relationship between 'retweet_count' and 'favorite_count'. The plot above also confirms this linear relationship.

5. Distribution of Rating



rating is calculated as rating_numerator divided by rating_denominator (always 10). Most dogs receive ratings between 10 and 13 or 1.0 and 1.3