Visualization

From the previous steps to analyze and clean, now the data is ready to be visualized and found insights from.

Since the data is merged from 3 different tables it is joined on the tweet id and a suffix has been added for each column depending on the table they are coming from.

There are a lot of different way of uploading posts these days. One could use different devices/platforms/applications to upload posts. In our dataset we try to analyze the number of types of sources people use to upload these posts. We find out that there are only 4 ways people have used to upload or at least the twitter algorithm has tagged them into 4 options.

1. Phone users
2. Vine application to upload the pictures
3. Twitter web client
4. Tweetdeck dashboard

Given the need to understand the platforms people use to access a service is key. Depending on the results the company might choose to focus or prioritize one over the other. Since there is a column which mentions the likes a tweet has received, we take the analysis to compare the average likes these posts have received depending on the applications they have used to upload the posts. It seems like that the posts uploaded through iPhone are more popular/liked than the web client. That might say something related to the kind of userbase each of these applications have.

I also wanted to see how the rating given by user who has made the post reflects the overall twitter population to react by liking/retweeting on it. This is a study focused towards behavioral science. How one rates their own dogs. Most likely people who posts are going to rate them high, but are they actually reflective of how much others like that post? It may be that some posts are just better appealing to a random user but then does the dog owner’s rating of the dog influence?

It seems like the posts having rated as 13 have been the highest liked as well as highest retweeted post. Although, there might be a multicollinearity in this but its worth looking at.