act_report

August 11, 2022

0.1 Report: act_report

- Create a **250-word-minimumwrittenreport** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

Analyzing and Visualizing Data For analysis, we took a look at the merged dataset and trying to find correlations between variables, we plotted a visualization in order to beteer understand the dataset. As a requirement, we arrived at 3 insights and at least 1 visualization.

Insight 1: using this function: twitter_master.rating_numerator.value_counts().sort_index(ascending=False)

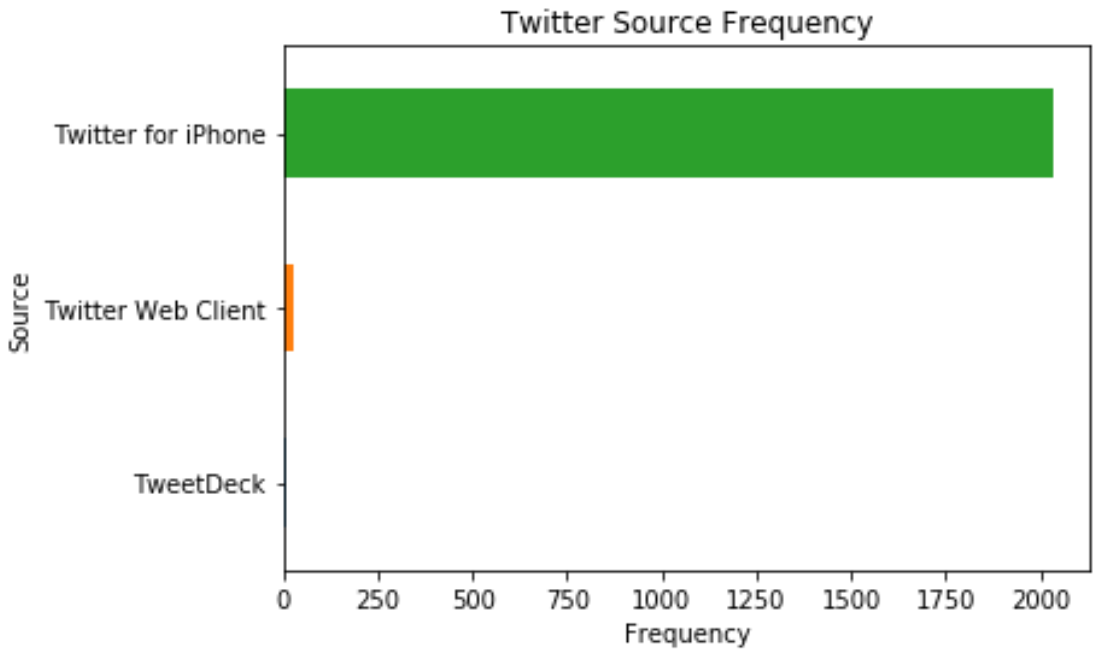
From the results gotten from the code above, it can be observed that the most frequent rating given to the dogs was 12, with 474 tweets containing that value as the numerator,this was followed closely by 10 with 429 occcurences respectively

Insight 2: The 'source' column contains information regarding the source of the tweet, it is hidden behind an html link which makes it difficult for us to see the source,after extraction we had an insight that the most common WeRateDogs tweets source is the iphone source.

Insight 3: lastly the most common breed predicted is the golden retriever

Visualization For our visualization, we plotted a graph with two variables to check the correlation between 'source' and count of tweet represented by 'tweet_id'

We did that to know which source produced the most tweets and by what margin. We pivoted the data to make a plot, but before we did that, we had to do some further cleaning by dropping the duplicate values in the 'tweet_id' column as it would prevent us from being able to pivot the dataset. From the visualization, we csaw that 'Twitter For Iphone' was the major source of all the tweets that we populated and analysed, while TweetDeck has the lowest number of users.


Twitter Source Frequency
Twitter for iPhone
Twitter Web Client
TweetDeck
Source
0
250
500
750
1000
1250
1500
1750
2000
Frequency

In [ ]: