Act_report

0.1 Report: act_report

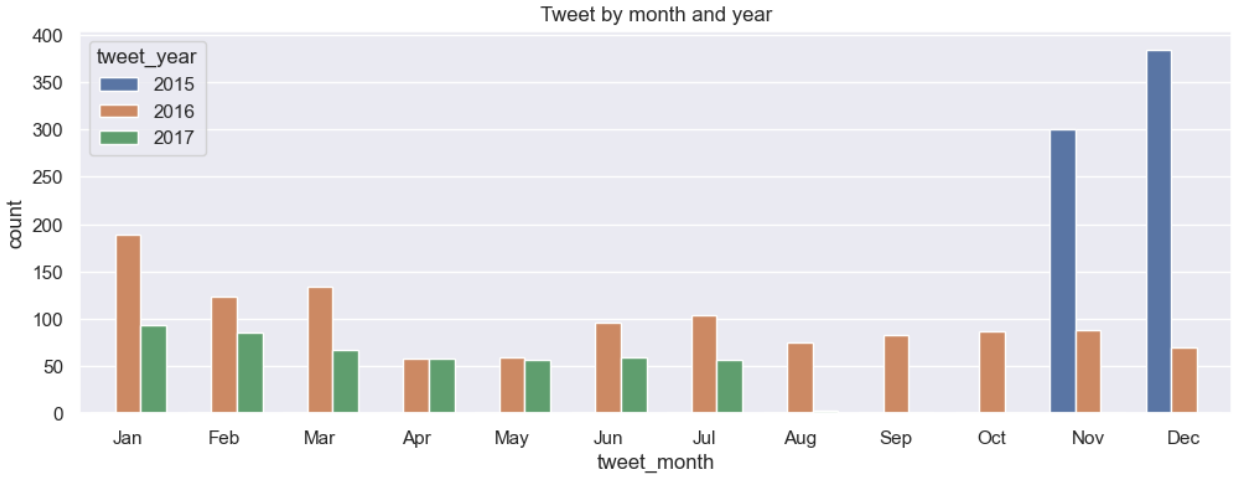
The following is an analytical walk-through of the Udacity data analyst nano-degree program's Data wrangling and analysis project.

Following the wrangling stage, the unified dataset was saved as a twitter archive master dataset and read into a pandas dataframe. Certain research questions were also considered to be addressed, including:

- What month and year had the most and least tweets?
- Which dogs (by name) had the highest average rating (%) in the top 15?
- Which dog has received the most retweets and likes (favorite)?
- Is there a link between the number of retweets and the number of favorites?

R1. Which month and year had the most and least tweets?

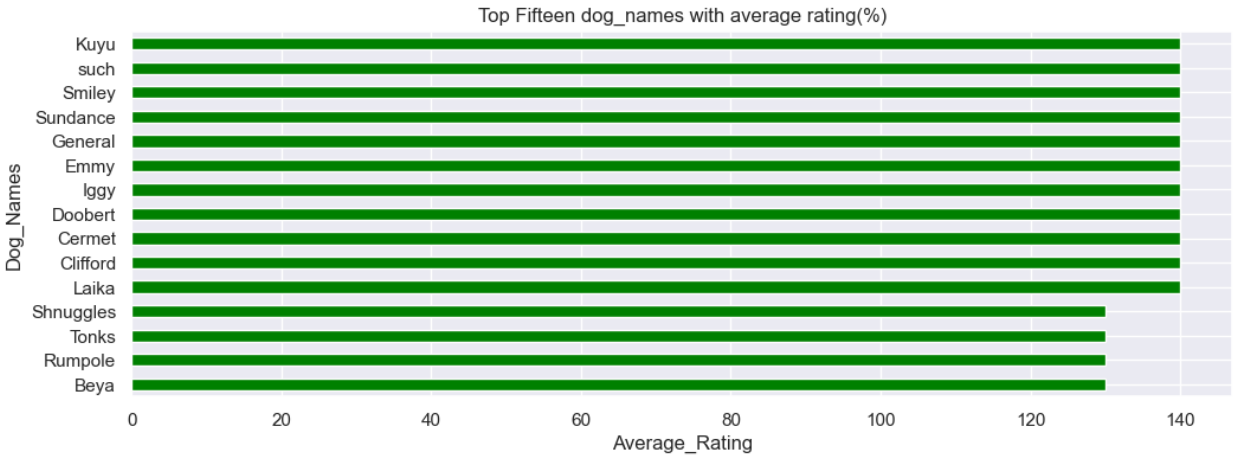
The timestamp column was divided to extract the tweets' day, month, and year to address this query. The countplot chart was then used to depict tweets produced over months, with the year as the colour, as seen in the diagram.



Insights: The WeRateDog tweets campaign had the highest interaction in the second month of its launch (December, 2015) and the lowest in the last month of the campaign (July, 2017).

R2. Which dogs (by name) had the highest average rating (%) among the top 15?

I used the. groupby () function to group the data by dog names and sort by rating in descending order to find the top 15 dogs with the highest average rate.

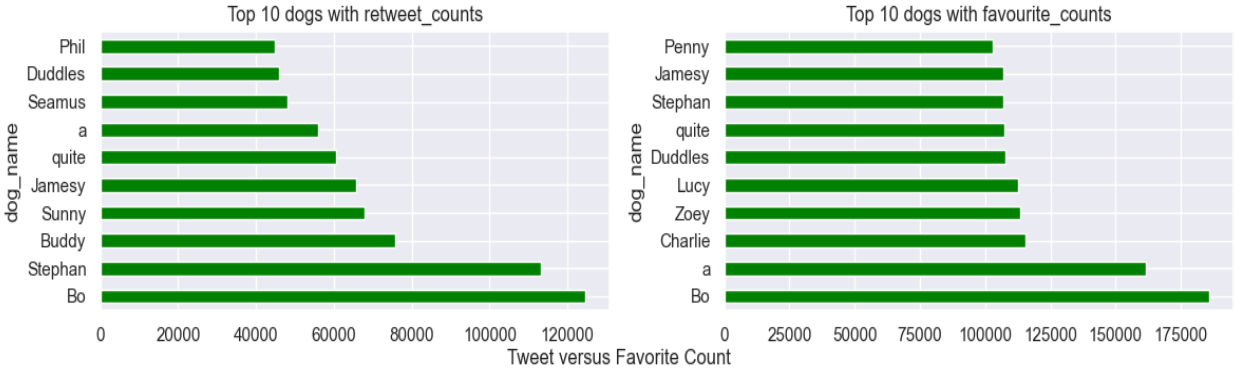


It was discovered that there is a level of tie in dog rating, we may simply state that this is due to the fact that that they were all great dogs and that rating them was a difficult chore

Meet some of the dogs in this category.

R3. Which dog has received the most retweets and likes (favorite)?

While assessing this, it was discovered that a significant number of dogs lacked names; so, my first method was to filter dogs whose name column was 'None' before moving on to visualize the filtered data; the outcome is shown below.



It turns out that Bo is the dog with the most tweets and likes (favorite). Is she not stunning?

Is there a link between the number of retweets and the number of favorites?

Trying to determine whether an increase in the number of tweets will result in an increase in the number of likes, as we saw with Bo. I created a scatter plot and examined the correlation between the two columns; the association appeared to be strong enough.