Data, Dogs, Tables, and Tweets

Ojo Michael,

Data Analysis and Visualization report

Using the merged and cleaned data set created from the wrangling result from the three data sources, a bar chart was created to show which platform is used the most to access the WeRateDog Twitter page.

Insights: From the bar chart it can be seen that the Twitter mobile app on iPhones is the most frequently used to view the WeRateDogs Twitter page over the period the data was collected, followed by the Web clients or browsers, with the TweetDeck tool been the least used.

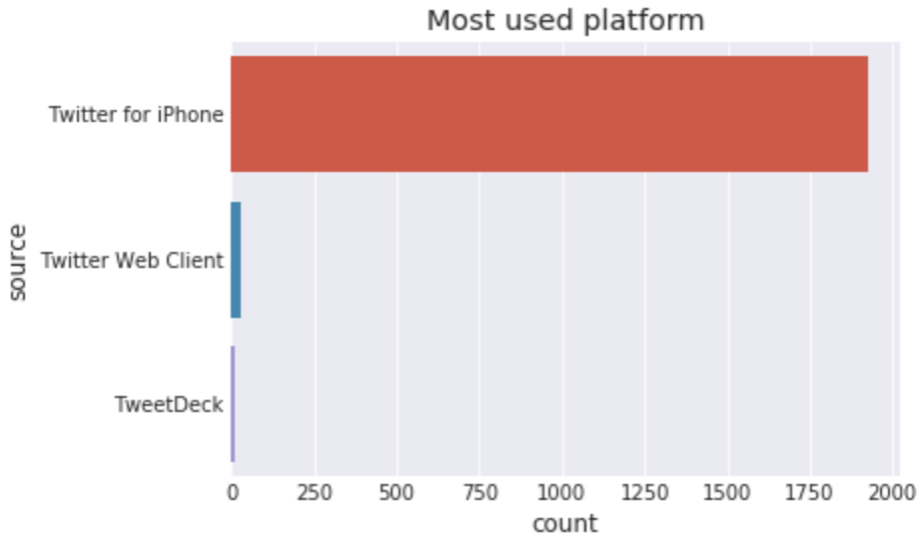


FIGURE 1 BAR CHART SHOWING THE MOST USED PLATFORMS TO VIEW THE WERATEDOG PAGE

Also, a bar chart was done to show the frequency of tweets for each day of the week over the given period.

Insights: from this visualization, Monday is the day where most tweets occurred on the WeRateDogs over the observed period.

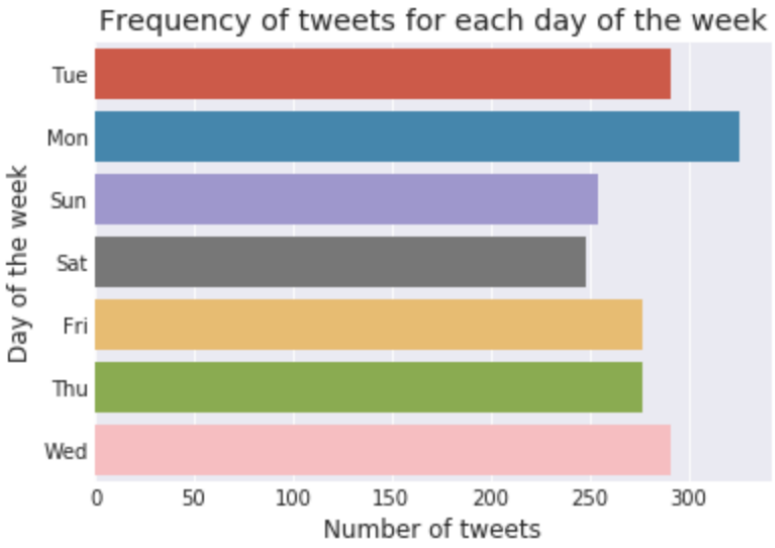
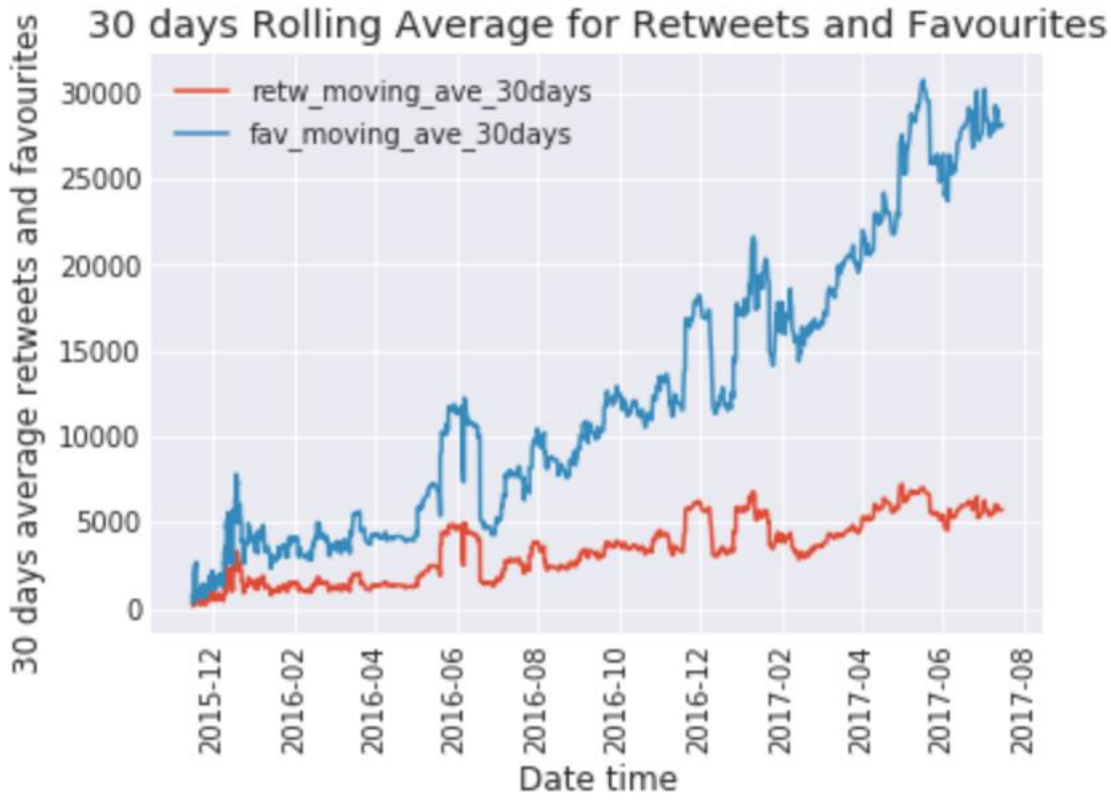


FIGURE 2 BAR CHART SHOWING THE FREQUENCY OF TWEETS FOR EACH WEEK

To represent the relationship between the number of retweets and favorites over the period captured by the datasets, a 30-day moving average line chart showing the smoother trend lines of favorites and retweets throughout the recorded data i.e. December 2015 to August 2017.

Insights: The line chart shows the trend of gradual increase in likes(favorites) is at a higher rate than the retweets over the period observed, with a steady period of spikes and drops in retweets and likes mirroring each other.



A pie chart showing the percentage of dogs at different stages;

Insight: From the pie chart, Dogs at the pupper stage are the most commonly mentioned in tweets over the period the data covered with 66% of the known dog stage totals.

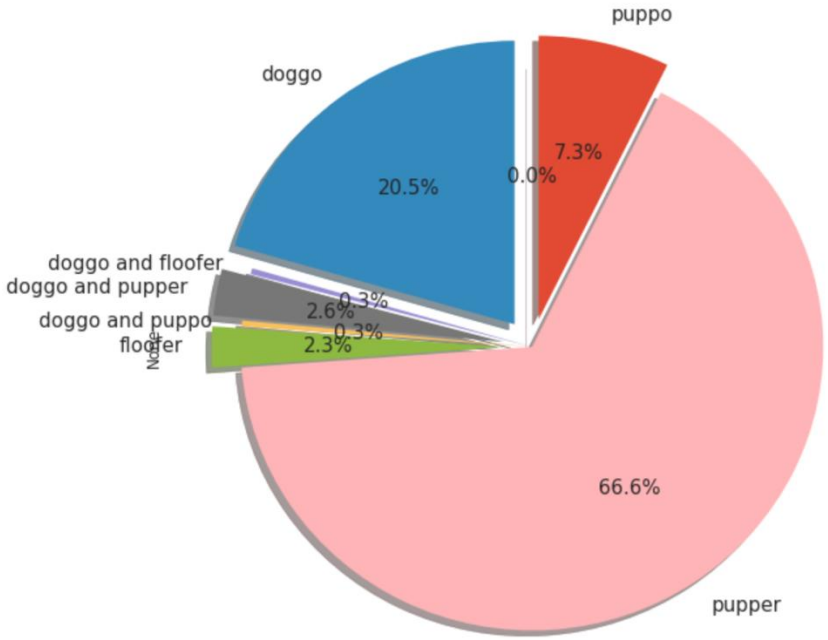
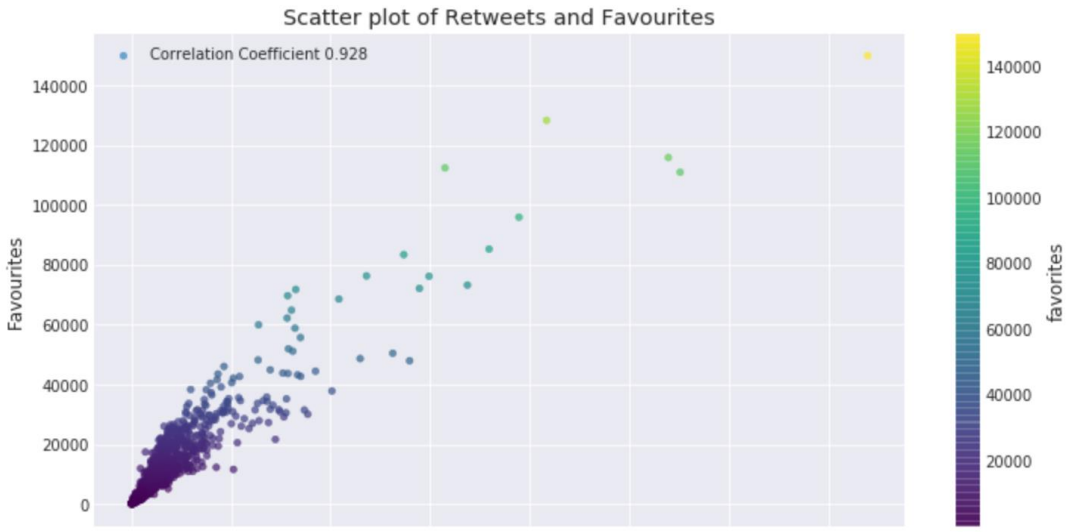


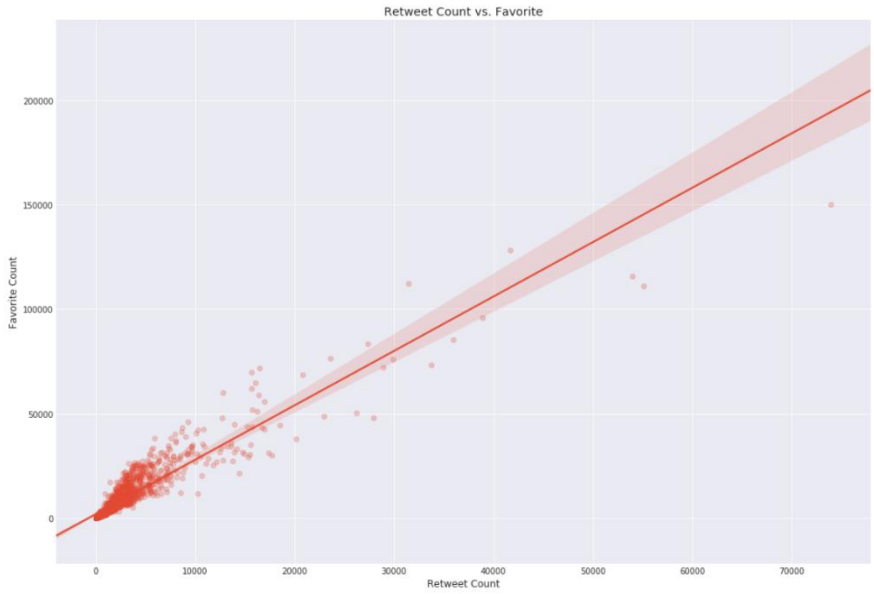
FIGURE 3 A PIE CHART SHOWING THE PERCENTAGE OF DOGS AT DIFFERENT STAGES

A scatter plot diagram showing the relationship between the number of retweets and favorites throughout the data.

Insights: From the scatter plot visualization and correlation coefficient value, the number of retweets has a positive and strong correlation with the number of favorites.

FIGURE 4 A SCATTER PLOT DIAGRAM SHOWING THE RELATIONSHIP BETWEEN THE NUMBER OF RETWEETS AND FAVOURITES





A word cloud diagram was used to display the most tweeted about dog names used on the Twitter page.

Insight: From the word cloud diagram Copper, Charlie and Oliver are the names most frequently used as dog names in the tweets.

FIGURE 5 A WORD CLOUD DIAGRAM DISPLAYING THE MOST TWEETED ABOUT DOG NAMES ON THE TWITTER PAGE.

The box plot diagram bellows showing the distribution of Rating numerator scores using only records that have a denominator value that is equal to 10 for uniformity. From the boxplot diagram, we can see that we have outliers in the data, with values greater than 350.

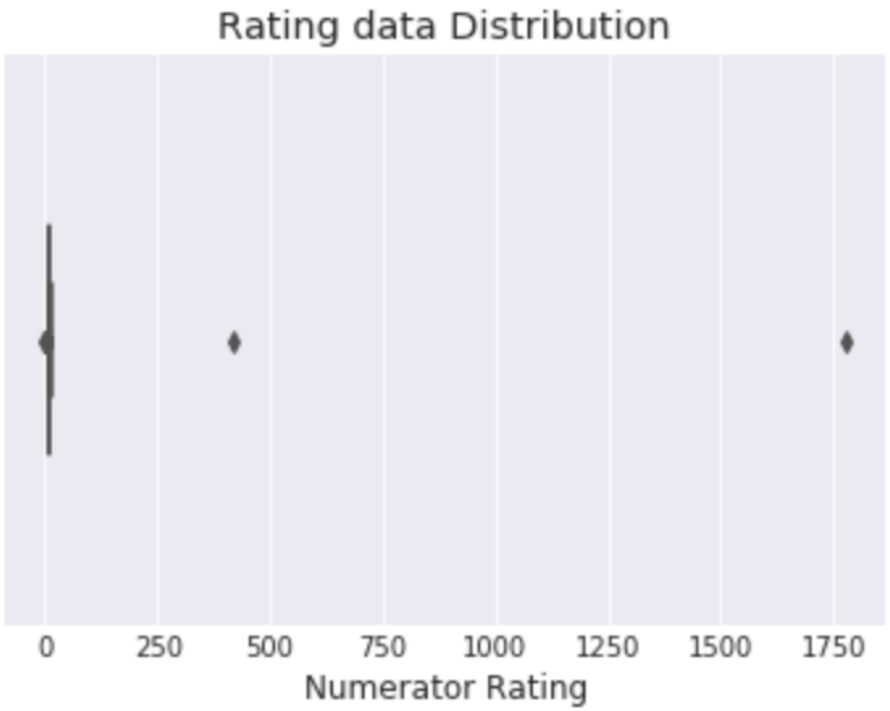


FIGURE 6 A BOX PLOT DIAGRAM SHOWING THE DISTRIBUTION OF RATING NUMERATOR SCORES USING ONLY RECORDS THAT HAVE A DENOMINATOR VALUE THAT IS EQUAL TO 10

Insight: From the images, the record with a 420 rating is not a dog, but a human, the rapper "Snoop Dogg" (Calvin Cordozar Broadus Jr) but seems to be in line with the humorous nature of the tweeter page. Also, the picture with the highest rating of 1776 is a dog in a disguise, with an even neural network algorithm predicting only the sunglasses and the bow tie on the dog and also interesting enough a microphone for a rapper.

```
plot diagram we can see that we have outliers ,with values greater than 350
```

```
d2[(df_merged2.rating_denominator2 == 10) & (df_merged2.rating_numerator2 >= 350)]
```

nf	retweets	favorites	dog_stage	day_of_week	rating_numerator2	rating_denominator2	prediction	fav_moving_ave_30days	retw_moving_ave_3
76	2370	4935		Mon	1776.0	10.0	Not likely a Dog	5205.333333	1731.6
13	7916	22855		Sun	420.0	10.0	Not likely a Dog	1871.000000	762.5

FIGURE 7 TABLE OF RECORDS WITH RATINGS NUMERATOR VALUES OF 420 AND 1776

FIGURE 9 DOG WITH HAT AND SUNGLASSES

FIGURE 8 SNOOP DOGG ”(CALVIN CORDOZAR BROADUS JR

References

- Mueller, A., 2020. Minimal Example — wordcloud 1.8.1 documentation. [online] Amueller.github.io. Available at: https://amueller.github.io/word_cloud/auto_examples/simple.html#sphx-glr-auto-examples-simple-py [Accessed 23 April 2021].
- Amipara, K., 2017. Better visualization of Pie charts by MatPlotLib. [online] Medium. Available at: https://medium.com/@kvnamipara/a-better-visualisation-of-pie-charts-by-matplotlib-935b7667d77f [Accessed 23 April 2021].
- khandelwal, p., 2021. display an image from web using IPython.display.image: sometimes work, sometimes not. [online] Stack Overflow. Available at: https://stackoverflow.com/questions/40645782/display-an-image-from-web-using-ipython-display-image-sometimes-work-sometimes [Accessed 23 April 2021].
- Richardson, L., 2020. Beautiful Soup Documentation — Beautiful Soup 4.9.0 documentation. [online] Crummy.com. Available at: https://www.crummy.com/software/BeautifulSoup/bs4/doc/ [Accessed 23 April 2021].