act_report

April 23, 2019

0.1 Visualizations and Analysis for "We Rate Dogs" Twitter Archive

0.1.1 What is "We Rate Dogs" Twitter Account ?

It is a Twitter account that rates people's dogs with a humorous comment about the dog. It started in 2015 by the college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter.

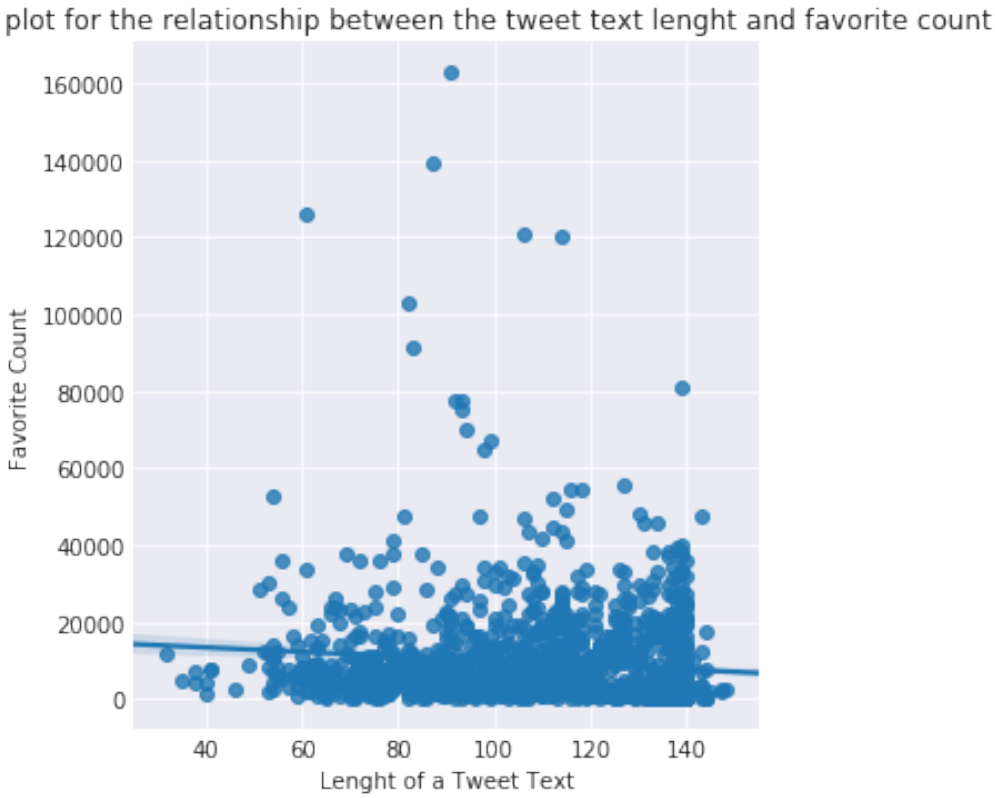
In the following sections, there are several visualizations and analysis related to the twitter account of "We Rate Dogs".

```
In [16]: import pandas as pd
         import csv
         import matplotlib.pyplot as plt
         import seaborn as sns
         sns.set_style('darkgrid')
         %matplotlib inline
         import warnings
         warnings.filterwarnings("ignore")
```

```
In [17]: master_df = pd.read_csv('twitter_archive_master.csv')
```

Is there a relationship between the tweet text lenght and the favourite count ?

```
In [21]: # setting the x and y axis values
         plot = sns.lmplot(x = 'text_letters_num', y = 'favorite_count', data = master_df )
         # add approptiate labels and title for the scatter plot
         plot = plot.set(ylabel= 'Favorite Count', xlabel='Lenght of a Tweet Text', title ='Scat
```



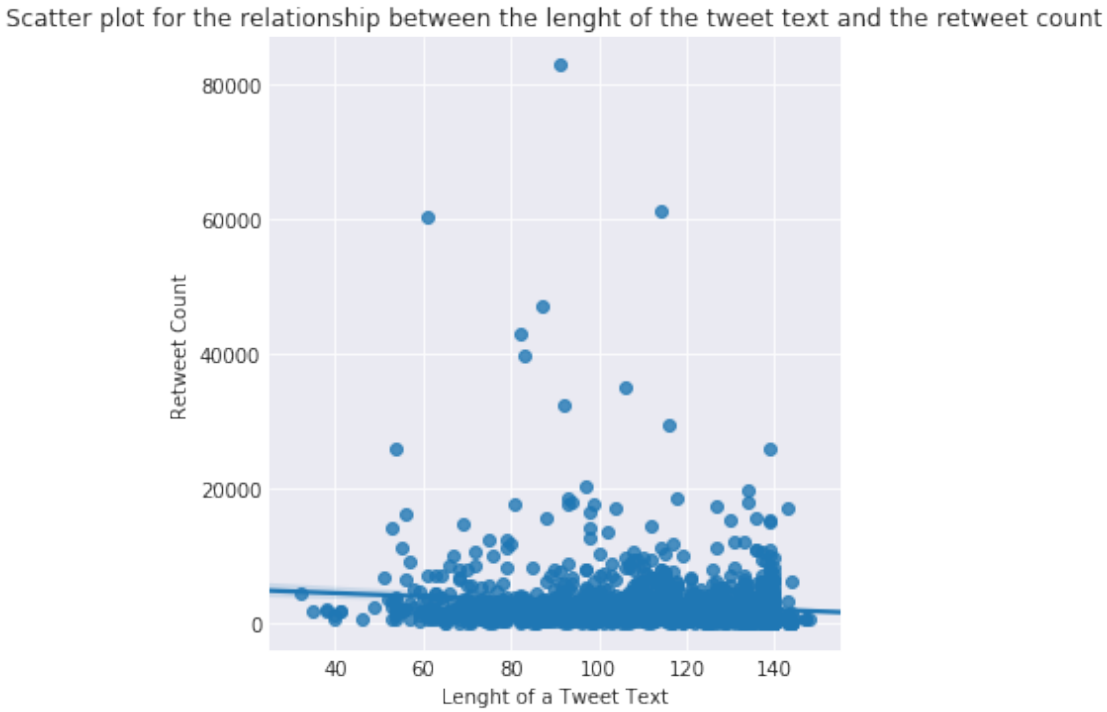

```
In [6]: # calculating the correlation coefficient between the lenght of a tweet text and the fav
        master_df['text_letters_num'].corr(master_df['favorite_count'])
```

```
Out[6]: -0.11083863911332734
```

Answer: I can observe that the relationship between the lenght of a tweet text and the favorite count of the tweet is almost doesn't exist. Therefore, I will test the retweet count of the tweet and check if there is a relationship with the number of letters in a tweet.

Is there a relationship between the tweet text lenght and the retweet count ?

```
In [22]: # setting the x and y axis values
         plot = sns.lmplot(x = 'text_letters_num', y = 'retweet_count', data = master_df )
         # add approptiate labels and title for the scatter plot
         plot = plot.set(ylabel= 'Retweet Count', xlabel='Lenght of a Tweet Text', title ='Scatt
```




```
In [23]: # calculating the correlation coefficient between the number of letters in a tweet and
         master_df['text_letters_num'].corr(master_df['retweet_count'])

Out[23]: -0.12833400504489664
```

Answer: I can notice that also there is almost no relationship between the number of letters and the retweet count.

Conclusion:

I came to the conclusion that the favourite count and retweet count have no relationship with the lenght of tweet text.

As a result, I suspect that the content of the tweet affects the engagement on the tweet.

What are the most popular dog names ?

```
In [10]: # finding the top four dog names that are popular
         master_df['dog_name'].value_counts().sort_values(ascending=False)

Out[10]: Lucy        10
         Cooper      10
         Oliver      10
         Charlie     10
         Penny        9
         Tucker       9
         Sadie        8
```

Winston	8
Lilly	7
Toby	7
Lola	7
Daisy	7
Stanley	6
Bella	6
Bo	6
Koda	6
Jax	6
Leo	5
Alfie	5
Chester	5
Dave	5
Ellie	5
Oscar	5
Scout	5
Buddy	5
Bailey	5
Rusty	5
Louis	5
Milo	5
Reggie	4
	..
Ron	1
Shooter	1
Hero	1
Grey	1
Maya	1
Karl	1
Michelangelope	1
Lizzie	1
Maddie	1
Arlo	1
Zeus	1
Cherokee	1
Heinrich	1
Bell	1
Franq	1
Alfredo	1
Rontu	1
Lugan	1
Snicku	1
Canela	1
Shiloh	1
Brudge	1
Brian	1
Kara	1

```
Alexander          1
Rufio              1
Levi               1
Darby              1
Wafer              1
Bruno              1
Name: dog_name, Length: 904, dtype: int64
```

Answer:
The top four popular dog names are : Oliver, Lucy, Charlie, and Cooper. I also noticed that generally the dog names are unique among the dogs that are in the data set.

What is the average dog rating for each dog stage ?

```
In [12]: # checking the average rating for each dog stage
         master_df.groupby('dog_stage')['rating_numerator'].mean()
```

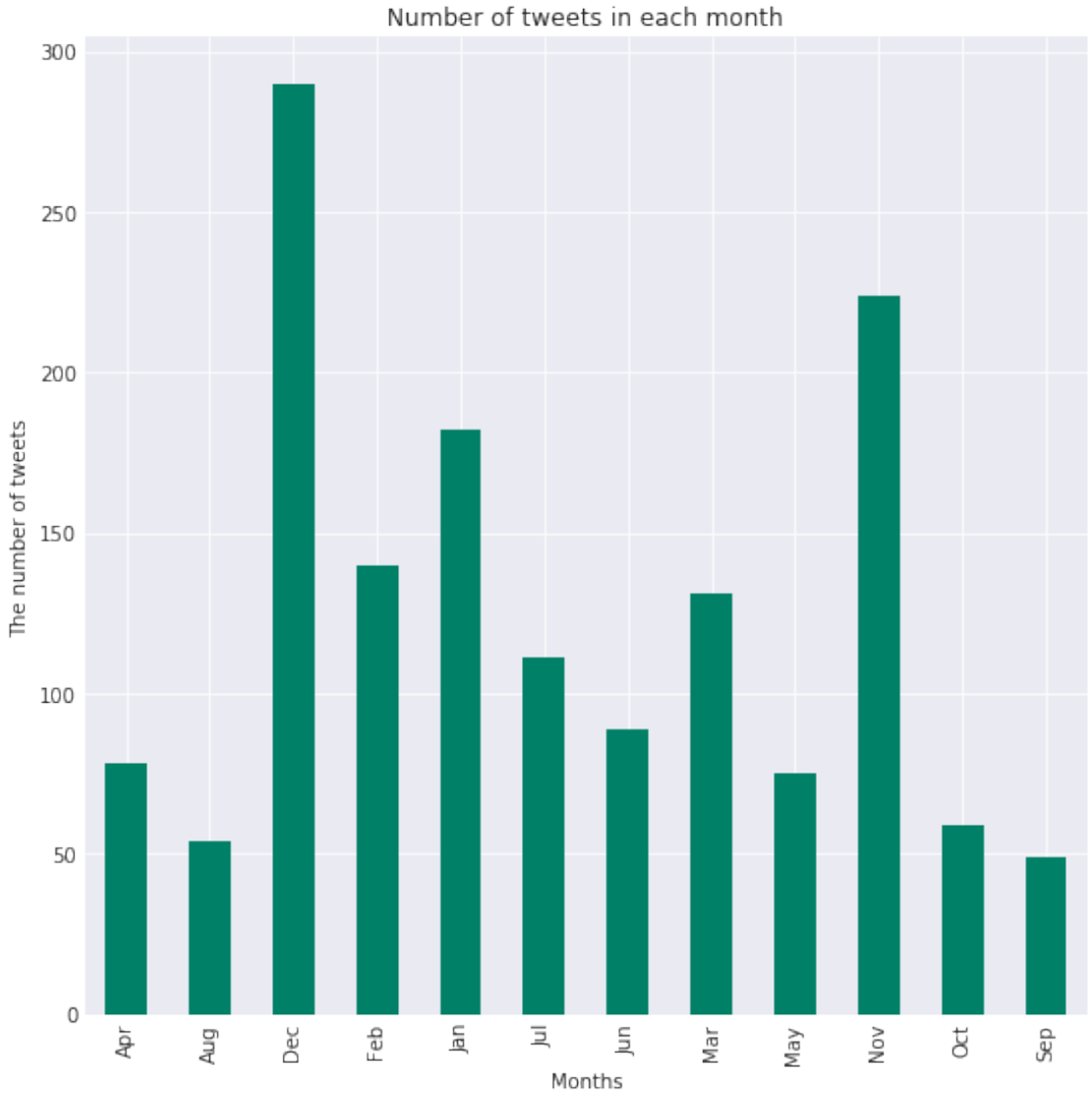
```
Out[12]: dog_stage
         doggo      11.888889
         floofer    11.875000
         pupper     10.634146
         puppo      12.043478
         Name: rating_numerator, dtype: float64
```

Answer: I can observe that the puppo stage has the highest average rating among other dog stages. And puppo has the lowest average rating among other dog stages.

In which months is the highest number of tweets are posted ?

```
In [14]: # checking the number of tweets in each month
         tweets_in_month = master_df.groupby('tweet_month')['tweet_id'].count()
         fig, ax = plt.subplots(figsize=(9,9))
         tweets_in_month.plot(ax=ax,kind="bar",colormap='summer')
         ax.set_xlabel('Months')
         ax.set_ylabel('The number of tweets')
         ax.set_title('Number of tweets in each month')
```

```
Out[14]: Text(0.5,1,'Number of tweets in each month')
```



Answer:

I can observe that the top three months that have the highest number of tweets are December, November, and January.

I suspect that this is related to the availability of the account owner. He might have vacation in this period of time.

What are the top dog names that got a high favourite count ?

```
In [15]: # finding the favorite count for each dog name
         master_df.groupby('dog_name')['favorite_count'].sum().sort_values(ascending=False)

Out[15]: dog_name
         Bo              177250
         Zoey            136506
```

Stephan	125767
Jamesy	120882
Lilly	119847
Lucy	105019
Charlie	103490
Duddles	103108
Penny	97909
Stanley	94652
Gary	93292
Earl	90359
Bella	88243
Phil	86906
Alfie	86809
Sunny	85161
Dexter	79576
Aja	77712
Kenneth	77478
Walter	76045
Winston	73791
Buddy	73756
Tucker	73673
Oliver	71807
Boomer	68536
Dave	68100
Canela	67134
Kevin	67104
Cooper	65924
Leo	63880
	...
Octaviath	379
Anthony	376
Saydee	370
Ronduh	369
Dook	366
Josep	346
Raphael	333
Cherokee	330
Sam	329
Clybe	329
Chuk	328
Lambeau	321
Timofy	320
Julio	304
Leroi	283
Kohl	254
Bradlay	254
Spark	252
Bloo	249

```
Marvin           246
Kollin           236
Jockson          236
Alphred          234
Lugan            231
Keet             216
Kallie           204
Klint            203
Fwed             193
Daryl            191
Christoper       185
Name: favorite_count, Length: 904, dtype: int64
```

Answer:
I can notice that the 10 top dog names that got the highest favourite count are :

- Bo - 177250
- Zoey - 136506
- Stephan - 125767
- Jamesy - 120882
- Lilly - 119847
- Lucy - 105019
- Charlie - 103490
- Duddles - 103108
- Penny - 97909
- Stanley - 94652

Note: Charlie, Lucy, Penny, Lilly, and Stanely were some of the names that are popular among the other dogs names.