

act_report

December 28, 2017

1 Data Analysis & Visualization

The first question I wanted to answer was which year had the most activity (tweets) for the weredogs twitter handle. I was able to answer this question using the following line of code using the pandas library:

```
tweets_master_df.groupby(tweets_master_df["timestamp"].apply(lambda x: x.year)) \
['timestamp'].count()
```

The results were that 2016 had the majority of tweets with 1021, while 2015 had 665 tweets and 2017 had 383 tweets. I was suprised that 2017 had such a low number. I know that I extracted tweets data up to August 2017 and as you'll see later in this notebook, the highest amount of activity for this twitter handle occured in November and December historically.

The second question I wanted to answer was which tweet was retweeted the most. To find this answer, I used the following line of code:

```
tweets_master_df.iloc[tweets_master_df['retweet_count'].nlargest(10).index, :]
```

This gave me the results for the top 10 tweets with the most retweets. Then I extracted out the text and link to the tweet using the following code:

```
tweets_master_df.ix[774, 'text']
tweets_master_df.ix[774, 'expanded_urls']
```

Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)
pic.twitter.com/7wE9LTEXC4

— WeRateDogs™ (@dog_rates) June 18, 2016

The third question I set out to answer as which tweets was favorited the most. To answer this question I used the following line of code:

```
tweets_master_df.iloc[tweets_master_df['favorite_count'].nlargest(10).index, :]
```

This gave me the top 10 tweets that were favorited the most and then I used the following code to extract out the text and link to the most favorited tweet:

```
tweets_master_df.ix[309, 'text']
tweets_master_df.ix[309, 'expanded_urls']
```



Most Retweeted Tweet



Most Favorited Tweet

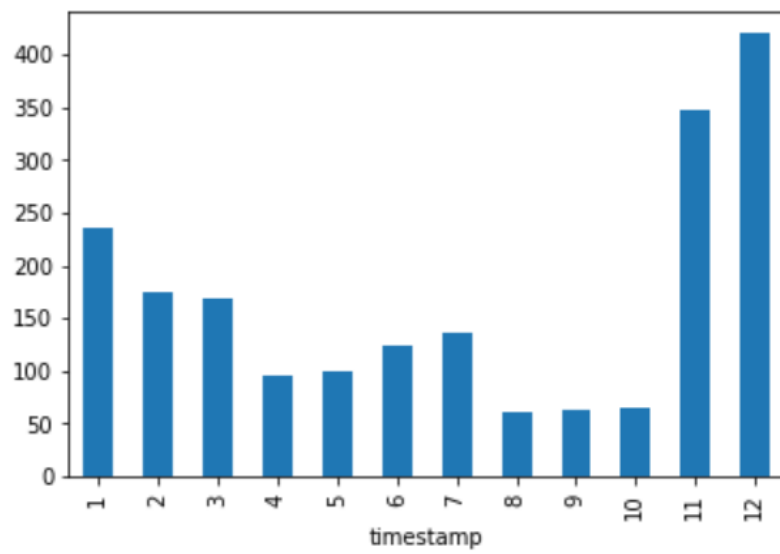
Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10
pic.twitter.com/nTz3FtorBc

— WeRateDogs™ (@dog_rates) January 21, 2017

For my visualization I used pandas to plot the tweets by month and find out which months throughout the year received the most activity:

```
tweets_master_df.groupby(tweets_master_df["timestamp"].apply(lambda x: x.month)) \
['timestamp'].count().plot(kind="bar")
```

This line of code illustrates the elegance and power of the pandas library. I had already converted the 'timestamp' column to type *datetime* so all I needed to do was group by each month and then create a bar chart that displays the count of the amount of tweets by each month. It turns out the most activity was in November and December but I'm not sure why.



bar chart