act_report

June 27, 2022

0.1 Report: act_report

- Create a **250-word-minimum written report** 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.

0.2 The data anslysis process

After getting the provided twitter_archive_enhanced.csv from Udacity which contains alot of information about the dog users which i downloaded it manually before reading it into a dataframe, i also downloaded the image_prediction.tsv programmatically using the link provided by Udacity and scrapped the twitter APi using the tweepy library from the code provided by Udacity. i assessed and cleanded the dataset then merged them to be ready for analysis.

Before analysing starts, i read the "twitter_archive_master.csv" file into a dataframe using the pandas .read_csv() function

```
In [1]: # importing libraries
        import pandas as pd
```

```
In [2]: #reading the twitter_archive_master.csv file using pd.read_csv() method
        data_clean = pd.read_csv("twitter_archive_master.csv")
        data_clean.head()
```

```
Out[2]:              tweet_id              source  \
        0  892420643555336193  Twitter for iPhone
        1  892177421306343426  Twitter for iPhone
        2  891815181378084864  Twitter for iPhone
        3  891689557279858688  Twitter for iPhone
        4  891327558926688256  Twitter for iPhone

                                                        text  \
        0  This is Phineas. He's a mystical boy. Only eve...
        1  This is Tilly. She's just checking pup on you...
        2  This is Archie. He is a rare Norwegian Pouncin...
        3  This is Darla. She commenced a snooze mid meal...
        4  This is Franklin. He would like you to stop ca...
```

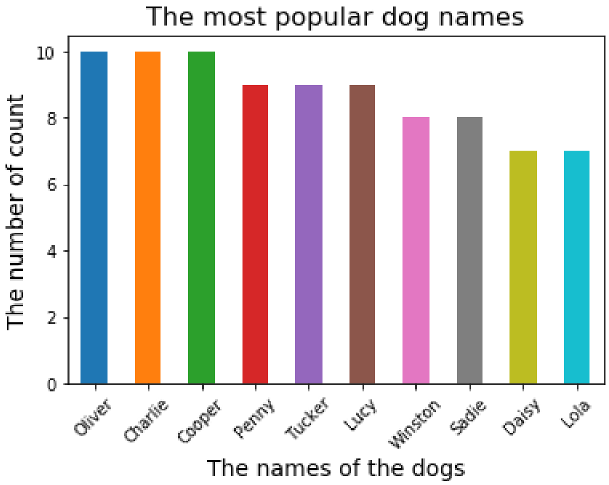
```
                                          expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                13   
1  https://twitter.com/dog_rates/status/892177421...                13   
2  https://twitter.com/dog_rates/status/891815181...                12   
3  https://twitter.com/dog_rates/status/891689557...                13   
4  https://twitter.com/dog_rates/status/891327558...                12   

   rating_denominator      name doggo floofer pupper      ...         p1_conf  \
0                  10   Phineas   NaN     NaN    NaN      ...        0.097049   
1                  10     Tilly   NaN     NaN    NaN      ...        0.323581   
2                  10    Archie   NaN     NaN    NaN      ...        0.716012   
3                  10     Darla   NaN     NaN    NaN      ...        0.170278   
4                  10  Franklin   NaN     NaN    NaN      ...        0.555712   

   p1_dog                  p2   p2_conf p2_dog                           p3  \
0   False               bagel  0.085851  False                       banana   
1    True            Pekinese  0.090647   True                     papillon   
2    True            malamute  0.078253   True                       kelpie   
3   False  Labrador_retriever  0.168086   True                      spatula   
4    True    English_springer  0.225770   True  German_short-haired_pointer   

    p3_conf  p3_dog  retweet_count favorite_count  
0  0.076110   False           7009          33819  
1  0.068957    True           5302          29336  
2  0.031379    True           3481          22060  
3  0.040836   False           7227          36945  
4  0.175219    True           7763          35314  

[5 rows x 27 columns]
```



0.2.1 Insights:

1. The names Charlie, Cooper and Oliver were more popular among dog users

2. The source Twitter for iphone was found more among dog users

3. The year 2016 had about 1206 of images posted

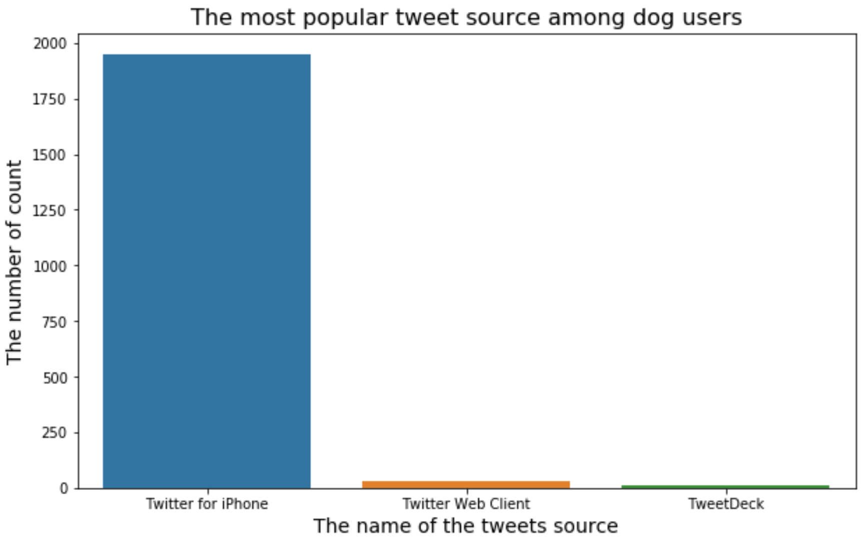
0.3 Visualization

```
In [3]: #getting the top 10 dog names used among dog users, using the value_counts() method
        df_name = data_clean['name'].value_counts().head(10)
        df_name
```

```
Out[3]: Charlie    10
        Cooper     10
        Oliver     10
        Penny       9
        Tucker      9
        Lucy        9
        Sadie       8
        Winston     8
        Lola        7
        Daisy       7
        Name: name, dtype: int64
```

1. These shows that the names Charlie, Cooper and Olive were more popular with a count of 10 among the dog users

2. These shows that twitter for iphone was more used to tweet among the dog users



```
In [4]: #getting the total count for each tweet source and putting into a data frame
        df_source =data_clean['source'].value_counts().to_frame().reset_index()
        df_source.rename({'index': 'source', 'source': 'count'}, axis = 1,  inplace = True)
        df_source
```

```
Out[4]:               source  count
        0  Twitter for iPhone   1948
        1  Twitter Web Client     28
        2           TweetDeck     10
```

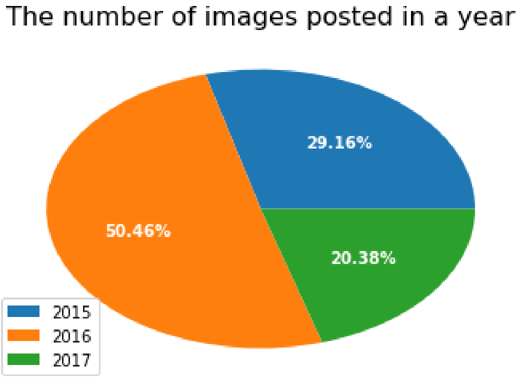
3. These shows that the year 2016 had more images posted with a percentage of 50.48%

```
In [5]: #getting the total count for each image_number with respect to the years and putting int
        df_year = data_clean.groupby(['year'])['image_number'].sum().to_frame().reset_index()
        df_year
```

```
Out[5]:    year  image_number
        0  2015           697
        1  2016          1206
        2  2017           487
```

1 that summarizes the analysis process

```
In [ ]:
```


The number of images posted in a year
29.16%
50.46%
20.38%
2015
2016
2017