Project 2

Act Report

ALX Nanodegree Data Analyst Program

Elaborated by: Oumayma Jribi

Table of content

1. Introduction:

WeRateDogs is a [Twitter](#) account that rates people's [dogs](#) with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention both for its popularity and for the attention drawn to [social media](#) [copyright law](#) when it was suspended by Twitter for breaking these aforementioned laws.

During this project, I gathered data about the WeRateDogs tweets from different sources using different methods, then assessed and cleaned it to get a final clean dataset containing different information about these tweets.
This dataset was then analyzed in order to derive insights about this account and its tweets.

So how are dog stages distributed? What are the top 20 dog breeds? How does the distribution of retweet counts and favorite counts look like?
These questions are more will be answered all along the insights and visualizations steps.

2. Insights:

- Distribution of dog stages:

```
Entrée [80]: dog_archive_clean['dog_stage'].value_counts()

Out[80]: pupper     222
         doggo       72
         puppo       23
         floofer      9
         Name: dog_stage, dtype: int64
```

⇒ The most dominant dog stage is "pupper"

- Top 20 dog breeds:

```
Entrée [81]: dog_archive_clean['dog_breed'].value_counts()[0:20]

Out[81]: Golden Retriever          158
         Labrador Retriever        108
         Pembroke                   95
         Chihuahua                  91
         Pug                        62
         Toy Poodle                 51
         Chow                       48
         Samoyed                    42
         Pomeranian                 42
         Malamute                   33
         Chesapeake Bay Retriever   31
         French Bulldog             31
         Cocker Spaniel             30
         Miniature Pinscher         25
         Eskimo Dog                 22
         Cardigan                   21
         German Shepherd            21
         Staffordshire Bullterrier  21
         Beagle                     20
         Shih-Tzu                   20
         Name: dog_breed, dtype: int64
```
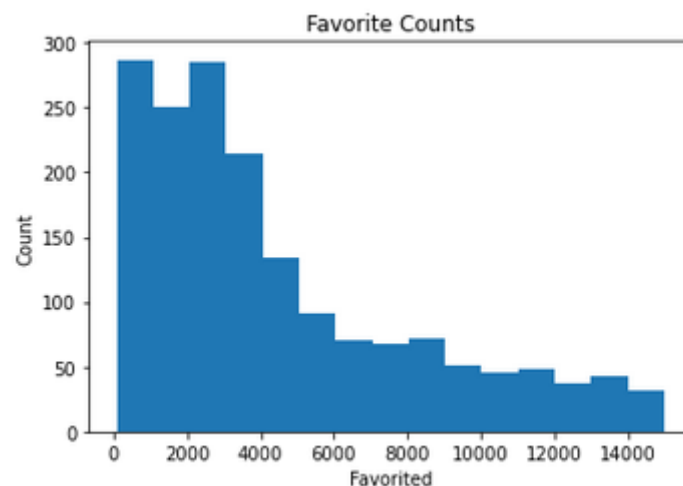
- Distribution of the 'source' of Twitter usage:

```
Entrée [82]: dog_archive_clean['source'].value_counts()

   Out[82]: Twitter-iPhone    1985
            Vine                91
            Twitter             30
            Tweetdeck           11
            Name: source, dtype: int64
```

⇒ Twitter for iPhone is the dominating source for Twitter usage
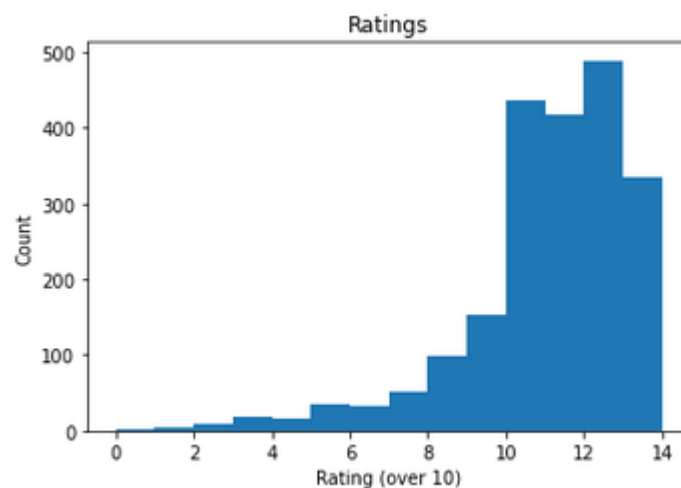
3. Visualizations:

- Plot favorite_count and retweet_count to get insights about the popularity of WeRateDogs

⇒ More than 275 Tweets were favorited between 0 and 1000 times
⇒ Almost 100 Tweets were favorited between 5000 and 6000 times

- Plot rating numerators to get insights about how dogs are rated

⇒ Almost 500 dogs have a rating between 12 and 13 over 10

4. Challenges and Limitations:

The major limitation I encountered during the insights and visualizations step is not being able to use all the variables for analysis mainly due to missing values.