5/28/2021

# Act Report

*Analysis of Tweeter user
"WeRateDogs"*

Andreja Ho
UDACITY'S PROJECT #4: WRANGLE AND ANALYZE DATA

**Project Overview**

Real-word data rarely comes clean, especially if data comes from variety of sources and formats. For this project I am analyzing tweet archive of Twitter user @dog_rates, also known as "WeRateDogs". WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. For this project I've used Python and its libraries in order to retrieve, clean and visualize data and draw meaningful conclusions.

This fun, yet comprehensive project required the following steps:

1. Data wrangling
    1.1. Gathering data
        ▪ **The WeRateDogs Twitter archive** dataset. File on hand and it is stored as a ***CSV file***.
        ▪ **The tweet image predictions** dataset was retrieved from Udacity's servers through the Python ***requests*** library. The file was stored in the ***TSV file***
        ▪ **Tweeter's retweet and favorite count** data was retrieved from Tweeter's website through ***API call***, using the ***Tweepy*** library. Retrieved data was stored in ***JSON format.***

    1.2. Assessing data
    After all data was gathered, I assessed it visually and programmatically for quality and tidiness issues. Complete documentation about assessing data is stored in "wrangle-repodt.pdf" file.

    1.3. Cleaning data
    Each of documented issue was cleaned using Python and its library. The end result was master pandas DataFrame.

2. Storing data
   Cleaned master DataFrame was exported and stored as csv file under the name: "twitter_archive_master.csv"

3. Analyzing & Visualizing
   The end results for this project are saved in Jupyter Notebook, *wrangle_act.ipynb:* complete code for gathering, assessing, cleaning and visualizing

4. Reporting Data
   The end result and complete reports are saved in
       - PDF file wrangle-repodt.pdf: complete documentation about data wrangling process
       - PDF file act_report.pdf (this file): complete documentation and visualization created for this project.

## Visualizations and Conclusions

For this project I have created several visualizations and draw conclusions about tweet archive of Twitter user @dog_rates, also known as "WeRateDogs".

**The most popular dog names**

From the Figure 1 & 2 we can see what are the most popular dog names on Tweeter's group "WeRateDogs". There is approximately 1390 data points. Dog names were grouped by name and its corresponding count was created with Python code. From the bar chart and table below, we can read that the most popular dog names are Charlie and Lucy (appeared 11 times), followed by Cooper and Oliver (appeared 10 times). This data can be used to give ideas to future dog owners to choose amongst popular or unique name for their dog.
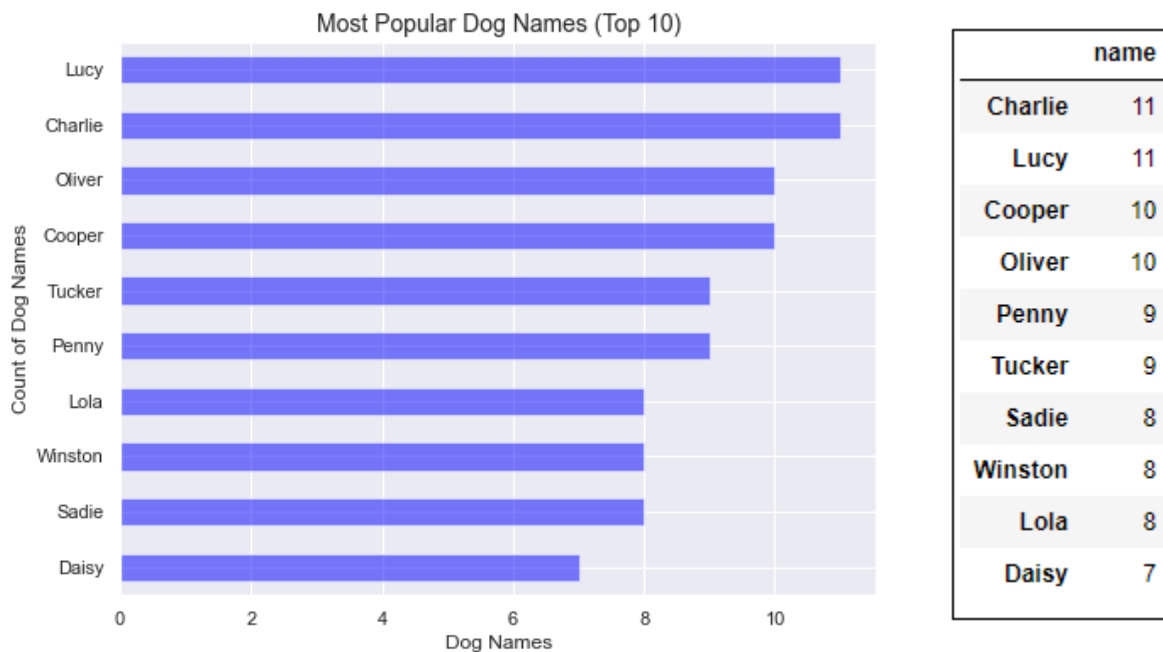


| | name |
|---|---|
| Charlie | 11 |
| Lucy | 11 |
| Cooper | 10 |
| Oliver | 10 |
| Penny | 9 |
| Tucker | 9 |
| Sadie | 8 |
| Winston | 8 |
| Lola | 8 |
| Daisy | 7 |

*Figure 1 & 2: Top 10 most popular dog names on "We Rate Dogs"*

Since there is about 1390 data points and most popular dog names doesn't exceed the count of 11, it seems that there is a lot of unique dog names. For better visualization I have created a pie chart that represent a part of popular dog names versus the group "Other" where all other dog names are stored.
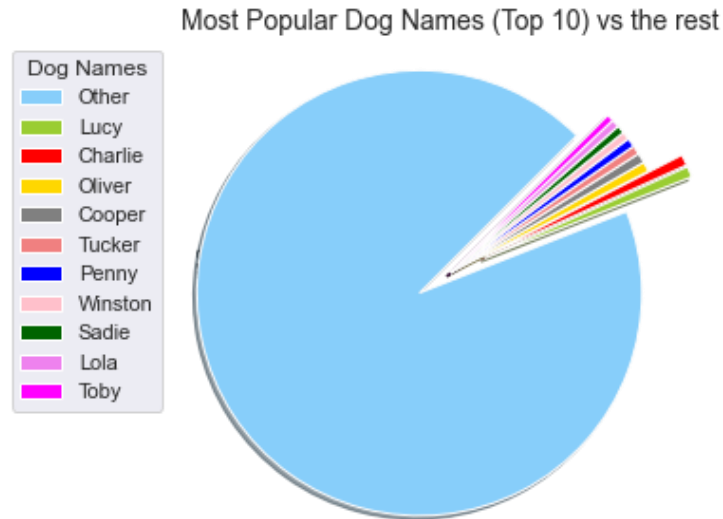
*Figure 3: Pie chart – Most popular dog names vs the rest.*

From the pie chart we can see that top 10 dog names represent just a small fraction of a whole. We can learn that dogs have quite unique names.

**Most retweeted dog breeds**

From the bar chart below, we can see what are the most popular dog breeds based on retweet count. The most retweeted dog breeds are Golden retriever, followed by Labrador retriever and Pembroke.
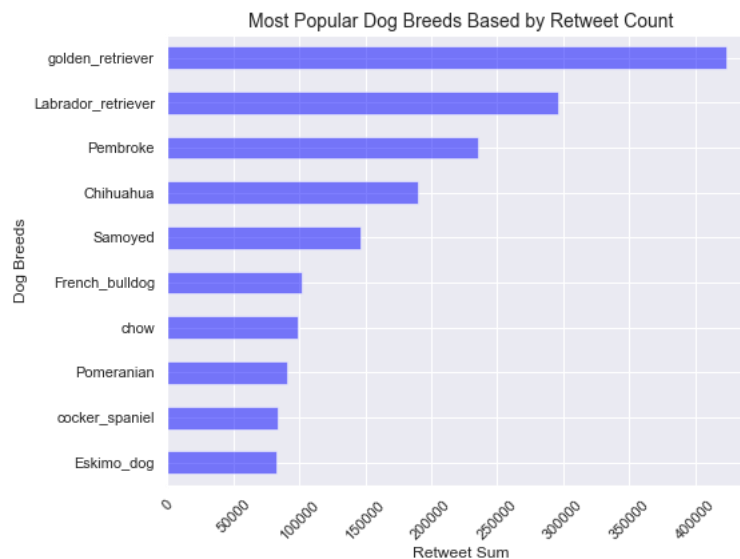


*Figure 4: Most popular dog breeds based on retweet count*

**Tweets Over Time**

From the line chart we can learn more about the tweets frequency over time. Tweets "WeRateDogs" had high traffic in the end of 2015. In 2016 the tweets started sharply declining and remained steadily low from May 2016 further. From the chart we can learn that there was a huge interest at the beginning and only few loyal users stayed active after the hype was gone.
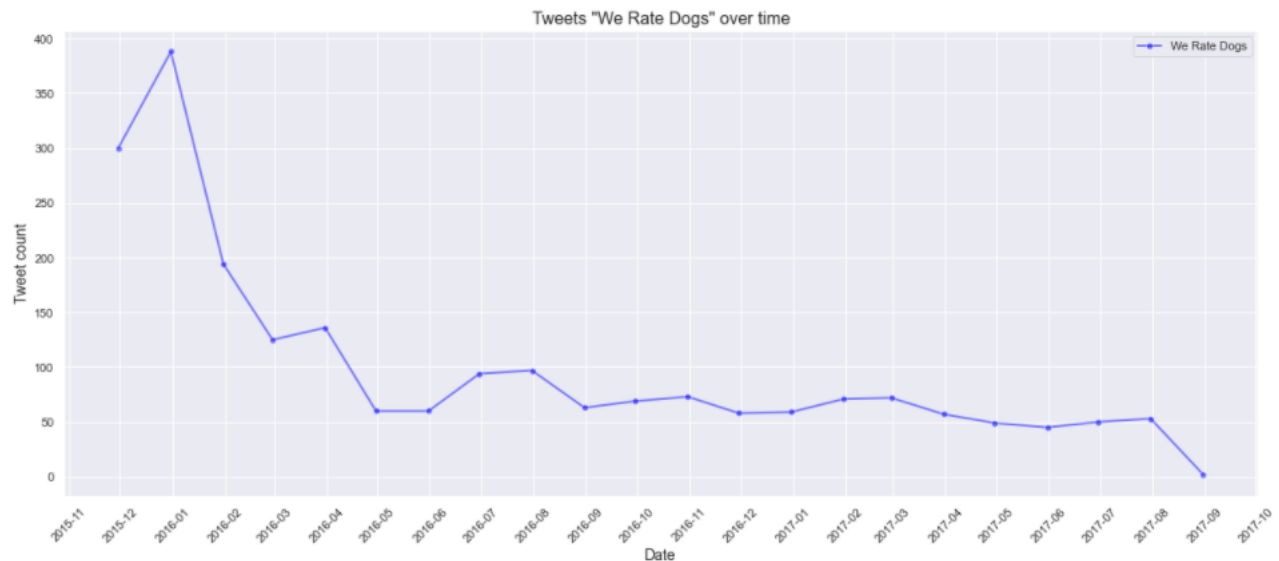


*Figure 5: Tweets Over Time*

**The 3 most rated tweets**

Since this project was all about dogs and I choose to use Python code to extract 3 most popular dog tweets, their text what owners have to say about them, their favorite and retweet count.

**The winner is:**

**Second place:**

Here's a super supportive puppo participating in the Toronto #WomensMarch today. My picture earned 127,683 likes and was retweeted 41,455-times!🏳️‍🌈

**Third place:**

This is Stephan. He just wants to help. such a good boy . My picture earned 115,251 likes and was retweeted 53,629-times!🧹