WeRateDogs

ANALYSIS AND INSIGHTS REPORT

By Paula Munoz | Udemy Project: Wrangle and Analyze data| 02/27/19

Introduction

Throughout this project I have gathered, assessed and cleaned different data sets related to the Twitter account @dog_rates, also knowns as WeRateDogs. The purpose of this project was to create an interesting and trustworthy analysis and visualizations.

ABOUT WERATEDOGS

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog, and these ratings almost always have a denominator of 10. and the numerator is almost always greater than 10. Some examples of ratings are: 11/10, 12/10, 13/10, etc.

# Analysis and Insights

WHAT IS THE MOST USED TWITTER SOURCE?

Based on the input data, there are three different Twitter sources:

- Twitter by iPhone
- Twitter Web Client
- TwitterDeck

And based on the results Twitter for iPhone has the most twitter counts which accounts for **98.13** % of the data.

Twitter Source counts

```
Twitter for iPhone    1572
Twitter Web Client      22
TweetDeck                8
```


Most used Twitter Source

Twitter for iPhone
Twitter Web Client
TweetDeck

98.13%
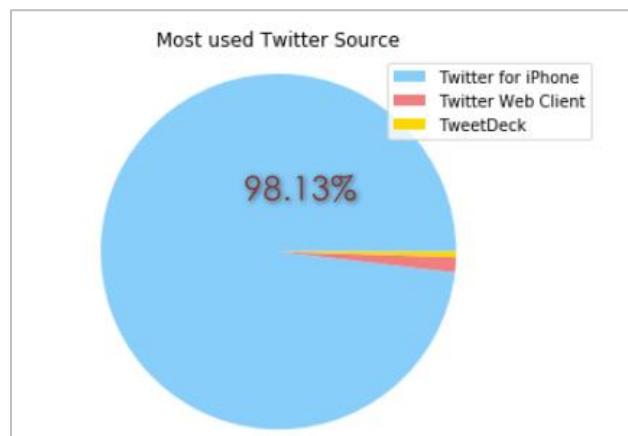
FINDIG THE TOP TEN DOG BREEDS

Based on the input data, there is a great variety of Dog Breeds mentioned in the tweets, however **Golden Retriever** is leading the tweet counts, followed by **Labrador Retriever.**

Top 10 Dog Breeds – Tweet counts

```
golden_retriever      153
Labrador_retriever    102
Pembroke               91
Chihuahua              89
pug                    58
toy_poodle             51
chow                   44
Pomeranian             41
Samoyed                37
malamute               32
```
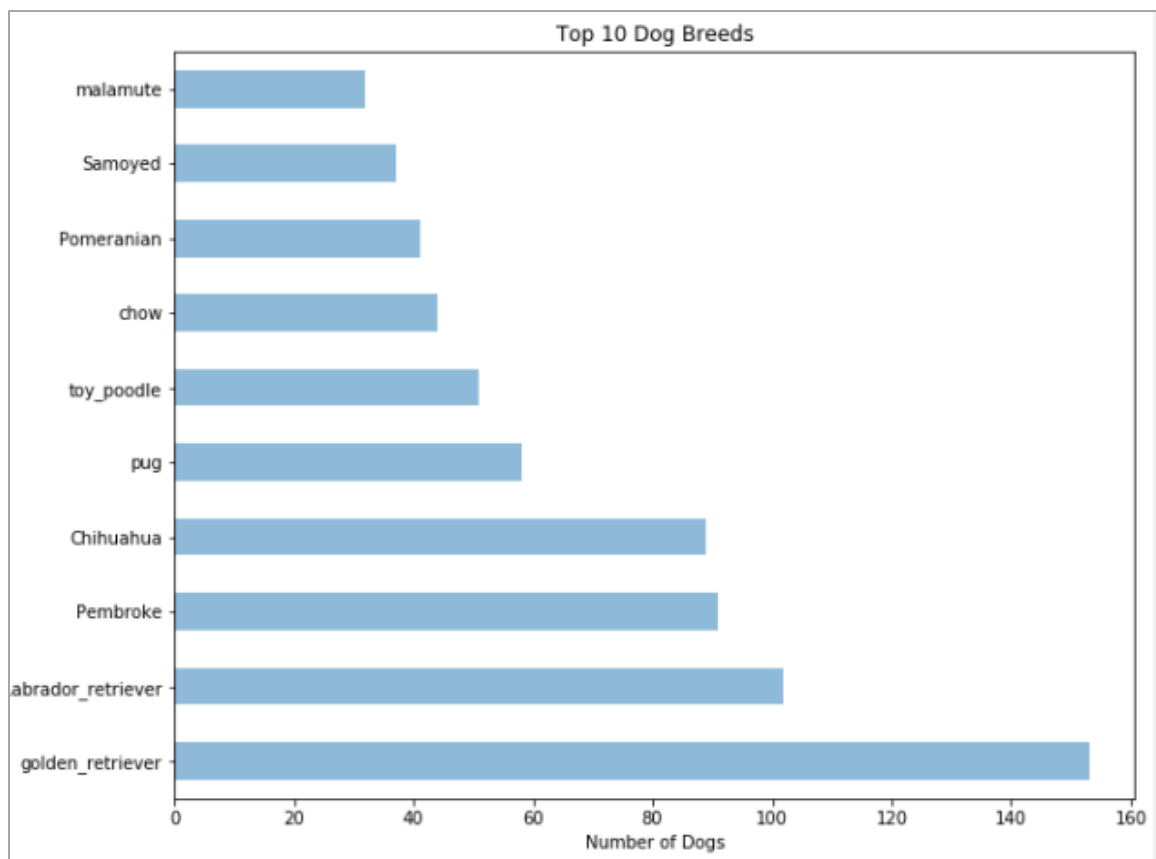
RETWEETS AND FAVORITE TWEETS OVER TIME

Two of the interesting variables in the dataset are retweet_count as well as favorite_counts.

We can clearly see from visualization below how the **favorite counts grow almost exponentially over time**, compared to retweet counts that even though it still shows some increment, it is not as high as favorite counts.

Retweet count Statistics

```
count      1602
mean       2630
std        4856
min          11
25%         592
50%        1305
75%        2954
max       83264
```

Favorite count Statistics

```
count      1602.
mean       8783.
std       13072.
min          78.
25%        1991.
50%        4004.
75%       10862.
max      163731.
```


Retweets and Favorite tweets over time
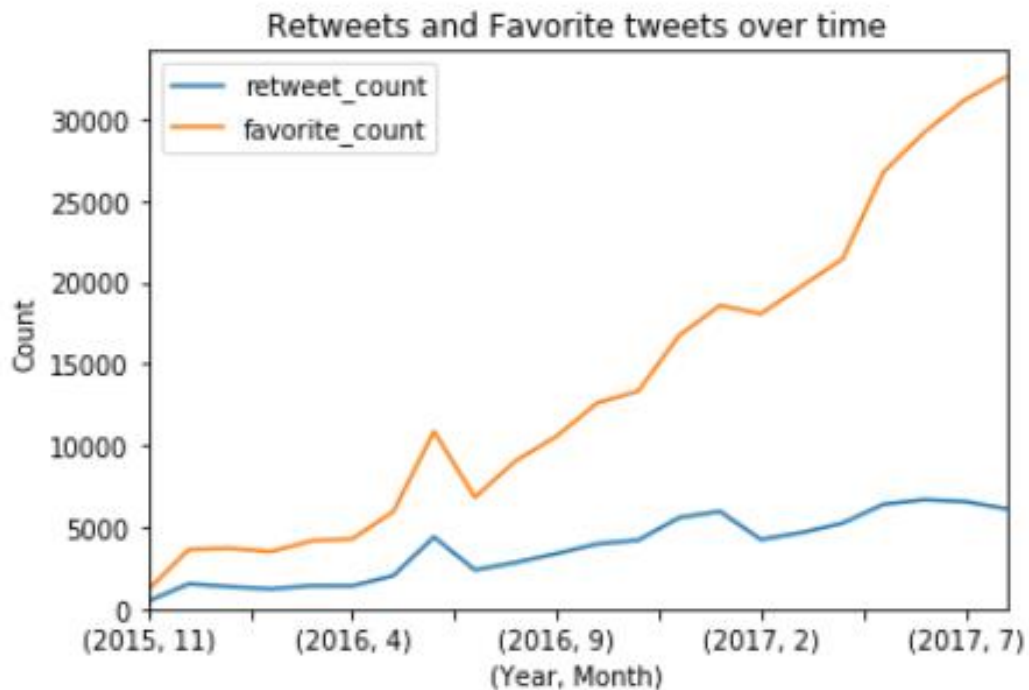
ABOUT RATING NUMERATOR

After performing data cleaning to the dataset and understanding the context of the rating, I have normalized the rating_numerator variable to only show values between 4 to 16.

Per Histogram visualization we can see the rating_numerator distribution is a little bit Negative Skewed, where the great majority of counts of rating numerators are above then, which make sense based on the way the Dogs are rated.

Rating Numerator Statistics

```
count    1602.00
mean       10.86
std         1.73
min         4.00
25%        10.00
50%        11.00
75%        12.00
max        16.00
```


Rating Numerator Distribution