

Data Analysis and Visualization Report

Prepared by: Ahmed Mohamed Hussein Unshur

The following report communicates insights and presents visualizations produced from a wrangling and analysis project completed as part of [Udacity's Data Analyst Nanodegree](#) program.

The dataset wrangled, analyzed, and visualized is the tweet archive of the Twitter user [@dog_rates](#), also known as [WeRateDog](#), which rates people's dogs with a humorous comment about the dog.

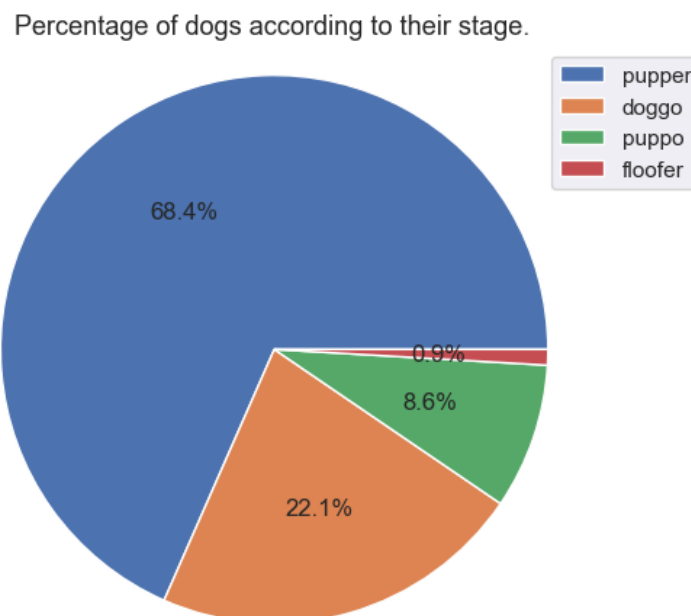
The project was completed using Anaconda, Python and some of its packages and libraries for data gathering, analysis and visualization (NumPy, Pandas, Matplotlib, Seaborn, Requests, Tweepy, and JSON), Jupyter Notebook, Sublime Text, and Microsoft Word.

The wrangled dataset was stored in a CSV file named: twitter_archive_master.csv. Then, we analyzed the dataset and produced insights about the following:

- I. The percentage of dogs in each stage.
- II. The tweet with the highest retweet count and the tweet with the lowest retweet count.
- III. The relationship between retweet and favorite counts.

I. The percentage of dogs in each stage.

We have visualized percentage of dogs in each stage and the following pie chart represents the findings:



We can conclude from the graph that:

1. Pupper has the highest percentage (68.4%) compared to other dog stages.
2. Floofer has the lowest percentage (0.9%) compared to other dog stages.

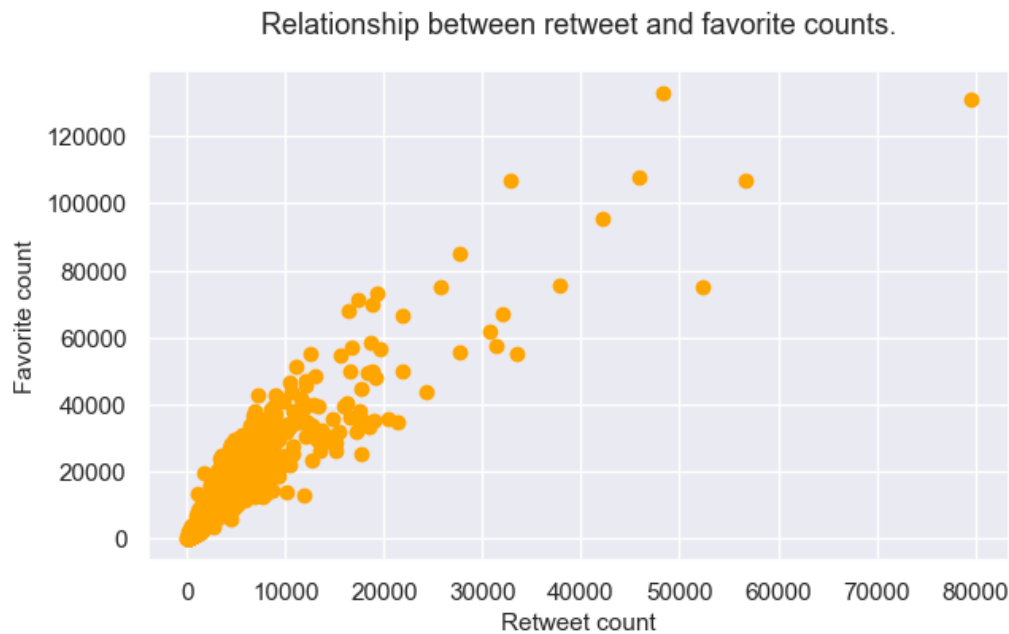
II. The tweet with the highest retweet count and the tweet with the lowest retweet count.

Analysis revealed that:

1. Highest retweeted tweet has the tweet id: 744234799360020481 and a retweet count of 79515.
2. Lowest retweeted tweet has the tweet id: 666102155909144576 and a retweet count of 16.

III. The relationship between retweet and favorite counts.

To find the relationship between retweet and favorite counts, we have produced the following graph.



The scatter plot above shows a positive correlation between retweet counts and favorite counts.

One important thing to note is that we cannot make causal inference from this finding. To do that, further investigation is required.