

# Data Wrangling Report

The WeRateDogs is a known twitter account that makes posts about dogs and rates them. It has about 9.3 million followers as of November 2022. This data comprises of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present. WeRateDogs has its own categories of dogs based on its age and appearance. The necessary file was provided by Udacity for the purpose of this study. After the wrangling process, I came out with the following analysis:

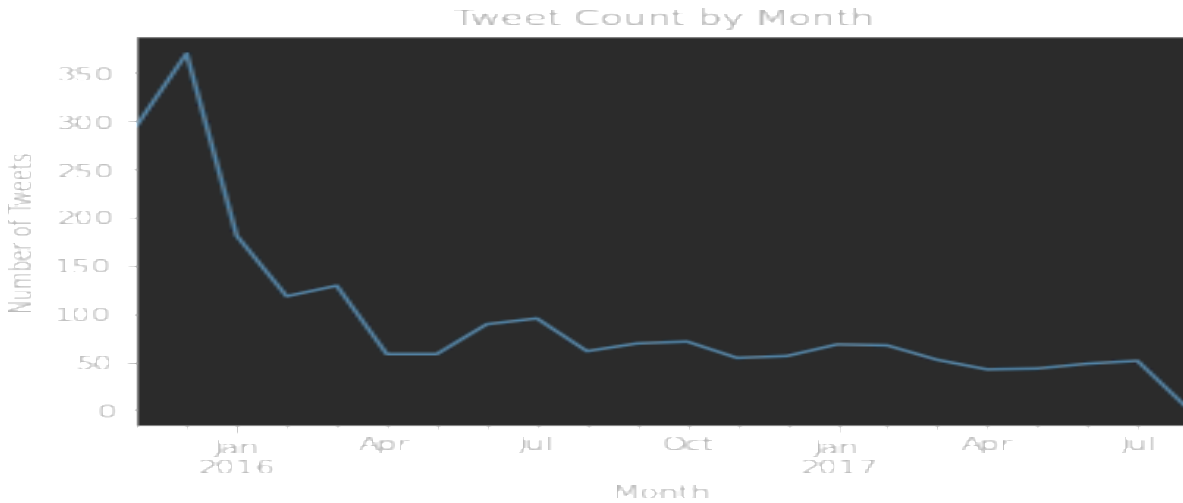
- \* The monthly trend of activities
- \* Posting trend
- \* Devices used by most followers of the page who interact with posts.
- \* Most common dog stage

**Data Gathering:** I downloaded the twitter\_archive.csv file from the link provided by Udacity. I also downloaded the .tsv file of image prediction and loaded it into a dataframe called imgpred. To get the tweet ids and counts, I downloaded the JSON file having this information via the twitter API.

**Data Assessing and Cleaning:** I was able to identify some data quality issues and fixed them in the main notebook which were clearly commented for readability purposes. I also made sure null values, invalid datatypes were fixed and removed unwanted columns for the analysis. I then merged the 3 files into one master data which has been provided as : `twitter_archive_master.csv`

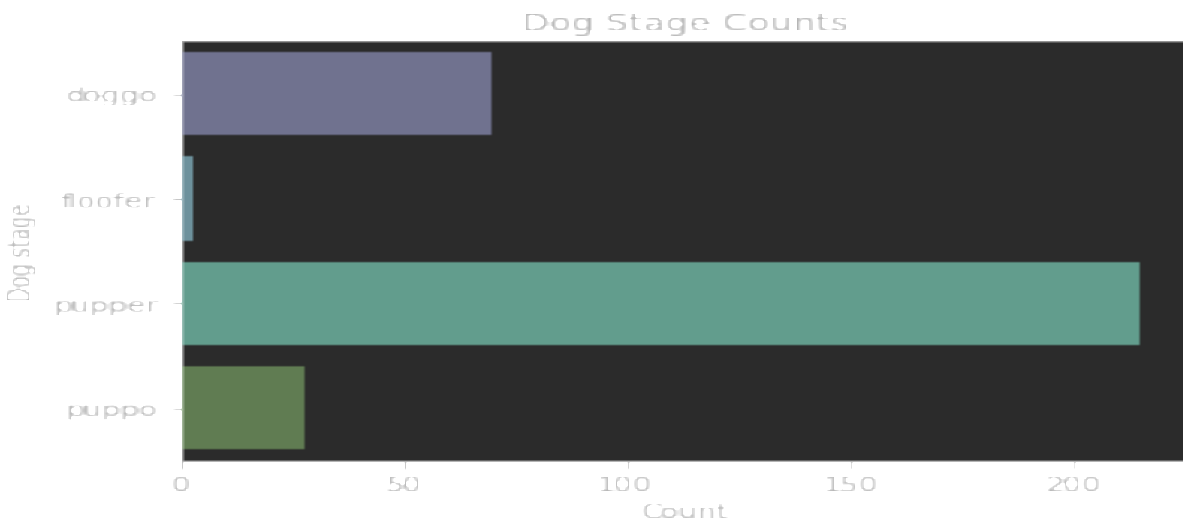
Below are some insights I gathered:

**Post trend:**



The number of posts on WeRateDogs has had a general downward trend by month since the account was first started, going from a peak of almost 350 tweets in one month down to around 50 in the latest months. There was a sharp decline only a few months after the account was created. However, this does not necessarily mean that the number of overall posts by the account has decreased, as my analysis only focused on original content with images. It is possible that the account has focused on other areas of content, such as retweets, videos, and interactions with followers.

#### Most common dog stage:



Clearly it is indicated that the most popular dog stage from the data is pupper as it has the highest count of 350 followed by doggo. The least popular dog stage was floofer.

