

# **Udacity Project 5**

## **Act Report**

### **Wrangle and Analyze Data**

#### **Objective**

The goal of the project: Gathering and combining multiple datasets of WeRateDogs from Twitter to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering using the Twitter API, then programmatically assessing and cleaning was required for great analyses and visualizations.

#### **Introduction**

WeRateDogs is a Twitter account that rates dogs picture with a nice comment twitted by users. It was started in 2015 by Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter. As of October 2017, the Twitter account has close to four million followers and Nelson sees 30,000 likes on a post as being viral. His most popular post was of a dog marching in the 2017 Women's March, which was retweeted more than 50,000 times and favouredited 134,000 times.

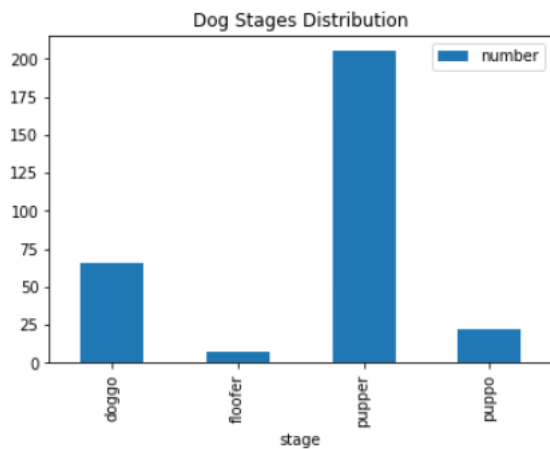
Tweepy package and tweeter API was used to collect the current status (retweet and favorite status) for the provided tweet IDs. At this stage, it is important to acknowledge that I gather the needed keys from the udacity students hub to use tweeter APIs. A software was used to predict the dog breed in the tweets. The predictions were supplied in a separate data file.

#### **Basic Statistics**

In this project we do data wrangling efforts in which, data were gathered, assessed, cleaned and finally combined in one master document to be used in data analysis and visualization.

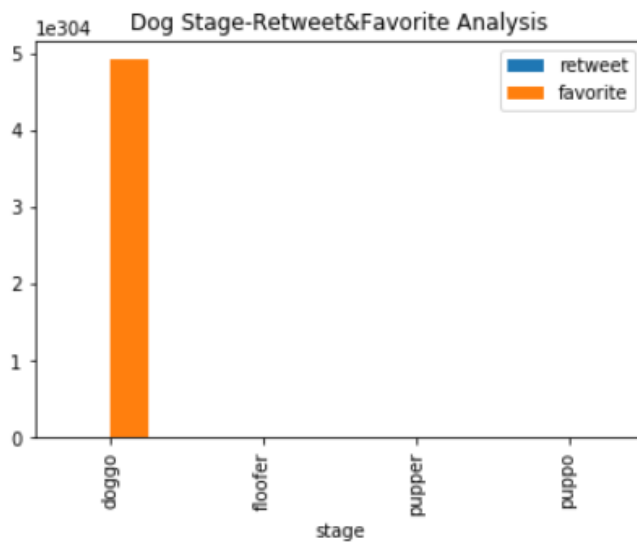
## Visualizations

**Chart 1: Dog Stages Distribution**



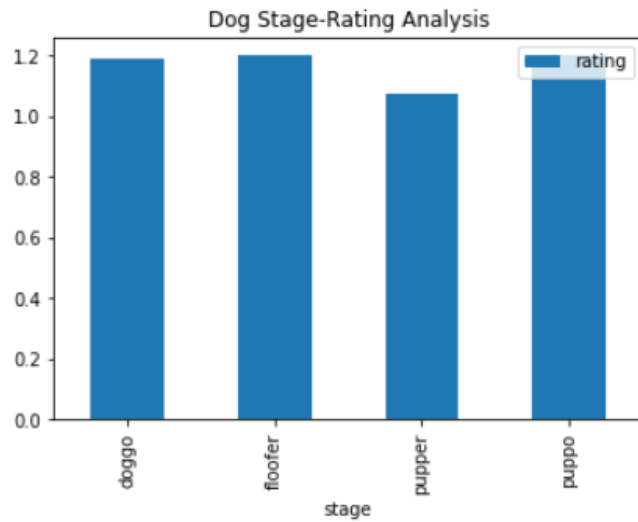
This chart shows the distribution of Dog Stages. Pupper and doggo are the most common.

**Chart 2: Dog Stage-Retweet & Favorite Analysis**



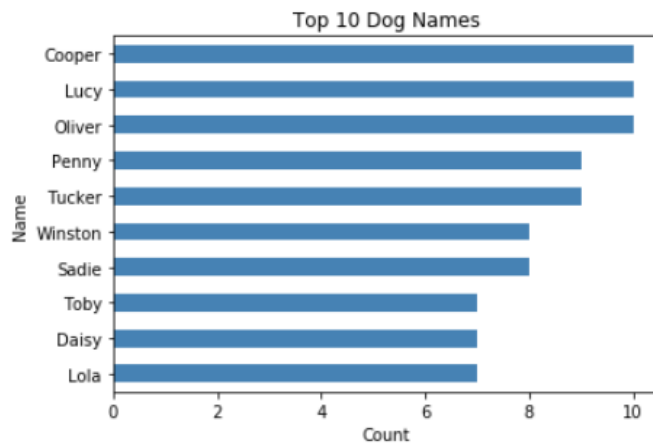
This chart shows the relationship between dog stage and retweet, favorite counts.

**Chart 3: Dog Stage-Rating Analysis**



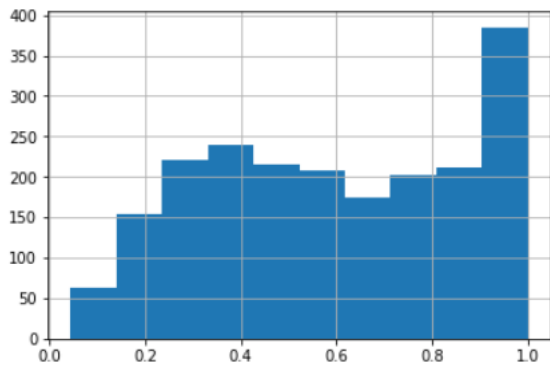
This chart shows the relationship between dog stage and rating.

**Chart 4: Top 10 Dog names**



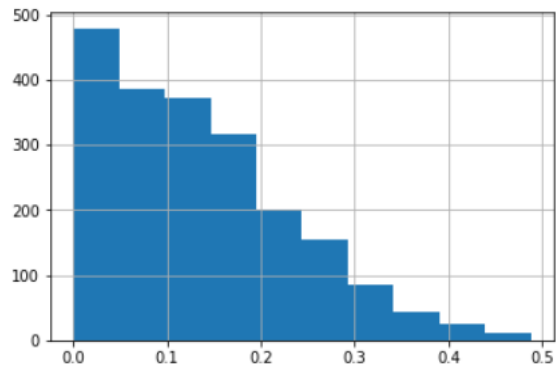
Cooper, Lucy, Oliver, Penny, Tucker, Winston are the most common dog names.

**Chart 5: p1\_conf**



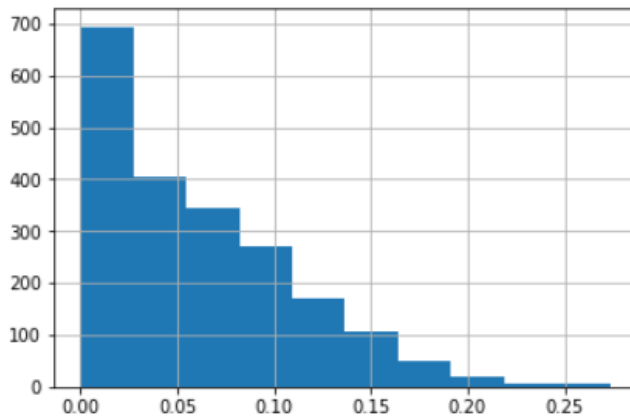
This chart shows the number of average ratings for p1\_conf, in which nearly 380 entries having average rating of 1.

**Chart 6: p2\_conf**



This chart shows the number of average ratings for p2\_conf, in which maximum ratings(nearly 480) are close or equal to zero.

**Chart 7: p3\_conf**



This chart shows the number of average ratings for p3\_conf, in which maximum ratings(nearly 700) are close or equal to zero.