

# Wrangle and Analyze Data

## Act Report

### Introduction:

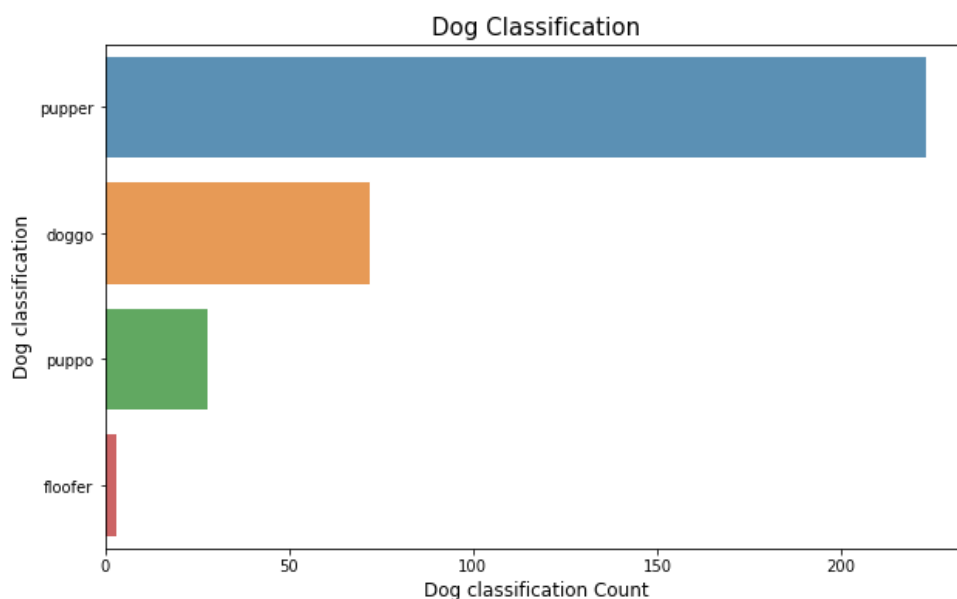
The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog\_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc.

The WeRateDogs Twitter archive contains basic tweet data for all 5000+ of their tweets, but not everything. One column the archive does contain though: each tweet's text, which I used to extract rating, dog name, and dog "stage" (i.e. doggo, floofer, pupper, and puppo)

### Analyzing and Visualizing Data

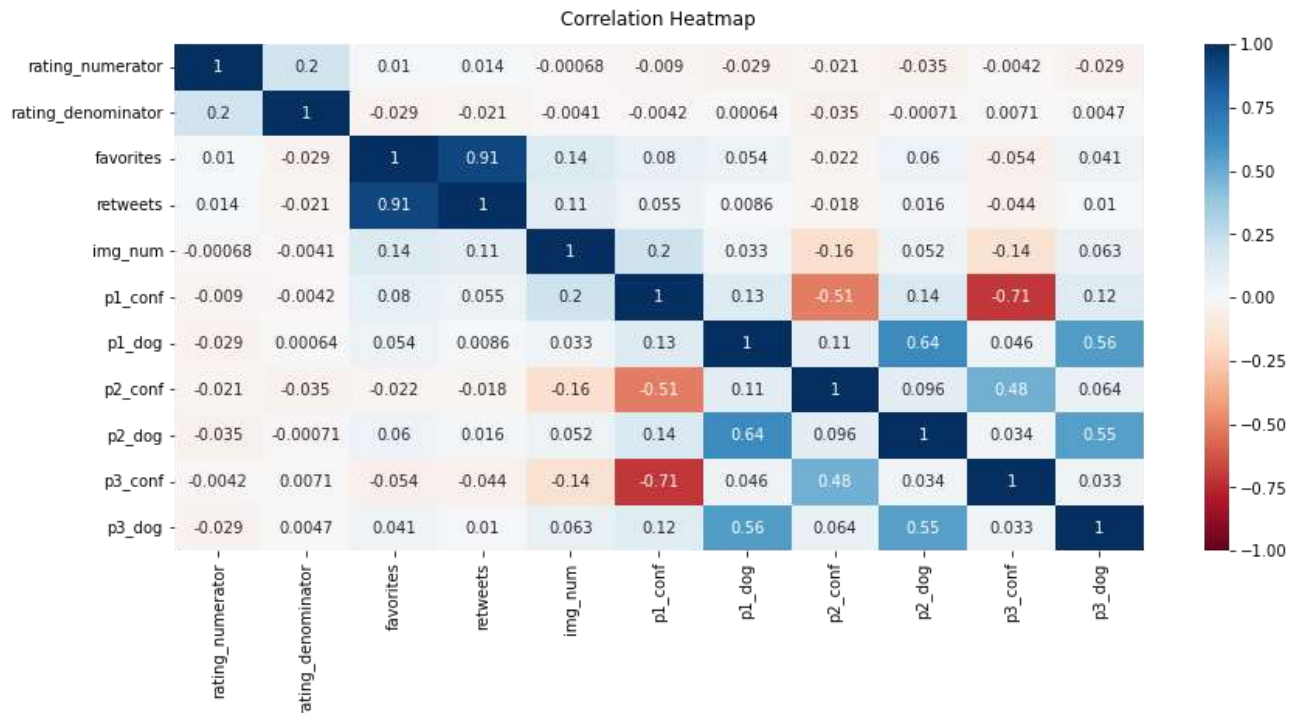
I have explored the dataset and came up with questions visualizations using (matplotlib and seaborn):

**Question 1: How many dogs we have for each category?**



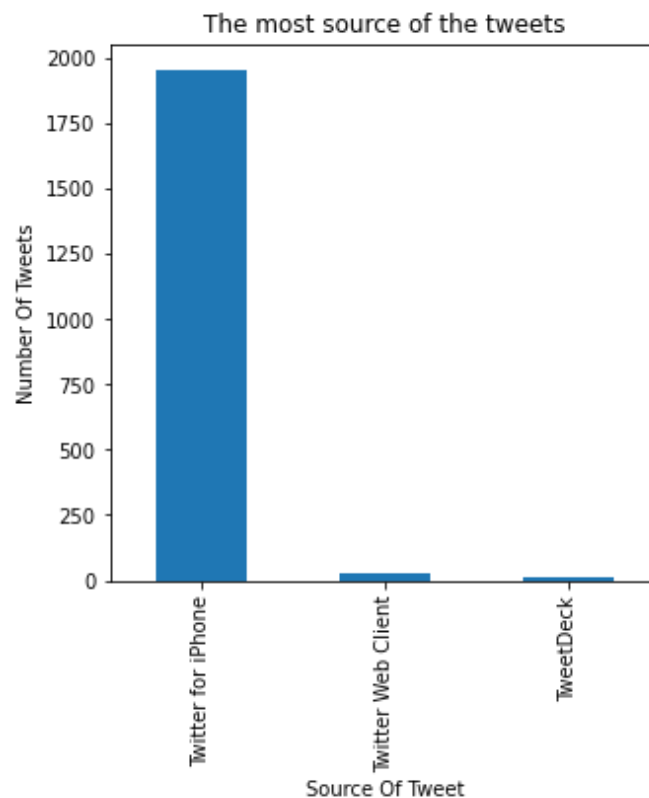
**Pupper is the most common dog category, followed by Doggo. Floofer is very rare.**

## Question 2: Correlation between each column



There is strong correlation between retweets and favorites

## Question 3: what are the most source of the tweets?



The most tweets have been tweeted from "Twitter application in iPhone"