

# Data Visualization - Using Twitter Data

## Introduction

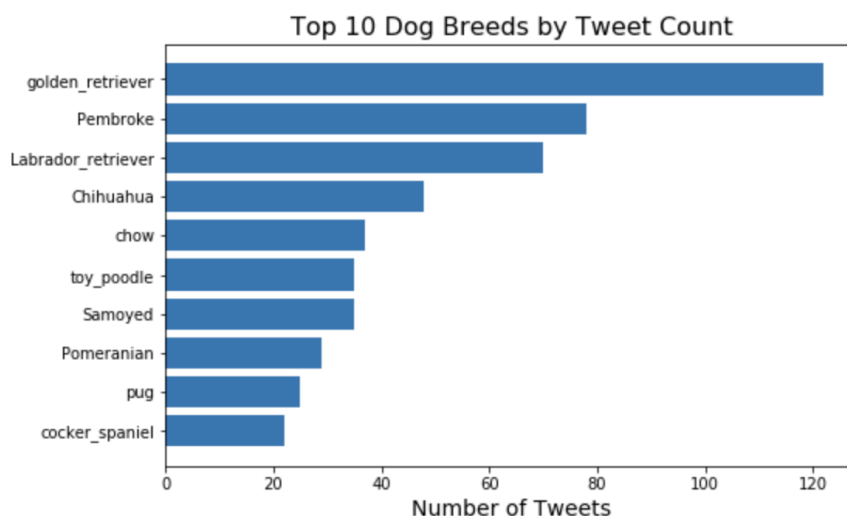
Real-world data rarely comes clean. Using Python and its libraries, you will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. You will document your wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

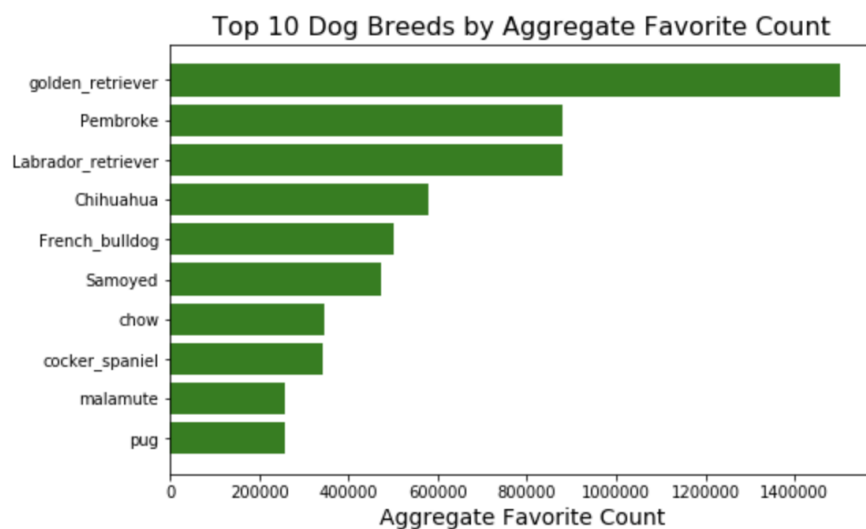
The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user [@dog\\_rates](#), also known as [WeRateDogs](#). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "[they're good dogs Brent](#)." WeRateDogs has over 4 million followers and has received international media coverage.

WeRateDogs [downloaded their Twitter archive](#) and sent it to Udacity via email exclusively for you to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

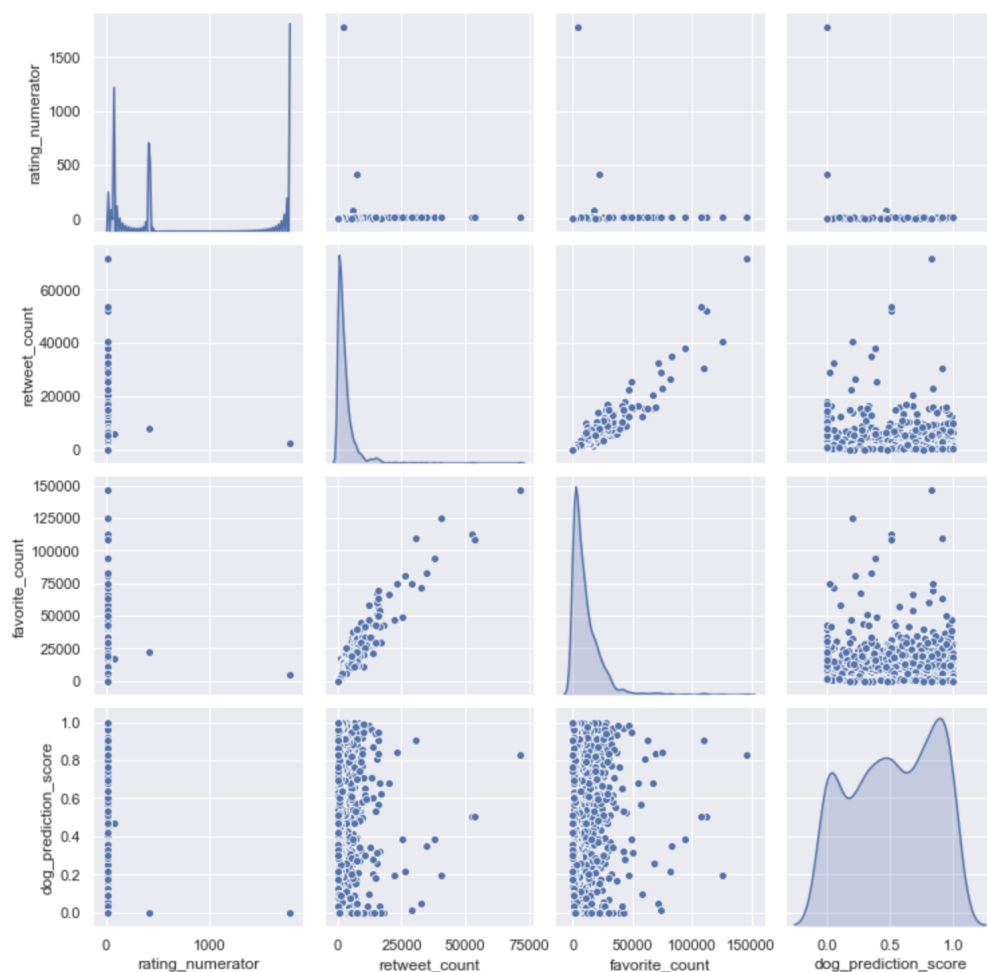
## Data Visualization Findings

Based on Data Wrangling steps followed by data visualization as shown in below 2 images we found out that Golden Retriever dog is the most favored dog breed on counts of number of tweets mentioning the dog and aggregated favorite counts.





We also see high correlation between retweet counts and aggregated favorite count through the following pair plot.



Thank You for making me your favorite.....

