

Wrangle and Analyze 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.

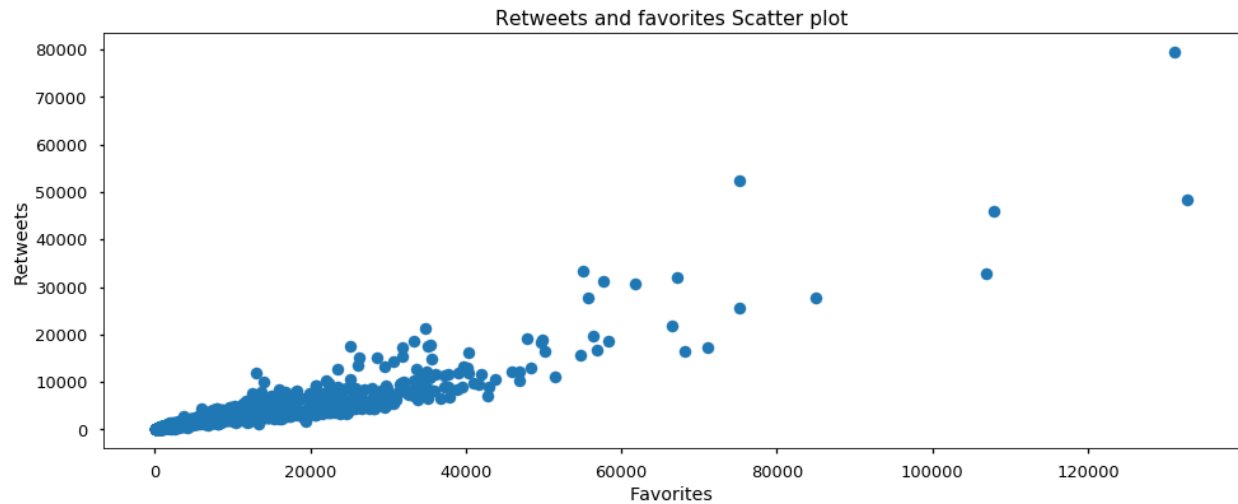
The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user [@dog_rates](#), also known as [WeRateDogs](#). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 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. More on this soon.

Data Visualization

Cleaned and improved data were used to visualize and infer essential takeaways to the tweet data .

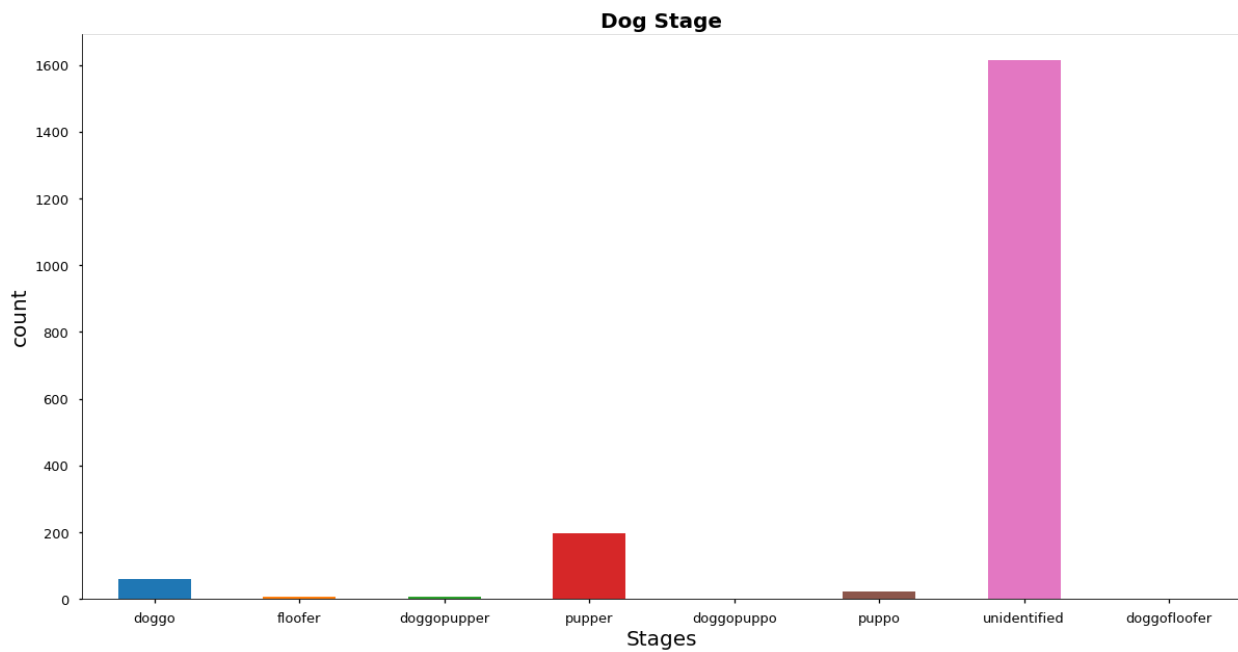
graph 1: Plotting a scatterplot between favorites and retweets.



The outcomes from this plot are as follows:

- A number of tweets happen to have less than 10,000 retweets.
- A lot of tweets tend to have fewer than 40,000 favorites.

graph 1: Bar Chart for Dog Stages.



The outcomes from this bar chart are as follows:

- Dog stages are not the same, the doggo count is over 200 on pupper .
- The pupper have the highest counts and the floofer have the fewest counts.