Udacity

Data Wrangling Project

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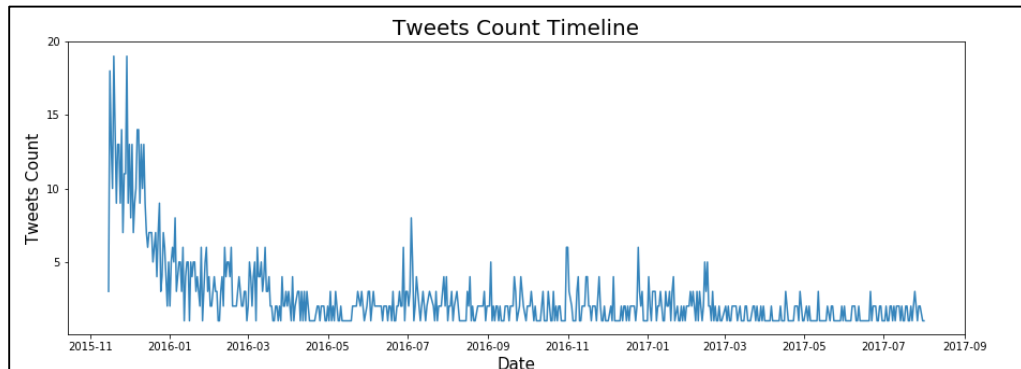
WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for you to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. More on this soon.
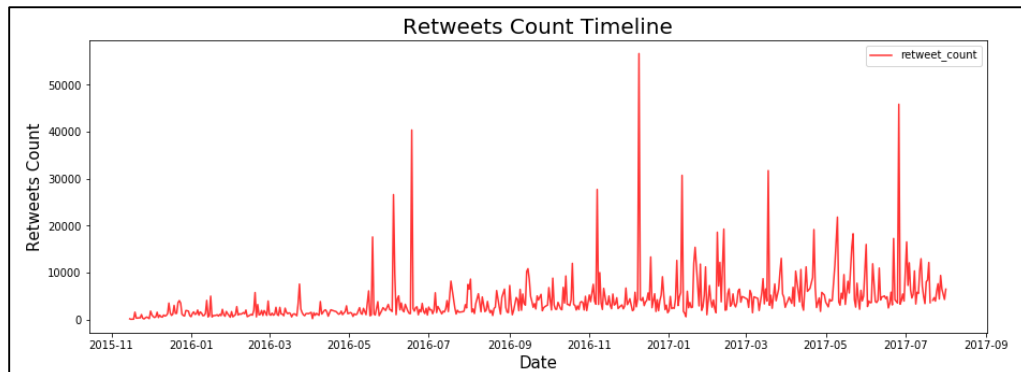
Data Analysis Conclusions

- The most common breed in tweets is: Golden Retriever.
- The most retweeted breed in tweets is: Standard Poodle.
- The most favorite breed in tweets is: Saluki.
- The most rated breed in tweets is: Saluki.
- The most tweeting hours of the highest interaction are: 6 am, 1 pm, and 6 pm.
- Doggolingo usage percentage in tweets is: 15.4%
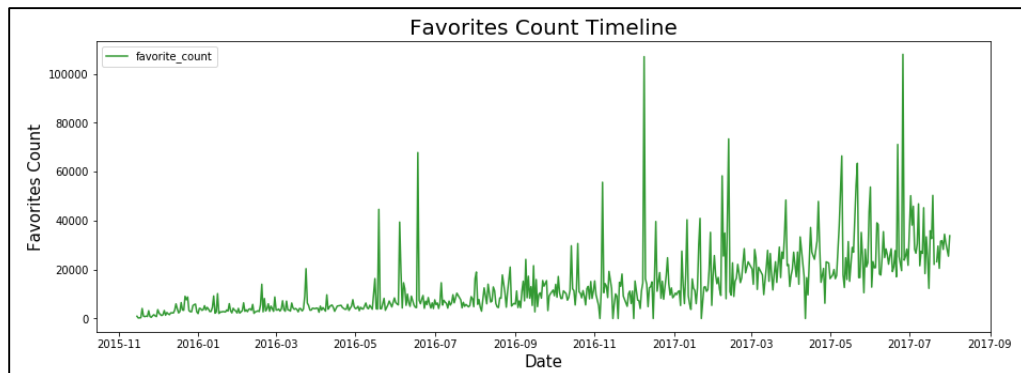
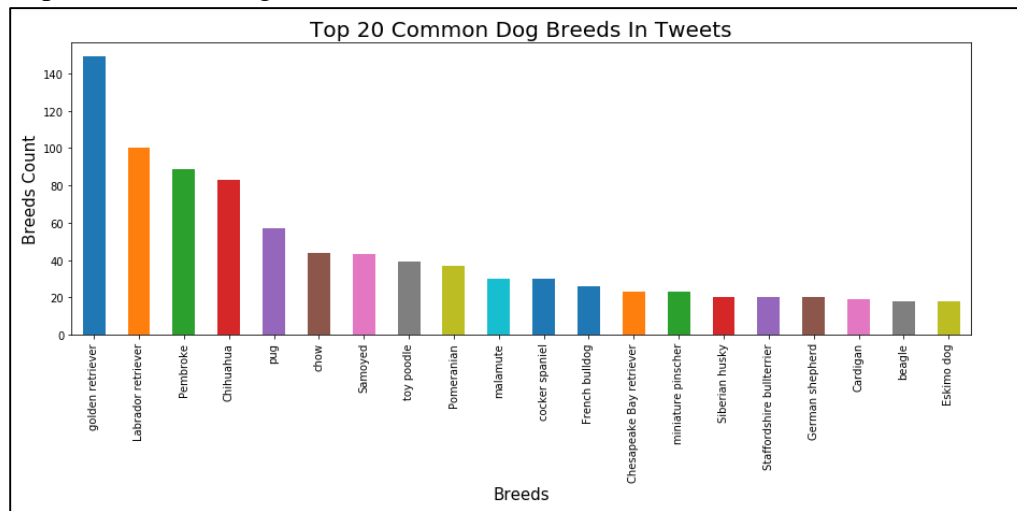# Data Visualization Highlights

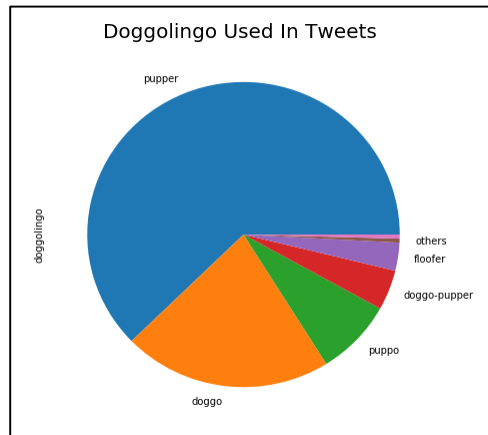- Tweets count timeline.

- Retweets count timeline.

- Favorite count timeline.

- Top 20 common dog breeds in tweets.

- Common Doggolingo used in tweets.

Data Visualization Conclusions

- The number of Tweets decreased while engagement increased.
- The most 20 common breeds in tweets doesn't contain the most interacting breeds.
- More than 50% of tweets which that contain doggolingo are popper.