

Wrangle and Analyze Data: WeRateDogs

Insights into the @dog_rates twitter handle [act_report]



(Source: https://twitter.com/dog_rates)

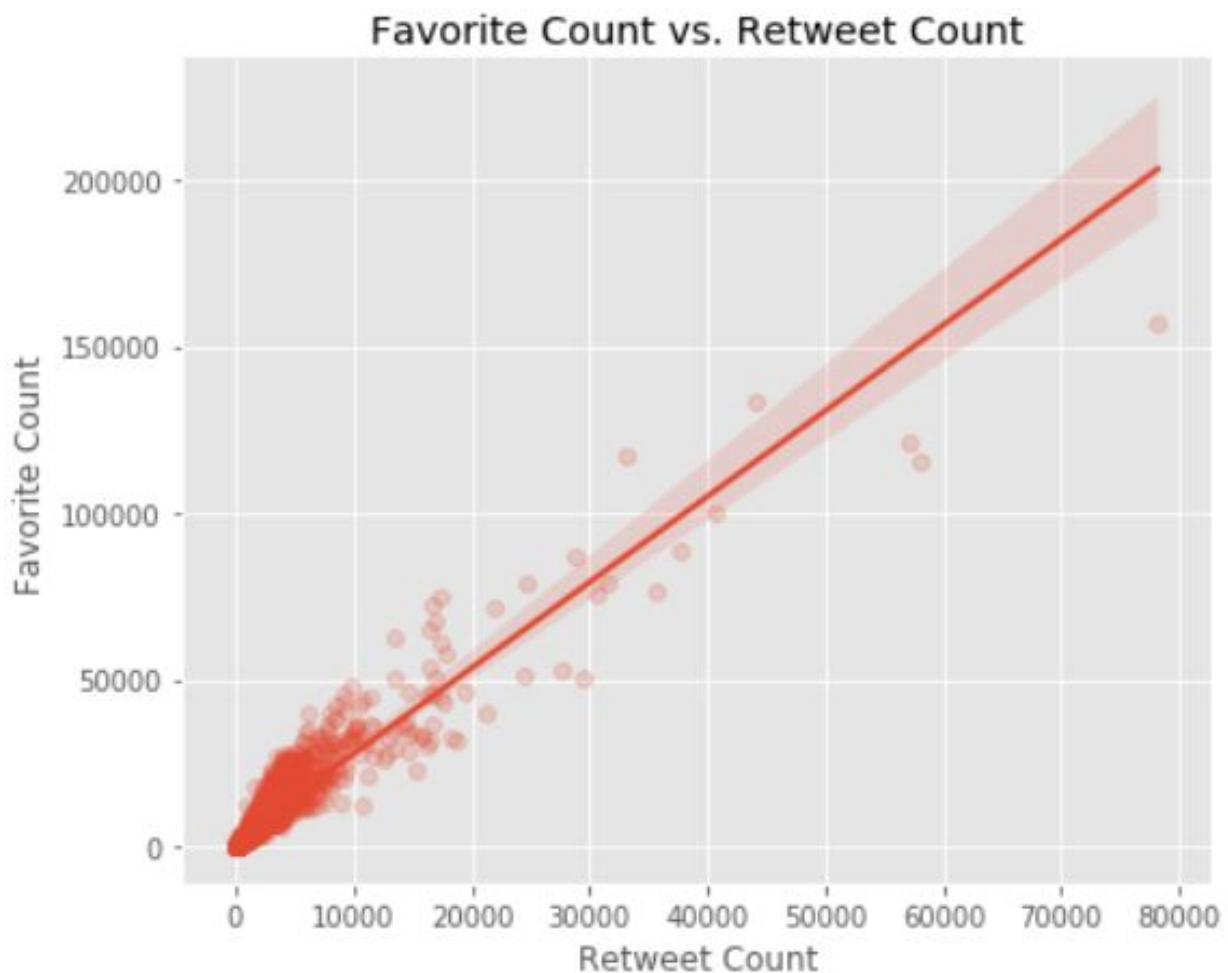
Introduction

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.

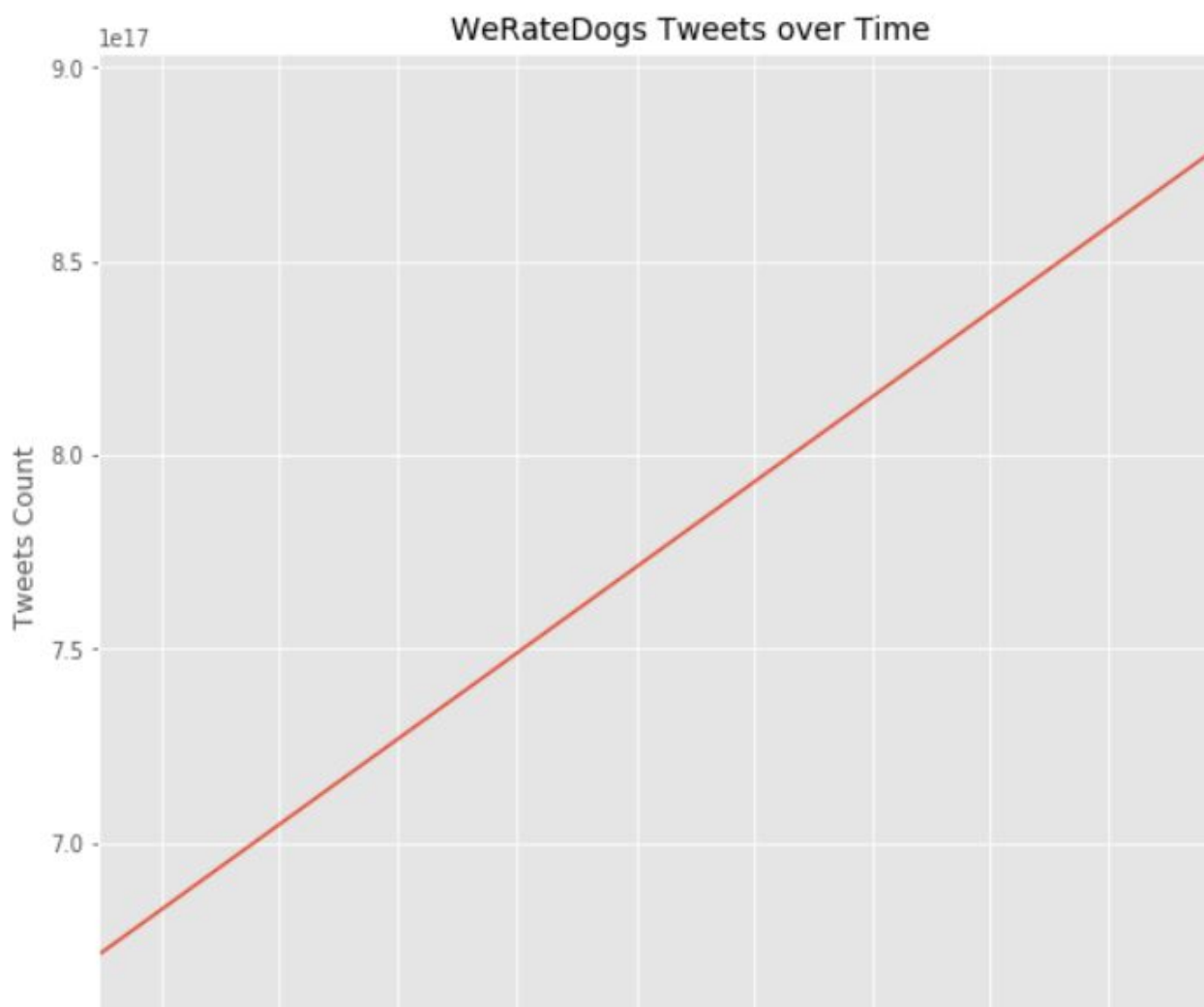
[Ref: *Project Overview section under concepts in Wrangle and Analyze Data*]

Favorite vs. Retweet Counts .



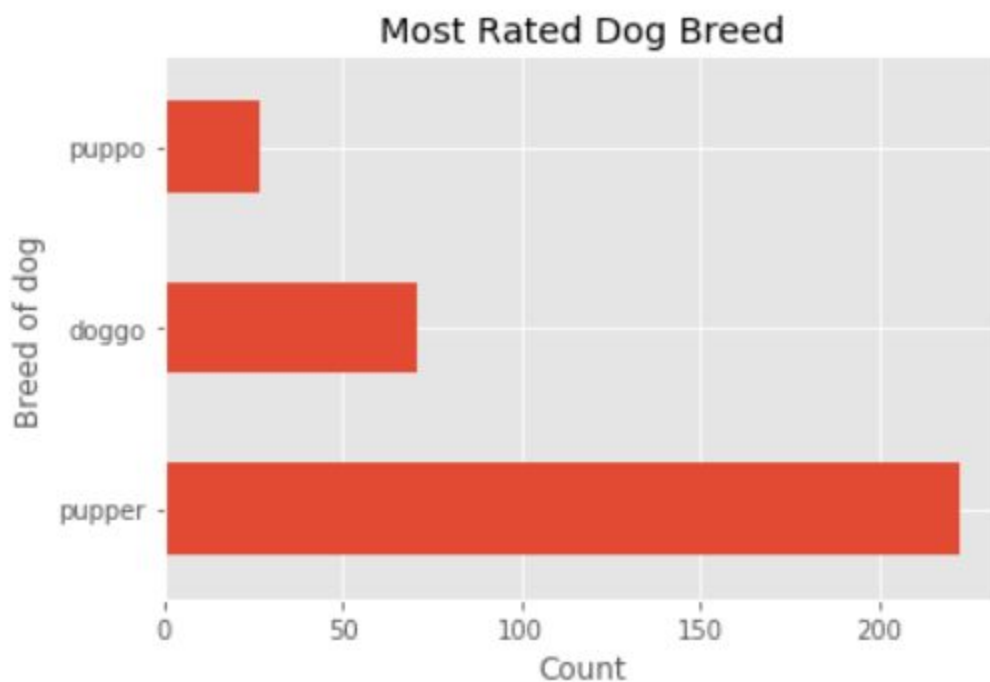
From the visualisation, we can openly figure out that there exists a complete Positive Correlation between favorite counts, and the retweet count. This specific positive correlation can be one of the important insights from the existing data, Where it could easily help to understand and determine the traffic over the twitter handle. Specifically the data analyst team, could further work on the recommended previous posts with either a high retweet count and a high favorite count. In which the page maintainer could follow up the further process based on the analysis for better and productive growth in the new follower acquisition and existing engagement.

Time over



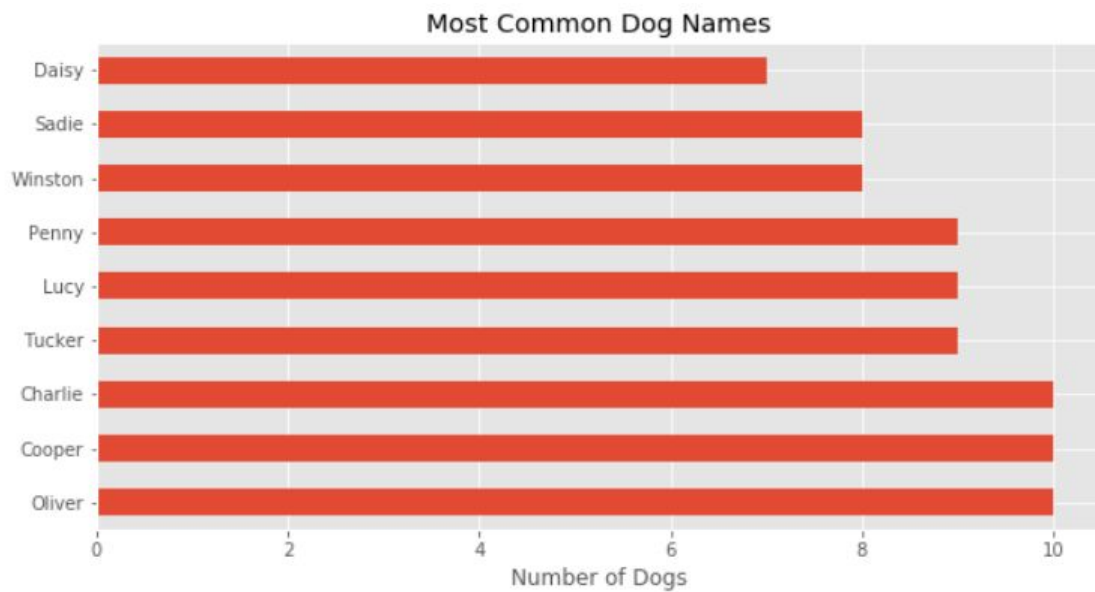
Over the time period of the tweets collected for this particular dataset, tweets decreased sharply starting in early 2016. While the tweets continue to decline over time, there are spikes in activity during the early spring of 2016, midsummer of 2016, but continues to generally decrease from there. This data set didn't provide corresponding data that could provide a reason for the sharp decrease in 2016, and steady decrease from there on out. The owner of the WeRateDogs Twitter account should be aware of this trend, and consider ways to increase user traffic on the page.

Dog Breed Popularity



From the very minimal visualisation of 'dog_type' in the dataset we can figure out that the Pupper breed of Dogs has the majority count.

Most popular dog names



From the ordered bar plot, we can find that that the five most popular names are

1. Oliver
2. Cooper
3. Carlie
4. Tucker
5. Lucky

Conclusion:

This report over the insights provides a narrow look at the Data Wrangling process that was done in the IPYNB file, Furthermore I conclude the report provides a short crisp insight analysis with visualizations.