Reporting: act_report

Introduction

Real-world data rarely comes clean. Using Python and its libraries, the goal of this project is to 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. The data wrangling efforts are documented in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that to be wrangled (and analysed and visualized) 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.

Analysis and Visualizations

Insights

- There a positive relationship between retweet_count and favorite_count
- Charlie is the most used dog name
- Pupper is the rated dog stage

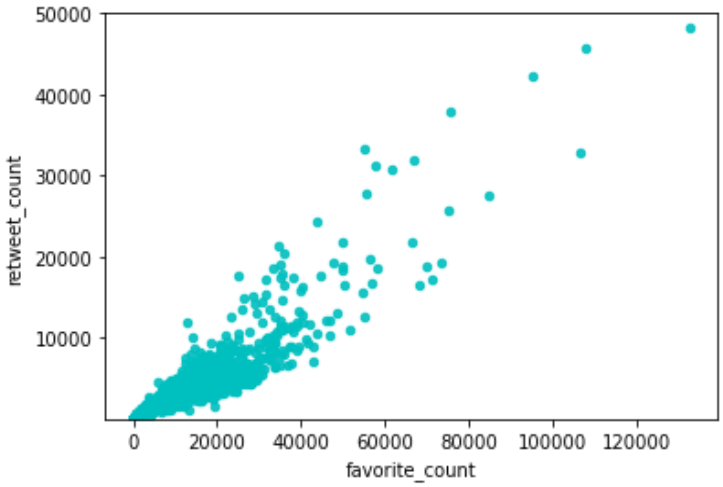

Figure 1

We can clearly see from figure1 that there is a positive relationship between retweet_count and favorite_count.

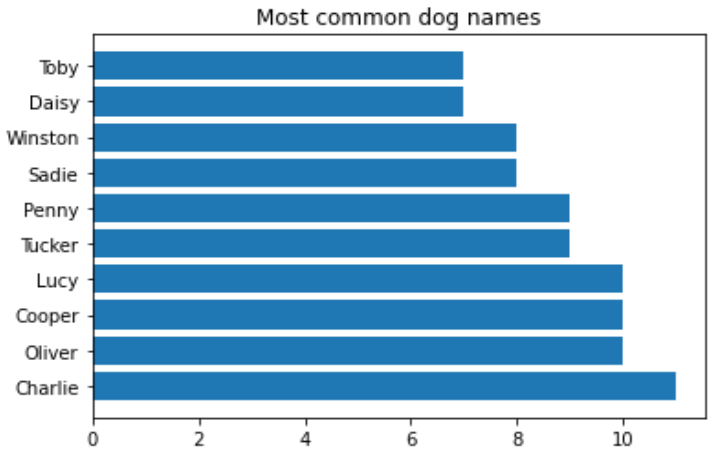

Figure 2

Figure 2 indicates that Charlie accounted for the most dog name used.

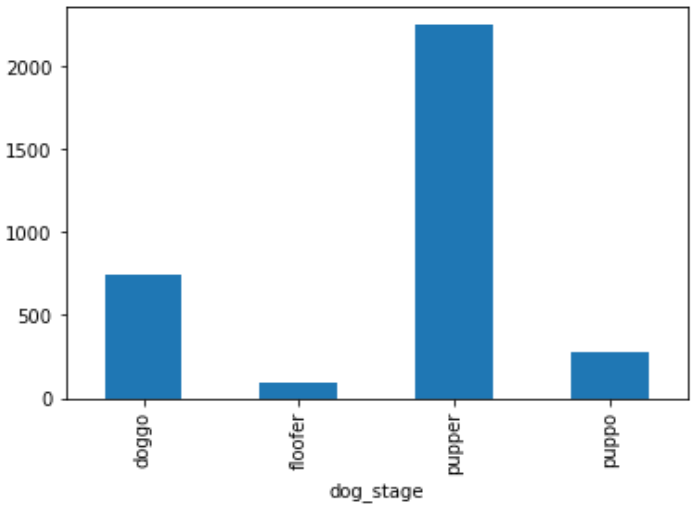

Figure 3

Figure 3 also indicates that pupper is the stage rated the most.