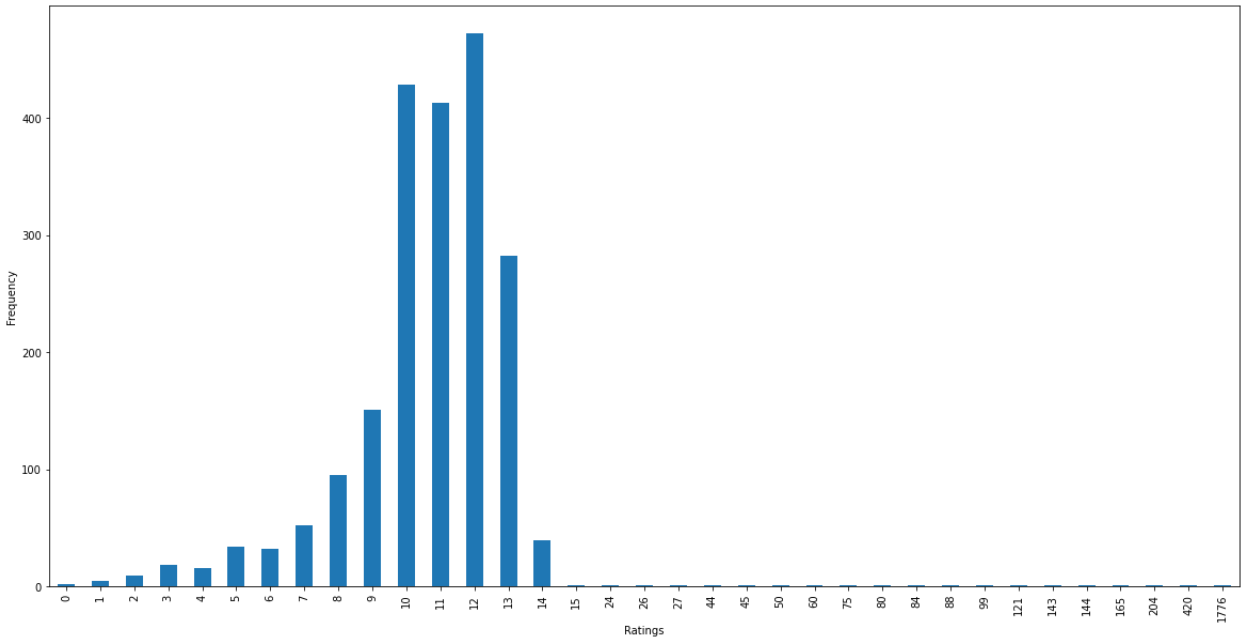
Act Report

Project Overview

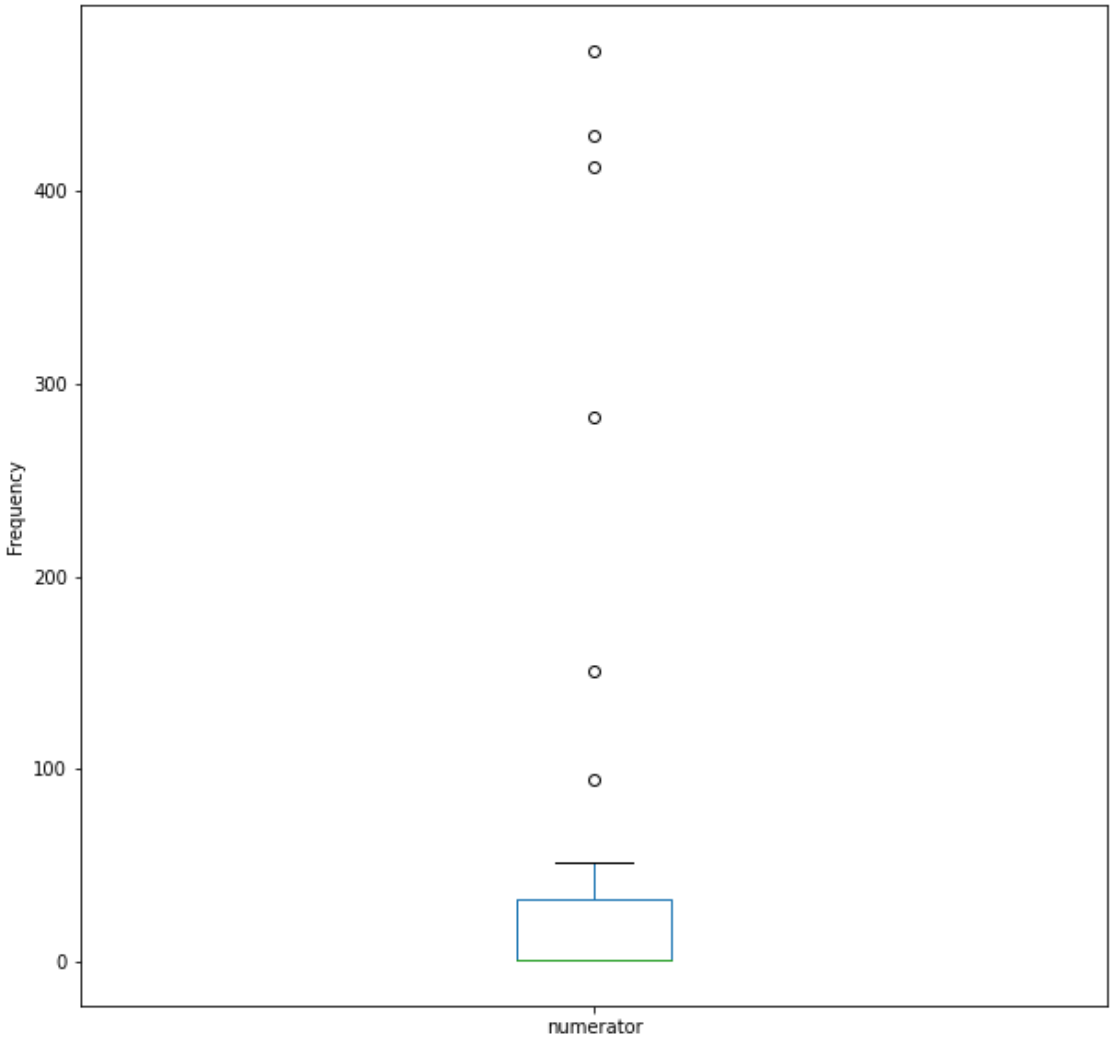
Real-world data rarely comes clean. Using Python and its libraries, we can gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. The task is to document our wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) .
The dataset that we are 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.

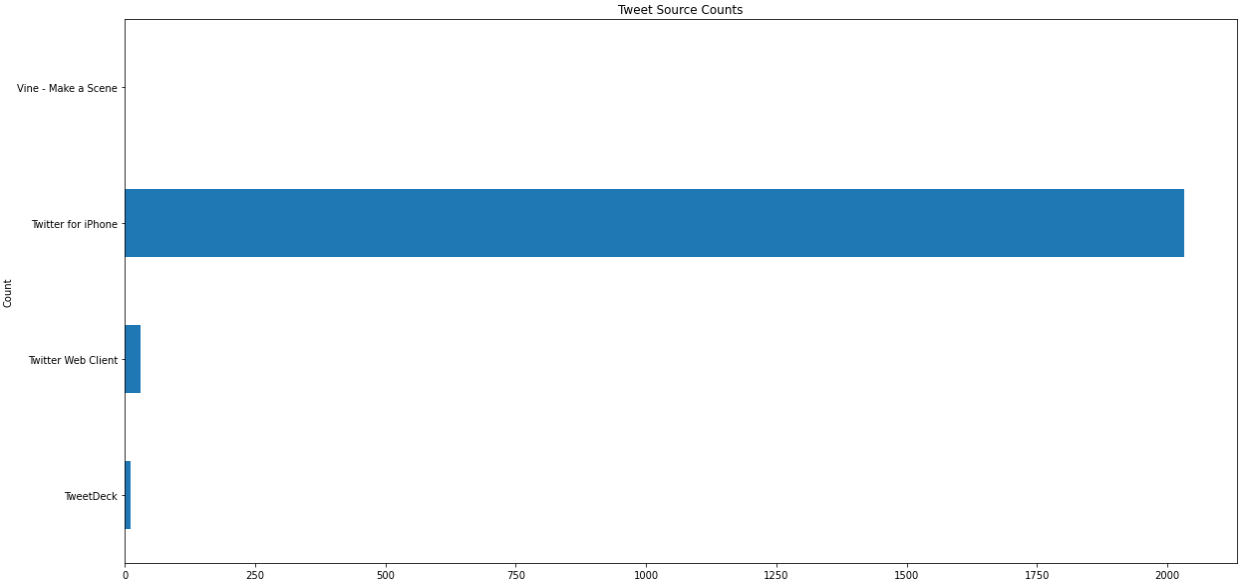
Visualization

- Plotting a bar chart for ratings distribution's frequency.



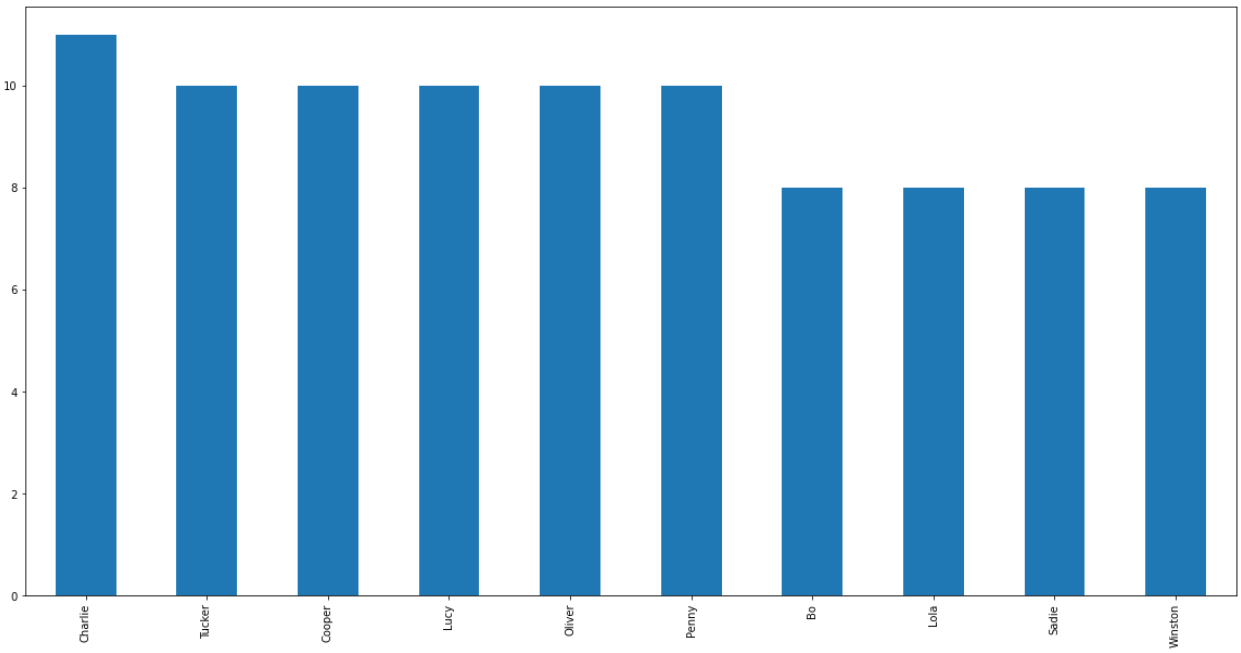
- same distribution in box plot





-plotting dog stages and noticing that pupper is the most common stage

-Most popular dogs names



- A relation between # of retweets and dog_stage, we notice that # of retweets is very high when dog stage is doggo or pupper

  This shows that most dogs are in Puppo category but highest retweets are about doggo category.