Act Report

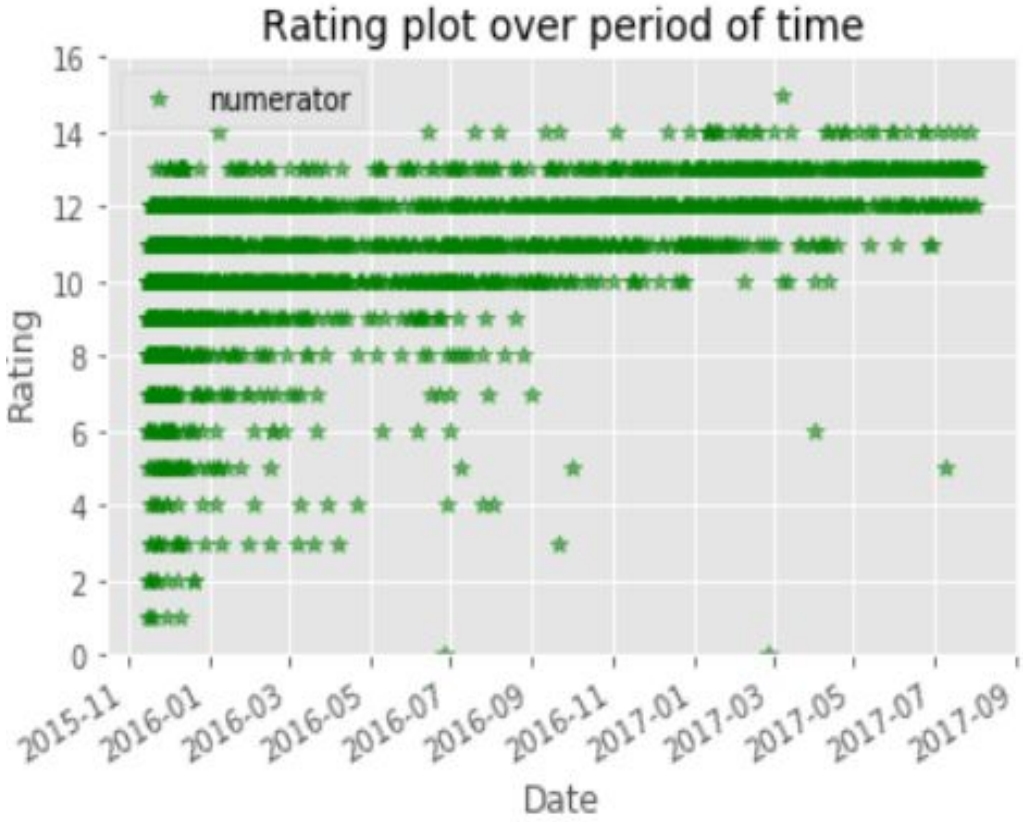
By: Mina Ghabbour

- Project Details:

Real-world data rarely comes clean. By 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. Here I will show my wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using different libraries in Python.

The dataset we are wrangling 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.

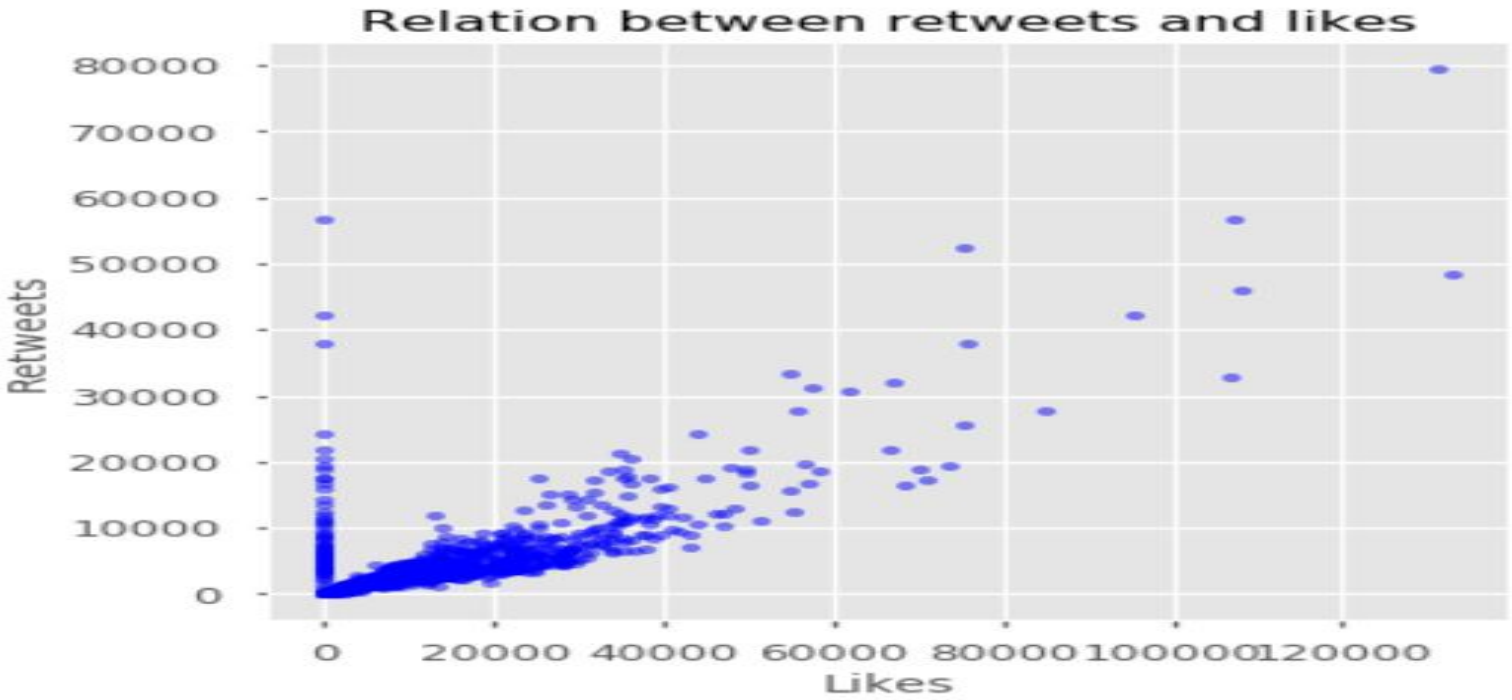


Here in this plot we can see the rating values over a period of time to discover which rating values are the most used.

So when we check the plot we can see that the rating value of 12 is the most common over a certain period of time.

- ANALYSIS AND VISUALS

Here we can see that there is a linear correlation between likes and retweets so when the retweets increase the likes increase too and vice versa.



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

There are many outliers as can be seen as trailing tails of boxplots. Plus it can be visualized that the count of retweets are in very large no. Means WeRateDogs is very popular among its followers. WOW!