Doggy Data Insights!

Let’s dig in…

As part of a Udacity Data Analyst Nanodegree project, WeRateDogs (@dog_rates) data was analyzed to gain insights on names, ratings, and whether a relationship might exist between ratings and combined retweet and favorite/like counts.

The project was fun and I learned a lot, but it was a bit of a bear at times!

Twitter Photo from WeRateDogs – Dog or Bear?

The main goal of the project was to wrangle data effectively so that insights could be drawn and a visualization prepared. Three sources of data were required, and the wrangling effort is comprised of 3 main steps, which are outline briefly below:

- Gathering – this step entailed collecting an “on-hand” file and a downloadable file from Udacity, while the third source of data was Twitter itself which was accessed via an API.

- Assessing – after the data is gathered it is given a high-level review and any peculiarities or seeming issues with the data are record. Issues can fall into one of two groups, Quality or Tidiness.

- Cleaning – Once the pertinent issues are identified, the data sets are worked and massaged to fix the issues and allow the data to be used for analysis and visualization.

Twitter Photo from WeRateDogs – Cute Puppy

Analysis and Visualization

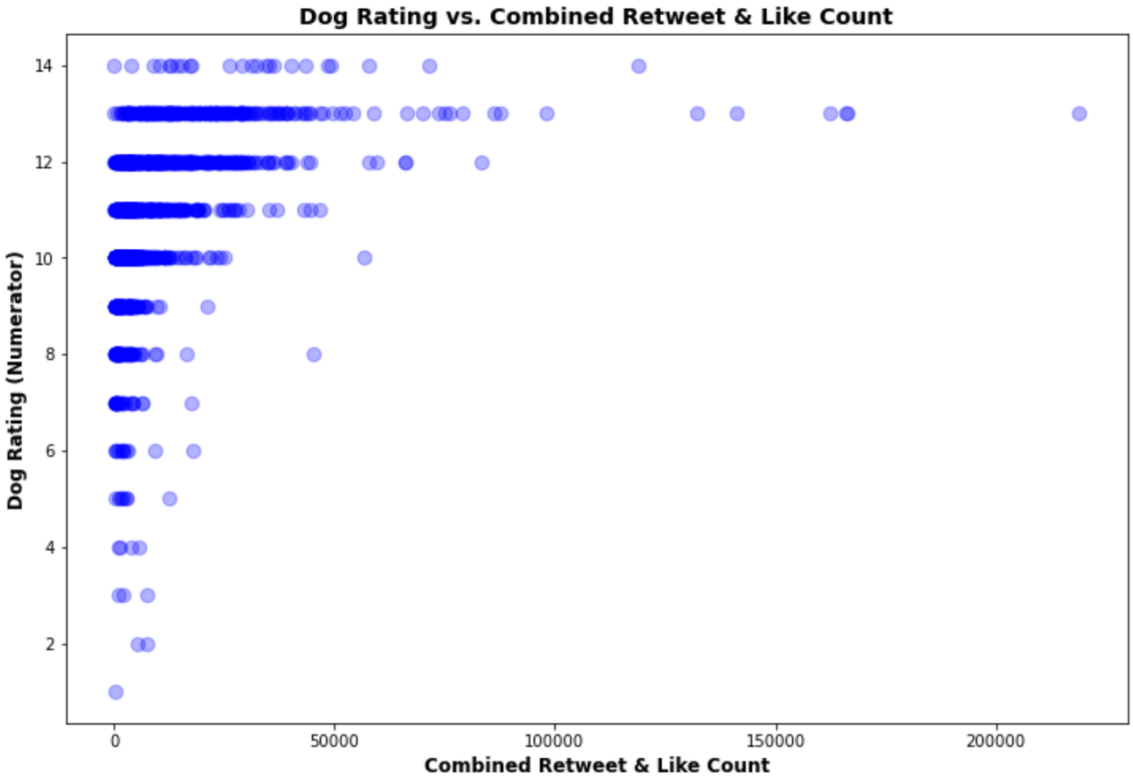
After summary cleaning work, summarized data revealed the following:

1. The most popular dog name appears to be "Cooper" and there are many unique names. (Tied for 2nd place are “Charlie” and “Oliver”, and in third “Tucker”).

2. Sadly many owners do not provide their dog’s names! (and in some cases, folks ruin the fun by submitting other animals….although, they can be very cute! …Did you happen to notice the bear on page 1?)

3. The top rating on WeRateDogs appears to be 14/10, and only 22 doggies received this high honor!

4. Do dogs with higher retweet and favorite/like counts have higher ratings? …Well this is something we aimed to find out as part of the project. What does the data tell us?



Well from the graphic above (aka scatter plot), it does indeed look like the dog ratings are positively influenced by retweet & like counts. Also of note, it looks like some dog tweets went viral!

*Source data and images provided courtesy of WeRateDogs, Udacity, and Twitter.