Act Report

My second project for the Udacity Nanodegree on data analysis is based on data wrangling. The dataset I wrangled, analyzed, 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 humorous comments about the dog. These ratings almost always have a denominator of 10. However, the numerators are almost always greater than 10, e.g 11/10, 12/10, 13/10, etc.

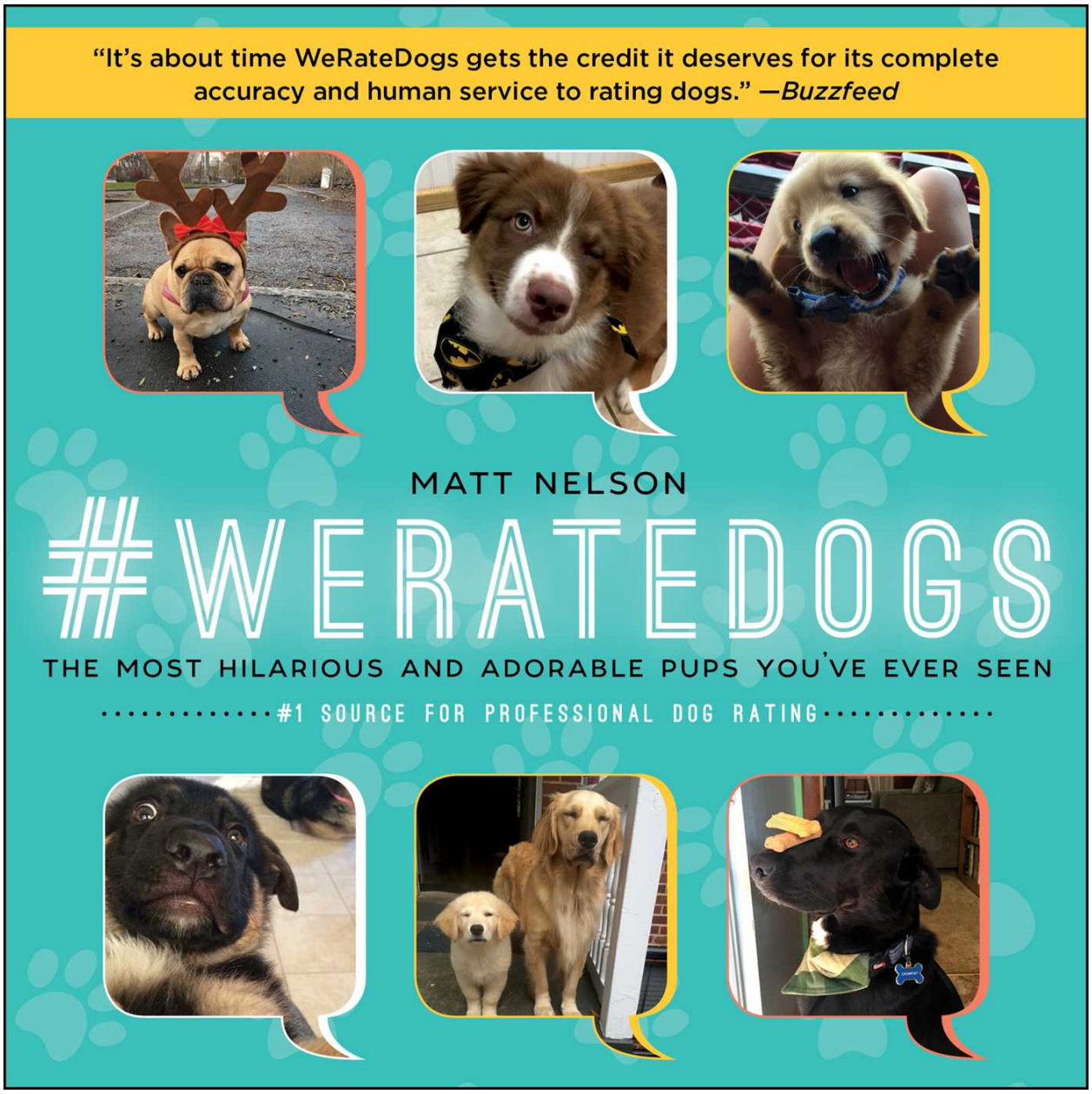


Image via simonandschuster.com

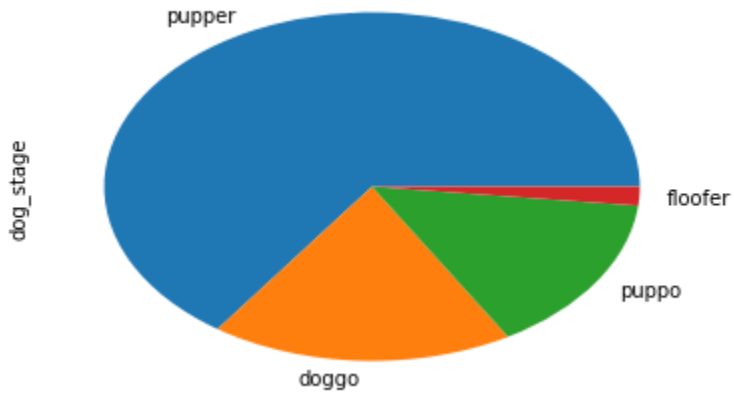
After cleaning the data, I analyzed, visualized and got some interesting insights. The insights answered some questions I had in mind before cleaning the dataset. These are the questions and some accompanying visualizations:

1. What dog names are most common? Charlie, Oliver, etc.

2. What tweets have the highest average favorite count? 807106840509214720 (tweet ID), etc.

3. What dog stage has the highest average favorite count? Floofer.

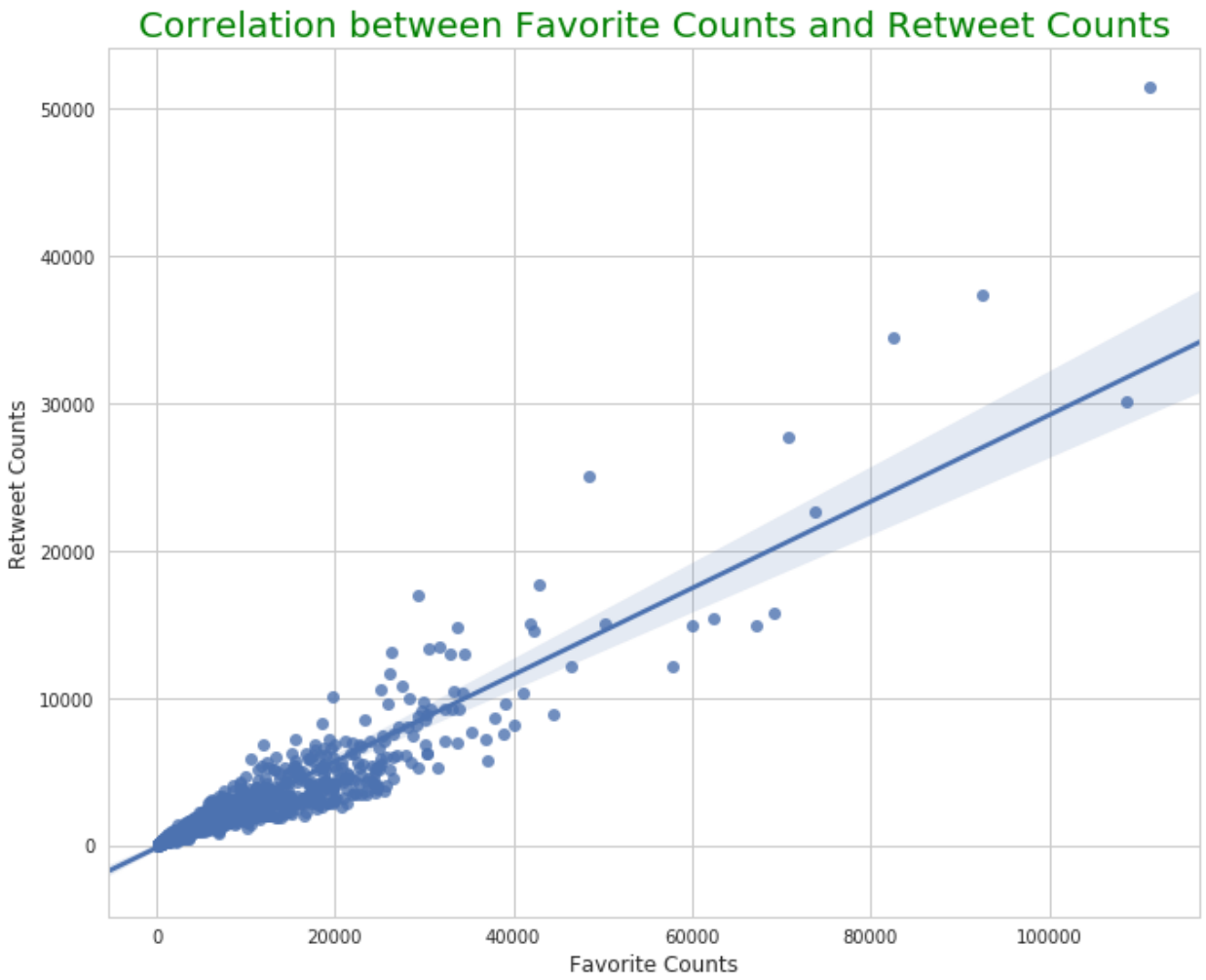
4. What is the most common dog stage? Pupper.



These are definitions of the various stages of dogs: doggo, pupper, puppo, and floof(er) (as explained by the #WeRateDogs book on Amazon)

- Doggo: A pupper that appears to have its life in order, a bigger pupper (usually older)
- Pupper: A small doggo, usually younger. Can be equally, if not more mature than some doggos.
- Puppo: A transitional phase between pupper and doggo, easily understood as the dog equivalent of a teenager.
- Floofer: Any dog with seemingly excess fur.

5. Is there any correlation between favorite count and number of retweets? Yes, positive correlation.



Completing this project was very interesting and enlightening. I look forward to completing more projects!