For this project, we worked on wrangling, analyzing and visualizing the data of a twitter account. The twitter user is @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs has over 4 million followers and has received international media coverage.

First, we began by gathering the data from three different sources:

1. The WeRateDogs Twitter archive, it was sent to Udacity via email exclusively to be used in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.
2. The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file is hosted on Udacity's servers and should be downloaded programmatically using the Requests library.
3. Scraping twitter for data using twitter API tweepy.

After that we worked on assessing the data using python libraries, we examined and took a closer look at the data to identify any quality or tidiness issues. We identified 8 quality issues and two tidiness issues.
The issues are as follow:

Quality Issues
1. Some of the dog names are incorrect/missing.
2. Columns timestamp and retweeted_status_timestamp are strings when they should be Datetime.
3. Get rid of outliers in rating_numerator and rating_denominator.
4. Drop replies and retweets as they are not original tweets.
5. Merge the dog breeds into one column, given the highest confidence.
6. Missing values in expanded_urls.
7. clean up the source column remove HTML tags.
8. Remove underscore in breed names.

Tidiness Issues
1. Dog type is represented using four columns: Doggo, Floofer, Puppo and Pupper, we can merge them into one column with the dog type.
2. Remove unnecessary columns.

Finally, we worked on solving each problem and bring to light three insights and a visualization.