# Read me

Dataset: WeRateDogs Twitter data and additional gathering, then assessing and cleaning.

1. twitter archive from WeRateDogs (file downloaded manually).
2. image_predictions.tsv file (file downloaded programmatically).
3. Query twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet_json.txt file.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Image predictions file contains the predictions a neural network does about the dogs in WeRateDogs: are they dogs? What breeds are them? etc.

We have two goals here: **to analyze the neural network and learn about the dogs that appear on WeRatgeDogs.**

Findings:
How reliable is this neural network?

- The first prediction is the most reliable, with 59%, followed by a large difference from the other two predictions, with 13% and 6% respectively.
- No correlation is observed between number of images and dog prediction in the #1 prediction.
- What the neural network identifies as not a dog seems to be part of a common scenario in the photos (e.g. patio or bow tie)

Learning about the dogs

- Each prediction shows a different ranking of breeds. However, they do share some breeds: Labrador Retriever, Chihuahua, and Golden Retriever.
- The breeds that receive the most likes and retweets are not breeds or dogs. In this case, the neural network identifies things.
- There is no recurring name. Ratings don't match the most named names, either.
- Twitter users only like the ratings between 11 and 14 out of 10.
- The double category of Doggo and puppo.

Process:

1. Data wrangling: gathering, assessing and cleaning.
2. Storing, analyzing, and visualizing your wrangled data
3. Reporting on 1) your data wrangling efforts and 2) your data analyses and visualizations