We Rate Dogs:

Wrangle, analyze and visualizations.

1. Importing Libraries --Done
2. Gather the Data -- Done
3. Using pandas merge all of the datasets (archive, api, and image recognition) based on “Twitter\_id” – need to double check to make sure we are not missing data and that the merge is working correctly
4. Create an SQL database and correct all of the datatype issues (Check Twitter ID may want to change this to int64).
5. Drop columns we don’t need.

**Data Inclusion Criteria**

We are expected to use the following criteria to select the required data:

* Do not include retweets
* Only tweets that have images

An additional comment was made in student discussions: *The archive also has reply tweets which in general contain upgraded/downgraded ratings of the dog. This means that in some cases there are two observations/ratings for the same dog*. As a result, I decided to only include original ratings and so developed an additional criteria:

* Do not include replies