My wrangling efforts:

For my first quality cleaning effort, I checked to see if there are any tweets that do not contain a URL in the text field. I “removed” those that did not contain a URL, however there were none. I then removed all tweets that contained the text ‘RT @’ as those were the tweets that were retweets. I changed all of the names that had lowercase values because those were all not names and I changed them to the word “None”. To elaborate, there were some names in the name column with singular letters or not names such as “quality”, “a”, etc. I changed the tweet_id from float64 to int as I was having issues converting those values to strings which I used in my next step. I concatenated the URL ‘https://twitter.com/dog_rates/status/’ and the tweet_id to create the tweet links. I then changed the name of that column from “external_urls” to “tweet_link” as it seemed more appropriate. I randomly stumbled upon a tweet that contained a moose (881268444196462592) and I removed it as it was not a dog. I removed the image links from the “text” column as I wanted to have just the literal text from the tweet. I then converted the timestamp column from a string to a datetime, which in turn removed the “+0000” from that column.

As for my tidiness efforts, I combined the tables ‘twitter_archive_enhanced’ and the ‘tweet_selected_attr’ to have retweets and favorites on the ‘twitter_archive_enhanced’ table. I then had to remove the duplicated timestamp column and rename the original timestamp column back to the original name (timestamp). I then removed the columns: 'source', 'in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp' as I deemed them unnecessary. Lastly, I removed the columns 'jpg_url', 'img_num', 'p3', 'p3_conf', 'p3_dog' from the image_predictions table. I then combined the newly merged table with the image_predictions table.