WeRateDogs Tweet Analysis

Introduction
The goal of this project was to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning was required for "Wow!"-worthy analyses and visualizations.

The Data
In this project, I worked on the following three datasets.

- Enhanced Twitter Archive
- Additional Data via the Twitter API
- Image Predictions File

Quality issues

Enhanced Twitter Archive

1. `tweet_id` has the wrong data type (int64 vs object).
2. `timestamp` has the wrong data type (object vs datetime).
3. Strange and unfamiliar pet names in the name column
4. some records are retweets and replies and wouldn't be needed. (78 replies and 181 retweets)
5. The source column has urls with the source at the end instead of just the source.
6. Replace 'none' names with something more descriptive

Image Predictions

7. `tweet_id` has the wrong data type (int64 vs object).

Additional Tweet Data

8. `id_str` has the wrong data type (int64 vs object).

Tidiness issues

Enhanced Twitter Archive

1. The dog stage `doggo`, `floofer`, `pupper`, `puppo` needs to be combined into 1 column.

2. We don't need `retweeted_status_id`, `retweeted_status_user_id`, `retweeted_status_timestamp`, `in_reply_to_status_id`, and `in_reply_to_user_id` colunms since the relevant data is in the additional data file.

Image Predictions

3. We only need the column with the highest probability as oposed to `p1`, `p2`, and `p3`.
4. There are records where the image the model predicts are not dogs
5. Some columns are not needed for the analysis like `img_num`

Additional Tweet Data

6. `id_str` needs to be changed to `tweet_id` so the 3 tables can be merged using it as key.
7. We only need 3 columns `id_str`, `retweet_count`, `favorite_count`.
8. Information is repeated across multiple tables

After carefully fixing all issues stated above, the 3 datasets were merged into one for the analysis phase.