Wrangle Report - We Rate Dogs

Stage 1: Gathering Data

```
The twitter-archive-enhance.csv file was provided by Udacity. I downloaded
it, and then uploaded it to my Jupyter Notebook environment.

Since this was a file containing data considered archived data, I named the
dataframe I loaded the .csv file into 'archive'

Udacity provided a URL to where a .tsv file named imagepredictions.tsv was
downloaded from via the Jupyter notebook.

I loaded the .tsv file into a dataframe named 'imgPredict'.

The Twitter API appears to no long work as it once did, so I had to resort to
using the JSON file provided by Udacity. I downloaded this file and stored it
as 'jsonlist'.
```

Stage 2: Assessing

```
I went through each of the files and checked for inconsistent or messy data.

I did this both visually and programmatically.

This issues found will be covered more thoroughly in the cleaning stage.
```

Stage 3: Cleaning

```
Initially, a copy of the data was created to maintain data integrity.

Then I began to address the issues reported.

Changed the timestamp to datetime datatype

Changed the numerator datatype to float

The tweet_id was changed to the string datatype in all three sets of data.

Denominators with a value higher than 10 were changed to 10 and the numerator
was changed accordingly.

Dog names that are invalid were changed to "No Name"

Deleted rows that had non-null values for tweets and replies
```

Once the non-null tweets and replies were removed, the following columns were removed:

- in_reply_to_status_id
- in_reply_to_user_id
- retweeted_status_id
- retweeted_status_user_id
- retweeted_status_timestamp

The HTML portion of the source was removed

The expanded_urls column was removed.

The columns of doggo, floofer, pupper, and puppo items were merged into a new column and then dropped.

The dataframes were copied, then merged into a new dataframe named df_main.

The data was then stored in a .csv file named twitter_archive_master.csv