

## Reporting: wrangle\_report

WeRateDogs wrangling efforts.

First and foremost, WeRateDog sends the archive dataset to Udacity for the purpose of this project. My first task in the wrangling process was loading and reading the dataset into Jupyter notebook using the pandas library.

The instruction in this project was that using the data provided by WeRateDogs is not enough for the project as it only contains basic information tweet information. And for this reason, using the requests library, I have downloaded WeRateDog's image prediction datasets hosted on the Udacity website.

Following, I use the Twitter API to extract additional data from Twitter, and unfortunately, the access was denied thus I decided to use the provided JSON file.

And I started assessing the data using panda's functions like `.head()`, `.info()`, and others. Upon assessing the data both visually and programmatically, I found that the data was of low quality and not tidy. Thus I documented both quality and tidiness issues.

After documenting quality and tidiness issues, I made copies of all the original data and started with the cleaning process. I checked for null values, and I also went on removing columns that in my mind thought were not necessary to complete the project. Also, I have changed the date format and renamed some columns to match with other tables. I also converted the bool values into int as well as checked the duplicate values.