

# wrangle\_report

August 21, 2023

## 0.1 Reporting: wrangle\_report

- Create a **300-600 word written report** called "wrangle\_report.pdf" or "wrangle\_report.html" that briefly describes your wrangling efforts. This is to be framed as an internal document.

Three different sources of data from twitter feed of WeRateDogs is gathered and loaded into pandas dataframes.

**Gathering** Twitter-archived-enhanced.csv was downloaded manually and loaded into a pandas dataframe. Image-predictions.tsv was downloaded programmatically and loaded into a dataframe. For the Twitter API data, tweet-json.txt file is manually downloaded. Tweet-json.txt required several operations before it could be loaded into a pandas dataframe

**Assessing** Assessing of data was done both visually and programmatically on all the three dataframes. Few quality issues and tidiness issues were identified in the data.

**Cleaning** Copies of original dataframes were made. All the quality issues and tidiness issues identified in the assess stage were defined, coded, and tested. Image Predictions and Twitter JSON API data were merged into the twitter archive table. It was later found that there was some valuable information embedded in the source column of the twitter archive table, which was extracted. All the other columns which are insignificant to analysis were dropped.