

# wrangle\_report

September 12, 2022

## 0.1 Reporting: wrangle\_report

I had three datasets to work with the first twitter\_archive dataset, twitter API data and image predictions tsv file. I first gathered all these data twitter\_archive dataset, image\_predictions and twitter API datasets I used requests method to download all them as files in my working directory.

From the gathering process I assessed these datasets. I was able to find couple number of problems in the dataset which include misrepresented data types, rating denominators below and above 10 whilst they're supposed to 10, I noticed tidiness issues where variable is to form a column but 1 variable had formed many columns.

from the assessment of the data I then cleaned the data,I placed doggo, floofer, pupper and puppo into a single column which i named dog stages. I changed timestamp datatype from object to datetime. I made sure that rating denominators where equal to 10 for all the rows in the dataset.