

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.

Introduction

This report presents and discusses the process of wrangling to assemble and clean the data that was required for Project 2: Wrangle and Analyze a Dataset of the WeRateDog twitter page and other data sources.

Wrangling is a vital process in the data analysis life cycle and it is made up of:

- Gathering
- Assessing
- Cleaning data.

1. Gathering Data

This was done in 3 ways in order to come up with 3 data sets.

The first method was downloading a dataset on WeRateDogs which was a csv file.

The second method was to programtically download and save a tsv file from the internet.

The third method was to use the Twiiter API to download data off twitter.

At the end of the process, i had three datasets and i had to commence the assessment and cleaning steps

2. Assessing the Data

Assessing the data was done visually and programatically and i came across quality issues like Keep original ratings (no retweets) that have images, Delete columns that won't be used for analysis, Incorrect columns case in point doggo, floofer, pupper and puppo columns, Separate timestamp into day - month - year (3 columns), Correct numerators with decimals, Correct denominators other than 10 among others as well as tidiness issues like the need to combine the three datasets into one.

3. Cleaning the Data

Cleaning the data became much easier due to the fact that the issues had already been identified. I then proceeded to tackle issue by issue by correcting the issue and then running a test to ensure the problem had been solved.

After the data had been cleaned, I then stored it in a new dataframe then converted into a csv "twitter archive master".

Outro

This wrangling process enabled me to build a foundation to carry out my insights and visualizations as they other report indicates.