

# **PROJECT: Wrangling And Analyze Data Report**

PROJECT: Wrangling And Analyze Data Report.....	1
.....	1
Gathering.....	1
Assessing Data.....	1
Data Cleaning.....	1
Storing data.....	2
Wrangling data.....	2
References .....	2

## Gathering

In this project, we gather, assess, and clean data and then act on it by wrangling the data.

WeRateDogs Twitter archive was downloaded from Ucadity (given to us by downloading it from the classroom)

Twitter Image Prediction (Image\_predictions.tsv) was downloaded programmatically from the URL using the Requests library

Additional data from Twitter was downloaded using Twitter API (Tweepy) and stored as tweets\_data.txt

Data was then assessed visually and programmatically before being cleaned for exploratory analysis

NB: API keys were removed from the work before submitting

## Assessing Data

Data gathered was assessed visually and programmatically, indicating quality and tidiness issues.

## Data Cleaning

Here, we were focused on :

1. First, make copies of all the datasets before cleaning them individually.
2. Data cleaning was done using the "Define-Code-Test" framework
3. Eradicating unnecessary columns from tables in each dataset
4. Renaming columns with more descriptive names for easy understanding of the column's representation
5. Removing or dropping rows with a missing value(s)
6. Changing the data types of columns to the appropriate datatype needed for our wrangling

## Storing data

At this step, we merged all the cleaned copies of the dataset (all three of them) and renamed it `twitter_archive_master`.

## Wrangling data

As the last step, we focused on finding insights from the data and visualizing the findings from the data using appropriate graphs to answer questions and tell the story the dataset presents.

## References

- <https://stackoverflow.com/questions/28384588/twitter-api-get-tweets-with-specific-id>
- <https://www.youtube.com/watch?v=bNDRiaFyLrs&t=28s>
- <https://www.youtube.com/watch?v=Lu1nskBkPJU>
- <https://medium.com/analytics-vidhya/fetch-tweets-using-their-ids-with-tweepy-twitter-api-and-python-ee7a22dcb845>
- <https://stackoverflow.com/questions/54407027/use-api-to-write-to-json-file>
- <https://www.geeksforgeeks.org/python-convert-tsv-to-csv-file/>

<https://sparkbyexamples.com/pandas/pandas-change-string-object-to-date-in-dataframe/>

<https://stackoverflow.com/questions/44116194/import-a-function-from-another-ipynb-file>