

## INTRODUCTION

- Data wrangling, which consists of:
  - Gathering data.
  - Assessing data.
  - Cleaning data.
- Storing, analyzing, and visualizing your wrangled data
- Reporting on:
  - 1) your data wrangling efforts.
  - 2) your data analyses and visualizations

## GATHERING DATA.

Gather each of the three pieces of data:

1. The WeRateDogs Twitter archive.
2. The tweet image predictions.
3. Each tweet's retweet count and favourite ("like") count at minimum, and any additional data you find interesting. Using the tweet IDs in the WeRateDogs Twitter archive, query the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called **tweet\_json.txt** file.

## ASSESSING DATA.

After gathering each of the above pieces of data, assess them visually and programmatically for quality and tidiness issues.

## CLEANING DATA.

The result should be a high quality and tidy master pandas DataFrame (or DataFrames, if appropriate). Again, the issues that satisfy the Project Motivation must be cleaned.

## STORING, ANALYZING, and VISUALIZING DATA.

Store the clean DataFrames in a CSV file with the main one named `twitter_archive_master.csv`. If additional files exist because multiple tables are required for tidiness, name these files appropriately. Additionally, you may store the cleaned data in a SQLite database.

Analyze and visualize your wrangled data in your **wrangle\_act.ipynb** Jupyter Notebook.