

# Udacity

# Data Analytics

Nanodegree Program  
Project: Data Wrangling

# Introduction

Data Wrangling is a crucial step for each data analysis project to have several resources related to the topic you are interested in analyzing and ensure that the insights are valid and clear of any mistakes.

All of these have been done through several steps which will be clarified in the coming pages projected on WeRateDogs Project.

# Data Gathering

Several resources used in this project:

- [Twitter API](#) as the base but I read the data from this file through [json](#) as I had an issue in the Twitter API account
- Comma Separated Value file which is [twitter\\_archive\\_enhanced](#) which I loaded and manipulated using [Pandas](#).
- Image\_Prediction which is obtained through url using [request](#) then saving the file as tab-separated values then reading the file similar to the previous approach.

# Data Assessing

Exploration of the data quickly using the power of pandas gave me quick insights about the issues in the data visually using head, and sample methods or programmatically using describe, info, and value\_counts methods.

The data had lots of quality issues like:-

- null values in several columns and sometimes existed in shape that reflects that no missing data
- Rows should be neglected as it was considered duplicated from a logic perspective like in the case of retweets
- Data type of several columns was wrong like timestamp existed as text
- Inaccurate values for rating needed to be handled.

# Data Assessing

The data had lots of tidiness issues like:-

- Several columns for the type were pivoted.
- Related observations that form one observation unit are separated into three different tables.

# Data Cleaning

An overview approach uses the previously mentioned issues by assessing steps and then defining the issues, followed by the code that solves the issue and testing the effectiveness of the solution.

pandas and numpy were great libraries for handling the data.

Important note that a copy of the files made before any cleaning steps.

# Data Storing

After cleaning the data it is saved for further Exploratory data analysis.

Thank You