

So, Today We are going to complete WeRateDogs project. Have you ever heard about it? No? That's alright! I also heard about it recently. It doesn't matter if you don't know anything about this. There are a lot of sources you can learn about. Do you have a dog? If so, take its picture and tweet about its beautiful, funny or other special habits. WeRateDogs! Next winner may be your dog in the future, who knows!

Wrangle and Analysis (WeRateDogs)

Data Wrangling

Data Wrangling includes Three steps

- **Gathering**
- **Assessing**
- **Cleaning**

So, let's dive deeper into **Data Wrangling** step's concepts

Gathering

- is a process of fetching data from different sources

We are also asked to gather data from at least three sources for this data. The first one given in .csv format, and we need to load this data simply and look at its structure quickly and understand its concepts generally.

The second one is needed to be programmatically downloaded from external source using requests library and must be loaded to our workspace.

Third one requires some more hardwork than others. At first you must sign up to twitter. Then apply to a developer account when your application is approved, you need to create an api and get consumer keys, then you need to use this authorization keys to gather data using python's tweepy. But I can't create a developer account so I used Udacity's handout json file to work on this project.

Assessing

- includes thinking like a detective and quickly spot some issues, and making decisions about them. So, let's dive in!

Assessing is the second step of data wrangling process and it includes two steps to fully complete. Visual assessing and Programmatic assessing.

Visual assessing finding issues and key points using visually, using spreadsheets or so on.

Programmatic assessment is using some programming environments to find the points I mentioned above. I also organized my wrangling process like this pattern which I have been using in this report.

Cleaning

- is a process of doing actions on our observed issues

That means it is time to fix the problems that we spot during assessment process. To do this we must use python's useful packages and algorithms. They are very helpful for us to complete even more complex problems.

At wrangling process I have used this like pattern:

Quick overview about the we are going to do

Gathering process

- 1) I have loaded 'twitter-archive-enhanced.csv' file and quick look through
- 2) I have loaded 'image-predictions.tsv' file and quick look through
- 3) I have loaded 'tweet-json.txt' file and quick look through

Assessing

- 1) First I look at visually for quick spots
- 2) Then I look at programmatically for further investigation

Cleaning

I have cleaned the issues which is arise during assessment process

After cleaning I have save the file to 'twitter-archive-master.csv' file for further use in

Visualization process

The experience was very helpful for me to learn a lot of things, I try to make more clean data in this project.