

# Data wrangling Report

## Intro :

Real-world data rarely comes clean. Using Python and its libraries, you will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. You will document your wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user [@dog\\_rates](#), also known as [WeRateDogs](#). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "[they're good dogs Brent](#)." WeRateDogs has over 4 million followers and has received international media coverage.

## Content :

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

## Gathering data:

There are 3 DataSet in this project :

- Twitter archive file : the `twitter_archive_enhanced.csv` was provided by

Udacity and downloaded manually.

- The tweet image predictions file: (image\_predictions.tsv) is hosted on Udacity.
- Twitter API & JSON: by using the tweet IDs in the WeRateDogs Twitter

## Assessing data :

We can See Data Describe and info about each Data Frame in value\_counts  
separated the issue encountered in quality issues and tidiness issues.

## Cleaning data:

-Make one Dataframe for 3 Files

-Delete 4 Columns 'doggo','floofer','pupper','puppo' and Relace it with one column

-Remove retweets the  
[retweeted\_status\_id ,retweeted\_status\_user\_id ,  
retweeted\_status\_timestamp ]

-Error in Datatype WE Handel it

-handle names in lowerCase and None to nan

## Conclusion :

Data wrangling is a very important part in data analyst .  
I am use Python programming language .

data analyst it is important for making Decision it use in business and and Adv like facebook and twitter .

It is better than excel analyze .

And strong dealing with Big Data .

Can dealing with APIs