# Wrangling Report : WeRateDogs Data Analysis

I wrangled WeeRateDogs Twitter data to create an analyses and visualizations. I will briefly describe the detail of wrangling.

## 1. Gathering data

There are three pieces of data as follows.

* 1. The WeRateDogs twitter archive   
     I manually downloaded from the Udacity website.
  2. The tweet image predictions  
     I downloaded it programmatically using the Requests library from the Udacity’s servers.
  3. Additional data from tweeter

I downloaded additional data from twitter such as retweet and favourite count. By using the tweet IDs in the WeRateDogs twitter archive and query the twitter API by Tweepy library, set of JSON data is downloaded and a pandas DataFrame is created based on this.

## Accessing and cleaning data

* 1. The WeRateDogs twitter archive  
     Missing data: some data relates to reply/retweet are missing, but these are nullable data.
  2. - These retweet/reply should be removed from this table as we want only original ratings. Instead, I will make additional columns of the number of retweet based on twitter\_count.csv.
  3. Dtype:
  4. - Timestamp is not time series data.
  5. - Each dog stage is not boolean.
  6. - Five columns relates to ID are not string/object.