# Wrangle report for Twitter @dogs_rates account

## Introduction:

There is three main phases on the data wrangling process we could illustrate it as the following:

- **Gathering data**: We must identify from where we will get the data ?, what sources ?, what are our needs? , what are the sources extraction mechanism's?
- **Assessing data**: Diving into the data , observing issues , write down these issues on their scope
- **Cleaning data**: Deal with the observed structural and quality issues from phases two by finding the suited ideal solution for them.

We will move into these phases step by step on our project for wrangling @dogs_rates twitter account tweets data.

## Gathering data:

At this phase, we have gather and upload tweets data from 3 sources.

- Twitter archive : already given .csv file to upload
- Image predictions: a .tsv file to download programmatically form a given URL
- Retweets & Likes : must be extracted through twitter API

- **Challenges**:

    - Mapping the columns in the for loop to its own corresponding data from tsv file to write it as csv and save it.
    - Convincing twitter that I'm not harmful and I need access to just extract some data from an account ☺ .
    - Extracting the data from the API before generating the master table.
    - Using Json library to parse the data from twitter API since it's the first time for me.

## Assessing data:

At this phase, We have assessed the gathered tables both visually and programmatically.

Besides the observations made in the notebook, it's worth to mention these main challenges:

- **Challenges**:
  - Detecting and observing by prioritizing the worst quality issues first.
  - Assessing the tables visually well using Jupiter. (Using excel in visual assessment is way much better. It helped me to observe more details)
  - Write the best code to check and assess.

## Cleaning data:

At this phase, We have cleaned the observed issues from phase two and more then finally, combined the three tables in one master table.

We have realize at this phase that more issues observed and appears while cleaning.

- **Challenges:**
  - Finding the best way for merging the four dog types columns in tidiness issue without missing any values.
  - Clean with the best practices upon our resources.