Introduction:

WeRateGogs it is a user account @dog_rates. We basically worked with the data that came from this account. We gather it, analyze it and visualize it. At the begging, the data was not clear and messier. So, we have done three steps to finalize our project, which is gathering the data, assessing the data & clean the data. Let's dive in to explore more.

The three steps:

- Gather the data.
- Assess the data.
- Clean the data.

Gathering data:

We three different data from three different source, twitter archive file from Udacity, image prediction also from Udacity server & json also a file downloaded from Udacity.

Assessing the data:

After gathering the data, it's time to assess it and find out more about it. There is several issues in quality and tidiness such as:

- **Quality:**
    - **1- source need to be enhanced since it is <a /a>**
    - **2- timestamp should be converted to datetime && tweet id to str**
    - **3- Delete 'in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timestamp', since its missing a lot of data**
    - **4- There is many names that is not corret such as (a,as..etc) and it should be renames to non**
    - **5- Drop duplicate image url**
    - **6- Combine p1, p2 & p3 and create dog breed**
    - **7- Rating not always correct**
    - **8- rename id in json to tweet_id**
- **Tidiness:**
    - **1- Dog stage**
    - **2- Create one master sheet**

Cleaning:

This is the part where any data analyst should have a full understanding of the data and start clean it and enhance it. First I start with the quality issues, I choose 8 and I fix it and I create one master sheet to make the data readable ready to do analysis and visualization.

Concluding

It was an exciting experience to work with such a data as this where we gather three data and combining them together then start analysis and visualization.