I’m going to describe my work in the wrangling report. My work is separated into five stages which are Gather, Assess, Clean, Store and Analysis and in each stage there are substages.
The first stage is Gather and, in this stage, I’ve gathered the data from many sources such as a given file, downloaded programmatically using request library and file constructed via API by many libraries such as tweepy, json, pandas, and request. Therefore, the second stage is to Assess and, in this stage, to find the quality and tidiness issues in each data frame. However, I’ve found each issue using the visual way and the programmatic way. I’ve using some methods from pandas in the programmatic way such as .value_counts(), .info() , .describe() , .sample() , .shape , .duplicated() , .head() , .tail() so on. The issues that I’ve found in each table are many. The first issues type is quality. So in the twitter_archive_enhanced table, I’ve discovered a lot of issues such as:
- Erroneous datatypes in these two columns (tweet_id and timestamp columns).
- there are some irrelevant columns.
- Erroneous in the rating_nominator values.
- change the values that showing as 'None' into ' NaN'.
- Display full content of 'text' column.
- make the source column more readable.

image_predictions table:
- Column names are not informative.
- capitalized each first litter in the breed of dogs and removes the underscore.
- Erroneous datatypes (tweet_id).

tweet_count table:
- Erroneous datatypes (tweet_id).

And The second issues type is tidiness which we have two issue the first one is about combining the three different data set into one data frame. The second one is combine the

different columns type of dogs [doggo, floofer, pupper and puppo ] into one column.
Moreover, the third stage is Cleaning and it’s separated into 3 substages which they are define: what we are going to do ,code: translate what we have said in the defining stage onto code and test: check that the code has been done. However in this stage, I’ve solved the second tidiness issue that is about the dog's type, then I dived in solving each quality issues I’ve found in the Assess stage. The last thing, I combined all the 3 sets into one master set and then store it as a CSV file and then analyze it.