In this project I did data wrangling process in several parts:

A - Data gathering

Data was gathered from three sources in three different ways.,

1) Udacity gave direct link of twitter archives file for the tweets of @WeRateDogs in csv format, I easily downloaded it manually.

2) image-predictions.tsv file was downloaded programmatically by using Python requests library to get a tsv (tab separated values) file consisting of image predictions of the dog breed for each of the tweets of the first data set.

3) Download by querying the Twitter API using a Python library . To use twitter API I created twitter developer account and followed instruction to use twitter API and copy required keys. After several hours, I finally download tweet's JSON data which includes the favorites and counts.

B- Data Assessment and cleaning

After gathering all of these three datas and open them in the data frame,I assessed them both manually by looking columns from the dataframe and programmatically. I have identified and documented several issues with data quality and tidiness. I tried to fix them programmatically to have a clean data. I always find an issue and start to fix it, because it is an iterative process. Finally, I used inner merging on the common column for three cleaned dataframe to give a result dataframe. This dataframe was also looked for tidiness and quality issues and was cleaned especially to avoid and redundancy in columns. I saved this final result dataframe in a csv format file.

The final cleaned result file was open in a result dataframe and used for visualization and data analysis.