Introduction:

This Wangle and Analyze Data Project is part of the Udacity Data Analyst Nanodegree program, the data set is from the twitter account WeRateDogs, as know as: @dog_rates, which provides information about dogs. The project is divided into three parts: Gathering, Assessing, and Cleaning.

Gathering:

The data sets are gathered from the files named:

twitter-archive-enhanced.csv

image_predictions.tsv

Assessing:

The following methods are employed to assess the data:

.info()

.value_counts()

.head()

.sample()

Cleaning:

The following are tidiness issues and quality issues found in the data set, which need to be cleaned.

Tidiness issue: Mergering data sets and simplifying dog stages;

Quality issue 1: Removing retweets, setting 'tweet_id' as string and 'timestamp' as datetime form;

Quality issue 2: Inaccuracy of names and ratings;

Quality issue 3: Renaming some columns and deleting some irrelevant ones.