

Wrangle Report

This project is the final project of Udacity Data Analyst program. The aim of this project is to wrangle and analyse WeRateDogs Twitter data. During performing those tasks, skills concerning gathering, assessing and analysing data would be increased dramatically. These skills are really necessary for people who want to seek for jobs in data analyst field.

In order to perform this project, first step is to gather data from three different sources: Twitter archive, image predictions and API data. Udacity provided WeRateDogs Twitter archive file and code were written to import this file to Jupiter notebook. The tweet image predictions file was hosted by Udacity's servers and need to be downloaded and imported by Request method. API data is stored in Python's Tweepy library and it can be downloaded by using Tweepy to query Twitter's API for additional data beyond the data included in the WeRateDogs Twitter archive. Because my computer can not run those API code provided, I need to import the file given from Udacity and import it to Jupiter.

After all data files have been added, data need to be assessed visually and programmatically for quality and tidiness issues. Eight quality issues and three tidiness issues have been found to be cleaned. Quality issues concern erroneous datatypes and incorrect values in some columns. Tidiness issue is about making the data to be easier to read and analyse.

Next step is to solve issues that have been found in assessing step. First step is to merge three data files into one dataframe. Then, many pandas methods, python loop and functions are used to address each issue. One issue is demonstrated by three parts: define, code and check. After all issues have been solved, the dataframe is saved in csv format.

Finally step is to analyse data and visualize data using many kinds of charts such as bar chart, line chart etc.