

# wrangle\_report

March 3, 2020

Wrangling report

## 0.1 Introduction:

This project is about Data Wrangling of tweeter account WeRateDogs. This account takes pictures of dogs posted on tweeter and they rate them with fancy ratings for fun. People react with these tweets and they like and retweet them. The provided data for this account came from many sources and they the data is not clean and needs to be prepared in order to be useful in analysis process.

## 0.2 Processes:

The processes included in this project such as gathering data, assessing data, cleaning data, storing data. (analyzing and visualization will be mentioned in a separate report) Brief description of each process:

### 0.2.1 Gathering data:

In this process we gathered data from different sources and in different formats such as CSV, TSV and JSON.

### 0.2.2 Assessing Data:

In this process we assessed our data using visual assessment by just skimming it in a program like spread sheet or pandas data frame and we made also programmatic assessment which we relied heavily on it using methods in pandas libraries like: `.info`, `.describe`, `.value_counts` and many more. We also categorized the issues as tidiness issues or quality issues.

### 0.2.3 Cleaning data:

after we categorized the issues as tidiness issues or quality issues in the previous process, in this process we cleaned the data in order to be ready for analysis. This process included many operations such as deleting & replacing values/rows/columns, converting data to the correct types for analyzing, restructuring and correcting the tidiness issues. In addition, using pandas cleaning methods such as: `replace`, `drop_duplicates`, `drop columns` or `rows`, `renaming columns` to be more descriptive and finally merging tables into one table to make them ready for analysis. Storing data: in this process we save the final result of our efforts and work of previous process, we save cleaned data frame(s) in files, to avoid running through the cleaning process each time.

```
In [ ]:
```