Data Wrangling Report

Introduction:

The process divided into 3 main steps:

1. Gathering the data
2. Assessing
3. Cleaning

1. Gathering Data

About the dataset ---->

The dataset I'll be wrangling is the tweet archive of Twitter user @dog_rates (https://twitter.com/dog_rates), also known as WeRateDogs. This archive/dataset consists of 2356 basic tweet data from November, 2015 to August, 2017.(reference Udacity)

Also there are lots of data about dog's images and its very useful the image predictions (the top three only) alongside each tweet ID, image URL, and the image number that corresponded to the most confident prediction.

There are 3 main csv file in our project:

1. **Twitter archive CSV file**
2. **tweet image predictions**
3. **Twitter API**

2. Assessing Data:

Each dataframe was assessed visually using the libraries like pandas so we used some function like head, tail, info, unique like that. Then finding the duplicate values and missing values then our main goal is that look for the quality and tidiness issues and understand the data.

3.Cleaning Data

The cleaning though was definitely the most time consuming part of wrangling the data. It was so surprising how much could had to be adjusted and how some of the issues really had to be looked at closely.

Cleaning of the data was necessary because resolving the tidiness issues.

Thank you.