# Date Wrangling Report

## Introduction

## WeRateDogs is a Twitter account that rates people's dogs with humorous comments about the dog. These ratings almost always have a denominator of 10, while the numerators can be greater than 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Sometimes, they mention the breed of the dog in their tweets, and other times, they don’t. WeRateDogs has over 4 million followers and has received international media coverage.

## Project objectives

The main objectives of this project are to:

* Gather data from 3 different sources.
* Perform data wrangling (gathering, assessing and cleaning) on provided data sets
* Store, analyze, and visualize the wrangled data.
* Reporting on the data wrangling efforts, analysis and visualizations.

## Data Gathering

The dataset used in this project is an archive of all tweets from the Twitter account @dog\_rates which was provided by Udacity. This archive contains basic tweet data for over 2000 tweets up till August 1, 2017. The Data Gathering process is comprises of 3 methods of gathering data:

* **WeRateDogs Twitter archive:** This 'twitter-archive-enhanced.csv' data was downloaded manually and loaded into a dataframe.
* **Tweet image predictions:** The ‘image\_predictions.tsv’ was downloaded programmatically using Python Requests library.
* **Additional data from the Twitter API**: Each tweet's retweet count and favorites count are supposed to be queried from Twitter's API into a ‘tweet\_json.txt’ file. However, due to my Twitter Developer account not being verified in time, I downloaded a JSON file containing the tweets which was provided by Udacity programmatically.

## Assessment and Data Cleaning

In order to make meaningful analysis out of the data, the data was first assessed and cleaned. In this phase, a number of observations were made. The observations and actions taken to clean the data is detailed in the table below: