**Wrangle and Analyze data**

**Wrangle report**

This project is about wrangling and analyzing the tweet archive of Twitter account “@dog\_rates”, also

known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous

comment about the dog. These ratings almost always have a denominator of 10. The numerators,

though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs

Brent." This was a quote from @dog\_rates account during a celebrated exchange in which the account

shut down a person taking issue with their rating system. WeRateDogs has over 4 million followers and

has received international media coverage.

In this project, I was challenged to wrangle and analyze a very messy data frame. Starting from gathering

data using three different ways, one where the file was a given and easily downloaded, second was

downloading the file programmatically from Udacity servers using Requests library, and finally the third

by using Twitter API by Tweepy library (Although in my project, I couldn’t get validation from Twitter to

use the API, so I just downloaded the file from Udacity). After gathering the data, I started to assess it,

where I should document any issues I can find visually and programmatically. After documenting the

issues, I started the cleaning process, where I will be fixing all (or most) of the issues that I documented

it in the assessing step.

Now I’ll briefly describe the process of wrangling:

**1- Gathering Data**

Depending on the source of your data, and what format it's in, the steps in gathering data

vary.

The high-level gathering process:

• **Obtaining data** (downloading a file from the internet, scraping a web page, querying an API,

etc.)

**• importing that data into your programming environment** (e.g. Jupyter Notebook)

**2- Assessing Data**

There are two types of issues you are looking for:

• **Quality**: Issues with content.

• **Tidiness**: Issues with structure that prevent easy analysis.

and you can assess by:

• **Visual assessment**: Scrolling through the data.

• **Programmatic assessment**: Using code to view specific portions and summaries of the data.

**3- Cleaning Data**

You can clean:

**• Manually**

**• Programmatically**

o **Define**: Convert our assessments into defined cleaning tasks.

o **Code**: Convert those definitions to code.

o **Test**: Test the dataset.