

## Wrangle Report

### *Internal Document*

My wrangling efforts started with gathering all the available data required for the project.

Getting the twitter-archive-enhanced.csv was the easy part of the process as the file was provided and all I had to do was download the file to my computer and then move it to my required directory using terminal on my Macbook.

Second I had to get the image\_predictions.tsv file and I had to get it programmatically for the project. I eventually struggled with this as I was using my Windows 10 machine and kept getting permission errors from Jupyter Notebook and Windows 10 when downloading the file programmatically. I then resorted to moving to using my Macbook and managed to easily download the file programmatically.

Getting the data from the twitter API was a bit difficult especially using JSON. I managed to strike it luck when I read the knowledge area on Udacity and found the way to download the data.

I stored my data for each dataset as follow:

```
Df_1 = twitter-archive-enhanced.csv
```

```
Df_2 = image-predictions.tsv
```

```
df_3 = pd.DataFrame(columns=['tweet_id', 'retweet_count', 'favorite_count'])
```

After I managed to successfully gather all my data I checked the data visually and noting any issues I can find. I then started “looking” programmatically by using the following functions mainly:

- df.head()
- df.tail
- df.info()
- Df.value\_counts and specifically using value\_counts on my columns to determine the values contained in each column

I did this for each dataset and making a list of problems and then moving on to the next dataset.

At the end of the **Gathering Data** section I created a list for each dataset with issues in the data that i have observed.

After that I created a copy of each dataset which I named \_clean for each dataset.

I then proceeded to list the heading of the dataset I will be working on and then listing what I am going to do in the dataset.

I then listed in a heading what is to be done and then proceeded with the code. After coding each block I tested my code to determine if what I did worked.