

wrangle_act

April 22, 2020

```
In [481]: import pandas as pd
import numpy as np
import datetime
import time
import json
import os
import requests
import string
import tweepy
```

1 Gathering Data

- 1) Manually uploading the file twitter_archive_enhanced.csv
- 2) Programmatically downloading the file image_predictions.tsv from the Udacity Servers
- 3) Using the Tweepy library to gather tweet data in JSON format tweet_json.txt, finally reading this text file data into a pandas dataframe

```
In [482]: # Step1
twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
```

```
In [483]: # Step2
# Get data
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions.tsv'
response = requests.get(url)
# Create file
with open('image_predictions.tsv', 'wb') as file:
    file.write(response.content)
```

```
In [484]: img_pred = pd.read_csv('image_predictions.tsv', sep='\t')
```

2 Step3 - Authentication

```
consumer_key = 'HIDDEN' consumer_secret = 'HIDDEN' access_token = 'HIDDEN' access_secret = 'HIDDEN'
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth)
```