wrangle_act

November 9, 2022

1 Project: Wrangling and Analyze Data

1.1 Data Gathering

In the cell below, gather **all** three pieces of data for this project and load them in the notebook. **Note:** the methods required to gather each data are different. 1. Directly download the WeRateDogs Twitter archive data (twitter_archive_enhanced.csv)

```
In [88]: import pandas as pd
         import numpy as np
         import requests
         import os
         import tweepy
         from tweepy import OAuthHandler
         import json
         from timeit import default_timer as timer
         import warnings
         warnings.filterwarnings('ignore')
         import seaborn as sb
         import matplotlib.pyplot as plt
         %matplotlib inline

In [89]: #Read the csv file into the dataframe
         df_tweet= pd.read_csv('twitter-archive-enhanced.csv')

In [90]: list(df_tweet)

Out[90]: ['tweet_id',
          'in_reply_to_status_id',
          'in_reply_to_user_id',
          'timestamp',
          'source',
          'text',
          'retweeted_status_id',
          'retweeted_status_user_id',
          'retweeted_status_timestamp',
          'expanded_urls',
          'rating_numerator',
```

```
 'rating_denominator',
 'name',
 'doggo',
 'floofer',
 'pupper',
 'puppo']
```

2. Use the Requests library to download the tweet image prediction (image_predictions.tsv)

```
In [91]: #this is the folder where the file for image_predictions.tsv will be downloaded
         url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predicti
         response = requests.get(url)
         response

Out[91]: <Response [200]>

In [92]: with open('image-prediction.tsv', 'wb') as file:
             file.write(response.content)

In [93]: #To read the image predictions table into a dataframe
         image = pd.read_csv('image-predictions.tsv', sep='\t')

In [94]: os.listdir()

Out[94]: ['image-prediction.tsv',
          'twitter-archive-enhanced.csv',
          'image',
          'act_report.ipynb',
          'tweet_json.txt',
          'image-predictions.tsv',
          '.ipynb_checkpoints',
          'wrangle_report.ipynb',
          'wrangle_act.ipynb']
```

3. Use the Tweepy library to query additional data via the Twitter API (tweet_json.txt)

```
In [9]: # Query Twitter API for each tweet in the Twitter archive and save JSON in a text file
        # These are hidden to comply with Twitter's API terms and conditions
        consumer_key = 'HIDDEN'
        consumer_secret ='HIDDEN'
        access_token = 'HIDDEN'
        access_secret = 'HIDDEN'
        auth = OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_secret)

        api = tweepy.API(auth, wait_on_rate_limit=True)

        # Tweet IDs for which to gather additional data via Twitter's API
        tweet_ids = df_tweet.tweet_id.values
```