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 WeRate-Dogs Twitter archive data (twitter_archive_enhanced.csv)

```python
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
```

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

```python
In [90]: list(df_tweet)
```

```python
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
```

```python
        len(tweet_ids)

        # Query Twitter's API for JSON data for each tweet ID in the Twitter archive
        count = 0
        fails_dict = {}
        start = timer()
        # Save each tweet's returned JSON as a new line in a .txt file
        with open('tweet_json.txt', 'w') as outfile:
            # This loop will likely take 20-30 minutes to run because of Twitter's rate limit
            for tweet_id in tweet_ids:
                count += 1
                print(str(count) + ": " + str(tweet_id))
                try:
                    tweet = api.get_status(tweet_id, tweet_mode='extended')
                    print("Success")
                    json.dump(tweet._json, outfile)
                    outfile.write('\n')
                except tweepy.TweepError as e:
                    print("Fail")
                    fails_dict[tweet_id] = e
                    pass
        end = timer()
        print(end - start)
        print(fails_dict)

1: 892420643555336193
Success
2: 892177421306343426
Success
3: 891815181378084864
Success
4: 891689557279858688
Success
5: 891327558926688256
Success
6: 891087950875897856
Success
7: 890971913173991426
Success
8: 890729181411237888
Success
9: 890609185150312448
Success
10: 890240255349198849
Success
11: 890006608113172480
Success
12: 889880896479866881
```

```
Success
13: 889665388333682689
Success
14: 889638837579907072
Success
15: 889531135344209921
Success
16: 889278841981685760
Success
17: 888917238123831296
Success
18: 888804989199671297
Success
19: 888554962724278272
Success
20: 888202515573088257
Fail
21: 888078434458587136
Success
22: 887705289381826560
Success
23: 887517139158093824
Success
24: 887473957103951883
Success
25: 887343217045368832
Success
26: 887101392804085760
Success
27: 886983233522544640
Success
28: 886736880519319552
Success
29: 886680336477933568
Success
30: 886366144734445568
Success
31: 886267009285017600
Success
32: 886258384151887873
Success
33: 886054160059072513
Success
34: 885984800019947520
Success
35: 885528943205470208
Success
36: 885518971528720385
```

```
Success
37: 885311592912609280
Success
38: 885167619883638784
Success
39: 884925521741709313
Success
40: 884876753390489601
Success
41: 884562892145688576
Success
42: 884441805382717440
Success
43: 884247878851493888
Success
44: 884162670584377345
Success
45: 883838122936631299
Success
46: 883482846933004288
Success
47: 883360690899218434
Success
48: 883117836046086144
Success
49: 882992080364220416
Success
50: 882762694511734784
Success
51: 882627270321602560
Success
52: 882268110199369728
Success
53: 882045870035918850
Success
54: 881906580714921986
Success
55: 881666595344535552
Success
56: 881633300179243008
Success
57: 881536004380872706
Success
58: 881268444196462592
Success
59: 880935762899988482
Success
60: 880872448815771648
```

```
Success
61: 880465832366813184
Success
62: 880221127280381952
Success
63: 880095782870896641
Success
64: 879862464715927552
Success
65: 879674319642796034
Success
66: 879492040517615616
Success
67: 879415818425184262
Success
68: 879376492567855104
Success
69: 879130579576475649
Success
70: 879050749262655488
Success
71: 879008229531029506
Success
72: 878776093423087618
Success
73: 878604707211726852
Success
74: 878404777348136964
Success
75: 878316110768087041
Success
76: 878281511006478336
Success
77: 878057613040115712
Success
78: 877736472329191424
Success
79: 877611172832227328
Success
80: 877556246731214848
Success
81: 877316821321428993
Success
82: 877201837425926144
Success
83: 876838120628539392
Success
84: 876537666061221889
```

Success
85: 876484053909872640
Success
86: 876120275196170240
Success
87: 875747767867523072
Success
88: 875144289856114688
Success
89: 875097192612077568
Success
90: 875021211251597312
Success
91: 874680097055178752
Success
92: 874434818259525634
Success
93: 874296783580663808
Success
94: 874057562936811520
Success
95: 874012996292530176
Success
96: 873697596434513921
Fail
97: 873580283840344065
Success
98: 873337748698140672
Success
99: 873213775632977920
Success
100: 872967104147763200
Success
101: 872820683541237760
Success
102: 872668790621863937
Fail
103: 872620804844003328
Success
104: 872486979161796608
Success
105: 872261713294495745
Fail
106: 872122724285648897
Success
107: 871879754684805121
Success
108: 871762521631449091

```
Success
109: 871515927908634625
Success
110: 871166179821445120
Success
111: 871102520638267392
Success
112: 871032628920680449
Success
113: 870804317367881728
Success
114: 870726314365509632
Success
115: 870656317836468226
Success
116: 870374049280663552
Success
117: 870308999962521604
Success
118: 870063196459192321
Success
119: 869988702071779329
Fail
120: 869772420881756160
Success
121: 869702957897576449
Success
122: 869596645499047938
Success
123: 869227993411051520
Success
124: 868880397819494401
Success
125: 868639477480148993
Success
126: 868622495443632128
Success
127: 868552278524837888
Success
128: 867900495410671616
Success
129: 867774946302451713
Success
130: 867421006826221569
Success
131: 867072653475098625
Success
132: 867051520902168576
```

```
Success
133: 866816280283807744
Fail
134: 866720684873056260
Success
135: 866686824827068416
Success
136: 866450705531457537
Success
137: 866334964761202691
Success
138: 866094527597207552
Success
139: 865718153858494464
Success
140: 865359393868664832
Success
141: 865006731092295680
Success
142: 864873206498414592
Success
143: 864279568663928832
Success
144: 864197398364647424
Success
145: 863907417377173506
Success
146: 863553081350529029
Success
147: 863471782782697472
Success
148: 863432100342583297
Success
149: 863427515083354112
Success
150: 863079547188785154
Success
151: 863062471531167744
Success
152: 862831371563274240
Success
153: 862722525377298433
Success
154: 862457590147678208
Success
155: 862096992088072192
Success
156: 861769973181624320
```

```
Fail
157: 861383897657036800
Success
158: 861288531465048066
Success
159: 861005113778896900
Success
160: 860981674716409858
Success
161: 860924035999428608
Success
162: 860563773140209665
Success
163: 860524505164394496
Success
164: 860276583193509888
Success
165: 860184849394610176
Success
166: 860177593139703809
Success
167: 859924526012018688
Success
168: 859851578198683649
Success
169: 859607811541651456
Success
170: 859196978902773760
Success
171: 859074603037188101
Success
172: 858860390427611136
Success
173: 858843525470990336
Success
174: 858471635011153920
Success
175: 858107933456039936
Success
176: 857989990357356544
Success
177: 857746408056729600
Success
178: 857393404942143489
Success
179: 857263160327368704
Success
180: 857214891891077121
```

```
Success
181: 857062103051644929
Success
182: 857029823797047296
Success
183: 856602993587888130
Fail
184: 856543823941562368
Success
185: 856526610513747968
Success
186: 856330835276025856
Fail
187: 856288084350160898
Success
188: 856282028240666624
Success
189: 855862651834028034
Success
190: 855860136149123072
Success
191: 855857698524602368
Success
192: 855851453814013952
Success
193: 855818117272018944
Success
194: 855459453768019968
Success
195: 855245323840757760
Success
196: 855138241867124737
Success
197: 854732716440526848
Success
198: 854482394044301312
Success
199: 854365224396361728
Success
200: 854120357044912130
Success
201: 854010172552949760
Success
202: 853760880890318849
Success
203: 853639147608842240
Success
204: 853299958564483072
```

Success
205: 852936405516943360
Success
206: 852912242202992640
Success
207: 852672615818899456
Success
208: 852553447878664193
Success
209: 852311364735569921
Success
210: 852226086759018497
Success
211: 852189679701164033
Success
212: 851953902622658560
Fail
213: 851861385021730816
Fail
214: 851591660324737024
Success
215: 851464819735769094
Success
216: 851224888060895234
Success
217: 850753642995093505
Success
218: 850380195714523136
Success
219: 850333567704068097
Success
220: 850145622816686080
Success
221: 850019790995546112
Success
222: 849776966551130114
Success
223: 849668094696017920
Success
224: 849412302885593088
Success
225: 849336543269576704
Success
226: 849051919805034497
Success
227: 848690551926992896
Success
228: 848324959059550208

```
Success
229: 848213670039564288
Success
230: 848212111729840128
Success
231: 847978865427394560
Success
232: 847971574464610304
Success
233: 847962785489326080
Success
234: 847842811428974592
Success
235: 847617282490613760
Success
236: 847606175596138505
Success
237: 847251039262605312
Success
238: 847157206088847362
Success
239: 847116187444137987
Success
240: 846874817362120707
Success
241: 846514051647705089
Success
242: 846505985330044928
Success
243: 846153765933735936
Success
244: 846139713627017216
Success
245: 846042936437604353
Success
246: 845812042753855489
Success
247: 845677943972139009
Success
248: 845459076796616705
Fail
249: 845397057150107648
Success
250: 845306882940190720
Success
251: 845098359547420673
Success
252: 844979544864018432
```

```
Success
253: 844973813909606400
Success
254: 844704788403113984
Fail
255: 844580511645339650
Success
256: 844223788422217728
Success
257: 843981021012017153
Success
258: 843856843873095681
Success
259: 843604394117681152
Success
260: 843235543001513987
Success
261: 842892208864923648
Fail
262: 842846295480000512
Success
263: 842765311967449089
Success
264: 842535590457499648
Success
265: 842163532590374912
Success
266: 842115215311396866
Success
267: 841833993020538882
Success
268: 841680585030541313
Success
269: 841439858740625411
Success
270: 841320156043304961
Success
271: 841314665196081154
Success
272: 841077006473256960
Success
273: 840761248237133825
Success
274: 840728873075638272
Success
275: 840698636975636481
Success
276: 840696689258311684
```

```
Success
277: 840632337062862849
Success
278: 840370681858686976
Success
279: 840268004936019968
Success
280: 839990271299457024
Success
281: 839549326359670784
Success
282: 839290600511926273
Success
283: 839239871831150596
Success
284: 838952994649550848
Success
285: 838921590096166913
Success
286: 838916489579200512
Success
287: 838831947270979586
Success
288: 838561493054533637
Success
289: 838476387338051585
Success
290: 838201503651401729
Success
291: 838150277551247360
Success
292: 838085839343206401
Success
293: 838083903487373313
Success
294: 837820167694528512
Success
295: 837482249356513284
Success
296: 837471256429613056
Success
297: 837366284874571778
Fail
298: 837110210464448512
Success
299: 837012587749474308
Fail
300: 836989968035819520
```

```
Success
301: 836753516572119041
Success
302: 836677758902222849
Success
303: 836648853927522308
Success
304: 836397794269200385
Success
305: 836380477523124226
Success
306: 836260088725786625
Success
307: 836001077879255040
Success
308: 835685285446955009
Success
309: 835574547218894849
Success
310: 835536468978302976
Success
311: 835309094223372289
Success
312: 835297930240217089
Success
313: 835264098648616962
Success
314: 835246439529840640
Success
315: 835172783151792128
Success
316: 835152434251116546
Success
317: 834931633769889797
Success
318: 834786237630337024
Success
319: 834574053763584002
Success
320: 834477809192075265
Success
321: 834458053273591808
Success
322: 834209720923721728
Success
323: 834167344700198914
Success
324: 834089966724603904
```

```
Success
325: 834086379323871233
Success
326: 833863086058651648
Success
327: 833826103416520705
Success
328: 833732339549220864
Success
329: 833722901757046785
Success
330: 833479644947025920
Success
331: 833124694597443584
Success
332: 832998151111966721
Success
333: 832769181346996225
Success
334: 832757312314028032
Success
335: 832682457690300417
Success
336: 832645525019123713
Success
337: 832636094638288896
Success
338: 832397543355072512
Success
339: 832369877331693569
Success
340: 832273440279240704
Success
341: 832215909146226688
Success
342: 832215726631055365
Success
343: 832088576586297345
Success
344: 832040443403784192
Success
345: 832032802820481025
Success
346: 831939777352105988
Success
347: 831926988323639298
Success
348: 831911600680497154
```

Success
349: 831670449226514432
Success
350: 831650051525054464
Success
351: 831552930092285952
Success
352: 831322785565769729
Success
353: 831315979191906304
Success
354: 831309418084069378
Success
355: 831262627380748289
Success
356: 830956169170665475
Success
357: 830583320585068544
Success
358: 830173239259324417
Success
359: 830097400375152640
Success
360: 829878982036299777
Success
361: 829861396166877184
Success
362: 829501995190984704
Success
363: 829449946868879360
Success
364: 829374341691346946
Fail
365: 829141528400556032
Success
366: 829011960981237760
Success
367: 828801551087042563
Success
368: 828770345708580865
Success
369: 828708714936930305
Success
370: 828650029636317184
Success
371: 828409743546925057
Success
372: 828408677031882754

```
Success
373: 828381636999917570
Success
374: 828376505180889089
Success
375: 828372645993398273
Success
376: 828361771580813312
Success
377: 828046555563323392
Success
378: 828011680017821696
Success
379: 827933404142436356
Success
380: 827653905312006145
Success
381: 827600520311402496
Success
382: 827324948884643840
Success
383: 827228250799742977
Fail
384: 827199976799354881
Success
385: 826958653328592898
Success
386: 826848821049180160
Success
387: 826615380357632002
Success
388: 826598799820865537
Success
389: 826598365270007810
Success
390: 826476773533745153
Success
391: 826240494070030336
Success
392: 826204788643753985
Success
393: 826115272272650244
Success
394: 825876512159186944
Success
395: 825829644528148480
Success
396: 825535076884762624
```

```
Success
397: 825147591692263424
Success
398: 825120256414846976
Success
399: 825026590719483904
Success
400: 824796380199809024
Success
401: 824775126675836928
Success
402: 824663926340194305
Success
403: 824325613288833024
Success
404: 824297048279236611
Success
405: 824025158776213504
Success
406: 823939628516474880
Success
407: 823719002937630720
Success
408: 823699002998870016
Success
409: 823581115634085888
Success
410: 823333489516937216
Success
411: 823322678127919110
Success
412: 823269594223824897
Success
413: 822975315408461824
Success
414: 822872901745569793
Success
415: 822859134160621569
Success
416: 822647212903690241
Success
417: 822610361945911296
Success
418: 822489057087389700
Success
419: 822462944365645825
Success
420: 822244816520155136
```

```
Success
421: 822163064745328640
Success
422: 821886076407029760
Success
423: 821813639212650496
Success
424: 821765923262631936
Success
425: 821522889702862852
Success
426: 821421320206483457
Success
427: 821407182352777218
Success
428: 821153421864615936
Success
429: 821149554670182400
Success
430: 821107785811234820
Success
431: 821044531881721856
Success
432: 820837357901512704
Success
433: 820749716845686786
Success
434: 820690176645140481
Success
435: 820494788566847489
Success
436: 820446719150292993
Success
437: 820314633777061888
Success
438: 820078625395449857
Success
439: 820013781606658049
Success
440: 819952236453363712
Success
441: 819924195358416896
Success
442: 819711362133872643
Success
443: 819588359383371776
Success
444: 819347104292290561
```

```
Success
445: 819238181065359361
Success
446: 819227688460238848
Success
447: 819015337530290176
Success
448: 819015331746349057
Success
449: 819006400881917954
Success
450: 819004803107983360
Success
451: 818646164899774465
Success
452: 818627210458333184
Success
453: 818614493328580609
Success
454: 818588835076603904
Success
455: 818536468981415936
Success
456: 818307523543449600
Success
457: 818259473185828864
Success
458: 818145370475810820
Success
459: 817908911860748288
Success
460: 817827839487737858
Success
461: 817777686764523521
Success
462: 817536400337801217
Success
463: 817502432452313088
Success
464: 817423860136083457
Success
465: 817415592588222464
Success
466: 817181837579653120
Success
467: 817171292965273600
Success
468: 817120970343411712
```

```
Success
469: 817056546584727552
Success
470: 816829038950027264
Success
471: 816816676327063552
Success
472: 816697700272001025
Success
473: 816450570814898180
Success
474: 816336735214911488
Success
475: 816091915477250048
Success
476: 816062466425819140
Success
477: 816014286006976512
Success
478: 815990720817401858
Success
479: 815966073409433600
Success
480: 815745968457060357
Success
481: 815736392542261248
Success
482: 815639385530101762
Success
483: 815390420867969024
Success
484: 814986499976527872
Success
485: 814638523311648768
Success
486: 814578408554463233
Success
487: 814530161257443328
Success
488: 814153002265309185
Success
489: 813944609378369540
Success
490: 813910438903693312
Success
491: 813812741911748608
Success
492: 813800681631023104
```

```
Success
493: 813217897535406080
Success
494: 813202720496779264
Success
495: 813187593374461952
Success
496: 813172488309972993
Success
497: 813157409116065792
Success
498: 813142292504645637
Success
499: 813130366689148928
Success
500: 813127251579564032
Success
501: 813112105746448384
Success
502: 813096984823349248
Success
503: 813081950185472002
Success
504: 813066809284972545
Success
505: 813051746834595840
Success
506: 812781120811126785
Success
507: 812747805718642688
Fail
508: 812709060537683968
Success
509: 812503143955202048
Success
510: 812466873996607488
Success
511: 812372279581671427
Success
512: 811985624773361665
Success
513: 811744202451197953
Success
514: 811647686436880384
Success
515: 811627233043480576
Success
516: 811386762094317568
```

```
Success
517:  810984652412424192
Success
518:  810896069567610880
Success
519:  810657578271330305
Success
520:  810284430598270976
Success
521:  810254108431155201
Success
522:  809920764300447744
Success
523:  809808892968534016
Success
524:  809448704142938112
Success
525:  809220051211603969
Success
526:  809084759137812480
Success
527:  808838249661788160
Success
528:  808733504066486276
Success
529:  808501579447930884
Success
530:  808344865868283904
Success
531:  808134635716833280
Success
532:  808106460588765185
Success
533:  808001312164028416
Success
534:  807621403335917568
Success
535:  807106840509214720
Success
536:  807059379405148160
Success
537:  807010152071229440
Success
538:  806629075125202948
Success
539:  806620845233815552
Success
540:  806576416489959424
```