# wrangle_act

November 17, 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
[870]: # importing the necessary library packages
       import numpy as np
       import pandas as pd
       import requests
       import tweepy
       import os
       import json
       import datetime
       import matplotlib.pyplot as plt
```

```python
[871]: df =  pd.read_csv('twitter-archive-enhanced.csv')
       df.sample(5)
```

```
[871]:               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
       836    767884188863397888 NaN                    NaN
       477    815990720817401858 NaN                    NaN
       64     879674319642796034  8.795538e+17            3.105441e+09
       2159   669571471778410496 NaN                    NaN
       705    785639753186217984 NaN                    NaN


                            timestamp  \
       836    2016-08-23 00:40:31 +0000
       477    2017-01-02 18:38:42 +0000
       64     2017-06-27 12:14:36 +0000
       2159   2015-11-25 17:40:53 +0000
       705    2016-10-11 00:34:48 +0000


         source  \
       836    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
```

```
477   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
64    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2159  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
705   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>

          text   \
836    This is Tito. He's on the lookout. Nobody knows for what. 10/10
https://t.co/Qai481H6RA
477    Meet Jack. He's one of the rare doggos that doesn't mind baths. 11/10
click the link to see how you can help Jack!\n\nhttps://t.co/r4W111FzAq
https://t.co/fQpYuMKG3p
64    @RealKentMurphy 14/10 confirmed
2159  This is Keith. He's had 13 DUIs. 7/10 that's too many Keith
https://t.co/fa7olwrF9Y
705    This is Pinot. He's a sophisticated doggo. You can tell by the hat. Also
pointier than your average pupper. Still 10/10 would pet cautiously
https://t.co/f2wmLZTPHd

      retweeted_status_id   retweeted_status_user_id  \
836  NaN                    NaN
477  NaN                    NaN
64   NaN                    NaN
2159 NaN                    NaN
705  NaN                    NaN

      retweeted_status_timestamp  \
836    NaN
477    NaN
64     NaN
2159   NaN
705    NaN

        expanded_urls  \
836   https://twitter.com/dog_rates/status/767884188863397888/photo/1,https://tw
itter.com/dog_rates/status/767884188863397888/photo/1,https://twitter.com/dog_ra
tes/status/767884188863397888/photo/1,https://twitter.com/dog_rates/status/76788
4188863397888/photo/1
477    https://www.gofundme.com/surgeryforjacktheminpin,https://twitter.com/dog_r
ates/status/815990720817401858/photo/1
64     NaN
2159  https://twitter.com/dog_rates/status/669571471778410496/photo/1
705    https://twitter.com/dog_rates/status/785639753186217984/photo/1,https://tw
itter.com/dog_rates/status/785639753186217984/photo/1
```

|      | rating_numerator | rating_denominator | name  | doggo | floofer | pupper | puppo |
|------|------------------|--------------------|-------|-------|---------|--------|-------|
| 836  | 10               | 10                 | Tito  | None  | None    | None   | None  |
| 477  | 11               | 10                 | Jack  | None  | None    | None   | None  |
| 64   | 14               | 10                 | None  | None  | None    | None   | None  |
| 2159 | 7                | 10                 | Keith | None  | None    | None   | None  |
| 705  | 10               | 10                 | Pinot | doggo | None    | pupper | None  |

[872]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   tweet_id                   2356 non-null   int64
 1   in_reply_to_status_id      78 non-null     float64
 2   in_reply_to_user_id        78 non-null     float64
 3   timestamp                  2356 non-null   object
 4   source                     2356 non-null   object
 5   text                       2356 non-null   object
 6   retweeted_status_id        181 non-null    float64
 7   retweeted_status_user_id   181 non-null    float64
 8   retweeted_status_timestamp 181 non-null    object
 9   expanded_urls              2297 non-null   object
 10  rating_numerator           2356 non-null   int64
 11  rating_denominator         2356 non-null   int64
 12  name                       2356 non-null   object
 13  doggo                      2356 non-null   object
 14  floofer                    2356 non-null   object
 15  pupper                     2356 non-null   object
 16  puppo                      2356 non-null   object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

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

[873]:
```python
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/
↪599fd2ad_image-predictions/image-predictions.tsv'
# url = 'http://localhost:8888/edit/imagepredictions.tsv'
response = requests.get(url)
with open(url.split('/')[-1], mode= 'wb') as outfile:
    outfile.write(response.content)

df1 = pd.read_csv('image-predictions.tsv', sep='\t', encoding = 'utf-8')

df1
```

```
[873]:              tweet_id                                            jpg_url  \
      0     666020888022790149   https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
      1     666029285002620928   https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
      2     666033412701032449   https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
      3     666044226329800704   https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
      4     666049248165822465   https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
      ...                   ...                                               ...
      2070  891327558926688256   https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
      2071  891689557279858688   https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
      2072  891815181378084864   https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
      2073  892177421306343426   https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
      2074  892420643555336193   https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

            img_num                          p1    p1_conf  p1_dog                 p2  \
      0     1         Welsh_springer_spaniel  0.465074    True      collie
      1     1         redbone                 0.506826    True      miniature_pinscher
      2     1         German_shepherd         0.596461    True      malinois
      3     1         Rhodesian_ridgeback     0.408143    True      redbone
      4     1         miniature_pinscher      0.560311    True      Rottweiler
      ...  ..                          ...         ...     ...                ...
      2070  2         basset                  0.555712    True      English_springer
      2071  1         paper_towel             0.170278    False     Labrador_retriever
      2072  1         Chihuahua               0.716012    True      malamute
      2073  1         Chihuahua               0.323581    True      Pekinese
      2074  1         orange                  0.097049    False     bagel

             p2_conf  p2_dog                          p3    p3_conf  p3_dog
      0     0.156665    True      Shetland_sheepdog       0.061428    True
      1     0.074192    True      Rhodesian_ridgeback     0.072010    True
      2     0.138584    True      bloodhound              0.116197    True
      3     0.360687    True      miniature_pinscher      0.222752    True
      4     0.243682    True      Doberman                0.154629    True
      ...        ...     ...                          ...        ...     ...
      2070  0.225770    True      German_short-haired_pointer  0.175219    True
      2071  0.168086    True      spatula                 0.040836    False
      2072  0.078253    True      kelpie                  0.031379    True
      2073  0.090647    True      papillon                0.068957    True
      2074  0.085851    False     banana                  0.076110    False

      [2075 rows x 12 columns]

[874]: df1.info()

      <class 'pandas.core.frame.DataFrame'>
      RangeIndex: 2075 entries, 0 to 2074
      Data columns (total 12 columns):
       #   Column    Non-Null Count  Dtype
      ---  ------    --------------  -----
```

```
 0   tweet_id  2075 non-null   int64
 1   jpg_url   2075 non-null   object
 2   img_num   2075 non-null   int64
 3   p1        2075 non-null   object
 4   p1_conf   2075 non-null   float64
 5   p1_dog    2075 non-null   bool
 6   p2        2075 non-null   object
 7   p2_conf   2075 non-null   float64
 8   p2_dog    2075 non-null   bool
 9   p3        2075 non-null   object
 10  p3_conf   2075 non-null   float64
 11  p3_dog    2075 non-null   bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

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

```
[875]:  consumer_key = '...'
        consumer_secret = '...'
        access_token = '...'
        access_token_secret = '...'


        auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_token_secret)

        api = tweepy.API(auth,
                         parser=tweepy.parsers.JSONParser(),
                         wait_on_rate_limit=True,
        #                 wait_on_rate_limit_notify=True
                         )
```

```
[876]:  file_name = 'tweet_json.txt'

        # Create a file, remove previous one if exists
        #open(file_name, 'w').close()

        # Create a file if it doesn't exist
        if not os.path.isfile(file_name):
            open(file_name, 'w').close()
```

```
[877]:  data_list=[]
        df_list=[]

        with open('tweet_json.txt' , 'r') as myfile:
            for line in myfile:
                json_line = json.loads(line)
                data_list.append(json_line)
```

```python
for i in data_list:
    tweet_id = i['id']
    retweet_count = i['retweet_count']
    favorite_count = i['favorite_count']
    df_list.append({'tweet_id':tweet_id , 'retweet_count': retweet_count ,↲
↪'favorite_count': favorite_count})

# Create a data Frame from Json extracted data:
json_df = pd.DataFrame(df_list , columns = ['tweet_id' , 'retweet_count' ,↲
↪'favorite_count'])

df_list
```

[877]: [{'tweet_id': 892420643555336193,
  'retweet_count': 8853,
  'favorite_count': 39467},
 {'tweet_id': 892177421306343426,
  'retweet_count': 6514,
  'favorite_count': 33819},
 {'tweet_id': 891815181378084864,
  'retweet_count': 4328,
  'favorite_count': 25461},
 {'tweet_id': 891689557279858688,
  'retweet_count': 8964,
  'favorite_count': 42908},
 {'tweet_id': 891327558926688256,
  'retweet_count': 9774,
  'favorite_count': 41048},
 {'tweet_id': 891087950875897856,
  'retweet_count': 3261,
  'favorite_count': 20562},
 {'tweet_id': 890971913173991426,
  'retweet_count': 2158,
  'favorite_count': 12041},
 {'tweet_id': 890729181411237888,
  'retweet_count': 16716,
  'favorite_count': 56848},
 {'tweet_id': 890609185150312448,
  'retweet_count': 4429,
  'favorite_count': 28226},
 {'tweet_id': 890240255349198849,
  'retweet_count': 7711,
  'favorite_count': 32467},
 {'tweet_id': 890006608113172480,

  'retweet_count': 7624,
  'favorite_count': 31166},
{'tweet_id': 889880896479866881,
  'retweet_count': 5156,
  'favorite_count': 28268},
{'tweet_id': 889665388333682689,
  'retweet_count': 8538,
  'favorite_count': 38818},
{'tweet_id': 889638837579907072,
  'retweet_count': 4735,
  'favorite_count': 27672},
{'tweet_id': 889531135344209921,
  'retweet_count': 2321,
  'favorite_count': 15359},
{'tweet_id': 889278841981685760,
  'retweet_count': 5637,
  'favorite_count': 25652},
{'tweet_id': 888917238123831296,
  'retweet_count': 4709,
  'favorite_count': 29611},
{'tweet_id': 888804989199671297,
  'retweet_count': 4559,
  'favorite_count': 26080},
{'tweet_id': 888554962724278272,
  'retweet_count': 3732,
  'favorite_count': 20290},
{'tweet_id': 888078434458587136,
  'retweet_count': 3653,
  'favorite_count': 22201},
{'tweet_id': 887705289381826560,
  'retweet_count': 5609,
  'favorite_count': 30779},
{'tweet_id': 887517139158093824,
  'retweet_count': 12082,
  'favorite_count': 46959},
{'tweet_id': 887473957103951883,
  'retweet_count': 18781,
  'favorite_count': 69871},
{'tweet_id': 887343217045368832,
  'retweet_count': 10737,
  'favorite_count': 34222},
{'tweet_id': 887101392804085760,
  'retweet_count': 6167,
  'favorite_count': 31061},
{'tweet_id': 886983233522544640,
  'retweet_count': 8084,
  'favorite_count': 35859},

{'tweet_id': 886736880519319552,
 'retweet_count': 3443,
 'favorite_count': 12306},
{'tweet_id': 886680336477933568,
 'retweet_count': 4610,
 'favorite_count': 22798},
{'tweet_id': 886366144734445568,
 'retweet_count': 3316,
 'favorite_count': 21524},
{'tweet_id': 886267009285017600, 'retweet_count': 4, 'favorite_count': 117},
{'tweet_id': 886258384151887873,
 'retweet_count': 6523,
 'favorite_count': 28469},
{'tweet_id': 886054160059072513, 'retweet_count': 108, 'favorite_count': 0},
{'tweet_id': 885984800019947520,
 'retweet_count': 7097,
 'favorite_count': 33382},
{'tweet_id': 885528943205470208,
 'retweet_count': 6683,
 'favorite_count': 36689},
{'tweet_id': 885518971528720385,
 'retweet_count': 3899,
 'favorite_count': 20788},
{'tweet_id': 885311592912609280, 'retweet_count': 19297, 'favorite_count': 0},
{'tweet_id': 885167619883638784,
 'retweet_count': 4556,
 'favorite_count': 22367},
{'tweet_id': 884925521741709313,
 'retweet_count': 16439,
 'favorite_count': 68152},
{'tweet_id': 884876753390489601,
 'retweet_count': 6096,
 'favorite_count': 28514},
{'tweet_id': 884562892145688576,
 'retweet_count': 5100,
 'favorite_count': 24765},
{'tweet_id': 884441805382717440,
 'retweet_count': 5856,
 'favorite_count': 27478},
{'tweet_id': 884247878851493888,
 'retweet_count': 21298,
 'favorite_count': 74423},
{'tweet_id': 884162670584377345,
 'retweet_count': 3128,
 'favorite_count': 20771},
{'tweet_id': 883838122936631299,
 'retweet_count': 3586,

'favorite_count': 22349},
{'tweet_id': 883482846933004288,
 'retweet_count': 10407,
 'favorite_count': 46860},
{'tweet_id': 883360690899218434,
 'retweet_count': 3825,
 'favorite_count': 22986},
{'tweet_id': 883117836046086144,
 'retweet_count': 6949,
 'favorite_count': 37914},
{'tweet_id': 882992080364220416,
 'retweet_count': 4122,
 'favorite_count': 24445},
{'tweet_id': 882762694511734784,
 'retweet_count': 5134,
 'favorite_count': 28903},
{'tweet_id': 882627270321602560,
 'retweet_count': 6342,
 'favorite_count': 28382},
{'tweet_id': 882268110199369728,
 'retweet_count': 12118,
 'favorite_count': 45880},
{'tweet_id': 882045870035918850,
 'retweet_count': 5203,
 'favorite_count': 29900},
{'tweet_id': 881906580714921986,
 'retweet_count': 3533,
 'favorite_count': 24773},
{'tweet_id': 881666595344535552,
 'retweet_count': 11099,
 'favorite_count': 51522},
{'tweet_id': 881633300179243008, 'retweet_count': 7, 'favorite_count': 129},
{'tweet_id': 881536004380872706,
 'retweet_count': 16570,
 'favorite_count': 50199},
{'tweet_id': 881268444196462592,
 'retweet_count': 5358,
 'favorite_count': 23501},
{'tweet_id': 880935762899988482,
 'retweet_count': 2886,
 'favorite_count': 17346},
{'tweet_id': 880872448815771648,
 'retweet_count': 3989,
 'favorite_count': 21734},
{'tweet_id': 880465832366813184,
 'retweet_count': 6546,
 'favorite_count': 29075},

```
{'tweet_id': 880221127280381952,
 'retweet_count': 4436,
 'favorite_count': 27640},
{'tweet_id': 880095782870896641,
 'retweet_count': 4533,
 'favorite_count': 28150},
{'tweet_id': 879862464715927552,
 'retweet_count': 3642,
 'favorite_count': 22667},
{'tweet_id': 879674319642796034, 'retweet_count': 10, 'favorite_count': 315},
{'tweet_id': 879492040517615616,
 'retweet_count': 3323,
 'favorite_count': 23822},
{'tweet_id': 879415818425184262,
 'retweet_count': 45849,
 'favorite_count': 107956},
{'tweet_id': 879376492567855104,
 'retweet_count': 3261,
 'favorite_count': 17099},
{'tweet_id': 879130579576475649, 'retweet_count': 7181, 'favorite_count': 0},
{'tweet_id': 879050749262655488,
 'retweet_count': 4941,
 'favorite_count': 23022},
{'tweet_id': 879008229531029506,
 'retweet_count': 2812,
 'favorite_count': 19317},
{'tweet_id': 878776093423087618,
 'retweet_count': 4319,
 'favorite_count': 19763},
{'tweet_id': 878604707211726852,
 'retweet_count': 7582,
 'favorite_count': 30931},
{'tweet_id': 878404777348136964, 'retweet_count': 1349, 'favorite_count': 0},
{'tweet_id': 878316110768087041, 'retweet_count': 6965, 'favorite_count': 0},
{'tweet_id': 878281511006478336,
 'retweet_count': 1349,
 'favorite_count': 7913},
{'tweet_id': 878057613040115712,
 'retweet_count': 7181,
 'favorite_count': 42876},
{'tweet_id': 877736472329191424,
 'retweet_count': 17300,
 'favorite_count': 71144},
{'tweet_id': 877611172832227328, 'retweet_count': 83, 'favorite_count': 0},
{'tweet_id': 877556246731214848,
 'retweet_count': 3994,
 'favorite_count': 23258},
```

{'tweet_id': 873337748698140672, 'retweet_count': 1667, 'favorite_count': 0},
{'tweet_id': 873213775632977920,
 'retweet_count': 1667,
 'favorite_count': 7467},
{'tweet_id': 872967104147763200,
 'retweet_count': 5669,
 'favorite_count': 28031},
{'tweet_id': 872820683541237760,
 'retweet_count': 3884,
 'favorite_count': 15029},
{'tweet_id': 872668790621863937, 'retweet_count': 31, 'favorite_count': 0},
{'tweet_id': 872620804844003328,
 'retweet_count': 3911,
 'favorite_count': 21309},
{'tweet_id': 872486979161796608,
 'retweet_count': 9429,
 'favorite_count': 41606},
{'tweet_id': 872261713294495745,
 'retweet_count': 6649,
 'favorite_count': 35085},
{'tweet_id': 872122724285648897,
 'retweet_count': 8549,
 'favorite_count': 35324},
{'tweet_id': 871879754684805121,
 'retweet_count': 11918,
 'favorite_count': 39090},
{'tweet_id': 871762521631449091,
 'retweet_count': 3678,
 'favorite_count': 20787},
{'tweet_id': 871515927908634625,
 'retweet_count': 3628,
 'favorite_count': 20730},
{'tweet_id': 871166179821445120, 'retweet_count': 5991, 'favorite_count': 0},
{'tweet_id': 871102520638267392,
 'retweet_count': 5764,
 'favorite_count': 21461},
{'tweet_id': 871032628920680449,
 'retweet_count': 3999,
 'favorite_count': 23255},
{'tweet_id': 870804317367881728,
 'retweet_count': 6393,
 'favorite_count': 33791},
{'tweet_id': 870726314365509632, 'retweet_count': 3, 'favorite_count': 121},
{'tweet_id': 870656317836468226,
 'retweet_count': 2817,
 'favorite_count': 12819},
{'tweet_id': 870374049280663552,

  'retweet_count': 27680,
  'favorite_count': 85011},
 {'tweet_id': 870308999962521604,
  'retweet_count': 4384,
  'favorite_count': 22453},
 {'tweet_id': 870063196459192321,
  'retweet_count': 8840,
  'favorite_count': 37771},
 {'tweet_id': 869988702071779329, 'retweet_count': 25661, 'favorite_count': 0},
 {'tweet_id': 869772420881756160,
  'retweet_count': 10663,
  'favorite_count': 43710},
 {'tweet_id': 869702957897576449,
  'retweet_count': 6728,
  'favorite_count': 29116},
 {'tweet_id': 869596645499047938,
  'retweet_count': 3327,
  'favorite_count': 16476},
 {'tweet_id': 869227993411051520,
  'retweet_count': 4023,
  'favorite_count': 21112},
 {'tweet_id': 868880397819494401,
  'retweet_count': 12518,
  'favorite_count': 55098},
 {'tweet_id': 868639477480148993, 'retweet_count': 2240, 'favorite_count': 0},
 {'tweet_id': 868622495443632128,
  'retweet_count': 6275,
  'favorite_count': 28295},
 {'tweet_id': 868552278524837888,
  'retweet_count': 2240,
  'favorite_count': 10539},
 {'tweet_id': 867900495410671616,
  'retweet_count': 4439,
  'favorite_count': 24964},
 {'tweet_id': 867774946302451713,
  'retweet_count': 7788,
  'favorite_count': 35179},
 {'tweet_id': 867421006826221569,
  'retweet_count': 2697,
  'favorite_count': 16755},
 {'tweet_id': 867072653475098625, 'retweet_count': 135, 'favorite_count': 0},
 {'tweet_id': 867051520902168576,
  'retweet_count': 8425,
  'favorite_count': 33420},
 {'tweet_id': 866816280283807744, 'retweet_count': 32883, 'favorite_count': 0},
 {'tweet_id': 866720684873056260,
  'retweet_count': 5168,

'favorite_count': 20888},
{'tweet_id': 866686824827068416,
 'retweet_count': 3727,
 'favorite_count': 20070},
{'tweet_id': 866450705531457537,
 'retweet_count': 32883,
 'favorite_count': 106827},
{'tweet_id': 866334964761202691,
 'retweet_count': 15546,
 'favorite_count': 54720},
{'tweet_id': 866094527597207552, 'retweet_count': 8952, 'favorite_count': 0},
{'tweet_id': 865718153858494464,
 'retweet_count': 6008,
 'favorite_count': 26640},
{'tweet_id': 865359393868664832,
 'retweet_count': 5384,
 'favorite_count': 27530},
{'tweet_id': 865006731092295680,
 'retweet_count': 8209,
 'favorite_count': 29063},
{'tweet_id': 864873206498414592,
 'retweet_count': 9361,
 'favorite_count': 33651},
{'tweet_id': 864279568663928832,
 'retweet_count': 3266,
 'favorite_count': 15195},
{'tweet_id': 864197398364647424,
 'retweet_count': 9616,
 'favorite_count': 31459},
{'tweet_id': 863907417377173506,
 'retweet_count': 4441,
 'favorite_count': 21477},
{'tweet_id': 863553081350529029,
 'retweet_count': 4489,
 'favorite_count': 15935},
{'tweet_id': 863471782782697472, 'retweet_count': 2687, 'favorite_count': 0},
{'tweet_id': 863432100342583297,
 'retweet_count': 5664,
 'favorite_count': 24829},
{'tweet_id': 863427515083354112,
 'retweet_count': 105,
 'favorite_count': 2363},
{'tweet_id': 863079547188785154,
 'retweet_count': 1195,
 'favorite_count': 9094},
{'tweet_id': 863062471531167744,
 'retweet_count': 2687,

'favorite_count': 8945},
{'tweet_id': 862831371563274240,
 'retweet_count': 5457,
 'favorite_count': 20011},
{'tweet_id': 862722525377298433,
 'retweet_count': 3809,
 'favorite_count': 17779},
{'tweet_id': 862457590147678208,
 'retweet_count': 5388,
 'favorite_count': 21492},
{'tweet_id': 862096992088072192,
 'retweet_count': 21840,
 'favorite_count': 66437},
{'tweet_id': 861769973181624320, 'retweet_count': 37911, 'favorite_count': 0},
{'tweet_id': 861383897657036800,
 'retweet_count': 11528,
 'favorite_count': 37744},
{'tweet_id': 861288531465048066,
 'retweet_count': 4479,
 'favorite_count': 18032},
{'tweet_id': 861005113778896900,
 'retweet_count': 4119,
 'favorite_count': 17538},
{'tweet_id': 860981674716409858, 'retweet_count': 2334, 'favorite_count': 0},
{'tweet_id': 860924035999428608, 'retweet_count': 882, 'favorite_count': 0},
{'tweet_id': 860563773140209665,
 'retweet_count': 2334,
 'favorite_count': 7878},
{'tweet_id': 860524505164394496,
 'retweet_count': 5698,
 'favorite_count': 24678},
{'tweet_id': 860276583193509888,
 'retweet_count': 3745,
 'favorite_count': 19154},
{'tweet_id': 860184849394610176,
 'retweet_count': 6295,
 'favorite_count': 17474},
{'tweet_id': 860177593139703809, 'retweet_count': 33421, 'favorite_count': 0},
{'tweet_id': 859924526012018688,
 'retweet_count': 4349,
 'favorite_count': 20021},
{'tweet_id': 859851578198683649,
 'retweet_count': 3780,
 'favorite_count': 16105},
{'tweet_id': 859607811541651456,
 'retweet_count': 1704,
 'favorite_count': 19476},

{'tweet_id': 859196978902773760,
 'retweet_count': 25661,
 'favorite_count': 75193},
{'tweet_id': 859074603037188101,
 'retweet_count': 14740,
 'favorite_count': 35553},
{'tweet_id': 858860390427611136, 'retweet_count': 8805, 'favorite_count': 0},
{'tweet_id': 858843525470990336,
 'retweet_count': 3771,
 'favorite_count': 16304},
{'tweet_id': 858471635011153920,
 'retweet_count': 5271,
 'favorite_count': 22640},
{'tweet_id': 858107933456039936,
 'retweet_count': 3154,
 'favorite_count': 16524},
{'tweet_id': 857989990357356544,
 'retweet_count': 2812,
 'favorite_count': 16952},
{'tweet_id': 857746408056729600,
 'retweet_count': 11524,
 'favorite_count': 36021},
{'tweet_id': 857393404942143489,
 'retweet_count': 1785,
 'favorite_count': 6236},
{'tweet_id': 857263160327368704,
 'retweet_count': 4934,
 'favorite_count': 21041},
{'tweet_id': 857214891891077121, 'retweet_count': 20, 'favorite_count': 242},
{'tweet_id': 857062103051644929, 'retweet_count': 183, 'favorite_count': 0},
{'tweet_id': 857029823797047296,
 'retweet_count': 4364,
 'favorite_count': 19910},
{'tweet_id': 856602993587888130, 'retweet_count': 11633, 'favorite_count': 0},
{'tweet_id': 856543823941562368,
 'retweet_count': 3131,
 'favorite_count': 17135},
{'tweet_id': 856526610513747968,
 'retweet_count': 2068,
 'favorite_count': 12446},
{'tweet_id': 856330835276025856, 'retweet_count': 731, 'favorite_count': 0},
{'tweet_id': 856288084350160898, 'retweet_count': 17, 'favorite_count': 545},
{'tweet_id': 856282028240666624,
 'retweet_count': 6841,
 'favorite_count': 29086},
{'tweet_id': 855862651834028034, 'retweet_count': 27, 'favorite_count': 320},
{'tweet_id': 855860136149123072,

  'retweet_count': 1058,
  'favorite_count': 4407},
{'tweet_id': 855857698524602368,
  'retweet_count': 2313,
  'favorite_count': 12498},
{'tweet_id': 855851453814013952,
  'retweet_count': 19196,
  'favorite_count': 47844},
{'tweet_id': 855818117272018944,
  'retweet_count': 5943,
  'favorite_count': 28056},
{'tweet_id': 855459453768019968,
  'retweet_count': 8987,
  'favorite_count': 31657},
{'tweet_id': 855245323840757760, 'retweet_count': 6568, 'favorite_count': 0},
{'tweet_id': 855138241867124737, 'retweet_count': 50, 'favorite_count': 0},
{'tweet_id': 854732716440526848,
  'retweet_count': 6690,
  'favorite_count': 24188},
{'tweet_id': 854482394044301312,
  'retweet_count': 7608,
  'favorite_count': 31131},
{'tweet_id': 854365224396361728,
  'retweet_count': 5159,
  'favorite_count': 20046},
{'tweet_id': 854120357044912130,
  'retweet_count': 8285,
  'favorite_count': 33911},
{'tweet_id': 854010172552949760,
  'retweet_count': 3433,
  'favorite_count': 17169},
{'tweet_id': 853760880890318849,
  'retweet_count': 6403,
  'favorite_count': 30414},
{'tweet_id': 853639147608842240,
  'retweet_count': 11265,
  'favorite_count': 37198},
{'tweet_id': 853299958564483072,
  'retweet_count': 3928,
  'favorite_count': 16508},
{'tweet_id': 852936405516943360, 'retweet_count': 2243, 'favorite_count': 0},
{'tweet_id': 852912242202992640,
  'retweet_count': 2037,
  'favorite_count': 9658},
{'tweet_id': 852672615818899456,
  'retweet_count': 2388,
  'favorite_count': 15939},

{'tweet_id': 852553447878664193,
 'retweet_count': 3885,
 'favorite_count': 17492},
{'tweet_id': 852311364735569921,
 'retweet_count': 10961,
 'favorite_count': 35325},
{'tweet_id': 852226086759018497,
 'retweet_count': 7570,
 'favorite_count': 21378},
{'tweet_id': 852189679701164033,
 'retweet_count': 1741,
 'favorite_count': 12217},
{'tweet_id': 851953902622658560, 'retweet_count': 10706, 'favorite_count': 0},
{'tweet_id': 851861385021730816, 'retweet_count': 23, 'favorite_count': 0},
{'tweet_id': 851591660324737024,
 'retweet_count': 3819,
 'favorite_count': 17300},
{'tweet_id': 851464819735769094,
 'retweet_count': 7855,
 'favorite_count': 25944},
{'tweet_id': 851224888060895234,
 'retweet_count': 6350,
 'favorite_count': 22090},
{'tweet_id': 850753642995093505,
 'retweet_count': 10352,
 'favorite_count': 33348},
{'tweet_id': 850380195714523136,
 'retweet_count': 2915,
 'favorite_count': 13994},
{'tweet_id': 850333567704068097,
 'retweet_count': 367,
 'favorite_count': 3647},
{'tweet_id': 850145622816686080,
 'retweet_count': 4244,
 'favorite_count': 17519},
{'tweet_id': 850019790995546112,
 'retweet_count': 5459,
 'favorite_count': 21944},
{'tweet_id': 849776966551130114,
 'retweet_count': 8404,
 'favorite_count': 32390},
{'tweet_id': 849668094696017920, 'retweet_count': 5513, 'favorite_count': 0},
{'tweet_id': 849412302885593088,
 'retweet_count': 3487,
 'favorite_count': 17039},
{'tweet_id': 849336543269576704,
 'retweet_count': 2101,

```
 'favorite_count': 12240},
{'tweet_id': 849051919805034497,
 'retweet_count': 7396,
 'favorite_count': 32617},
{'tweet_id': 848690551926992896,
 'retweet_count': 4826,
 'favorite_count': 27104},
{'tweet_id': 848324959059550208,
 'retweet_count': 4037,
 'favorite_count': 20229},
{'tweet_id': 848213670039564288,
 'retweet_count': 832,
 'favorite_count': 8834},
{'tweet_id': 848212111729840128,
 'retweet_count': 3444,
 'favorite_count': 17618},
{'tweet_id': 847978865427394560, 'retweet_count': 3652, 'favorite_count': 0},
{'tweet_id': 847971574464610304, 'retweet_count': 482, 'favorite_count': 0},
{'tweet_id': 847962785489326080,
 'retweet_count': 5730,
 'favorite_count': 25296},
{'tweet_id': 847842811428974592,
 'retweet_count': 1522,
 'favorite_count': 5935},
{'tweet_id': 847617282490613760,
 'retweet_count': 468,
 'favorite_count': 7558},
{'tweet_id': 847606175596138505,
 'retweet_count': 3774,
 'favorite_count': 20208},
{'tweet_id': 847251039262605312,
 'retweet_count': 4800,
 'favorite_count': 22036},
{'tweet_id': 847157206088847362,
 'retweet_count': 6572,
 'favorite_count': 21588},
{'tweet_id': 847116187444137987,
 'retweet_count': 3583,
 'favorite_count': 23108},
{'tweet_id': 846874817362120707,
 'retweet_count': 4404,
 'favorite_count': 21685},
{'tweet_id': 846514051647705089,
 'retweet_count': 13076,
 'favorite_count': 48410},
{'tweet_id': 846505985330044928,
 'retweet_count': 3492,
```

```
 'favorite_count': 15304},
{'tweet_id': 846153765933735936,
 'retweet_count': 10226,
 'favorite_count': 34394},
{'tweet_id': 846139713627017216,
 'retweet_count': 10972,
 'favorite_count': 33188},
{'tweet_id': 846042936437604353,
 'retweet_count': 3224,
 'favorite_count': 17256},
{'tweet_id': 845812042753855489,
 'retweet_count': 9894,
 'favorite_count': 31737},
{'tweet_id': 845677943972139009,
 'retweet_count': 5365,
 'favorite_count': 27154},
{'tweet_id': 845459076796616705, 'retweet_count': 15071, 'favorite_count': 0},
{'tweet_id': 845397057150107648,
 'retweet_count': 2072,
 'favorite_count': 8241},
{'tweet_id': 845306882940190720,
 'retweet_count': 6039,
 'favorite_count': 25225},
{'tweet_id': 845098359547420673, 'retweet_count': 8640, 'favorite_count': 0},
{'tweet_id': 844979544864018432,
 'retweet_count': 2909,
 'favorite_count': 14738},
{'tweet_id': 844973813909606400,
 'retweet_count': 3617,
 'favorite_count': 16361},
{'tweet_id': 844704788403113984,
 'retweet_count': 11633,
 'favorite_count': 42022},
{'tweet_id': 844580511645339650,
 'retweet_count': 3533,
 'favorite_count': 17871},
{'tweet_id': 844223788422217728,
 'retweet_count': 2450,
 'favorite_count': 14753},
{'tweet_id': 843981021012017153,
 'retweet_count': 3285,
 'favorite_count': 16327},
{'tweet_id': 843856843873095681,
 'retweet_count': 5220,
 'favorite_count': 23211},
{'tweet_id': 843604394117681152,
 'retweet_count': 3081,
```

 'favorite_count': 18310},
{'tweet_id': 843235543001513987,
 'retweet_count': 6852,
 'favorite_count': 23315},
{'tweet_id': 842892208864923648, 'retweet_count': 56625, 'favorite_count': 0},
{'tweet_id': 842846295480000512,
 'retweet_count': 4023,
 'favorite_count': 16440},
{'tweet_id': 842765311967449089,
 'retweet_count': 1439,
 'favorite_count': 7321},
{'tweet_id': 842535590457499648,
 'retweet_count': 3937,
 'favorite_count': 19637},
{'tweet_id': 842163532590374912,
 'retweet_count': 6568,
 'favorite_count': 26569},
{'tweet_id': 842115215311396866,
 'retweet_count': 3386,
 'favorite_count': 15204},
{'tweet_id': 841833993020538882, 'retweet_count': 17504, 'favorite_count': 0},
{'tweet_id': 841680585030541313,
 'retweet_count': 8748,
 'favorite_count': 27854},
{'tweet_id': 841439858740625411,
 'retweet_count': 4168,
 'favorite_count': 13755},
{'tweet_id': 841320156043304961,
 'retweet_count': 6080,
 'favorite_count': 21402},
{'tweet_id': 841314665196081154,
 'retweet_count': 5312,
 'favorite_count': 17305},
{'tweet_id': 841077006473256960,
 'retweet_count': 5991,
 'favorite_count': 24926},
{'tweet_id': 840761248237133825, 'retweet_count': 1972, 'favorite_count': 0},
{'tweet_id': 840728873075638272, 'retweet_count': 18285, 'favorite_count': 0},
{'tweet_id': 840698636975636481, 'retweet_count': 3, 'favorite_count': 197},
{'tweet_id': 840696689258311684,
 'retweet_count': 1116,
 'favorite_count': 13377},
{'tweet_id': 840632337062862849,
 'retweet_count': 1972,
 'favorite_count': 9761},
{'tweet_id': 840370681858686976,
 'retweet_count': 5146,

'favorite_count': 17918},
{'tweet_id': 840268004936019968,
 'retweet_count': 6497,
 'favorite_count': 20950},
{'tweet_id': 839990271299457024,
 'retweet_count': 2597,
 'favorite_count': 14640},
{'tweet_id': 839549326359670784,
 'retweet_count': 8805,
 'favorite_count': 29957},
{'tweet_id': 839290600511926273, 'retweet_count': 158, 'favorite_count': 0},
{'tweet_id': 839239871831150596,
 'retweet_count': 7422,
 'favorite_count': 29684},
{'tweet_id': 838952994649550848,
 'retweet_count': 4505,
 'favorite_count': 21289},
{'tweet_id': 838921590096166913,
 'retweet_count': 2357,
 'favorite_count': 12183},
{'tweet_id': 838916489579200512, 'retweet_count': 38, 'favorite_count': 0},
{'tweet_id': 838831947270979586, 'retweet_count': 12643, 'favorite_count': 0},
{'tweet_id': 838561493054533637,
 'retweet_count': 1504,
 'favorite_count': 11892},
{'tweet_id': 838476387338051585,
 'retweet_count': 5484,
 'favorite_count': 24664},
{'tweet_id': 838201503651401729, 'retweet_count': 11525, 'favorite_count': 0},
{'tweet_id': 838150277551247360,
 'retweet_count': 370,
 'favorite_count': 1824},
{'tweet_id': 838085839343206401, 'retweet_count': 0, 'favorite_count': 150},
{'tweet_id': 838083903487373313,
 'retweet_count': 3582,
 'favorite_count': 19183},
{'tweet_id': 837820167694528512,
 'retweet_count': 8952,
 'favorite_count': 37277},
{'tweet_id': 837482249356513284,
 'retweet_count': 495,
 'favorite_count': 4204},
{'tweet_id': 837471256429613056,
 'retweet_count': 2631,
 'favorite_count': 13967},
{'tweet_id': 837366284874571778,
 'retweet_count': 6005,

'favorite_count': 23074},
{'tweet_id': 837110210464448512,
 'retweet_count': 2731,
 'favorite_count': 17480},
{'tweet_id': 837012587749474308, 'retweet_count': 88, 'favorite_count': 0},
{'tweet_id': 836989968035819520,
 'retweet_count': 2610,
 'favorite_count': 13879},
{'tweet_id': 836753516572119041,
 'retweet_count': 5237,
 'favorite_count': 21029},
{'tweet_id': 836677758902222849,
 'retweet_count': 2522,
 'favorite_count': 13782},
{'tweet_id': 836648853927522308, 'retweet_count': 626, 'favorite_count': 0},
{'tweet_id': 836397794269200385, 'retweet_count': 31314, 'favorite_count': 0},
{'tweet_id': 836380477523124226,
 'retweet_count': 3337,
 'favorite_count': 16037},
{'tweet_id': 836260088725786625,
 'retweet_count': 4850,
 'favorite_count': 23177},
{'tweet_id': 836001077879255040,
 'retweet_count': 4935,
 'favorite_count': 20924},
{'tweet_id': 835685285446955009, 'retweet_count': 9327, 'favorite_count': 0},
{'tweet_id': 835574547218894849,
 'retweet_count': 4121,
 'favorite_count': 19447},
{'tweet_id': 835536468978302976, 'retweet_count': 1939, 'favorite_count': 0},
{'tweet_id': 835309094223372289, 'retweet_count': 24013, 'favorite_count': 0},
{'tweet_id': 835297930240217089,
 'retweet_count': 3381,
 'favorite_count': 17847},
{'tweet_id': 835264098648616962,
 'retweet_count': 1939,
 'favorite_count': 8503},
{'tweet_id': 835246439529840640, 'retweet_count': 83, 'favorite_count': 2259},
{'tweet_id': 835172783151792128,
 'retweet_count': 6516,
 'favorite_count': 28552},
{'tweet_id': 835152434251116546,
 'retweet_count': 3443,
 'favorite_count': 24574},
{'tweet_id': 834931633769889797,
 'retweet_count': 1878,
 'favorite_count': 11838},

{'tweet_id': 834786237630337024,
 'retweet_count': 6159,
 'favorite_count': 22943},
{'tweet_id': 834574053763584002,
 'retweet_count': 2882,
 'favorite_count': 14993},
{'tweet_id': 834477809192075265, 'retweet_count': 12224, 'favorite_count': 0},
{'tweet_id': 834458053273591808,
 'retweet_count': 1899,
 'favorite_count': 10512},
{'tweet_id': 834209720923721728,
 'retweet_count': 5476,
 'favorite_count': 22594},
{'tweet_id': 834167344700198914,
 'retweet_count': 4130,
 'favorite_count': 17194},
{'tweet_id': 834089966724603904,
 'retweet_count': 2427,
 'favorite_count': 10971},
{'tweet_id': 834086379323871233,
 'retweet_count': 2512,
 'favorite_count': 14296},
{'tweet_id': 833863086058651648,
 'retweet_count': 2729,
 'favorite_count': 14661},
{'tweet_id': 833826103416520705,
 'retweet_count': 3904,
 'favorite_count': 16728},
{'tweet_id': 833732339549220864, 'retweet_count': 251, 'favorite_count': 0},
{'tweet_id': 833722901757046785,
 'retweet_count': 3636,
 'favorite_count': 22585},
{'tweet_id': 833479644947025920,
 'retweet_count': 2357,
 'favorite_count': 16258},
{'tweet_id': 833124694597443584,
 'retweet_count': 5513,
 'favorite_count': 22133},
{'tweet_id': 832998151111966721,
 'retweet_count': 2522,
 'favorite_count': 14549},
{'tweet_id': 832769181346996225, 'retweet_count': 43, 'favorite_count': 0},
{'tweet_id': 832757312314028032,
 'retweet_count': 4127,
 'favorite_count': 18423},
{'tweet_id': 832682457690300417,
 'retweet_count': 3368,

 'favorite_count': 13017},
{'tweet_id': 832645525019123713,
 'retweet_count': 594,
 'favorite_count': 3195},
{'tweet_id': 832636094638288896,
 'retweet_count': 3220,
 'favorite_count': 17379},
{'tweet_id': 832397543355072512,
 'retweet_count': 2548,
 'favorite_count': 13126},
{'tweet_id': 832369877331693569,
 'retweet_count': 3652,
 'favorite_count': 18792},
{'tweet_id': 832273440279240704,
 'retweet_count': 2673,
 'favorite_count': 12385},
{'tweet_id': 832215909146226688, 'retweet_count': 7069, 'favorite_count': 0},
{'tweet_id': 832215726631055365, 'retweet_count': 10723, 'favorite_count': 0},
{'tweet_id': 832088576586297345, 'retweet_count': 3, 'favorite_count': 72},
{'tweet_id': 832040443403784192, 'retweet_count': 11131, 'favorite_count': 0},
{'tweet_id': 832032802820481025,
 'retweet_count': 4746,
 'favorite_count': 13887},
{'tweet_id': 831939777352105988,
 'retweet_count': 7031,
 'favorite_count': 26404},
{'tweet_id': 831926988323639298, 'retweet_count': 39, 'favorite_count': 369},
{'tweet_id': 831911600680497154,
 'retweet_count': 7458,
 'favorite_count': 30380},
{'tweet_id': 831670449226514432,
 'retweet_count': 2059,
 'favorite_count': 11469},
{'tweet_id': 831650051525054464,
 'retweet_count': 2243,
 'favorite_count': 7908},
{'tweet_id': 831552930092285952,
 'retweet_count': 2632,
 'favorite_count': 9872},
{'tweet_id': 831322785565769729,
 'retweet_count': 1744,
 'favorite_count': 10042},
{'tweet_id': 831315979191906304,
 'retweet_count': 1264,
 'favorite_count': 7117},
{'tweet_id': 831309418084069378,
 'retweet_count': 2786,

 'favorite_count': 12819},
{'tweet_id': 831262627380748289,
 'retweet_count': 2350,
 'favorite_count': 13066},
{'tweet_id': 830956169170665475,
 'retweet_count': 1735,
 'favorite_count': 8735},
{'tweet_id': 830583320585068544,
 'retweet_count': 19297,
 'favorite_count': 73397},
{'tweet_id': 830173239259324417, 'retweet_count': 6554, 'favorite_count': 0},
{'tweet_id': 830097400375152640,
 'retweet_count': 3455,
 'favorite_count': 10804},
{'tweet_id': 829878982036299777, 'retweet_count': 5757, 'favorite_count': 0},
{'tweet_id': 829861396166877184,
 'retweet_count': 2243,
 'favorite_count': 13441},
{'tweet_id': 829501995190984704,
 'retweet_count': 12224,
 'favorite_count': 34913},
{'tweet_id': 829449946868879360,
 'retweet_count': 2329,
 'favorite_count': 11519},
{'tweet_id': 829374341691346946,
 'retweet_count': 10706,
 'favorite_count': 38074},
{'tweet_id': 829141528400556032,
 'retweet_count': 8530,
 'favorite_count': 26952},
{'tweet_id': 829011960981237760,
 'retweet_count': 18627,
 'favorite_count': 58302},
{'tweet_id': 828801551087042563, 'retweet_count': 3901, 'favorite_count': 0},
{'tweet_id': 828770345708580865,
 'retweet_count': 6746,
 'favorite_count': 28085},
{'tweet_id': 828708714936930305,
 'retweet_count': 12882,
 'favorite_count': 40241},
{'tweet_id': 828650029636317184,
 'retweet_count': 1544,
 'favorite_count': 10467},
{'tweet_id': 828409743546925057,
 'retweet_count': 1305,
 'favorite_count': 6898},
{'tweet_id': 828408677031882754,

 'retweet_count': 1477,
 'favorite_count': 8398},
{'tweet_id': 828381636999917570,
 'retweet_count': 2554,
 'favorite_count': 13864},
{'tweet_id': 828376505180889089,
 'retweet_count': 1216,
 'favorite_count': 8112},
{'tweet_id': 828372645993398273,
 'retweet_count': 3344,
 'favorite_count': 13756},
{'tweet_id': 828361771580813312,
 'retweet_count': 195,
 'favorite_count': 2408},
{'tweet_id': 828046555563323392,
 'retweet_count': 3260,
 'favorite_count': 12923},
{'tweet_id': 828011680017821696,
 'retweet_count': 2451,
 'favorite_count': 11411},
{'tweet_id': 827933404142436356,
 'retweet_count': 5987,
 'favorite_count': 22180},
{'tweet_id': 827653905312006145,
 'retweet_count': 3433,
 'favorite_count': 16983},
{'tweet_id': 827600520311402496,
 'retweet_count': 1082,
 'favorite_count': 8143},
{'tweet_id': 827324948884643840,
 'retweet_count': 3510,
 'favorite_count': 17523},
{'tweet_id': 827228250799742977, 'retweet_count': 13517, 'favorite_count': 0},
{'tweet_id': 827199976799354881,
 'retweet_count': 2579,
 'favorite_count': 11659},
{'tweet_id': 826958653328592898,
 'retweet_count': 5757,
 'favorite_count': 23767},
{'tweet_id': 826848821049180160,
 'retweet_count': 11878,
 'favorite_count': 40325},
{'tweet_id': 826615380357632002, 'retweet_count': 4521, 'favorite_count': 0},
{'tweet_id': 826598799820865537,
 'retweet_count': 292,
 'favorite_count': 5637},
{'tweet_id': 826598365270007810,

  'retweet_count': 2709,
  'favorite_count': 11117},
{'tweet_id': 826476773533745153,
  'retweet_count': 4821,
  'favorite_count': 20275},
{'tweet_id': 826240494070030336,
  'retweet_count': 2965,
  'favorite_count': 14614},
{'tweet_id': 826204788643753985,
  'retweet_count': 1075,
  'favorite_count': 5361},
{'tweet_id': 826115272272650244,
  'retweet_count': 3599,
  'favorite_count': 17299},
{'tweet_id': 825876512159186944,
  'retweet_count': 2146,
  'favorite_count': 11525},
{'tweet_id': 825829644528148480,
  'retweet_count': 2848,
  'favorite_count': 14025},
{'tweet_id': 825535076884762624,
  'retweet_count': 19669,
  'favorite_count': 56413},
{'tweet_id': 825147591692263424,
  'retweet_count': 5244,
  'favorite_count': 20181},
{'tweet_id': 825120256414846976, 'retweet_count': 1483, 'favorite_count': 0},
{'tweet_id': 825026590719483904,
  'retweet_count': 1483,
  'favorite_count': 7020},
{'tweet_id': 824796380199809024, 'retweet_count': 6288, 'favorite_count': 0},
{'tweet_id': 824775126675836928,
  'retweet_count': 4069,
  'favorite_count': 16508},
{'tweet_id': 824663926340194305,
  'retweet_count': 1993,
  'favorite_count': 11113},
{'tweet_id': 824325613288833024,
  'retweet_count': 11848,
  'favorite_count': 12999},
{'tweet_id': 824297048279236611,
  'retweet_count': 4463,
  'favorite_count': 16625},
{'tweet_id': 824025158776213504,
  'retweet_count': 679,
  'favorite_count': 5255},
{'tweet_id': 823939628516474880,

 'retweet_count': 3123,
 'favorite_count': 11755},
{'tweet_id': 823719002937630720, 'retweet_count': 12953, 'favorite_count': 0},
{'tweet_id': 823699002998870016,
 'retweet_count': 2772,
 'favorite_count': 13826},
{'tweet_id': 823581115634085888,
 'retweet_count': 3031,
 'favorite_count': 14376},
{'tweet_id': 823333489516937216, 'retweet_count': 31, 'favorite_count': 777},
{'tweet_id': 823322678127919110,
 'retweet_count': 4637,
 'favorite_count': 17437},
{'tweet_id': 823269594223824897, 'retweet_count': 11421, 'favorite_count': 0},
{'tweet_id': 822975315408461824,
 'retweet_count': 3958,
 'favorite_count': 19139},
{'tweet_id': 822872901745569793,
 'retweet_count': 48265,
 'favorite_count': 132810},
{'tweet_id': 822859134160621569,
 'retweet_count': 2622,
 'favorite_count': 14576},
{'tweet_id': 822647212903690241, 'retweet_count': 7390, 'favorite_count': 0},
{'tweet_id': 822610361945911296,
 'retweet_count': 3423,
 'favorite_count': 16327},
{'tweet_id': 822489057087389700,
 'retweet_count': 7390,
 'favorite_count': 20083},
{'tweet_id': 822462944365645825,
 'retweet_count': 17209,
 'favorite_count': 31800},
{'tweet_id': 822244816520155136,
 'retweet_count': 11421,
 'favorite_count': 38832},
{'tweet_id': 822163064745328640, 'retweet_count': 5571, 'favorite_count': 0},
{'tweet_id': 821886076407029760,
 'retweet_count': 2692,
 'favorite_count': 12582},
{'tweet_id': 821813639212650496, 'retweet_count': 3839, 'favorite_count': 0},
{'tweet_id': 821765923262631936,
 'retweet_count': 1899,
 'favorite_count': 9317},
{'tweet_id': 821522889702862852,
 'retweet_count': 2030,
 'favorite_count': 8871},

{'tweet_id': 821421320206483457, 'retweet_count': 6470, 'favorite_count': 0},
{'tweet_id': 821407182352777218,
 'retweet_count': 5053,
 'favorite_count': 13075},
{'tweet_id': 821153421864615936, 'retweet_count': 10, 'favorite_count': 280},
{'tweet_id': 821149554670182400,
 'retweet_count': 2320,
 'favorite_count': 9718},
{'tweet_id': 821107785811234820,
 'retweet_count': 2487,
 'favorite_count': 10645},
{'tweet_id': 821044531881721856,
 'retweet_count': 2636,
 'favorite_count': 14021},
{'tweet_id': 820837357901512704, 'retweet_count': 7733, 'favorite_count': 0},
{'tweet_id': 820749716845686786,
 'retweet_count': 11525,
 'favorite_count': 34984},
{'tweet_id': 820690176645140481,
 'retweet_count': 3716,
 'favorite_count': 13518},
{'tweet_id': 820494788566847489, 'retweet_count': 648, 'favorite_count': 0},
{'tweet_id': 820446719150292993, 'retweet_count': 2980, 'favorite_count': 0},
{'tweet_id': 820314633777061888,
 'retweet_count': 648,
 'favorite_count': 3706},
{'tweet_id': 820078625395449857,
 'retweet_count': 7246,
 'favorite_count': 21979},
{'tweet_id': 820013781606658049, 'retweet_count': 1369, 'favorite_count': 0},
{'tweet_id': 819952236453363712,
 'retweet_count': 1369,
 'favorite_count': 5927},
{'tweet_id': 819924195358416896,
 'retweet_count': 5607,
 'favorite_count': 14305},
{'tweet_id': 819711362133872643,
 'retweet_count': 3604,
 'favorite_count': 14916},
{'tweet_id': 819588359383371776,
 'retweet_count': 2271,
 'favorite_count': 10606},
{'tweet_id': 819347104292290561,
 'retweet_count': 1383,
 'favorite_count': 8008},
{'tweet_id': 819238181065359361,
 'retweet_count': 462,

 'favorite_count': 2550},
{'tweet_id': 819227688460238848,
 'retweet_count': 7733,
 'favorite_count': 25652},
{'tweet_id': 819015337530290176, 'retweet_count': 42228, 'favorite_count': 0},
{'tweet_id': 819015331746349057, 'retweet_count': 21794, 'favorite_count': 0},
{'tweet_id': 819006400881917954,
 'retweet_count': 21794,
 'favorite_count': 49960},
{'tweet_id': 819004803107983360,
 'retweet_count': 42228,
 'favorite_count': 95450},
{'tweet_id': 818646164899774465, 'retweet_count': 24069, 'favorite_count': 0},
{'tweet_id': 818627210458333184,
 'retweet_count': 8564,
 'favorite_count': 24597},
{'tweet_id': 818614493328580609,
 'retweet_count': 2982,
 'favorite_count': 10971},
{'tweet_id': 818588835076603904, 'retweet_count': 7126, 'favorite_count': 0},
{'tweet_id': 818536468981415936,
 'retweet_count': 2873,
 'favorite_count': 12127},
{'tweet_id': 818307523543449600, 'retweet_count': 9366, 'favorite_count': 0},
{'tweet_id': 818259473185828864,
 'retweet_count': 2621,
 'favorite_count': 12197},
{'tweet_id': 818145370475810820,
 'retweet_count': 3014,
 'favorite_count': 13671},
{'tweet_id': 817908911860748288,
 'retweet_count': 906,
 'favorite_count': 5356},
{'tweet_id': 817827839487737858,
 'retweet_count': 31314,
 'favorite_count': 57622},
{'tweet_id': 817777686764523521,
 'retweet_count': 3084,
 'favorite_count': 11901},
{'tweet_id': 817536400337801217,
 'retweet_count': 3505,
 'favorite_count': 13105},
{'tweet_id': 817502432452313088, 'retweet_count': 3917, 'favorite_count': 0},
{'tweet_id': 817423860136083457,
 'retweet_count': 17504,
 'favorite_count': 38260},
{'tweet_id': 817415592588222464,

'retweet_count': 1131,
 'favorite_count': 6267},
{'tweet_id': 817181837579653120, 'retweet_count': 9907, 'favorite_count': 0},
{'tweet_id': 817171292965273600,
 'retweet_count': 2326,
 'favorite_count': 9690},
{'tweet_id': 817120970343411712,
 'retweet_count': 3011,
 'favorite_count': 13367},
{'tweet_id': 817056546584727552,
 'retweet_count': 1927,
 'favorite_count': 9517},
{'tweet_id': 816829038950027264, 'retweet_count': 5496, 'favorite_count': 0},
{'tweet_id': 816816676327063552,
 'retweet_count': 2361,
 'favorite_count': 11071},
{'tweet_id': 816697700272001025,
 'retweet_count': 2545,
 'favorite_count': 10905},
{'tweet_id': 816450570814898180,
 'retweet_count': 9366,
 'favorite_count': 33961},
{'tweet_id': 816336735214911488,
 'retweet_count': 2269,
 'favorite_count': 9564},
{'tweet_id': 816091915477250048,
 'retweet_count': 2500,
 'favorite_count': 9927},
{'tweet_id': 816062466425819140, 'retweet_count': 1207, 'favorite_count': 0},
{'tweet_id': 816014286006976512, 'retweet_count': 6154, 'favorite_count': 0},
{'tweet_id': 815990720817401858,
 'retweet_count': 1207,
 'favorite_count': 5545},
{'tweet_id': 815966073409433600,
 'retweet_count': 9907,
 'favorite_count': 25057},
{'tweet_id': 815745968457060357, 'retweet_count': 4797, 'favorite_count': 0},
{'tweet_id': 815736392542261248,
 'retweet_count': 2625,
 'favorite_count': 10937},
{'tweet_id': 815639385530101762,
 'retweet_count': 1918,
 'favorite_count': 9161},
{'tweet_id': 815390420867969024,
 'retweet_count': 4407,
 'favorite_count': 11467},
{'tweet_id': 814986499976527872,

  'retweet_count': 1505,
  'favorite_count': 8485},
{'tweet_id': 814638523311648768,
  'retweet_count': 3130,
  'favorite_count': 12511},
{'tweet_id': 814578408554463233, 'retweet_count': 6750, 'favorite_count': 0},
{'tweet_id': 814530161257443328,
  'retweet_count': 2156,
  'favorite_count': 9629},
{'tweet_id': 814153002265309185,
  'retweet_count': 10080,
  'favorite_count': 32000},
{'tweet_id': 813944609378369540, 'retweet_count': 3732, 'favorite_count': 0},
{'tweet_id': 813910438903693312,
  'retweet_count': 2194,
  'favorite_count': 10342},
{'tweet_id': 813812741911748608,
  'retweet_count': 16267,
  'favorite_count': 40402},
{'tweet_id': 813800681631023104,
  'retweet_count': 2060,
  'favorite_count': 9300},
{'tweet_id': 813217897535406080,
  'retweet_count': 8476,
  'favorite_count': 20783},
{'tweet_id': 813202720496779264,
  'retweet_count': 2090,
  'favorite_count': 10192},
{'tweet_id': 813187593374461952,
  'retweet_count': 5096,
  'favorite_count': 22085},
{'tweet_id': 813172488309972993,
  'retweet_count': 2236,
  'favorite_count': 10384},
{'tweet_id': 813157409116065792,
  'retweet_count': 2557,
  'favorite_count': 8588},
{'tweet_id': 813142292504645637,
  'retweet_count': 2728,
  'favorite_count': 9361},
{'tweet_id': 813130366689148928,
  'retweet_count': 513,
  'favorite_count': 4968},
{'tweet_id': 813127251579564032,
  'retweet_count': 3652,
  'favorite_count': 13242},
{'tweet_id': 813112105746448384,

'retweet_count': 3225,
 'favorite_count': 11515},
{'tweet_id': 813096984823349248,
 'retweet_count': 4207,
 'favorite_count': 11694},
{'tweet_id': 813081950185472002,
 'retweet_count': 3220,
 'favorite_count': 10989},
{'tweet_id': 813066809284972545,
 'retweet_count': 2276,
 'favorite_count': 8865},
{'tweet_id': 813051746834595840,
 'retweet_count': 8503,
 'favorite_count': 23337},
{'tweet_id': 812781120811126785,
 'retweet_count': 2191,
 'favorite_count': 8380},
{'tweet_id': 812747805718642688, 'retweet_count': 8067, 'favorite_count': 0},
{'tweet_id': 812709060537683968,
 'retweet_count': 1665,
 'favorite_count': 7373},
{'tweet_id': 812503143955202048,
 'retweet_count': 1424,
 'favorite_count': 6787},
{'tweet_id': 812466873996607488,
 'retweet_count': 2231,
 'favorite_count': 8900},
{'tweet_id': 812372279581671427,
 'retweet_count': 4252,
 'favorite_count': 15224},
{'tweet_id': 811985624773361665,
 'retweet_count': 1647,
 'favorite_count': 8102},
{'tweet_id': 811744202451197953,
 'retweet_count': 1884,
 'favorite_count': 8429},
{'tweet_id': 811647686436880384,
 'retweet_count': 872,
 'favorite_count': 6215},
{'tweet_id': 811627233043480576,
 'retweet_count': 3650,
 'favorite_count': 14265},
{'tweet_id': 811386762094317568,
 'retweet_count': 7444,
 'favorite_count': 23302},
{'tweet_id': 810984652412424192,
 'retweet_count': 1655,

  'favorite_count': 5927},
{'tweet_id': 810896069567610880,
 'retweet_count': 2090,
 'favorite_count': 10093},
{'tweet_id': 810657578271330305,
 'retweet_count': 3057,
 'favorite_count': 12192},
{'tweet_id': 810284430598270976,
 'retweet_count': 13369,
 'favorite_count': 39640},
{'tweet_id': 810254108431155201,
 'retweet_count': 3901,
 'favorite_count': 16380},
{'tweet_id': 809920764300447744,
 'retweet_count': 4521,
 'favorite_count': 17250},
{'tweet_id': 809808892968534016, 'retweet_count': 5711, 'favorite_count': 0},
{'tweet_id': 809448704142938112,
 'retweet_count': 1696,
 'favorite_count': 7727},
{'tweet_id': 809220051211603969,
 'retweet_count': 6554,
 'favorite_count': 22246},
{'tweet_id': 809084759137812480,
 'retweet_count': 4046,
 'favorite_count': 14685},
{'tweet_id': 808838249661788160,
 'retweet_count': 3536,
 'favorite_count': 11271},
{'tweet_id': 808733504066486276,
 'retweet_count': 2509,
 'favorite_count': 8784},
{'tweet_id': 808501579447930884,
 'retweet_count': 3007,
 'favorite_count': 12595},
{'tweet_id': 808344865868283904,
 'retweet_count': 24069,
 'favorite_count': 47281},
{'tweet_id': 808134635716833280, 'retweet_count': 6961, 'favorite_count': 0},
{'tweet_id': 808106460588765185,
 'retweet_count': 2525,
 'favorite_count': 9701},
{'tweet_id': 808001312164028416,
 'retweet_count': 4098,
 'favorite_count': 14015},
{'tweet_id': 807621403335917568,
 'retweet_count': 4288,

 'favorite_count': 16236},
{'tweet_id': 807106840509214720,
 'retweet_count': 56625,
 'favorite_count': 107015},
{'tweet_id': 807059379405148160, 'retweet_count': 8521, 'favorite_count': 0},
{'tweet_id': 807010152071229440,
 'retweet_count': 4418,
 'favorite_count': 14514},
{'tweet_id': 806629075125202948,
 'retweet_count': 37911,
 'favorite_count': 75639},
{'tweet_id': 806620845233815552, 'retweet_count': 6426, 'favorite_count': 0},
{'tweet_id': 806576416489959424,
 'retweet_count': 2230,
 'favorite_count': 5370},
{'tweet_id': 806542213899489280,
 'retweet_count': 2752,
 'favorite_count': 11363},
{'tweet_id': 806242860592926720, 'retweet_count': 13616, 'favorite_count': 0},
{'tweet_id': 806219024703037440,
 'retweet_count': 1388,
 'favorite_count': 7145},
{'tweet_id': 805958939288408065, 'retweet_count': 6238, 'favorite_count': 0},
{'tweet_id': 805932879469572096,
 'retweet_count': 2209,
 'favorite_count': 9178},
{'tweet_id': 805826884734976000,
 'retweet_count': 2132,
 'favorite_count': 7335},
{'tweet_id': 805823200554876929, 'retweet_count': 9374, 'favorite_count': 0},
{'tweet_id': 805520635690676224,
 'retweet_count': 1905,
 'favorite_count': 6368},
{'tweet_id': 805487436403003392,
 'retweet_count': 2929,
 'favorite_count': 9773},
{'tweet_id': 805207613751304193,
 'retweet_count': 1972,
 'favorite_count': 8680},
{'tweet_id': 804738756058218496,
 'retweet_count': 4480,
 'favorite_count': 15326},
{'tweet_id': 804475857670639616,
 'retweet_count': 2355,
 'favorite_count': 6886},
{'tweet_id': 804413760345620481, 'retweet_count': 3712, 'favorite_count': 0},
{'tweet_id': 804026241225523202,

 'retweet_count': 18876,
 'favorite_count': 49774},
{'tweet_id': 803773340896923648,
 'retweet_count': 3220,
 'favorite_count': 11203},
{'tweet_id': 803692223237865472, 'retweet_count': 8689, 'favorite_count': 0},
{'tweet_id': 803638050916102144,
 'retweet_count': 4828,
 'favorite_count': 12270},
{'tweet_id': 803380650405482500,
 'retweet_count': 2169,
 'favorite_count': 8601},
{'tweet_id': 803321560782307329, 'retweet_count': 6221, 'favorite_count': 0},
{'tweet_id': 803276597545603072,
 'retweet_count': 2887,
 'favorite_count': 11207},
{'tweet_id': 802952499103731712,
 'retweet_count': 2336,
 'favorite_count': 10085},
{'tweet_id': 802624713319034886, 'retweet_count': 3403, 'favorite_count': 0},
{'tweet_id': 802600418706604034,
 'retweet_count': 1714,
 'favorite_count': 7938},
{'tweet_id': 802572683846291456,
 'retweet_count': 2926,
 'favorite_count': 9959},
{'tweet_id': 802323869084381190,
 'retweet_count': 6808,
 'favorite_count': 18124},
{'tweet_id': 802265048156610565,
 'retweet_count': 1573,
 'favorite_count': 7039},
{'tweet_id': 802247111496568832, 'retweet_count': 5247, 'favorite_count': 0},
{'tweet_id': 802239329049477120,
 'retweet_count': 3040,
 'favorite_count': 10132},
{'tweet_id': 802185808107208704, 'retweet_count': 314, 'favorite_count': 0},
{'tweet_id': 801958328846974976,
 'retweet_count': 1992,
 'favorite_count': 8608},
{'tweet_id': 801854953262350336,
 'retweet_count': 275,
 'favorite_count': 1958},
{'tweet_id': 801538201127157760,
 'retweet_count': 2391,
 'favorite_count': 9141},
{'tweet_id': 801285448605831168,

 'retweet_count': 960,
 'favorite_count': 6802},
{'tweet_id': 801167903437357056,
 'retweet_count': 6961,
 'favorite_count': 27386},
{'tweet_id': 801127390143516673, 'retweet_count': 4627, 'favorite_count': 0},
{'tweet_id': 801115127852503040,
 'retweet_count': 2429,
 'favorite_count': 8992},
{'tweet_id': 800859414831898624, 'retweet_count': 113, 'favorite_count': 778},
{'tweet_id': 800855607700029440, 'retweet_count': 1889, 'favorite_count': 0},
{'tweet_id': 800751577355128832,
 'retweet_count': 3214,
 'favorite_count': 11701},
{'tweet_id': 800513324630806528,
 'retweet_count': 3495,
 'favorite_count': 14685},
{'tweet_id': 800459316964663297,
 'retweet_count': 2489,
 'favorite_count': 10538},
{'tweet_id': 800443802682937345, 'retweet_count': 5068, 'favorite_count': 0},
{'tweet_id': 800388270626521089,
 'retweet_count': 3265,
 'favorite_count': 12456},
{'tweet_id': 800188575492947969, 'retweet_count': 4581, 'favorite_count': 0},
{'tweet_id': 800141422401830912,
 'retweet_count': 2980,
 'favorite_count': 17092},
{'tweet_id': 800018252395122689,
 'retweet_count': 15351,
 'favorite_count': 31768},
{'tweet_id': 799774291445383169, 'retweet_count': 5488, 'favorite_count': 0},
{'tweet_id': 799757965289017345,
 'retweet_count': 2506,
 'favorite_count': 9390},
{'tweet_id': 799422933579902976,
 'retweet_count': 2213,
 'favorite_count': 8965},
{'tweet_id': 799308762079035393, 'retweet_count': 6478, 'favorite_count': 0},
{'tweet_id': 799297110730567681,
 'retweet_count': 3227,
 'favorite_count': 11065},
{'tweet_id': 799063482566066176,
 'retweet_count': 2863,
 'favorite_count': 9058},
{'tweet_id': 798933969379225600,
 'retweet_count': 5203,

 'favorite_count': 14712},
{'tweet_id': 798925684722855936,
 'retweet_count': 1663,
 'favorite_count': 8246},
{'tweet_id': 798705661114773508, 'retweet_count': 7724, 'favorite_count': 0},
{'tweet_id': 798701998996647937, 'retweet_count': 9126, 'favorite_count': 0},
{'tweet_id': 798697898615730177, 'retweet_count': 7602, 'favorite_count': 0},
{'tweet_id': 798694562394996736, 'retweet_count': 5812, 'favorite_count': 0},
{'tweet_id': 798686750113755136, 'retweet_count': 2702, 'favorite_count': 0},
{'tweet_id': 798682547630837760, 'retweet_count': 5515, 'favorite_count': 0},
{'tweet_id': 798673117451325440, 'retweet_count': 6485, 'favorite_count': 0},
{'tweet_id': 798665375516884993, 'retweet_count': 4519, 'favorite_count': 0},
{'tweet_id': 798644042770751489, 'retweet_count': 2153, 'favorite_count': 0},
{'tweet_id': 798628517273620480, 'retweet_count': 2272, 'favorite_count': 0},
{'tweet_id': 798585098161549313, 'retweet_count': 6618, 'favorite_count': 0},
{'tweet_id': 798576900688019456, 'retweet_count': 6871, 'favorite_count': 0},
{'tweet_id': 798340744599797760, 'retweet_count': 3924, 'favorite_count': 0},
{'tweet_id': 798209839306514432,
 'retweet_count': 2954,
 'favorite_count': 11548},
{'tweet_id': 797971864723324932,
 'retweet_count': 3652,
 'favorite_count': 13018},
{'tweet_id': 797545162159308800,
 'retweet_count': 5656,
 'favorite_count': 16198},
{'tweet_id': 797236660651966464,
 'retweet_count': 7726,
 'favorite_count': 22328},
{'tweet_id': 797165961484890113, 'retweet_count': 32, 'favorite_count': 256},
{'tweet_id': 796904159865868288, 'retweet_count': 10411, 'favorite_count': 0},
{'tweet_id': 796865951799083009,
 'retweet_count': 2198,
 'favorite_count': 8564},
{'tweet_id': 796759840936919040,
 'retweet_count': 3562,
 'favorite_count': 13256},
{'tweet_id': 796563435802726400, 'retweet_count': 8536, 'favorite_count': 0},
{'tweet_id': 796484825502875648,
 'retweet_count': 2042,
 'favorite_count': 8472},
{'tweet_id': 796387464403357696,
 'retweet_count': 4861,
 'favorite_count': 12334},
{'tweet_id': 796177847564038144, 'retweet_count': 16628, 'favorite_count': 0},
{'tweet_id': 796149749086875649,
 'retweet_count': 16628,

'favorite_count': 36177},
{'tweet_id': 796125600683540480,
 'retweet_count': 2079,
 'favorite_count': 5511},
{'tweet_id': 796116448414461957,
 'retweet_count': 2813,
 'favorite_count': 10139},
{'tweet_id': 796080075804475393,
 'retweet_count': 2703,
 'favorite_count': 9469},
{'tweet_id': 796031486298386433,
 'retweet_count': 4284,
 'favorite_count': 12071},
{'tweet_id': 795464331001561088,
 'retweet_count': 27728,
 'favorite_count': 55683},
{'tweet_id': 795400264262053889,
 'retweet_count': 3323,
 'favorite_count': 11270},
{'tweet_id': 795076730285391872,
 'retweet_count': 6288,
 'favorite_count': 18139},
{'tweet_id': 794983741416415232, 'retweet_count': 3942, 'favorite_count': 0},
{'tweet_id': 794926597468000259,
 'retweet_count': 2697,
 'favorite_count': 11492},
{'tweet_id': 794355576146903043, 'retweet_count': 12014, 'favorite_count': 0},
{'tweet_id': 794332329137291264,
 'retweet_count': 3088,
 'favorite_count': 10686},
{'tweet_id': 794205286408003585,
 'retweet_count': 3895,
 'favorite_count': 10314},
{'tweet_id': 793962221541933056,
 'retweet_count': 5711,
 'favorite_count': 18910},
{'tweet_id': 793845145112371200,
 'retweet_count': 2187,
 'favorite_count': 10295},
{'tweet_id': 793614319594401792, 'retweet_count': 3661, 'favorite_count': 0},
{'tweet_id': 793601777308463104,
 'retweet_count': 1908,
 'favorite_count': 8926},
{'tweet_id': 793500921481273345,
 'retweet_count': 2786,
 'favorite_count': 11953},
{'tweet_id': 793286476301799424,

 'retweet_count': 10723,
 'favorite_count': 27597},
{'tweet_id': 793271401113350145,
 'retweet_count': 2763,
 'favorite_count': 9677},
{'tweet_id': 793256262322548741,
 'retweet_count': 9714,
 'favorite_count': 22350},
{'tweet_id': 793241302385262592,
 'retweet_count': 3812,
 'favorite_count': 11780},
{'tweet_id': 793226087023144960,
 'retweet_count': 3338,
 'favorite_count': 11001},
{'tweet_id': 793210959003287553,
 'retweet_count': 3301,
 'favorite_count': 9997},
{'tweet_id': 793195938047070209,
 'retweet_count': 6547,
 'favorite_count': 17063},
{'tweet_id': 793180763617361921,
 'retweet_count': 2310,
 'favorite_count': 7740},
{'tweet_id': 793165685325201412,
 'retweet_count': 3238,
 'favorite_count': 10478},
{'tweet_id': 793150605191548928,
 'retweet_count': 1984,
 'favorite_count': 6909},
{'tweet_id': 793135492858580992,
 'retweet_count': 2893,
 'favorite_count': 7214},
{'tweet_id': 793120401413079041,
 'retweet_count': 4551,
 'favorite_count': 14202},
{'tweet_id': 792913359805018113,
 'retweet_count': 4715,
 'favorite_count': 16063},
{'tweet_id': 792883833364439040,
 'retweet_count': 4964,
 'favorite_count': 12666},
{'tweet_id': 792773781206999040,
 'retweet_count': 1963,
 'favorite_count': 8209},
{'tweet_id': 792394556390137856,
 'retweet_count': 4998,
 'favorite_count': 15128},

```
{'tweet_id': 792050063153438720,
 'retweet_count': 2088,
 'favorite_count': 8029},
{'tweet_id': 791821351946420224, 'retweet_count': 24514, 'favorite_count': 0},
{'tweet_id': 791784077045166082, 'retweet_count': 11271, 'favorite_count': 0},
{'tweet_id': 791780927877898241, 'retweet_count': 4432, 'favorite_count': 0},
{'tweet_id': 791774931465953280,
 'retweet_count': 21156,
 'favorite_count': 37818},
{'tweet_id': 791672322847637504,
 'retweet_count': 3661,
 'favorite_count': 13129},
{'tweet_id': 791406955684368384,
 'retweet_count': 4797,
 'favorite_count': 14670},
{'tweet_id': 791312159183634433,
 'retweet_count': 2963,
 'favorite_count': 9841},
{'tweet_id': 791026214425268224, 'retweet_count': 4858, 'favorite_count': 0},
{'tweet_id': 790987426131050500,
 'retweet_count': 2483,
 'favorite_count': 11089},
{'tweet_id': 790946055508652032,
 'retweet_count': 5496,
 'favorite_count': 18601},
{'tweet_id': 790723298204217344, 'retweet_count': 2704, 'favorite_count': 0},
{'tweet_id': 790698755171364864,
 'retweet_count': 2203,
 'favorite_count': 9158},
{'tweet_id': 790581949425475584,
 'retweet_count': 8183,
 'favorite_count': 22473},
{'tweet_id': 790337589677002753,
 'retweet_count': 2167,
 'favorite_count': 8740},
{'tweet_id': 790277117346975746,
 'retweet_count': 3732,
 'favorite_count': 14081},
{'tweet_id': 790227638568808452, 'retweet_count': 4190, 'favorite_count': 0},
{'tweet_id': 789986466051088384,
 'retweet_count': 2704,
 'favorite_count': 10369},
{'tweet_id': 789960241177853952, 'retweet_count': 4839, 'favorite_count': 0},
{'tweet_id': 789903600034189313,
 'retweet_count': 4858,
 'favorite_count': 11673},
{'tweet_id': 789628658055020548,
```

  'retweet_count': 2080,
  'favorite_count': 8448},
{'tweet_id': 789599242079838210,
  'retweet_count': 2279,
  'favorite_count': 7620},
{'tweet_id': 789530877013393408,
  'retweet_count': 3942,
  'favorite_count': 13188},
{'tweet_id': 789314372632018944,
  'retweet_count': 2666,
  'favorite_count': 9615},
{'tweet_id': 789280767834746880, 'retweet_count': 5747, 'favorite_count': 0},
{'tweet_id': 789268448748703744,
  'retweet_count': 3014,
  'favorite_count': 10196},
{'tweet_id': 789137962068021249,
  'retweet_count': 3244,
  'favorite_count': 10875},
{'tweet_id': 788908386943430656,
  'retweet_count': 14409,
  'favorite_count': 30653},
{'tweet_id': 788765914992902144,
  'retweet_count': 12014,
  'favorite_count': 30658},
{'tweet_id': 788552643979468800, 'retweet_count': 8407, 'favorite_count': 0},
{'tweet_id': 788412144018661376,
  'retweet_count': 5990,
  'favorite_count': 16060},
{'tweet_id': 788178268662984705,
  'retweet_count': 2488,
  'favorite_count': 8100},
{'tweet_id': 788150585577050112,
  'retweet_count': 1510,
  'favorite_count': 6865},
{'tweet_id': 788070120937619456, 'retweet_count': 3952, 'favorite_count': 0},
{'tweet_id': 788039637453406209,
  'retweet_count': 1535,
  'favorite_count': 6867},
{'tweet_id': 787810552592695296,
  'retweet_count': 3540,
  'favorite_count': 9717},
{'tweet_id': 787717603741622272,
  'retweet_count': 3240,
  'favorite_count': 11416},
{'tweet_id': 787397959788929025,
  'retweet_count': 3300,
  'favorite_count': 12120},

{'tweet_id': 787322443945877504,
 'retweet_count': 2032,
 'favorite_count': 8726},
{'tweet_id': 787111942498508800, 'retweet_count': 7798, 'favorite_count': 0},
{'tweet_id': 786963064373534720,
 'retweet_count': 9327,
 'favorite_count': 29725},
{'tweet_id': 786729988674449408, 'retweet_count': 2827, 'favorite_count': 0},
{'tweet_id': 786709082849828864,
 'retweet_count': 7069,
 'favorite_count': 20296},
{'tweet_id': 786664955043049472,
 'retweet_count': 2996,
 'favorite_count': 11957},
{'tweet_id': 786595970293370880,
 'retweet_count': 3601,
 'favorite_count': 10497},
{'tweet_id': 786363235746385920,
 'retweet_count': 4072,
 'favorite_count': 12189},
{'tweet_id': 786286427768250368,
 'retweet_count': 3352,
 'favorite_count': 8930},
{'tweet_id': 786233965241827333,
 'retweet_count': 5571,
 'favorite_count': 17178},
{'tweet_id': 786051337297522688,
 'retweet_count': 179,
 'favorite_count': 1614},
{'tweet_id': 786036967502913536, 'retweet_count': 8237, 'favorite_count': 0},
{'tweet_id': 785927819176054784,
 'retweet_count': 3652,
 'favorite_count': 12696},
{'tweet_id': 785872687017132033,
 'retweet_count': 2130,
 'favorite_count': 7489},
{'tweet_id': 785639753186217984,
 'retweet_count': 2561,
 'favorite_count': 8735},
{'tweet_id': 785533386513321988,
 'retweet_count': 2334,
 'favorite_count': 10183},
{'tweet_id': 785515384317313025,
 'retweet_count': 1477,
 'favorite_count': 6800},
{'tweet_id': 785264754247995392,
 'retweet_count': 1911,

'favorite_count': 8128},
{'tweet_id': 785170936622350336,
 'retweet_count': 5639,
 'favorite_count': 13491},
{'tweet_id': 784826020293709826,
 'retweet_count': 3712,
 'favorite_count': 11310},
{'tweet_id': 784517518371221505,
 'retweet_count': 2970,
 'favorite_count': 10039},
{'tweet_id': 784431430411685888,
 'retweet_count': 1491,
 'favorite_count': 6329},
{'tweet_id': 784183165795655680,
 'retweet_count': 9374,
 'favorite_count': 22513},
{'tweet_id': 784057939640352768,
 'retweet_count': 12953,
 'favorite_count': 33505},
{'tweet_id': 783839966405230592,
 'retweet_count': 12643,
 'favorite_count': 33689},
{'tweet_id': 783821107061198850,
 'retweet_count': 2269,
 'favorite_count': 8209},
{'tweet_id': 783695101801398276,
 'retweet_count': 3737,
 'favorite_count': 11650},
{'tweet_id': 783466772167098368,
 'retweet_count': 2608,
 'favorite_count': 9468},
{'tweet_id': 783391753726550016,
 'retweet_count': 6441,
 'favorite_count': 18908},
{'tweet_id': 783347506784731136, 'retweet_count': 6533, 'favorite_count': 0},
{'tweet_id': 783334639985389568,
 'retweet_count': 13616,
 'favorite_count': 32651},
{'tweet_id': 783085703974514689,
 'retweet_count': 2565,
 'favorite_count': 9112},
{'tweet_id': 782969140009107456,
 'retweet_count': 8521,
 'favorite_count': 26949},
{'tweet_id': 782747134529531904,
 'retweet_count': 1604,
 'favorite_count': 8310},

```
{'tweet_id': 782722598790725632,
 'retweet_count': 6238,
 'favorite_count': 19250},
{'tweet_id': 782598640137187329,
 'retweet_count': 2184,
 'favorite_count': 8694},
{'tweet_id': 782305867769217024,
 'retweet_count': 6470,
 'favorite_count': 18630},
{'tweet_id': 782021823840026624, 'retweet_count': 7236, 'favorite_count': 0},
{'tweet_id': 781955203444699136,
 'retweet_count': 3920,
 'favorite_count': 12357},
{'tweet_id': 781661882474196992,
 'retweet_count': 3129,
 'favorite_count': 11634},
{'tweet_id': 781655249211752448,
 'retweet_count': 1314,
 'favorite_count': 4466},
{'tweet_id': 781524693396357120,
 'retweet_count': 6426,
 'favorite_count': 23163},
{'tweet_id': 781308096455073793,
 'retweet_count': 2961,
 'favorite_count': 7973},
{'tweet_id': 781251288990355457,
 'retweet_count': 2458,
 'favorite_count': 9428},
{'tweet_id': 781163403222056960,
 'retweet_count': 3168,
 'favorite_count': 10895},
{'tweet_id': 780931614150983680,
 'retweet_count': 8536,
 'favorite_count': 24192},
{'tweet_id': 780858289093574656,
 'retweet_count': 2328,
 'favorite_count': 8033},
{'tweet_id': 780800785462489090,
 'retweet_count': 1505,
 'favorite_count': 6141},
{'tweet_id': 780601303617732608,
 'retweet_count': 3839,
 'favorite_count': 13525},
{'tweet_id': 780543529827336192,
 'retweet_count': 2032,
 'favorite_count': 7040},
{'tweet_id': 780496263422808064, 'retweet_count': 4319, 'favorite_count': 0},
```

{'tweet_id': 780476555013349377, 'retweet_count': 136, 'favorite_count': 0},
{'tweet_id': 780459368902959104,
 'retweet_count': 1224,
 'favorite_count': 5892},
{'tweet_id': 780192070812196864,
 'retweet_count': 2589,
 'favorite_count': 9712},
{'tweet_id': 780092040432480260, 'retweet_count': 2655, 'favorite_count': 0},
{'tweet_id': 780074436359819264,
 'retweet_count': 5815,
 'favorite_count': 13723},
{'tweet_id': 779834332596887552,
 'retweet_count': 8237,
 'favorite_count': 21252},
{'tweet_id': 779377524342161408,
 'retweet_count': 3613,
 'favorite_count': 9831},
{'tweet_id': 779124354206535695, 'retweet_count': 21324, 'favorite_count': 0},
{'tweet_id': 779123168116150273,
 'retweet_count': 4207,
 'favorite_count': 13206},
{'tweet_id': 779056095788752897,
 'retweet_count': 5247,
 'favorite_count': 16500},
{'tweet_id': 778990705243029504,
 'retweet_count': 8437,
 'favorite_count': 22342},
{'tweet_id': 778774459159379968, 'retweet_count': 11550, 'favorite_count': 0},
{'tweet_id': 778764940568104960, 'retweet_count': 426, 'favorite_count': 955},
{'tweet_id': 778748913645780993,
 'retweet_count': 1562,
 'favorite_count': 7717},
{'tweet_id': 778650543019483137,
 'retweet_count': 1729,
 'favorite_count': 6430},
{'tweet_id': 778624900596654080,
 'retweet_count': 1176,
 'favorite_count': 5177},
{'tweet_id': 778408200802557953,
 'retweet_count': 5023,
 'favorite_count': 15135},
{'tweet_id': 778396591732486144, 'retweet_count': 14198, 'favorite_count': 0},
{'tweet_id': 778383385161035776,
 'retweet_count': 1271,
 'favorite_count': 6515},
{'tweet_id': 778286810187399168,
 'retweet_count': 3836,

   'favorite_count': 11576},
{'tweet_id': 778039087836069888,
 'retweet_count': 3065,
 'favorite_count': 9417},
{'tweet_id': 778027034220126208,
 'retweet_count': 1885,
 'favorite_count': 7320},
{'tweet_id': 777953400541634568, 'retweet_count': 4080, 'favorite_count': 0},
{'tweet_id': 777885040357281792,
 'retweet_count': 1893,
 'favorite_count': 7078},
{'tweet_id': 777684233540206592,
 'retweet_count': 3403,
 'favorite_count': 12518},
{'tweet_id': 777641927919427584, 'retweet_count': 4947, 'favorite_count': 0},
{'tweet_id': 777621514455814149,
 'retweet_count': 2910,
 'favorite_count': 9742},
{'tweet_id': 777189768882946048,
 'retweet_count': 4999,
 'favorite_count': 15921},
{'tweet_id': 776819012571455488, 'retweet_count': 15839, 'favorite_count': 0},
{'tweet_id': 776813020089548800,
 'retweet_count': 1405,
 'favorite_count': 5444},
{'tweet_id': 776477788987613185,
 'retweet_count': 3249,
 'favorite_count': 9858},
{'tweet_id': 776249906839351296, 'retweet_count': 10673, 'favorite_count': 0},
{'tweet_id': 776218204058357768,
 'retweet_count': 18497,
 'favorite_count': 33345},
{'tweet_id': 776201521193218049,
 'retweet_count': 2919,
 'favorite_count': 10681},
{'tweet_id': 776113305656188928,
 'retweet_count': 5068,
 'favorite_count': 13102},
{'tweet_id': 776088319444877312,
 'retweet_count': 179,
 'favorite_count': 2045},
{'tweet_id': 775898661951791106, 'retweet_count': 17621, 'favorite_count': 0},
{'tweet_id': 775842724423557120,
 'retweet_count': 3116,
 'favorite_count': 13022},
{'tweet_id': 775733305207554048,
 'retweet_count': 4627,

'favorite_count': 15413},
{'tweet_id': 775729183532220416,
 'retweet_count': 5302,
 'favorite_count': 14361},
{'tweet_id': 775364825476165632,
 'retweet_count': 3472,
 'favorite_count': 8295},
{'tweet_id': 775350846108426240,
 'retweet_count': 4532,
 'favorite_count': 11309},
{'tweet_id': 775096608509886464, 'retweet_count': 9220, 'favorite_count': 0},
{'tweet_id': 775085132600442880,
 'retweet_count': 5488,
 'favorite_count': 17281},
{'tweet_id': 774757898236878852,
 'retweet_count': 2035,
 'favorite_count': 9462},
{'tweet_id': 774639387460112384,
 'retweet_count': 2013,
 'favorite_count': 7508},
{'tweet_id': 774314403806253056,
 'retweet_count': 6478,
 'favorite_count': 24167},
{'tweet_id': 773985732834758656,
 'retweet_count': 4492,
 'favorite_count': 11925},
{'tweet_id': 773922284943896577,
 'retweet_count': 1999,
 'favorite_count': 7110},
{'tweet_id': 773704687002451968,
 'retweet_count': 1891,
 'favorite_count': 7317},
{'tweet_id': 773670353721753600,
 'retweet_count': 1501,
 'favorite_count': 5935},
{'tweet_id': 773547596996571136,
 'retweet_count': 7126,
 'favorite_count': 24553},
{'tweet_id': 773336787167145985, 'retweet_count': 5912, 'favorite_count': 0},
{'tweet_id': 773308824254029826,
 'retweet_count': 8640,
 'favorite_count': 25846},
{'tweet_id': 773247561583001600,
 'retweet_count': 3699,
 'favorite_count': 10414},
{'tweet_id': 773191612633579521,
 'retweet_count': 4736,

```
 'favorite_count': 11117},
{'tweet_id': 772877495989305348,
 'retweet_count': 4501,
 'favorite_count': 9555},
{'tweet_id': 772826264096874500,
 'retweet_count': 2669,
 'favorite_count': 8842},
{'tweet_id': 772615324260794368, 'retweet_count': 3914, 'favorite_count': 0},
{'tweet_id': 772581559778025472,
 'retweet_count': 1968,
 'favorite_count': 7192},
{'tweet_id': 772193107915964416,
 'retweet_count': 1612,
 'favorite_count': 6665},
{'tweet_id': 772152991789019136,
 'retweet_count': 1300,
 'favorite_count': 4181},
{'tweet_id': 772117678702071809,
 'retweet_count': 848,
 'favorite_count': 4165},
{'tweet_id': 772114945936949249,
 'retweet_count': 546,
 'favorite_count': 3005},
{'tweet_id': 772102971039580160,
 'retweet_count': 1065,
 'favorite_count': 4448},
{'tweet_id': 771908950375665664,
 'retweet_count': 2181,
 'favorite_count': 7298},
{'tweet_id': 771770456517009408,
 'retweet_count': 3924,
 'favorite_count': 13356},
{'tweet_id': 771500966810099713,
 'retweet_count': 3018,
 'favorite_count': 9167},
{'tweet_id': 771380798096281600,
 'retweet_count': 5912,
 'favorite_count': 11746},
{'tweet_id': 771171053431250945, 'retweet_count': 8705, 'favorite_count': 0},
{'tweet_id': 771136648247640064,
 'retweet_count': 3357,
 'favorite_count': 10223},
{'tweet_id': 771102124360998913,
 'retweet_count': 1663,
 'favorite_count': 6898},
{'tweet_id': 771014301343748096,
 'retweet_count': 1782,
```

'favorite_count': 7032},
{'tweet_id': 770787852854652928,
 'retweet_count': 1415,
 'favorite_count': 5498},
{'tweet_id': 770772759874076672,
 'retweet_count': 1626,
 'favorite_count': 5749},
{'tweet_id': 770743923962707968, 'retweet_count': 52360, 'favorite_count': 0},
{'tweet_id': 770655142660169732,
 'retweet_count': 2013,
 'favorite_count': 8130},
{'tweet_id': 770414278348247044,
 'retweet_count': 2390,
 'favorite_count': 6987},
{'tweet_id': 770293558247038976,
 'retweet_count': 1718,
 'favorite_count': 6923},
{'tweet_id': 770093767776997377, 'retweet_count': 3520, 'favorite_count': 0},
{'tweet_id': 770069151037685760,
 'retweet_count': 2651,
 'favorite_count': 8385},
{'tweet_id': 769940425801170949,
 'retweet_count': 11131,
 'favorite_count': 34948},
{'tweet_id': 769695466921623552,
 'retweet_count': 1926,
 'favorite_count': 7101},
{'tweet_id': 769335591808995329, 'retweet_count': 8830, 'favorite_count': 0},
{'tweet_id': 769212283578875904,
 'retweet_count': 1969,
 'favorite_count': 5980},
{'tweet_id': 768970937022709760,
 'retweet_count': 7574,
 'favorite_count': 16017},
{'tweet_id': 768909767477751808, 'retweet_count': 3129, 'favorite_count': 0},
{'tweet_id': 768855141948723200,
 'retweet_count': 1034,
 'favorite_count': 4660},
{'tweet_id': 768609597686943744,
 'retweet_count': 1382,
 'favorite_count': 4580},
{'tweet_id': 768596291618299904,
 'retweet_count': 1473,
 'favorite_count': 5592},
{'tweet_id': 768554158521745409, 'retweet_count': 6719, 'favorite_count': 0},
{'tweet_id': 768473857036525572,
 'retweet_count': 3958,

 'favorite_count': 15110},
{'tweet_id': 768193404517830656,
 'retweet_count': 4080,
 'favorite_count': 12157},
{'tweet_id': 767884188863397888,
 'retweet_count': 1634,
 'favorite_count': 5309},
{'tweet_id': 767754930266464257,
 'retweet_count': 6221,
 'favorite_count': 17814},
{'tweet_id': 767500508068192258,
 'retweet_count': 2688,
 'favorite_count': 8295},
{'tweet_id': 767191397493538821,
 'retweet_count': 4406,
 'favorite_count': 13643},
{'tweet_id': 767122157629476866,
 'retweet_count': 3261,
 'favorite_count': 11227},
{'tweet_id': 766864461642756096, 'retweet_count': 6521, 'favorite_count': 0},
{'tweet_id': 766793450729734144,
 'retweet_count': 1565,
 'favorite_count': 5650},
{'tweet_id': 766714921925144576,
 'retweet_count': 438,
 'favorite_count': 2872},
{'tweet_id': 766693177336135680,
 'retweet_count': 918,
 'favorite_count': 4484},
{'tweet_id': 766423258543644672,
 'retweet_count': 1825,
 'favorite_count': 6671},
{'tweet_id': 766313316352462849,
 'retweet_count': 2166,
 'favorite_count': 7493},
{'tweet_id': 766078092750233600, 'retweet_count': 2970, 'favorite_count': 0},
{'tweet_id': 766069199026450432,
 'retweet_count': 1003,
 'favorite_count': 4765},
{'tweet_id': 766008592277377025,
 'retweet_count': 571,
 'favorite_count': 4149},
{'tweet_id': 765719909049503744,
 'retweet_count': 2475,
 'favorite_count': 8021},
{'tweet_id': 765669560888528897,
 'retweet_count': 1407,

 'favorite_count': 5760},
{'tweet_id': 765395769549590528,
 'retweet_count': 3127,
 'favorite_count': 20539},
{'tweet_id': 765371061932261376,
 'retweet_count': 2475,
 'favorite_count': 7842},
{'tweet_id': 765222098633691136,
 'retweet_count': 3914,
 'favorite_count': 12902},
{'tweet_id': 764857477905154048,
 'retweet_count': 2029,
 'favorite_count': 7099},
{'tweet_id': 764259802650378240,
 'retweet_count': 1745,
 'favorite_count': 6718},
{'tweet_id': 763956972077010945, 'retweet_count': 61, 'favorite_count': 812},
{'tweet_id': 763837565564780549,
 'retweet_count': 4858,
 'favorite_count': 14041},
{'tweet_id': 763183847194451968,
 'retweet_count': 1702,
 'favorite_count': 6004},
{'tweet_id': 763167063695355904, 'retweet_count': 3484, 'favorite_count': 0},
{'tweet_id': 763103485927849985,
 'retweet_count': 2602,
 'favorite_count': 8163},
{'tweet_id': 762699858130116608,
 'retweet_count': 4190,
 'favorite_count': 13518},
{'tweet_id': 762471784394268675,
 'retweet_count': 7612,
 'favorite_count': 12571},
{'tweet_id': 762464539388485633,
 'retweet_count': 4839,
 'favorite_count': 11503},
{'tweet_id': 762316489655476224,
 'retweet_count': 1298,
 'favorite_count': 5350},
{'tweet_id': 762035686371364864,
 'retweet_count': 17919,
 'favorite_count': 35400},
{'tweet_id': 761976711479193600,
 'retweet_count': 2310,
 'favorite_count': 5992},
{'tweet_id': 761750502866649088, 'retweet_count': 4535, 'favorite_count': 0},
{'tweet_id': 761745352076779520,

'retweet_count': 979,
 'favorite_count': 4707},
{'tweet_id': 761672994376806400,
 'retweet_count': 33421,
 'favorite_count': 55016},
{'tweet_id': 761599872357261312,
 'retweet_count': 1336,
 'favorite_count': 4578},
{'tweet_id': 761371037149827077, 'retweet_count': 20500, 'favorite_count': 0},
{'tweet_id': 761334018830917632,
 'retweet_count': 1669,
 'favorite_count': 5792},
{'tweet_id': 761292947749015552,
 'retweet_count': 1265,
 'favorite_count': 4957},
{'tweet_id': 761227390836215808,
 'retweet_count': 1775,
 'favorite_count': 5908},
{'tweet_id': 761004547850530816,
 'retweet_count': 3952,
 'favorite_count': 12482},
{'tweet_id': 760893934457552897,
 'retweet_count': 1104,
 'favorite_count': 4228},
{'tweet_id': 760656994973933572,
 'retweet_count': 2210,
 'favorite_count': 7343},
{'tweet_id': 760641137271070720,
 'retweet_count': 1481,
 'favorite_count': 5485},
{'tweet_id': 760539183865880579,
 'retweet_count': 4168,
 'favorite_count': 8399},
{'tweet_id': 760521673607086080,
 'retweet_count': 1608,
 'favorite_count': 4681},
{'tweet_id': 760290219849637889,
 'retweet_count': 13140,
 'favorite_count': 29618},
{'tweet_id': 760252756032651264,
 'retweet_count': 995,
 'favorite_count': 4338},
{'tweet_id': 760190180481531904,
 'retweet_count': 2079,
 'favorite_count': 6334},
{'tweet_id': 760153949710192640, 'retweet_count': 38, 'favorite_count': 0},
{'tweet_id': 759943073749200896,

```
 'retweet_count': 2382,
 'favorite_count': 6581},
{'tweet_id': 759923798737051648,
 'retweet_count': 6521,
 'favorite_count': 16284},
{'tweet_id': 759846353224826880,
 'retweet_count': 2265,
 'favorite_count': 7433},
{'tweet_id': 759793422261743616,
 'retweet_count': 2173,
 'favorite_count': 6620},
{'tweet_id': 759566828574212096, 'retweet_count': 24319, 'favorite_count': 0},
{'tweet_id': 759557299618865152,
 'retweet_count': 1341,
 'favorite_count': 5202},
{'tweet_id': 759447681597108224,
 'retweet_count': 2827,
 'favorite_count': 9418},
{'tweet_id': 759446261539934208,
 'retweet_count': 561,
 'favorite_count': 1846},
{'tweet_id': 759197388317847553,
 'retweet_count': 2221,
 'favorite_count': 6725},
{'tweet_id': 759159934323924993, 'retweet_count': 1359, 'favorite_count': 0},
{'tweet_id': 759099523532779520,
 'retweet_count': 4813,
 'favorite_count': 16101},
{'tweet_id': 759047813560868866,
 'retweet_count': 2302,
 'favorite_count': 7227},
{'tweet_id': 758854675097526272,
 'retweet_count': 1027,
 'favorite_count': 3904},
{'tweet_id': 758828659922702336,
 'retweet_count': 4376,
 'favorite_count': 12376},
{'tweet_id': 758740312047005698,
 'retweet_count': 1824,
 'favorite_count': 6339},
{'tweet_id': 758474966123810816,
 'retweet_count': 1132,
 'favorite_count': 4227},
{'tweet_id': 758467244762497024,
 'retweet_count': 2539,
 'favorite_count': 5316},
{'tweet_id': 758405701903519748,
```

  'retweet_count': 2204,
  'favorite_count': 5756},
{'tweet_id': 758355060040593408,
  'retweet_count': 1245,
  'favorite_count': 3797},
{'tweet_id': 758099635764359168,
  'retweet_count': 11550,
  'favorite_count': 21302},
{'tweet_id': 758041019896193024,
  'retweet_count': 433,
  'favorite_count': 3001},
{'tweet_id': 757741869644341248,
  'retweet_count': 3710,
  'favorite_count': 7613},
{'tweet_id': 757729163776290825, 'retweet_count': 9299, 'favorite_count': 0},
{'tweet_id': 757725642876129280,
  'retweet_count': 1391,
  'favorite_count': 5022},
{'tweet_id': 757611664640446465,
  'retweet_count': 1272,
  'favorite_count': 5026},
{'tweet_id': 757597904299253760, 'retweet_count': 336, 'favorite_count': 0},
{'tweet_id': 757596066325864448,
  'retweet_count': 1219,
  'favorite_count': 4808},
{'tweet_id': 757400162377592832,
  'retweet_count': 7759,
  'favorite_count': 16743},
{'tweet_id': 757393109802180609,
  'retweet_count': 2009,
  'favorite_count': 6462},
{'tweet_id': 757354760399941633,
  'retweet_count': 1637,
  'favorite_count': 4995},
{'tweet_id': 756998049151549440,
  'retweet_count': 2271,
  'favorite_count': 6923},
{'tweet_id': 756939218950160384,
  'retweet_count': 2295,
  'favorite_count': 7342},
{'tweet_id': 756651752796094464,
  'retweet_count': 1511,
  'favorite_count': 5612},
{'tweet_id': 756526248105566208,
  'retweet_count': 4169,
  'favorite_count': 11506},
{'tweet_id': 756303284449767430,

```
 'retweet_count': 1231,
 'favorite_count': 4359},
{'tweet_id': 756288534030475264,
 'retweet_count': 15071,
 'favorite_count': 28519},
{'tweet_id': 756275833623502848,
 'retweet_count': 1738,
 'favorite_count': 7114},
{'tweet_id': 755955933503782912,
 'retweet_count': 3285,
 'favorite_count': 8092},
{'tweet_id': 755206590534418437,
 'retweet_count': 6148,
 'favorite_count': 18212},
{'tweet_id': 755110668769038337,
 'retweet_count': 12621,
 'favorite_count': 23446},
{'tweet_id': 754874841593970688, 'retweet_count': 9193, 'favorite_count': 0},
{'tweet_id': 754856583969079297,
 'retweet_count': 2870,
 'favorite_count': 7616},
{'tweet_id': 754747087846248448,
 'retweet_count': 591,
 'favorite_count': 2854},
{'tweet_id': 754482103782404096,
 'retweet_count': 2220,
 'favorite_count': 5852},
{'tweet_id': 754449512966619136,
 'retweet_count': 846,
 'favorite_count': 4147},
{'tweet_id': 754120377874386944,
 'retweet_count': 2670,
 'favorite_count': 8655},
{'tweet_id': 754011816964026368,
 'retweet_count': 4079,
 'favorite_count': 9726},
{'tweet_id': 753655901052166144,
 'retweet_count': 2492,
 'favorite_count': 6458},
{'tweet_id': 753420520834629632,
 'retweet_count': 4071,
 'favorite_count': 8731},
{'tweet_id': 753398408988139520,
 'retweet_count': 2186,
 'favorite_count': 6384},
{'tweet_id': 753375668877008896,
 'retweet_count': 2655,
```

'favorite_count': 8411},
{'tweet_id': 753298634498793472, 'retweet_count': 6620, 'favorite_count': 0},
{'tweet_id': 753294487569522689,
 'retweet_count': 1191,
 'favorite_count': 3758},
{'tweet_id': 753039830821511168,
 'retweet_count': 24013,
 'favorite_count': 41080},
{'tweet_id': 753026973505581056,
 'retweet_count': 1109,
 'favorite_count': 4283},
{'tweet_id': 752932432744185856,
 'retweet_count': 7798,
 'favorite_count': 13970},
{'tweet_id': 752917284578922496,
 'retweet_count': 1790,
 'favorite_count': 7592},
{'tweet_id': 752701944171524096, 'retweet_count': 3291, 'favorite_count': 0},
{'tweet_id': 752682090207055872,
 'retweet_count': 1846,
 'favorite_count': 6642},
{'tweet_id': 752660715232722944,
 'retweet_count': 1794,
 'favorite_count': 4878},
{'tweet_id': 752568224206688256,
 'retweet_count': 2605,
 'favorite_count': 6140},
{'tweet_id': 752519690950500352,
 'retweet_count': 3917,
 'favorite_count': 8157},
{'tweet_id': 752334515931054080,
 'retweet_count': 1263,
 'favorite_count': 4238},
{'tweet_id': 752309394570878976, 'retweet_count': 18963, 'favorite_count': 0},
{'tweet_id': 752173152931807232,
 'retweet_count': 2106,
 'favorite_count': 6569},
{'tweet_id': 751950017322246144,
 'retweet_count': 1060,
 'favorite_count': 3415},
{'tweet_id': 751937170840121344,
 'retweet_count': 1530,
 'favorite_count': 5770},
{'tweet_id': 751830394383790080,
 'retweet_count': 2227,
 'favorite_count': 6428},
{'tweet_id': 751793661361422336,

```
 'retweet_count': 3320,
 'favorite_count': 6479},
{'tweet_id': 751598357617971201,
 'retweet_count': 3440,
 'favorite_count': 8699},
{'tweet_id': 751583847268179968,
 'retweet_count': 1265,
 'favorite_count': 4849},
{'tweet_id': 751538714308972544,
 'retweet_count': 1445,
 'favorite_count': 5547},
{'tweet_id': 751456908746354688,
 'retweet_count': 1127,
 'favorite_count': 3516},
{'tweet_id': 751251247299190784,
 'retweet_count': 6695,
 'favorite_count': 13791},
{'tweet_id': 751205363882532864,
 'retweet_count': 2086,
 'favorite_count': 6948},
{'tweet_id': 751132876104687617,
 'retweet_count': 1480,
 'favorite_count': 5610},
{'tweet_id': 750868782890057730,
 'retweet_count': 1595,
 'favorite_count': 5306},
{'tweet_id': 750719632563142656,
 'retweet_count': 5747,
 'favorite_count': 14621},
{'tweet_id': 750506206503038976,
 'retweet_count': 1316,
 'favorite_count': 4934},
{'tweet_id': 750429297815552001,
 'retweet_count': 4947,
 'favorite_count': 14569},
{'tweet_id': 750383411068534784,
 'retweet_count': 1309,
 'favorite_count': 5005},
{'tweet_id': 750381685133418496, 'retweet_count': 39, 'favorite_count': 758},
{'tweet_id': 750147208377409536,
 'retweet_count': 1095,
 'favorite_count': 3409},
{'tweet_id': 750132105863102464,
 'retweet_count': 1440,
 'favorite_count': 3990},
{'tweet_id': 750117059602808832,
 'retweet_count': 1466,
```

```
 'favorite_count': 4740},
{'tweet_id': 750101899009982464,
 'retweet_count': 959,
 'favorite_count': 3344},
{'tweet_id': 750086836815486976,
 'retweet_count': 613,
 'favorite_count': 2383},
{'tweet_id': 750071704093859840,
 'retweet_count': 3802,
 'favorite_count': 8653},
{'tweet_id': 750056684286914561,
 'retweet_count': 1011,
 'favorite_count': 3444},
{'tweet_id': 750041628174217216,
 'retweet_count': 703,
 'favorite_count': 3502},
{'tweet_id': 750026558547456000,
 'retweet_count': 888,
 'favorite_count': 2986},
{'tweet_id': 750011400160841729,
 'retweet_count': 1035,
 'favorite_count': 3568},
{'tweet_id': 749996283729883136,
 'retweet_count': 919,
 'favorite_count': 3331},
{'tweet_id': 749981277374128128,
 'retweet_count': 2772,
 'favorite_count': 5569},
{'tweet_id': 749774190421639168,
 'retweet_count': 1493,
 'favorite_count': 5114},
{'tweet_id': 749417653287129088,
 'retweet_count': 1904,
 'favorite_count': 6721},
{'tweet_id': 749403093750648834,
 'retweet_count': 622,
 'favorite_count': 2892},
{'tweet_id': 749395845976588288,
 'retweet_count': 3951,
 'favorite_count': 9488},
{'tweet_id': 749317047558017024,
 'retweet_count': 2509,
 'favorite_count': 6076},
{'tweet_id': 749075273010798592,
 'retweet_count': 2360,
 'favorite_count': 6353},
{'tweet_id': 749064354620928000,
```

'retweet_count': 1714,
 'favorite_count': 5277},
{'tweet_id': 749036806121881602,
 'retweet_count': 896,
 'favorite_count': 3425},
{'tweet_id': 748977405889503236,
 'retweet_count': 3759,
 'favorite_count': 11235},
{'tweet_id': 748932637671223296,
 'retweet_count': 2564,
 'favorite_count': 6461},
{'tweet_id': 748705597323898880,
 'retweet_count': 1067,
 'favorite_count': 3047},
{'tweet_id': 748699167502000129,
 'retweet_count': 1814,
 'favorite_count': 5213},
{'tweet_id': 748692773788876800,
 'retweet_count': 1504,
 'favorite_count': 4659},
{'tweet_id': 748575535303884801,
 'retweet_count': 2272,
 'favorite_count': 6696},
{'tweet_id': 748568946752774144,
 'retweet_count': 776,
 'favorite_count': 2411},
{'tweet_id': 748346686624440324,
 'retweet_count': 1413,
 'favorite_count': 5735},
{'tweet_id': 748337862848962560,
 'retweet_count': 4701,
 'favorite_count': 8462},
{'tweet_id': 748324050481647620,
 'retweet_count': 867,
 'favorite_count': 4078},
{'tweet_id': 748307329658011649,
 'retweet_count': 810,
 'favorite_count': 4027},
{'tweet_id': 748220828303695873,
 'retweet_count': 8825,
 'favorite_count': 15549},
{'tweet_id': 747963614829678593,
 'retweet_count': 2444,
 'favorite_count': 6397},
{'tweet_id': 747933425676525569,
 'retweet_count': 2894,
 'favorite_count': 7310},

```
        …]
```

`[878]:` `json_df.shape`

`[878]:` `(2354, 3)`

`[879]:` 
```python
# Save the dataFrame in file
json_df.to_csv('tweet_json.txt', encoding = 'utf-8', index=False)
```

`[880]:` 
```python
df2 = pd.read_csv('tweet_json.txt', encoding = 'utf-8')
df2
```

`[880]:`
```
              tweet_id  retweet_count  favorite_count
0      892420643555336193  8853            39467
1      892177421306343426  6514            33819
2      891815181378084864  4328            25461
3      891689557279858688  8964            42908
4      891327558926688256  9774            41048
…              …    …                 …
2349   666049248165822465  41              111
2350   666044226329800704  147             311
2351   666033412701032449  47              128
2352   666029285002620928  48              132
2353   666020888022790149  532             2535

[2354 rows x 3 columns]
```

## 1.2 Assessing Data

In this section, detect and document at least **eight (8) quality issues and two (2) tidiness issue**. You must use **both** visual assessment programmatic assessement to assess the data.

**Note:** pay attention to the following key points when you access the data.

- You only want original ratings (no retweets) that have images. Though there are 5000+ tweets in the dataset, not all are dog ratings and some are retweets.
- Assessing and cleaning the entire dataset completely would require a lot of time, and is not necessary to practice and demonstrate your skills in data wrangling. Therefore, the requirements of this project are only to assess and clean at least 8 quality issues and at least 2 tidiness issues in this dataset.
- The fact that the rating numerators are greater than the denominators does not need to be cleaned. This unique rating system is a big part of the popularity of WeRateDogs.
- You do not need to gather the tweets beyond August 1st, 2017. You can, but note that you won't be able to gather the image predictions for these tweets since you don't have access to the algorithm used.

`[881]:` `df.sample(5)`

```
[881]:                  tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
      799   772826264096874500 NaN                    NaN
      945   752660715232722944 NaN                    NaN
      1989  672828477930868736 NaN                    NaN
      1872  675146535592706048 NaN                    NaN
      356   830583320585068544 NaN                    NaN


                           timestamp  \
      799    2016-09-05 15:58:34 +0000
      945    2016-07-12 00:27:52 +0000
      1989   2015-12-04 17:23:04 +0000
      1872   2015-12-11 02:54:12 +0000
      356    2017-02-12 01:04:29 +0000


        source  \
      799    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
      iPhone</a>
      945    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
      iPhone</a>
      1989  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
      iPhone</a>
      1872  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
      iPhone</a>
      356   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
      iPhone</a>


                                                                          text  \
      799    Meet Roosevelt. He's preparing for takeoff. Make sure tray tables are in
      their full pupright &amp; licked position\n11/10 https://t.co/7CQkn3gHOQ
      945    Hooman used Pokeball\n*wiggle*\n*wiggle*\nDoggo broke free \n10/10
      https://t.co/bWSgqnwSHr
      1989  This is Jerry. He's a Timbuk Slytherin. Eats his pizza from the side
      first. Crushed that cup with his bare paws 9/10 https://t.co/fvxHL6cRRs
      1872  This is Coops. He's yelling at the carpet. Not very productive Coops. 7/10
      https://t.co/Uz52oYnHzF
      356    This is Lilly. She just parallel barked. Kindly requests a reward now.
      13/10 would pet so well https://t.co/SATN4If5H5


            retweeted_status_id  retweeted_status_user_id  \
      799  NaN                   NaN
      945  NaN                   NaN
      1989 NaN                   NaN
      1872 NaN                   NaN
      356  NaN                   NaN


            retweeted_status_timestamp  \
      799    NaN
```
63

```
945    NaN
1989   NaN
1872   NaN
356    NaN


                                     expanded_urls  \
799    https://twitter.com/dog_rates/status/772826264096874500/photo/1
945    https://twitter.com/dog_rates/status/752660715232722944/photo/1,https://tw
itter.com/dog_rates/status/752660715232722944/photo/1
1989   https://twitter.com/dog_rates/status/672828477930868736/photo/1
1872   https://twitter.com/dog_rates/status/675146535592706048/photo/1
356    https://twitter.com/dog_rates/status/830583320585068544/photo/1,https://tw
itter.com/dog_rates/status/830583320585068544/photo/1


       rating_numerator  rating_denominator       name  doggo floofer pupper  \
799    11                10                   Roosevelt None   None    None
945    10                10                   None      doggo  None    None
1989   9                 10                   Jerry     None   None    None
1872   7                 10                   Coops     None   None    None
356    13                10                   Lilly     None   None    None


       puppo
799    None
945    None
1989   None
1872   None
356    None
```

[882]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   tweet_id                   2356 non-null   int64
 1   in_reply_to_status_id      78 non-null     float64
 2   in_reply_to_user_id        78 non-null     float64
 3   timestamp                  2356 non-null   object
 4   source                     2356 non-null   object
 5   text                       2356 non-null   object
 6   retweeted_status_id        181 non-null    float64
 7   retweeted_status_user_id   181 non-null    float64
 8   retweeted_status_timestamp 181 non-null    object
 9   expanded_urls              2297 non-null   object
 10  rating_numerator           2356 non-null   int64
 11  rating_denominator         2356 non-null   int64
 12  name                       2356 non-null   object
```

```
13   doggo                         2356 non-null    object
14   floofer                       2356 non-null    object
15   pupper                        2356 non-null    object
16   puppo                         2356 non-null    object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

[883]: `df.describe()`

[883]:
```
              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
count    2.356000e+03           7.800000e+01         7.800000e+01
mean     7.427716e+17           7.455079e+17         2.014171e+16
std      6.856705e+16           7.582492e+16         1.252797e+17
min      6.660209e+17           6.658147e+17         1.185634e+07
25%      6.783989e+17           6.757419e+17         3.086374e+08
50%      7.196279e+17           7.038708e+17         4.196984e+09
75%      7.993373e+17           8.257804e+17         4.196984e+09
max      8.924206e+17           8.862664e+17         8.405479e+17


         retweeted_status_id  retweeted_status_user_id  rating_numerator  \
count           1.810000e+02              1.810000e+02       2356.000000
mean            7.720400e+17              1.241698e+16         13.126486
std             6.236928e+16              9.599254e+16         45.876648
min             6.661041e+17              7.832140e+05          0.000000
25%             7.186315e+17              4.196984e+09         10.000000
50%             7.804657e+17              4.196984e+09         11.000000
75%             8.203146e+17              4.196984e+09         12.000000
max             8.874740e+17              7.874618e+17       1776.000000


         rating_denominator
count           2356.000000
mean              10.455433
std                6.745237
min                0.000000
25%               10.000000
50%               10.000000
75%               10.000000
max              170.000000
```

[884]: `df[df.duplicated()]`

[884]: Empty DataFrame
Columns: [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]
Index: []

```
[885]: df[df.tweet_id.duplicated()]
```

```
[885]: Empty DataFrame
       Columns: [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]
       Index: []
```

```
[886]: df[df.isnull()]
```

```
[886]:        tweet_id  in_reply_to_status_id  in_reply_to_user_id timestamp source  \
       0     NaN        NaN                    NaN                  NaN       NaN
       1     NaN        NaN                    NaN                  NaN       NaN
       2     NaN        NaN                    NaN                  NaN       NaN
       3     NaN        NaN                    NaN                  NaN       NaN
       4     NaN        NaN                    NaN                  NaN       NaN
       …     ..         ..                     ..                   …         …
       2351  NaN        NaN                    NaN                  NaN       NaN
       2352  NaN        NaN                    NaN                  NaN       NaN
       2353  NaN        NaN                    NaN                  NaN       NaN
       2354  NaN        NaN                    NaN                  NaN       NaN
       2355  NaN        NaN                    NaN                  NaN       NaN

             text  retweeted_status_id  retweeted_status_user_id  \
       0     NaN NaN                    NaN
       1     NaN NaN                    NaN
       2     NaN NaN                    NaN
       3     NaN NaN                    NaN
       4     NaN NaN                    NaN
       …     …   ..                     ..
       2351  NaN NaN                    NaN
       2352  NaN NaN                    NaN
       2353  NaN NaN                    NaN
       2354  NaN NaN                    NaN
       2355  NaN NaN                    NaN

             retweeted_status_timestamp expanded_urls  rating_numerator  \
       0     NaN                        NaN            NaN
       1     NaN                        NaN            NaN
       2     NaN                        NaN            NaN
       3     NaN                        NaN            NaN
       4     NaN                        NaN            NaN
       …     …                          …              ..
       2351  NaN                        NaN            NaN
       2352  NaN                        NaN            NaN
       2353  NaN                        NaN            NaN
```

```
2354  NaN                              NaN            NaN
2355  NaN                              NaN            NaN


      rating_denominator name doggo floofer pupper puppo
0     NaN                 NaN  NaN   NaN     NaN     NaN
1     NaN                 NaN  NaN   NaN     NaN     NaN
2     NaN                 NaN  NaN   NaN     NaN     NaN
3     NaN                 NaN  NaN   NaN     NaN     NaN
4     NaN                 NaN  NaN   NaN     NaN     NaN
...   ..                  ...  ...   ...     ...     ...
2351  NaN                 NaN  NaN   NaN     NaN     NaN
2352  NaN                 NaN  NaN   NaN     NaN     NaN
2353  NaN                 NaN  NaN   NaN     NaN     NaN
2354  NaN                 NaN  NaN   NaN     NaN     NaN
2355  NaN                 NaN  NaN   NaN     NaN     NaN


[2356 rows x 17 columns]
```

[887]: `df.shape`

[887]: (2356, 17)

[888]: `df1.sample(10)`

[888]:
```
              tweet_id                                       jpg_url  \
1314  754856583969079297  https://pbs.twimg.com/media/CnnKCKNWgAAcOB8.jpg
745   687494652870668288  https://pbs.twimg.com/media/CYp4vFrVAAEs9AX.jpg
1270  749981277374128128  https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg
1299  752519690950500352  https://pbs.twimg.com/media/CnF8qVDWYAAhOg1.jpg
542   677228873407442944  https://pbs.twimg.com/media/CWYAEINW4AIuw8P.jpg
2007  877736472329191424  https://pbs.twimg.com/media/DC5YqoQWOAArOLH.jpg
916   701545186879471618  https://pbs.twimg.com/media/CbxjnyOWAAAWLUH.jpg
1799  831670449226514432  https://pbs.twimg.com/media/C4qv3JUW8AADirb.jpg
669   683078886620553216  https://pbs.twimg.com/media/CXrIntsUsAEkvOd.jpg
1391  766793450729734144  https://pbs.twimg.com/media/CqQykxrWYAAlD8g.jpg


      img_num                       p1   p1_conf  p1_dog                 p2  \
1314  2       golden_retriever         0.872385  True    Labrador_retriever
745   1       Rottweiler               0.391471  True    miniature_pinscher
1270  1       bow_tie                  0.533941  False   sunglasses
1299  3       swing                    0.999984  False   Labrador_retriever
542   1       common_iguana            0.566338  False   tennis_ball
2007  2       Chesapeake_Bay_retriever 0.837956  True    Labrador_retriever
916   1       Border_collie            0.280893  True    Cardigan
1799  1       Pembroke                 0.624802  True    Cardigan
669   1       koala                    0.141432  False   Eskimo_dog
1391  1       beagle                   0.451697  True    basset
```

|      | p2_conf  | p2_dog | p3             | p3_conf  | p3_dog |
|------|----------|--------|----------------|----------|--------|
| 1314 | 0.099963 | True   | cocker_spaniel | 0.006051 | True   |
| 745  | 0.273595 | True   | Tibetan_mastiff| 0.041692 | True   |
| 1270 | 0.080822 | False  | sunglass       | 0.050776 | False  |
| 1299 | 0.000010 | True   | Eskimo_dog     | 0.000001 | True   |
| 542  | 0.154646 | False  | green_lizard   | 0.044976 | False  |
| 2007 | 0.062034 | True   | Weimaraner     | 0.040599 | True   |
| 916  | 0.112550 | True   | toy_terrier    | 0.053317 | True   |
| 1799 | 0.362861 | True   | Appenzeller    | 0.003926 | True   |
| 669  | 0.094044 | True   | wallaby        | 0.085237 | False  |
| 1391 | 0.197513 | True   | bloodhound     | 0.072699 | True   |

[889]: `df1.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   tweet_id  2075 non-null   int64
 1   jpg_url   2075 non-null   object
 2   img_num   2075 non-null   int64
 3   p1        2075 non-null   object
 4   p1_conf   2075 non-null   float64
 5   p1_dog    2075 non-null   bool
 6   p2        2075 non-null   object
 7   p2_conf   2075 non-null   float64
 8   p2_dog    2075 non-null   bool
 9   p3        2075 non-null   object
 10  p3_conf   2075 non-null   float64
 11  p3_dog    2075 non-null   bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

[890]: `df1.describe()`

[890]:

|       | tweet_id     | img_num     | p1_conf     | p2_conf      | p3_conf      |
|-------|--------------|-------------|-------------|--------------|--------------|
| count | 2.075000e+03 | 2075.000000 | 2075.000000 | 2.075000e+03 | 2.075000e+03 |
| mean  | 7.384514e+17 | 1.203855    | 0.594548    | 1.345886e-01 | 6.032417e-02 |
| std   | 6.785203e+16 | 0.561875    | 0.271174    | 1.006657e-01 | 5.090593e-02 |
| min   | 6.660209e+17 | 1.000000    | 0.044333    | 1.011300e-08 | 1.740170e-10 |
| 25%   | 6.764835e+17 | 1.000000    | 0.364412    | 5.388625e-02 | 1.622240e-02 |
| 50%   | 7.119988e+17 | 1.000000    | 0.588230    | 1.181810e-01 | 4.944380e-02 |
| 75%   | 7.932034e+17 | 1.000000    | 0.843855    | 1.955655e-01 | 9.180755e-02 |
| max   | 8.924206e+17 | 4.000000    | 1.000000    | 4.880140e-01 | 2.734190e-01 |

[891]: `df1[df1.duplicated()]`

```
[891]: Empty DataFrame
      Columns: [tweet_id, jpg_url, img_num, p1, p1_conf, p1_dog, p2, p2_conf, p2_dog,
      p3, p3_conf, p3_dog]
      Index: []
```

```
[892]: df1[df1.jpg_url.duplicated()]
```

```
[892]:                tweet_id  \
      1297   752309394570878976
      1315   754874841593970688
      1333   757729163776290825
      1345   759159934323924993
      1349   759566828574212096
      ...                   ...
      1903   851953902622658560
      1944   861769973181624320
      1992   873697596434513921
      2041   885311592912609280
      2055   888202515573088257


             jpg_url  \
      1297   https://pbs.twimg.com/ext_tw_video_thumb/675354114423808004/pu/img/qL1R_nG
      Lqa6lmkOx.jpg
      1315   https://pbs.twimg.com/media/CWza7kpWcAAdYLc.jpg
      1333   https://pbs.twimg.com/media/CWyD2HGUYAQ1Xa7.jpg
      1345   https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg
      1349   https://pbs.twimg.com/media/CkNjahBXAAQ2kWo.jpg
      ...                                                 ...
      1903   https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg
      1944   https://pbs.twimg.com/media/CzG425nWgAAnP7P.jpg
      1992   https://pbs.twimg.com/media/DA7iHL5U0AA1OQo.jpg
      2041   https://pbs.twimg.com/media/C4bTH6nWMAAX_bJ.jpg
      2055   https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg


             img_num                          p1   p1_conf  p1_dog  \
      1297   1        upright                     0.303415  False
      1315   1        pug                         0.272205  True
      1333   2        cash_machine                0.802333  False
      1345   1        Irish_terrier               0.254856  True
      1349   1        Labrador_retriever          0.967397  True
      ...   ..                        ...             ...   ...
      1903   1        Staffordshire_bullterrier   0.757547  True
      1944   2        Arabian_camel               0.366248  False
      1992   1        laptop                      0.153718  False
      2041   1        Labrador_retriever          0.908703  True
      2055   2        Pembroke                    0.809197  True
```

69

```
                           p2     p2_conf  p2_dog  \
1297  golden_retriever              0.181351  True
1315  bull_mastiff                  0.251530  True
1333  schipperke                    0.045519  True
1345  briard                        0.227716  True
1349  golden_retriever              0.016641  True
...             ...                     ...   ...
1903  American_Staffordshire_terrier  0.149950  True
1944  house_finch                   0.209852  False
1992  French_bulldog                0.099984  True
2041  seat_belt                     0.057091  False
2055  Rhodesian_ridgeback           0.054950  True


                           p3     p3_conf  p3_dog
1297  Brittany_spaniel              0.162084  True
1315  bath_towel                    0.116806  False
1333  German_shepherd               0.023353  True
1345  soft-coated_wheaten_terrier   0.223263  True
1349  ice_bear                      0.014858  False
...          ...                       ...   ...
1903  Chesapeake_Bay_retriever      0.047523  True
1944  cocker_spaniel                0.046403  True
1992  printer                       0.077130  False
2041  pug                           0.011933  True
2055  beagle                        0.038915  True

[66 rows x 12 columns]
```

[893]: `df1[df1.isnull()]`

```
[893]:        tweet_id jpg_url  img_num   p1  p1_conf p1_dog   p2  p2_conf p2_dog  \
0        NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
1        NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
2        NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
3        NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
4        NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
...      ..       ...     ..      ...  ..      ...   ...  ..      ...
2070 NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
2071 NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
2072 NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
2073 NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN
2074 NaN      NaN     NaN     NaN   NaN     NaN   NaN   NaN     NaN

         p3  p3_conf p3_dog
0       NaN NaN      NaN
1       NaN NaN      NaN
2       NaN NaN      NaN
```

```
3       NaN NaN         NaN
4       NaN NaN         NaN

…       …   ..          …
2070    NaN NaN         NaN
2071    NaN NaN         NaN
2072    NaN NaN         NaN
2073    NaN NaN         NaN
2074    NaN NaN         NaN

[2075 rows x 12 columns]
```

[894]: `df1.shape`

[894]: (2075, 12)

[895]: `df2.sample(10)`

[895]:
```
            tweet_id  retweet_count  favorite_count
1741  679405845277462528  1364           2594
765   777684233540206592  3403           12518
176   857393404942143489  1785           6236
175   857746408056729600  11524          36021
1207  715696743237730304  1454           4257
8     890609185150312448  4429           28226
2091  670780561024270336  317            831
1159  723912936180330496  1374           4236
1750  678969228704284672  528            1796
2045  671518598289059840  319            1010
```

[896]: `df2.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        2354 non-null   int64
 1   retweet_count   2354 non-null   int64
 2   favorite_count  2354 non-null   int64
dtypes: int64(3)
memory usage: 55.3 KB
```

[897]: `df2.describe()`

[897]:
```
            tweet_id  retweet_count  favorite_count
count  2.354000e+03  2354.000000    2354.000000
mean   7.426978e+17  3164.797366    8080.968564
std    6.852812e+16  5284.770364    11814.771334
```

```
min     6.660209e+17  0.000000      0.000000
25%     6.783975e+17  624.500000    1415.000000
50%     7.194596e+17  1473.500000   3603.500000
75%     7.993058e+17  3652.000000   10122.250000
max     8.924206e+17  79515.000000  132810.000000
```

[898]: `df2[df2.duplicated()]`

[898]: Empty DataFrame
       Columns: [tweet_id, retweet_count, favorite_count]
       Index: []

[899]: `df2[df2.tweet_id.duplicated()]`

[899]: Empty DataFrame
       Columns: [tweet_id, retweet_count, favorite_count]
       Index: []

[900]: `df2[df2.isnull()]`

[900]:       tweet_id  retweet_count  favorite_count
       0     NaN       NaN            NaN
       1     NaN       NaN            NaN
       2     NaN       NaN            NaN
       3     NaN       NaN            NaN
       4     NaN       NaN            NaN
       …     ..        ..             ..
       2349  NaN       NaN            NaN
       2350  NaN       NaN            NaN
       2351  NaN       NaN            NaN
       2352  NaN       NaN            NaN
       2353  NaN       NaN            NaN

       [2354 rows x 3 columns]

[901]: `df2.shape`

[901]: (2354, 3)

### 1.2.1 Quality issues

1. The following columns have int64 datatypes instead of string datatype:

- in_reply_to_status_id
- in-reply_to_user_id
- retweeted_status_id
- retweeted_status_user_id
- tweet_id

2. The following columns are of object datatypes instead of datetime datatypes

- timestamp

3. The name,doggo, floofer, pupper and puppo columns have none null values ie. The none should be NAN instead.

4. The name column hava invalid names such as none, a, an, such, the, just etc. These invalid names are in lower case.

5. We only need original ratings with images and not all ratings have images.

6. We have missing values in the df1 dataset.There should be 2356 rows but we got 2075 rows.

7. In df1, some ratings are wrong.

8. Drop the five columns since they will not be used for analysis(in_reply_to_status_id, in-reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp)

9. Remove all rows in the `text` column that contains retweeted data.

10. There are rows with mulitiple stages for a tweet , so we are going to split them eg, the `doggopuppo` should be `doggo`, `puppo`

### 1.2.2 Tidiness issues

1. df2 should be part of df

2. Doggo, floofer, pupper, puppo columns show one variable.

3. Rating_numerator and denominator should be one variable rating.

4. The df2 dataset should be part of df

## 1.3 Cleaning Data

In this section, clean **all** of the issues you documented while assessing.

**Note:** Make a copy of the original data before cleaning. Cleaning includes merging individual pieces of data according to the rules of tidy data. The result should be a high-quality and tidy master pandas DataFrame (or DataFrames, if appropriate).

```
[902]:  # Make copies of original pieces of data
        df_clean = df.copy()
        df1_clean = df1.copy()
        df2_clean = df2.copy()
```

### 1.3.1 Issue #1:

## 1.4 1. The following columns have int datatypes instead of string datatype:

- in_reply_to_status_id
- in-reply_to_user_id
- retweeted_status_id
- retweeted_status_user_id

- tweet_id

**Define:** Covert datatypes in in_reply_to_status_id, in-reply_to_user_id, retweeted_status_id, retweeted_status_user_id columns to int

**Code**

```
[903]: df_clean.in_reply_to_status_id = df_clean.in_reply_to_status_id.astype('str')
       df_clean.in_reply_to_user_id = df_clean.in_reply_to_user_id.astype('str')
       df_clean.retweeted_status_id = df_clean.retweeted_status_id.astype('str')
       df_clean.retweeted_status_user_id = df_clean.retweeted_status_user_id.
         ↪astype('str')
       df_clean.tweet_id = df_clean.tweet_id.astype('str')
```

```
[904]: df2_clean.tweet_id = df2_clean.tweet_id.astype('str')
```

```
[905]: df1_clean.tweet_id = df1_clean.tweet_id.astype('str')
```

**Test**

```
[906]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   tweet_id                    2356 non-null   object
 1   in_reply_to_status_id       2356 non-null   object
 2   in_reply_to_user_id         2356 non-null   object
 3   timestamp                   2356 non-null   object
 4   source                      2356 non-null   object
 5   text                        2356 non-null   object
 6   retweeted_status_id         2356 non-null   object
 7   retweeted_status_user_id    2356 non-null   object
 8   retweeted_status_timestamp  181 non-null    object
 9   expanded_urls               2297 non-null   object
 10  rating_numerator            2356 non-null   int64
 11  rating_denominator          2356 non-null   int64
 12  name                        2356 non-null   object
 13  doggo                       2356 non-null   object
 14  floofer                     2356 non-null   object
 15  pupper                      2356 non-null   object
 16  puppo                       2356 non-null   object
dtypes: int64(2), object(15)
memory usage: 313.0+ KB
```

### 1.4.1 Issue #2:

- The name, doggo, floofer, pupper and puppo columns have none null values ie. The none should be NAN instead.

**Define**

- replace all none values with Null or empty

**Code**

```
[907]: df_clean.doggo=df_clean.doggo.replace('None','')
       df_clean.floofer=df_clean.floofer.replace('None','')
       df_clean.pupper=df_clean.pupper.replace('None','')
       df_clean.puppo=df_clean.puppo.replace('None','')
```

**Test**

```
[908]: df_clean.sample(10)
```

```
[908]:                tweet_id  in_reply_to_status_id in_reply_to_user_id  \
       1087   738156290900254721  nan                    nan
       1913   674372068062928900  nan                    nan
       2082   670815497391357952  nan                    nan
       188    855862651834028034  8.558615844633518e+17  194351775.0
       10     890006608113172480  nan                    nan
       1734   679736210798047232  nan                    nan
       329    833479644947025920  nan                    nan
       429    821107785811234820  nan                    nan
       837    767754930266464257  nan                    nan
       79     877556246731214848  nan                    nan


                           timestamp  \
       1087   2016-06-01 23:52:28 +0000
       1913   2015-12-08 23:36:44 +0000
       2082   2015-11-29 04:04:12 +0000
       188    2017-04-22 19:15:32 +0000
       10     2017-07-26 00:31:25 +0000
       1734   2015-12-23 18:51:56 +0000
       329    2017-02-20 00:53:27 +0000
       429    2017-01-16 21:32:06 +0000
       837    2016-08-22 16:06:54 +0000
       79     2017-06-21 15:58:08 +0000


         source  \
       1087  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       1913  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       2082  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
```

```
iPhone</a>
188    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
10     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1734   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
329    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
429    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
837    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
79     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>

      text  \
1087  This is Storkson. He's wet and sad. 10/10 cheer up pup
https://t.co/nrzvzPuTvC
1913  Meet Chesney. On the outside he stays calm &amp; collected. On the inside
he's having a complete mental breakdown. 10/10 https://t.co/G4mOTFY9uc
2082  This is Sage. He likes to burn shit. 10/10 https://t.co/nLYruSMRe6
188   @dhmontgomery We also gave snoop dogg a 420/10 but I think that predated
your research
10    This is Koda. He is a South Australian deckshark. Deceptively deadly.
Frighteningly majestic. 13/10 would risk a petting #BarkWeek
https://t.co/dVPWOBOMme
1734  This pup's name is Sabertooth (parents must be cool). Ears for days. Jumps
unannounced. 9/10 would pet diligently https://t.co/iazoiNUviP
329   This is Poppy. She just arrived. 13/10 would snug passionately
https://t.co/YGeSpyN8Gu
429   Here's a doggo who looks like he's about to give you a list of mythical
ingredients to go collect for his potion. 11/10 would obey
https://t.co/8SiwKDlRcl
837   This is Philbert. His toilet broke and he doesn't know what to do. Trying
not to panic. 11/10 furustrated af https://t.co/Nb68IsVb9O
79    This is Penny. She's both pupset and fired pup. Not pleased w your
barbaric attempts at cleanliness. 12/10 would enjoy more shampoo options
https://t.co/OYdDlfOGXP

      retweeted_status_id retweeted_status_user_id retweeted_status_timestamp  \
1087  nan                 nan                       NaN
1913  nan                 nan                       NaN
2082  nan                 nan                       NaN
188   nan                 nan                       NaN
10    nan                 nan                       NaN
1734  nan                 nan                       NaN
```

```
329  nan                   nan                   NaN
429  nan                   nan                   NaN
837  nan                   nan                   NaN
79   nan                   nan                   NaN


                    expanded_urls  \
1087  https://twitter.com/dog_rates/status/738156290900254721/photo/1
1913  https://twitter.com/dog_rates/status/674372068062928900/photo/1
2082  https://twitter.com/dog_rates/status/670815497391357952/photo/1
188   NaN
10    https://twitter.com/dog_rates/status/890006608113172480/photo/1,https://tw
itter.com/dog_rates/status/890006608113172480/photo/1
1734  https://twitter.com/dog_rates/status/679736210798047232/photo/1
329   https://twitter.com/dog_rates/status/833479644947025920/photo/1,https://tw
itter.com/dog_rates/status/833479644947025920/photo/1,https://twitter.com/dog_ra
tes/status/833479644947025920/photo/1
429   https://twitter.com/dog_rates/status/821107785811234820/photo/1
837   https://twitter.com/dog_rates/status/767754930266464257/photo/1
79    https://twitter.com/dog_rates/status/877556246731214848/photo/1


      rating_numerator  rating_denominator       name  doggo floofer pupper  \
1087  10                10                   Storkson
1913  10                10                   Chesney
2082  10                10                   Sage
188   420               10                   None
10    13                10                   Koda
1734  9                 10                   None
329   13                10                   Poppy
429   11                10                   None      doggo
837   11                10                   Philbert
79    12                10                   Penny


      puppo
1087
1913
2082
188
10
1734
329
429
837
79
```

### 1.4.2  Issue #3

- Doggo, floofer, pupper, puppo columns show one variable.

**Define**

- Create one column named dog_stage and drop columns doggo, floofer, pupper, puppo to show dog's stage.

**Code**

```
[909]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   tweet_id                    2356 non-null   object
 1   in_reply_to_status_id       2356 non-null   object
 2   in_reply_to_user_id         2356 non-null   object
 3   timestamp                   2356 non-null   object
 4   source                      2356 non-null   object
 5   text                        2356 non-null   object
 6   retweeted_status_id         2356 non-null   object
 7   retweeted_status_user_id    2356 non-null   object
 8   retweeted_status_timestamp  181 non-null    object
 9   expanded_urls               2297 non-null   object
 10  rating_numerator            2356 non-null   int64
 11  rating_denominator          2356 non-null   int64
 12  name                        2356 non-null   object
 13  doggo                       2356 non-null   object
 14  floofer                     2356 non-null   object
 15  pupper                      2356 non-null   object
 16  puppo                       2356 non-null   object
dtypes: int64(2), object(15)
memory usage: 313.0+ KB
```

```
[910]: df_clean['dog_stage']=df_clean.doggo.str.cat(df_clean.floofer).str.cat(df_clean.
       ↪pupper).str.cat(df_clean.puppo)
       df_clean=df_clean.drop(['doggo','floofer','pupper','puppo'],axis=1)
       df_clean.dog_stage=df_clean.dog_stage.replace('',np.nan)
```

**Test**

```
[911]: df_clean.sample(5)
```

```
[911]:                  tweet_id  in_reply_to_status_id in_reply_to_user_id  \
       409    823333489516937216  8.233263893362442e+17   1582853809.0
       789    773985732834758656  nan                     nan
       2108   670452855871037440  nan                     nan
       1039   744234799360020481  nan                     nan
       1569   687807801670897665  nan                     nan
```

78

```
                         timestamp  \
409    2017-01-23 00:56:15 +0000
789    2016-09-08 20:45:53 +0000
2108   2015-11-28 04:03:11 +0000
1039   2016-06-18 18:26:18 +0000
1569   2016-01-15 01:25:33 +0000


   source  \
409    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
789    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2108   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1039   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1569   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>


                                                                     text  \
409    @HistoryInPics 13/10
789    Meet Winnie. She just made awkward eye contact with the driver beside her.
Poor pupper panicked. 11/10 would comfort https://t.co/RFWtDqTnAz
2108   This dog can't see its haters. 11/10 https://t.co/35BcGFdEAK
1039   Here's a doggo realizing you can stand in a pool. 13/10 enlightened af
(vid by Tina Conrad) https://t.co/7wE9LTEXC4
1569   Meet Trooper &amp; Maya. Trooper protects Maya from bad things like
dognappers and Comcast. So touching. 11/10 for both https://t.co/c98k1IoZKy


     retweeted_status_id retweeted_status_user_id retweeted_status_timestamp  \
409    nan                    nan                    NaN
789    nan                    nan                    NaN
2108   nan                    nan                    NaN
1039   nan                    nan                    NaN
1569   nan                    nan                    NaN


        expanded_urls  \
409    NaN
789    https://twitter.com/dog_rates/status/773985732834758656/photo/1,https://tw
itter.com/dog_rates/status/773985732834758656/photo/1,https://twitter.com/dog_ra
tes/status/773985732834758656/photo/1,https://twitter.com/dog_rates/status/77398
5732834758656/photo/1
2108   https://twitter.com/dog_rates/status/670452855871037440/photo/1
1039   https://twitter.com/dog_rates/status/744234799360020481/video/1
1569   https://twitter.com/dog_rates/status/687807801670897665/photo/1


     rating_numerator  rating_denominator      name dog_stage
```

| 409 | 13 | 10 | None | NaN |
| 789 | 11 | 10 | Winnie | pupper |
| 2108 | 11 | 10 | None | NaN |
| 1039 | 13 | 10 | None | doggo |
| 1569 | 11 | 10 | Trooper | NaN |

### 1.4.3 Issue #4

Remove columns in_reply_to_status_id, in-reply_to_user_id, retweeted_status_id, retweeted_status_user_id,retweeted_status_timestamp.

**Define**  Use Drop to remove the five columns

**Code**

```
[912]: df_clean = df_clean.drop(columns=["in_reply_to_status_id",␣
       ↪"in_reply_to_user_id", "retweeted_status_id",␣
       ↪"retweeted_status_user_id","retweeted_status_timestamp"])
```

**Test**

```
[913]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 9 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   tweet_id           2356 non-null   object
 1   timestamp          2356 non-null   object
 2   source             2356 non-null   object
 3   text               2356 non-null   object
 4   expanded_urls      2297 non-null   object
 5   rating_numerator   2356 non-null   int64
 6   rating_denominator 2356 non-null   int64
 7   name               2356 non-null   object
 8   dog_stage          380 non-null    object
dtypes: int64(2), object(7)
memory usage: 165.8+ KB
```

```
[914]: df_clean.dog_stage.value_counts()
```

```
[914]: pupper         245
       doggo          83
       puppo          29
       doggopupper    12
       floofer        9
       doggopuppo     1
       doggofloofer   1
       Name: dog_stage, dtype: int64
```

```
[915]: df_clean.dog_stage.isnull().value_counts()
```

```
[915]: True     1976
       False    380
       Name: dog_stage, dtype: int64
```

```
[916]: df_clean.sample(5)
```

```
[916]:                tweet_id                 timestamp  \
       1487  693095443459342336  2016-01-29 15:36:45 +0000
       2336  666273097616637952  2015-11-16 15:14:19 +0000
       370   828409743546925057  2017-02-06 01:07:28 +0000
       1331  705591895322394625  2016-03-04 03:13:11 +0000
       2309  666804364988780544  2015-11-18 02:25:23 +0000


         source  \
       1487  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       2336  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       370   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       1331  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       2309  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>


                                                                        text  \
       1487  This is Milo. He doesn't understand your fancy human gestures. Will lick
       instead. 10/10 can't faze this pupper https://t.co/OhodPIDOpW
       2336  Can take selfies 11/10 https://t.co/ws2AMaNwPW
       370   This is Mutt Ryan. He's quite confident at the moment. 12/10 rise pup!
       https://t.co/ZH5CvRlCxt
       1331  "Ma'am, for the last time, I'm not authorized to make that type of
       transaction" 11/10 https://t.co/nPTBsdm3BF
       2309  This is Jockson. He is a Pinnacle Sagittarius. Fancy bandana. Enjoys
       lightly sucking on hot dog in nature. 8/10 https://t.co/RdKbAOEpDK


                                              expanded_urls  \
       1487  https://twitter.com/dog_rates/status/693095443459342336/photo/1
       2336  https://twitter.com/dog_rates/status/666273097616637952/photo/1
       370   https://twitter.com/dog_rates/status/828409743546925057/photo/1
       1331  https://twitter.com/dog_rates/status/705591895322394625/photo/1
       2309  https://twitter.com/dog_rates/status/666804364988780544/photo/1


             rating_numerator  rating_denominator      name dog_stage
       1487  10                10                       Milo    pupper
```

```
2336  11            10              None    NaN
370   12            10              Mutt    NaN
1331  11            10              None    NaN
2309  8             10              Jockson NaN
```

### 1.4.4 Issue #5:

- The df2 dataset should be part of df

**Define**

- Join df2 to df1 based on the tweet_id

**Code**

```
[917]: df2_clean = df2_clean[~(df2_clean.tweet_id.duplicated())]
       df_clean = pd.merge(df_clean, df2_clean, on='tweet_id', how='left')
```

**Test**

```
[918]: df_clean.sample(5)
```

```
[918]:            tweet_id              timestamp  \
       2234  668190681446379520  2015-11-21 22:14:07 +0000
       197   854482394044301312  2017-04-18 23:50:52 +0000
       2130  670093938074779648  2015-11-27 04:16:59 +0000
       776   776113305656188928  2016-09-14 17:40:06 +0000
       744   780192070812196864  2016-09-25 23:47:39 +0000


         source  \
       2234  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       197   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       2130  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       776   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>
       744   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>


                                                          text  \
       2234  This is Skittles. I would kidnap Skittles. Pink dog in back hasn't moved
       in days. 12/10 https://t.co/2wm0POA9N2
       197    This is Arya. She can barely contain her excitement for more peanut
       butter. Also patriotic af. 13/10 https://t.co/AL4Ahm1Rm5
       2130  This is Wally. He's a Flaccid Mitochondria. Going on vacation. Bag
       definitely full of treats. Great hat. 9/10 https://t.co/vYs9IVzHY9
       776    This is Sampson. He's about to get hit with a vicious draw 2. Has no idea.
       11/10 poor pupper https://t.co/FYT9QBEnKG
```

```
744    We only rate dogs. Pls stop sending in non-canines like this Urban Floof
       Giraffe. I can't handle this. 11/10 https://t.co/zHIqpM5Gni

                                                        expanded_urls  \
2234   https://twitter.com/dog_rates/status/668190681446379520/photo/1
197    https://twitter.com/dog_rates/status/854482394044301312/photo/1
2130   https://twitter.com/dog_rates/status/670093938074779648/photo/1
776    https://twitter.com/dog_rates/status/776113305656188928/photo/1
744    https://twitter.com/dog_rates/status/780192070812196864/photo/1

       rating_numerator  rating_denominator       name  dog_stage  retweet_count  \
2234   12                10                  Skittles    NaN        210.0
197    13                10                  Arya        NaN        7608.0
2130   9                 10                  Wally       NaN        365.0
776    11                10                  Sampson     pupper     5068.0
744    11                10                  None        NaN        2589.0

       favorite_count
2234   696.0
197    31131.0
2130   1106.0
776    13102.0
744    9712.0
```

### 1.4.5 Issue #6

- The datatype for timestamp column is erroneous.

**Define**

- Convert timestamp to datetime data type.

**Code**

```
[919]: df_clean.timestamp = pd.to_datetime(df_clean.timestamp)
```

**Test**

```
[920]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   tweet_id         2356 non-null   object
 1   timestamp        2356 non-null   datetime64[ns, UTC]
 2   source           2356 non-null   object
 3   text             2356 non-null   object
 4   expanded_urls    2297 non-null   object
```

```
 5   rating_numerator    2356 non-null   int64
 6   rating_denominator  2356 non-null   int64
 7   name                2356 non-null   object
 8   dog_stage           380 non-null    object
 9   retweet_count       2354 non-null   float64
 10  favorite_count      2354 non-null   float64
dtypes: datetime64[ns, UTC](1), float64(2), int64(2), object(6)
memory usage: 220.9+ KB
```

### 1.4.6 Issue 7

**In df1, some ratings are wrong.**

**Rating_numerator and denominator should be one variable rating.**

**define**

- Change the rating_numerator and rating_denominator for oberservations with wrong value
- Oberservations with tweet_id '810984652412424192'doesn't have a valid rating, so drop this row.
- Create new column rating=rating_numerator/rating_denominator. Drop rating_numerator and rating_denominator.
- Drop oberservations with extreme ratings.

**code**

```
[921]: pd.set_option('display.max_colwidth', -1)
       df_clean.loc[df_clean.rating_denominator!
        ↪=10,['tweet_id','text','rating_numerator','rating_denominator']]
```

```
C:\Users\ADMIN\AppData\Local\Temp\ipykernel_17332\1019973338.py:1:
FutureWarning: Passing a negative integer is deprecated in version 1.0 and will
not be supported in future version. Instead, use None to not limit the column
width.
  pd.set_option('display.max_colwidth', -1)
```

```
[921]:                tweet_id  \
       313    835246439529840640
       342    832088576586297345
       433    820690176645140481
       516    810984652412424192
       784    775096608509886464
       902    758467244762497024
       1068   740373189193256964
       1120   731156023742988288
       1165   722974582966214656
       1202   716439118184652801
       1228   713900603437621249
       1254   710658690886586372
       1274   709198395643068416
```

```
1351  704054845121142784
1433  697463031882764288
1598  686035780142297088
1634  684225744407494656
1635  684422868335505415
1662  682962037429899265
1663  682808988178739200
1779  677716515794329600
1843  675853064436391936
2335  666287406224695296
```

```
            text  \
313   @jonnysun @Lin_Manuel ok jomny I know you're excited but 960/00 isn't a
valid rating, 13/10 is tho
342   @docmisterio account started on 11/15/15
433   The floofs have been released I repeat the floofs have been released.
84/70 https://t.co/NIYC820tmd
516   Meet Sam. She smiles 24/7 &amp; secretly aspires to be a reindeer. \nKeep
Sam smiling by clicking and sharing this link:\nhttps://t.co/98tB8y7y7t
https://t.co/LouL5vdvxx
784   RT @dog_rates: After so many requests, this is Bretagne. She was the last
surviving 9/11 search dog, and our second ever 14/10. RIP https:/…
902   Why does this never happen at my front door… 165/150
https://t.co/HmwrdfEfUE
1068  After so many requests, this is Bretagne. She was the last surviving 9/11
search dog, and our second ever 14/10. RIP https://t.co/XAVDNDaVgQ
1120  Say hello to this unbelievably well behaved squad of doggos. 204/170 would
try to pet all at once https://t.co/yGQI3He3xv
1165  Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a
1202  This is Bluebert. He just saw that both #FinalFur match ups are split
50/50. Amazed af. 11/10 https://t.co/Kky1DPG4iq
1228  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody
https://t.co/mpvaVxKmc1
1254  Here's a brigade of puppers. All look very prepared for whatever happens
next. 80/80 https://t.co/0eb7R1Om12
1274  From left to right:\nCletus, Jerome, Alejandro, Burp, &amp; Titson\nNone
know where camera is. 45/50 would hug all at once https://t.co/sedre1ivTK
1351  Here is a whole flock of puppers.  60/50 I'll take the lot
https://t.co/9dpcw6MdWa
1433  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once
https://t.co/HppvrYuamZ
1598  Yes I do realize a rating of 4/20 would've been fitting. However, it would
be unjust to give these cooperative pups that low of a rating
1634  Two sneaky puppers were not initially seen, moving the rating to 143/130.
Please forgive us. Thank you https://t.co/kRK51Y5ac3
1635  Someone help the girl is being mugged. Several are distracting her while
two steal her shoes. Clever puppers 121/110 https://t.co/1zfnTJLt55
```

1662  This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was just spotted by the helicopter 10/10 https://t.co/7EsP8LmSp5
1663  I'm aware that I could've said 20/16, but here at WeRateDogs we are very professional. An inconsistent rating scale is simply irresponsible
1779  IT'S PUPPERGEDDON. Total of 144/120 …I think https://t.co/ZanVtAtvIq
1843  Here we have an entire platoon of puppers. Total score: 88/80 would pet all at once https://t.co/y93p6FLvVw
2335  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood flooring. Penis on the collar. 9/10 https://t.co/d9NcXFKwLv

|      | rating_numerator | rating_denominator |
|------|------------------|--------------------|
| 313  | 960              | 0                  |
| 342  | 11               | 15                 |
| 433  | 84               | 70                 |
| 516  | 24               | 7                  |
| 784  | 9                | 11                 |
| 902  | 165              | 150                |
| 1068 | 9                | 11                 |
| 1120 | 204              | 170                |
| 1165 | 4                | 20                 |
| 1202 | 50               | 50                 |
| 1228 | 99               | 90                 |
| 1254 | 80               | 80                 |
| 1274 | 45               | 50                 |
| 1351 | 60               | 50                 |
| 1433 | 44               | 40                 |
| 1598 | 4                | 20                 |
| 1634 | 143              | 130                |
| 1635 | 121              | 110                |
| 1662 | 7                | 11                 |
| 1663 | 20               | 16                 |
| 1779 | 144              | 120                |
| 1843 | 88               | 80                 |
| 2335 | 1                | 2                  |

```
[922]: df_clean=df_clean[df_clean.tweet_id!='810984652412424192']
       df_clean.loc[df_clean.
        ↪tweet_id=='740373189193256964',['rating_numerator','rating_denominator']]=[14,10]
       df_clean.loc[df_clean.
        ↪tweet_id=='729974582966214656',['rating_numerator','rating_denominator']]=[13,10]
       df_clean.loc[df_clean.
        ↪tweet_id=='716439118184652801',['rating_numerator','rating_denominator']]=[11,10]
       df_clean.loc[df_clean.
        ↪tweet_id=='682962037429899265',['rating_numerator','rating_denominator']]=[10,10]
       df_clean.loc[df_clean.
        ↪tweet_id=='666287406224695296',['rating_numerator','rating_denominator']]=[9,10]
```

```
df_clean['rating']=df_clean.rating_numerator/df_clean.rating_denominator
df_clean=df_clean.drop(['rating_numerator','rating_denominator'],axis=1)

df_clean.loc[df_clean.rating>2,['tweet_id','text','rating']]
```

[922]:
```
                 tweet_id  \
188     855862651834028034
189     855860136149123072
290     838150277551247360
313     835246439529840640
340     832215909146226688
695     786709082849828864
763     778027034220126208
979     749981277374128128
1712    680494726643068929
2074    670842764863651840


        text  \
188    @dhmontgomery We also gave snoop dogg a 420/10 but I think that predated
your research
189    @s8n You tried very hard to portray this good boy as not so good, but you
have ultimately failed. His goodness shines through. 666/10
290    @markhoppus 182/10
313    @jonnysun @Lin_Manuel ok jomny I know you're excited but 960/00 isn't a
valid rating, 13/10 is tho
340    RT @dog_rates: This is Logan, the Chow who lived. He solemnly swears he's
up to lots of good. H*ckin magical af 9.75/10 https://t.co/yBO5wu…
695    This is Logan, the Chow who lived. He solemnly swears he's up to lots of
good. H*ckin magical af 9.75/10 https://t.co/yBO5wuqaPS
763    This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears
at random just to smile at the locals. 11.27/10 would smile back
https://t.co/QFaUiIHxHq
979    This is Atticus. He's quite simply America af. 1776/10
https://t.co/GRXwMxLBkh
1712  Here we have uncovered an entire battalion of holiday puppers. Average of
11.26/10 https://t.co/eNm2S6p9BD
2074  After so many requests… here you go.\n\nGood dogg. 420/10
https://t.co/yfAAo1gdeY


        rating
188     42.0
189     66.6
290     18.2
313     inf
340     7.5
695     7.5
```

```
763    2.7
979    177.6
1712   2.6
2074   42.0
```

[923]:
```
df_clean.loc[df_clean.tweet_id=='786709082849828864','rating']=1.0
df_clean.loc[df_clean.tweet_id=='778027034220126208','rating']=1.1
df_clean.loc[df_clean.tweet_id=='680494726643068929','rating']=1.1
df_clean=df_clean[df_clean.rating<2]
```

**test**

[924]: `df_clean.rating.sort_values()`

[924]:
```
315     0.0
1016    0.0
2091    0.1
1869    0.1
2038    0.1
         …
446     1.4
101     1.4
291     1.5
285     1.5
55      1.7
Name: rating, Length: 2348, dtype: float64
```

### 1.4.7  Issue #8

The dog_stage column hava invalid names such as none, a and an.

**Define**  Replace all rows with values none, a and an to NAN

df_clean.in_reply_to_status_id        =        df_clean.in_reply_to_status_id.astype('Int64')
df_clean.in_reply_to_user_id          =        df_clean.in_reply_to_user_id.astype('Int64')
df_clean.retweeted_status_id          =        df_clean.retweeted_status_id.astype('Int64')
df_clean.retweeted_status_user_id = df_clean.retweeted_status_user_id.astype('Int64')

**Code**

[925]:
```
df_clean.dog_stage = df_clean.dog_stage.replace('a', 'NaN')
df_clean.dog_stage = df_clean.dog_stage.replace('None', 'NaN')
df_clean.dog_stage = df_clean.dog_stage.replace('an', 'NaN')
```

**test**

[926]: `df_clean.dog_stage.value_counts()`

[926]:
```
pupper          245
doggo           83
puppo           29
```

```
doggopupper      12
floofer           9
doggopuppo        1
doggofloofer      1
Name: dog_stage, dtype: int64
```

[927]: `df_clean.dog_stage[df_clean.dog_stage == 'an']`

[927]: `Series([], Name: dog_stage, dtype: object)`

### 1.4.8 Issue #9

The name column has none and a values which should instead be represented as NaN, and also have other invalid values which are not capitalized like the aactual dog names.

**Define**  Replace None and a with NaN

**Code**

[928]: `df_clean.name.value_counts()`

[928]:
```
None          740
a              55
Charlie        12
Oliver         11
Cooper         11
              ..
Dex             1
Ace             1
Tayzie          1
Grizzie         1
Christoper      1
Name: name, Length: 957, dtype: int64
```

[929]:
```
df_clean.name = df_clean.name.replace('None', 'NaN')
df_clean.name = df_clean.name.replace('a', 'NaN')
```

[930]:
```
# removing all value with lowercase in the name column

mask = df_clean.name.str.islower()
column_name = 'name'
df_clean.loc[mask, column_name] = np.nan
```

**Test**

[931]: `df_clean.name.value_counts()`

[931]:
```
NaN           795
Charlie        12
Lucy           11
```

```
Cooper          11
Oliver          11
                 ..
Devón            1
Gert             1
Dex              1
Ace              1
Christoper       1
Name: name, Length: 932, dtype: int64
```

[932]:
```python
df_clean.name[df_clean.name == 'NaN'].value_counts()
df_clean.timestamp = pd.to_datetime(df_clean.timestamp)
```

### 1.4.9  Issue 10

Remove all rows in the `text` column that contains retweeted data.

[933]:
```python
# remove retweeted text

df_clean = df_clean[df_clean["text"].str.contains("RT @") == False]
```

[934]:
```python
df_clean.sample(10)
```

[934]:
```
              tweet_id                   timestamp  \
400     824775126675836928 2017-01-27 00:24:48+00:00
299     836989968035819520 2017-03-01 17:22:13+00:00
378     827933404142436356 2017-02-04 17:34:40+00:00
1233    712809025985978368 2016-03-24 01:11:29+00:00
667     790337589677002753 2016-10-23 23:42:19+00:00
2020    672095186491711488 2015-12-02 16:49:14+00:00
1901    674646392044941312 2015-12-09 17:46:48+00:00
2261    667549055577362432 2015-11-20 03:44:31+00:00
1895    674742531037511680 2015-12-10 00:08:50+00:00
752     778990705243029504 2016-09-22 16:13:51+00:00

      source  \
400    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
299    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
378    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1233   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
667    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2020   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
```

```
1901  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2261  <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
1895  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
752   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>

text  \
400   This is Scooter. His lack of opposable thumbs is rendering his resistance
to tickling embarrassingly moot. 12/10 would keep tickling
https://t.co/F0VWg2GztI
299   This is Mookie. He really enjoys shopping but not from such high
altitudes. Doin him quite the concern. 12/10 someone lower him
https://t.co/beWUzGVKRM
378   This is Kona. Yesterday she stopped by the department to see what it takes
to be a police pupper. 12/10 vest was only a smidge too big
https://t.co/j8D3PQJvpJ
1233  This is Stubert. He just arrived. 10/10 https://t.co/HVGs5aAKAn
667   Meet Maggie. She can hear your cells divide. 12/10 can also probably fly
https://t.co/ovE2hqXryV
2020  This is Tuco. That's the toast that killed his father. 9/10
https://t.co/ujnWy26RMe
1901  Two gorgeous dogs here. Little waddling dog is a rebel. Refuses to look at
camera. Must be a preteen. 5/10 &amp; 8/10 https://t.co/YPfw7oahbD
2261  Never seen dog like this. Breathes heavy. Tilts head in a pattern. No
bark. Shitty at fetch. Not even cordless. 1/10 https://t.co/i9iSGNn3fx
1895  Some clarification is required. The dog is singing Cher and that is more
than worthy of an 11/10. Thank you
752   This is Jay. He's really h*ckin happy about the start of fall. Sneaky
tongue slip in 2nd pic. 11/10 snuggly af https://t.co/vyx1X5eyWI

                          expanded_urls  \
400   https://twitter.com/dog_rates/status/824775126675836928/photo/1,https://tw
itter.com/dog_rates/status/824775126675836928/photo/1
299   https://twitter.com/dog_rates/status/836989968035819520/photo/1
378   https://twitter.com/dog_rates/status/827933404142436356/photo/1,https://tw
itter.com/dog_rates/status/827933404142436356/photo/1,https://twitter.com/dog_ra
tes/status/827933404142436356/photo/1
1233  https://twitter.com/dog_rates/status/712809025985978368/photo/1
667   https://twitter.com/dog_rates/status/790337589677002753/photo/1
2020  https://twitter.com/dog_rates/status/672095186491711488/photo/1
1901  https://twitter.com/dog_rates/status/674646392044941312/photo/1
2261  https://twitter.com/dog_rates/status/667549055577362432/photo/1
1895  NaN
752   https://twitter.com/dog_rates/status/778990705243029504/photo/1,https://tw
itter.com/dog_rates/status/778990705243029504/photo/1
```

```
       name  dog_stage  retweet_count  favorite_count  rating
400    Scooter  NaN         4069.0         16508.0         1.2
299    Mookie   NaN         2610.0         13879.0         1.2
378    Kona     pupper      5987.0         22180.0         1.2
1233   Stubert  NaN         7602.0         20378.0         1.0
667    Maggie   NaN         2167.0         8740.0          1.2
2020   Tuco     NaN         395.0          1063.0          0.9
1901   NaN      NaN         557.0          1533.0          0.5
2261   NaN      NaN         2454.0         6138.0          0.1
1895   NaN      NaN         55.0           527.0           1.1
752    Jay      NaN         8437.0         22342.0         1.1
```

### 1.4.10   Issue 11

There are rows with mulitiple stages for a tweet , so we are going to split them eg, the `doggopuppo`
should be `doggo`, `puppo`

```
[935]: df_clean["dog_stage"].value_counts()
```

```
[935]: pupper          224
       doggo           75
       puppo           24
       doggopupper     10
       floofer         9
       doggopuppo      1
       doggofloofer    1
       Name: dog_stage, dtype: int64
```

```
[936]: df_clean.loc[df_clean.dog_stage == 'doggopupper', 'dog_stage'] = 'doggo, pupper'
       df_clean.loc[df_clean.dog_stage == 'doggopuppo', 'dog_stage'] = 'doggo, puppo'
       df_clean.loc[df_clean.dog_stage == 'doggofloofer', 'dog_stage'] = 'doggo,␣
        ↪floofer'
```

```
[937]: df_clean["dog_stage"].value_counts()
```

```
[937]: pupper           224
       doggo            75
       puppo            24
       doggo, pupper    10
       floofer          9
       doggo, puppo     1
       doggo, floofer   1
       Name: dog_stage, dtype: int64
```

```
[938]: df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2168 entries, 0 to 2355
```

```
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        2168 non-null   object
 1   timestamp       2168 non-null   datetime64[ns, UTC]
 2   source          2168 non-null   object
 3   text            2168 non-null   object
 4   expanded_urls   2114 non-null   object
 5   name            2119 non-null   object
 6   dog_stage       344 non-null    object
 7   retweet_count   2168 non-null   float64
 8   favorite_count  2168 non-null   float64
 9   rating          2168 non-null   float64
dtypes: datetime64[ns, UTC](1), float64(3), object(6)
memory usage: 186.3+ KB
```

## 1.5  Storing Data

Save gathered, assessed, and cleaned master dataset to a CSV file named "twitter_archive_master.csv".

[939]: `df_clean.sample(10)`

[939]:
```
                tweet_id                  timestamp  \
648   793120401413079041 2016-10-31 16:00:13+00:00
2324  666421158376562688 2015-11-17 01:02:40+00:00
1805  676942428000112642 2015-12-16 01:50:26+00:00
1854  675534494439489536 2015-12-12 04:35:48+00:00
1649  683742671509258241 2016-01-03 20:12:10+00:00
1357  703407252292673536 2016-02-27 02:32:12+00:00
1119  731285275100512256 2016-05-14 00:49:30+00:00
1604  685906723014619143 2016-01-09 19:31:20+00:00
1018  746818907684614144 2016-06-25 21:34:37+00:00
855   764857477905154048 2016-08-14 16:13:27+00:00

    source  \
648   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2324  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1805  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1854  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1649  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1357  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
```

```
1119   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1604   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1018   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
855    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>

        text  \
648    This is Robin. She's desperately trying to do me a frighten, but her
tongue drastically decreases her spook value. Still 11/10 great effort
https://t.co/sxMrsOZ8zb
2324   *internally screaming* 12/10 https://t.co/YMcrXC2Y6R
1805   Who leaves the last cupcake just sitting there? 9/10
https://t.co/PWMqAoEx2a
1854  Seriously guys?! Only send in dogs. I only rate dogs. This is a baby black
bear… 11/10 https://t.co/H7kpabTfLj
1649  Meet Sebastian. He's a womanizer. Romantic af. Always covered in flower
petals. Also a poet. 11/10 dreamy as hell https://t.co/eoL1bCpWCg
1357   This pupper doesn't understand gates. 10/10 so close
https://t.co/GUbFF4o6dZ
1119   This is Solomon. He's a Beneroo Cumberflop. 12/10 would hug passionately
https://t.co/5phLAnGPTP
1604   This is Olive. He's stuck in a sleeve. 9/10 damn it Olive
https://t.co/NnLjg6BgyF
1018   Guys… Dog Jesus 2.0\n13/10 buoyant af https://t.co/CuNA7OwfKQ
855    This is Aubie. He has paws for days. Nibbling tables is one of his
priorities. Second only to being cuddly af. 12/10 https://t.co/cBIFBsCRz6

                                  expanded_urls  \
648   https://twitter.com/dog_rates/status/793120401413079041/photo/1
2324  https://twitter.com/dog_rates/status/666421158376562688/photo/1
1805  https://twitter.com/dog_rates/status/676942428000112642/photo/1
1854  https://twitter.com/dog_rates/status/675534494439489536/photo/1
1649  https://twitter.com/dog_rates/status/683742671509258241/photo/1
1357  https://twitter.com/dog_rates/status/703407252292673536/photo/1
1119  https://twitter.com/dog_rates/status/731285275100512256/photo/1
1604  https://twitter.com/dog_rates/status/685906723014619143/photo/1,https://tw
itter.com/dog_rates/status/685906723014619143/photo/1
1018  https://twitter.com/dog_rates/status/746818907684614144/photo/1
855   https://twitter.com/dog_rates/status/764857477905154048/photo/1

          name dog_stage  retweet_count  favorite_count  rating
648   Robin       NaN       4551.0         14202.0        1.1
2324  NaN         NaN       118.0          327.0          1.2
1805  NaN         NaN       855.0          2261.0         0.9
```

```
1854  NaN        NaN      470.0    1953.0    1.1
1649  Sebastian  NaN      3781.0   7162.0    1.1
1357  NaN        pupper   786.0    2685.0    1.0
1119  Solomon    NaN      1090.0   3798.0    1.2
1604  Olive      NaN      3303.0   8277.0    0.9
1018  NaN        NaN      1944.0   5807.0    1.3
855   Aubie      NaN      2029.0   7099.0    1.2
```

[940]:  `df1_clean.sample(10)`

[940]:
```
                  tweet_id  \
1882  847157206088847362
535   676942428000112642
924   702321140488925184
1415  771102124360998913
811   692417313023332352
2052  887517139158093824
1632  805826884734976000
555   677662372920729601
657   682303737705140231
66    667176164155375616


         jpg_url  \
1882  https://pbs.twimg.com/media/C8GO_CMWsAAjjAY.jpg
535   https://pbs.twimg.com/media/CWT7imQXIAMwpQ2.jpg
924   https://pbs.twimg.com/media/Cb8lWafWEAA2q93.jpg
1415  https://pbs.twimg.com/media/CrOBSfgXgAABsTE.jpg
811   https://pbs.twimg.com/media/CZv13u5WYAA6wQe.jpg
2052  https://pbs.twimg.com/ext_tw_video_thumb/887517108413886465/pu/img/WanJKws
sZj4VJvL9.jpg
1632  https://pbs.twimg.com/ext_tw_video_thumb/805826823359631360/pu/img/yr_fFOT
ZCR-B70p2.jpg
555   https://pbs.twimg.com/media/CWeKTZTWsAA5R3Z.jpg
657   https://pbs.twimg.com/media/CXgHoLnWAAA8i52.jpg
66    https://pbs.twimg.com/media/CUJJLtWWsAE-go5.jpg


      img_num                         p1   p1_conf  p1_dog  \
1882  2        Staffordshire_bullterrier   0.219609  True
535   1        black-footed_ferret        0.707199  False
924   3        West_Highland_white_terrier 0.769159 True
1415  1        Labrador_retriever         0.568789  True
811   1        bison                      0.208922  False
2052  1        limousine                  0.130432  False
1632  1        Siberian_husky             0.248926  True
555   1        cowboy_hat                 0.256110  False
657   1        seat_belt                  0.997659  False
66    1        soft-coated_wheaten_terrier 0.318981 True
```

```
                                          p2    p2_conf  p2_dog  \
1882  American_Staffordshire_terrier  0.178671   True
535   polecat                         0.154630   False
924   Scotch_terrier                  0.064369   True
1415  pug                             0.179918   True
811   mink                            0.169945   False
2052  tow_truck                       0.029175   False
1632  American_Staffordshire_terrier  0.098313   True
555   trench_coat                     0.127086   False
657   Lakeland_terrier                0.001731   True
66    Lakeland_terrier                0.215218   True


                             p3    p3_conf  p3_dog
1882  pug                       0.123271   True
535   weasel                    0.097626   False
924   Old_English_sheepdog      0.043763   True
1415  Staffordshire_bullterrier 0.034437   True
811   polecat                   0.144494   False
2052  shopping_cart             0.026321   False
1632  Eskimo_dog                0.080188   True
555   cloak                     0.071439   False
657   Airedale                  0.000204   True
66    toy_poodle                0.106014   True
```

[941]: `df2_clean.sample(10)`

[941]:
```
              tweet_id  retweet_count  favorite_count
1494  692530551048294401  474            2021
336   832397543355072512  2548           13126
1589  686606069955735556  607            2071
735   780931614150983680  8536           24192
575   800859414831898624  113            778
729   781661882474196992  3129           11634
211   851861385021730816  23             0
608   797545162159308800  5656           16198
1208  715680795826982913  1813           4719
1145  726935089318363137  2738           7496
```

[942]: `df1_clean.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   tweet_id  2075 non-null   object
 1   jpg_url   2075 non-null   object
```

```
  2  img_num    2075 non-null    int64
  3  p1         2075 non-null    object
  4  p1_conf    2075 non-null    float64
  5  p1_dog     2075 non-null    bool
  6  p2         2075 non-null    object
  7  p2_conf    2075 non-null    float64
  8  p2_dog     2075 non-null    bool
  9  p3         2075 non-null    object
 10  p3_conf    2075 non-null    float64
 11  p3_dog     2075 non-null    bool
dtypes: bool(3), float64(3), int64(1), object(5)
memory usage: 152.1+ KB
```

[943]: ```
df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2168 entries, 0 to 2355
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        2168 non-null   object
 1   timestamp       2168 non-null   datetime64[ns, UTC]
 2   source          2168 non-null   object
 3   text            2168 non-null   object
 4   expanded_urls   2114 non-null   object
 5   name            2119 non-null   object
 6   dog_stage        344 non-null   object
 7   retweet_count   2168 non-null   float64
 8   favorite_count  2168 non-null   float64
 9   rating          2168 non-null   float64
dtypes: datetime64[ns, UTC](1), float64(3), object(6)
memory usage: 186.3+ KB
```

[944]: ```
master_df = pd.merge(df_clean, df1_clean)
```

[945]: ```
master_df
```

[945]: 
```
                tweet_id                  timestamp  \
0       892420643555336193 2017-08-01 16:23:56+00:00
1       892177421306343426 2017-08-01 00:17:27+00:00
2       891815181378084864 2017-07-31 00:18:03+00:00
3       891689557279858688 2017-07-30 15:58:51+00:00
4       891327558926688256 2017-07-29 16:00:24+00:00
...                    ...                       ...
1986    666049248165822465 2015-11-16 00:24:50+00:00
1987    666044226329800704 2015-11-16 00:04:52+00:00
1988    666033412701032449 2015-11-15 23:21:54+00:00
1989    666029285002620928 2015-11-15 23:05:30+00:00
```

```
1990  666020888022790149 2015-11-15 22:32:08+00:00

   source  \
0     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
3     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
4     <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
…
…
1986  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1987  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1988  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1989  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1990  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>


     text  \
0     This is Phineas. He's a mystical boy. Only ever appears in the hole of a
donut. 13/10 https://t.co/MgUWQ76dJU
1     This is Tilly. She's just checking pup on you. Hopes you're doing ok. If
not, she's available for pats, snugs, boops, the whole bit. 13/10
https://t.co/0Xxu71qeIV
2     This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in the tall
grass. You never know when one may strike. 12/10 https://t.co/wUnZnhtVJB
3     This is Darla. She commenced a snooze mid meal. 13/10 happens to the best
of us https://t.co/tD36da7qLQ
4     This is Franklin. He would like you to stop calling him "cute." He is a
very fierce shark and should be respected as such. 12/10 #BarkWeek
https://t.co/AtUZn91f7f
…
…
1986  Here we have a 1949 1st generation vulpix. Enjoys sweat tea and Fox News.
Cannot be phased. 5/10 https://t.co/4B7cOc1EDq
1987  This is a purebred Piers Morgan. Loves to Netflix and chill. Always looks
like he forgot to unplug the iron. 6/10 https://t.co/DWnyCjf2mx
1988  Here is a very happy pup. Big fan of well-maintained decks. Just look at
that tongue. 9/10 would cuddle af https://t.co/y671yMhoiR
```

```
1989  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2
dogs here. 7/10 would walk the shit out of https://t.co/r7mOb2m0UI
1990  Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of
relaxing on stair. 8/10 would pet https://t.co/BLDqew2Ijj


                                          expanded_urls  \
0        https://twitter.com/dog_rates/status/892420643555336193/photo/1
1        https://twitter.com/dog_rates/status/892177421306343426/photo/1
2        https://twitter.com/dog_rates/status/891815181378084864/photo/1
3        https://twitter.com/dog_rates/status/891689557279858688/photo/1
4        https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://tw
itter.com/dog_rates/status/891327558926688256/photo/1
…
…
1986  https://twitter.com/dog_rates/status/666049248165822465/photo/1
1987  https://twitter.com/dog_rates/status/666044226329800704/photo/1
1988  https://twitter.com/dog_rates/status/666033412701032449/photo/1
1989  https://twitter.com/dog_rates/status/666029285002620928/photo/1
1990  https://twitter.com/dog_rates/status/666020888022790149/photo/1

          name dog_stage  retweet_count  favorite_count  rating  … img_num  \
0      Phineas     NaN       8853.0          39467.0         1.3    …    1
1        Tilly     NaN       6514.0          33819.0         1.3    …    1
2       Archie     NaN       4328.0          25461.0         1.2    …    1
3        Darla     NaN       8964.0          42908.0         1.3    …    1
4     Franklin     NaN       9774.0          41048.0         1.2    …    2
…        …    …           …               …              …     … ..
1986       NaN     NaN         41.0            111.0         0.5    …    1
1987       NaN     NaN        147.0            311.0         0.6    …    1
1988       NaN     NaN         47.0            128.0         0.9    …    1
1989       NaN     NaN         48.0            132.0         0.7    …    1
1990       NaN     NaN        532.0           2535.0         0.8    …    1

                          p1   p1_conf p1_dog                    p2   p2_conf  \
0               orange   0.097049  False    bagel              0.085851
1            Chihuahua   0.323581   True  Pekinese            0.090647
2            Chihuahua   0.716012   True  malamute            0.078253
3           paper_towel  0.170278  False  Labrador_retriever  0.168086
4               basset   0.555712   True  English_springer    0.225770
…                …         …    …                               …         …
1986  miniature_pinscher  0.560311   True  Rottweiler          0.243682
1987  Rhodesian_ridgeback  0.408143  True  redbone             0.360687
1988     German_shepherd  0.596461   True  malinois            0.138584
1989             redbone  0.506826   True  miniature_pinscher  0.074192
1990  Welsh_springer_spaniel 0.465074 True  collie             0.156665

          p2_dog                      p3    p3_conf  p3_dog
```

```
0      False   banana                      0.076110  False
1      True    papillon                    0.068957  True
2      True    kelpie                      0.031379  True
3      True    spatula                     0.040836  False
4      True    German_short-haired_pointer 0.175219  True
...    ...     ...                         ...       ...
1986   True    Doberman                    0.154629  True
1987   True    miniature_pinscher          0.222752  True
1988   True    bloodhound                  0.116197  True
1989   True    Rhodesian_ridgeback         0.072010  True
1990   True    Shetland_sheepdog           0.061428  True

[1991 rows x 21 columns]
```

[946]: `master_df["dog_stage"].value_counts()`

[946]:
```
pupper          203
doggo           63
puppo           22
doggo, pupper   9
floofer         7
doggo, puppo    1
doggo, floofer  1
Name: dog_stage, dtype: int64
```

[947]:
```
master_df.loc[master_df.dog_stage == 'doggopupper', 'dog_stage'] = 'doggo,␣
 ↪pupper'
master_df.loc[master_df.dog_stage == 'doggopuppo', 'dog_stage'] = 'doggo, puppo'
master_df.loc[master_df.dog_stage == 'doggofloofer', 'dog_stage'] = 'doggo,␣
 ↪floofer'
```

[948]: `master_df["dog_stage"].value_counts()`

[948]:
```
pupper          203
doggo           63
puppo           22
doggo, pupper   9
floofer         7
doggo, puppo    1
doggo, floofer  1
Name: dog_stage, dtype: int64
```

[949]: `master_df["name"].value_counts()`

[949]:
```
NaN       600
Charlie   11
Lucy      10
```

```
Cooper        10
Oliver        10
              ..
Bookstore      1
Atticus        1
Jersey         1
Shiloh         1
Christoper     1
Name: name, Length: 914, dtype: int64
```

[950]: `master_df["name"].value_counts()`

[950]:
```
NaN          600
Charlie       11
Lucy          10
Cooper        10
Oliver        10
              ..
Bookstore      1
Atticus        1
Jersey         1
Shiloh         1
Christoper     1
Name: name, Length: 914, dtype: int64
```

[951]: `master_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1991 entries, 0 to 1990
Data columns (total 21 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        1991 non-null   object
 1   timestamp       1991 non-null   datetime64[ns, UTC]
 2   source          1991 non-null   object
 3   text            1991 non-null   object
 4   expanded_urls   1991 non-null   object
 5   name            1948 non-null   object
 6   dog_stage       306 non-null    object
 7   retweet_count   1991 non-null   float64
 8   favorite_count  1991 non-null   float64
 9   rating          1991 non-null   float64
 10  jpg_url         1991 non-null   object
 11  img_num         1991 non-null   int64
 12  p1              1991 non-null   object
 13  p1_conf         1991 non-null   float64
 14  p1_dog          1991 non-null   bool
 15  p2              1991 non-null   object
```

```
16  p2_conf          1991 non-null    float64
17  p2_dog           1991 non-null    bool
18  p3               1991 non-null    object
19  p3_conf          1991 non-null    float64
20  p3_dog           1991 non-null    bool
dtypes: bool(3), datetime64[ns, UTC](1), float64(6), int64(1), object(10)
memory usage: 301.4+ KB
```

[952]: `master_df.to_csv('twitter_archive_master.csv', index=True, encoding='utf-8')`

[953]: `df3 = pd.read_csv('twitter_archive_master.csv')`

[954]: 
```
df3.timestamp = pd.to_datetime(df3.timestamp)
df3.tweet_id = df3.tweet_id.astype('str')
```

[955]: `df3.info(10)`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 22 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Unnamed: 0      1991 non-null    int64
 1   tweet_id        1991 non-null    object
 2   timestamp       1991 non-null    datetime64[ns, UTC]
 3   source          1991 non-null    object
 4   text            1991 non-null    object
 5   expanded_urls   1991 non-null    object
 6   name            1348 non-null    object
 7   dog_stage       306 non-null     object
 8   retweet_count   1991 non-null    float64
 9   favorite_count  1991 non-null    float64
 10  rating          1991 non-null    float64
 11  jpg_url         1991 non-null    object
 12  img_num         1991 non-null    int64
 13  p1              1991 non-null    object
 14  p1_conf         1991 non-null    float64
 15  p1_dog          1991 non-null    bool
 16  p2              1991 non-null    object
 17  p2_conf         1991 non-null    float64
 18  p2_dog          1991 non-null    bool
 19  p3              1991 non-null    object
 20  p3_conf         1991 non-null    float64
 21  p3_dog          1991 non-null    bool
dtypes: bool(3), datetime64[ns, UTC](1), float64(6), int64(2), object(10)
memory usage: 301.5+ KB
```

[956]: `df3.sample(10)`

```
[956]:         Unnamed: 0            tweet_id                  timestamp  \
       1006  1006      707741517457260545  2016-03-10 01:35:01+00:00
       1181  1181      692530551048294401  2016-01-28 02:12:04+00:00
       366   366       813187593374461952  2016-12-26 01:00:05+00:00
       285   285       827600520311402496  2017-02-03 19:31:54+00:00
       611   611       767191397493538821  2016-08-21 02:47:37+00:00
       25    25        886983233522544640  2017-07-17 16:17:36+00:00
       460   460       793271401113350145  2016-11-01 02:00:14+00:00
       172   172       849412302885593088  2017-04-05 00:04:08+00:00
       579   579       772877495989305348  2016-09-05 19:22:09+00:00
       233   233       835264098648616962  2017-02-24 23:04:14+00:00

          source  \
1006  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
1181  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
366   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
285   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
611   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
25    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
460   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
172   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
579   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
233   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>

           text  \
1006  When your roommate eats your leftover Chili's but you pretend it's no big
deal cuz you fat anyway. 10/10 head up pup https://t.co/0nMgoue8IR
1181  Say hello to Cody. He's been to like 80 countries and is way more cultured
than you. He wanted me to say that. 10/10 https://t.co/Iv3flDTpXu
366   We only rate dogs. Please don't send in other things like this very good
Christmas tree. Thank you… 13/10 https://t.co/rvSANEsQZJ
285   Meet Tucker. It's his birthday. He's pupset with you because you're too
busy playing @GoodDogsGame to celebrate. 13/10 would put down phone
https://t.co/vrppizPGdb
611   I don't know any of the backstory behind this picture but for some reason
I'm crying. 13/10 for owner and doggo https://t.co/QOKZdus9TT
25    This is Maya. She's very shy. Rarely leaves her cup. 13/10 would find her
an environment to thrive in https://t.co/I6oNy0CgiT
```

460   Meet Dave. It's his favorite day of the year. He gets to fulfill his dream
of being a dinosaur. 12/10 inspirational af https://t.co/MgQSdfZGPN
172   This is Noosh. He noticed you were in the shower and thought you could use
some company. 12/10 h*ckin loyal https://t.co/Uq3ChFgWA3
579   You need to watch these two doggos argue through a cat door. Both 11/10
https://t.co/qEP31epKEV
233   Meet Lola. Her hobbies include being precious af and using her foot as a
toothbrush. 12/10 Lola requests your help\n\nhttps://t.co/FYFyHh7rir
https://t.co/IiB7ggduoU

         expanded_urls  \
1006  https://twitter.com/dog_rates/status/707741517457260545/photo/1
1181  https://twitter.com/dog_rates/status/692530551048294401/photo/1
366   https://twitter.com/dog_rates/status/813187593374461952/photo/1
285   https://twitter.com/dog_rates/status/827600520311402496/photo/1
611   https://twitter.com/dog_rates/status/767191397493538821/photo/1
25    https://twitter.com/dog_rates/status/886983233522544640/photo/1,https://tw
itter.com/dog_rates/status/886983233522544640/photo/1
460   https://twitter.com/dog_rates/status/793271401113350145/photo/1
172   https://twitter.com/dog_rates/status/849412302885593088/photo/1,https://tw
itter.com/dog_rates/status/849412302885593088/photo/1,https://twitter.com/dog_ra
tes/status/849412302885593088/photo/1,https://twitter.com/dog_rates/status/84941
2302885593088/photo/1
579   https://twitter.com/dog_rates/status/772877495989305348/video/1
233   https://www.gofundme.com/lolas-life-saving-surgery-funds,https://twitter.c
om/dog_rates/status/835264098648616962/photo/1,https://twitter.com/dog_rates/sta
tus/835264098648616962/photo/1

         name dog_stage  retweet_count  favorite_count  …  img_num  \
1006  NaN      NaN        696.0          2718.0          …  1
1181  Cody     NaN        474.0          2021.0          …  1
366   NaN      NaN        5096.0         22085.0         …  1
285   Tucker   NaN        1082.0         8143.0          …  1
611   NaN      doggo      4406.0         13643.0         …  1
25    Maya     NaN        8084.0         35859.0         …  2
460   Dave     NaN        2763.0         9677.0          …  1
172   Noosh    NaN        3487.0         17039.0         …  4
579   NaN      NaN        4501.0         9555.0          …  1
233   Lola     NaN        1939.0         8503.0          …  2

                    p1    p1_conf p1_dog                        p2   p2_conf  \
1006  whippet           0.738371  True   Italian_greyhound        0.191789
1181  Siberian_husky    0.486428  True   Eskimo_dog               0.448518
366   golden_retriever  0.888181  True   Labrador_retriever       0.042312
285   Pembroke          0.325638  True   golden_retriever         0.317235
611   patio             0.708665  False  boathouse                0.110056
25    Chihuahua         0.793469  True   toy_terrier              0.143528

```
460   Siberian_husky   0.231695  True   Eskimo_dog                 0.206749
172   schipperke       0.907559  True   crossword_puzzle           0.017934
579   tabby            0.218303  False  Norwegian_elkhound         0.138523
233   hyena            0.736871  False  Chesapeake_Bay_retriever   0.087503


       p2_dog                                   p3    p3_conf  p3_dog
1006   True    American_Staffordshire_terrier   0.020126  True
1181   True    white_wolf                       0.041506  False
366    True    Saluki                           0.009702  True
285    True    Labrador_retriever               0.116087  True
611    False   pier                             0.039532  False
25     True    can_opener                       0.032253  False
460    True    Pembroke                         0.070120  True
172    False   Chihuahua                        0.016191  True
579    True    wombat                           0.074217  False
233    True    meerkat                          0.042058  False


[10 rows x 22 columns]
```

[957]: `# df3 = df3.drop(columns = ['Unnamed: 0'])`

[958]: `df3.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 22 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      1991 non-null   int64
 1   tweet_id        1991 non-null   object
 2   timestamp       1991 non-null   datetime64[ns, UTC]
 3   source          1991 non-null   object
 4   text            1991 non-null   object
 5   expanded_urls   1991 non-null   object
 6   name            1348 non-null   object
 7   dog_stage       306 non-null    object
 8   retweet_count   1991 non-null   float64
 9   favorite_count  1991 non-null   float64
 10  rating          1991 non-null   float64
 11  jpg_url         1991 non-null   object
 12  img_num         1991 non-null   int64
 13  p1              1991 non-null   object
 14  p1_conf         1991 non-null   float64
 15  p1_dog          1991 non-null   bool
 16  p2              1991 non-null   object
 17  p2_conf         1991 non-null   float64
 18  p2_dog          1991 non-null   bool
 19  p3              1991 non-null   object
```

```
20  p3_conf         1991 non-null   float64
21  p3_dog          1991 non-null   bool
dtypes: bool(3), datetime64[ns, UTC](1), float64(6), int64(2), object(10)
memory usage: 301.5+ KB
```

## 1.6 Analyzing and Visualizing Data

In this section, analyze and visualize your wrangled data. You must produce at least **three (3) insights and one (1) visualization.**

### 1.6.1 Insights:

1. Among the 4 dog stages, pupper has the biggest frequency, but pupper also gets the lowest favorite counts and retweet counts and rating.
2. The mean for rating is 1.055 and the 3 most common ratings are 1.2,1.0,1.1, and rating frequency becomes smaller as the rating becomes extreme.
3. Posts with extreme ratings get more favorites and retweets.
4. Post with rating 0.6 gets the highest favorite counts and retweet counts.
5. Posts with rating 1.7 gets the lowest and equal favorite and retweet count

### 1.6.2 Visualization

[959]: `df3.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 22 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      1991 non-null   int64
 1   tweet_id        1991 non-null   object
 2   timestamp       1991 non-null   datetime64[ns, UTC]
 3   source          1991 non-null   object
 4   text            1991 non-null   object
 5   expanded_urls   1991 non-null   object
 6   name            1348 non-null   object
 7   dog_stage       306 non-null    object
 8   retweet_count   1991 non-null   float64
 9   favorite_count  1991 non-null   float64
 10  rating          1991 non-null   float64
 11  jpg_url         1991 non-null   object
 12  img_num         1991 non-null   int64
 13  p1              1991 non-null   object
 14  p1_conf         1991 non-null   float64
 15  p1_dog          1991 non-null   bool
 16  p2              1991 non-null   object
 17  p2_conf         1991 non-null   float64
 18  p2_dog          1991 non-null   bool
 19  p3              1991 non-null   object
```

```
 20   p3_conf          1991 non-null    float64
 21   p3_dog           1991 non-null    bool
dtypes: bool(3), datetime64[ns, UTC](1), float64(6), int64(2), object(10)
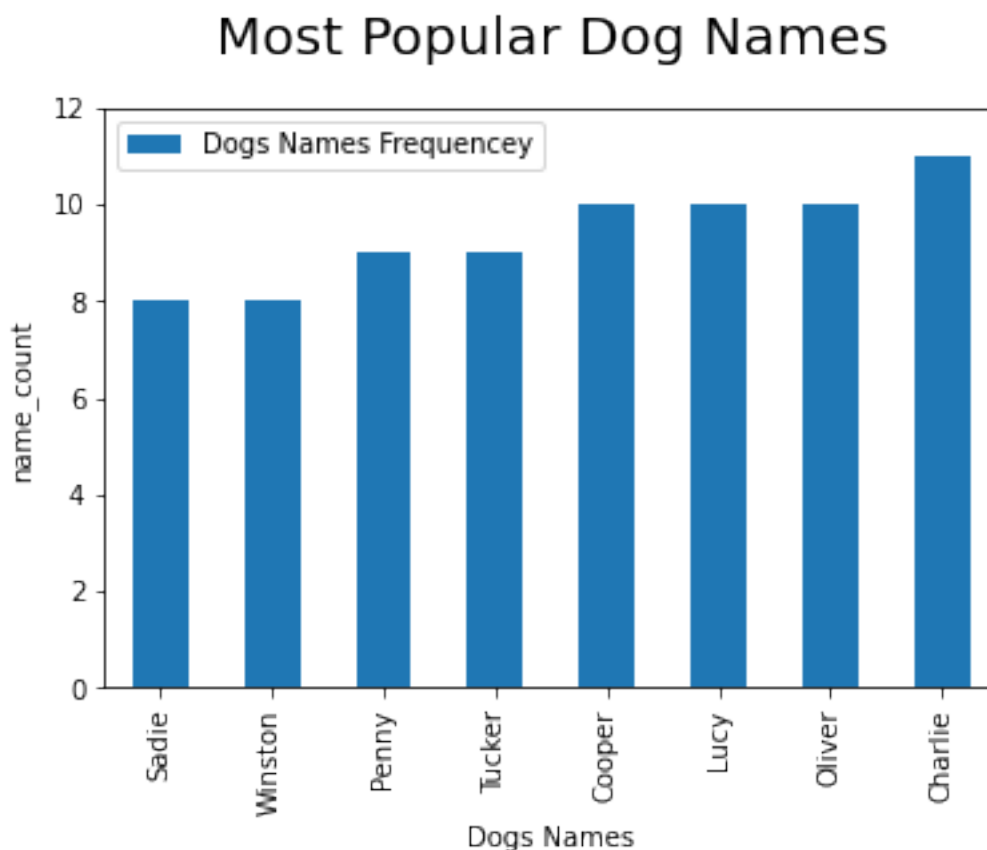memory usage: 301.5+ KB
```

[960]:
```
named_dogs = df3.query('name == name')
named_dogs_grouped = named_dogs.groupby('name').count()[['tweet_id']]
named_dogs_grouped.rename(columns={'tweet_id':'name_count'}, inplace=True)
named_dogs_grouped.query('name_count >= 8').sort_values(by=['name_count']).plot.
  ↪bar()
plt.ylim(top=12)
plt.title("Most Popular Dog Names",{'fontsize': 20},pad=20)
plt.xlabel("Dogs Names")
plt.ylabel("name_count")
plt.legend(["Dogs Names Frequencey"])
```

[960]:  <matplotlib.legend.Legend at 0x1cc1a36c8e0>



[961]:
```
tweet_sources = df3.groupby('source').count()[['tweet_id']]
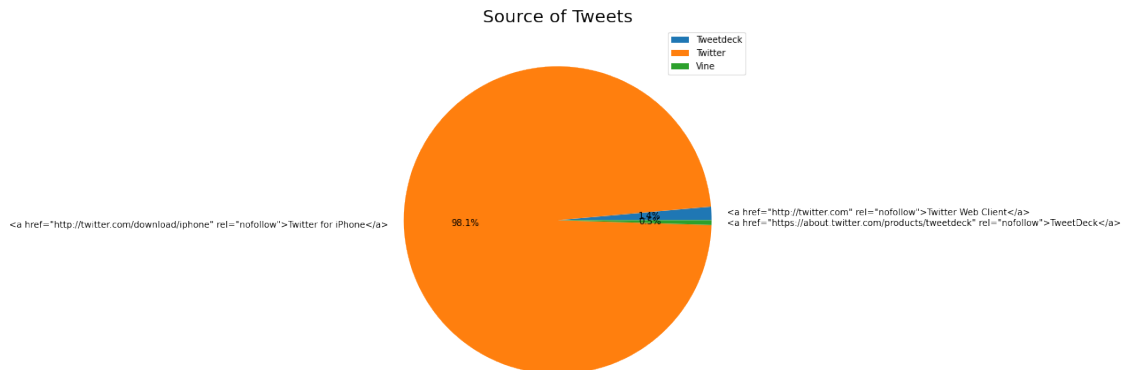tweet_sources.rename(columns={'tweet_id': 'source_count'}, inplace=True)
```

```
tweet_sources['source_percentage'] = tweet_sources.source_count / tweet_sources.
 ↪source_count.sum() * 100
tweet_sources['source_percentage'].plot.pie(figsize=(10,8), autopct='%1.1f%%',
        explode=None)
plt.title("Source of Tweets", {'fontsize': 20})
plt.legend(["Tweetdeck", "Twitter", "Vine", "iPhone"])
plt.ylabel("")
```

[961]: Text(0, 0.5, '')



[962]: 
```
values=np.sort(df3.rating.unique())
values
```

[962]: 
```
array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. , 1.1, 1.2,
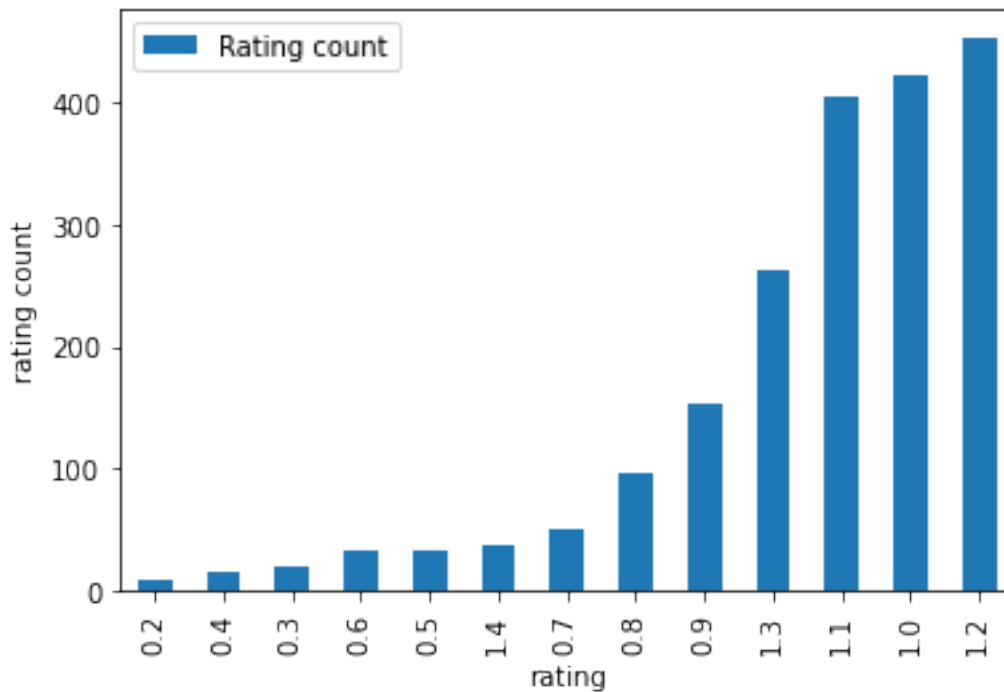       1.3, 1.4])
```

[963]: 
```
ratings = df3.query('rating == rating')
ratings_grouped = ratings.groupby('rating').count()[['tweet_id']]
ratings_grouped.rename(columns={'tweet_id':'rating_count'}, inplace=True)
ratings_grouped.query('rating_count >= 8').sort_values(by=['rating_count']).
 ↪plot.bar()
# plt.ylim(top=100)
plt.title("Rating distribution",{'fontsize': 20},pad=20)
plt.xlabel("rating")
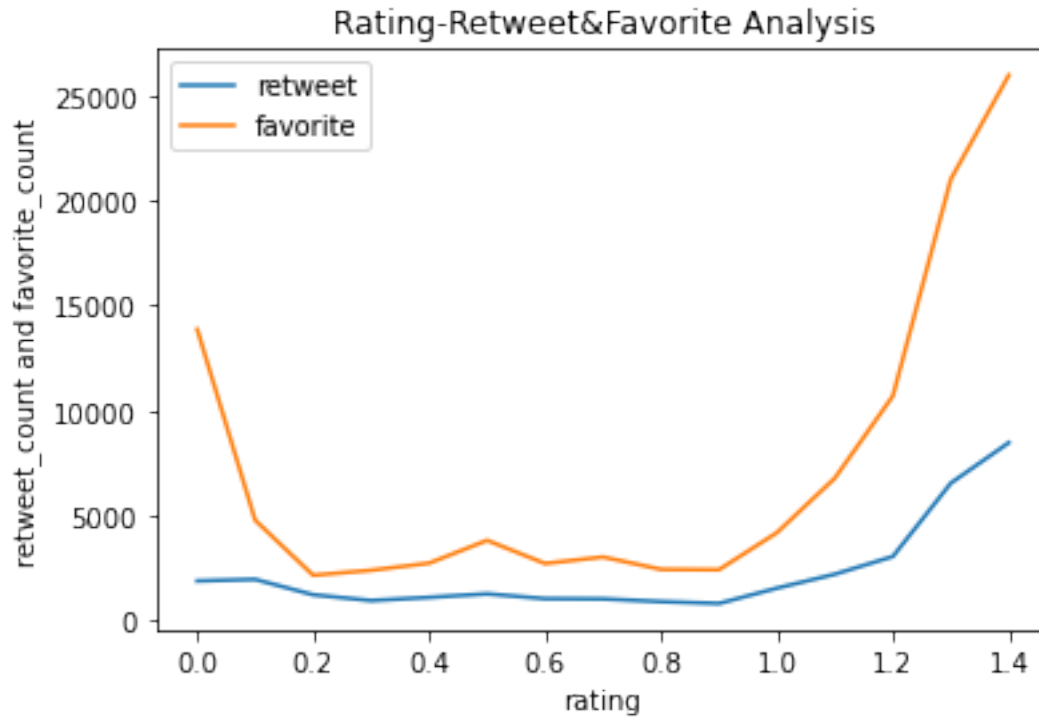plt.ylabel("rating count")
plt.legend(["Rating count"])
```

[963]: <matplotlib.legend.Legend at 0x1cc141d4160>

# Rating distribution



```
[964]: rating_analyze=pd.DataFrame(columns=['rating','number','retweet','favorite'])
       rating_analyze['rating']=values
       for x in rating_analyze['rating']:
           this_df=df_clean[df_clean.rating==x]
           rating_analyze.loc[rating_analyze.rating==x,'number']=this_df['tweet_id'].
        ↪count()
           rating_analyze.loc[rating_analyze.rating==x,'retweet']=this_df.
        ↪retweet_count.mean()
           rating_analyze.loc[rating_analyze.rating==x,'favorite']=this_df.
        ↪favorite_count.mean()
       rating_analyze[['rating','retweet','favorite']].plot(x='rating',␣
        ↪title='Rating-Retweet&Favorite Analysis')
       plt.ylabel("retweet_count and favorite_count")
```

[964]: Text(0, 0.5, 'retweet_count and favorite_count')

Rating-Retweet&Favorite Analysis

```
[965]: df3.describe()
```

```
[965]:         Unnamed: 0   retweet_count   favorite_count        rating      img_num  \
        count  1991.000000  1991.000000     1991.000000      1991.000000  1991.000000
        mean   995.000000   2766.526871     8899.343044      1.055450     1.203415
        std    574.896512   4678.024127     12221.969511     0.217822     0.561144
        min    0.000000     16.000000       81.000000        0.000000     1.000000
        25%    497.500000   623.500000      1976.000000      1.000000     1.000000
        50%    995.000000   1350.000000     4127.000000      1.100000     1.000000
        75%    1492.500000  3220.000000     11321.000000     1.200000     1.000000
        max    1990.000000  79515.000000    132810.000000    1.400000     4.000000


                   p1_conf       p2_conf       p3_conf
        count  1991.000000   1.991000e+03  1.991000e+03
        mean   0.594082      1.345157e-01  6.027009e-02
        std    0.271856      1.007215e-01  5.092272e-02
        min    0.044333      1.011300e-08  1.740170e-10
        25%    0.362880      5.397835e-02  1.614795e-02
        50%    0.587764      1.175660e-01  4.948690e-02
        75%    0.845942      1.953115e-01  9.164355e-02
        max    1.000000      4.880140e-01  2.734190e-01
```