wrangle_act

August 7, 2021

1 Wrangling and Analyze Data

1.1 Gathering

```
In [1]: # Importing the required modules
        import pandas as pd
        import numpy as np
        import tweepy
        import json
        import requests
        import seaborn as sns
        from matplotlib import pyplot as plt
        %matplotlib inline
```

```
In [2]: # Reading the required files(csv and tsv)
        df_twitter = pd.read_csv('twitter-archive-enhanced.csv')
        df_img_pred = pd.read_csv('image-predictions.tsv', sep = '\t')
```

```
In [3]: import tweepy
        from tweepy import OAuthHandler
        import json
        from timeit import default_timer as timer

        consumer_secret = 'HIDDEN'
        access_token = 'HIDDEN'
        access_secret = 'HIDDEN'

        auth = OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_secret)

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

        tweet_ids = df_1.tweet_id.values
        len(tweet_ids)

        count = 0
        fails_dict = {}
```

```python
        start = timer()

        with open('tweet_json.txt', 'w') as outfile:
            for tweet_id in tweet_ids:
                count += 1
                print(str(count) + ": " + str(tweet_id))
                try:
                    tweet = api.get_status(tweet_id, tweet_mode='extended')
                    print("Success")
                    json.dump(tweet._json, outfile)
                    outfile.write('\n')
                except tweepy.TweepError as e:
                    print("Fail")
                    fails_dict[tweet_id] = e
                    pass
        end = timer()
        print(end - start)
        print(fails_dict)
```

Out[3]: 'import tweepy\nfrom tweepy import OAuthHandler\nimport json\nfrom timeit import default

In [4]: # Reading json file
        df_tweets = pd.read_json("tweet-json.txt", lines = True)

1.2 Assess

In [5]: df_twitter.shape

Out[5]: (2356, 17)

In [6]: df_twitter

Out[6]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
        0   892420643555336193   NaN   NaN
        1   892177421306343426   NaN   NaN
        2   891815181378084864   NaN   NaN
        3   891689557279858688   NaN   NaN
        4   891327558926688256   NaN   NaN
        5   891087950875897856   NaN   NaN
        6   890971913173991426   NaN   NaN
        7   890729181411237888   NaN   NaN
        8   890609185150312448   NaN   NaN
        9   890240255349198849   NaN   NaN
        10   890006608113172480   NaN   NaN
        11   889880896479866881   NaN   NaN

12	889665388333682689	NaN	NaN
13	889638837579907072	NaN	NaN
14	889531135344209921	NaN	NaN
15	889278841981685760	NaN	NaN
16	888917238123831296	NaN	NaN
17	888804989199671297	NaN	NaN
18	888554962724278272	NaN	NaN
19	888202515573088257	NaN	NaN
20	888078434458587136	NaN	NaN
21	887705289381826560	NaN	NaN
22	887517139158093824	NaN	NaN
23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
...	...	...	...
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2336	666273097616637952	NaN	NaN
2337	666268910803644416	NaN	NaN
2338	661104133288665088	NaN	NaN
2339	661102155909144576	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2343	666073100786774016	NaN	NaN
2344	666071193221509120	NaN	NaN
2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2351	666049248165822465	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN

```
2355   666020888022790149                    NaN                    NaN

                        timestamp  \
0      2017-08-01 16:23:56 +0000
1      2017-08-01 00:17:27 +0000
2      2017-07-31 00:18:03 +0000
3      2017-07-30 15:58:51 +0000
4      2017-07-29 16:00:24 +0000
5      2017-07-29 00:08:17 +0000
6      2017-07-28 16:27:12 +0000
7      2017-07-28 00:22:40 +0000
8      2017-07-27 16:25:51 +0000
9      2017-07-26 15:59:51 +0000
10     2017-07-26 00:31:25 +0000
11     2017-07-25 16:11:53 +0000
12     2017-07-25 01:55:32 +0000
13     2017-07-25 00:10:02 +0000
14     2017-07-24 17:02:04 +0000
15     2017-07-24 00:19:32 +0000
16     2017-07-23 00:22:39 +0000
17     2017-07-22 16:56:37 +0000
18     2017-07-22 00:23:06 +0000
19     2017-07-21 01:02:36 +0000
20     2017-07-20 16:49:33 +0000
21     2017-07-19 16:06:48 +0000
22     2017-07-19 03:39:09 +0000
23     2017-07-19 00:47:34 +0000
24     2017-07-18 16:08:03 +0000
25     2017-07-18 00:07:08 +0000
26     2017-07-17 16:17:36 +0000
27     2017-07-16 23:58:41 +0000
28     2017-07-16 20:14:00 +0000
29     2017-07-15 23:25:31 +0000
...                          ...
2326   2015-11-17 00:24:19 +0000
2327   2015-11-17 00:06:54 +0000
2328   2015-11-16 23:23:41 +0000
2329   2015-11-16 21:54:18 +0000
2330   2015-11-16 21:10:36 +0000
2331   2015-11-16 20:32:58 +0000
2332   2015-11-16 20:01:42 +0000
2333   2015-11-16 19:31:45 +0000
2334   2015-11-16 16:37:02 +0000
2335   2015-11-16 16:11:11 +0000
2336   2015-11-16 15:14:19 +0000
2337   2015-11-16 14:57:41 +0000
2338   2015-11-16 04:02:55 +0000
2339   2015-11-16 03:55:04 +0000
```

```
2340   2015-11-16 03:44:34 +0000
2341   2015-11-16 03:22:39 +0000
2342   2015-11-16 02:38:37 +0000
2343   2015-11-16 01:59:36 +0000
2344   2015-11-16 01:52:02 +0000
2345   2015-11-16 01:22:45 +0000
2346   2015-11-16 01:01:59 +0000
2347   2015-11-16 00:55:59 +0000
2348   2015-11-16 00:49:46 +0000
2349   2015-11-16 00:35:11 +0000
2350   2015-11-16 00:30:50 +0000
2351   2015-11-16 00:24:50 +0000
2352   2015-11-16 00:04:52 +0000
2353   2015-11-15 23:21:54 +0000
2354   2015-11-15 23:05:30 +0000
2355   2015-11-15 22:32:08 +0000

                                                source  \
0      <a href="http://twitter.com/download/iphone" r...
1      <a href="http://twitter.com/download/iphone" r...
2      <a href="http://twitter.com/download/iphone" r...
3      <a href="http://twitter.com/download/iphone" r...
4      <a href="http://twitter.com/download/iphone" r...
5      <a href="http://twitter.com/download/iphone" r...
6      <a href="http://twitter.com/download/iphone" r...
7      <a href="http://twitter.com/download/iphone" r...
8      <a href="http://twitter.com/download/iphone" r...
9      <a href="http://twitter.com/download/iphone" r...
10     <a href="http://twitter.com/download/iphone" r...
11     <a href="http://twitter.com/download/iphone" r...
12     <a href="http://twitter.com/download/iphone" r...
13     <a href="http://twitter.com/download/iphone" r...
14     <a href="http://twitter.com/download/iphone" r...
15     <a href="http://twitter.com/download/iphone" r...
16     <a href="http://twitter.com/download/iphone" r...
17     <a href="http://twitter.com/download/iphone" r...
18     <a href="http://twitter.com/download/iphone" r...
19     <a href="http://twitter.com/download/iphone" r...
20     <a href="http://twitter.com/download/iphone" r...
21     <a href="http://twitter.com/download/iphone" r...
22     <a href="http://twitter.com/download/iphone" r...
23     <a href="http://twitter.com/download/iphone" r...
24     <a href="http://twitter.com/download/iphone" r...
25     <a href="http://twitter.com/download/iphone" r...
26     <a href="http://twitter.com/download/iphone" r...
27     <a href="http://twitter.com/download/iphone" r...
28     <a href="http://twitter.com/download/iphone" r...
29     <a href="http://twitter.com/download/iphone" r...
```

```
...                                              ...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2345  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
2348  <a href="http://twitter.com/download/iphone" r...
2349  <a href="http://twitter.com/download/iphone" r...
2350  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...

                                                   text  retweeted_status_id  \
0     This is Phineas. He's a mystical boy. Only eve...                  NaN
1     This is Tilly. She's just checking pup on you...                   NaN
2     This is Archie. He is a rare Norwegian Pouncin...                  NaN
3     This is Darla. She commenced a snooze mid meal...                  NaN
4     This is Franklin. He would like you to stop ca...                  NaN
5     Here we have a majestic great white breaching ...                  NaN
6     Meet Jax. He enjoys ice cream so much he gets ...                  NaN
7     When you watch your owner call another dog a g...                   NaN
8     This is Zoey. She doesn't want to be one of th...                  NaN
9     This is Cassie. She is a college pup. Studying...                  NaN
10    This is Koda. He is a South Australian decksha...                  NaN
11    This is Bruno. He is a service shark. Only get...                  NaN
12    Here's a puppo that seems to be on the fence a...                  NaN
13    This is Ted. He does his best. Sometimes that'...                  NaN
14    This is Stuart. He's sporting his favorite fan...                  NaN
```

```
15     This is Oliver. You're witnessing one of his m...              NaN
16     This is Jim. He found a fren. Taught him how t...              NaN
17     This is Zeke. He has a new stick. Very proud o...              NaN
18     This is Ralphus. He's powering up. Attempting ...              NaN
19     RT @dog_rates: This is Canela. She attempted s...     8.874740e+17
20     This is Gerald. He was just told he didn't get...              NaN
21     This is Jeffrey. He has a monopoly on the pool...              NaN
22     I've yet to rate a Venezuelan Hover Wiener. Th...              NaN
23     This is Canela. She attempted some fancy porch...              NaN
24     You may not have known you needed to see this ...              NaN
25     This... is a Jubilant Antarctic House Bear. We...              NaN
26     This is Maya. She's very shy. Rarely leaves he...              NaN
27     This is Mingus. He's a wonderful father to his...              NaN
28     This is Derek. He's late for a dog meeting. 13...              NaN
29     This is Roscoe. Another pupper fallen victim t...              NaN
...                                                  ...              ...
2326   This is quite the dog. Gets really excited whe...              NaN
2327   This is a southern Vesuvius bumblegruff. Can d...              NaN
2328   Oh goodness. A super rare northeast Qdoba kang...              NaN
2329   Those are sunglasses and a jean jacket. 11/10 ...              NaN
2330   Unique dog here. Very small. Lives in containe...              NaN
2331   Here we have a mixed Asiago from the Galápagos...              NaN
2332   Look at this jokester thinking seat belt laws ...              NaN
2333   This is an extremely rare horned Parthenon. No...              NaN
2334   This is a funny dog. Weird toes. Won't come do...              NaN
2335   This is an Albanian 3 1/2 legged  Episcopalian...              NaN
2336       Can take selfies 11/10 https://t.co/ws2AMaNwPW              NaN
2337   Very concerned about fellow dog trapped in com...              NaN
2338   Not familiar with this breed. No tail (weird)...              NaN
2339   Oh my. Here you are seeing an Adobe Setter giv...              NaN
2340   Can stand on stump for what seems like a while...              NaN
2341   This appears to be a Mongolian Presbyterian mi...              NaN
2342   Here we have a well-established sunblockerspan...              NaN
2343   Let's hope this flight isn't Malaysian (lol). ...              NaN
2344   Here we have a northern speckled Rhododendron...              NaN
2345   This is the happiest dog you will ever see. Ve...              NaN
2346   Here is the Rand Paul of retrievers folks! He'...              NaN
2347   My oh my. This is a rare blond Canadian terrie...              NaN
2348   Here is a Siberian heavily armored polar bear ...              NaN
2349   This is an odd dog. Hard on the outside but lo...              NaN
2350   This is a truly beautiful English Wilson Staff...              NaN
2351   Here we have a 1949 1st generation vulpix. Enj...              NaN
2352   This is a purebred Piers Morgan. Loves to Netf...              NaN
2353   Here is a very happy pup. Big fan of well-main...              NaN
2354   This is a western brown Mitsubishi terrier. Up...              NaN
2355   Here we have a Japanese Irish Setter. Lost eye...              NaN

       retweeted_status_user_id retweeted_status_timestamp  \
```

```
0                      NaN                               NaN
1                      NaN                               NaN
2                      NaN                               NaN
3                      NaN                               NaN
4                      NaN                               NaN
5                      NaN                               NaN
6                      NaN                               NaN
7                      NaN                               NaN
8                      NaN                               NaN
9                      NaN                               NaN
10                     NaN                               NaN
11                     NaN                               NaN
12                     NaN                               NaN
13                     NaN                               NaN
14                     NaN                               NaN
15                     NaN                               NaN
16                     NaN                               NaN
17                     NaN                               NaN
18                     NaN                               NaN
19            4.196984e+09   2017-07-19 00:47:34 +0000
20                     NaN                               NaN
21                     NaN                               NaN
22                     NaN                               NaN
23                     NaN                               NaN
24                     NaN                               NaN
25                     NaN                               NaN
26                     NaN                               NaN
27                     NaN                               NaN
28                     NaN                               NaN
29                     NaN                               NaN
...                    ...                               ...
2326                   NaN                               NaN
2327                   NaN                               NaN
2328                   NaN                               NaN
2329                   NaN                               NaN
2330                   NaN                               NaN
2331                   NaN                               NaN
2332                   NaN                               NaN
2333                   NaN                               NaN
2334                   NaN                               NaN
2335                   NaN                               NaN
2336                   NaN                               NaN
2337                   NaN                               NaN
2338                   NaN                               NaN
2339                   NaN                               NaN
2340                   NaN                               NaN
2341                   NaN                               NaN
2342                   NaN                               NaN
```

```
2343                        NaN                        NaN
2344                        NaN                        NaN
2345                        NaN                        NaN
2346                        NaN                        NaN
2347                        NaN                        NaN
2348                        NaN                        NaN
2349                        NaN                        NaN
2350                        NaN                        NaN
2351                        NaN                        NaN
2352                        NaN                        NaN
2353                        NaN                        NaN
2354                        NaN                        NaN
2355                        NaN                        NaN

                                        expanded_urls  rating_numerator  \
0      https://twitter.com/dog_rates/status/892420643...                13
1      https://twitter.com/dog_rates/status/892177421...                13
2      https://twitter.com/dog_rates/status/891815181...                12
3      https://twitter.com/dog_rates/status/891689557...                13
4      https://twitter.com/dog_rates/status/891327558...                12
5      https://twitter.com/dog_rates/status/891087950...                13
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...                13
7      https://twitter.com/dog_rates/status/890729181...                13
8      https://twitter.com/dog_rates/status/890609185...                13
9      https://twitter.com/dog_rates/status/890240255...                14
10     https://twitter.com/dog_rates/status/890006608...                13
11     https://twitter.com/dog_rates/status/889880896...                13
12     https://twitter.com/dog_rates/status/889665388...                13
13     https://twitter.com/dog_rates/status/889638837...                12
14     https://twitter.com/dog_rates/status/889531135...                13
15     https://twitter.com/dog_rates/status/889278841...                13
16     https://twitter.com/dog_rates/status/888917238...                12
17     https://twitter.com/dog_rates/status/888804989...                13
18     https://twitter.com/dog_rates/status/888554962...                13
19     https://twitter.com/dog_rates/status/887473957...                13
20     https://twitter.com/dog_rates/status/888078434...                12
21     https://twitter.com/dog_rates/status/887705289...                13
22     https://twitter.com/dog_rates/status/887517139...                14
23     https://twitter.com/dog_rates/status/887473957...                13
24     https://twitter.com/dog_rates/status/887343217...                13
25     https://twitter.com/dog_rates/status/887101392...                12
26     https://twitter.com/dog_rates/status/886983233...                13
27     https://www.gofundme.com/mingusneedsus,https:/...                13
28     https://twitter.com/dog_rates/status/886680336...                13
29     https://twitter.com/dog_rates/status/886366144...                12
...                                                  ...               ...
2326   https://twitter.com/dog_rates/status/666411507...                 2
2327   https://twitter.com/dog_rates/status/666407126...                 7
```

9

```
2328  https://twitter.com/dog_rates/status/666396247...                          9
2329  https://twitter.com/dog_rates/status/666373753...                         11
2330  https://twitter.com/dog_rates/status/666362758...                          6
2331  https://twitter.com/dog_rates/status/666353288...                          8
2332  https://twitter.com/dog_rates/status/666345417...                         10
2333  https://twitter.com/dog_rates/status/666337882...                          9
2334  https://twitter.com/dog_rates/status/666293911...                          3
2335  https://twitter.com/dog_rates/status/666287406...                          1
2336  https://twitter.com/dog_rates/status/666273097...                         11
2337  https://twitter.com/dog_rates/status/666268910...                         10
2338  https://twitter.com/dog_rates/status/666104133...                          1
2339  https://twitter.com/dog_rates/status/666102155...                         11
2340  https://twitter.com/dog_rates/status/666099513...                          8
2341  https://twitter.com/dog_rates/status/666094000...                          9
2342  https://twitter.com/dog_rates/status/666082916...                          6
2343  https://twitter.com/dog_rates/status/666073100...                         10
2344  https://twitter.com/dog_rates/status/666071193...                          9
2345  https://twitter.com/dog_rates/status/666063827...                         10
2346  https://twitter.com/dog_rates/status/666058600...                          8
2347  https://twitter.com/dog_rates/status/666057090...                          9
2348  https://twitter.com/dog_rates/status/666055525...                         10
2349  https://twitter.com/dog_rates/status/666051853...                          2
2350  https://twitter.com/dog_rates/status/666050758...                         10
2351  https://twitter.com/dog_rates/status/666049248...                          5
2352  https://twitter.com/dog_rates/status/666044226...                          6
2353  https://twitter.com/dog_rates/status/666033412...                          9
2354  https://twitter.com/dog_rates/status/666029285...                          7
2355  https://twitter.com/dog_rates/status/666020888...                          8

      rating_denominator       name  doggo floofer  pupper  puppo
0                     10   Phineas   None    None    None   None
1                     10     Tilly   None    None    None   None
2                     10    Archie   None    None    None   None
3                     10     Darla   None    None    None   None
4                     10  Franklin   None    None    None   None
5                     10      None   None    None    None   None
6                     10       Jax   None    None    None   None
7                     10      None   None    None    None   None
8                     10      Zoey   None    None    None   None
9                     10    Cassie  doggo    None    None   None
10                    10      Koda   None    None    None   None
11                    10     Bruno   None    None    None   None
12                    10      None   None    None    None  puppo
13                    10       Ted   None    None    None   None
14                    10    Stuart   None    None    None  puppo
15                    10    Oliver   None    None    None   None
16                    10       Jim   None    None    None   None
17                    10      Zeke   None    None    None   None
```

```
18              10    Ralphus    None    None    None    None
19              10    Canela     None    None    None    None
20              10    Gerald     None    None    None    None
21              10    Jeffrey    None    None    None    None
22              10      such     None    None    None    None
23              10    Canela     None    None    None    None
24              10      None     None    None    None    None
25              10      None     None    None    None    None
26              10      Maya     None    None    None    None
27              10    Mingus     None    None    None    None
28              10     Derek     None    None    None    None
29              10    Roscoe     None    None  pupper    None

...            ...      ...      ...     ...     ...     ...
2326            10     quite     None    None    None    None
2327            10        a      None    None    None    None
2328            10      None     None    None    None    None
2329            10      None     None    None    None    None
2330            10      None     None    None    None    None
2331            10      None     None    None    None    None
2332            10      None     None    None    None    None
2333            10        an     None    None    None    None
2334            10        a      None    None    None    None
2335             2        an     None    None    None    None
2336            10      None     None    None    None    None
2337            10      None     None    None    None    None
2338            10      None     None    None    None    None
2339            10      None     None    None    None    None
2340            10      None     None    None    None    None
2341            10      None     None    None    None    None
2342            10      None     None    None    None    None
2343            10      None     None    None    None    None
2344            10      None     None    None    None    None
2345            10       the     None    None    None    None
2346            10       the     None    None    None    None
2347            10        a      None    None    None    None
2348            10        a      None    None    None    None
2349            10        an     None    None    None    None
2350            10        a      None    None    None    None
2351            10      None     None    None    None    None
2352            10        a      None    None    None    None
2353            10        a      None    None    None    None
2354            10        a      None    None    None    None
2355            10      None     None    None    None    None

[2356 rows x 17 columns]
```

In [7]: # A concise summary of a DataFrame
        df_twitter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                    2356 non-null int64
in_reply_to_status_id       78 non-null float64
in_reply_to_user_id         78 non-null float64
timestamp                   2356 non-null object
source                      2356 non-null object
text                        2356 non-null object
retweeted_status_id         181 non-null float64
retweeted_status_user_id    181 non-null float64
retweeted_status_timestamp  181 non-null object
expanded_urls               2297 non-null object
rating_numerator            2356 non-null int64
rating_denominator          2356 non-null int64
name                        2356 non-null object
doggo                       2356 non-null object
floofer                     2356 non-null object
pupper                      2356 non-null object
puppo                       2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

In [8]: # Viewing data to determine appearance
        df_twitter.retweeted_status_timestamp.value_counts().head(3)

Out[8]: 2016-01-25 00:26:41 +0000   1
        2016-06-08 02:41:38 +0000   1
        2016-07-27 00:40:12 +0000   1
        Name: retweeted_status_timestamp, dtype: int64

In [9]: # Viewing data to determine appearance
        df_twitter.retweeted_status_id.value_counts().head(3)

Out[9]: 7.757333e+17   1
        7.507196e+17   1
        6.742918e+17   1
        Name: retweeted_status_id, dtype: int64

In [10]: # Descriptive statistics of a DataFrame
         df_twitter.describe()

Out[10]:   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
```

In [11]: df_twitter.source.value_counts()

Out[11]: Twitter for iPhone
         Vine - Make a Scene
         Twitter Web Client
         TweetDeck
         Name: source, dtype: int64

In [12]: df_twitter.rating_denominator.value_counts()

Out[12]: 10   2333
         11   3
         50   3
         80   2
         20   2
         2   1
         16   1
         40   1
         70   1
         15   1
         90   1
         110   1
         120   1
         130   1
         150   1
```

13

```
170        1
7          1
0          1
Name: rating_denominator, dtype: int64
```

In [13]: df_twitter.rating_numerator.value_counts()

Out[13]: 12   558
         11   464
         10   461
         13   351
         9   158
         8   102
         7   55
         14   54
         5   37
         6   32
         3   19
         4   17
         1   9
         2   9
         420   2
         0   2
         15   2
         75   2
         80   1
         20   1
         24   1
         26   1
         44   1
         50   1
         60   1
         165   1
         84   1
         88   1
         144   1
         182   1
         143   1
         666   1
         960   1
         1776   1
         17   1
         27   1
         45   1
         99   1
         121   1
         204   1
         Name: rating_numerator, dtype: int64

```
In [14]: df_img_pred.shape

Out[14]: (2075, 12)

In [15]: df_img_pred

Out[15]:               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
         5     666050758794694657      https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
         6     666051853826850816      https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
         7     666055525042405380      https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
         8     666057090499244032      https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
         9     666058600524156928      https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
         10    666063827256086533      https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
         11    666071193221509120      https://pbs.twimg.com/media/CT5cN_3WEAA1OoZ.jpg
         12    666073100786774016      https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
         13    666082916733198337      https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg
         14    666094000022159362      https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
         15    666099513787052032      https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
         16    666102155909144576      https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
         17    666104133288665088      https://pbs.twimg.com/media/CT56LSZWoAA1Jj2.jpg
         18    666268910803644416      https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
         19    666273097616637952      https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
         20    666287406224695296      https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
         21    666293911632134144      https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
         22    666337882303524864      https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
         23    666345417576210432      https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
         24    666353288456101888      https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
         25    666362758909284353      https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
         26    666373753744588802      https://pbs.twimg.com/media/CT9vZEYWUAA1Z05.jpg
         27    666396247373291520      https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
         28    666407126856765440      https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
         29    666411507551481857      https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
         ...                  ...                                                  ...
         2045  886366144734445568      https://pbs.twimg.com/media/DEOBTnQUwAApKEH.jpg
         2046  886680336477933568      https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
         2047  886736880519319552      https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
         2048  886983233522544640      https://pbs.twimg.com/media/DE8yicJWOAAAvBJ.jpg
         2049  887101392804085760      https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
         2050  887343217045368832   https://pbs.twimg.com/ext_tw_video_thumb/88734...
         2051  887473957103951883      https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
         2052  887517139158093824   https://pbs.twimg.com/ext_tw_video_thumb/88751...
         2053  887705289381826560      https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
```

```
2054  888078434458587136      https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  888202515573088257      https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  888554962724278272      https://pbs.twimg.com/media/DFTH_0-UQAACu20.jpg
2057  888804989199671297      https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296      https://pbs.twimg.com/media/DFYRgs0UQAARGh0.jpg
2059  889278841981685760      https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921      https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061  889638837579907072      https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689      https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881      https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480      https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849      https://pbs.twimg.com/media/DFrEyVuW0AA03t9.jpg
2066  890609185150312448      https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067  890729181411237888      https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068  890971913173991426      https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069  891087950875897856      https://pbs.twimg.com/media/DF3HwyEWsAABqE6.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  \
0          1    Welsh_springer_spaniel  0.465074     True
1          1                   redbone  0.506826     True
2          1           German_shepherd  0.596461     True
3          1        Rhodesian_ridgeback  0.408143     True
4          1         miniature_pinscher  0.560311     True
5          1       Bernese_mountain_dog  0.651137     True
6          1                box_turtle  0.933012    False
7          1                      chow  0.692517     True
8          1             shopping_cart  0.962465    False
9          1          miniature_poodle  0.201493     True
10         1          golden_retriever  0.775930     True
11         1             Gordon_setter  0.503672     True
12         1             Walker_hound   0.260857     True
13         1                      pug   0.489814     True
14         1                bloodhound  0.195217     True
15         1                    Lhasa   0.582330     True
16         1            English_setter  0.298617     True
17         1                      hen   0.965932    False
18         1          desktop_computer  0.086502    False
19         1         Italian_greyhound  0.176053     True
20         1               Maltese_dog  0.857531     True
21         1         three-toed_sloth   0.914671    False
22         1                      ox    0.416669    False
23         1          golden_retriever  0.858744     True
24         1                  malamute  0.336874     True
```

```
25        1                  guinea_pig  0.996496      False
26        1   soft-coated_wheaten_terrier  0.326467       True
27        1                    Chihuahua  0.978108       True
28        1       black-and-tan_coonhound  0.529139       True
29        1                         coho  0.404640      False
...     ...                          ...       ...        ...
2045      1                French_bulldog  0.999201       True
2046      1                   convertible  0.738995      False
2047      1                        kuvasz  0.309706       True
2048      2                     Chihuahua  0.793469       True
2049      1                       Samoyed  0.733942       True
2050      1              Mexican_hairless  0.330741       True
2051      2                      Pembroke  0.809197       True
2052      1                     limousine  0.130432      False
2053      1                        basset  0.821664       True
2054      1                French_bulldog  0.995026       True
2055      2                      Pembroke  0.809197       True
2056      3                 Siberian_husky  0.700377       True
2057      1               golden_retriever  0.469760       True
2058      1               golden_retriever  0.714719       True
2059      1                       whippet  0.626152       True
2060      1               golden_retriever  0.953442       True
2061      1                French_bulldog  0.991650       True
2062      1                      Pembroke  0.966327       True
2063      1                French_bulldog  0.377417       True
2064      1                       Samoyed  0.957979       True
2065      1                      Pembroke  0.511319       True
2066      1                  Irish_terrier  0.487574       True
2067      2                    Pomeranian  0.566142       True
2068      1                     Appenzeller  0.341703       True
2069      1        Chesapeake_Bay_retriever  0.425595       True
2070      2                        basset  0.555712       True
2071      1                    paper_towel  0.170278      False
2072      1                     Chihuahua  0.716012       True
2073      1                     Chihuahua  0.323581       True
2074      1                        orange  0.097049      False

                            p2   p2_conf  p2_dog                            p3  \
0                        collie  0.156665    True              Shetland_sheepdog
1            miniature_pinscher  0.074192    True             Rhodesian_ridgeback
2                       malinois  0.138584    True                      bloodhound
3                         redbone  0.360687    True              miniature_pinscher
4                     Rottweiler  0.243682    True                        Doberman
5               English_springer  0.263788    True     Greater_Swiss_Mountain_dog
6                      mud_turtle  0.045885   False                        terrapin
7               Tibetan_mastiff  0.058279    True                         fur_coat
8               shopping_basket  0.014594   False               golden_retriever
9                       komondor  0.192305    True    soft-coated_wheaten_terrier
```

10	Tibetan_mastiff	0.093718	True	Labrador_retriever
11	Yorkshire_terrier	0.174201	True	Pekinese
12	English_foxhound	0.175382	True	Ibizan_hound
13	bull_mastiff	0.404722	True	French_bulldog
14	German_shepherd	0.078260	True	malinois
15	Shih-Tzu	0.166192	True	Dandie_Dinmont
16	Newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy_terrier	0.111884	True	basenji
20	toy_poodle	0.063064	True	miniature_poodle
21	otter	0.015250	False	great_grey_owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
24	Siberian_husky	0.147655	True	Eskimo_dog
25	skunk	0.002402	False	hamster
26	Afghan_hound	0.259551	True	briard
27	toy_terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat-coated_retriever
29	barracouta	0.271485	False	gar
...	...	...	...	...
2045	Chihuahua	0.000361	True	Boston_bull
2046	sports_car	0.139952	False	car_wheel
2047	Great_Pyrenees	0.186136	True	Dandie_Dinmont
2048	toy_terrier	0.143528	True	can_opener
2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula

```
2072                    malamute  0.078253    True                    kelpie
2073                    Pekinese  0.090647    True                  papillon
2074                       bagel  0.085851    False                   banana


        p3_conf  p3_dog
0      0.061428    True
1      0.072010    True
2      0.116197    True
3      0.222752    True
4      0.154629    True
5      0.016199    True
6      0.017885   False
7      0.054449   False
8      0.007959    True
9      0.082086    True
10     0.072427    True
11     0.109454    True
12     0.097471    True
13     0.048960    True
14     0.075628    True
15     0.089688    True
16     0.133649    True
17     0.000052   False
18     0.079480   False
19     0.111152    True
20     0.025581    True
21     0.013207   False
22     0.102643    True
23     0.014241    True
24     0.093412    True
25     0.000461   False
26     0.206803    True
27     0.004577    True
28     0.173810    True
29     0.189945   False
...         ...     ...
2045   0.000076    True
2046   0.044173   False
2047   0.086346    True
2048   0.032253   False
2049   0.029705    True
2050   0.134203    True
2051   0.038915    True
2052   0.026321   False
2053   0.026236    True
2054   0.000903    True
2055   0.038915    True
2056   0.111411    True
```

```
2057  0.073482     True
2058  0.105506     True
2059  0.027351     True
2060  0.007958     True
2061  0.001498     True
2062  0.004633     True
2063  0.082981    False
2064  0.008167     True
2065  0.029248     True
2066  0.118184     True
2067  0.076507     True
2068  0.193548    False
2069  0.076902    False
2070  0.175219     True
2071  0.040836    False
2072  0.031379     True
2073  0.068957     True
2074  0.076110    False

[2075 rows x 12 columns]
```

In [16]: # A concise summary of a DataFrame
         df_img_pred.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

In [17]: # Descriptive statistics of a DataFrame
         df_img_pred.describe()

Out[17]:   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

20

```
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
```

In [18]: df_tweets.shape

Out[18]: (2354, 31)

In [19]: df_tweets

```
Out[19]:       contributors  coordinates           created_at display_text_range  \
        0               NaN          NaN  2017-08-01 16:23:56           [0, 85]
        1               NaN          NaN  2017-08-01 00:17:27          [0, 138]
        2               NaN          NaN  2017-07-31 00:18:03          [0, 121]
        3               NaN          NaN  2017-07-30 15:58:51           [0, 79]
        4               NaN          NaN  2017-07-29 16:00:24          [0, 138]
        5               NaN          NaN  2017-07-29 00:08:17          [0, 138]
        6               NaN          NaN  2017-07-28 16:27:12          [0, 140]
        7               NaN          NaN  2017-07-28 00:22:40          [0, 118]
        8               NaN          NaN  2017-07-27 16:25:51          [0, 122]
        9               NaN          NaN  2017-07-26 15:59:51          [0, 133]
        10              NaN          NaN  2017-07-26 00:31:25          [0, 130]
        11              NaN          NaN  2017-07-25 16:11:53          [0, 107]
        12              NaN          NaN  2017-07-25 01:55:32          [0, 106]
        13              NaN          NaN  2017-07-25 00:10:02           [0, 91]
        14              NaN          NaN  2017-07-24 17:02:04          [0, 118]
        15              NaN          NaN  2017-07-24 00:19:32          [0, 138]
        16              NaN          NaN  2017-07-23 00:22:39           [0, 86]
        17              NaN          NaN  2017-07-22 16:56:37          [0, 128]
        18              NaN          NaN  2017-07-22 00:23:06           [0, 87]
        19              NaN          NaN  2017-07-20 16:49:33          [0, 127]
        20              NaN          NaN  2017-07-19 16:06:48          [0, 127]
        21              NaN          NaN  2017-07-19 03:39:09          [0, 108]
        22              NaN          NaN  2017-07-19 00:47:34           [0, 99]
        23              NaN          NaN  2017-07-18 16:08:03           [0, 88]
        24              NaN          NaN  2017-07-18 00:07:08          [0, 129]
        25              NaN          NaN  2017-07-17 16:17:36          [0, 101]
        26              NaN          NaN  2017-07-16 23:58:41          [0, 121]
        27              NaN          NaN  2017-07-16 20:14:00           [0, 71]
        28              NaN          NaN  2017-07-15 23:25:31          [0, 131]
        29              NaN          NaN  2017-07-15 16:51:35         [27, 105]
        ...             ...          ...                  ...               ...
        2324            NaN          NaN  2015-11-17 00:24:19          [0, 140]
        2325            NaN          NaN  2015-11-17 00:06:54          [0, 139]
```

```
2326       NaN       NaN 2015-11-16 23:23:41            [0, 137]
2327       NaN       NaN 2015-11-16 21:54:18             [0, 81]
2328       NaN       NaN 2015-11-16 21:10:36            [0, 140]
2329       NaN       NaN 2015-11-16 20:32:58            [0, 135]
2330       NaN       NaN 2015-11-16 20:01:42            [0, 112]
2331       NaN       NaN 2015-11-16 19:31:45            [0, 139]
2332       NaN       NaN 2015-11-16 16:37:02            [0, 138]
2333       NaN       NaN 2015-11-16 16:11:11            [0, 136]
2334       NaN       NaN 2015-11-16 15:14:19             [0, 46]
2335       NaN       NaN 2015-11-16 14:57:41             [0, 82]
2336       NaN       NaN 2015-11-16 04:02:55            [0, 134]
2337       NaN       NaN 2015-11-16 03:55:04            [0, 128]
2338       NaN       NaN 2015-11-16 03:44:34            [0, 140]
2339       NaN       NaN 2015-11-16 03:22:39            [0, 132]
2340       NaN       NaN 2015-11-16 02:38:37            [0, 125]
2341       NaN       NaN 2015-11-16 01:59:36            [0, 137]
2342       NaN       NaN 2015-11-16 01:52:02            [0, 137]
2343       NaN       NaN 2015-11-16 01:22:45            [0, 107]
2344       NaN       NaN 2015-11-16 01:01:59            [0, 135]
2345       NaN       NaN 2015-11-16 00:55:59            [0, 124]
2346       NaN       NaN 2015-11-16 00:49:46            [0, 140]
2347       NaN       NaN 2015-11-16 00:35:11            [0, 138]
2348       NaN       NaN 2015-11-16 00:30:50            [0, 140]
2349       NaN       NaN 2015-11-16 00:24:50            [0, 120]
2350       NaN       NaN 2015-11-16 00:04:52            [0, 137]
2351       NaN       NaN 2015-11-15 23:21:54            [0, 130]
2352       NaN       NaN 2015-11-15 23:05:30            [0, 139]
2353       NaN       NaN 2015-11-15 22:32:08            [0, 131]

                                                entities  \
0      {'hashtags': [], 'symbols': [], 'user_mentions...
1      {'hashtags': [], 'symbols': [], 'user_mentions...
2      {'hashtags': [], 'symbols': [], 'user_mentions...
3      {'hashtags': [], 'symbols': [], 'user_mentions...
4      {'hashtags': [{'text': 'BarkWeek', 'indices': ...
5      {'hashtags': [{'text': 'BarkWeek', 'indices': ...
6      {'hashtags': [], 'symbols': [], 'user_mentions...
7      {'hashtags': [], 'symbols': [], 'user_mentions...
8      {'hashtags': [{'text': 'BarkWeek', 'indices': ...
9      {'hashtags': [], 'symbols': [], 'user_mentions...
10     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
11     {'hashtags': [], 'symbols': [], 'user_mentions...
12     {'hashtags': [], 'symbols': [], 'user_mentions...
13     {'hashtags': [], 'symbols': [], 'user_mentions...
14     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
15     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
16     {'hashtags': [], 'symbols': [], 'user_mentions...
17     {'hashtags': [], 'symbols': [], 'user_mentions...
```

```
18    {'hashtags': [], 'symbols': [], 'user_mentions...
19    {'hashtags': [], 'symbols': [], 'user_mentions...
20    {'hashtags': [], 'symbols': [], 'user_mentions...
21    {'hashtags': [], 'symbols': [], 'user_mentions...
22    {'hashtags': [], 'symbols': [], 'user_mentions...
23    {'hashtags': [], 'symbols': [], 'user_mentions...
24    {'hashtags': [], 'symbols': [], 'user_mentions...
25    {'hashtags': [], 'symbols': [], 'user_mentions...
26    {'hashtags': [], 'symbols': [], 'user_mentions...
27    {'hashtags': [], 'symbols': [], 'user_mentions...
28    {'hashtags': [], 'symbols': [], 'user_mentions...
29    {'hashtags': [], 'symbols': [], 'user_mentions...
...                                                  ...
2324  {'hashtags': [], 'symbols': [], 'user_mentions...
2325  {'hashtags': [], 'symbols': [], 'user_mentions...
2326  {'hashtags': [], 'symbols': [], 'user_mentions...
2327  {'hashtags': [], 'symbols': [], 'user_mentions...
2328  {'hashtags': [], 'symbols': [], 'user_mentions...
2329  {'hashtags': [], 'symbols': [], 'user_mentions...
2330  {'hashtags': [], 'symbols': [], 'user_mentions...
2331  {'hashtags': [], 'symbols': [], 'user_mentions...
2332  {'hashtags': [], 'symbols': [], 'user_mentions...
2333  {'hashtags': [], 'symbols': [], 'user_mentions...
2334  {'hashtags': [], 'symbols': [], 'user_mentions...
2335  {'hashtags': [], 'symbols': [], 'user_mentions...
2336  {'hashtags': [], 'symbols': [], 'user_mentions...
2337  {'hashtags': [], 'symbols': [], 'user_mentions...
2338  {'hashtags': [], 'symbols': [], 'user_mentions...
2339  {'hashtags': [], 'symbols': [], 'user_mentions...
2340  {'hashtags': [], 'symbols': [], 'user_mentions...
2341  {'hashtags': [], 'symbols': [], 'user_mentions...
2342  {'hashtags': [], 'symbols': [], 'user_mentions...
2343  {'hashtags': [], 'symbols': [], 'user_mentions...
2344  {'hashtags': [], 'symbols': [], 'user_mentions...
2345  {'hashtags': [], 'symbols': [], 'user_mentions...
2346  {'hashtags': [], 'symbols': [], 'user_mentions...
2347  {'hashtags': [], 'symbols': [], 'user_mentions...
2348  {'hashtags': [], 'symbols': [], 'user_mentions...
2349  {'hashtags': [], 'symbols': [], 'user_mentions...
2350  {'hashtags': [], 'symbols': [], 'user_mentions...
2351  {'hashtags': [], 'symbols': [], 'user_mentions...
2352  {'hashtags': [], 'symbols': [], 'user_mentions...
2353  {'hashtags': [], 'symbols': [], 'user_mentions...

                                    extended_entities  favorite_count  \
0     {'media': [{'id': 892420639486877696, 'id_str'...          39467
1     {'media': [{'id': 892177413194625024, 'id_str'...          33819
2     {'media': [{'id': 891815175371796480, 'id_str'...          25461
```

```
3      {'media': [{'id': 891689552724799489, 'id_str'...        42908
4      {'media': [{'id': 891327551943041024, 'id_str'...        41048
5      {'media': [{'id': 891087942176911360, 'id_str'...        20562
6      {'media': [{'id': 890971906207338496, 'id_str'...        12041
7      {'media': [{'id': 890729118844600320, 'id_str'...        56848
8      {'media': [{'id': 890609177319665665, 'id_str'...        28226
9      {'media': [{'id': 890240245463175168, 'id_str'...        32467
10     {'media': [{'id': 890006600089468928, 'id_str'...        31166
11     {'media': [{'id': 889880888800096258, 'id_str'...        28268
12     {'media': [{'id': 889665366129029120, 'id_str'...        38818
13     {'media': [{'id': 889638825424826374, 'id_str'...        27672
14     {'media': [{'id': 889531127467266049, 'id_str'...        15359
15     {'media': [{'id': 889278779352338437, 'id_str'...        25652
16     {'media': [{'id': 888917229776945152, 'id_str'...        29611
17     {'media': [{'id': 888804981515575296, 'id_str'...        26080
18     {'media': [{'id': 888554915546542081, 'id_str'...        20290
19     {'media': [{'id': 888078426338406400, 'id_str'...        22201
20     {'media': [{'id': 887705281597243393, 'id_str'...        30779
21     {'media': [{'id': 887517108413886465, 'id_str'...        46959
22     {'media': [{'id': 887473949361045505, 'id_str'...        69871
23     {'media': [{'id': 887343120832229379, 'id_str'...        34222
24     {'media': [{'id': 887101385971384320, 'id_str'...        31061
25     {'media': [{'id': 886983218871902208, 'id_str'...        35859
26     {'media': [{'id': 886736868116754432, 'id_str'...        12306
27     {'media': [{'id': 886680331239161856, 'id_str'...        22798
28     {'media': [{'id': 886366138128449536, 'id_str'...        21524
29                                                   NaN          117
...                                                   ...        ...
2324   {'media': [{'id': 666411498068123649, 'id_str'...          459
2325   {'media': [{'id': 666407121513275392, 'id_str'...          113
2326   {'media': [{'id': 666396240351993856, 'id_str'...          172
2327   {'media': [{'id': 666373746337402880, 'id_str'...          194
2328   {'media': [{'id': 666362717482020864, 'id_str'...          804
2329   {'media': [{'id': 666353280906170368, 'id_str'...          229
2330   {'media': [{'id': 666345414279471104, 'id_str'...          307
2331   {'media': [{'id': 666337857791987715, 'id_str'...          204
2332   {'media': [{'id': 666293909010702337, 'id_str'...          522
2333   {'media': [{'id': 666287399580733440, 'id_str'...          152
2334   {'media': [{'id': 666273081518768128, 'id_str'...          184
2335   {'media': [{'id': 666268904428277760, 'id_str'...          108
2336   {'media': [{'id': 666104129232740352, 'id_str'...        14765
2337   {'media': [{'id': 666102150364286977, 'id_str'...           81
2338   {'media': [{'id': 666099505364733952, 'id_str'...          164
2339   {'media': [{'id': 666093996847063040, 'id_str'...          169
2340   {'media': [{'id': 666082912819875840, 'id_str'...          121
2341   {'media': [{'id': 666073098362486784, 'id_str'...          335
2342   {'media': [{'id': 666071190449033216, 'id_str'...          154
2343   {'media': [{'id': 666063820255862784, 'id_str'...          496
```

```
2344  {'media': [{'id': 666058597072306176, 'id_str'...        115
2345  {'media': [{'id': 666057085227016192, 'id_str'...        304
2346  {'media': [{'id': 666055517517848576, 'id_str'...        448
2347  {'media': [{'id': 666051848592334848, 'id_str'...       1253
2348  {'media': [{'id': 666050754986266625, 'id_str'...        136
2349  {'media': [{'id': 666049244999131136, 'id_str'...        111
2350  {'media': [{'id': 666044217047650304, 'id_str'...        311
2351  {'media': [{'id': 666033409081393153, 'id_str'...        128
2352  {'media': [{'id': 666029276303482880, 'id_str'...        132
2353  {'media': [{'id': 666020881337073664, 'id_str'...       2535

      favorited                                          full_text  geo  \
0         False  This is Phineas. He's a mystical boy. Only eve...  NaN
1         False  This is Tilly. She's just checking pup on you...  NaN
2         False  This is Archie. He is a rare Norwegian Pouncin...  NaN
3         False  This is Darla. She commenced a snooze mid meal...  NaN
4         False  This is Franklin. He would like you to stop ca...  NaN
5         False  Here we have a majestic great white breaching ...  NaN
6         False  Meet Jax. He enjoys ice cream so much he gets ...  NaN
7         False  When you watch your owner call another dog a g...  NaN
8         False  This is Zoey. She doesn't want to be one of th...  NaN
9         False  This is Cassie. She is a college pup. Studying...  NaN
10        False  This is Koda. He is a South Australian decksha...  NaN
11        False  This is Bruno. He is a service shark. Only get...  NaN
12        False  Here's a puppo that seems to be on the fence a...  NaN
13        False  This is Ted. He does his best. Sometimes that'...  NaN
14        False  This is Stuart. He's sporting his favorite fan...  NaN
15        False  This is Oliver. You're witnessing one of his m...  NaN
16        False  This is Jim. He found a fren. Taught him how t...  NaN
17        False  This is Zeke. He has a new stick. Very proud o...  NaN
18        False  This is Ralphus. He's powering up. Attempting ...  NaN
19        False  This is Gerald. He was just told he didn't get...  NaN
20        False  This is Jeffrey. He has a monopoly on the pool...  NaN
21         True  I've yet to rate a Venezuelan Hover Wiener. Th...  NaN
22        False  This is Canela. She attempted some fancy porch...  NaN
23        False  You may not have known you needed to see this ...  NaN
24        False  This... is a Jubilant Antarctic House Bear. We...  NaN
25        False  This is Maya. She's very shy. Rarely leaves he...  NaN
26        False  This is Mingus. He's a wonderful father to his...  NaN
27        False  This is Derek. He's late for a dog meeting. 13...  NaN
28        False  This is Roscoe. Another pupper fallen victim t...  NaN
29        False  @NonWhiteHat @MayhewMayhem omg hello tanner yo...  NaN
...         ...                                                ...  ...
2324      False  This is quite the dog. Gets really excited whe...  NaN
2325      False  This is a southern Vesuvius bumblegruff. Can d...  NaN
2326      False  Oh goodness. A super rare northeast Qdoba kang...  NaN
2327      False  Those are sunglasses and a jean jacket. 11/10 ...  NaN
2328      False  Unique dog here. Very small. Lives in containe...  NaN
```

```
2329    False  Here we have a mixed Asiago from the Galápagos...    NaN
2330    False  Look at this jokester thinking seat belt laws ...    NaN
2331    False  This is an extremely rare horned Parthenon. No...    NaN
2332    False  This is a funny dog. Weird toes. Won't come do...    NaN
2333    False  This is an Albanian 3 1/2 legged  Episcopalian...    NaN
2334    False      Can take selfies 11/10 https://t.co/ws2AMaNwPW    NaN
2335    False  Very concerned about fellow dog trapped in com...    NaN
2336    False  Not familiar with this breed. No tail (weird)...    NaN
2337    False  Oh my. Here you are seeing an Adobe Setter giv...    NaN
2338    False  Can stand on stump for what seems like a while...    NaN
2339    False  This appears to be a Mongolian Presbyterian mi...    NaN
2340    False  Here we have a well-established sunblockerspan...    NaN
2341    False  Let's hope this flight isn't Malaysian (lol). ...    NaN
2342    False  Here we have a northern speckled Rhododendron...    NaN
2343    False  This is the happiest dog you will ever see. Ve...    NaN
2344    False  Here is the Rand Paul of retrievers folks! He'...    NaN
2345    False  My oh my. This is a rare blond Canadian terrie...    NaN
2346    False  Here is a Siberian heavily armored polar bear ...    NaN
2347    False  This is an odd dog. Hard on the outside but lo...    NaN
2348    False  This is a truly beautiful English Wilson Staff...    NaN
2349    False  Here we have a 1949 1st generation vulpix. Enj...    NaN
2350    False  This is a purebred Piers Morgan. Loves to Netf...    NaN
2351    False  Here is a very happy pup. Big fan of well-main...    NaN
2352    False  This is a western brown Mitsubishi terrier. Up...    NaN
2353    False  Here we have a Japanese Irish Setter. Lost eye...    NaN

                                   ...                         \
0                                  ...
1                                  ...
2                                  ...
3                                  ...
4                                  ...
5                                  ...
6                                  ...
7                                  ...
8                                  ...
9                                  ...
10                                 ...
11                                 ...
12                                 ...
13                                 ...
14                                 ...
15                                 ...
16                                 ...
17                                 ...
18                                 ...
19                                 ...
20                                 ...
```

```
21                         ...
22                         ...
23                         ...
24                         ...
25                         ...
26                         ...
27                         ...
28                         ...
29                         ...
...                        ...
2324                       ...
2325                       ...
2326                       ...
2327                       ...
2328                       ...
2329                       ...
2330                       ...
2331                       ...
2332                       ...
2333                       ...
2334                       ...
2335                       ...
2336                       ...
2337                       ...
2338                       ...
2339                       ...
2340                       ...
2341                       ...
2342                       ...
2343                       ...
2344                       ...
2345                       ...
2346                       ...
2347                       ...
2348                       ...
2349                       ...
2350                       ...
2351                       ...
2352                       ...
2353                       ...


      possibly_sensitive_appealable  quoted_status quoted_status_id  \
0                               0.0            NaN              NaN
1                               0.0            NaN              NaN
2                               0.0            NaN              NaN
3                               0.0            NaN              NaN
4                               0.0            NaN              NaN
5                               0.0            NaN              NaN
```

6	0.0	NaN	NaN
7	0.0	NaN	NaN
8	0.0	NaN	NaN
9	0.0	NaN	NaN
10	0.0	NaN	NaN
11	0.0	NaN	NaN
12	0.0	NaN	NaN
13	0.0	NaN	NaN
14	0.0	NaN	NaN
15	0.0	NaN	NaN
16	0.0	NaN	NaN
17	0.0	NaN	NaN
18	0.0	NaN	NaN
19	0.0	NaN	NaN
20	0.0	NaN	NaN
21	0.0	NaN	NaN
22	0.0	NaN	NaN
23	0.0	NaN	NaN
24	0.0	NaN	NaN
25	0.0	NaN	NaN
26	0.0	NaN	NaN
27	0.0	NaN	NaN
28	0.0	NaN	NaN
29	NaN	NaN	NaN
...	...	...	...
2324	0.0	NaN	NaN
2325	0.0	NaN	NaN
2326	0.0	NaN	NaN
2327	0.0	NaN	NaN
2328	0.0	NaN	NaN
2329	0.0	NaN	NaN
2330	0.0	NaN	NaN
2331	0.0	NaN	NaN
2332	0.0	NaN	NaN
2333	0.0	NaN	NaN
2334	0.0	NaN	NaN
2335	0.0	NaN	NaN
2336	0.0	NaN	NaN
2337	0.0	NaN	NaN
2338	0.0	NaN	NaN
2339	0.0	NaN	NaN
2340	0.0	NaN	NaN
2341	0.0	NaN	NaN
2342	0.0	NaN	NaN
2343	0.0	NaN	NaN
2344	0.0	NaN	NaN
2345	0.0	NaN	NaN
2346	0.0	NaN	NaN

```
2347                                    0.0           NaN              NaN
2348                                    0.0           NaN              NaN
2349                                    0.0           NaN              NaN
2350                                    0.0           NaN              NaN
2351                                    0.0           NaN              NaN
2352                                    0.0           NaN              NaN
2353                                    0.0           NaN              NaN

      quoted_status_id_str  retweet_count  retweeted  retweeted_status  \
0                      NaN           8853      False               NaN
1                      NaN           6514      False               NaN
2                      NaN           4328      False               NaN
3                      NaN           8964      False               NaN
4                      NaN           9774      False               NaN
5                      NaN           3261      False               NaN
6                      NaN           2158      False               NaN
7                      NaN          16716      False               NaN
8                      NaN           4429      False               NaN
9                      NaN           7711      False               NaN
10                     NaN           7624      False               NaN
11                     NaN           5156      False               NaN
12                     NaN           8538      False               NaN
13                     NaN           4735      False               NaN
14                     NaN           2321      False               NaN
15                     NaN           5637      False               NaN
16                     NaN           4709      False               NaN
17                     NaN           4559      False               NaN
18                     NaN           3732      False               NaN
19                     NaN           3653      False               NaN
20                     NaN           5609      False               NaN
21                     NaN          12082      False               NaN
22                     NaN          18781      False               NaN
23                     NaN          10737      False               NaN
24                     NaN           6167      False               NaN
25                     NaN           8084      False               NaN
26                     NaN           3443      False               NaN
27                     NaN           4610      False               NaN
28                     NaN           3316      False               NaN
29                     NaN              4      False               NaN
...                    ...            ...        ...               ...
2324                   NaN            339      False               NaN
2325                   NaN             44      False               NaN
2326                   NaN             92      False               NaN
2327                   NaN            100      False               NaN
2328                   NaN            595      False               NaN
2329                   NaN             77      False               NaN
2330                   NaN            146      False               NaN
2331                   NaN             96      False               NaN
```

29

```
2332                    NaN      368       False              NaN
2333                    NaN       71       False              NaN
2334                    NaN       82       False              NaN
2335                    NaN       37       False              NaN
2336                    NaN     6871       False              NaN
2337                    NaN       16       False              NaN
2338                    NaN       73       False              NaN
2339                    NaN       79       False              NaN
2340                    NaN       47       False              NaN
2341                    NaN      174       False              NaN
2342                    NaN       67       False              NaN
2343                    NaN      232       False              NaN
2344                    NaN       61       False              NaN
2345                    NaN      146       False              NaN
2346                    NaN      261       False              NaN
2347                    NaN      879       False              NaN
2348                    NaN       60       False              NaN
2349                    NaN       41       False              NaN
2350                    NaN      147       False              NaN
2351                    NaN       47       False              NaN
2352                    NaN       48       False              NaN
2353                    NaN      532       False              NaN

                                                source truncated  \
0    <a href="http://twitter.com/download/iphone" r...      False
1    <a href="http://twitter.com/download/iphone" r...      False
2    <a href="http://twitter.com/download/iphone" r...      False
3    <a href="http://twitter.com/download/iphone" r...      False
4    <a href="http://twitter.com/download/iphone" r...      False
5    <a href="http://twitter.com/download/iphone" r...      False
6    <a href="http://twitter.com/download/iphone" r...      False
7    <a href="http://twitter.com/download/iphone" r...      False
8    <a href="http://twitter.com/download/iphone" r...      False
9    <a href="http://twitter.com/download/iphone" r...      False
10   <a href="http://twitter.com/download/iphone" r...      False
11   <a href="http://twitter.com/download/iphone" r...      False
12   <a href="http://twitter.com/download/iphone" r...      False
13   <a href="http://twitter.com/download/iphone" r...      False
14   <a href="http://twitter.com/download/iphone" r...      False
15   <a href="http://twitter.com/download/iphone" r...      False
16   <a href="http://twitter.com/download/iphone" r...      False
17   <a href="http://twitter.com/download/iphone" r...      False
18   <a href="http://twitter.com/download/iphone" r...      False
19   <a href="http://twitter.com/download/iphone" r...      False
20   <a href="http://twitter.com/download/iphone" r...      False
21   <a href="http://twitter.com/download/iphone" r...      False
22   <a href="http://twitter.com/download/iphone" r...      False
23   <a href="http://twitter.com/download/iphone" r...      False
```

```
24     <a href="http://twitter.com/download/iphone" r...        False
25     <a href="http://twitter.com/download/iphone" r...        False
26     <a href="http://twitter.com/download/iphone" r...        False
27     <a href="http://twitter.com/download/iphone" r...        False
28     <a href="http://twitter.com/download/iphone" r...        False
29     <a href="http://twitter.com/download/iphone" r...        False
...                                                    ...          ...
2324   <a href="http://twitter.com/download/iphone" r...        False
2325   <a href="http://twitter.com/download/iphone" r...        False
2326   <a href="http://twitter.com/download/iphone" r...        False
2327   <a href="http://twitter.com/download/iphone" r...        False
2328   <a href="http://twitter.com/download/iphone" r...        False
2329   <a href="http://twitter.com/download/iphone" r...        False
2330   <a href="http://twitter.com/download/iphone" r...        False
2331   <a href="http://twitter.com/download/iphone" r...        False
2332   <a href="http://twitter.com/download/iphone" r...        False
2333   <a href="http://twitter.com/download/iphone" r...        False
2334   <a href="http://twitter.com/download/iphone" r...        False
2335   <a href="http://twitter.com/download/iphone" r...        False
2336   <a href="http://twitter.com/download/iphone" r...        False
2337   <a href="http://twitter.com/download/iphone" r...        False
2338   <a href="http://twitter.com/download/iphone" r...        False
2339   <a href="http://twitter.com/download/iphone" r...        False
2340   <a href="http://twitter.com/download/iphone" r...        False
2341   <a href="http://twitter.com/download/iphone" r...        False
2342   <a href="http://twitter.com/download/iphone" r...        False
2343   <a href="http://twitter.com/download/iphone" r...        False
2344   <a href="http://twitter.com/download/iphone" r...        False
2345   <a href="http://twitter.com/download/iphone" r...        False
2346   <a href="http://twitter.com/download/iphone" r...        False
2347   <a href="http://twitter.com/download/iphone" r...        False
2348   <a href="http://twitter.com/download/iphone" r...        False
2349   <a href="http://twitter.com/download/iphone" r...        False
2350   <a href="http://twitter.com/download/iphone" r...        False
2351   <a href="http://twitter.com/download/iphone" r...        False
2352   <a href="http://twitter.com/download/iphone" r...        False
2353   <a href="http://twitter.com/download/iphone" r...        False

                                                      user
0      {'id': 4196983835, 'id_str': '4196983835', 'na...
1      {'id': 4196983835, 'id_str': '4196983835', 'na...
2      {'id': 4196983835, 'id_str': '4196983835', 'na...
3      {'id': 4196983835, 'id_str': '4196983835', 'na...
4      {'id': 4196983835, 'id_str': '4196983835', 'na...
5      {'id': 4196983835, 'id_str': '4196983835', 'na...
6      {'id': 4196983835, 'id_str': '4196983835', 'na...
7      {'id': 4196983835, 'id_str': '4196983835', 'na...
8      {'id': 4196983835, 'id_str': '4196983835', 'na...
```

```
9     {'id': 4196983835, 'id_str': '4196983835', 'na...
10    {'id': 4196983835, 'id_str': '4196983835', 'na...
11    {'id': 4196983835, 'id_str': '4196983835', 'na...
12    {'id': 4196983835, 'id_str': '4196983835', 'na...
13    {'id': 4196983835, 'id_str': '4196983835', 'na...
14    {'id': 4196983835, 'id_str': '4196983835', 'na...
15    {'id': 4196983835, 'id_str': '4196983835', 'na...
16    {'id': 4196983835, 'id_str': '4196983835', 'na...
17    {'id': 4196983835, 'id_str': '4196983835', 'na...
18    {'id': 4196983835, 'id_str': '4196983835', 'na...
19    {'id': 4196983835, 'id_str': '4196983835', 'na...
20    {'id': 4196983835, 'id_str': '4196983835', 'na...
21    {'id': 4196983835, 'id_str': '4196983835', 'na...
22    {'id': 4196983835, 'id_str': '4196983835', 'na...
23    {'id': 4196983835, 'id_str': '4196983835', 'na...
24    {'id': 4196983835, 'id_str': '4196983835', 'na...
25    {'id': 4196983835, 'id_str': '4196983835', 'na...
26    {'id': 4196983835, 'id_str': '4196983835', 'na...
27    {'id': 4196983835, 'id_str': '4196983835', 'na...
28    {'id': 4196983835, 'id_str': '4196983835', 'na...
29    {'id': 4196983835, 'id_str': '4196983835', 'na...
...                                                 ...
2324  {'id': 4196983835, 'id_str': '4196983835', 'na...
2325  {'id': 4196983835, 'id_str': '4196983835', 'na...
2326  {'id': 4196983835, 'id_str': '4196983835', 'na...
2327  {'id': 4196983835, 'id_str': '4196983835', 'na...
2328  {'id': 4196983835, 'id_str': '4196983835', 'na...
2329  {'id': 4196983835, 'id_str': '4196983835', 'na...
2330  {'id': 4196983835, 'id_str': '4196983835', 'na...
2331  {'id': 4196983835, 'id_str': '4196983835', 'na...
2332  {'id': 4196983835, 'id_str': '4196983835', 'na...
2333  {'id': 4196983835, 'id_str': '4196983835', 'na...
2334  {'id': 4196983835, 'id_str': '4196983835', 'na...
2335  {'id': 4196983835, 'id_str': '4196983835', 'na...
2336  {'id': 4196983835, 'id_str': '4196983835', 'na...
2337  {'id': 4196983835, 'id_str': '4196983835', 'na...
2338  {'id': 4196983835, 'id_str': '4196983835', 'na...
2339  {'id': 4196983835, 'id_str': '4196983835', 'na...
2340  {'id': 4196983835, 'id_str': '4196983835', 'na...
2341  {'id': 4196983835, 'id_str': '4196983835', 'na...
2342  {'id': 4196983835, 'id_str': '4196983835', 'na...
2343  {'id': 4196983835, 'id_str': '4196983835', 'na...
2344  {'id': 4196983835, 'id_str': '4196983835', 'na...
2345  {'id': 4196983835, 'id_str': '4196983835', 'na...
2346  {'id': 4196983835, 'id_str': '4196983835', 'na...
2347  {'id': 4196983835, 'id_str': '4196983835', 'na...
2348  {'id': 4196983835, 'id_str': '4196983835', 'na...
2349  {'id': 4196983835, 'id_str': '4196983835', 'na...
```

```
         2350  {'id': 4196983835, 'id_str': '4196983835', 'na...
         2351  {'id': 4196983835, 'id_str': '4196983835', 'na...
         2352  {'id': 4196983835, 'id_str': '4196983835', 'na...
         2353  {'id': 4196983835, 'id_str': '4196983835', 'na...

         [2354 rows x 31 columns]

In [20]: # A concise summary of a DataFrame
         df_tweets.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 31 columns):
contributors                0 non-null float64
coordinates                 0 non-null float64
created_at                  2354 non-null datetime64[ns]
display_text_range          2354 non-null object
entities                    2354 non-null object
extended_entities           2073 non-null object
favorite_count              2354 non-null int64
favorited                   2354 non-null bool
full_text                   2354 non-null object
geo                         0 non-null float64
id                          2354 non-null int64
id_str                      2354 non-null int64
in_reply_to_screen_name     78 non-null object
in_reply_to_status_id       78 non-null float64
in_reply_to_status_id_str   78 non-null float64
in_reply_to_user_id         78 non-null float64
in_reply_to_user_id_str     78 non-null float64
is_quote_status             2354 non-null bool
lang                        2354 non-null object
place                       1 non-null object
possibly_sensitive          2211 non-null float64
possibly_sensitive_appealable  2211 non-null float64
quoted_status               28 non-null object
quoted_status_id            29 non-null float64
quoted_status_id_str        29 non-null float64
retweet_count               2354 non-null int64
retweeted                   2354 non-null bool
retweeted_status            179 non-null object
source                      2354 non-null object
truncated                   2354 non-null bool
user                        2354 non-null object
dtypes: bool(4), datetime64[ns](1), float64(11), int64(4), object(11)
memory usage: 505.8+ KB


In [21]: # Descriptive statistics of a DataFrame
```

```
df_tweets.describe()
```

Out[21]:
	contributors	coordinates	favorite_count	geo	id
count	0.0	0.0	2354.000000	0.0	2.354000e+03
mean	NaN	NaN	8080.968564	NaN	7.426978e+17
std	NaN	NaN	11814.771334	NaN	6.852812e+16
min	NaN	NaN	0.000000	NaN	6.660209e+17
25%	NaN	NaN	1415.000000	NaN	6.783975e+17
50%	NaN	NaN	3603.500000	NaN	7.194596e+17
75%	NaN	NaN	10122.250000	NaN	7.993058e+17
max	NaN	NaN	132810.000000	NaN	8.924206e+17

	id_str	in_reply_to_status_id	in_reply_to_status_id_str
count	2.354000e+03	7.800000e+01	7.800000e+01
mean	7.426978e+17	7.455079e+17	7.455079e+17
std	6.852812e+16	7.582492e+16	7.582492e+16
min	6.660209e+17	6.658147e+17	6.658147e+17
25%	6.783975e+17	6.757419e+17	6.757419e+17
50%	7.194596e+17	7.038708e+17	7.038708e+17
75%	7.993058e+17	8.257804e+17	8.257804e+17
max	8.924206e+17	8.862664e+17	8.862664e+17

	in_reply_to_user_id	in_reply_to_user_id_str	possibly_sensitive
count	7.800000e+01	7.800000e+01	2211.0
mean	2.014171e+16	2.014171e+16	0.0
std	1.252797e+17	1.252797e+17	0.0
min	1.185634e+07	1.185634e+07	0.0
25%	3.086374e+08	3.086374e+08	0.0
50%	4.196984e+09	4.196984e+09	0.0
75%	4.196984e+09	4.196984e+09	0.0
max	8.405479e+17	8.405479e+17	0.0

	possibly_sensitive_appealable	quoted_status_id	quoted_status_id_str
count	2211.0	2.900000e+01	2.900000e+01
mean	0.0	8.162686e+17	8.162686e+17
std	0.0	6.164161e+16	6.164161e+16
min	0.0	6.721083e+17	6.721083e+17
25%	0.0	7.888183e+17	7.888183e+17
50%	0.0	8.340867e+17	8.340867e+17
75%	0.0	8.664587e+17	8.664587e+17
max	0.0	8.860534e+17	8.860534e+17

	retweet_count
count	2354.000000
mean	3164.797366
std	5284.770364
min	0.000000
25%	624.500000

```
50%       1473.500000
75%       3652.000000
max      79515.000000
```

1.2.1 Quality

Twitter archive enhanced **table**:

- Invalid values in rating_denominator and rating_numerator columns.
- Inappropriate data types of timestamp, retweeted_status_timestamp, retweeted_status_id, retweeted_status_user_id columns.
- HTML tegs in source column.
- None instead of Null in some columns.
- Invalid names in name column.
- Unnecessary columns.

Image Predictions **table**: * Missing records(2075 rows instead of 2356). * Unnecessary columns.

Tweets **table**: * Unnecessary columns(contributors, coordinates, geo, place, ...).

1.2.2 Tidiness

- Four columns instead of one(doggo, floofer, pupper, puppo).
- Three columns for prediction, prediction confidence and prediction dog.
- No uniform DataFrames(for convinience).

1.3 Clean

```
In [22]: # Making copies of our initial DataFrames
         df_twitter_clean = df_twitter.copy()
         df_img_pred_clean = df_img_pred.copy()
         df_tweets_clean = df_tweets.copy()
```

1.3.1 Quality I

Tweets: Unnecessary columns(contributors, coordinates, geo, place, ...)

Define   Drop unnecessary columns for analysis

Code

```
In [23]: df_tweets_clean = df_tweets_clean[['id', 'favorite_count', 'retweet_count']]
         df_tweets_clean.rename(columns = {'id': 'tweet_id'}, inplace = True)
```

Test

```
In [24]: df_tweets_clean.head()

Out[24]:              tweet_id  favorite_count  retweet_count
         0   892420643555336193           39467           8853
         1   892177421306343426           33819           6514
         2   891815181378084864           25461           4328
         3   891689557279858688           42908           8964
         4   891327558926688256           41048           9774

In [25]: df_tweets_clean.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
tweet_id         2354 non-null int64
favorite_count   2354 non-null int64
retweet_count    2354 non-null int64
dtypes: int64(3)
memory usage: 55.2 KB
```

1.3.2 Tidiness I

No uniform DataFrames to convinience

Define   Merging three DataFrames into one

Code

```
In [26]: general = pd.merge(df_img_pred_clean, df_tweets_clean, on = 'tweet_id')
         general_clean = pd.merge(df_twitter_clean, general, on = 'tweet_id')
```

Test

```
In [27]: general_clean.head(3)

Out[27]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0   892420643555336193                    NaN                  NaN
         1   892177421306343426                    NaN                  NaN
         2   891815181378084864                    NaN                  NaN

                          timestamp  \
         0   2017-08-01 16:23:56 +0000
```

```
    1   2017-08-01 00:17:27 +0000
    2   2017-07-31 00:18:03 +0000

                                                    source  \
    0   <a href="http://twitter.com/download/iphone" r...
    1   <a href="http://twitter.com/download/iphone" r...
    2   <a href="http://twitter.com/download/iphone" r...

                                                      text  retweeted_status_id  \
    0   This is Phineas. He's a mystical boy. Only eve...                   NaN
    1   This is Tilly. She's just checking pup on you...                    NaN
    2   This is Archie. He is a rare Norwegian Pouncin...                   NaN

        retweeted_status_user_id retweeted_status_timestamp  \
    0                        NaN                        NaN
    1                        NaN                        NaN
    2                        NaN                        NaN

                                      expanded_urls     ...         p1_conf  \
    0   https://twitter.com/dog_rates/status/892420643...     ...         0.097049
    1   https://twitter.com/dog_rates/status/892177421...     ...         0.323581
    2   https://twitter.com/dog_rates/status/891815181...     ...         0.716012

        p1_dog         p2   p2_conf p2_dog        p3   p3_conf p3_dog  \
    0    False      bagel  0.085851  False    banana  0.076110  False
    1     True   Pekinese  0.090647   True  papillon  0.068957   True
    2     True   malamute  0.078253   True    kelpie  0.031379   True

        favorite_count  retweet_count
    0            39467           8853
    1            33819           6514
    2            25461           4328

    [3 rows x 30 columns]

In [28]: general_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2073 entries, 0 to 2072
Data columns (total 30 columns):
tweet_id                    2073 non-null int64
in_reply_to_status_id       23 non-null float64
in_reply_to_user_id         23 non-null float64
timestamp                   2073 non-null object
source                      2073 non-null object
text                        2073 non-null object
retweeted_status_id         79 non-null float64
retweeted_status_user_id    79 non-null float64
```

```
retweeted_status_timestamp    79 non-null object
expanded_urls               2073 non-null object
rating_numerator            2073 non-null int64
rating_denominator          2073 non-null int64
name                        2073 non-null object
doggo                       2073 non-null object
floofer                     2073 non-null object
pupper                      2073 non-null object
puppo                       2073 non-null object
jpg_url                     2073 non-null object
img_num                     2073 non-null int64
p1                          2073 non-null object
p1_conf                     2073 non-null float64
p1_dog                      2073 non-null bool
p2                          2073 non-null object
p2_conf                     2073 non-null float64
p2_dog                      2073 non-null bool
p3                          2073 non-null object
p3_conf                     2073 non-null float64
p3_dog                      2073 non-null bool
favorite_count              2073 non-null int64
retweet_count               2073 non-null int64
dtypes: bool(3), float64(7), int64(6), object(14)
memory usage: 459.5+ KB
```

Four columns instead of one(doggo, floofer, pupper, puppo)

Define   Transform doggo, floofer, pupper, puppo columns into one(stages_of_dog) using melt method

Code

```
In [29]: general_clean.columns

Out[29]: Index(['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',
                'jpg_url', 'img_num', 'p1', 'p1_conf', 'p1_dog', 'p2', 'p2_conf',
                'p2_dog', 'p3', 'p3_conf', 'p3_dog', 'favorite_count', 'retweet_count'],
               dtype='object')

In [30]: general_clean = pd.melt(general_clean, id_vars=['tweet_id', 'in_reply_to_status_id', 'i
                'source', 'text', 'retweeted_status_id', 'retweeted_status_user_id',
                'retweeted_status_timestamp', 'expanded_urls', 'rating_numerator',
                'rating_denominator', 'name', 'jpg_url', 'img_num', 'p1', 'p1_conf', 'p1_dog', '
                'p2_dog', 'p3', 'p3_conf', 'p3_dog', 'favorite_count', 'retweet_count'], var_nam
```

Test

In [31]: general_clean

Out[31]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
        0   892420643555336193   NaN   NaN
        1   892177421306343426   NaN   NaN
        2   891815181378084864   NaN   NaN
        3   891689557279858688   NaN   NaN
        4   891327558926688256   NaN   NaN
        5   891087950875897856   NaN   NaN
        6   890971913173991426   NaN   NaN
        7   890729181411237888   NaN   NaN
        8   890609185150312448   NaN   NaN
        9   890240255349198849   NaN   NaN
        10   890006608113172480   NaN   NaN
        11   889880896479866881   NaN   NaN
        12   889665388333682689   NaN   NaN
        13   889638837579907072   NaN   NaN
        14   889531135344209921   NaN   NaN
        15   889278841981685760   NaN   NaN
        16   888917238123831296   NaN   NaN
        17   888804989199671297   NaN   NaN
        18   888554962724278272   NaN   NaN
        19   888078434458587136   NaN   NaN
        20   887705289381826560   NaN   NaN
        21   887517139158093824   NaN   NaN
        22   887473957103951883   NaN   NaN
        23   887343217045368832   NaN   NaN
        24   887101392804085760   NaN   NaN
        25   886983233522544640   NaN   NaN
        26   886736880519319552   NaN   NaN
        27   886680336477933568   NaN   NaN
        28   886366144734445568   NaN   NaN
        29   886258384151887873   NaN   NaN
        ...   .
        8262   666411507551481857   NaN   NaN
        8263   666407126856765440   NaN   NaN
        8264   666396247373291520   NaN   NaN
        8265   666373753744588802   NaN   NaN
        8266   666362758909284353   NaN   NaN
        8267   666353288456101888   NaN   NaN
        8268   666345417576210432   NaN   NaN
        8269   666337882303524864   NaN   NaN
        8270   666293911632134144   NaN   NaN
        8271   666287406224695296   NaN   NaN
        8272   666273097616637952   NaN   NaN
        8273   666268910803644416   NaN   NaN
        8274   666104133288665088   NaN   NaN

8275	666102155909144576	NaN	NaN
8276	666099513787052032	NaN	NaN
8277	666094000022159362	NaN	NaN
8278	666082916733198337	NaN	NaN
8279	666073100786774016	NaN	NaN
8280	666071193221509120	NaN	NaN
8281	666063827256086533	NaN	NaN
8282	666058600524156928	NaN	NaN
8283	666057090499244032	NaN	NaN
8284	666055525042405380	NaN	NaN
8285	666051853826850816	NaN	NaN
8286	666050758794694657	NaN	NaN
8287	666049248165822465	NaN	NaN
8288	666044226329800704	NaN	NaN
8289	666033412701032449	NaN	NaN
8290	666029285002620928	NaN	NaN
8291	666020888022790149	NaN	NaN

```
                          timestamp  \
0      2017-08-01 16:23:56 +0000
1      2017-08-01 00:17:27 +0000
2      2017-07-31 00:18:03 +0000
3      2017-07-30 15:58:51 +0000
4      2017-07-29 16:00:24 +0000
5      2017-07-29 00:08:17 +0000
6      2017-07-28 16:27:12 +0000
7      2017-07-28 00:22:40 +0000
8      2017-07-27 16:25:51 +0000
9      2017-07-26 15:59:51 +0000
10     2017-07-26 00:31:25 +0000
11     2017-07-25 16:11:53 +0000
12     2017-07-25 01:55:32 +0000
13     2017-07-25 00:10:02 +0000
14     2017-07-24 17:02:04 +0000
15     2017-07-24 00:19:32 +0000
16     2017-07-23 00:22:39 +0000
17     2017-07-22 16:56:37 +0000
18     2017-07-22 00:23:06 +0000
19     2017-07-20 16:49:33 +0000
20     2017-07-19 16:06:48 +0000
21     2017-07-19 03:39:09 +0000
22     2017-07-19 00:47:34 +0000
23     2017-07-18 16:08:03 +0000
24     2017-07-18 00:07:08 +0000
25     2017-07-17 16:17:36 +0000
26     2017-07-16 23:58:41 +0000
27     2017-07-16 20:14:00 +0000
28     2017-07-15 23:25:31 +0000
```

```
29      2017-07-15 16:17:19 +0000
...                             ...
8262    2015-11-17 00:24:19 +0000
8263    2015-11-17 00:06:54 +0000
8264    2015-11-16 23:23:41 +0000
8265    2015-11-16 21:54:18 +0000
8266    2015-11-16 21:10:36 +0000
8267    2015-11-16 20:32:58 +0000
8268    2015-11-16 20:01:42 +0000
8269    2015-11-16 19:31:45 +0000
8270    2015-11-16 16:37:02 +0000
8271    2015-11-16 16:11:11 +0000
8272    2015-11-16 15:14:19 +0000
8273    2015-11-16 14:57:41 +0000
8274    2015-11-16 04:02:55 +0000
8275    2015-11-16 03:55:04 +0000
8276    2015-11-16 03:44:34 +0000
8277    2015-11-16 03:22:39 +0000
8278    2015-11-16 02:38:37 +0000
8279    2015-11-16 01:59:36 +0000
8280    2015-11-16 01:52:02 +0000
8281    2015-11-16 01:22:45 +0000
8282    2015-11-16 01:01:59 +0000
8283    2015-11-16 00:55:59 +0000
8284    2015-11-16 00:49:46 +0000
8285    2015-11-16 00:35:11 +0000
8286    2015-11-16 00:30:50 +0000
8287    2015-11-16 00:24:50 +0000
8288    2015-11-16 00:04:52 +0000
8289    2015-11-15 23:21:54 +0000
8290    2015-11-15 23:05:30 +0000
8291    2015-11-15 22:32:08 +0000


                                                  source  \
0       <a href="http://twitter.com/download/iphone" r...
1       <a href="http://twitter.com/download/iphone" r...
2       <a href="http://twitter.com/download/iphone" r...
3       <a href="http://twitter.com/download/iphone" r...
4       <a href="http://twitter.com/download/iphone" r...
5       <a href="http://twitter.com/download/iphone" r...
6       <a href="http://twitter.com/download/iphone" r...
7       <a href="http://twitter.com/download/iphone" r...
8       <a href="http://twitter.com/download/iphone" r...
9       <a href="http://twitter.com/download/iphone" r...
10      <a href="http://twitter.com/download/iphone" r...
11      <a href="http://twitter.com/download/iphone" r...
12      <a href="http://twitter.com/download/iphone" r...
13      <a href="http://twitter.com/download/iphone" r...
```

```
14      <a href="http://twitter.com/download/iphone" r...
15      <a href="http://twitter.com/download/iphone" r...
16      <a href="http://twitter.com/download/iphone" r...
17      <a href="http://twitter.com/download/iphone" r...
18      <a href="http://twitter.com/download/iphone" r...
19      <a href="http://twitter.com/download/iphone" r...
20      <a href="http://twitter.com/download/iphone" r...
21      <a href="http://twitter.com/download/iphone" r...
22      <a href="http://twitter.com/download/iphone" r...
23      <a href="http://twitter.com/download/iphone" r...
24      <a href="http://twitter.com/download/iphone" r...
25      <a href="http://twitter.com/download/iphone" r...
26      <a href="http://twitter.com/download/iphone" r...
27      <a href="http://twitter.com/download/iphone" r...
28      <a href="http://twitter.com/download/iphone" r...
29      <a href="http://twitter.com/download/iphone" r...
...                                                  ...
8262    <a href="http://twitter.com/download/iphone" r...
8263    <a href="http://twitter.com/download/iphone" r...
8264    <a href="http://twitter.com/download/iphone" r...
8265    <a href="http://twitter.com/download/iphone" r...
8266    <a href="http://twitter.com/download/iphone" r...
8267    <a href="http://twitter.com/download/iphone" r...
8268    <a href="http://twitter.com/download/iphone" r...
8269    <a href="http://twitter.com/download/iphone" r...
8270    <a href="http://twitter.com/download/iphone" r...
8271    <a href="http://twitter.com/download/iphone" r...
8272    <a href="http://twitter.com/download/iphone" r...
8273    <a href="http://twitter.com/download/iphone" r...
8274    <a href="http://twitter.com/download/iphone" r...
8275    <a href="http://twitter.com/download/iphone" r...
8276    <a href="http://twitter.com/download/iphone" r...
8277    <a href="http://twitter.com/download/iphone" r...
8278    <a href="http://twitter.com/download/iphone" r...
8279    <a href="http://twitter.com/download/iphone" r...
8280    <a href="http://twitter.com/download/iphone" r...
8281    <a href="http://twitter.com/download/iphone" r...
8282    <a href="http://twitter.com/download/iphone" r...
8283    <a href="http://twitter.com/download/iphone" r...
8284    <a href="http://twitter.com/download/iphone" r...
8285    <a href="http://twitter.com/download/iphone" r...
8286    <a href="http://twitter.com/download/iphone" r...
8287    <a href="http://twitter.com/download/iphone" r...
8288    <a href="http://twitter.com/download/iphone" r...
8289    <a href="http://twitter.com/download/iphone" r...
8290    <a href="http://twitter.com/download/iphone" r...
8291    <a href="http://twitter.com/download/iphone" r...
```

```
                                                          text  retweeted_status_id  \
0     This is Phineas. He's a mystical boy. Only eve...                  NaN
1     This is Tilly. She's just checking pup on you...                   NaN
2     This is Archie. He is a rare Norwegian Pouncin...                  NaN
3     This is Darla. She commenced a snooze mid meal...                  NaN
4     This is Franklin. He would like you to stop ca...                  NaN
5     Here we have a majestic great white breaching ...                  NaN
6     Meet Jax. He enjoys ice cream so much he gets ...                   NaN
7     When you watch your owner call another dog a g...                   NaN
8     This is Zoey. She doesn't want to be one of th...                   NaN
9     This is Cassie. She is a college pup. Studying...                   NaN
10    This is Koda. He is a South Australian decksha...                   NaN
11    This is Bruno. He is a service shark. Only get...                   NaN
12    Here's a puppo that seems to be on the fence a...                   NaN
13    This is Ted. He does his best. Sometimes that'...                   NaN
14    This is Stuart. He's sporting his favorite fan...                   NaN
15    This is Oliver. You're witnessing one of his m...                   NaN
16    This is Jim. He found a fren. Taught him how t...                    NaN
17    This is Zeke. He has a new stick. Very proud o...                    NaN
18    This is Ralphus. He's powering up. Attempting ...                   NaN
19    This is Gerald. He was just told he didn't get...                   NaN
20    This is Jeffrey. He has a monopoly on the pool...                   NaN
21    I've yet to rate a Venezuelan Hover Wiener. Th...                    NaN
22    This is Canela. She attempted some fancy porch...                   NaN
23    You may not have known you needed to see this ...                    NaN
24    This... is a Jubilant Antarctic House Bear. We...                    NaN
25    This is Maya. She's very shy. Rarely leaves he...                    NaN
26    This is Mingus. He's a wonderful father to his...                   NaN
27    This is Derek. He's late for a dog meeting. 13...                    NaN
28    This is Roscoe. Another pupper fallen victim t...                    NaN
29    This is Waffles. His doggles are pupside down...                     NaN
...                                                 ...                    ...
8262  This is quite the dog. Gets really excited whe...                   NaN
8263  This is a southern Vesuvius bumblegruff. Can d...                   NaN
8264  Oh goodness. A super rare northeast Qdoba kang...                   NaN
8265  Those are sunglasses and a jean jacket. 11/10 ...                   NaN
8266  Unique dog here. Very small. Lives in containe...                   NaN
8267  Here we have a mixed Asiago from the Galápagos...                    NaN
8268  Look at this jokester thinking seat belt laws ...                    NaN
8269  This is an extremely rare horned Parthenon. No...                   NaN
8270  This is a funny dog. Weird toes. Won't come do...                    NaN
8271  This is an Albanian 3 1/2 legged  Episcopalian...                   NaN
8272      Can take selfies 11/10 https://t.co/ws2AMaNwPW                   NaN
8273  Very concerned about fellow dog trapped in com...                   NaN
8274  Not familiar with this breed. No tail (weird)...                     NaN
8275  Oh my. Here you are seeing an Adobe Setter giv...                    NaN
8276  Can stand on stump for what seems like a while...                   NaN
8277  This appears to be a Mongolian Presbyterian mi...                   NaN
```

```
8278  Here we have a well-established sunblockerspan...         NaN
8279  Let's hope this flight isn't Malaysian (lol). ...        NaN
8280  Here we have a northern speckled Rhododendron...    NaN
8281  This is the happiest dog you will ever see. Ve...        NaN
8282  Here is the Rand Paul of retrievers folks! He'...        NaN
8283  My oh my. This is a rare blond Canadian terrie...        NaN
8284  Here is a Siberian heavily armored polar bear ...        NaN
8285  This is an odd dog. Hard on the outside but lo...        NaN
8286  This is a truly beautiful English Wilson Staff...        NaN
8287  Here we have a 1949 1st generation vulpix. Enj...        NaN
8288  This is a purebred Piers Morgan. Loves to Netf...        NaN
8289  Here is a very happy pup. Big fan of well-main...        NaN
8290  This is a western brown Mitsubishi terrier. Up...        NaN
8291  Here we have a Japanese Irish Setter. Lost eye...        NaN

      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                        NaN
1                          NaN                        NaN
2                          NaN                        NaN
3                          NaN                        NaN
4                          NaN                        NaN
5                          NaN                        NaN
6                          NaN                        NaN
7                          NaN                        NaN
8                          NaN                        NaN
9                          NaN                        NaN
10                         NaN                        NaN
11                         NaN                        NaN
12                         NaN                        NaN
13                         NaN                        NaN
14                         NaN                        NaN
15                         NaN                        NaN
16                         NaN                        NaN
17                         NaN                        NaN
18                         NaN                        NaN
19                         NaN                        NaN
20                         NaN                        NaN
21                         NaN                        NaN
22                         NaN                        NaN
23                         NaN                        NaN
24                         NaN                        NaN
25                         NaN                        NaN
26                         NaN                        NaN
27                         NaN                        NaN
28                         NaN                        NaN
29                         NaN                        NaN
...                        ...                        ...
8262                       NaN                        NaN
```

```
8263                          NaN                          NaN
8264                          NaN                          NaN
8265                          NaN                          NaN
8266                          NaN                          NaN
8267                          NaN                          NaN
8268                          NaN                          NaN
8269                          NaN                          NaN
8270                          NaN                          NaN
8271                          NaN                          NaN
8272                          NaN                          NaN
8273                          NaN                          NaN
8274                          NaN                          NaN
8275                          NaN                          NaN
8276                          NaN                          NaN
8277                          NaN                          NaN
8278                          NaN                          NaN
8279                          NaN                          NaN
8280                          NaN                          NaN
8281                          NaN                          NaN
8282                          NaN                          NaN
8283                          NaN                          NaN
8284                          NaN                          NaN
8285                          NaN                          NaN
8286                          NaN                          NaN
8287                          NaN                          NaN
8288                          NaN                          NaN
8289                          NaN                          NaN
8290                          NaN                          NaN
8291                          NaN                          NaN

                                         expanded_urls  ...      \
0      https://twitter.com/dog_rates/status/892420643...  ...
1      https://twitter.com/dog_rates/status/892177421...  ...
2      https://twitter.com/dog_rates/status/891815181...  ...
3      https://twitter.com/dog_rates/status/891689557...  ...
4      https://twitter.com/dog_rates/status/891327558...  ...
5      https://twitter.com/dog_rates/status/891087950...  ...
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...  ...
7      https://twitter.com/dog_rates/status/890729181...  ...
8      https://twitter.com/dog_rates/status/890609185...  ...
9      https://twitter.com/dog_rates/status/890240255...  ...
10     https://twitter.com/dog_rates/status/890006608...  ...
11     https://twitter.com/dog_rates/status/889880896...  ...
12     https://twitter.com/dog_rates/status/889665388...  ...
13     https://twitter.com/dog_rates/status/889638837...  ...
14     https://twitter.com/dog_rates/status/889531135...  ...
15     https://twitter.com/dog_rates/status/889278841...  ...
16     https://twitter.com/dog_rates/status/888917238...  ...
```

```
17     https://twitter.com/dog_rates/status/888804989...   ...
18     https://twitter.com/dog_rates/status/888554962...   ...
19     https://twitter.com/dog_rates/status/888078434...   ...
20     https://twitter.com/dog_rates/status/887705289...   ...
21     https://twitter.com/dog_rates/status/887517139...   ...
22     https://twitter.com/dog_rates/status/887473957...   ...
23     https://twitter.com/dog_rates/status/887343217...   ...
24     https://twitter.com/dog_rates/status/887101392...   ...
25     https://twitter.com/dog_rates/status/886983233...   ...
26     https://www.gofundme.com/mingusneedsus,https:/...   ...
27     https://twitter.com/dog_rates/status/886680336...   ...
28     https://twitter.com/dog_rates/status/886366144...   ...
29     https://twitter.com/dog_rates/status/886258384...   ...
...                                                   ...   ...
8262   https://twitter.com/dog_rates/status/666411507...   ...
8263   https://twitter.com/dog_rates/status/666407126...   ...
8264   https://twitter.com/dog_rates/status/666396247...   ...
8265   https://twitter.com/dog_rates/status/666373753...   ...
8266   https://twitter.com/dog_rates/status/666362758...   ...
8267   https://twitter.com/dog_rates/status/666353288...   ...
8268   https://twitter.com/dog_rates/status/666345417...   ...
8269   https://twitter.com/dog_rates/status/666337882...   ...
8270   https://twitter.com/dog_rates/status/666293911...   ...
8271   https://twitter.com/dog_rates/status/666287406...   ...
8272   https://twitter.com/dog_rates/status/666273097...   ...
8273   https://twitter.com/dog_rates/status/666268910...   ...
8274   https://twitter.com/dog_rates/status/666104133...   ...
8275   https://twitter.com/dog_rates/status/666102155...   ...
8276   https://twitter.com/dog_rates/status/666099513...   ...
8277   https://twitter.com/dog_rates/status/666094000...   ...
8278   https://twitter.com/dog_rates/status/666082916...   ...
8279   https://twitter.com/dog_rates/status/666073100...   ...
8280   https://twitter.com/dog_rates/status/666071193...   ...
8281   https://twitter.com/dog_rates/status/666063827...   ...
8282   https://twitter.com/dog_rates/status/666058600...   ...
8283   https://twitter.com/dog_rates/status/666057090...   ...
8284   https://twitter.com/dog_rates/status/666055525...   ...
8285   https://twitter.com/dog_rates/status/666051853...   ...
8286   https://twitter.com/dog_rates/status/666050758...   ...
8287   https://twitter.com/dog_rates/status/666049248...   ...
8288   https://twitter.com/dog_rates/status/666044226...   ...
8289   https://twitter.com/dog_rates/status/666033412...   ...
8290   https://twitter.com/dog_rates/status/666029285...   ...
8291   https://twitter.com/dog_rates/status/666020888...   ...

                      p2   p2_conf p2_dog                 p3  \
0                  bagel  0.085851  False              banana
1               Pekinese  0.090647   True            papillon
```

2	malamute	0.078253	True	kelpie
3	Labrador_retriever	0.168086	True	spatula
4	English_springer	0.225770	True	German_short-haired_pointer
5	Irish_terrier	0.116317	True	Indian_elephant
6	Border_collie	0.199287	True	ice_lolly
7	Eskimo_dog	0.178406	True	Pembroke
8	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
9	Cardigan	0.451038	True	Chihuahua
10	Pomeranian	0.013884	True	chow
11	Labrador_retriever	0.151317	True	muzzle
12	Cardigan	0.027356	True	basenji
13	boxer	0.002129	True	Staffordshire_bullterrier
14	Labrador_retriever	0.013834	True	redbone
15	borzoi	0.194742	True	Saluki
16	Tibetan_mastiff	0.120184	True	Labrador_retriever
17	Labrador_retriever	0.184172	True	English_setter
18	Eskimo_dog	0.166511	True	malamute
19	pug	0.000932	True	bull_mastiff
20	redbone	0.087582	True	Weimaraner
21	tow_truck	0.029175	False	shopping_cart
22	Rhodesian_ridgeback	0.054950	True	beagle
23	sea_lion	0.275645	False	Weimaraner
24	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
25	toy_terrier	0.143528	True	can_opener
26	Great_Pyrenees	0.186136	True	Dandie_Dinmont
27	sports_car	0.139952	False	car_wheel
28	Chihuahua	0.000361	True	Boston_bull
29	shower_cap	0.025286	False	Siamese_cat
...	...	...	...	...
8262	barracouta	0.271485	False	gar
8263	bloodhound	0.244220	True	flat-coated_retriever
8264	toy_terrier	0.009397	True	papillon
8265	Afghan_hound	0.259551	True	briard
8266	skunk	0.002402	False	hamster
8267	Siberian_husky	0.147655	True	Eskimo_dog
8268	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
8269	Newfoundland	0.278407	True	groenendael
8270	otter	0.015250	False	great_grey_owl
8271	toy_poodle	0.063064	True	miniature_poodle
8272	toy_terrier	0.111884	True	basenji
8273	desk	0.085547	False	bookcase
8274	cock	0.033919	False	partridge
8275	Newfoundland	0.149842	True	borzoi
8276	Shih-Tzu	0.166192	True	Dandie_Dinmont
8277	German_shepherd	0.078260	True	malinois
8278	bull_mastiff	0.404722	True	French_bulldog
8279	English_foxhound	0.175382	True	Ibizan_hound
8280	Yorkshire_terrier	0.174201	True	Pekinese

47

```
8281          Tibetan_mastiff  0.093718   True        Labrador_retriever
8282               komondor  0.192305   True  soft-coated_wheaten_terrier
8283        shopping_basket  0.014594  False           golden_retriever
8284        Tibetan_mastiff  0.058279   True                   fur_coat
8285            mud_turtle  0.045885  False                   terrapin
8286       English_springer  0.263788   True  Greater_Swiss_Mountain_dog
8287            Rottweiler  0.243682   True                   Doberman
8288               redbone  0.360687   True          miniature_pinscher
8289               malinois  0.138584   True                 bloodhound
8290     miniature_pinscher  0.074192   True          Rhodesian_ridgeback
8291                collie  0.156665   True           Shetland_sheepdog

        p3_conf p3_dog  favorite_count  retweet_count stage_of_dog  value
0      0.076110  False           39467           8853        doggo   None
1      0.068957   True           33819           6514        doggo   None
2      0.031379   True           25461           4328        doggo   None
3      0.040836  False           42908           8964        doggo   None
4      0.175219   True           41048           9774        doggo   None
5      0.076902  False           20562           3261        doggo   None
6      0.193548  False           12041           2158        doggo   None
7      0.076507   True           56848          16716        doggo   None
8      0.118184   True           28226           4429        doggo   None
9      0.029248   True           32467           7711        doggo  doggo
10     0.008167   True           31166           7624        doggo   None
11     0.082981  False           28268           5156        doggo   None
12     0.004633   True           38818           8538        doggo   None
13     0.001498   True           27672           4735        doggo   None
14     0.007958   True           15359           2321        doggo   None
15     0.027351   True           25652           5637        doggo   None
16     0.105506   True           29611           4709        doggo   None
17     0.073482   True           26080           4559        doggo   None
18     0.111411   True           20290           3732        doggo   None
19     0.000903   True           22201           3653        doggo   None
20     0.026236   True           30779           5609        doggo   None
21     0.026321  False           46959          12082        doggo   None
22     0.038915   True           69871          18781        doggo   None
23     0.134203   True           34222          10737        doggo   None
24     0.029705   True           31061           6167        doggo   None
25     0.032253  False           35859           8084        doggo   None
26     0.086346   True           12306           3443        doggo   None
27     0.044173  False           22798           4610        doggo   None
28     0.000076   True           21524           3316        doggo   None
29     0.002849  False           28469           6523        doggo   None
...         ...    ...             ...            ...          ...    ...
8262   0.189945  False             459            339        puppo   None
8263   0.173810   True             113             44        puppo   None
8264   0.004577   True             172             92        puppo   None
8265   0.206803   True             194            100        puppo   None
```

```
8266   0.000461   False          804          595          puppo   None
8267   0.093412   True           229           77          puppo   None
8268   0.014241   True           307          146          puppo   None
8269   0.102643   True           204           96          puppo   None
8270   0.013207   False          522          368          puppo   None
8271   0.025581   True           152           71          puppo   None
8272   0.111152   True           184           82          puppo   None
8273   0.079480   False          108           37          puppo   None
8274   0.000052   False        14765         6871          puppo   None
8275   0.133649   True            81           16          puppo   None
8276   0.089688   True           164           73          puppo   None
8277   0.075628   True           169           79          puppo   None
8278   0.048960   True           121           47          puppo   None
8279   0.097471   True           335          174          puppo   None
8280   0.109454   True           154           67          puppo   None
8281   0.072427   True           496          232          puppo   None
8282   0.082086   True           115           61          puppo   None
8283   0.007959   True           304          146          puppo   None
8284   0.054449   False          448          261          puppo   None
8285   0.017885   False         1253          879          puppo   None
8286   0.016199   True           136           60          puppo   None
8287   0.154629   True           111           41          puppo   None
8288   0.222752   True           311          147          puppo   None
8289   0.116197   True           128           47          puppo   None
8290   0.072010   True           132           48          puppo   None
8291   0.061428   True          2535          532          puppo   None

[8292 rows x 28 columns]
```

In [32]: `general_clean.value.value_counts()`

Out[32]:
```
None      7959
pupper     221
doggo       80
puppo       24
floofer      8
Name: value, dtype: int64
```

1.3.3   Quality II

Unnecessary columns

Define   Drop unnecessary columns

Code

```
In [33]: general_clean.drop(['in_reply_to_status_id', 'in_reply_to_user_id', 'retweeted_status_i
                             'retweeted_status_user_id', 'retweeted_status_timestamp'], axis=1,
```

Test

```
In [34]: general_clean.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8292 entries, 0 to 8291
Data columns (total 23 columns):
tweet_id             8292 non-null int64
timestamp            8292 non-null object
source               8292 non-null object
text                 8292 non-null object
expanded_urls        8292 non-null object
rating_numerator     8292 non-null int64
rating_denominator   8292 non-null int64
name                 8292 non-null object
jpg_url              8292 non-null object
img_num              8292 non-null int64
p1                   8292 non-null object
p1_conf              8292 non-null float64
p1_dog               8292 non-null bool
p2                   8292 non-null object
p2_conf              8292 non-null float64
p2_dog               8292 non-null bool
p3                   8292 non-null object
p3_conf              8292 non-null float64
p3_dog               8292 non-null bool
favorite_count       8292 non-null int64
retweet_count        8292 non-null int64
stage_of_dog         8292 non-null object
value                8292 non-null object
dtypes: bool(3), float64(3), int64(6), object(11)
memory usage: 1.3+ MB
```

Data type of timestamp column

Define   Change the type of data using to_datetime() function

Code

```
In [35]: general_clean.timestamp = pd.to_datetime(general_clean.timestamp)
```

Test

```
In [36]: general_clean.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8292 entries, 0 to 8291
Data columns (total 23 columns):
tweet_id              8292 non-null int64
timestamp             8292 non-null datetime64[ns]
source                8292 non-null object
text                  8292 non-null object
expanded_urls         8292 non-null object
rating_numerator      8292 non-null int64
rating_denominator    8292 non-null int64
name                  8292 non-null object
jpg_url               8292 non-null object
img_num               8292 non-null int64
p1                    8292 non-null object
p1_conf               8292 non-null float64
p1_dog                8292 non-null bool
p2                    8292 non-null object
p2_conf               8292 non-null float64
p2_dog                8292 non-null bool
p3                    8292 non-null object
p3_conf               8292 non-null float64
p3_dog                8292 non-null bool
favorite_count        8292 non-null int64
retweet_count         8292 non-null int64
stage_of_dog          8292 non-null object
value                 8292 non-null object
dtypes: bool(3), datetime64[ns](1), float64(3), int64(6), object(10)
memory usage: 1.3+ MB
```

Invalid values in rating_denominator and rating_numerator columns

Define Find the quotient of rating_denominator and rating_numerator columns to find the rating(It seems that the invalid values in the rating_denominator and rating_numerator columns are not random, because when a large value appears in some column, approximately the same value appears in another column(f.e. 140 and 134, 76 and 70. Not like 90 and 10 or 12 and 50). Because of this, i will find the quotient of these columns so that there is not much difference between the values)

Code

```
In [37]: general_clean['rating'] = general_clean.rating_numerator / general_clean.rating_denomin

In [38]: general_clean.drop(['rating_numerator', 'rating_denominator'], axis = 1, inplace = True
```

Test

In [39]: general_clean

Out[39]:   tweet_id   timestamp \
         0   892420643555336193 2017-08-01 16:23:56
         1   892177421306343426 2017-08-01 00:17:27
         2   891815181378084864 2017-07-31 00:18:03
         3   891689557279858688 2017-07-30 15:58:51
         4   891327558926688256 2017-07-29 16:00:24
         5   891087950875897856 2017-07-29 00:08:17
         6   890971913173991426 2017-07-28 16:27:12
         7   890729181411237888 2017-07-28 00:22:40
         8   890609185150312448 2017-07-27 16:25:51
         9   890240255349198849 2017-07-26 15:59:51
         10   890006608113172480 2017-07-26 00:31:25
         11   889880896479866881 2017-07-25 16:11:53
         12   889665388333682689 2017-07-25 01:55:32
         13   889638837579907072 2017-07-25 00:10:02
         14   889531135344209921 2017-07-24 17:02:04
         15   889278841981685760 2017-07-24 00:19:32
         16   888917238123831296 2017-07-23 00:22:39
         17   888804989199671297 2017-07-22 16:56:37
         18   888554962724278272 2017-07-22 00:23:06
         19   888078434458587136 2017-07-20 16:49:33
         20   887705289381826560 2017-07-19 16:06:48
         21   887517139158093824 2017-07-19 03:39:09
         22   887473957103951883 2017-07-19 00:47:34
         23   887343217045368832 2017-07-18 16:08:03
         24   887101392804085760 2017-07-18 00:07:08
         25   886983233522544640 2017-07-17 16:17:36
         26   886736880519319552 2017-07-16 23:58:41
         27   886680336477933568 2017-07-16 20:14:00
         28   886366144734445568 2017-07-15 23:25:31
         29   886258384151887873 2017-07-15 16:17:19
         ...   .
         8262   666411507551481857 2015-11-17 00:24:19
         8263   666407126856765440 2015-11-17 00:06:54
         8264   666396247373291520 2015-11-16 23:23:41
         8265   666373753744588802 2015-11-16 21:54:18
         8266   666362758909284353 2015-11-16 21:10:36
         8267   666353288456101888 2015-11-16 20:32:58
         8268   666345417576210432 2015-11-16 20:01:42
         8269   666337882303524864 2015-11-16 19:31:45
         8270   666293911632134144 2015-11-16 16:37:02
         8271   666287406224695296 2015-11-16 16:11:11
         8272   666273097616637952 2015-11-16 15:14:19
         8273   666268910803644416 2015-11-16 14:57:41
         8274   666104133288665088 2015-11-16 04:02:55

```
8275  666102155909144576  2015-11-16 03:55:04
8276  666099513787052032  2015-11-16 03:44:34
8277  666094000022159362  2015-11-16 03:22:39
8278  666082916733198337  2015-11-16 02:38:37
8279  666073100786774016  2015-11-16 01:59:36
8280  666071193221509120  2015-11-16 01:52:02
8281  666063827256086533  2015-11-16 01:22:45
8282  666058600524156928  2015-11-16 01:01:59
8283  666057090499244032  2015-11-16 00:55:59
8284  666055525042405380  2015-11-16 00:49:46
8285  666051853826850816  2015-11-16 00:35:11
8286  666050758794694657  2015-11-16 00:30:50
8287  666049248165822465  2015-11-16 00:24:50
8288  666044226329800704  2015-11-16 00:04:52
8289  666033412701032449  2015-11-15 23:21:54
8290  666029285002620928  2015-11-15 23:05:30
8291  666020888022790149  2015-11-15 22:32:08

                                                     source  \
0     <a href="http://twitter.com/download/iphone" r...
1     <a href="http://twitter.com/download/iphone" r...
2     <a href="http://twitter.com/download/iphone" r...
3     <a href="http://twitter.com/download/iphone" r...
4     <a href="http://twitter.com/download/iphone" r...
5     <a href="http://twitter.com/download/iphone" r...
6     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
8     <a href="http://twitter.com/download/iphone" r...
9     <a href="http://twitter.com/download/iphone" r...
10    <a href="http://twitter.com/download/iphone" r...
11    <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
13    <a href="http://twitter.com/download/iphone" r...
14    <a href="http://twitter.com/download/iphone" r...
15    <a href="http://twitter.com/download/iphone" r...
16    <a href="http://twitter.com/download/iphone" r...
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
```

```
29     <a href="http://twitter.com/download/iphone" r...
...                                                  ...
8262   <a href="http://twitter.com/download/iphone" r...
8263   <a href="http://twitter.com/download/iphone" r...
8264   <a href="http://twitter.com/download/iphone" r...
8265   <a href="http://twitter.com/download/iphone" r...
8266   <a href="http://twitter.com/download/iphone" r...
8267   <a href="http://twitter.com/download/iphone" r...
8268   <a href="http://twitter.com/download/iphone" r...
8269   <a href="http://twitter.com/download/iphone" r...
8270   <a href="http://twitter.com/download/iphone" r...
8271   <a href="http://twitter.com/download/iphone" r...
8272   <a href="http://twitter.com/download/iphone" r...
8273   <a href="http://twitter.com/download/iphone" r...
8274   <a href="http://twitter.com/download/iphone" r...
8275   <a href="http://twitter.com/download/iphone" r...
8276   <a href="http://twitter.com/download/iphone" r...
8277   <a href="http://twitter.com/download/iphone" r...
8278   <a href="http://twitter.com/download/iphone" r...
8279   <a href="http://twitter.com/download/iphone" r...
8280   <a href="http://twitter.com/download/iphone" r...
8281   <a href="http://twitter.com/download/iphone" r...
8282   <a href="http://twitter.com/download/iphone" r...
8283   <a href="http://twitter.com/download/iphone" r...
8284   <a href="http://twitter.com/download/iphone" r...
8285   <a href="http://twitter.com/download/iphone" r...
8286   <a href="http://twitter.com/download/iphone" r...
8287   <a href="http://twitter.com/download/iphone" r...
8288   <a href="http://twitter.com/download/iphone" r...
8289   <a href="http://twitter.com/download/iphone" r...
8290   <a href="http://twitter.com/download/iphone" r...
8291   <a href="http://twitter.com/download/iphone" r...


                                                    text  \
0      This is Phineas. He's a mystical boy. Only eve...
1      This is Tilly. She's just checking pup on you...
2      This is Archie. He is a rare Norwegian Pouncin...
3      This is Darla. She commenced a snooze mid meal...
4      This is Franklin. He would like you to stop ca...
5      Here we have a majestic great white breaching ...
6      Meet Jax. He enjoys ice cream so much he gets ...
7      When you watch your owner call another dog a g...
8      This is Zoey. She doesn't want to be one of th...
9      This is Cassie. She is a college pup. Studying...
10     This is Koda. He is a South Australian decksha...
11     This is Bruno. He is a service shark. Only get...
12     Here's a puppo that seems to be on the fence a...
13     This is Ted. He does his best. Sometimes that'...
```

```
14     This is Stuart. He's sporting his favorite fan...
15     This is Oliver. You're witnessing one of his m...
16     This is Jim. He found a fren. Taught him how t...
17     This is Zeke. He has a new stick. Very proud o...
18     This is Ralphus. He's powering up. Attempting ...
19     This is Gerald. He was just told he didn't get...
20     This is Jeffrey. He has a monopoly on the pool...
21     I've yet to rate a Venezuelan Hover Wiener. Th...
22     This is Canela. She attempted some fancy porch...
23     You may not have known you needed to see this ...
24     This... is a Jubilant Antarctic House Bear. We...
25     This is Maya. She's very shy. Rarely leaves he...
26     This is Mingus. He's a wonderful father to his...
27     This is Derek. He's late for a dog meeting. 13...
28     This is Roscoe. Another pupper fallen victim t...
29     This is Waffles. His doggles are pupside down...
...                                                 ...
8262   This is quite the dog. Gets really excited whe...
8263   This is a southern Vesuvius bumblegruff. Can d...
8264   Oh goodness. A super rare northeast Qdoba kang...
8265   Those are sunglasses and a jean jacket. 11/10 ...
8266   Unique dog here. Very small. Lives in containe...
8267   Here we have a mixed Asiago from the Galápagos...
8268   Look at this jokester thinking seat belt laws ...
8269   This is an extremely rare horned Parthenon. No...
8270   This is a funny dog. Weird toes. Won't come do...
8271   This is an Albanian 3 1/2 legged  Episcopalian...
8272      Can take selfies 11/10 https://t.co/ws2AMaNwPW
8273   Very concerned about fellow dog trapped in com...
8274   Not familiar with this breed. No tail (weird)...
8275   Oh my. Here you are seeing an Adobe Setter giv...
8276   Can stand on stump for what seems like a while...
8277   This appears to be a Mongolian Presbyterian mi...
8278   Here we have a well-established sunblockerspan...
8279   Let's hope this flight isn't Malaysian (lol). ...
8280   Here we have a northern speckled Rhododendron...
8281   This is the happiest dog you will ever see. Ve...
8282   Here is the Rand Paul of retrievers folks! He'...
8283   My oh my. This is a rare blond Canadian terrie...
8284   Here is a Siberian heavily armored polar bear ...
8285   This is an odd dog. Hard on the outside but lo...
8286   This is a truly beautiful English Wilson Staff...
8287   Here we have a 1949 1st generation vulpix. Enj...
8288   This is a purebred Piers Morgan. Loves to Netf...
8289   Here is a very happy pup. Big fan of well-main...
8290   This is a western brown Mitsubishi terrier. Up...
8291   Here we have a Japanese Irish Setter. Lost eye...
```

```
                                      expanded_urls       name  \
0      https://twitter.com/dog_rates/status/892420643...    Phineas
1      https://twitter.com/dog_rates/status/892177421...      Tilly
2      https://twitter.com/dog_rates/status/891815181...     Archie
3      https://twitter.com/dog_rates/status/891689557...      Darla
4      https://twitter.com/dog_rates/status/891327558...   Franklin
5      https://twitter.com/dog_rates/status/891087950...       None
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...        Jax
7      https://twitter.com/dog_rates/status/890729181...       None
8      https://twitter.com/dog_rates/status/890609185...       Zoey
9      https://twitter.com/dog_rates/status/890240255...     Cassie
10     https://twitter.com/dog_rates/status/890006608...       Koda
11     https://twitter.com/dog_rates/status/889880896...      Bruno
12     https://twitter.com/dog_rates/status/889665388...       None
13     https://twitter.com/dog_rates/status/889638837...        Ted
14     https://twitter.com/dog_rates/status/889531135...     Stuart
15     https://twitter.com/dog_rates/status/889278841...     Oliver
16     https://twitter.com/dog_rates/status/888917238...        Jim
17     https://twitter.com/dog_rates/status/888804989...       Zeke
18     https://twitter.com/dog_rates/status/888554962...    Ralphus
19     https://twitter.com/dog_rates/status/888078434...     Gerald
20     https://twitter.com/dog_rates/status/887705289...    Jeffrey
21     https://twitter.com/dog_rates/status/887517139...       such
22     https://twitter.com/dog_rates/status/887473957...     Canela
23     https://twitter.com/dog_rates/status/887343217...       None
24     https://twitter.com/dog_rates/status/887101392...       None
25     https://twitter.com/dog_rates/status/886983233...       Maya
26     https://www.gofundme.com/mingusneedsus,https:/...     Mingus
27     https://twitter.com/dog_rates/status/886680336...      Derek
28     https://twitter.com/dog_rates/status/886366144...     Roscoe
29     https://twitter.com/dog_rates/status/886258384...    Waffles
...                                                 ...        ...
8262   https://twitter.com/dog_rates/status/666411507...      quite
8263   https://twitter.com/dog_rates/status/666407126...          a
8264   https://twitter.com/dog_rates/status/666396247...       None
8265   https://twitter.com/dog_rates/status/666373753...       None
8266   https://twitter.com/dog_rates/status/666362758...       None
8267   https://twitter.com/dog_rates/status/666353288...       None
8268   https://twitter.com/dog_rates/status/666345417...       None
8269   https://twitter.com/dog_rates/status/666337882...         an
8270   https://twitter.com/dog_rates/status/666293911...          a
8271   https://twitter.com/dog_rates/status/666287406...         an
8272   https://twitter.com/dog_rates/status/666273097...       None
8273   https://twitter.com/dog_rates/status/666268910...       None
8274   https://twitter.com/dog_rates/status/666104133...       None
8275   https://twitter.com/dog_rates/status/666102155...       None
8276   https://twitter.com/dog_rates/status/666099513...       None
8277   https://twitter.com/dog_rates/status/666094000...       None
```

```
8278   https://twitter.com/dog_rates/status/666082916...          None
8279   https://twitter.com/dog_rates/status/666073100...          None
8280   https://twitter.com/dog_rates/status/666071193...          None
8281   https://twitter.com/dog_rates/status/666063827...           the
8282   https://twitter.com/dog_rates/status/666058600...           the
8283   https://twitter.com/dog_rates/status/666057090...             a
8284   https://twitter.com/dog_rates/status/666055525...             a
8285   https://twitter.com/dog_rates/status/666051853...            an
8286   https://twitter.com/dog_rates/status/666050758...             a
8287   https://twitter.com/dog_rates/status/666049248...          None
8288   https://twitter.com/dog_rates/status/666044226...             a
8289   https://twitter.com/dog_rates/status/666033412...             a
8290   https://twitter.com/dog_rates/status/666029285...             a
8291   https://twitter.com/dog_rates/status/666020888...          None

                                             jpg_url  img_num  \
0         https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg         1
1         https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg         1
2         https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg         1
3         https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg         1
4         https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg         2
5         https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg         1
6         https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg         1
7         https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg         2
8         https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg         1
9         https://pbs.twimg.com/media/DFrEyVuWOAAO3t9.jpg         1
10        https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg         1
11        https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg         1
12        https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg         1
13        https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg         1
14        https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg         1
15    https://pbs.twimg.com/ext_tw_video_thumb/88927...         1
16        https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg         1
17        https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg         1
18        https://pbs.twimg.com/media/DFTH_O-UQAACu2O.jpg         3
19        https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg         1
20        https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg         1
21    https://pbs.twimg.com/ext_tw_video_thumb/88751...         1
22        https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg         2
23    https://pbs.twimg.com/ext_tw_video_thumb/88734...         1
24        https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg         1
25        https://pbs.twimg.com/media/DE8yicJWOAAAvBJ.jpg         2
26        https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg         1
27        https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg         1
28        https://pbs.twimg.com/media/DEOBTnQUwAApKEH.jpg         1
29        https://pbs.twimg.com/media/DEyfTG4UMAE4aE9.jpg         1
...                                               ...       ...
8262      https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg         1
```

```
8263    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg            1
8264    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg            1
8265    https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.jpg            1
8266    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg            1
8267    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg            1
8268    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg            1
8269    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg            1
8270    https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg            1
8271    https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg            1
8272    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg            1
8273    https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg            1
8274    https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg            1
8275    https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg            1
8276    https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg            1
8277    https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg            1
8278    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg            1
8279    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg            1
8280    https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg            1
8281    https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg            1
8282    https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg            1
8283    https://pbs.twimg.com/media/CT5PY9OWoAAQGLo.jpg            1
8284    https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg            1
8285    https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg            1
8286    https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg            1
8287    https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg            1
8288    https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg            1
8289    https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg            1
8290    https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg            1
8291    https://pbs.twimg.com/media/CT4udnOWwAAOaMy.jpg            1

                            p1   p1_conf  ...    p2_conf p2_dog  \
0                       orange  0.097049  ...   0.085851  False
1                    Chihuahua  0.323581  ...   0.090647   True
2                    Chihuahua  0.716012  ...   0.078253   True
3                  paper_towel  0.170278  ...   0.168086   True
4                       basset  0.555712  ...   0.225770   True
5     Chesapeake_Bay_retriever  0.425595  ...   0.116317   True
6                   Appenzeller  0.341703  ...   0.199287   True
7                   Pomeranian  0.566142  ...   0.178406   True
8                Irish_terrier  0.487574  ...   0.193054   True
9                     Pembroke  0.511319  ...   0.451038   True
10                     Samoyed  0.957979  ...   0.013884   True
11              French_bulldog  0.377417  ...   0.151317   True
12                    Pembroke  0.966327  ...   0.027356   True
13              French_bulldog  0.991650  ...   0.002129   True
14             golden_retriever 0.953442  ...   0.013834   True
15                     whippet  0.626152  ...   0.194742   True
16             golden_retriever 0.714719  ...   0.120184   True
```

58

17	golden_retriever	0.469760	...	0.184172	True
18	Siberian_husky	0.700377	...	0.166511	True
19	French_bulldog	0.995026	...	0.000932	True
20	basset	0.821664	...	0.087582	True
21	limousine	0.130432	...	0.029175	False
22	Pembroke	0.809197	...	0.054950	True
23	Mexican_hairless	0.330741	...	0.275645	False
24	Samoyed	0.733942	...	0.035029	True
25	Chihuahua	0.793469	...	0.143528	True
26	kuvasz	0.309706	...	0.186136	True
27	convertible	0.738995	...	0.139952	False
28	French_bulldog	0.999201	...	0.000361	True
29	pug	0.943575	...	0.025286	False
...	...	...	...	...	...
8262	coho	0.404640	...	0.271485	False
8263	black-and-tan_coonhound	0.529139	...	0.244220	True
8264	Chihuahua	0.978108	...	0.009397	True
8265	soft-coated_wheaten_terrier	0.326467	...	0.259551	True
8266	guinea_pig	0.996496	...	0.002402	False
8267	malamute	0.336874	...	0.147655	True
8268	golden_retriever	0.858744	...	0.054787	True
8269	ox	0.416669	...	0.278407	True
8270	three-toed_sloth	0.914671	...	0.015250	False
8271	Maltese_dog	0.857531	...	0.063064	True
8272	Italian_greyhound	0.176053	...	0.111884	True
8273	desktop_computer	0.086502	...	0.085547	False
8274	hen	0.965932	...	0.033919	False
8275	English_setter	0.298617	...	0.149842	True
8276	Lhasa	0.582330	...	0.166192	True
8277	bloodhound	0.195217	...	0.078260	True
8278	pug	0.489814	...	0.404722	True
8279	Walker_hound	0.260857	...	0.175382	True
8280	Gordon_setter	0.503672	...	0.174201	True
8281	golden_retriever	0.775930	...	0.093718	True
8282	miniature_poodle	0.201493	...	0.192305	True
8283	shopping_cart	0.962465	...	0.014594	False
8284	chow	0.692517	...	0.058279	True
8285	box_turtle	0.933012	...	0.045885	False
8286	Bernese_mountain_dog	0.651137	...	0.263788	True
8287	miniature_pinscher	0.560311	...	0.243682	True
8288	Rhodesian_ridgeback	0.408143	...	0.360687	True
8289	German_shepherd	0.596461	...	0.138584	True
8290	redbone	0.506826	...	0.074192	True
8291	Welsh_springer_spaniel	0.465074	...	0.156665	True

	p3	p3_conf	p3_dog	favorite_count \
0	banana	0.076110	False	39467
1	papillon	0.068957	True	33819

59

2	kelpie	0.031379	True	25461
3	spatula	0.040836	False	42908
4	German_short-haired_pointer	0.175219	True	41048
5	Indian_elephant	0.076902	False	20562
6	ice_lolly	0.193548	False	12041
7	Pembroke	0.076507	True	56848
8	Chesapeake_Bay_retriever	0.118184	True	28226
9	Chihuahua	0.029248	True	32467
10	chow	0.008167	True	31166
11	muzzle	0.082981	False	28268
12	basenji	0.004633	True	38818
13	Staffordshire_bullterrier	0.001498	True	27672
14	redbone	0.007958	True	15359
15	Saluki	0.027351	True	25652
16	Labrador_retriever	0.105506	True	29611
17	English_setter	0.073482	True	26080
18	malamute	0.111411	True	20290
19	bull_mastiff	0.000903	True	22201
20	Weimaraner	0.026236	True	30779
21	shopping_cart	0.026321	False	46959
22	beagle	0.038915	True	69871
23	Weimaraner	0.134203	True	34222
24	Staffordshire_bullterrier	0.029705	True	31061
25	can_opener	0.032253	False	35859
26	Dandie_Dinmont	0.086346	True	12306
27	car_wheel	0.044173	False	22798
28	Boston_bull	0.000076	True	21524
29	Siamese_cat	0.002849	False	28469
...	...	...	...	...
8262	gar	0.189945	False	459
8263	flat-coated_retriever	0.173810	True	113
8264	papillon	0.004577	True	172
8265	briard	0.206803	True	194
8266	hamster	0.000461	False	804
8267	Eskimo_dog	0.093412	True	229
8268	Labrador_retriever	0.014241	True	307
8269	groenendael	0.102643	True	204
8270	great_grey_owl	0.013207	False	522
8271	miniature_poodle	0.025581	True	152
8272	basenji	0.111152	True	184
8273	bookcase	0.079480	False	108
8274	partridge	0.000052	False	14765
8275	borzoi	0.133649	True	81
8276	Dandie_Dinmont	0.089688	True	164
8277	malinois	0.075628	True	169
8278	French_bulldog	0.048960	True	121
8279	Ibizan_hound	0.097471	True	335
8280	Pekinese	0.109454	True	154

8281	Labrador_retriever	0.072427	True	496
8282	soft-coated_wheaten_terrier	0.082086	True	115
8283	golden_retriever	0.007959	True	304
8284	fur_coat	0.054449	False	448
8285	terrapin	0.017885	False	1253
8286	Greater_Swiss_Mountain_dog	0.016199	True	136
8287	Doberman	0.154629	True	111
8288	miniature_pinscher	0.222752	True	311
8289	bloodhound	0.116197	True	128
8290	Rhodesian_ridgeback	0.072010	True	132
8291	Shetland_sheepdog	0.061428	True	2535

	retweet_count	stage_of_dog	value	rating
0	8853	doggo	None	1.3
1	6514	doggo	None	1.3
2	4328	doggo	None	1.2
3	8964	doggo	None	1.3
4	9774	doggo	None	1.2
5	3261	doggo	None	1.3
6	2158	doggo	None	1.3
7	16716	doggo	None	1.3
8	4429	doggo	None	1.3
9	7711	doggo	doggo	1.4
10	7624	doggo	None	1.3
11	5156	doggo	None	1.3
12	8538	doggo	None	1.3
13	4735	doggo	None	1.2
14	2321	doggo	None	1.3
15	5637	doggo	None	1.3
16	4709	doggo	None	1.2
17	4559	doggo	None	1.3
18	3732	doggo	None	1.3
19	3653	doggo	None	1.2
20	5609	doggo	None	1.3
21	12082	doggo	None	1.4
22	18781	doggo	None	1.3
23	10737	doggo	None	1.3
24	6167	doggo	None	1.2
25	8084	doggo	None	1.3
26	3443	doggo	None	1.3
27	4610	doggo	None	1.3
28	3316	doggo	None	1.2
29	6523	doggo	None	1.3
...	...	...	...	...
8262	339	puppo	None	0.2
8263	44	puppo	None	0.7
8264	92	puppo	None	0.9
8265	100	puppo	None	1.1

```
8266              595        puppo   None   0.6
8267               77        puppo   None   0.8
8268              146        puppo   None   1.0
8269               96        puppo   None   0.9
8270              368        puppo   None   0.3
8271               71        puppo   None   0.5
8272               82        puppo   None   1.1
8273               37        puppo   None   1.0
8274             6871        puppo   None   0.1
8275               16        puppo   None   1.1
8276               73        puppo   None   0.8
8277               79        puppo   None   0.9
8278               47        puppo   None   0.6
8279              174        puppo   None   1.0
8280               67        puppo   None   0.9
8281              232        puppo   None   1.0
8282               61        puppo   None   0.8
8283              146        puppo   None   0.9
8284              261        puppo   None   1.0
8285              879        puppo   None   0.2
8286               60        puppo   None   1.0
8287               41        puppo   None   0.5
8288              147        puppo   None   0.6
8289               47        puppo   None   0.9
8290               48        puppo   None   0.7
8291              532        puppo   None   0.8

[8292 rows x 22 columns]
```

HTML tags in source column

Define   Remove HTML tags from the begin and the end of the data using strip method

Code

```
In [40]: general_clean.source = general_clean.source.str.strip('< =":/.abcdefghijklmnopqrstuvwxy
```

```
In [41]: general_clean.source =general_clean.source.str.strip('</a>')
```

Test

```
In [42]: general_clean.head()

Out[42]:             tweet_id            timestamp              source  \
         0  892420643555336193  2017-08-01 16:23:56  Twitter for iPhone
         1  892177421306343426  2017-08-01 00:17:27  Twitter for iPhone
```

```
2  891815181378084864 2017-07-31 00:18:03  Twitter for iPhone
3  891689557279858688 2017-07-30 15:58:51  Twitter for iPhone
4  891327558926688256 2017-07-29 16:00:24  Twitter for iPhone

                                                      text  \
0  This is Phineas. He's a mystical boy. Only eve...
1  This is Tilly. She's just checking pup on you...
2  This is Archie. He is a rare Norwegian Pouncin...
3  This is Darla. She commenced a snooze mid meal...
4  This is Franklin. He would like you to stop ca...

                                       expanded_urls       name  \
0  https://twitter.com/dog_rates/status/892420643...    Phineas
1  https://twitter.com/dog_rates/status/892177421...      Tilly
2  https://twitter.com/dog_rates/status/891815181...     Archie
3  https://twitter.com/dog_rates/status/891689557...      Darla
4  https://twitter.com/dog_rates/status/891327558...   Franklin

                                        jpg_url  img_num          p1  \
0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg        1      orange
1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg        1   Chihuahua
2  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg        1   Chihuahua
3  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg        1  paper_towel
4  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg        2      basset

      p1_conf  ...     p2_conf p2_dog                          p3   p3_conf  \
0   0.097049  ...    0.085851  False                      banana  0.076110
1   0.323581  ...    0.090647   True                    papillon  0.068957
2   0.716012  ...    0.078253   True                      kelpie  0.031379
3   0.170278  ...    0.168086   True                     spatula  0.040836
4   0.555712  ...    0.225770   True  German_short-haired_pointer  0.175219

   p3_dog  favorite_count  retweet_count  stage_of_dog  value rating
0   False           39467           8853         doggo   None    1.3
1    True           33819           6514         doggo   None    1.3
2    True           25461           4328         doggo   None    1.2
3   False           42908           8964         doggo   None    1.3
4    True           41048           9774         doggo   None    1.2

[5 rows x 22 columns]
```

In [43]: general_clean.source.value_counts()

Out[43]: Twitter for iPhone   8128
         Twitter Web Client   120
         TweetDeck   44
         Name: source, dtype: int64

Invalid names in name column

Define   Change the invalid names to the None

Code

```
In [44]: def name(n):
             return general_clean.name == n

In [45]: general_clean.loc[(general_clean['name'].str.len() < 2) |
                           (name('an')) |
                           (name('by')) |
                           (name('my')) |
                           (name('Al')) |
                           (name('JD')) |
                           (name('Mo')) |
                           (name('the')), 'name'] = "None"
```

Test

```
In [46]: general_clean.name.value_counts()

Out[46]: None        2608
         Charlie       44
         Tucker        40
         Oliver        40
         Penny         40
         Lucy          40
         Cooper        40
         Lola          32
         Bo            32
         Winston       32
         Sadie         32
         Toby          28
         Daisy         28
         Koda          24
         Dave          24
         Bella         24
         Milo          24
         Rusty         24
         Stanley       24
         Scout         24
         Jax           24
         Bailey        24
         Louis         20
         Oscar         20
         Buddy         20
         Chester       20
```

```
Alfie         20
Leo           20
Larry         20
Sunny         16
              ...
Darby          4
Rupert         4
Edmund         4
Trip           4
Josep          4
Maks           4
Tupawc         4
Rey            4
Richie         4
Severus        4
Bungalo        4
Evy            4
Mauve          4
Jaycob         4
Boots          4
Maya           4
Buckley        4
Jackie         4
Eugene         4
Shikha         4
Cedrick        4
Ridley         4
Keet           4
Rinna          4
Julius         4
Spencer        4
Mary           4
Edd            4
Pawnd          4
Jarod          4
Name: name, Length: 927, dtype: int64
```

None instead of Null in some columns

Define   Change None to the NaN

Code

```
In [47]: general_clean.loc[(general_clean.value == 'None'), 'value'] = 'NaN'
```

Test

```
In [48]: general_clean.value
```

```
Out[48]: 0          NaN
         1          NaN
         2          NaN
         3          NaN
         4          NaN
         5          NaN
         6          NaN
         7          NaN
         8          NaN
         9        doggo
         10         NaN
         11         NaN
         12         NaN
         13         NaN
         14         NaN
         15         NaN
         16         NaN
         17         NaN
         18         NaN
         19         NaN
         20         NaN
         21         NaN
         22         NaN
         23         NaN
         24         NaN
         25         NaN
         26         NaN
         27         NaN
         28         NaN
         29         NaN
                   ...
         8262       NaN
         8263       NaN
         8264       NaN
         8265       NaN
         8266       NaN
         8267       NaN
         8268       NaN
         8269       NaN
         8270       NaN
         8271       NaN
         8272       NaN
         8273       NaN
         8274       NaN
         8275       NaN
```

```
8276       NaN
8277       NaN
8278       NaN
8279       NaN
8280       NaN
8281       NaN
8282       NaN
8283       NaN
8284       NaN
8285       NaN
8286       NaN
8287       NaN
8288       NaN
8289       NaN
8290       NaN
8291       NaN
Name: value, Length: 8292, dtype: object
```

1.3.4   Tidiness II

Three columns of prediction, prediction confidence and prediction dog

Define   Create new lists: prediction and confidence. If p1_dog, p2_dog or p3_dog(or all) is True then append data from p1, p1_conf, p2, p2_conf, p3, p3_conf columns to the lists that we created. After add this lists to our DataFrame.

Code

```
In [49]: general_clean.iloc[:, 8:17].head(4)

Out[49]:                  p1    p1_conf  p1_dog                   p2   p2_conf  p2_dog  \
         0        orange  0.097049   False                bagel  0.085851   False
         1     Chihuahua  0.323581    True             Pekinese  0.090647    True
         2     Chihuahua  0.716012    True              malamute  0.078253    True
         3   paper_towel  0.170278   False  Labrador_retriever  0.168086    True


                 p3    p3_conf   p3_dog
         0    banana  0.076110    False
         1  papillon  0.068957     True
         2    kelpie  0.031379     True
         3   spatula  0.040836    False

In [50]: prediction = []
         confidence = []
```
```

```
        for i in range(len(general_clean.p1_dog)):
            if general_clean.p1_dog[i] == True:
                prediction.append(general_clean.p1[i])
                confidence.append(general_clean.p1_conf[i])
            elif general_clean.p2_dog[i] == True:
                prediction.append(general_clean.p2[i])
                confidence.append(general_clean.p2_conf[i])
            elif general_clean.p3_dog[i] == True:
                prediction.append(general_clean.p3[i])
                confidence.append(general_clean.p3_conf[i])
            else:
                prediction.append('unknown')
                confidence.append(0.0)

        general_clean['prediction'] = prediction
        general_clean['confidence'] = confidence

In [51]: general_clean.drop(['p1', 'p1_conf', 'p1_dog', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_con
         'p3_dog'], axis=1, inplace=True)
```

Test

```
In [52]: general_clean

Out[52]:          tweet_id            timestamp               source  \
         0    892420643555336193  2017-08-01 16:23:56  Twitter for iPhone
         1    892177421306343426  2017-08-01 00:17:27  Twitter for iPhone
         2    891815181378084864  2017-07-31 00:18:03  Twitter for iPhone
         3    891689557279858688  2017-07-30 15:58:51  Twitter for iPhone
         4    891327558926688256  2017-07-29 16:00:24  Twitter for iPhone
         5    891087950875897856  2017-07-29 00:08:17  Twitter for iPhone
         6    890971913173991426  2017-07-28 16:27:12  Twitter for iPhone
         7    890729181411237888  2017-07-28 00:22:40  Twitter for iPhone
         8    890609185150312448  2017-07-27 16:25:51  Twitter for iPhone
         9    890240255349198849  2017-07-26 15:59:51  Twitter for iPhone
         10   890006608113172480  2017-07-26 00:31:25  Twitter for iPhone
         11   889880896479866881  2017-07-25 16:11:53  Twitter for iPhone
         12   889665388333682689  2017-07-25 01:55:32  Twitter for iPhone
         13   889638837579907072  2017-07-25 00:10:02  Twitter for iPhone
         14   889531135344209921  2017-07-24 17:02:04  Twitter for iPhone
         15   889278841981685760  2017-07-24 00:19:32  Twitter for iPhone
         16   888917238123831296  2017-07-23 00:22:39  Twitter for iPhone
         17   888804989199671297  2017-07-22 16:56:37  Twitter for iPhone
         18   888554962724278272  2017-07-22 00:23:06  Twitter for iPhone
         19   888078434458587136  2017-07-20 16:49:33  Twitter for iPhone
         20   887705289381826560  2017-07-19 16:06:48  Twitter for iPhone
         21   887517139158093824  2017-07-19 03:39:09  Twitter for iPhone
         22   887473957103951883  2017-07-19 00:47:34  Twitter for iPhone
```

```
23    887343217045368832  2017-07-18  16:08:03  Twitter for iPhone
24    887101392804085760  2017-07-18  00:07:08  Twitter for iPhone
25    886983233522544640  2017-07-17  16:17:36  Twitter for iPhone
26    886736880519319552  2017-07-16  23:58:41  Twitter for iPhone
27    886680336477933568  2017-07-16  20:14:00  Twitter for iPhone
28    886366144734445568  2017-07-15  23:25:31  Twitter for iPhone
29    886258384151887873  2017-07-15  16:17:19  Twitter for iPhone
...                 ...         ...        ...                 ...
8262  666411507551481857  2015-11-17  00:24:19  Twitter for iPhone
8263  666407126856765440  2015-11-17  00:06:54  Twitter for iPhone
8264  666396247373291520  2015-11-16  23:23:41  Twitter for iPhone
8265  666373753744588802  2015-11-16  21:54:18  Twitter for iPhone
8266  666362758909284353  2015-11-16  21:10:36  Twitter for iPhone
8267  666353288456101888  2015-11-16  20:32:58  Twitter for iPhone
8268  666345417576210432  2015-11-16  20:01:42  Twitter for iPhone
8269  666337882303524864  2015-11-16  19:31:45  Twitter for iPhone
8270  666293911632134144  2015-11-16  16:37:02  Twitter for iPhone
8271  666287406224695296  2015-11-16  16:11:11  Twitter for iPhone
8272  666273097616637952  2015-11-16  15:14:19  Twitter for iPhone
8273  666268910803644416  2015-11-16  14:57:41  Twitter for iPhone
8274  666104133288665088  2015-11-16  04:02:55  Twitter for iPhone
8275  666102155909144576  2015-11-16  03:55:04  Twitter for iPhone
8276  666099513787052032  2015-11-16  03:44:34  Twitter for iPhone
8277  666094000022159362  2015-11-16  03:22:39  Twitter for iPhone
8278  666082916733198337  2015-11-16  02:38:37  Twitter for iPhone
8279  666073100786774016  2015-11-16  01:59:36  Twitter for iPhone
8280  666071193221509120  2015-11-16  01:52:02  Twitter for iPhone
8281  666063827256086533  2015-11-16  01:22:45  Twitter for iPhone
8282  666058600524156928  2015-11-16  01:01:59  Twitter for iPhone
8283  666057090499244032  2015-11-16  00:55:59  Twitter for iPhone
8284  666055525042405380  2015-11-16  00:49:46  Twitter for iPhone
8285  666051853826850816  2015-11-16  00:35:11  Twitter for iPhone
8286  666050758794694657  2015-11-16  00:30:50  Twitter for iPhone
8287  666049248165822465  2015-11-16  00:24:50  Twitter for iPhone
8288  666044226329800704  2015-11-16  00:04:52  Twitter for iPhone
8289  666033412701032449  2015-11-15  23:21:54  Twitter for iPhone
8290  666029285002620928  2015-11-15  23:05:30  Twitter for iPhone
8291  666020888022790149  2015-11-15  22:32:08  Twitter for iPhone

                                                          text  \
0     This is Phineas. He's a mystical boy. Only eve...
1     This is Tilly. She's just checking pup on you...
2     This is Archie. He is a rare Norwegian Pouncin...
3     This is Darla. She commenced a snooze mid meal...
4     This is Franklin. He would like you to stop ca...
5     Here we have a majestic great white breaching ...
6     Meet Jax. He enjoys ice cream so much he gets ...
7     When you watch your owner call another dog a g...
```

```
8      This is Zoey. She doesn't want to be one of th...
9      This is Cassie. She is a college pup. Studying...
10     This is Koda. He is a South Australian decksha...
11     This is Bruno. He is a service shark. Only get...
12     Here's a puppo that seems to be on the fence a...
13     This is Ted. He does his best. Sometimes that'...
14     This is Stuart. He's sporting his favorite fan...
15     This is Oliver. You're witnessing one of his m...
16     This is Jim. He found a fren. Taught him how t...
17     This is Zeke. He has a new stick. Very proud o...
18     This is Ralphus. He's powering up. Attempting ...
19     This is Gerald. He was just told he didn't get...
20     This is Jeffrey. He has a monopoly on the pool...
21     I've yet to rate a Venezuelan Hover Wiener. Th...
22     This is Canela. She attempted some fancy porch...
23     You may not have known you needed to see this ...
24     This... is a Jubilant Antarctic House Bear. We...
25     This is Maya. She's very shy. Rarely leaves he...
26     This is Mingus. He's a wonderful father to his...
27     This is Derek. He's late for a dog meeting. 13...
28     This is Roscoe. Another pupper fallen victim t...
29     This is Waffles. His doggles are pupside down...
...                                               ...
8262   This is quite the dog. Gets really excited whe...
8263   This is a southern Vesuvius bumblegruff. Can d...
8264   Oh goodness. A super rare northeast Qdoba kang...
8265   Those are sunglasses and a jean jacket. 11/10 ...
8266   Unique dog here. Very small. Lives in containe...
8267   Here we have a mixed Asiago from the Galápagos...
8268   Look at this jokester thinking seat belt laws ...
8269   This is an extremely rare horned Parthenon. No...
8270   This is a funny dog. Weird toes. Won't come do...
8271   This is an Albanian 3 1/2 legged  Episcopalian...
8272      Can take selfies 11/10 https://t.co/ws2AMaNwPW
8273   Very concerned about fellow dog trapped in com...
8274   Not familiar with this breed. No tail (weird)...
8275   Oh my. Here you are seeing an Adobe Setter giv...
8276   Can stand on stump for what seems like a while...
8277   This appears to be a Mongolian Presbyterian mi...
8278   Here we have a well-established sunblockerspan...
8279   Let's hope this flight isn't Malaysian (lol). ...
8280   Here we have a northern speckled Rhododendron...
8281   This is the happiest dog you will ever see. Ve...
8282   Here is the Rand Paul of retrievers folks! He'...
8283   My oh my. This is a rare blond Canadian terrie...
8284   Here is a Siberian heavily armored polar bear ...
8285   This is an odd dog. Hard on the outside but lo...
8286   This is a truly beautiful English Wilson Staff...
```

```
8287   Here we have a 1949 1st generation vulpix. Enj...
8288   This is a purebred Piers Morgan. Loves to Netf...
8289   Here is a very happy pup. Big fan of well-main...
8290   This is a western brown Mitsubishi terrier. Up...
8291   Here we have a Japanese Irish Setter. Lost eye...

                                      expanded_urls       name  \
0      https://twitter.com/dog_rates/status/892420643...    Phineas
1      https://twitter.com/dog_rates/status/892177421...      Tilly
2      https://twitter.com/dog_rates/status/891815181...     Archie
3      https://twitter.com/dog_rates/status/891689557...      Darla
4      https://twitter.com/dog_rates/status/891327558...   Franklin
5      https://twitter.com/dog_rates/status/891087950...       None
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...        Jax
7      https://twitter.com/dog_rates/status/890729181...       None
8      https://twitter.com/dog_rates/status/890609185...       Zoey
9      https://twitter.com/dog_rates/status/890240255...     Cassie
10     https://twitter.com/dog_rates/status/890006608...       Koda
11     https://twitter.com/dog_rates/status/889880896...      Bruno
12     https://twitter.com/dog_rates/status/889665388...       None
13     https://twitter.com/dog_rates/status/889638837...        Ted
14     https://twitter.com/dog_rates/status/889531135...     Stuart
15     https://twitter.com/dog_rates/status/889278841...     Oliver
16     https://twitter.com/dog_rates/status/888917238...        Jim
17     https://twitter.com/dog_rates/status/888804989...       Zeke
18     https://twitter.com/dog_rates/status/888554962...    Ralphus
19     https://twitter.com/dog_rates/status/888078434...     Gerald
20     https://twitter.com/dog_rates/status/887705289...    Jeffrey
21     https://twitter.com/dog_rates/status/887517139...       such
22     https://twitter.com/dog_rates/status/887473957...     Canela
23     https://twitter.com/dog_rates/status/887343217...       None
24     https://twitter.com/dog_rates/status/887101392...       None
25     https://twitter.com/dog_rates/status/886983233...       Maya
26     https://www.gofundme.com/mingusneedsus,https:/...     Mingus
27     https://twitter.com/dog_rates/status/886680336...      Derek
28     https://twitter.com/dog_rates/status/886366144...     Roscoe
29     https://twitter.com/dog_rates/status/886258384...    Waffles
...                                                 ...        ...
8262   https://twitter.com/dog_rates/status/666411507...      quite
8263   https://twitter.com/dog_rates/status/666407126...       None
8264   https://twitter.com/dog_rates/status/666396247...       None
8265   https://twitter.com/dog_rates/status/666373753...       None
8266   https://twitter.com/dog_rates/status/666362758...       None
8267   https://twitter.com/dog_rates/status/666353288...       None
8268   https://twitter.com/dog_rates/status/666345417...       None
8269   https://twitter.com/dog_rates/status/666337882...       None
8270   https://twitter.com/dog_rates/status/666293911...       None
8271   https://twitter.com/dog_rates/status/666287406...       None
```

```
8272    https://twitter.com/dog_rates/status/666273097...          None
8273    https://twitter.com/dog_rates/status/666268910...          None
8274    https://twitter.com/dog_rates/status/666104133...          None
8275    https://twitter.com/dog_rates/status/666102155...          None
8276    https://twitter.com/dog_rates/status/666099513...          None
8277    https://twitter.com/dog_rates/status/666094000...          None
8278    https://twitter.com/dog_rates/status/666082916...          None
8279    https://twitter.com/dog_rates/status/666073100...          None
8280    https://twitter.com/dog_rates/status/666071193...          None
8281    https://twitter.com/dog_rates/status/666063827...          None
8282    https://twitter.com/dog_rates/status/666058600...          None
8283    https://twitter.com/dog_rates/status/666057090...          None
8284    https://twitter.com/dog_rates/status/666055525...          None
8285    https://twitter.com/dog_rates/status/666051853...          None
8286    https://twitter.com/dog_rates/status/666050758...          None
8287    https://twitter.com/dog_rates/status/666049248...          None
8288    https://twitter.com/dog_rates/status/666044226...          None
8289    https://twitter.com/dog_rates/status/666033412...          None
8290    https://twitter.com/dog_rates/status/666029285...          None
8291    https://twitter.com/dog_rates/status/666020888...          None

                                            jpg_url  img_num  \
0         https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg        1
1         https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg        1
2         https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg        1
3         https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg        1
4         https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg        2
5         https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg        1
6         https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg        1
7         https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg        2
8         https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg        1
9         https://pbs.twimg.com/media/DFrEyVuWOAAO3t9.jpg        1
10        https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg        1
11        https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg        1
12        https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg        1
13        https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg        1
14        https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg        1
15    https://pbs.twimg.com/ext_tw_video_thumb/88927...        1
16        https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg        1
17        https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg        1
18        https://pbs.twimg.com/media/DFTH_O-UQAACu2O.jpg        3
19        https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg        1
20        https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg        1
21    https://pbs.twimg.com/ext_tw_video_thumb/88751...        1
22        https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg        2
23    https://pbs.twimg.com/ext_tw_video_thumb/88734...        1
24        https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg        1
25        https://pbs.twimg.com/media/DE8yicJWOAAAvBJ.jpg        2
```

```
26      https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg          1
27      https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg          1
28      https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg          1
29      https://pbs.twimg.com/media/DEyfTG4UMAE4aE9.jpg          1
...                                                      ...     ...
8262    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg          1
8263    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg          1
8264    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg          1
8265    https://pbs.twimg.com/media/CT9vZEYWUAAlZ05.jpg          1
8266    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg          1
8267    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg          1
8268    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg          1
8269    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg          1
8270    https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg          1
8271    https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg          1
8272    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg          1
8273    https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg          1
8274    https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg          1
8275    https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg          1
8276    https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg          1
8277    https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg          1
8278    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg          1
8279    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg          1
8280    https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg          1
8281    https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg          1
8282    https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg          1
8283    https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg          1
8284    https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg          1
8285    https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg          1
8286    https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg          1
8287    https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg          1
8288    https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg          1
8289    https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg          1
8290    https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg          1
8291    https://pbs.twimg.com/media/CT4udnOWwAAOaMy.jpg          1

      favorite_count   retweet_count   stage_of_dog   value   rating  \
0             39467            8853          doggo     NaN      1.3
1             33819            6514          doggo     NaN      1.3
2             25461            4328          doggo     NaN      1.2
3             42908            8964          doggo     NaN      1.3
4             41048            9774          doggo     NaN      1.2
5             20562            3261          doggo     NaN      1.3
6             12041            2158          doggo     NaN      1.3
7             56848           16716          doggo     NaN      1.3
8             28226            4429          doggo     NaN      1.3
9             32467            7711          doggo   doggo      1.4
10            31166            7624          doggo     NaN      1.3
```

11	28268	5156	doggo	NaN	1.3
12	38818	8538	doggo	NaN	1.3
13	27672	4735	doggo	NaN	1.2
14	15359	2321	doggo	NaN	1.3
15	25652	5637	doggo	NaN	1.3
16	29611	4709	doggo	NaN	1.2
17	26080	4559	doggo	NaN	1.3
18	20290	3732	doggo	NaN	1.3
19	22201	3653	doggo	NaN	1.2
20	30779	5609	doggo	NaN	1.3
21	46959	12082	doggo	NaN	1.4
22	69871	18781	doggo	NaN	1.3
23	34222	10737	doggo	NaN	1.3
24	31061	6167	doggo	NaN	1.2
25	35859	8084	doggo	NaN	1.3
26	12306	3443	doggo	NaN	1.3
27	22798	4610	doggo	NaN	1.3
28	21524	3316	doggo	NaN	1.2
29	28469	6523	doggo	NaN	1.3
...	...	...	...	...	...
8262	459	339	puppo	NaN	0.2
8263	113	44	puppo	NaN	0.7
8264	172	92	puppo	NaN	0.9
8265	194	100	puppo	NaN	1.1
8266	804	595	puppo	NaN	0.6
8267	229	77	puppo	NaN	0.8
8268	307	146	puppo	NaN	1.0
8269	204	96	puppo	NaN	0.9
8270	522	368	puppo	NaN	0.3
8271	152	71	puppo	NaN	0.5
8272	184	82	puppo	NaN	1.1
8273	108	37	puppo	NaN	1.0
8274	14765	6871	puppo	NaN	0.1
8275	81	16	puppo	NaN	1.1
8276	164	73	puppo	NaN	0.8
8277	169	79	puppo	NaN	0.9
8278	121	47	puppo	NaN	0.6
8279	335	174	puppo	NaN	1.0
8280	154	67	puppo	NaN	0.9
8281	496	232	puppo	NaN	1.0
8282	115	61	puppo	NaN	0.8
8283	304	146	puppo	NaN	0.9
8284	448	261	puppo	NaN	1.0
8285	1253	879	puppo	NaN	0.2
8286	136	60	puppo	NaN	1.0
8287	111	41	puppo	NaN	0.5
8288	311	147	puppo	NaN	0.6
8289	128	47	puppo	NaN	0.9

8290	132	48	puppo	NaN	0.7
8291	2535	532	puppo	NaN	0.8

	prediction	confidence
0	unknown	0.000000
1	Chihuahua	0.323581
2	Chihuahua	0.716012
3	Labrador_retriever	0.168086
4	basset	0.555712
5	Chesapeake_Bay_retriever	0.425595
6	Appenzeller	0.341703
7	Pomeranian	0.566142
8	Irish_terrier	0.487574
9	Pembroke	0.511319
10	Samoyed	0.957979
11	French_bulldog	0.377417
12	Pembroke	0.966327
13	French_bulldog	0.991650
14	golden_retriever	0.953442
15	whippet	0.626152
16	golden_retriever	0.714719
17	golden_retriever	0.469760
18	Siberian_husky	0.700377
19	French_bulldog	0.995026
20	basset	0.821664
21	unknown	0.000000
22	Pembroke	0.809197
23	Mexican_hairless	0.330741
24	Samoyed	0.733942
25	Chihuahua	0.793469
26	kuvasz	0.309706
27	unknown	0.000000
28	French_bulldog	0.999201
29	pug	0.943575
...	...	...
8262	unknown	0.000000
8263	black-and-tan_coonhound	0.529139
8264	Chihuahua	0.978108
8265	soft-coated_wheaten_terrier	0.326467
8266	unknown	0.000000
8267	malamute	0.336874
8268	golden_retriever	0.858744
8269	Newfoundland	0.278407
8270	unknown	0.000000
8271	Maltese_dog	0.857531
8272	Italian_greyhound	0.176053
8273	unknown	0.000000
8274	unknown	0.000000

```
8275              English_setter    0.298617
8276                       Lhasa    0.582330
8277                   bloodhound    0.195217
8278                          pug    0.489814
8279                 Walker_hound    0.260857
8280                 Gordon_setter    0.503672
8281             golden_retriever    0.775930
8282             miniature_poodle    0.201493
8283             golden_retriever    0.007959
8284                         chow    0.692517
8285                      unknown    0.000000
8286         Bernese_mountain_dog    0.651137
8287            miniature_pinscher    0.560311
8288           Rhodesian_ridgeback    0.408143
8289               German_shepherd    0.596461
8290                      redbone    0.506826
8291        Welsh_springer_spaniel    0.465074

[8292 rows x 15 columns]
```

1.3.5 Quality III

Some breeds with a lowercase and with _ sign

Define   Replace _ sign with space and change the capital letters to uppercase letter

Code

```
In [53]: general_clean.prediction = general_clean.prediction.str.replace('_', ' ').str.capitaliz
```

Test

```
In [54]: general_clean.prediction
```

```
Out[54]: 0                          Unknown
         1                        Chihuahua
         2                        Chihuahua
         3                Labrador retriever
         4                           Basset
         5         Chesapeake bay retriever
         6                       Appenzeller
         7                        Pomeranian
         8                    Irish terrier
         9                         Pembroke
```

```
10                          Samoyed
11                    French bulldog
12                          Pembroke
13                    French bulldog
14                   Golden retriever
15                           Whippet
16                   Golden retriever
17                   Golden retriever
18                    Siberian husky
19                    French bulldog
20                            Basset
21                           Unknown
22                          Pembroke
23                   Mexican hairless
24                          Samoyed
25                         Chihuahua
26                            Kuvasz
27                           Unknown
28                    French bulldog
29                               Pug
                       ...
8262                          Unknown
8263          Black-and-tan coonhound
8264                        Chihuahua
8265      Soft-coated wheaten terrier
8266                          Unknown
8267                         Malamute
8268                 Golden retriever
8269                     Newfoundland
8270                          Unknown
8271                      Maltese dog
8272                Italian greyhound
8273                          Unknown
8274                          Unknown
8275                    English setter
8276                            Lhasa
8277                       Bloodhound
8278                              Pug
8279                     Walker hound
8280                    Gordon setter
8281                 Golden retriever
8282                 Miniature poodle
8283                 Golden retriever
8284                             Chow
8285                          Unknown
8286            Bernese mountain dog
8287               Miniature pinscher
8288              Rhodesian ridgeback
```

```
8289                German shepherd
8290                        Redbone
8291          Welsh springer spaniel
Name: prediction, Length: 8292, dtype: object
```

1.4   Storing a clean DataFrame

```
In [55]: twitter_archive_master = general_clean.copy()

In [56]: twitter_archive_master.head(2)

Out[56]:              tweet_id            timestamp              source  \
         0  892420643555336193 2017-08-01 16:23:56  Twitter for iPhone
         1  892177421306343426 2017-08-01 00:17:27  Twitter for iPhone


                                               text  \
         0  This is Phineas. He's a mystical boy. Only eve...
         1  This is Tilly. She's just checking pup on you...


                                     expanded_urls     name  \
         0  https://twitter.com/dog_rates/status/892420643...  Phineas
         1  https://twitter.com/dog_rates/status/892177421...    Tilly


                                            jpg_url  img_num  favorite_count  \
         0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg        1           39467
         1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg        1           33819


            retweet_count stage_of_dog value  rating prediction  confidence
         0           8853        doggo   NaN     1.3    Unknown    0.000000
         1           6514        doggo   NaN     1.3  Chihuahua    0.323581

In [57]: twitter_archive_master.to_csv('twitter_archive_master.csv', index=False)
```

2   Analyze and Visualization Data.

- *What are the most common names?*

- *Which breed gets the highest rating on average??*

- *What is the most common stage of the dog in a DataFrame?*

  To get the appropriate DafaFrames to answer the questions above, i used SQLite program. My SQL queries for all questions are given in each section of the relevant question below.

2.0.1 What are the most common names?

```sql
SELECT count(name) / 4 as num_of_names, name
FROM twitter_archive_master
GROUP BY name
HAVING name NOT LIKE "None"
ORDER BY num_of_names DESC
LIMIT 20
```

```python
In [58]: the_most_common_names = pd.read_csv('the_most_common_names.csv')
```
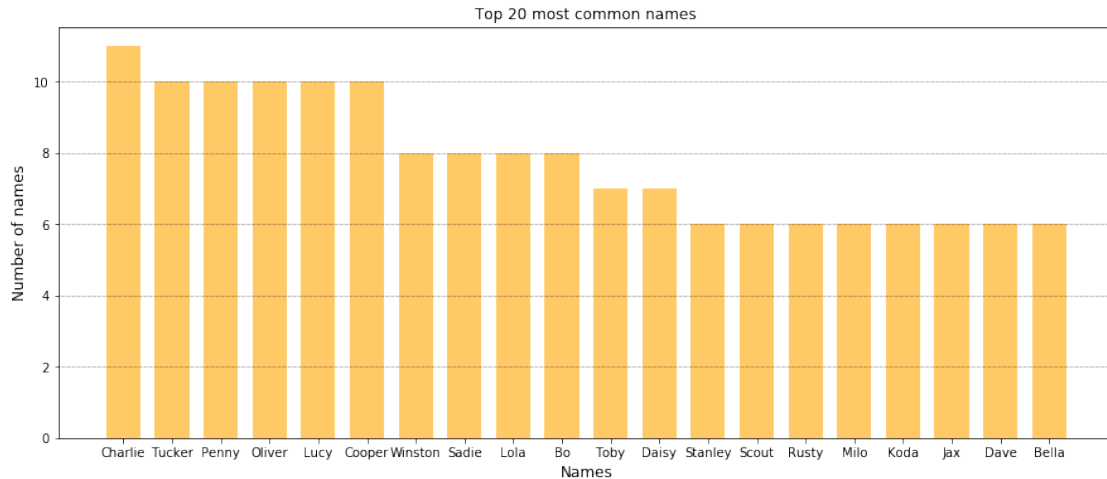
```python
In [59]: the_most_common_names
```

```
Out[59]:      num_of_names       name
         0              11  Charlie
         1              10   Tucker
         2              10    Penny
         3              10   Oliver
         4              10     Lucy
         5              10   Cooper
         6               8  Winston
         7               8    Sadie
         8               8     Lola
         9               8       Bo
         10              7     Toby
         11              7    Daisy
         12              6  Stanley
         13              6    Scout
         14              6    Rusty
         15              6     Milo
         16              6     Koda
         17              6      Jax
         18              6     Dave
         19              6    Bella
```

```python
In [60]: index = range(len(the_most_common_names.name))
         index
```

```
Out[60]: range(0, 20)
```

```python
In [61]: plt.figure(figsize = (15,6))
         plt.bar(the_most_common_names.name.sort_values(), the_most_common_names.num_of_names, w
         plt.xticks(index, the_most_common_names.name)
         plt.xlabel('Names', fontsize = 12)
         plt.ylabel('Number of names', fontsize = 12)
         plt.title('Top 20 most common names')
         plt.grid(axis = 'y', linestyle=':', linewidth=0.5, color='black');
```

Top 20 most common names

_____

2.0.2 Which breed gets the highest rating on average??

```
SELECT sum(rating) / count(prediction) as avg_rating, prediction, count(prediction) / 4 as numbe
FROM twitter_archive_master
GROUP BY prediction
HAVING avg_rating NOT LIKE 1.58699690402476 AND number_of_appearances > 1
ORDER BY avg_rating DESC
LIMIT 15
```

I have pulled data in which breeds are found and evaluated more than once.

```
In [62]: avg_rating = pd.read_csv('avg_rating_prediction.csv')

In [63]: avg_rating

Out[63]:      avg_rating                 prediction  number_of_appearances
         0     1.250000                      Saluki                      4
         1     1.247619                   Pomeranian                     42
         2     1.240000               Tibetan mastiff                     5
         3     1.233333                      Briard                      3
         4     1.214286               Border terrier                      7
         5     1.183333                 Irish setter                      6
         6     1.175000                Gordon setter                      4
         7     1.171739                     Samoyed                     46
         8     1.168478              Golden retriever                   173
         9     1.163158                      Kuvasz                     19
         10    1.150000               Giant schnauzer                     4
         11    1.150000        Wire-haired fox terrier                     2
         12    1.150000             Australian terrier                     2
         13    1.147059                Cocker spaniel                     34
         14    1.146667                Great pyrenees                     15
```
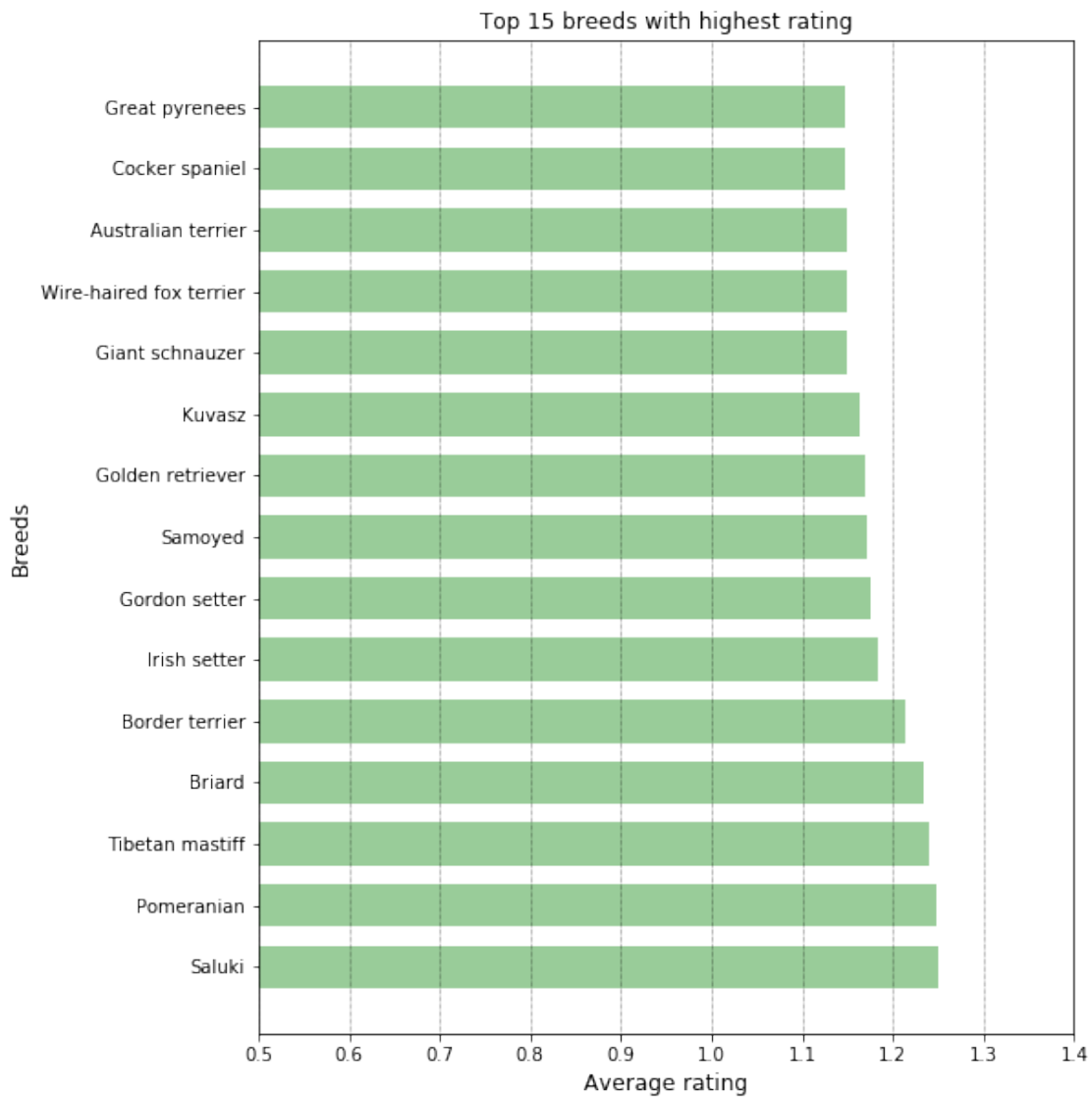
80

```
In [64]: index2 = range(len(avg_rating.prediction))
         index2
```

```
Out[64]: range(0, 15)
```

```
In [65]: plt.figure(figsize = (8, 10))
         plt.barh(avg_rating.prediction.sort_values(), avg_rating.avg_rating, height = 0.7, colo
         plt.yticks(index2, avg_rating.prediction)
         plt.xlabel('Average rating', fontsize = 12)
         plt.ylabel('Breeds', fontsize = 12)
         plt.title('Top 15 breeds with highest rating')
         plt.xlim(0.5, 1.4)
         plt.grid(axis = 'x', linestyle=':', linewidth=0.5, color='black');
```

2.0.3 What is the most common stage of the dog in a DataFrame?

```sql
SELECT count(value) as num_of_appearances, value as stage_of_dogs
FROM twitter_archive_master
GROUP BY stage_of_dogs
HAVING stage_of_dogs NOT LIKE "NaN"
```

```
In [66]: num_app_dog_stages = pd.read_csv('share_of_stages_of_dogs.csv')

In [67]: num_app_dog_stages

Out[67]:    num_of_appearances stage_of_dogs
         0                  80         doggo
         1                   8       floofer
         2                 221        pupper
         3                  24         puppo

In [68]: prop_of_dog_stages = num_app_dog_stages.copy()

In [69]: prop_of_dog_stages.num_of_appearances = prop_of_dog_stages.num_of_appearances / 333

In [70]: prop_of_dog_stages.rename(columns = {'num_of_appearances': 'propotion'}, inplace = True

In [71]: prop_of_dog_stages

Out[71]:    propotion stage_of_dogs
         0   0.240240         doggo
         1   0.024024       floofer
         2   0.663664        pupper
         3   0.072072         puppo

In [72]: propotions = prop_of_dog_stages.propotion
         labels=prop_of_dog_stages.stage_of_dogs
         explode = [0.08, 0.08, 0.1, 0.08]
         radius = 1.1
         colors = ['#ABDED7', '#FFC98B', '#8BBA48', '#DE5B6D']

         plt.pie(propotions, labels = labels, explode = explode,
                 startangle = 290, autopct = '%1.2f%%', radius = radius,
                 colors = colors, shadow = True,
                 wedgeprops = {'edgecolor' : 'black'});


         plt.title('Percentage of dog stages among all dogs in DataFrame');
         plt.tight_layout();
```

Percentage of dog stages among all dogs in DataFrame