wrangle

June 16, 2020

1 Gathering Data

Import packages and assess datasets.

```
In [1]: import pandas as pd
        import numpy as np
        import requests as r
        import os
        import tweepy as tp
        import json
        %matplotlib inline
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [2]: #read "twitter-archive-enhanced.csv" dataset.
        dfTwitter = pd.read_csv('twitter-archive-enhanced.csv')
        dfTwitter
```

```
Out[2]:                 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  666104133288665088                    NaN                  NaN
2339  666102155909144576                    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
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 [3]: #download file from internet and read "image-predictions.tsv".
        file_path = r'https://raw.githubusercontent.com/udacity/new-dand-advanced-china/master/%
        response = r.get(file_path)
        with open(file_path.split('/')[-1],mode='wb') as file:
            file.write(response.content)
        dfImage_Pred = pd.read_csv('image-predictions.tsv',sep='\t')
        dfImage_Pred
```

```
Out[3]:                 tweet_id                                            jpg_url  \
0     666020888022790149   https://pbs.twimg.com/media/CT4udnOWwAAOaMy.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_3WEAAlOoZ.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/CT56LSZWoAAlJj2.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/CT9vZEYWUAAlZ05.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/DE0BTnQUwAApKEH.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/DE8yicJW0AAAvBJ.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_O-UQAACu20.jpg
2057  888804989199671297   https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296   https://pbs.twimg.com/media/DFYRgsOUQAARGhO.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/DFrEyVuW0AAO3t9.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 [4]: #Download from Twitter.

        #consumer_key = 'YOUR CONSUMER KEY'
        #consumer_secret = 'YOUR CONSUMER SECRET'
        #access_token = 'YOUR ACCESS TOKEN'
        #access_secret = 'YOUR ACCESS SECRET'

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

        #api = tp.API(auth)

        #Print other users' contents from timeline
        #public_tweets = api.user_timeline('WeRateDogs')

        #for tweet in public_tweets:
        #    print(tweet.text)
```

```
In [ ]:
```

Becasue of the contrain in the region, I am not able to use Twitter, so I read the json file directly.

```
In [5]: import zipfile
        with open('tweet-json.zip','rb') as f:
            z_tweets = zipfile.ZipFile(f)
            z_tweets.extractall()

        # check for the extracted file
        z_tweets.namelist()

Out[5]: ['tweet-json copy']
```

```
In [6]: # read the file in DataFrame
        with open('tweet-json copy', 'r') as f:
            dfTweet_json = pd.read_json(f, lines= True, encoding = 'utf-8')
        # check the data
        dfTweet_json
        # select the columns of interest : 'id', 'favorite_count','retweet_count'
        dfTweet_json = dfTweet_json.loc[:,['id','favorite_count','retweet_count']]

        #rename column id as tweet_id
        dfTweet_json = dfTweet_json.rename(columns = {"id": "tweet_id"})
        dfTweet_json.head()
```

```
Out[6]:              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
```

2 Assessing Data

2.1 Detect quality issues and tidiness issues in "dfTwitter" dataset.

```
In [7]: dfTwitter.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]: #check rating_denominator
        dfTwitter['rating_denominator'].value_counts()
```

```
Out[8]: 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
        170       1
        7         1
        0         1
        Name: rating_denominator, dtype: int64
```

2.1.1 1. The values in "rating_denominator" are not all integral tens digit, which also inclues some other integers, such as 11, 2, 16, 15, and 7.

```
In [9]: #check names
        dfTwitter['name'].value_counts().head(20)
```

```
Out[9]: None       745
        a           55
        Charlie     12
        Lucy        11
        Oliver      11
        Cooper      11
        Penny       10
        Tucker      10
        Lola        10
        Bo           9
        Winston      9
        the          8
        Sadie        8
        Toby         7
        Bailey       7
        Daisy        7
        Buddy        7
        an           7
        Dave         6
        Bella        6
        Name: name, dtype: int64
```

2.1.2 2. In dogs’ names, there are some words with a, an, and the (articles), which is not a good way to identify the dogs’ names.

```
In [10]: (dfTwitter.iloc[:,-4:]=='None').astype(int).sum(axis=1).value_counts()

Out[10]: 4    1976
         3     366
         2      14
         dtype: int64
```

2.1.3 3. There is a mistake in dog’s rates and some dogs have more thhan two rates.

```
In [11]: #check missing in names
         (dfTwitter.loc[:,'name']=='None').astype(int).sum()

Out[11]: 745
```

2.1.4 4. There are lots of missing in dogs’ names. The data only have 745 input dog’s names.

```
In [12]: #check duplicatation
         dfTwitter['tweet_id'].duplicated().sum()

Out[12]: 0

In [13]: #Check for tweets with no image
         dfTwitter['expanded_urls'].isnull().value_counts()

Out[13]: False    2297
         True       59
         Name: expanded_urls, dtype: int64
```

2.1.5 5. There are many tweets do not have images.

2.2 Detect quality issues and tidiness issues in "dfImage_Pred" dataset.

```
In [14]: dfImage_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 [15]: #check duplications
         dfImage_Pred['jpg_url'].duplicated().sum()
```

```
Out[15]: 66
```

```
In [ ]:
```

2.3 Detect quality issues and tidiness issues in "dfTweet_json" dataset.

```
In [16]: dfTweet_json.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
```

2.4 Document quality issues and tidiness issues in these three datasets.

2.4.1 Quality Issues:

dfTweeter:

1. The datasets include infomation about retweet and forward.

2. There are missing records in "expanded_urls" variable

3. The values in "rating_denominator" are not all integral tens digit, which also inclues some other integers, such as 11, 2, 16, 15, and 7.

4. In dogs' names, there are some words with a, an, and the (articles), which is not a good way to identify the dogs' names.

5. There are lots of missing in dogs' names. The data only have 745 inputs about dog's names.

6. There are many tweets do not have images.

7. In the column " source", there are record with the formats in HTML.

dfImage_Pred:

1. There are duplicated record in the dataset.

2. In p1, p2, and p3, there is a mixed usage of upper case and lower case. The seperation of each word was not consistent.

2.4.2 Tidiness Issues:

In dfTwitterthe "rate" of dogs used four variables to measure, they are: doggo,floofer,pupper, and puppo.

The observations of these three dataframes are the same group of people, so we need to combine them into one dataframe.

```
In [ ]:
```

3 Cleaning Data

```
In [17]: # make copies of the datasets.
         dfTwitter_C = dfTwitter.copy()
         dfImage_Pred_C= dfImage_Pred.copy()
         dfTweet_json_C = dfTweet_json.copy()
```

3.1 dfTweeter:

3.1.1 1. The datasets include infomation about retweet and forward.

- Delete retweet and forward records.

```
In [18]: dfTwitter_C = dfTwitter_C[dfTwitter_C['retweeted_status_id'].isnull()]
         dfTwitter_C = dfTwitter_C[dfTwitter_C['in_reply_to_user_id'].isnull()]
```

```
In [19]: #test
         dfTwitter_C.info()
         #delete extra useless columns
         dfTwitter_C.drop(['in_reply_to_status_id','in_reply_to_user_id','retweeted_status_id','
         dfTwitter_C.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2097 non-null int64
in_reply_to_status_id         0 non-null float64
in_reply_to_user_id           0 non-null float64
timestamp                     2097 non-null object
```

```
source                        2097 non-null object
text                          2097 non-null object
retweeted_status_id           0 non-null float64
retweeted_status_user_id      0 non-null float64
retweeted_status_timestamp    0 non-null object
expanded_urls                 2094 non-null object
rating_numerator              2097 non-null int64
rating_denominator            2097 non-null int64
name                          2097 non-null object
doggo                         2097 non-null object
floofer                       2097 non-null object
pupper                        2097 non-null object
puppo                         2097 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 294.9+ KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id              2097 non-null int64
timestamp             2097 non-null object
source                2097 non-null object
text                  2097 non-null object
expanded_urls         2094 non-null object
rating_numerator      2097 non-null int64
rating_denominator    2097 non-null int64
name                  2097 non-null object
doggo                 2097 non-null object
floofer               2097 non-null object
pupper                2097 non-null object
puppo                 2097 non-null object
dtypes: int64(3), object(9)
memory usage: 213.0+ KB
```

3.1.2 2. There are missing records in "expanded_urls" variable

-Delete missing records in "expanded_urls" variable

```
In [20]: dfTwitter_C = dfTwitter_C[dfTwitter_C['expanded_urls'].notnull()]
```

```
In [21]: #test
         dfTwitter_C.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2094 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id              2094 non-null int64
timestamp             2094 non-null object
source                2094 non-null object
```

```
text                   2094 non-null object
expanded_urls          2094 non-null object
rating_numerator       2094 non-null int64
rating_denominator     2094 non-null int64
name                   2094 non-null object
doggo                  2094 non-null object
floofer                2094 non-null object
pupper                 2094 non-null object
puppo                  2094 non-null object
dtypes: int64(3), object(9)
memory usage: 212.7+ KB
```

3.1.3 3. The values in "rating_denominator" are integral tens digit, which also includes some other integers, such as 11, 2, 16, 15, and 7.

- For the data have two rating record, keep the first rate as the principle. If the cleaned data is still not in integral tens digit, change that number into the closest integral tens digit.

```
In [22]: dfTwitter_C['rating_numerator'],dfTwitter_C['rating_denominator'] = dfTwitter_C['text']


         #Change one of the record manually
         index = dfTwitter_C[dfTwitter_C['rating_numerator'].isnull()].index[0]
         dfTwitter_C.loc[index,'rating_numerator']=24
         dfTwitter_C.loc[index,'rating_denominator']=7

         #change data type as float
         dfTwitter_C['rating_numerator'] = dfTwitter_C['rating_numerator'].astype(float)
         dfTwitter_C['rating_denominator'] = dfTwitter_C['rating_denominator'].astype(float)
```

```
In [23]: #test
         dfTwitter_C['rating_denominator'].value_counts()
```

```
Out[23]: 10.0     2080
         50.0        3
         80.0        2
         150.0       1
         110.0       1
         90.0        1
         70.0        1
         170.0       1
         120.0       1
         40.0        1
         20.0        1
         7.0         1
         Name: rating_denominator, dtype: int64
```

```
In [24]: dfTwitter_C['rating_numerator'].value_counts()
```

```
Out[24]: 12.00      485
         10.00      435
         11.00      413
         13.00      287
         9.00       153
         8.00        98
         7.00        51
         14.00       39
         5.00        33
         6.00        32
         3.00        19
         4.00        16
         2.00         9
         1.00         4
         13.50        1
         0.00         1
         24.00        1
         84.00        1
         420.00       1
         1776.00      1
         80.00        1
         60.00        1
         44.00        1
         144.00       1
         88.00        1
         11.26        1
         11.27        1
         121.00       1
         9.75         1
         99.00        1
         204.00       1
         45.00        1
         165.00       1
         50.00        1
         Name: rating_numerator, dtype: int64
```

3.1.4 4. In dogs' names, there are some words with a, an, and the (articles), which is not a good way to identify the dogs' names.

- Retrive back to the original post and gussing it is recorded after "This is...". Get the dogs' names from the original records.

```
In [25]: dfTwitter_C['name'] = dfTwitter_C['text'].str.extract(r'\S*[This is|Here is|Here\'s|nam
```

```
In [26]: #test
         dfTwitter_C['name'].value_counts()
```

```
Out[26]: Oliver           11
         Charlie          11
```

```
Lucy              11
Cooper            10
Penny              9
Tucker             9
Winston            8
Lola               8
Christmas          8
Sadie              8
Toby               8
Daisy              7
Bo                 7
Bella              6
Oscar              6
Bailey             6
Stanley            6
Koda               6
Jax                6
Milo               5
Rusty              5
Louis              5
Leo                5
Dave               5
Scout              5
Chester            5
Buddy              5
Bentley            5
Boomer             5
Zoey               5
                  ..
Mike               1
Very               1
Winifred           1
Sailor             1
Kayla              1
Orion              1
East               1
Schnozz            1
Bobb               1
Hero               1
Bangladeshi        1
Ferg               1
Lili               1
Andru              1
Banjo              1
Trump              1
Aubie              1
Tyrannosaurus      1
Anakin             1
```

```
Ralphie          1
Augie            1
Cali             1
Chubbs           1
Skye             1
Mauve            1
Bruno            1
Rinna            1
Mack             1
Brat             1
Logan            1
Name: name, Length: 1063, dtype: int64
```

3.1.5 5. There are lots of missing in dogs' names. The data only have 745 input dog's names

- I am not able to add any more information here, since the users didn't provide any inputs here.

3.1.6 6. There are many tweets do not have images.

-There is no status data available from the Twitter API and not all tweets have an image. I did not confirm that all tweets with an image stored the image.

3.1.7 7. In the column " source", there are record with the formats in HTML.

- Get the url

```
In [27]: dfTwitter_C['source'] = dfTwitter_C['source'].str.extract(r'>(.+)<',expand=True)
```

```
In [28]: #test
         dfTwitter_C.head()
```

```
Out[28]:             tweet_id                  timestamp             source  \
         0  892420643555336193  2017-08-01 16:23:56 +0000  Twitter for iPhone
         1  892177421306343426  2017-08-01 00:17:27 +0000  Twitter for iPhone
         2  891815181378084864  2017-07-31 00:18:03 +0000  Twitter for iPhone
         3  891689557279858688  2017-07-30 15:58:51 +0000  Twitter for iPhone
         4  891327558926688256  2017-07-29 16:00:24 +0000  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  rating_numerator  \
         0  https://twitter.com/dog_rates/status/892420643...              13.0
         1  https://twitter.com/dog_rates/status/892177421...              13.0
```

```
2  https://twitter.com/dog_rates/status/891815181...              12.0
3  https://twitter.com/dog_rates/status/891689557...              13.0
4  https://twitter.com/dog_rates/status/891327558...              12.0

   rating_denominator      name  doggo  floofer  pupper  puppo
0                10.0   Phineas   None     None    None   None
1                10.0     Tilly   None     None    None   None
2                10.0    Archie   None     None    None   None
3                10.0     Darla   None     None    None   None
4                10.0  Franklin   None     None    None   None
```

3.2 dfImage_Pred:

3.2.1 1. There are duplicated record in the dataset.

- Remove the duplicated records

```
In [29]: dfImage_Pred_C.drop_duplicates(subset='jpg_url',inplace=True)
```

```
In [30]: #test
         dfImage_Pred_C['jpg_url'].duplicated().sum()
         #Cool! The duplication is 0 now!
```

```
Out[30]: 0
```

3.2.2 2. In p1, p2, and p3, there is a mixed usage of upper case and lower case. The seperation of each word was not consistent.

- Change all the letters in lower case and change all the seprations as underscore.

```
In [31]: dfImage_Pred_C[['p1','p2','p3']] = dfImage_Pred_C[['p1','p2','p3']].applymap(str.lower)
         dfImage_Pred_C[['p1','p2','p3']] = dfImage_Pred_C[['p1','p2','p3']].replace(' ','_').re
```

```
In [32]: #test
         dfImage_Pred_C.head(5)
```

```
Out[32]:             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

            img_num                      p1   p1_conf  p1_dog                  p2  \
         0        1  welsh_springer_spaniel  0.465074    True              collie
         1        1                 redbone  0.506826    True  miniature_pinscher
         2        1         german_shepherd  0.596461    True            malinois
         3        1     rhodesian_ridgeback  0.408143    True             redbone
         4        1      miniature_pinscher  0.560311    True          rottweiler
```

```
    p2_conf  p2_dog                   p3   p3_conf  p3_dog
0  0.156665    True    shetland_sheepdog  0.061428    True
1  0.074192    True  rhodesian_ridgeback  0.072010    True
2  0.138584    True           bloodhound  0.116197    True
3  0.360687    True   miniature_pinscher  0.222752    True
4  0.243682    True             doberman  0.154629    True
```

3.3 In dfTwitterthe "rate" of dogs used four variables to measure, they are: doggo,floofer,pupper, and puppo.

3.3.1 - Combine the columns (doggo,floofer,pupper,puppo) and create a new column called "growth". Delete doggo,floofer,pupper, and puppo.

```
In [33]: #Combine
         dfTwitter_C['growth'] = dfTwitter_C['doggo']+dfTwitter_C['floofer']+dfTwitter_C['pupper
         dfTwitter_C['growth'] = dfTwitter_C['growth'].str.replace('None','')
         dfTwitter_C = dfTwitter_C.replace(({'growth':{'':np.nan}}))

         #Create new column and delete the old four columns.
         dfTwitter_C.drop(['doggo','floofer','pupper','puppo'],axis=1,inplace=True)
         dfTwitter_C[dfTwitter_C['growth'].notnull()]
```

```
Out[33]:                tweet_id                  timestamp              source  \
         9    890240255349198849  2017-07-26 15:59:51 +0000  Twitter for iPhone
         12   889665388333682689  2017-07-25 01:55:32 +0000  Twitter for iPhone
         14   889531135344209921  2017-07-24 17:02:04 +0000  Twitter for iPhone
         29   886366144734445568  2017-07-15 23:25:31 +0000  Twitter for iPhone
         43   884162670584377345  2017-07-09 21:29:42 +0000  Twitter for iPhone
         46   883360690899218434  2017-07-07 16:22:55 +0000  Twitter for iPhone
         49   882762694511734784  2017-07-06 00:46:41 +0000  Twitter for iPhone
         56   881536004380872706  2017-07-02 15:32:16 +0000  Twitter for iPhone
         71   878776093423087618  2017-06-25 00:45:22 +0000  Twitter for iPhone
         82   876838120628539392  2017-06-19 16:24:33 +0000  Twitter for iPhone
         92   874296783580663808  2017-06-12 16:06:11 +0000  Twitter for iPhone
         94   874012996292530176  2017-06-11 21:18:31 +0000  Twitter for iPhone
         98   873213775632977920  2017-06-09 16:22:42 +0000  Twitter for iPhone
         99   872967104147763200  2017-06-09 00:02:31 +0000  Twitter for iPhone
         107  871762521631449091  2017-06-05 16:15:56 +0000  Twitter for iPhone
         108  871515927908634625  2017-06-04 23:56:03 +0000  Twitter for iPhone
         110  871102520638267392  2017-06-03 20:33:19 +0000  Twitter for iPhone
         121  869596645499047938  2017-05-30 16:49:31 +0000  Twitter for iPhone
         129  867421006826221569  2017-05-24 16:44:18 +0000  Twitter for iPhone
         135  866450705531457537  2017-05-22 00:28:40 +0000  Twitter for iPhone
         168  859607811541651456  2017-05-03 03:17:27 +0000  Twitter for iPhone
         172  858843525470990336  2017-05-01 00:40:27 +0000  Twitter for iPhone
         191  855851453814013952  2017-04-22 18:31:02 +0000  Twitter for iPhone
         199  854120357044912130  2017-04-17 23:52:16 +0000  Twitter for iPhone
         200  854010172552949760  2017-04-17 16:34:26 +0000  Twitter for iPhone
         220  850019790995546112  2017-04-06 16:18:05 +0000  Twitter for iPhone
```

```
240   846514051647705089  2017-03-28 00:07:32 +0000  Twitter for iPhone
248   845397057150107648  2017-03-24 22:08:59 +0000  Twitter for iPhone
249   845306882940190720  2017-03-24 16:10:40 +0000  Twitter for iPhone
293   837820167694528512  2017-03-04 00:21:08 +0000  Twitter for iPhone
...                  ...                        ...                 ...
1875  675113801096802304  2015-12-11 00:44:07 +0000  Twitter for iPhone
1880  675006312288268288  2015-12-10 17:37:00 +0000  Twitter for iPhone
1889  674774481756377088  2015-12-10 02:15:47 +0000  Twitter for iPhone
1897  674737130913071104  2015-12-09 23:47:22 +0000  Twitter for iPhone
1903  674638615994089473  2015-12-09 17:15:54 +0000  Twitter for iPhone
1907  674447403907457024  2015-12-09 04:36:06 +0000  Twitter for iPhone
1915  674318007229923329  2015-12-08 20:01:55 +0000  Twitter for iPhone
1921  674262580978937856  2015-12-08 16:21:41 +0000  Twitter for iPhone
1930  674038233588723717  2015-12-08 01:30:12 +0000  Twitter for iPhone
1936  673956914389192708  2015-12-07 20:07:04 +0000  Twitter for iPhone
1937  673919437611909120  2015-12-07 17:38:09 +0000  Twitter for iPhone
1945  673707060090052608  2015-12-07 03:34:14 +0000  Twitter for iPhone
1948  673697980713705472  2015-12-07 02:58:09 +0000  Twitter for iPhone
1954  673656262056419329  2015-12-07 00:12:23 +0000  Twitter for iPhone
1956  673612854080196609  2015-12-06 21:19:54 +0000  Twitter for iPhone
1960  673363615379013632  2015-12-06 04:49:31 +0000  Twitter for iPhone
1967  673342308415348736  2015-12-06 03:24:51 +0000  Twitter for iPhone
1970  673295268553605120  2015-12-06 00:17:55 +0000  Twitter for iPhone
1974  673148804208660480  2015-12-05 14:35:56 +0000  Twitter for iPhone
1977  672988786805112832  2015-12-05 04:00:04 +0000  Twitter for iPhone
1980  672975131468300288  2015-12-05 03:05:49 +0000  Twitter for iPhone
1981  672970152493887488  2015-12-05 02:46:02 +0000  Twitter for iPhone
1985  672898206762672129  2015-12-04 22:00:08 +0000  Twitter for iPhone
1991  672622327801233409  2015-12-04 03:43:54 +0000  Twitter for iPhone
1992  672614745925664768  2015-12-04 03:13:46 +0000  Twitter for iPhone
1995  672594978741354496  2015-12-04 01:55:13 +0000  Twitter for iPhone
2002  672481316919734272  2015-12-03 18:23:34 +0000  Twitter for iPhone
2009  672254177670729728  2015-12-03 03:21:00 +0000  Twitter for iPhone
2015  672205392827572224  2015-12-03 00:07:09 +0000  Twitter for iPhone
2017  672160042234327040  2015-12-02 21:06:56 +0000  Twitter for iPhone

                                                   text  \
9     This is Cassie. She is a college pup. Studying...
12    Here's a puppo that seems to be on the fence a...
14    This is Stuart. He's sporting his favorite fan...
29    This is Roscoe. Another pupper fallen victim t...
43    Meet Yogi. He doesn't have any important dog m...
46    Meet Grizzwald. He may be the floofiest floofe...
49    This is Gus. He's quite the cheeky pupper. Alr...
56    Here is a pupper approaching maximum borkdrive...
71    This is Snoopy. He's a proud #PrideMonthPuppo...
82    This is Ginger. She's having a ruff Monday. To...
92    This is Jed. He may be the fanciest pupper in ...
```

```
94    This is Sebastian. He can't see all the colors...
98    This is Sierra. She's one precious pupper. Abs...
99    Here's a very large dog. He has a date later. ...
107   This is Rover. As part of pupper protocol he h...
108   This is Napolean. He's a Raggedy East Nicaragu...
110   Never doubt a doggo 14/10 https://t.co/AbBLh2FZCH
121   This is Scout. He just graduated. Officially a...
129   This is Shikha. She just watched you drop a sk...
135   This is Jamesy. He gives a kiss to every other...
168   Sorry for the lack of posts today. I came home...
172   I have stumbled puppon a doggo painting party...
191   Here's a puppo participating in the #ScienceMa...
199   Sometimes you guys remind me just how impactfu...
200   At first I thought this was a shy doggo, but i...
220   Say hello to Boomer. He's a sandy pupper. Havi...
240   This is Barney. He's an elder doggo. Hitches a...
248   Say hello to Mimosa. She's an emotional suppor...
249   This is Pickles. She's a silly pupper. Thinks ...
293   Here's a pupper before and after being asked "...
...                                                 ...
1875  Meet Zuzu. He just graduated college. Astute p...
1880  Say hello to Mollie. This pic was taken after ...
1889  This is Superpup. His head isn't proportional ...
1897  Meet Rufio. He is unaware of the pink legless ...
1903  This pupper is fed up with being tickled. 12/1...
1907  This pupper just wants a belly rub. This puppe...
1915  This is Lennon. He's in quite the predicament...
1921  This is Gus. He's super stoked about being an ...
1930  This is Kaiya. She's an aspiring shoe model. 1...
1936  This is one esteemed pupper. Just graduated co...
1937  This is Obie. He is on guard watching for evil...
1945  This is Raymond. He's absolutely terrified of ...
1948  This is Pickles. She's a tiny pointy pupper. A...
1954  This is Albert AKA King Banana Peel. He's a ki...
1956  This is Jeffri. He's a speckled ice pupper. Ve...
1960  This little pupper can't wait for Christmas. H...
1967  This is Django. He's a skilled assassin pupper...
1970  Meet Eve. She's a raging alcoholic 8/10 (would...
1974  This is Fletcher. He's had a ruff night. No mo...
1977  This is Schnozz. He's had a blurred tail since...
1980  This is Chuckles. He is one skeptical pupper. ...
1981  This is Chet. He's having a hard time. Really ...
1985  This is Cheryl AKA Queen Pupper of the Skies. ...
1991  This lil pupper is sad because we haven't foun...
1992  This is Norman. Doesn't bark much. Very docile...
1995  Meet Scott. Just trying to catch his train to ...
2002  Say hello to Jazz. She should be on the cover ...
2009  This is Rolf. He's having the time of his life...
```

```
2015  This is Opal. He's a Royal John Coctostan. Rea...
2017  This is Bubba. He's a Titted Peebles Aorta. Ev...

                                          expanded_urls  rating_numerator  \
9     https://twitter.com/dog_rates/status/890240255...              14.0
12    https://twitter.com/dog_rates/status/889665388...              13.0
14    https://twitter.com/dog_rates/status/889531135...              13.0
29    https://twitter.com/dog_rates/status/886366144...              12.0
43    https://twitter.com/dog_rates/status/884162670...              12.0
46    https://twitter.com/dog_rates/status/883360690...              13.0
49    https://twitter.com/dog_rates/status/882762694...              12.0
56    https://twitter.com/dog_rates/status/881536004...              14.0
71    https://twitter.com/dog_rates/status/878776093...              13.0
82    https://twitter.com/dog_rates/status/876838120...              12.0
92    https://twitter.com/dog_rates/status/874296783...              13.0
94    https://twitter.com/dog_rates/status/874012996...              13.0
98    https://www.gofundme.com/help-my-baby-sierra-g...              12.0
99    https://twitter.com/dog_rates/status/872967104...              12.0
107   https://twitter.com/dog_rates/status/871762521...              12.0
108   https://twitter.com/dog_rates/status/871515927...              12.0
110   https://twitter.com/animalcog/status/871075758...              14.0
121   https://twitter.com/dog_rates/status/869596645...              12.0
129   https://twitter.com/dog_rates/status/867421006...              12.0
135   https://twitter.com/dog_rates/status/866450705...              13.0
168   https://twitter.com/dog_rates/status/859607811...              13.0
172   https://twitter.com/dog_rates/status/858843525...              13.0
191   https://twitter.com/dog_rates/status/855851453...              13.0
199   https://twitter.com/dog_rates/status/854120357...              14.0
200   https://twitter.com/dog_rates/status/854010172...              11.0
220   https://twitter.com/dog_rates/status/850019790...              12.0
240   https://twitter.com/dog_rates/status/846514051...              13.0
248   https://www.gofundme.com/help-save-a-pup,https...              13.0
249   https://twitter.com/dog_rates/status/845306882...              12.0
293   https://twitter.com/dog_rates/status/837820167...              12.0
...                                                 ...               ...
1875  https://twitter.com/dog_rates/status/675113801...              10.0
1880  https://twitter.com/dog_rates/status/675006312...              10.0
1889  https://twitter.com/dog_rates/status/674774481...              11.0
1897  https://twitter.com/dog_rates/status/674737130...              10.0
1903  https://twitter.com/dog_rates/status/674638615...              12.0
1907  https://twitter.com/dog_rates/status/674447403...              10.0
1915  https://twitter.com/dog_rates/status/674318007...               8.0
1921  https://twitter.com/dog_rates/status/674262580...               9.0
1930  https://twitter.com/dog_rates/status/674038233...              12.0
1936  https://twitter.com/dog_rates/status/673956914...              10.0
1937  https://twitter.com/dog_rates/status/673919437...              11.0
1945  https://twitter.com/dog_rates/status/673707060...              10.0
1948  https://twitter.com/dog_rates/status/673697980...               8.0
```

```
1954  https://twitter.com/dog_rates/status/673656262...              10.0
1956  https://twitter.com/dog_rates/status/673612854...               7.0
1960  https://twitter.com/dog_rates/status/673363615...              11.0
1967  https://twitter.com/dog_rates/status/673342308...              10.0
1970  https://twitter.com/dog_rates/status/673295268...               8.0
1974  https://twitter.com/dog_rates/status/673148804...               8.0
1977  https://twitter.com/dog_rates/status/672988786...              10.0
1980  https://twitter.com/dog_rates/status/672975131...              10.0
1981  https://twitter.com/dog_rates/status/672970152...               7.0
1985  https://twitter.com/dog_rates/status/672898206...              11.0
1991  https://twitter.com/dog_rates/status/672622327...              12.0
1992  https://twitter.com/dog_rates/status/672614745...               6.0
1995  https://twitter.com/dog_rates/status/672594978...               9.0
2002  https://twitter.com/dog_rates/status/672481316...              12.0
2009  https://twitter.com/dog_rates/status/672254177...              11.0
2015  https://twitter.com/dog_rates/status/672205392...               9.0
2017  https://twitter.com/dog_rates/status/672160042...               8.0

      rating_denominator       name         growth
9                   10.0     Cassie          doggo
12                  10.0        NaN          puppo
14                  10.0     Stuart          puppo
29                  10.0     Roscoe         pupper
43                  10.0       Yogi          doggo
46                  10.0  Grizzwald        floofer
49                  10.0        Gus         pupper
56                  10.0        NaN         pupper
71                  10.0     Snoopy          puppo
82                  10.0     Ginger         pupper
92                  10.0        Jed         pupper
94                  10.0  Sebastian          puppo
98                  10.0     Sierra         pupper
99                  10.0        NaN          doggo
107                 10.0      Rover         pupper
108                 10.0   Napolean          doggo
110                 10.0        NaN          doggo
121                 10.0      Scout          doggo
129                 10.0     Shikha          puppo
135                 10.0     Jamesy         pupper
168                 10.0       Zoey          puppo
172                 10.0   Pupcasso          doggo
191                 10.0        NaN     doggopuppo
199                 10.0        NaN         pupper
200                 10.0       Rare   doggofloofer
220                 10.0     Boomer         pupper
240                 10.0     Barney          doggo
248                 10.0     Mimosa          doggo
249                 10.0    Pickles         pupper
```

```
293                   10.0        NaN        pupper
...                    ...        ...           ...
1875                  10.0       Zuzu        pupper
1880                  10.0     Mollie        pupper
1889                  10.0   Superpup        pupper
1897                  10.0      Rufio        pupper
1903                  10.0        NaN        pupper
1907                  10.0        NaN        pupper
1915                  10.0     Lennon        pupper
1921                  10.0        Gus        pupper
1930                  10.0      Kaiya        pupper
1936                  10.0        NaN        pupper
1937                  10.0       Obie        pupper
1945                  10.0    Raymond        pupper
1948                  10.0    Pickles        pupper
1954                  10.0     Albert        pupper
1956                  10.0     Jeffri        pupper
1960                  10.0  Christmas        pupper
1967                  10.0     Django        pupper
1970                  10.0        Eve        pupper
1974                  10.0   Fletcher        pupper
1977                  10.0    Schnozz        pupper
1980                  10.0   Chuckles        pupper
1981                  10.0       Chet        pupper
1985                  10.0     Cheryl        pupper
1991                  10.0       Kony        pupper
1992                  10.0     Norman        pupper
1995                  10.0      Scott        pupper
2002                  10.0       Jazz        pupper
2009                  10.0       Rolf        pupper
2015                  10.0       Opal        pupper
2017                  10.0      Bubba        pupper

[335 rows x 9 columns]
```

```
In [34]: dfTwitter_C['growth'].value_counts()
```

```
Out[34]: pupper          220
         doggo            72
         puppo            23
         doggopupper       9
         floofer           9
         doggofloofer      1
         doggopuppo        1
         Name: growth, dtype: int64
```

3.4 The observations of these three dataframes are the same group of people, so we need to combine them into one dataframe.

3.4.1 Use merge to combine three dataframes.

```
In [35]: dfCombine = pd.merge(dfTwitter_C,dfImage_Pred_C,how='inner',on='tweet_id').merge(dfTwee
```

```
In [36]: #test
         dfCombine.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1971 entries, 0 to 1970
Data columns (total 22 columns):
tweet_id              1971 non-null int64
timestamp             1971 non-null object
source                1971 non-null object
text                  1971 non-null object
expanded_urls         1971 non-null object
rating_numerator      1971 non-null float64
rating_denominator    1971 non-null float64
name                  1520 non-null object
growth                303 non-null object
jpg_url               1971 non-null object
img_num               1971 non-null int64
p1                    1971 non-null object
p1_conf               1971 non-null float64
p1_dog                1971 non-null bool
p2                    1971 non-null object
p2_conf               1971 non-null float64
p2_dog                1971 non-null bool
p3                    1971 non-null object
p3_conf               1971 non-null float64
p3_dog                1971 non-null bool
favorite_count        1971 non-null int64
retweet_count         1971 non-null int64
dtypes: bool(3), float64(5), int64(4), object(10)
memory usage: 313.7+ KB
```

```
In [37]: #test
         dfCombine.head()
```

```
Out[37]:              tweet_id                  timestamp             source  \
         0  892420643555336193  2017-08-01 16:23:56 +0000  Twitter for iPhone
         1  892177421306343426  2017-08-01 00:17:27 +0000  Twitter for iPhone
         2  891815181378084864  2017-07-31 00:18:03 +0000  Twitter for iPhone
         3  891689557279858688  2017-07-30 15:58:51 +0000  Twitter for iPhone
         4  891327558926688256  2017-07-29 16:00:24 +0000  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  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...              13.0
1  https://twitter.com/dog_rates/status/892177421...              13.0
2  https://twitter.com/dog_rates/status/891815181...              12.0
3  https://twitter.com/dog_rates/status/891689557...              13.0
4  https://twitter.com/dog_rates/status/891327558...              12.0

   rating_denominator      name growth  \
0                10.0   Phineas    NaN
1                10.0     Tilly    NaN
2                10.0    Archie    NaN
3                10.0     Darla    NaN
4                10.0  Franklin    NaN

                                           jpg_url      ...         p1_conf  \
0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg      ...        0.097049
1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg      ...        0.323581
2  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg      ...        0.716012
3  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg      ...        0.170278
4  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg      ...        0.555712

  p1_dog                 p2   p2_conf p2_dog                           p3  \
0  False              bagel  0.085851  False                       banana
1   True           pekinese  0.090647   True                     papillon
2   True           malamute  0.078253   True                       kelpie
3  False  labrador_retriever  0.168086   True                      spatula
4   True    english_springer  0.225770   True  german_short-haired_pointer

    p3_conf p3_dog  favorite_count  retweet_count
0  0.076110  False           39467           8853
1  0.068957   True           33819           6514
2  0.031379   True           25461           4328
3  0.040836  False           42908           8964
4  0.175219   True           41048           9774

[5 rows x 22 columns]
```

The data looks good to evaluate now.

```
In [38]: #save the cleaned dataframe.
         dfCombine.to_csv('twitter_archive_master.csv', index=False)
```

```
In [ ]:
```