wrangle_act

December 5, 2019

1 Data Wrangling Project

```
In [85]: import pandas as pd
         import matplotlib.pyplot as plt
         import requests
         import numpy as np
         from io import StringIO
         import tweepy as tp
         import json
         import seaborn as sb
```

```
In [2]: # load csv file into df
        df= pd.read_csv('twitter-archive-enhanced.csv')
```

```
In [3]: df.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 [4]: # checking number of rows and columns
        df.shape

Out[4]: (2356, 17)

In [5]: df.head

Out[5]: <bound method NDFrame.head of   tweet_id in_reply_to_status_id in_reply_
        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 [6]: images= requests.get("https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_

        images=StringIO(images.text)

        img= pd.read_csv(images, sep="\t")
        img

Out[6]:                    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/CT5PY9OWoAAQGLo.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/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/CT9cxOtUEAAhNN_.jpg
25      666362758909284353      https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26      666373753744588802      https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.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/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_O-UQAACu2O.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_2PVWOAEHN3p.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/DFrEyVuWOAAO3t9.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
```

12

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
```

14

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 [7]: #Loading in given file
df2=pd.read_csv('image-predictions.tsv', delimiter= '\t', encoding = 'utf-8')

In [8]: # Checking to see if it downloaded correctly
df2.info

Out[8]: <bound method DataFrame.info of   tweet_id
        0   666020888022790149   https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg

16

```
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/CT9cxOtUEAAhNN_.jpg
25     666362758909284353      https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26     666373753744588802      https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.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_O-UQAACu2O.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_2PVWOAEHN3p.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/DFrEyVuWOAAO3t9.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 [9]: #Setting up API
        consumer_key= 'DyyXXwTUCQ8YNb7gPWApcnqGP'
        consumer_secret= 'UJtetZf3yJR2GqdBs6GISqqxFpa7mJENobIfZv4O3VuBTz94zi'
        access_token= '1084429021998432257-cuWaI6usQHFN7MXXDHj0ktk56ukZdX'
        access_token_secret='VbUOLxGKo4qAIhoUtxHsXtWxyZ8XnWeIgfp72SPDXlCWE'

In [10]: auth= tp.OAuthHandler(consumer_key, consumer_secret)
         api = tp.API(auth, wait_on_rate_limit= True, wait_on_rate_limit_notify= True)

In [11]: df_list= []
         errors= []
         for id in df2.tweet_id:
             try:
                 page= api.get_status(id, tweet_mode= 'extended')
                 favorites=page.favorite_count
                 retweets=page.retweet_count
                 date_time=page.created_at
                 df_list.append({'tweet_id':int(id),
                                 'favorites': int(favorites),
                                 'retweets':int(retweets),
                                 'date_time': pd.to_datetime(date_time)})
             except Exception as e:
                 print(str(id)+ "_" + str(e))
                 errors.append(id)

680055455951884288_[{'code': 144, 'message': 'No status found with that ID.'}]
Rate limit reached. Sleeping for: 742
754011816964026368_[{'code': 144, 'message': 'No status found with that ID.'}]
779123168116150273_[{'code': 144, 'message': 'No status found with that ID.'}]
802247111496568832_[{'code': 144, 'message': 'No status found with that ID.'}]
829374341691346946_[{'code': 144, 'message': 'No status found with that ID.'}]
Rate limit reached. Sleeping for: 741
837012587749474308_[{'code': 144, 'message': 'No status found with that ID.'}]
837366284874571778_[{'code': 144, 'message': 'No status found with that ID.'}]
842892208864923648_[{'code': 144, 'message': 'No status found with that ID.'}]
844704788403113984_[{'code': 144, 'message': 'No status found with that ID.'}]

22
```

```
851953902622658560_[{'code': 144, 'message': 'No status found with that ID.'}]
861769973181624320_[{'code': 144, 'message': 'No status found with that ID.'}]
872261713294495745_[{'code': 144, 'message': 'No status found with that ID.'}]
873697596434513921_[{'code': 144, 'message': 'No status found with that ID.'}]
888202515573088257_[{'code': 144, 'message': 'No status found with that ID.'}]


In [12]: error_list2= []
         for e in errors:
             try:
                 favorites=page.favorite_count
                 retweets=page.retweet_count
                 date_time=page['created_at']
                 df_list.append({'tweet_id':int(tweet_id),
                                 'favorites': int(favorites),
                                 'retweets':int(retweets),
                                 'date_time': pd.to_datetime(date_time)})
             except Exception as e:
                 print(str(id)+ "_" + str(e))
                 error_list2.append(id)

892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable
892420643555336193_'Status' object is not subscriptable


In [13]: #Viewing error list, noting duplicates
         error_list2

Out[13]: [892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
          892420643555336193,
```

```
            892420643555336193,
            892420643555336193,
            892420643555336193,
            892420643555336193,
            892420643555336193]
```

In [14]: *#Writing text to a json file*
```python
with open('tweet_json.txt', encoding='utf-8') as file:
    for id in df2['tweet_id']:
        try:
            file.write(json.dumps(page))
            file.write('\n')
        except: pass
file.close
```

Out[14]: <function TextIOWrapper.close()>

In [15]: *#Converting the list into a dataframe*
```python
df3= pd.DataFrame(df_list)
```

In [16]: *#Fixing tidiness issue #2*
```python
df_master=df2.merge(df3)
```

In [17]: *#Viewing first merge*
```python
df_master.head()
```

Out[17]:
```
              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

           date_time  favorites   retweets
```

```
0  2015-11-15 22:32:08          2463       474
1  2015-11-15 23:05:30           121        44
2  2015-11-15 23:21:54           118        43
3  2015-11-16 00:04:52           279       134
4  2015-11-16 00:24:50            99        41
```

In [18]: *#Executing second merge and viewing new dataframe*
         main2= df_master.merge(df)
         main2.head()

```
Out[18]:              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  ...  retweeted_status_user_id  \
         0  0.156665    True    Shetland_sheepdog  ...                       NaN
         1  0.074192    True  Rhodesian_ridgeback  ...                       NaN
         2  0.138584    True           bloodhound  ...                       NaN
         3  0.360687    True   miniature_pinscher  ...                       NaN
         4  0.243682    True             Doberman  ...                       NaN

            retweeted_status_timestamp  \
         0                         NaN
         1                         NaN
         2                         NaN
         3                         NaN
         4                         NaN

                                         expanded_urls  rating_numerator  \
         0  https://twitter.com/dog_rates/status/666020888...                 8
         1  https://twitter.com/dog_rates/status/666029285...                 7
         2  https://twitter.com/dog_rates/status/666033412...                 9
         3  https://twitter.com/dog_rates/status/666044226...                 6
         4  https://twitter.com/dog_rates/status/666049248...                 5

            rating_denominator  name  doggo floofer pupper puppo
         0                  10  None   None    None   None  None
         1                  10     a   None    None   None  None
```

```
2              10       a    None    None    None    None
3              10       a    None    None    None    None
4              10    None    None    None    None    None

[5 rows x 31 columns]
```

In [19]: *#Verifying text is in the dataframe*
         main2.text

Out[19]: 0   Here we have a Japanese Irish Setter. Lost eye...
         1   This is a western brown Mitsubishi terrier. Up...
         2   Here is a very happy pup. Big fan of well-main...
         3   This is a purebred Piers Morgan. Loves to Netf...
         4   Here we have a 1949 1st generation vulpix. Enj...
         5   This is a truly beautiful English Wilson Staff...
         6   This is an odd dog. Hard on the outside but lo...
         7   Here is a Siberian heavily armored polar bear ...
         8   My oh my. This is a rare blond Canadian terrie...
         9   Here is the Rand Paul of retrievers folks! He'...
         10   This is the happiest dog you will ever see. Ve...
         11   Here we have a northern speckled Rhododendron...
         12   Let's hope this flight isn't Malaysian (lol). ...
         13   Here we have a well-established sunblockerspan...
         14   This appears to be a Mongolian Presbyterian mi...
         15   Can stand on stump for what seems like a while...
         16   Oh my. Here you are seeing an Adobe Setter giv...
         17   Not familiar with this breed. No tail (weird)...
         18   Very concerned about fellow dog trapped in com...
         19   Can take selfies 11/10 https://t.co/ws2AMaNwPW
         20   This is an Albanian 3 1/2 legged Episcopalian...
         21   This is a funny dog. Weird toes. Won't come do...
         22   This is an extremely rare horned Parthenon. No...
         23   Look at this jokester thinking seat belt laws ...
         24   Here we have a mixed Asiago from the Galápagos...
         25   Unique dog here. Very small. Lives in containe...
         26   Those are sunglasses and a jean jacket. 11/10 ...
         27   Oh goodness. A super rare northeast Qdoba kang...
         28   This is a southern Vesuvius bumblegruff. Can d...
         29   This is quite the dog. Gets really excited whe...
                                     ...
         2031   This is Waffles. His doggles are pupside down...
         2032   This is Roscoe. Another pupper fallen victim t...
         2033   This is Derek. He's late for a dog meeting. 13...
         2034   This is Mingus. He's a wonderful father to his...
         2035   This is Maya. She's very shy. Rarely leaves he...
         2036   This... is a Jubilant Antarctic House Bear. We...
         2037   You may not have known you needed to see this ...
         2038   This is Canela. She attempted some fancy porch...

```
2039    I've yet to rate a Venezuelan Hover Wiener. Th...
2040    This is Jeffrey. He has a monopoly on the pool...
2041    This is Gerald. He was just told he didn't get...
2042    This is Ralphus. He's powering up. Attempting ...
2043    This is Zeke. He has a new stick. Very proud o...
2044    This is Jim. He found a fren. Taught him how t...
2045    This is Oliver. You're witnessing one of his m...
2046    This is Stuart. He's sporting his favorite fan...
2047    This is Ted. He does his best. Sometimes that'...
2048    Here's a puppo that seems to be on the fence a...
2049    This is Bruno. He is a service shark. Only get...
2050    This is Koda. He is a South Australian decksha...
2051    This is Cassie. She is a college pup. Studying...
2052    This is Zoey. She doesn't want to be one of th...
2053    When you watch your owner call another dog a g...
2054    Meet Jax. He enjoys ice cream so much he gets ...
2055    Here we have a majestic great white breaching ...
2056    This is Franklin. He would like you to stop ca...
2057    This is Darla. She commenced a snooze mid meal...
2058    This is Archie. He is a rare Norwegian Pouncin...
2059    This is Tilly. She's just checking pup on you...
2060    This is Phineas. He's a mystical boy. Only eve...
Name: text, Length: 2061, dtype: object
```

2 Assessing data

In [20]: *#testing*
         main2.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2061 entries, 0 to 2060
Data columns (total 31 columns):
tweet_id                2061 non-null int64
jpg_url                 2061 non-null object
img_num                 2061 non-null int64
p1                      2061 non-null object
p1_conf                 2061 non-null float64
p1_dog                  2061 non-null bool
p2                      2061 non-null object
p2_conf                 2061 non-null float64
p2_dog                  2061 non-null bool
p3                      2061 non-null object
p3_conf                 2061 non-null float64
p3_dog                  2061 non-null bool
date_time               2061 non-null datetime64[ns]
favorites               2061 non-null int64
retweets                2061 non-null int64
```

```
in_reply_to_status_id       23 non-null float64
in_reply_to_user_id         23 non-null float64
timestamp                   2061 non-null object
source                      2061 non-null object
text                        2061 non-null object
retweeted_status_id         74 non-null float64
retweeted_status_user_id    74 non-null float64
retweeted_status_timestamp  74 non-null object
expanded_urls               2061 non-null object
rating_numerator            2061 non-null int64
rating_denominator          2061 non-null int64
name                        2061 non-null object
doggo                       2061 non-null object
floofer                     2061 non-null object
pupper                      2061 non-null object
puppo                       2061 non-null object
dtypes: bool(3), datetime64[ns](1), float64(7), int64(6), object(14)
memory usage: 473.0+ KB
```

In [21]: *#testing*
         main2.describe()

Out[21]:   tweet_id   img_num   p1_conf   p2_conf   p3_conf \
       count 2.061000e+03 2061.000000 2061.000000 2.061000e+03 2.061000e+03
       mean   7.378612e+17   1.203785   0.594270 1.346576e-01 6.040991e-02
       std   6.755694e+16   0.562765   0.271183 1.006952e-01 5.097150e-02
       min   6.660209e+17   1.000000   0.044333 1.011300e-08 1.740170e-10
       25%   6.762636e+17   1.000000   0.363272 5.387110e-02 1.624560e-02
       50%   7.113638e+17   1.000000   0.587764 1.181810e-01 4.947150e-02
       75%   7.928838e+17   1.000000   0.843911 1.955730e-01 9.219820e-02
       max   8.924206e+17   4.000000   1.000000 4.880140e-01 2.734190e-01

                 favorites   retweets in_reply_to_status_id \
       count   2061.000000   2061.000000   2.300000e+01
       mean   8119.295488   2647.380398   6.978112e+17
       std   12264.777774   4628.012425   4.359384e+16
       min   0.000000   11.000000   6.671522e+17
       25%   1548.000000   566.000000   6.732411e+17
       50%   3578.000000   1250.000000   6.757073e+17
       75%   10138.000000   3038.000000   7.031489e+17
       max   159247.000000   79478.000000   8.558181e+17

               in_reply_to_user_id retweeted_status_id retweeted_status_user_id \
       count   2.300000e+01   7.400000e+01   7.400000e+01
       mean   4.196984e+09   7.619394e+17   1.064138e+16
       std   0.000000e+00   5.800000e+16   9.154057e+16
       min   4.196984e+09   6.675094e+17   7.832140e+05
```

```
25%        4.196984e+09        7.117707e+17        4.196984e+09
50%        4.196984e+09        7.755992e+17        4.196984e+09
75%        4.196984e+09        7.958815e+17        4.196984e+09
max        4.196984e+09        8.768508e+17        7.874618e+17

       rating_numerator  rating_denominator
count       2061.000000         2061.000000
mean          12.262494           10.514799
std           40.818203            7.201294
min            0.000000            2.000000
25%           10.000000           10.000000
50%           11.000000           10.000000
75%           12.000000           10.000000
max         1776.000000          170.000000
```

In [22]: *#testing*
main2.retweets.describe()

Out[22]:
```
count     2061.000000
mean      2647.380398
std       4628.012425
min         11.000000
25%        566.000000
50%       1250.000000
75%       3038.000000
max      79478.000000
Name: retweets, dtype: float64
```

In [23]: *#reviewing head of table to asssess for issues*
main2.head(10)
#noting that name column has two quality issues:
#1. name= None
#2. name= a, an, etc.

Out[23]:
```
            tweet_id                                         jpg_url  \
0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAAOaMy.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

   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
```

29

		p1	p1_conf	p1_dog	p2
2	1	German_shepherd	0.596461	True	malinois
3	1	Rhodesian_ridgeback	0.408143	True	redbone
4	1	miniature_pinscher	0.560311	True	Rottweiler
5	1	Bernese_mountain_dog	0.651137	True	English_springer
6	1	box_turtle	0.933012	False	mud_turtle
7	1	chow	0.692517	True	Tibetan_mastiff
8	1	shopping_cart	0.962465	False	shopping_basket
9	1	miniature_poodle	0.201493	True	komondor

	p2_conf	p2_dog	p3	...	\
0	0.156665	True	Shetland_sheepdog	...	
1	0.074192	True	Rhodesian_ridgeback	...	
2	0.138584	True	bloodhound	...	
3	0.360687	True	miniature_pinscher	...	
4	0.243682	True	Doberman	...	
5	0.263788	True	Greater_Swiss_Mountain_dog	...	
6	0.045885	False	terrapin	...	
7	0.058279	True	fur_coat	...	
8	0.014594	False	golden_retriever	...	
9	0.192305	True	soft-coated_wheaten_terrier	...	

	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	

	expanded_urls	rating_numerator	\
0	https://twitter.com/dog_rates/status/666020888...	8	
1	https://twitter.com/dog_rates/status/666029285...	7	
2	https://twitter.com/dog_rates/status/666033412...	9	
3	https://twitter.com/dog_rates/status/666044226...	6	
4	https://twitter.com/dog_rates/status/666049248...	5	
5	https://twitter.com/dog_rates/status/666050758...	10	
6	https://twitter.com/dog_rates/status/666051853...	2	
7	https://twitter.com/dog_rates/status/666055525...	10	
8	https://twitter.com/dog_rates/status/666057090...	9	
9	https://twitter.com/dog_rates/status/666058600...	8	

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	None	None	None	None	None
1	10	a	None	None	None	None

2	10	a	None	None	None	None
3	10	a	None	None	None	None
4	10	None	None	None	None	None
5	10	a	None	None	None	None
6	10	an	None	None	None	None
7	10	a	None	None	None	None
8	10	a	None	None	None	None
9	10	the	None	None	None	None

```
[10 rows x 31 columns]

In [24]: #Viewing number of predictions for each breed
         main2['p1'].value_counts()[:10]

Out[24]: golden_retriever      150
         Labrador_retriever     98
         Pembroke               88
         Chihuahua              81
         pug                    57
         chow                   44
         Samoyed                42
         Pomeranian             38
         toy_poodle             38
         malamute               30
         Name: p1, dtype: int64

In [25]: #noting denominators other than 10
         main2['rating_denominator'].unique()

Out[25]: array([ 10,    2,  80, 120,  11, 110, 130,  40,  50,  90,  20, 170, 150,
                 7,  70])

In [26]: #noting numerators >15
         main2['rating_numerator'].unique()

Out[26]: array([   8,    7,    9,    6,    5,   10,    2,   11,    1,    3,   12,
                  4,   13,  420,   88,  144,   26,  121,  143,   14,   44,   60,
                 45,   80,   99,   50,  204,    0, 1776,  165,   27,   75,   24,
                 84,   15])
```

2.1 Quality Issues:

1. tweet_id should be a string
2. duplications of tweet_id
3. 'None' instead of NA in name column
4. retweets included in the data set
5. replies to tweets included in data set
6. numerators greater than 15
7. denominators other than 10
8. a, an, the listed as names

2.2 Tidiness Issues

1. Dog Stage is spread across columns
2. Data is spread across three tables
3. Unnecessary columns

3 Cleaning Data

3.1 Quality Issues

```
In [27]: #1.Changing tweet_id to a string
         main2.tweet_id= main2.tweet_id.astype(str)
```

```
In [28]: #checking work
         main2.columns
```

```
Out[28]: Index(['tweet_id', 'jpg_url', 'img_num', 'p1', 'p1_conf', 'p1_dog', 'p2',
                'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3_dog', 'date_time',
                'favorites', 'retweets', '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'],
               dtype='object')
```

```
In [29]: #2. Getting rid of duplicated Ids
         main2 = main2[ ~main2.tweet_id.duplicated() ]
```

```
In [30]: #checking work
         len(main2)
```

```
Out[30]: 2061
```

```
In [31]: #3. Changing 'none' for name to 'NaN'
         main2['name']= main2['name'].replace('None', np.nan)
```

```
In [32]: #4. Deleting retweets
         main2= main2[pd.isnull(main2.retweeted_status_id)]
```

```
In [33]: #checking work
         len(main2)
```

```
Out[33]: 1987
```

```
In [34]: #5. Deleting replies
         main2= main2[pd.isnull(main2.in_reply_to_status_id)]
```

```
In [35]: main2= main2[pd.isnull(main2.in_reply_to_user_id)]
```

```
In [36]: #checking work
         len(main2)
```

```
Out[36]: 1964

In [37]: #6. Searching for index numbers of abnormal numerators
         main2.query('rating_numerator == "501"')
         main2.query('rating_numerator == "277"')
         main2.query('rating_numerator == "420"')
         main2.query('rating_numerator == "144"')
         main2.query('rating_numerator == "121"')
         main2.query('rating_numerator == "44"')
         main2.query('rating_numerator == "60"')
         main2.query('rating_numerator == "45"')
         main2.query('rating_numerator == "80"')
         main2.query('rating_numerator == "50"')
         main2.query('rating_numerator == "204"')
         main2.query('rating_numerator == "1776"')
         main2.query('rating_numerator == "165"')
         main2.query('rating_numerator == "27"')
         main2.query('rating_numerator == "75"')
         main2.query('rating_numerator == "24"')
         main2.query('rating_numerator == "84"')
         main2.query('rating_numerator == "0"')

Out[37]:                    tweet_id                                      jpg_url  \
         1819   835152434251116546   https://pbs.twimg.com/media/C5cOtWVWMAEjO5p.jpg


                img_num     p1   p1_conf  p1_dog                             p2  \
         1819         3   swing  0.967066   False   American_Staffordshire_terrier


                 p2_conf  p2_dog                           p3  ...  \
         1819   0.012731    True   Staffordshire_bullterrier  ...


                retweeted_status_user_id  retweeted_status_timestamp  \
         1819                        NaN                         NaN


                                            expanded_urls  rating_numerator  \
         1819   https://twitter.com/dog_rates/status/835152434...                 0


                rating_denominator  name  doggo floofer pupper puppo
         1819                   10   NaN   None    None   None  None


         [1 rows x 31 columns]

In [38]: #using indices to drop the rows with abnormal numerators
         main2= main2.drop(index= 501)
         main2= main2.drop(index= 277)
         main2= main2.drop(index= 420)
         main2= main2.drop(index= 560)
         main2= main2.drop(index= 691)
```

```
main2= main2.drop(index= 864)
main2= main2.drop(index= 940)
main2= main2.drop(index= 1006)
main2= main2.drop(index= 1024)
main2= main2.drop(index= 1049)
main2= main2.drop(index= 1070)
main2= main2.drop(index= 1147)
main2= main2.drop(index= 1269)
main2= main2.drop(index= 1336)
main2= main2.drop(index= 1457)
main2= main2.drop(index= 1512)
main2= main2.drop(index= 1654)
main2= main2.drop(index= 1724)
main2= main2.drop(index= 1819)
main2= main2.drop(index= 620)
```

In [39]: *#checking work*
         len(main2)

Out[39]: 1944

In [40]: *#7. Noting indices for denominators other than 10*
         main2[main2['rating_denominator'] !=10]

Out[40]:   tweet_id   jpg_url \
         20   666287406224695296   https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
         666   682962037429899265   https://pbs.twimg.com/media/CXpeVzQW8AApKYb.jpg
         1104 722974582966214656   https://pbs.twimg.com/media/CgiFjIpWgAA4wVp.jpg
         1195 740373189193256964   https://pbs.twimg.com/media/CkZVdJ6WYAAXZ5A.jpg

               img_num   p1   p1_conf p1_dog   p2 \
         20   1   Maltese_dog 0.857531   True   toy_poodle
         666   1   dingo 0.278600   False   Chihuahua
         1104   1   Great_Dane 0.246762   True Greater_Swiss_Mountain_dog
         1195   3 golden_retriever 0.807644   True   kuvasz

                 p2_conf p2_dog   p3 ...   retweeted_status_user_id \
         20   0.063064   True   miniature_poodle ...   NaN
         666   0.155207   True   loupe ...   NaN
         1104   0.126131   True   Weimaraner ...   NaN
         1195   0.101286   True   Labrador_retriever ...   NaN

               retweeted_status_timestamp \
         20   NaN
         666   NaN
         1104   NaN
         1195   NaN

                                        expanded_urls rating_numerator \
```
34
```

```
20     https://twitter.com/dog_rates/status/666287406...                    1
666    https://twitter.com/dog_rates/status/682962037...                    7
1104   https://twitter.com/dog_rates/status/722974582...                    4
1195   https://twitter.com/dog_rates/status/740373189...                    9


       rating_denominator    name  doggo floofer pupper puppo
20                      2      an   None    None   None  None
666                    11  Darrel   None    None   None  None
1104                   20     NaN   None    None   None  None
1195                   11     NaN   None    None   None  None


[4 rows x 31 columns]
```

In [41]: *#using indices to drop rows with denominators other than 10*
         main2= main2.drop(index= 20)
         main2= main2.drop(index= 666)
         main2= main2.drop(index= 1104)
         main2= main2.drop(index= 1195)

In [42]: main2.head()

Out[42]:
```
                    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


       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  ...  retweeted_status_user_id  \
    0  0.156665    True   Shetland_sheepdog  ...                       NaN
    1  0.074192    True  Rhodesian_ridgeback  ...                       NaN
    2  0.138584    True          bloodhound  ...                       NaN
    3  0.360687    True  miniature_pinscher  ...                       NaN
    4  0.243682    True            Doberman  ...                       NaN


       retweeted_status_timestamp  \
    0                         NaN
    1                         NaN
    2                         NaN
    3                         NaN
    4                         NaN
```

```
                                      expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/666020888...                 8
1  https://twitter.com/dog_rates/status/666029285...                 7
2  https://twitter.com/dog_rates/status/666033412...                 9
3  https://twitter.com/dog_rates/status/666044226...                 6
4  https://twitter.com/dog_rates/status/666049248...                 5

   rating_denominator  name doggo floofer pupper puppo
0                  10   NaN  None    None   None  None
1                  10     a  None    None   None  None
2                  10     a  None    None   None  None
3                  10     a  None    None   None  None
4                  10   NaN  None    None   None  None

[5 rows x 31 columns]
```

In [43]: *#8. Deleting a, an, the, from name column*
```
main2['name'] = main2['name'].replace(['a', 'an', 'the'], np.NaN)
```

In [44]: *#checking work*
```
main2['name'].value_counts()
```

Out[44]:
```
Cooper      10
Charlie     10
Oliver      10
Lucy         9
Tucker       9
Penny        9
Sadie        8
Winston      8
Daisy        7
Lola         7
Toby         7
Bella        6
Jax          6
Stanley      6
Koda         6
Bo           6
Dave         5
Buddy        5
Chester      5
Rusty        5
Leo          5
Milo         5
Bailey       5
Louis        5
Oscar        5
```

```
             Scout         5
             Phil          4
             Jack          4
             George        4
             Bentley       4
                          ..
             Laela         1
             Zara          1
             Dale          1
             Willy         1
             Traviss       1
             such          1
             Fwed          1
             Coopson       1
             Acro          1
             Cal           1
             Ralph         1
             Hazel         1
             Franq         1
             DonDon        1
             Malikai       1
             Duchess       1
             Sundance      1
             Halo          1
             Superpup      1
             Arnold        1
             Torque        1
             Todo          1
             Quinn         1
             Odin          1
             Brady         1
             all           1
             Jameson       1
             Lance         1
             Lupe          1
             Rey           1
             Name: name, Length: 925, dtype: int64
```

In [45]: *#9. Editing dog breed column so that everything is lowercase and dashes are removed*
```python
main2['p1'] = main2['p1'].astype('str')
main2['p1'] = main2['p1'].str.lower().str.replace('_', ' ').str.replace('-', ' ')
main2['p1'] = main2['p1'].replace('nan', np.nan)
```

3.2 Tidiness Issues

In [46]: *#creating new column to merge data from columns 'doggo', 'floofer', 'pupper', 'puppo'*

```python
main2['stage']= 0
```

```
        main2['stage'] = ['doggo' if 'doggo' in i
                           else 'pupper' if 'pupper' in i
                           else 'floofer' if 'floofer' in i
                           else 'puppo' if 'puppo' in i
                           else np.nan for i in main2['text'].str.lower()]
        #changing into categorical
        main2['stage'] = main2['stage'].astype('category')

In [47]: #checking work
        main2['stage'].value_counts()

Out[47]: pupper     217
        doggo       77
        puppo       26
        floofer      7
        Name: stage, dtype: int64
```

4   2. Cleaned in earlier code. Data was spread among 3 tables. I first joined df2 and df3. After that I joined df to get my final dataframe with all of the acquired data in it.

```
In [48]: #Viewing columns to sort
        main2.columns

Out[48]: Index(['tweet_id', 'jpg_url', 'img_num', 'p1', 'p1_conf', 'p1_dog', 'p2',
               'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3_dog', 'date_time',
               'favorites', 'retweets', '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', 'stage'],
              dtype='object')

In [49]: #3. Dropping unnecessary columns to make data more readable
        main2 = main2.drop(columns= ['img_num', 'p2', 'p2_conf', 'p2_dog', 'p3', 'p3_conf', 'p3
        main2.head()

Out[49]:            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


                              p1   p1_conf  p1_dog             date_time  favorites  \
        0  welsh springer spaniel  0.465074    True  2015-11-15 22:32:08       2463
        1                 redbone  0.506826    True  2015-11-15 23:05:30        121
        2         german shepherd  0.596461    True  2015-11-15 23:21:54        118
```

```
3      rhodesian ridgeback  0.408143      True 2015-11-16 00:04:52           279
4      miniature pinscher   0.560311      True 2015-11-16 00:24:50            99

    retweets                    timestamp  \
0        474  2015-11-15 22:32:08 +0000
1         44  2015-11-15 23:05:30 +0000
2         43  2015-11-15 23:21:54 +0000
3        134  2015-11-16 00:04:52 +0000
4         41  2015-11-16 00:24:50 +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...

                                                text  \
0  Here we have a Japanese Irish Setter. Lost eye...
1  This is a western brown Mitsubishi terrier. Up...
2  Here is a very happy pup. Big fan of well-main...
3  This is a purebred Piers Morgan. Loves to Netf...
4  Here we have a 1949 1st generation vulpix. Enj...

                                       expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/666020888...                 8
1  https://twitter.com/dog_rates/status/666029285...                 7
2  https://twitter.com/dog_rates/status/666033412...                 9
3  https://twitter.com/dog_rates/status/666044226...                 6
4  https://twitter.com/dog_rates/status/666049248...                 5

   rating_denominator name stage
0                  10  NaN   NaN
1                  10  NaN   NaN
2                  10  NaN   NaN
3                  10  NaN   NaN
4                  10  NaN   NaN
```

In [50]: *#checking to make sure columns were dropped*
         main2.columns

Out[50]: Index(['tweet_id', 'jpg_url', 'p1', 'p1_conf', 'p1_dog', 'date_time',
                'favorites', 'retweets', 'timestamp', 'source', 'text', 'expanded_urls',
                'rating_numerator', 'rating_denominator', 'name', 'stage'],
               dtype='object')

In [57]: *#dropping more columns*
         main2 = main2.drop(columns= ['p1_dog', 'Stage'])
         main2.head()

```
Out[57]:              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


                            p1    p1_conf  p1_dog            date_time  favorites  \
       0  welsh springer spaniel  0.465074    True  2015-11-15 22:32:08       2463
       1                 redbone  0.506826    True  2015-11-15 23:05:30        121
       2          german shepherd  0.596461    True  2015-11-15 23:21:54        118
       3      rhodesian ridgeback  0.408143    True  2015-11-16 00:04:52        279
       4       miniature pinscher  0.560311    True  2015-11-16 00:24:50         99


          retweets                  timestamp  \
       0       474  2015-11-15 22:32:08 +0000
       1        44  2015-11-15 23:05:30 +0000
       2        43  2015-11-15 23:21:54 +0000
       3       134  2015-11-16 00:04:52 +0000
       4        41  2015-11-16 00:24:50 +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...


                                                    text  \
       0  Here we have a Japanese Irish Setter. Lost eye...
       1  This is a western brown Mitsubishi terrier. Up...
       2  Here is a very happy pup. Big fan of well-main...
       3  This is a purebred Piers Morgan. Loves to Netf...
       4  Here we have a 1949 1st generation vulpix. Enj...


                                        expanded_urls  rating_numerator  \
       0  https://twitter.com/dog_rates/status/666020888...                 8
       1  https://twitter.com/dog_rates/status/666029285...                 7
       2  https://twitter.com/dog_rates/status/666033412...                 9
       3  https://twitter.com/dog_rates/status/666044226...                 6
       4  https://twitter.com/dog_rates/status/666049248...                 5


          rating_denominator name stage
       0                  10  NaN   NaN
       1                  10  NaN   NaN
       2                  10  NaN   NaN
       3                  10  NaN   NaN
       4                  10  NaN   NaN
```

In [58]: *#checking for any other unneccessary columns.*
         main2.columns

Out[58]: Index(['tweet_id', 'jpg_url', 'p1', 'p1_conf', 'p1_dog', 'date_time',
                'favorites', 'retweets', 'timestamp', 'source', 'text', 'expanded_urls',
                'rating_numerator', 'rating_denominator', 'name', 'stage'],
               dtype='object')

In [59]: main2.head()

Out[59]:   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

                            p1   p1_conf p1_dog   date_time favorites \
         0 welsh springer spaniel 0.465074   True 2015-11-15 22:32:08   2463
         1   redbone 0.506826   True 2015-11-15 23:05:30   121
         2   german shepherd 0.596461   True 2015-11-15 23:21:54   118
         3   rhodesian ridgeback 0.408143   True 2015-11-16 00:04:52   279
         4   miniature pinscher 0.560311   True 2015-11-16 00:24:50   99

            retweets   timestamp \
         0   474 2015-11-15 22:32:08 +0000
         1   44 2015-11-15 23:05:30 +0000
         2   43 2015-11-15 23:21:54 +0000
         3   134 2015-11-16 00:04:52 +0000
         4   41 2015-11-16 00:24:50 +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...

                                               text \
         0 Here we have a Japanese Irish Setter. Lost eye...
         1 This is a western brown Mitsubishi terrier. Up...
         2 Here is a very happy pup. Big fan of well-main...
         3 This is a purebred Piers Morgan. Loves to Netf...
         4 Here we have a 1949 1st generation vulpix. Enj...

                                       expanded_urls rating_numerator \
         0 https://twitter.com/dog_rates/status/666020888...   8
         1 https://twitter.com/dog_rates/status/666029285...   7

```
        2  https://twitter.com/dog_rates/status/666033412...                        9
        3  https://twitter.com/dog_rates/status/666044226...                        6
        4  https://twitter.com/dog_rates/status/666049248...                        5


           rating_denominator name stage
        0                  10  NaN   NaN
        1                  10  NaN   NaN
        2                  10  NaN   NaN
        3                  10  NaN   NaN
        4                  10  NaN   NaN
```

In [60]: *#Storing end dataframe*
         main2.to_csv('twitter_archive_master.csv', index= False)

5   Insights and visualization

In [61]: main2['retweets'].describe()

Out[61]: count   1940.000000
         mean   2517.566495
         std   4512.568827
         min   11.000000
         25%   560.750000
         50%   1213.500000
         75%   2876.500000
         max   79478.000000
         Name: retweets, dtype: float64

In [64]: main2.query('retweets == "79478"')

Out[64]:   tweet_id   jpg_url \
         1220   744234799360020481   https://pbs.twimg.com/ext_tw_video_thumb/74423...

                             p1   p1_conf p1_dog   date_time favorites \
         1220 labrador retriever 0.825333   True 2016-06-18 18:26:18   159247

               retweets   timestamp \
         1220   79478   2016-06-18 18:26:18 +0000

                                                       source \
         1220 <a href="http://twitter.com/download/iphone" r...

                                                         text \
         1220 Here's a doggo realizing you can stand in a po...

                                      expanded_urls rating_numerator \
         1220 https://twitter.com/dog_rates/status/744234799...   13
```

```
         rating_denominator name  stage
1220                     10  NaN  doggo
```

Insight 1. The post with the most retweets is a labrador retriever. Labrador retrievers are also the mode of the predictions.

```
In [65]: #searching for post with lowest amount of retweets
         main2.query('retweets == "11"')

Out[65]:                 tweet_id                                    jpg_url  \
         16   666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg


                        p1    p1_conf  p1_dog          date_time  favorites  retweets  \
         16   english setter  0.298617    True 2015-11-16 03:55:04         73        11


                        timestamp  \
         16   2015-11-16 03:55:04 +0000


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


                                                   text  \
         16   Oh my. Here you are seeing an Adobe Setter giv...


                                          expanded_urls  rating_numerator  \
         16   https://twitter.com/dog_rates/status/666102155...                11


             rating_denominator name stage
         16                   10  NaN   NaN

In [66]: #searching for english setters
         main2.query('p1 == "english setter"')

Out[66]:                 tweet_id                                    jpg_url  \
         16     666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
         46     666804364988780544  https://pbs.twimg.com/media/CUD3A7YWoAA82N0.jpg
         389    673359818736984064  https://pbs.twimg.com/media/CVhBLohWEAAXtYl.jpg
         541    677187300187611136  https://pbs.twimg.com/media/CWXaQMBWcAAATDi.jpg
         664    682750546109968385  https://pbs.twimg.com/media/CXmd_bsWkAEEXck.jpg
         1149   732005617171337216  https://pbs.twimg.com/media/CiibOMzUYAA9Mxz.jpg
         1749   824663926340194305  https://pbs.twimg.com/media/C3HLdOHXUAAUI2b.jpg


                          p1    p1_conf  p1_dog          date_time  favorites  \
         16     english setter  0.298617    True 2015-11-16 03:55:04         73
         46     english setter  0.328792    True 2015-11-18 02:25:23        225
         389    english setter  0.696568    True 2015-12-06 04:34:25       1427
         541    english setter  0.282396    True 2015-12-16 18:03:28       2755
         664    english setter  0.947198    True 2016-01-01 02:29:49       1536
         1149   english setter  0.677408    True 2016-05-16 00:31:53      15109
```

```
1749    english setter   0.526488      True 2017-01-26 17:02:56          10377

         retweets                       timestamp  \
16             11   2015-11-16 03:55:04 +0000
46             86   2015-11-18 02:25:23 +0000
389           660   2015-12-06 04:34:25 +0000
541           925   2015-12-16 18:03:28 +0000
664           445   2016-01-01 02:29:49 +0000
1149         5496   2016-05-16 00:31:53 +0000
1749         1799   2017-01-26 17:02:56 +0000

                                             source  \
16      <a href="http://twitter.com/download/iphone" r...
46      <a href="http://twitter.com/download/iphone" r...
389     <a href="http://twitter.com/download/iphone" r...
541     <a href="http://twitter.com/download/iphone" r...
664     <a href="http://twitter.com/download/iphone" r...
1149    <a href="http://twitter.com/download/iphone" r...
1749    <a href="http://twitter.com/download/iphone" r...

                                               text  \
16      Oh my. Here you are seeing an Adobe Setter giv...
46      This is Jockson. He is a Pinnacle Sagittarius...
389     This is Steve. He was just relaxing in hot tub...
541     Here we see a Byzantine Rigatoni. Very aerodyn...
664     Meet Taco. He's a speckled Garnier Fructis. Lo...
1149    This is Larry. He has no self control. Tongue ...
1749    This is Wilson. Named after the volleyball. He...

                                      expanded_urls  rating_numerator  \
16      https://twitter.com/dog_rates/status/666102155...                11
46      https://twitter.com/dog_rates/status/666804364...                 8
389     https://twitter.com/dog_rates/status/673359818...                 8
541     https://twitter.com/dog_rates/status/677187300...                 9
664     https://twitter.com/dog_rates/status/682750546...                 9
1149    https://twitter.com/dog_rates/status/732005617...                11
1749    https://twitter.com/dog_rates/status/824663926...                13

         rating_denominator       name stage
16                      10        NaN   NaN
46                      10    Jockson   NaN
389                     10      Steve   NaN
541                     10        NaN   NaN
664                     10       Taco   NaN
1149                    10      Larry   NaN
1749                    10     Wilson   NaN
```

Insight 2. The post with the least amount of retweets is an english setter. It is not the only one

in the data set. The average number of retweets for an English Setter is 1,346. That means that this post, with only 11 retweets, is an outlier.

```
In [67]: #getting information on the ratings
         main2['rating_numerator'].describe()

Out[67]: count    1940.000000
         mean       10.536598
         std         2.162840
         min         1.000000
         25%        10.000000
         50%        11.000000
         75%        12.000000
         max        14.000000
         Name: rating_numerator, dtype: float64

In [69]: ten= main2.query('rating_numerator == "10"')
         len(ten)

Out[69]: 416

In [68]: eleven= main2.query('rating_numerator == "11"')
         len(eleven)

Out[68]: 392

In [72]: twelve= main2.query('rating_numerator == "12"')
         len(twelve)

Out[72]: 443

In [73]: thirteen= main2.query('rating_numerator == "13"')
         len(thirteen)

Out[73]: 249

In [74]: fourteen= main2.query('rating_numerator == "14"')
         len(fourteen)

Out[74]: 33
```

Insight 3. The most frequently occurring numerator is 12. On either side of it, the frequencies go down like a bell curve. The numerator 10, however, has a high frequency. 13/10 is a common rating, but generally the highest rating they give. Few dogs made 14/10.

5.1 Visualization

```
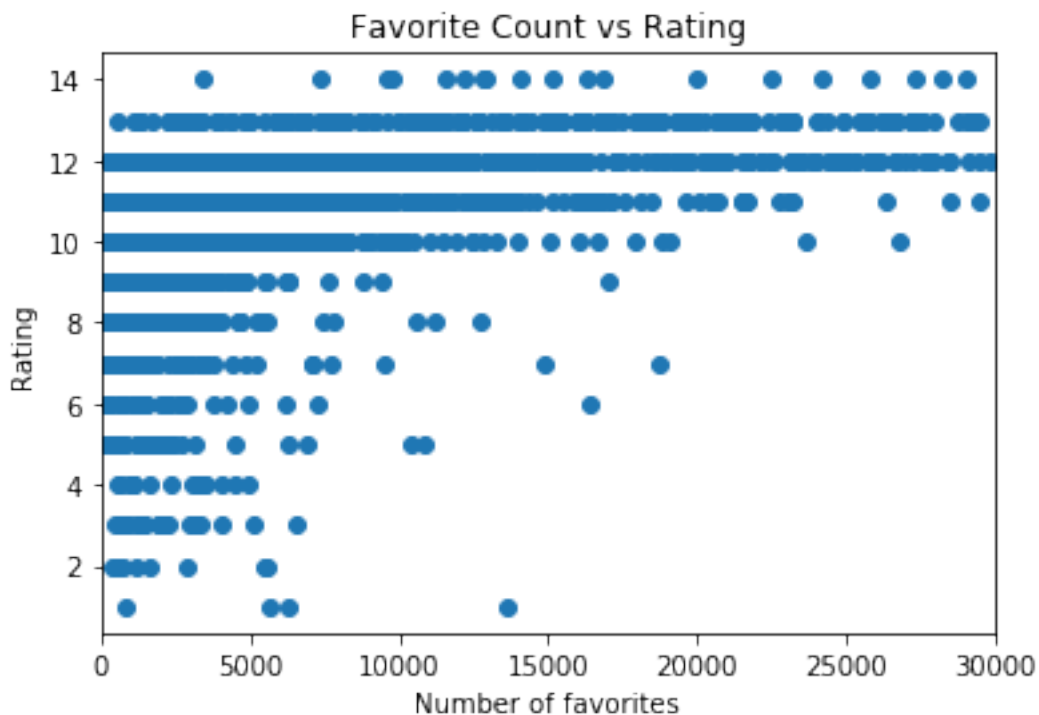In [75]: main2['favorites'].describe()
```

```
Out[75]: count        1940.000000
         mean         8469.191753
         std         12467.796501
         min            73.000000
         25%          1803.750000
         50%          3819.000000
         75%         10550.750000
         max        159247.000000
         Name: favorites, dtype: float64
```

```
In [89]: # the distribution of favorites on a histogram
         favorites2= main2['favorites']
         rating_numerator2= main2['rating_numerator']
         plt.scatter(x= favorites2, y=rating_numerator2)
         plt.xlim(0, 30000)
         plt.xlabel('Number of favorites')
         plt.ylabel('Rating')
         plt.title('Favorite Count vs Rating')
```

```
Out[89]: Text(0.5,1,'Favorite Count vs Rating')
```


The above scatterplot suggests a positive correlation between the rating a dog receives and the number of favorites on the tweet.

6 Reports

```
In [ ]: #installing a pdf reader
        !pip install PyPDF2

In [ ]: #reading in pdf
        from PyPDF2 import PdfFileReader
        wrangle_report= PdfFileReader('wrangle_report.pdf')

In [ ]: main2
```

resources: https://bekleydata.com/wp-content/uploads/files/wrangle_act.html
https://a-chong.github.io/data_wrangling_dog_ratings.html
https://www.geeksforgeeks.org/adding-new-column-to-existing-dataframe-in-pandas/
https://thispointer.com/python-pandas-how-to-drop-rows-in-dataframe-by-index-labels/