wrangle_act

April 24, 2018

WeRateDogs Tweets' Data Wrangling Project

```
In [2]: import pandas as pd
        import numpy as np
        import requests
        import tweepy
        import json
        import re
        import matplotlib.pyplot as plt
        import warnings


        url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictic
        r = requests.get(url)

In [3]: df_img_pred = pd.read_csv(url, sep = '\t')

In [4]: df_img_pred.head(10)

Out[4]:          tweet_id                                    jpg_url  \
        0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
        1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
        2  666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
        3  666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
        4  666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
        5  666050758794694657  https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
        6  666051853826850816  https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
        7  666055525042405380  https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
        8  666057090499244032  https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
        9  666058600524156928  https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg

           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
        5        1     Bernese_mountain_dog  0.651137    True    English_springer
```

1

```
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   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
5     0.263788    True    Greater_Swiss_Mountain_dog 0.016199     True
6     0.045885   False                      terrapin 0.017885    False
7     0.058279    True                      fur_coat 0.054449    False
8     0.014594   False             golden_retriever  0.007959     True
9     0.192305    True    soft-coated_wheaten_terrier 0.082086    True
```

In [5]:
```python
import tweepy

consumer_key = ''
consumer_secret = ''
access_token = ''
access_secret = ''

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

api = tweepy.API(auth)
```

In [6]:
```python
twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
```

In [7]:
```python
# For loop which will add each available tweet to a new line of tweet_json.txt
with open('tweet_json.txt', 'a', encoding='utf8') as f:
    for tweet_id in twitter_archive['tweet_id']:
        try:
            tweet = api.get_status(tweet_id, tweet_mode='extended')
            json.dump(tweet._json, f)
            f.write('\n')
        except:
            continue
```

In [8]:
```python
# For loop to append each tweet into a list
tweets_data = []

tweet_file = open('tweet_json.txt', "r")

for line in tweet_file:
    try:
        tweet = json.loads(line)
```

```
                tweets_data.append(tweet)
            except:
                continue

        tweet_file.close()

In [9]: # Create tweet_info DataFrame
        tweet_info = pd.DataFrame()

In [10]: # Add selected variables to tweet_info DataFrame
         tweet_info['id'] = list(map(lambda tweet: tweet['id'], tweets_data))
         tweet_info['retweet_count'] = list(map(lambda tweet: tweet['retweet_count'], tweets_dat
         tweet_info['favorite_count'] = list(map(lambda tweet: tweet['favorite_count'], tweets_d
```

Assess

```
In [11]: # View first 10 rows of twitter_archive DataFrame
         twitter_archive.head(10)

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

                             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

                                                      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...


                                                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


   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/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


   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
```

In [12]: # View last 5 rows of twitter_archive DataFrame
         twitter_archive.tail()

Out[12]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         2351 666049248165822465   NaN   NaN
         2352 666044226329800704   NaN   NaN
         2353 666033412701032449   NaN   NaN
         2354 666029285002620928   NaN   NaN
         2355 666020888022790149   NaN   NaN

                              timestamp \
         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 \
         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 \
         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 \
         2351   NaN   NaN
         2352   NaN   NaN
         2353   NaN   NaN
         2354   NaN   NaN
         2355   NaN   NaN

                                             expanded_urls rating_numerator \
         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
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
```

In [13]: # View twitter_archive DataFrame info
         twitter_archive.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 [14]: # View twitter_archive DataFrame descriptive statistics
         twitter_archive.describe()

```
Out[14]:          tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         count  2.356000e+03           7.800000e+01         7.800000e+01
         mean   7.427716e+17           7.455079e+17         2.014171e+16
         std    6.856705e+16           7.582492e+16         1.252797e+17
         min    6.660209e+17           6.658147e+17         1.185634e+07
         25%    6.783989e+17           6.757419e+17         3.086374e+08
         50%    7.196279e+17           7.038708e+17         4.196984e+09
         75%    7.993373e+17           8.257804e+17         4.196984e+09
         max    8.924206e+17           8.862664e+17         8.405479e+17
```

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


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

In [15]: *# View first rows of image_predictions DataFrame*
         df_img_pred.head()

Out[15]:   tweet_id   jpg_url \
         0 666020888022790149 https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
         1 666029285002620928 https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
         2 666033412701032449 https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
         3 666044226329800704 https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
         4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

```
   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
```

In [16]: *# View last 5 rows of image_predictions DataFrame*
         df_img_pred.tail()

Out[16]:   tweet_id   jpg_url \
         2070 891327558926688256 https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg

```
2071  891689557279858688  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  891815181378084864  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  892177421306343426  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  892420643555336193  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

        img_num            p1  p1_conf  p1_dog                  p2   p2_conf  \
2070          2        basset  0.555712    True     English_springer  0.225770
2071          1  paper_towel  0.170278   False   Labrador_retriever  0.168086
2072          1     Chihuahua  0.716012    True             malamute  0.078253
2073          1     Chihuahua  0.323581    True             Pekinese  0.090647
2074          1        orange  0.097049   False                bagel  0.085851

        p2_dog                          p3   p3_conf  p3_dog
2070     True  German_short-haired_pointer  0.175219    True
2071     True                      spatula  0.040836   False
2072     True                       kelpie  0.031379    True
2073     True                     papillon  0.068957    True
2074    False                       banana  0.076110   False
```

In [17]: *# View image_predictions DataFrame info*
         df_img_pred.info()

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

In [18]: *# View descriptive statistics of image_predictions DataFrame*
         df_img_pred.describe()

Out[18]:
```
              tweet_id       img_num       p1_conf       p2_conf       p3_conf
count    2.075000e+03   2075.000000   2075.000000  2.075000e+03  2.075000e+03
mean     7.384514e+17      1.203855      0.594548  1.345886e-01  6.032417e-02
std      6.785203e+16      0.561875      0.271174  1.006657e-01  5.090593e-02
min      6.660209e+17      1.000000      0.044333  1.011300e-08  1.740170e-10
```

```
        25%    6.764835e+17     1.000000    0.364412   5.388625e-02  1.622240e-02
        50%    7.119988e+17     1.000000    0.588230   1.181810e-01  4.944380e-02
        75%    7.932034e+17     1.000000    0.843855   1.955655e-01  9.180755e-02
        max    8.924206e+17     4.000000    1.000000   4.880140e-01  2.734190e-01
```

In [19]: # View first 5 rows of tweet_info DataFrame
         tweet_info.head()

Out[19]:
```
                          id  retweet_count  favorite_count
        0  892420643555336193           8642           38965
        1  892177421306343426           6351           33353
        2  891815181378084864           4213           25139
        3  891689557279858688           8759           42300
        4  891327558926688256           9522           40471
```

In [20]: # View last 5 rows of tweet_info DataFrame
         tweet_info.tail()

Out[20]:
```
                            id  retweet_count  favorite_count
        8195  759099523532779520           4681           15783
        8196  759047813560868866           2246            7080
        8197  758854675097526272            995            3850
        8198  758828659922702336           4268           12148
        8199  674019345211760640            324            1184
```

In [21]: # View info of tweet_info DataFrame
         tweet_info.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8200 entries, 0 to 8199
Data columns (total 3 columns):
id                8200 non-null int64
retweet_count     8200 non-null int64
favorite_count    8200 non-null int64
dtypes: int64(3)
memory usage: 192.3 KB
```

In [22]: # View descriptive statistics of tweet_info DataFrame
         tweet_info.describe()

Out[22]:
```
                       id  retweet_count  favorite_count
        count  8.200000e+03    8200.000000     8200.000000
        mean   8.168399e+17    5256.821585    14543.098171
        std    4.229340e+16    6093.743445    15539.191382
        min    6.660209e+17       0.000000        0.000000
        25%    7.869631e+17    2147.750000     5103.000000
        50%    8.174156e+17    3557.500000    11184.500000
        75%    8.479628e+17    6173.000000    20653.250000
        max    8.924206e+17   61445.000000   143962.000000
```

```
In [23]: # View rows in twitter_archive which contain '&amp;' instead of '&' in 'text' column
         twitter_archive[twitter_archive.text.str.contains('&amp;')]

Out[23]:                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         262   842765311967449089                    NaN                  NaN
         273   840728873075638272                    NaN                  NaN
         320   834458053273591808                    NaN                  NaN
         461   817536400337801217                    NaN                  NaN
         485   814578408554463233                    NaN                  NaN
         516   810984652412424192                    NaN                  NaN
         799   772826264096874500                    NaN                  NaN
         889   759793422261743616                    NaN                  NaN
         898   758854675097526272                    NaN                  NaN
         976   750026558547456000                    NaN                  NaN
         1104  735137028879360001                    NaN                  NaN
         1179  719367763014393856                    NaN                  NaN
         1199  716791146589110272                    NaN                  NaN
         1222  714258258790387713                    NaN                  NaN
         1274  709198395643068416                    NaN                  NaN
         1366  702671118226825216                    NaN                  NaN
         1421  698195409219559425                    NaN                  NaN
         1465  694352839993344000                    NaN                  NaN
         1481  693280720173801472                    NaN                  NaN
         1508  691483041324204033                    NaN                  NaN
         1524  690597161306841088                    NaN                  NaN
         1538  689835978131935233                    NaN                  NaN
         1569  687807801670897665                    NaN                  NaN
         1593  686386521809772549                    NaN                  NaN
         1621  684926975086034944                    NaN                  NaN
         1646  683834909291606017                    NaN                  NaN
         1707  680801747103793152                    NaN                  NaN
         1763  678446151570427904                    NaN                  NaN
         1795  677314812125323265                    NaN                  NaN
         1812  676811746707918848                    NaN                  NaN
         1817  676603393314578432                    NaN                  NaN
         1842  675870721063669760           6.757073e+17         4.196984e+09
         1897  674737130913071104                    NaN                  NaN
         1899  674670581682434048                    NaN                  NaN
         1901  674646392044941312                    NaN                  NaN
         1913  674372068062928900                    NaN                  NaN
         1931  674036086168010753                    NaN                  NaN
         2031  671768281401958400                    NaN                  NaN
         2037  671561002136281088                    NaN                  NaN
         2064  671154572044468225                    NaN                  NaN
         2084  670807719151067136                    NaN                  NaN
         2096  670755717859713024                    NaN                  NaN
         2137  670046952931721218                    NaN                  NaN
         2177  669037058363662336                    NaN                  NaN
```

```
2190   668960084974809088                      NaN              NaN
2196   668852170888998912                      NaN              NaN
2207   668627278264475648                      NaN              NaN
2210   668620235289837568                      NaN              NaN
2216   668537837512433665                      NaN              NaN
2232   668221241640230912                      NaN              NaN
2246   667878741721415682                      NaN              NaN
2268   667517642048163840                      NaN              NaN
2293   667152164079423490                      NaN              NaN
2306   666835007768551424                      NaN              NaN


                        timestamp  \
262    2017-03-17 15:51:22 +0000
273    2017-03-12 00:59:17 +0000
320    2017-02-22 17:41:18 +0000
461    2017-01-07 01:00:41 +0000
485    2016-12-29 21:06:41 +0000
516    2016-12-19 23:06:23 +0000
799    2016-09-05 15:58:34 +0000
889    2016-07-31 16:50:42 +0000
898    2016-07-29 02:40:28 +0000
976    2016-07-04 18:00:41 +0000
1104   2016-05-24 15:55:00 +0000
1179   2016-04-11 03:33:34 +0000
1199   2016-04-04 00:55:01 +0000
1222   2016-03-28 01:10:13 +0000
1274   2016-03-14 02:04:08 +0000
1366   2016-02-25 01:47:04 +0000
1421   2016-02-12 17:22:12 +0000
1465   2016-02-02 02:53:12 +0000
1481   2016-01-30 03:52:58 +0000
1508   2016-01-25 04:49:38 +0000
1524   2016-01-22 18:09:28 +0000
1538   2016-01-20 15:44:48 +0000
1569   2016-01-15 01:25:33 +0000
1593   2016-01-11 03:17:53 +0000
1621   2016-01-07 02:38:10 +0000
1646   2016-01-04 02:18:42 +0000
1707   2015-12-26 17:25:59 +0000
1763   2015-12-20 05:25:42 +0000
1795   2015-12-17 02:30:09 +0000
1812   2015-12-15 17:11:09 +0000
1817   2015-12-15 03:23:14 +0000
1842   2015-12-13 02:51:51 +0000
1897   2015-12-09 23:47:22 +0000
1899   2015-12-09 19:22:56 +0000
1901   2015-12-09 17:46:48 +0000
1913   2015-12-08 23:36:44 +0000
```

```
1931    2015-12-08 01:21:40 +0000
2031    2015-12-01 19:10:13 +0000
2037    2015-12-01 05:26:34 +0000
2064    2015-11-30 02:31:34 +0000
2084    2015-11-29 03:33:17 +0000
2096    2015-11-29 00:06:39 +0000
2137    2015-11-27 01:10:17 +0000
2177    2015-11-24 06:17:19 +0000
2190    2015-11-24 01:11:27 +0000
2196    2015-11-23 18:02:38 +0000
2207    2015-11-23 03:09:00 +0000
2210    2015-11-23 02:41:01 +0000
2216    2015-11-22 21:13:35 +0000
2232    2015-11-22 00:15:33 +0000
2246    2015-11-21 01:34:35 +0000
2268    2015-11-20 01:39:42 +0000
2293    2015-11-19 01:27:25 +0000
2306    2015-11-18 04:27:09 +0000


                                                      source  \
262     <a href="http://twitter.com/download/iphone" r...
273     <a href="http://twitter.com/download/iphone" r...
320     <a href="http://twitter.com/download/iphone" r...
461     <a href="http://twitter.com/download/iphone" r...
485     <a href="http://twitter.com/download/iphone" r...
516     <a href="http://twitter.com/download/iphone" r...
799     <a href="http://twitter.com/download/iphone" r...
889     <a href="http://twitter.com/download/iphone" r...
898     <a href="http://twitter.com/download/iphone" r...
976     <a href="https://about.twitter.com/products/tw...
1104    <a href="http://twitter.com/download/iphone" r...
1179    <a href="http://twitter.com/download/iphone" r...
1199    <a href="http://twitter.com/download/iphone" r...
1222    <a href="http://twitter.com/download/iphone" r...
1274    <a href="http://twitter.com/download/iphone" r...
1366    <a href="http://twitter.com/download/iphone" r...
1421    <a href="http://twitter.com/download/iphone" r...
1465    <a href="http://twitter.com/download/iphone" r...
1481    <a href="http://twitter.com/download/iphone" r...
1508    <a href="http://twitter.com/download/iphone" r...
1524    <a href="http://twitter.com/download/iphone" r...
1538    <a href="http://twitter.com/download/iphone" r...
1569    <a href="http://twitter.com/download/iphone" r...
1593    <a href="http://twitter.com/download/iphone" r...
1621    <a href="http://twitter.com/download/iphone" r...
1646    <a href="http://twitter.com/download/iphone" r...
1707    <a href="http://twitter.com/download/iphone" r...
1763    <a href="http://twitter.com/download/iphone" r...
```

```
1795  <a href="http://twitter.com/download/iphone" r...
1812  <a href="http://twitter.com/download/iphone" r...
1817  <a href="http://twitter.com/download/iphone" r...
1842  <a href="http://twitter.com/download/iphone" r...
1897  <a href="http://twitter.com/download/iphone" r...
1899  <a href="http://twitter.com/download/iphone" r...
1901  <a href="http://twitter.com/download/iphone" r...
1913  <a href="http://twitter.com/download/iphone" r...
1931  <a href="http://twitter.com/download/iphone" r...
2031  <a href="http://twitter.com/download/iphone" r...
2037  <a href="http://twitter.com/download/iphone" r...
2064  <a href="http://twitter.com/download/iphone" r...
2084  <a href="http://twitter.com/download/iphone" r...
2096  <a href="http://twitter.com/download/iphone" r...
2137  <a href="http://twitter.com/download/iphone" r...
2177  <a href="http://twitter.com/download/iphone" r...
2190  <a href="http://twitter.com/download/iphone" r...
2196  <a href="http://twitter.com/download/iphone" r...
2207  <a href="http://twitter.com/download/iphone" r...
2210  <a href="http://twitter.com/download/iphone" r...
2216  <a href="http://twitter.com/download/iphone" r...
2232  <a href="http://twitter.com/download/iphone" r...
2246  <a href="http://twitter.com/download/iphone" r...
2268  <a href="http://twitter.com" rel="nofollow">Tw...
2293  <a href="http://twitter.com/download/iphone" r...
2306  <a href="http://twitter.com/download/iphone" r...


                                               text  retweeted_status_id  \
262   Meet Indie. She's not a fan of baths but she's...                  NaN
273   RT @dog_rates: This is Pipsy. He is a fluffbal...         6.671522e+17
320   Meet Chester (bottom) &amp; Harold (top). They...                  NaN
461   Say hello to Eugene &amp; Patti Melt. No matte...                  NaN
485   RT @dog_rates: Meet Beau &amp; Wilbur. Wilbur ...         6.981954e+17
516   Meet Sam. She smiles 24/7 &amp; secretly aspir...                  NaN
799   Meet Roosevelt. He's preparing for takeoff. Ma...                  NaN
889   Meet Maggie &amp; Lila. Maggie is the doggo, L...                  NaN
898   This is Lilli Bee &amp; Honey Bear. Unfortunat...                  NaN
976   Meet Jax &amp; Jil. Jil is yelling the pledge ...                  NaN
1104  Meet Buckley. His family &amp; some neighbors ...                  NaN
1179  Meet Sid &amp; Murphy. Murphy floats alongside...                  NaN
1199  Meet Jennifur. She's supposed to be navigating...                  NaN
1222  Meet Travis and Flurp. Travis is pretty chill ...                  NaN
1274  From left to right:\nCletus, Jerome, Alejandro...                  NaN
1366  Meet Rambo &amp; Kiwi. Rambo's the pup with th...                  NaN
1421  Meet Beau &amp; Wilbur. Wilbur stole Beau's be...                  NaN
1465  Meet Oliviér. He takes killer selfies. Has a d...                  NaN
1481  This is Sadie and her 2 pups Shebang &amp; Ruf...                  NaN
1508  When bae says they can't go out but you see th...                  NaN
```

```
1524  This is Lolo. She's America af. Behind in scie...                    NaN
1538  Meet Fynn &amp; Taco. Fynn is an all-powerful ...                    NaN
1569  Meet Trooper &amp; Maya. Trooper protects Maya...                    NaN
1593  Say hello to Crimson. He's a Speckled Winnebag...                    NaN
1621  Meet Bruiser &amp; Charlie. They are the best ...                    NaN
1646  Here we see a faulty pupper. Might need to rep...                    NaN
1707  Great picture here. Dog on the right panicked ...                    NaN
1763  Touching scene here. Really stirs up the emoti...                    NaN
1795  Meet Tassy &amp; Bee. Tassy is pretty chill, b...                    NaN
1812  Say hello to Penny &amp; Gizmo. They are pract...                    NaN
1817  This is Godzilla pupper. He had a ruff childho...                    NaN
1842  &amp; this is Yoshi. Another world record cont...                    NaN
1897  Meet Rufio. He is unaware of the pink legless ...                    NaN
1899  Meet Jeb &amp; Bush. Jeb is somehow stuck in t...                    NaN
1901  Two gorgeous dogs here. Little waddling dog is...                    NaN
1913  Meet Chesney. On the outside he stays calm &am...                    NaN
1931  Meet Daisy. She has no eyes &amp; her face has...                    NaN
2031  When you try to recreate the scene from Lady &...                    NaN
2037  This is the best thing I've ever seen so sprea...                    NaN
2064  Meet Holly. She's trying to teach small human-...                    NaN
2084  Say hello to Andy. He can balance on one foot,...                    NaN
2096  Say hello to Gin &amp; Tonic. They're having a...                    NaN
2137  This is Ben &amp; Carson. It's impossible for ...                    NaN
2177  Here we have Pancho and Peaches. Pancho is a C...                    NaN
2190  Meet Jaycob. He got scared of the vacuum. Hide...                    NaN
2196  Say hello to Bobb. Bobb is a Golden High Fescu...                    NaN
2207  This is Timofy. He's a pilot for Southwest. It...                    NaN
2210  Say hello to Kallie. There was a tornado in th...                    NaN
2216  This is Spark. He's nervous. Other dog hasn't ...                    NaN
2232  These two dogs are Bo &amp; Smittens. Smittens...                    NaN
2246  This is Tedrick. He lives on the edge. Needs s...                    NaN
2268  This is Dook &amp; Milo. Dook is struggling to...                    NaN
2293  This is Pipsy. He is a fluffball. Enjoys trave...                    NaN
2306  These are Peruvian Feldspars. Their names are ...                    NaN

      retweeted_status_user_id retweeted_status_timestamp  \
262                        NaN                        NaN
273               4.196984e+09  2015-11-19 01:27:25 +0000
320                        NaN                        NaN
461                        NaN                        NaN
485               4.196984e+09  2016-02-12 17:22:12 +0000
516                        NaN                        NaN
799                        NaN                        NaN
889                        NaN                        NaN
898                        NaN                        NaN
976                        NaN                        NaN
1104                       NaN                        NaN
1179                       NaN                        NaN
```

```
1199                            NaN                         NaN
1222                            NaN                         NaN
1274                            NaN                         NaN
1366                            NaN                         NaN
1421                            NaN                         NaN
1465                            NaN                         NaN
1481                            NaN                         NaN
1508                            NaN                         NaN
1524                            NaN                         NaN
1538                            NaN                         NaN
1569                            NaN                         NaN
1593                            NaN                         NaN
1621                            NaN                         NaN
1646                            NaN                         NaN
1707                            NaN                         NaN
1763                            NaN                         NaN
1795                            NaN                         NaN
1812                            NaN                         NaN
1817                            NaN                         NaN
1842                            NaN                         NaN
1897                            NaN                         NaN
1899                            NaN                         NaN
1901                            NaN                         NaN
1913                            NaN                         NaN
1931                            NaN                         NaN
2031                            NaN                         NaN
2037                            NaN                         NaN
2064                            NaN                         NaN
2084                            NaN                         NaN
2096                            NaN                         NaN
2137                            NaN                         NaN
2177                            NaN                         NaN
2190                            NaN                         NaN
2196                            NaN                         NaN
2207                            NaN                         NaN
2210                            NaN                         NaN
2216                            NaN                         NaN
2232                            NaN                         NaN
2246                            NaN                         NaN
2268                            NaN                         NaN
2293                            NaN                         NaN
2306                            NaN                         NaN

                                     expanded_urls  rating_numerator  \
262    https://www.gofundme.com/get-indie-home/,https...                12
273    https://twitter.com/dog_rates/status/667152164...                12
320    https://twitter.com/dog_rates/status/834458053...                12
461    https://twitter.com/dog_rates/status/817536400...                12
```

485	https://twitter.com/dog_rates/status/698195409...	9
516	https://www.gofundme.com/sams-smile,https://tw...	24
799	https://twitter.com/dog_rates/status/772826264...	11
889	https://twitter.com/dog_rates/status/759793422...	12
898	https://twitter.com/dog_rates/status/758854675...	11
976	https://twitter.com/dog_rates/status/750026558...	10
1104	https://twitter.com/dog_rates/status/735137028...	9
1179	https://twitter.com/dog_rates/status/719367763...	11
1199	https://twitter.com/dog_rates/status/716791146...	11
1222	https://twitter.com/dog_rates/status/714258258...	10
1274	https://twitter.com/dog_rates/status/709198395...	45
1366	https://twitter.com/dog_rates/status/702671118...	10
1421	https://twitter.com/dog_rates/status/698195409...	9
1465	https://twitter.com/dog_rates/status/694352839...	10
1481	https://twitter.com/dog_rates/status/693280720...	10
1508	https://twitter.com/dog_rates/status/691483041...	5
1524	https://twitter.com/dog_rates/status/690597161...	11
1538	https://twitter.com/dog_rates/status/689835978...	11
1569	https://twitter.com/dog_rates/status/687807801...	11
1593	https://twitter.com/dog_rates/status/686386521...	11
1621	https://twitter.com/dog_rates/status/684926975...	11
1646	https://twitter.com/dog_rates/status/683834909...	9
1707	https://twitter.com/dog_rates/status/680801747...	10
1763	https://twitter.com/dog_rates/status/678446151...	10
1795	https://twitter.com/dog_rates/status/677314812...	10
1812	https://twitter.com/dog_rates/status/676811746...	9
1817	https://twitter.com/dog_rates/status/676603393...	9
1842	https://twitter.com/dog_rates/status/675870721...	11
1897	https://twitter.com/dog_rates/status/674737130...	10
1899	https://twitter.com/dog_rates/status/674670581...	9
1901	https://twitter.com/dog_rates/status/674646392...	5
1913	https://twitter.com/dog_rates/status/674372068...	10
1931	https://twitter.com/dog_rates/status/674036086...	9
2031	https://twitter.com/dog_rates/status/671768281...	10
2037	https://twitter.com/dog_rates/status/671561002...	13
2064	https://twitter.com/dog_rates/status/671154572...	11
2084	https://twitter.com/dog_rates/status/670807719...	11
2096	https://twitter.com/dog_rates/status/670755717...	9
2137	https://twitter.com/dog_rates/status/670046952...	11
2177	https://twitter.com/dog_rates/status/669037058...	10
2190	https://twitter.com/dog_rates/status/668960084...	10
2196	https://twitter.com/dog_rates/status/668852170...	11
2207	https://twitter.com/dog_rates/status/668627278...	9
2210	https://twitter.com/dog_rates/status/668620235...	10
2216	https://twitter.com/dog_rates/status/668537837...	8
2232	https://twitter.com/dog_rates/status/668221241...	10
2246	https://twitter.com/dog_rates/status/667878741...	2
2268	https://twitter.com/dog_rates/status/667517642...	8

```
2293    https://twitter.com/dog_rates/status/667152164...                    12
2306    https://twitter.com/dog_rates/status/666835007...                    10

        rating_denominator      name   doggo floofer  pupper puppo
262                     10     Indie    None    None    None  None
273                     10     Pipsy    None    None    None  None
320                     10    Chester   None    None    None  None
461                     10     Eugene   None    None    None  None
485                     10      Beau    None    None    None  None
516                      7       Sam    None    None    None  None
799                     10  Roosevelt   None    None    None  None
889                     10     Maggie  doggo    None  pupper  None
898                     10     Lilli    None    None    None  None
976                     10       Jax    None    None    None  None
1104                    10    Buckley   None    None  pupper  None
1179                    10       Sid    None    None    None  None
1199                    10    Jennifur   None    None    None  None
1222                    10     Travis   None    None    None  None
1274                    50      None    None    None    None  None
1366                    10     Rambo    None    None    None  None
1421                    10      Beau    None    None    None  None
1465                    10    Oliviér   None    None    None  None
1481                    10     Sadie    None    None    None  None
1508                    10      None    None    None    None  None
1524                    10      Lolo    None    None    None  None
1538                    10      Fynn    None    None    None  None
1569                    10    Trooper   None    None    None  None
1593                    10    Crimson   None    None    None  None
1621                    10    Bruiser   None    None    None  None
1646                    10      None    None    None  pupper  None
1707                    10      None    None    None    None  None
1763                    10      None    None    None    None  None
1795                    10     Tassy    None    None    None  None
1812                    10     Penny    None    None    None  None
1817                    10   Godzilla   None    None  pupper  None
1842                    10      None    None    None    None  None
1897                    10     Rufio    None    None  pupper  None
1899                    10       Jeb    None    None    None  None
1901                    10      None    None    None    None  None
1913                    10    Chesney   None    None    None  None
1931                    10     Daisy    None    None    None  None
2031                    10      None    None    None    None  None
2037                    10       the    None    None    None  None
2064                    10     Holly    None    None    None  None
2084                    10      Andy    None    None    None  None
2096                    10       Gin    None    None    None  None
2137                    10       Ben    None    None    None  None
2177                    10      None    None    None    None  None
```

```
         2190                10    Jaycob   None    None    None  None
         2196                10      Bobb   None    None    None  None
         2207                10    Timofy   None    None    None  None
         2210                10    Kallie   None    None    None  None
         2216                10     Spark   None    None    None  None
         2232                10      None   None    None    None  None
         2246                10   Tedrick   None    None    None  None
         2268                10      Dook   None    None    None  None
         2293                10     Pipsy   None    None    None  None
         2306                10      None   None    None    None  None
```

In [24]: *# Sort values of 'name' column alphabetically - lowercase values appear at the bottom*
         twitter_archive.name.sort_values()

Out[24]: 1035   Abby
         1021   Abby
         938   Ace
         1933   Acro
         1327   Adele
         1934   Aiden
         77   Aja
         480   Akumi
         820   Al
         875   Albert
         1954   Albert
         144   Albus
         412   Albus
         1115   Aldrick
         2046   Alejandro
         374   Alexander
         1189   Alexanderson
         486   Alf
         2238   Alfie
         661   Alfie
         367   Alfie
         1616   Alfie
         858   Alfie
         51   Alfy
         201   Alice
         1701   Alice
         1495   Amber
         1334   Ambrose
         2078   Amy
         2195   Amélie
                          ...
         852   my
         988   not
         335   not
```

```
1747        officially
1206              old
993               one
1936              one
924               one
369               one
2326            quite
169             quite
118             quite
193             quite
2030            space
22               such
2345              the
2346              the
1603              the
2037              the
2212              the
1815              the
1797              the
1527              the
1120             this
1121     unacceptable
1031             very
773              very
1097             very
819              very
1385             very
Name: name, Length: 2356, dtype: object
```

In [25]: # View number of entries for each source
         twitter_archive.source.value_counts()

Out[25]: Twitter for iPhone
         Vine - Make a Scene
         Twitter Web Client
         TweetDeck
         Name: source, dtype: int64

In [26]: # View rows where the value of 'name' is lowercase - indicates that it is not an actual
         twitter_archive.loc[(twitter_archive['name'].str.islower())]

Out[26]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         22   887517139158093824   NaN   NaN
         56   881536004380872706   NaN   NaN
         118   869988702071779329   NaN   NaN
         169   859196978902773760   NaN   NaN
         193   855459453768019968   NaN   NaN
         335   832645525019123713   NaN   NaN
         369   828650029636317184   NaN   NaN
```

| | | | |
|---|---|---|---|
| 542 | 806219024703037440 | NaN | NaN |
| 649 | 792913359805018113 | NaN | NaN |
| 682 | 788552643979468800 | NaN | NaN |
| 759 | 778396591732486144 | NaN | NaN |
| 773 | 776249906839351296 | NaN | NaN |
| 801 | 772581559778025472 | NaN | NaN |
| 819 | 770655142660169732 | NaN | NaN |
| 822 | 770093767776997377 | NaN | NaN |
| 852 | 765395769549590528 | NaN | NaN |
| 924 | 755206590534418437 | NaN | NaN |
| 988 | 748977405889503236 | NaN | NaN |
| 992 | 748692773788876800 | NaN | NaN |
| 993 | 748575535303884801 | NaN | NaN |
| 1002 | 747885874273214464 | NaN | NaN |
| 1004 | 747816857231626240 | NaN | NaN |
| 1017 | 746872823977771008 | NaN | NaN |
| 1025 | 746369468511756288 | NaN | NaN |
| 1031 | 745422732645535745 | NaN | NaN |
| 1040 | 744223424764059648 | NaN | NaN |
| 1049 | 743222593470234624 | NaN | NaN |
| 1063 | 741067306818797568 | NaN | NaN |
| 1071 | 740214038584557568 | NaN | NaN |
| 1095 | 736392552031657984 | NaN | NaN |
| ... | ... | ... | ... |
| 2191 | 668955713004314625 | NaN | NaN |
| 2198 | 668815180734689280 | NaN | NaN |
| 2204 | 668636665813057536 | NaN | NaN |
| 2211 | 668614819948453888 | NaN | NaN |
| 2212 | 668587383441514497 | NaN | NaN |
| 2218 | 668507509523615744 | NaN | NaN |
| 2222 | 668466899341221888 | NaN | NaN |
| 2235 | 668171859951755264 | NaN | NaN |
| 2249 | 667861340749471744 | NaN | NaN |
| 2255 | 667773195014021121 | NaN | NaN |
| 2264 | 667538891197542400 | NaN | NaN |
| 2273 | 667470559035432960 | NaN | NaN |
| 2287 | 667177989038297088 | NaN | NaN |
| 2304 | 666983947667116034 | NaN | NaN |
| 2311 | 666781792255496192 | NaN | NaN |
| 2314 | 666701168228331520 | NaN | NaN |
| 2326 | 666411507551481857 | NaN | NaN |
| 2327 | 666407126856765440 | NaN | NaN |
| 2333 | 666337882303524864 | NaN | NaN |
| 2334 | 666293911632134144 | NaN | NaN |
| 2335 | 666287406224695296 | 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
2352  666044226329800704                    NaN              NaN
2353  666033412701032449                    NaN              NaN
2354  666029285002620928                    NaN              NaN


                        timestamp  \
22    2017-07-19 03:39:09 +0000
56    2017-07-02 15:32:16 +0000
118   2017-05-31 18:47:24 +0000
169   2017-05-02 00:04:57 +0000
193   2017-04-21 16:33:22 +0000
335   2017-02-17 17:38:57 +0000
369   2017-02-06 17:02:17 +0000
542   2016-12-06 19:29:28 +0000
649   2016-10-31 02:17:31 +0000
682   2016-10-19 01:29:35 +0000
759   2016-09-21 00:53:04 +0000
773   2016-09-15 02:42:54 +0000
801   2016-09-04 23:46:12 +0000
819   2016-08-30 16:11:18 +0000
822   2016-08-29 03:00:36 +0000
852   2016-08-16 03:52:26 +0000
924   2016-07-19 01:04:16 +0000
988   2016-07-01 20:31:43 +0000
992   2016-07-01 01:40:41 +0000
993   2016-06-30 17:54:50 +0000
1002  2016-06-28 20:14:22 +0000
1004  2016-06-28 15:40:07 +0000
1017  2016-06-26 01:08:52 +0000
1025  2016-06-24 15:48:42 +0000
1031  2016-06-22 01:06:43 +0000
1040  2016-06-18 17:41:06 +0000
1049  2016-06-15 23:24:09 +0000
1063  2016-06-10 00:39:48 +0000
1071  2016-06-07 16:09:13 +0000
1095  2016-05-28 03:04:00 +0000
...                         ...
2191  2015-11-24 00:54:05 +0000
2198  2015-11-23 15:35:39 +0000
2204  2015-11-23 03:46:18 +0000
2211  2015-11-23 02:19:29 +0000
2212  2015-11-23 00:30:28 +0000
2218  2015-11-22 19:13:05 +0000
2222  2015-11-22 16:31:42 +0000
2235  2015-11-21 20:59:20 +0000
2249  2015-11-21 00:25:26 +0000
```

```
2255   2015-11-20 18:35:10 +0000
2264   2015-11-20 03:04:08 +0000
2273   2015-11-19 22:32:36 +0000
2287   2015-11-19 03:10:02 +0000
2304   2015-11-18 14:18:59 +0000
2311   2015-11-18 00:55:42 +0000
2314   2015-11-17 19:35:19 +0000
2326   2015-11-17 00:24:19 +0000
2327   2015-11-17 00:06:54 +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
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
2352   2015-11-16 00:04:52 +0000
2353   2015-11-15 23:21:54 +0000
2354   2015-11-15 23:05:30 +0000

                                                source  \
22    <a href="http://twitter.com/download/iphone" r...
56    <a href="http://twitter.com/download/iphone" r...
118   <a href="http://twitter.com/download/iphone" r...
169   <a href="http://twitter.com/download/iphone" r...
193   <a href="http://twitter.com/download/iphone" r...
335   <a href="http://twitter.com" rel="nofollow">Tw...
369   <a href="http://twitter.com/download/iphone" r...
542   <a href="http://twitter.com/download/iphone" r...
649   <a href="http://twitter.com/download/iphone" r...
682   <a href="http://twitter.com/download/iphone" r...
759   <a href="http://twitter.com/download/iphone" r...
773   <a href="http://twitter.com/download/iphone" r...
801   <a href="http://twitter.com/download/iphone" r...
819   <a href="http://twitter.com/download/iphone" r...
822   <a href="http://twitter.com/download/iphone" r...
852   <a href="http://twitter.com/download/iphone" r...
924   <a href="http://twitter.com/download/iphone" r...
988   <a href="http://twitter.com/download/iphone" r...
992   <a href="http://twitter.com/download/iphone" r...
993   <a href="http://twitter.com/download/iphone" r...
1002  <a href="http://twitter.com/download/iphone" r...
1004  <a href="http://twitter.com/download/iphone" r...
1017  <a href="http://twitter.com/download/iphone" r...
1025  <a href="http://twitter.com/download/iphone" r...
1031  <a href="http://twitter.com/download/iphone" r...
```

```
1040  <a href="http://twitter.com/download/iphone" r...
1049  <a href="http://twitter.com/download/iphone" r...
1063  <a href="http://twitter.com/download/iphone" r...
1071  <a href="http://twitter.com/download/iphone" r...
1095  <a href="http://vine.co" rel="nofollow">Vine -...
...                                                 ...
2191  <a href="http://twitter.com/download/iphone" r...
2198  <a href="http://twitter.com/download/iphone" r...
2204  <a href="http://twitter.com/download/iphone" r...
2211  <a href="http://twitter.com/download/iphone" r...
2212  <a href="http://vine.co" rel="nofollow">Vine -...
2218  <a href="http://twitter.com/download/iphone" r...
2222  <a href="http://twitter.com/download/iphone" r...
2235  <a href="http://twitter.com/download/iphone" r...
2249  <a href="http://twitter.com/download/iphone" r...
2255  <a href="http://twitter.com" rel="nofollow">Tw...
2264  <a href="http://twitter.com" rel="nofollow">Tw...
2273  <a href="http://twitter.com" rel="nofollow">Tw...
2287  <a href="http://twitter.com/download/iphone" r...
2304  <a href="http://twitter.com/download/iphone" r...
2311  <a href="http://twitter.com/download/iphone" r...
2314  <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...
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...
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...
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...

                                                  text  retweeted_status_id  \
22     I've yet to rate a Venezuelan Hover Wiener. Th...                  NaN
56     Here is a pupper approaching maximum borkdrive...                  NaN
118    RT @dog_rates: We only rate dogs. This is quit...         8.591970e+17
169    We only rate dogs. This is quite clearly a smo...                  NaN
193    Guys, we only rate dogs. This is quite clearly...                  NaN
335    There's going to be a dog terminal at JFK Airp...                  NaN
369    Occasionally, we're sent fantastic stories. Th...                  NaN
542    We only rate dogs. Please stop sending in non-...                  NaN
649    Here is a perfect example of someone who has t...                  NaN
682    RT @dog_rates: Say hello to mad pupper. You kn...         7.363926e+17
```

```
759    RT @dog_rates: This is an East African Chalupa...    7.030419e+17
773    RT @dog_rates: We only rate dogs. Pls stop sen...    7.007478e+17
801    Guys this is getting so out of hand. We only r...          NaN
819    We only rate dogs. Pls stop sending in non-can...          NaN
822    RT @dog_rates: This is just downright precious...    7.410673e+17
852    This is my dog. Her name is Zoey. She knows I'...          NaN
924    This is one of the most inspirational stories ...          NaN
988    What jokester sent in a pic without a dog in i...          NaN
992    That is Quizno. This is his beach. He does not...          NaN
993    This is one of the most reckless puppers I've ...         NaN
1002   This is a mighty rare blue-tailed hammer sherk...          NaN
1004   Viewer discretion is advised. This is a terrib...          NaN
1017   This is a carrot. We only rate dogs. Please on...          NaN
1025   This is an Iraqi Speed Kangaroo. It is not a d...          NaN
1031   We only rate dogs. Pls stop sending in non-can...          NaN
1040   This is actually a pupper and I'd pet it so we...          NaN
1049   This is a very rare Great Alaskan Bush Pupper...           NaN
1063   This is just downright precious af. 12/10 for ...          NaN
1071   This is getting incredibly frustrating. This i...          NaN
1095   Say hello to mad pupper. You know what you did...          NaN
...                                          ...                  ...
2191   This is a Slovakian Helter Skelter Feta named ...          NaN
2198   This is a wild Toblerone from Papua New Guinea...          NaN
2204   This is an Irish Rigatoni terrier named Berta...           NaN
2211   Here is a horned dog. Much grace. Can jump ove...          NaN
2212   Never forget this vine. You will not stop watc...          NaN
2218   This is a Birmingham Quagmire named Chuk. Love...          NaN
2222   Here is a mother dog caring for her pups. Snaz...          NaN
2235   This is a Trans Siberian Kellogg named Alfonso...          NaN
2249   This is a Shotokon Macadamia mix named Cheryl...           NaN
2255   This is a rare Hungarian Pinot named Jessiga. ...          NaN
2264   This is a southwest Coriander named Klint. Hat...          NaN
2273   This is a northern Wahoo named Kohl. He runs t...          NaN
2287   This is a Dasani Kingfisher from Maine. His na...          NaN
2304   This is a curly Ticonderoga named Pepe. No fee...          NaN
2311   This is a purebred Bacardi named Octaviath. Ca...          NaN
2314   This is a golden Buckminsterfullerene named Jo...          NaN
2326   This is quite the dog. Gets really excited whe...          NaN
2327   This is a southern Vesuvius bumblegruff. Can d...          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
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
```

```
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


      retweeted_status_user_id retweeted_status_timestamp  \
22                         NaN                         NaN
56                         NaN                         NaN
118               4.196984e+09  2017-05-02 00:04:57 +0000
169                        NaN                         NaN
193                        NaN                         NaN
335                        NaN                         NaN
369                        NaN                         NaN
542                        NaN                         NaN
649                        NaN                         NaN
682               4.196984e+09  2016-05-28 03:04:00 +0000
759               4.196984e+09  2016-02-26 02:20:37 +0000
773               4.196984e+09  2016-02-19 18:24:26 +0000
801                        NaN                         NaN
819                        NaN                         NaN
822               4.196984e+09  2016-06-10 00:39:48 +0000
852                        NaN                         NaN
924                        NaN                         NaN
988                        NaN                         NaN
992                        NaN                         NaN
993                        NaN                         NaN
1002                       NaN                         NaN
1004                       NaN                         NaN
1017                       NaN                         NaN
1025                       NaN                         NaN
1031                       NaN                         NaN
1040                       NaN                         NaN
1049                       NaN                         NaN
1063                       NaN                         NaN
1071                       NaN                         NaN
1095                       NaN                         NaN
...                        ...                         ...
2191                       NaN                         NaN
2198                       NaN                         NaN
2204                       NaN                         NaN
2211                       NaN                         NaN
2212                       NaN                         NaN
2218                       NaN                         NaN
2222                       NaN                         NaN
2235                       NaN                         NaN
2249                       NaN                         NaN
2255                       NaN                         NaN
2264                       NaN                         NaN
2273                       NaN                         NaN
```

```
2287                      NaN                      NaN
2304                      NaN                      NaN
2311                      NaN                      NaN
2314                      NaN                      NaN
2326                      NaN                      NaN
2327                      NaN                      NaN
2333                      NaN                      NaN
2334                      NaN                      NaN
2335                      NaN                      NaN
2345                      NaN                      NaN
2346                      NaN                      NaN
2347                      NaN                      NaN
2348                      NaN                      NaN
2349                      NaN                      NaN
2350                      NaN                      NaN
2352                      NaN                      NaN
2353                      NaN                      NaN
2354                      NaN                      NaN

                                            expanded_urls  rating_numerator  \
22     https://twitter.com/dog_rates/status/887517139...                14
56     https://twitter.com/dog_rates/status/881536004...                14
118    https://twitter.com/dog_rates/status/859196978...                12
169    https://twitter.com/dog_rates/status/859196978...                12
193    https://twitter.com/dog_rates/status/855459453...                12
335    http://us.blastingnews.com/news/2017/02/jfk-an...                10
369    https://twitter.com/dog_rates/status/828650029...                14
542    https://twitter.com/dog_rates/status/806219024...                11
649    https://twitter.com/dog_rates/status/792913359...                13
682    https://vine.co/v/iEggaEOiLO3,https://vine.co/...                13
759    https://twitter.com/dog_rates/status/703041949...                10
773    https://twitter.com/dog_rates/status/700747788...                11
801    https://twitter.com/dog_rates/status/772581559...                10
819    https://twitter.com/dog_rates/status/770655142...                11
822    https://twitter.com/dog_rates/status/741067306...                12
852    https://twitter.com/dog_rates/status/765395769...                13
924    https://twitter.com/dog_rates/status/755206590...                14
988    https://twitter.com/dog_rates/status/748977405...                10
992    https://twitter.com/dog_rates/status/748692773...                10
993    https://twitter.com/dog_rates/status/748575535...                 6
1002   https://twitter.com/dog_rates/status/747885874...                 8
1004   https://twitter.com/dog_rates/status/747816857...                 4
1017   https://twitter.com/dog_rates/status/746872823...                11
1025   https://twitter.com/dog_rates/status/746369468...                 9
1031   https://twitter.com/dog_rates/status/745422732...                 9
1040   https://twitter.com/strange_animals/status/672...                12
1049   https://twitter.com/dog_rates/status/743222593...                12
1063   https://twitter.com/dog_rates/status/741067306...                12
```

```
1071     https://twitter.com/dog_rates/status/740214038...              10
1095                        https://vine.co/v/iEggaEOiL03              13
...                                                        ...          ...
2191     https://twitter.com/dog_rates/status/668955713...              10
2198     https://twitter.com/dog_rates/status/668815180...               7
2204     https://twitter.com/dog_rates/status/668636665...              10
2211     https://twitter.com/dog_rates/status/668614819...               7
2212                        https://vine.co/v/ea0OwvPTx9l              13
2218     https://twitter.com/dog_rates/status/668507509...              10
2222     https://twitter.com/dog_rates/status/668466899...               4
2235     https://twitter.com/dog_rates/status/668171859...               7
2249     https://twitter.com/dog_rates/status/667861340...               9
2255     https://twitter.com/dog_rates/status/667773195...               8
2264     https://twitter.com/dog_rates/status/667538891...               9
2273     https://twitter.com/dog_rates/status/667470559...              11
2287     https://twitter.com/dog_rates/status/667177989...               8
2304     https://twitter.com/dog_rates/status/666983947...              11
2311     https://twitter.com/dog_rates/status/666781792...              10
2314     https://twitter.com/dog_rates/status/666701168...               8
2326     https://twitter.com/dog_rates/status/666411507...               2
2327     https://twitter.com/dog_rates/status/666407126...               7
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
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
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

      rating_denominator        name  doggo floofer  pupper puppo
22                    10        such   None    None    None  None
56                    10           a   None    None  pupper  None
118                   10       quite   None    None    None  None
169                   10       quite   None    None    None  None
193                   10       quite   None    None    None  None
335                   10         not   None    None    None  None
369                   10         one   None    None    None  None
542                   10  incredibly   None    None    None  None
649                   10           a   None    None    None  None
682                   10         mad   None    None  pupper  None
759                   10          an   None    None    None  None
773                   10        very   None    None    None  None
801                   10           a   None    None    None  None
```

819	10	very	None	None	None	None
822	10	just	doggo	None	pupper	None
852	10	my	None	None	None	None
924	10	one	doggo	None	None	None
988	10	not	None	None	None	None
992	10	his	doggo	None	None	None
993	10	one	None	None	None	None
1002	10	a	None	None	None	None
1004	10	a	None	None	None	None
1017	10	a	None	None	None	None
1025	10	an	None	None	None	None
1031	10	very	None	None	None	None
1040	10	actually	None	None	pupper	None
1049	10	a	None	None	pupper	None
1063	10	just	doggo	None	pupper	None
1071	10	getting	None	None	None	None
1095	10	mad	None	None	pupper	None
...	...	...	...	...	...	...
2191	10	a	None	None	None	None
2198	10	a	None	None	None	None
2204	10	an	None	None	None	None
2211	10	a	None	None	None	None
2212	10	the	None	None	None	None
2218	10	a	None	None	None	None
2222	10	a	None	None	None	None
2235	10	a	None	None	None	None
2249	10	a	None	None	None	None
2255	10	a	None	None	None	None
2264	10	a	None	None	None	None
2273	10	a	None	None	None	None
2287	10	a	None	None	None	None
2304	10	a	None	None	None	None
2311	10	a	None	None	None	None
2314	10	a	None	None	None	None
2326	10	quite	None	None	None	None
2327	10	a	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
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
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None

```
          [109 rows x 17 columns]

In [27]: # View rows where the value of 'name' is lowercase and the word 'named' appears in the
         # there is an actual dog name in the text
         twitter_archive.loc[(twitter_archive['name'].str.islower()) & (twitter_archive['text'].

Out[27]:                    tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         1853  675706639471788032                    NaN                  NaN
         1955  673636718965334016                    NaN                  NaN
         2034  671743150407421952                    NaN                  NaN
         2066  671147085991960577                    NaN                  NaN
         2116  670427002554466305                    NaN                  NaN
         2125  670361874861563904                    NaN                  NaN
         2128  670303360680108032                    NaN                  NaN
         2146  669923323644657664                    NaN                  NaN
         2161  669564461267722241                    NaN                  NaN
         2191  668955713004314625                    NaN                  NaN
         2204  668636665813057536                    NaN                  NaN
         2218  668507509523615744                    NaN                  NaN
         2235  668171859951755264                    NaN                  NaN
         2249  667861340749471744                    NaN                  NaN
         2255  667773195014021121                    NaN                  NaN
         2264  667538891197542400                    NaN                  NaN
         2273  667470559035432960                    NaN                  NaN
         2304  666983947667116034                    NaN                  NaN
         2311  666781792255496192                    NaN                  NaN
         2314  666701168228331520                    NaN                  NaN

                                 timestamp  \
         1853  2015-12-12 15:59:51 +0000
         1955  2015-12-06 22:54:44 +0000
         2034  2015-12-01 17:30:22 +0000
         2066  2015-11-30 02:01:49 +0000
         2116  2015-11-28 02:20:27 +0000
         2125  2015-11-27 22:01:40 +0000
         2128  2015-11-27 18:09:09 +0000
         2146  2015-11-26 16:59:01 +0000
         2161  2015-11-25 17:13:02 +0000
         2191  2015-11-24 00:54:05 +0000
         2204  2015-11-23 03:46:18 +0000
         2218  2015-11-22 19:13:05 +0000
         2235  2015-11-21 20:59:20 +0000
         2249  2015-11-21 00:25:26 +0000
         2255  2015-11-20 18:35:10 +0000
         2264  2015-11-20 03:04:08 +0000
         2273  2015-11-19 22:32:36 +0000
         2304  2015-11-18 14:18:59 +0000
```

```
2311   2015-11-18 00:55:42 +0000
2314   2015-11-17 19:35:19 +0000


                                                    source  \
1853   <a href="http://twitter.com/download/iphone" r...
1955   <a href="http://twitter.com/download/iphone" r...
2034   <a href="http://twitter.com/download/iphone" r...
2066   <a href="http://twitter.com/download/iphone" r...
2116   <a href="http://twitter.com/download/iphone" r...
2125   <a href="http://twitter.com/download/iphone" r...
2128   <a href="http://twitter.com/download/iphone" r...
2146   <a href="http://twitter.com/download/iphone" r...
2161   <a href="http://twitter.com/download/iphone" r...
2191   <a href="http://twitter.com/download/iphone" r...
2204   <a href="http://twitter.com/download/iphone" r...
2218   <a href="http://twitter.com/download/iphone" r...
2235   <a href="http://twitter.com/download/iphone" r...
2249   <a href="http://twitter.com/download/iphone" r...
2255   <a href="http://twitter.com" rel="nofollow">Tw...
2264   <a href="http://twitter.com" rel="nofollow">Tw...
2273   <a href="http://twitter.com" rel="nofollow">Tw...
2304   <a href="http://twitter.com/download/iphone" r...
2311   <a href="http://twitter.com/download/iphone" r...
2314   <a href="http://twitter.com/download/iphone" r...


                                                    text  retweeted_status_id  \
1853   This is a Sizzlin Menorah spaniel from Brookly...                  NaN
1955   This is a Lofted Aphrodisiac Terrier named Kip...                  NaN
2034   This is a Tuscaloosa Alcatraz named Jacob (Yac...                  NaN
2066   This is a Helvetica Listerine named Rufus. Thi...                  NaN
2116   This is a Deciduous Trimester mix named Spork...                   NaN
2125   This is a Rich Mahogany Seltzer named Cherokee...                  NaN
2128   This is a Speckled Cauliflower Yosemite named ...                  NaN
2146   This is a spotted Lipitor Rumpelstiltskin name...                  NaN
2161   This is a Coriander Baton Rouge named Alfredo...                   NaN
2191   This is a Slovakian Helter Skelter Feta named ...                  NaN
2204   This is an Irish Rigatoni terrier named Berta...                   NaN
2218   This is a Birmingham Quagmire named Chuk. Love...                  NaN
2235   This is a Trans Siberian Kellogg named Alfonso...                  NaN
2249   This is a Shotokon Macadamia mix named Cheryl...                   NaN
2255   This is a rare Hungarian Pinot named Jessiga. ...                  NaN
2264   This is a southwest Coriander named Klint. Hat...                  NaN
2273   This is a northern Wahoo named Kohl. He runs t...                  NaN
2304   This is a curly Ticonderoga named Pepe. No fee...                  NaN
2311   This is a purebred Bacardi named Octaviath. Ca...                  NaN
2314   This is a golden Buckminsterfullerene named Jo...                  NaN


       retweeted_status_user_id retweeted_status_timestamp  \
```

```
1853                           NaN                    NaN
1955                           NaN                    NaN
2034                           NaN                    NaN
2066                           NaN                    NaN
2116                           NaN                    NaN
2125                           NaN                    NaN
2128                           NaN                    NaN
2146                           NaN                    NaN
2161                           NaN                    NaN
2191                           NaN                    NaN
2204                           NaN                    NaN
2218                           NaN                    NaN
2235                           NaN                    NaN
2249                           NaN                    NaN
2255                           NaN                    NaN
2264                           NaN                    NaN
2273                           NaN                    NaN
2304                           NaN                    NaN
2311                           NaN                    NaN
2314                           NaN                    NaN

                                       expanded_urls  rating_numerator  \
1853  https://twitter.com/dog_rates/status/675706639...                10
1955  https://twitter.com/dog_rates/status/673636718...                10
2034  https://twitter.com/dog_rates/status/671743150...                11
2066  https://twitter.com/dog_rates/status/671147085...                 9
2116  https://twitter.com/dog_rates/status/670427002...                 9
2125  https://twitter.com/dog_rates/status/670361874...                 9
2128  https://twitter.com/dog_rates/status/670303360...                 9
2146  https://twitter.com/dog_rates/status/669923323...                10
2161  https://twitter.com/dog_rates/status/669564461...                10
2191  https://twitter.com/dog_rates/status/668955713...                10
2204  https://twitter.com/dog_rates/status/668636665...                10
2218  https://twitter.com/dog_rates/status/668507509...                10
2235  https://twitter.com/dog_rates/status/668171859...                 7
2249  https://twitter.com/dog_rates/status/667861340...                 9
2255  https://twitter.com/dog_rates/status/667773195...                 8
2264  https://twitter.com/dog_rates/status/667538891...                 9
2273  https://twitter.com/dog_rates/status/667470559...                11
2304  https://twitter.com/dog_rates/status/666983947...                11
2311  https://twitter.com/dog_rates/status/666781792...                10
2314  https://twitter.com/dog_rates/status/666701168...                 8

      rating_denominator name doggo floofer pupper puppo
1853                  10    a  None    None   None  None
1955                  10    a  None    None   None  None
2034                  10    a  None    None   None  None
2066                  10    a  None    None   None  None
```

```
2116                        10     a  None    None   None  None
2125                        10     a  None    None   None  None
2128                        10     a  None    None   None  None
2146                        10     a  None    None   None  None
2161                        10     a  None    None   None  None
2191                        10     a  None    None   None  None
2204                        10    an  None    None   None  None
2218                        10     a  None    None   None  None
2235                        10     a  None    None   None  None
2249                        10     a  None    None   None  None
2255                        10     a  None    None   None  None
2264                        10     a  None    None   None  None
2273                        10     a  None    None   None  None
2304                        10     a  None    None   None  None
2311                        10     a  None    None   None  None
2314                        10     a  None    None   None  None
```

In [28]: *# View rows where the value of 'name' is lowercase and the words 'name is' appears in t*
 # there is an actual dog name in the text
`twitter_archive.loc[(twitter_archive['name'].str.islower()) & (twitter_archive['text'].`

Out[28]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         852   765395769549590528   NaN   NaN
         2287 667177989038297088   NaN   NaN

                               timestamp \
         852   2016-08-16 03:52:26 +0000
         2287   2015-11-19 03:10:02 +0000

                                                     source \
         852   <a href="http://twitter.com/download/iphone" r...
         2287   <a href="http://twitter.com/download/iphone" r...

                                                       text retweeted_status_id \
         852   This is my dog. Her name is Zoey. She knows I'...   NaN
         2287   This is a Dasani Kingfisher from Maine. His na...   NaN

                retweeted_status_user_id retweeted_status_timestamp \
         852   NaN   NaN
         2287   NaN   NaN

                                           expanded_urls rating_numerator \
         852   https://twitter.com/dog_rates/status/765395769...   13
         2287   https://twitter.com/dog_rates/status/667177989...   8

                rating_denominator name doggo floofer pupper puppo
         852   10   my None   None   None None
         2287   10   a None   None   None None

32

```
In [29]: # View row where dog name is 'O' but we can see in the 'text' column that the dog's nam
         twitter_archive[twitter_archive.name == "O"]

Out[29]:               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         775  776201521193218049                    NaN                  NaN

                          timestamp  \
         775  2016-09-14 23:30:38 +0000

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

                                                          text  retweeted_status_id  \
         775  This is O'Malley. That is how he sleeps. Doesn...                  NaN

              retweeted_status_user_id retweeted_status_timestamp  \
         775                       NaN                        NaN

                                            expanded_urls  rating_numerator  \
         775  https://twitter.com/dog_rates/status/776201521...                10

              rating_denominator name doggo floofer pupper puppo
         775                  10    O  None    None   None  None

In [30]: #disable warnings
         warnings.simplefilter('ignore')

         # View rows where text column contains #.#/# indicating a decimal for the rating numera
         # however they do not appear in the 'rating_numerator' column
         twitter_archive[twitter_archive.text.str.contains(r"(\d+\.\d*\/\d+)")]

Out[30]:               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         45     883482846933004288                    NaN                  NaN
         340    832215909146226688                    NaN                  NaN
         695    786709082849828864                    NaN                  NaN
         763    778027034220126208                    NaN                  NaN
         1689   681340665377193984           6.813394e+17         4.196984e+09
         1712   680494726643068929                    NaN                  NaN

                          timestamp  \
         45     2017-07-08 00:28:19 +0000
         340    2017-02-16 13:11:49 +0000
         695    2016-10-13 23:23:56 +0000
         763    2016-09-20 00:24:34 +0000
         1689   2015-12-28 05:07:27 +0000
         1712   2015-12-25 21:06:00 +0000

                                                        source  \
```

```
45     <a href="http://twitter.com/download/iphone" r...
340    <a href="http://twitter.com/download/iphone" r...
695    <a href="http://twitter.com/download/iphone" r...
763    <a href="http://twitter.com/download/iphone" r...
1689   <a href="http://twitter.com/download/iphone" r...
1712   <a href="http://twitter.com/download/iphone" r...

                                             text  retweeted_status_id  \
45     This is Bella. She hopes her smile made you sm...                NaN
340    RT @dog_rates: This is Logan, the Chow who liv...       7.867091e+17
695    This is Logan, the Chow who lived. He solemnly...                NaN
763    This is Sophie. She's a Jubilant Bush Pupper. ...                NaN
1689   I've been told there's a slight possibility he...                NaN
1712   Here we have uncovered an entire battalion of ...                NaN

       retweeted_status_user_id retweeted_status_timestamp  \
45                          NaN                        NaN
340                4.196984e+09    2016-10-13 23:23:56 +0000
695                         NaN                        NaN
763                         NaN                        NaN
1689                        NaN                        NaN
1712                        NaN                        NaN

                                       expanded_urls  rating_numerator  \
45     https://twitter.com/dog_rates/status/883482846...                 5
340    https://twitter.com/dog_rates/status/786709082...                75
695    https://twitter.com/dog_rates/status/786709082...                75
763    https://twitter.com/dog_rates/status/778027034...                27
1689                                             NaN                 5
1712   https://twitter.com/dog_rates/status/680494726...                26

       rating_denominator    name doggo floofer  pupper  puppo
45                     10   Bella  None    None    None   None
340                    10   Logan  None    None    None   None
695                    10   Logan  None    None    None   None
763                    10  Sophie  None    None  pupper   None
1689                   10    None  None    None    None   None
1712                   10    None  None    None    None   None

In [31]:  # View row of specific tweet using tweet_id of a tweet that doesn't have a rating
          twitter_archive[twitter_archive.tweet_id == 810984652412424192]

Out[31]:            tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
          516  810984652412424192                    NaN                  NaN

                          timestamp  \
          516  2016-12-19 23:06:23 +0000
```

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

                                                           text  retweeted_status_id  \
        516  Meet Sam. She smiles 24/7 &amp; secretly aspir...                  NaN

             retweeted_status_user_id retweeted_status_timestamp  \
        516                       NaN                        NaN

                                           expanded_urls  rating_numerator  \
        516  https://www.gofundme.com/sams-smile,https://tw...                24

             rating_denominator name doggo floofer pupper puppo
        516                   7  Sam  None    None   None  None
```

1 Quality

-Tweets without any images
   -Contents of 'text' cutoff
   -Dataset contains retweets
   -Incorrect dog names
   -Erroneous datatypes (tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, dog stages)
   -Missing values in 'name' and dog stages showing as 'None'

- Rating numerators with decimals not showing full float

-Tweet ID 810984652412424192 doesn't have a rating
   -Extra characters after '&'
   -Sources difficult to read
   -Tweet with more than one #/# sometimes have the first occurence erroneously used for the rating numerators and -denominators

2 Tidiness

- Dog "stage" variable in four columns: doggo, floofer, pupper, puppo

- Join 'tweet_info' and 'image_predictions' to 'twitter_archive'

3 Clean

```
In [32]: # Create copies of original DataFrames to work off of
         twitter_archive_clean = twitter_archive.copy()
         image_predictions_clean = df_img_pred.copy()
         tweet_info_clean = tweet_info.copy()
```

Define
Create dog stage variable and remove individual dog stage columns

4 Code

```
In [33]: # Create 'dog_stage' variable which is made by extracting the dog stage variables from
         twitter_archive_clean['dog_stage'] = twitter_archive_clean['text'].str.extract('(puppo|
```

```
In [34]: # Create variable of columns that are no longer needed and drop them from the DataFrame
         columns = ['doggo', 'floofer', 'pupper', 'puppo']
         twitter_archive_clean = twitter_archive_clean.drop(columns, axis=1)
```

5 Test

```
In [35]: twitter_archive_clean.head(20)
```

```
Out[35]:            tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0   892420643555336193                    NaN                  NaN
         1   892177421306343426                    NaN                  NaN
         2   891815181378084864                    NaN                  NaN
         3   891689557279858688                    NaN                  NaN
         4   891327558926688256                    NaN                  NaN
         5   891087950875897856                    NaN                  NaN
         6   890971913173991426                    NaN                  NaN
         7   890729181411237888                    NaN                  NaN
         8   890609185150312448                    NaN                  NaN
         9   890240255349198849                    NaN                  NaN
         10  890006608113172480                    NaN                  NaN
         11  889880896479866881                    NaN                  NaN
         12  889665388333682689                    NaN                  NaN
         13  889638837579907072                    NaN                  NaN
         14  889531135344209921                    NaN                  NaN
         15  889278841981685760                    NaN                  NaN
         16  888917238123831296                    NaN                  NaN
         17  888804989199671297                    NaN                  NaN
         18  888554962724278272                    NaN                  NaN
         19  888202515573088257                    NaN                  NaN

                           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


                                            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...


                                            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

    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

                                          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
```

```
        rating_denominator      name dog_stage
0                        10   Phineas       NaN
1                        10     Tilly       NaN
2                        10    Archie       NaN
3                        10     Darla       NaN
4                        10  Franklin       NaN
5                        10      None       NaN
6                        10       Jax       NaN
7                        10      None       NaN
8                        10      Zoey       NaN
9                        10    Cassie     doggo
10                       10      Koda       NaN
11                       10     Bruno       NaN
12                       10      None     puppo
13                       10       Ted       NaN
14                       10    Stuart     puppo
15                       10    Oliver       NaN
16                       10       Jim       NaN
17                       10      Zeke       NaN
18                       10   Ralphus       NaN
19                       10    Canela       NaN
```

Define
Add tweet_info and image_predictions to twitter_archive table.

6   Code

```
In [36]: twitter_archive_clean = pd.merge(left=twitter_archive_clean, right=tweet_info_clean, le

In [37]: twitter_archive_clean = twitter_archive_clean.merge(image_predictions_clean, on='tweet_

In [38]: twitter_archive_clean = twitter_archive_clean.drop('id', axis=1)
```

7   Test

```
In [39]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 6743 entries, 0 to 6742
Data columns (total 27 columns):
tweet_id                      6743 non-null int64
in_reply_to_status_id         36 non-null float64
in_reply_to_user_id           36 non-null float64
timestamp                     6743 non-null object
source                        6743 non-null object
text                          6743 non-null object
```

```
retweeted_status_id           617 non-null float64
retweeted_status_user_id      617 non-null float64
retweeted_status_timestamp    617 non-null object
expanded_urls                 6743 non-null object
rating_numerator              6743 non-null int64
rating_denominator            6743 non-null int64
name                          6743 non-null object
dog_stage                     1129 non-null object
retweet_count                 6743 non-null int64
favorite_count                6743 non-null int64
jpg_url                       6743 non-null object
img_num                       6743 non-null int64
p1                            6743 non-null object
p1_conf                       6743 non-null float64
p1_dog                        6743 non-null bool
p2                            6743 non-null object
p2_conf                       6743 non-null float64
p2_dog                        6743 non-null bool
p3                            6743 non-null object
p3_conf                       6743 non-null float64
p3_dog                        6743 non-null bool
dtypes: bool(3), float64(7), int64(6), object(11)
memory usage: 1.3+ MB
```

8 Define

Remove rows where there are no images (expanded_urls).

9 Code

```
In [40]: twitter_archive_clean = twitter_archive_clean.dropna(subset=['expanded_urls'])
```

10 Test

```
In [41]: sum(twitter_archive_clean['expanded_urls'].isnull())

Out[41]: 0
```

11 Define

Remove retweets.

12 Code

13 Test

In [43]: twitter_archive_clean.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6126 entries, 0 to 6742
Data columns (total 27 columns):
tweet_id                    6126 non-null int64
in_reply_to_status_id       36 non-null float64
in_reply_to_user_id         36 non-null float64
timestamp                   6126 non-null object
source                      6126 non-null object
text                        6126 non-null object
retweeted_status_id         0 non-null float64
retweeted_status_user_id    0 non-null float64
retweeted_status_timestamp  0 non-null object
expanded_urls               6126 non-null object
rating_numerator            6126 non-null int64
rating_denominator          6126 non-null int64
name                        6126 non-null object
dog_stage                   1029 non-null object
retweet_count               6126 non-null int64
favorite_count              6126 non-null int64
jpg_url                     6126 non-null object
img_num                     6126 non-null int64
p1                          6126 non-null object
p1_conf                     6126 non-null float64
p1_dog                      6126 non-null bool
p2                          6126 non-null object
p2_conf                     6126 non-null float64
p2_dog                      6126 non-null bool
p3                          6126 non-null object
p3_conf                     6126 non-null float64
p3_dog                      6126 non-null bool
dtypes: bool(3), float64(7), int64(6), object(11)
memory usage: 1.2+ MB
```

14 Define

Remove retweeted columns.

15 Code

```
In [44]: # Update columns variable and drop columns related to retweets
         columns = ['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_status_timesta
         twitter_archive_clean = twitter_archive_clean.drop(columns, axis=1)
```

16 Test

```
In [45]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 6126 entries, 0 to 6742
Data columns (total 24 columns):
tweet_id                6126 non-null int64
in_reply_to_status_id   36 non-null float64
in_reply_to_user_id     36 non-null float64
timestamp               6126 non-null object
source                  6126 non-null object
text                    6126 non-null object
expanded_urls           6126 non-null object
rating_numerator        6126 non-null int64
rating_denominator      6126 non-null int64
name                    6126 non-null object
dog_stage               1029 non-null object
retweet_count           6126 non-null int64
favorite_count          6126 non-null int64
jpg_url                 6126 non-null object
img_num                 6126 non-null int64
p1                      6126 non-null object
p1_conf                 6126 non-null float64
p1_dog                  6126 non-null bool
p2                      6126 non-null object
p2_conf                 6126 non-null float64
p2_dog                  6126 non-null bool
p3                      6126 non-null object
p3_conf                 6126 non-null float64
p3_dog                  6126 non-null bool
dtypes: bool(3), float64(5), int64(6), object(10)
memory usage: 1.0+ MB
```

17 Define

Display full content of 'text' column.

18 Code

```
In [46]: # Set column width to infinite so entire content of 'text' column is displayed
         pd.set_option('display.max_colwidth', -1)
```

19 Test

```
In [47]: twitter_archive_clean.sample(5)
```

```
Out[47]:                    tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         1896  842163532590374912 NaN                        NaN
         3360  816450570814898180 NaN                        NaN
         951   867900495410671616 NaN                        NaN
         4336  798209839306514432 NaN                        NaN
         3461  814530161257443328 NaN                        NaN


                            timestamp  \
         1896  2017-03-16 00:00:07 +0000
         3360  2017-01-04 01:05:59 +0000
         951   2017-05-26 00:29:37 +0000
         4336  2016-11-14 17:03:50 +0000
         3461  2016-12-29 17:54:58 +0000


                                                                                 sourc
         1896  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
         3360  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
         951   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
         4336  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
         3461  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


         1896  Meet George. He looks slightly deflated but overall quite powerful. Not sure how
         3360  Meet Moose. He doesn't want his friend to go back to college. 13/10 looks like yo
         951   Unbelievable. We only rate dogs. Please don't send in non-canines like the "I" fr
         4336  This is Cooper. His bow tie was too heavy for the front so he moved it to the sid
         3461  This is Alf. Someone just rubbed a balloon on his head. He's only a little pupset


         1896  https://twitter.com/dog_rates/status/842163532590374912/photo/1,https://twitter.c
         3360  https://twitter.com/dog_rates/status/816450570814898180/photo/1,https://twitter.c
         951   https://twitter.com/dog_rates/status/867900495410671616/photo/1
         4336  https://twitter.com/dog_rates/status/798209839306514432/photo/1
         3461  https://twitter.com/dog_rates/status/814530161257443328/photo/1


               rating_numerator  rating_denominator    name    ...    img_num  \
         1896  12                10                    George  ...    2
         3360  13                10                    Moose   ...    1
         951   12                10                    None    ...    1
```

```
4336  13                  10                    Cooper   ...    1
3461  12                  10                    Alf      ...    1

                            p1      p1_conf  p1_dog            p2      p2_conf  p2_dog  \
1896   French_bulldog          0.891227  True    soccer_ball  0.022811  False
3360   web_site                0.352857  False   envelope     0.060107  False
951    Labrador_retriever      0.522644  True    kuvasz       0.332461  True
4336   Pekinese                0.524583  True    Shih-Tzu     0.102931  True
3461   miniature_poodle        0.626913  True    toy_poodle   0.265582  True

                              p3      p3_conf  p3_dog
1896   bull_mastiff                 0.012852  True
3360   nail                         0.031291  False
951    dalmatian                    0.032008  True
4336   Pomeranian                   0.097893  True
3461   soft-coated_wheaten_terrier  0.041614  True

[5 rows x 24 columns]
```

Define

Change incorrect dog names.

20   Code

```python
In [48]: # Save locations where 'name' column is lowercase, lowercase and 'text' column contains
         # column contains the words 'name is'
         named_to_replace = twitter_archive_clean.loc[(twitter_archive_clean['name'].str.islower
         name_is_to_replace = twitter_archive_clean.loc[(twitter_archive_clean['name'].str.islow
         not_named_to_replace = twitter_archive_clean.loc[(twitter_archive_clean['name'].str.isl

         # Save these locations as lists
         named_to_replace_list = named_to_replace['text'].tolist()
         name_is_to_replace_list = name_is_to_replace['text'].tolist()
         not_named_to_replace_list = not_named_to_replace['text'].tolist()

         # For loop to iterate through locations where name is lowercase and the words 'named' a
         # value to be the word that appears after 'named'
         for entry in named_to_replace_list:
             mask = twitter_archive_clean.text == entry
             name_column = 'name'
             twitter_archive_clean.loc[mask, name_column] = re.findall(r"named\s(\w+)", entry)

         # For loop to iterate through locations where name is lowercase and the words 'name is'
         # value to be the word that appears after 'name is'
         for entry in name_is_to_replace_list:
             mask = twitter_archive_clean.text == entry
             name_column = 'name'
```

```
            twitter_archive_clean.loc[mask, name_column] = re.findall(r"name is\s(\w+)", entry)

            # For loop to iterate through locations where name is lowercase and replace the name va
            for entry in not_named_to_replace_list:
                mask = twitter_archive_clean.text == entry
                name_column = 'name'
                twitter_archive_clean.loc[mask, name_column] = "None"
```

In [49]: # Replace the occurence of "O" with "O'Malley"
         twitter_archive_clean.name = twitter_archive_clean.name.replace("O", "O'Malley")

21   Test

In [50]: twitter_archive_clean.name.sort_values()

Out[50]: 6600   Acro
         600   Aja
         602   Aja
         596   Aja
         597   Aja
         598   Aja
         599   Aja
         595   Aja
         601   Aja
         594   Aja
         3416   Akumi
         3411   Akumi
         3415   Akumi
         3419   Akumi
         3418   Akumi
         3412   Akumi
         3413   Akumi
         3417   Akumi
         3414   Akumi
         5939   Al
         5932   Al
         5933   Al
         5934   Al
         5936   Al
         5937   Al
         5931   Al
         5935   Al
         5938   Al
         6364   Albert
         6365   Albert
                  ...

                          45
```

```
       1332    Zeke
       153     Zeke
       1336    Zeke
       855     Zoey
       856     Zoey
       858     Zoey
       859     Zoey
       860     Zoey
       861     Zoey
       862     Zoey
       863     Zoey
       857     Zoey
       73      Zoey
       79      Zoey
       72      Zoey
       75      Zoey
       76      Zoey
       78      Zoey
       74      Zoey
       80      Zoey
       77      Zoey
       1121    Zooey
       1117    Zooey
       1116    Zooey
       1122    Zooey
       1118    Zooey
       1119    Zooey
       1123    Zooey
       1120    Zooey
       1124    Zooey
       Name: name, Length: 6126, dtype: object

In [51]: twitter_archive_clean.loc[(twitter_archive_clean['name'].str.islower())]

Out[51]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text
         Index: []

         [0 rows x 24 columns]

In [52]: twitter_archive_clean[twitter_archive_clean.name == "O'Malley"]

Out[52]:                  tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         5571   776201521193218049 NaN                    NaN
         5572   776201521193218049 NaN                    NaN
         5573   776201521193218049 NaN                    NaN
         5574   776201521193218049 NaN                    NaN
         5575   776201521193218049 NaN                    NaN
         5576   776201521193218049 NaN                    NaN
```

```
5577  776201521193218049 NaN                    NaN
5578  776201521193218049 NaN                    NaN
5579  776201521193218049 NaN                    NaN


                        timestamp  \
5571  2016-09-14 23:30:38 +0000
5572  2016-09-14 23:30:38 +0000
5573  2016-09-14 23:30:38 +0000
5574  2016-09-14 23:30:38 +0000
5575  2016-09-14 23:30:38 +0000
5576  2016-09-14 23:30:38 +0000
5577  2016-09-14 23:30:38 +0000
5578  2016-09-14 23:30:38 +0000
5579  2016-09-14 23:30:38 +0000


                                                                              sourc
5571  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5572  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5573  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5574  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5575  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5576  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5577  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5578  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5579  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


5571  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5572  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5573  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5574  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5575  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5576  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5577  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5578  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10
5579  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10

                                                           expanded_urls  \
5571  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5572  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5573  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5574  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5575  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5576  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5577  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5578  https://twitter.com/dog_rates/status/776201521193218049/photo/1
5579  https://twitter.com/dog_rates/status/776201521193218049/photo/1
```

```
          rating_numerator  rating_denominator      name  ...   img_num  \
5571  10                10                      O'Malley  ...   1
5572  10                10                      O'Malley  ...   1
5573  10                10                      O'Malley  ...   1
5574  10                10                      O'Malley  ...   1
5575  10                10                      O'Malley  ...   1
5576  10                10                      O'Malley  ...   1
5577  10                10                      O'Malley  ...   1
5578  10                10                      O'Malley  ...   1
5579  10                10                      O'Malley  ...   1

               p1    p1_conf p1_dog                        p2   p2_conf  p2_dog  \
5571  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5572  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5573  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5574  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5575  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5576  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5577  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5578  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True
5579  Rottweiler  0.502228  True    black-and-tan_coonhound  0.154594  True

               p3    p3_conf  p3_dog
5571  bloodhound  0.135176  True
5572  bloodhound  0.135176  True
5573  bloodhound  0.135176  True
5574  bloodhound  0.135176  True
5575  bloodhound  0.135176  True
5576  bloodhound  0.135176  True
5577  bloodhound  0.135176  True
5578  bloodhound  0.135176  True
5579  bloodhound  0.135176  True

[9 rows x 24 columns]
```

22 Define

Change missing values in 'name' from 'None' to NaN.

23 Code

```
In [53]: twitter_archive_clean['name'] = twitter_archive_clean['name'].replace('None', np.NaN)
```

24 Test

```
In [54]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6126 entries, 0 to 6742
Data columns (total 24 columns):
tweet_id                6126 non-null int64
in_reply_to_status_id   36 non-null float64
in_reply_to_user_id     36 non-null float64
timestamp               6126 non-null object
source                  6126 non-null object
text                    6126 non-null object
expanded_urls           6126 non-null object
rating_numerator        6126 non-null int64
rating_denominator      6126 non-null int64
name                    4888 non-null object
dog_stage               1029 non-null object
retweet_count           6126 non-null int64
favorite_count          6126 non-null int64
jpg_url                 6126 non-null object
img_num                 6126 non-null int64
p1                      6126 non-null object
p1_conf                 6126 non-null float64
p1_dog                  6126 non-null bool
p2                      6126 non-null object
p2_conf                 6126 non-null float64
p2_dog                  6126 non-null bool
p3                      6126 non-null object
p3_conf                 6126 non-null float64
p3_dog                  6126 non-null bool
dtypes: bool(3), float64(5), int64(6), object(10)
memory usage: 1.0+ MB
```

25 Define

Fix rating numerator and denominators that are not actually ratings.

26 Code

```
In [55]: # View all occurences where there are more than one #/# in 'text' column
         twitter_archive_clean[twitter_archive_clean.text.str.contains( r"(\d+\.?\d*\/\d+\.?\d*\

Out[55]:                     tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         5499  777684233540206592 NaN                         NaN
         5500  777684233540206592 NaN                         NaN
         5501  777684233540206592 NaN                         NaN
         5502  777684233540206592 NaN                         NaN
         5503  777684233540206592 NaN                         NaN
         5504  777684233540206592 NaN                         NaN
```

```
5505  777684233540206592 NaN                    NaN
5506  777684233540206592 NaN                    NaN
5507  777684233540206592 NaN                    NaN
6611  668537837512433665 NaN                    NaN
6657  667544320556335104 NaN                    NaN
6664  667491009379606528 NaN                    NaN
6697  666835007768551424 NaN                    NaN
6724  662287406224695296 NaN                    NaN


                            timestamp  \
5499   2016-09-19 01:42:24 +0000
5500   2016-09-19 01:42:24 +0000
5501   2016-09-19 01:42:24 +0000
5502   2016-09-19 01:42:24 +0000
5503   2016-09-19 01:42:24 +0000
5504   2016-09-19 01:42:24 +0000
5505   2016-09-19 01:42:24 +0000
5506   2016-09-19 01:42:24 +0000
5507   2016-09-19 01:42:24 +0000
6611   2015-11-22 21:13:35 +0000
6657   2015-11-20 03:25:43 +0000
6664   2015-11-19 23:53:52 +0000
6697   2015-11-18 04:27:09 +0000
6724   2015-11-16 16:11:11 +0000


                                                                               sourc
5499   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5500   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5501   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5502   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5503   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5504   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5505   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5506   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5507   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
6611   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
6657   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
6664   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
6697   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
6724   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


5499   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5500   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5501   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5502   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5503   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5504   "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
```

50

```
5505  "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5506  "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
5507  "Yep... just as I suspected. You're not flossing." 12/10 and 11/10 for the pup no
6611  This is Spark. He's nervous. Other dog hasn't moved in a while. Won't come when c
6657  This is Kial. Kial is either wearing a cape, which would be rad, or flashing us,
6664  Two dogs in this one. Both are rare Jujitsu Pythagoreans. One slightly whiter tha
6697  These are Peruvian Feldspars. Their names are Cupit and Prencer. Both resemble Ra
6724  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo

                                                         expanded_urls  \
5499  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5500  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5501  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5502  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5503  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5504  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5505  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5506  https://twitter.com/dog_rates/status/777684233540206592/photo/1
5507  https://twitter.com/dog_rates/status/777684233540206592/photo/1
6611  https://twitter.com/dog_rates/status/668537837512433665/photo/1
6657  https://twitter.com/dog_rates/status/667544320556335104/photo/1
6664  https://twitter.com/dog_rates/status/667491009379606528/photo/1
6697  https://twitter.com/dog_rates/status/666835007768551424/photo/1
6724  https://twitter.com/dog_rates/status/666287406224695296/photo/1


      rating_numerator  rating_denominator   name    ...    img_num  \
5499  12                10                   NaN    ...    1
5500  12                10                   NaN    ...    1
5501  12                10                   NaN    ...    1
5502  12                10                   NaN    ...    1
5503  12                10                   NaN    ...    1
5504  12                10                   NaN    ...    1
5505  12                10                   NaN    ...    1
5506  12                10                   NaN    ...    1
5507  12                10                   NaN    ...    1
6611  8                 10                   Spark  ...    1
6657  10                10                   Kial   ...    1
6664  7                 10                   NaN    ...    1
6697  10                10                   NaN    ...    1
6724  1                 2                    NaN    ...    1


                   p1     p1_conf p1_dog               p2    p2_conf  p2_dog  \
5499  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
5500  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
5501  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
5502  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
5503  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
5504  cocker_spaniel   0.253442   True    golden_retriever  0.162850  True
```

```
      5505   cocker_spaniel    0.253442  True   golden_retriever  0.162850  True
      5506   cocker_spaniel    0.253442  True   golden_retriever  0.162850  True
      5507   cocker_spaniel    0.253442  True   golden_retriever  0.162850  True
      6611   Lakeland_terrier  0.372988  True   toy_poodle        0.250445  True
      6657   Pomeranian        0.412893  True   Pembroke          0.312958  True
      6664   borzoi            0.852088  True   ice_bear          0.132264  False
      6697   Airedale          0.448459  True   toy_poodle        0.124030  True
      6724   Maltese_dog       0.857531  True   toy_poodle        0.063064  True


                          p3   p3_conf  p3_dog
      5499   otterhound         0.110921  True
      5500   otterhound         0.110921  True
      5501   otterhound         0.110921  True
      5502   otterhound         0.110921  True
      5503   otterhound         0.110921  True
      5504   otterhound         0.110921  True
      5505   otterhound         0.110921  True
      5506   otterhound         0.110921  True
      5507   otterhound         0.110921  True
      6611   Chihuahua          0.189737  True
      6657   Chihuahua          0.071960  True
      6664   weasel             0.005730  False
      6697   teddy              0.110183  False
      6724   miniature_poodle   0.025581  True


      [14 rows x 24 columns]
```

In [56]: `# Save the text where the rating numerator and denominators were incorrectly extracted`
```
ratings_to_fix = ['After so many requests, this is Bretagne. She was the last surviving
 'Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a',
 'This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amazed a
 'This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was just s
 'This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood flooring
```

In [57]: `# Loop through the list of ratings to fix and extract the second occurence of #/ to sau`
```
# occurences of the actual ratings in the ratings to fix list have a denominator of 10,
#entry instead of extracting it.
for entry in ratings_to_fix:
    mask = twitter_archive_clean.text == entry
    column_name1 = 'rating_numerator'
    column_name2 = 'rating_denominator'
    twitter_archive_clean.loc[mask, column_name1] = re.findall(r"\d+\.?\d*\/\d+\.?\d*\D
    twitter_archive_clean.loc[mask, column_name2] = 10
```

27 Test

In [58]: `twitter_archive_clean[twitter_archive_clean.text.isin(ratings_to_fix)]`

```
Out[58]:                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         6724  666287406224695296 NaN                    NaN


                                  timestamp  \
         6724  2015-11-16 16:11:11 +0000


                                                                            sourc
         6724  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


         6724  This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floo


                                            expanded_urls  \
         6724  https://twitter.com/dog_rates/status/666287406224695296/photo/1


               rating_numerator  rating_denominator name    ...    img_num           p1  \
         6724  9                 10                  NaN     ...    1         Maltese_dog


                 p1_conf p1_dog          p2    p2_conf  p2_dog                p3  \
         6724    0.857531  True   toy_poodle  0.063064  True   miniature_poodle


                 p3_conf  p3_dog
         6724    0.025581  True


         [1 rows x 24 columns]
```

28   Define

Fix rating numerator that have decimals.

```
In [59]: # View tweets with decimals in rating in 'text' column
         twitter_archive_clean[twitter_archive_clean.text.str.contains(r"(\d+\.\d*\/\d+)")]

Out[59]:               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         360   883482846933004288 NaN                    NaN
         361   883482846933004288 NaN                    NaN
         362   883482846933004288 NaN                    NaN
         363   883482846933004288 NaN                    NaN
         364   883482846933004288 NaN                    NaN
         365   883482846933004288 NaN                    NaN
         366   883482846933004288 NaN                    NaN
         367   883482846933004288 NaN                    NaN
         368   883482846933004288 NaN                    NaN
         4977  786709082849828864 NaN                    NaN
         4978  786709082849828864 NaN                    NaN
         4979  786709082849828864 NaN                    NaN
         4980  786709082849828864 NaN                    NaN
         4981  786709082849828864 NaN                    NaN
```

```
4982  786709082849828864 NaN                    NaN
4983  786709082849828864 NaN                    NaN
4984  786709082849828864 NaN                    NaN
4985  786709082849828864 NaN                    NaN
5481  778027034220126208 NaN                    NaN
5482  778027034220126208 NaN                    NaN
5483  778027034220126208 NaN                    NaN
5484  778027034220126208 NaN                    NaN
5485  778027034220126208 NaN                    NaN
5486  778027034220126208 NaN                    NaN
5487  778027034220126208 NaN                    NaN
5488  778027034220126208 NaN                    NaN
5489  778027034220126208 NaN                    NaN


                            timestamp  \
360    2017-07-08 00:28:19 +0000
361    2017-07-08 00:28:19 +0000
362    2017-07-08 00:28:19 +0000
363    2017-07-08 00:28:19 +0000
364    2017-07-08 00:28:19 +0000
365    2017-07-08 00:28:19 +0000
366    2017-07-08 00:28:19 +0000
367    2017-07-08 00:28:19 +0000
368    2017-07-08 00:28:19 +0000
4977   2016-10-13 23:23:56 +0000
4978   2016-10-13 23:23:56 +0000
4979   2016-10-13 23:23:56 +0000
4980   2016-10-13 23:23:56 +0000
4981   2016-10-13 23:23:56 +0000
4982   2016-10-13 23:23:56 +0000
4983   2016-10-13 23:23:56 +0000
4984   2016-10-13 23:23:56 +0000
4985   2016-10-13 23:23:56 +0000
5481   2016-09-20 00:24:34 +0000
5482   2016-09-20 00:24:34 +0000
5483   2016-09-20 00:24:34 +0000
5484   2016-09-20 00:24:34 +0000
5485   2016-09-20 00:24:34 +0000
5486   2016-09-20 00:24:34 +0000
5487   2016-09-20 00:24:34 +0000
5488   2016-09-20 00:24:34 +0000
5489   2016-09-20 00:24:34 +0000


                                                                        sourc
360    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
361    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
362    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
363    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
```

```
364    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
365    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
366    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
367    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
368    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4977   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4978   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4979   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4980   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4981   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4982   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4983   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4984   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4985   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5481   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5482   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5483   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5484   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5485   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5486   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5487   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5488   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5489   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


360    This is Bella. She hopes her smile made you smile. If not, she is also offering y
361    This is Bella. She hopes her smile made you smile. If not, she is also offering y
362    This is Bella. She hopes her smile made you smile. If not, she is also offering y
363    This is Bella. She hopes her smile made you smile. If not, she is also offering y
364    This is Bella. She hopes her smile made you smile. If not, she is also offering y
365    This is Bella. She hopes her smile made you smile. If not, she is also offering y
366    This is Bella. She hopes her smile made you smile. If not, she is also offering y
367    This is Bella. She hopes her smile made you smile. If not, she is also offering y
368    This is Bella. She hopes her smile made you smile. If not, she is also offering y
4977   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4978   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4979   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4980   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4981   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4982   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4983   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4984   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4985   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
5481   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5482   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5483   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5484   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5485   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
```

```
5486   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5487   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5488   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5489   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando


360    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
361    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
362    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
363    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
364    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
365    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
366    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
367    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
368    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
4977   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4978   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4979   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4980   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4981   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4982   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4983   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4984   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4985   https://twitter.com/dog_rates/status/786709082849828864/photo/1
5481   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5482   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5483   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5484   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5485   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5486   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5487   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5488   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5489   https://twitter.com/dog_rates/status/778027034220126208/photo/1
```

	rating_numerator	rating_denominator	name	...	img_num \
360	5	10	Bella	...	1
361	5	10	Bella	...	1
362	5	10	Bella	...	1
363	5	10	Bella	...	1
364	5	10	Bella	...	1
365	5	10	Bella	...	1
366	5	10	Bella	...	1
367	5	10	Bella	...	1
368	5	10	Bella	...	1
4977	75	10	Logan	...	1
4978	75	10	Logan	...	1
4979	75	10	Logan	...	1
4980	75	10	Logan	...	1

```
4981  75              10                     Logan    ...    1
4982  75              10                     Logan    ...    1
4983  75              10                     Logan    ...    1
4984  75              10                     Logan    ...    1
4985  75              10                     Logan    ...    1
5481  27              10                     Sophie   ...    1
5482  27              10                     Sophie   ...    1
5483  27              10                     Sophie   ...    1
5484  27              10                     Sophie   ...    1
5485  27              10                     Sophie   ...    1
5486  27              10                     Sophie   ...    1
5487  27              10                     Sophie   ...    1
5488  27              10                     Sophie   ...    1
5489  27              10                     Sophie   ...    1


                    p1   p1_conf p1_dog                   p2   p2_conf  p2_dog  \
360   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
361   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
362   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
363   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
364   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
365   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
366   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
367   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
368   golden_retriever  0.943082   True    Labrador_retriever  0.032409   True
4977  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4978  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4979  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4980  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4981  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4982  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4983  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4984  Pomeranian        0.467321   True    Persian_cat         0.122978  False
4985  Pomeranian        0.467321   True    Persian_cat         0.122978  False
5481  clumber           0.946718   True    cocker_spaniel      0.015950   True
5482  clumber           0.946718   True    cocker_spaniel      0.015950   True
5483  clumber           0.946718   True    cocker_spaniel      0.015950   True
5484  clumber           0.946718   True    cocker_spaniel      0.015950   True
5485  clumber           0.946718   True    cocker_spaniel      0.015950   True
5486  clumber           0.946718   True    cocker_spaniel      0.015950   True
5487  clumber           0.946718   True    cocker_spaniel      0.015950   True
5488  clumber           0.946718   True    cocker_spaniel      0.015950   True
5489  clumber           0.946718   True    cocker_spaniel      0.015950   True


          p3    p3_conf  p3_dog
360   kuvasz  0.005501   True
361   kuvasz  0.005501   True
362   kuvasz  0.005501   True
```

```
363    kuvasz    0.005501    True
364    kuvasz    0.005501    True
365    kuvasz    0.005501    True
366    kuvasz    0.005501    True
367    kuvasz    0.005501    True
368    kuvasz    0.005501    True
4977   chow      0.102654    True
4978   chow      0.102654    True
4979   chow      0.102654    True
4980   chow      0.102654    True
4981   chow      0.102654    True
4982   chow      0.102654    True
4983   chow      0.102654    True
4984   chow      0.102654    True
4985   chow      0.102654    True
5481   Lhasa     0.006519    True
5482   Lhasa     0.006519    True
5483   Lhasa     0.006519    True
5484   Lhasa     0.006519    True
5485   Lhasa     0.006519    True
5486   Lhasa     0.006519    True
5487   Lhasa     0.006519    True
5488   Lhasa     0.006519    True
5489   Lhasa     0.006519    True

[27 rows x 24 columns]
```

In [60]: # Change datatype of rating_numerator and denominator to float
         twitter_archive_clean['rating_numerator'] = twitter_archive_clean['rating_numerator'].a
         twitter_archive_clean['rating_denominator'] = twitter_archive_clean['rating_denominator

In [61]: # Set correct numerators for specific tweets
         twitter_archive_clean.loc[(twitter_archive_clean['tweet_id'] == 883482846933004288) & (
         twitter_archive_clean.loc[(twitter_archive_clean['tweet_id'] == 786709082849828864) & (
         twitter_archive_clean.loc[(twitter_archive_clean['tweet_id'] == 778027034220126208) & (
         twitter_archive_clean.loc[(twitter_archive_clean['tweet_id'] == 680494726643068929) & (

29 Test

In [62]: twitter_archive_clean[twitter_archive_clean.text.str.contains(r"(\d+\.\d*\/\d+)")]

Out[62]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         360   883482846933004288 NaN   NaN
         361   883482846933004288 NaN   NaN
         362   883482846933004288 NaN   NaN
         363   883482846933004288 NaN   NaN
         364   883482846933004288 NaN   NaN
         365   883482846933004288 NaN   NaN

```

```
366   883482846933004288 NaN   NaN
367   883482846933004288 NaN   NaN
368   883482846933004288 NaN   NaN
4977   786709082849828864 NaN   NaN
4978   786709082849828864 NaN   NaN
4979   786709082849828864 NaN   NaN
4980   786709082849828864 NaN   NaN
4981   786709082849828864 NaN   NaN
4982   786709082849828864 NaN   NaN
4983   786709082849828864 NaN   NaN
4984   786709082849828864 NaN   NaN
4985   786709082849828864 NaN   NaN
5481   778027034220126208 NaN   NaN
5482   778027034220126208 NaN   NaN
5483   778027034220126208 NaN   NaN
5484   778027034220126208 NaN   NaN
5485   778027034220126208 NaN   NaN
5486   778027034220126208 NaN   NaN
5487   778027034220126208 NaN   NaN
5488   778027034220126208 NaN   NaN
5489   778027034220126208 NaN   NaN

                         timestamp \
360   2017-07-08 00:28:19 +0000
361   2017-07-08 00:28:19 +0000
362   2017-07-08 00:28:19 +0000
363   2017-07-08 00:28:19 +0000
364   2017-07-08 00:28:19 +0000
365   2017-07-08 00:28:19 +0000
366   2017-07-08 00:28:19 +0000
367   2017-07-08 00:28:19 +0000
368   2017-07-08 00:28:19 +0000
4977   2016-10-13 23:23:56 +0000
4978   2016-10-13 23:23:56 +0000
4979   2016-10-13 23:23:56 +0000
4980   2016-10-13 23:23:56 +0000
4981   2016-10-13 23:23:56 +0000
4982   2016-10-13 23:23:56 +0000
4983   2016-10-13 23:23:56 +0000
4984   2016-10-13 23:23:56 +0000
4985   2016-10-13 23:23:56 +0000
5481   2016-09-20 00:24:34 +0000
5482   2016-09-20 00:24:34 +0000
5483   2016-09-20 00:24:34 +0000
5484   2016-09-20 00:24:34 +0000
5485   2016-09-20 00:24:34 +0000
5486   2016-09-20 00:24:34 +0000
5487   2016-09-20 00:24:34 +0000
```

```
5488   2016-09-20 00:24:34 +0000
5489   2016-09-20 00:24:34 +0000


                                                                                sourc
360    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
361    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
362    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
363    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
364    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
365    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
366    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
367    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
368    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4977   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4978   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4979   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4980   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4981   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4982   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4983   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4984   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
4985   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5481   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5482   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5483   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5484   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5485   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5486   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5487   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5488   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a
5489   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a


360    This is Bella. She hopes her smile made you smile. If not, she is also offering y
361    This is Bella. She hopes her smile made you smile. If not, she is also offering y
362    This is Bella. She hopes her smile made you smile. If not, she is also offering y
363    This is Bella. She hopes her smile made you smile. If not, she is also offering y
364    This is Bella. She hopes her smile made you smile. If not, she is also offering y
365    This is Bella. She hopes her smile made you smile. If not, she is also offering y
366    This is Bella. She hopes her smile made you smile. If not, she is also offering y
367    This is Bella. She hopes her smile made you smile. If not, she is also offering y
368    This is Bella. She hopes her smile made you smile. If not, she is also offering y
4977   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4978   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4979   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4980   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4981   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4982   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
```

```
4983   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4984   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
4985   This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*
5481   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5482   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5483   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5484   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5485   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5486   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5487   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5488   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando
5489   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rando


360    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
361    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
362    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
363    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
364    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
365    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
366    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
367    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
368    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.c
4977   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4978   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4979   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4980   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4981   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4982   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4983   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4984   https://twitter.com/dog_rates/status/786709082849828864/photo/1
4985   https://twitter.com/dog_rates/status/786709082849828864/photo/1
5481   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5482   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5483   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5484   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5485   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5486   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5487   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5488   https://twitter.com/dog_rates/status/778027034220126208/photo/1
5489   https://twitter.com/dog_rates/status/778027034220126208/photo/1


      rating_numerator  rating_denominator   name    ...    img_num  \
360   13.50             10.0                 Bella   ...    1
361   13.50             10.0                 Bella   ...    1
362   13.50             10.0                 Bella   ...    1
363   13.50             10.0                 Bella   ...    1
364   13.50             10.0                 Bella   ...    1
```

```
365   13.50            10.0            Bella    ...    1
366   13.50            10.0            Bella    ...    1
367   13.50            10.0            Bella    ...    1
368   13.50            10.0            Bella    ...    1
4977  9.75             10.0            Logan    ...    1
4978  9.75             10.0            Logan    ...    1
4979  9.75             10.0            Logan    ...    1
4980  9.75             10.0            Logan    ...    1
4981  9.75             10.0            Logan    ...    1
4982  9.75             10.0            Logan    ...    1
4983  9.75             10.0            Logan    ...    1
4984  9.75             10.0            Logan    ...    1
4985  9.75             10.0            Logan    ...    1
5481  11.27            10.0            Sophie   ...    1
5482  11.27            10.0            Sophie   ...    1
5483  11.27            10.0            Sophie   ...    1
5484  11.27            10.0            Sophie   ...    1
5485  11.27            10.0            Sophie   ...    1
5486  11.27            10.0            Sophie   ...    1
5487  11.27            10.0            Sophie   ...    1
5488  11.27            10.0            Sophie   ...    1
5489  11.27            10.0            Sophie   ...    1

                    p1     p1_conf p1_dog                 p2    p2_conf p2_dog  \
360    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
361    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
362    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
363    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
364    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
365    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
366    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
367    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
368    golden_retriever  0.943082  True  Labrador_retriever  0.032409  True
4977   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4978   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4979   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4980   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4981   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4982   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4983   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4984   Pomeranian        0.467321  True  Persian_cat         0.122978  False
4985   Pomeranian        0.467321  True  Persian_cat         0.122978  False
5481   clumber           0.946718  True  cocker_spaniel      0.015950  True
5482   clumber           0.946718  True  cocker_spaniel      0.015950  True
5483   clumber           0.946718  True  cocker_spaniel      0.015950  True
5484   clumber           0.946718  True  cocker_spaniel      0.015950  True
5485   clumber           0.946718  True  cocker_spaniel      0.015950  True
5486   clumber           0.946718  True  cocker_spaniel      0.015950  True
```

```
5487   clumber              0.946718  True   cocker_spaniel      0.015950  True
5488   clumber              0.946718  True   cocker_spaniel      0.015950  True
5489   clumber              0.946718  True   cocker_spaniel      0.015950  True


             p3   p3_conf  p3_dog
360    kuvasz  0.005501  True
361    kuvasz  0.005501  True
362    kuvasz  0.005501  True
363    kuvasz  0.005501  True
364    kuvasz  0.005501  True
365    kuvasz  0.005501  True
366    kuvasz  0.005501  True
367    kuvasz  0.005501  True
368    kuvasz  0.005501  True
4977   chow    0.102654  True
4978   chow    0.102654  True
4979   chow    0.102654  True
4980   chow    0.102654  True
4981   chow    0.102654  True
4982   chow    0.102654  True
4983   chow    0.102654  True
4984   chow    0.102654  True
4985   chow    0.102654  True
5481   Lhasa   0.006519  True
5482   Lhasa   0.006519  True
5483   Lhasa   0.006519  True
5484   Lhasa   0.006519  True
5485   Lhasa   0.006519  True
5486   Lhasa   0.006519  True
5487   Lhasa   0.006519  True
5488   Lhasa   0.006519  True
5489   Lhasa   0.006519  True

[27 rows x 24 columns]
```

Define
Remove tweet without rating.

30   Code

```
In [63]: twitter_archive_clean = twitter_archive_clean[twitter_archive_clean.tweet_id != 8109846
```

31   Test

```
In [64]: twitter_archive_clean[twitter_archive_clean.tweet_id == 810984652412424192]

Out[64]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text
```

```
Index: []

[0 rows x 24 columns]
```

Define
Remove extra characters after '&' in twitter_archive_clean['text'].

32 Code

```
In [65]: twitter_archive_clean['text'] = twitter_archive_clean['text'].str.replace('&amp;', '&')
```

33 Test

```
In [66]: twitter_archive_clean[twitter_archive_clean.text.str.contains('&amp;')]

Out[66]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, text
         Index: []

         [0 rows x 24 columns]
```

34 Define

Change sources to more readable categories.

35 Code

```
In [67]: # Remove url from sources
         twitter_archive_clean['source'] = twitter_archive_clean['source'].str.replace('<a href=
         twitter_archive_clean['source'] = twitter_archive_clean['source'].str.replace('<a href=
         twitter_archive_clean['source'] = twitter_archive_clean['source'].str.replace('<a href=
         twitter_archive_clean['source'] = twitter_archive_clean['source'].str.replace('<a href=

In [68]: # Change datatype to category
         twitter_archive_clean['source'] = twitter_archive_clean['source'].astype('category')
```

36 Test

```
In [69]: twitter_archive_clean.source.value_counts()

Out[69]: Twitter for iPhone    6047
         Twitter Web Client    70
         Name: source, dtype: int64

In [70]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6117 entries, 0 to 6742
Data columns (total 24 columns):
tweet_id                6117 non-null int64
in_reply_to_status_id   36 non-null float64
in_reply_to_user_id     36 non-null float64
timestamp               6117 non-null object
source                  6117 non-null category
text                    6117 non-null object
expanded_urls           6117 non-null object
rating_numerator        6117 non-null float64
rating_denominator      6117 non-null float64
name                    4879 non-null object
dog_stage               1029 non-null object
retweet_count           6117 non-null int64
favorite_count          6117 non-null int64
jpg_url                 6117 non-null object
img_num                 6117 non-null int64
p1                      6117 non-null object
p1_conf                 6117 non-null float64
p1_dog                  6117 non-null bool
p2                      6117 non-null object
p2_conf                 6117 non-null float64
p2_dog                  6117 non-null bool
p3                      6117 non-null object
p3_conf                 6117 non-null float64
p3_dog                  6117 non-null bool
dtypes: bool(3), category(1), float64(7), int64(4), object(9)
memory usage: 1.0+ MB
```

37 Define

Change datatypes of timestamp to datetime, dog_stage to categorical, and tweet_id, in_reply_to_status_id, and in_reply_to_user_id to strings.

38 Code

```
In [71]: twitter_archive_clean['dog_stage'] = twitter_archive_clean['dog_stage'].astype('categor
         twitter_archive_clean['timestamp'] = pd.to_datetime(twitter_archive_clean['timestamp'])
         twitter_archive_clean['tweet_id'] = twitter_archive_clean['tweet_id'].astype('str')
         twitter_archive_clean['in_reply_to_status_id'] = twitter_archive_clean['in_reply_to_sta
         twitter_archive_clean['in_reply_to_user_id'] = twitter_archive_clean['in_reply_to_user_
```

39   Test

```
In [72]: twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 6117 entries, 0 to 6742
Data columns (total 24 columns):
tweet_id                 6117 non-null object
in_reply_to_status_id    6117 non-null object
in_reply_to_user_id      6117 non-null object
timestamp                6117 non-null datetime64[ns]
source                   6117 non-null category
text                     6117 non-null object
expanded_urls            6117 non-null object
rating_numerator         6117 non-null float64
rating_denominator       6117 non-null float64
name                     4879 non-null object
dog_stage                1029 non-null category
retweet_count            6117 non-null int64
favorite_count           6117 non-null int64
jpg_url                  6117 non-null object
img_num                  6117 non-null int64
p1                       6117 non-null object
p1_conf                  6117 non-null float64
p1_dog                   6117 non-null bool
p2                       6117 non-null object
p2_conf                  6117 non-null float64
p2_dog                   6117 non-null bool
p3                       6117 non-null object
p3_conf                  6117 non-null float64
p3_dog                   6117 non-null bool
dtypes: bool(3), category(2), datetime64[ns](1), float64(5), int64(3), object(10)
memory usage: 985.9+ KB
```

40   Store Data

```
In [73]: # Save clean DataFrame to csv file
         twitter_archive_clean.to_csv('twitter_archive_master.csv')
```

41   Explotary Data Analysis

```
In [74]: # Create copy of twitter_archive_clean to work off of using only my variables of intere
         time_df = twitter_archive_clean[['timestamp', 'retweet_count', 'favorite_count', 'ratin

In [75]: # Set the index to be the timestamp so time is displayed properly in plots
         time_df.set_index('timestamp', inplace=True)
```

```
In [76]:  # Create rating_ration variable by dividing the rating numerator by the deonominator to
          # out of 10
          time_df['rating_ratio'] = time_df['rating_numerator']/time_df['rating_denominator']

In [77]:  time_df['retweet_count'].plot(color = 'red', label='Retweets')
          time_df['favorite_count'].plot(color = 'green', label='Favorites')
          plt.style.use('seaborn-darkgrid')
          plt.legend(loc='upper left')
          plt.xlabel('Tweet timestamp')
          plt.ylabel('Count')
          plt.title('Retweets and favorites over time')
          plt.savefig('retweets_favorites.png')
          plt.show()
```
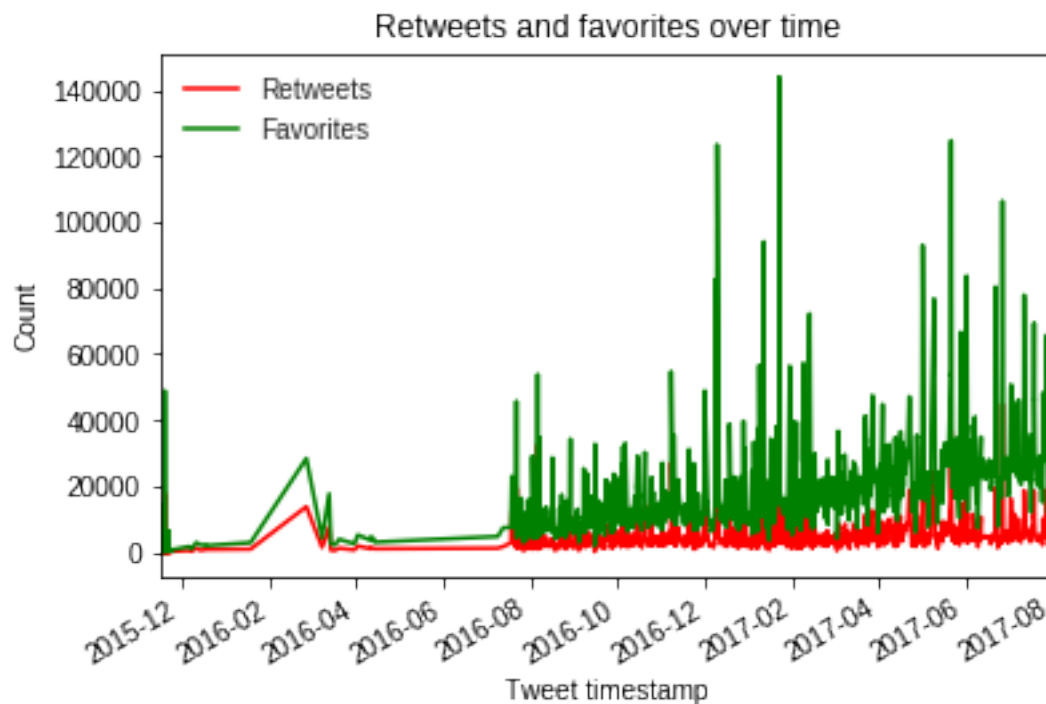


```
In [78]:  time_df['rating_ratio'].plot()
          plt.style.use('seaborn-darkgrid')
          plt.xlabel('Time')
          plt.ylabel('Ratio')
          plt.title('Rating ratio over time')
          plt.savefig('ratio.png')
          plt.show()
```

Rating ratio over time
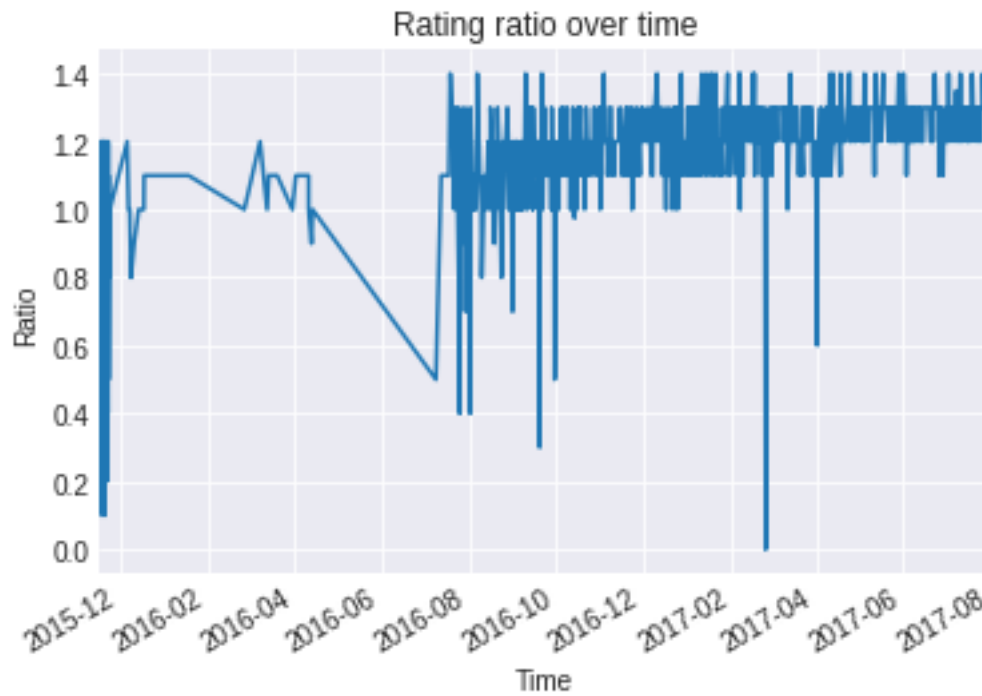
In [79]: # Limit y axis to zoom in on data and ignore outliers
```python
time_df['rating_ratio'].plot()
plt.ylim(-1, 3)
plt.style.use('seaborn-darkgrid')
plt.xlabel('Time')
plt.ylabel('Ratio')
plt.title('Rating ratio over time')
plt.savefig('ratio_zoom.png')
plt.show()
```

Rating ratio over time
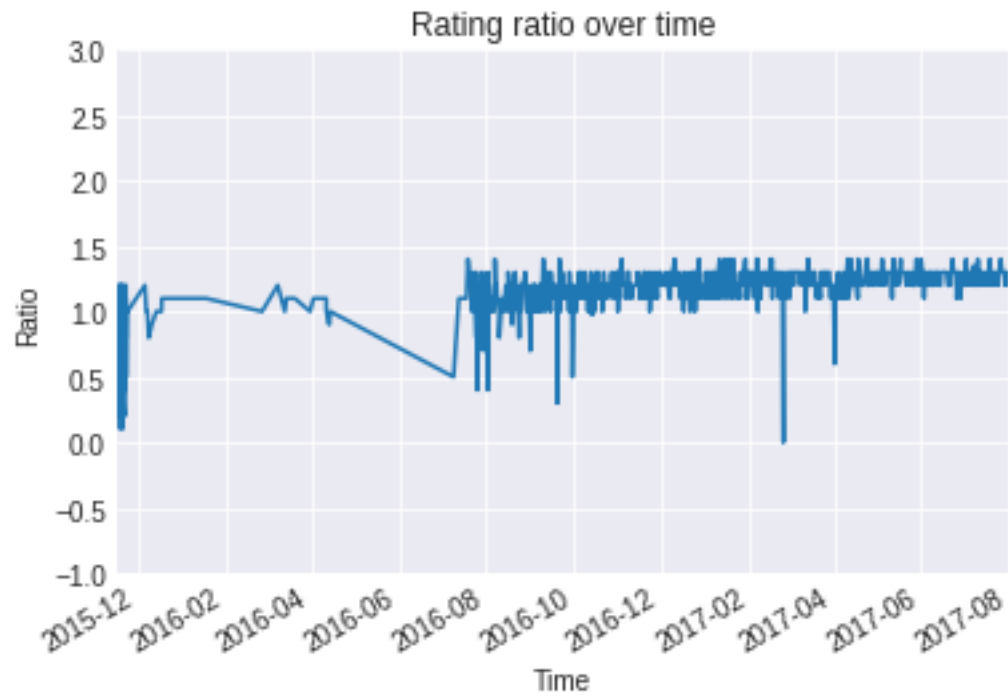
```
In [80]: twitter_archive_clean.plot(kind = 'scatter', x = 'favorite_count', y = 'retweet_count',
         plt.xlabel('favorite_count')
         plt.ylabel('retweet_count')
         plt.title('Retweets and favorites Scatter plot')

Out[80]: Text(0.5,1,'Retweets and favorites Scatter plot')
```

Retweets and favorites Scatter plot
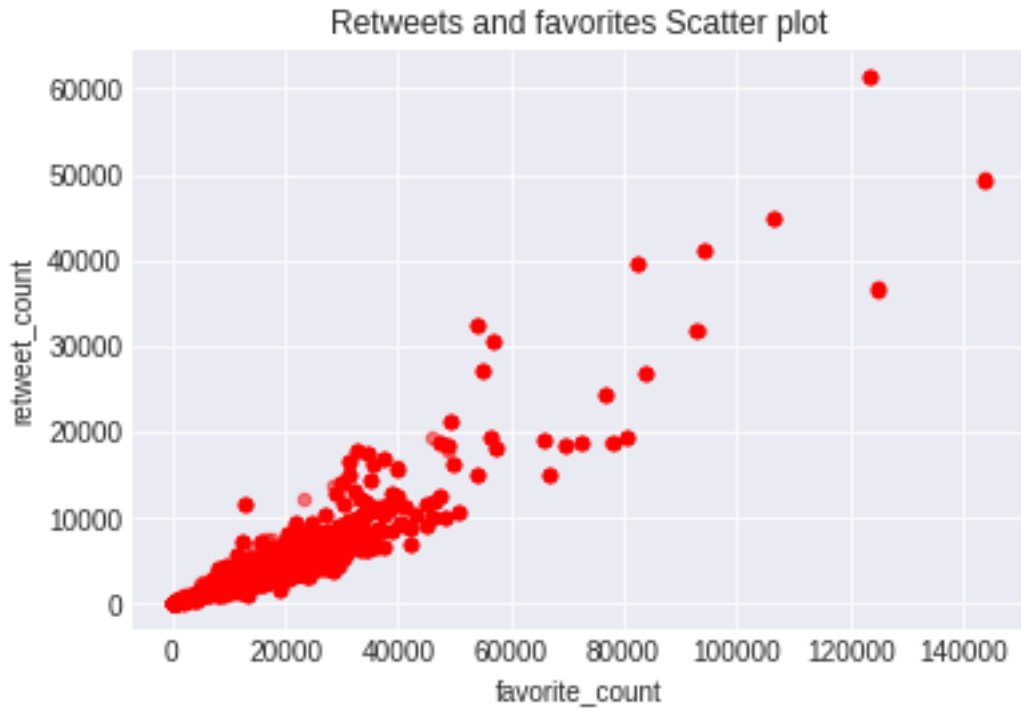
In [81]: `# Our range will be [0,16] taking of the two ouliers (1776 and 420)`
`twitter_archive_clean.plot(x= 'timestamp', y ='rating_numerator', ylim=[0,16], style =`
`plt.title('Rating plot over Time')`
`plt.xlabel('Date')`
`plt.ylabel('Rating')`

Out[81]: `Text(0,0.5,'Rating')`

Rating plot over Time