wrangle_act

July 1, 2018

```
In [180]: import pandas as pd
          import numpy as np
          import requests
          import tweepy
          import json
          import re
          import warnings
          import matplotlib.pyplot as plt
          from IPython.display import Image
          from IPython.core.display import HTML
```

0.1 Gathering

```
In [2]: # Read csv file as Pandas DataFrame
        twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
```

```
In [14]: # Download image-predictions.tsv file from a url using requests library
         url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predic

         filename = list(url.split('/'))[-1]
         response = requests.get(url)

         with open(filename, 'wb') as file:
             file.write(response.content)
```

```
In [3]: # Read tsv file as a Pandas DataFrame
        filename = 'image-predictions.tsv'
        image_predictions = pd.read_csv(filename, sep='\t')
```

```
In [7]: # Personal API keys, secrets, and tokens have been replaced with placeholders
        consumer_key = 'MY CONSUMER KEY'
        consumer_secret = 'MY CONSUMER SECRET'
        access_token = 'MY ACCESS TOKEN'
        access_secret = 'MY ACCESS SECRET'
```

```
In [8]: # Variables created for tweepy query
        import tweepy
        auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_secret)
        api = tweepy.API(auth, wait_on_rate_limit = True, wait_on_rate_limit_notify = True)
```

```
In [4]: # For loop which will add each available tweet to a new line of tweet_json.txt
        counter = 0
        with open('tweet_json.txt', 'a', encoding='utf-8') 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')

                    counter += 1
                except Exception as e:
                    continue
        print(counter)
```

```
0
```

```
In [5]: # 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 [6]: # Create tweet_info DataFrame
        tweet_info = pd.DataFrame()
```

```
In [7]: # 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_da
        tweet_info['favorite_count'] = list(map(lambda tweet: tweet['favorite_count'], tweets_
```

0.2 Assess

```
In [30]: # Set column width to 1000 to display full content of 'text' column
         pd.set_option('display.max_colwidth', 1000)
```

```
In [31]: # View first 20 rows of twitter_archive DataFrame
         twitter_archive.head()
```

```
Out[31]:             tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0  892420643555336193                    NaN                  NaN
         1  892177421306343426                    NaN                  NaN
         2  891815181378084864                    NaN                  NaN
         3  891689557279858688                    NaN                  NaN
         4  891327558926688256                    NaN                  NaN

                            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

                                                                                     source
         0  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         1  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         2  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         3  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         4  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>


         0                                                        This is Phineas. He's a mystic
         1  This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she's
         2                   This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in th
         3                                                              This is Darla. She comme
         4  This is Franklin. He would like you to stop calling him "cute." He is a very fierce

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


         0                                                                  https://twitter.com
         1                                                                  https://twitter.com
         2                                                                  https://twitter.com
         3                                                                  https://twitter.com
         4  https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.com

            rating_numerator  rating_denominator      name doggo floofer pupper puppo
         0                13                  10   Phineas  None    None   None  None
         1                13                  10     Tilly  None    None   None  None
         2                12                  10    Archie  None    None   None  None
         3                13                  10     Darla  None    None   None  None
         4                12                  10  Franklin  None    None   None  None
```

```
In [32]: print((twitter_archive['rating_numerator']/twitter_archive['rating_denominator'])*100)
```

```
0       130.0
1       130.0
2       120.0
3       130.0
4       120.0
5       130.0
6       130.0
7       130.0
8       130.0
9       140.0
10      130.0
11      130.0
12      130.0
13      120.0
14      130.0
15      130.0
16      120.0
17      130.0
18      130.0
19      130.0
20      120.0
21      130.0
22      140.0
23      130.0
24      130.0
25      120.0
26      130.0
27      130.0
28      130.0
29      120.0
        ...
2326     20.0
2327     70.0
2328     90.0
2329    110.0
2330     60.0
2331     80.0
2332    100.0
2333     90.0
2334     30.0
2335     50.0
2336    110.0
2337    100.0
2338     10.0
2339    110.0
2340     80.0
```

```
2341     90.0
2342     60.0
2343    100.0
2344     90.0
2345    100.0
2346     80.0
2347     90.0
2348    100.0
2349     20.0
2350    100.0
2351     50.0
2352     60.0
2353     90.0
2354     70.0
2355     80.0
Length: 2356, dtype: float64
```

```
In [33]: twitter_archive.name.value_counts()
```

```
Out[33]: None            745
         a                55
         Charlie          12
         Lucy             11
         Oliver           11
         Cooper           11
         Lola             10
         Tucker           10
         Penny            10
         Winston           9
         Bo                9
         Sadie             8
         the               8
         Daisy             7
         Toby              7
         Buddy             7
         an                7
         Bailey            7
         Oscar             6
         Scout             6
         Leo               6
         Rusty             6
         Jax               6
         Bella             6
         Jack              6
         Milo              6
         Dave              6
         Koda              6
```

```
Stanley         6
Louis           5
             ...
Ricky           1
my              1
Hubertson       1
officially      1
Jebberson       1
Dylan           1
Kellogg         1
Gin             1
Hazel           1
Geoff           1
Karma           1
Lassie          1
Chubbs          1
Carper          1
Tanner          1
Acro            1
Malikai         1
Zara            1
Chaz            1
Biden           1
Finnegus        1
Petrick         1
Huxley          1
Lugan           1
Kramer          1
Millie          1
Willow          1
Tripp           1
Diogi           1
Ralphson        1
Name: name, Length: 957, dtype: int64
```

In [34]:

```
# View last 5 rows of twitter_archive DataFrame
twitter_archive.tail()
```

Out[34]:

```
                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

                                                                                  sou
2351  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2352  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2353  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2354  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2355  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<


2351                     Here we have a 1949 1st generation vulpix. Enjoys sweat tea a
2352    This is a purebred Piers Morgan. Loves to Netflix and chill. Always looks like
2353           Here is a very happy pup. Big fan of well-maintained decks. Just look a
2354  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 dogs he
2355          Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan o

      retweeted_status_id  retweeted_status_user_id  \
2351                  NaN                       NaN
2352                  NaN                       NaN
2353                  NaN                       NaN
2354                  NaN                       NaN
2355                  NaN                       NaN

     retweeted_status_timestamp  \
2351                        NaN
2352                        NaN
2353                        NaN
2354                        NaN
2355                        NaN

                                                        expanded_urls  \
2351  https://twitter.com/dog_rates/status/666049248165822465/photo/1
2352  https://twitter.com/dog_rates/status/666044226329800704/photo/1
2353  https://twitter.com/dog_rates/status/666033412701032449/photo/1
2354  https://twitter.com/dog_rates/status/666029285002620928/photo/1
2355  https://twitter.com/dog_rates/status/666020888022790149/photo/1

      rating_numerator  rating_denominator  name doggo floofer pupper puppo
2351                 5                  10  None  None    None   None  None
2352                 6                  10     a  None    None   None  None
2353                 9                  10     a  None    None   None  None
2354                 7                  10     a  None    None   None  None
2355                 8                  10  None  None    None   None  None
```

```
In [35]: # View info of twitter_archive DataFrame
         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 [36]: # View descriptive statistics of twitter_archive DataFrame
         twitter_archive.describe()
```

```
Out[36]:            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 [37]:
```
# View first 5 rows of image_predictions DataFrame
image_predictions.head()
```

Out[37]:
```
             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 [38]:
```
# View last 5 rows of image_predictions DataFrame
image_predictions.tail()
```

Out[38]:
```
                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 [39]: # View info of image_predictions DataFrame
         image_predictions.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 [40]: # View descriptive statistics of image_predictions DataFrame
         image_predictions.describe()
```

```
Out[40]:            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 [41]: # View first 5 rows of tweet_info DataFrame
         tweet_info.head()
```

```
Out[41]:                    id  retweet_count  favorite_count
         0  892420643555336193           8560           38693
         1  892177421306343426           6293           33168
```

```
        2  891815181378084864          4176          24967
        3  891689557279858688          8691          42082
        4  891327558926688256          9452          40229
```

```
In [42]: # View last 5 rows of tweet_info DataFrame
         tweet_info.tail()
```

```
Out[42]:                       id  retweet_count  favorite_count
         2340  666049248165822465             41             109
         2341  666044226329800704            141             298
         2342  666033412701032449             45             125
         2343  666029285002620928             47             129
         2344  666020888022790149            517            2560
```

```
In [43]: # View info of tweet_info DataFrame
         tweet_info.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2345 entries, 0 to 2344
Data columns (total 3 columns):
id                2345 non-null int64
retweet_count     2345 non-null int64
favorite_count    2345 non-null int64
dtypes: int64(3)
memory usage: 55.0 KB
```

```
In [44]: # View descriptive statistics of tweet_info DataFrame
         tweet_info.describe()
```

```
Out[44]:                  id  retweet_count  favorite_count
         count  2.345000e+03    2345.000000     2345.000000
         mean   7.422940e+17    3015.556077     8045.715565
         std    6.833642e+16    5016.235535    12107.546778
         min    6.660209e+17       0.000000        0.000000
         25%    6.783802e+17     605.000000     1403.000000
         50%    7.189392e+17    1405.000000     3527.000000
         75%    7.986979e+17    3511.000000     9950.000000
         max    8.924206e+17   77143.000000   143024.000000
```

```
In [45]: # View rows in twitter_archive which contain '&amp;' instead of '&' in 'text' column
         twitter_archive[twitter_archive.text.str.contains('&amp;')]
```

```
Out[45]:                 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

                                                                              sou
262    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
273    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
320    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
461    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
485    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
516    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
799    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
889    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
898    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
976   <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck<
1104   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1179   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1199   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1222   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1274   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1366   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1421   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1465   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1481   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1508   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1524   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1538   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1569   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1593   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1621   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1646   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1707   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1763   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1795   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1812   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1817   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1842   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1897   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1899   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1901   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1913   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
1931   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
```

```
2031   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2037   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2064   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2084   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2096   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2137   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2177   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2190   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2196   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2207   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2210   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2216   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2232   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2246   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2268                   <a href="http://twitter.com" rel="nofollow">Twitter Web Client
2293   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
2306   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone


262   Meet Indie. She's not a fan of baths but she's definitely a fan of hide &amp; se
273                          RT @dog_rates: This is Pipsy. He is a fluffball. Enjoys
320       Meet Chester (bottom) &amp; Harold (top). They are different dogs not only i
461     Say hello to Eugene &amp; Patti Melt. No matter how dysfunctional they get, th
485                          RT @dog_rates: Meet Beau &amp; Wilbur. Wilbur stole Bea
516   Meet Sam. She smiles 24/7 &amp; secretly aspires to be a reindeer. \nKeep Sam sm
799                         Meet Roosevelt. He's preparing for takeoff. Make sure ti
889                          Meet Maggie &amp; Lila. Maggie is the doggo, Lila is th
898                                    This is Lilli Bee &amp; Honey Bear. Unfortuna
976                           Meet Jax &amp; Jil. Jil is yelling the pledge of alleg
1104                         Meet Buckley. His family &amp; some neighbors came over
1179                           Meet Sid &amp; Murphy. Murphy floats alongside Sid an
1199                         Meet Jennifur. She's supposed to be navigating. Not eve
1222                         Meet Travis and Flurp. Travis is pretty chill but Flurp
1274                       From left to right:\nCletus, Jerome, Alejandro, Burp, &am
1366                     Meet Rambo &amp; Kiwi. Rambo's the pup with the sharp toes
1421                         Meet Beau &amp; Wilbur. Wilbur stole Beau's bed from hi
1465                      Meet Oliviér. He takes killer selfies. Has a dog of his ow
1481                          This is Sadie and her 2 pups Shebang &amp; Ruffalo. Sa
1508                         When bae says they can't go out but you see them with s
1524                          This is Lolo. She's America af. Behind in science &amp
1538                      Meet Fynn &amp; Taco. Fynn is an all-powerful leaf lord an
1569                          Meet Trooper &amp; Maya. Trooper protects Maya from ba
1593                          Say hello to Crimson. He's a Speckled Winnebago. Main
1621                             Meet Bruiser &amp; Charlie. They are the best of pa
1646                          Here we see a faulty pupper. Might need to replace bat
1707                         Great picture here. Dog on the right panicked &amp; for
1763                          Touching scene here. Really stirs up the emotions. The
1795                        Meet Tassy &amp; Bee. Tassy is pretty chill, but Bee is
```

```
1812                            Say hello to Penny &amp; Gizmo. They are practicing th
1817                           This is Godzilla pupper. He had a ruff childhood &amp;
1842                           &amp; this is Yoshi. Another world record contender 11/
1897                           Meet Rufio. He is unaware of the pink legless pupper wi
1899                            Meet Jeb &amp; Bush. Jeb is somehow stuck in that fenc
1901                              Two gorgeous dogs here. Little waddling dog is a reb
1913                           Meet Chesney. On the outside he stays calm &amp; collec
1931                           Meet Daisy. She has no eyes &amp; her face has been blu
2031                           When you try to recreate the scene from Lady &amp; The
2037                           This is the best thing I've ever seen so spread it like
2064                            Meet Holly. She's trying to teach small human-like pup
2084                            Say hello to Andy. He can balance on one foot, obliter
2096                                                   Say hello to Gin &amp; Tonic. T
2137                           This is Ben &amp; Carson. It's impossible for them to t
2177                              Here we have Pancho and Peaches. Pancho is a Condole
2190                           Meet Jaycob. He got scared of the vacuum. Hide &amp; se
2196                           Say hello to Bobb. Bobb is a Golden High Fescue &amp; a
2207                           This is Timofy. He's a pilot for Southwest. It's Christ
2210                           Say hello to Kallie. There was a tornado in the area &a
2216                           This is Spark. He's nervous. Other dog hasn't moved in
2232                            These two dogs are Bo &amp; Smittens. Smittens is tryi
2246                                This is Tedrick. He lives on the edge. Needs someo
2268                            This is Dook &amp; Milo. Dook is struggling to find wh
2293                            This is Pipsy. He is a fluffball. Enjoys traveling the
2306                           These are Peruvian Feldspars. Their names are Cupit and

      retweeted_status_id  retweeted_status_user_id  \
262                   NaN                       NaN   
273          6.671522e+17              4.196984e+09   
320                   NaN                       NaN   
461                   NaN                       NaN   
485          6.981954e+17              4.196984e+09   
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

     retweeted_status_timestamp  \
262                         NaN
273   2015-11-19 01:27:25 +0000
320                         NaN
461                         NaN
485   2016-02-12 17:22:12 +0000
516                         NaN
799                         NaN
889                         NaN
898                         NaN
976                         NaN
1104                        NaN
1179                        NaN
1199                        NaN
```

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


262
273
320
461   https://twitter.com/dog_rates/status/817536400337801217/photo/1,https://twitter
485
```

```
516
799
889
898   https://twitter.com/dog_rates/status/758854675097526272/photo/1,https://twitter
976
1104
1179
1199
1222
1274
1366
1421
1465  https://twitter.com/dog_rates/status/694352839993344000/photo/1,https://twitter
1481
1508  https://twitter.com/dog_rates/status/691483041324204033/photo/1,https://twitter
1524
1538
1569
1593
1621
1646
1707
1763
1795
1812
1817
1842
1897
1899
1901
1913
1931
2031
2037
2064
2084                                                                  https://twitter
2096
2137
2177
2190
2196
2207
2210
2216
2232
2246
2268
2293
```

```
2306

      rating_numerator  rating_denominator       name  doggo floofer  pupper  \
262                 12                  10      Indie   None    None    None
273                 12                  10      Pipsy   None    None    None
320                 12                  10    Chester   None    None    None
461                 12                  10     Eugene   None    None    None
485                  9                  10       Beau   None    None    None
516                 24                   7        Sam   None    None    None
799                 11                  10  Roosevelt   None    None    None
889                 12                  10     Maggie  doggo    None  pupper
898                 11                  10      Lilli   None    None    None
976                 10                  10        Jax   None    None    None
1104                 9                  10    Buckley   None    None  pupper
1179                11                  10        Sid   None    None    None
1199                11                  10   Jennifur   None    None    None
1222                10                  10     Travis   None    None    None
1274                45                  50       None   None    None    None
1366                10                  10      Rambo   None    None    None
1421                 9                  10       Beau   None    None    None
1465                10                  10    Oliviér   None    None    None
1481                10                  10      Sadie   None    None    None
1508                 5                  10       None   None    None    None
1524                11                  10       Lolo   None    None    None
1538                11                  10       Fynn   None    None    None
1569                11                  10    Trooper   None    None    None
1593                11                  10    Crimson   None    None    None
1621                11                  10    Bruiser   None    None    None
1646                 9                  10       None   None    None  pupper
1707                10                  10       None   None    None    None
1763                10                  10       None   None    None    None
1795                10                  10      Tassy   None    None    None
1812                 9                  10      Penny   None    None    None
1817                 9                  10   Godzilla   None    None  pupper
1842                11                  10       None   None    None    None
1897                10                  10      Rufio   None    None  pupper
1899                 9                  10        Jeb   None    None    None
1901                 5                  10       None   None    None    None
1913                10                  10    Chesney   None    None    None
1931                 9                  10      Daisy   None    None    None
2031                10                  10       None   None    None    None
2037                13                  10        the   None    None    None
2064                11                  10      Holly   None    None    None
2084                11                  10       Andy   None    None    None
2096                 9                  10        Gin   None    None    None
2137                11                  10        Ben   None    None    None
2177                10                  10       None   None    None    None
2190                10                  10     Jaycob   None    None    None
```

```
2196                11                  10       Bobb   None    None    None
2207                 9                  10     Timofy   None    None    None
2210                10                  10     Kallie   None    None    None
2216                 8                  10      Spark   None    None    None
2232                10                  10       None   None    None    None
2246                 2                  10    Tedrick   None    None    None
2268                 8                  10       Dook   None    None    None
2293                12                  10      Pipsy   None    None    None
2306                10                  10       None   None    None    None

     puppo
262   None
273   None
320   None
461   None
485   None
516   None
799   None
889   None
898   None
976   None
1104  None
1179  None
1199  None
1222  None
1274  None
1366  None
1421  None
1465  None
1481  None
1508  None
1524  None
1538  None
1569  None
1593  None
1621  None
1646  None
1707  None
1763  None
1795  None
1812  None
1817  None
1842  None
1897  None
1899  None
1901  None
1913  None
1931  None
```

```
         2031  None
         2037  None
         2064  None
         2084  None
         2096  None
         2137  None
         2177  None
         2190  None
         2196  None
         2207  None
         2210  None
         2216  None
         2232  None
         2246  None
         2268  None
         2293  None
         2306  None
```

```
In [46]: # Sort values of 'name' column decending alphabetically
         twitter_archive.name.sort_values(ascending=False)
```

```
Out[46]: 1385                very
         819                 very
         1097                very
         773                 very
         1031                very
         1121        unacceptable
         1120                this
         1527                 the
         1797                 the
         1815                 the
         2212                 the
         2037                 the
         1603                 the
         2346                 the
         2345                 the
         22                  such
         2030               space
         193                quite
         118                quite
         169                quite
         2326               quite
         369                  one
         924                  one
         1936                 one
         993                  one
         1206                 old
         1747          officially
```

```
335              not
988              not
852               my
            ...
2195          Amélie
2078             Amy
1334         Ambrose
1495           Amber
1701           Alice
201            Alice
51              Alfy
858            Alfie
1616           Alfie
367            Alfie
661            Alfie
2238           Alfie
486              Alf
1189    Alexanderson
374        Alexander
2046       Alejandro
1115         Aldrick
412            Albus
144            Albus
1954          Albert
875           Albert
820               Al
480            Akumi
77               Aja
1934           Aiden
1327           Adele
1933            Acro
938              Ace
1021            Abby
1035            Abby
Name: name, Length: 2356, dtype: object
```

```
In [47]: # View number of entries for each source
         twitter_archive.source.value_counts()
```

```
Out[47]: <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>
         <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
         <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
         Name: source, dtype: int64
```

```
In [48]: # View rows where the value of 'name' is lowercase - indicates that it is not an actu
         twitter_archive.loc[(twitter_archive['name'].str.islower())]
```

```
Out[48]:                 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

                                                                                  sou
22    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
56    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
118   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
169   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
193   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
335                   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</
369   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
542   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
649   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
682   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
759   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
773   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
801   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
819   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
822   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
852   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
924   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
988   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
992   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
```

```
993   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1002  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1004  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1017  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1025  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1031  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1040  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1049  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1063  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1071  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1095                     <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</
...
2191  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2198  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2204  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2211  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2212                     <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</
2218  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2222  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2235  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2249  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2255                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client</
2264                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client</
2273                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client</
2287  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2304  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2311  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2314  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2326  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2327  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2333  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2334  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2335  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2345  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2346  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2347  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2348  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2349  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2350  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2352  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2353  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2354  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
```

```
22                                  I've yet to rate a Venezuelan Hover Wiener. This i
56           Here is a pupper approaching maximum borkdrive. Zooming at never before s
118                         RT @dog_rates: We only rate dogs. This is quite clearly a
169                 We only rate dogs. This is quite clearly a smol broken polar bear
```

```
193   Guys, we only rate dogs. This is quite clearly a bulbasaur. Please only send dog
335                                                               There's going to be a
369                                                               Occasionally, we're se
542                  We only rate dogs. Please stop sending in non-canines like this
649                                     Here is a perfect example of someone who has
682                                   RT @dog_rates: Say hello to mad pupper. You know
759                         RT @dog_rates: This is an East African Chalupa Seal. We on
773                         RT @dog_rates: We only rate dogs. Pls stop sending in non-
801                         Guys this is getting so out of hand. We only rate dogs. Th
819                          We only rate dogs. Pls stop sending in non-canines like
822                                                             RT @dog_rates: This is j
852                               This is my dog. Her name is Zoey. She knows I've bee
924                         This is one of the most inspirational stories I've ever co
988                             What jokester sent in a pic without a dog in it? This
992                          That is Quizno. This is his beach. He does not tolerate
993                         This is one of the most reckless puppers I've ever seen.
1002                        This is a mighty rare blue-tailed hammer sherk. Human almo
1004                          Viewer discretion is advised. This is a terrible attack
1017                              This is a carrot. We only rate dogs. Please only sen
1025                          This is an Iraqi Speed Kangaroo. It is not a dog. Please
1031                        We only rate dogs. Pls stop sending in non-canines like th
1040
1049                          This is a very rare Great Alaskan Bush Pupper. Hard to
1063                                                                         This is j
1071                         This is getting incredibly frustrating. This is a Mexican
1095                                                 Say hello to mad pupper. You know
...
2191                        This is a Slovakian Helter Skelter Feta named Leroi. Likes
2198                           This is a wild Toblerone from Papua New Guinea. Mouth a
2204                         This is an Irish Rigatoni terrier named Berta. Completely
2211                         Here is a horned dog. Much grace. Can jump over moons (da
2212                              Never forget this vine. You will not stop watching
2218                          This is a Birmingham Quagmire named Chuk. Loves to relax
2222                            Here is a mother dog caring for her pups. Snazzy red
2235                                   This is a Trans Siberian Kellogg named Alfonso
2249                        This is a Shotokon Macadamia mix named Cheryl. Sophisticat
2255                            This is a rare Hungarian Pinot named Jessiga. She is e
2264                                              This is a southwest Coriander named
2273                          This is a northern Wahoo named Kohl. He runs this town.
2287                              This is a Dasani Kingfisher from Maine. His name is
2304                                            This is a curly Ticonderoga named Pepe
2311                                                         This is a purebred Bacard
2314                         This is a golden Buckminsterfullerene named Johm. Drives
2326                        This is quite the dog. Gets really excited when not in wat
2327                         This is a southern Vesuvius bumblegruff. Can drive a truc
2333                         This is an extremely rare horned Parthenon. Not amused.
2334                          This is a funny dog. Weird toes. Won't come down. Loves
2335                            This is an Albanian 3 1/2 legged  Episcopalian. Loves
```

```
2345                                                    This is the happiest dog
2346                              Here is the Rand Paul of retrievers folks! He's proba
2347                                          My oh my. This is a rare blond Canadian te
2348                         Here is a Siberian heavily armored polar bear mix. Strong
2349                           This is an odd dog. Hard on the outside but loving on th
2350                         This is a truly beautiful English Wilson Staff retriever.
2352                            This is a purebred Piers Morgan. Loves to Netflix and c
2353                                   Here is a very happy pup. Big fan of well-mainta
2354                          This is a western brown Mitsubishi terrier. Upset about I

      retweeted_status_id  retweeted_status_user_id  \
22                    NaN                       NaN
56                    NaN                       NaN
118          8.591970e+17              4.196984e+09
169                   NaN                       NaN
193                   NaN                       NaN
335                   NaN                       NaN
369                   NaN                       NaN
542                   NaN                       NaN
649                   NaN                       NaN
682          7.363926e+17              4.196984e+09
759          7.030419e+17              4.196984e+09
773          7.007478e+17              4.196984e+09
801                   NaN                       NaN
819                   NaN                       NaN
822          7.410673e+17              4.196984e+09
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

     retweeted_status_timestamp  \
22                          NaN
56                          NaN
118   2017-05-02 00:04:57 +0000
169                         NaN
193                         NaN
335                         NaN
369                         NaN
542                         NaN
649                         NaN
682   2016-05-28 03:04:00 +0000
759   2016-02-26 02:20:37 +0000
773   2016-02-19 18:24:26 +0000
801                         NaN
819                         NaN
822   2016-06-10 00:39:48 +0000
852                         NaN
924                         NaN
988                         NaN
992                         NaN
993                         NaN
1002                        NaN
1004                        NaN
```

```
1017                        NaN
1025                        NaN
1031                        NaN
1040                        NaN
1049                        NaN
1063                        NaN
1071                        NaN
1095                        NaN
...                         ...
2191                        NaN
2198                        NaN
2204                        NaN
2211                        NaN
2212                        NaN
2218                        NaN
2222                        NaN
2235                        NaN
2249                        NaN
2255                        NaN
2264                        NaN
2273                        NaN
2287                        NaN
2304                        NaN
2311                        NaN
2314                        NaN
2326                        NaN
2327                        NaN
2333                        NaN
2334                        NaN
2335                        NaN
2345                        NaN
2346                        NaN
2347                        NaN
2348                        NaN
2349                        NaN
2350                        NaN
2352                        NaN
2353                        NaN
2354                        NaN


22
56
118
169
193
335
369                                                                   https://twitter
```

```
542
649   https://twitter.com/dog_rates/status/792913359805018113/photo/1,https://twitter
682
759
773
801                                                                   https://twitter
819
822
852
924   https://twitter.com/dog_rates/status/755206590534418437/photo/1,https://twitter
988
992
993
1002
1004
1017
1025
1031
1040
1049
1063
1071
1095
...
2191
2198
2204
2211
2212
2218
2222
2235
2249
2255
2264
2273
2287
2304
2311
2314
2326
2327
2333
2334
2335
2345
2346
2347
```

```
2348
2349
2350
2352
2353
2354

      rating_numerator  rating_denominator        name  doggo floofer  pupper  \
22                  14                  10        such   None    None    None
56                  14                  10           a   None    None  pupper
118                 12                  10       quite   None    None    None
169                 12                  10       quite   None    None    None
193                 12                  10       quite   None    None    None
335                 10                  10         not   None    None    None
369                 14                  10         one   None    None    None
542                 11                  10  incredibly   None    None    None
649                 13                  10           a   None    None    None
682                 13                  10         mad   None    None  pupper
759                 10                  10          an   None    None    None
773                 11                  10        very   None    None    None
801                 10                  10           a   None    None    None
819                 11                  10        very   None    None    None
822                 12                  10        just  doggo    None  pupper
852                 13                  10          my   None    None    None
924                 14                  10         one  doggo    None    None
988                 10                  10         not   None    None    None
992                 10                  10         his  doggo    None    None
993                  6                  10         one   None    None    None
1002                 8                  10           a   None    None    None
1004                 4                  10           a   None    None    None
1017                11                  10           a   None    None    None
1025                 9                  10          an   None    None    None
1031                 9                  10        very   None    None    None
1040                12                  10    actually   None    None  pupper
1049                12                  10           a   None    None  pupper
1063                12                  10        just  doggo    None  pupper
1071                10                  10     getting   None    None    None
1095                13                  10         mad   None    None  pupper
...                ...                 ...         ...    ...     ...     ...
2191                10                  10           a   None    None    None
2198                 7                  10           a   None    None    None
2204                10                  10          an   None    None    None
2211                 7                  10           a   None    None    None
2212                13                  10         the   None    None    None
2218                10                  10           a   None    None    None
2222                 4                  10           a   None    None    None
2235                 7                  10           a   None    None    None
2249                 9                  10           a   None    None    None
```

```
2255                 8                  10            a   None    None    None
2264                 9                  10            a   None    None    None
2273                11                  10            a   None    None    None
2287                 8                  10            a   None    None    None
2304                11                  10            a   None    None    None
2311                10                  10            a   None    None    None
2314                 8                  10            a   None    None    None
2326                 2                  10        quite   None    None    None
2327                 7                  10            a   None    None    None
2333                 9                  10           an   None    None    None
2334                 3                  10            a   None    None    None
2335                 1                   2           an   None    None    None
2345                10                  10          the   None    None    None
2346                 8                  10          the   None    None    None
2347                 9                  10            a   None    None    None
2348                10                  10            a   None    None    None
2349                 2                  10           an   None    None    None
2350                10                  10            a   None    None    None
2352                 6                  10            a   None    None    None
2353                 9                  10            a   None    None    None
2354                 7                  10            a   None    None    None

     puppo
22    None
56    None
118   None
169   None
193   None
335   None
369   None
542   None
649   None
682   None
759   None
773   None
801   None
819   None
822   None
852   None
924   None
988   None
992   None
993   None
1002  None
1004  None
1017  None
1025  None
1031  None
```

```
1040  None
1049  None
1063  None
1071  None
1095  None
...    ...
2191  None
2198  None
2204  None
2211  None
2212  None
2218  None
2222  None
2235  None
2249  None
2255  None
2264  None
2273  None
2287  None
2304  None
2311  None
2314  None
2326  None
2327  None
2333  None
2334  None
2335  None
2345  None
2346  None
2347  None
2348  None
2349  None
2350  None
2352  None
2353  None
2354  None

[109 rows x 17 columns]
```

```
In [49]: # View rows where the value of 'name' is lowercase and the word 'named' appears in th
         # there is an actual dog name in the text
         twitter_archive.loc[(twitter_archive['name'].str.islower()) & (twitter_archive['text']
```

```
Out[49]:                 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

                                                                                 sou
1853  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1955  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2034  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2066  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2116  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2125  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2128  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
2146  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
```

```
2161  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2191  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2204  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2218  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2235  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2249  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2255                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client<
2264                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client<
2273                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client<
2304  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2311  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<
2314  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone<


1853  This is a Sizzlin Menorah spaniel from Brooklyn named Wylie. Lovable eyes. Chil
1955   This is a Lofted Aphrodisiac Terrier named Kip. Big fan of bed n breakfasts. F
2034      This is a Tuscaloosa Alcatraz named Jacob (Yacb). Loves to sit in swing. St
2066  This is a Helvetica Listerine named Rufus. This time Rufus will be ready for th
2116   This is a Deciduous Trimester mix named Spork. Only 1 ear works. No seat belt.
2125     This is a Rich Mahogany Seltzer named Cherokee. Just got destroyed by a snow
2128    This is a Speckled Cauliflower Yosemite named Hemry. He's terrified of intrud
2146    This is a spotted Lipitor Rumpelstiltskin named Alphred. He can't wait for th
2161    This is a Coriander Baton Rouge named Alfredo. Loves to cuddle with smaller w
2191  This is a Slovakian Helter Skelter Feta named Leroi. Likes to skip on roofs. Go
2204   This is an Irish Rigatoni terrier named Berta. Completely made of rope. No eye
2218    This is a Birmingham Quagmire named Chuk. Loves to relax and watch the game w
2235             This is a Trans Siberian Kellogg named Alfonso. Huge ass eyeballs. A
2249  This is a Shotokon Macadamia mix named Cheryl. Sophisticated af. Looks like a d
2255      This is a rare Hungarian Pinot named Jessiga. She is either mid-stroke or g
2264                        This is a southwest Coriander named Klint. Hat looks expe
2273    This is a northern Wahoo named Kohl. He runs this town. Chases tumbleweeds. D
2304                      This is a curly Ticonderoga named Pepe. No feet. Loves to j
2311                                   This is a purebred Bacardi named Octaviath. Ca
2314   This is a golden Buckminsterfullerene named Johm. Drives trucks. Lumberjack (?

      retweeted_status_id  retweeted_status_user_id  \
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

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

                                                        expanded_urls  \
1853  https://twitter.com/dog_rates/status/675706639471788032/photo/1
1955  https://twitter.com/dog_rates/status/673636718965334016/photo/1
2034  https://twitter.com/dog_rates/status/671743150407421952/photo/1
2066  https://twitter.com/dog_rates/status/671147085991960577/photo/1
2116  https://twitter.com/dog_rates/status/670427002554466305/photo/1
2125  https://twitter.com/dog_rates/status/670361874861563904/photo/1
2128  https://twitter.com/dog_rates/status/670303360680108032/photo/1
2146  https://twitter.com/dog_rates/status/669923323644657664/photo/1
2161  https://twitter.com/dog_rates/status/669564461267722241/photo/1
2191  https://twitter.com/dog_rates/status/668955713004314625/photo/1
2204  https://twitter.com/dog_rates/status/668636665813057536/photo/1
2218  https://twitter.com/dog_rates/status/668507509523615744/photo/1
2235  https://twitter.com/dog_rates/status/668171859951755264/photo/1
2249  https://twitter.com/dog_rates/status/667861340749471744/photo/1
2255  https://twitter.com/dog_rates/status/667773195014021121/photo/1
2264  https://twitter.com/dog_rates/status/667538891197542400/photo/1
```

```
2273  https://twitter.com/dog_rates/status/667470559035432960/photo/1
2304  https://twitter.com/dog_rates/status/666983947667116034/photo/1
2311  https://twitter.com/dog_rates/status/666781792255496192/photo/1
2314  https://twitter.com/dog_rates/status/666701168228331520/photo/1

      rating_numerator  rating_denominator name doggo floofer pupper puppo
1853                10                  10    a  None    None   None  None
1955                10                  10    a  None    None   None  None
2034                11                  10    a  None    None   None  None
2066                 9                  10    a  None    None   None  None
2116                 9                  10    a  None    None   None  None
2125                 9                  10    a  None    None   None  None
2128                 9                  10    a  None    None   None  None
2146                10                  10    a  None    None   None  None
2161                10                  10    a  None    None   None  None
2191                10                  10    a  None    None   None  None
2204                10                  10   an  None    None   None  None
2218                10                  10    a  None    None   None  None
2235                 7                  10    a  None    None   None  None
2249                 9                  10    a  None    None   None  None
2255                 8                  10    a  None    None   None  None
2264                 9                  10    a  None    None   None  None
2273                11                  10    a  None    None   None  None
2304                11                  10    a  None    None   None  None
2311                10                  10    a  None    None   None  None
2314                 8                  10    a  None    None   None  None
```

```
In [50]: # View rows where the value of 'name' is lowercase and the words 'name is' appears in
         # there is an actual dog name in the text
         twitter_archive.loc[(twitter_archive['name'].str.islower()) & (twitter_archive['text']
```

```
Out[50]:                 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


                                                                                   sou
         852   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
         2287  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</


         852   This is my dog. Her name is Zoey. She knows I've been rating other dogs. She's r
         2287  This is a Dasani Kingfisher from Maine. His name is Daryl. Daryl doesn't like be

               retweeted_status_id  retweeted_status_user_id  \
852                   NaN                       NaN
2287                  NaN                       NaN

     retweeted_status_timestamp  \
852                         NaN
2287                        NaN

                                                        expanded_urls  \
852   https://twitter.com/dog_rates/status/765395769549590528/photo/1
2287  https://twitter.com/dog_rates/status/667177989038297088/photo/1

      rating_numerator  rating_denominator name doggo floofer pupper puppo
852                 13                  10   my  None    None   None  None
2287                 8                  10    a  None    None   None  None
```

In [51]:

```
# View row where dog name is 'O' but the dogs name given in 'text' column is 'O'Malle
twitter_archive[twitter_archive.name == "O"]
```

Out[51]:

```
               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

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


775  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 10

     retweeted_status_id  retweeted_status_user_id retweeted_status_timestamp  \
775                  NaN                       NaN                        NaN

                                                       expanded_urls  \
775  https://twitter.com/dog_rates/status/776201521193218049/photo/1

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

In [52]:

```
#disable warnings
warnings.simplefilter('ignore')
```

In [53]:

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

Out[53]:

```
                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

                                                                                  sou
45    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
340   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
695   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
763   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1689  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
1712  <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</


45                            This is Bella. She hopes her smile made you smile. If n
340                       RT @dog_rates: This is Logan, the Chow who lived. He solemn
695                                     This is Logan, the Chow who lived. He solemnly
763   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rand
1689                                                           I've been told there's a
1712                                                           Here we have uncovered an

      retweeted_status_id  retweeted_status_user_id  \
45                    NaN                       NaN
340          7.867091e+17              4.196984e+09
695                   NaN                       NaN
763                   NaN                       NaN
1689                  NaN                       NaN
1712                  NaN                       NaN

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


45    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter
340                                                                 https://twitter
```

```
695                                                                https://twitter
763                                                                https://twitter
1689
1712                                                               https://twitter

      rating_numerator  rating_denominator    name doggo floofer  pupper puppo
45                   5                  10   Bella  None    None    None  None
340                 75                  10   Logan  None    None    None  None
695                 75                  10   Logan  None    None    None  None
763                 27                  10  Sophie  None    None  pupper  None
1689                 5                  10    None  None    None    None  None
1712                26                  10    None  None    None    None  None
```

```
In [54]: # tweet_id = 810984652412424192 doesn't have a rating
         twitter_archive[twitter_archive.tweet_id == 810984652412424192]
```

```
Out[54]:                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

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


         516  Meet Sam. She smiles 24/7 &amp; secretly aspires to be a reindeer. \nKeep Sam smi

              retweeted_status_id  retweeted_status_user_id retweeted_status_timestamp  \
         516                  NaN                       NaN                        NaN


         516  https://www.gofundme.com/sams-smile,https://twitter.com/dog_rates/status/8109846

              rating_numerator  rating_denominator name doggo floofer pupper puppo
         516                24                   7  Sam  None    None   None  None
```

0.2.1 Quality issues

- Dataset contains retweets.
- Tweets with no images
- Source contains whole tag info. Can extract source name from it.
- Contents of 'text' cutoff (fixed already by increasing the display width during assessing data).
- Extra characters after '&'
- Incorrect dog names
- Missing values in 'name' and dog stages showing as 'None'
- Rating numerators with decimals not showing full float

- Tweet with more than one #/# sometimes have the first occurence used for the rating numerators and denominators
- Tweet ID# 810984652412424192 doesn't contain a rating
- Erroneous datatypes (timestamp, source, dog stages, tweet_id, in_reply_to_status_id, in_reply_to_user_id)

0.2.2 Tidyness issues

- Dog "stage" variable in four columns: doggo, floofer, pupper, puppo
- Join 'tweet_info' and 'image_predictions' to 'twitter_archive'

0.3 Clean

```
In [64]: # Create copies of original DataFrames to work off of
         twitter_archive_clean = twitter_archive.copy()
         image_predictions_clean = image_predictions.copy()
         tweet_info_clean = tweet_info.copy()
```

0.3.1 Define

Remove rows with missing images.

Code

```
In [65]: twitter_archive_clean = twitter_archive_clean.dropna(subset=['expanded_urls'])
```

Test

```
In [66]: sum(twitter_archive_clean['expanded_urls'].isnull())

Out[66]: 0
```

0.3.2 Define

- Remove retweets.
- Remove retweeted columns.

Code

```
In [67]: # Select rows where 'retweeted_status_id' is null to save to twitter_archive_clean
         twitter_archive_clean = twitter_archive_clean[twitter_archive_clean['retweeted_status

         retweet_columns = ['retweeted_status_id', 'retweeted_status_user_id', 'retweeted_statu
         twitter_archive_clean = twitter_archive_clean.drop(retweet_columns, axis=1)
```

Test

```
In [68]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2117 entries, 0 to 2355
Data columns (total 14 columns):
tweet_id                 2117 non-null int64
in_reply_to_status_id    23 non-null float64
in_reply_to_user_id      23 non-null float64
timestamp                2117 non-null object
source                   2117 non-null object
text                     2117 non-null object
expanded_urls            2117 non-null object
rating_numerator         2117 non-null int64
rating_denominator       2117 non-null int64
name                     2117 non-null object
doggo                    2117 non-null object
floofer                  2117 non-null object
pupper                   2117 non-null object
puppo                    2117 non-null object
dtypes: float64(2), int64(3), object(9)
memory usage: 248.1+ KB
```

0.3.3 Define

Source column contains full html tag info. Correct source values.

Code

```
In [69]: # Update source column with source name only
         source_dict = {
                         '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter i
                         '<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>': 'Vi
                         '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>':
                         '<a href="https://about.twitter.com/products/tweetdeck" rel="nofollow
                         }

         for key,value in source_dict.items():
             twitter_archive_clean.source = twitter_archive_clean.source.str.replace(key,value
```

Test

```
In [70]: twitter_archive_clean.source.value_counts()
```

```
Out[70]: Twitter for iPhone     1985
         Vine - Make a Scene      91
         Twitter Web Client       30
         TweetDeck                11
         Name: source, dtype: int64
```

0.3.4 Define

Combine individual dog stage columns 'puppo','pupper','floofer', 'doggo' into single variable dog stage and remove individual dog stage columns.

Code

```
In [71]: # extracting the dog stage variable into 'dog_stage' variable from the text column
         twitter_archive_clean['dog_stage'] = twitter_archive_clean.text.str.extract('(puppo|pu
```

```
In [72]: # drop individual dog stage columns
         dog_stage_columns = ['doggo', 'floofer', 'pupper', 'puppo']
         twitter_archive_clean = twitter_archive_clean.drop(dog_stage_columns, axis=1)
```

Test

```
In [73]: twitter_archive_clean.dog_stage.value_counts()
```

```
Out[73]: pupper     242
         doggo       81
         puppo       29
         floofer      4
         Name: dog_stage, dtype: int64
```

```
In [74]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2117 entries, 0 to 2355
Data columns (total 11 columns):
tweet_id                 2117 non-null int64
in_reply_to_status_id    23 non-null float64
in_reply_to_user_id      23 non-null float64
timestamp                2117 non-null object
source                   2117 non-null object
text                     2117 non-null object
expanded_urls            2117 non-null object
rating_numerator         2117 non-null int64
rating_denominator       2117 non-null int64
name                     2117 non-null object
dog_stage                356 non-null object
dtypes: float64(2), int64(3), object(6)
memory usage: 198.5+ KB
```

0.3.5 Define

Remove extra characters after '&' in text column.

Code

```
In [75]: twitter_archive_clean.text = twitter_archive_clean.text.str.replace('&amp;', '&')
```

Test

```
In [76]: twitter_archive_clean[twitter_archive_clean.text.str.contains('&amp;')]
```

```
Out[76]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, tex
         Index: []
```

0.3.6 Define

Extract correct dog names from ’text’ column and update column ’name’.

Code

```
In [77]: # Save locations where 'name' column is lowercase, lowercase and 'text' column contai
         # column contains the words 'named', 'name is'
         dog_named = twitter_archive_clean.loc[(twitter_archive_clean.name.str.islower()) \
                                               & (twitter_archive_clean.text.str.contains('name
         dog_name_is = twitter_archive_clean.loc[(twitter_archive_clean.name.str.islower()) \
                                                 & (twitter_archive_clean.text.str.contains('na
         dog_no_name = twitter_archive_clean.loc[(twitter_archive_clean.name.str.islower()) \
                                                  & ~(twitter_archive_clean.text.str.contains(
                                                  & ~(twitter_archive_clean.text.str.contains(

         # Save these locations as lists
         dog_named = dog_named['text'].tolist()
         dog_name_is = dog_name_is['text'].tolist()
         dog_no_name = dog_no_name['text'].tolist()
```

```
In [78]: # get the name from 'text' where name is lowercase and 'text' contains the word 'name
         # name of the dog would be the word appeared after the word 'named' or 'name is'
         column_to_update = 'name'
         for content in dog_named:
             mask = twitter_archive_clean.text == content
             twitter_archive_clean.loc[mask, column_to_update] = re.findall(r'named\s(\w+)', co

         # get the name from 'text' where name is lowercase and 'text' contains the words 'nam
         # name of the dog would be the word appeared after the words 'name is'
         for content in dog_name_is:
             mask = twitter_archive_clean.text == content
             twitter_archive_clean.loc[mask, column_to_update] = re.findall(r'name is\s(\w+)',

         # set the name as 'None' where name is lowercase and name not found in 'text'
         for content in dog_no_name:
             mask = twitter_archive_clean.text == content
             twitter_archive_clean.loc[mask, column_to_update] = "None"
```

```
In [79]: # Replace dog named "O" with "O'Malley"
         twitter_archive_clean.name = twitter_archive_clean.name.replace("O", "O'Malley")
```

Test

```
In [80]: twitter_archive_clean.name.sort_values(ascending = False)
```

```
Out[80]: 1875            Zuzu
         151            Zooey
         2141            Zoey
         115             Zoey
         8               Zoey
         852             Zoey
         966              Zoe
         1124            Ziva
         1210            Zeus
         547             Zeke
         181             Zeke
         17              Zeke
         2206            Zeek
         1332            Zara
         1409           Yukon
         43              Yogi
         622             Yogi
         1378            Yoda
         1853           Wylie
         174            Wyatt
         410            Wyatt
         1451           Wyatt
         877           Wishes
         986          Winston
         280          Winston
         816          Winston
         1243         Winston
         559          Winston
         2133         Winston
         407          Winston
                     ...
         1334         Ambrose
         1495           Amber
         2146         Alphred
         1701           Alice
         201            Alice
         51              Alfy
         2161         Alfredo
         2235         Alfonso
         367            Alfie
         858            Alfie
         2238           Alfie
         1616           Alfie
         486              Alf
         1189    Alexanderson
         374        Alexander
         2046       Alejandro
         1115         Aldrick
         144            Albus
         412            Albus
         875           Albert
         1954          Albert
         820               Al
         480            Akumi
         77               Aja
         1934           Aiden
         1327           Adele
         1933            Acro
         938              Ace
         1035            Abby
         1021            Abby
         Name: name, Length: 2117, dtype: object
```

```
In [81]: twitter_archive_clean[twitter_archive_clean.name.str.islower()]
```

```
Out[81]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, te
         Index: []
```

```
In [82]: twitter_archive_clean[twitter_archive_clean.name == "O'Malley"]
```

```
Out[82]:                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         775  776201521193218049                    NaN                  NaN

                                  timestamp              source  \
         775  2016-09-14 23:30:38 +0000  Twitter for iPhone


         775  This is O'Malley. That is how he sleeps. Doesn't care what you think about it. 1

                                                                expanded_urls  \
         775  https://twitter.com/dog_rates/status/776201521193218049/photo/1

              rating_numerator  rating_denominator      name dog_stage
         775                10                  10  O'Malley       NaN
```

0.3.7 Define

Change missing values in 'name' from 'None' to NaN. Missing dog stages values already resolved while correcting dog stages.

Code

```
In [83]: twitter_archive_clean.name = twitter_archive_clean.name.replace('None', np.NaN)
```

Test

```
In [84]: twitter_archive_clean[twitter_archive_clean.name == 'None']
```

```
Out[84]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, te
         Index: []
```

0.3.8 Define

Combine tweet_info and image_predictions tables with twitter_archive table.

Code

```
In [85]: # Join tweet_info_clean and image_predictions_clean with twitter_archive
         twitter_archive_clean = pd.merge(pd.merge(twitter_archive_clean, tweet_info_clean, le
                                                        , right_on='id', how='inner')\
                                               , image_predictions_clean, on='tweet_id', how='
```

```
In [86]: # As 'tweet_id' and 'id' columns contains same values. So can drop 'id' column
         twitter_archive_clean = twitter_archive_clean.drop('id', axis=1)
```

Test

```
In [87]: twitter_archive_clean.head()
```

```
Out[87]:              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

                          timestamp              source  \
         0  2017-08-01 16:23:56 +0000  Twitter for iPhone
         1  2017-08-01 00:17:27 +0000  Twitter for iPhone
         2  2017-07-31 00:18:03 +0000  Twitter for iPhone
         3  2017-07-30 15:58:51 +0000  Twitter for iPhone
         4  2017-07-29 16:00:24 +0000  Twitter for iPhone


         0                                                     This is Phineas. He's a mysti
         1  This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she'
         2                    This is Archie. He is a rare Norwegian Pouncing Corgo. Lives in th
         3                                                           This is Darla. She comme
         4  This is Franklin. He would like you to stop calling him "cute." He is a very fierc


         0                                                                  https://twitter.co
```

```
1                                                                  https://twitter.co
2                                                                  https://twitter.co
3                                                                  https://twitter.co
4  https://twitter.com/dog_rates/status/891327558926688256/photo/1,https://twitter.co

   rating_numerator  rating_denominator      name   ...   img_num  \
0                13                  10   Phineas   ...         1
1                13                  10     Tilly   ...         1
2                12                  10    Archie   ...         1
3                13                  10     Darla   ...         1
4                12                  10  Franklin   ...         2

            p1   p1_conf p1_dog                 p2   p2_conf  p2_dog  \
0       orange  0.097049  False              bagel  0.085851   False
1    Chihuahua  0.323581   True           Pekinese  0.090647    True
2    Chihuahua  0.716012   True           malamute  0.078253    True
3  paper_towel  0.170278  False  Labrador_retriever  0.168086    True
4       basset  0.555712   True    English_springer  0.225770    True

                            p3   p3_conf  p3_dog
0                       banana  0.076110   False
1                     papillon  0.068957    True
2                       kelpie  0.031379    True
3                      spatula  0.040836   False
4  German_short-haired_pointer  0.175219    True

[5 rows x 24 columns]
```

```
In [88]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 24 columns):
tweet_id                 1994 non-null int64
in_reply_to_status_id    23 non-null float64
in_reply_to_user_id      23 non-null float64
timestamp                1994 non-null object
source                   1994 non-null object
text                     1994 non-null object
expanded_urls            1994 non-null object
rating_numerator         1994 non-null int64
rating_denominator       1994 non-null int64
name                     1372 non-null object
dog_stage                326 non-null object
retweet_count            1994 non-null int64
favorite_count           1994 non-null int64
jpg_url                  1994 non-null object
img_num                  1994 non-null int64
```

```
p1                        1994 non-null object
p1_conf                   1994 non-null float64
p1_dog                    1994 non-null bool
p2                        1994 non-null object
p2_conf                   1994 non-null float64
p2_dog                    1994 non-null bool
p3                        1994 non-null object
p3_conf                   1994 non-null float64
p3_dog                    1994 non-null bool
dtypes: bool(3), float64(5), int64(6), object(10)
memory usage: 348.6+ KB
```

0.3.9 Define

Correct rating numerator and denominators.

Code

```
In [89]: # View all occurences where there are more than one #/# in 'text' column
         fix_rating_of = twitter_archive_clean.loc[twitter_archive_clean.text.str.contains( r"
         fix_rating_of = fix_rating_of['text'].tolist()
```

```
In [90]: # Loop through the list of ratings to fix and extract the second occurence of #/ to g
         # And set rating_denominator as 10 because the actual ratings are based on scale of 1
         #entry instead of extracting it.
         col1 = 'rating_numerator'
         col2 = 'rating_denominator'

         for content in fix_rating_of:
             mask = twitter_archive_clean.text == content
             twitter_archive_clean.loc[mask, col1] = re.findall(r"\d+\.?\d*\/\d+\.?\d*\D+(\d+\
             twitter_archive_clean.loc[mask, col2] = 10
```

Test

```
In [91]: twitter_archive_clean[twitter_archive_clean.text.isin(fix_rating_of)]
```

```
Out[91]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         555   777684233540206592                    NaN                  NaN
         749   747600769478692864                    NaN                  NaN
         800   740373189193256964                    NaN                  NaN
         891   722974582966214656                    NaN                  NaN
         925   716439118184652801                    NaN                  NaN
         941   714258258790387713                    NaN                  NaN
         1062  703356393781329922                    NaN                  NaN
         1152  695064344191721472                    NaN                  NaN
         1157  694352839993344000                    NaN                  NaN
         1192  691483041324204033                    NaN                  NaN
```

```
1207  690400367696297985                    NaN                  NaN
1218  689835978131935233                    NaN                  NaN
1328  682962037429899265                    NaN                  NaN
1450  677314812125323265                    NaN                  NaN
1484  676191832485810177                    NaN                  NaN
1546  674737130913071104                    NaN                  NaN
1550  674646392044941312                    NaN                  NaN
1615  673295268553605120                    NaN                  NaN
1655  672248013293752320                    NaN                  NaN
1708  671154572044468225                    NaN                  NaN
1757  670434127938719744                    NaN                  NaN
1820  669037058363662336                    NaN                  NaN
1857  668537837512433665                    NaN                  NaN
1902  667544320556335104                    NaN                  NaN
1911  667491009379606528                    NaN                  NaN
1944  666835007768551424                    NaN                  NaN
1973  666287406224695296                    NaN                  NaN

                      timestamp              source  \
555   2016-09-19 01:42:24 +0000  Twitter for iPhone
749   2016-06-28 01:21:27 +0000  Twitter for iPhone
800   2016-06-08 02:41:38 +0000  Twitter for iPhone
891   2016-04-21 02:25:47 +0000  Twitter for iPhone
925   2016-04-03 01:36:11 +0000  Twitter for iPhone
941   2016-03-28 01:10:13 +0000  Twitter for iPhone
1062  2016-02-26 23:10:06 +0000  Twitter for iPhone
1152  2016-02-04 02:00:27 +0000  Twitter for iPhone
1157  2016-02-02 02:53:12 +0000  Twitter for iPhone
1192  2016-01-25 04:49:38 +0000  Twitter for iPhone
1207  2016-01-22 05:07:29 +0000  Twitter for iPhone
1218  2016-01-20 15:44:48 +0000  Twitter for iPhone
1328  2016-01-01 16:30:13 +0000  Twitter for iPhone
1450  2015-12-17 02:30:09 +0000  Twitter for iPhone
1484  2015-12-14 00:07:50 +0000  Twitter for iPhone
1546  2015-12-09 23:47:22 +0000  Twitter for iPhone
1550  2015-12-09 17:46:48 +0000  Twitter for iPhone
1615  2015-12-06 00:17:55 +0000  Twitter for iPhone
1655  2015-12-03 02:56:30 +0000  Twitter for iPhone
1708  2015-11-30 02:31:34 +0000  Twitter for iPhone
1757  2015-11-28 02:48:46 +0000  Twitter for iPhone
1820  2015-11-24 06:17:19 +0000  Twitter for iPhone
1857  2015-11-22 21:13:35 +0000  Twitter for iPhone
1902  2015-11-20 03:25:43 +0000  Twitter Web Client
1911  2015-11-19 23:53:52 +0000  Twitter Web Client
1944  2015-11-18 04:27:09 +0000  Twitter for iPhone
1973  2015-11-16 16:11:11 +0000  Twitter for iPhone
```

```
555                               "Yep... just as I suspected. You're not flossing." 12
749                 This is Bookstore and Seaweed. Bookstore is tired and Seaweed is a
800     After so many requests, this is Bretagne. She was the last surviving 9/11 sear
891                                                                                 Ha
925                             This is Bluebert. He just saw that both #FinalFur match
941     Meet Travis and Flurp. Travis is pretty chill but Flurp can't lie down properl
1062          This is Socks. That water pup w the super legs just splashed him. Socks
1152      This may be the greatest video I've ever been sent. 4/10 for Charles the pup
1157     Meet Oliviér. He takes killer selfies. Has a dog of his own. It leaps at rand
1192    When bae says they can't go out but you see them with someone else that same n
1207  This is Eriq. His friend just reminded him of last year's super bowl. Not cool f
1218     Meet Fynn & Taco. Fynn is an all-powerful leaf lord and Taco is in the wrong
1328     This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Wa
1450       Meet Tassy & Bee. Tassy is pretty chill, but Bee is convinced the Ruffles a
1484       These two pups just met and have instantly bonded. Spectacular scene. Mesme
1546    Meet Rufio. He is unaware of the pink legless pupper wrapped around him. Might
1550       Two gorgeous dogs here. Little waddling dog is a rebel. Refuses to look at
1615    Meet Eve. She's a raging alcoholic 8/10 (would b 11/10 but pupper alcoholism i
1655                                                     10/10 for dog. 7/10 for cat. 12/
1708     Meet Holly. She's trying to teach small human-like pup about blocks but he's
1757         Meet Hank and Sully. Hank is very proud of the pumpkin they found and Sul
1820       Here we have Pancho and Peaches. Pancho is a Condoleezza Gryffindor, and Pe
1857    This is Spark. He's nervous. Other dog hasn't moved in a while. Won't come whe
1902     This is Kial. Kial is either wearing a cape, which would be rad, or flashing
1911      Two dogs in this one. Both are rare Jujitsu Pythagoreans. One slightly white
1944    These are Peruvian Feldspars. Their names are Cupit and Prencer. Both resemble
1973        This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwo


555
749
800   https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter
891
925
941
1062
1152
1157  https://twitter.com/dog_rates/status/694352839993344000/photo/1,https://twitter
1192  https://twitter.com/dog_rates/status/691483041324204033/photo/1,https://twitter
1207
1218
1328
1450
1484                                                                  https://twitter
1546
1550
1615
1655
```

```
1708
1757
1820
1857
1902
1911
1944
1973

     rating_numerator  rating_denominator       name   ...   img_num  \
555                11                  10        NaN   ...         1
749                 7                  10  Bookstore   ...         1
800                14                  10        NaN   ...         3
891                13                  10        NaN   ...         1
925                11                  10   Bluebert   ...         1
941                 8                  10     Travis   ...         1
1062                2                  10      Socks   ...         1
1152               13                  10        NaN   ...         1
1157                5                  10    Oliviér   ...         2
1192               10                  10        NaN   ...         1
1207                6                  10       Eriq   ...         1
1218               10                  10       Fynn   ...         1
1328               10                  10     Darrel   ...         1
1450               11                  10      Tassy   ...         2
1484                7                  10        NaN   ...         2
1546                4                  10      Rufio   ...         1
1550                8                  10        NaN   ...         1
1615               11                  10        Eve   ...         1
1655                7                  10        NaN   ...         1
1708                8                  10      Holly   ...         1
1757                8                  10       Hank   ...         1
1820                7                  10        NaN   ...         1
1857                1                  10      Spark   ...         1
1902                4                  10       Kial   ...         1
1911                8                  10        NaN   ...         1
1944               10                  10        NaN   ...         1
1973                9                  10        NaN   ...         1

                            p1   p1_conf p1_dog                          p2  \
555             cocker_spaniel  0.253442   True            golden_retriever
749   Chesapeake_Bay_retriever  0.804363   True                  Weimaraner
800           golden_retriever  0.807644   True                      kuvasz
891                 Great_Dane  0.246762   True  Greater_Swiss_Mountain_dog
925             Siberian_husky  0.396495   True                    malamute
941                     collie  0.176758   True    Chesapeake_Bay_retriever
1062             Border_collie  0.894842   True                      collie
1152                 seat_belt  0.522211  False                  sunglasses
1157        Australian_terrier  0.407886   True           Yorkshire_terrier
```

```
1192                bloodhound  0.886232   True     black-and-tan_coonhound
1207                  Pembroke  0.426459   True                    papillon
1218                    collie  0.600186   True           Shetland_sheepdog
1328                     dingo  0.278600  False                   Chihuahua
1450          Blenheim_spaniel  0.924127   True            Japanese_spaniel
1484                 Chihuahua  0.376741   True           Italian_greyhound
1546                Pomeranian  0.948537   True                  schipperke
1550     flat-coated_retriever  0.837448   True                 groenendael
1615          golden_retriever  0.889241   True          Labrador_retriever
1655             Irish_terrier  0.413173   True                    Airedale
1708        Labrador_retriever  0.495047   True    Chesapeake_Bay_retriever
1757            jack-o'-lantern  0.919140  False    Chesapeake_Bay_retriever
1820                 Chihuahua  0.803528   True                  Pomeranian
1857          Lakeland_terrier  0.372988   True                  toy_poodle
1902                Pomeranian  0.412893   True                    Pembroke
1911                    borzoi  0.852088   True                    ice_bear
1944                  Airedale  0.448459   True                  toy_poodle
1973               Maltese_dog  0.857531   True                  toy_poodle

       p2_conf  p2_dog                  p3   p3_conf  p3_dog
555   0.162850    True          otterhound  0.110921    True
749   0.054431    True  Labrador_retriever  0.043268    True
800   0.101286    True  Labrador_retriever  0.023785    True
891   0.126131    True          Weimaraner  0.085297    True
925   0.317053    True          Eskimo_dog  0.273419    True
941   0.101834    True              beagle  0.101294    True
1062  0.097364    True    English_springer  0.003037    True
1152  0.077552   False           ice_lolly  0.051774   False
1157  0.328173    True       silky_terrier  0.108404    True
1192  0.077420    True       Gordon_setter  0.009826    True
1207  0.317368    True   Shetland_sheepdog  0.077616    True
1218  0.298939    True              borzoi  0.022616    True
1328  0.155207    True               loupe  0.153598   False
1450  0.054790    True           Chihuahua  0.008204    True
1484  0.173114    True              muzzle  0.071485   False
1546  0.014310    True           Chihuahua  0.008120    True
1550  0.086166    True  Labrador_retriever  0.016052    True
1615  0.064683    True      Great_Pyrenees  0.012613    True
1655  0.335616    True          toy_poodle  0.027952    True
1708  0.350188    True    golden_retriever  0.142400    True
1757  0.027351    True  Labrador_retriever  0.020081    True
1820  0.053871    True                chow  0.032257    True
1857  0.250445    True           Chihuahua  0.189737    True
1902  0.312958    True           Chihuahua  0.071960    True
1911  0.132264   False              weasel  0.005730   False
1944  0.124030    True               teddy  0.110183   False
1973  0.063064    True    miniature_poodle  0.025581    True
```

```
[27 rows x 24 columns]
```

0.3.10 Define

Fix rating numerator that have decimals.

Code

```
In [92]: # View tweets with decimals in rating in 'text' column
         twitter_archive_clean[twitter_archive_clean.text.str.contains(r"(\d+\.\d*\/\d+)")]
```

```
Out[92]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         39    883482846933004288                    NaN                  NaN
         503   786709082849828864                    NaN                  NaN
         553   778027034220126208                    NaN                  NaN
         1374  680494726643068929                    NaN                  NaN

                              timestamp             source  \
         39    2017-07-08 00:28:19 +0000  Twitter for iPhone
         503   2016-10-13 23:23:56 +0000  Twitter for iPhone
         553   2016-09-20 00:24:34 +0000  Twitter for iPhone
         1374  2015-12-25 21:06:00 +0000  Twitter for iPhone


         39                               This is Bella. She hopes her smile made you smile. If r
         503                                      This is Logan, the Chow who lived. He solemnly
         553   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at rand
         1374                                                         Here we have uncovered an


         39    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter
         503                                                                   https://twitter
         553                                                                   https://twitter
         1374                                                                  https://twitter

              rating_numerator  rating_denominator    name   ...   img_num  \
         39                  5                  10   Bella   ...         1
         503                75                  10   Logan   ...         1
         553                27                  10  Sophie   ...         1
         1374               26                  10     NaN   ...         1

                             p1   p1_conf p1_dog                  p2   p2_conf  p2_dog  \
         39    golden_retriever  0.943082   True  Labrador_retriever  0.032409    True
         503         Pomeranian  0.467321   True          Persian_cat  0.122978   False
         553            clumber  0.946718   True      cocker_spaniel  0.015950    True
         1374            kuvasz  0.438627   True             Samoyed  0.111622    True

                            p3   p3_conf  p3_dog
         39            kuvasz  0.005501    True
         503             chow  0.102654    True
         553            Lhasa  0.006519    True
         1374  Great_Pyrenees  0.064061    True

         [4 rows x 24 columns]
```

```
In [93]: # Change datatype of rating_numerator and denominator to float
         twitter_archive_clean.rating_numerator = twitter_archive_clean.rating_numerator.astype
         twitter_archive_clean.rating_denominator = twitter_archive_clean.rating_denominator.as
```

```
In [94]: # update the correct values of rating_numerator for against tweet_id's found above
         fix_rating_decimal = {883482846933004288: 13.5,
                               786709082849828864: 9.75,
                               778027034220126208: 11.27,
                               680494726643068929: 11.26
                                 }


         for id, value in fix_rating_decimal.items():
             twitter_archive_clean.loc[twitter_archive_clean.tweet_id == id, 'rating_numerator
```

Test

```
In [95]: twitter_archive_clean[twitter_archive_clean.text.str.contains(r"(\d+\.\d*\/\d+)")]
```

```
Out[95]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         39    883482846933004288                    NaN                  NaN
         503   786709082849828864                    NaN                  NaN
         553   778027034220126208                    NaN                  NaN
         1374  680494726643068929                    NaN                  NaN

                               timestamp              source  \
         39    2017-07-08 00:28:19 +0000  Twitter for iPhone
         503   2016-10-13 23:23:56 +0000  Twitter for iPhone
         553   2016-09-20 00:24:34 +0000  Twitter for iPhone
         1374  2015-12-25 21:06:00 +0000  Twitter for iPhone



         39                             This is Bella. She hopes her smile made you smile. If 
         503                                    This is Logan, the Chow who lived. He solemnly
         553   This is Sophie. She's a Jubilant Bush Pupper. Super h*ckin rare. Appears at ranc
         1374                                                               Here we have uncovered a


         39    https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter
         503                                                                  https://twitter
         553                                                                  https://twitter
         1374                                                                 https://twitter
```

```
      rating_numerator  rating_denominator    name   ...   img_num  \
39               13.50                10.0   Bella   ...         1
503               9.75                10.0   Logan   ...         1
553              11.27                10.0  Sophie   ...         1
1374             11.26                10.0     NaN   ...         1

                    p1   p1_conf p1_dog                  p2   p2_conf  p2_dog  \
39    golden_retriever  0.943082   True  Labrador_retriever  0.032409    True
503         Pomeranian  0.467321   True         Persian_cat  0.122978   False
553            clumber  0.946718   True      cocker_spaniel  0.015950    True
1374            kuvasz  0.438627   True             Samoyed  0.111622    True

                  p3   p3_conf  p3_dog
39            kuvasz  0.005501    True
503             chow  0.102654    True
553            Lhasa  0.006519    True
1374  Great_Pyrenees  0.064061    True

[4 rows x 24 columns]
```

0.3.11 Define

Remove tweet with tweet_id = 810984652412424192 because text does not contain rating in this case.

Code

```
In [96]: twitter_archive_clean = twitter_archive_clean[twitter_archive_clean.tweet_id != 81098
```

Test

```
In [97]: twitter_archive_clean[twitter_archive_clean.tweet_id == 810984652412424192]
```

```
Out[97]: Empty DataFrame
         Columns: [tweet_id, in_reply_to_status_id, in_reply_to_user_id, timestamp, source, te
         Index: []

         [0 rows x 24 columns]
```

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

Code

```
In [98]: twitter_archive_clean.timestamp = pd.to_datetime(twitter_archive_clean.timestamp)
         twitter_archive_clean.dog_stage = twitter_archive_clean.dog_stage.astype('category')
         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_statu
         twitter_archive_clean.in_reply_to_user_id = twitter_archive_clean.in_reply_to_user_id
```

Test

```
In [99]: twitter_archive_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1993 entries, 0 to 1993
Data columns (total 24 columns):
tweet_id                 1993 non-null object
in_reply_to_status_id    1993 non-null object
in_reply_to_user_id      1993 non-null object
timestamp                1993 non-null datetime64[ns]
source                   1993 non-null object
text                     1993 non-null object
expanded_urls            1993 non-null object
rating_numerator         1993 non-null float64
rating_denominator       1993 non-null float64
name                     1371 non-null object
dog_stage                326 non-null category
retweet_count            1993 non-null int64
favorite_count           1993 non-null int64
jpg_url                  1993 non-null object
img_num                  1993 non-null int64
p1                       1993 non-null object
p1_conf                  1993 non-null float64
p1_dog                   1993 non-null bool
p2                       1993 non-null object
p2_conf                  1993 non-null float64
p2_dog                   1993 non-null bool
p3                       1993 non-null object
p3_conf                  1993 non-null float64
p3_dog                   1993 non-null bool
dtypes: bool(3), category(1), datetime64[ns](1), float64(5), int64(3), object(11)
memory usage: 334.9+ KB
```

0.4 Export

```
In [155]: # Save clean DataFrame to csv file
          twitter_archive_clean.to_csv('twitter_archive_master.csv', index=False)
```

0.5 Analyze

```
In [157]: import matplotlib
          df = pd.read_csv('twitter_archive_master.csv')
```

```
In [158]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1993 entries, 0 to 1992
```

```
Data columns (total 24 columns):
tweet_id                 1993 non-null int64
in_reply_to_status_id    23 non-null float64
in_reply_to_user_id      23 non-null float64
timestamp                1993 non-null object
source                   1993 non-null object
text                     1993 non-null object
expanded_urls            1993 non-null object
rating_numerator         1993 non-null float64
rating_denominator       1993 non-null float64
name                     1371 non-null object
dog_stage                326 non-null object
retweet_count            1993 non-null int64
favorite_count           1993 non-null int64
jpg_url                  1993 non-null object
img_num                  1993 non-null int64
p1                       1993 non-null object
p1_conf                  1993 non-null float64
p1_dog                   1993 non-null bool
p2                       1993 non-null object
p2_conf                  1993 non-null float64
p2_dog                   1993 non-null bool
p3                       1993 non-null object
p3_conf                  1993 non-null float64
p3_dog                   1993 non-null bool
dtypes: bool(3), float64(7), int64(4), object(10)
memory usage: 332.9+ KB
```

```
In [159]: df.head(2)
```

```
Out[159]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
          0  892420643555336193                    NaN                  NaN
          1  892177421306343426                    NaN                  NaN

                       timestamp              source  \
          0  2017-08-01 16:23:56  Twitter for iPhone
          1  2017-08-01 00:17:27  Twitter for iPhone


          0                                                     This is Phineas. He's a myst
          1  This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not, she

                                                              expanded_urls  \
          0  https://twitter.com/dog_rates/status/892420643555336193/photo/1
          1  https://twitter.com/dog_rates/status/892177421306343426/photo/1

             rating_numerator  rating_denominator     name   ...   img_num         p1  \
          0              13.0                10.0  Phineas   ...         1     orange
          1              13.0                10.0    Tilly   ...         1  Chihuahua

              p1_conf p1_dog        p2   p2_conf  p2_dog        p3   p3_conf   p3_dog
          0  0.097049  False     bagel  0.085851   False    banana  0.076110    False
          1  0.323581   True  Pekinese  0.090647    True  papillon  0.068957     True

          [2 rows x 24 columns]
```

0.5.1 Define

- Who has the most favorited dog?
- What does their picture look like?

```
In [188]: pd.set_option('display.max_columns', None)
```

```
In [189]: df[df["favorite_count"]== 143024]
```

Out[189]:

```
               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
309  822872901745569793                    NaN                  NaN

               timestamp              source  \
309  2017-01-21 18:26:02  Twitter for iPhone


309  Here's a super supportive puppo participating in the Toronto  #WomensMarch toda

                                                       expanded_urls  \
309  https://twitter.com/dog_rates/status/822872901745569793/photo/1

     rating_numerator  rating_denominator name dog_stage  retweet_count  \
309              13.0                10.0  NaN     puppo          48971

     favorite_count                                          jpg_url  img_num  \
309          143024  https://pbs.twimg.com/media/C2tugXLXgAArJ04.jpg        1

                   p1  p1_conf  p1_dog                  p2  p2_conf  p2_dog  \
309  Lakeland_terrier  0.196015    True  Labrador_retriever  0.160329    True

                p3  p3_conf  p3_dog
309  Irish_terrier  0.069126    True
```

```
In [215]: #Let's pull his picture the dataset
          img_url = str(df[df['tweet_id']==822872901745569793].jpg_url).split()[1]
          print(img_url)
          Image(img_url,width=300, height=300)
```

```
https://pbs.twimg.com/media/C2tugXLXgAArJ04.jpg
```

Out[215]:

0.5.2 Define

- What are the top 5 most popular dog names?

```
In [213]: from collections import Counter

          common_5_names = df[df.name.notnull()].name
          count = Counter(common_5_names)
          count.most_common(5)
```

```
Out[213]: [('Charlie', 11), ('Oliver', 10), ('Cooper', 10), ('Lucy', 10), ('Penny', 9)]
```

- Charlie, Oliver, Cooper, Lucy and Penny are the five most common name

0.6 Descriptive Statistical Analysis

```
In [160]: # Descriptive statistics
          stats= df.drop('tweet_id', axis=1)
          stats.describe()
```

```
Out[160]:        in_reply_to_status_id  in_reply_to_user_id  rating_numerator  \
          count           2.300000e+01         2.300000e+01       1993.000000
          mean            6.978112e+17         4.196984e+09         12.206613
          std             4.359384e+16         0.000000e+00         41.473096
          min             6.671522e+17         4.196984e+09          0.000000
          25%             6.732411e+17         4.196984e+09         10.000000
          50%             6.757073e+17         4.196984e+09         11.000000
          75%             7.031489e+17         4.196984e+09         12.000000
          max             8.558181e+17         4.196984e+09       1776.000000

                 rating_denominator  retweet_count  favorite_count      img_num  \
          count         1993.000000    1993.000000     1993.000000  1993.000000
          mean            10.511791    2708.934772     8827.983944     1.203211
          std              7.262919    4677.697123    12537.586518     0.560899
          min             10.000000      13.000000       80.000000     1.000000
          25%             10.000000     606.000000     1913.000000     1.000000
          50%             10.000000    1304.000000     4032.000000     1.000000
          75%             10.000000    3119.000000    11113.000000     1.000000
          max            170.000000   77143.000000   143024.000000     4.000000

                     p1_conf       p2_conf       p3_conf
          count  1993.000000  1.993000e+03  1.993000e+03
          mean      0.593802  1.344685e-01  6.026575e-02
          std       0.271951  1.006821e-01  5.089760e-02
          min       0.044333  1.011300e-08  1.740170e-10
          25%       0.362835  5.405530e-02  1.619070e-02
          50%       0.587507  1.175080e-01  4.952370e-02
          75%       0.845256  1.952180e-01  9.160200e-02
          max       1.000000  4.880140e-01  2.734190e-01
```

Key points:

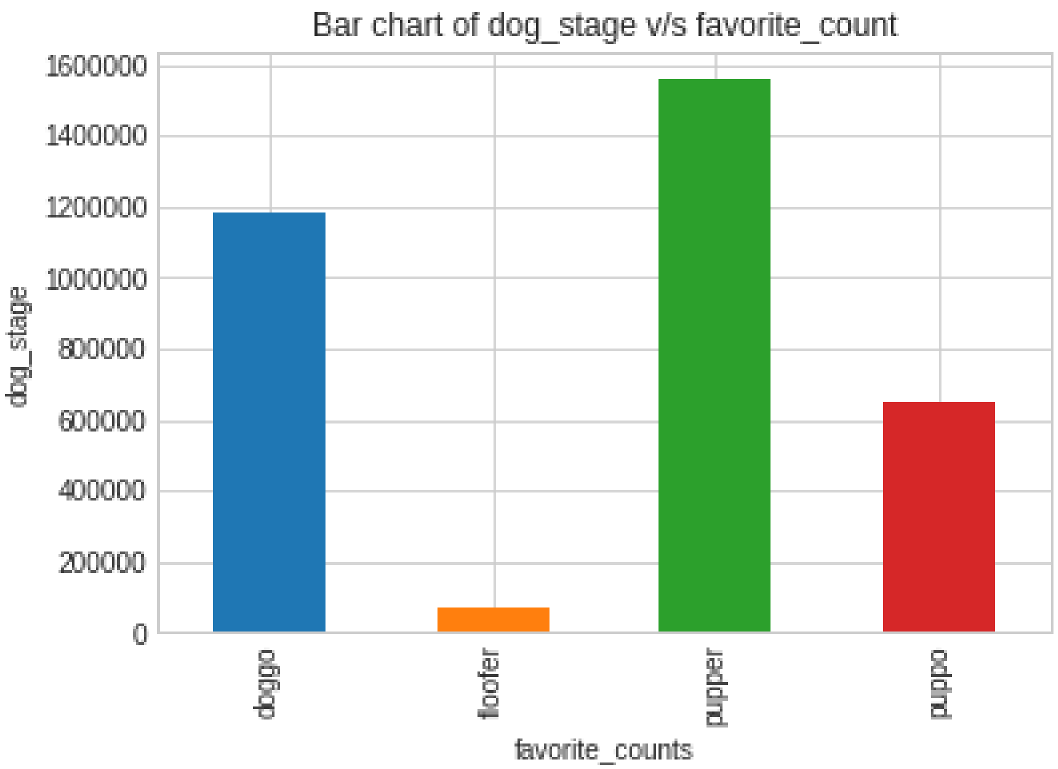
- The neural network performed the best on the 1st iteration with a mean prediciton of 0.59
- Mean rating for a dog is 12.207/10 with an outlier of 1776/10
- Mean retweet count for an original tweet was 2708 and a maximum value of 77143.
- Mean favorite count for an original tweet was 8827 and a maximum value of 143024.

0.6.1 dog_stage analysis

- Which dog_stage has got most favorite counts ?

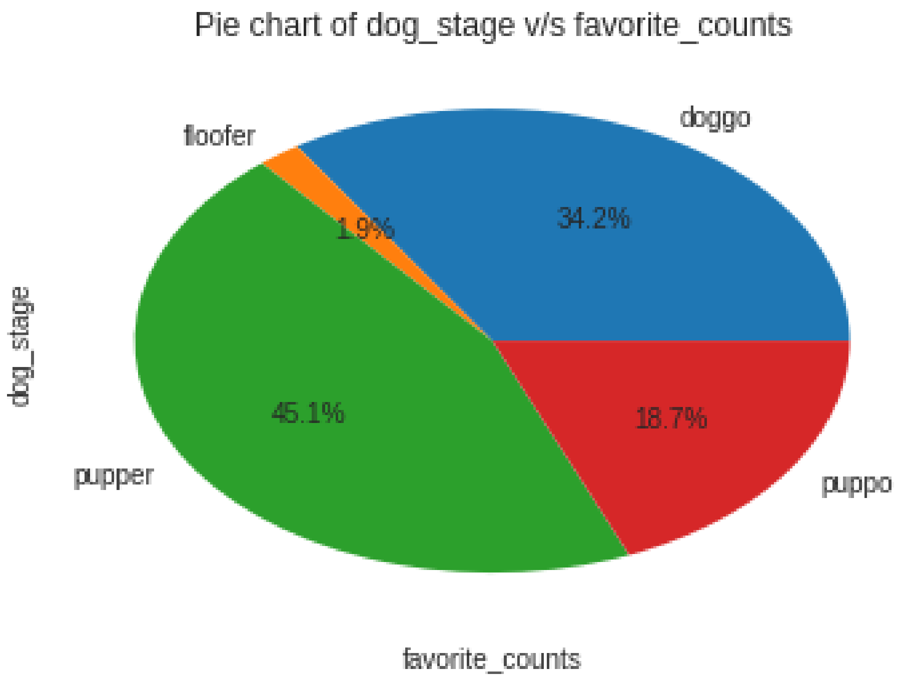
```
In [143]: top_dog_stage = df.groupby('dog_stage')['favorite_count'].sum()
          top_dog_stage.plot.bar()
          plt.title('Bar chart of dog_stage v/s favorite_count')
          plt.xlabel('favorite_counts')
          plt.ylabel('dog_stage')

Out[143]: Text(0,0.5,'dog_stage')
```



```
In [144]: top_dog_stage.plot(kind = 'pie', autopct='%1.1f%%')
          plt.title('Pie chart of dog_stage v/s favorite_counts')
          plt.xlabel('favorite_counts')
          plt.ylabel('dog_stage')

Out[144]: Text(0,0.5,'dog_stage')
```



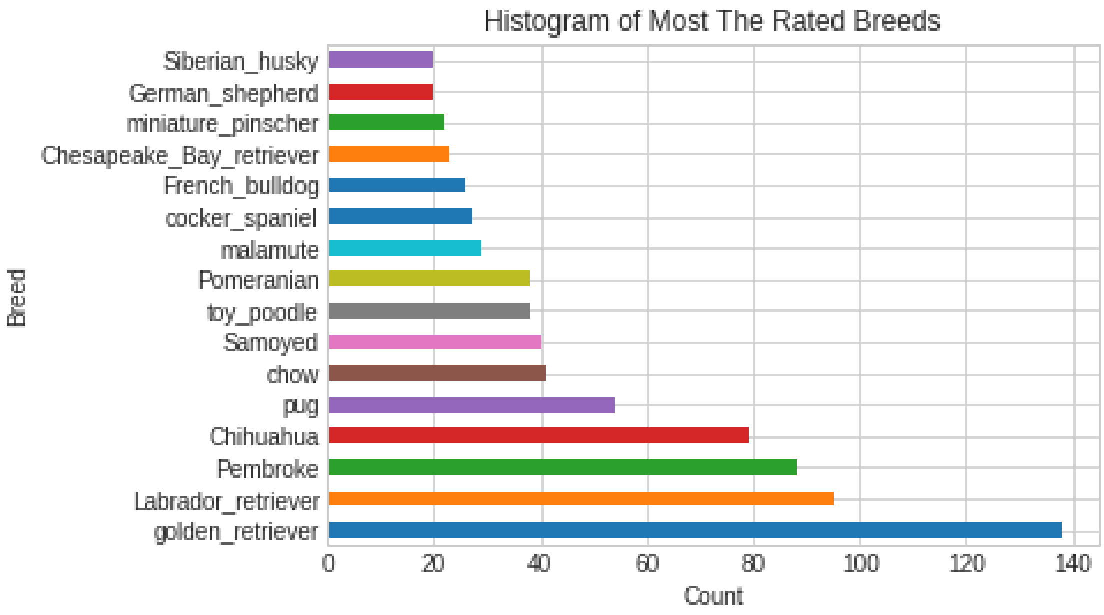
dog_stage pupper has the most favorite counts.

0.6.2 Most liked Breed

- What breed is having most favorite counts ?

```
In [214]: # Most liked breed
          top_breeds=df[df.p1_dog == True].groupby('p1').filter(lambda x: len(x) >= 20)
          top_breeds.p1.value_counts().plot(kind = 'barh')
          plt.title('Histogram of Most The Rated Breeds')
          plt.xlabel('Count')
          plt.ylabel('Breed')

Out[214]: Text(0,0.5,'Breed')
```



Golden_retriever is the most rated breed.

0.7 Conclusion

- The neural network performed the best on the 1st iteration with a mean prediciton of 0.59
- Mean rating for a dog is 12.207/10 with an outlier of 1776/10
- Mean retweet count for an original tweet was 2708 and a maximum value of 77143.
- Mean favorite count for an original tweet was 8827 and a maximum value of 143024.
- Most favorite dog tweet_id = 822872901745569793 with maximum value of favorite counts.
- Charlie, Oliver, Cooper, Lucy and Penny are the five most common name
- dog_stage pupper has the most favorite counts.
- Golden_retriever is the most rated breed.