wrangle_act

March 19, 2019

1 Wrangle and Analyze Data Project

1.1 Introduction

The purpose of this project is to wrangle data about WeRateDogs Twitter account. First, data is going to be obtained from multiple sources. Next, the data is going to be assessed in terms of both quality and tidiness (https://ryanwingate.com/purpose/tidy-data/). Finally, I will clean the data and perform simple analysis.

1.2 Gathering Data

```
In [1]: # import necessary libraries
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        % matplotlib inline
        import requests
        import os
```

```
In [2]: # load Twitter archive data and display 1 row
        df_archive_original = pd.read_csv('twitter-archive-enhanced.csv')
        df_archive_original.head(1)
```

```
Out[2]:          tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0  892420643555336193                    NaN                  NaN

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

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

                                                      text  retweeted_status_id  \
        0  This is Phineas. He's a mystical boy. Only eve...                  NaN

           retweeted_status_user_id retweeted_status_timestamp  \
        0                       NaN                        NaN
```

```
                                                expanded_urls  rating_numerator  \
        0  https://twitter.com/dog_rates/status/892420643...                13

           rating_denominator     name doggo floofer pupper puppo
        0                  10  Phineas  None    None   None  None
```

In [3]: *# Download the tweet image predictions*
        url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictio
        local_filename = url.split('/')[-1]
        r = requests.get(url)
        f = open(local_filename, 'wb')
        f.write(r.content)

Out[3]: 335079

In [4]: *# open and look at the predictions*
        df_pred_original = pd.read_csv('image-predictions.tsv', sep='\t')
        df_pred_original.head()

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

           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 [5]: *# Check for tweet_json.txt*
        # if tweet_json.txt doesn't exist, download programatically
        url = 'https://s3.amazonaws.com/video.udacity-data.com/topher/2018/November/5be5fb7d_twe
        json_txt_file_path = 'tweet_json.txt'
        if not os.path.isfile(json_txt_file_path):
            with open(json_txt_file_path, mode = 'wb') as file:
                file.write(requests.get(url).content)
        # Load into DataFrame
        df_json_original = pd.read_json(json_txt_file_path, lines=True)

```
         df_json_original = df_json_original[['id', 'retweet_count', 'favorite_count']]
         df_json_original.head()

Out[5]:                       id  retweet_count  favorite_count
         0  892420643555336193           8853           39467
         1  892177421306343426           6514           33819
         2  891815181378084864           4328           25461
         3  891689557279858688           8964           42908
         4  891327558926688256           9774           41048
```

Make copies of the dataframes for the analysis.

```
In [6]: df_archive = df_archive_original.copy()
        df_pred = df_pred_original.copy()
        df_json = df_json_original.copy()
```

1.3 Assessing Data

```
In [7]: df_archive.head(1)

Out[7]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0  892420643555336193                    NaN                  NaN

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

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

                                                      text  retweeted_status_id  \
        0  This is Phineas. He's a mystical boy. Only eve...                  NaN

           retweeted_status_user_id retweeted_status_timestamp  \
        0                       NaN                        NaN

                                      expanded_urls  rating_numerator  \
        0  https://twitter.com/dog_rates/status/892420643...                13

           rating_denominator     name doggo floofer pupper puppo
        0                  10  Phineas  None    None   None  None
```

```
In [8]: df_pred.head(5)

Out[8]:              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 [9]: df_json.head(1)

Out[9]:   id retweet_count favorite_count
       0   892420643555336193   8853   39467

In [10]: df_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 [11]: df_pred.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
```

```
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB


In [12]: df_json.info()

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


In [13]: df_archive.rating_denominator.value_counts()

Out[13]: 10      2333
         11         3
         50         3
         80         2
         20         2
         2          1
         16         1
         40         1
         70         1
         15         1
         90         1
         110        1
         120        1
         130        1
         150        1
         170        1
         7          1
```

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

In [14]: df_archive.name.value_counts()

Out[14]: None            745
         a               55
         Charlie         12
         Cooper          11
         Oliver          11
         Lucy            11
         Tucker          10
         Lola            10
         Penny           10
         Winston          9
         Bo               9
         the              8
         Sadie            8
         an               7
         Daisy            7
         Toby             7
         Bailey           7
         Buddy            7
         Scout            6
         Leo              6
         Jax              6
         Koda             6
         Stanley          6
         Milo             6
         Dave             6
         Bella            6
         Oscar            6
         Jack             6
         Rusty            6
         Sammy            5
                        ...
         Meera            1
         Kirk             1
         Sailor           1
         Zoe              1
         Gabby            1
         Millie           1
         Cleopatricia     1
         Kota             1
         Mojo             1
         Damon            1
         Barclay          1
         Hamrick          1
```

```
Stu               1
Chubbs            1
Ralphé            1
Frönq             1
Jeb               1
Arnold            1
Tycho             1
Edmund            1
Hector            1
Rhino             1
Ole               1
Duddles           1
Strider           1
my                1
Brockly           1
Mark              1
Beckham           1
Dook              1
Name: name, Length: 957, dtype: int64
```

1.3.1 Quality

1. Change the **id** column in `df_json` dataframe to **tweet_id** in order to match it with other dataframes. This will make is easier to merge dataframes at the last steps. In addition, I will change the data type from int to string because ids should not be numeric and they aren't intended to perform calculations.
2. Change **timestamp** data type in `df_archive`. This step will help if I was to analyze the data based on time or date.
3. Drop **retweets** rows in `df_archive`. As per project instructions, only original tweets must be included in the final master dataset.
4. Drop unnecessary columns in `df_archive`. It is not properly an issue, but it will make it easier to analyze the data. That is why I have 9 bullet points in this section.
5. Drop null values from **expanded_urls** in `df_archive`. The **expanded_urls** is the only column that has a few missing values. I could not determine why that was the case nor I could fill these values out; therefore, I will drop these values.
6. Drop tweets that are not dogs. This analysis should only incorporate tweets of dogs; therefore, the next step is to delete rows that contain non-dog tweets. In order to do that, I will merge the df_pred dataframe with df_archive and create a new dataframe df_archive2. The df_pred contains 3 predictions from the neural network whether a picture shows a dog. I assume that if any of the predictions suggest that it is a dog, I treated it as a dog. I checked a few tweets 'by hand' and it appeared to be true.
7. Clean **rating_denominator** in `df_archive`. Extract the values from text, delete tweets without ratings and tweets with multiple dogs. Change the data type to float.
8. Clean **rating_nominator** in `df_archive`. Extract the values from text, delete tweets without ratings and tweets with multiple dogs. Change the data type to float.
9. Clean names using regular expression in `df_archive`. A useful pattern to find these "fake names" is that they started with lowercase, instead, the "real names" always started with uppercase, moreover, the real ones usually are placed after words like "named", "is" etc...so

7

it'is possible to clean this issue using a regex.

1.3.2 Tidiness

1. Melt 'doggo' 'floofer' 'pupper' 'puppo' in `df_archive`.
2. Join all data frames into `twitter_archive_master.csv`.

1.4 Cleaning Data

1.4.1 Quality

Define 1. Change the **id** column in `df_json` dataframe to **tweet_id** in order to match it with other dataframes. This will make is easier to merge dataframes at the last steps. In addition, I will change the data type from int to string because ids should not be numeric and they aren't intended to perform calculations.
   Code

```
In [15]: # rename to `tweet_id`
         df_json = df_json.rename(columns={'id': 'tweet_id'})

In [16]: # change `tweet_id` datatype to string in all 3 data frames
         df_archive['tweet_id']= df_archive['tweet_id'].astype(str)
         df_pred['tweet_id']= df_pred['tweet_id'].astype(str)
         df_json['tweet_id']= df_json['tweet_id'].astype(str)
```

   Test

```
In [17]: df_json.head(1)

Out[17]:            tweet_id  retweet_count  favorite_count
         0  892420643555336193           8853           39467

In [18]: df_json.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 3 columns):
tweet_id         2354 non-null object
retweet_count    2354 non-null int64
favorite_count   2354 non-null int64
dtypes: int64(2), object(1)
memory usage: 55.2+ KB
```

Define 2. Change **timestamp** data type in `df_archive`. This step will help if I was to analyze the data based on time or date.
   Code

```
In [19]: # change data type to datetime
         df_archive['timestamp'] = pd.to_datetime(df_archive.timestamp)
```

8

Test

```
In [20]: df_archive.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2356 non-null object
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2356 non-null datetime64[ns]
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: datetime64[ns](1), float64(4), int64(2), object(10)
memory usage: 313.0+ KB
```

Define 3. Drop **retweets** rows in df_archive. As per project instructions, only original tweets must be included in the final master dataset.
Code

```
In [21]: # check how many retweets are in the dataframe
         df_archive.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2356 non-null object
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2356 non-null datetime64[ns]
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
```

9

```
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: datetime64[ns](1), float64(4), int64(2), object(10)
memory usage: 313.0+ KB
```

There are 181 retweets in the dataframe. Next, I will delete the retweets.

In [22]: *# select retweets by checking the retweets columns (if they are empty it's an original*
         df_archive_notnull = df_archive[df_archive.retweeted_status_id.notnull()]
         df_archive_notnull

Out[22]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         19   888202515573088257   NaN   NaN
         32   886054160059072513   NaN   NaN
         36   885311592912609280   NaN   NaN
         68   879130579576475649   NaN   NaN
         73   878404777348136964   NaN   NaN
         74   878316110768087041   NaN   NaN
         78   877611172832227328   NaN   NaN
         91   874434818259525634   NaN   NaN
         95   873697596434513921   NaN   NaN
         97   873337748698140672   NaN   NaN
         101   872668790621863937   NaN   NaN
         109   871166179821445120   NaN   NaN
         118   869988702071779329   NaN   NaN
         124   868639477480148993   NaN   NaN
         130   867072653475098625   NaN   NaN
         132   866816280283807744   NaN   NaN
         137   866094527597207552   NaN   NaN
         146   863471782782697472   NaN   NaN
         155   861769973181624320   NaN   NaN
         159   860981674716409858   NaN   NaN
         160   860924035999428608   NaN   NaN
         165   860177593139703809   NaN   NaN
         171   858860390427611136   NaN   NaN
         180   857062103051644929   NaN   NaN
         182   856602993587888130   NaN   NaN
         185   856330835276025856   NaN   NaN
         194   855245323840757760   NaN   NaN
         195   855138241867124737   NaN   NaN
         204   852936405516943360   NaN   NaN
         211   851953902622658560   NaN   NaN
         ...   .
```

```
784    775096608509886464                              NaN               NaN
794    773336787167145985                              NaN               NaN
800    772615324260794368                              NaN               NaN
811    771171053431250945                              NaN               NaN
815    771004394259247104                              NaN               NaN
818    770743923962707968                              NaN               NaN
822    770093767776997377                              NaN               NaN
826    769335591808995329                              NaN               NaN
829    768909767477751808                              NaN               NaN
833    768554158521745409                              NaN               NaN
841    766864461642756096                              NaN               NaN
847    766078092750233600                              NaN               NaN
860    763167063695355904                              NaN               NaN
868    761750502866649088                              NaN               NaN
872    761371037149827077                              NaN               NaN
885    760153949710192640                              NaN               NaN
890    759566828574212096                              NaN               NaN
895    759159934323924993                              NaN               NaN
908    757729163776290825                              NaN               NaN
911    757597904299253760                              NaN               NaN
926    754874841593970688                              NaN               NaN
937    753298634498793472                              NaN               NaN
943    752701944171524096                              NaN               NaN
949    752309394570878976                              NaN               NaN
1012   747242308580548608                              NaN               NaN
1023   746521445350707200                              NaN               NaN
1043   743835915802583040                              NaN               NaN
1242   711998809858043904                              NaN               NaN
2259   667550904950915073                              NaN               NaN
2260   667550882905632768                              NaN               NaN

                 timestamp                                          source  \
19     2017-07-21 01:02:36  <a href="http://twitter.com/download/iphone" r...
32     2017-07-15 02:45:48  <a href="http://twitter.com/download/iphone" r...
36     2017-07-13 01:35:06  <a href="http://twitter.com/download/iphone" r...
68     2017-06-26 00:13:58  <a href="http://twitter.com/download/iphone" r...
73     2017-06-24 00:09:53  <a href="http://twitter.com/download/iphone" r...
74     2017-06-23 18:17:33  <a href="http://twitter.com/download/iphone" r...
78     2017-06-21 19:36:23  <a href="http://twitter.com/download/iphone" r...
91     2017-06-13 01:14:41  <a href="http://twitter.com/download/iphone" r...
95     2017-06-11 00:25:14  <a href="http://twitter.com/download/iphone" r...
97     2017-06-10 00:35:19  <a href="http://twitter.com/download/iphone" r...
101    2017-06-08 04:17:07  <a href="http://twitter.com/download/iphone" r...
109    2017-06-04 00:46:17  <a href="http://twitter.com/download/iphone" r...
118    2017-05-31 18:47:24  <a href="http://twitter.com/download/iphone" r...
124    2017-05-28 01:26:04  <a href="http://twitter.com/download/iphone" r...
130    2017-05-23 17:40:04  <a href="http://twitter.com/download/iphone" r...
132    2017-05-23 00:41:20  <a href="http://twitter.com/download/iphone" r...
```

```
137  2017-05-21 00:53:21  <a href="http://twitter.com/download/iphone" r...
146  2017-05-13 19:11:30  <a href="http://twitter.com/download/iphone" r...
155  2017-05-09 02:29:07  <a href="http://twitter.com/download/iphone" r...
159  2017-05-06 22:16:42  <a href="http://twitter.com/download/iphone" r...
160  2017-05-06 18:27:40  <a href="http://twitter.com/download/iphone" r...
165  2017-05-04 17:01:34  <a href="http://twitter.com/download/iphone" r...
171  2017-05-01 01:47:28  <a href="http://twitter.com/download/iphone" r...
180  2017-04-26 02:41:43  <a href="http://twitter.com/download/iphone" r...
182  2017-04-24 20:17:23  <a href="http://twitter.com/download/iphone" r...
185  2017-04-24 02:15:55  <a href="http://twitter.com/download/iphone" r...
194  2017-04-21 02:22:29  <a href="http://twitter.com/download/iphone" r...
195  2017-04-20 19:16:59  <a href="http://twitter.com/download/iphone" r...
204  2017-04-14 17:27:40  <a href="http://twitter.com/download/iphone" r...
211  2017-04-12 00:23:33  <a href="http://twitter.com/download/iphone" r...
...                  ...                                                  ...
784  2016-09-11 22:20:06  <a href="http://twitter.com/download/iphone" r...
794  2016-09-07 01:47:12  <a href="http://twitter.com/download/iphone" r...
800  2016-09-05 02:00:22  <a href="http://twitter.com/download/iphone" r...
811  2016-09-01 02:21:21  <a href="http://twitter.com/download/iphone" r...
815  2016-08-31 15:19:06  <a href="http://twitter.com/download/iphone" r...
818  2016-08-30 22:04:05  <a href="http://twitter.com/download/iphone" r...
822  2016-08-29 03:00:36  <a href="http://twitter.com/download/iphone" r...
826  2016-08-27 00:47:53  <a href="http://twitter.com/download/iphone" r...
829  2016-08-25 20:35:48  <a href="http://twitter.com/download/iphone" r...
833  2016-08-24 21:02:45  <a href="http://twitter.com/download/iphone" r...
841  2016-08-20 05:08:29  <a href="http://twitter.com/download/iphone" r...
847  2016-08-18 01:03:45  <a href="http://twitter.com/download/iphone" r...
860  2016-08-10 00:16:21  <a href="http://twitter.com/download/iphone" r...
868  2016-08-06 02:27:27  <a href="http://twitter.com/download/iphone" r...
872  2016-08-05 01:19:35  <a href="http://twitter.com/download/iphone" r...
885  2016-08-01 16:43:19  <a href="http://twitter.com/download/iphone" r...
890  2016-07-31 01:50:18  <a href="http://twitter.com/download/iphone" r...
895  2016-07-29 22:53:27  <a href="http://twitter.com/download/iphone" r...
908  2016-07-26 00:08:05  <a href="http://twitter.com/download/iphone" r...
911  2016-07-25 15:26:30  <a href="http://twitter.com/download/iphone" r...
926  2016-07-18 03:06:01  <a href="http://twitter.com/download/iphone" r...
937  2016-07-13 18:42:44  <a href="http://twitter.com/download/iphone" r...
943  2016-07-12 03:11:42  <a href="http://twitter.com/download/iphone" r...
949  2016-07-11 01:11:51  <a href="http://twitter.com/download/iphone" r...
1012 2016-06-27 01:37:04  <a href="http://twitter.com/download/iphone" r...
1023 2016-06-25 01:52:36  <a href="http://twitter.com/download/iphone" r...
1043 2016-06-17 16:01:16  <a href="http://twitter.com/download/iphone" r...
1242 2016-03-21 19:31:59  <a href="http://twitter.com/download/iphone" r...
2259 2015-11-20 03:51:52  <a href="http://twitter.com" rel="nofollow">Tw...
2260 2015-11-20 03:51:47  <a href="http://twitter.com" rel="nofollow">Tw...


                                                   text  retweeted_status_id  \
19      RT @dog_rates: This is Canela. She attempted s...         8.874740e+17
```

```
32      RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...        8.860537e+17
36      RT @dog_rates: This is Lilly. She just paralle...       8.305833e+17
68      RT @dog_rates: This is Emmy. She was adopted t...       8.780576e+17
73      RT @dog_rates: Meet Shadow. In an attempt to r...       8.782815e+17
74      RT @dog_rates: Meet Terrance. He's being yelle...       6.690004e+17
78      RT @rachel2195: @dog_rates the boyfriend and h...       8.768508e+17
91      RT @dog_rates: This is Coco. At first I though...       8.663350e+17
95      RT @dog_rates: This is Walter. He won't start ...       8.688804e+17
97      RT @dog_rates: This is Sierra. She's one preci...       8.732138e+17
101     RT @loganamnosis: Penelope here is doing me qu...       8.726576e+17
109     RT @dog_rates: This is Dawn. She's just checki...       8.410770e+17
118     RT @dog_rates: We only rate dogs. This is quit...       8.591970e+17
124     RT @dog_rates: Say hello to Cooper. His expres...       8.685523e+17
130     RT @rachaeleasler: these @dog_rates hats are 1...       8.650134e+17
132     RT @dog_rates: This is Jamesy. He gives a kiss...       8.664507e+17
137     RT @dog_rates: Here's a pupper before and afte...       8.378202e+17
146     RT @dog_rates: Say hello to Quinn. She's quite...       8.630625e+17
155     RT @dog_rates: "Good afternoon class today we'...       8.066291e+17
159     RT @dog_rates: Meet Lorenzo. He's an avid nift...       8.605638e+17
160     RT @tallylott: h*ckin adorable promposal. 13/1...       8.609145e+17
165     RT @dog_rates: Ohboyohboyohboyohboyohboyohboyo...       7.616730e+17
171     RT @dog_rates: Meet Winston. He knows he's a l...       8.395493e+17
180     RT @AaronChewning: First time wearing my @dog_...       8.570611e+17
182     RT @dog_rates: This is Luna. It's her first ti...       8.447048e+17
185     RT @Jenna_Marbles: @dog_rates Thanks for ratin...       8.563302e+17
194     RT @dog_rates: Meet George. He looks slightly ...       8.421635e+17
195     RT @frasercampbell_: oh my... what's that... b...       8.551225e+17
204     RT @dog_rates: I usually only share these on F...       8.316501e+17
211     RT @dog_rates: This is Astrid. She's a guide d...       8.293743e+17
...                              ...                              ...
784     RT @dog_rates: After so many requests, this is...       7.403732e+17
794     RT @dog_rates: Meet Fizz. She thinks love is a...       7.713808e+17
800     RT @dog_rates: This is Gromit. He's pupset bec...       7.652221e+17
811     RT @dog_rates: This is Frankie. He's wearing b...       6.733201e+17
815     RT @katieornah: @dog_rates learning a lot at c...       7.710021e+17
818     RT @dog_rates: Here's a doggo blowing bubbles...       7.392382e+17
822     RT @dog_rates: This is just downright precious...       7.410673e+17
826     RT @dog_rates: Ever seen a dog pet another dog...       7.069045e+17
829     RT @dog_rates: When it's Janet from accounting...       7.001438e+17
833     RT @dog_rates: This is Nollie. She's waving at...       7.399792e+17
841     RT @dog_rates: We only rate dogs... this is a ...       7.599238e+17
847     RT @dog_rates: This is Colby. He's currently r...       7.258423e+17
860     RT @dog_rates: Meet Eve. She's a raging alcoho...       6.732953e+17
868     RT @dog_rates: "Tristan do not speak to me wit...       6.853251e+17
872     RT @dog_rates: Oh. My. God. 13/10 magical af h...       7.116948e+17
885     RT @hownottodraw: The story/person behind @dog...       7.601538e+17
890     RT @dog_rates: This... is a Tyrannosaurus rex...       7.395441e+17
895     RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...       6.703191e+17
```

```
908    RT @dog_rates: This is Chompsky. He lives up t...        6.790626e+17
911    RT @jon_hill987: @dog_rates There is a cunning...       7.575971e+17
926    RT @dog_rates: This is Rubio. He has too much ...       6.791584e+17
937    RT @dog_rates: This is Carly. She's actually 2...       6.815232e+17
943    RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...       6.835159e+17
949    RT @dog_rates: Everyone needs to watch this. 1...       6.753544e+17
1012   RT @dog_rates: This pupper killed this great w...       7.047611e+17
1023   RT @dog_rates: This is Shaggy. He knows exactl...       6.678667e+17
1043   RT @dog_rates: Extremely intelligent dog here...        6.671383e+17
1242   RT @twitter: @dog_rates Awesome Tweet! 12/10. ...       7.119983e+17
2259   RT @dogratingrating: Exceptional talent. Origi...       6.675487e+17
2260   RT @dogratingrating: Unoriginal idea. Blatant ...       6.675484e+17


       retweeted_status_user_id retweeted_status_timestamp  \
19               4.196984e+09   2017-07-19 00:47:34 +0000
32               1.960740e+07   2017-07-15 02:44:07 +0000
36               4.196984e+09   2017-02-12 01:04:29 +0000
68               4.196984e+09   2017-06-23 01:10:23 +0000
73               4.196984e+09   2017-06-23 16:00:04 +0000
74               4.196984e+09   2015-11-24 03:51:38 +0000
78               5.128045e+08   2017-06-19 17:14:49 +0000
91               4.196984e+09   2017-05-21 16:48:45 +0000
95               4.196984e+09   2017-05-28 17:23:24 +0000
97               4.196984e+09   2017-06-09 16:22:42 +0000
101              1.547674e+08   2017-06-08 03:32:35 +0000
109              4.196984e+09   2017-03-13 00:02:39 +0000
118              4.196984e+09   2017-05-02 00:04:57 +0000
124              4.196984e+09   2017-05-27 19:39:34 +0000
130              7.874618e+17   2017-05-18 01:17:25 +0000
132              4.196984e+09   2017-05-22 00:28:40 +0000
137              4.196984e+09   2017-03-04 00:21:08 +0000
146              4.196984e+09   2017-05-12 16:05:02 +0000
155              4.196984e+09   2016-12-07 22:38:52 +0000
159              4.196984e+09   2017-05-05 18:36:06 +0000
160              3.638908e+08   2017-05-06 17:49:42 +0000
165              4.196984e+09   2016-08-05 21:19:27 +0000
171              4.196984e+09   2017-03-08 18:52:12 +0000
180              5.870972e+07   2017-04-26 02:37:47 +0000
182              4.196984e+09   2017-03-23 00:18:10 +0000
185              6.669901e+07   2017-04-24 02:13:14 +0000
194              4.196984e+09   2017-03-16 00:00:07 +0000
195              7.475543e+17   2017-04-20 18:14:33 +0000
204              4.196984e+09   2017-02-14 23:43:18 +0000
211              4.196984e+09   2017-02-08 17:00:26 +0000
...                       ...                         ...
784              4.196984e+09   2016-06-08 02:41:38 +0000
794              4.196984e+09   2016-09-01 16:14:48 +0000
800              4.196984e+09   2016-08-15 16:22:20 +0000
```

```
811              4.196984e+09  2015-12-06 01:56:44 +0000
815              1.732729e+09  2016-08-31 15:10:07 +0000
818              4.196984e+09  2016-06-04 23:31:25 +0000
822              4.196984e+09  2016-06-10 00:39:48 +0000
826              4.196984e+09  2016-03-07 18:09:06 +0000
829              4.196984e+09  2016-02-18 02:24:13 +0000
833              4.196984e+09  2016-06-07 00:36:02 +0000
841              4.196984e+09  2016-08-01 01:28:46 +0000
847              4.196984e+09  2016-04-29 00:21:01 +0000
860              4.196984e+09  2015-12-06 00:17:55 +0000
868              4.196984e+09  2016-01-08 05:00:14 +0000
872              4.196984e+09  2016-03-20 23:23:54 +0000
885              1.950368e+08  2016-08-01 16:42:51 +0000
890              4.196984e+09  2016-06-05 19:47:03 +0000
895              4.196984e+09  2015-11-27 19:11:49 +0000
908              4.196984e+09  2015-12-21 22:15:18 +0000
911              2.804798e+08  2016-07-25 15:23:28 +0000
926              4.196984e+09  2015-12-22 04:35:49 +0000
937              4.196984e+09  2015-12-28 17:12:42 +0000
943              4.196984e+09  2016-01-03 05:11:12 +0000
949              4.196984e+09  2015-12-11 16:40:19 +0000
1012             4.196984e+09  2016-03-01 20:11:59 +0000
1023             4.196984e+09  2015-11-21 00:46:50 +0000
1043             4.196984e+09  2015-11-19 00:32:12 +0000
1242             7.832140e+05  2016-03-21 19:29:52 +0000
2259             4.296832e+09  2015-11-20 03:43:06 +0000
2260             4.296832e+09  2015-11-20 03:41:59 +0000


                                      expanded_urls  rating_numerator  \
19    https://twitter.com/dog_rates/status/887473957...                13
32    https://twitter.com/dog_rates/status/886053434...                12
36    https://twitter.com/dog_rates/status/830583320...                13
68    https://twitter.com/dog_rates/status/878057613...                14
73    https://www.gofundme.com/3yd6y1c,https://twitt...                13
74    https://twitter.com/dog_rates/status/669000397...                11
78    https://twitter.com/rachel2195/status/87685077...                14
91    https://twitter.com/dog_rates/status/866334964...                12
95    https://twitter.com/dog_rates/status/868880397...                14
97    https://www.gofundme.com/help-my-baby-sierra-g...                12
101   https://twitter.com/loganamnosis/status/872657...                14
109   https://twitter.com/dog_rates/status/841077006...                12
118   https://twitter.com/dog_rates/status/859196978...                12
124   https://www.gofundme.com/3ti3nps,https://twitt...                12
130   https://twitter.com/rachaeleasler/status/86501...                13
132   https://twitter.com/dog_rates/status/866450705...                13
137   https://twitter.com/dog_rates/status/837820167...                12
146   https://www.gofundme.com/helpquinny,https://tw...                13
155   https://twitter.com/dog_rates/status/806629075...                13
```

159	https://www.gofundme.com/help-lorenzo-beat-can...	13
160	https://twitter.com/tallylott/status/860914485...	13
165	https://twitter.com/dog_rates/status/761672994...	10
171	https://twitter.com/dog_rates/status/839549326...	12
180	https://twitter.com/AaronChewning/status/85706...	13
182	https://twitter.com/dog_rates/status/844704788...	13
185	NaN	14
194	https://twitter.com/dog_rates/status/842163532...	12
195	https://twitter.com/frasercampbell_/status/855...	14
204	http://www.gofundme.com/bluethewhitehusky,http...	13
211	https://twitter.com/dog_rates/status/829374341...	13
...	...	...
784	https://twitter.com/dog_rates/status/740373189...	9
794	https://twitter.com/dog_rates/status/771380798...	11
800	https://twitter.com/dog_rates/status/765222098...	10
811	https://twitter.com/dog_rates/status/673320132...	11
815	https://twitter.com/katieornah/status/77100213...	12
818	https://twitter.com/dog_rates/status/739238157...	13
822	https://twitter.com/dog_rates/status/741067306...	12
826	https://vine.co/v/iXQAm5Lrgrh,https://vine.co/...	13
829	https://twitter.com/dog_rates/status/700143752...	10
833	https://twitter.com/dog_rates/status/739979191...	12
841	https://twitter.com/dog_rates/status/759923798...	10
847	https://twitter.com/dog_rates/status/725842289...	12
860	https://twitter.com/dog_rates/status/673295268...	8
868	https://twitter.com/dog_rates/status/685325112...	10
872	https://twitter.com/dog_rates/status/711694788...	13
885	https://weratedogs.com/pages/about-us,https://...	11
890	https://twitter.com/dog_rates/status/739544079...	10
895	https://twitter.com/dog_rates/status/670319130...	11
908	https://twitter.com/dog_rates/status/679062614...	11
911	https://twitter.com/jon_hill987/status/7575971...	11
926	https://twitter.com/dog_rates/status/679158373...	11
937	https://twitter.com/dog_rates/status/681523177...	12
943	https://vine.co/v/ibvnzrauFuV,https://vine.co/...	11
949	https://twitter.com/dog_rates/status/675354435...	13
1012	https://twitter.com/dog_rates/status/704761120...	13
1023	https://twitter.com/dog_rates/status/667866724...	10
1043	https://twitter.com/dog_rates/status/667138269...	10
1242	https://twitter.com/twitter/status/71199827977...	12
2259	https://twitter.com/dogratingrating/status/667...	12
2260	https://twitter.com/dogratingrating/status/667...	5

	rating_denominator	name	doggo	floofer	pupper	puppo
19	10	Canela	None	None	None	None
32	10	None	None	None	None	None
36	10	Lilly	None	None	None	None
68	10	Emmy	None	None	None	None

73	10	Shadow	None	None	None	None
74	10	Terrance	None	None	None	None
78	10	None	None	None	pupper	None
91	10	Coco	None	None	None	None
95	10	Walter	None	None	None	None
97	10	Sierra	None	None	pupper	None
101	10	None	None	None	None	None
109	10	Dawn	None	None	None	None
118	10	quite	None	None	None	None
124	10	Cooper	None	None	None	None
130	10	None	None	None	None	None
132	10	Jamesy	None	None	pupper	None
137	10	None	None	None	pupper	None
146	10	Quinn	None	None	None	None
155	10	None	None	None	None	None
159	10	Lorenzo	None	None	None	None
160	10	None	None	None	None	None
165	10	None	None	None	None	None
171	10	Winston	None	None	None	None
180	10	None	None	None	None	None
182	10	Luna	None	None	None	None
185	10	None	None	None	None	None
194	10	George	None	None	None	None
195	10	None	None	None	None	None
204	10	None	None	None	None	None
211	10	Astrid	doggo	None	None	None
...	...	...	...	...	...	...
784	11	None	None	None	None	None
794	10	Fizz	None	None	None	None
800	10	Gromit	None	None	None	None
811	10	Frankie	None	None	None	None
815	10	None	None	None	pupper	None
818	10	None	doggo	None	None	None
822	10	just	doggo	None	pupper	None
826	10	None	None	None	None	None
829	10	None	None	None	pupper	None
833	10	Nollie	None	None	None	None
841	10	None	None	None	None	None
847	10	Colby	None	None	None	None
860	10	Eve	None	None	pupper	None
868	10	None	None	None	None	None
872	10	None	None	None	None	None
885	10	None	None	None	None	None
890	10	None	None	None	None	None
895	10	None	None	None	None	None
908	10	Chompsky	None	None	None	None
911	10	None	None	None	pupper	None
926	10	Rubio	None	None	None	None

```
937                    10    Carly    None    None    None   None
943                    10     None    None    None    None   None
949                    10     None    None    None    None   None
1012                   10     None    None    None  pupper   None
1023                   10   Shaggy    None    None    None   None
1043                   10     None    None    None    None   None
1242                   10     None    None    None    None   None
2259                   10     None    None    None    None   None
2260                   10     None    None    None    None   None

[181 rows x 17 columns]
```

In [23]: # drop retweets (df_archive.retweeted_status rows that are not empty)
         df_archive = df_archive.drop(df_archive_notnull.index, axis=0)
         df_archive

Out[23]:   tweet_id in_reply_to_status_id in_reply_to_user_id \
         0   892420643555336193   NaN   NaN
         1   892177421306343426   NaN   NaN
         2   891815181378084864   NaN   NaN
         3   891689557279858688   NaN   NaN
         4   891327558926688256   NaN   NaN
         5   891087950875897856   NaN   NaN
         6   890971913173991426   NaN   NaN
         7   890729181411237888   NaN   NaN
         8   890609185150312448   NaN   NaN
         9   890240255349198849   NaN   NaN
         10   890006608113172480   NaN   NaN
         11   889880896479866881   NaN   NaN
         12   889665388333682689   NaN   NaN
         13   889638837579907072   NaN   NaN
         14   889531135344209921   NaN   NaN
         15   889278841981685760   NaN   NaN
         16   888917238123831296   NaN   NaN
         17   888804989199671297   NaN   NaN
         18   888554962724278272   NaN   NaN
         20   888078434458587136   NaN   NaN
         21   887705289381826560   NaN   NaN
         22   887517139158093824   NaN   NaN
         23   887473957103951883   NaN   NaN
         24   887343217045368832   NaN   NaN
         25   887101392804085760   NaN   NaN
         26   886983233522544640   NaN   NaN
         27   886736880519319552   NaN   NaN
         28   886680336477933568   NaN   NaN
         29   886366144734445568   NaN   NaN
         30   886267009285017600   8.862664e+17   2.281182e+09
         ...   .
```

```
2326   666411507551481857                           NaN                    NaN
2327   666407126856765440                           NaN                    NaN
2328   666396247373291520                           NaN                    NaN
2329   666373753744588802                           NaN                    NaN
2330   666362758909284353                           NaN                    NaN
2331   666353288456101888                           NaN                    NaN
2332   666345417576210432                           NaN                    NaN
2333   666337882303524864                           NaN                    NaN
2334   666293911632134144                           NaN                    NaN
2335   666287406224695296                           NaN                    NaN
2336   666273097616637952                           NaN                    NaN
2337   666268910803644416                           NaN                    NaN
2338   666104133288665088                           NaN                    NaN
2339   666102155909144576                           NaN                    NaN
2340   666099513787052032                           NaN                    NaN
2341   666094000022159362                           NaN                    NaN
2342   666082916733198337                           NaN                    NaN
2343   666073100786774016                           NaN                    NaN
2344   666071193221509120                           NaN                    NaN
2345   666063827256086533                           NaN                    NaN
2346   666058600524156928                           NaN                    NaN
2347   666057090499244032                           NaN                    NaN
2348   666055525042405380                           NaN                    NaN
2349   666051853826850816                           NaN                    NaN
2350   666050758794694657                           NaN                    NaN
2351   666049248165822465                           NaN                    NaN
2352   666044226329800704                           NaN                    NaN
2353   666033412701032449                           NaN                    NaN
2354   666029285002620928                           NaN                    NaN
2355   666020888022790149                           NaN                    NaN


                 timestamp                                             source  \
0      2017-08-01 16:23:56   <a href="http://twitter.com/download/iphone" r...
1      2017-08-01 00:17:27   <a href="http://twitter.com/download/iphone" r...
2      2017-07-31 00:18:03   <a href="http://twitter.com/download/iphone" r...
3      2017-07-30 15:58:51   <a href="http://twitter.com/download/iphone" r...
4      2017-07-29 16:00:24   <a href="http://twitter.com/download/iphone" r...
5      2017-07-29 00:08:17   <a href="http://twitter.com/download/iphone" r...
6      2017-07-28 16:27:12   <a href="http://twitter.com/download/iphone" r...
7      2017-07-28 00:22:40   <a href="http://twitter.com/download/iphone" r...
8      2017-07-27 16:25:51   <a href="http://twitter.com/download/iphone" r...
9      2017-07-26 15:59:51   <a href="http://twitter.com/download/iphone" r...
10     2017-07-26 00:31:25   <a href="http://twitter.com/download/iphone" r...
11     2017-07-25 16:11:53   <a href="http://twitter.com/download/iphone" r...
12     2017-07-25 01:55:32   <a href="http://twitter.com/download/iphone" r...
13     2017-07-25 00:10:02   <a href="http://twitter.com/download/iphone" r...
14     2017-07-24 17:02:04   <a href="http://twitter.com/download/iphone" r...
15     2017-07-24 00:19:32   <a href="http://twitter.com/download/iphone" r...
```

```
16   2017-07-23 00:22:39  <a href="http://twitter.com/download/iphone" r...
17   2017-07-22 16:56:37  <a href="http://twitter.com/download/iphone" r...
18   2017-07-22 00:23:06  <a href="http://twitter.com/download/iphone" r...
20   2017-07-20 16:49:33  <a href="http://twitter.com/download/iphone" r...
21   2017-07-19 16:06:48  <a href="http://twitter.com/download/iphone" r...
22   2017-07-19 03:39:09  <a href="http://twitter.com/download/iphone" r...
23   2017-07-19 00:47:34  <a href="http://twitter.com/download/iphone" r...
24   2017-07-18 16:08:03  <a href="http://twitter.com/download/iphone" r...
25   2017-07-18 00:07:08  <a href="http://twitter.com/download/iphone" r...
26   2017-07-17 16:17:36  <a href="http://twitter.com/download/iphone" r...
27   2017-07-16 23:58:41  <a href="http://twitter.com/download/iphone" r...
28   2017-07-16 20:14:00  <a href="http://twitter.com/download/iphone" r...
29   2017-07-15 23:25:31  <a href="http://twitter.com/download/iphone" r...
30   2017-07-15 16:51:35  <a href="http://twitter.com/download/iphone" r...
...                  ...                                                 ...
2326 2015-11-17 00:24:19  <a href="http://twitter.com/download/iphone" r...
2327 2015-11-17 00:06:54  <a href="http://twitter.com/download/iphone" r...
2328 2015-11-16 23:23:41  <a href="http://twitter.com/download/iphone" r...
2329 2015-11-16 21:54:18  <a href="http://twitter.com/download/iphone" r...
2330 2015-11-16 21:10:36  <a href="http://twitter.com/download/iphone" r...
2331 2015-11-16 20:32:58  <a href="http://twitter.com/download/iphone" r...
2332 2015-11-16 20:01:42  <a href="http://twitter.com/download/iphone" r...
2333 2015-11-16 19:31:45  <a href="http://twitter.com/download/iphone" r...
2334 2015-11-16 16:37:02  <a href="http://twitter.com/download/iphone" r...
2335 2015-11-16 16:11:11  <a href="http://twitter.com/download/iphone" r...
2336 2015-11-16 15:14:19  <a href="http://twitter.com/download/iphone" r...
2337 2015-11-16 14:57:41  <a href="http://twitter.com/download/iphone" r...
2338 2015-11-16 04:02:55  <a href="http://twitter.com/download/iphone" r...
2339 2015-11-16 03:55:04  <a href="http://twitter.com/download/iphone" r...
2340 2015-11-16 03:44:34  <a href="http://twitter.com/download/iphone" r...
2341 2015-11-16 03:22:39  <a href="http://twitter.com/download/iphone" r...
2342 2015-11-16 02:38:37  <a href="http://twitter.com/download/iphone" r...
2343 2015-11-16 01:59:36  <a href="http://twitter.com/download/iphone" r...
2344 2015-11-16 01:52:02  <a href="http://twitter.com/download/iphone" r...
2345 2015-11-16 01:22:45  <a href="http://twitter.com/download/iphone" r...
2346 2015-11-16 01:01:59  <a href="http://twitter.com/download/iphone" r...
2347 2015-11-16 00:55:59  <a href="http://twitter.com/download/iphone" r...
2348 2015-11-16 00:49:46  <a href="http://twitter.com/download/iphone" r...
2349 2015-11-16 00:35:11  <a href="http://twitter.com/download/iphone" r...
2350 2015-11-16 00:30:50  <a href="http://twitter.com/download/iphone" r...
2351 2015-11-16 00:24:50  <a href="http://twitter.com/download/iphone" r...
2352 2015-11-16 00:04:52  <a href="http://twitter.com/download/iphone" r...
2353 2015-11-15 23:21:54  <a href="http://twitter.com/download/iphone" r...
2354 2015-11-15 23:05:30  <a href="http://twitter.com/download/iphone" r...
2355 2015-11-15 22:32:08  <a href="http://twitter.com/download/iphone" r...

                                                    text  retweeted_status_id  \
0      This is Phineas. He's a mystical boy. Only eve...                  NaN
```

```
1      This is Tilly. She's just checking pup on you...              NaN
2      This is Archie. He is a rare Norwegian Pouncin...             NaN
3      This is Darla. She commenced a snooze mid meal...             NaN
4      This is Franklin. He would like you to stop ca...            NaN
5      Here we have a majestic great white breaching ...            NaN
6      Meet Jax. He enjoys ice cream so much he gets ...            NaN
7      When you watch your owner call another dog a g...            NaN
8      This is Zoey. She doesn't want to be one of th...            NaN
9      This is Cassie. She is a college pup. Studying...            NaN
10     This is Koda. He is a South Australian decksha...            NaN
11     This is Bruno. He is a service shark. Only get...            NaN
12     Here's a puppo that seems to be on the fence a...            NaN
13     This is Ted. He does his best. Sometimes that'...            NaN
14     This is Stuart. He's sporting his favorite fan...            NaN
15     This is Oliver. You're witnessing one of his m...            NaN
16     This is Jim. He found a fren. Taught him how t...            NaN
17     This is Zeke. He has a new stick. Very proud o...            NaN
18     This is Ralphus. He's powering up. Attempting ...            NaN
20     This is Gerald. He was just told he didn't get...            NaN
21     This is Jeffrey. He has a monopoly on the pool...           NaN
22     I've yet to rate a Venezuelan Hover Wiener. Th...           NaN
23     This is Canela. She attempted some fancy porch...           NaN
24     You may not have known you needed to see this ...           NaN
25     This... is a Jubilant Antarctic House Bear. We...           NaN
26     This is Maya. She's very shy. Rarely leaves he...           NaN
27     This is Mingus. He's a wonderful father to his...           NaN
28     This is Derek. He's late for a dog meeting. 13...           NaN
29     This is Roscoe. Another pupper fallen victim t...           NaN
30     @NonWhiteHat @MayhewMayhem omg hello tanner yo...           NaN
...                                                          ...            ...
2326   This is quite the dog. Gets really excited whe...           NaN
2327   This is a southern Vesuvius bumblegruff. Can d...           NaN
2328   Oh goodness. A super rare northeast Qdoba kang...           NaN
2329   Those are sunglasses and a jean jacket. 11/10 ...           NaN
2330   Unique dog here. Very small. Lives in containe...           NaN
2331   Here we have a mixed Asiago from the Galápagos...           NaN
2332   Look at this jokester thinking seat belt laws ...           NaN
2333   This is an extremely rare horned Parthenon. No...           NaN
2334   This is a funny dog. Weird toes. Won't come do...           NaN
2335   This is an Albanian 3 1/2 legged  Episcopalian...           NaN
2336      Can take selfies 11/10 https://t.co/ws2AMaNwPW           NaN
2337   Very concerned about fellow dog trapped in com...           NaN
2338   Not familiar with this breed. No tail (weird)...          NaN
2339   Oh my. Here you are seeing an Adobe Setter giv...           NaN
2340   Can stand on stump for what seems like a while...           NaN
2341   This appears to be a Mongolian Presbyterian mi...           NaN
2342   Here we have a well-established sunblockerspan...           NaN
2343   Let's hope this flight isn't Malaysian (lol). ...           NaN
```

```
2344  Here we have a northern speckled Rhododendron...              NaN
2345  This is the happiest dog you will ever see. Ve...             NaN
2346  Here is the Rand Paul of retrievers folks! He'...             NaN
2347  My oh my. This is a rare blond Canadian terrie...             NaN
2348  Here is a Siberian heavily armored polar bear ...             NaN
2349  This is an odd dog. Hard on the outside but lo...             NaN
2350  This is a truly beautiful English Wilson Staff...             NaN
2351  Here we have a 1949 1st generation vulpix. Enj...             NaN
2352  This is a purebred Piers Morgan. Loves to Netf...             NaN
2353  Here is a very happy pup. Big fan of well-main...             NaN
2354  This is a western brown Mitsubishi terrier. Up...             NaN
2355  Here we have a Japanese Irish Setter. Lost eye...             NaN


      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                        NaN
1                          NaN                        NaN
2                          NaN                        NaN
3                          NaN                        NaN
4                          NaN                        NaN
5                          NaN                        NaN
6                          NaN                        NaN
7                          NaN                        NaN
8                          NaN                        NaN
9                          NaN                        NaN
10                         NaN                        NaN
11                         NaN                        NaN
12                         NaN                        NaN
13                         NaN                        NaN
14                         NaN                        NaN
15                         NaN                        NaN
16                         NaN                        NaN
17                         NaN                        NaN
18                         NaN                        NaN
20                         NaN                        NaN
21                         NaN                        NaN
22                         NaN                        NaN
23                         NaN                        NaN
24                         NaN                        NaN
25                         NaN                        NaN
26                         NaN                        NaN
27                         NaN                        NaN
28                         NaN                        NaN
29                         NaN                        NaN
30                         NaN                        NaN
...                        ...                        ...
2326                       NaN                        NaN
2327                       NaN                        NaN
2328                       NaN                        NaN
```

```
2329                          NaN                   NaN
2330                          NaN                   NaN
2331                          NaN                   NaN
2332                          NaN                   NaN
2333                          NaN                   NaN
2334                          NaN                   NaN
2335                          NaN                   NaN
2336                          NaN                   NaN
2337                          NaN                   NaN
2338                          NaN                   NaN
2339                          NaN                   NaN
2340                          NaN                   NaN
2341                          NaN                   NaN
2342                          NaN                   NaN
2343                          NaN                   NaN
2344                          NaN                   NaN
2345                          NaN                   NaN
2346                          NaN                   NaN
2347                          NaN                   NaN
2348                          NaN                   NaN
2349                          NaN                   NaN
2350                          NaN                   NaN
2351                          NaN                   NaN
2352                          NaN                   NaN
2353                          NaN                   NaN
2354                          NaN                   NaN
2355                          NaN                   NaN


                                        expanded_urls  rating_numerator  \
0      https://twitter.com/dog_rates/status/892420643...                13
1      https://twitter.com/dog_rates/status/892177421...                13
2      https://twitter.com/dog_rates/status/891815181...                12
3      https://twitter.com/dog_rates/status/891689557...                13
4      https://twitter.com/dog_rates/status/891327558...                12
5      https://twitter.com/dog_rates/status/891087950...                13
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...                13
7      https://twitter.com/dog_rates/status/890729181...                13
8      https://twitter.com/dog_rates/status/890609185...                13
9      https://twitter.com/dog_rates/status/890240255...                14
10     https://twitter.com/dog_rates/status/890006608...                13
11     https://twitter.com/dog_rates/status/889880896...                13
12     https://twitter.com/dog_rates/status/889665388...                13
13     https://twitter.com/dog_rates/status/889638837...                12
14     https://twitter.com/dog_rates/status/889531135...                13
15     https://twitter.com/dog_rates/status/889278841...                13
16     https://twitter.com/dog_rates/status/888917238...                12
17     https://twitter.com/dog_rates/status/888804989...                13
18     https://twitter.com/dog_rates/status/888554962...                13
```

20	https://twitter.com/dog_rates/status/888078434...	12
21	https://twitter.com/dog_rates/status/887705289...	13
22	https://twitter.com/dog_rates/status/887517139...	14
23	https://twitter.com/dog_rates/status/887473957...	13
24	https://twitter.com/dog_rates/status/887343217...	13
25	https://twitter.com/dog_rates/status/887101392...	12
26	https://twitter.com/dog_rates/status/886983233...	13
27	https://www.gofundme.com/mingusneedsus,https:/...	13
28	https://twitter.com/dog_rates/status/886680336...	13
29	https://twitter.com/dog_rates/status/886366144...	12
30	NaN	12
...	...	...
2326	https://twitter.com/dog_rates/status/666411507...	2
2327	https://twitter.com/dog_rates/status/666407126...	7
2328	https://twitter.com/dog_rates/status/666396247...	9
2329	https://twitter.com/dog_rates/status/666373753...	11
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2332	https://twitter.com/dog_rates/status/666345417...	10
2333	https://twitter.com/dog_rates/status/666337882...	9
2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2343	https://twitter.com/dog_rates/status/666073100...	10
2344	https://twitter.com/dog_rates/status/666071193...	9
2345	https://twitter.com/dog_rates/status/666063827...	10
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2348	https://twitter.com/dog_rates/status/666055525...	10
2349	https://twitter.com/dog_rates/status/666051853...	2
2350	https://twitter.com/dog_rates/status/666050758...	10
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6
2353	https://twitter.com/dog_rates/status/666033412...	9
2354	https://twitter.com/dog_rates/status/666029285...	7
2355	https://twitter.com/dog_rates/status/666020888...	8

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None

24

4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
30	10	None	None	None	None	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None

```
        2347                        10          a      None      None      None      None
        2348                        10          a      None      None      None      None
        2349                        10         an      None      None      None      None
        2350                        10          a      None      None      None      None
        2351                        10       None      None      None      None      None
        2352                        10          a      None      None      None      None
        2353                        10          a      None      None      None      None
        2354                        10          a      None      None      None      None
        2355                        10       None      None      None      None      None

        [2175 rows x 17 columns]
```

Test

```
In [24]: # confirm changes (retweeted_status_id should be 0)
         df_archive.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                    2175 non-null object
in_reply_to_status_id       78 non-null float64
in_reply_to_user_id         78 non-null float64
timestamp                   2175 non-null datetime64[ns]
source                      2175 non-null object
text                        2175 non-null object
retweeted_status_id         0 non-null float64
retweeted_status_user_id    0 non-null float64
retweeted_status_timestamp  0 non-null object
expanded_urls               2117 non-null object
rating_numerator            2175 non-null int64
rating_denominator          2175 non-null int64
name                        2175 non-null object
doggo                       2175 non-null object
floofer                     2175 non-null object
pupper                      2175 non-null object
puppo                       2175 non-null object
dtypes: datetime64[ns](1), float64(4), int64(2), object(10)
memory usage: 305.9+ KB
```

Define 4. Drop unnecessary columns in `df_archive`. It is not properly an issue, but it will make it easier to analyze the data. That is why I have 9 bullet points in this section.
Code

```
In [25]: # drop columns
         df_archive.drop(['in_reply_to_status_id','in_reply_to_user_id','retweeted_status_id', '
```

Test

```
In [26]: # confirm changes
         df_archive

Out[26]:             tweet_id           timestamp  \
         0     892420643555336193 2017-08-01 16:23:56
         1     892177421306343426 2017-08-01 00:17:27
         2     891815181378084864 2017-07-31 00:18:03
         3     891689557279858688 2017-07-30 15:58:51
         4     891327558926688256 2017-07-29 16:00:24
         5     891087950875897856 2017-07-29 00:08:17
         6     890971913173991426 2017-07-28 16:27:12
         7     890729181411237888 2017-07-28 00:22:40
         8     890609185150312448 2017-07-27 16:25:51
         9     890240255349198849 2017-07-26 15:59:51
         10    890006608113172480 2017-07-26 00:31:25
         11    889880896479866881 2017-07-25 16:11:53
         12    889665388333682689 2017-07-25 01:55:32
         13    889638837579907072 2017-07-25 00:10:02
         14    889531135344209921 2017-07-24 17:02:04
         15    889278841981685760 2017-07-24 00:19:32
         16    888917238123831296 2017-07-23 00:22:39
         17    888804989199671297 2017-07-22 16:56:37
         18    888554962724278272 2017-07-22 00:23:06
         20    888078434458587136 2017-07-20 16:49:33
         21    887705289381826560 2017-07-19 16:06:48
         22    887517139158093824 2017-07-19 03:39:09
         23    887473957103951883 2017-07-19 00:47:34
         24    887343217045368832 2017-07-18 16:08:03
         25    887101392804085760 2017-07-18 00:07:08
         26    886983233522544640 2017-07-17 16:17:36
         27    886736880519319552 2017-07-16 23:58:41
         28    886680336477933568 2017-07-16 20:14:00
         29    886366144734445568 2017-07-15 23:25:31
         30    886267009285017600 2017-07-15 16:51:35
         ...                  ...                 ...
         2326  666411507551481857 2015-11-17 00:24:19
         2327  666407126856765440 2015-11-17 00:06:54
         2328  666396247373291520 2015-11-16 23:23:41
         2329  666373753744588802 2015-11-16 21:54:18
         2330  666362758909284353 2015-11-16 21:10:36
         2331  666353288456101888 2015-11-16 20:32:58
         2332  666345417576210432 2015-11-16 20:01:42
         2333  666337882303524864 2015-11-16 19:31:45
         2334  666293911632134144 2015-11-16 16:37:02
         2335  666287406224695296 2015-11-16 16:11:11
         2336  666273097616637952 2015-11-16 15:14:19
         2337  666268910803644416 2015-11-16 14:57:41
         2338  666104133288665088 2015-11-16 04:02:55
```

27

```
2339  666102155909144576 2015-11-16 03:55:04
2340  666099513787052032 2015-11-16 03:44:34
2341  666094000022159362 2015-11-16 03:22:39
2342  666082916733198337 2015-11-16 02:38:37
2343  666073100786774016 2015-11-16 01:59:36
2344  666071193221509120 2015-11-16 01:52:02
2345  666063827256086533 2015-11-16 01:22:45
2346  666058600524156928 2015-11-16 01:01:59
2347  666057090499244032 2015-11-16 00:55:59
2348  666055525042405380 2015-11-16 00:49:46
2349  666051853826850816 2015-11-16 00:35:11
2350  666050758794694657 2015-11-16 00:30:50
2351  666049248165822465 2015-11-16 00:24:50
2352  666044226329800704 2015-11-16 00:04:52
2353  666033412701032449 2015-11-15 23:21:54
2354  666029285002620928 2015-11-15 23:05:30
2355  666020888022790149 2015-11-15 22:32:08


                                                   text  \
0      This is Phineas. He's a mystical boy. Only eve...
1      This is Tilly. She's just checking pup on you...
2      This is Archie. He is a rare Norwegian Pouncin...
3      This is Darla. She commenced a snooze mid meal...
4      This is Franklin. He would like you to stop ca...
5      Here we have a majestic great white breaching ...
6      Meet Jax. He enjoys ice cream so much he gets ...
7      When you watch your owner call another dog a g...
8      This is Zoey. She doesn't want to be one of th...
9      This is Cassie. She is a college pup. Studying...
10     This is Koda. He is a South Australian decksha...
11     This is Bruno. He is a service shark. Only get...
12     Here's a puppo that seems to be on the fence a...
13     This is Ted. He does his best. Sometimes that'...
14     This is Stuart. He's sporting his favorite fan...
15     This is Oliver. You're witnessing one of his m...
16     This is Jim. He found a fren. Taught him how t...
17     This is Zeke. He has a new stick. Very proud o...
18     This is Ralphus. He's powering up. Attempting ...
20     This is Gerald. He was just told he didn't get...
21     This is Jeffrey. He has a monopoly on the pool...
22     I've yet to rate a Venezuelan Hover Wiener. Th...
23     This is Canela. She attempted some fancy porch...
24     You may not have known you needed to see this ...
25     This... is a Jubilant Antarctic House Bear. We...
26     This is Maya. She's very shy. Rarely leaves he...
27     This is Mingus. He's a wonderful father to his...
28     This is Derek. He's late for a dog meeting. 13...
29     This is Roscoe. Another pupper fallen victim t...
```

```
30     @NonWhiteHat @MayhewMayhem omg hello tanner yo...
...                                                    ...
2326   This is quite the dog. Gets really excited whe...
2327   This is a southern Vesuvius bumblegruff. Can d...
2328   Oh goodness. A super rare northeast Qdoba kang...
2329   Those are sunglasses and a jean jacket. 11/10 ...
2330   Unique dog here. Very small. Lives in containe...
2331   Here we have a mixed Asiago from the Galápagos...
2332   Look at this jokester thinking seat belt laws ...
2333   This is an extremely rare horned Parthenon. No...
2334   This is a funny dog. Weird toes. Won't come do...
2335   This is an Albanian 3 1/2 legged  Episcopalian...
2336      Can take selfies 11/10 https://t.co/ws2AMaNwPW
2337   Very concerned about fellow dog trapped in com...
2338   Not familiar with this breed. No tail (weird)...
2339   Oh my. Here you are seeing an Adobe Setter giv...
2340   Can stand on stump for what seems like a while...
2341   This appears to be a Mongolian Presbyterian mi...
2342   Here we have a well-established sunblockerspan...
2343   Let's hope this flight isn't Malaysian (lol). ...
2344   Here we have a northern speckled Rhododendron...
2345   This is the happiest dog you will ever see. Ve...
2346   Here is the Rand Paul of retrievers folks! He'...
2347   My oh my. This is a rare blond Canadian terrie...
2348   Here is a Siberian heavily armored polar bear ...
2349   This is an odd dog. Hard on the outside but lo...
2350   This is a truly beautiful English Wilson Staff...
2351   Here we have a 1949 1st generation vulpix. Enj...
2352   This is a purebred Piers Morgan. Loves to Netf...
2353   Here is a very happy pup. Big fan of well-main...
2354   This is a western brown Mitsubishi terrier. Up...
2355   Here we have a Japanese Irish Setter. Lost eye...

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

```
14      https://twitter.com/dog_rates/status/889531135...              13
15      https://twitter.com/dog_rates/status/889278841...              13
16      https://twitter.com/dog_rates/status/888917238...              12
17      https://twitter.com/dog_rates/status/888804989...              13
18      https://twitter.com/dog_rates/status/888554962...              13
20      https://twitter.com/dog_rates/status/888078434...              12
21      https://twitter.com/dog_rates/status/887705289...              13
22      https://twitter.com/dog_rates/status/887517139...              14
23      https://twitter.com/dog_rates/status/887473957...              13
24      https://twitter.com/dog_rates/status/887343217...              13
25      https://twitter.com/dog_rates/status/887101392...              12
26      https://twitter.com/dog_rates/status/886983233...              13
27      https://www.gofundme.com/mingusneedsus,https:/...              13
28      https://twitter.com/dog_rates/status/886680336...              13
29      https://twitter.com/dog_rates/status/886366144...              12
30                                                   NaN              12
...                                                  ...             ...
2326    https://twitter.com/dog_rates/status/666411507...               2
2327    https://twitter.com/dog_rates/status/666407126...               7
2328    https://twitter.com/dog_rates/status/666396247...               9
2329    https://twitter.com/dog_rates/status/666373753...              11
2330    https://twitter.com/dog_rates/status/666362758...               6
2331    https://twitter.com/dog_rates/status/666353288...               8
2332    https://twitter.com/dog_rates/status/666345417...              10
2333    https://twitter.com/dog_rates/status/666337882...               9
2334    https://twitter.com/dog_rates/status/666293911...               3
2335    https://twitter.com/dog_rates/status/666287406...               1
2336    https://twitter.com/dog_rates/status/666273097...              11
2337    https://twitter.com/dog_rates/status/666268910...              10
2338    https://twitter.com/dog_rates/status/666104133...               1
2339    https://twitter.com/dog_rates/status/666102155...              11
2340    https://twitter.com/dog_rates/status/666099513...               8
2341    https://twitter.com/dog_rates/status/666094000...               9
2342    https://twitter.com/dog_rates/status/666082916...               6
2343    https://twitter.com/dog_rates/status/666073100...              10
2344    https://twitter.com/dog_rates/status/666071193...               9
2345    https://twitter.com/dog_rates/status/666063827...              10
2346    https://twitter.com/dog_rates/status/666058600...               8
2347    https://twitter.com/dog_rates/status/666057090...               9
2348    https://twitter.com/dog_rates/status/666055525...              10
2349    https://twitter.com/dog_rates/status/666051853...               2
2350    https://twitter.com/dog_rates/status/666050758...              10
2351    https://twitter.com/dog_rates/status/666049248...               5
2352    https://twitter.com/dog_rates/status/666044226...               6
2353    https://twitter.com/dog_rates/status/666033412...               9
2354    https://twitter.com/dog_rates/status/666029285...               7
2355    https://twitter.com/dog_rates/status/666020888...               8
```

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
30	10	None	None	None	None	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None

```
2342                 10    None  None  None  None  None
2343                 10    None  None  None  None  None
2344                 10    None  None  None  None  None
2345                 10     the  None  None  None  None
2346                 10     the  None  None  None  None
2347                 10       a  None  None  None  None
2348                 10       a  None  None  None  None
2349                 10      an  None  None  None  None
2350                 10       a  None  None  None  None
2351                 10    None  None  None  None  None
2352                 10       a  None  None  None  None
2353                 10       a  None  None  None  None
2354                 10       a  None  None  None  None
2355                 10    None  None  None  None  None

[2175 rows x 11 columns]
```

Define 5. Drop null values from **expanded_urls** in df_archive. The **expanded_urls** is the only column that has a few missing values. I could not determine why that was the case nor I could fill these values out; therefore, I will drop these values.

Code

```
In [27]: # drop null values
         df_archive.dropna(inplace=True)
```

Test

```
In [28]: # confirm changes
         df_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2117 entries, 0 to 2355
Data columns (total 11 columns):
tweet_id              2117 non-null object
timestamp             2117 non-null datetime64[ns]
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: datetime64[ns](1), int64(2), object(8)
memory usage: 198.5+ KB
```

Define 6. Drop tweets that are not dogs. This analysis should only incorporate tweets of dogs; therefore, the next step is to delete rows that contain non-dog tweets. In order to do that, I

will merge the df_pred dataframe with df_archive and create a new dataframe df_archive2. The df_pred contains 3 predictions from the neural network whether a picture shows a dog. I assume that if any of the predictions suggest that it is a dog, I treated it as a dog. I checked a few tweets 'by hand' and it appeared to be true.

Code

```
In [29]: # merge predictions into the archive
         df_archive2 = df_archive.merge(df_pred, on=['tweet_id'], how='left')

In [30]: # check if the dataframes were merged correctly
         df_archive2.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2117 entries, 0 to 2116
Data columns (total 22 columns):
tweet_id             2117 non-null object
timestamp            2117 non-null datetime64[ns]
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
jpg_url              1994 non-null object
img_num              1994 non-null float64
p1                   1994 non-null object
p1_conf              1994 non-null float64
p1_dog               1994 non-null object
p2                   1994 non-null object
p2_conf              1994 non-null float64
p2_dog               1994 non-null object
p3                   1994 non-null object
p3_conf              1994 non-null float64
p3_dog               1994 non-null object
dtypes: datetime64[ns](1), float64(4), int64(2), object(15)
memory usage: 380.4+ KB


In [31]: # select pictures if any of the predictions suggest that it is a dog
         df_archive2 = df_archive2.query('p1_dog == True | p2_dog == True | p3_dog == True')
```

Test

```
In [32]: # check how many dog tweets are left
         df_archive2.info()
```

33

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1686 entries, 1 to 2116
Data columns (total 22 columns):
tweet_id               1686 non-null object
timestamp              1686 non-null datetime64[ns]
text                   1686 non-null object
expanded_urls          1686 non-null object
rating_numerator       1686 non-null int64
rating_denominator     1686 non-null int64
name                   1686 non-null object
doggo                  1686 non-null object
floofer                1686 non-null object
pupper                 1686 non-null object
puppo                  1686 non-null object
jpg_url                1686 non-null object
img_num                1686 non-null float64
p1                     1686 non-null object
p1_conf                1686 non-null float64
p1_dog                 1686 non-null object
p2                     1686 non-null object
p2_conf                1686 non-null float64
p2_dog                 1686 non-null object
p3                     1686 non-null object
p3_conf                1686 non-null float64
p3_dog                 1686 non-null object
dtypes: datetime64[ns](1), float64(4), int64(2), object(15)
memory usage: 303.0+ KB
```

Define 7. Clean **rating_denominator** in df_archive. Extract the values from text, delete tweets without ratings and tweets with multiple dogs. Change the data type to float.

8. Clean **rating_denominator** in df_archive. Extract the values from text, delete tweets without ratings and tweets with multiple dogs. Change the data type to float.

Code

```
In [33]: # change column width to view full text
         pd.options.display.max_colwidth = 200

In [34]: # import regular expression
         import re

In [35]: # extract ratings from the text column into a new column 'ratings'.
         # Some ratings need to be fixed manually because there are multiple xx/xx occurrences i
         # or there are multiple dogs in one picture
         df_archive2['ratings'] = df_archive2['text'].str.extract(r'(\d+\.?\d*/\d+){1}', expand

In [36]: # Extract rating numerators from the new column 'ratings' as float
         df_archive2['rating_numerator'] = df_archive2['ratings'].str.split('/').str.get(0).asty
```

```
In [37]: # Extract rating denominators from the new column 'ratings' as float
         df_archive2['rating_denominator'] = df_archive2['ratings'].str.split('/').str.get(1).as
```

```
In [38]: # drop ratings column after the extraction
         df_archive2.drop(['ratings'], axis=1, inplace=True)
```

In this section, I'm going to manually clean incorrecntly extracted ratings. I will delete tweets
with multiple dogs.

```
In [39]: df_archive2['rating_denominator'].value_counts()
```

```
Out[39]: 10.0      1669
         50.0         3
         80.0         2
         11.0         2
         130.0        1
         150.0        1
         2.0          1
         120.0        1
         110.0        1
         40.0         1
         90.0         1
         20.0         1
         7.0          1
         70.0         1
         Name: rating_denominator, dtype: int64
```

```
In [40]: df_archive2[df_archive2.rating_denominator == 70]
```

```
Out[40]:                    tweet_id              timestamp  \
         341   820690176645140481  2017-01-15 17:52:40


         341  The floofs have been released I repeat the floofs have been released. 84/70 https:


         341  https://twitter.com/dog_rates/status/820690176645140481/photo/1,https://twitter.co

              rating_numerator  rating_denominator  name doggo floofer pupper  ...      \
         341              84.0                70.0  None  None    None   None  ...

              img_num                          p1    p1_conf p1_dog      p2    p2_conf  \
         341      2.0  West_Highland_white_terrier   0.872064    True  kuvasz  0.059526

              p2_dog        p3 p3_conf p3_dog
         341    True   Samoyed  0.0374    True

         [1 rows x 22 columns]
```

```
In [41]: # drop row 341 because it is multiple dogs
         df_archive2.drop(341, inplace=True)

In [42]: df_archive2[df_archive2.rating_denominator == 7]

Out[42]:                tweet_id           timestamp  \
         405  810984652412424192 2016-12-19 23:06:23


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


         405  https://www.gofundme.com/sams-smile,https://twitter.com/dog_rates/status/810984652

              rating_numerator  rating_denominator name doggo floofer pupper  ...    \
         405              24.0                 7.0  Sam  None    None   None  ...

              img_num               p1   p1_conf p1_dog               p2   p2_conf  \
         405     1.0  golden_retriever  0.871342   True  Tibetan_mastiff  0.036708

              p2_dog                p3   p3_conf p3_dog
         405    True  Labrador_retriever  0.025823   True

         [1 rows x 22 columns]

In [43]: # drop row 405 as there is no rating provided
         df_archive2.drop(405, inplace=True)

In [44]: df_archive2[df_archive2.rating_denominator == 20]

Out[44]:                tweet_id           timestamp  \
         953  722974582966214656 2016-04-21 02:25:47

                                                             text  \
         953  Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a

                                                     expanded_urls  \
         953  https://twitter.com/dog_rates/status/722974582966214656/photo/1

              rating_numerator  rating_denominator  name doggo floofer pupper  ...    \
         953               4.0                20.0  None  None    None   None  ...

              img_num          p1   p1_conf p1_dog                          p2  \
         953     1.0  Great_Dane  0.246762   True  Greater_Swiss_Mountain_dog

              p2_conf p2_dog         p3   p3_conf p3_dog
         953  0.126131   True  Weimaraner  0.085297   True

         [1 rows x 22 columns]
```

```
In [45]: df_archive2.loc[953, 'rating_denominator'] = 10
         df_archive2.loc[953, 'rating_numerator'] = 13

In [46]: df_archive2[df_archive2.rating_denominator == 40]

Out[46]:                tweet_id           timestamp  \
         1218   697463031882764288 2016-02-10 16:51:59


         1218  Happy Wednesday here's a bucket of pups. 44/40 would pet all at once https://t.co

                                                            expanded_urls  \
         1218  https://twitter.com/dog_rates/status/697463031882764288/photo/1

               rating_numerator  rating_denominator  name doggo floofer pupper  ...     \
         1218              44.0                40.0  None  None    None  None  ...

               img_num                  p1   p1_conf p1_dog                 p2   p2_conf  \
         1218     1.0  Labrador_retriever  0.999885   True  golden_retriever  0.000098

               p2_dog   p3   p3_conf p3_dog
         1218    True  pug  0.000008   True

         [1 rows x 22 columns]

In [47]: # drop row 1218 because it is multiple dogs
         df_archive2.drop(1218, inplace=True)

In [48]: df_archive2[df_archive2.rating_denominator == 90]

Out[48]:                tweet_id           timestamp  \
         1016   713900603437621249 2016-03-27 01:29:02


         1016  Happy Saturday here's 9 puppers on a bench. 99/90 good work everybody https://t.c

                                                            expanded_urls  \
         1016  https://twitter.com/dog_rates/status/713900603437621249/photo/1

               rating_numerator  rating_denominator  name doggo floofer pupper  ...     \
         1016              99.0                90.0  None  None    None  None  ...

               img_num                  p1   p1_conf p1_dog                 p2   p2_conf  \
         1016     1.0  golden_retriever  0.371816   True  cocker_spaniel  0.177413

               p2_dog            p3   p3_conf p3_dog
         1016    True  Irish_setter  0.092725   True

         [1 rows x 22 columns]
```

```
In [49]: # drop row 1016 because it is multiple dogs
         df_archive2.drop(1016, inplace=True)

In [50]: df_archive2[df_archive2.rating_denominator == 110]

Out[50]:                   tweet_id             timestamp  \
         1411   684222868335505415 2016-01-05 04:00:18


         1411   Someone help the girl is being mugged. Several are distracting her while two stea

                                                          expanded_urls  \
         1411   https://twitter.com/dog_rates/status/684222868335505415/photo/1

                rating_numerator  rating_denominator  name doggo floofer pupper  ...    \
         1411              121.0               110.0  None  None   None   None  ...

                img_num                           p1   p1_conf p1_dog              p2  \
         1411       1.0  soft-coated_wheaten_terrier  0.791182   True  cocker_spaniel

                p2_conf p2_dog     p3   p3_conf p3_dog
         1411  0.072444   True  teddy  0.071486  False

         [1 rows x 22 columns]

In [51]: # drop row 1411 because it is multiple dogs
         df_archive2.drop(1411, inplace=True)

In [52]: df_archive2[df_archive2.rating_denominator == 120]

Out[52]:                   tweet_id             timestamp  \
         1552   677716515794329600 2015-12-18 05:06:23


                                                                   text  \
         1552   IT'S PUPPERGEDDON. Total of 144/120 ...I think https://t.co/ZanVtAtvIq

                                                          expanded_urls  \
         1552   https://twitter.com/dog_rates/status/677716515794329600/photo/1

                rating_numerator  rating_denominator  name doggo floofer pupper  ...    \
         1552              144.0               120.0  None  None   None   None  ...

                img_num     p1   p1_conf p1_dog    p2   p2_conf p2_dog    p3   p3_conf  \
         1552       1.0  teddy  0.662908  False  crib  0.031891  False  chow  0.025438

                p3_dog
         1552    True

         [1 rows x 22 columns]
```

```
In [53]: # drop row 1552 because it is multiple dogs
         df_archive2.drop(1552, inplace=True)

In [54]: df_archive2[df_archive2.rating_denominator == 130]

Out[54]:              tweet_id          timestamp  \
         1410  684225744407494656 2016-01-05 04:11:44


         1410  Two sneaky puppers were not initially seen, moving the rating to 143/130. Please


         1410  https://twitter.com/dog_rates/status/684225744407494656/photo/1,https://twitter.c

               rating_numerator  rating_denominator  name doggo floofer pupper  ...    \
         1410             143.0               130.0  None  None    None   None  ...

               img_num               p1   p1_conf p1_dog       p2   p2_conf p2_dog  \
         1410      2.0  golden_retriever  0.203249   True  Samoyed  0.067958   True

                         p3   p3_conf p3_dog
         1410  Great_Pyrenees  0.065327   True

         [1 rows x 22 columns]

In [55]: # drop row 1410 because it is multiple dogs
         df_archive2.drop(1410, inplace=True)

In [56]: df_archive2[df_archive2.rating_denominator == 150]

Out[56]:           tweet_id          timestamp  \
         702  758467244762497024 2016-07-28 01:00:57


                                                                                       text  \
         702  Why does this never happen at my front door... 165/150 https://t.co/HmwrdfEfUE

                                                        expanded_urls  \
         702  https://twitter.com/dog_rates/status/758467244762497024/video/1

              rating_numerator  rating_denominator  name doggo floofer pupper  ...    \
         702             165.0               150.0  None  None    None   None  ...

              img_num                 p1   p1_conf p1_dog         p2   p2_conf p2_dog  \
         702      1.0  Labrador_retriever  0.436377   True  Chihuahua  0.113956   True

                                      p3   p3_conf p3_dog
         702  American_Staffordshire_terrier  0.099689   True

         [1 rows x 22 columns]
```

```
In [57]:  # drop row 1410 because it is multiple dogs
          df_archive2.drop(702, inplace=True)

In [58]:  df_archive2[df_archive2.rating_denominator == 2]

Out[58]:                tweet_id             timestamp  \
          2096   666287406224695296 2015-11-16 16:11:11


          2096   This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished hardwood floc

                                                             expanded_urls  \
          2096   https://twitter.com/dog_rates/status/666287406224695296/photo/1

                 rating_numerator  rating_denominator name doggo floofer pupper   ...      \
          2096                1.0                 2.0   an  None    None   None   ...

                 img_num          p1    p1_conf p1_dog          p2   p2_conf p2_dog  \
          2096      1.0  Maltese_dog   0.857531   True  toy_poodle  0.063064   True

                              p3    p3_conf p3_dog
          2096   miniature_poodle   0.025581   True

          [1 rows x 22 columns]

In [59]:  # change row 2096
          df_archive2.loc[2096, 'rating_denominator'] = 10
          df_archive2.loc[2096, 'rating_numerator'] = 9

In [60]:  df_archive2[df_archive2.rating_denominator == 11]

Out[60]:                tweet_id             timestamp  \
          857    740373189193256964 2016-06-08 02:41:38
          1438   682962037429899265 2016-01-01 16:30:13


          857    After so many requests, this is Bretagne. She was the last surviving 9/11 search
          1438     This is Darrel. He just robbed a 7/11 and is in a high speed police chase. Was j


          857    https://twitter.com/dog_rates/status/740373189193256964/photo/1,https://twitter.c
          1438

                 rating_numerator  rating_denominator    name doggo floofer pupper  \
          857                 9.0                11.0    None  None    None   None
          1438                7.0                11.0  Darrel  None    None   None

                   ...  img_num              p1    p1_conf p1_dog      p2   p2_conf  \
          857      ...      3.0  golden_retriever   0.807644   True  kuvasz  0.101286
```

40

```
       1438   ...          1.0                dingo  0.278600  False  Chihuahua  0.155207

             p2_dog                      p3    p3_conf p3_dog
       857     True  Labrador_retriever  0.023785    True
       1438    True               loupe  0.153598   False

       [2 rows x 22 columns]
```

In [61]: df_archive2.loc[857, 'rating_denominator'] = 10
         df_archive2.loc[857, 'rating_numerator'] = 14

In [62]: df_archive2.loc[1438, 'rating_denominator'] = 10
         df_archive2.loc[1438, 'rating_numerator'] = 10

In [63]: df_archive2[df_archive2.rating_denominator == 80]

Out[63]:   tweet_id   timestamp \
         1041 710658690886586372 2016-03-18 02:46:49
         1615 675853064436391936 2015-12-13 01:41:41

         1041 Here's a brigade of puppers. All look very prepared for whatever happens next. 80
         1615   Here we have an entire platoon of puppers. Total score: 88/80 would pet all at o

         1041   https://twitter.c
         1615 https://twitter.com/dog_rates/status/675853064436391936/photo/1,https://twitter.c

               rating_numerator rating_denominator name doggo floofer pupper ...   \
         1041   80.0   80.0 None None   None None ...
         1615   88.0   80.0 None None   None None ...

               img_num   p1   p1_conf p1_dog   p2 \
         1041   1.0 soft-coated_wheaten_terrier 0.948617   True   Dandie_Dinmont
         1615   1.0   Labrador_retriever 0.868367   True   golden_retriever

               p2_conf p2_dog   p3   p3_conf p3_dog
         1041 0.018664   True   cairn 0.015943   True
         1615 0.043305   True vizsla 0.028207   True

         [2 rows x 22 columns]

In [64]: # drop rows 1041 and 1615 because it is multiple dogs
         df_archive2.drop([1041,1615], inplace=True)

In [65]: df_archive2[df_archive2.rating_denominator == 50]

Out[65]:   tweet_id   timestamp \
         990   716439118184652801 2016-04-03 01:36:11
```

```
      1061   709198395643068416 2016-03-14 02:04:08
      1136   704054845121142784 2016-02-28 21:25:30


      990                                     This is Bluebert. He just saw that both #FinalFur ma
      1061   From left to right:\nCletus, Jerome, Alejandro, Burp, &amp; Titson\nNone know whe
      1136                                                              Here is a whole f

                                                  expanded_urls  \
      990    https://twitter.com/dog_rates/status/716439118184652801/photo/1
      1061   https://twitter.com/dog_rates/status/709198395643068416/photo/1
      1136   https://twitter.com/dog_rates/status/704054845121142784/photo/1


            rating_numerator  rating_denominator        name doggo floofer pupper  \
      990                50.0                50.0    Bluebert  None    None  None
      1061               45.0                50.0        None  None    None  None
      1136               60.0                50.0           a  None    None  None


            ...    img_num            p1    p1_conf p1_dog                     p2  \
      990    ...       1.0  Siberian_husky  0.396495   True               malamute
      1061   ...       1.0          borzoi  0.490783   True  wire-haired_fox_terrier
      1136   ...       1.0   Great_Pyrenees  0.667939   True                 kuvasz


            p2_conf p2_dog                p3    p3_conf p3_dog
      990   0.317053   True       Eskimo_dog  0.273419   True
      1061  0.083513   True    English_setter  0.083184   True
      1136  0.228764   True  golden_retriever  0.043885   True

      [3 rows x 22 columns]
```

In [66]: df_archive2.loc[990, 'rating_denominator'] = 10
         df_archive2.loc[990, 'rating_numerator'] = 11

In [67]: # drop rows 1061 and 1136 because it is multiple dogs
         df_archive2.drop([1061,1136], inplace=True)

Clean Rating Numerator

In [68]: df_archive2.rating_numerator.value_counts()

Out[68]: 12.00   423
         10.00   361
         11.00   357
         13.00   230
         9.00   133
         8.00   68
         7.00   31
         14.00   24
         6.00   16

```
         5.00     13
         4.00      6
         3.00      5
         2.00      2
        13.50      1
        11.27      1
         0.00      1
        11.26      1
         9.75      1
        Name: rating_numerator, dtype: int64

In [69]: df_archive2[df_archive2.rating_numerator == 0]

Out[69]:              tweet_id           timestamp  \
        249  835152434251116546 2017-02-24 15:40:31


        249  When you're so blinded by your systematic plagiarism that you forget what day it i


        249  https://twitter.com/dog_rates/status/835152434251116546/photo/1,https://twitter.cc


             rating_numerator  rating_denominator  name doggo floofer pupper  ...     \
        249               0.0                10.0  None  None    None   None  ...


             img_num    p1   p1_conf p1_dog                              p2   p2_conf  \
        249      3.0  swing  0.967066  False  American_Staffordshire_terrier  0.012731


             p2_dog                     p3   p3_conf p3_dog
        249    True  Staffordshire_bullterrier  0.007039   True

        [1 rows x 22 columns]

In [70]: df_archive2[df_archive2.rating_numerator == 2]

Out[70]:               tweet_id           timestamp  \
        1538  678424312106393600 2015-12-20 03:58:55
        2010  667878741721415682 2015-11-21 01:34:35


        1538  This is Crystal. She's a shitty fireman. No sense of urgency. People could be dyi
        2010   This is Tedrick. He lives on the edge. Needs someone to hit the gas tho. Other t


                                                  expanded_urls  \
        1538  https://twitter.com/dog_rates/status/678424312106393600/photo/1
        2010  https://twitter.com/dog_rates/status/667878741721415682/photo/1


              rating_numerator  rating_denominator     name doggo floofer pupper  \
        1538               2.0                10.0  Crystal  None    None   None
```

```
2010                 2.0                 10.0  Tedrick  None     None    None


        ...   img_num           p1   p1_conf p1_dog                       p2  \
1538    ...       1.0  Maltese_dog  0.759945    True          toy_poodle
2010    ...       1.0    seat_belt  0.200373   False  miniature_pinscher


      p2_conf p2_dog          p3   p3_conf p3_dog
1538  0.101194   True    Shih-Tzu  0.056037    True
2010  0.106003   True  schipperke  0.104733    True


[2 rows x 22 columns]
```

In [71]: df_archive2[df_archive2.rating_numerator == 3]

```
Out[71]:                tweet_id            timestamp  \
        589   777885040357281792  2016-09-19 15:00:20
        977   718246886998687744  2016-04-08 01:19:36
        1036  711306686208872448  2016-03-19 21:41:44
        1706  673906403526995968  2015-12-07 16:46:21
        1948  668989615043424256  2015-11-24 03:08:48



        589   This is Wesley. He's clearly trespassing. Seems rather h*ckin violent too. Weapon
        977           This is Alexanderson. He's got a weird ass birth mark. Dreadful at fetch.
        1036     What hooligan sent in pictures w/out a dog in them? Churlish af. 3/10 just bc
        1706   Guys I'm getting real tired of this. We only rate dogs. Please don't send in oth
        1948  This is Bernie. He's taking his Halloween costume very seriously. Wants to be bak



        589   https://twitter.com/dog_rates/status/777885040357281792/photo/1,https://twitter.c
        977                                                              https://twitter.c
        1036  https://twitter.com/dog_rates/status/711306686208872448/photo/1,https://twitter.c
        1706                                                             https://twitter.c
        1948                                                             https://twitter.c


        rating_numerator  rating_denominator          name doggo floofer pupper  \
589                  3.0                10.0        Wesley  None    None   None
977                  3.0                10.0  Alexanderson  None    None   None
1036                 3.0                10.0          None  None    None   None
1706                 3.0                10.0          None  None    None   None
1948                 3.0                10.0        Bernie  None    None   None


        ...   img_num                 p1   p1_conf p1_dog  \
589     ...       1.0       Afghan_hound  0.123529    True
977     ...       1.0          Chihuahua  0.354488    True
1036    ...       1.0  leatherback_turtle  0.280835   False
1706    ...       1.0        toilet_seat  0.683319   False
1948    ...       1.0                pug  0.917326    True
```

```
                                    p2    p2_conf p2_dog              p3    p3_conf  \
589                              basset   0.119682   True   Siberian_husky   0.108709
977                              carton   0.159672  False   Siberian_husky   0.057498
1036                         loggerhead   0.123290  False   Dandie_Dinmont   0.086792
1706  soft-coated_wheaten_terrier   0.048928   True   Siberian_husky   0.030386
1948                         waffle_iron   0.014918  False        Chihuahua   0.013524


      p3_dog
589     True
977     True
1036    True
1706    True
1948    True


[5 rows x 22 columns]
```

In [72]: df_archive2[df_archive2.rating_numerator == 4]

```
Out[72]:               tweet_id            timestamp  \
797    747816857231626240  2016-06-28 15:40:07
1089   707420581654872064  2016-03-09 04:19:44
1475   680940246314430465  2015-12-27 02:36:20
1836   671122204919246848  2015-11-30 00:22:57
2050   667176164155375616  2015-11-19 03:02:47
2077   666649482315059201  2015-11-17 16:09:56



797        Viewer discretion is advised. This is a terrible attack in progress. Not even i
1089   This is Keurig. He's a rare dog. Laughs like an idiot tho. Head is basically a we
1475
1836        Two miniature golden retrievers here. Webbed paws. Don't walk very efficientl
2050   These are strange dogs. All have toupees. Long neck for dogs. In a shed of sorts?
2077     Cool dog. Enjoys couch. Low monotone bark. Very nice kicks. Pisses milk (must b

                                                expanded_urls  \
797    https://twitter.com/dog_rates/status/747816857231626240/photo/1
1089   https://twitter.com/dog_rates/status/707420581654872064/photo/1
1475   https://twitter.com/dog_rates/status/680940246314430465/photo/1
1836   https://twitter.com/dog_rates/status/671122204919246848/photo/1
2050   https://twitter.com/dog_rates/status/667176164155375616/photo/1
2077   https://twitter.com/dog_rates/status/666649482315059201/photo/1


      rating_numerator  rating_denominator    name doggo floofer pupper  \
797                4.0                10.0       a  None    None   None
1089               4.0                10.0  Keurig  None    None   None
1475               4.0                10.0   Alice  None    None   None
1836               4.0                10.0    None  None    None   None
```

```
2050                    4.0                    10.0   None   None     None   None
2077                    4.0                    10.0   None   None     None   None


         ...   img_num                         p1   p1_conf  p1_dog  \
797      ...      1.0                   Pembroke  0.768923    True
1089     ...      1.0                        ram  0.518215   False
1475     ...      1.0   soft-coated_wheaten_terrier  0.289598    True
1836     ...      1.0                      goose  0.351957   False
2050     ...      1.0   soft-coated_wheaten_terrier  0.318981    True
2077     ...      1.0                Border_collie  0.447803    True


                                 p2   p2_conf  p2_dog                 p3  \
797                       Chihuahua  0.029053    True   Shetland_sheepdog
1089                         kuvasz  0.149391    True      Great_Pyrenees
1475   West_Highland_white_terrier  0.157195    True         toy_poodle
1836                      Chihuahua  0.101228    True                hen
2050                Lakeland_terrier  0.215218    True         toy_poodle
2077               English_springer  0.170497    True             collie


       p3_conf  p3_dog
797   0.029035    True
1089  0.106003    True
1475  0.074435    True
1836  0.065818   False
2050  0.106014    True
2077  0.139206    True


[6 rows x 22 columns]

In [73]: df_archive2[df_archive2.rating_numerator == 5]

Out[73]:                 tweet_id             timestamp  \
        561   781661882474196992  2016-09-30 01:08:10
        750   751583847268179968  2016-07-09 01:08:47
        1288  691483041324204033  2016-01-25 04:49:38
        1421  683849932751646720  2016-01-04 03:18:23
        1569  677301033169788928  2015-12-17 01:35:24
        1592  676588346097852417  2015-12-15 02:23:26
        1671  674646392044941312  2015-12-09 17:46:48
        1746  672980819271634944  2015-12-05 03:28:25
        1858  670782429121134593  2015-11-29 01:52:48
        1946  668994913074286592  2015-11-24 03:29:51
        1970  668631377374486528  2015-11-23 03:25:17
        2073  666776908487630848  2015-11-18 00:36:17
        2112  666049248165822465  2015-11-16 00:24:50


        561                         Who keeps sending in pictures without dogs in them? This ne
```

```
750           Please stop sending it pictures that don't even have a doggo or pupper in
1288  When bae says they can't go out but you see them with someone else that same nigh
1421                          This is Jiminy. He's not the brightest dog. Needs to
1569                                        This is Juckson. He's totally on his
1592    This is Bubbles. He kinda resembles a fish. Always makes eye contact with u n
1671    Two gorgeous dogs here. Little waddling dog is a rebel. Refuses to look at cam
1746                      Extraordinary dog here. Looks large. Just a head. No body.
1858                                                            This dude slap
1946    Two gorgeous pups here. Both have cute fake horns(adorable). Barn in the back
1970          Meet Zeek. He is a grey Cumulonimbus. Zeek is hungry. Someone should fee
2073      This is Josep. He is a Rye Manganese mix. Can drive w eyes closed. Very irre
2112                        Here we have a 1949 1st generation vulpix. Enjoys sweat t


561
750
1288  https://twitter.com/dog_rates/status/691483041324204033/photo/1,https://twitter.c
1421
1569
1592
1671
1746
1858
1946
1970
2073
2112


      rating_numerator  rating_denominator     name doggo floofer  pupper  \
561                5.0                10.0     None  None    None    None
750                5.0                10.0     None doggo    None  pupper
1288               5.0                10.0     None  None    None    None
1421               5.0                10.0   Jiminy  None    None    None
1569               5.0                10.0  Juckson  None    None    None
1592               5.0                10.0  Bubbles  None    None    None
1671               5.0                10.0     None  None    None    None
1746               5.0                10.0     None  None    None    None
1858               5.0                10.0     None  None    None    None
1946               5.0                10.0     None  None    None    None
1970               5.0                10.0     Zeek  None    None    None
2073               5.0                10.0    Josep  None    None    None
2112               5.0                10.0     None  None    None    None

      ...  img_num                 p1   p1_conf  p1_dog  \
561   ...      1.0           Pembroke  0.438087    True
750   ...      1.0          dalmatian  0.868304    True
1288  ...      1.0         bloodhound  0.886232    True
1421  ...      1.0                hog  0.458855   False
```

47

```
1569    ...          1.0          Japanese_spaniel  0.661178    True
1592    ...          1.0              Boston_bull  0.976577    True
1671    ...          1.0    flat-coated_retriever  0.837448    True
1746    ...          1.0               car_mirror  0.232754   False
1858    ...          1.0                Chihuahua  0.952963    True
1946    ...          1.0                      hog  0.113789   False
1970    ...          1.0        miniature_schnauzer  0.904549    True
2073    ...          1.0                 seat_belt  0.375057   False
2112    ...          1.0        miniature_pinscher  0.560311    True


                              p2    p2_conf p2_dog                 p3    p3_conf  \
561            golden_retriever  0.226954    True             collie  0.070652
750                studio_couch  0.059623   False       snow_leopard  0.013876
1288  black-and-tan_coonhound  0.077420    True       Gordon_setter  0.009826
1421          Mexican_hairless  0.164906    True           wild_boar  0.111700
1569                  Pekinese  0.150119    True           Chihuahua  0.119720
1592            French_bulldog  0.014324    True           Chihuahua  0.002302
1671              groenendael  0.086166    True  Labrador_retriever  0.016052
1746                    basset  0.219461    True             beagle  0.112397
1858            French_bulldog  0.036575    True         Boston_bull  0.001977
1946          English_springer  0.089763    True      French_bulldog  0.082186
1970        Australian_terrier  0.022529    True        silky_terrier  0.015243
2073        miniature_pinscher  0.167175    True           Chihuahua  0.086951
2112              Rottweiler  0.243682    True            Doberman  0.154629


        p3_dog
561      True
750     False
1288     True
1421    False
1569     True
1592     True
1671     True
1746     True
1858     True
1946     True
1970     True
2073     True
2112     True


[13 rows x 22 columns]

In [74]: df_archive2[df_archive2.rating_numerator == 6]

Out[74]:                  tweet_id            timestamp  \
        186   848212111729840128  2017-04-01 16:35:01
        786   748575535303884801  2016-06-30 17:54:50
        867   739485634323156992  2016-06-05 15:54:48
```

```
1173   700462010979500032  2016-02-18 23:28:52
1243   695051054296211456  2016-02-04 01:07:39
1533   678740035362037760  2015-12-21 00:53:29
1560   677565715327688705  2015-12-17 19:07:09
1611   675891555769696257  2015-12-13 04:14:39
1739   673240798075449344  2015-12-05 20:41:29
1837   671115716440031232  2015-11-29 23:57:10
1866   670704688707301377  2015-11-28 20:43:53
1928   669375718304980992  2015-11-25 04:43:02
2002   668113020489474048  2015-11-21 17:05:31
2083   666430724426358785  2015-11-17 01:40:41
2103   666082916733198337  2015-11-16 02:38:37
2113   666044226329800704  2015-11-16 00:04:52
```

```
186    This is Jerry. He's doing a distinguished tongue slip. Slightly patronizing tbh.
786                           This is one of the most reckless puppers I've ever seen. Ho
867                            This is Kyle. He's a heavy drinker and an avid pot user.
1173                           This is Murphy. He's a mini golden retriever. Missing two
1243                        Meet Brian (pronounced "Kirk"). He's not amused by ur churl
1533                         Meet Tango. He's a large dog. Doesn't care much for pers
1560                      Contortionist pup here. Inside pentagram. Clearly worships
1611                           This is Donny. He's summoning the demon monster Babadook.
1739                        Magical floating dog here. Very calm. Always hangs by the p
1837                        Meet Phred. He isn't steering, looking at the road, or wear
1866                          Meet Danny. He's too good to look at the road when he'
1928                                 This is Billl. He's trying to
2002                       This is Alfie. He's that one hypocritical gym teach
2083                      Oh boy what a pup! Sunglasses take this one to t
2103                          Here we have a well-established sunblockersp
2113                     This is a purebred Piers Morgan. Loves to Netflix and ch
```

```
186                                                               https://twitter.c
786                                                               https://twitter.c
867    https://twitter.com/dog_rates/status/739485634323156992/photo/1,https://twitter.c
1173                                                              https://twitter.c
1243   https://twitter.com/dog_rates/status/695051054296211456/photo/1,https://twitter.c
1533                                                              https://twitter.c
1560                                                              https://twitter.c
1611                                                              https://twitter.c
1739                                                              https://twitter.c
1837                                                              https://twitter.c
1866                                                              https://twitter.c
1928                                                              https://twitter.c
2002                                                              https://twitter.c
2083                                                              https://twitter.c
2103                                                              https://twitter.c
```

	rating_numerator	rating_denominator	name	doggo	floofer	pupper
186	6.0	10.0	Jerry	None	None	None
786	6.0	10.0	one	None	None	None
867	6.0	10.0	Kyle	None	None	None
1173	6.0	10.0	Murphy	None	None	None
1243	6.0	10.0	Brian	None	None	None
1533	6.0	10.0	Tango	None	None	None
1560	6.0	10.0	None	None	None	None
1611	6.0	10.0	Donny	None	None	None
1739	6.0	10.0	None	None	None	None
1837	6.0	10.0	Phred	None	None	None
1866	6.0	10.0	Danny	None	None	None
1928	6.0	10.0	Billl	None	None	None
2002	6.0	10.0	Alfie	None	None	None
2083	6.0	10.0	None	None	None	None
2103	6.0	10.0	None	None	None	None
2113	6.0	10.0	a	None	None	None

	...	img_num	p1	p1_conf	p1_dog
186	...	1.0	Bedlington_terrier	0.333486	True
786	...	1.0	muzzle	0.176172	False
867	...	2.0	Walker_hound	0.640256	True
1173	...	1.0	hamster	0.678651	False
1243	...	1.0	Boston_bull	0.761454	True
1533	...	1.0	seat_belt	0.787164	False
1560	...	1.0	basset	0.397295	True
1611	...	1.0	Italian_greyhound	0.305637	True
1739	...	1.0	Airedale	0.443004	True
1837	...	1.0	malinois	0.406341	True
1866	...	1.0	Norwich_terrier	0.419838	True
1928	...	1.0	Airedale	0.168762	True
2002	...	1.0	Pembroke	0.548896	True
2083	...	1.0	llama	0.505184	False
2103	...	1.0	pug	0.489814	True
2113	...	1.0	Rhodesian_ridgeback	0.408143	True

	p2	p2_conf	p2_dog	p3
186	Ibizan_hound	0.245797	True	wallaby
786	seat_belt	0.160953	False	soft-coated_wheaten_terrier
867	English_foxhound	0.229799	True	beagle
1173	Pomeranian	0.110268	True	Angora
1243	pug	0.075395	True	Chihuahua
1533	sunglasses	0.045739	False	beagle
1560	Welsh_springer_spaniel	0.199554	True	purse
1611	whippet	0.232057	True	Great_Dane
1739	brown_bear	0.114162	False	Chesapeake_Bay_retriever

```
1837                   kelpie  0.143366   True                    dingo
1866                    cairn  0.351876   True          Norfolk_terrier
1928          Norfolk_terrier  0.107479   True         Lakeland_terrier
2002                 Cardigan  0.191101   True                   collie
2083            Irish_terrier  0.104109   True                    dingo
2103              bull_mastiff  0.404722   True            French_bulldog
2113                  redbone  0.360687   True        miniature_pinscher


        p3_conf  p3_dog
186     0.131647  False
786     0.086499   True
867     0.037754   True
1173    0.104139  False
1243    0.041598   True
1533    0.022525   True
1560    0.105641  False
1611    0.117806   True
1739    0.094639   True
1837    0.129802  False
1866    0.051094   True
1928    0.097846   True
2002    0.059814   True
2083    0.062071  False
2103    0.048960   True
2113    0.222752   True


[16 rows x 22 columns]
```

In [75]: df_archive2[df_archive2.rating_numerator == 7]

Out[75]:   tweet_id   timestamp \
        696   759099523532779520 2016-07-29 18:53:24
        819   745712589599014916 2016-06-22 20:18:30
        1111 706166467411222528 2016-03-05 17:16:20
        1200 698703483621523456 2016-02-14 03:01:06
        1231 696405997980676096 2016-02-07 18:51:43
        1352 687494652870668288 2016-01-14 04:41:12
        1358 687312378585812992 2016-01-13 16:36:55
        1385 685641971164143616 2016-01-09 01:59:19
        1504 679854723806179328 2015-12-24 02:42:51
        1511 679530280114372609 2015-12-23 05:13:38
        1513 679511351870550016 2015-12-23 03:58:25
        1528 678798276842360832 2015-12-21 04:44:55
        1643 675146535592706048 2015-12-11 02:54:12
        1723 673612854080196609 2015-12-06 21:19:54
        1732 673345638550134785 2015-12-06 03:38:05
        1765 672538107540070400 2015-12-03 22:09:14
        1779 672239279297454080 2015-12-03 02:21:48

```
1789   672068090318987265 2015-12-02 15:01:33
1843   670833812859932673 2015-11-29 05:16:59
1854   670792680469889025 2015-11-29 02:33:32
1859   670780561024270336 2015-11-29 01:45:22
1962   668815180734689280 2015-11-23 15:35:39
1999   668171859951755264 2015-11-21 20:59:20
2034   667491009379606528 2015-11-19 23:53:52
2036   667455448082227200 2015-11-19 21:32:34
2041   667405339315146752 2015-11-19 18:13:27
2058   667090893657276420 2015-11-18 21:23:57
2081   666437273139982337 2015-11-17 02:06:42
2084   666428276349472768 2015-11-17 01:30:57
2088   666407126856765440 2015-11-17 00:06:54
2115   666029285002620928 2015-11-15 23:05:30
```

```
696    Meet Toby. He has a drinking problem. Inflatable marijuana plant in the back is a
819                      This is Percy. He fell asleep at the wheel. Irresponsible af.
1111        This is Dexter. He's a shy pup. Doesn't bark much. Dreadful fetcher. Has r
1200                  This is Rusty. He has no respect for POULTRY products. Un
1231   This is Berb. He just found out that they have made 31 Kidz Bop CD's. Downright
1352                                                         This is Marq
1358                                "You got any games on your phone" 7/10
1385                         This is Otis. He just passed a cop while going 61
1504   This is Bruce. He's a rare pup. Covered in Frosted Flakes. Nifty gold teeth. Over
1511   Guys this really needs to stop. We've been over this way too many times. This is
1513        Say hello to William. He makes fun of others because he's terrified of his
1528                                                         This is L
1643                       This is Coops. He's yelling at the carp
1723      This is Jeffri. He's a speckled ice pupper. Very lazy. Enjoys the occasional
1732      This is Gerald. He's a fluffy lil yellow pup. Always looks like his favorite
1765                        Meet Darby. He's a Fiscal Tutankhamen
1779                                                  This is Ridley.
1789      Say hello to Gizmo. He's upset because he's not sure if he's really big or the
1843                                                      This is Jett
1854   This is Antony. He's a Sheraton Tetrahedron. Skips awkwardly. Doesn't look when h
1859        This is Traviss. He has no ears. Two rare dogs in background. I bet they all
1962      This is a wild Toblerone from Papua New Guinea. Mouth always open. Addicted to
1999             This is a Trans Siberian Kellogg named Alfonso. Huge ass eyeballs. Act
2034     Two dogs in this one. Both are rare Jujitsu Pythagoreans. One slightly whiter t
2036      This is Reese and Twips. Reese protects Twips. Both think they're too good for
2041                                                         This i
2058   This is Clybe. He is an Anemone Valdez. One ear works. Can look in 2 different di
2081     Here we see a lone northeastern Cumberbatch. Half ladybug. Only builds with bri
2084                       Here we have an Austrian Pulitzer. Collectors edition. Le
2088    This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made friends w
2115    This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 dogs he
```

```
                                                    expanded_urls  \
696    https://twitter.com/dog_rates/status/759099523532779520/photo/1
819    https://twitter.com/dog_rates/status/745712589599014916/photo/1
1111   https://twitter.com/dog_rates/status/706166467411222528/photo/1
1200   https://twitter.com/dog_rates/status/698703483621523456/photo/1
1231   https://twitter.com/dog_rates/status/696405997980676096/photo/1
1352   https://twitter.com/dog_rates/status/687494652870668288/photo/1
1358   https://twitter.com/dog_rates/status/687312378585812992/photo/1
1385   https://twitter.com/dog_rates/status/685641971164143616/photo/1
1504   https://twitter.com/dog_rates/status/679854723806179328/photo/1
1511   https://twitter.com/dog_rates/status/679530280114372609/photo/1
1513   https://twitter.com/dog_rates/status/679511351870550016/photo/1
1528   https://twitter.com/dog_rates/status/678798276842360832/photo/1
1643   https://twitter.com/dog_rates/status/675146535592706048/photo/1
1723   https://twitter.com/dog_rates/status/673612854080196609/photo/1
1732   https://twitter.com/dog_rates/status/673345638550134785/photo/1
1765   https://twitter.com/dog_rates/status/672538107540070400/photo/1
1779   https://twitter.com/dog_rates/status/672239279297454080/photo/1
1789   https://twitter.com/dog_rates/status/672068090318987265/photo/1
1843   https://twitter.com/dog_rates/status/670833812859932673/photo/1
1854   https://twitter.com/dog_rates/status/670792680469889025/photo/1
1859   https://twitter.com/dog_rates/status/670780561024270336/photo/1
1962   https://twitter.com/dog_rates/status/668815180734689280/photo/1
1999   https://twitter.com/dog_rates/status/668171859951755264/photo/1
2034   https://twitter.com/dog_rates/status/667491009379606528/photo/1
2036   https://twitter.com/dog_rates/status/667455448082227200/photo/1
2041   https://twitter.com/dog_rates/status/667405339315146752/photo/1
2058   https://twitter.com/dog_rates/status/667090893657276420/photo/1
2081   https://twitter.com/dog_rates/status/666437273139982337/photo/1
2084   https://twitter.com/dog_rates/status/666428276349472768/photo/1
2088   https://twitter.com/dog_rates/status/666407126856765440/photo/1
2115   https://twitter.com/dog_rates/status/666029285002620928/photo/1

      rating_numerator  rating_denominator     name doggo floofer  pupper  \
696                7.0                10.0     Toby  None    None    None
819                7.0                10.0    Percy  None    None    None
1111               7.0                10.0   Dexter  None    None    None
1200               7.0                10.0    Rusty  None    None    None
1231               7.0                10.0     Berb  None    None    None
1352               7.0                10.0     Marq  None    None    None
1358               7.0                10.0     None  None    None  pupper
1385               7.0                10.0     Otis  None    None  pupper
1504               7.0                10.0    Bruce  None    None    None
1511               7.0                10.0        a  None    None    None
1513               7.0                10.0  William  None    None    None
1528               7.0                10.0    Linda  None    None    None
1643               7.0                10.0    Coops  None    None    None
1723               7.0                10.0   Jeffri  None    None  pupper
```

```
1732                 7.0              10.0    Gerald  None      None      None
1765                 7.0              10.0     Darby  None      None      None
1779                 7.0              10.0    Ridley  None      None      None
1789                 7.0              10.0     Gizmo  None      None      None
1843                 7.0              10.0      Jett  None      None      None
1854                 7.0              10.0    Antony  None      None      None
1859                 7.0              10.0   Traviss  None      None      None
1962                 7.0              10.0         a  None      None      None
1999                 7.0              10.0         a  None      None      None
2034                 7.0              10.0      None  None      None      None
2036                 7.0              10.0     Reese  None      None      None
2041                 7.0              10.0     Biden  None      None      None
2058                 7.0              10.0     Clybe  None      None      None
2081                 7.0              10.0      None  None      None      None
2084                 7.0              10.0      None  None      None      None
2088                 7.0              10.0         a  None      None      None
2115                 7.0              10.0         a  None      None      None

         ...    img_num                        p1   p1_conf p1_dog  \
696      ...        1.0         Shetland_sheepdog  0.129034   True
819      ...        1.0                 seat_belt  0.379055  False
1111     ...        1.0                   Samoyed  0.430418   True
1200     ...        1.0          Brittany_spaniel  0.931963   True
1231     ...        1.0                    borzoi  0.132845   True
1352     ...        1.0                Rottweiler  0.391471   True
1358     ...        1.0                 seat_belt  0.703561  False
1385     ...        1.0          Lakeland_terrier  0.253839   True
1504     ...        1.0                     llama  0.887963  False
1511     ...        1.0                 dalmatian  0.750256   True
1513     ...        1.0                 Chihuahua  0.761972   True
1528     ...        1.0                  Airedale  0.583122   True
1643     ...        1.0                     dingo  0.288447  False
1723     ...        1.0              Newfoundland  0.223101   True
1732     ...        1.0                   hamster  0.761025  False
1765     ...        1.0               Siamese_cat  0.383937  False
1779     ...        1.0                       pug  0.332536   True
1789     ...        1.0                       pug  0.863385   True
1843     ...        1.0                  Pekinese  0.609853   True
1854     ...        1.0                brown_bear  0.882426  False
1859     ...        1.0        Labrador_retriever  0.244889   True
1962     ...        1.0                   redbone  0.461172   True
1999     ...        1.0                 Chihuahua  0.664834   True
2034     ...        1.0                    borzoi  0.852088   True
2036     ...        1.0           Tibetan_terrier  0.676376   True
2041     ...        1.0             Saint_Bernard  0.381377   True
2058     ...        1.0                 Chihuahua  0.959514   True
2081     ...        1.0                 Chihuahua  0.671853   True
2084     ...        1.0                  Pembroke  0.371361   True
```

```
2088    ...         1.0    black-and-tan_coonhound   0.529139     True
2115    ...         1.0                   redbone   0.506826     True
```

	p2	p2_conf	p2_dog	p3	\
696	kelpie	0.117508	True	Siberian_husky	
819	chow	0.062754	True	minibus	
1111	kuvasz	0.279600	True	Great_Pyrenees	
1200	Welsh_springer_spaniel	0.030695	True	beagle	
1231	Walker_hound	0.086005	True	Great_Pyrenees	
1352	miniature_pinscher	0.273595	True	Tibetan_mastiff	
1358	Great_Dane	0.139909	True	Weimaraner	
1385	Airedale	0.213349	True	three-toed_sloth	
1504	ram	0.095021	False	chow	
1511	jaguar	0.169007	False	zebra	
1513	black-footed_ferret	0.150605	False	squirrel_monkey	
1528	silky_terrier	0.129567	True	Lakeland_terrier	
1643	Cardigan	0.229944	True	Pembroke	
1723	Leonberg	0.111106	True	shovel	
1732	weasel	0.069362	False	Pomeranian	
1765	Chihuahua	0.160274	True	giant_panda	
1779	French_bulldog	0.258124	True	bull_mastiff	
1789	shopping_cart	0.125746	False	Border_terrier	
1843	Persian_cat	0.265442	False	Japanese_spaniel	
1854	toy_poodle	0.031355	True	miniature_poodle	
1859	American_black_bear	0.056994	False	brown_bear	
1962	Italian_greyhound	0.270733	True	miniature_pinscher	
1999	cowboy_boot	0.060343	False	giant_panda	
2034	ice_bear	0.132264	False	weasel	
2036	Irish_terrier	0.054933	True	Yorkshire_terrier	
2041	Leonberg	0.127998	True	golden_retriever	
2058	Italian_greyhound	0.005370	True	Pomeranian	
2081	beagle	0.124680	True	Saluki	
2084	chow	0.249394	True	Pomeranian	
2088	bloodhound	0.244220	True	flat-coated_retriever	
2115	miniature_pinscher	0.074192	True	Rhodesian_ridgeback	

	p3_conf	p3_dog
696	0.106708	True
819	0.052423	False
1111	0.117480	True
1200	0.012896	True
1231	0.065582	True
1352	0.041692	True
1358	0.021112	True
1385	0.083834	False
1504	0.002307	True
1511	0.006481	False
1513	0.028148	False

```
      1528  0.094727    True
      1643  0.190407    True
      1723  0.085626   False
      1732  0.064627    True
      1765  0.147745   False
      1779  0.120873    True
      1789  0.002972    True
      1843  0.027460    True
      1854  0.025743    True
      1859  0.053993   False
      1962  0.109752    True
      1999  0.059838   False
      2034  0.005730   False
      2036  0.040576    True
      2041  0.069357    True
      2058  0.002641    True
      2081  0.044094    True
      2084  0.241878    True
      2088  0.173810    True
      2115  0.072010    True

      [31 rows x 22 columns]
```

In [76]: df_archive2[df_archive2.rating_numerator == 13.5]

Out[76]:   tweet_id   timestamp \
         41   883482846933004288 2017-07-08 00:28:19

         41   This is Bella. She hopes her smile made you smile. If not, she is also offering you

         41   https://twitter.com/dog_rates/status/883482846933004288/photo/1,https://twitter.com

              rating_numerator rating_denominator   name doggo floofer pupper   ...   \
         41   13.5   10.0 Bella None   None   None   ...

              img_num   p1   p1_conf p1_dog   p2   p2_conf \
         41   1.0 golden_retriever 0.943082   True Labrador_retriever   0.032409

              p2_dog   p3   p3_conf p3_dog
         41   True kuvasz 0.005501   True

         [1 rows x 22 columns]

Test

In [77]: # confirm that all denominators are 10
         df_archive2['rating_denominator'].value_counts()

```
Out[77]: 10.0    1674
         Name: rating_denominator, dtype: int64

In [78]: # check for changed data type (float for ratings)
         df_archive2.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1674 entries, 1 to 2116
Data columns (total 22 columns):
tweet_id              1674 non-null object
timestamp             1674 non-null datetime64[ns]
text                  1674 non-null object
expanded_urls         1674 non-null object
rating_numerator      1674 non-null float64
rating_denominator    1674 non-null float64
name                  1674 non-null object
doggo                 1674 non-null object
floofer               1674 non-null object
pupper                1674 non-null object
puppo                 1674 non-null object
jpg_url               1674 non-null object
img_num               1674 non-null float64
p1                    1674 non-null object
p1_conf               1674 non-null float64
p1_dog                1674 non-null object
p2                    1674 non-null object
p2_conf               1674 non-null float64
p2_dog                1674 non-null object
p3                    1674 non-null object
p3_conf               1674 non-null float64
p3_dog                1674 non-null object
dtypes: datetime64[ns](1), float64(6), object(15)
memory usage: 300.8+ KB


In [79]: # confirm changes
         df_archive2.rating_numerator.value_counts()

Out[79]: 12.00    423
         10.00    361
         11.00    357
         13.00    230
         9.00     133
         8.00      68
         7.00      31
         14.00     24
         6.00      16
         5.00      13
         4.00       6
```

```
3.00       5
2.00       2
13.50      1
11.27      1
0.00       1
11.26      1
9.75       1
Name: rating_numerator, dtype: int64
```

All rating numerators look clean.

Define 8. Clean names using regular expression in df_archive. A useful pattern to find these "fake names" is that they started with lowercase, instead, the "real names" always started with uppercase, moreover, the real ones usually are placed after words like "named", "is" etc...so it'is possible to clean this issue using a regex

Code

```
In [80]: # extract correct names from text
         df_archive2['name'] = df_archive2['text'].str.extract(r'([is|named|to|meet]+\s[A-Z][\W]
```

```
In [81]: # Change 'I' to NaN
         df_archive2.loc[df_archive2['name'] == "I", 'name'] = np.nan
```

Test

```
In [82]: df_archive2['name'].value_counts()
```

```
Out[82]: Cooper      10
         Charlie     10
         Lucy        10
         Tucker       9
         Oliver       9
         Penny        8
         Daisy        7
         Sadie        7
         Winston      7
         Toby         7
         Lola         6
         Koda         6
         Jax          6
         Bo           6
         Bella        5
         Oscar        5
         Zoey         5
         Rusty        5
         Stanley      5
         Reggie       5
         Leo          5
         Louis        4
         Jack         4
```

```
             Gus          4
             Maggie       4
             Oakley       4
             Sophie       4
             Milo         4
             Finn         4
             Boomer       4
                          ..
             Gingivitis   1
             Kona         1
             Rand         1
             Yoshi        1
             Pumpkin      1
             Rueben       1
             Jim          1
             Bobble       1
             Napolean     1
             Lili         1
             Murphy       1
             Diogi        1
             Franq        1
             Kuyu         1
             Stephanus    1
             Molly        1
             Winifred     1
             Timison      1
             Mollie       1
             Bobbay       1
             Blanket      1
             Pawnd        1
             Lilli        1
             Bradley      1
             Crimson      1
             Kevon        1
             Orion        1
             Kayla        1
             Eriq         1
             Kenny        1
             Name: name, Length: 942, dtype: int64
```

In [83]: df_archive2.head()

Out[83]:   tweet_id   timestamp \
         1 892177421306343426 2017-08-01 00:17:27
         2 891815181378084864 2017-07-31 00:18:03
         3 891689557279858688 2017-07-30 15:58:51
         4 891327558926688256 2017-07-29 16:00:24
         5 891087950875897856 2017-07-29 00:08:17

```
       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 the
       3                                                    This is Darla. She commen
       4  This is Franklin. He would like you to stop calling him "cute." He is a very fierce
       5  Here we have a majestic great white breaching off South Africa's coast. Absolutely h


       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/
       5                                                    https://twitter.com/

          rating_numerator  rating_denominator       name doggo floofer pupper  ...      \
       1              13.0                10.0      Tilly  None    None   None  ...
       2              12.0                10.0     Archie  None    None   None  ...
       3              13.0                10.0      Darla  None    None   None  ...
       4              12.0                10.0   Franklin  None    None   None  ...
       5              13.0                10.0        NaN  None    None   None  ...

         img_num                       p1    p1_conf p1_dog                    p2  \
       1     1.0                Chihuahua   0.323581   True             Pekinese
       2     1.0                Chihuahua   0.716012   True              malamute
       3     1.0              paper_towel   0.170278  False   Labrador_retriever
       4     2.0                   basset   0.555712   True      English_springer
       5     1.0  Chesapeake_Bay_retriever   0.425595   True        Irish_terrier

          p2_conf p2_dog                         p3    p3_conf p3_dog
       1  0.090647   True                   papillon   0.068957   True
       2  0.078253   True                      kelpie   0.031379   True
       3  0.168086   True                    spatula   0.040836  False
       4  0.225770   True  German_short-haired_pointer   0.175219   True
       5  0.116317   True               Indian_elephant   0.076902  False

       [5 rows x 22 columns]

In [84]: df_archive2.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1674 entries, 1 to 2116
Data columns (total 22 columns):
tweet_id              1674 non-null object
timestamp             1674 non-null datetime64[ns]
text                  1674 non-null object
expanded_urls         1674 non-null object
rating_numerator      1674 non-null float64
```

```
rating_denominator    1674 non-null float64
name                  1321 non-null object
doggo                 1674 non-null object
floofer               1674 non-null object
pupper                1674 non-null object
puppo                 1674 non-null object
jpg_url               1674 non-null object
img_num               1674 non-null float64
p1                    1674 non-null object
p1_conf               1674 non-null float64
p1_dog                1674 non-null object
p2                    1674 non-null object
p2_conf               1674 non-null float64
p2_dog                1674 non-null object
p3                    1674 non-null object
p3_conf               1674 non-null float64
p3_dog                1674 non-null object
dtypes: datetime64[ns](1), float64(6), object(15)
memory usage: 300.8+ KB
```

1.4.2 Tidiness

1. Melt 'doggo' 'floofer' 'pupper' 'puppo' in `df_archive`.

```
In [85]: df_archive2.head(1)

Out[85]:               tweet_id           timestamp  \
         1  892177421306343426 2017-08-01 00:17:27


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

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

            rating_numerator  rating_denominator   name doggo floofer pupper  ...    \
         1              13.0                10.0  Tilly  None    None   None  ...

           img_num         p1   p1_conf p1_dog        p2   p2_conf p2_dog        p3  \
         1     1.0  Chihuahua  0.323581   True  Pekinese  0.090647   True  papillon

             p3_conf p3_dog
         1  0.068957   True


         [1 rows x 22 columns]

In [86]: df_archive2['stage'] = df_archive2.doggo + df_archive2.floofer + df_archive2.pupper + d
         df_archive2.head(1)
```

```
Out[86]:                     tweet_id            timestamp  \
         1  892177421306343426 2017-08-01 00:17:27


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

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

            rating_numerator  rating_denominator   name doggo floofer pupper  \
         1              13.0                10.0  Tilly  None    None   None

                      ...                  p1   p1_conf  p1_dog        p2   p2_conf p2_dog  \
         1            ...           Chihuahua  0.323581    True  Pekinese  0.090647   True

                 p3   p3_conf p3_dog              stage
         1  papillon  0.068957   True  NoneNoneNoneNone

         [1 rows x 23 columns]
```

In [87]: `df_archive2.stage.value_counts()`

```
Out[87]: NoneNoneNoneNone       1414
         NoneNonepupperNone      168
         doggoNoneNoneNone        54
         NoneNoneNonepuppo        21
         doggoNonepupperNone       8
         NoneflooferNoneNone       7
         doggoflooferNoneNone      1
         doggoNoneNonepuppo        1
         Name: stage, dtype: int64
```

In [88]: `df_archive2.loc[df_archive2.stage == 'NoneNoneNoneNone', 'stage'] = 'None'`
         `df_archive2.loc[df_archive2.stage == 'NoneNonepupperNone', 'stage'] = 'pupper'`
         `df_archive2.loc[df_archive2.stage == 'doggoNoneNoneNone', 'stage'] = 'doggo'`
         `df_archive2.loc[df_archive2.stage == 'NoneNoneNonepuppo', 'stage'] = 'puppo'`
         `df_archive2.loc[df_archive2.stage == 'doggoNonepupperNone', 'stage'] = 'multiple'`
         `df_archive2.loc[df_archive2.stage == 'NoneflooferNoneNone', 'stage'] = 'floofer'`
         `df_archive2.loc[df_archive2.stage == 'doggoflooferNoneNone', 'stage'] = 'multiple'`
         `df_archive2.loc[df_archive2.stage == 'doggoNoneNonepuppo', 'stage'] = 'multiple'`

In [89]: `df_archive2.stage.value_counts()`

```
Out[89]: None        1414
         pupper       168
         doggo         54
         puppo         21
         multiple      10
         floofer        7
         Name: stage, dtype: int64
```

```
In [90]: # drop individual stages columns
         df_archive2.drop(['floofer','doggo','pupper', 'puppo'], axis=1, inplace = True)
         df_archive2.head(1)

Out[90]:               tweet_id            timestamp  \
         1  892177421306343426 2017-08-01 00:17:27


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

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

            rating_numerator  rating_denominator   name  \
         1              13.0                10.0  Tilly

                                              jpg_url  img_num         p1  \
         1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg      1.0  Chihuahua

            p1_conf p1_dog        p2   p2_conf p2_dog        p3   p3_conf p3_dog stage
         1  0.323581   True  Pekinese  0.090647   True  papillon  0.068957   True  None

In [91]: df_pred.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id     2075 non-null object
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(1), object(5)
memory usage: 152.1+ KB
```

2. Join all data frames into `twitter_archive_master.csv`.

```
In [92]: twitter_archive_master = df_archive2.merge(df_json, on=['tweet_id'], how='left')

In [93]: twitter_archive_master.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1674 entries, 0 to 1673
Data columns (total 21 columns):
tweet_id             1674 non-null object
timestamp            1674 non-null datetime64[ns]
text                 1674 non-null object
expanded_urls        1674 non-null object
rating_numerator     1674 non-null float64
rating_denominator   1674 non-null float64
name                 1321 non-null object
jpg_url              1674 non-null object
img_num              1674 non-null float64
p1                   1674 non-null object
p1_conf              1674 non-null float64
p1_dog               1674 non-null object
p2                   1674 non-null object
p2_conf              1674 non-null float64
p2_dog               1674 non-null object
p3                   1674 non-null object
p3_conf              1674 non-null float64
p3_dog               1674 non-null object
stage                1674 non-null object
retweet_count        1674 non-null int64
favorite_count       1674 non-null int64
dtypes: datetime64[ns](1), float64(6), int64(2), object(12)
memory usage: 287.7+ KB
```

In [94]: twitter_archive_master.head(1)

Out[94]:   tweet_id   timestamp \
        0 892177421306343426 2017-08-01 00:17:27

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

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

           rating_numerator rating_denominator   name \
        0   13.0   10.0 Tilly

                                               jpg_url img_num   p1 \
        0 https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg   1.0 Chihuahua

                   ...   p1_dog   p2 p2_conf p2_dog   p3 p3_conf \
        0   ...   True Pekinese 0.090647   True papillon 0.068957

                                           64
```

```
      p3_dog stage  retweet_count  favorite_count
0     True  None            6514           33819

[1 rows x 21 columns]
```

In [95]: # save to csv
         twitter_archive_master.to_csv('twitter_archive_master.csv', index=False)

1.5 Analysis

In [96]: ax = twitter_archive_master.groupby('stage').favorite_count.mean().sort_values(ascendin
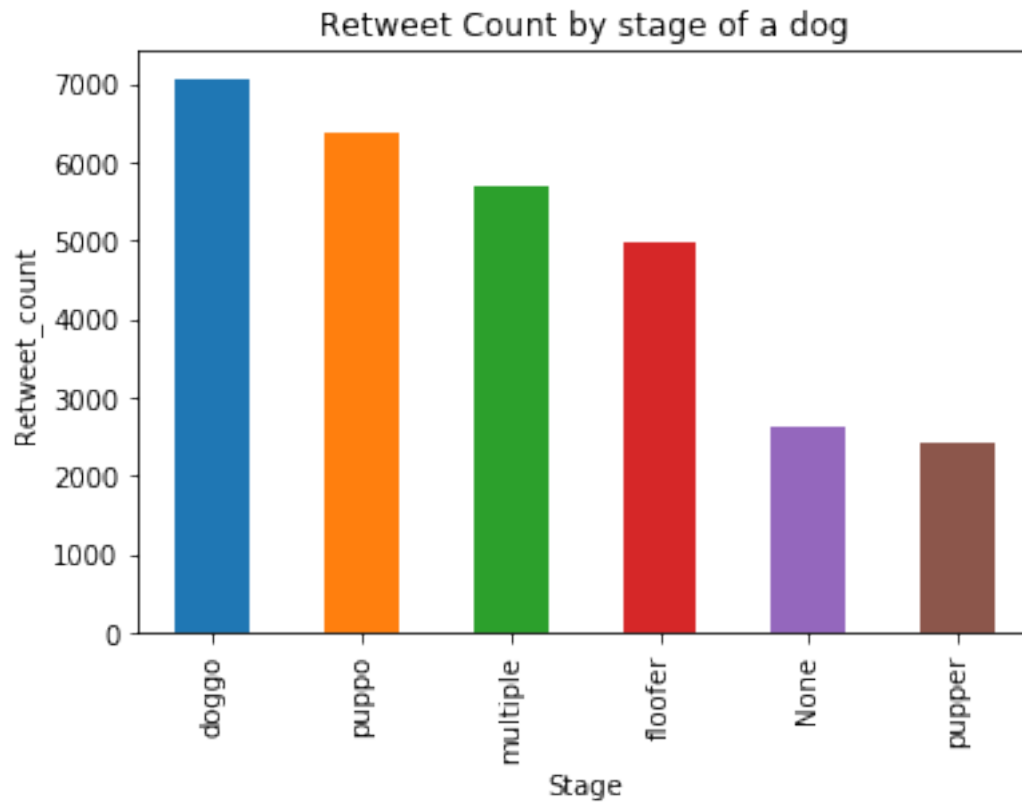   ax.set_xlabel("Stage");
   ax.set_ylabel("Favorite_count");
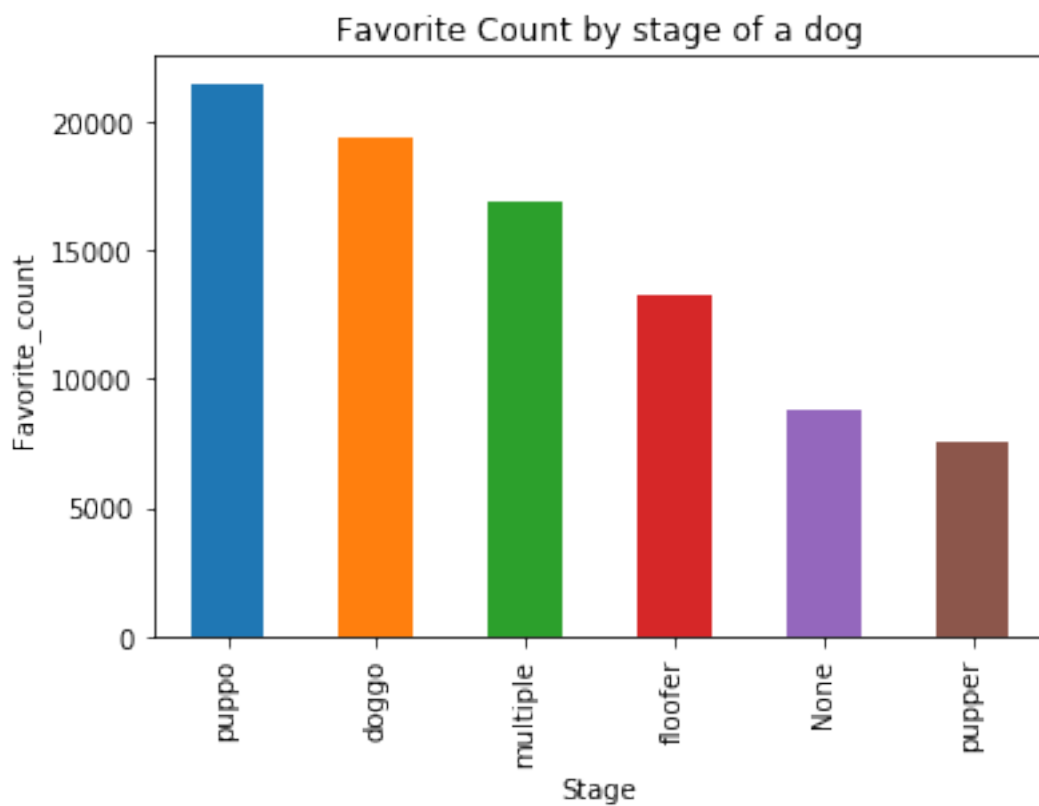
In [97]: ax = twitter_archive_master.groupby('stage').retweet_count.mean().sort_values(ascending
         ax.set_xlabel("Stage");
         ax.set_ylabel("Retweet_count");

Retweet Count by stage of a dog

```
In [98]: ax = twitter_archive_master.groupby('stage').rating_numerator.mean().sort_values(ascend
         ax.set_xlabel("Stage");
         ax.set_ylabel("rating_numerator");
```

Rating numerator by stage of a dog
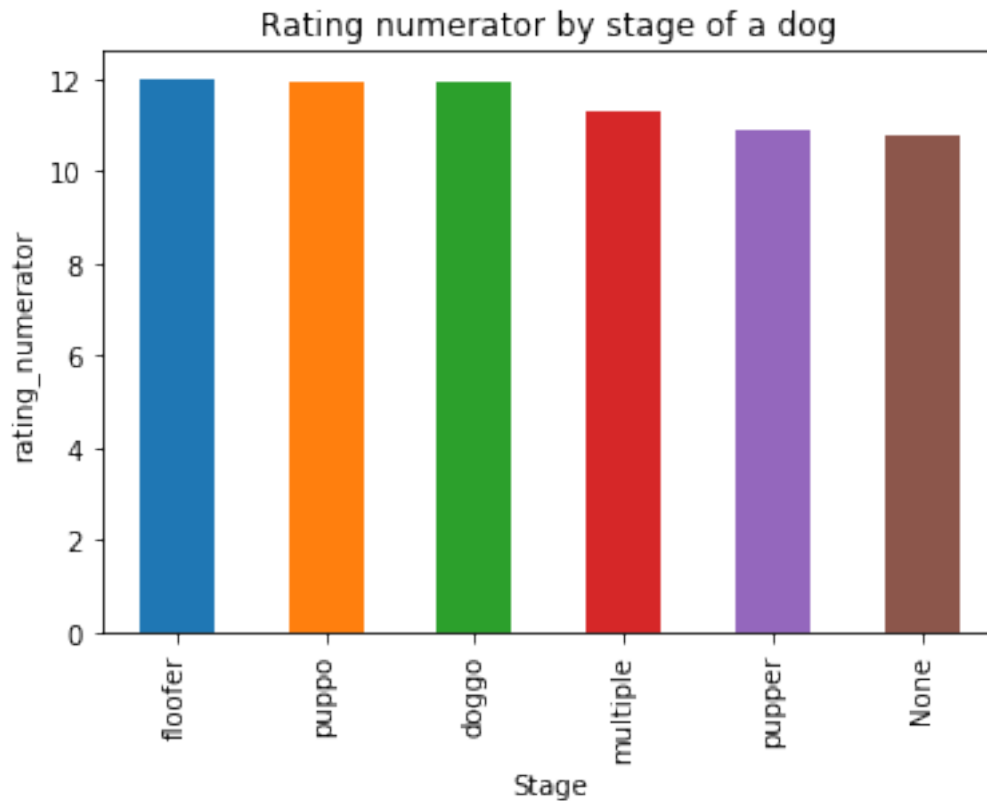
The chart shows that tweets that contain dogs' stages like 'puppo' or 'doggo' tend to have higher count favorite count. On the other hand, 'puppers' have the lowest favorite count, even lower then 'None'. The same patter is present in the 'retweet_count'. In addition, dogs that have 'stage' added to the tweets, tend to have higher rating numerators. 'Floofers' tend to have the greatest rating numerators.

```
In [99]: twitter_archive_master.rating_numerator.mean()

Out[99]: 10.840370370370369
```

Mean rating_numerator for the tweets is 10.84.

```
In [100]: twitter_archive_master[twitter_archive_master.stage == 'floofer'].rating_numerator.mea

Out[100]: 12.0
```

Mean rating_numerator for floofers is 12.

```
In [101]: ax = twitter_archive_master.groupby('rating_numerator').favorite_count.mean().sort_val
          ax.set_xlabel("rating_numerator");
          ax.set_ylabel("favorite_count");
```

Favorite Count by rating numerator
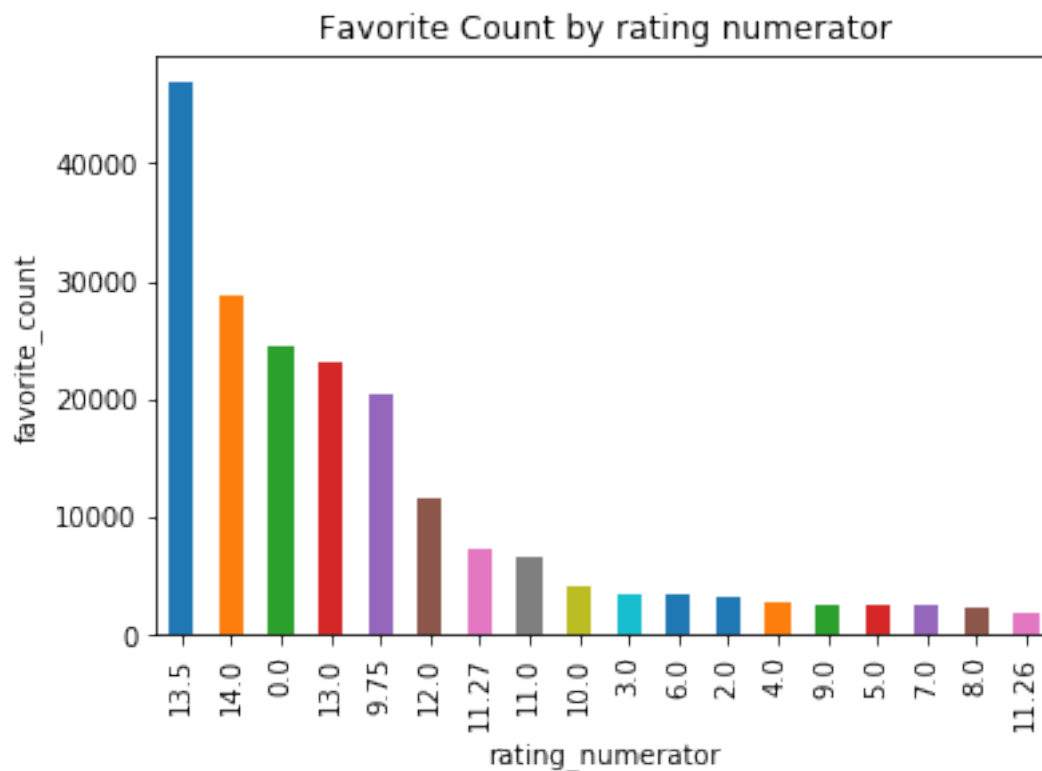

```
In [102]: twitter_archive_master.rating_numerator.value_counts()

Out[102]: 12.00    423
          10.00    361
          11.00    357
          13.00    230
          9.00     133
          8.00      68
          7.00      31
          14.00     24
          6.00      16
          5.00      13
          4.00       6
          3.00       5
          2.00       2
          13.50      1
          11.27      1
          0.00       1
          11.26      1
          9.75       1
          Name: rating_numerator, dtype: int64
```

The chart shows that dogs with higher ratings also tend to get higher favorite count. An exception is rating numerator of 0; however, only one dog deceived that rating; therefore, we can ignore that results as the sample size is not big enough.

```
In [103]: twitter_archive_master.favorite_count.mean()

Out[103]: 9268.3655913978491

In [104]: twitter_archive_master.groupby('rating_numerator').favorite_count.mean().sort_values(a

Out[104]: rating_numerator
          13.50    46860.000000
          14.00    28772.416667
          0.00     24574.000000
          13.00    23103.182609
          9.75     20296.000000
          12.00    11473.617021
          11.27     7320.000000
          11.00     6661.815126
          10.00     4124.229917
          3.00      3384.200000
          6.00      3382.312500
          2.00      3162.500000
          4.00      2780.166667
          9.00      2524.849624
          5.00      2518.923077
          7.00      2422.290323
          8.00      2233.882353
          11.26     1879.000000
          Name: favorite_count, dtype: float64

In [105]: ax = twitter_archive_master.groupby('name').favorite_count.mean().sort_values(ascendin
          ax.set_xlabel("Name");
          ax.set_ylabel("Favorite Count");
```
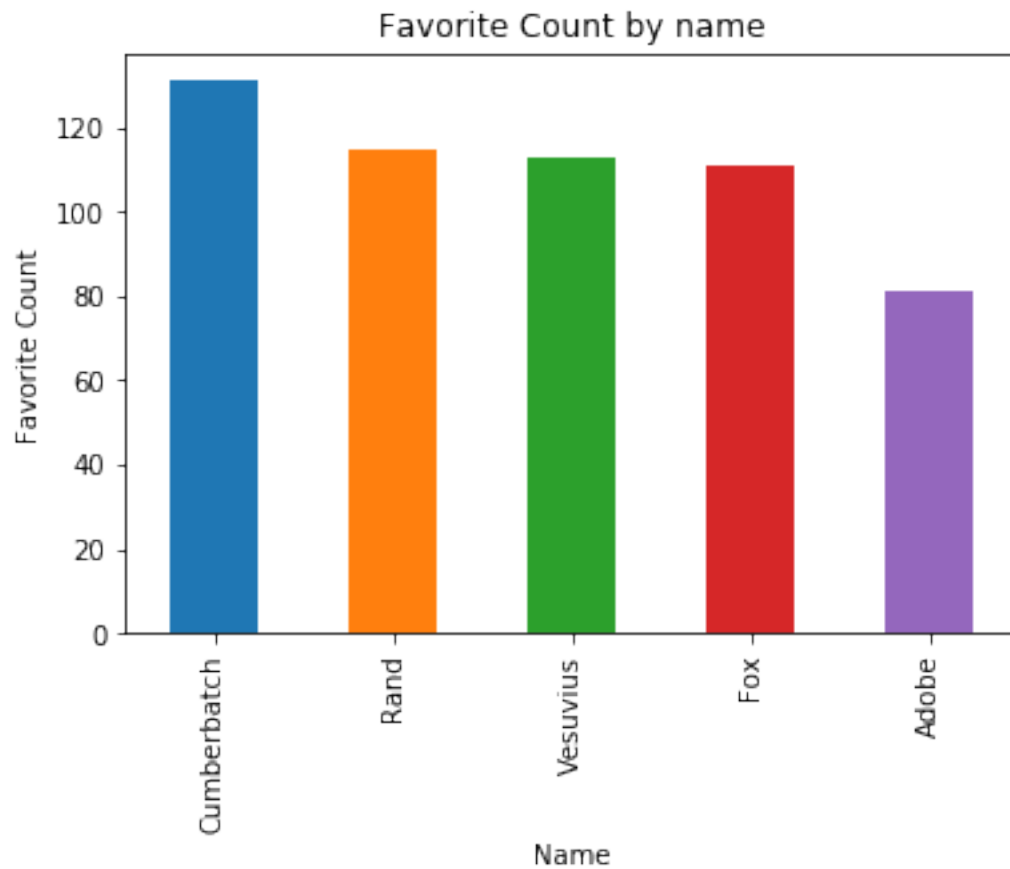
Favorite Count by name
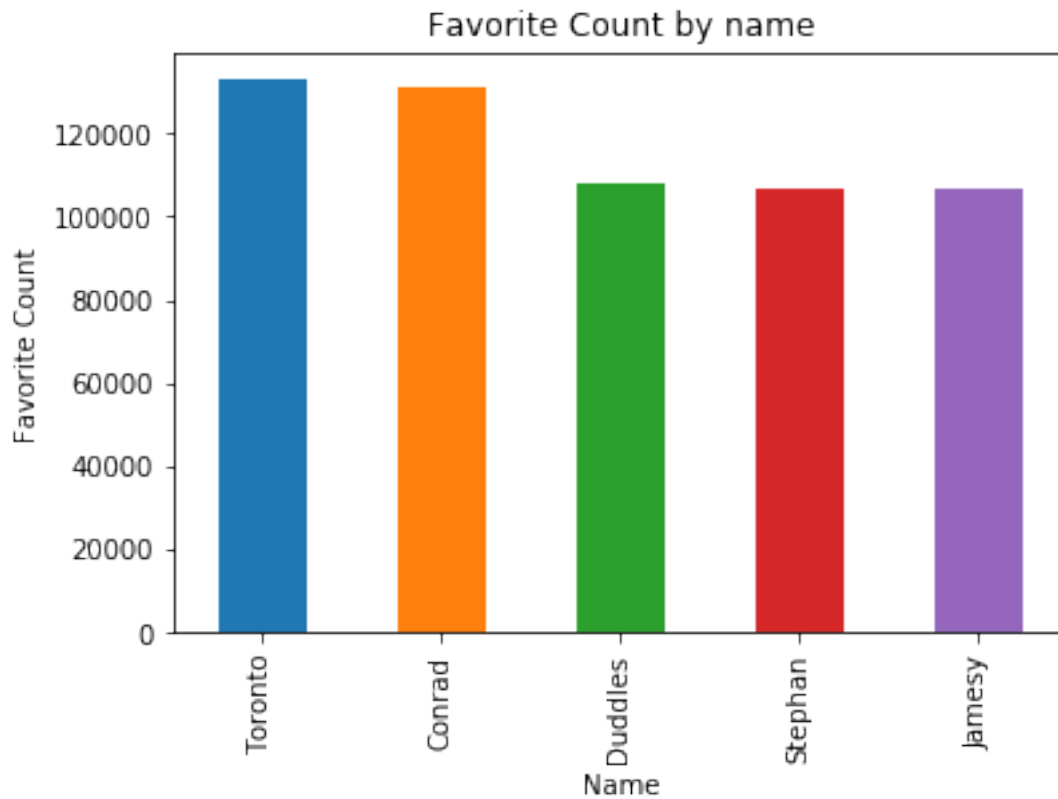

```
In [106]: twitter_archive_master.groupby('name').favorite_count.mean().sort_values(ascending=Fal

Out[106]: name
          Toronto     132810.0
          Conrad      131075.0
          Duddles     107956.0
          Stephan     107015.0
          Jamesy      106827.0
          Name: favorite_count, dtype: float64

In [107]: ax = twitter_archive_master.groupby('name').favorite_count.mean().sort_values(ascendin
          ax.set_xlabel("Name");
          ax.set_ylabel("Favorite Count");
```

Favorite Count by name


```
In [108]: twitter_archive_master.groupby('name').favorite_count.mean().sort_values(ascending=Fal

Out[108]: name
          Cumberbatch    131.0
          Rand           115.0
          Vesuvius       113.0
          Fox            111.0
          Adobe           81.0
          Name: favorite_count, dtype: float64
```