wrangle_act

March 8, 2022

# Project: -Wrangling and Analyze Data- ## Analyzed BY *Samiha Amroune*

1 Introduction:

This is the 4th project provided by **Udacity** for **DNA** students

In this project; The dataset that we will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user [@dog_rates](https://twitter.com/dog_rates), also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

1.0.1 Project Steps Overview

- Gathering data
- Assessing data
- Cleaning data
- Storing data
- Analyzing, and visualizing data

## Data Gathering In the cell below, gather **all** three pieces of data for this project and load them in the notebook. **Note:** the methods required to gather each data are different. 1. Directly download the WeRateDogs Twitter archive data (twitter_archive_enhanced.csv)

[1]:

```
# Import all libraries that we need
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
from bs4 import BeautifulSoup as bs
import requests
import tweepy
from tweepy import OAuthHandler
import json
from timeit import default_timer as timer
%matplotlib inline
```

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.set_option.html?highlight=pd%20set_optio

[2]:
```
#I Used these options to see all dataframe content uncomment them to see more
'''
# set up display area to show dataframe in jupyter qtconsole
#display all text in a cell without truncation
pd.set_option('display.max_rows', None) # set value to ‘None’ value means␣
 ↪unlimited
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
'''
```

[2]:
```
"\n# set up display area to show dataframe in jupyter qtconsole\n#display all
text in a cell without truncation\npd.set_option('display.max_rows', None) # set
value to ‘None’ value means unlimited\npd.set_option('display.max_columns',
None)\npd.set_option('display.width', None)\n"
```

[3]:
```
# Open the csv file
df_twt_arch = pd.read_csv('twitter-archive-enhanced.csv')
df_twt_arch.head()
```

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

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

                                            source  \
0  <a href="http://twitter.com/download/iphone" r…
1  <a href="http://twitter.com/download/iphone" r…
2  <a href="http://twitter.com/download/iphone" r…
3  <a href="http://twitter.com/download/iphone" r…
4  <a href="http://twitter.com/download/iphone" r…

                                              text  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
```

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

   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
```

2. Use the Requests library to download the tweet image prediction (image_predictions.tsv)

[4]:
```
# Download the image prediction file using the link provided to Udacity students
url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/
 ↪599fd2ad_image-predictions/image-predictions.tsv'
image_request = requests.get(url, allow_redirects=True)

open('image_predictions.tsv', 'wb').write(image_request.content)
```

[4]:
```
335079
```

[5]:
```
# read image-predictions.tsv file
df_img = pd.read_csv('image-predictions.tsv', sep='\t')
```

3. Use the Tweepy library to query additional data via the Twitter API (tweet_json.txt)

[6]:
```
# Query Twitter API for each tweet in the Twitter archive and save JSON in a␣
 ↪text file
# These are hidden to comply with Twitter's API terms and conditions
'''
consumer_key = 'Put your keys here'
consumer_secret = 'Put your keys here'
access_token = 'Put your keys here'
access_secret = 'Put your keys here'

auth = OAuthHandler(consumer_key, consumer_secret)
```

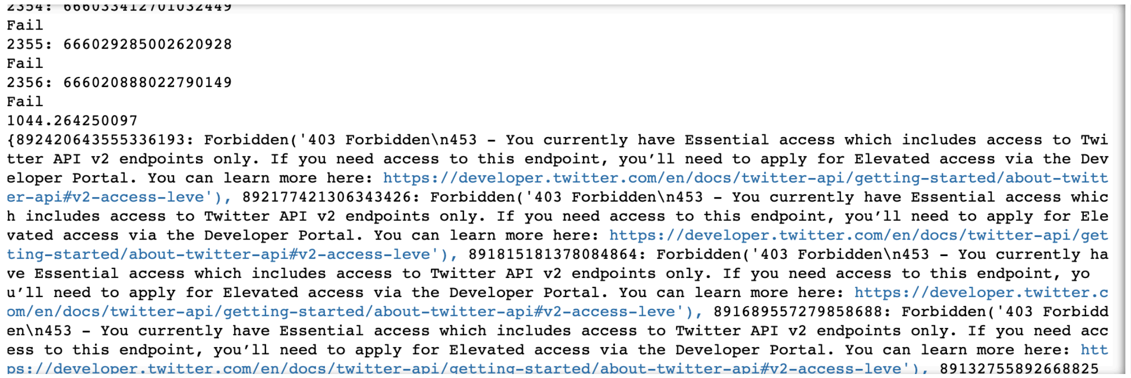
```
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth, wait_on_rate_limit=True)

# NOTE TO STUDENT WITH MOBILE VERIFICATION ISSUES:
# df_1 is a DataFrame with the twitter_archive_enhanced.csv file. You may have␣
 ↪to
# change line 17 to match the name of your DataFrame with␣
 ↪twitter_archive_enhanced.csv
# NOTE TO REVIEWER: this student had mobile verification issues so the following
# Twitter API code was sent to this student from a Udacity instructor
# Tweet IDs for which to gather additional data via Twitter's API
tweet_ids = df_twt_arch.tweet_id.values
len(tweet_ids)

# Query Twitter's API for JSON data for each tweet ID in the Twitter archive
count = 0
fails_dict = {}
start = timer()
# Save each tweet's returned JSON as a new line in a .txt file
with open('tweet_json.txt', 'w') as outfile:
    # This loop will likely take 20-30 minutes to run because of Twitter's rate␣
 ↪limit
    for tweet_id in tweet_ids:
        count += 1
        print(str(count) + ": " + str(tweet_id))
        try:
            tweet = api.get_status(tweet_id, tweet_mode='extended')
            print("Success")
            json.dump(tweet._json, outfile)
            outfile.write('\n')
            #AttributeError: module 'tweepy' has no attribute 'TweepError'
            #https://knowledge.udacity.com/questions/743507
        except Exception as e:
            print("Fail")
            fails_dict[tweet_id] = e
            pass
end = timer()
print(end - start)
print(fails_dict)
'''
```

[6]:

```
'\nconsumer_key = \'Put your keys here\'\nconsumer_secret = \'Put your keys
here\'\naccess_token = \'Put your keys here\'\naccess_secret = \'Put your keys
here\'\n\nauth = OAuthHandler(consumer_key,
consumer_secret)\nauth.set_access_token(access_token, access_secret)\n\napi =
tweepy.API(auth, wait_on_rate_limit=True)\n\n# NOTE TO STUDENT WITH MOBILE
```

```
VERIFICATION ISSUES:\n# df_1 is a DataFrame with the
twitter_archive_enhanced.csv file. You may have to\n# change line 17 to match
the name of your DataFrame with twitter_archive_enhanced.csv\n# NOTE TO
REVIEWER: this student had mobile verification issues so the following\n#
Twitter API code was sent to this student from a Udacity instructor\n# Tweet IDs
for which to gather additional data via Twitter\'s API\ntweet_ids =
df_twt_arch.tweet_id.values\nlen(tweet_ids)\n\n# Query Twitter\'s API for JSON
data for each tweet ID in the Twitter archive\ncount = 0\nfails_dict = {}\nstart
= timer()\n# Save each tweet\'s returned JSON as a new line in a .txt file\nwith
open(\'tweet_json.txt\', \'w\') as outfile:\n    # This loop will likely take
20-30 minutes to run because of Twitter\'s rate limit\n    for tweet_id in
tweet_ids:\n        count += 1\n        print(str(count) + ": " +
str(tweet_id))\n        try:\n            tweet = api.get_status(tweet_id,
tweet_mode=\'extended\')\n            print("Success")\n
json.dump(tweet._json, outfile)\n            outfile.write(\'\n\')\n
#AttributeError: module \'tweepy\' has no attribute \'TweepError\'\n
#https://knowledge.udacity.com/questions/743507\n        except Exception as
e:\n            print("Fail")\n            fails_dict[tweet_id] = e\n
pass\nend = timer()\nprint(end - start)\nprint(fails_dict)\n'
```

I got this issue from tweeter, regarding to the fail status, tweeter rejecting my access to the request API. I will use the file provided by **Udacity**



```
2354: 666033412701032449
Fail
2355: 666029285002620928
Fail
2356: 666020888022790149
Fail
1044.264250097
{892420643555336193: Forbidden('403 Forbidden\n453 - You currently have Essential access which includes access to Twi
tter API v2 endpoints only. If you need access to this endpoint, you'll need to apply for Elevated access via the Dev
eloper Portal. You can learn more here: https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitt
er-api#v2-access-leve'), 892177421306343426: Forbidden('403 Forbidden\n453 - You currently have Essential access whic
h includes access to Twitter API v2 endpoints only. If you need access to this endpoint, you'll need to apply for Ele
vated access via the Developer Portal. You can learn more here: https://developer.twitter.com/en/docs/twitter-api/get
ting-started/about-twitter-api#v2-access-leve'), 891815181378084864: Forbidden('403 Forbidden\n453 - You currently ha
ve Essential access which includes access to Twitter API v2 endpoints only. If you need access to this endpoint, yo
u'll need to apply for Elevated access via the Developer Portal. You can learn more here: https://developer.twitter.c
om/en/docs/twitter-api/getting-started/about-twitter-api#v2-access-leve'), 891689557279858688: Forbidden('403 Forbidd
en\n453 - You currently have Essential access which includes access to Twitter API v2 endpoints only. If you need acc
ess to this endpoint, you'll need to apply for Elevated access via the Developer Portal. You can learn more here: htt
ps://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api#v2-access-leve'), 891327558926688256
```

[7]:

```python
# Create dataframe from tweet-json.txt file
js_data = []

with open('tweet-json.txt', encoding='utf-8') as js_file:
    # loop line by line and append values to the created dataframe

    for line in js_file:
        data = json.loads(line)
        tweet_id = data['id']
        tweet = data['full_text']
        url_tweet = tweet[tweet.find('https'):]
        favorite_count = data['favorite_count']
        retweet_count = data['retweet_count']
        retweet_status = data['retweet_status'] = data.get('retweet_status',␣
 ↪'Original tweet')
        if retweet_status == 'Original tweet':
            url = url_tweet
        else:
            retweeted_status = 'A retweet'
            url = 'A retweet'


        js_data.append({'tweet_id': tweet_id,'favorite_count':␣
 ↪favorite_count,'retweet_status': retweet_status,
                     'url': url,'retweet_count': retweet_count})

# create a new DataFrame

df_js = pd.DataFrame(js_data, columns = ['tweet_id',␣
 ↪'favorite_count','retweet_count', 'retweet_status', 'url'])


df_js.head()

df_js.to_csv('tweet_json.txt', index=False)
```

```
[8]: # Read the dataframe created
     df_json = pd.read_csv('tweet_json.txt')
```

## Assessing Data In this section, detect and document at least **eight (8) quality issues and two (2) tidiness issue**. You must use **both** visual assessment programmatic assessement to assess the data.

Note: pay attention to the following key points when you access the data.

- You only want original ratings (no retweets) that have images. Though there are 5000+ tweets in the dataset, not all are dog ratings and some are retweets.
- Assessing and cleaning the entire dataset completely would require a lot of time, and is not necessary to practice and demonstrate your skills in data wrangling. Therefore, the requirements of this project are only to assess and clean at least 8 quality issues and at least 2 tidiness issues in this dataset.
- The fact that the rating numerators are greater than the denominators does not need to be cleaned. This unique rating system is a big part of the popularity of WeRateDogs.
- You do not need to gather the tweets beyond August 1st, 2017. You can, but note that you won't be able to gather the image predictions for these tweets since you don't have access to the algorithm used.

## Quality Issue

1.0.2 From df_twt_arch :

1 - In columns: (‘doggo’, ‘floofer’, ‘pupper’, ‘puppo’, ‘name’) ‘None’ assigned instead of ‘NaN’ for empty missing data **{visual assessment}** See Solution Noticed it in **2** - columns not needed: (‘in_reply_to_status_id’, ‘in_reply_to_user_id’,‘retweeted_status_id’, ‘retweeted_status_user_id’,‘retweeted_status_timestamp’) - columns (‘source’ ,‘text’,‘name) need to rename to be familliar with users **{visual assessment}** See Solution Noticed it in **3** - column timestamp dtype should be datetime and split into two columns date and time for better visualisation**{programmatic assessment}** See Solution Noticed it in **4** - ’tweet_id’ must be a string.**{programmatic assessment}** See Solution Noticed it in **5** - ‘source’ column contains tag html. **{visual assessment}** See Solution Noticed it in **6** - column ‘name’ has values: ‘None’, ‘a’, ‘O’, ‘Devón’. **{programmatic assessment}** See Solution Noticed it in **7** - expanded_urls has missing value and inccorrect urls **{programmatic assessment}** and **{visual assessment}** See Solution Noticed it in **8** - Rating dinominator must be equal to 10 there are other values: (0, 15, 70, 7, 11, 150, 170, 20, 50, 90, 80, 40, 130, 110, 16, 120, 2) **{programmatic assessment** See Solution Noticed it in

1.0.3 From df_img:

9- The predictions (‘P1’, ‘P2’, ‘P3’) columns are not clear and familiar to the reader and have strange predictions (spatula, barrow, minibus, etc) **{programmatic assessment}** See Solution Noticed it in **10** - Some “tweet_ids” have same “jpg_url”, after checking using the urls: (https://twitter.com/dog_rates/status/803692223237865472) (https://twitter.com/dog_rates/status/691416866452082688) and changing the ids they were the same tweet **{programmatic assessment}** - ids img does not exist “Hmm…this page doesn’t exist. Try searching for something else”: **{visual assessment}** - 759566828574212096 - 802247111496568832 - 851953902622658560 - 842892208864923648 - 861769973181624320 - 873697596434513921 - 888202515573088257 See Solution Noticed it in

1.0.4 From df_json:

11 - Ivalid urls: (https://…)(https:/…) (https:/t.c…) - 175 duplicated url **{programmatic assessment}** See Solution Noticed it in **12**- retweet_status has one value ‘Original tweet’, no need it See Solution Noticed it in **13** - Tweets missing retweet count and favorite count **{programmatic assessment}** See Solution Noticed it in

## Tidiness Issue

1- doggo, floofer, pupper, puppo these 4 variables shoule be combined into one categorical variable ‘dogtionary’. **{visual assessment}** See Solution Noticed it in **2**- rating nominator, rating dinominator should be one column since rating dinominator always be 10 **{visual assessment}** See Solution Noticed it in **3**- Dataframes: twitter_archive, image_predictions, and tweet_json, Should be one df (twitter_master_df) **{visual assessment}** See Solution Noticed it in **4**- in twitter_master_df: expanded_urls and url have same values **{visual assessment}** See Solution Noticed it in

[9]:

```
# Create functions to avoid repeating the same code
# value counts function of each column
def val_count(df):
    for i in df.columns:
        print(i,'\n', df[i].value_counts())

# duplicated function of each column
def duplic(df):
    for i in df.columns:
        print(i,'\n', df[i].duplicated().sum())
```

issue 3 Tidi

[10]:

```
# Display the dataframe and print number of columns and rows
print(df_json.shape)
df_json
```

```
(2354, 5)
```

[10]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
0     892420643555336193           39467           8853  Original tweet
1     892177421306343426           33819           6514  Original tweet
2     891815181378084864           25461           4328  Original tweet
3     891689557279858688           42908           8964  Original tweet
4     891327558926688256           41048           9774  Original tweet
…                    …             …            …               …
2349  666049248165822465             111             41  Original tweet
2350  666044226329800704             311            147  Original tweet
2351  666033412701032449             128             47  Original tweet
2352  666029285002620928             132             48  Original tweet
2353  666020888022790149            2535            532  Original tweet

                          url
0     https://t.co/MgUWQ76dJU
1     https://t.co/0Xxu71qeIV
2     https://t.co/wUnZnhtVJB
3     https://t.co/tD36da7qLQ
4     https://t.co/AtUZn91f7f
…                         …
2349  https://t.co/4B7c0c1EDq
2350  https://t.co/DWnyCjf2mx
2351  https://t.co/y671yMhoiR
2352  https://t.co/r7m0b2m0UI
2353  https://t.co/BLDqew2Ijj

[2354 rows x 5 columns]
```

[11]:
```
# Display the dataframe and print number of columns and rows
print(df_img.shape)
df_img
```

```
(2075, 12)
```

[11]:
```
                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
…                    …                                                …
2070  891327558926688256  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071  891689557279858688  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  891815181378084864  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  892177421306343426  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  892420643555336193  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

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

       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
…          …     …                          …       …       …
2070  0.225770    True  German_short-haired_pointer  0.175219    True
2071  0.168086    True                    spatula  0.040836   False
2072  0.078253    True                     kelpie  0.031379    True
2073  0.090647    True                   papillon  0.068957    True
2074  0.085851   False                     banana  0.076110   False

[2075 rows x 12 columns]
```

issue 1 Quality issue 1 Tidi issue 2 Tidi

[12]:

```
# Display the dataframe and print number of columns and rows
print(df_twt_arch.shape)
df_twt_arch
```

```
(2356, 17)
```

[12]:

```
                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
…                    …                     …                  …
2351  666049248165822465                    NaN                  NaN
2352  666044226329800704                    NaN                  NaN
2353  666033412701032449                    NaN                  NaN
2354  666029285002620928                    NaN                  NaN
2355  666020888022790149                    NaN                  NaN

                       timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
…                          …
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

                                                  source  \
0     <a href="http://twitter.com/download/iphone" r…
1     <a href="http://twitter.com/download/iphone" r…
2     <a href="http://twitter.com/download/iphone" r…
3     <a href="http://twitter.com/download/iphone" r…
4     <a href="http://twitter.com/download/iphone" r…
…                                                 …
2351  <a href="http://twitter.com/download/iphone" r…
2352  <a href="http://twitter.com/download/iphone" r…
2353  <a href="http://twitter.com/download/iphone" r…
2354  <a href="http://twitter.com/download/iphone" r…
2355  <a href="http://twitter.com/download/iphone" r…

                                                    text  retweeted_status_id  \
0     This is Phineas. He's a mystical boy. Only eve…                      NaN
1     This is Tilly. She's just checking pup on you…                      NaN
```

```
2     This is Archie. He is a rare Norwegian Pouncin…                  NaN
3     This is Darla. She commenced a snooze mid meal…                  NaN
4     This is Franklin. He would like you to stop ca…                  NaN
…                                                 …                  …
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
…                        …                        …
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
…                                                 …                …
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
…                  …       …   …     …    …   …
2351                  10      None  None    None   None  None
2352                  10         a  None    None   None  None
2353                  10         a  None    None   None  None
2354                  10         a  None    None   None  None
```

```
2355                    10       None  None    None   None  None

[2356 rows x 17 columns]
```

issue2 Quality issue3 Quality issue4 Quality

```
[13]: # Get for more info
df_twt_arch.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   tweet_id                    2356 non-null   int64
 1   in_reply_to_status_id       78 non-null     float64
 2   in_reply_to_user_id         78 non-null     float64
 3   timestamp                   2356 non-null   object
 4   source                      2356 non-null   object
 5   text                        2356 non-null   object
 6   retweeted_status_id         181 non-null    float64
 7   retweeted_status_user_id    181 non-null    float64
 8   retweeted_status_timestamp  181 non-null    object
 9   expanded_urls               2297 non-null   object
 10  rating_numerator            2356 non-null   int64
 11  rating_denominator          2356 non-null   int64
 12  name                        2356 non-null   object
 13  doggo                       2356 non-null   object
 14  floofer                     2356 non-null   object
 15  pupper                      2356 non-null   object
 16  puppo                       2356 non-null   object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

Back To Quality list

issue7 Quality

```
[14]: df_twt_arch.expanded_urls.unique()
```

```
[14]: array(['https://twitter.com/dog_rates/status/892420643555336193/photo/1',
       'https://twitter.com/dog_rates/status/892177421306343426/photo/1',
       'https://twitter.com/dog_rates/status/891815181378084864/photo/1',
       …,
       'https://twitter.com/dog_rates/status/666033412701032449/photo/1',
       'https://twitter.com/dog_rates/status/666029285002620928/photo/1',
       'https://twitter.com/dog_rates/status/666020888022790149/photo/1'],
      dtype=object)
```

[15]:
```
df_twt_arch['expanded_urls'].isnull().sum()
```

[15]:
```
59
```

[16]:
```
df_twt_arch[df_twt_arch['expanded_urls'].isnull()]
```

[16]:
```
                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
30    886267009285017600           8.862664e+17         2.281182e+09
55    881633300179243008           8.816070e+17         4.738443e+07
64    879674319642796034           8.795538e+17         3.105441e+09
113   870726314365509632           8.707262e+17         1.648776e+07
148   863427515083354112           8.634256e+17         7.759620e+07
179   857214891891077121           8.571567e+17         1.806710e+08
185   856330835276025856                    NaN                  NaN
186   856288084350160898           8.562860e+17         2.792810e+08
188   855862651834028034           8.558616e+17         1.943518e+08
189   855860136149123072           8.558585e+17         1.361572e+07
218   850333567704068097           8.503288e+17         2.195506e+07
228   848213670039564288           8.482121e+17         4.196984e+09
234   847617282490613760           8.476062e+17         4.196984e+09
274   840698636975636481           8.406983e+17         8.405479e+17
290   838150277551247360           8.381455e+17         2.195506e+07
291   838085839343206401           8.380855e+17         2.894131e+09
313   835246439529840640           8.352460e+17         2.625958e+07
342   832088576586297345           8.320875e+17         3.058208e+07
346   831926988323639298           8.319030e+17         2.068372e+07
375   828361771580813312                    NaN                  NaN
387   826598799820865537           8.265984e+17         4.196984e+09
409   823333489516937216           8.233264e+17         1.582854e+09
427   821153421864615936           8.211526e+17         1.132119e+08
498   813130366689148928           8.131273e+17         4.196984e+09
513   811647686436880384           8.116272e+17         4.196984e+09
570   801854953262350336           8.018543e+17         1.185634e+07
576   800859414831898624           8.008580e+17         2.918590e+08
611   797165961484890113           7.971238e+17         2.916630e+07
701   786051337297522688           7.727430e+17         7.305050e+17
707   785515384317313025                    NaN                  NaN
843   766714921925144576           7.667118e+17         4.196984e+09
857   763956972077010945           7.638652e+17         1.584641e+07
967   750381685133418496           7.501805e+17         4.717297e+09
1005  747651430853525504           7.476487e+17         4.196984e+09
1080  738891149612572673           7.384119e+17         3.589728e+08
1295  707983188426153984           7.079801e+17         2.319108e+09
1345  704491224099647488           7.044857e+17         2.878549e+07
1445  696518437233913856                    NaN                  NaN
1446  696490539101908992           6.964887e+17         4.196984e+09
1474  693644216740769793           6.936422e+17         4.196984e+09
```

```
1479  693582294167244802           6.935722e+17         1.198989e+09
1497  692423280028966913           6.924173e+17         4.196984e+09
1523  690607260360429569           6.903413e+17         4.670367e+08
1598  686035780142297088           6.860340e+17         4.196984e+09
1605  685681090388975616           6.855479e+17         4.196984e+09
1618  684969860808454144           6.849598e+17         4.196984e+09
1663  682808988178739200           6.827884e+17         4.196984e+09
1689  681340665377193984           6.813394e+17         4.196984e+09
1774  678023323247357953           6.780211e+17         4.196984e+09
1819  676590572941893632           6.765883e+17         4.196984e+09
1844  675849018447167488           6.758457e+17         4.196984e+09
1895  674742531037511680           6.747400e+17         4.196984e+09
1905  674606911342424069           6.744689e+17         4.196984e+09
1914  674330906434379776           6.658147e+17         1.637468e+07
1940  673716320723169284           6.737159e+17         4.196984e+09
2038  671550332464455680           6.715449e+17         4.196984e+09
2149  669684865554620416           6.693544e+17         4.196984e+09
2189  668967877119254528           6.689207e+17         2.143566e+07
2298  667070482143944705           6.670655e+17         4.196984e+09

                      timestamp  \
30    2017-07-15 16:51:35 +0000
55    2017-07-02 21:58:53 +0000
64    2017-06-27 12:14:36 +0000
113   2017-06-02 19:38:25 +0000
148   2017-05-13 16:15:35 +0000
179   2017-04-26 12:48:51 +0000
185   2017-04-24 02:15:55 +0000
186   2017-04-23 23:26:03 +0000
188   2017-04-22 19:15:32 +0000
189   2017-04-22 19:05:32 +0000
218   2017-04-07 13:04:55 +0000
228   2017-04-01 16:41:12 +0000
234   2017-03-31 01:11:22 +0000
274   2017-03-11 22:59:09 +0000
290   2017-03-04 22:12:52 +0000
291   2017-03-04 17:56:49 +0000
313   2017-02-24 21:54:03 +0000
342   2017-02-16 04:45:50 +0000
346   2017-02-15 18:03:45 +0000
375   2017-02-05 21:56:51 +0000
387   2017-02-01 01:11:25 +0000
409   2017-01-23 00:56:15 +0000
427   2017-01-17 00:33:26 +0000
498   2016-12-25 21:12:41 +0000
513   2016-12-21 19:01:02 +0000
570   2016-11-24 18:28:13 +0000
```

```
576   2016-11-22 00:32:18 +0000
611   2016-11-11 19:55:50 +0000
701   2016-10-12 03:50:17 +0000
707   2016-10-10 16:20:36 +0000
843   2016-08-19 19:14:16 +0000
857   2016-08-12 04:35:10 +0000
967   2016-07-05 17:31:49 +0000
1005  2016-06-28 04:42:46 +0000
1080  2016-06-04 00:32:32 +0000
1295  2016-03-10 17:35:20 +0000
1345  2016-03-01 02:19:31 +0000
1445  2016-02-08 02:18:30 +0000
1446  2016-02-08 00:27:39 +0000
1474  2016-01-31 03:57:23 +0000
1479  2016-01-30 23:51:19 +0000
1497  2016-01-27 19:05:49 +0000
1523  2016-01-22 18:49:36 +0000
1598  2016-01-10 04:04:10 +0000
1605  2016-01-09 04:34:45 +0000
1618  2016-01-07 05:28:35 +0000
1663  2016-01-01 06:22:03 +0000
1689  2015-12-28 05:07:27 +0000
1774  2015-12-19 01:25:31 +0000
1819  2015-12-15 02:32:17 +0000
1844  2015-12-13 01:25:37 +0000
1895  2015-12-10 00:08:50 +0000
1905  2015-12-09 15:09:55 +0000
1914  2015-12-08 20:53:11 +0000
1940  2015-12-07 04:11:02 +0000
2038  2015-12-01 04:44:10 +0000
2149  2015-11-26 01:11:28 +0000
2189  2015-11-24 01:42:25 +0000
2298  2015-11-18 20:02:51 +0000

                                                source  \
30    <a href="http://twitter.com/download/iphone" r…
55    <a href="http://twitter.com/download/iphone" r…
64    <a href="http://twitter.com/download/iphone" r…
113   <a href="http://twitter.com/download/iphone" r…
148   <a href="http://twitter.com/download/iphone" r…
179   <a href="http://twitter.com/download/iphone" r…
185   <a href="http://twitter.com/download/iphone" r…
186   <a href="http://twitter.com/download/iphone" r…
188   <a href="http://twitter.com/download/iphone" r…
189   <a href="http://twitter.com/download/iphone" r…
218   <a href="http://twitter.com/download/iphone" r…
228   <a href="http://twitter.com/download/iphone" r…
```

```
234   <a href="http://twitter.com/download/iphone" r…
274   <a href="http://twitter.com/download/iphone" r…
290   <a href="http://twitter.com/download/iphone" r…
291   <a href="http://twitter.com/download/iphone" r…
313   <a href="http://twitter.com/download/iphone" r…
342   <a href="http://twitter.com/download/iphone" r…
346   <a href="http://twitter.com/download/iphone" r…
375   <a href="http://twitter.com" rel="nofollow">Tw…
387   <a href="http://twitter.com/download/iphone" r…
409   <a href="http://twitter.com/download/iphone" r…
427   <a href="http://twitter.com/download/iphone" r…
498   <a href="http://twitter.com/download/iphone" r…
513   <a href="http://twitter.com/download/iphone" r…
570   <a href="http://twitter.com/download/iphone" r…
576   <a href="http://twitter.com/download/iphone" r…
611   <a href="http://twitter.com/download/iphone" r…
701   <a href="http://twitter.com/download/iphone" r…
707   <a href="http://twitter.com/download/iphone" r…
843   <a href="http://twitter.com/download/iphone" r…
857   <a href="http://twitter.com/download/iphone" r…
967   <a href="http://twitter.com/download/iphone" r…
1005  <a href="http://twitter.com/download/iphone" r…
1080  <a href="http://twitter.com/download/iphone" r…
1295  <a href="http://twitter.com/download/iphone" r…
1345  <a href="http://twitter.com/download/iphone" r…
1445  <a href="http://twitter.com/download/iphone" r…
1446  <a href="http://twitter.com/download/iphone" r…
1474  <a href="http://twitter.com/download/iphone" r…
1479  <a href="http://twitter.com/download/iphone" r…
1497  <a href="http://twitter.com/download/iphone" r…
1523  <a href="http://twitter.com/download/iphone" r…
1598  <a href="http://twitter.com/download/iphone" r…
1605  <a href="http://twitter.com/download/iphone" r…
1618  <a href="http://twitter.com/download/iphone" r…
1663  <a href="http://twitter.com/download/iphone" r…
1689  <a href="http://twitter.com/download/iphone" r…
1774  <a href="http://twitter.com/download/iphone" r…
1819  <a href="http://twitter.com/download/iphone" r…
1844  <a href="http://twitter.com/download/iphone" r…
1895  <a href="http://twitter.com/download/iphone" r…
1905  <a href="http://twitter.com/download/iphone" r…
1914  <a href="http://twitter.com/download/iphone" r…
1940  <a href="http://twitter.com/download/iphone" r…
2038  <a href="http://twitter.com/download/iphone" r…
2149  <a href="http://twitter.com/download/iphone" r…
2189  <a href="http://twitter.com/download/iphone" r…
2298  <a href="http://twitter.com/download/iphone" r…
```

```
                                                    text  retweeted_status_id  \
30    @NonWhiteHat @MayhewMayhem omg hello tanner yo…                  NaN
55    @roushfenway These are good dogs but 17/10 is …                  NaN
64                       @RealKentMurphy 14/10 confirmed                  NaN
113   @ComplicitOwl @ShopWeRateDogs &gt;10/10 is res…                  NaN
148   @Jack_Septic_Eye I'd need a few more pics to p…                  NaN
179                         @Marc_IRL pixelated af 12/10                  NaN
185   RT @Jenna_Marbles: @dog_rates Thanks for ratin…         8.563302e+17
186   @xianmcguire @Jenna_Marbles Kardashians wouldn…                  NaN
188   @dhmontgomery We also gave snoop dogg a 420/10…                  NaN
189   @s8n You tried very hard to portray this good …                  NaN
218   @markhoppus MARK THAT DOG HAS SEEN AND EXPERIE…                  NaN
228   Jerry just apuppologized to me. He said there …                  NaN
234   .@breaannanicolee PUPDATE: Cannon has a heart …                  NaN
274   @0_kelvin_0 &gt;10/10 is reserved for puppos s…                  NaN
290                                   @markhoppus 182/10                  NaN
291   @bragg6of8 @Andy_Pace_ we are still looking fo…                  NaN
313   @jonnysun @Lin_Manuel ok jomny I know you're e…                  NaN
342             @docmisterio account started on 11/15/15                  NaN
346                               @UNC can confirm 12/10                  NaN
375   Beebop and Doobert should start a band 12/10 w…                  NaN
387   I was going to do 007/10, but the joke wasn't …                  NaN
409                                 @HistoryInPics 13/10                  NaN
427   @imgur for a polar bear tho I'd say 13/10 is a…                  NaN
498   I've been informed by multiple sources that th…                  NaN
513   PUPDATE: I've been informed that Augie was act…                  NaN
570   .@NBCSports OMG THE TINY HAT I'M GOING TO HAVE…                  NaN
576   @SkyWilliams doggo simply protecting you from …                  NaN
611   @JODYHiGHROLLER it may be an 11/10 but what do…                  NaN
701              13/10 for breakdancing puppo @shibbnbot                  NaN
707       Today, 10/10, should be National Dog Rates Day                  NaN
843   His name is Charley and he already has a new s…                  NaN
857   @TheEllenShow I'm not sure if you know this bu…                  NaN
967                  13/10 such a good doggo\n@spaghemily                  NaN
1005  Other pupper asked not to have his identity sh…                  NaN
1080                                  @mount_alex3 13/10                  NaN
1295  @serial @MrRoles OH MY GOD I listened to all o…                  NaN
1345                                  13/10 hero af\n@ABC                  NaN
1445     Oh my god 10/10 for every little hot dog pupper                  NaN
1446  After reading the comments I may have overesti…                  NaN
1474  BREAKING PUPDATE: I've just been notified that…                  NaN
1479  Personally I'd give him an 11/10. Not sure why…                  NaN
1497  PUPDATE: just noticed this dog has some extra …                  NaN
1523                               12/10 @LightningHoltt                  NaN
1598  Yes I do realize a rating of 4/20 would've bee…                  NaN
1605  Jack deserves another round of applause. If yo…                  NaN
```

```
1618  For those who claim this is a goat, u are wron…                  NaN
1663  I'm aware that I could've said 20/16, but here…                  NaN
1689  I've been told there's a slight possibility he…                  NaN
1774  After getting lost in Reese's eyes for several…                  NaN
1819  After some outrage from the crowd. Bubbles is …                  NaN
1844  This dog is being demoted to a 9/10 for not we…                  NaN
1895  Some clarification is required. The dog is sin…                  NaN
1905  The 13/10 also takes into account this impecca…                  NaN
1914                                       13/10\n@ABC7                  NaN
1940  The millennials have spoken and we've decided …                  NaN
2038  After 22 minutes of careful deliberation this …                  NaN
2149  After countless hours of research and hundreds…                  NaN
2189                    12/10 good shit Bubka\n@wane15                  NaN
2298  After much debate this dog is being upgraded t…                  NaN

      retweeted_status_user_id retweeted_status_timestamp expanded_urls  \
30                         NaN                        NaN           NaN
55                         NaN                        NaN           NaN
64                         NaN                        NaN           NaN
113                        NaN                        NaN           NaN
148                        NaN                        NaN           NaN
179                        NaN                        NaN           NaN
185                 66699013.0  2017-04-24 02:13:14 +0000           NaN
186                        NaN                        NaN           NaN
188                        NaN                        NaN           NaN
189                        NaN                        NaN           NaN
218                        NaN                        NaN           NaN
228                        NaN                        NaN           NaN
234                        NaN                        NaN           NaN
274                        NaN                        NaN           NaN
290                        NaN                        NaN           NaN
291                        NaN                        NaN           NaN
313                        NaN                        NaN           NaN
342                        NaN                        NaN           NaN
346                        NaN                        NaN           NaN
375                        NaN                        NaN           NaN
387                        NaN                        NaN           NaN
409                        NaN                        NaN           NaN
427                        NaN                        NaN           NaN
498                        NaN                        NaN           NaN
513                        NaN                        NaN           NaN
570                        NaN                        NaN           NaN
576                        NaN                        NaN           NaN
611                        NaN                        NaN           NaN
701                        NaN                        NaN           NaN
707                        NaN                        NaN           NaN
843                        NaN                        NaN           NaN
```

```
857                        NaN                        NaN           NaN
967                        NaN                        NaN           NaN
1005                       NaN                        NaN           NaN
1080                       NaN                        NaN           NaN
1295                       NaN                        NaN           NaN
1345                       NaN                        NaN           NaN
1445                       NaN                        NaN           NaN
1446                       NaN                        NaN           NaN
1474                       NaN                        NaN           NaN
1479                       NaN                        NaN           NaN
1497                       NaN                        NaN           NaN
1523                       NaN                        NaN           NaN
1598                       NaN                        NaN           NaN
1605                       NaN                        NaN           NaN
1618                       NaN                        NaN           NaN
1663                       NaN                        NaN           NaN
1689                       NaN                        NaN           NaN
1774                       NaN                        NaN           NaN
1819                       NaN                        NaN           NaN
1844                       NaN                        NaN           NaN
1895                       NaN                        NaN           NaN
1905                       NaN                        NaN           NaN
1914                       NaN                        NaN           NaN
1940                       NaN                        NaN           NaN
2038                       NaN                        NaN           NaN
2149                       NaN                        NaN           NaN
2189                       NaN                        NaN           NaN
2298                       NaN                        NaN           NaN

      rating_numerator  rating_denominator  name  doggo floofer  pupper  puppo
30                  12                  10  None   None    None    None   None
55                  17                  10  None   None    None    None   None
64                  14                  10  None   None    None    None   None
113                 10                  10  None   None    None    None   None
148                 12                  10  None   None    None    None   None
179                 12                  10  None   None    None    None   None
185                 14                  10  None   None    None    None   None
186                 14                  10  None   None    None    None   None
188                420                  10  None   None    None    None   None
189                666                  10  None   None    None    None   None
218                 13                  10  None   None    None    None   None
228                 11                  10  None   None    None    None   None
234                 13                  10  None   None    None    None   None
274                 10                  10  None   None    None    None   None
290                182                  10  None   None    None    None   None
291                 15                  10  None   None    None    None   None
313                960                   0  None   None    None    None   None
```

```
342                 11                  15  None   None    None    None   None
346                 12                  10  None   None    None    None   None
375                 12                  10  None   None    None    None   None
387                  7                  10  None   None    None    None   None
409                 13                  10  None   None    None    None   None
427                 13                  10  None   None    None    None   None
498                 12                  10  None   None    None    None   None
513                 11                  10  None   None    None    None   None
570                 11                  10  None   None    None    None   None
576                 11                  10  None  doggo    None    None   None
611                 11                  10  None   None    None    None   None
701                 13                  10  None   None    None    None  puppo
707                 10                  10  None   None    None    None   None
843                 13                  10  None   None    None    None   None
857                 12                  10  None  doggo    None    None   None
967                 13                  10  None  doggo    None    None   None
1005                12                  10  None   None    None  pupper   None
1080                13                  10  None   None    None    None   None
1295                12                  10  None   None    None    None   None
1345                13                  10  None   None    None    None   None
1445                10                  10  None   None    None  pupper   None
1446                 1                  10  None   None    None    None   None
1474                10                  10  None   None    None    None   None
1479                11                  10  None   None    None    None   None
1497                 9                  10  None   None    None    None   None
1523                12                  10  None   None    None    None   None
1598                 4                  20  None   None    None    None   None
1605                14                  10  None   None    None    None   None
1618                 5                  10  None   None    None    None   None
1663                20                  16  None   None    None    None   None
1689                 5                  10  None   None    None    None   None
1774                13                  10  None   None    None    None   None
1819                 7                  10  None   None    None    None   None
1844                 9                  10  None   None    None    None   None
1895                11                  10  None   None    None    None   None
1905                13                  10  None   None    None    None   None
1914                13                  10  None   None    None    None   None
1940                 1                  10  None   None    None    None   None
2038                 1                  10  None   None    None    None   None
2149                11                  10  None   None    None    None   None
2189                12                  10  None   None    None    None   None
2298                10                  10  None   None    None    None   None
```

Back To Quality list

```
[17]: # Get for more info
      df_json.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2354 entries, 0 to 2353
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   tweet_id        2354 non-null   int64
 1   favorite_count  2354 non-null   int64
 2   retweet_count   2354 non-null   int64
 3   retweet_status  2354 non-null   object
 4   url             2354 non-null   object
dtypes: int64(3), object(2)
memory usage: 92.1+ KB
```

[18]:

```
# Get for more info
df_img.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   tweet_id  2075 non-null   int64
 1   jpg_url   2075 non-null   object
 2   img_num   2075 non-null   int64
 3   p1        2075 non-null   object
 4   p1_conf   2075 non-null   float64
 5   p1_dog    2075 non-null   bool
 6   p2        2075 non-null   object
 7   p2_conf   2075 non-null   float64
 8   p2_dog    2075 non-null   bool
 9   p3        2075 non-null   object
 10  p3_conf   2075 non-null   float64
 11  p3_dog    2075 non-null   bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

[19]:

```
# Check duplicates
df_twt_arch.duplicated().sum()
```

[19]: 0

[20]:

```
# Check duplicates
df_json.duplicated().sum()
```

[20]: 0

[21]:

```
# Check duplicates
df_img.duplicated().sum()
```

[21]:

```
0
```

[22]:

```
# Get a descriptive statistics summary
df_twt_arch.describe()
```

[22]:

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

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

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

[23]:

```
# Get a descriptive statistics summary
df_json.describe()
```

[23]:

```
           tweet_id  favorite_count  retweet_count
count  2.354000e+03     2354.000000    2354.000000
mean   7.426978e+17     8080.968564    3164.797366
std    6.852812e+16    11814.771334    5284.770364
min    6.660209e+17        0.000000       0.000000
25%    6.783975e+17     1415.000000     624.500000
50%    7.194596e+17     3603.500000    1473.500000
75%    7.993058e+17    10122.250000    3652.000000
max    8.924206e+17   132810.000000   79515.000000
```

[24]:
```
# Get a descriptive statistics summary
df_img.describe()
```

[24]:
```
           tweet_id      img_num      p1_conf       p2_conf       p3_conf
count  2.075000e+03  2075.000000  2075.000000  2.075000e+03  2.075000e+03
mean   7.384514e+17     1.203855     0.594548  1.345886e-01  6.032417e-02
std    6.785203e+16     0.561875     0.271174  1.006657e-01  5.090593e-02
min    6.660209e+17     1.000000     0.044333  1.011300e-08  1.740170e-10
25%    6.764835e+17     1.000000     0.364412  5.388625e-02  1.622240e-02
50%    7.119988e+17     1.000000     0.588230  1.181810e-01  4.944380e-02
75%    7.932034e+17     1.000000     0.843855  1.955655e-01  9.180755e-02
max    8.924206e+17     4.000000     1.000000  4.880140e-01  2.734190e-01
```

[25]:
```
# check value_counts in each column in df_json
val_count(df_json)
```

```
tweet_id
 892420643555336193    1
687102708889812993    1
687826841265172480    1
687818504314159109    1
687807801670897665    1
                     ..
774757898236878852    1
774639387460112384    1
774314403806253056    1
773985732834758656    1
666020888022790149    1
Name: tweet_id, Length: 2354, dtype: int64
favorite_count
 0       179
2706      3
2918      3
610       3
2768      3
       …
7493      1
6671      1
4484      1
2872      1
2535      1
Name: favorite_count, Length: 2007, dtype: int64
retweet_count
 1972    5
3652    5
83      5
1207    4
336     4
```

```
       ..
3357    1
3018    1
2181    1
848     1
147     1
Name: retweet_count, Length: 1724, dtype: int64
retweet_status
 Original tweet    2354
Name: retweet_status, dtype: int64
url
 https:/…                  17
0                         17
https://…                 15
…                         11
https://t.c…               7
                          ..
https://t.co/HmwrdfEfUE    1
https://t.co/uPUAotqQtM    1
https://t.co/rfWQ4T9iQj    1
https://t.co/9T2h14yn4Q    1
https://t.co/BLDqew2Ijj    1
Name: url, Length: 2179, dtype: int64
```

issue11 Quality

[26]:
```
# Get list of duplicated urls
np.array(df_json[df_json['url'].duplicated()])
```

[26]:
```
array([[873337748698140672, 0, 1667, 'Original tweet', 'https:/…'],
       [870726314365509632, 121, 3, 'Original tweet', 's'],
       [869988702071779329, 0, 25661, 'Original tweet', '…'],
       [868880397819494401, 55098, 12518, 'Original tweet',
        'https://t.co/r28jFx9uyF'],
       [868639477480148993, 0, 2240, 'Original tweet', 'https://…'],
       [863471782782697472, 0, 2687, 'Original tweet', 'https://…'],
       [860177593139703809, 0, 33421, 'Original tweet', 'https://t.c…'],
       [858860390427611136, 0, 8805, 'Original tweet', '…'],
       [857214891891077121, 242, 20, 'Original tweet', '0'],
       [856602993587888130, 0, 11633, 'Original tweet', '…'],
       [856288084350160898, 545, 17, 'Original tweet', '0'],
       [855862651834028034, 320, 27, 'Original tweet', 'h'],
       [855860136149123072, 4407, 1058, 'Original tweet', '0'],
       [855245323840757760, 0, 6568, 'Original tweet', '…'],
       [855138241867124737, 0, 50, 'Original tweet', '…'],
       [852936405516943360, 0, 2243, 'Original tweet', 'https://t…'],
       [847978865427394560, 0, 3652, 'Original tweet', 'https://t.c…'],
       [847617282490613760, 7558, 468, 'Original tweet', '0'],
       [845459076796616705, 0, 15071, 'Original tweet', 'https://…'],
       [845098359547420673, 0, 8640, 'Original tweet', 'https://t…'],
       [840761248237133825, 0, 1972, 'Original tweet', 'https:/…'],
       [840728873075638272, 0, 18285, 'Original tweet', 'https://…'],
       [840698636975636481, 197, 3, 'Original tweet', 'n'],
       [838150277551247360, 1824, 370, 'Original tweet', '0'],
       [838085839343206401, 150, 0, 'Original tweet', '0'],
       [836648853927522308, 0, 626, 'Original tweet', '…'],
       [836397794269200385, 0, 31314, 'Original tweet', 'https:/…'],
       [835685285446955009, 0, 9327, 'Original tweet', 'https://t.co/…'],
       [835536468978302976, 0, 1939, 'Original tweet', 'https://…'],
       [834477809192075265, 0, 12224, 'Original tweet', '…'],
       [833732339549220864, 0, 251, 'Original tweet', '…'],
       [831926988323639298, 369, 39, 'Original tweet', '0'],
       [830583320585068544, 73397, 19297, 'Original tweet',
        'https://t.co/SATN4If5H5'],
       [830173239259324417, 0, 6554, 'Original tweet', '…'],
       [829374341691346946, 38074, 10706, 'Original tweet',
        'https://t.co/xo7FZFIAao'],
       [828801551087042563, 0, 3901, 'Original tweet', 'https…'],
       [828361771580813312, 2408, 195, 'Original tweet', 'n'],
       [827228250799742977, 0, 13517, 'Original tweet', 'https://…'],
       [826958653328592898, 23767, 5757, 'Original tweet',
        'https://t.co/QV5nx6otZR'],
       [826615380357632002, 0, 4521, 'Original tweet', 'https:/…'],
       [825120256414846976, 0, 1483, 'Original tweet', 'https:/…'],
       [823333489516937216, 777, 31, 'Original tweet', '0'],
       [822489057087389700, 20083, 7390, 'Original tweet',
        'https://t.co/liJGwMp17E'],
       [822244816520155136, 38832, 11421, 'Original tweet',
        'https://t.co/rLi83ZyCL5'],
       [822163064745328640, 0, 5571, 'Original tweet', 'https:…'],
       [820494788566847489, 0, 648, 'Original tweet', 'https://t…'],
       [820013781606658049, 0, 1369, 'Original tweet', 'https://…'],
       [819006400881917954, 49960, 21794, 'Original tweet',
        'https://t.co/YOC1fHFCSb'],
       [819004803107983360, 95450, 42228, 'Original tweet',
        'https://t.co/AdPKrI8BZ1'],
       [817502432452313088, 0, 3917, 'Original tweet', 'https:/…'],
       [817423860136083457, 38260, 17504, 'Original tweet',
        'https://t.co/btzf1zTDeQ'],
       [816062466425819140, 0, 1207, 'Original tweet', 'https://…'],
       [815966073409433600, 25057, 9907, 'Original tweet',
        'https://t.co/MiMKtsLN6k'],
       [815745968457060357, 0, 4797, 'Original tweet', '…'],
       [814578408554463233, 0, 6750, 'Original tweet', 'https:/…'],
       [813130366689148928, 4968, 513, 'Original tweet', '0'],
       [812747805718642688, 0, 8067, 'Original tweet', 'https://t.c…'],
       [811647686436880384, 6215, 872, 'Original tweet', '0'],
       [808344865868283904, 47281, 24069, 'Original tweet',
        'https://t.co/hfi264QwYM'],
       [807106840509214720, 107015, 56625, 'Original tweet',
        'https://t.co/DkBYaCAg2d'],
       [806629075125202948, 75639, 37911, 'Original tweet',
        'https://t.co/f1h2Fsalv9'],
       [806620845233815552, 0, 6426, 'Original tweet', 'https://t.c…'],
       [802185808107208704, 0, 314, 'Original tweet', 'https:/…'],
       [801167903437357056, 27386, 6961, 'Original tweet',
        'https://t.co/R6wJyC2Tey'],
       [801127390143516673, 0, 4627, 'Original tweet', 'https://…'],
       [800859414831898624, 778, 113, 'Original tweet', 's'],
       [800188575492947969, 0, 4581, 'Original tweet', 'https://…'],
       [800141422401830912, 17092, 2980, 'Original tweet',
        'https://t.co/pbKOesr8Tg'],
       [798705661114773508, 0, 7724, 'Original tweet', 'https://…'],
       [798701998996647937, 0, 9126, 'Original tweet', 'https://t…'],
       [798585098161549313, 0, 6618, 'Original tweet', 'https://t.…'],
       [798576900688019456, 0, 6871, 'Original tweet', 'https://t.co/…'],
       [796904159865868288, 0, 10411, 'Original tweet', 'https:/…'],
       [796149749086875649, 36177, 16628, 'Original tweet',
        'https://t.co/y5KarNXWXt'],
       [795076730285391872, 18139, 6288, 'Original tweet',
        'https://t.co/ivPWhspN3E'],
       [793962221541933056, 18910, 5711, 'Original tweet',
        'https://t.co/xIfrsMNLBR'],
       [793286476301799424, 27597, 10723, 'Original tweet',
        'https://t.co/rLHGx3VAF3'],
       [791672322847637504, 13129, 3661, 'Original tweet',
        'https://t.co/O5IUmRHRIh'],
       [790946055508652032, 18601, 5496, 'Original tweet',
        'https://t.co/dgvTPZ9tgI'],
       [790277117346975746, 14081, 3732, 'Original tweet',
        'https://t.co/aI7umZHIq7'],
       [789986466051088384, 10369, 2704, 'Original tweet',
        'https://t.co/Ln89R4FP7v'],
       [789960241177853952, 0, 4839, 'Original tweet', 'https://t.…'],
       [789530877013393408, 13188, 3942, 'Original tweet',
        'https://t.co/TU4sZogVIq'],
       [789280767834746880, 0, 5747, 'Original tweet', 'https://t…'],
       [788765914992902144, 30658, 12014, 'Original tweet',
        'https://t.co/x5gXRS1abq'],
       [786729988674449408, 0, 2827, 'Original tweet', 'https://t.…'],
       [786051337297522688, 1614, 179, 'Original tweet', 't'],
       [784826020293709826, 11310, 3712, 'Original tweet',
        'https://t.co/AsykOwMrXQ'],
       [784183165795655680, 22513, 9374, 'Original tweet',
        'https://t.co/t0cEeRbcXJ'],
       [784057939640352768, 33505, 12953, 'Original tweet',
        'https://t.co/T7Jr4Gw4sC'],
       [783334639985389568, 32651, 13616, 'Original tweet',
        'https://t.co/nfprKAXqwu'],
       [782969140009107456, 26949, 8521, 'Original tweet',
        'https://t.co/vOBV1ZqVcX'],
       [782722598790725632, 19250, 6238, 'Original tweet',
        'https://t.co/zrMVdfFej6'],
       [780601303617732608, 13525, 3839, 'Original tweet',
        'https://t.co/mH0IOyFdIG'],
       [780092040432480260, 0, 2655, 'Original tweet', 'https://t.c…'],
       [779834332596887552, 21252, 8237, 'Original tweet',
        'https://t.co/x0gV2Ck3AD'],
       [779124354206535695, 0, 21324, 'Original tweet', 'https://t.co/…'],
       [779056095788752897, 16500, 5247, 'Original tweet',
        'https://t.co/I1bJUzUEW5'],
       [777953400541634568, 0, 4080, 'Original tweet', 'https://…'],
       [777684233540206592, 12518, 3403, 'Original tweet',
        'https://t.co/SuXcI9B7pQ'],
       [776249906839351296, 0, 10673, 'Original tweet', 'https://…'],
       [776113305656188928, 13102, 5068, 'Original tweet',
        'https://t.co/FYT9QBEnKG'],
       [775096608509886464, 0, 9220, 'Original tweet', 'https:/…'],
       [775085132600442880, 17281, 5488, 'Original tweet',
        'https://t.co/wdgY9oHPrT'],
       [773547596996571136, 24553, 7126, 'Original tweet',
        'https://t.co/nBJ5RE4yHb'],
       [773336787167145985, 0, 5912, 'Original tweet', 'https:/…'],
       [771770456517009408, 13356, 3924, 'Original tweet',
        'https://t.co/E6bGWf9EOm'],
       [770743923962707968, 0, 52360, 'Original tweet', 'https://t…'],
       [769940425801170949, 34948, 11131, 'Original tweet',
        'https://t.co/guDkLrX8zV'],
       [768909767477751808, 0, 3129, 'Original tweet', 'https:/…'],
       [768554158521745409, 0, 6719, 'Original tweet', 'https://t.c…'],
       [766864461642756096, 0, 6521, 'Original tweet', 'https:/…'],
       [766714921925144576, 2872, 438, 'Original tweet', 'y'],
       [765222098633691136, 12902, 3914, 'Original tweet',
        'https://t.co/eSvz4EapHH'],
       [763956972077010945, 812, 61, 'Original tweet', '0'],
       [763837565564780549, 14041, 4858, 'Original tweet',
        'https://t.co/gx3xF5mZbo'],
       [763167063695355904, 0, 3484, 'Original tweet', 'https:/…'],
       [761004547850530816, 12482, 3952, 'Original tweet',
        'https://t.co/1acHQS8rvK'],
       [758099635764359168, 21302, 11550, 'Original tweet',
        'https://t.co/iy7ylyZcsE'],
       [753298634498793472, 0, 6620, 'Original tweet', 'https:/…'],
       [753039830821511168, 41080, 24013, 'Original tweet',
        'https://t.co/dsv4xAtfv7'],
       [752932432744185856, 13970, 7798, 'Original tweet',
        'https://t.co/v5m2bIukXc'],
       [750429297815552001, 14569, 4947, 'Original tweet',
        'https://t.co/lprdOylVpS'],
       [750381685133418496, 758, 39, 'Original tweet', 'y'],
       [747242308580548608, 0, 3257, 'Original tweet', 'https://…'],
       [746757706116112384, 10520, 4432, 'Original tweet',
        'https://t.co/bROYbceZ1u'],
       [746521445350707200, 0, 1110, 'Original tweet', 'https:/…'],
       [742423170473463808, 10812, 4319, 'Original tweet',
        'https://t.co/BXIuvkQO9b'],
       [741067306818797568, 10342, 3520, 'Original tweet',
        'https://t.co/o5J479bZUC'],
       [739544079319588864, 43694, 24319, 'Original tweet',
        'https://t.co/zxw8d5g94P'],
       [738891149612572673, 115, 6, 'Original tweet', '0'],
       [736392552031657984, 19450, 8407, 'Original tweet',
        'https://t.co/u1ulQ5heLX'],
       [733109485275860992, 44619, 17621, 'Original tweet',
        'https://t.co/pG2inLaOda'],
       [732005617171337216, 16324, 6154, 'Original tweet',
        'https://t.co/ghyT4Ubk1r'],
       [725842289046749185, 7691, 2970, 'Original tweet',
        'https://t.co/vtVHtKFtBH'],
       [712809025985978368, 20378, 7602, 'Original tweet',
        'https://t.co/HVGs5aAKAn'],
       [711694788429553666, 35865, 20500, 'Original tweet',
        'https://t.co/Ezu6jQrKAZ'],
       [707610948723478529, 18557, 7236, 'Original tweet',
        'https://t.co/oHKJHghrp5'],
       [706904523814649856, 15961, 8830, 'Original tweet',
        'https://t.co/3PoKf6cw7f'],
       [704491224099647488, 840, 97, 'Original tweet', 'C'],
       [703041949650034688, 28996, 14198, 'Original tweet',
        'https://t.co/iHe6liLwWR'],
       [701214700881756160, 13475, 5812, 'Original tweet',
        'https://t.co/JKXh0NbBNL'],
       [696490539101908992, 1339, 170, 'Original tweet', 'e'],
       [692423280028966913, 1618, 265, 'Original tweet', '0'],
       [691416866452082688, 21253, 8689, 'Original tweet',
        'https://t.co/fR1P3g5I6k'],
       [690607260360429569, 313, 19, 'Original tweet', 't'],
       [686035780142297088, 1274, 138, 'Original tweet', 'g'],
       [685325112850124800, 10471, 4535, 'Original tweet',
        'https://t.co/VGPHOTfESw'],
       [684830982659280897, 38551, 24514, 'Original tweet',
        'https://t.co/e3LPMAHj7p'],
       [683515932363329536, 8058, 3291, 'Original tweet',
        'https://t.co/XQ1tRUmO3z'],
       [682808988178739200, 1953, 210, 'Original tweet', 'e'],
       [682088079302213632, 20108, 11271, 'Original tweet',
        'https://t.co/4Kn9GEHXiE'],
       [681340665377193984, 1803, 313, 'Original tweet', 'e'],
       [679828447187857408, 39726, 15839, 'Original tweet',
        'https://t.co/hyAC5Hq9GC'],
       [679158373988876288, 23568, 9193, 'Original tweet',
        'https://t.co/NLOHmlENag'],
       [679062614270468097, 18712, 9299, 'Original tweet',
        'https://t.co/Xl37lQEWd0'],
       [678023323247357953, 2095, 439, 'Original tweet', 'O'],
       [676590572941893632, 1001, 142, 'Original tweet', 'u'],
       [675849018447167488, 1027, 172, 'Original tweet', 'u'],
       [675501075957489664, 18482, 6485, 'Original tweet',
        'https://t.co/WId5lxNdPH'],
       [675354435921575936, 35178, 18963, 'Original tweet',
        'https://t.co/Bb3xnpsWBC'],
       [674742531037511680, 527, 55, 'Original tweet', 'u'],
       [674606911342424069, 1013, 115, 'Original tweet', 's'],
       [674291837063053312, 15817, 6533, 'Original tweet',
        'https://t.co/2S6oye3XqK'],
       [673716320723169284, 3289, 843, 'Original tweet', 'u'],
       [673320132811366400, 14441, 8705, 'Original tweet',
        'https://t.co/iJABMhVidf'],
       [671896809300709376, 9016, 4519, 'Original tweet',
        'https://t.co/eqmkr2gmPH'],
       [671550332464455680, 967, 229, 'Original tweet', 's'],
       [670444955656130560, 7120, 2153, 'Original tweet',
        'https://t.co/J5Mqn8VeYq'],
       [670319130621435904, 4110, 1359, 'Original tweet',
        'https://t.co/QnfO7HEQGA'],
       [669684865554620416, 551, 99, 'Original tweet', 'O'],
       [668967877119254528, 161, 25, 'Original tweet', '5'],
       [668587383441514497, 1760, 1174, 'Original tweet',
        'https://t.co/roqIxCvEB3'],
       [667509364010450944, 7148, 2272, 'Original tweet',
        'https://t.co/0aDF62KVP7'],
       [667070482143944705, 88, 14, 'Original tweet', 'O']], dtype=object)
```

2 Duplicated url:

“…”, “s”, “0”, “5”, “h”, “n”, “e”, “g”, “y”, “t”, “https://…”, “https://t.c…”, “https://t.c…”, “https://…”, “https://t.co/…”, ‘https://t.co/R6wJyC2Tey’, ‘https://t.co/pbKOesr8Tg’, ‘https://t.co/y5KarNXWXt’, ‘https://t.co/SATN4If5H5’, ‘https://t.co/xo7FZFIAao’, ‘https://t.co/QV5nx6otZR’, ‘https://t.co/Xl37lQEWd0’, ‘https://t.co/WId5lxNdPH’, ‘https://t.co/Bb3xnpsWBC’, ‘https://t.co/2S6oye3XqK’, ‘https://t.co/iJABMhVidf’, ‘https://t.co/eqmkr2gmPH’, ‘https://t.co/J5Mqn8VeYq’, ‘https://t.co/QnfO7HEQGA’, ‘https://t.co/0aDF62KVP7’, ‘https://t.co/HVGs5aAKAn’, ‘https://t.co/Ezu6jQrKAZ’, ‘https://t.co/oHKJHghrp5’, ‘https://t.co/iHe6liLwWR’, ‘https://t.co/JKXh0NbBNL’, ‘https://t.co/fR1P3g5I6k’, ‘https://t.co/VGPH0TfESw’, ‘https://t.co/e3LPMAHj7p’, ‘https://t.co/zxw8d5g94P’, ‘https://t.co/u1ulQ5heLX’, ‘https://t.co/pG2inLaOda’, ‘https://t.co/ghyT4Ubk1r’, ‘https://t.co/vtVHtKFtBH’, ‘https://t.co/XQ1tRUmO3z’, ‘https://t.co/4Kn9GEHXiE’, ‘https://t.co/hyAC5Hq9GC’, ‘https://t.co/NLOHmlENag’, ‘https://t.co/o5J479bZUC’, ‘https://t.co/BXIuvkQO9b’, ‘https://t.co/bROYbceZ1u’, ‘https://t.co/lprdOylVpS’, ‘https://t.co/v5m2bIukXc’, ‘https://t.co/dsv4xAtfv7’, ‘https://t.co/iy7ylyZcsE’, ‘https://t.co/1acHQS8rvK’, ‘https://t.co/SuXcI9B7pQ’, ‘https://t.co/gx3xF5mZbo’, ‘https://t.co/eSvz4EapHH’, ‘https://t.co/guDkLrX8zV’, ‘https://t.co/E6bGWf9EOm’, ‘https://t.co/nBJ5RE4yHb’, ‘https://t.co/wdgY9oHPrT’, ‘https://t.co/I1bJUzUEW5’, ‘https://t.co/AsykOwMrXQ’, ‘https://t.co/t0cEeRbcXJ’, ‘https://t.co/T7Jr4Gw4sC’, ‘https://t.co/nfprKAXqwu’, ‘https://t.co/zrMVdfFej6’, ‘https://t.co/mH0IOyFdIG’, ‘https://t.co/x0gV2Ck3AD’, ‘https://t.co/xIfrsMNLBR’, ‘https://t.co/rLHGx3VAF3’, ‘https://t.co/O5IUmRHRIh’, ‘https://t.co/dgvTPZ9tgI’, ‘https://t.co/aI7umZHIq7’, ‘https://t.co/Ln89R4FP7v’, ‘https://t.co/TU4sZogVIq’, ‘https://t.co/x5gXRS1abq’, ‘https://t.co/r28jFx9uyF’, ‘https://t.co/liJGwMp17E’, ‘https://t.co/rLi83ZyCL5’, ‘https://t.co/YOC1fHFCSb’, ‘https://t.co/AdPKrI8BZ1’, ‘https://t.co/btzf1zTDeQ’, ‘https://t.co/MiMKtsLN6k’, ‘https://t.co/hfi264QwYM’, ‘https://t.co/DkBYaCAg2d’, ‘https://t.co/f1h2Fsalv9’

Next i will check each link using ‘https://twitter.com/dog_rates/status/tweet_id’ to figure what kinds of these tweets

```
[27]: df_json.query('url == "…"')
```

```
[27]:               tweet_id  favorite_count  retweet_count  retweet_status url
      100  872668790621863937               0             31  Original tweet   …
      117  869988702071779329               0          25661  Original tweet   …
      170  858860390427611136               0           8805  Original tweet   …
      181  856602993587888130               0          11633  Original tweet   …
      193  855245323840757760               0           6568  Original tweet   …
      194  855138241867124737               0             50  Original tweet   …
      301  836648853927522308               0            626  Original tweet   …
      318  834477809192075265               0          12224  Original tweet   …
      326  833732339549220864               0            251  Original tweet   …
      356  830173239259324417               0           6554  Original tweet   …
      478  815745968457060357               0           4797  Original tweet   …
```

After verification: in this ‘url == “…” ’ some are retweets and others are not dog ratings

```
[28]: df_json.query('url == "s"')
```

```
[28]:                tweet_id  favorite_count  retweet_count  retweet_status url
      54    881633300179243008             129              7  Original tweet   s
      112   870726314365509632             121              3  Original tweet   s
      575   800859414831898624             778            113  Original tweet   s
      1903  674606911342424069            1013            115  Original tweet   s
      2036  671550332464455680             967            229  Original tweet   s
```

After verification: in this 'url == "s" ' some are Replying and others are not dog ratings

```
[29]: df_json.query('url == "0"')
```

```
[29]:                tweet_id  favorite_count  retweet_count  retweet_status url
      147   863427515083354112            2363            105  Original tweet   0
      178   857214891891077121             242             20  Original tweet   0
      185   856288084350160898             545             17  Original tweet   0
      188   855860136149123072            4407           1058  Original tweet   0
      233   847617282490613760            7558            468  Original tweet   0
      289   838150277551247360            1824            370  Original tweet   0
      290   838085839343206401             150              0  Original tweet   0
      345   831926988323639298             369             39  Original tweet   0
      408   823333489516937216             777             31  Original tweet   0
      497   813130366689148928            4968            513  Original tweet   0
      512   811647686436880384            6215            872  Original tweet   0
      855   763956972077010945             812             61  Original tweet   0
      1078  738891149612572673             115              6  Original tweet   0
      1495  692423280028966913            1618            265  Original tweet   0
      1772  678023323247357953            2095            439  Original tweet   0
      2147  669684865554620416             551             99  Original tweet   0
      2296  667070482143944705              88             14  Original tweet   0
```

After verification: in this 'url == "0" ' are Replying

```
[30]: df_json.query('url == "5"')
```

```
[30]:                tweet_id  favorite_count  retweet_count  retweet_status url
      341   832088576586297345              72              3  Original tweet   5
      2187  668967877119254528             161             25  Original tweet   5
```

After verification: in this 'url == "5" ' are Replying

```
[31]: df_json.query('url == "https://…"')
```

```
[31]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      67    879130579576475649               0           7181  Original tweet
      123   868639477480148993               0           2240  Original tweet
      145   863471782782697472               0           2687  Original tweet
      246   845459076796616705               0          15071  Original tweet
```

```
272   840728873075638272               0          18285  Original tweet
308   835536468978302976               0           1939  Original tweet
381   827228250799742977               0          13517  Original tweet
437   820013781606658049               0           1369  Original tweet
474   816062466425819140               0           1207  Original tweet
573   801127390143516673               0           4627  Original tweet
582   800188575492947969               0           4581  Original tweet
593   798705661114773508               0           7724  Original tweet
763   777953400541634568               0           4080  Original tweet
772   776249906839351296               0          10673  Original tweet
1010  747242308580548608               0           3257  Original tweet

            url
67    https://…
123   https://…
145   https://…
246   https://…
272   https://…
308   https://…
381   https://…
437   https://…
474   https://…
573   https://…
582   https://…
593   https://…
763   https://…
772   https://…
1010  https://…
```

After verification: in this ‘url == “https://…”’ are WeRateDogs® Retweeted

[32]:

```
df_json.query('url == "https://t.c…"')
```

[32]:

```
               tweet_id  favorite_count  retweet_count  retweet_status  \
90   874434818259525634               0          15546  Original tweet
164  860177593139703809               0          33421  Original tweet
229  847978865427394560               0           3652  Original tweet
505  812747805718642688               0           8067  Original tweet
537  806620845233815552               0           6426  Original tweet
744  780092040432480260               0           2655  Original tweet
831  768554158521745409               0           6719  Original tweet

              url
90   https://t.c…
164  https://t.c…
229  https://t.c…
505  https://t.c…
537  https://t.c…
```

```
744  https://t.c…
831  https://t.c…
```

After verification: in this ‘url == “https://t.c…”’ are Retweeted

[33]:
```
df_json.query('url == "h"')
```

[33]:
```
                tweet_id  favorite_count  retweet_count  retweet_status url
184  856330835276025856               0            731  Original tweet   h
187  855862651834028034             320             27  Original tweet   h
```

After verification: in this ‘url == “h”’: - This Tweet id 856330835276025856 is from an account that no longer exists - 855862651834028034 not a dog rating

[34]:
```
df_json.query('url == "https://…"')
```

[34]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
67    879130579576475649               0           7181  Original tweet
123   868639477480148993               0           2240  Original tweet
145   863471782782697472               0           2687  Original tweet
246   845459076796616705               0          15071  Original tweet
272   840728873075638272               0          18285  Original tweet
308   835536468978302976               0           1939  Original tweet
381   827228250799742977               0          13517  Original tweet
437   820013781606658049               0           1369  Original tweet
474   816062466425819140               0           1207  Original tweet
573   801127390143516673               0           4627  Original tweet
582   800188575492947969               0           4581  Original tweet
593   798705661114773508               0           7724  Original tweet
763   777953400541634568               0           4080  Original tweet
772   776249906839351296               0          10673  Original tweet
1010  747242308580548608               0           3257  Original tweet

            url
67    https://…
123   https://…
145   https://…
246   https://…
272   https://…
308   https://…
381   https://…
437   https://…
474   https://…
573   https://…
582   https://…
593   https://…
763   https://…
772   https://…
```

```
1010  https://…
```

After verification: in this 'url == "https://…" ' are Retweeted

[35]:

```
df_json.query('url == "https://t…"')
```

[35]:

```
               tweet_id  favorite_count  retweet_count  retweet_status  \
158  860981674716409858               0           2334  Original tweet
203  852936405516943360               0           2243  Original tweet
249  845098359547420673               0           8640  Original tweet
433  820494788566847489               0            648  Original tweet
594  798701998996647937               0           9126  Original tweet
676  789280767834746880               0           5747  Original tweet
816  770743923962707968               0          52360  Original tweet

           url
158  https://t…
203  https://t…
249  https://t…
433  https://t…
594  https://t…
676  https://t…
816  https://t…
```

After verification: in this 'url == "https://t…" ' are Retweeted

[36]:

```
df_json.query('url == "https://t.co/…"')
```

[36]:

```
               tweet_id  favorite_count  retweet_count  retweet_status  \
179  857062103051644929               0            183  Original tweet
306  835685285446955009               0           9327  Original tweet
604  798576900688019456               0           6871  Original tweet
748  779124354206535695               0          21324  Original tweet

               url
179  https://t.co/…
306  https://t.co/…
604  https://t.co/…
748  https://t.co/…
```

After verification: in this 'url == "https://t.co/…" ' some are Retweeted and others are not dog ratings

[37]:

```
df_json.query('url == "n"')
```

[37]:

```
               tweet_id  favorite_count  retweet_count  retweet_status url
29   886267009285017600             117              4  Original tweet   n
273  840698636975636481             197              3  Original tweet   n
374  828361771580813312            2408            195  Original tweet   n
```

After verification: in this ‘url == “n” ’ some are Replying and others are not dog ratings

```
[38]: df_json.query('url == "https://t.co/roqIxCvEB3"')
```

```
[38]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      656   791774931465953280           37818          21156  Original tweet
      2210  668587383441514497            1760           1174  Original tweet

                               url
      656   https://t.co/roqIxCvEB3
      2210  https://t.co/roqIxCvEB3
```

After cheking this ‘url == “https://t.co/roqIxCvEB3” ’ is the same dog voting for the second time

```
[39]: df_json.query('url == "e"')
```

```
[39]:                  tweet_id  favorite_count  retweet_count  retweet_status url
      426   821153421864615936             280             10  Original tweet   e
      1444  696490539101908992            1339            170  Original tweet   e
      1661  682808988178739200            1953            210  Original tweet   e
      1687  681340665377193984            1803            313  Original tweet   e
```

After checking this ‘url == “e” ’ Are replaying tweets

```
[40]: df_json.query('url == "g"')
```

```
[40]:                  tweet_id  favorite_count  retweet_count  retweet_status url
      386   826598799820865537            5637            292  Original tweet   g
      1596  686035780142297088            1274            138  Original tweet   g
```

After checking this ‘url == “g” ’ Are replaying tweets

```
[41]: df_json.query('url == "y"')
```

```
[41]:                 tweet_id  favorite_count  retweet_count  retweet_status url
      706  785515384317313025            6800           1477  Original tweet   y
      841  766714921925144576            2872            438  Original tweet   y
      965  750381685133418496             758             39  Original tweet   y
```

After checking this ‘url == “y” ’ are Not a dog rating

```
[42]: df_json.query('url == "t"')
```

```
[42]:                  tweet_id  favorite_count  retweet_count  retweet_status url
      227   848213670039564288            8834            832  Original tweet   t
      700   786051337297522688            1614            179  Original tweet   t
      1521  690607260360429569             313             19  Original tweet   t
```

After checking this ‘url == “t” ’ are replaying

3 Next

After verification all next links: first one is Retweeted and the second the origin tweet

```
[43]: df_json.query('url == "https://t.co/R6wJyC2Tey"')
```

```
[43]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      529  808134635716833280               0           6961  Original tweet
      572  801167903437357056           27386           6961  Original tweet

                                url
      529  https://t.co/R6wJyC2Tey
      572  https://t.co/R6wJyC2Tey
```

```
[44]: df_json.query('url == "https://t.co/pbK0esr8Tg"')
```

```
[44]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      434  820446719150292993               0           2980  Original tweet
      583  800141422401830912           17092           2980  Original tweet

                                url
      434  https://t.co/pbK0esr8Tg
      583  https://t.co/pbK0esr8Tg
```

```
[45]: df_json.query('url == "https://t.co/y5KarNXWXt"')
```

```
[45]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      617  796177847564038144               0          16628  Original tweet
      618  796149749086875649           36177          16628  Original tweet

                                url
      617  https://t.co/y5KarNXWXt
      618  https://t.co/y5KarNXWXt
```

```
[46]: df_json.query('url == "https://t.co/SATN4If5H5"')
```

```
[46]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      35   885311592912609280               0          19297  Original tweet
      355  830583320585068544           73397          19297  Original tweet

                                url
      35   https://t.co/SATN4If5H5
      355  https://t.co/SATN4If5H5
```

```
[47]: df_json.query('url == "https://t.co/xo7FZFIAao"')
```

```
[47]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      210  851953902622658560               0          10706  Original tweet
```

```
362  829374341691346946          38074          10706  Original tweet

                         url
210  https://t.co/xo7FZFIAao
362  https://t.co/xo7FZFIAao
```

[48]:
```
df_json.query('url == "https://t.co/QV5nx6otZR"')
```

[48]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
358  829878982036299777               0           5757  Original tweet
383  826958653328592898           23767           5757  Original tweet

                         url
358  https://t.co/QV5nx6otZR
383  https://t.co/QV5nx6otZR
```

[49]:
```
df_json.query('url == "https://t.co/X1371QEWd0"')
```

[49]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
906   757729163776290825               0           9299  Original tweet
1746  679062614270468097           18712           9299  Original tweet

                          url
906   https://t.co/X1371QEWd0
1746  https://t.co/X1371QEWd0
```

[50]:
```
df_json.query('url == "https://t.co/WId5lxNdPH"')
```

[50]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
599   798673117451325440               0           6485  Original tweet
1856  675501075957489664           18482           6485  Original tweet

                          url
599   https://t.co/WId5lxNdPH
1856  https://t.co/WId5lxNdPH
```

[51]:
```
df_json.query('url == "https://t.co/Bb3xnpsWBC"')
```

[51]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
947   752309394570878976               0          18963  Original tweet
1863  675354435921575936           35178          18963  Original tweet

                          url
947   https://t.co/Bb3xnpsWBC
1863  https://t.co/Bb3xnpsWBC
```

[52]:
```
df_json.query('url == "https://t.co/2S6oye3XqK"')
```

```
[52]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      719   783347506784731136               0           6533  Original tweet
      1915  674291837063053312           15817           6533  Original tweet

                               url
      719   https://t.co/2S6oye3XqK
      1915  https://t.co/2S6oye3XqK
```

```
[53]: df_json.query('url == "https://t.co/iJABMhVidf"')
```

```
[53]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      810   771171053431250945               0           8705  Original tweet
      1966  673320132811366400           14441           8705  Original tweet

                               url
      810   https://t.co/iJABMhVidf
      1966  https://t.co/iJABMhVidf
```

```
[54]: df_json.query('url == "https://t.co/eqmkr2gmPH"')
```

```
[54]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      600   798665375516884993               0           4519  Original tweet
      2021  671896809300709376            9016           4519  Original tweet

                               url
      600   https://t.co/eqmkr2gmPH
      2021  https://t.co/eqmkr2gmPH
```

```
[55]: df_json.query('url == "https://t.co/J5Mqn8VeYq"')
```

```
[55]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      601   798644042770751489               0           2153  Original tweet
      2108  670444955656130560            7120           2153  Original tweet

                               url
      601   https://t.co/J5Mqn8VeYq
      2108  https://t.co/J5Mqn8VeYq
```

```
[56]: df_json.query('url == "https://t.co/Qnf07HEQGA"')
```

```
[56]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      893   759159934323924993               0           1359  Original tweet
      2125  670319130621435904            4110           1359  Original tweet

                               url
      893   https://t.co/Qnf07HEQGA
      2125  https://t.co/Qnf07HEQGA
```

[57]:
```
df_json.query('url == "https://t.co/0aDF62KVP7"')
```

[57]:
```
                 tweet_id  favorite_count  retweet_count retweet_status  \
602   798628517273620480               0           2272  Original tweet
2267  667509364010450944            7148           2272  Original tweet

                          url
602   https://t.co/0aDF62KVP7
2267  https://t.co/0aDF62KVP7
```

[58]:
```
df_json.query('url == "https://t.co/HVGs5aAKAn"')
```

[58]:
```
                 tweet_id  favorite_count  retweet_count retweet_status  \
595   798697898615730177               0           7602  Original tweet
1231  712809025985978368           20378           7602  Original tweet

                          url
595   https://t.co/HVGs5aAKAn
1231  https://t.co/HVGs5aAKAn
```

[59]:
```
df_json.query('url == "https://t.co/Ezu6jQrKAZ"')
```

[59]:
```
                 tweet_id  favorite_count  retweet_count retweet_status  \
870   761371037149827077               0          20500  Original tweet
1244  711694788429553666           35865          20500  Original tweet

                          url
870   https://t.co/Ezu6jQrKAZ
1244  https://t.co/Ezu6jQrKAZ
```

[60]:
```
df_json.query('url == "https://t.co/oHKJHghrp5"')
```

[60]:
```
                 tweet_id  favorite_count  retweet_count retweet_status  \
727   782021823840026624               0           7236  Original tweet
1300  707610948723478529           18557           7236  Original tweet

                          url
727   https://t.co/oHKJHghrp5
1300  https://t.co/oHKJHghrp5
```

[61]:
```
df_json.query('url == "https://t.co/iHe6liLwWR"')
```

[61]:
```
                 tweet_id  favorite_count  retweet_count retweet_status  \
758   778396591732486144               0          14198  Original tweet
1360  703041949650034688           28996          14198  Original tweet

                          url
758   https://t.co/iHe6liLwWR
```

```
1360   https://t.co/iHe6liLwWR
```

```
[62]: df_json.query('url == "https://t.co/JKXh0NbBNL"')
```

```
[62]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      596   798694562394996736               0           5812  Original tweet
      1378  701214700881756160           13475           5812  Original tweet

                                url
      596   https://t.co/JKXh0NbBNL
      1378  https://t.co/JKXh0NbBNL
```

```
[63]: df_json.query('url == "https://t.co/fR1P3g5I6k"')
```

```
[63]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      554   803692223237865472               0           8689  Original tweet
      1509  691416866452082688           21253           8689  Original tweet

                                url
      554   https://t.co/fR1P3g5I6k
      1509  https://t.co/fR1P3g5I6k
```

```
[64]: df_json.query('url == "https://t.co/VGPH0TfESw"')
```

```
[64]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      866   761750502866649088               0           4535  Original tweet
      1609  685325112850124800           10471           4535  Original tweet

                                url
      866   https://t.co/VGPH0TfESw
      1609  https://t.co/VGPH0TfESw
```

```
[65]: df_json.query('url == "https://t.co/e3LPMAHj7p"')
```

```
[65]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      653   791821351946420224               0          24514  Original tweet
      1623  684830982659280897           38551          24514  Original tweet

                                url
      653   https://t.co/e3LPMAHj7p
      1623  https://t.co/e3LPMAHj7p
```

```
[66]: df_json.query('url == "https://t.co/zxw8d5g94P"')
```

```
[66]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      888   759566828574212096               0          24319  Original tweet
      1075  739544079319588864           43694          24319  Original tweet
```

```
                          url
888   https://t.co/zxw8d5g94P
1075  https://t.co/zxw8d5g94P
```

[67]:

```
df_json.query('url == "https://t.co/u1ulQ5heLX"')
```

[67]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
681   788552643979468800               0           8407  Original tweet
1093  736392552031657984           19450           8407  Original tweet

                          url
681   https://t.co/u1ulQ5heLX
1093  https://t.co/u1ulQ5heLX
```

[68]:

```
df_json.query('url == "https://t.co/pG2inLaOda"')
```

[68]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
777   775898661951791106               0          17621  Original tweet
1111  733109485275860992           44619          17621  Original tweet

                          url
777   https://t.co/pG2inLaOda
1111  https://t.co/pG2inLaOda
```

[69]:

```
df_json.query('url == "https://t.co/ghyT4Ubk1r"')
```

[69]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
475   816014286006976512               0           6154  Original tweet
1116  732005617171337216           16324           6154  Original tweet

                          url
475   https://t.co/ghyT4Ubk1r
1116  https://t.co/ghyT4Ubk1r
```

[70]:

```
df_json.query('url == "https://t.co/vtVHtKFtBH"')
```

[70]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
845   766078092750233600               0           2970  Original tweet
1149  725842289046749185            7691           2970  Original tweet

                          url
845   https://t.co/vtVHtKFtBH
1149  https://t.co/vtVHtKFtBH
```

[71]:

```
df_json.query('url == "https://t.co/XQ1tRUm03z"')
```

[71]:

```
                tweet_id  favorite_count  retweet_count  retweet_status  \
941   752701944171524096               0           3291  Original tweet
```

```
1648  683515932363329536           8058           3291  Original tweet

                          url
941   https://t.co/XQ1tRUm03z
1648  https://t.co/XQ1tRUm03z
```

```
[72]: df_json.query('url == "https://t.co/4Kn9GEHXiE"')
```

```
[72]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      654   791784077045166082               0          11271  Original tweet
      1674  682088079302213632           20108          11271  Original tweet

                                url
      654   https://t.co/4Kn9GEHXiE
      1674  https://t.co/4Kn9GEHXiE
```

```
[73]: df_json.query('url == "https://t.co/hyAC5Hq9GC"')
```

```
[73]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      769   776819012571455488               0          15839  Original tweet
      1730  679828447187857408           39726          15839  Original tweet

                                url
      769   https://t.co/hyAC5Hq9GC
      1730  https://t.co/hyAC5Hq9GC
```

```
[74]: df_json.query('url == "https://t.co/NLOHmlENag"')
```

```
[74]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      924   754874841593970688               0           9193  Original tweet
      1742  679158373988876288           23568           9193  Original tweet

                                url
      924   https://t.co/NLOHmlENag
      1742  https://t.co/NLOHmlENag
```

```
[75]: df_json.query('url == "https://t.co/o5J479bZUC"')
```

```
[75]:                tweet_id  favorite_count  retweet_count  retweet_status  \
      820   770093767776997377               0           3520  Original tweet
      1061  741067306818797568           10342           3520  Original tweet

                                url
      820   https://t.co/o5J479bZUC
      1061  https://t.co/o5J479bZUC
```

```
[76]: df_json.query('url == "https://t.co/BXIuvkQ09b"')
```

```
[76]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      740   780496263422808064               0           4319  Original tweet
      1052  742423170473463808           10812           4319  Original tweet

                                url
      740   https://t.co/BXIuvkQ09b
      1052  https://t.co/BXIuvkQ09b
```

```
[77]: df_json.query('url == "https://t.co/bR0YbceZ1u"')
```

```
[77]:                  tweet_id  favorite_count  retweet_count  retweet_status  \
      655   791780927877898241               0           4432  Original tweet
      1018  746757706116112384           10520           4432  Original tweet

                                url
      655   https://t.co/bR0YbceZ1u
      1018  https://t.co/bR0YbceZ1u
```

```
[78]: df_json.query('url == "https://t.co/lprd0ylVpS"')
```

```
[78]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      766  777641927919427584               0           4947  Original tweet
      963  750429297815552001           14569           4947  Original tweet

                               url
      766  https://t.co/lprd0ylVpS
      963  https://t.co/lprd0ylVpS
```

```
[79]: df_json.query('url == "https://t.co/v5m2bIukXc"')
```

```
[79]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      691  787111942498508800               0           7798  Original tweet
      939  752932432744185856           13970           7798  Original tweet

                               url
      691  https://t.co/v5m2bIukXc
      939  https://t.co/v5m2bIukXc
```

```
[80]: df_json.query('url == "https://t.co/dsv4xAtfv7"')
```

```
[80]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      309  835309094223372289               0          24013  Original tweet
      937  753039830821511168           41080          24013  Original tweet

                               url
      309  https://t.co/dsv4xAtfv7
      937  https://t.co/dsv4xAtfv7
```

```
[81]: df_json.query('url == "https://t.co/iy7ylyZcsE"')
```

```
[81]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      752  778774459159379968               0          11550  Original tweet
      903  758099635764359168           21302          11550  Original tweet

                               url
      752  https://t.co/iy7ylyZcsE
      903  https://t.co/iy7ylyZcsE
```

```
[82]: df_json.query('url == "https://t.co/1acHQS8rvK"')
```

```
[82]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      685  788070120937619456               0           3952  Original tweet
      874  761004547850530816           12482           3952  Original tweet

                               url
      685  https://t.co/1acHQS8rvK
      874  https://t.co/1acHQS8rvK
```

```
[83]: df_json.query('url == "https://t.co/SuXcI9B7pQ"')
```

```
[83]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      560  802624713319034886               0           3403  Original tweet
      765  777684233540206592           12518           3403  Original tweet

                               url
      560  https://t.co/SuXcI9B7pQ
      765  https://t.co/SuXcI9B7pQ
```

```
[84]: df_json.query('url == "https://t.co/gx3xF5mZbo"')
```

```
[84]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      660  791026214425268224               0           4858  Original tweet
      856  763837565564780549           14041           4858  Original tweet

                               url
      660  https://t.co/gx3xF5mZbo
      856  https://t.co/gx3xF5mZbo
```

```
[85]: df_json.query('url == "https://t.co/eSvz4EapHH"')
```

```
[85]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      799  772615324260794368               0           3914  Original tweet
      852  765222098633691136           12902           3914  Original tweet

                               url
      799  https://t.co/eSvz4EapHH
```

```
852  https://t.co/eSvz4EapHH
```

```
[86]: df_json.query('url == "https://t.co/guDkLrX8zV"')
```

```
[86]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      342  832040443403784192               0          11131  Original tweet
      822  769940425801170949           34948          11131  Original tweet

                               url
      342  https://t.co/guDkLrX8zV
      822  https://t.co/guDkLrX8zV
```

```
[87]: df_json.query('url == "https://t.co/E6bGWf9EOm"')
```

```
[87]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      605  798340744599797760               0           3924  Original tweet
      807  771770456517009408           13356           3924  Original tweet

                               url
      605  https://t.co/E6bGWf9EOm
      807  https://t.co/E6bGWf9EOm
```

```
[88]: df_json.query('url == "https://t.co/nBJ5RE4yHb"')
```

```
[88]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      452  818588835076603904               0           7126  Original tweet
      792  773547596996571136           24553           7126  Original tweet

                               url
      452  https://t.co/nBJ5RE4yHb
      792  https://t.co/nBJ5RE4yHb
```

```
[89]: df_json.query('url == "https://t.co/wdgY9oHPrT"')
```

```
[89]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      585  799774291445383169               0           5488  Original tweet
      784  775085132600442880           17281           5488  Original tweet

                               url
      585  https://t.co/wdgY9oHPrT
      784  https://t.co/wdgY9oHPrT
```

```
[90]: df_json.query('url == "https://t.co/I1bJUzUEW5"')
```

```
[90]:               tweet_id  favorite_count  retweet_count  retweet_status  \
      565  802247111496568832               0           5247  Original tweet
      750  779056095788752897           16500           5247  Original tweet
```

```
                          url
565  https://t.co/I1bJUzUEW5
750  https://t.co/I1bJUzUEW5
```

[91]:
```
df_json.query('url == "https://t.co/AsykOwMrXQ"')
```

[91]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
551  804413760345620481               0           3712  Original tweet
709  784826020293709826           11310           3712  Original tweet

                          url
551  https://t.co/AsykOwMrXQ
709  https://t.co/AsykOwMrXQ
```

[92]:
```
df_json.query('url == "https://t.co/t0cEeRbcXJ"')
```

[92]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
545  805823200554876929               0           9374  Original tweet
712  784183165795655680           22513           9374  Original tweet

                          url
545  https://t.co/t0cEeRbcXJ
712  https://t.co/t0cEeRbcXJ
```

[93]:
```
df_json.query('url == "https://t.co/T7Jr4Gw4sC"')
```

[93]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
405  823719002937630720               0          12953  Original tweet
713  784057939640352768           33505          12953  Original tweet

                          url
405  https://t.co/T7Jr4Gw4sC
713  https://t.co/T7Jr4Gw4sC
```

[94]:
```
df_json.query('url == "https://t.co/nfprKAXqwu"')
```

[94]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
540  806242860592926720               0          13616  Original tweet
720  783334639985389568           32651          13616  Original tweet

                          url
540  https://t.co/nfprKAXqwu
720  https://t.co/nfprKAXqwu
```

[95]:
```
df_json.query('url == "https://t.co/zrMVdfFej6"')
```

[95]:
```
                tweet_id  favorite_count  retweet_count  retweet_status  \
542  805958939288408065               0           6238  Original tweet
```

```
724  782722598790725632          19250           6238  Original tweet

                        url
542  https://t.co/zrMVdfFej6
724  https://t.co/zrMVdfFej6
```

```
[96]: df_json.query('url == "https://t.co/mH0IOyFdIG"')
```

```
[96]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      421  821813639212650496               0           3839  Original tweet
      738  780601303617732608           13525           3839  Original tweet

                              url
      421  https://t.co/mH0IOyFdIG
      738  https://t.co/mH0IOyFdIG
```

```
[97]: df_json.query('url == "https://t.co/x0gV2Ck3AD"')
```

```
[97]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      701  786036967502913536               0           8237  Original tweet
      746  779834332596887552           21252           8237  Original tweet

                              url
      701  https://t.co/x0gV2Ck3AD
      746  https://t.co/x0gV2Ck3AD
```

```
[98]: df_json.query('url == "https://t.co/xIfrsMNLBR"')
```

```
[98]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      521  809808892968534016               0           5711  Original tweet
      631  793962221541933056           18910           5711  Original tweet

                              url
      521  https://t.co/xIfrsMNLBR
      631  https://t.co/xIfrsMNLBR
```

```
[99]: df_json.query('url == "https://t.co/rLHGx3VAF3"')
```

```
[99]:                 tweet_id  favorite_count  retweet_count  retweet_status  \
      340  832215726631055365               0          10723  Original tweet
      636  793286476301799424           27597          10723  Original tweet

                              url
      340  https://t.co/rLHGx3VAF3
      636  https://t.co/rLHGx3VAF3
```

```
[100]: df_json.query('url == "https://t.co/O5IUmRHRIh"')
```

```
[100]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       633  793614319594401792               0           3661  Original tweet
       657  791672322847637504           13129           3661  Original tweet

                                  url
       633  https://t.co/O5IUmRHRIh
       657  https://t.co/O5IUmRHRIh
```

```
[101]: df_json.query('url == "https://t.co/dgvTPZ9tgI"')
```

```
[101]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       468  816829038950027264               0           5496  Original tweet
       662  790946055508652032           18601           5496  Original tweet

                                  url
       468  https://t.co/dgvTPZ9tgI
       662  https://t.co/dgvTPZ9tgI
```

```
[102]: df_json.query('url == "https://t.co/aI7umZHIq7"')
```

```
[102]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       487  813944609378369540               0           3732  Original tweet
       667  790277117346975746           14081           3732  Original tweet

                                  url
       487  https://t.co/aI7umZHIq7
       667  https://t.co/aI7umZHIq7
```

```
[103]: df_json.query('url == "https://t.co/Ln89R4FP7v"')
```

```
[103]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       663  790723298204217344               0           2704  Original tweet
       669  789986466051088384           10369           2704  Original tweet

                                  url
       663  https://t.co/Ln89R4FP7v
       669  https://t.co/Ln89R4FP7v
```

```
[104]: df_json.query('url == "https://t.co/TU4sZogVIq"')
```

```
[104]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       626  794983741416415232               0           3942  Original tweet
       674  789530877013393408           13188           3942  Original tweet

                                  url
       626  https://t.co/TU4sZogVIq
       674  https://t.co/TU4sZogVIq
```

[105]:
```
df_json.query('url == "https://t.co/x5gXRS1abq"')
```

[105]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
628  794355576146903043               0          12014  Original tweet
680  788765914992902144           30658          12014  Original tweet

                         url
628  https://t.co/x5gXRS1abq
680  https://t.co/x5gXRS1abq
```

[106]:
```
df_json.query('url == "https://t.co/r28jFx9uyF"')
```

[106]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
94   873697596434513921               0          12518  Original tweet
122  868880397819494401           55098          12518  Original tweet

                         url
94   https://t.co/r28jFx9uyF
122  https://t.co/r28jFx9uyF
```

[107]:
```
df_json.query('url == "https://t.co/liJGwMp17E"')
```

[107]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
414  822647212903690241               0           7390  Original tweet
416  822489057087389700           20083           7390  Original tweet

                         url
414  https://t.co/liJGwMp17E
416  https://t.co/liJGwMp17E
```

[108]:
```
df_json.query('url == "https://t.co/rLi83ZyCL5"')
```

[108]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
410  823269594223824897               0          11421  Original tweet
418  822244816520155136           38832          11421  Original tweet

                         url
410  https://t.co/rLi83ZyCL5
418  https://t.co/rLi83ZyCL5
```

[109]:
```
df_json.query('url == "https://t.co/Y0C1fHFCSb"')
```

[109]:
```
               tweet_id  favorite_count  retweet_count  retweet_status  \
446  819015331746349057               0          21794  Original tweet
447  819006400881917954           49960          21794  Original tweet

                         url
446  https://t.co/Y0C1fHFCSb
```

```
447  https://t.co/YOC1fHFCSb
```

```
[110]: df_json.query('url == "https://t.co/AdPKrI8BZ1"')
```

```
[110]:               tweet_id  favorite_count  retweet_count  retweet_status  \
445  819015337530290176               0          42228  Original tweet
448  819004803107983360           95450          42228  Original tweet

                         url
445  https://t.co/AdPKrI8BZ1
448  https://t.co/AdPKrI8BZ1
```

```
[111]: df_json.query('url == "https://t.co/btzf1zTDeQ"')
```

```
[111]:               tweet_id  favorite_count  retweet_count  retweet_status  \
265  841833993020538882               0          17504  Original tweet
462  817423860136083457           38260          17504  Original tweet

                         url
265  https://t.co/btzf1zTDeQ
462  https://t.co/btzf1zTDeQ
```

```
[112]: df_json.query('url == "https://t.co/MiMKtsLN6k"')
```

```
[112]:               tweet_id  favorite_count  retweet_count  retweet_status  \
464  817181837579653120               0           9907  Original tweet
477  815966073409433600           25057           9907  Original tweet

                         url
464  https://t.co/MiMKtsLN6k
477  https://t.co/MiMKtsLN6k
```

```
[113]: df_json.query('url == "https://t.co/hfi264QwYM"')
```

```
[113]:               tweet_id  favorite_count  retweet_count  retweet_status  \
449  818646164899774465               0          24069  Original tweet
528  808344865868283904           47281          24069  Original tweet

                         url
449  https://t.co/hfi264QwYM
528  https://t.co/hfi264QwYM
```

```
[114]: df_json.query('url == "https://t.co/DkBYaCAg2d"')
```

```
[114]:               tweet_id  favorite_count  retweet_count  retweet_status  \
259  842892208864923648               0          56625  Original tweet
533  807106840509214720          107015          56625  Original tweet
```

```
                     url
259  https://t.co/DkBYaCAg2d
533  https://t.co/DkBYaCAg2d
```

```
[115]: df_json.query('url == "https://t.co/f1h2Fsalv9"')
```

```
[115]:               tweet_id  favorite_count  retweet_count  retweet_status  \
       154  861769973181624320               0          37911  Original tweet
       536  806629075125202948           75639          37911  Original tweet

                            url
       154  https://t.co/f1h2Fsalv9
       536  https://t.co/f1h2Fsalv9
```

Back To Quality list

```
[116]: # check duplicates in each column in df_json
       duplic(df_json)
```

```
tweet_id
 0
favorite_count
 347
retweet_count
 630
retweet_status
 2353
url
 175
```

issue12 Quality

```
[117]: df_json.retweet_status.value_counts()
```

```
[117]: Original tweet    2354
       Name: retweet_status, dtype: int64
```

issue13 Quality

```
[118]: df_json.favorite_count.value_counts()
```

```
[118]: 0       179
       2706      3
       2918      3
       610       3
       2768      3
              …
       7493      1
       6671      1
```

```
4484      1
2872      1
2535      1
Name: favorite_count, Length: 2007, dtype: int64
```

[119]:
```
df_json.retweet_count.value_counts()
```

[119]:
```
1972    5
3652    5
83      5
1207    4
336     4
       ..
3357    1
3018    1
2181    1
848     1
147     1
Name: retweet_count, Length: 1724, dtype: int64
```

[120]:
```
# check value_counts in each column in df_img
val_count(df_img)
```

```
tweet_id
 666020888022790149    1
761292947749015552    1
767122157629476866    1
766793450729734144    1
766693177336135680    1
                     ..
684097758874210310    1
683857920510050305    1
683852578183077888    1
683849932751646720    1
892420643555336193    1
Name: tweet_id, Length: 2075, dtype: int64
jpg_url
 https://pbs.twimg.com/media/CZhn-QAWwAASQan.jpg
2
https://pbs.twimg.com/media/Cq9guJ5WgAADfpF.jpg
2
https://pbs.twimg.com/ext_tw_video_thumb/807106774843039744/pu/img/8XZg1xW35Xp2J
6JW.jpg    2
https://pbs.twimg.com/media/CU1zsMSUAAAS0qW.jpg
2
https://pbs.twimg.com/media/CsrjryzWgAAZY0O.jpg
2
          ..
```

```
https://pbs.twimg.com/media/CXrmMSpUwAAdeRj.jpg
1
https://pbs.twimg.com/media/CXrawAhWkAAWSxC.jpg
1
https://pbs.twimg.com/media/CXrIntsUsAEkv0d.jpg
1
https://pbs.twimg.com/media/CXqc0HCUQAAugTB.jpg
1
https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg
1
Name: jpg_url, Length: 2009, dtype: int64
img_num
 1    1780
2     198
3      66
4      31
Name: img_num, dtype: int64
p1
 golden_retriever      150
Labrador_retriever    100
Pembroke               89
Chihuahua              83
pug                    57
                     …
pillow                  1
carousel                1
bald_eagle              1
lorikeet                1
orange                  1
Name: p1, Length: 378, dtype: int64
p1_conf
 0.721188    2
0.676439    2
0.505370    2
0.777468    2
0.254856    2
           ..
0.141432    1
0.722218    1
0.278600    1
0.375610    1
0.097049    1
Name: p1_conf, Length: 2006, dtype: int64
p1_dog
 True     1532
False     543
Name: p1_dog, dtype: int64
p2
```

```
 Labrador_retriever    104
golden_retriever       92
Cardigan               73
Chihuahua              44
Pomeranian             42
                     …
medicine_chest          1
quail                   1
horse_cart              1
waffle_iron             1
bagel                   1
Name: p2, Length: 405, dtype: int64
p2_conf
 0.069362    3
0.001186    2
0.181351    2
0.016641    2
0.112943    2
           ..
0.053008    1
0.094044    1
0.193804    1
0.155207    1
0.085851    1
Name: p2_conf, Length: 2004, dtype: int64
p2_dog
 True     1553
False     522
Name: p2_dog, dtype: int64
p3
 Labrador_retriever    79
Chihuahua             58
golden_retriever      48
Eskimo_dog            38
kelpie                35
                      ..
ox                     1
assault_rifle          1
axolotl                1
pot                    1
banana                 1
Name: p3, Length: 408, dtype: int64
p3_conf
 0.023353    2
0.046403    2
0.077008    2
0.017492    2
0.223263    2
```

```
           ..
0.037881    1
0.085237    1
0.055194    1
0.153598    1
0.076110    1
Name: p3_conf, Length: 2006, dtype: int64
p3_dog
 True     1499
False     576
Name: p3_dog, dtype: int64
```

Back To Quality list

issue10 Quality

[121]:
```
# Check a specific jpg_url to get more info
df_img.query('jpg_url == "https://pbs.twimg.com/media/CZhn-QAWwAASQan.jpg"')
```

[121]:
```
                 tweet_id                                          jpg_url  \
800   691416866452082688  https://pbs.twimg.com/media/CZhn-QAWwAASQan.jpg
1624  803692223237865472  https://pbs.twimg.com/media/CZhn-QAWwAASQan.jpg

      img_num                p1    p1_conf  p1_dog              p2   p2_conf  \
800         1  Lakeland_terrier  0.530104    True  Irish_terrier  0.197314
1624        1  Lakeland_terrier  0.530104    True  Irish_terrier  0.197314

      p2_dog        p3   p3_conf  p3_dog
800     True  Airedale  0.082515    True
1624    True  Airedale  0.082515    True
```

[122]:
```
# Get list of duplicates
ids_dup = np.array(df_img["tweet_id"][df_img["jpg_url"].duplicated()])
ids_dup
```

[122]:
```
array([752309394570878976, 754874841593970688, 757729163776290825,
       759159934323924993, 759566828574212096, 761371037149827077,
       761750502866649088, 766078092750233600, 770093767776997377,
       771171053431250945, 772615324260794368, 775898661951791106,
       776819012571455488, 777641927919427584, 778396591732486144,
       780496263422808064, 782021823840026624, 783347506784731136,
       786036967502913536, 788070120937619456, 790723298204217344,
       791026214425268224, 793614319594401792, 794355576146903043,
       794983741416415232, 796177847564038144, 798340744599797760,
       798628517273620480, 798644042770751489, 798665375516884993,
       798673117451325440, 798694562394996736, 798697898615730177,
       799774291445383169, 800443802682937345, 802247111496568832,
       802624713319034886, 803692223237865472, 804413760345620481,
       805958939288408065, 806242860592926720, 807059379405148160,
       808134635716833280, 809808892968534016, 813944609378369540,
       816014286006976512, 816829038950027264, 817181837579653120,
       818588835076603904, 819015331746349057, 819015337530290176,
       820446719150292993, 821813639212650496, 822647212903690241,
       823269594223824897, 824796380199809024, 829878982036299777,
       832040443403784192, 832215726631055365, 841833993020538882,
       842892208864923648, 851953902622658560, 861769973181624320,
       873697596434513921, 885311592912609280, 888202515573088257])
```

[123]:

```
# Print and Check each duplicates ids manually
for i in ids_dup:
    print('https://twitter.com/dog_rates/status/'+ str(i))
```

```
https://twitter.com/dog_rates/status/752309394570878976
https://twitter.com/dog_rates/status/754874841593970688
https://twitter.com/dog_rates/status/757729163776290825
https://twitter.com/dog_rates/status/759159934323924993
https://twitter.com/dog_rates/status/759566828574212096
https://twitter.com/dog_rates/status/761371037149827077
https://twitter.com/dog_rates/status/761750502866649088
https://twitter.com/dog_rates/status/766078092750233600
https://twitter.com/dog_rates/status/770093767776997377
https://twitter.com/dog_rates/status/771171053431250945
https://twitter.com/dog_rates/status/772615324260794368
https://twitter.com/dog_rates/status/775898661951791106
https://twitter.com/dog_rates/status/776819012571455488
https://twitter.com/dog_rates/status/777641927919427584
https://twitter.com/dog_rates/status/778396591732486144
https://twitter.com/dog_rates/status/780496263422808064
https://twitter.com/dog_rates/status/782021823840026624
https://twitter.com/dog_rates/status/783347506784731136
https://twitter.com/dog_rates/status/786036967502913536
https://twitter.com/dog_rates/status/788070120937619456
https://twitter.com/dog_rates/status/790723298204217344
https://twitter.com/dog_rates/status/791026214425268224
https://twitter.com/dog_rates/status/793614319594401792
https://twitter.com/dog_rates/status/794355576146903043
https://twitter.com/dog_rates/status/794983741416415232
https://twitter.com/dog_rates/status/796177847564038144
https://twitter.com/dog_rates/status/798340744599797760
https://twitter.com/dog_rates/status/798628517273620480
https://twitter.com/dog_rates/status/798644042770751489
https://twitter.com/dog_rates/status/798665375516884993
https://twitter.com/dog_rates/status/798673117451325440
https://twitter.com/dog_rates/status/798694562394996736
https://twitter.com/dog_rates/status/798697898615730177
https://twitter.com/dog_rates/status/799774291445383169
https://twitter.com/dog_rates/status/800443802682937345
```

```
https://twitter.com/dog_rates/status/802247111496568832
https://twitter.com/dog_rates/status/802624713319034886
https://twitter.com/dog_rates/status/803692223237865472
https://twitter.com/dog_rates/status/804413760345620481
https://twitter.com/dog_rates/status/805958939288408065
https://twitter.com/dog_rates/status/806242860592926720
https://twitter.com/dog_rates/status/807059379405148160
https://twitter.com/dog_rates/status/808134635716833280
https://twitter.com/dog_rates/status/809808892968534016
https://twitter.com/dog_rates/status/813944609378369540
https://twitter.com/dog_rates/status/816014286006976512
https://twitter.com/dog_rates/status/816829038950027264
https://twitter.com/dog_rates/status/817181837579653120
https://twitter.com/dog_rates/status/818588835076603904
https://twitter.com/dog_rates/status/819015331746349057
https://twitter.com/dog_rates/status/819015337530290176
https://twitter.com/dog_rates/status/820446719150292993
https://twitter.com/dog_rates/status/821813639212650496
https://twitter.com/dog_rates/status/822647212903690241
https://twitter.com/dog_rates/status/823269594223824897
https://twitter.com/dog_rates/status/824796380199809024
https://twitter.com/dog_rates/status/829878982036299777
https://twitter.com/dog_rates/status/832040443403784192
https://twitter.com/dog_rates/status/832215726631055365
https://twitter.com/dog_rates/status/841833993020538882
https://twitter.com/dog_rates/status/842892208864923648
https://twitter.com/dog_rates/status/851953902622658560
https://twitter.com/dog_rates/status/861769973181624320
https://twitter.com/dog_rates/status/873697596434513921
https://twitter.com/dog_rates/status/885311592912609280
https://twitter.com/dog_rates/status/888202515573088257
```

After checking all these urls some are retweeted and otheres Page not existing

[124]:
```
# check duplicates in each column in df_img
duplic(df_img)
```

```
tweet_id
 0
jpg_url
 66
img_num
 2071
p1
 1697
p1_conf
 69
p1_dog
 2073
```

```
p2
 1670
p2_conf
 71
p2_dog
 2073
p3
 1667
p3_conf
 69
p3_dog
 2073
```

Back To Quality list

issue9 Quality

[125]:

```
#Check values in p1 col
df_img.p1.value_counts()
```

[125]:

```
golden_retriever      150
Labrador_retriever    100
Pembroke               89
Chihuahua              83
pug                    57
                     …
pillow                  1
carousel                1
bald_eagle              1
lorikeet                1
orange                  1
Name: p1, Length: 378, dtype: int64
```

[126]:

```
#Check values in p2 col
df_img.p2.value_counts()
```

[126]:

```
Labrador_retriever    104
golden_retriever       92
Cardigan               73
Chihuahua              44
Pomeranian             42
                     …
medicine_chest          1
quail                   1
horse_cart              1
waffle_iron             1
bagel                   1
Name: p2, Length: 405, dtype: int64
```

[127]:
```
#Check values in p3 col
df_img.p3.value_counts()
```

[127]:
```
Labrador_retriever    79
Chihuahua             58
golden_retriever      48
Eskimo_dog            38
kelpie                35
                      ..
ox                     1
assault_rifle          1
axolotl                1
pot                    1
banana                 1
Name: p3, Length: 408, dtype: int64
```

[128]:
```
# check duplicates in each column in df_twt_arch
duplic(df_twt_arch)
```

```
tweet_id
 0
in_reply_to_status_id
 2278
in_reply_to_user_id
 2324
timestamp
 0
source
 2352
text
 0
retweeted_status_id
 2174
retweeted_status_user_id
 2330
retweeted_status_timestamp
 2174
expanded_urls
 137
rating_numerator
 2316
rating_denominator
 2338
name
 1399
doggo
 2354
floofer
```

```
 2354
pupper
 2354
puppo
 2354
```

[129]:
```
# check value_counts in each column in df_twt_arch
val_count(df_twt_arch)
```

```
tweet_id
 892420643555336193    1
687102708889812993    1
687826841265172480    1
687818504314159109    1
687807801670897665    1
                     ..
775085132600442880    1
774757898236878852    1
774639387460112384    1
774314403806253056    1
666020888022790149    1
Name: tweet_id, Length: 2356, dtype: int64
in_reply_to_status_id
 6.671522e+17    2
8.862664e+17    1
6.920419e+17    1
6.827884e+17    1
6.842229e+17    1
               ..
8.116272e+17    1
8.131273e+17    1
8.211526e+17    1
8.233264e+17    1
6.670655e+17    1
Name: in_reply_to_status_id, Length: 77, dtype: int64
in_reply_to_user_id
 4.196984e+09    47
2.195506e+07     2
2.281182e+09     1
1.132119e+08     1
1.637468e+07     1
4.670367e+08     1
1.198989e+09     1
2.878549e+07     1
2.319108e+09     1
3.589728e+08     1
4.717297e+09     1
1.584641e+07     1
```

```
7.305050e+17     1
2.916630e+07     1
2.918590e+08     1
1.185634e+07     1
2.068372e+07     1
1.582854e+09     1
4.738443e+07     1
3.058208e+07     1
2.625958e+07     1
2.894131e+09     1
8.405479e+17     1
1.361572e+07     1
1.943518e+08     1
2.792810e+08     1
1.806710e+08     1
7.759620e+07     1
1.648776e+07     1
3.105441e+09     1
2.143566e+07     1
Name: in_reply_to_user_id, dtype: int64
timestamp
 2017-08-01 16:23:56 +0000    1
2016-01-13 02:43:46 +0000    1
2016-01-15 02:41:12 +0000    1
2016-01-15 02:08:05 +0000    1
2016-01-15 01:25:33 +0000    1
                            ..
2016-09-11 21:34:30 +0000    1
2016-09-10 23:54:11 +0000    1
2016-09-10 16:03:16 +0000    1
2016-09-09 18:31:54 +0000    1
2015-11-15 22:32:08 +0000    1
Name: timestamp, Length: 2356, dtype: int64
source
 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>     2221
<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>
91
<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
33
<a href="https://about.twitter.com/products/tweetdeck"
rel="nofollow">TweetDeck</a>      11
Name: source, dtype: int64
text
 This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut.
13/10 https://t.co/MgUWQ76dJU                                              1
Army of water dogs here. None of them know where they're going. Have no real
purpose. Aggressive barks. 5/10 for all https://t.co/A88x73TwMN       1
```

```
This is Louis. He's a rollercoaster of emotions. Incalculably fluffy. 12/10
would pet firmly https://t.co/17RGvOZO9P                                1
With great pupper comes great responsibility. 12/10 https://t.co/hK6xB042EP
1
Meet Trooper &amp; Maya. Trooper protects Maya from bad things like dognappers
and Comcast. So touching. 11/10 for both https://t.co/c98k1IoZKy    1
                                                                   ..
This is Tucker. He would like a hug. 13/10 someone hug him
https://t.co/wdgY9oHPrT
1
This is Finley. She's a Beneboop Cumbersplash. 12/10 I'd do unspeakable things
for Finley https://t.co/dS8SCbNF9P                                   1
This is Sprinkles. He's trapped in light jail. 10/10 would post bail for him
https://t.co/4s5Xlijogu                                                1
I WAS SENT THE ACTUAL DOG IN THE PROFILE PIC BY HIS OWNER THIS IS SO WILD. 14/10
ULTIMATE LEGEND STATUS https://t.co/7oQ1wpfxIH                     1
Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of
relaxing on stair. 8/10 would pet https://t.co/BLDqew2Ijj                 1
Name: text, Length: 2356, dtype: int64
retweeted_status_id
 8.874740e+17    1
6.704450e+17    1
6.671828e+17    1
6.661041e+17    1
7.717705e+17    1
               ..
7.950767e+17    1
7.840579e+17    1
8.222448e+17    1
8.224891e+17    1
6.675484e+17    1
Name: retweeted_status_id, Length: 181, dtype: int64
retweeted_status_user_id
 4.196984e+09    156
4.296832e+09      2
5.128045e+08      1
7.124572e+17      1
7.832140e+05      1
2.804798e+08      1
1.950368e+08      1
1.732729e+09      1
1.228326e+09      1
7.992370e+07      1
2.488557e+07      1
4.871977e+08      1
4.466750e+07      1
7.266347e+08      1
1.547674e+08      1
```

```
1.960740e+07      1
4.119842e+07      1
5.970642e+08      1
3.410211e+08      1
7.475543e+17      1
6.669901e+07      1
5.870972e+07      1
3.638908e+08      1
7.874618e+17      1
8.117408e+08      1
Name: retweeted_status_user_id, dtype: int64
retweeted_status_timestamp
 2017-07-19 00:47:34 +0000    1
2015-11-28 03:31:48 +0000    1
2015-11-19 03:29:07 +0000    1
2015-11-16 04:02:55 +0000    1
2016-09-02 18:03:10 +0000    1
                            ..
2016-11-06 01:33:58 +0000    1
2016-10-06 15:49:14 +0000    1
2017-01-20 00:50:15 +0000    1
2017-01-20 17:00:46 +0000    1
2015-11-20 03:41:59 +0000    1
Name: retweeted_status_timestamp, Length: 181, dtype: int64
expanded_urls
 https://twitter.com/dog_rates/status/667152164079423490/photo/1
2
https://twitter.com/dog_rates/status/718631497683582976/photo/1
2
https://twitter.com/dog_rates/status/783839966405230592/photo/1,https://twitter.
com/dog_rates/status/783839966405230592/photo/1,https://twitter.com/dog_rates/st
atus/783839966405230592/photo/1
2
https://twitter.com/dog_rates/status/774314403806253056/photo/1,https://twitter.
com/dog_rates/status/774314403806253056/photo/1,https://twitter.com/dog_rates/st
atus/774314403806253056/photo/1,https://twitter.com/dog_rates/status/77431440380
6253056/photo/1    2
https://twitter.com/dog_rates/status/781524693396357120/photo/1
2
                  ..
https://twitter.com/dog_rates/status/763837565564780549/photo/1
1
https://twitter.com/dog_rates/status/764259802650378240/photo/1,https://twitter.
com/dog_rates/status/764259802650378240/photo/1
1
https://twitter.com/dog_rates/status/764857477905154048/photo/1
1
https://twitter.com/dog_rates/status/765222098633691136/photo/1
```

```
1
https://twitter.com/dog_rates/status/666020888022790149/photo/1
1
Name: expanded_urls, Length: 2218, dtype: int64
rating_numerator
 12      558
11      464
10      461
13      351
9       158
8       102
7        55
14       54
5        37
6        32
3        19
4        17
2         9
1         9
75        2
15        2
420       2
0         2
80        1
144       1
17        1
26        1
20        1
121       1
143       1
44        1
60        1
45        1
50        1
99        1
204       1
1776      1
165       1
666       1
27        1
182       1
24        1
960       1
84        1
88        1
Name: rating_numerator, dtype: int64
rating_denominator
 10     2333
```

```
11        3
50        3
20        2
80        2
70        1
7         1
15        1
150       1
170       1
0         1
90        1
40        1
130       1
110       1
16        1
120       1
2         1
Name: rating_denominator, dtype: int64
name
 None          745
a              55
Charlie        12
Cooper         11
Lucy           11
             …
Dex             1
Ace             1
Tayzie          1
Grizzie         1
Christoper      1
Name: name, Length: 957, dtype: int64
doggo
 None     2259
doggo      97
Name: doggo, dtype: int64
floofer
 None       2346
floofer      10
Name: floofer, dtype: int64
pupper
 None      2099
pupper     257
Name: pupper, dtype: int64
puppo
 None     2326
puppo      30
Name: puppo, dtype: int64
```

Back To quality list

issue6 Quality

[130]:
```
#check name column in data frame tweeter archive
names =[]
for i in df_twt_arch.name:
    names.append(i)

names
```

[130]:
```
['Phineas',
 'Tilly',
 'Archie',
 'Darla',
 'Franklin',
 'None',
 'Jax',
 'None',
 'Zoey',
 'Cassie',
 'Koda',
 'Bruno',
 'None',
 'Ted',
 'Stuart',
 'Oliver',
 'Jim',
 'Zeke',
 'Ralphus',
 'Canela',
 'Gerald',
 'Jeffrey',
 'such',
 'Canela',
 'None',
 'None',
 'Maya',
 'Mingus',
 'Derek',
 'Roscoe',
 'None',
 'Waffles',
 'None',
 'Jimbo',
 'Maisey',
 'None',
 'Lilly',
 'None',
 'Earl',
 'Lola',
 'Kevin',
 'None',
 'None',
 'Yogi',
 'Noah',
 'Bella',
 'Grizzwald',
 'None',
 'Rusty',
 'Gus',
 'Stanley',
 'Alfy',
 'Koko',
 'Rey',
 'Gary',
 'None',
 'a',
 'Elliot',
 'Louis',
 'None',
 'Bella',
 'Jesse',
 'None',
 'Romeo',
 'None',
 'Bailey',
 'Duddles',
 'Jack',
 'Emmy',
 'Steven',
 'Beau',
 'Snoopy',
 'None',
 'Shadow',
 'Terrance',
 'Shadow',
 'Emmy',
 'Aja',
 'None',
 'Penny',
 'Dante',
 'Nelly',
 'Ginger',
 'None',
 'Benedict',
 'Venti',
 'Goose',
 'Nugget',
 'None',
 'None',
 'Cash',
 'Coco',
 'Jed',
 'None',
 'Sebastian',
 'Walter',
 'None',
 'Sierra',
 'Sierra',
 'None',
 'None',
 'None',
 'Monkey',
 'None',
 'Harry',
 'Kody',
 'Lassie',
 'Rover',
 'Napolean',
 'Dawn',
 'None',
 'Boomer',
 'None',
 'None',
 'Cody',
 'Zoey',
 'Rumble',
 'Clifford',
 'quite',
 'Dewey',
 'Stanley',
 'Scout',
 'Gizmo',
 'Walter',
 'Cooper',
 'None',
 'Cooper',
 'None',
 'Harold',
 'Shikha',
 'None',
 'None',
 'Jamesy',
 'None',
 'Lili',
 'Jamesy',
 'Coco',
 'None',
 'Boomer',
 'Sammy',
 'Nelly',
 'None',
 'Meatball',
 'Paisley',
 'Albus',
 'Neptune',
 'Quinn',
 'Belle',
 'None',
 'None',
 'Quinn',
 'Zooey',
 'Dave',
 'Jersey',
 'None',
 'None',
 'Hobbes',
 'None',
 'Burt',
 'Lorenzo',
 'None',
 'Lorenzo',
 'Carl',
 'Jordy',
 'None',
 'None',
 'Milky',
 'Trooper',
 'None',
 'quite',
 'None',
 'Winston',
 'None',
 'Sophie',
 'Wyatt',
 'Rosie',
 'Thor',
 'None',
 'Oscar',
 'None',
 'None',
 'Zeke',
 'Luna',
 'Callie',
 'None',
 'None',
 'None',
 'Cermet',
 'None',
 'None',
 'None',
 'None',
 'None',
 'quite',
 'George',
 'None',
 'Marlee',
 'Arya',
 'Einstein',
 'None',
 'None',
 'Alice',
 'None',
 'Rumpole',
 'None',
 'Benny',
 'Aspen',
 'Jarod',
 'Wiggles',
 'General',
 'Sailor',
 'Astrid',
 'None',
 'None',
 'Iggy',
 'Snoop',
 'Kyle',
 'Leo',
 'None',
 'Riley',
 'Boomer',
 'None',
 'Gidget',
 'Noosh',
 'None',
 'Kevin',
 'None',
 'Odin',
 'None',
 'Jerry',
 'Charlie',
 'None',
 'Georgie',
 'Rontu',
 'None',
 'Cannon',
 'Furzey',
 'Daisy',
 'None',
 'Tuck',
 'Barney',
 'None',
 'Vixen',
 'None',
 'Jarvis',
 'None',
 'None',
 'None',
 'Mimosa',
 'Pickles',
 'Bungalo',
 'None',
 'Brady',
 'Luna',
 'Charlie',
 'Margo',
 'None',
 'Sadie',
 'Hank',
 'Tycho',
 'Stephan',
 'Charlie',
 'Indie',
 'Winnie',
 'George',
 'Bentley',
 'Ken',
 'Penny',
 'None',
 'None',
 'Max',
 'Dawn',
 'Maddie',
 'Pipsy',
 'None',
 'None',
 'Maddie',
 'None',
 'Monty',
 'Sojourner',
 'Winston',
 'None',
 'Odie',
 'None',
 'Arlo',
 'None',
 'Riley',
 'Walter',
 'Stanley',
 'Sunny',
 'None',
 'None',
 'Daisy',
 'None',
 'Waffles',
 'Vincent',
 'Lucy',
 'Clark',
 'None',
 'Mookie',
 'Meera',
 'Oliver',
 'None',
 'Buddy',
 'Ava',
 'Lucy',
 'None',
 'Rory',
 'Eli',
 'Lola',
 'None',
 'Ash',
 'Lola',
 'None',
 'None',
 'None',
 'Tucker',
 'Tobi',
 'None',
 'Leo',
 'Chester',
 'Wilson',
 'Sunshine',
 'None',
 'Lipton',
 'Bentley',
 'Charlie',
 'Gabby',
 'Bronte',
 'Poppy',
 'Gidget',
 'Rhino',
 'None',
 'Willow',
 'None',
 'not',
 'Orion',
 'Eevee',
 'Charlie',
 'Smiley',
 'Logan',
 'Moreton',
 'None',
 'Klein',
 'Miguel',
 'Emanuel',
 'None',
 'Kuyu',
 'Daisy',
 'None',
 'Dutch',
 'Pete',
 'None',
 'Scooter',
 'Tucker',
 'Reggie',
 'Lilly',
 'Kyro',
 'Samson',
 'Loki',
 'Mia',
 'Leo',
 'None',
 'Astrid',
 'Malcolm',
 'Dexter',
 'Gus',
 'Alfie',
 'Fiona',
 'one',
 'Mutt',
 'Bear',
 'Doobert',
 'Beebop',
 'Alexander',
 'None',
 'Sailer',
 'Brutus',
 'Kona',
 'Boots',
 'Tucker',
 'Ralphie',
 'Phil',
 'Charlie',
 'Loki',
 'Cupid',
 'None',
 'None',
 'Pawnd',
 'Pilot',
 'None',
 'None',
 'Ike',
 'Mo',
 'Toby',
 'None',
 'Sweet',
 'Pablo',
 'Pablo',
 'Bailey',
 'Scooter',
 'Wilson',
 'None',
 'Nala',
 'None',
 'Cash',
 'Balto',
 'Winston',
 'Crawford',
 'None',
 'Wyatt',
 'None',
 'Albus',
 'None',
 'Hobbes',
 'Paisley',
 'None',
 'Paisley',
 'Gabe',
 'None',
 'Mattie',
 'Jimison',
 'Hercules',
 'Duchess',
 'Harlso',
 'Sampson',
 'Sundance',
 'None',
 'Luca',
 'None',
 'Flash',
 'Finn',
 'Sunny',
 'None',
 'None',
 'Peaches',
 'None',
 'None',
 'Oliver',
 'Oliver',
 'None',
 'Howie',
 'Jazzy',
 'Anna',
 'None',
 'Finn',
 'Bo',
 'Sunny',
 'Sunny',
 'Bo',
 'Seamus',
 'Wafer',
 'Bear',
 'Chelsea',
 'Tom',
 'Moose',
 'Florence',
 'Autumn',
 'None',
 'Buddy',
 'Dido',
 'Eugene',
 'Herschel',
 'Ken',
 'Strudel',
 'None',
 'Tebow',
 'None',
 'Chloe',
 'Betty',
 'Timber',
 'Binky',
 'Moose',
 'Dudley',
 'Comet',
 'Jack',
 'Larry',
 'Jack',
 'None',
 'Levi',
 'Akumi',
 'Titan',
 'None',
 'Cooper',
 'Olivia',
 'Beau',
 'Alf',
 'Oshie',
 'Bruce',
 'Chubbs',
 'Gary',
 'Sky',
 'Atlas',
 'None',
 'None',
 'Eleanor',
 'Layla',
 'None',
 'None',
 'None',
 'Toby',
 'Rocky',
 'Baron',
 'Tyr',
 'Bauer',
 'Swagger',
 'Sammy',
 'Brandi',
 'None',
 'Mary',
 'Moe',
 'Ted',
 'Halo',
 'None',
 'Augie',
 'Craig',
 'Sam',
 'Hunter',
 'Pavlov',
 'Phil',
 'Gus',
 'None',
 'Maximus',
 'None',
 'Kyro',
 'Wallace',
 'Ito',
 'None',
 'Koda',
 'Seamus',
 'Milo',
 'None',
 'Cooper',
 'Ollie',
 'Stephan',
 'Cali',
 'Lennon',
 'None',
 'None',
 'None',
 'Waffles',
 'Dave',
 'incredibly',
 'Penny',
 'Major',
 'Duke',
 'Reginald',
 'Zeke',
 'Sansa',
 'Shooter',
 'Django',
 'None',
 'Rusty',
 'Bo',
 'Diogi',
 'None',
 'None',
 'Sonny',
 'Philbert',
 'Winston',
 'Marley',
 'None',
 'Bailey',
 'Winnie',
 'Severus',
 'None',
 'None',
 'Loki',
 'None',
 'Ronnie',
 'None',
 'Wallace',
 'None',
 'Milo',
 'Anakin',
 'Bones',
 'None',
 'None',
 'Mauve',
 'Chef',
 'None',
 'Sampson',
 'Doc',
 'Bo',
 'Peaches',
 'None',
 'Tucker',
 'Sobe',
 'Longfellow',
 'None',
 'Jeffrey',
 'Mister',
 'Iroh',
 'Shadow',
 'Baloo',
 'None',
 'Stubert',
 'None',
 'Jack',
 'None',
 'None',
 'Lola',
 'Paull',
 'None',
 'Timison',
 'None',
 'Davey',
 'Cooper',
 'None',
 'Cassie',
 'Pancake',
 'None',
 'Tyrone',
 'Tyr',
 'Romeo',
 'None',
 'None',
 'Snicku',
 'Ruby',
 'Ruby',
 'None',
 'None',
 'Yogi',
 'Daisy',
 'None',
 'Brody',
 'Bailey',
 'Rizzy',
 'Mack',
 'Butter',
 'Nimbus',
 'Laika',
 'Maximus',
 'Clark',
 'None',
 'Dobby',
 'Fiona',
 'Moreton',
 'Dave',
 'None',
 'Tucker',
 'Juno',
 'Maude',
 'Lily',
 'Newt',
 'Benji',
 'Nida',
 'None',
 'Robin',
 'a',
 'Bailey',
 'Monster',
 'BeBe',
 'Remus',
 'None',
 'None',
 'Maddie',
 'None',
 'None',
 'Levi',
 'Mabel',
 'Alfie',
 'Misty',
 'Betty',
 'Happy',
 'Mosby',
 'Duke',
 'Maggie',
 'Bruce',
 'Leela',
 'Happy',
 'Buddy',
 'Ralphy',
 'Eli',
 'Brownie',
 'Rizzy',
 'None',
 'Meyer',
 'Stella',
 'Bo',
 'Lucy',
 'Butter',
 'mad',
 'Dexter',
 'None',
 'Leo',
 'Bo',
 'None',
 'Frank',
 'Tonks',
 'Moose',
 'Lincoln',
 'Carl',
 'Rory',
 'Oakley',
 'Logan',
 'None',
 'Dale',
 'Rizzo',
 'Arnie',
 'Mattie',
 'None',
 'Scout',
 'Lucy',
 'Rusty',
 'Pinot',
 'Dallas',
 'None',
 'Doc',
 'Hero',
 'Rusty',
 'Frankie',
 'Stormy',
 'Reginald',
 'Balto',
 'Riley',
 'Mairi',
 'Loomis',
 'Finn',
 'Godi',
 'Kenny',
 'Dave',
 'Earl',
 'Cali',
 'Deacon',
 'Penny',
 'Timmy',
 'Sampson',
 'Harper',
 'Chipson',
 'None',
 'Combo',
 'None',
 'None',
 'Oakley',
 'None',
 'None',
 'Dash',
 'Koda',
 'Hercules',
 'None',
 'Bell',
 'None',
 'Bear',
 'None',
 'Hank',
 'None',
 'Scout',
 'None',
 'Hurley',
 'Reggie',
 'None',
 'Jay',
 'None',
 'None',
 'Mya',
 'Strider',
 'Penny',
 'None',
 'an',
 'Nala',
 'Stanley',
 'None',
 'Sophie',
 'Gerald',
 'Wesley',
 'None',
 'Arnie',
 'Derek',
 'Jeffrey',
 'None',
 'Solomon',
 'Huck',
 'very',
 'None',
 'O',
 'Sampson',
 'None',
 'None',
 'Blue',
 'Anakin',
 'None',
 'Finley',
 'Maximus',
 'None',
 'Tucker',
 'Finley',
 'Sprinkles',
 'None',
 'Winnie',
 'Heinrich',
 'Loki',
 'Shakespeare',
 'Chelsea',
 'Fizz',
 'Bungalo',
 'Chip',
 'Grey',
 'None',
 'Roosevelt',
 'Gromit',
 'a',
 'Willem',
 'None',
 'Jack',
 'Finn',
 'Penny',
 'None',
 'Davey',
 'Dakota',
 'Fizz',
 'Frankie',
 'Dixie',
 'Charlie',
 'None',
 'None',
 'Winston',
 'Sebastian',
 'None',
 'very',
 'Al',
 'Jackson',
 'just',
 'Carbon',
 'Klein',
 'Titan',
 'None',
 'DonDon',
 'Kirby',
 'None',
 'Jesse',
 'Lou',
 'Oakley',
 'Nollie',
 'Chevy',
 'Gerald',
 'Tito',
 'Philbert',
 'Louie',
 'None',
 'Rupert',
 'None',
 'Rufus',
 'None',
 'Brudge',
 'Shadoe',
 'Oscar',
 'Colby',
 'Juno',
 'Angel',
 'Brat',
 'Tove',
 'my',
 'Louie',
 'Gromit',
 'Aubie',
 'Kota',
 'None',
 'Alfie',
 'Clark',
 'Eve',
 'Belle',
 'Leela',
 'Glenn',
 'Buddy',
 'Scout',
 'None',
 'Shelby',
 'None',
 'None',
 'None',
 'Sephie',
 'None',
 'Bruce',
 'Bonaparte',
 'Albert',
 'Bo',
 'Wishes',
 'Rose',
 'Theo',
 'Atlas',
 'None',
 'Rocco',
 'Fido',
 'Sadie',
 'None',
 'None',
 'None',
 'Kirby',
 'Maggie',
 'None',
 'Emma',
 'Oakley',
 'None',
 'Luna',
 'None',
 'Toby',
 'Spencer',
 'Lilli',
 'None',
 'Boston',
 'Brandonald',
 'None',
 'Odie',
 'Corey',
 'None',
 'None',
 'Leonard',
 'Chompsky',
 'Beckham',
 'Cooper',
 'None',
 'None',
 'None',
 'None',
 'Devón',
 'Oliver',
 'Jax',
 'Gert',
 'None',
 'None',
 'None',
 'None',
 'None',
 'one',
 'Watson',
 'Rubio',
 'Winnie',
 'Keith',
 'Milo',
 'Dex',
 'None',
 'Charlie',
 'None',
 'None',
 'Scout',
 'Hank',
 'Carly',
 'Ace',
 'None',
 'Tayzie',
 'Carl',
 'Grizzie',
 'None',
 'None',
 'None',
 'None',
 'None',
 'None',
 'None',
 'Brody',
 'Lola',
 'Ruby',
 'Tucker',
 'Fred',
 'Toby',
 'None',
 'Max',
 'None',
 'Gilbert',
 'None',
 'Cooper',
 'Milo',
 'Meyer',
 'Malcolm',
 'Arnie',
 'Zoe',
 'None',
 'None',
 'Stewie',
 'Calvin',
 'Lilah',
 'Spanky',
 'None',
 'Jameson',
 'Beau',
 'Jax',
 'Piper',
 'Bo',
 'Atticus',
 'Lucy',
 'Finn',
 'None',
 'George',
 'Blu',
 'Boomer',
 'Winston',
 'Dietrich',
 'not',
 'Divine',
 'None',
 'Tripp',
 'his',
 'one',
 'Cora',
 'None',
 'None',
 'Duke',
 'None',
 'None',
 …]
```

[131]:
```
df_twt_arch.query('name == "O"')
```

[131]:
```
               tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
775  776201521193218049                    NaN                  NaN

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

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

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

     retweeted_status_user_id retweeted_status_timestamp  \
775                       NaN                        NaN

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

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

```
[132]: # Check for unique values in retweet_status
       df_json.retweet_status.unique()
```

```
[132]: array(['Original tweet'], dtype=object)
```

Back To Quality list

issue8 Quality

```
[133]: # Check for unique values rating_numerator
       df_twt_arch.rating_numerator.unique()
```

```
[133]: array([  13,   12,   14,    5,   17,   11,   10,  420,  666,    6,   15,
               182,  960,    0,   75,    7,   84,    9,   24,    8,    1,   27,
                 3,    4,  165, 1776,  204,   50,   99,   80,   45,   60,   44,
               143,  121,   20,   26,    2,  144,   88])
```

```
[134]: # Check for unique values rating_numerator
       df_twt_arch.rating_denominator.unique()
```

```
[134]: array([ 10,   0,  15,  70,   7,  11, 150, 170,  20,  50,  90,  80,  40,
              130, 110,  16, 120,   2])
```

issue5 Quality

```
[135]: # Get source column values
       np.array(df_twt_arch.source)
```

```
[135]: array(['<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              …,
              '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>'],
             dtype=object)
```

```
[136]: # Check of unique value in source column
       df_twt_arch.source.unique()
```

```
[136]: array(['<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
       iPhone</a>',
              '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
        '<a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>',
        '<a href="https://about.twitter.com/products/tweetdeck"
rel="nofollow">TweetDeck</a>'],
      dtype=object)
```

Back To Quality list

## Cleaning Data In this section, clean **all** of the issues you documented while assessing.

Note: Make a copy of the original data before cleaning. Cleaning includes merging individual pieces of data according to the rules of tidy data. The result should be a high-quality and tidy master pandas DataFrame (or DataFrames, if appropriate).

[137]:

```
# Make copies of original dataframes
df_tw = df_twt_arch.copy()
df_js = df_json.copy()
df_mg = df_img.copy()
```

[138]:

```
df_tw
```

[138]:

```
                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
…                    …                    …                  …
2351  666049248165822465                    NaN                  NaN
2352  666044226329800704                    NaN                  NaN
2353  666033412701032449                    NaN                  NaN
2354  666029285002620928                    NaN                  NaN
2355  666020888022790149                    NaN                  NaN

                      timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
…                           …
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

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

```
2     <a href="http://twitter.com/download/iphone" r…
3     <a href="http://twitter.com/download/iphone" r…
4     <a href="http://twitter.com/download/iphone" r…
…                                                 …
2351  <a href="http://twitter.com/download/iphone" r…
2352  <a href="http://twitter.com/download/iphone" r…
2353  <a href="http://twitter.com/download/iphone" r…
2354  <a href="http://twitter.com/download/iphone" r…
2355  <a href="http://twitter.com/download/iphone" r…

                                                   text  retweeted_status_id  \
0     This is Phineas. He's a mystical boy. Only eve…                  NaN
1     This is Tilly. She's just checking pup on you…                  NaN
2     This is Archie. He is a rare Norwegian Pouncin…                  NaN
3     This is Darla. She commenced a snooze mid meal…                  NaN
4     This is Franklin. He would like you to stop ca…                  NaN
…                                                 …                  …
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
…                        …                        …
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
…                                                 …               …
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
…                    …        …   …     …     …   …
2351                  10      None  None    None   None  None
2352                  10         a  None    None   None  None
2353                  10         a  None    None   None  None
2354                  10         a  None    None   None  None
2355                  10      None  None    None   None  None

[2356 rows x 17 columns]
```

# Quality issue fixed

### Issue #1:

In columns: (‘doggo’, ‘floofer’, ‘pupper’, ‘puppo’, ‘name’) ‘None’ assigned instead of ‘NaN’ for empty missing data

Define:

Replace None value by Nan using

Code

[139]:
```
#replace None value
df_tw = df_tw.replace('None', np.nan)
```

Test

[140]:
```
df_tw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   tweet_id                    2356 non-null   int64
 1   in_reply_to_status_id       78 non-null     float64
 2   in_reply_to_user_id         78 non-null     float64
 3   timestamp                   2356 non-null   object
 4   source                      2356 non-null   object
 5   text                        2356 non-null   object
 6   retweeted_status_id         181 non-null    float64
 7   retweeted_status_user_id    181 non-null    float64
 8   retweeted_status_timestamp  181 non-null    object
 9   expanded_urls               2297 non-null   object
 10  rating_numerator            2356 non-null   int64
 11  rating_denominator          2356 non-null   int64
 12  name                        1611 non-null   object
 13  doggo                       97 non-null     object
 14  floofer                     10 non-null     object
 15  pupper                      257 non-null    object
 16  puppo                       30 non-null     object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

Back to Quality list

### Issue #2: > Columns not needed: (‘in_reply_to_status_id’, ‘in_reply_to_user_id’,‘retweeted_status_id’, ‘retweeted_status_user_id’,‘retweeted_status_timestamp’)

Define

> Drop the columns (‘in_reply_to_status_id’, ‘in_reply_to_user_id’,‘retweeted_status_id’, ‘retweeted_status_user_id’,‘retweeted_status_timestamp’)

Code

[141]:
```
# First we Remove Retweet rows from df_tw
df_tw = df_tw[df_tw.retweeted_status_id.isnull()]
df_tw = df_tw[df_tw.retweeted_status_user_id.isnull()]
df_tw = df_tw[df_tw.retweeted_status_timestamp.isnull()]
df_tw = df_tw[df_tw.in_reply_to_status_id.isnull()]
df_tw = df_tw[df_tw.in_reply_to_user_id.isnull()]
```

[142]:
```
#Drop columns
df_tw.drop(columns=[
    'in_reply_to_status_id',
    'in_reply_to_user_id',
    'retweeted_status_id',
    'retweeted_status_user_id',
    'retweeted_status_timestamp'
], inplace=True)
```

[143]:
```
# Rename col in df_tw
columns ={
    'source': 'tweet_source',
    'text':'tweet_text',
    'name':'dog_name',
}
df_tw.rename(columns=columns, inplace=True)
```

Test

[144]:
```
df_tw.head(1)
```

```
[144]:              tweet_id                  timestamp  \
        0  892420643555336193  2017-08-01 16:23:56 +0000

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

                                                tweet_text  \
        0  This is Phineas. He's a mystical boy. Only eve…

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

           rating_denominator dog_name doggo floofer pupper puppo
        0                  10  Phineas   NaN     NaN    NaN   NaN
```

Back to Quality list

### Issue #3: > Column timestamp dtype should be datetime

Define

> Change ‘timestamp’ datatype from int to datetime

Code

```
[145]: # Convert 'timestamp' dtype to datetime To facilitate the split
       df_tw['timestamp'] = pd.to_datetime(df_tw.timestamp)
```

```
[146]: df_tw['tweet_date'] = [d.date() for d in df_tw['timestamp']]
       df_tw['tweet_time'] = df_tw['timestamp'].dt.hour
```

```
[147]: # Convert 'tweet_date' dtype to datetime
       df_tw['tweet_date'] = pd.to_datetime(df_tw.tweet_date)
```

```
[148]: #Drop timestamp col
       df_tw.drop(columns='timestamp', inplace=True)
```

Test

```
[149]: #Check dtypes
       df_tw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2097 entries, 0 to 2355
Data columns (total 13 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   tweet_id            2097 non-null   int64
 1   tweet_source        2097 non-null   object
 2   tweet_text          2097 non-null   object
 3   expanded_urls       2094 non-null   object
 4   rating_numerator    2097 non-null   int64
 5   rating_denominator  2097 non-null   int64
 6   dog_name            1494 non-null   object
 7   doggo               83 non-null     object
 8   floofer             10 non-null     object
 9   pupper              230 non-null    object
 10  puppo               24 non-null     object
 11  tweet_date          2097 non-null   datetime64[ns]
 12  tweet_time          2097 non-null   int64
dtypes: datetime64[ns](1), int64(4), object(8)
memory usage: 229.4+ KB
```

[150]:
```
#check new cols
df_tw.head()
```

[150]:
```
             tweet_id                                        tweet_source  \
0  892420643555336193  <a href="http://twitter.com/download/iphone" r…
1  892177421306343426  <a href="http://twitter.com/download/iphone" r…
2  891815181378084864  <a href="http://twitter.com/download/iphone" r…
3  891689557279858688  <a href="http://twitter.com/download/iphone" r…
4  891327558926688256  <a href="http://twitter.com/download/iphone" r…

                                         tweet_text  \
0  This is Phineas. He's a mystical boy. Only eve…
1  This is Tilly. She's just checking pup on you…
2  This is Archie. He is a rare Norwegian Pouncin…
3  This is Darla. She commenced a snooze mid meal…
4  This is Franklin. He would like you to stop ca…

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

   rating_denominator  dog_name doggo floofer pupper puppo tweet_date  \
0                  10   Phineas   NaN     NaN    NaN   NaN 2017-08-01
1                  10     Tilly   NaN     NaN    NaN   NaN 2017-08-01
2                  10    Archie   NaN     NaN    NaN   NaN 2017-07-31
3                  10     Darla   NaN     NaN    NaN   NaN 2017-07-30
4                  10  Franklin   NaN     NaN    NaN   NaN 2017-07-29

   tweet_time
0          16
1           0
2           0
```

```
3            15
4            16
```

Back to Quality list

### Issue #4: >‘tweet_id’ must be a string

Define

> Change ‘tweet_id’ datatype from int to string in all dfs

Code

```
[151]: #  Convert 'tweet_id' dtype to str in all dfs
       df_tw['tweet_id'] = df_tw['tweet_id'].astype(str)
       df_js['tweet_id'] = df_js['tweet_id'].astype(str)
       df_mg['tweet_id'] = df_mg['tweet_id'].astype(str)
```

Test

```
[152]: # Check dtypes
       df_tw['tweet_id'].dtypes, df_js['tweet_id'].dtypes, df_mg['tweet_id'].dtypes
```

```
[152]: (dtype('O'), dtype('O'), dtype('O'))
```

Back to Quality list

### Issue #5: >‘source’ column contains tag html

Define

> Replace the Hyperlink by the innerHTML of it

Code

```
[153]: # Replace tag link with str
       df_tw['tweet_source'].replace('<a href="http://twitter.com/download/iphone"␣
        ↪rel="nofollow">Twitter for iPhone</a>'
                                               ,'Twitter for iPhone', inplace = True)
       df_tw['tweet_source'].replace('<a href="http://twitter.com"␣
        ↪rel="nofollow">Twitter Web Client</a>'
                                               ,'Twitter Web Client', inplace = True)
       df_tw['tweet_source'].replace('<a href="http://vine.co" rel="nofollow">Vine -␣
        ↪Make a Scene</a>'
                                               ,'Vine - Make a Scene', inplace = True)
       df_tw['tweet_source'].replace('<a href="https://about.twitter.com/products/
        ↪tweetdeck" rel="nofollow">TweetDeck</a>'
                                               ,'TweetDeck', inplace = True)
```

Test

```
[154]: df_tw['tweet_source'].unique()
```

```
[154]: array(['Twitter for iPhone', 'Twitter Web Client', 'Vine - Make a Scene',
              'TweetDeck'], dtype=object)
```

Back to Quality list

### Issue #6: > column 'name' has values: 'None', 'a', 'O', 'Mo'.

Define

- None value has changed in previous section to NaN
- Correct wired names

Code

```
[155]: # Check unique name value
       wirred_name = df_tw[df_tw['dog_name'].str.len() <= 2]['dog_name'].unique()
       wirred_name
```

```
[155]: array(['a', 'Mo', 'Bo', 'O', 'Al', 'my', 'an', 'by', 'Ed', 'JD', 'Jo'],
             dtype=object)
```

https://knowledge.udacity.com/questions/459296

```
[156]: # 1. Fill in NaN values with the word lower
       # 2. Check for all observations that are lowercase
       # 3. Use that as your mask for the next stage
       # mask = df_tw.name.fillna("lower").str.islower()
       lower = []
       for i in df_tw['dog_name'].unique():
           #convert i to str 'float' has no attribute 'islower'
           if str(i).islower() == True:
               lower.append(i)
       lower
```

```
[156]: [nan,
        'such',
        'a',
        'quite',
        'not',
        'one',
        'incredibly',
        'very',
        'my',
        'his',
        'an',
        'actually',
        'just',
        'getting',
        'mad',
        'this',
        'unacceptable',
        'all',
        'old',
        'infuriating',
        'the',
        'by',
        'officially',
        'life',
        'light',
        'space']
```

[157]:

```python
# Replacing these names with NaN value
df_tw['dog_name'].replace(lower, np.nan, inplace = True)
```

[158]:

```python
df_tw['dog_name'].unique()
```

[158]:

```
array(['Phineas', 'Tilly', 'Archie', 'Darla', 'Franklin', nan, 'Jax',
       'Zoey', 'Cassie', 'Koda', 'Bruno', 'Ted', 'Stuart', 'Oliver',
       'Jim', 'Zeke', 'Ralphus', 'Gerald', 'Jeffrey', 'Canela', 'Maya',
       'Mingus', 'Derek', 'Roscoe', 'Waffles', 'Jimbo', 'Maisey', 'Earl',
       'Lola', 'Kevin', 'Yogi', 'Noah', 'Bella', 'Grizzwald', 'Rusty',
       'Gus', 'Stanley', 'Alfy', 'Koko', 'Rey', 'Gary', 'Elliot', 'Louis',
       'Jesse', 'Romeo', 'Bailey', 'Duddles', 'Jack', 'Steven', 'Beau',
       'Snoopy', 'Shadow', 'Emmy', 'Aja', 'Penny', 'Dante', 'Nelly',
       'Ginger', 'Benedict', 'Venti', 'Goose', 'Nugget', 'Cash', 'Jed',
       'Sebastian', 'Sierra', 'Monkey', 'Harry', 'Kody', 'Lassie',
       'Rover', 'Napolean', 'Boomer', 'Cody', 'Rumble', 'Clifford',
       'Dewey', 'Scout', 'Gizmo', 'Walter', 'Cooper', 'Harold', 'Shikha',
       'Lili', 'Jamesy', 'Coco', 'Sammy', 'Meatball', 'Paisley', 'Albus',
       'Neptune', 'Belle', 'Quinn', 'Zooey', 'Dave', 'Jersey', 'Hobbes',
       'Burt', 'Lorenzo', 'Carl', 'Jordy', 'Milky', 'Trooper', 'Sophie',
       'Wyatt', 'Rosie', 'Thor', 'Oscar', 'Callie', 'Cermet', 'Marlee',
       'Arya', 'Einstein', 'Alice', 'Rumpole', 'Benny', 'Aspen', 'Jarod',
       'Wiggles', 'General', 'Sailor', 'Iggy', 'Snoop', 'Kyle', 'Leo',
       'Riley', 'Noosh', 'Odin', 'Jerry', 'Georgie', 'Rontu', 'Cannon',
       'Furzey', 'Daisy', 'Tuck', 'Barney', 'Vixen', 'Jarvis', 'Mimosa',
       'Pickles', 'Brady', 'Luna', 'Charlie', 'Margo', 'Sadie', 'Hank',
       'Tycho', 'Indie', 'Winnie', 'George', 'Bentley', 'Max', 'Dawn',
       'Maddie', 'Monty', 'Sojourner', 'Winston', 'Odie', 'Arlo',
       'Vincent', 'Lucy', 'Clark', 'Mookie', 'Meera', 'Ava', 'Eli', 'Ash',
       'Tucker', 'Tobi', 'Chester', 'Wilson', 'Sunshine', 'Lipton',
       'Bronte', 'Poppy', 'Gidget', 'Rhino', 'Willow', 'Orion', 'Eevee',
       'Smiley', 'Miguel', 'Emanuel', 'Kuyu', 'Dutch', 'Pete', 'Scooter',
       'Reggie', 'Lilly', 'Samson', 'Mia', 'Astrid', 'Malcolm', 'Dexter',
       'Alfie', 'Fiona', 'Mutt', 'Bear', 'Doobert', 'Beebop', 'Alexander',
       'Sailer', 'Brutus', 'Kona', 'Boots', 'Ralphie', 'Loki', 'Cupid',
       'Pawnd', 'Pilot', 'Ike', 'Mo', 'Toby', 'Sweet', 'Pablo', 'Nala',
       'Crawford', 'Gabe', 'Jimison', 'Duchess', 'Harlso', 'Sundance',
       'Luca', 'Flash', 'Sunny', 'Howie', 'Jazzy', 'Anna', 'Finn', 'Bo',
       'Wafer', 'Tom', 'Florence', 'Autumn', 'Buddy', 'Dido', 'Eugene',
       'Ken', 'Strudel', 'Tebow', 'Chloe', 'Timber', 'Binky', 'Moose',
       'Dudley', 'Comet', 'Akumi', 'Titan', 'Olivia', 'Alf', 'Oshie',
       'Chubbs', 'Sky', 'Atlas', 'Eleanor', 'Layla', 'Rocky', 'Baron',
       'Tyr', 'Bauer', 'Swagger', 'Brandi', 'Mary', 'Moe', 'Halo',
       'Augie', 'Craig', 'Sam', 'Hunter', 'Pavlov', 'Phil', 'Kyro',
       'Wallace', 'Ito', 'Seamus', 'Ollie', 'Stephan', 'Lennon', 'Major',
       'Duke', 'Sansa', 'Shooter', 'Django', 'Diogi', 'Sonny', 'Marley',
       'Severus', 'Ronnie', 'Milo', 'Bones', 'Mauve', 'Chef', 'Doc',
       'Peaches', 'Sobe', 'Longfellow', 'Mister', 'Iroh', 'Pancake',
       'Snicku', 'Ruby', 'Brody', 'Mack', 'Nimbus', 'Laika', 'Maximus',
       'Dobby', 'Moreton', 'Juno', 'Maude', 'Lily', 'Newt', 'Benji',
       'Nida', 'Robin', 'Monster', 'BeBe', 'Remus', 'Levi', 'Mabel',
       'Misty', 'Betty', 'Mosby', 'Maggie', 'Bruce', 'Happy', 'Ralphy',
       'Brownie', 'Rizzy', 'Stella', 'Butter', 'Frank', 'Tonks',
       'Lincoln', 'Rory', 'Logan', 'Dale', 'Rizzo', 'Arnie', 'Mattie',
       'Pinot', 'Dallas', 'Hero', 'Frankie', 'Stormy', 'Reginald',
       'Balto', 'Mairi', 'Loomis', 'Godi', 'Cali', 'Deacon', 'Timmy',
       'Sampson', 'Chipson', 'Combo', 'Oakley', 'Dash', 'Hercules', 'Jay',
       'Mya', 'Strider', 'Wesley', 'Solomon', 'Huck', 'O', 'Blue',
       'Anakin', 'Finley', 'Sprinkles', 'Heinrich', 'Shakespeare',
       'Chelsea', 'Bungalo', 'Chip', 'Grey', 'Roosevelt', 'Willem',
       'Davey', 'Dakota', 'Fizz', 'Dixie', 'Al', 'Jackson', 'Carbon',
       'Klein', 'DonDon', 'Kirby', 'Lou', 'Chevy', 'Tito', 'Philbert',
       'Louie', 'Rupert', 'Rufus', 'Brudge', 'Shadoe', 'Angel', 'Brat',
       'Tove', 'Gromit', 'Aubie', 'Kota', 'Leela', 'Glenn', 'Shelby',
       'Sephie', 'Bonaparte', 'Albert', 'Wishes', 'Rose', 'Theo', 'Rocco',
       'Fido', 'Emma', 'Spencer', 'Lilli', 'Boston', 'Brandonald',
       'Corey', 'Leonard', 'Beckham', 'Devón', 'Gert', 'Watson', 'Keith',
       'Dex', 'Ace', 'Tayzie', 'Grizzie', 'Fred', 'Gilbert', 'Meyer',
       'Zoe', 'Stewie', 'Calvin', 'Lilah', 'Spanky', 'Jameson', 'Piper',
       'Atticus', 'Blu', 'Dietrich', 'Divine', 'Tripp', 'Cora', 'Huxley',
       'Keurig', 'Bookstore', 'Linus', 'Abby', 'Shiloh', 'Gustav',
       'Arlen', 'Percy', 'Lenox', 'Sugar', 'Harvey', 'Blanket', 'Geno',
       'Stark', 'Beya', 'Kilo', 'Kayla', 'Maxaroni', 'Bell', 'Doug',
       'Edmund', 'Aqua', 'Theodore', 'Baloo', 'Chase', 'Nollie', 'Rorie',
       'Simba', 'Charles', 'Bayley', 'Axel', 'Storkson', 'Remy',
       'Chadrick', 'Kellogg', 'Buckley', 'Livvie', 'Terry', 'Hermione',
       'Ralpher', 'Aldrick', 'Larry', 'Rooney', 'Crystal', 'Ziva',
       'Stefan', 'Pupcasso', 'Puff', 'Flurpson', 'Coleman', 'Enchilada',
       'Raymond', 'Rueben', 'Cilantro', 'Karll', 'Sprout', 'Blitz',
       'Bloop', 'Colby', 'Lillie', 'Ashleigh', 'Kreggory', 'Sarge',
       'Luther', 'Ivar', 'Jangle', 'Schnitzel', 'Panda', 'Berkeley',
       'Ralphé', 'Charleson', 'Clyde', 'Harnold', 'Sid', 'Pippa', 'Otis',
       'Carper', 'Bowie', 'Alexanderson', 'Suki', 'Barclay', 'Skittle',
       'Ebby', 'Flávio', 'Smokey', 'Link', 'Jennifur', 'Ozzy', 'Bluebert',
       'Stephanus', 'Bubbles', 'Zeus', 'Bertson', 'Nico',
       'Michelangelope', 'Siba', 'Calbert', 'Curtis', 'Travis', 'Thumas',
       'Kanu', 'Lance', 'Opie', 'Stubert', 'Kane', 'Olive', 'Chuckles',
       'Staniel', 'Sora', 'Beemo', 'Gunner', 'Lacy', 'Tater', 'Olaf',
       'Cecil', 'Vince', 'Karma', 'Billy', 'Walker', 'Rodney', 'Klevin',
       'Malikai', 'Bobble', 'River', 'Jebberson', 'Remington', 'Farfle',
       'Jiminus', 'Harper', 'Clarkus', 'Finnegus', 'Cupcake', 'Kathmandu',
       'Ellie', 'Katie', 'Kara', 'Adele', 'Zara', 'Ambrose', 'Jimothy',
       'Bode', 'Terrenth', 'Reese', 'Chesterson', 'Lucia', 'Bisquick',
       'Ralphson', 'Socks', 'Rambo', 'Rudy', 'Fiji', 'Rilo', 'Bilbo',
       'Coopson', 'Yoda', 'Millie', 'Chet', 'Crouton', 'Daniel', 'Kaia',
       'Murphy', 'Dotsy', 'Eazy', 'Coops', 'Fillup', 'Miley', 'Charl',
       'Reagan', 'Yukon', 'CeCe', 'Cuddles', 'Claude', 'Jessiga',
       'Carter', 'Ole', 'Pherb', 'Blipson', 'Reptar', 'Trevith', 'Berb',
       'Bob', 'Colin', 'Brian', 'Oliviér', 'Grady', 'Kobe', 'Freddery',
       'Bodie', 'Dunkin', 'Wally', 'Tupawc', 'Amber', 'Herschel', 'Edgar',
       'Teddy', 'Kingsley', 'Brockly', 'Richie', 'Molly', 'Vinscent',
       'Cedrick', 'Hazel', 'Lolo', 'Eriq', 'Phred', 'Oddie', 'Maxwell',
       'Geoff', 'Covach', 'Durg', 'Fynn', 'Ricky', 'Herald', 'Lucky',
       'Ferg', 'Trip', 'Clarence', 'Hamrick', 'Brad', 'Pubert', 'Frönq',
       'Derby', 'Lizzie', 'Ember', 'Blakely', 'Opal', 'Marq', 'Kramer',
       'Barry', 'Tyrone', 'Gordon', 'Baxter', 'Mona', 'Horace', 'Crimson',
       'Birf', 'Hammond', 'Lorelei', 'Marty', 'Brooks', 'Petrick',
       'Hubertson', 'Gerbald', 'Oreo', 'Bruiser', 'Perry', 'Bobby',
       'Jeph', 'Obi', 'Tino', 'Kulet', 'Sweets', 'Lupe', 'Tiger',
       'Jiminy', 'Griffin', 'Banjo', 'Brandy', 'Lulu', 'Darrel', 'Taco',
       'Joey', 'Patrick', 'Kreg', 'Todo', 'Tess', 'Ulysses', 'Toffee',
       'Apollo', 'Carly', 'Asher', 'Glacier', 'Chuck', 'Champ', 'Ozzie',
       'Griswold', 'Cheesy', 'Moofasa', 'Hector', 'Goliath', 'Kawhi',
       'Emmie', 'Penelope', 'Willie', 'Rinna', 'Mike', 'William',
       'Dwight', 'Evy', 'Hurley', 'Rubio', 'Chompsky', 'Rascal', 'Linda',
       'Tug', 'Tango', 'Grizz', 'Jerome', 'Crumpet', 'Jessifer', 'Izzy',
       'Ralph', 'Sandy', 'Humphrey', 'Tassy', 'Juckson', 'Chuq', 'Tyrus',
       'Karl', 'Godzilla', 'Vinnie', 'Kenneth', 'Herm', 'Bert', 'Striker',
       'Donny', 'Pepper', 'Bernie', 'Buddah', 'Lenny', 'Arnold', 'Zuzu',
       'Mollie', 'Laela', 'Tedders', 'Superpup', 'Rufio', 'Jeb', 'Rodman',
       'Jonah', 'Chesney', 'Kenny', 'Henry', 'Bobbay', 'Mitch', 'Kaiya',
       'Acro', 'Aiden', 'Obie', 'Dot', 'Shnuggles', 'Kendall', 'Jeffri',
       'Steve', 'Eve', 'Mac', 'Fletcher', 'Kenzie', 'Pumpkin', 'Schnozz',
       'Gustaf', 'Cheryl', 'Ed', 'Leonidas', 'Norman', 'Caryl', 'Scott',
       'Taz', 'Darby', 'Jackie', 'Jazz', 'Franq', 'Pippin', 'Rolf',
       'Snickers', 'Ridley', 'Cal', 'Bradley', 'Bubba', 'Tuco', 'Patch',
       'Mojo', 'Batdog', 'Dylan', 'Mark', 'JD', 'Alejandro', 'Scruffers',
       'Pip', 'Julius', 'Tanner', 'Sparky', 'Anthony', 'Holly', 'Jett',
       'Amy', 'Sage', 'Andy', 'Mason', 'Trigger', 'Antony', 'Creg',
       'Traviss', 'Gin', 'Jeffrie', 'Danny', 'Ester', 'Pluto', 'Bloo',
       'Edd', 'Paull', 'Willy', 'Herb', 'Damon', 'Peanut', 'Nigel',
       'Butters', 'Sandra', 'Fabio', 'Randall', 'Liam', 'Tommy', 'Ben',
       'Raphael', 'Julio', 'Andru', 'Kloey', 'Shawwn', 'Skye', 'Kollin',
       'Ronduh', 'Billl', 'Saydee', 'Dug', 'Sully', 'Kirk', 'Ralf',
       'Clarq', 'Jaspers', 'Samsom', 'Terrance', 'Harrison', 'Chaz',
       'Jeremy', 'Jaycob', 'Lambeau', 'Ruffles', 'Amélie', 'Bobb',
       'Banditt', 'Kevon', 'Winifred', 'Hanz', 'Churlie', 'Zeek',
       'Timofy', 'Maks', 'Jomathan', 'Kallie', 'Marvin', 'Spark',
       'Gòrdón', 'Jo', 'DayZ', 'Jareld', 'Torque', 'Ron', 'Skittles',
       'Cleopatricia', 'Erik', 'Stu', 'Tedrick', 'Shaggy', 'Filup',
       'Kial', 'Naphaniel', 'Dook', 'Hall', 'Philippe', 'Biden', 'Fwed',
       'Genevieve', 'Joshwa', 'Timison', 'Bradlay', 'Pipsy', 'Clybe',
       'Keet', 'Carll', 'Jockson', 'Josep', 'Lugan', 'Christoper'],
      dtype=object)
```

[159]:

```
# Display wired names
df_tw[df_tw['dog_name'].
 ↪isin(wirred_name)][['dog_name','tweet_text','expanded_urls']]
```

[159]:

```
     dog_name                                         tweet_text  \
393        Mo  This is Mo. No one will push him around in the…
449        Bo  This is Bo. He was a very good First Doggo. 14…
553        Bo  This is Bo. He's going to make me cry. 13/10 p…
679        Bo  This is Bo. He's a West Congolese Bugaboop Snu…
775         O  This is O'Malley. That is how he sleeps. Doesn…
820        Al  Meet Al Cabone. He's a gangsta puppa. Rather h…
876        Bo  This is Bo and Ty. Bo eats paper and Ty felt l…
978        Bo  This is Bo. He emanates happiness. 12/10 I cou…
1683       Bo  This is Bo. He's a Benedoop Cumbersnatch. Seem…
1988       Ed  This is Ed. He's not mad, just disappointed. 1…
2041       JD  This is JD (stands for "just dog"). He's like …
2219       Jo  This is Jo. Jo is a Swedish Queso. Tongue bigg…

                                          expanded_urls
393   https://twitter.com/dog_rates/status/825876512…
449   https://twitter.com/dog_rates/status/819004803…
553   https://twitter.com/dog_rates/status/804026241…
679   https://twitter.com/dog_rates/status/789137962…
775   https://twitter.com/dog_rates/status/776201521…
820   https://twitter.com/dog_rates/status/770414278…
876   https://twitter.com/dog_rates/status/761004547…
978   https://twitter.com/dog_rates/status/749996283…
1683  https://twitter.com/dog_rates/status/681694085…
1988  https://twitter.com/dog_rates/status/672834301…
2041  https://twitter.com/dog_rates/status/671542985…
2219  https://twitter.com/dog_rates/status/668496999…
```

```
[160]: # Get all wired names indeces
       w_name_index = (df_tw[df_tw['dog_name'].
        ↪isin(wirred_name)][['dog_name','tweet_text']]).index
       w_name_index
```

```
[160]: Int64Index([393, 449, 553, 679, 775, 820, 876, 978, 1683, 1988, 2041, 2219],
       dtype='int64')
```

```
[161]: # Print each name in w_name_index with thier indecies
       for i in w_name_index:
           print(i, df_tw.loc[i,'dog_name'],':', df_tw.loc[i,'tweet_text'], '\n')
```

```
393 Mo : This is Mo. No one will push him around in the grocery cart. He's quite
pupset about it. 11/10 I volunteer https://t.co/feNwTq12S5

449 Bo : This is Bo. He was a very good First Doggo. 14/10 would be an absolute
honor to pet https://t.co/AdPKrI8BZ1

553 Bo : This is Bo. He's going to make me cry. 13/10 please get off the bus for
him Carly https://t.co/U7FvBZo6Bq

679 Bo : This is Bo. He's a West Congolese Bugaboop Snuggle. Rather exotic.
Master of the head tilt. 12/10 would pay to pet https://t.co/2jwxxtNzoN

775 O : This is O'Malley. That is how he sleeps. Doesn't care what you think
about it. 10/10 comfy af https://t.co/Pq150LeRaC

820 Al : Meet Al Cabone. He's a gangsta puppa. Rather h*ckin ruthless. Shows no
mercy sometimes. 11/10 pet w extreme caution https://t.co/OUwWbEKOUV

876 Bo : This is Bo and Ty. Bo eats paper and Ty felt left out. 11/10 for both
https://t.co/1acHQS8rvK

978 Bo : This is Bo. He emanates happiness. 12/10 I could cut the freedom with a
knife https://t.co/c7LNFt39eR

1683 Bo : This is Bo. He's a Benedoop Cumbersnatch. Seems frustrated with own
feet. Portable as hell. 11/10 very solid pupper https://t.co/TONMhRoQh7

1988 Ed : This is Ed. He's not mad, just disappointed. 10/10
https://t.co/BIljUOzhLN

2041 JD : This is JD (stands for "just dog"). He's like Airbud but with trading
card games instead of sports. 10/10 much skill https://t.co/zzueJV9jCF

2219 Jo : This is Jo. Jo is a Swedish Queso. Tongue bigger than face. Tiny lil
legs. Still no seatbelt. Simply careless. 8/10 https://t.co/Edy7B5vOp2
```

I checked each link from the previous cell I found that there are just two requiered changes for: - 'O' - 'Al'

All the other like ('Mo', 'Bo', 'Ed', 'JD', 'Jo') are real names

```
[162]: df_tw["dog_name"].replace({'O':"O'Malley",'Al':'Al Cabone'}, inplace=True)
```

```
[163]: # Drop tweets based on observed text
       # using ~ before df show the opposite
       df_tw = df_tw[~df_tw.tweet_text.str.contains(
           "without dogs|Please only send dogs|Please don't send|please stop sending␣
        ↪pictures without"
       )]
       df_tw
```

```
[163]:                 tweet_id        tweet_source  \
       0     892420643555336193  Twitter for iPhone
       1     892177421306343426  Twitter for iPhone
       2     891815181378084864  Twitter for iPhone
       3     891689557279858688  Twitter for iPhone
       4     891327558926688256  Twitter for iPhone
       …                    …                   …
       2351  666049248165822465  Twitter for iPhone
       2352  666044226329800704  Twitter for iPhone
       2353  666033412701032449  Twitter for iPhone
       2354  666029285002620928  Twitter for iPhone
       2355  666020888022790149  Twitter for iPhone

                                                      tweet_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…
       …                                                   …
       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
       …                                                   …               …
```

```
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  dog_name doggo floofer pupper puppo tweet_date  \
0                     10   Phineas   NaN     NaN    NaN   NaN 2017-08-01
1                     10     Tilly   NaN     NaN    NaN   NaN 2017-08-01
2                     10    Archie   NaN     NaN    NaN   NaN 2017-07-31
3                     10     Darla   NaN     NaN    NaN   NaN 2017-07-30
4                     10  Franklin   NaN     NaN    NaN   NaN 2017-07-29
…                  …       …   …     …    …   …        …
2351                  10       NaN   NaN     NaN    NaN   NaN 2015-11-16
2352                  10       NaN   NaN     NaN    NaN   NaN 2015-11-16
2353                  10       NaN   NaN     NaN    NaN   NaN 2015-11-15
2354                  10       NaN   NaN     NaN    NaN   NaN 2015-11-15
2355                  10       NaN   NaN     NaN    NaN   NaN 2015-11-15

      tweet_time
0             16
1              0
2              0
3             15
4             16
…          …
2351           0
2352           0
2353          23
2354          23
2355          22

[2077 rows x 13 columns]
```

Test

```
[164]: #Those are real names of dogs according to thier tweet links
df_tw[df_tw['dog_name'].str.len() <= 2]['dog_name'].unique()
```

```
[164]: array(['Mo', 'Bo', 'Ed', 'JD', 'Jo'], dtype=object)
```

```
[165]: # Check changing names
df_tw.loc[775,'dog_name'] , df_tw.loc[820,'dog_name']
```

```
[165]: ("O'Malley", 'Al Cabone')
```

```
[166]: #check name column values
df_tw['dog_name'].value_counts()
```

```
[166]: Charlie       11
       Lucy          11
       Oliver        10
       Cooper        10
       Penny          9
                     ..
       Shelby         1
       Sephie         1
       Bonaparte      1
       Wishes         1
       Christoper     1
       Name: dog_name, Length: 929, dtype: int64
```

Back to Quality list

### Issue #7: > expanded_urls has missing value and inccorrect urls

Define

> drop null value and fix other urls

Code

```
[167]: #Drop null expanded_urls
       df_tw.dropna(subset=["expanded_urls"], inplace=True)
```

```
[168]: df_tw["expanded_urls"].isnull().sum()
```

```
[168]: 0
```

```
[169]: # replace exist value with new value to combine them later
       df_tw["expanded_urls"]="https://twitter.com/dog_rates/status/"
```

```
[170]: # Check if they replaced
       df_tw.expanded_urls
```

```
[170]: 0       https://twitter.com/dog_rates/status/
       1       https://twitter.com/dog_rates/status/
       2       https://twitter.com/dog_rates/status/
       3       https://twitter.com/dog_rates/status/
       4       https://twitter.com/dog_rates/status/
                                ...
       2351    https://twitter.com/dog_rates/status/
       2352    https://twitter.com/dog_rates/status/
       2353    https://twitter.com/dog_rates/status/
       2354    https://twitter.com/dog_rates/status/
       2355    https://twitter.com/dog_rates/status/
       Name: expanded_urls, Length: 2074, dtype: object
```

[171]:

```
# Combine the new expanded_urls with thier ids and "/photo/1" to fix the urls
df_tw["expanded_urls"] = df_tw["expanded_urls"] + df_tw["tweet_id"]+ "/photo/1"
```

Test

[172]:

```
# Check if urls changed
df_tw.expanded_urls
```

[172]:

```
0       https://twitter.com/dog_rates/status/892420643…
1       https://twitter.com/dog_rates/status/892177421…
2       https://twitter.com/dog_rates/status/891815181…
3       https://twitter.com/dog_rates/status/891689557…
4       https://twitter.com/dog_rates/status/891327558…
                              …
2351    https://twitter.com/dog_rates/status/666049248…
2352    https://twitter.com/dog_rates/status/666044226…
2353    https://twitter.com/dog_rates/status/666033412…
2354    https://twitter.com/dog_rates/status/666029285…
2355    https://twitter.com/dog_rates/status/666020888…
Name: expanded_urls, Length: 2074, dtype: object
```

Back to Quality list

### Issue #8: > Rating dinominator must be equal to 10 there are other values: (0, 15, 70, 7, 11, 150, 170, 20, 50, 90, 80, 40, 130, 110, 16, 120, 2) {programmatic assessment}

Define:

Replace rating_denominator values by 10 and Fix rating_numerator

Code

[173]:

```
df_tw.query('rating_denominator != 10')
```

[173]:

```
                tweet_id        tweet_source  \
433   820690176645140481  Twitter for iPhone
516   810984652412424192  Twitter for iPhone
902   758467244762497024  Twitter for iPhone
1068  740373189193256964  Twitter for iPhone
1120  731156023742988288  Twitter for iPhone
1165  722974582966214656  Twitter for iPhone
1202  716439118184652801  Twitter for iPhone
1228  713900603437621249  Twitter for iPhone
1254  710658690886586372  Twitter for iPhone
1274  709198395643068416  Twitter for iPhone
1351  704054845121142784  Twitter for iPhone
1433  697463031882764288  Twitter for iPhone
1635  684222868335505415  Twitter for iPhone
1662  682962037429899265  Twitter for iPhone
1779  677716515794329600  Twitter for iPhone
```

```
1843  675853064436391936  Twitter for iPhone
2335  666287406224695296  Twitter for iPhone

                                              tweet_text  \
433   The floofs have been released I repeat the flo…
516   Meet Sam. She smiles 24/7 &amp; secretly aspir…
902   Why does this never happen at my front door…
1068  After so many requests, this is Bretagne. She …
1120  Say hello to this unbelievably well behaved sq…
1165  Happy 4/20 from the squad! 13/10 for all https…
1202  This is Bluebert. He just saw that both #Final…
1228  Happy Saturday here's 9 puppers on a bench. 99…
1254  Here's a brigade of puppers. All look very pre…
1274  From left to right:\nCletus, Jerome, Alejandro…
1351  Here is a whole flock of puppers.  60/50 I'll …
1433  Happy Wednesday here's a bucket of pups. 44/40…
1635  Someone help the girl is being mugged. Several…
1662  This is Darrel. He just robbed a 7/11 and is i…
1779  IT'S PUPPERGEDDON. Total of 144/120 …I think…
1843  Here we have an entire platoon of puppers. Tot…
2335  This is an Albanian 3 1/2 legged  Episcopalian…

                                           expanded_urls  rating_numerator  \
433   https://twitter.com/dog_rates/status/820690176…                84
516   https://twitter.com/dog_rates/status/810984652…                24
902   https://twitter.com/dog_rates/status/758467244…               165
1068  https://twitter.com/dog_rates/status/740373189…                 9
1120  https://twitter.com/dog_rates/status/731156023…               204
1165  https://twitter.com/dog_rates/status/722974582…                 4
1202  https://twitter.com/dog_rates/status/716439118…                50
1228  https://twitter.com/dog_rates/status/713900603…                99
1254  https://twitter.com/dog_rates/status/710658690…                80
1274  https://twitter.com/dog_rates/status/709198395…                45
1351  https://twitter.com/dog_rates/status/704054845…                60
1433  https://twitter.com/dog_rates/status/697463031…                44
1635  https://twitter.com/dog_rates/status/684222868…               121
1662  https://twitter.com/dog_rates/status/682962037…                 7
1779  https://twitter.com/dog_rates/status/677716515…               144
1843  https://twitter.com/dog_rates/status/675853064…                88
2335  https://twitter.com/dog_rates/status/666287406…                 1

      rating_denominator  dog_name doggo floofer pupper puppo tweet_date  \
433                   70       NaN   NaN     NaN    NaN   NaN 2017-01-15
516                    7       Sam   NaN     NaN    NaN   NaN 2016-12-19
902                  150       NaN   NaN     NaN    NaN   NaN 2016-07-28
1068                  11       NaN   NaN     NaN    NaN   NaN 2016-06-08
1120                 170       NaN   NaN     NaN    NaN   NaN 2016-05-13
```

```
1165                      20       NaN   NaN     NaN     NaN    NaN 2016-04-21
1202                      50  Bluebert   NaN     NaN     NaN    NaN 2016-04-03
1228                      90       NaN   NaN     NaN     NaN    NaN 2016-03-27
1254                      80       NaN   NaN     NaN     NaN    NaN 2016-03-18
1274                      50       NaN   NaN     NaN     NaN    NaN 2016-03-14
1351                      50       NaN   NaN     NaN     NaN    NaN 2016-02-28
1433                      40       NaN   NaN     NaN     NaN    NaN 2016-02-10
1635                     110       NaN   NaN     NaN     NaN    NaN 2016-01-05
1662                      11    Darrel   NaN     NaN     NaN    NaN 2016-01-01
1779                     120       NaN   NaN     NaN     NaN    NaN 2015-12-18
1843                      80       NaN   NaN     NaN     NaN    NaN 2015-12-13
2335                       2       NaN   NaN     NaN     NaN    NaN 2015-11-16

      tweet_time
433           17
516           23
902            1
1068           2
1120          16
1165           2
1202           1
1228           1
1254           2
1274           2
1351          21
1433          16
1635           4
1662          16
1779           5
1843           1
2335          16
```

[174]:
```
df_tw.loc[(df_tw.tweet_id == '835246439529840640'), 'rating_numerator'] = 13
```

[175]:
```
#replacing all values in rating_denominator column by 10
df_tw['rating_denominator'] = 10
```

[176]:
```
#832088576586297345 not a rating dog
df_tw.drop(df_tw.query('tweet_id == "832088576586297345"').index, inplace=True)
```

[177]:
```
df_tw.loc[(df_tw.tweet_id == '820690176645140481'), 'rating_numerator'] = 12
```

[178]:
```
# '810984652412424192' not a rating dog
df_tw.drop(df_tw.query('tweet_id == "810984652412424192"').index, inplace=True)
```

[179]:
```
#'775096608509886464'page not existing
df_tw.drop(df_tw.query('tweet_id == "775096608509886464"').index, inplace=True)
```

```
[180]: df_tw.loc[(df_tw.tweet_id == '758467244762497024'), 'rating_numerator'] = 11
```

```
[181]: df_tw.loc[(df_tw.tweet_id == '740373189193256964'), 'rating_numerator'] = 14
```

```
[182]: df_tw.loc[(df_tw.tweet_id == '731156023742988288'), 'rating_numerator'] = 12
```

```
[183]: df_tw.loc[(df_tw.tweet_id == '722974582966214656'), 'rating_numerator'] = 13
```

```
[184]: df_tw.loc[(df_tw.tweet_id == '716439118184652801'), 'rating_numerator'] = 11
```

```
[185]: df_tw.loc[(df_tw.tweet_id == '713900603437621249'), 'rating_numerator'] = 11
```

```
[186]: df_tw.loc[(df_tw.tweet_id == '710658690886586372'), 'rating_numerator'] = 10
```

```
[187]: df_tw.loc[(df_tw.tweet_id == '709198395643068416'), 'rating_numerator'] = 9
```

```
[188]: df_tw.loc[(df_tw.tweet_id == '704054845121142784'), 'rating_numerator'] = 12
```

```
[189]: df_tw.loc[(df_tw.tweet_id == '697463031882764288'), 'rating_numerator'] = 11
```

```
[190]: #'686035780142297088'  not a rating dog
       df_tw.drop(df_tw.query('tweet_id == "686035780142297088"').index, inplace=True)
```

```
[191]: df_tw.loc[(df_tw.tweet_id == '684225744407494656'), 'rating_numerator'] = 11
```

```
[192]: df_tw.loc[(df_tw.tweet_id == '684222868335505415'), 'rating_numerator'] = 11
```

```
[193]: df_tw.loc[(df_tw.tweet_id == '682962037429899265'), 'rating_numerator'] = 10
```

```
[194]: #'682808988178739200' not a rating
       df_tw.drop(df_tw.query('tweet_id == "682808988178739200"').index, inplace=True)
```

```
[195]: df_tw.loc[(df_tw.tweet_id == '677716515794329600'), 'rating_numerator'] = 12
```

```
[196]: df_tw.loc[(df_tw.tweet_id == '675853064436391936'), 'rating_numerator'] = 10
```

```
[197]: df_tw.loc[(df_tw.tweet_id == '666287406224695296'), 'rating_numerator'] = 9
```

```
[198]: df_tw.rating_numerator.unique()
```

```
[198]: array([  13,   12,   14,    5,   11,    6,   10,    0,   75,   27,    3,
                 7,    8,    9,    4, 1776,   26,    2,    1,  420])
```

```
[199]: df_tw.rating_numerator.value_counts()
```

```
[199]: 12      481
       10      436
       11      415
```

```
13      285
9       152
8        98
7        51
14       38
5        33
6        32
3        18
4        15
2         9
1         4
27        1
75        1
0         1
1776      1
26        1
420       1
Name: rating_numerator, dtype: int64
```

Check small value_counts in rating_numerator

```
[200]: df_tw.query('rating_numerator == 0')
```

```
[200]:                tweet_id        tweet_source  \
315  835152434251116546  Twitter for iPhone

                                            tweet_text  \
315  When you're so blinded by your systematic plag…

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

     rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
315                  10      NaN   NaN     NaN    NaN   NaN 2017-02-24

     tweet_time
315          15
```

```
[201]: # Not a dog rating
       df_tw.drop(df_tw.query('rating_numerator == 0').index, inplace=True)
```

```
[202]: df_tw.query('rating_numerator == 1')
```

```
[202]:                 tweet_id        tweet_source  \
1869  675153376133427200  Twitter for iPhone
2091  670783437142401025  Twitter for iPhone
2261  667549055577362432  Twitter Web Client
2338  666104133288665088  Twitter for iPhone
```

```
                                             tweet_text  \
1869  What kind of person sends in a picture without…
2091  Flamboyant pup here. Probably poisonous. Won't…
2261  Never seen dog like this. Breathes heavy. Tilt…
2338  Not familiar with this breed. No tail (weird)…

                                          expanded_urls  rating_numerator  \
1869  https://twitter.com/dog_rates/status/675153376…                 1
2091  https://twitter.com/dog_rates/status/670783437…                 1
2261  https://twitter.com/dog_rates/status/667549055…                 1
2338  https://twitter.com/dog_rates/status/666104133…                 1

      rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
1869                  10      NaN   NaN     NaN    NaN   NaN 2015-12-11
2091                  10      NaN   NaN     NaN    NaN   NaN 2015-11-29
2261                  10      NaN   NaN     NaN    NaN   NaN 2015-11-20
2338                  10      NaN   NaN     NaN    NaN   NaN 2015-11-16

      tweet_time
1869           3
2091           1
2261           3
2338           4
```

[203]:

```
#Not a dog rating
df_tw.drop(df_tw.query('rating_numerator == 1').index, inplace=True)
```

[204]:

```
df_tw.query('rating_numerator == 2')
```

[204]:

```
                tweet_id        tweet_source  \
1761  678675843183484930  Twitter for iPhone
1764  678424312106393600  Twitter for iPhone
1920  674265582246694913  Twitter for iPhone
2079  670826280409919488  Twitter for iPhone
2237  668142349051129856  Twitter for iPhone
2246  667878741721415682  Twitter for iPhone
2310  666786068205871104  Twitter for iPhone
2326  666411507551481857  Twitter for iPhone
2349  666051853826850816  Twitter for iPhone

                                             tweet_text  \
1761  Exotic pup here. Tail long af. Throat looks sw…
1764  This is Crystal. She's a shitty fireman. No se…
1920  This is Henry. He's a shit dog. Short pointy e…
2079  Scary dog here. Too many legs. Extra tail. Not…
2237  This lil pup is Oliver. Hops around. Has wings…
```

```
2246  This is Tedrick. He lives on the edge. Needs s…
2310  Unfamiliar with this breed. Ears pointy af. Wo…
2326  This is quite the dog. Gets really excited whe…
2349  This is an odd dog. Hard on the outside but lo…

                                         expanded_urls  rating_numerator  \
1761  https://twitter.com/dog_rates/status/678675843…                 2
1764  https://twitter.com/dog_rates/status/678424312…                 2
1920  https://twitter.com/dog_rates/status/674265582…                 2
2079  https://twitter.com/dog_rates/status/670826280…                 2
2237  https://twitter.com/dog_rates/status/668142349…                 2
2246  https://twitter.com/dog_rates/status/667878741…                 2
2310  https://twitter.com/dog_rates/status/666786068…                 2
2326  https://twitter.com/dog_rates/status/666411507…                 2
2349  https://twitter.com/dog_rates/status/666051853…                 2

      rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
1761                  10      NaN   NaN     NaN    NaN   NaN 2015-12-20
1764                  10  Crystal   NaN     NaN    NaN   NaN 2015-12-20
1920                  10    Henry   NaN     NaN    NaN   NaN 2015-12-08
2079                  10      NaN   NaN     NaN    NaN   NaN 2015-11-29
2237                  10      NaN   NaN     NaN    NaN   NaN 2015-11-21
2246                  10  Tedrick   NaN     NaN    NaN   NaN 2015-11-21
2310                  10      NaN   NaN     NaN    NaN   NaN 2015-11-18
2326                  10      NaN   NaN     NaN    NaN   NaN 2015-11-17
2349                  10      NaN   NaN     NaN    NaN   NaN 2015-11-16

      tweet_time
1761          20
1764           3
1920          16
2079           4
2237          19
2246           1
2310           1
2326           0
2349           0
```

[205]:

```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "678675843183484930"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674265582246694913"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670826280409919488"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668142349051129856"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666786068205871104"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666411507551481857"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666051853826850816"').index, inplace=True)
```

```
[206]: df_tw.query('rating_numerator == 3')
```

```
[206]:                 tweet_id        tweet_source  \
       765   777885040357281792  Twitter for iPhone
       1189  718246886998687744  Twitter for iPhone
       1249  711306686208872448  Twitter for iPhone
       1314  706901761596989440  Twitter for iPhone
       1406  699088579889332224  Twitter for iPhone
       1478  693590843962331137  Twitter for iPhone
       1601  686003207160610816  Twitter for iPhone
       1692  681302363064414209  Twitter for iPhone
       1836  676098748976615425  Twitter for iPhone
       1898  674690135443775488  Twitter for iPhone
       1928  674045139690631169  Twitter for iPhone
       1947  673700254269775872  Twitter for iPhone
       2136  670055038660800512  Twitter for iPhone
       2183  668989615043424256  Twitter for iPhone
       2202  668643542311546881  Twitter for iPhone
       2239  667937095915278337  Twitter for iPhone
       2305  666837028449972224  Twitter for iPhone
       2334  666293911632134144  Twitter for iPhone

                                                    tweet_text  \
       765   This is Wesley. He's clearly trespassing. Seem…
       1189  This is Alexanderson. He's got a weird ass bir…
       1249  What hooligan sent in pictures w/out a dog in …
       1314  This is Elliot. He's blocking the roadway. Dow…
       1406  This is Charl. He's a bully. Chucks that dumbb…
       1478  Meet Phil. He's big af. Currently destroying t…
       1601  This is Hammond. He's a peculiar pup. Loves lo…
       1692  This is Chuck. He's a neat dog. Very flexible…
       1836  Extremely rare pup here. Very religious. Alway…
       1898  Meet Patrick. He's an exotic pup. Jumps great …
       1928  Herd of wild dogs here. Not sure what they're …
       1947  Large blue dog here. Cool shades. Flipping us …
       2136  This is Tommy. He's a cool dog. Hard not to st…
       2183  This is Bernie. He's taking his Halloween cost…
       2202  Fascinating dog here. Loves beach. Oddly long …
       2239  This dog resembles a baked potato. Bed looks u…
       2305  My goodness. Very rare dog here. Large. Tail d…
       2334  This is a funny dog. Weird toes. Won't come do…

                                                 expanded_urls  rating_numerator  \
       765   https://twitter.com/dog_rates/status/777885040…                 3
       1189  https://twitter.com/dog_rates/status/718246886…                 3
       1249  https://twitter.com/dog_rates/status/711306686…                 3
       1314  https://twitter.com/dog_rates/status/706901761…                 3
```

```
1406  https://twitter.com/dog_rates/status/699088579…                 3
1478  https://twitter.com/dog_rates/status/693590843…                 3
1601  https://twitter.com/dog_rates/status/686003207…                 3
1692  https://twitter.com/dog_rates/status/681302363…                 3
1836  https://twitter.com/dog_rates/status/676098748…                 3
1898  https://twitter.com/dog_rates/status/674690135…                 3
1928  https://twitter.com/dog_rates/status/674045139…                 3
1947  https://twitter.com/dog_rates/status/673700254…                 3
2136  https://twitter.com/dog_rates/status/670055038…                 3
2183  https://twitter.com/dog_rates/status/668989615…                 3
2202  https://twitter.com/dog_rates/status/668643542…                 3
2239  https://twitter.com/dog_rates/status/667937095…                 3
2305  https://twitter.com/dog_rates/status/666837028…                 3
2334  https://twitter.com/dog_rates/status/666293911…                 3

      rating_denominator      dog_name doggo floofer  pupper puppo tweet_date  \
765                   10        Wesley   NaN     NaN     NaN   NaN 2016-09-19
1189                  10  Alexanderson   NaN     NaN     NaN   NaN 2016-04-08
1249                  10           NaN   NaN     NaN     NaN   NaN 2016-03-19
1314                  10        Elliot   NaN     NaN     NaN   NaN 2016-03-07
1406                  10         Charl   NaN     NaN     NaN   NaN 2016-02-15
1478                  10          Phil   NaN     NaN  pupper   NaN 2016-01-31
1601                  10       Hammond   NaN     NaN     NaN   NaN 2016-01-10
1692                  10         Chuck   NaN     NaN     NaN   NaN 2015-12-28
1836                  10           NaN   NaN     NaN     NaN   NaN 2015-12-13
1898                  10       Patrick   NaN     NaN     NaN   NaN 2015-12-09
1928                  10           NaN   NaN     NaN     NaN   NaN 2015-12-08
1947                  10           NaN   NaN     NaN     NaN   NaN 2015-12-07
2136                  10         Tommy   NaN     NaN     NaN   NaN 2015-11-27
2183                  10        Bernie   NaN     NaN     NaN   NaN 2015-11-24
2202                  10           NaN   NaN     NaN     NaN   NaN 2015-11-23
2239                  10           NaN   NaN     NaN     NaN   NaN 2015-11-21
2305                  10           NaN   NaN     NaN     NaN   NaN 2015-11-18
2334                  10           NaN   NaN     NaN     NaN   NaN 2015-11-16

      tweet_time
765           15
1189           1
1249          21
1314          17
1406           4
1478           0
1601           1
1692           2
1836          17
1898          20
1928           1
```

```
1947          3
2136          1
2183          3
2202          4
2239          5
2305          4
2334         16
```

[207]:

```
#Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "777885040357281792"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "718246886998687744"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "706901761596989440"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "699088579889332224"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "693590843962331137"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "686003207160610816"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "681302363064414209"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "676098748976615425"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674690135443775488"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674045139690631169"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673700254269775872"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670055038660800512"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668643542311546881"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667937095915278337"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666837028449972224"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666293911632134144"').index, inplace=True)
```

[208]:

```
df_tw.query('rating_numerator == 4')
```

[208]:

```
                tweet_id        tweet_source  \
883   760252756032651264  Twitter for iPhone
912   757596066325864448  Twitter for iPhone
1004  747816857231626240  Twitter for iPhone
1219  714631576617938945  Twitter for iPhone
1303  707420581654872064  Twitter for iPhone
1459  695064344191721472  Twitter for iPhone
1629  684567543613382656  Twitter for iPhone
1701  680940246314430465  Twitter for iPhone
1941  673715861853720576  Twitter for iPhone
2070  671122204919246848  Twitter for iPhone
2076  670838202509447168  Twitter for iPhone
2186  668981893510119424  Twitter for iPhone
2222  668466899341221888  Twitter for iPhone
2288  667176164155375616  Twitter for iPhone
2316  666649482315059201  Twitter for iPhone

                                             tweet_text  \
883   This is Fido. He can tell the weather. Not goo…
```

```
912   Here's another picture without a dog in it. Id…
1004  Viewer discretion is advised. This is a terrib…
1219  This is Benedict. He's a feisty pup. Needs a b…
1303  This is Keurig. He's a rare dog. Laughs like a…
1459  This may be the greatest video I've ever been …
1629  This is Bobby. He doesn't give a damn about pe…
1701  This is Alice. She's an idiot. 4/10 https://t…
1941  This is a heavily opinionated dog. Loves walls…
2070  Two miniature golden retrievers here. Webbed p…
2076  Pink dogs here. Unreasonably long necks. Left …
2186  Unique dog here. Oddly shaped tail. Long pink …
2222  Here is a mother dog caring for her pups. Snaz…
2288  These are strange dogs. All have toupees. Long…
2316  Cool dog. Enjoys couch. Low monotone bark. Ver…

                                          expanded_urls  rating_numerator  \
883   https://twitter.com/dog_rates/status/760252756…                 4
912   https://twitter.com/dog_rates/status/757596066…                 4
1004  https://twitter.com/dog_rates/status/747816857…                 4
1219  https://twitter.com/dog_rates/status/714631576…                 4
1303  https://twitter.com/dog_rates/status/707420581…                 4
1459  https://twitter.com/dog_rates/status/695064344…                 4
1629  https://twitter.com/dog_rates/status/684567543…                 4
1701  https://twitter.com/dog_rates/status/680940246…                 4
1941  https://twitter.com/dog_rates/status/673715861…                 4
2070  https://twitter.com/dog_rates/status/671122204…                 4
2076  https://twitter.com/dog_rates/status/670838202…                 4
2186  https://twitter.com/dog_rates/status/668981893…                 4
2222  https://twitter.com/dog_rates/status/668466899…                 4
2288  https://twitter.com/dog_rates/status/667176164…                 4
2316  https://twitter.com/dog_rates/status/666649482…                 4

      rating_denominator  dog_name doggo floofer pupper puppo tweet_date  \
883                   10      Fido   NaN     NaN    NaN   NaN 2016-08-01
912                   10       NaN   NaN     NaN    NaN   NaN 2016-07-25
1004                  10       NaN   NaN     NaN    NaN   NaN 2016-06-28
1219                  10  Benedict   NaN     NaN    NaN   NaN 2016-03-29
1303                  10    Keurig   NaN     NaN    NaN   NaN 2016-03-09
1459                  10       NaN   NaN     NaN    NaN   NaN 2016-02-04
1629                  10     Bobby   NaN     NaN    NaN   NaN 2016-01-06
1701                  10     Alice   NaN     NaN    NaN   NaN 2015-12-27
1941                  10       NaN   NaN     NaN    NaN   NaN 2015-12-07
2070                  10       NaN   NaN     NaN    NaN   NaN 2015-11-30
2076                  10       NaN   NaN     NaN    NaN   NaN 2015-11-29
2186                  10       NaN   NaN     NaN    NaN   NaN 2015-11-24
2222                  10       NaN   NaN     NaN    NaN   NaN 2015-11-22
2288                  10       NaN   NaN     NaN    NaN   NaN 2015-11-19
```

```
2316                  10       NaN   NaN     NaN    NaN   NaN 2015-11-17

      tweet_time
883           23
912           15
1004          15
1219           1
1303           4
1459           2
1629           2
1701           2
1941           4
2070           0
2076           5
2186           2
2222          16
2288           3
2316          16
```

[209]:
```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "760252756032651264"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "757596066325864448"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "714631576617938945"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "707420581654872064"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "695064344191721472"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "684567543613382656"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "680940246314430465"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673715861853720576"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671122204919246848"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670838202509447168"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668981893510119424"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667176164155375616"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666649482315059201"').index, inplace=True)
```

[210]:
```
df_tw.query('rating_numerator == 5')
```

[210]:
```
                tweet_id         tweet_source  \
45    883482846933004288   Twitter for iPhone
956   751583847268179968   Twitter for iPhone
1399  699691744225525762   Twitter for iPhone
1461  694925794720792577  Vine - Make a Scene
1508  691483041324204033   Twitter for iPhone
1583  687102708889812993   Twitter for iPhone
1619  684959798585110529   Twitter for iPhone
1624  684880619965411328   Twitter for iPhone
1645  683849932751646720   Twitter for iPhone
1680  682003177596559360   Twitter for iPhone
```

```
1727  679877062409191424   Twitter for iPhone
1796  677301033169788928   Twitter for iPhone
1808  676897532954456065   Twitter for iPhone
1820  676588346097852417   Twitter for iPhone
1861  675483430902214656   Twitter for iPhone
1874  675135153782571009   Twitter for iPhone
1901  674646392044941312   Twitter for iPhone
1904  674632714662858753   Twitter for iPhone
1925  674063288070742018   Twitter for iPhone
1979  672980819271634944   Twitter for iPhone
2013  672231046314901505   Twitter for iPhone
2026  671879137494245376   Twitter for iPhone
2063  671159727754231808   Twitter for iPhone
2092  670782429121134593   Twitter for iPhone
2109  670449342516494336   Twitter for iPhone
2134  670069087419133954   Twitter for iPhone
2139  670037189829525505   Twitter for iPhone
2153  669661792646373376   Twitter for iPhone
2181  668994913074286592   Twitter for iPhone
2206  668631377374486528   Twitter for iPhone
2242  667911425562669056   Twitter for iPhone
2312  666776908487630848   Twitter for iPhone
2351  666049248165822465   Twitter for iPhone

                                              tweet_text  \
45    This is Bella. She hopes her smile made you sm…
956   Please stop sending it pictures that don't eve…
1399  This is Dave. He's a tropical pup. Short lil l…
1461  Please only send in dogs. This t-rex is very s…
1508  When bae says they can't go out but you see th…
1583  Army of water dogs here. None of them know whe…
1619  This is Jerry. He's a neat dog. No legs (tragi…
1624  Here we have a basking dino pupper. Looks powe…
1645  This is Jiminy. He's not the brightest dog. Ne…
1680  Unique dog here. Wrinkly as hell. Weird segmen…
1727  Meet Penelope. She's a bacon frise. Total babe…
1796  This is Juckson. He's totally on his way to a …
1808  Exotic handheld dog here. Appears unathletic. …
1820  This is Bubbles. He kinda resembles a fish. Al…
1861  Rare shielded battle dog here. Very happy abou…
1874  This is Steven. He got locked outside. Damn it…
1901  Two gorgeous dogs here. Little waddling dog is…
1904  Rare submerged pup here. Holds breath for a lo…
1925  This is Earl. Earl is lost. Someone help Earl…
1979  Extraordinary dog here. Looks large. Just a he…
2013  Exotic underwater dog here. Very shy. Wont ret…
2026  This is Brad. He's a chubby lil pup. Doesn't r…
```

```
2063  This is Anthony. He just finished up his maste…
2092  This dude slaps your girl's ass what do you do…
2109  Vibrant dog here. Fabulous tail. Only 2 legs t…
2134  This is Randall. He's from Chernobyl. Built pl…
2139  Awesome dog here. Not sure where it is tho. Sp…
2153  This is a brave dog. Excellent free climber. T…
2181  Two gorgeous pups here. Both have cute fake ho…
2206  Meet Zeek. He is a grey Cumulonimbus. Zeek is …
2242  Wow. Armored dog here. Ready for battle. Face …
2312  This is Josep. He is a Rye Manganese mix. Can …
2351  Here we have a 1949 1st generation vulpix. Enj…

                                          expanded_urls  rating_numerator  \
45    https://twitter.com/dog_rates/status/883482846…                 5
956   https://twitter.com/dog_rates/status/751583847…                 5
1399  https://twitter.com/dog_rates/status/699691744…                 5
1461  https://twitter.com/dog_rates/status/694925794…                 5
1508  https://twitter.com/dog_rates/status/691483041…                 5
1583  https://twitter.com/dog_rates/status/687102708…                 5
1619  https://twitter.com/dog_rates/status/684959798…                 5
1624  https://twitter.com/dog_rates/status/684880619…                 5
1645  https://twitter.com/dog_rates/status/683849932…                 5
1680  https://twitter.com/dog_rates/status/682003177…                 5
1727  https://twitter.com/dog_rates/status/679877062…                 5
1796  https://twitter.com/dog_rates/status/677301033…                 5
1808  https://twitter.com/dog_rates/status/676897532…                 5
1820  https://twitter.com/dog_rates/status/676588346…                 5
1861  https://twitter.com/dog_rates/status/675483430…                 5
1874  https://twitter.com/dog_rates/status/675135153…                 5
1901  https://twitter.com/dog_rates/status/674646392…                 5
1904  https://twitter.com/dog_rates/status/674632714…                 5
1925  https://twitter.com/dog_rates/status/674063288…                 5
1979  https://twitter.com/dog_rates/status/672980819…                 5
2013  https://twitter.com/dog_rates/status/672231046…                 5
2026  https://twitter.com/dog_rates/status/671879137…                 5
2063  https://twitter.com/dog_rates/status/671159727…                 5
2092  https://twitter.com/dog_rates/status/670782429…                 5
2109  https://twitter.com/dog_rates/status/670449342…                 5
2134  https://twitter.com/dog_rates/status/670069087…                 5
2139  https://twitter.com/dog_rates/status/670037189…                 5
2153  https://twitter.com/dog_rates/status/669661792…                 5
2181  https://twitter.com/dog_rates/status/668994913…                 5
2206  https://twitter.com/dog_rates/status/668631377…                 5
2242  https://twitter.com/dog_rates/status/667911425…                 5
2312  https://twitter.com/dog_rates/status/666776908…                 5
2351  https://twitter.com/dog_rates/status/666049248…                 5
```

```
      rating_denominator  dog_name  doggo floofer  pupper puppo tweet_date  \
45                    10     Bella    NaN     NaN     NaN   NaN 2017-07-08
956                   10       NaN  doggo     NaN  pupper   NaN 2016-07-09
1399                  10      Dave    NaN     NaN     NaN   NaN 2016-02-16
1461                  10       NaN    NaN     NaN     NaN   NaN 2016-02-03
1508                  10       NaN    NaN     NaN     NaN   NaN 2016-01-25
1583                  10       NaN    NaN     NaN     NaN   NaN 2016-01-13
1619                  10     Jerry    NaN     NaN     NaN   NaN 2016-01-07
1624                  10       NaN    NaN     NaN  pupper   NaN 2016-01-06
1645                  10    Jiminy    NaN     NaN     NaN   NaN 2016-01-04
1680                  10       NaN    NaN     NaN     NaN   NaN 2015-12-30
1727                  10  Penelope    NaN     NaN     NaN   NaN 2015-12-24
1796                  10   Juckson    NaN     NaN     NaN   NaN 2015-12-17
1808                  10       NaN    NaN     NaN     NaN   NaN 2015-12-15
1820                  10   Bubbles    NaN     NaN     NaN   NaN 2015-12-15
1861                  10       NaN    NaN     NaN     NaN   NaN 2015-12-12
1874                  10    Steven    NaN     NaN     NaN   NaN 2015-12-11
1901                  10       NaN    NaN     NaN     NaN   NaN 2015-12-09
1904                  10       NaN    NaN     NaN     NaN   NaN 2015-12-09
1925                  10      Earl    NaN     NaN     NaN   NaN 2015-12-08
1979                  10       NaN    NaN     NaN     NaN   NaN 2015-12-05
2013                  10       NaN    NaN     NaN     NaN   NaN 2015-12-03
2026                  10      Brad    NaN     NaN     NaN   NaN 2015-12-02
2063                  10   Anthony    NaN     NaN     NaN   NaN 2015-11-30
2092                  10       NaN    NaN     NaN     NaN   NaN 2015-11-29
2109                  10       NaN    NaN     NaN     NaN   NaN 2015-11-28
2134                  10   Randall    NaN     NaN     NaN   NaN 2015-11-27
2139                  10       NaN    NaN     NaN     NaN   NaN 2015-11-27
2153                  10       NaN    NaN     NaN     NaN   NaN 2015-11-25
2181                  10       NaN    NaN     NaN     NaN   NaN 2015-11-24
2206                  10      Zeek    NaN     NaN     NaN   NaN 2015-11-23
2242                  10       NaN    NaN     NaN     NaN   NaN 2015-11-21
2312                  10     Josep    NaN     NaN     NaN   NaN 2015-11-18
2351                  10       NaN    NaN     NaN     NaN   NaN 2015-11-16

      tweet_time
45             0
956            1
1399          20
1461          16
1508           4
1583           2
1619           4
1624          23
1645           3
1680           1
1727           4
```

```
1796           1
1808          22
1820           2
1861           1
1874           2
1901          17
1904          16
1925           3
1979           3
2013           1
2026           2
2063           2
2092           1
2109           3
2134           2
2139           0
2153          23
2181           3
2206           3
2242           3
2312           0
2351           0
```

[211]:
```
# Not A rating dog
df_tw.drop(df_tw.query('tweet_id == "699691744225525762"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "694925794720792577"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "687102708889812993"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "684959798585110529"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "684880619965411328"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "683849932751646720"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "682003177596559360"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "679877062409191424"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "676897532954456065"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "675483430902214656"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "675135153782571009"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674646392044941312"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674632714662858753"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "674063288070742018"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "672980819271634944"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "672231046314901505"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671879137494245376"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671159727754231808"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670449342516494336"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670069087419133954"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670037189829525505"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "669661792646373376"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668994913074286592"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667911425562669056"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667550882905632768"').index, inplace=True)
```

```
[212]: df_tw.query('rating_numerator == 6')
```

```
[212]:                 tweet_id        tweet_source  \
229   848212111729840128  Twitter for iPhone
993   748575535303884801  Twitter for iPhone
1078  739485634323156992  Twitter for iPhone
1125  730196704625098752  Twitter for iPhone
1241  712065007010385924  Twitter for iPhone
1363  702932127499816960  Twitter for iPhone
1388  700462010979500032  Twitter for iPhone
1392  700062718104104960  Twitter for iPhone
1460  695051054296211456  Twitter for iPhone
1636  684200372118904832  Twitter for iPhone
1708  680798457301471234  Twitter for iPhone
1759  678740035362037760  Twitter for iPhone
1787  677565715327688705  Twitter for iPhone
1803  676948236477857792  Twitter for iPhone
1839  675891555769696257  Twitter for iPhone
1972  673240798075449344  Twitter for iPhone
1986  672884426393653248  Twitter for iPhone
1992  672614745925664768  Twitter for iPhone
2033  671744970634719232  Twitter for iPhone
2040  671544874165002241  Twitter for iPhone
2044  671533943490011136  Twitter for iPhone
2071  671115716440031232  Twitter for iPhone
2100  670704688707301377  Twitter for iPhone
2119  670417414769758208  Twitter for iPhone
2157  669583744538451968  Twitter for iPhone
2163  669375718304980992  Twitter for iPhone
2238  668113020489474048  Twitter for iPhone
2276  667443425659232256  Twitter for iPhone
2322  666430724426358785  Twitter for iPhone
2330  666362758909284353  Twitter for iPhone
2342  666082916733198337  Twitter for iPhone
2352  666044226329800704  Twitter for iPhone

                                             tweet_text  \
229   This is Jerry. He's doing a distinguished tong…
993   This is one of the most reckless puppers I've …
1078  This is Kyle. He's a heavy drinker and an avid…
1125  This is Charles. He's camera shy. Tail longer …
1241  This is Chester. He's clearly in charge of the…
1363  This is Chip. He's an Upper West Nile Pantaloo…
1388  This is Murphy. He's a mini golden retriever. …
```

```
1392  This is Eazy-E. He's colorful af. Must be rare…
1460  Meet Brian (pronounced "Kirk"). He's not amuse…
1636  Gang of fearless hoofed puppers here. Straight…
1708  Say hello to Moofasa. He must be a powerful do…
1759  Meet Tango. He's a large dog. Doesn't care muc…
1787  Contortionist pup here. Inside pentagram. Clea…
1803  This is Karl. Karl thinks he's slick. 6/10 sne…
1839  This is Donny. He's summoning the demon monste…
1972  Magical floating dog here. Very calm. Always h…
1986  Marvelous dog here. Rad ears. Not very soft. L…
1992  This is Norman. Doesn't bark much. Very docile…
2033  Very fit horned dog here. Looks powerful. Not …
2040  Interesting dog here. Very large. Purple. Mani…
2044  Super rare dog here. Spiffy mohawk. Sharp mout…
2071  Meet Phred. He isn't steering, looking at the …
2100  Meet Danny. He's too good to look at the road …
2119  Sharp dog here. Introverted. Loves purple. Not…
2157  Special dog here. Pretty big. Neck kinda long …
2163  This is Billl. He's trying to be a ghost but h…
2238  This is Alfie. He's that one hypocritical gym …
2276  Exotic dog here. Long neck. Weird paws. Obsess…
2322  Oh boy what a pup! Sunglasses take this one to…
2330  Unique dog here. Very small. Lives in containe…
2342  Here we have a well-established sunblockerspan…
2352  This is a purebred Piers Morgan. Loves to Netf…

                                          expanded_urls  rating_numerator  \
229   https://twitter.com/dog_rates/status/848212111…                 6
993   https://twitter.com/dog_rates/status/748575535…                 6
1078  https://twitter.com/dog_rates/status/739485634…                 6
1125  https://twitter.com/dog_rates/status/730196704…                 6
1241  https://twitter.com/dog_rates/status/712065007…                 6
1363  https://twitter.com/dog_rates/status/702932127…                 6
1388  https://twitter.com/dog_rates/status/700462010…                 6
1392  https://twitter.com/dog_rates/status/700062718…                 6
1460  https://twitter.com/dog_rates/status/695051054…                 6
1636  https://twitter.com/dog_rates/status/684200372…                 6
1708  https://twitter.com/dog_rates/status/680798457…                 6
1759  https://twitter.com/dog_rates/status/678740035…                 6
1787  https://twitter.com/dog_rates/status/677565715…                 6
1803  https://twitter.com/dog_rates/status/676948236…                 6
1839  https://twitter.com/dog_rates/status/675891555…                 6
1972  https://twitter.com/dog_rates/status/673240798…                 6
1986  https://twitter.com/dog_rates/status/672884426…                 6
1992  https://twitter.com/dog_rates/status/672614745…                 6
2033  https://twitter.com/dog_rates/status/671744970…                 6
2040  https://twitter.com/dog_rates/status/671544874…                 6
```

```
2044  https://twitter.com/dog_rates/status/671533943…                 6
2071  https://twitter.com/dog_rates/status/671115716…                 6
2100  https://twitter.com/dog_rates/status/670704688…                 6
2119  https://twitter.com/dog_rates/status/670417414…                 6
2157  https://twitter.com/dog_rates/status/669583744…                 6
2163  https://twitter.com/dog_rates/status/669375718…                 6
2238  https://twitter.com/dog_rates/status/668113020…                 6
2276  https://twitter.com/dog_rates/status/667443425…                 6
2322  https://twitter.com/dog_rates/status/666430724…                 6
2330  https://twitter.com/dog_rates/status/666362758…                 6
2342  https://twitter.com/dog_rates/status/666082916…                 6
2352  https://twitter.com/dog_rates/status/666044226…                 6

      rating_denominator dog_name doggo floofer  pupper puppo tweet_date  \
229                   10    Jerry   NaN     NaN     NaN   NaN 2017-04-01
993                   10      NaN   NaN     NaN     NaN   NaN 2016-06-30
1078                  10     Kyle   NaN     NaN     NaN   NaN 2016-06-05
1125                  10  Charles   NaN     NaN     NaN   NaN 2016-05-11
1241                  10  Chester   NaN     NaN     NaN   NaN 2016-03-21
1363                  10     Chip   NaN     NaN     NaN   NaN 2016-02-25
1388                  10   Murphy   NaN     NaN     NaN   NaN 2016-02-18
1392                  10     Eazy   NaN     NaN     NaN   NaN 2016-02-17
1460                  10    Brian   NaN     NaN     NaN   NaN 2016-02-04
1636                  10      NaN   NaN     NaN     NaN   NaN 2016-01-05
1708                  10  Moofasa   NaN     NaN     NaN   NaN 2015-12-26
1759                  10    Tango   NaN     NaN     NaN   NaN 2015-12-21
1787                  10      NaN   NaN     NaN     NaN   NaN 2015-12-17
1803                  10     Karl   NaN     NaN     NaN   NaN 2015-12-16
1839                  10    Donny   NaN     NaN     NaN   NaN 2015-12-13
1972                  10      NaN   NaN     NaN     NaN   NaN 2015-12-05
1986                  10      NaN   NaN     NaN     NaN   NaN 2015-12-04
1992                  10   Norman   NaN     NaN  pupper   NaN 2015-12-04
2033                  10      NaN   NaN     NaN     NaN   NaN 2015-12-01
2040                  10      NaN   NaN     NaN     NaN   NaN 2015-12-01
2044                  10      NaN   NaN     NaN     NaN   NaN 2015-12-01
2071                  10    Phred   NaN     NaN     NaN   NaN 2015-11-29
2100                  10    Danny   NaN     NaN     NaN   NaN 2015-11-28
2119                  10      NaN   NaN     NaN     NaN   NaN 2015-11-28
2157                  10      NaN   NaN     NaN     NaN   NaN 2015-11-25
2163                  10    Billl   NaN     NaN     NaN   NaN 2015-11-25
2238                  10    Alfie   NaN     NaN     NaN   NaN 2015-11-21
2276                  10      NaN   NaN     NaN     NaN   NaN 2015-11-19
2322                  10      NaN   NaN     NaN     NaN   NaN 2015-11-17
2330                  10      NaN   NaN     NaN     NaN   NaN 2015-11-16
2342                  10      NaN   NaN     NaN     NaN   NaN 2015-11-16
2352                  10      NaN   NaN     NaN     NaN   NaN 2015-11-16
```

```
      tweet_time
229           16
993           17
1078          15
1125           0
1241          23
1363          19
1388          23
1392          21
1460           1
1636           2
1708          17
1759           0
1787          19
1803           2
1839           4
1972          20
1986          21
1992           3
2033          17
2040           4
2044           3
2071          23
2100          20
2119           1
2157          18
2163           4
2238          17
2276          20
2322           1
2330          21
2342           2
2352           0
```

[213]:

```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "848212111729840128"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "730196704625098752"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "712065007010385924"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "702932127499816960"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "700462010979500032"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "700062718104104960"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "684200372118904832"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "680798457301471234"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "678740035362037760"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "677565715327688705"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "676948236477857792"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673240798075449344"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "672884426393653248"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "672614745925664768"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671744970634719232"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671544874165002241"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671533943490011136"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670417414769758208"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "669583744538451968"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667443425659232256"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666430724426358785"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "666362758909284353"').index, inplace=True)
```

[214]:

```
df_tw.query('rating_numerator == 7')
```

[214]:

```
                tweet_id        tweet_source  \
814   771014301343748096  Twitter for iPhone
896   759099523532779520  Twitter for iPhone
998   748307329658011649  Twitter for iPhone
1029  745712589599014916  Twitter for iPhone
1045  743595368194129920  Twitter for iPhone
1239  712092745624633345  Twitter for iPhone
1300  707693576495472641  Twitter for iPhone
1325  706166467411222528  Twitter for iPhone
1342  704819833553219584  Twitter for iPhone
1384  700796979434098688  Twitter for iPhone
1415  698703483621523456  Twitter for iPhone
1448  696405997980676096  Twitter for iPhone
1473  693647888581312512  Twitter for iPhone
1498  692417313023332352  Twitter for iPhone
1555  688894073864884227  Twitter for iPhone
1573  687494652870668288  Twitter for iPhone
1579  687312378585812992  Twitter for iPhone
1608  685641971164143616  Twitter for iPhone
1670  682406705142087680  Twitter for iPhone
1684  681679526984871937  Twitter for iPhone
1730  679854723806179328  Twitter for iPhone
1737  679530280114372609  Twitter for iPhone
1739  679511351870550016  Twitter for iPhone
1754  678798276842360832  Twitter for iPhone
1872  675146535592706048  Twitter for iPhone
1956  673612854080196609  Twitter for iPhone
1959  673576835670777856  Twitter for iPhone
1965  673345638550134785  Twitter for iPhone
1981  672970152493887488  Twitter for iPhone
1998  672538107540070400  Twitter for iPhone
2012  672239279297454080  Twitter for iPhone
2022  672068090318987265  Twitter for iPhone
2054  671390180817915904  Twitter for iPhone
```

```
2077  670833812859932673  Twitter for iPhone
2088  670792680469889025  Twitter for iPhone
2093  670780561024270336  Twitter for iPhone
2107  670465786746662913  Twitter for iPhone
2159  669571471778410496  Twitter for iPhone
2184  668988183816871936  Twitter for iPhone
2198  668815180734689280  Twitter for iPhone
2211  668614819948453888  Twitter for iPhone
2235  668171859951755264  Twitter for iPhone
2258  667724302356258817  Twitter Web Client
2272  667491009379606528  Twitter Web Client
2274  667455448082227200  Twitter Web Client
2279  667405339315146752  Twitter for iPhone
2296  667090893657276420  Twitter for iPhone
2320  666437273139982337  Twitter for iPhone
2323  666428276349472768  Twitter for iPhone
2327  666407126856765440  Twitter for iPhone
2354  666029285002620928  Twitter for iPhone

                                             tweet_text  \
814   Another pic without a dog in it? What am I sup…
896   Meet Toby. He has a drinking problem. Inflatab…
998   This sherk must've leapt out of the water and …
1029  This is Percy. He fell asleep at the wheel. Ir…
1045  This is Harold. He looks slippery af. Probably…
1239  This is Steven. He's inverted af. Also very he…
1300  This is Jiminus. He's in a tub for some reason…
1325  This is Dexter. He's a shy pup. Doesn't bark m…
1342  This is Chesterson. He's a Bolivian Scoop Dog…
1384  This is Daniel. He's a neat pup. Exotic af. Cu…
1415  This is Rusty. He has no respect for POULTRY p…
1448  This is Berb. He just found out that they have…
1473  What kind of person sends in a pic without a d…
1498  Meet Herschel. He's slightly bigger than ur av…
1555  This is Hamrick. He's covered in corn flakes. …
1573  This is Marq. He stole this car. 7/10 wtf Marq…
1579  "You got any games on your phone" 7/10 for inv…
1608  This is Otis. He just passed a cop while going…
1670  This is Patrick. He's a bigass pupper. 7/10 ht…
1684  Can you spot Toby the guilty pupper? 7/10 woul…
1730  This is Bruce. He's a rare pup. Covered in Fro…
1737  Guys this really needs to stop. We've been ove…
1739  Say hello to William. He makes fun of others b…
1754  This is Linda. She fucking hates trees. 7/10 h…
1872  This is Coops. He's yelling at the carpet. Not…
1956  This is Jeffri. He's a speckled ice pupper. Ve…
1959  Sun burnt dog here. Quite large. Wants to prom…
```

```
1965  This is Gerald. He's a fluffy lil yellow pup. …
1981  This is Chet. He's having a hard time. Really …
1998  Meet Darby. He's a Fiscal Tutankhamen Waxbeard…
2012  This is Ridley. He doesn't know how to couch. …
2022  Say hello to Gizmo. He's upset because he's no…
2054  Striped dog here. Having fun playing on back. …
2077  This is Jett. He is unimpressed by flower. 7/1…
2088  This is Antony. He's a Sheraton Tetrahedron. S…
2093  This is Traviss. He has no ears. Two rare dogs…
2107  Silly dog here. Wearing bunny ears. Nice long …
2159  This is Keith. He's had 13 DUIs. 7/10 that's t…
2184  Honor to rate this dog. Lots of fur on him. Tw…
2198  This is a wild Toblerone from Papua New Guinea…
2211  Here is a horned dog. Much grace. Can jump ove…
2235  This is a Trans Siberian Kellogg named Alfonso…
2258  What a dog to start the day with. Very calm. L…
2272  Two dogs in this one. Both are rare Jujitsu Py…
2274  This is Reese and Twips. Reese protects Twips…
2279  This is Biden. Biden just tripped… 7/10 http…
2296  This is Clybe. He is an Anemone Valdez. One ea…
2320  Here we see a lone northeastern Cumberbatch. H…
2323  Here we have an Austrian Pulitzer. Collectors …
2327  This is a southern Vesuvius bumblegruff. Can d…
2354  This is a western brown Mitsubishi terrier. Up…

                                        expanded_urls  rating_numerator  \
814   https://twitter.com/dog_rates/status/771014301…                 7
896   https://twitter.com/dog_rates/status/759099523…                 7
998   https://twitter.com/dog_rates/status/748307329…                 7
1029  https://twitter.com/dog_rates/status/745712589…                 7
1045  https://twitter.com/dog_rates/status/743595368…                 7
1239  https://twitter.com/dog_rates/status/712092745…                 7
1300  https://twitter.com/dog_rates/status/707693576…                 7
1325  https://twitter.com/dog_rates/status/706166467…                 7
1342  https://twitter.com/dog_rates/status/704819833…                 7
1384  https://twitter.com/dog_rates/status/700796979…                 7
1415  https://twitter.com/dog_rates/status/698703483…                 7
1448  https://twitter.com/dog_rates/status/696405997…                 7
1473  https://twitter.com/dog_rates/status/693647888…                 7
1498  https://twitter.com/dog_rates/status/692417313…                 7
1555  https://twitter.com/dog_rates/status/688894073…                 7
1573  https://twitter.com/dog_rates/status/687494652…                 7
1579  https://twitter.com/dog_rates/status/687312378…                 7
1608  https://twitter.com/dog_rates/status/685641971…                 7
1670  https://twitter.com/dog_rates/status/682406705…                 7
1684  https://twitter.com/dog_rates/status/681679526…                 7
1730  https://twitter.com/dog_rates/status/679854723…                 7
```

```
1737  https://twitter.com/dog_rates/status/679530280…                 7
1739  https://twitter.com/dog_rates/status/679511351…                 7
1754  https://twitter.com/dog_rates/status/678798276…                 7
1872  https://twitter.com/dog_rates/status/675146535…                 7
1956  https://twitter.com/dog_rates/status/673612854…                 7
1959  https://twitter.com/dog_rates/status/673576835…                 7
1965  https://twitter.com/dog_rates/status/673345638…                 7
1981  https://twitter.com/dog_rates/status/672970152…                 7
1998  https://twitter.com/dog_rates/status/672538107…                 7
2012  https://twitter.com/dog_rates/status/672239279…                 7
2022  https://twitter.com/dog_rates/status/672068090…                 7
2054  https://twitter.com/dog_rates/status/671390180…                 7
2077  https://twitter.com/dog_rates/status/670833812…                 7
2088  https://twitter.com/dog_rates/status/670792680…                 7
2093  https://twitter.com/dog_rates/status/670780561…                 7
2107  https://twitter.com/dog_rates/status/670465786…                 7
2159  https://twitter.com/dog_rates/status/669571471…                 7
2184  https://twitter.com/dog_rates/status/668988183…                 7
2198  https://twitter.com/dog_rates/status/668815180…                 7
2211  https://twitter.com/dog_rates/status/668614819…                 7
2235  https://twitter.com/dog_rates/status/668171859…                 7
2258  https://twitter.com/dog_rates/status/667724302…                 7
2272  https://twitter.com/dog_rates/status/667491009…                 7
2274  https://twitter.com/dog_rates/status/667455448…                 7
2279  https://twitter.com/dog_rates/status/667405339…                 7
2296  https://twitter.com/dog_rates/status/667090893…                 7
2320  https://twitter.com/dog_rates/status/666437273…                 7
2323  https://twitter.com/dog_rates/status/666428276…                 7
2327  https://twitter.com/dog_rates/status/666407126…                 7
2354  https://twitter.com/dog_rates/status/666029285…                 7

      rating_denominator    dog_name doggo floofer  pupper puppo tweet_date  \
814                   10         NaN   NaN     NaN     NaN   NaN 2016-08-31
896                   10        Toby   NaN     NaN     NaN   NaN 2016-07-29
998                   10         NaN   NaN     NaN     NaN   NaN 2016-06-30
1029                  10       Percy   NaN     NaN     NaN   NaN 2016-06-22
1045                  10      Harold   NaN     NaN     NaN   NaN 2016-06-17
1239                  10      Steven   NaN     NaN     NaN   NaN 2016-03-22
1300                  10     Jiminus   NaN     NaN     NaN   NaN 2016-03-09
1325                  10      Dexter   NaN     NaN     NaN   NaN 2016-03-05
1342                  10  Chesterson   NaN     NaN     NaN   NaN 2016-03-02
1384                  10      Daniel   NaN     NaN     NaN   NaN 2016-02-19
1415                  10       Rusty   NaN     NaN     NaN   NaN 2016-02-14
1448                  10        Berb   NaN     NaN     NaN   NaN 2016-02-07
1473                  10         NaN   NaN     NaN     NaN   NaN 2016-01-31
1498                  10    Herschel   NaN     NaN  pupper   NaN 2016-01-27
1555                  10     Hamrick   NaN     NaN  pupper   NaN 2016-01-18
```

```
1573                  10        Marq   NaN     NaN     NaN   NaN 2016-01-14
1579                  10         NaN   NaN     NaN  pupper   NaN 2016-01-13
1608                  10        Otis   NaN     NaN  pupper   NaN 2016-01-09
1670                  10     Patrick   NaN     NaN  pupper   NaN 2015-12-31
1684                  10         NaN   NaN     NaN  pupper   NaN 2015-12-29
1730                  10       Bruce   NaN     NaN     NaN   NaN 2015-12-24
1737                  10         NaN   NaN     NaN     NaN   NaN 2015-12-23
1739                  10     William   NaN     NaN     NaN   NaN 2015-12-23
1754                  10       Linda   NaN     NaN     NaN   NaN 2015-12-21
1872                  10       Coops   NaN     NaN     NaN   NaN 2015-12-11
1956                  10      Jeffri   NaN     NaN  pupper   NaN 2015-12-06
1959                  10         NaN   NaN     NaN     NaN   NaN 2015-12-06
1965                  10      Gerald   NaN     NaN     NaN   NaN 2015-12-06
1981                  10        Chet   NaN     NaN  pupper   NaN 2015-12-05
1998                  10       Darby   NaN     NaN     NaN   NaN 2015-12-03
2012                  10      Ridley   NaN     NaN     NaN   NaN 2015-12-03
2022                  10       Gizmo   NaN     NaN     NaN   NaN 2015-12-02
2054                  10         NaN   NaN     NaN     NaN   NaN 2015-11-30
2077                  10        Jett   NaN     NaN     NaN   NaN 2015-11-29
2088                  10      Antony   NaN     NaN     NaN   NaN 2015-11-29
2093                  10     Traviss   NaN     NaN     NaN   NaN 2015-11-29
2107                  10         NaN   NaN     NaN     NaN   NaN 2015-11-28
2159                  10       Keith   NaN     NaN     NaN   NaN 2015-11-25
2184                  10         NaN   NaN     NaN     NaN   NaN 2015-11-24
2198                  10         NaN   NaN     NaN     NaN   NaN 2015-11-23
2211                  10         NaN   NaN     NaN     NaN   NaN 2015-11-23
2235                  10         NaN   NaN     NaN     NaN   NaN 2015-11-21
2258                  10         NaN   NaN     NaN     NaN   NaN 2015-11-20
2272                  10         NaN   NaN     NaN     NaN   NaN 2015-11-19
2274                  10       Reese   NaN     NaN     NaN   NaN 2015-11-19
2279                  10       Biden   NaN     NaN     NaN   NaN 2015-11-19
2296                  10       Clybe   NaN     NaN     NaN   NaN 2015-11-18
2320                  10         NaN   NaN     NaN     NaN   NaN 2015-11-17
2323                  10         NaN   NaN     NaN     NaN   NaN 2015-11-17
2327                  10         NaN   NaN     NaN     NaN   NaN 2015-11-17
2354                  10         NaN   NaN     NaN     NaN   NaN 2015-11-15

      tweet_time
814           15
896           18
998            0
1029          20
1045           0
1239           1
1300          22
1325          17
1342           0
```

```
1384     21
1415      3
1448     18
1473      4
1498     18
1555      1
1573      4
1579     16
1608      1
1670      3
1684      3
1730      2
1737      5
1739      3
1754      4
1872      2
1956     21
1959     18
1965      3
1981      2
1998     22
2012      2
2022     15
2054     18
2077      5
2088      2
2093      1
2107      4
2159     17
2184      3
2198     15
2211      2
2235     20
2258     15
2272     23
2274     21
2279     18
2296     21
2320      2
2323      1
2327      0
2354     23
```

[215]:

```
# Not A dog rating
df_tw.drop(df_tw.query('tweet_id == "817502432452313088"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "771014301343748096"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "743595368194129920"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "712092745624633345"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "706166467411222528"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "704819833553219584"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "700796979434098688"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "693647888581312512"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "692417313023332352"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "688894073864884227"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "687312378585812992"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "682406705142087680"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "679854723806179328"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "679530280114372609"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "679511351870550016"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673612854080196609"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673576835670777856"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "673345638550134785"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "672970152493887488"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "671390180817915904"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670792680469889025"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "670465786746662913"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668988183816871936"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "668614819948453888"').index, inplace=True)
df_tw.drop(df_tw.query('tweet_id == "667724302356258817"').index, inplace=True)
```

[216]:

```
df_tw.query('rating_numerator == 14')
```

[216]:

```
                tweet_id         tweet_source  \
9     890240255349198849   Twitter for iPhone
22    887517139158093824   Twitter for iPhone
35    885518971528720385   Twitter for iPhone
41    884441805382717440   Twitter for iPhone
56    881536004380872706   Twitter for iPhone
76    878057613040115712   Twitter for iPhone
83    876537666061221889   Twitter for iPhone
110   871102520638267392   Twitter for iPhone
117   870063196459192321   Twitter for iPhone
123   868880397819494401   Twitter for iPhone
164   860184849394610176   Twitter for iPhone
187   856282028240666624   Twitter for iPhone
199   854120357044912130   Twitter for iPhone
209   852226086759018497   Twitter Web Client
214   851464819735769094   Twitter for iPhone
268   841439858740625411   Twitter for iPhone
339   832273440279240704   Twitter for iPhone
347   831911600680497154   Twitter for iPhone
369   828650029636317184   Twitter for iPhone
372   828381636999917570   Twitter for iPhone
395   825535076884762624   Twitter for iPhone
```

```
404   824025158776213504   Twitter for iPhone
418   822462944365645825   Twitter for iPhone
426   821407182352777218   Twitter for iPhone
436   820314633777061888   Twitter for iPhone
448   819006400881917954   Twitter for iPhone
449   819004803107983360   Twitter for iPhone
490   813812741911748608   Twitter for iPhone
533   807621403335917568   Twitter for iPhone
631   794205286408003585   Twitter for iPhone
657   791774931465953280  Vine - Make a Scene
758   778408200802557953   Twitter for iPhone
788   774314403806253056   Twitter for iPhone
866   762035686371364864   Twitter for iPhone
924   755206590534418437   Twitter for iPhone
1053  742465774154047488   Twitter for iPhone
1068  740373189193256964   Twitter for iPhone
1609  685547936038666240   Twitter for iPhone

                                             tweet_text  \
9     This is Cassie. She is a college pup. Studying…
22    I've yet to rate a Venezuelan Hover Wiener. Th…
35    I have a new hero and his name is Howard. 14/1…
41    I present to you, Pup in Hat. Pup in Hat is gr…
56    Here is a pupper approaching maximum borkdrive…
76    This is Emmy. She was adopted today. Massive r…
83    I can say with the pupmost confidence that the…
110   Never doubt a doggo 14/10 https://t.co/AbBLh2FZCH
117   Meet Clifford. He's quite large. Also red. Goo…
123   This is Walter. He won't start hydrotherapy wi…
164   Here we have perhaps the wisest dog of all. Ab…
187   This is Cermet, Paesh, and Morple. They are ab…
199   Sometimes you guys remind me just how impactfu…
209   Meet General. He wasn't content with the quali…
214   This is Iggy. He was a rescue dog killed in th…
268   Here we have some incredible doggos for #K9Vet…
339   Say hello to Smiley. He's a blind therapy dogg…
347   Meet Kuyu. He was trapped in a well for 10 day…
369   Occasionally, we're sent fantastic stories. Th…
372   Meet Doobert. He's a deaf doggo. Didn't stop h…
395   Here's a very loving and accepting puppo. Appe…
404   "I wish we were dogs" 14/10 for @BadlandsNPS h…
418   This is Gabe. He was the unequivocal embodimen…
426   This is Sundance. He's a doggo drummer. Even s…
436   We are proud to support @LoveYourMelon on thei…
448   This is Sunny. She was also a very good First …
449   This is Bo. He was a very good First Doggo. 14…
490   Meet Gary, Carrie Fisher's dog. Idk what I can…
```

```
533   This is Ollie Vue. He was a 3 legged pupper on…
631   This is Laika. She was a space pupper. The fir…
657   Vine will be deeply missed. This was by far my…
758   RIP Loki. Thank you for the good times. You wi…
788   I WAS SENT THE ACTUAL DOG IN THE PROFILE PIC B…
866   This left me speechless. 14/10 heckin heroic a…
924   This is one of the most inspirational stories …
1053  Was just informed about this hero pupper and o…
1068  After so many requests, this is Bretagne. She …
1609  Everybody needs to read this. Jack is our firs…

                                        expanded_urls  rating_numerator  \
9     https://twitter.com/dog_rates/status/890240255…                 14
22    https://twitter.com/dog_rates/status/887517139…                 14
35    https://twitter.com/dog_rates/status/885518971…                 14
41    https://twitter.com/dog_rates/status/884441805…                 14
56    https://twitter.com/dog_rates/status/881536004…                 14
76    https://twitter.com/dog_rates/status/878057613…                 14
83    https://twitter.com/dog_rates/status/876537666…                 14
110   https://twitter.com/dog_rates/status/871102520…                 14
117   https://twitter.com/dog_rates/status/870063196…                 14
123   https://twitter.com/dog_rates/status/868880397…                 14
164   https://twitter.com/dog_rates/status/860184849…                 14
187   https://twitter.com/dog_rates/status/856282028…                 14
199   https://twitter.com/dog_rates/status/854120357…                 14
209   https://twitter.com/dog_rates/status/852226086…                 14
214   https://twitter.com/dog_rates/status/851464819…                 14
268   https://twitter.com/dog_rates/status/841439858…                 14
339   https://twitter.com/dog_rates/status/832273440…                 14
347   https://twitter.com/dog_rates/status/831911600…                 14
369   https://twitter.com/dog_rates/status/828650029…                 14
372   https://twitter.com/dog_rates/status/828381636…                 14
395   https://twitter.com/dog_rates/status/825535076…                 14
404   https://twitter.com/dog_rates/status/824025158…                 14
418   https://twitter.com/dog_rates/status/822462944…                 14
426   https://twitter.com/dog_rates/status/821407182…                 14
436   https://twitter.com/dog_rates/status/820314633…                 14
448   https://twitter.com/dog_rates/status/819006400…                 14
449   https://twitter.com/dog_rates/status/819004803…                 14
490   https://twitter.com/dog_rates/status/813812741…                 14
533   https://twitter.com/dog_rates/status/807621403…                 14
631   https://twitter.com/dog_rates/status/794205286…                 14
657   https://twitter.com/dog_rates/status/791774931…                 14
758   https://twitter.com/dog_rates/status/778408200…                 14
788   https://twitter.com/dog_rates/status/774314403…                 14
866   https://twitter.com/dog_rates/status/762035686…                 14
924   https://twitter.com/dog_rates/status/755206590…                 14
```

```
1053  https://twitter.com/dog_rates/status/742465774…                14
1068  https://twitter.com/dog_rates/status/740373189…                14
1609  https://twitter.com/dog_rates/status/685547936…                14

      rating_denominator  dog_name  doggo floofer  pupper  puppo tweet_date  \
9                     10    Cassie  doggo     NaN     NaN    NaN 2017-07-26
22                    10       NaN    NaN     NaN     NaN    NaN 2017-07-19
35                    10       NaN    NaN     NaN     NaN    NaN 2017-07-13
41                    10       NaN    NaN     NaN     NaN    NaN 2017-07-10
56                    10       NaN    NaN     NaN  pupper    NaN 2017-07-02
76                    10      Emmy    NaN     NaN     NaN    NaN 2017-06-23
83                    10       NaN    NaN     NaN     NaN    NaN 2017-06-18
110                   10       NaN  doggo     NaN     NaN    NaN 2017-06-03
117                   10  Clifford    NaN     NaN     NaN    NaN 2017-05-31
123                   10    Walter    NaN     NaN     NaN    NaN 2017-05-28
164                   10       NaN    NaN     NaN     NaN    NaN 2017-05-04
187                   10    Cermet    NaN     NaN     NaN    NaN 2017-04-23
199                   10       NaN    NaN     NaN  pupper    NaN 2017-04-17
209                   10   General    NaN     NaN     NaN    NaN 2017-04-12
214                   10      Iggy    NaN     NaN     NaN    NaN 2017-04-10
268                   10       NaN    NaN     NaN     NaN    NaN 2017-03-14
339                   10    Smiley  doggo     NaN     NaN    NaN 2017-02-16
347                   10      Kuyu    NaN     NaN     NaN    NaN 2017-02-15
369                   10       NaN    NaN     NaN     NaN    NaN 2017-02-06
372                   10   Doobert  doggo     NaN     NaN    NaN 2017-02-05
395                   10       NaN    NaN     NaN     NaN  puppo 2017-01-29
404                   10       NaN    NaN     NaN     NaN    NaN 2017-01-24
418                   10      Gabe    NaN     NaN  pupper    NaN 2017-01-20
426                   10  Sundance  doggo     NaN     NaN    NaN 2017-01-17
436                   10       NaN    NaN     NaN     NaN    NaN 2017-01-14
448                   10     Sunny  doggo     NaN     NaN    NaN 2017-01-11
449                   10        Bo  doggo     NaN     NaN    NaN 2017-01-11
490                   10      Gary    NaN     NaN     NaN    NaN 2016-12-27
533                   10     Ollie    NaN     NaN  pupper    NaN 2016-12-10
631                   10     Laika    NaN     NaN  pupper    NaN 2016-11-03
657                   10       NaN    NaN     NaN     NaN    NaN 2016-10-27
758                   10       NaN    NaN     NaN     NaN    NaN 2016-09-21
788                   10       NaN    NaN     NaN     NaN    NaN 2016-09-09
866                   10       NaN    NaN     NaN     NaN    NaN 2016-08-06
924                   10       NaN  doggo     NaN     NaN    NaN 2016-07-19
1053                  10       NaN    NaN     NaN  pupper    NaN 2016-06-13
1068                  10       NaN    NaN     NaN     NaN    NaN 2016-06-08
1609                  10       NaN    NaN     NaN  pupper    NaN 2016-01-08

      tweet_time
9             15
22             3
```

```
35      15
41      15
56      15
76       1
83      20
110     20
117     23
123     17
164     17
187     23
199     23
209     18
214     16
268      0
339     17
347     17
369     17
372     23
395      2
404     22
418     15
426     17
436     17
448      2
449      2
490     18
533     16
631     15
657     22
758      1
788     18
866     21
924      1
1053    21
1068     2
1609    19
```

[217]:
```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "885518971528720385"').index, inplace=True)
```

[218]:
```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "879674319642796034"').index, inplace=True)
```

[219]:
```
# Not a dog rating
df_tw.drop(df_tw.query('tweet_id == "879130579576475649"').index, inplace=True)
```

[220]:
```
df_tw.query('rating_numerator == 26')
```

[220]:

```
                tweet_id        tweet_source  \
1712  680494726643068929  Twitter for iPhone

                                             tweet_text  \
1712  Here we have uncovered an entire battalion of …

                                          expanded_urls  rating_numerator  \
1712  https://twitter.com/dog_rates/status/680494726…                26

      rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
1712                  10      NaN   NaN     NaN    NaN   NaN 2015-12-25

      tweet_time
1712          21
```

[221]:

```
# Fix rating_nominator
df_tw.loc[(df_tw.tweet_id == '680494726643068929'), 'rating_numerator'] = 11.26
```

[222]:

```
df_tw.query('rating_numerator == 27')
```

[222]:

```
               tweet_id        tweet_source  \
763  778027034220126208  Twitter for iPhone

                                            tweet_text  \
763  This is Sophie. She's a Jubilant Bush Pupper. …

                                         expanded_urls  rating_numerator  \
763  https://twitter.com/dog_rates/status/778027034…              27.0

     rating_denominator dog_name doggo floofer  pupper puppo tweet_date  \
763                  10   Sophie   NaN     NaN  pupper   NaN 2016-09-20

     tweet_time
763           0
```

[223]:

```
# Fix rating_numerator
df_tw.loc[(df_tw.tweet_id == '778027034220126208'), 'rating_numerator'] = 11.27
```

[224]:

```
df_tw.query('rating_numerator == 75')
```

[224]:

```
               tweet_id        tweet_source  \
695  786709082849828864  Twitter for iPhone

                                            tweet_text  \
695  This is Logan, the Chow who lived. He solemnly…

                                         expanded_urls  rating_numerator  \
695  https://twitter.com/dog_rates/status/786709082…             75.0

     rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
695                  10    Logan   NaN     NaN    NaN   NaN 2016-10-13

     tweet_time
695          23
```

```
[225]: # Retweet id
       df_tw.drop(df_tw.query('tweet_id == "832215909146226688"').index, inplace=True)
```

```
[226]: df_tw.query('rating_numerator == 420')
```

```
[226]:                 tweet_id        tweet_source  \
       2074  670842764863651840  Twitter for iPhone

                                                     tweet_text  \
       2074  After so many requests… here you go.\n\nGood…

                                                  expanded_urls  rating_numerator  \
       2074  https://twitter.com/dog_rates/status/670842764…             420.0

             rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
       2074                  10      NaN   NaN     NaN    NaN   NaN 2015-11-29

             tweet_time
       2074           5
```

```
[227]: # Not a dog rating
       df_tw.drop(df_tw.query('rating_numerator == 420').index, inplace=True)
```

```
[228]: df_tw.query('rating_numerator == 420')
```

```
[228]: Empty DataFrame
       Columns: [tweet_id, tweet_source, tweet_text, expanded_urls, rating_numerator,
       rating_denominator, dog_name, doggo, floofer, pupper, puppo, tweet_date,
       tweet_time]
       Index: []
```

```
[229]: df_tw.query('rating_numerator == 1776')
```

```
[229]:                tweet_id tweet_source  \
       979  749981277374128128    TweetDeck

                                                    tweet_text  \
       979  This is Atticus. He's quite simply America af…

                                                 expanded_urls  rating_numerator  \
       979  https://twitter.com/dog_rates/status/749981277…           1776.0

            rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
       979                  10  Atticus   NaN     NaN    NaN   NaN 2016-07-04

            tweet_time
       979          15
```

This is original right tweet

Test

```
[230]: # check if not rating ids dropped
       df_tw.query('tweet_id == "682808988178739200"')
```

```
[230]: Empty DataFrame
       Columns: [tweet_id, tweet_source, tweet_text, expanded_urls, rating_numerator,
       rating_denominator, dog_name, doggo, floofer, pupper, puppo, tweet_date,
       tweet_time]
       Index: []
```

```
[231]: #check if rating_numerator replaced
       df_tw.query('tweet_id == "666287406224695296"')
```

```
[231]:                 tweet_id        tweet_source  \
       2335  666287406224695296  Twitter for iPhone

                                                  tweet_text  \
       2335  This is an Albanian 3 1/2 legged  Episcopalian…

                                                expanded_urls  rating_numerator  \
       2335  https://twitter.com/dog_rates/status/666287406…               9.0

             rating_denominator dog_name doggo floofer pupper puppo tweet_date  \
       2335                  10      NaN   NaN     NaN    NaN   NaN 2015-11-16

             tweet_time
       2335          16
```

```
[232]: # Check if ids droped
       df_tw.rating_numerator.value_counts()
```

```
[232]: 12.00      481
       10.00      436
       11.00      415
       13.00      285
       9.00       152
       8.00        98
       14.00       37
```

```
7.00        27
6.00        10
5.00         9
4.00         2
3.00         2
2.00         2
75.00        1
11.27        1
1776.00      1
11.26        1
Name: rating_numerator, dtype: int64
```

Back to Quality list

### Issue #9: >The predictions (‘P1’, ‘P2’, ‘P3’) columns are not clear and familiar to the reader and have strange predictions (spatula, barrow, minibus, etc) **{programmatic assessment}**

Define

Create 2 Columns that handle the confidence prediction and predection dog

Code

[233]:

```
# create a general function which takes a row as input and work on that row
def best_dog(row):
    if row['p1_dog']: # if first_prediction_dog true
        return row['p1']
    elif row['p2_dog']: # if second_prediction_dog is true
        return row['p2']
    else:
        return row['p3']
```

[234]:

```
def conf_dog(row):
    if row['p1_dog']: # if first_prediction_dog true
        return row['p1_conf']
    elif row['p2_dog']: # if second_prediction_dog is true
        return row['p2_conf']
    else:
        return row['p3_conf']
```

[235]:

```
# Create new col prediction_confidence
df_mg['prediction_dog'] = df_mg.apply(best_dog, axis=1)
```

[236]:

```
# Create new col prediction_confidence
df_mg['prediction_confidence'] = df_mg.apply(conf_dog, axis=1)
```

[237]:

```
# Drop columns
dcol = ['p1',
    'p1_conf',
    'p1_dog',
    'p2',
    'p2_conf',
    'p2_dog',
    'p3',
    'p3_conf',
    'p3_dog']
df_mg.drop(columns=dcol, inplace=True)
```

Test

[238]:

```
df_mg.head(0)
```

[238]:

```
Empty DataFrame
Columns: [tweet_id, jpg_url, img_num, prediction_dog, prediction_confidence]
Index: []
```

Back to Quality list

### Issue #10: >Some "tweet_ids" have same "img_url", after checking using the urls: > - (https://twitter.com/dog_rates/status/803692223237865472) > - (https://twitter.com/dog_rates/status/691416866452082688) > - ids img does not exist "Hmm...this page doesn't exist. Try searching for something else": **{visual assessment}** - 759566828574212096 - 802247111496568832 - 851953902622658560 - 842892208864923648 - 861769973181624320 - 873697596434513921 - 888202515573088257

Define

drop duplicates since the second are "WeRateDogs® Retweeted" from the origin img

Code

[239]:

```
df_mg.jpg_url.duplicated().sum()
```

[239]:

```
66
```

[240]:

```
df_mg.query('jpg_url == "https://pbs.twimg.com/media/CWza7kpWcAAdYLc.jpg"')
```

[240]:

```
                tweet_id                                          jpg_url  \
591   679158373988876288  https://pbs.twimg.com/media/CWza7kpWcAAdYLc.jpg
1315  754874841593970688  https://pbs.twimg.com/media/CWza7kpWcAAdYLc.jpg

      img_num prediction_dog  prediction_confidence
591         1            pug               0.272205
1315        1            pug               0.272205
```

[241]:

```
df_mg.query('jpg_url == "https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg"')
```

[241]:

```
                tweet_id                                          jpg_url  \
1785  829374341691346946  https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg
1903  851953902622658560  https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg
```

```
      img_num             prediction_dog  prediction_confidence
1785        1  Staffordshire_bullterrier               0.757547
1903        1  Staffordshire_bullterrier               0.757547
```

[242]:
```
df_mg.drop_duplicates(subset="jpg_url", inplace=True)
```

[243]:
```
# drop page not found
df_mg.drop(df_mg.query('tweet_id == "759566828574212096"').index, inplace=True)
```

[244]:
```
df_mg.drop(df_mg.query('tweet_id == "802247111496568832"').index, inplace=True)
```

[245]:
```
df_mg.drop(df_mg.query('tweet_id == "851953902622658560"').index, inplace=True)
```

[246]:
```
df_mg.drop(df_mg.query('tweet_id == "842892208864923648"').index, inplace=True)
```

[247]:
```
df_mg.drop(df_mg.query('tweet_id == "861769973181624320"').index, inplace=True)
```

[248]:
```
df_mg.drop(df_mg.query('tweet_id == "873697596434513921"').index, inplace=True)
```

[249]:
```
df_mg.drop(df_mg.query('tweet_id == "888202515573088257"').index, inplace=True)
```

Test

[250]:
```
df_mg.jpg_url.duplicated().sum()
```

[250]:
```
0
```

[251]:
```
df_mg.shape
```

[251]:
```
(2009, 5)
```

[252]:
```
df_mg.query('jpg_url == "https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg"')
```

[252]:
```
                tweet_id                                          jpg_url  \
1785  829374341691346946  https://pbs.twimg.com/media/C4KHj-nWQAA3poV.jpg

      img_num             prediction_dog  prediction_confidence
1785        1  Staffordshire_bullterrier               0.757547
```

[253]:
```
df_mg.query('tweet_id == "873697596434513921"')
```

[253]:
```
Empty DataFrame
Columns: [tweet_id, jpg_url, img_num, prediction_dog, prediction_confidence]
Index: []
```

Back to Quality list

### Issue #11: - Ivalid urls: (https://…)(https:/…) (https:/t.c…) - 175 duplicated url

Define

- Drop invalid urls
- Drop duplicated urls

Code

[254]:

```
# Collect lists of urls to drop and check
list_drop_all = [
    "…",
    "s",
    "0",
    "5",
    "h",
    "n",
    "e",
    "g",
    "y",
    "t",
    "https://…",
    "https://t.c…",
    "https://t.c…",
    "https://…",
    "https://t.co/…",
]
```

[255]:

```
# Drop not valid urls
for i in list_drop_all:
    df_js = df_js[df_js.url != i]
```

[256]:

```
# Drop duplicates and keep last one
df_js.drop_duplicates(subset=['url'], keep='last', inplace=True)
```

Test

[257]:

```
# check if invalid urls was droped
df_js[df_js.url == "https://…"]
```

[257]:

```
Empty DataFrame
Columns: [tweet_id, favorite_count, retweet_count, retweet_status, url]
Index: []
```

[258]:

```
# Check if droped duplicated urls keep last
df_js[df_js.url == "https://t.co/f1h2Fsalv9"]
```

[258]:

```
               tweet_id  favorite_count  retweet_count  retweet_status  \
536  806629075125202948           75639          37911  Original tweet

                         url
536  https://t.co/f1h2Fsalv9
```

```
[259]: # Check if there is any duplicated url
       df_js.url.duplicated().sum()
```

```
[259]: 0
```

Back to Quality list

### Issue #12: > retweet_status has one value 'Original tweet', no need it

Define

Drop retweet_status column

Code

```
[260]: # drop retweet_status column
       df_js.drop('retweet_status', inplace=True, axis=1)
```

Test

```
[261]: # Check if retweet_status exist
       df_js.head()
```

```
[261]:              tweet_id  favorite_count  retweet_count                      url
       0  892420643555336193           39467           8853  https://t.co/MgUWQ76dJU
       1  892177421306343426           33819           6514  https://t.co/0Xxu71qeIV
       2  891815181378084864           25461           4328  https://t.co/wUnZnhtVJB
       3  891689557279858688           42908           8964  https://t.co/tD36da7qLQ
       4  891327558926688256           41048           9774  https://t.co/AtUZn91f7f
```

Back to Quality list

### Issue #13: > Tweets missing retweet count and favorite count

Define :

check and drop

Code

```
[262]: # Check (retweet_count) and (favorite_count) columns == 0
       df_js.query('retweet_count == 0 or favorite_count == 0')
```

```
[262]:                tweet_id  favorite_count  retweet_count  \
       31   886054160059072513               0            108
       73   878316110768087041               0           6965
       77   877611172832227328               0             83
       108  871166179821445120               0           5991
       129  867072653475098625               0            135
       131  866816280283807744               0          32883
       136  866094527597207552               0           8952
       159  860924035999428608               0            882
       211  851861385021730816               0             23
```

```
230   847971574464610304               0            482
280   839290600511926273               0            158
284   838916489579200512               0             38
288   838201503651401729               0          11525
297   837012587749474308               0             88
331   832769181346996225               0             43
339   832215909146226688               0           7069
365   828801551087042563               0           3901
419   822163064745328640               0           5571
424   821421320206483457               0           6470
430   820837357901512704               0           7733
454   818307523543449600               0           9366
557   803321560782307329               0           6221
576   800855607700029440               0           1889
588   799308762079035393               0           6478
597   798686750113755136               0           2702
614   796563435802726400               0           8536
668   790227638568808452               0           4190
693   786729988674449408               0           2827
741   780476555013349377               0            136
816   770743923962707968               0          52360
883   760153949710192640               0             38
909   757597904299253760               0            336
1021  746521445350707200               0           1110
1041  743835915802583040               0           2387
1240  711998809858043904               0            138
2257  667550904950915073               0             37
2258  667550882905632768               0             34

                                                  url
31                            https://t.co/WxwJmvjfxo
73                                    https://t.co/i…
77                            https://t.co/dJx4Gzc50G
108                                   https://t.co/X…
129                           https://t.co/nRCdq4g9gG
131                                https://t.co/wk7T…
136                                https://t.co/ORiK…
159                           https://t.co/6n8hzNihJ9
211                           https://t.co/bO0rThDlXI
230                           https://t.co/0bYtoL7Wwt
280                           https://t.co/act3duiqbL
284                           https://t.co/6UCGFczlOI
288                                  https://t.co/EM…
297                           https://t.co/rHLoU5gS30
331                           https://t.co/zFEfMTaHqU
339                              https://t.co/yBO5wu…
365                                            https…
```

```
419                                                   https:…
424                                             https://t.co…
430                                       https://t.co/ii5eNX5…
454                                       https://t.co/LIhmM7i…
557                                            https://t.co/Nb…
576                                     https://t.co/pnUF69K4xk
588                                       https://t.co/7oQ1wpf…
597                                             https://t.co/Y…
614                                            https://t.co/5d…
668                                         https://t.co/21Bg3…
693                                                 https://t.…
741   https://t.co/rnKvzt6RJs https://t.co/v4e2ywe8iO
816                                                  https://t…
883                                     https://t.co/AG5UnRrmzJ
909                                     https://t.co/7boff8zojZ
1021                                                  https:/…
1041                                          https://t.co/0Dv…
1240                                    https://t.co/j6FQGhxYuN
2257                                    https://t.co/uarnTjBeVA
2258                                    https://t.co/r7XzeQZWzb
```

After checking tweets missing retweet count and favorite count some are retweets and others are not dog ratings so we need to drop them

[263]:
```
# drop rows
df_js.drop(df_js.query('retweet_count == 0 or favorite_count == 0').index,␣
 ↪inplace=True)
```

Test

[264]:
```
df_js.query('retweet_count == 0 or favorite_count == 0')
```

[264]:
```
Empty DataFrame
Columns: [tweet_id, favorite_count, retweet_count, url]
Index: []
```

Back to Quality list

# Tidiness Solution

### Issue #1: >doggo, floofer, pupper, puppo these 4 variables shoule be combined into one categorical variable 'dogtionary'. {programmatic assessment}

Define :

Frirst create dogtionary column to handle extract stages from 'tweet_text' column Then drop ('doggo','floofer','pupper','puppo') cols

Code https://knowledge.udacity.com/questions/30062

[265]:

```
#handle none
df_tw.doggo.replace(np.nan, '', inplace=True)
df_tw.floofer.replace(np.nan, '', inplace=True)
df_tw.pupper.replace(np.nan, '', inplace=True)
df_tw.puppo.replace(np.nan, '', inplace=True)

# Merge into a column
df_tw['dogtionary'] = df_tw.doggo + df_tw.floofer + df_tw.pupper + df_tw.puppo

# handle multiple stages
df_tw.loc[df_tw.dogtionary == 'doggopupper', 'dogtionary'] = 'doggo, pupper'
df_tw.loc[df_tw.dogtionary == 'doggopuppo', 'dogtionary'] = 'doggo, puppo'
df_tw.loc[df_tw.dogtionary == 'doggofloofer', 'dogtionary'] = 'doggo, floofer'
# Handle missing values
df_tw.loc[df_tw.dogtionary == '', 'dogtionary'] = np.nan
```

[266]:

```
# Get sample to check dog_stages column
df_tw[['doggo','floofer','pupper','puppo','dogtionary']].sample(15)
```

[266]:

```
      doggo floofer  pupper puppo     dogtionary
1057                 pupper               pupper
1540                                         NaN
54                                           NaN
956   doggo          pupper        doggo, pupper
1643                                         NaN
1796                                         NaN
1000                 pupper               pupper
1573                                         NaN
708                                          NaN
388                                          NaN
1174                                         NaN
360                                          NaN
1261                                         NaN
1439                                         NaN
56                   pupper               pupper
```

[267]:

```
# Check values in dog_stages column
df_tw.dogtionary.value_counts()
```

[267]:

```
pupper            211
doggo              72
puppo              23
floofer             9
doggo, pupper       9
doggo, puppo        1
doggo, floofer      1
Name: dogtionary, dtype: int64
```

Fix the double dogtionary

```
[268]: df_tw.query('dogtionary == "doggo, pupper"')
```

```
[268]:                 tweet_id         tweet_source  \
      460   817777686764523521   Twitter for iPhone
      531   808106460588765185   Twitter for iPhone
      575   801115127852503040   Twitter for iPhone
      705   785639753186217984   Twitter for iPhone
      733   781308096455073793  Vine - Make a Scene
      889   759793422261743616   Twitter for iPhone
      956   751583847268179968   Twitter for iPhone
      1063  741067306818797568   Twitter for iPhone
      1113  733109485275860992   Twitter for iPhone

                                                   tweet_text  \
      460   This is Dido. She's playing the lead role in "…
      531   Here we have Burke (pupper) and Dexter (doggo)…
      575   This is Bones. He's being haunted by another d…
      705   This is Pinot. He's a sophisticated doggo. You…
      733   Pupper butt 1, Doggo 0. Both 12/10 https://t.c…
      889   Meet Maggie &amp; Lila. Maggie is the doggo, L…
      956   Please stop sending it pictures that don't eve…
      1063  This is just downright precious af. 12/10 for …
      1113  Like father (doggo), like son (pupper). Both 1…

                                                expanded_urls  rating_numerator  \
      460   https://twitter.com/dog_rates/status/817777686…              13.0
      531   https://twitter.com/dog_rates/status/808106460…              12.0
      575   https://twitter.com/dog_rates/status/801115127…              12.0
      705   https://twitter.com/dog_rates/status/785639753…              10.0
      733   https://twitter.com/dog_rates/status/781308096…              12.0
      889   https://twitter.com/dog_rates/status/759793422…              12.0
      956   https://twitter.com/dog_rates/status/751583847…               5.0
      1063  https://twitter.com/dog_rates/status/741067306…              12.0
      1113  https://twitter.com/dog_rates/status/733109485…              12.0

            rating_denominator dog_name  doggo floofer  pupper puppo tweet_date  \
      460                   10     Dido  doggo          pupper       2017-01-07
      531                   10      NaN  doggo          pupper       2016-12-12
      575                   10    Bones  doggo          pupper       2016-11-22
      705                   10    Pinot  doggo          pupper       2016-10-11
      733                   10      NaN  doggo          pupper       2016-09-29
      889                   10   Maggie  doggo          pupper       2016-07-31
      956                   10      NaN  doggo          pupper       2016-07-09
      1063                  10      NaN  doggo          pupper       2016-06-10
      1113                  10      NaN  doggo          pupper       2016-05-19
            tweet_time     dogtionary
      460           16  doggo, pupper
      531            0  doggo, pupper
      575           17  doggo, pupper
      705            0  doggo, pupper
      733            1  doggo, pupper
      889           16  doggo, pupper
      956            1  doggo, pupper
      1063           0  doggo, pupper
      1113           1  doggo, pupper
```

[269]:
```
df_tw.loc[(df_tw.tweet_id == '817777686764523521'), 'dogtionary'] = 'pupper'
```

[270]:
```
df_tw.loc[(df_tw.tweet_id == '808106460588765185'), 'dogtionary'] = 'pupper'
```

[271]:
```
df_tw.loc[(df_tw.tweet_id == '801115127852503040'), 'dogtionary'] = 'pupper'
```

[272]:
```
# Not a dog
df_tw.drop(df_tw.query('tweet_id == "785639753186217984"').index, inplace=True)
```

[273]:
```
# Not a dog rating 781308096455073793
df_tw.drop(df_tw.query('tweet_id == "781308096455073793"').index, inplace=True)
```

[274]:
```
df_tw.loc[(df_tw.tweet_id == '759793422261743616'), 'dogtionary'] = 'pupper'
```

[275]:
```
# Not a dog rating 751583847268179968
df_tw.drop(df_tw.query('tweet_id == "751583847268179968"').index, inplace=True)
```

[276]:
```
df_tw.loc[(df_tw.tweet_id == '733109485275860992'), 'dogtionary'] = 'pupper'
```

[277]:
```
df_tw.loc[(df_tw.tweet_id == '741067306818797568'), 'dogtionary'] = 'pupper'
```

[278]:
```
df_tw.query('dogtionary == "doggo, puppo"')
```

[278]:
```
               tweet_id        tweet_source  \
191  855851453814013952  Twitter for iPhone

                                          tweet_text  \
191  Here's a puppo participating in the #ScienceMa…

                                       expanded_urls  rating_numerator  \
191  https://twitter.com/dog_rates/status/855851453…              13.0

     rating_denominator dog_name  doggo floofer pupper  puppo tweet_date  \
191                  10      NaN  doggo                 puppo 2017-04-22

     tweet_time    dogtionary
```

```
191          18  doggo, puppo
```

[279]:

```python
df_tw.loc[(df_tw.tweet_id == '855851453814013952'), 'dogtionary'] = 'puppo'
```

[280]:

```python
df_tw.query('dogtionary == "doggo, floofer"')
```

[280]:

```
               tweet_id        tweet_source  \
200  854010172552949760  Twitter for iPhone

                                            tweet_text  \
200  At first I thought this was a shy doggo, but i…

                                         expanded_urls  rating_numerator  \
200  https://twitter.com/dog_rates/status/854010172…              11.0

     rating_denominator dog_name  doggo  floofer pupper puppo tweet_date  \
200                  10      NaN  doggo  floofer              2017-04-17

     tweet_time      dogtionary
200          16  doggo, floofer
```

[281]:

```python
df_tw.loc[(df_tw.tweet_id == '854010172552949760'), 'dogtionary'] = 'floofer'
```

[282]:

```python
# Drop 'doggo','floofer','pupper','puppo' columns
df_tw.drop(columns=['doggo', 'floofer', 'pupper', 'puppo'], inplace=True)
```

Test

[283]:

```python
#Check column created
df_tw.head()
```

[283]:

```
             tweet_id        tweet_source  \
0  892420643555336193  Twitter for iPhone
1  892177421306343426  Twitter for iPhone
2  891815181378084864  Twitter for iPhone
3  891689557279858688  Twitter for iPhone
4  891327558926688256  Twitter for iPhone

                                          tweet_text  \
0  This is Phineas. He's a mystical boy. Only eve…
1  This is Tilly. She's just checking pup on you…
2  This is Archie. He is a rare Norwegian Pouncin…
3  This is Darla. She commenced a snooze mid meal…
4  This is Franklin. He would like you to stop ca…

                                       expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643…              13.0
1  https://twitter.com/dog_rates/status/892177421…              13.0
```

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

   rating_denominator  dog_name tweet_date  tweet_time dogtionary
0                  10   Phineas 2017-08-01          16        NaN
1                  10     Tilly 2017-08-01           0        NaN
2                  10    Archie 2017-07-31           0        NaN
3                  10     Darla 2017-07-30          15        NaN
4                  10  Franklin 2017-07-29          16        NaN
```

[284]:
```
# Get sample from dogtionary column
df_tw['dogtionary'].value_counts()
```

[284]:
```
pupper     217
doggo       72
puppo       24
floofer     10
Name: dogtionary, dtype: int64
```

[285]:
```
df_tw.query('dogtionary == "doggo, pupper"')
```

[285]:
```
Empty DataFrame
Columns: [tweet_id, tweet_source, tweet_text, expanded_urls, rating_numerator,
rating_denominator, dog_name, tweet_date, tweet_time, dogtionary]
Index: []
```

Back To Tidiness list

### Issue #2: >rating nominator, rating dinominator should be one column since rating dinominator always be 10 {visual assessment}

Define

Drop rating_dinominator column Rename rating_nominator

Code

[286]:
```
#Get columns name
df_tw.columns
```

[286]:
```
Index(['tweet_id', 'tweet_source', 'tweet_text', 'expanded_urls',
       'rating_numerator', 'rating_denominator', 'dog_name', 'tweet_date',
       'tweet_time', 'dogtionary'],
      dtype='object')
```

[287]:
```
# drop rating_denominator
df_tw.drop(['rating_denominator'], axis=1, inplace=True)
```

```
[288]: # rename column rating_numerator
df_tw.rename(
    columns={"rating_numerator":"rating_out_of_ten"}
          ,inplace=True)
```

Test

```
[289]: #Check columns droped and renamed
df_tw.head(2)
```

```
[289]:              tweet_id        tweet_source  \
0  892420643555336193  Twitter for iPhone
1  892177421306343426  Twitter for iPhone

                                          tweet_text  \
0  This is Phineas. He's a mystical boy. Only eve…
1  This is Tilly. She's just checking pup on you…

                                     expanded_urls  rating_out_of_ten  \
0  https://twitter.com/dog_rates/status/892420643…               13.0
1  https://twitter.com/dog_rates/status/892177421…               13.0

  dog_name tweet_date  tweet_time dogtionary
0  Phineas 2017-08-01          16        NaN
1    Tilly 2017-08-01           0        NaN
```

Back To Tidiness list

### Issue #3: > Dataframes: twitter_archive, image_predictions, and tweet_json, Should be one df {visual assessment}

Define

> merge all dataframes in one (twitter_master_df)

Code

```
[290]: # Merge df_tw and df_js to be twitter_master_df
twitter_master_df = df_tw.merge(df_js, how='left')
```

```
[291]: # Merge df_mg and twitter_master_df
twitter_master_df = twitter_master_df.merge(df_mg, how='left')
```

Test issue 4 Tidi

```
[292]: print(twitter_master_df.shape)
twitter_master_df.head(2)
```

```
(1957, 16)
```

[292]:
```
              tweet_id        tweet_source  \
0  892420643555336193  Twitter for iPhone
1  892177421306343426  Twitter for iPhone

                                          tweet_text  \
0  This is Phineas. He's a mystical boy. Only eve…
1  This is Tilly. She's just checking pup on you…

                                       expanded_urls  rating_out_of_ten  \
0  https://twitter.com/dog_rates/status/892420643…               13.0
1  https://twitter.com/dog_rates/status/892177421…               13.0

  dog_name tweet_date  tweet_time dogtionary  favorite_count  retweet_count  \
0  Phineas 2017-08-01          16        NaN         39467.0         8853.0
1    Tilly 2017-08-01           0        NaN         33819.0         6514.0

                       url                                          jpg_url  \
0  https://t.co/MgUWQ76dJU  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg
1  https://t.co/0Xxu71qeIV  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg

   img_num prediction_dog  prediction_confidence
0      1.0         banana               0.076110
1      1.0      Chihuahua               0.323581
```

[293]:
```
twitter_master_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1957 entries, 0 to 1956
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   tweet_id               1957 non-null   object
 1   tweet_source           1957 non-null   object
 2   tweet_text             1957 non-null   object
 3   expanded_urls          1957 non-null   object
 4   rating_out_of_ten      1957 non-null   float64
 5   dog_name               1340 non-null   object
 6   tweet_date             1957 non-null   datetime64[ns]
 7   tweet_time             1957 non-null   int64
 8   dogtionary             323 non-null    object
 9   favorite_count         1956 non-null   float64
 10  retweet_count          1956 non-null   float64
 11  url                    1956 non-null   object
 12  jpg_url                1837 non-null   object
 13  img_num                1837 non-null   float64
 14  prediction_dog         1837 non-null   object
 15  prediction_confidence  1837 non-null   float64
dtypes: datetime64[ns](1), float64(5), int64(1), object(9)
```

```
memory usage: 259.9+ KB
```

Back To Tidiness list

### Issue #4: > **url** column and **expanded_urls** has the same value

Define:

> Since **url** column and **expanded_urls** has the same value and **expanded_urls** has not missing value so we drop **url** col

Code

```
[294]: #Drop column 'url'
       twitter_master_df.drop(columns='url', inplace=True )
```

Test

```
[295]: #Check column droped
       twitter_master_df.head(0)
```

```
[295]: Empty DataFrame
       Columns: [tweet_id, tweet_source, tweet_text, expanded_urls, rating_out_of_ten,
       dog_name, tweet_date, tweet_time, dogtionary, favorite_count, retweet_count,
       jpg_url, img_num, prediction_dog, prediction_confidence]
       Index: []
```

```
[296]: twitter_master_df.isnull().sum()
```

```
[296]: tweet_id                    0
       tweet_source                0
       tweet_text                  0
       expanded_urls               0
       rating_out_of_ten           0
       dog_name                  617
       tweet_date                  0
       tweet_time                  0
       dogtionary               1634
       favorite_count              1
       retweet_count               1
       jpg_url                   120
       img_num                   120
       prediction_dog            120
       prediction_confidence     120
       dtype: int64
```

Back To Tidiness list

## Storing Data Save gathered, assessed, and cleaned master dataset to a CSV file named “twitter_archive_master.csv”.

[297]:

```
# Store the clean DataFrame in a CSV file
twitter_master_df.to_csv('twitter_archive_master.csv', index=False, encoding =␣
 ↪'utf-8')
```

## Analyzing and Visualizing Data In this section, analyze and visualize your wrangled data. You must produce at least **three (3) insights and one (1) visualization.**

[298]:

```
# Read stored cleaned data
df = pd.read_csv('twitter_archive_master.csv')
print(df.shape)
df.head()
```

```
(1957, 15)
```

[298]:

```
             tweet_id        tweet_source  \
0  892420643555336193  Twitter for iPhone
1  892177421306343426  Twitter for iPhone
2  891815181378084864  Twitter for iPhone
3  891689557279858688  Twitter for iPhone
4  891327558926688256  Twitter for iPhone

                                          tweet_text  \
0  This is Phineas. He's a mystical boy. Only eve…
1  This is Tilly. She's just checking pup on you…
2  This is Archie. He is a rare Norwegian Pouncin…
3  This is Darla. She commenced a snooze mid meal…
4  This is Franklin. He would like you to stop ca…

                                     expanded_urls  rating_out_of_ten  \
0  https://twitter.com/dog_rates/status/892420643…               13.0
1  https://twitter.com/dog_rates/status/892177421…               13.0
2  https://twitter.com/dog_rates/status/891815181…               12.0
3  https://twitter.com/dog_rates/status/891689557…               13.0
4  https://twitter.com/dog_rates/status/891327558…               12.0

   dog_name  tweet_date  tweet_time dogtionary  favorite_count  retweet_count  \
0   Phineas  2017-08-01          16        NaN         39467.0         8853.0
1     Tilly  2017-08-01           0        NaN         33819.0         6514.0
2    Archie  2017-07-31           0        NaN         25461.0         4328.0
3     Darla  2017-07-30          15        NaN         42908.0         8964.0
4  Franklin  2017-07-29          16        NaN         41048.0         9774.0

                                           jpg_url  img_num  \
0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg      1.0
1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg      1.0
2  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg      1.0
3  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg      1.0
4  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg      2.0
```

```
      prediction_dog  prediction_confidence
0             banana               0.076110
1          Chihuahua               0.323581
2          Chihuahua               0.716012
3  Labrador_retriever              0.168086
4             basset               0.555712
```

[299]:

```
df.describe()
```

[299]:

```
           tweet_id  rating_out_of_ten  tweet_time  favorite_count  \
count  1.957000e+03        1957.000000 1957.000000     1956.000000
mean   7.384000e+17          11.870480    9.914154     9140.824131
std    6.694702e+16          39.958275    8.645411    12319.956203
min    6.660209e+17           2.000000    0.000000       81.000000
25%    6.777000e+17          10.000000    1.000000     2119.000000
50%    7.139006e+17          11.000000    5.000000     4343.500000
75%    7.902771e+17          12.000000   18.000000    11652.250000
max    8.924206e+17        1776.000000   23.000000   132810.000000

       retweet_count      img_num  prediction_confidence
count    1956.000000  1837.000000           1.837000e+03
mean     2891.254601     1.212303           4.959476e-01
std      4802.323116     0.573185           3.237651e-01
min        16.000000     1.000000           1.740170e-10
25%       647.750000     1.000000           1.966550e-01
50%      1440.000000     1.000000           4.900680e-01
75%      3324.000000     1.000000           7.940870e-01
max     79515.000000     4.000000           9.999560e-01
```

[300]:

```
# Most Popular dog name
df.dog_name.value_counts()
```

[300]:

```
Lucy          11
Charlie       11
Cooper        10
Oliver        10
Penny          9
              ..
Angel          1
Brat           1
Tove           1
Gromit         1
Christoper     1
Name: dog_name, Length: 904, dtype: int64
```

[301]:

```
# Most Popular tweet_time
df.tweet_time.value_counts()
```

[301]:

```
1     270
0     242
2     206
16    186
17    163
3     163
23    109
18    103
15     89
4      87
19     82
20     80
22     72
21     67
5      23
14      9
6       4
13      2
Name: tweet_time, dtype: int64
```

[302]:

```
# Most Popular tweet_date
df.tweet_date.value_counts()
```

[302]:

```
2015-11-28    23
2015-11-16    21
2015-11-19    21
2015-11-29    21
2015-11-22    20
              ..
2016-04-20     1
2016-04-17     1
2017-01-26     1
2016-04-14     1
2016-09-16     1
Name: tweet_date, Length: 593, dtype: int64
```

[303]:

```
# Most Popular dogtionary
df.dogtionary.value_counts()
```

[303]:

```
pupper     217
doggo       72
puppo       24
floofer     10
Name: dogtionary, dtype: int64
```

[304]:

```
# Most Popular tweet_source
df.tweet_source.value_counts()
```

[304]:

```
Twitter for iPhone     1829
Vine - Make a Scene      89
Twitter Web Client       28
TweetDeck                11
Name: tweet_source, dtype: int64
```

[305]:

```
df.rating_out_of_ten.value_counts()
```

[305]:

```
12.00      480
10.00      435
11.00      415
13.00      285
9.00       152
8.00        98
14.00       37
7.00        27
6.00        10
5.00         8
4.00         2
3.00         2
2.00         2
75.00        1
11.27        1
1776.00      1
11.26        1
Name: rating_out_of_ten, dtype: int64
```

[306]:

```
df.prediction_dog.value_counts()
```

[306]:

```
golden_retriever      153
Labrador_retriever    105
Pembroke               92
Chihuahua              87
pug                    62
                     …
snorkel                 1
desktop_computer        1
parachute               1
barrow                  1
coyote                  1
Name: prediction_dog, Length: 281, dtype: int64
```

[307]:

```
df.query('prediction_dog == "golden_retriever"')
```

[307]:

```
                tweet_id        tweet_source  \
14    889531135344209921  Twitter for iPhone
16    888917238123831296  Twitter for iPhone
17    888804989199671297  Twitter for iPhone
39    883482846933004288  Twitter for iPhone
44    882268110199369728  Twitter for iPhone
…                      …                  …
1903  667160273090932737  Twitter for iPhone
1911  667044094246576128  Twitter for iPhone
1936  666345417576210432  Twitter for iPhone
1947  666063827256086533  Twitter for iPhone
1949  666057090499244032  Twitter for iPhone

                                             tweet_text  \
14    This is Stuart. He's sporting his favorite fan…
16    This is Jim. He found a fren. Taught him how t…
17    This is Zeke. He has a new stick. Very proud o…
39    This is Bella. She hopes her smile made you sm…
44    This is Alfy. You're witnessing his first wate…
…                                                  …
1903  This is Bradlay. He is a Ronaldinho Matsuyama …
1911            12/10 gimme now https://t.co/QZAnwgnOMB
1936  Look at this jokester thinking seat belt laws …
1947  This is the happiest dog you will ever see. Ve…
1949  My oh my. This is a rare blond Canadian terrie…

                                          expanded_urls  rating_out_of_ten  \
14    https://twitter.com/dog_rates/status/889531135…               13.0
16    https://twitter.com/dog_rates/status/888917238…               12.0
17    https://twitter.com/dog_rates/status/888804989…               13.0
39    https://twitter.com/dog_rates/status/883482846…                5.0
44    https://twitter.com/dog_rates/status/882268110…               13.0
…                                                  …                 …
1903  https://twitter.com/dog_rates/status/667160273…               11.0
1911  https://twitter.com/dog_rates/status/667044094…               12.0
1936  https://twitter.com/dog_rates/status/666345417…               10.0
1947  https://twitter.com/dog_rates/status/666063827…               10.0
1949  https://twitter.com/dog_rates/status/666057090…                9.0

     dog_name  tweet_date  tweet_time dogtionary  favorite_count  \
14     Stuart  2017-07-24          17      puppo         15359.0
16        Jim  2017-07-23           0        NaN         29611.0
17       Zeke  2017-07-22          16        NaN         26080.0
39      Bella  2017-07-08           0        NaN         46860.0
44       Alfy  2017-07-04          16        NaN         45880.0
…         …         …         …       …          …
1903  Bradlay  2015-11-19           1        NaN           268.0
```

```
1911      NaN  2015-11-18          18         NaN           198.0
1936      NaN  2015-11-16          20         NaN           307.0
1947      NaN  2015-11-16           1         NaN           496.0
1949      NaN  2015-11-16           0         NaN           304.0

      retweet_count                                          jpg_url  img_num  \
14           2321.0  https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg      1.0
16           4709.0  https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg      1.0
17           4559.0  https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg      1.0
39          10407.0  https://pbs.twimg.com/media/DELC9dZXUAADqUk.jpg      1.0
44          12118.0  https://pbs.twimg.com/media/DD5yKdPW0AArzX8.jpg      1.0
…             …                                              …      …
1903           66.0  https://pbs.twimg.com/media/CUI6uuaW4AAvCIs.jpg      1.0
1911           54.0  https://pbs.twimg.com/media/CUHREBXXAAE6A9b.jpg      1.0
1936          146.0  https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg      1.0
1947          232.0  https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg      1.0
1949          146.0  https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg      1.0

        prediction_dog  prediction_confidence
14    golden_retriever               0.953442
16    golden_retriever               0.714719
17    golden_retriever               0.469760
39    golden_retriever               0.943082
44    golden_retriever               0.762211
…                …                    …
1903  golden_retriever               0.471351
1911  golden_retriever               0.765266
1936  golden_retriever               0.858744
1947  golden_retriever               0.775930
1949  golden_retriever               0.007959

[153 rows x 15 columns]
```

```
[308]: df.sort_values(["retweet_count"], ascending=False).head(1)
```

```
[308]:                tweet_id        tweet_source  \
       796  744234799360020481  Twitter for iPhone

                                                   tweet_text  \
       796  Here's a doggo realizing you can stand in a po…

                                                expanded_urls  rating_out_of_ten  \
       796  https://twitter.com/dog_rates/status/744234799…                 13.0

           dog_name  tweet_date  tweet_time dogtionary  favorite_count  \
       796      NaN  2016-06-18          18      doggo        131075.0
```

```
     retweet_count                                            jpg_url  \
796        79515.0  https://pbs.twimg.com/ext_tw_video_thumb/74423…

     img_num      prediction_dog  prediction_confidence
796      1.0  Labrador_retriever               0.825333
```

[309]:

```
df.sort_values(["favorite_count"], ascending=False).head(1)
```

[309]:

```
                tweet_id        tweet_source  \
309  822872901745569793  Twitter for iPhone

                                           tweet_text  \
309  Here's a super supportive puppo participating …

                                        expanded_urls  rating_out_of_ten  \
309  https://twitter.com/dog_rates/status/822872901…               13.0

    dog_name  tweet_date  tweet_time dogtionary  favorite_count  \
309      NaN  2017-01-21          18      puppo        132810.0

     retweet_count                                          jpg_url  img_num  \
309        48265.0  https://pbs.twimg.com/media/C2tugXLXgAArJO4.jpg      1.0

       prediction_dog  prediction_confidence
309  Lakeland_terrier               0.196015
```

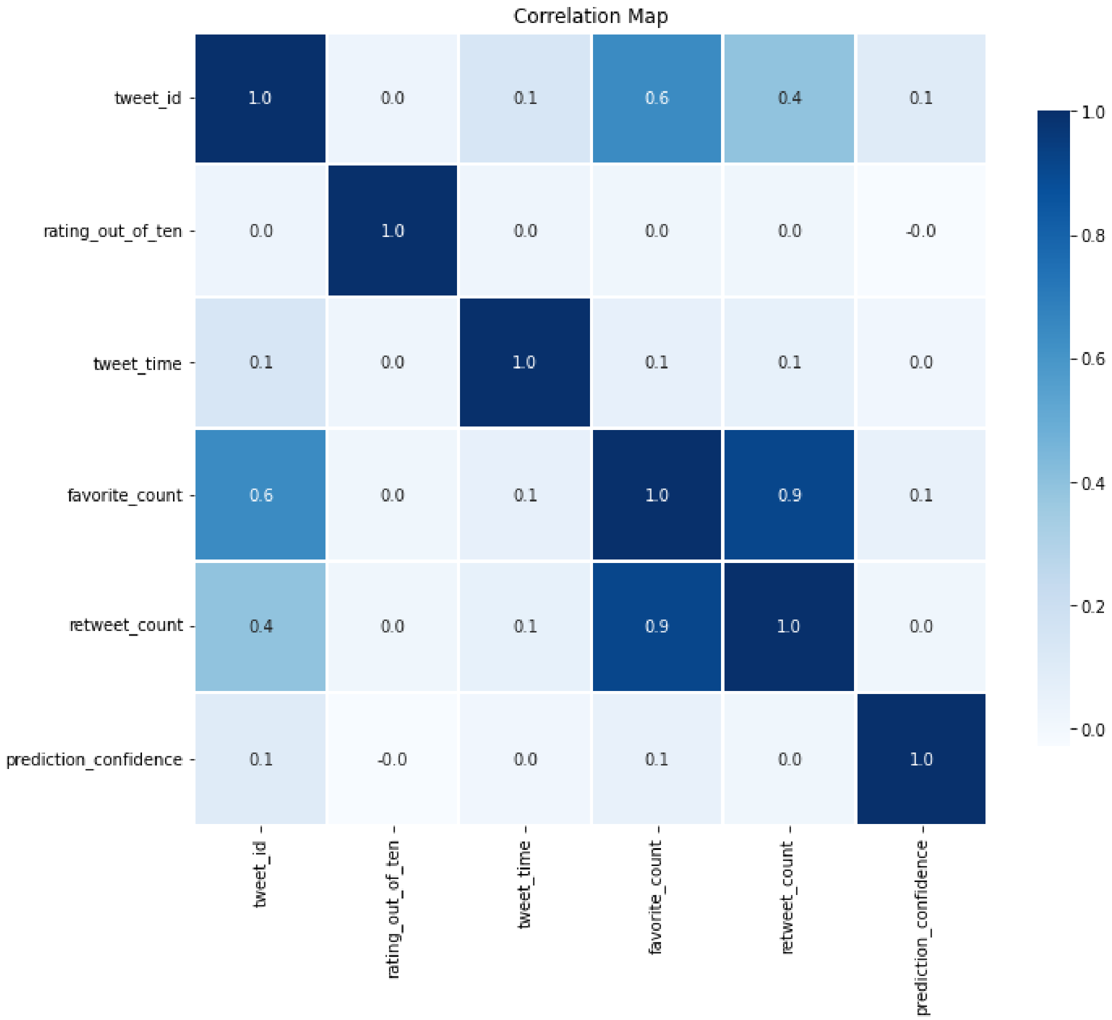
3.0.1 Insights:

1. Most Popular dogtionary: "pupper"
2. Most Popular tweet_source: "Twitter for iPhone"
3. Most Popular tweet_date: "2015-11-28"
4. Most Popular tweet_time: at "1 o'clock"
5. Most Popular dog_name: "Charlie"
6. Most prediction_dog: "golden_retriever"
7. The most retweet: 79515 times
8. The most favorite: 132810 times

3.0.2 Visualization

[310]:

```
import seaborn as sns
f,ax = plt.subplots(figsize=(11, 9));
sns.heatmap(df[[
    'tweet_id', 'tweet_source',
    'rating_out_of_ten','tweet_time','favorite_count',
    'retweet_count','prediction_confidence'
]].corr(),
            cmap="Blues", square=True, annot=True, fmt= '.1f',ax=ax,
           linewidth=0.2, cbar_kws={"shrink": .8});
plt.title('Correlation Map');
```



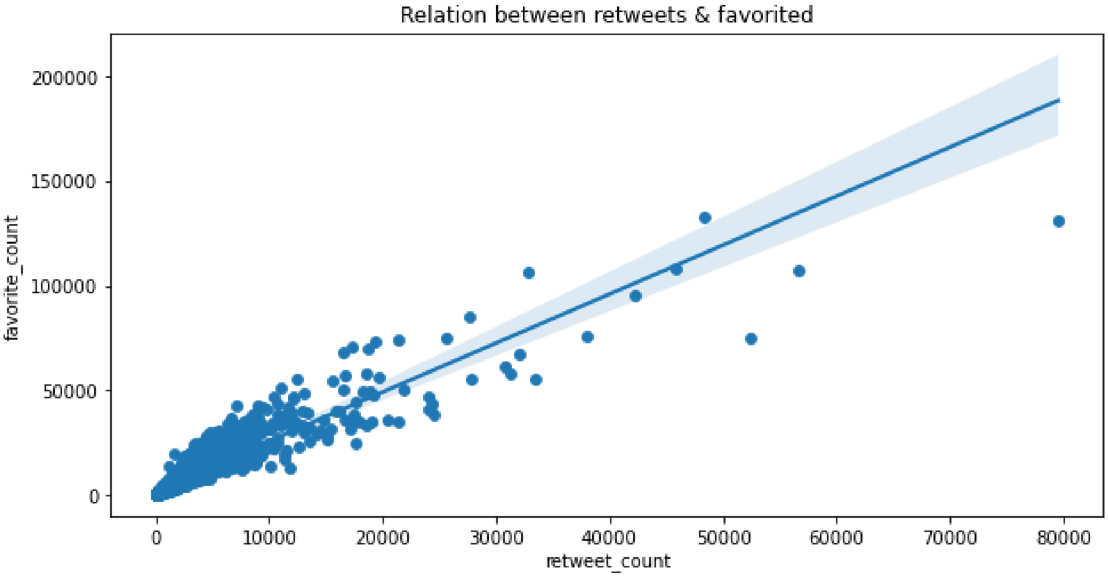
4 Analyze heatmap

1. From the map There is only one strong correlation between **'favorite_count'** and **'retweet_count'** : This is obviously the most favorite , the most retweeted
2. From the map There is moderator correlation between **'favorite_count'** and **'tweet_id'** : This is Normal the most favorite , the more tweets
3. 'prediction_confidence' don't have any relation between with any other ploted variable
4. 'rating_out_of_ten' don't have any relation between with any other ploted vari-

able

5. ‘tweet_time’ have week corelation with

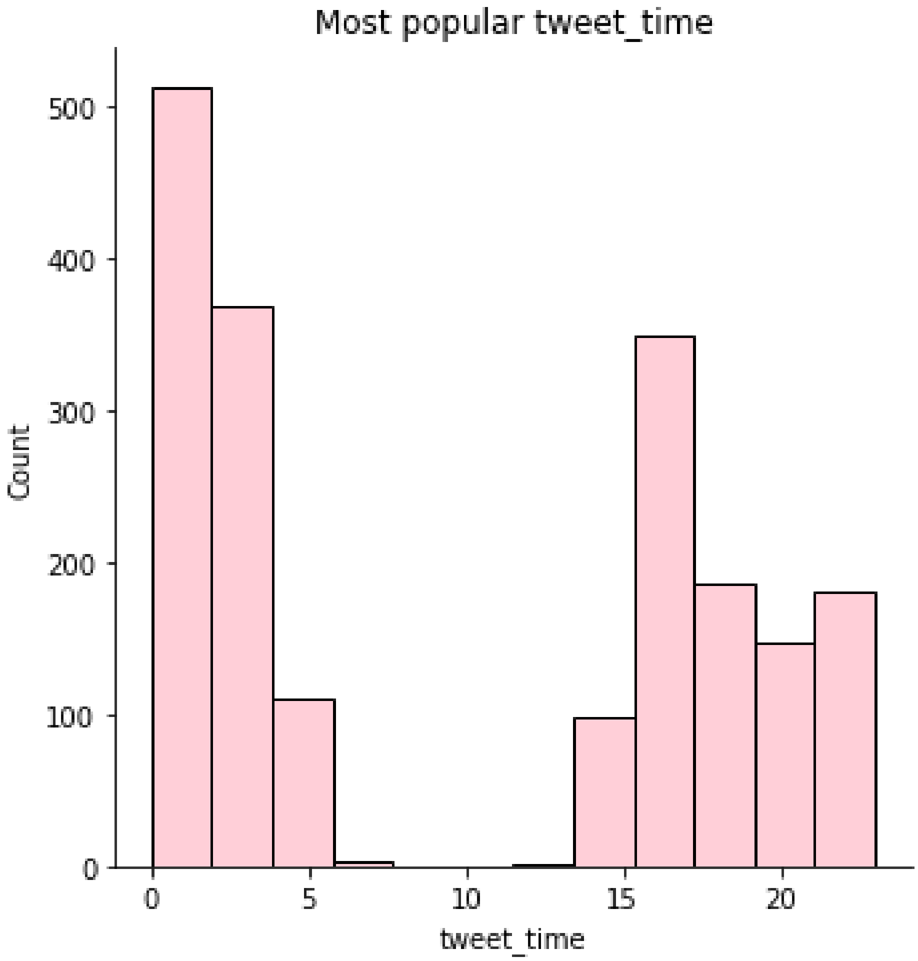
[311]:
```
# plot the relationship between 'favorite_count' and 'retweet_count'
#  define plot size
plt.figure(figsize=(10, 5));
sns.scatterplot(data = df, x = 'retweet_count', y = 'favorite_count')
sns.regplot(x="retweet_count",
            y="favorite_count",
            data=df);
plt.title("Relation between retweets & favorited");

plt.show()
```



5 Analyze scatterplot

Positive linear relationship between the number of retweets and the number of favorited tweets

[312]:
```
# Plot 'tweet_time'
sns.displot(df, x='tweet_time', color="pink");
plt.title("Most popular tweet_time");
```

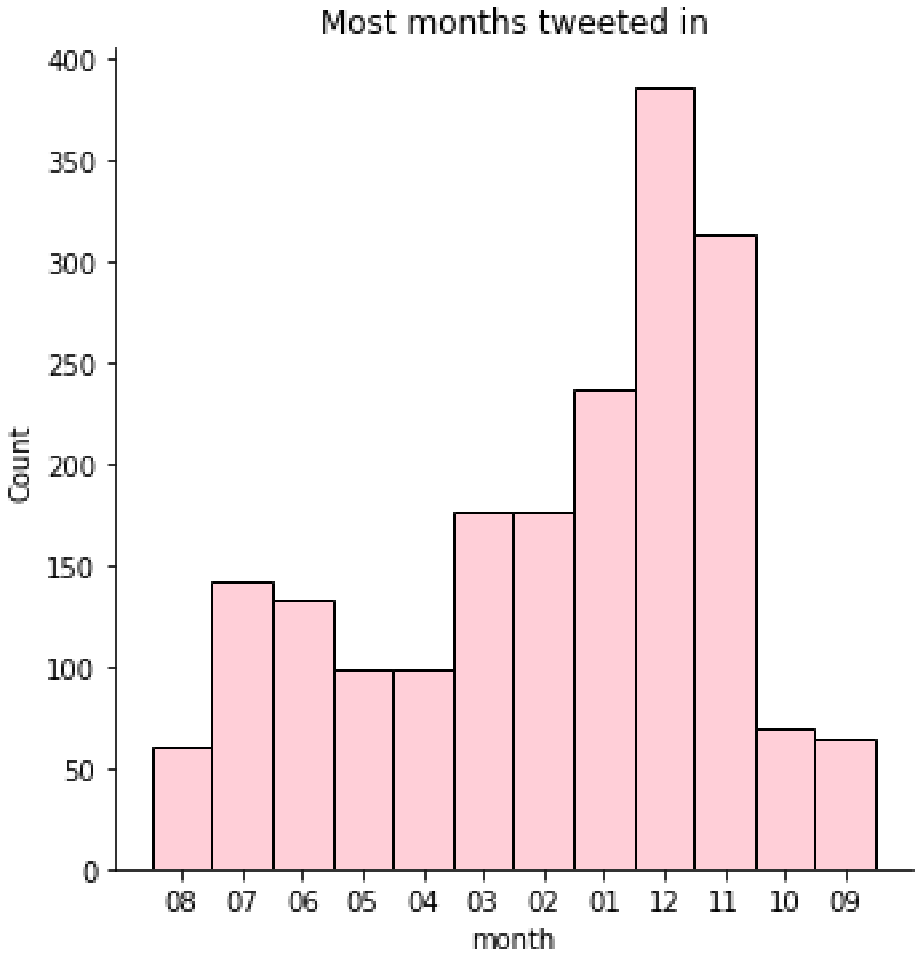


6 Analyze Histogram

From the hist plot we observed: 1. Tweets are highest between 0:00 and 3:00 a.m 2. There are no tweets between 6 am and 2 pm

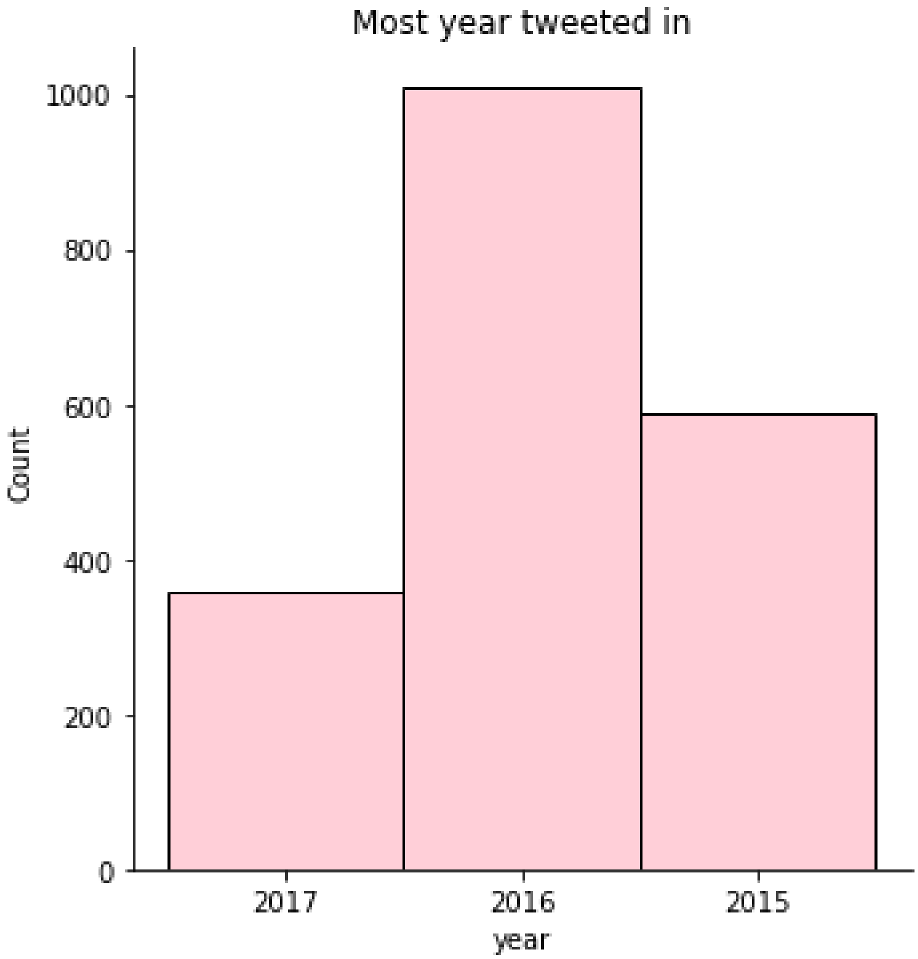
[313]:
```
#Check most days and months and years tweeted in
df[["year", "month", "day"]] = df["tweet_date"].str.split("-", expand = True)
```

[314]:
```
#plot Most months tweeted in
sns.displot(df, x="month", color="pink");
plt.title("Most months tweeted in");
```



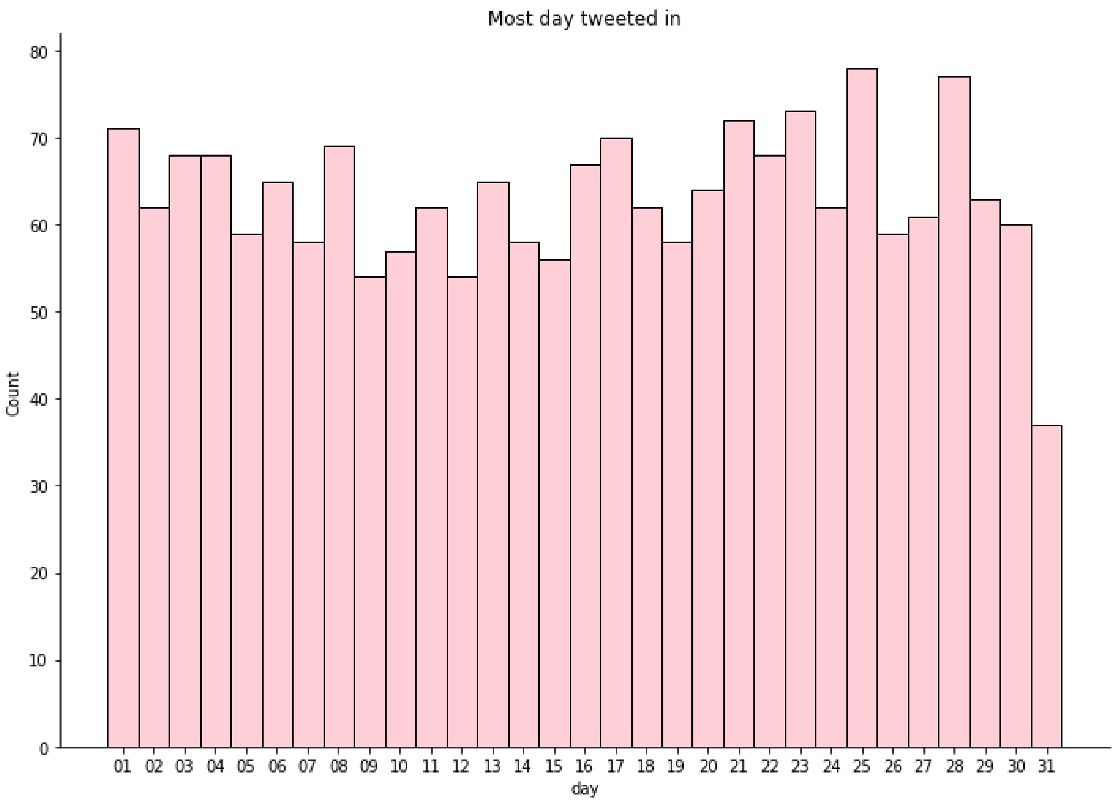
December is the Most Month tweeted in

[315]:
```
#plot Most year tweeted in
sns.displot(df, x="year", color="pink");
plt.title("Most year tweeted in");
```



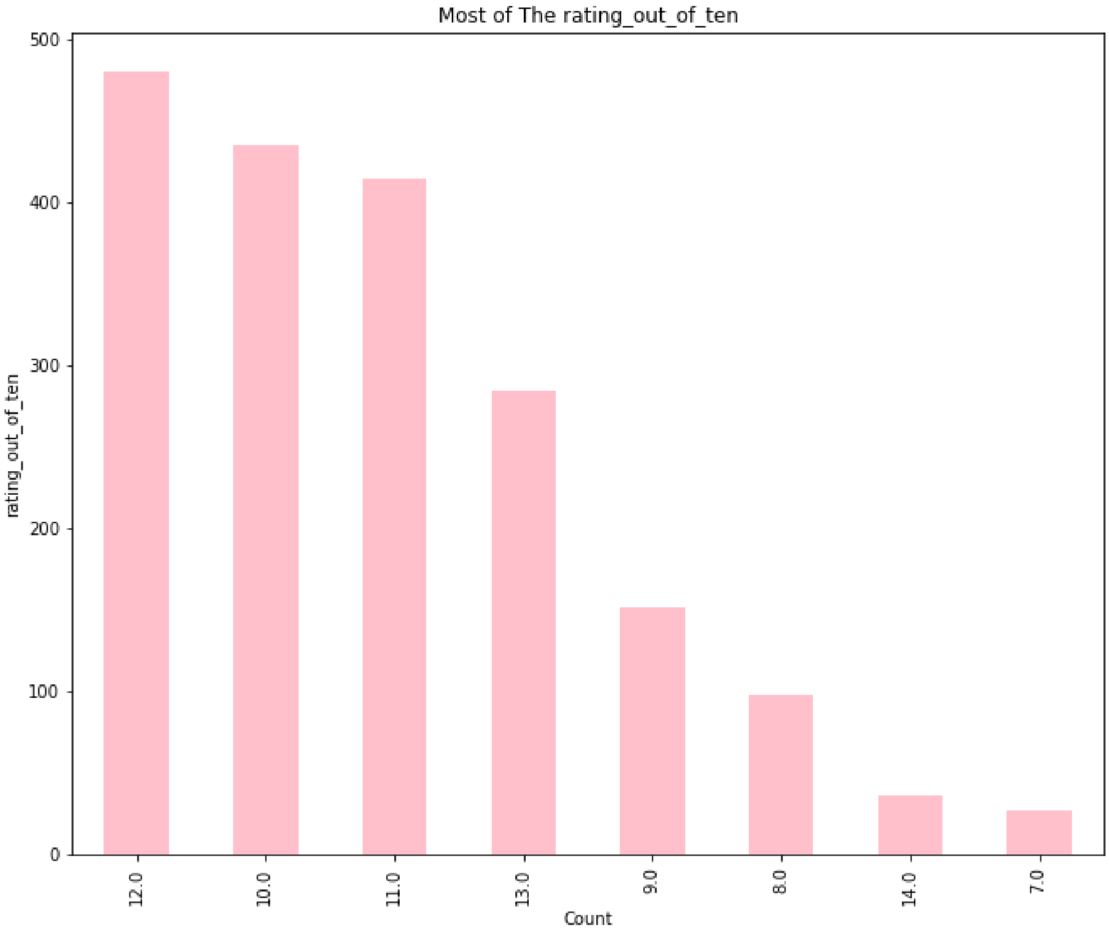
2016 is the Most year tweeted in

[316]:
```
#plot Most day tweeted in
days = df.sort_values("day", ascending=True)
sns.displot(days, x="day", height=7, aspect=10/7, color="pink");
plt.title("Most day tweeted in");
```



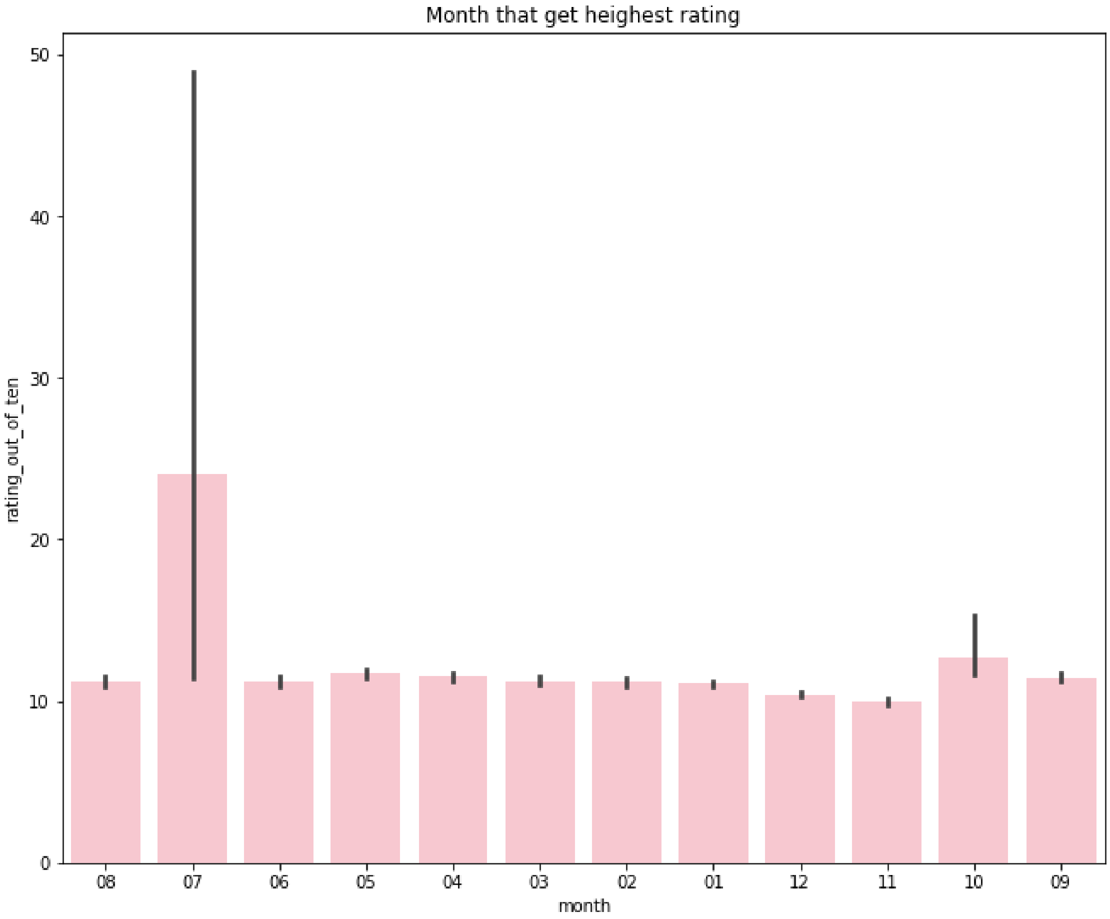
days 25, 28 , 01 Are the most days tweeted in

[317]:
```
#plot rating_out_of_ten
plt.figure(figsize=(11, 9));
rating_out_of_ten = df.groupby('rating_out_of_ten').filter(lambda x: len(x) >=␣
 ↪20)
rating_out_of_ten['rating_out_of_ten'].value_counts().plot(kind = 'bar',␣
 ↪color='pink')
plt.title('Most of The rating_out_of_ten');
plt.xlabel('Count');
plt.ylabel('rating_out_of_ten');
```



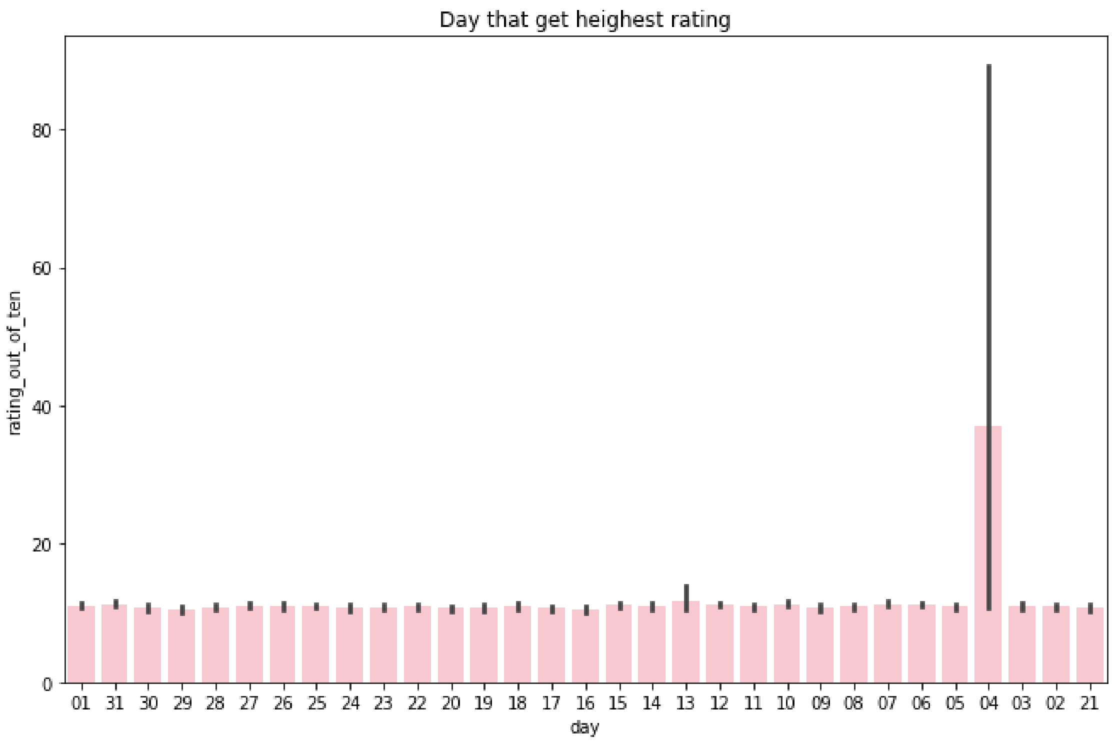
More then 75% of tweet has 12/10 rating The min rating is 2/10

```
[318]: # plot the month that get heighest rating
plt.figure(figsize=(11, 9));
sns.set_color_codes("pastel")
sns.barplot(x="month", y="rating_out_of_ten", data=df,
            label="Rating", color="pink");
plt.title("Month that get heighest rating");
```



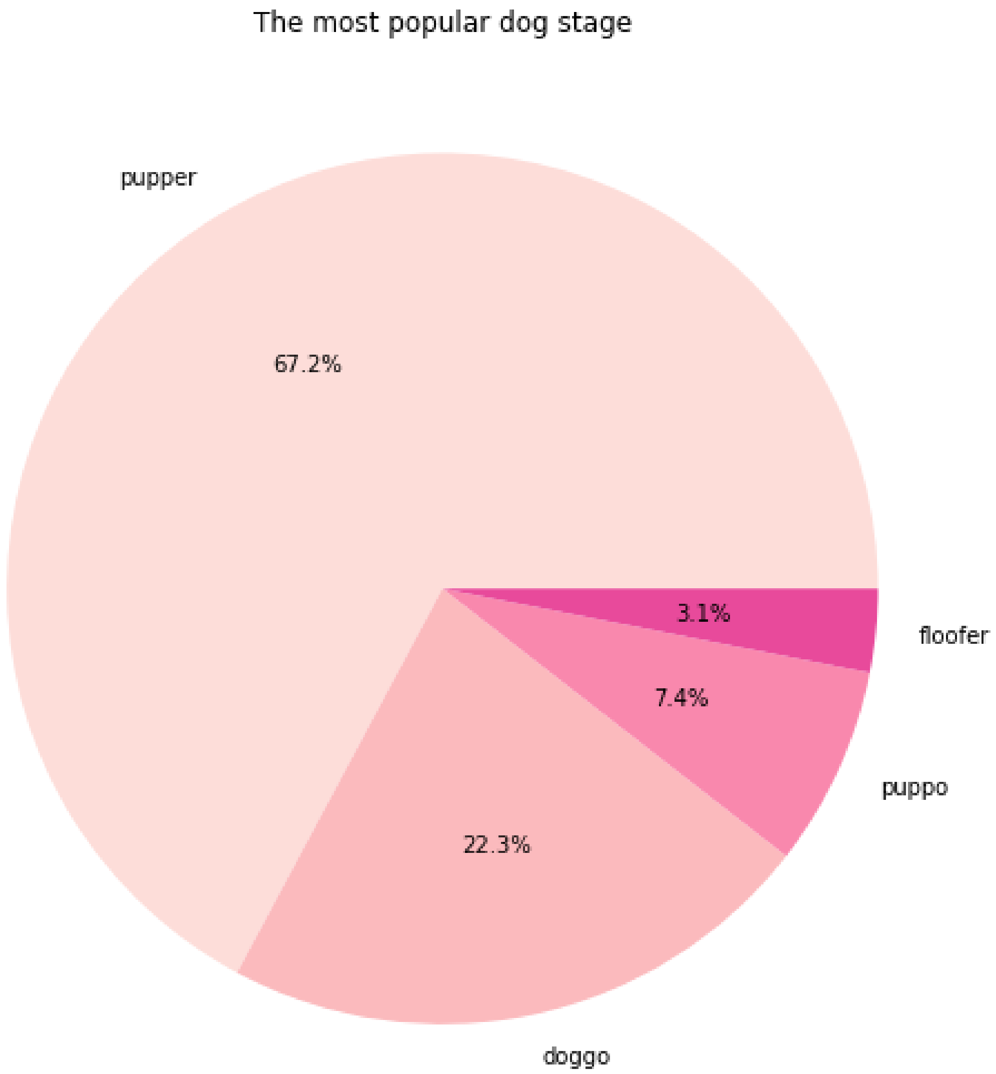
July is the month that get highest rating

[319]:
```
# plot the day that get heighest rating
plt.figure(figsize=(11, 7));
sns.set_color_codes("pastel")
sns.barplot(x="day", y="rating_out_of_ten", data=df,
            label="Rating", color="pink");
plt.title("Day that get heighest rating");
```



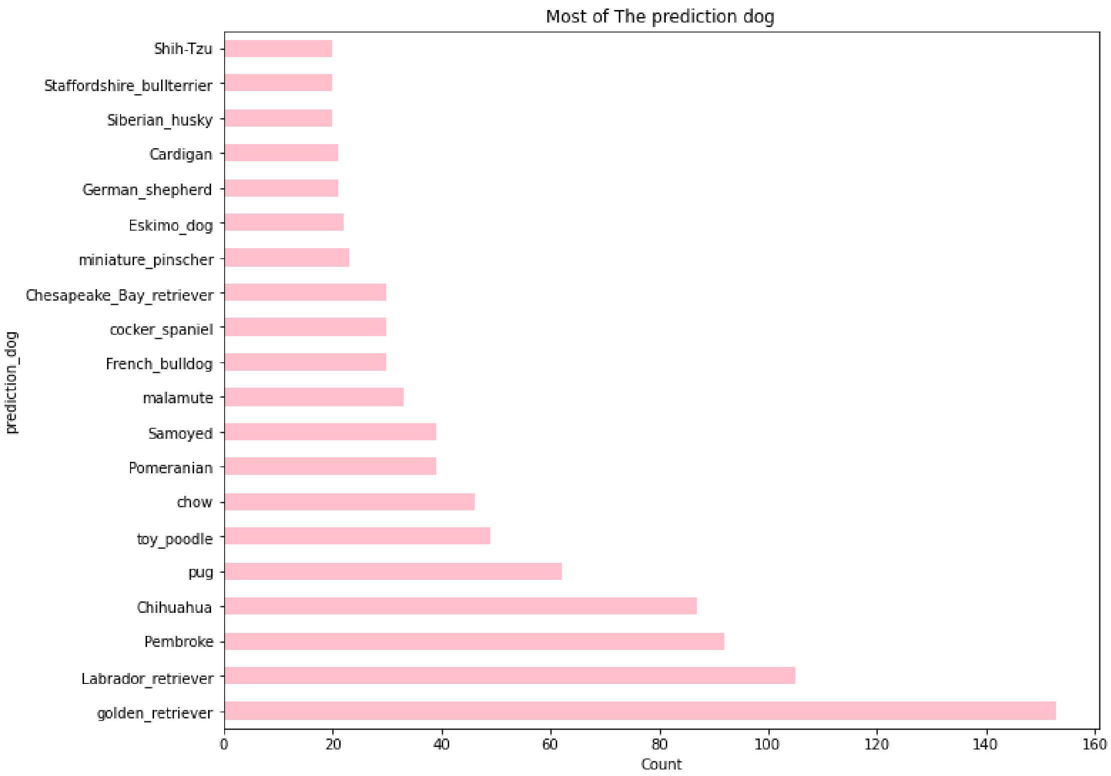
The 4th day of the month is the day that get Highest rating

[320]:
```
#plot the most popular dog stage
plt.figure(figsize=(11, 9));
colors = sns.color_palette('RdPu')
labels= ['pupper','doggo','puppo','floofer']
plt.pie(df['dogtionary'].value_counts(),labels=labels, colors = colors,␣
 ↪autopct='%1.1f%%')
plt.title('The most popular dog stage')
plt.show()
```



Pupper is the most popular dog stage

```
[321]: # specify the lengh to get good result
       plt.figure(figsize=(11, 9));
       prediction_dog = df.groupby('prediction_dog').filter(lambda x: len(x) >= 20)
       prediction_dog['prediction_dog'].value_counts().plot(kind = 'barh',␣
        ↪color='pink')
       plt.title('Most of The prediction dog');
       plt.xlabel('Count');
       plt.ylabel('prediction_dog');
```



Golden_retriever is the most prediction dog

https://amueller.github.io/word_cloud/auto_examples/masked.html#sphx-glr-download-auto-examples-masked-py

[322]:

```python
#  creating wordcloud with most frequent names for Dogs
from os import path
from PIL import Image
import numpy as np
import matplotlib.pyplot as plt
import os

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator

# get data directory (using getcwd() is needed to support running example in generated IPython notebook)
d = path.dirname(__file__) if "__file__" in locals() else os.getcwd()


#  create the word cloud text
text = []
for name in df.dog_name.values:
    # excluding NaN
    if name != np.nan:
        text.append(name)
text = ' '.join(map(str, text))
text = text.replace('nan','')


# read the image from directory path
dog_mask = np.array(Image.open(path.join(d, "img/golden-retriever.png")))

wc = WordCloud(
    width = 500,
    height = 480,
    background_color="#fcf5f9",
    max_words=5000,
    mask = dog_mask,
    colormap='cool',
    max_font_size=100,
    random_state=42
)
# generate word cloud
wc.generate(text)

#  show word cloud
plt.figure(figsize=(20, 10));
plt.axis("off")
plt.title("Popular Dog Names");
plt.imshow(wc, interpolation="bilinear");
```

Popular Dog Names
Doc
Coco
Maxaroni
Bentley
Rocco
Rory
Chip
Beau
Riley
Fred
Frankie
Otis
Colby
Opal
Hunter
Penny
Jack
Belle
Curtis
Vincent
Earl
Buddy
Cash
Shadow
Daisy
George
Jimothy
Sampson
Axel
Hank
Butter
Griffin
Flavio
Loki
Stanley
Max
Raymond
Ruby
Sophie
Maggie
Terry
Olivia
Franklin
Kilo
Ava
Charlie
Finley
Sansa
Coops
Yogi
Skittle
Nala
Calbert
Moose
Arnie
Sebastian
Milo
Scout
Oscar
Fiona
Frank
Smokey
Wyatt
Kevin
Pickles
Zeke
Rosie
Albert
Chuckles
Benji
Theodore
Neptune
Hobbes
Mia
Lucy
Scooter
Aspen
Bo
Gus
Oliver
Gary
Abby
Rufus
Zoey
Jimison
Blitz
Winnie
Cupcake
Chipson
Baxter
Watson
Oshie
Nelly
Koda
Bailey
Cooper
Sweet
Luca
Brody
Malcolm
Larry
Clark
Lennon
Louis
Maddie
Jesse
Ellie
Lily
Sammy
Jeffrey
Bruce
Lou
Reginald
Carl
Rocky
Lola
Sadie
Cody
Walter
Keith
Misty
Klevin
Dexter
Dash
Dakota
Bear
Dave
Moe
Archie
Ted
Winston
Peaches
Atlas
Samson
Rusty
Mister
Calvin
Odie
Layla
Lorenzo
Chester
Boomer
Sunny
Wilson
Kyle
Lincoln
Phil
Toby
Gizmo
Percy
Olive
Wallace
Tucker
Herald
Reggie
Olivier
Duke
Trooper
Bella
Finn
Roosevelt
Romeo
Eli
Louie
Waffles
Luna
Alfie
Jax
Albus
Indie
Cassie
Juno
Derek
Gabe
Clarence
Ash
Leo
Pippa
Bubbles
Gerald
Oakley
Kenneth
Wally
Maximus
Phineas
Doug
Reese
Piper
Atticus
Sugar
Phred
Linda
Kirby
Titan
CeCe
Stuart

Go up