wrangle_act

April 3, 2018

1

WeRateDogs “Wow!”
* * * * * * 1) 2)

2

2.1 Uda

2.1.1

-
  - DONE“”3
  - DONE“”3 pandas

2.1.2

-
  - DONE Jupyter Notebook Excel
  - DONE pandas /
-
  - DONE 82

2.1.3

-
  - DONE
-
  - DONE
  - DONE Python pandas
  - DONE

2.1.4

-
    - DONE CSV SQLite
- /)
    - DONE Jupyter Notebook pandas SQL 3
    - DONE Jupyter Notebook Python Tableau 1
    - DONE

2.1.5

-
    - DONE wrangle_report.pdf300-600
-
    - DONE31act_report.pdf 250

2.1.6

-
    - DONEwrangle_act.ipynb
    - DONEwrangle_report.pdf
    - DONEact_report.pdf
    - DONE

2.2

- /
-
-
-
-
-

tweet-archive-master * tweet_id df_tweeter * text df_tweeter * timestamp df_tweeter * retweet-count df_json * favorite-count df_json * rating df_tweeter * dog-name df_tweeter * dog-status df_tweeter * dog-types (df_pred) * dog-prediction probabily (df_pred) — # Update 2018_04_03 ## : * 2 1 2

- text
- This is | Meet | name is | Say hello to | named namedThis is | Meet | namenamedname

```
In [1]: #
        import pandas as pd
        import numpy as np
        import requests
import tweepy
import json
import os
import seaborn as sns
from IPython.display import display
import matplotlib.pyplot as plt
import re
import statsmodels.formula.api as smf
%matplotlib inline
%config InlineBackend.figure_format='retina'
#import wordcloud
```

3

3.1

```
In [2]: # image_predictions.tsv
        url1 = 'https://raw.githubusercontent.com/udacity/new-dand-advanced-china/master/%E6%95%
        with open('image_predictions.tsv' , 'wb') as file:
            image_f = requests.get(url1)
            file.write(image_f.content)
        image_predictions = pd.read_csv('image_predictions.tsv' , sep = '\t')
        #image_predictions.head(5)
```

```
In [3]: # twitter-archive-enhanced.csv
        url2 = 'https://raw.githubusercontent.com/udacity/new-dand-advanced-china/master/%E6%95%
        with open('twitter-archive-enhanced.csv' , 'wb') as file:
            twitter_f = requests.get(url2)
            file.write(twitter_f.content)
        twitter_archive_enhanced = pd.read_csv('twitter-archive-enhanced.csv' , sep = ',')
        #twitter_archive_enhanced.head(5)
```

```
In [4]: # tweeterudacityjsonTXTprojecttweet_json.txt
        url3 = 'https://raw.githubusercontent.com/udacity/new-dand-advanced-china/master/%E6%95%

        with open('tweet_json.txt' , 'wb') as file:
            tweet_f = requests.get(url3)
            file.write(tweet_f.content)
        #test = open('tweet_json.txt' , 'r')
        #a = test.readlines()
        #a

        tweet_data = pd.DataFrame(columns=['tweet_id','retweet_count','favorite_count'])
        with open('tweet_json.txt','r')as file:
            for line in file.readlines():
                dic = json.loads(line)
                #print(dic['id'])
                tweet_id = dic['id']
                retweet_count = dic['retweet_count']
                favorite_count = dic['favorite_count']
                tweet_data = tweet_data.append({'tweet_id' :tweet_id,'retweet_count':retweet_cou
        #tweet_data.head(5)
```

```
In [5]: #
        df_pred= image_predictions.copy()
        df_twitter = twitter_archive_enhanced.copy()
        df_json = tweet_data.copy()
```

4

df_twitter, df_pred, df_json

4.1 df_twitter

```
In [6]: df_twitter
```

```
Out[6]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0     892420643555336193                    NaN                  NaN
        1     892177421306343426                    NaN                  NaN
        2     891815181378084864                    NaN                  NaN
        3     891689557279858688                    NaN                  NaN
        4     891327558926688256                    NaN                  NaN
        5     891087950875897856                    NaN                  NaN
        6     890971913173991426                    NaN                  NaN
        7     890729181411237888                    NaN                  NaN
        8     890609185150312448                    NaN                  NaN
        9     890240255349198849                    NaN                  NaN
        10    890006608113172480                    NaN                  NaN
        11    889880896479866881                    NaN                  NaN
        12    889665388333682689                    NaN                  NaN
        13    889638837579907072                    NaN                  NaN
        14    889531135344209921                    NaN                  NaN
        15    889278841981685760                    NaN                  NaN
        16    888917238123831296                    NaN                  NaN
        17    888804989199671297                    NaN                  NaN
        18    888554962724278272                    NaN                  NaN
        19    888202515573088257                    NaN                  NaN
        20    888078434458587136                    NaN                  NaN
        21    887705289381826560                    NaN                  NaN
        22    887517139158093824                    NaN                  NaN
        23    887473957103951883                    NaN                  NaN
        24    887343217045368832                    NaN                  NaN
        25    887101392804085760                    NaN                  NaN
        26    886983233522544640                    NaN                  NaN
        27    886736880519319552                    NaN                  NaN
        28    886680336477933568                    NaN                  NaN
```

```
29    886366144734445568                    NaN                  NaN
...                  ...                    ...                  ...
2326  666411507551481857                    NaN                  NaN
2327  666407126856765440                    NaN                  NaN
2328  666396247373291520                    NaN                  NaN
2329  666373753744588802                    NaN                  NaN
2330  666362758909284353                    NaN                  NaN
2331  666353288456101888                    NaN                  NaN
2332  666345417576210432                    NaN                  NaN
2333  666337882303524864                    NaN                  NaN
2334  666293911632134144                    NaN                  NaN
2335  666287406224695296                    NaN                  NaN
2336  666273097616637952                    NaN                  NaN
2337  666268910803644416                    NaN                  NaN
2338  666104133288665088                    NaN                  NaN
2339  666102155909144576                    NaN                  NaN
2340  666099513787052032                    NaN                  NaN
2341  666094000022159362                    NaN                  NaN
2342  666082916733198337                    NaN                  NaN
2343  666073100786774016                    NaN                  NaN
2344  666071193221509120                    NaN                  NaN
2345  666063827256086533                    NaN                  NaN
2346  666058600524156928                    NaN                  NaN
2347  666057090499244032                    NaN                  NaN
2348  666055525042405380                    NaN                  NaN
2349  666051853826850816                    NaN                  NaN
2350  666050758794694657                    NaN                  NaN
2351  666049248165822465                    NaN                  NaN
2352  666044226329800704                    NaN                  NaN
2353  666033412701032449                    NaN                  NaN
2354  666029285002620928                    NaN                  NaN
2355  666020888022790149                    NaN                  NaN

                      timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
5     2017-07-29 00:08:17 +0000
6     2017-07-28 16:27:12 +0000
7     2017-07-28 00:22:40 +0000
8     2017-07-27 16:25:51 +0000
9     2017-07-26 15:59:51 +0000
10    2017-07-26 00:31:25 +0000
11    2017-07-25 16:11:53 +0000
12    2017-07-25 01:55:32 +0000
13    2017-07-25 00:10:02 +0000
```

```
14    2017-07-24 17:02:04 +0000
15    2017-07-24 00:19:32 +0000
16    2017-07-23 00:22:39 +0000
17    2017-07-22 16:56:37 +0000
18    2017-07-22 00:23:06 +0000
19    2017-07-21 01:02:36 +0000
20    2017-07-20 16:49:33 +0000
21    2017-07-19 16:06:48 +0000
22    2017-07-19 03:39:09 +0000
23    2017-07-19 00:47:34 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
26    2017-07-17 16:17:36 +0000
27    2017-07-16 23:58:41 +0000
28    2017-07-16 20:14:00 +0000
29    2017-07-15 23:25:31 +0000
...                         ...
2326  2015-11-17 00:24:19 +0000
2327  2015-11-17 00:06:54 +0000
2328  2015-11-16 23:23:41 +0000
2329  2015-11-16 21:54:18 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2332  2015-11-16 20:01:42 +0000
2333  2015-11-16 19:31:45 +0000
2334  2015-11-16 16:37:02 +0000
2335  2015-11-16 16:11:11 +0000
2336  2015-11-16 15:14:19 +0000
2337  2015-11-16 14:57:41 +0000
2338  2015-11-16 04:02:55 +0000
2339  2015-11-16 03:55:04 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
2343  2015-11-16 01:59:36 +0000
2344  2015-11-16 01:52:02 +0000
2345  2015-11-16 01:22:45 +0000
2346  2015-11-16 01:01:59 +0000
2347  2015-11-16 00:55:59 +0000
2348  2015-11-16 00:49:46 +0000
2349  2015-11-16 00:35:11 +0000
2350  2015-11-16 00:30:50 +0000
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000
```

```
                                                 source  \
0     <a href="http://twitter.com/download/iphone" r...
1     <a href="http://twitter.com/download/iphone" r...
2     <a href="http://twitter.com/download/iphone" r...
3     <a href="http://twitter.com/download/iphone" r...
4     <a href="http://twitter.com/download/iphone" r...
5     <a href="http://twitter.com/download/iphone" r...
6     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
8     <a href="http://twitter.com/download/iphone" r...
9     <a href="http://twitter.com/download/iphone" r...
10    <a href="http://twitter.com/download/iphone" r...
11    <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
13    <a href="http://twitter.com/download/iphone" r...
14    <a href="http://twitter.com/download/iphone" r...
15    <a href="http://twitter.com/download/iphone" r...
16    <a href="http://twitter.com/download/iphone" r...
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
29    <a href="http://twitter.com/download/iphone" r...
...                                                 ...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
```

```
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2345  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
2348  <a href="http://twitter.com/download/iphone" r...
2349  <a href="http://twitter.com/download/iphone" r...
2350  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...

                                                   text  retweeted_status_id  \
0     This is Phineas. He's a mystical boy. Only eve...                  NaN
1     This is Tilly. She's just checking pup on you...                  NaN
2     This is Archie. He is a rare Norwegian Pouncin...                  NaN
3     This is Darla. She commenced a snooze mid meal...                  NaN
4     This is Franklin. He would like you to stop ca...                  NaN
5     Here we have a majestic great white breaching ...                  NaN
6     Meet Jax. He enjoys ice cream so much he gets ...                  NaN
7     When you watch your owner call another dog a g...                  NaN
8     This is Zoey. She doesn't want to be one of th...                  NaN
9     This is Cassie. She is a college pup. Studying...                  NaN
10    This is Koda. He is a South Australian decksha...                  NaN
11    This is Bruno. He is a service shark. Only get...                  NaN
12    Here's a puppo that seems to be on the fence a...                  NaN
13    This is Ted. He does his best. Sometimes that'...                  NaN
14    This is Stuart. He's sporting his favorite fan...                  NaN
15    This is Oliver. You're witnessing one of his m...                  NaN
16    This is Jim. He found a fren. Taught him how t...                  NaN
17    This is Zeke. He has a new stick. Very proud o...                  NaN
18    This is Ralphus. He's powering up. Attempting ...                  NaN
19    RT @dog_rates: This is Canela. She attempted s...         8.874740e+17
20    This is Gerald. He was just told he didn't get...                  NaN
21    This is Jeffrey. He has a monopoly on the pool...                  NaN
22    I've yet to rate a Venezuelan Hover Wiener. Th...                  NaN
23    This is Canela. She attempted some fancy porch...                  NaN
24    You may not have known you needed to see this ...                  NaN
25    This... is a Jubilant Antarctic House Bear. We...                  NaN
26    This is Maya. She's very shy. Rarely leaves he...                  NaN
27    This is Mingus. He's a wonderful father to his...                  NaN
28    This is Derek. He's late for a dog meeting. 13...                  NaN
29    This is Roscoe. Another pupper fallen victim t...                  NaN
...                                                 ...                  ...
2326  This is quite the dog. Gets really excited whe...                  NaN
```

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

      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                        NaN
1                          NaN                        NaN
2                          NaN                        NaN
3                          NaN                        NaN
4                          NaN                        NaN
5                          NaN                        NaN
6                          NaN                        NaN
7                          NaN                        NaN
8                          NaN                        NaN
9                          NaN                        NaN
10                         NaN                        NaN
11                         NaN                        NaN
12                         NaN                        NaN
13                         NaN                        NaN
14                         NaN                        NaN
15                         NaN                        NaN
16                         NaN                        NaN
```

```
17                           NaN                        NaN
18                           NaN                        NaN
19                  4.196984e+09  2017-07-19 00:47:34 +0000
20                           NaN                        NaN
21                           NaN                        NaN
22                           NaN                        NaN
23                           NaN                        NaN
24                           NaN                        NaN
25                           NaN                        NaN
26                           NaN                        NaN
27                           NaN                        NaN
28                           NaN                        NaN
29                           NaN                        NaN
...                          ...                        ...
2326                         NaN                        NaN
2327                         NaN                        NaN
2328                         NaN                        NaN
2329                         NaN                        NaN
2330                         NaN                        NaN
2331                         NaN                        NaN
2332                         NaN                        NaN
2333                         NaN                        NaN
2334                         NaN                        NaN
2335                         NaN                        NaN
2336                         NaN                        NaN
2337                         NaN                        NaN
2338                         NaN                        NaN
2339                         NaN                        NaN
2340                         NaN                        NaN
2341                         NaN                        NaN
2342                         NaN                        NaN
2343                         NaN                        NaN
2344                         NaN                        NaN
2345                         NaN                        NaN
2346                         NaN                        NaN
2347                         NaN                        NaN
2348                         NaN                        NaN
2349                         NaN                        NaN
2350                         NaN                        NaN
2351                         NaN                        NaN
2352                         NaN                        NaN
2353                         NaN                        NaN
2354                         NaN                        NaN
2355                         NaN                        NaN

                                          expanded_urls  rating_numerator  \
0     https://twitter.com/dog_rates/status/892420643...                13
1     https://twitter.com/dog_rates/status/892177421...                13
```

```
2     https://twitter.com/dog_rates/status/891815181...            12
3     https://twitter.com/dog_rates/status/891689557...            13
4     https://twitter.com/dog_rates/status/891327558...            12
5     https://twitter.com/dog_rates/status/891087950...            13
6     https://gofundme.com/ydvmve-surgery-for-jax,ht...            13
7     https://twitter.com/dog_rates/status/890729181...            13
8     https://twitter.com/dog_rates/status/890609185...            13
9     https://twitter.com/dog_rates/status/890240255...            14
10    https://twitter.com/dog_rates/status/890006608...            13
11    https://twitter.com/dog_rates/status/889880896...            13
12    https://twitter.com/dog_rates/status/889665388...            13
13    https://twitter.com/dog_rates/status/889638837...            12
14    https://twitter.com/dog_rates/status/889531135...            13
15    https://twitter.com/dog_rates/status/889278841...            13
16    https://twitter.com/dog_rates/status/888917238...            12
17    https://twitter.com/dog_rates/status/888804989...            13
18    https://twitter.com/dog_rates/status/888554962...            13
19    https://twitter.com/dog_rates/status/887473957...            13
20    https://twitter.com/dog_rates/status/888078434...            12
21    https://twitter.com/dog_rates/status/887705289...            13
22    https://twitter.com/dog_rates/status/887517139...            14
23    https://twitter.com/dog_rates/status/887473957...            13
24    https://twitter.com/dog_rates/status/887343217...            13
25    https://twitter.com/dog_rates/status/887101392...            12
26    https://twitter.com/dog_rates/status/886983233...            13
27    https://www.gofundme.com/mingusneedsus,https:/...            13
28    https://twitter.com/dog_rates/status/886680336...            13
29    https://twitter.com/dog_rates/status/886366144...            12
...                                                 ...           ...
2326  https://twitter.com/dog_rates/status/666411507...             2
2327  https://twitter.com/dog_rates/status/666407126...             7
2328  https://twitter.com/dog_rates/status/666396247...             9
2329  https://twitter.com/dog_rates/status/666373753...            11
2330  https://twitter.com/dog_rates/status/666362758...             6
2331  https://twitter.com/dog_rates/status/666353288...             8
2332  https://twitter.com/dog_rates/status/666345417...            10
2333  https://twitter.com/dog_rates/status/666337882...             9
2334  https://twitter.com/dog_rates/status/666293911...             3
2335  https://twitter.com/dog_rates/status/666287406...             1
2336  https://twitter.com/dog_rates/status/666273097...            11
2337  https://twitter.com/dog_rates/status/666268910...            10
2338  https://twitter.com/dog_rates/status/666104133...             1
2339  https://twitter.com/dog_rates/status/666102155...            11
2340  https://twitter.com/dog_rates/status/666099513...             8
2341  https://twitter.com/dog_rates/status/666094000...             9
2342  https://twitter.com/dog_rates/status/666082916...             6
2343  https://twitter.com/dog_rates/status/666073100...            10
2344  https://twitter.com/dog_rates/status/666071193...             9
```

```
2345  https://twitter.com/dog_rates/status/666063827...                 10
2346  https://twitter.com/dog_rates/status/666058600...                  8
2347  https://twitter.com/dog_rates/status/666057090...                  9
2348  https://twitter.com/dog_rates/status/666055525...                 10
2349  https://twitter.com/dog_rates/status/666051853...                  2
2350  https://twitter.com/dog_rates/status/666050758...                 10
2351  https://twitter.com/dog_rates/status/666049248...                  5
2352  https://twitter.com/dog_rates/status/666044226...                  6
2353  https://twitter.com/dog_rates/status/666033412...                  9
2354  https://twitter.com/dog_rates/status/666029285...                  7
2355  https://twitter.com/dog_rates/status/666020888...                  8

      rating_denominator      name  doggo floofer  pupper  puppo
0                     10   Phineas   None    None    None   None
1                     10     Tilly   None    None    None   None
2                     10    Archie   None    None    None   None
3                     10     Darla   None    None    None   None
4                     10  Franklin   None    None    None   None
5                     10      None   None    None    None   None
6                     10       Jax   None    None    None   None
7                     10      None   None    None    None   None
8                     10      Zoey   None    None    None   None
9                     10    Cassie  doggo    None    None   None
10                    10      Koda   None    None    None   None
11                    10     Bruno   None    None    None   None
12                    10      None   None    None    None  puppo
13                    10       Ted   None    None    None   None
14                    10    Stuart   None    None    None  puppo
15                    10    Oliver   None    None    None   None
16                    10       Jim   None    None    None   None
17                    10      Zeke   None    None    None   None
18                    10   Ralphus   None    None    None   None
19                    10    Canela   None    None    None   None
20                    10    Gerald   None    None    None   None
21                    10   Jeffrey   None    None    None   None
22                    10      such   None    None    None   None
23                    10    Canela   None    None    None   None
24                    10      None   None    None    None   None
25                    10      None   None    None    None   None
26                    10      Maya   None    None    None   None
27                    10    Mingus   None    None    None   None
28                    10     Derek   None    None    None   None
29                    10    Roscoe   None    None  pupper   None
...                  ...       ...    ...     ...     ...    ...
2326                  10     quite   None    None    None   None
2327                  10         a   None    None    None   None
2328                  10      None   None    None    None   None
2329                  10      None   None    None    None   None
```

```
        2330                  10    None   None    None    None   None
        2331                  10    None   None    None    None   None
        2332                  10    None   None    None    None   None
        2333                  10      an   None    None    None   None
        2334                  10       a   None    None    None   None
        2335                   2      an   None    None    None   None
        2336                  10    None   None    None    None   None
        2337                  10    None   None    None    None   None
        2338                  10    None   None    None    None   None
        2339                  10    None   None    None    None   None
        2340                  10    None   None    None    None   None
        2341                  10    None   None    None    None   None
        2342                  10    None   None    None    None   None
        2343                  10    None   None    None    None   None
        2344                  10    None   None    None    None   None
        2345                  10     the   None    None    None   None
        2346                  10     the   None    None    None   None
        2347                  10       a   None    None    None   None
        2348                  10       a   None    None    None   None
        2349                  10      an   None    None    None   None
        2350                  10       a   None    None    None   None
        2351                  10    None   None    None    None   None
        2352                  10       a   None    None    None   None
        2353                  10       a   None    None    None   None
        2354                  10       a   None    None    None   None
        2355                  10    None   None    None    None   None

        [2356 rows x 17 columns]
```

```
In [7]: df_twitter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2356 non-null int64
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2356 non-null object
source                        2356 non-null object
text                          2356 non-null object
retweeted_status_id           181 non-null float64
retweeted_status_user_id      181 non-null float64
retweeted_status_timestamp    181 non-null object
expanded_urls                 2297 non-null object
rating_numerator              2356 non-null int64
rating_denominator            2356 non-null int64
name                          2356 non-null object
doggo                         2356 non-null object
```

```
floofer                       2356 non-null object
pupper                        2356 non-null object
puppo                         2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

```
In [8]: #tweet_id
        temp = df_twitter.drop(['in_reply_to_status_id','in_reply_to_user_id','timestamp','sourc
        #temp
        temp_dog = pd.melt(temp, id_vars=["tweet_id"], value_vars=["doggo","floofer","pupper","p
        #temp_dog
        temp_dog = temp_dog[temp_dog.value!='None']
        temp_dog = temp_dog.drop(['variable'],axis=1)
        a = temp_dog.drop_duplicates(subset=['tweet_id'],keep='first')
        b = temp_dog.drop_duplicates(subset=['tweet_id'],keep=False)
        dupl_dog = a.append(b).drop_duplicates(subset=['tweet_id'],keep=False)
        dupl_dog_list = list(dupl_dog.tweet_id)
        dupl_dog.tweet_id
```

```
Out[8]: 191     855851453814013952
        200     854010172552949760
        460     817777686764523521
        531     808106460588765185
        565     802265048156610565
        575     801115127852503040
        705     785639753186217984
        733     781308096455073793
        778     775898661951791106
        822     770093767776997377
        889     759793422261743616
        956     751583847268179968
        1063    741067306818797568
        1113    733109485275860992
        Name: tweet_id, dtype: int64
```

```
In [9]: #
        none_list = temp['tweet_id'][(temp.doggo == 'None') & (temp.floofer == 'None') & (temp.p
        none_list
```

```
Out[9]: 0       892420643555336193
        1       892177421306343426
        2       891815181378084864
        3       891689557279858688
        4       891327558926688256
        5       891087950875897856
        6       890971913173991426
        7       890729181411237888
        8       890609185150312448
```

10	890006608113172480
11	889880896479866881
13	889638837579907072
15	889278841981685760
16	888917238123831296
17	888804989199671297
18	888554962724278272
19	888202515573088257
20	888078434458587136
21	887705289381826560
22	887517139158093824
23	887473957103951883
24	887343217045368832
25	887101392804085760
26	886983233522544640
27	886736880519319552
28	886680336477933568
30	886267009285017600
31	886258384151887873
32	886054160059072513
33	885984800019947520
	...
2326	666411507551481857
2327	666407126856765440
2328	666396247373291520
2329	666373753744588802
2330	666362758909284353
2331	666353288456101888
2332	666345417576210432
2333	666337882303524864
2334	666293911632134144
2335	666287406224695296
2336	666273097616637952
2337	666268910803644416
2338	666104133288665088
2339	666102155909144576
2340	666099513787052032
2341	666094000022159362
2342	666082916733198337
2343	666073100786774016
2344	666071193221509120
2345	666063827256086533
2346	666058600524156928
2347	666057090499244032
2348	666055525042405380
2349	666051853826850816
2350	666050758794694657
2351	666049248165822465

```
2352    666044226329800704
2353    666033412701032449
2354    666029285002620928
2355    666020888022790149
Name: tweet_id, Length: 1976, dtype: int64
```

```
In [10]: #
         temp_dog.value.value_counts()
         df_twitter.name.value_counts()
         df_twitter.rating_numerator.value_counts()
         df_twitter.rating_denominator.value_counts()
```

```
Out[10]: 10     2333
         11        3
         50        3
         80        2
         20        2
         2         1
         16        1
         40        1
         70        1
         15        1
         90        1
         110       1
         120       1
         130       1
         150       1
         170       1
         7         1
         0         1
         Name: rating_denominator, dtype: int64
```

```
In [11]: df_twitter.text.sample(20)
```

```
Out[11]: 734     This is Oakley. He just got yelled at for goin...
         1647    Breathtaking pupper here. Should be on the cov...
         2188    This is Jeremy. He hasn't grown into his skin ...
         1309    Say hello to Cupcake. She's an Icelandic Dippe...
         376     This is Sailer. He waits on the roof for his o...
         548     Meet Sansa and Gary. They run along the fence ...
         380     Meet Tucker. It's his birthday. He's pupset wi...
         921     Here's a heartwarming scene of a single father...
         44      This is Noah. He can't believe someone made th...
         1006    This is Keurig. He apparently headbutts other ...
         2034    This is a Tuscaloosa Alcatraz named Jacob (Yac...
         629     RT @dog_rates: This is Butter. She can have wh...
         1260    The squad is back for St. Patrick's Day!  \n...
         985     This is Boomer. He's self-baptizing. Other dog...
         1278    This is Lucy. She doesn't understand fetch. 8/...
```

```
1168    This is Oliver. Bath time is upon him. His fea...
1617    Meet Gerbald. He just found out he's adopted. ...
1357    This pupper doesn't understand gates. 10/10 so...
1735    This is Hunter. He was playing with his ball m...
1505    We usually don't rate penguins but this one is...
Name: text, dtype: object
```

- **tweet_idint64in_reply_to_status_id & in_reply_to_user_id 64IDQ**
- **timestampobjecttimedate** (Q)
- **name** (Q)
- (Q)
- (Q)
- **expanded_urls2297 non-null object2356** (Q)
- T
- **retweeted_count favorite_count df_jsonQ**
- **sourceT**
- **RatingQ**
- **np.nan'None'Q**

4.2 df_pred

```
In [12]: df_pred
```

```
Out[12]:                  tweet_id                                          jpg_url  \
         0      666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
         1      666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
         2      666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
         3      666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
         4      666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
         5      666050758794694657  https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
         6      666051853826850816  https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
         7      666055525042405380  https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
         8      666057090499244032  https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
         9      666058600524156928  https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
         10     666063827256086533  https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
         11     666071193221509120  https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg
         12     666073100786774016  https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
         13     666082916733198337  https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg
         14     666094000022159362  https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
         15     666099513787052032  https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
         16     666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
         17     666104133288665088  https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg
         18     666268910803644416  https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
         19     666273097616637952  https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
         20     666287406224695296  https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
         21     666293911632134144  https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
         22     666337882303524864  https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
         23     666345417576210432  https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
```

```
24    666353288456101888    https://pbs.twimg.com/media/CT9cx0tUEAAhNN_.jpg
25    666362758909284353    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26    666373753744588802    https://pbs.twimg.com/media/CT9vZEYWUAAlZ05.jpg
27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    666411507551481857    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...                  ...                                                ...
2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046  886680336477933568    https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg
2049  887101392804085760    https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050  887343217045368832  https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051  887473957103951883    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052  887517139158093824  https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053  887705289381826560    https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054  888078434458587136    https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  888202515573088257    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  888554962724278272    https://pbs.twimg.com/media/DFTH_O-UQAACu20.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  889278841981685760  https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921    https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061  889638837579907072    https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881    https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.jpg
2066  890609185150312448    https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067  890729181411237888    https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068  890971913173991426    https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069  891087950875897856    https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070  891327558926688256    https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071  891689557279858688    https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  891815181378084864    https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  892177421306343426    https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  892420643555336193    https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

      img_num                           p1   p1_conf  p1_dog  \
0           1       Welsh_springer_spaniel  0.465074    True
1           1                      redbone  0.506826    True
2           1              German_shepherd  0.596461    True
3           1          Rhodesian_ridgeback  0.408143    True
4           1           miniature_pinscher  0.560311    True
5           1         Bernese_mountain_dog  0.651137    True
6           1                   box_turtle  0.933012   False
7           1                         chow  0.692517    True
8           1                shopping_cart  0.962465   False
```

```
9          1             miniature_poodle  0.201493    True
10         1             golden_retriever  0.775930    True
11         1                Gordon_setter  0.503672    True
12         1                 Walker_hound  0.260857    True
13         1                          pug  0.489814    True
14         1                   bloodhound  0.195217    True
15         1                        Lhasa  0.582330    True
16         1               English_setter  0.298617    True
17         1                          hen  0.965932   False
18         1             desktop_computer  0.086502   False
19         1            Italian_greyhound  0.176053    True
20         1                  Maltese_dog  0.857531    True
21         1             three-toed_sloth  0.914671   False
22         1                           ox  0.416669   False
23         1             golden_retriever  0.858744    True
24         1                     malamute  0.336874    True
25         1                   guinea_pig  0.996496   False
26         1  soft-coated_wheaten_terrier  0.326467    True
27         1                    Chihuahua  0.978108    True
28         1      black-and-tan_coonhound  0.529139    True
29         1                         coho  0.404640   False
...      ...                          ...       ...     ...
2045       1               French_bulldog  0.999201    True
2046       1                  convertible  0.738995   False
2047       1                       kuvasz  0.309706    True
2048       2                    Chihuahua  0.793469    True
2049       1                      Samoyed  0.733942    True
2050       1             Mexican_hairless  0.330741    True
2051       2                     Pembroke  0.809197    True
2052       1                    limousine  0.130432   False
2053       1                       basset  0.821664    True
2054       1               French_bulldog  0.995026    True
2055       2                     Pembroke  0.809197    True
2056       3               Siberian_husky  0.700377    True
2057       1             golden_retriever  0.469760    True
2058       1             golden_retriever  0.714719    True
2059       1                      whippet  0.626152    True
2060       1             golden_retriever  0.953442    True
2061       1               French_bulldog  0.991650    True
2062       1                     Pembroke  0.966327    True
2063       1               French_bulldog  0.377417    True
2064       1                      Samoyed  0.957979    True
2065       1                     Pembroke  0.511319    True
2066       1                Irish_terrier  0.487574    True
2067       2                   Pomeranian  0.566142    True
2068       1                  Appenzeller  0.341703    True
2069       1     Chesapeake_Bay_retriever  0.425595    True
2070       2                       basset  0.555712    True
```

```
2071        1                  paper_towel  0.170278   False
2072        1                    Chihuahua  0.716012    True
2073        1                    Chihuahua  0.323581    True
2074        1                       orange  0.097049   False

                            p2  p2_conf  p2_dog                           p3  \
0                       collie  0.156665    True            Shetland_sheepdog
1           miniature_pinscher  0.074192    True          Rhodesian_ridgeback
2                     malinois  0.138584    True                   bloodhound
3                      redbone  0.360687    True           miniature_pinscher
4                   Rottweiler  0.243682    True                     Doberman
5             English_springer  0.263788    True   Greater_Swiss_Mountain_dog
6                   mud_turtle  0.045885   False                     terrapin
7              Tibetan_mastiff  0.058279    True                     fur_coat
8              shopping_basket  0.014594   False             golden_retriever
9                     komondor  0.192305    True  soft-coated_wheaten_terrier
10             Tibetan_mastiff  0.093718    True           Labrador_retriever
11           Yorkshire_terrier  0.174201    True                     Pekinese
12            English_foxhound  0.175382    True                 Ibizan_hound
13                bull_mastiff  0.404722    True               French_bulldog
14             German_shepherd  0.078260    True                     malinois
15                    Shih-Tzu  0.166192    True               Dandie_Dinmont
16                Newfoundland  0.149842    True                       borzoi
17                        cock  0.033919   False                    partridge
18                        desk  0.085547   False                     bookcase
19                 toy_terrier  0.111884    True                      basenji
20                  toy_poodle  0.063064    True             miniature_poodle
21                       otter  0.015250   False               great_grey_owl
22                Newfoundland  0.278407    True                  groenendael
23    Chesapeake_Bay_retriever  0.054787    True           Labrador_retriever
24              Siberian_husky  0.147655    True                   Eskimo_dog
25                       skunk  0.002402   False                      hamster
26                Afghan_hound  0.259551    True                       briard
27                 toy_terrier  0.009397    True                     papillon
28                  bloodhound  0.244220    True        flat-coated_retriever
29                  barracouta  0.271485   False                          gar
...                        ...       ...     ...                          ...
2045                 Chihuahua  0.000361    True                  Boston_bull
2046                sports_car  0.139952   False                    car_wheel
2047            Great_Pyrenees  0.186136    True               Dandie_Dinmont
2048               toy_terrier  0.143528    True                   can_opener
2049                Eskimo_dog  0.035029    True    Staffordshire_bullterrier
2050                  sea_lion  0.275645   False                   Weimaraner
2051       Rhodesian_ridgeback  0.054950    True                       beagle
2052                 tow_truck  0.029175   False                shopping_cart
2053                   redbone  0.087582    True                   Weimaraner
2054                       pug  0.000932    True                 bull_mastiff
2055       Rhodesian_ridgeback  0.054950    True                       beagle
```

```
2056           Eskimo_dog  0.166511    True                     malamute
2057   Labrador_retriever  0.184172    True               English_setter
2058      Tibetan_mastiff  0.120184    True           Labrador_retriever
2059               borzoi  0.194742    True                       Saluki
2060   Labrador_retriever  0.013834    True                      redbone
2061                boxer  0.002129    True   Staffordshire_bullterrier
2062             Cardigan  0.027356    True                      basenji
2063   Labrador_retriever  0.151317    True                       muzzle
2064           Pomeranian  0.013884    True                         chow
2065             Cardigan  0.451038    True                    Chihuahua
2066         Irish_setter  0.193054    True    Chesapeake_Bay_retriever
2067           Eskimo_dog  0.178406    True                     Pembroke
2068        Border_collie  0.199287    True                    ice_lolly
2069        Irish_terrier  0.116317    True               Indian_elephant
2070     English_springer  0.225770    True  German_short-haired_pointer
2071   Labrador_retriever  0.168086    True                      spatula
2072             malamute  0.078253    True                       kelpie
2073             Pekinese  0.090647    True                     papillon
2074                bagel  0.085851   False                       banana

       p3_conf  p3_dog
0     0.061428    True
1     0.072010    True
2     0.116197    True
3     0.222752    True
4     0.154629    True
5     0.016199    True
6     0.017885   False
7     0.054449   False
8     0.007959    True
9     0.082086    True
10    0.072427    True
11    0.109454    True
12    0.097471    True
13    0.048960    True
14    0.075628    True
15    0.089688    True
16    0.133649    True
17    0.000052   False
18    0.079480   False
19    0.111152    True
20    0.025581    True
21    0.013207   False
22    0.102643    True
23    0.014241    True
24    0.093412    True
25    0.000461   False
26    0.206803    True
```

```
27    0.004577    True
28    0.173810    True
29    0.189945   False
...        ...     ...
2045  0.000076    True
2046  0.044173   False
2047  0.086346    True
2048  0.032253   False
2049  0.029705    True
2050  0.134203    True
2051  0.038915    True
2052  0.026321   False
2053  0.026236    True
2054  0.000903    True
2055  0.038915    True
2056  0.111411    True
2057  0.073482    True
2058  0.105506    True
2059  0.027351    True
2060  0.007958    True
2061  0.001498    True
2062  0.004633    True
2063  0.082981   False
2064  0.008167    True
2065  0.029248    True
2066  0.118184    True
2067  0.076507    True
2068  0.193548   False
2069  0.076902   False
2070  0.175219    True
2071  0.040836   False
2072  0.031379    True
2073  0.068957    True
2074  0.076110   False

[2075 rows x 12 columns]
```

```
In [13]: df_pred.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
```

```
p2           2075 non-null object
p2_conf      2075 non-null float64
p2_dog       2075 non-null bool
p3           2075 non-null object
p3_conf      2075 non-null float64
p3_dog       2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

```
In [14]: #
         df_pred['tweet_id'][(df_pred.p1_dog==False) & (df_pred.p2_dog==False) & (df_pred.p3_dog
```

```
Out[14]: 6       666051853826850816
         17      666104133288665088
         18      666268910803644416
         21      666293911632134144
         25      666362758909284353
         29      666411507551481857
         45      666786068205871104
         50      666837028449972224
         51      666983947667116034
         53      667012601033924608
         56      667065535570550784
         69      667188689915760640
         73      667369227918143488
         77      667437278097252352
         78      667443425659232256
         93      667549055577362432
         94      667550882905632768
         96      667724302356258817
         98      667766675769573376
         100     667782464991965184
         106     667866724293877760
         107     667873844930215936
         112     667911425562669056
         115     667937095915278337
         117     668142349051129856
         118     668154635664932864
         123     668226093875376128
         130     668291999406125056
         132     668466899341221888
         140     668544745690562560
                        ...
         1839    837482249356513284
         1844    838916489579200512
         1847    839290600511926273
         1851    840370681858686976
```

```
1853    840696689258311684
1869    844580511645339650
1886    847962785489326080
1887    847971574464610304
1891    849051919805034497
1892    849336543269576704
1900    851464819735769094
1902    851861385021730816
1905    852226086759018497
1906    852311364735569921
1910    853299958564483072
1931    859074603037188101
1936    860184849394610176
1937    860276583193509888
1940    860924035999428608
1946    862457590147678208
1953    863907417377173506
1956    864873206498414592
1975    870063196459192321
1979    870804317367881728
2012    879050749262655488
2021    880935762899988482
2022    881268444196462592
2046    886680336477933568
2052    887517139158093824
2074    892420643555336193
Name: tweet_id, Length: 324, dtype: int64
```

- **tweet_id img_numint64Q**

4.3 df_json

```
In [15]: df_json
```

```
Out[15]:                tweet_id  retweet_count  favorite_count
         0     892420643555336193           8842           39492
         1     892177421306343426           6480           33786
         2     891815181378084864           4301           25445
         3     891689557279858688           8925           42863
         4     891327558926688256           9721           41016
         5     891087950875897856           3240           20548
         6     890971913173991426           2142           12053
         7     890729181411237888          19548           66596
         8     890609185150312448           4403           28187
         9     890240255349198849           7684           32467
         10    890006608113172480           7584           31127
         11    889880896479866881           5116           28208
         12    889665388333682689           8502           38745
```

13	889638837579907072	4705	27633
14	889531135344209921	2309	15329
15	889278841981685760	5635	25712
16	888917238123831296	4681	29555
17	888804989199671297	4535	26021
18	888554962724278272	3722	20267
19	888078434458587136	3637	22144
20	887705289381826560	5584	30690
21	887517139158093824	12053	46940
22	887473957103951883	18813	70007
23	887343217045368832	10713	34223
24	887101392804085760	6147	31045
25	886983233522544640	8045	35786
26	886736880519319552	3420	12286
27	886680336477933568	4597	22802
28	886366144734445568	3297	21488
29	886267009285017600	4	117
...	...	...	...
2322	666411507551481857	337	457
2323	666407126856765440	43	113
2324	666396247373291520	91	171
2325	666373753744588802	99	194
2326	666362758909284353	590	801
2327	666353288456101888	76	228
2328	666345417576210432	146	308
2329	666337882303524864	96	203
2330	666293911632134144	365	519
2331	666287406224695296	71	152
2332	666273097616637952	81	183
2333	666268910803644416	37	108
2334	666104133288665088	6835	14703
2335	666102155909144576	15	81
2336	666099513787052032	73	160
2337	666094000022159362	78	168
2338	666082916733198337	47	121
2339	666073100786774016	173	334
2340	666071193221509120	67	154
2341	666063827256086533	230	494
2342	666058600524156928	61	117
2343	666057090499244032	146	304
2344	666055525042405380	261	449
2345	666051853826850816	877	1250
2346	666050758794694657	60	136
2347	666049248165822465	41	111
2348	666044226329800704	147	309
2349	666033412701032449	47	128
2350	666029285002620928	48	132
2351	666020888022790149	530	2528

```
        [2352 rows x 3 columns]
```

```
In [16]: df_json.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2352 entries, 0 to 2351
Data columns (total 3 columns):
tweet_id          2352 non-null object
retweet_count     2352 non-null object
favorite_count    2352 non-null object
dtypes: object(3)
memory usage: 55.2+ KB
```

- **int64**Q

4.4

df_twitter

- Done C1_retweeted_count favorite_count df_json
- Done C2_df_pred

df_twitter

- Done Q1_name
- Done Q2_tweet_idint64in_reply_to_status_id & in_reply_to_user_id 64ID
- Done Q3_timestampobjecttimedate
- Done Q4_
- Done Q5_Rating
- Done Q6_np.nan'None'
- Done Q7_***update_20180403***

df_pred

- Done Q7_tweet_id img_numint64

df_json

- Done Q8_retweet_count&favorite_countint64

df_twitter

- Done T1_source,expanded_urls,in_reply_to_status_id,in_reply_to_user_id,retweeted_status_id,ret
- Done T2_

5

5.1 &

5.1.1

C1_retweeted_count favorite_count df_json

5.1.2

```
In [17]: df_twitter_clean = pd.merge(df_twitter, df_json, how='left', on='tweet_id')
```

5.1.3

```
In [18]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2356 entries, 0 to 2355
Data columns (total 19 columns):
tweet_id                      2356 non-null object
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2356 non-null object
source                        2356 non-null object
text                          2356 non-null object
retweeted_status_id           181 non-null float64
retweeted_status_user_id      181 non-null float64
retweeted_status_timestamp    181 non-null object
expanded_urls                 2297 non-null object
rating_numerator              2356 non-null int64
rating_denominator            2356 non-null int64
name                          2356 non-null object
doggo                         2356 non-null object
floofer                       2356 non-null object
pupper                        2356 non-null object
puppo                         2356 non-null object
retweet_count                 2352 non-null object
favorite_count                2352 non-null object
dtypes: float64(4), int64(2), object(13)
memory usage: 368.1+ KB
```

5.1.4

- Done Q7_***update_20180403***

5.1.5

```
In [19]: df_twitter_clean = df_twitter_clean.drop(df_twitter_clean[df_twitter_clean['retweeted_s
```

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 19 columns):
tweet_id                      2175 non-null object
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2175 non-null object
source                        2175 non-null object
text                          2175 non-null object
retweeted_status_id           0 non-null float64
retweeted_status_user_id      0 non-null float64
retweeted_status_timestamp    0 non-null object
expanded_urls                 2117 non-null object
rating_numerator              2175 non-null int64
rating_denominator            2175 non-null int64
name                          2175 non-null object
doggo                         2175 non-null object
floofer                       2175 non-null object
pupper                        2175 non-null object
puppo                         2175 non-null object
retweet_count                 2175 non-null object
favorite_count                2175 non-null object
dtypes: float64(4), int64(2), object(13)
memory usage: 339.8+ KB
```

5.1.6

T1_source,expanded_urls,in_reply_to_status_id,in_reply_to_user_id,retweeted_status_id,retweeted_sta

5.1.7

```
In [21]: df_twitter_clean.drop(['source','expanded_urls','in_reply_to_status_id', 'in_reply_to_u
```

5.1.8

```
In [22]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2175 entries, 0 to 2355
Data columns (total 12 columns):
tweet_id              2175 non-null object
timestamp             2175 non-null object
```

```
text                    2175 non-null object
rating_numerator        2175 non-null int64
rating_denominator      2175 non-null int64
name                    2175 non-null object
doggo                   2175 non-null object
floofer                 2175 non-null object
pupper                  2175 non-null object
puppo                   2175 non-null object
retweet_count           2175 non-null object
favorite_count          2175 non-null object
dtypes: int64(2), object(10)
memory usage: 220.9+ KB
```

5.1.9

C2_df_pred

Q7_***update_20180403***

5.1.10

```
In [23]: df_pred_clean = df_pred.copy()
         df_pred_clean.drop(['jpg_url','img_num'],axis = 1, inplace = True)
         df_twitter_clean = pd.merge(df_twitter_clean, df_pred_clean)#20180403merge
```

5.1.11

```
In [24]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 21 columns):
tweet_id                1994 non-null object
timestamp               1994 non-null object
text                    1994 non-null object
rating_numerator        1994 non-null int64
rating_denominator      1994 non-null int64
name                    1994 non-null object
doggo                   1994 non-null object
floofer                 1994 non-null object
pupper                  1994 non-null object
puppo                   1994 non-null object
retweet_count           1994 non-null object
favorite_count          1994 non-null object
p1                      1994 non-null object
p1_conf                 1994 non-null float64
p1_dog                  1994 non-null bool
p2                      1994 non-null object
```

```
p2_conf                 1994 non-null float64
p2_dog                  1994 non-null bool
p3                      1994 non-null object
p3_conf                 1994 non-null float64
p3_dog                  1994 non-null bool
dtypes: bool(3), float64(3), int64(2), object(13)
memory usage: 301.8+ KB
```

5.1.12

T2_

Q4_

Update_20180403_text

5.1.13

```
In [25]: #text
         df_twitter_clean['type'] = 'None'
         dog_lists = ['pupper', 'puppo', 'doggo', 'floofer']
         for i in range(0,len(df_twitter_clean)):
             text = df_twitter_clean.text[i].lower()#20180403text
             for dog_status in dog_lists:
                 if dog_status in text:
                     df_twitter_clean.type[i] = dog_status
         df_twitter_clean
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
```

```
Out[25]:                 tweet_id                  timestamp  \
         0     892420643555336193  2017-08-01 16:23:56 +0000
         1     892177421306343426  2017-08-01 00:17:27 +0000
         2     891815181378084864  2017-07-31 00:18:03 +0000
         3     891689557279858688  2017-07-30 15:58:51 +0000
         4     891327558926688256  2017-07-29 16:00:24 +0000
         5     891087950875897856  2017-07-29 00:08:17 +0000
         6     890971913173991426  2017-07-28 16:27:12 +0000
         7     890729181411237888  2017-07-28 00:22:40 +0000
         8     890609185150312448  2017-07-27 16:25:51 +0000
         9     890240255349198849  2017-07-26 15:59:51 +0000
         10    890006608113172480  2017-07-26 00:31:25 +0000
         11    889880896479866881  2017-07-25 16:11:53 +0000
         12    889665388333682689  2017-07-25 01:55:32 +0000
         13    889638837579907072  2017-07-25 00:10:02 +0000
```

```
14      889531135344209921  2017-07-24 17:02:04 +0000
15      889278841981685760  2017-07-24 00:19:32 +0000
16      888917238123831296  2017-07-23 00:22:39 +0000
17      888804989199671297  2017-07-22 16:56:37 +0000
18      888554962724278272  2017-07-22 00:23:06 +0000
19      888078434458587136  2017-07-20 16:49:33 +0000
20      887705289381826560  2017-07-19 16:06:48 +0000
21      887517139158093824  2017-07-19 03:39:09 +0000
22      887473957103951883  2017-07-19 00:47:34 +0000
23      887343217045368832  2017-07-18 16:08:03 +0000
24      887101392804085760  2017-07-18 00:07:08 +0000
25      886983233522544640  2017-07-17 16:17:36 +0000
26      886736880519319552  2017-07-16 23:58:41 +0000
27      886680336477933568  2017-07-16 20:14:00 +0000
28      886366144734445568  2017-07-15 23:25:31 +0000
29      886258384151887873  2017-07-15 16:17:19 +0000
...                    ...                        ...
1964    666411507551481857  2015-11-17 00:24:19 +0000
1965    666407126856765440  2015-11-17 00:06:54 +0000
1966    666396247373291520  2015-11-16 23:23:41 +0000
1967    666373753744588802  2015-11-16 21:54:18 +0000
1968    666362758909284353  2015-11-16 21:10:36 +0000
1969    666353288456101888  2015-11-16 20:32:58 +0000
1970    666345417576210432  2015-11-16 20:01:42 +0000
1971    666337882303524864  2015-11-16 19:31:45 +0000
1972    666293911632134144  2015-11-16 16:37:02 +0000
1973    666287406224695296  2015-11-16 16:11:11 +0000
1974    666273097616637952  2015-11-16 15:14:19 +0000
1975    666268910803644416  2015-11-16 14:57:41 +0000
1976    666104133288665088  2015-11-16 04:02:55 +0000
1977    666102155909144576  2015-11-16 03:55:04 +0000
1978    666099513787052032  2015-11-16 03:44:34 +0000
1979    666094000022159362  2015-11-16 03:22:39 +0000
1980    666082916733198337  2015-11-16 02:38:37 +0000
1981    666073100786774016  2015-11-16 01:59:36 +0000
1982    666071193221509120  2015-11-16 01:52:02 +0000
1983    666063827256086533  2015-11-16 01:22:45 +0000
1984    666058600524156928  2015-11-16 01:01:59 +0000
1985    666057090499244032  2015-11-16 00:55:59 +0000
1986    666055525042405380  2015-11-16 00:49:46 +0000
1987    666051853826850816  2015-11-16 00:35:11 +0000
1988    666050758794694657  2015-11-16 00:30:50 +0000
1989    666049248165822465  2015-11-16 00:24:50 +0000
1990    666044226329800704  2015-11-16 00:04:52 +0000
1991    666033412701032449  2015-11-15 23:21:54 +0000
1992    666029285002620928  2015-11-15 23:05:30 +0000
1993    666020888022790149  2015-11-15 22:32:08 +0000
```

```
                                                   text  rating_numerator  \
0     This is Phineas. He's a mystical boy. Only eve...                13
1     This is Tilly. She's just checking pup on you...                13
2     This is Archie. He is a rare Norwegian Pouncin...                12
3     This is Darla. She commenced a snooze mid meal...                13
4     This is Franklin. He would like you to stop ca...                12
5     Here we have a majestic great white breaching ...                13
6     Meet Jax. He enjoys ice cream so much he gets ...                13
7     When you watch your owner call another dog a g...                13
8     This is Zoey. She doesn't want to be one of th...                13
9     This is Cassie. She is a college pup. Studying...                14
10    This is Koda. He is a South Australian decksha...                13
11    This is Bruno. He is a service shark. Only get...                13
12    Here's a puppo that seems to be on the fence a...                13
13    This is Ted. He does his best. Sometimes that'...                12
14    This is Stuart. He's sporting his favorite fan...                13
15    This is Oliver. You're witnessing one of his m...                13
16    This is Jim. He found a fren. Taught him how t...                12
17    This is Zeke. He has a new stick. Very proud o...                13
18    This is Ralphus. He's powering up. Attempting ...                13
19    This is Gerald. He was just told he didn't get...                12
20    This is Jeffrey. He has a monopoly on the pool...                13
21    I've yet to rate a Venezuelan Hover Wiener. Th...                14
22    This is Canela. She attempted some fancy porch...                13
23    You may not have known you needed to see this ...                13
24    This... is a Jubilant Antarctic House Bear. We...                12
25    This is Maya. She's very shy. Rarely leaves he...                13
26    This is Mingus. He's a wonderful father to his...                13
27    This is Derek. He's late for a dog meeting. 13...                13
28    This is Roscoe. Another pupper fallen victim t...                12
29    This is Waffles. His doggles are pupside down...                13
...                                                 ...               ...
1964  This is quite the dog. Gets really excited whe...                 2
1965  This is a southern Vesuvius bumblegruff. Can d...                 7
1966  Oh goodness. A super rare northeast Qdoba kang...                 9
1967  Those are sunglasses and a jean jacket. 11/10 ...                11
1968  Unique dog here. Very small. Lives in containe...                 6
1969  Here we have a mixed Asiago from the Galápagos...                 8
1970  Look at this jokester thinking seat belt laws ...                10
1971  This is an extremely rare horned Parthenon. No...                 9
1972  This is a funny dog. Weird toes. Won't come do...                 3
1973  This is an Albanian 3 1/2 legged  Episcopalian...                 1
1974     Can take selfies 11/10 https://t.co/ws2AMaNwPW                11
1975  Very concerned about fellow dog trapped in com...                10
1976  Not familiar with this breed. No tail (weird)...                 1
1977  Oh my. Here you are seeing an Adobe Setter giv...                11
1978  Can stand on stump for what seems like a while...                 8
1979  This appears to be a Mongolian Presbyterian mi...                 9
```

```
1980  Here we have a well-established sunblockerspan...                 6
1981  Let's hope this flight isn't Malaysian (lol). ...                10
1982  Here we have a northern speckled Rhododendron...                 9
1983  This is the happiest dog you will ever see. Ve...                10
1984  Here is the Rand Paul of retrievers folks! He'...                 8
1985  My oh my. This is a rare blond Canadian terrie...                 9
1986  Here is a Siberian heavily armored polar bear ...                10
1987  This is an odd dog. Hard on the outside but lo...                 2
1988  This is a truly beautiful English Wilson Staff...                10
1989  Here we have a 1949 1st generation vulpix. Enj...                 5
1990  This is a purebred Piers Morgan. Loves to Netf...                 6
1991  Here is a very happy pup. Big fan of well-main...                 9
1992  This is a western brown Mitsubishi terrier. Up...                 7
1993  Here we have a Japanese Irish Setter. Lost eye...                 8

      rating_denominator      name  doggo floofer  pupper  puppo   ...    \
0                     10   Phineas   None    None    None   None   ...
1                     10     Tilly   None    None    None   None   ...
2                     10    Archie   None    None    None   None   ...
3                     10     Darla   None    None    None   None   ...
4                     10  Franklin   None    None    None   None   ...
5                     10      None   None    None    None   None   ...
6                     10       Jax   None    None    None   None   ...
7                     10      None   None    None    None   None   ...
8                     10      Zoey   None    None    None   None   ...
9                     10    Cassie  doggo    None    None   None   ...
10                    10      Koda   None    None    None   None   ...
11                    10     Bruno   None    None    None   None   ...
12                    10      None   None    None    None  puppo   ...
13                    10       Ted   None    None    None   None   ...
14                    10    Stuart   None    None    None  puppo   ...
15                    10    Oliver   None    None    None   None   ...
16                    10       Jim   None    None    None   None   ...
17                    10      Zeke   None    None    None   None   ...
18                    10   Ralphus   None    None    None   None   ...
19                    10    Gerald   None    None    None   None   ...
20                    10   Jeffrey   None    None    None   None   ...
21                    10      such   None    None    None   None   ...
22                    10    Canela   None    None    None   None   ...
23                    10      None   None    None    None   None   ...
24                    10      None   None    None    None   None   ...
25                    10      Maya   None    None    None   None   ...
26                    10    Mingus   None    None    None   None   ...
27                    10     Derek   None    None    None   None   ...
28                    10    Roscoe   None    None  pupper   None   ...
29                    10   Waffles   None    None    None   None   ...
...                  ...       ...    ...     ...     ...    ...   ...
1964                  10     quite   None    None    None   None   ...
```

```
1965                 10         a   None    None    None   None   ...
1966                 10      None   None    None    None   None   ...
1967                 10      None   None    None    None   None   ...
1968                 10      None   None    None    None   None   ...
1969                 10      None   None    None    None   None   ...
1970                 10      None   None    None    None   None   ...
1971                 10        an   None    None    None   None   ...
1972                 10         a   None    None    None   None   ...
1973                  2        an   None    None    None   None   ...
1974                 10      None   None    None    None   None   ...
1975                 10      None   None    None    None   None   ...
1976                 10      None   None    None    None   None   ...
1977                 10      None   None    None    None   None   ...
1978                 10      None   None    None    None   None   ...
1979                 10      None   None    None    None   None   ...
1980                 10      None   None    None    None   None   ...
1981                 10      None   None    None    None   None   ...
1982                 10      None   None    None    None   None   ...
1983                 10       the   None    None    None   None   ...
1984                 10       the   None    None    None   None   ...
1985                 10         a   None    None    None   None   ...
1986                 10         a   None    None    None   None   ...
1987                 10        an   None    None    None   None   ...
1988                 10         a   None    None    None   None   ...
1989                 10      None   None    None    None   None   ...
1990                 10         a   None    None    None   None   ...
1991                 10         a   None    None    None   None   ...
1992                 10         a   None    None    None   None   ...
1993                 10      None   None    None    None   None   ...

                             p1   p1_conf p1_dog                   p2  \
0                        orange  0.097049  False                bagel
1                     Chihuahua  0.323581   True             Pekinese
2                     Chihuahua  0.716012   True             malamute
3                   paper_towel  0.170278  False   Labrador_retriever
4                        basset  0.555712   True     English_springer
5      Chesapeake_Bay_retriever  0.425595   True        Irish_terrier
6                   Appenzeller  0.341703   True        Border_collie
7                    Pomeranian  0.566142   True           Eskimo_dog
8                 Irish_terrier  0.487574   True         Irish_setter
9                      Pembroke  0.511319   True             Cardigan
10                      Samoyed  0.957979   True           Pomeranian
11               French_bulldog  0.377417   True   Labrador_retriever
12                     Pembroke  0.966327   True             Cardigan
13               French_bulldog  0.991650   True                boxer
14             golden_retriever  0.953442   True   Labrador_retriever
15                      whippet  0.626152   True               borzoi
16             golden_retriever  0.714719   True       Tibetan_mastiff
```

```
17               golden_retriever  0.469760   True        Labrador_retriever
18                 Siberian_husky  0.700377   True                Eskimo_dog
19                 French_bulldog  0.995026   True                       pug
20                         basset  0.821664   True                   redbone
21                      limousine  0.130432  False                 tow_truck
22                       Pembroke  0.809197   True       Rhodesian_ridgeback
23               Mexican_hairless  0.330741   True                  sea_lion
24                        Samoyed  0.733942   True                Eskimo_dog
25                      Chihuahua  0.793469   True               toy_terrier
26                         kuvasz  0.309706   True            Great_Pyrenees
27                    convertible  0.738995  False                sports_car
28                 French_bulldog  0.999201   True                 Chihuahua
29                            pug  0.943575   True                shower_cap
...                           ...       ...    ...                       ...
1964                         coho  0.404640  False                barracouta
1965      black-and-tan_coonhound  0.529139   True                bloodhound
1966                    Chihuahua  0.978108   True               toy_terrier
1967  soft-coated_wheaten_terrier  0.326467   True              Afghan_hound
1968                   guinea_pig  0.996496  False                     skunk
1969                     malamute  0.336874   True            Siberian_husky
1970             golden_retriever  0.858744   True  Chesapeake_Bay_retriever
1971                           ox  0.416669  False              Newfoundland
1972             three-toed_sloth  0.914671  False                     otter
1973                  Maltese_dog  0.857531   True                toy_poodle
1974            Italian_greyhound  0.176053   True               toy_terrier
1975             desktop_computer  0.086502  False                      desk
1976                          hen  0.965932  False                      cock
1977               English_setter  0.298617   True              Newfoundland
1978                        Lhasa  0.582330   True                  Shih-Tzu
1979                   bloodhound  0.195217   True           German_shepherd
1980                          pug  0.489814   True              bull_mastiff
1981                 Walker_hound  0.260857   True          English_foxhound
1982                Gordon_setter  0.503672   True         Yorkshire_terrier
1983             golden_retriever  0.775930   True           Tibetan_mastiff
1984             miniature_poodle  0.201493   True                  komondor
1985                shopping_cart  0.962465  False           shopping_basket
1986                         chow  0.692517   True           Tibetan_mastiff
1987                   box_turtle  0.933012  False                mud_turtle
1988         Bernese_mountain_dog  0.651137   True          English_springer
1989           miniature_pinscher  0.560311   True                Rottweiler
1990          Rhodesian_ridgeback  0.408143   True                   redbone
1991              German_shepherd  0.596461   True                  malinois
1992                      redbone  0.506826   True        miniature_pinscher
1993       Welsh_springer_spaniel  0.465074   True                    collie

       p2_conf p2_dog                           p3   p3_conf p3_dog    type
0     0.085851  False                       banana  0.076110  False    None
1     0.090647   True                     papillon  0.068957   True    None
```

```
2     0.078253   True                        kelpie  0.031379   True    None
3     0.168086   True                       spatula  0.040836  False    None
4     0.225770   True  German_short-haired_pointer  0.175219   True    None
5     0.116317   True                Indian_elephant  0.076902  False    None
6     0.199287   True                     ice_lolly  0.193548  False    None
7     0.178406   True                      Pembroke  0.076507   True    None
8     0.193054   True      Chesapeake_Bay_retriever  0.118184   True    None
9     0.451038   True                     Chihuahua  0.029248   True   doggo
10    0.013884   True                          chow  0.008167   True    None
11    0.151317   True                        muzzle  0.082981  False    None
12    0.027356   True                       basenji  0.004633   True   puppo
13    0.002129   True    Staffordshire_bullterrier  0.001498   True    None
14    0.013834   True                       redbone  0.007958   True   puppo
15    0.194742   True                        Saluki  0.027351   True    None
16    0.120184   True            Labrador_retriever  0.105506   True    None
17    0.184172   True                English_setter  0.073482   True    None
18    0.166511   True                      malamute  0.111411   True    None
19    0.000932   True                   bull_mastiff  0.000903   True    None
20    0.087582   True                    Weimaraner  0.026236   True    None
21    0.029175  False                 shopping_cart  0.026321  False    None
22    0.054950   True                        beagle  0.038915   True    None
23    0.275645  False                    Weimaraner  0.134203   True    None
24    0.035029   True    Staffordshire_bullterrier  0.029705   True    None
25    0.143528   True                    can_opener  0.032253  False    None
26    0.186136   True                Dandie_Dinmont  0.086346   True    None
27    0.139952  False                     car_wheel  0.044173  False    None
28    0.000361   True                   Boston_bull  0.000076   True  pupper
29    0.025286  False                   Siamese_cat  0.002849  False    None
...        ...    ...                           ...       ...    ...     ...
1964  0.271485  False                           gar  0.189945  False    None
1965  0.244220   True         flat-coated_retriever  0.173810   True    None
1966  0.009397   True                      papillon  0.004577   True    None
1967  0.259551   True                        briard  0.206803   True    None
1968  0.002402  False                       hamster  0.000461  False    None
1969  0.147655   True                    Eskimo_dog  0.093412   True    None
1970  0.054787   True            Labrador_retriever  0.014241   True    None
1971  0.278407   True                    groenendael  0.102643   True    None
1972  0.015250  False                 great_grey_owl  0.013207  False    None
1973  0.063064   True               miniature_poodle  0.025581   True    None
1974  0.111884   True                       basenji  0.111152   True    None
1975  0.085547  False                      bookcase  0.079480  False    None
1976  0.033919  False                     partridge  0.000052  False    None
1977  0.149842   True                        borzoi  0.133649   True    None
1978  0.166192   True                Dandie_Dinmont  0.089688   True    None
1979  0.078260   True                      malinois  0.075628   True    None
1980  0.404722   True                French_bulldog  0.048960   True    None
1981  0.175382   True                   Ibizan_hound  0.097471   True    None
1982  0.174201   True                      Pekinese  0.109454   True    None
```

```
1983  0.093718   True           Labrador_retriever  0.072427   True    None
1984  0.192305   True  soft-coated_wheaten_terrier  0.082086   True    None
1985  0.014594  False             golden_retriever  0.007959   True    None
1986  0.058279   True                     fur_coat  0.054449  False    None
1987  0.045885  False                     terrapin  0.017885  False    None
1988  0.263788   True   Greater_Swiss_Mountain_dog  0.016199   True    None
1989  0.243682   True                     Doberman  0.154629   True    None
1990  0.360687   True           miniature_pinscher  0.222752   True    None
1991  0.138584   True                   bloodhound  0.116197   True    None
1992  0.074192   True          Rhodesian_ridgeback  0.072010   True    None
1993  0.156665   True            Shetland_sheepdog  0.061428   True    None

[1994 rows x 22 columns]
```

```
In [26]: none_list = df_twitter_clean['tweet_id'][(df_twitter_clean.doggo == 'None') & (df_twitt
         none_dog_list = list(none_list)
         len(none_dog_list)
```

```
Out[26]: 1688
```

```
In [27]: #
         with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[df_twitter_clean['tweet_id'].isin(dupl_dog_list)][['tweet_
```

```
               tweet_id    name  \
148  855851453814013952    None
154  854010172552949760    None
340  817777686764523521    Dido
397  808106460588765185    None
419  802265048156610565    None
425  801115127852503040   Bones
510  785639753186217984   Pinot
652  759793422261743616  Maggie
704  751583847268179968    None
795  741067306818797568    just
841  733109485275860992    None


148       Here's a puppo participating in the #ScienceMarch. Cleverly disguising her own doggo
154  At first I thought this was a shy doggo, but it's actually a Rare Canadian Floofer Owl. Ama
340  This is Dido. She's playing the lead role in "Pupper Stops to Catch Snow Before Resuming Sh
397                      Here we have Burke (pupper) and Dexter (doggo). Pupper wants to be exac
419
425                      This is Bones. He's being haunted by another doggo of roughly the same
510  This is Pinot. He's a sophisticated doggo. You can tell by the hat. Also pointier than your
652                      Meet Maggie &amp; Lila. Maggie is the doggo, Lila is the pupper. They a
704                              Please stop sending it pictures that don't even have a doggo or
795                                                                           This is just downr
841                                                                                            L
```

```
     doggo  floofer  pupper  puppo
148  doggo     None    None  puppo
154  doggo  floofer    None   None
340  doggo     None  pupper   None
397  doggo     None  pupper   None
419  doggo     None  pupper   None
425  doggo     None  pupper   None
510  doggo     None  pupper   None
652  doggo     None  pupper   None
704  doggo     None  pupper   None
795  doggo     None  pupper   None
841  doggo     None  pupper   None
```

```
In [28]: #with pd.option_context('max_colwidth', 200):
             #display(df_twitter_clean[df_twitter_clean['tweet_id'].isin(none_dog_list)][['tweet
```

```
In [29]: #df_twitter_clean[df_twitter_clean['tweet_id'].isin(dupl_dog_list)].index
```

```
In [30]: with pd.option_context('max_colwidth', 200):
             display(df_pred[df_pred['tweet_id'].isin(dupl_dog_list)][['tweet_id','jpg_url','p1_
```

```
                tweet_id  \
1155  733109485275860992
1201  741067306818797568
1292  751583847268179968
1350  759793422261743616
1407  770093767776997377
1446  775898661951791106
1507  785639753186217984
1609  801115127852503040
1616  802265048156610565
1644  808106460588765185
1707  817777686764523521
1913  854010172552949760
1919  855851453814013952


                                                                                      jpg_url  \
1155                                          https://pbs.twimg.com/media/CiyHLocU4AI2pJu.jpg
1201                                          https://pbs.twimg.com/media/CkjMx99UoAM2B1a.jpg
1292                                          https://pbs.twimg.com/media/Cm4phTpWcAAgLsr.jpg
1350                                          https://pbs.twimg.com/media/CotUFZEWcAA2Pku.jpg
1407                                          https://pbs.twimg.com/media/CkjMx99UoAM2B1a.jpg
1446                                          https://pbs.twimg.com/media/CiyHLocU4AI2pJu.jpg
1507                                          https://pbs.twimg.com/media/CucnLmeWAAALOSC.jpg
1609                                          https://pbs.twimg.com/media/Cx4h7zHUsAAqaJd.jpg
1616                                          https://pbs.twimg.com/media/CyI3zXgWEAACQfB.jpg
1644                                          https://pbs.twimg.com/media/Czb4iFRXgAIUMiN.jpg
```

```
1707  https://pbs.twimg.com/ext_tw_video_thumb/817777588030476288/pu/img/KbuLpE4krHF4VdPf.jpg
1913                                          https://pbs.twimg.com/media/C9oNt91WAAAFSLS.jpg
1919                                          https://pbs.twimg.com/media/C-CYWrvWAAU8AXH.jpg

      p1_dog  p2_dog  p3_dog
1155    True    True   False
1201    True    True    True
1292    True   False   False
1350    True    True    True
1407    True    True    True
1446    True    True   False
1507   False   False   False
1609    True    True    True
1616    True    True    True
1644    True    True    True
1707    True    True    True
1913    True    True    True
1919    True    True    True
```

```
In [31]: #textdf_pre
         #
         #855851453814013952854010172552949760817777686764523521801115127852503040doggo
         # 751583847268179968 pupper
         #785639753186217984pupper & doggo
         #

         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 855851453814013952),'doggo'] = 'None
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 854010172552949760),'doggo'] = 'None
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 817777686764523521),'doggo'] = 'None
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 801115127852503040),'doggo'] = 'None
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 751583847268179968),'pupper'] = 'Non
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 785639753186217984),('doggo','pupper

         with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[df_twitter_clean['tweet_id'].isin(dupl_dog_list)][['tweet_
```

```
               tweet_id  doggo  floofer  pupper  puppo
148  855851453814013952   None     None    None  puppo
154  854010172552949760   None  floofer    None   None
340  817777686764523521   None     None  pupper   None
397  808106460588765185  doggo     None  pupper   None
419  802265048156610565  doggo     None  pupper   None
425  801115127852503040   None     None  pupper   None
510  785639753186217984   None     None    None   None
652  759793422261743616  doggo     None  pupper   None
704  751583847268179968  doggo     None    None   None
795  741067306818797568  doggo     None  pupper   None
```

```
841  733109485275860992  doggo     None  pupper   None
```

```
In [32]: df_twitter_clean['dog_status'] = 'None'
         for i in range(0,len(df_twitter_clean)):
             if df_twitter_clean.doggo[i] != 'None':
                 df_twitter_clean.dog_status[i] = 'doggo'
             elif df_twitter_clean.floofer[i] != 'None':
                 df_twitter_clean.dog_status[i] = 'floofer'
             elif df_twitter_clean.pupper[i] != 'None':
                 df_twitter_clean.dog_status[i] = 'pupper'
             elif df_twitter_clean.puppo[i] != 'None':
                 df_twitter_clean.dog_status[i] = 'puppo'
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  after removing the cwd from sys.path.
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  # Remove the CWD from sys.path while we load stuff.
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
```

```
In [33]: #df_twitter_clean
```

```
In [34]: df_twitter_clean.type = df_twitter_clean.type.replace('None','')
         df_twitter_clean.dog_status = df_twitter_clean.dog_status.replace('None','')
```

```
In [35]: df_twitter_clean.type.value_counts()
```

```
Out[35]:            1652
         pupper      228
         doggo        79
         puppo        27
         floofer       8
         Name: type, dtype: int64
```

```
In [36]: df_twitter_clean.dog_status.value_counts()
         #df_twitter_clean.dog_status
```

```
Out[36]:            1689
         pupper      205
         doggo        69
         puppo        23
         floofer       8
         Name: dog_status, dtype: int64
```

```
In [37]: # typedog_statusstatus
         df_twitter_clean['status'] = ''
         for i in range(0,len(df_twitter_clean)):
             if df_twitter_clean.type[i] != '':
                 if df_twitter_clean.dog_status[i] != '':
                     if df_twitter_clean.dog_status[i] == df_twitter_clean.type[i]:
                         df_twitter_clean.status[i] = df_twitter_clean.type[i]
                     else: df_twitter_clean.status[i] = 'needcheck'
             else: df_twitter_clean.status[i] = df_twitter_clean.type[i] + df_twitter_clean.dog_
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  if __name__ == '__main__':
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  import sys
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
```

```
In [38]: #dog_statustexttypedog_status
         with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[(df_twitter_clean['status']=='needcheck')][['tweet_id','st
```

```
               tweet_id     status   type dog_status  \
148  855851453814013952  needcheck  doggo      puppo
340  817777686764523521  needcheck  doggo     pupper
425  801115127852503040  needcheck  doggo     pupper


148       Here's a puppo participating in the #ScienceMarch. Cleverly disguising her own doggo
340  This is Dido. She's playing the lead role in "Pupper Stops to Catch Snow Before Resuming Sh
```

```
425                      This is Bones. He's being haunted by another doggo of roughly the same
```

```
In [39]: #dog_statusstatus
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 855851453814013952),'status'] = 'pup
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 817777686764523521),'status'] = 'pup
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 801115127852503040),'status'] = 'pup
```

```
In [40]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 24 columns):
tweet_id              1994 non-null object
timestamp             1994 non-null object
text                  1994 non-null object
rating_numerator      1994 non-null int64
rating_denominator    1994 non-null int64
name                  1994 non-null object
doggo                 1994 non-null object
floofer               1994 non-null object
pupper                1994 non-null object
puppo                 1994 non-null object
retweet_count         1994 non-null object
favorite_count        1994 non-null object
p1                    1994 non-null object
p1_conf               1994 non-null float64
p1_dog                1994 non-null bool
p2                    1994 non-null object
p2_conf               1994 non-null float64
p2_dog                1994 non-null bool
p3                    1994 non-null object
p3_conf               1994 non-null float64
p3_dog                1994 non-null bool
type                  1994 non-null object
dog_status            1994 non-null object
status                1994 non-null object
dtypes: bool(3), float64(3), int64(2), object(16)
memory usage: 428.6+ KB
```

```
In [41]: #
         df_twitter_clean = df_twitter_clean.drop(['doggo','floofer','pupper','puppo','dog_statu
```

5.1.14

```
In [42]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
```

```
Data columns (total 18 columns):
tweet_id              1994 non-null object
timestamp             1994 non-null object
text                  1994 non-null object
rating_numerator      1994 non-null int64
rating_denominator    1994 non-null int64
name                  1994 non-null object
retweet_count         1994 non-null object
favorite_count        1994 non-null object
p1                    1994 non-null object
p1_conf               1994 non-null float64
p1_dog                1994 non-null bool
p2                    1994 non-null object
p2_conf               1994 non-null float64
p2_dog                1994 non-null bool
p3                    1994 non-null object
p3_conf               1994 non-null float64
p3_dog                1994 non-null bool
status                1994 non-null object
dtypes: bool(3), float64(3), int64(2), object(10)
memory usage: 335.1+ KB
```

5.2

5.2.1

- Q2_tweet_idint64in_reply_to_status_id & in_reply_to_user_id 64ID
- Q3_timestampobjecttimedate
- Q7_tweet_id img_numint64
- Q8_retweet_count&favorite_count

5.2.2

```
In [43]: df_twitter_clean.timestamp=pd.to_datetime(df_twitter_clean.timestamp)
         df_twitter_clean.retweet_count = df_twitter_clean.retweet_count.astype(float)
         df_twitter_clean.favorite_count = df_twitter_clean.favorite_count.astype(float)
```

5.2.3

```
In [44]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 18 columns):
tweet_id              1994 non-null object
timestamp             1994 non-null datetime64[ns]
text                  1994 non-null object
rating_numerator      1994 non-null int64
```

```
rating_denominator    1994 non-null int64
name                  1994 non-null object
retweet_count         1994 non-null float64
favorite_count        1994 non-null float64
p1                    1994 non-null object
p1_conf               1994 non-null float64
p1_dog                1994 non-null bool
p2                    1994 non-null object
p2_conf               1994 non-null float64
p2_dog                1994 non-null bool
p3                    1994 non-null object
p3_conf               1994 non-null float64
p3_dog                1994 non-null bool
status                1994 non-null object
dtypes: bool(3), datetime64[ns](1), float64(5), int64(2), object(7)
memory usage: 335.1+ KB
```

5.2.4

- Q5_Rating,rating

5.2.5

```
In [45]: with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[df_twitter_clean.rating_denominator != 10][['tweet_id','na
```

```
                tweet_id      name  \
323   820690176645140481      None
385   810984652412424192       Sam
662   758467244762497024      None
800   740373189193256964      None
848   731156023742988288      this
891   722974582966214656      None
925   716439118184652801  Bluebert
946   713900603437621249      None
970   710658690886586372      None
988   709198395643068416      None
1054  704054845121142784         a
1130  697463031882764288      None
1302  684225744407494656      None
1303  684222868335505415      None
1328  682962037429899265    Darrel
1435  677716515794329600      None
1494  675853064436391936      None
1973  666287406224695296        an
```

```
323                                                                           The floofs have been
385   Meet Sam. She smiles 24/7 &amp; secretly aspires to be a reindeer. \nKeep Sam smiling by c
662
800                                  After so many requests, this is Bretagne. She was the last su
848                                                      Say hello to this unbelievably well behave
891
925                                                          This is Bluebert. He just saw that bot
946                                                                                  Happy Saturday
970                                                                    Here's a brigade of puppers.
988                           From left to right:\nCletus, Jerome, Alejandro, Burp, &amp; Titson\
1054                                                                                          Her
1130                                                                               Happy Wednesd
1302                                             Two sneaky puppers were not initially seen, mov
1303                               Someone help the girl is being mugged. Several are distractin
1328                                This is Darrel. He just robbed a 7/11 and is in a high speed
1435
1494                                                                Here we have an entire plato
1973                                   This is an Albanian 3 1/2 legged  Episcopalian. Loves wel

      rating_numerator  rating_denominator
323                 84                  70
385                 24                   7
662                165                 150
800                  9                  11
848                204                 170
891                  4                  20
925                 50                  50
946                 99                  90
970                 80                  80
988                 45                  50
1054                60                  50
1130                44                  40
1302               143                 130
1303               121                 110
1328                 7                  11
1435               144                 120
1494                88                  80
1973                 1                   2
```

```
In [46]: #
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 835246439529840640),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 832088576586297345),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 810984652412424192),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 775096608509886464),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 740373189193256964),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 722974582966214656),('rating_numerat
         df_twitter_clean.loc[(df_twitter_clean.tweet_id == 716439118184652801),('rating_numerat
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 682962037429899265),('rating_numerat
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 666287406224695296),('rating_numerat

        #
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 832215909146226688),('rating_numerat
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 786709082849828864),('rating_numerat
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 778027034220126208),('rating_numerat
        df_twitter_clean.loc[(df_twitter_clean.tweet_id == 680494726643068929),('rating_numerat
```

```
In [47]: df_twitter_clean['rating'] = df_twitter_clean['rating_numerator'] / df_twitter_clean['r
         df_twitter_clean.rating.value_counts()
```

```
Out[47]: 1.200      454
         1.000      421
         1.100      403
         1.300      262
         0.900      152
         0.800       95
         0.700       51
         1.400       36
         0.500       33
         0.600       32
         0.300       19
         0.400       15
         0.200        9
         0.100        4
         0.000        2
         1.127        1
         1.126        1
         0.975        1
         42.000       1
         177.600      1
         Name: rating, dtype: int64
```

```
In [48]: with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[(df_twitter_clean.rating>2) | (df_twitter_clean.rating<0.2
```

```
                tweet_id  rating  \
235   835152434251116546     0.0
726   749981277374128128   177.6
756   746906459439529985     0.0
1519  675153376133427200     0.1
1718  670842764863651840    42.0
1735  670783437142401025     0.1
1900  667549055577362432     0.1
1976  666104133288665088     0.1


235                               When you're so blinded by your systematic plagiarism that you f
```

```
726                                                                 This is Atticus. He's quite
756                         PUPDATE: can't see any. Even if I could, I couldn't reach them to pet
1519                     What kind of person sends in a picture without a dog in it? 1/10 just b
1718                                                           After so many requests... here yo
1735      Flamboyant pup here. Probably poisonous. Won't eat kibble. Doesn't bark. Slow af. Pett
1900  Never seen dog like this. Breathes heavy. Tilts head in a pattern. No bark. Shitty at fetc
1976        Not familiar with this breed. No tail (weird). Only 2 legs. Doesn't bark. Surprising

      rating_numerator  rating_denominator
235                0.0                10.0
726             1776.0                10.0
756                0.0                10.0
1519               1.0                10.0
1718             420.0                10.0
1735               1.0                10.0
1900               1.0                10.0
1976               1.0                10.0
```

```
In [49]: df_twitter_clean = df_twitter_clean.drop(['rating_numerator','rating_denominator'],axis
```

5.2.6

```
In [50]: df_twitter_clean.rating.value_counts()
```

```
Out[50]: 1.200      454
         1.000      421
         1.100      403
         1.300      262
         0.900      152
         0.800       95
         0.700       51
         1.400       36
         0.500       33
         0.600       32
         0.300       19
         0.400       15
         0.200        9
         0.100        4
         0.000        2
         1.127        1
         1.126        1
         0.975        1
         42.000       1
         177.600      1
         Name: rating, dtype: int64
```

```
In [51]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 17 columns):
tweet_id          1994 non-null object
timestamp         1994 non-null datetime64[ns]
text              1994 non-null object
name              1994 non-null object
retweet_count     1994 non-null float64
favorite_count    1994 non-null float64
p1                1994 non-null object
p1_conf           1994 non-null float64
p1_dog            1994 non-null bool
p2                1994 non-null object
p2_conf           1994 non-null float64
p2_dog            1994 non-null bool
p3                1994 non-null object
p3_conf           1994 non-null float64
p3_dog            1994 non-null bool
status            1994 non-null object
rating            1993 non-null float64
dtypes: bool(3), datetime64[ns](1), float64(6), object(7)
memory usage: 319.5+ KB
```

5.2.7

- Q1_name

5.2.8

```
In [52]: #
         df_twitter_clean.name.value_counts()
```

```
Out[52]: None        546
         a            55
         Charlie      11
         Oliver       10
         Lucy         10
         Cooper       10
         Penny         9
         Tucker        9
         Winston       8
         Sadie         8
         Lola          7
         Daisy         7
         Toby          7
         the           7
         Jax           6
```

```
Bo            6
Koda          6
Stanley       6
Bella         6
an            6
Rusty         5
Bailey        5
Louis         5
Scout         5
Chester       5
Dave          5
Milo          5
Leo           5
Oscar         5
Buddy         5
           ...
Jarod         1
Spencer       1
Kollin        1
Sailor        1
Shikha        1
Zooey         1
Bauer         1
Arya          1
Wafer         1
Rorie         1
Cheryl        1
Coleman       1
Crimson       1
Berb          1
Mitch         1
Pippin        1
Vince         1
Emanuel       1
Aja           1
Jockson       1
Florence      1
Rumpole       1
Kallie        1
Sully         1
Bobbay        1
Mingus        1
Lucky         1
Nigel         1
Swagger       1
Bradlay       1
Name: name, Length: 936, dtype: int64
```

```
In [53]: with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[(df_twitter_clean.name == 'a') | (df_twitter_clean.name ==
```

```
                tweet_id  \
49    881536004380872706
472   792913359805018113
582   772581559778025472
746   747885874273214464
748   747816857231626240
757   746872823977771008
762   746369468511756288
783   743222593470234624
919   717537687239008257
929   715733265223708672
1045  704859558691414016
1054  704054845121142784
1064  703079050210877440
1065  703041949650034688
1070  702539513671897089
1082  700864154249383937
1185  692187005137076224
1209  690360449368465409
1275  685943807276412928
1398  679530280114372609
1441  677644091929329666
1452  677269281705472000
1469  676613908052996102
1503  675706639471788032
1504  675534494439489536
1527  675109292475830276
1528  675047298674663426
1569  674082852460433408
1586  673715861853720576
1600  673636718965334016
...                  ...
1796  669661792646373376
1804  669564461267722241
1833  668955713004314625
1840  668815180734689280
1846  668636665813057536
1853  668614819948453888
1859  668507509523615744
1863  668466899341221888
1876  668171859951755264
1890  667861340749471744
1896  667773195014021121
1903  667538891197542400
1912  667470559035432960
```

```
1926  667177989038297088
1942  666983947667116034
1949  666781792255496192
1952  666701168228331520
1965  666407126856765440
1971  666337882303524864
1972  666293911632134144
1973  666287406224695296
1983  666063827256086533
1984  666058600524156928
1985  666057090499244032
1986  666055525042405380
1987  666051853826850816
1988  666050758794694657
1990  666044226329800704
1991  666033412701032449
1992  666029285002620928


49    Here is a pupper approaching maximum borkdrive. Zooming at never before seen speeds. 14/10
472                              Here is a perfect example of someone who has their priorities i
582                  Guys this is getting so out of hand. We only rate dogs. This is a Galapagos
746                  This is a mighty rare blue-tailed hammer sherk. Human almost lost a limb tr
748                    Viewer discretion is advised. This is a terrible attack in progress. Not
757                        This is a carrot. We only rate dogs. Please only send in dogs. You al
762                    This is an Iraqi Speed Kangaroo. It is not a dog. Please only send in dog
783                    This is a very rare Great Alaskan Bush Pupper. Hard to stumble upon witho
919                   People please. This is a Deadly Mediterranean Plop T-Rex. We only rate dog
929                     This is a taco. We only rate dogs. Please only send in dogs. Dogs are wh
1045                                         Here is a heartbreaking scene of an incredible pu
1054                                                                 Here is a whole f
1064                  This is a Butternut Cumberfloof. It's not windy they just look like that.
1065                                This is an East African Chalupa Seal. We only rate dogs. Ple
1070                     This is a Wild Tuscan Poofwiggle. Careful not to startle. Rare tongue s
1082                                                   "Pupper is a present to world. Here is
1185                  This is a rare Arctic Wubberfloof. Unamused by the happenings. No longer h
1209                                          Stop sending in lobsters. This is the final warn
1275                 This is the newly formed pupper a capella group. They're just starting out
1398                 Guys this really needs to stop. We've been over this way too many times. Th
1441                                                    This is a dog swinging. I really enjoy
1452                                                           This is the happiest pupper I'v
1469                                                                 This is the saddest/sweet
1503                 This is a Sizzlin Menorah spaniel from Brooklyn named Wylie. Lovable eyes.
1504                                           Seriously guys?! Only send in dogs. I only rate
1527                    C'mon guys. We've been over this. We only rate dogs. This is a cow. Plea
1528                                      This is a fluffy albino Bacardi Columbia mix. Excell
1569                                                             This is a Sagitariot Baklava
1586                       This is a heavily opinionated dog. Loves walls. Nobody knows how the
```

```
1600                  This is a Lofted Aphrodisiac Terrier named Kip. Big fan of bed n breakfast
...
1796                  This is a brave dog. Excellent free climber. Trying to get closer to God.
1804                   This is a Coriander Baton Rouge named Alfredo. Loves to cuddle with small
1833                 This is a Slovakian Helter Skelter Feta named Leroi. Likes to skip on roofs
1840                    This is a wild Toblerone from Papua New Guinea. Mouth always open. Addic
1846                  This is an Irish Rigatoni terrier named Berta. Completely made of rope. No
1853                  Here is a horned dog. Much grace. Can jump over moons (dam!). Paws not sof
1859                   This is a Birmingham Quagmire named Chuk. Loves to relax and watch the ga
1863                      Here is a mother dog caring for her pups. Snazzy red mohawk. Doesn't w
1876                            This is a Trans Siberian Kellogg named Alfonso. Huge ass eyeball
1890                 This is a Shotokon Macadamia mix named Cheryl. Sophisticated af. Looks like
1896                     This is a rare Hungarian Pinot named Jessiga. She is either mid-stroke
1903                                       This is a southwest Coriander named Klint. Hat looks
1912                   This is a northern Wahoo named Kohl. He runs this town. Chases tumbleweed
1926                       This is a Dasani Kingfisher from Maine. His name is Daryl. Daryl does
1942                                     This is a curly Ticonderoga named Pepe. No feet. Loves
1949                                                  This is a purebred Bacardi named Octaviath
1952                  This is a golden Buckminsterfullerene named Johm. Drives trucks. Lumberjac
1965                  This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made fri
1971                  This is an extremely rare horned Parthenon. Not amused. Wears shoes. Overa
1972                   This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses t
1973                     This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished har
1983                                                  This is the happiest dog you will ever see
1984                      Here is the Rand Paul of retrievers folks! He's probably good at poker
1985                                 My oh my. This is a rare blond Canadian terrier on wheels.
1986                 Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I wo
1987                   This is an odd dog. Hard on the outside but loving on the inside. Petting
1988                 This is a truly beautiful English Wilson Staff retriever. Has a nice phone.
1990                    This is a purebred Piers Morgan. Loves to Netflix and chill. Always look
1991                           Here is a very happy pup. Big fan of well-maintained decks. Just
1992                  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 d

     name
49      a
472     a
582     a
746     a
748     a
757     a
762    an
783     a
919     a
929     a
1045    a
1054    a
1064    a
1065   an
```

```
1070    a
1082    a
1185    a
1209  the
1275  the
1398    a
1441    a
1452  the
1469  the
1503    a
1504    a
1527    a
1528    a
1569    a
1586    a
1600    a
...   ...
1796    a
1804    a
1833    a
1840    a
1846   an
1853    a
1859    a
1863    a
1876    a
1890    a
1896    a
1903    a
1912    a
1926    a
1942    a
1949    a
1952    a
1965    a
1971   an
1972    a
1973   an
1983  the
1984  the
1985    a
1986    a
1987   an
1988    a
1990    a
1991    a
1992    a
```

```
[68 rows x 3 columns]
```

```
In [54]: #namednamed https://regex101.com/
         for i in range(0,len(df_twitter_clean)):
             if "named" in df_twitter_clean.text[i]:
                 df_twitter_clean.name[i] = ''.join(re.findall(r"named\s(\w*)", df_twitter_clean
                 print(i)
```

```
1503
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  after removing the cwd from sys.path.
```

```
1600
1679
1710
1760
1769
1772
1790
1804
1809
1833
1846
1859
1868
1876
1890
1896
1903
1908
1912
1942
1949
1952
```

```
In [55]: df_twitter_clean.name.value_counts()
```

```
Out[55]: None        543
         a            36
         Charlie      11
         Lucy         10
```

```
Oliver       10
Cooper       10
Penny         9
Tucker        9
Winston       8
Sadie         8
Lola          7
the           7
Daisy         7
Toby          7
Stanley       6
Bo            6
Jax           6
Bella         6
Koda          6
Oscar         5
Milo          5
an            5
Leo           5
Rusty         5
Buddy         5
Dave          5
Scout         5
Bailey        5
Louis         5
Chester       5
           ...
Shikha        1
Sailor        1
Bauer         1
Zooey         1
Philippe      1
Laela         1
Ginger        1
Wafer         1
Brat          1
Rorie         1
Coleman       1
Crimson       1
Berb          1
Mitch         1
Pippin        1
Vince         1
Emanuel       1
Aja           1
Jockson       1
Rumpole       1
Karma         1
```

```
Kallie        1
Sully         1
Bobbay        1
Mingus        1
Lucky         1
Nigel         1
Swagger       1
Florence      1
Bradlay       1
Name: name, Length: 955, dtype: int64
```

reviwer This is | Meet | name is | Say hello to | named namedThis is | Meet | namenamedname

```
In [56]: with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[(df_twitter_clean.name == 'a') | (df_twitter_clean.name ==
```

```
                tweet_id  \
49    881536004380872706
472   792913359805018113
582   772581559778025472
746   747885874273214464
748   747816857231626240
757   746872823977771008
762   746369468511756288
783   743222593470234624
919   717537687239008257
929   715733265223708672
1045  704859558691414016
1054  704054845121142784
1064  703079050210877440
1065  703041949650034688
1070  702539513671897089
1082  700864154249383937
1185  692187005137076224
1209  690360449368465409
1275  685943807276412928
1398  679530280114372609
1441  677644091929329666
1452  677269281705472000
1469  676613908052996102
1504  675534494439489536
1527  675109292475830276
1528  675047298674663426
1569  674082852460433408
1586  673715861853720576
1639  672604026190569472
1682  671561002136281088
1796  669661792646373376
```

```
1840  668815180734689280
1853  668614819948453888
1863  668466899341221888
1926  667177989038297088
1965  666407126856765440
1971  666337882303524864
1972  666293911632134144
1973  666287406224695296
1983  666063827256086533
1984  666058600524156928
1985  666057090499244032
1986  666055525042405380
1987  666051853826850816
1988  666050758794694657
1990  666044226329800704
1991  666033412701032449
1992  666029285002620928
```

```
49    Here is a pupper approaching maximum borkdrive. Zooming at never before seen speeds. 14/1
472                              Here is a perfect example of someone who has their priorities i
582                  Guys this is getting so out of hand. We only rate dogs. This is a Galapagos
746                  This is a mighty rare blue-tailed hammer sherk. Human almost lost a limb tr
748                    Viewer discretion is advised. This is a terrible attack in progress. Not
757                        This is a carrot. We only rate dogs. Please only send in dogs. You al
762                    This is an Iraqi Speed Kangaroo. It is not a dog. Please only send in dog
783                    This is a very rare Great Alaskan Bush Pupper. Hard to stumble upon witho
919                   People please. This is a Deadly Mediterranean Plop T-Rex. We only rate dog
929                     This is a taco. We only rate dogs. Please only send in dogs. Dogs are wh
1045                                           Here is a heartbreaking scene of an incredible pu
1054                                                                           Here is a whole f
1064                  This is a Butternut Cumberfloof. It's not windy they just look like that.
1065                                This is an East African Chalupa Seal. We only rate dogs. Ple
1070                     This is a Wild Tuscan Poofwiggle. Careful not to startle. Rare tongue s
1082                                                     "Pupper is a present to world. Here is
1185                  This is a rare Arctic Wubberfloof. Unamused by the happenings. No longer h
1209                                            Stop sending in lobsters. This is the final warn
1275                 This is the newly formed pupper a capella group. They're just starting out
1398                 Guys this really needs to stop. We've been over this way too many times. Th
1441                                                      This is a dog swinging. I really enjoy
1452                                                             This is the happiest pupper I'v
1469                                                                   This is the saddest/sweet
1504                                             Seriously guys?! Only send in dogs. I only rate
1527                    C'mon guys. We've been over this. We only rate dogs. This is a cow. Plea
1528                                        This is a fluffy albino Bacardi Columbia mix. Excell
1569                                                               This is a Sagitariot Baklava
1586                       This is a heavily opinionated dog. Loves walls. Nobody knows how the
1639                                                         This is a baby Rand Paul. Curls for
```

```
1682             This is the best thing I've ever seen so spread it like wildfire &amp; maybe we
1796                  This is a brave dog. Excellent free climber. Trying to get closer to God.
1840                    This is a wild Toblerone from Papua New Guinea. Mouth always open. Addic
1853                  Here is a horned dog. Much grace. Can jump over moons (dam!). Paws not sof
1863                      Here is a mother dog caring for her pups. Snazzy red mohawk. Doesn't w
1926                       This is a Dasani Kingfisher from Maine. His name is Daryl. Daryl does
1965                  This is a southern Vesuvius bumblegruff. Can drive a truck (wow). Made fri
1971                  This is an extremely rare horned Parthenon. Not amused. Wears shoes. Overa
1972                   This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses t
1973                     This is an Albanian 3 1/2 legged  Episcopalian. Loves well-polished har
1983                                                  This is the happiest dog you will ever see
1984                      Here is the Rand Paul of retrievers folks! He's probably good at poker
1985                                 My oh my. This is a rare blond Canadian terrier on wheels.
1986                 Here is a Siberian heavily armored polar bear mix. Strong owner. 10/10 I wo
1987                   This is an odd dog. Hard on the outside but loving on the inside. Petting
1988                 This is a truly beautiful English Wilson Staff retriever. Has a nice phone.
1990                    This is a purebred Piers Morgan. Loves to Netflix and chill. Always look
1991                           Here is a very happy pup. Big fan of well-maintained decks. Just
1992                  This is a western brown Mitsubishi terrier. Upset about leaf. Actually 2 d

     name
49      a
472     a
582     a
746     a
748     a
757     a
762    an
783     a
919     a
929     a
1045    a
1054    a
1064    a
1065   an
1070    a
1082    a
1185    a
1209  the
1275  the
1398    a
1441    a
1452  the
1469  the
1504    a
1527    a
1528    a
1569    a
```

```
1586      a
1639      a
1682    the
1796      a
1840      a
1853      a
1863      a
1926      a
1965      a
1971     an
1972      a
1973     an
1983    the
1984    the
1985      a
1986      a
1987     an
1988      a
1990      a
1991      a
1992      a
```

```
In [57]: df_twitter_clean.loc[(df_twitter_clean.name == 'a'),'name'] = 'None'
         df_twitter_clean.loc[(df_twitter_clean.name == 'an'),'name'] = 'None'
         df_twitter_clean.loc[(df_twitter_clean.name == 'the'),'name'] = 'None'
```

5.2.9

```
In [58]: df_twitter_clean.name.value_counts()
```

```
Out[58]: None        591
         Charlie      11
         Oliver       10
         Lucy         10
         Cooper       10
         Tucker        9
         Penny         9
         Winston       8
         Sadie         8
         Daisy         7
         Toby          7
         Lola          7
         Bo            6
         Stanley       6
         Bella         6
         Jax           6
         Koda          6
```

```
Oscar         5
Leo           5
Rusty         5
Milo          5
Buddy         5
Dave          5
Chester       5
Louis         5
Bailey        5
Scout         5
Finn          4
Gus           4
Scooter       4
           ...
Spencer       1
Shikha        1
Sailor        1
Bauer         1
Zooey         1
Philippe      1
Wafer         1
Florence      1
Miley         1
Swagger       1
Trip          1
Berb          1
Coleman       1
Crimson       1
Mitch         1
Pippin        1
Vince         1
Emanuel       1
Aja           1
Jockson       1
Karma         1
Rumpole       1
Cupid         1
Kallie        1
Sully         1
Bobbay        1
Mingus        1
Lucky         1
Nigel         1
Bradlay       1
Name: name, Length: 952, dtype: int64
```

5.2.10

- P1~P3

5.2.11

```
In [59]: df_twitter_clean['p_conf'] = 'None'
         df_twitter_clean['p_dog'] = 'None'
         df_twitter_clean['p'] = 'None'

         #print(df_twitter_clean.p1_dog[1] == True)

         for i in range(0,len(df_twitter_clean)):
             if df_twitter_clean.p1_dog[i] == True:
                 df_twitter_clean['p'][i] = df_twitter_clean['p1'][i]
                 df_twitter_clean['p_conf'][i] = df_twitter_clean['p1_conf'][i]
                 df_twitter_clean['p_dog'][i] = df_twitter_clean['p1_dog'][i]
             elif df_twitter_clean.p2_dog[i] == True:
                 df_twitter_clean['p'][i] = df_twitter_clean['p2'][i]
                 df_twitter_clean['p_conf'][i] = df_twitter_clean['p2_conf'][i]
                 df_twitter_clean['p_dog'][i] = df_twitter_clean['p2_dog'][i]
             elif df_twitter_clean.p3_dog[i] == True:
                 df_twitter_clean['p'][i] = df_twitter_clean['p3'][i]
                 df_twitter_clean['p_conf'][i] = df_twitter_clean['p3_conf'][i]
                 df_twitter_clean['p_dog'][i] = df_twitter_clean['p3_dog'][i]
             else: df_twitter_clean.p_dog[i] = df_twitter_clean.p3_dog[i]
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:20: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  if __name__ == '__main__':
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  # Remove the CWD from sys.path while we load stuff.
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  # This is added back by InteractiveShellApp.init_path()
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  del sys.path[0]
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:14: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:15: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  from ipykernel import kernelapp as app
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:17: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:18: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:19: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
```

```
In [60]: with pd.option_context('max_colwidth', 200):
             display(df_twitter_clean[200:250][['tweet_id','p','p_dog','p_conf','p1','p1_dog','p
```

```
               tweet_id                           p  p_dog     p_conf  \
200  842846295480000512          Labrador_retriever   True   0.461076
201  842765311967449089          Labrador_retriever   True  0.0293399
202  842535590457499648                    Pembroke   True   0.685084
203  842163532590374912              French_bulldog   True   0.891227
204  842115215311396866                        chow   True   0.293493
205  841680585030541313                   Chihuahua   True   0.547401
206  841439858740625411          Labrador_retriever   True  0.0481999
207  841314665196081154                Afghan_hound   True   0.903712
208  841077006473256960            Brittany_spaniel   True   0.962985
209  840696689258311684                        None  False       None
210  840632337062862849            golden_retriever   True   0.711148
211  840370681858686976                        None  False       None
212  840268004936019968     Chesapeake_Bay_retriever   True   0.863987
213  839990271299457024   Staffordshire_bullterrier   True   0.604938
214  839549326359670784             Norwich_terrier   True    0.05248
215  839239871831150596                    Leonberg   True   0.927021
216  838921590096166913               Border_terrier   True   0.664538
217  838561493054533637                      kelpie   True   0.216562
```

```
218  838476387338051585                  Great_Pyrenees   True   0.997692
219  838083903487373313                            chow   True   0.800975
220  837820167694528512                golden_retriever   True   0.887625
221  837482249356513284                            None  False       None
222  837471256429613056              Norwegian_elkhound   True   0.976255
223  837366284874571778  American_Staffordshire_terrier   True   0.660085
224  837110210464448512                  Siberian_husky   True   0.767696
225  836989968035819520                      toy_poodle   True  0.0058873
226  836753516572119041                      schipperke   True  0.0115635
227  836677758902222849                            None  False       None
228  836380477523124226                            None  False       None
229  836260088725786625                          borzoi   True   0.564688
230  836001077879255040                         Samoyed   True   0.963558
231  835574547218894849       Staffordshire_bullterrier   True   0.610655
232  835297930240217089                      Rottweiler   True   0.341276
233  835264098648616962        Chesapeake_Bay_retriever   True  0.0875033
234  835172783151792128                   Border_collie   True   0.663138
235  835152434251116546  American_Staffordshire_terrier   True  0.0127309
236  834931633769889797     soft-coated_wheaten_terrier   True   0.196476
237  834786237630337024                  Border_terrier   True   0.156276
238  834574053763584002                golden_retriever   True   0.226564
239  834458053273591808             Rhodesian_ridgeback   True   0.468619
240  834209720923721728                golden_retriever   True   0.754799
241  834167344700198914                            None  False       None
242  834086379323871233              Labrador_retriever   True  0.0452627
243  833863086058651648                          kuvasz   True   0.494969
244  833826103416520705                       Chihuahua   True   0.438054
245  833722901757046785     West_Highland_white_terrier   True   0.918144
246  833479644947025920                golden_retriever   True   0.727039
247  833124694597443584                        Cardigan   True   0.710523
248  832998151111966721                           boxer   True   0.539036
249  832757312314028032                        Cardigan   True   0.160888

                                 p1  p1_dog   p1_conf  \
200              Labrador_retriever    True  0.461076
201                             tub   False  0.665238
202                        Pembroke    True  0.685084
203                  French_bulldog    True  0.891227
204                            chow    True  0.293493
205                       Chihuahua    True  0.547401
206                military_uniform   False  0.853684
207                    Afghan_hound    True  0.903712
208                Brittany_spaniel    True  0.962985
209                        web_site   False  0.841768
210                golden_retriever    True  0.711148
211                          teapot   False  0.981819
212        Chesapeake_Bay_retriever    True  0.863987
213       Staffordshire_bullterrier    True  0.604938
```

```
214                           swing   False  0.393527
215                        Leonberg    True  0.927021
216                  Border_terrier    True  0.664538
217                          kelpie    True  0.216562
218                  Great_Pyrenees    True  0.997692
219                            chow    True  0.800975
220                golden_retriever    True  0.887625
221                       birdhouse   False  0.541196
222              Norwegian_elkhound    True  0.976255
223  American_Staffordshire_terrier    True  0.660085
224                  Siberian_husky    True  0.767696
225                   shopping_cart   False  0.572422
226                     mortarboard   False  0.936882
227                         leopard   False  0.797410
228                    wooden_spoon   False  0.082489
229                          borzoi    True  0.564688
230                         Samoyed    True  0.963558
231       Staffordshire_bullterrier    True  0.610655
232                      Rottweiler    True  0.341276
233                           hyena   False  0.736871
234                   Border_collie    True  0.663138
235                           swing   False  0.967066
236                        ice_bear   False  0.330573
237                  Border_terrier    True  0.156276
238                   toilet_tissue   False  0.262936
239             Rhodesian_ridgeback    True  0.468619
240                golden_retriever    True  0.754799
241                              ox   False  0.991682
242                      bath_towel   False  0.736759
243                          kuvasz    True  0.494969
244                       Chihuahua    True  0.438054
245     West_Highland_white_terrier    True  0.918144
246                golden_retriever    True  0.727039
247                        Cardigan    True  0.710523
248                           boxer    True  0.539036
249                        Cardigan    True  0.160888

                                 p2  p2_dog   p2_conf  \
200                golden_retriever    True  0.154946
201                          bucket   False  0.105166
202                        Cardigan    True  0.314608
203                     soccer_ball   False  0.022811
204                    Newfoundland    True  0.181336
205                         bow_tie   False  0.198361
206              Labrador_retriever    True  0.048200
207                          Saluki    True  0.035215
208                Blenheim_spaniel    True  0.014820
209                            rule   False  0.007087
```

```
210                  cocker_spaniel    True  0.157929
211                             cup   False  0.014026
212              Labrador_retriever    True  0.052632
213  American_Staffordshire_terrier    True  0.311540
214                 Norwich_terrier    True  0.052480
215                    Newfoundland    True  0.050009
216               Brabancon_griffon    True  0.170451
217                         doormat   False  0.139994
218                          kuvasz    True  0.001001
219                       seat_belt   False  0.164133
220              Labrador_retriever    True  0.068718
221                      can_opener   False  0.121094
222                        keeshond    True  0.013990
223       Staffordshire_bullterrier    True  0.334947
224                      Eskimo_dog    True  0.217079
225                 shopping_basket   False  0.414002
226                   academic_gown   False  0.020815
227                          jaguar   False  0.095487
228                    sliding_door   False  0.061017
229                        ice_bear   False  0.078267
230                      white_wolf   False  0.019848
231                          muzzle   False  0.132138
232                  Border_terrier    True  0.336220
233        Chesapeake_Bay_retriever    True  0.087503
234                          collie    True  0.152494
235  American_Staffordshire_terrier    True  0.012731
236     soft-coated_wheaten_terrier    True  0.196476
237              Norwegian_elkhound    True  0.125912
238                golden_retriever    True  0.226564
239                         whippet    True  0.177531
240                        Pekinese    True  0.197861
241                           bison   False  0.005335
242                    sleeping_bag   False  0.062959
243                  Great_Pyrenees    True  0.312632
244                          kelpie    True  0.149706
245                     Maltese_dog    True  0.025721
246                  cocker_spaniel    True  0.071140
247                          kelpie    True  0.106102
248                  French_bulldog    True  0.317617
249       Staffordshire_bullterrier    True  0.159441

                                 p3  p3_dog   p3_conf
200                       Chihuahua    True  0.110249
201              Labrador_retriever    True  0.029340
202                         basenji    True  0.000160
203                    bull_mastiff    True  0.012852
204                      schipperke    True  0.125152
205                        Pembroke    True  0.058493
```

```
206                     groenendael    True  0.015394
207                      bloodhound    True  0.026565
208                         clumber    True  0.009557
209                        envelope   False  0.006820
210              Labrador_retriever    True  0.059582
211                       coffeepot   False  0.002421
212                          kelpie    True  0.032574
213                     Boston_bull    True  0.037159
214                        Pembroke    True  0.049901
215                   Saint_Bernard    True  0.010728
216               Yorkshire_terrier    True  0.087824
217                       dalmatian    True  0.132820
218                    Newfoundland    True  0.000405
219                      Pomeranian    True  0.017981
220                          kuvasz    True  0.030387
221                          carton   False  0.056137
222                       seat_belt   False  0.002111
223                       dalmatian    True  0.002697
224                        malamute    True  0.011657
225                      toy_poodle    True  0.005887
226                      schipperke    True  0.011564
227                    snow_leopard   False  0.079694
228                     grand_piano   False  0.055086
229                        Pembroke    True  0.057916
230                        malamute    True  0.005904
231  American_Staffordshire_terrier    True  0.109544
232                   Gordon_setter    True  0.045448
233                         meerkat   False  0.042058
234                        Cardigan    True  0.035471
235       Staffordshire_bullterrier    True  0.007039
236                   Irish_terrier    True  0.073097
237                     Boston_bull    True  0.096624
238                         bathtub   False  0.078879
239                         redbone    True  0.106552
240              Labrador_retriever    True  0.008654
241                   water_buffalo   False  0.001130
242              Labrador_retriever    True  0.045263
243                golden_retriever    True  0.141736
244                        Pembroke    True  0.096480
245                Lakeland_terrier    True  0.020211
246                 Tibetan_mastiff    True  0.048694
247                   shopping_cart   False  0.055475
248                    bull_mastiff    True  0.093928
249                     Boston_bull    True  0.154368
```

```
In [61]: df_twitter_clean = df_twitter_clean.drop(['p1','p1_conf','p1_dog','p2','p2_conf','p2_do
```

5.2.12

```
In [62]: df_twitter_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 1993
Data columns (total 11 columns):
tweet_id          1994 non-null object
timestamp         1994 non-null datetime64[ns]
text              1994 non-null object
name              1994 non-null object
retweet_count     1994 non-null float64
favorite_count    1994 non-null float64
status            1994 non-null object
rating            1993 non-null float64
p_conf            1994 non-null object
p_dog             1994 non-null object
p                 1994 non-null object
dtypes: datetime64[ns](1), float64(3), object(7)
memory usage: 266.9+ KB
```

5.2.13

Q6_np.nan’None’

5.2.14

```
In [63]: print('None' in df_twitter_clean.tweet_id.value_counts())
         print('None' in df_twitter_clean.timestamp.value_counts())
         print('None' in df_twitter_clean.text.value_counts())
         print('None' in df_twitter_clean.name.value_counts())
         print('None' in df_twitter_clean.retweet_count.value_counts())
         print('None' in df_twitter_clean.favorite_count.value_counts())
         print('None' in df_twitter_clean.status.value_counts())
         print('None' in df_twitter_clean.rating.value_counts())
         print('None' in df_twitter_clean.p_conf.value_counts())
         print('None' in df_twitter_clean.p_dog.value_counts())
         print('None' in df_twitter_clean.p.value_counts())
```

```
False
False
False
True
False
False
False
False
True
```

```
False
True
```

```
In [64]: df_twitter_clean.name = df_twitter_clean.name.replace('None',np.nan)
         df_twitter_clean.p_conf = df_twitter_clean.p_conf.replace('None',np.nan)
         df_twitter_clean.p = df_twitter_clean.p.replace('None',np.nan)
```

5.2.15

```
In [65]: print('None' in df_twitter_clean.tweet_id.value_counts())
         print('None' in df_twitter_clean.timestamp.value_counts())
         print('None' in df_twitter_clean.text.value_counts())
         print('None' in df_twitter_clean.name.value_counts())
         print('None' in df_twitter_clean.retweet_count.value_counts())
         print('None' in df_twitter_clean.favorite_count.value_counts())
         print('None' in df_twitter_clean.status.value_counts())
         print('None' in df_twitter_clean.rating.value_counts())
         print('None' in df_twitter_clean.p_conf.value_counts())
         print('None' in df_twitter_clean.p_dog.value_counts())
         print('None' in df_twitter_clean.p.value_counts())
```

```
False
False
False
False
False
False
False
False
False
False
False
```

5.3

tweet-archive-master

tweet_id df_tweeter text df_tweeter timestamp df_tweeter retweet_count df_json favorite_count df_json rating df_tweeter name df_tweeter status df_tweeter p (df_pred) p_conf (df_pred)

```
In [66]: df_twitter_clean.to_csv('twitter_archive_master.csv', index=False)
```

6

* / * * * (wordcloud * *

```
In [67]: #sns.set()
         df_twitter_drop = df_twitter_clean.dropna(axis=0)
         df_twitter_drop.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1229 entries, 1 to 1963
Data columns (total 11 columns):
tweet_id          1229 non-null object
timestamp         1229 non-null datetime64[ns]
text              1229 non-null object
name              1229 non-null object
retweet_count     1229 non-null float64
favorite_count    1229 non-null float64
status            1229 non-null object
rating            1229 non-null float64
p_conf            1229 non-null float64
p_dog             1229 non-null object
p                 1229 non-null object
dtypes: datetime64[ns](1), float64(4), object(6)
memory usage: 115.2+ KB
```

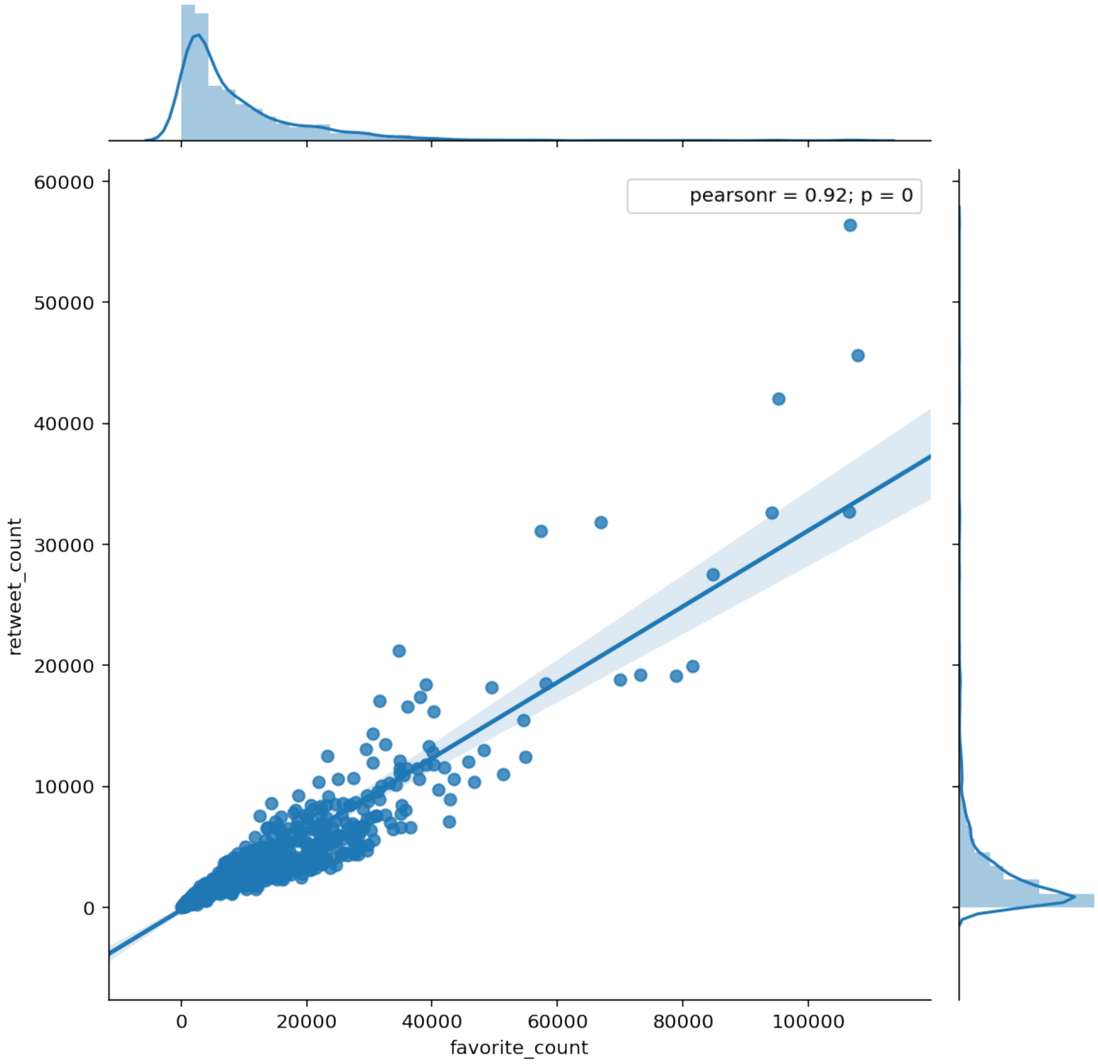
6.1 1

- 1-1 R-squared0.855
- 1-230%

```
In [68]: #favorite_count & retweet_count
         f_and_r = sns.jointplot('favorite_count','retweet_count', data = df_twitter_drop,kind =
         mod = smf.ols(formula='retweet_count ~ favorite_count', data=df_twitter_drop)
         res = mod.fit()
         print(res.summary())
         #f_and_r = f_and_r.plot(sns.regplot, sns.distplot)
         #f_and_r = f_and_r.annotate(f_and_r.pearsonr,rsquare)
         #https://blog.csdn.net/ice_martin/article/details/61617053
         #http://seaborn.pydata.org/index.html
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          retweet_count   R-squared:                       0.855
Model:                            OLS   Adj. R-squared:                  0.855
Method:                 Least Squares   F-statistic:                     7241.
Date:                Tue, 03 Apr 2018   Prob (F-statistic):               0.00
Time:                        13:40:54   Log-Likelihood:                -10785.
No. Observations:                1229   AIC:                         2.157e+04
Df Residuals:                    1227   BIC:                         2.158e+04
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept      -224.8620     56.971     -3.947      0.000    -336.633    -113.091
favorite_count    0.3138      0.004     85.092      0.000       0.307       0.321
==============================================================================
Omnibus:                     1194.006   Durbin-Watson:                   1.364
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           147714.683
Skew:                           4.140   Prob(JB):                         0.00
Kurtosis:                      56.066   Cond. No.                     1.97e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.97e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

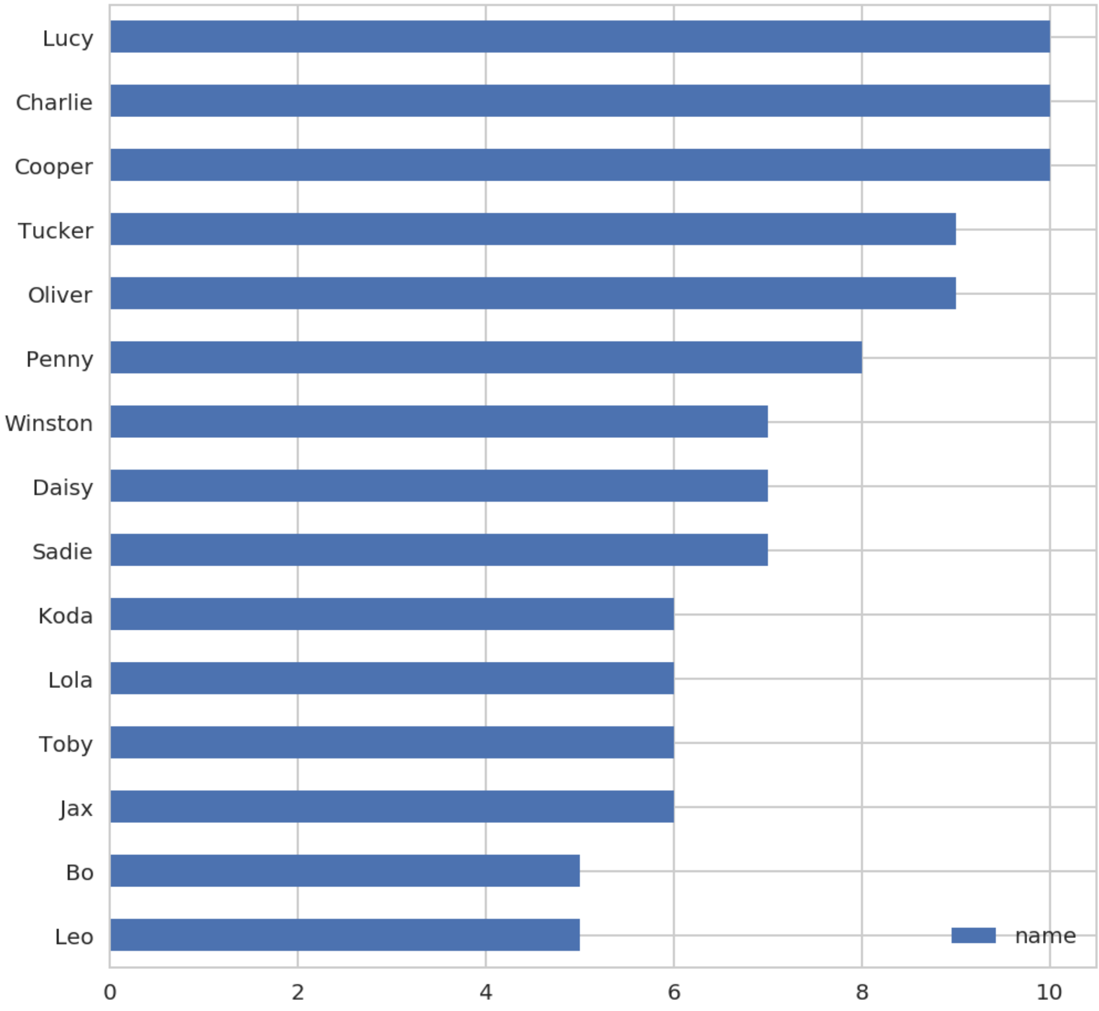


6.2 2

- 2-1 TuckerLucyCharlieCooperPennyOliverTop5

```
In [69]: #
         sns.set(style='whitegrid')
         sns_named = pd.DataFrame(df_twitter_drop.name.value_counts().sort_values().tail(15))
         sns_named.plot(kind='barh',figsize = (8,8))
```

```
Out[69]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3bfbba06d8>
```



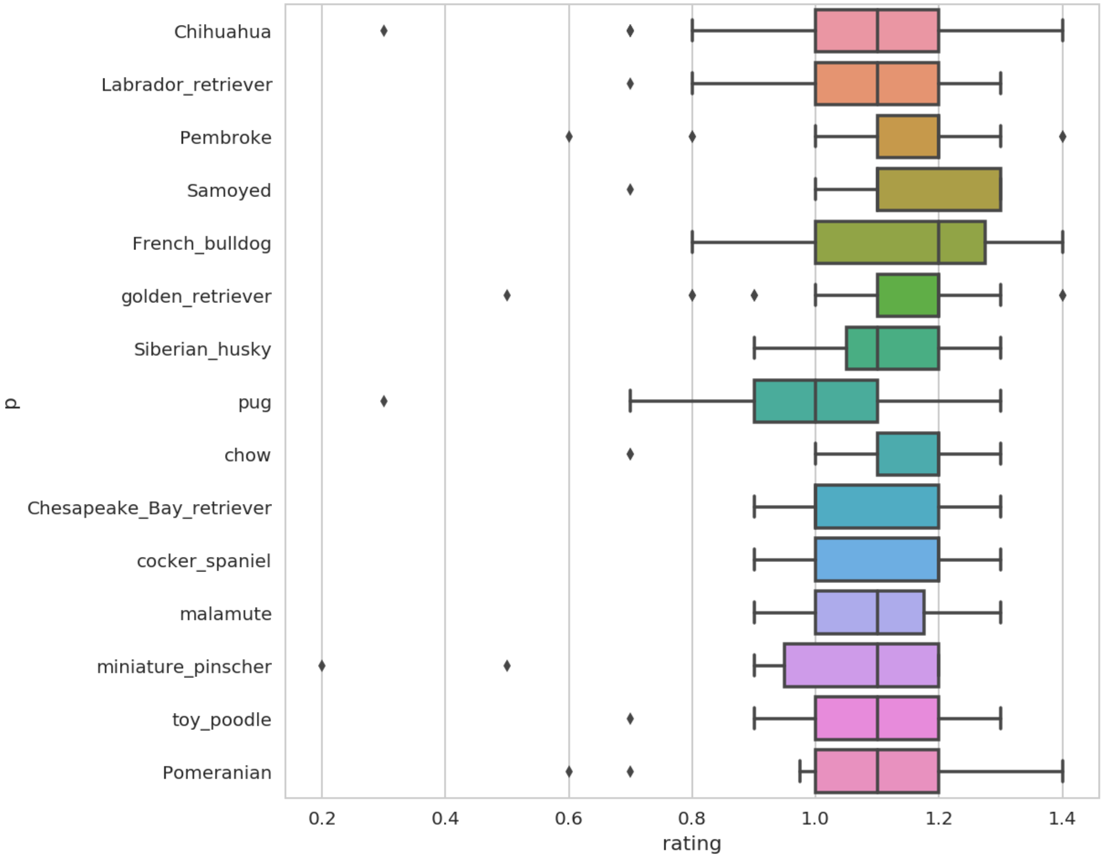
6.3 3

- 3-1 Chihuahua

- 3-2 French Bulldog
- 3-3

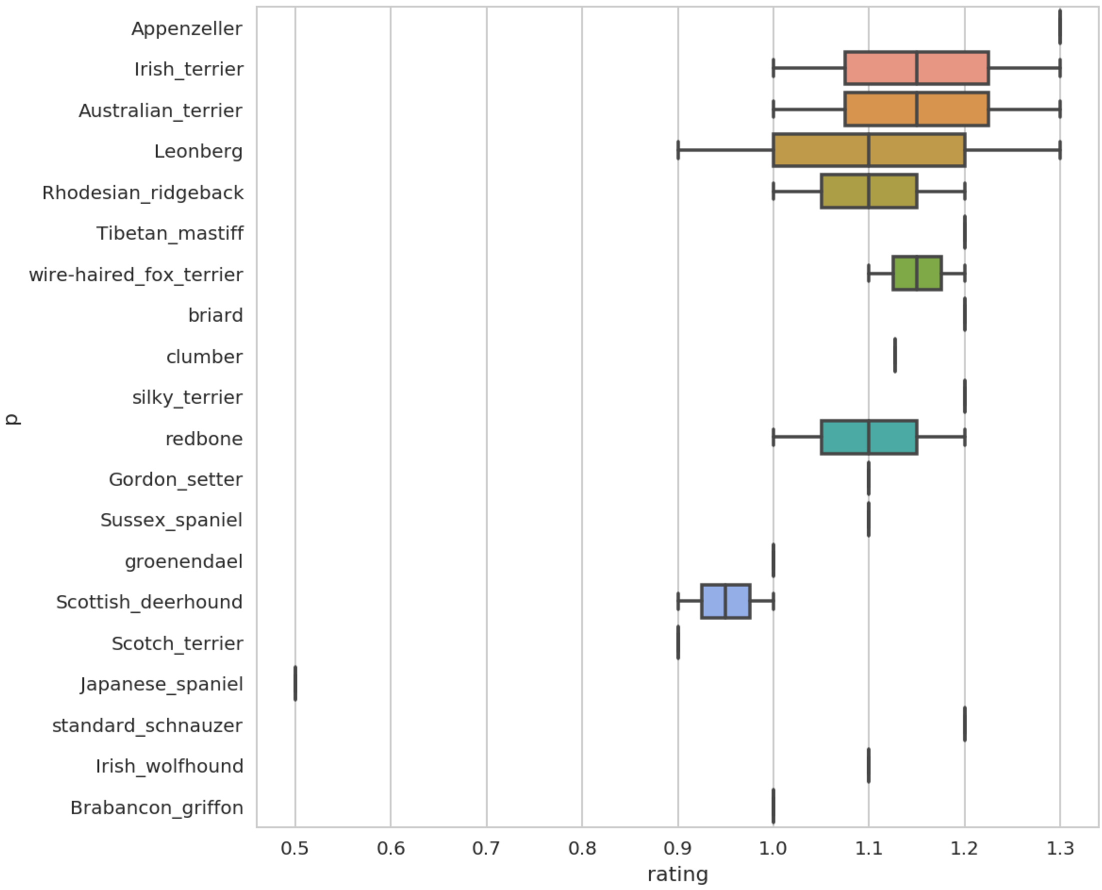
```
In [70]: sns_p = pd.DataFrame(df_twitter_drop.p.value_counts().sort_values().tail(15))
         df_p = df_twitter_drop[df_twitter_drop.p.isin(sns_p.index)]
         plt.subplots(figsize=(8, 8))
         sns.boxplot(x='rating',y='p',data = df_p)
```

```
Out[70]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3bfbca2550>
```



```
In [71]: sns_p_2 = pd.DataFrame(df_twitter_drop.p.value_counts().sort_values().head(20))
         df_p_2 = df_twitter_drop[df_twitter_drop.p.isin(sns_p_2.index)]
         plt.subplots(figsize=(8, 8))
         sns.boxplot(x='rating',y='p',data = df_p_2)
```

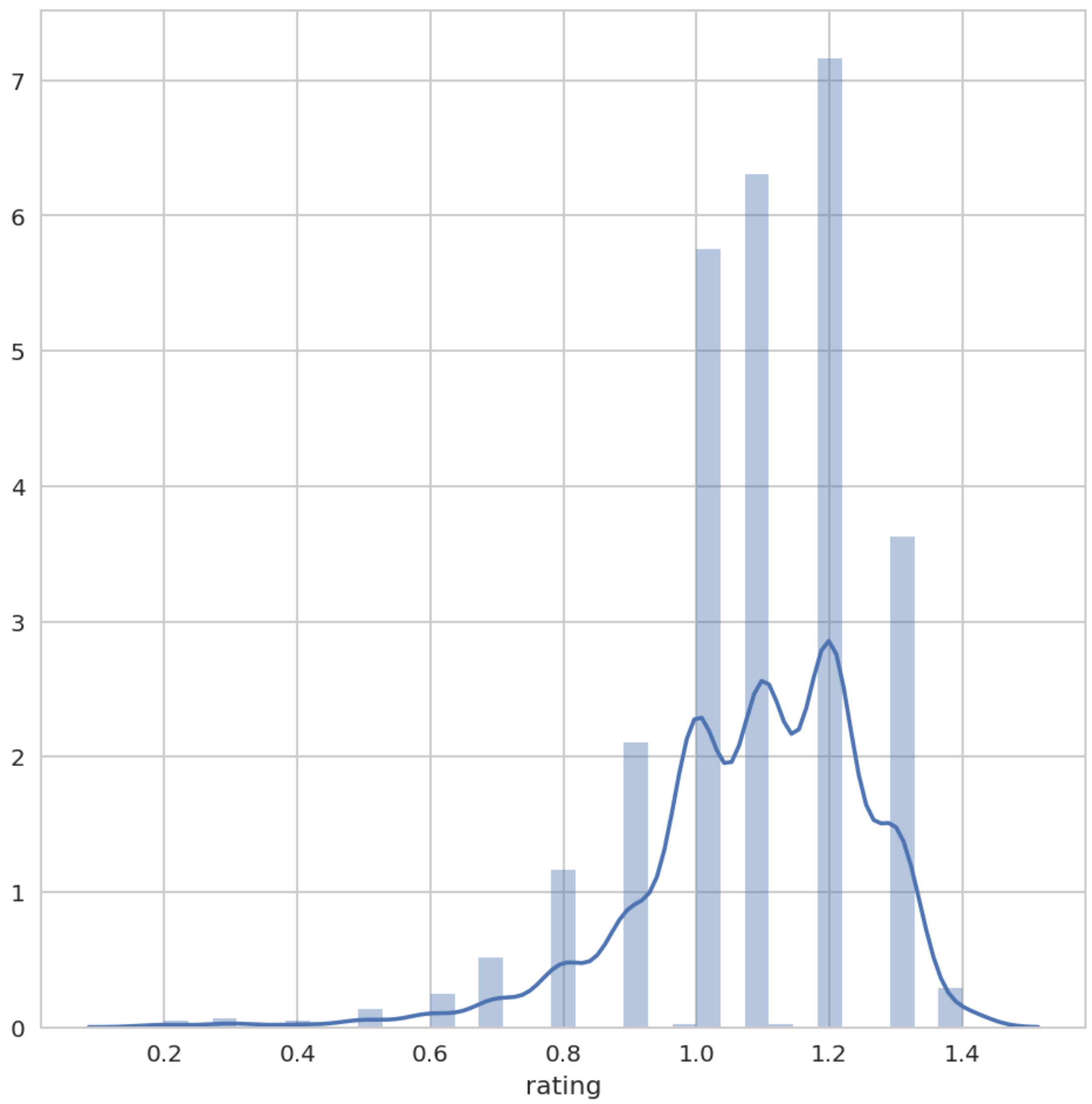
```
Out[71]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3bfb59d390>
```



6.4 4

- 4-1 1.0~1.2
- 4-2

```
In [72]: plt.subplots(figsize=(8, 8))
         sns_rating = sns.distplot(df_twitter_drop.rating,hist=True,)
```

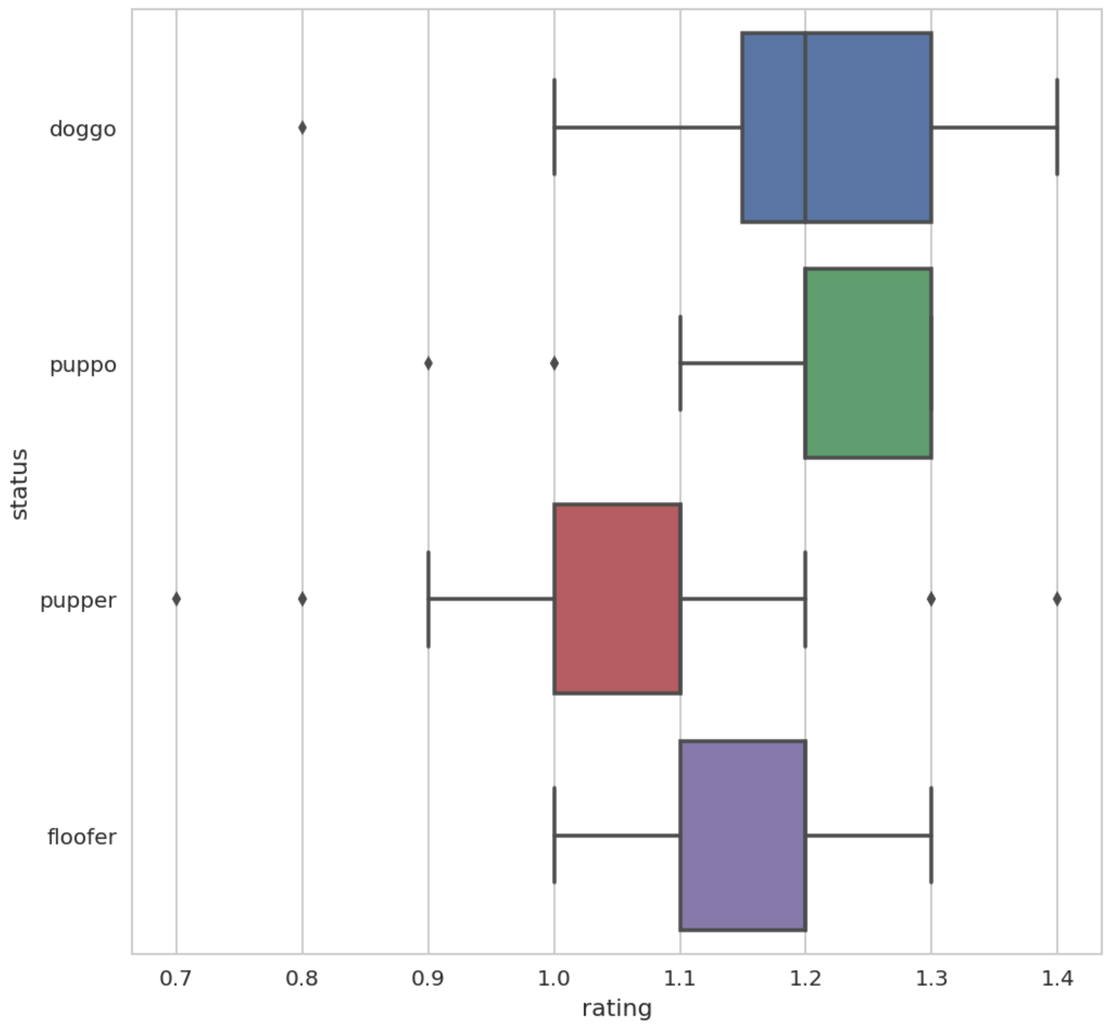


6.5 5

- 5-1 doggo
- 5-2 pupper
- 5-3

```
In [73]: sns_status = pd.DataFrame(df_twitter_drop.status.value_counts().sort_values().head(4))
         df_status = df_twitter_drop[df_twitter_drop.status.isin(sns_status.index)]
         sns.set(style='whitegrid')
         plt.subplots(figsize=(8, 8))
         sns.boxplot(x='rating',y='status',data = df_status)
```

```
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3bf87ae0b8>
```


doggo
puppo
pupper
floofer
status
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
rating