wrangle_act

October 22, 2018

1 GATHERING :

```
In [41]: import requests
         import pandas as pd
         import numpy as np
         import tweepy
         import json
         import time
         import matplotlib.pyplot as plt
         import os
         import re
```

GATHER TWITTER ARCHIVE

```
In [42]: twitter_archive = pd.read_csv('twitter-archive-enhanced.csv',encoding='cp1252')
         twitter_archive.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2517 entries, 0 to 2516
Data columns (total 17 columns):
tweet_id                    2467 non-null object
in_reply_to_status_id       94 non-null object
in_reply_to_user_id         94 non-null object
timestamp                   2372 non-null object
source                      2447 non-null object
text                        2451 non-null object
retweeted_status_id         261 non-null float64
retweeted_status_user_id    261 non-null object
retweeted_status_timestamp  261 non-null object
expanded_urls               2302 non-null object
rating_numerator            2356 non-null object
rating_denominator          2356 non-null object
name                        2261 non-null object
doggo                       2261 non-null object
floofer                     2261 non-null object
pupper                      2261 non-null object
puppo                       2261 non-null object
```

```
dtypes: float64(1), object(16)
memory usage: 334.4+ KB
```

GETTING IMAGES PREDICTION USING URL

```
In [43]: url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predicti
         response = requests.get(url)
         folder = 'image-predictions'
         if not os.path.exists(folder):
             os.makedirs(folder)

         with open(os.path.join(folder, url.split('/')[-1]), mode='wb') as file:
             file.write(response.content)
```

GATHERING TWITTER STATUS USING API

```
In [ ]: consumer_key = "htHF71TQ6FzX75nC0cNDVwOL8"
        consumer_secret = "ZDVtGSvSOkvsZzjC961Ctb5EUHnHdrV8iMp21wKgOA60Yt53AY"
        access_token = "2280373254-29A3Ov7Iv9KvX9ePwOHDeg4bAnI973ufu3A5lyF"
        access_secret = "1fKZ7IJErOIhbocJUR1Hk6rXMTHKoKVDCmrmEOxCOGurI"

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

        api = tweepy.API(auth, parser=tweepy.parsers.JSONParser(),
                         wait_on_rate_limit=True, wait_on_rate_limit_notify=True)
```

GATHERING TWEETS STATUS IN JSON FORMAT

```
In [ ]: twitter_list = []
        errors = []
        ta_tweet_id_list = twitter_archive.tweet_id.tolist()
        status_list = range(1, len(ta_tweet_id_list), 100)

        for index, id in enumerate(ta_tweet_id_list):
            try:
                page = api.get_status(id)
                favorites = page['favorite_count']
                retweets = page['retweet_count']
                twitter_list.append({'tweet_id': int(id),
                                     'favorites': int(favorites),
                                     'retweets': int(retweets)})
                if index in status_list:
                    print('Completed ~', index, ' Length of twitter_list is',
                          len(twitter_list), ' Length of errors list is',
                          len(errors))
                else:
```

```
                    pass
            except Exception as e:
                errors.append((id, e))

In [ ]: with open('tweet_json.txt', 'w') as file:
            file.write(json.dumps(twitter_list))
```

2 ASSESSING :

ASSESSING JSON TXT through dataframe.

```
In [44]: status = pd.read_json('tweet_json.txt')
         status.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2341 entries, 0 to 2340
Data columns (total 3 columns):
favorites    2341 non-null int64
retweets     2341 non-null int64
tweet_id     2341 non-null int64
dtypes: int64(3)
memory usage: 73.2 KB


In [ ]: ASSESSING IMAGE PREDICTION

In [5]: image_prediction = pd.read_csv(folder + '/image-predictions.tsv', sep='\t')
        image_prediction.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
```

ASSESSING TWITTER ARCHIVE

```
In [6]: twitter_archive.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2517 entries, 0 to 2516
Data columns (total 17 columns):
tweet_id                    2467 non-null object
in_reply_to_status_id       94 non-null object
in_reply_to_user_id         94 non-null object
timestamp                   2372 non-null object
source                      2447 non-null object
text                        2451 non-null object
retweeted_status_id         261 non-null float64
retweeted_status_user_id    261 non-null object
retweeted_status_timestamp  261 non-null object
expanded_urls               2302 non-null object
rating_numerator            2356 non-null object
rating_denominator          2356 non-null object
name                        2261 non-null object
doggo                       2261 non-null object
floofer                     2261 non-null object
pupper                      2261 non-null object
puppo                       2261 non-null object
dtypes: float64(1), object(16)
memory usage: 334.4+ KB
```

3   CLEANING

Cleaning has to be done for 8 Quality issues and 2 tidy issues .

   1.Retweets to be discarded as we require only original tweets.

```
In [45]: twitter_archive_clean=twitter_archive[twitter_archive.retweeted_status_id.isnull()]
         twitter_archive_clean=twitter_archive_clean.drop(['retweeted_status_id','retweeted_stat
         twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2256 entries, 0 to 2516
Data columns (total 14 columns):
tweet_id             2206 non-null object
in_reply_to_status_id  79 non-null object
in_reply_to_user_id    79 non-null object
timestamp            2191 non-null object
source               2191 non-null object
text                 2190 non-null object
expanded_urls        2042 non-null object
rating_numerator     2095 non-null object
rating_denominator   2095 non-null object
name                 2095 non-null object
```

```
doggo                   2095 non-null object
floofer                 2095 non-null object
pupper                  2095 non-null object
puppo                   2095 non-null object
dtypes: object(14)
memory usage: 264.4+ KB
```

2.Tweets id in twitter_archive has blank values and unneeded values.

```
In [8]: twitter_archive_clean.dropna(subset=['tweet_id'])
        twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2256 entries, 0 to 2516
Data columns (total 14 columns):
tweet_id                2206 non-null object
in_reply_to_status_id   79 non-null object
in_reply_to_user_id     79 non-null object
timestamp               2191 non-null object
source                  2191 non-null object
text                    2190 non-null object
expanded_urls           2042 non-null object
rating_numerator        2095 non-null object
rating_denominator      2095 non-null object
name                    2095 non-null object
doggo                   2095 non-null object
floofer                 2095 non-null object
pupper                  2095 non-null object
puppo                   2095 non-null object
dtypes: object(14)
memory usage: 264.4+ KB
```

```
In [9]: twitter_archive_clean=twitter_archive_clean.drop_duplicates(subset=['tweet_id'])
        twitter_archive_clean.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2204 entries, 0 to 2516
Data columns (total 14 columns):
tweet_id                2203 non-null object
in_reply_to_status_id   79 non-null object
in_reply_to_user_id     79 non-null object
timestamp               2191 non-null object
source                  2191 non-null object
text                    2190 non-null object
expanded_urls           2042 non-null object
rating_numerator        2095 non-null object
rating_denominator      2095 non-null object
```

```
name                    2095 non-null object
doggo                   2095 non-null object
floofer                 2095 non-null object
pupper                  2095 non-null object
puppo                   2095 non-null object
dtypes: object(14)
memory usage: 258.3+ KB
```

3.Drop extreme outlier values of ratings which is calculated by dividing rating_numerator and rating_denominator.

```
In [10]: twitter_archive_clean['rating_numerator']=pd.to_numeric(twitter_archive_clean['rating_n
         twitter_archive_clean['rating_denominator']=pd.to_numeric(twitter_archive_clean['rating
         twitter_archive_clean['rating_ratio']=twitter_archive_clean['rating_numerator']/twitter
         twitter_archive_clean = twitter_archive_clean[twitter_archive_clean['rating_ratio']< 1.
         twitter_archive_clean.rating_ratio.value_counts()

Out[10]: 1.200000    483
         1.000000    429
         1.100000    422
         1.300000    281
         0.900000    153
         0.800000     97
         0.700000     53
         1.400000     41
         0.500000     36
         0.600000     32
         0.300000     19
         0.400000     15
         0.200000     11
         0.100000      7
         0.000000      2
         0.818182      1
         1.700000      1
         0.733333      1
         0.636364      1
         1.250000      1
         1.500000      1
         Name: rating_ratio, dtype: int64
```

4.Identifying wrong names of dogs and replacing with 'NaN'in twitter_archive

```
In [11]: twitter_archive_clean.name=twitter_archive_clean.name.replace('None',np.nan)
         twitter_archive_clean.name=twitter_archive_clean.name.replace('a',np.nan)
         twitter_archive_clean.name=twitter_archive_clean.name.replace('the',np.nan)
         twitter_archive_clean.name=twitter_archive_clean.name.replace('an',np.nan)

                    tweet_id in_reply_to_status_id in_reply_to_user_id  \
0      892420643555336193                   NaN                 NaN
```

1	8921774213063434426	NaN	NaN
2	891815181378084864	NaN	NaN
3	891689557279858688	NaN	NaN
4	891327558926688256	NaN	NaN
5	891087950875897856	NaN	NaN
9	890729181411237888	NaN	NaN
10	890609185150312448	NaN	NaN
11	890240255349198849	NaN	NaN
12	890006608113172480	NaN	NaN
13	889880896479866881	NaN	NaN
14	889665388333682689	NaN	NaN
15	889638837579907072	NaN	NaN
16	889531135344209921	NaN	NaN
17	889278841981685760	NaN	NaN
18	888917238123831296	NaN	NaN
19	888804989199671297	NaN	NaN
20	888554962724278272	NaN	NaN
22	888078434458587136	NaN	NaN
23	887705289381826560	NaN	NaN
24	887517139158093824	NaN	NaN
25	887473957103951883	NaN	NaN
26	887343217045368832	NaN	NaN
27	887101392804085760	NaN	NaN
28	886983233522544640	NaN	NaN
32	886680336477933568	NaN	NaN
33	886366144734445568	NaN	NaN
34	886267009285017600	8.86E+17	2281181600
35	886258384151887873	NaN	NaN
37	885984800019947520	NaN	NaN
...	...	...	...
2487	666411507551481857	NaN	NaN
2488	666407126856765440	NaN	NaN
2489	666396247373291520	NaN	NaN
2490	666373753744588802	NaN	NaN
2491	666362758909284353	NaN	NaN
2492	666353288456101888	NaN	NaN
2493	666345417576210432	NaN	NaN
2494	666337882303524864	NaN	NaN
2495	666293911632134144	NaN	NaN
2496	666287406224695296	NaN	NaN
2497	666273097616637952	NaN	NaN
2498	666268910803644416	NaN	NaN
2499	666104133288665088	NaN	NaN
2500	666102155909144576	NaN	NaN
2501	666099513787052032	NaN	NaN
2502	666094000022159362	NaN	NaN
2503	666082916733198337	NaN	NaN
2504	666073100786774016	NaN	NaN

```
2505   6660711932215091204                        NaN              NaN
2506   666063827256086533                         NaN              NaN
2507   666058600524156928                         NaN              NaN
2508   666057090499244032                         NaN              NaN
2509   666055525042405380                         NaN              NaN
2510   666051853826850816                         NaN              NaN
2511   666050758794694657                         NaN              NaN
2512   666049248165822465                         NaN              NaN
2513   666044226329800704                         NaN              NaN
2514   666033412701032449                         NaN              NaN
2515   666029285002620928                         NaN              NaN
2516   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
9      2017-07-28 00:22:40 +0000
10     2017-07-27 16:25:51 +0000
11     2017-07-26 15:59:51 +0000
12     2017-07-26 00:31:25 +0000
13     2017-07-25 16:11:53 +0000
14     2017-07-25 01:55:32 +0000
15     2017-07-25 00:10:02 +0000
16     2017-07-24 17:02:04 +0000
17     2017-07-24 00:19:32 +0000
18     2017-07-23 00:22:39 +0000
19     2017-07-22 16:56:37 +0000
20     2017-07-22 00:23:06 +0000
22     2017-07-20 16:49:33 +0000
23     2017-07-19 16:06:48 +0000
24     2017-07-19 03:39:09 +0000
25     2017-07-19 00:47:34 +0000
26     2017-07-18 16:08:03 +0000
27     2017-07-18 00:07:08 +0000
28     2017-07-17 16:17:36 +0000
32     2017-07-16 20:14:00 +0000
33     2017-07-15 23:25:31 +0000
34     2017-07-15 16:51:35 +0000
35     2017-07-15 16:17:19 +0000
37     2017-07-14 22:10:11 +0000
...                          ...
2487   2015-11-17 00:24:19 +0000
2488   2015-11-17 00:06:54 +0000
2489   2015-11-16 23:23:41 +0000
```

```
2490   2015-11-16 21:54:18 +0000
2491   2015-11-16 21:10:36 +0000
2492   2015-11-16 20:32:58 +0000
2493   2015-11-16 20:01:42 +0000
2494   2015-11-16 19:31:45 +0000
2495   2015-11-16 16:37:02 +0000
2496   2015-11-16 16:11:11 +0000
2497   2015-11-16 15:14:19 +0000
2498   2015-11-16 14:57:41 +0000
2499   2015-11-16 04:02:55 +0000
2500   2015-11-16 03:55:04 +0000
2501   2015-11-16 03:44:34 +0000
2502   2015-11-16 03:22:39 +0000
2503   2015-11-16 02:38:37 +0000
2504   2015-11-16 01:59:36 +0000
2505   2015-11-16 01:52:02 +0000
2506   2015-11-16 01:22:45 +0000
2507   2015-11-16 01:01:59 +0000
2508   2015-11-16 00:55:59 +0000
2509   2015-11-16 00:49:46 +0000
2510   2015-11-16 00:35:11 +0000
2511   2015-11-16 00:30:50 +0000
2512   2015-11-16 00:24:50 +0000
2513   2015-11-16 00:04:52 +0000
2514   2015-11-15 23:21:54 +0000
2515   2015-11-15 23:05:30 +0000
2516   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...
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...
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...
32      <a href="http://twitter.com/download/iphone" r...
33      <a href="http://twitter.com/download/iphone" r...
34      <a href="http://twitter.com/download/iphone" r...
35      <a href="http://twitter.com/download/iphone" r...
37      <a href="http://twitter.com/download/iphone" r...
...                                                   ...
2487    <a href="http://twitter.com/download/iphone" r...
2488    <a href="http://twitter.com/download/iphone" r...
2489    <a href="http://twitter.com/download/iphone" r...
2490    <a href="http://twitter.com/download/iphone" r...
2491    <a href="http://twitter.com/download/iphone" r...
2492    <a href="http://twitter.com/download/iphone" r...
2493    <a href="http://twitter.com/download/iphone" r...
2494    <a href="http://twitter.com/download/iphone" r...
2495    <a href="http://twitter.com/download/iphone" r...
2496    <a href="http://twitter.com/download/iphone" r...
2497    <a href="http://twitter.com/download/iphone" r...
2498    <a href="http://twitter.com/download/iphone" r...
2499    <a href="http://twitter.com/download/iphone" r...
2500    <a href="http://twitter.com/download/iphone" r...
2501    <a href="http://twitter.com/download/iphone" r...
2502    <a href="http://twitter.com/download/iphone" r...
2503    <a href="http://twitter.com/download/iphone" r...
2504    <a href="http://twitter.com/download/iphone" r...
2505    <a href="http://twitter.com/download/iphone" r...
2506    <a href="http://twitter.com/download/iphone" r...
2507    <a href="http://twitter.com/download/iphone" r...
2508    <a href="http://twitter.com/download/iphone" r...
2509    <a href="http://twitter.com/download/iphone" r...
2510    <a href="http://twitter.com/download/iphone" r...
2511    <a href="http://twitter.com/download/iphone" r...
2512    <a href="http://twitter.com/download/iphone" r...
2513    <a href="http://twitter.com/download/iphone" r...
2514    <a href="http://twitter.com/download/iphone" r...
2515    <a href="http://twitter.com/download/iphone" r...
2516    <a href="http://twitter.com/download/iphone" r...

                                                    text  \
0       This is Phineas. He's a mystical boy. Only eve...
1       This is Tilly. She's just checking pup on you...
2       This is Archie. He is a rare Norwegian Pouncin...
3       This is Darla. She commenced a snooze mid meal...
```

```
4     This is Franklin. He would like you to stop ca...
5     Here we have a majestic great white breaching ...
9     When you watch your owner call another dog a g...
10    This is Zoey. She doesn't want to be one of th...
11    This is Cassie. She is a college pup. Studying...
12    This is Koda. He is a South Australian decksha...
13    This is Bruno. He is a service shark. Only get...
14    Here's a puppo that seems to be on the fence a...
15    This is Ted. He does his best. Sometimes that'...
16    This is Stuart. He's sporting his favorite fan...
17    This is Oliver. You're witnessing one of his m...
18    This is Jim. He found a fren. Taught him how t...
19    This is Zeke. He has a new stick. Very proud o...
20    This is Ralphus. He's powering up. Attempting ...
22    This is Gerald. He was just told he didn't get...
23    This is Jeffrey. He has a monopoly on the pool...
24    I've yet to rate a Venezuelan Hover Wiener. Th...
25    This is Canela. She attempted some fancy porch...
26    You may not have known you needed to see this ...
27    This... is a Jubilant Antarctic House Bear. We...
28    This is Maya. She's very shy. Rarely leaves he...
32    This is Derek. He's late for a dog meeting. 13...
33    This is Roscoe. Another pupper fallen victim t...
34    @NonWhiteHat @MayhewMayhem omg hello tanner yo...
35    This is Waffles. His doggles are pupside down...
37    Viewer discretion advised. This is Jimbo. He w...
...                                             ...
2487  This is quite the dog. Gets really excited whe...
2488  This is a southern Vesuvius bumblegruff. Can d...
2489  Oh goodness. A super rare northeast Qdoba kang...
2490  Those are sunglasses and a jean jacket. 11/10 ...
2491  Unique dog here. Very small. Lives in containe...
2492  Here we have a mixed Asiago from the Gal?ípago...
2493  Look at this jokester thinking seat belt laws ...
2494  This is an extremely rare horned Parthenon. No...
2495  This is a funny dog. Weird toes. Won't come do...
2496  This is an Albanian 3 1/2 legged  Episcopalian...
2497     Can take selfies 11/10 https://t.co/ws2AMaNwPW
2498  Very concerned about fellow dog trapped in com...
2499  Not familiar with this breed. No tail (weird)...
2500  Oh my. Here you are seeing an Adobe Setter giv...
2501  Can stand on stump for what seems like a while...
2502  This appears to be a Mongolian Presbyterian mi...
2503  Here we have a well-established sunblockerspan...
2504  Let's hope this flight isn't Malaysian (lol). ...
2505  Here we have a northern speckled Rhododendron...
2506  This is the happiest dog you will ever see. Ve...
2507  Here is the Rand Paul of retrievers folks! He'...
```

```
2508  My oh my. This is a rare blond Canadian terrie...
2509  Here is a Siberian heavily armored polar bear ...
2510  This is an odd dog. Hard on the outside but lo...
2511  This is a truly beautiful English Wilson Staff...
2512  Here we have a 1949 1st generation vulpix. Enj...
2513  This is a purebred Piers Morgan. Loves to Netf...
2514  Here is a very happy pup. Big fan of well-main...
2515  This is a western brown Mitsubishi terrier. Up...
2516  Here we have a Japanese Irish Setter. Lost eye...


                                        expanded_urls  rating_numerator  \
0      https://twitter.com/dog_rates/status/892420643...              13.0
1      https://twitter.com/dog_rates/status/892177421...              13.0
2      https://twitter.com/dog_rates/status/891815181...              12.0
3      https://twitter.com/dog_rates/status/891689557...              13.0
4      https://twitter.com/dog_rates/status/891327558...              12.0
5      https://twitter.com/dog_rates/status/891087950...              13.0
9      https://twitter.com/dog_rates/status/890729181...              13.0
10     https://twitter.com/dog_rates/status/890609185...              13.0
11     https://twitter.com/dog_rates/status/890240255...              14.0
12     https://twitter.com/dog_rates/status/890006608...              13.0
13     https://twitter.com/dog_rates/status/889880896...              13.0
14     https://twitter.com/dog_rates/status/889665388...              13.0
15     https://twitter.com/dog_rates/status/889638837...              12.0
16     https://twitter.com/dog_rates/status/889531135...              13.0
17     https://twitter.com/dog_rates/status/889278841...              13.0
18     https://twitter.com/dog_rates/status/888917238...              12.0
19     https://twitter.com/dog_rates/status/888804989...              13.0
20     https://twitter.com/dog_rates/status/888554962...              13.0
22     https://twitter.com/dog_rates/status/888078434...              12.0
23     https://twitter.com/dog_rates/status/887705289...              13.0
24     https://twitter.com/dog_rates/status/887517139...              14.0
25     https://twitter.com/dog_rates/status/887473957...              13.0
26     https://twitter.com/dog_rates/status/887343217...              13.0
27     https://twitter.com/dog_rates/status/887101392...              12.0
28     https://twitter.com/dog_rates/status/886983233...              13.0
32     https://twitter.com/dog_rates/status/886680336...              13.0
33     https://twitter.com/dog_rates/status/886366144...              12.0
34                                                NaN              12.0
35     https://twitter.com/dog_rates/status/886258384...              13.0
37     https://twitter.com/dog_rates/status/885984800...              12.0
...                                                ...               ...
2487   https://twitter.com/dog_rates/status/666411507...               2.0
2488   https://twitter.com/dog_rates/status/666407126...               7.0
2489   https://twitter.com/dog_rates/status/666396247...               9.0
2490   https://twitter.com/dog_rates/status/666373753...              11.0
2491   https://twitter.com/dog_rates/status/666362758...               6.0
2492   https://twitter.com/dog_rates/status/666353288...               8.0
```

```
2493  https://twitter.com/dog_rates/status/666345417...                10.0
2494  https://twitter.com/dog_rates/status/666337882...                 9.0
2495  https://twitter.com/dog_rates/status/666293911...                 3.0
2496  https://twitter.com/dog_rates/status/666287406...                 1.0
2497  https://twitter.com/dog_rates/status/666273097...                11.0
2498  https://twitter.com/dog_rates/status/666268910...                10.0
2499  https://twitter.com/dog_rates/status/666104133...                 1.0
2500  https://twitter.com/dog_rates/status/666102155...                11.0
2501  https://twitter.com/dog_rates/status/666099513...                 8.0
2502  https://twitter.com/dog_rates/status/666094000...                 9.0
2503  https://twitter.com/dog_rates/status/666082916...                 6.0
2504  https://twitter.com/dog_rates/status/666073100...                10.0
2505  https://twitter.com/dog_rates/status/666071193...                 9.0
2506  https://twitter.com/dog_rates/status/666063827...                10.0
2507  https://twitter.com/dog_rates/status/666058600...                 8.0
2508  https://twitter.com/dog_rates/status/666057090...                 9.0
2509  https://twitter.com/dog_rates/status/666055525...                10.0
2510  https://twitter.com/dog_rates/status/666051853...                 2.0
2511  https://twitter.com/dog_rates/status/666050758...                10.0
2512  https://twitter.com/dog_rates/status/666049248...                 5.0
2513  https://twitter.com/dog_rates/status/666044226...                 6.0
2514  https://twitter.com/dog_rates/status/666033412...                 9.0
2515  https://twitter.com/dog_rates/status/666029285...                 7.0
2516  https://twitter.com/dog_rates/status/666020888...                 8.0

      rating_denominator      name  doggo floofer  pupper  puppo  rating_ratio
0                   10.0   Phineas   None    None    None   None           1.3
1                   10.0     Tilly   None    None    None   None           1.3
2                   10.0    Archie   None    None    None   None           1.2
3                   10.0     Darla   None    None    None   None           1.3
4                   10.0  Franklin   None    None    None   None           1.2
5                   10.0       NaN   None    None    None   None           1.3
9                   10.0       NaN   None    None    None   None           1.3
10                  10.0      Zoey   None    None    None   None           1.3
11                  10.0    Cassie  doggo    None    None   None           1.4
12                  10.0      Koda   None    None    None   None           1.3
13                  10.0     Bruno   None    None    None   None           1.3
14                  10.0       NaN   None    None    None  puppo           1.3
15                  10.0       Ted   None    None    None   None           1.2
16                  10.0    Stuart   None    None    None  puppo           1.3
17                  10.0    Oliver   None    None    None   None           1.3
18                  10.0       Jim   None    None    None   None           1.2
19                  10.0      Zeke   None    None    None   None           1.3
20                  10.0   Ralphus   None    None    None   None           1.3
22                  10.0    Gerald   None    None    None   None           1.2
23                  10.0   Jeffrey   None    None    None   None           1.3
24                  10.0      such   None    None    None   None           1.4
25                  10.0    Canela   None    None    None   None           1.3
```

13

```
26             10.0     NaN    None    None    None    None          1.3
27             10.0     NaN    None    None    None    None          1.2
28             10.0    Maya    None    None    None    None          1.3
32             10.0   Derek    None    None    None    None          1.3
33             10.0  Roscoe    None    None  pupper    None          1.2
34             10.0     NaN    None    None    None    None          1.2
35             10.0 Waffles    None    None    None    None          1.3
37             10.0   Jimbo    None    None    None    None          1.2
...             ...     ...     ...     ...     ...     ...          ...
2487           10.0   quite    None    None    None    None          0.2
2488           10.0     NaN    None    None    None    None          0.7
2489           10.0     NaN    None    None    None    None          0.9
2490           10.0     NaN    None    None    None    None          1.1
2491           10.0     NaN    None    None    None    None          0.6
2492           10.0     NaN    None    None    None    None          0.8
2493           10.0     NaN    None    None    None    None          1.0
2494           10.0     NaN    None    None    None    None          0.9
2495           10.0     NaN    None    None    None    None          0.3
2496            2.0     NaN    None    None    None    None          0.5
2497           10.0     NaN    None    None    None    None          1.1
2498           10.0     NaN    None    None    None    None          1.0
2499           10.0     NaN    None    None    None    None          0.1
2500           10.0     NaN    None    None    None    None          1.1
2501           10.0     NaN    None    None    None    None          0.8
2502           10.0     NaN    None    None    None    None          0.9
2503           10.0     NaN    None    None    None    None          0.6
2504           10.0     NaN    None    None    None    None          1.0
2505           10.0     NaN    None    None    None    None          0.9
2506           10.0     NaN    None    None    None    None          1.0
2507           10.0     NaN    None    None    None    None          0.8
2508           10.0     NaN    None    None    None    None          0.9
2509           10.0     NaN    None    None    None    None          1.0
2510           10.0     NaN    None    None    None    None          0.2
2511           10.0     NaN    None    None    None    None          1.0
2512           10.0     NaN    None    None    None    None          0.5
2513           10.0     NaN    None    None    None    None          0.6
2514           10.0     NaN    None    None    None    None          0.9
2515           10.0     NaN    None    None    None    None          0.7
2516           10.0     NaN    None    None    None    None          0.8

[2087 rows x 15 columns]
```

5.convert the timestamp to datetime datatype in twitter_archive

```
In [12]: twitter_archive_clean.timestamp=pd.to_datetime(twitter_archive_clean.timestamp)
```

6.Dropping the unnecessary column in the twitter archive.

14

```
In [40]: twitter_archive_clean.drop(['in_reply_to_status_id', 'in_reply_to_user_id'], axis=1, in
         print(twitter_archive_clean.columns)


         ---------------------------------------------------------------------------

         ValueError                                Traceback (most recent call last)

         <ipython-input-40-e0a2129a5678> in <module>()
    ----> 1 twitter_archive_clean.drop(['in_reply_to_status_id', 'in_reply_to_user_id'], axis=1,
            2 print(twitter_archive_clean.columns)


         /opt/conda/lib/python3.6/site-packages/pandas/core/generic.py in drop(self, labels, axis
         2159                  new_axis = axis.drop(labels, level=level, errors=errors)
         2160             else:
      -> 2161                  new_axis = axis.drop(labels, errors=errors)
         2162             dropped = self.reindex(**{axis_name: new_axis})
         2163             try:


         /opt/conda/lib/python3.6/site-packages/pandas/core/indexes/base.py in drop(self, labels,
         3622             if errors != 'ignore':
         3623                 raise ValueError('labels %s not contained in axis' %
      -> 3624                                  labels[mask])
         3625             indexer = indexer[~mask]
         3626         return self.delete(indexer)


         ValueError: labels ['in_reply_to_status_id' 'in_reply_to_user_id'] not contained in axis
```

7.Remove the unnecessary part of string in the source column of twitter archive

```
In [14]: twitter_archive_clean.source.replace({'<a href="http://twitter.com/download/iphone" rel

Out[14]: 0        Twitter for iPhone</a>
         1        Twitter for iPhone</a>
         2        Twitter for iPhone</a>
         3        Twitter for iPhone</a>
         4        Twitter for iPhone</a>
         5        Twitter for iPhone</a>
         9        Twitter for iPhone</a>
         10       Twitter for iPhone</a>
         11       Twitter for iPhone</a>
         12       Twitter for iPhone</a>
         13       Twitter for iPhone</a>
         14       Twitter for iPhone</a>
         15       Twitter for iPhone</a>
```

```
16      Twitter for iPhone</a>
17      Twitter for iPhone</a>
18      Twitter for iPhone</a>
19      Twitter for iPhone</a>
20      Twitter for iPhone</a>
22      Twitter for iPhone</a>
23      Twitter for iPhone</a>
24      Twitter for iPhone</a>
25      Twitter for iPhone</a>
26      Twitter for iPhone</a>
27      Twitter for iPhone</a>
28      Twitter for iPhone</a>
32      Twitter for iPhone</a>
33      Twitter for iPhone</a>
34      Twitter for iPhone</a>
35      Twitter for iPhone</a>
37      Twitter for iPhone</a>
                    ...
2487    Twitter for iPhone</a>
2488    Twitter for iPhone</a>
2489    Twitter for iPhone</a>
2490    Twitter for iPhone</a>
2491    Twitter for iPhone</a>
2492    Twitter for iPhone</a>
2493    Twitter for iPhone</a>
2494    Twitter for iPhone</a>
2495    Twitter for iPhone</a>
2496    Twitter for iPhone</a>
2497    Twitter for iPhone</a>
2498    Twitter for iPhone</a>
2499    Twitter for iPhone</a>
2500    Twitter for iPhone</a>
2501    Twitter for iPhone</a>
2502    Twitter for iPhone</a>
2503    Twitter for iPhone</a>
2504    Twitter for iPhone</a>
2505    Twitter for iPhone</a>
2506    Twitter for iPhone</a>
2507    Twitter for iPhone</a>
2508    Twitter for iPhone</a>
2509    Twitter for iPhone</a>
2510    Twitter for iPhone</a>
2511    Twitter for iPhone</a>
2512    Twitter for iPhone</a>
2513    Twitter for iPhone</a>
2514    Twitter for iPhone</a>
2515    Twitter for iPhone</a>
2516    Twitter for iPhone</a>
```

```
        Name: source, Length: 2087, dtype: object
```

8.Multiple stages mentioned for a single dog.Finding the correct stage from text and replacing
it

```
In [15]: twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='817777686764523521']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='817777686764523521',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='855851453814013952']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='855851453814013952',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='808106460588765185']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='808106460588765185',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='802265048156610565']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='802265048156610565',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='801115127852503040']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='801115127852503040',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='785639753186217984']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='785639753186217984',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='781308096455073793']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='781308096455073793',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='759793422261743616']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='759793422261743616',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='751583847268179968']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='751583847268179968',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='751583847268179968',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='741067306818797568']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='741067306818797568',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='741067306818797568',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='733109485275860992']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='733109485275860992',['pupper
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='733109485275860992',['doggo'
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='854010172552949760']
         twitter_archive_clean.loc[twitter_archive_clean.tweet_id=='854010172552949760',['floofe
```

4   TIDINESS

1.Merging of the twitter_archive_clean and status dataframe so that variable favorites and tweets
can be used in analysis effectively.

```
In [16]: status.tweet_id=status.tweet_id.astype(str)
         twitter_archive_clean=pd.merge(twitter_archive_clean,status,on='tweet_id')
         print(twitter_archive_clean)

                 tweet_id            timestamp  \
0     892420643555336193 2017-08-01 16:23:56
1     892177421306343426 2017-08-01 00:17:27
2     891815181378084864 2017-07-31 00:18:03
3     891689557279858688 2017-07-30 15:58:51
4     891327558926688256 2017-07-29 16:00:24
```

```
5       8910879508758897856  2017-07-29 00:08:17
6       890729181411237888   2017-07-28 00:22:40
7       890609185150312448   2017-07-27 16:25:51
8       890240255349198849   2017-07-26 15:59:51
9       890006608113172480   2017-07-26 00:31:25
10      889880896479866881   2017-07-25 16:11:53
11      889665388333682689   2017-07-25 01:55:32
12      889638837579907072   2017-07-25 00:10:02
13      889531135344209921   2017-07-24 17:02:04
14      889278841981685760   2017-07-24 00:19:32
15      888917238123831296   2017-07-23 00:22:39
16      888804989199671297   2017-07-22 16:56:37
17      888554962724278272   2017-07-22 00:23:06
18      888078434458587136   2017-07-20 16:49:33
19      887705289381826560   2017-07-19 16:06:48
20      887517139158093824   2017-07-19 03:39:09
21      887473957103951883   2017-07-19 00:47:34
22      887343217045368832   2017-07-18 16:08:03
23      887101392804085760   2017-07-18 00:07:08
24      886983233522544640   2017-07-17 16:17:36
25      886680336477933568   2017-07-16 20:14:00
26      886366144734445568   2017-07-15 23:25:31
27      886267009285017600   2017-07-15 16:51:35
28      886258384151887873   2017-07-15 16:17:19
29      885984800019947520   2017-07-14 22:10:11
...                ...                     ...
2056    666411507551481857   2015-11-17 00:24:19
2057    666407126856765440   2015-11-17 00:06:54
2058    666396247373291520   2015-11-16 23:23:41
2059    666373753744588802   2015-11-16 21:54:18
2060    666362758909284353   2015-11-16 21:10:36
2061    666353288456101888   2015-11-16 20:32:58
2062    666345417576210432   2015-11-16 20:01:42
2063    666337882303524864   2015-11-16 19:31:45
2064    666293911632134144   2015-11-16 16:37:02
2065    666287406224695296   2015-11-16 16:11:11
2066    666273097616637952   2015-11-16 15:14:19
2067    666268910803644416   2015-11-16 14:57:41
2068    666104133288665088   2015-11-16 04:02:55
2069    666102155909144576   2015-11-16 03:55:04
2070    666099513787052032   2015-11-16 03:44:34
2071    666094000022159362   2015-11-16 03:22:39
2072    666082916733198337   2015-11-16 02:38:37
2073    666073100786774016   2015-11-16 01:59:36
2074    666071193221509120   2015-11-16 01:52:02
2075    666063827256086533   2015-11-16 01:22:45
2076    666058600524156928   2015-11-16 01:01:59
2077    666057090499244032   2015-11-16 00:55:59
```

```
2078   666055525042405380  2015-11-16 00:49:46
2079   666051853826850816  2015-11-16 00:35:11
2080   666050758794694657  2015-11-16 00:30:50
2081   666049248165822465  2015-11-16 00:24:50
2082   666044226329800704  2015-11-16 00:04:52
2083   666033412701032449  2015-11-15 23:21:54
2084   666029285002620928  2015-11-15 23:05:30
2085   666020888022790149  2015-11-15 22:32:08


                                                   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...
...                                                  ...
2056   <a href="http://twitter.com/download/iphone" r...
2057   <a href="http://twitter.com/download/iphone" r...
2058   <a href="http://twitter.com/download/iphone" r...
2059   <a href="http://twitter.com/download/iphone" r...
2060   <a href="http://twitter.com/download/iphone" r...
2061   <a href="http://twitter.com/download/iphone" r...
2062   <a href="http://twitter.com/download/iphone" r...
```

```
2063    <a href="http://twitter.com/download/iphone" r...
2064    <a href="http://twitter.com/download/iphone" r...
2065    <a href="http://twitter.com/download/iphone" r...
2066    <a href="http://twitter.com/download/iphone" r...
2067    <a href="http://twitter.com/download/iphone" r...
2068    <a href="http://twitter.com/download/iphone" r...
2069    <a href="http://twitter.com/download/iphone" r...
2070    <a href="http://twitter.com/download/iphone" r...
2071    <a href="http://twitter.com/download/iphone" r...
2072    <a href="http://twitter.com/download/iphone" r...
2073    <a href="http://twitter.com/download/iphone" r...
2074    <a href="http://twitter.com/download/iphone" r...
2075    <a href="http://twitter.com/download/iphone" r...
2076    <a href="http://twitter.com/download/iphone" r...
2077    <a href="http://twitter.com/download/iphone" r...
2078    <a href="http://twitter.com/download/iphone" r...
2079    <a href="http://twitter.com/download/iphone" r...
2080    <a href="http://twitter.com/download/iphone" r...
2081    <a href="http://twitter.com/download/iphone" r...
2082    <a href="http://twitter.com/download/iphone" r...
2083    <a href="http://twitter.com/download/iphone" r...
2084    <a href="http://twitter.com/download/iphone" r...
2085    <a href="http://twitter.com/download/iphone" r...


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

```
23      This... is a Jubilant Antarctic House Bear. We...
24      This is Maya. She's very shy. Rarely leaves he...
25      This is Derek. He's late for a dog meeting. 13...
26      This is Roscoe. Another pupper fallen victim t...
27      @NonWhiteHat @MayhewMayhem omg hello tanner yo...
28      This is Waffles. His doggles are pupside down...
29      Viewer discretion advised. This is Jimbo. He w...
...                                                  ...
2056    This is quite the dog. Gets really excited whe...
2057    This is a southern Vesuvius bumblegruff. Can d...
2058    Oh goodness. A super rare northeast Qdoba kang...
2059    Those are sunglasses and a jean jacket. 11/10 ...
2060    Unique dog here. Very small. Lives in containe...
2061    Here we have a mixed Asiago from the Gal?ípago...
2062    Look at this jokester thinking seat belt laws ...
2063    This is an extremely rare horned Parthenon. No...
2064    This is a funny dog. Weird toes. Won't come do...
2065    This is an Albanian 3 1/2 legged  Episcopalian...
2066        Can take selfies 11/10 https://t.co/ws2AMaNwPW
2067    Very concerned about fellow dog trapped in com...
2068    Not familiar with this breed. No tail (weird)...
2069    Oh my. Here you are seeing an Adobe Setter giv...
2070    Can stand on stump for what seems like a while...
2071    This appears to be a Mongolian Presbyterian mi...
2072    Here we have a well-established sunblockerspan...
2073    Let's hope this flight isn't Malaysian (lol). ...
2074    Here we have a northern speckled Rhododendron...
2075    This is the happiest dog you will ever see. Ve...
2076    Here is the Rand Paul of retrievers folks! He'...
2077    My oh my. This is a rare blond Canadian terrie...
2078    Here is a Siberian heavily armored polar bear ...
2079    This is an odd dog. Hard on the outside but lo...
2080    This is a truly beautiful English Wilson Staff...
2081    Here we have a 1949 1st generation vulpix. Enj...
2082    This is a purebred Piers Morgan. Loves to Netf...
2083    Here is a very happy pup. Big fan of well-main...
2084    This is a western brown Mitsubishi terrier. Up...
2085    Here we have a Japanese Irish Setter. Lost eye...


                                  expanded_urls  rating_numerator  \
0     https://twitter.com/dog_rates/status/892420643...              13.0
1     https://twitter.com/dog_rates/status/892177421...              13.0
2     https://twitter.com/dog_rates/status/891815181...              12.0
3     https://twitter.com/dog_rates/status/891689557...              13.0
4     https://twitter.com/dog_rates/status/891327558...              12.0
5     https://twitter.com/dog_rates/status/891087950...              13.0
6     https://twitter.com/dog_rates/status/890729181...              13.0
7     https://twitter.com/dog_rates/status/890609185...              13.0
```

```
8      https://twitter.com/dog_rates/status/890240255...                14.0
9      https://twitter.com/dog_rates/status/890006608...                13.0
10     https://twitter.com/dog_rates/status/889880896...                13.0
11     https://twitter.com/dog_rates/status/889665388...                13.0
12     https://twitter.com/dog_rates/status/889638837...                12.0
13     https://twitter.com/dog_rates/status/889531135...                13.0
14     https://twitter.com/dog_rates/status/889278841...                13.0
15     https://twitter.com/dog_rates/status/888917238...                12.0
16     https://twitter.com/dog_rates/status/888804989...                13.0
17     https://twitter.com/dog_rates/status/888554962...                13.0
18     https://twitter.com/dog_rates/status/888078434...                12.0
19     https://twitter.com/dog_rates/status/887705289...                13.0
20     https://twitter.com/dog_rates/status/887517139...                14.0
21     https://twitter.com/dog_rates/status/887473957...                13.0
22     https://twitter.com/dog_rates/status/887343217...                13.0
23     https://twitter.com/dog_rates/status/887101392...                12.0
24     https://twitter.com/dog_rates/status/886983233...                13.0
25     https://twitter.com/dog_rates/status/886680336...                13.0
26     https://twitter.com/dog_rates/status/886366144...                12.0
27                                                  NaN                 12.0
28     https://twitter.com/dog_rates/status/886258384...                13.0
29     https://twitter.com/dog_rates/status/885984800...                12.0
...                                                  ...                 ...
2056   https://twitter.com/dog_rates/status/666411507...                 2.0
2057   https://twitter.com/dog_rates/status/666407126...                 7.0
2058   https://twitter.com/dog_rates/status/666396247...                 9.0
2059   https://twitter.com/dog_rates/status/666373753...                11.0
2060   https://twitter.com/dog_rates/status/666362758...                 6.0
2061   https://twitter.com/dog_rates/status/666353288...                 8.0
2062   https://twitter.com/dog_rates/status/666345417...                10.0
2063   https://twitter.com/dog_rates/status/666337882...                 9.0
2064   https://twitter.com/dog_rates/status/666293911...                 3.0
2065   https://twitter.com/dog_rates/status/666287406...                 1.0
2066   https://twitter.com/dog_rates/status/666273097...                11.0
2067   https://twitter.com/dog_rates/status/666268910...                10.0
2068   https://twitter.com/dog_rates/status/666104133...                 1.0
2069   https://twitter.com/dog_rates/status/666102155...                11.0
2070   https://twitter.com/dog_rates/status/666099513...                 8.0
2071   https://twitter.com/dog_rates/status/666094000...                 9.0
2072   https://twitter.com/dog_rates/status/666082916...                 6.0
2073   https://twitter.com/dog_rates/status/666073100...                10.0
2074   https://twitter.com/dog_rates/status/666071193...                 9.0
2075   https://twitter.com/dog_rates/status/666063827...                10.0
2076   https://twitter.com/dog_rates/status/666058600...                 8.0
2077   https://twitter.com/dog_rates/status/666057090...                 9.0
2078   https://twitter.com/dog_rates/status/666055525...                10.0
2079   https://twitter.com/dog_rates/status/666051853...                 2.0
2080   https://twitter.com/dog_rates/status/666050758...                10.0
```

```
2081   https://twitter.com/dog_rates/status/666049248...          5.0
2082   https://twitter.com/dog_rates/status/666044226...          6.0
2083   https://twitter.com/dog_rates/status/666033412...          9.0
2084   https://twitter.com/dog_rates/status/666029285...          7.0
2085   https://twitter.com/dog_rates/status/666020888...          8.0

       rating_denominator      name   doggo floofer  pupper   puppo  \
0                    10.0   Phineas    None    None    None    None
1                    10.0     Tilly    None    None    None    None
2                    10.0    Archie    None    None    None    None
3                    10.0     Darla    None    None    None    None
4                    10.0  Franklin    None    None    None    None
5                    10.0       NaN    None    None    None    None
6                    10.0       NaN    None    None    None    None
7                    10.0      Zoey    None    None    None    None
8                    10.0    Cassie   doggo    None    None    None
9                    10.0      Koda    None    None    None    None
10                   10.0     Bruno    None    None    None    None
11                   10.0       NaN    None    None    None   puppo
12                   10.0       Ted    None    None    None    None
13                   10.0    Stuart    None    None    None   puppo
14                   10.0    Oliver    None    None    None    None
15                   10.0       Jim    None    None    None    None
16                   10.0      Zeke    None    None    None    None
17                   10.0   Ralphus    None    None    None    None
18                   10.0    Gerald    None    None    None    None
19                   10.0   Jeffrey    None    None    None    None
20                   10.0      such    None    None    None    None
21                   10.0    Canela    None    None    None    None
22                   10.0       NaN    None    None    None    None
23                   10.0       NaN    None    None    None    None
24                   10.0      Maya    None    None    None    None
25                   10.0     Derek    None    None    None    None
26                   10.0    Roscoe    None    None  pupper    None
27                   10.0       NaN    None    None    None    None
28                   10.0   Waffles    None    None    None    None
29                   10.0     Jimbo    None    None    None    None
...                   ...       ...     ...     ...     ...     ...
2056                 10.0     quite    None    None    None    None
2057                 10.0       NaN    None    None    None    None
2058                 10.0       NaN    None    None    None    None
2059                 10.0       NaN    None    None    None    None
2060                 10.0       NaN    None    None    None    None
2061                 10.0       NaN    None    None    None    None
2062                 10.0       NaN    None    None    None    None
2063                 10.0       NaN    None    None    None    None
2064                 10.0       NaN    None    None    None    None
2065                  2.0       NaN    None    None    None    None
```

```
2066              10.0     NaN   None    None    None    None
2067              10.0     NaN   None    None    None    None
2068              10.0     NaN   None    None    None    None
2069              10.0     NaN   None    None    None    None
2070              10.0     NaN   None    None    None    None
2071              10.0     NaN   None    None    None    None
2072              10.0     NaN   None    None    None    None
2073              10.0     NaN   None    None    None    None
2074              10.0     NaN   None    None    None    None
2075              10.0     NaN   None    None    None    None
2076              10.0     NaN   None    None    None    None
2077              10.0     NaN   None    None    None    None
2078              10.0     NaN   None    None    None    None
2079              10.0     NaN   None    None    None    None
2080              10.0     NaN   None    None    None    None
2081              10.0     NaN   None    None    None    None
2082              10.0     NaN   None    None    None    None
2083              10.0     NaN   None    None    None    None
2084              10.0     NaN   None    None    None    None
2085              10.0     NaN   None    None    None    None

     rating_ratio   favorites   retweets
0             1.3       38298       8404
1             1.3       32826       6196
2             1.2       24719       4100
3             1.3       41631       8531
4             1.2       39795       9249
5             1.3       19984       3076
6             1.3       64645      18636
7             1.3       27474       4215
8             1.4       31523       7294
9             1.3       30284       7242
10            1.3       27441       4914
11            1.3       47520       9933
12            1.2       26837       4487
13            1.3       14905       2212
14            1.3       24943       5315
15            1.2       28723       4443
16            1.3       25243       4245
17            1.3       19605       3508
18            1.2       21490       3451
19            1.3       29795       5322
20            1.4       45693      11520
21            1.3       68258      17961
22            1.3       33260      10278
23            1.2       30179       5889
24            1.3       34728       7679
25            1.3       22151       4408
```

```
26               1.2        20927        3162
27               1.2          116           4
28               1.3        27649        6208
29               1.2        32288        6722
...              ...          ...         ...
2056             0.2          440         324
2057             0.7          107          40
2058             0.9          167          84
2059             1.1          186          92
2060             0.6          772         565
2061             0.8          219          71
2062             1.0          291         133
2063             0.9          197          90
2064             0.3          495         352
2065             0.5          146          64
2066             1.1          174          76
2067             1.0          102          35
2068             0.1        14195        6510
2069             1.1           80          12
2070             0.8          153          67
2071             0.9          163          73
2072             0.6          116          44
2073             1.0          319         163
2074             0.9          145          59
2075             1.0          471         216
2076             0.8          111          57
2077             0.9          294         141
2078             1.0          433         246
2079             0.2         1208         843
2080             1.0          132          59
2081             0.5          107          40
2082             0.6          295         139
2083             0.9          125          44
2084             0.7          128          47
2085             0.8         2551         508

[2086 rows x 15 columns]
```

2.merging of four columns'doggo','floofer','pupper' and 'puppo' in the twitter archive in to single column 'stage' to reduce redundancy of display of data .

```
In [17]: twitter_archive_clean.doggo=twitter_archive_clean.doggo.replace('None','')
         twitter_archive_clean.floofer=twitter_archive_clean.floofer.replace('None','')
         twitter_archive_clean.pupper=twitter_archive_clean.pupper.replace('None','')
         twitter_archive_clean.puppo=twitter_archive_clean.puppo.replace('None','')
         twitter_archive_clean['stage']=(twitter_archive_clean.doggo).str.cat(twitter_archive_cl
         twitter_archive_clean.stage.value_counts()
```

```
twitter_archive_clean.drop(['doggo','floofer','pupper','puppo'], axis=1, inplace=True)
twitter_archive_clean.stage=twitter_archive_clean.stage.replace('',np.nan)
```

3.Creation of new column "Prediction" from details of other columns p1,p1_dog,p2,p2_dog,p3,p3_dog to simplify analysis of data.

```
In [18]: image_prediction['predicted']=image_prediction.p1
         image_prediction.predicted[image_prediction.p1_dog==False]=image_prediction.p2
         image_prediction.predicted[(image_prediction.p1_dog==False)&(image_prediction.p2_dog==F
         image_prediction.predicted[(image_prediction.p1_dog==False)&(image_prediction.p2_dog==F
         image_prediction.predicted=image_prediction.predicted.replace('None',np.nan)
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:2: 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:3: 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 separate from the ipykernel package so we can avoid doing imports until
/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.
```

```
In [ ]: print(image_prediction)
```

5   DATA STORING:

```
In [19]: twitter_archive_clean.to_csv('twitter_archive_master.csv', index=False)
         image_prediction.to_csv('image_predictions_CLEAN.csv', index=False)
```

6   ANALYSIS AND VISUALISATION

```
In [20]: twitter_archive_clean.describe()
```

```
Out[20]:        rating_numerator  rating_denominator  rating_ratio      favorites  \
        count        2086.000000         2086.000000   2086.000000    2086.000000
        mean           11.127517           10.500959      1.057113    8759.275168
        std             8.416100            7.110080      0.222591   12817.506128
        min             0.000000            2.000000      0.000000      51.000000
        25%            10.000000           10.000000      1.000000    1863.500000
        50%            11.000000           10.000000      1.100000    3891.500000
        75%            12.000000           10.000000      1.200000   10999.500000
```

26

```
max            204.000000        170.000000     1.700000   165074.000000

               retweets
count     2086.000000
mean      2712.573346
std       4791.244160
min          0.000000
25%        583.250000
50%       1283.500000
75%       3105.500000
max      84382.000000
```

Average rating_ratio is 1.05 and maximum is observed around 1.7.Average favorites is found to be 8760 and maximum favorites is found to be 165074.Average retweets is found to be 2712 and maximum retweets is found to be 84382 .

```
In [21]: twitter_archive_clean=twitter_archive_clean.round({'rating_ratio': 1})
         print(twitter_archive_clean.columns)
         values=np.sort(twitter_archive_clean.rating_ratio.unique())
         values

Index(['tweet_id', 'timestamp', 'source', 'text', 'expanded_urls',
       'rating_numerator', 'rating_denominator', 'name', 'rating_ratio',
       'favorites', 'retweets', 'stage'],
     dtype='object')


Out[21]: array([ 0. ,  0.1,  0.2,  0.3,  0.4,  0.5,  0.6,  0.7,  0.8,  0.9,  1. ,
                 1.1,  1.2,  1.3,  1.4,  1.5,  1.7])
```

7   DOG STAGE ANALYSIS

Below show the distribution of the dogs stage.We can see of the stages described in the tweets ,Majority cases belong to pupper stage.It is followed by doggo with second highest count.While floofer is least mentioned stage in tweet.

```
In [22]: dogstage=twitter_archive_clean[['tweet_id','rating_ratio','favorites','retweets','stage
         dogstage['stage'].value_counts().plot(kind='pie')
         dogstage['stage'].describe()

Out[22]: count        324
         unique         4
         top       pupper
         freq         218
         Name: stage, dtype: object
```
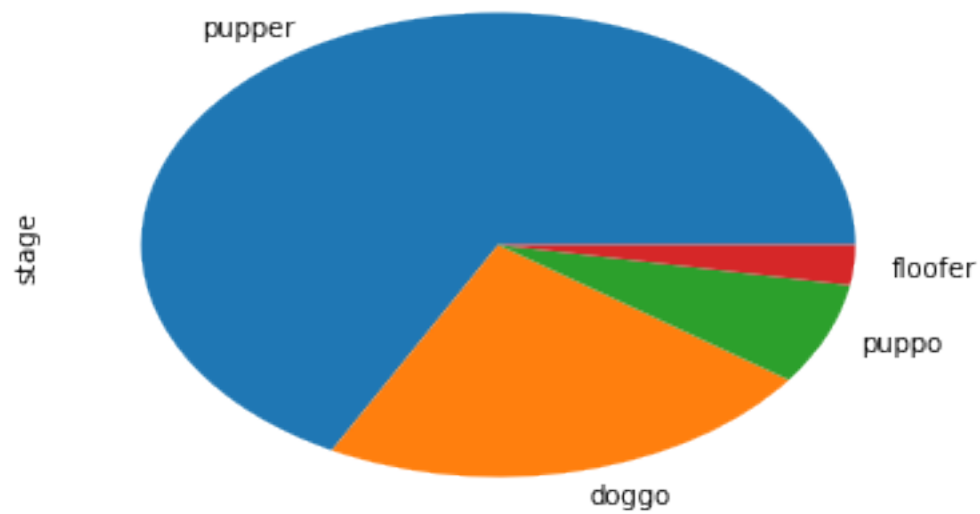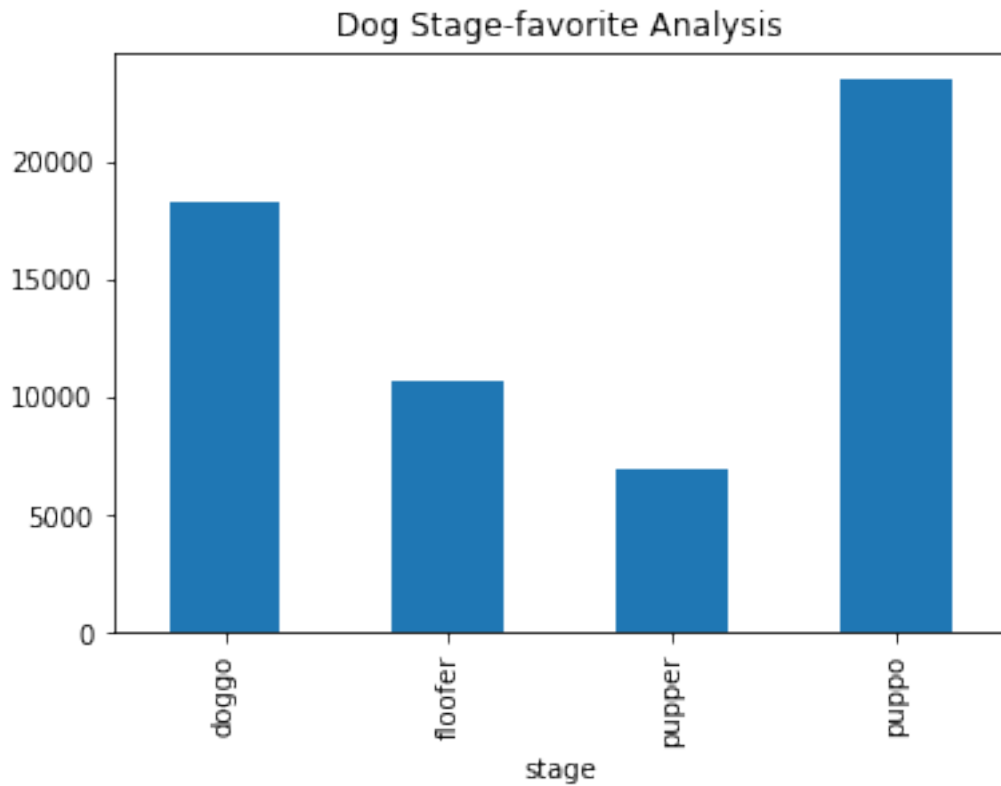
```
In [23]: dogstage.groupby('stage')['retweets'].mean().plot(title='Dog Stage-Retweet Analysis',ki
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2ce0e714e0>
```
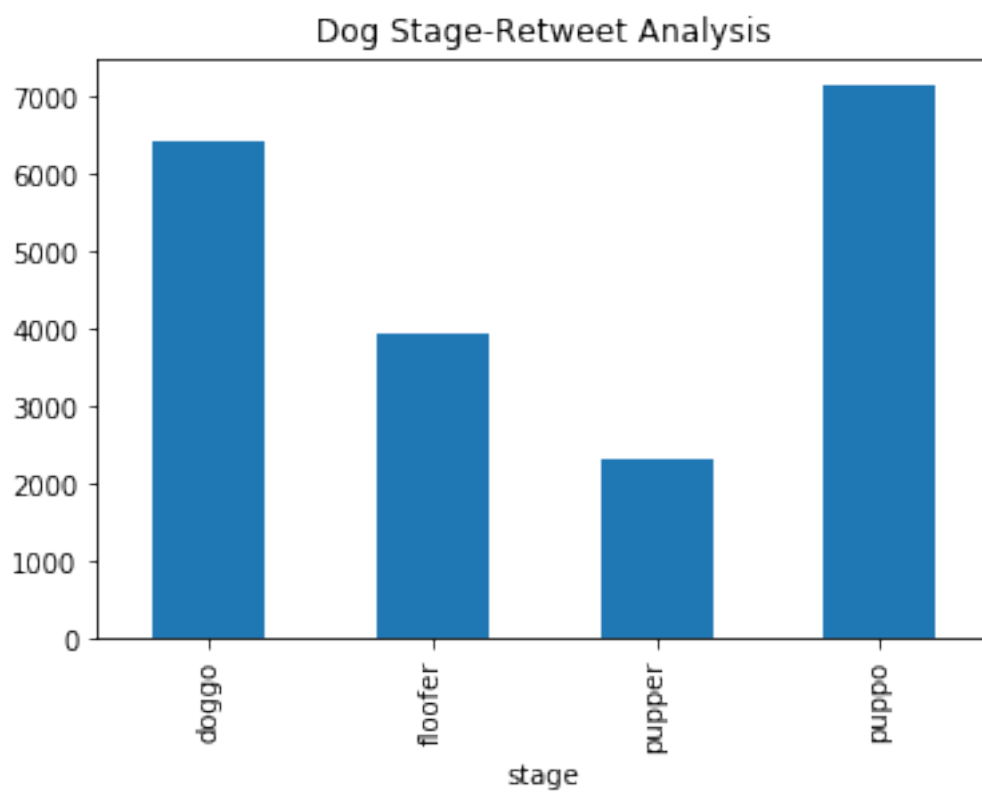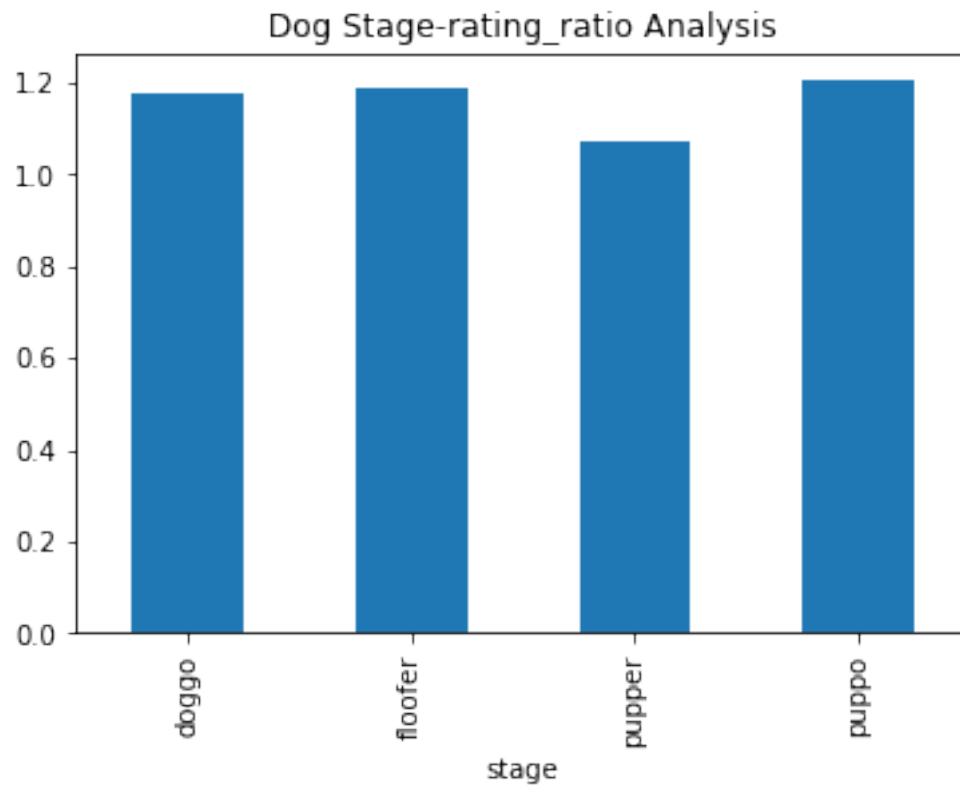
```
In [38]: dogstage.groupby('stage')['favorites'].mean().plot(title='Dog Stage-favorite Analysis',
```

```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2cd9e5dd30>
```


Dog Stage-favorite Analysis

```
In [27]: dogstage.groupby('stage')['rating_ratio'].mean().plot(title='Dog Stage-rating_ratio Ana
```

```
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2cdeb48f28>
```

Dog Stage-rating_ratio Analysis

8   RATING ANALYSIS:

Below shows the Distribution of rating ratio of tweets.

```
In [36]: twitter_archive_clean.rating_ratio.value_counts().plot(title='Rating Distribution',kind
```

```
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2cdaffe668>
```

Rating Distribution
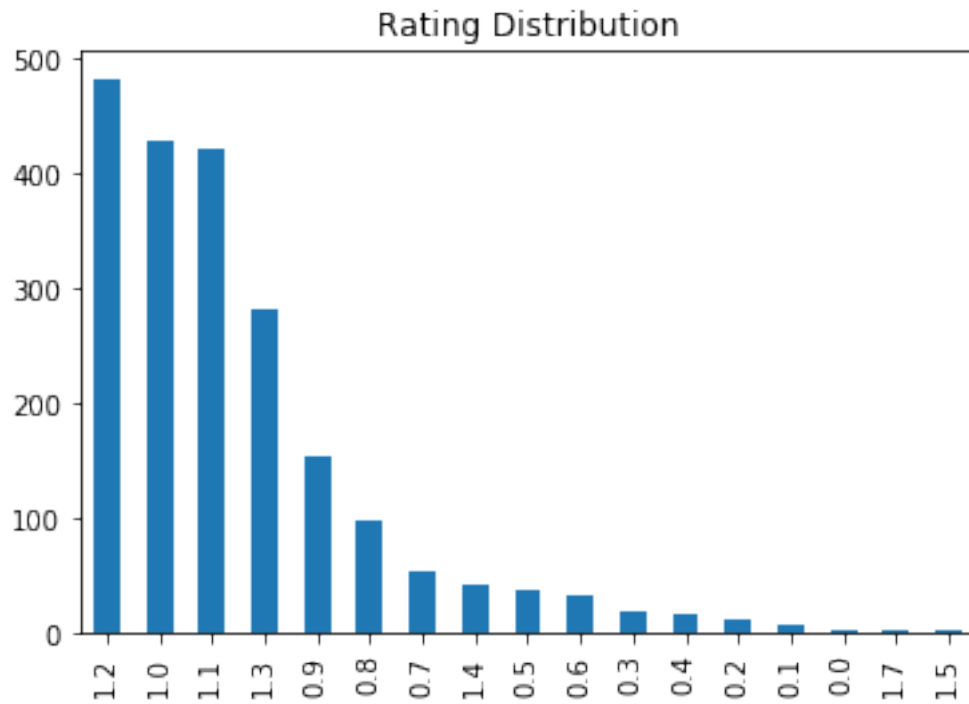
In [ ]: Results shows that more rating are in the range of 1 to 1.3.when ratings decreases ,freq

RATING VS TIME ANALYSIS

In [31]: `twitter_archive_clean.groupby('timestamp')['rating_ratio'].mean().plot(title='Time-rati`
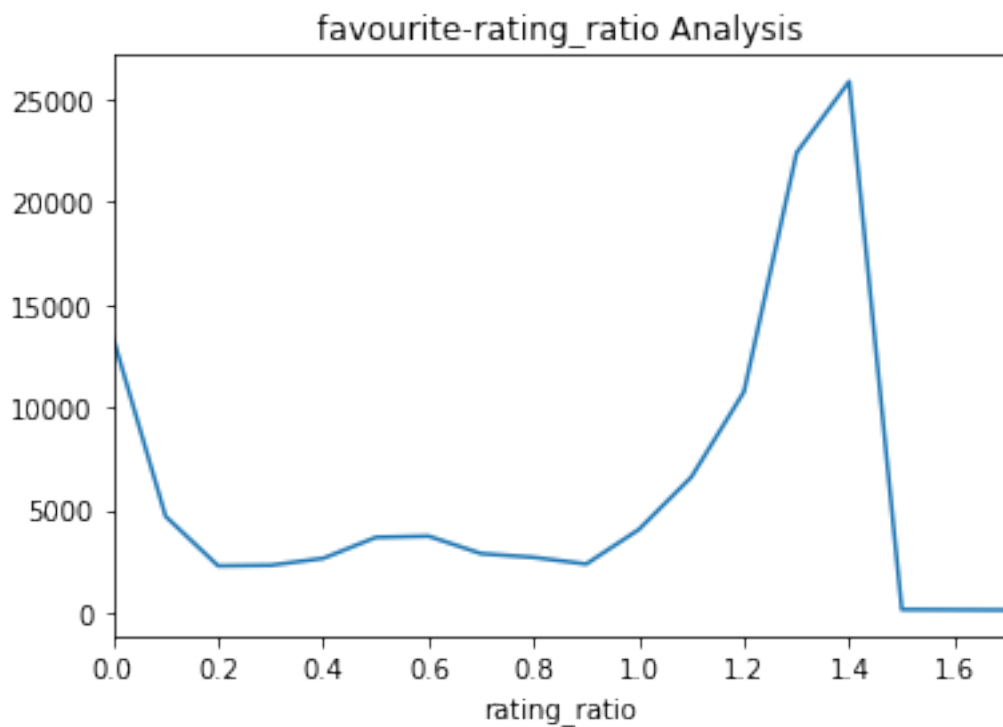
Out[31]: `<matplotlib.axes._subplots.AxesSubplot at 0x7f2cddc6f550>`

Time-rating_ratio Analysis
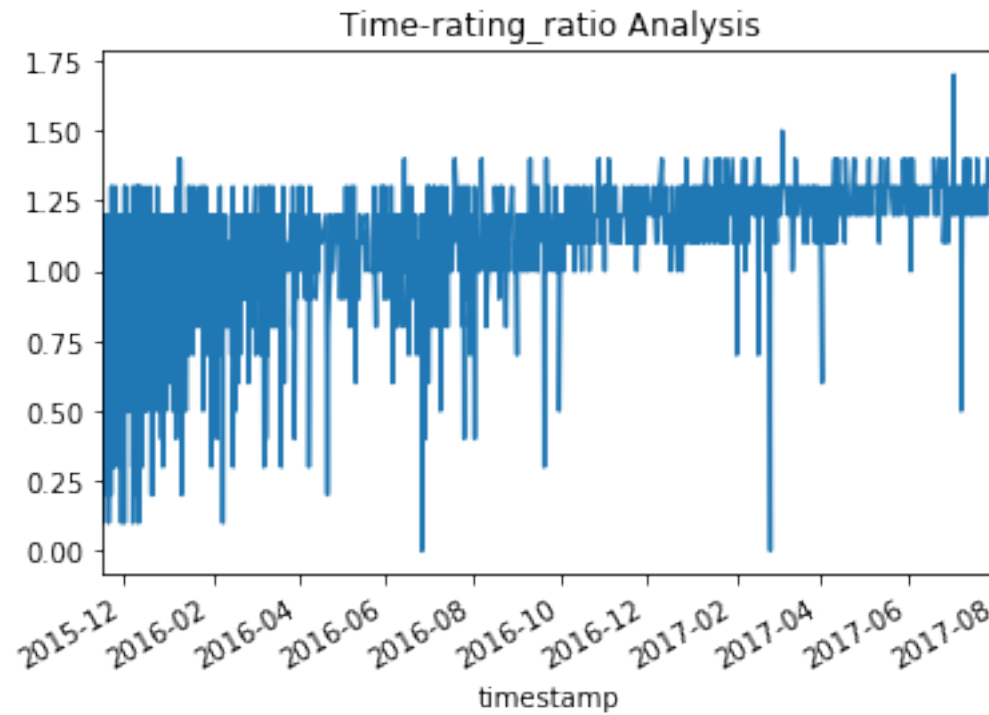
In [ ]: Results shows that rating gets improved over time from 2015 to 2017.

RATING VS FAVOURITE AND RETWEET ANALYSIS:

In [32]: twitter_archive_clean.groupby('rating_ratio')['favorites'].mean().plot(title='favourite

Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2cdca9d828>

favourite-rating_ratio Analysis

In [33]: twitter_archive_clean.groupby('rating_ratio')['retweets'].mean().plot(title='retweets-r

Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2cdcb74f60>

retweets-rating_ratio Analysis
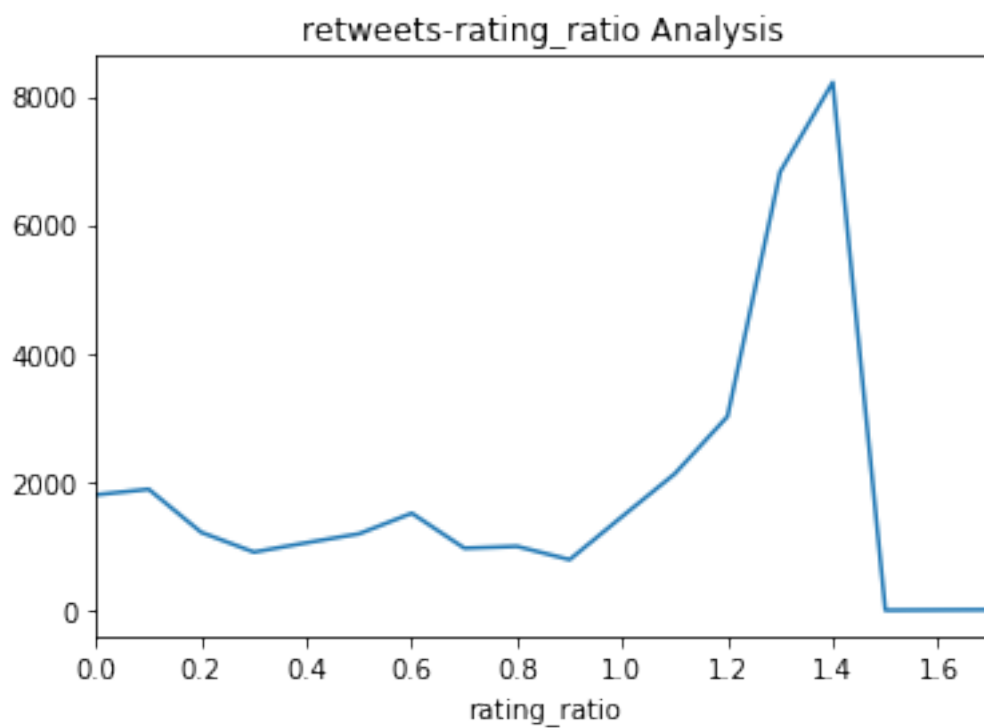
```
In [ ]: Higher the rating ,favourites and retweets are higher.
```

9 Insights:

1.Of all the stages of dog,Pupper has highest frequency stage mentioned in tweets .Doggo has the second highest frequency mentioned stage.While floofer is least mentioned stage in tweet. 2.Average rating Ratio is around 1.05.Most ratio of rating are between 1 to 1.3.Retweets and favorites are higher for posts with higher rating and retweets and favourites decreases when rating decreases. 3.Puppo has most favorites and retweets even though it is not frequently mentioned in tweets .Pupper has lowest favorites and retweets eventhough it is highest mentioned frequency dog stage. 4.Rating ratio increased with time .Intially there are more poor rating around 2015 which decreased in 2016 and further improved in 2017.