wrangle_act

January 10, 2019

1 Project: WeRateDogs Data Analysis

1.1 Introduction

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

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for you to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

```
In [4]: #import libraries
        import pandas as pd
        import numpy as np
        import seaborn as sb
        import matplotlib.pyplot as plt
        % matplotlib inline
```

```
In [5]: # Read and check the twitter archive
        twitter_archive = pd.read_csv('twitter-archive-enhanced.csv')
        twitter_archive.head(3)
```

```
Out[5]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0  892420643555336193                    NaN                  NaN
        1  892177421306343426                    NaN                  NaN
        2  891815181378084864                    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

                                                   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...

                                                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

   retweeted_status_user_id retweeted_status_timestamp  \
0                       NaN                        NaN
1                       NaN                        NaN
2                       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

   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
```

In [6]:
```
# Programmatically download the dog image prediction files from
# the Udacity server using Request library

# Reference : http://docs.python-requests.org/en/master/user/quickstart/#response-cont
# Reference : https://docs.python.org/3/library/os.path.html
# Reference : https://docs.python.org/3/library/os.html?highlight=makedirs#os.makedirs
# Reference : https://stat.ethz.ch/pipermail/r-devel/2012-August/064739.html
# Reference : https://docs.python.org/3/tutorial/inputoutput.html

# import the os library for accessing filesystem
import os
# import the request module which allows us to send organic, grass-fed HTTP/1.1 reques
import requests

# Save to a file called folder_name
folder_name = 'image-predictions'

# Make sure the file existed
# Check if the file is not existed which return True then it will execute Recursive di
# Like mkdir(), but makes all intermediate-level directories needed to contain the lea
if not os.path.exists(folder_name) :
    os.makedirs(folder_name)

url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predicti

# Get the webpage from url. Now , we have a response object called response
```

```
        response = requests.get(url)

        # Open() returns the file object which is
        # a file which Join folder_name and url path components intelligently  then
        # writes the contents of response to the file, returning the number of characters writ
        with open(os.path.join(folder_name , url.split('/')[-1]) , mode = 'wb') as file :
            file.write(response.content)
```

```
In [7]: #open tsv file
        images = pd.read_table('image-predictions/image-predictions.tsv',
                               sep='\t')
```

```
In [8]: images.head(2)
```

Out[8]:

	tweet_id	jpg_url	\
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg	
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	

	img_num	p1	p1_conf	p1_dog	p2	\
0	1	Welsh_springer_spaniel	0.465074	True	collie	
1	1	redbone	0.506826	True	miniature_pinscher	

	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	0.156665	True	Shetland_sheepdog	0.061428	True
1	0.074192	True	Rhodesian_ridgeback	0.072010	True

1.2 Retrieve tweets from Twitter API

Use Python's Tweepy Library to query Twitter API for each tweet's JSON data. Then, store each tweet's JSON data into a file

```
In [9]: # Ref : http://www.tweepy.org
        # Ref : https://developer.twitter.com/en/docs/basics/developer-portal/overview
        # Ref : https://stackoverflow.com/questions/28384588/twitter-api-get-tweets-with-speci
        import tweepy
        from tweepy import OAuthHandler
        import json
        import csv
        import sys
        import os
        import time

        # authentication
        consumer_key = ''
        consumer_secret = ''
        access_token = ''
        access_secret = ''

        auth = tweepy.OAuthHandler(consumer_key , consumer_secret)
```

```
        auth.set_access_token(access_token , access_secret)

        api = tweepy.API(auth_handler = auth ,
                         wait_on_rate_limit = True ,
                         wait_on_rate_limit_notify = True)
```

In [10]:
```
# So now we are ready to start the process.
# We create the list of ids that we gonna use to fetch the required tweets.
# We use the ids of the twitter_archive DataFrame:
list_of_ids_in_twittter_archive = twitter_archive['tweet_id'].tolist()
```

In [11]:
```
# We then fetch all tweets from the twitter API using the following loop:
list_of_tweets = []
# Tweets that can't be found are saved in the list below:
cant_find_tweets_for_those_ids = []
for each_id in list_of_ids_in_twittter_archive:
    try:
        start = time.time()
        list_of_tweets.append(api.get_status(each_id))
        end = time.time()
        print(end - start)
    except Exception as e:
        cant_find_tweets_for_those_ids.append(each_id)
        print(str(each_id))
```

```
0.44844794273376465
0.39110589027404785
0.40046215057373047
0.39612913131713867
0.3946690559387207
0.3797566890716553
0.42017483711242676
0.4222836494445801
0.4060170650482178
0.3942992687225342
0.40502095222473145
0.436643123626709
0.4070901870727539
0.3954949378967285
0.40032482147216797
0.4062190055847168
0.420853853225708
0.39242124557495117
0.41161489486694336
888202515573088257
0.4080069065093994
0.3999922275543213
0.393474817276001
```

0.4025869369506836
0.4124610424041748
0.40041112899780273
0.4188361167907715
0.42183876037597656
0.40070295333862305
0.4015960693359375
0.3880608081817627
0.405163049697876
0.41303491592407227
0.4232771396636963
0.41156697273254395
0.37395787239074707
0.39792680740356445
0.4032127857208252
0.4138927459716797
0.4032151699066162
0.3818807601928711
0.42504286766052246
0.43329691886901855
0.3741917610168457
0.39508867263793945
0.3879423141479492
0.4103727340698242
0.4214191436767578
0.4040651321411133
0.42829227447509766
0.39633822441101074
0.4126150608062744
0.4038362503051758
0.41495323181152344
0.40180301666259766
0.3945438861846924
0.41541409492492676
0.40806078910827637
0.40170717239379883
0.3949260711669922
0.4168400764465332
0.397672176361084
0.41214799880981445
0.37308692932128906
0.4068601131439209
0.39145398139953613
0.4270503520965576
0.41336607933044434
0.3936350345611572
0.40088510513305664
0.38749098777770996

0.48560094833374023
0.4161851406097412
0.5144577026367188
0.8026270866394043
0.827812910079956
1.0029819011688232
0.9513840675354004
1.2029898166656494
0.40238308906555176
0.5951058864593506
0.4257831573486328
0.4131889343261719
0.4420499801635742
0.38691282272338867
0.4217827320098877
0.4225008487701416
0.39919519424438477
0.4272341728210449
0.392108678817749
0.3958859443664551
0.39287877082824707
0.3949317932128906
0.3961632251739502
0.46027588844299316
873697596434513921
0.39434289932250977
0.4190220832824707
0.4461669921875
0.3897738456726074
0.4178009033203125
872668790621863937
0.3905670642852783
0.5243866443634033
0.4205150604248047
0.43100404739379883
0.39984774589538574
0.39499592781066895
0.3889322280883789
0.4301929473876953
0.40305161476135254
0.40985798835754395
0.4012291431427002
0.42612719535827637
0.42822790145874023
0.42779994010925293
0.41284799575805664
0.42871618270874023
869988702071779329

0.40761613845825195
0.4378318786621094
0.39476990699768066
0.4381241798400879
0.40869593620300293
0.38720107078552246
0.3935680389404297
0.5460290908813477
0.40524816513061523
0.37836289405822754
0.40155482292175293
0.4423952102661133
0.4078350067138672
866816280283807744
0.3904690742492676
0.4325129985809326
0.4056730270385742
0.43982911109924316
0.4045741558074951
0.41309380531311035
0.41519880294799805
0.40515613555908203
0.38928890228271484
0.41312670707702637
0.42711305618286133
0.4131131172180176
0.41774702072143555
0.41588282585144043
0.533911943435669
0.3902561664581299
0.3867008686065674
0.3978118896484375
0.4074709415435791
0.4101860523223877
0.36957716941833496
0.4170072078704834
861769973181624320
0.3823680877685547
0.4023630619049072
0.3997981548309326
0.4062831401824951
0.412883996963501
0.41092920303344727
0.4116511344909668
0.407163143157959
0.39921998977661133
0.4608588218688965
0.3989241123199463

0.39810609817504883
0.3929309844970703
0.5239408016204834
0.3911259174346924
0.4123549461364746
0.4370288848876953
0.3978240489959717
0.3949558734893799
0.3837282657623291
0.3996717929840088
0.416583776473999
0.3808598518371582
0.39542293548583984
0.4034900665283203
0.4013941287994385
0.4266018867492676
0.4091217517852783
0.3817920684814453
0.39406585693359375
0.4074389934539795
0.40555286407470703
0.38860011100769043
0.40088891983032227
0.3936798572540283
0.4222702980041504
0.40699005126953125
0.4014921188354492
0.4477689266204834
0.41201305389404297
0.4165380001068115
0.40913891792297363
0.40273523330688477
0.42267274856567383
0.4009389877319336
0.4085378646850586
0.4129927158355713
0.41619873046875
0.3850979804992676
0.3905799388885498
0.5260794162750244
0.42264699935913086
0.40192198753356934
0.4080212116241455
0.40596890449523926
0.4158592224121094
0.4222440719604492
0.4196479320526123
0.4052598476409912

0.39641284942626953
0.3967609405517578
0.3951129913330078
0.40029430389404297
0.40448999404907227
0.45061588287353516
0.42722082138061523
0.41825389862060547
0.39569997787475586
0.3962087631225586
0.573559045791626
0.637908935546875
0.39767003059387207
0.42932605743408203
0.41325902938842773
0.4013338088989258
0.4000430107116699
0.41434383392333984
0.3980090618133545
0.39629077911376953
0.4072229862213135
0.4232199192047119
0.4057447910308838
0.38604021072387695
0.3995957374572754
0.40307140350341797
0.40102505683898926
0.4174678325653076
0.42092227935791016
0.37206006050109863
0.39129090309143066
0.41196131706237793
845459076796616705
0.38585495948791504
0.4006071090698242
0.3894929885864258
0.41072893142700195
0.413877010345459
0.3905761241912842
0.3997807502746582
0.40552687644958496
0.4814009666442871
0.40485095977783203
0.4140760898590088
0.4145197868347168
842892208864923648
0.38137197494506836
0.42514681816101074

0.40516209602355957
0.3754878044128418
0.40227675437927246
0.4202120304107666
0.4091379642486572
0.36699700355529785
0.39545392990112305
0.5520389080047607
0.4191751480102539
0.39556193351745605
0.4085350036621094
0.40422582626342773
0.3927152156829834
0.38515806198120117
0.4725308418273926
0.45234203338623047
0.40531396865844727
0.3821260929107666
0.40383005142211914
0.40421605110168457
0.40946221351623535
0.40460991859436035
0.38582706451416016
0.4785580635070801
0.4061131477355957
0.4189748764038086
0.39324331283569336
0.39137887954711914
0.4052720069885254
0.4101860523223877
0.3937840461730957
0.4141840934753418
0.4070091247558594
0.3883359432220459
0.36949920654296875
837012587749474308
0.39888811111450195
0.3974893093109131
0.4156219959259033
0.41373538970947266
0.4830968379974365
0.390545129776001
0.391387939453125
0.41584205627441406
0.3932969570159912
0.4100508689880371
0.39290690422058105
0.40207409858703613

0.41814589500427246
0.4140617847442627
0.41115522384643555
0.4125049114227295
0.39634203910827637
0.4166991710662842
0.39015984535217285
0.40349483489990234
0.4112887382507324
0.386732816696167
0.4007549285888672
0.42630815505981445
0.3879239559173584
0.3979341983795166
0.39763402938842773
0.37825894355773926
0.40833520889282227
0.41034531593322754
0.4047389030456543
0.4121220111846924
0.4057488441467285
0.4225442409515381
0.38773417472839355
0.41638970375061035
0.3879122734069824
0.39400577545166016
0.40737485885620117
0.414201021194458
0.3958468437194824
0.4071519374847412
0.42972493171691895
0.40166807174682617
0.47132110595703125
0.384052038192749
0.43106627464294434
0.4195067882537842
0.40584397315979004
0.40098094940185547
0.39232397079467773
0.44989013671875
0.3741281032562256
0.42307400703430176
0.4495658874511719
0.4048440456390381
0.4943990707397461
0.3977489471435547
0.4078061580657959
0.4054441452026367

0.38264012336730957
0.38912534713745117
0.4105799198150635
0.39424729347229004
0.396090030670166
0.4073050022125244
0.39584898948669434
0.4095780849456787
0.3904759883880615
0.38312387466430664
0.3711972236633301
0.4354689121246338
0.4245419502258301
0.43042993545532227
0.3919408321380615
0.4161989688873291
0.4140908718109131
0.39697790145874023
0.4076700210571289
0.3961200714111328
0.4086582660675049
1.348038911819458
0.3999192714691162
827228250799742977
0.3901679515838623
0.4052891731262207
0.405472993850708
0.39049696922302246
0.4228229522705078
0.40935206413269043
0.3823690414428711
0.40199971199035645
0.412769079208374
0.4169919490814209
0.3746497631072998
0.43204498291015625
0.41562390327453613
0.4140360355377197
0.3986368179321289
0.4085960388183594
0.4173099994659424
0.4053928852081299
0.39504122734069824
0.4149651527404785
0.41031885147094727
0.42614197731018066
0.3765280246734619
0.4211697578430176

0.42054295539855957
0.3901028633117676
0.38929200172424316
0.42382097244262695
0.40636324882507324
0.3996908664703369
0.40090012550354004
0.4038407802581787
0.4228858947753906
0.3907139301300049
0.4055778980255127
0.40941691398620605
0.4115619659423828
0.4484426975250244
0.42723584175109863
0.4116957187652588
0.384505033493042
0.3936460018157959
0.392714262008667
0.40982913970947266
0.3987541198730469
0.4073779582977295
0.41744375228881836
0.40364789962768555
0.3895301818847656
0.427325963973999
0.4126396179199219
0.4368250370025635
0.4104330539703369
0.4152798652648926
0.3719451427459717
0.39784884452819824
0.4049186706542969
0.39401793479919434
0.3776869773864746
0.3842649459838867
0.4052388668060303
0.4178290367126465
0.37264299392700195
0.38169074058532715
0.407991886138916
0.4423348903656006
0.3866140842437744
0.40946269035339355
0.41643595695495605
0.42125606536865234
0.44347214698791504
0.388775110244751

0.41704511642456055
0.4038419723510742
0.3902721405029297
0.42692017555236816
0.4114358425140381
0.4037139415740967
0.39855384826660156
0.42258691787719727
0.40297603607177734
0.4833340644836426
0.39997005462646484
0.4060850143432617
0.38611602783203125
0.4271409511566162
0.41950297355651855
0.4081089496612549
0.39699816703796387
0.5021927356719971
0.4023709297180176
0.406404972076416
0.4094381332397461
0.3920760154724121
0.4011058807373047
0.3954441547393799
0.432542085647583
0.39100003242492676
0.39824390411376953
0.4009079933166504
0.39087986946105957
0.4035961627960205
0.3919558525085449
0.41721081733703613
0.4161989688873291
0.38662004470825195
0.4080371856689453
0.3903920650482178
0.48902297019958496
0.3993997573852539
0.4077732563018799
0.3925149440765381
0.39383983612060547
0.39722490310668945
0.4011669158935547
0.4717240333557129
0.4060380458831787
0.4293999671936035
0.4233551025390625
0.39423704147338867

0.39459705352783203
0.4107699394226074
0.41626811027526855
812747805718642688
0.4117591381072998
0.38677287101745605
0.40173912048339844
0.3898751735687256
0.39641499519348145
0.3897268772125244
0.40781331062316895
0.40665292739868164
0.3880350589752197
0.41931986808776855
0.4234199523925781
0.3951852321624756
0.43490004539489746
0.43475770950317383
0.39209508895874023
0.4078071117401123
0.4173290729522705
0.4101450443267822
0.580615758895874
0.5195529460906982
0.3921520709991455
0.4213988780975342
0.402493953704834
0.39194703102111816
0.4036071300506592
0.4180631637573242
0.4182291030883789
0.40213799476623535
0.41336607933044434
0.4277961254119873
0.4132978916168213
0.3913612365722656
0.45650482177734375
0.39545607566833496
0.41331005096435547
0.4282059669494629
0.41180992126464844
0.43223094940185547
0.42188310623168945
0.3915112018585205
0.429060697555542
0.40703701972961426
0.3917262554168701
0.401552677154541

0.4267711639404297
0.40413999557495117
0.4100379943847656
0.3704099655151367
0.4035170078277588
0.39912891387939453
0.3937647342681885
0.41480493545532227
0.4167287349700928
0.3931751251220703
0.39644694328308105
0.41963696479797363
0.4743819236755371
0.3781862258911133
0.3972599506378174
802247111496568832
0.4215717315673828
0.39060187339782715
0.3883347511291504
0.39249277114868164
0.4120481014251709
0.3974430561065674
0.38880109786987305
0.41109609603881836
0.4209620952606201
0.38933491706848145
0.4000840187072754
0.4133589267730713
0.4088318347930908
0.3736698627471924
0.37559080123901367
0.40871500968933105
0.4103100299835205
0.3943932056427002
0.41136789321899414
0.42531895637512207
0.4086337089538574
0.40055394172668457
0.41863083839416504
0.3738400936126709
0.39539098739624023
0.39740896224975586
0.4071969985961914
0.40127086639404297
0.4038519859313965
0.4018361568450928
0.4017517566680908
0.39813995361328125

0.40444374084472656
0.41605401039123535
0.413485050201416
0.4085350036621094
0.4174368381500244
0.4058680534362793
0.4000866413116455
0.38619184494018555
0.4018208980560303
0.4120657444000244
0.48082995414733887
0.36902618408203125
0.4062037467956543
0.4108009338378906
0.3966338634490967
0.375169038772583
0.4865550994873047
0.4093339443206787
0.3930089473724365
0.39012908935546875
0.4027698040008545
0.4159860610961914
0.4030609130859375
0.43001890182495117
0.39499521255493164
0.392780065536499
0.3898942470550537
0.3947410583496094
0.4004402160644531
0.38075923919677734
0.4083890914916992
0.4279301166534424
0.3932023048400879
0.3740348815917969
0.4195249080657959
0.41619300842285156
0.4031369686126709
0.37697482109069824
0.40900421142578125
0.39158201217651367
0.43181419372558594
0.3991551399230957
0.39217305183410645
0.40802502632141113
0.4030270576477051
0.4213578701019287
0.41246581077575684
0.38419294357299805

0.3933069705963135
0.3964962959289551
0.4057199954986572
0.39600300788879395
0.376018762588501
0.41281890869140625
0.4132661819458008
0.4044497013092041
0.400256872177124
0.3923001289367676
0.3944110870361328
0.3841879367828369
0.410966157913208
0.41179871559143066
0.4100809097290039
0.38698601722717285
0.43175196647644043
0.41477084159851074
0.4085681438446045
0.3952798843383789
0.42403101921081543
0.41029882431030273
0.39246487617492676
0.4187130928039551
0.41867709159851074
0.40541911125183105
0.3870372772216797
0.40488696098327637
0.4131340980529785
0.4203782081604004
0.40033888816833496
0.39663195610046387
0.40708398818969727
0.38500213623046875
0.40026021003723145
0.40076613426208496
0.448638916015625
0.39097094535827637
0.39981698989868164
0.3978250026702881
0.41509103775024414
0.42124176025390625
0.41257500648498535
0.3916642665863037
0.40929102897644043
0.4298219680786133
0.3820328712463379
0.4160912036895752

0.40757131576538086
0.38843870162963867
0.42113518714904785
0.42178797721862793
0.3893749713897705
0.39775681495666504
0.39401698112487793
0.4060227870941162
0.40039515495300293
0.3796398639678955
0.40632104873657227
0.4059786796569824
0.40538811683654785
0.3962690830230713
0.4336061477661133
0.4147789478302002
0.44893312454223633
0.394258975982666
0.4109160900115967
0.4156360626220703
0.4190866947174072
0.4043879508972168
0.3921499252319336
0.4247770309448242
0.4011240005493164
0.38869237899780273
0.4061698913574219
0.406649112701416
0.3891570568084717
0.3823888301849365
0.4058041572570801
0.3888099193572998
0.4005119800567627
0.4259181022644043
0.42752599716186523
0.41623592376708984
0.39716482162475586
0.41062307357788086
0.4268782138824463
0.3880448341369629
0.4092140197753906
0.5027358531951904
0.40763378143310547
0.41102099418640137
0.41700291633605957
0.40125536918640137
0.4317638874053955
0.43213486671447754

0.3763437271118164
0.397747278213501
0.40944719314575195
0.42333197593688965
0.38204407691955566
0.4235880374908447
0.40512990951538086
0.40996503829956055
0.4303760528564453
0.42157506942749023
0.3812401294708252
0.398317813873291
0.402616024017334
0.41495704650878906
0.4201641082763672
0.39302897453308105
0.4206860065460205
0.4810659885406494
0.38253283500671387
0.39102792739868164
0.43135619163513184
0.4136629104614258
0.3797628879547119
0.3935427665710449
0.4104921817779541
0.41678619384765625
0.40940093994140625
0.394406795501709
0.3937649726867676
0.38733625411987305
0.4153130054473877
0.4183528423309326
0.39658308029174805
0.4143033027648926
0.4160919189453125
0.4169962406158447
0.40640711784362793
0.396151065826416
0.4474930763244629
0.43724703788757324
0.3910808563232422
775096608509886464
0.4555039405822754
0.384627103805542
0.4156160354614258
0.4157898426055908
0.3967552185058594
0.40115904808044434

0.383986234664917
0.40695619583129883
0.45580530166625977
0.3883249759674072
0.3832249641418457
0.4019339084625244
0.4129307270050049
0.3883488178253174
0.3908412456512451
0.39797401428222656
0.6232020854949951
0.40859484672546387
0.408494234085083
0.3914451599121094
0.41135191917419434
0.43039488792419434
0.417266845703125
0.42551589012145996
0.3940141201019287
0.3912520408630371
0.6259772777557373
0.5143511295318604
0.39792585372924805
0.4178791046142578
0.41629481315612793
0.3984830379486084
0.4025540351867676
770743923962707968
0.40137171745300293
0.4008300304412842
0.40170717239379883
0.41649603843688965
0.40012693405151367
0.3808920383453369
0.3914198875427246
0.42406678199768066
0.39155101776123047
0.37813282012939453
0.3887948989868164
0.3998699188232422
0.40133023262023926
0.38895201683044434
0.4589118957519531
0.3947138786315918
0.3932681083679199
0.38008594512939453
0.40308618545532227
0.40837788581848145

0.39742016792297363
0.4111170768737793
0.4070861339569092
0.4113950729370117
0.40702009201049805
0.4013340473175049
0.39280200004577637
0.39651918411254883
0.42130208015441895
0.4101722240447998
0.40618085861206055
0.45511293411254883
0.4017157554626465
0.555413007736206
0.39684391021728516
0.42584705352783203
0.4094831943511963
0.3786928653717041
0.4196290969848633
0.41626906394958496
0.3957529067993164
0.41722917556762695
0.39134979248046875
0.3969569206237793
0.41811609268188477
0.3781750202178955
0.4153158664703369
0.4097578525543213
0.3769698143005371
0.40613722801208496
0.3983731269836426
0.4038889408111572
0.40460705757141113
0.3899390697479248
0.5568768978118896
0.38155698776245117
0.3971390724182129
0.4086189270019531
0.4155740737915039
0.40774083137512207
0.39327502250671387
0.41074395179748535
0.42626214027404785
0.426192045211792
0.4073038101196289
0.39910340309143066
0.4286937713623047
0.4172999858856201

```
0.3959341049194336
0.4297981262207031
0.42588305473327637
0.39990782737731934
0.3814249038696289
0.40569019317626953
0.4232640266418457
0.3804149627685547
0.4228401184082031
0.4287707805633545
0.4043142795562744
0.3947110176086426
0.41080594062805176


Rate limit reached. Sleeping for: 528


533.696469783783
0.41982603073120117
0.3906371593475342
0.4201548099517822
0.4344909191131592
0.46900463104248047
0.3933830261230469
0.3890368938446045
0.44300127029418945
0.38306593894958496
0.41452908515930176
0.3888988494873047
0.42049312591552734
0.41269803047180176
0.40535593032836914
0.4304921627044678
0.416762113571167
0.40422677993774414
0.3977317810058594
0.395035982131958
0.3974339962005615
0.3816249370574951
0.4118993282318115
0.491041898727417
0.42212605476379395
0.4410710334777832
0.4362947940826416
0.43041396141052246
0.4917008876800537
0.4734489917755127
```

0.42803096771240234
0.4296751022338867
754011816964026368
0.42095375061035156
0.39218711853027344
0.39488720893859863
0.4405970573425293
0.4122469425201416
0.40545105934143066
0.3941538333892822
0.39253926277160645
0.420332670211792
0.40805721282958984
0.40773773193359375
0.39913487434387207
0.4059107303619385
0.4094390869140625
0.3924601078033447
0.40836501121520996
0.4122471809387207
0.3992130756378174
0.382997989654541
0.4034538269042969
0.4294471740722656
0.39753198623657227
0.38103318214416504
0.3924570083618164
0.41290283203125
0.3733100891113281
0.4014139175415039
0.41132688522338867
0.40610504150390625
0.38492369651794434
0.4039132595062256
0.5262389183044434
0.39483189582824707
0.3817427158355713
0.38583803176879883
0.43019914627075195
0.3806459903717041
0.41138505935668945
0.4239170551300049
0.40364694595336914
0.41727113723754883
0.3894531726837158
0.40542101860046387
0.37897491455078125
0.41606903076171875

0.3756699562072754
0.417863130569458
0.41210412979125977
0.42891693115234375
0.40718793869018555
0.4329841136932373
0.38678503036499023
0.40952086448669434
0.4091019630432129
0.4087638854980469
0.39345622062683105
0.4104282855987549
0.43704700469970703
0.39813923835754395
0.4145781993865967
0.41300272941589355
0.4121208190917969
0.4177069664001465
0.40093183517456055
0.3951601982116699
0.4056830406188965
0.46006226539611816
0.3722560405731201
0.42118310928344727
0.4115488529205322
0.40794992446899414
0.3984708786010742
0.39827585220336914
0.39309024810791016
0.3865959644317627
0.3967750072479248
0.4396529197692871
0.40779614448547363
0.38242411613464355
0.39962315559387207
0.3976173400878906
0.3972008228302002
0.3746480941772461
0.37180399894714355
0.4022550582885742
0.3774888515472412
0.40918493270874023
0.39057302474975586
0.3821728229522705
0.3812699317932129
0.40898990631103516
0.42240190505981445
0.40148401260375977

0.4111039638519287
0.4006941318511963
0.3735320568084717
0.4245150089263916
0.4364349842071533
0.39411306381225586
0.39481496810913086
0.40706491470336914
0.39852213859558105
0.4001171588897705
0.39890193939208984
0.4215061664581299
0.3938171863555908
0.3738059997558594
0.40177392959594727
0.447174072265625
0.43394899368286133
0.38010120391845703
0.3955812454223633
0.4213838577270508
0.4115569591522217
0.4158480167388916
0.4078669548034668
0.443418025970459
0.4009242057800293
0.3849506378173828
0.3732450008392334
0.4181067943572998
0.5272390842437744
0.37607502937316895
0.4006233215332031
0.43724584579467773
0.392453670501709
0.5463607311248779
0.4076361656188965
0.3868980407714844
0.39548373222351074
0.39813995361328125
0.39571690559387207
0.398406982421875
0.40588808059692383
0.41140317916870117
0.40700697898864746
0.4195430278778076
0.40713024139404297
0.4144251346588135
0.4224698543548584
0.4108998775482178

0.40494799613952637
0.4008667469024658
0.47807812690734863
0.4168679714202881
0.43717002868652344
0.4876887798309326
0.42349720001220703
0.46271491050720215
0.41052794456481934
0.40006494522094727
0.3990819454193115
0.39990973472595215
0.3935530185699463
0.3909280300140381
0.41953492164611816
0.4158139228820801
0.4044339656829834
0.43183183670043945
0.3909797668457031
0.41182994842529297
0.4731299877166748
0.3877599239349365
0.39336299896240234
0.449465274810791
0.37571120262145996
0.3812727928161621
0.43341898918151855
0.40342283248901367
0.3767580986022949
0.41335606575012207
0.3842611312866211
0.40389132499694824
0.3921060562133789
0.41559386253356934
0.3948988914489746
0.4124617576599121
0.4104442596435547
0.4122731685638428
0.4081108570098877
0.38167238235473633
0.41007304191589355
0.4105720520019531
0.3951380252838135
0.40267300605773926
0.41301894187927246
0.3867208957672119
0.40514492988586426
0.4691281318664551

0.413402795791626
0.37755703926086426
0.3814709186553955
0.42997312545776367
0.42121195793151855
0.39792823791503906
0.3895609378814697
0.41291093826293945
0.40257835388183594
0.4064059257507324
0.39279794692993164
0.4155120849609375
0.3872811794281006
0.40871095657348633
0.40488505363464355
0.40006399154663086
0.37688493728637695
0.39771580696105957
0.4035968780517578
0.41471171379089355
0.39733195304870605
0.43759799003601074
0.4039640426635742
0.40297484397888184
0.39366722106933594
0.42966294288635254
0.42717695236206055
0.39873313903808594
0.39714694023132324
0.4210779666900635
0.4101991653442383
0.3912949562072754
0.4206361770629883
0.4178318977355957
0.41649794578552246
0.3967721462249756
0.38736486434936523
0.4294252395629883
0.41747117042541504
0.4129040241241455
0.3935701847076416
0.3965780735015869
0.39851975440979004
0.38947296142578125
0.40888381004333496
0.41286516189575195
0.38015222549438477
0.40566086769104004

0.4203300476074219
0.38902902603149414
0.4209558963775635
0.3926839828491211
0.39199113845825195
0.3725011348724365
0.40455126762390137
0.4265589714050293
0.4055509567260742
0.39391183853149414
0.3937520980834961
0.39873194694519043
0.42335057258605957
0.38228726387023926
0.39787793159484863
0.4789578914642334
0.4201819896697998
0.39908504486083984
0.408843994140625
0.40700697898864746
0.41396307945251465
0.3922102451324463
0.41326189041137695
0.4225499629974365
0.4201488494873047
0.39737582206726074
0.404158353805542
0.4198329448699951
0.4189300537109375
0.3985447883605957
0.41950178146362305
0.4209611415863037
0.38062524795532227
0.4014749526977539
0.3905770778656006
0.4253840446472168
0.43366312980651855
0.40521788597106934
0.38344311714172363
0.3939070701599121
0.427142858505249
0.39987778663635254
0.3931143283843994
0.48000597953796387
0.38135600090026855
0.4253079891204834
0.4209609031677246
0.38526296615600586

0.39324402809143066
0.4208371639251709
0.4022181034088135
0.3934319019317627
0.4112210273742676
0.38692712783813477
0.4111032485961914
0.3956260681152344
0.41321396827697754
0.4044961929321289
0.39748406410217285
0.4077341556549072
0.3955850601196289
0.3938329219818115
0.39171409606933594
0.40696096420288086
0.40805602073669434
0.43592381477355957
0.4037487506866455
0.41661500930786133
0.37345099449157715
0.40458106994628906
0.39624595642089844
0.4170231819152832
0.39017319679260254
0.41201186180114746
0.4006972312927246
0.3933751583099365
0.39347100257873535
0.5081601142883301
0.5126028060913086
0.38377904891967773
0.38564205169677734
0.41510605812072754
0.40403127670288086
0.4009730815887451
0.40053486824035645
0.4027392864227295
0.41867780685424805
0.4256880283355713
0.38191795349121094
0.3958699703216553
0.3851299285888672
0.3914520740509033
0.400911808013916
0.4128241539001465
0.39851808547973633
0.4196507930755615

0.4166240692138672
0.40457916259765625
0.4165313243865967
0.44234299659729004
0.4525110721588135
0.3785982131958008
0.3944261074066162
0.40880608558654785
0.38095903396606445
0.4117579460144043
0.40964818000793457
0.39135098457336426
0.4257698059082031
0.396589994430542
0.4171297550201416
0.4069631099700928
0.4037139415740967
0.4220468997955322
0.3990061283111572
0.40191102027893066
0.41977787017822266
0.5416238307952881
0.4035320281982422
0.4408853054046631
0.42232799530029297
0.4039139747619629
0.38654375076293945
0.3883330821990967
0.4212479591369629
0.40774106979370117
0.3811922073364258
0.40169811248779297
0.39426112174987793
0.41795802116394043
0.38476991653442383
0.39731621742248535
0.4099428653717041
0.4084038734436035
0.37459373474121094
0.41147804260253906
0.40906810760498047
0.41255712509155273
0.38889098167419434
0.4011199474334717
0.41509413719177246
0.427915096282959
0.4203500747680664
0.4056098461151123

0.4113471508026123
0.46361279487609863
0.38927173614501953
0.39119696617126465
0.39066410064697266
0.40000200271606445
0.4136509895324707
0.3920612335205078
0.4194009304046631
0.39782214164733887
0.40769195556640625
0.41240811347961426
0.40777087211608887
0.3987250328063965
0.3921020030975342
0.4240119457244873
0.39169931411743164
0.40143299102783203
0.3931248188018799
0.43440818786621094
0.42011499404907227
0.4274308681488037
0.3911259174346924
0.482982873916626
0.395766019821167
0.41338610649108887
0.39658212661743164
0.3982052803039551
0.39358091354370117
0.4149940013885498
0.4040408134460449
0.3958401679992676
0.39458394050598145
0.38413095474243164
0.41870689392089844
0.3991880416870117
0.44045090675354004
0.41866397857666016
0.40090203285217285
0.4104428291320801
0.4000091552734375
0.42139220237731934
0.4313089847564697
0.43747901916503906
0.38886117935180664
0.4127368927001953
0.4113152027130127
0.4642179012298584

0.377946138381958
0.38771581649780273
0.38950085639953613
0.42186999320983887
0.42226696014404297
0.3936338424682617
0.3962700366973877
0.4308328628540039
0.392636775970459
0.39809679985046387
0.39145803451538086
0.4367706775665283
0.3910667896270752
0.3989090919494629
0.3775808811187744
0.3961641788482666
0.38957905769348145
0.3938863277435303
0.3995089530944824
0.40653181076049805
0.40502309799194336
0.3802371025085449
0.3957839012145996
0.4468820095062256
0.381976842880249
0.41533493995666504
0.37560200691223145
0.6874771118164062
0.39170312881469727
0.39213013648986816
0.39481329917907715
0.4245471954345703
0.4010753631591797
0.6100082397460938
0.40859007835388184
0.42430710792541504
0.4075350761413574
0.4272129535675049
0.3930819034576416
0.3891899585723877
0.45786476135253906
0.4103119373321533
0.40751099586486816
0.4166688919067383
0.40555500984191895
0.40775394439697266
0.40590786933898926
0.416027307510376

0.40821409225463867
0.39731478691101074
0.3968069553375244
0.428333044052124
0.3965480327606201
0.40442609786987305
0.381791353225708
0.3876378536224365
0.4501218795776367
0.39687609672546387
0.4015309810638428
0.3860170841217041
0.3886539936065674
0.3771212100982666
0.3958249092102051
0.39658212661743164
0.40894603729248047
0.4020547866821289
0.3947608470916748
0.4142460823059082
0.38645386695861816
0.3853440284729004
0.40082502365112305
0.4497082233428955
0.403656005859375
0.3943970203399658
0.41382813453674316
0.3921630382537842
0.4078347682952881
0.4779520034790039
0.38655900955200195
0.4120776653289795
0.3993949890136719
0.3727140426635742
0.4192080497741699
0.41001176834106445
0.40985679626464844
0.37697815895080566
0.40143513679504395
0.38576221466064453
0.414808988571167
0.3838841915130615
0.42688584327697754
0.392808198928833
0.3831369876861572
0.39447498321533203
0.41504406929016113
0.38808703422546387

0.38695573806762695
0.39553403854370117
0.4088597297668457
0.4127628803253174
0.5405230522155762
0.3982279300689697
0.4102489948272705
0.40881919860839844
0.3960709571838379
0.4162440299987793
0.4220712184906006
0.4020960330963135
0.39743471145629883
0.4134540557861328
0.3871591091156006
0.3834259510040283
0.4568440914154053
0.4036262035369873
0.38344407081604004
0.37180495262145996
0.4062211513519287
0.38748812675476074
0.40200376510620117
0.39406275749206543
0.39859604835510254
0.40634608268737793
0.4313998222351074
0.3861119747161865
0.4059488773345947
0.41081809997558594
0.5667979717254639
0.38776302337646484
0.4225621223449707
0.40411901473999023
0.413923978805542
0.39571309089660645
0.4200000762939453
0.4100821018218994
0.4070761203765869
0.3995392322540283
0.39939022064208984
0.40146827697753906
0.385692834854126
0.4232020378112793
0.39554905891418457
0.40929484367370605
0.40273618698120117
0.40633296966552734

0.4079411029815674
0.3982977867126465
0.38791799545288086
0.42812180519104004
0.4007248878479004
0.4013099670410156
0.40536999702453613
0.40995192527770996
0.4055788516998291
0.386411190032959
0.39375782012939453
0.4095900058746338
0.4007449150085449
0.4031379222869873
0.38178515434265137
0.4011979103088379
0.4110147953033447
0.39444470405578613
0.4108729362487793
0.37874817848205566
0.3993220329284668
0.39733314514160156
0.413754940032959
0.40047621726989746
0.45326900482177734
0.39850282669067383
0.5495648384094238
0.3816850185394287
0.3996891975402832
0.38754820823669434
0.398435115814209
0.400479793548584
0.3932211399078369
0.4066288471221924
0.4081919193267822
0.3924109935760498
0.39941883087158203
0.40895509719848633
0.4044158458709717
0.3955357074737549
0.39501094818115234
0.39164113998413086
0.4333069324493408
0.39452528953552246
0.38065671920776367
0.40923309326171875
0.4073758125305176
0.39582109451293945

0.3801398277282715
0.4149930477142334
0.420025110244751
0.37285399436950684
0.38617801666259766
0.3893871307373047
0.5135679244995117
0.3757476806640625
0.4024219512939453
0.43769001960754395
0.4189488887786865
0.37797975540161133
0.4333610534667969
0.4457669258117676
0.39644908905029297
0.38739919662475586
0.40257906913757324
0.4088010787963867
0.4082660675048828
0.402987003326416
0.39652204513549805
0.41268301010131836
0.402637243270874
0.39826202392578125
0.38515710830688477
0.4088907241821289
0.4315361976623535
0.39205002784729004
0.4276440143585205
0.39231085777282715
0.4086272716522217
0.4023728370666504
0.3814530372619629
0.413909912109375
0.4102609157562256
0.40920114517211914
0.39387083053588867
0.37688493728637695
0.38540196418762207
0.4165990352630615
0.39573073387145996
0.38964295387268066
0.41132020950317383
0.40873217582702637
0.4046599864959717
0.4065699577331543
0.3855319023132324
0.4206218719482422

0.40631890296936035
0.40545010566711426
0.3948678970336914
0.4130690097808838
0.4009532928466797
0.3870272636413574
0.3972511291503906
0.41681909561157227
0.4245729446411133
0.3841512203216553
0.40688371658325195
0.4315471649169922
0.3907759189605713
0.38614916801452637
0.37888598442077637
0.4225740432739258
0.39250993728637695
0.38388514518737793
0.4084761142730713
0.401900053024292
0.38746023178100586
0.5502870082855225
0.39990711212158203
0.39812564849853516
0.4039311408996582
0.3975410461425781
0.4036233425140381
0.4612269401550293
0.3947930335998535
0.4165489673614502
0.376467227935791
0.403641939163208
0.40158700942993164
0.41680908203125
0.3720428943634033
0.41631579399108887
0.4004521369934082
0.5356428623199463
0.3999979496002197
0.40866518020629883
0.3955190181732178
0.4327511787414551
0.4041001796722412
0.40790605545043945
0.404155969619751
0.3966047763824463
0.4002799987792969
0.4063730239868164

0.4305911064147949
0.390178918838501
0.39256906509399414
0.40095996856689453
0.39331507682800293
0.3951120376586914
0.4088311195373535
0.4204092025756836
0.3755319118499756
0.5253312587738037
0.4136190414428711
0.4121401309967041
0.3798971176147461
0.37073421478271484
0.3909759521484375
0.41260480880737305
0.4021928310394287
0.37987518310546875
0.41890883445739746
0.415463924407959
0.38988590240478516
0.3966829776763916
0.39458298683166504
0.4288959503173828
0.3892698287963867
0.3883368968963623
0.40394115447998047
0.4655139446258545
0.4113473892211914
0.40253305435180664
0.39966702461242676
0.42087769508361816
0.39397573471069336
0.41519689559936523
0.5397689342498779
0.40652894973754883
0.3789100646972656
0.4065999984741211
0.39100003242492676
0.38526296615600586
0.3981921672821045
0.3984651565551758
0.41580820083618164
0.3986470699310303
0.4004988670349121
0.41240596771240234
0.3921940326690674
0.3955247402191162

0.4025299549102783
0.4278140068054199
0.3899989128112793
0.40634608268737793
0.38325023651123047
0.405548095703125
0.44092798233032227
0.39335179328918457
0.3859250545501709
0.49795985221862793
0.4185810089111328
0.40445995330810547
0.4106900691986084
0.3956410884857178
0.43580007553100586
0.3990461826324463
0.38178515434265137
0.4807000160217285
0.3717329502105713
0.3967750072479248
0.40256834030151367
0.38025808334350586
0.4019479751586914
0.41541624069213867
0.40505385398864746
0.38658833503723145
0.38523221015930176
0.4151580333709717
680055455951884288
0.39989399909973145
0.39191222190856934
0.40288519859313965
0.4193301200866699
0.408419132232666
0.3933119773864746
0.4055788516998291
0.4362471103668213
0.39279794692993164
0.3808748722076416
0.4078490734100342
0.3997471332550049
0.4059138298034668
0.4024958610534668
0.40323519706726074
0.39683008193969727
0.38040685653686523
0.40731287002563477
0.42921924591064453

0.41673922538757324
0.3917050361633301
0.40058112144470215
0.4002699851989746
0.42331504821777344
0.3861269950866699
0.38376307487487793
0.4195587635040283
0.4174158573150635
0.42966699600219727
0.3782939910888672
0.41554903984069824
0.4140598773956299
0.40574169158935547
0.38569211959838867
0.4104928970336914
0.4402308464050293
0.3921489715576172
0.39474916458129883
0.42341184616088867
0.4084348678588867
0.4048187732696533
0.37360405921936035
0.3896477222442627
0.4171440601348877
0.41306591033935547
0.3934669494628906
0.40399694442749023
0.43709683418273926
0.37696313858032227
0.3763859272003174
0.38724398612976074
0.5361301898956299
0.41016387939453125
0.4059560298919678
0.4034099578857422
0.40496110916137695
0.40001773834228516
0.39097023010253906
0.388261079788208
0.4030940532684326
0.4062778949737549
0.39223599433898926
0.404400110244751
0.4291701316833496
0.4012620449066162
0.3997008800506592
0.4085400104522705

```
0.40943408012390137
0.3858969211578369
0.41537976264953613
0.38872408866882324
0.37599992752075195
0.3930470943450928


Rate limit reached. Sleeping for: 532


537.7044861316681
0.39557790756225586
0.43133020401000977
0.40377068519592285
0.4120931625366211
0.40102314949035645
0.38814783096313477
0.39780092239379883
0.3973262310028076
0.40305590629577637
0.4423530101776123
0.4450070858001709
0.38541579246520996
0.3867981433868408
0.39977407455444336
0.420055627822876
0.3922092914581299
0.3865821361541748
0.399259090423584
0.42824220657348633
0.4170968532562256
0.39939188957214355
0.396712064743042
0.43874192237854004
0.3886599540710449
0.38070082664489746
0.38977503776550293
0.4257500171661377
0.3890988826751709
0.41750574111938477
0.4170267581939697
0.41814184188842773
0.3824150562286377
0.4123849868774414
0.3956489562988281
0.430682897567749
0.37461090087890625
```

0.4121053218841553
0.4361600875854492
0.39039182662963867
0.4047551155090332
0.4162909984588623
0.401716947555542
0.40638113021850586
0.5384738445281982
0.4078688621520996
0.44277477264404297
0.41198182106018066
0.3993949890136719
0.48601508140563965
0.40104198455810547
0.4103519916534424
0.40755701065063477
0.39685487747192383
0.4080660343170166
0.4165840148925781
0.3849320411682129
0.41853809356689453
0.4113621711730957
0.414337158203125
0.38210201263427734
0.38813304901123047
0.4029262065887451
0.5963020324707031
0.40412473678588867
0.3929910659790039
0.3941628932952881
0.38802099227905273
0.3875691890716553
0.4355342388153076
0.41066408157348633
0.4150419235229492
0.4120509624481201
0.41535496711730957
0.418201208114624
0.39838504791259766
0.37919092178344727
0.4241979122161865
0.4271430969238281
0.3867940902709961
0.3870687484741211
0.4115149974822998
0.39061999320983887
0.4116077423095703
0.3798062801361084

0.4015798568725586
0.41318392753601074
0.42594099044799805
0.40198302268981934
0.40772104263305664
0.3918459415435791
0.40303897857666016
0.39708495140075684
0.3892951011657715
0.3973078727722168
0.44281983375549316
0.4129006862640381
0.40250396728515625
0.5689167976379395
0.4005312919616699
0.4018111228942871
0.41796040534973145
0.38997912406921387
0.39932703971862793
0.4196758270263672
0.38936781883239746
0.38782429695129395
0.40694713592529297
0.41490936279296875
0.3948500156402588
0.38556790351867676
0.4111628532409668
0.41470980644226074
0.410322904586792
0.4039480686187744
0.4168093204498291
0.3780553340911865
0.3700900077819824
0.405764102935791
0.43591904640197754
0.4163978099822998
0.3891010284423828
0.42813801765441895
0.5485057830810547
0.5089268684387207
0.40346789360046387
0.3991820812225342
0.38176584243774414
0.39301490783691406
0.43707799911499023
0.41747188568115234
0.40181493759155273
0.42261195182800293

0.40724778175354004
0.41772007942199707
0.40583229064941406
0.40827012062072754
0.4086132049560547
0.40392088890075684
0.4143979549407959
0.3974630832672119
0.4114351272583008
0.4109458923339844
0.4132840633392334
0.4328160285949707
0.40946316719055176
0.3822908401489258
0.44188594818115234
0.37592291831970215
0.4176340103149414
0.41518688201904297
0.4219379425048828
0.40627217292785645
0.3930058479309082
0.3999488353729248
0.4506499767303467
0.3966550827026367
0.38857102394104004
0.39569616317749023
0.3808152675628662
0.41456103324890137
0.39023900032043457
0.3836331367492676
0.42231082916259766
0.40931105613708496
0.40556788444519043
0.4105720520019531
0.4355473518371582
0.4027090072631836
0.39353299140930176
0.3995931148529053
0.40808987617492676
0.4102780818939209
0.41352391242980957
0.3831610679626465
0.3996453285217285
0.40119099617004395
0.39006900787353516
0.4178321361541748
0.4015669822692871
0.408825159072876

0.4152567386627197
0.40825486183166504
0.39711618423461914
0.387286901473999
0.4191551208496094
0.4017910957336426
0.40656208992004395
0.3982851505279541
0.3947739601135254
0.39034295082092285
0.4325289726257324
0.39981532096862793
0.4001138210296631
0.41100597381591797
0.39986395835876465
0.4143400192260742
0.42026400566101074
0.39508700370788574
0.41261911392211914
0.3838350772857666
0.40486598014831543
0.40442991256713867
0.39224696159362793
0.402163028717041
0.41335391998291016
0.41016411781311035
0.5027048587799072
0.46680498123168945
0.4030318260192871
0.4029390811920166
0.41666102409362793
0.4154832363128662
0.4074277877807617
0.4056720733642578
0.4195690155029297
0.3863401412963867
0.40230798721313477
0.40619421005249023
0.41358184814453125
0.574174165725708
0.4110841751098633
0.414870023727417
0.4023098945617676
0.37157297134399414
0.39559292793273926
0.39304685592651367
0.40147900581359863
0.39512109756469727

0.39086103439331055
0.40525102615356445
0.4005587100982666
0.39243197441101074
0.3971259593963623
0.4084811210632324
0.4004178047180176
0.38048386573791504
0.38288402557373047
0.4209311008453369
0.42809605598449707
0.3970448970794678
0.3980369567871094
0.4161999225616455
0.41254591941833496
0.39134812355041504
0.422055721282959
0.43935227394104004
0.43894219398498535
0.45637083053588867
0.46597719192504883
0.4452629089355469
0.4049999713897705
0.3827838897705078
0.4244241714477539
0.4228241443634033
0.408735990524292
0.3993957042694092
0.407886266708374
0.41446781158447266
0.3906989097595215
0.3928260803222656
0.39919519424438477
0.41440701484680176
0.42455291748046875
0.4011530876159668
0.3866157531738281
0.420835018157959
0.41689610481262207
0.4103569984436035
0.3953070640563965
0.3995242118835449
0.585662841796875
0.3897249698638916
0.42176198959350586
0.41843509674072266
0.4050579071044922
0.41661715507507324

0.383530855178833
0.42020392417907715
0.4180271625518799
0.40612101554870605
0.4271581172943115
0.37627196311950684
0.41857099533081055
0.406451940536499
0.4044029712677002
0.39467310905456543
0.40781688690185547
0.40219593048095703
0.3771698474884033
0.404693603515625
0.42031311988830566
0.4146409034729004
0.4109818935394287
0.40160393714904785
0.42335009574890137
0.3956921100616455
0.41924571990966797
0.4040699005126953
0.4222259521484375
0.40200090408325195
0.38059091567993164
0.39236903190612793
0.41512012481689453
0.3870971202850342
0.38688206672668457
0.3902740478515625
0.4032471179962158
0.4141509532928467
0.39711904525756836
0.4027400016784668
0.4343540668487549
0.4033961296081543
0.46837711334228516
0.4320950508117676
0.39219188690185547
0.39914584159851074
0.42019104957580566
0.41172003746032715
0.40883421897888184
0.3860290050506592
0.39222192764282227
0.39867305755615234
0.3956780433654785
0.3933892250061035

0.38425230979919434
0.4178941249847412
0.4117000102996826
0.4384000301361084
0.3941648006439209
0.40867185592651367
0.4286630153656006
0.3898589611053467
0.38652682304382324
0.4050099849700928
0.41960906982421875
0.38486218452453613
0.3773529529571533
0.3971269130706787
0.43350672721862793
0.40251708030700684
0.39379191398620605
0.4038238525390625
0.41840600967407227
0.4157140254974365
0.381026029586792
0.3963019847869873
0.4084460735321045
0.4186208248138428
0.3700838088989258
0.3995661735534668
0.3937678337097168
0.403609037399292
0.3871941566467285
0.39901089668273926
0.49506139755249023
0.4088921546936035
0.3977668285369873
0.39827609062194824
0.4602370262145996
0.3951759338378906
0.3941068649291992
0.3984677791595459
0.39782023429870605
0.40485715866088867
0.39061927795410156
0.4554319381713867
0.41194963455200195
0.3903980255126953
0.38320302963256836
0.39787912368774414
0.4388911724090576
0.41100287437438965

0.3740880489349365
0.3969721794128418
0.4154322147369385
0.4591999053955078
0.43266820907592773
0.4074890613555908
0.39392828941345215
0.36893796920776367
0.3918581008911133
0.40187811851501465
0.4028298854827881
0.3992729187011719
0.39690089225769043
0.42848706245422363
0.4041781425476074
0.38935089111328125
0.3877410888671875
0.3973968029022217
0.39339399337768555
0.40334200859069824
0.3941068649291992
0.407073974609375
0.4174678325653076
0.3960130214691162
0.42090320587158203
0.4169652462005615
0.42449402809143066
0.4053351879119873
0.3953080177307129
0.4019052982330322
0.48618316650390625
0.4339780807495117
0.41375303268432617
0.4112370014190674
0.44798994064331055
0.41446590423583984
0.43597888946533203
0.47113680839538574
0.40992307662963867
0.40634703636169434
0.3799269199371338
0.39136791229248047
0.4019441604614258
0.40163111686706543
0.39417123794555664
0.4014272689819336
0.5380239486694336
0.39463210105895996

0.3841240406036377
0.4022090435028076
0.3866841793060303
0.428041934967041
0.4041478633880615
0.40088796615600586
0.3951230049133301
0.39229393005371094
0.43436598777770996
0.39810991287231445
0.40714192390441895
0.4035766124725342
0.4142179489135742
0.42672109603881836
0.4139988422393799
0.37366676330566406
0.41399407386779785
0.39174699783325195
0.5210790634155273
0.3924241065979004
0.42762112617492676
0.40189218521118164
0.3953559398651123
0.39574575424194336
0.40625786781311035
0.3952031135559082
0.40604281425476074
0.4317348003387451
0.39064598083496094
0.408627986907959
0.41179990768432617
0.38982200622558594
0.3904228210449219
0.4168057441711426
0.45279693603515625
0.36584901809692383
0.3935511112213135
0.40250492095947266
0.4264488220214844
0.3867921829223633
0.39254307746887207
0.38954806327819824
0.4287691116333008
0.396007776260376
0.3805429935455322
0.4234120845794678
0.42143797874450684
0.39247798919677734

0.4006690979003906
0.41377997398376465
0.4225459098815918
0.39682888984680176
0.3964071273803711
0.5894718170166016
0.4269087314605713
0.488922119140625
0.40975117683410645
0.4216020107269287
0.40609216690063477
0.38025903701782227
0.40430712699890137
0.4090588092803955
0.3870236873626709
0.38736891746520996
0.4278452396392822
0.4167449474334717
0.3972151279449463
0.39708876609802246
0.39876508712768555
0.42029523849487305
0.4369950294494629
0.3944072723388672
0.3879737854003906
0.41683387756347656
0.38680195808410645
0.4092848300933838
0.4161238670349121
0.4153940677642822
0.3975982666015625
0.5022940635681152
0.41460180282592773
0.7122368812561035
0.4159121513366699
0.4058380126953125
0.3907201290130615
0.42345285415649414
0.3996310234069824
0.3975529670715332
0.5353169441223145
0.389606237411499
0.4113643169403076
0.3894989490509033
0.4151740074157715
0.3728008270263672
0.4205801486968994
0.4081242084503174

```
0.4022519588470459
0.38953495025634766
0.4174509048461914
0.3842189311981201
0.3938028812408447
0.403411865234375
0.43556714057922363
0.42803406715393066
0.39058899879455566
0.5053391456604004
0.413226842880249
0.38147616386413574
0.39585304260253906
0.40962791442871094
0.4076230525970459
0.37947893142700195
0.4057137966156006
0.408128023147583
0.4247860908508301
0.43572402000427246
0.41420912742614746
0.41088294982910156
0.43650078773498535
0.4025442600250244
0.41912102699279785
0.4172658920288086
0.40434980392456055
0.38060808181762695
0.4072728157043457
0.3998889923095703
0.3972001075744629
0.41284990310668945
0.39113688468933105
0.4141707420349121
0.43420910835266113
0.4129981994628906
0.41970205307006836
0.4255411624908447
0.3946101665496826
```

```
In [12]: # Then in this code block we isolate the json part of each tweepy status object
         # that we have downloaded and we add them all into a list....
         my_list_of_dicts = []
         for each_json_tweet in list_of_tweets:
             my_list_of_dicts.append(each_json_tweet._json)
         #...and then we write this list into a txt file:
         with open('tweet_json.txt', 'w') as file:
```

```
                file.write(json.dumps(my_list_of_dicts, indent=4))
```

```
In [13]: # Twitter Query using twitter_id information from twitter_archive

         # Ref for rate limit - https://developer.twitter.com/en/docs/basics/rate-limiting
         # Ref for Measuring time elapsed in Python(code timer) - https://stackoverflow.com/qu
         # Ref for Completely open API - https://github.com/siznax/wptools/wiki
         # Ref for try-except blocks - https://wiki.python.org/moin/HandlingExceptions
         # Ref for Reading and Writing JSON to a file in Python - https://stackabuse.com/readi
         # Now we are going to create a DataFrame from the tweet_json.txt file:
         my_demo_list = []
         with open('tweet_json.txt', encoding='utf-8') as json_file:
             all_data = json.load(json_file)
             for each_dictionary in all_data:
                 tweet_id = each_dictionary['id']
                 whole_tweet = each_dictionary['text']
                 only_url = whole_tweet[whole_tweet.find('https'):]
                 favorite_count = each_dictionary['favorite_count']
                 retweet_count = each_dictionary['retweet_count']
                 created_at = each_dictionary['created_at']
                 whole_source = each_dictionary['source']
                 only_device = whole_source[whole_source.find('rel="nofollow">') + 15:-4]
                 source = only_device
                 retweeted_status = each_dictionary['retweeted_status'] = each_dictionary.get(
                 if retweeted_status == 'Original tweet':
                     url = only_url
                 else:
                     retweeted_status = 'This is a retweet'
                     url = 'This is a retweet'

                 my_demo_list.append({'tweet_id': str(tweet_id),
                                      'favorite_count': int(favorite_count),
                                      'retweet_count': int(retweet_count),
                                      'url': url,
                                      'created_at': created_at,
                                      'source': source,
                                      'retweeted_status': retweeted_status,
                                     })
                 tweet_json = pd.DataFrame(my_demo_list, columns = ['tweet_id', 'favorite_coun
                                                                    'retweet_count', 'created_at',
                                                                    'source', 'retweeted_status',
```

2 Data Wrangling

2.1 Assess

Identify Quality and tidiness issues visually and programmatically. Assess ,Clean and Test at least (8) Quality issues and at least (2) tidiness issues

```
In [14]: twitter_archive.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                      2356 non-null int64
in_reply_to_status_id         78 non-null float64
in_reply_to_user_id           78 non-null float64
timestamp                     2356 non-null object
source                        2356 non-null object
text                          2356 non-null object
retweeted_status_id           181 non-null float64
retweeted_status_user_id      181 non-null float64
retweeted_status_timestamp    181 non-null object
expanded_urls                 2297 non-null object
rating_numerator              2356 non-null int64
rating_denominator            2356 non-null int64
name                          2356 non-null object
doggo                         2356 non-null object
floofer                       2356 non-null object
pupper                        2356 non-null object
puppo                         2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB
```

1. What the hell is doggo, floofer , pupper and puppo maybe it should be in integer or boolean
2. Not all have same number of entries , a lot of null values

```
In [15]: images.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
```

2.1.1 Quality Issues - commonly referred to as dirty data. Dirty data has issues with its content. Issues that only requires you to remove rows or columns or to update values to fix them (and does not require you to change the structure of the columns) are hence should be considered Quality issues.

In twitter_archive table, there are a lot of names that should not be included in the Names Column. We can cross-reference the text data with Names Column.

retweet_status : this variable provide the source tweet which provides us information about the status of the tweets and can be use to identify tweets that are retweets.

In twitter_archive table , Date and Time columns need to be converted to datetime objects.

In twitter_archive table , drop the inessential columns to reduce confusion.

The date, time and stage columns should be near the front columns for easy read in twitter_archive_clean

tweet_id has an integer datatype when it is not numeric in twitter_archive_clean

date and time columns has to change the datatypes to time-related datatypes in twitter_archive_clean

If the rating_numerator or rating_denominator have decimal values , it could not work in twitter_archive_clean

In tweet_json table, rename the id column to "tweet_id" to match the other 2 tables for merging purposes.

In twitter_json table , convert id column from a number to a string

Inessential underscore in p1,p2,p3 columns in images table

Some of the text in p1, p2, p3 columns has big letter and some has small letter. Making the data harder to analyse

The datatypes of tweet_id in twitter_archive_clean and tweet_json_clean is string while for images_clean is integer

The 3 datasets are discrete and confusing when we need to analyze the different correlation in each dataset

There is time info in the Time column that is unessential when the data is being retrieved using Twitter API

Inessential column called 'Unnamed:0'

2.1.2 Tidiness Issues - commonly referred to as messy data. Messy data has issues with its structure.

In twitter_archive table, each dog stage should forms a columns not few different columns for doggo, fluffer, pupper, puppo .

In twitter_archive table, the date and time are put into one column instead it should be split into two different columns

2.2 Clean

```
In [16]: # Copy the dataframes to avoid from causing unwanted data loss in the original datafr
         twitter_archive_clean = twitter_archive.copy()
         images_clean = images.copy()
         tweet_json_clean = tweet_json.copy()
```

Problem In twitter_archive table, there are a lot of names that should not be included in the Names Column. We can cross-reference the text data with Names Column.

Define Replace the name with correct names or replace it as a None value

Code

```
In [17]: # Replace the name with correct names or replace it as a None value (Based on your op
         # Drop rather than replace, try dig deeper to understand later)

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.repla

         twitter_archive_clean['name'].replace('the' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('light' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('life' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('an' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('a' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('by' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('actually' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('just' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('getting' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('infuriating' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('old' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('all' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('this' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('very' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('mad' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('not' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('one' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('my' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('O' , "O'Malley" , inplace = True)
         twitter_archive_clean['name'].replace('quite' , 'None' , inplace = True)
         twitter_archive_clean['name'].replace('such' , 'None' , inplace = True)
```

Test

```
In [18]: twitter_archive_clean.head(5)
```

Out[18]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
0	892420643555336193	NaN	NaN	
1	892177421306343426	NaN	NaN	
2	891815181378084864	NaN	NaN	
3	891689557279858688	NaN	NaN	
4	891327558926688256	NaN	NaN	

	timestamp	\
0	2017-08-01 16:23:56 +0000	
1	2017-08-01 00:17:27 +0000	

```
2  2017-07-31 00:18:03 +0000
3  2017-07-30 15:58:51 +0000
4  2017-07-29 16:00:24 +0000

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

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

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

                                       expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                13
1  https://twitter.com/dog_rates/status/892177421...                13
2  https://twitter.com/dog_rates/status/891815181...                12
3  https://twitter.com/dog_rates/status/891689557...                13
4  https://twitter.com/dog_rates/status/891327558...                12

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

Problem retweet_status : this variable provide the source tweet which provides us information about the status of the tweets and can be use to identify tweets that are retweets.

Define Identify and drop the tweets that have a retweeted_status because those are not original

Code

```
In [19]: # Identify the tweets that have a retweeted_status columns
         twitter_archive_clean['retweeted_status_id'].head(5)
```

```
Out[19]: 0   NaN
         1   NaN
         2   NaN
         3   NaN
         4   NaN
         Name: retweeted_status_id, dtype: float64

In [20]: # Show the tweets that have a retweeted_status columns

         # Ref : https://pandas.pydata.org/pandas-docs/version/0.23.4/generated/pandas.isnull.
         twitter_archive_clean[twitter_archive_clean['retweeted_status_id'].isnull() == False]
```

Out[20]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
19	888202515573088257	NaN	NaN	
32	886054160059072513	NaN	NaN	
36	885311592912609280	NaN	NaN	
68	879130579576475649	NaN	NaN	
73	878404777348136964	NaN	NaN	
74	878316110768087041	NaN	NaN	
78	877611172832227328	NaN	NaN	
91	874434818259525634	NaN	NaN	
95	873697596434513921	NaN	NaN	
97	873337748698140672	NaN	NaN	
101	872668790621863937	NaN	NaN	
109	871166179821445120	NaN	NaN	
118	869988702071779329	NaN	NaN	
124	868639477480148993	NaN	NaN	
130	867072653475098625	NaN	NaN	
132	866816280283807744	NaN	NaN	
137	866094527597207552	NaN	NaN	
146	863471782782697472	NaN	NaN	
155	861769973181624320	NaN	NaN	
159	860981674716409858	NaN	NaN	
160	860924035999428608	NaN	NaN	
165	860177593139703809	NaN	NaN	
171	858860390427611136	NaN	NaN	
180	857062103051644929	NaN	NaN	
182	856602993587888130	NaN	NaN	
185	856330835276025856	NaN	NaN	
194	855245323840757760	NaN	NaN	
195	855138241867124737	NaN	NaN	
204	852936405516943360	NaN	NaN	
211	851953902622658560	NaN	NaN	
...	...	...	...	
784	775096608509886464	NaN	NaN	
794	773336787167145985	NaN	NaN	
800	772615324260794368	NaN	NaN	
811	771171053431250945	NaN	NaN	

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

                      timestamp  \
19    2017-07-21 01:02:36 +0000
32    2017-07-15 02:45:48 +0000
36    2017-07-13 01:35:06 +0000
68    2017-06-26 00:13:58 +0000
73    2017-06-24 00:09:53 +0000
74    2017-06-23 18:17:33 +0000
78    2017-06-21 19:36:23 +0000
91    2017-06-13 01:14:41 +0000
95    2017-06-11 00:25:14 +0000
97    2017-06-10 00:35:19 +0000
101   2017-06-08 04:17:07 +0000
109   2017-06-04 00:46:17 +0000
118   2017-05-31 18:47:24 +0000
124   2017-05-28 01:26:04 +0000
130   2017-05-23 17:40:04 +0000
132   2017-05-23 00:41:20 +0000
137   2017-05-21 00:53:21 +0000
146   2017-05-13 19:11:30 +0000
155   2017-05-09 02:29:07 +0000
159   2017-05-06 22:16:42 +0000
```

```
160   2017-05-06 18:27:40 +0000
165   2017-05-04 17:01:34 +0000
171   2017-05-01 01:47:28 +0000
180   2017-04-26 02:41:43 +0000
182   2017-04-24 20:17:23 +0000
185   2017-04-24 02:15:55 +0000
194   2017-04-21 02:22:29 +0000
195   2017-04-20 19:16:59 +0000
204   2017-04-14 17:27:40 +0000
211   2017-04-12 00:23:33 +0000
...                        ...
784   2016-09-11 22:20:06 +0000
794   2016-09-07 01:47:12 +0000
800   2016-09-05 02:00:22 +0000
811   2016-09-01 02:21:21 +0000
815   2016-08-31 15:19:06 +0000
818   2016-08-30 22:04:05 +0000
822   2016-08-29 03:00:36 +0000
826   2016-08-27 00:47:53 +0000
829   2016-08-25 20:35:48 +0000
833   2016-08-24 21:02:45 +0000
841   2016-08-20 05:08:29 +0000
847   2016-08-18 01:03:45 +0000
860   2016-08-10 00:16:21 +0000
868   2016-08-06 02:27:27 +0000
872   2016-08-05 01:19:35 +0000
885   2016-08-01 16:43:19 +0000
890   2016-07-31 01:50:18 +0000
895   2016-07-29 22:53:27 +0000
908   2016-07-26 00:08:05 +0000
911   2016-07-25 15:26:30 +0000
926   2016-07-18 03:06:01 +0000
937   2016-07-13 18:42:44 +0000
943   2016-07-12 03:11:42 +0000
949   2016-07-11 01:11:51 +0000
1012  2016-06-27 01:37:04 +0000
1023  2016-06-25 01:52:36 +0000
1043  2016-06-17 16:01:16 +0000
1242  2016-03-21 19:31:59 +0000
2259  2015-11-20 03:51:52 +0000
2260  2015-11-20 03:51:47 +0000

                                                  source  \
19    <a href="http://twitter.com/download/iphone" r...
32    <a href="http://twitter.com/download/iphone" r...
36    <a href="http://twitter.com/download/iphone" r...
68    <a href="http://twitter.com/download/iphone" r...
73    <a href="http://twitter.com/download/iphone" r...
```

```
74    <a href="http://twitter.com/download/iphone" r...
78    <a href="http://twitter.com/download/iphone" r...
91    <a href="http://twitter.com/download/iphone" r...
95    <a href="http://twitter.com/download/iphone" r...
97    <a href="http://twitter.com/download/iphone" r...
101   <a href="http://twitter.com/download/iphone" r...
109   <a href="http://twitter.com/download/iphone" r...
118   <a href="http://twitter.com/download/iphone" r...
124   <a href="http://twitter.com/download/iphone" r...
130   <a href="http://twitter.com/download/iphone" r...
132   <a href="http://twitter.com/download/iphone" r...
137   <a href="http://twitter.com/download/iphone" r...
146   <a href="http://twitter.com/download/iphone" r...
155   <a href="http://twitter.com/download/iphone" r...
159   <a href="http://twitter.com/download/iphone" r...
160   <a href="http://twitter.com/download/iphone" r...
165   <a href="http://twitter.com/download/iphone" r...
171   <a href="http://twitter.com/download/iphone" r...
180   <a href="http://twitter.com/download/iphone" r...
182   <a href="http://twitter.com/download/iphone" r...
185   <a href="http://twitter.com/download/iphone" r...
194   <a href="http://twitter.com/download/iphone" r...
195   <a href="http://twitter.com/download/iphone" r...
204   <a href="http://twitter.com/download/iphone" r...
211   <a href="http://twitter.com/download/iphone" r...
...                                                  ...
784   <a href="http://twitter.com/download/iphone" r...
794   <a href="http://twitter.com/download/iphone" r...
800   <a href="http://twitter.com/download/iphone" r...
811   <a href="http://twitter.com/download/iphone" r...
815   <a href="http://twitter.com/download/iphone" r...
818   <a href="http://twitter.com/download/iphone" r...
822   <a href="http://twitter.com/download/iphone" r...
826   <a href="http://twitter.com/download/iphone" r...
829   <a href="http://twitter.com/download/iphone" r...
833   <a href="http://twitter.com/download/iphone" r...
841   <a href="http://twitter.com/download/iphone" r...
847   <a href="http://twitter.com/download/iphone" r...
860   <a href="http://twitter.com/download/iphone" r...
868   <a href="http://twitter.com/download/iphone" r...
872   <a href="http://twitter.com/download/iphone" r...
885   <a href="http://twitter.com/download/iphone" r...
890   <a href="http://twitter.com/download/iphone" r...
895   <a href="http://twitter.com/download/iphone" r...
908   <a href="http://twitter.com/download/iphone" r...
911   <a href="http://twitter.com/download/iphone" r...
926   <a href="http://twitter.com/download/iphone" r...
937   <a href="http://twitter.com/download/iphone" r...
```

```
943   <a href="http://twitter.com/download/iphone" r...
949   <a href="http://twitter.com/download/iphone" r...
1012  <a href="http://twitter.com/download/iphone" r...
1023  <a href="http://twitter.com/download/iphone" r...
1043  <a href="http://twitter.com/download/iphone" r...
1242  <a href="http://twitter.com/download/iphone" r...
2259  <a href="http://twitter.com" rel="nofollow">Tw...
2260  <a href="http://twitter.com" rel="nofollow">Tw...

                                                  text  retweeted_status_id  \
19    RT @dog_rates: This is Canela. She attempted s...         8.874740e+17
32    RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...         8.860537e+17
36    RT @dog_rates: This is Lilly. She just paralle...         8.305833e+17
68    RT @dog_rates: This is Emmy. She was adopted t...         8.780576e+17
73    RT @dog_rates: Meet Shadow. In an attempt to r...         8.782815e+17
74    RT @dog_rates: Meet Terrance. He's being yelle...         6.690004e+17
78    RT @rachel2195: @dog_rates the boyfriend and h...         8.768508e+17
91    RT @dog_rates: This is Coco. At first I though...         8.663350e+17
95    RT @dog_rates: This is Walter. He won't start ...         8.688804e+17
97    RT @dog_rates: This is Sierra. She's one preci...         8.732138e+17
101   RT @loganamnosis: Penelope here is doing me qu...         8.726576e+17
109   RT @dog_rates: This is Dawn. She's just checki...         8.410770e+17
118   RT @dog_rates: We only rate dogs. This is quit...         8.591970e+17
124   RT @dog_rates: Say hello to Cooper. His expres...         8.685523e+17
130   RT @rachaeleasler: these @dog_rates hats are 1...         8.650134e+17
132   RT @dog_rates: This is Jamesy. He gives a kiss...         8.664507e+17
137   RT @dog_rates: Here's a pupper before and afte...         8.378202e+17
146   RT @dog_rates: Say hello to Quinn. She's quite...         8.630625e+17
155   RT @dog_rates: "Good afternoon class today we'...         8.066291e+17
159   RT @dog_rates: Meet Lorenzo. He's an avid nift...         8.605638e+17
160   RT @tallylott: h*ckin adorable promposal. 13/1...         8.609145e+17
165   RT @dog_rates: Ohboyohboyohboyohboyohboyohboyo...         7.616730e+17
171   RT @dog_rates: Meet Winston. He knows he's a l...         8.395493e+17
180   RT @AaronChewning: First time wearing my @dog_...         8.570611e+17
182   RT @dog_rates: This is Luna. It's her first ti...         8.447048e+17
185   RT @Jenna_Marbles: @dog_rates Thanks for ratin...         8.563302e+17
194   RT @dog_rates: Meet George. He looks slightly ...         8.421635e+17
195   RT @frasercampbell_: oh my... what's that... b...         8.551225e+17
204   RT @dog_rates: I usually only share these on F...         8.316501e+17
211   RT @dog_rates: This is Astrid. She's a guide d...         8.293743e+17
...                                                 ...                  ...
784   RT @dog_rates: After so many requests, this is...         7.403732e+17
794   RT @dog_rates: Meet Fizz. She thinks love is a...         7.713808e+17
800   RT @dog_rates: This is Gromit. He's pupset bec...         7.652221e+17
811   RT @dog_rates: This is Frankie. He's wearing b...         6.733201e+17
815   RT @katieornah: @dog_rates learning a lot at c...         7.710021e+17
818   RT @dog_rates: Here's a doggo blowing bubbles...          7.392382e+17
822   RT @dog_rates: This is just downright precious...         7.410673e+17
```

```
826   RT @dog_rates: Ever seen a dog pet another dog...         7.069045e+17
829   RT @dog_rates: When it's Janet from accounting...         7.001438e+17
833   RT @dog_rates: This is Nollie. She's waving at...         7.399792e+17
841   RT @dog_rates: We only rate dogs... this is a ...         7.599238e+17
847   RT @dog_rates: This is Colby. He's currently r...         7.258423e+17
860   RT @dog_rates: Meet Eve. She's a raging alcoho...         6.732953e+17
868   RT @dog_rates: "Tristan do not speak to me wit...         6.853251e+17
872   RT @dog_rates: Oh. My. God. 13/10 magical af h...         7.116948e+17
885   RT @hownottodraw: The story/person behind @dog...         7.601538e+17
890   RT @dog_rates: This... is a Tyrannosaurus rex...          7.395441e+17
895   RT @dog_rates: AT DAWN...\nWE RIDE\n\n11/10 ht...         6.703191e+17
908   RT @dog_rates: This is Chompsky. He lives up t...         6.790626e+17
911   RT @jon_hill987: @dog_rates There is a cunning...         7.575971e+17
926   RT @dog_rates: This is Rubio. He has too much ...         6.791584e+17
937   RT @dog_rates: This is Carly. She's actually 2...         6.815232e+17
943   RT @dog_rates: HEY PUP WHAT'S THE PART OF THE ...         6.835159e+17
949   RT @dog_rates: Everyone needs to watch this. 1...         6.753544e+17
1012  RT @dog_rates: This pupper killed this great w...         7.047611e+17
1023  RT @dog_rates: This is Shaggy. He knows exactl...         6.678667e+17
1043  RT @dog_rates: Extremely intelligent dog here...          6.671383e+17
1242  RT @twitter: @dog_rates Awesome Tweet! 12/10. ...         7.119983e+17
2259  RT @dogratingrating: Exceptional talent. Origi...         6.675487e+17
2260  RT @dogratingrating: Unoriginal idea. Blatant ...         6.675484e+17

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

```
180               5.870972e+07  2017-04-26 02:37:47 +0000
182               4.196984e+09  2017-03-23 00:18:10 +0000
185               6.669901e+07  2017-04-24 02:13:14 +0000
194               4.196984e+09  2017-03-16 00:00:07 +0000
195               7.475543e+17  2017-04-20 18:14:33 +0000
204               4.196984e+09  2017-02-14 23:43:18 +0000
211               4.196984e+09  2017-02-08 17:00:26 +0000
...                        ...                        ...
784               4.196984e+09  2016-06-08 02:41:38 +0000
794               4.196984e+09  2016-09-01 16:14:48 +0000
800               4.196984e+09  2016-08-15 16:22:20 +0000
811               4.196984e+09  2015-12-06 01:56:44 +0000
815               1.732729e+09  2016-08-31 15:10:07 +0000
818               4.196984e+09  2016-06-04 23:31:25 +0000
822               4.196984e+09  2016-06-10 00:39:48 +0000
826               4.196984e+09  2016-03-07 18:09:06 +0000
829               4.196984e+09  2016-02-18 02:24:13 +0000
833               4.196984e+09  2016-06-07 00:36:02 +0000
841               4.196984e+09  2016-08-01 01:28:46 +0000
847               4.196984e+09  2016-04-29 00:21:01 +0000
860               4.196984e+09  2015-12-06 00:17:55 +0000
868               4.196984e+09  2016-01-08 05:00:14 +0000
872               4.196984e+09  2016-03-20 23:23:54 +0000
885               1.950368e+08  2016-08-01 16:42:51 +0000
890               4.196984e+09  2016-06-05 19:47:03 +0000
895               4.196984e+09  2015-11-27 19:11:49 +0000
908               4.196984e+09  2015-12-21 22:15:18 +0000
911               2.804798e+08  2016-07-25 15:23:28 +0000
926               4.196984e+09  2015-12-22 04:35:49 +0000
937               4.196984e+09  2015-12-28 17:12:42 +0000
943               4.196984e+09  2016-01-03 05:11:12 +0000
949               4.196984e+09  2015-12-11 16:40:19 +0000
1012              4.196984e+09  2016-03-01 20:11:59 +0000
1023              4.196984e+09  2015-11-21 00:46:50 +0000
1043              4.196984e+09  2015-11-19 00:32:12 +0000
1242              7.832140e+05  2016-03-21 19:29:52 +0000
2259              4.296832e+09  2015-11-20 03:43:06 +0000
2260              4.296832e+09  2015-11-20 03:41:59 +0000

                                                expanded_urls  rating_numerator  \
19    https://twitter.com/dog_rates/status/887473957...                 13
32    https://twitter.com/dog_rates/status/886053434...                 12
36    https://twitter.com/dog_rates/status/830583320...                 13
68    https://twitter.com/dog_rates/status/878057613...                 14
73    https://www.gofundme.com/3yd6y1c,https://twitt...                 13
74    https://twitter.com/dog_rates/status/669000397...                 11
78    https://twitter.com/rachel2195/status/87685077...                 14
91    https://twitter.com/dog_rates/status/866334964...                 12
```

95	https://twitter.com/dog_rates/status/868880397...	14
97	https://www.gofundme.com/help-my-baby-sierra-g...	12
101	https://twitter.com/loganamnosis/status/872657...	14
109	https://twitter.com/dog_rates/status/841077006...	12
118	https://twitter.com/dog_rates/status/859196978...	12
124	https://www.gofundme.com/3ti3nps,https://twitt...	12
130	https://twitter.com/rachaeleasler/status/86501...	13
132	https://twitter.com/dog_rates/status/866450705...	13
137	https://twitter.com/dog_rates/status/837820167...	12
146	https://www.gofundme.com/helpquinny,https://tw...	13
155	https://twitter.com/dog_rates/status/806629075...	13
159	https://www.gofundme.com/help-lorenzo-beat-can...	13
160	https://twitter.com/tallylott/status/860914485...	13
165	https://twitter.com/dog_rates/status/761672994...	10
171	https://twitter.com/dog_rates/status/839549326...	12
180	https://twitter.com/AaronChewning/status/85706...	13
182	https://twitter.com/dog_rates/status/844704788...	13
185	NaN	14
194	https://twitter.com/dog_rates/status/842163532...	12
195	https://twitter.com/frasercampbell_/status/855...	14
204	http://www.gofundme.com/bluethewhitehusky,http...	13
211	https://twitter.com/dog_rates/status/829374341...	13
...	...	...
784	https://twitter.com/dog_rates/status/740373189...	9
794	https://twitter.com/dog_rates/status/771380798...	11
800	https://twitter.com/dog_rates/status/765222098...	10
811	https://twitter.com/dog_rates/status/673320132...	11
815	https://twitter.com/katieornah/status/77100213...	12
818	https://twitter.com/dog_rates/status/739238157...	13
822	https://twitter.com/dog_rates/status/741067306...	12
826	https://vine.co/v/iXQAm5Lrgrh,https://vine.co/...	13
829	https://twitter.com/dog_rates/status/700143752...	10
833	https://twitter.com/dog_rates/status/739979191...	12
841	https://twitter.com/dog_rates/status/759923798...	10
847	https://twitter.com/dog_rates/status/725842289...	12
860	https://twitter.com/dog_rates/status/673295268...	8
868	https://twitter.com/dog_rates/status/685325112...	10
872	https://twitter.com/dog_rates/status/711694788...	13
885	https://weratedogs.com/pages/about-us,https://...	11
890	https://twitter.com/dog_rates/status/739544079...	10
895	https://twitter.com/dog_rates/status/670319130...	11
908	https://twitter.com/dog_rates/status/679062614...	11
911	https://twitter.com/jon_hill987/status/7575971...	11
926	https://twitter.com/dog_rates/status/679158373...	11
937	https://twitter.com/dog_rates/status/681523177...	12
943	https://vine.co/v/ibvnzrauFuV,https://vine.co/...	11
949	https://twitter.com/dog_rates/status/675354435...	13
1012	https://twitter.com/dog_rates/status/704761120...	13

```
1023  https://twitter.com/dog_rates/status/667866724...               10
1043  https://twitter.com/dog_rates/status/667138269...               10
1242  https://twitter.com/twitter/status/71199827977...               12
2259  https://twitter.com/dogratingrating/status/667...               12
2260  https://twitter.com/dogratingrating/status/667...                5

      rating_denominator      name  doggo floofer  pupper puppo
19                    10    Canela   None    None    None  None
32                    10      None   None    None    None  None
36                    10     Lilly   None    None    None  None
68                    10      Emmy   None    None    None  None
73                    10    Shadow   None    None    None  None
74                    10  Terrance   None    None    None  None
78                    10      None   None    None  pupper  None
91                    10      Coco   None    None    None  None
95                    10    Walter   None    None    None  None
97                    10    Sierra   None    None  pupper  None
101                   10      None   None    None    None  None
109                   10      Dawn   None    None    None  None
118                   10      None   None    None    None  None
124                   10    Cooper   None    None    None  None
130                   10      None   None    None    None  None
132                   10    Jamesy   None    None  pupper  None
137                   10      None   None    None  pupper  None
146                   10     Quinn   None    None    None  None
155                   10      None   None    None    None  None
159                   10   Lorenzo   None    None    None  None
160                   10      None   None    None    None  None
165                   10      None   None    None    None  None
171                   10   Winston   None    None    None  None
180                   10      None   None    None    None  None
182                   10      Luna   None    None    None  None
185                   10      None   None    None    None  None
194                   10    George   None    None    None  None
195                   10      None   None    None    None  None
204                   10      None   None    None    None  None
211                   10    Astrid  doggo    None    None  None
...                  ...       ...    ...     ...     ...   ...
784                   11      None   None    None    None  None
794                   10      Fizz   None    None    None  None
800                   10    Gromit   None    None    None  None
811                   10   Frankie   None    None    None  None
815                   10      None   None    None  pupper  None
818                   10      None  doggo    None    None  None
822                   10      None  doggo    None  pupper  None
826                   10      None   None    None    None  None
829                   10      None   None    None  pupper  None
833                   10    Nollie   None    None    None  None
```

```
841                  10       None    None    None    None  None
847                  10      Colby    None    None    None  None
860                  10        Eve    None    None  pupper  None
868                  10       None    None    None    None  None
872                  10       None    None    None    None  None
885                  10       None    None    None    None  None
890                  10       None    None    None    None  None
895                  10       None    None    None    None  None
908                  10    Chompsky    None    None    None  None
911                  10       None    None    None  pupper  None
926                  10      Rubio    None    None    None  None
937                  10      Carly    None    None    None  None
943                  10       None    None    None    None  None
949                  10       None    None    None    None  None
1012                 10       None    None    None  pupper  None
1023                 10     Shaggy    None    None    None  None
1043                 10       None    None    None    None  None
1242                 10       None    None    None    None  None
2259                 10       None    None    None    None  None
2260                 10       None    None    None    None  None

[181 rows x 17 columns]
```

In [21]:
```
# Drop the tweets that have a retweeted_status_id

# Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.drop.

twitter_archive_clean.drop(twitter_archive_clean[twitter_archive_clean['retweeted_sta
```

Test

In [22]:
```
# Check the tweets that have retweeted_status
twitter_archive_clean.head(5)
```

Out[22]:
```
             tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
0  892420643555336193                    NaN                  NaN
1  892177421306343426                    NaN                  NaN
2  891815181378084864                    NaN                  NaN
3  891689557279858688                    NaN                  NaN
4  891327558926688256                    NaN                  NaN

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

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

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

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

                                       expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                13
1  https://twitter.com/dog_rates/status/892177421...                13
2  https://twitter.com/dog_rates/status/891815181...                12
3  https://twitter.com/dog_rates/status/891689557...                13
4  https://twitter.com/dog_rates/status/891327558...                12

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

Problem In twitter_archive table , Date and Time columns need to be converted to datetime objects.

Define Convert the Date and Time datatypes from String to Datetime objects

Code

```
In [23]: # import datetime and timedelta libraries from datetime

         # Ref : https://docs.python.org/2/library/datetime.html
         from datetime import datetime, timedelta
```

```
# Check the timestamp datatypes
twitter_archive_clean['timestamp'].dtypes
```

```
Out[23]: dtype('O')
```

```
In [24]: # Convert the Date and time datatypes from String to Datetime objects

         # Ref : https://pandas.pydata.org/pandas-docs/version/0.20/generated/pandas.to_dateti

         twitter_archive_clean['timestamp'] = pd.to_datetime(twitter_archive_clean['timestamp']

         # Create two more columns for Date and Time respectively for easy data exploration in

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.apply
         # Test this if free , Ref : https://stackoverflow.com/questions/35595710/splitting-ti
         # Ref for lambda and time is needed

         twitter_archive_clean['date'] = twitter_archive_clean['timestamp'].apply(lambda time
         twitter_archive_clean['time'] = twitter_archive_clean['timestamp'].apply(lambda time
```

Test

```
In [25]: # Check to make sure there are two additional dataset
         twitter_archive_clean.head(5)
```

```
Out[25]:             tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0  892420643555336193                    NaN                  NaN
         1  892177421306343426                    NaN                  NaN
         2  891815181378084864                    NaN                  NaN
         3  891689557279858688                    NaN                  NaN
         4  891327558926688256                    NaN                  NaN

                    timestamp                                             source  \
         0 2017-08-01 16:23:56  <a href="http://twitter.com/download/iphone" r...
         1 2017-08-01 00:17:27  <a href="http://twitter.com/download/iphone" r...
         2 2017-07-31 00:18:03  <a href="http://twitter.com/download/iphone" r...
         3 2017-07-30 15:58:51  <a href="http://twitter.com/download/iphone" r...
         4 2017-07-29 16:00:24  <a href="http://twitter.com/download/iphone" r...

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

            retweeted_status_user_id retweeted_status_timestamp  \
         0                       NaN                        NaN
```

```
1                       NaN                       NaN
2                       NaN                       NaN
3                       NaN                       NaN
4                       NaN                       NaN

                                          expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                 13
1  https://twitter.com/dog_rates/status/892177421...                 13
2  https://twitter.com/dog_rates/status/891815181...                 12
3  https://twitter.com/dog_rates/status/891689557...                 13
4  https://twitter.com/dog_rates/status/891327558...                 12

   rating_denominator      name doggo floofer pupper puppo        date   time
0                  10   Phineas  None    None   None  None  08-01-2017  16:23
1                  10     Tilly  None    None   None  None  08-01-2017  00:17
2                  10    Archie  None    None   None  None  07-31-2017  00:18
3                  10     Darla  None    None   None  None  07-30-2017  15:58
4                  10  Franklin  None    None   None  None  07-29-2017  16:00
```

In [26]:
```
# Check to see the datatypes of the newly added dataset and timestamp's datatypes
twitter_archive_clean.dtypes
```

Out[26]:
```
tweet_id                             int64
in_reply_to_status_id              float64
in_reply_to_user_id                float64
timestamp                   datetime64[ns]
source                              object
text                                object
retweeted_status_id                float64
retweeted_status_user_id           float64
retweeted_status_timestamp          object
expanded_urls                       object
rating_numerator                     int64
rating_denominator                   int64
name                                object
doggo                               object
floofer                             object
pupper                              object
puppo                               object
date                                object
time                                object
dtype: object
```

Problem In twitter_archive table, Combine each dog stage column into a single column named "stage"

Define Combine each dog stage column into a single column named "stage"

Code

```
In [27]: twitter_archive_clean['stage'] = twitter_archive_clean[['doggo', 'floofer','pupper','p
         twitter_archive_clean['stage'].replace("NoneNoneNoneNone" , "None" , inplace = True)
         twitter_archive_clean['stage'].replace("doggoNoneNoneNone" , "doggo" , inplace = True)
         twitter_archive_clean['stage'].replace("NoneflooferNoneNone" , "floofer" , inplace = T
         twitter_archive_clean['stage'].replace("NoneNonepupperNone" , "pupper" , inplace = Tru
         twitter_archive_clean['stage'].replace("NoneNoneNonepuppo" , "puppo" , inplace = True)
```

Test

```
In [28]: twitter_archive_clean.head(5)
```

Out[28]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	retweeted_status_id	retweeted_status_user_id	retweeted_status_timestamp	expanded_urls	rating_numerator	rating_denominator	name	doggo	floofer	pupper	puppo	date	time
0	892420643555336193	NaN	NaN	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...	This is Phineas. He's a mystical boy. Only eve...	NaN	NaN	NaN	https://twitter.com/dog_rates/status/892420643...	13	10	None	None	None	None	None		
1	892177421306343426	NaN	NaN	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you...	NaN	NaN	NaN	https://twitter.com/dog_rates/status/892177421...	13	10	None	None	None	None	None		
2	891815181378084864	NaN	NaN	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	NaN	NaN	NaN	https://twitter.com/dog_rates/status/891815181...	12	10	None	None	None	None	None		
3	891689557279858688	NaN	NaN	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...	This is Darla. She commenced a snooze mid meal...	NaN	NaN	NaN	https://twitter.com/dog_rates/status/891689557...	13	10	None	None	None	None	None		
4	891327558926688256	NaN	NaN	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...	This is Franklin. He would like you to stop ca...	NaN	NaN	NaN	https://twitter.com/dog_rates/status/891327558...	12	10	None	None	None	None	None		

```
0                  10    Phineas  None    None    None  None  08-01-2017  16:23
1                  10      Tilly  None    None    None  None  08-01-2017  00:17
2                  10     Archie  None    None    None  None  07-31-2017  00:18
3                  10      Darla  None    None    None  None  07-30-2017  15:58
4                  10   Franklin  None    None    None  None  07-29-2017  16:00

  stage
0  None
1  None
2  None
3  None
4  None
```

Problem In twitter_archive table , drop the inessential columns to reduce confusion.

Define Drop the inessential columns to reduce confusion

Code

```
In [29]: # Drop the inessential columns using the drop function
         twitter_archive_clean.drop(['timestamp' , 'retweeted_status_user_id', 'retweeted_statu
                                     'in_reply_to_status_id' , 'in_reply_to_user_id' , 'in_reply_to_s
                                     'source' , 'doggo' , 'floofer' , 'pupper' , 'puppo' , 'text'] ,
```

Test

```
In [30]: # Check if all the inessential columns have been dropped
         twitter_archive_clean.head(5)
```

```
Out[30]:             tweet_id  rating_numerator  rating_denominator      name  \
         0  892420643555336193                13                  10   Phineas
         1  892177421306343426                13                  10     Tilly
         2  891815181378084864                12                  10    Archie
         3  891689557279858688                13                  10     Darla
         4  891327558926688256                12                  10  Franklin

                 date   time stage
         0  08-01-2017  16:23  None
         1  08-01-2017  00:17  None
         2  07-31-2017  00:18  None
         3  07-30-2017  15:58  None
         4  07-29-2017  16:00  None
```

Problem The date, time and stage columns should be near the front columns for easy read in twitter_archive_clean

Define Change the order of the columns making sure date, time and stage columns are near the front

Code

```
In [31]: # Change the order of the columns and reindex the columns

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.reind
         columnHeaders = ['tweet_id' , 'date' , 'time' , 'name' , 'stage' , 'rating_numerator'

         twitter_archive_clean = twitter_archive_clean.reindex(columns = columnHeaders)
```

Test

```
In [32]: # Check the reindex and order of the columns
         twitter_archive_clean.head(5)
```

Out[32]:

	tweet_id	date	time	name	stage	rating_numerator	\
0	892420643555336193	08-01-2017	16:23	Phineas	None	13	
1	892177421306343426	08-01-2017	00:17	Tilly	None	13	
2	891815181378084864	07-31-2017	00:18	Archie	None	12	
3	891689557279858688	07-30-2017	15:58	Darla	None	13	
4	891327558926688256	07-29-2017	16:00	Franklin	None	12	

	rating_denominator
0	10
1	10
2	10
3	10
4	10

Problem tweet_id has an integer datatype when it is not numeric in twitter_archive_clean

Define tweet_id columns has to change datatype from integer to string datatype in twitter_archive_clean

Code

```
In [33]: # Check the datatypes of tweet_id columns
         # Ref : https://stackoverflow.com/questions/29530232/how-to-check-if-any-value-is-nan
         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.dtype

         twitter_archive_clean['tweet_id'].dtypes
```

```
Out[33]: dtype('int64')
```

```
In [34]: # Change the datatype of twitter_id from integer to string

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.astyp

         twitter_archive_clean['tweet_id'] = twitter_archive_clean.astype(str)
```

Test

```
In [35]: twitter_archive_clean['tweet_id'].dtypes
```

```
Out[35]: dtype('O')
```

Problem date and time columns has to change the datatypes to time-related datatypes in twitter_archive_clean

Define Change the datatypes of date and time columns to datetime dtypes

Code

```
In [36]: # Change the datatypes of date and time columns to datetime dtypes

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.to_datetime.htm

         twitter_archive_clean['date'] = pd.to_datetime(twitter_archive_clean['date'])
         twitter_archive_clean['time'] = pd.to_datetime(twitter_archive_clean['time'])
```

Test

```
In [37]: twitter_archive_clean['date'].dtypes
         twitter_archive_clean['time'].dtypes
```

```
Out[37]: dtype('<M8[ns]')
```

Problem If the rating_numerator or rating_denominator have decimal values , it could not work in twitter_archive_clean

Define Change the datatypes of rating_numerator and rating_denominator from integer dtypes to float dtypes

Code

```
In [38]: # Change the datatypes of rating_numerator and rating_denominator from integer dtypes

         twitter_archive_clean['rating_numerator'] = twitter_archive_clean['rating_numerator']
         twitter_archive_clean['rating_denominator'] = twitter_archive_clean['rating_denominato
```

Test

```
In [39]: twitter_archive_clean['rating_numerator'].dtypes
         twitter_archive_clean['rating_denominator'].dtypes
```

```
Out[39]: dtype('float64')
```

3 Tweet_json table

```
In [40]: tweet_json_clean.head(10)
```

Out[40]:

	tweet_id	favorite_count	retweet_count	created_at	source	retweeted_status	url
0	892420643555336193	38071	8327	Tue Aug 01 16:23:56 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/MgUWQ76dJU
1	892177421306343426	32681	6151	Tue Aug 01 00:17:27 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/aQFSeaCu9L
2	891815181378084864	24589	4069	Mon Jul 31 00:18:03 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/r0YlrsGCgy
3	891689557279858688	41429	8463	Sun Jul 30 15:58:51 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/tD36da7qLQ
4	891327558926688256	39599	9166	Sat Jul 29 16:00:24 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/0g0KMIVXZ3
5	891087950875897856	19883	3047	Sat Jul 29 00:08:17 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/xx5cilW0Dd
6	890971913173991426	11624	2021	Fri Jul 28 16:27:12 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/MV01Q820LT
7	890729181411237888	64233	18473	Fri Jul 28 00:22:40 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/hrcFOGi12V
8	890609185150312448	27322	4189	Thu Jul 27 16:25:51 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/UkrdQyoYxV
9	890240255349198849	31350	7220	Wed Jul 26 15:59:51 +0000 2017	Twitter for iPhone	Original tweet	https://t.co/l3TSS3o2M0

Problem In tweet_json table, rename the id column to "tweet_id" to match the other 2 tables for merging purposes.

Define Rename the 'id' in tweet_json to 'tweet_id' to match the other 2 tables for merging purposes

Code

```
In [41]: # Rename id to tweet_id

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.renam

         tweet_json_clean.rename(columns = {'id' : 'tweet_id'} , inplace = True)
```

Test

```
In [42]: # Check the columns 'id' has changed to 'tweet_id'
         tweet_json_clean.head(1)

Out[42]:              tweet_id  favorite_count  retweet_count  \
         0  892420643555336193           38071           8327

                               created_at              source retweeted_status  \
         0  Tue Aug 01 16:23:56 +0000 2017  Twitter for iPhone   Original tweet

                               url
         0  https://t.co/MgUWQ76dJU
```

Problem In twitter_json table , convert id column from a number to a string

Define Convert the datatypes of 'tweet_id' from integer to string dtypes in twitter_json table

Code

```
In [43]: # Check the datatypes of tweet_id

         tweet_json_clean['tweet_id'].dtypes

Out[43]: dtype('O')

In [44]: # Convert datatypes from integer to string

         tweet_json_clean['tweet_id'] = tweet_json_clean['tweet_id'].astype(str)
```

Test

```
In [45]: # Check the datatypes of tweet_id

         tweet_json_clean['tweet_id'].dtypes

Out[45]: dtype('O')
```

3.1 Images_clean df

```
In [46]: images_clean.head(10)
```

Out[46]:

	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

	img_num	p1	p1_conf	p1_dog	p2
0	1	Welsh_springer_spaniel	0.465074	True	collie
1	1	redbone	0.506826	True	miniature_pinscher
2	1	German_shepherd	0.596461	True	malinois
3	1	Rhodesian_ridgeback	0.408143	True	redbone
4	1	miniature_pinscher	0.560311	True	Rottweiler
5	1	Bernese_mountain_dog	0.651137	True	English_springer
6	1	box_turtle	0.933012	False	mud_turtle
7	1	chow	0.692517	True	Tibetan_mastiff
8	1	shopping_cart	0.962465	False	shopping_basket
9	1	miniature_poodle	0.201493	True	komondor

	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	0.156665	True	Shetland_sheepdog	0.061428	True
1	0.074192	True	Rhodesian_ridgeback	0.072010	True
2	0.138584	True	bloodhound	0.116197	True
3	0.360687	True	miniature_pinscher	0.222752	True
4	0.243682	True	Doberman	0.154629	True
5	0.263788	True	Greater_Swiss_Mountain_dog	0.016199	True
6	0.045885	False	terrapin	0.017885	False
7	0.058279	True	fur_coat	0.054449	False
8	0.014594	False	golden_retriever	0.007959	True
9	0.192305	True	soft-coated_wheaten_terrier	0.082086	True

Problem Inessential underscore in p1,p2,p3 columns in images table

Define Replace the '_' in p1, p2, p3 columns with spaces in images table

Code

```
In [47]: # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.str.repl
```

```
images_clean['p1'] = images_clean['p1'].str.replace('_' , ' ')
images_clean['p2'] = images_clean['p2'].str.replace('_' , ' ')
images_clean['p3'] = images_clean['p3'].str.replace('_' , ' ')
```

Test

```
In [48]: # Check the replacement of _ into empty space
         images_clean.head(5)
```

Out[48]:

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	666020888022790149	https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg	1	Welsh springer spaniel	0.465074	True	collie	0.156665	True	Shetland sheepdog	0.061428	True
1	666029285002620928	https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg	1	redbone	0.506826	True	miniature pinscher	0.074192	True	Rhodesian ridgeback	0.072010	True
2	666033412701032449	https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg	1	German shepherd	0.596461	True	malinois	0.138584	True	bloodhound	0.116197	True
3	666044226329800704	https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg	1	Rhodesian ridgeback	0.408143	True	redbone	0.360687	True	miniature pinscher	0.222752	True
4	666049248165822465	https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg	1	miniature pinscher	0.560311	True	Rottweiler	0.243682	True	Doberman	0.154629	True

Problem Some of the text in p1, p2, p3 columns has big letter and some has small letter. Making the data harder to analyse

Define Converts first character of each word to uppercase and remaining to lowercase in p1, p2 , p3 columns in images table

Code

```
In [49]: # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.str.titl

         images_clean['p1'] = images_clean['p1'].str.title()
         images_clean['p2'] = images_clean['p2'].str.title()
         images_clean['p3'] = images_clean['p3'].str.title()
```

Test

```
In [50]: images_clean.head(5)
```

```
Out[50]:              tweet_id                                          jpg_url  \
         0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
         1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
         2  666033412701032449  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
         3  666044226329800704  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
         4  666049248165822465  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

            img_num                     p1   p1_conf  p1_dog                 p2  \
         0        1  Welsh Springer Spaniel  0.465074    True             Collie
         1        1                Redbone  0.506826    True  Miniature Pinscher
         2        1         German Shepherd  0.596461    True            Malinois
         3        1     Rhodesian Ridgeback  0.408143    True             Redbone
         4        1      Miniature Pinscher  0.560311    True          Rottweiler

             p2_conf  p2_dog                   p3   p3_conf  p3_dog
         0  0.156665    True    Shetland Sheepdog  0.061428    True
         1  0.074192    True  Rhodesian Ridgeback  0.072010    True
         2  0.138584    True           Bloodhound  0.116197    True
         3  0.360687    True   Miniature Pinscher  0.222752    True
         4  0.243682    True             Doberman  0.154629    True
```

Problem The datatypes of tweet_id in twitter_archive_clean and tweet_json_clean is string while for images_clean is integer

Define Change the datatypes of tweet_id in images_clean into string

Code

```
In [51]: # Change the datatype
         images_clean['tweet_id'] = images_clean['tweet_id'].astype(str)
```

Test

```
In [52]: # Check the datatypes of tweet_id in images_clean
         images_clean['tweet_id'].dtypes
```

```
Out[52]: dtype('O')
```

Problem The 3 datasets are discrete and confusing when we need to analyze the different correlation in each dataset

Define Combine the three datasets into one datasets

Code

In [53]:
```
# Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.merge

# Merge twitter_archive_clean with tweet_json_clean and join on 'tweet_id' using the
# datasets and preserve the order in the left keys
tweets_merge = pd.merge(twitter_archive_clean , tweet_json_clean , on = 'tweet_id' , h
```

In [54]:
```
# Merge the images_clean and tweets_merge and join on 'tweet_id' using the intersecti
# and preserve the order in the left keys

tweets_merge = pd.merge(tweets_merge, images_clean , on = 'tweet_id' , how = 'inner')
```

Test

In [55]:
```
tweets_merge.head(5)
```

Out[55]:

	tweet_id	date	time	name	stage	rating_numerator	rating_denominator	favorite_count	retweet_count	created_at	...	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog
0	892420643555336193	2017-08-01	2019-01-04 16:23:00	Phineas	None	13.0	10.0	38071	8327	Tue Aug 01 16:23:56 +0000 2017	...	1	Orange	0.097049	False	Bagel	0.085851	False	Banana	0.076110	False
1	892177421306343426	2017-08-01	2019-01-04 00:17:00	Tilly	None	13.0	10.0	32681	6151	Tue Aug 01 00:17:27 +0000 2017	...	1	Chihuahua	0.323581	True	Pekinese	0.090647	True	Papillon	0.068957	True
2	891815181378084864	2017-07-31	2019-01-04 00:18:00	Archie	None	12.0	10.0	24589	4069	Mon Jul 31 00:18:03 +0000 2017	...	1	Chihuahua	0.716012	True	Malamute	0.078253	True	Kelpie	0.031379	True
3	891689557279858688	2017-07-30	2019-01-04 15:58:00	Darla	None	13.0	10.0	41429	8463	Sun Jul 30 15:58:51 +0000 2017	...	1	Paper Towel	0.170278	False	Labrador Retriever	0.168086	True	Spatula	0.040836	False
4	891327558926688256	2017-07-29	2019-01-04 16:00:00	Franklin	None	12.0	10.0	39599	9166	Sat Jul 29 16:00:24 +0000 2017	...	2	Basset	0.555712	True	English Springer	0.225770	True	German Short-Haired Pointer		

```
4  0.175219    True

[5 rows x 24 columns]
```

Problem There is time info in the Time column that is unessential when the data is being retrieved using Twitter API

Define Drop the time column in tweets_merge

Code

```
In [56]: # Drop the time column in tweets_merge
         tweets_merge.drop(columns = ['time'] , axis = 1 , inplace = True)
```

Test

```
In [57]: tweets_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1992 entries, 0 to 1991
Data columns (total 23 columns):
tweet_id              1992 non-null object
date                  1992 non-null datetime64[ns]
name                  1992 non-null object
stage                 1992 non-null object
rating_numerator      1992 non-null float64
rating_denominator    1992 non-null float64
favorite_count        1992 non-null int64
retweet_count         1992 non-null int64
created_at            1992 non-null object
source                1992 non-null object
retweeted_status      1992 non-null object
url                   1992 non-null object
jpg_url               1992 non-null object
img_num               1992 non-null int64
p1                    1992 non-null object
p1_conf               1992 non-null float64
p1_dog                1992 non-null bool
p2                    1992 non-null object
p2_conf               1992 non-null float64
p2_dog                1992 non-null bool
p3                    1992 non-null object
p3_conf               1992 non-null float64
p3_dog                1992 non-null bool
dtypes: bool(3), datetime64[ns](1), float64(5), int64(3), object(11)
memory usage: 332.6+ KB
```

Problem Duplicate date information in both date columns and created_at columns

Define Drop the date information in date columns and split the created_at columns information into both date and time and drop the created_at column

Code

```
In [58]: # Drop the date columns
         tweets_merge.drop('date' , axis = 1, inplace = True)

In [59]: # Split the created_at columns into both date and time
         # Convert the Date and time datatypes from String to Datetime objects

         # Ref : https://pandas.pydata.org/pandas-docs/version/0.20/generated/pandas.to_dateti

         tweets_merge['created_at'] = pd.to_datetime(tweets_merge['created_at'])

         # Create two more columns for Date and Time respectively for easy data exploration in

         # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.apply
         # Test this if free , Ref : https://stackoverflow.com/questions/35595710/splitting-ti
         # Ref for lambda and time is needed

         tweets_merge['date'] = tweets_merge['created_at'].apply(lambda time : time.strftime('
         tweets_merge['time'] = tweets_merge['created_at'].apply(lambda time : time.strftime('

In [60]: # Drop the created_at column in tweets_merge
         tweets_merge.drop('created_at' , axis = 1, inplace = True)
```

Test

```
In [61]: # Check the correction
         tweets_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1992 entries, 0 to 1991
Data columns (total 23 columns):
tweet_id              1992 non-null object
name                  1992 non-null object
stage                 1992 non-null object
rating_numerator      1992 non-null float64
rating_denominator    1992 non-null float64
favorite_count        1992 non-null int64
retweet_count         1992 non-null int64
source                1992 non-null object
retweeted_status      1992 non-null object
url                   1992 non-null object
jpg_url               1992 non-null object
```

```
img_num               1992 non-null int64
p1                    1992 non-null object
p1_conf               1992 non-null float64
p1_dog                1992 non-null bool
p2                    1992 non-null object
p2_conf               1992 non-null float64
p2_dog                1992 non-null bool
p3                    1992 non-null object
p3_conf               1992 non-null float64
p3_dog                1992 non-null bool
date                  1992 non-null object
time                  1992 non-null object
dtypes: bool(3), float64(5), int64(3), object(12)
memory usage: 332.6+ KB
```

```
In [62]: tweets_merge.head(2)
```

```
Out[62]:              tweet_id     name stage  rating_numerator  rating_denominator  \
         0  892420643555336193  Phineas  None              13.0                10.0
         1  892177421306343426    Tilly  None              13.0                10.0

            favorite_count  retweet_count              source retweeted_status  \
         0           38071           8327  Twitter for iPhone   Original tweet
         1           32681           6151  Twitter for iPhone   Original tweet

                                url  ...     p1_conf  p1_dog        p2   p2_conf  \
         0  https://t.co/MgUWQ76dJU  ...    0.097049   False     Bagel  0.085851
         1  https://t.co/aQFSeaCu9L  ...    0.323581    True  Pekinese  0.090647

            p2_dog        p3   p3_conf  p3_dog        date   time
         0   False    Banana  0.076110   False  08-01-2017  16:23
         1    True  Papillon  0.068957    True  08-01-2017  00:17

         [2 rows x 23 columns]
```

4 Data Storage

Store the clean DataFrame(s) in a CSV file with the main one named twitter_archive_master.csv. If additional files exist because multiple tables are required for tidiness, name these files appropriately. Additionally, you may store the cleaned data in a SQLite database (which is to be submitted as well if you

```
In [64]: # Save it in a file
```

```
         folder_name = 'Final_Documents'
         if not os.path.exists(folder_name) :
             os.makedirs(folder_name)
```

```
twitter_archive_clean.to_csv('Final_Documents/twitter_archive_master.csv')
tweet_json_clean.to_csv('Final_Documents/tweet_query_master.csv')
images_clean.to_csv('Final_Documents/image_prediction_master.csv')
tweets_merge.to_csv('Final_Documents/final_master.csv')
```

5 Data Analysis

```
In [65]: final_master= pd.read_csv("Final_Documents/final_master.csv")
         image_pred_master = pd.read_csv("Final_Documents/image_prediction_master.csv")
```

```
In [66]: final_master.head(5)
```

Out[66]:

	Unnamed: 0	tweet_id	name	stage	rating_numerator	rating_denominator	favorite_count	retweet_count	source	retweeted_status	...	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog	date	time
0	0	892420643555336193	Phineas	None	13.0	10.0	38071	8327	Twitter for iPhone	Original tweet	...	0.097049	False	Bagel	0.085851	False	Banana	0.076110	False	08-01-2017	16:23
1	1	892177421306343426	Tilly	None	13.0	10.0	32681	6151	Twitter for iPhone	Original tweet	...	0.323581	True	Pekinese	0.090647	True	Papillon	0.068957	True	08-01-2017	00:17
2	2	891815181378084864	Archie	None	12.0	10.0	24589	4069	Twitter for iPhone	Original tweet	...	0.716012	True	Malamute	0.078253	True	Kelpie	0.031379	True	07-31-2017	00:18
3	3	891689557279858688	Darla	None	13.0	10.0	41429	8463	Twitter for iPhone	Original tweet	...	0.170278	False	Labrador Retriever	0.168086	True	Spatula	0.040836	False	07-30-2017	15:58
4	4	891327558926688256	Franklin	None	12.0	10.0	39599	9166	Twitter for iPhone	Original tweet	...	0.555712	True	English Springer	0.225770	True	German Short-Haired Pointer	0.175219	True	07-29-2017	16:00

[5 rows x 24 columns]

```
In [67]: image_pred_master.head(5)
```

Out[67]:

	Unnamed: 0	tweet_id
0	0	666020888022790149
1	1	666029285002620928

```
2           2  666033412701032449
3           3  666044226329800704
4           4  666049248165822465

                                        jpg_url  img_num  \
0  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg        1
1  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg        1
2  https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg        1
3  https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg        1
4  https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg        1

                     p1   p1_conf  p1_dog                 p2   p2_conf  \
0  Welsh Springer Spaniel  0.465074    True             Collie  0.156665
1                Redbone  0.506826    True  Miniature Pinscher  0.074192
2        German Shepherd  0.596461    True            Malinois  0.138584
3    Rhodesian Ridgeback  0.408143    True             Redbone  0.360687
4     Miniature Pinscher  0.560311    True          Rottweiler  0.243682

   p2_dog                  p3   p3_conf  p3_dog
0    True   Shetland Sheepdog  0.061428    True
1    True  Rhodesian Ridgeback  0.072010    True
2    True           Bloodhound  0.116197    True
3    True   Miniature Pinscher  0.222752    True
4    True             Doberman  0.154629    True
```

Problem Inessential Unnamed: 0 columns in both image_pred_master and final_master dataset

Define Drop the Unnamed: 0 columns in both dataset

Code

```
In [68]: # Ref : https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.drop.

         image_pred_master.drop(columns = ['Unnamed: 0'] , axis = 1 , inplace = True)
         final_master.drop(columns = ['Unnamed: 0'] , axis = 1 , inplace = True)
```

Test

```
In [69]: image_pred_master.head(2)

Out[69]:             tweet_id                                        jpg_url  \
         0  666020888022790149  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
         1  666029285002620928  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg

            img_num                     p1   p1_conf  p1_dog                 p2  \
         0        1  Welsh Springer Spaniel  0.465074    True             Collie
         1        1                Redbone  0.506826    True  Miniature Pinscher
```

```
    p2_conf  p2_dog                    p3   p3_conf  p3_dog
0  0.156665    True     Shetland Sheepdog  0.061428    True
1  0.074192    True  Rhodesian Ridgeback  0.072010    True
```

```
In [70]: final_master.head(2)
```

```
Out[70]:             tweet_id     name stage  rating_numerator  rating_denominator  \
0  892420643555336193  Phineas  None              13.0                10.0
1  892177421306343426    Tilly  None              13.0                10.0

   favorite_count  retweet_count             source retweeted_status  \
0           38071           8327  Twitter for iPhone   Original tweet
1           32681           6151  Twitter for iPhone   Original tweet

                        url  ...     p1_conf  p1_dog        p2   p2_conf  \
0  https://t.co/MgUWQ76dJU  ...    0.097049   False     Bagel  0.085851
1  https://t.co/aQFSeaCu9L  ...    0.323581    True  Pekinese  0.090647

   p2_dog        p3   p3_conf  p3_dog        date   time
0   False    Banana  0.076110   False  08-01-2017  16:23
1    True  Papillon  0.068957    True  08-01-2017  00:17

[2 rows x 23 columns]
```

```
In [71]: # Check the descriptive statistics of final_master df
         final_master.describe()
```

```
Out[71]:           tweet_id  rating_numerator  rating_denominator  favorite_count  \
count  1.992000e+03       1992.000000         1992.000000     1992.000000
mean   7.358697e+17         12.282129           10.532631     8737.326305
std    6.749923e+16         41.518523            7.324367    12812.229908
min    6.660209e+17          0.000000            2.000000       79.000000
25%    6.758399e+17         10.000000           10.000000     1877.750000
50%    7.084748e+17         11.000000           10.000000     3951.000000
75%    7.878678e+17         12.000000           10.000000    10884.750000
max    8.924206e+17       1776.000000          170.000000   164423.000000

       retweet_count      img_num      p1_conf       p2_conf       p3_conf
count    1992.000000  1992.000000  1992.000000  1.992000e+03  1.992000e+03
mean     2654.109940     1.203313     0.593736  1.344164e-01  6.028013e-02
std      4731.362619     0.561022     0.271942  1.006394e-01  5.089864e-02
min        12.000000     1.000000     0.044333  1.011300e-08  1.740170e-10
25%       592.000000     1.000000     0.362775  5.401683e-02  1.619708e-02
50%      1274.000000     1.000000     0.587440  1.174550e-01  4.950530e-02
75%      3038.250000     1.000000     0.845599  1.949773e-01  9.162278e-02
max     83742.000000     4.000000     1.000000  4.880140e-01  2.734190e-01
```

The neural network performed the best on the 1st iteration with a mean prediciton of 0.593 Mean rating for a dog was 11/10 with an outlier of 1776/10 Mean retweet count for an original

tweet was 1274 and a maximum value of 83749. Mean favorite count for an original tweet was 3952 and a maximum value of 164443

5.1 1. What is the type of the dog that has the highest retweets count?

```
In [78]: final_master[final_master['retweet_count'] == 83742]
```

```
Out[78]:              tweet_id  name  stage  rating_numerator  rating_denominator  \
         774  744234799360020481  None  doggo              13.0                10.0

              favorite_count  retweet_count             source retweeted_status  \
         774          164423          83742  Twitter for iPhone   Original tweet

                                  url  ...     p1_conf  p1_dog        p2   p2_conf  \
         774  https://t.co/7wE9LTEXC4  ...    0.825333    True  Ice Bear  0.044681

              p2_dog       p3   p3_conf  p3_dog        date   time
         774   False  Whippet  0.018442    True  06-18-2016  18:26

         [1 rows x 23 columns]
```

```
In [79]: final_master[final_master['retweet_count'] == 83742].p1
```

```
Out[79]: 774    Labrador Retriever
         Name: p1, dtype: object
```

```
In [74]: # Check if his picture is correctly predicted by the neural network

         final_master[final_master['tweet_id']== 744234799360020481].url
```

```
Out[74]: 774    https://t.co/7wE9LTEXC4
         Name: url, dtype: object
```

The type of dog that has the highest retweet count is Labrador Retriever with a neural network prediction confidence level of 82.533%. By checking its picture, it is truly a Labrador Retriever

5.2 2. What are the most common dog names

```
In [75]: from collections import Counter

         name = final_master['name']

         count = Counter(name)
         count.most_common(5)
```

```
Out[75]: [('None', 639), ('Oliver', 10), ('Cooper', 10), ('Charlie', 10), ('Lucy', 10)]
```

4 dog names tied at the first place with Oliver, Cooper, Charlie and Lucy with 10 counts each

5.3 3. What are the most common dog ratings

```
In [76]: rating_numerator = final_master['rating_numerator']
         count = Counter(rating_numerator)
         count.most_common(1)

Out[76]: [(12.0, 449)]
```

12 is the most common Dog rating with 449 times rated.

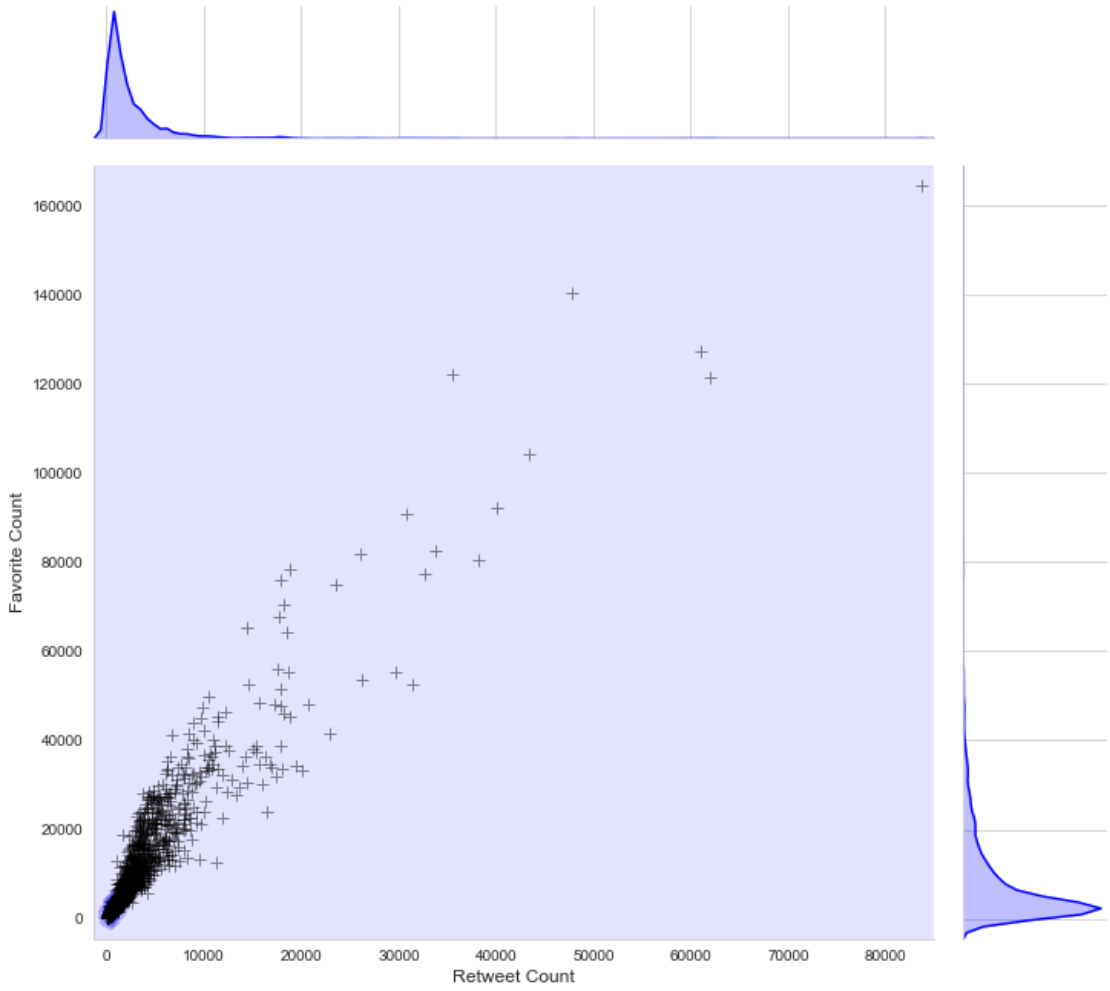
6 Data Visualization

6.1 1. Does the increase in retweet_count increases favorite_count

```
In [77]: import seaborn as sns
         sns.set_style('whitegrid')
         g = sns.jointplot(x="retweet_count",
                           y="favorite_count",
                           data=final_master,
                           color="blue",
                           kind="kde",
                           size=10)
         g.plot_joint(plt.scatter,
                      c="black",
                      s=80,
                      linewidth=1,
                      marker="+",
                      alpha=0.45)
         g.set_axis_labels("Retweet Count", "Favorite Count",fontsize=12)

         g.fig.subplots_adjust(top=0.9)
```

```
/anaconda3/lib/python3.7/site-packages/seaborn/axisgrid.py:2262: UserWarning: The `size` paramt
  warnings.warn(msg, UserWarning)
/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tu
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```



There is strong correlation between retweet_count and favourite count below 10, 000 retweeted_count. Then the correclation started weaken from 10, 000 onwards.

Hence the increase in retweet_count does increases favourite_count below 10 ,000 retweet count. Then, starting from 10, 000 retweet count there is marginal effect when there is increase in retweet count.