wrangle_act

September 17, 2020

1 Data Wrangling and Visualization - WeRateDogs Twitter account

1.1 Import the necessary libraries

```
In [5]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        import requests
        from bs4 import BeautifulSoup
        import os
        import tweepy
        import json
        %matplotlib inline
```

1.2 Gathering Data

Data is gathered from three different sources: Twitter archive (downloading the csv file manually), tweet image predictions (downloading the tsv file programmatically using Python's Requests library) and data scraped using the Twitter API using Python's Tweepy library . Each data is in a different file format. Each piece of data is imported into a separate pandas DataFrame at first.

1.2.1 i) Data from Twitter archive - manually downloading csv file

```
In [6]: df_archive = pd.read_csv('twitter-archive-enhanced.csv')

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

```
Out[7]:             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
```

1.2.2 ii) Data from twitt image predictions file - programmatically downloading tsv file

```
In [8]: # Make directory if it doesn't already exist
        folder_name = 'image_predictions'
        if not os.path.exists(folder_name):
            os.makedirs(folder_name)
```

```
In [9]: # Getting the data from the URL
        url = 'https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictio
```

```
        response = requests.get(url)
        response
```

```
Out[9]: <Response [200]>
```

```
In [10]: # Writing the data to a file
         with open (os.path.join(folder_name,url.split('/')[-1]), mode ='wb') as file:
             file.write(response.content)
```

```
In [11]: os.listdir('image_predictions')
```

```
Out[11]: ['image-predictions.tsv']
```

```
In [12]: df_predictions = pd.read_csv('image_predictions/image-predictions.tsv',sep='\t')
```

```
In [13]: df_predictions.head()
```

```
Out[13]:              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
```

1.2.3 iii) Data scraped using the Twitter API - using Python's Tweepy library

```
In [104]: consumer_key = 'XXXXXXXXXXX'
          consumer_secret = 'XXXXXXXXXXX'
          access_token = 'XXXXXXXXXXX'
          access_secret = 'XXXXXXXXXXX'

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

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

```
In [15]: tweet_id = df_archive.tweet_id
         len(tweet_id)

Out[15]: 2356

In [17]: working_ids=[]
         nonworking_ids=[]

In [18]: with open ('tweet_json.txt','w') as outfile:
             for tweet_id in tweet_id:
                 try:
                     tweet = api.get_status(tweet_id, tweet_mode='extended')
                     json.dump(tweet._json,outfile)
                     outfile.write('\n')
                     working_ids.append(tweet_id)
                 except:
                     print ("Error for: " + str(tweet_id))
                     nonworking_ids.append(tweet_id)

Error for: 888202515573088257
Error for: 873697596434513921
Error for: 872668790621863937
Error for: 869988702071779329
Error for: 866816280283807744
Error for: 861769973181624320
Error for: 845459076796616705
Error for: 842892208864923648
Error for: 837012587749474308
Error for: 827228250799742977
Error for: 812747805718642688
Error for: 802247111496568832
Error for: 775096608509886464
Error for: 770743923962707968


Rate limit reached. Sleeping for: 541


Error for: 758740312047005698
Error for: 754011816964026368
Error for: 680055455951884288


Rate limit reached. Sleeping for: 541


Error for: 676957860086095872


In [138]: working_ids
```

```
Out[138]: [8924206435553361 93,
           8921774213063434 26,
           8918151813780848 64,
           8916895572798586 88,
           8913275589266882 56,
           8910879508758978 56,
           8909719131739914 26,
           8907291814112378 88,
           8906091851503124 48,
           8902402553491988 49,
           8900066081131724 80,
           8898808964798668 81,
           8896653883336826 89,
           8896388375799070 72,
           8895311353442099 21,
           8892788419816857 60,
           8889172381238312 96,
           8888049891996712 97,
           8885549627242782 72,
           8880784344585871 36,
           8877052893818265 60,
           8875171391580938 24,
           8874739571039518 83,
           8873432170453688 32,
           8871013928040857 60,
           8869832335225446 40,
           8867368805193195 52,
           8866803364779335 68,
           8863661447344455 68,
           8862670092850176 00,
           8862583841518878 73,
           8860541600590725 13,
           8859848000199475 20,
           8855289432054702 08,
           8855189715287203 85,
           8853115929126092 80,
           8851676198836387 84,
           8849255217417093 13,
           8848767533904896 01,
           8845628921456885 76,
           8844418053827174 40,
           8842478788514938 88,
           8841626705843773 45,
           8838381229366312 99,
           8834828469330042 88,
           8833606908992184 34,
           8831178360460861 44,
           8829920803642204 16,
```

882762694511734784,
882627270321602560,
882268110199369728,
882045870035918850,
881906580714921986,
881666595344535552,
881633300179243008,
881536004380872706,
881268444196462592,
880935762899988482,
880872448815771648,
880465832366813184,
880221127280381952,
880095782870896641,
879862464715927552,
879674319642796034,
879492040517615616,
879415818425184262,
879376492567855104,
879130579576475649,
879050749262655488,
879008229531029506,
878776093423087618,
878604707211726852,
878404777348136964,
878316110768087041,
878281511006478336,
878057613040115712,
877736472329191424,
877611172832227328,
877556246731214848,
877316821321428993,
877201837425926144,
876838120628539392,
876537666061221889,
876484053909872640,
876120275196170240,
875747767867523072,
875144289856114688,
875097192612077568,
875021211251597312,
874680097055178752,
874434818259525634,
874296783580663808,
874057562936811520,
874012996292530176,
873580283840344065,
873337748698140672,

873213775632977920,
872967104147763200,
872820683541237760,
872620804844003328,
872486979161796608,
872261713294495745,
872122724285648897,
871879754684805121,
871762521631449091,
871515927908634625,
871166179821445120,
871102520638267392,
871032628920680449,
870804317367881728,
870726314365509632,
870656317836468226,
870374049280663552,
870308999962521604,
870063196459192321,
869772420881756160,
869702957897576449,
869596645499047938,
869227993411051520,
868880397819494401,
868639477480148993,
868622495443632128,
868552278524837888,
867900495410671616,
867774946302451713,
867421006826221569,
867072653475098625,
867051520902168576,
866720684873056260,
866686824827068416,
866450705531457537,
866334964761202691,
866094527597207552,
865718153858494464,
865359393868664832,
865006731092295680,
864873206498414592,
864279568663928832,
864197398364647424,
863907417377173506,
863553081350529029,
863471782782697472,
863432100342583297,
863427515083354112,

863079547188785154,
863062471531167744,
862831371563274240,
862722525377298433,
862457590147678208,
862096992088072192,
861383897657036800,
861288531465048066,
861005113778896900,
860981674716409858,
860924035999428608,
860563773140209665,
860524505164394496,
860276583193509888,
860184849394610176,
860177593139703809,
859924526012018688,
859851578198683649,
859607811541651456,
859196978902773760,
859074603037188101,
858860390427611136,
858843525470990336,
858471635011153920,
858107933456039936,
857989990357356544,
857746408056729600,
857393404942143489,
857263160327368704,
857214891891077121,
857062103051644929,
857029823797047296,
856602993587888130,
856543823941562368,
856526610513747968,
856330835276025856,
856288084350160898,
856282028240666624,
855862651834028034,
855860136149123072,
855857698524602368,
855851453814013952,
855818117272018944,
855459453768019968,
855245323840757760,
855138241867124737,
854732716440526848,
854482394044301312,

854365224396361728,
854120357044912130,
854010172552949760,
853760880890318849,
853639147608842240,
853299958564483072,
852936405516943360,
852912242202992640,
852672615818899456,
852553447878664193,
852311364735569921,
852226086759018497,
852189679701164033,
851953902622658560,
851861385021730816,
851591660324737024,
851464819735769094,
851224888060895234,
850753642995093505,
850380195714523136,
850333567704068097,
850145622816686080,
850019790995546112,
849776966551130114,
849668094696017920,
849412302885593088,
849336543269576704,
849051919805034497,
848690551926992896,
848324959059550208,
848213670039564288,
848212111729840128,
847978865427394560,
847971574464610304,
847962785489326080,
847842811428974592,
847617282490613760,
847606175596138505,
847251039262605312,
847157206088847362,
847116187444137987,
846874817362120707,
846514051647705089,
846505985330044928,
846153765933735936,
846139713627017216,
846042936437604353,
845812042753855489,

845677943972139009,
84539705715010764 8,
845306882940190720,
845098359547420673,
844979544864018432,
844973813909606400,
844704788403113984,
844580511645339650,
844223788422217728,
843981021012017153,
843856843873095681,
843604394117681152,
843235543001513987,
842846295480000512,
842765311967449089,
842535590457499648,
842163532590374912,
842115215311396866,
841833993020538882,
841680585030541313,
841439858740625411,
841320156043304961,
841314665196081154,
841077006473256960,
840761248237133825,
840728873075638272,
840698636975636481,
840696689258311684,
840632337062862849,
840370681858686976,
840268004936019968,
839990271299457024,
839549326359670784,
839290600511926273,
839239871831150596,
838952994649550848,
838921590096166913,
838916489579200512,
838831947270979586,
838561493054533637,
838476387338051585,
838201503651401729,
838150277551247360,
838085839343206401,
838083903487373313,
837820167694528512,
837482249356513284,
837471256429613056,

837366284874571778,
837110210464448512,
836989968035819520,
836753516572119041,
836677758902222849,
836648853927522308,
836397794269200385,
836380477523124226,
836260088725786625,
836001077879255040,
835685285446955009,
835574547218894849,
835536468978302976,
835309094223372289,
835297930240217089,
835264098648616962,
835246439529840640,
835172783151792128,
835152434251116546,
834931633769889797,
834786237630337024,
834574053763584002,
834477809192075265,
834458053273591808,
834209720923721728,
834167344700198914,
834089966724603904,
834086379323871233,
833863086058651648,
833826103416520705,
833732339549220864,
833722901757046785,
833479644947025920,
833124694597443584,
832998151111966721,
832769181346996225,
832757312314028032,
832682457690300417,
832645525019123713,
832636094638288896,
832397543355072512,
832369877331693569,
832273440279240704,
832215909146226688,
832215726631055365,
832088576586297345,
832040443403784192,
832032802820481025,

831939777352105988,
831926988323639298,
831911600680497154,
831670449226514432,
831650051525054464,
831552930092285952,
831322785565769729,
831315979191906304,
831309418084069378,
831262627380748289,
830956169170665475,
830583320585068544,
830173239259324417,
830097400375152640,
829878982036299777,
829861396166877184,
829501995190984704,
829449946868879360,
829374341691346946,
829141528400556032,
829011960981237760,
828801551087042563,
828770345708580865,
828708714936930305,
828650029636317184,
828409743546925057,
828408677031882754,
828381636999917570,
828376505180889089,
828372645993398273,
828361771580813312,
828046555563323392,
828011680017821696,
827933404142436356,
827653905312006145,
827600520311402496,
827324948884643840,
827199976799354881,
826958653328592898,
826848821049180160,
826615380357632002,
826598799820865537,
826598365270007810,
826476773533745153,
826240494070030336,
826204788643753985,
826115272272650244,
825876512159186944,

825829644528148480,
82553507688476262 4,
82514759169226342 4,
82512025641484697 6,
82502659071948390 4,
82479638019980902 4,
82477512667583692 8,
82466392634019430 5,
82432561328883302 4,
82429704827923661 1,
82402515877621350 4,
82393962851647488 0,
82371900293763072 0,
82369900299887001 6,
82358111563408588 8,
82333348951693721 6,
82332267812791911 0,
82326959422382489 7,
82297531540846182 4,
82287290174556979 3,
82285913416062156 9,
82264721290369024 1,
82261036194591129 6,
82248905708738970 0,
82246294436564582 5,
82224481652015513 6,
82216306474532864 0,
82188607640702976 0,
82181363921265049 6,
82176592326263193 6,
82152288970286285 2,
82142132020648345 7,
82140718235277721 8,
82115342186461593 6,
82114955467018240 0,
82110778581123482 0,
82104453188172185 6,
82083735790151270 4,
82074971684568678 6,
82069017664514048 1,
82049478856684748 9,
82044671915029299 3,
82031463377706188 8,
82007862539544985 7,
82001378160665804 9,
81995223645336371 2,
81992419535841689 6,
81971136213387264 3,

819588359383371776,
819347104292290561,
819238181065359361,
819227688460238848,
819015337530290176,
819015331746349057,
819006400881917954,
819004803107983360,
818646164899774465,
818627210458333184,
818614493328580609,
818588835076603904,
818536468981415936,
818307523543449600,
818259473185828864,
818145370475810820,
817908911860748288,
817827839487737858,
817777686764523521,
817536400337801217,
817502432452313088,
817423860136083457,
817415592588222464,
817181837579653120,
817171292965273600,
817120970343411712,
817056546584727552,
816829038950027264,
816816676327063552,
816697700272001025,
816450570814898180,
816336735214911488,
816091915477250048,
816062466425819140,
816014286006976512,
815990720817401858,
815966073409433600,
815745968457060357,
815736392542261248,
815639385530101762,
815390420867969024,
814986499976527872,
814638523311648768,
814578408554463233,
814530161257443328,
814153002265309185,
813944609378369540,
813910438903693312,

813812741911748608,
813800681631023104,
813217897535406080,
813202720496779264,
813187593374461952,
813172488309972993,
813157409116065792,
813142292504645637,
813130366689148928,
813127251579564032,
813112105746448384,
813096984823349248,
813081950185472002,
813066809284972545,
813051746834595840,
812781120811126785,
812709060537683968,
812503143955202048,
812466873996607488,
812372279581671427,
811985624773361665,
811744202451197953,
811647686436880384,
811627233043480576,
811386762094317568,
810984652412424192,
810896069567610880,
810657578271330305,
810284430598270976,
810254108431155201,
809920764300447744,
809808892968534016,
809448704142938112,
809220051211603969,
809084759137812480,
808838249661788160,
808733504066486276,
808501579447930884,
808344865868283904,
808134635716833280,
808106460588765185,
808001312164028416,
807621403335917568,
807106840509214720,
807059379405148160,
807010152071229440,
806629075125202948,
806620845233815552,

806576416489959424,
806542213899489280,
806242860592926720,
806219024703037440,
805958939288408065,
805932879469572096,
805826884734976000,
805823200554876929,
805520635690676224,
805487436403003392,
805207613751304193,
804738756058218496,
804475857670639616,
804413760345620481,
804026241225523202,
803773340896923648,
803692223237865472,
803638050916102144,
803380650405482500,
803321560782307329,
803276597545603072,
802952499103731712,
802624713319034886,
802600418706604034,
802572683846291456,
802323869084381190,
802265048156610565,
802239329049477120,
802185808107208704,
801958328846974976,
801854953262350336,
801538201127157760,
801285448605831168,
801167903437357056,
801127390143516673,
801115127852503040,
800859414831898624,
800855607700029440,
800751577355128832,
800513324630806528,
800459316964663297,
800443802682937345,
800388270626521089,
800188575492947969,
800141422401830912,
800018252395122689,
799774291445383169,
799757965289017345,

799422933579902976,
799308762079035393,
799297110730567681,
799063482566066176,
798933969379225600,
798925684722855936,
798705661114773508,
798701998996647937,
798697898615730177,
798694562394996736,
798686750113755136,
798682547630837760,
798673117451325440,
798665375516884993,
798644042770751489,
798628517273620480,
798585098161549313,
798576900688019456,
798340744599797760,
798209839306514432,
797971864723324932,
797545162159308800,
797236660651966464,
797165961484890113,
796904159865868288,
796865951799083009,
796759840936919040,
796563435802726400,
796484825502875648,
796387464403357696,
796177847564038144,
796149749086875649,
796125600683540480,
796116448414461957,
796080075804475393,
796031486298386433,
795464331001561088,
795400264262053889,
795076730285391872,
794983741416415232,
794926597468000259,
794355576146903043,
794332329137291264,
794205286408003585,
793962221541933056,
793845145112371200,
793614319594401792,
793601777308463104,

793500921481273345,
793286476301799424,
793271401113350145,
793256262322548741,
793241302385262592,
793226087023144960,
793210959003287553,
793195938047070209,
793180763617361921,
793165685325201412,
793150605191548928,
793135492858580992,
793120401413079041,
792913359805018113,
792883833364439040,
792773781206999040,
792394556390137856,
792050063153438720,
791821351946420224,
791784077045166082,
791780927877898241,
791774931465953280,
791672322847637504,
791406955684368384,
791312159183634433,
791026214425268224,
790987426131050500,
790946055508652032,
790723298204217344,
790698755171364864,
790581949425475584,
790337589677002753,
790277117346975746,
790227638568808452,
789986466051088384,
789960241177853952,
789903600034189313,
789628658055020548,
789599242079838210,
789530877013393408,
789314372632018944,
789280767834746880,
789268448748703744,
789137962068021249,
788908386943430656,
788765914992902144,
788552643979468800,
788412144018661376,

788178268662984705,
788150585577050112,
788070120937619456,
788039637453406209,
787810552592695296,
787717603741622272,
787397959788929025,
787322443945877504,
787111942498508800,
786963064373534720,
786729988674449408,
786709082849828864,
786664955043049472,
786595970293370880,
786363235746385920,
786286427768250368,
786233965241827333,
786051337297522688,
786036967502913536,
785927819176054784,
785872687017132033,
785639753186217984,
785533386513321988,
785515384317313025,
785264754247995392,
785170936622350336,
784826020293709826,
784517518371221505,
784431430411685888,
784183165795655680,
784057939640352768,
783839966405230592,
783821107061198850,
783695101801398276,
783466772167098368,
783391753726550016,
783347506784731136,
783334639985389568,
783085703974514689,
782969140009107456,
782747134529531904,
782722598790725632,
782598640137187329,
782305867769217024,
782021823840026624,
781955203444699136,
781661882474196992,
781655249211752448,

781524693396357120,
781308096455073793,
781251288990355457,
781163403222056960,
780931614150983680,
780858289093574656,
780800785462489090,
780601303617732608,
780543529827336192,
780496263422808064,
780476555013349377,
780459368902959104,
780192070812196864,
780092040432480260,
780074436359819264,
779834332596887552,
779377524342161408,
779124354206535695,
779123168116150273,
779056095788752897,
778990705243029504,
778774459159379968,
778764940568104960,
778748913645780993,
778650543019483137,
778624900596654080,
778408200802557953,
778396591732486144,
778383385161035776,
778286810187399168,
778039087836069888,
778027034220126208,
777953400541634568,
777885040357281792,
777684233540206592,
777641927919427584,
777621514455814149,
777189768882946048,
776819012571455488,
776813020089548800,
776477788987613185,
776249906839351296,
776218204058357768,
776201521193218049,
776113305656188928,
776088319444877312,
775898661951791106,
775842724423557120,

775733305207554048,
775729183532220416,
775364825476165632,
775350846108426240,
775085132600442880,
774757898236878852,
774639387460112384,
774314403806253056,
773985732834758656,
773922284943896577,
773704687002451968,
773670353721753600,
773547596996571136,
773336787167145985,
773308824254029826,
773247561583001600,
773191612633579521,
772877495989305348,
772826264096874500,
772615324260794368,
772581559778025472,
772193107915964416,
772152991789019136,
772117678702071809,
772114945936949249,
772102971039580160,
771908950375665664,
771770456517009408,
771500966810099713,
771380798096281600,
771171053431250945,
771136648247640064,
771102124360998913,
771014301343748096,
771004394259247104,
770787852854652928,
770772759874076672,
770655142660169732,
770414278348247044,
770293558247038976,
770093767776997377,
770069151037685760,
769940425801170949,
769695466921623552,
769335591808995329,
769212283578875904,
768970937022709760,
768909767477751808,

768855141948723200,
768609597686943744,
768596291618299904,
768554158521745409,
768473857036525572,
768193404517830656,
767884188863397888,
767754930266464257,
767500508068192258,
767191397493538821,
767122157629476866,
766864461642756096,
766793450729734144,
766714921925144576,
766693177336135680,
766423258543644672,
766313316352462849,
766078092750233600,
766069199026450432,
766008592277377025,
765719909049503744,
765669560888528897,
765395769549590528,
765371061932261376,
765222098633691136,
764857477905154048,
764259802650378240,
763956972077010945,
763837565564780549,
763183847194451968,
763167063695355904,
763103485927849985,
762699858130116608,
762471784394268675,
762464539388485633,
762316489655476224,
762035686371364864,
761976711479193600,
761750502866649088,
761745352076779520,
761672994376806400,
761599872357261312,
761371037149827077,
761334018830917632,
761292947749015552,
761227390836215808,
761004547850530816,
760893934457552897,

760656994973933572,
760641137271070720,
760539183865880579,
760521673607086080,
760290219849637889,
760252756032651264,
760190180481531904,
760153949710192640,
759943073749200896,
759923798737051648,
759846353224826880,
759793422261743616,
759566828574212096,
759557299618865152,
759447681597108224,
759446261539934208,
759197388317847553,
759159934323924993,
759099523532779520,
759047813560868866,
758854675097526272,
758828659922702336,
758474966123810816,
758467244762497024,
758405701903519748,
758355060040593408,
758099635764359168,
758041019896193024,
757741869644341248,
757729163776290825,
757725642876129280,
757611664640446465,
757597904299253760,
757596066325864448,
757400162377592832,
757393109802180609,
757354760399941633,
756998049151549440,
756939218950160384,
756651752796094464,
756526248105566208,
756303284449767430,
756288534030475264,
756275833623502848,
755955933503782912,
755206590534418437,
755110668769038337,
754874841593970688,

754856583969079297,
754747087846248448,
754482103782404096,
754449512966619136,
754120377874386944,
753655901052166144,
753420520834629632,
753398408988139520,
753375668877008896,
753298634498793472,
753294487569522689,
753039830821511168,
753026973505581056,
752932432744185856,
752917284578922496,
752701944171524096,
752682090207055872,
752660715232722944,
752568224206688256,
752519690950500352,
752334515931054080,
752309394570878976,
752173152931807232,
751950017322246144,
751937170840121344,
751830394383790080,
751793661361422336,
751598357617971201,
751583847268179968,
751538714308972544,
751456908746354688,
751251247299190784,
751205363882532864,
751132876104687617,
750868782890057730,
750719632563142656,
750506206503038976,
750429297815552001,
750383411068534784,
750381685133418496,
750147208377409536,
750132105863102464,
750117059602808832,
750101899009982464,
750086836815486976,
750071704093859840,
750056684286914561,
750041628174217216,

```
            750026558547456000,
            750011400160841729,
            749996283729883136,
            749981277374128128,
            749774190421639168,
            749417653287129088,
            749403093750648834,
            749395845976588288,
            749317047558017024,
            749075273010798592,
            749064354620928000,
            749036806121881602,
            748977405889503236,
            748932637671223296,
            748705597323898880,
            748699167502000129,
            748692773788876800,
            748575535303884801,
            748568946752774144,
            748346686624440324,
            748337862848962560,
            748324050481647620,
            748307329658011649,
            748220828303695873,
            747963614829678593,
            747933425676525569,
            747885874273214464,
            747844099428986880,
            747816857231626240,
            747651430853525504,
            747648653817413632,
            747600769478692864,
            747594051852075008,
            747512671126323200,
            747461612269887489,
            747439450712596480,
            747242308580548608,
            747219827526344708,
            747204161125646336,
            747103485104099331,
            ...]
```

In [139]:
```
len(working_ids)
```

Out[139]: 2338

In [90]:
```
nonworking_ids
```

Out[90]:
```
[888202515573088257,
 873697596434513921,
```

```
 872668790621863937,
 869988702071779329,
 866816280283807744,
 861769973181624320,
 845459076796616705,
 842892208864923648,
 837012587749474308,
 827228250799742977,
 812747805718642688,
 802247111496568832,
 775096608509886464,
 770743923962707968,
 758740312047005698,
 754011816964026368,
 680055455951884288,
 676957860086095872]
```

In [19]:
```
len(nonworking_ids)
```

Out[19]:
```
18
```

In [20]:
```
df_tweetdata = pd.read_json('tweet_json.txt',lines=True)
```

In [21]:
```
df_tweetdata.head()
```

Out[21]:

	contributors	coordinates	created_at	display_text_range	\
0	NaN	NaN	2017-08-01 16:23:56	[0, 85]	
1	NaN	NaN	2017-08-01 00:17:27	[0, 138]	
2	NaN	NaN	2017-07-31 00:18:03	[0, 121]	
3	NaN	NaN	2017-07-30 15:58:51	[0, 79]	
4	NaN	NaN	2017-07-29 16:00:24	[0, 138]	

	entities	\
0	{'hashtags': [], 'symbols': [], 'user_mentions...	
1	{'hashtags': [], 'symbols': [], 'user_mentions...	
2	{'hashtags': [], 'symbols': [], 'user_mentions...	
3	{'hashtags': [], 'symbols': [], 'user_mentions...	
4	{'hashtags': [{'text': 'BarkWeek', 'indices': ...	

	extended_entities	favorite_count	\
0	{'media': [{'id': 892420639486877696, 'id_str'...	37950	
1	{'media': [{'id': 892177413194625024, 'id_str'...	32588	
2	{'media': [{'id': 891815175371796480, 'id_str'...	24536	
3	{'media': [{'id': 891689552724799489, 'id_str'...	41291	
4	{'media': [{'id': 891327551943041024, 'id_str'...	39478	

	favorited	full_text	geo	\
0	False	This is Phineas. He's a mystical boy. Only eve...	NaN	
1	False	This is Tilly. She's just checking pup on you...	NaN	

```
2      False  This is Archie. He is a rare Norwegian Pouncin...  NaN
3      False  This is Darla. She commenced a snooze mid meal...  NaN
4      False  This is Franklin. He would like you to stop ca...  NaN

                         ...                          quoted_status  \
0                        ...                                    NaN
1                        ...                                    NaN
2                        ...                                    NaN
3                        ...                                    NaN
4                        ...                                    NaN

   quoted_status_id quoted_status_id_str  quoted_status_permalink  \
0               NaN                  NaN                      NaN
1               NaN                  NaN                      NaN
2               NaN                  NaN                      NaN
3               NaN                  NaN                      NaN
4               NaN                  NaN                      NaN

   retweet_count  retweeted  retweeted_status  \
0           8291      False               NaN
1           6123      False               NaN
2           4054      False               NaN
3           8428      False               NaN
4           9130      False               NaN

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

                                                user
0  {'id': 4196983835, 'id_str': '4196983835', 'na...
1  {'id': 4196983835, 'id_str': '4196983835', 'na...
2  {'id': 4196983835, 'id_str': '4196983835', 'na...
3  {'id': 4196983835, 'id_str': '4196983835', 'na...
4  {'id': 4196983835, 'id_str': '4196983835', 'na...

[5 rows x 32 columns]
```

1.3 Assessing Data

After gathering the data, the next step is to assess the data visually and programmatically for quality and tidiness issues. We assess each of the three dataframes: df_archive, df_predictions and df_tweetdata, detect the issues and document them.

In [22]:
```
# display the df_archive dataframe
df_archive
```

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	
5	891087950875897856	NaN	NaN	
6	890971913173991426	NaN	NaN	
7	890729181411237888	NaN	NaN	
8	890609185150312448	NaN	NaN	
9	890240255349198849	NaN	NaN	
10	890006608113172480	NaN	NaN	
11	889880896479866881	NaN	NaN	
12	889665388333682689	NaN	NaN	
13	889638837579907072	NaN	NaN	
14	889531135344209921	NaN	NaN	
15	889278841981685760	NaN	NaN	
16	888917238123831296	NaN	NaN	
17	888804989199671297	NaN	NaN	
18	888554962724278272	NaN	NaN	
19	888202515573088257	NaN	NaN	
20	887978434458587136	NaN	NaN	
21	887705289381826560	NaN	NaN	
22	887517139158093824	NaN	NaN	
23	887473957103951883	NaN	NaN	
24	887343217045368832	NaN	NaN	
25	887101392804085760	NaN	NaN	
26	886983233522544640	NaN	NaN	
27	886736880519319552	NaN	NaN	
28	886680336477933568	NaN	NaN	
29	886366144734445568	NaN	NaN	
...	...	...	...	
2326	666411507551481857	NaN	NaN	
2327	666407126856765440	NaN	NaN	
2328	666396247373291520	NaN	NaN	
2329	666373753744588802	NaN	NaN	
2330	666362758909284353	NaN	NaN	
2331	666353288456101888	NaN	NaN	
2332	666345417576210432	NaN	NaN	
2333	666337882303524864	NaN	NaN	
2334	666293911632134144	NaN	NaN	
2335	666287406224695296	NaN	NaN	
2336	666273097616637952	NaN	NaN	
2337	666268910803644416	NaN	NaN	
2338	666104133288665088	NaN	NaN	

```
2339  666102155909144576                  NaN                  NaN
2340  666099513787052032                  NaN                  NaN
2341  666094000022159362                  NaN                  NaN
2342  666082916733198337                  NaN                  NaN
2343  666073100786774016                  NaN                  NaN
2344  666071193221509120                  NaN                  NaN
2345  666063827256086533                  NaN                  NaN
2346  666058600524156928                  NaN                  NaN
2347  666057090499244032                  NaN                  NaN
2348  666055525042405380                  NaN                  NaN
2349  666051853826850816                  NaN                  NaN
2350  666050758794694657                  NaN                  NaN
2351  666049248165822465                  NaN                  NaN
2352  666044226329800704                  NaN                  NaN
2353  666033412701032449                  NaN                  NaN
2354  666029285002620928                  NaN                  NaN
2355  666020888022790149                  NaN                  NaN

                      timestamp  \
0     2017-08-01 16:23:56 +0000
1     2017-08-01 00:17:27 +0000
2     2017-07-31 00:18:03 +0000
3     2017-07-30 15:58:51 +0000
4     2017-07-29 16:00:24 +0000
5     2017-07-29 00:08:17 +0000
6     2017-07-28 16:27:12 +0000
7     2017-07-28 00:22:40 +0000
8     2017-07-27 16:25:51 +0000
9     2017-07-26 15:59:51 +0000
10    2017-07-26 00:31:25 +0000
11    2017-07-25 16:11:53 +0000
12    2017-07-25 01:55:32 +0000
13    2017-07-25 00:10:02 +0000
14    2017-07-24 17:02:04 +0000
15    2017-07-24 00:19:32 +0000
16    2017-07-23 00:22:39 +0000
17    2017-07-22 16:56:37 +0000
18    2017-07-22 00:23:06 +0000
19    2017-07-21 01:02:36 +0000
20    2017-07-20 16:49:33 +0000
21    2017-07-19 16:06:48 +0000
22    2017-07-19 03:39:09 +0000
23    2017-07-19 00:47:34 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
26    2017-07-17 16:17:36 +0000
27    2017-07-16 23:58:41 +0000
28    2017-07-16 20:14:00 +0000
```

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

                                                  source  \
0     <a href="http://twitter.com/download/iphone" r...
1     <a href="http://twitter.com/download/iphone" r...
2     <a href="http://twitter.com/download/iphone" r...
3     <a href="http://twitter.com/download/iphone" r...
4     <a href="http://twitter.com/download/iphone" r...
5     <a href="http://twitter.com/download/iphone" r...
6     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
8     <a href="http://twitter.com/download/iphone" r...
9     <a href="http://twitter.com/download/iphone" r...
10    <a href="http://twitter.com/download/iphone" r...
11    <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
13    <a href="http://twitter.com/download/iphone" r...
```

```
14    <a href="http://twitter.com/download/iphone" r...
15    <a href="http://twitter.com/download/iphone" r...
16    <a href="http://twitter.com/download/iphone" r...
17    <a href="http://twitter.com/download/iphone" r...
18    <a href="http://twitter.com/download/iphone" r...
19    <a href="http://twitter.com/download/iphone" r...
20    <a href="http://twitter.com/download/iphone" r...
21    <a href="http://twitter.com/download/iphone" r...
22    <a href="http://twitter.com/download/iphone" r...
23    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
26    <a href="http://twitter.com/download/iphone" r...
27    <a href="http://twitter.com/download/iphone" r...
28    <a href="http://twitter.com/download/iphone" r...
29    <a href="http://twitter.com/download/iphone" r...
...                                                  ...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2345  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
2348  <a href="http://twitter.com/download/iphone" r...
2349  <a href="http://twitter.com/download/iphone" r...
2350  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...
```

	text	retweeted_status_id	\
0	This is Phineas. He's a mystical boy. Only eve...	NaN	
1	This is Tilly. She's just checking pup on you...	NaN	
2	This is Archie. He is a rare Norwegian Pouncin...	NaN	
3	This is Darla. She commenced a snooze mid meal...	NaN	
4	This is Franklin. He would like you to stop ca...	NaN	
5	Here we have a majestic great white breaching ...	NaN	
6	Meet Jax. He enjoys ice cream so much he gets ...	NaN	
7	When you watch your owner call another dog a g...	NaN	
8	This is Zoey. She doesn't want to be one of th...	NaN	
9	This is Cassie. She is a college pup. Studying...	NaN	
10	This is Koda. He is a South Australian decksha...	NaN	
11	This is Bruno. He is a service shark. Only get...	NaN	
12	Here's a puppo that seems to be on the fence a...	NaN	
13	This is Ted. He does his best. Sometimes that'...	NaN	
14	This is Stuart. He's sporting his favorite fan...	NaN	
15	This is Oliver. You're witnessing one of his m...	NaN	
16	This is Jim. He found a fren. Taught him how t...	NaN	
17	This is Zeke. He has a new stick. Very proud o...	NaN	
18	This is Ralphus. He's powering up. Attempting ...	NaN	
19	RT @dog_rates: This is Canela. She attempted s...	8.874740e+17	
20	This is Gerald. He was just told he didn't get...	NaN	
21	This is Jeffrey. He has a monopoly on the pool...	NaN	
22	I've yet to rate a Venezuelan Hover Wiener. Th...	NaN	
23	This is Canela. She attempted some fancy porch...	NaN	
24	You may not have known you needed to see this ...	NaN	
25	This... is a Jubilant Antarctic House Bear. We...	NaN	
26	This is Maya. She's very shy. Rarely leaves he...	NaN	
27	This is Mingus. He's a wonderful father to his...	NaN	
28	This is Derek. He's late for a dog meeting. 13...	NaN	
29	This is Roscoe. Another pupper fallen victim t...	NaN	
...	...	...	
2326	This is quite the dog. Gets really excited whe...	NaN	
2327	This is a southern Vesuvius bumblegruff. Can d...	NaN	
2328	Oh goodness. A super rare northeast Qdoba kang...	NaN	
2329	Those are sunglasses and a jean jacket. 11/10 ...	NaN	
2330	Unique dog here. Very small. Lives in containe...	NaN	
2331	Here we have a mixed Asiago from the Galápagos...	NaN	
2332	Look at this jokester thinking seat belt laws ...	NaN	
2333	This is an extremely rare horned Parthenon. No...	NaN	
2334	This is a funny dog. Weird toes. Won't come do...	NaN	
2335	This is an Albanian 3 1/2 legged Episcopalian...	NaN	
2336	Can take selfies 11/10 https://t.co/ws2AMaNwPW	NaN	
2337	Very concerned about fellow dog trapped in com...	NaN	
2338	Not familiar with this breed. No tail (weird)...	NaN	
2339	Oh my. Here you are seeing an Adobe Setter giv...	NaN	
2340	Can stand on stump for what seems like a while...	NaN	
2341	This appears to be a Mongolian Presbyterian mi...	NaN	

```
2342  Here we have a well-established sunblockerspan...                NaN
2343  Let's hope this flight isn't Malaysian (lol). ...                 NaN
2344  Here we have a northern speckled Rhododendron...                 NaN
2345  This is the happiest dog you will ever see. Ve...                 NaN
2346  Here is the Rand Paul of retrievers folks! He'...                 NaN
2347  My oh my. This is a rare blond Canadian terrie...                 NaN
2348  Here is a Siberian heavily armored polar bear ...                 NaN
2349  This is an odd dog. Hard on the outside but lo...                 NaN
2350  This is a truly beautiful English Wilson Staff...                 NaN
2351  Here we have a 1949 1st generation vulpix. Enj...                 NaN
2352  This is a purebred Piers Morgan. Loves to Netf...                 NaN
2353  Here is a very happy pup. Big fan of well-main...                 NaN
2354  This is a western brown Mitsubishi terrier. Up...                 NaN
2355  Here we have a Japanese Irish Setter. Lost eye...                 NaN

      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                         NaN
1                          NaN                         NaN
2                          NaN                         NaN
3                          NaN                         NaN
4                          NaN                         NaN
5                          NaN                         NaN
6                          NaN                         NaN
7                          NaN                         NaN
8                          NaN                         NaN
9                          NaN                         NaN
10                         NaN                         NaN
11                         NaN                         NaN
12                         NaN                         NaN
13                         NaN                         NaN
14                         NaN                         NaN
15                         NaN                         NaN
16                         NaN                         NaN
17                         NaN                         NaN
18                         NaN                         NaN
19                4.196984e+09  2017-07-19 00:47:34 +0000
20                         NaN                         NaN
21                         NaN                         NaN
22                         NaN                         NaN
23                         NaN                         NaN
24                         NaN                         NaN
25                         NaN                         NaN
26                         NaN                         NaN
27                         NaN                         NaN
28                         NaN                         NaN
29                         NaN                         NaN
...                        ...                         ...
2326                       NaN                         NaN
```

```
2327                        NaN                         NaN
2328                        NaN                         NaN
2329                        NaN                         NaN
2330                        NaN                         NaN
2331                        NaN                         NaN
2332                        NaN                         NaN
2333                        NaN                         NaN
2334                        NaN                         NaN
2335                        NaN                         NaN
2336                        NaN                         NaN
2337                        NaN                         NaN
2338                        NaN                         NaN
2339                        NaN                         NaN
2340                        NaN                         NaN
2341                        NaN                         NaN
2342                        NaN                         NaN
2343                        NaN                         NaN
2344                        NaN                         NaN
2345                        NaN                         NaN
2346                        NaN                         NaN
2347                        NaN                         NaN
2348                        NaN                         NaN
2349                        NaN                         NaN
2350                        NaN                         NaN
2351                        NaN                         NaN
2352                        NaN                         NaN
2353                        NaN                         NaN
2354                        NaN                         NaN
2355                        NaN                         NaN

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

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

      rating_denominator      name  doggo floofer  pupper  puppo
0                     10   Phineas   None    None    None   None
1                     10     Tilly   None    None    None   None
```

2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None

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

[2356 rows x 17 columns]
```

```
In [23]: df_archive.info()
```

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

```
In [24]: df_archive.describe()
```

```
Out[24]:            tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         count  2.356000e+03           7.800000e+01         7.800000e+01
         mean   7.427716e+17           7.455079e+17         2.014171e+16
         std    6.856705e+16           7.582492e+16         1.252797e+17
         min    6.660209e+17           6.658147e+17         1.185634e+07
         25%    6.783989e+17           6.757419e+17         3.086374e+08
         50%    7.196279e+17           7.038708e+17         4.196984e+09
```

```
75%    7.993373e+17           8.257804e+17         4.196984e+09
max    8.924206e+17           8.862664e+17         8.405479e+17

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

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

In [97]:
```
# display the df_predictions dataframe
df_predictions
```

Out[97]:

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

```
21    666293911632134144    https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22    666337882303524864    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
23    666345417576210432    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
24    666353288456101888    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg
25    666362758909284353    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26    666373753744588802    https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.jpg
27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    666411507551481857    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...                  ...                                               ...
2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046  886680336477933568    https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJW0AAAvBJ.jpg
2049  887101392804085760    https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050  887343217045368832  https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051  887473957103951883    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052  887517139158093824  https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053  887705289381826560    https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054  888078434458587136    https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  888202515573088257    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  888554962724278272    https://pbs.twimg.com/media/DFTH_0-UQAACu2O.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  889278841981685760  https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921    https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg
2061  889638837579907072    https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881    https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
2065  890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.jpg
2066  890609185150312448    https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067  890729181411237888    https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068  890971913173991426    https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069  891087950875897856    https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070  891327558926688256    https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071  891689557279858688    https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072  891815181378084864    https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073  892177421306343426    https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074  892420643555336193    https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

      img_num                          p1   p1_conf  p1_dog  \
0           1       Welsh_springer_spaniel  0.465074    True
1           1                     redbone  0.506826    True
2           1              German_shepherd  0.596461    True
3           1          Rhodesian_ridgeback  0.408143    True
4           1           miniature_pinscher  0.560311    True
5           1         Bernese_mountain_dog  0.651137    True
```

6	1	box_turtle	0.933012	False
7	1	chow	0.692517	True
8	1	shopping_cart	0.962465	False
9	1	miniature_poodle	0.201493	True
10	1	golden_retriever	0.775930	True
11	1	Gordon_setter	0.503672	True
12	1	Walker_hound	0.260857	True
13	1	pug	0.489814	True
14	1	bloodhound	0.195217	True
15	1	Lhasa	0.582330	True
16	1	English_setter	0.298617	True
17	1	hen	0.965932	False
18	1	desktop_computer	0.086502	False
19	1	Italian_greyhound	0.176053	True
20	1	Maltese_dog	0.857531	True
21	1	three-toed_sloth	0.914671	False
22	1	ox	0.416669	False
23	1	golden_retriever	0.858744	True
24	1	malamute	0.336874	True
25	1	guinea_pig	0.996496	False
26	1	soft-coated_wheaten_terrier	0.326467	True
27	1	Chihuahua	0.978108	True
28	1	black-and-tan_coonhound	0.529139	True
29	1	coho	0.404640	False
...	...	...	...	...
2045	1	French_bulldog	0.999201	True
2046	1	convertible	0.738995	False
2047	1	kuvasz	0.309706	True
2048	2	Chihuahua	0.793469	True
2049	1	Samoyed	0.733942	True
2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True

```
2068        1                 Appenzeller  0.341703    True
2069        1     Chesapeake_Bay_retriever  0.425595    True
2070        2                       basset  0.555712    True
2071        1                  paper_towel  0.170278   False
2072        1                    Chihuahua  0.716012    True
2073        1                    Chihuahua  0.323581    True
2074        1                       orange  0.097049   False

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

2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

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

```
24    0.093412    True
25    0.000461   False
26    0.206803    True
27    0.004577    True
28    0.173810    True
29    0.189945   False
...        ...     ...
2045  0.000076    True
2046  0.044173   False
2047  0.086346    True
2048  0.032253   False
2049  0.029705    True
2050  0.134203    True
2051  0.038915    True
2052  0.026321   False
2053  0.026236    True
2054  0.000903    True
2055  0.038915    True
2056  0.111411    True
2057  0.073482    True
2058  0.105506    True
2059  0.027351    True
2060  0.007958    True
2061  0.001498    True
2062  0.004633    True
2063  0.082981   False
2064  0.008167    True
2065  0.029248    True
2066  0.118184    True
2067  0.076507    True
2068  0.193548   False
2069  0.076902   False
2070  0.175219    True
2071  0.040836   False
2072  0.031379    True
2073  0.068957    True
2074  0.076110   False

[2075 rows x 12 columns]
```

```
In [25]: df_predictions.info()
```

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

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

In [26]:
```
df_predictions.describe()
```

Out[26]:

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

In [27]:
```
# display the df_tweetdata dataframe
df_tweetdata
```

Out[27]:

	contributors	coordinates	created_at	display_text_range	\
0	NaN	NaN	2017-08-01 16:23:56	[0, 85]	
1	NaN	NaN	2017-08-01 00:17:27	[0, 138]	
2	NaN	NaN	2017-07-31 00:18:03	[0, 121]	
3	NaN	NaN	2017-07-30 15:58:51	[0, 79]	
4	NaN	NaN	2017-07-29 16:00:24	[0, 138]	
5	NaN	NaN	2017-07-29 00:08:17	[0, 138]	
6	NaN	NaN	2017-07-28 16:27:12	[0, 140]	
7	NaN	NaN	2017-07-28 00:22:40	[0, 118]	
8	NaN	NaN	2017-07-27 16:25:51	[0, 122]	
9	NaN	NaN	2017-07-26 15:59:51	[0, 133]	
10	NaN	NaN	2017-07-26 00:31:25	[0, 130]	
11	NaN	NaN	2017-07-25 16:11:53	[0, 107]	
12	NaN	NaN	2017-07-25 01:55:32	[0, 106]	
13	NaN	NaN	2017-07-25 00:10:02	[0, 91]	
14	NaN	NaN	2017-07-24 17:02:04	[0, 118]	
15	NaN	NaN	2017-07-24 00:19:32	[0, 138]	
16	NaN	NaN	2017-07-23 00:22:39	[0, 86]	
17	NaN	NaN	2017-07-22 16:56:37	[0, 128]	
18	NaN	NaN	2017-07-22 00:23:06	[0, 87]	
19	NaN	NaN	2017-07-20 16:49:33	[0, 127]	

```
20             NaN          NaN 2017-07-19 16:06:48           [0, 127]
21             NaN          NaN 2017-07-19 03:39:09           [0, 108]
22             NaN          NaN 2017-07-19 00:47:34            [0, 99]
23             NaN          NaN 2017-07-18 16:08:03            [0, 88]
24             NaN          NaN 2017-07-18 00:07:08           [0, 129]
25             NaN          NaN 2017-07-17 16:17:36           [0, 101]
26             NaN          NaN 2017-07-16 23:58:41           [0, 121]
27             NaN          NaN 2017-07-16 20:14:00            [0, 71]
28             NaN          NaN 2017-07-15 23:25:31           [0, 131]
29             NaN          NaN 2017-07-15 16:51:35          [27, 105]
...            ...          ...                 ...               ...
2308           NaN          NaN 2015-11-17 00:24:19           [0, 140]
2309           NaN          NaN 2015-11-17 00:06:54           [0, 139]
2310           NaN          NaN 2015-11-16 23:23:41           [0, 137]
2311           NaN          NaN 2015-11-16 21:54:18            [0, 81]
2312           NaN          NaN 2015-11-16 21:10:36           [0, 140]
2313           NaN          NaN 2015-11-16 20:32:58           [0, 135]
2314           NaN          NaN 2015-11-16 20:01:42           [0, 112]
2315           NaN          NaN 2015-11-16 19:31:45           [0, 139]
2316           NaN          NaN 2015-11-16 16:37:02           [0, 138]
2317           NaN          NaN 2015-11-16 16:11:11           [0, 136]
2318           NaN          NaN 2015-11-16 15:14:19            [0, 46]
2319           NaN          NaN 2015-11-16 14:57:41            [0, 82]
2320           NaN          NaN 2015-11-16 04:02:55           [0, 134]
2321           NaN          NaN 2015-11-16 03:55:04           [0, 128]
2322           NaN          NaN 2015-11-16 03:44:34           [0, 140]
2323           NaN          NaN 2015-11-16 03:22:39           [0, 132]
2324           NaN          NaN 2015-11-16 02:38:37           [0, 125]
2325           NaN          NaN 2015-11-16 01:59:36           [0, 137]
2326           NaN          NaN 2015-11-16 01:52:02           [0, 137]
2327           NaN          NaN 2015-11-16 01:22:45           [0, 107]
2328           NaN          NaN 2015-11-16 01:01:59           [0, 135]
2329           NaN          NaN 2015-11-16 00:55:59           [0, 124]
2330           NaN          NaN 2015-11-16 00:49:46           [0, 140]
2331           NaN          NaN 2015-11-16 00:35:11           [0, 138]
2332           NaN          NaN 2015-11-16 00:30:50           [0, 140]
2333           NaN          NaN 2015-11-16 00:24:50           [0, 120]
2334           NaN          NaN 2015-11-16 00:04:52           [0, 137]
2335           NaN          NaN 2015-11-15 23:21:54           [0, 130]
2336           NaN          NaN 2015-11-15 23:05:30           [0, 139]
2337           NaN          NaN 2015-11-15 22:32:08           [0, 131]

                                                 entities  \
0     {'hashtags': [], 'symbols': [], 'user_mentions...
1     {'hashtags': [], 'symbols': [], 'user_mentions...
2     {'hashtags': [], 'symbols': [], 'user_mentions...
3     {'hashtags': [], 'symbols': [], 'user_mentions...
4     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
```

```
5     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
6     {'hashtags': [], 'symbols': [], 'user_mentions...
7     {'hashtags': [], 'symbols': [], 'user_mentions...
8     {'hashtags': [{'text': 'BarkWeek', 'indices': ...
9     {'hashtags': [], 'symbols': [], 'user_mentions...
10    {'hashtags': [{'text': 'BarkWeek', 'indices': ...
11    {'hashtags': [], 'symbols': [], 'user_mentions...
12    {'hashtags': [], 'symbols': [], 'user_mentions...
13    {'hashtags': [], 'symbols': [], 'user_mentions...
14    {'hashtags': [{'text': 'BarkWeek', 'indices': ...
15    {'hashtags': [{'text': 'BarkWeek', 'indices': ...
16    {'hashtags': [], 'symbols': [], 'user_mentions...
17    {'hashtags': [], 'symbols': [], 'user_mentions...
18    {'hashtags': [], 'symbols': [], 'user_mentions...
19    {'hashtags': [], 'symbols': [], 'user_mentions...
20    {'hashtags': [], 'symbols': [], 'user_mentions...
21    {'hashtags': [], 'symbols': [], 'user_mentions...
22    {'hashtags': [], 'symbols': [], 'user_mentions...
23    {'hashtags': [], 'symbols': [], 'user_mentions...
24    {'hashtags': [], 'symbols': [], 'user_mentions...
25    {'hashtags': [], 'symbols': [], 'user_mentions...
26    {'hashtags': [], 'symbols': [], 'user_mentions...
27    {'hashtags': [], 'symbols': [], 'user_mentions...
28    {'hashtags': [], 'symbols': [], 'user_mentions...
29    {'hashtags': [], 'symbols': [], 'user_mentions...
...                                                 ...
2308  {'hashtags': [], 'symbols': [], 'user_mentions...
2309  {'hashtags': [], 'symbols': [], 'user_mentions...
2310  {'hashtags': [], 'symbols': [], 'user_mentions...
2311  {'hashtags': [], 'symbols': [], 'user_mentions...
2312  {'hashtags': [], 'symbols': [], 'user_mentions...
2313  {'hashtags': [], 'symbols': [], 'user_mentions...
2314  {'hashtags': [], 'symbols': [], 'user_mentions...
2315  {'hashtags': [], 'symbols': [], 'user_mentions...
2316  {'hashtags': [], 'symbols': [], 'user_mentions...
2317  {'hashtags': [], 'symbols': [], 'user_mentions...
2318  {'hashtags': [], 'symbols': [], 'user_mentions...
2319  {'hashtags': [], 'symbols': [], 'user_mentions...
2320  {'hashtags': [], 'symbols': [], 'user_mentions...
2321  {'hashtags': [], 'symbols': [], 'user_mentions...
2322  {'hashtags': [], 'symbols': [], 'user_mentions...
2323  {'hashtags': [], 'symbols': [], 'user_mentions...
2324  {'hashtags': [], 'symbols': [], 'user_mentions...
2325  {'hashtags': [], 'symbols': [], 'user_mentions...
2326  {'hashtags': [], 'symbols': [], 'user_mentions...
2327  {'hashtags': [], 'symbols': [], 'user_mentions...
2328  {'hashtags': [], 'symbols': [], 'user_mentions...
2329  {'hashtags': [], 'symbols': [], 'user_mentions...
```

```
2330  {'hashtags': [], 'symbols': [], 'user_mentions...
2331  {'hashtags': [], 'symbols': [], 'user_mentions...
2332  {'hashtags': [], 'symbols': [], 'user_mentions...
2333  {'hashtags': [], 'symbols': [], 'user_mentions...
2334  {'hashtags': [], 'symbols': [], 'user_mentions...
2335  {'hashtags': [], 'symbols': [], 'user_mentions...
2336  {'hashtags': [], 'symbols': [], 'user_mentions...
2337  {'hashtags': [], 'symbols': [], 'user_mentions...

                                        extended_entities  favorite_count  \
0     {'media': [{'id': 892420639486877696, 'id_str'...            37950
1     {'media': [{'id': 892177413194625024, 'id_str'...            32588
2     {'media': [{'id': 891815175371796480, 'id_str'...            24536
3     {'media': [{'id': 891689552724799489, 'id_str'...            41291
4     {'media': [{'id': 891327551943041024, 'id_str'...            39478
5     {'media': [{'id': 891087942176911360, 'id_str'...            19839
6     {'media': [{'id': 890971906207338496, 'id_str'...            11599
7     {'media': [{'id': 890729118844600320, 'id_str'...            64022
8     {'media': [{'id': 890609177319665665, 'id_str'...            27250
9     {'media': [{'id': 890240245463175168, 'id_str'...            31269
10    {'media': [{'id': 890006600089468928, 'id_str'...            30042
11    {'media': [{'id': 889880888800096258, 'id_str'...            27249
12    {'media': [{'id': 889665366129029120, 'id_str'...            47146
13    {'media': [{'id': 889638825424826374, 'id_str'...            26608
14    {'media': [{'id': 889531127467266049, 'id_str'...            14810
15    {'media': [{'id': 889278779352338437, 'id_str'...            24738
16    {'media': [{'id': 888917229776945152, 'id_str'...            28532
17    {'media': [{'id': 888804981515575296, 'id_str'...            25053
18    {'media': [{'id': 888554915546542081, 'id_str'...            19420
19    {'media': [{'id': 888078426338406400, 'id_str'...            21313
20    {'media': [{'id': 887705281597243393, 'id_str'...            29565
21    {'media': [{'id': 887517108413886465, 'id_str'...            45347
22    {'media': [{'id': 887473949361045505, 'id_str'...            67599
23    {'media': [{'id': 887343120832229379, 'id_str'...            33001
24    {'media': [{'id': 887101385971384320, 'id_str'...            29961
25    {'media': [{'id': 886983218871902208, 'id_str'...            34432
26    {'media': [{'id': 886736868116754432, 'id_str'...            11817
27    {'media': [{'id': 886680331239161856, 'id_str'...            21978
28    {'media': [{'id': 886366138128449536, 'id_str'...            20766
29                                                   NaN              116
...                                                  ...              ...
2308  {'media': [{'id': 666411498068123649, 'id_str'...              435
2309  {'media': [{'id': 666407121513275392, 'id_str'...              105
2310  {'media': [{'id': 666396240351993856, 'id_str'...              162
2311  {'media': [{'id': 666373746337402880, 'id_str'...              182
2312  {'media': [{'id': 666362717482020864, 'id_str'...              760
2313  {'media': [{'id': 666353280906170368, 'id_str'...              215
2314  {'media': [{'id': 666345414279471104, 'id_str'...              286
```

```
2315  {'media': [{'id': 666337857791987715, 'id_str'...             191
2316  {'media': [{'id': 666293909010702337, 'id_str'...             489
2317  {'media': [{'id': 666287399580733440, 'id_str'...             144
2318  {'media': [{'id': 666273081518768128, 'id_str'...             171
2319  {'media': [{'id': 666268904428277760, 'id_str'...             100
2320  {'media': [{'id': 666104129232740352, 'id_str'...           14052
2321  {'media': [{'id': 666102150364286977, 'id_str'...              78
2322  {'media': [{'id': 666099505364733952, 'id_str'...             151
2323  {'media': [{'id': 666093996847063040, 'id_str'...             162
2324  {'media': [{'id': 666082912819875840, 'id_str'...             114
2325  {'media': [{'id': 666073098362486784, 'id_str'...             315
2326  {'media': [{'id': 666071190449033216, 'id_str'...             143
2327  {'media': [{'id': 666063820255862784, 'id_str'...             466
2328  {'media': [{'id': 666058597072306176, 'id_str'...             109
2329  {'media': [{'id': 666057085227016192, 'id_str'...             290
2330  {'media': [{'id': 666055517517848576, 'id_str'...             429
2331  {'media': [{'id': 666051848592334848, 'id_str'...            1200
2332  {'media': [{'id': 666050754986266625, 'id_str'...             130
2333  {'media': [{'id': 666049244999131136, 'id_str'...             106
2334  {'media': [{'id': 666044217047650304, 'id_str'...             293
2335  {'media': [{'id': 666033409081393153, 'id_str'...             123
2336  {'media': [{'id': 666029276303482880, 'id_str'...             125
2337  {'media': [{'id': 666020881337073664, 'id_str'...            2532

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

```
23        False  You may not have known you needed to see this ...  NaN
24        False  This... is a Jubilant Antarctic House Bear. We...  NaN
25        False  This is Maya. She's very shy. Rarely leaves he...  NaN
26        False  This is Mingus. He's a wonderful father to his...  NaN
27        False  This is Derek. He's late for a dog meeting. 13...  NaN
28        False  This is Roscoe. Another pupper fallen victim t...  NaN
29        False  @NonWhiteHat @MayhewMayhem omg hello tanner yo...  NaN
...         ...                                                ...  ...
2308      False  This is quite the dog. Gets really excited whe...  NaN
2309      False  This is a southern Vesuvius bumblegruff. Can d...  NaN
2310      False  Oh goodness. A super rare northeast Qdoba kang...  NaN
2311      False  Those are sunglasses and a jean jacket. 11/10 ...  NaN
2312      False  Unique dog here. Very small. Lives in containe...  NaN
2313      False  Here we have a mixed Asiago from the Galápagos...  NaN
2314      False  Look at this jokester thinking seat belt laws ...  NaN
2315      False  This is an extremely rare horned Parthenon. No...  NaN
2316      False  This is a funny dog. Weird toes. Won't come do...  NaN
2317      False  This is an Albanian 3 1/2 legged  Episcopalian...  NaN
2318      False     Can take selfies 11/10 https://t.co/ws2AMaNwPW  NaN
2319      False  Very concerned about fellow dog trapped in com...  NaN
2320      False  Not familiar with this breed. No tail (weird)...  NaN
2321      False  Oh my. Here you are seeing an Adobe Setter giv...  NaN
2322      False  Can stand on stump for what seems like a while...  NaN
2323      False  This appears to be a Mongolian Presbyterian mi...  NaN
2324      False  Here we have a well-established sunblockerspan...  NaN
2325      False  Let's hope this flight isn't Malaysian (lol). ...  NaN
2326      False  Here we have a northern speckled Rhododendron...  NaN
2327      False  This is the happiest dog you will ever see. Ve...  NaN
2328      False  Here is the Rand Paul of retrievers folks! He'...  NaN
2329      False  My oh my. This is a rare blond Canadian terrie...  NaN
2330      False  Here is a Siberian heavily armored polar bear ...  NaN
2331      False  This is an odd dog. Hard on the outside but lo...  NaN
2332      False  This is a truly beautiful English Wilson Staff...  NaN
2333      False  Here we have a 1949 1st generation vulpix. Enj...  NaN
2334      False  This is a purebred Piers Morgan. Loves to Netf...  NaN
2335      False  Here is a very happy pup. Big fan of well-main...  NaN
2336      False  This is a western brown Mitsubishi terrier. Up...  NaN
2337      False  Here we have a Japanese Irish Setter. Lost eye...  NaN

                              ...                 quoted_status  \
0                             ...                           NaN
1                             ...                           NaN
2                             ...                           NaN
3                             ...                           NaN
4                             ...                           NaN
5                             ...                           NaN
6                             ...                           NaN
7                             ...                           NaN
```

8	...	NaN
9	...	NaN
10	...	NaN
11	...	NaN
12	...	NaN
13	...	NaN
14	...	NaN
15	...	NaN
16	...	NaN
17	...	NaN
18	...	NaN
19	...	NaN
20	...	NaN
21	...	NaN
22	...	NaN
23	...	NaN
24	...	NaN
25	...	NaN
26	...	NaN
27	...	NaN
28	...	NaN
29	...	NaN
...	...	...
2308	...	NaN
2309	...	NaN
2310	...	NaN
2311	...	NaN
2312	...	NaN
2313	...	NaN
2314	...	NaN
2315	...	NaN
2316	...	NaN
2317	...	NaN
2318	...	NaN
2319	...	NaN
2320	...	NaN
2321	...	NaN
2322	...	NaN
2323	...	NaN
2324	...	NaN
2325	...	NaN
2326	...	NaN
2327	...	NaN
2328	...	NaN
2329	...	NaN
2330	...	NaN
2331	...	NaN
2332	...	NaN

```
2333                       ...                                  NaN
2334                       ...                                  NaN
2335                       ...                                  NaN
2336                       ...                                  NaN
2337                       ...                                  NaN

      quoted_status_id quoted_status_id_str  quoted_status_permalink  \
0                  NaN                  NaN                      NaN
1                  NaN                  NaN                      NaN
2                  NaN                  NaN                      NaN
3                  NaN                  NaN                      NaN
4                  NaN                  NaN                      NaN
5                  NaN                  NaN                      NaN
6                  NaN                  NaN                      NaN
7                  NaN                  NaN                      NaN
8                  NaN                  NaN                      NaN
9                  NaN                  NaN                      NaN
10                 NaN                  NaN                      NaN
11                 NaN                  NaN                      NaN
12                 NaN                  NaN                      NaN
13                 NaN                  NaN                      NaN
14                 NaN                  NaN                      NaN
15                 NaN                  NaN                      NaN
16                 NaN                  NaN                      NaN
17                 NaN                  NaN                      NaN
18                 NaN                  NaN                      NaN
19                 NaN                  NaN                      NaN
20                 NaN                  NaN                      NaN
21                 NaN                  NaN                      NaN
22                 NaN                  NaN                      NaN
23                 NaN                  NaN                      NaN
24                 NaN                  NaN                      NaN
25                 NaN                  NaN                      NaN
26                 NaN                  NaN                      NaN
27                 NaN                  NaN                      NaN
28                 NaN                  NaN                      NaN
29                 NaN                  NaN                      NaN
...                ...                  ...                      ...
2308               NaN                  NaN                      NaN
2309               NaN                  NaN                      NaN
2310               NaN                  NaN                      NaN
2311               NaN                  NaN                      NaN
2312               NaN                  NaN                      NaN
2313               NaN                  NaN                      NaN
2314               NaN                  NaN                      NaN
2315               NaN                  NaN                      NaN
2316               NaN                  NaN                      NaN
2317               NaN                  NaN                      NaN
```

```
2318               NaN                  NaN                        NaN
2319               NaN                  NaN                        NaN
2320               NaN                  NaN                        NaN
2321               NaN                  NaN                        NaN
2322               NaN                  NaN                        NaN
2323               NaN                  NaN                        NaN
2324               NaN                  NaN                        NaN
2325               NaN                  NaN                        NaN
2326               NaN                  NaN                        NaN
2327               NaN                  NaN                        NaN
2328               NaN                  NaN                        NaN
2329               NaN                  NaN                        NaN
2330               NaN                  NaN                        NaN
2331               NaN                  NaN                        NaN
2332               NaN                  NaN                        NaN
2333               NaN                  NaN                        NaN
2334               NaN                  NaN                        NaN
2335               NaN                  NaN                        NaN
2336               NaN                  NaN                        NaN
2337               NaN                  NaN                        NaN

      retweet_count  retweeted  retweeted_status  \
0              8291      False               NaN
1              6123      False               NaN
2              4054      False               NaN
3              8428      False               NaN
4              9130      False               NaN
5              3039      False               NaN
6              2010      False               NaN
7             18389      False               NaN
8              4176      False               NaN
9              7192      False               NaN
10             7147      False               NaN
11             4853      False               NaN
12             9787      False               NaN
13             4413      False               NaN
14             2193      False               NaN
15             5234      False               NaN
16             4393      False               NaN
17             4178      False               NaN
18             3456      False               NaN
19             3407      False               NaN
20             5258      False               NaN
21            11399      False               NaN
22            17726      False               NaN
23            10157      False               NaN
24             5814      False               NaN
25             7582      False               NaN
```

```
26             3197      False               NaN
27             4354      False               NaN
28             3126      False               NaN
29                4      False               NaN
...             ...        ...               ...
2308            322      False               NaN
2309             40      False               NaN
2310             83      False               NaN
2311             89      False               NaN
2312            561      False               NaN
2313             71      False               NaN
2314            132      False               NaN
2315             90      False               NaN
2316            344      False               NaN
2317             64      False               NaN
2318             76      False               NaN
2319             35      False               NaN
2320           6398      False               NaN
2321             11      False               NaN
2322             66      False               NaN
2323             72      False               NaN
2324             44      False               NaN
2325            159      False               NaN
2326             58      False               NaN
2327            213      False               NaN
2328             57      False               NaN
2329            138      False               NaN
2330            235      False               NaN
2331            831      False               NaN
2332             59      False               NaN
2333             43      False               NaN
2334            138      False               NaN
2335             43      False               NaN
2336             47      False               NaN
2337            499      False               NaN

                                                    source truncated  \
0     <a href="http://twitter.com/download/iphone" r...     False
1     <a href="http://twitter.com/download/iphone" r...     False
2     <a href="http://twitter.com/download/iphone" r...     False
3     <a href="http://twitter.com/download/iphone" r...     False
4     <a href="http://twitter.com/download/iphone" r...     False
5     <a href="http://twitter.com/download/iphone" r...     False
6     <a href="http://twitter.com/download/iphone" r...     False
7     <a href="http://twitter.com/download/iphone" r...     False
8     <a href="http://twitter.com/download/iphone" r...     False
9     <a href="http://twitter.com/download/iphone" r...     False
10    <a href="http://twitter.com/download/iphone" r...     False
```

11	<a href="http://twitter.com/download/iphone" r...		False
12	<a href="http://twitter.com/download/iphone" r...		False
13	<a href="http://twitter.com/download/iphone" r...		False
14	<a href="http://twitter.com/download/iphone" r...		False
15	<a href="http://twitter.com/download/iphone" r...		False
16	<a href="http://twitter.com/download/iphone" r...		False
17	<a href="http://twitter.com/download/iphone" r...		False
18	<a href="http://twitter.com/download/iphone" r...		False
19	<a href="http://twitter.com/download/iphone" r...		False
20	<a href="http://twitter.com/download/iphone" r...		False
21	<a href="http://twitter.com/download/iphone" r...		False
22	<a href="http://twitter.com/download/iphone" r...		False
23	<a href="http://twitter.com/download/iphone" r...		False
24	<a href="http://twitter.com/download/iphone" r...		False
25	<a href="http://twitter.com/download/iphone" r...		False
26	<a href="http://twitter.com/download/iphone" r...		False
27	<a href="http://twitter.com/download/iphone" r...		False
28	<a href="http://twitter.com/download/iphone" r...		False
29	<a href="http://twitter.com/download/iphone" r...		False
...		...	...
2308	<a href="http://twitter.com/download/iphone" r...		False
2309	<a href="http://twitter.com/download/iphone" r...		False
2310	<a href="http://twitter.com/download/iphone" r...		False
2311	<a href="http://twitter.com/download/iphone" r...		False
2312	<a href="http://twitter.com/download/iphone" r...		False
2313	<a href="http://twitter.com/download/iphone" r...		False
2314	<a href="http://twitter.com/download/iphone" r...		False
2315	<a href="http://twitter.com/download/iphone" r...		False
2316	<a href="http://twitter.com/download/iphone" r...		False
2317	<a href="http://twitter.com/download/iphone" r...		False
2318	<a href="http://twitter.com/download/iphone" r...		False
2319	<a href="http://twitter.com/download/iphone" r...		False
2320	<a href="http://twitter.com/download/iphone" r...		False
2321	<a href="http://twitter.com/download/iphone" r...		False
2322	<a href="http://twitter.com/download/iphone" r...		False
2323	<a href="http://twitter.com/download/iphone" r...		False
2324	<a href="http://twitter.com/download/iphone" r...		False
2325	<a href="http://twitter.com/download/iphone" r...		False
2326	<a href="http://twitter.com/download/iphone" r...		False
2327	<a href="http://twitter.com/download/iphone" r...		False
2328	<a href="http://twitter.com/download/iphone" r...		False
2329	<a href="http://twitter.com/download/iphone" r...		False
2330	<a href="http://twitter.com/download/iphone" r...		False
2331	<a href="http://twitter.com/download/iphone" r...		False
2332	<a href="http://twitter.com/download/iphone" r...		False
2333	<a href="http://twitter.com/download/iphone" r...		False
2334	<a href="http://twitter.com/download/iphone" r...		False
2335	<a href="http://twitter.com/download/iphone" r...		False

```
2336  <a href="http://twitter.com/download/iphone" r...     False
2337  <a href="http://twitter.com/download/iphone" r...     False

                                                    user
0     {'id': 4196983835, 'id_str': '4196983835', 'na...
1     {'id': 4196983835, 'id_str': '4196983835', 'na...
2     {'id': 4196983835, 'id_str': '4196983835', 'na...
3     {'id': 4196983835, 'id_str': '4196983835', 'na...
4     {'id': 4196983835, 'id_str': '4196983835', 'na...
5     {'id': 4196983835, 'id_str': '4196983835', 'na...
6     {'id': 4196983835, 'id_str': '4196983835', 'na...
7     {'id': 4196983835, 'id_str': '4196983835', 'na...
8     {'id': 4196983835, 'id_str': '4196983835', 'na...
9     {'id': 4196983835, 'id_str': '4196983835', 'na...
10    {'id': 4196983835, 'id_str': '4196983835', 'na...
11    {'id': 4196983835, 'id_str': '4196983835', 'na...
12    {'id': 4196983835, 'id_str': '4196983835', 'na...
13    {'id': 4196983835, 'id_str': '4196983835', 'na...
14    {'id': 4196983835, 'id_str': '4196983835', 'na...
15    {'id': 4196983835, 'id_str': '4196983835', 'na...
16    {'id': 4196983835, 'id_str': '4196983835', 'na...
17    {'id': 4196983835, 'id_str': '4196983835', 'na...
18    {'id': 4196983835, 'id_str': '4196983835', 'na...
19    {'id': 4196983835, 'id_str': '4196983835', 'na...
20    {'id': 4196983835, 'id_str': '4196983835', 'na...
21    {'id': 4196983835, 'id_str': '4196983835', 'na...
22    {'id': 4196983835, 'id_str': '4196983835', 'na...
23    {'id': 4196983835, 'id_str': '4196983835', 'na...
24    {'id': 4196983835, 'id_str': '4196983835', 'na...
25    {'id': 4196983835, 'id_str': '4196983835', 'na...
26    {'id': 4196983835, 'id_str': '4196983835', 'na...
27    {'id': 4196983835, 'id_str': '4196983835', 'na...
28    {'id': 4196983835, 'id_str': '4196983835', 'na...
29    {'id': 4196983835, 'id_str': '4196983835', 'na...
...                                                   ...
2308  {'id': 4196983835, 'id_str': '4196983835', 'na...
2309  {'id': 4196983835, 'id_str': '4196983835', 'na...
2310  {'id': 4196983835, 'id_str': '4196983835', 'na...
2311  {'id': 4196983835, 'id_str': '4196983835', 'na...
2312  {'id': 4196983835, 'id_str': '4196983835', 'na...
2313  {'id': 4196983835, 'id_str': '4196983835', 'na...
2314  {'id': 4196983835, 'id_str': '4196983835', 'na...
2315  {'id': 4196983835, 'id_str': '4196983835', 'na...
2316  {'id': 4196983835, 'id_str': '4196983835', 'na...
2317  {'id': 4196983835, 'id_str': '4196983835', 'na...
2318  {'id': 4196983835, 'id_str': '4196983835', 'na...
2319  {'id': 4196983835, 'id_str': '4196983835', 'na...
2320  {'id': 4196983835, 'id_str': '4196983835', 'na...
```

```
2321  {'id': 4196983835, 'id_str': '4196983835', 'na...
2322  {'id': 4196983835, 'id_str': '4196983835', 'na...
2323  {'id': 4196983835, 'id_str': '4196983835', 'na...
2324  {'id': 4196983835, 'id_str': '4196983835', 'na...
2325  {'id': 4196983835, 'id_str': '4196983835', 'na...
2326  {'id': 4196983835, 'id_str': '4196983835', 'na...
2327  {'id': 4196983835, 'id_str': '4196983835', 'na...
2328  {'id': 4196983835, 'id_str': '4196983835', 'na...
2329  {'id': 4196983835, 'id_str': '4196983835', 'na...
2330  {'id': 4196983835, 'id_str': '4196983835', 'na...
2331  {'id': 4196983835, 'id_str': '4196983835', 'na...
2332  {'id': 4196983835, 'id_str': '4196983835', 'na...
2333  {'id': 4196983835, 'id_str': '4196983835', 'na...
2334  {'id': 4196983835, 'id_str': '4196983835', 'na...
2335  {'id': 4196983835, 'id_str': '4196983835', 'na...
2336  {'id': 4196983835, 'id_str': '4196983835', 'na...
2337  {'id': 4196983835, 'id_str': '4196983835', 'na...

[2338 rows x 32 columns]
```

```
In [28]: df_tweetdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2338 entries, 0 to 2337
Data columns (total 32 columns):
contributors                    0 non-null float64
coordinates                     0 non-null float64
created_at                      2338 non-null datetime64[ns]
display_text_range              2338 non-null object
entities                        2338 non-null object
extended_entities               2065 non-null object
favorite_count                  2338 non-null int64
favorited                       2338 non-null bool
full_text                       2338 non-null object
geo                             0 non-null float64
id                              2338 non-null int64
id_str                          2338 non-null int64
in_reply_to_screen_name         77 non-null object
in_reply_to_status_id           77 non-null float64
in_reply_to_status_id_str       77 non-null float64
in_reply_to_user_id             77 non-null float64
in_reply_to_user_id_str         77 non-null float64
is_quote_status                 2338 non-null bool
lang                            2338 non-null object
place                           1 non-null object
possibly_sensitive              2203 non-null float64
possibly_sensitive_appealable   2203 non-null float64
quoted_status                   24 non-null object
```

```
quoted_status_id                26 non-null float64
quoted_status_id_str            26 non-null float64
quoted_status_permalink         26 non-null object
retweet_count                   2338 non-null int64
retweeted                       2338 non-null bool
retweeted_status                167 non-null object
source                          2338 non-null object
truncated                       2338 non-null bool
user                            2338 non-null object
dtypes: bool(4), datetime64[ns](1), float64(11), int64(4), object(12)
memory usage: 520.6+ KB
```

```
In [29]: columns_of_interest = ['id','retweet_count','retweeted','retweeted_status','favorite_co
         df_tweetdata = df_tweetdata[columns_of_interest]

In [30]: df_tweetdata.head()
```

Out[30]:

	id	retweet_count	retweeted	retweeted_status	favorite_count
0	892420643555336193	8291	False	NaN	37950
1	892177421306343426	6123	False	NaN	32588
2	891815181378084864	4054	False	NaN	24536
3	891689557279858688	8428	False	NaN	41291
4	891327558926688256	9130	False	NaN	39478

```
In [31]: df_tweetdata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2338 entries, 0 to 2337
Data columns (total 5 columns):
id                  2338 non-null int64
retweet_count       2338 non-null int64
retweeted           2338 non-null bool
retweeted_status    167 non-null object
favorite_count      2338 non-null int64
dtypes: bool(1), int64(3), object(1)
memory usage: 75.4+ KB
```

```
In [32]: df_tweetdata.describe()
```

Out[32]:

	id	retweet_count	favorite_count
count	2.338000e+03	2338.000000	2338.000000

```
mean   7.422385e+17    2920.497861     7946.362703
std    6.834068e+16    4919.812857    12308.340561
min    6.660209e+17       0.000000        0.000000
25%    6.783509e+17     584.250000     1368.750000
50%    7.186224e+17    1362.000000     3453.500000
75%    7.986971e+17    3401.250000     9721.000000
max    8.924206e+17   83352.000000   163892.000000
```

In [33]:
```
# check the 'name' column in the df_archive table for any 'None' values.
df_archive[df_archive['name']=='None']
```

Out[33]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
5	891087950875897856	NaN	NaN	
7	890729181411237888	NaN	NaN	
12	889665388333682689	NaN	NaN	
24	887343217045368832	NaN	NaN	
25	887101392804085760	NaN	NaN	
30	886267009285017600	8.862664e+17	2.281182e+09	
32	886054160059072513	NaN	NaN	
35	885518971528720385	NaN	NaN	
37	885167619883638784	NaN	NaN	
41	884441805382717440	NaN	NaN	
42	884247878851493888	NaN	NaN	
47	883117836046086144	NaN	NaN	
55	881633300179243008	8.816070e+17	4.738443e+07	
59	880872448815771648	NaN	NaN	
62	880095782870896641	NaN	NaN	
64	879674319642796034	8.795538e+17	3.105441e+09	
72	878604707211726852	NaN	NaN	
78	877611172832227328	NaN	NaN	
83	876537666061221889	NaN	NaN	
88	875097192612077568	NaN	NaN	
89	875021211251597312	NaN	NaN	
93	874057562936811520	NaN	NaN	
96	873580283840344065	NaN	NaN	
99	872967104147763200	NaN	NaN	
100	872820683541237760	NaN	NaN	
101	872668790621863937	NaN	NaN	
103	872486979161796608	NaN	NaN	
110	871102520638267392	NaN	NaN	
112	870804317367881728	NaN	NaN	
113	870726314365509632	8.707262e+17	1.648776e+07	
...	...	...	...	
2294	667138269671505920	NaN	NaN	
2298	667070482143944705	6.670655e+17	4.196984e+09	
2299	667065535570550784	NaN	NaN	
2301	667044094246576128	NaN	NaN	
2305	666837028449972224	NaN	NaN	

```
2306  666835007768551424                  NaN                  NaN
2307  666826780179869698                  NaN                  NaN
2310  666786068205871104                  NaN                  NaN
2316  666649482315059201                  NaN                  NaN
2320  666437273139982337                  NaN                  NaN
2321  666435652385423360                  NaN                  NaN
2322  666430724426358785                  NaN                  NaN
2323  666428276349472768                  NaN                  NaN
2324  666421158376562688                  NaN                  NaN
2328  666396247373291520                  NaN                  NaN
2329  666373753744588802                  NaN                  NaN
2330  666362758909284353                  NaN                  NaN
2331  666353288456101888                  NaN                  NaN
2332  666345417576210432                  NaN                  NaN
2336  666273097616637952                  NaN                  NaN
2337  666268910803644416                  NaN                  NaN
2338  666104133288665088                  NaN                  NaN
2339  666102155909144576                  NaN                  NaN
2340  666099513787052032                  NaN                  NaN
2341  666094000022159362                  NaN                  NaN
2342  666082916733198337                  NaN                  NaN
2343  666073100786774016                  NaN                  NaN
2344  666071193221509120                  NaN                  NaN
2351  666049248165822465                  NaN                  NaN
2355  666020888022790149                  NaN                  NaN

                      timestamp  \
5     2017-07-29 00:08:17 +0000
7     2017-07-28 00:22:40 +0000
12    2017-07-25 01:55:32 +0000
24    2017-07-18 16:08:03 +0000
25    2017-07-18 00:07:08 +0000
30    2017-07-15 16:51:35 +0000
32    2017-07-15 02:45:48 +0000
35    2017-07-13 15:19:09 +0000
37    2017-07-12 16:03:00 +0000
41    2017-07-10 15:58:53 +0000
42    2017-07-10 03:08:17 +0000
47    2017-07-07 00:17:54 +0000
55    2017-07-02 21:58:53 +0000
59    2017-06-30 19:35:32 +0000
62    2017-06-28 16:09:20 +0000
64    2017-06-27 12:14:36 +0000
72    2017-06-24 13:24:20 +0000
78    2017-06-21 19:36:23 +0000
83    2017-06-18 20:30:39 +0000
88    2017-06-14 21:06:43 +0000
89    2017-06-14 16:04:48 +0000
```

```
93    2017-06-12 00:15:36 +0000
96    2017-06-10 16:39:04 +0000
99    2017-06-09 00:02:31 +0000
100   2017-06-08 14:20:41 +0000
101   2017-06-08 04:17:07 +0000
103   2017-06-07 16:14:40 +0000
110   2017-06-03 20:33:19 +0000
112   2017-06-03 00:48:22 +0000
113   2017-06-02 19:38:25 +0000
...                        ...
2294  2015-11-19 00:32:12 +0000
2298  2015-11-18 20:02:51 +0000
2299  2015-11-18 19:43:11 +0000
2301  2015-11-18 18:17:59 +0000
2305  2015-11-18 04:35:11 +0000
2306  2015-11-18 04:27:09 +0000
2307  2015-11-18 03:54:28 +0000
2310  2015-11-18 01:12:41 +0000
2316  2015-11-17 16:09:56 +0000
2320  2015-11-17 02:06:42 +0000
2321  2015-11-17 02:00:15 +0000
2322  2015-11-17 01:40:41 +0000
2323  2015-11-17 01:30:57 +0000
2324  2015-11-17 01:02:40 +0000
2328  2015-11-16 23:23:41 +0000
2329  2015-11-16 21:54:18 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2332  2015-11-16 20:01:42 +0000
2336  2015-11-16 15:14:19 +0000
2337  2015-11-16 14:57:41 +0000
2338  2015-11-16 04:02:55 +0000
2339  2015-11-16 03:55:04 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
2343  2015-11-16 01:59:36 +0000
2344  2015-11-16 01:52:02 +0000
2351  2015-11-16 00:24:50 +0000
2355  2015-11-15 22:32:08 +0000

                                                  source  \
5     <a href="http://twitter.com/download/iphone" r...
7     <a href="http://twitter.com/download/iphone" r...
12    <a href="http://twitter.com/download/iphone" r...
24    <a href="http://twitter.com/download/iphone" r...
25    <a href="http://twitter.com/download/iphone" r...
30    <a href="http://twitter.com/download/iphone" r...
```

```
32    <a href="http://twitter.com/download/iphone" r...
35    <a href="http://twitter.com/download/iphone" r...
37    <a href="http://twitter.com/download/iphone" r...
41    <a href="http://twitter.com/download/iphone" r...
42    <a href="http://twitter.com/download/iphone" r...
47    <a href="http://twitter.com/download/iphone" r...
55    <a href="http://twitter.com/download/iphone" r...
59    <a href="http://twitter.com/download/iphone" r...
62    <a href="http://twitter.com/download/iphone" r...
64    <a href="http://twitter.com/download/iphone" r...
72    <a href="http://twitter.com/download/iphone" r...
78    <a href="http://twitter.com/download/iphone" r...
83    <a href="http://twitter.com/download/iphone" r...
88    <a href="http://twitter.com/download/iphone" r...
89    <a href="http://twitter.com/download/iphone" r...
93    <a href="http://twitter.com/download/iphone" r...
96    <a href="http://twitter.com/download/iphone" r...
99    <a href="http://twitter.com/download/iphone" r...
100   <a href="http://twitter.com/download/iphone" r...
101   <a href="http://twitter.com/download/iphone" r...
103   <a href="http://twitter.com/download/iphone" r...
110   <a href="http://twitter.com/download/iphone" r...
112   <a href="http://twitter.com/download/iphone" r...
113   <a href="http://twitter.com/download/iphone" r...
...                                                  ...
2294  <a href="http://twitter.com/download/iphone" r...
2298  <a href="http://twitter.com/download/iphone" r...
2299  <a href="http://twitter.com/download/iphone" r...
2301  <a href="http://twitter.com/download/iphone" r...
2305  <a href="http://twitter.com/download/iphone" r...
2306  <a href="http://twitter.com/download/iphone" r...
2307  <a href="http://twitter.com/download/iphone" r...
2310  <a href="http://twitter.com/download/iphone" r...
2316  <a href="http://twitter.com/download/iphone" r...
2320  <a href="http://twitter.com/download/iphone" r...
2321  <a href="http://twitter.com/download/iphone" r...
2322  <a href="http://twitter.com/download/iphone" r...
2323  <a href="http://twitter.com/download/iphone" r...
2324  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2329  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2332  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
```

```
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...

                                                  text  retweeted_status_id  \
5     Here we have a majestic great white breaching ...                  NaN
7     When you watch your owner call another dog a g...                  NaN
12    Here's a puppo that seems to be on the fence a...                  NaN
24    You may not have known you needed to see this ...                  NaN
25    This... is a Jubilant Antarctic House Bear. We...                  NaN
30    @NonWhiteHat @MayhewMayhem omg hello tanner yo...                  NaN
32    RT @Athletics: 12/10 #BATP https://t.co/WxwJmv...        8.860537e+17
35    I have a new hero and his name is Howard. 14/1...                  NaN
37    Here we have a corgi undercover as a malamute...                  NaN
41    I present to you, Pup in Hat. Pup in Hat is gr...                  NaN
42    OMG HE DIDN'T MEAN TO HE WAS JUST TRYING A LIT...                  NaN
47    Please only send dogs. We don't rate mechanics...                  NaN
55    @roushfenway These are good dogs but 17/10 is ...                  NaN
59    Ugh not again. We only rate dogs. Please don't...                  NaN
62    Please don't send in photos without dogs in th...                  NaN
64                      @RealKentMurphy 14/10 confirmed                  NaN
72    Martha is stunning how h*ckin dare you. 13/10 ...                  NaN
78    RT @rachel2195: @dog_rates the boyfriend and h...        8.768508e+17
83    I can say with the pupmost confidence that the...                  NaN
88    You'll get your package when that precious man...                  NaN
89    Guys please stop sending pictures without any ...                  NaN
93    I can't believe this keeps happening. This, is...                  NaN
96    We usually don't rate Deck-bound Saskatoon Bla...                  NaN
99    Here's a very large dog. He has a date later. ...                  NaN
100   Here are my favorite #dogsatpollingstations \n...                  NaN
101   RT @loganamnosis: Penelope here is doing me qu...        8.726576e+17
103   We. Only. Rate. Dogs. Do not send in other thi...                  NaN
110   Never doubt a doggo 14/10 https://t.co/AbBLh2FZCH                  NaN
112   Real funny guys. Sending in a pic without a do...                  NaN
113   @ComplicitOwl @ShopWeRateDogs &gt;10/10 is res...                  NaN
...                                                 ...                  ...
2294  Extremely intelligent dog here. Has learned to...                  NaN
2298  After much debate this dog is being upgraded t...                  NaN
2299  Here we have a Hufflepuff. Loves vest. Eyes wi...                  NaN
2301            12/10 gimme now https://t.co/QZAnwgnOMB                  NaN
2305  My goodness. Very rare dog here. Large. Tail d...                  NaN
2306  These are Peruvian Feldspars. Their names are ...                  NaN
2307  12/10 simply brilliant pup https://t.co/V6ZzG4...                  NaN
2310  Unfamiliar with this breed. Ears pointy af. Wo...                  NaN
```

```
2316  Cool dog. Enjoys couch. Low monotone bark. Ver...                  NaN
2320  Here we see a lone northeastern Cumberbatch. H...                  NaN
2321  "Can you behave? You're ruining my wedding day...                  NaN
2322  Oh boy what a pup! Sunglasses take this one to...                  NaN
2323  Here we have an Austrian Pulitzer. Collectors ...                  NaN
2324  *internally screaming* 12/10 https://t.co/YMcr...                  NaN
2328  Oh goodness. A super rare northeast Qdoba kang...                  NaN
2329  Those are sunglasses and a jean jacket. 11/10 ...                  NaN
2330  Unique dog here. Very small. Lives in containe...                  NaN
2331  Here we have a mixed Asiago from the Galápagos...                  NaN
2332  Look at this jokester thinking seat belt laws ...                  NaN
2336     Can take selfies 11/10 https://t.co/ws2AMaNwPW                  NaN
2337  Very concerned about fellow dog trapped in com...                  NaN
2338  Not familiar with this breed. No tail (weird)...                  NaN
2339  Oh my. Here you are seeing an Adobe Setter giv...                  NaN
2340  Can stand on stump for what seems like a while...                  NaN
2341  This appears to be a Mongolian Presbyterian mi...                  NaN
2342  Here we have a well-established sunblockerspan...                  NaN
2343  Let's hope this flight isn't Malaysian (lol). ...                  NaN
2344  Here we have a northern speckled Rhododendron...                   NaN
2351  Here we have a 1949 1st generation vulpix. Enj...                  NaN
2355  Here we have a Japanese Irish Setter. Lost eye...                  NaN

      retweeted_status_user_id retweeted_status_timestamp  \
5                          NaN                         NaN
7                          NaN                         NaN
12                         NaN                         NaN
24                         NaN                         NaN
25                         NaN                         NaN
30                         NaN                         NaN
32                19607400.0  2017-07-15 02:44:07 +0000
35                         NaN                         NaN
37                         NaN                         NaN
41                         NaN                         NaN
42                         NaN                         NaN
47                         NaN                         NaN
55                         NaN                         NaN
59                         NaN                         NaN
62                         NaN                         NaN
64                         NaN                         NaN
72                         NaN                         NaN
78               512804507.0  2017-06-19 17:14:49 +0000
83                         NaN                         NaN
88                         NaN                         NaN
89                         NaN                         NaN
93                         NaN                         NaN
96                         NaN                         NaN
99                         NaN                         NaN
```

```
100                      NaN                         NaN
101             154767397.0  2017-06-08 03:32:35 +0000
103                      NaN                         NaN
110                      NaN                         NaN
112                      NaN                         NaN
113                      NaN                         NaN
...                      ...                         ...
2294                     NaN                         NaN
2298                     NaN                         NaN
2299                     NaN                         NaN
2301                     NaN                         NaN
2305                     NaN                         NaN
2306                     NaN                         NaN
2307                     NaN                         NaN
2310                     NaN                         NaN
2316                     NaN                         NaN
2320                     NaN                         NaN
2321                     NaN                         NaN
2322                     NaN                         NaN
2323                     NaN                         NaN
2324                     NaN                         NaN
2328                     NaN                         NaN
2329                     NaN                         NaN
2330                     NaN                         NaN
2331                     NaN                         NaN
2332                     NaN                         NaN
2336                     NaN                         NaN
2337                     NaN                         NaN
2338                     NaN                         NaN
2339                     NaN                         NaN
2340                     NaN                         NaN
2341                     NaN                         NaN
2342                     NaN                         NaN
2343                     NaN                         NaN
2344                     NaN                         NaN
2351                     NaN                         NaN
2355                     NaN                         NaN

                                              expanded_urls  rating_numerator  \
5     https://twitter.com/dog_rates/status/891087950...                 13
7     https://twitter.com/dog_rates/status/890729181...                 13
12    https://twitter.com/dog_rates/status/889665388...                 13
24    https://twitter.com/dog_rates/status/887343217...                 13
25    https://twitter.com/dog_rates/status/887101392...                 12
30                                                  NaN                 12
32    https://twitter.com/dog_rates/status/886053434...                 12
35    https://twitter.com/4bonds2carbon/status/88551...                 14
37    https://twitter.com/dog_rates/status/885167619...                 13
```

41	https://twitter.com/dog_rates/status/884441805...	14
42	https://twitter.com/kaijohnson_19/status/88396...	13
47	https://twitter.com/dog_rates/status/883117836...	13
55	NaN	17
59	https://twitter.com/dog_rates/status/880872448...	12
62	https://twitter.com/dog_rates/status/880095782...	11
64	NaN	14
72	https://twitter.com/bbcworld/status/8785998685...	13
78	https://twitter.com/rachel2195/status/87685077...	14
83	https://twitter.com/mpstowerham/status/8761629...	14
88	https://twitter.com/drboondoc/status/874413398...	13
89	https://twitter.com/dog_rates/status/875021211...	12
93	https://twitter.com/dog_rates/status/874057562...	12
96	https://twitter.com/dog_rates/status/873580283...	13
99	https://twitter.com/dog_rates/status/872967104...	12
100	https://twitter.com/dog_rates/status/872820683...	13
101	https://twitter.com/loganamnosis/status/872657...	14
103	https://twitter.com/dog_rates/status/872486979...	12
110	https://twitter.com/animalcog/status/871075758...	14
112	https://twitter.com/dog_rates/status/870804317...	11
113	NaN	10
...	...	...
2294	https://twitter.com/dog_rates/status/667138269...	10
2298	NaN	10
2299	https://twitter.com/dog_rates/status/667065535...	8
2301	https://twitter.com/dog_rates/status/667044094...	12
2305	https://twitter.com/dog_rates/status/666837028...	3
2306	https://twitter.com/dog_rates/status/666835007...	10
2307	https://twitter.com/dog_rates/status/666826780...	12
2310	https://twitter.com/dog_rates/status/666786068...	2
2316	https://twitter.com/dog_rates/status/666649482...	4
2320	https://twitter.com/dog_rates/status/666437273...	7
2321	https://twitter.com/dog_rates/status/666435652...	10
2322	https://twitter.com/dog_rates/status/666430724...	6
2323	https://twitter.com/dog_rates/status/666428276...	7
2324	https://twitter.com/dog_rates/status/666421158...	12
2328	https://twitter.com/dog_rates/status/666396247...	9
2329	https://twitter.com/dog_rates/status/666373753...	11
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2332	https://twitter.com/dog_rates/status/666345417...	10
2336	https://twitter.com/dog_rates/status/666273097...	11
2337	https://twitter.com/dog_rates/status/666268910...	10
2338	https://twitter.com/dog_rates/status/666104133...	1
2339	https://twitter.com/dog_rates/status/666102155...	11
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6

```
2343  https://twitter.com/dog_rates/status/666073100...                10
2344  https://twitter.com/dog_rates/status/666071193...                 9
2351  https://twitter.com/dog_rates/status/666049248...                 5
2355  https://twitter.com/dog_rates/status/666020888...                 8

      rating_denominator  name  doggo floofer  pupper  puppo
5                     10  None   None    None    None   None
7                     10  None   None    None    None   None
12                    10  None   None    None    None  puppo
24                    10  None   None    None    None   None
25                    10  None   None    None    None   None
30                    10  None   None    None    None   None
32                    10  None   None    None    None   None
35                    10  None   None    None    None   None
37                    10  None   None    None    None   None
41                    10  None   None    None    None   None
42                    10  None   None    None    None   None
47                    10  None   None    None    None   None
55                    10  None   None    None    None   None
59                    10  None   None    None    None   None
62                    10  None   None    None    None   None
64                    10  None   None    None    None   None
72                    10  None   None    None    None   None
78                    10  None   None    None  pupper   None
83                    10  None   None    None    None   None
88                    10  None   None    None    None   None
89                    10  None   None    None    None   None
93                    10  None   None    None    None   None
96                    10  None   None    None    None   None
99                    10  None  doggo    None    None   None
100                   10  None   None    None    None   None
101                   10  None   None    None    None   None
103                   10  None   None    None    None   None
110                   10  None  doggo    None    None   None
112                   10  None   None    None    None   None
113                   10  None   None    None    None   None
...                  ...   ...    ...     ...     ...    ...
2294                  10  None   None    None    None   None
2298                  10  None   None    None    None   None
2299                  10  None   None    None    None   None
2301                  10  None   None    None    None   None
2305                  10  None   None    None    None   None
2306                  10  None   None    None    None   None
2307                  10  None   None    None    None   None
2310                  10  None   None    None    None   None
2316                  10  None   None    None    None   None
2320                  10  None   None    None    None   None
2321                  10  None   None    None    None   None
```

```
2322                  10  None   None    None    None   None
2323                  10  None   None    None    None   None
2324                  10  None   None    None    None   None
2328                  10  None   None    None    None   None
2329                  10  None   None    None    None   None
2330                  10  None   None    None    None   None
2331                  10  None   None    None    None   None
2332                  10  None   None    None    None   None
2336                  10  None   None    None    None   None
2337                  10  None   None    None    None   None
2338                  10  None   None    None    None   None
2339                  10  None   None    None    None   None
2340                  10  None   None    None    None   None
2341                  10  None   None    None    None   None
2342                  10  None   None    None    None   None
2343                  10  None   None    None    None   None
2344                  10  None   None    None    None   None
2351                  10  None   None    None    None   None
2355                  10  None   None    None    None   None

[745 rows x 17 columns]
```

In [34]:
```
# check the 'expanded_urls' column in the df_archive table for any 'None' values to ide
df_archive[df_archive['expanded_urls'].isnull()]
```

Out[34]:
```
                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
30    886267009285017600           8.862664e+17         2.281182e+09
55    881633300179243008           8.816070e+17         4.738443e+07
64    879674319642796034           8.795538e+17         3.105441e+09
113   870726314365509632           8.707262e+17         1.648776e+07
148   863427515083354112           8.634256e+17         7.759620e+07
179   857214891891077121           8.571567e+17         1.806710e+08
185   856330835276025856                    NaN                  NaN
186   856288084350160898           8.562860e+17         2.792810e+08
188   855862651834028034           8.558616e+17         1.943518e+08
189   855860136149123072           8.558585e+17         1.361572e+07
218   850333567704068097           8.503288e+17         2.195506e+07
228   848213670039564288           8.482121e+17         4.196984e+09
234   847617282490613760           8.476062e+17         4.196984e+09
274   840698636975636481           8.406983e+17         8.405479e+17
290   838150277551247360           8.381455e+17         2.195506e+07
291   838085839343206401           8.380855e+17         2.894131e+09
313   835246439529840640           8.352460e+17         2.625958e+07
342   832088576586297345           8.320875e+17         3.058208e+07
346   831926988323639298           8.319030e+17         2.068372e+07
375   828361771580813312                    NaN                  NaN
387   826598799820865537           8.265984e+17         4.196984e+09
409   823334489516937216           8.233264e+17         1.582854e+09
```

```
427   821153421864615936          8.211526e+17        1.132119e+08
498   813130366689148928          8.131273e+17        4.196984e+09
513   811647686436880384          8.116272e+17        4.196984e+09
570   801854953262350336          8.018543e+17        1.185634e+07
576   800859414831898624          8.008580e+17        2.918590e+08
611   797165961484890113          7.971238e+17        2.916630e+07
701   786051337297522688          7.727430e+17        7.305050e+17
707   785515384317313025                   NaN                 NaN
843   766714921925144576          7.667118e+17        4.196984e+09
857   763956972077010945          7.638652e+17        1.584641e+07
967   750381685133418496          7.501805e+17        4.717297e+09
1005  747651430853525504          7.476487e+17        4.196984e+09
1080  738891149612572673          7.384119e+17        3.589728e+08
1295  707983188426153984          7.079801e+17        2.319108e+09
1345  704491224099647488          7.044857e+17        2.878549e+07
1445  696518437233913856                   NaN                 NaN
1446  696490539101908992          6.964887e+17        4.196984e+09
1474  693644216740769793          6.936422e+17        4.196984e+09
1479  693582294167244802          6.935722e+17        1.198989e+09
1497  692423280028966913          6.924173e+17        4.196984e+09
1523  690607260360429569          6.903413e+17        4.670367e+08
1598  686035780142297088          6.860340e+17        4.196984e+09
1605  685681090388975616          6.855479e+17        4.196984e+09
1618  684969860808454144          6.849598e+17        4.196984e+09
1663  682808988178739200          6.827884e+17        4.196984e+09
1689  681340665377193984          6.813394e+17        4.196984e+09
1774  678023323247357953          6.780211e+17        4.196984e+09
1819  676590572941893632          6.765883e+17        4.196984e+09
1844  675849018447167488          6.758457e+17        4.196984e+09
1895  674742531037511680          6.747400e+17        4.196984e+09
1905  674606911342424069          6.744689e+17        4.196984e+09
1914  674330906434379776          6.658147e+17        1.637468e+07
1940  673716320723169284          6.737159e+17        4.196984e+09
2038  671550332464455680          6.715449e+17        4.196984e+09
2149  669684865554620416          6.693544e+17        4.196984e+09
2189  668967877119254528          6.689207e+17        2.143566e+07
2298  667070482143944705          6.670655e+17        4.196984e+09

                      timestamp  \
30    2017-07-15 16:51:35 +0000
55    2017-07-02 21:58:53 +0000
64    2017-06-27 12:14:36 +0000
113   2017-06-02 19:38:25 +0000
148   2017-05-13 16:15:35 +0000
179   2017-04-26 12:48:51 +0000
185   2017-04-24 02:15:55 +0000
186   2017-04-23 23:26:03 +0000
188   2017-04-22 19:15:32 +0000
```

```
189   2017-04-22 19:05:32 +0000
218   2017-04-07 13:04:55 +0000
228   2017-04-01 16:41:12 +0000
234   2017-03-31 01:11:22 +0000
274   2017-03-11 22:59:09 +0000
290   2017-03-04 22:12:52 +0000
291   2017-03-04 17:56:49 +0000
313   2017-02-24 21:54:03 +0000
342   2017-02-16 04:45:50 +0000
346   2017-02-15 18:03:45 +0000
375   2017-02-05 21:56:51 +0000
387   2017-02-01 01:11:25 +0000
409   2017-01-23 00:56:15 +0000
427   2017-01-17 00:33:26 +0000
498   2016-12-25 21:12:41 +0000
513   2016-12-21 19:01:02 +0000
570   2016-11-24 18:28:13 +0000
576   2016-11-22 00:32:18 +0000
611   2016-11-11 19:55:50 +0000
701   2016-10-12 03:50:17 +0000
707   2016-10-10 16:20:36 +0000
843   2016-08-19 19:14:16 +0000
857   2016-08-12 04:35:10 +0000
967   2016-07-05 17:31:49 +0000
1005  2016-06-28 04:42:46 +0000
1080  2016-06-04 00:32:32 +0000
1295  2016-03-10 17:35:20 +0000
1345  2016-03-01 02:19:31 +0000
1445  2016-02-08 02:18:30 +0000
1446  2016-02-08 00:27:39 +0000
1474  2016-01-31 03:57:23 +0000
1479  2016-01-30 23:51:19 +0000
1497  2016-01-27 19:05:49 +0000
1523  2016-01-22 18:49:36 +0000
1598  2016-01-10 04:04:10 +0000
1605  2016-01-09 04:34:45 +0000
1618  2016-01-07 05:28:35 +0000
1663  2016-01-01 06:22:03 +0000
1689  2015-12-28 05:07:27 +0000
1774  2015-12-19 01:25:31 +0000
1819  2015-12-15 02:32:17 +0000
1844  2015-12-13 01:25:37 +0000
1895  2015-12-10 00:08:50 +0000
1905  2015-12-09 15:09:55 +0000
1914  2015-12-08 20:53:11 +0000
1940  2015-12-07 04:11:02 +0000
2038  2015-12-01 04:44:10 +0000
2149  2015-11-26 01:11:28 +0000
```

```
2189  2015-11-24 01:42:25 +0000
2298  2015-11-18 20:02:51 +0000

                                                  source  \
30    <a href="http://twitter.com/download/iphone" r...
55    <a href="http://twitter.com/download/iphone" r...
64    <a href="http://twitter.com/download/iphone" r...
113   <a href="http://twitter.com/download/iphone" r...
148   <a href="http://twitter.com/download/iphone" r...
179   <a href="http://twitter.com/download/iphone" r...
185   <a href="http://twitter.com/download/iphone" r...
186   <a href="http://twitter.com/download/iphone" r...
188   <a href="http://twitter.com/download/iphone" r...
189   <a href="http://twitter.com/download/iphone" r...
218   <a href="http://twitter.com/download/iphone" r...
228   <a href="http://twitter.com/download/iphone" r...
234   <a href="http://twitter.com/download/iphone" r...
274   <a href="http://twitter.com/download/iphone" r...
290   <a href="http://twitter.com/download/iphone" r...
291   <a href="http://twitter.com/download/iphone" r...
313   <a href="http://twitter.com/download/iphone" r...
342   <a href="http://twitter.com/download/iphone" r...
346   <a href="http://twitter.com/download/iphone" r...
375   <a href="http://twitter.com" rel="nofollow">Tw...
387   <a href="http://twitter.com/download/iphone" r...
409   <a href="http://twitter.com/download/iphone" r...
427   <a href="http://twitter.com/download/iphone" r...
498   <a href="http://twitter.com/download/iphone" r...
513   <a href="http://twitter.com/download/iphone" r...
570   <a href="http://twitter.com/download/iphone" r...
576   <a href="http://twitter.com/download/iphone" r...
611   <a href="http://twitter.com/download/iphone" r...
701   <a href="http://twitter.com/download/iphone" r...
707   <a href="http://twitter.com/download/iphone" r...
843   <a href="http://twitter.com/download/iphone" r...
857   <a href="http://twitter.com/download/iphone" r...
967   <a href="http://twitter.com/download/iphone" r...
1005  <a href="http://twitter.com/download/iphone" r...
1080  <a href="http://twitter.com/download/iphone" r...
1295  <a href="http://twitter.com/download/iphone" r...
1345  <a href="http://twitter.com/download/iphone" r...
1445  <a href="http://twitter.com/download/iphone" r...
1446  <a href="http://twitter.com/download/iphone" r...
1474  <a href="http://twitter.com/download/iphone" r...
1479  <a href="http://twitter.com/download/iphone" r...
1497  <a href="http://twitter.com/download/iphone" r...
1523  <a href="http://twitter.com/download/iphone" r...
1598  <a href="http://twitter.com/download/iphone" r...
```

```
1605  <a href="http://twitter.com/download/iphone" r...
1618  <a href="http://twitter.com/download/iphone" r...
1663  <a href="http://twitter.com/download/iphone" r...
1689  <a href="http://twitter.com/download/iphone" r...
1774  <a href="http://twitter.com/download/iphone" r...
1819  <a href="http://twitter.com/download/iphone" r...
1844  <a href="http://twitter.com/download/iphone" r...
1895  <a href="http://twitter.com/download/iphone" r...
1905  <a href="http://twitter.com/download/iphone" r...
1914  <a href="http://twitter.com/download/iphone" r...
1940  <a href="http://twitter.com/download/iphone" r...
2038  <a href="http://twitter.com/download/iphone" r...
2149  <a href="http://twitter.com/download/iphone" r...
2189  <a href="http://twitter.com/download/iphone" r...
2298  <a href="http://twitter.com/download/iphone" r...

                                                  text  retweeted_status_id  \
30    @NonWhiteHat @MayhewMayhem omg hello tanner yo...                   NaN
55    @roushfenway These are good dogs but 17/10 is ...                   NaN
64                      @RealKentMurphy 14/10 confirmed                   NaN
113   @ComplicitOwl @ShopWeRateDogs &gt;10/10 is res...                   NaN
148   @Jack_Septic_Eye I'd need a few more pics to p...                   NaN
179                        @Marc_IRL pixelated af 12/10                   NaN
185   RT @Jenna_Marbles: @dog_rates Thanks for ratin...         8.563302e+17
186   @xianmcguire @Jenna_Marbles Kardashians wouldn...                   NaN
188   @dhmontgomery We also gave snoop dogg a 420/10...                   NaN
189   @s8n You tried very hard to portray this good ...                   NaN
218   @markhoppus MARK THAT DOG HAS SEEN AND EXPERIE...                   NaN
228   Jerry just apuppologized to me. He said there ...                   NaN
234   .@breaannanicolee PUPDATE: Cannon has a heart ...                   NaN
274   @0_kelvin_0 &gt;10/10 is reserved for puppos s...                   NaN
290                                 @markhoppus 182/10                   NaN
291   @bragg6of8 @Andy_Pace_ we are still looking fo...                   NaN
313   @jonnysun @Lin_Manuel ok jomny I know you're e...                   NaN
342           @docmisterio account started on 11/15/15                   NaN
346                              @UNC can confirm 12/10                   NaN
375   Beebop and Doobert should start a band 12/10 w...                   NaN
387   I was going to do 007/10, but the joke wasn't ...                   NaN
409                                @HistoryInPics 13/10                   NaN
427   @imgur for a polar bear tho I'd say 13/10 is a...                   NaN
498   I've been informed by multiple sources that th...                   NaN
513   PUPDATE: I've been informed that Augie was act...                   NaN
570   .@NBCSports OMG THE TINY HAT I'M GOING TO HAVE...                   NaN
576   @SkyWilliams doggo simply protecting you from ...                   NaN
611   @JODYHiGHROLLER it may be an 11/10 but what do...                   NaN
701             13/10 for breakdancing puppo @shibbnbot                   NaN
707      Today, 10/10, should be National Dog Rates Day                   NaN
843   His name is Charley and he already has a new s...                   NaN
```

```
857   @TheEllenShow I'm not sure if you know this bu...                 NaN
967                13/10 such a good doggo\n@spaghemily                 NaN
1005  Other pupper asked not to have his identity sh...                 NaN
1080                                 @mount_alex3 13/10                 NaN
1295  @serial @MrRoles OH MY GOD I listened to all o...                 NaN
1345                               13/10 hero af\n@ABC                  NaN
1445    Oh my god 10/10 for every little hot dog pupper                 NaN
1446  After reading the comments I may have overesti...                 NaN
1474  BREAKING PUPDATE: I've just been notified that...                 NaN
1479  Personally I'd give him an 11/10. Not sure why...                 NaN
1497  PUPDATE: just noticed this dog has some extra ...                 NaN
1523                             12/10 @LightningHoltt                 NaN
1598  Yes I do realize a rating of 4/20 would've bee...                 NaN
1605  Jack deserves another round of applause. If yo...                 NaN
1618  For those who claim this is a goat, u are wron...                 NaN
1663  I'm aware that I could've said 20/16, but here...                 NaN
1689  I've been told there's a slight possibility he...                 NaN
1774  After getting lost in Reese's eyes for several...                 NaN
1819  After some outrage from the crowd. Bubbles is ...                 NaN
1844  This dog is being demoted to a 9/10 for not we...                 NaN
1895  Some clarification is required. The dog is sin...                 NaN
1905  The 13/10 also takes into account this impecca...                 NaN
1914                                   13/10\n@ABC7                    NaN
1940  The millennials have spoken and we've decided ...                 NaN
2038  After 22 minutes of careful deliberation this ...                 NaN
2149  After countless hours of research and hundreds...                 NaN
2189                    12/10 good shit Bubka\n@wane15                 NaN
2298  After much debate this dog is being upgraded t...                 NaN

      retweeted_status_user_id retweeted_status_timestamp expanded_urls  \
30                         NaN                         NaN           NaN
55                         NaN                         NaN           NaN
64                         NaN                         NaN           NaN
113                        NaN                         NaN           NaN
148                        NaN                         NaN           NaN
179                        NaN                         NaN           NaN
185                66699013.0  2017-04-24 02:13:14 +0000           NaN
186                        NaN                         NaN           NaN
188                        NaN                         NaN           NaN
189                        NaN                         NaN           NaN
218                        NaN                         NaN           NaN
228                        NaN                         NaN           NaN
234                        NaN                         NaN           NaN
274                        NaN                         NaN           NaN
290                        NaN                         NaN           NaN
291                        NaN                         NaN           NaN
313                        NaN                         NaN           NaN
342                        NaN                         NaN           NaN
```

346	NaN	NaN	NaN
375	NaN	NaN	NaN
387	NaN	NaN	NaN
409	NaN	NaN	NaN
427	NaN	NaN	NaN
498	NaN	NaN	NaN
513	NaN	NaN	NaN
570	NaN	NaN	NaN
576	NaN	NaN	NaN
611	NaN	NaN	NaN
701	NaN	NaN	NaN
707	NaN	NaN	NaN
843	NaN	NaN	NaN
857	NaN	NaN	NaN
967	NaN	NaN	NaN
1005	NaN	NaN	NaN
1080	NaN	NaN	NaN
1295	NaN	NaN	NaN
1345	NaN	NaN	NaN
1445	NaN	NaN	NaN
1446	NaN	NaN	NaN
1474	NaN	NaN	NaN
1479	NaN	NaN	NaN
1497	NaN	NaN	NaN
1523	NaN	NaN	NaN
1598	NaN	NaN	NaN
1605	NaN	NaN	NaN
1618	NaN	NaN	NaN
1663	NaN	NaN	NaN
1689	NaN	NaN	NaN
1774	NaN	NaN	NaN
1819	NaN	NaN	NaN
1844	NaN	NaN	NaN
1895	NaN	NaN	NaN
1905	NaN	NaN	NaN
1914	NaN	NaN	NaN
1940	NaN	NaN	NaN
2038	NaN	NaN	NaN
2149	NaN	NaN	NaN
2189	NaN	NaN	NaN
2298	NaN	NaN	NaN

	rating_numerator	rating_denominator	name	doggo	floofer	pupper	puppo
30	12	10	None	None	None	None	None
55	17	10	None	None	None	None	None
64	14	10	None	None	None	None	None
113	10	10	None	None	None	None	None
148	12	10	None	None	None	None	None

179	12	10	None	None	None	None	None
185	14	10	None	None	None	None	None
186	14	10	None	None	None	None	None
188	420	10	None	None	None	None	None
189	666	10	None	None	None	None	None
218	13	10	None	None	None	None	None
228	11	10	None	None	None	None	None
234	13	10	None	None	None	None	None
274	10	10	None	None	None	None	None
290	182	10	None	None	None	None	None
291	15	10	None	None	None	None	None
313	960	0	None	None	None	None	None
342	11	15	None	None	None	None	None
346	12	10	None	None	None	None	None
375	12	10	None	None	None	None	None
387	7	10	None	None	None	None	None
409	13	10	None	None	None	None	None
427	13	10	None	None	None	None	None
498	12	10	None	None	None	None	None
513	11	10	None	None	None	None	None
570	11	10	None	None	None	None	None
576	11	10	None	doggo	None	None	None
611	11	10	None	None	None	None	None
701	13	10	None	None	None	None	puppo
707	10	10	None	None	None	None	None
843	13	10	None	None	None	None	None
857	12	10	None	doggo	None	None	None
967	13	10	None	doggo	None	None	None
1005	12	10	None	None	None	pupper	None
1080	13	10	None	None	None	None	None
1295	12	10	None	None	None	None	None
1345	13	10	None	None	None	None	None
1445	10	10	None	None	None	pupper	None
1446	1	10	None	None	None	None	None
1474	10	10	None	None	None	None	None
1479	11	10	None	None	None	None	None
1497	9	10	None	None	None	None	None
1523	12	10	None	None	None	None	None
1598	4	20	None	None	None	None	None
1605	14	10	None	None	None	None	None
1618	5	10	None	None	None	None	None
1663	20	16	None	None	None	None	None
1689	5	10	None	None	None	None	None
1774	13	10	None	None	None	None	None
1819	7	10	None	None	None	None	None
1844	9	10	None	None	None	None	None
1895	11	10	None	None	None	None	None
1905	13	10	None	None	None	None	None

```
1914                13                  10  None  None  None  None  None
1940                 1                  10  None  None  None  None  None
2038                 1                  10  None  None  None  None  None
2149                11                  10  None  None  None  None  None
2189                12                  10  None  None  None  None  None
2298                10                  10  None  None  None  None  None
```

In [35]:
```
#checking the rating_numerator column for outliers for values less than 10 or greater t
df_archive[(df_archive.rating_numerator > 20) | (df_archive.rating_numerator < 10)]
```

Out[35]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
45	883482846933004288	NaN	NaN	
188	855862651834028034	8.558616e+17	1.943518e+08	
189	855860136149123072	8.558585e+17	1.361572e+07	
229	848212111729840128	NaN	NaN	
290	838150277551247360	8.381455e+17	2.195506e+07	
313	835246439529840640	8.352460e+17	2.625958e+07	
315	835152434251116546	NaN	NaN	
340	832215909146226688	NaN	NaN	
387	826598799820865537	8.265984e+17	4.196984e+09	
433	820690176645140481	NaN	NaN	
462	817502432452313088	NaN	NaN	
485	814578408554463233	NaN	NaN	
516	810984652412424192	NaN	NaN	
599	798682547630837760	NaN	NaN	
605	798576900688019456	NaN	NaN	
695	786709082849828864	NaN	NaN	
730	781661882474196992	NaN	NaN	
745	780092040432480260	NaN	NaN	
763	778027034220126208	NaN	NaN	
764	777953400541634568	NaN	NaN	
765	777885040357281792	NaN	NaN	
784	775096608509886464	NaN	NaN	
814	771014301343748096	NaN	NaN	
835	768193404517830656	NaN	NaN	
845	766423258543644672	NaN	NaN	
859	763183847194451968	NaN	NaN	
860	763167063695355904	NaN	NaN	
883	760252756032651264	NaN	NaN	
896	759099523532779520	NaN	NaN	
902	758467244762497024	NaN	NaN	
...	...	...	...	
2312	666776908487630848	NaN	NaN	
2314	666701168228331520	NaN	NaN	
2315	666691418707132416	NaN	NaN	
2316	666649482315059201	NaN	NaN	
2317	666644823164719104	NaN	NaN	
2319	666447344410484738	NaN	NaN	

2320	666437273139982337	NaN	NaN
2322	666430724426358785	NaN	NaN
2323	666428276349472768	NaN	NaN
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2338	666104133288665088	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2344	666071193221509120	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2349	666051853826850816	NaN	NaN
2351	666049248165822465	NaN	NaN
2352	666044226329800704	NaN	NaN
2353	666033412701032449	NaN	NaN
2354	666029285002620928	NaN	NaN
2355	666020888022790149	NaN	NaN

	timestamp \
45	2017-07-08 00:28:19 +0000
188	2017-04-22 19:15:32 +0000
189	2017-04-22 19:05:32 +0000
229	2017-04-01 16:35:01 +0000
290	2017-03-04 22:12:52 +0000
313	2017-02-24 21:54:03 +0000
315	2017-02-24 15:40:31 +0000
340	2017-02-16 13:11:49 +0000
387	2017-02-01 01:11:25 +0000
433	2017-01-15 17:52:40 +0000
462	2017-01-06 22:45:43 +0000
485	2016-12-29 21:06:41 +0000
516	2016-12-19 23:06:23 +0000
599	2016-11-16 00:22:12 +0000
605	2016-11-15 17:22:24 +0000
695	2016-10-13 23:23:56 +0000
730	2016-09-30 01:08:10 +0000
745	2016-09-25 17:10:10 +0000
763	2016-09-20 00:24:34 +0000
764	2016-09-19 19:31:59 +0000
765	2016-09-19 15:00:20 +0000
784	2016-09-11 22:20:06 +0000

```
814   2016-08-31 15:58:28 +0000
835   2016-08-23 21:09:14 +0000
845   2016-08-18 23:55:18 +0000
859   2016-08-10 01:23:03 +0000
860   2016-08-10 00:16:21 +0000
883   2016-08-01 23:15:56 +0000
896   2016-07-29 18:53:24 +0000
902   2016-07-28 01:00:57 +0000
...                         ...
2312  2015-11-18 00:36:17 +0000
2314  2015-11-17 19:35:19 +0000
2315  2015-11-17 18:56:35 +0000
2316  2015-11-17 16:09:56 +0000
2317  2015-11-17 15:51:26 +0000
2319  2015-11-17 02:46:43 +0000
2320  2015-11-17 02:06:42 +0000
2322  2015-11-17 01:40:41 +0000
2323  2015-11-17 01:30:57 +0000
2326  2015-11-17 00:24:19 +0000
2327  2015-11-17 00:06:54 +0000
2328  2015-11-16 23:23:41 +0000
2330  2015-11-16 21:10:36 +0000
2331  2015-11-16 20:32:58 +0000
2333  2015-11-16 19:31:45 +0000
2334  2015-11-16 16:37:02 +0000
2335  2015-11-16 16:11:11 +0000
2338  2015-11-16 04:02:55 +0000
2340  2015-11-16 03:44:34 +0000
2341  2015-11-16 03:22:39 +0000
2342  2015-11-16 02:38:37 +0000
2344  2015-11-16 01:52:02 +0000
2346  2015-11-16 01:01:59 +0000
2347  2015-11-16 00:55:59 +0000
2349  2015-11-16 00:35:11 +0000
2351  2015-11-16 00:24:50 +0000
2352  2015-11-16 00:04:52 +0000
2353  2015-11-15 23:21:54 +0000
2354  2015-11-15 23:05:30 +0000
2355  2015-11-15 22:32:08 +0000

                                                  source  \
45    <a href="http://twitter.com/download/iphone" r...
188   <a href="http://twitter.com/download/iphone" r...
189   <a href="http://twitter.com/download/iphone" r...
229   <a href="http://twitter.com/download/iphone" r...
290   <a href="http://twitter.com/download/iphone" r...
313   <a href="http://twitter.com/download/iphone" r...
315   <a href="http://twitter.com/download/iphone" r...
```

```
340   <a href="http://twitter.com/download/iphone" r...
387   <a href="http://twitter.com/download/iphone" r...
433   <a href="http://twitter.com/download/iphone" r...
462   <a href="http://twitter.com/download/iphone" r...
485   <a href="http://twitter.com/download/iphone" r...
516   <a href="http://twitter.com/download/iphone" r...
599   <a href="http://twitter.com/download/iphone" r...
605   <a href="http://twitter.com/download/iphone" r...
695   <a href="http://twitter.com/download/iphone" r...
730   <a href="http://twitter.com/download/iphone" r...
745   <a href="http://twitter.com/download/iphone" r...
763   <a href="http://twitter.com/download/iphone" r...
764   <a href="http://twitter.com/download/iphone" r...
765   <a href="http://twitter.com/download/iphone" r...
784   <a href="http://twitter.com/download/iphone" r...
814   <a href="http://twitter.com/download/iphone" r...
835   <a href="http://twitter.com/download/iphone" r...
845   <a href="http://twitter.com/download/iphone" r...
859   <a href="http://twitter.com/download/iphone" r...
860   <a href="http://twitter.com/download/iphone" r...
883   <a href="http://twitter.com/download/iphone" r...
896   <a href="http://twitter.com/download/iphone" r...
902   <a href="http://twitter.com/download/iphone" r...
...                                                  ...
2312  <a href="http://twitter.com/download/iphone" r...
2314  <a href="http://twitter.com/download/iphone" r...
2315  <a href="http://twitter.com/download/iphone" r...
2316  <a href="http://twitter.com/download/iphone" r...
2317  <a href="http://twitter.com/download/iphone" r...
2319  <a href="http://twitter.com/download/iphone" r...
2320  <a href="http://twitter.com/download/iphone" r...
2322  <a href="http://twitter.com/download/iphone" r...
2323  <a href="http://twitter.com/download/iphone" r...
2326  <a href="http://twitter.com/download/iphone" r...
2327  <a href="http://twitter.com/download/iphone" r...
2328  <a href="http://twitter.com/download/iphone" r...
2330  <a href="http://twitter.com/download/iphone" r...
2331  <a href="http://twitter.com/download/iphone" r...
2333  <a href="http://twitter.com/download/iphone" r...
2334  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
```

```
2349  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...

                                                  text  retweeted_status_id  \
45    This is Bella. She hopes her smile made you sm...                  NaN
188   @dhmontgomery We also gave snoop dogg a 420/10...                  NaN
189   @s8n You tried very hard to portray this good ...                  NaN
229   This is Jerry. He's doing a distinguished tong...                  NaN
290                                 @markhoppus 182/10                   NaN
313   @jonnysun @Lin_Manuel ok jomny I know you're e...                  NaN
315   When you're so blinded by your systematic plag...                  NaN
340   RT @dog_rates: This is Logan, the Chow who liv...        7.867091e+17
387   I was going to do 007/10, but the joke wasn't ...                  NaN
433   The floofs have been released I repeat the flo...                  NaN
462   RT @dog_rates: Meet Herschel. He's slightly bi...        6.924173e+17
485   RT @dog_rates: Meet Beau &amp;amp; Wilbur. Wilbur ...        6.981954e+17
516   Meet Sam. She smiles 24/7 &amp;amp; secretly aspir...              NaN
599   RT @dog_rates: Here we see a rare pouched pupp...        6.769365e+17
605   RT @dog_rates: Not familiar with this breed. N...        6.661041e+17
695   This is Logan, the Chow who lived. He solemnly...                  NaN
730   Who keeps sending in pictures without dogs in ...                  NaN
745   RT @dog_rates: This is Hank. He's mischievous ...        7.533757e+17
763   This is Sophie. She's a Jubilant Bush Pupper. ...                  NaN
764   RT @dog_rates: Meet Gerald. He's a fairly exot...        7.681934e+17
765   This is Wesley. He's clearly trespassing. Seem...                  NaN
784   RT @dog_rates: After so many requests, this is...        7.403732e+17
814   Another pic without a dog in it? What am I sup...                  NaN
835   Meet Gerald. He's a fairly exotic doggo. Floof...                  NaN
845   This is Shadoe. Her tongue flies out of her mo...                  NaN
859   This is Clark. He collects teddy bears. It's a...                  NaN
860   RT @dog_rates: Meet Eve. She's a raging alcoho...        6.732953e+17
883   This is Fido. He can tell the weather. Not goo...                  NaN
896   Meet Toby. He has a drinking problem. Inflatab...                  NaN
902   Why does this never happen at my front door...                    NaN
...                                                 ...                  ...
2312  This is Josep. He is a Rye Manganese mix. Can ...                  NaN
2314  This is a golden Buckminsterfullerene named Jo...                  NaN
2315  This is Christoper. He is a spotted Penne. Can...                  NaN
2316  Cool dog. Enjoys couch. Low monotone bark. Ver...                  NaN
2317  This is Jimothy. He is a Botwanian Gouda. Can ...                  NaN
2319  This is Scout. She is a black Downton Abbey. I...                  NaN
2320  Here we see a lone northeastern Cumberbatch. H...                  NaN
2322  Oh boy what a pup! Sunglasses take this one to...                  NaN
2323  Here we have an Austrian Pulitzer. Collectors ...                  NaN
```

```
2326  This is quite the dog. Gets really excited whe...                NaN
2327  This is a southern Vesuvius bumblegruff. Can d...                NaN
2328  Oh goodness. A super rare northeast Qdoba kang...                NaN
2330  Unique dog here. Very small. Lives in containe...                NaN
2331  Here we have a mixed Asiago from the Galápagos...                NaN
2333  This is an extremely rare horned Parthenon. No...                NaN
2334  This is a funny dog. Weird toes. Won't come do...                NaN
2335  This is an Albanian 3 1/2 legged  Episcopalian...                NaN
2338  Not familiar with this breed. No tail (weird)...                NaN
2340  Can stand on stump for what seems like a while...                NaN
2341  This appears to be a Mongolian Presbyterian mi...                NaN
2342  Here we have a well-established sunblockerspan...                NaN
2344  Here we have a northern speckled Rhododendron...                 NaN
2346  Here is the Rand Paul of retrievers folks! He'...                NaN
2347  My oh my. This is a rare blond Canadian terrie...                NaN
2349  This is an odd dog. Hard on the outside but lo...                NaN
2351  Here we have a 1949 1st generation vulpix. Enj...                NaN
2352  This is a purebred Piers Morgan. Loves to Netf...                NaN
2353  Here is a very happy pup. Big fan of well-main...                NaN
2354  This is a western brown Mitsubishi terrier. Up...                NaN
2355  Here we have a Japanese Irish Setter. Lost eye...                NaN

      retweeted_status_user_id retweeted_status_timestamp  \
45                         NaN                         NaN
188                        NaN                         NaN
189                        NaN                         NaN
229                        NaN                         NaN
290                        NaN                         NaN
313                        NaN                         NaN
315                        NaN                         NaN
340               4.196984e+09  2016-10-13 23:23:56 +0000
387                        NaN                         NaN
433                        NaN                         NaN
462               4.196984e+09  2016-01-27 18:42:06 +0000
485               4.196984e+09  2016-02-12 17:22:12 +0000
516                        NaN                         NaN
599               4.196984e+09  2015-12-16 01:27:03 +0000
605               4.196984e+09  2015-11-16 04:02:55 +0000
695                        NaN                         NaN
730                        NaN                         NaN
745               4.196984e+09  2016-07-13 23:48:51 +0000
763                        NaN                         NaN
764               4.196984e+09  2016-08-23 21:09:14 +0000
765                        NaN                         NaN
784               4.196984e+09  2016-06-08 02:41:38 +0000
814                        NaN                         NaN
835                        NaN                         NaN
845                        NaN                         NaN
```

```
859                      NaN                         NaN
860             4.196984e+09  2015-12-06 00:17:55 +0000
883                      NaN                         NaN
896                      NaN                         NaN
902                      NaN                         NaN
...                      ...                         ...
2312                     NaN                         NaN
2314                     NaN                         NaN
2315                     NaN                         NaN
2316                     NaN                         NaN
2317                     NaN                         NaN
2319                     NaN                         NaN
2320                     NaN                         NaN
2322                     NaN                         NaN
2323                     NaN                         NaN
2326                     NaN                         NaN
2327                     NaN                         NaN
2328                     NaN                         NaN
2330                     NaN                         NaN
2331                     NaN                         NaN
2333                     NaN                         NaN
2334                     NaN                         NaN
2335                     NaN                         NaN
2338                     NaN                         NaN
2340                     NaN                         NaN
2341                     NaN                         NaN
2342                     NaN                         NaN
2344                     NaN                         NaN
2346                     NaN                         NaN
2347                     NaN                         NaN
2349                     NaN                         NaN
2351                     NaN                         NaN
2352                     NaN                         NaN
2353                     NaN                         NaN
2354                     NaN                         NaN
2355                     NaN                         NaN

                                              expanded_urls  rating_numerator  \
45    https://twitter.com/dog_rates/status/883482846...                   5
188                                                  NaN                 420
189                                                  NaN                 666
229   https://twitter.com/dog_rates/status/848212111...                   6
290                                                  NaN                 182
313                                                  NaN                 960
315   https://twitter.com/dog_rates/status/835152434...                   0
340   https://twitter.com/dog_rates/status/786709082...                  75
387                                                  NaN                   7
433   https://twitter.com/dog_rates/status/820690176...                  84
```

462	https://twitter.com/dog_rates/status/692417313...	7
485	https://twitter.com/dog_rates/status/698195409...	9
516	https://www.gofundme.com/sams-smile,https://tw...	24
599	https://twitter.com/dog_rates/status/676936541...	8
605	https://twitter.com/dog_rates/status/666104133...	1
695	https://twitter.com/dog_rates/status/786709082...	75
730	https://twitter.com/dog_rates/status/781661882...	5
745	https://twitter.com/dog_rates/status/753375668...	8
763	https://twitter.com/dog_rates/status/778027034...	27
764	https://twitter.com/dog_rates/status/768193404...	8
765	https://twitter.com/dog_rates/status/777885040...	3
784	https://twitter.com/dog_rates/status/740373189...	9
814	https://twitter.com/dog_rates/status/771014301...	7
835	https://twitter.com/dog_rates/status/768193404...	8
845	https://twitter.com/dog_rates/status/766423258...	9
859	https://twitter.com/dog_rates/status/763183847...	8
860	https://twitter.com/dog_rates/status/673295268...	8
883	https://twitter.com/dog_rates/status/760252756...	4
896	https://twitter.com/dog_rates/status/759099523...	7
902	https://twitter.com/dog_rates/status/758467244...	165
...	...	...
2312	https://twitter.com/dog_rates/status/666776908...	5
2314	https://twitter.com/dog_rates/status/666701168...	8
2315	https://twitter.com/dog_rates/status/666691418...	8
2316	https://twitter.com/dog_rates/status/666649482...	4
2317	https://twitter.com/dog_rates/status/666644823...	9
2319	https://twitter.com/dog_rates/status/666447344...	9
2320	https://twitter.com/dog_rates/status/666437273...	7
2322	https://twitter.com/dog_rates/status/666430724...	6
2323	https://twitter.com/dog_rates/status/666428276...	7
2326	https://twitter.com/dog_rates/status/666411507...	2
2327	https://twitter.com/dog_rates/status/666407126...	7
2328	https://twitter.com/dog_rates/status/666396247...	9
2330	https://twitter.com/dog_rates/status/666362758...	6
2331	https://twitter.com/dog_rates/status/666353288...	8
2333	https://twitter.com/dog_rates/status/666337882...	9
2334	https://twitter.com/dog_rates/status/666293911...	3
2335	https://twitter.com/dog_rates/status/666287406...	1
2338	https://twitter.com/dog_rates/status/666104133...	1
2340	https://twitter.com/dog_rates/status/666099513...	8
2341	https://twitter.com/dog_rates/status/666094000...	9
2342	https://twitter.com/dog_rates/status/666082916...	6
2344	https://twitter.com/dog_rates/status/666071193...	9
2346	https://twitter.com/dog_rates/status/666058600...	8
2347	https://twitter.com/dog_rates/status/666057090...	9
2349	https://twitter.com/dog_rates/status/666051853...	2
2351	https://twitter.com/dog_rates/status/666049248...	5
2352	https://twitter.com/dog_rates/status/666044226...	6

```
2353  https://twitter.com/dog_rates/status/666033412...                9
2354  https://twitter.com/dog_rates/status/666029285...                7
2355  https://twitter.com/dog_rates/status/666020888...                8

      rating_denominator        name  doggo floofer  pupper puppo
45                    10       Bella   None    None    None  None
188                   10        None   None    None    None  None
189                   10        None   None    None    None  None
229                   10       Jerry   None    None    None  None
290                   10        None   None    None    None  None
313                    0        None   None    None    None  None
315                   10        None   None    None    None  None
340                   10       Logan   None    None    None  None
387                   10        None   None    None    None  None
433                   70        None   None    None    None  None
462                   10    Herschel   None    None  pupper  None
485                   10        Beau   None    None    None  None
516                    7         Sam   None    None    None  None
599                   10        None   None    None  pupper  None
605                   10        None   None    None    None  None
695                   10       Logan   None    None    None  None
730                   10        None   None    None    None  None
745                   10        Hank   None    None    None  None
763                   10      Sophie   None    None  pupper  None
764                   10      Gerald  doggo    None    None  None
765                   10      Wesley   None    None    None  None
784                   11        None   None    None    None  None
814                   10        None   None    None    None  None
835                   10      Gerald  doggo    None    None  None
845                   10      Shadoe   None    None    None  None
859                   10       Clark   None    None    None  None
860                   10         Eve   None    None  pupper  None
883                   10        Fido   None    None    None  None
896                   10        Toby   None    None    None  None
902                  150        None   None    None    None  None
...                  ...         ...    ...     ...     ...   ...
2312                  10       Josep   None    None    None  None
2314                  10           a   None    None    None  None
2315                  10  Christoper   None    None    None  None
2316                  10        None   None    None    None  None
2317                  10     Jimothy   None    None    None  None
2319                  10       Scout   None    None    None  None
2320                  10        None   None    None    None  None
2322                  10        None   None    None    None  None
2323                  10        None   None    None    None  None
2326                  10       quite   None    None    None  None
2327                  10           a   None    None    None  None
2328                  10        None   None    None    None  None
```

```
2330                  10        None   None    None    None  None
2331                  10        None   None    None    None  None
2333                  10          an   None    None    None  None
2334                  10           a   None    None    None  None
2335                   2          an   None    None    None  None
2338                  10        None   None    None    None  None
2340                  10        None   None    None    None  None
2341                  10        None   None    None    None  None
2342                  10        None   None    None    None  None
2344                  10        None   None    None    None  None
2346                  10         the   None    None    None  None
2347                  10           a   None    None    None  None
2349                  10          an   None    None    None  None
2351                  10        None   None    None    None  None
2352                  10           a   None    None    None  None
2353                  10           a   None    None    None  None
2354                  10           a   None    None    None  None
2355                  10        None   None    None    None  None

[464 rows x 17 columns]
```

In [36]:
```
#checking the rating_denominator column for outliers with values less than 5 or greater
df_archive[(df_archive.rating_denominator > 15) | (df_archive.rating_denominator < 5)]
```

Out[36]:
```
                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
313   835246439529840640           8.352460e+17         2.625958e+07
433   820690176645140481                    NaN                  NaN
902   758467244762497024                    NaN                  NaN
1120  731156023742988288                    NaN                  NaN
1165  722974582966214656                    NaN                  NaN
1202  716439118184652801                    NaN                  NaN
1228  713900603437621249                    NaN                  NaN
1254  710658690886586372                    NaN                  NaN
1274  709198395643068416                    NaN                  NaN
1351  704054845121142784                    NaN                  NaN
1433  697463031882764288                    NaN                  NaN
1598  686035780142297088           6.860340e+17         4.196984e+09
1634  684225744407494656           6.842229e+17         4.196984e+09
1635  684222868335505415                    NaN                  NaN
1663  682808988178739200           6.827884e+17         4.196984e+09
1779  677716515794329600                    NaN                  NaN
1843  675853064436391936                    NaN                  NaN
2335  666287406224695296                    NaN                  NaN

                     timestamp  \
313   2017-02-24 21:54:03 +0000
433   2017-01-15 17:52:40 +0000
902   2016-07-28 01:00:57 +0000
```

```
1120  2016-05-13 16:15:54 +0000
1165  2016-04-21 02:25:47 +0000
1202  2016-04-03 01:36:11 +0000
1228  2016-03-27 01:29:02 +0000
1254  2016-03-18 02:46:49 +0000
1274  2016-03-14 02:04:08 +0000
1351  2016-02-28 21:25:30 +0000
1433  2016-02-10 16:51:59 +0000
1598  2016-01-10 04:04:10 +0000
1634  2016-01-05 04:11:44 +0000
1635  2016-01-05 04:00:18 +0000
1663  2016-01-01 06:22:03 +0000
1779  2015-12-18 05:06:23 +0000
1843  2015-12-13 01:41:41 +0000
2335  2015-11-16 16:11:11 +0000

                                                 source  \
313   <a href="http://twitter.com/download/iphone" r...
433   <a href="http://twitter.com/download/iphone" r...
902   <a href="http://twitter.com/download/iphone" r...
1120  <a href="http://twitter.com/download/iphone" r...
1165  <a href="http://twitter.com/download/iphone" r...
1202  <a href="http://twitter.com/download/iphone" r...
1228  <a href="http://twitter.com/download/iphone" r...
1254  <a href="http://twitter.com/download/iphone" r...
1274  <a href="http://twitter.com/download/iphone" r...
1351  <a href="http://twitter.com/download/iphone" r...
1433  <a href="http://twitter.com/download/iphone" r...
1598  <a href="http://twitter.com/download/iphone" r...
1634  <a href="http://twitter.com/download/iphone" r...
1635  <a href="http://twitter.com/download/iphone" r...
1663  <a href="http://twitter.com/download/iphone" r...
1779  <a href="http://twitter.com/download/iphone" r...
1843  <a href="http://twitter.com/download/iphone" r...
2335  <a href="http://twitter.com/download/iphone" r...

                                                   text  retweeted_status_id  \
313   @jonnysun @Lin_Manuel ok jomny I know you're e...                  NaN
433   The floofs have been released I repeat the flo...                  NaN
902   Why does this never happen at my front door...                 NaN
1120  Say hello to this unbelievably well behaved sq...                  NaN
1165  Happy 4/20 from the squad! 13/10 for all https...                  NaN
1202  This is Bluebert. He just saw that both #Final...                  NaN
1228  Happy Saturday here's 9 puppers on a bench. 99...                  NaN
1254  Here's a brigade of puppers. All look very pre...                  NaN
1274  From left to right:\nCletus, Jerome, Alejandro...                  NaN
1351  Here is a whole flock of puppers.  60/50 I'll ...                  NaN
1433  Happy Wednesday here's a bucket of pups. 44/40...                  NaN
```

```
1598  Yes I do realize a rating of 4/20 would've bee...                 NaN
1634  Two sneaky puppers were not initially seen, mo...                 NaN
1635  Someone help the girl is being mugged. Several...                 NaN
1663  I'm aware that I could've said 20/16, but here...                 NaN
1779  IT'S PUPPERGEDDON. Total of 144/120 ...I think...                 NaN
1843  Here we have an entire platoon of puppers. Tot...                 NaN
2335  This is an Albanian 3 1/2 legged  Episcopalian...                 NaN

      retweeted_status_user_id retweeted_status_timestamp  \
313                        NaN                         NaN
433                        NaN                         NaN
902                        NaN                         NaN
1120                       NaN                         NaN
1165                       NaN                         NaN
1202                       NaN                         NaN
1228                       NaN                         NaN
1254                       NaN                         NaN
1274                       NaN                         NaN
1351                       NaN                         NaN
1433                       NaN                         NaN
1598                       NaN                         NaN
1634                       NaN                         NaN
1635                       NaN                         NaN
1663                       NaN                         NaN
1779                       NaN                         NaN
1843                       NaN                         NaN
2335                       NaN                         NaN

                                           expanded_urls  rating_numerator  \
313                                                  NaN               960
433   https://twitter.com/dog_rates/status/820690176...                84
902   https://twitter.com/dog_rates/status/758467244...               165
1120  https://twitter.com/dog_rates/status/731156023...               204
1165  https://twitter.com/dog_rates/status/722974582...                 4
1202  https://twitter.com/dog_rates/status/716439118...                50
1228  https://twitter.com/dog_rates/status/713900603...                99
1254  https://twitter.com/dog_rates/status/710658690...                80
1274  https://twitter.com/dog_rates/status/709198395...                45
1351  https://twitter.com/dog_rates/status/704054845...                60
1433  https://twitter.com/dog_rates/status/697463031...                44
1598                                                 NaN                 4
1634  https://twitter.com/dog_rates/status/684225744...               143
1635  https://twitter.com/dog_rates/status/684222868...               121
1663                                                 NaN                20
1779  https://twitter.com/dog_rates/status/677716515...               144
1843  https://twitter.com/dog_rates/status/675853064...                88
2335  https://twitter.com/dog_rates/status/666287406...                 1
```

```
      rating_denominator      name doggo floofer pupper puppo
313                    0      None  None    None   None  None
433                   70      None  None    None   None  None
902                  150      None  None    None   None  None
1120                 170      this  None    None   None  None
1165                  20      None  None    None   None  None
1202                  50  Bluebert  None    None   None  None
1228                  90      None  None    None   None  None
1254                  80      None  None    None   None  None
1274                  50      None  None    None   None  None
1351                  50         a  None    None   None  None
1433                  40      None  None    None   None  None
1598                  20      None  None    None   None  None
1634                 130      None  None    None   None  None
1635                 110      None  None    None   None  None
1663                  16      None  None    None   None  None
1779                 120      None  None    None   None  None
1843                  80      None  None    None   None  None
2335                   2        an  None    None   None  None
```

In [37]:
```
#identify retweets from df_archive
df_archive[df_archive['retweeted_status_id'].notnull()]
```

Out[37]:
```
                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	quite	None	None	None	None
124	10	Cooper	None	None	None	None
130	10	None	None	None	None	None
132	10	Jamesy	None	None	pupper	None
137	10	None	None	None	pupper	None
146	10	Quinn	None	None	None	None
155	10	None	None	None	None	None
159	10	Lorenzo	None	None	None	None
160	10	None	None	None	None	None
165	10	None	None	None	None	None
171	10	Winston	None	None	None	None
180	10	None	None	None	None	None
182	10	Luna	None	None	None	None
185	10	None	None	None	None	None
194	10	George	None	None	None	None
195	10	None	None	None	None	None
204	10	None	None	None	None	None
211	10	Astrid	doggo	None	None	None

```
...                  ...       ...    ...     ...      ...   ...
784                   11      None   None    None     None  None
794                   10      Fizz   None    None     None  None
800                   10    Gromit   None    None     None  None
811                   10   Frankie   None    None     None  None
815                   10      None   None    None   pupper  None
818                   10      None  doggo    None     None  None
822                   10      just  doggo    None   pupper  None
826                   10      None   None    None     None  None
829                   10      None   None    None   pupper  None
833                   10    Nollie   None    None     None  None
841                   10      None   None    None     None  None
847                   10     Colby   None    None     None  None
860                   10       Eve   None    None   pupper  None
868                   10      None   None    None     None  None
872                   10      None   None    None     None  None
885                   10      None   None    None     None  None
890                   10      None   None    None     None  None
895                   10      None   None    None     None  None
908                   10  Chompsky   None    None     None  None
911                   10      None   None    None   pupper  None
926                   10     Rubio   None    None     None  None
937                   10     Carly   None    None     None  None
943                   10      None   None    None     None  None
949                   10      None   None    None     None  None
1012                  10      None   None    None   pupper  None
1023                  10    Shaggy   None    None     None  None
1043                  10      None   None    None     None  None
1242                  10      None   None    None     None  None
2259                  10      None   None    None     None  None
2260                  10      None   None    None     None  None

[181 rows x 17 columns]
```

In [38]:
```
#identify reply tweets from df_archive
df_archive[df_archive['in_reply_to_status_id'].notnull()]
```

Out[38]:
```
                tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
30    886267009285017600           8.862664e+17         2.281182e+09
55    881633300179243008           8.816070e+17         4.738443e+07
64    879674319642796034           8.795538e+17         3.105441e+09
113   870726314365509632           8.707262e+17         1.648776e+07
148   863427515083354112           8.634256e+17         7.759620e+07
149   863079547188785154           6.671522e+17         4.196984e+09
179   857214891891077121           8.571567e+17         1.806710e+08
184   856526610513747968           8.558181e+17         4.196984e+09
186   856288084350160898           8.562860e+17         2.792810e+08
188   855862651834028034           8.558616e+17         1.943518e+08
```

189	855860136149123072	8.558585e+17	1.361572e+07
218	850333567704068097	8.503288e+17	2.195506e+07
228	848213670039564288	8.482121e+17	4.196984e+09
234	847617282490613760	8.476062e+17	4.196984e+09
251	844979544864018432	7.590995e+17	4.196984e+09
274	840698636975636481	8.406983e+17	8.405479e+17
290	838150277551247360	8.381455e+17	2.195506e+07
291	838085839343206401	8.380855e+17	2.894131e+09
313	835246439529840640	8.352460e+17	2.625958e+07
342	832088576586297345	8.320875e+17	3.058208e+07
346	831926988323639298	8.319030e+17	2.068372e+07
387	826598799820865537	8.265984e+17	4.196984e+09
409	823333489516937216	8.233264e+17	1.582854e+09
427	821153421864615936	8.211526e+17	1.132119e+08
498	813130366689148928	8.131273e+17	4.196984e+09
513	811647686436880384	8.116272e+17	4.196984e+09
565	802265048156610565	7.331095e+17	4.196984e+09
570	801854953262350336	8.018543e+17	1.185634e+07
576	800859414831898624	8.008580e+17	2.918590e+08
611	797165961484890113	7.971238e+17	2.916630e+07
...	...	...	...
1479	693582294167244802	6.935722e+17	1.198989e+09
1497	692423280028966913	6.924173e+17	4.196984e+09
1501	692142790915014657	6.920419e+17	4.196984e+09
1523	690607260360429569	6.903413e+17	4.670367e+08
1598	686035780142297088	6.860340e+17	4.196984e+09
1605	685681090388975616	6.855479e+17	4.196984e+09
1618	684969860808454144	6.849598e+17	4.196984e+09
1630	684538444857667585	6.844811e+17	4.196984e+09
1634	684225744407494656	6.842229e+17	4.196984e+09
1663	682808988178739200	6.827884e+17	4.196984e+09
1689	681340665377193984	6.813394e+17	4.196984e+09
1774	678023323247357953	6.780211e+17	4.196984e+09
1819	676590572941893632	6.765883e+17	4.196984e+09
1842	675870721063669760	6.757073e+17	4.196984e+09
1844	675849018447167488	6.758457e+17	4.196984e+09
1852	675707330206547968	6.754971e+17	4.196984e+09
1866	675349384339542016	6.749998e+17	4.196984e+09
1882	674999807681908736	6.747934e+17	4.196984e+09
1885	674793399141146624	6.717299e+17	4.196984e+09
1892	674754018082705410	6.747522e+17	4.196984e+09
1895	674742531037511680	6.747400e+17	4.196984e+09
1905	674606911342424069	6.744689e+17	4.196984e+09
1914	674330906434379776	6.658147e+17	1.637468e+07
1940	673716320723169284	6.737159e+17	4.196984e+09
2036	671729906628341761	6.715610e+17	4.196984e+09
2038	671550332464455680	6.715449e+17	4.196984e+09
2149	669684865554620416	6.693544e+17	4.196984e+09

```
2169  669353438988365824           6.678065e+17         4.196984e+09
2189  668967877119254528           6.689207e+17         2.143566e+07
2298  667070482143944705           6.670655e+17         4.196984e+09

                      timestamp  \
30    2017-07-15 16:51:35 +0000
55    2017-07-02 21:58:53 +0000
64    2017-06-27 12:14:36 +0000
113   2017-06-02 19:38:25 +0000
148   2017-05-13 16:15:35 +0000
149   2017-05-12 17:12:53 +0000
179   2017-04-26 12:48:51 +0000
184   2017-04-24 15:13:52 +0000
186   2017-04-23 23:26:03 +0000
188   2017-04-22 19:15:32 +0000
189   2017-04-22 19:05:32 +0000
218   2017-04-07 13:04:55 +0000
228   2017-04-01 16:41:12 +0000
234   2017-03-31 01:11:22 +0000
251   2017-03-23 18:29:57 +0000
274   2017-03-11 22:59:09 +0000
290   2017-03-04 22:12:52 +0000
291   2017-03-04 17:56:49 +0000
313   2017-02-24 21:54:03 +0000
342   2017-02-16 04:45:50 +0000
346   2017-02-15 18:03:45 +0000
387   2017-02-01 01:11:25 +0000
409   2017-01-23 00:56:15 +0000
427   2017-01-17 00:33:26 +0000
498   2016-12-25 21:12:41 +0000
513   2016-12-21 19:01:02 +0000
565   2016-11-25 21:37:47 +0000
570   2016-11-24 18:28:13 +0000
576   2016-11-22 00:32:18 +0000
611   2016-11-11 19:55:50 +0000
...                        ...
1479  2016-01-30 23:51:19 +0000
1497  2016-01-27 19:05:49 +0000
1501  2016-01-27 00:31:15 +0000
1523  2016-01-22 18:49:36 +0000
1598  2016-01-10 04:04:10 +0000
1605  2016-01-09 04:34:45 +0000
1618  2016-01-07 05:28:35 +0000
1630  2016-01-06 00:54:18 +0000
1634  2016-01-05 04:11:44 +0000
1663  2016-01-01 06:22:03 +0000
1689  2015-12-28 05:07:27 +0000
1774  2015-12-19 01:25:31 +0000
```

```
1819  2015-12-15 02:32:17 +0000
1842  2015-12-13 02:51:51 +0000
1844  2015-12-13 01:25:37 +0000
1852  2015-12-12 16:02:36 +0000
1866  2015-12-11 16:20:15 +0000
1882  2015-12-10 17:11:09 +0000
1885  2015-12-10 03:30:58 +0000
1892  2015-12-10 00:54:28 +0000
1895  2015-12-10 00:08:50 +0000
1905  2015-12-09 15:09:55 +0000
1914  2015-12-08 20:53:11 +0000
1940  2015-12-07 04:11:02 +0000
2036  2015-12-01 16:37:44 +0000
2038  2015-12-01 04:44:10 +0000
2149  2015-11-26 01:11:28 +0000
2169  2015-11-25 03:14:30 +0000
2189  2015-11-24 01:42:25 +0000
2298  2015-11-18 20:02:51 +0000

                                                  source  \
30    <a href="http://twitter.com/download/iphone" r...
55    <a href="http://twitter.com/download/iphone" r...
64    <a href="http://twitter.com/download/iphone" r...
113   <a href="http://twitter.com/download/iphone" r...
148   <a href="http://twitter.com/download/iphone" r...
149   <a href="http://twitter.com/download/iphone" r...
179   <a href="http://twitter.com/download/iphone" r...
184   <a href="http://twitter.com/download/iphone" r...
186   <a href="http://twitter.com/download/iphone" r...
188   <a href="http://twitter.com/download/iphone" r...
189   <a href="http://twitter.com/download/iphone" r...
218   <a href="http://twitter.com/download/iphone" r...
228   <a href="http://twitter.com/download/iphone" r...
234   <a href="http://twitter.com/download/iphone" r...
251   <a href="http://twitter.com/download/iphone" r...
274   <a href="http://twitter.com/download/iphone" r...
290   <a href="http://twitter.com/download/iphone" r...
291   <a href="http://twitter.com/download/iphone" r...
313   <a href="http://twitter.com/download/iphone" r...
342   <a href="http://twitter.com/download/iphone" r...
346   <a href="http://twitter.com/download/iphone" r...
387   <a href="http://twitter.com/download/iphone" r...
409   <a href="http://twitter.com/download/iphone" r...
427   <a href="http://twitter.com/download/iphone" r...
498   <a href="http://twitter.com/download/iphone" r...
513   <a href="http://twitter.com/download/iphone" r...
565   <a href="http://twitter.com/download/iphone" r...
570   <a href="http://twitter.com/download/iphone" r...
```

```
576   <a href="http://twitter.com/download/iphone" r...
611   <a href="http://twitter.com/download/iphone" r...
...                                                 ...
1479  <a href="http://twitter.com/download/iphone" r...
1497  <a href="http://twitter.com/download/iphone" r...
1501  <a href="http://twitter.com/download/iphone" r...
1523  <a href="http://twitter.com/download/iphone" r...
1598  <a href="http://twitter.com/download/iphone" r...
1605  <a href="http://twitter.com/download/iphone" r...
1618  <a href="http://twitter.com/download/iphone" r...
1630  <a href="http://twitter.com/download/iphone" r...
1634  <a href="http://twitter.com/download/iphone" r...
1663  <a href="http://twitter.com/download/iphone" r...
1689  <a href="http://twitter.com/download/iphone" r...
1774  <a href="http://twitter.com/download/iphone" r...
1819  <a href="http://twitter.com/download/iphone" r...
1842  <a href="http://twitter.com/download/iphone" r...
1844  <a href="http://twitter.com/download/iphone" r...
1852  <a href="http://twitter.com/download/iphone" r...
1866  <a href="http://twitter.com/download/iphone" r...
1882  <a href="http://twitter.com/download/iphone" r...
1885  <a href="http://twitter.com/download/iphone" r...
1892  <a href="http://twitter.com/download/iphone" r...
1895  <a href="http://twitter.com/download/iphone" r...
1905  <a href="http://twitter.com/download/iphone" r...
1914  <a href="http://twitter.com/download/iphone" r...
1940  <a href="http://twitter.com/download/iphone" r...
2036  <a href="http://twitter.com/download/iphone" r...
2038  <a href="http://twitter.com/download/iphone" r...
2149  <a href="http://twitter.com/download/iphone" r...
2169  <a href="http://twitter.com/download/iphone" r...
2189  <a href="http://twitter.com/download/iphone" r...
2298  <a href="http://twitter.com/download/iphone" r...

                                                  text  retweeted_status_id  \
30    @NonWhiteHat @MayhewMayhem omg hello tanner yo...                   NaN
55    @roushfenway These are good dogs but 17/10 is ...                   NaN
64                        @RealKentMurphy 14/10 confirmed                 NaN
113   @ComplicitOwl @ShopWeRateDogs &gt;10/10 is res...                   NaN
148   @Jack_Septic_Eye I'd need a few more pics to p...                   NaN
149   Ladies and gentlemen... I found Pipsy. He may ...                   NaN
179                          @Marc_IRL pixelated af 12/10                 NaN
184   THIS IS CHARLIE, MARK. HE DID JUST WANT TO SAY...                   NaN
186   @xianmcguire @Jenna_Marbles Kardashians wouldn...                   NaN
188   @dhmontgomery We also gave snoop dogg a 420/10...                   NaN
189   @s8n You tried very hard to portray this good ...                   NaN
218   @markhoppus MARK THAT DOG HAS SEEN AND EXPERIE...                   NaN
228   Jerry just apuppologized to me. He said there ...                   NaN
```

234	.@breaannanicolee PUPDATE: Cannon has a heart ...	NaN
251	PUPDATE: I'm proud to announce that Toby is 23...	NaN
274	@0_kelvin_0 >10/10 is reserved for puppos s...	NaN
290	@markhoppus 182/10	NaN
291	@bragg6of8 @Andy_Pace_ we are still looking fo...	NaN
313	@jonnysun @Lin_Manuel ok jomny I know you're e...	NaN
342	@docmisterio account started on 11/15/15	NaN
346	@UNC can confirm 12/10	NaN
387	I was going to do 007/10, but the joke wasn't ...	NaN
409	@HistoryInPics 13/10	NaN
427	@imgur for a polar bear tho I'd say 13/10 is a...	NaN
498	I've been informed by multiple sources that th...	NaN
513	PUPDATE: I've been informed that Augie was act...	NaN
565	Like doggo, like pupper version 2. Both 11/10 ...	NaN
570	.@NBCSports OMG THE TINY HAT I'M GOING TO HAVE...	NaN
576	@SkyWilliams doggo simply protecting you from ...	NaN
611	@JODYHiGHROLLER it may be an 11/10 but what do...	NaN
...	...	...
1479	Personally I'd give him an 11/10. Not sure why...	NaN
1497	PUPDATE: just noticed this dog has some extra ...	NaN
1501	These are some pictures of Teddy that further ...	NaN
1523	12/10 @LightningHoltt	NaN
1598	Yes I do realize a rating of 4/20 would've bee...	NaN
1605	Jack deserves another round of applause. If yo...	NaN
1618	For those who claim this is a goat, u are wron...	NaN
1630	After watching this video, we've determined th...	NaN
1634	Two sneaky puppers were not initially seen, mo...	NaN
1663	I'm aware that I could've said 20/16, but here...	NaN
1689	I've been told there's a slight possibility he...	NaN
1774	After getting lost in Reese's eyes for several...	NaN
1819	After some outrage from the crowd. Bubbles is ...	NaN
1842	& this is Yoshi. Another world record cont...	NaN
1844	This dog is being demoted to a 9/10 for not we...	NaN
1852	We've got ourselves a battle here. Watch out R...	NaN
1866	Yea I lied. Here's more. All 13/10 https://t.c...	NaN
1882	Ok last one of these. I may try to make some m...	NaN
1885	I have found another. 13/10 https://t.co/HwroP...	NaN
1892	Just received another perfect photo of dogs an...	NaN
1895	Some clarification is required. The dog is sin...	NaN
1905	The 13/10 also takes into account this impecca...	NaN
1914	13/10\n@ABC7	NaN
1940	The millennials have spoken and we've decided ...	NaN
2036	I'm just going to leave this one here as well...	NaN
2038	After 22 minutes of careful deliberation this ...	NaN
2149	After countless hours of research and hundreds...	NaN
2169	This is Tessa. She is also very pleased after ...	NaN
2189	12/10 good shit Bubka\n@wane15	NaN
2298	After much debate this dog is being upgraded t...	NaN

```
      retweeted_status_user_id retweeted_status_timestamp  \
30                         NaN                         NaN
55                         NaN                         NaN
64                         NaN                         NaN
113                        NaN                         NaN
148                        NaN                         NaN
149                        NaN                         NaN
179                        NaN                         NaN
184                        NaN                         NaN
186                        NaN                         NaN
188                        NaN                         NaN
189                        NaN                         NaN
218                        NaN                         NaN
228                        NaN                         NaN
234                        NaN                         NaN
251                        NaN                         NaN
274                        NaN                         NaN
290                        NaN                         NaN
291                        NaN                         NaN
313                        NaN                         NaN
342                        NaN                         NaN
346                        NaN                         NaN
387                        NaN                         NaN
409                        NaN                         NaN
427                        NaN                         NaN
498                        NaN                         NaN
513                        NaN                         NaN
565                        NaN                         NaN
570                        NaN                         NaN
576                        NaN                         NaN
611                        NaN                         NaN
...                        ...                         ...
1479                       NaN                         NaN
1497                       NaN                         NaN
1501                       NaN                         NaN
1523                       NaN                         NaN
1598                       NaN                         NaN
1605                       NaN                         NaN
1618                       NaN                         NaN
1630                       NaN                         NaN
1634                       NaN                         NaN
1663                       NaN                         NaN
1689                       NaN                         NaN
1774                       NaN                         NaN
1819                       NaN                         NaN
1842                       NaN                         NaN
1844                       NaN                         NaN
```

```
1852                      NaN                      NaN
1866                      NaN                      NaN
1882                      NaN                      NaN
1885                      NaN                      NaN
1892                      NaN                      NaN
1895                      NaN                      NaN
1905                      NaN                      NaN
1914                      NaN                      NaN
1940                      NaN                      NaN
2036                      NaN                      NaN
2038                      NaN                      NaN
2149                      NaN                      NaN
2169                      NaN                      NaN
2189                      NaN                      NaN
2298                      NaN                      NaN

                                          expanded_urls  rating_numerator  \
30                                                  NaN                12
55                                                  NaN                17
64                                                  NaN                14
113                                                 NaN                10
148                                                 NaN                12
149   https://twitter.com/dog_rates/status/863079547...                14
179                                                 NaN                12
184   https://twitter.com/dog_rates/status/856526610...                14
186                                                 NaN                14
188                                                 NaN               420
189                                                 NaN               666
218                                                 NaN                13
228                                                 NaN                11
234                                                 NaN                13
251   https://twitter.com/dog_rates/status/844979544...                13
274                                                 NaN                10
290                                                 NaN               182
291                                                 NaN                15
313                                                 NaN               960
342                                                 NaN                11
346                                                 NaN                12
387                                                 NaN                 7
409                                                 NaN                13
427                                                 NaN                13
498                                                 NaN                12
513                                                 NaN                11
565   https://twitter.com/dog_rates/status/802265048...                11
570                                                 NaN                11
576                                                 NaN                11
611                                                 NaN                11
...                                                 ...               ...
```

```
1479                                                  NaN                11
1497                                                  NaN                 9
1501  https://twitter.com/dog_rates/status/692142790...                 13
1523                                                  NaN                12
1598                                                  NaN                 4
1605                                                  NaN                14
1618                                                  NaN                 5
1630  https://twitter.com/dog_rates/status/684538444...                 12
1634  https://twitter.com/dog_rates/status/684225744...                143
1663                                                  NaN                20
1689                                                  NaN                 5
1774                                                  NaN                13
1819                                                  NaN                 7
1842  https://twitter.com/dog_rates/status/675870721...                 11
1844                                                  NaN                 9
1852  https://twitter.com/dog_rates/status/675707330...                 11
1866  https://twitter.com/dog_rates/status/675349384...                 13
1882  https://twitter.com/dog_rates/status/674999807...                 13
1885  https://twitter.com/dog_rates/status/674793399...                 13
1892  https://twitter.com/dog_rates/status/674754018...                 12
1895                                                  NaN                11
1905                                                  NaN                13
1914                                                  NaN                13
1940                                                  NaN                 1
2036  https://twitter.com/dog_rates/status/671729906...                 13
2038                                                  NaN                 1
2149                                                  NaN                11
2169  https://twitter.com/dog_rates/status/669353438...                 10
2189                                                  NaN                12
2298                                                  NaN                10

      rating_denominator   name  doggo floofer  pupper puppo
30                    10   None   None    None    None  None
55                    10   None   None    None    None  None
64                    10   None   None    None    None  None
113                   10   None   None    None    None  None
148                   10   None   None    None    None  None
149                   10   None   None    None    None  None
179                   10   None   None    None    None  None
184                   10   None   None    None    None  None
186                   10   None   None    None    None  None
188                   10   None   None    None    None  None
189                   10   None   None    None    None  None
218                   10   None   None    None    None  None
228                   10   None   None    None    None  None
234                   10   None   None    None    None  None
251                   10   None   None    None    None  None
274                   10   None   None    None    None  None
```

290	10	None	None	None	None	None
291	10	None	None	None	None	None
313	0	None	None	None	None	None
342	15	None	None	None	None	None
346	10	None	None	None	None	None
387	10	None	None	None	None	None
409	10	None	None	None	None	None
427	10	None	None	None	None	None
498	10	None	None	None	None	None
513	10	None	None	None	None	None
565	10	None	doggo	None	pupper	None
570	10	None	None	None	None	None
576	10	None	doggo	None	None	None
611	10	None	None	None	None	None
...	...	...	...	...	...	...
1479	10	None	None	None	None	None
1497	10	None	None	None	None	None
1501	10	None	None	None	None	None
1523	10	None	None	None	None	None
1598	20	None	None	None	None	None
1605	10	None	None	None	None	None
1618	10	None	None	None	None	None
1630	10	None	None	None	None	None
1634	130	None	None	None	None	None
1663	16	None	None	None	None	None
1689	10	None	None	None	None	None
1774	10	None	None	None	None	None
1819	10	None	None	None	None	None
1842	10	None	None	None	None	None
1844	10	None	None	None	None	None
1852	10	None	None	None	None	None
1866	10	None	None	None	None	None
1882	10	None	None	None	None	None
1885	10	None	None	None	None	None
1892	10	None	None	None	None	None
1895	10	None	None	None	None	None
1905	10	None	None	None	None	None
1914	10	None	None	None	None	None
1940	10	None	None	None	None	None
2036	10	None	None	None	None	None
2038	10	None	None	None	None	None
2149	10	None	None	None	None	None
2169	10	Tessa	None	None	None	None
2189	10	None	None	None	None	None
2298	10	None	None	None	None	None

[78 rows x 17 columns]

```
In [39]: #check the values in the source column
         df_archive.source.value_counts()

Out[39]: <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         <a href="http://vine.co" rel="nofollow">Vine - Make a Scene</a>
         <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
         <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
         Name: source, dtype: int64

In [40]: df_archive.info()

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

1.3.1 Quality

1. The following columns have incorrect data types: df_archive table: tweet_id, in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id. These have numeric values(int/float) instead of string type. The 'timestamp' and 'retweeted_status_timestamp' columns have datatype 'string' instead of 'datetime'. df_predictions table: incorrect data types: tweet_id (int instead of string) df_tweetdata table: incorrect data types: tweet_id (int instead of string)
2. The date and time information is contained in a single column.
3. The 'name' column in df_archive table has several 'None' or incorrect values.
4. The 'expanded_urls' column in df_archive table has several 'None' values (no images present).

5. The 'rating_denominator' column in df_archive table has several incorrect values, including 0.
6. The df_archive and df_tweetdata tables contain retweets and replies.
7. The 'source' column in df_archive table has information that is difficult to read and comprehend.
8. The 'text' column in df_archive is truncated, so the full information in this column cannot be read.
9. The name of the 'id' column in the df_tweetdata table is inconsistent with the column name 'tweet_id' in the other two dataframes.

1.3.2 Tidiness

1. There are 4 different columns for dog stage instead of a single column.
2. The columns not relevant for our analysis may be removed.
3. The three datasets can be merged into a single dataset according to the rules of tidy data.

1.4 Cleaning Data

```
In [41]: # Make copies of the original datasets
         df_archive_clean = df_archive.copy()
         df_predictions_clean = df_predictions.copy()
         df_tweetdata_clean = df_tweetdata.copy()
```

We now start cleaning the data to address the quality and tidiness issues identified above.

Define Rename the 'id' column in the df_tweetdata_clean to 'tweet_id' for consistency with the other two dataframes

Code

```
In [42]: # rename column 'id' to 'tweet_id' in df_tweetdata_clean
         df_tweetdata_clean.rename(columns={'id':'tweet_id'}, inplace=True)
```

Test

```
In [43]: df_tweetdata_clean.head()
```

```
Out[43]:              tweet_id  retweet_count  retweeted retweeted_status  \
         0  892420643555336193           8291      False              NaN
         1  892177421306343426           6123      False              NaN
         2  891815181378084864           4054      False              NaN
         3  891689557279858688           8428      False              NaN
         4  891327558926688256           9130      False              NaN

            favorite_count
```

```
0           37950
1           32588
2           24536
3           41291
4           39478
```

Define Merge the three datasets into a single dataset according to the rules of tidy data.

Code

```
In [44]: df_final = pd.merge(df_archive_clean, df_predictions_clean, on = 'tweet_id', how = 'inn
         df_final = pd.merge(df_final, df_tweetdata_clean, on = 'tweet_id', how = 'inner')
```

Test

```
In [45]: df_final.head()
```

```
Out[45]:              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      ...        \
0  https://twitter.com/dog_rates/status/892420643...      ...
1  https://twitter.com/dog_rates/status/892177421...      ...
2  https://twitter.com/dog_rates/status/891815181...      ...
3  https://twitter.com/dog_rates/status/891689557...      ...
4  https://twitter.com/dog_rates/status/891327558...      ...

                    p2   p2_conf p2_dog                         p3   p3_conf  \
0               bagel  0.085851  False                     banana  0.076110
1            Pekinese  0.090647   True                   papillon  0.068957
2            malamute  0.078253   True                     kelpie  0.031379
3  Labrador_retriever  0.168086   True                    spatula  0.040836
4    English_springer  0.225770   True  German_short-haired_pointer  0.175219

  p3_dog retweet_count retweeted  retweeted_status favorite_count
0  False          8291     False               NaN          37950
1   True          6123     False               NaN          32588
2   True          4054     False               NaN          24536
3  False          8428     False               NaN          41291
4   True          9130     False               NaN          39478

[5 rows x 32 columns]
```

```
In [46]: df_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2065 entries, 0 to 2064
Data columns (total 32 columns):
tweet_id                      2065 non-null int64
in_reply_to_status_id         23 non-null float64
in_reply_to_user_id           23 non-null float64
timestamp                     2065 non-null object
source                        2065 non-null object
text                          2065 non-null object
retweeted_status_id           75 non-null float64
retweeted_status_user_id      75 non-null float64
retweeted_status_timestamp    75 non-null object
expanded_urls                 2065 non-null object
rating_numerator              2065 non-null int64
rating_denominator            2065 non-null int64
name                          2065 non-null object
doggo                         2065 non-null object
floofer                       2065 non-null object
pupper                        2065 non-null object
```

```
puppo                         2065 non-null object
jpg_url                       2065 non-null object
img_num                       2065 non-null int64
p1                            2065 non-null object
p1_conf                       2065 non-null float64
p1_dog                        2065 non-null bool
p2                            2065 non-null object
p2_conf                       2065 non-null float64
p2_dog                        2065 non-null bool
p3                            2065 non-null object
p3_conf                       2065 non-null float64
p3_dog                        2065 non-null bool
retweet_count                 2065 non-null int64
retweeted                     2065 non-null bool
retweeted_status              75 non-null object
favorite_count                2065 non-null int64
dtypes: bool(4), float64(7), int64(6), object(15)
memory usage: 475.9+ KB
```

Define Merge the 4 different columns for dog stage to a single column.

Code

```
In [47]: # merge the columns 'doggo', 'floofer', 'pupper', 'puppo' into a single column called '
         df_final['stage'] = df_final[df_final.columns[13:17]].apply(
             lambda x: ','.join(x.astype(str)),
             axis=1
         )

In [48]: df_final['stage'].value_counts()

Out[48]: None,None,None,None        1744
         None,None,pupper,None       211
         doggo,None,None,None         67
         None,None,None,puppo         23
         doggo,None,pupper,None       11
         None,floofer,None,None        7
         doggo,None,None,puppo         1
         doggo,floofer,None,None       1
         Name: stage, dtype: int64

In [49]: #replace the column values as seen from the above value_counts results by removing mult
         df_final['stage']
         replace_dict = dict([
                 ('None,None,None,None', 'None'), ('None,None,pupper,None', 'pupper'), ('doggo,N
             ('None,None,None,puppo', 'puppo'), ('doggo,None,pupper,None', 'doggo,pupper'), ('No
             ('doggo,None,None,puppo', 'doggo,puppo'), ('doggo,floofer,None,None', 'floofer')
```

```
    ])

df_final=df_final.replace(replace_dict, regex=True)

# delete the separate dog stage columns from the dataset
columns_to_drop = ['doggo', 'floofer', 'pupper', 'puppo']
df_final = df_final.drop(columns_to_drop, axis=1)

#set the datatype of the 'stage' column as category
df_final['stage'] = df_final.stage.astype('category')
```

Test

```
In [50]: df_final.sample(5)
```

Out[50]:

	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
722	759197388317847553	NaN	NaN	2016-07-30 01:22:17 +0000	<a href="http://twitter.com/download/iphone" r...	This is Luna. She's just heckin precious af I ...	NaN	NaN	NaN
1028	712065007010385924	NaN	NaN	2016-03-21 23:55:01 +0000	<a href="http://twitter.com/download/iphone" r...	This is Chester. He's clearly in charge of the...	NaN	NaN	NaN
330	822462944365645825	NaN	NaN	2017-01-20 15:17:01 +0000	<a href="http://twitter.com/download/iphone" r...	This is Gabe. He was the unequivocal embodimen...	NaN	NaN	NaN
1480	678991772295516161	NaN	NaN	2015-12-21 17:33:48 +0000	<a href="http://twitter.com/download/iphone" r...	If your Monday isn't going so well just take a...	NaN	NaN	NaN
1628	674410619106390016	NaN	NaN	2015-12-09 02:09:56 +0000	<a href="http://twitter.com/download/iphone" r...	This is Lenny. He wants to be a sprinkler. 10/...	NaN	NaN	NaN

```
1628                        NaN                         NaN

                                              expanded_urls   ...     p2_conf  \
722   https://twitter.com/dog_rates/status/759197388...   ...    0.076899
1028  https://twitter.com/dog_rates/status/712065007...   ...    0.084253
330   https://twitter.com/dog_rates/status/822462944...   ...    0.023056
1480  https://twitter.com/dog_rates/status/678991772...   ...    0.187003
1628  https://twitter.com/dog_rates/status/674410619...   ...    0.046475

      p2_dog           p3   p3_conf  p3_dog retweet_count  retweeted  \
722     True   white_wolf  0.063269   False          2097      False
1028   False        pizza  0.080168   False           662      False
330     True  Maltese_dog  0.008945    True         15884      False
1480    True    Chihuahua  0.101420    True          1251      False
1628   False       beagle  0.019427    True           482      False

      retweeted_status favorite_count   stage
722                NaN           6383    None
1028               NaN           2353    None
330                NaN          30137  pupper
1480               NaN           2420    None
1628               NaN           1225    None

[5 rows x 29 columns]
```

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

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2065 entries, 0 to 2064
Data columns (total 29 columns):
tweet_id                      2065 non-null int64
in_reply_to_status_id         23 non-null float64
in_reply_to_user_id           23 non-null float64
timestamp                     2065 non-null object
source                        2065 non-null object
text                          2065 non-null object
retweeted_status_id           75 non-null float64
retweeted_status_user_id      75 non-null float64
retweeted_status_timestamp    75 non-null object
expanded_urls                 2065 non-null object
rating_numerator              2065 non-null int64
rating_denominator            2065 non-null int64
name                          2065 non-null object
jpg_url                       2065 non-null object
img_num                       2065 non-null int64
p1                            2065 non-null object
p1_conf                       2065 non-null float64
p1_dog                        2065 non-null bool
```

```
p2                          2065 non-null object
p2_conf                     2065 non-null float64
p2_dog                      2065 non-null bool
p3                          2065 non-null object
p3_conf                     2065 non-null float64
p3_dog                      2065 non-null bool
retweet_count               2065 non-null int64
retweeted                   2065 non-null bool
retweeted_status            75 non-null object
favorite_count              2065 non-null int64
stage                       2065 non-null category
dtypes: bool(4), category(1), float64(7), int64(6), object(11)
memory usage: 413.8+ KB
```

Define Convert the incorrect datatypes into strings for the following columns

Code

```
In [52]: df_final['tweet_id']=df_final['tweet_id'].astype(str)
         df_final['in_reply_to_status_id']=df_final['in_reply_to_status_id'].astype(str)
         df_final['in_reply_to_user_id']=df_final['in_reply_to_user_id'].astype(str)
         df_final['retweeted_status_id']=df_final['retweeted_status_id'].astype(str)
         df_final['retweeted_status_user_id']=df_final['retweeted_status_user_id'].astype(str)
```

Test

```
In [53]: df_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2065 entries, 0 to 2064
Data columns (total 29 columns):
tweet_id                    2065 non-null object
in_reply_to_status_id       2065 non-null object
in_reply_to_user_id         2065 non-null object
timestamp                   2065 non-null object
source                      2065 non-null object
text                        2065 non-null object
retweeted_status_id         2065 non-null object
retweeted_status_user_id    2065 non-null object
retweeted_status_timestamp  75 non-null object
expanded_urls               2065 non-null object
rating_numerator            2065 non-null int64
rating_denominator          2065 non-null int64
name                        2065 non-null object
jpg_url                     2065 non-null object
img_num                     2065 non-null int64
p1                          2065 non-null object
```

```
p1_conf                     2065 non-null float64
p1_dog                      2065 non-null bool
p2                          2065 non-null object
p2_conf                     2065 non-null float64
p2_dog                      2065 non-null bool
p3                          2065 non-null object
p3_conf                     2065 non-null float64
p3_dog                      2065 non-null bool
retweet_count               2065 non-null int64
retweeted                   2065 non-null bool
retweeted_status            75 non-null object
favorite_count              2065 non-null int64
stage                       2065 non-null category
dtypes: bool(4), category(1), float64(3), int64(5), object(16)
memory usage: 413.8+ KB
```

Define Convert the 'timestamp' and 'retweeted_status_timestamp' columns into datetime datatypes, and split into two separate columns for date and time.

Code

```
In [54]: from datetime import datetime

         #convert datatypes to datetime
         date_time = pd.to_datetime(df_final.timestamp, errors='coerce')
         retweet_date_time = pd.to_datetime(df_final.retweeted_status_timestamp, errors='coerce'

         #split each column into separate date and time columns
         df_final = df_final.assign(tweet_date=date_time.dt.date, tweet_time=date_time.dt.time)
         df_final = df_final.assign(retweeted_status_date=retweet_date_time.dt.date, retweeted_s

         #delete the original columns
         datecolumns_to_drop=['timestamp','retweeted_status_timestamp']
         df_final = df_final.drop(datecolumns_to_drop, axis=1)
```

Test

```
In [56]: df_final.sample(5)
```

```
Out[56]:                tweet_id in_reply_to_status_id in_reply_to_user_id  \
         573   783347506784731136                   nan                  nan
         162   852912242202992640                   nan                  nan
         1210  696754882863349760                   nan                  nan
         224   839239871831150596                   nan                  nan
         1119  704761120771465216                   nan                  nan

                                                source  \
```

```
573   <a href="http://twitter.com/download/iphone" r...
162   <a href="http://twitter.com/download/iphone" r...
1210  <a href="http://twitter.com/download/iphone" r...
224   <a href="http://twitter.com/download/iphone" r...
1119  <a href="http://twitter.com/download/iphone" r...

                                                  text  \
573   RT @dog_rates: This is Kenny. He just wants to...
162   Meet Benny. He likes being adorable and making...
1210  This is Reptar. He specifically asked for his ...
224   This is Odie. He's big. 13/10 would attempt to...
1119  This pupper killed this great white in an epic...

       retweeted_status_id retweeted_status_user_id  \
573   6.742918370630533e+17             4196983835.0
162                     nan                      nan
1210                    nan                      nan
224                     nan                      nan
1119                    nan                      nan

                                       expanded_urls  rating_numerator  \
573   https://twitter.com/dog_rates/status/674291837...                11
162   https://www.gofundme.com/bennys-medical-bills,...                12
1210  https://twitter.com/dog_rates/status/696754882...                10
224   https://twitter.com/dog_rates/status/839239871...                13
1119  https://twitter.com/dog_rates/status/704761120...                13

      rating_denominator          ...          p3_dog retweet_count  \
573                   10          ...            True          6102
162                   10          ...            True          1908
1210                  10          ...            True           377
224                   10          ...            True          6969
1119                  10          ...            True          3074

      retweeted                                       retweeted_status  \
573       False  {'created_at': 'Tue Dec 08 18:17:56 +0000 2015...
162       False                                                   NaN
1210      False                                                   NaN
224       False                                                   NaN
1119      False                                                   NaN

      favorite_count   stage  tweet_date  tweet_time  retweeted_status_date  \
573                0    None  2016-10-04    16:46:14             2015-12-08
162             9271    None  2017-04-14    15:51:39                    NaT
1210            1543    None  2016-02-08    17:58:03                    NaT
224            28400    None  2017-03-07    22:22:32                    NaT
1119            6913  pupper  2016-03-01    20:11:59                    NaT
```

```
     retweeted_status_time
573               18:17:56
162                    NaT
1210                   NaT
224                    NaT
1119                   NaT

[5 rows x 31 columns]
```

Define Remove the retweets from the dataset

Code

```
In [57]: df_final = df_final[df_final['retweeted_status_id']=='nan']
```

Test

```
In [58]: df_final['retweeted_status_id'].value_counts()

Out[58]: nan    1990
         Name: retweeted_status_id, dtype: int64
```

Define Remove the reply tweets from the dataset

Code

```
In [59]: df_final = df_final[df_final['in_reply_to_status_id']=='nan']
```

Test

```
In [60]: df_final['in_reply_to_status_id'].value_counts()

Out[60]: nan    1967
         Name: in_reply_to_status_id, dtype: int64

In [473]: df_final.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1967 entries, 0 to 2064
Data columns (total 31 columns):
tweet_id                    1967 non-null object
in_reply_to_status_id       1967 non-null object
in_reply_to_user_id         1967 non-null object
source                      1967 non-null object
text                        1967 non-null object
retweeted_status_id         1967 non-null object
retweeted_status_user_id    1967 non-null object
expanded_urls               1967 non-null object
rating_numerator            1967 non-null int64
```

```
rating_denominator          1967 non-null int64
name                        1967 non-null object
jpg_url                     1967 non-null object
img_num                     1967 non-null int64
p1                          1967 non-null object
p1_conf                     1967 non-null float64
p1_dog                      1967 non-null bool
p2                          1967 non-null object
p2_conf                     1967 non-null float64
p2_dog                      1967 non-null bool
p3                          1967 non-null object
p3_conf                     1967 non-null float64
p3_dog                      1967 non-null bool
retweet_count               1967 non-null int64
retweeted                   1967 non-null bool
retweeted_status            0 non-null object
favorite_count              1967 non-null int64
stage                       1967 non-null object
tweet_date                  1967 non-null object
tweet_time                  1967 non-null object
retweeted_status_date       0 non-null object
retweeted_status_time       0 non-null object
dtypes: bool(4), float64(3), int64(5), object(19)
memory usage: 438.0+ KB
```

Define Check and remove the following columns as they do not contain any relevant information for our analysis: in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id,retweeted_status, retweeted_status_date,retweeted, retweeted_status_time

Code

```
In [61]: df_final['in_reply_to_status_id'].value_counts()

Out[61]: nan    1967
         Name: in_reply_to_status_id, dtype: int64

In [62]: df_final['in_reply_to_user_id'].value_counts()

Out[62]: nan    1967
         Name: in_reply_to_user_id, dtype: int64

In [484]: df_final['retweeted'].value_counts()

Out[484]: False    1967
          Name: retweeted, dtype: int64

In [477]: df_final['retweeted_status_id'].value_counts()
```

```
Out[477]: nan    1967
          Name: retweeted_status_id, dtype: int64

In [478]: df_final['retweeted_status_user_id'].value_counts()

Out[478]: nan    1967
          Name: retweeted_status_user_id, dtype: int64

In [479]: df_final['retweeted_status'].value_counts()

Out[479]: Series([], Name: retweeted_status, dtype: int64)

In [480]: df_final['retweeted_status_date'].value_counts()

Out[480]: Series([], Name: retweeted_status_date, dtype: int64)

In [481]: df_final['retweeted_status_time'].value_counts()

Out[481]: Series([], Name: retweeted_status_time, dtype: int64)

In [63]: columns_to_be_dropped = ['in_reply_to_status_id','in_reply_to_user_id','retweeted','ret
         df_final = df_final.drop(columns_to_be_dropped, axis=1)
```

Test

```
In [64]: df_final.sample(5)

Out[64]:                tweet_id                                                source  \
         182   848324959059550208  <a href="http://twitter.com/download/iphone" r...
         1719  672272411274932228  <a href="http://twitter.com/download/iphone" r...
         1982  667491009379606528  <a href="http://twitter.com" rel="nofollow">Tw...
         1840  670319130621435904  <a href="http://twitter.com/download/iphone" r...
         90    871762521631449091  <a href="http://twitter.com/download/iphone" r...

                                                          text  \
         182   Meet Odin. He's supposed to be giving directio...
         1719  This is Pippin. He is terrified of his new lit...
         1982  Two dogs in this one. Both are rare Jujitsu Py...
         1840  AT DAWN...\nWE RIDE\n\n11/10 https://t.co/QnfO...
         90    This is Rover. As part of pupper protocol he h...

                                                 expanded_urls  rating_numerator  \
         182   https://twitter.com/dog_rates/status/848324959...                12
         1719  https://twitter.com/dog_rates/status/672272411...                11
         1982  https://twitter.com/dog_rates/status/667491009...                 7
         1840  https://twitter.com/dog_rates/status/670319130...                11
         90    https://twitter.com/dog_rates/status/871762521...                12

               rating_denominator    name  \
         182                   10    Odin
```

```
1719                  10  Pippin
1982                  10    None
1840                  10    None
90                    10   Rover

                                                  jpg_url                  p1  \
182   https://pbs.twimg.com/media/C8XbDR1WAAAxND8.jpg            malamute
1719  https://pbs.twimg.com/media/CVRkLuJWUAAhhYp.jpg                 pug
1982  https://pbs.twimg.com/media/CUNniSlUYAEj1Jl.jpg              borzoi
1840  https://pbs.twimg.com/media/CU1zsMSUAAASOqW.jpg       Irish_terrier
90    https://pbs.twimg.com/media/DBkfY58XcAEdzZy.jpg  Labrador_retriever

       p1_conf    ...       p2_conf p2_dog                        p3  \
182   0.544576    ...      0.290268   True                  Eskimo_dog
1719  0.914685    ...      0.014982   True                 Siamese_cat
1982  0.852088    ...      0.132264  False                      weasel
1840  0.254856    ...      0.227716   True  soft-coated_wheaten_terrier
90    0.921393    ...      0.064608   True                   bloodhound

       p3_conf p3_dog  retweet_count  favorite_count   stage  tweet_date  \
182   0.154421   True           4824           25537    None  2017-04-02
1719  0.009221  False           3478            6538    None  2015-12-03
1982  0.005730  False            231             533    None  2015-11-19
1840  0.223263   True           1247            3905    None  2015-11-27
90    0.003383   True           3438           19985  pupper  2017-06-05

     tweet_time
182    00:03:26
1719   04:33:27
1982   23:53:52
1840   19:11:49
90     16:15:56

[5 rows x 22 columns]
```

Define Convert the source column into readable text

```
In [65]: df_final['source'].value_counts()
```

```
Out[65]: <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
         <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
         <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
         Name: source, dtype: int64
```

Code

```
In [66]: replace_source_dict = dict([
                ('<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhon
```

```
    ])
df_final=df_final.replace(replace_source_dict)
```

Test

In [67]:
```
df_final.sample(5)
```

Out[67]:

```
                tweet_id              source  \
48    881906580714921986  Twitter for iPhone
361   817827839487737858  Twitter for iPhone
1448  680221482581123072  Twitter for iPhone
1284  690005060500217858  Twitter for iPhone
2045  666273097616637952  Twitter for iPhone

                                                   text  \
48    This is Rey. He's a Benebop Cumberfloof. 12/10...
361   This is Buddy. He ran into a glass door once. ...
1448  This is CeCe. She's patiently waiting for Sant...
1284  "I'm the only one that ever does anything in t...
2045     Can take selfies 11/10 https://t.co/ws2AMaNwPW

                                          expanded_urls  rating_numerator  \
48    https://twitter.com/dog_rates/status/881906580...                12
361   https://twitter.com/dog_rates/status/817827839...                13
1448  https://twitter.com/dog_rates/status/680221482...                10
1284  https://twitter.com/dog_rates/status/690005060...                10
2045  https://twitter.com/dog_rates/status/666273097...                11

      rating_denominator   name  \
48                    10    Rey
361                   10  Buddy
1448                  10   CeCe
1284                  10   None
2045                  10   None

                                                 jpg_url                p1  \
48      https://pbs.twimg.com/media/DDOpWm9XcAAeSBL.jpg         Weimaraner
361   https://pbs.twimg.com/ext_tw_video_thumb/81782...     cocker_spaniel
1448    https://pbs.twimg.com/media/CXChOQZW8AALdXm.jpg             bubble
1284    https://pbs.twimg.com/media/CZNj8N-WQAMXASZ.jpg            Samoyed
2045    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg  Italian_greyhound

       p1_conf    ...       p2_conf p2_dog        p3   p3_conf p3_dog  \
48    0.291539    ...      0.278966   True     koala  0.127017  False
361   0.387608    ...      0.264844   True  Pekinese  0.122123   True
1448  0.240173    ...      0.146222  False     abaya  0.139342  False
1284  0.270287    ...      0.114027   True     teddy  0.072475  False
2045  0.176053    ...      0.111884   True   basenji  0.111152   True
```

```
      retweet_count  favorite_count  stage  tweet_date tweet_time
48             3315           23822   None  2017-07-03   16:04:48
361           29567           55035   None  2017-01-07   20:18:46
1448            306            1519   None  2015-12-25   03:00:14
1284           1794            3838   None  2016-01-21   02:56:40
2045             76             171   None  2015-11-16   15:14:19

[5 rows x 22 columns]
```

Define Rearrange the columns according to our requirement

Code

```
In [68]: cols = list(df_final.columns.values)
         cols

Out[68]: ['tweet_id',
          'source',
          'text',
          'expanded_urls',
          'rating_numerator',
          'rating_denominator',
          'name',
          'jpg_url',
          'p1',
          'p1_conf',
          'p1_dog',
          'p2',
          'p2_conf',
          'p2_dog',
          'p3',
          'p3_conf',
          'p3_dog',
          'retweet_count',
          'favorite_count',
          'stage',
          'tweet_date',
          'tweet_time']

In [69]: df_final = df_final[['tweet_id','tweet_date','tweet_time','name','stage','retweet_count
                              'rating_denominator','source','expanded_urls','jpg_url',\
                              'p1','p1_conf','p1_dog','p2','p2_conf','p2_dog','p3','p3_conf','p3
```

Test

```
In [70]: df_final.sample(5)
```

```
Out[70]:                tweet_id  tweet_date tweet_time     name stage  retweet_count  \
1761  671511350426865664  2015-12-01   02:09:16  Hammond  None            741
1667  673662677122719744  2015-12-07   00:37:52  Kendall  None            370
46    882268110199369728  2017-07-04   16:01:23     Alfy  None          11327
1309  688547210804498433  2016-01-17   02:23:42    Frönq  None            744
1825  670435821946826752  2015-11-28   02:55:30     None  None            529

      favorite_count  rating_numerator  rating_denominator  \
1761            1629                 8                  10
1667            1274                12                  10
46             44025                13                  10
1309            2737                 9                  10
1825            1066                10                  10

                  source                          ...                          \
1761  Twitter for iPhone                          ...
1667  Twitter for iPhone                          ...
46    Twitter for iPhone                          ...
1309  Twitter for iPhone                          ...
1825  Twitter for iPhone                          ...

                     p1   p1_conf p1_dog                  p2   p2_conf  \
1761         hermit_crab  0.625409  False               tick  0.127333
1667  Labrador_retriever  0.957670   True              beagle  0.012413
46      golden_retriever  0.762211   True  Labrador_retriever  0.098985
1309            papillon  0.531279   True    Blenheim_spaniel  0.214197
1825              sorrel  0.460370  False             basenji  0.135767

     p2_dog                p3   p3_conf p3_dog  \
1761  False             snail  0.097916  False
1667   True  golden_retriever  0.005689   True
46     True    cocker_spaniel  0.017199   True
1309   True     Border_collie  0.053840   True
1825   True          Cardigan  0.099174   True

                                                 text
1761  Say hello to Hammond. He's just a wee lil pup...
1667  This is Kendall. 12/10 would cuddle the hell o...
46    This is Alfy. You're witnessing his first wate...
1309  This is Frönq. He got caught stealing a waffle...
1825  Two unbelievably athletic dogs here. Great for...

[5 rows x 22 columns]
```

Define Remove invalid names and replace by 'Not known'. We check the names that start with a lower case letter, these names can be assumed to be invalid and are replaced by 'Not known'. We also replace the names 'A' and 'None', by 'Not known'.

Code

In [71]:
```
# identify the names starting with lowercase letter and convert them to a list.
invalid_names = list(df_final[df_final['name'].str[0].str.islower()]['name'])
invalid_names
```

Out[71]:
```
['such',
 'a',
 'quite',
 'quite',
 'one',
 'incredibly',
 'a',
 'a',
 'very',
 'my',
 'one',
 'not',
 'his',
 'one',
 'a',
 'a',
 'a',
 'an',
 'very',
 'a',
 'just',
 'getting',
 'very',
 'this',
 'unacceptable',
 'all',
 'a',
 'a',
 'infuriating',
 'a',
 'a',
 'a',
 'an',
 'a',
 'a',
 'very',
 'getting',
 'just',
 'a',
 'the',
 'the',
 'actually',
 'by',
```

```
'a',
'officially',
'a',
'the',
'the',
'a',
'a',
'a',
'a',
'a',
'one',
'a',
'a',
'a',
'light',
'just',
'space',
'a',
'the',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'an',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'a',
'quite',
'a',
'an',
'a',
'an',
'the',
'the',
```

```
          'a',
          'a',
          'an',
          'a',
          'a',
          'a',
          'a']
```

```
In [72]: df_final['name'] = df_final['name'].replace(invalid_names, 'Not known')
         df_final['name'] = df_final['name'].replace('A', 'Not known')
         df_final['name'] = df_final['name'].replace('None', 'Not known')
```

Test

```
In [73]: df_final['name'].value_counts()
```

```
Out[73]: Not known    621
         Cooper        10
         Oliver        10
         Charlie       10
         Lucy          10
         Penny          9
         Tucker         9
         Winston        8
         Sadie          8
         Lola           7
         Toby           7
         Daisy          7
         Koda           6
         Bo             6
         Stanley        6
         Bella          6
         Jax            6
         Milo           5
         Buddy          5
         Chester        5
         Bailey         5
         Leo            5
         Louis          5
         Scout          5
         Oscar          5
         Rusty          5
         Dave           5
         Brody          4
         Finn           4
         Oakley         4
                      ...
         Bell           1
```

```
Sora          1
Murphy        1
Moofasa       1
Amélie        1
Fiji          1
Samsom        1
Clybe         1
Perry         1
Flash         1
Travis        1
Ron           1
Darby         1
Dale          1
Mack          1
Marvin        1
Sage          1
Emanuel       1
Crumpet       1
Nida          1
Joey          1
Rumpole       1
Traviss       1
Duddles       1
Ralphie       1
Snicku        1
Finnegus      1
Severus       1
Jiminus       1
Blakely       1
Name: name, Length: 912, dtype: int64
```

Define Format the 'text' column in the dataframe so that the information in this column can be read in full without being truncated.

Code

```
In [74]: #check the display width of the column and increase it accordingly so that the full tea
         pd.options.display.max_colwidth

Out[74]: 50

In [75]: pd.options.display.max_colwidth = 200
```

Test

```
In [76]: df_final

Out[76]:                 tweet_id  tweet_date tweet_time       name   stage  \
         0     892420643555336193  2017-08-01   16:23:56    Phineas    None
```

1	892177421306343426	2017-08-01	00:17:27	Tilly	None
2	891815181378084864	2017-07-31	00:18:03	Archie	None
3	891689557279858688	2017-07-30	15:58:51	Darla	None
4	891327558926688256	2017-07-29	16:00:24	Franklin	None
5	891087950875897856	2017-07-29	00:08:17	Not known	None
6	890971913173991426	2017-07-28	16:27:12	Jax	None
7	890729181411237888	2017-07-28	00:22:40	Not known	None
8	890609185150312448	2017-07-27	16:25:51	Zoey	None
9	890240255349198849	2017-07-26	15:59:51	Cassie	doggo
10	890006608113172480	2017-07-26	00:31:25	Koda	None
11	889880896479866881	2017-07-25	16:11:53	Bruno	None
12	889665388333682689	2017-07-25	01:55:32	Not known	puppo
13	889638837579907072	2017-07-25	00:10:02	Ted	None
14	889531135344209921	2017-07-24	17:02:04	Stuart	puppo
15	889278841981685760	2017-07-24	00:19:32	Oliver	None
16	888917238123831296	2017-07-23	00:22:39	Jim	None
17	888804989199671297	2017-07-22	16:56:37	Zeke	None
18	888554962724278272	2017-07-22	00:23:06	Ralphus	None
19	888078434458587136	2017-07-20	16:49:33	Gerald	None
20	887705289381826560	2017-07-19	16:06:48	Jeffrey	None
21	887517139158093824	2017-07-19	03:39:09	Not known	None
22	887473957103951883	2017-07-19	00:47:34	Canela	None
23	887343217045368832	2017-07-18	16:08:03	Not known	None
24	887101392804085760	2017-07-18	00:07:08	Not known	None
25	886983233522544640	2017-07-17	16:17:36	Maya	None
26	886736880519319552	2017-07-16	23:58:41	Mingus	None
27	886680336477933568	2017-07-16	20:14:00	Derek	None
28	886366144734445568	2017-07-15	23:25:31	Roscoe	pupper
29	886258384151887873	2017-07-15	16:17:19	Waffles	None
...	...	...	...	...	...
2035	666411507551481857	2015-11-17	00:24:19	Not known	None
2036	666407126856765440	2015-11-17	00:06:54	Not known	None
2037	666396247373291520	2015-11-16	23:23:41	Not known	None
2038	666373753744588802	2015-11-16	21:54:18	Not known	None
2039	666362758909284353	2015-11-16	21:10:36	Not known	None
2040	666353288456101888	2015-11-16	20:32:58	Not known	None
2041	666345417576210432	2015-11-16	20:01:42	Not known	None
2042	666337882303524864	2015-11-16	19:31:45	Not known	None
2043	666293911632134144	2015-11-16	16:37:02	Not known	None
2044	666287406224695296	2015-11-16	16:11:11	Not known	None
2045	666273097616637952	2015-11-16	15:14:19	Not known	None
2046	666268910803644416	2015-11-16	14:57:41	Not known	None
2047	666104133288665088	2015-11-16	04:02:55	Not known	None
2048	666102155909144576	2015-11-16	03:55:04	Not known	None
2049	666099513787052032	2015-11-16	03:44:34	Not known	None
2050	666094000022159362	2015-11-16	03:22:39	Not known	None
2051	666082916733198337	2015-11-16	02:38:37	Not known	None
2052	666073100786774016	2015-11-16	01:59:36	Not known	None

```
2053  666071193221509120  2015-11-16   01:52:02  Not known    None
2054  666063827256086533  2015-11-16   01:22:45  Not known    None
2055  666058600524156928  2015-11-16   01:01:59  Not known    None
2056  666057090499244032  2015-11-16   00:55:59  Not known    None
2057  666055525042405380  2015-11-16   00:49:46  Not known    None
2058  666051853826850816  2015-11-16   00:35:11  Not known    None
2059  666050758794694657  2015-11-16   00:30:50  Not known    None
2060  666049248165822465  2015-11-16   00:24:50  Not known    None
2061  666044226329800704  2015-11-16   00:04:52  Not known    None
2062  666033412701032449  2015-11-15   23:21:54  Not known    None
2063  666029285002620928  2015-11-15   23:05:30  Not known    None
2064  666020888022790149  2015-11-15   22:32:08  Not known    None

      retweet_count  favorite_count  rating_numerator  rating_denominator  \
0              8291           37950                13                  10
1              6123           32588                13                  10
2              4054           24536                12                  10
3              8428           41291                13                  10
4              9130           39478                12                  10
5              3039           19839                13                  10
6              2010           11599                13                  10
7             18389           64022                13                  10
8              4176           27250                13                  10
9              7192           31269                14                  10
10             7147           30042                13                  10
11             4853           27249                13                  10
12             9787           47146                13                  10
13             4413           26608                12                  10
14             2193           14810                13                  10
15             5234           24738                13                  10
16             4393           28532                12                  10
17             4178           25053                13                  10
18             3456           19420                13                  10
19             3407           21313                12                  10
20             5258           29565                13                  10
21            11399           45347                14                  10
22            17726           67599                13                  10
23            10157           33001                13                  10
24             5814           29961                12                  10
25             7582           34432                13                  10
26             3197           11817                13                  10
27             4354           21978                13                  10
28             3126           20766                12                  10
29             6135           27428                13                  10
...             ...             ...               ...                 ...
2035            322             435                 2                  10
2036             40             105                 7                  10
2037             83             162                 9                  10
```

2038	89	182	11	10
2039	561	760	6	10
2040	71	215	8	10
2041	132	286	10	10
2042	90	191	9	10
2043	344	489	3	10
2044	64	144	1	2
2045	76	171	11	10
2046	35	100	10	10
2047	6398	14052	1	10
2048	11	78	11	10
2049	66	151	8	10
2050	72	162	9	10
2051	44	114	6	10
2052	159	315	10	10
2053	58	143	9	10
2054	213	466	10	10
2055	57	109	8	10
2056	138	290	9	10
2057	235	429	10	10
2058	831	1200	2	10
2059	59	130	10	10
2060	43	106	5	10
2061	138	293	6	10
2062	43	123	9	10
2063	47	125	7	10
2064	499	2532	8	10

	source	\
0	Twitter for iPhone	
1	Twitter for iPhone	
2	Twitter for iPhone	
3	Twitter for iPhone	
4	Twitter for iPhone	
5	Twitter for iPhone	
6	Twitter for iPhone	
7	Twitter for iPhone	
8	Twitter for iPhone	
9	Twitter for iPhone	
10	Twitter for iPhone	
11	Twitter for iPhone	
12	Twitter for iPhone	
13	Twitter for iPhone	
14	Twitter for iPhone	
15	Twitter for iPhone	
16	Twitter for iPhone	
17	Twitter for iPhone	
18	Twitter for iPhone	

```
19    Twitter for iPhone
20    Twitter for iPhone
21    Twitter for iPhone
22    Twitter for iPhone
23    Twitter for iPhone
24    Twitter for iPhone
25    Twitter for iPhone
26    Twitter for iPhone
27    Twitter for iPhone
28    Twitter for iPhone
29    Twitter for iPhone
...                   ...
2035  Twitter for iPhone
2036  Twitter for iPhone
2037  Twitter for iPhone
2038  Twitter for iPhone
2039  Twitter for iPhone
2040  Twitter for iPhone
2041  Twitter for iPhone
2042  Twitter for iPhone
2043  Twitter for iPhone
2044  Twitter for iPhone
2045  Twitter for iPhone
2046  Twitter for iPhone
2047  Twitter for iPhone
2048  Twitter for iPhone
2049  Twitter for iPhone
2050  Twitter for iPhone
2051  Twitter for iPhone
2052  Twitter for iPhone
2053  Twitter for iPhone
2054  Twitter for iPhone
2055  Twitter for iPhone
2056  Twitter for iPhone
2057  Twitter for iPhone
2058  Twitter for iPhone
2059  Twitter for iPhone
2060  Twitter for iPhone
2061  Twitter for iPhone
2062  Twitter for iPhone
2063  Twitter for iPhone
2064  Twitter for iPhone


0
1
2
3
```

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
...
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055

```
2056
2057
2058
2059
2060
2061
2062
2063
2064
```

	p1	p1_conf	p1_dog	p2	\
0	orange	0.097049	False	bagel	
1	Chihuahua	0.323581	True	Pekinese	
2	Chihuahua	0.716012	True	malamute	
3	paper_towel	0.170278	False	Labrador_retriever	
4	basset	0.555712	True	English_springer	
5	Chesapeake_Bay_retriever	0.425595	True	Irish_terrier	
6	Appenzeller	0.341703	True	Border_collie	
7	Pomeranian	0.566142	True	Eskimo_dog	
8	Irish_terrier	0.487574	True	Irish_setter	
9	Pembroke	0.511319	True	Cardigan	
10	Samoyed	0.957979	True	Pomeranian	
11	French_bulldog	0.377417	True	Labrador_retriever	
12	Pembroke	0.966327	True	Cardigan	
13	French_bulldog	0.991650	True	boxer	
14	golden_retriever	0.953442	True	Labrador_retriever	
15	whippet	0.626152	True	borzoi	
16	golden_retriever	0.714719	True	Tibetan_mastiff	
17	golden_retriever	0.469760	True	Labrador_retriever	
18	Siberian_husky	0.700377	True	Eskimo_dog	
19	French_bulldog	0.995026	True	pug	
20	basset	0.821664	True	redbone	
21	limousine	0.130432	False	tow_truck	
22	Pembroke	0.809197	True	Rhodesian_ridgeback	
23	Mexican_hairless	0.330741	True	sea_lion	
24	Samoyed	0.733942	True	Eskimo_dog	
25	Chihuahua	0.793469	True	toy_terrier	
26	kuvasz	0.309706	True	Great_Pyrenees	
27	convertible	0.738995	False	sports_car	
28	French_bulldog	0.999201	True	Chihuahua	
29	pug	0.943575	True	shower_cap	
...	...	...	...	...	
2035	coho	0.404640	False	barracouta	
2036	black-and-tan_coonhound	0.529139	True	bloodhound	
2037	Chihuahua	0.978108	True	toy_terrier	
2038	soft-coated_wheaten_terrier	0.326467	True	Afghan_hound	
2039	guinea_pig	0.996496	False	skunk	
2040	malamute	0.336874	True	Siberian_husky	

2041	golden_retriever	0.858744	True	Chesapeake_Bay_retriever
2042	ox	0.416669	False	Newfoundland
2043	three-toed_sloth	0.914671	False	otter
2044	Maltese_dog	0.857531	True	toy_poodle
2045	Italian_greyhound	0.176053	True	toy_terrier
2046	desktop_computer	0.086502	False	desk
2047	hen	0.965932	False	cock
2048	English_setter	0.298617	True	Newfoundland
2049	Lhasa	0.582330	True	Shih-Tzu
2050	bloodhound	0.195217	True	German_shepherd
2051	pug	0.489814	True	bull_mastiff
2052	Walker_hound	0.260857	True	English_foxhound
2053	Gordon_setter	0.503672	True	Yorkshire_terrier
2054	golden_retriever	0.775930	True	Tibetan_mastiff
2055	miniature_poodle	0.201493	True	komondor
2056	shopping_cart	0.962465	False	shopping_basket
2057	chow	0.692517	True	Tibetan_mastiff
2058	box_turtle	0.933012	False	mud_turtle
2059	Bernese_mountain_dog	0.651137	True	English_springer
2060	miniature_pinscher	0.560311	True	Rottweiler
2061	Rhodesian_ridgeback	0.408143	True	redbone
2062	German_shepherd	0.596461	True	malinois
2063	redbone	0.506826	True	miniature_pinscher
2064	Welsh_springer_spaniel	0.465074	True	collie

	p2_conf	p2_dog	p3	p3_conf	p3_dog	\
0	0.085851	False	banana	0.076110	False	
1	0.090647	True	papillon	0.068957	True	
2	0.078253	True	kelpie	0.031379	True	
3	0.168086	True	spatula	0.040836	False	
4	0.225770	True	German_short-haired_pointer	0.175219	True	
5	0.116317	True	Indian_elephant	0.076902	False	
6	0.199287	True	ice_lolly	0.193548	False	
7	0.178406	True	Pembroke	0.076507	True	
8	0.193054	True	Chesapeake_Bay_retriever	0.118184	True	
9	0.451038	True	Chihuahua	0.029248	True	
10	0.013884	True	chow	0.008167	True	
11	0.151317	True	muzzle	0.082981	False	
12	0.027356	True	basenji	0.004633	True	
13	0.002129	True	Staffordshire_bullterrier	0.001498	True	
14	0.013834	True	redbone	0.007958	True	
15	0.194742	True	Saluki	0.027351	True	
16	0.120184	True	Labrador_retriever	0.105506	True	
17	0.184172	True	English_setter	0.073482	True	
18	0.166511	True	malamute	0.111411	True	
19	0.000932	True	bull_mastiff	0.000903	True	
20	0.087582	True	Weimaraner	0.026236	True	
21	0.029175	False	shopping_cart	0.026321	False	

```
22    0.054950   True                       beagle  0.038915   True
23    0.275645  False                   Weimaraner  0.134203   True
24    0.035029   True    Staffordshire_bullterrier  0.029705   True
25    0.143528   True                   can_opener  0.032253  False
26    0.186136   True               Dandie_Dinmont  0.086346   True
27    0.139952  False                    car_wheel  0.044173  False
28    0.000361   True                  Boston_bull  0.000076   True
29    0.025286  False                  Siamese_cat  0.002849  False
...        ...    ...                          ...       ...    ...
2035  0.271485  False                          gar  0.189945  False
2036  0.244220   True        flat-coated_retriever  0.173810   True
2037  0.009397   True                     papillon  0.004577   True
2038  0.259551   True                       briard  0.206803   True
2039  0.002402  False                      hamster  0.000461  False
2040  0.147655   True                  Eskimo_dog  0.093412   True
2041  0.054787   True           Labrador_retriever  0.014241   True
2042  0.278407   True                  groenendael  0.102643   True
2043  0.015250  False                great_grey_owl  0.013207  False
2044  0.063064   True              miniature_poodle  0.025581   True
2045  0.111884   True                       basenji  0.111152   True
2046  0.085547  False                      bookcase  0.079480  False
2047  0.033919  False                     partridge  0.000052  False
2048  0.149842   True                       borzoi  0.133649   True
2049  0.166192   True                Dandie_Dinmont  0.089688   True
2050  0.078260   True                      malinois  0.075628   True
2051  0.404722   True                French_bulldog  0.048960   True
2052  0.175382   True                  Ibizan_hound  0.097471   True
2053  0.174201   True                      Pekinese  0.109454   True
2054  0.093718   True           Labrador_retriever  0.072427   True
2055  0.192305   True  soft-coated_wheaten_terrier  0.082086   True
2056  0.014594  False              golden_retriever  0.007959   True
2057  0.058279   True                      fur_coat  0.054449  False
2058  0.045885  False                      terrapin  0.017885  False
2059  0.263788   True   Greater_Swiss_Mountain_dog  0.016199   True
2060  0.243682   True                      Doberman  0.154629   True
2061  0.360687   True           miniature_pinscher  0.222752   True
2062  0.138584   True                    bloodhound  0.116197   True
2063  0.074192   True          Rhodesian_ridgeback  0.072010   True
2064  0.156665   True              Shetland_sheepdog  0.061428   True


0                                                   This is Phineas. He's a
1         This is Tilly. She's just checking pup on you. Hopes you're doing ok. If not,
2                           This is Archie. He is a rare Norwegian Pouncing Corgo. Lives
3                                                                 This is Darla. She
4         This is Franklin. He would like you to stop calling him "cute." He is a very
5         Here we have a majestic great white breaching off South Africa's coast. Absol
6     Meet Jax. He enjoys ice cream so much he gets nervous around it. 13/10 help Jax e
```

```
7                            When you watch your owner call another dog a good boy but
8                       This is Zoey. She doesn't want to be one of the scary sharks.
9              This is Cassie. She is a college pup. Studying international doggo commu
10                This is Koda. He is a South Australian deckshark. Deceptively deadly.
11                                    This is Bruno. He is a service shark. Only get
12                                     Here's a puppo that seems to be on the fence
13                                                  This is Ted. He does his best.
14                            This is Stuart. He's sporting his favorite fanny pack. Se
15       This is Oliver. You're witnessing one of his many brutal attacks. Seems to be
16                                                        This is Jim. He found a f
17                 This is Zeke. He has a new stick. Very proud of it. Would like you
18                                                       This is Ralphus. He's powe
19                  This is Gerald. He was just told he didn't get the job he intervie
20                  This is Jeffrey. He has a monopoly on the pool noodles. Currently
21                                    I've yet to rate a Venezuelan Hover Wiener. Thi
22                                           This is Canela. She attempted some fan
23                                                    You may not have known you
24                 This... is a Jubilant Antarctic House Bear. We only rate dogs. Pleas
25                                         This is Maya. She's very shy. Rarely lea
26                       This is Mingus. He's a wonderful father to his smol pup. Confi
27                                                                    This is De
28              This is Roscoe. Another pupper fallen victim to spontaneous tongue eje
29                                                 This is Waffles. His doggles are
...
2035                           This is quite the dog. Gets really excited when not in
2036                            This is a southern Vesuvius bumblegruff. Can drive a t
2037                              Oh goodness. A super rare northeast Qdoba kangaroo m
2038
2039                           Unique dog here. Very small. Lives in container of Fros
2040                                Here we have a mixed Asiago from the Galápagos Isl
2041                                                      Look at this jokester think
2042                            This is an extremely rare horned Parthenon. Not amused
2043                             This is a funny dog. Weird toes. Won't come down. Lov
2044                               This is an Albanian 3 1/2 legged  Episcopalian. Lov
2045
2046
2047                                 Not familiar with this breed. No tail (weird). On
2048                                       Oh my. Here you are seeing an Adobe Setter
2049                           Can stand on stump for what seems like a while. Built t
2050                                   This appears to be a Mongolian Presbyterian mix
2051                                          Here we have a well-established sunblock
2052                              Let's hope this flight isn't Malaysian (lol). What a
2053                              Here we have a northern speckled Rhododendron. Much
2054                                                        This is the happiest d
2055                                Here is the Rand Paul of retrievers folks! He's pr
2056                                           My oh my. This is a rare blond Canadian
2057                           Here is a Siberian heavily armored polar bear mix. Stro
2058                             This is an odd dog. Hard on the outside but loving on
```

```
2059                          This is a truly beautiful English Wilson Staff retrieve
2060                                                Here we have a 1949 1st generation
2061                             This is a purebred Piers Morgan. Loves to Netflix an
2062                                     Here is a very happy pup. Big fan of well-mai
2063                           This is a western brown Mitsubishi terrier. Upset abou
2064                                   Here we have a Japanese Irish Setter. Lost eye

[1967 rows x 22 columns]
```

In [77]:
```
df_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1967 entries, 0 to 2064
Data columns (total 22 columns):
tweet_id              1967 non-null object
tweet_date            1967 non-null object
tweet_time            1967 non-null object
name                  1967 non-null object
stage                 1967 non-null category
retweet_count         1967 non-null int64
favorite_count        1967 non-null int64
rating_numerator      1967 non-null int64
rating_denominator    1967 non-null int64
source                1967 non-null object
expanded_urls         1967 non-null object
jpg_url               1967 non-null object
p1                    1967 non-null object
p1_conf               1967 non-null float64
p1_dog                1967 non-null bool
p2                    1967 non-null object
p2_conf               1967 non-null float64
p2_dog                1967 non-null bool
p3                    1967 non-null object
p3_conf               1967 non-null float64
p3_dog                1967 non-null bool
text                  1967 non-null object
dtypes: bool(3), category(1), float64(3), int64(4), object(11)
memory usage: 300.0+ KB
```

1.5 Storing the cleaned dataframe

In [78]:
```
# Store the clean DataFrame to a csv file
df_final.to_csv('twitter_archive_master.csv')
```

1.6 Analysis and Visualisation

Identify the 10 most common dog names

```
In [79]: #df_final['name'].value_counts()
         pd.DataFrame(df_final['name'].value_counts().head(11))
```

```
Out[79]:            name
         Not known   621
         Cooper       10
         Oliver       10
         Charlie      10
         Lucy         10
         Penny         9
         Tucker        9
         Winston       8
         Sadie         8
         Lola          7
         Toby          7
```

Statistical Summary table
Find the average dog rating, average favourite count and average retweet count

```
In [80]: df_final.describe()
```

```
Out[80]:        retweet_count  favorite_count  rating_numerator  rating_denominator  \
         count    1967.000000     1967.000000       1967.000000         1967.000000
         mean     2660.608541     8765.766650         12.225724           10.478393
         std      4736.170250    12837.312318         41.676305            6.860210
         min        11.000000       78.000000          0.000000            2.000000
         25%       591.000000     1882.000000         10.000000           10.000000
         50%      1281.000000     3951.000000         11.000000           10.000000
         75%      3038.500000    10934.000000         12.000000           10.000000
         max     83352.000000   163892.000000       1776.000000          170.000000

                    p1_conf       p2_conf       p3_conf
         count  1967.000000  1.967000e+03  1.967000e+03
         mean      0.594131  1.346345e-01  6.023006e-02
         std       0.272163  1.010457e-01  5.096530e-02
         min       0.044333  1.011300e-08  1.740170e-10
         25%       0.362880  5.339800e-02  1.608055e-02
         50%       0.587372  1.174020e-01  4.947150e-02
         75%       0.847139  1.955655e-01  9.158675e-02
         max       1.000000  4.880140e-01  2.734190e-01
```

Thus the average (mean) rating across the whole dataset is found to be 12.226, average favourite count is 8765.77 and average retweet count is 2660.61

Dog stages count Plot a bar chart to show the number of dogs according to the different dog stages - Pupper, Doggo, Puppo, Floofer, Doggo-pupper and Doggo-puppo

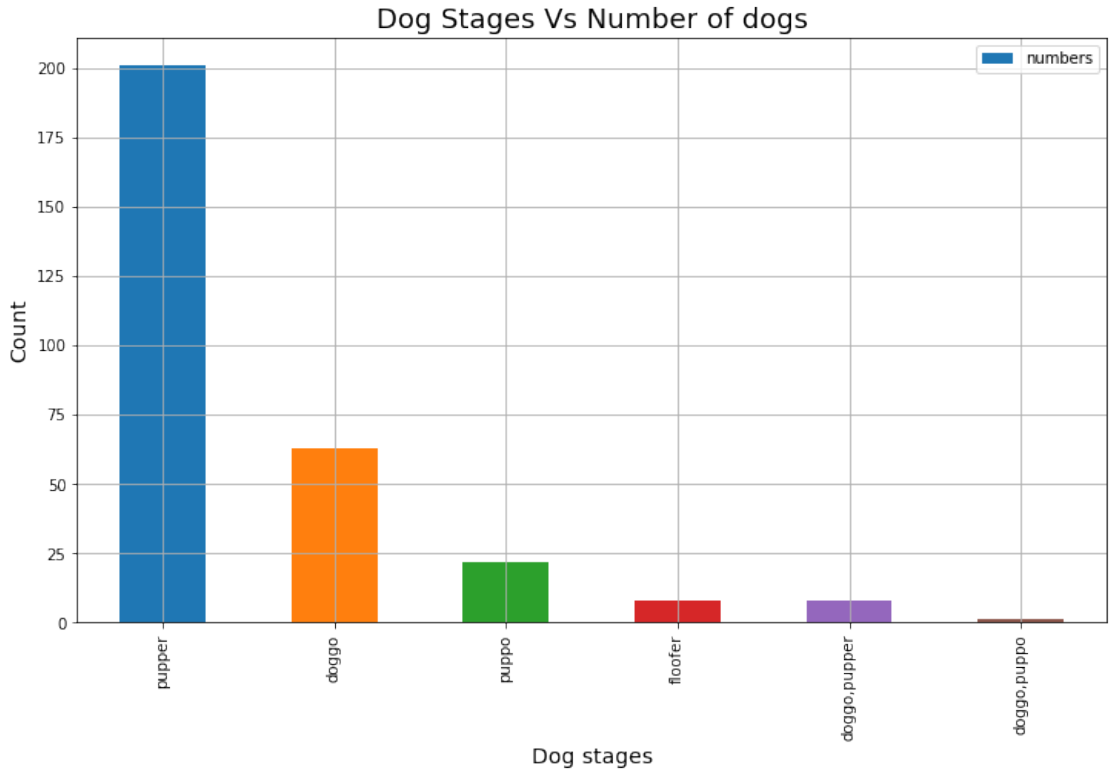
```
In [85]: count_stage = pd.DataFrame(df_final['stage'].value_counts())
         count_stage = count_stage.reset_index()
```

```
count_stage.columns = ['stage','numbers']
count_stage
```

Out[85]:

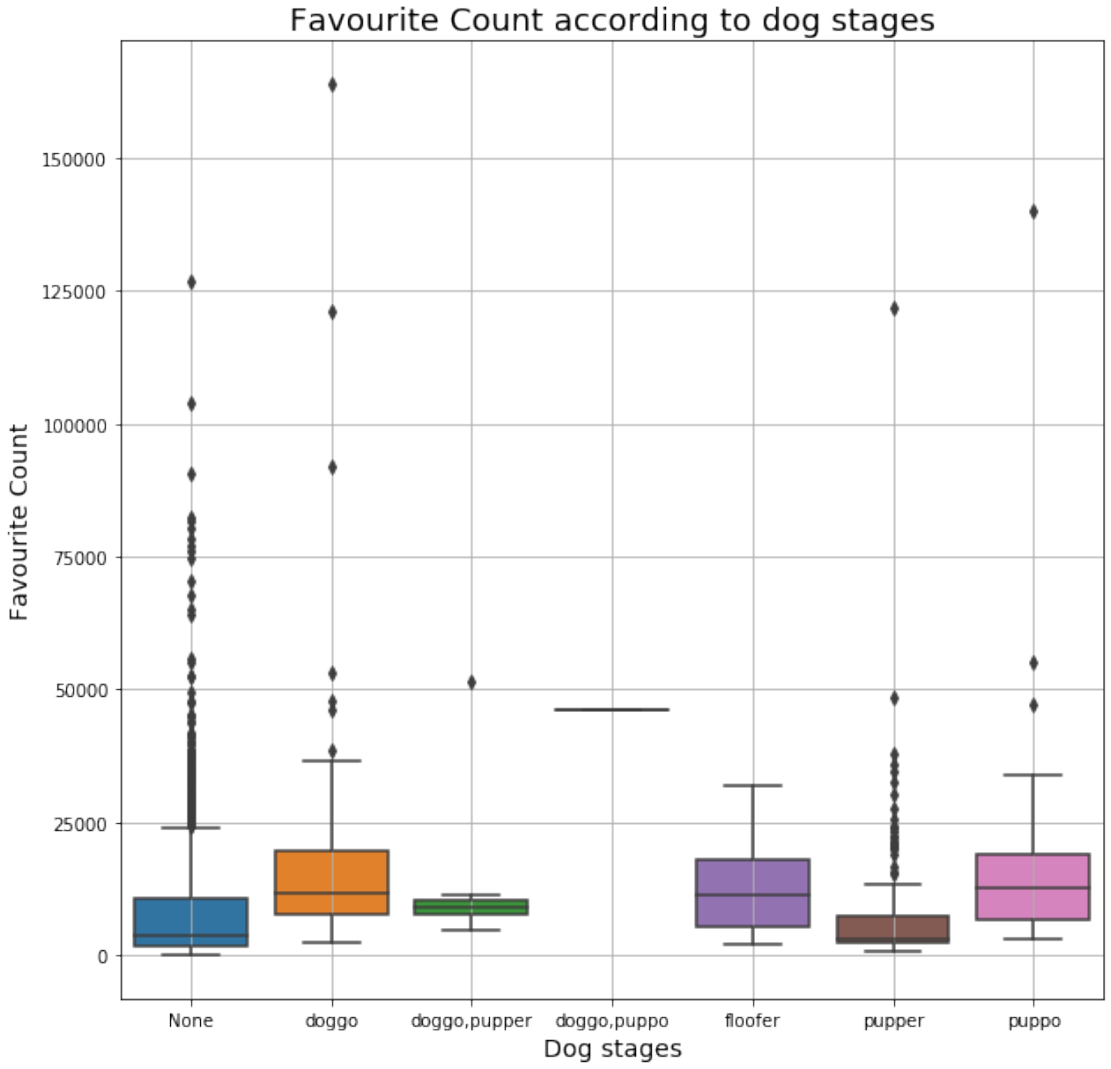
	stage	numbers
0	None	1664
1	pupper	201
2	doggo	63
3	puppo	22
4	floofer	8
5	doggo,pupper	8
6	doggo,puppo	1

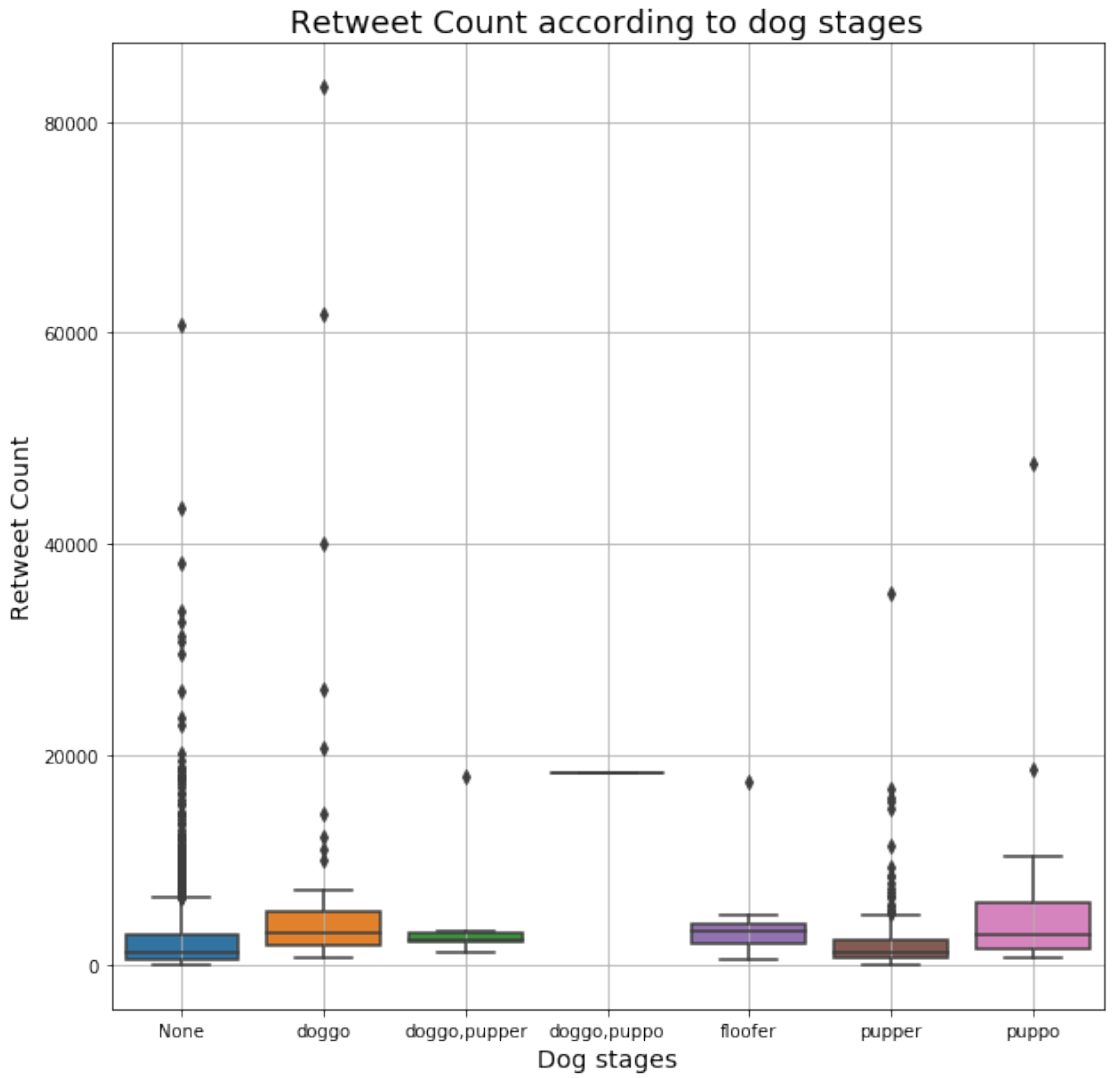
```
In [86]: count_stage = count_stage.drop(0) #remove the observations without any stages
         count_stage.plot(kind='bar',x='stage',y='numbers',figsize=(12, 7))
         plt.ylabel('Count',fontsize=14)
         plt.xlabel('Dog stages',fontsize=14)
         plt.title('Dog Stages Vs Number of dogs',fontsize=18)
         plt.grid(True)
         plt.show()
```



Analysis of favourite_count and retweets according to dog stages We plot a boxplot to show the favourite_count and retweets according to the different dog stages

```
In [87]: fig1=sns.boxplot(x=df_final.stage, y=df_final.favorite_count)
         plt.gcf().set_size_inches(10,10)
         plt.ylabel('Favourite Count',fontsize=14)
         plt.xlabel('Dog stages',fontsize=14)
         plt.title('Favourite Count according to dog stages',fontsize=18)
         plt.grid(True)
         plt.show()
         fig2=sns.boxplot(x=df_final.stage, y=df_final.retweet_count)
         plt.gcf().set_size_inches(10, 10)
         plt.ylabel('Retweet Count',fontsize=14)
         plt.xlabel('Dog stages',fontsize=14)
         plt.title('Retweet Count according to dog stages',fontsize=18)
         plt.grid(True)
         plt.show()
```



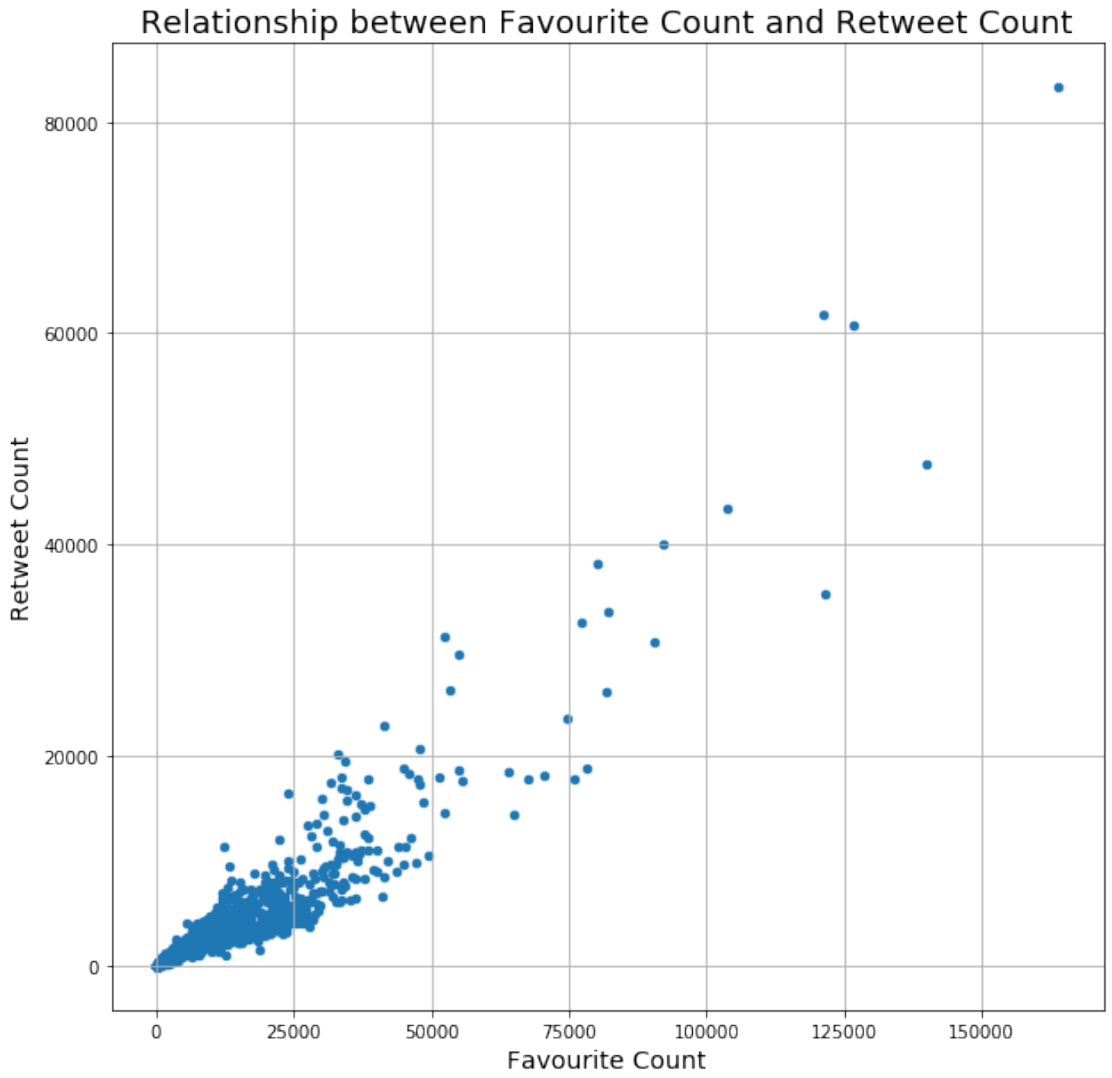


Thus it appears that although the puppers have the highest population, their average favourite_counts and retweet_counts are lower compared to the other stage groups. The average number of retweets for floofers are the highest, whereas for average favourite count, puppos and floofers lead the rankings.

Favourite Count and Retweet Count We will analyse the relation between the favourite_count and the retweet_count by plotting a scatterplot. We will then validate our findings by calculating the correlation coefficient.

```
In [88]: df_final.plot(kind='scatter',x='favorite_count',y='retweet_count',figsize=(10,10));
         plt.ylabel('Retweet Count',fontsize=14);
         plt.xlabel('Favourite Count',fontsize=14);
         plt.title('Relationship between Favourite Count and Retweet Count',fontsize=18);
         plt.grid(True)
         plt.show()
```

```
correlation_coefficient = df_final['favorite_count'].corr(df_final['retweet_count'])
print('Correlation_coefficient between favorite_count and retweet_count is {}'.format(c
```




```
Correlation_coefficient between favorite_count and retweet_count is 0.9296357865320213
```

From the above scatterplot we can observe that there is a strong positive correlation between the favorite_count and the retweet_count. Thus the tweets with higher favourite count also have a higher retweet count, and vice versa. This is further validated by the value of Correlation coefficient which is found to be **0.93**

Find the dog with the highest rating From the statistical summary table, the highest rating is found to be 1776.

```
In [89]: df_final[df_final['rating_numerator']==1776]
```

Out[89]:
```
              tweet_id  tweet_date tweet_time     name stage  retweet_count  \
796  749981277374128128  2016-07-04   15:00:45  Atticus  None           2623

     favorite_count  rating_numerator  rating_denominator     source  \
796            5364              1776                  10  TweetDeck

                                          ...                                  \
796                                       ...

          p1   p1_conf p1_dog          p2   p2_conf p2_dog        p3  \
796  bow_tie  0.533941  False  sunglasses  0.080822  False  sunglass

      p3_conf p3_dog  \
796  0.050776  False

                                                                            text
796  This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh

[1 rows x 22 columns]
```

In [90]:
```
url_highest = df_final[df_final['rating_numerator']==1776]['jpg_url']
url_highest
```

Out[90]:
```
796    https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg
Name: jpg_url, dtype: object
```

In [91]:
```
#import necessary libraries for displaying our images
import requests
from IPython.display import Image as IPythonImage
from PIL import Image
from PIL import ImageDraw
```

In [92]:
```
# Function to return the image for a given URL
def display_images(myurl):
    name='dog_image.png'
    dog_image = requests.get(myurl)
    with open(name,'wb') as image_file:
       image_file.write(dog_image.content)
    img = Image.open("dog_image.png")
    draw = ImageDraw.Draw(img)

    return img
```

In [93]:
```
#display the image of the highest rated dog named Atticus
img=display_images('https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg')
img = img.resize((300,400))
img.save('sample-out.png')
IPythonImage(filename='sample-out.png')
```

Out[93]:

Find the dog with the lowest rating From the table, the lowest rating is found to be 0

In [94]:
```
df_final[df_final['rating_numerator']==0]
```

Out[94]:

	tweet_id	tweet_date	tweet_time	name	stage	retweet_count	favorite_count	rating_numerator	rating_denominator	source
245	835152434251116546	2017-02-24	15:40:31	Not known	None	3232	23532	0	10	Twitter for iPhone

```
                                                                  ...
245                                                               ...

        p1   p1_conf p1_dog                                p2   p2_conf p2_dog  \
245  swing  0.967066  False  American_Staffordshire_terrier  0.012731   True

                            p3   p3_conf p3_dog  \
245  Staffordshire_bullterrier  0.007039   True


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

[1 rows x 22 columns]
```

In [95]:
```
url_lowest = df_final[df_final['rating_numerator']==0]['jpg_url']
url_lowest
```

Out[95]:
```
245    https://pbs.twimg.com/media/C5cOtWVWMAEjO5p.jpg
Name: jpg_url, dtype: object
```

In [96]:
```
#display the image of the lowest rated dog
img=display_images('https://pbs.twimg.com/media/C5cOtWVWMAEjO5p.jpg')
img = img.resize((300,400))
img.save('sample-out.png')
IPythonImage(filename='sample-out.png')
```

Out[96]:

Find the dog with the highest favourite count From the statistical summary table, the highest favourite count is found to be 163917.

```
In [98]: df_final[df_final['favorite_count']==163892]
```

```
Out[98]:                tweet_id  tweet_date tweet_time       name  stage  \
         845  744234799360020481  2016-06-18   18:26:18  Not known  doggo

              retweet_count  favorite_count  rating_numerator  rating_denominator  \
         845          83352          163892                13                  10

                        source  \
         845  Twitter for iPhone
```

```
                                                                  ...
845                                                               ...

                     p1   p1_conf p1_dog        p2   p2_conf p2_dog       p3  \
845  Labrador_retriever  0.825333   True  ice_bear  0.044681  False  whippet

      p3_conf p3_dog  \
845  0.018442   True


845  Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tin

[1 rows x 22 columns]
```

In [99]:
```
url_favorite = df_final[df_final['favorite_count']==163892]['jpg_url']
url_favorite
```

Out[99]:
```
845    https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img/1GaWmtJtdqzZW
Name: jpg_url, dtype: object
```

In [100]:
```
#display the image of the dog with highest favorite count
img=display_images('https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img
img = img.resize((300,400))
img.save('sample-out.png')
IPythonImage(filename='sample-out.png')
```

Out[100]:

Find the dog with the highest retweet count From the same table, the highest retweet count is found to be 83377.

```
In [101]: url_retweet = df_final[df_final['retweet_count']==83352]['jpg_url']
          url_retweet

Out[101]: 845    https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img/1GaWmtJtdqzZ
          Name: jpg_url, dtype: object

In [103]: #display the image of the dog with highest retweet count
          img=display_images('https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img
          img = img.resize((300,400))
          img.save('sample-out.png')
          IPythonImage(filename='sample-out.png')
```

Out[103]:

This 'doggo' with a video image has the highest favourite count and is also found to have the highest number of retweets. We had earlier found that favourite count and retweet count have a high correlation between them.