

---

## WRANGLING

---

## STEPS

---

### 1- GATHERING THE DATA

- KNOW WHAT DATA IS NEED TO BE COLLECTED IN WHATAT FORM , AND FROM WHERE TO COLLECT IT

### 2- Asserting

- Reading the datasets
- EXAMINIG THE DATA AND THE QULITY OR TIDNY ISSUES WITH IT

### 3- Cleaning

- Phase one : Data Cleaning
  - 1- Dealing with duplicate
  - 2- Removing Non-exexting Data
- Phase Two : Data Analysis in Python
- Phase Three : Visulaization and sharts

### 4- OUTPUT AND FINAL DATA AND Visulaization

---

## WRANGLING

### PART 1 :: GATHERING THE DATA

---

IN THIS PROJECT I HAVE TO DEAL WITH THREE DATA FROM THREE SEPREATS ORGINS

#### \* First

BY including the provided link of the first tow data and throw request in pytohn i was able to download the tow files and save both in ".csv" file format , and start to wokr on them.

#### \* Second

And to get the tweets from tweeter on the accoun of "dog\_rate" , started by creating tweeter accoun with more than five repalys on my request to get the approval on the "tweeter Developer accoun" and examin how to use the tweepy api to get tweets from certin ids that is proviede by the previlu downloaded in step one and save thes tweets with in ".csv" file format to start work on that file

Type *Markdown* and *LaTeX*:  $\alpha^2$

---

### PART 2 :: Asserting

# Data Assessment:

## Visualy and Programmatic Examining the Data

### A - Visualy Assessment

#### Visual assessment:

- By Openin the FILES [ 'api\_df.csv' , 'archive\_df.csv' , 'image\_predictions\_df.csv' ] , In Exel APP and VS CODE APP ON MY WINDOWS DESKTOP AND CHECK THE DATA

<https://t.co/IsVwOKSkd5b...>, <https://www.gofundme.com/minigusneedsus>, [https://twitter.com/dog\\_rates/status/886736880519319552/photo/1](https://twitter.com/dog_rates/status/886736880519319552/photo/1), [https://twitter.com/dog\\_rates/status/886736880519319552/photo/1](https://twitter.com/dog_rates/status/886736880519319552/photo/1), 13,10,Mingus 886680336477933568..., 2017-07-16 20:14:00 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Derek. He's late for a dog meeting. 13/10 pet..\_al to the metal <https://t.co/BQclueo0Aa>, 886361473445568..., 2017-07-15 23:53:31 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Roscoe. Another pupper fallen victim to spontaneous tongue ejections. Gets the BleepPen in 88626709285017600, 8,862663570751283e+17, 2281181600, 0, 2017-07-15 16:51:35 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, @NoWhiteHat @MayhemWayhem omg hell banner you are a scary good 886253841518873..., 2017-07-15 16:17:19 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Waffles. His doggies are upside down. Unsure how to fix. 13/10 someone assist Waffles ht 8866541600859672513..., 2017-07-14 02:45:17 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, #Athletics: 12/18 #B4TP <https://t.co/wkWemyFxo>, 8,86053734421026e+17, 19607400, 0, 2017-07-15 02:45:17 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Viewer discretion advised. This is Jimbo. He will rip your finger right h\*ckin off. Other dog clear 88598480019947528..., 2017-07-14 22:10:11 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Maisley. She fell asleep mid-excavation. Happens to the best of us. 13/10 will put noggins 885519481752093536..., 2017-07-13 19:59:47 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Maisley. She fell asleep mid-excavation. Happens to the best of us. 13/10 will put noggins 88311592012690284..., 2017-07-13 01:35:06 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, I have a new name and my name is Howdy. 12/10 <https://t.co/grlhBou7Se>, <https://t.co/4qzJGmDf>, 88516761983636784..., 2017-07-12 16:03:06 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, RT @Trollface\_12: This is Little. He's a little dog. He's parallel to the ground. He's not a dog. He's a dog. 13/10 8849552174179031..., 2017-07-12 00:01:00 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Earl. He found a hat. Nervous about what you think of it. 13/10 it's delightful. Earl ht 884876753390489601..., 2017-07-11 20:47:12 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Lola. It's her first time. Must test the earth and taste the atmosphere. 13/10 you 884562892145688576..., 2017-07-11 00:00:02 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Kevin. He's just so happy. 13/10 what is your secret Kevin <https://t.co/TrdMEFcX5...>, <https://t.co/TrdMEFcX5...>, 8844410532217440..., 2017-07-10 20:50:53 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, I present to you, Pup in Nat. Pup in Nat is great for all occasions. Extremely versatile. Compa 88424787851493888..., 2017-07-10 03:00:17 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Omg HE IS SO SORRY 13/10 WILL FORGIVE IM 88416267058437345..., 2017-07-09 21:29:42 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Meet Yogi. He doesn't have any important dog meetings today he just enjoys looking his best at all 883818122936631299..., 2017-07-09 00:00:04 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Noah. He can't believe someone made this mess. Got the vacuum out for you though. Offered 8848284693308286..., 2017-07-08 00:28:19 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Bella. She hopes her smile made you smile. If not, she is also offering you her favorite 883360690899218434..., 2017-07-07 16:52:55 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Meet Grizzelbow. He may be the floofiest floofier I ever did see. Lost eyes saving a schoolbus from 881183864686144..., 2017-07-07 17:15:54 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Please only send dogs. We don't rate mechanics, no matter how h\*ckin good. Thank you. 13/10 wo 88292080364220416..., 2017-07-06 15:58:11 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Rusty. He wasn't ready for the first pic. Clearly puffed for the second. 13/10 confirmed 88272694511734784..., 2017-07-06 00:46:41 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Gus. He's quite the cheeky pup. Already perfect the distanced wink. 12/10 would 88262720321602560..., 2017-07-05 15:45:34 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Stanley. He has his first swim lesson today. Doggie straps adjusted. Ready to go. 13/10 P 882681181099369728..., 2017-07-04 16:01:23 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Alf. You're witnessing his first watermelon experience. I think it was a success. 13/10 882045870635918589..., 2017-07-04 01:18:17 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Koko. Her owner, Inspired by Barney, recently built a cart for her to use during walks i 8819658987149721986..., 2017-07-03 16:00:45 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Rey. He's a Benobop Cumberflod. 12/10 dangerously pettable <https://t.co/5083QbhdQd...>, <https://t.co/5083QbhdQd...>, 881666595344535552..., 2017-07-03 00:11:00 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Gary. He couldn't miss this opportunity for a selfie. Flawless focusing skills. 13/10 wo 881613300179430080..., 2017-07-02 21:18:19 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, #pushkinway These are good dogs but 17/10 is an emotional impulse 88155604880672076..., 2017-07-02 15:52:45 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Here is a pupper approaching maximum barkdrive. Zooming at never before seen speeds. 14/10 paw-i (<https://t.co/0D001E1QZT>)..., <https://t.co/0D001E1QZT>..., <https://t.co/0D001E1QZT>..., 8811684419642592..., 2017-07-02 17:14:34 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Meet Elliott. He's a Canadian Forrest Pup. Unusual number of antlers for a dog. Sneaky tongue clip 880931576299988482..., 2017-06-30 23:47:07 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Louis. He's crossing. It's a big deal. 13/10 h\*ckin breathtaking <https://t.co/00bGIkAt>, 88887244881571648..., 2017-06-30 16:39:47 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Ugh not again. We only rate dogs. Please don't send in well-dressed floppy-tongued street penguin 8804658323668113184..., 2017-06-29 16:39:47 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Bella. She had her first beach experience this morning. Complete success. 12/10 would per 8822117723809552..., 2017-06-28 00:27:25 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Meet Jesse. He's a Fatty Woof. His tongue ejects without warning. A true bleptomaniac. 12/10 would 88090578278096641..., 2017-06-28 16:09:00 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Please don't send in photos without dogs in them. We're not g\*orch rates. Insubordinate and churl 87967431964279634..., 2017-06-27 22:42:13 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Romeo. He would like to do an entrance. Requesting your immediate assistance. 13/10 https://t.co/14\_10...Nones, 8795538273341727, 3105440746..., 2017-06-27 12:14:34 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, @RealKentMurphy 34/10 confirmed, ..., 14,10,None,None,None,None, 8794916609185150312448..., 2017-06-27 00:10:17 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Bailey. He thinks you should measure ear length for signs of growth instead. 12/10 https://t.co/79953821135344209921, 2321, 15359, 3200889 879415811825184262..., 2017-06-26 19:07:24 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Doodles. He did an attempt. 13/10 someone help him <https://t.co/0idbGeLci> <https://t.co/0idbGeLci> 87937462958149265..., 2017-06-26 16:51:39 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Jack AKA Stephen Furry. You're not scoring on him. Unless he slips down the slide. 12/10 879130579576457649..., 2017-06-26 00:13:58 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, RT @doge\_rates: This is Emmy. She was adopted today. Massive round of puppylove for Emmy and her ne 87905749262655488..., 2017-06-25 18:56:45 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Steven. He has trouble relating to other dogs. Quite shy. Neck longer than average. Tropic 8790882329182956..., 2017-06-25 04:07:45 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Beau. That is Beau's balloon. He takes it everywhere. 13/10 would protect at all costs ht 87776693423087618..., 2017-06-25 00:45:22 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, This is Snooty. He's a proud #PrideMonthPupper. Impeccable handwriting for not having thumbs. 13/10 87866470711726852..., 2017-06-24 13:24:20 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, Martha is stunning how h\*ckin dare you. 13/10 <https://t.co/9uABQxjMw...>, <https://t.co/9uABQxjMw...>, 87866470711726852..., 2017-06-24 13:24:20 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, RT @doge\_rates: Mac Shadow. In an attempt to reach maximum zooming barkdrive he force his ... 87866470711726852..., 2017-06-24 13:24:20 +0000 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, RT @doge\_rates: Mac Shadow. In an attempt to reach maximum zooming barkdrive he force his ...

```
You, an hour ago | 1 author (you)
1 | tweet_id, retweet_count, favorite_count, user_count
2 | 0,892420643555331693,8853,39467,3200889
3 | 1,89217421306343426,6514,33819,3200889
4 | 2,89181518137808464,4328,25461,3200889
5 | 3,891689557279858688,8964,42908,3200889
6 | 4,89132755892668256,9774,41048,3200889
7 | 5,891897958875897856,3261,28562,3200889
8 | 6,890971913173991426,2158,12041,3200889
9 | 7,890729181411237888,16716,56848,3200889
10 | 8,890609185150312448,4429,28226,3200889
11 | 9,890240255349198849,7711,32467,3200889
12 | 10,89006608113172480,7624,31166,3200889
13 | 11,88880896479866881,5156,28268,3200889
14 | 12,889665388333682689,8538,38818,3200889
15 | 13,889638837579907072,4735,27672,3200889
16 | 14,889531135344209921,2321,15359,3200889
17 | 15,889278841981685760,5637,25652,3200889
18 | 16,888912738123831296,4709,29611,3200889
19 | 17,888804989199671297,4559,26080,3200889
20 | 18,888554962724278272,3732,20290,3200889
21 | 19,888078434458587136,3653,22201,3200889
22 | 20,888705289381826560,5609,30779,3200889
23 | 21,887517139158093824,12082,46959,3200889
24 | 22,887473957103951883,18781,69871,3200889
25 | 23,887343217045368832,10737,34222,3200889
26 | 24,887101392804085766,6167,31061,3200889
27 | 25,88698323352544640,8084,3589,3200889
28 | 26,886736880519319552,3443,12306,3200889
29 | 27,886680336477933568,4610,22798,3200889
30 | 28,88636614473445568,3316,21524,3200889
31 | 29,886267009285017600,4,117,3200889
32 | 30,8862583841518873,6523,28469,3200889
33 | 31,88605160059072513,108,0,3200889
34 | 32,885984800019947520,7097,33382,3200889
35 | 33,8855289432805470208,6683,36689,3200889
36 | 34,885518971528720385,3899,20788,3200889
37 | 35,885311592912609280,19297,0,3200889
38 | 36,885167619883638784,4556,22367,3200889
39 | 37,884925521741709313,16439,68152,3200889
40 | 38,884876753390489601,6096,28514,3200889
41 | 39,884562892145688576,5100,24765,3200889
42 | 40,88441805382717440,5856,27478,3200889
43 | 41,884247878851493888,21298,74423,3200889
44 | 42,884162670584377345,3128,20771,3200889
45 | 43,883838122936631299,3586,22349,3200889
46 | 44,883482846933004288,10407,46860,3200889
47 | 45,883360690899218434,3825,22986,3200889
48 | 46,883117836046086144,6949,37914,3200889
```

1	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog	You, an hour ago · Finished Assessing data lesson
2	66602088822790149	https://pbs.twimg.com/media/CT4audn0QwAAyMly.jpg	1	Wheat_springer	0.465074	True	collie	0.156665	True	Shetland_sheepdog	0.0614285	True	
3	666029285002629928	https://pbs.twimg.com/media/CT42GrqYUAA5lDo.jpg	1	redbone	0.506826	True	miniature_pinscher	0.07419169999999995	True	Rhodesian_ridgeback	0.07201	True	
4	666833412701032449	https://pbs.twimg.com/media/CT4521TlwAEWyu.jpg	1	German_shepherd	0.596461	True	malinois	0.1385839999999998	True	bloodhound	0.11619700000000001	True	
5	666044226329800794	https://pbs.twimg.com/media/CT45D8HUEAAT-1EU-.jpg	1	Rhodesian_ridgeback	0.480813	True	redbone	0.360087	True	miniature_pinscher	0.22725	True	
6	6666942416582465	https://pbs.twimg.com/media/CT5QsXIAAX4A-.jpg	1	miniature_pinscher	0.560311	True	Rottweiler	0.243682	True	Doberman	0.154629	True	
7	66669875894694657	https://pbs.twimg.com/media/CT5QsXIAAX4A-.jpg	1	Bernese_mountain_dog	0.651137	True	English_springer	0.263788	True	Greater_Swiss_Mountain_dog	0.6	True	
8	666651853826850816	https://pbs.twimg.com/media/CT5QsXIAAxJlAoA1ash.jpg	1	box_turtle	0.9330120000000001	False	mud_turtle	0.04585480000000001	False	terrapin	0.01788529999999996	True	
9	66669552042405388	https://pbs.twimg.com/media/CT5Npt0tXIAAf1s.jpg	1	chow	0.692517	True	Tibetan_mastiff	0.05827939999999995	True	fur_coat	0.0544486	False	
10	666057090499244032	https://pbs.twimg.com/media/CT5PY90w0AAQGQo.jpg	1	shopping_cart	0.962465	False	shopping_basket	0.01459379999999999	False	golden_retriever	0.09795896	True	
11	666658600524156928	https://pbs.twimg.com/media/CT5Qg94XAA4_2dp.jpg	1	miniature_poodle	0.201493	True	komondor	0.192305	True	soft-coated_wheaten_terrier	0.0820861000000	True	
12	666663827256865533	https://pbs.twimg.com/media/CT5Vg_wXIAAxNjF.jpg	1	golden_retriever	0.77593	True	Tibetan_mastiff	0.0597178	True	Labrador_retriever	0.24266000000000001	True	
13	666671192321509120	https://pbs.twimg.com/media/CT5N_3WEA0I0Z.jpg	1	Gordon_setter	0.503672	True	Yorkshire_terrrier	0.174281	True	Pekinese	0.109454	True	
14	666673100786774016	https://pbs.twimg.com/media/CT5Qjof1WUEAvNw.jpg	1	Walker_kerry_blue	0.260857	True	English_foxhound	0.111884	True	Ibizan_hound	0.0974705	True	
15	666682916733198337	https://pbs.twimg.com/media/CT5MgVGWEAAkt8c.jpg	1	pug	0.489814	True	bulldog	0.4047219999999997	True	French_bulldog	0.0489595	True	
16	66669400022159362	https://pbs.twimg.com/media/CT5Np7K4AAfBWN.jpg	1	bloodhound	0.195217	True	German_shepherd	0.0782598	True	malinois	0.07562780000000001	True	
17	666699513787052032	https://pbs.twimg.com/media/CT51_JJUEAAHv8.jpg	1	Lhasa	0.58233	True	Shih-Tzu	0.166192	True	Dandie_Dimont	0.0896883	True	
18	66610215909144576	https://pbs.twimg.com/media/CT54Y6iWAEZnKo.jpg	1	English_setter	0.298617	True	Newfoundland	0.149842	True	borzoi	0.133649	True	
19	66610413238865088	https://pbs.twimg.com/media/CT8Qd1WEAA0Ds.jpg	1	hen	0.963932	False	cock	0.0339194	False	partridge	0.20568e-05	False	
20	66626891080364416	https://pbs.twimg.com/media/CT8Qd1WEAA0Ds.jpg	1	desktop_computer	0.085602	False	desk	0.0855474	False	bookcase	0.0794797	False	
21	66627309716637952	https://pbs.twimg.com/media/CT5Qd1WEAA0Qm.jpg	1	Italian_greyhound	0.176083	True	toy_toyrier	0.111884	True	basenji	0.111152	True	
22	666274806224695296	https://pbs.twimg.com/media/CT8QgBpUEAAfFjg.jpg	1	Maltese_dog	0.8575309999999999	True	toy_poodle	0.0630638	True	miniature_poodle	0.0255806	True	
23	666293916732131444	https://pbs.twimg.com/media/CT5Wm7K4AAfBWN.jpg	1	three-toed_sloth	0.9146789999999999	False	otter	0.01525	False	great_prey_owl	0.0132072	False	
24	666337882303524864	https://pbs.twimg.com/media/CT9QnFtWEAMRje.jpg	1	ox	0.4166889999999999	False	Newfoundland	0.278407	True	groenendael	0.10264300000000001	True	
25	666345417576210432	https://pbs.twimg.com/media/CT9Vn7Pw0AA_ZCM.jpg	1	golden_retriever	0.8587440000000001	True	Chesapeake_Bay_retriever	0.05478680000000004	True	Labrador_retriever	0.24266000000000001	True	
26	66635328456101888	https://pbs.twimg.com/media/CT5Qd1WEAA0Qm.jpg	1	maimata	0.33687399999999995	True	Siberian_husky	0.147655	True	Eskimo_dog	0.0934123999999999	True	
27	666362758999284353	https://pbs.twimg.com/media/CT9X1GsuCAVUpT.jpg	1	guinea_pig	0.9964959999999999	False	skunk	0.00240425	False	hamster	0.0004608630000000005	True	
28	66637373744588802	https://pbs.twimg.com/media/CT5Qd1WEAA0Qm.jpg	1	soft-coated_wheaten_terrier	0.326467	True	Afghan_hound	0.25951000000000003	True	briard	0.20680300000000002	True	
29	666396247373291520	https://pbs.twimg.com/media/CT-DZHH1AAfXg1k.jpg	1	Chiuhuahua	0.978108	True	toy_toyrier	0.00939697	True	papillon	0.00457681	True	
30	6664097173279398237	https://pbs.twimg.com/media/CT-pkmB17AAfXg1k.jpg	1	black-and-tan_coonhound	0.529139	True	bloodhound	0.244222	True	flat-coated_retriever	0.17381	True	
31	666411507551481857	https://pbs.twimg.com/media/CT-RugiWAAfLEaqZ.jpg	1	coho	0.40464	False	barracouta	0.271485	False	gar	0.189954	False	
32	666418789513265929	https://pbs.twimg.com/media/CT-YmB7UAAfQnW.jpg	1	toy_terrier	0.14968	True	papillon	0.148285	True	Chiuhuahua	0.14286	True	
33	666421515873626688	https://pbs.twimg.com/media/CT-agzXAAfMT3J2.jpg	1	Blenheim_spaniel	0.986777	True	cocker_spaniel	0.0903464	True	Shih-Tzu	0.00111687	True	
34	66642872649377476	https://pbs.twimg.com/media/CT-g-0DUwAQd5n.jpg	1	Pembroke	0.371361	True	chow	0.249394	True	Pomeranian	0.1421878	True	
35	66643072426358785	https://pbs.twimg.com/media/CT-nTOWEAfDn.jpg	1	llama	0.505184	False	Irish_terrier	0.104109	True	dingo	0.0620712	False	
36	666435652385423360	https://pbs.twimg.com/media/CT-nTOWEAfDn.jpg	1	Chesapeake_Bay_retriever	0.184113	True	chain_saw	0.05675730000000001	False	power_drill	0.0367634	False	
37	66643727313998237	https://pbs.twimg.com/media/CT-pkmB17AAfXg1k.jpg	1	Chiuhuahua	0.6718529999999999	True	beagle	0.124688	True	Satuki	0.0449941999999999	True	
38	6664734441048738	https://pbs.twimg.com/media/CT-y5QwAAfJLX5.jpg	1	curly-coated_retriever	0.32208400000000004	True	giant_schnauzer	0.287955	True	Labrador_retriever	0.16663	True	
39	666454714377183233	https://pbs.twimg.com/media/CT-5ws-MUAAl2Ec.jpg	1	dalatian	0.278954	True	Labrador_retriever	0.237612	True	Great_Pyrenees	0.171106	True	
40	66646448231647191804	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	Ibizan_hound	0.044334	True	Pembroke	0.0432093	True	West_Highland_white_terrier	0.0389065	True	
41	666464948231509281	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	Border_colie	0.44780299999999995	True	English_springer	0.170497	True	collie	0.139206	True	
42	666691418707132416	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	German_shepherd	0.9754010000000001	True	beagle	0.0868727	True	bloodhound	0.00539404	True	
43	6670116282335120	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	Labrador_retriever	0.887707	True	Chiuhuahua	0.092397	True	French_bulldog	0.0207563	True	
44	66673932793983658	https://pbs.twimg.com/media/CT-UCAat0WAtXg1zey.jpg	1	miniature_poodle	0.546933	True	cocker_spaniel	0.1652549999999999	True	toy_poodle	0.0959589	True	
45	66677690847630848	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	sest_beit	0.37505700000000003	False	miniature_pinscher	0.167175	True	Chiuhuahua	0.0869506	True	
46	666781797255496192	https://pbs.twimg.com/media/CT-Ub1WAAfJLX5.jpg	1	Italian_greyhound	0.618136	True	Weimaraner	0.151363	True	vizsla	0.08598100000000001	True	
47	66678068205871104	https://pbs.twimg.com/media/CTUDMzIkWAAfATPe.jpg	1	snail	0.9998879999999999	False	slug	5.51417e-05	False	acorn	2.6258e-05	False	
48	666804364988780544	https://pbs.twimg.com/media/CTUD3AYm0A82H0.jpg	1	English_setter	0.32879200000000003	True	Brittany_spaniel	0.283545	True	Ibizan_hound	0.0574615	True	

You, an hour ago · Author (You)

"Created at: Aug 01 16:23:56 +0000 2017", "id": 8924206435536193, "id\_str": "8924206435536193", "full\_text": "This is Phineas. He's a mystical boy. Only ever appears in the hole of a donut. 13/10 https://t.co/M6Uu76D6jU", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Aug 01 16:23:56 +0000 2017", "id": "8924206435536193", "id\_str": "8924206435536193", "full\_text": "This is Shelly. She's just checking pup on you. Hopes you're doing ok. If not, she's available for pats, snugs, barks, and belly rubs.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 31 00:18:03 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Sun Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +0000 2017", "id": "89181518378084864", "id\_str": "89181518378084864", "full\_text": "This is Shelly. She's a rare Norwegian Pouncing Corgo. Lives in the tall grass. You never know when one may strike.", "source": "Twitter for iPhone", "truncated": false, "in\_reply\_to\_status\_id": null, "in\_reply\_to\_status\_id\_str": null, "in\_reply\_to\_user\_id": null, "in\_reply\_to\_user\_id\_str": null, "in\_reply\_to\_screen\_name": null, "in\_reply\_to\_user\_name": null, "geo": null, "coordinates": null, "place": null, "contributors": null}, {"created\_at": "Mon Jul 30 13:58:23 +

Below i provided some photos of the results of using code to visually Programmatic the data given :

```
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Unnamed: 0        2356 non-null   int64  
 1   tweet_id          2356 non-null   int64  
 2   in_reply_to_status_id  78 non-null   float64 
 3   in_reply_to_user_id  78 non-null   float64 
 4   timestamp          2356 non-null   object  
 5   source             2356 non-null   object  
 6   text               2356 non-null   object  
 7   retweeted_status_id 181 non-null   float64 
 8   retweeted_status_user_id 181 non-null   float64 
 9   retweeted_status_timestamp 181 non-null   object  
 10  expanded_urls      2297 non-null   object  
 11  rating_numerator  2356 non-null   int64  
 12  rating_denominator 2356 non-null   int64  
 13  name               2356 non-null   object  
 14  doggo              2356 non-null   object  
 15  floofier            2356 non-null   object  
 16  pupper              2356 non-null   object  
 17  puppo              2356 non-null   object  
dtypes: float64(4), int64(4), object(10)
memory usage: 331.4+ KB
```

[11] > ML

```
print('finished in : ' + str((time.time() - start)))
max    2355.000000  8.924206e+1/      8.862664e+1/      8.40547e+1/
                                                 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
*****
Count of actual data in each column against the total number of rows:
```

[11] > ML

```
Count of actual data in each column against the total number of rows:

Unnamed: 0          2356
tweet_id            2356
in_reply_to_status_id  78
in_reply_to_user_id  78
timestamp           2356
source              2356
text                2356
retweeted_status_id 181
retweeted_status_user_id 181
retweeted_status_timestamp 181
expanded_urls       2297
rating_numerator   2356
rating_denominator 2356
name                2356
doggo               2356
floofier             2356
pupper               2356
puppo                2356
dtype: int64
*****
```

[11] > ML

Non existing data in each row:

```
Unnamed: 0          0
tweet_id            0
in_reply_to_status_id  2278
in_reply_to_user_id  2278
timestamp           0
source              0
text                0
retweeted_status_id 2175
retweeted_status_user_id 2175
retweeted_status_timestamp 2175
expanded_urls       59
rating_numerator   0
rating_denominator 0
name                0
doggo               0
floofier             0
pupper               0
puppo                0
dtype: int64
*****
```

Duplicated data :

[11] > ML

```

*****
Shape of the data:
(2075, 13)
*****
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 13 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Unnamed: 0    2075 non-null   int64  
 1   tweet_id     2075 non-null   int64  
 2   jpg_url      2075 non-null   object  
 3   img_num      2075 non-null   int64  
 4   p1           2075 non-null   object  
 5   p1_conf      2075 non-null   float64 
 6   p1_dog       2075 non-null   bool    
 7   p2           2075 non-null   object  
 8   p2_conf      2075 non-null   float64 
 9   p2_dog       2075 non-null   bool    
 10  p3           2075 non-null   object  
 11  p3_conf      2075 non-null   float64 
 12  p3_dog       2075 non-null   bool    
dtypes: bool(3), float64(3), int64(3), object(4)

```

## C - Asserting issues with Data

### Quality and tidiness issues

#### Quality

##### 1- Missing values :

A - Missing values in "twitter-archive-enhanced" Data set , AS AN EXAMPLE - in :

- "in\_reply\_to\_status\_id"
- "in\_reply\_to\_user\_id"
- "retweeted\_status\_id"
- "retweeted\_status\_user\_id"
- 'name'
- "source"

columns .

B - Missing values in "twitter-archive-enhanced" , AS AN EXAMPLE - in :

```

* Row # "2351" with tweet id    "666049248165822465" , have name with "None" vlaue .
* Row # "2352" with tweet id    "666044226329800704" , have name with "a" letter vlaue .
* Row # "2353" with tweet id    "666033412701032449" , have name with "a" letter vlaue .
* Row # "2354" with tweet id    "666029285002620928" , have name with "a" letter vlaue .
* Row # "2355" with tweet id    "666020888022790149" , have name with "None" vlaue .

```

ROWS .

##### 2- Duplicated Tweets in "twitter-archive-enhanced" Data in

```

"retweeted_status_id"      ,with count of    181
"retweeted_status_user_id" ,with count of    181
"retweeted_status_timestamp",with count of    181

```

columns.

##### 3- DATA TYPES

- IN "twitter-archive-enhanced" , THE "retweeted\_status\_timestamp" COLUMN DATA TYPE IS 'object' NOT "TIME STAMP" OR "DATE TIME " TYPE.
- IN "twitter-archive-enhanced" , "tweet\_id" column is "int" type with no need to do analysis on it so it may convert to string .

##### 4- DATA VALUES

- Validity:

IN "twitter-archive-enhanced" there is :

```

* "rating_denominator" HAVE DATA VLUES LESS THAN 10 WICH IS THE MINMIMUM RATING FOR DOG , with count of 3 rat
ings

* "rating_numerator"   HAVE DATA VLUES LESS THAN 10 WICH IS THE MINMIMUM RATING FOR DOG , with count of 440 r
atings

```

5 - More than one column in "image\_predictions" , for the smae vlaue and measure in AKA ( DOGS IS seperated IN 4 COLUMNS )

```
[p1, p2 ,p3]  
[ p1_dog , p2_dog , p3_dog ]  
[ p1_conf , p2_conf , p3_conf ]  
6 - In "image_predictions" , columns header are vales not variables name
```

---

## Tidiness

- 1 - All the data is in relation in each other but seperated into three tables
- 2 - COLUMN with no useful , or repated data , must be droped from the final data

Type *Markdown* and *LaTeX*:  $\alpha^2$

---

## PART 3 :: Cleaning

- Phase one : Data Cleaning
    - 1- Dealing with duplicate
    - 2- Removing Non-existing Data
  - Phase Two : Data Analysis in Python
  - Phase Three : Visualization and charts
- 

### Data Cleaning:

- AT THE END OF THE process we need data that contain the "dog name" if available, the "dog type" AKA "dog type" if available, and "rating tweets" , that dog type get over the time .

In Cleaning either drop the repeated columns the give the same answer or merge them each catigory in one column

in order to do so I need a data with only "tweet id" that is not duplicated , and original tweet no retweets and also a rating with the most predection available,

- To do so i will do the follow process :

- 1- merge the types of the dog in one column "dog type" contain the type rather than (False , True) boolean column
- 2- will drop the other tow less predection , columns , drop text name after extraxting the dog type from the txt column
- 3- drop the ratings that violate the rating measures
- 4- creating a clean data and save it to new ".csv" data type .

Type *Markdown* and *LaTeX*:  $\alpha^2$

---

## WRANGLING

## THE END RESULT

---

### Output:

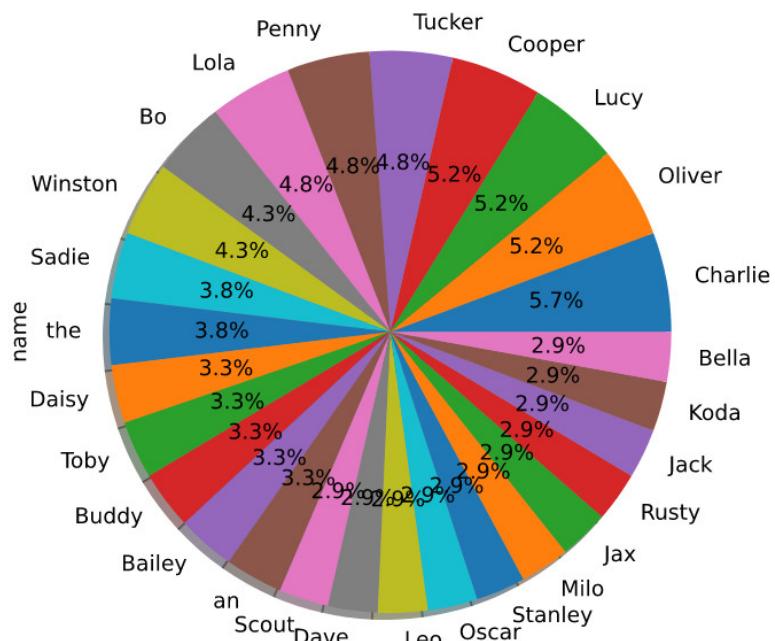
The file created as result of this process.

THE OUTPUT FINAL FILE IS "twitter\_archive\_master.csv" DATA WHICH IS HAVE IN MY VIEW THE OPTIMAL DATA CLEANED

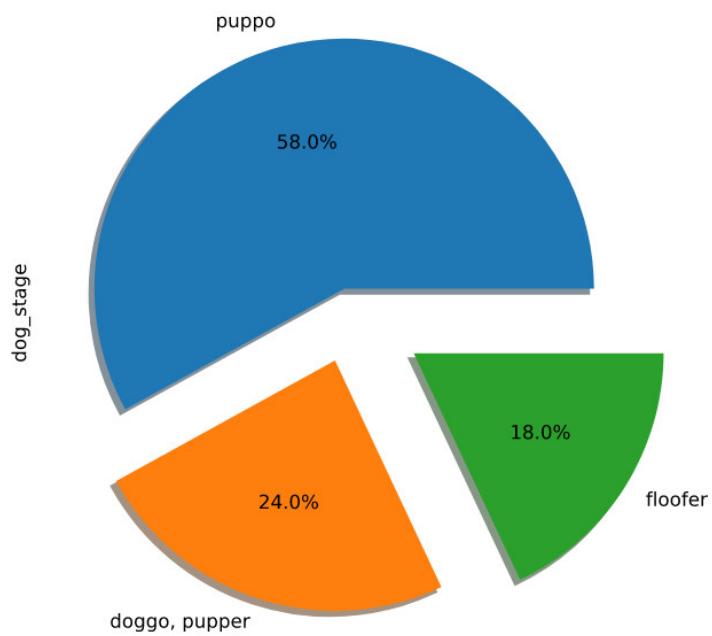
## WICH WILL BE USED IN THE NEXT PHASES :

- Phase Two : Data Analysis in Python
- Phase Three : Visulaization and sharts

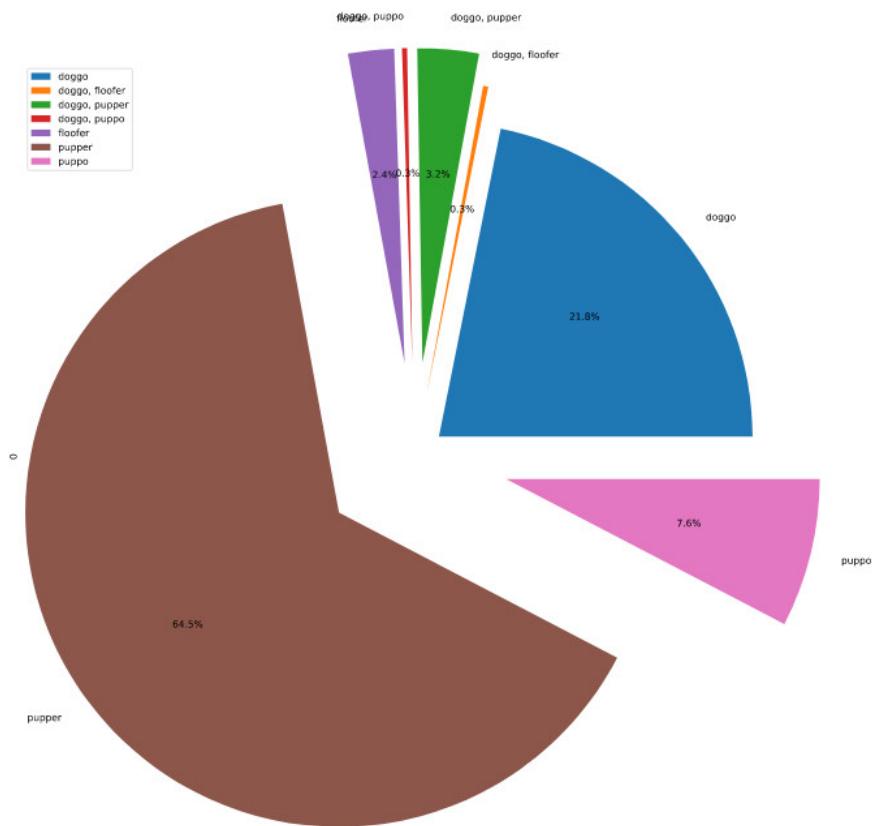
Most names , tweeted for dogs, percentage



DOG STAGE , apperd int tweets for dogs, percentage



Rating numerator , apperd in tweets for dogs, percentage



In [ ]:

1