

该项目基于推特 WeRateDogs 档案数据，结合对档案图像预测数据以及推特的转发和点赞数据，通过对数据质量和整洁度的一系列评估与清洗工作，我们最终得到了可用于数据分析和可视化的数据集 twitter\_archive\_master.csv。在 Jupyter Notebook 中以 DataFrame 格式载入该数据集，分析如下四个问题：

A. 获取点赞和转发总数排名前三的狗狗推特信息；

	tweet_id	timestamp	text	expanded_urls	rating_numerator	rating_denominator		
629	744234799360020481	2016-06-18 18:26:18	Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad) https://t.co/7wE9LTEXC4	https://twitter.com/dog_rates/status/744234799360020481/video/1	13	10		
1676	822872901745569793	2017-01-21 18:26:02	Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10 https://t.co/nTz3FtorBc	https://twitter.com/dog_rates/status/822872901745569793/photo/1	13	10		
329	807106840509214720	2016-12-09 06:17:20	This is Stephan. He just wants to help. 13/10 such a good boy https://t.co/DkBYaCag2d	https://twitter.com/dog_rates/status/807106840509214720/video/1	13	10		

name	stage	prediction1_result	prediction1_confidence	prediction1_is_dog	prediction2_result	prediction2_confidence	prediction2_is_dog	prediction3_result
NaN	doggo	Labrador_retriever	0.825333	True	ice_bear	0.044681	False	whippet
NaN	puppo	Lakeland_terrier	0.196015	True	Labrador_retriever	0.160329	True	Irish_terrier
Stephan	NaN	Chihuahua	0.505370	True	Pomeranian	0.120358	True	toy_terrier

prediction3_confidence	prediction3_is_dog	retweet_count	favorite_count	total_retweet_favorite_count
0.018442	True	77236	128101	205337
0.069126	True	49045	143195	192240
0.077008	True	61007	122842	183849

由上表可知转发和点赞总数排名前三的狗狗推特信息如下：

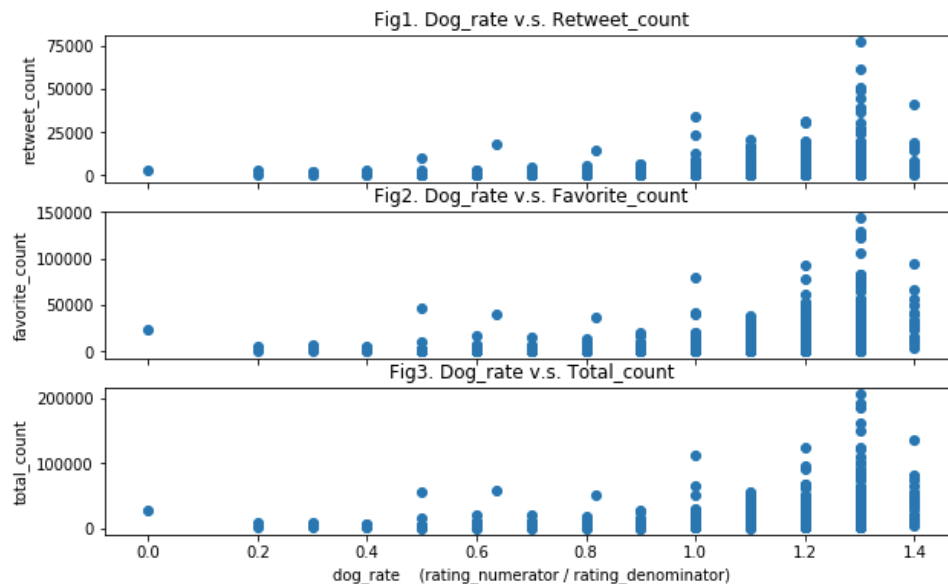
- 排名第一的狗狗推特**是 2016 年 6 月 18 日用户（推特 ID:744234799360020481）发表的关于拉布拉多 doggo（名字不详）在游泳池游泳的一段视频——>  
([https://twitter.com/dog\\_rates/status/744234799360020481/video/1](https://twitter.com/dog_rates/status/744234799360020481/video/1)),  
获得评分 13/10，转发数 77236，点赞数 128101；
- 排名第二的狗狗推特**是 2017 年 1 月 21 日用户（推特 ID:822872901745569793）发表的关于拉布拉多 puppo（名字不详）参加多伦多妇女游行的图片——>  
([https://twitter.com/dog\\_rates/status/822872901745569793/photo/1](https://twitter.com/dog_rates/status/822872901745569793/photo/1)),  
获得评分 13/10，转发数 49045，点赞数 143195；
- 排名第三的狗狗推特**是 2016 年 12 月 9 日用户（推特 ID:807106840509214720）发表的关

于吉娃娃 Stephan(地位不详)主动当作抹布帮主人拖地的视频——>

([https://twitter.com/dog\\_rates/status/807106840509214720/video/1](https://twitter.com/dog_rates/status/807106840509214720/video/1)),

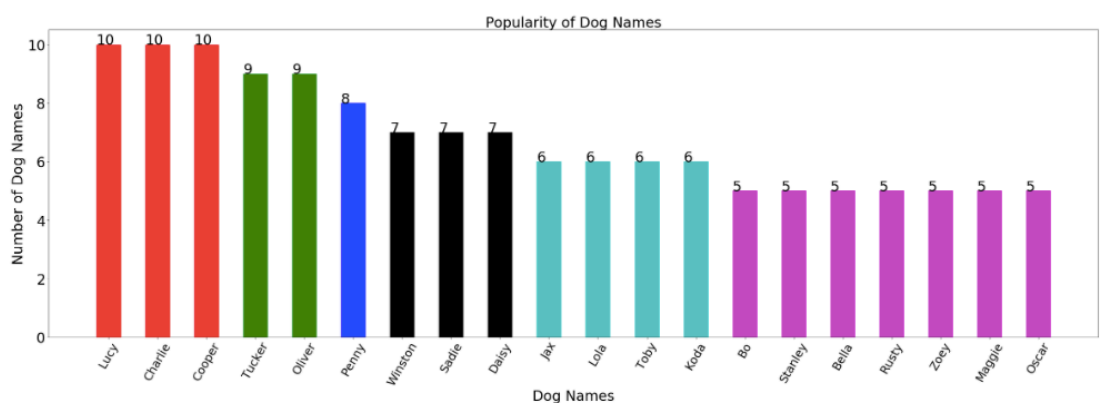
获得评分 13/10, 转发数 61007, 点赞数 122842.

B. 点赞和转发数与狗狗评分是否成正比关系? 点赞和转发数高的狗狗评分集中在什么区域?



由上图可知, 点赞和转发总数似乎并不是与狗狗评分成正比。虽然评分低 ( $<10/10$ ) 的狗狗获得转发和点赞数比评分高 ( $\geq 12/10$ ) 的狗狗少, 但并不是评分越高的狗狗获得的转发和点赞越多。可以发现评分为 13/10 的狗狗获得转发和点赞数最多。原因有可能是部分狗狗较高的评分 (例如  $\geq 14/10$ ) 太主观, 没有实至名归, 也可能从评分者的角度来说, 13/10 已经是很高的评分了, 获得此评分的狗狗里面有很多值得转发和点赞。

C. 最受欢迎的狗狗名字有哪些?



由上图可知, 最受欢迎的狗狗名字如下:

**并列第一名**有 3 个名字, 均在数据集中出现次数 10 次, 分别是 'Lucy', 'Charlie', 'Cooper';

**并列第二名**的有 2 个名字, 均在数据集中出现次数 9 次, 分别是 'Tucker', 'Oliver';

**第三名**是 'Penny', 在数据集中出现次数 8 次, ;

并列第四名的有 3 个名字，均在数据集中出现次数 7 次，分别是'Sadie','Winston','Daisy'；

并列第五名的有 4 个名字，均在数据集中出现次数 6 次，分别是'koda','Lola','Toby','Jax'；

并列第六名的有 7 个名字，均在数据集中出现次数 5 次，分别是

'Oscar','Zoey','Stanley','Bella','Maggie','Bo','Rusty'.

D. 哪个地位(stage)的狗狗更受欢迎与青睐？

Fig1. Retweet and favorite total counts for different dog stages

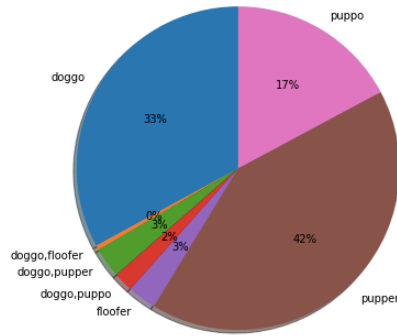
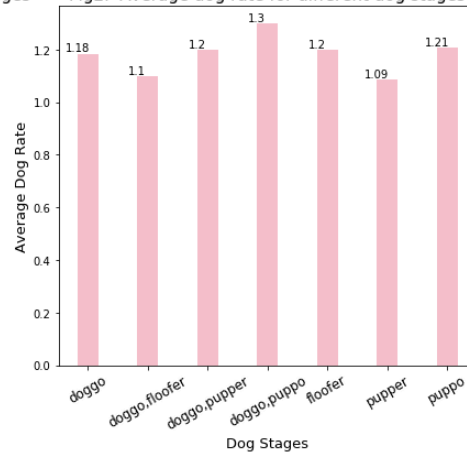


Fig2. Average dog rate for different dog stages



由上图可知，狗狗地位的转发和点赞总数占比依次为'pupper'(42%) > 'doggo'(33%) > 'puppo'(17%) > 'floofer'(3%) = pupper,doggo(3%) > puppo,doggo(2%)；

但是各个狗狗地位的平均评分相差不大，平均分最高分是 puppo,doggo 组合(13/10)，最低分是 pupper (11/10)。其中 doggo、floofer、pupper 和 doggo 组合、puppo 的平均得分均约等于 12/10；

综合来看，pupper 虽然得分不高但是较受欢迎，其次是 doggo。floofer 受欢迎程度低，可能并不是每个人都能接受皮毛过量的狗狗。各种组合的狗狗毕竟少数，所以得到的赞数和转发数有限，但是评分不低，狗多力量大啊！