

Remove Watermark Now

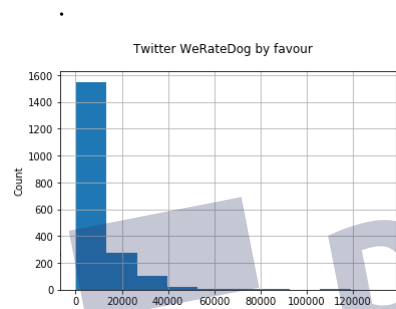
通过对推特用户 WeRateDogs的数据档案进行数据分析，探索宠物狗评分高的因素

看一下数据集中，用户点赞的统计：

总体的用户点赞统计如下：

count	1971.000000
mean	8977.030949
std	12456.033983
min	81.000000
25%	1992.000000
50%	4128.000000
75%	11367.000000
max	132318.000000

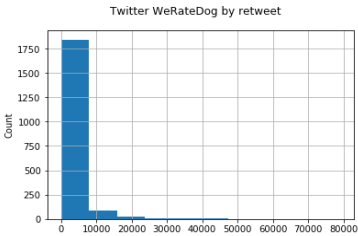
统计直方图：



总体的转发统计如下：

count	1971.000000
mean	2787.830036
std	4739.182501
min	15.000000
25%	627.000000
50%	1362.000000
75%	3224.500000
max	79116.000000

总体转发的直方图如图：



Remove Watermark Now

统计的1971条信息中，平均点赞数是8977，标准偏差std=12456，最小的点赞数是81，最大的是132318，由直方图可以看出，绝大多数的狗狗点赞数在20000以下；平均转发数是2787，标准偏差std=4739.18，同样的，绝大多数狗狗的转发数在10000以下。那么点赞数超两万的狗狗具备什么特征呢？它们的转发数是否一样居高。

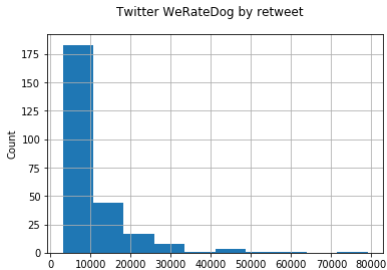
其中点赞数高于20000的推文统计如下：

count	259.000000
mean	33781.042471
std	17951.492512
min	20002.000000
25%	23087.000000
50%	28187.000000
75%	35844.000000
max	132318.000000

点赞数高于20000的推文转发量统计：

count	259.000000
mean	10534.332046
std	9345.093286
min	3099.000000
25%	5337.500000
50%	7555.000000
75%	11692.500000
max	79116.000000

统计直方图如图：



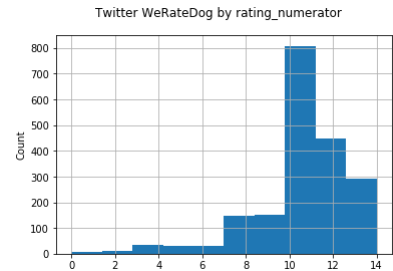
可以看出，点赞数高于20000的，相应的转发数也高，平均值为10534，标准偏差std=9345，最大值79116。

那么点赞高，转发量跟狗狗的评分是否成正比， WeRateDog的评分是否跟用户的喜好一致

分别对总体的评分和点赞数高于20000的推文的评分进行统计分析

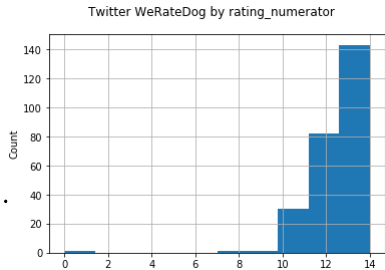
在分母一样的情况下，样本总体的评分统计如下：

count	1957.000000
mean	10.540511
std	2.182353
min	0.000000
25%	10.000000
50%	11.000000
75%	12.000000
max	14.000000



在分母一样的情况下，样本中点赞数超过20000的推文评分统计如下：

count	258.000000
mean	12.415698
std	1.242454
min	0.000000
25%	12.000000
50%	13.000000
75%	13.000000
max	14.000000



通过对1957个数据的评分进行统计，发现平均值为10.54，标准偏差std=2.18，评分在11左右分布，比较集中；

其中点赞数超过20000的狗狗的评分平均值为12.42，标准偏差std=1.24，频率分布直方图跟总体的一样，都是集中分布同一区间，相对点赞数跟转发数来说，这点差别并不明显。可见那么点赞高，转发量跟狗狗的评分不成正比，WeRateDog的评分跟用户的喜好没多大关系。

本次分析的结果是基于所收集的WeRateDog用户档案数据分析得出