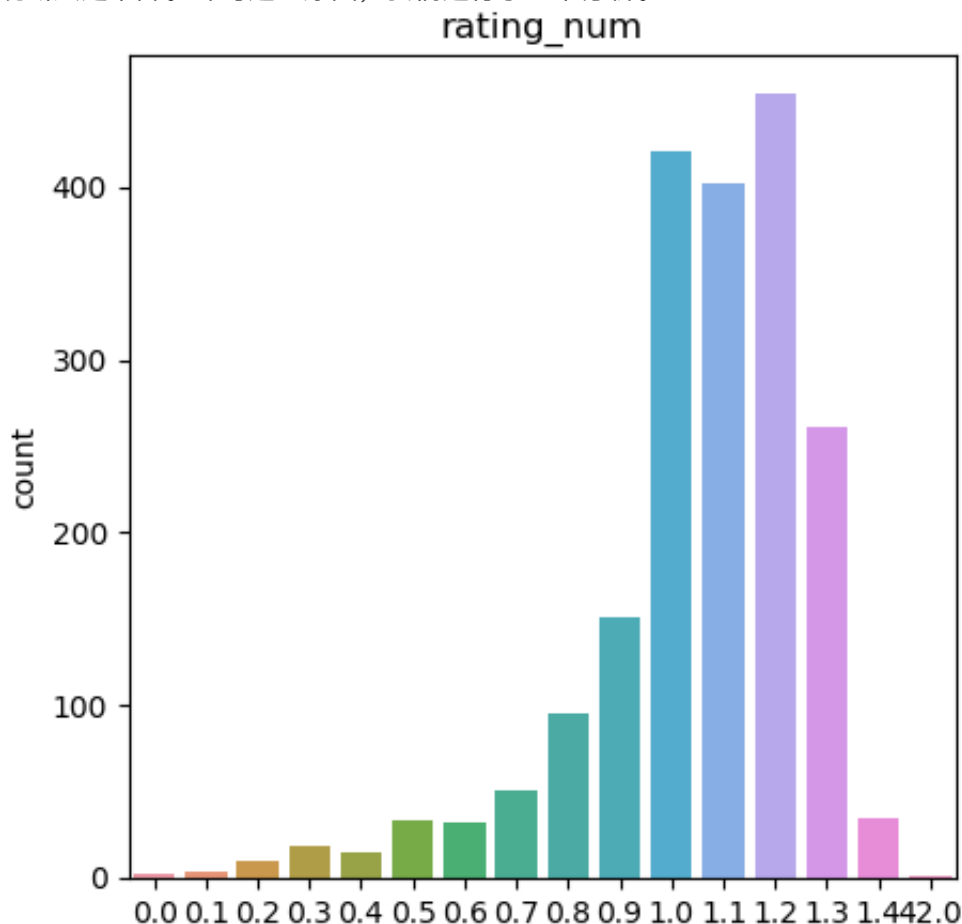


# 数据分析及结果

## 1. 该推特主对狗狗打分分布情况。

可以看出分数主要集中在1.0-1.3分之间，其中1.2分数量最多，整个分布情况符合正态分布。1.0分一下的数量出现断层是下降。针对这一原因，我们进行了一下分析。

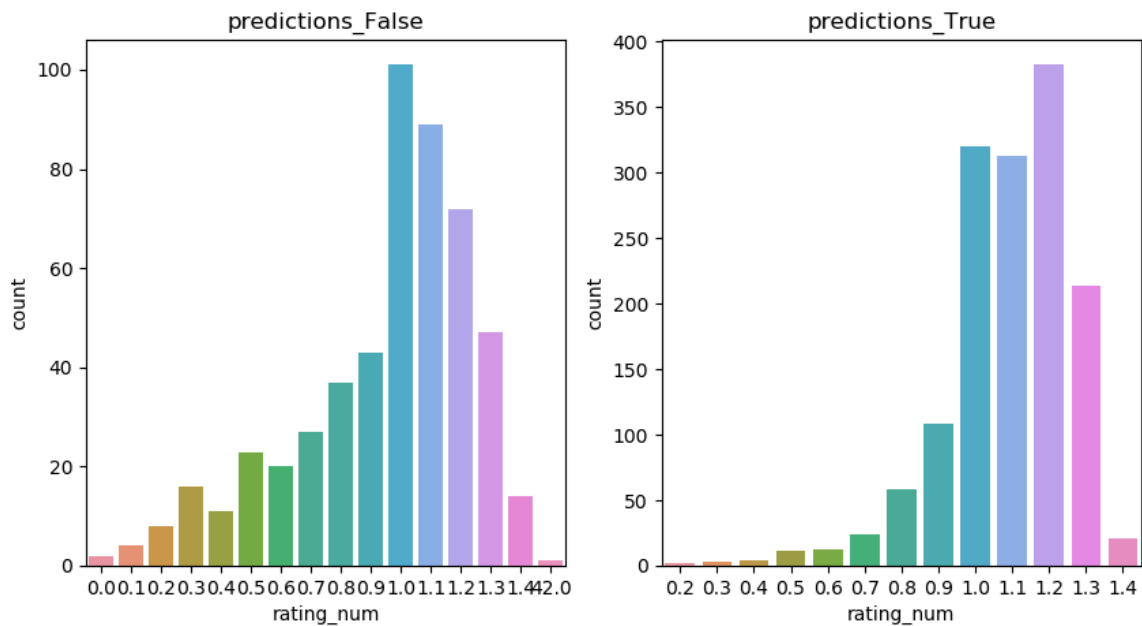


## 探索什么样的原因可能导致该推特主打更高的分数。

**猜测：**可能和图片拍摄的方式有关。能够更好识别狗狗的图片可能会被打更高的分数。

**分析：**通过对狗狗图片进行神经网络识别的结果和推特主打的分数来分析。将识别结果为True 和 False 的分别分成两组。对每一组统计推特主打分的分布情况。

首先对两组绘制了直方图进行观看，然后进行了描述统计分析。



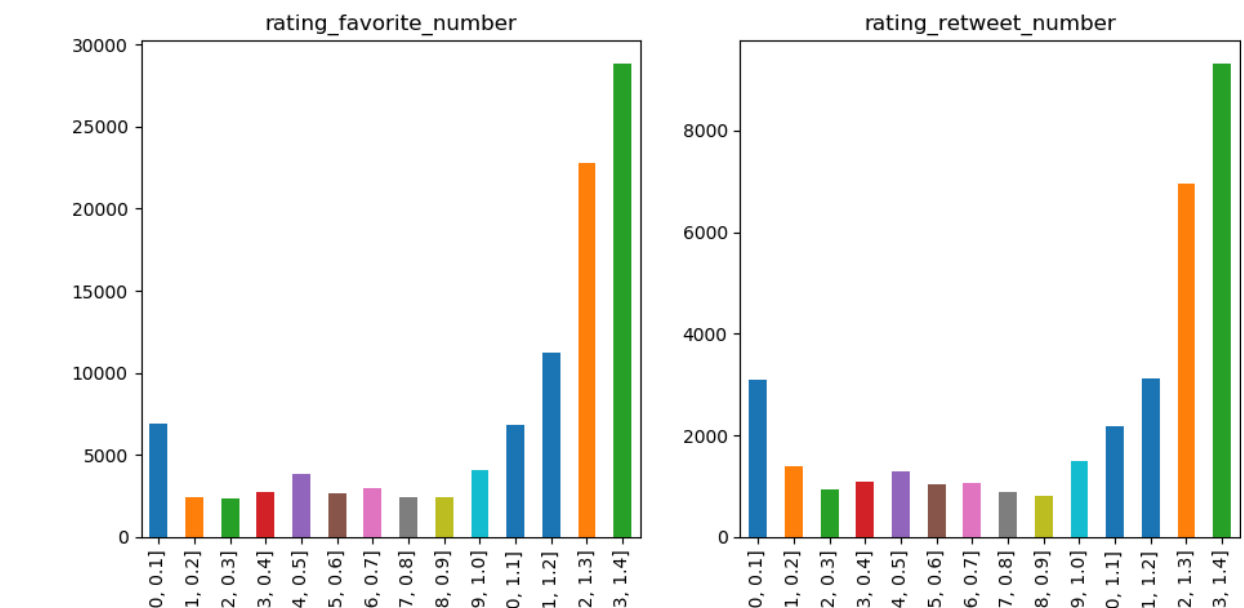
**结论：**神经网络识别为False的图片，给出的评分多集中在1.0分，平均数及中位数分别为1.02分和1.0分。神经网络识别为True的图片，给出的评分多集中在1.2分，平均数和中位数分别为1.09分和1.1分。查看对应的图片可以发现，未能识别的图片大多为离狗狗距离较远进行拍摄的，或者没有拍到狗狗的正脸。故如果想要获得高的分数，需要尽可能的拍清狗狗的正脸。

## 探索推特主的评分与点赞率及转发率的关系。

**猜测：**推特主评分主要集中在1.2分附近，该评分的狗狗应该也是大众比较喜爱的，所以点赞率与转发率最多的应该也集中在该评分附近。

**分析：**将该推特主的评分 rating\_numerator 列切割成15份。然后统计对应评分区间的点赞数 favorite\_count，转发数 retweet\_count 和 推特文章数，然后统计每个区间对应的推特文章数。用 点赞数/文章数 = 每篇文章点赞数，转发数/文章数 = 每篇文章转发数。

然后将结果绘制成直方图，可以看出无论是点赞率还是转发率都主要集中在1.3分和1.4分两列。

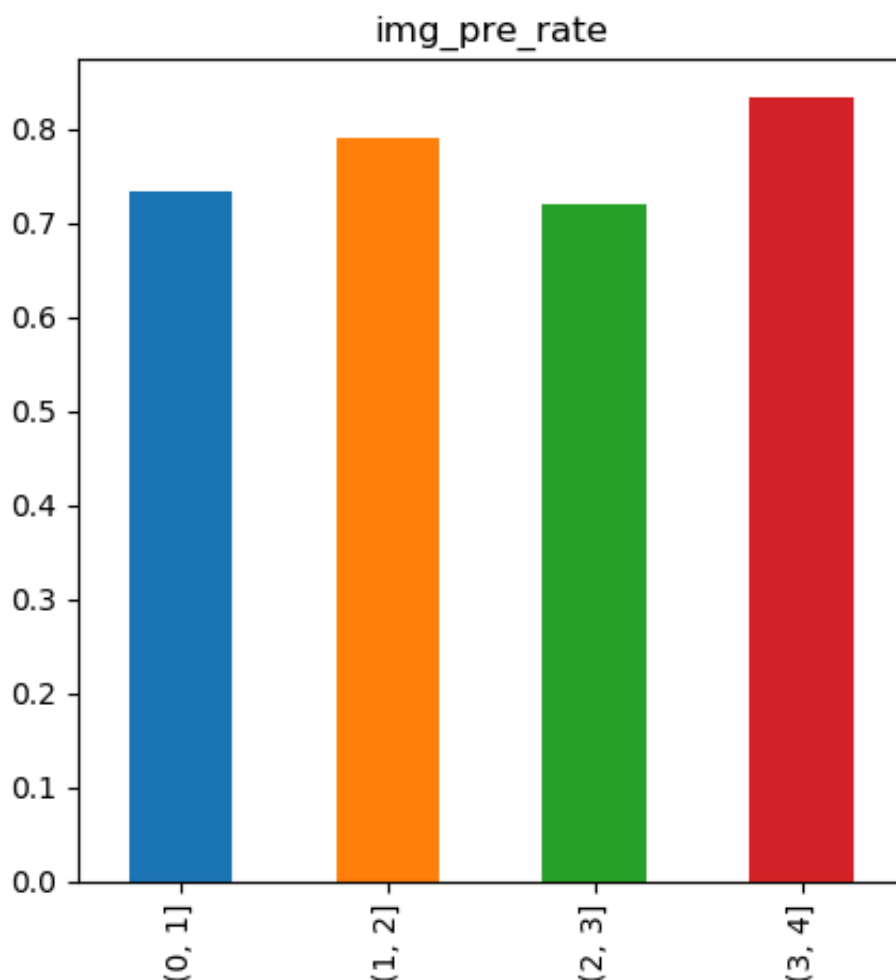


结论：大众更偏爱评分高的狗狗，随着评分越高，点赞数以及转发数也会更高。

## 神经网络识别成功率与文章中的图片数量的关系。

猜想：提供的图片越多，预测的准确率越高。

分析：由于 img\_num 列的最大值为4，也就是说最多只有4张照片的文章，所以将 img\_num 列分为4组，间隔为1。然后对每一组的 p1\_dog 列求和得出每组识别为 True 的数量,即正确识别数。然后将每一组的推特文章数量求和。用 正确识别数/每组推特文章数量=识别正确率。



结论：可以看出图片的数量与识别成功率没有太大的关系，可能主要还是与拍着的角度，距离等因素有关。

## 总结

本次分析的数据并不全面，并且狗狗图片的评分具有很强的主观性。该分析和结论存在一定的偶然性，并不具备相关性。仅供参考。