

分析与可视化

现在，将整理好的 `twitter_archive_master.csv` 导入

（推特档案数据 `twitter_archive_enhanced.csv`，推特图像的预测数据 `image-predictions.tsv`，和推特 API 附加数据合并后）

现在想探讨的问题是狗狗的受喜爱程度和 WeRateDog twitter 上评分之间的关系

从转发量 (`retweet_count`)，点赞数 (`favorite_count`)，评分分子(`rating_numerator`)这三列来进行分析和探讨，因为评分分母(`rating_denominator`)都是 10，不用考虑评分分母这个因素。

- 首先，我们筛出点赞数、转发量、评分分子这几列，新数据集命名 `favorite_retweet_rating`，将评分分子改名为评分。
- 再将数据集按评分排列

用 `seaborn` 的 `Swarmplot` 画出点赞数和评分，转发量和评分之间关系的散点图。

然后再用 `seaborn` 画出点赞数和评分，转发量和评分之间关系的直方图。

再用 `seaborn` 的 `Implot` 来看看是否有正向相关。

最后，再用箱形图了解它们各自的集中区域。

在分母为 10 的基础上，从以上第一个散点图来看，每个评分区间都有一定的转发量，点赞数和转发量最多的数量都集中在 10 到 14 分之间，点赞数明显大于转发量；

第二个直方图也可以得出和散点图相似的结论，但是在直方图中，两个直方图评分 65-76 之间都出现了较高的数量，可能是由于纵坐标区间较小造成。

在第三个图中，点赞数和转发量各自对于狗狗评分都呈正向相关，其中，点赞数和评分相关直线角度明显大于转发量和评分相关直线。

从箱形图中，我们更直观地看到点赞数和转发量各自的集中范围分别在 25000-50000，8000-22000 之间，这在前几个图中也可以很明显地看到。

总结

通过围绕狗狗的受喜爱程度和评分之间的关系进行可视化分析，我们很直观地可以知道，狗狗受喜爱程度和 WeRateDog twitter 上给出的评分呈一定的正向关系，评分在 10-14 之间

的狗狗最受人们欢迎，点赞数总量大于转发数总量；造成这个现象的原因可能是 WeRateDog twitter 的评分有一些不客观，人们无法自己判断有些狗狗的真实评分，而 WeRateDog 的评分又被广泛关注，它很容易就对喜欢看它推文的人造成影响。

如果没有受到 WeRateDog 的影响，结果可能是另一个样子。