

twitter 数据分析

三个问题:

1. 评分最高的狗名字是什么?

twitter_archive_enhanced.csv 包含狗的评分信息, 但是有一个问题, 使用的总分不一致, 这里用评分除以总分, 并新建列名为 score 存储这个分数, 用 score 作为狗的评分得到评分最高的狗的名字。由于打分比较无厘头, 分数不能代表狗的可爱漂亮等。打开评分最高的图片, 可能是狗的装扮比较受欢迎, 导致分数高。

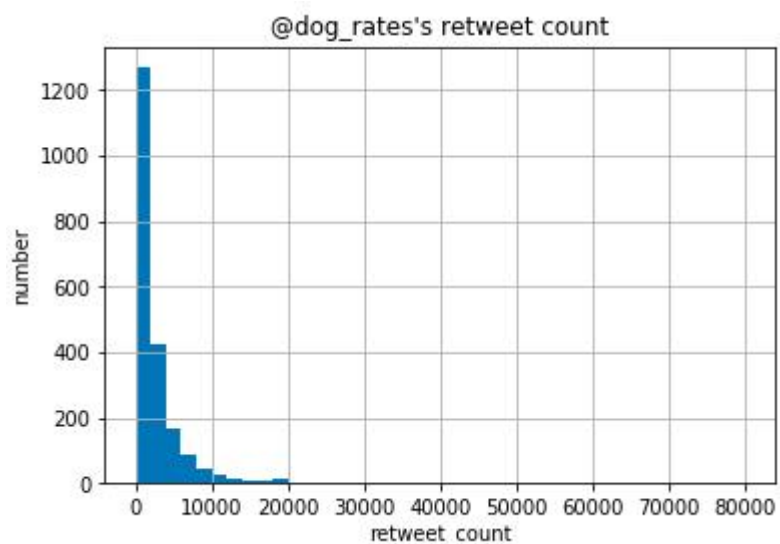
This is Atticus. He's quite simply America af.
1776/10



上午8:00 - 2016年7月4日

2. 转发的中位数是多少?

大多数帖子转发量在 0-2000, 但是也存在转发量非常大的将近 8 万, 平均转发量 2845



3. 点赞量和转发量是否有关系？

转发数据和点赞数从图中看来有强的相关性，相关系数是 0.913，如过帖子转发数多，那么点赞数据通常也多，这听上去也很合理。因为转发多可能是在互联网上火了的，看到的人多了，点赞就多了。

