

Twitter 推主 WeRateDogs 狗狗评分研究

V1.1

黄超 | Udacity | 2019.08.13

报告简介

如果你上过 Twitter，那么你可能看到过 WeRateDogs (@dog_rates) 推主。粉丝把他们的狗狗照片发到@dog_rates 上，然后，Matt Nelson，年仅 21 岁的推主，为他们的狗狗从 1-10 评分。事实上，很多狗评分高于 10 分，因为 Nelson 曾说，他们是美好的狗狗。

WeRateDogs 目前有 830 万粉丝，Nelson 结合了互联网上人们最喜欢的两个元素：狗和诙谐，这是他成功的基础。

- 报告基于 WeRateDogs 的 Twitter 数据，我们主要爬取了 Twitter 上的评分，地位，转发数，喜欢数等数据。
- 原始数据很混乱，我们对其进行了基本的清理，将数据整理为 Tidy Data。清理后的数据一共有 2171 条。
- 对整理后的数据进行分析 and 可视化，得出结论。

报告正文

报告主要从三个方面着手，分别是总体分布情况，地位研究以及相关性探索。

总体分布情况

这里统计了 1981 条 Twitter：

- **Rating (评分) 的分布情况：**评分中最高的有 1776，最低为 0 分（有一个负分的样本，但是不是对狗的评分）。平均得分为 11.65 分，中位数是 11，四分位分别是 10,12。值得注意的是，在 15 分及以上的评分非常少，统计了下一共才 2 个。
- **Favorite (喜欢数) 的分布情况：**喜欢数最高 132318，最低 52，平均为 8954，中位数是 4123。
- **Retweet (转发数) 的分布情况：**Twitter 转发最高 79116，最低 0，平均 2779，中位数为 1348。

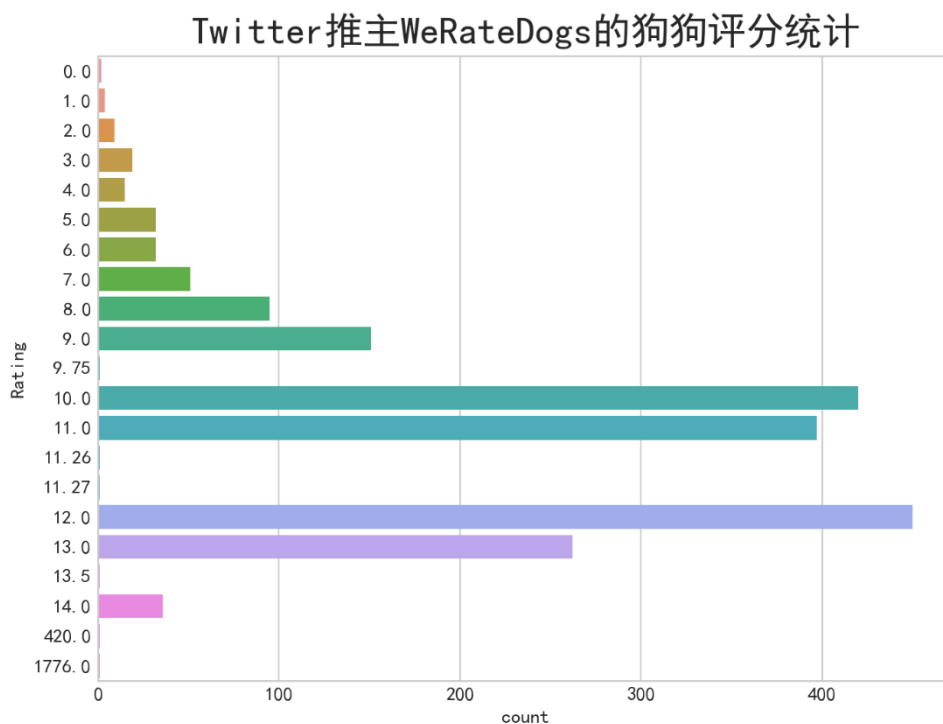


图 1

狗的 STAGE (地位) 研究

狗的称谓 (即 doggo、floof、pupper 和 puppo) 反映了狗的“地位”。样本中涉及到狗的 Stage 的 Twitter 一共有 337 条, 其中 pupper 占了大多数, 占到了 67.7。就评分来看, pupper 评分最低, 平均只有 10.82, 与其他三个地位 doggo (11.84)、floof (11.78)、puppo (12.04) 相差较多。二从 Retweet 和 Favorite 数量统计来看, 也和评分接近。Pupper 无论是喜欢还是 Favorite 都是最低的。其余

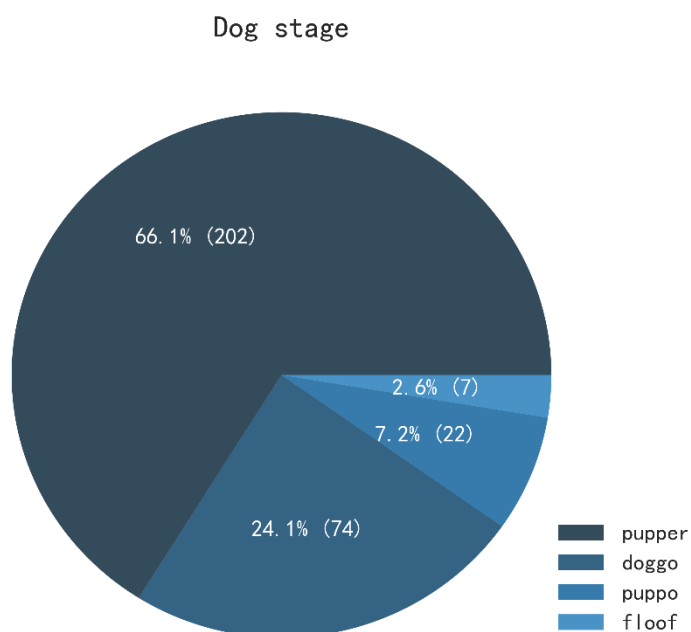


图 2

三个地位接近，但总的来说，puppo 最高，看来年轻充满活力的狗最惹人喜爱。

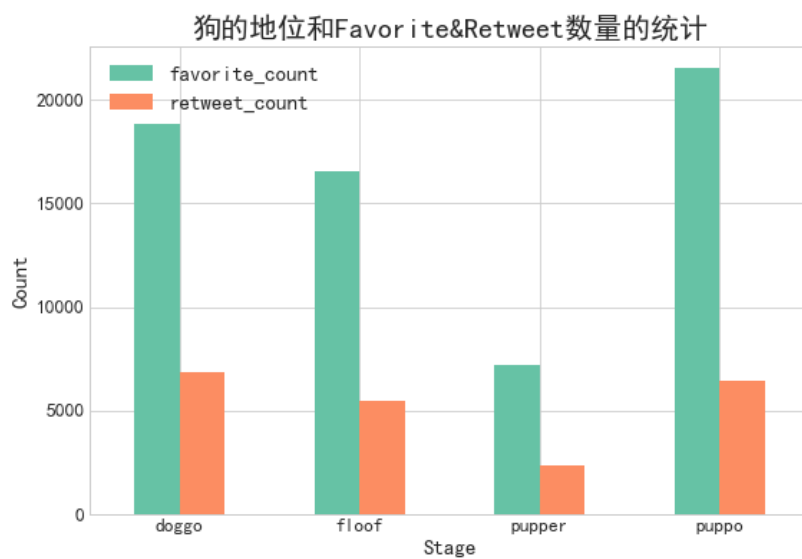


图 3

相关性探索

一般来说，Favorite 和 Retweet 存在正相关的关系。对于 WeRateDogs，更是如此。两者的 pearson 相关系数达到了 0.91，P 值为 0，有很强的相关性。

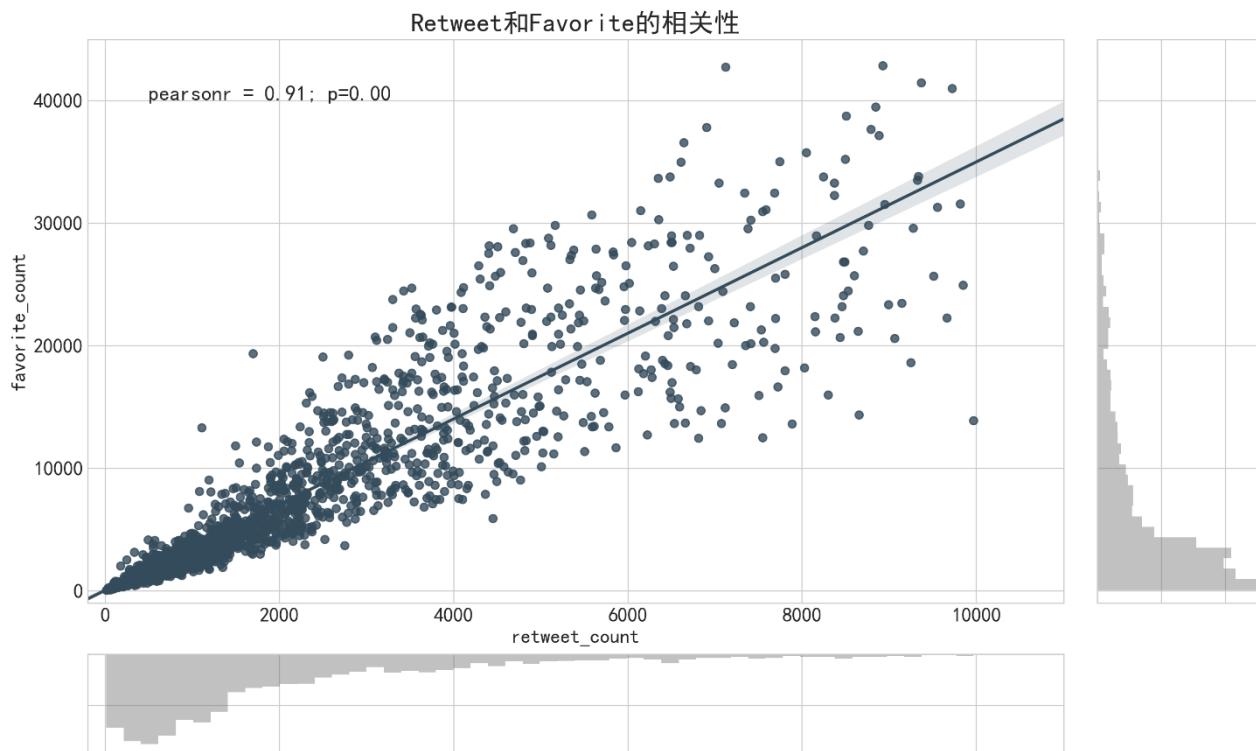


图 4

图 5 按照月份重采样取平均，可以看到，转发和喜欢数随着时间逐渐变大，因为 WeRateDogs 的粉丝越来越多，影响力也越来越大。所以研究评分和转发，喜欢数的关系，要考虑到时间的影响。这个是相对困难的，观察到 Retweet 数量随时间变化不是很明显，所以我们只研究 Retweet 数量和评分的关系。

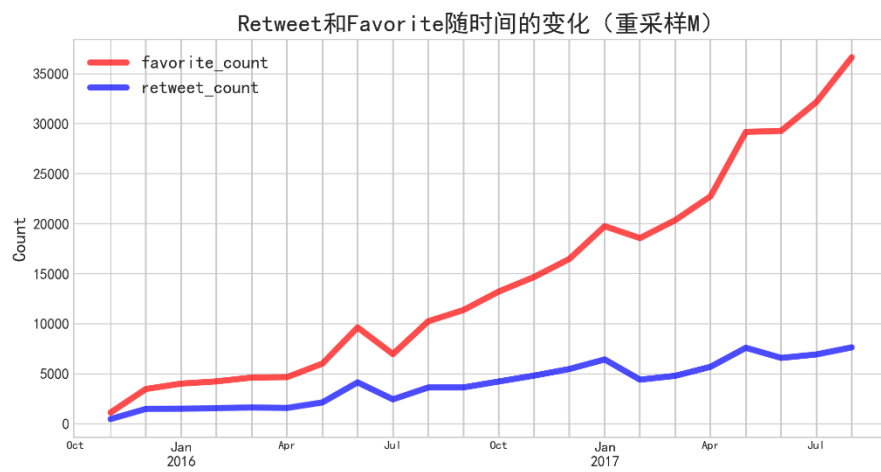


图 5

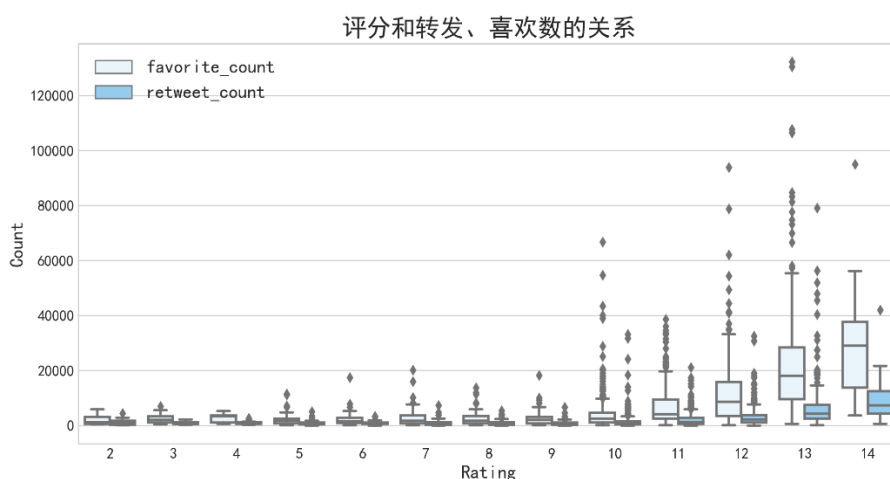


图 6

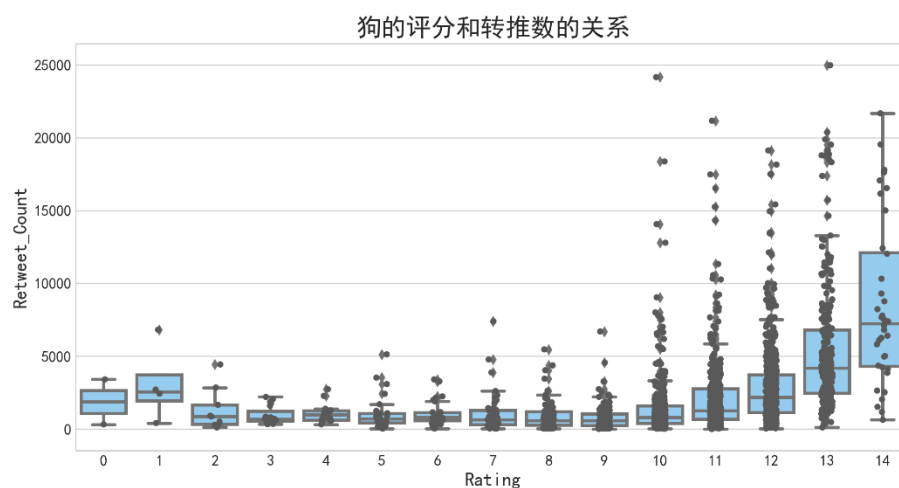


图 7

图 6 为评分和转发、喜欢数的关系，可以看到，评分越高，转发数和喜欢数的中位数越高(这里对小数的评分进行取整)。考虑到 Favorite 和 Retweet 的强正相关，这里我们选取随时间变化不明显的 Retweet 数来研究。图 7 显示，在评分 8-13 之间，狗的评分越高，Retweet 数量也越高。但是总体来看，趋势不是很明显，pearson 相关系数也只有 0.32。

彩蛋

想知道评分最高的狗是怎么样的么？

是的，他的名字叫 Atticus, 1776/10, amazing!



不过到目前为止，这个狗只有 2.5k 的 Retweets, 5.3k 的 Favorite =。=