

对萌宠推主 WeRateDogs 的档案分析

我们整理了推特昵称为 WeRateDogs 的推特用户 @dog_rates 的档案。WeRateDogs 是一个推特主，他以诙谐幽默的方式对人们的宠物狗评分。这些评分通常以 10 作为分母。但是分子则一般大于 10：11/10、12/10、13/10 等等。这个档案是基本的推特数据（推特 ID、时间戳、推特文本等），包含了截止到 2017 年 4 月 1 日的 5000 多条推特，我们筛选出了 2356 条包含评分的推特数据进行数据清洗和分析。

通过对数据评估，我们发现了 15 个质量问题和 3 个整洁度问题，通过一系列的数据清洗，我们得到了 2015 年 11 月-2017 年 8 月 1 日期间共 1992 条推文及其相关数据。通过分析，我们发现了以下现象：

1. Charlie 狗狗最常用的名字，Cooper、Oliver 等名字也很受欢迎。

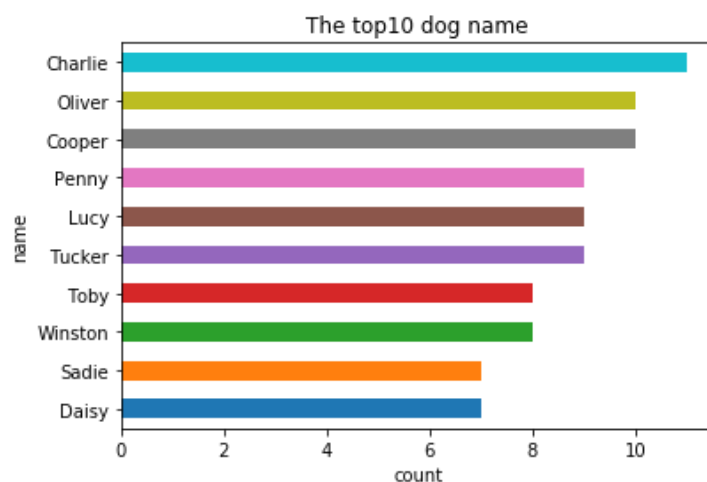


图 1. 狗狗名字受欢迎程度 top10

2. 我们将狗的评分分为 high、medium、low 三类，其评分高于等于 10 的为 high，低于等于 5 的为 low，然后绘制了转发数与评分的箱型图以及点赞量与评分的箱型图。不出意外，评分高的狗狗，其推文的转发量和点赞量最多。然而一个有趣现象是评分低狗狗推文的转发量和点赞量要高于中等评分的狗狗推文，这也许就是萝卜青菜各有所爱。

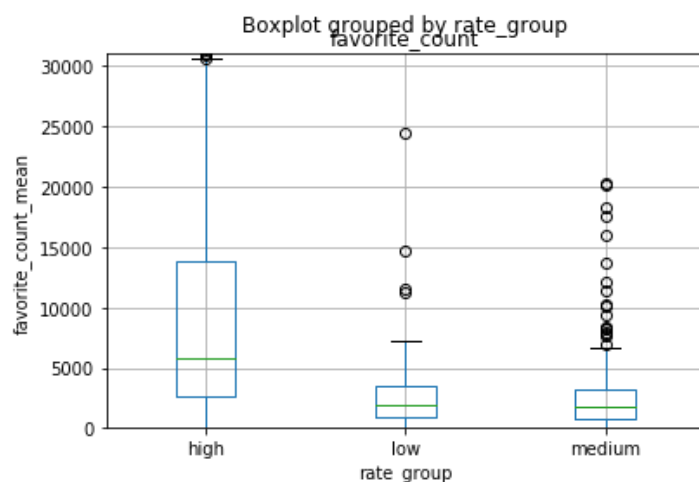


图 2. 点赞量和狗狗评分之间的关系

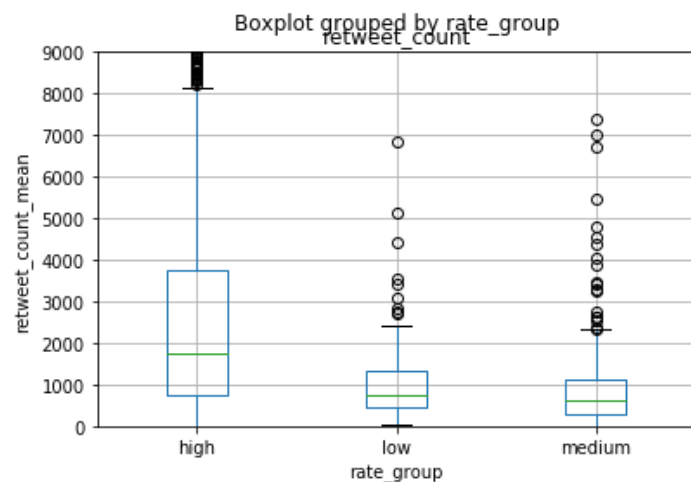


图 3. 转发量和狗狗评分之间的关系

- 我们运用神经网络，根据推文中的图片对狗的种类进行分类，得到一份图像预测结果。根据预测结果，我们对狗的种类的平均转发量和平均点赞量进行排序，可以看到 Bedlington terrier 品种的狗无论是平均转发量还是点赞量都排在前两位，Afghan hound、English springer、Saluki、French bulldog 在平均转发量前十排行榜和平均点赞量前十排行榜中均有其位。

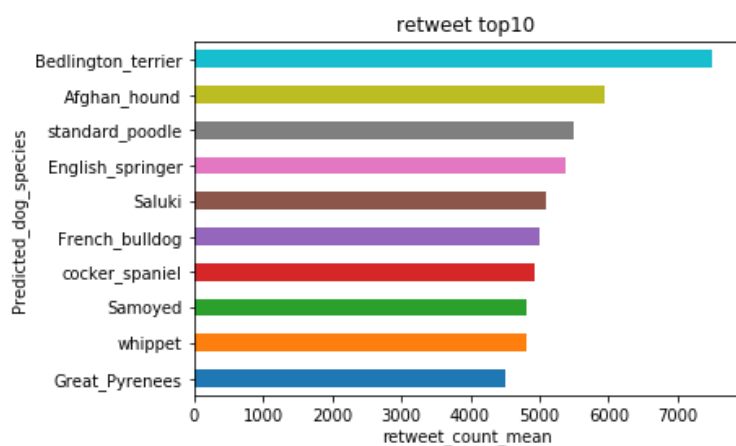


图 4. 平均转发量和狗狗的种类之间的关系

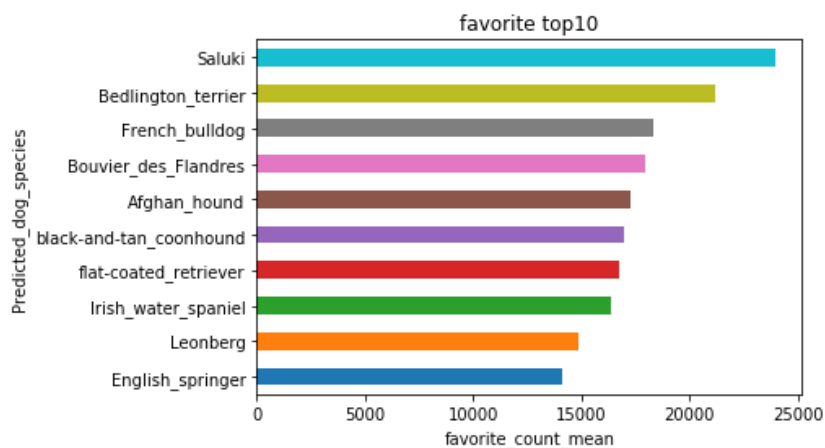


图 5. 平均点赞量和狗狗的种类之间的关系