

act_report

2020 年 4 月 27 日

1 推特狗狗分析 Act Report

2 项目背景

2.1 项目名称

推特狗狗评分数据集清洗与分析

2.2 项目内容

本项目需要基于推特狗狗评分的三个原始数据集整理出分析和生成可视化。数据集是推特用户用于狗狗评分的数据资料。

数据集包含数千条推特用户的基本信息，但信息并不完整，其中包含一些重复以及空列数据。在经过清洗和整理后，我从数千条数据里筛选出 2356 条相对完整适合分析的推特数据。

3 数据文件

- tweet_json.txt 收集的数据，用于修正原始数据。
- twitter-archive-enhanced.csv 原始数据
- image-predictions.tsv 图像预测结果
- 最终生成数据集的主要数据结构: 'tweet_id', 'rating_numerator', 'rating_denominator', 'name', 'retweet_count', 'fa

4 favorite 数据上涨的同时 retweet 数据同时也在上扬

- favorite 的数据上涨数量比 retweet 的数据上涨数量多。
- retweet 数据的上扬速度越后期越缓慢，与 favorite 数据列的差异越明显。
- 可以看出推特用户更愿意付出时间和用心去接触 favorite 的信息，相对需要维护 retweet 的时间就相应减少。

5 最受推特用户喜爱的 10 种狗狗

- 使用柱形可视化后前 10 最受喜爱的狗狗种类对比明显。
- 评分最高也是最受欢迎的狗狗种类是 golden retriever, 由此可以判断推特用户更喜欢性格温顺的中大型犬种。