

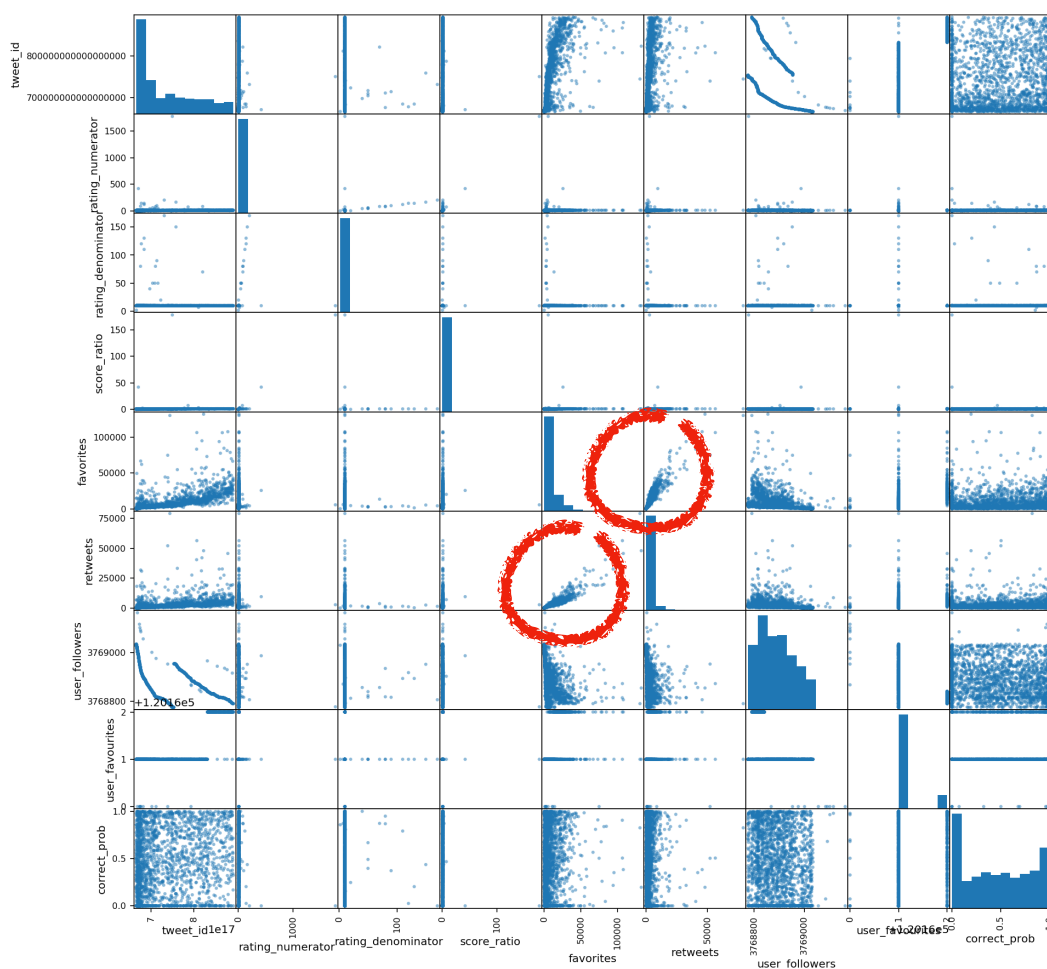
分析方向

根据前面的数据清理流程，发现数据有这些特点：

- 收集的数据跨度有2年
- 对于不同的tweet信息，有对应的点赞和转发信息
- tweet内容的质量（狗狗可爱不）也有一定的依据（有一个评分体系）

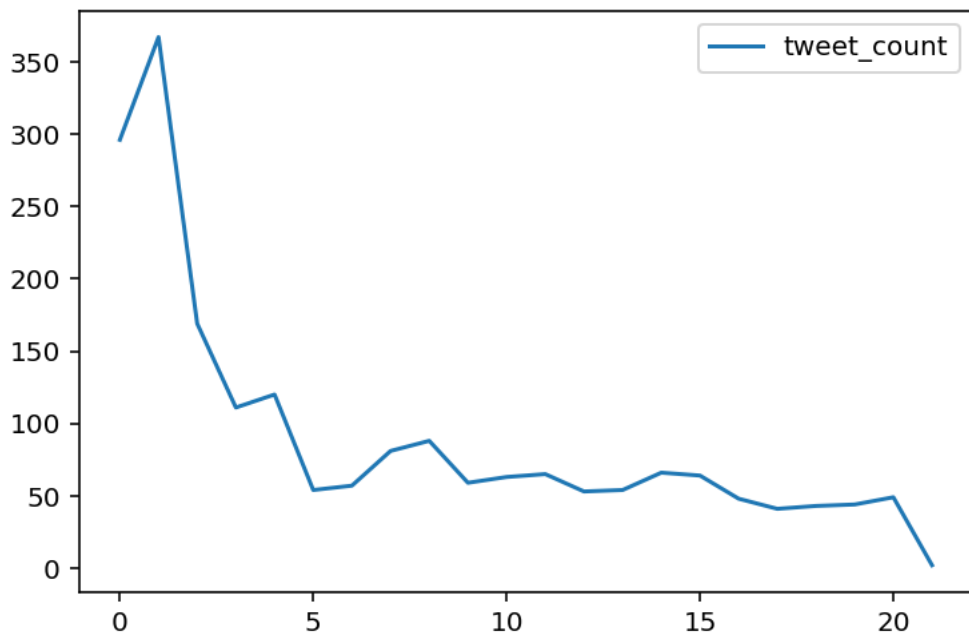
所以这里会深入的分析tweet内容与转发等之间的关系及得出一些猜想

1. 首先了解下数据之间的大概关系

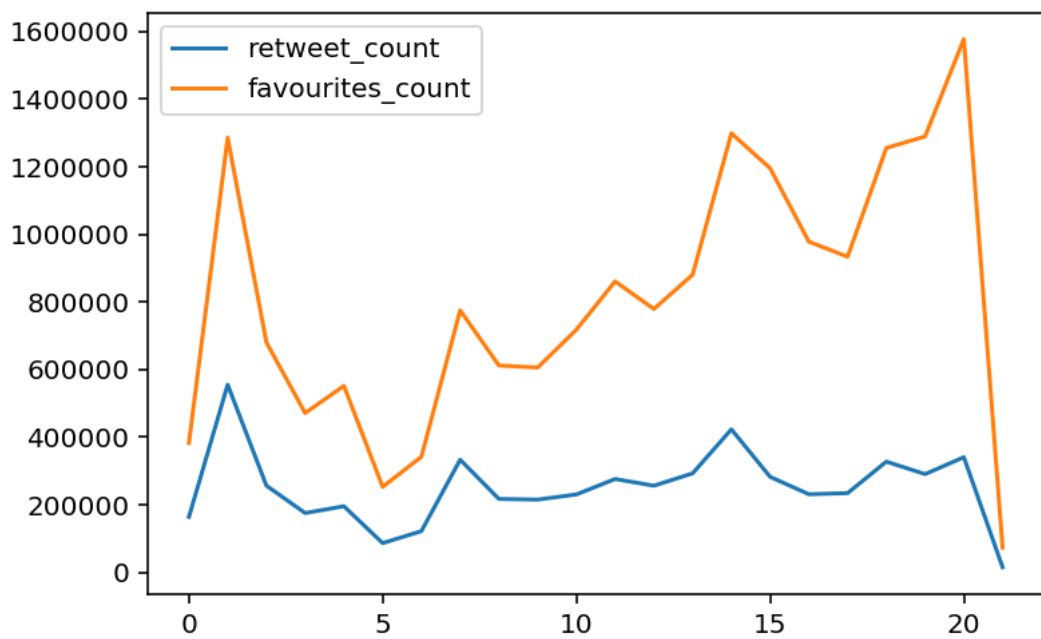


其中转发和点赞之间是存在相关性的。

2. 转发和点赞之间的关系¶



tweet在不同的月份发布的总数



tweet在不同的月份点赞和转发的数量

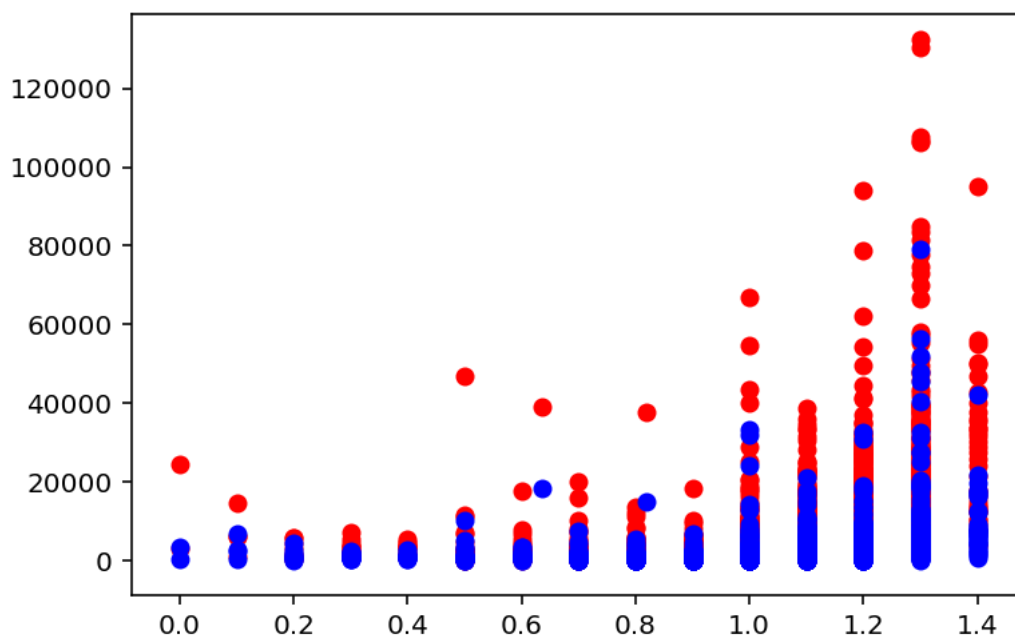
结论：

- 点赞数和转发数是正相关的
- 用户更喜欢点赞，转发的行为相对会弱一点
- 前期用户少时需要更多的运营，发送更多的内容，才能维持大家的喜爱
- 用户数达到临界值后，可以只进行日常的维护，来保证账号的持续产出

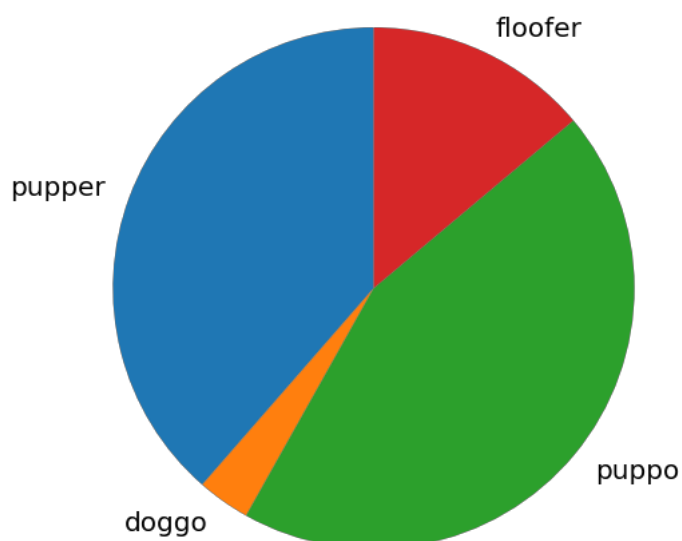
- “网络效应”通过这个例子得到证明

3. 转发（点赞）和内容质量的关系

为了图像的可读性，将超出范围的评分先去除掉。



分享数量和狗狗的评分的关系



狗狗不同种类和分享的总次数饼图

结论：

- 当狗狗的评分相对高一点时，对应tweet的转发和点赞数普遍会高很多
- 狗狗评分中，pupper和puppo类型的狗狗会得到更多的转发及点赞

特别的地方：

分数最高的两张"狗狗"的照片，可以看到用户对有创意的图情有独钟



4. 猜想

tweets_json.txt由于使用的是udacity提供的内容，最多是一天之内就准备好了。但是在获取用户的关注数时，就增加了387个粉丝。

当时是2017-08-01，有3769163的粉丝。

现在是2018-12-15，有7510000的粉丝。

平均每天需要增长7959个用户，再次证明了互联网头部资源的“网络效应”