

探索推特用户 WeRateDogs 对宠物狗的评分

一、项目背景

整理推特用户 WeRateDogs 对宠物狗评分，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。为什么会有这样的评分？因为 “[They’re good dogs Brent.](#)” WeRateDogs 拥有四百多万关注者，曾受到国际媒体的报道。

二、分析问题

a) Q1-哪种体型的宠物狗最受欢迎？

回答这个问题之前，首先要了解 doggo、floofer、pupper 是描述宠物狗的哪种属性，根据我在因特网上查阅的资料，

(<https://www.youtube.com/watch?reload=9&v=ah6fmNEtXFI>)(需翻墙)，doggo、floofer、pupper 为描述狗体型的单词，其中体型由大到小排列为 floofer、doggo、pupper。

确认宠物狗体型描述之后，通过对数据集的探索性数据分析，得到如下内容：

```
pupper      220
doggo       66
floof       33
puppo       27
pupper,doggo 9
doggo,puppo 2
doggo,floof 2
pupper,floof 1
Name: dogtype, dtype: int64
```

在 360 条数据中，pupper 体型的宠物狗，最受欢迎。问题是，pupper 到底是什么体型的狗，见下图：



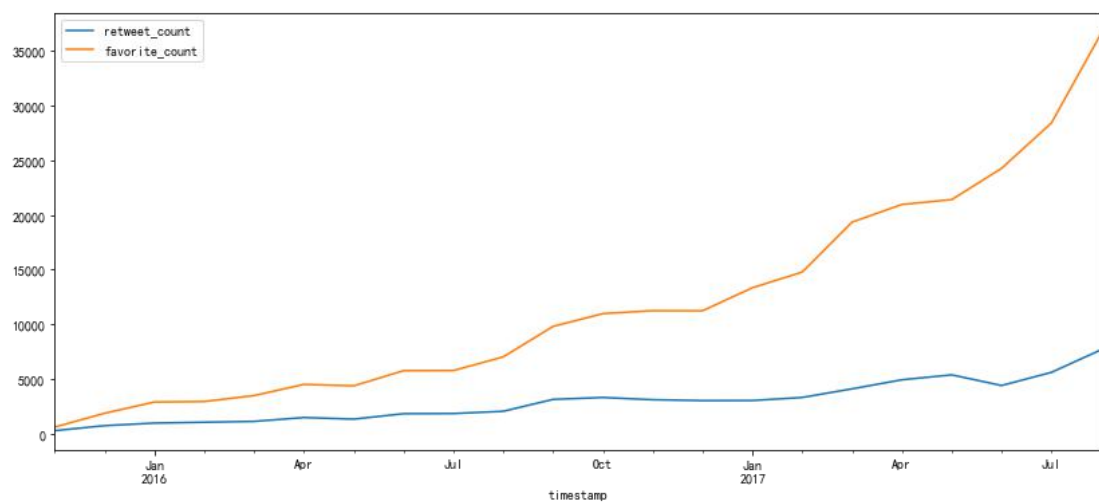
b) Q2-哪种品种的宠物狗获得较高评分？

通过分析数据集，排除异常数据后，得到如下内容：

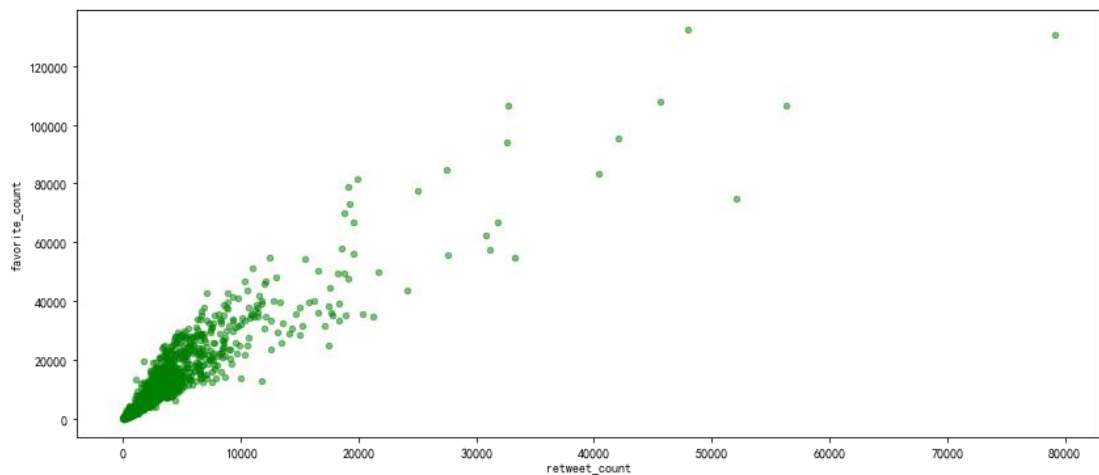
rating_numerator		p1
501	75	pomeranian
551	27	clumber
1362	26	kuvasz
49	14	samoyed
36	14	pembroke
278	14	bedlington_terrier
146	14	old_english_sheepdog
147	14	chihuahua
547	14	pembroke
297	14	rottweiler

根据评分分子排序后，得分前三名的狗分别为 pomeranian（博美犬）、clumber（克伦博猎犬）、kuvasz（匈牙利库维斯犬）。

c) Q3-宠物狗推特转发量和喜爱量之间的关系
以发推时间为索引排序后，得到下图：



2016 年 4 月以前喜爱量是转推量的两倍，之后喜爱量的增长速度明显高于转推量。
使用散点图获得转推量与喜爱量的相关性结论



转推数量在 0-10000 之间，转推数量与喜爱数量的相关性很强，转推数量超过 20000 后相关性越来越弱。

三、推测

1.根据以上数据分析结论，我推测大部分人更喜欢体型小巧的宠物狗。

2.2016 年 4 月以前喜爱量是转推量的两倍，之后喜爱量的增长速度明显高于转推量，推测导致这种现象的原因是喜爱比转推操作上更加方便，再加上推特被用户所接受，用户大量增长。

3.转推数量在 0-10000 之间，转推数量与喜爱数量的相关性很强，转推数量超过 20000 后相关性越来越弱，推测是 0-10000 转推量所占比例比 10000-80000 多的多，导致之后的散点越来越稀疏，毕竟超过 10000 的转推量与喜爱数量确实比较困难，不是一般的宠物狗获得不了这么高的浏览量。