act_report

简介

本次项目选取推特昵称为 WeRateDogs 的档案为数据集进行数据清洗分析和可视化。因为这份档案只包含基础信息，还需要另外收集相关数据，一起进行评估和清洗，得出结论。

通过对 twitter_archive_enhanced.csv、tweet_json.txt、image_predicitions.tsv 这三个数据集进行整理合并，将其保存在 twitter_archive_master.csv 数据集里面。对这个数据集进行分析可视化。

分析

问题 1： favorite_count 点赞数和 retweet_count 转发数表现如何？

```
df.favorite_count.describe()
```

```
count      1994.000000
mean       8923.133400
std       12400.238808
min          81.000000
25%        1972.250000
50%        4117.000000
75%       11275.500000
max      132318.000000
Name: favorite_count, dtype: float64
```

```
df.retweet_count.describe()
```

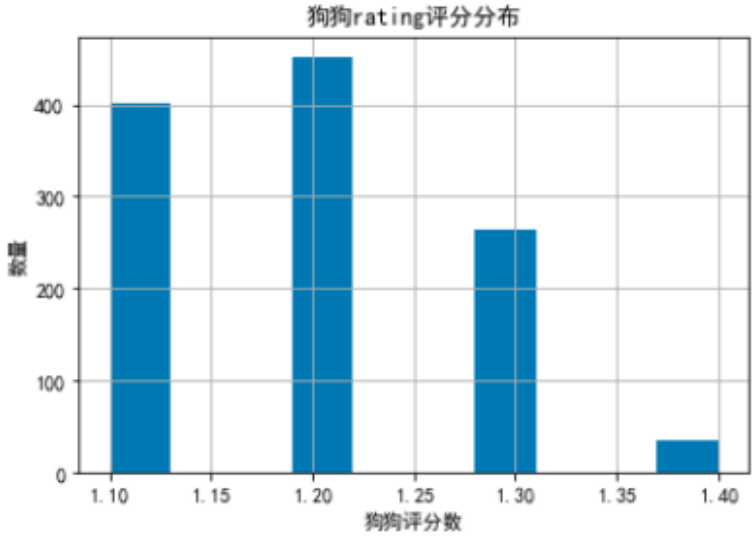
```
count     1994.000000
mean      2770.021063
std       4715.961325
min         15.000000
25%        622.250000
50%       1348.500000
75%       3202.750000
max      79116.000000
Name: retweet_count, dtype: float64
```

发现

favorite_count 点赞数和 retweet_count 转发数不一致。

- favorite_count 中位数大约是 4117，retweet_count 中位数大约是 1348。
- favorite_count 最大值大约是 132318，retweet_count 中位数大约是 79116。
- 说明人们相对于转发，可能更偏爱点赞。

问题 2：狗狗 rating 评分主要集中在哪里？
df 生成新的一列 rating,用 rating_numerator 除以 rating_denominator 表示。



发现

狗狗的 rating 评分大部分集中在 1.2,评分，1.4 已经比较少了。

问题 3：狗狗 stage 地位主要集中在哪里？

```
df.stage.value_counts()
```

```
pupper            221
doggo              66
puppo              26
pupper, doggo       8
floofer             3
puppo, doggo        2
Name: stage, dtype: int64
```

狗狗 stage 地位主要集中在 pupper。

结论

- favorite_count 点赞数和 retweet_count 转发数不一致，相对于转发，人们更喜欢点赞。
- 狗狗的 rating 评分大部分集中在 1.2,评分 1.4 已经比较少了。
- 狗狗 stage 地位主要集中在 pupper。