

act-report

2018 年 12 月 25 日

0.1 探索数据

0.1.1 问题

- 1 喜欢数和转发数和最高的前 5 种狗的品种?
- 2 评分数和最高的前 5 种狗的品种?
- 3 狗的地位中喜欢数和转发数和的排序?

0.1.2 数据整理

- 1) 为 dog_cl 创建 count 列为 favorite_count 列和 retweet_count 列之和
- 2) 为 dog_cl 创建 rating 列为 rating_numerator 列和 rating_denominator 列之比
- 3) 保留 “rating” 列, “stage” 列, “count” 列, “p1” 列, “p1_conf” 列, “p1_dog” 列, 并创建 dog_val
- 4) 检查结果

0.1.3 数据可视化

- 1) 导入可视化工具包

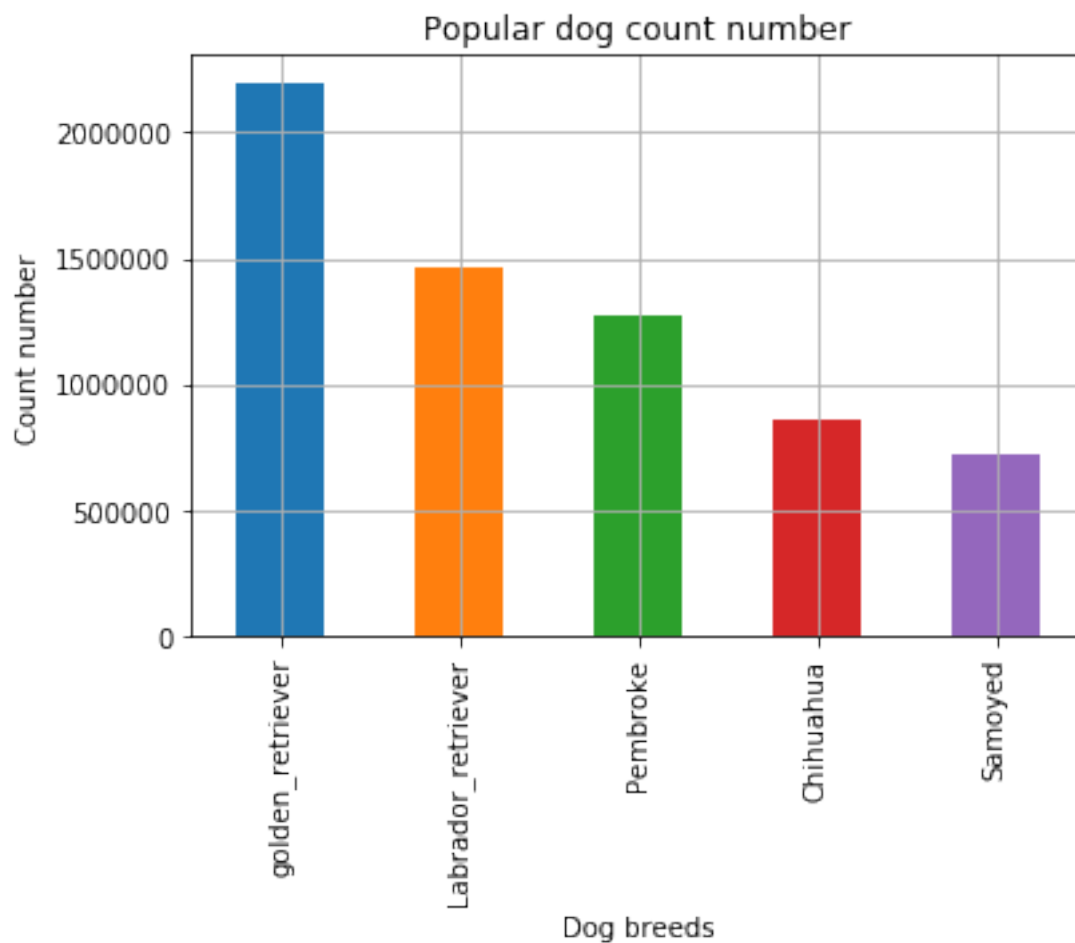
问题 1 1) 喜欢数和转发数和最高的前 5 种狗的品种?

- 1) 打印统计结果
- 2) 可视化结果
- 3) 设置可视化图标签及标题

count	5.000000e+00
mean	1.303086e+06
std	5.847181e+05
min	7.252420e+05
25%	8.532420e+05
50%	1.270737e+06
75%	1.466747e+06
max	2.199460e+06

Name: count, dtype: float64

Out[466]: Text(0.5, 1.0, 'Popular dog count number')



结论 1: 喜欢数和转发数和最高的前 5 种狗的品种从 到低依次为 golden_retriever、Labrador_retriever、Pembroke、Chihuahua、Samoyed。

问题 2 2) 评分数和最高的前 5 种狗的品种?

- 1) 打印统计结果
- 2) 可视化结果
- 3) 设置可视化图标签及标题

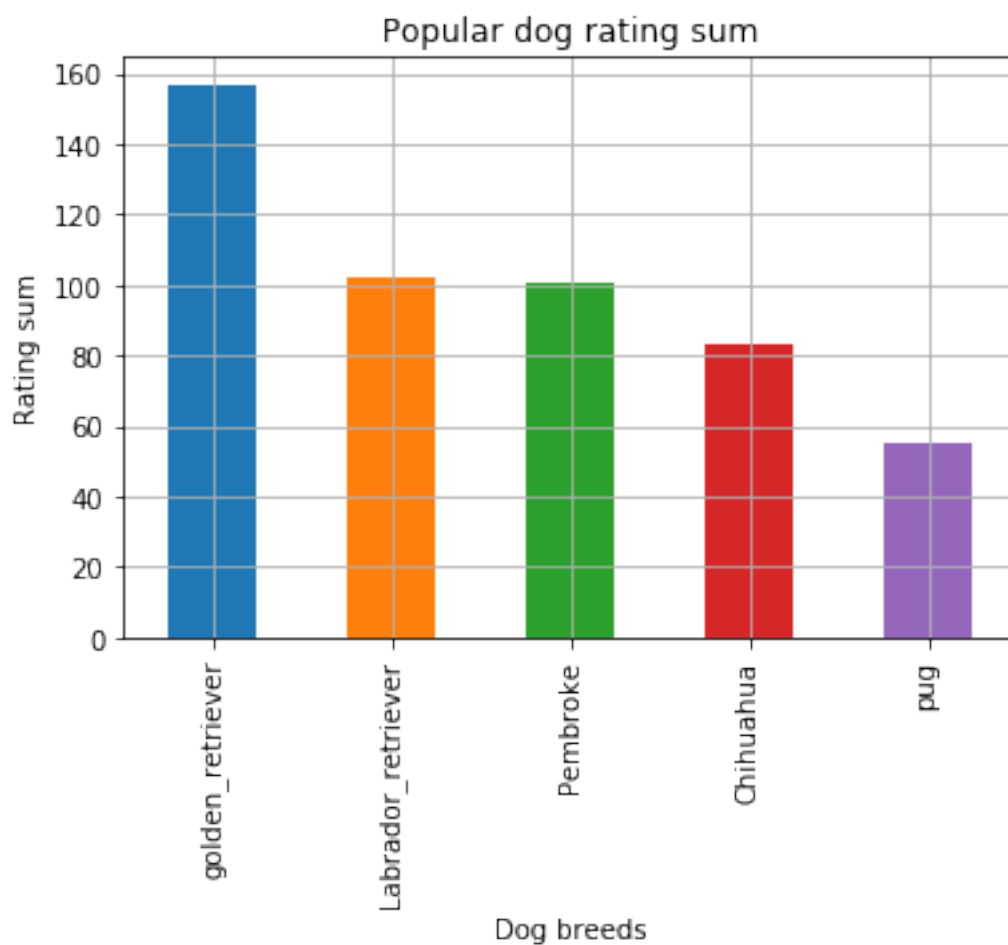
count 5.000000
mean 99.723636

```

std      37.156855
min      55.300000
25%      83.300000
50%     100.700000
75%     102.400000
max     156.918182
Name: rating, dtype: float64

```

Out[467]: Text(0.5, 1.0, 'Popular dog rating sum')



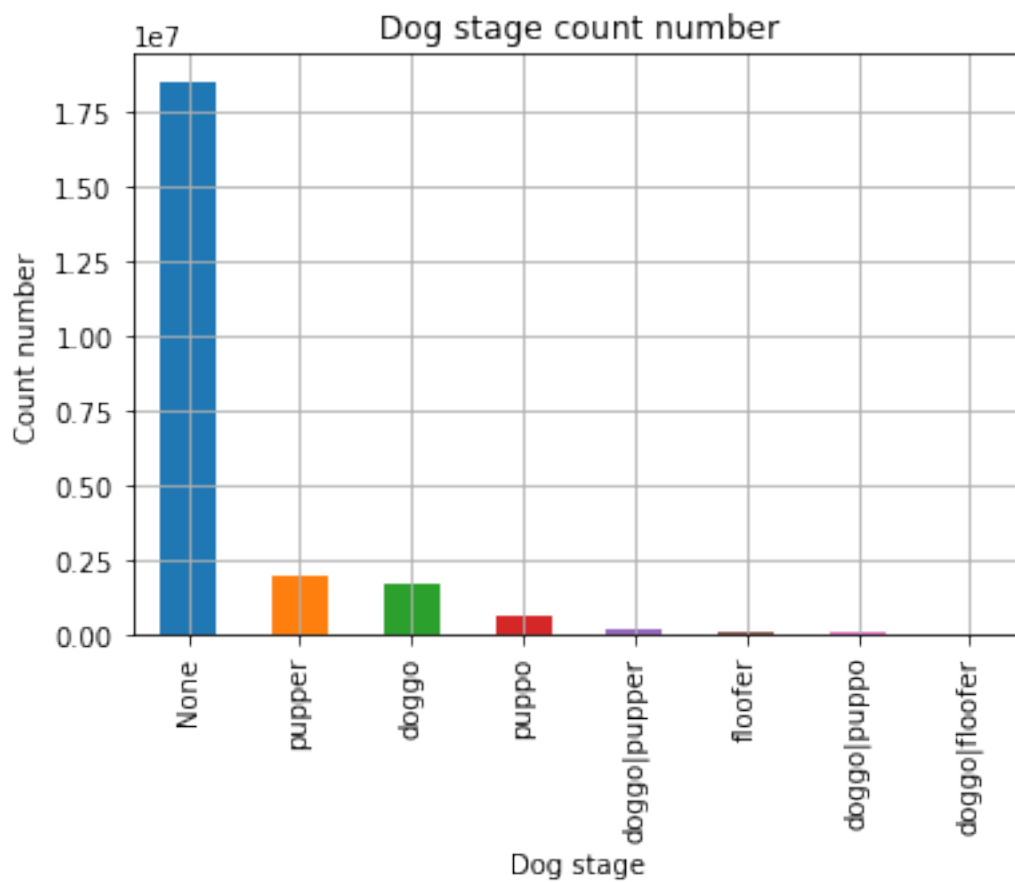
结论 2： 评分数和最高的前 5 种狗的品种从 到低依次为 golden_retriever、Labrador_retriever、Pembroke、Chihuahua、Pug。

问题 3 3) 狗的地位中喜欢数和转发数和的排序?

- 1) 打印统计结果
- 2) 可视化结果
- 3) 设置可视化图标签及标题

```
count      8.000000e+00
mean       2.882755e+06
std        6.353587e+06
min        2.053900e+04
25%        1.116815e+05
50%        3.818035e+05
75%        1.727383e+06
max        1.849787e+07
Name: count, dtype: float64
```

Out[468]: Text(0.5, 1.0, 'Dog stage count number')



结论 3: 除去未标明狗的地位, 标明狗地位中喜欢数和转发数和从 到低依次为 pupper、doggo、puppo、doggo|pupper、floofer、doggo|puppo、doggo|floofer。