

最终数据观察分析

提出问题

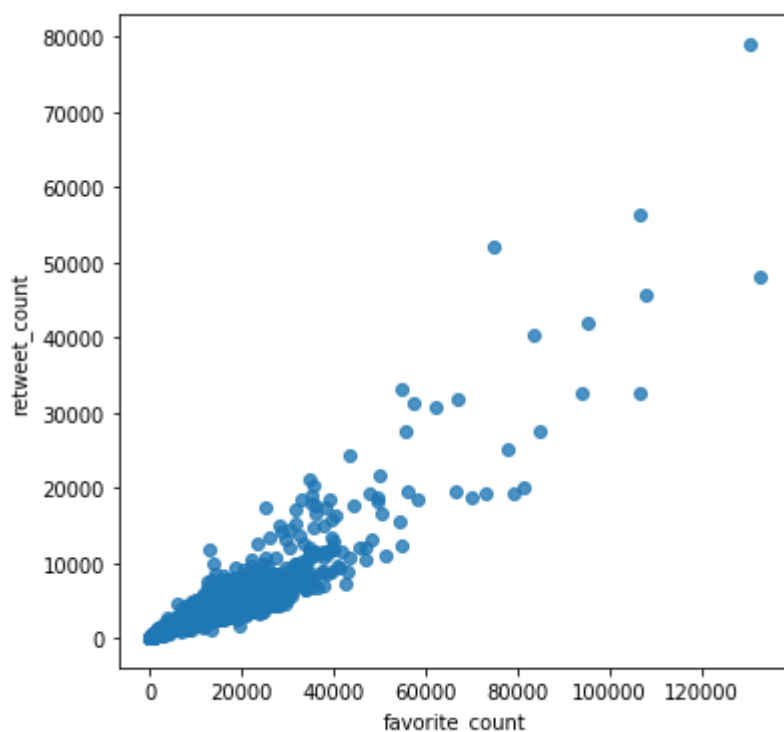
加载进行完清理的最终数据，首先提出三个问题进行分析：

- 狗的点赞数和转发数有什么关系吗？
- 不同成长阶段的狗的评分相同吗？
- 那个品种的狗更受欢迎？

分析和可视化

问题1：狗的点赞数和转发数有什么关系吗？

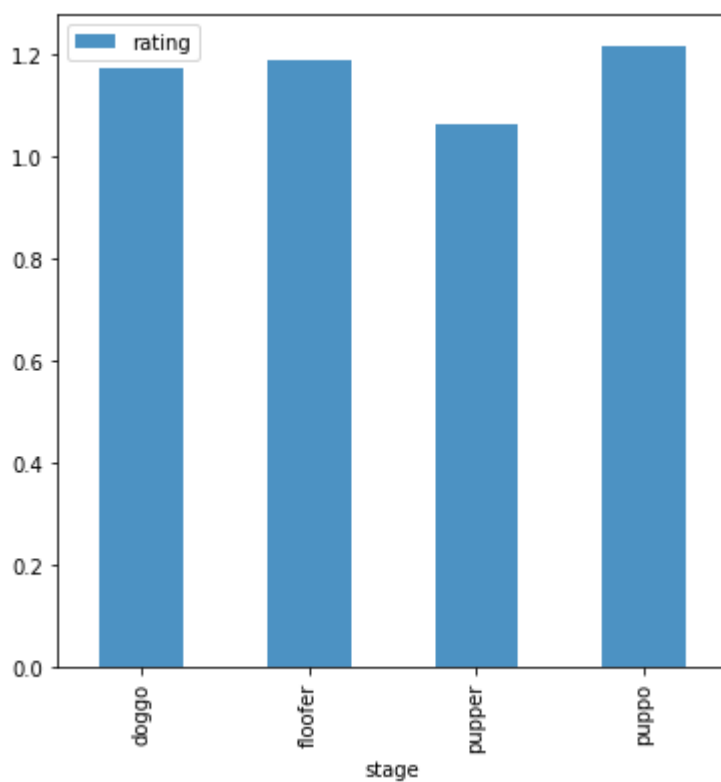
提取数据集中的点赞数和转发数，绘制散点图如下：



从图中可以观察到，点赞数favorite_count和转发数retweet_count有很强的正相关关系，favorite_count越高，retweet_count越大。

问题2：不同成长阶段的狗的评分相同吗？

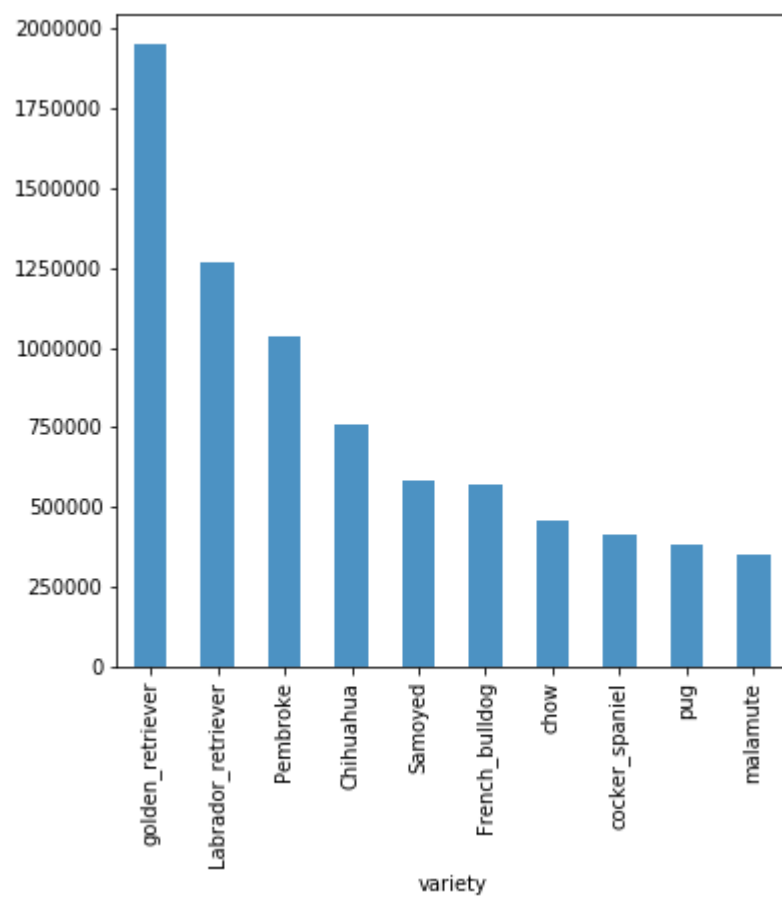
提取狗成长阶段的数据以及评分数据，将包含多个阶段数据的数据进行处理，并使用groupby将评分按照不同阶段分组，求不同阶段狗评分的平均分，绘制条形图：



从上条形图可以看出，所有阶段的狗狗评分很接近，都在1.1-1.2左右，处于'puppero'阶段的狗狗的评分均分最高，其次是'doggo'阶段的狗狗，处于'pupper'阶段的狗狗评分均分最低

问题3：那些品种的狗更受欢迎？

提取所有品种预测数据和点赞数，对预测数据进行处理，提取其中狗的品种数据，使用groupby对数据进行分组，计算不同品种点赞数总和，排序后取前十名绘制条形图：



前十名的数据为：

```
variety
golden_retriever    1950826
Labrador_retriever  1269771
Pembroke            1036321
Chihuahua           756715
Samoyed             582082
French_bulldog      568978
chow                456699
cocker_spaniel      413968
pug                 382463
malamute            350710
```

条形图和数据集可以看出，品种为golden_retriever的狗狗点赞数最多，其次是品种为Pembroke的狗狗。