

Twitter 数据观察与分析

现拥有 WeRateDogs 推特档案的清洗后的数据集，其中包含狗狗名字、地位、评分、点赞转发数、品种预测等数据。通过观察数据，拟探索问题如下：

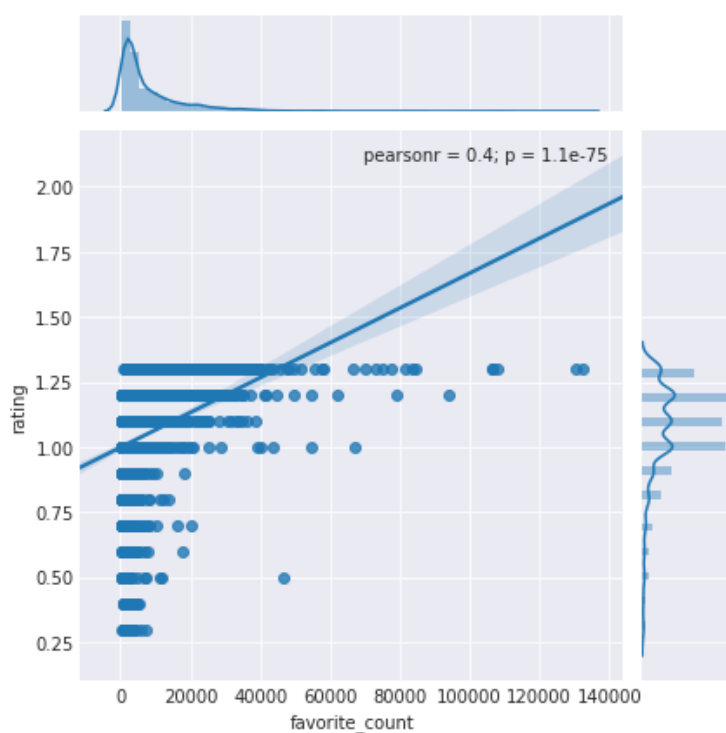
①retweet_count、favorite_count 与评分 rating 有无相关性？

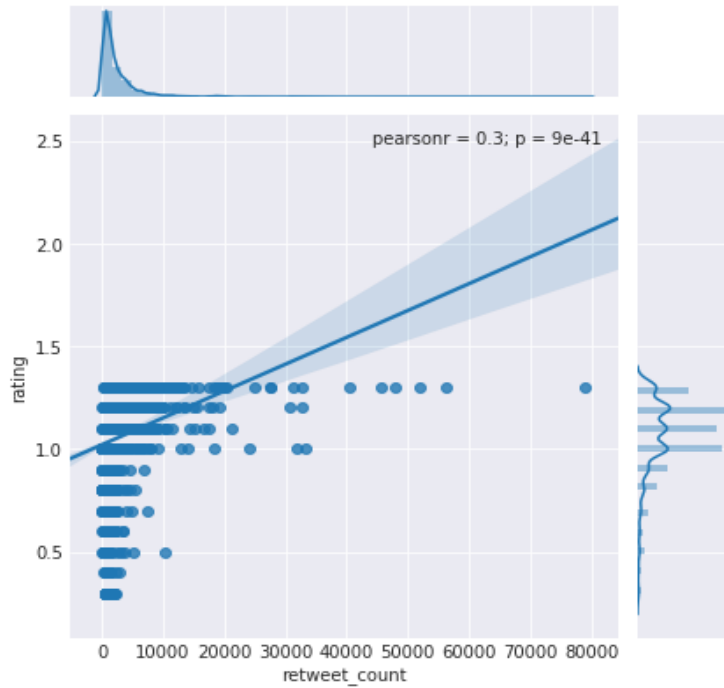
②不同品种的狗狗的受喜爱程度是怎样的？

③狗狗的什么名字最流行？

问题①：

分别绘制 retweet_count、rating 与 favorite_count、rating 的散点图添加线性回归曲线。

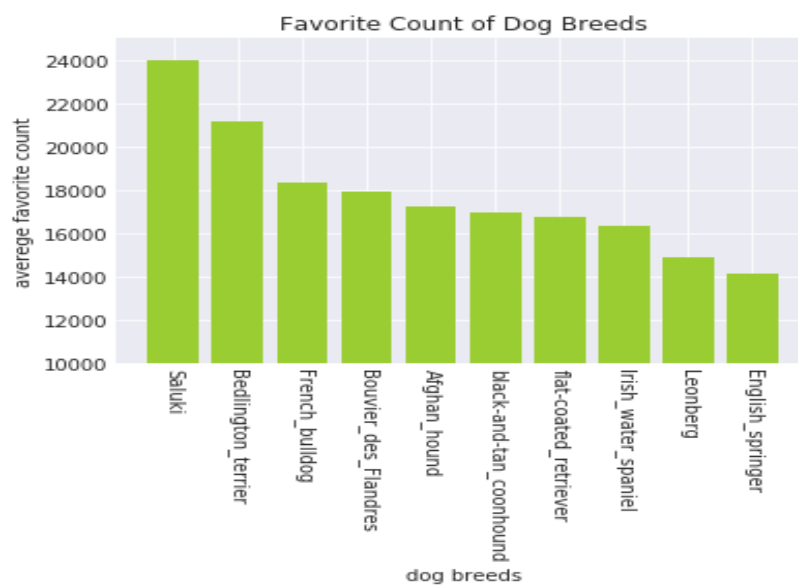




结论：从以上两幅图可以看出，retweet_count、favorite_count 与评分 rating 的关系成正相关。说明狗狗的评分对狗狗的喜爱程度（人们的关注度）有影响，亦或是狗狗的喜爱程度（人们的关注度）影响了狗狗的评分。

问题②：

从狗狗品种预测数据中提取到每个狗狗的最可能的品种类别，对每种类狗狗收获的平均点赞数进行排名，选出排名前十，绘制成柱状图，如下：



结论：平均点赞数排名前 10 的狗狗品种 Saluki、Bedlington_terrier、French_bulldog、Bouvier_des_Flandres、Afghan_hound、black-and-tan_coonhound、flat-coated_retriever、Irish_water_spaniel、Leonberg、English_springer，其中 Saluki 平均点赞数最高，最受喜爱。

问题③：

此处使用数据集中狗狗的名字绘制成词云，可以比较直观看到狗狗名字的使用的流行情况，为契合狗狗主题，添加词云背景也是一只狗狗。



结论：词云中显示 Charlie、Oliver、Cooper、Lucy 等字体较大，说明这些名字使用得较多，在狗狗中比较流行。