数据分析及可视化

前期我对数据集进行了数据评估，清洗与整理之后，三个数据集合为一个，也相对清洁了。在这个基础上，我提出了感兴趣的问题，并针对问题进行了数据分析和可视化。

在这部分中，我提出了以下问题：

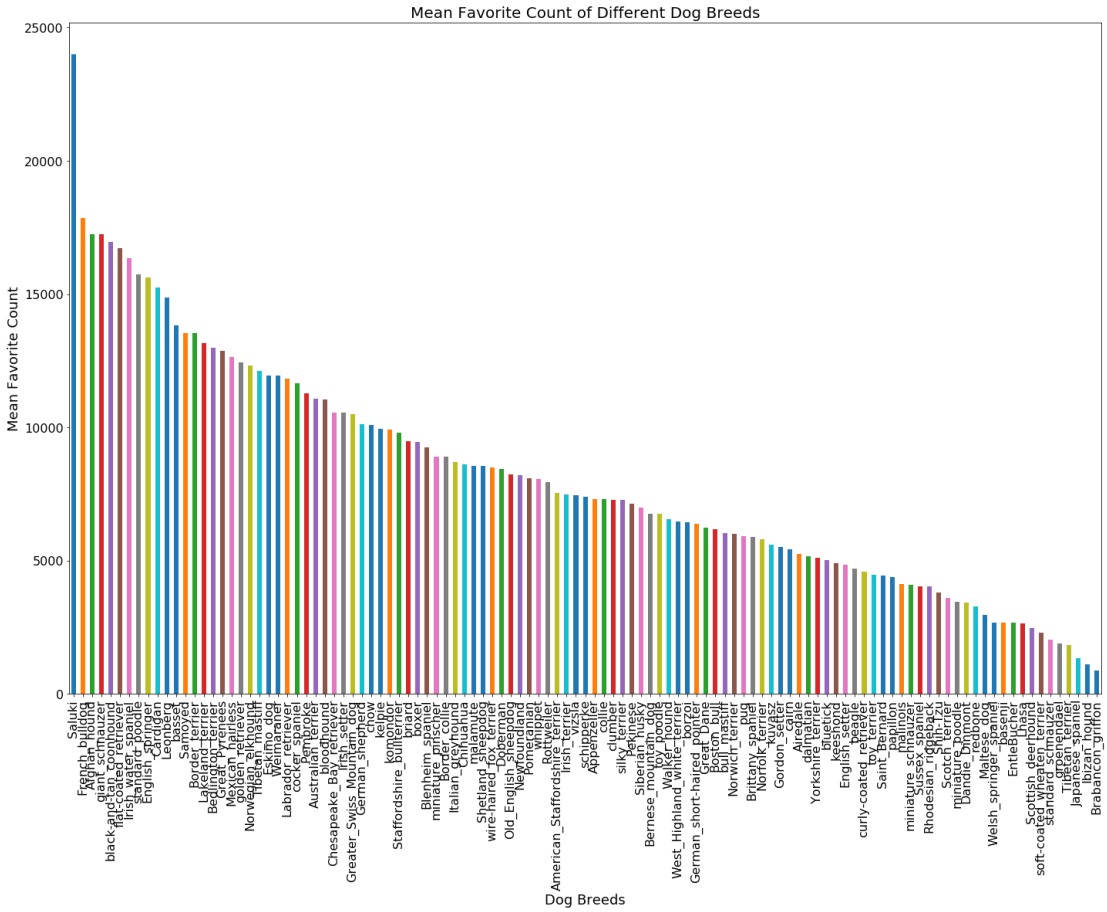
1. 被点赞数最多的前 10 大狗的排名及信息如何？
2. 哪种品种获得的点赞数最多，公众是否对某一品种存在特别好感？
3. 点赞数和转发数是否有强相关性？
4. 点赞数和博主的打分是否有强相关性？

针对第 1 个问题进行分析：1. 被点赞数最多的前 10 大狗的排名及信息如何？
首先针对数据集中点赞数列进行排序，排出前 10 大点赞数最多的狗狗。然后根据 jpg_url 列中的链接，编程下载前 10 大狗狗的图片，存在名为 Top10_dogs_images 的文件夹中。前十大狗狗的图片名分别为 Top1.jpg, Top2.jpg……，以下为图片的文件夹。

Top10_dogs_images.zip

分析第二个问题：2. 哪种品种获得的点赞数最多，公众是否对某一品种存在特别好感？

首先通过 p1 列的狗的品种所获得的平均点赞数，依次从大到小排序并画柱状图。
由图可见，最受欢迎的前三大品种为 Saluki，French bulldog 和 Afghan hound。而且第一名的 Saluki 所获得点赞数远超第二名之后的点赞数。



于是我很好奇的去下载了一张最受欢迎的 Saluki 品种的狗狗的图片，如下。

针对第三个问题：3. 点赞数和转发数是否有强相关性？

通过分析点赞数列和转发数列之间的相关系数，得出的 0.9117 说明，点赞数和转发数的相关性较强，即通常点赞数越多，转发数也越多。

针对第四个问题：4. 点赞数和博主的打分是否有强相关性？
通过分析点赞数列和博主打分列之间的相关系数，0.4531 的相关系数说明，点赞数和博主的打分相关性并不太强。博主极具娱乐性的主观打分和大众的普遍感受不完全一致。娱乐性也是博主的一大特色。而且博主文字很幽默，能够博主确为专业能力很强的爱狗之人。

不过此次修改过的系数 0.4531，已经高于上次报告的 0.3040。这个改变的原因在于提取打分的时候，根据建议，考虑到了更多提取不正确的因素而修改了提取正则表达式。