
数据分析和可视化

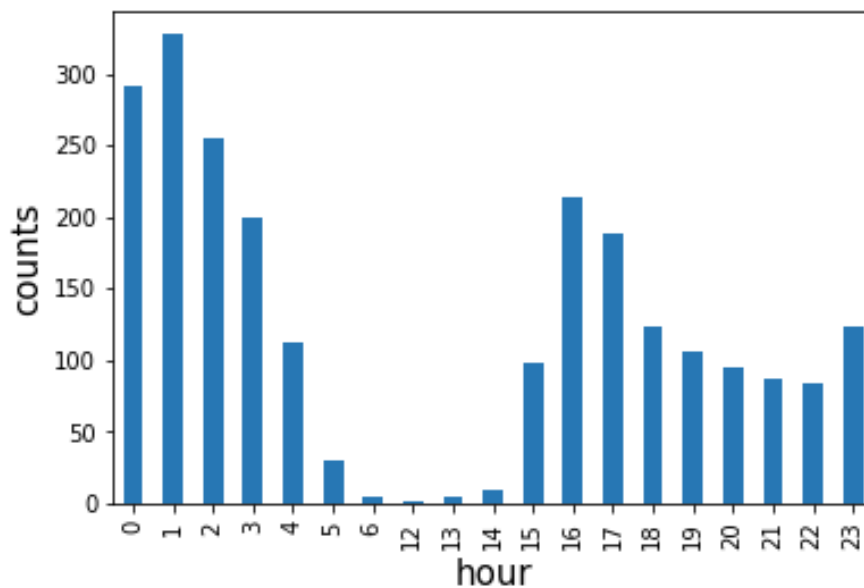
我对数据集进行了评估，清洗与整理之后，我提出了感兴趣的问题，并针对问题进行了数据分析和可视化。

我提出了以下问题:

- 1.在一天中的哪个时段，发推文的用户比较多？
- 2.哪些种类的狗狗获得了较高的转发数？
- 3.得到了高评分的狗狗得到的转发数就一定多吗？
- 4.数据中狗狗评分的高低和所处于的地位有什么关系？

针对第 1 个问题进行分析：1.在一天中的哪个时段，发推文的用户比较多？

首先我们把时间栏提取出来，并单独提取数据中的‘小时’数据，用value_counts() 查看该数据的值分布，并用条形图可视化数据。



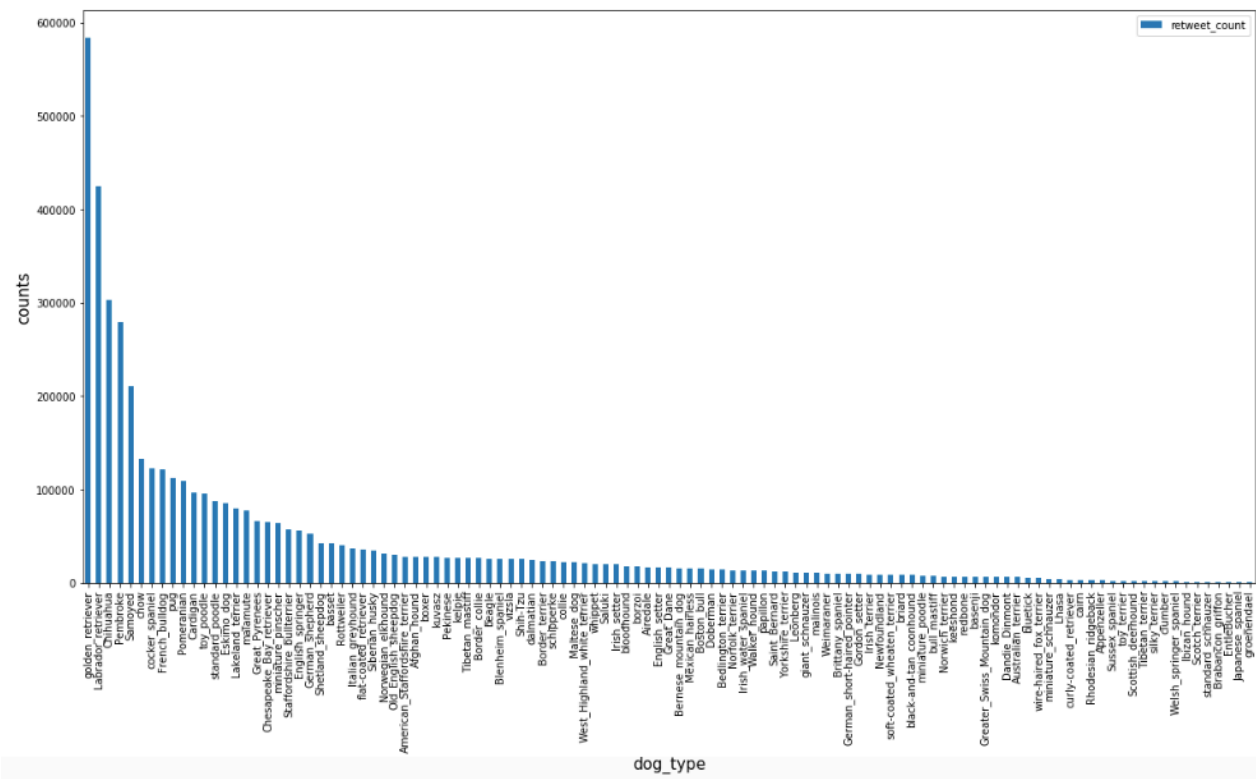
从图中我们可以看出， 凌晨时段和下午傍晚时段， 发推特的用户比较多

针对第 2 个问题进行分析：2.哪些种类的狗狗获得了较高的转发数？

首先我们观察预测数据， 使用同一行三种预测情况中选出最大概率的预测结果， 我们将预测结果分类不是狗的数据删除。

针对转发数， 清除较多没有数据的行， 调整转发数的数据类型为float。

提取预测结果狗的种类列与转发数这一列， 根据种类分组， 用条形图可视化， 观察每种种类获得多少次转发。

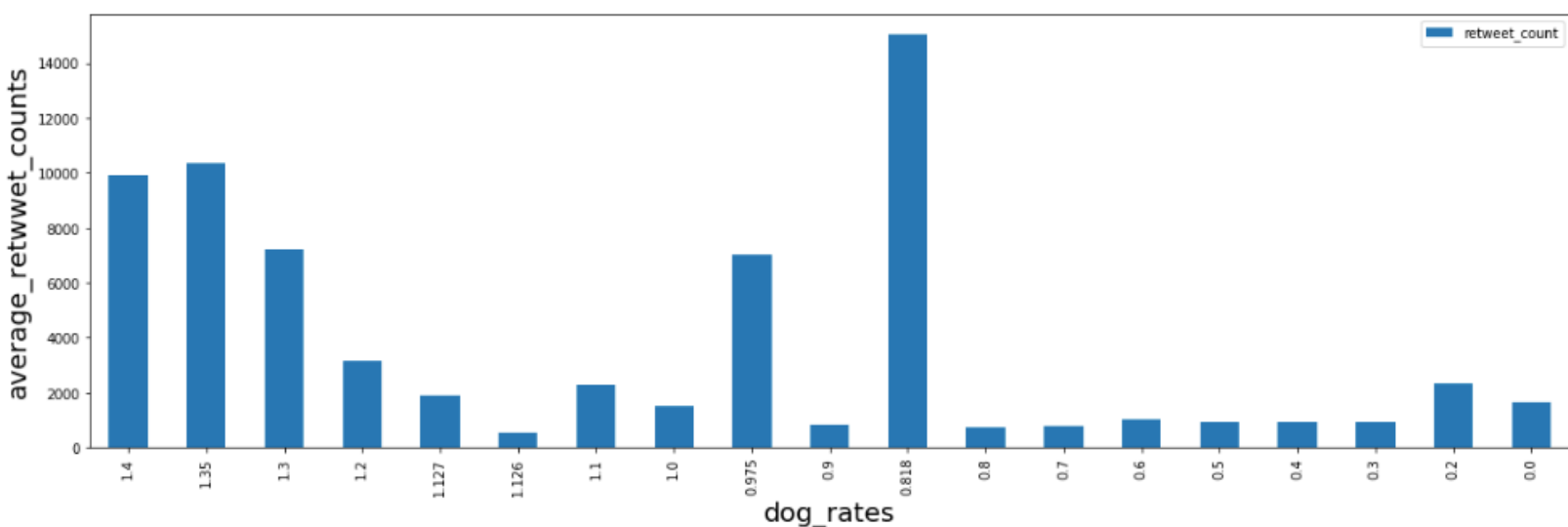


我们可以看到golden_retriever比较受欢迎， 得到了最多的转发数。

针对第 3 个问题进行分析: 3.得到了高分的狗狗得到的转发数就一定多吗?

清理质量问题后, 根据数据中的分子和分母, 相除计算评分的最终结果, 提取该列和转发数的列。

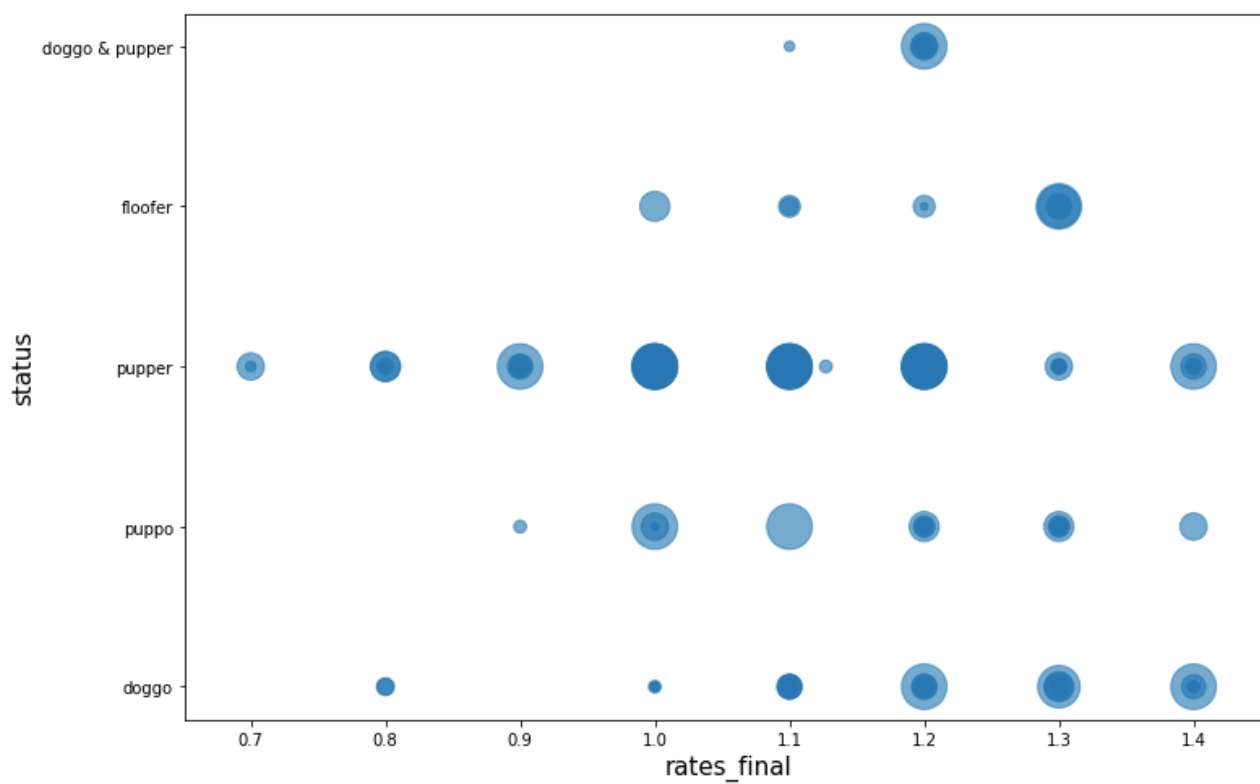
以评分数的各个特殊值, 对两列组成的DF进行分组, 求出每个评分所对应转发数的均值, 并以柱状图可视化。



我们从图中可以看到, 总体上来说 评分越高得到的转发数相应也高。

针对第 4 个问题进行分析: 4.数据中狗狗评分的高低和所处于的地位有什么关系?

我们删除掉地位栏内为空的数据, 提取出地位栏和评分栏, 观察每个评分下所对应的 狗的地位分布情况, 并以气泡图可视化。



观察数据发现，数据中狗狗评分的高低和所处于的地位关系不大，评分中的地位大多被 pupper和doggo占据,评分高的狗狗，处于doggo地位更多。

BUDGET

Ut vehicula nunc mattis pede

Curabitur labore. Ac augue donec, sed a dolor luctus, congue arcu id diam praesent, pretium ac, ullamcorper non hac in quisque hac. Magna amet libero maecenas justo.

Description	Quantity	Unit Price	Cost
Item 1	55	\$ 100	\$ 5,500
Item 2	13	\$ 90	\$ 1,170
Item 3	25	\$ 50	\$ 1,250
Total			\$ 7,920
