

Tweet dog rates 数据的清理与分析

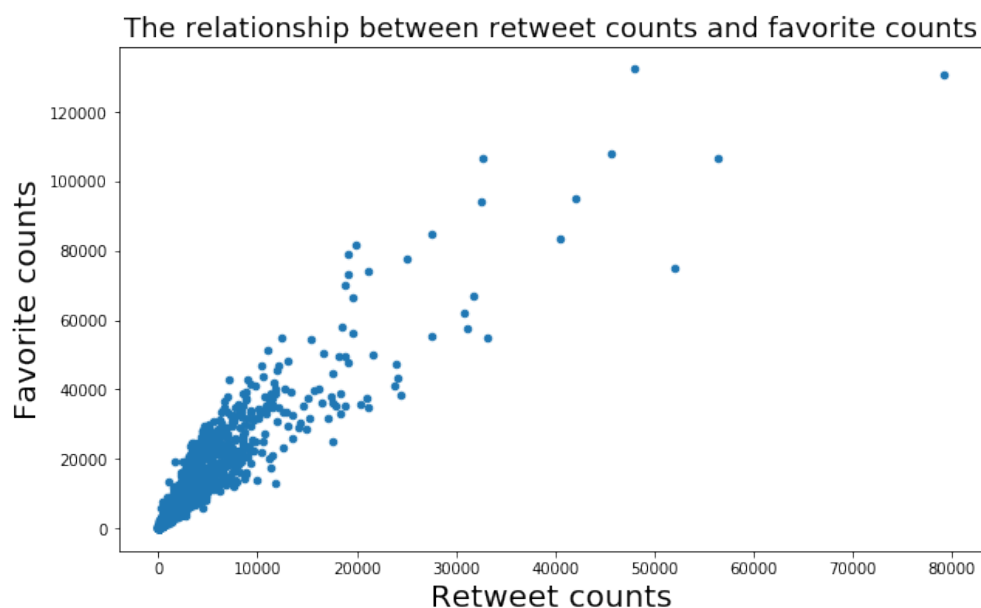
本数据集包含 2000 多个狗狗的数据，包括狗狗名字、分类、品种、评分等数据。在经过数据清理和可视化之后，提出如下几个问题，并通过分析，得出了结论。

问题：

- 1，Twitter 的点赞数和转发数呈什么关系？相关系数是多少？
- 2，在哪些时间段发送 Twitter，会有较高的点赞数预期？
- 3，四个种类的狗狗哪种最受欢迎，平均评分最高？
- 4，哪些品种的狗狗最受欢迎？它们中哪些品种的评分较高？
- 5，最受欢迎的狗狗名字是哪几个？

针对以上问题的分析和结论：

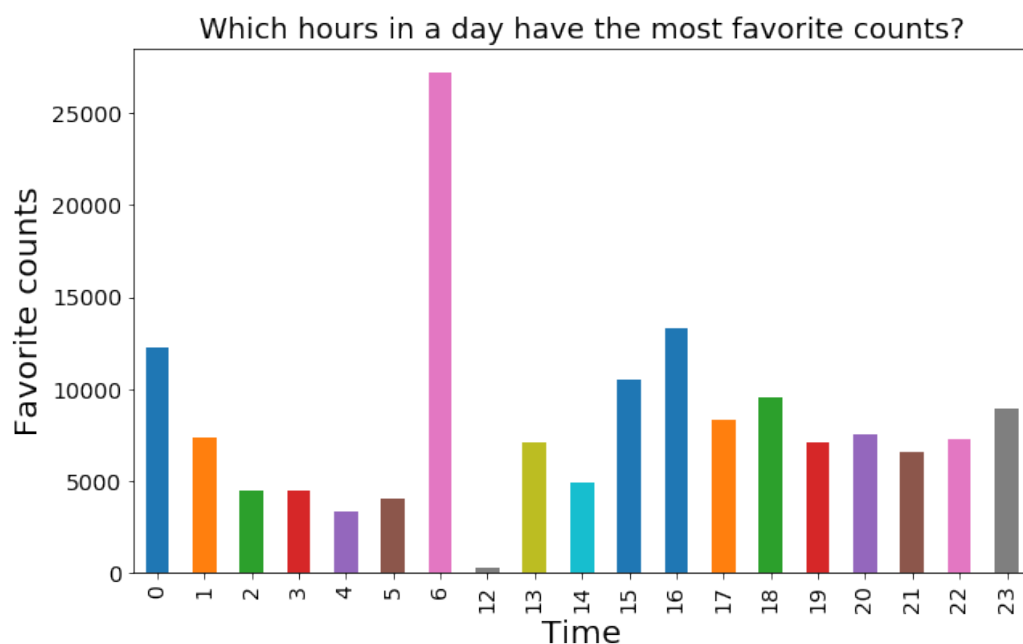
1，通过绘制散点图，可以较清晰的看出，点赞数与转发数呈线性相关关系，且为正相关，皮尔逊相关系数为 0.91。所以二者为强相关关系。转发数越多，会有越多点赞数的预期。



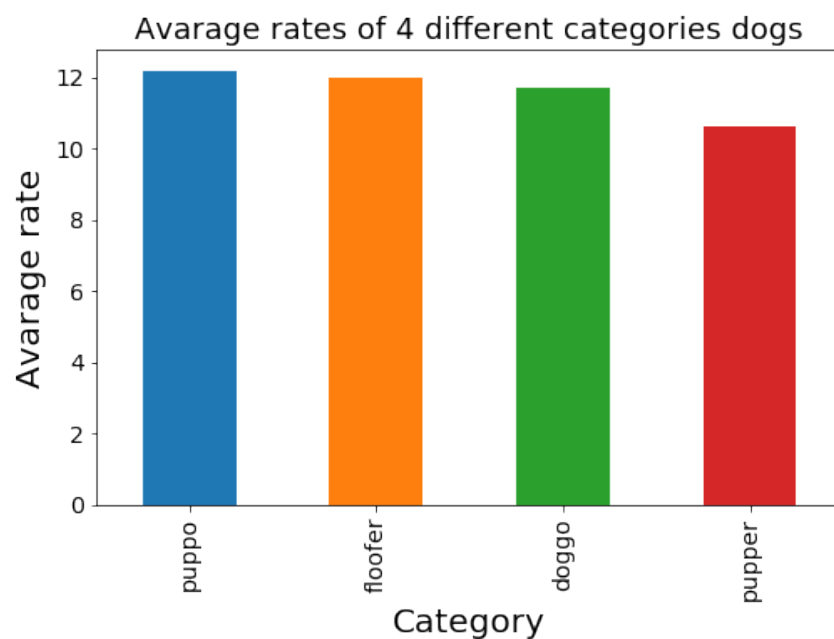
2, 通过拆解时间戳, 将 Twitter 发送的小时进行汇总分析, 发现在早上 **6 点-7 点**之间, 发送 Twitter 获得了最多的点赞数, 推测可能该时间为大多数人早上起床和上班时间, 在经过一夜睡眠起床之后, 习惯性打开手机观看 Twitter 并点赞, 抑或在上下班通勤的路上, 会将更多注意力集中在手机和电脑上。在此时间段发送的 Twitter 会得到较多的用户转发和点赞。

此外, 在下午 **15 点-17 点**之间, 点赞量也明显较高, 该时间段与上班族的下午茶时间对应, 推测此时会有更多用户在下午茶期间浏览 Twitter。

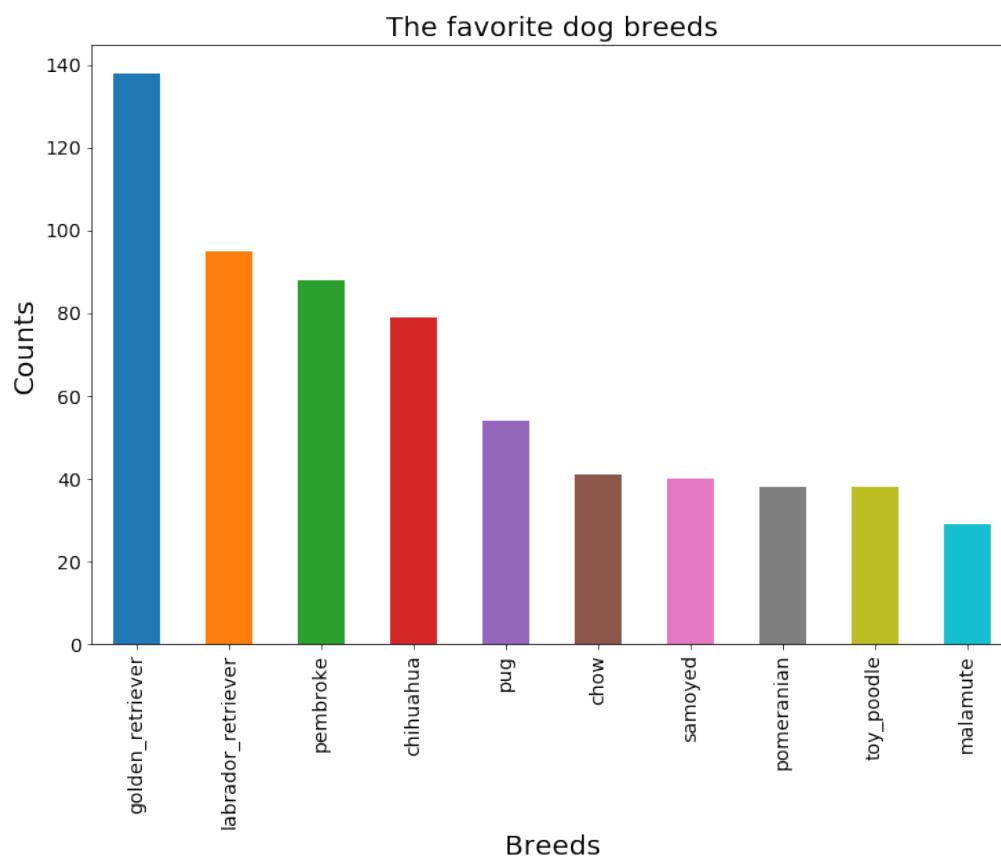
而且在晚上 **23 点-1 点**之间, 也有较多点赞量, 此时是大多数用户的临睡时间, 推测此时人们习惯在睡前浏览 Twitter 并点赞。



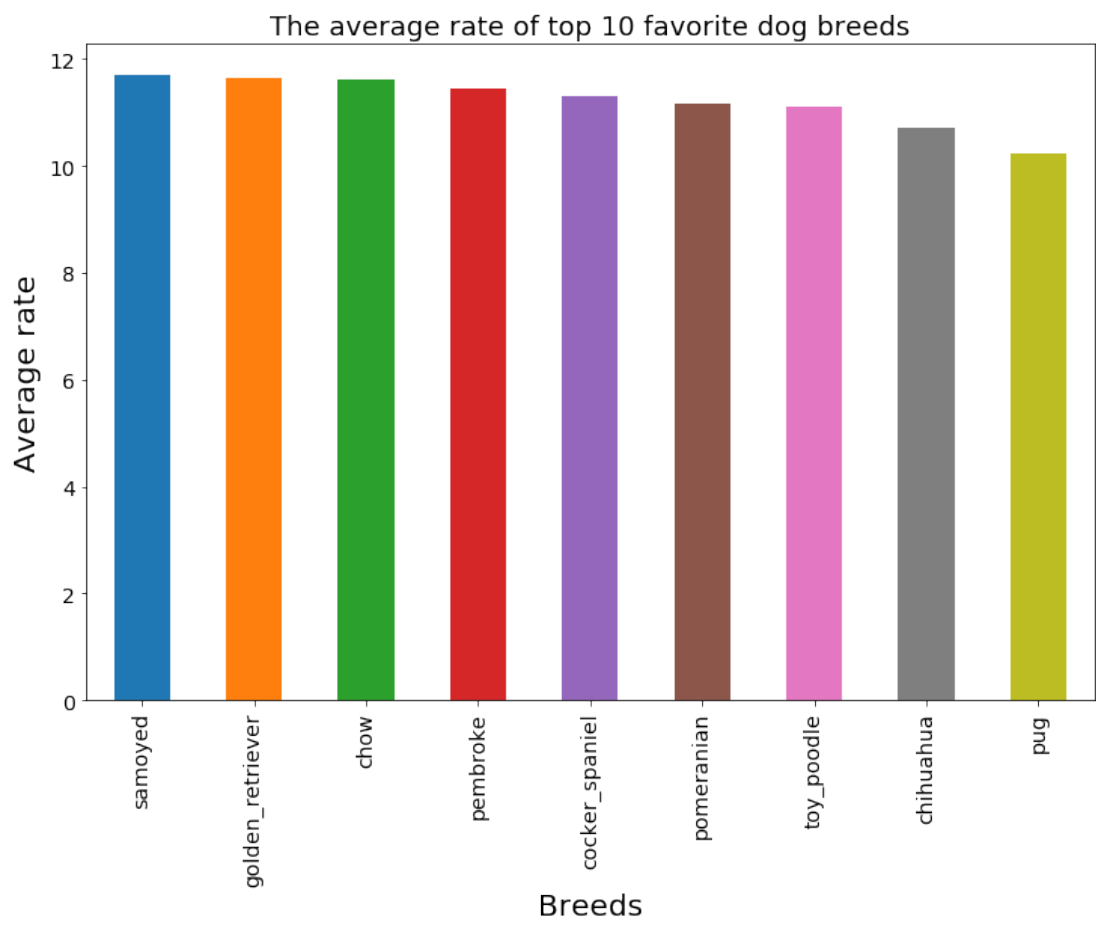
3, 四个种类的狗狗中, 最受欢迎的是 puppo, 平均评分超过 12 分, 其次是 floofer, doggo, pupper。



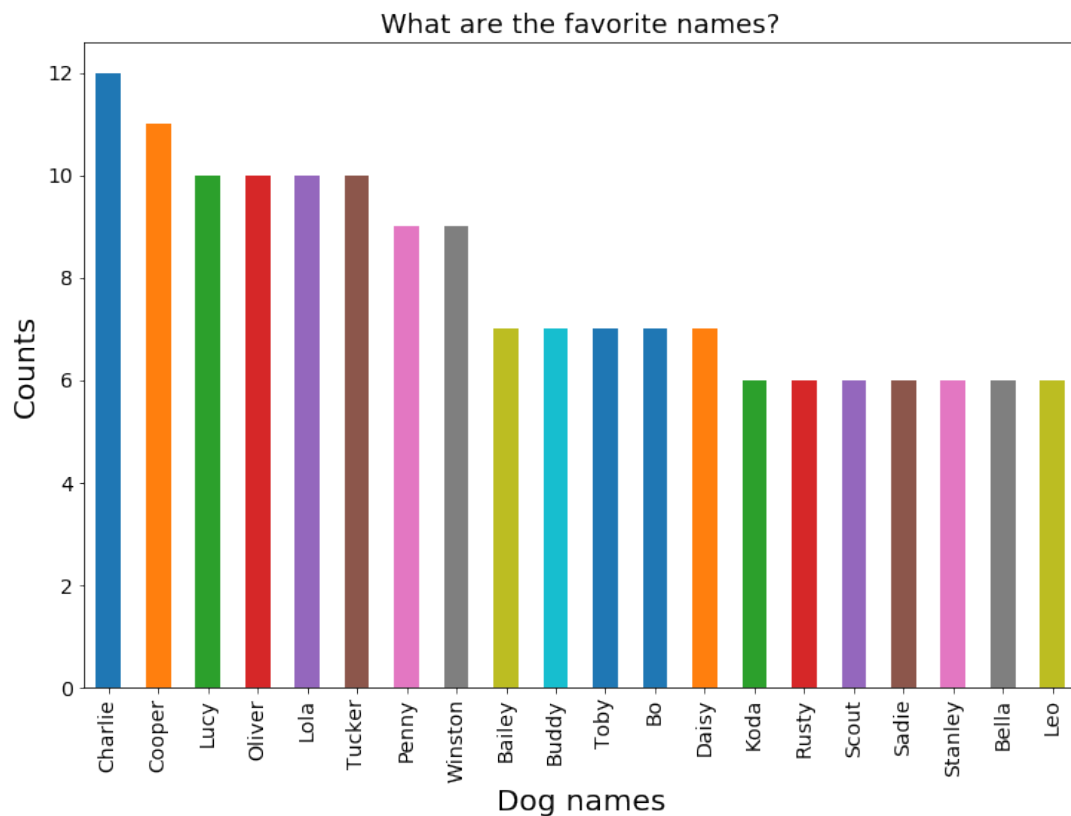
4，通过分析最受欢迎/饲养最多的 10 种狗狗，第一名的品种是 golden_retriever 统计显示有 150 个用户养的是 golden_retriever，其次，有 100 用户养的是 labrador_retriever，89 个用户养的是 pembroke 等等，其他品种狗狗的数量情况如图。



而在 top10 的狗狗品种中 ,评分最高的是 samoyed ,平均 11.7 分 ,其次是 golden_retriever 和 chow 的 11.6 分 , pembroke 的 11.4 分 , cocker spaniel 的 11.3 分 , 其他排名如下图。



5 ,最受欢迎的狗狗名称/出现频率最高的名字是 Charlie ,出现了 12 次 ,其次是 Cooper 11 次 , Lucy、 Oliver、 Lola、 Tucker 10 次 , 其它受欢迎的名字还有 Penny、 Winston、 Bailey、 Buddy、 Toby、 Bo、 Daisy、 Koda、 Rusty、 Scout、 Sadie、 Stanley、 Bella、 Leo 等。具体排名如下图所示。



总结和说明：

- 1, 本数据集总数据量并不足够大, 且在进行清理之后, 仍然存在缺失值, 所以以上分析和结论仅为初步结果, 有待数据丰富后获得更精确的分析和结论。
- 2, 在分析过程中, 仅针对部分数据之间的关系, 进行汇总和归纳, 数据集中仍存在有价值的信息有待进一步发掘。
- 3, 本分析结论仅做为数据归纳总结, 为相关性的层面, 并由此引发思考, 得出的部分推测, 并非出于因果关系, 即结论并非绝对正确的事实。

20190831

崔传敏