

Act Report

背景资料:

这次项目的数据集是推特用户 @dog_rates 的档案, 推特昵称为 WeRateDogs。WeRateDogs 是一个推特主, 他以诙谐幽默的方式对人们的宠物狗评分。这些评分通常以 10 作为分母。但是分子则一般大于 10: 11/10、12/10、13/10 等等。清洗 WeRateDogs 推特数据, 创建有趣且可靠的分析和可视化。以下的可视化分析包含了 1. 随着时间的变化流行度的趋势, 2. 预测频率最高的狗品种, 3. 狗地位的评分分布。

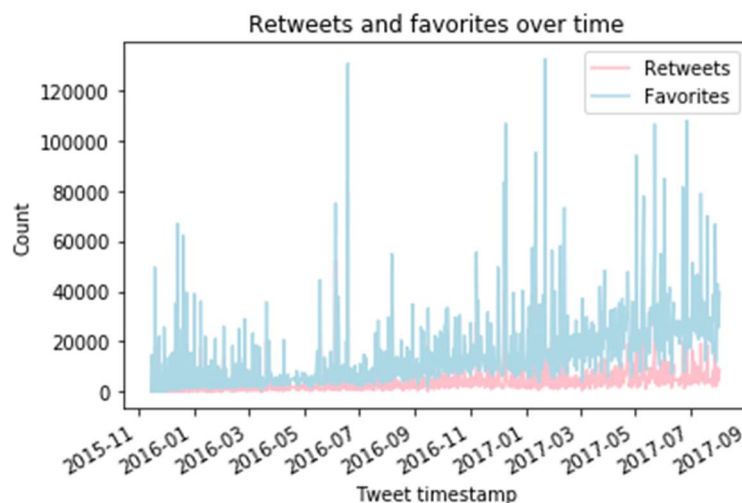
数据分析结论:

通过整理, 分析以及可视化数据集, 有以下的观察和结论。

1. 随着时间的变化流行度的趋势

在以下 Retweets 和 Favorites 数量 vs.时间的柱状图中, 可以发现两者随着时间的推移有什么变化。

Favorites 喜爱度在 2016 年 6-7 月及 2017 年 1-2 月的期间里达到最高峰, 而转发度 retweets 则在 2017 年 3-4 月的期间里达到最高峰。就以下观察, 转发和喜爱度随 2 年时间的推移越来越高, 表示 we rate dogs 越来越用户的受欢迎和关注。

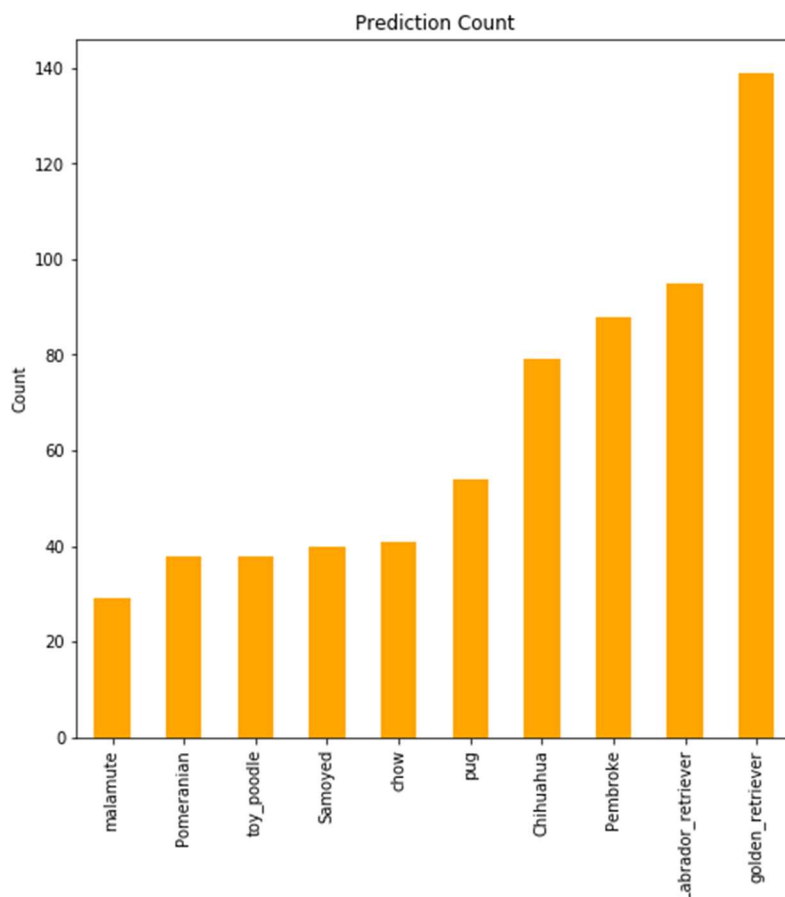


2. 预测频率最高的狗品种

在以下预测狗品种数量柱状图中, 列出了前十名第一位被预测最多的狗品种。

被预测最多的品种是金毛寻回犬, 其次是拉布拉多, 柯基犬, 吉娃娃, 八哥, 松狮狗, 萨摩耶, 贵宾犬, 博美犬, 以及雪橇犬。根据以下分析, 最受欢迎的狗品种是金毛, 这也和一般的认知相吻合。





3. 狗地位的评分分布

在以下狗地位评分的箱线图中，可以看到 **Pupper** 的评分差距较 **doggo**, **floofer** 和 **puppo** 大，中位数较其余三者低。最高点为 14 分，最低点为 7 分，中位数为 11 分。而 **Puppo** 得到的评分分布差距较小，评分普遍较高，最高点为 14 分，最低点为 11 分，中位数为 12 分。根据 **The Dogtitionary** 的文章对 **pupper** 和 **puppo** 的解释，可以看到青少年阶段的狗，**puppo** 普遍更受欢迎，而幼犬，**pupper** 相对不受欢迎。

