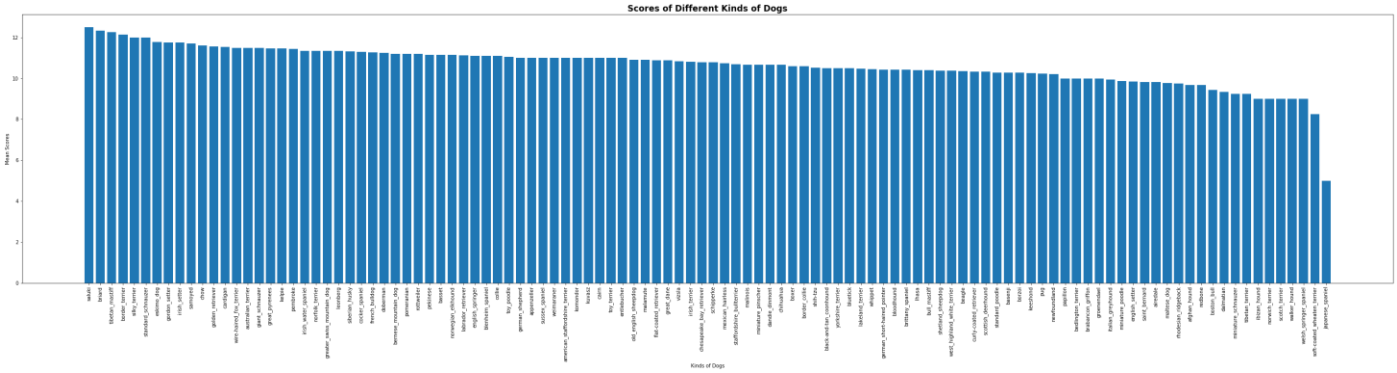
数据整理报告

1. 分析不同种类狗的评分情况

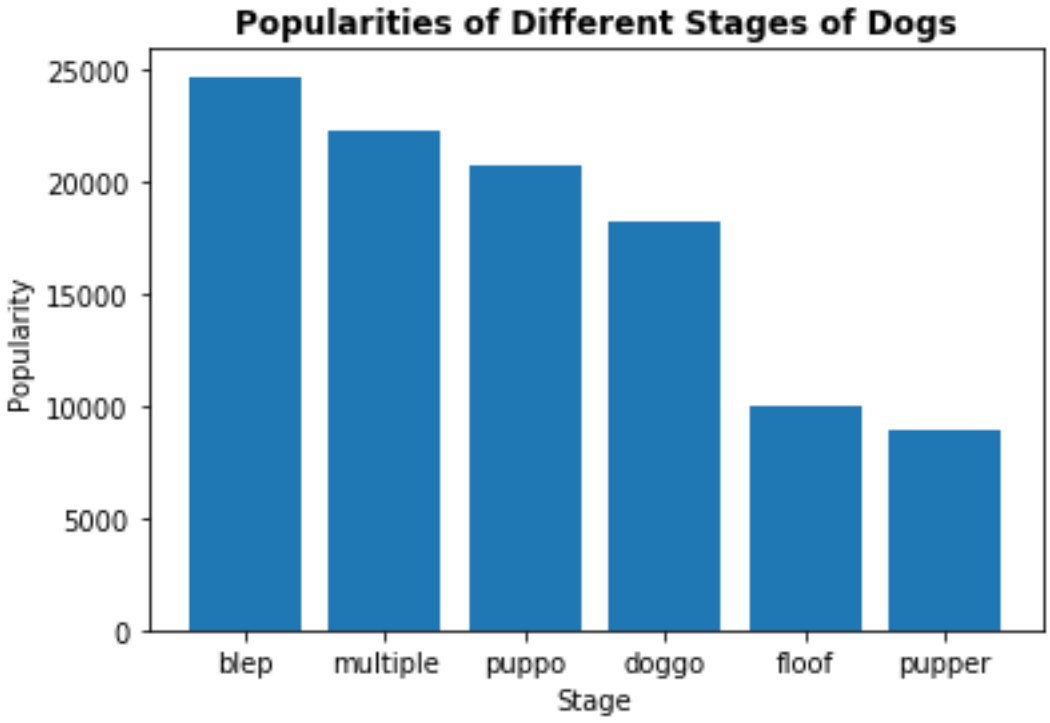
合并 WeRateDogs 提供的推特数据和图像预测数据，得到数据集 df_dogscore. 使用 matplotlib.pyplot.bar 绘制柱状图，x 轴为狗种类，y 轴为狗评分均值。结果如下：



可以看到，各个种类的狗的评分总体区别不大，排名前三位的狗依次是 saluki，briard 和 tibetan_mastiff。

2. 分析不同‘地位’狗的受欢迎程度

合并 WeRateDogs 提供的推特数据和推特 API 下载数据，得到数据集 df_stagepoplular. 使用 matplotlib.pyplot.bar 绘制柱状图，x 轴为狗“地位”，y 轴为狗所在推特被喜爱和转发次数之和。结果如下：



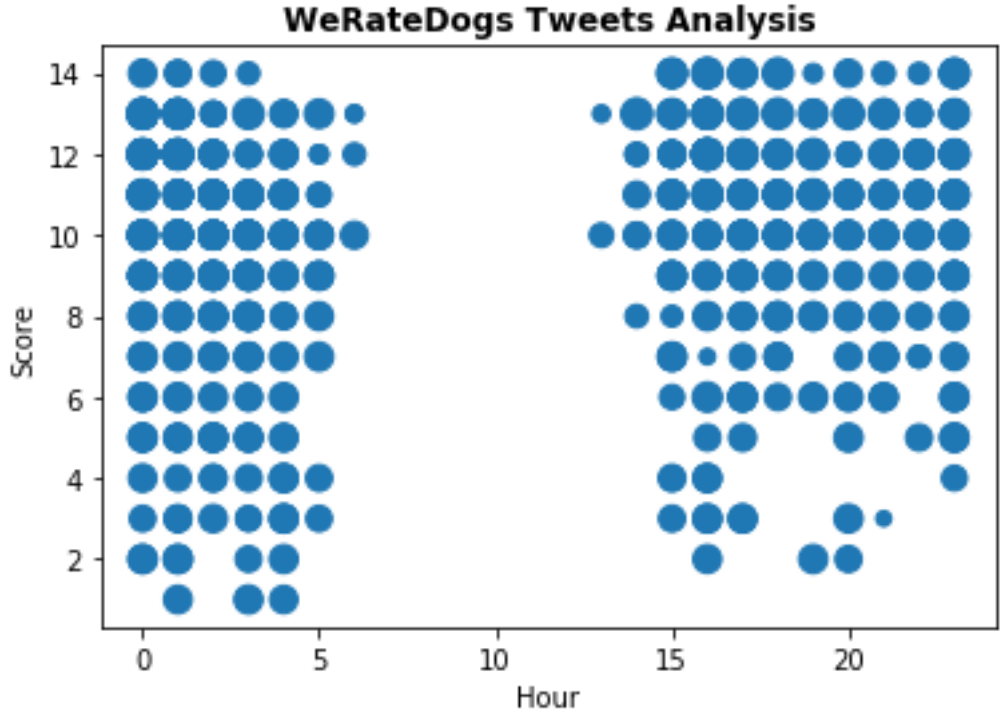
可以看出，“地位”是 blep, multiple 和 puppo 的狗在 tweeter 上最受欢迎（喜爱和转发次数最多）。

3.分析每天不同时段 WeRateDogs 对狗的评价是否有规律

直接使用 WeRateDogs 提供的推特数据。

针对这个问题，考虑到文字数和评分分别能够间接和直接反映 WeRateDogs 的评价习惯，因此使用这两个变量和每天的不同小时作为时段，探索评价规律。

使用 matplotlib.pyplot.scatter 绘制散点图，x 轴为小时时段，y 轴为狗评分，点大小为推文长度。结果如下：



可以看到：WeRateDogs 发推没有出现在 7 点到 12 点，发文字数没有特别明显的规律。相对来说，低分较多地出现在 0 点到 5 点发的推文当中，在 10 分以上的推文中，WeRateDogs 用的“笔墨”比较多。

有限性阐述

1.在数据整理时，去除了部分空值和异常值，这将会对结果造成一定误差。

2.在分析不同‘地位’狗的受欢迎程度时，未考虑用户的文字水平，图片质量等因素，而是只考虑狗的“地位”。

3.在分析不同种类的狗的评分时，默认图像预测的第一个结果为实际情况。