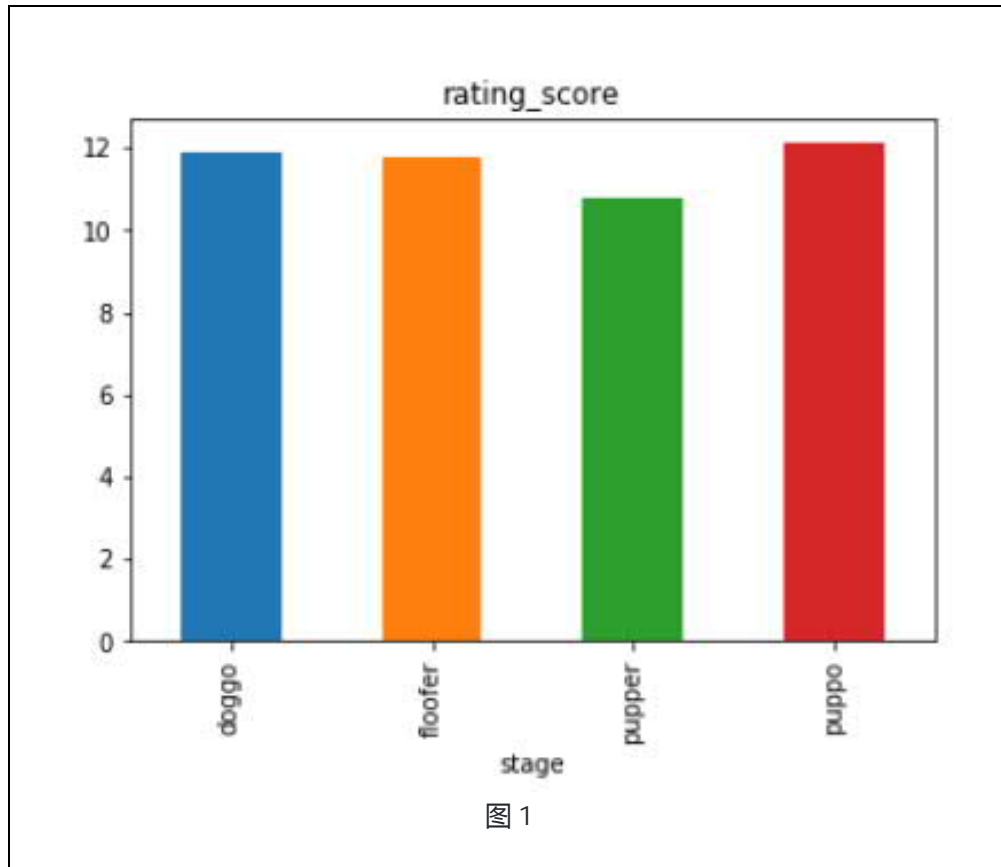


数据分析与可视化

`twitter_archive_master.csv` 是一份关于狗狗评分的推特数据集，其中包含推特 ID，狗狗地位，推特点赞数和转发数等数据。以下是对本数据集进行简单分析与可视化的过程。

1. 观察不同狗狗地位的获得的平均分数（图 1），可以看出 4 种地位的狗狗获得的评分差别不大。平均分数都超过 10 分，表明这 4 种地位的狗狗都比较受到人们的喜欢。



2. 从狗狗地位与相应的推特点赞数的关系来看（图2），puppo 狗狗推特的点赞数最高，pupper 狗狗推特的点赞数最少。但是这没有绝对的关系，点赞数可能与推特内容其他方面相关。

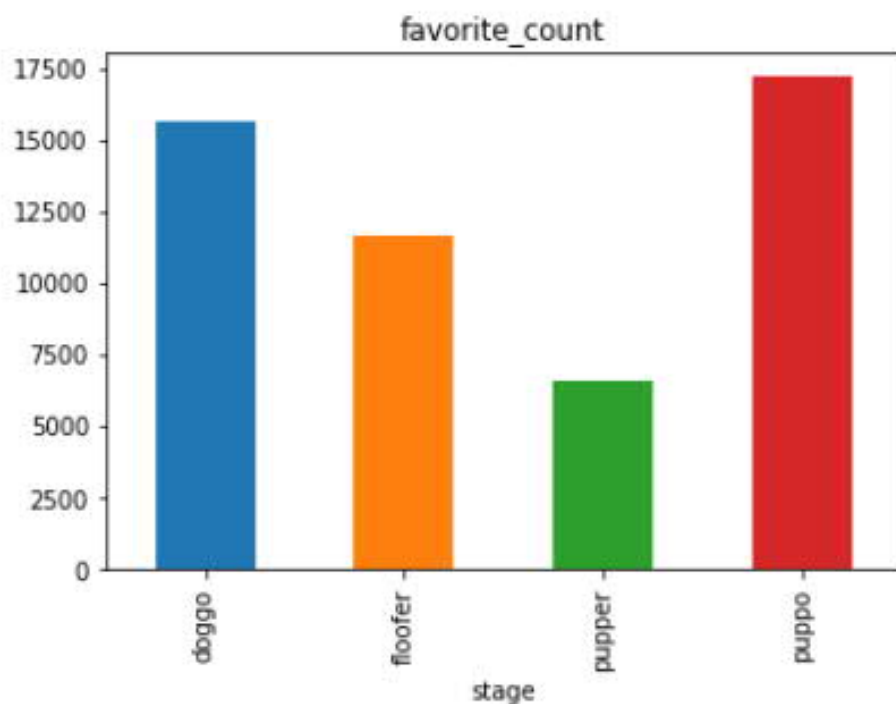
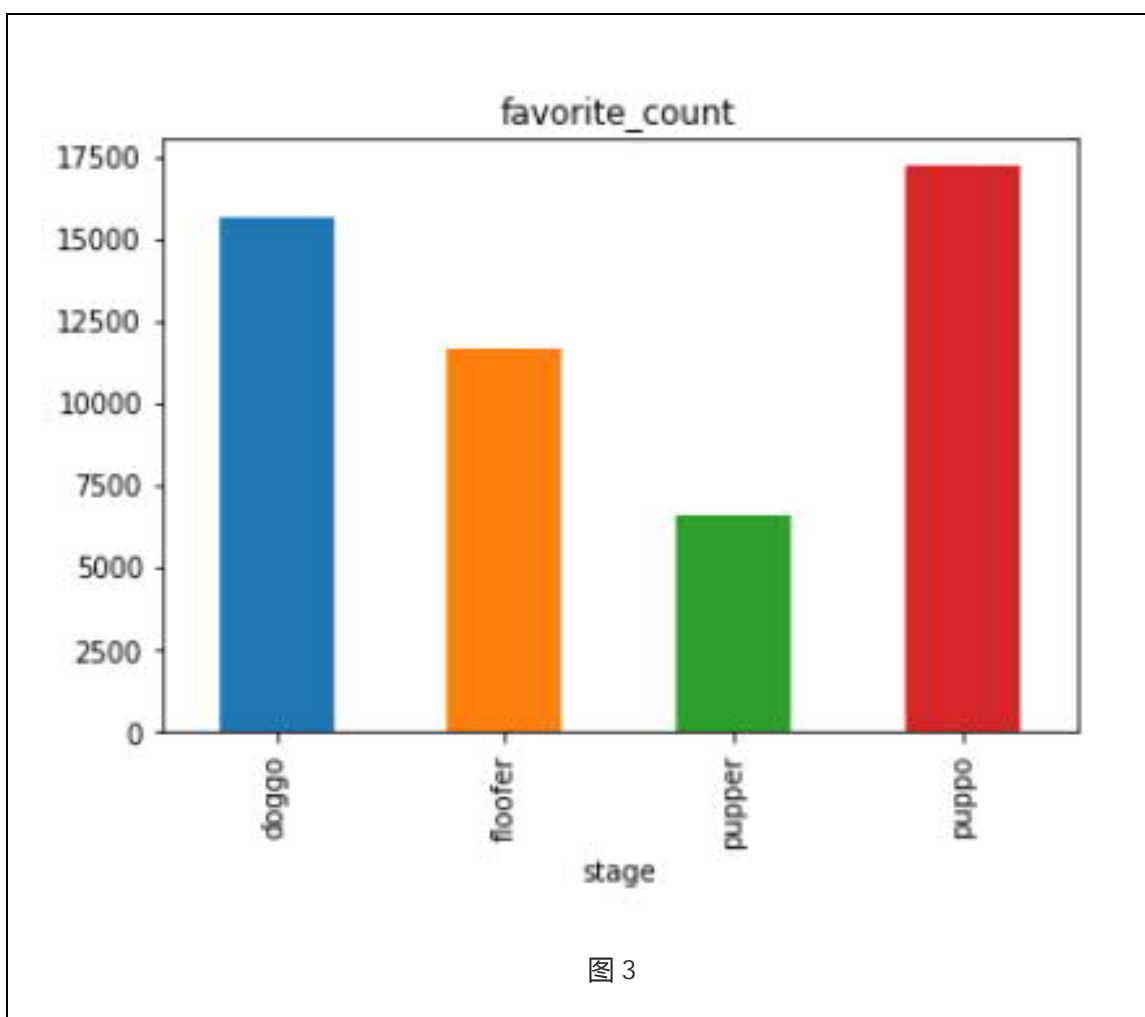


图 2

3. 从推特的点赞数和转发数的关系来看（图 3），二者有轻微的正相关，大致可以反映出，点赞越多的推特也越多人转发。



4. 观察评分值与推文发表的时间的关系（图 4），比较惊奇地发现，早上 6 点到中午 1 点这段时间出现较高的评分（缺少中间 7 点-11 点时间的数据）

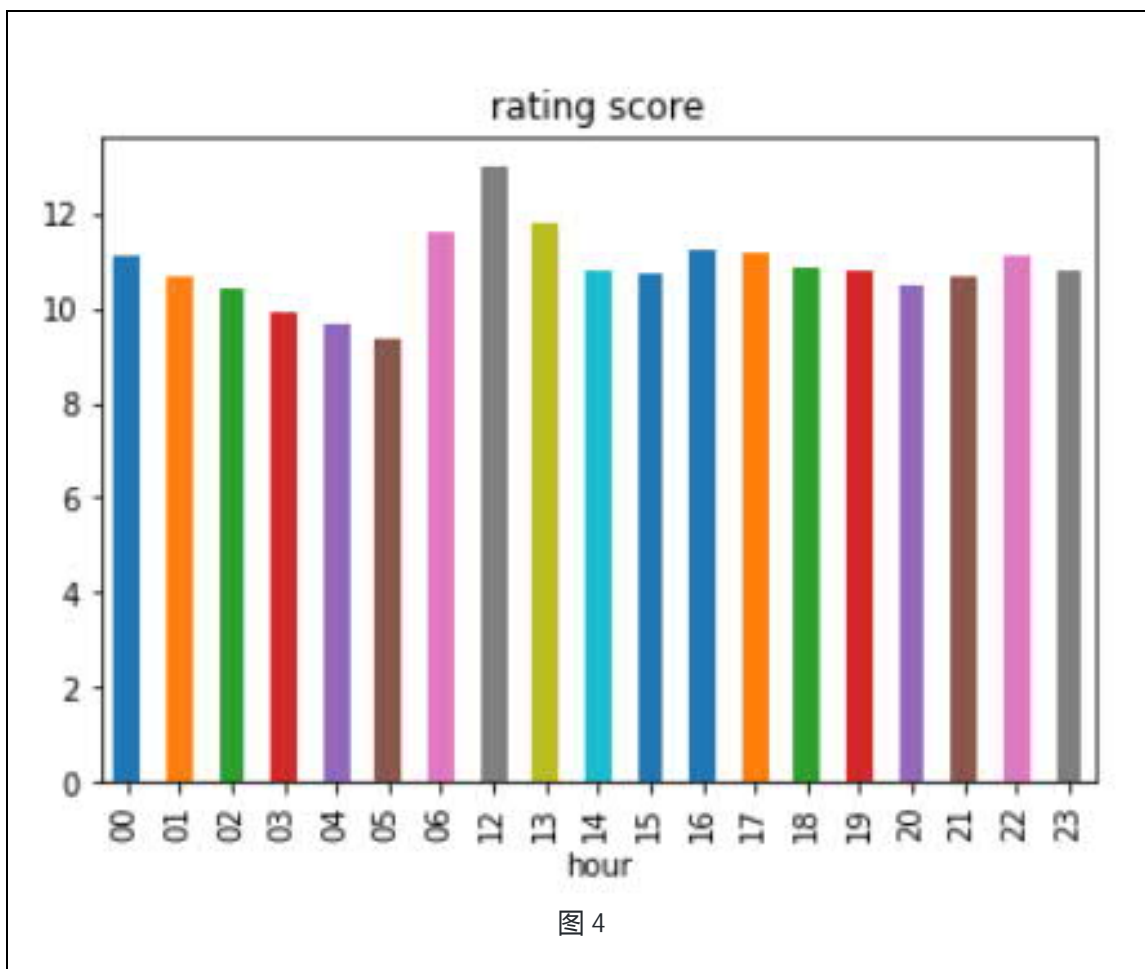


图 4

5. 观察预测结果出现频率最高的前 5 个狗狗品种（图 5），其中 golden_retriever 出现次数遥遥领先。

