

关于清洗 WeRateDogs 推特数据

大家好，我是 YimingZheng,这次的项目主要是数据清洗，所以最主要的时间都花在了数据清洗上了，这个过程对我这种数据菜鸟来说感觉真有点儿不容易，特别是我才开始清洗的时候就听到助教说已经有三个人通过了项目二。可想而知，我又紧张又害怕，然后陆陆续续赶进度，赶回来了，时刻提醒自己，要高效。向厉害的同学看齐，加油。

以下是我清洗完数据后进行的一些简单的分析，分析的问题包括：

- 1.狗狗的 stage 中哪种身份最多？
- 2.数据集中，排名前 10 的最常见的 10 个名字是哪些？
- 3.数据集中，狗狗的评分大部分是多少？

下面依次为我的简单可视化结果：

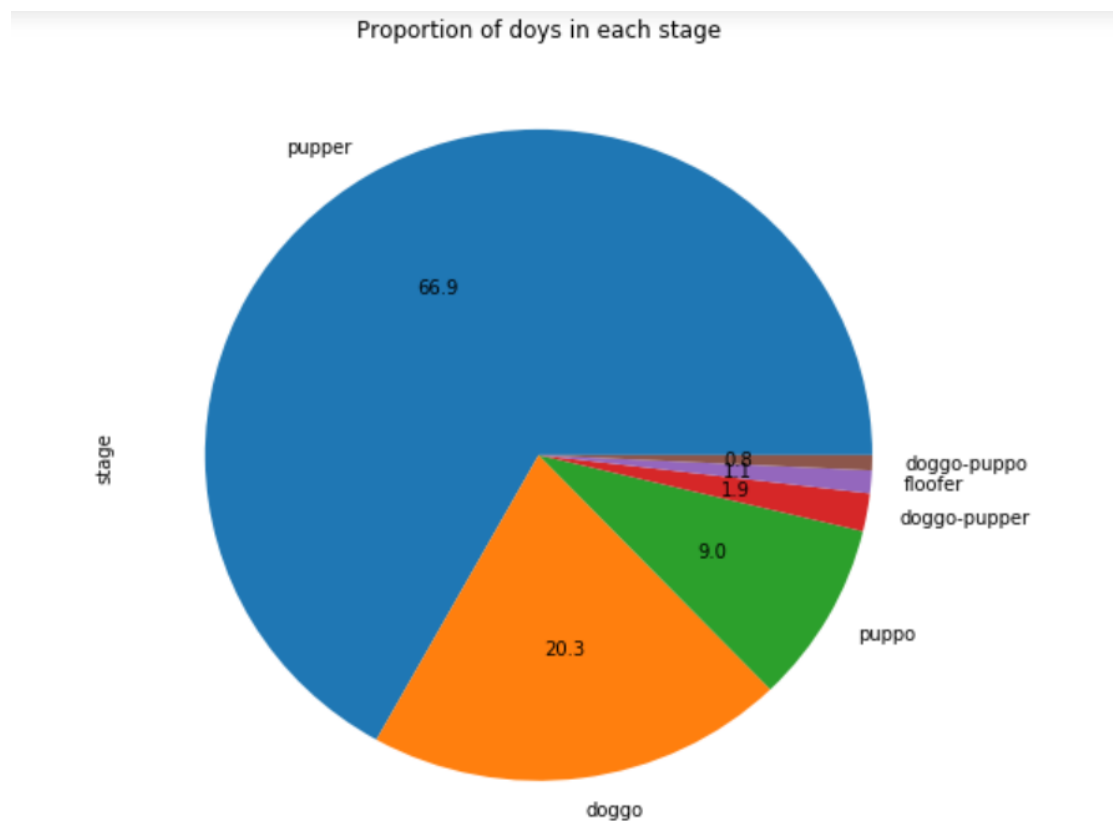


图 1 狗狗的 stage 中哪种身份最多？

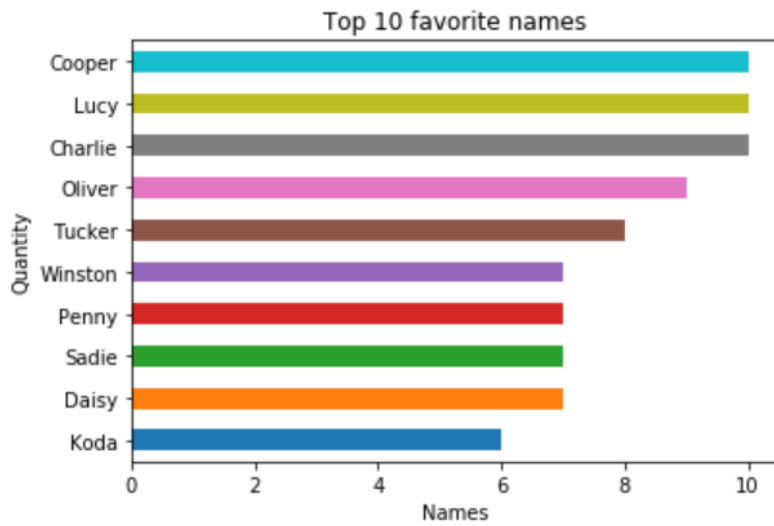


图 2 # 分析 数据集中，大家最喜欢给狗狗取什么名字

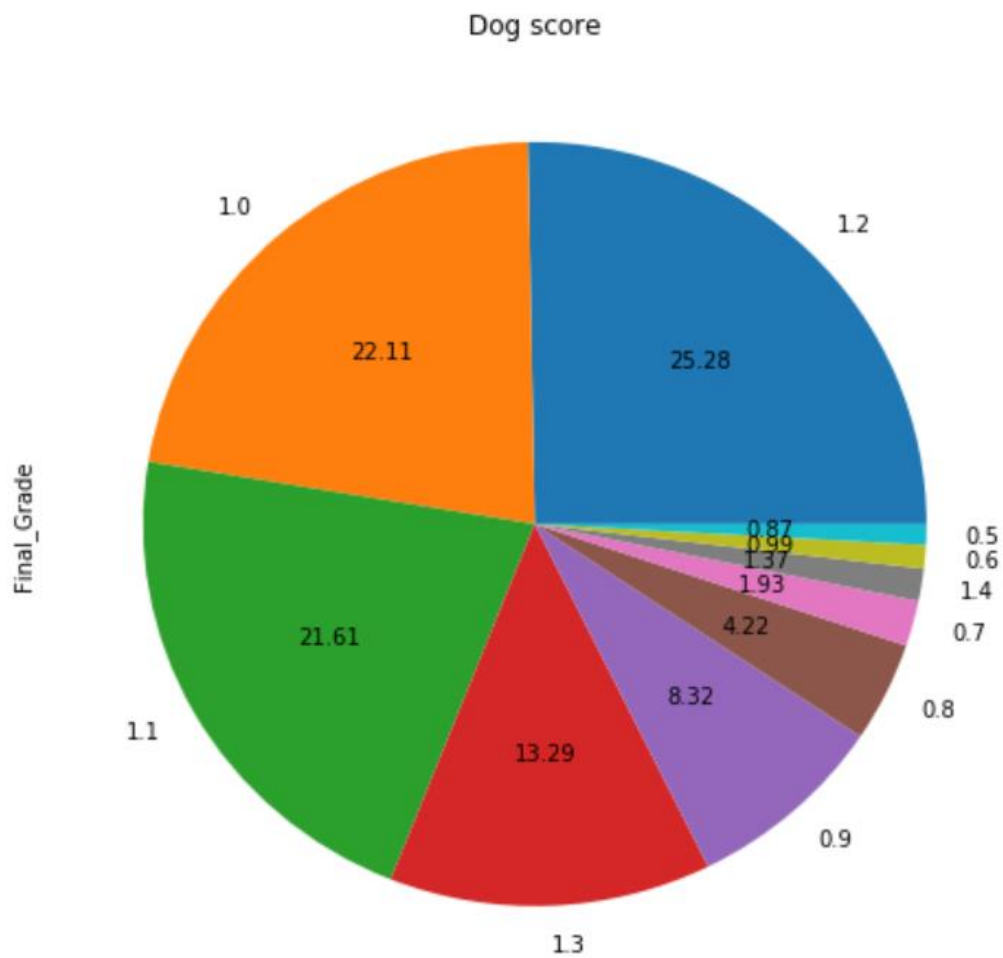


图 3 # 分析：数据集中，狗狗的评分大部分是多少？

谢谢大家。