基于Spark ALS 推荐

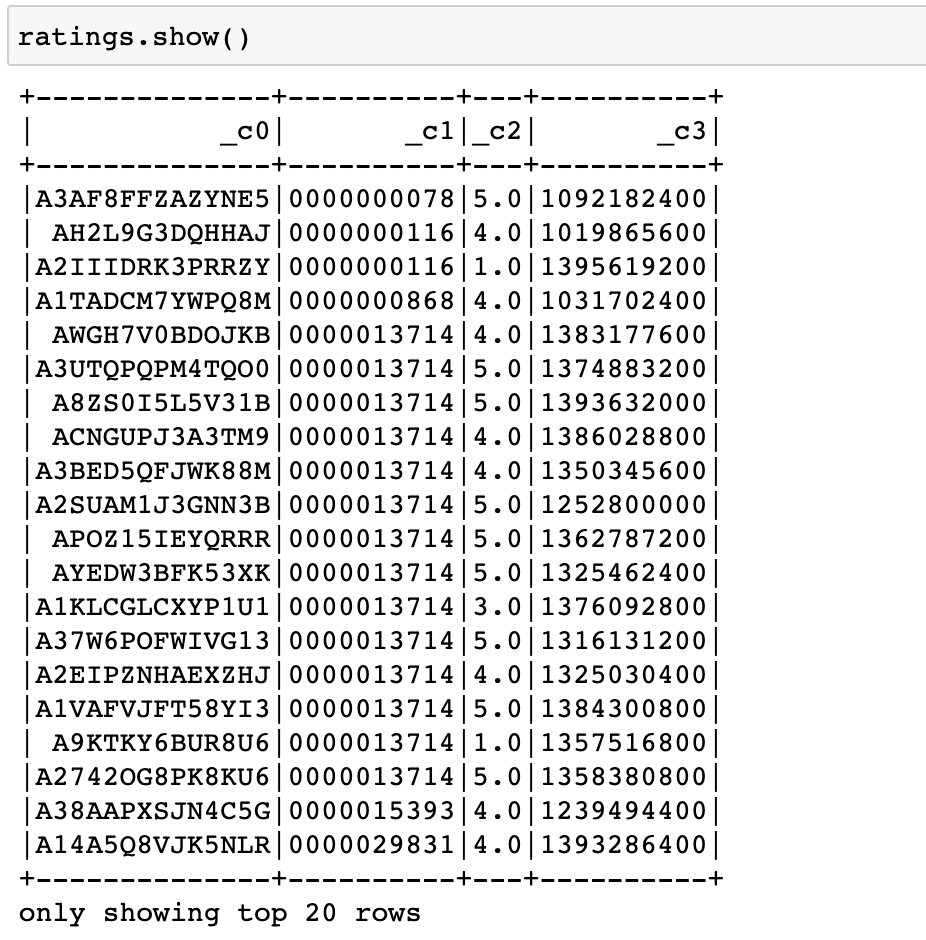
逍阳 - 2022年7月15日



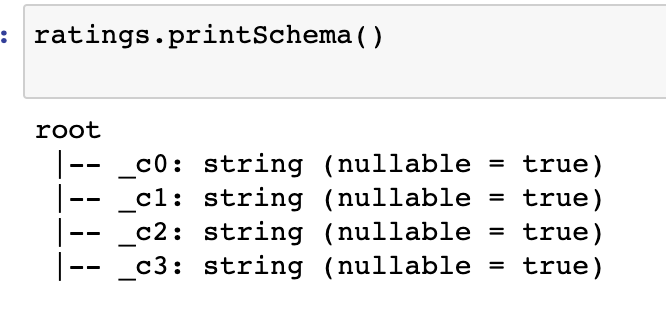
介绍

加载必要的pyspark库和数据

然后进行初步数据探索：

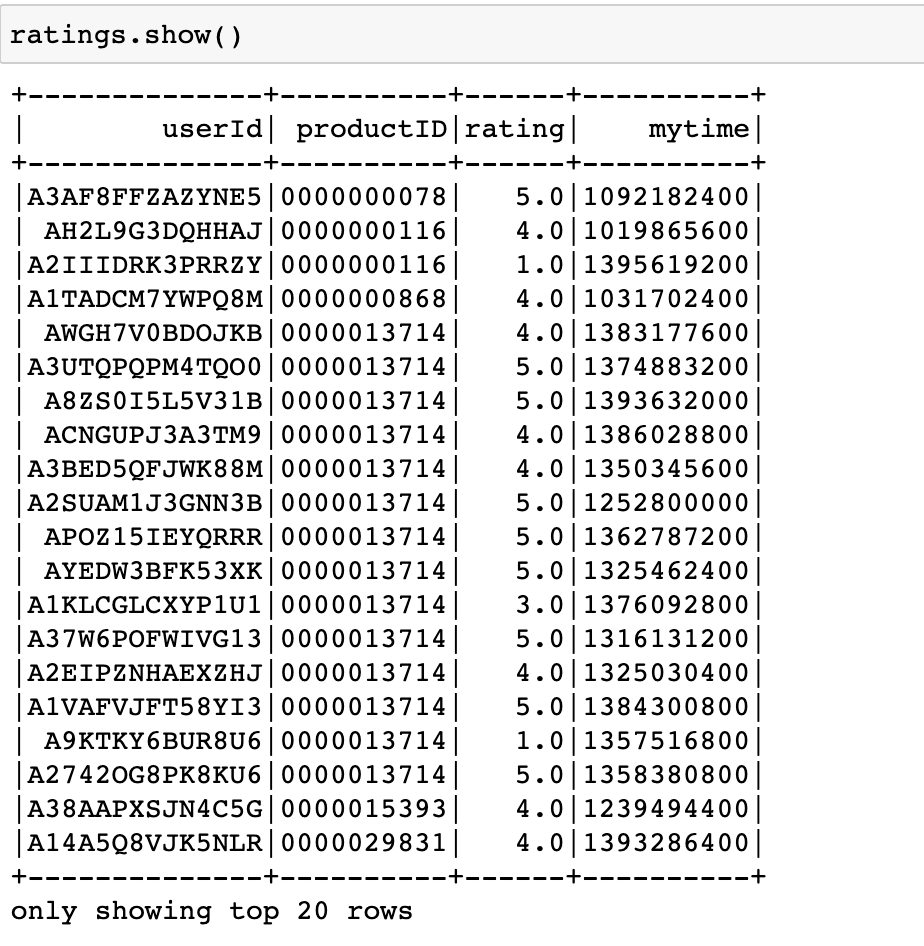


属性检查：



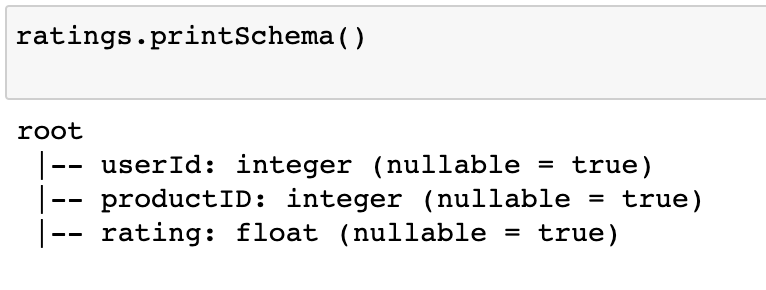
我们添加列名，根据数据集描述：



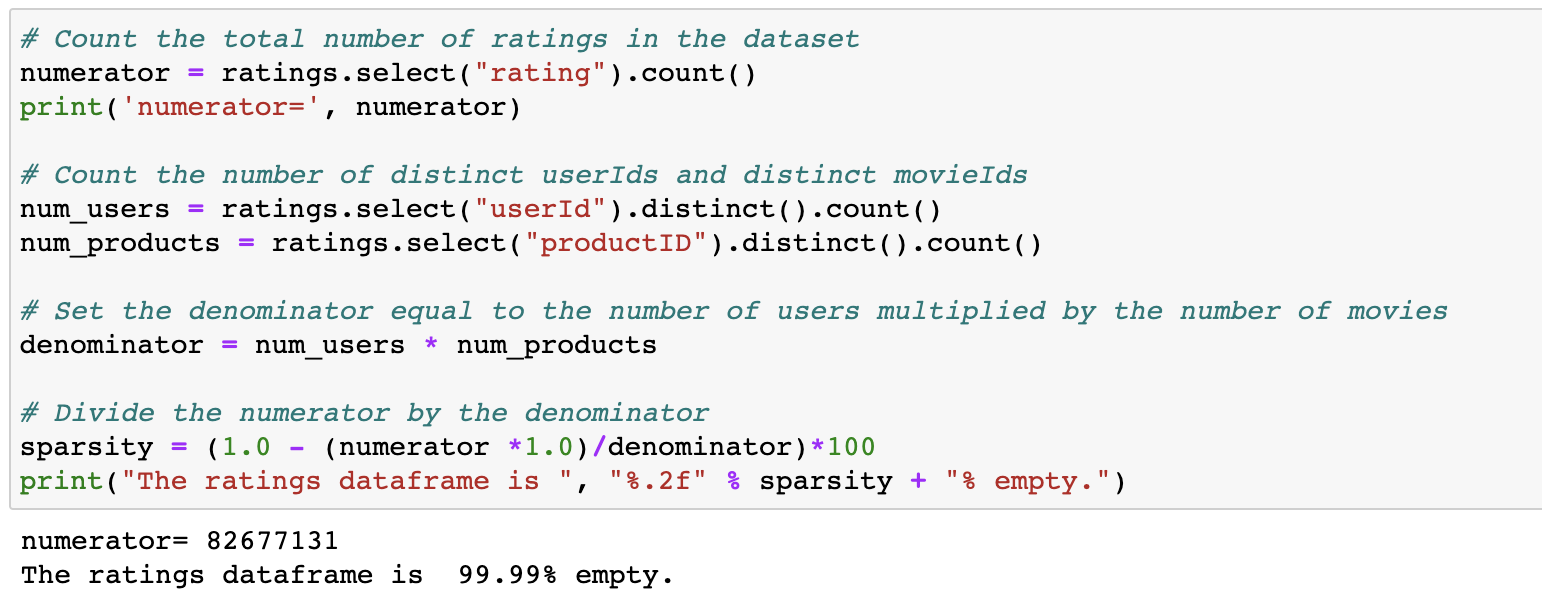


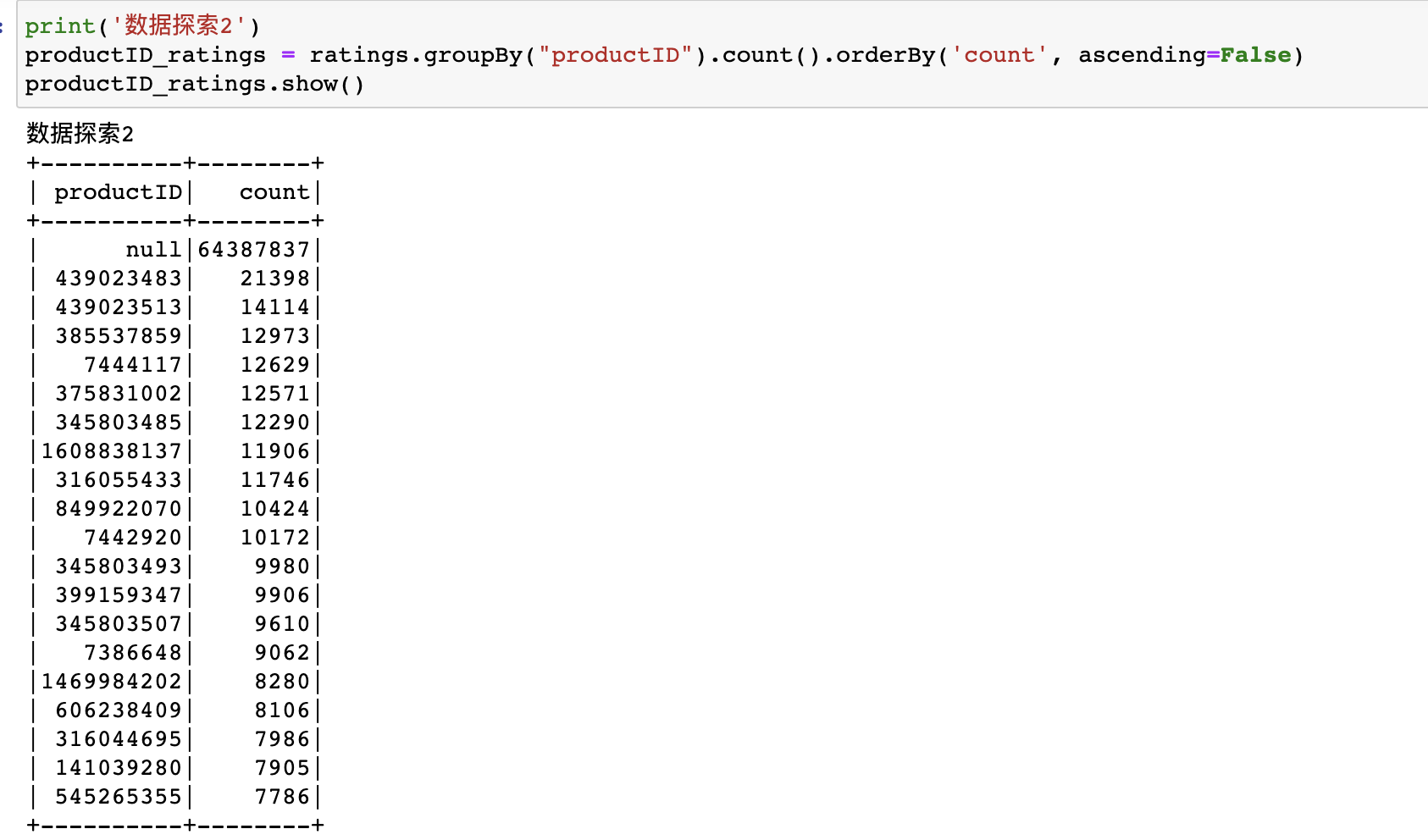
进行列的类型转化：

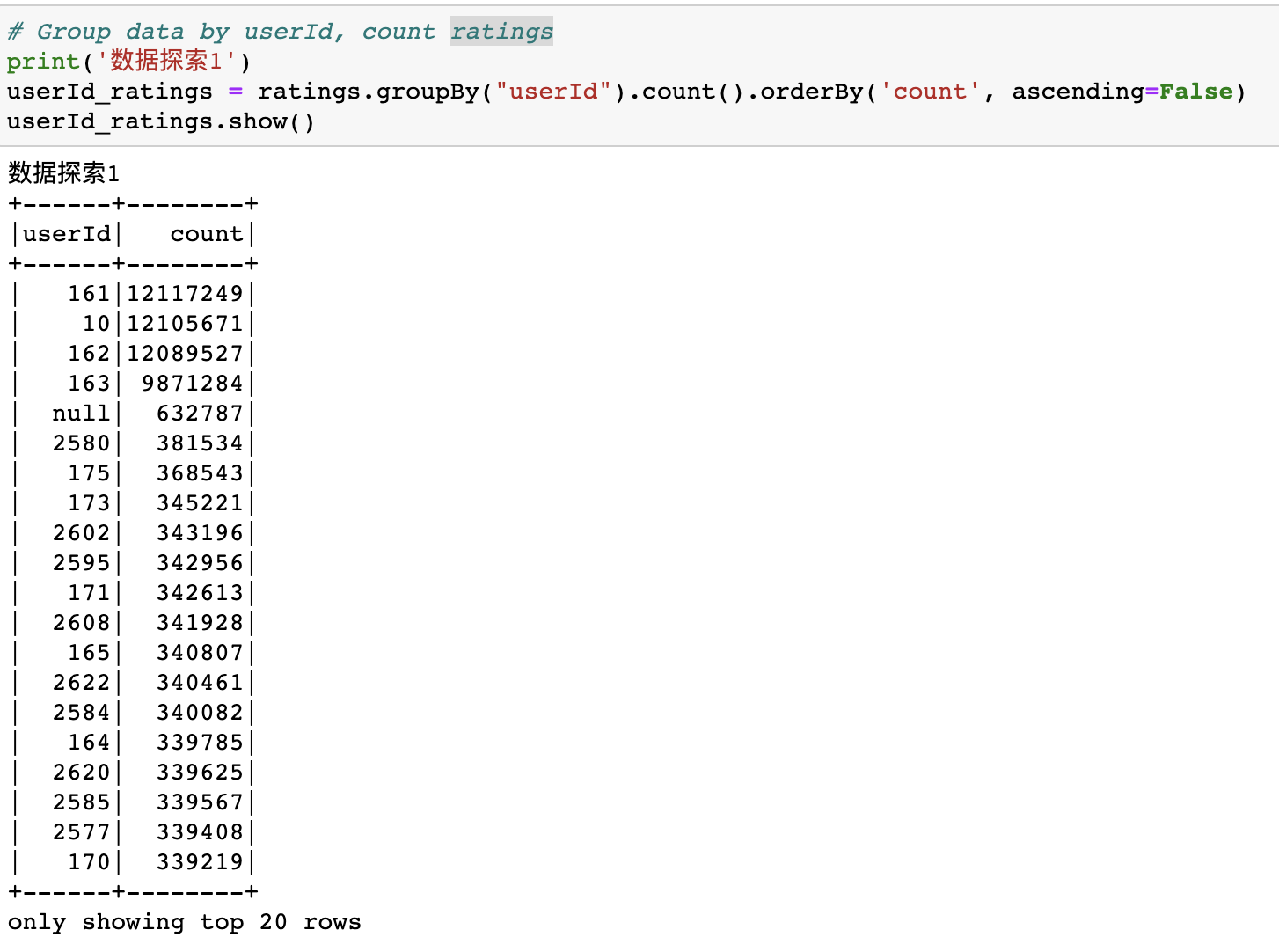


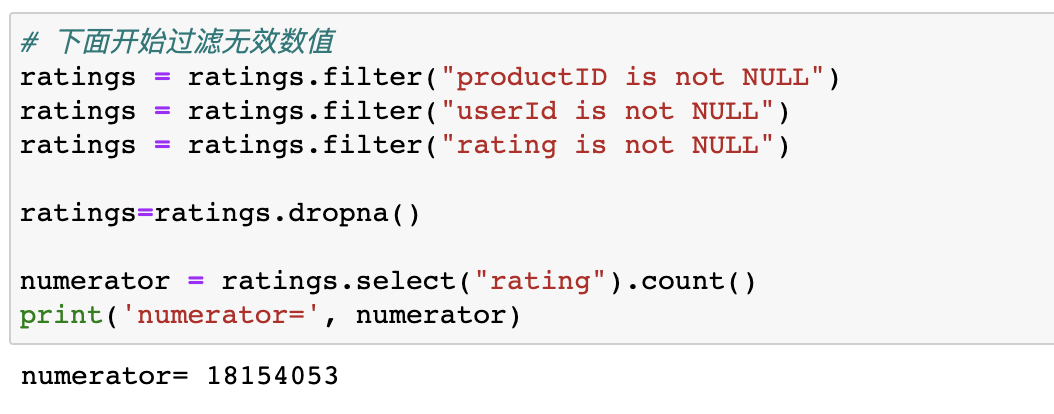
检查转换是否正确：

进行数据分析和探索：

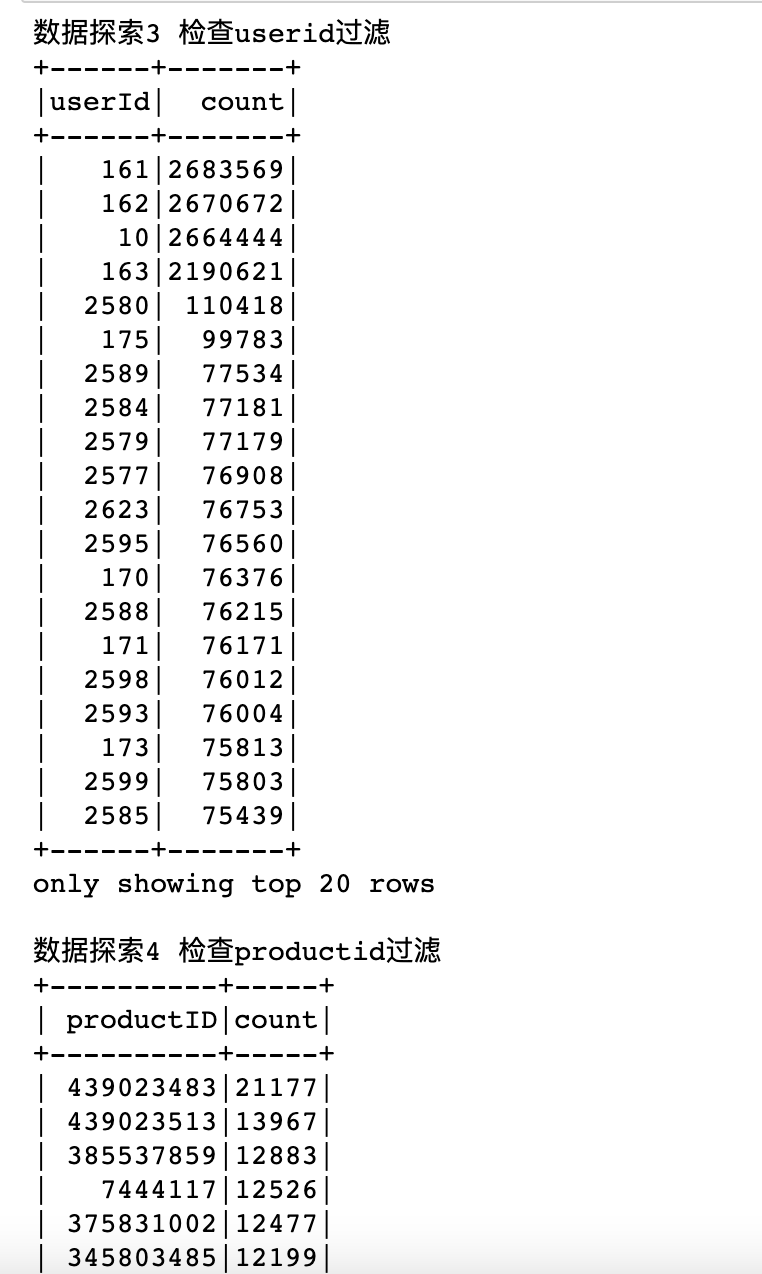


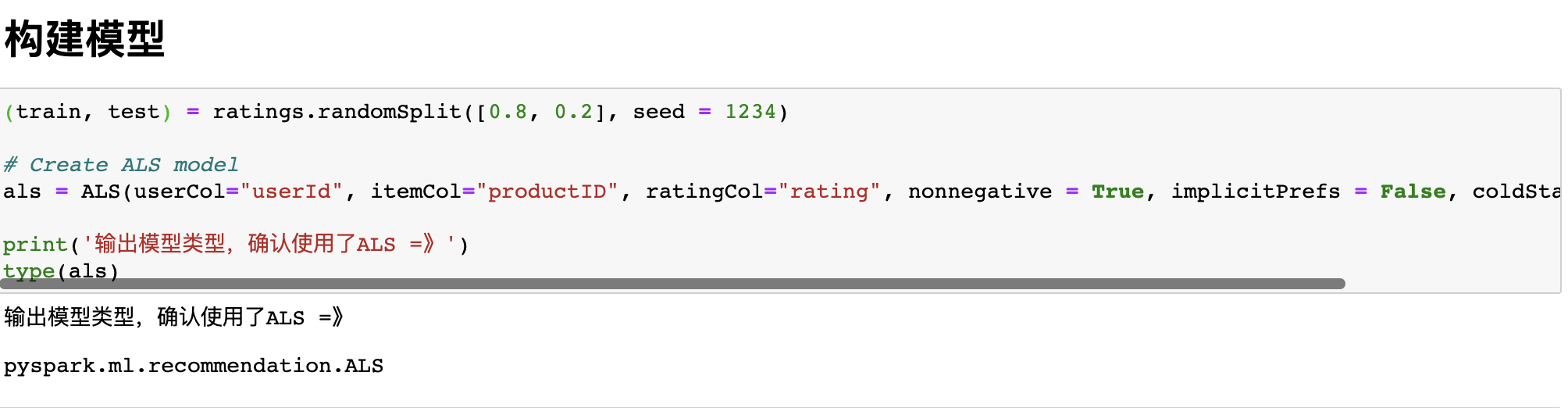


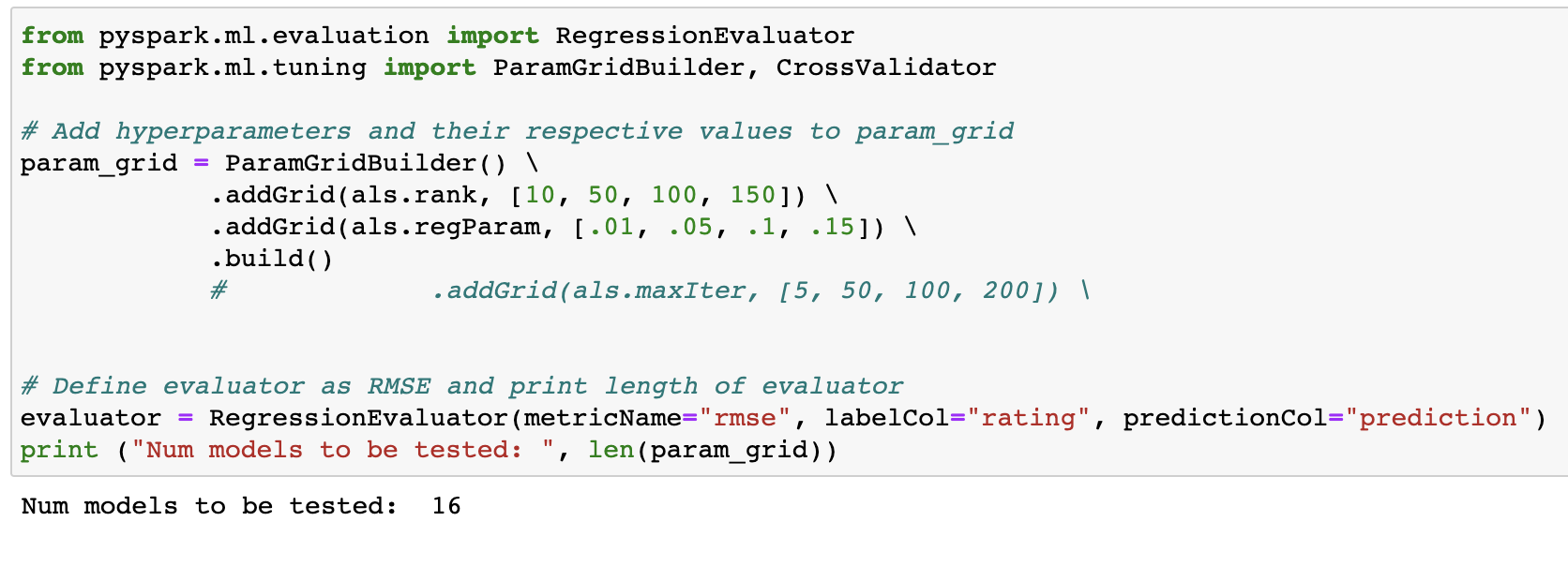


处理无效值并过滤之：

检查过滤和处理是否正确：



构建模型：



模型调优和验证