

1. 将数据导入
2. 看数据：重点看元数据，即对字段解释、数据来源等信息；导入数据后，提取部分数据进行查看
3. 缺失值清洗
  - 根据需要对缺失值进行处理，可以删除数据或填充数据
  - 重新取数：如果某些非常重要的字段缺失，需要和负责采集数据的人沟通，是否可以再获得
4. 数据格式清洗：统一数据的时间、日期、全半角等显示格式
5. 逻辑错误的数  
据
  - 重复的数据
  - 不合理的值
6. 不一致错误的处理：指对矛盾内容的修正，最常见的如身份证号和出生年月日不对应

不同业务中数据清洗的任务略有不同，比如数据有不同来源的话，数据格式清洗和不一致错误的处理就尤为突出。数据预处理是数据类岗位工作内容中重要的部分。