

模块二: 为数据探索做准备

可用数据文件

本模块提供多种格式的原始数据,帮助你理解数据收集和准备的多样性:

- raw_student_records.json - JSON格式学生基础信息
- survey_responses.txt - 非结构化文本格式问卷数据
- lms_access_logs.csv - 结构化LMS访问日志
- data_dictionary.md - 数据字典文档

核心教学内容: 多源异构数据处理

场景: 学校信息系统中存在多种格式的学生数据,如何统一整合?

演示内容:

- 展示JSON、CSV、TXT三种格式的差异
- 说明结构化 vs 非结构化数据的区别
- 讲解数据字典的重要性
- 演示如何将多种格式转换为统一的分析格式

教育场景延伸

结合教育数据特点进行讲解:

- 学生隐私保护: 展示数据中的敏感信息如何脱敏处理
- 数据伦理讨论: 收集学生学习行为数据的边界在哪里?
- 多源数据整合: 教务系统、LMS、在线平台数据如何匹配关联?