

模块六: 使用Python进行数据分析

可用数据文件

本模块提供真实规模的数据和完整代码示例:

- large_student_dataset.csv - 大规模学生数据集(可扩展至1000+)
- learning_logs_nested.json - 嵌套JSON格式学习日志
- clickstream_data.csv - 点击流行为数据
- python_analysis_tasks.py - 完整分析代码(6个任务)

任务1: 数据读取与基本探索

学习内容: pandas读取CSV, info()和describe(), 缺失值检查

任务2: 数据清洗与特征工程

学习内容: apply()数据转换, 成绩等级分类, 学习投入度指数构建

任务3: 分组聚合分析

学习内容: groupby分组统计, agg多维度聚合, 透视表概念

任务4: 数据可视化

学习内容: matplotlib直方图/箱线图, seaborn热力图, 多子图布局

任务5: JSON数据处理

学习内容: 读取解析嵌套JSON, 数据扁平化, 时序分析

任务6: 简单预测模型(可选)

学习内容: scikit-learn线性回归, 训练测试划分, 模型评估

AI时代延伸 - 重点!

AI辅助编程现场演示:

- 给AI一个提示词,让其生成数据清洗代码
- 展示AI如何解释代码和调试错误
- 对比AI生成代码 vs 自己写代码的效率

批判性思考:

- AI生成的代码可能存在什么问题?
- 如何验证AI代码的正确性?
- 理解原理 vs 使用工具,哪个更重要?