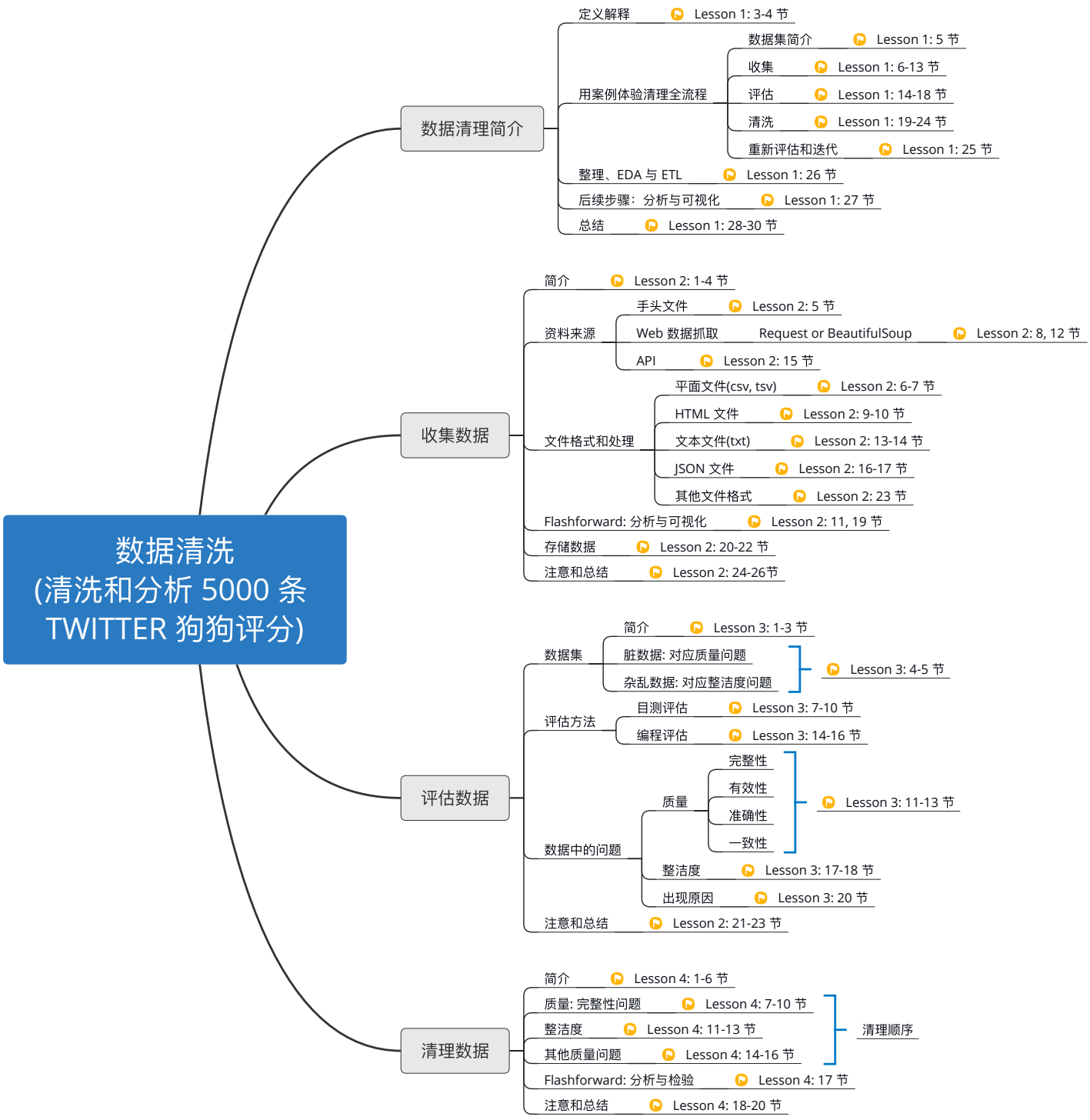
数据清洗
(清洗和分析 5000 条
TWITTER 狗狗评分)
数据清理简介
定义解释
Lesson 1: 3-4 节
用案例体验清理全流程
数据集简介
Lesson 1: 5 节
收集
Lesson 1: 6-13 节
评估
Lesson 1: 14-18 节
清洗
Lesson 1: 19-24 节
重新评估和迭代
Lesson 1: 25 节
整理、EDA 与 ETL
Lesson 1: 26 节
后续步骤：分析与可视化
Lesson 1: 27 节
总结
Lesson 1: 28-30 节
收集数据
简介
Lesson 2: 1-4 节
资料来源
手头文件
Lesson 2: 5 节
Web 数据抓取
Request or BeautifulSoup
Lesson 2: 8, 12 节
API
Lesson 2: 15 节
文件格式和处理
平面文件(csv, tsv)
Lesson 2: 6-7 节
HTML 文件
Lesson 2: 9-10 节
文本文件(txt)
Lesson 2: 13-14 节
JSON 文件
Lesson 2: 16-17 节
其他文件格式
Lesson 2: 23 节
Flashforward: 分析与可视化
Lesson 2: 11, 19 节
存储数据
Lesson 2: 20-22 节
注意和总结
Lesson 2: 24-26节
评估数据
数据集
简介
Lesson 3: 1-3 节
脏数据: 对应质量问题
杂乱数据: 对应整洁度问题
Lesson 3: 4-5 节
评估方法
目测评估
Lesson 3: 7-10 节
编程评估
Lesson 3: 14-16 节
数据中的问题
质量
完整性
有效性
准确性
一致性
Lesson 3: 11-13 节
整洁度
Lesson 3: 17-18 节
出现原因
Lesson 3: 20 节
注意和总结
Lesson 2: 21-23 节
清理数据
简介
Lesson 4: 1-6 节
质量: 完整性问题
Lesson 4: 7-10 节
整洁度
Lesson 4: 11-13 节
其他质量问题
Lesson 4: 14-16 节
清理顺序
Flashforward: 分析与检验
Lesson 4: 17 节
注意和总结
Lesson 4: 18-20 节