

肖正午

20岁 | 手机: 13537504508 | 邮箱: xzw13537504508@gmail.com | GitHub: github.com/Zev55555

教育背景

美国加利福尼亚大学圣地亚哥分校 University of California, San Diego

预计毕业时间: 2027

概率与统计 本科 | 大三 | GPA 3.7/4.0 (专业前15%)

相关课程: 人工智能、数据科学实践、数据分析与推断、概率论、随机过程、数理统计、回归分析、线性代数、应用线性代数、微积分与向量微积分、数学证明、计算机科学基础、Java 编程、商业基础

专业技能

AI 产品与 Agent workflow: 了解大语言模型 (LLM)、AI Agent、提示词工程 (Prompt Engineering)、RAG、工具调用 (Tool Calling) 等基础概念; 具备将模糊业务需求拆解为 AI workflow、指标口径、交互流程和功能模块的项目经验。

AI 应用评测与优化: 熟悉 Prompt 调优、输出质量评估、指标口径校验和异常案例回归测试; 能够围绕模型输出稳定性、可解释性和业务可用性设计测试样例。

数据分析与指标体系: 熟悉 SQL、Python、Pandas、DuckDB、Excel; 掌握指标拆解、漏斗分析、转化率分析、分子 / 分母口径定义、Top Movers 异动分析和辅助指标对比。

产品设计与需求分析: 具备需求分析、竞品调研、用户流程设计、PRD 撰写和原型设计基础; 能够将业务问题转化为可执行的产品方案, 并结合数据结果优化功能设计。

工具与协作: 熟悉 Figma、GitHub、VS Code、Jupyter Notebook、Cursor / Codex / Claude Code 等工具; 具备使用 AI 编程工具辅助完成原型开发、功能调试和产品验证的经验。

项目经历

SOVA AI | AI 指标异动分析 Agent

AI 产品 / Agent workflow / 数据分析工具 / 大模型应用

GitHub: <https://github.com/Zev55555/Sova-ai>

- 场景抽象:** 针对业务团队常见的“指标下跌但原因不明”问题, 将模糊业务需求拆解为指标澄清、字段识别、Metric Spec 生成、指标计算、异动拆解、证据链生成和报告输出等环节, 形成从问题输入到分析结论的结构化分析流程。
- AI 交互设计:** 设计 Step 1-10 分步式 AI 分析流程, 引导用户依次确认指标口径、对比周期、拆解维度、近期变化因素和数据字段需求, 降低用户直接写 SQL 或自行设计分析路径的门槛。
- 指标计算设计:** 将自然语言业务问题转化为结构化 Metric Spec, 明确指标公式、分子 / 分母字段、时间字段、拆解维度和辅助指标, 避免直接让 LLM 自由生成 SQL, 提高计算过程的可控性和可复查性。
- 数据分析链路:** 基于 Pandas 完成 CSV / Excel 数据读取与字段识别, 通过 DuckDB 执行安全聚合分析, 输出本期 vs 上期指标变化、分子分母变化、Top Movers、维度拆解和辅助指标对比。
- 多场景验证:** 基于物流、SaaS、客服、游戏、营销等场景构造测试案例, 验证 Agent 在不同行业指标下的字段识别、Metric Spec 生成、DuckDB 指标计算、证据链生成和报告草稿输出能力, 并记录字段误判、场景泛化和输出稳定性问题。

UCSD Triton Transit 校园班车晚间服务缺口诊断与排班优化

Python / SQL / DuckDB / Pandas / GTFS / 数据分析

GitHub: github.com/Zev55555/ucsd-triton-transit-service-gap

- 问题定义:** 针对 UCSD 校园班车晚间服务供给不足的问题, 将“学生晚间出行体验”拆解为时间段、路线、站点和资源优先级四个分析维度, 明确 17:00-22:00 的服务缺口诊断目标。
- 数据处理:** 基于 UCSD Triton Transit GTFS static feed, 使用 routes、trips、stop_times、stops、calendar 等核心表构建路线 - 站点 - 时间维度的服务分析数据集。
- 指标分析:** 使用 SQL / DuckDB 与 Pandas 完成数据连接、字段清洗、服务时间提取和分组统计, 计算每小时计划班次、估算发车间隔、晚间服务占比、末班车时间等核心指标。
- 可视化分析:** 通过路线 × 小时热力图、路线缺口评分图和站点覆盖分析, 发现晚间服务缺口主要集中在 20:00 后, 尤其是 21:00-22:00 时段。
- 业务建议:** 设计 Evening Service Gap Score, 结合发车间隔、末班车时间和晚间班次占比识别高优先级路线, 并提出增加晚间班次、延后末班车和补强重点站点覆盖等排班优化建议。

项目经历

AI Exposure 与编程职业就业前景分析

Python / Pandas / 数据分析 / 职业数据研究 | 进行中

- **研究设计:** 围绕“AI 暴露度是否与编程相关职业的薪资、就业规模和增长预期有关”设计分析问题，比较不同技术岗位在 AI 影响下的就业前景差异。
- **数据整合:** 整合 BLS、O*NET 和 AI exposure 数据，基于职业名称与 SOC code 构建职业层面的分析数据集，并完成字段清洗、变量标准化、缺失值检查和分组统计。
- **指标比较:** 从薪资水平、就业规模、职业增长率和 AI 暴露度等维度比较不同职业群体，观察高 AI 暴露职业与低 AI 暴露职业之间的差异。
- **结果解释:** 使用可视化和相关性分析解释 AI exposure 与就业指标之间的关系，并说明样本量、代理变量和非因果解释等局限，避免将相关性直接解释为因果关系。

InsightFlow AI 数据分析业务流程助手

AI 数据分析工具 / Prompt Engineering / 业务分析流程设计

GitHub: github.com/Zev55555/insightflow-ai

- **产品原型设计:** 面向数据分析新手设计 AI 工具型产品原型，帮助用户将模糊业务需求转化为分析问题、指标体系、业务假设、分析路径和汇报大纲。
- **Prompt 框架设计:** 设计结构化 Prompt 输出框架，使 AI 结果围绕业务问题重写、指标拆解、分析维度、可能假设、分析步骤和风险提醒展开。
- **功能实现:** 支持用户输入业务场景、分析问题和 CSV 字段信息，自动生成更接近真实数据分析工作的分析方案，并通过网页端完成输入、结果展示和基础交互流程。
- **项目延展:** 接入大模型能力搭建 AI 分析助手，重点参与产品定位、分析框架设计、Prompt 设计和测试反馈，为后续 SOVA AI 项目提供早期产品验证经验。