



联络方式

电话: +86 153-9702-6802
邮件: cheng.w@columbia.edu

个人网站: <https://chengwu-data.github.io/>
Github: <https://github.com/ChengWu-Data>

教育经历

哥伦比亚大学, 工程学院 | 纽约, 美国
数据科学硕士 (MS in Data Science)
方向: 数据工程、统计建模、预测分析、商业决策建模、AWS 与 BI 实战

2024.09 - 2026.12
GPA: 4.00 / 4.00

伊利诺伊大学香槟分校 | 伊利诺伊, 美国
计量经济学与定量经济学学士 (Bachelor of Science in Econometrics & Quant Econ)
辅修统计学 | 数据科学证书
方向: 因果推断、计量建模、时间序列、Python/R 实证分析

2021.09 - 2025.05
GPA: 3.90 / 4.00

工作经历

Tarte Cosmetics | 纽约, 美国
合规数据分析实习生 (Data Analyst Intern, Compliance)

2025.10 - 至今

- 深度梳理 7 套系统联通的SKU-HTS-报关-清关-审计全链路数据流 (ERP/报关系统/供应商系统/审计系统), 构建字段级血缘关系图 + 异常传播路径图, 定位12类结构性数据缺陷, 解释约40%审计延误根因
- 基于Python构建多规则校验引擎(HTS 语义匹配、价格区间约束、规格映射、字段完整度一致性)并结合异常检测模型(规则打分 + 离群阈值), 在~300条/周报关数据中实现100% 高风险错误自动拦截
- 设计合规预警与差异追踪仪表盘(供应商错误热力图、时间漂移监控、红旗分级机制、人工复核入口),支持 Ops 与合规团队实时定位源头责任方,将审核周期压缩35%
- 建立可溯源审计数据链路(时间戳、人工覆盖逻辑、规则触发路径、决策版本记录),支持复现级取证、审计追责与合规复盘自动化

Donglai Natural BioTech | 伊利诺伊, 美国
数据分析实习生, 市场与消费者行为

2024.06 - 2024.09

- 从 Web 日志与订单数据重构180 万行级完整用户行为路径, 统一埋点口径, 构建标准化转化漏斗(访问→浏览→加购→下单),精确定位关键流失节点与路径断裂点
- 基于价格弹性、探索深度、复购频率、渠道来源、页面停留时长构建 K-Means + DBSCAN 复合分群模型, 稳定识别5类高区分度行为人群
- 融合 N≈327 份消费者心理问卷 构建 行为-认知错配矩阵, 定量解释两类“高意向低转化”人群 的价值认知偏差机制
- 设计 准A/B实验 (叙事重构、价格透明、信任提示), 使用 Bootstrap + 比例检验 + 置信区间稳健性验证→ 转化率提升 10-12%, 核心客单价提升约 8%

Innovative Bloom Foundation | 上海, 中国
政策与影响力数据分析师

2019.05 - 2024.06

- 构建 6 年期多源纵向数据面板 (交易+家庭调研+区域宏观指标), 覆盖1200+ 农户, 实现户级-村级-区域级联动分析
- 采用 倾向得分匹配 (PSM) + 双重差分 (DiD) 进行政策效果评估:
→ 教育支出占比提升 18-25% → 家庭收入波动率显著下降
- 将 60+ 份深度访谈 编码为 行为约束因子 (性别规范、流动性约束、教育障碍), 补足纯量化模型的因果解释缺口
- 构建 “风险-回报-资源配置” 决策框架, 用于捐赠资金与教育资源最优配置, 使高风险家庭流失率下降 19%

浙商证券 | 杭州, 中国
数据分析师, 机构金融

2023.05 - 2023.09

- 构建 多因子选股模型 (财务质量、行为异常、市场微结构、政策事件因子), 用于中短期超额收益挖掘
- 进行 政策、再融资、并购事件 AR / CAR 事件研究, 输出 组合扰动区间与回撤风险评估
- 自动化 PE / M&A 月度风险监控系統 (VBA + SQL), 整合 敞口、流动性、集中度、阈值预警, 人工对账时间减少 70%
- 输出 风控与组合稳定性备忘录, 用于投资委员会仓位与再平衡决策

科研经历

哥伦比亚大学 | 多文档 RAG 智能决策系统 (项目负责人)

2025 - 至今

- 构建 端到端 RAG 系统 (PDF→分块→向量化→FAISS 检索→LLM 推理), 服务于 多源研报、合同、策略文档自动综合
- 引入 交叉编码器重排 + 幻觉校验 + 证据一致性约束, 事实一致性由 62% 提升至 87%
- 建立 系统性评估框架 (Recall@k、MRR、nDCG、幻觉率、噪声干扰度), 支持模型选型与检索策略对比

UIUC | Elizabeth T. Powers 教授研究助理

2023.04 - 2024.02

- 使用 DiD / IV / PSM 研究托育补贴对 劳动力供给与教育决策 影响
- 清洗与构建 50,000+ 户家庭纵向追踪数据集
- 输出多篇 政策效果评估与机制解释类分析简报

技能

分析方法: 因果推断 (DiD、PSM、IV)、回报率/风险/场景分析、行为分群、转化漏斗、留存分析、A/B 测试、事件研究

技术工具: Python、SQL、R、ETL、Spark、Dask、AWS (S3/Redshift)、Tableau、Power BI、Dash、Plotly、VBA

业务能力: 合规与风控分析、产品与市场策略分析、KPI 体系设计、异常监控与预警、跨部门沟通与数据决策支持