

《人工智能与旅游创新应用》

旅游数据分析 与预测实训

模块四 生成式人工智能的旅游应用实训

旅游管理专业核心课程

主讲教师： ____ · 2026 秋季学期

MODULE 04 — PROJECT 08



CONTENTS · 本讲目录

让数据 开口说话

ASK · CLEAN
MODEL · DECIDE

01 实训导航

任务书 · 考核标准 · 分析三问 · 数据准备

02 分析流水线

分析五步 · 方法对照 · 清洗四查 · 好差之辨

03 预测实战

流程对比 · 预测三法 · 三周节奏 · 实战复盘

04 汇报复盘

反面案例 · 课堂辩论 · 红线底线 · 成果

实训目标

成果导向

真数据 真结论

REAL DATA
REAL DECISION

知识目标

能说出**分析五步流程**与预测三法的适用边界。

能力目标

拿一份真实经营数据**走完全流程**：从提问到结论，产出可执行的运营建议。

素养目标

建立「**结论有边界**」意识——相关性不是因果，数据会说谎：严谨是分析师的职业道德。

一份表格救了一个淡季

山间民宿老板娘的纠结——十一月到底该歇业还是硬撑？**往届学生拉出三年数据**：周末入住率68%，**塌的是周中不是周末**。不该关店，该做周中团建套餐。

老板娘为什么判断错了三年？

数据没变，为什么结论变了？

如果你只有一份 Excel，从哪问起？

实训主线：**提问 - 取数 - 清洗 - 分析 - 结论**，三周一份真数据。



01

PART

实训 导航

THE BR

① 任务书 ② 考核标准 ③ 分析三问 ④ 数据准备

实训任务书：三周一份真结论

数据实训不是「跑个模型」——**一份能落地运营建议的数据报告**才是交付物：决策者会用行动投票。

—— 综合主流教材与业界通行表述

任务

DATA

小组认领**一份真实经营数据**：合作民宿、公开统计或自采问卷均可。

流程

FLOW

提问 - 取数 - 清洗 - 分析 - 结论**五步全走**：每步留过程文档。

交付

REPORT

产出**数据报告加运营建议**：一条建议必须落到具体动作。

成绩 = 过程文档 30%+ 分析质量 40%+ 汇报答辩 30%—— **只有图表没有结论，直接不合格。**

考核怎么算：三组数字

可执行的运营建议

1 条

结论必须落到动作——「**建议优化**」不算数：什么时候对谁做什么，三要素齐。

5 步

流程环节留痕数

提问到结论五步文档——**过程占 30%**：跳步 = 不合格。

3 版

核心图表迭代版数

主图表至少三版——**好图表是改出来的**：一稿过的多半没想清。

8 分钟

汇报答辩时长

结论先行限时讲——**超时扣分**：讲不清说明没想清。

分析三问：从看数到算命

问一 · 发生了什么

描述

入门

- 均值、占比、趋势
- 先把数看懂
- 多数报告停在这

问二 · 为什么

诊断

进阶

- 对比、细分、归因
- 找到差异的来源
- 业务的听诊器

问三 · 接下来呢

预测

高阶

- 从规律到预判
- 支撑今天的决策
- 分析的终点

实训要求 三问都要走到**预测**——描述是及格、诊断是良好：本实训的考核点压在第三问。

数据准备：三件事

拿数据 · 三个来源

合作企业脱敏数据、文旅局公开统计、自己发的问卷——**先有小数据**

再谈大数据：够用就好。

认数据 · 口径为王

每个字段什么意思、怎么统计的——**口径不清全盘皆输**：「游客」和「人次」差一个世界。

提问题 · 分析之始

带着问题看数据——**没有问题的分析是漫游**：老板娘的问题就一个，十一月该不该关。

数据体检三问

这份数据能回答我的问题吗？

字段口径都问清了吗？

时间范围覆盖完整吗？——**三问过了再动手**

新手最大的坑是**先动手后提问**——图表做了一堆，问题还是那个问题。

数据素养：四条军规

分析师修养

四条军规

工具可以速成——**素养决定你能走多远。**

四条军规全程适用。

01

口径先行

先问数怎么来的——**同一指标不同口径**：比较是分析的地基。

02

样本意识

看到数先问谁没被统计——**沉默的数据不说话**：幸存者偏差防不胜防。

03

相关慎言

相关不等于因果——**冰淇淋和溺水同涨**：夏天才是共同原因。

04

结论有界

数据只覆盖它的范围——**外推要打折**：三年的数据预测不了黑天鹅。

实训导航：四问四答

Q1 实训的交付物是什么？

一份有可执行建议的数据报告 + 五步过程文档 + 限时汇报。

Q2 分析三问指什么？

发生了什么、为什么、接下来呢——描述、诊断、预测：考核压在第三问。

Q3 数据从哪来？

合作企业、公开统计、自采问卷——先有小数据：够用就好。

Q4 分析师的四条军规？

口径先行、样本意识、相关慎言、结论有界——素养决定走多远。

02

PART 1

分析 流水线

THE PIPE

① 分析五步 ② 方法对照 ③ 清洗四查 ④ 好差之辨

分析五步：从问题到结论

步 ①

提问

定方向

把业务问题翻译成数据问题——「**该不该关店**」变成「**入住率结构如何**」：问对了就成了一半。

步 ②

取数

备原料

选字段定范围——**三年十一月逐日**：范围围着问题转。

步 ③

清洗

保质量

四查缺失重复异常口径——**脏数据进垃圾出**：这步占全流程六成时间。

步 ④

分析

出洞察

对比细分找差异——**周末 68% 对周中 22%**：差异即洞察的矿脉。

步 ⑤

结论

给动作

一句话结论一条建议——**什么时候对谁做什么**：决策者只看这页。

五步是流水线——**两头是业务，中间才是技术**：从业务中来，回业务中去。

方法对照：问题对方法

问题类型	常用方法	典型问题	产出
有多少	描述统计	入住率、客单价多少	一张经营底片
差在哪	对比细分	周末和周中差多少	差异定位
为什么	归因分析	差的原因是什么	因果线索
会怎样	趋势外推	下月客流多少	预测区间

方法跟着问题走——**别拿着锤子找钉子**：先问要回答什么，再选什么方法。

清洗四查：脏数据的体检

质检流程

四查过关

清洗占全流程六成时间——**最枯燥也最要命。**

四查一项不能省。

01

查缺失

哪些字段是空的——**空得有没有规律**：系统性缺失是预警信号。

02

查重复

一单录了两次吗——**重复即虚增**：先把水分挤干净。

03

查异常

负数房价、千岁游客——**异常值先查录入错**：别急着删。

04

查口径

今年含税去年不含——**口径不齐比较失效**：统一后再分析。

好差之辨：两张图

差图表 · 技巧炫技

"3D 饼图十二个色块"

立体效果、彩虹配色、图例盖住数据——**看图三分钟找不到重点**：炫技是给同行看的。

好图表 · 结论直达

"周末撑住了整个十一月"

标题即结论、一图一观点、色不过三——**三秒读到洞察**：图表是给决策者看的。

图表的唯一标准：**外行三秒能懂**——决策者没空研究你的图例：把话说清楚是本分。

AI 分析：能替与替不了

能替的 · 体力环节

" 代码已生成，请运行 "

写清洗代码、跑统计模型、生成图表初稿——**技术门槛塌方**：
不会代码也能做分析。

替不了的 · 判断环节

" 这个数该信吗 "

提对问题、选对指标、读出业务含义——**AI 给输出，人给解读**：数字背后的生意只有你懂。

AI 塌掉的是技术门槛——**问题意识与业务解读成了新门槛**：这是分析师的护城河。

先算账：一份报告的时间

传统：手工清洗 + 查表计算 + 排版制图 **1 周**

1 天

AI 辅助：代码生成 + 自动可视化（学生实训实测带）

一份报告从 1 周到 1 天

快在

清洗 四查代码一键生成——**六成时间压缩到一成**：枯燥环节大幅自动化。

快在

制图 图表初稿自动出——**人只管改标题**：标题即结论那句话。

花在

解读 省下的时间砸进业务——**数字背后的原因**：AI 算得出差 32% 说不出为什么。

省下的时间不是让你早下班——**是用来把结论磨准**：分析的价值全在最后一步。

分析流水线：四问四答

Q1 分析五步的顺序？

提问 - 取数 - 清洗 - 分析 - 结论——两头是业务：从业务中来回业务中去。

Q2 清洗为什么占六成时间？

脏数据进垃圾出——四查一项不能省：地基不牢全盘皆输。

Q3 好图表的标准？

外行三秒能懂——标题即结论、一图一观点：炫技给同行看没意义。

Q4 AI 分析的边界？

替体力不替判断——问题、指标、解读在人：这是护城河。

03

PART TI

预测 实战

FOREC /

-
- ① 流程对比
 - ② 预测三法
 - ③ 三周节奏
 - ④ 实战复盘

工作方式：传统 vs AI 辅助

传统分析 · 手艺密集

环节	做法
清洗	Excel 手工逐行筛
计算	查表按计算器
建模	统计软件高门槛
制图	一图一图手工调

AI 辅助 · 判断密集

环节	做法
清洗	四查代码一键生成
计算	自然语言出结果
建模	AI 搭模型人验假设
制图	初稿自动出人改标题

方式变了，**严谨的标准没变**——口径、样本、边界：变的只是抵达速度。

预测三法：按样本量选

法一 · 移动平均

看趋势

最简单

- 近几期均值外推
- 数据少也能用
- 波动被抹平是弱点

法二 · 回归分析

找关系

最常用

- 找自变量定系数
- 解释性强好汇报
- 样本几十条起步

法三 · 机器学习

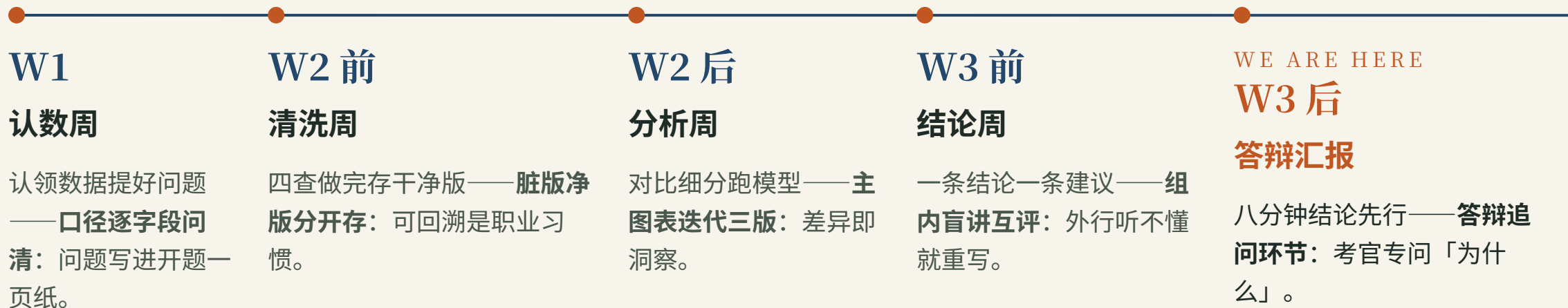
喂样本

最黑箱

- 大样本才发挥
- 精度高但难解释
- 决策者常不买账

选型建议 实训数据量**推荐回归起步**——能解释、好汇报：机器学习留给样本够的小组，简单方法先跑通。

三周节奏：分析日历



节奏的意义：**分析不过夜、结论不过周**——拖到最后一周的组，连清洗都做不完。

答辩评分：三个维度

汇报答辩时长上限

8 分钟

结论先行限时讲——**超时即扣分**：
电梯里讲不清的结论不算想清了。

严谨

方法与口径

四查做没做、口径问没问——**地基分占大头**：歪楼再漂亮也是歪的。

洞察

结论的价值

有没有「没想到」——**常识复读机不得分**：数据说了别人不知道的事。

落地

建议可执行

什么时候对谁做什么——**三要素齐**：建议落不了地等于零。

实战复盘：淡季不该关

CASE · 实战实录

任务：往届小组「山间民宿淡季决策」全流程分析

任务 三年入住数据，回答「十一月该不该歇业」

成果 结论推翻老板直觉，建议落地后淡季增收

- **提问翻译（定问题）**：「该不该关」翻译成「收入结构能否覆盖成本」——**业务问题变数据问题**：这一步定生死。
- **细分分析（找差异）**：按周末周中拆开——**周末 68% 对周中 22%**：塌的是周中，关店是误杀。
- **落地建议（给动作）**：周中做团建套餐、周末保价——**两个动作两个客群**：建议具体到执行人。

效果

套餐上线后周中入住翻倍——**数据结论变成真金白银**：这是分析的最高褒奖。

教训

初版只在总量打转——**平均入住率 40% 看着确实该关**：不细分就等着误判。

启示

总数会骗人结构不会——**细分是洞察的第一来源**：一个维度拆下去答案自现。

两种决策：关店与拆分

拍脑袋决策：十一月歇业（错过周末客流示意）**亏 8 万**

增收 5 万

数据决策：周中套餐 + 周末保价（落地后实测）

从亏 8 万到增收 5 万

差在

视角 把淡季当一个数——平均数掩盖了结构：一刀切决策必误伤。

赢在

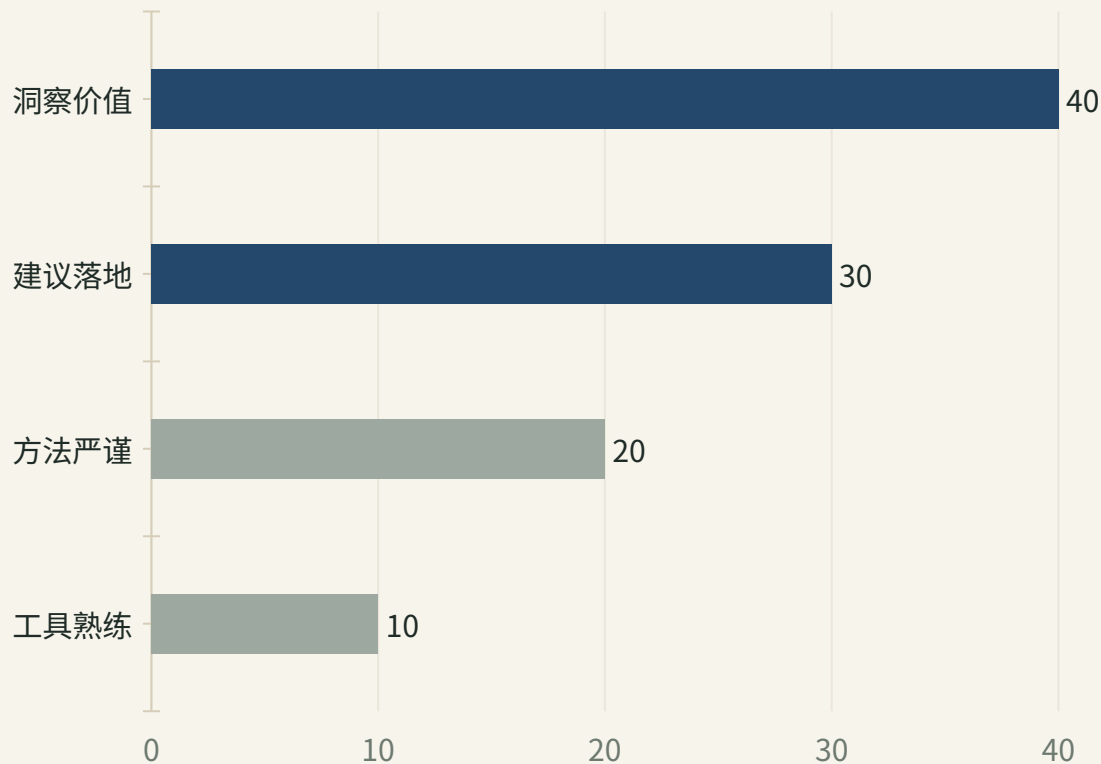
细分 按客群拆开看——该保的保该救的救：精准施策才是数据分析的本意。

提醒

边界 结论基于历史三年——疫情等黑天鹅不在其中：预测要留置信区间。

数据不是替老板做决定——**是把选项摆清楚**：决策权在人，知情权在数据。

答辩评分权重



看清权重再分配精力——**洞察与落地是大头**：别把三周全花在跑模型上。

40%

洞察价值占四成——**说出了别人不知道的**：数据分析的存在理由

30%

建议落地占三成——**三要素齐才算数**：落不了地等于零

10%

工具熟练只占一成——**AI 代劳了**：值钱的从来不是手速

往届三坑，对症下药

坑① 图表炫技

十八张图塞进汇报——**每张都半懂不懂**：图表是论据不是装饰。

坑② 口径混乱

今年含税去年不含——**对比建立在流沙上**：结论越漂亮越危险。

坑③ 结论悬空

「建议加强营销」——**正确的废话**：没有动作的建议是安慰剂。

对症下药

一图一论→ 砍到三张图以内

口径卡→ 每个指标建一张口径卡

三要素→ 时间对象动作缺一不可

预测实战：四问四答

Q1 预测方法怎么选？

按样本量——回归起步最稳：能解释好汇报，机器学习留给大样本。

Q2 答辩评分看什么？

洞察 40%+ 落地 30%——工具只占 10%：别把时间全花在模型上。

Q3 民宿案例赢在哪？

细分出结构——周末周中拆开：总数会骗人结构不会。

Q4 结论的边界？

基于历史留区间——黑天鹅不可测：预测要带置信度。

04

PART F

汇报 与复盘

REPORT & DE

① 反面案例 ② 课堂辩论 ③ 红线底线 ④ 成果三件

反面复盘：一个被推翻的结论

CASE · 反面教材

任务：「年卡客群最该主推」为什么被答辩组推翻（往届实录）

现象 会员数据显示年卡客消费最高，
小组建议 All in 年卡

代价 答辩现场被一个提问击穿，结论
全盘返工

- **原逻辑（看似成立）**：年卡客客单价最高——**高价值客群锁定年卡**：数据支持逻辑自治，答辩前信心满满。
- **致命提问（现场翻车）**：考官问「没办年卡的流失客人，在数据里吗」——**全场沉默**：数据里只有留下来的人。
- **重新审视（补视角）**：这是典型幸存者偏差——**流失者的沉默被当成了不存在**：样本边界没问过自己。

教训

数据里的高消费是结果不是原因——**先有忠诚才办卡**：因果搞反了方向。

效果

补充流失访谈后重做——**结论改为分层运营**：翻过车的报告才扎实。

启示

看到数先问谁不在数里——**沉默的数据最响**：样本意识是分析师的防弹衣。

课堂辩论：信数据还是信经验

老店长凭三十年直觉说「这批客人不会来」，数据模型说「复购概率 72%」——听谁的？

——议题背景

数据派

- 直觉说不清为什么——**数据可复验可追溯**
- 人记不住三万条记录——**模型记得住**
- 经验会过时——**数据反映的是当下**

经验派

- 数据只有历史——**直觉里藏着没量化的变量**
- 口径陷阱防不胜防——**老法师一眼看出数不对**
- 决策要担责——**模型不签字**

VS

讨论规则：双方必须回答——「**预算只够养一个岗位，你要数据分析师还是金牌店长**」？**落到用人决策场景讨论。**

数据红线：三条底线

隐私

客人信息先脱敏

姓名手机号身份证——分析前一律匿名化处理

来源

数据获取须合规

企业数据签授权、公开数据看许可——不碰灰色爬虫

表述

相关不得说成因

结论措辞守边界——「与…相关」不写成「因为」

一条红线

比结论更贵

隐私与合规是职业生命线——

红线意识从第一份数据养成。

汇报四步：八分钟讲透

汇报流程

四步讲完

汇报不是展示劳动量——**是推销你的结论。**

结论先行贯穿始终。

01 抛结论

第一句话就是结论——**「淡季不该关该拆分」**：先给答案再给证据。

02 给证据

三张图各撑一个论点——**标题即结论**：图表是论据不是装饰。

03 说边界

主动交代局限——**样本范围与不适用场景**：诚实加分不扣分。

04 留追问

备好数据附录——**考官追问「为什么」时**：底牌厚才敢让人提问。

作品集三件套

真实数据的分析报告

1 份

一份真数据的真结论——**胜过十份模拟练习**：面试时讲得出细节。

报告

数据报告与建议

结论、证据、边界三段式——**一条可执行建议**：这是主件。

文档

五步过程档案

提问到结论全留痕——**口径卡与清洗日志**：严谨性的证据。

数据

脱敏数据集

清洗后的净版数据——**附数据字典**：可复现的分析才可信。

带走的清单：三周六件事

数据分析冲刺，六件事按序做：

01 提问题 业务问题翻译成数据问题

02 认数据 口径逐字段问清建卡

03 做清洗 四查过关，脏净分存

04 跑分析 对比细分，主图三版

05 磨结论 一条结论一条建议三要素

06 练答辩 八分钟盲讲，外行能懂

四个问题，四个答案

Q1 数据分析的五步？

提问 - 取数 - 清洗 - 分析 - 结论——两头是业务：从业务中来回业务中去。

Q2 AI 改变了分析什么？

技术门槛塌方——问题意识与业务解读成新门槛：这是护城河。

Q3 最容易翻车的点？

样本与口径——幸存者偏差防不胜防：先问谁不在数里。

Q4 这段经历怎么写进简历？

写**决策影响**——从亏 8 万到增收 5 万：数据变成钱的故事。

作业：报告与复盘

思考题 · 下次课开篇讨论

想一想 01 民宿案例里，如果你只能再要一个字段，你要什么？为什么？

想一想 02 幸存者偏差的反面案例，你生活中还在哪见过（App 评分、好评榜）？

课后作业 · 实践任务

数据报告

提交完整数据报告：问题定义、口径卡、四查记录、三版主图表、一条三要素齐全的运营建议
脱敏数据与字典必附

答辩录像

提交八分钟汇报录像：结论先行、证据三图、边界交代与追问预案
组内互评分须附

课程平台提交，优秀报告推荐至合作企业

实训资源包

工具

表格与 AI 分析

电子表格 +AI 代码助手——各练熟一套。

数据

文旅公开数据

文旅统计公报——练手的真实数据源。

范例

优秀报告拆解库

往届答辩最佳——看结论怎么写。

理论

统计思维入门

偏差与谬误清单——防身必备。

每周拆一份文旅统计公报，写三行洞察——数据感是喂出来的。

项目一：促进旅游业发展的应用展望

MODULE 05 · PROJECT 01

抬头 看路

从实训到展望

01 行业蓝图

AI 重塑旅游业的全景图——**看清你在哪一站。**

02 趋势研判

哪些岗位在消失哪些在生长——**提前卡位。**

03 职业规划

把二十六项目连成你的路线图——**毕业去向早想。**