



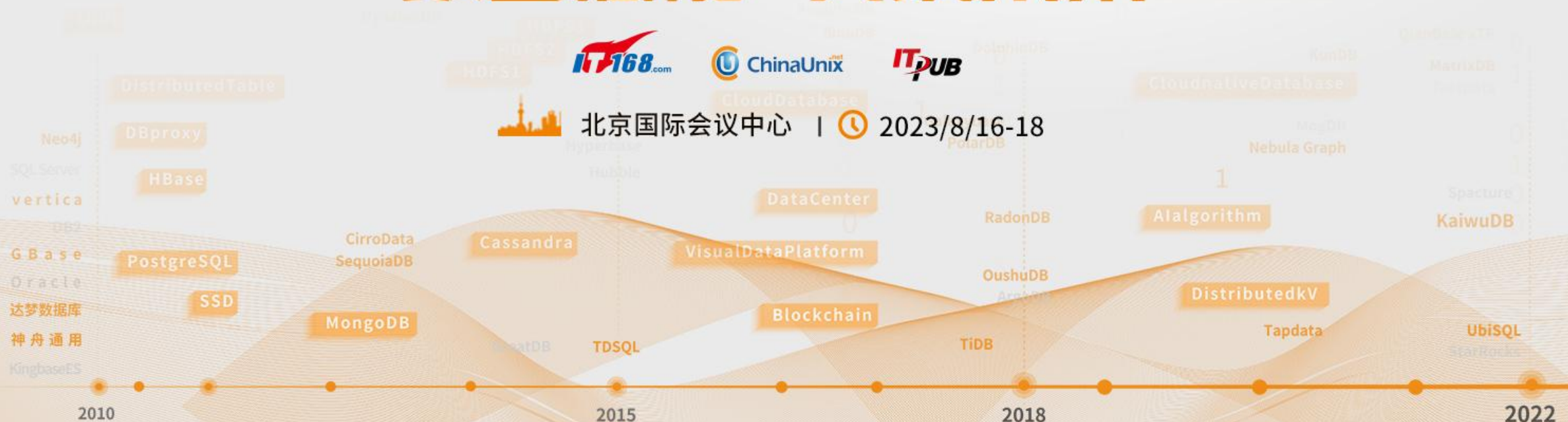
第十四届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA

数智赋能 共筑未来



北京国际会议中心 | 2023/8/16-18



归因分析在去哪儿网的应用落地

公司：去哪儿

职位：数据挖掘方向负责人

姓名：齐凯凡

CONTENTS

为什么做归因

背景
痛点
价值

01

归因方法论

探索归因
归因方法论
关键技术点

02

归因应用场景

方法论的流程落地
转化率指标的归因分析

03

未来规划

业务
技术
收益

04

为什么做归因

“分析一下这样做能不能成功？”

“为什么别人做能成功，我一做就失败？”

营销也有大学问

新品上市，该如何分析

用户流失率高了，该如何分析

如何提高销售业绩

“帮忙分析一下我的财务状况吧。”



如何归因？



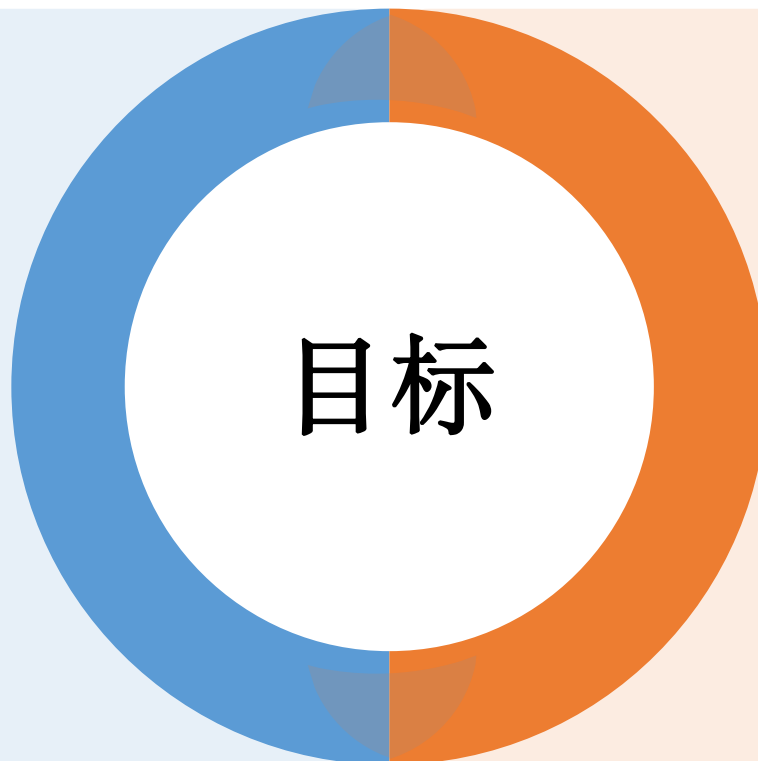
归因流程及痛点



归因目标

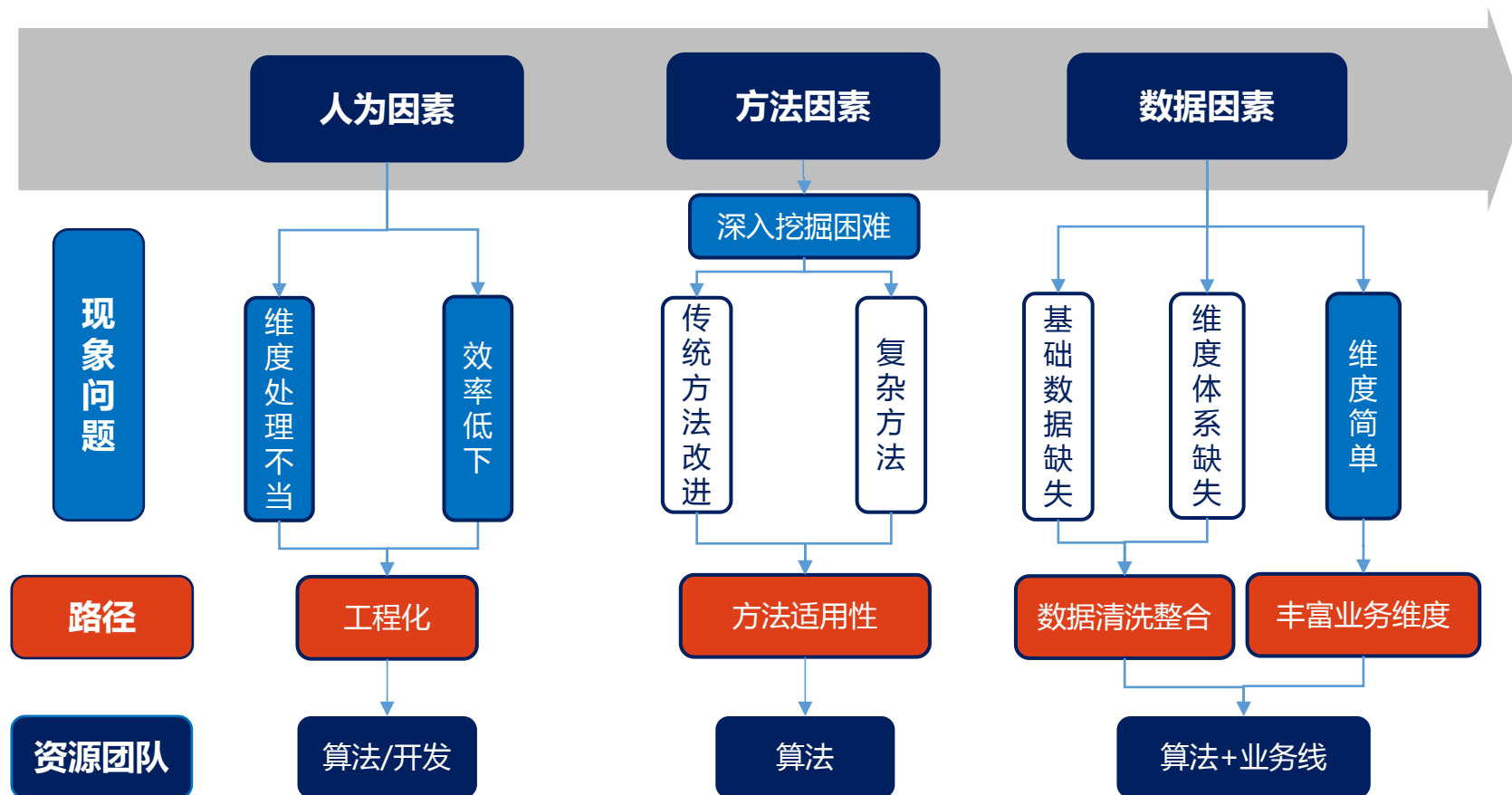


提效率

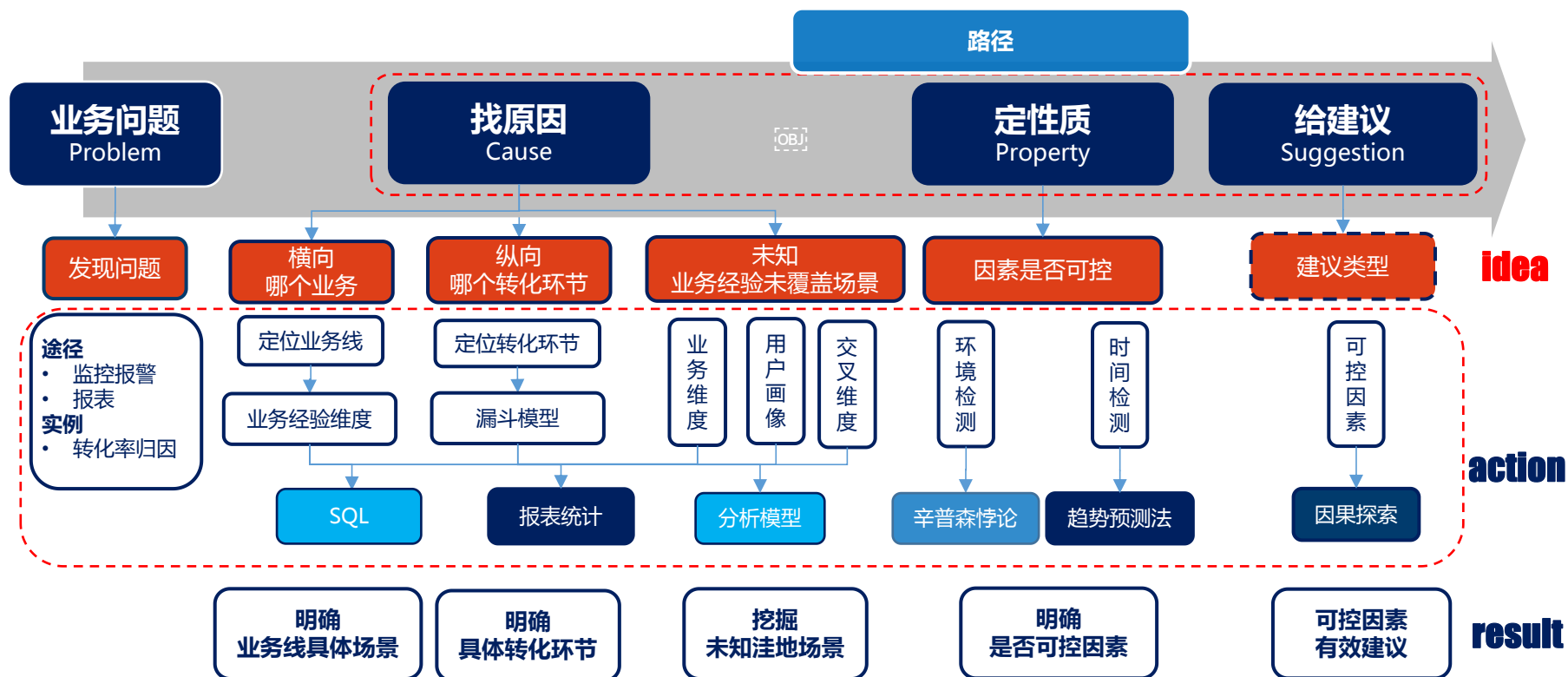


提质量

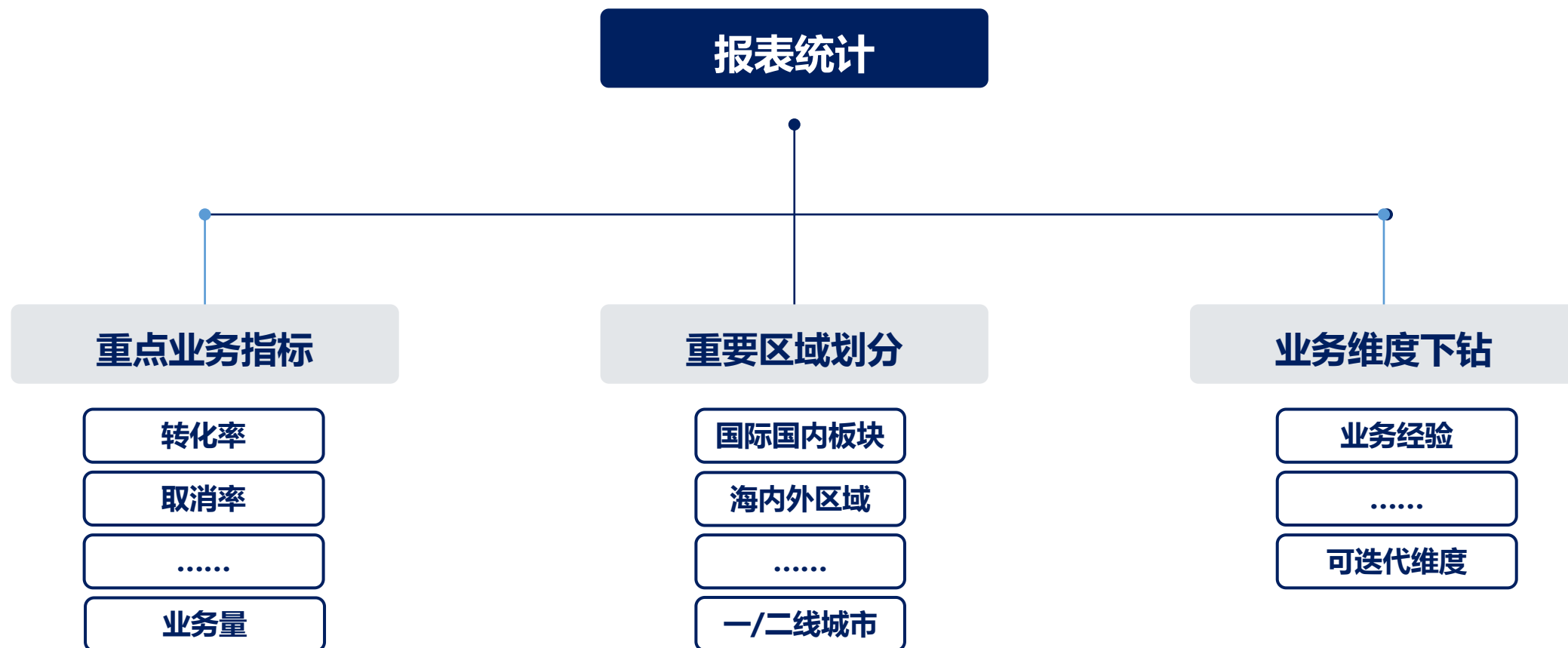
探索归因问题对应解决路径



归因方法论流程



定位问题原因-报表统计



识别环境因素-辛普森悖论

定义：英国统计学家E.H.辛普森（E.H.Simpson）于1951年提出的悖论，即在某个条件下的两组数据，分别讨论时都会满足某种趋势，可是一旦合并考虑，却可能导致相反的结论。

辛普森悖论经典案例

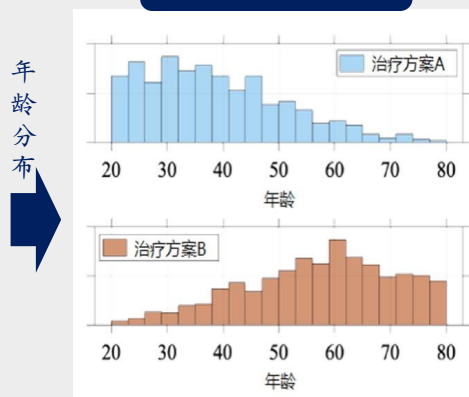
新治疗方案的治愈率是否更高？

治愈率	新方案A	旧方案B
年轻人	234/270=87%	81/87=92%
老年人	55/80=69%	192/263=73%
总体	289/350=83%	273/350=78%

总体：方案A是优于方案B
分年龄：方案B优于方案A

辛普森悖论应用

人群结构分布



年龄分布存在明显差异

解释辛普森悖论

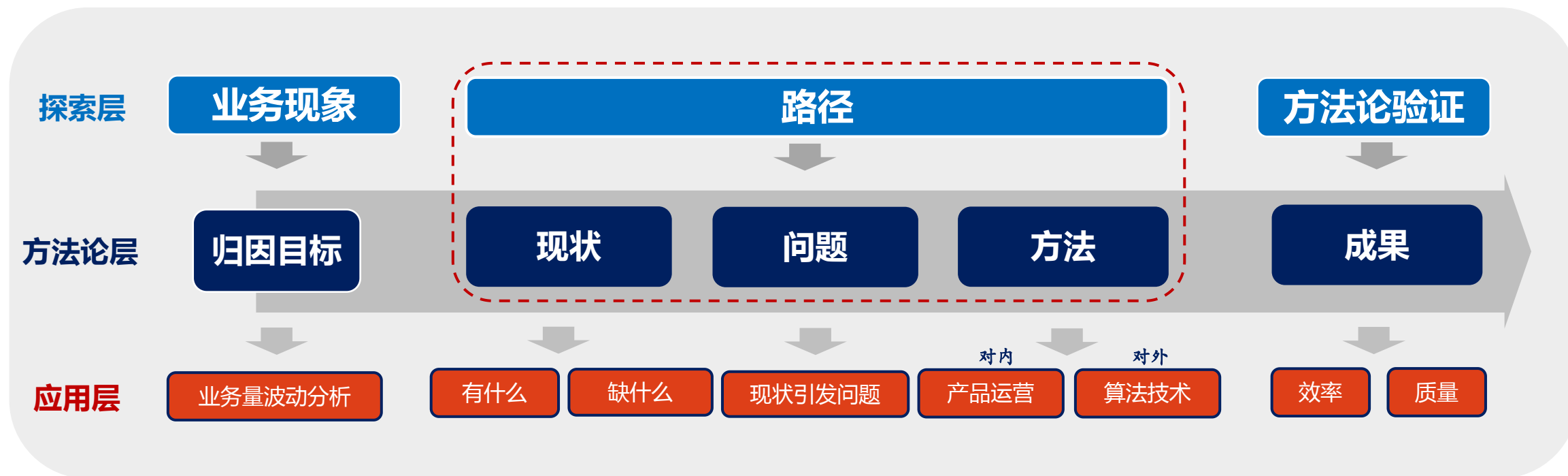
方案A更偏年轻化，
而年轻人的药物治疗效果更好，
所以出现方案A优于方案B的结果

人群结构分布的差异导致相反结论

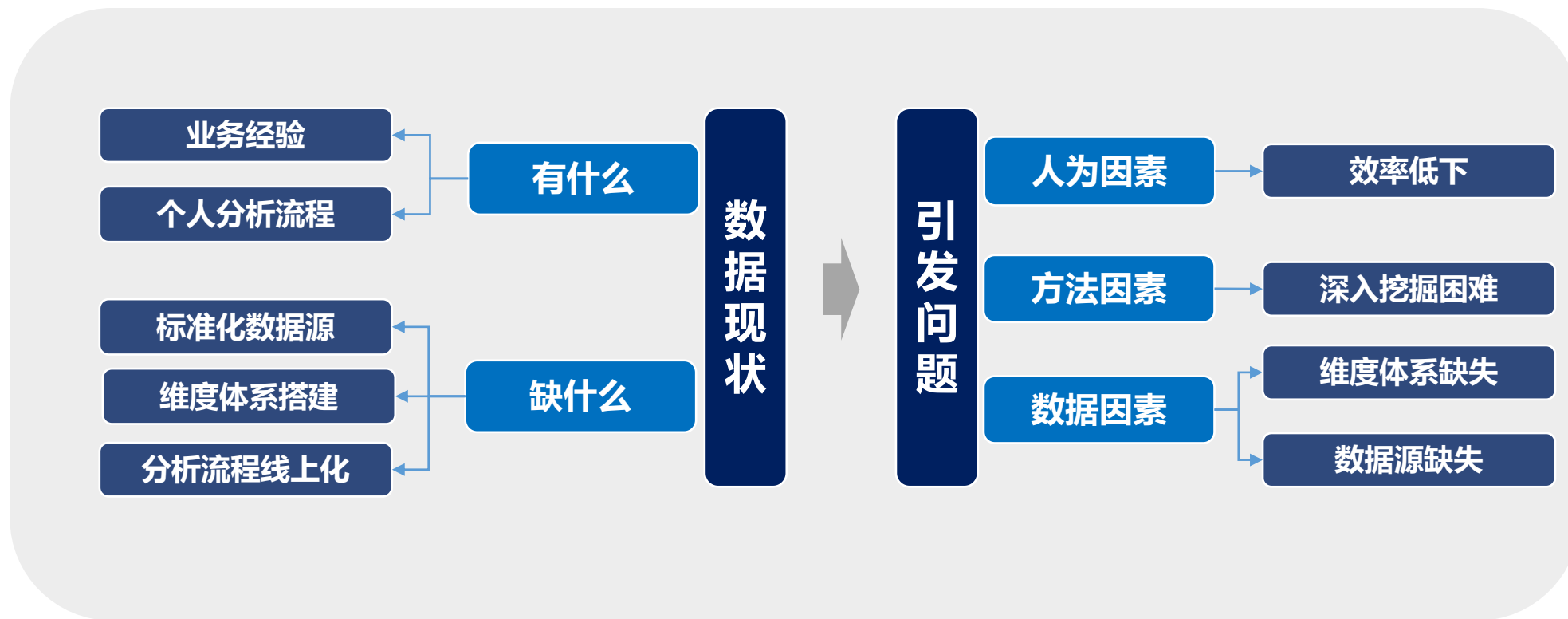
方案A和方案B存在差异
这个差异中哪些是环境因素，哪些是非环境因素引起的
运用辛普森悖论识别环境因素

实例一-业务量波动归因

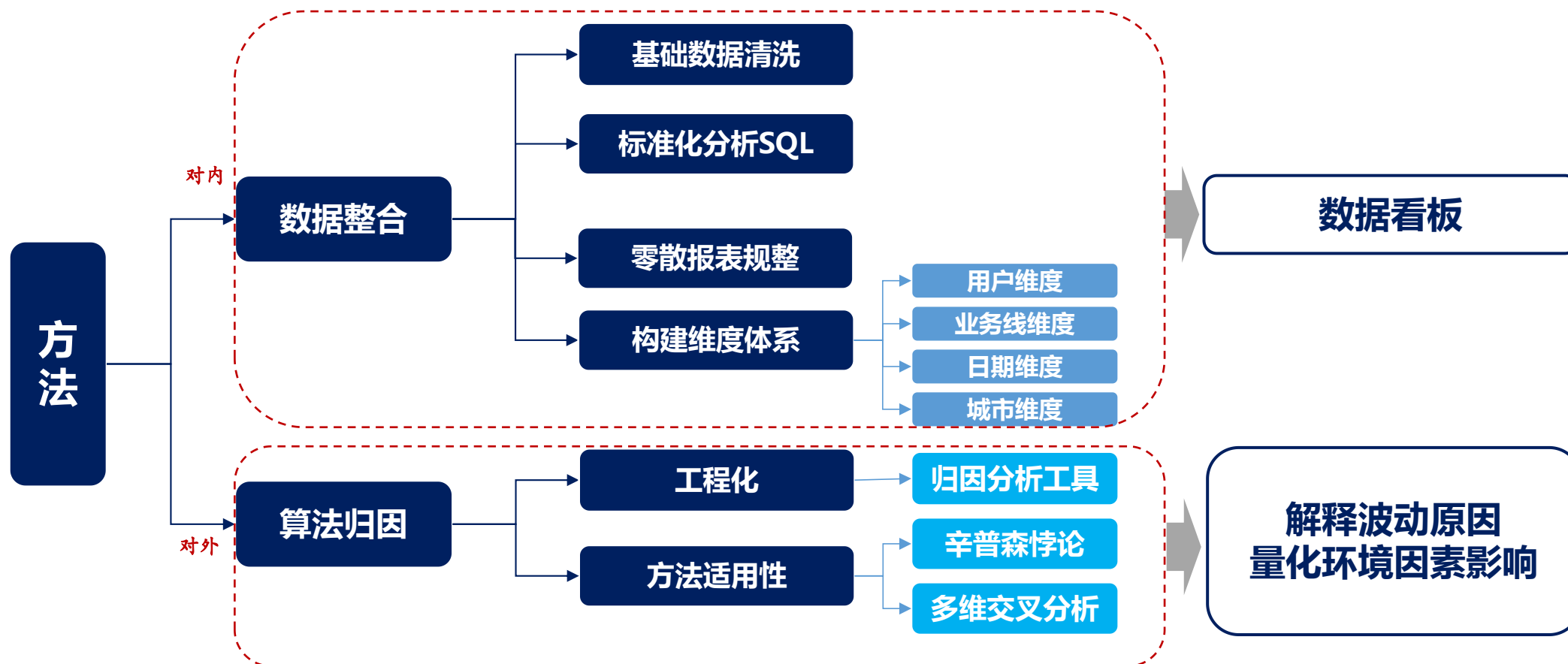
需求背景：基于归因分析探索到方法论建立的流程，对业务量波动指标归因分析



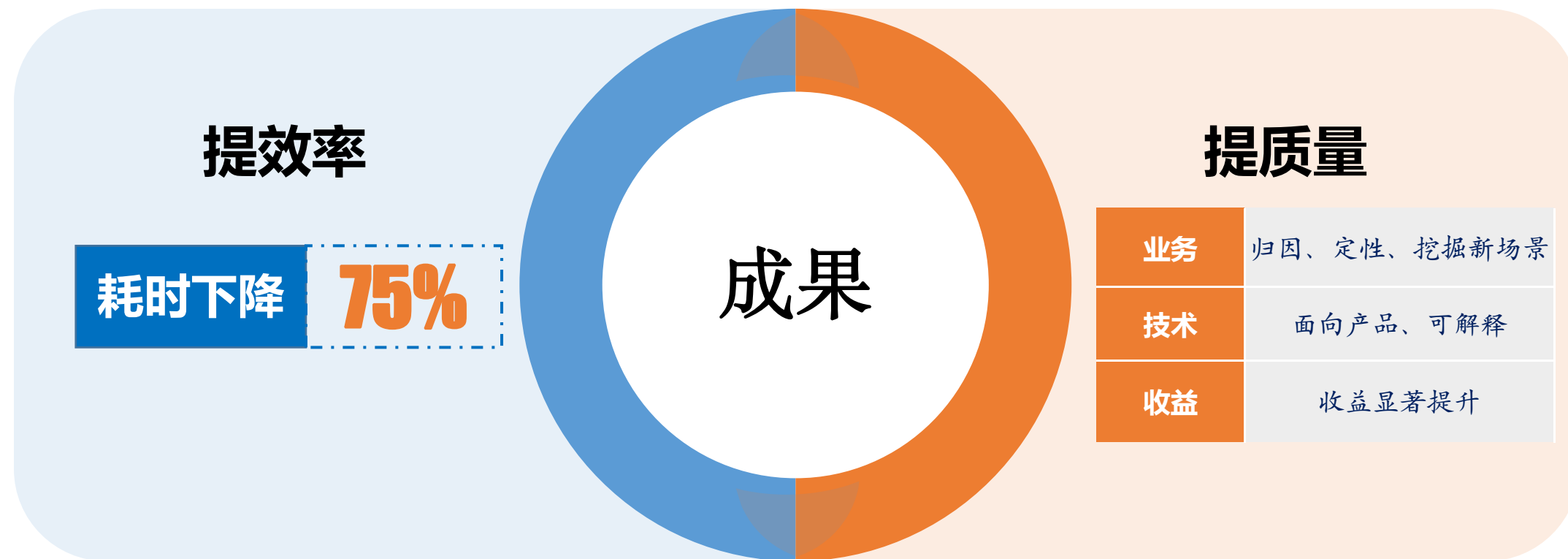
实例一存在问题



实例一-解决方法

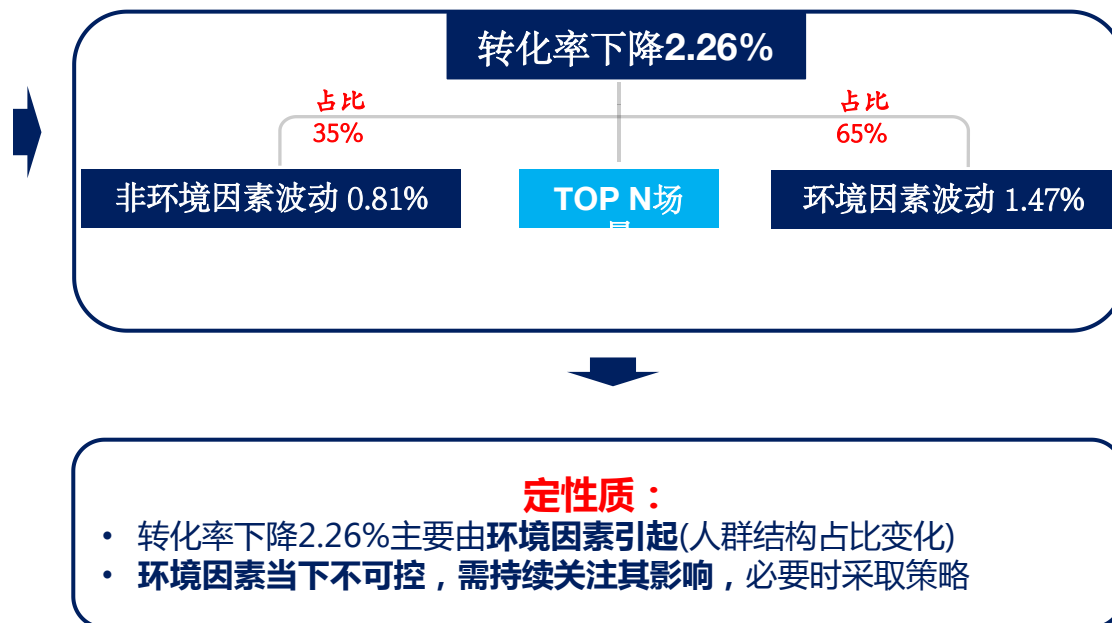
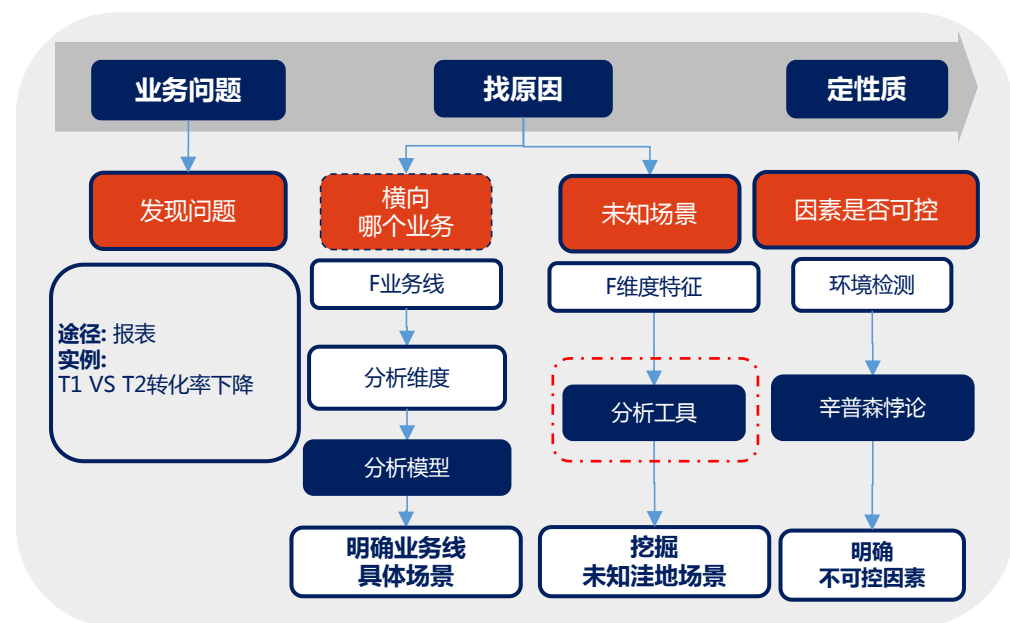


实例一-归因成果



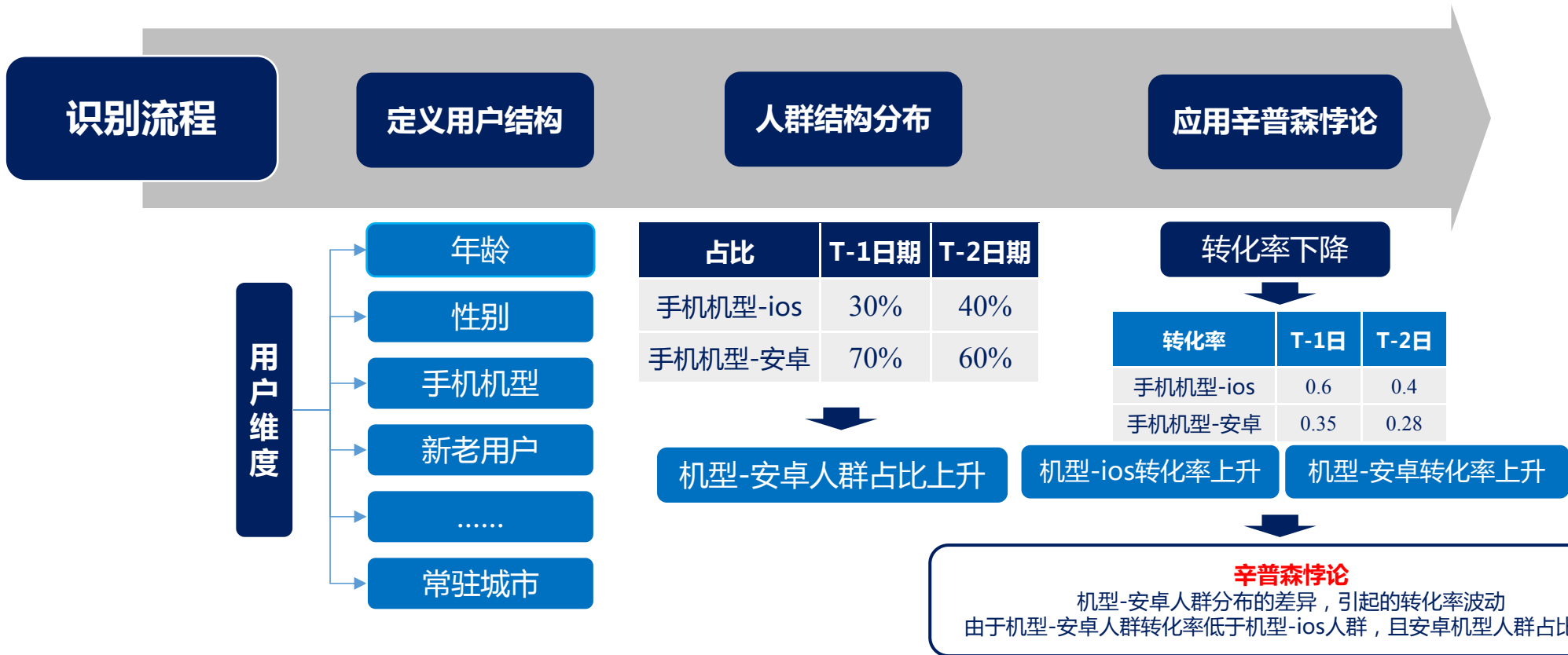
实例二-转化率波动归因

需求背景：通过归因分析，解释业务线的T-1日比T-2日转化率变化的原因



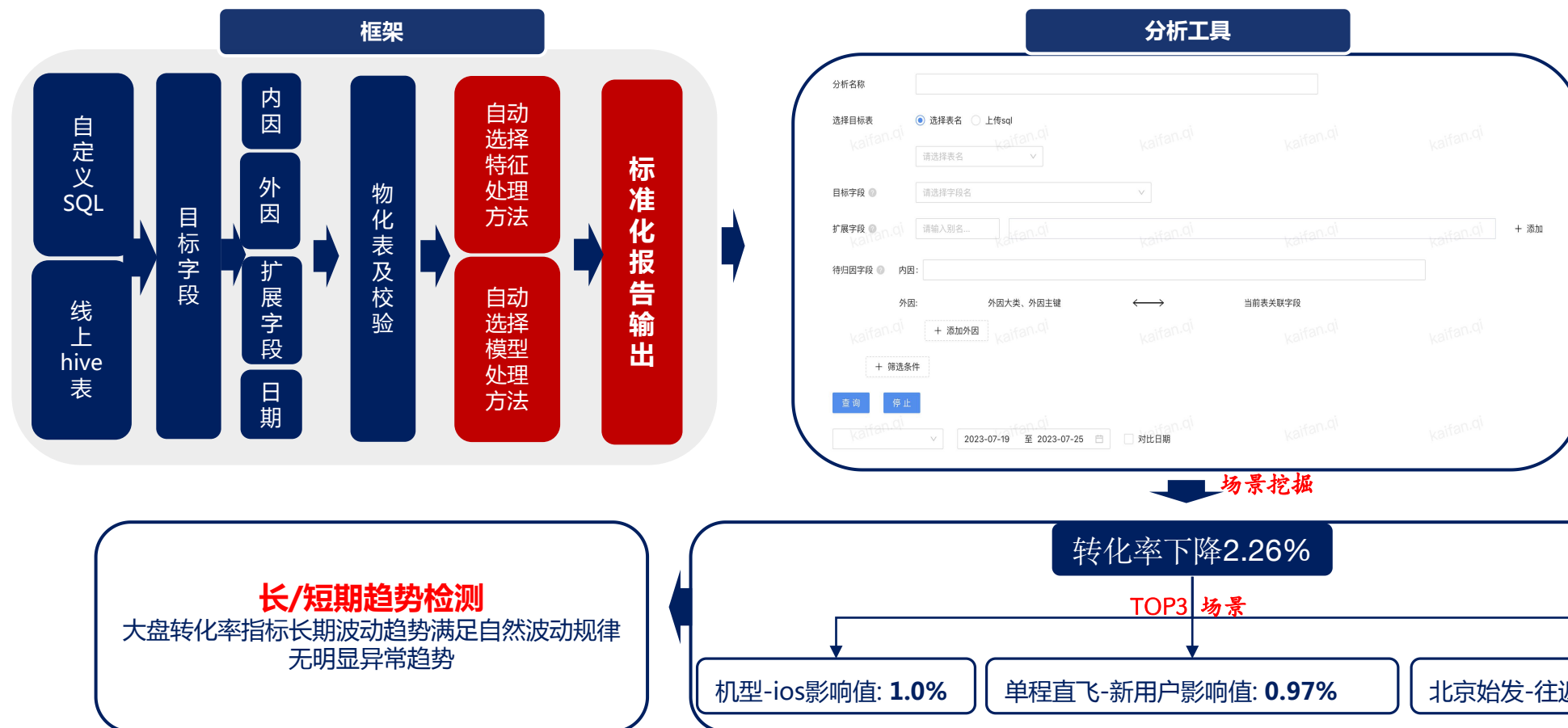
实例二-环境因素识别

环境因素定义：用户结构分布差异，所引起的转化率波动



去哪儿旅行
总 有 你 要 的 低 价

实例二-挖掘归因场景



总结与建议



基础建设

数据整合，基础方法的重要性不可忽视



提问者心态

对什么归因
怎么归因
有了归因结果能做什么



实践出真知

贴合业务实践的归因才有价值

未来规划



THANKS

TDDL

DistributedTable

DBproxy

HBase

PostgreSQL

SSD

MongoDB

GreatDB

Cassandra

Hyperbase

Hubble

DataCenter

VisualDataPlatform

Blockchain

ArgoDB

Distributed

DatabaseKernel

TemporalData

CloudnativeData

AIalgorithm