

# 大数据驱动业务增长：

## 百度外卖商业智能推荐系统实践

蒋凡

百度外卖技术委员会主席

北京

伦敦

纽约

旧金山

圣保罗

上海

东京

# QCon

## 全球软件开发大会

### [上海站]

主办方 **Geekbang** 极客邦科技 **InfoQ**

信息安全

机器学习

人工智能

黑产

互联网金融 (FinTech)

团队管理

云计算

基础设施

软件性能

硅谷

微服务

互联网架构

2017年10月17-19日  
上海·宝华万豪酒店

——> 扫描二维码  
开启软件开发新思路





Geekbang> | EGO EXTRA GEEKS' ORGANIZATION NETWORKS  
极客邦科技

# EGO会员招募季

EGO旨在组建全球最具影响力的技术领导者社交网络，联结杰出的技术领导者学习和成长。

2017年6月30-7月10

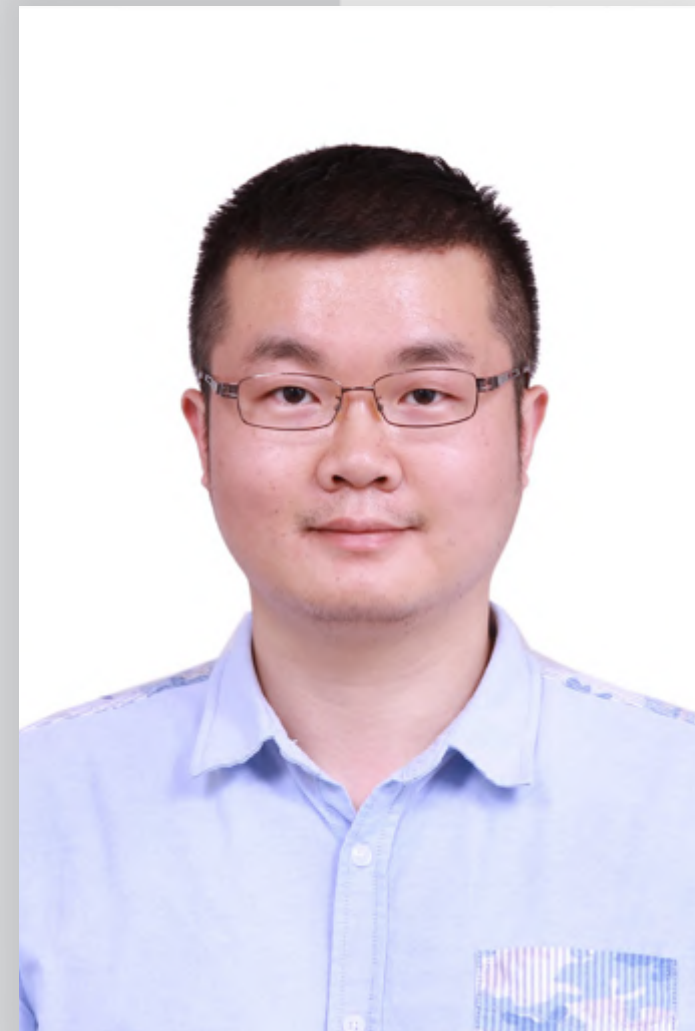


扫码报名

# SPEAKER INTRODUCE

## 蒋 凡 百度外卖技术委员会主席

- 2006年硕士毕业于中国科学技术大学，加入百度网页搜索团队，设计开发相关性核心算法和跨语言搜索基础架构。
- 2012年加入百度知心团队，设计开发新一代的知识图谱推荐引擎，改变百度搜索右侧的展现形态。
- 2014年加入百度外卖团队，负责物流调度、个性化推荐、智能营销和画像建模等多个技术方向。
- 译有《推荐系统》（人民邮电出版社）和《推荐系统：技术、评估及高效算法》（机械工业出版社）。



SPEAKER  
ArchSummit 2017 ShenZhen

# 为什么需要数据驱动增长

单量 × 利润

## 积累成长

- 搭建生态
- 教育用户
- 线下推广
- 丰富功能

## 野蛮生长

- 市场增容
- 团队管理
- 资金投放
- 快速迭代

## 精益增长

- 大数据标签
- 平台化能力
- 分场景建模
- 针对性打法

商业智能  
推荐系统

时间

TABLE OF  
**CONTENTS 大纲**

---

- **建模商业画像**
- 开放平台接口
- 区分业务场景
- 运用增长工具

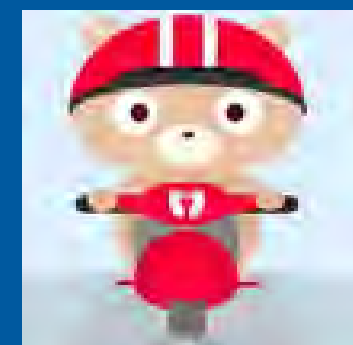
百度外卖

百度糯米

百度地图

百度大数据

第三方



用户画像

商户画像

骑士画像

1000+项 全准细



1

## 人口属性

- 1) 了解我们的用户是谁？
- 2) 性别、年龄、身份、职业、设备信息等

2

## 使用场景

- 1) 了解用户什么时候用，在哪里用？
- 2) 时间、地点、场景等

3

## 用户偏好

- 1) 了解用户喜欢吃什么、喜欢买什么、都去哪些店买？
- 2) 菜品、商品、口味、品牌、商家、品类等

4

## 消费行为

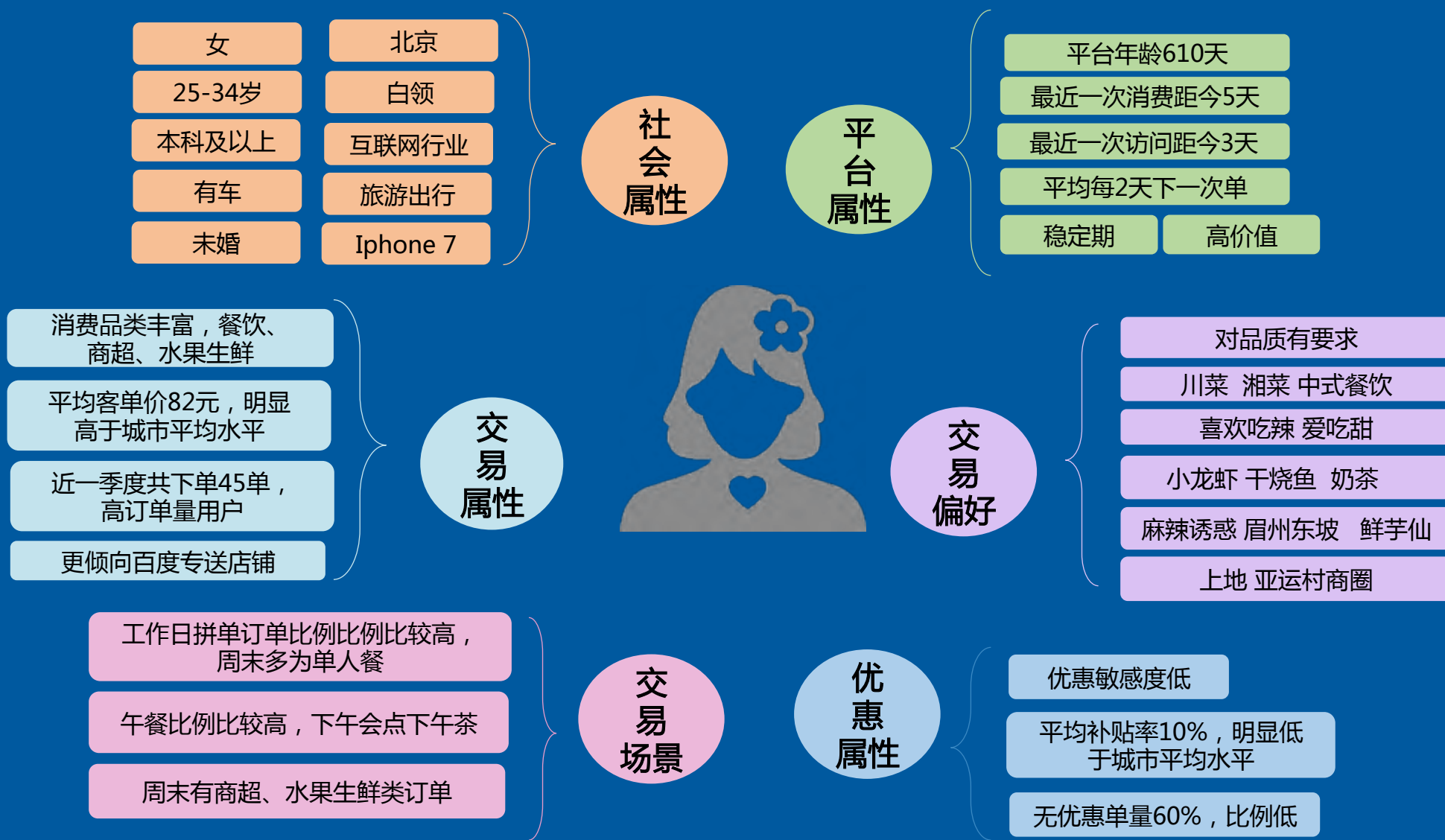
- 1) 了解用户消费习惯、每次花多少钱、多久消费一次？
- 2) 优惠敏感度、价格、频次、消费决策等

5

## 平台属性

- 1) 了解用户的平台情况是什么样子？
- 2) 用户生命周期、用户价值、用户分层等





北京白领女性，平台粘性高，工作日拼单比例高，周末点餐品丰富，不注重补贴，喜欢吃辣的川菜、湘菜同时也喜欢中式的餐饮，追求生活品质，对优惠不敏感。

# 用户价值分级

- 提供高价值的用户
- 提供中等价值的用户
- 提供低价值的用户
- 不能提供价值的用户



鲸鱼用户

海豚用户

小鱼用户

海藻用户

- 以维稳为主、根据用户喜欢，提供贴心的VIP服务

- 以转化为主，维护为辅。
- 结合用户画像挖掘用户潜在能力，不断提升这两个部分的用户潜在价值，额度利润最大化

- 重转化，对于多次转化失败的用户尽早舍弃

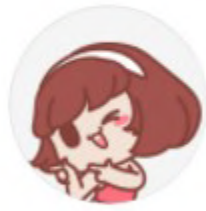
TABLE OF  
**CONTENTS 大纲**

---

- 建模商业画像
- **开放平台接口**
- 区分业务场景
- 运用增长工具

# 用户画像平台

## 基础属性



ID:1277953951

女

25-34岁

广州

本科及以上

无车

设备信息: OPPOR9tm

平台年龄: 590天

消费水平: 高

收餐地址:

餐饮消费水平: 高

兴趣爱好: 软件应用, 影视音乐

餐饮美食: 口味/甜; 品牌偏好/天美食

消费日期

收藏商户

夜宵8次(33.3%; 平均客单价: 44.5元)  
下午茶8次(33.3%; 平均客单价: 28.8元)  
午餐5次(20.8%; 平均客单价: 23.6元)  
早餐2次(8.3%; 平均客单价: 26.5元)  
晚餐1次(4.2%; 平均客单价: 25元)

支付方式

渠道来源

消费时间

消费日期

收藏商户

聚合收银台40次 | 外置独立收银台28次

iphone客户端68次

午餐32次(47.1%; 平均客单价: 83... 更多

工作日40次(58.8%; 平均客单价: 9... 更多

紫光园(红庙总店) 更多

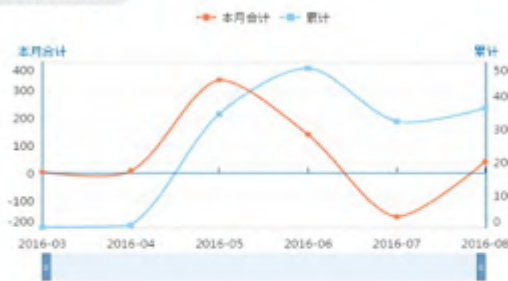
紫光园(红庙总店)

鼎泰丰(沈阳店)

鼎泰丰(沈阳店)

## 交易属性 (近一年数据)

### 平台总收益趋势



### 月单量趋势



### 客单价 (优惠前)



### 客单价 (优惠后)



## 交易偏好 (近90天的数据)

### 口味

麻辣烫

### 食材

小龙虾水饺猪肉卷  
龙虾 冰沙 豆腐 汤圆

商户集中度: 1.7

支付方式

渠道来源

消费时间

消费日期

微信支付29次 | 余额+钱包7次 | 余额3次 | 线下支付1次

iphone客户端40次

下午茶13次(32.5%) | 午餐13次(32.5%) | 晚餐6次(15%) | 早餐4次(10%) | 夜宵4次(10%)

工作日26次(65%) | 节假日14次(35%)

商家名称

17饭(西二旗店) 4次(10%) 更多

17饭(西二旗店) 4次(10%)

蓝花食尚4次(10%)

美湘厨(上地店)3次(7.5%)

东方饺子王(西二旗店)3次(7.5%)

讨小海鲜海鲜饭(西二旗店)2次(5%)

邻家(上地三街店)2次(5%)

坑大夜宵(西二旗店)2次(5%)

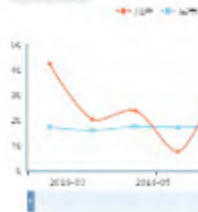
良记麻辣小龙虾(西二旗店)2次(5%)

麻辣烤鱼·麻辣小龙虾·小外食2次(5%)

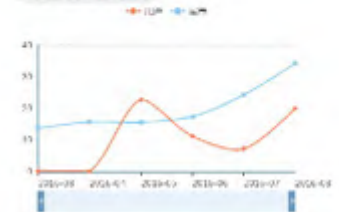
华源超市(上地店)2次(5%)

### 优惠属性

#### 补贴趋势



#### 无优惠订单比例趋势



# 商户画像平台



金三顺紫菜包饭(裕彤店)

## 优势指标 ?

用户粘性 **高** 用户投诉 **高**

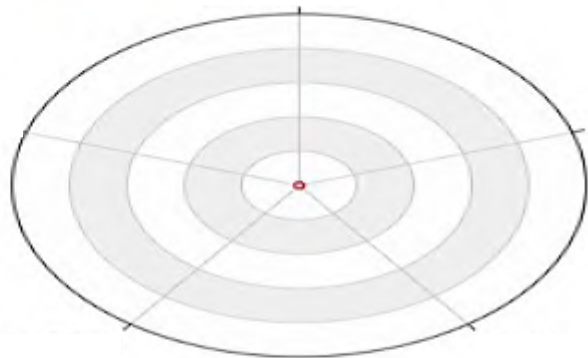
## 中等指标 ?

## 弱势指标 ?

商户价值	低	客单价(优惠前)	低	日均流水(优惠前)	低	订单转化率	低	周复购率	低
曝光人数	低	曝光量	低	访问人数	低	访问量	低	下单人数	低
下单量	低	完成单人数	低	完成单量	低	曝光-访问UV转化率	低	曝光-访问PV转化率	低
访问-下单UV转化率	低	访问-下单PV转化率	低	下单-完成UV转化率	低	下单-完成PV转化率	低	平台收益	低
客单价(优惠后)	低	用户评分	低						

## 商户价值

商户价值 **21.43分** 低价值商户



## 口味食材 ?

紫菜包饭 泡菜  
紫菜包饭 菜包  
补充蛋白质 肉松  
补充维生素

显示列/隐藏列

指标	原值/商圈中位数/城市中位数	商圈排名	城市排名
客单价(优惠前)	¥ 0.00/ ¥ 12.60/ ¥ 8.60	412/575 <span style="color: red;">——</span>	7,204/10,455 <span style="color: blue;">——</span>
日均流水(优惠前)	¥ 0.00/ ¥ 17.50/ ¥ 11.80	412/575 <span style="color: red;">——</span>	7,204/10,455 <span style="color: blue;">——</span>



# 群体筛选平台

**基础属性**

城市 已选1个 北京 ×

商圈 已选0个 请选择商圈, 支持搜索

商户名称 已选0个 请选择商户, 支持搜索

首单时间 早于 2016-08-01

最后一单时间 早于 2016-11-22

客单价(优惠后)

客单价(优惠前) 近1年 大于 70 元

**其他属性**

+添加指标

用户分布 / 用户特征分布

城市 全国 商圈 全部

预览结果

**16,908** 位用户符合以上筛选条件

**43,786,681** 位用户符合以上筛选条件

数据截止时间: 2016-11-28

**整体用户分层**

**43,822,141** 位用户符合以上筛选条件

数据截止时间: 2016-11-29

用户分层比例

- 活跃用户
- 次活跃用户
- 流失老用户
- 次新用户
- 流失新用户
- 沉默用户

分层	口径	比例	人数
活跃用户	近一年订单量 >=4 & 近6个月订单量 >=4	29.25%	12,818,944
次活跃用户	近一年订单量 >=4 & 近6个月1<=订单量<=3	5.27%	2,311,562
次新用户	近一年订单量 >=4 & 近6个月订单量=0	8.91%	3,905,940
流失老用户	1<=近一年订单量<=3 & 平台年龄<=42天	4.96%	2,173,074
沉默用户	1<=近一年订单量<=3 & 平台年龄>42天 & 近6个月订单量>0	29.77%	13,045,485
流失新用户	1<=近一年订单量<=3 & 平台年龄>42天 & 近6个月订单量=0	21.83%	9,567,136

ID	城市
867657951	北京
833499163	北京
1620116610	北京
1921041543	北京
184094462	北京

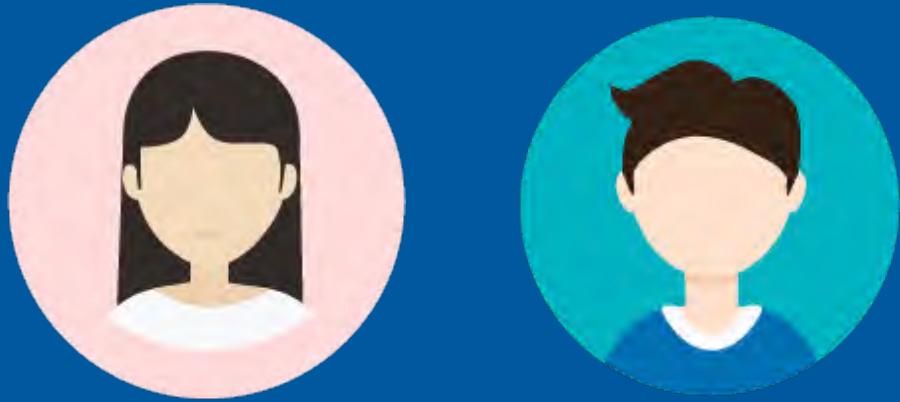
TABLE OF  
**CONTENTS 大纲**

---

- 建模商业画像
- 开放平台接口
- **区分业务场景**
- 运用增长工具



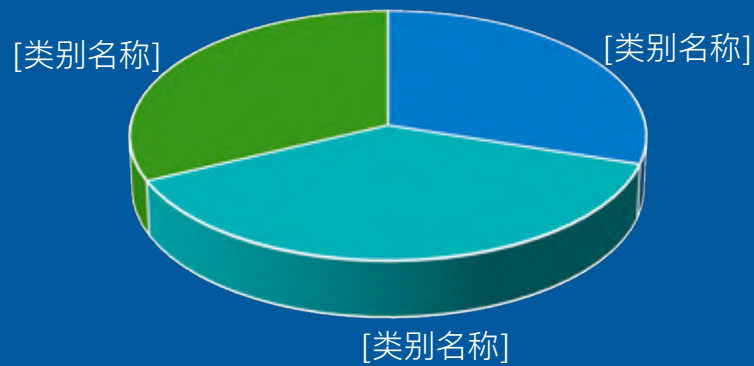
# 区分标准：差异化人群



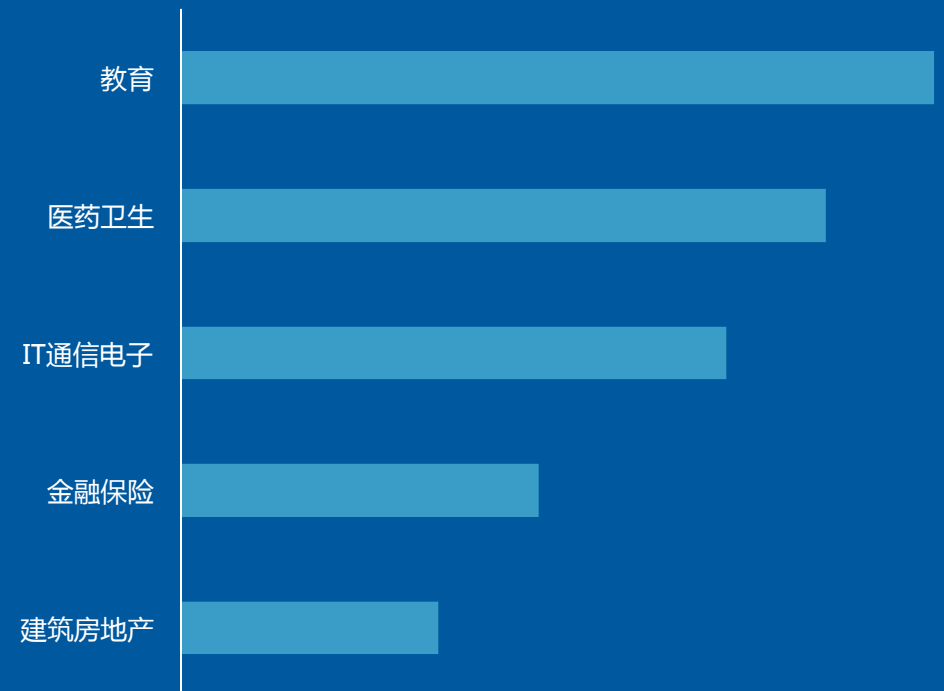
性别 年龄



学历 行业



■本科及以上 ■大专 ■高中及以下

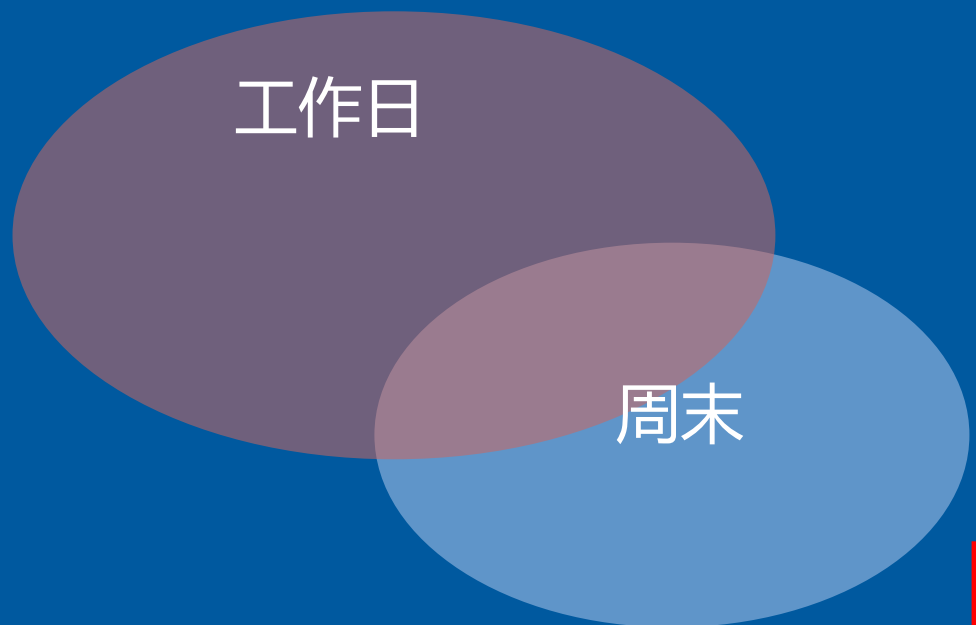


# 人群聚类

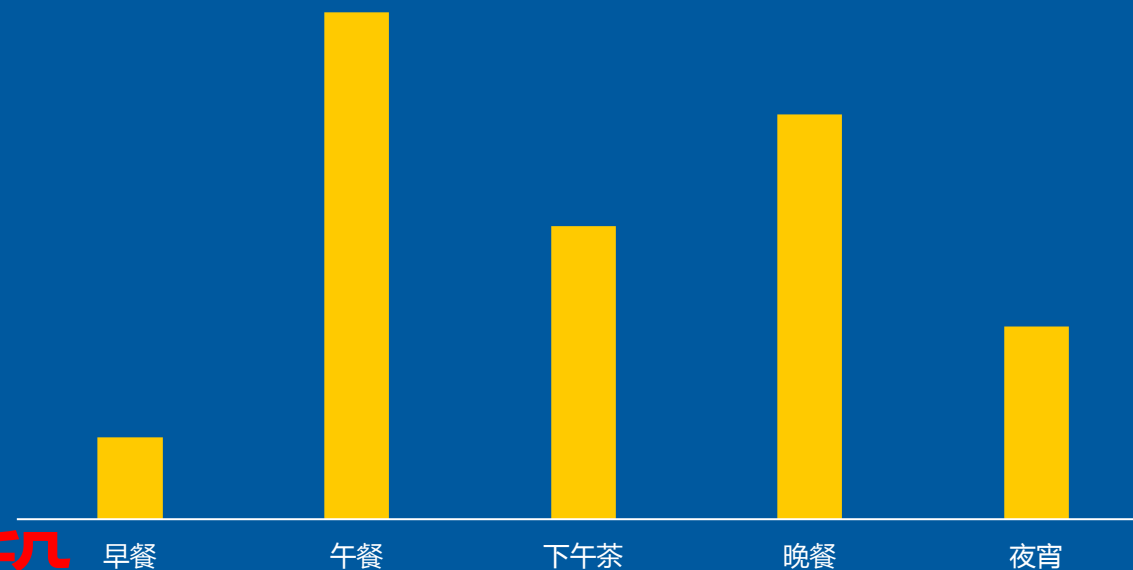
通过用户聚类算法，自下而上区分出彼此互斥的若干类用户集合



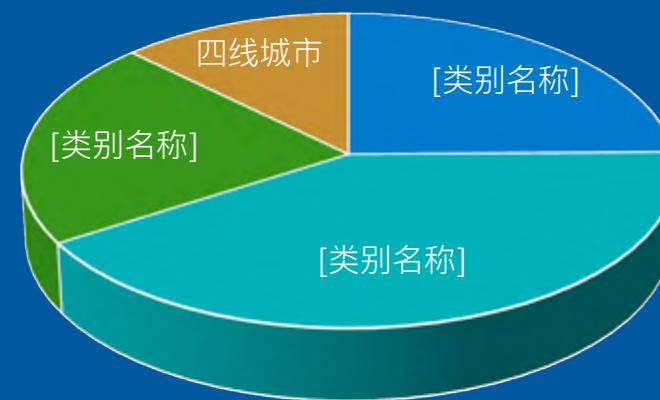
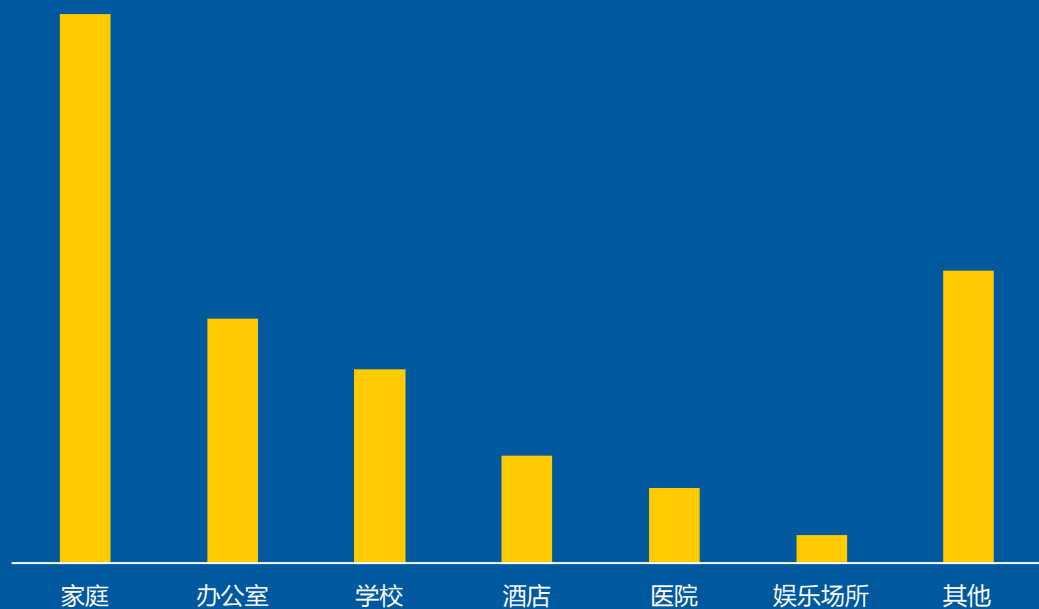
# 区分标准：时间+空间



日期 时段



地点 城市



□一线城市 □二线城市 □三线城市 □四线城市

# 特定场景

依照时间、地点和用户自上而下切分出特定场景的细分人群

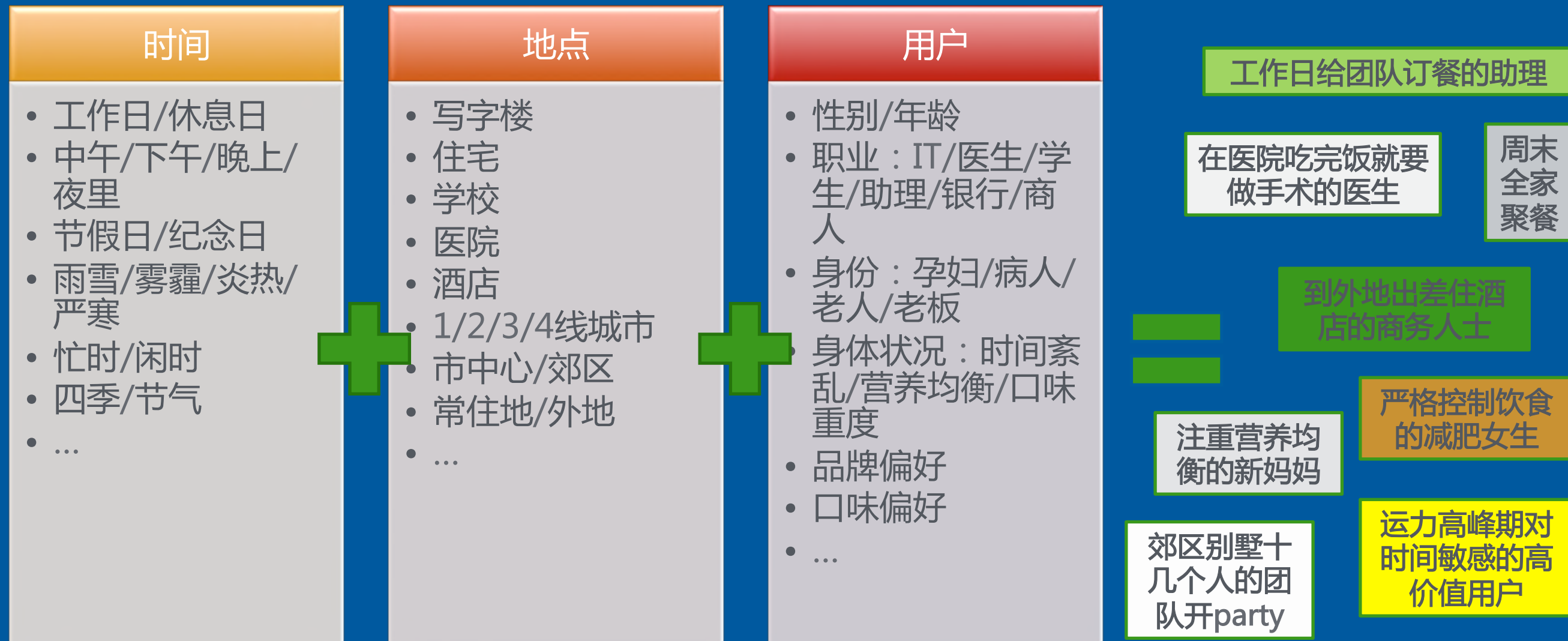


TABLE OF  
**CONTENTS 大纲**

---

- 建模商业画像
- 开发平台接口
- 区分业务场景
- **运用增长工具**

# 拉新

从0到1000，圈定种子用户  
从1万到1千万，构建增长机制

# 定价

服务区分  
用户区分  
时间区分



## 增长工具

# 留存

留存分析工具  
挽回流失用户

# 风控

加强数据校验  
发现数据异常  
制止作弊行为

# 拉新

## 圈定种子用户

从0到10000  
采用邀请机制  
引起社区关注

## 构建增长机制

从1万到1000万  
竞争群体增长  
延伸品类增长  
相近地址增长  
社交关系增长

## 评估与反馈

降低竞品用户的获取成本  
提高附近用户的激活比例  
提高延伸用户的活跃程度  
提高社交用户的影响能力



# 拉新

从1万到1000万，构建增长机制

增长领域	数据源	推荐物品	技术点
竞品拉新	竞品用户授权数据	潜在新用户	自然语言处理 模式识别
品类扩展拉新	用户消费行为	新品类用户	协同过滤
地址扩展拉新	POI语义标签	潜在新用户	语义相似度
社交关系增长	用户关联矩阵	潜在新用户	社交图谱挖掘

# 留存

## 固定用户群-时序演化

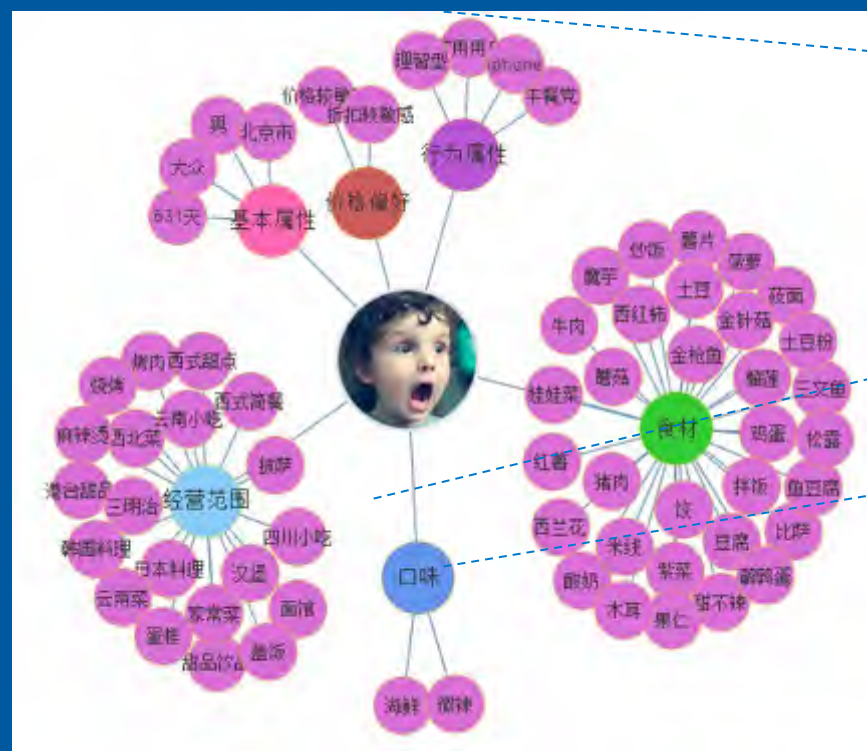
- 将固定用户群的画像属性放在动态的时间流变化过程中考虑
- 对比分析随**时间变化**导致用户行为变化而呈现出的**特征差异**
- 可视化呈现固定人群随时间演进在数量比例、属性指标上变化

## 目标用户群-定格测算

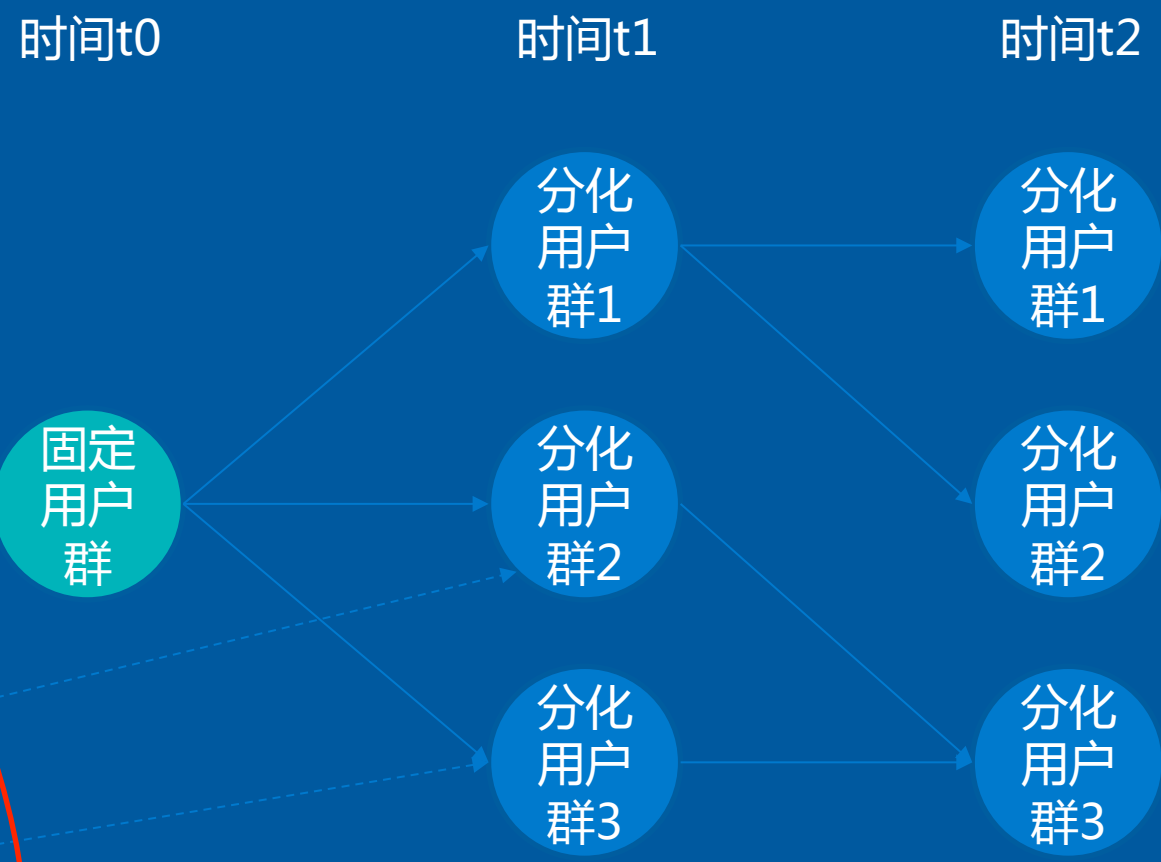
- 利用描述标签的倒排索引对目标用户群进行筛选分析
- 观测发现**特定类别人群**组成、转化趋势背后潜在**关键原因**
- 可视化呈现目标人群的分布统计信息、消费体验指标、变化趋势及提前预警

# 留存

## 固定用户群的时序演化

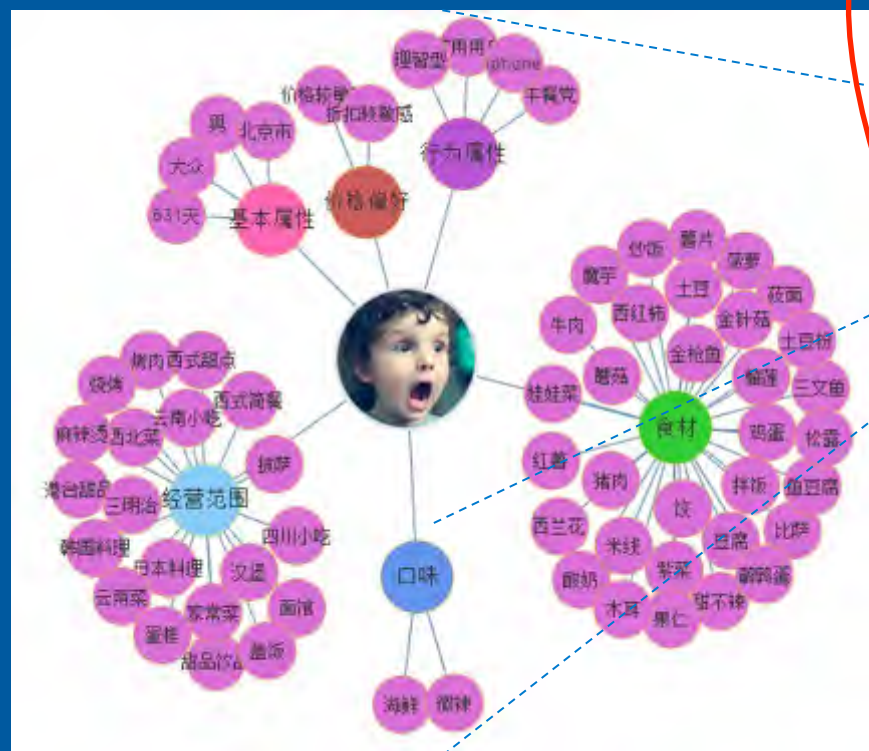


推荐相关因素

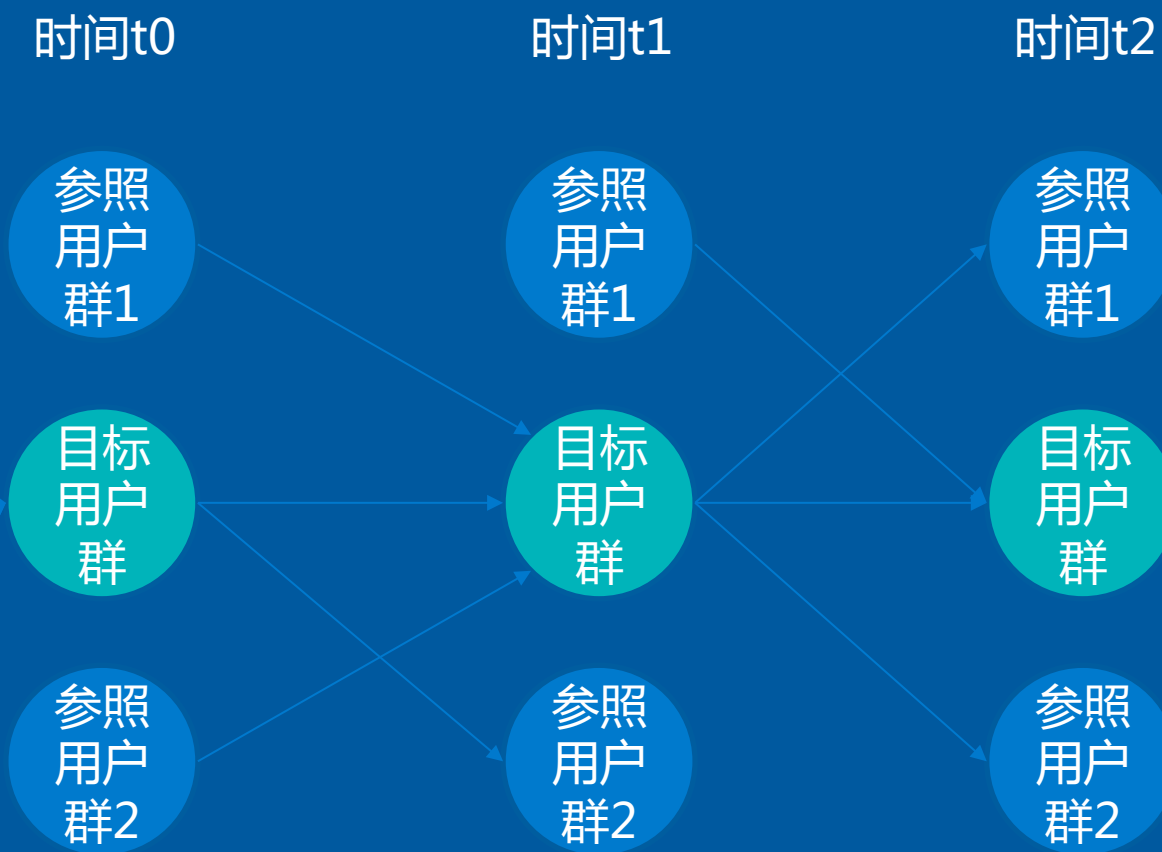


# 留存

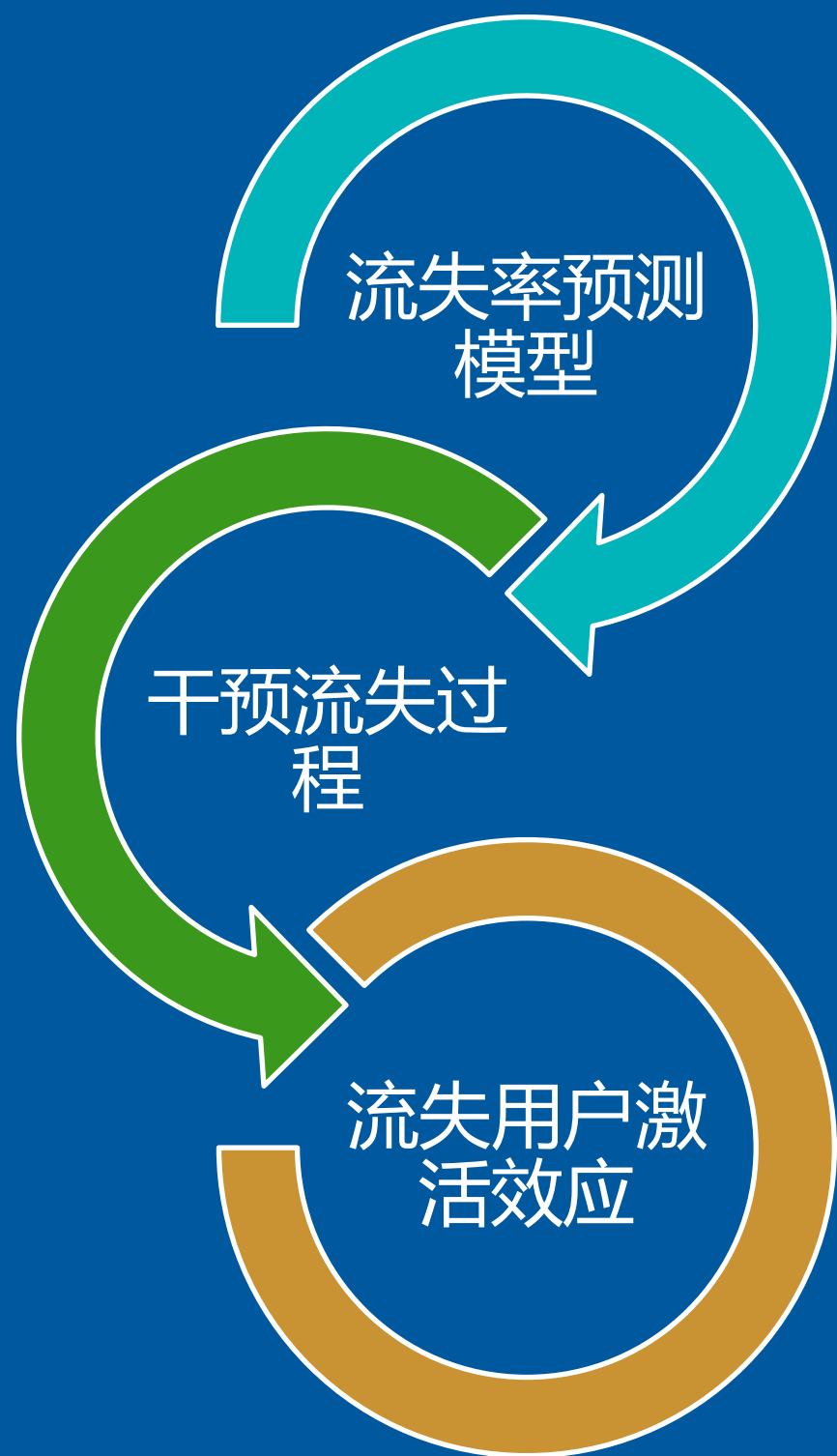
## 目标用户群的定格测算



推演相关因素



# 留存



## 挽回流失用户

机器学习模型预测每个用户的未来消费意愿

用红包、优惠券、返现等形式降价促销，影响消费者决策意愿

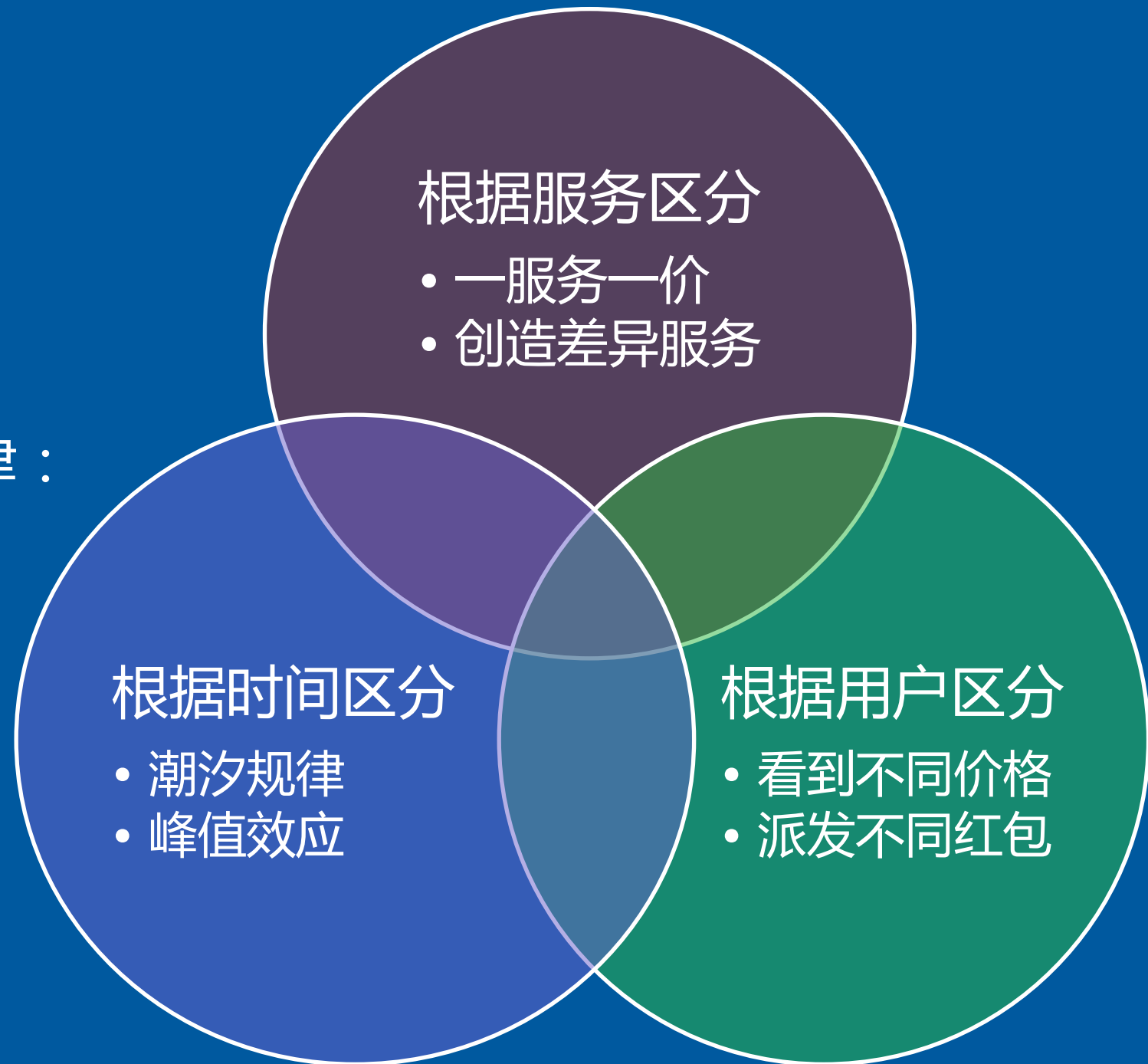
给定补贴金额总量情况下，整体情况下召回总概率最大

# 定价

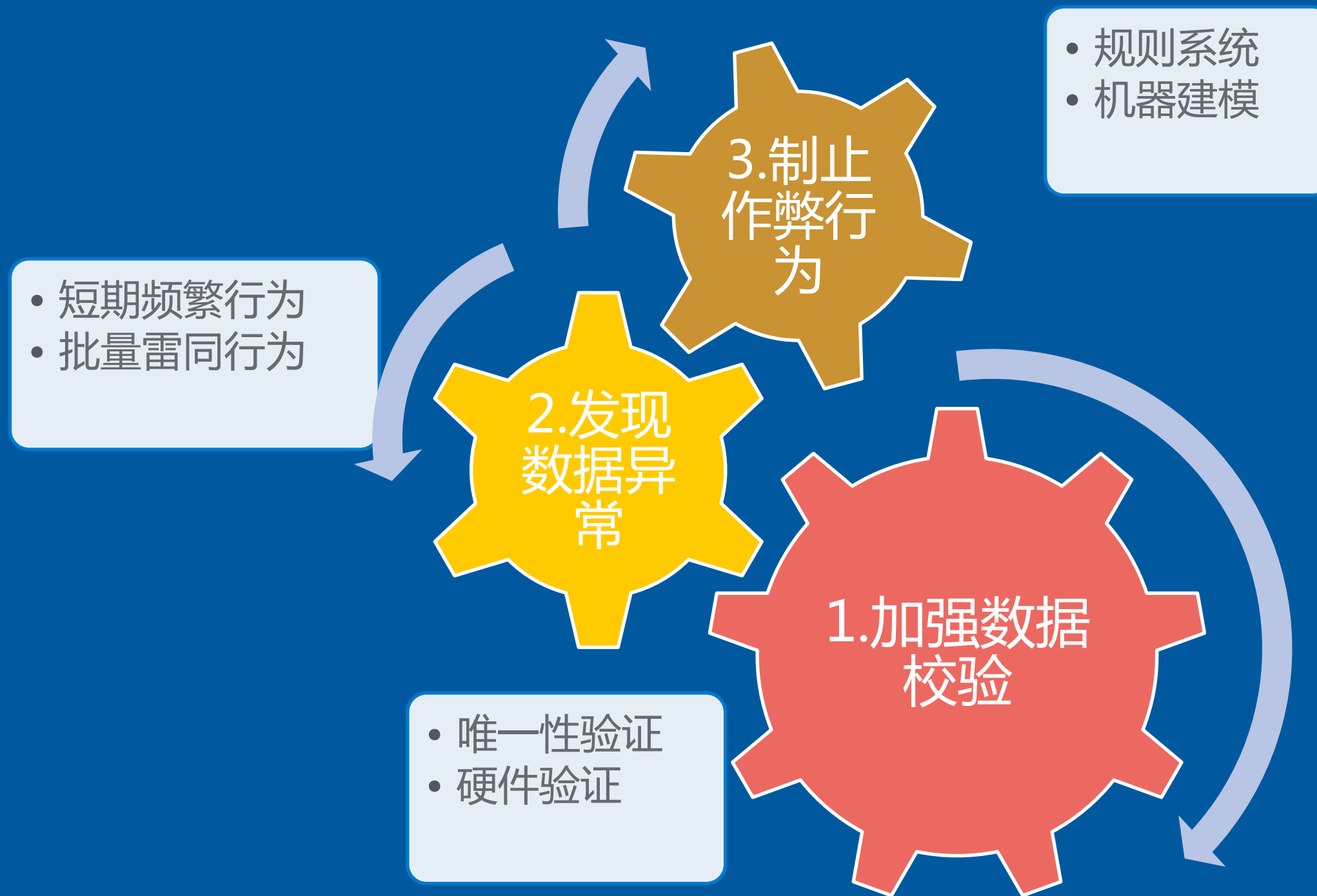
机器学习模型预测：  
用户的价格敏感度/补贴敏感度

用户接受补贴方案的行为变化规律：  
购买登录比/投资收益比

注入资金改变订单的时间分布，  
起到削峰填谷的作用



# 风控

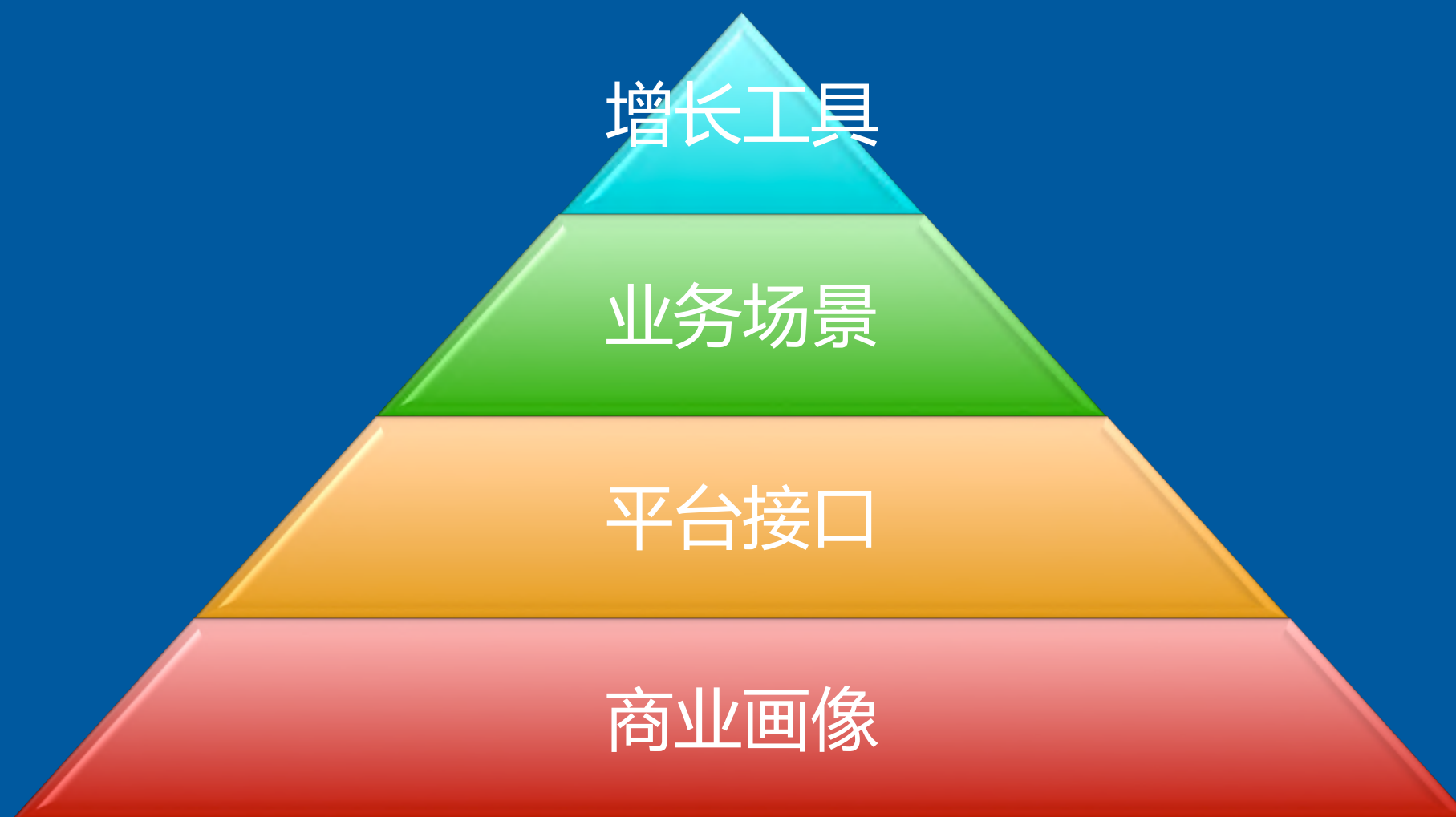




# 总结

经典推荐系统：根据历史数据提升现在用户的体验

商业智能推荐系统：利用大数据预测未来，改变用户行为



# 总结

在不同业务场景将大数据转化为增长动力

单量 × 利润



# THANKS!

让创新技术推动社会进步

HELP TO BUILD A BETTER SOCIETY WITH  
INNOVATIVE TECHNOLOGIES

# Geekbang >

## 极客邦科技

**InfoQ**  
ueue

专注中高端技术人员的技术媒体



**EGO** EXTRA GEEKS' ORGANIZATION  
NETWORKS

高端技术人员学习型社交平台



**StuQ**  
ueue

斯达克学院

实践驱动的 IT 教育平台

