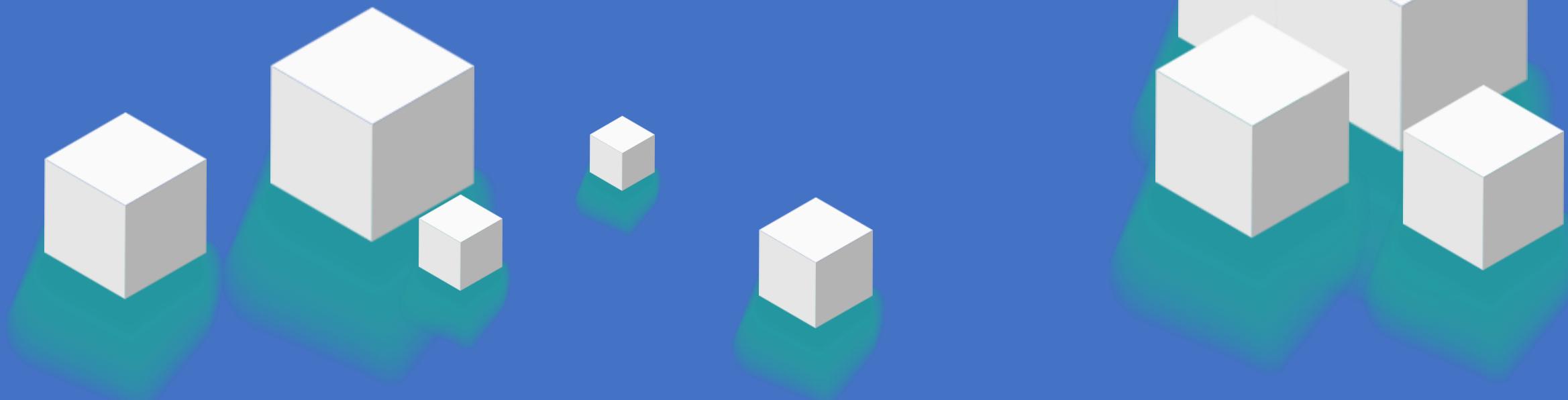


# 数据分析思维



# 学习方法

---

- Thinking: behind the theory, original from the real problem
- Action: solve problems by tools, present the results

# Self-introduction

---

## 和计算机、算法相关

(10岁，清华计算机博士，NOI，ACM比赛，ACM，IEEE，中国人工智能协会，阿里云MVP，CCF专委)

## 和培训、企业服务相关

(专栏付费订阅人数超过17万，企业客户包括：蚂蚁金服，美的，汽车之家，上汽大众，航天信息，中国银行，中国银联，花旗银行，杭州银行，泰隆银行，中原银行，长沙银行，渤海银行，平安银行，平安保险，阳光保险，兴业数金，上海银科，尚诚金融，中泰证券，中国外汇交易中心，联通软研院，香港马会，融创中国，大恒集团，麦当劳，西门子等)



# >> 今天的学习目标

---

## 数据驱动业务增长

- 分析式AI与生成式AI
- 老奶奶 vs Google的思维方式
- 用数据定义市场空间
- 用数据定义运营成本
- 用数据发现商业机会
- 用数据发现羊毛党
- 业务增长的四象限

## 数据思维分析框架

分类思维：BCG矩阵

分类思维的目的，哪些场景用到分类思维

分层思维：RFM顾客分层

分层思维的目的，哪些场景用到分层思维

# 分析式AI与生成式AI

Thinking: 分析式AI与生成式AI的区别是什么？

	分析式AI (Analytical AI)	生成式AI (Generative AI)
目标	分析和解释数据，提供洞察、预测或决策支持。	生成新的内容，这些内容在形式和风格上与训练数据相似。
用途	数据分析、预测建模、风险评估、分类任务等。	文本生成、图像创作、音频生成、内容创作等。
输出	报告、图表、预测结果、决策建议等。	新的文本、图像、音频、视频等内容。
技术原理	基于统计分析、机器学习算法（如回归分析、决策树、神经网络等）。	基于生成式大模型，如DeepSeek, Qwen, ChatGPT (Transformer架构)。
数据需求	需要高质量、标注清晰的数据，以学习数据中的规律。	需要大量未标注数据，以学习数据分布。
应用场景	金融风险评估、医疗诊断、市场预测、客户行为分析等。	广告设计、内容创作、游戏开发、艺术创作等。
风险挑战	模型可能过度拟合训练数据，导致在新数据上表现不佳；数据偏差可能导致错误决策。	生成内容可能包含错误、误导性信息或不符合伦理的内容需防止内容被滥用。

数据驱动业务增长

# 预测是一种能力

《远见》（2019年11月出版）介绍过一个案例：

我的爷爷参加了解放战争和抗美援朝战争，为了新中国光荣负伤，于是国家给了我奶奶特别的待遇，让她可以在医生、教师、裁缝中选择任意的职业，于是我奶奶选择了裁缝。

在我爸爸和叔叔找工作的时候，因为根正苗红，爸爸被保送军官，叔叔被保送警察，但是一家之主的奶奶死活不同意，于是他们两个都双双进了工厂。

我父亲下岗后决定卖房买车，然后在货车和出租车中间选择了货车，接着是大家都知道的房价和出租车牌照价格蹭蹭往上涨

我叔叔把发家致富的眼光投在了彩票上，天天研究走势，我堂妹天天在家砸核桃，给她爸补脑.....

期数	开奖号码	开奖号码																		指标区	0-9分布																								
		百位						十位						个位																															
		0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9														
2011220	999	17	7	21	15	1	10	2	9	8	9	7	3	16	12	2	5	4	9	1	9	10	6	8	21	14	1	28	5	2	9	27													
2011221	069	0	4	22	18	2	11	3	10	9	1	8	4	17	13	3	6	8	10	2	1	11	7	9	22	15	2	29	6	3	9	15													
2011222	713	1	9	23	17	3	12	1	7	10	2	9	1	18	14	4	7	1	11	3	2	12	8	10	3	16	3	30	7	4	1	11													
2011223	345	2	10	24	3	13	5	1	11	3	10	1	19	15	4	8	2	12	4	3	13	9	11	1	17	5	31	8	5	2	12	4	2	1	4	2									
2011224	610	3	11	25	1	5	14	6	2	12	4	11	1	20	16	1	9	3	13	5	4	0	12	2	18	1	32	9	6	3	7	1	2	5	3										
2011225	863	4	12	26	2	8	15	1	3	8	5	12	1	21	17	2	10	6	14	5	1	11	13	3	19	2	33	10	7	4	17	1	2	1	13	3	2	6	3	8	4				
2011226	242	5	13	2	3	7	16	2	4	1	6	13	2	22	18	4	11	1	15	7	5	2	12	2	1	20	3	34	11	8	5	8	0	3	2	2	1	4	3	1	4	1	5		
2011227	084	0	14	1	4	8	17	3	5	2	7	14	3	23	19	1	12	2	10	8	7	3	13	1	2	4	4	35	12	9	6	12	0	3	0	3	1	2	4	4	2	5	8	6	
2011228	820	1	15	2	5	9	18	4	6	8	8	15	4	2	0	12	2	13	3	17	1	8	0	2	3	1	5	36	13	10	7	10	10	0	3	0	4	2	3	1	5	3	6	8	7
2011229	176	2	1	4	5	10	19	5	7	1	9	16	5	1	21	3	14	3	7	2	9	1	15	3	4	2	6	6	14	11	8	14	2	1	1	1	4	2	6	8	7	1	8		
2011230	936	3	1	4	7	11	20	6	8	2	9	17	5	2	3	1	15	5	1	3	10	2	18	4	5	3	7	6	15	12	9	18	2	1	2	1	2	3	3	7	6	1	2	9	
2011231	827	4	2	5	8	12	21	7	9	8	1	18	7	2	1	5	16	6	2	4	11	3	17	5	6	4	8	1	7	13	10	17	1	2	3	2	1	4	8	1	7	8	1		
2011232	452	5	3	6	9	14	22	6	10	1	2	19	8	1	2	1	2	5	7	3	5	12	4	18	2	1	5	9	2	1	14	11	11	1	2	4	3	2	2	4	5	2	1	1	2
2011233	002	0	4	7	10	1	23	9	11	2	3	0	3	2	3	7	折线	8	13	5	19	2	8	6	10	3	2	15	12	2	2	0	3	0	4	2	3	1	3	2	3	2	3		
2011234	211	1	5	2	11	2	24	10	12	3	4	1	1	3	4	8	2	8	6	7	14	6	1	4	9	7	11	4	3	16	13	4	2	1	1	1	2	4	2	2	4	3	3		
2011235	056	0	5	1	12	3	25	11	13	4	5	2	1	4	5	9	5	10	6	8	15	7	1	2	10	8	12	6	4	17	14	11	1	2	0	1	1	5	3	5	6	4	4	5	
2011236	634	1	7	2	13	4	29	6	14	5	6	3	2	5	3	0	1	11	7	9	16	8	2	3	11	4	13	1	5	18	15	13	1	2	1	2	2	3	4	1	6	5	5	6	
2011237	662	2	8	3	14	5	27	6	15	6	7	4	3	6	1	11	9	8	10	17	9	3	2	1	14	2	6	19	16	14	0	3	2	3	1	2	6	6	7	1	2				
2011238	301	3	9	4	3	28	1	16	7	8	0	7	2	12	3	1	9	11	18	10	1	1	13	2	15	3	7	20	17	4	2	1	0	1	1	3	2	3	1	7	8				
2011239	787	4	10	5	1	7	29	2	7	8	9	1	5	8	3	13	4	2	16	8	19	11	2	14	3	16	4	7	21	18	22	2	1	1	2	1	3	4	2	7	8	9			
2011240	975	5	11	6	2	8	30	3	1	9	9	2	6	9	4	14	5	3	7	1	20	12	2	3	15	4	5	1	22	19	21	3	0	2	2	3	4	5	3	7	1	9			
2011241	868	6	12	7	3	9	31	4	2	8	1	3	7	10	5	15	6	6	2	21	13	3	4	16	5	1	6	2	8	20	22	0	3	3	4	3	5	1	6	1	8	1			
2011242	697	7	13	8	4	10	32	6	3	1	2	4	8	11	6	16	7	1	2	3	9	14	4	5	17	6	2	7	7	1	21	22	2	1	4	5	4	6	2	6	7	1	9		
2011243	866	8	14	9	5	11	33	1	4	8	3	5	9	12	7	17	8	6	3	4	1	15	5	6	18	7	3	6	1	2	20	20	0	3	5	5	6	5	7	3	6	1	8	1	
2011244	716	9	15	10	6	12	34	2	7	1	4	8	1	3	18	9	1	4	5	2	18	6	7	19	8	4	6	2	3	23	14	2	2	1	6	1	7	5	8	4	6	7	1	2	
2011245	080	0	10	11	7	13	35	3	1	2	5	7	1	14	9	19	10	2	5	8	3	0	1	8	20	9	5	1	3	4	24	8	0	3	0	1	8	7	9	5	1	1	8	3	
2011246	650	1	17	12	8	14	36	6	2	3	6	8	2	15	10	20	5	3	6	1	4	0	8	9	21	10	6	2	4	5	25	11	1	2	0	2	9	8	10	5	6	2	1	4	
2011247	684	2	18	13	9	15	37	6	3	4	7	9	3	16	11	21	1	4	7	8	5	1	9	10	22	4	7	3	5	6	26	18	0	3	1	3	10	9	4	1	6	3	8	5	
2011248	370	3	19	14	3	16	38	1	4	5	8	10	4	17	12	22	2	5	7	1	6	0	11	23	1	8	4	6	7	27	10	2	1	0	4	11	3	1	2	1	7	1	6		
2011249	358	4	20	15	3	17	39	2	5	6	9	11	5	18	13	23	5	6	1	2	7	1	11	12	24	2	9	5	1	8	28	16	2	1	1	5	12	3	2	5	2	1	8		

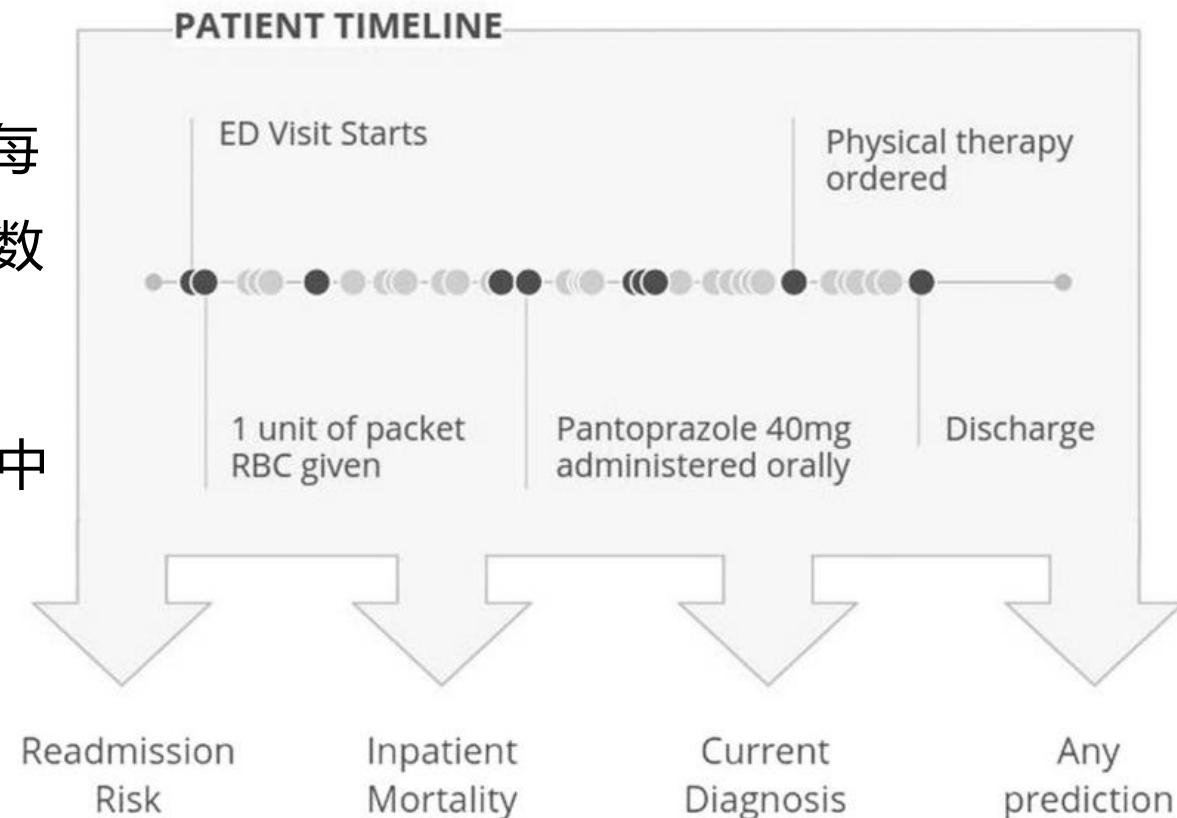
# 预测是一种能力

Google的AI预测医院病人的死亡时间，准确率达95%

数据来自两个医学中心 21 万 6 千名患者，这些患者每位都在医院停留了至少 24 小时，共收集了175639个数据点

所有电子病历中的事件全都被Embedding到低维空间中

《Nature》论文：Scalable and accurate deep learning with electronic health records, 2018



善于利用数据，和拍脑门做决策  
是两种不同的能力

# 用数据定义市场空间

---

数据是一种思维方式，可以让复杂问题简单化

Thinking: 北京特斯拉汽车的数量？

采用逻辑树分析法 => 将问题拆成子问题

- 1) 北京汽车数量是多少？
- 2) 特斯拉占比？

除了拆成 北京汽车数量 \* 特斯拉占比，还可以怎样拆解？



# 用数据定义市场空间

不同的思考角度，得出不同的结果

Thinking: 公司主营数据分析培训，年营收为1000万，  
5年的发展规划是怎样？

- 1) 基于往年业务发展的推演
- 2) 基于市场空间的战略定位



我们做数据分析，需要先思考，再动手 => 类似DeepSeek-R1

DeepSeek-R1对我们的数据思维是辅助加强，而不是替代决策。

如果你没有经历过完整的数据分析思维训练，在项目中就很难驾驭AI

# 用数据定义运营成本

Thinking: 很多企业都会用数据定义运营成本，数量维度很多，该看哪个？

渠道、成本、收益（不等于收入，用户的行为是立体的，不仅是付费，还可以是注册数，口碑数，主动分享数等）

## 内容运营

内容展示: PV, UV, 停留时长

内容粘性: 单位用户阅读数, 用户重复活跃次数

内容扩散: 分享次数, 回流率

内容转化: 付费人数, 付费金额

## 活动运营

宣传效果: 分渠道投放成本, 分渠道引流数

活动效果: 页面展示量, 参与人数, 满足活动目的的参与人数

成本控制: 预估成本, 实际成本, 单位用户成本

## 用户运营

用户规模: 新用户数, 老用户数, 流失用户数

用户粘性: 活跃登录用户, 复购用户, 点赞用户数, 收藏用户数

用户价值: 单位用户贡献, 分层用户贡献

# 用数据发现商业机会（经典数据分析案例）

---

美国明尼苏达州一家Target被客户投诉，一位中年男子指控Target将婴儿产品优惠券寄给他的女儿（高中生）。但没多久他却来电道歉，因为女儿经他逼问后坦承自己真的怀孕了

# 用数据发现商业机会（经典数据分析案例）

---

啤酒和尿布：

沃尔玛在分析销售记录时，发现啤酒和尿布经常一起被购买，于是他们调整了货架，把两者放在一起，结果真的提升了啤酒的销量。

原因解释：爸爸在给宝宝买尿布的时候，会顺便给自己买点啤酒？

沃尔玛是最早通过大数据分析而受益的传统零售企业，对消费者购物行为进行跟踪和分析

# 用数据发现商业机会

- 数据有助于消费者行为研究，消费者购买的时间、年龄、性别、购买保险的体验等数据，对于保险公司很有价值

Thinking：健康险，一般会在哪个时段购买的人多？

蚂蚁金服保险团队通过数据分析发现，很多消费者是在半夜12点 – 凌晨2点之间购买健康险（夜晚更容易引发健康问题的思考）

Thinking：车险，在哪个时间段购买的人多？

车险大部分是在上班时间购买（购买车险花费时间相对较大，很多人不愿意占用自己的休息时间）

=> 在不同时间段，向客户推送不同的产品

蚂蚁金服：从帮保险公司卖保险，到研究保险买卖，做好技术

通过技术与保险公司互为生态



# 用数据发现羊毛党

2015-2019年，空姐李某购买航班延误险 => 获利300余万元

这期间延误航班近900次，使用不同护照号身份证号，在各大保险公司频繁申请航空延误险，多次进行理赔

理赔金赔付款到具体保险人后，都会将理赔金转到李某手里

Thinking：为什么能获利成功？

在购买航班之前，李某会对航班以及当地天气进行分析。综合历史数据找到延误率非常高的航班，起飞的时候再去看它的天气

Thinking：是否构成保险欺诈？

=> 不论是否构成欺诈，这种行为不应该被倡导

=> 可以通过数据分析，识别“羊毛党”（保险购买异常，并获利）



不同身份投保“延误险”（保险公司发现）

# 用数据发现羊毛党

---

天猫复购用户预测比赛：

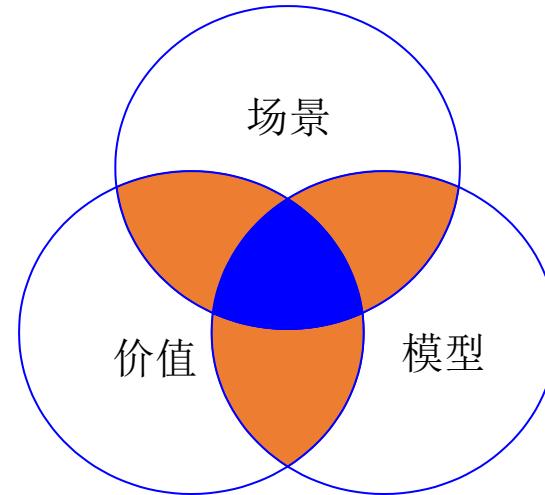
- <https://tianchi.aliyun.com/competition/entrance/231576/information>
- 电商经常会做大促，尤其是在特殊的节日，比如黑色星期五，双十一等，目的是为了吸引大量的新用户。然而，许多吸引过来的购买者只是一次性交易，这些促销对于转换为长期顾客来说收效甚微
- 为了解决这个问题，对于复购用户的预测就尤为重要，因为关注潜在的忠诚消费者，对于商家来说可以降低促销成本，增加投资回报率ROI
- 在线广告领域，发现有价值的客户非常有挑战，尤其是对于新购买者来说。但对于天猫来说，已经积累了大量用户行为日志，可以用于解决客户价值定位的问题



# Summary

---

- 预测是一种能力（老奶奶 VS Google）
- 用数据定义市场空间
- 用数据定义运营成本
- 用数据发现商业机会
- 用数据发现羊毛党



数据驱动业务增长

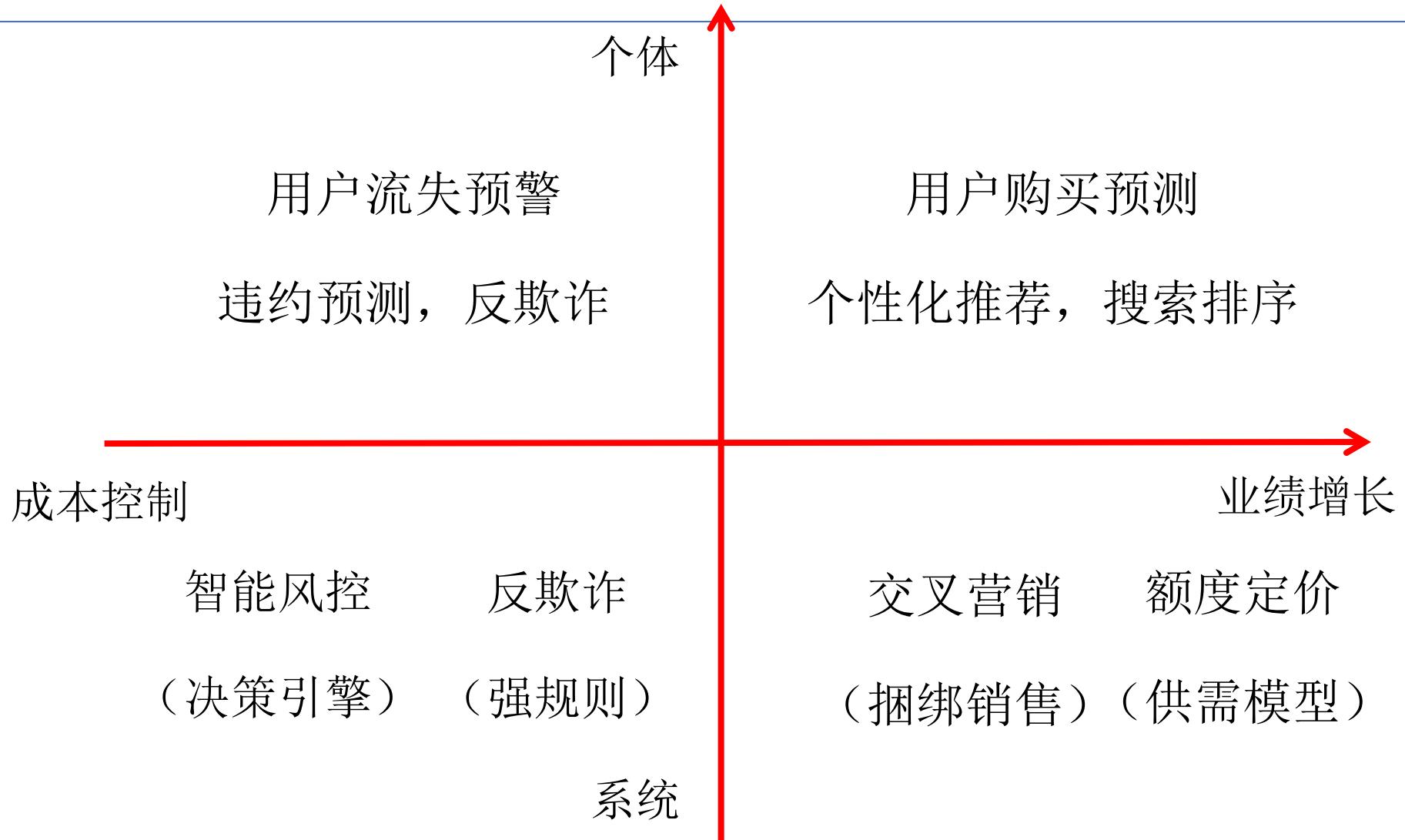
# 数据驱动业务增长

---

- 数据分析不仅是技术工具，更应该是全员必备的思维方式
- 面向业务流，采用精细化运营的方式给企业带来利润来源
- 组织内部，使用数据解决问题的行动指南

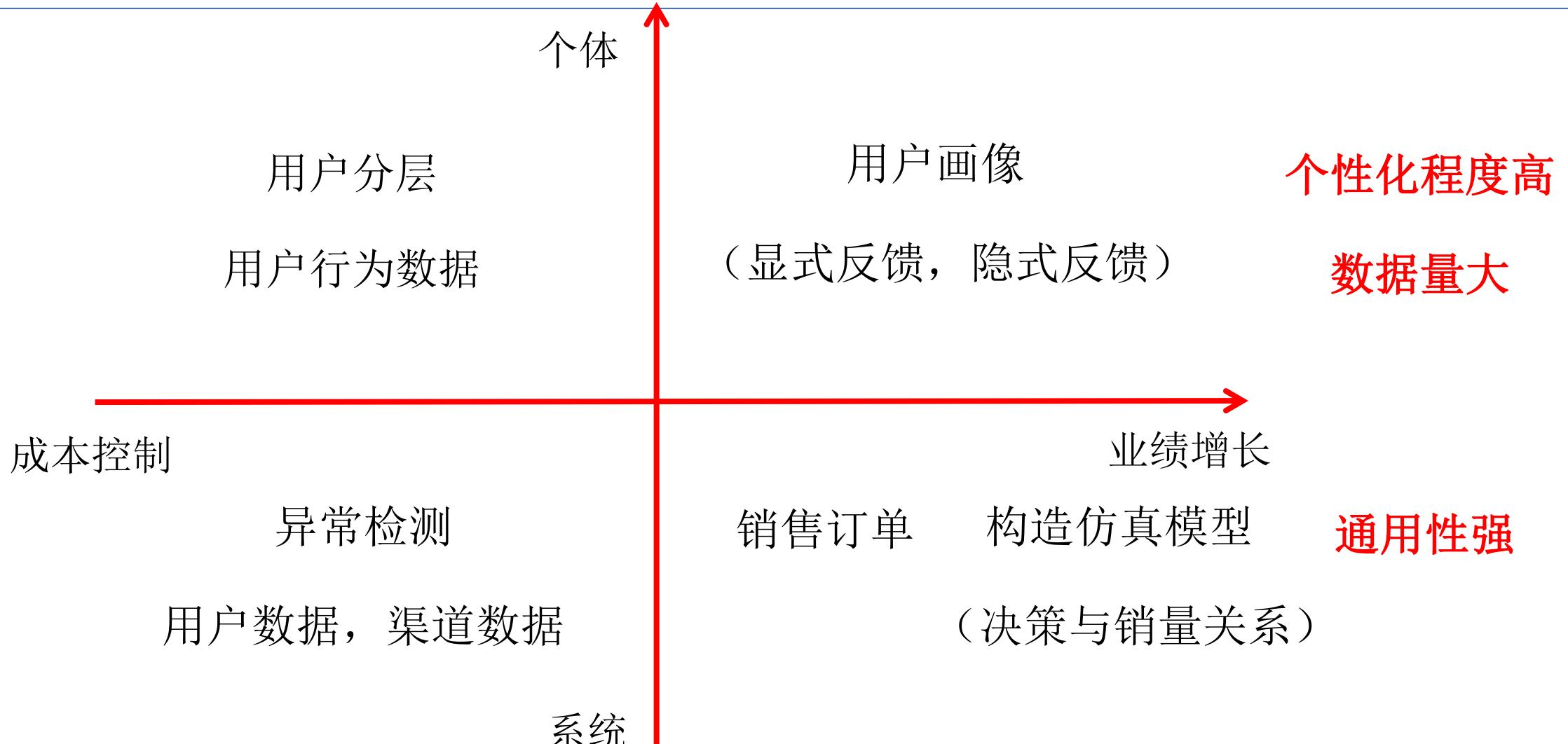
数据分析的目的是给企业提供业绩增长，全员的思维方式及行动指南

# 价值性：明确驱动目标



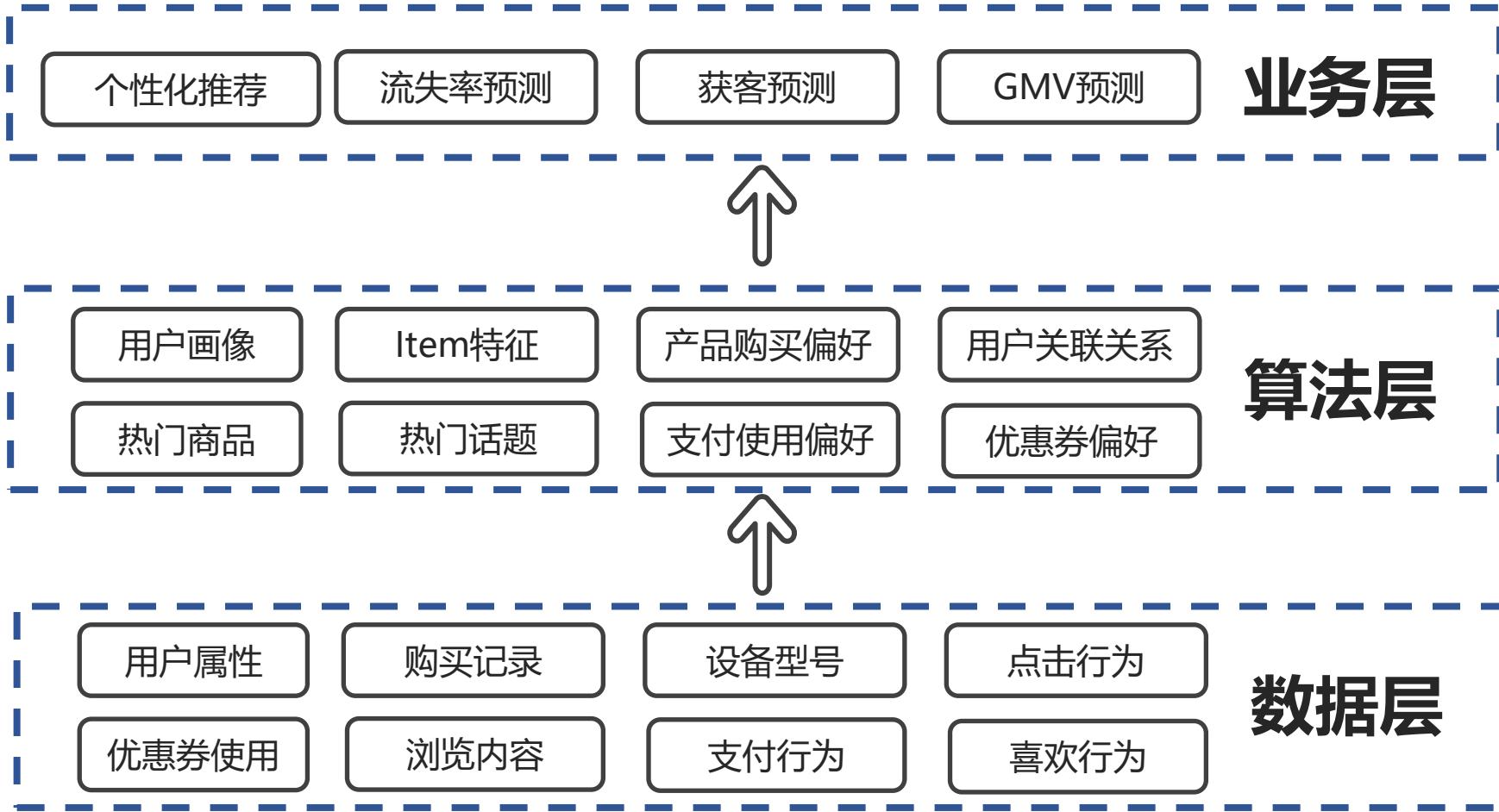
从4个象限中挖掘数据业务价值：从一线人员到系统整体，从成本控制到业绩增长

# 落地性：明确数据基础



业务价值目标导向：基于业务价值目标，制定数据及分析解决方案

# 按照数据流处理阶段划分



# 数据分析思维框架

# 分类思维：BCG矩阵分析



- **明星产品：销售增长率高，市场占有率高**

产品潜力大，现金流也充足，应该增加投资份额，扩大经济规模，用于巩固市场地位

- **现金牛产品：销售增长率低，市场占有率高**

市场增长缓慢，产品已经进入了成熟期，但该产品利润大，可以为企业提供资金的后盾

- **问题产品：销售增长率高，市场占有率低**

高增长代表行业前景好，但低市场占有率代表运营有问题，可能产品利润率低，需要改善运营手段

- **瘦狗产品：销售增长率低，市场占有率低**

销量不高，利润小，基本处于亏损状态，果断放弃，或整合到公司其他业务上

分类是一种思维，在不同场景中，可以灵活运用  
服务我们的目标：降本增效

# Kraljic采购定位模型

## 战略物资 (Strategic Items)

- 对产品生产流程至关重要的采购项目，往往由于供给稀缺或运输困难而具有较高的供应风险
- 买卖双方地位：力量均衡，相互依赖性较高
- 采购战略推荐：战略联盟，紧密联系，供应商尽早介入，共同创造，并充分考虑垂直整合，关注长期价值

## 瓶颈物资 (Bottleneck Items)

- 只能由某一特定供应商提供、运输不便、财务影响较低的采购项目
- 买卖双方地位：卖方主动，相互依赖性一般。
- 采购战略推荐：数量保险合同，供应商管理库存，确保额外库存，寻找潜在供应商



- 利润潜力：采购对可盈利性所做贡献的潜力，供应物品对利润的影响
- 供应风险：采购物品获得的难易程度，如果供应商没能准时供货，客户将遭受的损失程度

# Kraljic采购定位模型

## 杠杆物资 (Strategic Items)

- 可选供应商较多、能够为买家带来较高利润的采购项目，替换供应商较为容易，具有标准化的产品质量标准
- 买卖双方地位：买方主动，相互依赖性一般
- 采购战略推荐：采购招标，供应商选择，目标订价，与首选供应商达成一揽子协议，最后按正常供应程序执行、处理分订单

## 一般物资 (Non-Critical Items)

- 供给丰富、采购容易、财务影响较低的采购项目。具有标准化的产品质量标准
- 买卖双方地位：力量均衡，相互依赖性较低
- 采购战略推荐：通过提高产品标准和改进生产流程，减少对此类项目的采购投入



- 利润潜力：**采购对可盈利性所做贡献的潜力，  
供应物品对利润的影响
- 供应风险：**采购物品获得的难易程度，如果供  
应商没能准时供货，客户将遭受的损失程度

# 分类思维

---



分类思维的目的是什么?



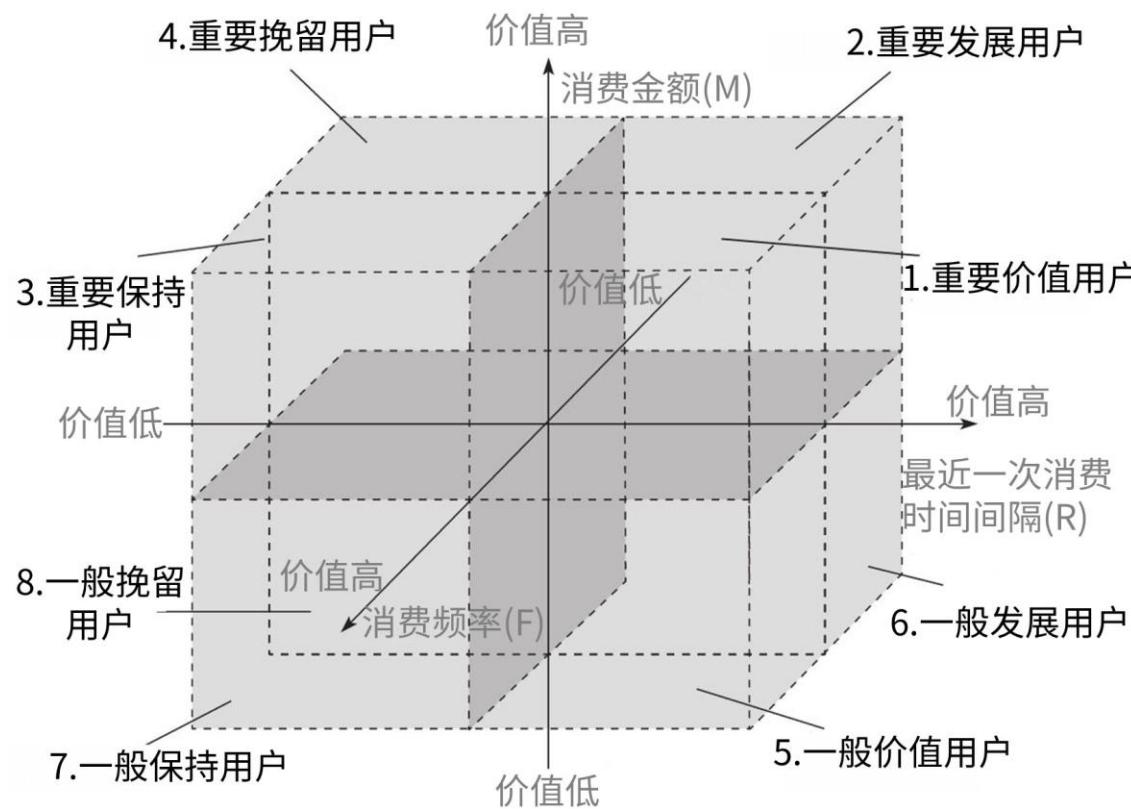
你的工作中有哪些场景用到分类思维?

# 分层思维：RFM顾客分层

Recency: 最近一次消费时间间隔

Frequency: 消费频率，一段时间（比如1年）内的消费次数

Monetary: 消费金额，一段时间（比如1年）内的消费金额



用户分类	最近一次消费时间间隔 (R)	消费频率 (F)	消费金额 (M)
1、重要价值用户	高	高	高
2、重要发展用户	高	低	高
3、重要保持用户	低	高	高
4、重要挽留用户	低	低	高
5、一般价值用户	高	高	低
6、一般发展用户	高	低	低
7、一般保持用户	低	高	低
8、一般挽留用户	低	低	低

# RFM用户价值模型

---

RFM指标使用场景：

- 一个店铺，某个月收入大幅下跌60%，通过数据分析，发现原来几个重要的用户被竞争对手挖走了（这几个用户贡献了店铺60%的收入）

Thinking: 出现这个问题的原因是什么？

因为店主没有对用户分类，也就是所有用户都采取相同的运营决策 => 用户分类，识别有价值的用户

比如第1类用户，RFM每个指标都很高的用户，是最重要的用户（最近一次消费较近，消费频率高，消费金额高）

=> 要提供vip服务

# RFM用户价值模型

RFM指标计算：

- Step1，计算用户的R、F、M打分

用户ID	最近一次消费间隔(R)	消费频率(F)	消费金额(M)	R值打分	F值打分	M值打分
1	4	1	4500			
2	1	12	1200			

R、F、M三个指标的打分：

按价值打分	最近一次消费时间间隔(R)	消费频率(F)	消费金额(M)
1	20天以上	2次以内	1000元以内
2	10-20天	2-6次	1000-1500元
3	5-10天	6-8次	1500-3000元
4	3-5天	8-15次	3000-5000元
5	3天以内	15次以上	5000元以上

# RFM用户价值模型

RFM指标计算：

- Step2，计算平均值

用户ID	最近一次消费间隔(R)	消费频率(F)	消费金额(M)	R值打分	F值打分	M值打分
1	4	1	4500	4	1	4
2	1	12	1200	5	4	2
价值平均值				4.5	2.5	3

- Step3，用户分类

用户ID	R值打分	F值打分	M值打分	R值评价	F值评价	M值评价
1	4	1	4	低	低	高
2	5	4	2	高	高	低

- Step4，对应表格，找到用户属于哪个分类

用户1：重要挽留用户（分类4）

用户2：一般价值用户（分类5）

# 帕累托 2/8法则

帕累托 2/8法则：

- 意大利经济学家帕累托发现

Thinking: 在国内，是否也有同样的规律？

招商银行 占比1.9%的金葵花客户，占所有零售客户资产的82%

日均资产在50万以上的客户 => 金葵花

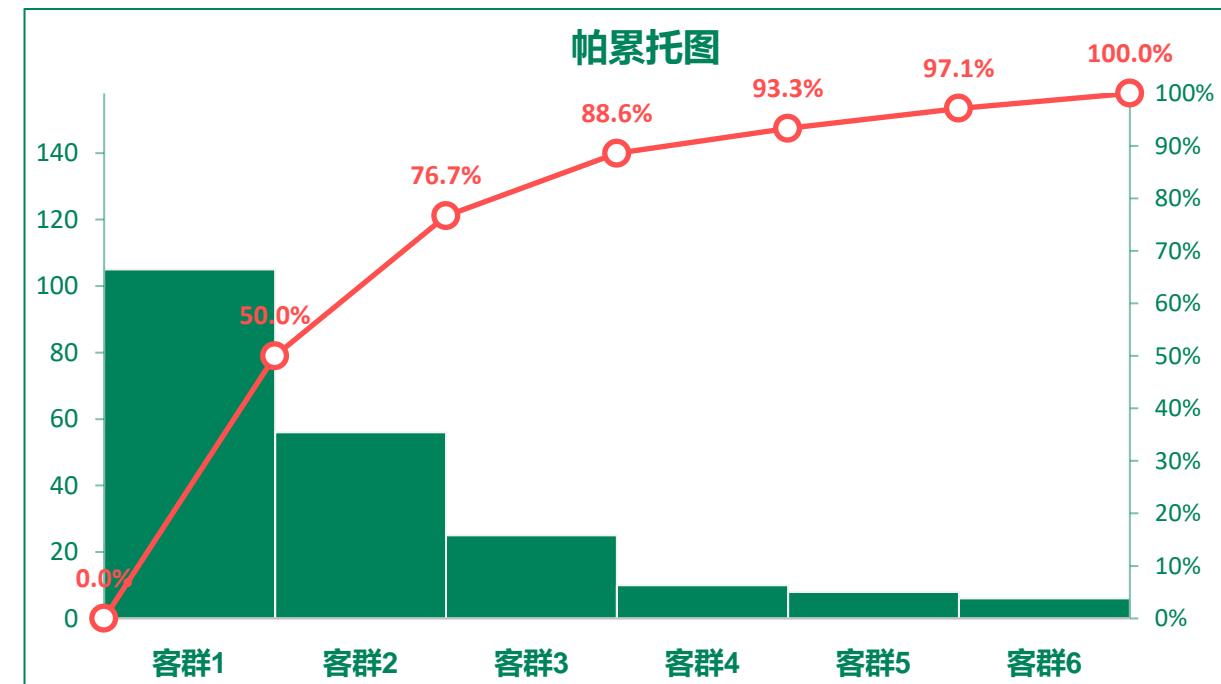
招商银行金葵花及以上客户资产演变							
年份	零售客户数(万户)	金葵花客户数(万户)	金葵花客户占比	零售客户总资产(亿)	金葵花客户资产(亿)	金葵花客户资产占比	金葵花户均资产(万元)
2007	3375	29	0.9%	8548	4710	55.1%	162
2008	3844	39	1.0%	10657	6593	61.9%	169
2009	4223	55	1.3%	14096	9444	67.0%	172
2010	4520	67	1.5%	17411	11716	67.3%	175
2011	5008	78	1.6%	20688	14363	69.4%	184
2012	5383	91	1.7%	23804	16720	70.2%	184
2013	4763	107	2.2%	28252	20525	72.6%	192
2014	5625	129	2.3%	34699	26007	75.0%	202
2015	6694	165	2.5%	47496	37296	78.5%	226
2016	9106	191	2.1%	55305	45408	82.1%	238
2017	10663	213	2.0%	61643	50614	82.1%	238
2018	12541	236	1.9%	68021	55082	81.0%	233
2019	14400	265	1.8%	74940	60852	81.2%	230
2020	15800	310	1.9%	89417	73455	82.1%	236

金葵花指招行月日均总资产在 50 万元及以上的零售客户。

# 帕累托 2/8法则

Thinking: 在客户服务中，是否也有2/8法则？

在我们的日常工作中，是否存在2/8法则？



# 分层思维

---



分层思维的目的是什么？



你的工作中有哪些场景用到分层思维？

# Summary

---

数据能干什么：

- 用数据定义市场空间
- 用数据定义运营成本
- 用数据发现商业机会
- 用数据发现羊毛党

常用数据分析思维：

- 分类思维，比如BCG矩阵
- 分层思维：比如RFM价值模型

Thinking: 在企业中的应用？

# 学员发言和呈现



在日常工作中，您认为有哪些数据驱动业务增长的场景和案例

- 发言的同学有积分哦！

No	业务场景	价值性	可行性



Thank You  
Using data to solve problems