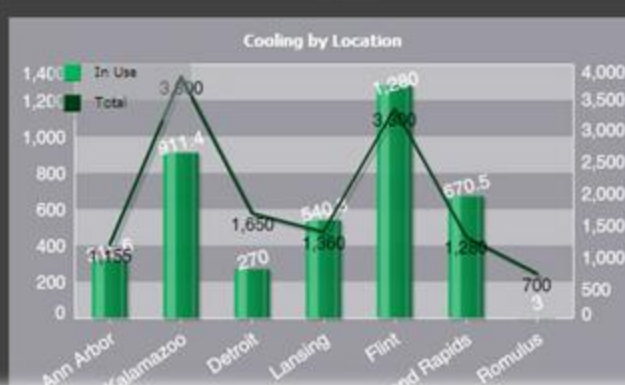
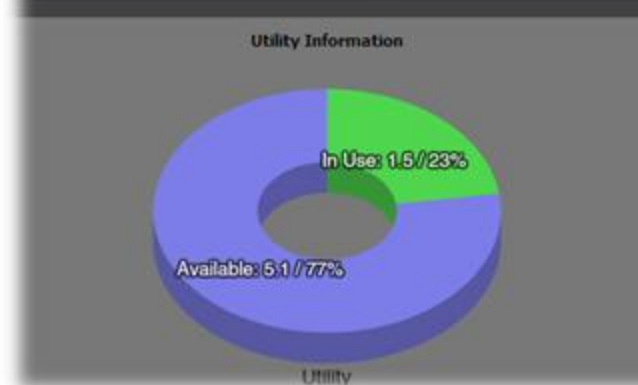
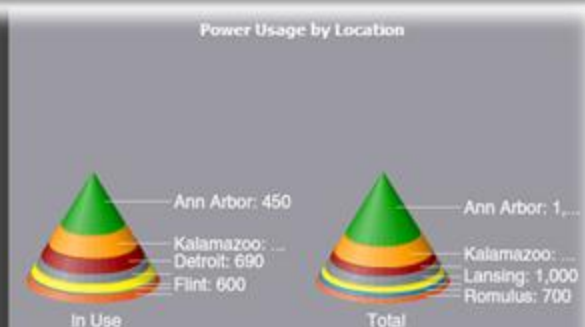
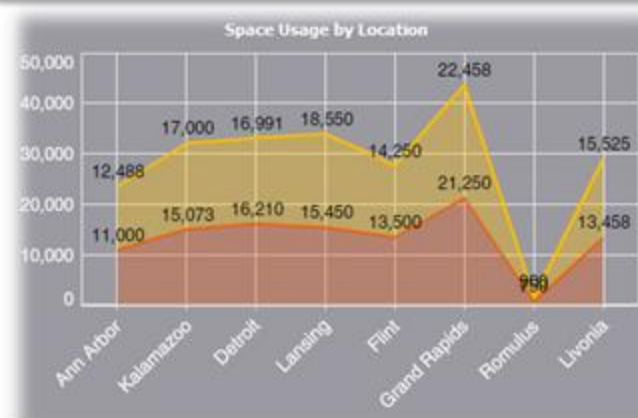


数据分析交流

新浪微博: @数据挖掘与数据分析 spss15.0@163.com



蓝精灵体数据分析师版：“在那山的那边海的那边，

有一群数据分析师，他们苦逼又聪明，他们每天看数据，他们呕心沥血不分昼夜都在赶报告，他们年复一年盼着涨工资。噢苦命的数据分析师，噢苦命的数据分析师，他们齐心协力开动脑筋找出数据背后的规律，他们的分析结果还是不被重视。”

目录

- 数据分析就这样产生了？
- 数据分析思路
- 数据分析方法
- 数据分析应用及工具





我们公司的
数据越来越多，
‘老大’整天问我
要这数据，要哪数
据，可俺不是做数
据分析的？我只会
种地。。。

‘数据big了’

数据分析驱动业务发展？

- 数据

- 大量的，运营型数据

- 用户和需求

- 业务决策支持的需求

- 技术和方法

- 计算能力的飞速发展

- 多学科领域研究的发展

计算机便宜了！



‘数据融合’

数据分析与业务的结合开战。。。

目标客户

市场推广

定价优化

用户保持



产品优化

受众细分

.....

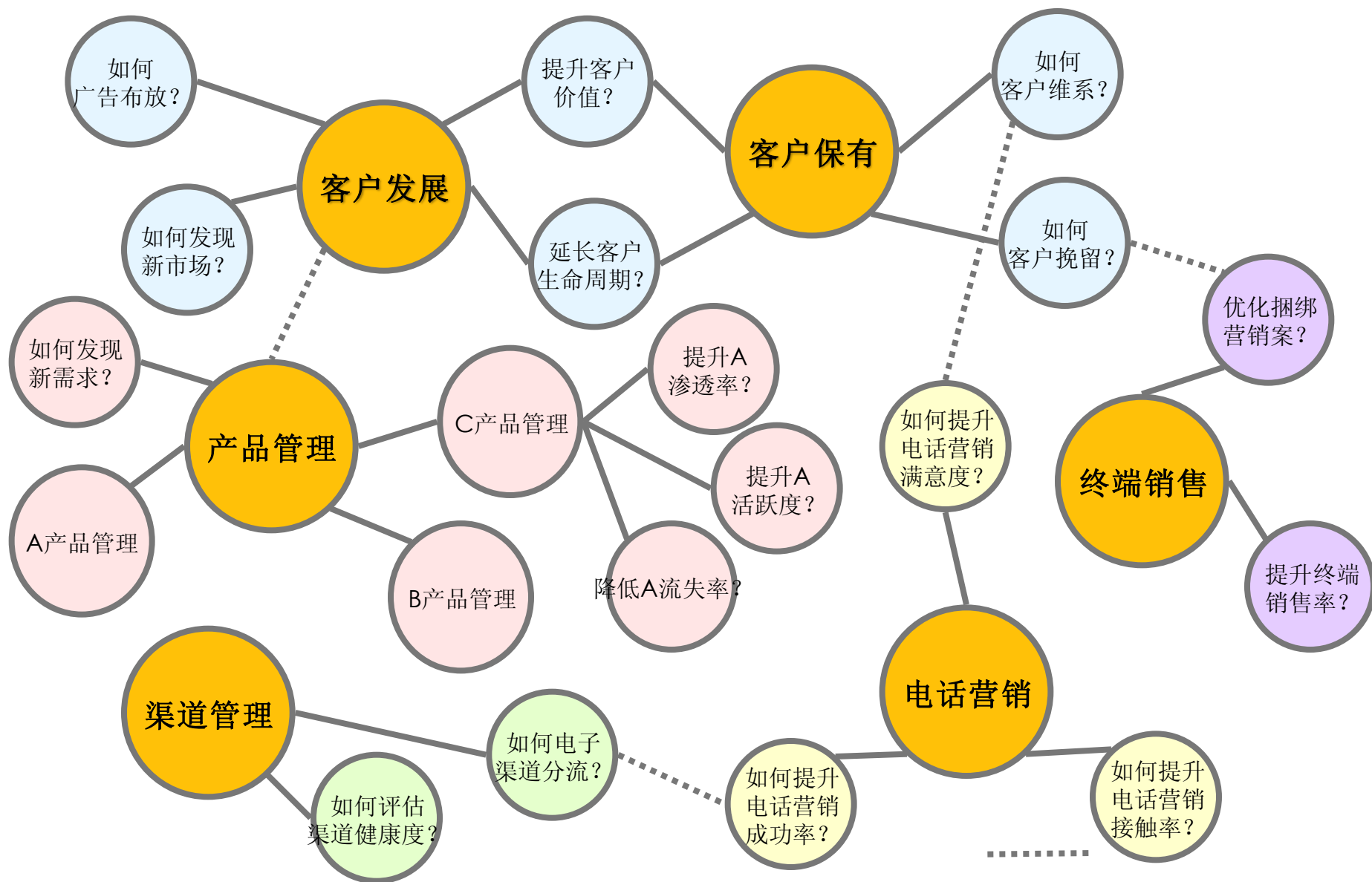
内容优化

数据分析木有思路，烦。。。。

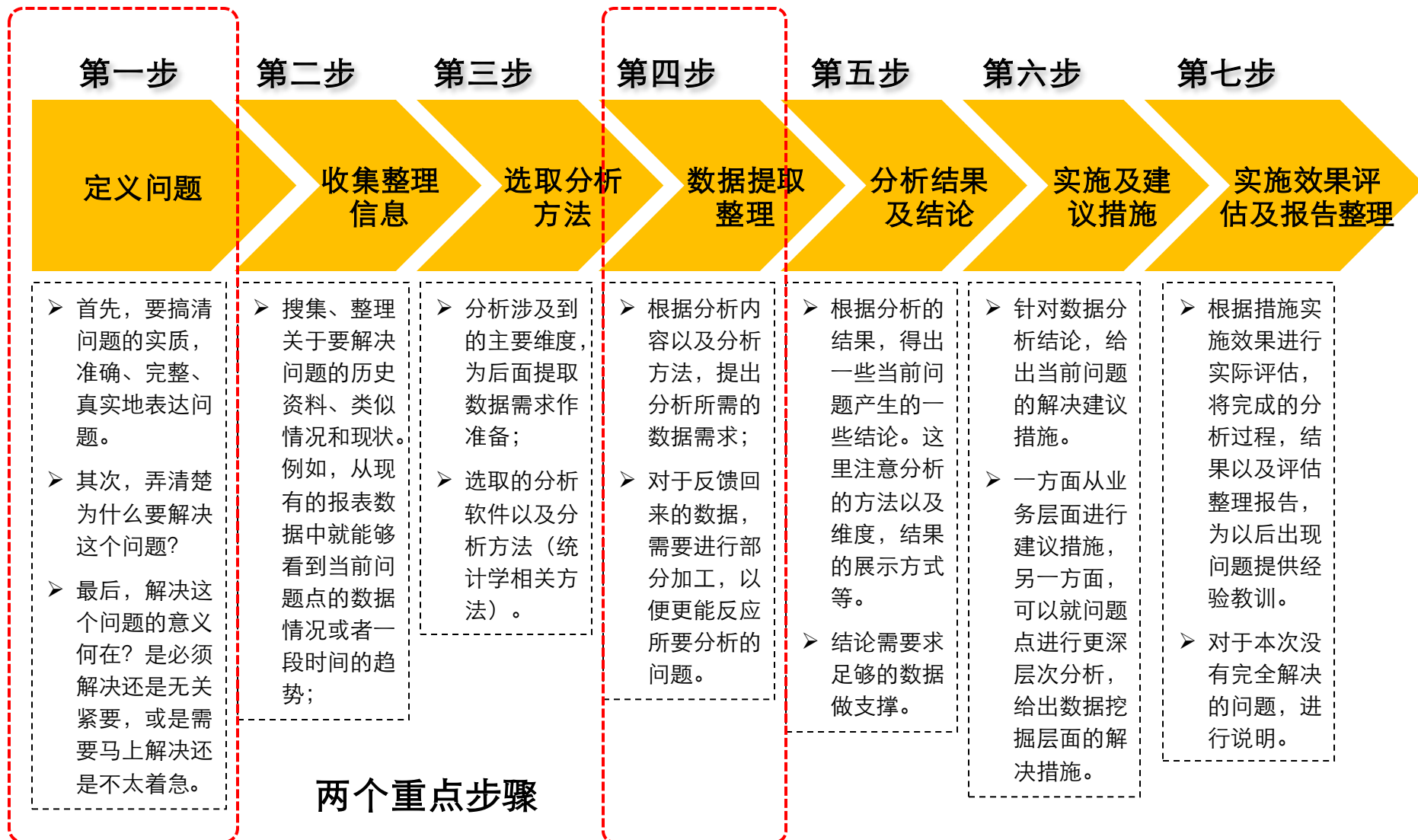


‘业务复杂’

不知从何而云？打破沙锅问到底！



‘思路决定价值’ 按规矩办事不会错的！



‘场景’ 从这里开始我们的旅行.....



市场总监:昨天咱们做EDM与push mail分别700万和1000万用户，但是实际使用

电子优惠券的用户不到5万，请告诉我原因是什么？



分析师A（技术出身）：

接到需求就开始去技术部，亲自操刀数据库，开始捞数据了。。。一会一个sql，一会一个sum、average，2天过去了。。。无果！



分析师B（业务出身）：

制作了问题分解图，罗列了许多因素，但是不知道提取什么数据？找不到真正驱动的因素；



分析师C（打杂出身）：

首先针对此次营销罗列了详细的操作环节，针对每个环节都列出主要的监控数据，最后找到了此次营销失败的原因！

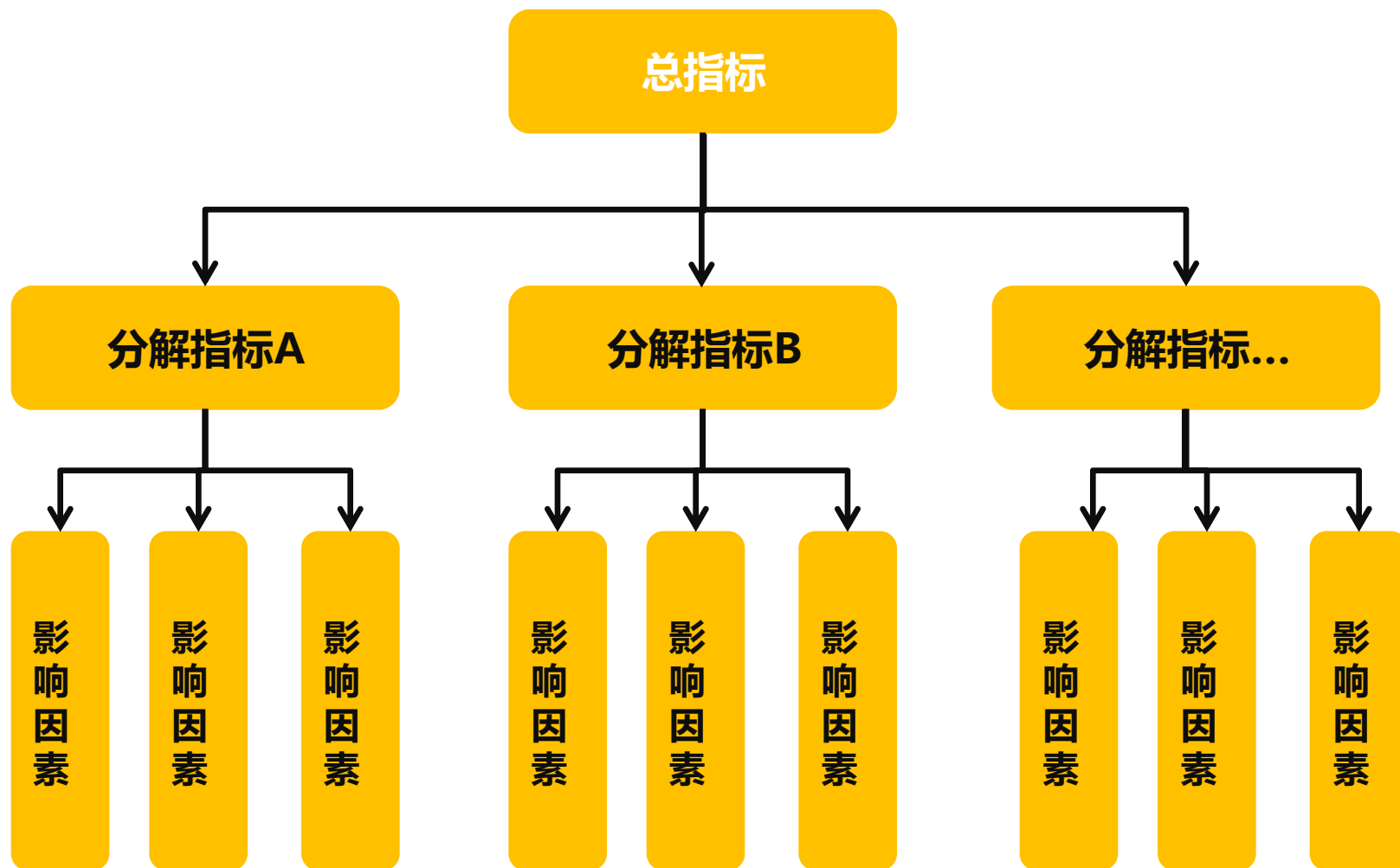
‘先动脑’ 别着急动手！

数据分析前，我们怎么去思考？

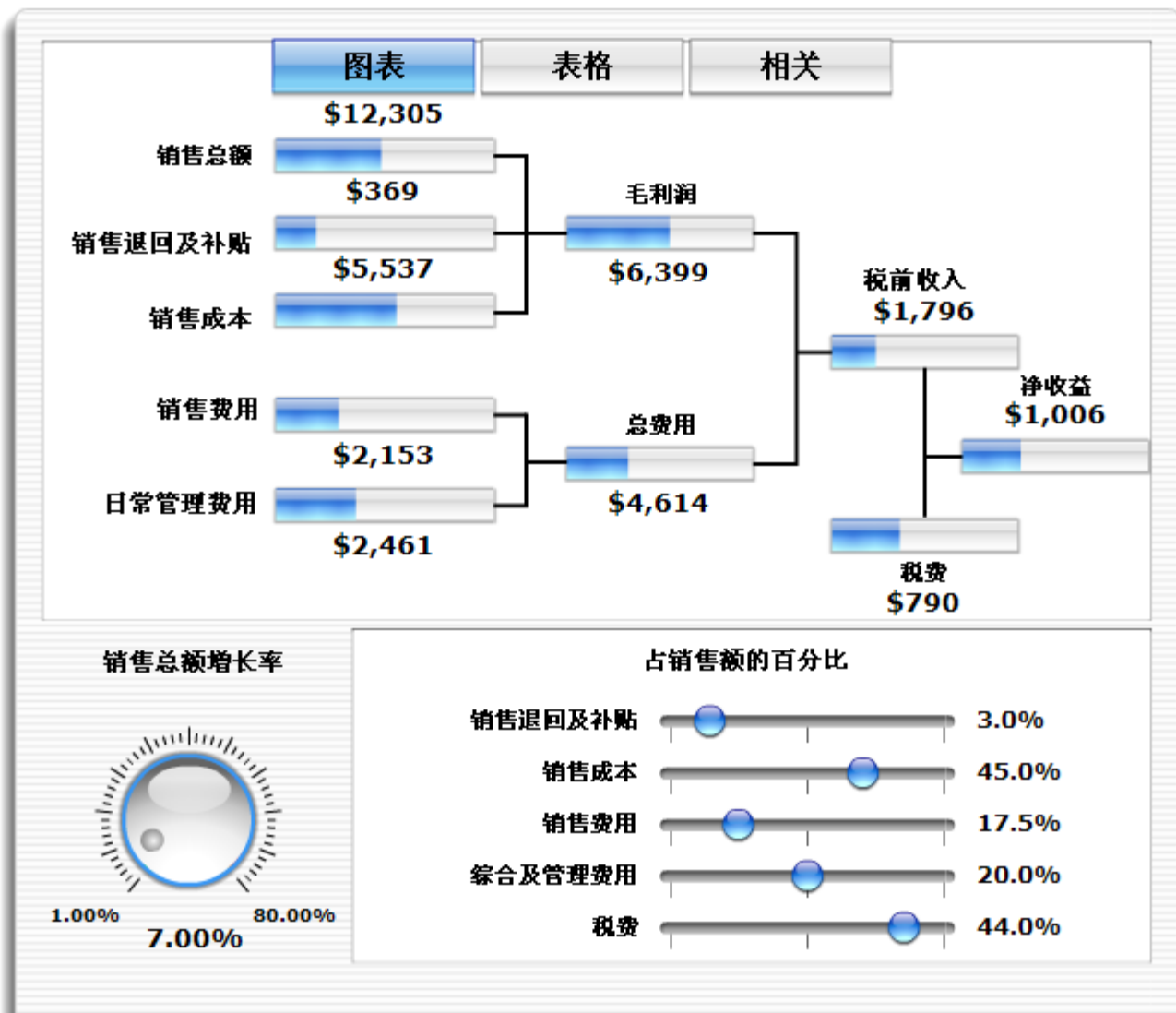


每一个步骤可能面临的问题以及需要准备的东西！！！！

‘指标拆分法’——数据分析师常用利器



‘锁核心、抓重点、握主干’ —— 数据分析师常用利器



数据分析方法杂谈



‘常用必杀技’ 要熟练掌握的技能

趋势分析

从时间序列分析随时间的变化趋势，找出其规律，如移动平均、同比、环比等。

对比分析

从相同的角度去对不同数据集合进行对比，找出差异所在，并可进一步深入挖掘差异原因；

异常分析

从大量历史数据中找出太高、太低、变化幅度过大等异常情况数据，支持预警显示、预警提醒；并可进一步进行相关影响原因的数据挖掘

排名分析

从大量数据中找出按某种分类方法的Top N或Bottom N数据，这些数据代表了需要特别关注的程度；

相关分析

测定现象之间相关关系

结构分析

在统计分组的基础上，计算各组成部分所占比重，进而分析某一总体现象的内部结构特征、总体的性质、总体内部结构变化规律的分析方法。

预测分析

利用决策树、回归分析、神经网络、时间序列等数据挖掘算法对客户响应、流失预警、收入预测、业务量预测、欠费预测等进行预测分析

‘趋势分析法’ 了解过去，脚踏今日，展望未来

同比

平均增长



时间序列、回归

环比

移动平滑

‘对比分析法’ 没有绝对的事，都是相对的！

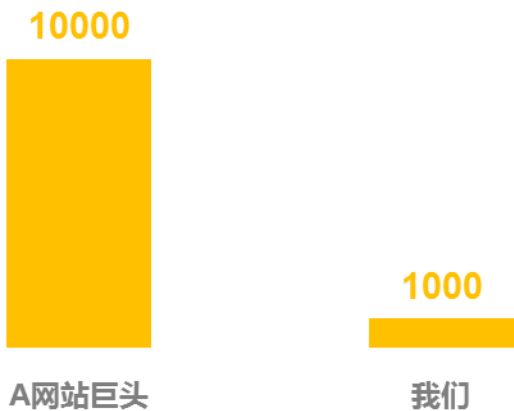
我们网站的日访问量为1000？

A说 ‘太菜’

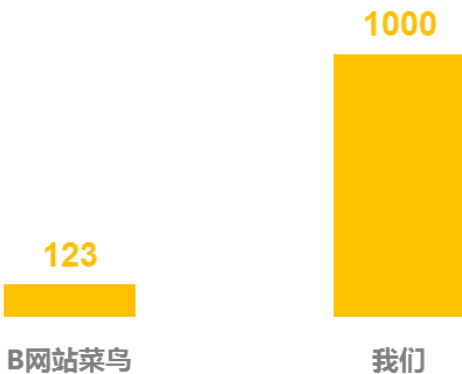
B说 ‘很牛’

C说 ‘还不错’

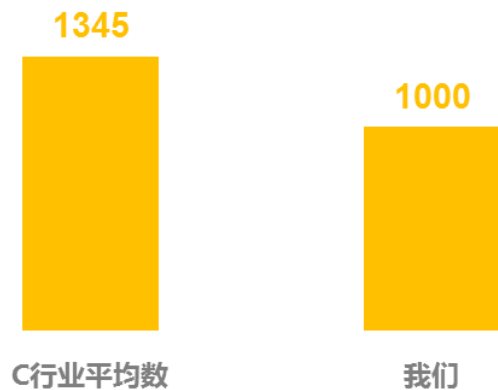
与巨头比？



与菜鸟比？



与平均比？

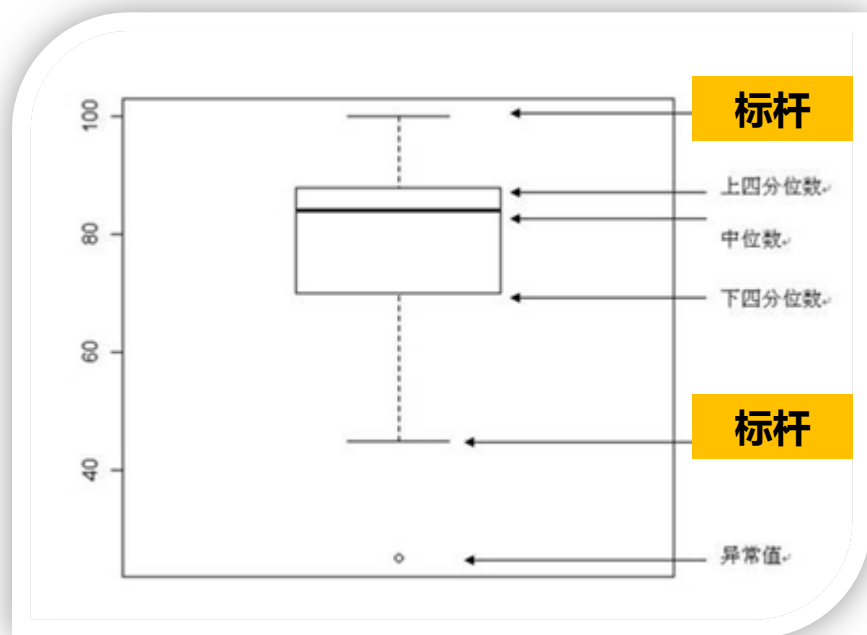


‘异常分析法’ 数据抽筋了吗？

- 有人说管理的核心就是异常情况管理。对企业经营管理中的各种经济现象中比较突出的部分进行分析就基本可以找到问题所在和解决办法。这种方法可以称为异常分析法。

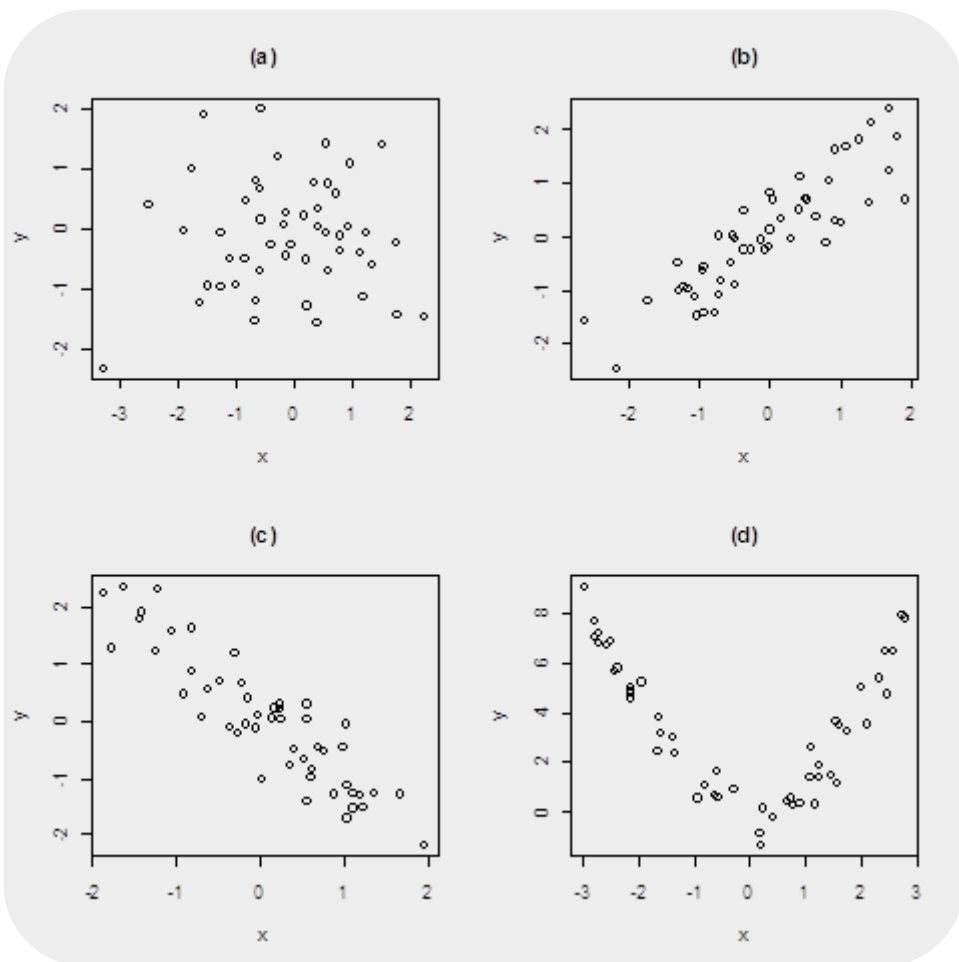
方法、设定标杆

按照“同一把尺子”观察分析不同的对象，因为业务一般在一定区间内浮动，建议设为区间，但要注意区间不易太大！

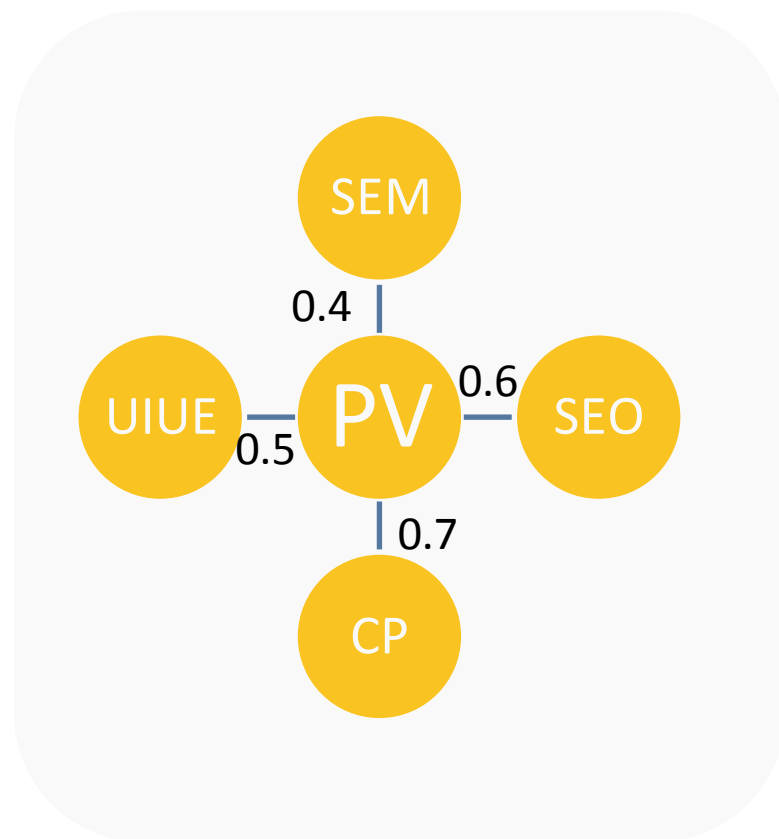


‘相关性分析’ 找关系好的

散点图



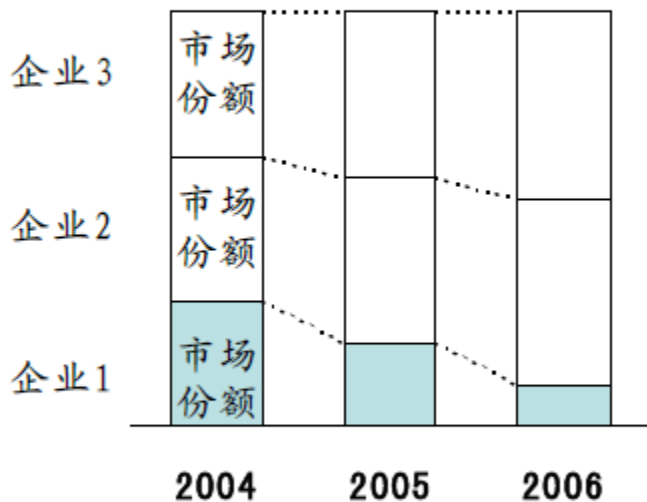
相关系数



‘结构分析法’ 咱们的蛋糕够大吗？

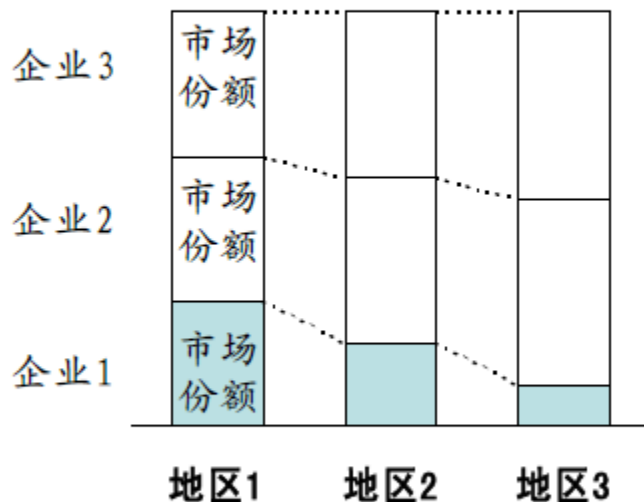
- 结构分析法是在统计分组的基础上，计算各组成部分所占比重，进而分析某一总体现象的内部结构特征、总体的性质、总体内部结构变化规律的分析方法。

总体份额对比分析



各企业的扩张与衰退情况

地区份额对比



各企业的优势地区及劣势地区



这些我都懂，太
Easy、太Easy！

知道这些分析
方法我就可以成为
优秀的数据分析师
了？

还有神马方法呢？

‘技术图谱’ “实用的方法才是最好的方法！”

	主要分析方法	应用优劣势
简单的数据分析	<ul style="list-style-type: none">■ 集中趋势分析：众数、中位数、均数■ 离散趋势分析：极差、方差、标准差■ 交叉表■ 数据加权	<ul style="list-style-type: none">■ 简单分析变量间关系■ 相对应用较广，较容易掌握和使用■ 对数据和使用者的要求较低
多元统计分析	<ul style="list-style-type: none">■ 相关分析■ 回归分析■ 因子分析■ 聚类分析■ 对应分析■ 联合分析	<ul style="list-style-type: none">■ 分析变量间的因果关系、相似度等■ 多用于预测、用户细分等场景■ 对数据要求：数据量要足够、数据周期要足够；■ 对用户的要求：需要掌握基本的统计学知识和对业务有一定理解；
数据挖掘	<ul style="list-style-type: none">■ 神经网络■ 决策树■ 等等	<ul style="list-style-type: none">■ 优势在于可以定制算法满足个性化需求和具备自适应和自学习性；■ 对数据要求：■ 对数据量和数据健壮性要求均很高；■ 对用户要求：对算法、业务的理解度均高

这些只是冰山一角！

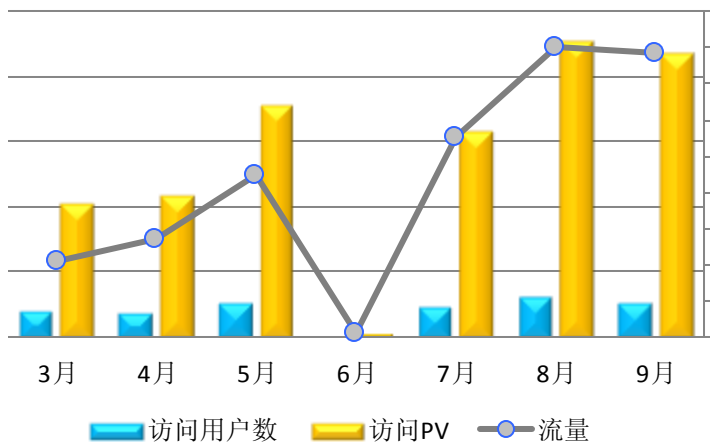
数据分析应用及工具



“数据分析应用”——一份报告中多数是？

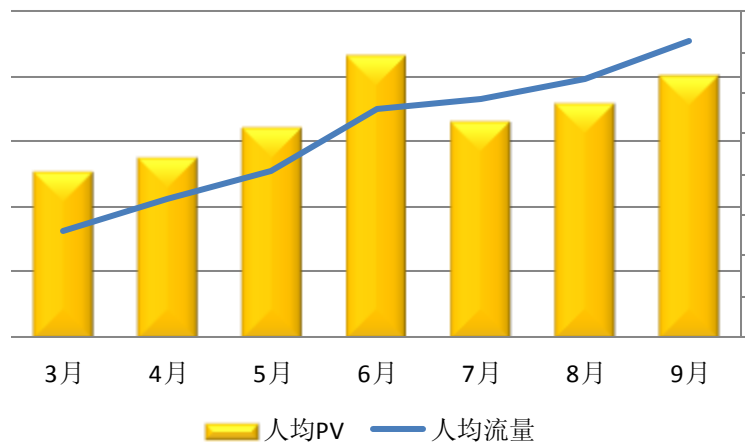
新闻频道访问用户数、PV、流量

流量: Mb 次/人



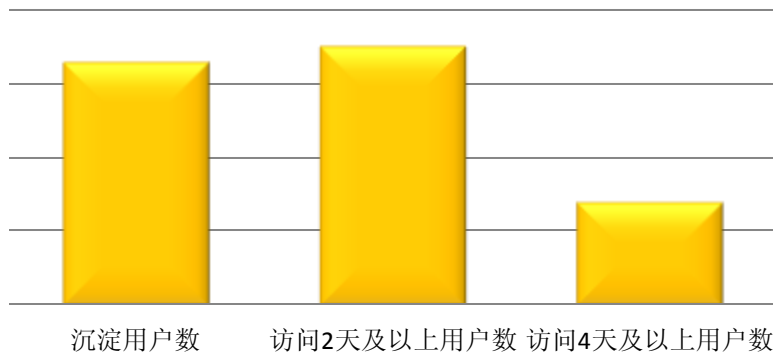
新闻频道人均PV、人均流量

K/人



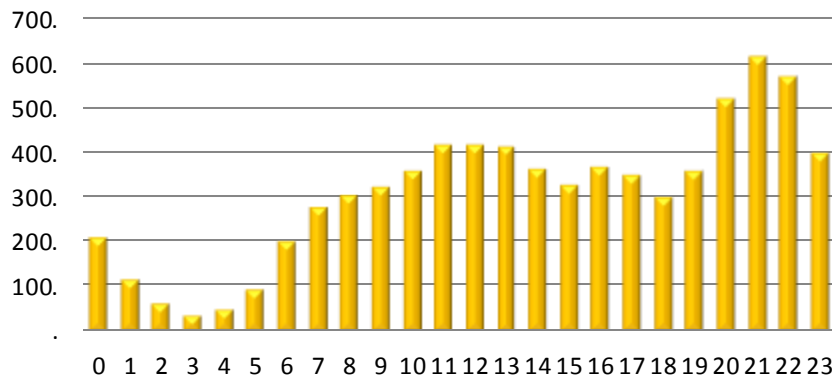
新闻频道重复回访、沉淀用户数

人



新闻频道各时段用户数

人



■新闻频道的PV和流量逐渐增高，其访问两次及以上用户数较多。

“数据分析应用”数据挖掘模型

- 通过数据挖掘决策树/LOGISTIC回归建模等方法，利用多数据整合对未直接表现出某项特征的客户进行行为预测（潜在偏好/活跃/异动）。

以潜在偏好模型为例

样本选取

选择表现出某项强烈特征或行为的客户做为样本（如：新闻类内容偏好），预测其余客户具有同类行为趋势的评分。

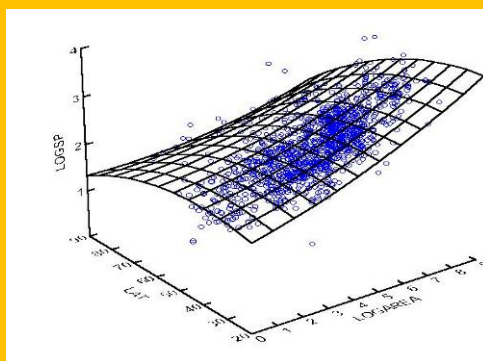
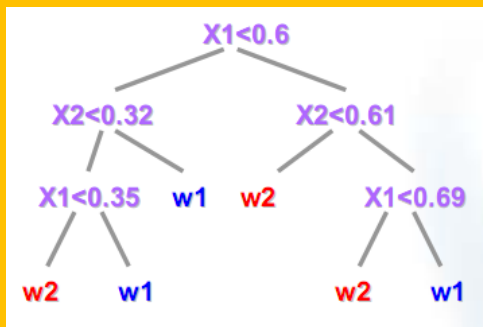
表现偏好飞信客户

客户号码	飞信客户	内容偏好	终端	交往圈	...
139**	1	聊天类(87%)	功能	23	
136**	1	聊天类(89%)	智能	35	
136**	1	娱乐类(72%)	智能	46	

未表现偏好飞信客户

138**	0	聊天类(69%)	智能	10	
137**	0	新闻类(87%)	智能	24	
135**	0	娱乐类(58%)	功能	8	
...	

模型选择



预测结果

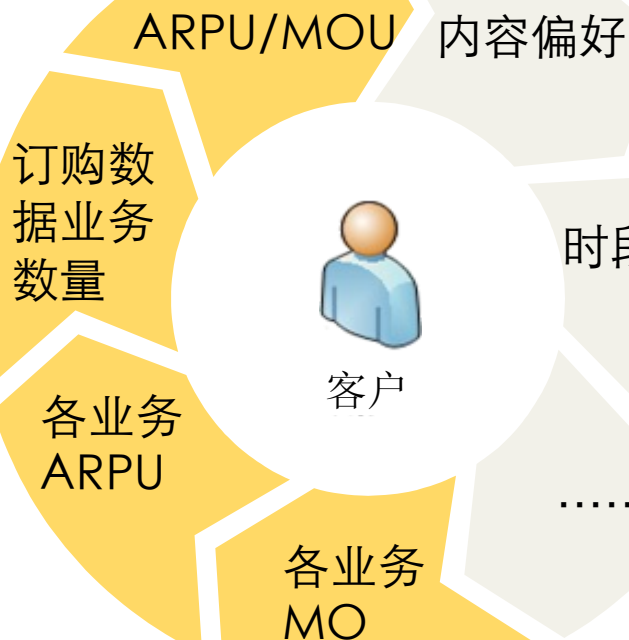
挖掘模型可以对未表现显性偏好的客户进行评分
评分越高的客户偏好新闻类内容的可能性越大

客户号码	新闻类页面浏览时长	评分
138**	0	0.65
137**	2.0	0.87
135**	0	0.81
158**	0	0.17
...

“数据分析应用”推荐系统

基于客户消费的数据

过去我们的分析中大量使用这类数据，它帮助我们了解客户的消费能力。



基于客户行为的数据

越来越多的互联网企业致力于挖掘客户行为数据，并使其发挥出了魅力。

豆瓣读书

豆瓣猜

数据源：用户的读过、想读、在读、评价行为（偏客户行为的数据源）

什么是豆瓣猜

这是豆瓣给你的个人推荐。通过每天分析你的读过、想读、在读、评价行为，豆瓣会从海量数据里挑选你会感兴趣的内容给你。

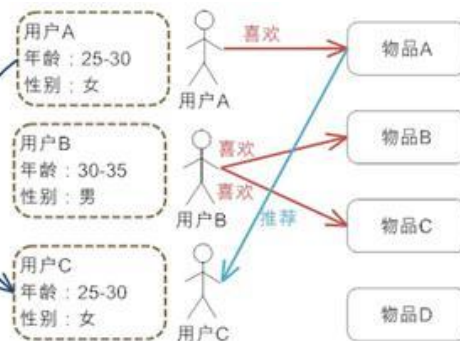
你用得越多，豆瓣猜得就越准确。

“你可以试着从更多的角度了解你的客户，你越了解他们，便越能获利。”

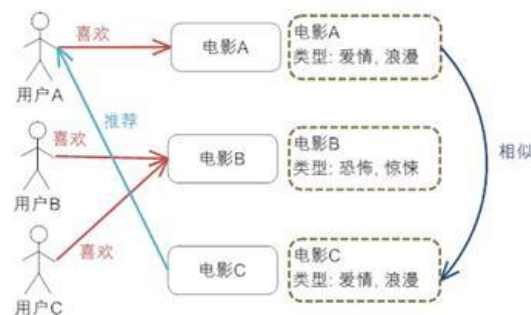
——【美】数据分析专家，Walter Janowski

简单的

人口统计学



内容



‘兵器厂房’ 你懂几个？



‘微博作坊’

业余时间要看看？

数据分析

@沈浩老师 @文彤老师
@刘德寰 @车品觉
@数据化管理 @william_ou
@张磊IDMer
@小蚊子乐园
@数据挖掘与数据分析
@郑来轶
@庖丁的刀
@sxlion
@文心武行
@夏之朝露
@朝阳小魔

数据分析与展现

@刘万祥ExcelPro
@MrExcel
@演绎生动Excel图表
@每日图表
@图表汇
@数据知天下
@EasyTrace网站分析
@全国数据分析业协会
@ SAS中文论坛
@数据挖掘研究院
@中国统计网
@网站分析星期三

PPT制作

@蝇子
@栗子面
@秋叶语录
@西岭夜雪
@曹将PPTao
@Lonely_Fish
@刘俊0730
@小田_PPT
@我爱PPT
@PPT精选
@无敌的面包
@教育技术秋记

‘好书连连看’ 书非借而不能读书也？

