



2014中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2014



大数据技术探索和价值发现

天猫商品详情页满意度模型

周祥军

天猫 TMALL.COM



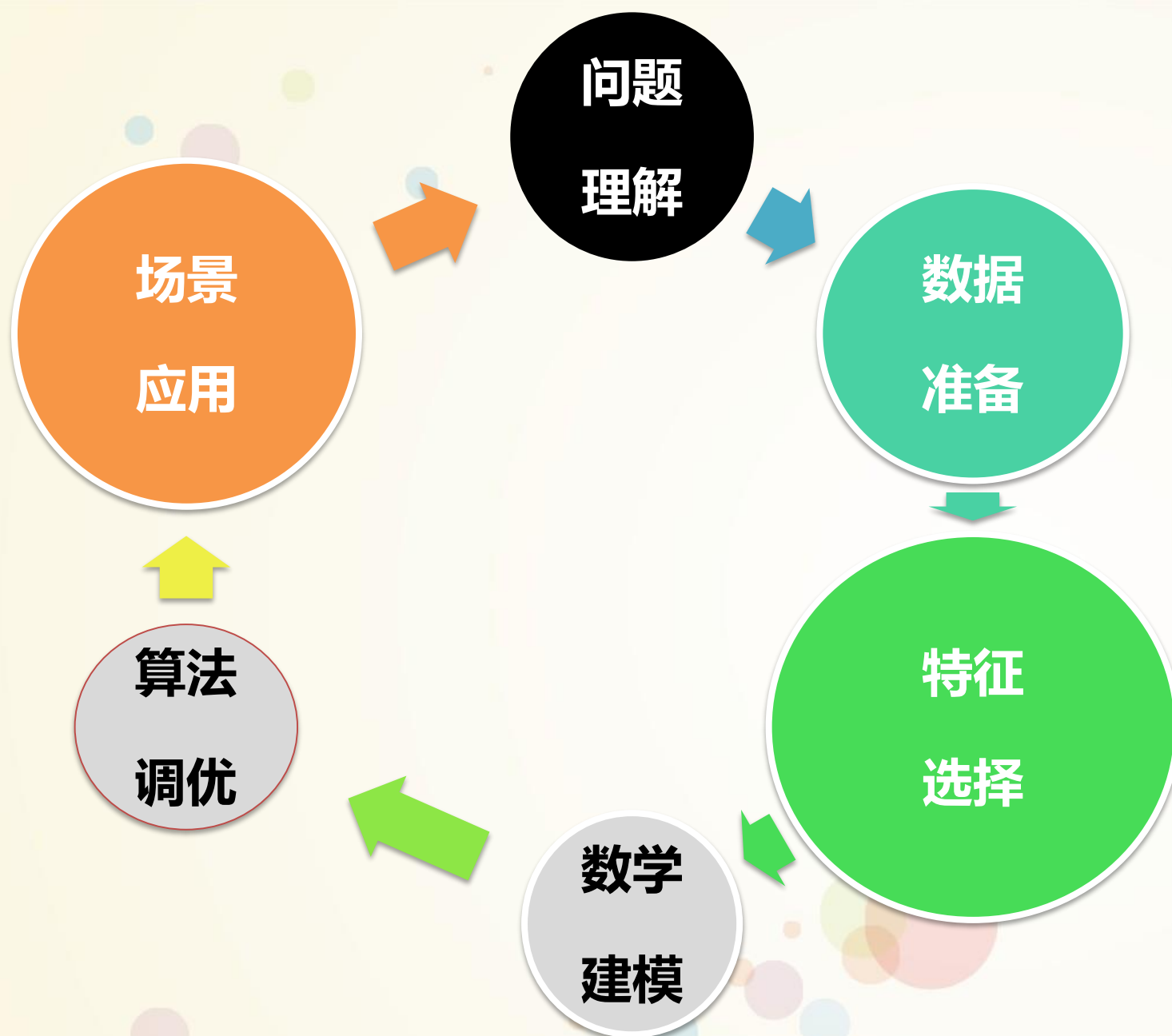
关于

- 个人

- 2009~2011：阿里妈妈，计算广告
- 2011~2013：百度商务部，竞价策略优化
- 2013~至今：天猫搜索

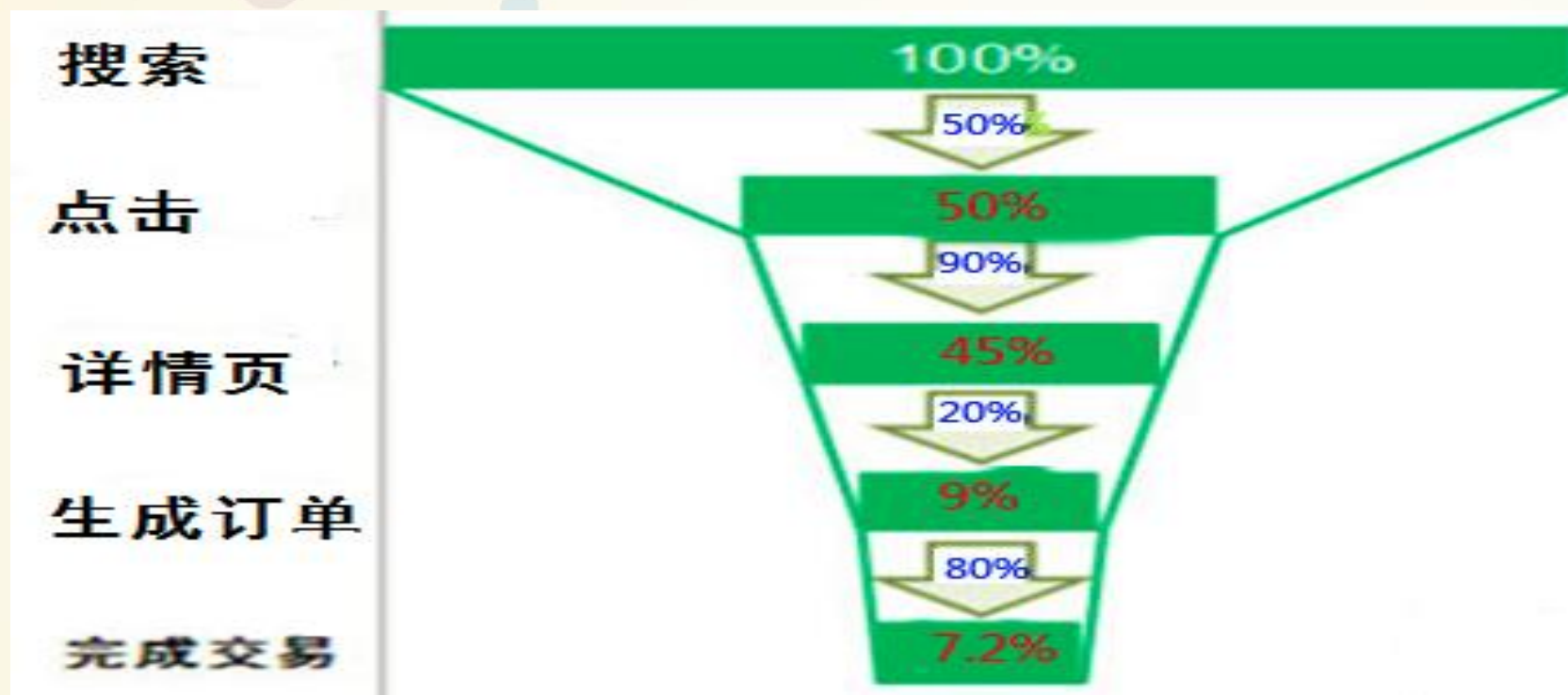
- 分享主题

- 基于商品详情页用户行为建模及应用



问题理解

天猫 TMALL.COM



	浏览	购物车	订单	支付	完成交易
人数	2071	622	284	235	223
上一步转化率	100%	30.0%	45.7%	82.7%	94.9%
总体转化率	100%	30.0%	13.7%	11.3%	10.8%

问题

- **成交稀疏**

- 搜索购物路径：Query → SRP → Detail → 成交
- CTR：浏览到点击 50%
- CVR：详情页到购买 **3%**
 - 那么对于那97%没有成交的IPV是不是就是不好的商品呢？

- **目标**

- 提升转化率：**更好的**商品排序靠前
- 基于用户每次在商品详情页面的各种**行为量化**用户对当前一次浏览的**满意程度**

Decathlon/迪卡侬

DECATHLON 迪卡侬



迪卡侬正品 户外 冲锋衣 男 单层薄款 情侣款 防风防雨 QUECHUA

新款即将上市,新旧款胸前logo处稍有区别 请亲们注意哦

价格 **¥199.00**

江苏苏州 至 杭州 快递: 5.00

店铺优惠 满299元,包邮;

更多

604 月销量

3736 累计评价
★★★★★ 4.8

99 送天猫积分

尺码

S M L XL XXL

颜色分类



数量

1 件 库存158件

立刻购买

加入购物车

服务承诺

按时发货

极速退款

退货保障卡

支付方式











看了又看

DECATHLON 迪卡侬



¥249.00

DECATHLON 迪卡侬



¥279.00

DECATHLON 迪卡侬



¥199.00

自由搭配



迪卡侬正品 户外 冲
本店
¥199.00

迪卡侬 户外帽子男
本店
¥84.00
搭配省5.00元



迪卡侬 皮肤风衣 男
本店
¥145.00
搭配省4.00元



迪卡侬 正品 户外
本店
¥68.90
搭配省0.10元



迪卡侬 蓝色货品 户
本店
¥39.00
搭配省0.90元



迪卡侬正品 户外登
本店
¥29.90

套餐价: ¥199.00 × 1

立即购买套餐

加入购物车

店铺
优惠

全国包邮
2014.03.09 - 2014.05.31
活动商品

满 299 元, 包邮

不包邮地区: 台湾, 海外;

商家信息

迪卡侬旗舰店

店铺动态评分 与同行业相比
描述相符: 4.8 高于 21.95%
服务态度: 4.8 高于 16.95%
发货速度: 4.8 高于 36.67%

进入店铺

收藏店铺

商品详情

累计评价 3736

月成交记录604件

扫一扫, 手机购买



快速直达

品牌名称: Decathlon/迪卡侬

关注

产品参数:

是否商场同款: 是

上市时间: 2013年冬季

吊牌价: 199

是否现货: 现货

适用对象: 男

价格区间: 101-200元

货号: 8171815

主流品牌: Decathlon/迪卡侬

商品系列: ARP300

材质: 涤纶

里料材质: 涤纶

适用季节: 冬季 春季 秋季

防水透气指数: 5000mm及以下

款式: 2.5层压胶夹克

服饰工艺: 接缝全压胶 无缝对接

产地: 中国

颜色分类: 新款黑色 卡其色 新款蓝色 ...

主要功能: 防水 防风 透气

运动户外项目: 郊游 野营 登山 其他 徒...

面料科技: 其他

尺码: S M L XL XXL XXXL

- 优惠信息
- 模特图
- 功能展示
- 细节图
- 自定义
- 服务质量

意见反馈



0





户外山地运动

男式选择 女式选择
儿童选择

上装

抓绒衣 冲锋衣 羽绒服
T恤 衬衫 休闲夹克

下装

长裤 短裤/中裤

服装附件

帽/围巾/手套/眼镜
户外雨披

户外鞋袜

徒步登山鞋 保暖鞋靴
户外凉鞋 户外袜

户外包

登山背包 日常背包
背包附件

越野装备

登山杖及附件

野营运动

两秒快开帐篷 登山帐篷
睡袋睡床 户外休闲家具
野营及野营配件

防水透气指数: 5000mm及以下

款式: 2.5层压胶夹壳

服饰工艺: 接缝全压胶 无缝对接

产地: 中国

颜色分类: 新款黑色 卡其色 新款蓝色 ...

主要功能: 防水 防风 透气

运动户外项目: 郊游 野营 登山 其他 徒...

面料科技: 其他

尺码: S M L XL XXL XXXL

SPU功能特征: 防水, 防风



关于发货 请您注意:

亲爱的顾客您好, 为了更快速的将商品送到您的手中, 我们已在河北燕郊开设了新的仓库, 现在正在试运行中, 在此您的订单可能会出现从两个地方发出, 如有对您造成不便, 请您谅解. 希望我们的改变能给您更多的便捷。



轻装趣徒步

户外男式徒步鞋
轻盈 | 缓震 | 抓地 | 高性价比

更低价格 99元 >> 查看详情



1 2 3 4

随手带一件



成人户外棉袜 2双装



成人户外运动帽



成人户外防晒帽



成人户外多功能头巾

- 优惠信息
- 模特图
- 功能展示
- 细节图
- 自定义
- 服务质量



商家信息

迪卡侬旗舰店

店铺动态评分 与同行相比
描述相符: 4.8 高于 21.95%
服务态度: 4.8 高于 16.95%
发货速度: 4.8 高于 36.67%

进入店铺 收藏店铺

10 YEARS

登山背包
十年质保

查看详情 >

户外山地运动

男式选择 女式选择
儿童选择

上装
抓绒衣 冲锋衣 羽绒服
T恤 衬衫 休闲夹克

下装
长裤 短裤/中裤

服装附件
帽/围巾/手套/眼镜
户外雨披

商品详情

累计评价 3736

月成交记录604件

扫一扫,手机购买

与描述相符

4.8

非常不满 不满意 一般 满意 非常满意

大家认为

偏小 适合 偏大

尺码: 78%

有色差 无色差

色差: 92%

查看追加 (127)

有内容评价

按时间

按信用

按推荐

防水性不是很优秀,但做工不错

今天

颜色分类: 新款绿色

尺码: S

j***4 (匿名)

商品: 衣服不错 推荐 大小也合适 就是袖子有点长了 不过 还是喜欢老款 可惜没有 了 新款 也是不错的喔

服务: 很好的

今天

颜色分类: 新款黑色

尺码: M

身高: 170cm

体重: 70kg

海***文 (匿名)

商品: 很好, 颜色很鲜亮, 和图片一样, 袖子偏长, 总体很好

服务: 很好

今天

颜色分类: 新款蓝色

尺码: M

历***7 (匿名)

一直想买一件冲锋衣, 其实我不是户外就是平时穿。看到别的评价有的说衣服袖子长。我想是不是自己身材的问题吧。确实是比平时休闲装长一点, 但是我觉得还好吧, 看朋友们穿的这类衣服都是这样的。反正我很满意。版型什么的都不错。满5分。

04.08

颜色分类: 新款黑色

尺码: M

身高: 177cm

体重: 67kg

九日不出门

很好

04.08

颜色分类: 新款绿色

尺码: XL

x***3 (匿名)

衣服不错, 就是袖子过长!

颜色分类: 新款黑色

尺码: S

大唐来的施主

本店同类好评商品

DECMON 迪卡侬

¥249

DECMON 迪卡侬

¥249

DECMON 迪卡侬

¥279

DECMON 迪卡侬

¥149

DECMON 迪卡侬

意见反馈

热力图



天猫积分 送 2 倍 共 248 | 支付 信用卡分期 快捷支付 19038

服务承诺 商品: 按时发货 · 退款: T3 极速退款 · 退货: 上门退货 赠运费险

商家信息

甘甘的世界服饰旗舰店

店铺动态评分 与同行业相比

描述相符: 4.8 高于 22.75%

服务态度: 4.8 高于 23.37%

发货速度: 4.7 低于 0.84%

进入店铺

收藏店铺

搭配套餐1



雪纺连衣裙 ...
原价: 399.00

+



小西装女修...
原价: 556.00

6166415.0

+

+

套餐价格: 546.00

节省: 409.00

查看套餐

5800219.0 情

2.0117297E7 992

3137048.0 记录 1236 件

看了又看

商品详情tmalldetail-tab-成交(新)
tmalldetail.4.13

用户PV: 2790557.0
用户UV: 1372430.0
访问PV: 347091.0
访问UV: 240995.0

店铺介绍

模特平铺图

细节图展示

服务质量

单笔订单满 1 元, 免

999 元, 免运费;

499 元, 免运费; 单笔订单满 999 元减 100 元,

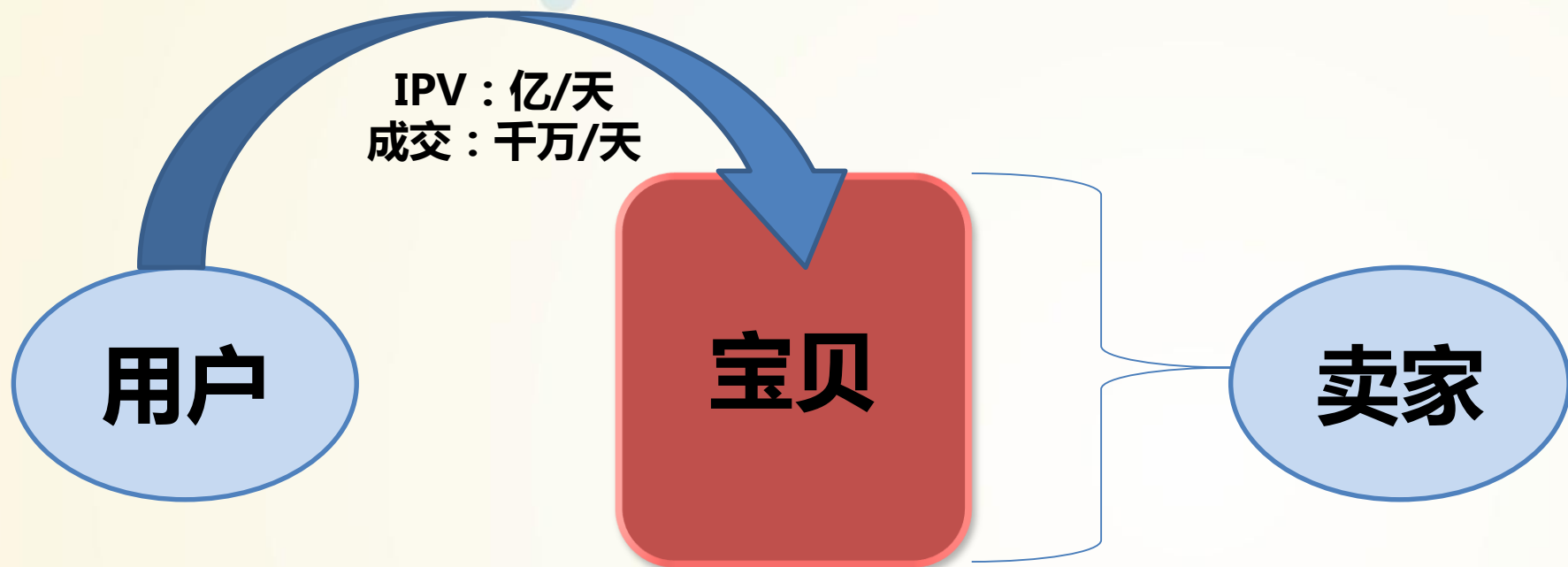
详情页框架





数据准备

天猫 TMALL.COM



- 4亿+
- 千万(在线)

- 商品 : 8000W+
- 品牌 : 10W+
- 行为 : 上百种

- 10W+

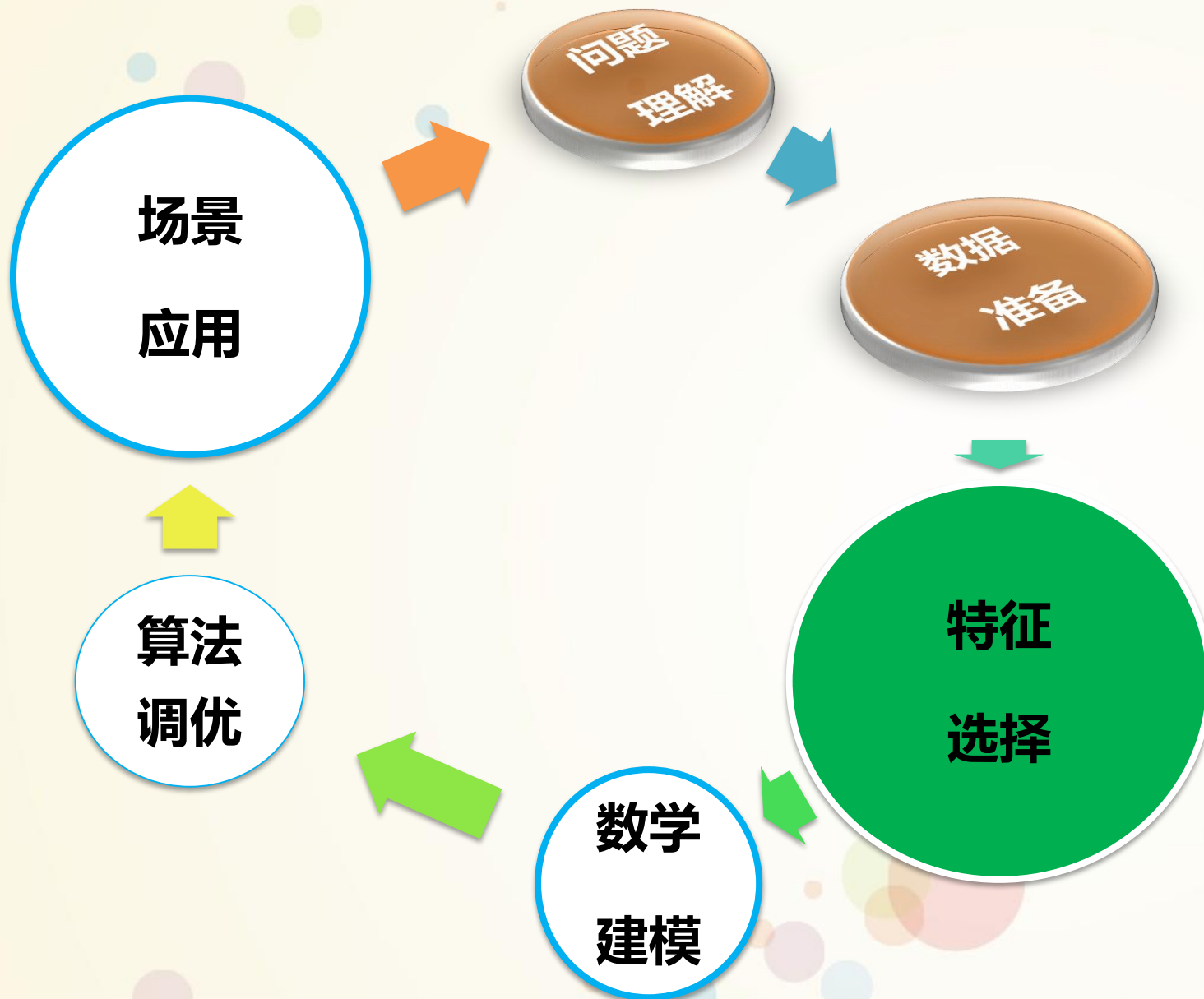
收集数据

- **历史数据**

- 用户
- 商家
- 商品

- **新数据**

- 获取：根据经验定义新类型数据，前端和系统配合
- 积累：模型需要一定的数据积累(7/30/90/180天)



特征选择

天猫 TMALL.COM



用户特征	店铺特征	宝贝特征	详情页特征	流量特征
行为 <ul style="list-style-type: none">• 页面点击• 点击次数• 停留时间• 分享• 收藏• 购买• 访问时间 信息 <ul style="list-style-type: none">• 性别• 年龄• 等级	NID DSR 城市	NID 类目 价格分档 好评率 退款率 周销量 月销量	页面长度 页面宽度 页面分屏数 浏览器 设备	流量来源 <ul style="list-style-type: none">• 搜索• 活动• App• <u>Wap</u>

- 单维特征

- 特征组合

- <类目, X>

- <来源, X>

- 特征处理

- 去噪(反作弊)

- 离散化

- 归一化

- 特殊变形

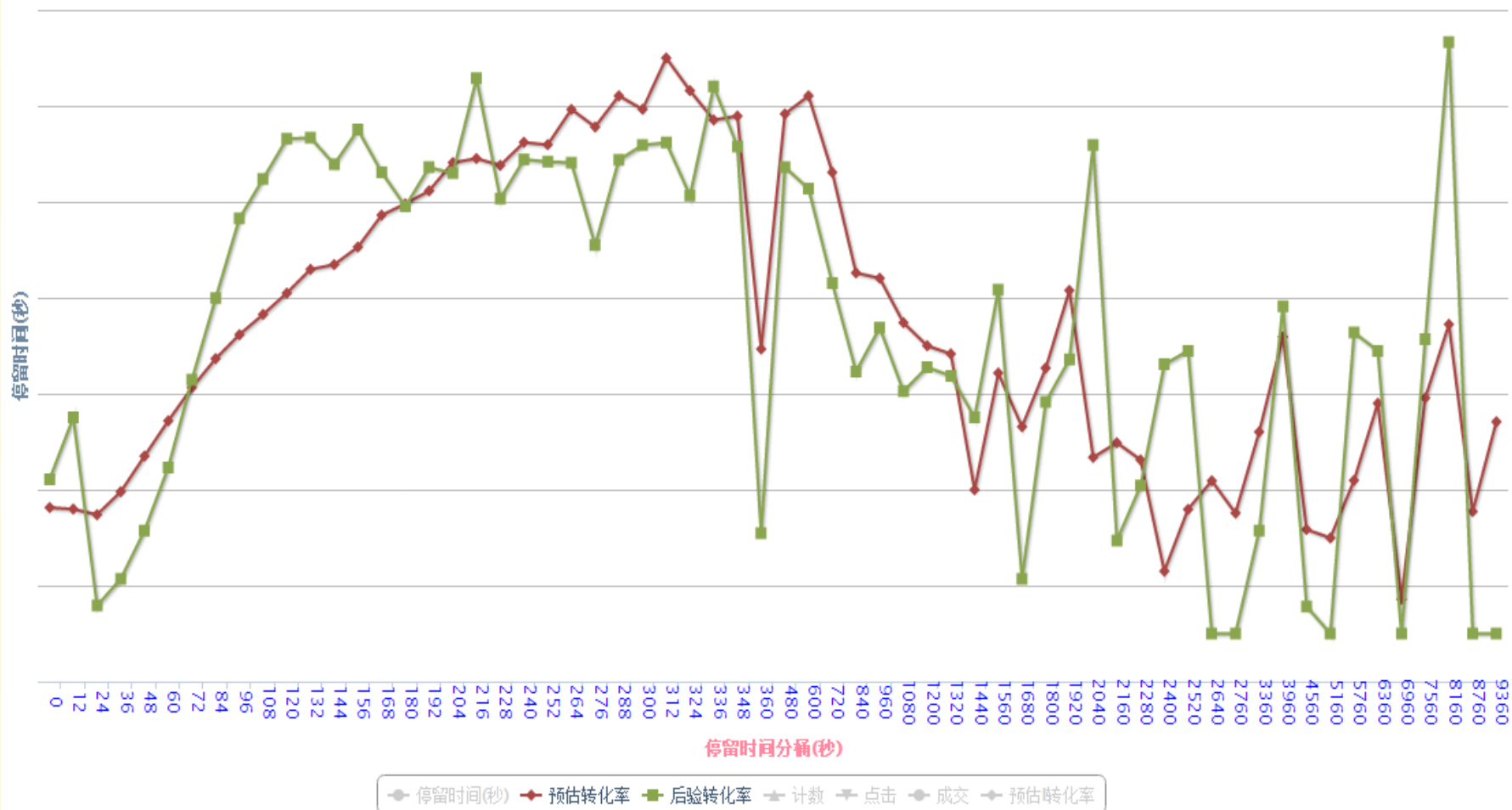


用户行为 - 特征分析

		记录条数	成交笔数	成交概率	记录条数	成交笔数	成交概率	
总		123210243	2559076	2.08%				
行为特征		发生了该行为			没有发生该行为			覆盖率
☆ 精选市场		38609416	2191964	5.68%	84600827	367112	0.43%	31.34%
👤 女装 / 内衣	0001	20431667	944885	4.62%	102778576	1614191	1.57%	16.58%
👤 男装 / 运动	0001	16647035	691363	4.15%	106563208	1867713	1.75%	13.51%
👤 女鞋 / 男鞋	0001	4970137	668781	13.46%	118240106	1890295	1.60%	4.03%
👤 化妆品 / 香水	0001	4399286	549674	12.49%	118810957	2009402	1.69%	3.57%
📱 手机数码	0001	3228897	131503	4.07%	119981346	2427573	2.02%	2.62%
👶 母婴玩具	0001	3097311	120619	3.89%	120112932	2438457	2.03%	2.51%
🍴 食品	0001	3064551	210360	6.86%	120145692	2348716	1.95%	2.49%
🏠 大家电 / 生活	0001	3008826	227453	7.56%	120201417	2331623	1.94%	2.44%
🏠 家具建材	0001	1657822	991755	59.82%	121552421	1567321	1.29%	1.35%
💎 珠宝首饰	001	1185054	95751	8.08%	122025189	2463325	2.02%	0.96%
🚗 全新整车	001	944177	34427	3.65%	122266066	2524649	2.06%	0.77%
🏠 家纺 / 家纺	001	804367	36441	4.53%	122405876	2522635	2.06%	0.65%
🏠 医药保健	001	801410	41728	5.21%	122408833	2517348	2.06%	0.65%
🏠 居家百货	0001	550832	370781	67.31%	122659411	2188295	1.78%	0.45%
📖 图书音像	0001	517334	34850	6.74%	122692909	2524226	2.06%	0.42%
	0001	32836	5268	16.04%	123177407	2553808	2.07%	0.03%
	0001	22083	3391	15.36%	123188160	2555685	2.07%	0.02%

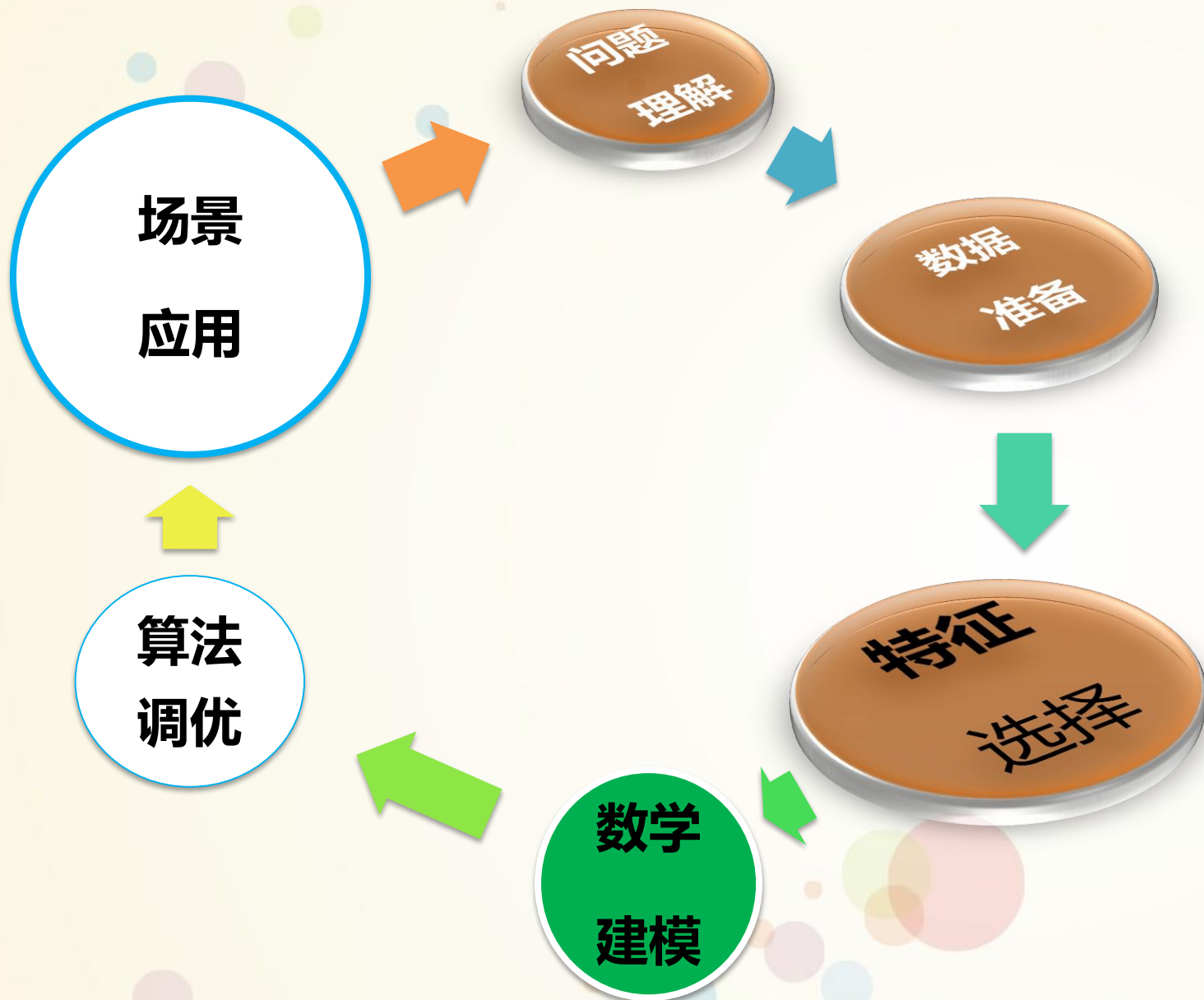
停留时间 - 特征分析

停留时间/转化率关系分析



详情页屏数 - 特征分析

浏览屏数	pv占比
2屏以上	48.02%
4屏以上	34.91%
8屏以上	20.31%
13屏以上	10.73%
19屏以上	5.23%
35屏以上	1%





- 回归函数

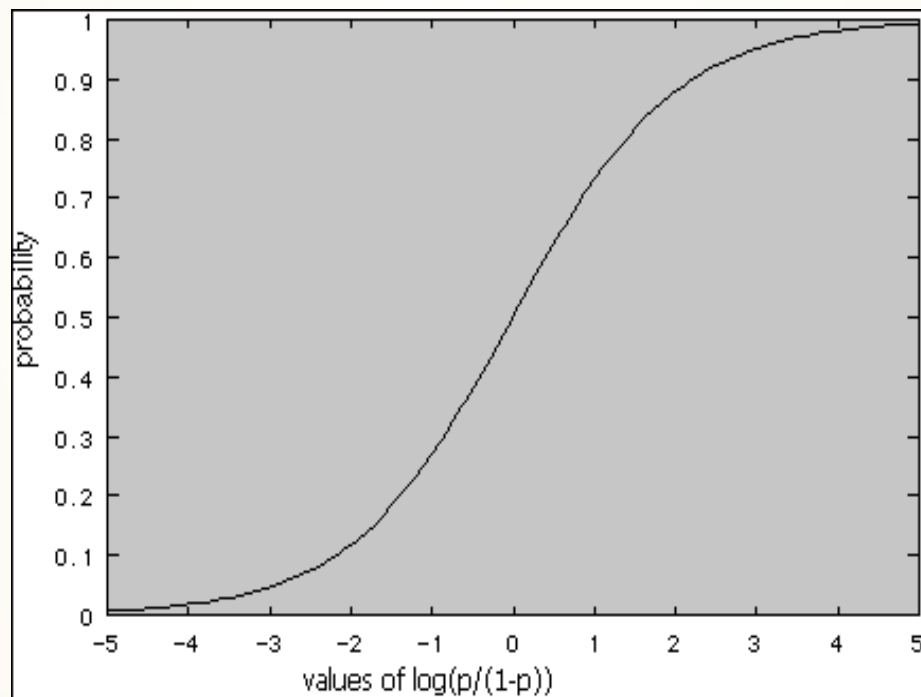
$$P(y_i = 1 \mid x_i) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i * x_i)}}$$

- 发生比(odds)

$$\frac{p_i}{1 - p_i} = e^{(\beta_0 + \sum \beta_i * x_i)}$$

- Logit形式

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \sum \beta_i * x_i$$



模型评估

- **模型评估**

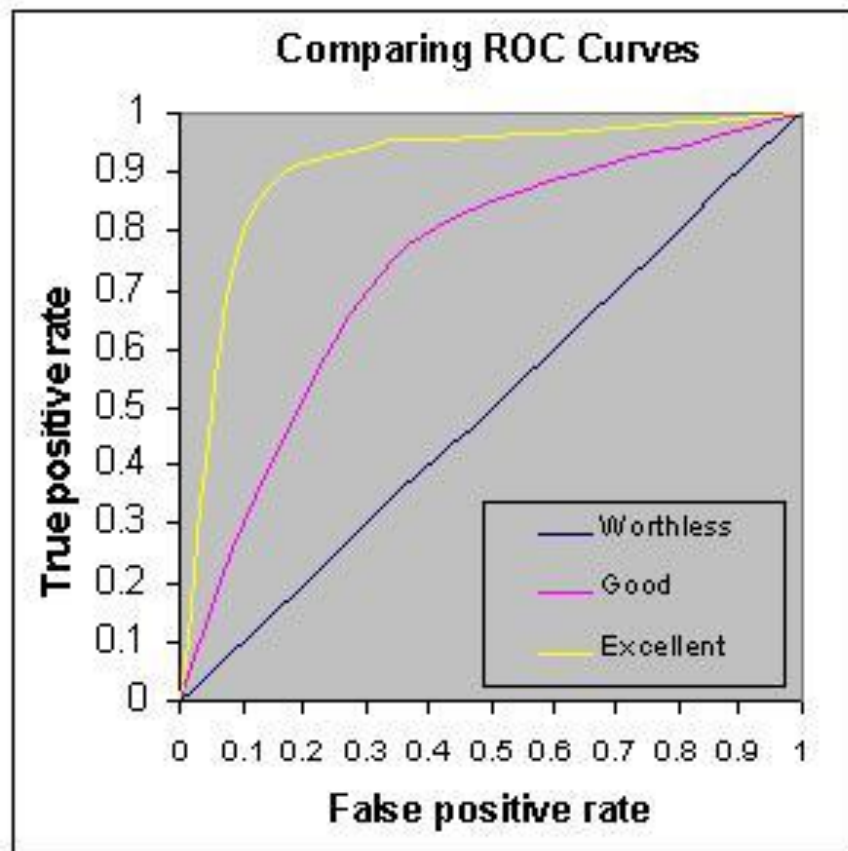
- AUC: 真实值和预估值是否一致（保序性）
- MAE/MSE：真实值和预估值Gap有多大（精确性）
- 具体case

- **ABTest**

AUC

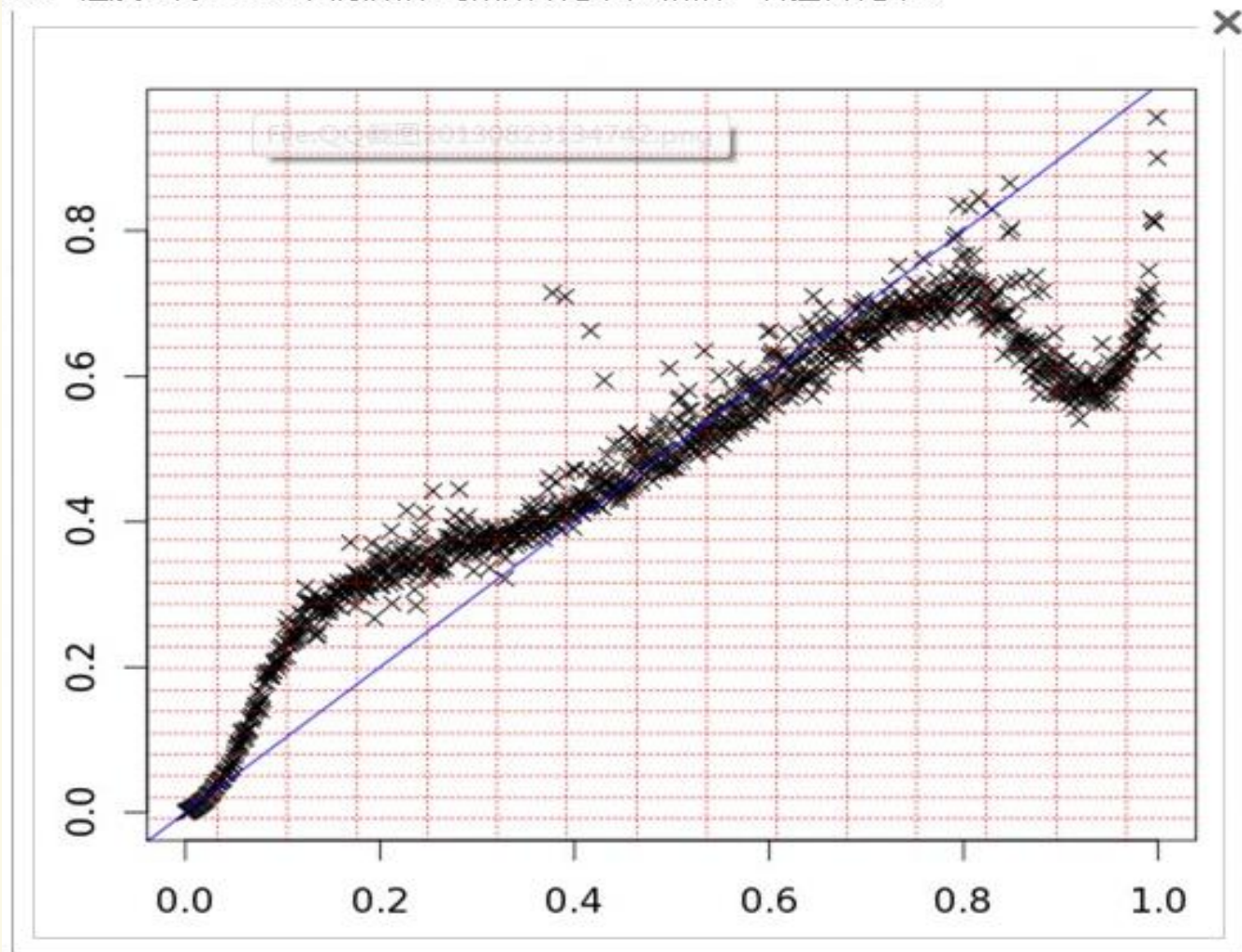
```
# trainset 07.01~07.31
# evalset 08.01~08.05
# L1 weight = 3.000000
# L2 weight = 3.000000
# http://mpi.alibaba-inc.com/log\_detail?job\_id=995
total_Clk=6108314.000000
total_nonClk=385445228.000000
total_ins=121600066
MEAN_CTR=0.015600(total_clk/total_pv)
MEAN_CTR=0.025231(1/n sum(ctr))
MEAN_PCTR=0.021880(1/n sum(pctr))
MSE=0.019240
MAE=0.032888
AUC=0.918956
WAUC=0.918956
```

```
#test auc
total_Clk=1171295.000000
total_nonClk=77879222.000000
total_ins=24732546
MEAN_CTR=0.014817(total_clk/total_pv)
MEAN_CTR=0.026209(1/n sum(ctr))
MEAN_PCTR=0.023119(1/n sum(pctr))
MSE=0.020057
MAE=0.034857
AUC=0.927818
WAUC=0.927818
```

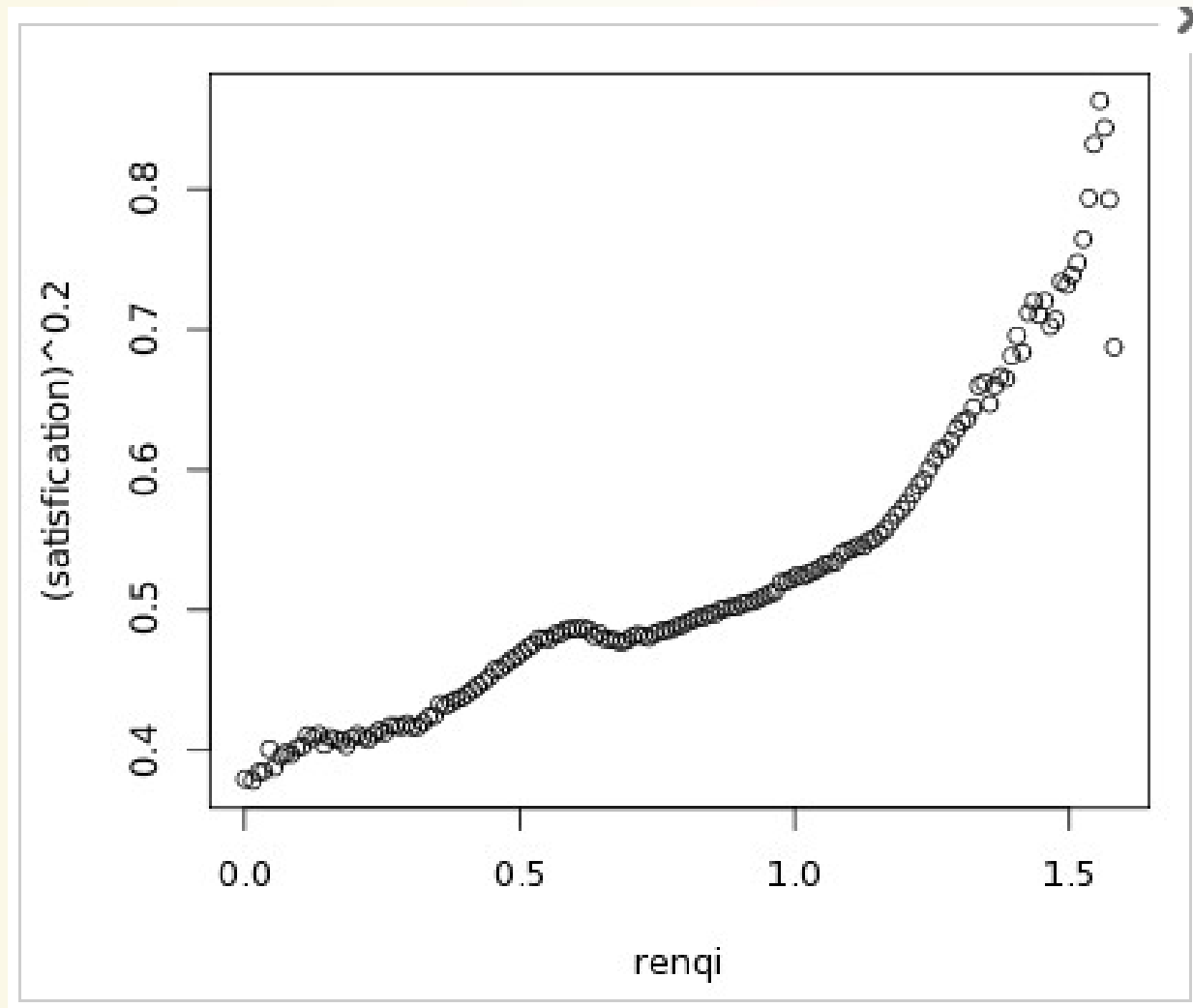


真实转化率 vs 满意度

IPV 粒度 (分1000个桶横轴: 预估转化率, 纵轴: 后验转化率)

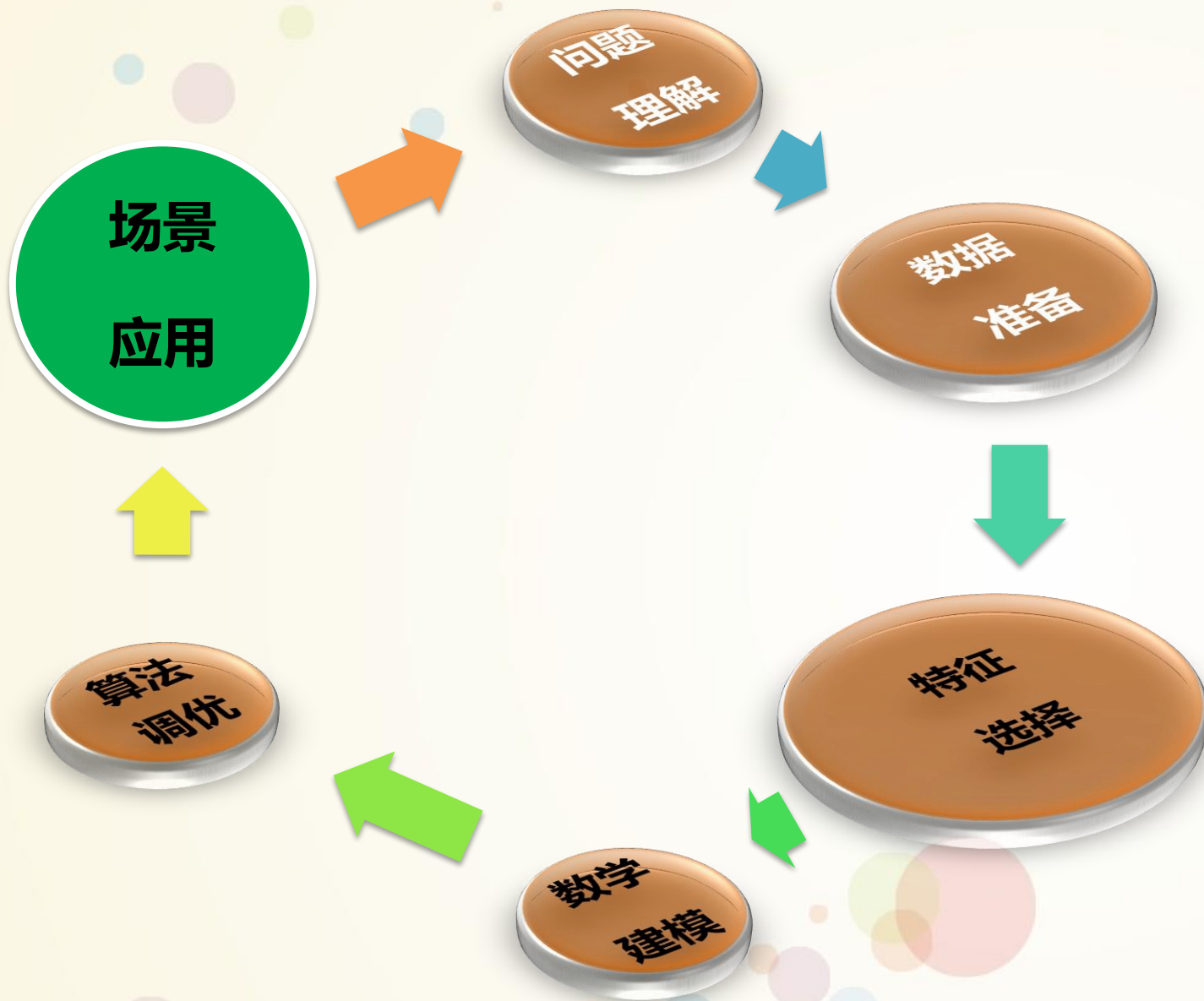


人气 vs 满意度



ABTest 效果

日期	Bucket	pV	uv	ipV	ipVuv	支付宝成交金额	支付宝成交笔数	购买UV	客单价	CTR	uv购买转化率
20140408	测试桶	1,056,747	132,150	422,103	83,710	792,491.92	6,360	4,641	170.76	39.94%	3.51%
	基准桶	1,060,230	132,260	433,218	84,716	774,392.21	6,242	4,548	170.27	40.86%	3.44%
	GAP%	↓ 0.33%	↓ 0.08%	↓ 2.57%	↓ 1.19%	↑ 2.34%	↑ 1.89%	↑ 2.04%	↑ 0.29%	↓ 2.24%	↑ 2.13%
20140407	测试桶	994,879	122,616	401,796	77,938	794,324.06	6,307	4,611	172.27	40.39%	3.76%
	基准桶	1,004,824	122,843	411,188	78,683	740,934.26	5,846	4,357	170.06	40.92%	3.55%
	GAP%	↓ 0.99%	↓ 0.18%	↓ 2.28%	↓ 0.95%	↑ 7.21%	↑ 7.89%	↑ 5.83%	↑ 1.30%	↓ 1.31%	↑ 6.03%
20140406	测试桶	986,219	118,680	393,040	75,665	730,139.90	5,832	4,276	170.75	39.85%	3.60%
	基准桶	996,124	119,229	408,194	76,137	723,875.77	5,420	4,008	180.59	40.98%	3.36%
	GAP%	↓ 0.99%	↓ 0.46%	↓ 3.71%	↓ 0.62%	↑ 0.87%	↑ 7.60%	↑ 6.69%	↓ 5.44%	↓ 2.75%	↑ 7.17%
20140405	测试桶	956,181	115,349	379,712	72,663	684,691.17	5,396	4,003	171.04	39.71%	3.47%
	基准桶	946,418	115,334	383,319	72,915	681,933.74	5,158	3,796	179.62	40.50%	3.29%
	GAP%	↑ 1.03%	↑ 0.01%	↓ 0.94%	↓ 0.35%	↑ 0.40%	↑ 4.61%	↑ 5.45%	↓ 4.78%	↓ 1.95%	↑ 5.43%





- 搜索排序
- 推荐选品
- 信息披露
- 详情页优化

一些建议

- **数据(50%)、特征(20%)、应用(20%)、优化模型(10%)**
- **要什么数据，做什么，要先想清楚**
- **数据正确性校验**
 - 抽样，统计单维度数据是否正确
- **基础数据**尽量保持原始数据，各种变形，组合，离散化等尽量放到最后
- **样本抽样**：因为正负样本差异比较大，丢掉部分负样本也许对模型有利
- **人工调权**：模型训练出来后的feature weight, 可以根据业务和经验需要人工设置权重。

Q&A

THANKS

