

# 客户流失预警模型分析报告

项目时间：2015-10-06

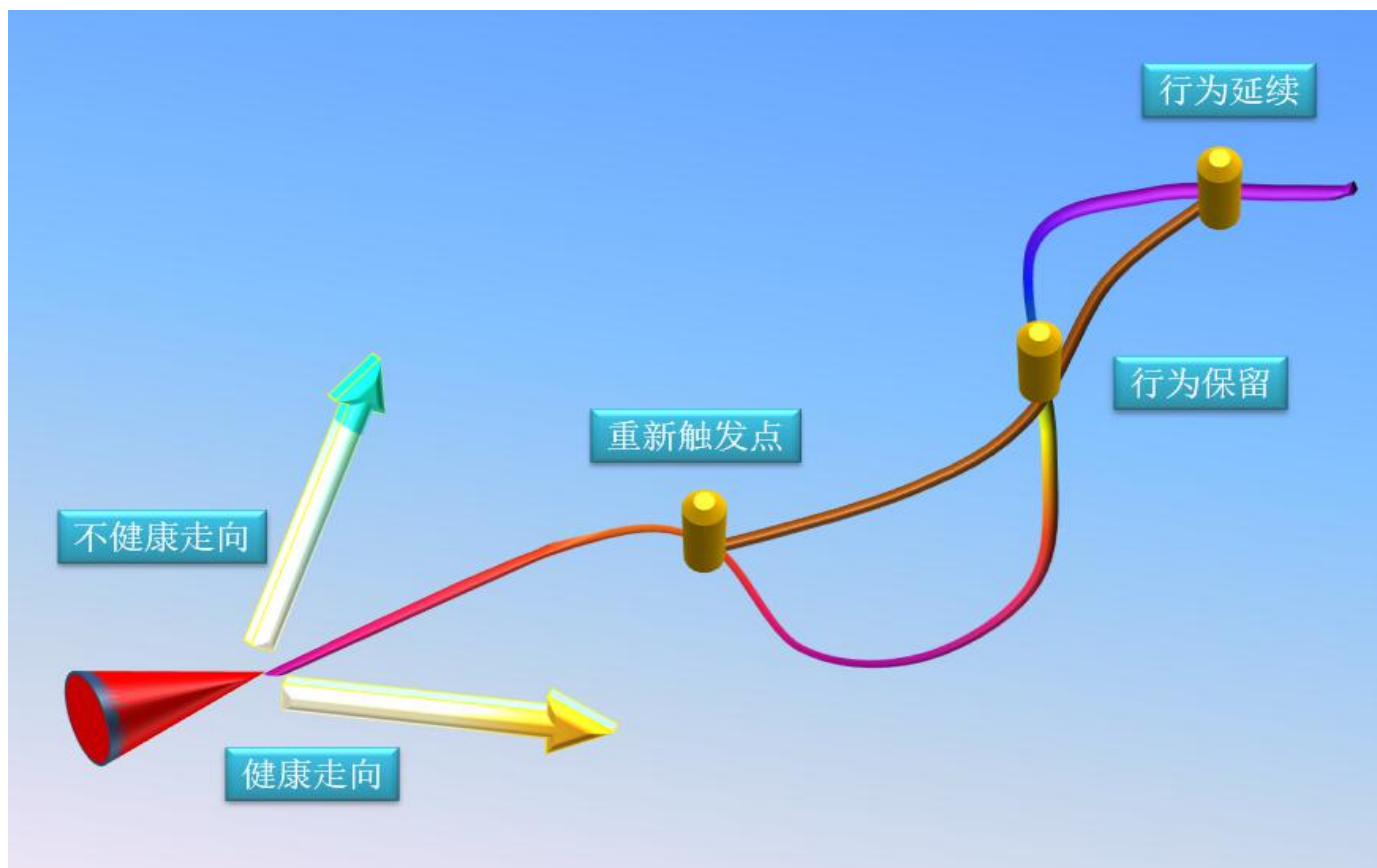
# 主要内容

---

- 背景介绍
- 分析目标
- 模型定义
- 数据分析
- 模型构建
- 模型评估

# 背景

2014年底，贵公司业务重组启动，客户的移动市场转化将带来全新的业务模式和客户行为模式，尤其是中高端的价值客户，面临消费黏性和消费惯性的不稳定，如何确保其在新的消费模式下，维持价值稳定性是贵公司当前的主要战略目标之一。

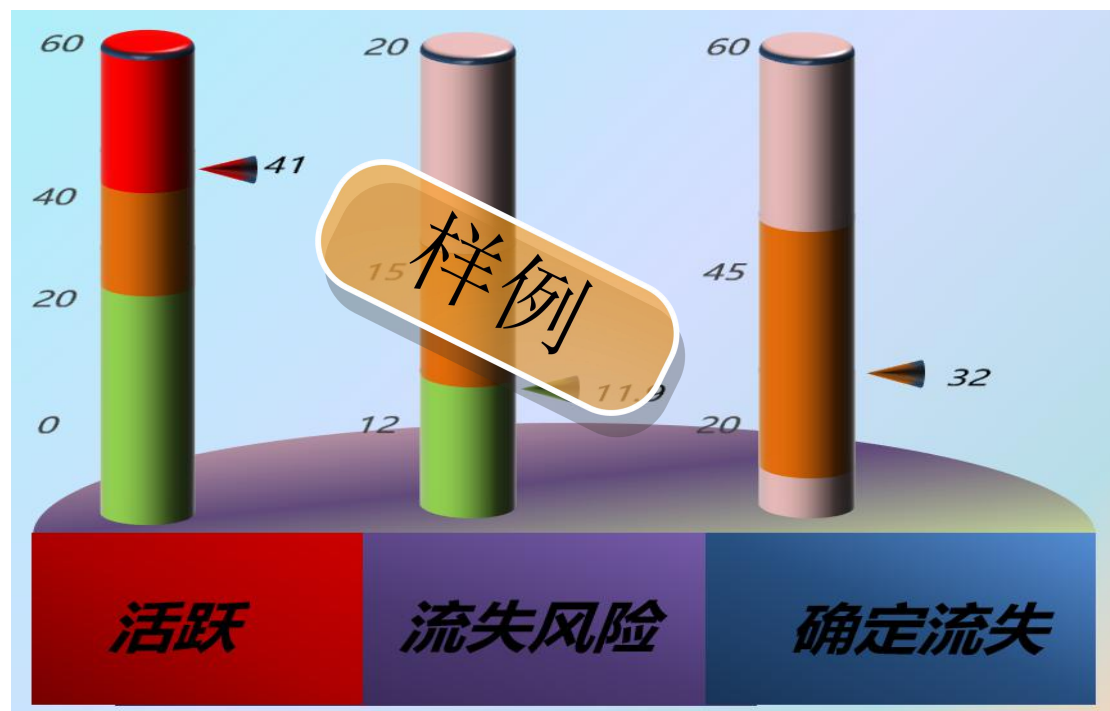


# 目标用户行为链

目标用户：高价值、新客户、忠诚度

现状：活跃客户不足（41%）、确定流失客户高达32%

针对客户整个生命周期的行为，建立客户的价值流失预警模型，预测贵公司客户流失概率以及中高价值客户的价值下降问题，使中高端客户保有更有针对性，更有效地支撑客户挽留活动和预警效果。



# 模型定义

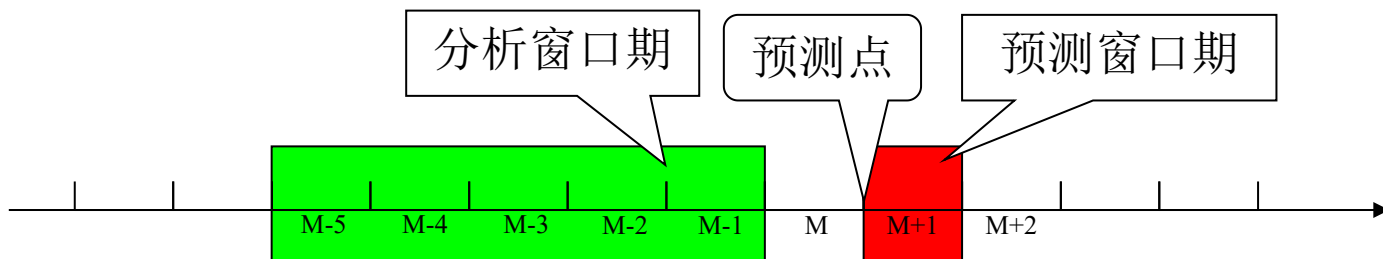
## 客户群分析：

- 贵公司下发的客户卡：02月份在线注册的C类卡8万张，网上销售的同类卡片有20万张，但根据注  
销的2万张C卡来看，同比和环比符合公司领导层的预期。
- 中高端客户：最近六个月账单总金额不小于200元的B卡类用户

## 模型目标：

- 预测哪些客户的在预测窗口会成为高流失的客户，将此类用户定义为重点处理对象。

## 时间窗口：



## 构建模型的时间窗口：

- 分析窗口：M表示30天，截至M5表示分析期到达150天。
- 预测窗口：当前分析日期的30天。

# 数据分析

- 目标:

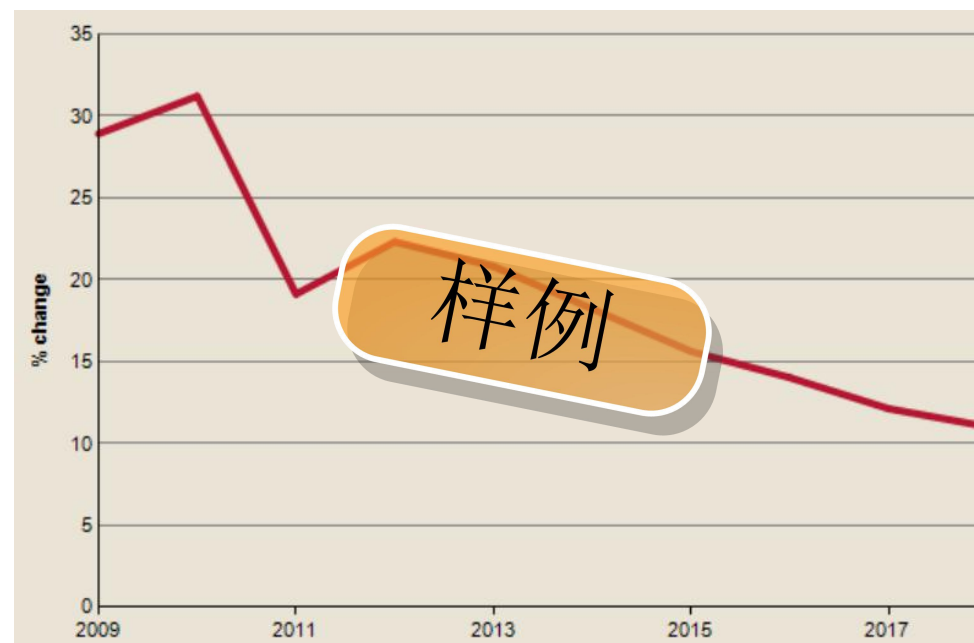
- 通过对数据质量的检查、各个变量关系的探索, 选取进入逐步逻辑回归模型的变量
- 右图显示, 2011年的错误战略直接导致连续6年的客户增整下滑。

- 对象:

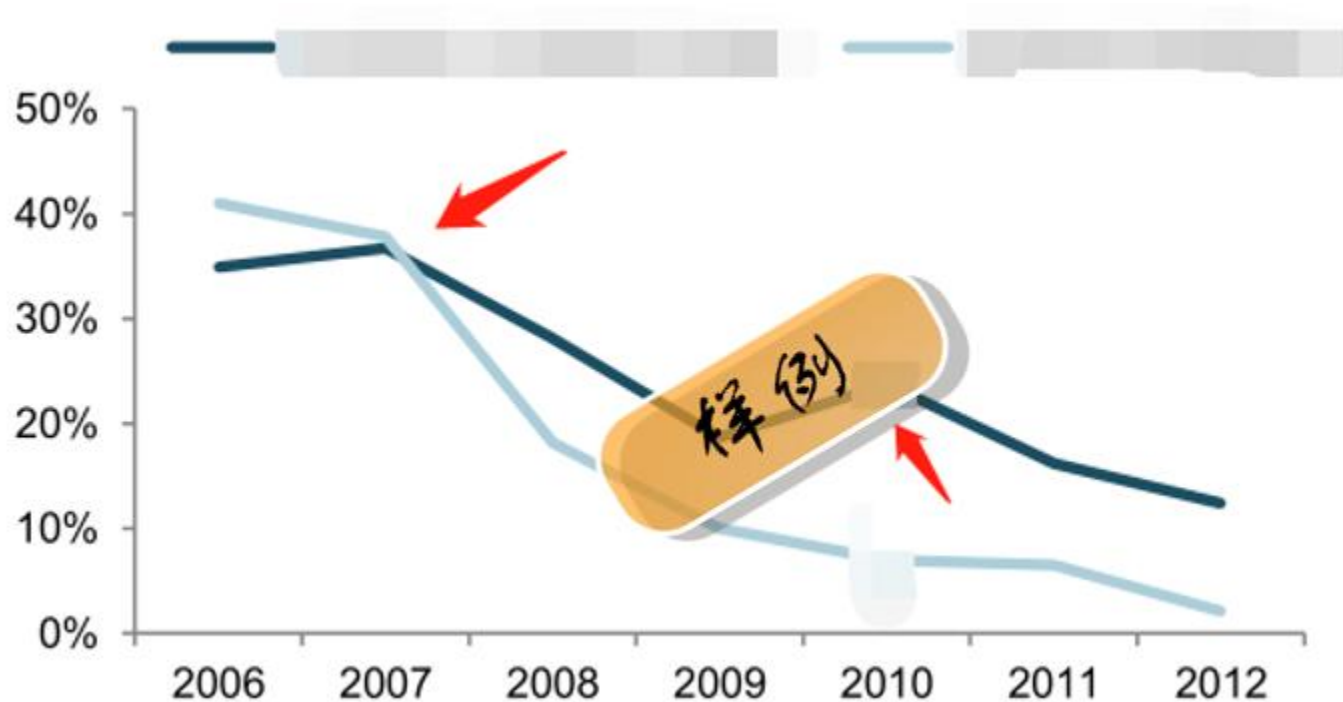
- 用户: \*\*\*\*\*用户特征
- 数据: 第三方、集团内数据中心（长途）、网络通话, 视频数据等

- 方法:

- 数据质量监控:
- 相关分析: 探索流失客户的触发的点。
- 客户业务流程分析: 典型的业务链条。
- 客户矩阵分析: 行为属性关联

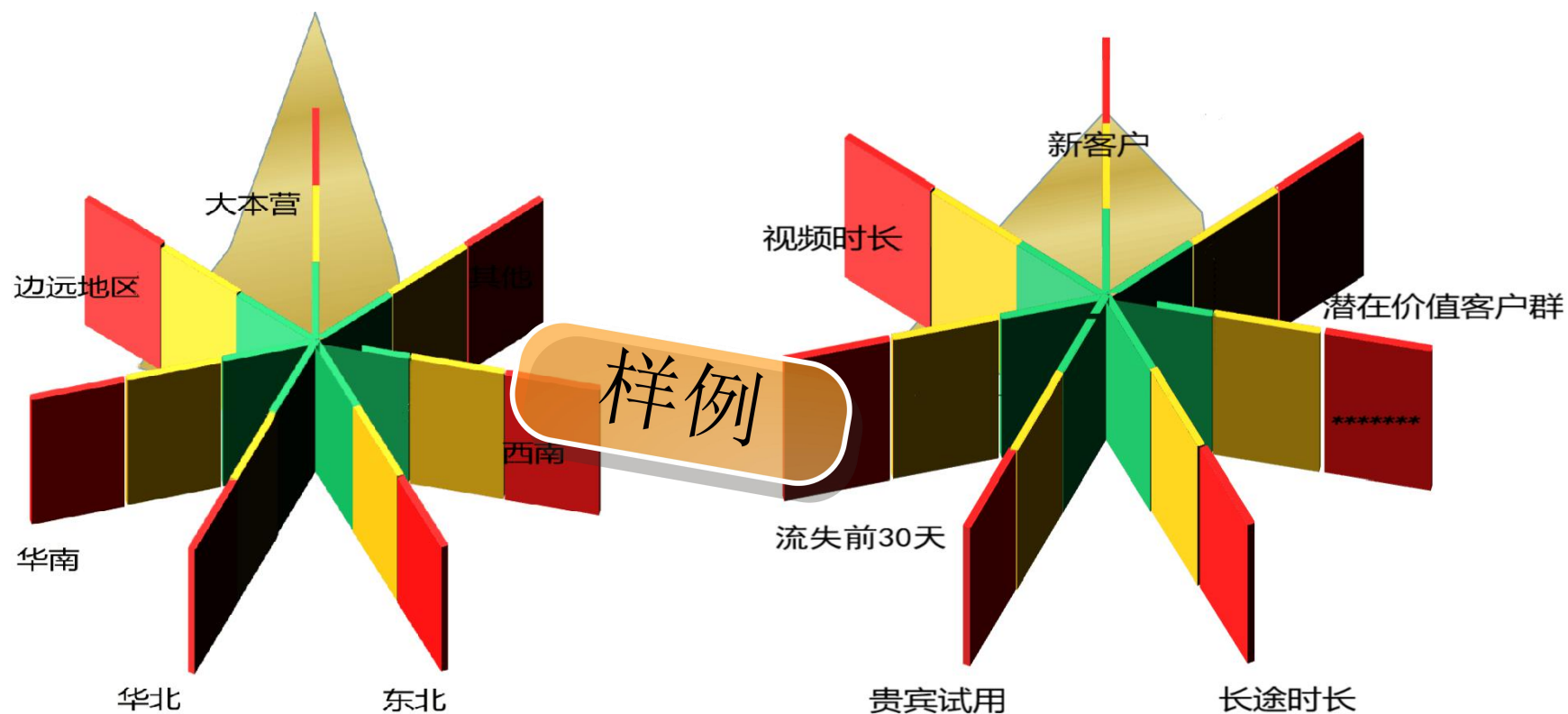


# 数据分析：第三方数据



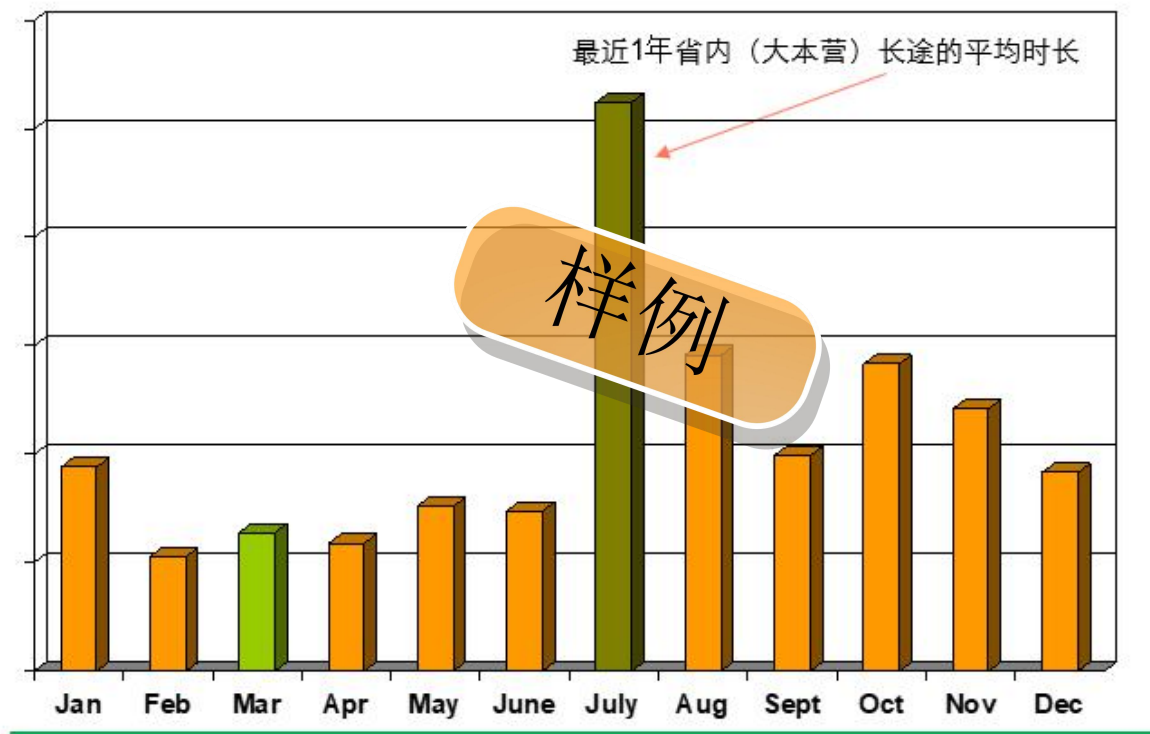
上图：第三方数据只能提供到2012年，贵公司主营业务和价值用户群的流失情况，其中07年的两类客户群的在主营业务上的使用大体相当，但随后流失差异越来越大。

# 数据分析：主营客户群



大本营的业务占比预期60%，但实际上去年同期已经下滑到52.6%，今年同期数据51%也无法达标，因此重点处理的客户群或在今年下半年希望通过“潜在流失客户群”和“新客户”两类客户执行新的营销和数据分析，希望短期能维持主营客户群50%以上的占比。

# 数据分析：长途电话



最近1年的国内长途数据变换情况：

- 数据显示与业务预期基本相似，7月出现长途时长大幅增加，主要原因是大本营的数据贡献（72%贡献率）。
- 从3月份开始实施15项营销方法以来，7月份尤为突出，8月至年底，收效甚微。

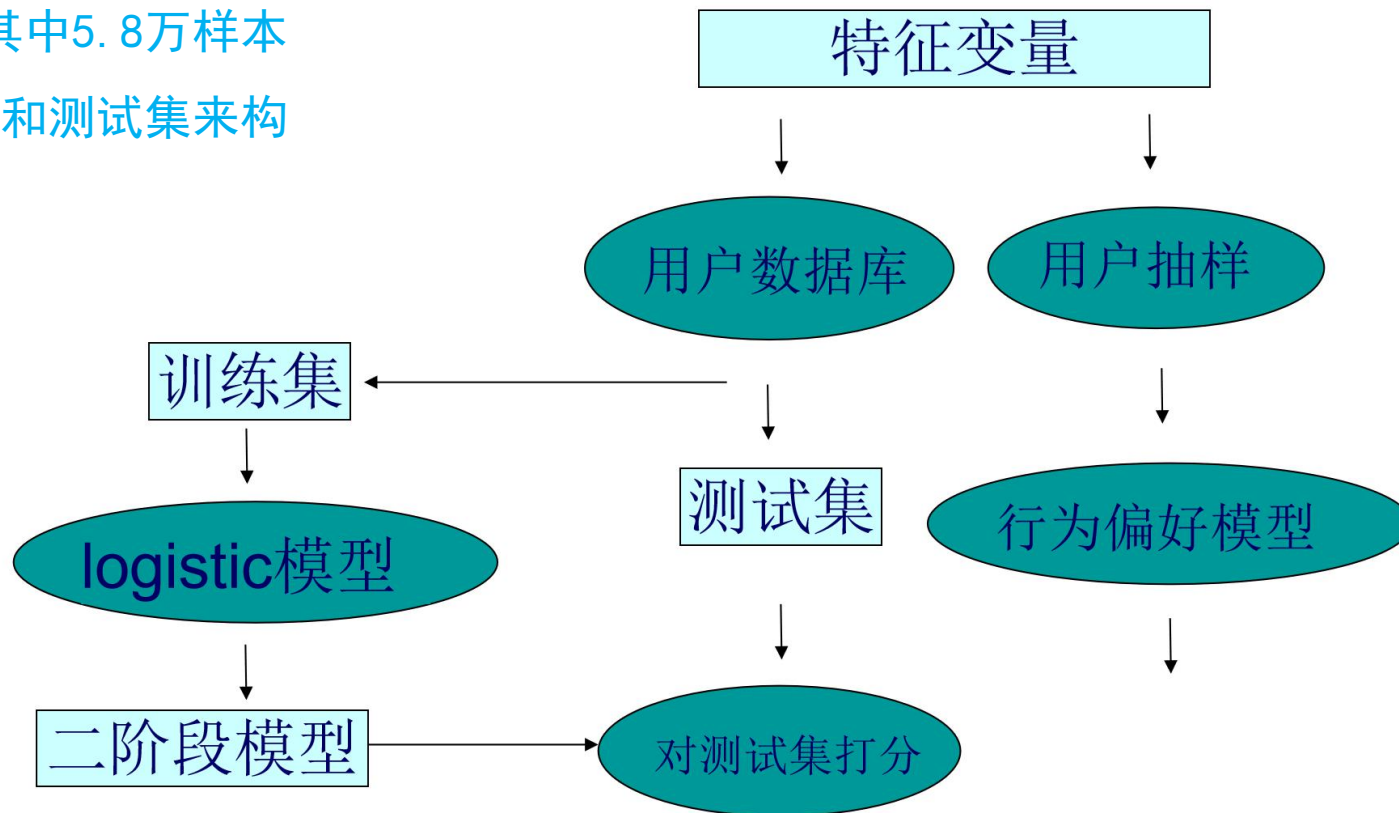
# 数据分析：特征筛选

- 在网时长在业务是很重要的指标。
- 数据描述发现，账户余额占比过高，流失率很低。
- 最近30天和90天主叫和被叫时长可以有效预测存留率。
- 上午和下午对应商务办公用话。
- .....

|                          |
|--------------------------|
| 常数                       |
| 在网时长                     |
| 账户余额                     |
| 被叫网内对端号码数                |
| 最近三个月主叫电信固网号码数的平均值       |
| 最近三个月晚上本地主叫计费时长          |
| 被叫电信固网号码数                |
| 本月国内漫游网内被叫次数与最近三个月平均值的比值 |
| 本月ARPU是否小与等于80           |
| 本地下午主叫次数                 |
| 本地早上被叫次数                 |

# 模型构建：建模流程

使用6万用户（其中1200个样本为抽样样本+数据汇总二次数据），对其中5.8万样本按照6：4的比例分成训练集和测试集来构建和评估模型。



# 模型构建：建模算法

---

首先，构建以大本营客户为载体的营销方案，将潜在的流失客户群转向手机端的分类预测，并通过概率预测，对其施以手机端下载的优惠营销方案。

使用logistic回归模型完成建模工作。

其次，由于数据缺失（或第三方数据完整性考虑），无法使用所有数据，因此忠诚客户单独构建模型（生存分析），并使用生存分析模型的忠诚度评分和logistic回归的流失率评分进行加权，并最终获得所有客户的流失率评分构建客户生命周期。

最后，客户行为分析作为客户生命周期辅助解释模型。

# 模型评估

---

使用ROC、命中率（%）、提升率（Lift）、覆盖率（%）来对模型进行评估：

- ROC=敏感度和特异性的组合。
- 命中率=提取的客户中流失的数目/提取的人数。
- 提升率=命中率/总体的流失率。
- 覆盖率=提取的客户中流失的数目/总体的流失的人数。

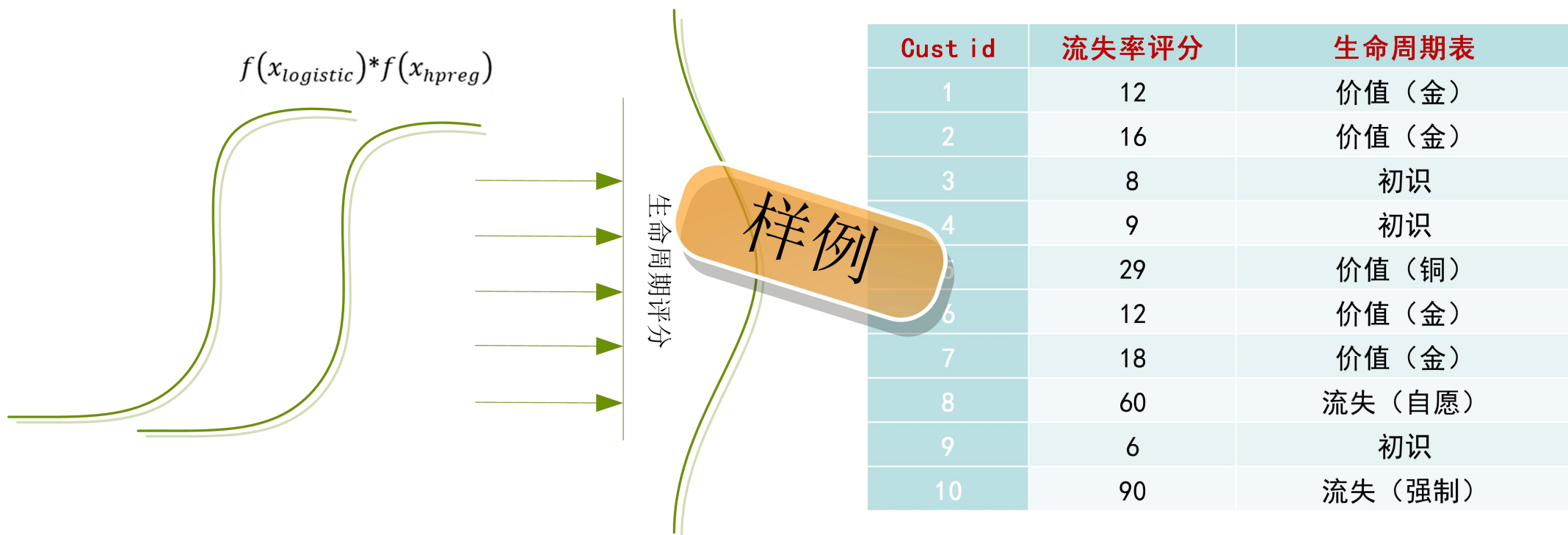
# 模型结果分析

| 特征                       | 业务权重 | 统计权重 | 综合权重 |
|--------------------------|------|------|------|
| 在网时长                     | ***  | ***  | 0.01 |
| 账户余额                     | ***  | ***  | 0.16 |
| 被叫网内对端号码数                | ***  | ***  | 0.08 |
| 最近30天主叫电信固网号码数的平均值       | ***  | ***  | 0.09 |
| 最近30天晚上本地主叫计费时长          | ***  | ***  | 0.19 |
| 被叫电信固网号码数                | ***  | ***  | 0.12 |
| 本月国内漫游国内被叫次数与最近30天平均值的比值 | ***  | ***  | 0.18 |
| 本月GRF是否小于等于38            | ***  | ***  | 0.01 |
| 本地下午主叫次数                 | ***  | ***  | 0.06 |
| 本地早上被叫次数                 | ***  | ***  | 0.1  |

样例

最终我们发现，主要的几个特征变量：**账户余额、最近30天晚上本地主叫计费时长、被叫电信固网号码数、本月国内漫游国内被叫次数与最近30天平均值的比值**为主要影响变量。

# 模型结果分析



- 模型经过加权整合，并输出综合评分，以作为生命周期评分。
- 每个客户都被赋予一个流失率评分，然后根据流失率评分的分布及其他统计特征
- 进行分箱化处理，并制作业务标签，这就是生命周期表

# 模型部署

---

1. 新客户例行评分，赋值和营销准备，主要针对高价值客户的毛利价值。
2. 客户流失率高的客户，定期维护，以修改和调整生命周期表。
3. 生命周期表为中心，构建客户关系管理，将客户画像、产品关联、营销匹配等商业支持进行整合，以获取更全面的客户描述。
4. 数据分析业务标签和评分的常态化。
5. 数据更新和模型维护，定期对模型进行评估，当模型历经3个月“老了”，必须重新建模。