

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：大数据分析挖掘

■ 新浪微博：ChinaHadoop



1

银行业客户流失预警模型的介绍

目录

银行业客群和产品的类别

客户流失预警模型的业务意义

数据介绍和描述

银行业客群和产品的类别

□ 银行的个人客户

银行对个人客户的业务主要是以合理安排客户的个人财物为手段，为之提供存取款、小额贷款、代理投资理财、信息咨询及其他各类中介服务，由此为客户取得收益并帮助其防范风险，同时提高银行自身效益。

□ 银行的公司客户

公司客户主要指与银行发生业务关系的各企事业单位及政府机关，其中以企业单位为主体。
。公司客户能为银行带来大量存款、贷款和收费业务，并成为银行利润的重要来源。
。

银行业客群和产品的类别

□ 零售客户主要分以下五种：

一般客户：其接受的金融服务以负债、中间业务为主，包括存折、存单、账户、借记卡业务、像普通居民提供小额存款、取款、转账、缴费等传统服务。

消费信贷客户：业务以资产业务为主，只在购买物品、服务时申请私人用途的贷款，主要指住房信贷、汽车信贷、耐用消费品信贷等。

信用卡客户：其接受的金融服务不含借记卡业务，一般是面向社会各阶层人士提供的小额短期信贷、支付、转账业务，集资产、负债、中间业务于一身。

贵宾理财客户：其接受的金融服务属于代客理财范畴，一般指向收入高或者财富多的中产阶级以上者提供资产保值、增值服务，如股票、债券、结构性产品等。

私人银行客户：其接受的金融服务属于资产管理业务，一般是面向新兴的富豪或家族性富豪提供的专属、私密、量身定做的金融服务，包括海外基金、信托基金、私募基金等，目的是帮助其财富增值、避税等。

银行业客群和产品的类别

□ 银行信贷类资产

- 信用贷款
- 抵押贷款
- 保证书担保贷款
- 贷款证券化

银行负债业务

- 活期存款
- 定期存款
- 储蓄存款
- 可转让定期存单
- 其他种类

目录

银行业客群和产品的类别

客户流失预警模型的
业务意义

数据介绍和描述

客户流失预警模型的业务意义

- 严格地讲，客户流失指的是客户在该行所有业务终止，并销号。但是具体业务部门可单独定义在该部门的全部或某些业务上，客户的终止行为。
- 对专家及金融业业内人士的走访及调研结果表明，商业银行客户流失较为严重。国内商业银行，客户流失率可达20%甚至更高。而获得新客户的成本，可达维护现有客户的5倍。
- 因此，从海量客户交易记录中挖掘出对流失有影响的信息，建立高效的客户流失预警体系尤为重要。

客户流失预警模型的业务意义

客户流失主要原因包括：

价格流失

产品流失

服务流失

市场流失

促销流失

技术流失

政治流失

一个表面上看来是由于价格因素流失的客户，实际上可能存在很多种流失原因

- 不现实利润目标、价格结构的不合理
- 过度复杂的业务流程、组织结构的不合理、
- 决策工具的应用失误、信息系统的缺陷
- 银行不恰当的雇佣原则、培训的缺乏、落后的技术
- 员工职业生涯发展不利和对员工的激励不足

客户流失预警模型的业务意义

□ 维护客户关系的基本方法

- 追踪制度：
- 产品跟进
- 扩大销售
- 维护访问
- 机制维护

客户流失预警模型的业务意义

□ 建立量化模型，合理预测客群的潜在流失风险

- 常用的风险因子
- 客户持有的产品数量、种类
- 客户的年龄、性别
- 受地理区域的影响
- 受产品类别的影响
- 交易的间隔时间
- 营销、促销手段
- 银行的服务方式和态度

目录

银行业客群和产品的类别

客户流失预警模型的业务意义

数据介绍和描述

数据介绍和描述

□ 数据介绍

本案例搜集了17,241例数据，其中有1,741例流失样本，总流失率达到10.10%

银行自有字段

- 账户类信息
- 个人类信息
- 存款类信息
- 消费、交易类信息
- 理财、基金类信息
- 柜台服务、网银类信息

外部三方数据

- 外呼客服数据
- 资产类数据
- 其他消费类数据

数据介绍和描述

□ 单因子分析之连续变量

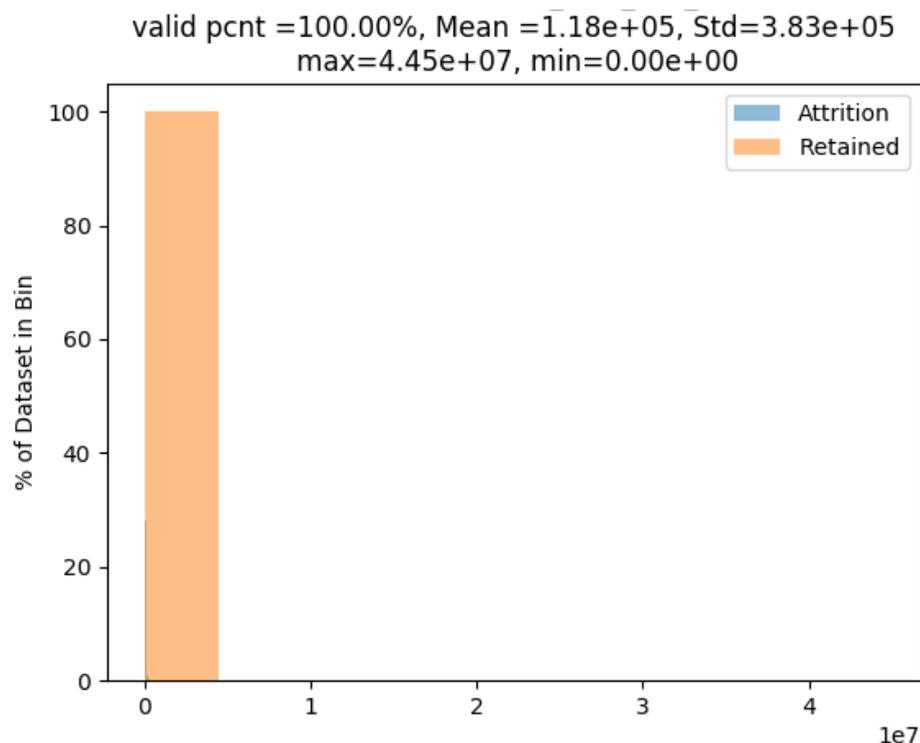
- 有效记录的占比
- 整体分布
 - 初始分布
 - 截断分布
- 按目标变量分布的差异

数据介绍和描述

□ 单因子分析之连续变量(续)

“资产当前总余额”初始分布

几乎看不出区别

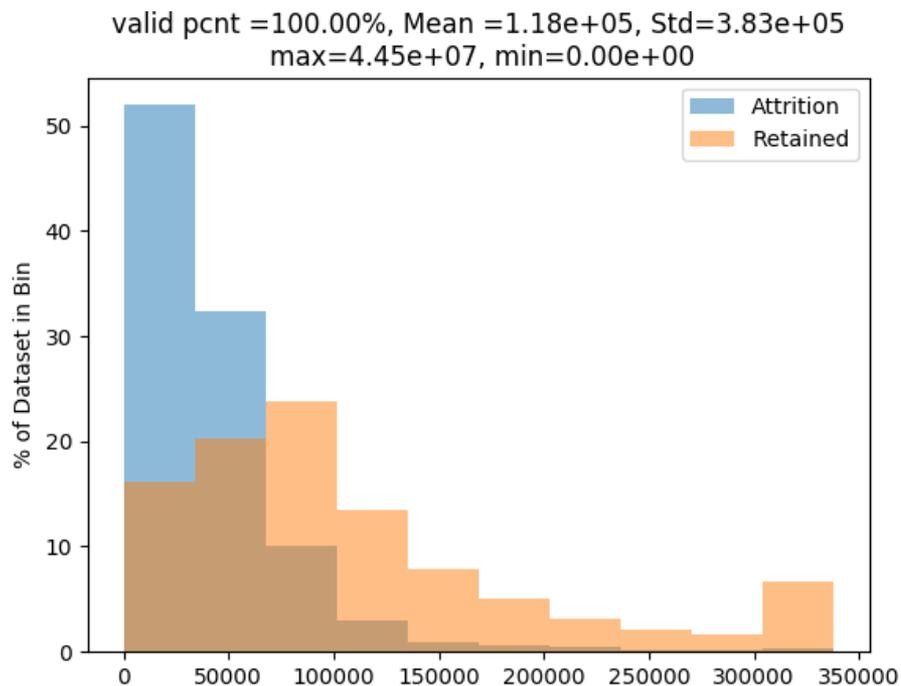


数据介绍和描述

□ 单因子分析之连续变量(续)

“资产当前总余额”截断分布：0.95分位点以内

差异很明显：高资产客户流失率显著低于低资产客户



数据介绍和描述

□ 单因子分析之连续变量(续)

差异的量化

单因子量化分析，用于两个及两个以上样本均数差异的显著性检验

基本原理

$$\text{总差异} = \text{组间差异} + \text{组内差异}$$

```
from statsmodels.formula.api import ols
```

```
from statsmodels.stats.anova import anova_lm
```

```
anova_results =
```

```
anova_lm(ols('ASSET_MON_AVG_BAL~CHURN_CUST_IND',AllData).fit())
```

结果：p值小于0.05，统计差异显著

```
>>> anova_results
```

```
...
```

	df	sum_sq	mean_sq	F	PR(>F)
CHURN_CUST_IND	1.0	1.013756e+13	1.013756e+13	70.891562	4.066894e-17
Residual	17239.0	2.465192e+15	1.430009e+11	NaN	NaN

数据介绍和描述

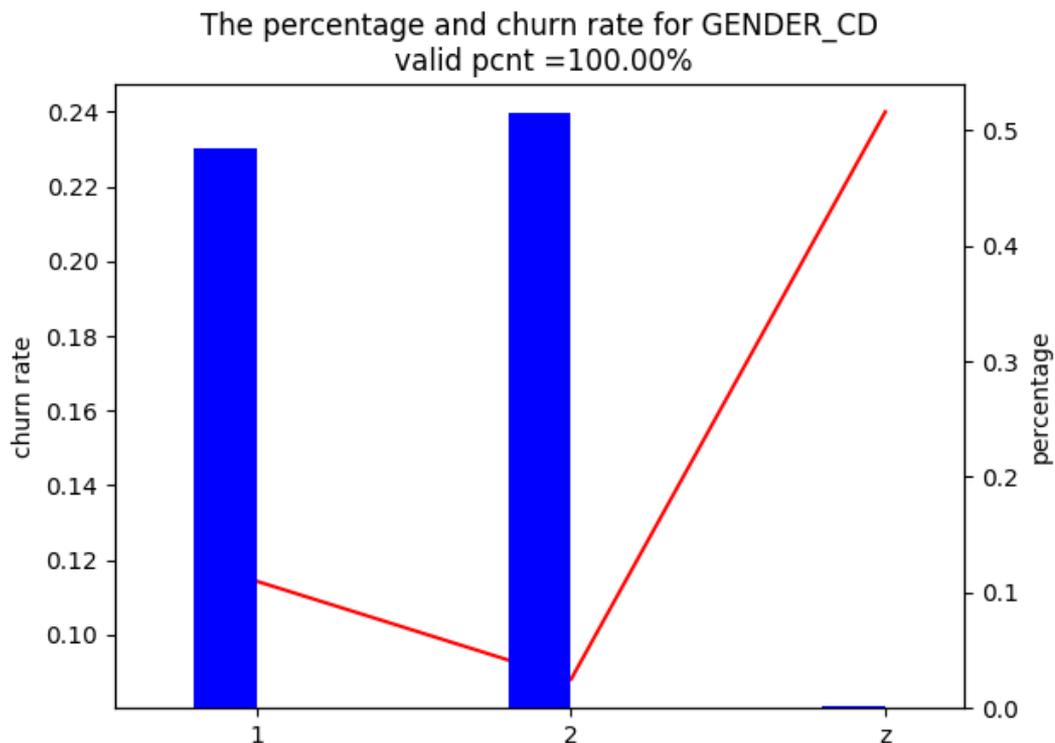
□ 单因子分析之类别变量

- 有效记录的占比
- 种类
- 整体分布
- 按目标变量分布的差异

数据介绍和描述

□ 单因子分析之类别变量(续)

- 绝大多数性别已知
- 已知性别中，男性占比较高
- 男性流失率低于女性
- 未知性别的客户的流失率显著高于性别已知的客户



数据介绍和描述

□ 单因子分析之类别变量(续)

差异的量化

卡方检验：用来衡量某因子对目标变量的偏好

原理

如果某因子跟目标变量独立，则目标变量在该因子上的分布是均匀的。卡方检验用来衡量分布的差异性

性别	总数	流失数	期望流失数
男	8875	781	896
女	8341	954	842
未知	25	6	3

$$\chi^2 = \frac{(896 - 781)^2}{896} + \frac{(842 - 965)^2}{842} + \frac{(3 - 6)^2}{3} = 32.66$$

卡方值大于自由度2、置信度0.05下的卡方分位点5.99，因此GENDER显著

疑问

- 问题答疑：<http://www.xxwenda.com/>
 - 可邀请老师或者其他人回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop

