

# 2021中国大学生保险 数字挑战赛数字赛道

团队名：scu\_team

成员：谢东霖、陆琪

学校：四川大学

1. 题目理解
2. 数据探索
3. 特征构造
4. 模型框架
5. 业务应用

# 1.赛题背景

## 背景

2020年，疫情和车险综合改革下，车险保费增速下降，单非车险保费保持较高增速；且60%的非车险是在车险销售环境下购买。  
而车险作为刚需产品，因此为非车险等多种保险产品销售提供了有效通道。

## 需求

**初赛：**预测购买非车险的意向（分类问题）  
**决赛：**预测购买非车险得保费（回归问题）

## 业务应用

结合挖掘建模得到的可解释性信息进行业务应用

# 1.赛题数据

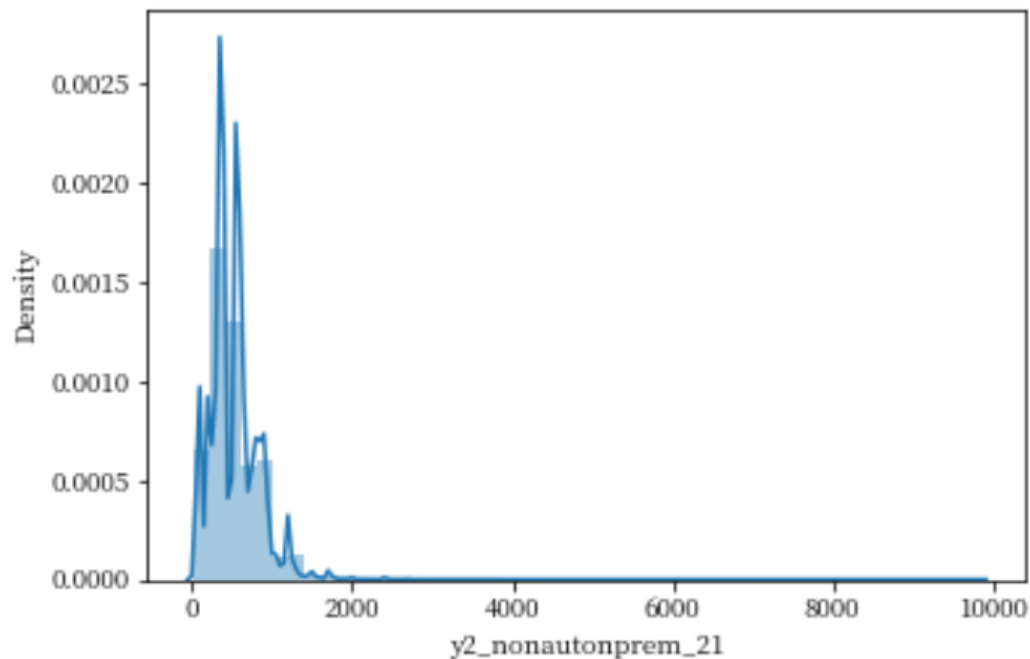
用户基本信息

字段名	含义
p1_gender	性别
p1_age	年龄
p1_census_register	籍贯
birth_month	出生月份
p2_marital_status	婚姻状况
f1_child_flag	是否有子女
p2_is_child_under_15_family	家庭是否有 15 岁以下小孩
p2_is_adult_over_55_family	家庭是否有 55 岁以上老人
p1_is_bank_eff	是否银行有效客户
p2_is_enterprise_owner	是否为企业主
p2_is_smeowner	是否为小微企业主
f2_posses_house_flag	是否有房
f2_cust_housing_price_total	房产总价
newvalue	新车购置价
.....	

保险基本信息

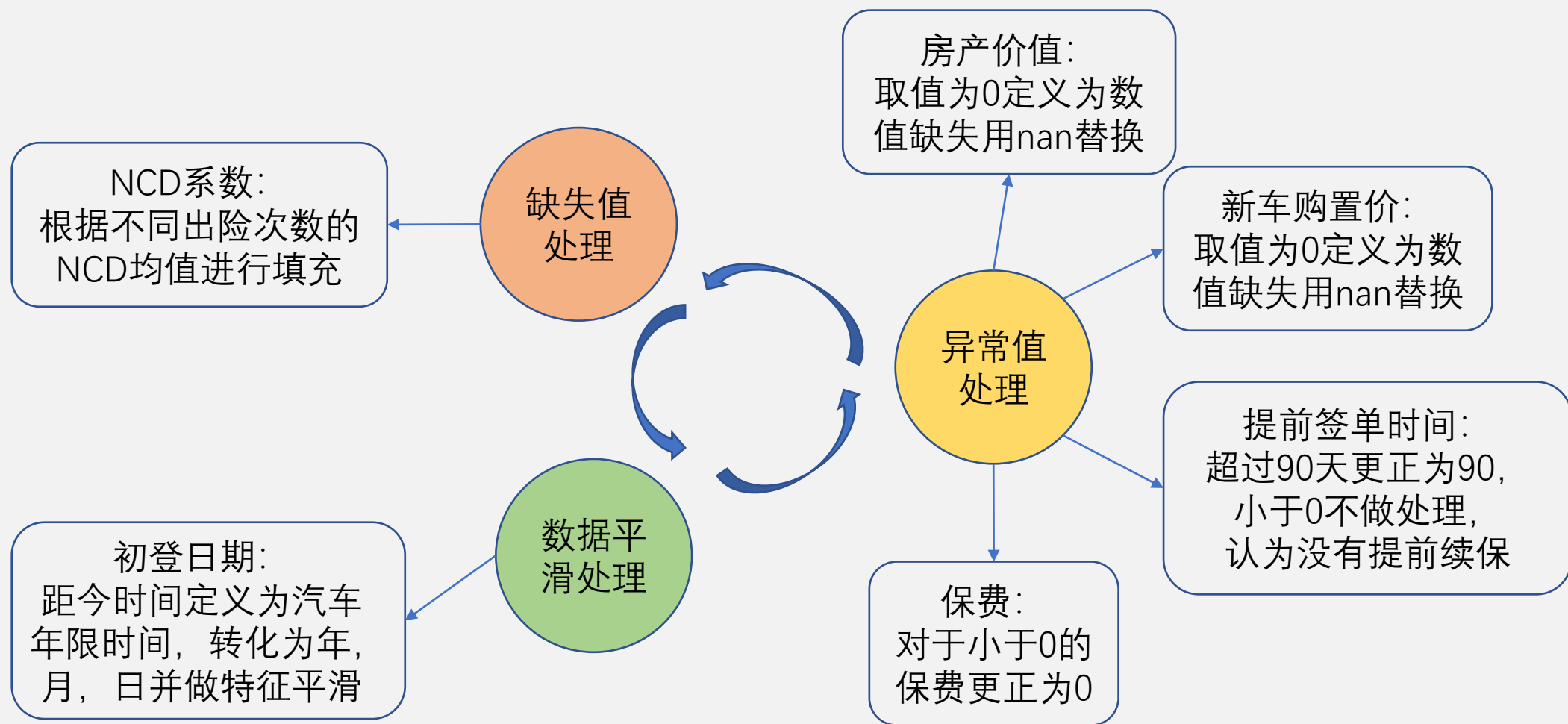
字段名	含义
dpt	保险机构
xz/xb	险种/险别
nprem_ly	上年的车险保费
ncd_ly	上年ncd系数(无赔偿优待)
nprem_od/si_od	车损险保费/保额
nprem_bt/si_bt	盗抢险保费/保额
nprem_tp/si_tp	商业三者险
nprem_vld/si_vld	司机座位责任险保费/保额
nprem_vlp/si_vlp	乘客座位责任险保费/保额
.....	
suiche_nonauto_nprem(16-20)	16-20年随车非车保费
suiche_nonauto_amount(16-20)	16-20年随车非车保额
dur_personal_insurance_90	点击时长非车险

## 2.数据探索

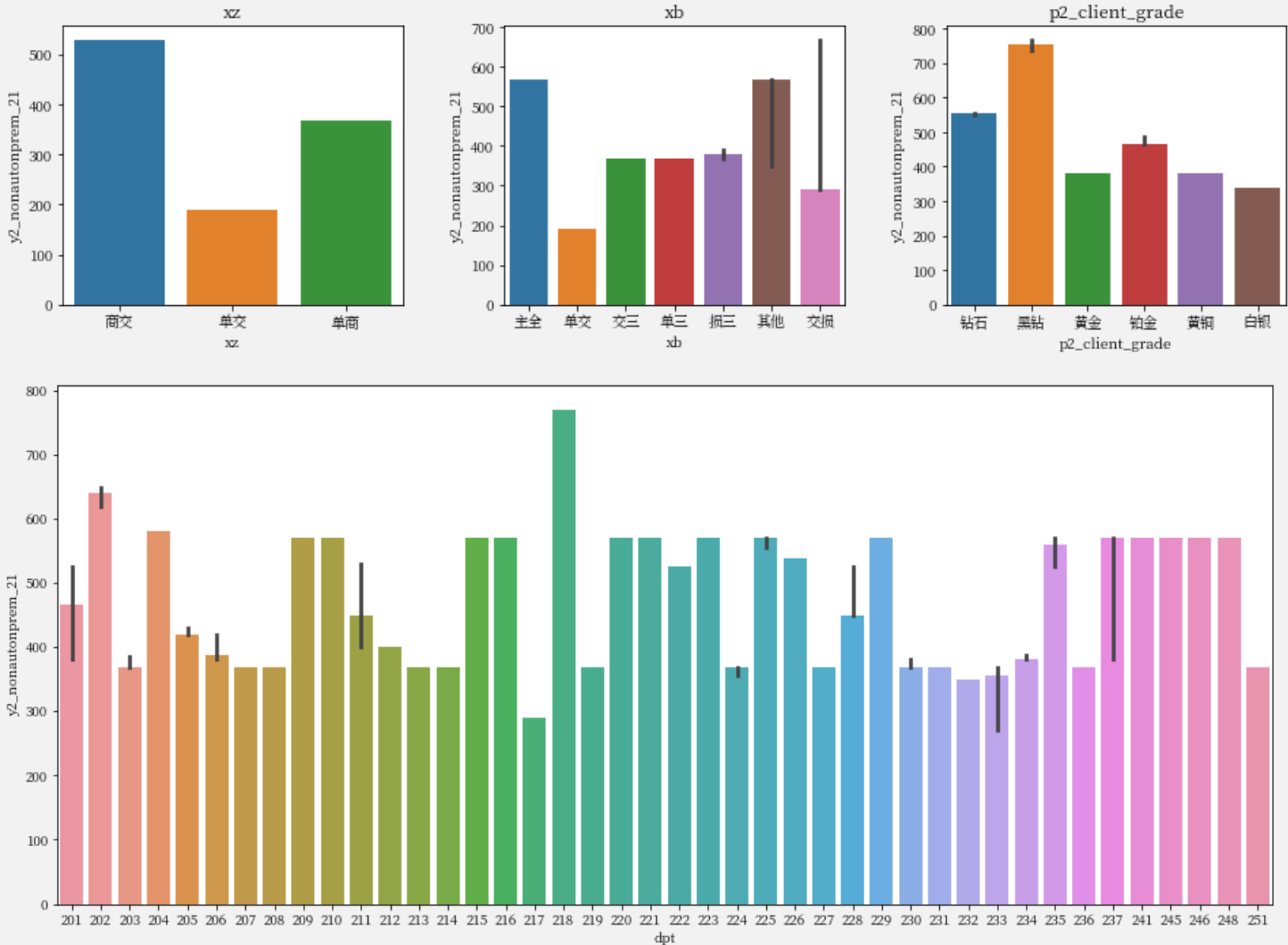


总的训练集有60w左右，其中用户购买非车险的保费总体呈现长尾分布。  
90%的用户非车险保费都小于1000，  
只有1%的用户非车险保费大于1500，  
极少数用户非车保费达到5000甚至1w。

## 2.数据探索——数据预处理



## 2.数据探索——单因子分析

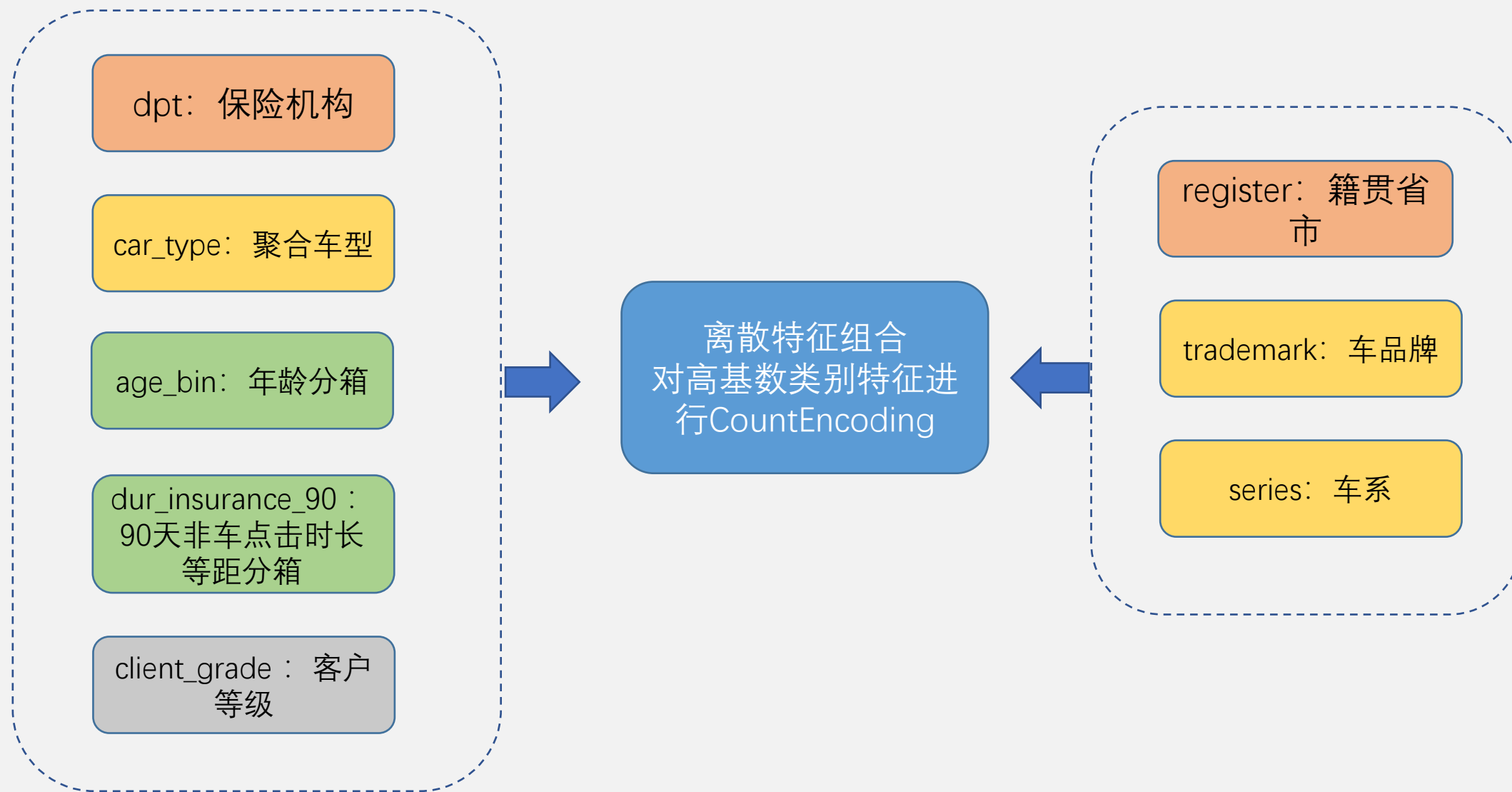


### 1.离散属性

对于类别特征不同类别购买非车险保费的中位数可视化（如左图所示），

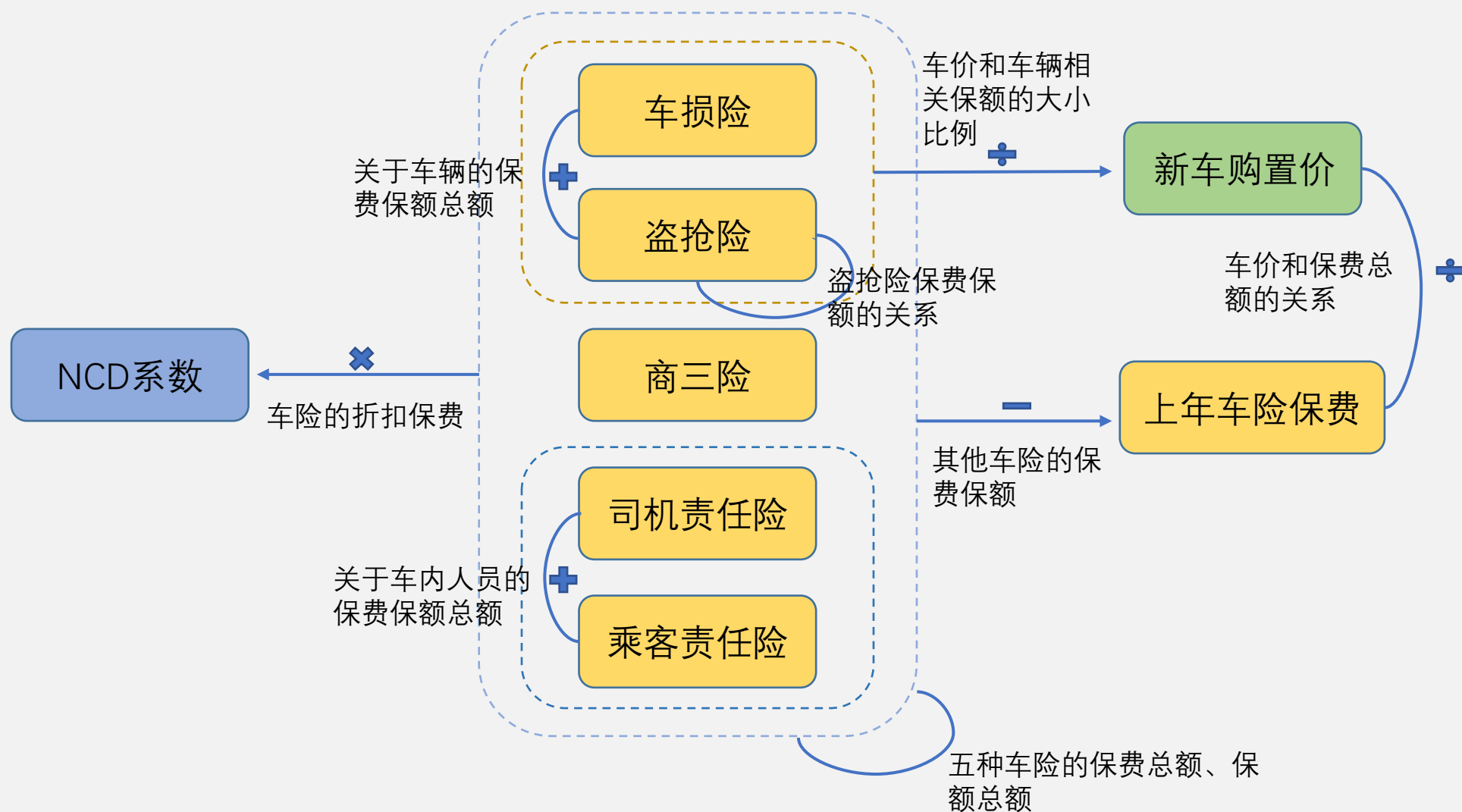
其中，对于**险别、险种、机构以及用户俱乐部等级**这几个类别特征而言，在每个类别上购买非车险保费的**中位数差异**比较明显，表明这几个类别特征用户选择不同类别对于其购买非车险有着明显的影响。

### 3.特征工程——用户画像特征

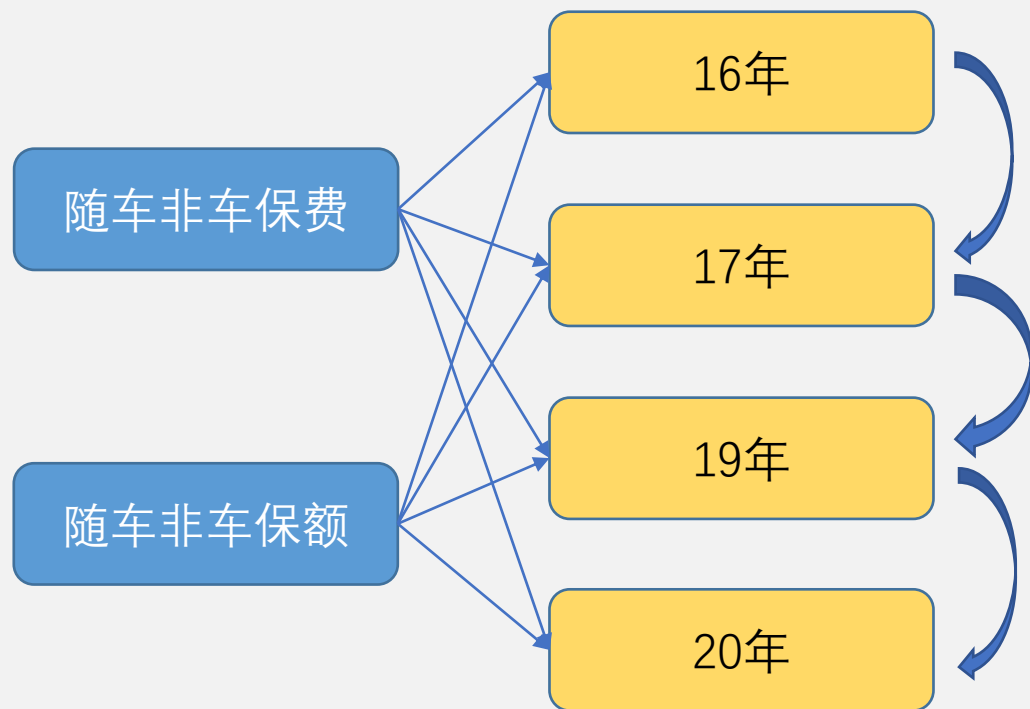




### 3.特征工程——连续特征构造



### 3.特征工程——历史非车险特征



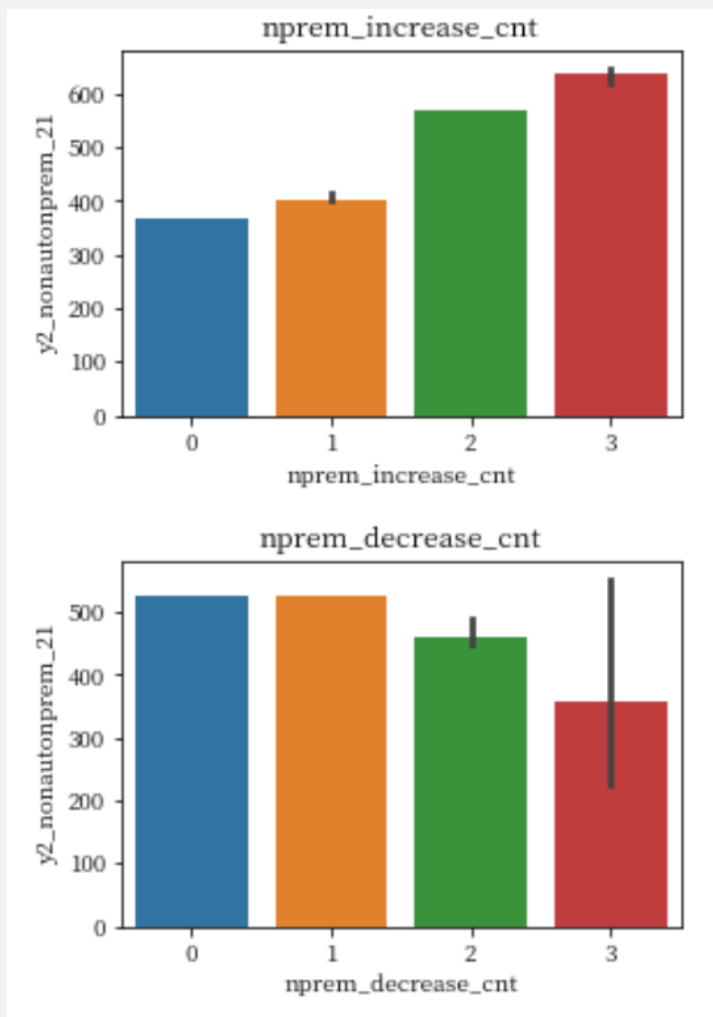
考虑历史购买非车险的保费保额的数值大小可能会直接影响用户现在购买非车险保费。

*(18年的数据存在大量的缺失且只有少量非零值, 因此最后去掉了18年的数据)*

构造历史非车险保费的统计特征, 用于刻画用户历史购买非车险的情况。

- 每年相比前一年非车保费 (保额) 变化的差值, 以及变化的方差、均值;
- 每年相比与前一年, 保费 (保额) 增加/下降的年份计数;
- 每年非车保费和保额的比例 (只统计了20、19年);
- 近五年保费, 保额, 保费/保额的变化趋势 (逐年增加/逐年减少/有增有减)。

### 3.特征工程——历史非车险特征



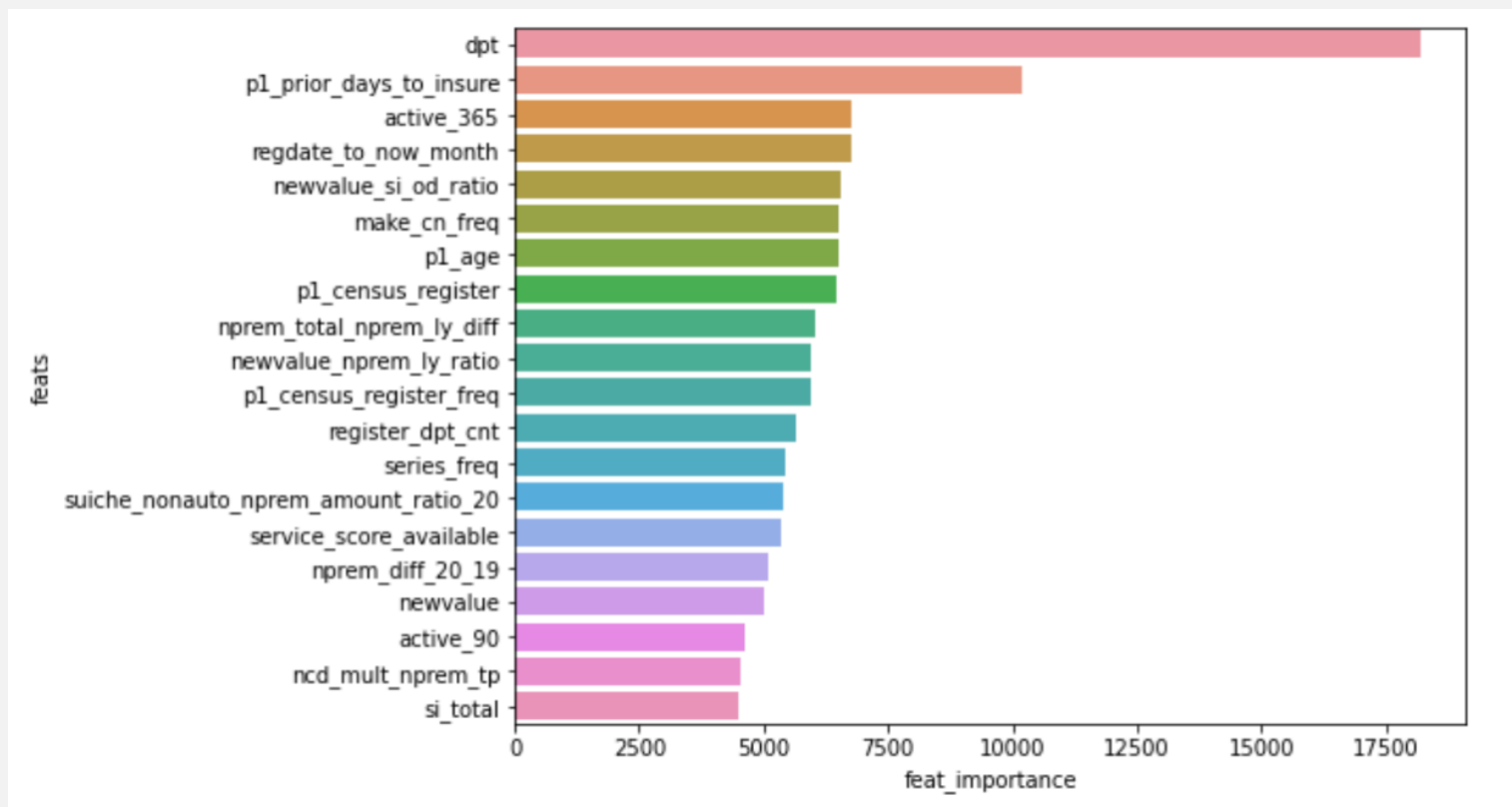
近五年内，随车非车保费相比前一年保费 增加/减少的年份数。

其中0表示过去五年内每年的非车险保费相比前一年都没有增加，可能是部分用户近五年都可能没有随车非车保费的购买。

通过分析增加年份数1，2，3发现，**保费增加的年份数越高，购买非车险的保费也会更高。**（左边上面的图）  
反之同理，对于保费减少的年份数而言，**减少的年份数越多，购买非车险的保费是会更低。**（左边下面的图）

### 3.特征工程——特征选择

Lightgbm模型输出特征重要度，降序排序，选择top20可视化



1.机构

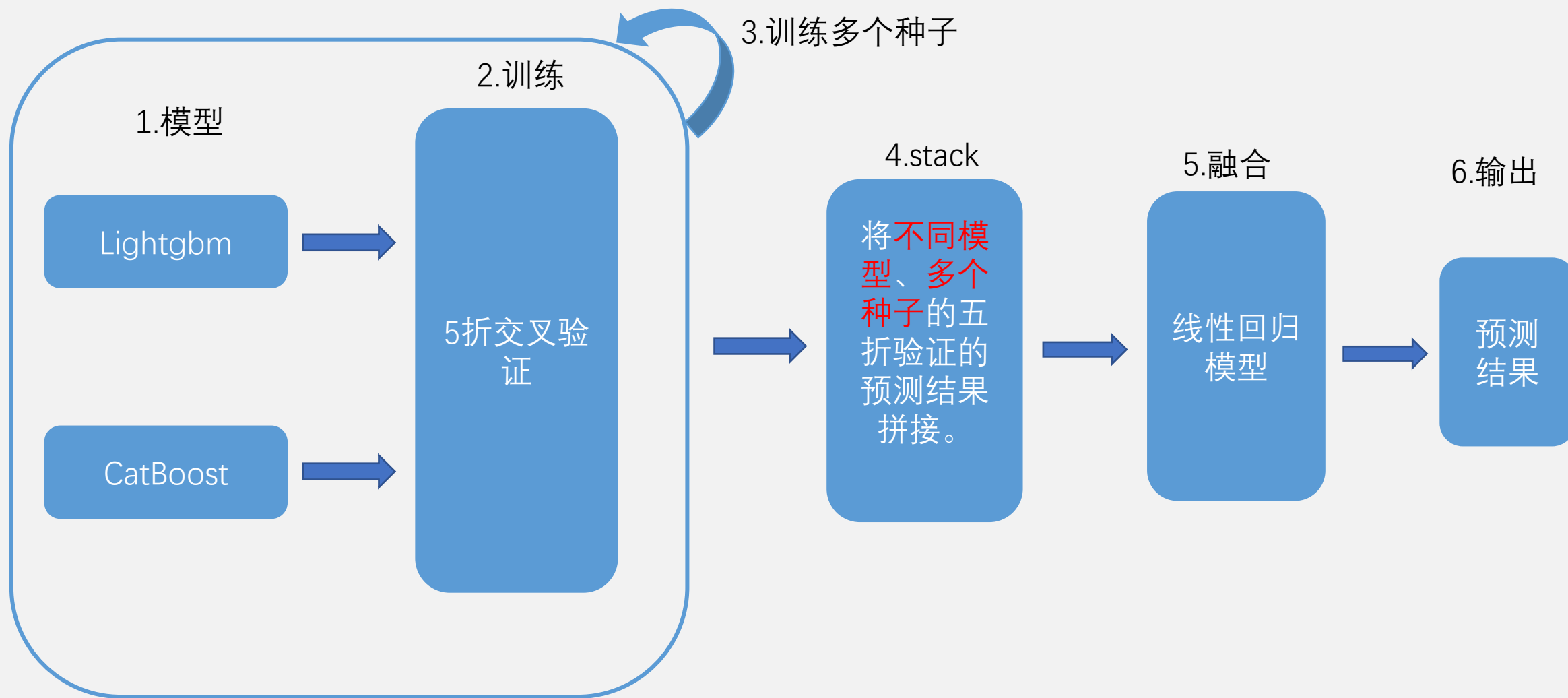
2.提前续保时间

3.近一年的活跃  
天数

4.新车价格和车  
损保额的比例

5.车型的  
CountEncoding

## 4.模型框架



## 5.业务应用——客群划分

### 1.车险购买情况：

购买的车险**险种、险别**，以及购买车险的**机构**。

对于某些机构的非车险购买整体消费更高，以及购买不同险种、险别时用户非车险保费也有明显差异，**可以根据不同种类选择推荐非车险的不同套餐**。

### 3.历史非车险购买情况：

过去5年内是否购买过非车险，以及历史非车险购买保费的趋势，如果呈现上升趋势，**可以基于去年非车险保费推荐相同或更高等级的套餐**。

客群划分

### 2.用户的基本信息：

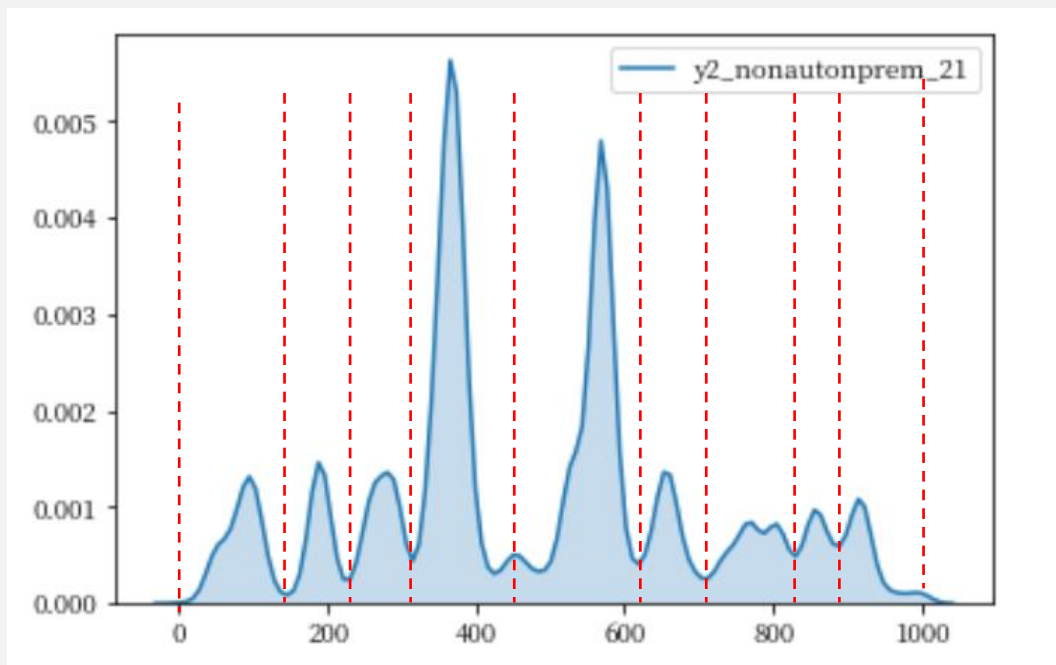
a.财产信息(车辆的价值、品牌、车辆年限，家庭房产价值等)，**推荐不同价格的非车险套餐**；

b.以及根据家庭(是否有老人小孩等)，**推荐一些“老人意外险”、“熊孩子险”之类非车险**。

c.根据用户活跃度(过去7、30天活跃度以及点击时长等)，对于**活跃用户**根据其他信息进行**推荐非车险**、**不活跃用户**则可以进行**非车险的基本介绍**。

具体如何选择不同套餐的非车险？

## 5.业务应用——非车险保费区间划分



因为90%的用户非车险保费都小于1000，因此对于1000以下和1000以上的分开考虑。

(1) 如左图所示，1000以下的非车险保费呈现多峰分布，即购买非车险保费有明显的不同区间聚集情况，可以以此进行划分区间（如左图中的红色虚线所示）。

(2) 1000-5000的保费则进行等距离划分，1000-2000、2000-3000等。

**具体推荐非车险的不同套餐时，可根据模型最终预测值坐落在的区间，选择相应区间的不同套餐进行推荐。**

## 5.业务应用——销售场景

### 销售场景

**1.线上销售**(移动端、pc端等):  
根据用户可能购买的非车险品种和套餐等级, 在用户浏览非车险的场景下进行**绑定的销售(可选)**, 一起购买则会提供**一定的优惠**。

**2.线下销售**: 可**结合不同的销售机构、地区, 在车险场景下进行一些特定品种的非车险推荐**, 例如当地台风、暴雨等比较多, 可推荐“家财宝”、“意外险”之类的险种。



感谢各位老师的聆听！