

统计分析与大数据分布式计算原理

李丰

feng.li@cufe.edu.cn

中央财经大学

事实

- 48 %的高管认为，大数据是一个有用的工具，而另外23 %的人认为大数据将彻底改变企业的管理方式。
- 缺乏了解如何利用大数据直接阻碍了大数据在企业中的应用。

想像中的大数据



实际的大数据



数据分布

- 据 IBM 称， 90 % 的组织收集的数据为商业数据。
- 社会化媒体在这份名单上排名最低， 仅为 39 %。
- 全世界 90% 的数据在过去两年被创造。

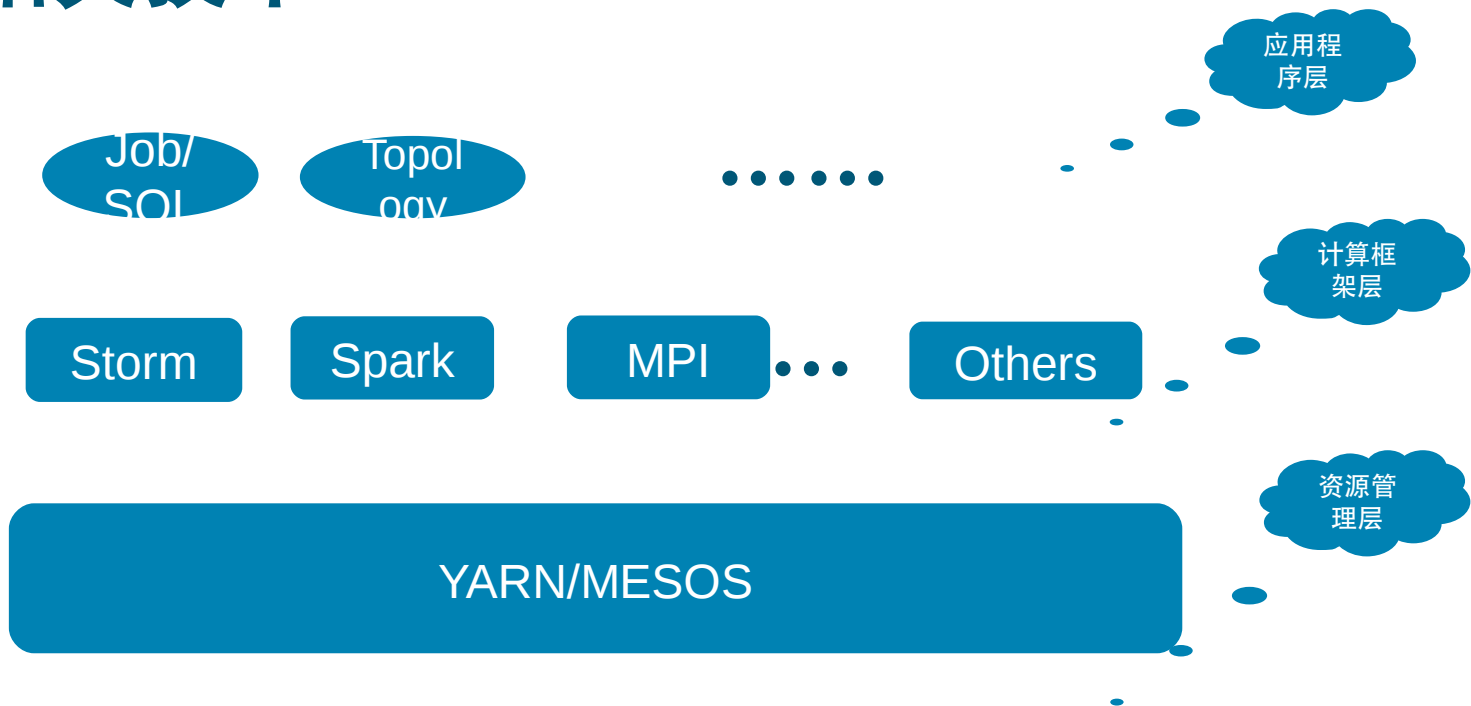
如何将传统统计工具应用到大数据

- MapReduce：用于大规模数据集（大于 1TB）的并行运算策略。
- **主要思想**：将一个巨大的统计任务拆分（map）计算并最终合并结果（reduce）
- 假设任务之间是相互独立的。

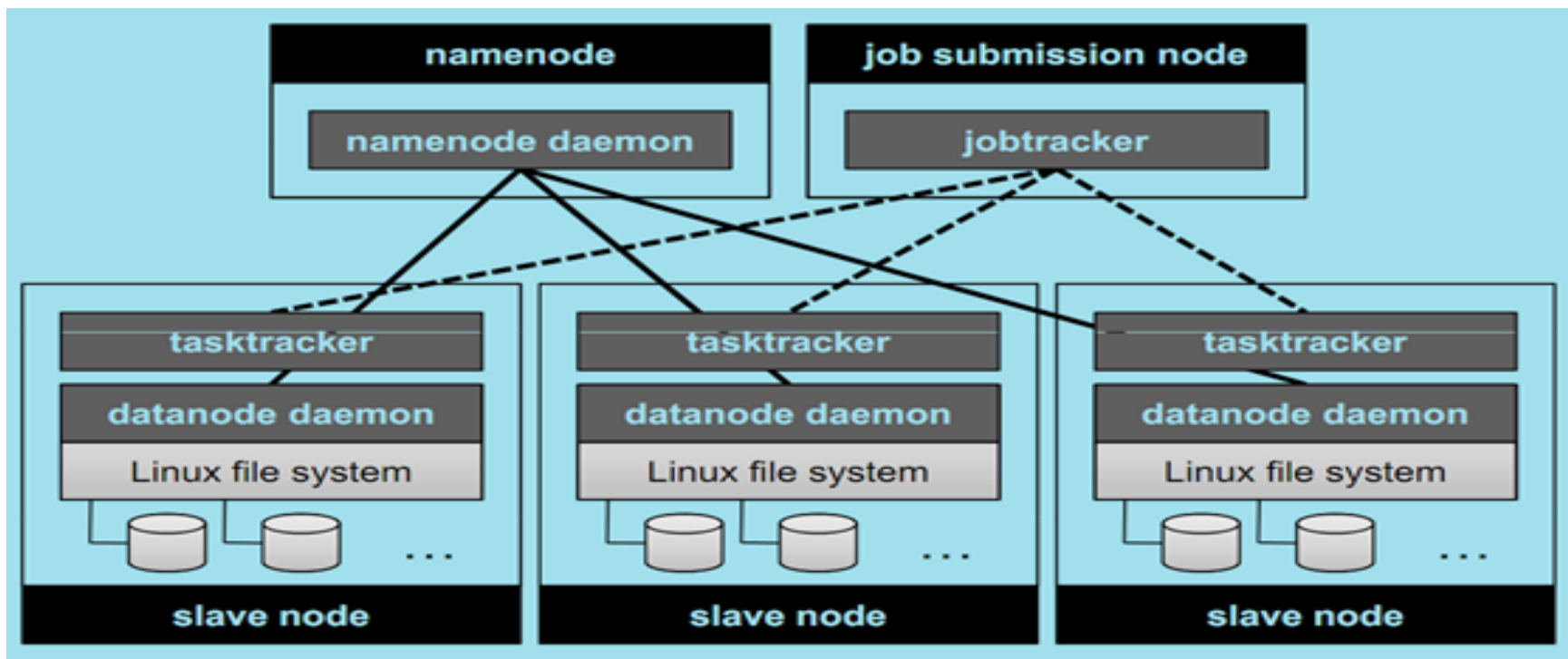
MapReduce 示例

- 计算全国所有火力发电场的平均发电量。
- 分别计算各个省的火电发电平均量（map）
- 将所有的平均量汇总（reduce）

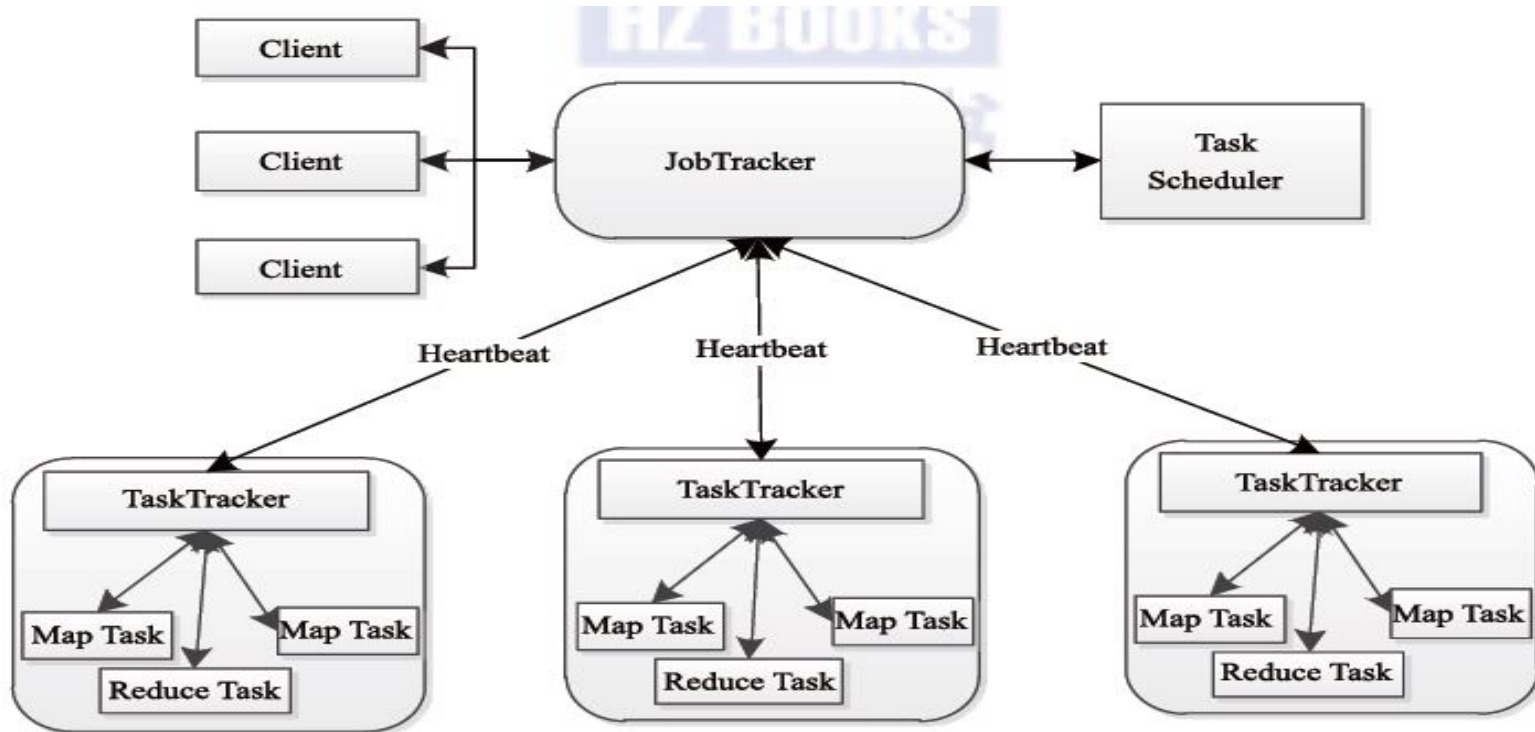
大数据相关技术



大数据相关技术 ---Hadoop

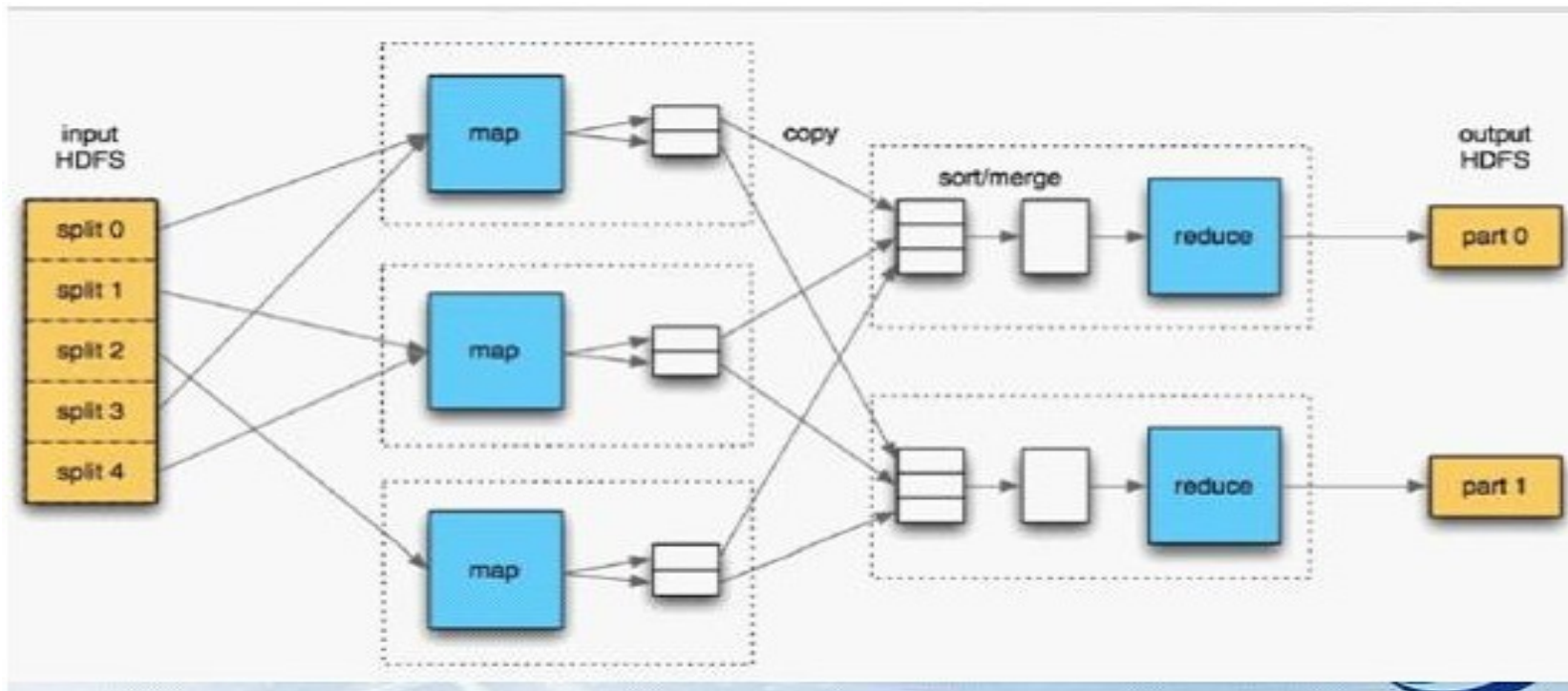


大数据相关技术 ---Hadoop



大数据相关技术 ---Hadoop

Hadoop job 数据处理流程



大数据相关技术 ---Hadoop

Hadoop 数据处理实例

```
#!/bin/bash
↵
.. /bin/hadoop fs -rmr /example/data/output↵
↵
.. /bin/hadoop jar ↵
/home/administrator/softwares/hadoop-2.6.0/share/hadoop/tools/lib/hadoop-streami
ng-2.6.0.jar ↵
-D stream.map.output.field.separator=',' ↵
-D stream.num.map.output.key.fields=1 ↵
-partitioner org.apache.hadoop.mapred.lib.KeyFieldBasedPartitioner ↵
-input /example/data/input ↵
-output /example/data/output ↵
-mapper /bin/cat ↵
-reducer reduce ↵
-numReduceTasks 1 ↵
-file reduce↵
```

Hadoop streaming 实现
mapper 为 shell 命令
reducer 为 c++ 程序

李伟 | 大数据

大数据相关技术 --- 总结

大数据处理核心思路：大而化小，按一定规则的分布式处理



分布式计算技术：map/reduce spark storm 等等

分布式存储：分布式文件系统（hdfs），
分布式 NoSql KV 存储（hbase 等）

以大数据的视角考虑问题

- 数据特有的时效性。
- 数据的隐私安全。
 - 企业数据安全
 - 个人存在企业的数据安全
 - 利用大数据对个人隐私的侵犯
- 如果数据不使用即便再大也是死数据。
- 我们实际面对的更多的是复杂数据。
- 大数据和小问题

实例分析：基于大数据的P2P网络借贷平台 散标投融资行为决策研究

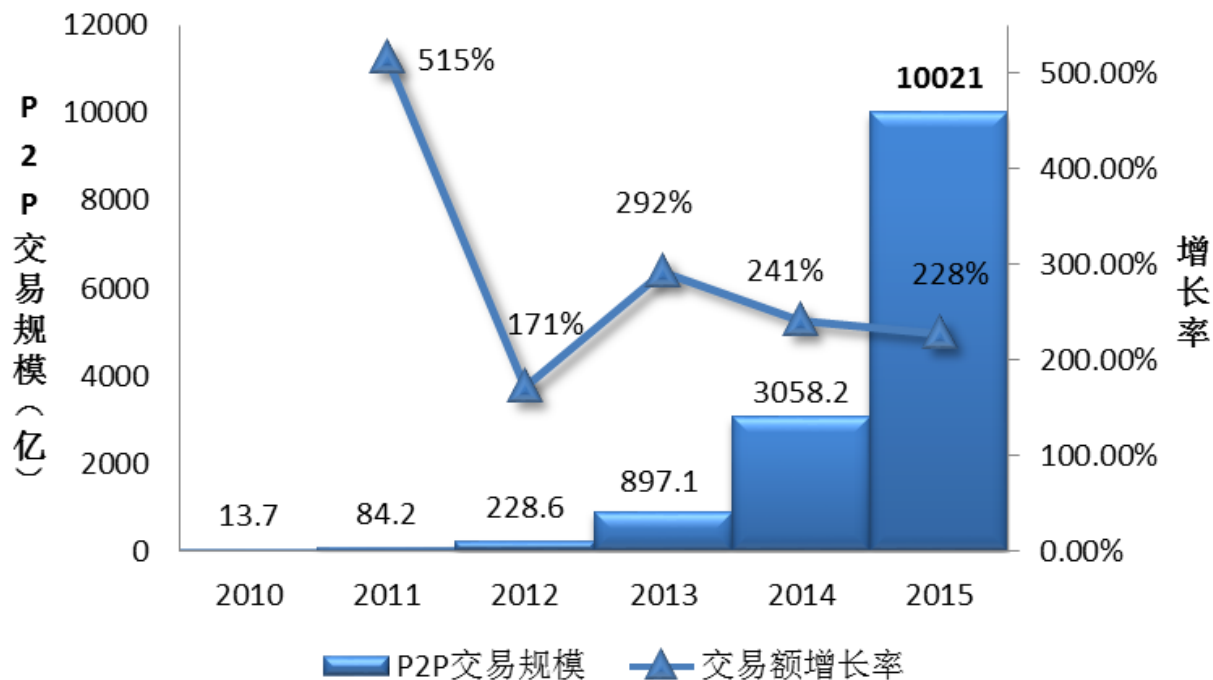
背景

- 在各国中小微企业在国民经济中扮演了非常重要的角色。
- 中小微企业及个体的小额贷款因面临**信息披露制度不完善**以及其**风险评估存在困难**等问题，面临较大困境。
- 随着计算机技术与互联网技术的普及以及电子商务的飞速发展，小额信贷开始逐渐与互联网技术相融合，基于网络技术的**P2P网络借贷平台**便应运而生。



中国P2P市场

- 国内的P2P网络借贷平台虽然行业起步较晚，但短短几年时间内发展迅速。据悉，2008年以来P2P网络借贷平台数量由十几家大幅增长到将近4000家。2015年，我国P2P网贷平台数量为**3844**，同比增长**107.34%**，另外网贷交易规模为**10021亿**，同比增长**228%**。



- 尽管新成立平台的增速趋于平缓，但是整个网贷行业的成交额则呈指数型增长的态势。预计未来两年内仍然保持200%左右的增速发展



三、数据来源介绍

（一）P2P网络借贷平台选择——人人贷（www.we.com）

- 中国最早一批基于互联网的P2P信用借贷服务平台：2010年10月网站正式上线。
- 截至2015年12月，公司成立仅5年的人人贷网站累计成交量已达到**132.1亿人民币**。同时人人贷曾多次在“中国网贷评价体系”中获得**网贷AAA级最高评级**。
- 至今，人人贷的服务已覆盖了全国**30余个省的2000多个地区**，服务了**一百五十万**名客户，成功帮助他们通过信用申请获得融资借款，或通过自主出借获得稳定收益。



（二）平台选择理由

- 1、“人人贷”**用户量还是贷款规模都处于领先地位**，且“人人贷”更早的为用户提供**了本金或是本金+利息**的保障服务，这点显然吸引了更多用户使用该平台。其平台出于实际需要的交易更多。
- 2、“人人贷”对**用户收取的费用**（包括理财人的充值费用、提现费用、债权转让费用及借款人的借款服务费、借款管理费、提现费用、充值费用、逾期罚息和逾期管理费）**均较低**，因此更容易有客户的实际参与交易。
- 3、“人人贷”平台综合评级高，**用户较为信赖平台**。
- 4、“人人贷”平台成立已较长时间，积累了大量借款交易信息，适合研究



（三）数据获取情况

- 目前“人人贷”平台已有80余万条交易记录，但由于网站对于连续时间内大量爬取其平台数据有防御措施，会限制该IP访问，另结合计算机等因素，本文随机采集了编号在000001至800000间人人贷平台的共40000笔交易的数据。
- （利用Hadoop MapReduce获取这些交易数据，这里省略10,000字）



（四）数据获取思路

- 选定平台后，查看交易页面范例。

<http://www.we.com/lend/detailPage.action?loanId=0701816>



借款交易页面范例

客服电话: 400-090-6600

立即登录 | 快速注册 | 关于我们 | 帮助中心 | 理财论坛 | 移动客户端

人人贷Wealth Evolution

人人贷旗下理财品牌

首页 | 基金 | U计划 | 薪计划 | 债权 | 新手专区 | 我的账户

散标投资 | 债权转让

信 用于家庭装修及家具购置

借款协议 (范本)

30,000元

标的总额

12.00%

年利率

9个月

还款期限

3月

剩余期数

2016-03-24

下一合约还款日

保障方式 本金 ①

提前还款费率 1.00%

还款方式 按月还款/等额本息 ①

散标详情

投标记录

还款表现

债权信息

转让记录

借贷人信息

请 登录 或 注册 后查看

审核状态

审核项目	状态	通过日期
------	----	------



借款交易页面范例

请 [登录](#) 或 [注册](#) 后查看

审核状态

审核项目	状态	通过日期
信用报告	✓ 已完成	2015-08-24
身份认证	✓ 已完成	2015-08-14
工作认证 (工薪阶层)	✓ 已完成	2015-08-24
收入认证	✓ 已完成	2015-08-24
房产认证	✓ 已完成	2015-08-14

- 1.人人贷及其合作机构将始终秉持客观公正的原则，严控风险，最大程度的尽力确保借入者信息的真实性，但不保证审核信息100%无误。
- 2.借入者若长期逾期，其个人信息将被公布。

借款描述

本人在IT行业工作8年，是公司主层管理人员，年收入15W左右。近期由于要装修房子及新房购置家具需要资金周转。详见征集报告。

☰

🖨️

📄

⬆️



登录账户

人人贷

Wealth Evolution
人人贷旗下理财品牌

登录

人人贷WE理财全面升级

一样的账户 更好的服务

人人贷WE理财资金由民生银行存管

请输入手机号/邮箱

请输入密码

☐ 记住用户名 [忘记密码](#)

立即登录

[没有账号？免费注册](#)

© 2016 人人贷 All rights reserved | 人人贷商务顾问(北京)有限公司 | 京公网安备11010502020657 | 京ICP备12025643号-1

查看进一步信息

散标详情

投标记录

还款表现

债权信息

转让记录

借贷人信息

昵称 maovv

信用评级 

基础信息

年龄 34

学历 大专

婚姻 已婚

信用信息

申请借款 1笔

信用额度 30,000.00元

逾期金额 0.00元

成功借款 1笔

借款总额 30,000.00元

逾期次数 1次

还清笔数 0笔

待还本息 10,494.00元

严重逾期 0笔

资产信息

收入 5000-10000元

房产 有房产

房贷 有房贷

车产 有车产

车贷 无车贷

工作信息

公司行业 IT

公司规模 10-100人

岗位职位 部门经理

工作城市 湖北 武汉

工作时间 5年以上

审核状态

审核项目	状态	通过日期
信用报告	 已完成	2015-08-24



查看源代码

```
248
249
250
251 <div id="loan-basic">
252   <div class="container_12_1080 mt10 color-white-bg pg-details" id="loan-basic-panel">
253     <div class="ui-box-white-bg fn-clear">
254       <div class="ui-box-title fn-clear">
255         <div class="fn-left fn-text-overflow pl25" title="信用认证标"><em class="icon-loantype XYRZ mr5"></em><em class="title-text">用于家庭装修及家具购置</em></div>
256
257
258         <a class="fn-right text ui-box-text pr25" href="javascript:void(0)" onclick="javascript:window.open('../account/emptyContract.action?loanId=701816', '_blank', 'height=768,
width=1024,toolbar=no,scrollbars=yes,menubar=no,status=no')">借款协议（范本）</a>
259       </div>
260       <input type="hidden" id="invest-type" value="loan" />
261       <div class="pl25 pr25 fn-clear">
262         <div class="fn-left pt10 loaninfo">
263           <div class="fn-clear mb25">
264             <dl class="fn-left w300">
265               <dd class="basic-all color-orange-text">
266                 <em class="text-xxx1 num-family">30,000</em><span class="font-18px basic-text">元</span>
267               </dd>
268               <dt class="minus-mt-5">标的总额</dt>
269             </dl>
270             <dl class="fn-left w240">
271               <dd class="basic-all">
272                 <em class="text-xxx1 num-family color-dark-text" style="margin-left:-4px;">12.00</em><em class="font-24px">%</em>
273               </dd>
274               <dt class="minus-mt-5">年利率</dt>
275             </dl>
276             <dl class="fn-left w140">
277               <dd class="basic-all">
278                 <em class="text-xxx1 num-family color-dark-text">9</em><span class="font-18px basic-text">个月</span>
279               </dd>
280               <dt class="minus-mt-5">还款期限</dt>
281             </dl>
282           </div>
283           <ul>
284             <li class="fn-clear">
285               <span class="fn-left basic-label w70">保障方式</span>
286               <span class="fn-left basic-value last">本金 <a id="tips_0" class="icon-info color-gray-text" href="/help/security/security!detail.action?flag=bjbz" target="_blank"></a>
287
288               <span class="fn-left basic-label mr30">提前还款费率</span>
289               <span class="fn-left basic-value num"><em>1.00</em>%</span>
290             </li>
291             <li class="fn-clear">
292               <span class="fn-left basic-label w70">还款方式</span>
293               <span class="fn-left basic-value">按月还款/等额本息 <a id="tips_1" class="icon-info color-gray-text" href="/help/terms/terms!terms.action#equalInterest"
target="_blank"></a></span>
294
295               <span class="fn-left basic-label mr30">月还本息（元）</span>
296               <span class="fn-left basic-value num last"><em>3,504.00</em></span>-->
297             </li>
298           </ul>
299         </div>
300       </div>
301     </div>
302   </div>
303 </div>
```



```
496         <td colspan="3"><div class="color-tab-title-bg">工作信息</div></td></tr>
497         <td class="basic-filed-1"><div class="basic-filed"><span>公司行业</span><em> IT</em></div></td>
498         <td class="basic-filed-2"><div class="basic-filed"><span>公司规模</span><em>10-100人</em></div></td>
499     <td class="basic-filed-3"><div class="basic-filed"><span>岗位职位</span><em> 部门经理</em></div></td>
500     </tr><tr>
501         <td class="basic-filed-1"><div class="basic-filed"><span>工作城市</span><em>湖北 武汉</em></div></td>
502         <td colspan="2"><div class="basic-filed"><span>工作时间</span><em>5年以上</em></div></td>
503
504
505 </table>
506
507
508 </div>
509
510
511 <div class="ui-tab-content-auditoria border-bottom p25 minus-mt-10 plr60">
512     <h4 class="ui-tab-content-title color-dark-text">审核状态</h4>
513     <div class="details-verfication-list mt10 color-lightdark-text" id="verification-container"></div>
514     <ul>
515         <li>1.人人贷及其合作机构将始终秉持客观公正的原则，严控风险，最大程度的尽力确保借入者信息的真实性，但不保证审核信息100%无误。</li>
516         <li>2.借入者若长期逾期，其个人信息将被公布。</li>
517     </ul>
518 </div>
519
520 <div class="ui-tab-content-description border-bottom p25">
521     <h4 class="ui-tab-content-title color-dark-text">借款描述</h4>
522     <div class="ui-tab-list color-dark-text">
523         本人在IT行业工作8年，是公司主层管理人员，年收入15W左右。近期由于要装修房子及新房购置家具需要资金周转。详见征集报告。
524     </div>
525 </div>
526
527     <div class="ui-tab-content" data-name="investment">
528
529
530
531
532
533 <div class="text-right text color-lightdark-text mb10 pt10 pr25">
534     <span class="mr50">加入人次 <em id="investor-count"></em> 人</span>
535     <span class="mr10">投标总额 <em>30,000</em>元</span>
536 </div>
537 <div class="ui-list ui-list-s" id="investment-records"></div>
538
539
540 </div>
541
542     <div class="ui-tab-content" data-name="repayment">
543
544
545
546
```

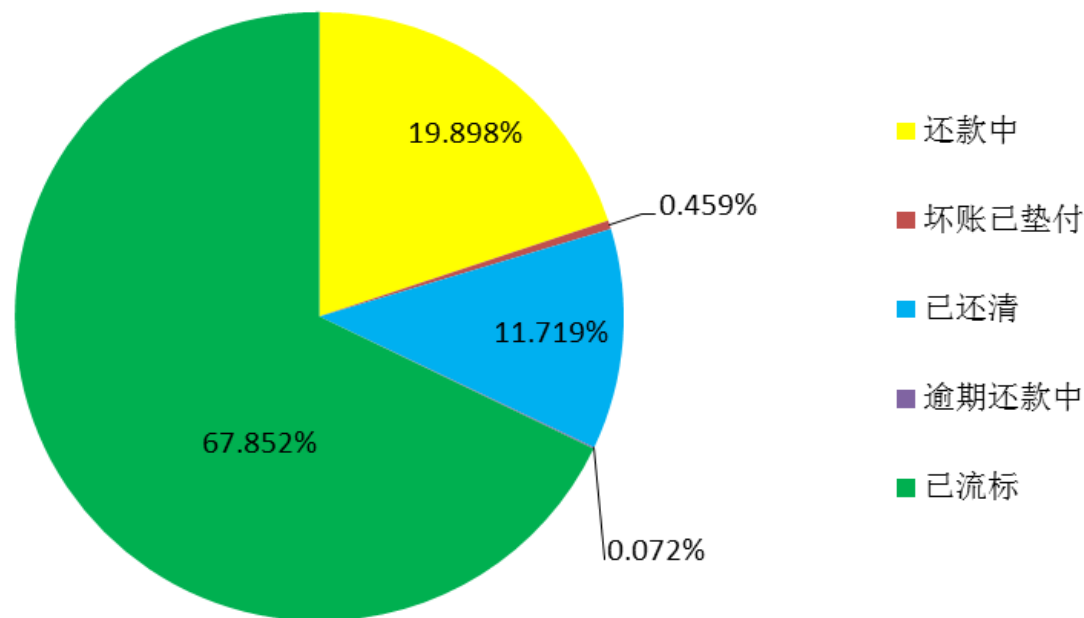
数据范例

信息类别	变量	交易编号	150000	150001	150002
交易基本信息	状态	流标	流标	流标	流标
	用途	其他借款	个人消费	购车借款	购车借款
	标的类型	信用认证标	信用认证标	信用认证标	信用认证标
	借款金额	30, 000	50, 000	50, 000	50, 000
	年利率	15	20	15	15
	还款期限月	12	24	24	24
	保障方式	本金	本金	本金	本金
	还款方式	等额本息	等额本息	等额本息	等额本息
	投标总额	0	0	0	0
	担保人偿还	0	0	0	0
	借款人id	570403	570400	570134	570134
借款人基本信息	年龄	32	24	32	32
	学历	本科	大专	高中或以下	高中或以下
	公司行业	政府机关	教育/培训	制造业	制造业
	公司规模	10-100人	10-100人	10-100人	10-100人
	岗位职位	副镇长	营销经理	调机师	调机师
	工作城市	广西 河池	上海 嘉定	浙江 台州	浙江 台州
	工作时间	3-5年（含）	3-5年（含）	1-3年（含）	1-3年（含）
	收入范围	2000-5000元	5000-10000元	2000-5000元	2000-5000元
	房产	1	0	0	0
	房贷	1	0	0	0
	车产	0	0	0	0
	车贷	0	0	0	0
	信用分数	0	20	0	0
	申请借款笔数	1	1	1	1
借款人平台信息	信用额度元	0	0	0	0
	逾期金额元	0	0	0	0
	成功借款笔数	0	0	0	0
	借款总额元	0	0	0	0
	逾期次数	0	0	0	0
	还清笔数	0	0	0	0
	待还本息元	0	0	0	0
	严重逾期笔数	0	0	0	0



四、借款信息及受众群体特征分析

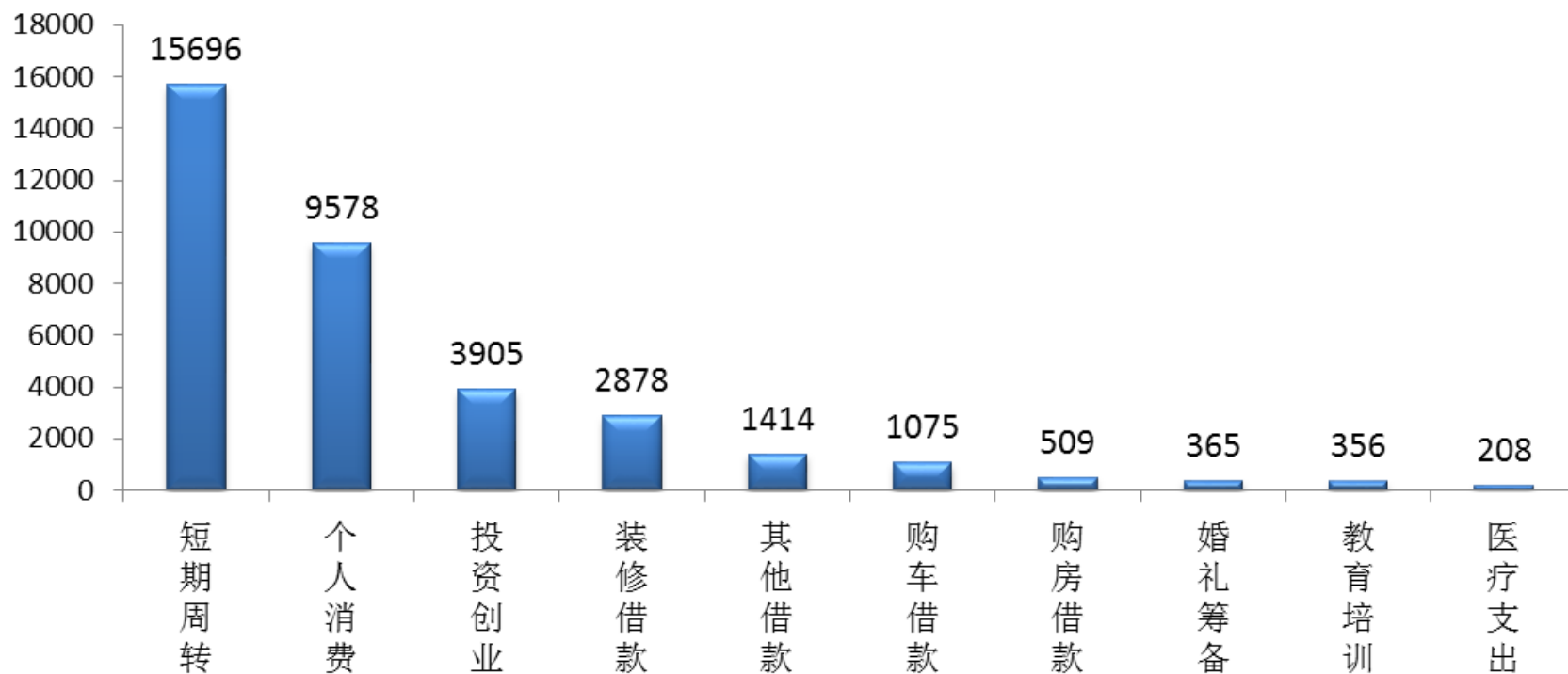
- （一）借款信息的描述性统计
- 1、借款交易状态



- 这显示了“人人贷”平台借款成功率约为32%，不是很理想，另一方面该平台逾期还款与坏账垫付的概率较小，说明该平台用户信用相对可靠。



- 2、借款用途



- 如图所示，43.62%即近一半的借款用途都为短期周转，其次是个人的消费与投资创业，分别占了26.62%和10.85%，可见借款用户主要资金需求在工作与消费方面。

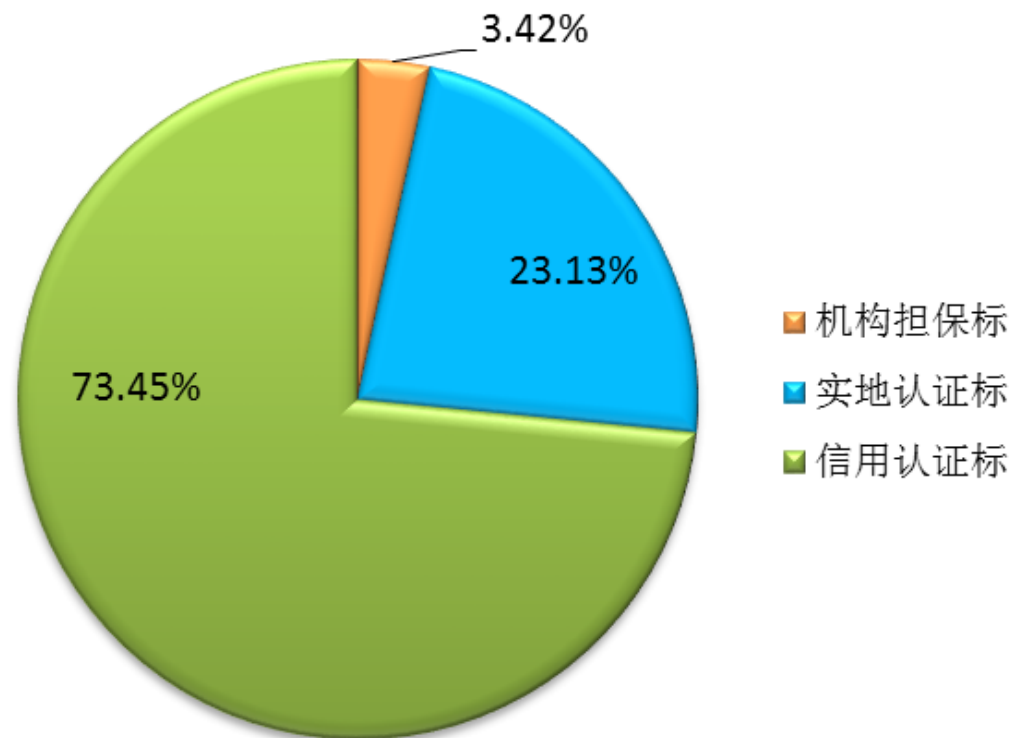


- 3、标的类型及保障方式
- “人人贷”平台共有三种标的类型：信用认证标、实地认证标与机构担保标。
- 信用认证标是人人贷通过对借款用户的个人信用资质进行全面审核后，允许用户发布的借款标。
- 实地认证标是人人贷与友众信业金融信息服务（上海）有限公司（以下简称“友信”）共同推出的产品，在原有严格审核的基础上，增加了友信前端工作人员对借款人情况的实地走访、审核调查以及后续的贷中、贷后服务环节，进一步加强风险管理控制，达到了双重保障的效果。



- 3、标的类型及保障方式
- 实地认证标相对信用认证标增添了实地认证审核，进一步保障了理财用户资金安全；同时采用本息保障的赔付方式，使用户理财更加省心、放心。机构担保标是指人人贷的合作伙伴为相应的借款承担连带保证责任的借款标的。所谓连带保证责任即连带保证人对债务人负连带责任，无论主债务人的财产是否能够清偿债务，债权人均有权要求保证人履行保证义务。针对机构担保标借款申请人，人人贷会通过严格的审核系统进行双重审核，严控风险。此外，一旦合作伙伴违背其应承担的连带保证责任，根据合作协议人人贷有权通过法律手段进行追偿。

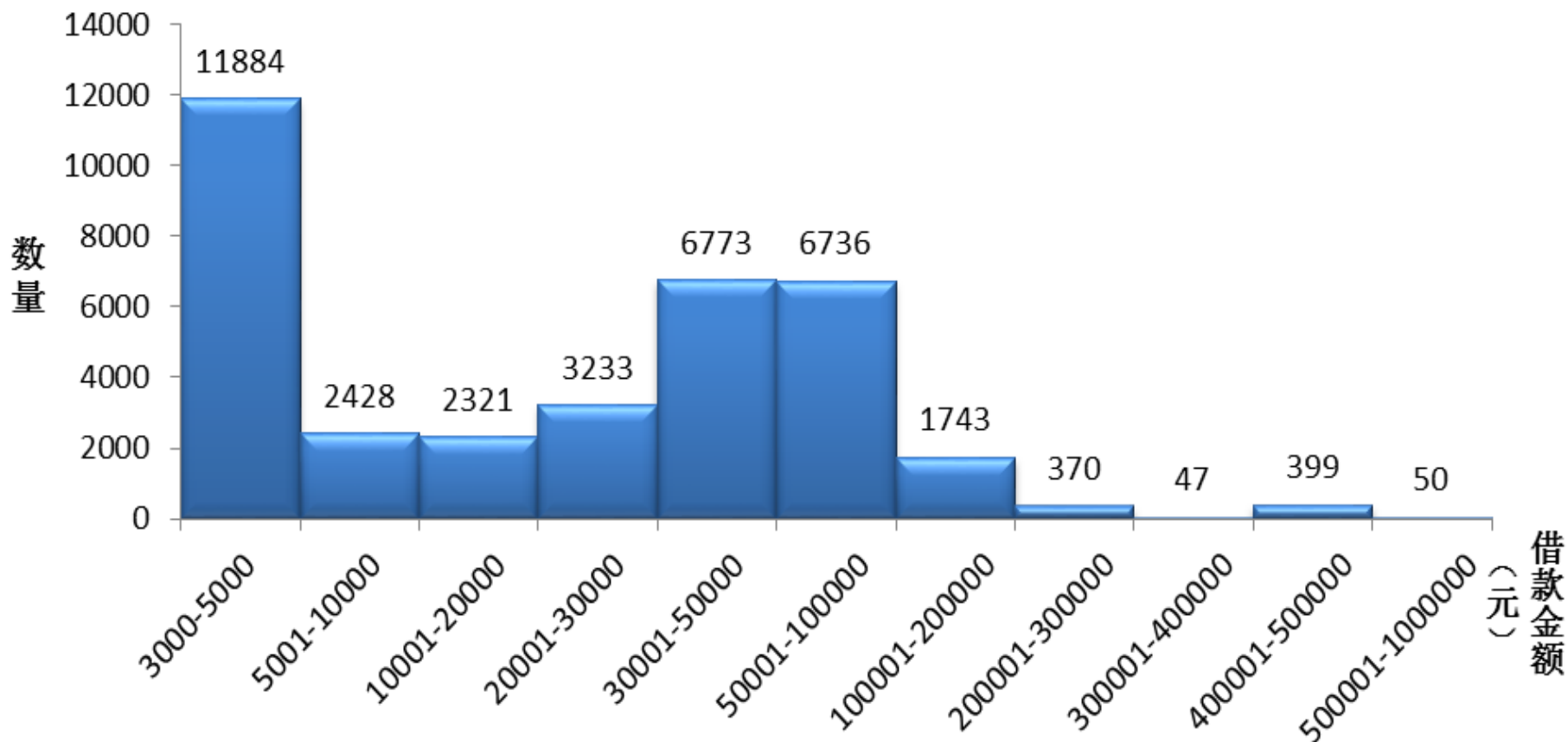




- 信用认证标的保障方式是本金保障，而实地认证标与机构担保标的保障方式均为本息保障。其中，信用认证标有26431笔，实地认证标与机构担保标分别有8322与1231笔。可以看出，大部分借款的标的类型均为信用认证标，其保障性相对来说较为一般。



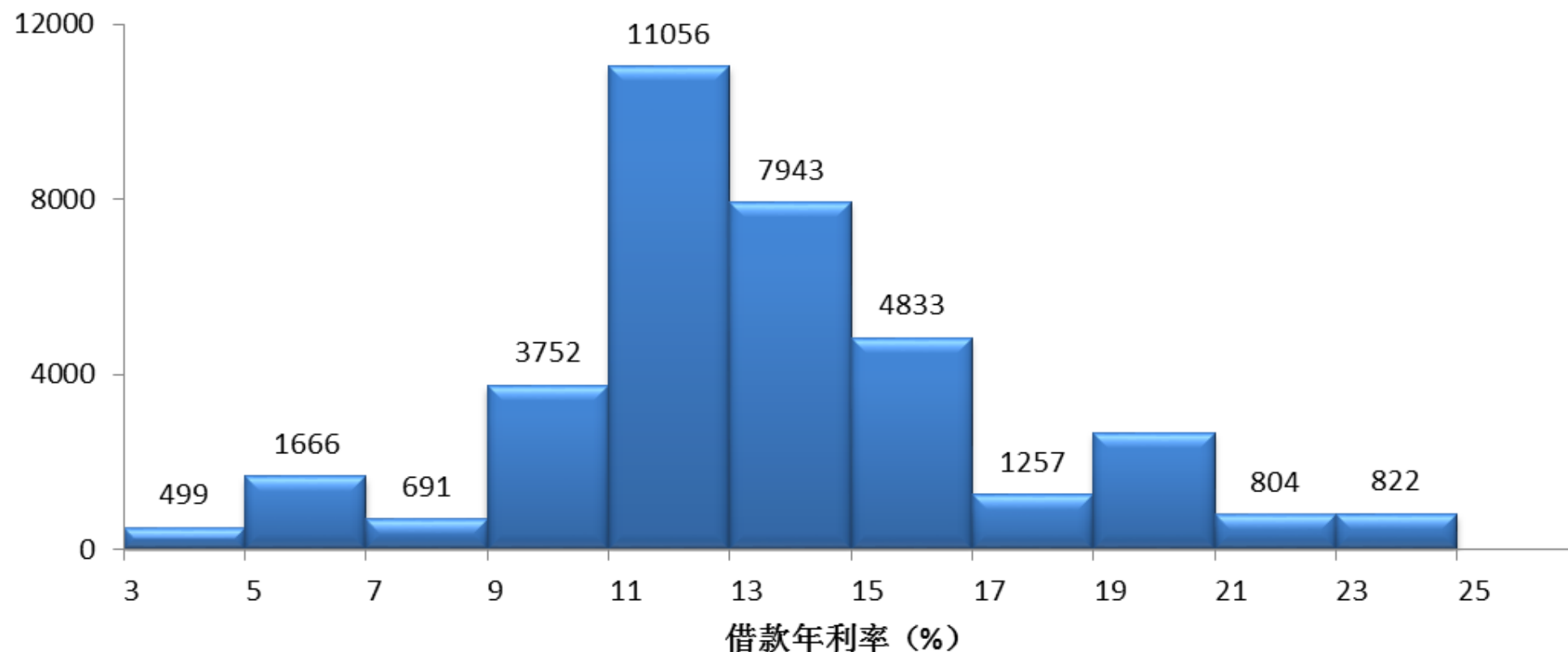
- 4、借款金额



综合本文35987个样本，借款金额平均为46502.50元，其中用户发起的借款金额大多小于100000元，



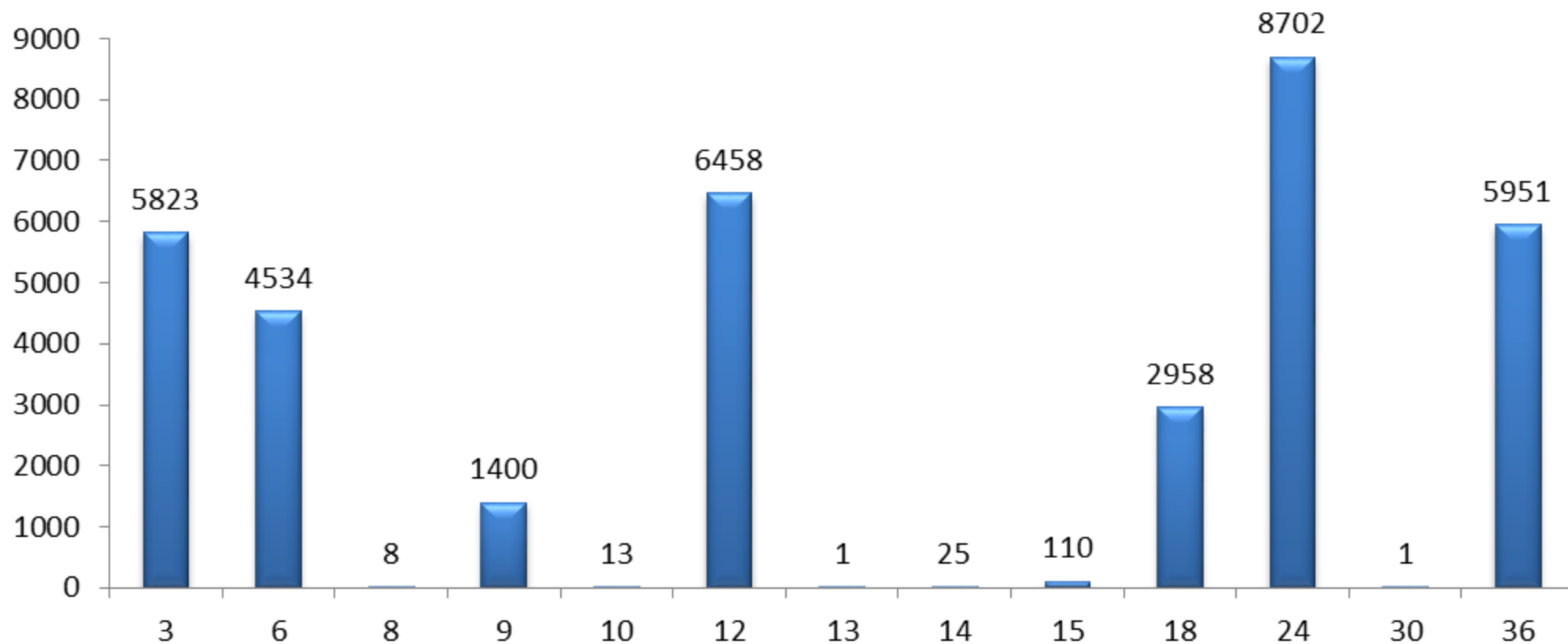
- 5、年利率



根据样本可得，该平台观测记录内发起的借款交易年利率平均为13.23%，其中，在[11%，17%)中的数量最多，占到了66.23%。



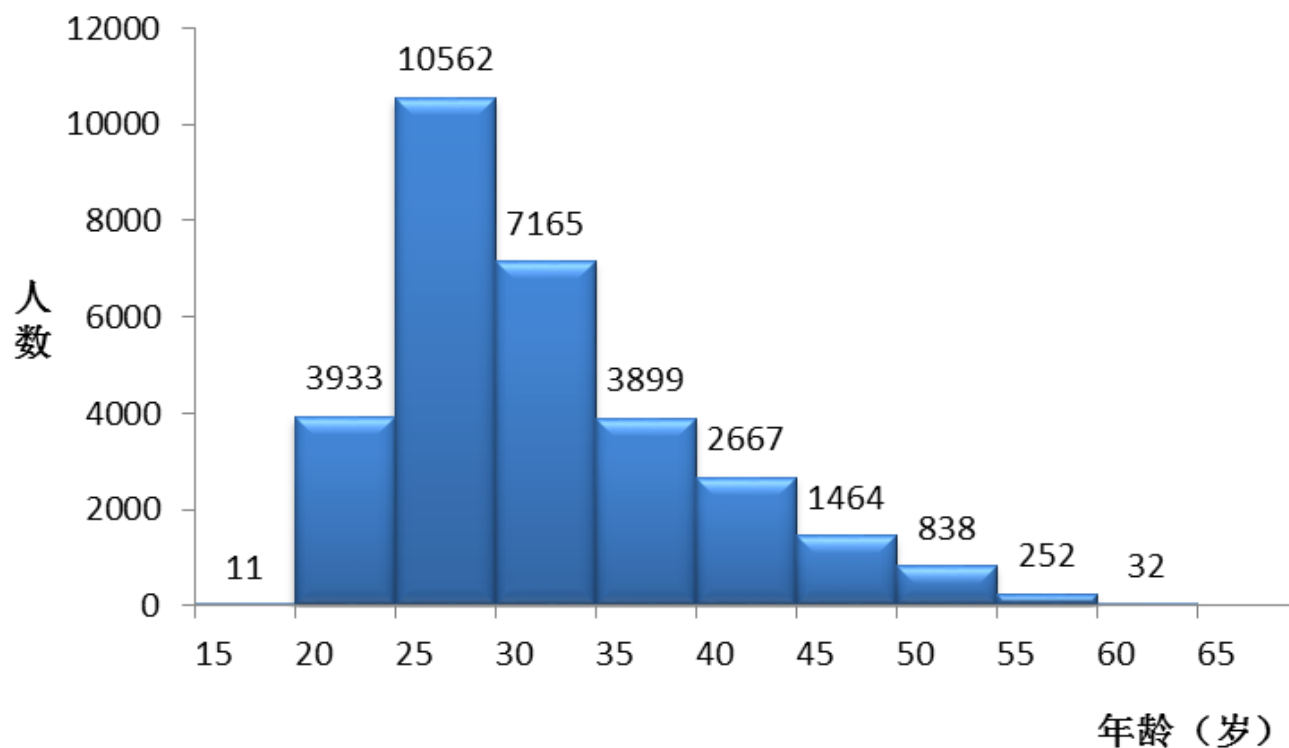
- 6、还款期限



在本文35984条借款记录中，还款期限平均为17.04个月，最小值为3个月，最大值为36个月，中位数为12个月，众数为24个月。标准差为11.21个月。



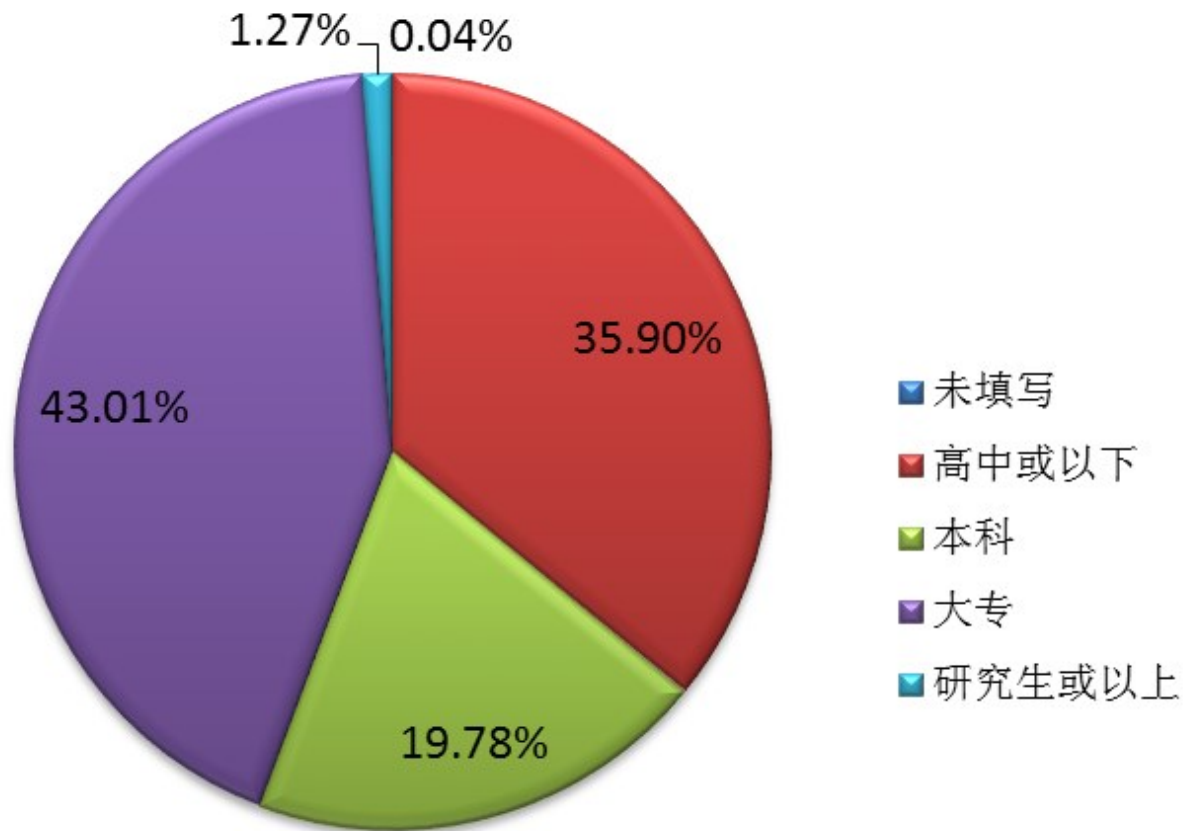
- （二）借款用户的描述性统计
- 在35984个借款样本中，共有30823名用户发起借款。
- 1、年龄



借款用户的平均年龄为33岁。借款用户的平均年龄为33岁。其中20至30岁用户占70.27%，总体来说借款用户属于较年轻群体。



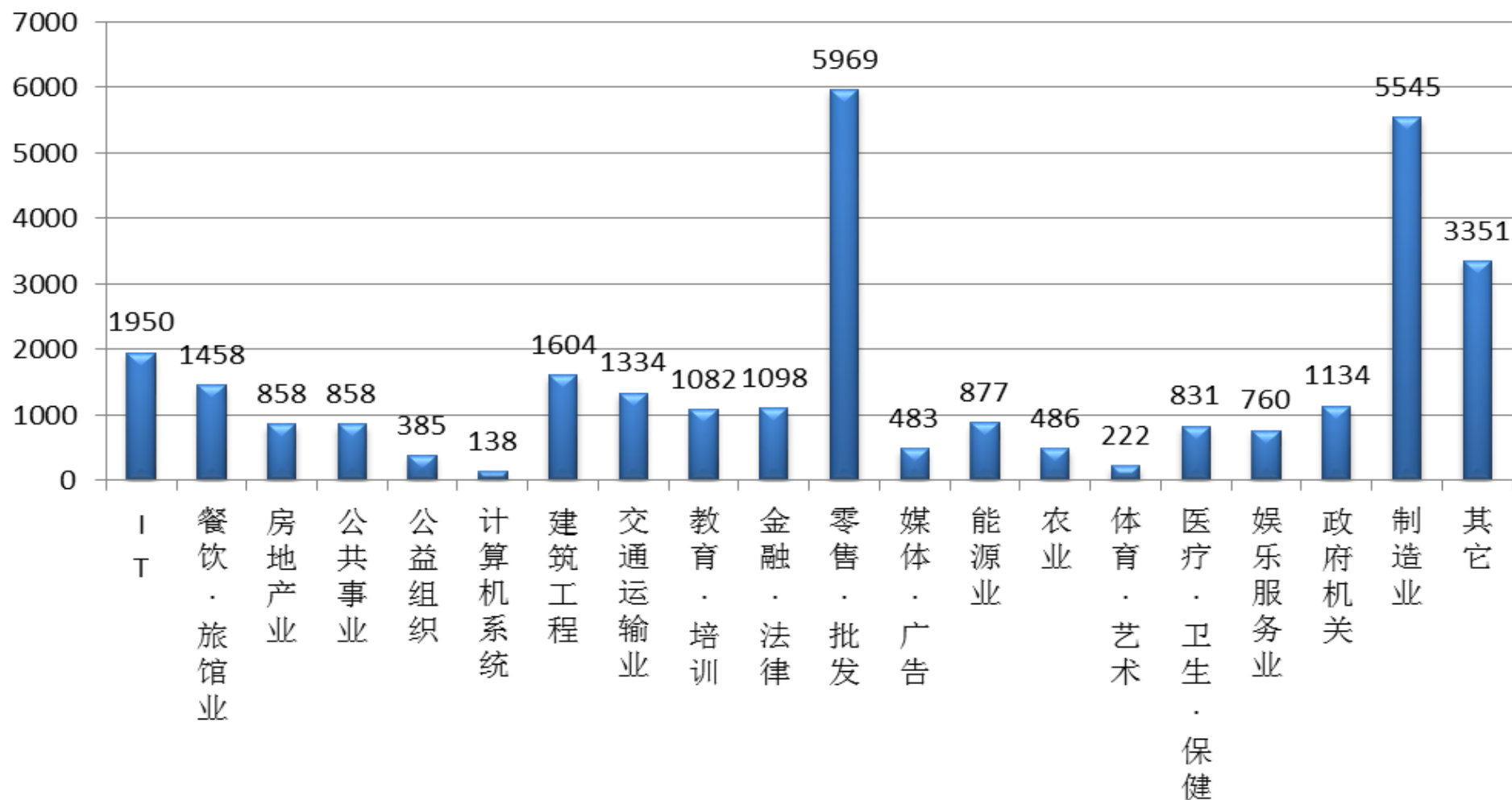
- 2、最高学历



借款用户中有12名借款用户未填写学历，高中或以下学历的用户有11064名，大专毕业的用户有13258名。说明大部分借款用户的学历集中在高中或以下只大专学历，本科及以上学历的用户相对来说较少。



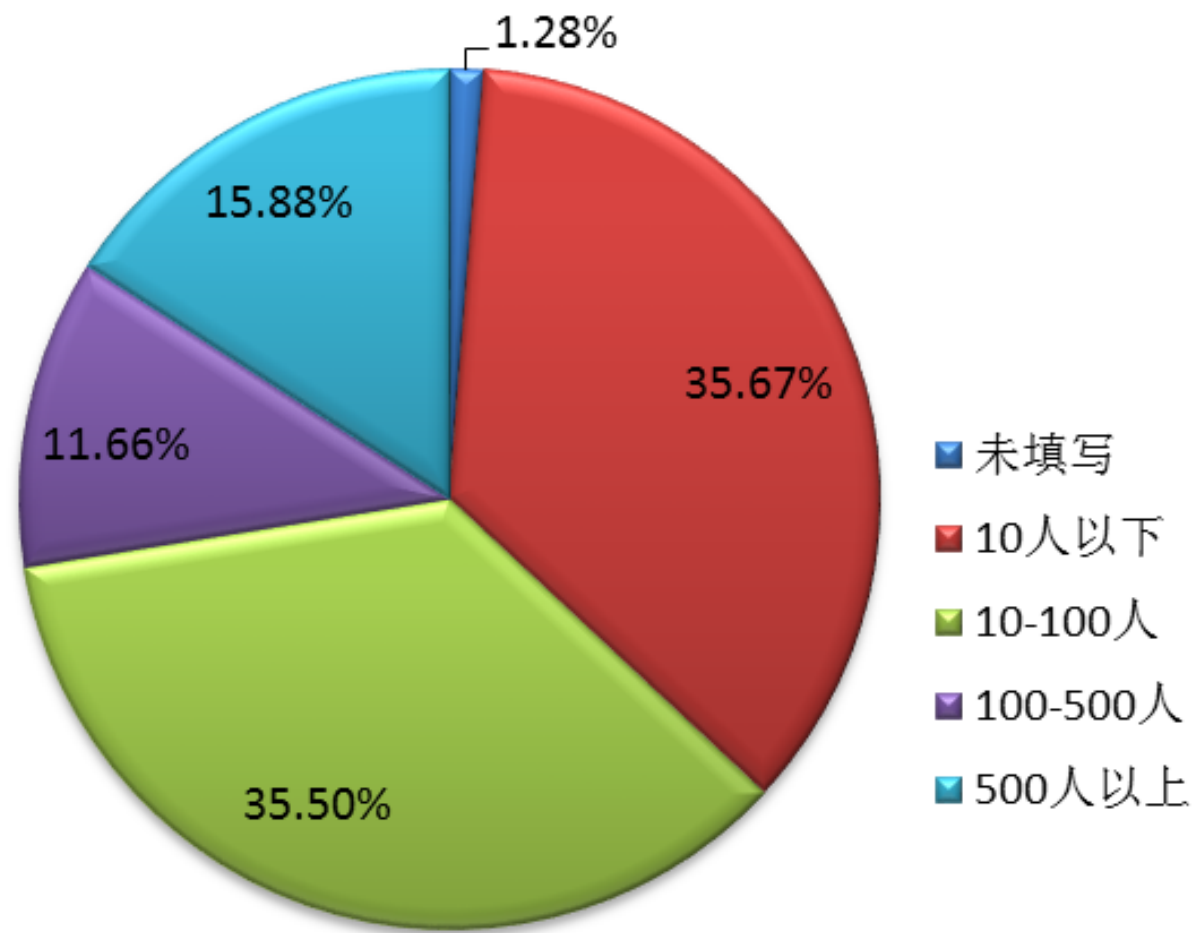
- 3、所在公司行业



可以看出，很大比例的借款用户都在从事零售/批发及制造业行业，共占37.36%，而从事公益组织、体育/艺术、计算机系统等行业用户鲜少在该平台发起借款。



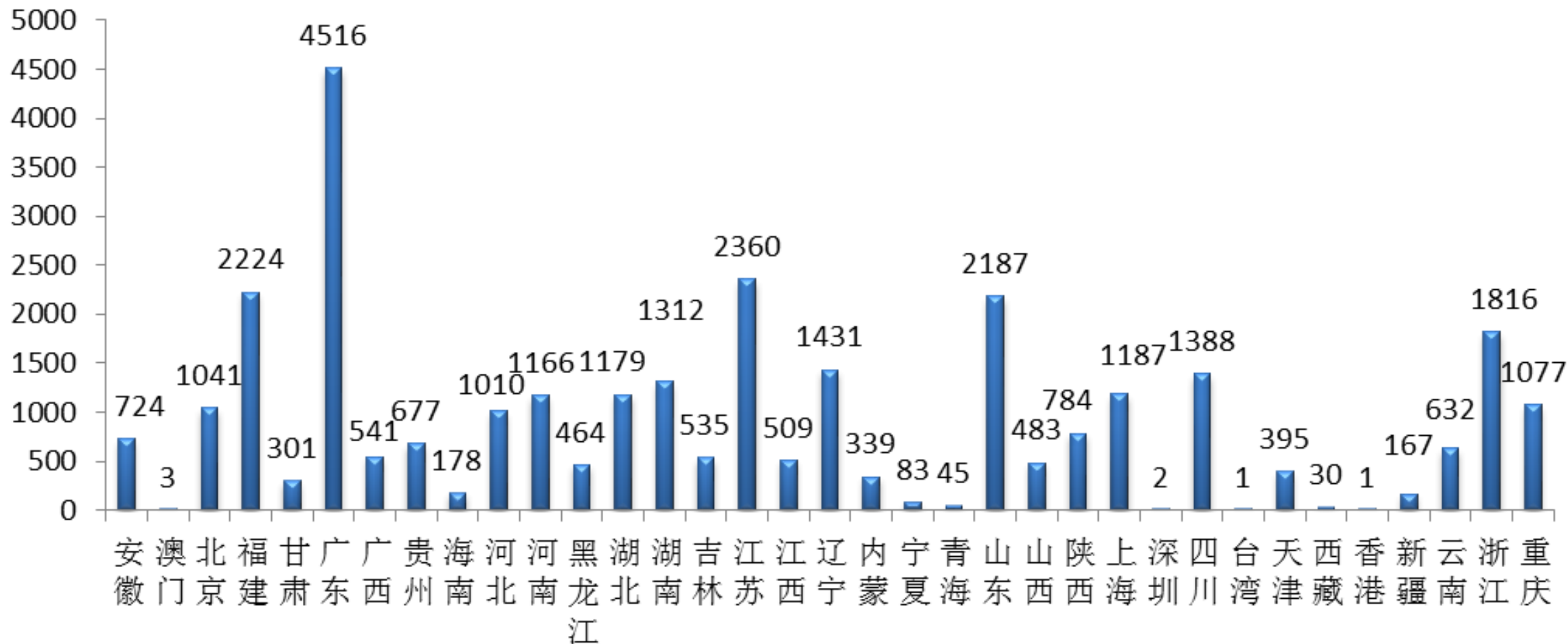
- 4、所在公司规模



样本所含用户中，有394人未填写，公司规模在10人以下的有10996名，在10-100人规模公司工作的有10943名，公司规模在100-500人的有3595名用户。



- 5、所在省份



深圳、台湾、西藏、香港、新建等地发起借款的用户相当少。



- 6、工作时间

参与工作时间	借款用户数量	所占百分比
未填写	14	0.04%
1年（含）以下	11590	33.70%
1-3年（含）	11321	32.92%
3-5年（含）	3570	10.38%
5年以上	4328	12.58%

- 工作时间在3年以下的用户发起借款较多。



- 6、收入范围

收入范围	借款用户数量	所占百分比
1000元以下	141	0.41%
1001-2000元	1379	4.01%
2000-5000元	12239	35.59%
5000-10000元	9523	27.69%
10000-20000元	3815	11.09%
20000-50000元	2340	6.80%
50000元以上	1386	4.03%

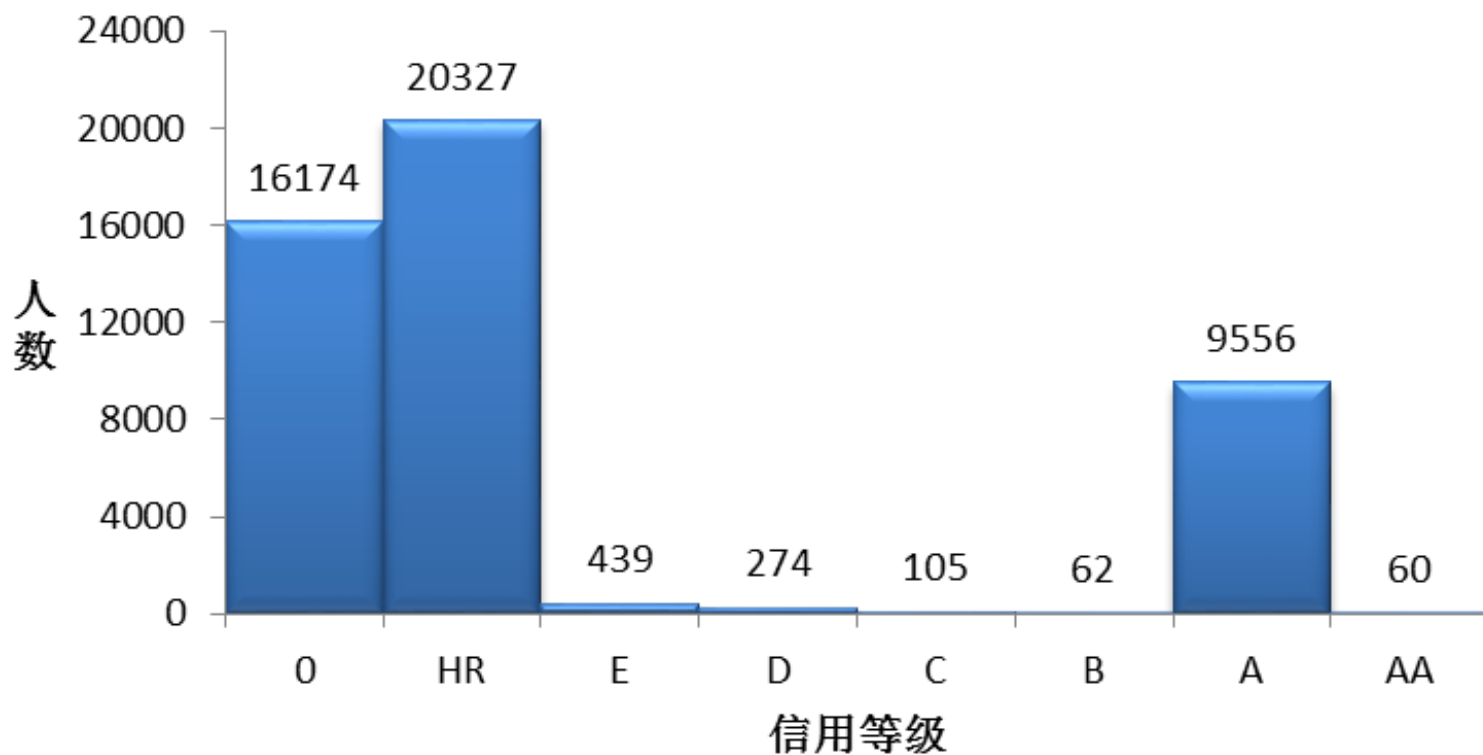
- 可以看出，收入范围在2000至20000的用户借款需求较大，而收入在50000以上的用户借款需求相对来说少许多。



- 7、信用分数
- 人人贷认证体系包括信用等级和信用额度。用户可以选择“我要借款”，根据自身情况，选择一种借款产品申请并提交相应材料，当用户的借款申请通过审核后即可获得信用等级及信用额度。此外，提交“可选信用认证”以及用户在人人贷网站上良好的借贷记录也可以提高信用分数。



- 7、信用分数



- 借款用户的
没有在平台经过可靠信用认证。这对与放款用户来说会降低对借款交易的信任，
影响融资获得率与成功率



五、融资可获得性模型的构建

- 通过样本我们可以发现，标的类型为信用认证标的借款，其保障方式仅为本金保障，因此放款用户会通过参考借款人的身份信息特征、借款的风险与收益性以及网站所提供的借款人信用评价等方面来做出是否出借资金的决策。



- 在选取解释变量时，首先根据所学理论挑选指标作为解释变量。借款的成功与否首先取决于放款人的信任。放款用户一般会将借款人的**个人基本情况、个人资产负债情况、历史借款还款逾期情况**以及**平台对借款人的信用评价**作为重要参考信息。
- 在认为出借资金的**回报损失不会太大**的情况下，放款用户会在同等情况下通过**比较借款信息**来选择借款交易。本文选取了借款金额、年利率、还款期限、提取还款费率、担保人是否赔偿以及借款用途作为借款信息的自变量代表。



- 选取的全部变量为：
- 借款人年龄 (Age)、学历 (Education)、婚姻状况 (Marriage)、收入范围 (Income)、所在行业 (Industry)、公司规模 (Company Scale)、工作时间 (Work Experience)、工作所在地 (Province)
- 是否有房产 (Households)、是否有房贷 (Mortgage)、是否有车产 (Car)、是否有车贷 (Auto Loan)
- 共申请借款笔数 (Application Time)、成功借款笔数 (Success Time)、还清次数 (Pay-off Time)、逾期次数 (Overdue Time)、严重逾期笔数 (Serious Overdue Time)、借款总额 (Success Amount)、逾期还款金额 (Overdue Amount)、待还本息 (Remain Debt)、平台给出的信用额度 (Credit Limit)、信用分数 (Credit Score)
- 借款交易中的借款金额 (Amount)、年利率 (Rate)、还款期限 (Loan Term)、提取还款费率 (Advance Rate)、担保人是否赔偿 (Warrantor)、借款用途 (Aim)



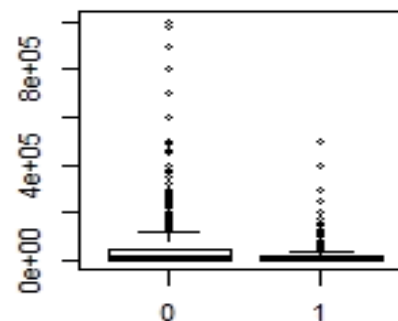
variable	chi-squared	df	p-value
advance.rate	1.3814	1	0.2399
warrantor	2249.498	1	2.20E-16
aim	150.2164	9	2.20E-16
education	473.573	4	2.20E-16
marriage	228.3698	3	2.20E-16
income	156.7462	6	2.20E-16
industry	293.6828	20	2.20E-16
company.scale	128.5291	4	2.20E-16
working.time	426.1538	4	2.20E-16
province	442.5915	34	2.20E-16
households	180.1289	1	2.20E-16
mortgage	145.6274	1	2.20E-16
car	287.7018	1	2.20E-16
auto.loan	57.6485	1	3.13E-14
application.time	3444.315	44	2.20E-16
success.time	15198.94	35	2.20E-16
pay.off.time	13080.98	34	2.20E-16
overdue.time	9330.927	33	2.20E-16
serious.overdue.time	3039.492	5	2.20E-16

- 对因变量与结构型变量制作列联表并进行卡方检验，除了提前还款费率与融资结果无显著相关，其他结构型变量均可能与融资结果存在一定关联。

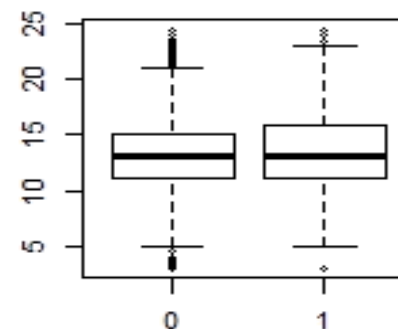


- 可以看出，成功融资的**借款金额、利率和还款期限**分布范围较**小**，多处于中等水平，借款人年龄较融资失败借款人来说更大一些，**信用额度和信用分数**普遍偏高，与常

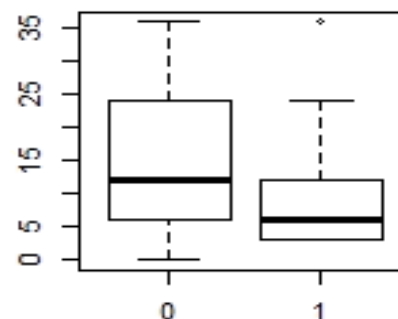
借款金额箱线图



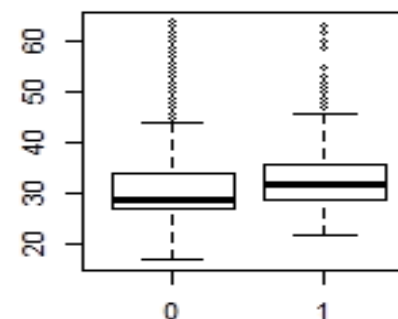
借款利率箱线图



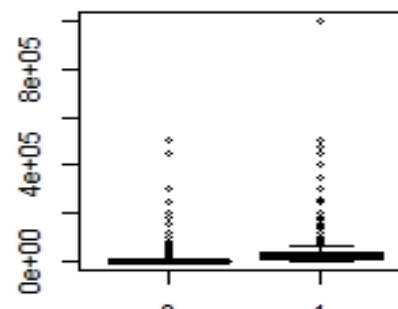
还款期限箱线图



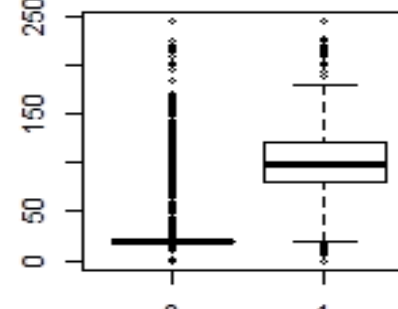
借款人年龄箱线图



借款人信用额度箱线图



借款人信用分数箱线图



- 由于借款成功与否的数据为0-1变量，而自变量中年龄等变量为连续变量，为使模型成立，对数据作Logit转换使因变量值转为鉴定借款融资成功的可能性。

$$Z = \left(b_0 + B_1^T P + B_2^T R + B_3^T L \right) + \varepsilon \quad Z = \ln \frac{P(Y_1 = 1)}{P(Y_1 = 0)}$$

- P、R、L分别代表个人信息变量（Personal）、历史信用变量（Record）、借款信息变量（Loan）的相应变量向量，而B1、B2、B3则分别代表相应的系数向量，b0为常数项， ε 为误残差项，代表除自变量外其它因素对Y1的影响以及实验数据误差。



空模型与全模型的方差分析表

	残差自由度	残差越轨	自由度	偏差
模型1：空模型	21144	11525.5		
模型2：全模型	21035	3460.2	110	8065.3

用卡方检验计算偏差的P值（ $1-\text{pchisq}(8065.3, 110)$ ）得到结果为0，可以看出，全模型与空模型有显著差别，全模型中有至少一个自变量对因变量融资成功率有一定的解释作用，即有一定的预测能力。



- 从Z检验结果可得，**借款金额、利率、还款期限及借款人历史信用信息**对融资成功率有显著的影响，一定程度上对融资成功率有解释作用，而其他变量中**借款用途、借款人收入范围、所在行业、公司规模**的部分取值对因变量有一定解释力，剩余变量则不具有显著的解释性。
- 在0.1水平下，剔除上述变量不显著的变量，重新建模得到结果



回归系数	参数估计值	参数标准差	z检验值	p值	显著性
截距	-3.44E+00	2.64E-01	-13.043	<2e-16	***
借款金额	-2.15E-05	1.76E-06	-12.242	<2e-16	***
借款利率	3.44E-02	9.38E-03	3.665	0.000247	***
还款期限	-1.14E-02	5.47E-03	-2.073	0.03818	*
aim.个人消费	-3.56E-01	1.02E-01	-3.499	0.000468	***
aim.装修借款	9.66E-01	1.32E-01	7.315	2.57E-13	***
aim.购车借款	6.30E-01	2.17E-01	2.898	0.00375	**
income.20000.50000元	5.09E-01	1.78E-01	2.857	0.00428	**
income.50000元以上	1.83E+00	1.85E-01	9.889	<2e-16	***
industry.金融.法律	-7.97E-01	2.09E-01	-3.81	0.000139	***
industry.教育.培训	2.89E-01	1.69E-01	1.711	0.087158	.
industry.零售.批发	2.53E-01	1.26E-01	2.012	0.044246	*
industry.医疗.卫生.保健	5.51E-01	2.01E-01	2.744	0.006067	**
companyscale.100.500人	-1.10E+00	2.37E-01	-4.657	3.21E-06	***
companyscale.10.100人	-1.01E+00	2.24E-01	-4.516	6.30E-06	***
companyscale.10人以下	-1.02E+00	2.38E-01	-4.289	1.79E-05	***
companyscale.500人以上	-8.71E-01	2.30E-01	-3.789	0.000151	***
worktime.1年.含.以下	-4.71E-01	1.26E-01	-3.727	0.000194	***
借款申请次数	-4.14E-01	2.00E-02	-20.686	<2e-16	***
借款成功次数	4.06E-01	2.41E-02	16.846	<2e-16	***
逾期次数	-3.24E-01	1.18E-02	27.392	<2e-16	***
信用分数	0.04412	0.001036	42.592	<2e-16	***



模型比较

- 评判标准为：
- True Positive Rate (TPR) : 把真实的融资成功案例正确预测为 $\text{status}=1$ 的概率
- False Positive Rate (FPR) : 把真实的融资失败案例错误地预测为 $\text{status}=1$ 的概率



模型比较

融资成功情况logit模型预测			
真实借款 融资情况		失败	成功
	失败	4810	83
	成功	136	257

TPR= 0.654， FPR= 0.01696。



六、对P2P网络借贷用户及平台的建议

- （一）针对P2P网络借贷平台
- 结合本文的研究，首先针对P2P网络借贷平台而言，本文的建议如下：
- 在借款人发布借款标的之前，可以应用模型预测融资可获得性的概率，以提高平台运营效率：先对借款交易能否被投标，被投标概率为多少进行预测。对于预测成功可能性非常低的借款列表，根据其对应的参数大小，建议借款用户通过增加认证或上传更多资料等方式，提升借款成功率，以避免流标浪费系统资源以及借款人和放款用户的时间。



- （二）针对P2P网络借贷平台用户
- 针对P2P网络借贷平台的用户而言，根据本文分析给出如下建议：
- 从样本分析以及融资获得性模型可以看出，借款交易中借款信息、个人特征信息对于放款用户的决策非常重要。在P2P网络借贷平台中，经常会出现一些借款交易，作为发起者的借款用户上传的资料极少，也极少进行认证，这种借款交易极易流标。而这种流标一方面会对新的潜在借款人造成不能借到款的负面影响，同时也浪费了极大的降低了P2P网络借贷的运营效率。
- 因此，如实填写并完善借款交易中每一项信息并在P2P借贷平台进行各项信用认证可以大幅度提高放款用户对借款交易的信任度，从而可能提高融资获得率。





谢谢