

For this case study, you will use Python to **solve a real interview case**. Because of the confidentiality, the company's name has been removed.  
Please read through the instruction as below:

---

## Machine Learning CoE: Case Study

### Assignment:

The assignment is to investigate and analyze the attached data, producing a model that can be used to predict whether a loan will be paid in full or charged off (this occurs when the loan is unpaid or delinquent for a period of time so the financial institution determines that the debt is unlikely to be collected). You will use the `loan_status` column in the dataset for the outcome, and you can use the remaining variables in the dataset to predict the loan status (although beware of variables that may contain data leakage).

We prefer that you complete this case study in R or Python, but if you are not comfortable in either of those then please feel free to use whichever tool/language you have the most fluency in.

### Deliverables:

1. The code you used to perform the work (please use comments to describe your choices/steps).

[REDACTED]

### Expectations:

We are looking to understand your familiarity with data preparation and model building. We would also like to see your comfort level with creating business value from analytics and model results.

We are not expecting you to work on this task for more than 8 hours. If you find that you are spending too much time on the data preparation/model building section, please feel free to explain in comments any additional steps that you would like to take if this project were assigned to you in the role of a Data Scientist [REDACTED]

Other useful reading materials:

<https://www.lendingclub.com/public/how-peer-lending-works.action>

<https://zhuanlan.zhihu.com/p/54193765>

## **Lending Club 的深度报告及其商业模式**

Lending Club 是全球最大的撮合借款人和投资人的线上金融平台，它利用互联网模式建立了一种比传统银行系统更有效率的、能够在借款人和投资人之间自由配置资本的机制。

通过 Lending Club 的平台，借款人和小微企业可以获得更低的利率，投资人则可以获得较好的收益。

公司于 2006 年 10 月在特拉华州以“SocBank”的名字成立，之后于 2006 年 11 月更名为 Lending Club。

截至 2014 年 6 月 30 日，Lending Club 累计贷款金额超过 50 亿美元，支付给投资人的利息超过 4.9 亿美元。

自 2007 年成立以来，Lending Club 每年的贷款金额的增长率都在 100% 以上，贷款数量的增长率（除 2011 年为 70% 以外）均大于 100%。

Lending Club 自 2007 年 5 月 24 日正式上线以来共经历了三个阶段：

- 1) 2007 年 5 月 – 2008 年 4 月的运营初期；
- 2) 2008 年 4 月 – 2008 年 10 月的“静默期”；
- 3) 2008 年 10 月至今的新模式运营期。

### **运营初期**

在运营初期，即 2007 年 5 月 – 2008 年 4 月，Lending Club 不发行凭证，而是由贷款人成员直接购买成员贷款，贷款人收到对应其所买贷款金额的匿名个人凭证，赋予 Lending Club 为成员贷款提供服务的权利。

2007 年 5 月 24 日 Lending Club 在 Facebook 上线“合作性的 P2P 贷款服务”，该服务提供给 Facebook 用户一条不需要银行参与、借贷双方直接联系、拥有更优惠利率的贷款渠道。

Facebook 用户在 Lending Club 上注册，注册信息就会显示在用户的 Facebook 页面上，访问者知道他是 Lending Club 的成员，并且有资金出借（或借入）的需求。

Lending Club 负责用户身份认证，银行账户认证，信用检查，信用报告以及资金的转移与收回。通过 Lending Club，借款人可以向一个或多个贷款人申请 \$1,000 – \$25,000 的个人贷款。

基于亲和力准则，Lending Club 开发了一套搜索和识别贷款机会的 LendingMatch 系统。

该系统能够帮助用户迅速发现之前不知道的关系，使他们在有信任基础的人群中借贷资金。

用户通过设定的标准（比如区域、工作地点、教育情况、专业背景、社会关系等）识别贷款，进而构建一个符合个人风险偏好的分散化的贷款组合（包括 10 – 30 个借款人）。

LendingMatch 系统包含了多种关系，包括 47,000 个美国城市；1,000 家公司；500 个政府机构；6,300 所大专院校；1,600 个国家组织和其他组织以及数以千计的 Facebook 关系群和网络。

为了鼓励负责任的借款行为，Lending Club 只供信用分数（FICO 分数）在 640 分以上的借款人申请贷款。

Lending Club 分别向借款人收取一次性的贷款金额 0.75% – 2% 的服务费和向贷款人收取分期还款数额 1% 的服务费。

Lending Club 选择在 Facebook 平台上线产生了互惠互利的效果。

一方面，LendingClub 为 Facebook 用户提供了简单便捷的融资和投资渠道；

另一方面，Facebook 为 Lending Club 未来业务的有效开展提供了有利条件：

- Lending Club 能够获得 Facebook 用户的朋友圈数据，产生大量的潜在客户群；
- Facebook 为 Lending Club 提供用户身份认证的依据；
- Facebook 社交平台营造了借贷双方并非完全陌生人的环境，更有利于 P2P 贷款业务的开展。

自 2007 年 5 月 24 日到 2007 年 8 月 23 日，共有 13000 个 Facebook 用户登录 Lending Club，获得 75 万美元的贷款。

2007 年 9 月 13 日 Lending Club 宣布开始使用自己的网站 LendingClub.com 提供服务，并在网站上提供 LendingMatch 系统。

Lending Club 网站还开办论坛以供金融专家与 Lending Club 的成员分享经验。Lending Club 同时拥有自身的网站和 Facebook 两个平台提供服务。

除了利用 Facebook 的同学、同事等社会关系借贷外，Lending Club 的另一个举措是利用校友关系借贷。

2007 年 10 月 31 日，Lending Club 宣布推出“基于熟人关系的借贷服务”，该服务是通过高等院校校友组织的网站来提供资金的借贷。

Lending Club 是 P2P 贷款领域第一个向线上社区（包括乔治亚科技大学、堪萨斯州立大学和密歇根大学的校友组织）提供联合品牌贷款申请的公司。

这项服务能够通过借贷双方的信任关系有效地降低借贷成本，同时也实现了 Lending Club 与知名大学的合作。

## 静默期

2008 年伴随着金融危机的发生，SEC 也加强了对 P2P 贷款公司的监管，收益权凭证被 SEC 定义为证券，Lending Club 需要对收益权凭证注册后，才能向贷款人出售。

自 2008 年 4 月开始，Lending Club 不提供给投资者购买收益权凭证的服务，同时也不接受贷款人成员的注册和已存在贷款人成员的提供资金的合约。

但是 Lending Club 仍将继续为之前筹集的贷款提供服务，贷款人成员仍可以查询账户，管理之前的贷款和撤出可用的资金。

2008 年 6 月 20 日，Lending Club 向 SEC 提交 S-1 表格，申请发行 6 亿美元的会员支付凭证。

2008 年 8 月 1 日，Lending Club 针对 S-1 表格提出了修正案，提出新的利率计算方法以及关于“转售交易系统”的更多细节。

2008 年 10 月 14 日，Lending Club 完成注册程序，并公布了募集说明书，恢复业务。

在 Lending Club 业务暂停的 6 个月里，借款方仍旧可以申请成员贷款，只是资金完全由 Lending Club 自身提供，因此借款方并没有受到影响。

## 新模式运营期

自 2008 年 10 月 14 日 Lending Club 成功向 SEC 注册后，Lending Club 就开始通过发行会员支付凭证来向借款人贷款。

Lending Club 通过与美国联邦存款保险公司（FDIC）担保的犹他州特许银行 WebBank 合作，由 WebBank 向通过审核的借款用户放贷，WebBank 再将贷款以凭证形式卖给 Lending Club，并获得由 Lending Club 发行会员支付凭证的资金。

投资人和借款人没有直接的债务债权关系，Lending Club 和 WebBank 也不承担贷款违约的风险。借款人每月通过银行账户自动将还款转给 Lending Club，再由 Lending Club 转给投资人。

投资人的每笔投资不得低于 25 美元。投资人需要满足一定的条件才能投资：

1、年收入在 7 万美元以上并且净资产（不包括房产和汽车）在 7 万美元以上；

或

2、净资产（不包括房产和汽车）在 25 万美元以上。

加利福尼亚州的投资者需要具备的条件：

1、年收入在 8.5 万美元以上并且净资产（不包括房产和汽车）在 8.5 万美元以上；

或

2、净资产（不包括房产和汽车）在 20 万美元以上；不满足条件者只能做 2,500 美元以下的投资。肯塔基州的投资者需要满足证券法关于合格投资人的条件。

## 产品及服务

自成立到 2014 年 2 月，Lending Club 只提供个人信贷服务，2014 年 3 月开始上线商业信贷服务，2014 年 4 月公司收购 Springstone 公司，开始提供教育和医疗信贷服务。

更多金融业资讯，案例、法规、分析、创新、PE/VC、IPO、三板、并购重组、银行、证券、信托、债市、项目融资、资产管理、资产证券化、风险管控、基金管理、融资租赁、小贷、典当、担保、保险、互联网金融、行业知识等，学习、交流，尽在《中国资本联盟》平台微信公众平台号：CACNORG。让我们以诚携手共进，缔造中国金融未来。（投稿、商务合作、发送邮箱：cacnorg163@.com）

（1）个人信贷 Lending Club 提供额度从 1,000 美元 – 35,000 美元的个人无抵押贷款，1,000 美元 9,975 美元的贷款只能申请 3 年的期限。

初始到期日为三年的贷款如果不能按时还款，可以将最终到期日扩展到 5 年；初始到期日为五年的贷款，最终到期日仍为五年且不可扩展。

个人贷款分为四类：合并债务贷款、信用卡支付贷款、家庭改善贷款和游泳池贷款。

其中，游泳池贷款是 Lending Club 的一种创新。

2011 年 5 月 2 日 Lending Club 宣布 Pool orporation 将通过 Lending Club 在个人信贷方面投资 200 万美元，用以刺激游泳池和 Spa 的销售。

Lending Club 应用创新的理念成为上市公司为其潜在客户融资的平台。

（2）商业信贷商业贷款的额度为 15,000 美元-100,000 美元，期限为 1-5 年。

（3）教育和医疗信贷服务 Lending Club 目前只能通过 Springstone 公司提供教育和医疗信贷，但公司计划未来能够直接通过 Lending Club 平台提供教育和医疗融资产品。目前教育和医疗融资额度为 499 美元-40,000 美元。

除此之外，Lending Club 的子公司 LC Advisors，专门为合格投资人提供投资管理服务。LC Advisors 作为普通合伙人成立了五个投资基金，同时成立信托来管理这些基金。

这些基金通过不同的投资策略来购买 Lending Club 发行的凭证。据了解，这些基金最多只能有 100 位投资人，而且每人的最低出资额为 50 万美元。

除了成立了投资基金，LCA 还设立了分离管理账户（SMAs）来增加合格投资人的投资途径。使用分离管理账户的投资者对投资策略有更强的控制力。

LCA 还为通过第三方机构在 Lending Club 平台上管理投资的合格投资人提供服务。

## 盈利模式

Lending Club 的收入来源为交易手续费、服务费和管理费。

交易手续费是向借款人收取的费用；服务费是向投资者收取的费用；管理费是管理投资基金和其他管理账户的费用。

交易手续费是 Lending Club 收入的主要来源。根据招股说明书的数据，2014 年上半年交易手续费收入占总收入的 93%。

除手续费外，借款人还可能需支付以下费用：

1、未成功支付的费用：借款人每月的还款都是通过借款人账户向 Lending Club 自动转账，如果借款人账户未能支付（账户余额不足，账户关闭或者账户暂停），借款人需支付每次 15 美元的费用；

2、逾期还款的费用：如果借款人还款晚于 15 天的宽限期，从第 16 天开始需支付未付分期还款金额的 5% 和 15 美元之间的高者，每次逾期支付一次；

3、处理支票的费用：如果借款人选择使用支票付款，需支付每笔 15 美元的费用。

服务费率是借款人还款金额的 1%，用于补偿 Lending Club 对贷款的服务、对凭证的支付和对投资人账户的维护。

除服务费外，投资人还可能需支付催款费。催款费按收到金额的百分比支付：

1、逾期超过 60 天且短于 90 天，按 30% 的比例收取；

2、其他情况（诉讼除外）按 35% 的比例收取；

3、在诉讼的情况下，按 30% 的比例或者诉讼律师费加上其他费用；

4、未催收成功不收费，催收费不超过收回金额。

管理费率为每年 LCA 管理资产的 0.7%-1.25%。

## 信用评级及利率计算

借款人提交申请后，Lending Club 会根据 WebBank 的贷款标准进行初步审查。贷款人需满足以下标准才能借款：

1、FICO 分数（根据 Experian、TransUnion、Equifax 三大征信局信用报告）在 660 分以上；

2、债务收入比例低于 40%；

3、信用报告反应以下情况：至少有两个循环账户正在使用，最近 6 个月不超过 5 次被调查，至少 36 个月的信用记录。

当借款人通过了初始信用审查，Lending Club 的专有评分模型会对申请人进行评估，并决定同意或拒绝申请。

初始评分模型提供给申请人一个分数，再连同申请人 FICO 的分数和其他信用属性一起输入 Model Rank 中。

Model Rank 根据内部开发的算法来分析借款人成员的表现，并考虑申请人的 FICO 分数，信用属性和其他申请数据。

Model Rank 介于 1 – 25 之间并对应一个基础风险子级。

Lending Club 再根据贷款金额和贷款期限调整基础风险子级得到最终子级。

最终子级共分为 35 个级别，归入 A、B、C、D、E、F、G7 个等级，每个等级又包含了 1、2、3、4、5 五个子级。

Lending Club 设定的利率考虑了信用风险和市场情况，每个等级最终的贷款利率由以下公式计算：

贷款利率 = 基准利率 + 风险和波动率调整

风险和波动率调整涵盖了期望损失并为从 A1 到 G5 各个等级的贷款提供递增的风险调整收益。

贷款等级的确定不只考虑了借款人的信用分数，还包含了借款人信用报告和贷款申请中的信用风险指标。

在 Lending Club 七个等级的贷款中，A、B 两类贷款占比较大，C、D 两类贷款的占比不断增多，E、F 和 G 三类贷款的占比最小，且呈不断缩小的趋势。

截至 2014 年 6 月底，A、B、C 和 D 四类等级贷款的总占比为 87%，所有贷款平均的利率为 14.66%。

根据 Lending Club 对平台上借款人的问卷调查，2014 年前七个月借款人原有债务的平均利率为 21.7%，而从 Lending Club 借款的平均利率为 14.8%。

### 各类等级贷款的占比及平均利率

基于 Lending Club 多层严谨的信用风险审核体系，截至 2012 年底只有 10.7%的借款人通过审核可以借款，贷款违约率也保持在 3.5%左右的较低水平。

截至 2014 年 6 月 30 日，Lending Club 的借款人平均 FICO 分数为 700，平均债务 – 收入比率为 16.7%，平均 15.7 年的信用记录，借款人平均年收入为 72,751 美元，而平均贷款额为 14,056 美元。

E、F 和 G 等级贷款占比较小，而且有逐渐减少的趋势。

### 全球 P2P 龙头 Lending Club 的商业模式

随着 2014 年福布斯排名第五的最有潜力公司 Lending Club 的上市，整个世界会开始关注“这个新类别的资产”。Lending Club 以债权形式交易的证券是有别于股票、各种优先、劣后的企业债、可转债等常见金融产品，是另一种资产类型——个人债。如果 Lending Club 上市成功，或许意味着 P2P 债权至少在美国完全有可能成为一种新的主流金融产品。

过去一年所谓的 P2P (peer-to-peer) 人人贷业务出现了爆炸式的发展，连前一段《经济学人》杂志都发表了一个专题报告 Banking without Banks。P2P 是移动互联网对传统行业改造下的产物。简单而言，P2P 就是通过网上竞拍直接让借方和贷方相匹配。虽然国内 P2P 业务还处在野蛮生长的阶段（我曾经听说朋友隔壁的 P2P 公司瞬间倒闭），但在海外已经有了比较成熟的商业模式。

P2P 的借贷在各国成长迅速。在英国，贷出量每六个月翻倍一次。它们刚刚跨过 10 亿英镑（17 亿美元）大关，虽然与这个国家 1.2 万亿英镑的小额存款比起来这还是很少。在美国，两个最大的点对点贷方，借贷俱乐部(Lending Club)和普洛斯普(Prosper)拥有 98%的市场份额。它们在 2013 年发出了 24 亿美元的贷款，从 2012 年的 8.71 亿美元增长起来的。小公司们做的更好，虽然他们由更低基础发展起来。

记得去年在加州的时候看到许多 Lending Club 的网店，如同小银行一样散落在旧金山。去年 12 月初，LendingClub 的贷款达到 30 亿美元，并不断推出如“Lending Robot”的新服务，表现出稳健的发展趋势。公司在去年完成了 1.25 亿融资后估值已经达到 15 亿美元，并且将在今年上市。如果翻开 Lending Club 的董事会名单时，你会发现互联网女皇 Mary Meeker，摩根斯坦利前 CEO John Mack，以及美国前财政部长 Larry Summers 都在里面。

**美国的个人信贷市场及 Prosper 和 Lending Club 的运营模式有以下几个特点：**

- 1) 美国利率市场化已经完成，这点和中国利率市场化还没有完成却大力发展 P2P 的背景完全不一样。在美国，信用卡的贷款年化利率都高达 18%，个人的无抵押借贷更是在 15-25%之间。而 Lending Club 这种 P2P 平台上的利率只有 11%；
- 2) 美国和中国另一个最大的不同是有完整的信用体系 FICO 评分。这种针对小企业和个人的风险定价已经很完善。这也是为什么 08 年信贷危机时，许多美国人在房子低于水位价格之后，依然没有短供的原因。一旦你的信用评分下来了，你下辈子可能再也无法贷款，申请信用卡了。同样，完整的信用体系让 Lending Club 很容易区分各种风险。而他需要做的仅仅是通过互联网下沉到各个地区。Lending Club 2011 年的综合违约率是 3.9%，而出借人的综合收益是 7.83%；
- 3) 美国是过度消费而又金融发达的国家。过度消费意味着居民没有什么钱用来在 P2P 平台上投资，金融发达意味着美国居民有相当多的投资渠道，P2P 仅是渠道之一。这点又和中国不同。许多中国人把 P2P 平台当做了自己主要理财的渠道，占据家庭金融资产很大比重，而美国人更多只是在 Lending Club 上“玩一玩”。Lending Club 的平均贷款额为 5500 美元。最低 1000 美元，最高 2.5 万美元。

据 Lending Club 创始人兼 CEO Renaud laplanche 所说，他们现有针对个人的风控技术和信贷产品可以很轻易地移植到车贷、企业经营贷、住房抵押贷款等其他贷款领域。而他们的终极目标，则是“覆盖全部信贷产业”。有趣的是，Lending Club 上市的主承销商是摩根大通，它给 Lending Club 定的上市“股票故事”为“A New Asset Class”——一个新的资产类型。即 Lending Club 以债权形式交易的证券是有别于股票、各种优先、劣后的企业债、可转债等常见金融产品之外的另一种资产类型——个人债。如果 Lending Club 上市成功，或许意味着 P2P 债权至少在美国完全有可能成为一种新的主流金融产品。

今天，虽然 P2P 业务还有各种问题，特别是缺少监管，欧洲政府甚至没有准备好接受这个新事物。但我们不得不承认，银行业也会被互联网所改造。十年之后，我们会惊讶过去十年互联网对我们生活的颠覆。而随着 2014 年福布斯排名第五的最有潜力公司 Lending Club 的上市，整个世界会开始关注“这个新类别的资产”。