A Comparative Analysis Of Credit Rating Model Used In Peer to Peer Lending

Rik Choudhury

IUBH, Berlin Campus

Under advisement of Prof. Kunal Saigal

Abstract

The aim of the thesis is to study the credit rating system used in P2P lending platforms like Zopa. This study specifically investigates the factors which makes a loan to be successful or default. It examines the critical factors in assessing credit quality and probability of default from the perspective of the investor. This study will explore the datasets of two large peer to peer platforms, namely Zopa and RateSetter. This study shows a correlation matrix to see which factors play an important role in determining the nature of the loan. A logistic regression model is used to see the variables which effect the status of the loan and by how much. It will take a dynamic review of the current developments of P2P lending based on previous literature.

Keywords: credit rating, peer to peer lending (P2P), correlations, regression

Table of Contents

Definitions

SME	Small and Medium Enterprise
IFISA	Innovation Finance Individual Savings Account.
FSCS	Financial Service Compensation Scheme.
FCA	Financial Conduct Authority.

Introduction

P2P refers to peer-to-peer or peer -to-person. Peer to Peer lending companies provides a platform which act as an intermediary to match lenders and borrowers to transact. P2P platforms reduce overhead and infrastructure cost compared to that of traditional banks. Financial transactions which bypass conventional intermediaries by directly connecting the borrower to the lender classify as P2P lending[1]. They can offer low interest rate for the borrowers and higher returns for investors.

Borrowers provide information such as occupation, financial leverage, loan purpose, collateral, and loan history which are used for the qualitative evaluation by the P2P platform. Quantitative attributes such as loan amount, credit scores, debt-to-income ratio, monthly income are used to classify borrowers into various risk groups. Investors can browse this information to decide whether to invest and the amount for investment.

It cuts the middlemen, a traditional financial institution, and connects multiple investors and potential borrowers to invest capital and to borrow credit. For the investors, this market provides the opportunities to receive a greater return on their capital as compared to depositing it in a savings account in a traditional bank and mitigating the risk. However, there is a great risk of borrower defaulting and not repaying the loan with interest. Therefore, there is a need to identify which characteristics make a borrower’s loan bad and enables investors to make better decisions.

The largest players globally in the peer-to-peer lending market are Lending Club, Prosper, Zopa, Funding Circle based on the amount of loans generated. The rapid growth in P2P lending, along with an apparent investor appetite to provide them with equity funding and use them to channel funds directly into consumer and SME lending, has led some to predict profound disruption of the traditional banking model.

The emergence of more P2P lending platform makes the marketplace for such financial practices stronger. With the IPO of Lending Club[2] in 2014, has been witness to the amount of growth in P2P lending. This growth has forced traditional institutions such as banks to take notice and develop relationships with the industry. Hedge funds show interest in purchasing loan

---

[1] Peer-to-peer lending wiki
[2] Lending Club wiki

products from P2P lending platform. The interest is due to attractive returns as well as stable products offered by P2P lending platforms. There is also a trend of financial institutions to open their own department focusing on P2P lending to be with the times. Consolidation of credit is one of the main reasons for borrowing loans from P2P platforms.

Asymmetric information is found everywhere in P2P lending. The P2P lending market lacks historical data. Therefore, chances of investors being more prone to borrower's default is quite high. P2P lending is a two-sided market. In order to further boost market growth, P2P lending platforms also need to enhance the ability of investors to assess credit risks. By doing this, platforms can offer higher return, and thus, attract more participation of investors.
in lending activity.

This study attempts to evaluate the critical risk factors which could help the investor to assess credit risk of their investments. To achieve the objective, data from prominent P2P lending platforms (publicly available) is utilized. The rationale of the study is that the investor would properly vet the credit risk of the potential borrower therefore, a model to calculate will aid in loan selection. This model will act to mitigate information asymmetry on P2P lending and gaming philosophy of borrowers. Besides, this study will also take a review of the current development of P2P lending built on previous literature.

It has been almost 15 years since the first P2P lending platform was founded in the UK. While P2P lending has been growing rapidly within the past 15 years, it is still in the infant stage compared to the traditional banking industry. There are over 100 academic papers about P2P lending between 2008 and 2018, but from different perspectives, including analyses of determinants of a loan to be successfully funded by investors, regulations, credit risks, determinants of credit quality and default probability, business model of P2P lending across countries, internal information system and literature reviews.

Even though a handful of papers did research on credit risks using data mining methodologies, most of them were focused on explaining the determinants of a loan being successfully funded.

It is seen in recent years, there has been a scene in peer-to-peer lending. The purpose of these alternative financing seems to work well. Many investors find a new source to make profit and borrowers are managing to get loans easily. This is a legitimate area of investing which is being made secured through providing credit scores as well as information on borrowers but there exists a risk of losing money.

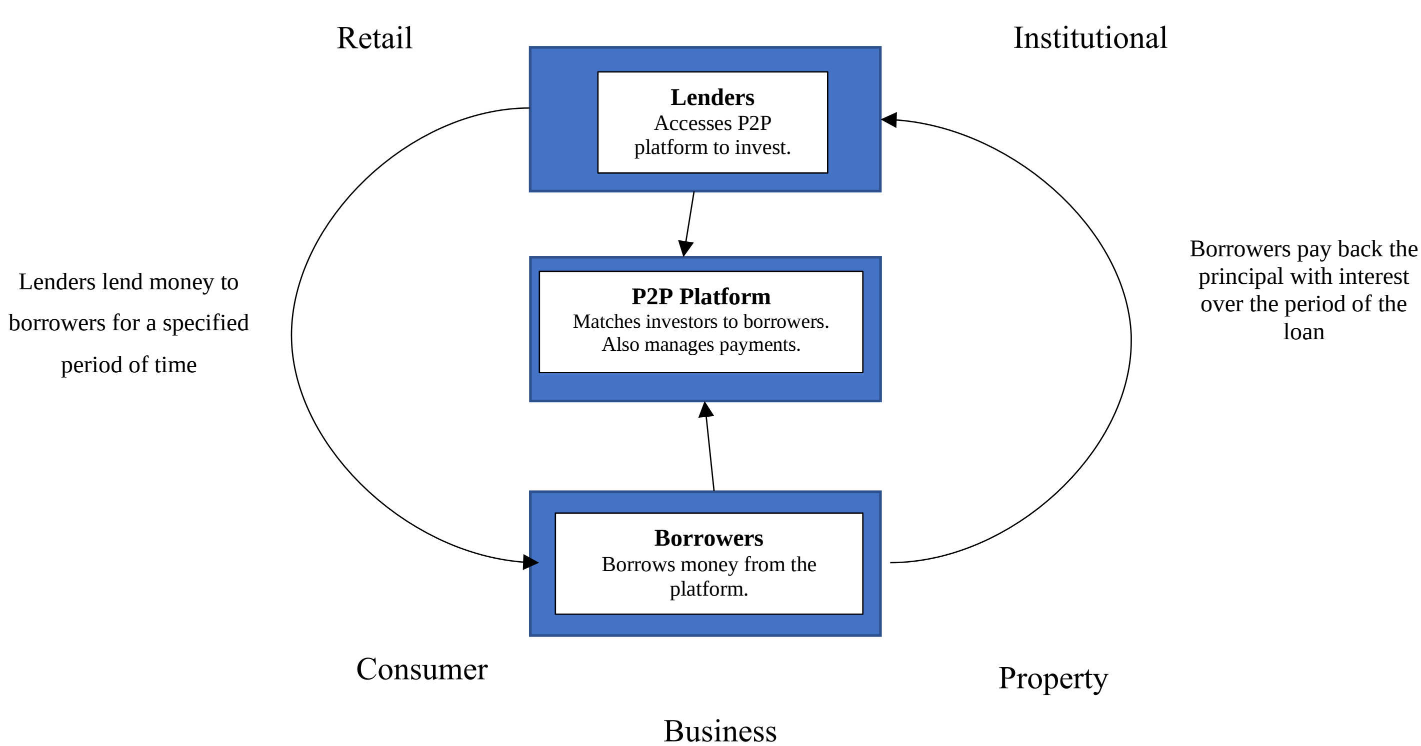
The study is organized as follows. Section 2 presents process of peer-to-peer lending. Section 3 presents the market view of the peer-to-peer lending market. Section 4 presents the literature review which give a picture of the previous research done on the subject. Section 5 presents the loan book from Zopa and RateSetter. It also provides the analysis of the data. Section 6 compares the features between Zopa and RateSetter. Section 7 concludes.

Process

P2P lending platform acquires borrowers by asking them for their private information. Information such as loan amount, credit rating, monthly income, occupation, purpose etc. are provided by the borrowers to the platform. The assessment of the documents in done by the platform. A fixed interest rate is decided upon for the loan. Then the listing goes on to the platform for the investors to watch. The investors can access all the loan details and decide on whether to invest on a loan or not.

P2P platforms allow lending to borrowers without any prior knowledge. Zopa platform is the oldest and credit providing company in UK since 2005.

Figure 1:Peer-to-Peer Process



Borrowers:

Consumer lending: The loans are granted for a variety of purposes like car financing, home renovations, debt consolidation like credit card bills and weddings. Typical loan size is between 260 to 35000 pounds. Loan origination is one of the main issues in this sector. High quality borrowers are always sought after by platforms to provide loans.

Business lending: This type of lending accounts for 35% of overall P2P market[3]. It usually lends to manufacturing, engineering, transport, utilities, finance and retail. Loan size can be large, for example, Funding Circle[4], a P2P platform started in UK in 2010, funds around 5000 up to 350000 pounds to businesses.

Property lending: Products such as long term commercial and residential mortgages, lending for construction and development are included in this type of lending. Companies like LendInvest[5] provide a range of financial models and products.

Lenders:

Retail: Big companies and institutional normally do not get involved in emerging market segments like the P2P lending market. Large institutions can lend large amounts of money, which in turn could lead to a large proportional of a platform. Major P2P platforms are in the process of getting regulated and getting the stamp of approval. Big companies can stay away from platforms like that. Since P2P lending is still in its baby stages, independent research and necessary due diligence is unavailable. This is a possibility of institutions choosing the best loans available for themselves as being large proportional investor.

Corporate banks: Corporate banks often invest through P2P platforms to credit small and medium sized companies. For example, the British Business bank[6] invests in small loans through Funding Circle.

---

[3] https://www.statista.com/statistics/461645/smes-share-stock-outstanding-business-loans-in-totals-uk/
[4] https://www.fundingcircle.com/uk/
[55] https://www.lendinvest.com/
[6] https://www.british-business-bank.co.uk/

Investment Trusts:

Investment trusts like P2P Global Investments[7] provide high yield to institutional investors. A negative can be, it rates fluctuate significantly from net asset value (NAV), that may cause large losses.

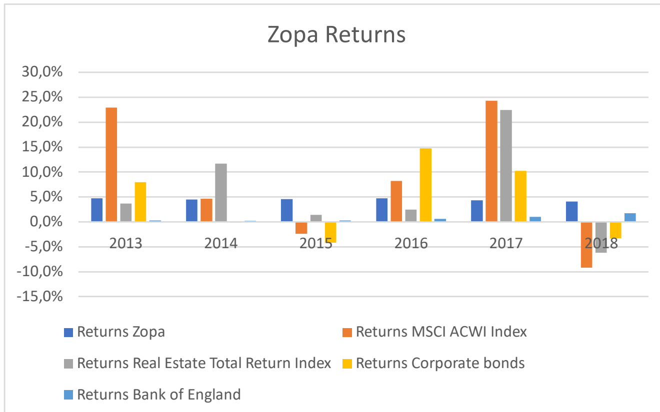


Figure 2:Comparision of returns of Zopa platform to other popular investments.

Growth of Peer-to-Peer Lending:

Banks became inaccessible: The consumer lending is becoming more crowded with the traditional institutions branching out into peer-to-peer lending services. This has reduced the yield of the investor. All though it is good for the borrowers but bad for the lenders. The increased competition can slow down the P2P market. To counter this, platforms like Funding Circle provide loans on an average below £100000, which usually is below what large institute look for.

Most investments return lower yields:

10-year government bonds and cash in banks provides near zero returns. The investor looks for alternative finance for making profits.

Innovations in technology:

Technology has the power to disrupt traditional business. Like Uber created new opportunities for taxi service and Airbnb provided alternative solutions to travellers, P2P platforms connects lenders and borrowers around the world. The advantage of better user

[7] https://www.bloomberg.com/profile/company/1367667D:LN

experience and speed drives the P2P market. Traditional institutions are stuck with legacy systems and organizational structures which obstruct business to excel.

Low trust on banks:

There has been a growing public distrust towards the traditional financial institutes after the 2008 credit crisis. An example can be seen in the UK's Edelman's Trust Barometer[8], which runs a continuous survey on public's trust on financial institutions. People are frustrated with their banks and are in search for alternatives.

Regulation of Peer-to-Peer platforms:

Innovative Finance ISA:

Launched in 2016, provided P2P platforms with tax-efficient solutions to their products. This shows the commitment of the regulator on the P2P market and it provides a secure product for the investor.

FCA Regulation:

The regulations will become stricter as the market matures and grows. Due to early stages of the market, many intermediaries do not participate in the market. Independent third party like investor education is likely to create hopeful scenarios in the future.

[8] https://www.edelman.com/sites/g/files/aatuss191/files/2021-01/2021-edelman-trust-barometer.pdf

Market review

The main market of P2P lending is the size of unsecured loan like the unsecured personal loans. The US consumer credit is 4.1 trillion from 2010-2020 as assessed by the Federal Reserve. China has a large market for new P2P lending platforms. It manages 99.7 billion USD in 2015 compared to UK which has a market of 4.1 billion USD [9].

The following section shows loans issued in prominent P2P lending marketplace around the world.

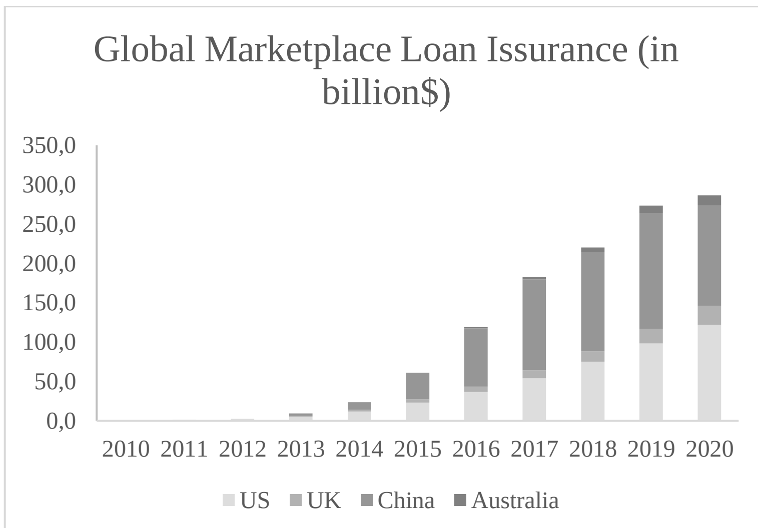


Figure 3:Global number of loan issued.

10

Table to the graph is provided at the appendix of the study here

---

[9] Source: The Global Alternative Finance Market Benchmarking Report by Cambridge Centre for Alternative Finance (2020).

[10] Source: Company Data, Morgan Stanley Research estimates

The following graph shows the awareness among public segregated by age.

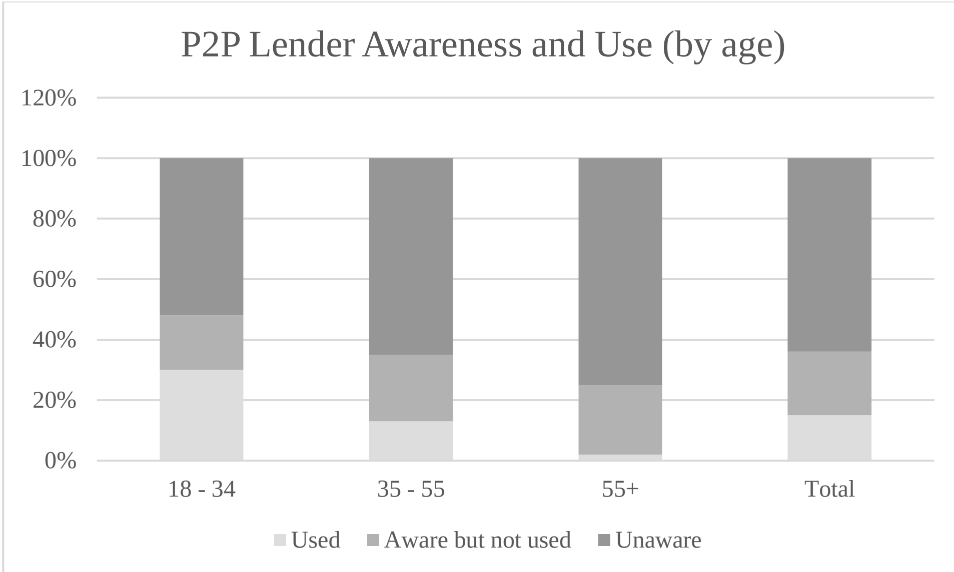


Figure 4:Peer-to-peer awareness.

[11]

Table to the graph is provided at the appendix of the study here

Key players and Respective Marketplace

Platform	Year of conception	Total Loan Issued (in bn)	Crowdfunding Type	Currency	Country
Lending Club[12]	2007	59,60	Personal lending	USD	US
Prosper[13]	2006	17	Personal lending	USD	US
Zopa[14]	2005	6,04	Personal lending	GBP	UK
Plenti/RateSetter[15]	2009	4,03	Personal lending	GBP/AUD	UK (RateSetter)/AUS
LendingWorks[16]	2012	0,21	Personal lending	GBP	UK
Funding Circle[17]	2015	10	Business lending	GBP	Ger/UK
Auxmoney[18]	2007	2.0	Personal lending	Euro	Ger

[11] Sources: AlphaWise, Morgan Stanley Research
[12] Sources: Lending Club https://www.lendingclub.com/info/statistics.action
[13] Sources: Prosper: https://www.prosper.com/;
[14] Sources: Zopa: https://www.zopa.com/invest/historical-performance
[15] RateSetter: https://www.ratesetter.com/invest/statistics
[16] LendingWorks: https://www.lendingworks.co.uk/about-us/statistics
[17] Funding Circle: https://www.fundingcircle.com/uk/statistics/
[18] Auxmoney: https://www.auxmoney.com/infos/statistiken

These are some of the players in the market. It shows Lending Club of US to be more matured compared to the UK platforms like Zopa although it started operations in 2005 before any other platform. German P2P marketplace is in its infancy.

Literature Review

Banks have been a solution at an institutional level to the issue of asymmetric information in the credit market when we look at the history of lending as ascertained by (Stiglitz, 1981) and (Akerlof, 1970). The credit market is between the provider of funds and the receiver of those funds. Banks can mitigate adverse selection and the moral hazard asymmetric as they have the expertise in screening and monitoring borrowers at a reduced cost compared to individual lenders, stressed by (Leland, 1977).

Therefore, in comparison to individual lenders, in a hypothetical scenario of perfect screening and monitoring, banks can extend funding opportunities to many firms and individuals who may not have a chance. In such a case, the borrowers should pay the right price as well as not be rationed to obtain the loan. All consumers who have the right price to pay (interest) would have access to credit and there would be no scope for new business ideas to complete with the banking sector. But contrary to that, in recent years with the growth of internet have led to crowd-based platforms which match lenders and borrowers progressively. In this way of lending the power is put in the hands of private lenders and borrowers. The decision making in the process of loan origination is between the two parties.

As (Herzenstein, 2011), (Galloway, 2009) and (Morse, 2015) stressed that in this kind of lending model, the mediation of banks in not required. Borrowers (either households or firms) post their loan request online and provide information on their current financial situation, e.g., income and open credit lines (other loans). On the other hand, lenders investigate a pool of credit seekers and search the ones that best fit their risk-return preferences. The final decision in the screening process rests with the lenders.

The success of P2P lending is all over the world. There is a positive trend for demand of credit through P2P platforms at the same time, comparable segments of the credit market via the banking sector are showing a negative trend.

Several studies have been conducted for the P2P lending market in the US, but there are very few papers which investigate this question in a developed country other than the US. Moreover, the interest rate comparison between the P2P platforms versus the banks have not been done explicitly due to the unavailability of public statistics on interest rates on consumer loan. Furthermore, US has more mixed market which includes non-banking loans (payday loans). Several studies have been conducted on the mechanism of transmission of rates. The meaning of transmission rate here is the changes in the market rate affects the retail rate. But not much research has been conducted on flexibility of credit demand with respect to interest rates on loans. This area of study is critical as in recent years there have been changes like lower interest rates, non-traditional monetary policy and more constraints for acquiring bank capital.

The interest in borrowing and lending via web-based platforms has been investigated by looking at the Google Trends database provided by Google.

As a sector P2P lending is relatively young which started in 2005 with the launch of Zopa, there is a continuous trend of research in this topic. The interest was triggered when a competitor of Zopa, US based Prosper made its platform's data public in 2007 as evidenced by (Ravina, 2012) and (Pope, 2011). P2P platforms have online presence which provides access to minute data on credit provision and provides an opportunity to investigate new aspects. In previous literature scholars have researched to what extent do customers' characteristics affect their interest rates. The impact of trustful pictures on interest rates have been evaluated by (Duarte, 2012) and (Lin, 2013) evaluated the impact of friendship connections on interest rates, and (Ravina, 2012) evaluates the effect of beauty and skin colour on interest rates. Information facilitation and institutional design may magnify chances of credit provision to a certain extent, is topic for research. (Hildebrand T. M., 2015) investigates the extent of the change of interest rate from auctions to rates that are pre-determined by the P2P platforms affects the amount of credit provision. (Herzenstein, 2011) analyses to what extent investors herd with each other when analysing which loan to fund.

A borrower applies for loans by providing private information such as loan amount, term, credit rating score, monthly income, occupation and the loan purpose. Platforms will then assess the information and decide a fixed interest rate for the loan. After the interest is agreed on by the borrower, the loan will be listed on the platform for investors to browse. Then investors can browse loan information and decide whether to invest and how much to invest.

Between 2008 and 2018, many papers were on the topic of P2P lending. Out of them 21 papers discussed how to increase the possibility of loans being successfully funded and what are the key determinants. (Gregor, 2010) says verified variables play a much more significant role in determining whether to invest in a loan compared to unverified variables. Also, borrowers who are willing to disclose more information normally pay less interest rate (Bohme, 2010). Social ties will increase the chances of having the loan fully funded (Herrero-Lopez, 2009) (Greiner, 2009) (Hildebrand T. P., 2010) thereby reducing the interest charged on the loan, and also decrease the default risk associated with the loan (Lin, 2013). Furthermore, some research is focused on the contribution of demographic information of borrowers on loan funding such as appearance and gender.

Research shows that appearance also does influence the decision of lenders to fund a loan or not (Jefferson Duarte, 2012). Female borrowers are less likely to get loans funded than are male borrowers. The objective of lending money on P2P platforms is to gain high return and mitigate default risk. Investors on P2P lending platforms are inclined to invest in loans with higher return, which also carry higher default risk. Assessing default risks based on previous loans' performance is another focus of academic papers.

There are 10 papers that built models to investigate what are the key determinants of default risk, so investors can use this as a guideline to avoid adverse selection. Loans with lower credit grade and longer terms will result in higher default risk (Riza Emekter, 2015). There are discrepancies between risk premiums charged and real default risk associated with loans on P2P lending platforms (Kumar, 2007). This conclusion is supported by the fact that the proof shows that the premium charged by P2P platforms is not enough to cover the potential loss of investors (Riza Emekter, 2015). Recommendations were also imposed that another way to mitigate default risk of loans is to set up a social reputation system in P2P lending platforms (Everett, 2010). Platforms will charge borrowers a loan origination fee once the loan is successfully funded. Investors will also be charged a service fee of managing instalment payments from borrowers. A handful of papers were focused on building the internal information system of P2P platforms. For instance, (Collier, 2010) informed practice and theory on developing community reputation that can improve information asymmetry on Prosper and mitigate adverse selection. Also, as an intermediary in the financial market, platforms are regulated by both Securities and Exchange Commission (SEC) and Consumer Financial Protection Bureau (CFPB). Five papers uncovered the current regulations on P2P lending and inform implications for further development of specific regulation for P2P lending. A multi-agency regulatory approach of P2P lending should be implemented that intimates the approach applied to regulate traditional lending (Rapp, 2012).

Borrowers need to pay monthly instalments until the loans reach maturity. If desired, they can also choose to repay all principal payments ahead of the loan's maturity by paying a service fee. Platforms also provide a trading system to investors who want to sell holding loans with a certain discount. This trading system helps platforms to provide more flexibility to investors. However, some loans default in early stages of instalment payments. This causes a huge loss for investors. Investors are inclined not to hire an agency to collect net principal loss due to the small amount of investment (Jin, 2014). Further research into after-default management of P2P lending is an urgent need because it can help mitigate net principal loss of investors and improve the risk-adjusted return of platforms as a whole.

There are different types of P2P lending platforms based on the activity they do as shown in the table below:

Type of loan	Platform
General loans/Personal loans	Lending Club, Prosper,
World Poverty reduction loans	Kiva, Zidisha
Family and Friends loans	TrustLeaf, LoanKin
Others	Funding Circle, Kabbage, Sofi

The next section presents the types of business models which exists in the P2P market currently.

Business Models of P2P Lending[19]

1. One to Many Model (diffused): The platform actively decides in loan selection as well as matching borrower and investors. The platform collect money from the investor and allocated it to several loans depending on the investors agreed upon amount to lend, expected returns and risk appetite (depends on the risk expected from the investor's portfolio). Advantage of this model is borrowers get money quickly and many borrowers can get their request of funds.
2. One to One Model(direct): Here the investor selects the loan, according to the documents submitted during loan application. The investor decides the loan amount to fund the borrower. It is more time consuming and does not assure diversification of portfolio of investor. In this model, borrowers may not be fully satisfied because of being selected partially.

Finance processing Models[20]

1. Client segregated model: The platform matches investors and borrowers, and the money is collected in a separate account different from platform's account. The financing is done through reverse auction (Milne A, 2016). The quotes of the common fund are displayed on the platform. This is done to preserve the funds if the platform fails. Reverse Auction: The benefit is the lowest bid(s) get the loan. In a conventional auction (at least in theory) each bidder makes an individual judgement on how much the item is worth to them and bids up to that limit. The item is then won by the person who values it highest. In a conventional auction items usually go for above the reserve price. So in a reverse auction for a loan each lender decides the minimum rate they are willing to accept from that borrower and the loan is funded by the lender(s) who are prepared to lend at the lowest rate.

2. Partnering Model: Loans are generated by partner banks. Platforms only acts as a broker.

---

[19]Information from different websites and papers on peer-to-peer lending.

[20] Information from different websites and papers on peer-to-peer lending.

3. Guaranteed return Model: Platforms performs the assessment of credit worthiness of the borrower and sets up the interest. It collects funds from the investor. There are two ways of execution of the model:
    a. The research and pre-screening of financiers is done outside the platform. The loan request is displayed on the platform and lenders can make offers. This model works when there are more supply than demand for loans.
    b. An algorithm invests the collected fund depending on risk credit. In this method there is remuneration depending on the rate and loan period.
4. Balance sheet Model: Platform takes the loan and shows it in their balance sheet so it can sell them to institutions for investments. The platforms collect the money, provide it to borrowers who pay interest to the platform. This can be dangerous if the platform fails, the investor may not be able to get money back.

Methodology

Data collection and preparation: Zopa

Zopa being the oldest[21] financial peer-to-peer lending platform company has a risk of losing due to credit risk. Now the credit risk has been improved. To understand better what makes these loans riskier, I analysed the loan book[22] of Zopa. Some patterns behind borrowers are observed and how they link to default or successful loan.

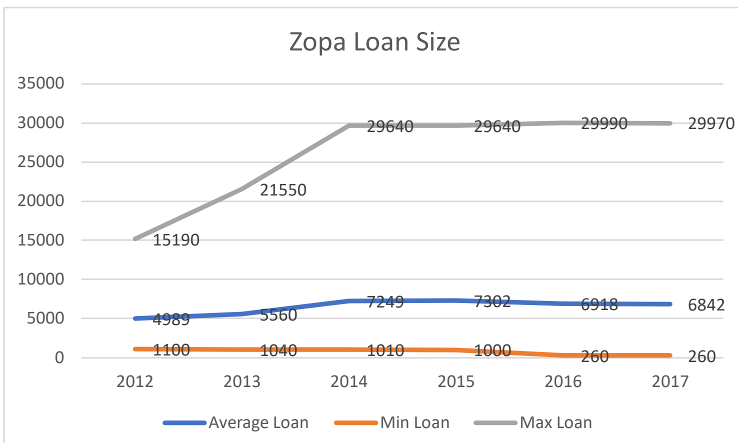


Figure 5: Loan size of Zopa

In traditional lending, a loan is evaluated based on the borrower's credit score from an independent credit assessing agency[23]. These loans are provided by banks traditionally. In contrast, the P2P lending platforms provide more riskier loans. If these loans are assessed by the traditional credit agencies, they will always be categorized as high-risk loan. These factors which make these loans high risk could be questionable to investors. So, the metrics used in Peer-to-Peer loans properly are different from traditional credit metrics.

---

[21] Zopa wiki
[22] loanbook
[23] peer-to-peer lending

Zopa started in 2005[24]. The total amount of loans provided by Zopa till 2021 is close to 5 billion pounds[25]. Zopa provides personal loans to individuals. These loans are repaid within 5 years. The rate of return for lenders is between 3.5 to 3.7 percent. This excludes bad loans.

Since it is the oldest peer-to-peer lending platform, it has a huge set of loan data which makes it easier to approve new loans come in and fix interest rates based on how the borrowers have done in the past.

Even though the stock market has seen a significant low ever since the peer-to-peer lending platform has started, Zopa has shown an excellent positive return every year for its lenders, even though the financial crisis of 2008. It is a good choice for any lending portfolio.

The minimum amount of money investors can lend is £ 1000. The borrowers repay back the money lent to them with interest. The interest repaid is between 3-34%, which depends on strength of the possibility of repayment. Zopa makes this assessment on the basis of the data they acquire from the credit reference agencies as well as its previous experience providing more than 600,000 loans.

The money lent out by the investor is spread across many loans, at least 100 loans, to reduce the risk of losses due to default. Investors have the option to lent out more money in the first 3 years and relent the repayments received to counter losses.

Zopa has reliable products such as Zopa Plus, where investors can lend to a mix of creditworthy borrowers. The interest earned is fixed at 5.2% regardless of losses incurred. Zopa provides other products such as Zopa Core, with a possibility of earning 4.5% interest excluding the bad loans. Zopa Core includes highest grade borrowers. There is also the possibility to earn more if there are less bad debts and earn lower interest than target rate if there are more bad debts. This is designed to provide satisfactory returns to majority of investors.

---

[24] Zopa wiki
[25] Zopa

Zopa performs certain tasks to maintain satisfactory performance through years. It finds new borrowers. They assess the loan applications with data provided by the borrower. After approval, investors are assigned small loan amounts. It collects repayments by the borrowers which it pays back to the investor. In case of bad loans, it tracks the borrower repayment patterns. It also restricts applications and provide less loans in weak economic conditions as well as in crisis.

Zopa also provides the possibility to earn £1000 interest tax free using personal savings allowance or if invested through a Zopa IFISA.

Zopa earns income in the form of loan serving fee from the borrower. Investors also have to pay a 1% sale fee in case they exit early.

Investors can exit loans by turning off auto-relend to get back money normally from borrower repayments. There is also a possibility to sell existing loans other investors, if any investor wants to exit early. Zopa finds investors for them. When an investor sells the loans, they get the outstanding loan amount back. But bad loans cannot be sold. Loans can be sold early depending on availability of other investors ready to buy loans.

The investment which has not been lent yet is protected by FSCS. But after the loan is lent out, it’s not regulated by FSCS. Loan closure plans are already set before giving out loans, in case of such cases. Money is set aside to process closures. The whole process is regulated under United Kingdom’s financial regulator FCA.

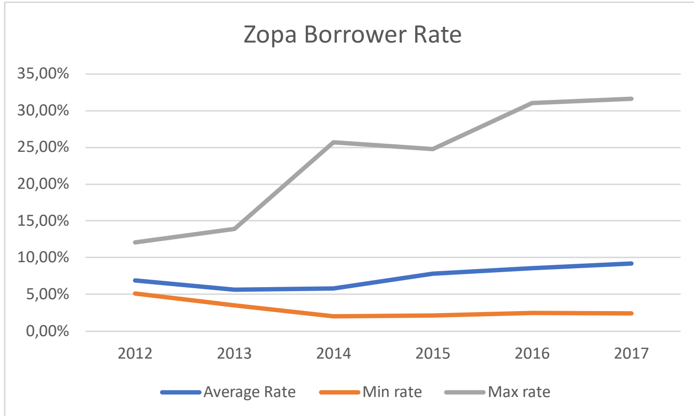


Figure 6: Borrower rates for loans from Zopa over the years.

Zopa accepts a borrower of a wide range, borrowers who had under average financial position. But borrowers have to show good track record of repayments. It usually accepts 20% of its loan applications, at least during normal times. Borrowers need to have a good credit history, a repayment record and if they can afford the loan. During COVID-19 pandemic, Zopa has tightened lending criteria to ensure the borrowers have lower rate and lower risks due to the economic fallouts of lockdown.

Recovery of bad loans is usually not a big mention is personal loan platforms like Zopa. But the bad debt is in line with other peer-to-peer lending companies in the industry and close to the recovery for traditional banks.

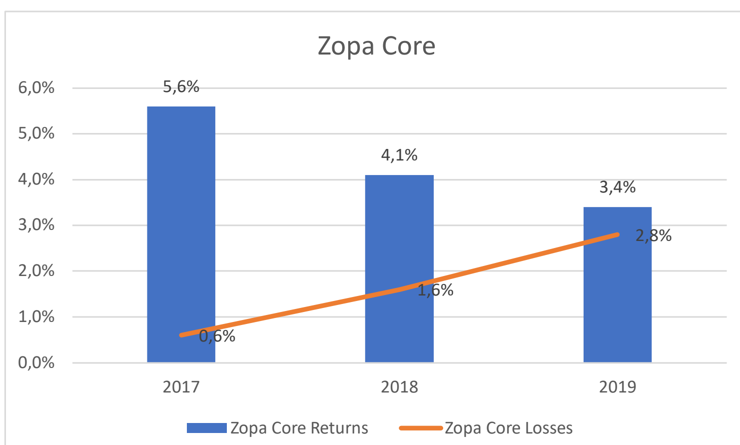


Figure 7: Zopa Core returns vs losses.

Zopa Core[26] lending account provides a return of 2-19% interest to lenders. The average lending rate is estimated to be 8.63% in Zopa Core loans excluding the bad loans. A minor data is omitted due to irregularities. The target return for Zopa is 3.2% in Zopa Core product.

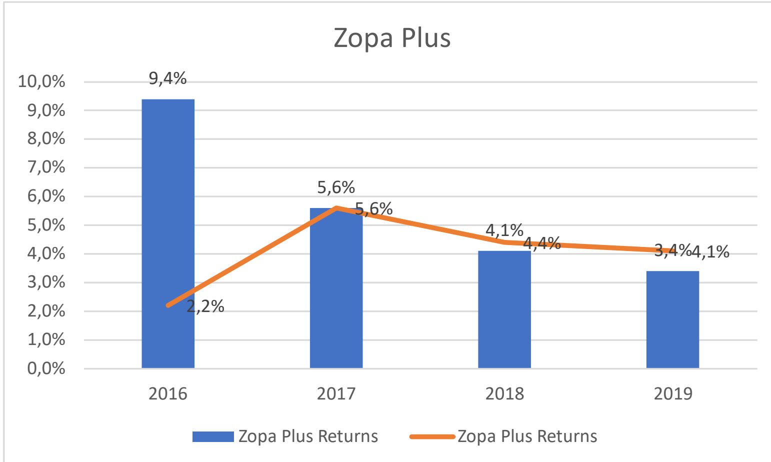


Figure 8: Zopa Plus returns vs losses

In Zopa Plus[27], a return of 34% in some loans is provided. The average rate till the pandemic started was 11.73%. After the bad debts, the target return for Zopa Plus is 3.7%

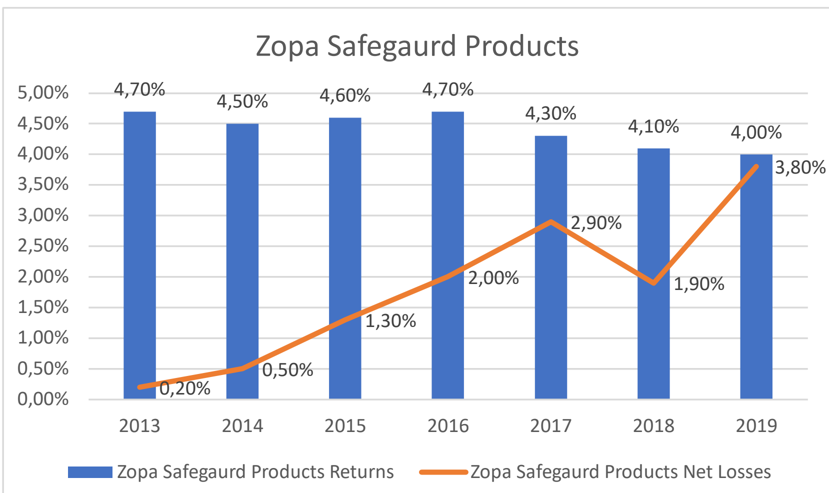


Figure 9: Zopa products with Safeguard returns vs losses

[26] zopa
[27] zopa

Data processing: Zopa

The Zopa dataset was obtained from their website[28] under the statistics section available for public downloads. It can be downloaded here. There are 16 variables and 734079 observations for each loan during the period 2005 to 2020, which can be divided into three categories:

- Loan Status: It indicated the status of the loan such as Active, Late, Default or complete.
- Borrower Data: These show the data provided by the borrower such as PostCode
- Loan Data: Basics of the loan such as term of the loan, Lending rate, disbursal date, original loan amount, principal collected, interest collected, total no. of payments made, term of the loan, date of last payment, latest status of the loan and in case of default, the date of default is also mentioned.

The Zopa[29] variable table which are of interest for analysis is as follows:

Key Variables:

Variable name	Type	Definition
Original loan amount	Quantitative	The amount of the loan
Principal collected	Quantitative	Amount of the principal of the loan collected.
Interest collected	Quantitative	Amount of the interest collected
Total no. of payments	Quantitative	No. of instalments
Term	Quantitative	The length of the term
Lending rate	Quantitative	Interest rate the borrower needs to pay
Latest status	Qualitative	Whether the loan is active, complete or default
Date of Default	Quantitative	Date on which the loan of defaulted.
Post Code	Qualitative	Location of the loan taker.

Initial question of interest is can investors always make money on riskier loans?

Taking a first look on their website indicates that credit assessment metrics on P2P lending platforms have some value, which is been referred when making a decision in favor of

[28] Zopa loan book
[29] Source: Zopa Loan book collected from their website: https://www.zopa.com/invest

the loan. Something similar can be observed in other successful lending platforms such as Lending Club. These is not a lot of information about the borrowers on their platform, except for the metrics provided by them. Although credit metrics used to be part of the loan book they provide on the website but it has been scraped for a few years. The loanbook does not contain any personal information on the credit rating or credit score of the individual borrowers. This information is for the investors for seizing opportunities to earn money. So a better question is what properties of borrowers in P2P platform are relevant to identifying the nature of the loan and how they are linked to default loans? If there are some trends in these properties, maybe that can be important points in calculating a loan as good or bad.

I analysed the Zopa loanbook in R software. Firstly, I loaded the necessary packages in R studio. The packages included ggplot2, tidyr, dplyr, ggthemes, gridExtra, RColorBrewer, knitr, GGally, psych, corrplot libraries. Thereafter, I loaded the loanbook of Zopa in R. I changed it from .csv format to .xlsx . The dimensions of loanbook is that it contains 14 variables and 734079 observations for each loan data.

First, I looked at the loan status of the loans in the dataset. The same could be found at the 'Latest Status' column of the loanbook. The variables in the column are as follows:

1. Active
2. Completed
3. Default
4. Late

There are four loan status. I wanted to know about the risky loans in the dataset, so I focused on the Default and Late variables in the Latest Status column of the dataset. So I combined the two categories into one category. I changed the format of the column into variable. The number of observations (loans) in each category is as follows:

Category	No. of observations
Active	257302

Category	No. of observations
Completed	416772
Default	37533
Late	22472

The barplot of distribution of the loans are as follows:

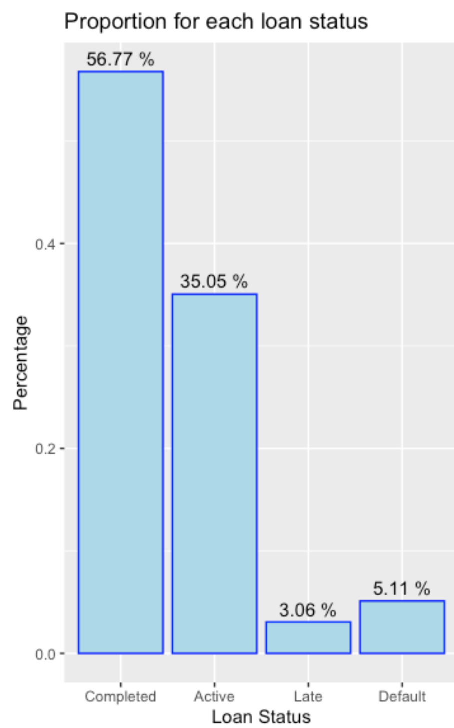


Figure 10: Distribution of all loans of Zopa

I created a new variable called the CompleteOrRisk to assign to risky loans included late and default. The high risk category included loans which are late or default and the completed loans are assigned to completed category. The active loans have been removed from this new variable as the nature of loan could not be determined at the time of analysis.

I created a bar chart to see the distribution of the terms of the loans. The 'Term' column indicates the no. of terms the loan has been provided for.

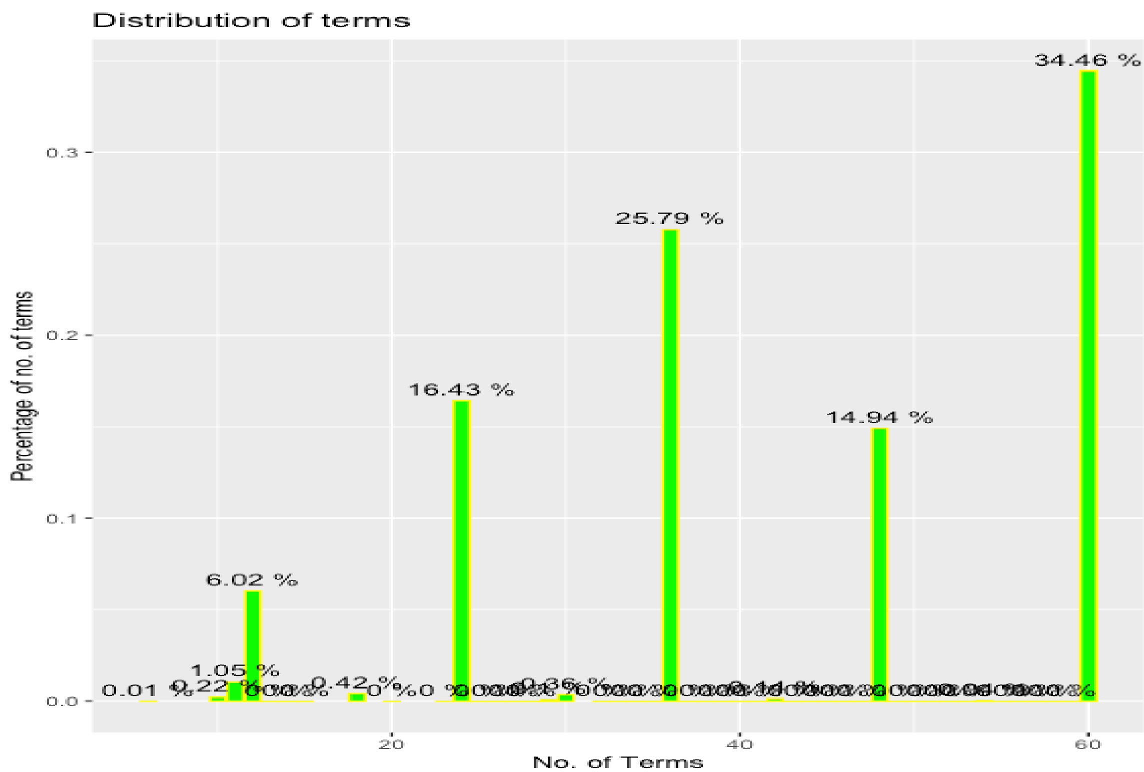


Figure 11: Distribution of terms of loans of Zopa

This shows the higher frequency of loans been given for 12 months, 24 months, 36 months, 48 months and 60 months. The various terms for the loans are as follows:

Months	No. of observations
6	56
10	1629
11	7703
12	44216
18	3101
24	120629
29	698
30	2646
36	189329
42	1024
48	109639
54	287
60	252992

The distribution of the loans based on different location in UK is as follows. I have included here a table to indicate the locations to the largest number of loans from UK.

Location	Post Code	No. of observations
Croydon	CR0	1853
Greater London	E17	1261
Leicestershire	LE3	1150
Lancashire	PR7	1072
Nottingham	NG5	1050

I created a subset of the data which contains only Completed and HighRisk loans. The no. of observations in the subset has reduced to 476777, containing 15 variables.

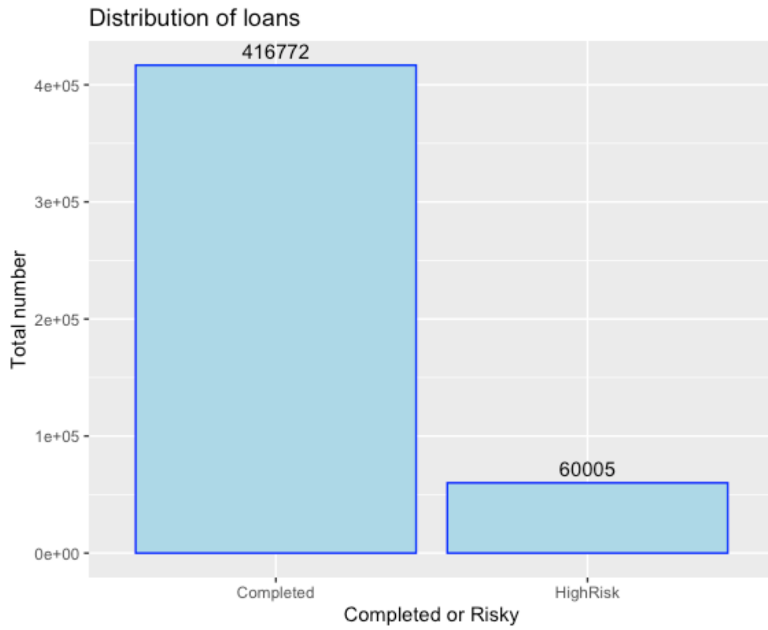


Figure 12: Distribution of loans which are completed or at high risk to default.

After removing active loans, the no. of loans for each category is as follows:

Status	No. of loans
Completed	416770
Late	22472
Default	37533

The proportion of each loan category is shown in the following bar chart.

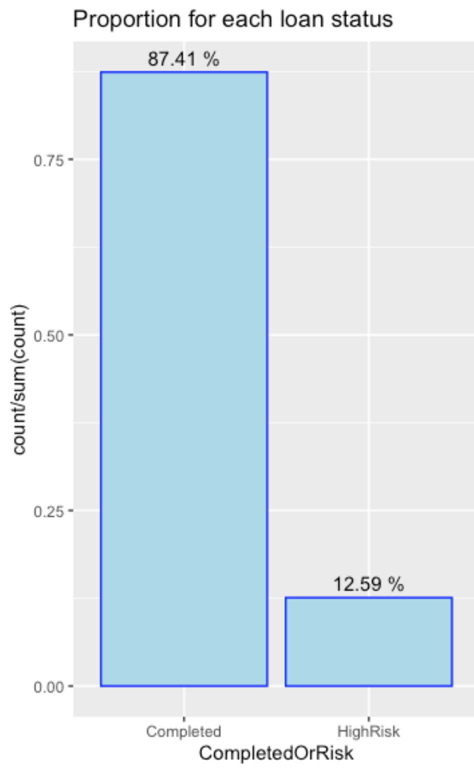


Figure 13:Prportion of loans in each category of Zopa

The proportion of the active loans are 35.05%. The unknown nature of these loans make it unusable for this analysis. We don't know if they are defaulted or completed. There are 56.77% loans completed. There are 3.06% loans which are late, means there is a possibility that the investor may lose money.

Next, we look at the lending rates of Zopa dataset.

The loan data will give a better picture about the dataset. The lending rate for the loans is summarised as follows:

No. of observations	734079
Mean	9%

No. of observations	734079
Standard deviation	+-6%
Median	7%
Median absolute deviation	6%
Minimum	1%
Maximum	34%
Skew	1.23
Kurtosis	0.85

The above table shows the average lending rate is 9% in all of the loans. The range is about 1% – 34%, with slightly right skewed. It has a spike at 2.94%.

The histogram of the lending rates is as follows:

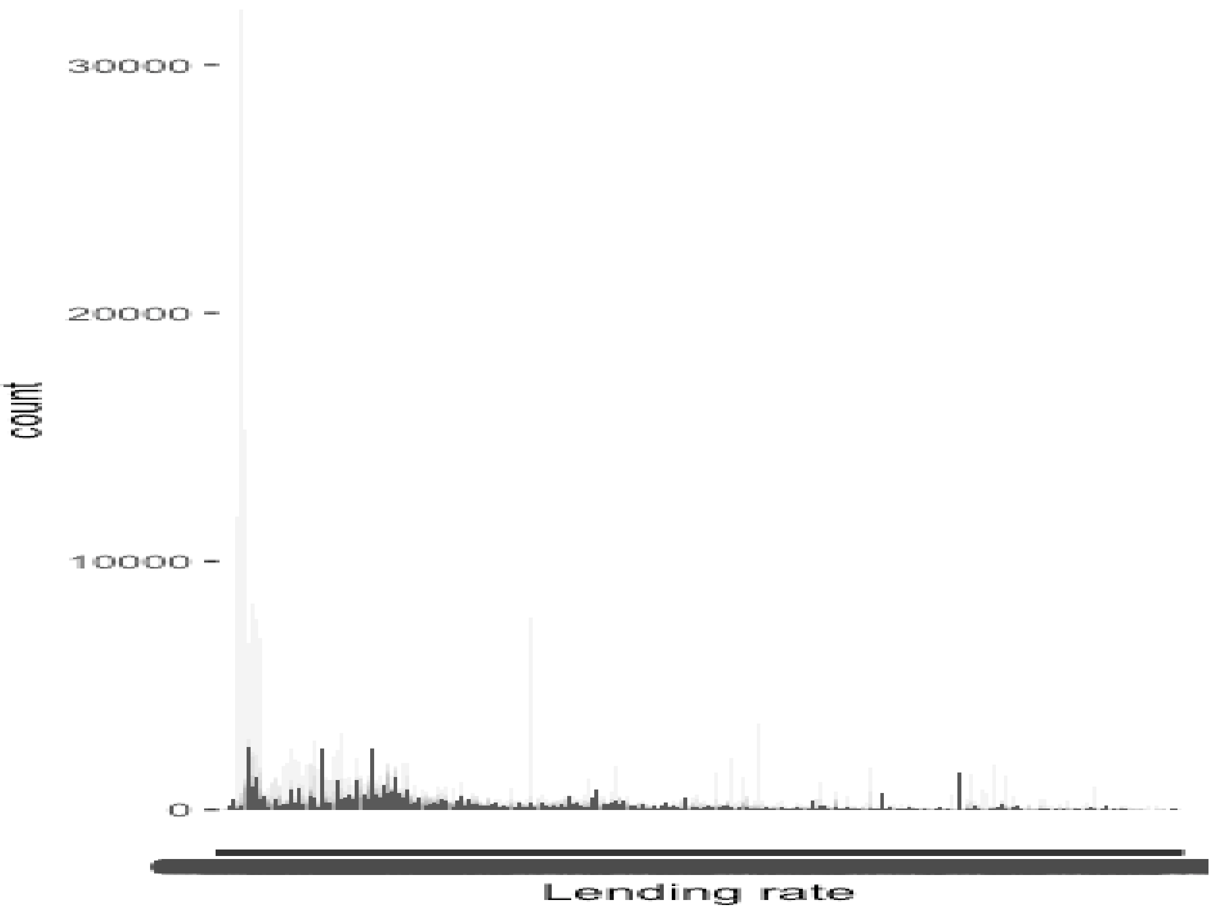


Figure 14: Distribution of lending rates of Zopa

The frequency of the lending rate often given to borrowers are as follows:

Lending Rate	No. of observations
2.94%	32241
3.04%	15286
2.84%	11823
3.24%	8291
11.61%	7758

The following section shows the lending rate distribution within the completed or risky loan category.

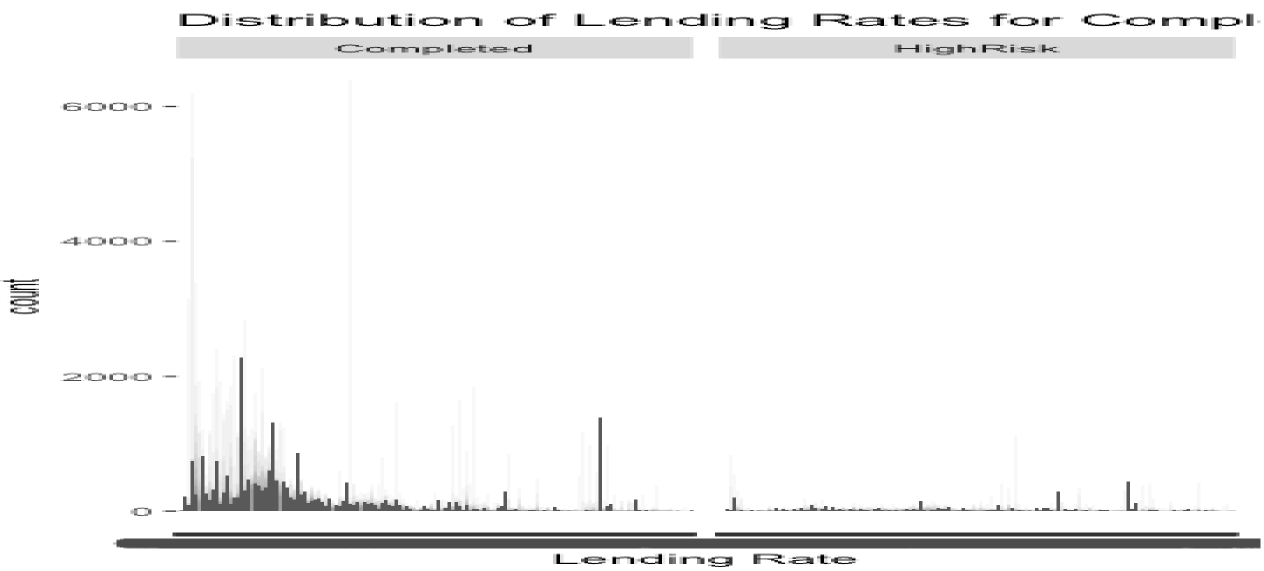


Figure 15: Distribution of loans in each category i.e. completed loans or high risk loans.

The summary of lending rates in the completed and risky loans are as follows:

No. of observations	476777
Mean	9%
Standard deviation	+-6%
Median	7%
Median absolute deviation	4%
Minimum	1%
Maximum	34%
Skew	1.31

No. of observations	476777
Kurtosis	1.09

The most frequent interest rates is 3.04% with 6742 observations.

Lending Rate	No. of observations
3.04%	6742
11.61%	6483
2.94%	6075
3.24%	3685
2.84%	3516

The following shows the summary of the lending rates in the completed and high risk categories:

Feature	Completed	HighRisk
Minimum rate	1%	2.42%
Maximum rate	33.78%	33.78%
Average rate	8.73%	14.60%
Median	7.06%	14%

The average lending rate for high risk loans is 14.6%, with spikes at 3.04% and the rate with completed loans are around 8.73%, which is slightly right-skewed. Generally, completed loans have lower lending rates compared to high risk loans. The highest rates in high risk loans go up to 33.78%. This may be specific to a particular year. There could be a financial crisis in a particular year to make the rates so high. And the rates are sensitive to financial conditions.

The following shows the summary of Original Loan amount for the Completed and high risk category.

The lending rates for each year could give a better understanding of the situation.

I created a column of year to check that.

Figure 16: Distribution of lending rates of Zopa over the years from inception.

Year	2005		2006		2007		2008		2009		2010		2011		2012
Category	All	Comp. or Risk	All	Comp. or Risk	All	Comp. or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All
Min rate	4.4%	4.4%	3.9%	3.9%	4.3%	4.3%	4.85%	4.85%	5%	5.01%	1%	1%	1%	1%	5.12%
Average	6%	6%	5.65%	5.65%	7.09%	7.09%	9.03%	9.03%	8.61%	8.61%	8.63%	8.63%	7.15%	7.15%	6.83%
Max rate	14%	14%	13.4%	13.4%	14.33%	14.33%	18.5%	18.53%	15.19%	18.2%	15.19%	15.19%	13.32%	13.31%	12.09%

Year	2013		2014		2015		2016		2017		2018		2019		2020
Category	All	Comp. or Risk	All	Comp. or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All
Min rate	3.5%	3.5%	2%	2%	2.14%	2.14%	2.49%	2.49%	2.42%	2.42%	2.84%	2.84%	2.8%	2.84%	2.88
Average	5.78%	5.78%	6.11%	6.11%	8.6%	8.68%	9.76%	10.08%	9.8%	10.63%	9.27%	10.87%	10.72%	12.94%	10.28
Max rate	13.9%	13.9%	25.7%	25.7%	24.75%	24.75%	31.03%	31.03%	31.62%	31.62%	33.61	33.61%	33.78%	33.78%	31.55

During the financial crisis of 2008-2009, we can observe that the minimum lending rate has decreased to 1%. There is a gradual decrease in the average lending rate over the years. The rates increased again in 2015 onwards. In average lending rate of completed or high risk loans increased from 2016 onwards with regards to the average lending rate of all loans. The interest rate for the loans have a clear pattern over the years for high risk loans compared to all loans. It seems the rise in the interest rate is due to the higher risky nature of the loans and the possibility of default.

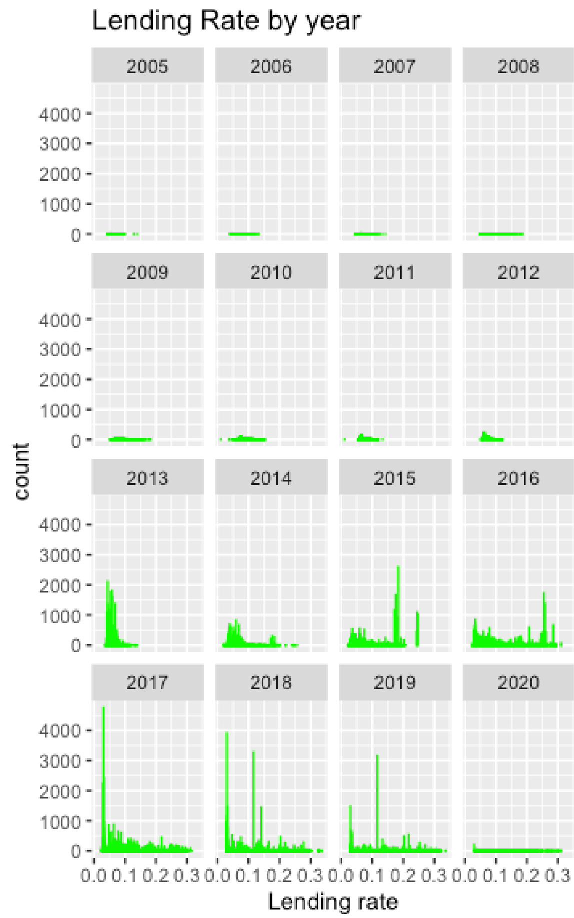


Figure 17: Distribution of lending rates of completed or high risk loans of Zopa.

In 2011 and 2012, there are fewer risky or completed loans. A significant factor could be the aftereffects of the financial crisis of 2008 and the high inflation market environment in UK. Next, we have a look at the amount of loan applied for over all loans as well as risky and completed loans.

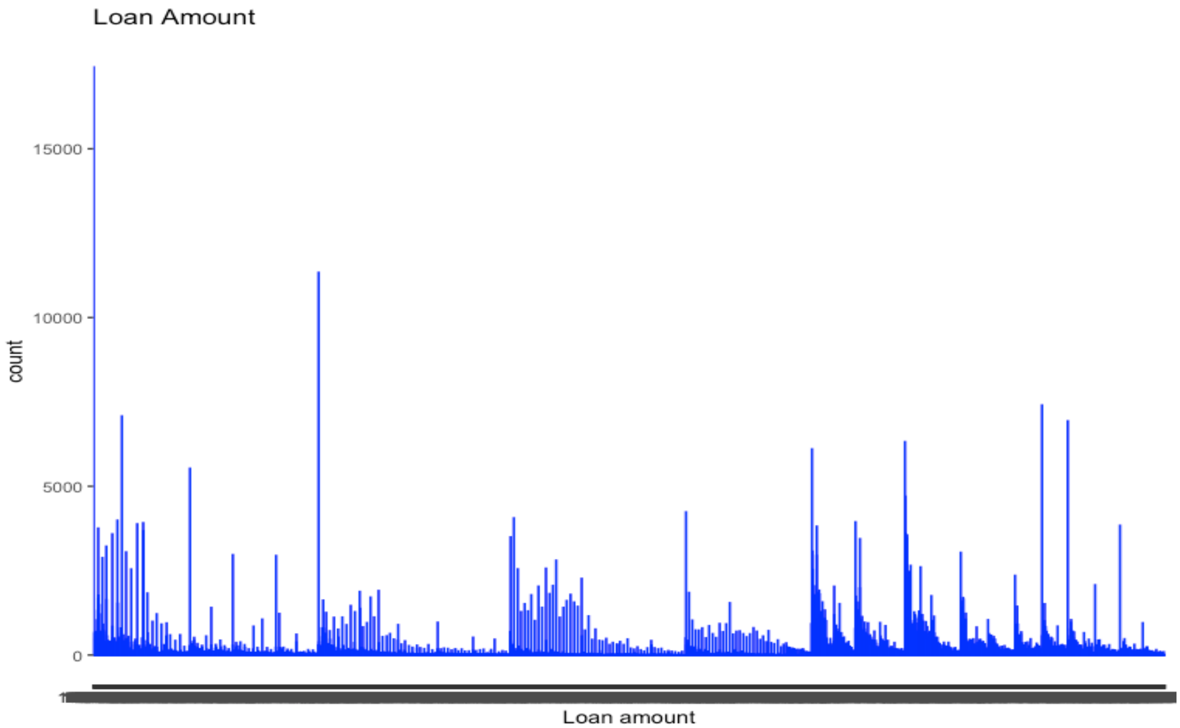


Figure 18: Loan amount for all Zopa loans

Looking at the graph shows loans of ten thousand pounds to spike. Average loan amount applied for is £7183.

Feature	All loans
Minimum amount (in pounds)	260
Maximum amount (in pounds)	35000
Average rate (in pounds)	7183
Median (in pounds)	5430

The following table shows the occurrences of popular amount applied as loans.

Original Loan Amount (in pounds)	No. of observations
10000	17424
15000	11338
7500	7407
1060	7082
8000	6938

The following shows the summary of Original Loan amount for the Completed and high risk category. The amount of money borrowed often as loan in the Completed or HighRisk category is as follows:

Original Loan Amount (in pounds)	No. of observations
10000	5835
1060	5301
3010	4406
1010	3673
3130	3580

Looking at both categories, all loans as well as risky or complete loans, we see a loan of £ 10000 to be common for both categories.

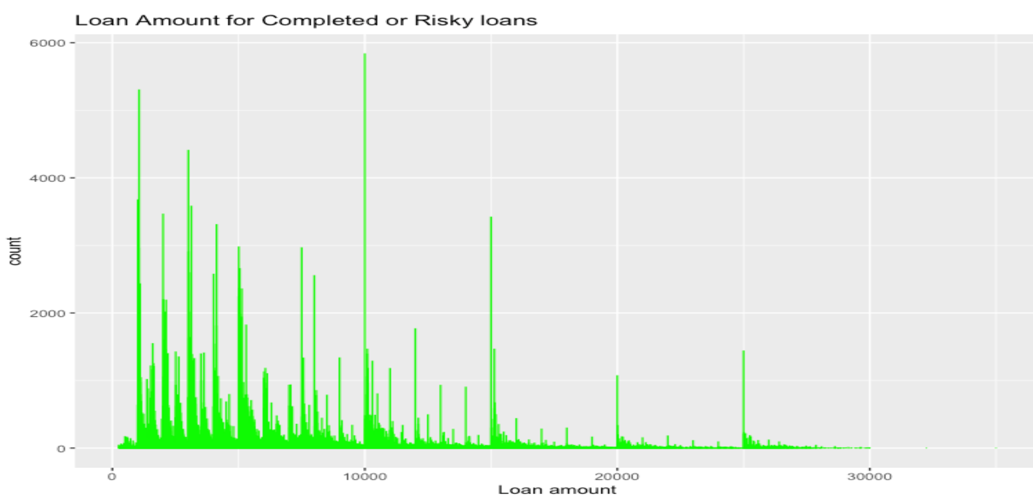


Figure 19:Distribution of total loan amount for completed or risky loans of Zopa

Feature	Completed	HighRisk
Minimum amount (in pounds)	260	260
Maximum amount (in pounds)	35000	29910
Average rate (in pounds)	6230	7947
Median (in pounds)	5010	6250

There is a clear trend that the average loan amount for riskier loans are considerably higher compared to successful loans.

The following shows the summary of Principal collected for the Completed and high risk category.

Feature	Completed	HighRisk
Minimum amount (in pounds)	260	0
Maximum amount (in pounds)	35000	2.964e+10
Average rate (in pounds)	6230	2.082e+09
Median (in pounds)	5010	1.093e+09

Now, we look at the total no. of payments collected for all loans.

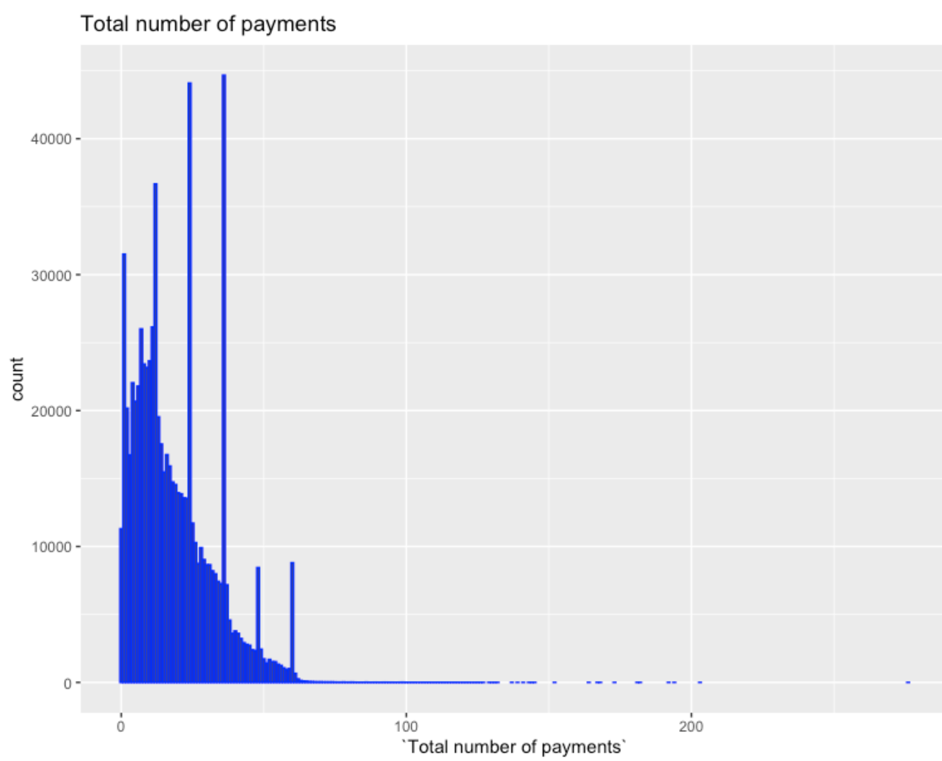


Figure 20: Number of payments of loan

Feature	All loans
Minimum no.	0
Maximum no.	276
Average no.	19.2
Median	16

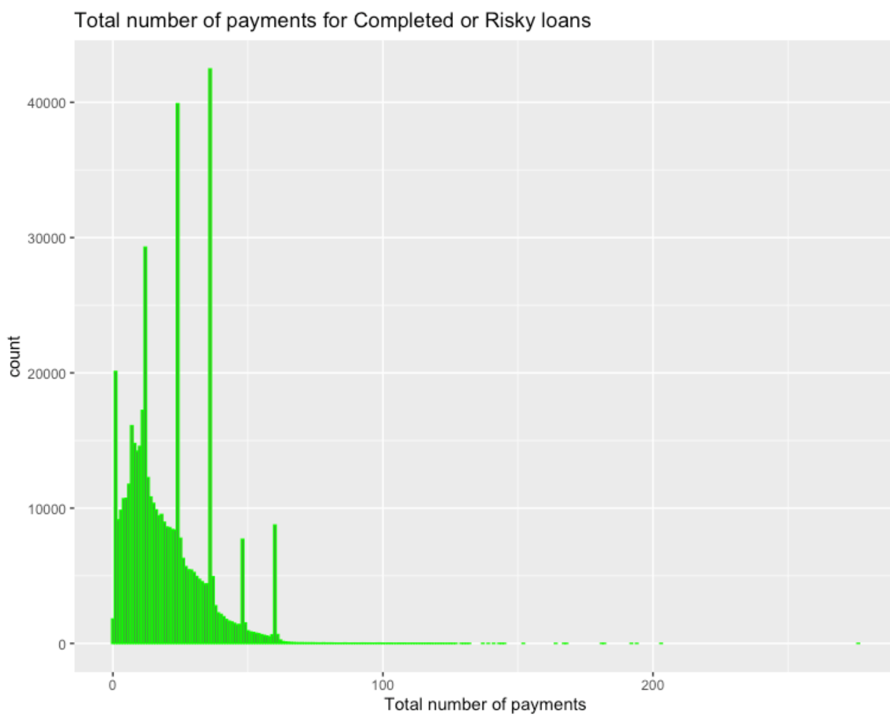


Figure 21: Distribution of total number of payments made by borrowers for completed or risky loans of Zopa.

Feature	Completed	High Risk
Minimum no.	0	0
Maximum no.	141	276
Average no.	21	17.6
Median	19	17

The average number of times of payment of loans made is higher in case of successful loans compared to high-risk loans.

Data collection and preparation: RateSetter

RateSetter started in 2010. One of the striking features of RateSetter is it distributes the risk associated with loans by putting a fraction of investment on all loan. It diversifies its loan profile by using all investment and distributing it to many loans, thereby reducing the losses associated with default of a particular loan.

Ratesetter also has a Provident Fund. This fund is used as a protection against bad loans. It's size has increased over time as the mix of loans have changed, thereby providing a secure future.

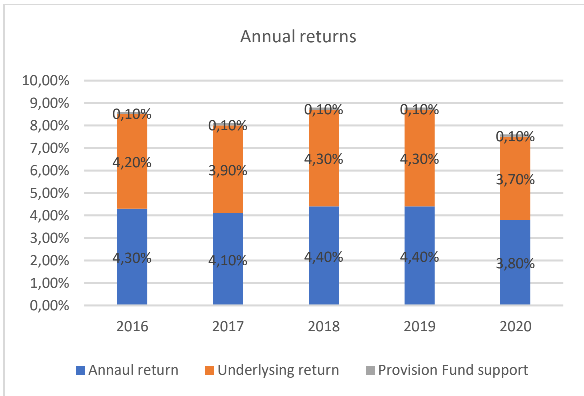


Figure 22:RateSetter Annual returns

The platform also allows early exit for lenders. In case of emergencies, the lender can sell their loans to other lenders, getting their investments back.

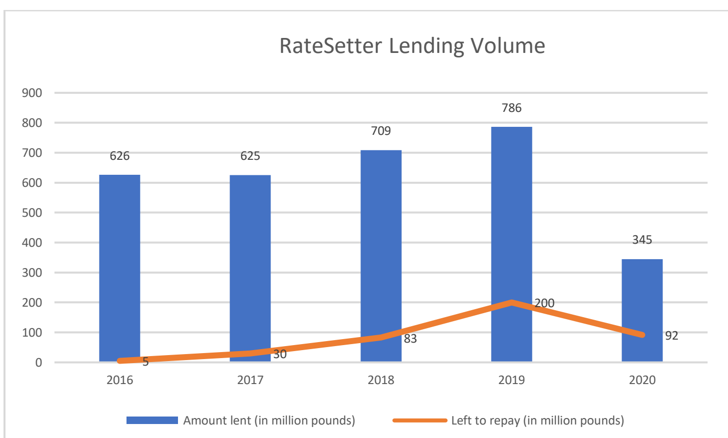


Figure 23: RateSetter Lending Volume

The current rate paid to lenders is 1.5% and 1.75% in Access IFISA account.

Data processing: RateSetter

Next, we analyse the loan book of RateSetter.

The key variable found in the loan book

Variable name	Type	Definition
Finance Amount	Quantitative	The amount of the loan
Principal outstanding	Quantitative	Amount of the principal of the loan not paid.
Finance purpose	Qualitative	Purpose of the loan
Loan term	Quantitative	The length of the term
Annual rate	Quantitative	Interest rates the borrower needs to pay
Repayment status	Qualitative	Whether the loan is active, complete or default
Borrower age	Quantitative	Age of the borrower
Employment status	Qualitative	Full time or contract or self employed
Borrower income	Quantitative	Income of borrower
Housing status	Qualitative	If the borrower has a mortgage or not
Early Payments made status	Qualitative	If the borrower has made an early payment (yes/no)
Borrower state	Qualitative	State the borrowers belongs

The RateSetter loan book[30] dataset will be explored here. I will try to find some patterns behind the loan takers properties and how they are related to default of loans or a successful loan.

The dataset has some basic properties. I have divided into three major categories:

1. Repayment Status: This indicates the status of the loan. The values which can indicate the status of the loan are loans which are in time indicated by “On Schedule”, loans which are less than 30 days old are indicated by “<30 days late”, loans which are more than 30 days late are indicated by “>30 days late”, loans which are in hardships

[30] Appendix

are indicated by “Hardship”, loans which are already defaulted are indicated by “In Default” and lastly, the loans which have already being paid up are indicated by “Repaid”.

2. Borrower’s Data: Some of the variables in the loan book show personal data regarding the borrower, such as the purpose of applying for the loan, the status of employment of the borrower, the income range of the borrower, the age range of the borrower and the housing status of the borrower.
3. Loan Data: This variables show the characteristics of the loan. It included variables like amount of the loan, indicated by “Finance Amount”, the amount of the principal of the loan outstanding, indicated by “PrincipalOutstanding”, the yearly rate of the loan, indicated by “Annual rate” of the loan, if the borrower has made some early payments, it’s indicated by the variable “EarlyPaymentsMade”, also date of starting the loan is indicated by “ContractDate”. The term of the loan is indicated by “LoanTerm”.

The first query that comes to mind looking at the dataset, is which variables could be connected to defaults of some of the loans. And what else factors indicate the successful completion of the loan.

I explored the data using the R software. I start with loading common libraries. Then the data set is loaded into the software. The file available on the RateSetter website is in .csv format. (now not available). I converted it into .xls format before loading in the software.

The dataset has 14 variables with loan 49942 observations.

Analysis:

First, I started with the status of repayment variable. There are 6 repayment status. My focus was to figure out the loans which has potential to be high risk loans. I changed the variables into characters. I plotted out the distribution of the loans based on their status. The graph shows 47.67% of loans, have been repaid. The current ongoing loans are at 46.32%. The high risk loans such as loan which are less than 30 days late are 0.98%, more than 30 days are 0.73% and loans under hardships are 1.47% and the percentage of loans defaulted are 2.83%.

Next, I created a variable called "CompleteOrRisk" to indicated all loans which are either completed, meaning which has been successfully paid off or the loans which are under stress, like the loans which are late, in hardships or have defaulted. I created a column with the variable name CompleteOrRisk. I excluded the loans which are currently active as well as repayment is on time.

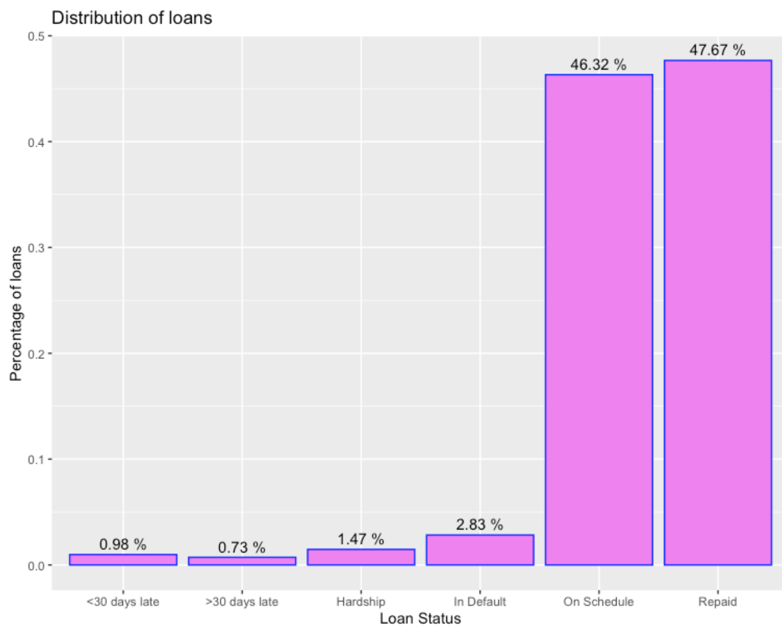


Figure 24:Loan distribution based on status of repayment of RateSetter.

I separated the data form the dataset in a subset which included data only from completed and high risk loans. The new dataset is called the sub_loans. The new dataset has 15 variables with 26807 loan observations.

I plotted the data from the CompletedOrRisk category loans. The plot below shows, 88.81% of loans have been repaid and 11.19% of the loans is under the high risk loans.

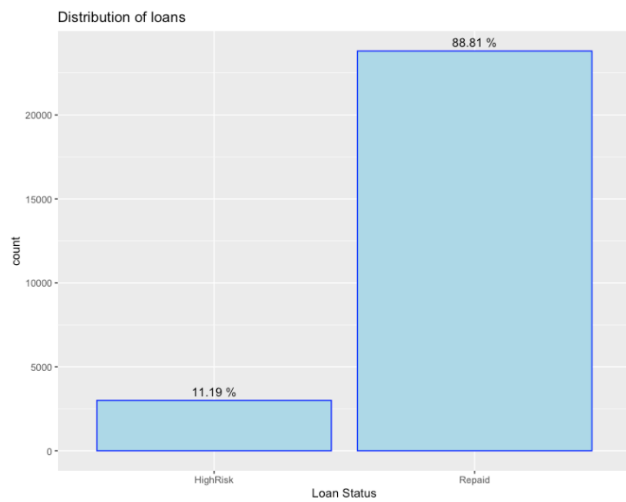


Figure 25:Loan proportion which have been repaid or are in high risk category

Now, I look at the annual rate of the loans of RateSetter. The annual rates of all loans are as follows:

Feature	Value
Minimum rate	2.05%
Average rate	6.65%
Maximum rate	16.5%
Median	6.92%

The annual rates which have maximum occurrences for the loans in the dataset are as follows:

Interest rate	No. of loans
6.92%	2563
7.02%	1731
6.5%	1668
9.61%	1080
8.48%	976

The figure below shows the distribution of annual rates for all loans in the data set.

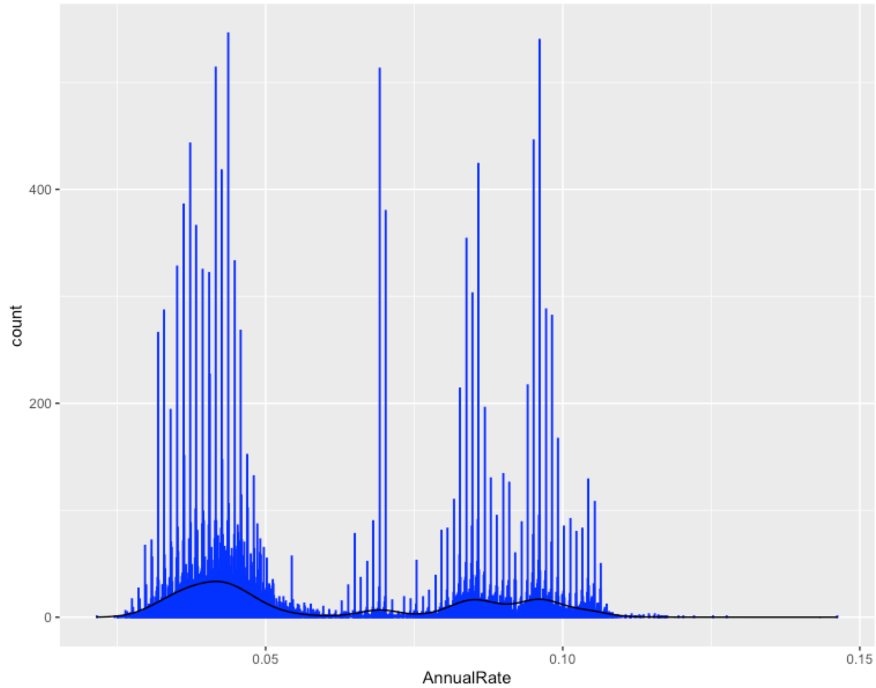


Figure 26:Distribution of annual rates of loans of Ratesetter

The annual rate on an average is 6.65%. The range of the interest lie between 2%-16.5%. It has a spike close to 7% and 9.6%.

Now I can check the annual rate distribution of the loans in the sub_loan data. The sub_loan consists of all loans which are either completed and risky in nature. Looking at the loans, we can observe the following.

Feature	Value
Minimum rate	2.16%
Average rate	6.17%
Maximum rate	14.6%
Median	4.8%

The interest rates which have been higher occurrences are as follows:

Interest rate	No. of loans
4.37%	546
9.61%	540
4.16%	514
6.92%	513
9.5%	446

The figure below shows that the annual rate of the high risk loans, with the increase in the annual rate, the nature of the loans to be more riskier also increase. This trend is opposite to that of the completed loans which show, higher no. of loans with lower annual rates.

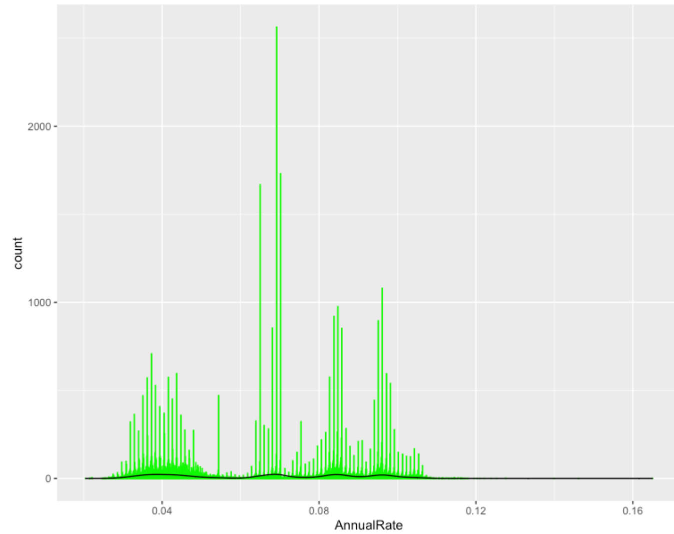


Figure 27:Annual rate of risky and repaid loans together.

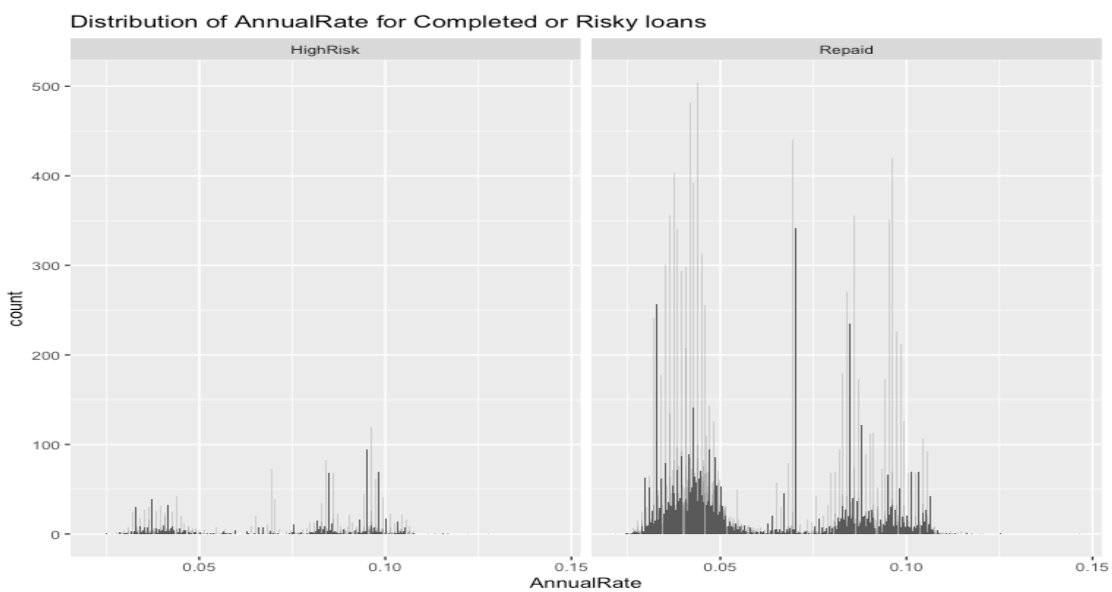


Figure 28:Annual rates of repaid and high risk loans of Ratesetter

On average the interest rate for borrowers in the category of CompletedOrRisk is 6.17%. The interest rate for high risk loans is 7.33% on average and repaid loans have an annual rate of 6%. But maximum annual rates for high risk loans are 12.76% whereas repaid loans have an annual rate of 14.62%.

Feature	Repaid	High risk
Minimum rate	2.16%	2.5%
Average rate	6%	7.33%
Maximum rate	14.6%	12.76%
Median	4.6%	8.3%

To understand better, the annual rates changes over time, I analysed the annual rate based on different years from inception. I created a column in the dataset with year in mind. I changed the contract date of the loan into a numeric and formatted it to years. The annual rates per year is as below:

Year	2014		2015		2016		2017		2018		2019		2020	
Category	All	Comp. or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk	All	Comp. Or Risk
Min rate	5.9%	5.9%	3.37%	3.37%	3%	3%	2.87%	2.87%	2.16%	2.16%	2.15%	2.46%	2%	2.64%
Average	8.4%	8.4%	7.9%	7.8%	7.5%	7.3%	6.37%	5.92%	6.76%	5.81%	6.51%	6%	6.35%	5.59%
Max rate	11.2%	11.2%	11.9%	11.9%	11.7%	11.7%	10.23%	10.23%	14.6%	14.6%	16.5%	9.44%	10.44%	9.21%

On an average, the annual rate of RateSetter have decreased over time, both all loans as well as repaid and risky loans. The annual rate of repaid and risky loans is lower steady compared to all loans for all years.

The chart below shows the annual rate distribution over the years.

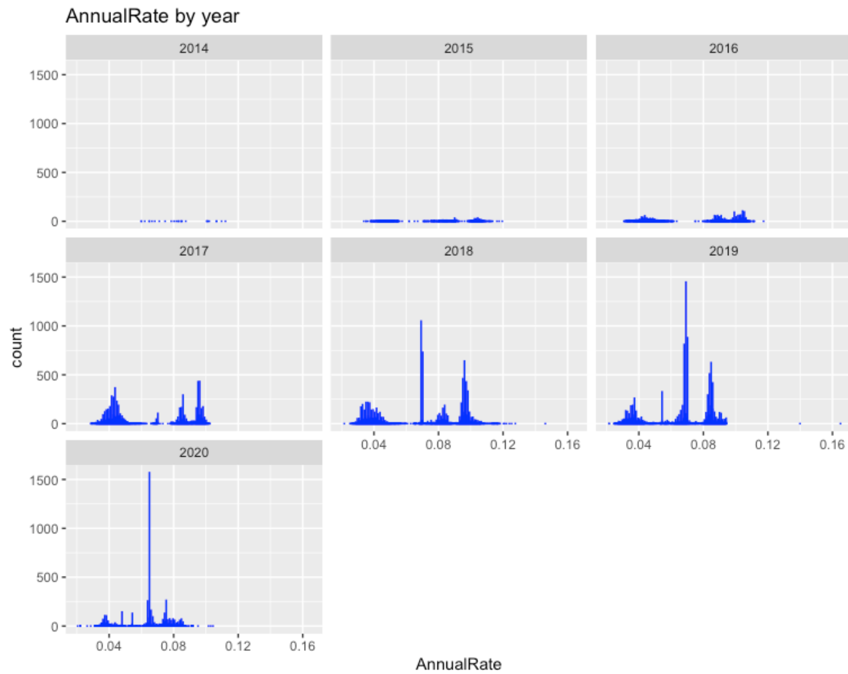


Figure 29:Annual rate by year of Ratesetter.

The following chart shows the annual rate of high risk or repaid loans over the years from its inception.

Figure 30:Annual rate of repaid or risky loans by RateSetter

Next, I look at the borrower data such as the income of the borrow and the housing status.

The distribution of the loans based on different location in Australia is as follows.

Location	Post Code	No. of observations
New South Wales	NSW	7270
Greater London	QLD	6998
Leicestershire	VIC	6615
Lancashire	WA	3459
Nottingham	SA	1397

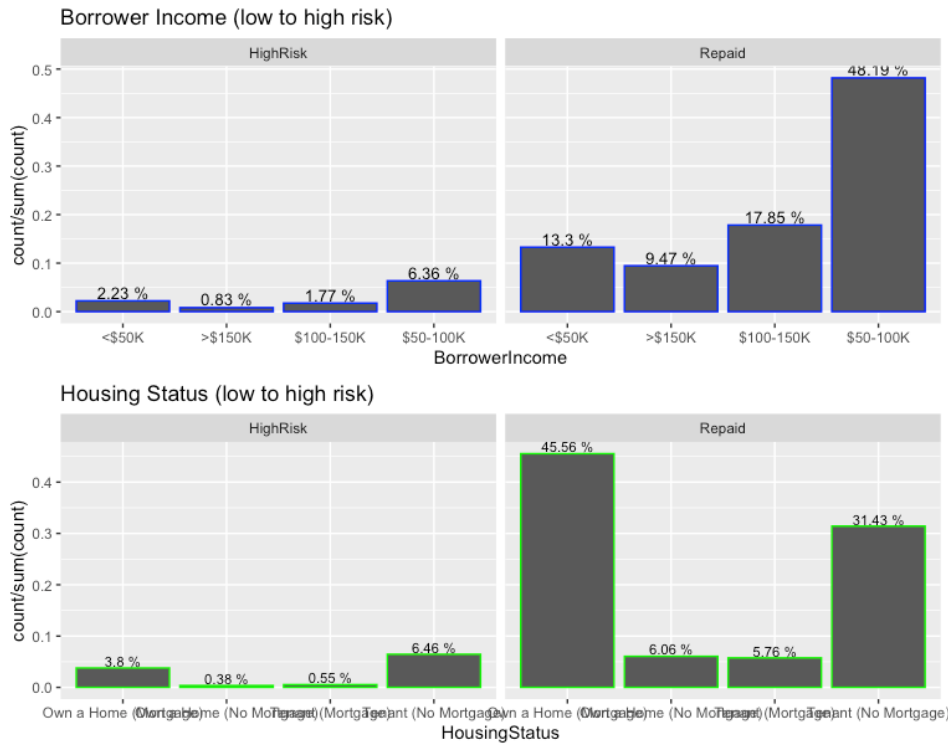


Figure 31:Borrower Income and Housing Status of high risk and repaid loans of RateSetter.

In the figure above shows, borrower income and housing status of the borrower in risky or repaid category. In the high risk loan category, most of the borrowers earn an income between 50K-100K, which is 6.36% of the all borrowers in that category. This has similarities with the repaid loans category. Most of the borrowers who have successful paid their loans, earn in the range of 50K -100K.

Also a similar trend in the housing status of the borrowers. The borrowers in the high risk loan category do not have a mortgage and they are a tenant. On the other hand, the borrowers who have repaid their loans, have a house and a mortgage. This suggests having a home is a good indicator of a loan been successfully repaid compared to borrowers who do not have a home or a mortgage.

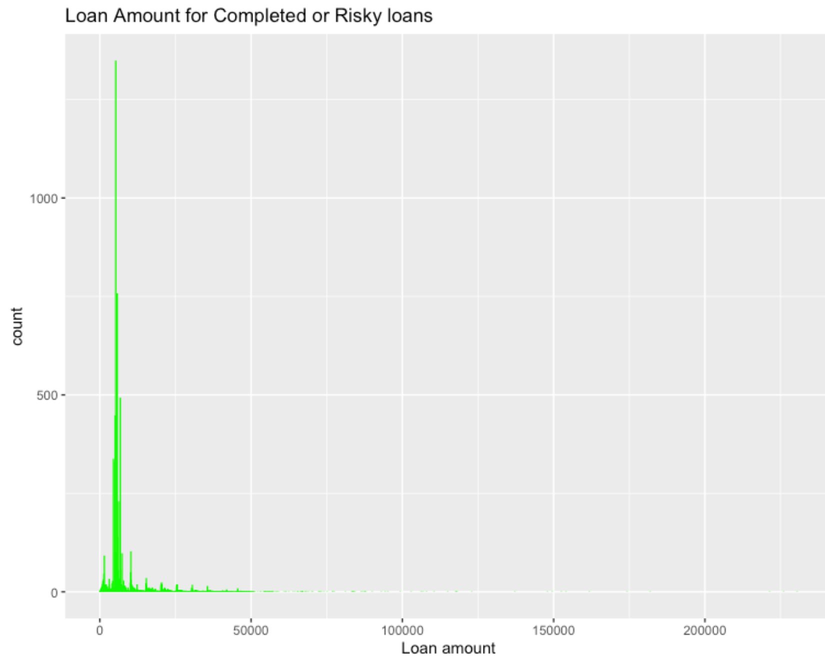


Figure 32:Total loan amount distribution of high risk or repaid loans of RateSetter.

Feature	Repaid	High risk
Minimum amount	110	219
Average amount	12120	14615.3
Maximum amount	230473	122965
Median	11842.5	8413

The average loan amount borrowed in the risky category is higher compared to loan repaid category. Higher amount of loan borrowed could be a possibility of not been able to pay back.

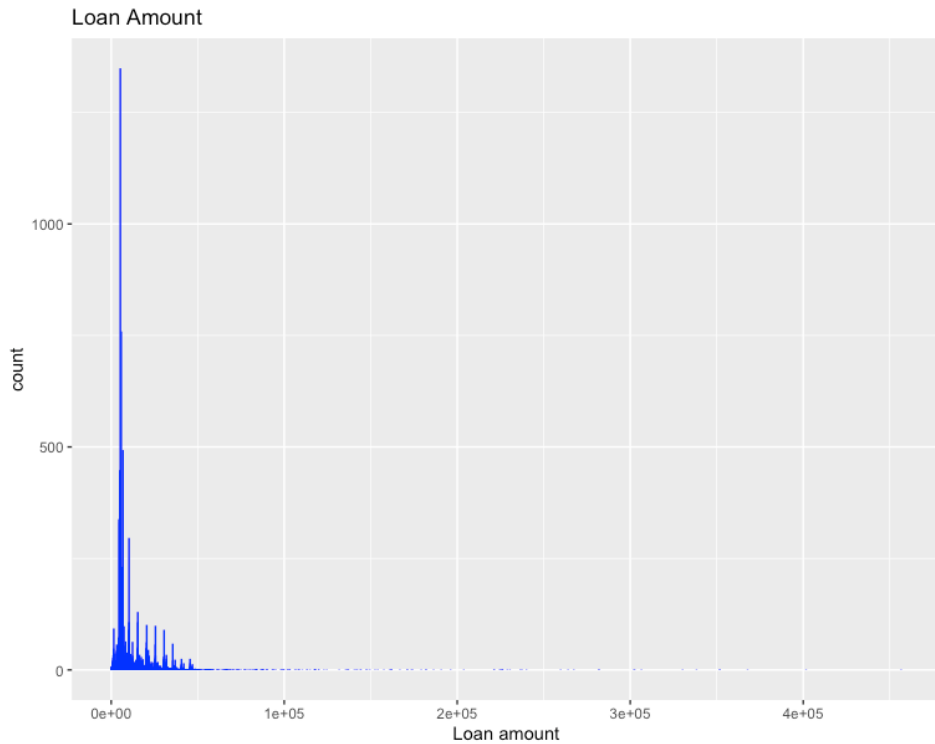


Figure 33:Total loan amount distribution of all loans of RateSetter

Feature	Loan Amount (in pounds)
Min amount	110
Avg. amount	13843
Max amount	456609
Median	10501

The average loan amount borrowed is 13843 pounds which is lower than the high risk category loans of 14615.3.

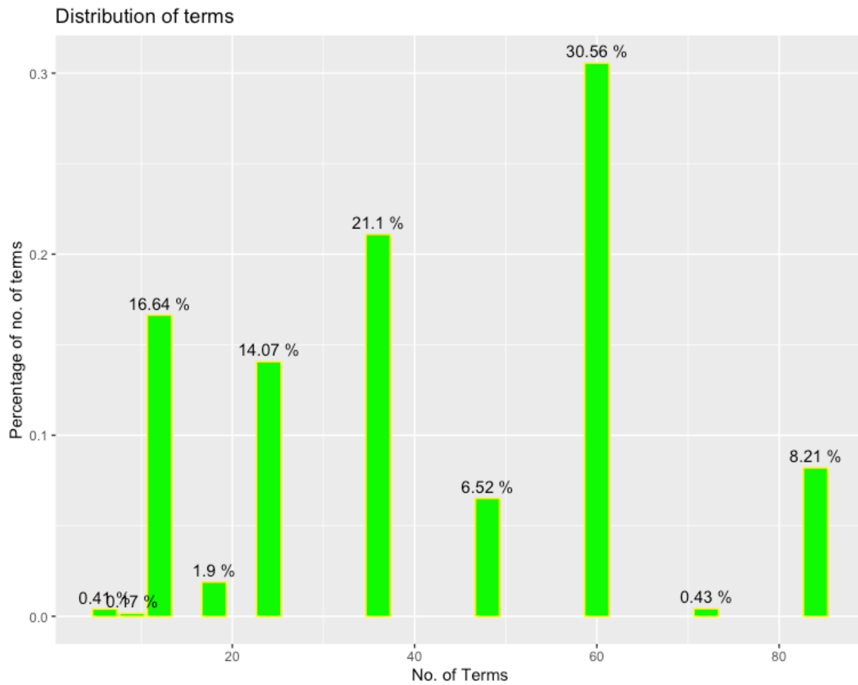


Figure 34:Distribution of terms of all loans of RateSetter

Feature	Loan term (in months)
Min term	6
Avg. amount	42.02
Max amount	84
Median	36

The most popular term for borrowers of loan is 60 months.

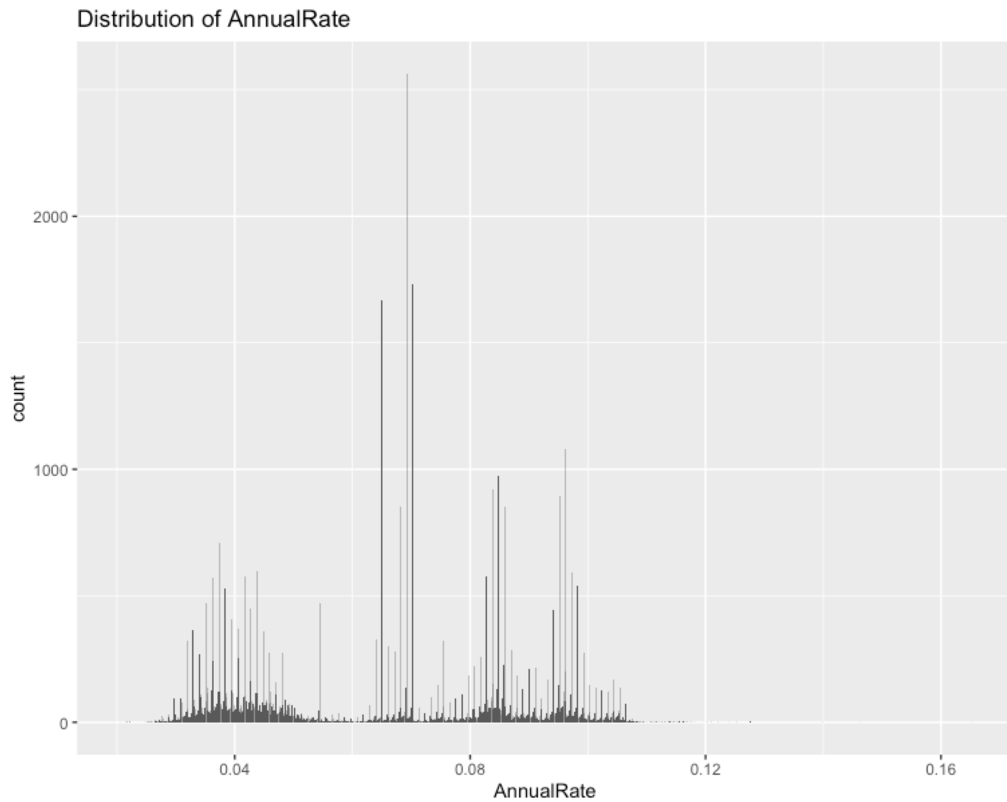


Figure 35:Distribution of annual rate of all loans of RateSetter.

Feature	Annual rates
Min term	2%
Avg. amount	6.6%
Max amount	16.5%
Median	6.9%

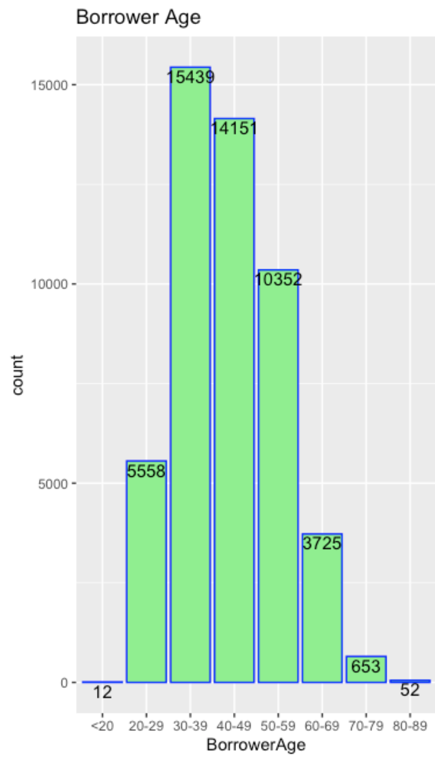


Figure 36:Distribution of Borrowers Age of RateSetter

The graph above suggests the maximum number of loan takers are in the age range 30-39 years. These are middle age applicants who have certain small requirements of money which is been met with these loans.

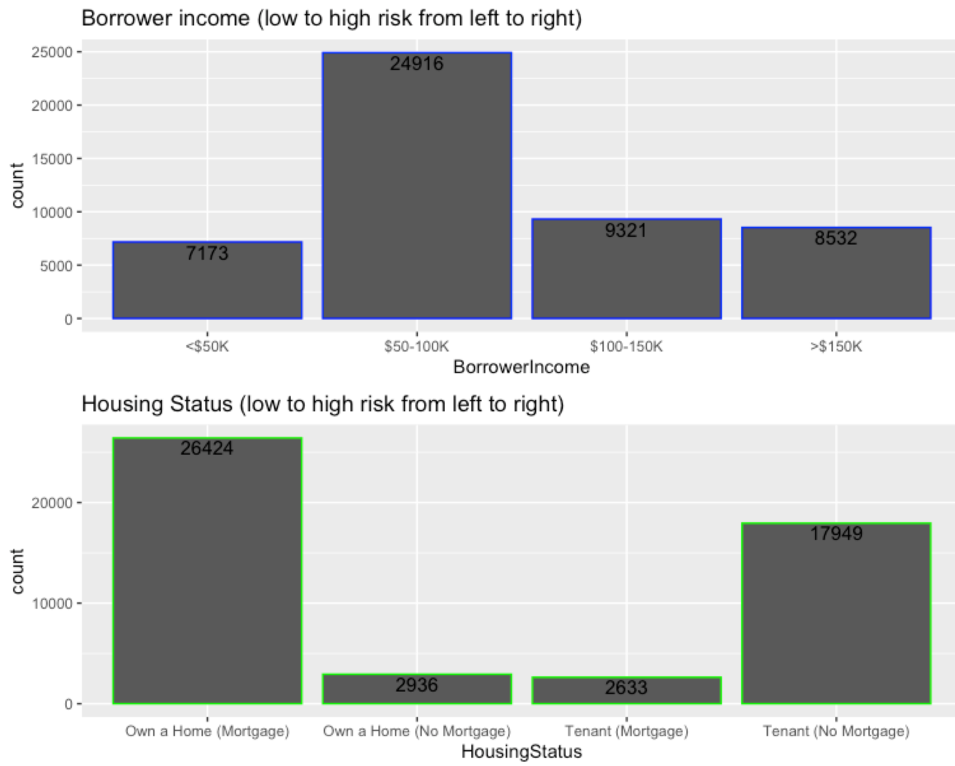


Figure 37: Borrower's Housing status and income range distribution of all loans.

Most of the borrowers have an annual income in the range of 50K -100K and majority of them owns a house with a mortgage.

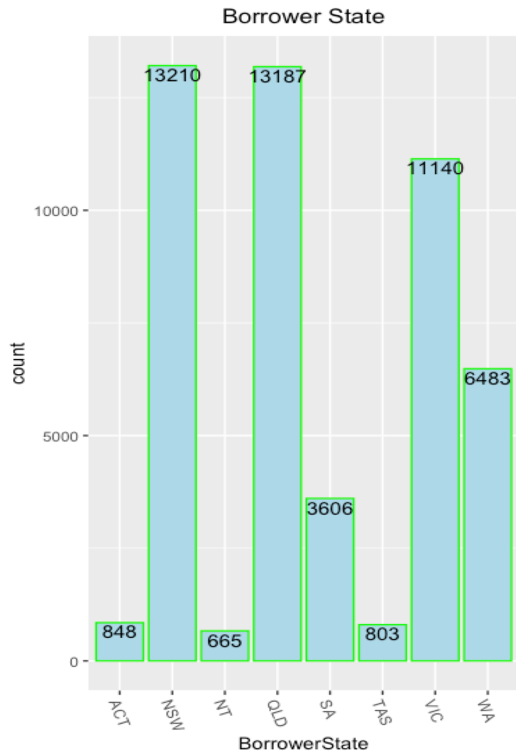


Figure 38:Distribution of Borrowers of loans from each state.

The most popular location for borrower of loans is from Australia's New South Wales and Queensland.

Code	State
ACT	Australian Capital Territory
NSW	New South Wales
NT	Northern Territory
QLD	Queensland
SA	South Australia
TAS	Tasmania
VIC	Victoria
WA	Western Australia

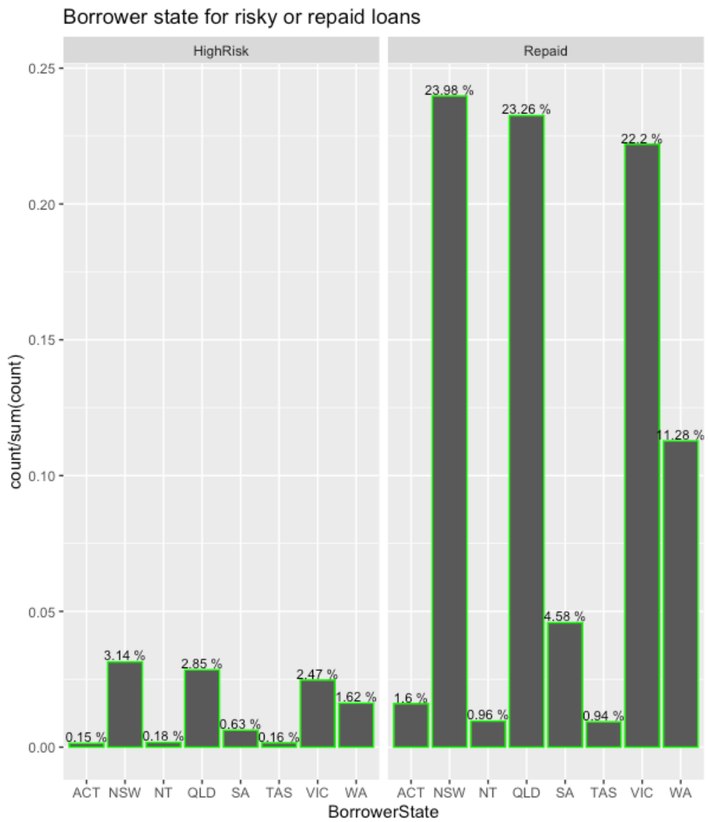


Figure 39:Distribution of Borrower of repaid or high risk loans from each state.

From the graph above we can see, the highest rate of default is observed in the state of New South Wales, Australia. Then comes Queensland and Victory. But the pattern is similar in repaid loans category. The proportion of high risk loans in the state of New South Wales is 1 in 13 borrowers.

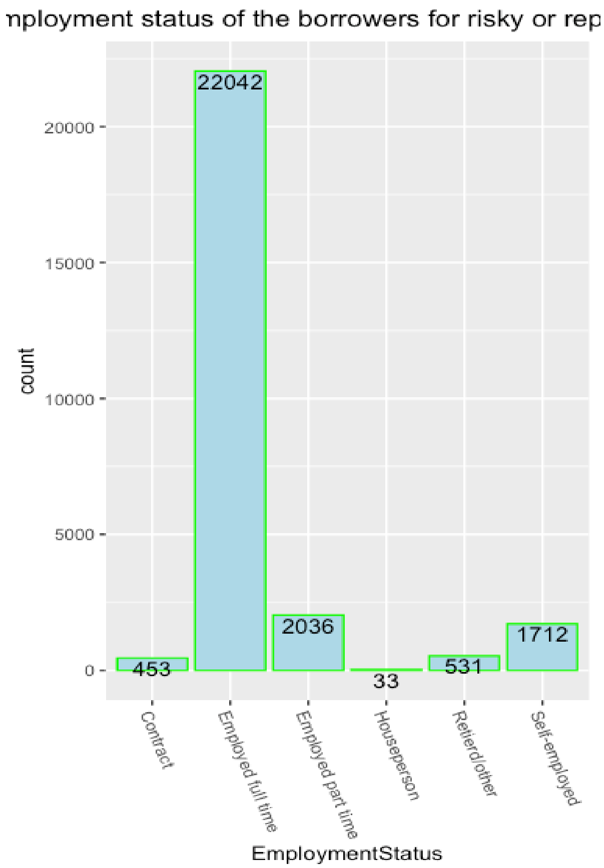


Figure 40: Employment status of all Borrowers for high risk and repaid loans.

Most of the borrowers are fully employed in the risk or repaid category. A small portion of borrowers are self-employed as well as partly employed.

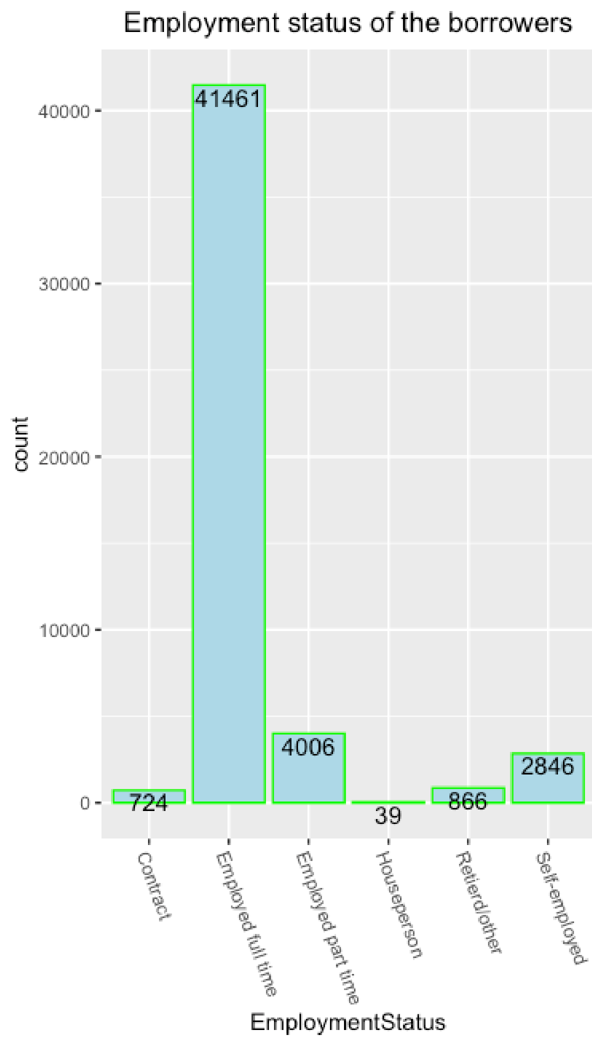


Figure 41: Employment status of Borrowers of all loans of RateSetter

Borrowers in all loan category, have full employment as was the trend seen above in the risk or repaid category.

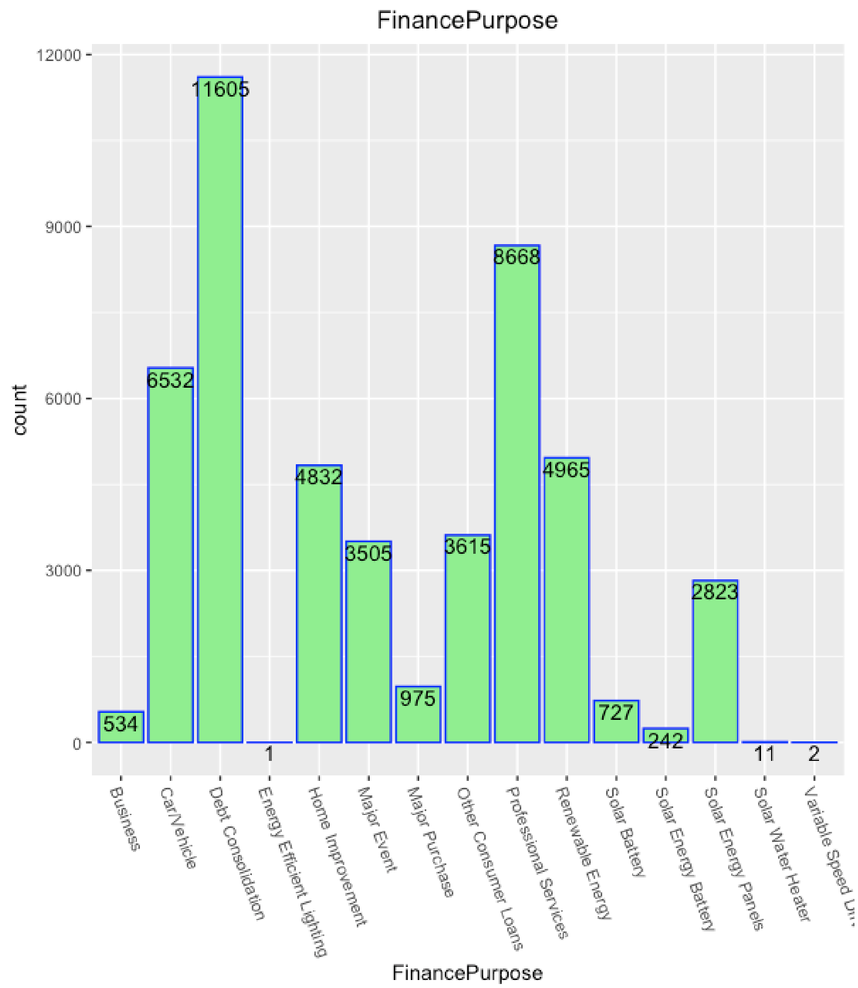


Figure 42: Purpose of loan for Borrowers of RateSetter

The most common purpose to get a loan for the borrowers are debt consolidation, professional services and to buy a car or a vehicle.

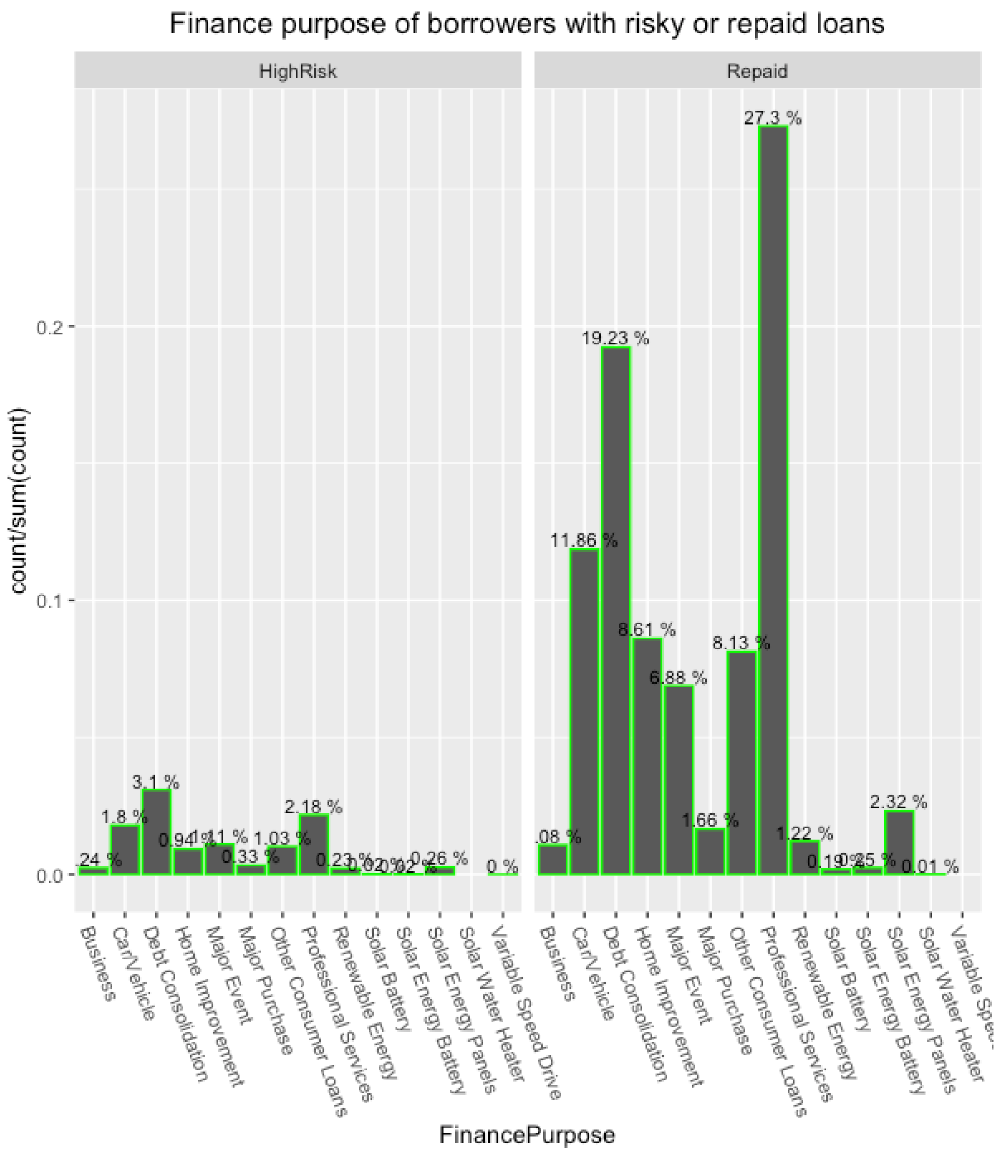


Figure 43: Purpose of loan for high risk and repaid loans of borrowers

In case of high risk loans, borrowers get loans for debt consolidation and professional services. This trend continues in repaid loans category.

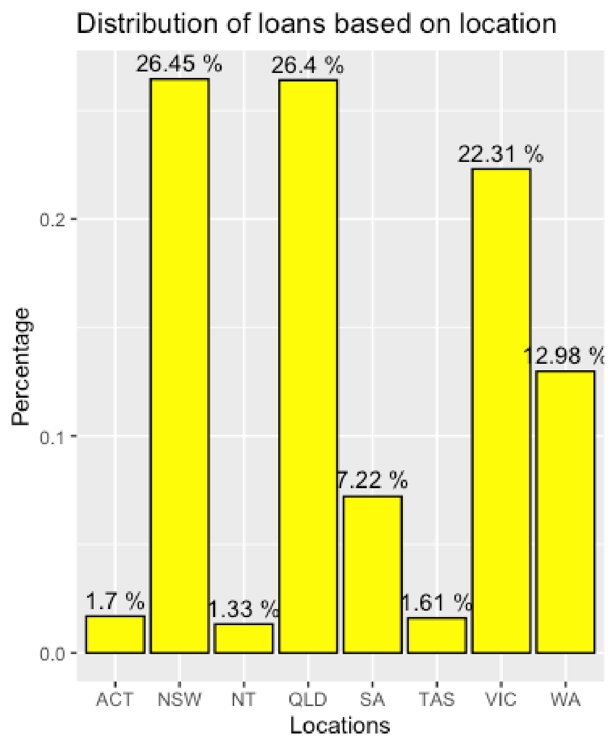


Figure 44 Distribution of all loans from particular states

Code	State
ACT	Australian Capital Territory
NSW	New South Wales
NT	Northern Territory
QLD	Queensland
SA	South Australia
TAS	Tasmania
VIC	Victoria
WA	Western Australia

Most of the loans are from New South Wales and Victoria of Australia.

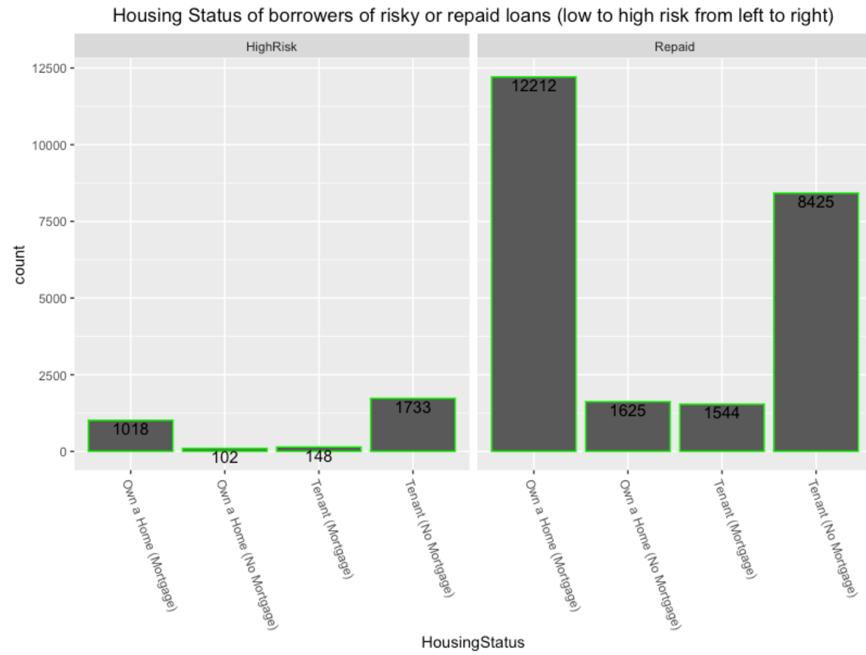


Figure 45: Housing status of Borrowers at RateSetter

High risk loan borrowers are tenants with no mortgage whereas repaid loan borrowers have a house with a mortgage.

Data Analysis and Modeling

The intent of building the model is to determine the conditions for the possibility of default loans.

Zopa

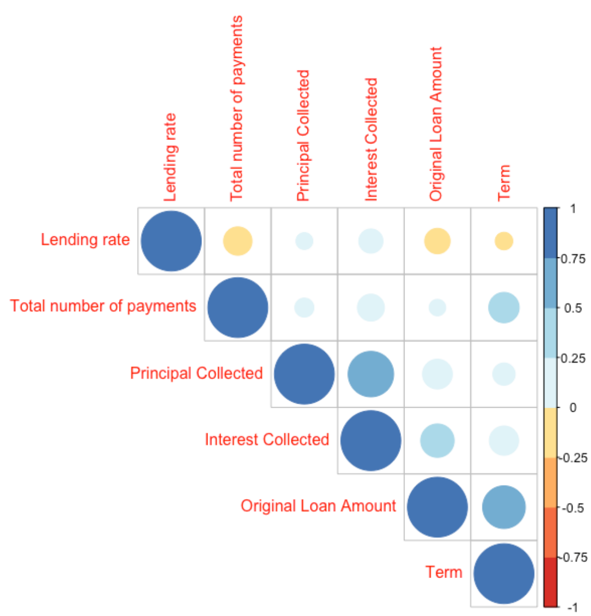


Figure 46: Correlational matrix for all loans of Zopa

The correlation matrix of Zopa for all loans show that, Total number of payments have a negative correlation with lending rate. It means that when the lending rate is higher, the possibility to make payments for the loan becomes less. So loans with higher lending rates will find it difficult to make payments to repay the loan.

The principal amount collected is slightly positively related to lending rate. So higher lending rates may slightly influence the principal amount collected. The same goes with the interest collected. But the lending rate is more related to interest collected compared to principal collected. It shows similar trend.

The loan amount is slightly negatively related to lending rate. This means that if the amount of loan applied for is higher, the loan may get a lower lending rate. This is more favourable to borrowers compared to investors.

The term of the loan is negatively related to lending rate. The longer term of loan results into a slight lower lending rate. This is commonly seen in loans provided around the world, when the interest rates are lower, if the loan is applied for a longer period of time.

The amount of the principal collected is slightly positively related to no. of payments made. This is quite intuitive. More payments made will result into more principal collected.

Similarly, the amount of the interest collected is higher when the no. of payments made is higher.

The longer the term of the loan, the higher are the number of payments collected over time.

Principal collected is strongly positively related to interest collected. If the principal collection is more so will be the interest collection.

The higher the loan amount, the higher will be the principal collected.

The principal collected is slightly positively connected to term of the loan. It means that if the loan is applied for a longer period, the amount of principal collected is more.

Similar is the case for interest collected. If the loan amount is higher, the interest collected will be higher. It is also positively related to the term of the loan.

The bigger the amount of the loan applied for, the term of the loan is higher correspondingly.

So, important relationship is observed between lending rate and total number of payments made, principal collected and interest collected and loan amount with the term of the loan.

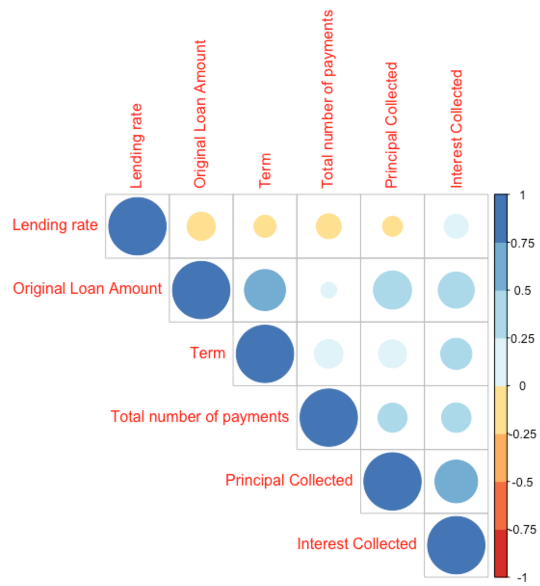


Figure 47: Correlational matrix of risky or completed loans of Zopa

In case of risky loans or completed loans, we see that lending rate is negatively related to amount of the loan applied for. So higher the loan amount, lower is the lending rate.

The lending rate is also slightly negatively related to term of the loan, principal and total no. of payments collected. The longer the loan is applied for the lower is the lending rate. There is an inverse relationship between principal collected as well as interest collected.

The longer the loan is applied for the longer is the loan amount. The loan amount is positively related to principal and interest collected.

Here too, there is a strong relationship between principal collected to interest collected.

An important distinction between the correlation matrix of all loans compared to risky or completed loans is the negative correlation between lending rate and principal collected as well as interest collected in case of risky or completed loan.

Next I looked at the relationship between lending rate with the rate of default. I created a list on indices where the defaulters data matches up with all loan data. I created a table of lending rates with the proportion of the default loans. The graph below shows the relationship between defaulted with others based on lending rate.

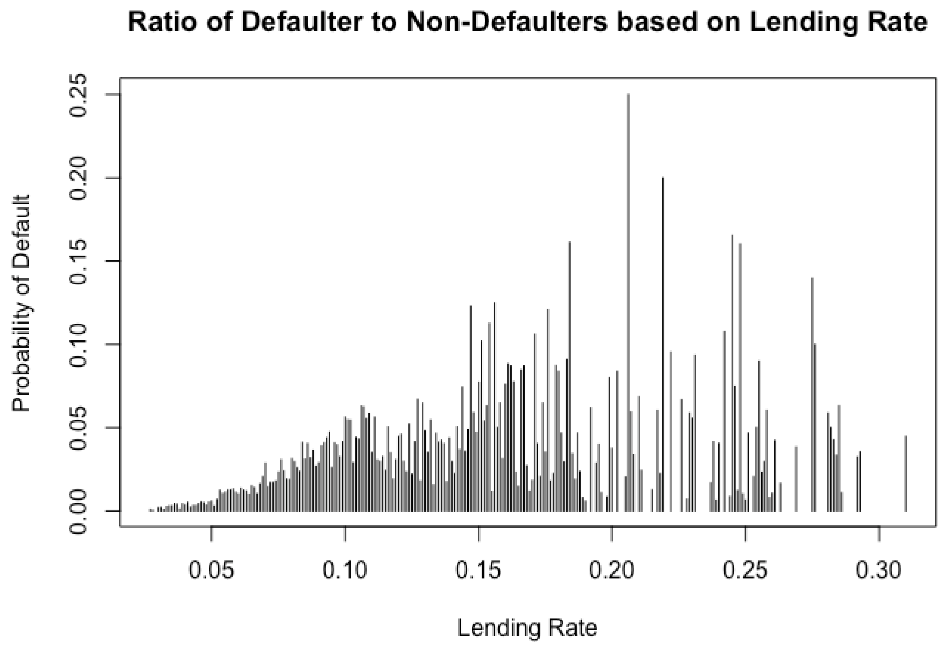


Figure 48: Ratio of defaulters to non-defaulters of loan borrowers of Zopa

The no. of observations is relatively smaller at around 9000. This may account to instability in larger loans as higher rates are given to low number of loans.

I used the logarithmic scale to give me a better view of the defaults per lending rate.

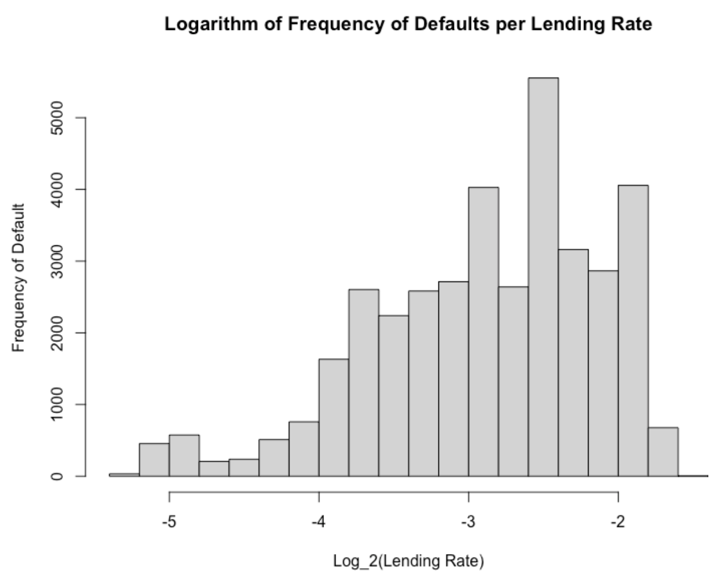


Figure 49: Frequency of default per rate (log)

Next, I looked for the amount of money as loans given out per month. I wanted to see if there is some economic condition present such as an inflation or a strong up market. I plotted the disbursal value i.e. amount of money out per month. So the following graph gives the loan amount with the disbursal date relationship.

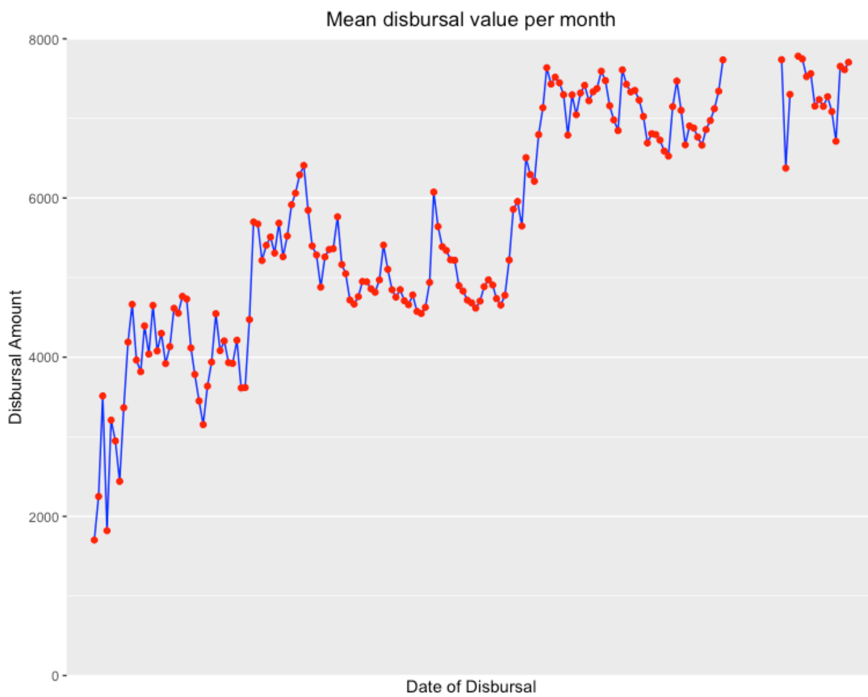


Figure 50: Mean disbursal amount per month of loans of Zopa

I smooth out the graph to see a better trend.

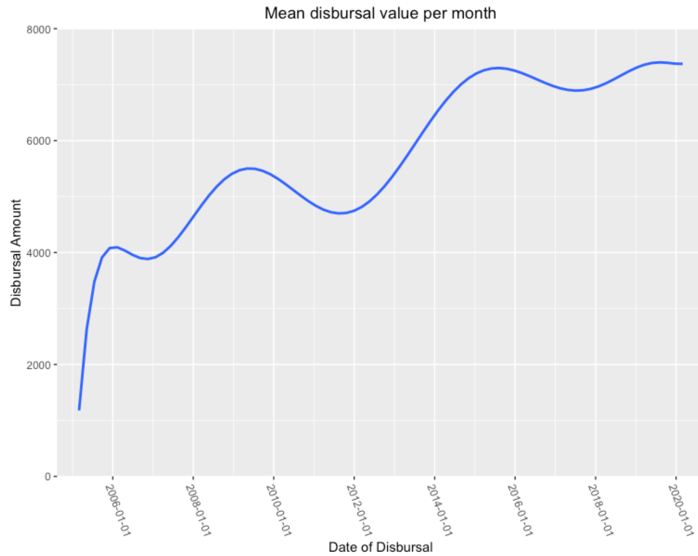
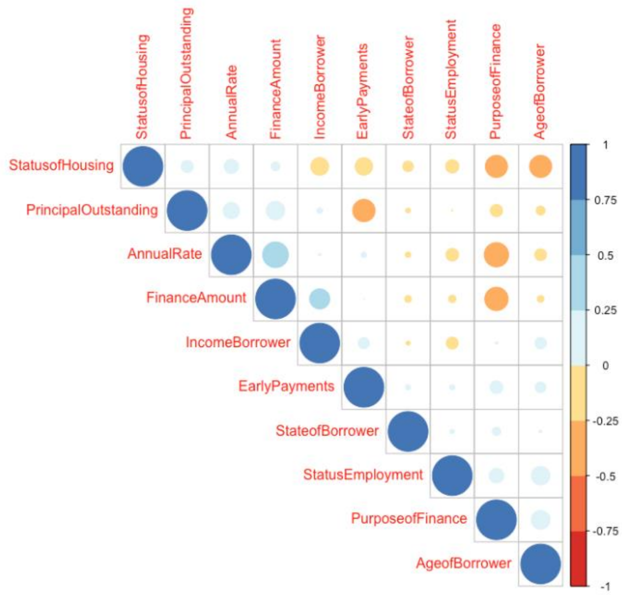


Figure 51: Amount of loan per month of loans of Zopa

The figure above gives the amount of the loans given out has increased over the years. There are some heights and tufts in the amount of the loan which can be interpreted as effect of the financial markets and overall environment.

Next we look at the correlation matrix of variables in the RateSetter loans.

RateSetter



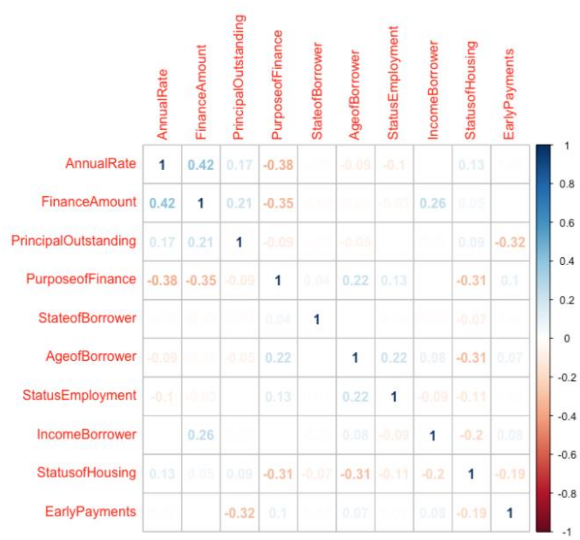


Figure 52:Correlational matrix of RateSetter.

Taking a look at the correlation matrix of RateSetter for all loans, the annual rate is positively related to the amount of the loan. The higher the loan amount, higher is the interest rate. This trend continues for the principal outstading amount with the annual rate.
There is a strong positive relation with the loan amount and the principal outstanding amount.

We can see a similar pattern even for risky or completed loans of RateSetter. The annual rate is positively related to amount of loan.

The housing status of the borrower of loan is negatively correlated to age of the borrower.

Housing Status	Mapping
Owns a Home (Mortgage)	1
Owns a Home (No Mortgage)	2
Tenant (Mortgage)	3
Tenant(No Mortgage)	4

Borrower Age	Mapping
<20	1
20-29	2
30-39	3
40-49	4
50-59	5
60-69	6
70-79	7
80-89	8

The higher the mapping no. on housing status of the borrower, the lower is the mapping no. on borrower’s age. This is intuitively correct. A young person in the age range of <20 years, does not own a house and probably does not have a mortgage to pay. The young person is most likely to be a tenant. On the other hand, an elderly person in age range of 80-89, own a house and most likely a mortgage or no mortgage.

Next we look at the finance amount compared to the date of loan provided. I changes the date into months.

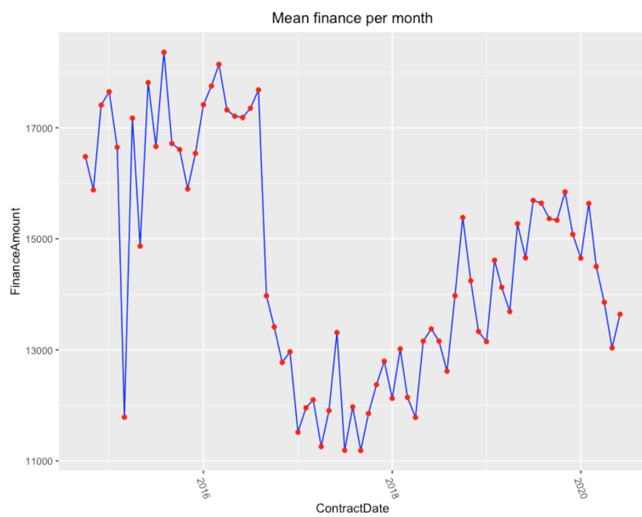


Figure 53 Mean finance amount per month of RateSetter

I smoothed out the graph to see a trend

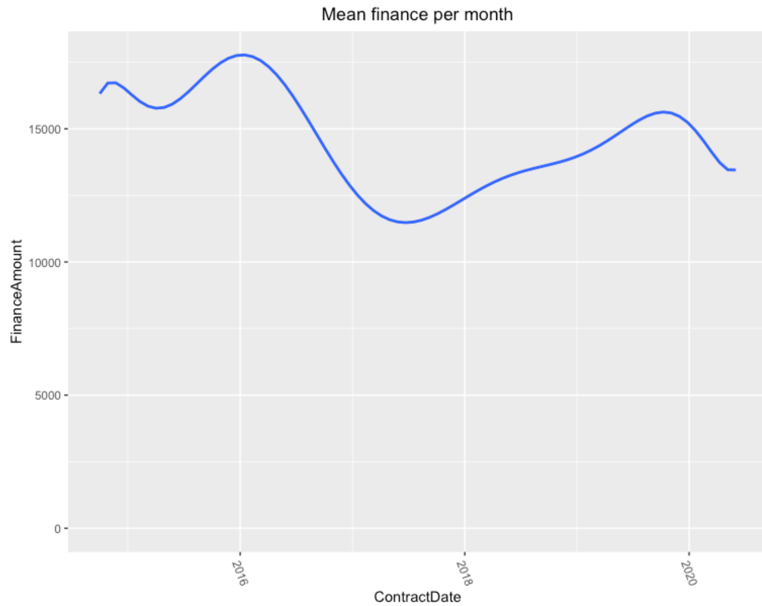


Figure 54 Mean of finance amount per month for RateSetter(smooth)

There is a significant change in in amount of the loan provided over time. It deepen a bit in the middle. A reason could be due to financial situations. The lending amount increase again when the situation improved over time.

Model Building and Interpretation

I wanted to check how some critical variables could possibly effect the nature of the loan from being successful to being risky. To do that I created a logistic regression model. Firstly, I created a variable called the Status. The variable has two values 0 and 1. 1 is for High risk loans and 0 is for completed loans.

Zopa

In the correlational matrix, we have seen that some variable have a relationship with the some other variables in the dataset. Now I want to check out of these variable which one has implications on a loan been risky or successful. I choose 4 variables from the data set. Lending rate, Original loan amount, principal collected and interest collected. I want to see if these variables have impact on status of the loan and by how much.

I ran a linear logistic function on the data. I observed the following results.

There is a negative coefficient for interest collected. It predicts that if all other variables are kept constant, the higher the interest collected is less likely to a risky loan. The positive coefficient for the lending rate, principal collected and loan amount suggests that the higher these variables are there is a greater possibility of the loans to be risky. Sa loans with higher lending rates may be more risky to default compared to lower lending rates. The higher the loan amount, the higher are the chances of the loan to be risky. A important relationship is observed in principal collected. The higher the principal collected may result the loan to be more risky to default.

Next I wanted to see the impact of each of these variables on the status of the loan. If we keep all other factors constant, if the lending rate goes up by 1 point, the changes of the loan to be risky is 1.8%. Similarly, keeping other variables constant, if the loan amount increases by a point, then risk of the loan increases by 5.3%.

The principal collected is positively related to the loan being risky. If there is a single point increase in the value of principal collected, the risk odds of the loan increase by 5%, keeping all other variables constant.

And in case of interest collected, if there is 1 point increase in the value of interest collected, the loan becomes more secure by 5.4%, keeping all other variable constant.

RateSetter

Firstly, looking at the data, it shows some NA (Not Applicable) values. I filled them median values to generate continuous values for the model. I also created a variable called "Status" which can hold values of only 1 or 0. For high-risk loans it assigns 1 and for completed loans it assigns 0. From the analysis above, it seems some variables are more important than others. From the dataset, variables such as Annual Rate, Principal Outstanding Borrower Income, Borrower Age and Finance Amount seems to have a better change to predict whether a loan could default or not in the future. So, I used these variables as my predictor variables in the model.

I made sure that I did not have any undefined values in the predictor variables. If there were any, I changed it with the median values. I tried to fit the logistic model with the values.

The results of the model shows negative coefficients for Finance Amount and Borrower Income where as positive coefficient for Principal Outstanding, Annual rate as well as Borrower Age. The negative coefficient for Finance Amount predictor suggests that, when all other variables are equal, The higher the loan amount is less likely to be of high risk. The same goes for Borrower's income too. Taking all other factors to be same, if a borrower's income is higher, its less likely to be a high risk loan for the borrower.

On the other hand, it turns out if the Annual rate of a loan is high, the chances of the loan being default increases. This can be also said for the principal outstanding. If the principal outstanding of a borrower is high, the loan has a higher possibility to go bad with all other variable be same. Another surprising point, turns out to be the borrowers age, if the borrowers age is higher, changes are the loan may not be repaid. This fall in the high risk category.

Furthermore, if we consider all other factors constant, for a single point increase in annual rate, we can expect the loan to be a high risk by 5.4%. With one point increase in principal outstanding for a loan, the chances of the loan being high risk increase by 2.71%. Also with the case of borrowers age, the loan to be more riskier is higher if there in increase in age of the borrower. For borrowers with the age range of 20-29, the loans are riskier by 3.3%, in the age range of 30-39, the loans are risker by 1.9%, in the age range of 40-49, the loans are riskier by

2.3%, in the age range of 50-59, the loans can be riskier by 2.3%, I. the age range of 60-69, the loans becomes riskier by 1.9%, in the age range of 70-79, the loans get riskier by 2.3% and in the age range of 80-89, the loans get riskier by 2.6%.

On the other hand, Taking all other factors to be constant, if there is a single point of increase in amount of the loan, the loan becomes less likely to be high risk by 1.7%. Even when the borrowers income increases by one point, the chances of the loan becoming high risk reduces, for the income range of 50K-100K, the risk decreases by 3.3%, for the income range of 100K-150K, the risk decreases by 5% and for the income range of above 150K, the risk decreases by 5.7%.

Comparative study

This section will compare findings from Zopa and RateSetter in terms of similarities and dissimilarities and try to interpret the reasons behind it. Both the peer-to-peer companies operate in different countries. Zopa operates in UK where RateSetter operates in Australia.

The loan book of Zopa has less variables compared to that of RateSetter. Some critical variables as the credit rating or credit score was missing from the loan book. Although, these P2P platform loan book used to have details like this but over time they have decided to remove personal credit information from the loan book. Having variables like credit details could have getting the existing study more depth. The credit data on individuals are not available in public without any assist from the credit companies. So these variables are outside the scope of this study.

The rate of default in case of Zopa is 5.11% whereas RateSetter has a default rate of 2.83%. Although, the length of observations from the time of inception is higher in case of Zopa from 2005-2020 but in case of RateSetter, it is 2014-2020. The proportion of high risk loans in Zopa is 12.59% compared to 11.19% at RateSetter.

The number of locations for the loans in case of Zopa is much larger all around UK compared to RateSetter, which dominated in 4 major states.

The loan amount given by Zopa is much more uniform and spread around over a period of time as compared to RateSetter which is much more concentrated. This has been the trend is both all loans as well as complete or risky loans.

There is a gradual growth in the amount of loans given per month i.e. the mean disbursal value per month is more in case of Zopa compared to RateSetter over the years. Both have ups and downs in loan amount but Zopa have shown a better performance compared to RateSetter.

The interest rates for both Zopa and RateSetter have been comparatively higher in the last 4 years from 2017-2020 for all loans. The interest rates for the high risk and completed loans show similar patterns in distribution over the years.

There seems to be more similarities than differences in the two peer-to-peer platforms.

Conclusion

With the advent of Web 3.0 and blockchain technology, more confidence has been seen in the P2P lending marketplace. New start-ups are making use of these high secure technologies to bring in the P2P market. Lenders and investors feel more secure with their investments. These technologies will revolutionise the traditional lending institute like the banks. It's been 16years from the inception of P2P lending in the world. This industry is still in its growing phase.

Better legislation, rules and connected ecosystem is expected in the future. More third party companies are getting involved in this market like payments, data cleaning and sorting, research analysis are slowing coming into the ecosystem.

Analysing the data of the P2P platforms, the rates provided are in terms with the market sensitive environment. But there is a fear of artificial rates due to inflation which can cause problems in the market. Also, comparatively newer P2P platforms tend to present lower default rates to attract customers when can be a trap for investors and put a bad name for the peer-to-peer industry.

Conclusions drawn from the analysis of the two P2P platform are: Firstly, with the increase in the interest rate of the loans, the chances of the loan to be risky is higher in both the platforms. So higher annual rates will lead to loans to be risky and possibly default.

Secondly, when amount of principal of the loan collected is higher, the chances of the loan been more risky in case of Zopa on the other the hand, when the principal outstanding is higher, the loan is more risky in case of RateSetter. This is unusual with normal intuition.

Thirdly, the older the borrower, the loan becomes more risky but with higher income range of the borrower, the loan becomes less risky in case of RateSetter. So relatively younger, middle age borrower with high income makes the loans less riskier.

Fourthly, the higher the amount of the loan, less risky it becomes. This is also unusual observation in case of RateSetter.

There are many benefits of P2P lending platforms: Risk adjusted Return, Efficient asset class, uncorrelated, diversification and some risks such as performance risk, platform risk, market risk, liquidity risks.

Appendix

Table for global marketplace loan issuance

	US	UK	China	Australia
2010	1,0	0,0	0,0	0,0
2011	1,4	0,0	0,0	0,0
2012	2,4	0,0	0,0	0,0
2013	5,2	0,8	3,0	0,3
2014	12,0	2,3	8,9	0,5
2015	23,2	4,0	33,2	0,8
2016	36,7	7,0	73,8	1,6
2017	54,2	9,9	115,3	3,4
2018	75,2	13,1	126,3	5,7
2019	98,2	18,1	147,7	9,0
2020	122,1	24,1	127,8	12,3

Compounded Annual Growth rate = 123% (2010 – 2014), 51%(2015 – 2020)

Table for global awareness of peer-to-peer lending

Age	Used	Aware but not used	Unaware
18 -34	30%	18%	52%
35 - 55	13%	22%	65%
55+	2%	23%	75%
Total	15%	21%	64%

Mapping between Zopa and RateSetter

Plenti(RateSetter)	Zopa
Loan Term	Term
Annual Rate	Lending Rate
Finance Amount	Original Loan Amount
Principal Outstanding	Disbursal date
Finance purpose	Interest collected
Borrower state	Post Code
Borrower age	Principal collected
Employment status	Total no. of payments
Borrower income	Latest status
Housing Status	Date of default
Early payments made	
Repayment status	

Provider	Cumulative Total (in mill, 2017)	2016 Total (in million)
Zopa	2529	688
RateSetter	2010	665
LendingWorks	76	21

Zopa Loan Size	2012	2013	2014	2015	2016	2017
Average Loan	4989	5560	7249	7302	6918	6842
Min Loan	1100	1040	1010	1000	260	260
Max Loan	15190	21550	29640	29640	29990	29970
Zopa Borrower Rate	2012	2013	2014	2015	2016	2017
Average Rate	6,90%	5,62%	5,80%	7,83%	8,56%	9,20%
Min rate	5,12%	3,50%	2,00%	2,14%	2,49%	2,42%
Max rate	12,09%	13,90%	25,70%	24,75%	31,03%	31,62%

	Zopa Core	Zopa Plus
Rate	3,90%	6,10%

	Zopa Core	Zopa Plus
Safeguard Fund Protection	No	No
Early Access Fee	1%	1%
Minimum Investment	1000	1000
Loan Type	A* - C	A* - E

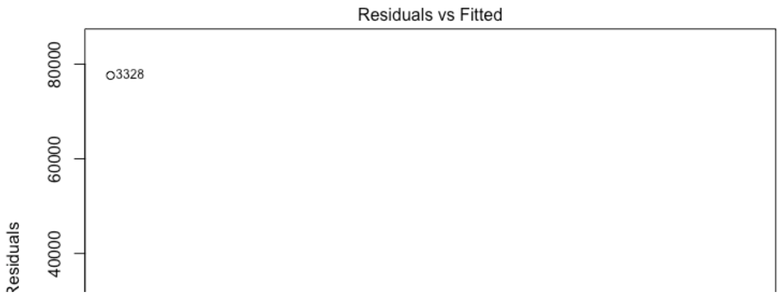
Zopa Core		
Year	Returns	Losses
2017	5,6%	0,6%
2018	4,1%	1,6%
2019	3,4%	2,8%

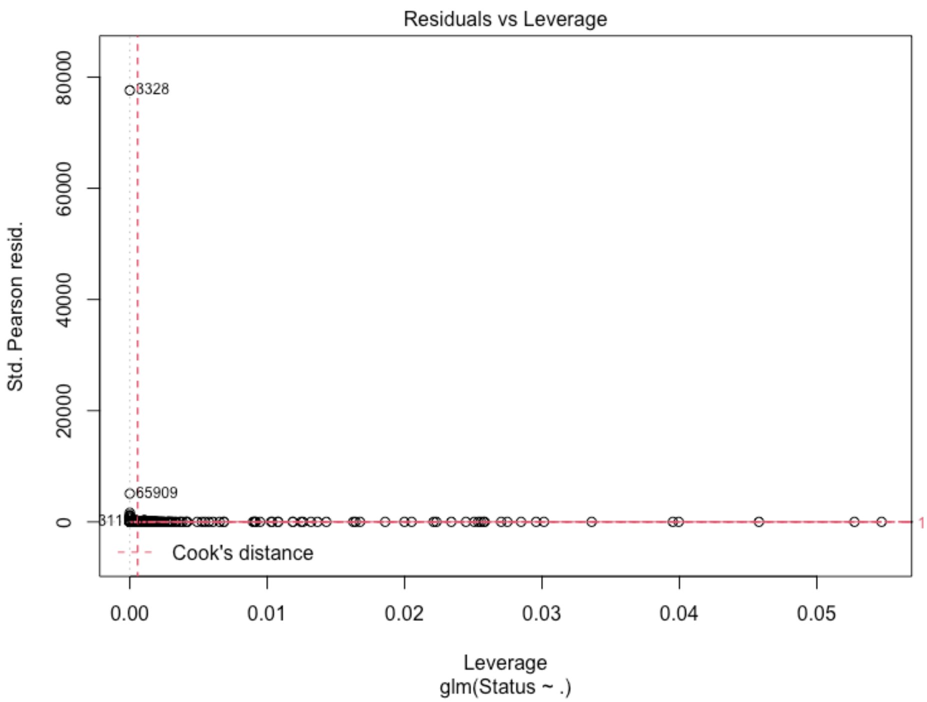
Year(RateSetter)	2016	2017	2018	2019	2020
Annaul return	4,30%	4,10%	4,40%	4,40%	3,80%
Underlysing return	4,20%	3,90%	4,30%	4,30%	3,70%
Provision Fund support	0,10%	0,10%	0,10%	0,10%	0,10%

Zopa Plus		
Year	Returns	
2016	9,4%	2,2%
2017	5,6%	5,6%
2018	4,1%	4,4%
2019	3,4%	4,1%
Zopa Safeguard Products		
Year	Returns	Net Losses
2013	4,70%	0,20%
2014	4,50%	0,50%
2015	4,60%	1,30%
2016	4,70%	2,00%
2017	4,30%	2,90%
2018	4,10%	1,90%
2019	4,00%	3,80%

Returns					
Year	Zopa	MSCI ACWI Index	Real Estate Total Return Index	Corporate bonds	Bank of England
2013	4,7%	22,91%	3,7%	8,0%	0,3%
2014	4,5%	4,64%	11,70%	-0,10%	0,2%
2015	4,6%	-2,39%	1,40%	-4,20%	0,3%
2016	4,7%	8,22%	2,50%	14,80%	0,6%
2017	4,3%	24,35%	22,50%	10,20%	1,0%
2018	4,1%	-9,15%	-6,20%	-3,30%	1,7%
2019	4,0%	26,70%			

Year (RateSetter)	2016	2017	2018	2019	2020
Amount lent (in million pounds)	626	625	709	786	345
Left to repay (in million pounds)	5	30	83	200	92


Residuals vs Fitted
Residuals
80000
60000
40000
3328


Residuals vs Leverage
Std. Pearson resid.
80000
60000
40000
20000
0
3328
65909
Cook's distance
0.00
0.01
0.02
0.03
0.04
0.05
Leverage
glm(Status ~ .)

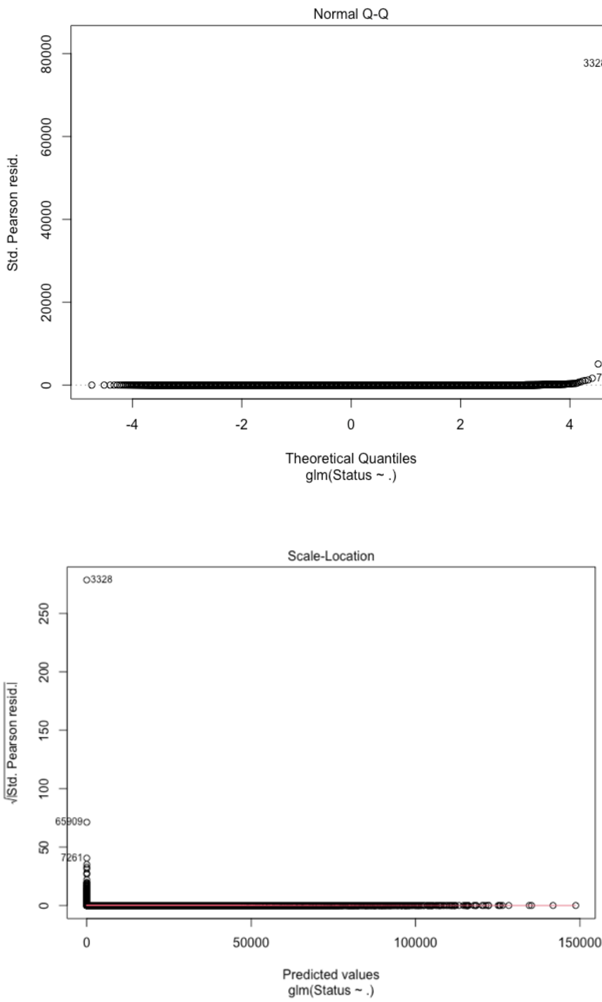


Figure 55: Zopa Logistic Regression plots

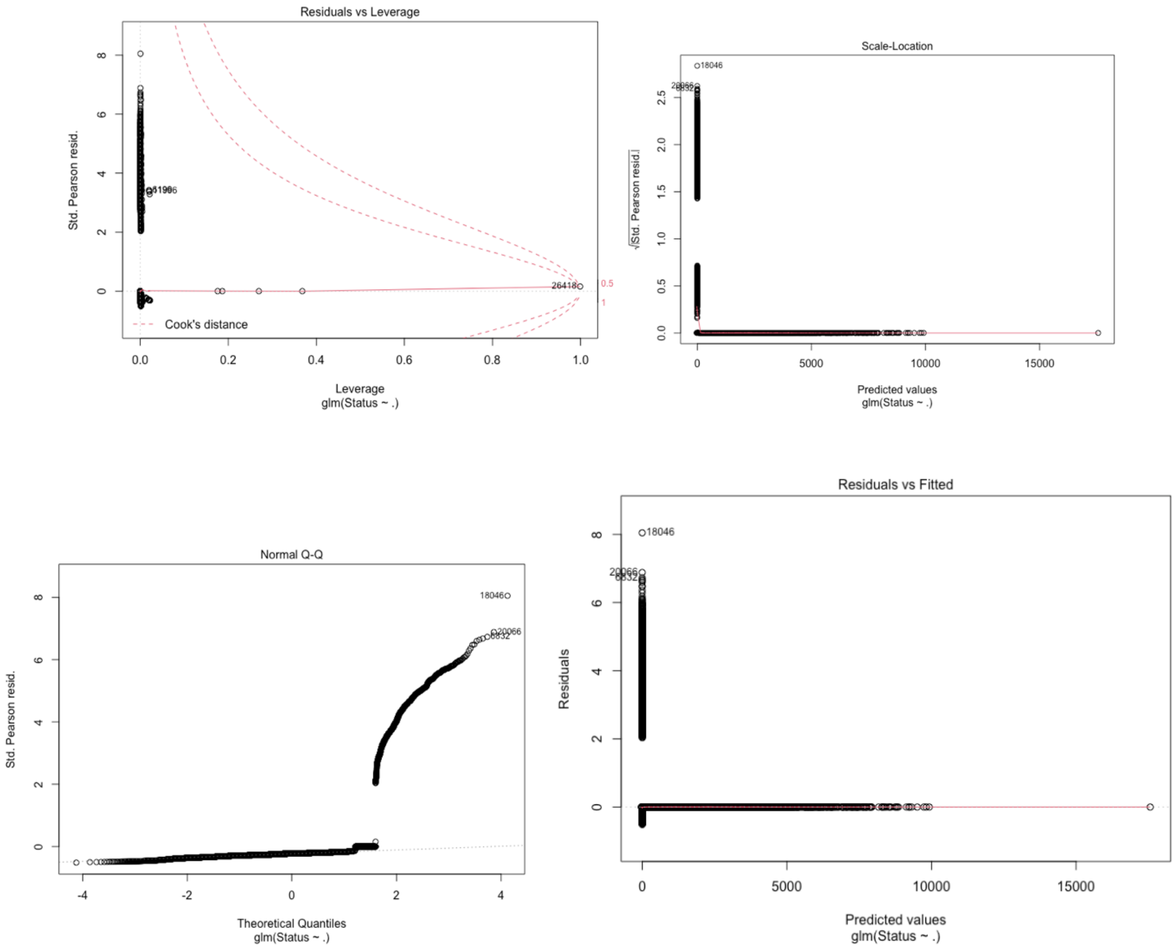


Figure 56: RateSetter Logistic Regression plots

Code used in R.

```
title: "A Comparative Analysis Of Credit Rating Model Used In Peer to Peer Lending"
author: "Rik Choudhury"
date: "January 23, 2021"
#Zopa loan book analysis

library(ggplot2)
library(tidyr)
library(dplyr)
library(ggthemes)
library(gridExtra)
library(RColorBrewer)
library(knitr)
library(GGally)
library(psych)
library('corrplot')
library(tidyverse)
library(scales)
library(RColorBrewer)
library(tidyverse)
library(grid)
library(ggthemes)
library(pander)
library(readxl)
Zopa_Loan_book <- read_excel("/Volumes/MacSD/Documents/Thesis/Credit Rating Modeling/Loan Book/Zopa Loan book.xlsm")
View(Zopa_Loan_book)

summary(Zopa_Loan_book)
glimpse(Zopa_Loan_book)
```

```
loan <- Zopa_Loan_book
dim(loan)
table(loan$`Latest Status`)
loan$`Latest Status` <- as.character(loan$`Latest Status`)
Completed <- c("Completed")
Late <- c("Late")
Default <- c("Default")
loan$`Latest Status`[loan$`Latest Status` %in% Completed] <- "Completed"
loan$`Latest Status`[loan$`Latest Status` %in% Late] <- "Late"
loan$`Latest Status`[loan$`Latest Status` %in% Default] <- "Default"
loan$`Latest Status` <- factor(loan$`Latest Status`,
                        levels = c("Completed",  "Active",
                              "Late", "Default"),
                        ordered = T)
table(loan$`Latest Status`)
```

Figure Latest Status

```
ggplot(aes(x= `Latest Status`, y = ..count../sum(..count..)), data = loan) +
  geom_bar(color="blue", fill="violet") +
  geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%"))
            , vjust=-0.5)+ggtitle("Distribution of loans")+
  ylab("Percentage of loans")+xlab("Loan Status")
```

Figure Terms

```
ggplot(aes(x= Term, y = ..count../sum(..count..)), data = loan) +
  geom_bar(color="yellow", fill="green") +
  geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%"))
            , vjust=-0.5)+ggtitle("Distribution of terms")+
  ylab("Percentage of no. of terms")+xlab("No. of Terms")+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))

sort(table(loan$PostCode), decreasing=TRUE)[1:20]
# create CompletedOrRisk column
loan$CompletedOrRisk<- ifelse(
  loan$`Latest Status` %in% c("Default","Late"),
```

```
  "HighRisk", as.character(loan$`Latest Status`))

sub_loan<- subset(loan, CompletedOrRisk!= "Active")
dim(sub_loan)
table(sub_loan$`Latest Status`)
```

Figure for risky or completed loans

```
ggplot(aes(x= CompletedOrRisk), data = sub_loan) +
  geom_bar(color="blue", fill="light blue") +
  geom_text(stat='count', aes(label=..count..), vjust=-0.5)+ggtitle("Distribution of loans")+
  ylab("Total number")+xlab("Completed or Risky")
```

Figure loan proportion

```
ggplot(aes(x= CompletedOrRisk, y = ..count../sum(..count..)), data = sub_loan) +
  geom_bar(color="blue", fill="light blue") +
  geom_text(stat='count',
          aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
          vjust=-0.5) + ggtitle("Proportion for each loan status")
  labs(x = "Loan Status", y = "Percentage")
```

Figure latest status proportion

```
ggplot(aes(x= `Latest Status`, y = ..count../sum(..count..)), data = loan) +
    geom_bar(color="blue", fill="light blue") +
    geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
            vjust=-0.5) + ggtitle("Proportion for each loan status")+
  labs(x = "Loan Status", y = "Percentage")

loan$`Lending rate`<-as.numeric(loan$`Lending rate`)
summary(loan$`Lending rate`)
describe(loan$`Lending rate`)
sort(table(loan$`Lending rate`), decreasing=TRUE)[1:5]
```

Figure lending rate histogram

```
ggplot(aes(x = `Lending rate`), data = loan) +
  geom_histogram(stat="count",binwidth = 0.0005)+ggtitle("Distribution of Lending Rates")+
  labs(x = "Lending Rate")
sub_loan$`Lending rate`<-as.numeric(sub_loan$`Lending rate`)
describe(sub_loan$`Lending rate`)
```

```
sort(table(sub_loan$`Lending rate`), decreasing=TRUE)[1:5]
```

Figure lending rate histogram for risky or completed loans.

```
ggplot(aes(x = `Lending rate`), data = sub_loan) +
  geom_histogram(stat="count",binwidth = 0.005) +
  facet_wrap(~CompletedOrRisk, ncol = 2)+ggtitle("Distribution of Lending Rates for Completed or Risky loans")+
  labs(x = "Lending Rate")
glimpse(sub_loan)

by(sub_loan$`Lending rate`, sub_loan$CompletedOrRisk, summary)
sub_loan$`Original Loan Amount`<-as.numeric(sub_loan$`Original Loan Amount`)

by(loan$`Original Loan Amount`, loan$CompletedOrRisk, summary)
describe(loan$`Original Loan Amount`)
sort(table(loan$`Original Loan Amount`), decreasing=TRUE)[1:5]




by(sub_loan$`Original Loan Amount`, sub_loan$CompletedOrRisk, summary)
describe(sub_loan$`Original Loan Amount`)
sort(table(sub_loan$`Original Loan Amount`), decreasing=TRUE)[1:5]

sub_loan$`Principal Collected`<-as.numeric(sub_loan$`Principal Collected`)
by(sub_loan$`Principal Collected`, sub_loan$CompletedOrRisk, summary)
sort(table(sub_loan$`Principal Collected`), decreasing=TRUE)[1:5]
table(sub_loan$`Principal Collected`)

table(sub_loan$`Interest Collected`)
sub_loan$`Interest Collected`<-as.numeric(sub_loan$`Interest Collected`)
by(sub_loan$`Interest Collected`, sub_loan$CompletedOrRisk, summary)
sort(table(sub_loan$`Interest Collected`), decreasing=TRUE)[1:5]

sub_loan$`Total number of payments`<-as.numeric(sub_loan$`Total number of payments`)
by(sub_loan$`Total number of payments`, sub_loan$CompletedOrRisk, summary)
sort(table(sub_loan$`Total number of payments`), decreasing=TRUE)[1:5]
summary(loan$`Total number of payments`)
```

```
sort(table(loan$`Lending rate`), decreasing=TRUE)[1:5]

loan$`Disbursal date` <- as.Date(loan$`Disbursal date`)

loan$year <- as.numeric(format(loan$`Disbursal date`, format = "%Y"))
sub_loan$year <- as.numeric(format(sub_loan$`Disbursal date`, format = "%Y"))

loan$year <- factor(loan$year)
sub_loan$year <- factor(sub_loan$year)
table(loan$year)
by(sub_loan$`Lending rate`, sub_loan$year, summary)
```

Figure lending rate per year

```
ggplot(aes(x = `Lending rate`), data = loan) +
  geom_bar(color="blue",binwidth = 0.005) +
  facet_wrap(~year) + ggtitle("Lending Rate by year")+
  labs(x = "Lending rate")
```

Figure lending rate for risky or completed loans

```
ggplot(aes(x = `Lending rate`), data = sub_loan) +
  geom_bar(color="green",binwidth = 0.005) +
  facet_wrap(~year) + ggtitle("Lending Rate by year for CompleteOrRisk loans")+
  labs(x = "Lending rate")

loan$`Date of Default` <- as.Date(loan$`Date of Default`)
table(loan$`Date of Default`)

sub_loan$`Date of Default`<- sapply(sub_loan$`Date of Default`,function(x){x=1})
loan1<-loan%>%select(`Lending rate`,`Date of Default`)
zopa_all$Date.of.Default <- sapply(zopa_all$Date.of.Default,function(x){if (x != ""){x=1} else if (x
== ""){x=0}})
```

Figure lending rate with trendline

```
ggplot(loan, aes(x=`Lending rate`)) +
  geom_histogram(stat="count", binwidth=.005, position="dodge") +
  geom_density(alpha=.3)
```

Figure lending rate with risky or completed loans with trendline

```
ggplot(sub_loan, aes(x=`Lending rate`)) +
  geom_histogram(stat="count",binwidth=.005, position="dodge") +
  geom_density(alpha=.3)
describe(loan)
describe(sub_loan)

summary(loan$`Original Loan Amount`)
```

Figure loan amount distribution

```
ggplot(aes(x = `Original Loan Amount`), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("Loan Amount")+
  labs(x = "Loan amount")
```

Figure loan amount for risky or completed loans

```
ggplot(aes(x = `Original Loan Amount`), data = sub_loan) +
  geom_histogram(stat="count", color = "green")+ ggtitle("Loan Amount for Completed or Risky loans")+
  labs(x = "Loan amount")

summary(loan$`Principal Collected`)
```

Figure principal collected histogram

```
ggplot(aes(x = `Principal Collected`), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("Principal Collected")+
  labs(x = "Principal amount")
```

Figure principal collected histogram with risky or completed loans

```
ggplot(aes(x = `Principal Collected`), data = sub_loan) +
  geom_histogram(stat="count", color = "green")+ ggtitle("Principal Collected for Completed or Risky loans")+
  labs(x = "Principal amount")

summary(loan$`Interest Collected`)
```

Figure interest collected distribution.

```
ggplot(aes(x = `Interest Collected`), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("Interest Collected")+
  labs(x = "Interest amount")
```

Figure interest collected for risky or completed loans

```
ggplot(aes(x = `Interest Collected`), data = sub_loan) +
```

```
  geom_histogram(stat="count", color = "green")+ ggtitle("Interest Collected for Completed or Risky loans")+
  labs(x = "Interest amount")

summary(loan$`Total number of payments`)
```

Figure Interest collected histogram

```
ggplot(aes(x = `Total number of payments`), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("Total number of payments")+
  labs(x = "`Total number of payments`")
```

Figure interest collected for risky or completed loans dist.

```
ggplot(aes(x = `Total number of payments`), data = sub_loan) +
  geom_histogram(stat="count", color = "green")+ ggtitle("Total number of payments for Completed or Risky loans")+
  labs(x = "Total number of payments")



ds$`Lending rate` <- as.numeric(ds$`Lending rate`)
ds$`Original Loan Amount` <- as.numeric(ds$`Original Loan Amount`)
ds$`Principal Collected` <- as.numeric(ds$`Principal Collected`)
ds$`Interest Collected` <- as.numeric(ds$`Interest Collected`)
```

Model

```
glimpse(sub_loan)
sub_loan$Status <- ifelse(
  sub_loan$CompletedOrRisk == "HighRisk", 1, 0)
ds <- sub_loan[, c("Status", "Lending rate","Original Loan Amount", "Principal Collected", "Interest Collected")]
sapply(ds, function(x) sum(is.na(x)))
model <- glm(Status ~.,family=binomial(link='logit'),data=ds)
summary(model)
print(model)
```

```
coefs <- coef(model)
exp(coefs)
(exp(coefs)-1)*100
```

```
ds$`Lending rate` <- as.numeric(ds$`Lending rate`)
d$`Original Loan Amount` <- as.numeric(d$`Original Loan Amount`)
d$`Principal Collected` <- as.numeric(d$`Principal Collected`)
d$`Interest Collected` <- as.numeric(d$`Interest Collected`)
d$`Total number of payments` <- as.numeric(d$`Total number of payments`)
ds$`Lending rate` <- ifelse(
  is.na(ds$`Lending rate`), 0, ds$`Lending rate`)
ds$`Original Loan Amount` <- ifelse(
  is.na(ds$`Original Loan Amount`), 0, ds$`Original Loan Amount`)
ds$`Principal Collected` <- ifelse(
  is.na(ds$`Principal Collected`), 0, ds$`Principal Collected`)
ds$`Interest Collected` <- ifelse(
  is.na(ds$`Principal Collected`), 0, ds$`Interest Collected`)
ds$`Principal Collected` <- ifelse(
  is.na(ds$`Principal Collected`), 0, ds$`Principal Collected`)
```

Correlation Matrix

```
loan$`Original Loan Amount`<-as.numeric(loan$`Original Loan Amount`)
loan$`Principal Collected`<-as.numeric(loan$`Principal Collected`)
loan$`Interest Collected`<-as.numeric(loan$`Interest Collected`)
#Corr with loan(Fig19)
corr_loan = cor(loan[,c(5,6,7,8,10,11)])
plot(corr_loan)
corrplot(corr_loan, type="upper", order="hclust",col = brewer.pal(n=8,name = "RdYlBu"))
```

Correlation Matrix with risky or completed loans

```
corr_sub_loan = cor(sub_loan[,c(5,6,7,8,10,11)])
corrplot(corr_sub_loan, type="upper", order="hclust",col = brewer.pal(n=8,name = "RdYlBu"))
```

```
sub_loan_new<-subset(sub_loan,select = c(11,13))
sub_loan_new<-subset(sub_loan_new,subset =sub_loan_new$`Date of Default`!="NA" )
sub_loan_new$`Date of Default`<-sapply(sub_loan_new$`Date of Default`,function(x){x=1})
sub_loan1<-subset(loan,select = c(11,13))
```

Figure lending rate with default

```
hist(log(sub_loan_new$`Lending rate`,2), xlab = "Log_2(Lending Rate)", ylab = "Frequency of Default", main = "Logarithm of Frequency of Defaults per Lending Rate")

zopa_date_amount <- subset(loan,select = c(5,4))
zopa_date_amount <- zopa_date_amount[order(zopa_date_amount$`Disbursal date`),]
zopa_date_amount$`Disbursal date`<-lubridate::ymd(zopa_date_amount$`Disbursal date`)

zopa_date_amount$month <- lubridate::floor_date(zopa_date_amount$`Disbursal date`,"month")


zopa_date_amount$`Original Loan Amount`<-as.numeric(zopa_date_amount$`Original Loan Amount`)
table(zopa_date_amount$`Original Loan Amount`)
zopa_aggregate <- aggregate(`Original Loan Amount`~ month, zopa_date_amount,mean)
names(zopa_aggregate) <- c("DisbursalDate", "OriginalLoanAmount")
```

Figure disbursal rate

```
ggplot(data = zopa_aggregate, aes(x=DisbursalDate, y = OriginalLoanAmount)) +
  geom_line(color="blue") +
  geom_point(color="red") +
  expand_limits(y=0) +
```

```
scale_x_date(expand = c(0,200),breaks = pretty(zopa_aggregate$`Disbursal date`, n = 10)) +
scale_y_continuous(expand=c(0,0),limits = c(0,8000)) +
xlab("Date of Disbursal") + ylab("Disbursal Amount")+
ggtitle("Mean disbursal value per month") +
theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure Disbursal rate

```
ggplot(data = zopa_aggregate, aes(x=DisbursalDate, y = OriginalLoanAmount)) +
stat_smooth(method = "lm", formula = y ~ poly(x, 9), size = 1,se=FALSE)+
expand_limits(y=0) +
scale_x_date(expand = c(0,200),breaks = pretty(zopa_aggregate$DisbursalDate, n = 10)) +
scale_y_continuous(expand=c(0,0),limits = c(0,8000)) +
xlab("Date of Disbursal") + ylab("Disbursal Amount")+
ggtitle("Mean disbursal value per month") +
theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

```
title: "A Comparative Analysis Of Credit Rating Model Used In Peer to Peer Lending"
author: "Rik Choudhury"
date: "January 5, 2021"
"RateSetter loan book analysis"
# Load all necessary packages
library(ggplot2)
library(tidyr)
library(dplyr)
library(ggthemes)
library(gridExtra)
library(RColorBrewer)
library(knitr)
library(GGally)
library(psych)
library('corrplot')
library(tidyverse)
library(scales)
library(RColorBrewer)
library(tidyverse)
library(grid)
library(ggthemes)
library(pander)
library(readxl)
Plenti_Loan_Book <- read_excel("/Volumes/MacSD/Documents/Thesis/Credit Rating Modeling/Loan
Book/Plenti Loan Book.xlsx")
View(Plenti_Loan_Book)
attach(Plenti_Loan_Book)
summary(Plenti_Loan_Book)
loan <- Plenti_Loan_Book
dim(loan)
```

```
table(Plenti_Loan_Book$RepaymentStatus)
loan$RepaymentStatus <- as.character(loan$RepaymentStatus)
loan$RepaymentStatus <- factor(loan$RepaymentStatus,
                          levels = c("Repaid",  "On Schedule", "Hardship",
                               "<30 days late", ">30 days late", "In Default"),
                          ordered = T)

Plenti_Loan_Book$CompletedOrRisk<- ifelse(
  Plenti_Loan_Book$RepaymentStatus %in% c("<30 days late",">30 days late", "Hardship","In Default"),
  "HighRisk", as.character(loan$RepaymentStatus))


sub_loan<- subset(Plenti_Loan_Book, CompletedOrRisk!= "On Schedule")
dim(sub_loan)
```

Figure loan status

```
ggplot(aes(x= RepaymentStatus, y = ..count../sum(..count..)), data = loan) +
  geom_bar(color="blue", fill="violet") +
  geom_text(stat='count',
        aes(label= paste(round(..count../sum(..count..)*100, 2),"%"))
        , vjust=-0.5)+ggtitle("Distribution of loans")+
  ylab("Percentage of loans")+xlab("Loan Status")
```

Figure terms

```
ggplot(aes(x= LoanTerm, y = ..count../sum(..count..)), data = loan) +
  geom_bar(color="yellow", fill="green") +
  geom_text(stat='count',
        aes(label= paste(round(..count../sum(..count..)*100, 2),"%"))
        , vjust=-0.5)+ggtitle("Distribution of terms")+
  ylab("Percentage of no. of terms")+xlab("No. of Terms")
```

Figure post code

```
ggplot(aes(x= BorrowerState, y = ..count../sum(..count..)), data = loan) +
  geom_bar(color="black", fill="yellow") +
```

```
  geom_text(stat='count',
          aes(label= paste(round(..count../sum(..count..)*100, 2),"%"))
          , vjust=-0.5)+ggtitle("Distribution of loans based on location")+
  ylab("Percentage")+xlab("Locations")
```

Figure loan dist. with risky or complete loans.

```
ggplot(sub_loan,aes(CompletedOrRisk)) +
  geom_bar(color="blue", fill="light blue") +
  geom_text(stat='count', aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
          vjust=-0.5) +
  labs(x = "Loan Status")+ggtitle("Distribution of loans")


loan$`AnnualRate`<-as.numeric(loan$AnnualRate)
summary(loan$AnnualRate)
describe(loan$AnnualRate)
sort(table(loan$AnnualRate), decreasing=TRUE)[1:5]
```

Figure annual rate histogram

```
ggplot(aes(x = AnnualRate), data = loan) +
  geom_histogram(stat="count",binwidth = 0.0005)+ggtitle("Distribution of AnnualRate")+
  labs(x = "AnnualRate")

sub_loan$AnnualRate<-as.numeric(sub_loan$AnnualRate)
summary(sub_loan$AnnualRate)
describe(sub_loan$AnnualRate)
sort(table(sub_loan$AnnualRate), decreasing=TRUE)[1:5]
```

Figure annual rate of risky or repaid loans

```
ggplot(aes(x = AnnualRate), data = sub_loan) +
  geom_histogram(stat="count",binwidth = 0.005) +
  facet_wrap(~CompletedOrRisk, ncol = 2)+ggtitle("Distribution of AnnualRate for Completed or Risky
loans")+
  labs(x = "AnnualRate")
```

```
by(sub_loan$AnnualRate, sub_loan$CompletedOrRisk, summary)

sub_loan$FinanceAmount<-as.numeric(sub_loan$FinanceAmount)
by(sub_loan$FinanceAmount, sub_loan$CompletedOrRisk, summary)
describe(sub_loan$FinanceAmount)
sort(table(sub_loan$FinanceAmount), decreasing=TRUE)[1:5]

sub_loan$PrincipalOutstanding<-as.numeric(sub_loan$PrincipalOutstanding)
by(sub_loan$PrincipalOutstanding, sub_loan$CompletedOrRisk, summary)
sort(table(sub_loan$PrincipalOutstanding), decreasing=TRUE)[1:5]


sort(table(sub_loan$BorrowerState), decreasing=TRUE)[1:5]
table(loan$BorrowerState)

sort(table(loan$AnnualRate), decreasing=TRUE)[1:5]

loan$ContractDate <- as.Date(loan$ContractDate)

loan$year <- as.numeric(format(loan$ContractDate, format = "%Y"))
sub_loan$year <- as.numeric(format(sub_loan$ContractDate, format = "%Y"))

loan$year <- factor(loan$year)
sub_loan$year <- factor(sub_loan$year)
table(loan$year)
by(sub_loan$AnnualRate, sub_loan$year, summary)
by(loan$AnnualRate, loan$year, summary)
```

Figure annual rate per year

```
ggplot(aes(x = AnnualRate), data = loan) +
  geom_bar(color="blue",binwidth = 0.005) +
  facet_wrap(~year) + ggtitle("AnnualRate by year")+
  labs(x = "AnnualRate")
```

Figure annual rate in risky or repaid loans per year

```
ggplot(aes(x = AnnualRate), data = sub_loan) +
  geom_bar(color="green",binwidth = 0.005) +
  facet_wrap(~year) + ggtitle("AnnualRate by year for CompleteOrRisk loans")+
```

```
  labs(x = "AnnualRate")
```

Figure annual rate with trendline

```
ggplot(loan, aes(x=AnnualRate)) +
  geom_histogram(stat="count", binwidth=.005, position="dodge",color="green") +
  geom_density(alpha=.3)
```

Figure annual rate of risky or repaid loans with trendline

```
ggplot(sub_loan, aes(x=AnnualRate)) +
  geom_histogram(stat="count",binwidth=.005, position="dodge",color="blue") +
  geom_density(alpha=.3)

describe(loan)
describe(sub_loan)

summary(loan$FinanceAmount)
```

Figure finance amount histogram

```
ggplot(aes(x = FinanceAmount), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("Loan Amount")+
  labs(x = "Loan amount")
```

Figure finance amount for risky or repaid loans

```
ggplot(aes(x = FinanceAmount), data = sub_loan) +
  geom_histogram(stat="count", color = "green")+ ggtitle("Loan Amount for Completed or Risky loans")+
  labs(x = "Loan amount")

summary(loan$PrincipalOutstanding)
```

Figure principal outstanding dist.

```
ggplot(aes(x = PrincipalOutstanding), data = loan) +
  geom_histogram(stat="count", color = "blue")+ ggtitle("PrincipalOutstanding")+
  labs(x = "Principal amount")
```

Figure principal outstanding of risky or repaid loans.

```
ggplot(aes(x = PrincipalOutstanding), data = sub_loan) +
  geom_histogram(stat="count", color = "green")+ ggtitle("PrincipalOutstanding for Completed or Risky
loans")+
```

```
  labs(x = "Principal amount")

sub_loan$Status <- ifelse(
  sub_loan$CompletedOrRisk == "HighRisk", 1, 0)
```

Model

```
ds <- sub_loan[, c("Status","AnnualRate","FinanceAmount", "PrincipalOutstanding", "BorrowerIncome",
"BorrowerAge")]
sapply(ds, function(x) sum(is.na(x)))
model <- glm(Status ~.,family=binomial(link='logit'),data=ds)
summary(model)
print(model)
plot(model)
coefs <- coef(model)
exp(coefs)
(exp(coefs)-1)*100
```

Correlation Matrix with risky or repaid loans

```
d <- sub_loan[, c("AnnualRate","FinanceAmount", "PrincipalOutstanding")]
d$AnnualRate <- as.numeric(d$AnnualRate)
d$FinanceAmount <- as.numeric(d$FinanceAmount)
d$PrincipalOutstanding <- as.numeric(d$PrincipalOutstanding)
d$AnnualRate <- ifelse(
  is.na(d$AnnualRate), 0, d$AnnualRate)
d$FinanceAmount <- ifelse(
  is.na(d$FinanceAmount), 0, d$FinanceAmount)
d$PrincipalOutstanding <- ifelse(
  is.na(d$PrincipalOutstanding), 0, d$PrincipalOutstanding)
d <- na.omit(d)
M_sub_loan <- cor(d)
corrplot(M_sub_loan, type="upper", order="hclust",col = brewer.pal(n=8,name = "RdYlBu"))
```

Correlation Matrix with all loans

```
d <- loan[, c("AnnualRate","FinanceAmount", "PrincipalOutstanding")]
d$AnnualRate <- as.numeric(d$AnnualRate)
d$FinanceAmount <- as.numeric(d$FinanceAmount)
d$PrincipalOutstanding <- as.numeric(d$PrincipalOutstanding)
```

```
d$AnnualRate <- ifelse(
  is.na(d$AnnualRate), 0, d$AnnualRate)
d$FinanceAmount <- ifelse(
  is.na(d$FinanceAmount), 0, d$FinanceAmount)
d$PrincipalOutstanding <- ifelse(
  is.na(d$PrincipalOutstanding), 0, d$PrincipalOutstanding)
d <- na.omit(d)
M_loan <- cor(d)
corrplot(M_loan, type="upper", order="hclust",col = brewer.pal(n=8,name = "RdYlBu"))
```

Figure amount with contract date

```
rs_date_amount <- subset(loan,select = c(2,5))
rs_date_amount <- rs_date_amount[order(rs_date_amount$ContractDate),]
rs_date_amount$ContractDate<-lubridate::ymd(rs_date_amount$ContractDate)
rs_date_amount$month <- lubridate::floor_date(rs_date_amount$ContractDate,"month")

rs_date_amount$FinanceAmount<-as.numeric(rs_date_amount$FinanceAmount)
rs_aggregate <- aggregate(FinanceAmount~ month, rs_date_amount,mean)
names(rs_aggregate) <- c("ContractDate", "FinanceAmount")
sub_loan$FinanceAmount<-as.numeric(sub_loan$FinanceAmount)
ggplot(data = rs_aggregate, aes(x=ContractDate, y = FinanceAmount)) +
  geom_line(color="blue") +
  geom_point(color="red")  +
  xlab("ContractDate") + ylab("FinanceAmount")+
  ggtitle("Mean finance per month") +
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure amount with date

```
ggplot(data = rs_aggregate, aes(x=ContractDate, y = FinanceAmount)) +
  stat_smooth(method = "lm", formula = y ~ poly(x, 9), size = 1,se=FALSE)+
  expand_limits(y=0) +
  xlab("ContractDate") + ylab("FinanceAmount")+
  ggtitle("Mean finance per month") +
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure annual rate for risky or repaid loans

```
ggplot(Plenti_Loan_Book,aes(AnnualRate)) +
  geom_histogram(stat = "count", binwidth = 0.5)+
  facet_wrap(~CompletedOrRisk, ncol = 2)
by(sub_loan$AnnualRate, sub_loan$CompletedOrRisk, summary)
sort(table(Plenti_Loan_Book$AnnualRate), decreasing=TRUE)[1:5]
```

Figure Borrower Income for all loans

```
Plenti_Loan_Book$BorrowerIncome <- factor(Plenti_Loan_Book$BorrowerIncome,
                          levels =c("<$50K", "$50-100K", "$100-150K", ">$150K"),
                          order = T)
ggplot(subset(Plenti_Loan_Book,BorrowerIncome  != ""), aes(BorrowerIncome)) +
  geom_bar(color="blue") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Borrower income (low to high risk from left to right)")
```

Figure Borrower income of risky or repaid loans

```
sub_loan$BorrowerIncome <- factor(sub_loan$BorrowerIncome,
                                        levels =c("<$50K", "$50-100K", "$100-150K", ">$150K"),
                                        order = T)
ggplot(subset(sub_loan,BorrowerIncome  != ""), aes(BorrowerIncome)) +
  geom_bar(color="green") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Borrower income of risky or repaid loans (low to high risk from left to right)")+
  facet_wrap(~CompletedOrRisk, ncol = 2)+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure HousingStatus for all loans

```
Plenti_Loan_Book$HousingStatus <-
  factor(Plenti_Loan_Book$HousingStatus,
         levels =c("Own a Home (Mortgage)", "Own a Home (No Mortgage)", "Tenant (Mortgage)", "Tenant (No
Mortgage)"),
```

```
        order = T)

ggplot(subset(Plenti_Loan_Book, HousingStatus != ""),
          aes(HousingStatus)) +
  geom_bar(color="blue") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Housing Status (low to high risk from left to right)")+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure HousingStatus for risky or repaid loans

```
sub_loan$HousingStatus <-
  factor(sub_loan$HousingStatus,
        levels =c("Own a Home (Mortgage)", "Own a Home (No Mortgage)", "Tenant (Mortgage)", "Tenant (No
Mortgage)"),
        order = T)

ggplot(subset(sub_loan, HousingStatus != ""),
          aes(HousingStatus)) +
  geom_bar(color="green") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Housing Status of borrowers of risky or repaid loans (low to high risk from left to right)")+
  facet_wrap(~CompletedOrRisk, ncol = 2)+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure Finance purpose for all loans

```
Plenti_Loan_Book$FinancePurpose <-
  factor(Plenti_Loan_Book$FinancePurpose,
        levels =c("Business", "Car/Vehicle", "Debt Consolidation", "Energy Efficient Lighting","Home
Improvement",
                  "Investment ", "Major Event", "Major Purchase","Other Consumer Loans", "Professional
Services",
                  "Renewable Energy", "Solar Battery","Solar Energy Battery", "Solar Energy Equipment ",
                  "Solar Energy Panels", "Solar Water Heater", "Variable Speed Drive"),
        order = T)
ggplot(subset(Plenti_Loan_Book,FinancePurpose  != ""), aes(FinancePurpose)) +
  geom_bar(color="blue",fill="light green") +
```

```
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("FinancePurpose")+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure Finance purpose for risky or repaid loans

```
sub_loan$FinancePurpose <-
  factor(sub_loan$FinancePurpose,
         levels =c("Business", "Car/Vehicle", "Debt Consolidation", "Energy Efficient Lighting","Home
Improvement",
                   "Investment ", "Major Event", "Major Purchase","Other Consumer Loans", "Professional
Services",
                   "Renewable Energy", "Solar Battery","Solar Energy Battery", "Solar Energy Equipment ",
                   "Solar Energy Panels", "Solar Water Heater", "Variable Speed Drive"),
         order = T)
ggplot(subset(sub_loan,FinancePurpose != "" ),
       aes(FinancePurpose,..count../sum(..count..))) +
  geom_bar(color="green") +
  geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
            vjust=-.2, size = 3) +
  ggtitle("Finance purpose of borrowers with risky or repaid loans") +
  facet_wrap(~CompletedOrRisk, ncol = 2)+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure Borrowers State All loans

```
Plenti_Loan_Book$BorrowerState <-
  factor(Plenti_Loan_Book$BorrowerState,
         levels =c("ACT", "NSW", "NT", "QLD","SA",
                   "TAS", "VIC", "WA"),
         order = T)
ggplot(subset(Plenti_Loan_Book,BorrowerState  != ""), aes(BorrowerState)) +
  geom_bar(color="green",fill="light blue") +
```

```
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Borrower State")+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure Borrower State (Risky or Repaid)

```
sub_loan$BorrowerState <-
  factor(sub_loan$BorrowerState,
         levels =c("ACT", "NSW", "NT", "QLD","SA",
                   "TAS", "VIC", "WA"),
         order = T)
ggplot(subset(sub_loan,BorrowerState != "" ),
       aes(BorrowerState,..count../sum(..count..))) +
  geom_bar(color="green") +
  geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
            vjust=-.2, size = 3) +
  ggtitle("Borrower state for risky or repaid loans") +
  facet_wrap(~CompletedOrRisk, ncol = 2)
```

Figure Borrower Age (All loans)

```
Plenti_Loan_Book$BorrowerAge <- factor(Plenti_Loan_Book$BorrowerAge,
                                       levels =c("<20", "20-29", "30-39", "40-49",
                                                 "50-59", "60-69", "70-79", "80-89"),
                                       order = T)

ggplot(subset(Plenti_Loan_Book,BorrowerAge  != ""), aes(BorrowerAge)) +
  geom_bar(color="blue",fill="light green") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Borrower Age")
```

Figure Borrower Age (Risky or Repaid)

```
sub_loan$BorrowerAge <- factor(sub_loan$BorrowerAge,
                               levels =c("<20", "20-29", "30-39", "40-49",
                                         "50-59", "60-69", "70-79", "80-89"),
                               order = T)
```

```
ggplot(subset(sub_loan,BorrowerAge != "" ),
       aes(HousingStatus,..count../sum(..count..))) +
  geom_bar(color="green") +
  geom_text(stat='count',
            aes(label= paste(round(..count../sum(..count..)*100, 2),"%")),
            vjust=-.2, size = 3) +
  ggtitle("Borrower's Age") +
  facet_wrap(~CompletedOrRisk, ncol = 2)+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

Figure EmploymentStatus

```
table(loan$EmploymentStatus)
table(sub_loan$EmploymentStatus)

Plenti_Loan_Book$EmploymentStatus <-
  factor(Plenti_Loan_Book$EmploymentStatus,
         levels =c("Contract", "Employed full time", "Employed part time", "Houseperson","Retierd/other",
                   "Self-employed"),
         order = T)
ggplot(subset(Plenti_Loan_Book,EmploymentStatus  != ""), aes(EmploymentStatus)) +
  geom_bar(color="green",fill="light blue") +
  geom_text(stat='count', aes(label=..count..), vjust=1.2, size = 4) +
  ggtitle("Employment status of the borrowers")+
  theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))

sub_loan$EmploymentStatus <-
  factor(sub_loan$EmploymentStatus,
         levels =c("Contract", "Employed full time", "Employed part time", "Houseperson","Retierd/other",
                   "Self-employed"),
         order = T)
ggplot(subset(sub_loan,EmploymentStatus  != ""), aes(EmploymentStatus)) +
  geom_bar(stat='count',color="green",fill="light blue") +
```

```
facet_wrap(~CompletedOrRisk, ncol = 2)
ggtitle("Employment status of the borrowers for risky or repaid loans")+
theme(plot.title = element_text(hjust = 0.5),axis.text.x = element_text(angle = 290,hjust =0))
```

RateSetter loan book. As RateSetter has decided on some changes, its loan book is not available online for download. I had done my analysis on the loan book that used to be available at the point of the study. I have attached an embedded excel page to give a glimpse of the loan book which is available with me. I can provide the file if interested.

ContractID	ContractDate	LoanTerm	AnnualRate	FinanceAmount	PrincipalOutstanding	FinancePurpose	BorrowerState	BorrowerAge	EmploymentStatus	BorrowerIncome	HousingStatus	EarlyPaymentsMade	RepaymentStatus
E7D81DD7-55334	30.10.14	24	7,47%	$30.470	$0,00	Car/Vehicle	NSW	60-69	Employed full time	>$150K	Tenant (No Mortgage)	Yes	Repaid
97DAED44-89634	31.10.14	60	10,13%	$2.489	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
FFB78354-199537	11.11.14	36	8,30%	$15.631	$0,00	Investment	NSW	40-49	Employed full time	>$150K	Tenant (No Mortgage)	No	Repaid
3F4FFC33-940203	14.11.14	9	6,49%	$10.430	$0,00	Major Event	WA	30-39	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
FE1CE069-19746	18.11.14	60	10,09%	$18.473	$0,00	Car/Vehicle	QLD	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
89F79EDB-44008	19.11.14	60	10,67%	$32.175	$0,00	Home Improvemen	VIC	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
19A17B8D-90424	19.11.14	12	5,99%	$13.243	$0,00	Home Improvemen	TAS	50-59	Employed full time	$50-100K	Own a Home (No Mortga	Yes	Repaid
B13507A3-93660	26.11.14	24	6,21%	$12.708	$0,00	Investment	SA	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
2FD8F1B5-27744	27.11.14	24	6,50%	$3.388	$0,00	Other Consumer Lo	NSW	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
B5D9D696-14269	28.11.14	36	8,29%	$20.884	$0,00	Debt Consolidation	NSW	40-49	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
904D1628-74807	28.11.14	36	8,44%	$15.998	$0,00	Car/Vehicle	WA	60-69	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
08793A1D-76150	01.12.14	24	6,66%	$35.410	$0,00	Business	VIC	50-59	Self-employed	$50-100K	Own a Home (Mortgage)	No	Repaid
0AF111F7-214644	02.12.14	36	8,48%	$3.978	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
16251565-665918	02.12.14	60	10,64%	$27.550	$0,00	Car/Vehicle	VIC	40-49	Employed full time	>$150K	Tenant (No Mortgage)	Yes	Repaid
8F93B190-916791	03.12.14	60	11,21%	$30.840	$0,00	Debt Consolidation	NSW	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
581ADDE8-58695	03.12.14	60	10,97%	$28.300	$0,00	Home Improvemen	VIC	30-39	Employed full time	$100-150K	Tenant (Mortgage)	Yes	Repaid
53F7A212-62297	03.12.14	36	8,52%	$13.460	$0,00	Other Consumer Lo	NSW	50-59	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
A8F83BEC-89404	08.12.14	12	6,01%	$18.506	$0,00	Car/Vehicle	WA	40-49	Contract	$100-150K	Own a Home (Mortgage)	No	Repaid
FB7931DB-50512	17.12.14	36	7,85%	$5.599	$0,00	Business	VIC	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
BA3074ED-04282	18.12.14	24	7,11%	$10.680	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
D1B409E2-24271	19.12.14	60	10,05%	$10.630	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
A00B9F1E-87998	19.12.14	36	8,02%	$21.606	$0,00	Debt Consolidation	VIC	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
70FE1819-910741	23.12.14	36	8,16%	$13.044	$0,00	Debt Consolidation	QLD	20-29	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
D232FD3E-16651	24.12.14	24	6,83%	$26.464	$0,00	Business	VIC	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
B8ADC0FF-23736	30.12.14	60	10,21%	$5.765	$0,00	Other Consumer Lo	VIC	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
82B810A9-12555	30.12.14	60	10,05%	$15.885	$0,00	Other Consumer Lo	QLD	50-59	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
7199139B-11096	31.12.14	36	8,75%	$10.789	$0,00	Other Consumer Lo	WA	30-39	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
7097F978-402137	01.01.15	60	10,35%	$31.470	$0,00	Major Event	NT	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
3ED34046-84241	06.01.15	24	6,51%	$5.478	$0,00	Other Consumer Lo	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
34DCB326-70087	06.01.15	24	6,71%	$37.143	$0,00	Home Improvemen	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
9B6E9129-55778	06.01.15	60	10,57%	$32.010	$0,00	Other Consumer Lo	VIC	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
2589FBCD-29349	07.01.15	36	9,45%	$30.656	$0,00	Debt Consolidation	VIC	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
FA289FEC-91319	08.01.15	48	10,92%	$32.439	$0,00	Home Improvemen	QLD	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
D6CC8454-22879	08.01.15	60	11,29%	$27.925	$0,00	Debt Consolidation	QLD	50-59	Self-employed	$50-100K	Tenant (No Mortgage)	Yes	Repaid
D12CF768-96772	09.01.15	36	9,19%	$15.571	$0,00	Car/Vehicle	WA	60-69	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
024BA8F8-61401	10.01.15	36	9,21%	$10.314	$0,00	Car/Vehicle	VIC	30-39	Employed full time	$50-100K	Tenant (Mortgage)	Yes	Repaid
92E529CF-63393	13.01.15	24	6,18%	$6.461	$0,00	Home Improvemen	NSW	50-59	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
146AF3CD-17087	13.01.15	36	9,21%	$15.314	$0,00	Other Consumer Lo	SA	60-69	Employed full time	<$50K	Own a Home (No Mortga	No	Repaid
59D5E4FB-69439	13.01.15	36	8,48%	$10.903	$0,00	Debt Consolidation	WA	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
61262BCE-09985	14.01.15	24	6,19%	$15.605	$0,00	Debt Consolidation	WA	70-79	Employed full time	$50-100K	Tenant (Mortgage)	Yes	Repaid
E93615A3-73724	14.01.15	60	10,81%	$26.824	$0,00	Debt Consolidation	NSW	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
1D065602-72410	20.01.15	24	5,75%	$5.350	$0,00	Car/Vehicle	QLD	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
5B96BB0C-92716	20.01.15	48	11,16%	$25.760	$0,00	Major Event	QLD	60-69	Employed full time	$100-150K	Tenant (Mortgage)	Yes	Repaid
5FFF9123-208544	21.01.15	48	11,36%	$10.775	$0,00	Major Event	NSW	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
DFF87D11-66204	22.01.15	60	11,62%	$24.389	$0,00	Debt Consolidation	NSW	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
A661DF2E-10033	27.01.15	24	5,46%	$14.642	$0,00	Debt Consolidation	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
382E85A1-18847	27.01.15	24	5,50%	$2.997	$0,00	Car/Vehicle	VIC	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
3ACF1101-99862	27.01.15	36	8,86%	$13.294	$0,00	Debt Consolidation	NSW	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
55051B61-62949	28.01.15	36	8,31%	$10.903	$0,00	Major Event	QLD	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
6E47EA94-79515	30.01.15	24	5,41%	$5.268	$0,00	Debt Consolidation	NSW	30-39	Employed full time	$100-150K	Tenant (Mortgage)	No	Repaid
9B8B169E-25707	30.01.15	36	8,77%	$10.157	$0,00	Car/Vehicle	NSW	20-29	Employed full time	<$50K	Tenant (No Mortgage)	Yes	Repaid
373B9A32-11329	30.01.15	60	11,14%	$31.740	$0,00	Debt Consolidation	VIC	60-69	Self-employed	$100-150K	Tenant (No Mortgage)	No	Repaid
43815E63-460721	31.01.15	48	11,15%	$5.511	$0,00	Car/Vehicle	QLD	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
F960ECF5-78201	02.02.15	60	11,22%	$17.759	$0,00	Home Improvemen	ACT	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
228A5ABF-30358	03.02.15	36	8,74%	$25.810	$0,00	Business	SA	70-79	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
8B8AF2CF-94087	03.02.15	36	8,79%	$26.005	$0,00	Debt Consolidation	ACT	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
FB5E4623-36500	03.02.15	60	11,25%	$11.497	$0,00	Car/Vehicle	QLD	30-39	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
42C5CC54-65797	04.02.15	60	10,49%	$16.335	$0,00	Investment	NSW	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
[illegible]													
3438EC17-13838	21.04.15	24	5,34%	$15.358	$0,00	Home Improvemen	WA	60-69	Employed full time	$100-150K	Own a Home (No Mortga	Yes	Repaid
EA8A9909-06882	21.04.15	36	7,73%	$12.770	$0,00	Car/Vehicle	VIC	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
333A56C1-14650	21.04.15	60	10,72%	$15.795	$0,00	Major Event	NSW	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
FEB275D2-20101	22.04.15	36	7,72%	$7.426	$0,00	Car/Vehicle	VIC	60-69	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
0FCE9C76-41886	22.04.15	48	10,78%	$25.530	$0,00	Home Improvemen	QLD	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
DA36B527-65747	22.04.15	6	4,90%	$3.232	$0,00	Major Event	NSW	50-59	Employed full time	$50-100K	Tenant (Mortgage)	No	Repaid
5D9ABCF9-60868	22.04.15	60	10,64%	$26.263	$0,00	Other Consumer Lo	NSW	50-59	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
A0B5EBCC-42679	23.04.15	48	10,74%	$13.629	$0,00	Debt Consolidation	WA	50-59	Employed full time	>$150K	Tenant (Mortgage)	No	Repaid
305CE00D-29294	23.04.15	36	7,72%	$11.555	$0,00	Debt Consolidation	SA	30-39	Employed full time	<$50K	Tenant (No Mortgage)	No	Repaid
0166DD4E-83663	23.04.15	36	7,68%	$20.694	$0,00	Home Improvemen	NSW	40-49	Self-employed	$100-150K	Own a Home (Mortgage)	No	Repaid
33A05E88-73244	24.04.15	36	7,77%	$3.607	$0,00	Car/Vehicle	VIC	50-59	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
858BF96A-40771	24.04.15	60	10,66%	$35.510	$0,00	Debt Consolidation	SA	60-69	Employed full time	$100-150K	Own a Home (No Mortga	Yes	Repaid
6BFD2578-15965	24.04.15	36	7,75%	$19.055	$0,00	Major Event	VIC	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
CAC1DD77-82238	24.04.15	60	10,78%	$5.675	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
976E5FA4-14347	28.04.15	36	7,54%	$7.942	$0,00	Car/Vehicle	QLD	60-69	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
FCF39A58-47111	28.04.15	60	10,74%	$25.600	$0,00	Other Consumer Lo	WA	50-59	Employed full time	$100-150K	Tenant (No Mortgage)	No	Repaid
509A563D-42569	28.04.15	48	10,84%	$10.461	$0,00	Major Event	NSW	60-69	Self-employed	<$50K	Tenant (No Mortgage)	Yes	Repaid
5D8D327E-97296	29.04.15	24	4,22%	$10.575	$0,00	Car/Vehicle	NSW	40-49	Employed full time	>$150K	Tenant (No Mortgage)	No	Repaid
1A5B895D-37851	29.04.15	60	10,78%	$31.110	$0,00	Other Consumer Lo	NSW	40-49	Self-employed	$50-100K	Tenant (Mortgage)	Yes	Repaid
285696AD-10751	01.05.15	24	4,24%	$20.950	$0,00	Major Event	QLD	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
9EBEB12B-57111	04.05.15	60	10,64%	$21.020	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
1FA3F9D8-50368	04.05.15	24	4,12%	$5.388	$0,00	Major Event	QLD	20-29	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
8106AF9F-54164	04.05.15	60	10,63%	$27.175	$0,00	Other Consumer Lo	WA	30-39	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
325C3F86-21241	05.05.15	36	7,13%	$11.388	$0,00	Home Improvemen	QLD	30-39	Employed full time	<$50K	Own a Home (Mortgage)	No	Repaid
B799D2D2-78904	05.05.15	60	10,60%	$12.768	$0,00	Major Event	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
5FC712FA-88568	05.05.15	12	3,83%	$3.275	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
74F59D63-39039	06.05.15	60	10,61%	$27.626	$0,00	Major Event	QLD	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
18FD8779-25987	06.05.15	24	4,09%	$35.909	$0,00	Other Consumer Lo	VIC	30-39	Employed full time	>$150K	Tenant (Mortgage)	No	Repaid
2FB80605-57450	06.05.15	36	7,07%	$5.400	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
0B1BAB49-00573	07.05.15	24	4,16%	$4.266	$0,00	Car/Vehicle	WA	40-49	Employed full time	<$50K	Own a Home (Mortgage)	No	Repaid
142158AD-08451	07.05.15	36	7,12%	$12.360	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
C19FFAD4-80237	07.05.15	60	10,62%	$12.076	$0,00	Other Consumer Lo	VIC	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
645B5223-38453	07.05.15	60	10,63%	$25.675	$0,00	Car/Vehicle	WA	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
3D233D0D-19638	08.05.15	18	4,07%	$14.553	$0,00	Other Consumer Lo	SA	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
C7072352-86881	08.05.15	60	10,64%	$36.770	$0,00	Business	WA	40-49	Self-employed	$50-100K	Own a Home (No Mortga	No	Repaid
503636ED-85067	11.05.15	36	7,05%	$15.913	$0,00	Car/Vehicle	ACT	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
C4C54050-47215	11.05.15	60	10,48%	$19.813	$0,00	Debt Consolidation	QLD	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
1CC65180-43890	11.05.15	60	10,64%	$10.420	$0,00	Major Purchase	TAS	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
70C6A22C-08260	12.05.15	60	10,60%	$15.544	$0,00	Major Event	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
5D103EC9-41077	12.05.15	36	7,13%	$4.360	$0,00	Home Improvemen	VIC	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
F3AFFB5D-41030	13.05.15	24	4,09%	$17.583	$0,00	Debt Consolidation	QLD	50-59	Employed full time	$100-150K	Own a Home (Mortgage)	No	Repaid
1D96CD45-70158	13.05.15	24	4,14%	$6.425	$0,00	Car/Vehicle	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
865765B9-86947	14.05.15	36	7,13%	$8.398	$0,00	Car/Vehicle	QLD	50-59	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
46471797-40911	14.05.15	24	4,12%	$10.365	$0,00	Car/Vehicle	ACT	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
13AF15A8-75539	14.05.15	36	7,20%	$12.071	$0,00	Debt Consolidation	WA	50-59	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
C13CBA11-84509	14.05.15	48	10,64%	$24.623	$0,00	Debt Consolidation	WA	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
BF69EE84-639017	15.05.15	18	4,58%	$5.839	$0,00	Investment	NSW	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
CE007CD4-42987	15.05.15	60	10,54%	$20.720	$0,00	Professional Service	WA	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
AD5AF08C-25566	15.05.15	36	7,23%	$13.818	$0,00	Car/Vehicle	WA	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
F34069F2-680343	15.05.15	36	7,25%	$10.523	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
D630B253-67926	15.05.15	48	10,33%	$8.267	$0,00	Car/Vehicle	QLD	50-59	Employed part time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
44E413FA-29409	18.05.15	36	7,43%	$10.333	$0,00	Car/Vehicle	NSW	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
46C7E704-48872	18.05.15	36	8,02%	$18.008	$0,00	Debt Consolidation	QLD	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
0241183B-47190	19.05.15	36	8,00%	$15.799	$0,00	Debt Consolidation	QLD	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
40FF78C0-82041	19.05.15	36	8,17%	$8.519	$0,00	Car/Vehicle	QLD	40-49	Employed full time	<$50K	Own a Home (Mortgage)	Yes	Repaid
461451BC-41183	19.05.15	60	10,11%	$7.853	$0,00	Professional Service	WA	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
70BD6C13-08590	19.05.15	36	8,70%	$5.014	$0,00	Car/Vehicle	NSW	30-39	Employed part time	<$50K	Tenant (No Mortgage)	No	Repaid
4568D9B0-99527	20.05.15	60	10,17%	$10.810	$0,00	Debt Consolidation	NSW	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
609B4873-78855	20.05.15	48	10,38%	$11.578	$0,00	Car/Vehicle	WA	30-39	Employed full time	$100-150K	Tenant (Mortgage)	Yes	Repaid
4236250E-97223	20.05.15	24	4,25%	$5.388	$0,00	Professional Service	VIC	50-59	Employed full time	<$50K	Tenant (No Mortgage)	Yes	Repaid
DB44BBD6-74943	20.05.15	24	4,25%	$20.470	$0,00	Home Improvemen	WA	30-39	Self-employed	$50-100K	Own a Home (Mortgage)	No	Repaid
486B6F1E-11832	21.05.15	36	7,72%	$15.685	$0,00	Home Improvemen	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
D32893EF-63838	21.05.15	60	10,45%	$29.812	$0,00	Debt Consolidation	VIC	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
E5BFC979-63851	22.05.15	48	10,33%	$30.783	$0,00	Debt Consolidation	NSW	50-59	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
0DABE8C8-37813	22.05.15	36	7,74%	$20.390	$0,00	Car/Vehicle	SA	60-69	Employed full time	$50-100K	Tenant (Mortgage)	Yes	Repaid
BCEE2962-99108	22.05.15	24	3,94%	$14.518	$0,00	Debt Consolidation	NSW	40-49	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
FF4C2DBB-95119	25.05.15	18	3,87%	$6.380	$0,00	Investment	VIC	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
7C309FD6-96629	26.05.15	24	3,83%	$8.284	$0,00	Home Improvemen	VIC	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
FC9C3215-62966	26.05.15	36	7,67%	$20.694	$0,00	Home Improvemen	NSW	50-59	Employed full time	$50-100K	Tenant (Mortgage)	No	Repaid
C7FF70A7-93538	26.05.15	36	7,65%	$10.151	$0,00	Other Consumer Lo	WA	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
829393DD-07679	26.05.15	36	7,74%	$12.246	$0,00	Debt Consolidation	NSW	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
5F6E69FA-63537	26.05.15	60	10,23%	$8.900	$0,00	Car/Vehicle	QLD	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
4D60071D-63375	26.05.15	60	10,30%	$26.275	$0,00	Debt Consolidation	NSW	30-39	Employed full time	$50-100K	Tenant (No Mortgage)	Yes	Repaid
FFFC0FF8-399776	27.05.15	36	7,65%	$10.409	$0,00	Investment	WA	50-59	Employed full time	$100-150K	Own a Home (Mortgage)	No	Repaid
59B1E8BB-13072	27.05.15	60	10,22%	$7.890	$0,00	Car/Vehicle	NSW	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
E5F5018B-60236	27.05.15	60	10,34%	$33.356	$0,00	Car/Vehicle	VIC	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
BAFA7ACE-51221	27.05.15	12	3,83%	$3.290	$0,00	Home Improvemen	NSW	30-39	Employed full time	$100-150K	Tenant (No Mortgage)	No	Repaid
05279631-78564	27.05.15	24	3,83%	$12.650	$0,00	Debt Consolidation	SA	30-39	Employed full time	<$50K	Tenant (No Mortgage)	No	Repaid
1E14FD3A-28421	27.05.15	48	10,29%	$21.928	$0,00	Debt Consolidation	WA	30-39	Employed full time	>$150K	Tenant (No Mortgage)	Yes	Repaid
9E10FC82-26605	28.05.15	36	7,54%	$20.998	$0,00	Home Improvemen	WA	40-49	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
F626677C-35433	28.05.15	60	10,40%	$13.729	$0,00	Debt Consolidation	WA	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
8B466BB6-32176	29.05.15	24	3,97%	$11.075	$0,00	Car/Vehicle	WA	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
362BAF32-38883	29.05.15	24	4,12%	$30.863	$0,00	Business	VIC	40-49	Employed full time	>$150K	Own a Home (Mortgage)	Yes	Repaid
6C31DDB6-84939	29.05.15	36	7,57%	$7.999	$0,00	Debt Consolidation	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
9095F177-56430	29.05.15	36	7,65%	$7.999	$0,00	Car/Vehicle	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
8CDF4E5F-65689	01.06.15	60	10,40%	$37.855	$0,00	Professional Service	WA	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	No	Repaid
38C3BF34-70827	02.06.15	36	7,54%	$11.789	$0,00	Debt Consolidation	QLD	60-69	Employed full time	$100-150K	Own a Home (Mortgage)	No	Repaid
1B7D1A4B-91549	02.06.15	60	10,42%	$10.660	$0,00	Professional Service	NSW	40-49	Employed full time	$100-150K	Own a Home (Mortgage)	Yes	Repaid
EEC9AC27-11977	02.06.15	60	10,43%	$21.080	$0,00	Home Improvemen	VIC	40-49	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
568E03D9-37798	02.06.15	36	7,60%	$13.703	$0,00	Car/Vehicle	VIC	30-39	Employed full time	$100-150K	Tenant (No Mortgage)	Yes	Repaid
8A66CECD-72081	03.06.15	36	7,78%	$21.150	$0,00	Debt Consolidation	VIC	40-49	Contract	$50-100K	Tenant (No Mortgage)	Yes	Repaid
BF38274A-38966	03.06.15	60	10,30%	$11.115	$0,00	Major Purchase	VIC	50-59	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
4D65FF23-95280	03.06.15	12	3,37%	$17.465	$0,00	Major Event	WA	40-49	Employed full time	>$150K	Tenant (Mortgage)	Yes	Repaid
52FA4031-51596	03.06.15	36	7,58%	$14.865	$0,00	Car/Vehicle	QLD	20-29	Employed full time	<$50K	Tenant (No Mortgage)	Yes	Repaid
7E4F1536-27719	03.06.15	36	7,65%	$8.580	$0,00	Debt Consolidation	QLD	40-49	Employed full time	$50-100K	Tenant (No Mortgage)	No	Repaid
ADCEE4D9-80102	05.06.15	60	10,11%	$4.552	$0,00	Home Improvemen	QLD	30-39	Employed full time	$50-100K	Own a Home (Mortgage)	Yes	Repaid
15B377FF-45435	05.06.15	12	3,64%	$36.166	$0,00	Debt Consolidation	VIC	50-59	Self-employed	>$150K	Own a Home (Mortgage)	No	Repaid

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism.

Herzenstein, M. U. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing*.

Galloway, I. (2009). *Peer-to-peer lending and community development finance.*

Morse, A. (2015). Peer-to-peer crowdfunding: information and the potential for disruption in consumer lending. *NBER Working Papers*.

Ravina, E. (2012). Love & loans: the effect of beauty and personal characteristics in credit markets.

Pope, D. G. (2011). What's in a picture? Evidence of discrimination from prosper.com. *Jounal of Human Resources*.

Company Data, M. S. (2020). Global Marketplace Loan Issurance. Morgan Stanley Research.

Club, L. (n.d.). Retrieved from https://www.lendingclub.com/info/statistics.action

Milne A, P. P. (2016). The Business Models and Economics of Peer-to-Peer Lending. *Euro Credit Research Institute*.

Duarte, J. S. (2012). Trust and credit: the role of appearance in peer-to-peer lending. . *Review of Financial Studies*.

Lin, M. N. (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. . *Management Sciennces*.

Hildebrand, T. M. (2015). Adverse incentives in crowdfunding. *Management Science Forthcoming*.

Gregor, N. F. (2010). Mitigating Adverse Selection In P2p Lending Empirical Evidence From Prosper.Com .

Bohme, R. a. (2010). Privacy in online social lending. *Spring Symposium on Intelligent Privacy Management*.

Herrero-Lopez, S. (2009). Social Interactions in P2P Lending.

Hildebrand, T. P. (2010). Skin in the Game : Evidence from the Online Social Lending Market .

Greiner, M. E. (2009). The Role of Social Capital in People-to-People Lending Marketplaces . *Association for Information Systems*.

Jefferson Duarte, S. S. (2012). Trust and Credit: The Role of Appearance in Peer-to-peer Lending . *Society of Financial Studies*.

Stiglitz, J. E. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*.

Leland, H. a. (1977). Informational asymmetries, financial structure, and financial intermediation . *Journal of Finance*.

Riza Emekter, Y. T. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending . *Applied Economics*.

Kumar, S. (2007). Empirical Analysis of Peer-to-Peer Financial Marketplace. *Bank of One.* Colorado: American Conference on Information Systems.

Everett, C. R. (2010). Group membership, relationship banking and loan default risk: the case of online social lending. West Lafayette, IN: Group.

Collier, B. a. (2010). Sending Mixed Signals: Multilevel Reputation Effects in Peer-to-Peer Lending Markets. . ACM Conference on Computer Supported Cooperative Work.

Rapp, E. C. (2012). Regulating Online Peer-to-Peer Lending in the Aftermath of Dodd-Frank: In Search of an Evolving Regulatory Regime for an Evolving Industry .

Jin, F. S. (2014). THE INFORMATION VALUE OF ONLINE SOCIAL NETWORKS: ESSONS FROM PEER-TO-PEER LENDING . *Working paper*.

Dataset for RateSetter collected from *http://www.plenti.com.au/loan-book-june-2020/*