***Lending insights from LendingClub loan data***

LendingClub was one of the initial pioneers of the peer-to-peer online lending platform. It championed the way in which an everyday Joe the lender could invest/lend funds directly to everyday Joe the borrower without any hassles in return for attractive investment returns.

Like any banking platform, online peer-to-peer lending model comes withs own risks and rewards. Even though LendingClub is able to remove the banker from its banking system there are some caveats that an investor must be aware before one begins investing in this platform. LendingClub conveniently allows the investor to choose from a variety of loan types where risk commensurately matches the reward, i.e., higher the risk higher the reward and vice versa.

The main risks that a lender/investor faces in such a platform are: loan default risk and loan prepayment risk. Loan default occurs when a borrower abruptly stops making payments to the lender, in such a scenario the investor loses out on both the principal and interest payments from the borrower. The is the most significant risk that any lender faces and LendingClub tries to mitigate this risk by only allowing borrowers with very good credit history to apply for loans.

The second most significant risk is the loan prepayment risk where the borrower prepays the principal in order to reduce his interest burden. In such a scenario even though the lender recovers the entire principal amount, he loses out on the interest payments that he would have received had the borrower made all payments as per the payment schedule. In order to mitigate this scenario a substantial portion of the initial payments goes towards interest recovery.

In this study, we use the loan dataset made available by LendingClub to gain data insights about which features play a pivotal role in predicting the final status of a loan. The dataset used for this study contains roughly 2.2M entries of approved loans with 151 features to describe each loan entry.

In the subsequent data analysis section, we try to study the effects of some key features on the final outcome of a loan status. Following is the distribution of loans in our study of loan outcome:

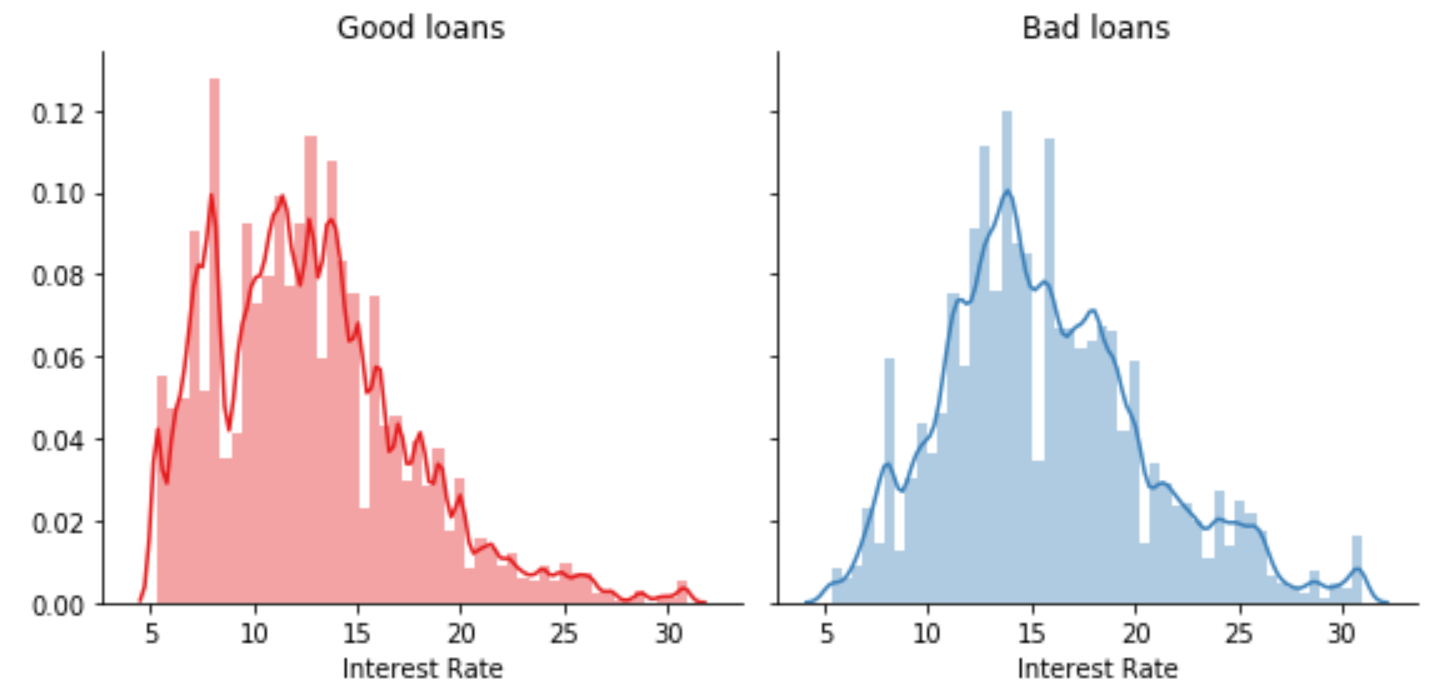
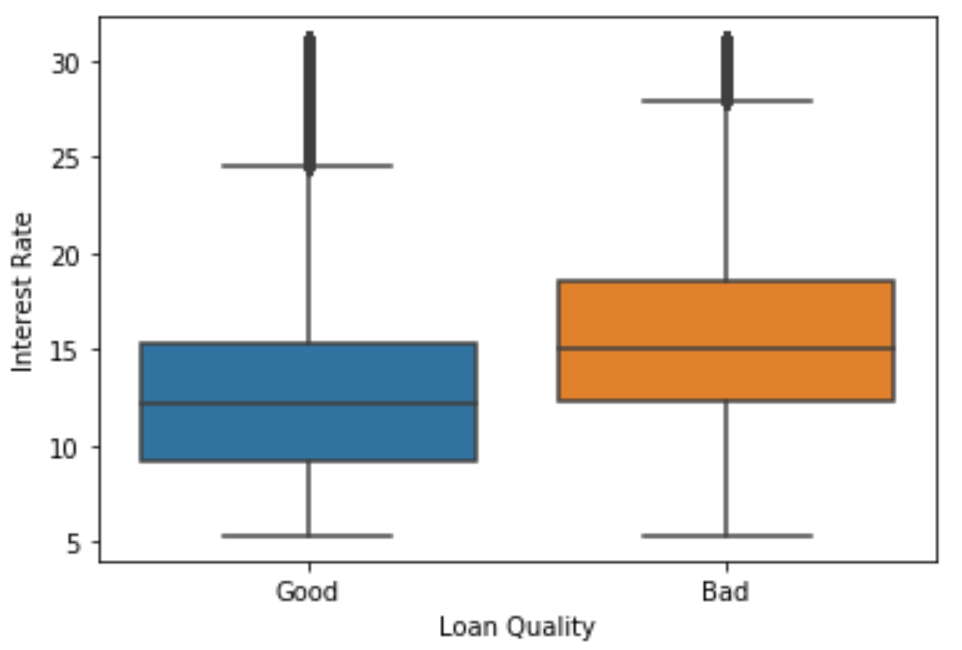
***Loan Default: Loan Prepayment:***

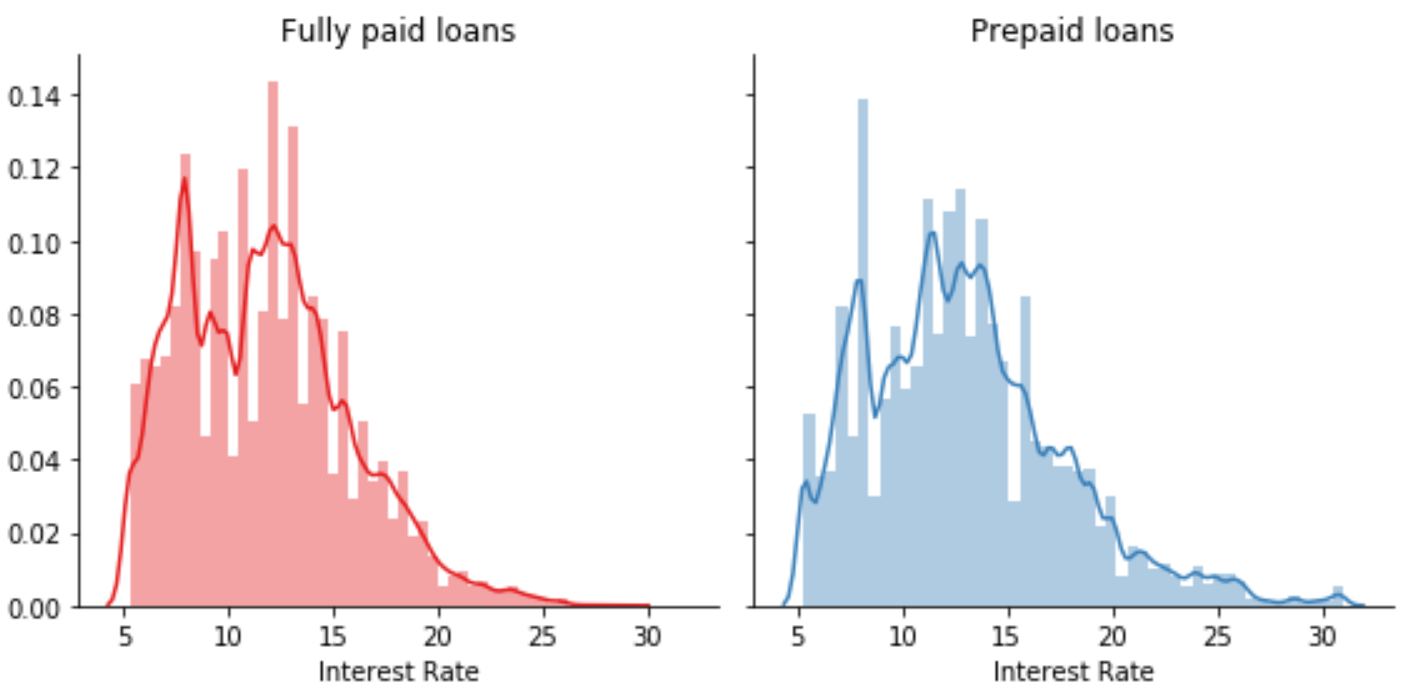
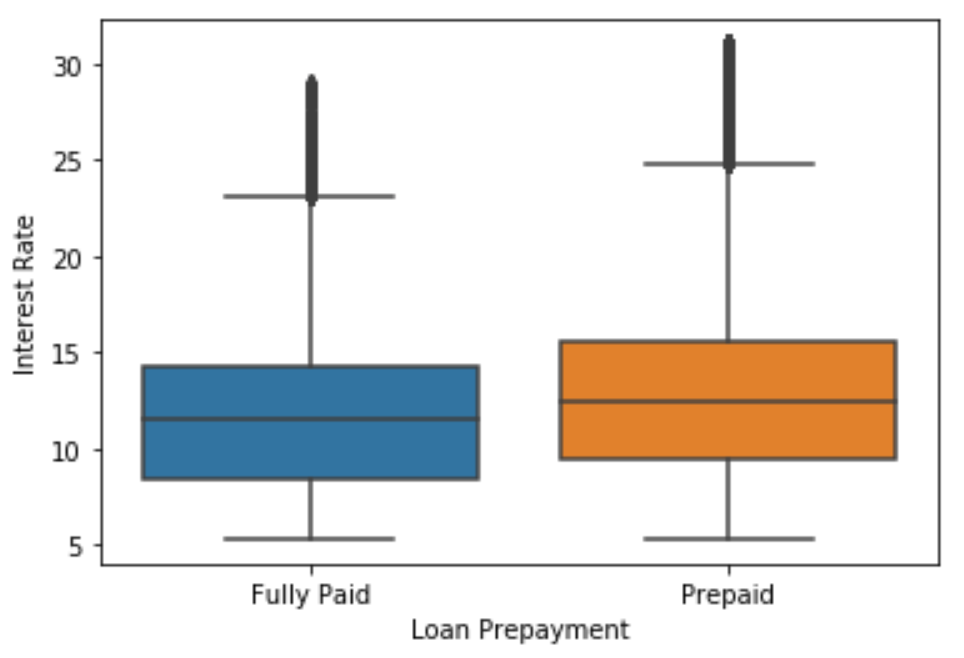
Good loans: ~1966000 (87%) Fully-paid loans: ~273000 (25%)

Bad loans: ~295000 (13%) Pre-paid loans: ~806000 (75%)

***Exploratory Data Analysis:***

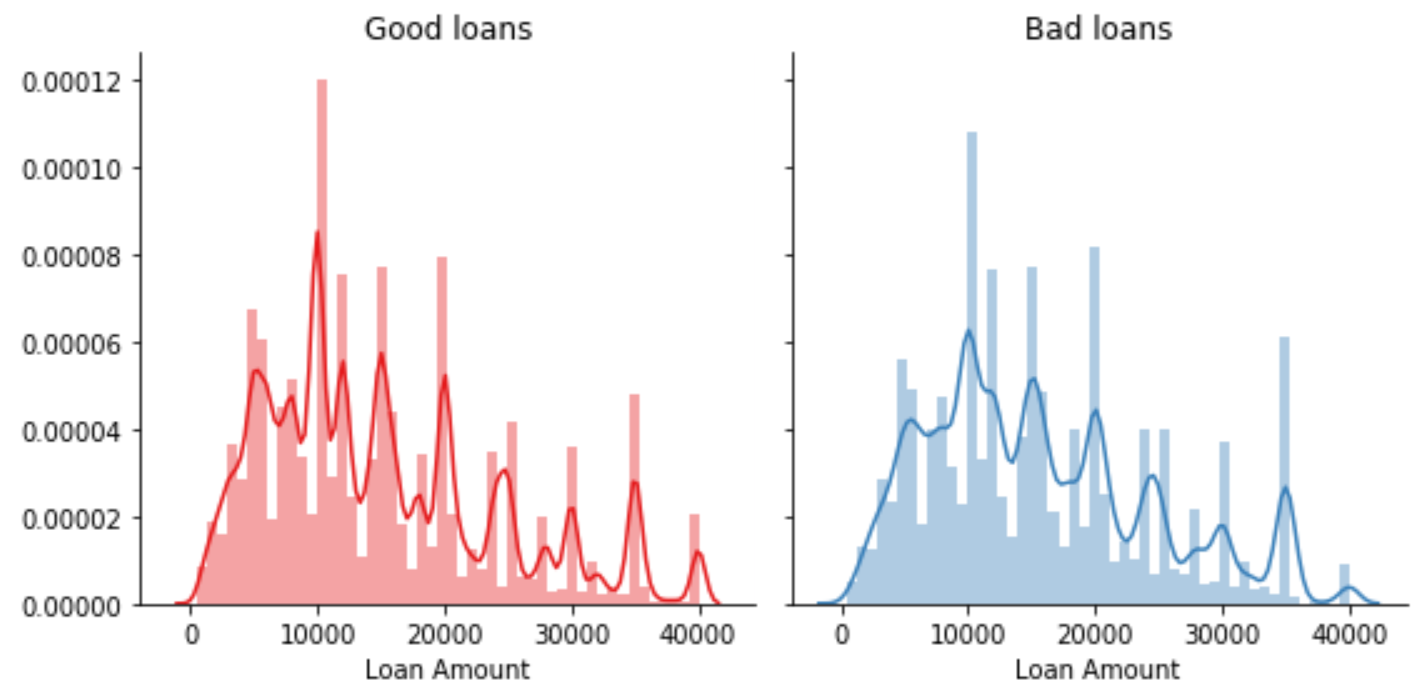
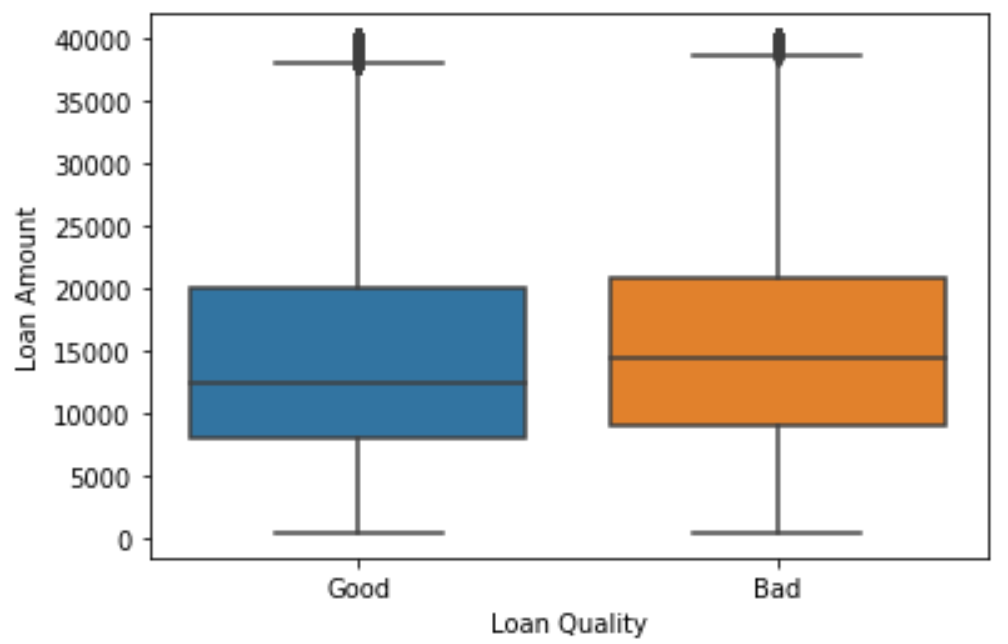
**Interest rates:**

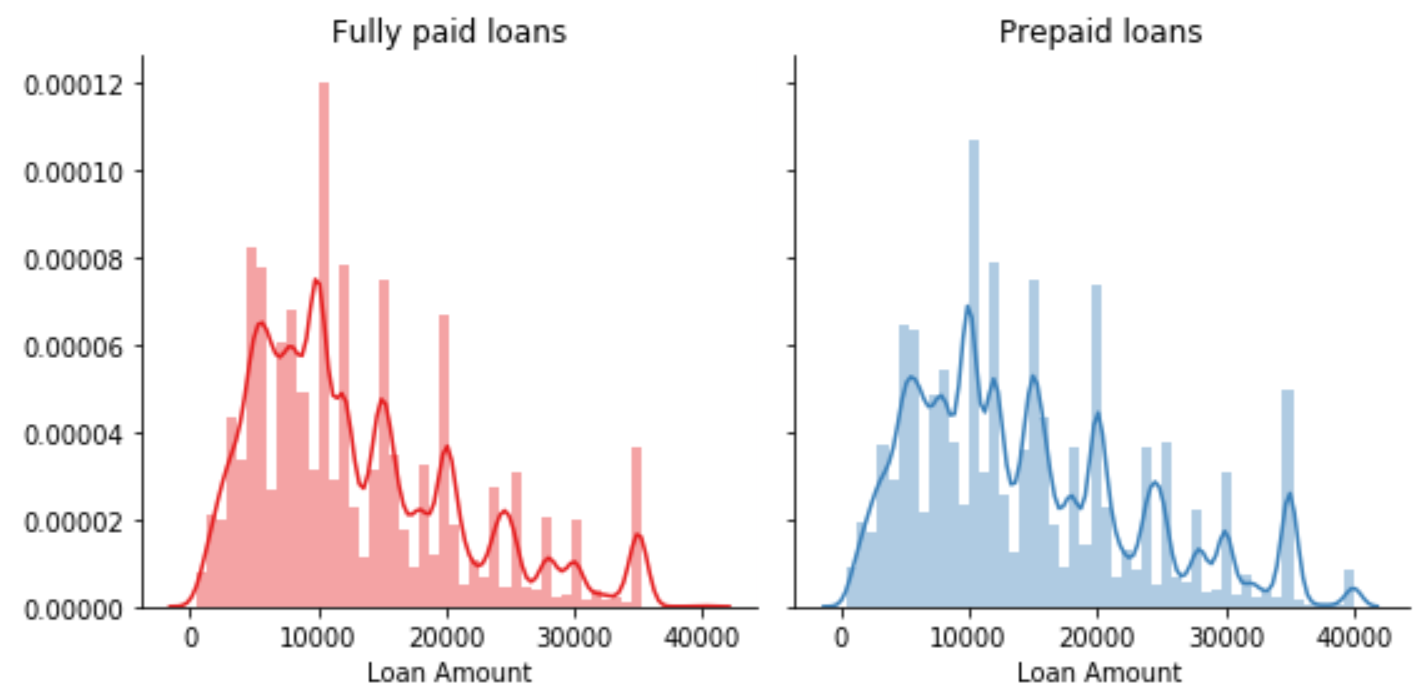
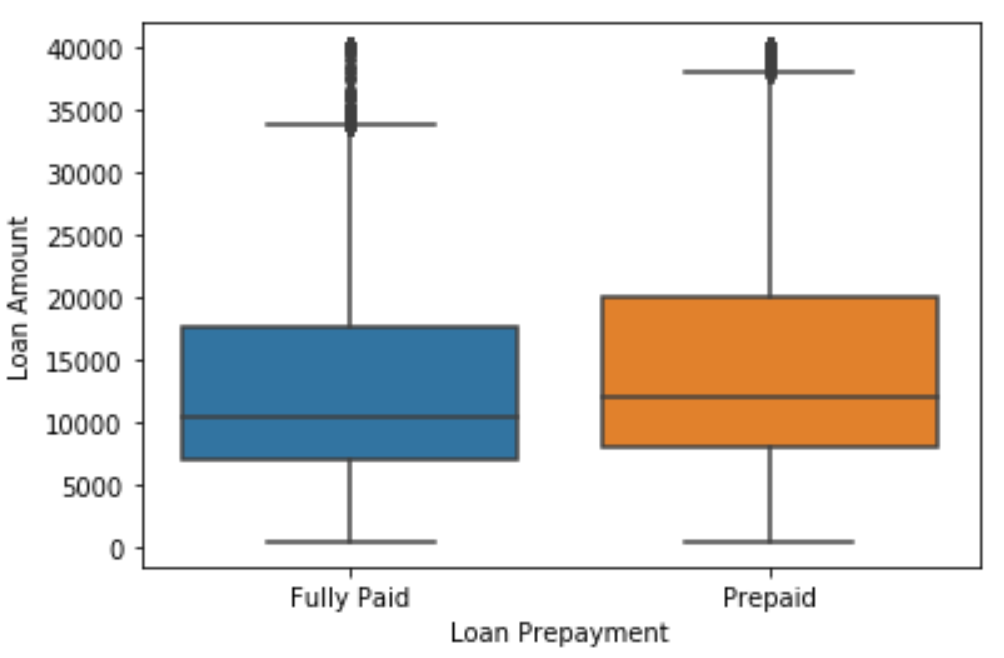




Interest rates play one of most key roles in predicting the outcome of loan. Interest rates vary across loan types based on variety of factors such as credit history of borrower, tenure of loan and current prevailing interest rate environment. Looking at the distribution of interest rates one can easily surmise that interest rates have a direct effect on the outcome of the loan, i.e., higher the interest rate of the loan higher the likelihood of it to default or be prepaid.

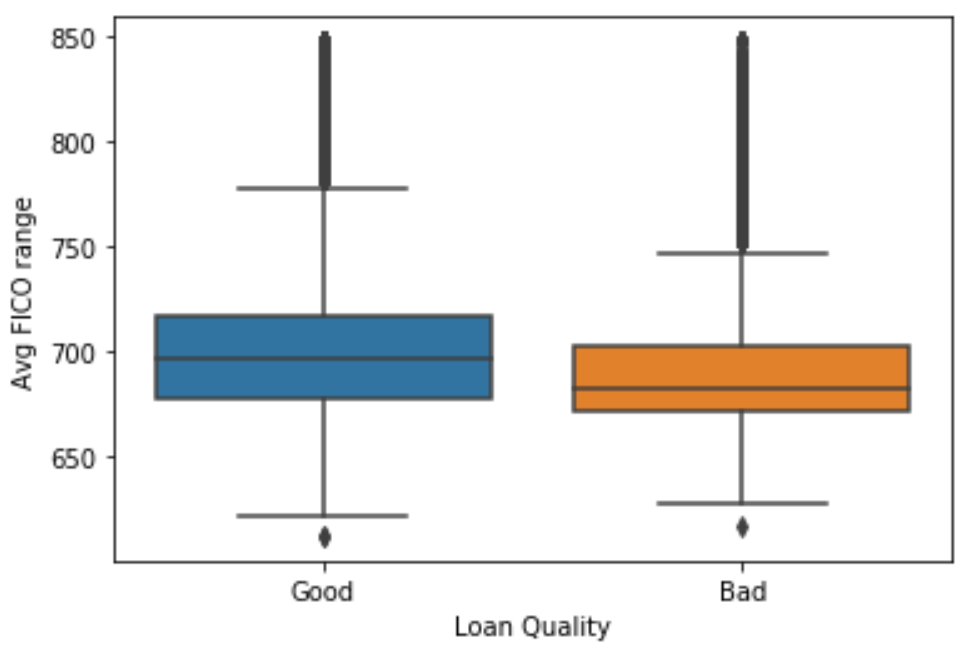
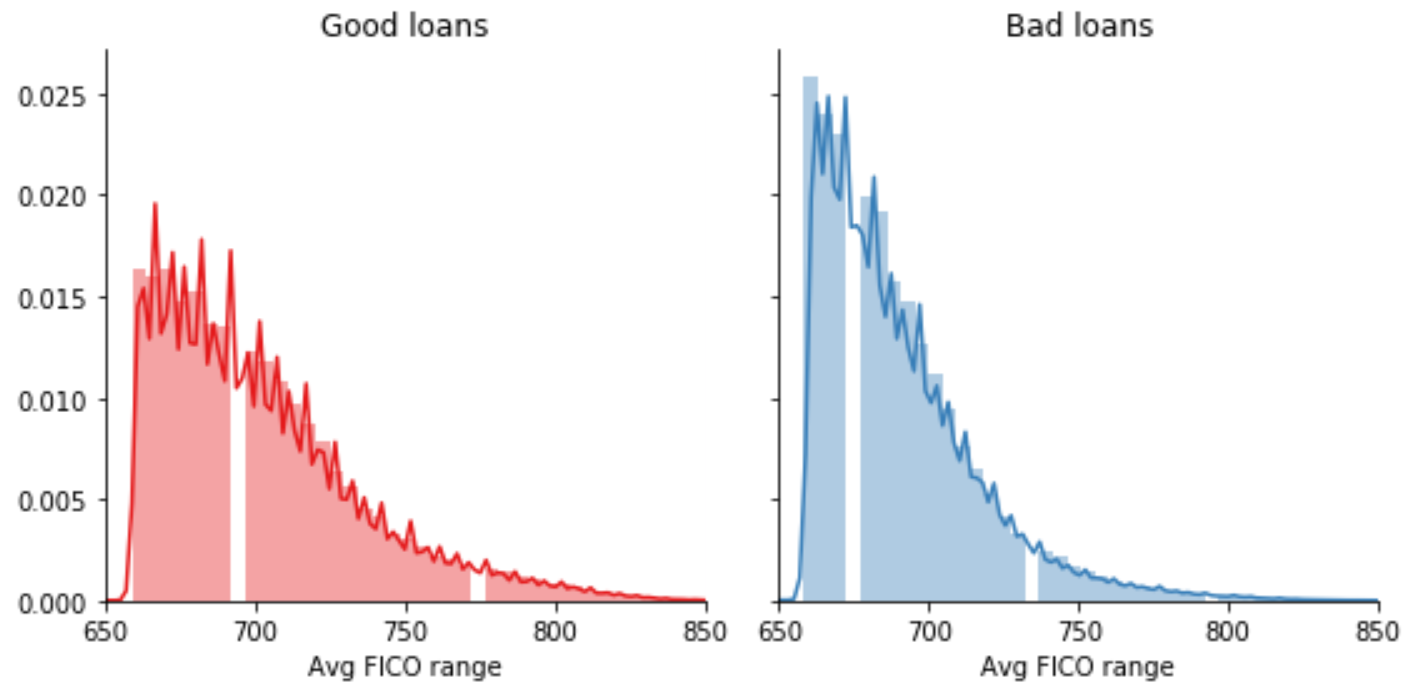
**Loan amount:**

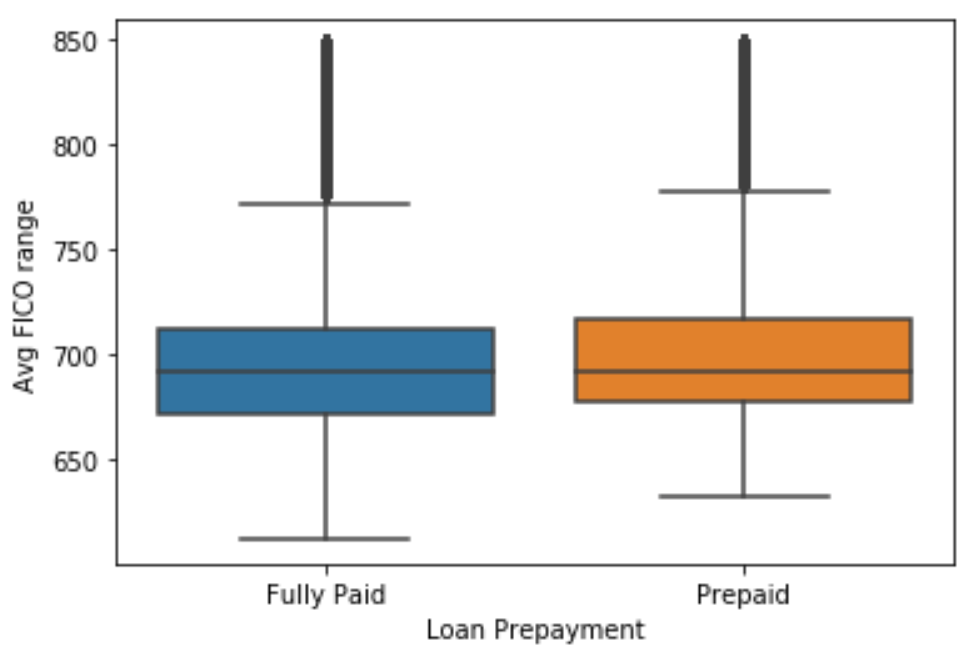
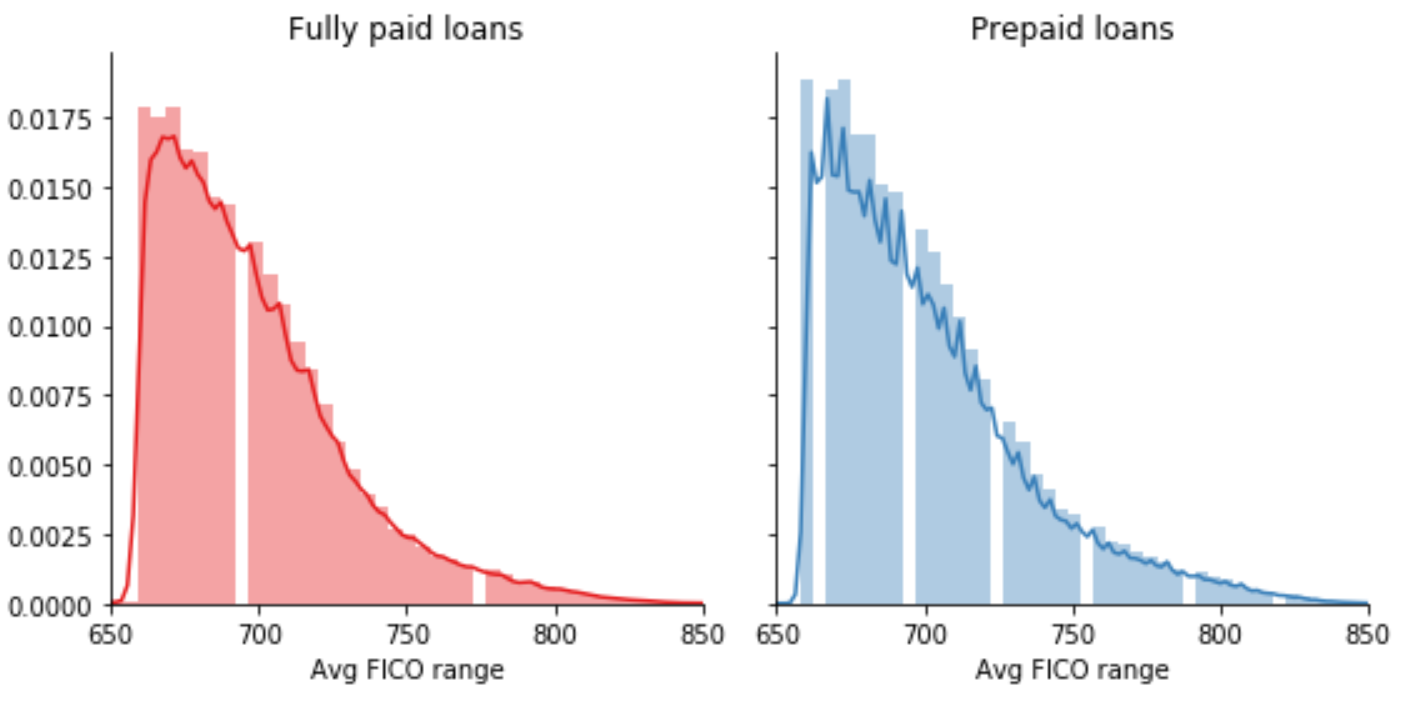




Loan amount plays an important role in predicting the outcome of a loan in conjunction with other features like interest rate and fico score. The distribution plots clearly depict the direct impact the loan amount on the loan status, i.e., higher the loan amount higher the likelihood for the loan to default or be prepaid.

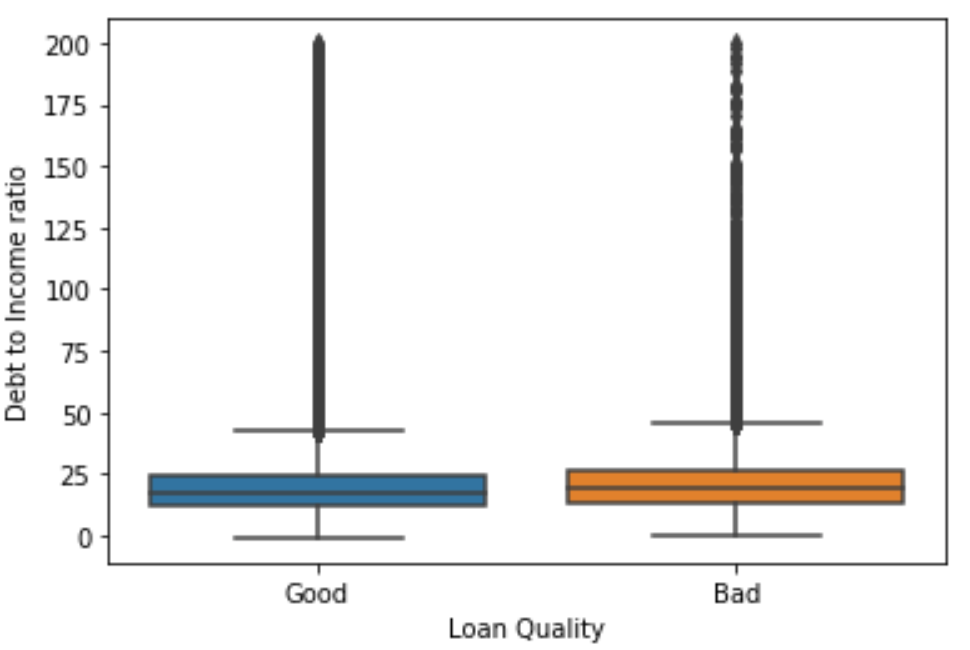
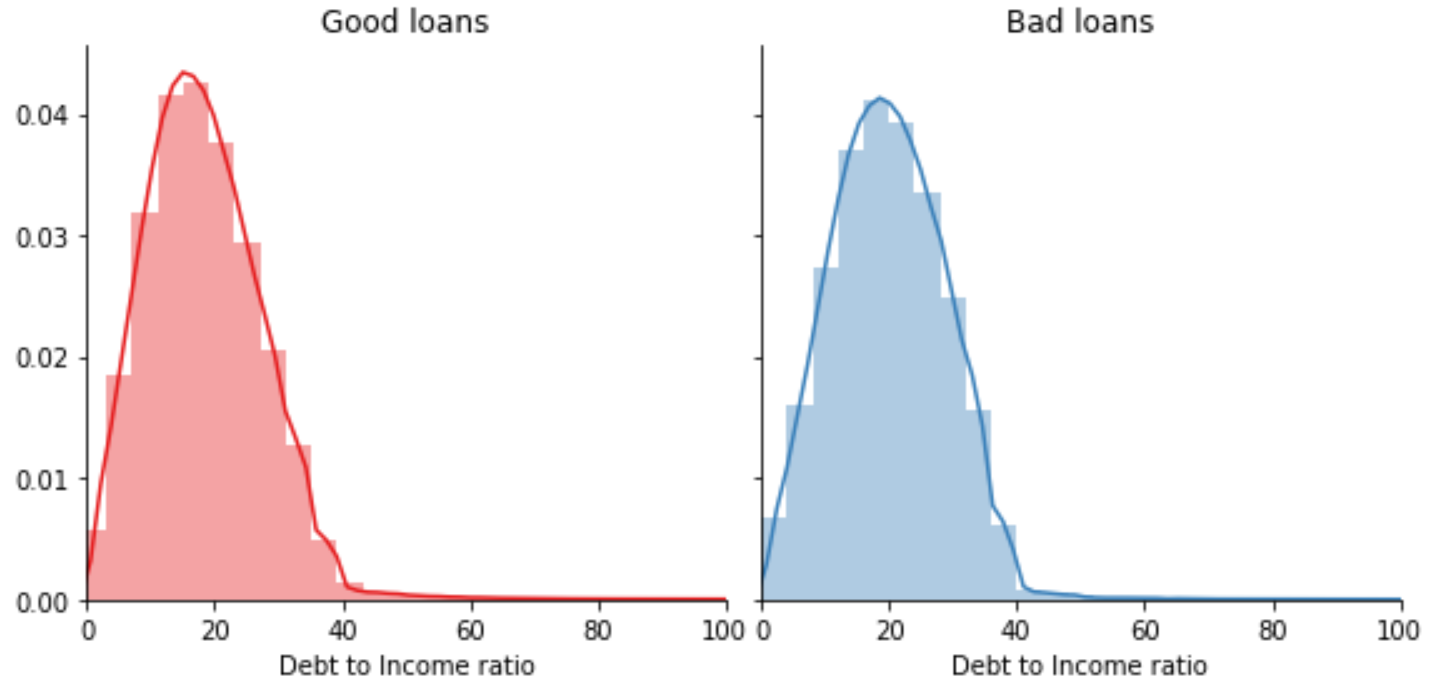
**FICO Score:**

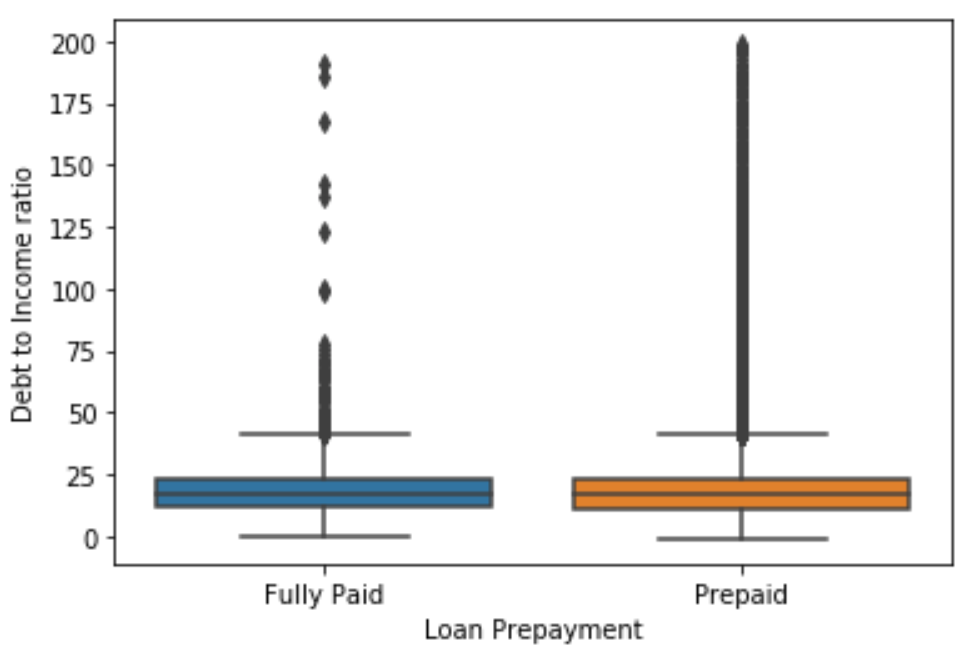
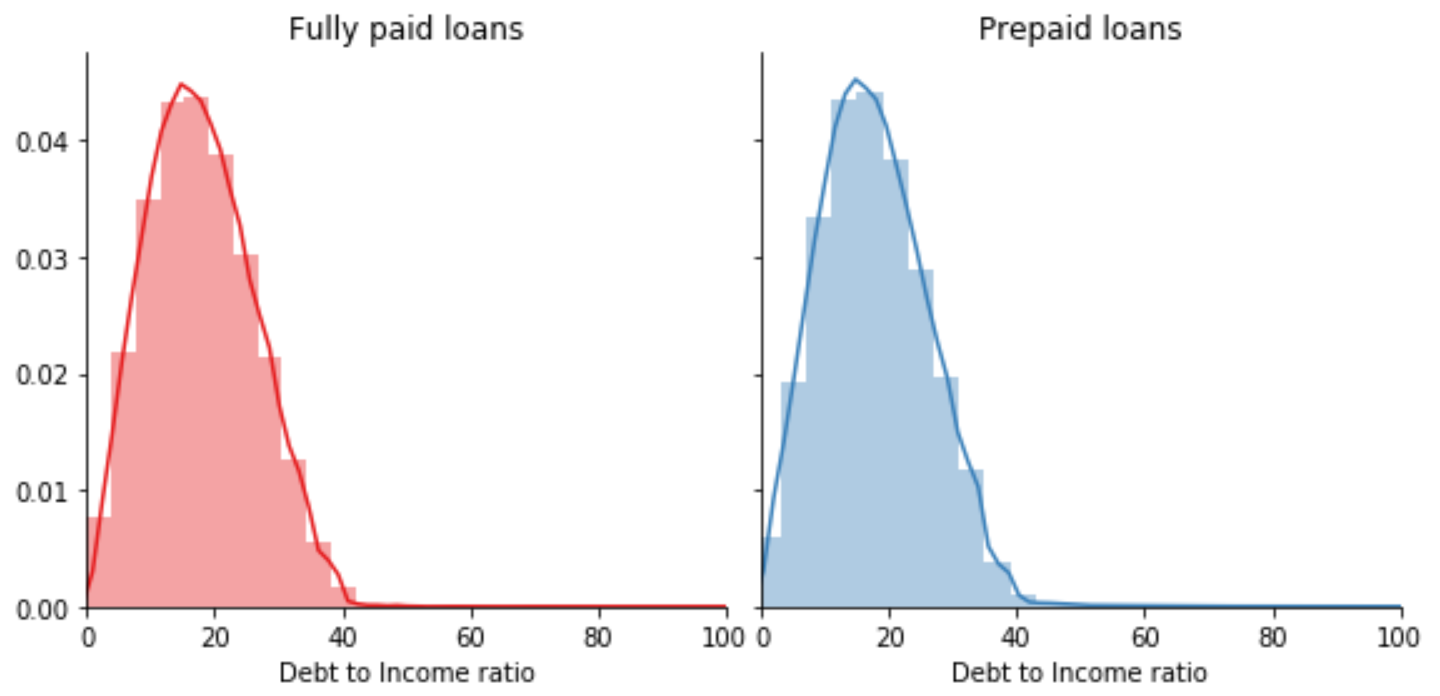
**** ****

**** ****

The credit score tells lenders about the credit worthiness of a borrower and is calculated based on the information made available from the credit reports. Thus credit score plays a pivotal role in predicting the outcome of the loan. Looking at the distribution plots in the two scenarios one can see the inverse role played by FICO score. Higher the FICO score higher the likelihood of the loan getting prepaid while lower the FICO score higher the likelihood of the borrower to default on the loan payments.

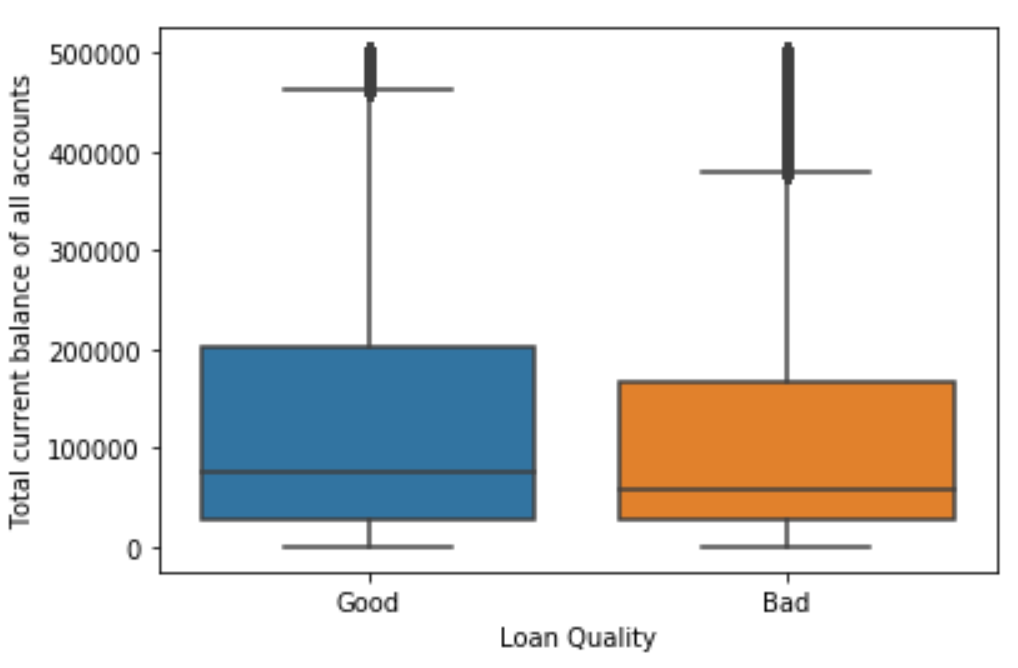
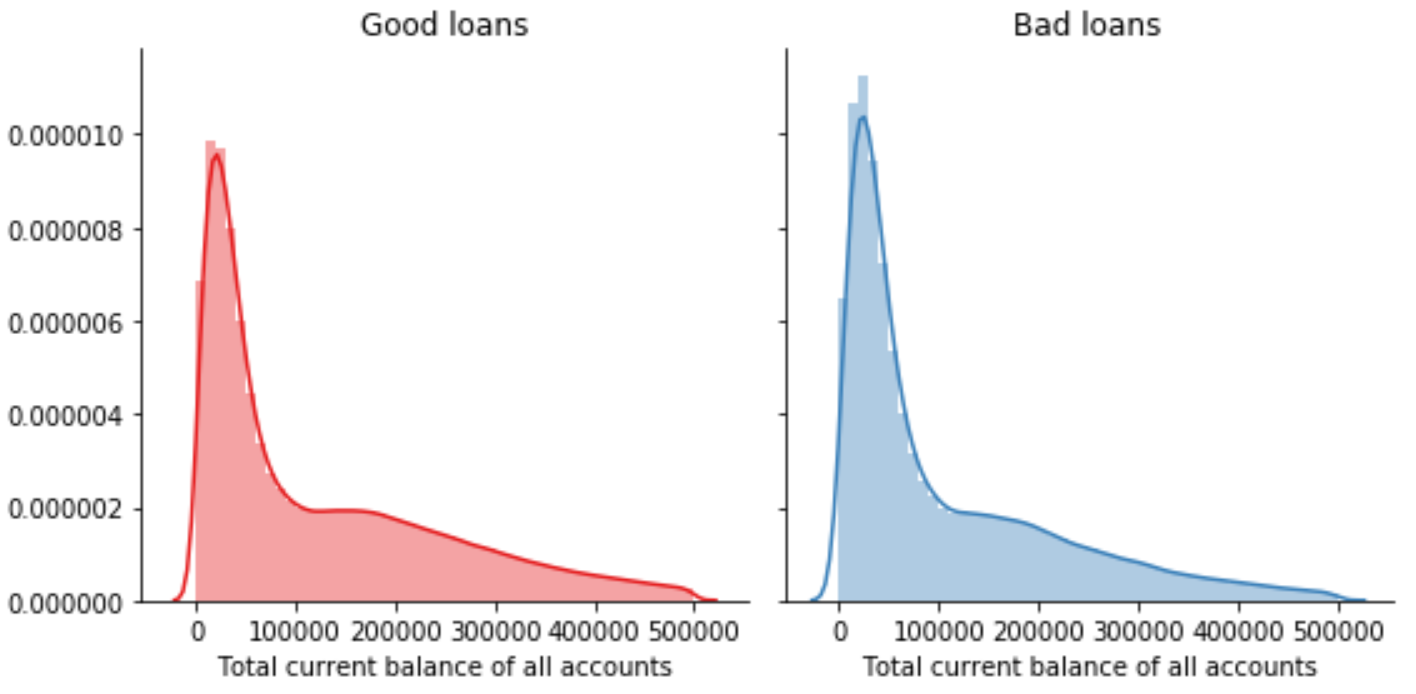
**Debt-to-Income ratio:**

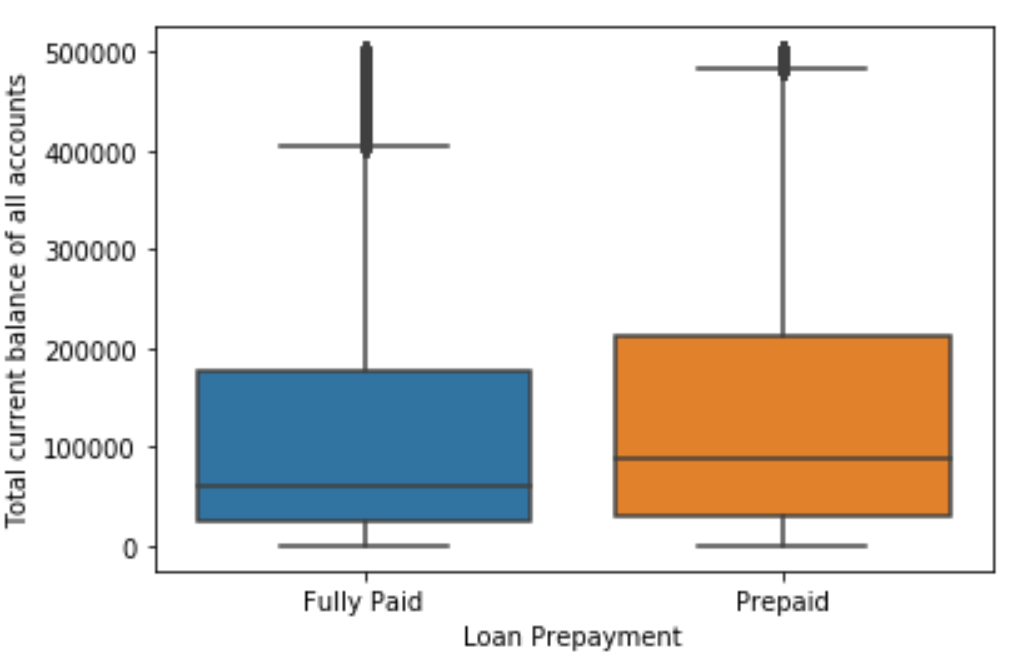
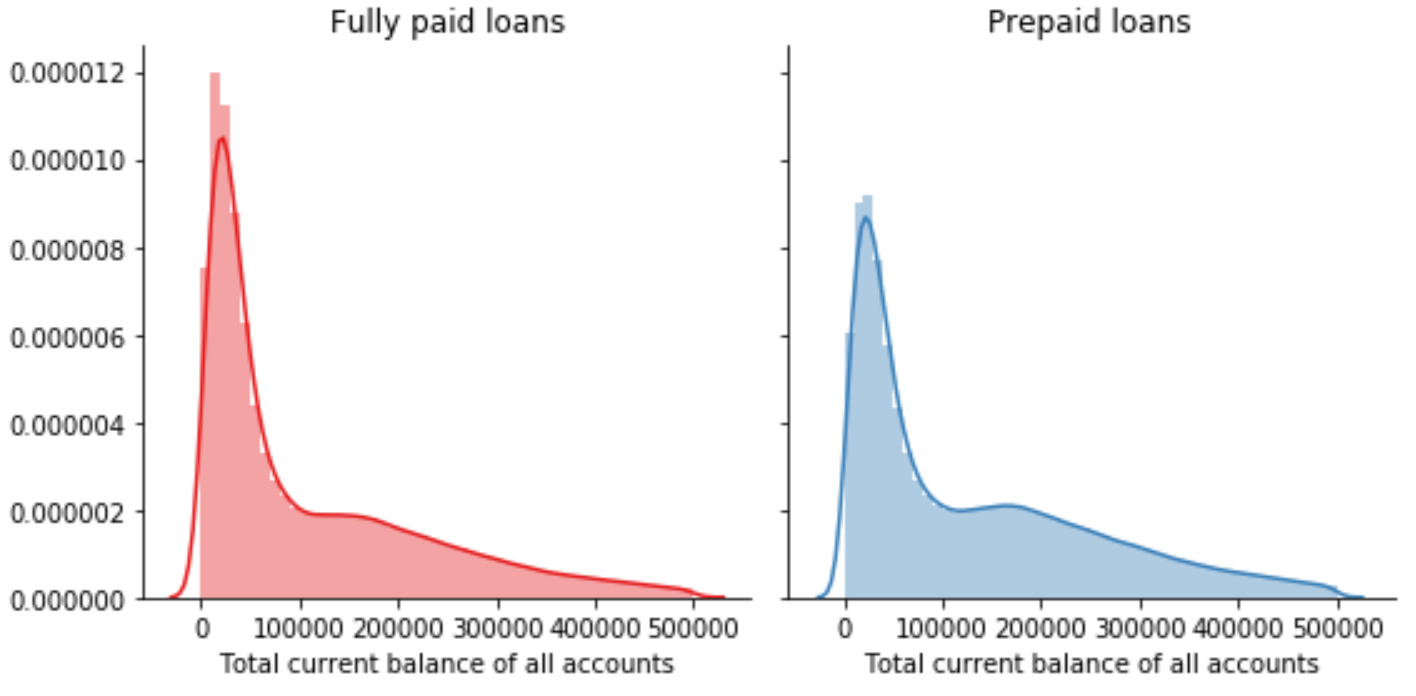
 

The debt-to-income ratio (dti) is the sum of all monthly debt payments of a borrower divided by his monthly gross income. It is a measure to use to judge the ability of the borrower to manage the monthly payments on the loan. LendingClub did an excellent job in screening applications based on dti and focused mostly on applications in the range 0-43. Evidence from studies of mortgage loans suggest that borrowers with a higher debt-to-income ratio are more likely to run into trouble making monthly payments. The 43 percent debt-to-income ratio is important because, in most cases, that is the highest ratio a borrower can have and still get a loan approved.

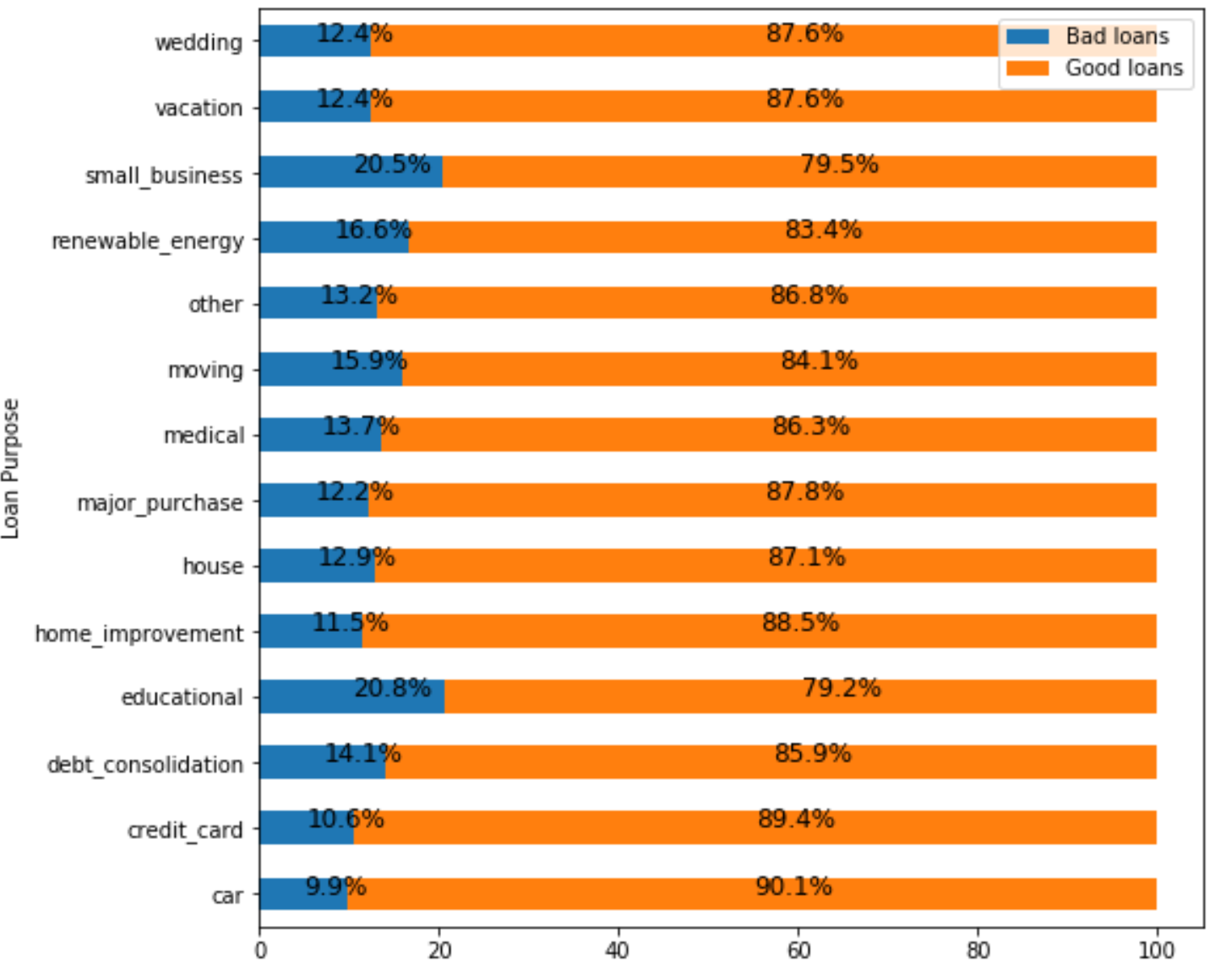
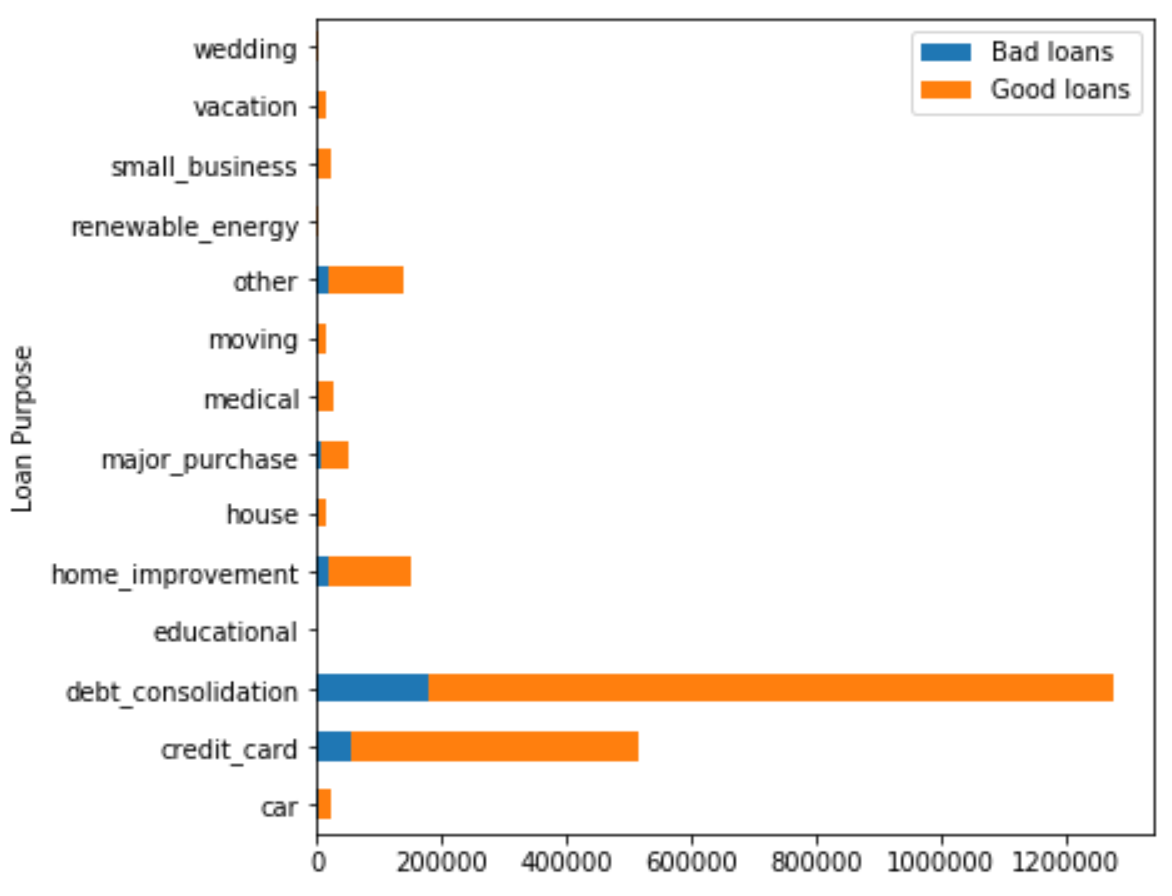
**Total current balance of all accounts:**

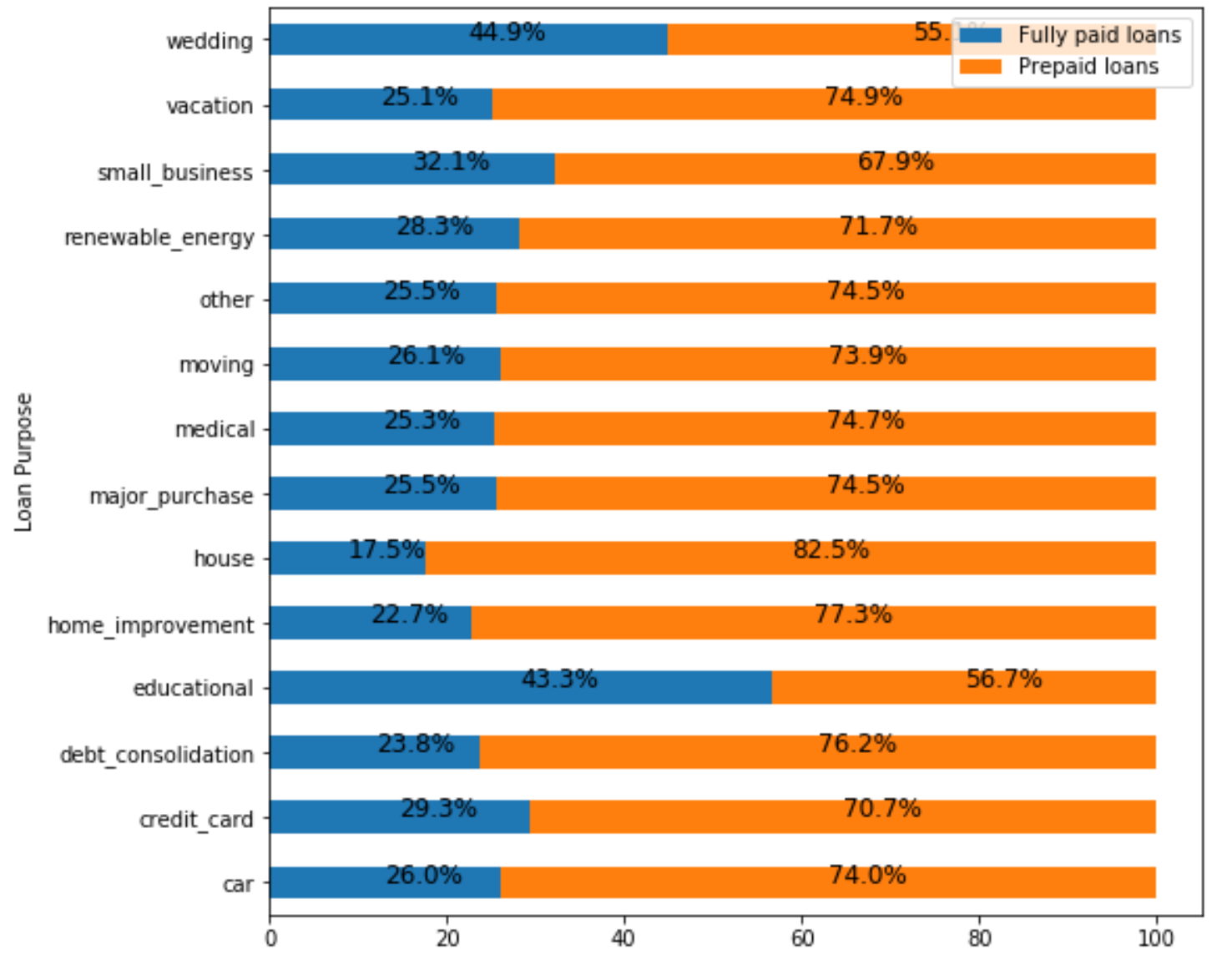
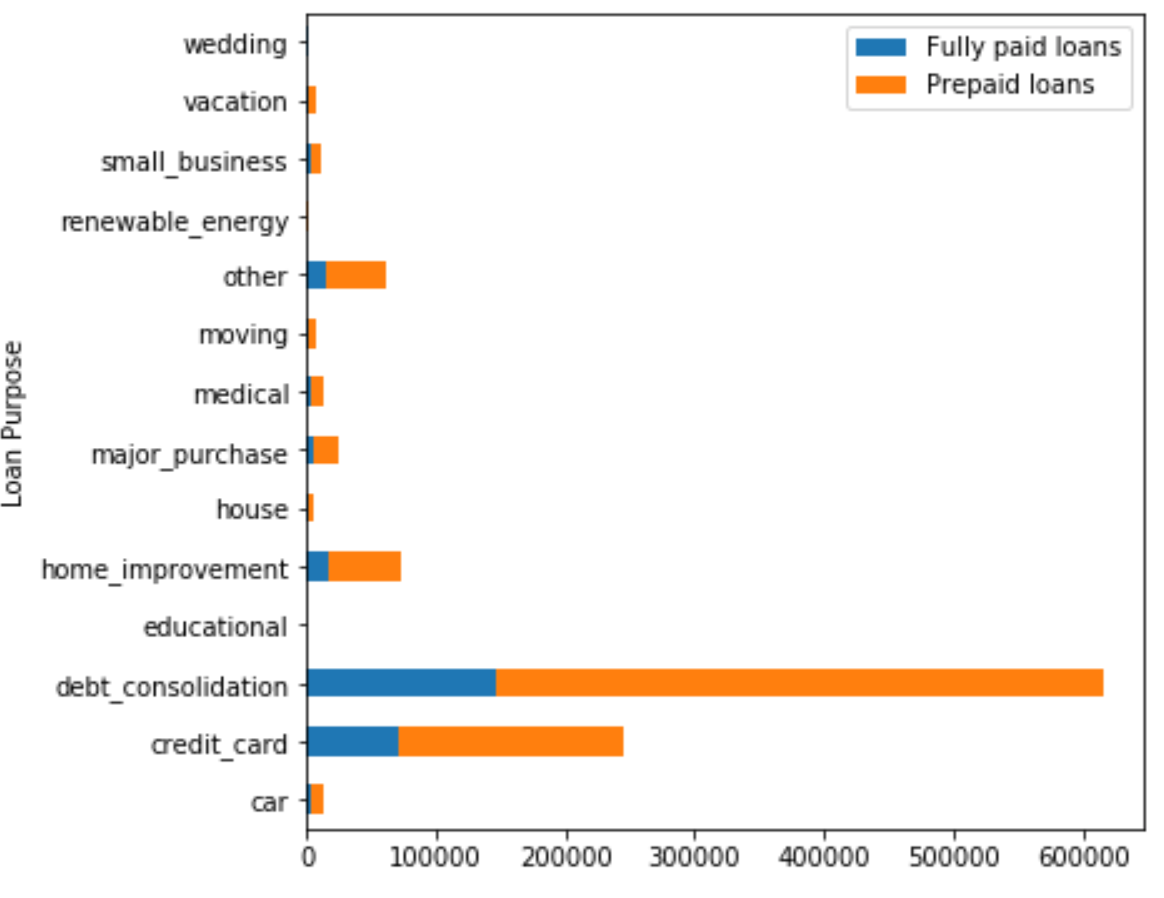
 

As the name of the feature suggests, this feature also plays an important role in predicting the outcome if a loan. Higher the total current balance available on all accounts higher the likelihood to prepay the loan using these funds. Lower the balance higher the likelihood to default on the loan amount.

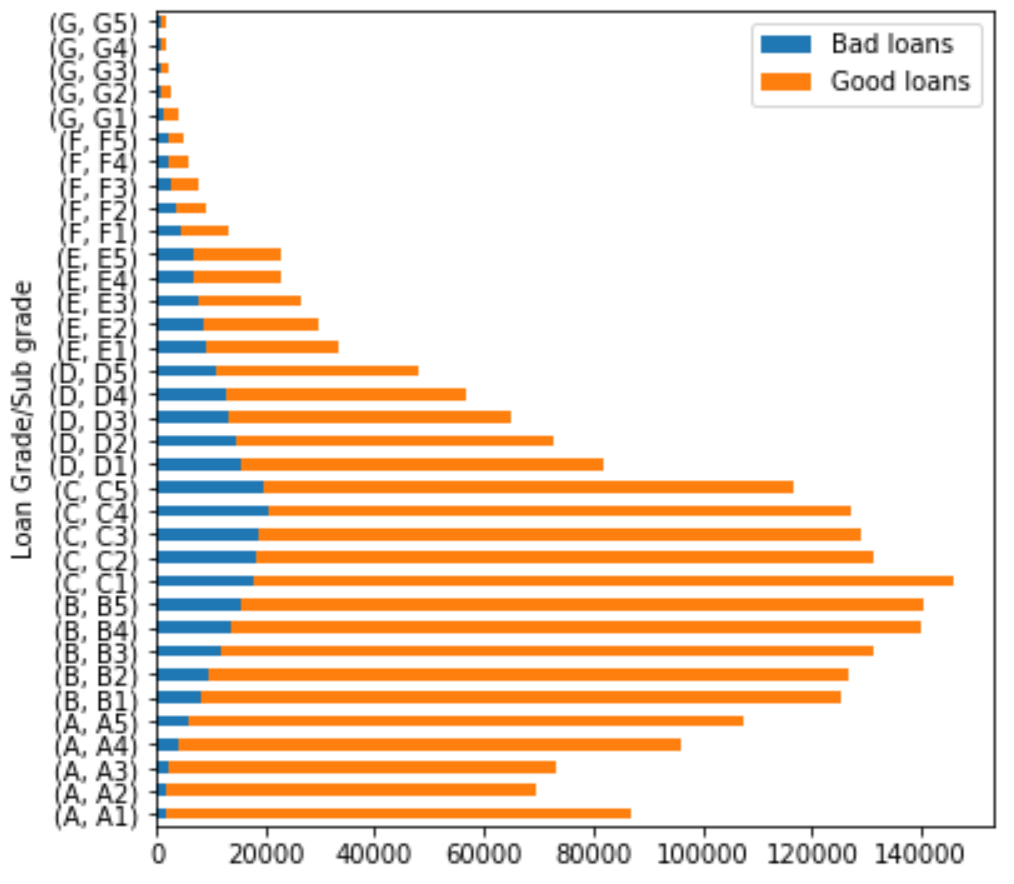
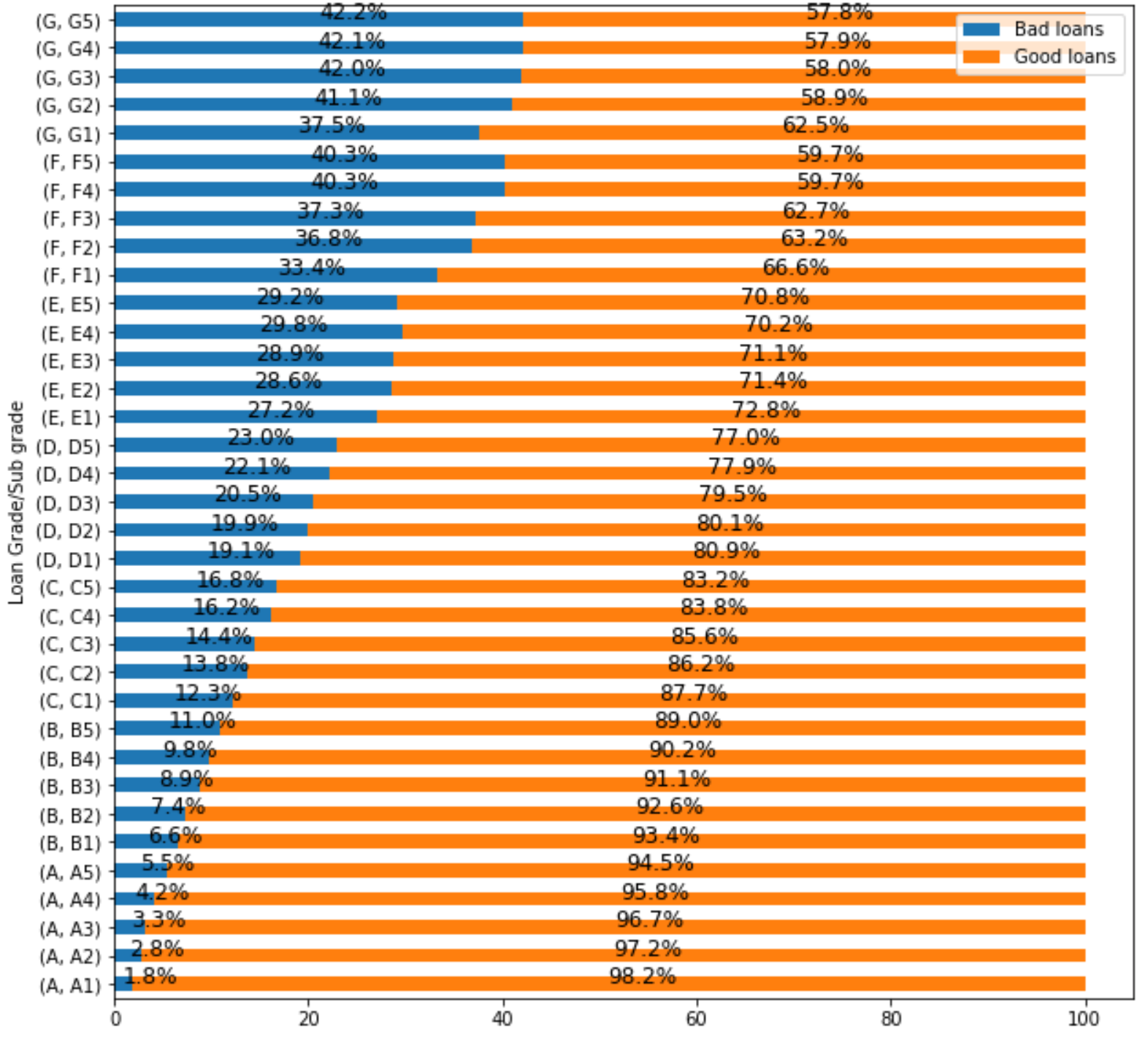
**Loan Purpose:**

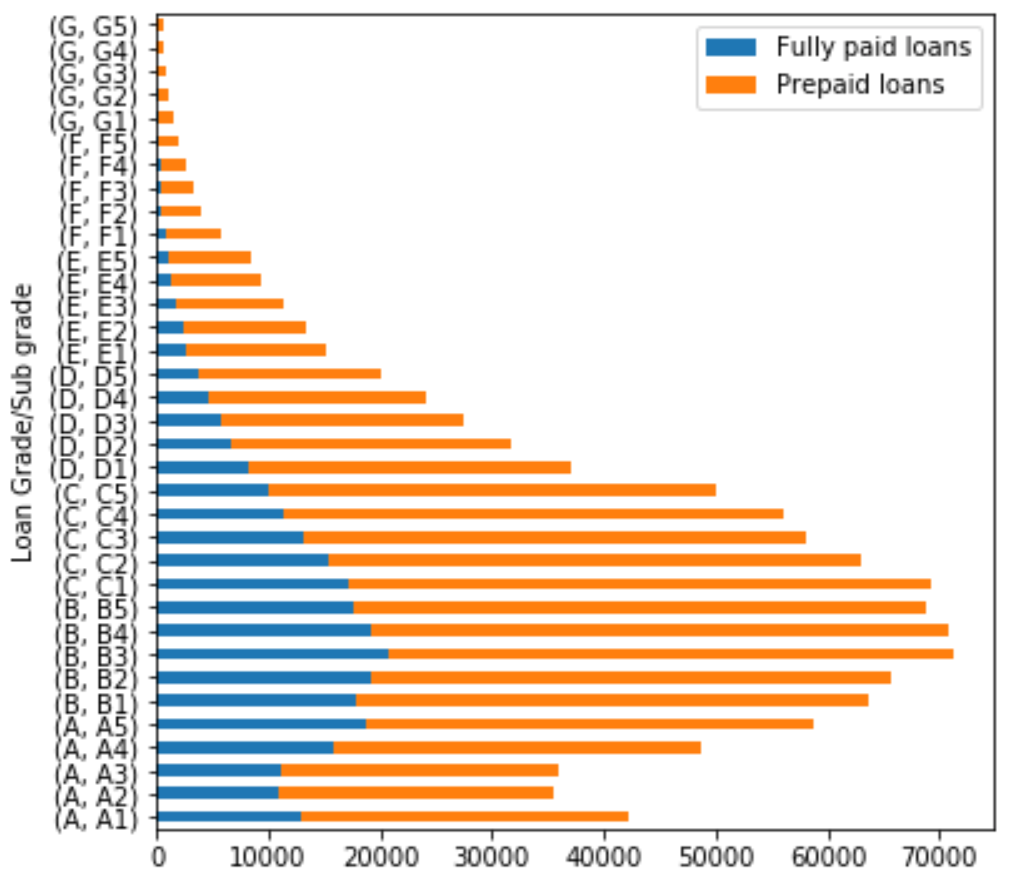
****

****

In this section we evaluate the effect of loan purpose on the outcome of loan status. Debt-consolidation, credit card and home improvement seem to be the most common reason why borrowers apply for loan at LendingClub. Not only are these 3 categories the most commonly stated reason to apply for a loan, they are also relatively safer both in terms in default probability and prepayment when compared to other categories. On the contrary loans taken for the purpose of education, small business and renewable energy seem to be the riskiest both in terms of default and prepayment.

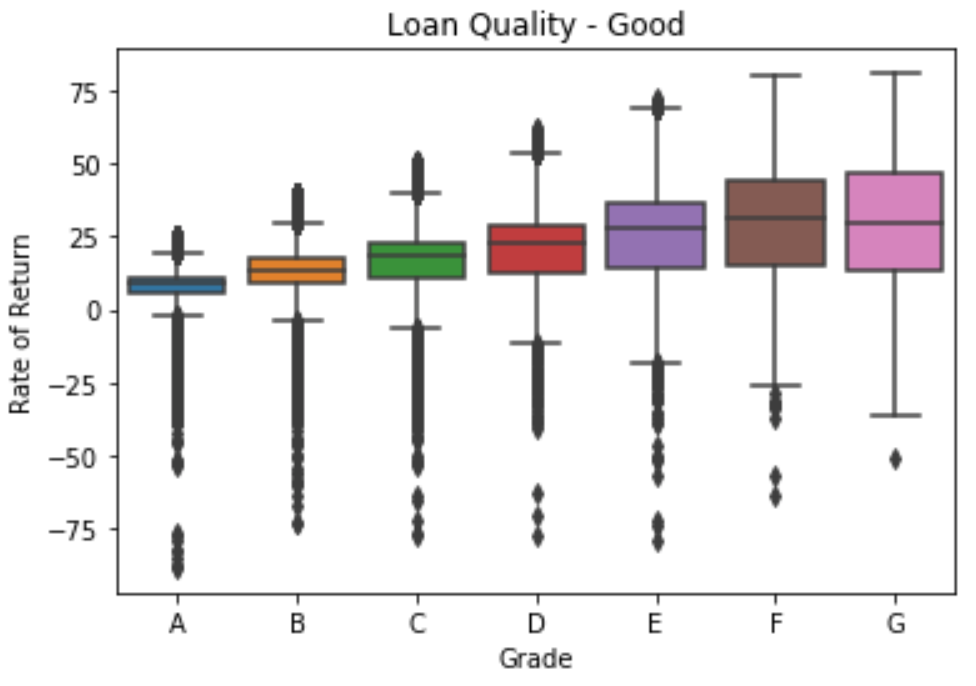
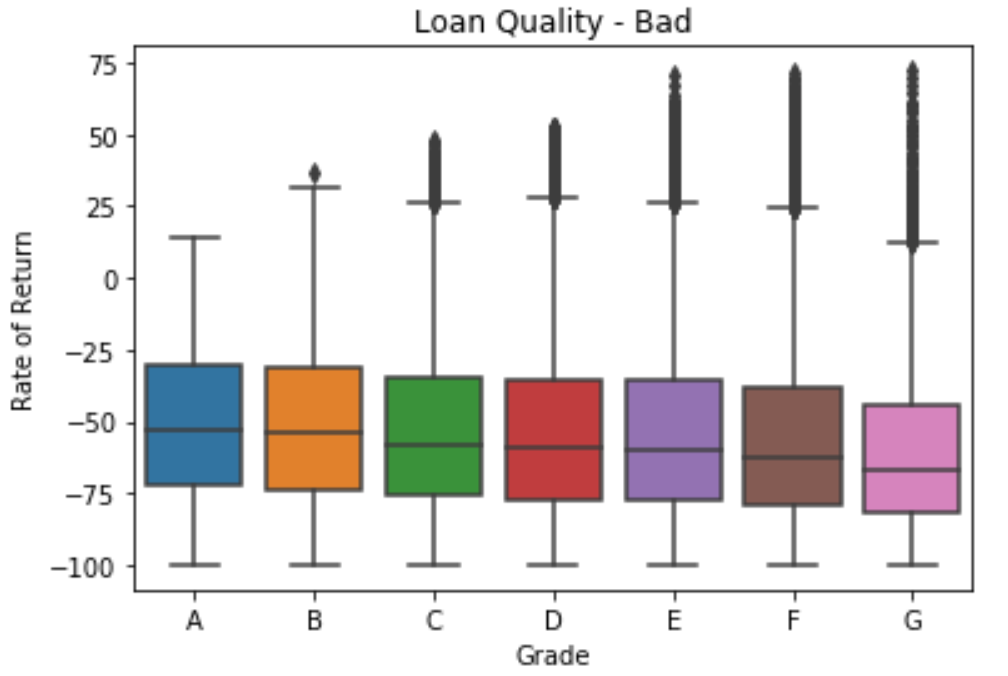
**Loan Grade/Sub-grade:**

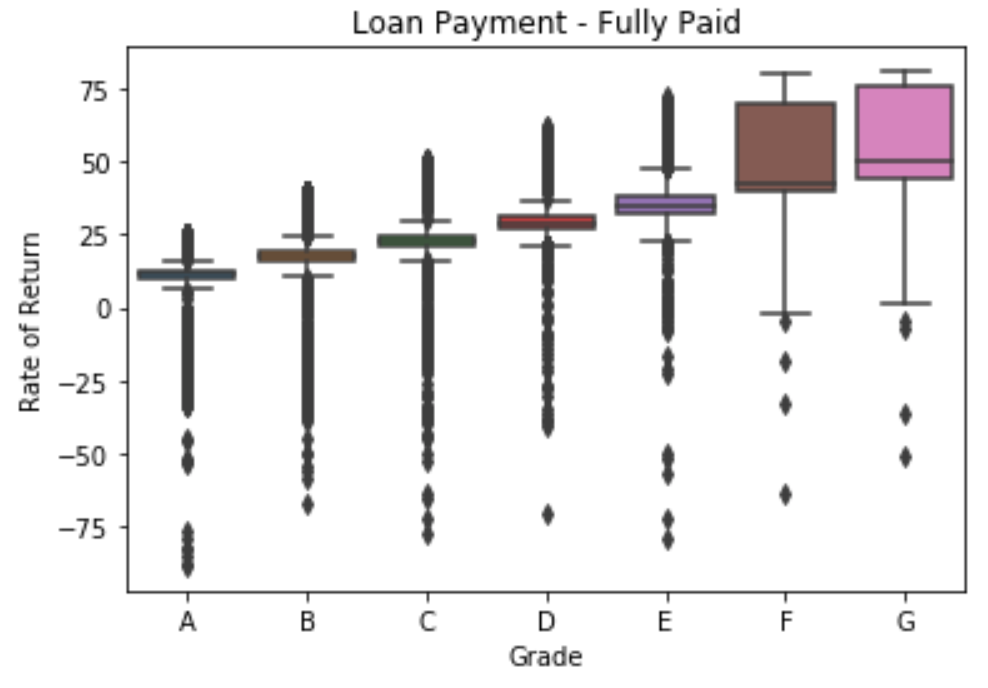
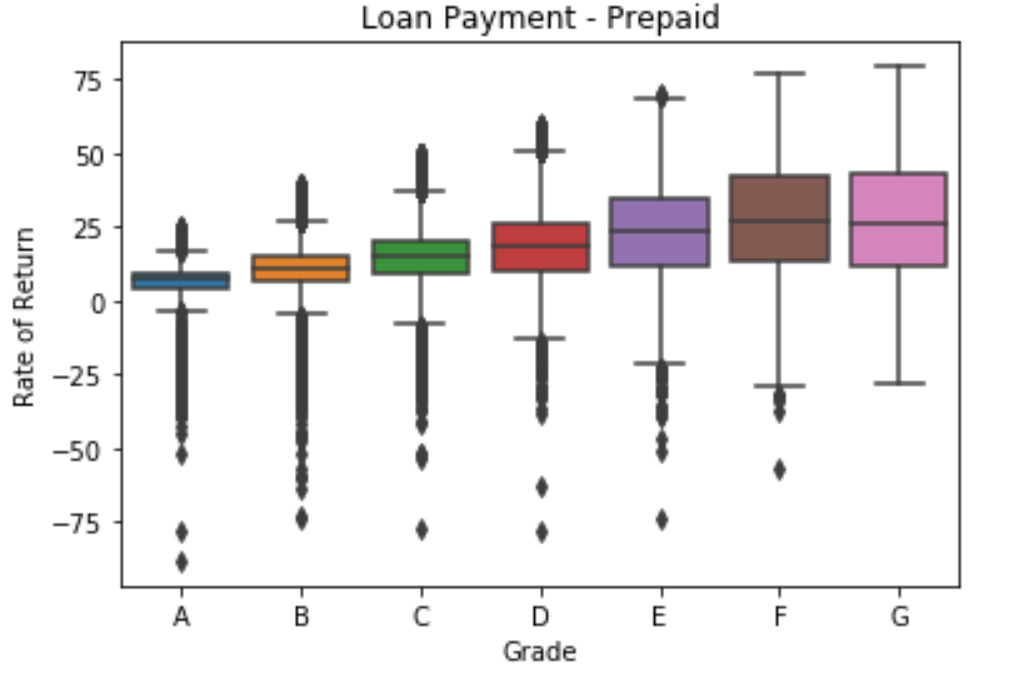
**** ****

**** ****

We now study how the combination of loan grade/subgrade effect the outcome of a loan. In both the scenarios, default/prepayment, majority of loans are allocated in the B and C grade category. As the loan grade diminishes from A to G, we observe a corresponding rise in defaults and vice versa as the loan grade improves from G to A, we observe a rise in prepayments. It is interesting to note how borrowers of loans in the lowest G grade that managed to fully pay find it almost impossible to prepay due to the huge interest rate burden. Conversely, borrowers of loans in highest grade manage to prepay due to the lower interest rate burden.

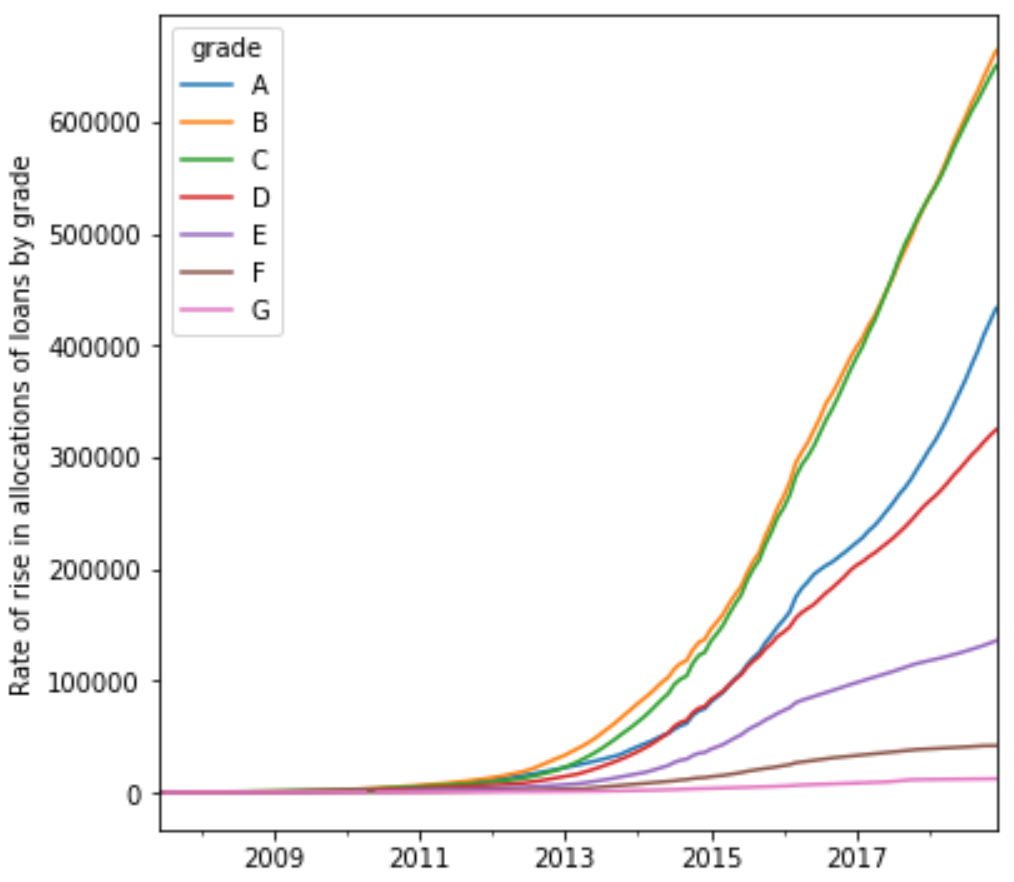
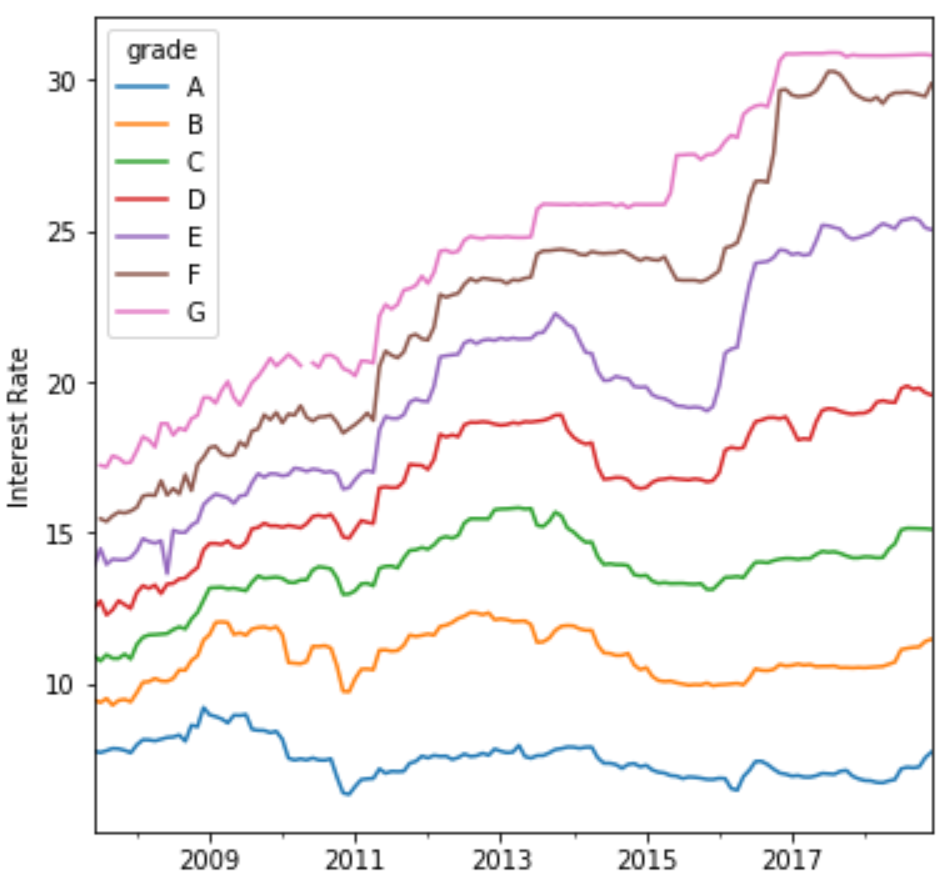
**Rate of return:**

****

The following graphs show the distribution of rate of return across loan grade categories in the events of default, prepayment and full payment. Without any surprises the rate of return is highest when loans are fully paid and lowest when the borrower defaults. Also, we see the differences in return when loans get prepaid instead of being fully paid due to lost interest payments. The returns of loans in grade G vary from -75% to 75%, clearly highlighting the importance of loan grades in assessing the risk of investments.

**Relationship between interest rates and allocation of loans across grades:**



In this section we study the variation of interest rates across loan grades over the period of time and rise in allocation of loans across loan grades over the same period of time. The interest rate gap between highest and lowest grade categories more than doubled over the period of time. There are two instances where interest rates in lower grade categories received a significant bump upwards. Over the period of time, allocation of loan across categories B and C increased the most, with investors finding these grades as sweet spot with the ideal risk-return characteristics. Most of the investors avoided loans in grades F and G, with some recent increase in interest in loans in category D and E.



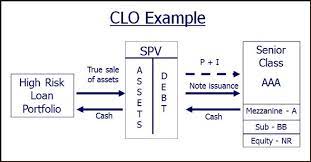
Due to increased allocation of loans in category C there was an increase in defaults with the rate of rise in defaults much greater than the rate of rise in allocation in this category. Categories A and B also saw a rise in defaults but at a much slower pace compared to the rate of allocations.

***A machine learning approach:***

We use a machine learning approach to build models that will assist investors in predicting the likelihood that a loan will default or be prepaid. The following machine learning approaches: Decision Trees, XGBoost and Neural Networks, were used to train on the available dataset and build two hybrid models that would work in our case. Due to the non-linear nature of this process, we found these three modeling techniques serve the purpose both in terms of computational accuracy and efficiency.

While one model will be used in assisting the investor to calculate the likelihood of default of a loan, the other will assist the investor in calculating the likelihood of prepayment. We believe based on these available inputs the investor will be able to make a more educated guess in deciding which loans match their risk-reward characteristics and invest in them. It will also allow the investors to diversify their portfolios between loan types thus maximizing returns while minimizing the risk.

***A Collateralized Loan Obligation application:***



**Collateralized loan obligations** (**CLOs**) are a form of securitization where payments from multiple middle sized and large business loans are pooled together and passed on to different classes of owners in various tranches. A CLO is a type of collateralized debt obligation (Source: Wikipedia). The loans generated by LendingClub fall in the subprime or junk category, especially considering that fact that they are not backed by any collateral makes investing in them extremely risky. In this study we explore the advantages of applying a CLO based approach in diversifying the risk involved with such loans.

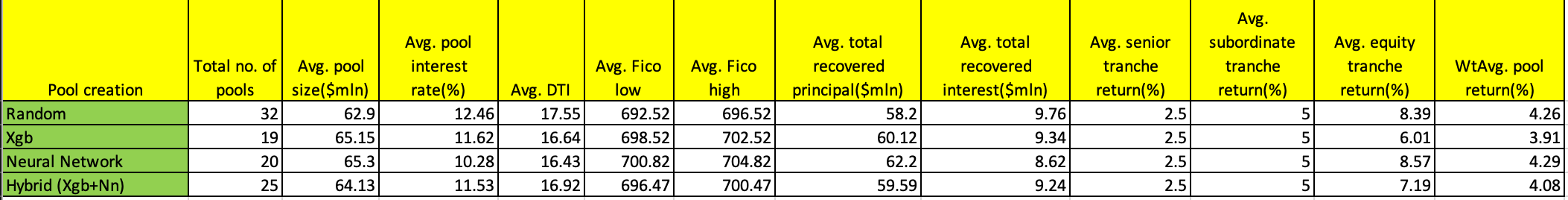
We pooled together loans created in year 2014 with 36 months term length, with each pool consisting of 5000 individual loans. A CLO is a complex marked instrument, in order to simplify calculations, we assumed the CLO consists of only 3 tranches. The senior tranche having the first priority in payments with 2.5% rate of return and 50% contribution of the loan portfolio. The subordinate tranche having the second priority in payments with 5% rate of return and 35% contribution of the loan portfolio. Finally, the equity tranche with 15% contribution of the loan portfolio and the last tranche to get interest payments after all the previous tranches have been paid, making it the riskiest tranche to invest in.

In the first simulation we randomly selected the loans that went into the pool and there were 32 pools in total. As is evident from the figure below both the higher priority tranches, senior and subordinate, receive their promised interest payments. The equity tranche has an average return of 8.39%.

In the second simulation we filter out loans predicted as bad by XGBoost, we end with only 19 pools in total. Both the high priority tranches received their promised interest payments. The equity tranche had an average return of 6.01%.

In the third simulation with used neural networks to filter out loans with a high probability of default. In this case we ended up with 20 pools in total with both the senior tranches receiving their share. The equity tranche had an average return of 8.57%.

In the final simulation we filter out the loans that have been predicted by both the XGBoost and Neural Network models as most likely to default. We end up with 25 pools in total, and even in the case both the senior and subordinate tranches receive their promised rate of return of 2.5 and 5% respectively. The equity tranches in the case have 7.01% average rate of return.



In the future and based on the current gained insights we would like to automate the process of loan selections and pool these loans together to create instruments that can be traded in secondary markets similar to the way mortgage loans get traded. These market instruments will have varied risk-return profile that can be used to attract fixed income investors that are looking to gain exposure to this asset class and thus diversify their portfolio.