LENDING CLUB CASE STUDY

SUBMISSION

Group Members:

Abhinav Joshi (Group Facilitator)

Tasheer Hussain B

Lending Club Case Study

Business Understanding:

You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

When a person applies for a loan, there are two types of decisions that could be taken by the company:

- **Loan accepted**: If the company approves the loan, there are 3 possible scenarios described below:

  Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

  Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

  Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

- **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Objective:

- To identify these risky loan applicants, loans can be reduced thereby cutting down the amount of credit loss.
- To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Problem Solving Methodology

Importing Data

Importing the data from the csv files

Data Cleaning

- Handling NULL Columns
- Handling NULL Rows

Data Manipulation

- Removing Columns that are not important for analysis.
- Formatting the data in columns to one data type
- Creating derived columns from available columns to support with the analysis

Univariate Analysis

- Performing Univariate Analysis on each of the columns to understand the distribution of data and to gather insights.

Bivariate Analysis

- Performing Bivariate Analysis to understand the effect of one variable with the other and their relationship.
- Gather insights based on the above analysis

Conclusion

- Evaluate and aggregate the insights obtained from the previous steps to derive at specific conclusions accomplishing the objective of the analysis.

Data Sources

Source of Data are as follows,

1. Loan Dataset, A file containing the complete loan data for all loans issued through the time period 2007 t0 2011 (Load Data Set.csv)

2. Data Dictionary, A file describing the meaning of each variable

Data Cleaning and Manipulation Performed

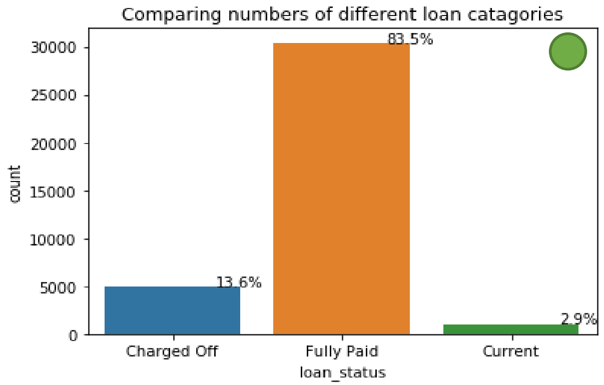
- Dropping all the columns that contain NULL values for more than 50% of the data.
- Dropping other columns that are not necessary for our analysis, eg: desc, id, member_id, url, zip_code
- Dropping columns that contain only a single value.
- Dropping all the rows having NULL values.
- Dropping Columns containing redundant data, eg: funded_amnt,funded_amnt_inv
- Performing Data Formatting on specific columns to retain the necessary data and also deriving at new columns from available columns aiding in the analysis.
- Some analysis is on the complete data, while some is on just the defaulter data.

Note:
Complete Data : The graphs containing this color consists of the complete data.
Defaulter Data : The graphs containing this color consists of just the defaulter data

Univariate Analysis

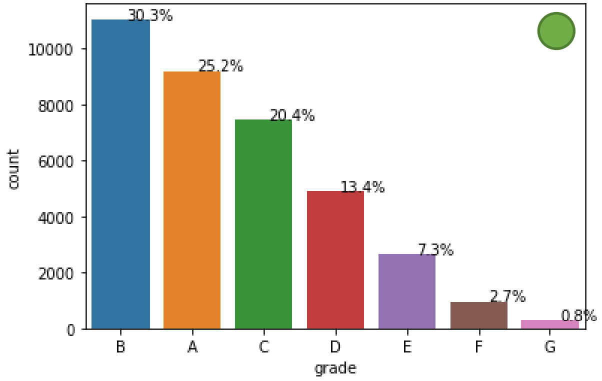
UpGrad

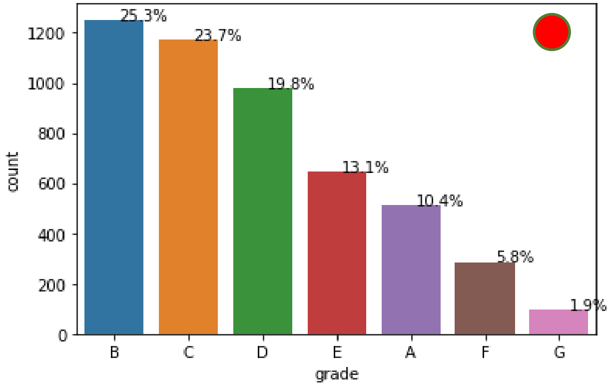
Loan Status



Out of the number of loans the company has given away, 83.5% have been paid back, while 13.6% have been charged off.

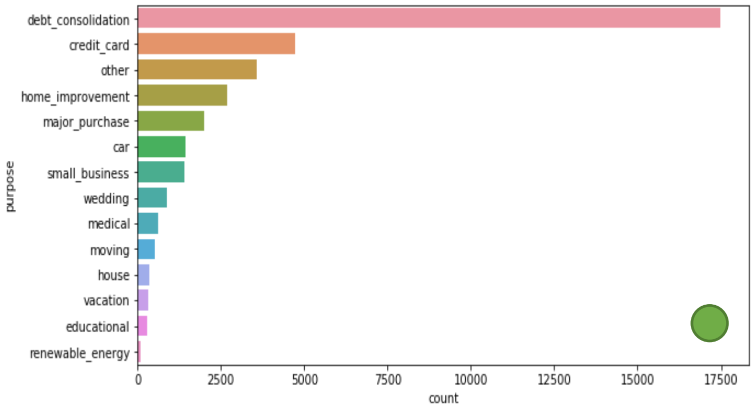
Grade

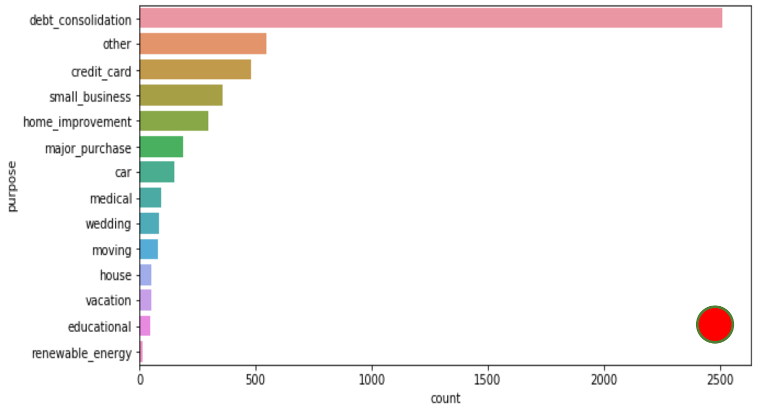




Majority of the loans have been distributed in the higher grades itself, the similar trend is observed in defaulter data, where the defaulters are also in the higher Grade range.

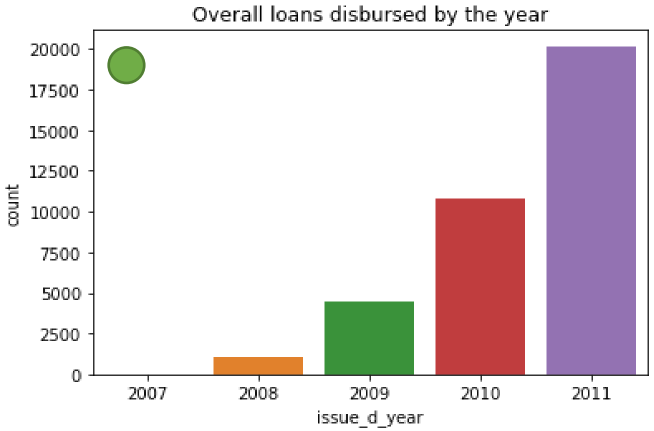
Purpose





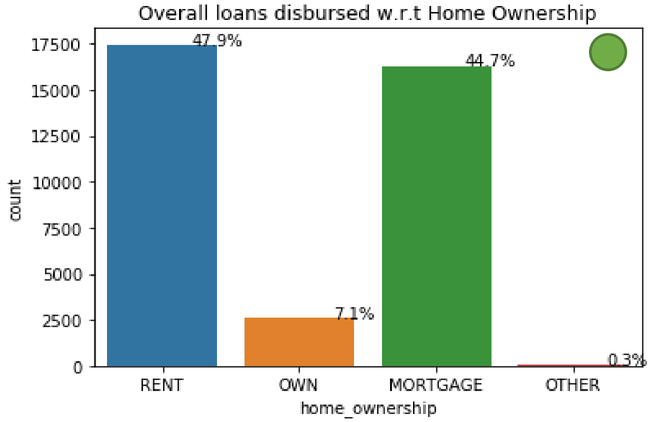
Majority of the loans have been taken for debt consolidation and the majority of loan defaulters are also from the people who have taken a loan for the purpose of debt consolidation

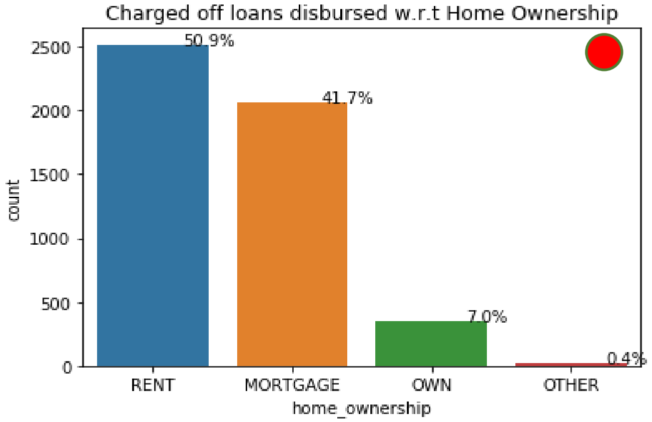
Loan Year



The number of loans provided by the company increases gradually every year.

Home Ownership

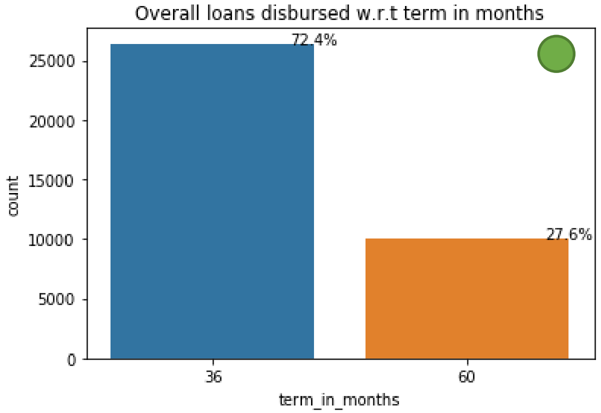


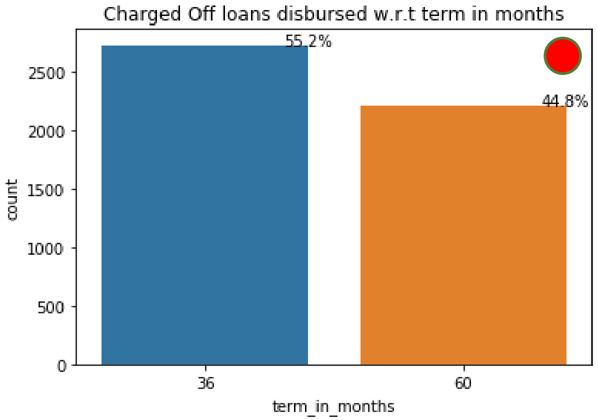


Majority of the loans are taken by the one's who do not own a house, implying they do not have enough income/assets.

Majority of Loan Defaulters are also people who do not own a house

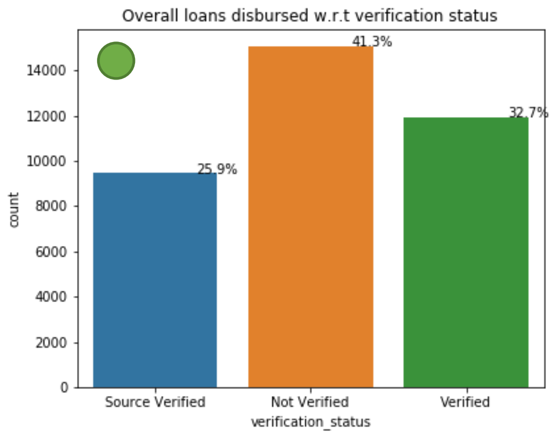
Term

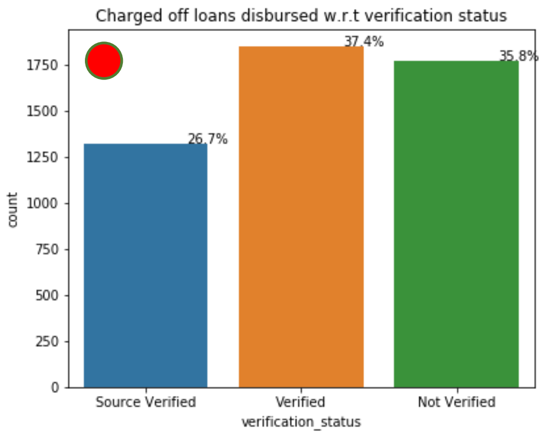




Majority of the loans have been taken for 36 months term, and 55% of the defaulters are the people who have taken the loan for 36 months term

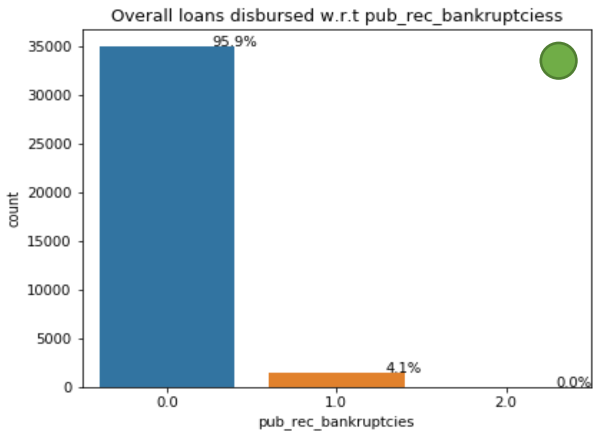
Verification Status

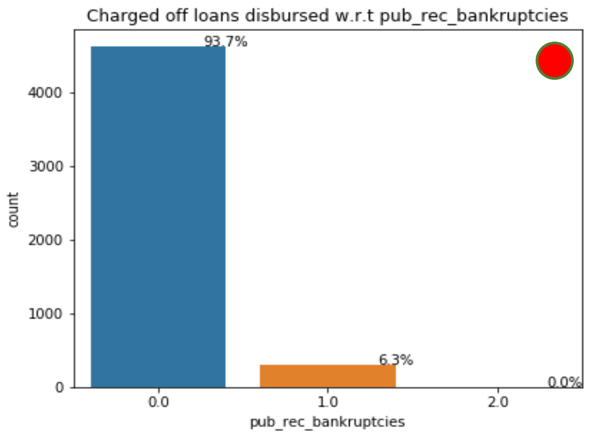




The number of loans that are not verified exceed both the source verified and verified loan categories. And these not verified loans have a very high rate of being charged off. This is a matter of concern and the company should take due diligence before advancing loans.

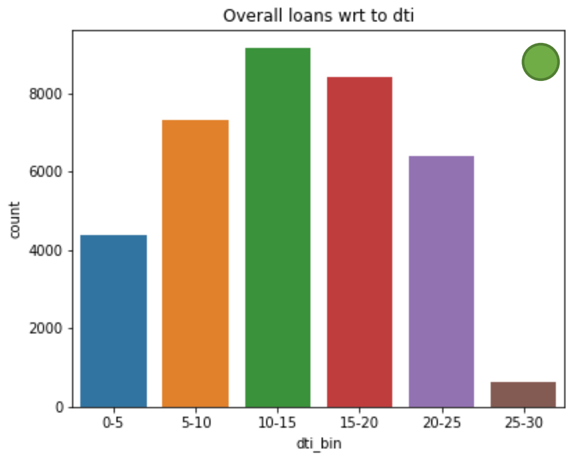
Public Record of Bankruptcy

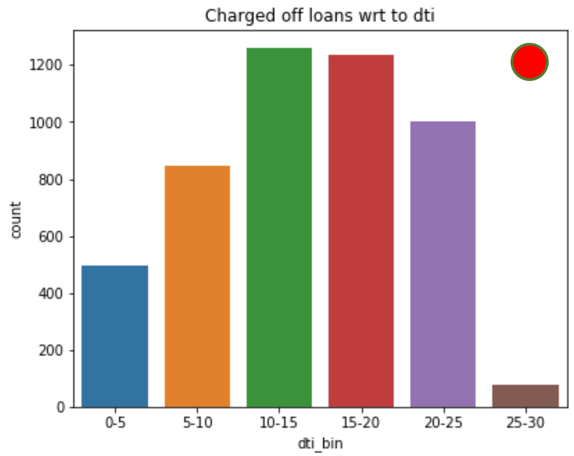




Very few accounts with prior bankruptcy record are given loans. The accounts with prior bankruptcy record that do get loans, have a high probability of being charged off.

DTI



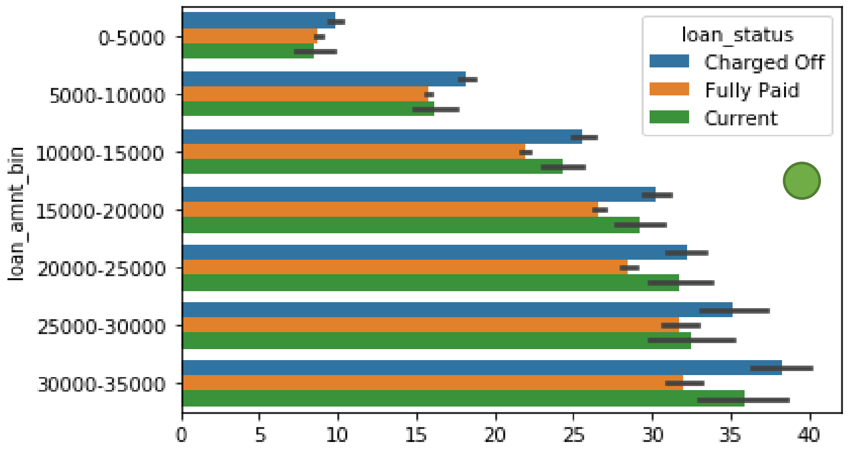


Trends in DTI for charged off loans follow the trends for overall loans, except in the bin 20-25, which means that for DTI between 20-25, there is a higher probability of loans being charged off.

Bivariate Analysis

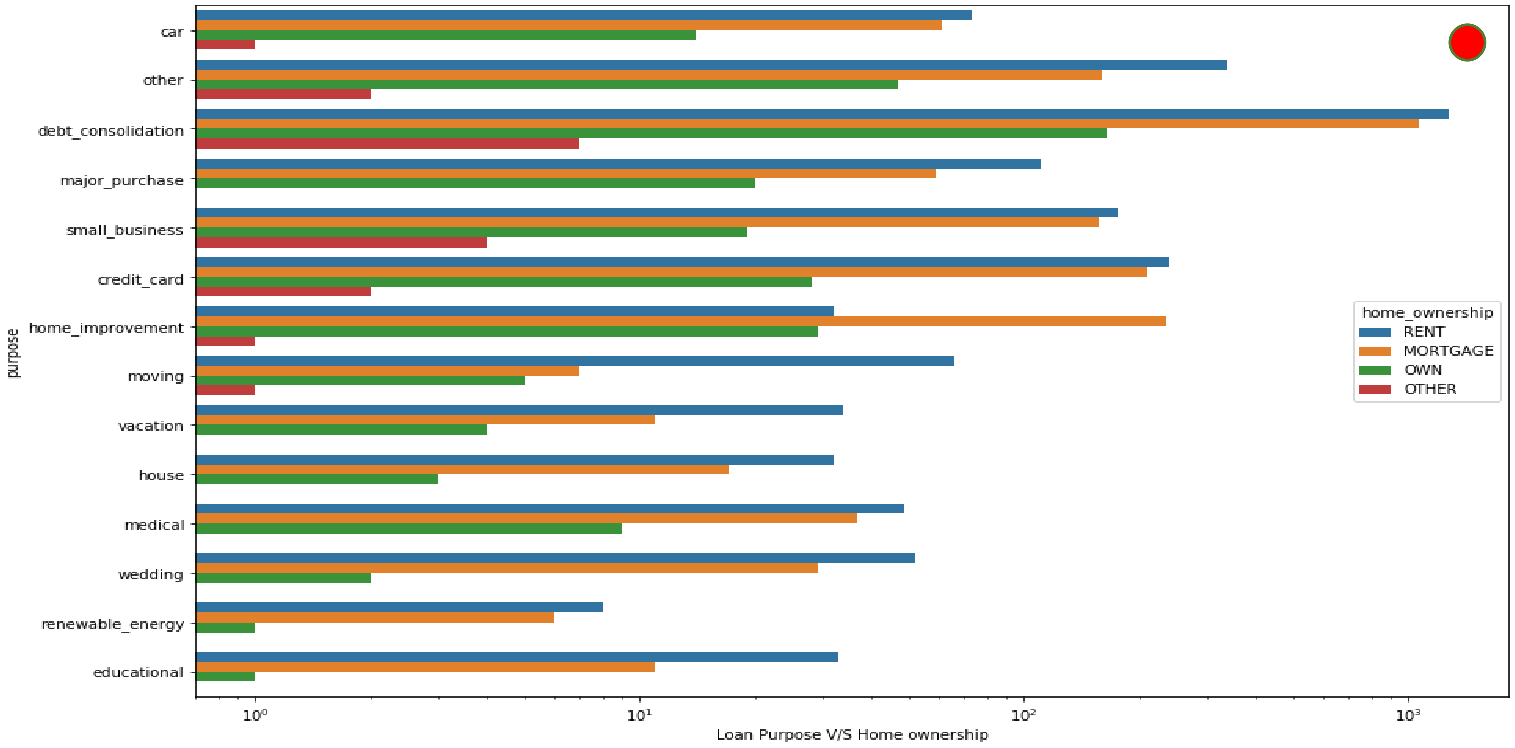
UpGrad

Ratio(LoanAmount/AnnualIncome) v/s Loan Amount



We can see that when the ratio between loan amount and annual income exceeds 20%, the numbers of charged off loans increases

Ratio(LoanAmount/AnnualIncome) v/s Loan Amount

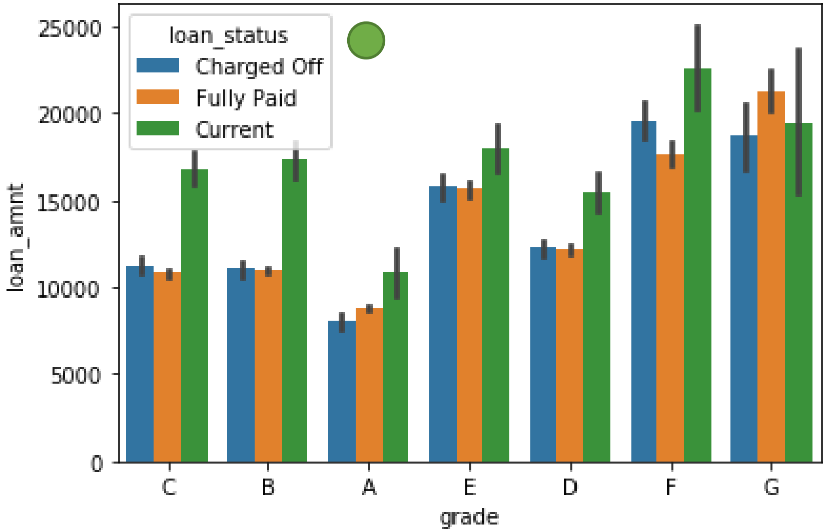


Those living on rent and mortgage have a high propensity to seek out loans. These could be potential target customers.

Likelihood of default for those who live on rent and mortgage is high among all the purposes. This strongly indicates a risky borrowing behaviour.

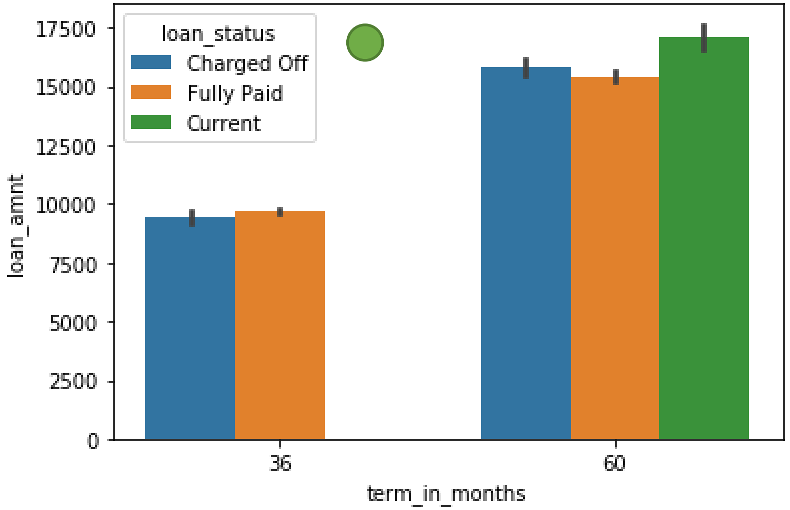
Peculiar observation: Those who live on mortgage and take loans for home improvement have a very high chance of their loans being charged off!

Grade v/s Loan Amount



We can see from here that lower grade loans are of a higher amount. These loans are also showing very high rates of being charged off. This is a very risky lending behaviour and LC should curb this practice in future.

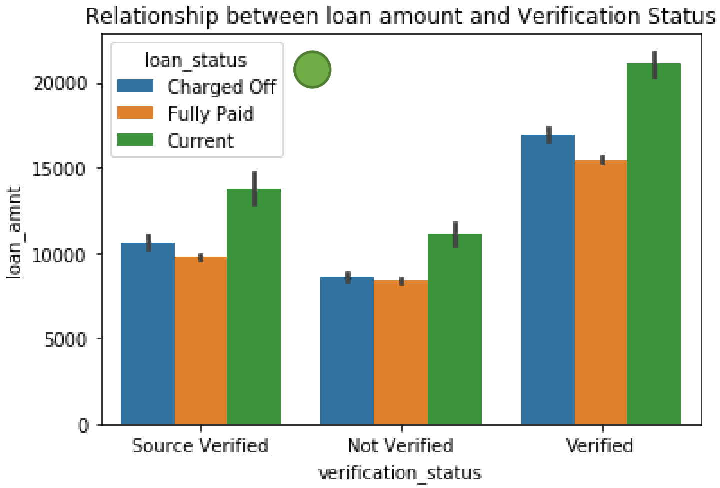
Term v/s Loan Amount



The 60 month tenure loans are more popular for higher loan amounts, but they also have a very high rate of loans being charged off. In fact for 60 month tenure, the charged off loans are slightly higher than the fully paid loans in terms of loan amounts.

We can say that for high valued loans, the loan term of 60 months is a very risky proposition.

Loan Amount v/s Verification Status

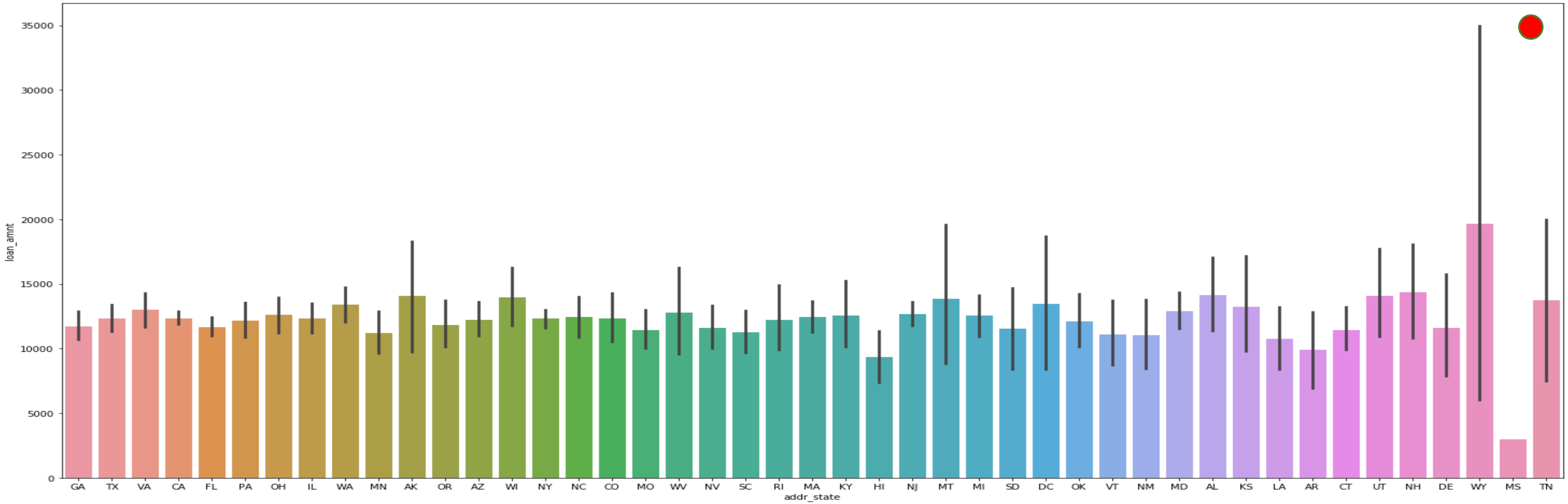


In terms of value of loan amount, the non-verified loans are the lowest. But among these not-verified loan category, the charged off loans value exceeds the fully paid loans value.

Though the higher loans are less in numbers, they have a high probability of being charged off.

More due diligence is displayed in terms of verification for higher valued loans

Loan Amount v/s Address State

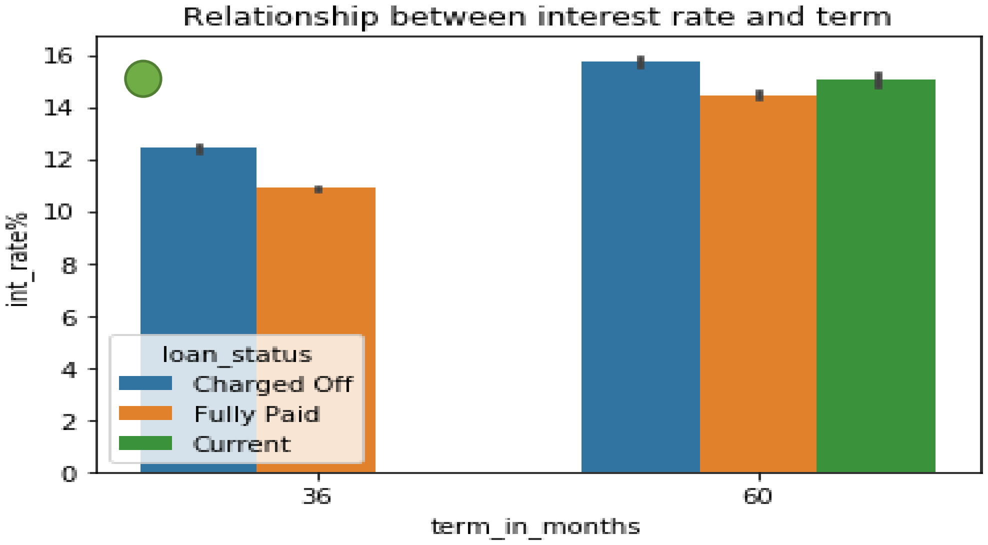


In terms of loan amount, the state of Wyoming has the highest average loan amount that was charged off. So this state clearly represents a risk category.

The top 5 risky states in terms of loan amounts being charged off are:

1. Wyoming 2. Tennessee 3. NH 4. Arkansas 5. Alaska

Interest Rate v/s Term



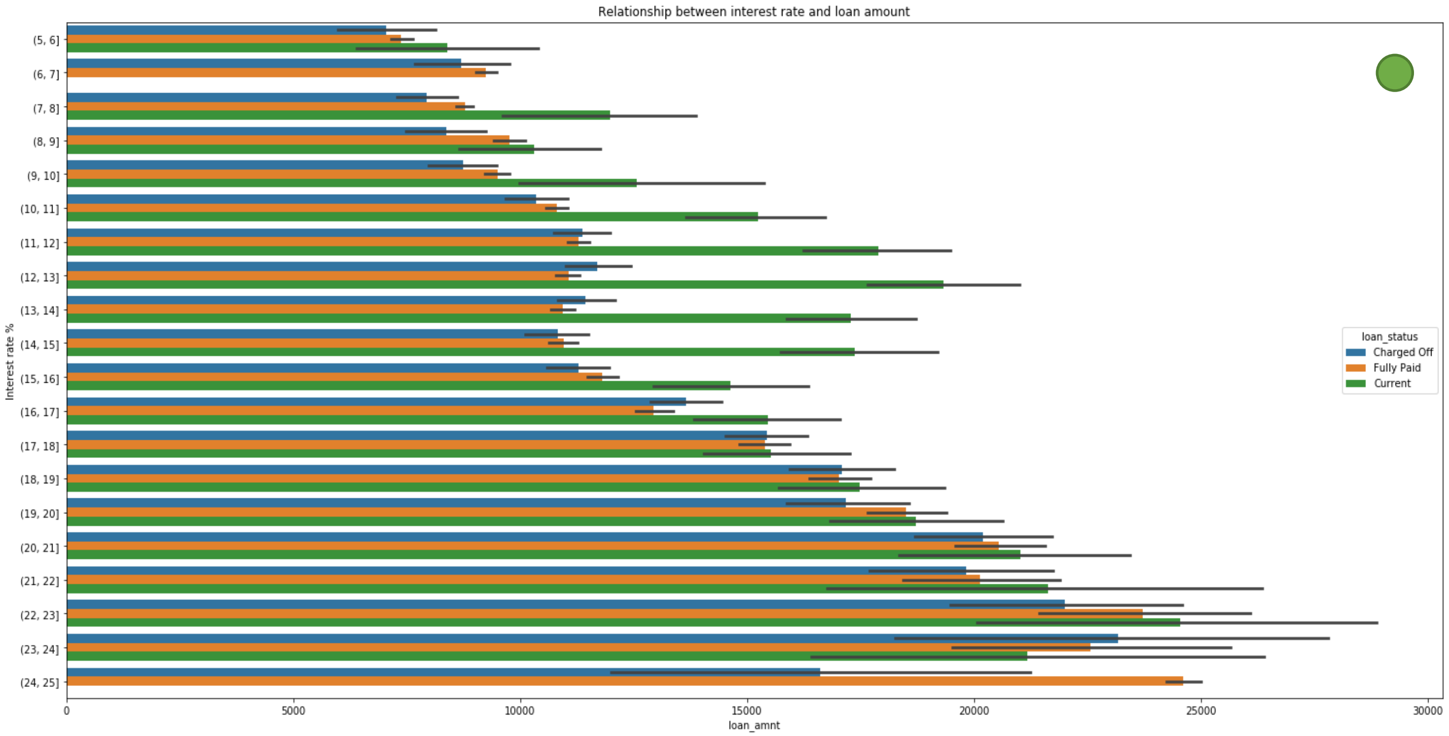
Higher rate of interests are being charged for longer term loans

For the 36 month tenured loans, the charged off loans start exceeding the fully paid loans in the interest rate region of about 10.5%-11%.

For the 60 month tenured loans, the charged off loans start exceeding the fully paid loans in the interest rate region of about 14%.

So, considering other things, the maximum interest rates to avoid charge offs could be below10% for 36 months loans and Below 14% for 60 month loans

Interest Rate v/s Loan Amount



Interest rates are rising with the increase in loan amounts.

In the range of 11%-14% and 16%-19%, we see the charged off loans exceeding the fully paid loans.

From 19% onwards we see the fully paid loans exceeding the charged off loans with the exception in the range 23%-24%.

We can see that the loans with values higher than 15000 are generally paid back more, even though they charge a higher rate of interest, with the exception in the range 23%-24%.

Conclusions

- Low Grade Loans have a tendency to default.
- Those living on rent and mortgage have a high propensity to seek out loans. These could be potential target customers.
- Likelihood of default for those who live on rent and mortgage is high among all the purposes. This strongly indicates a risky borrowing behaviour.
- In terms of loan amount, the state of Wyoming has the highest average loan amount that was charged off. So this state clearly represents a risk category
- The number of loans that are not verified exceed both the source verified and verified loan categories. And these not verified loans have a very high rate of being charged off. This is a matter of concern and the company should take due diligence before advancing loans.