Lending Club Case Study

Siddharth Singh <siddyy04@gmail.com> (Group Facilitator)

Vernica Ahuja <vernica.ahuja07@gmail.com>

Batch MLC45

Abstract

Lending Club is a marketplace which lends various types of Personal loans , Business loans and Financing of medical procedures. When company receives a loan application, the company must take decision of loan approval based on the applicant's profile.

As an employee of Lending Company our aim is to analyze the past data . The data contains all the information about Applicant like annual income home ownership and past loan whether they 'defaulted' or not.

The aim is to identify patterns which indicate if an applicant is likely to be default, which may be used for taking decisions such as denying loan, reducing the amount of loan, lending (risky applicants) at a higher interest rate etc.

Problem Solving Approach

1 **Data Understanding** — Read loan dataset, check sample data, size and datatype, then find out the Target variable

2 **Data Cleaning** — Cleaning the dataset by dealing with missing values, removing redundant columns and correcting data types.

3 **Univariate Analysis** — Analyze one variable at a time and plot with loan status to find out inferences driving the target variable.

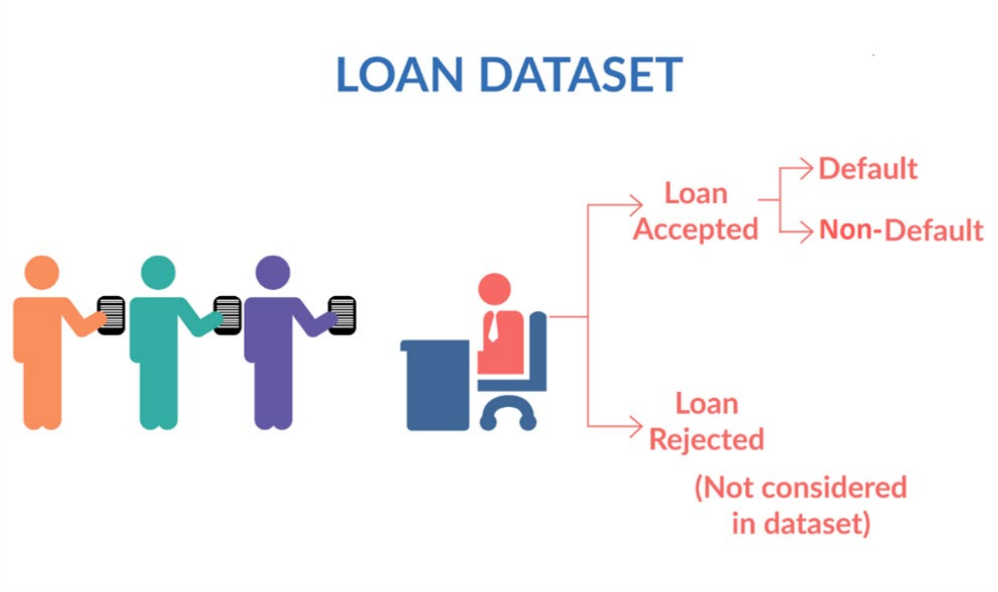
4 **Bivariate Analysis** — Use two variables to plot the charts, heat map, scatter plots and ascertain the impact of combination of variables on loan status.

5 **Recommendations & Insights** — Conclude high impacting variables that result in loan default.

1 Data Understanding

Let's take a thorough look at the dataset

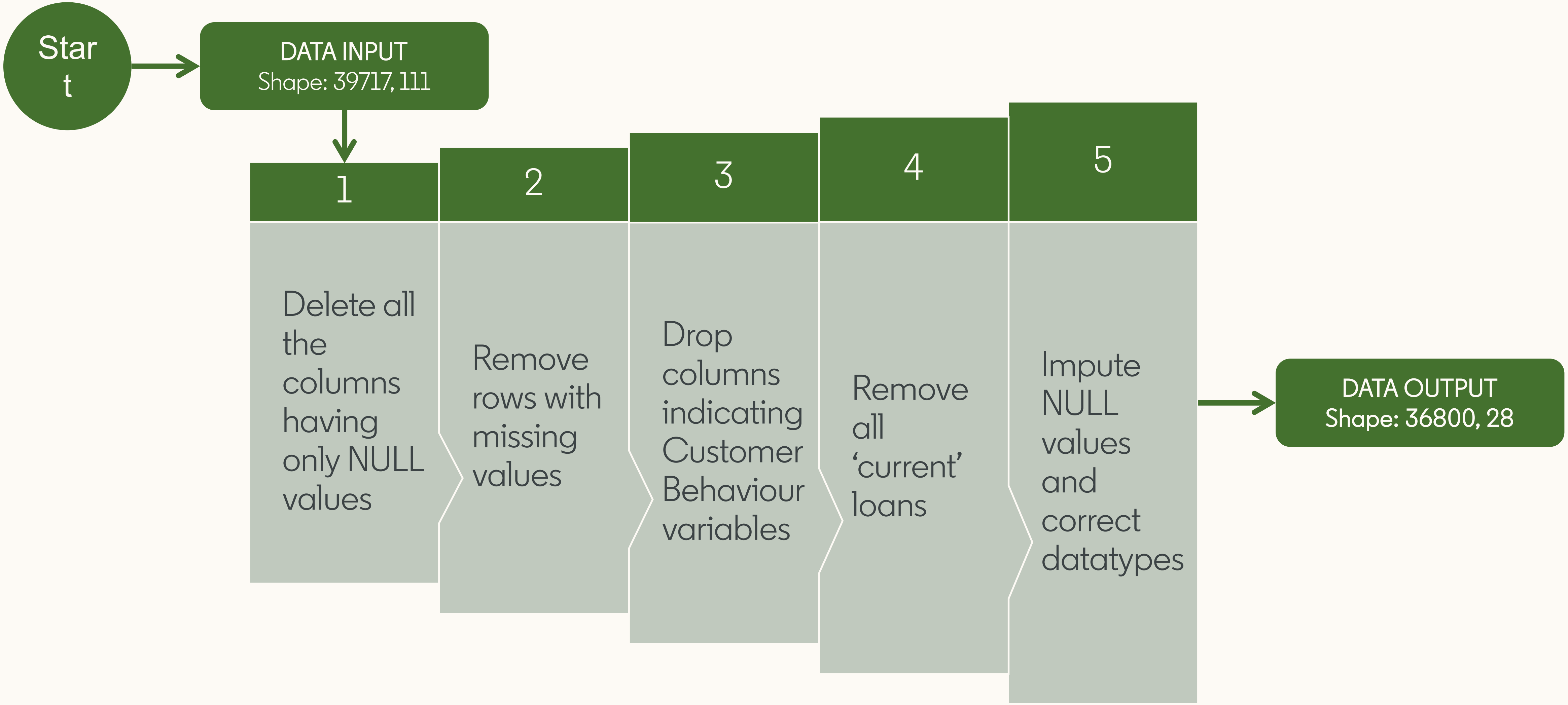
- Loan data is imported from "loan.csv" file. The raw data contains 39717 rows and 111 columns.
- Checked the datatype of various columns.
- There are different types of variables present in the dataset like:
  - ✓ Applicant Demographics (Employment length, Employment title, Annual income, etc.)
  - ✓ Loan information & its characteristics (Loan amount, Funded amount, Loan grade, Loan status, Interest rate, etc.)
  - ✓ Customer behavior (application type, loan purpose, delinquency 2-year, revolving balance, etc.)



- Now, applicant behavior variables are not available at the time of loan application , thus it can not be used for prediction of loan approval.
- We also checked the columns which contains numerical data and extract it for further data insights.
- Understanding of date attributes and convert them to the correct date format.
- We identified target variable columns , in this case "loan_status" is target column.

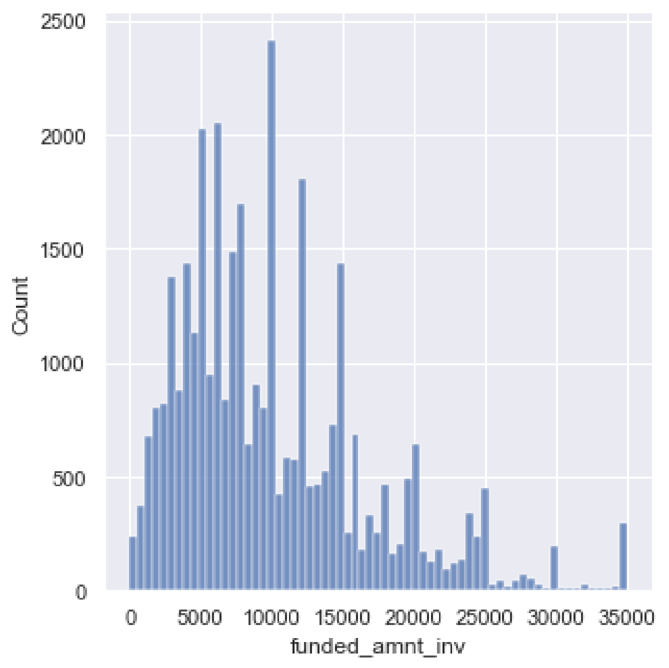
2 Data Cleaning

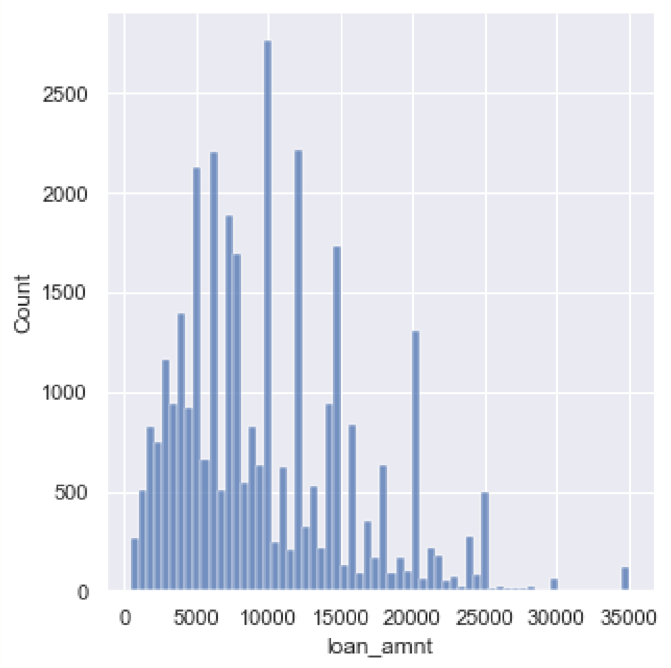
Here are the steps we took to clean data

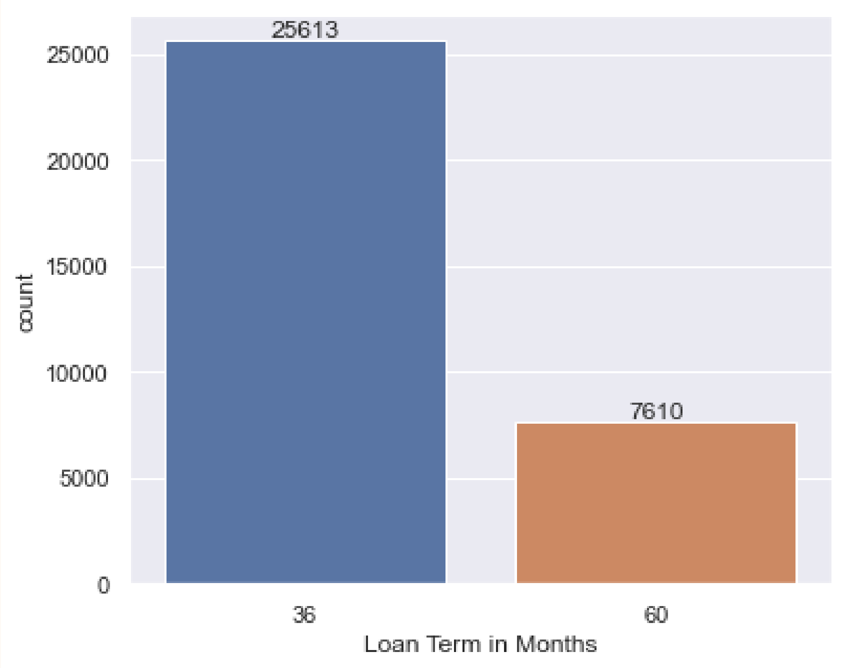


3 Univariate and Segmented Univariate Analysis

Comparing amount value of funded vs loan; quick look at loan term split



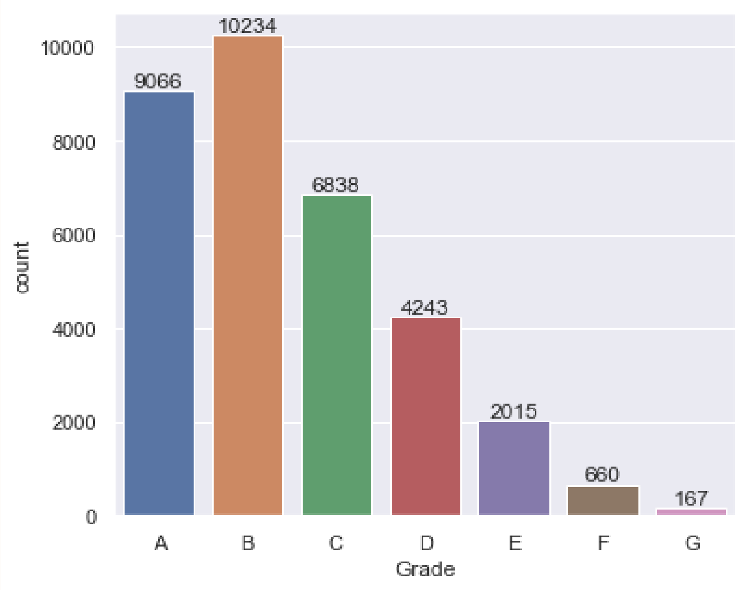


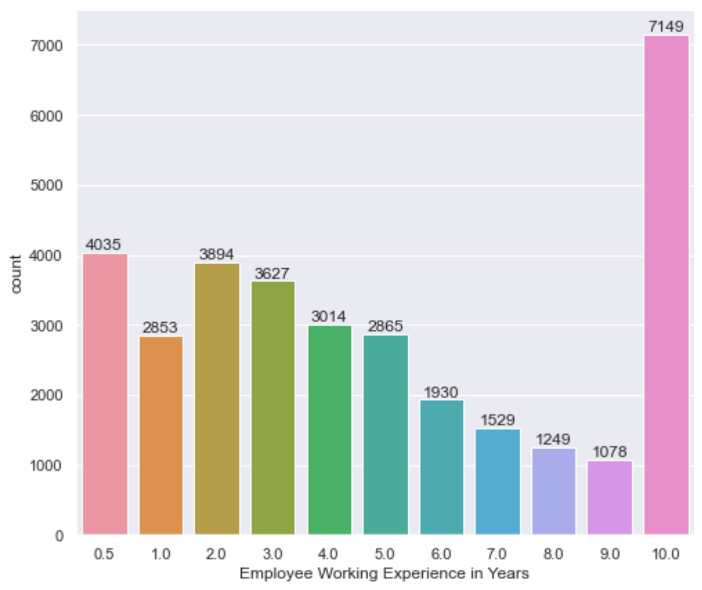


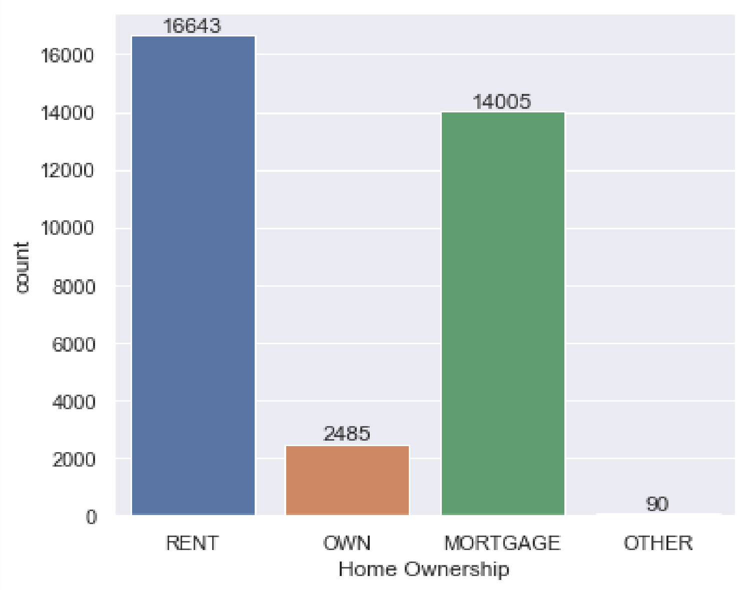
Like loan amount, number of loans sanctioned were near 5000, while the loan sanctioned amount of 10000 is almost equal to 5000.

Most of the loans sanctioned were for the term of 36 months (about 3 years)

Loan grades, borrowers' work experience and home ownership





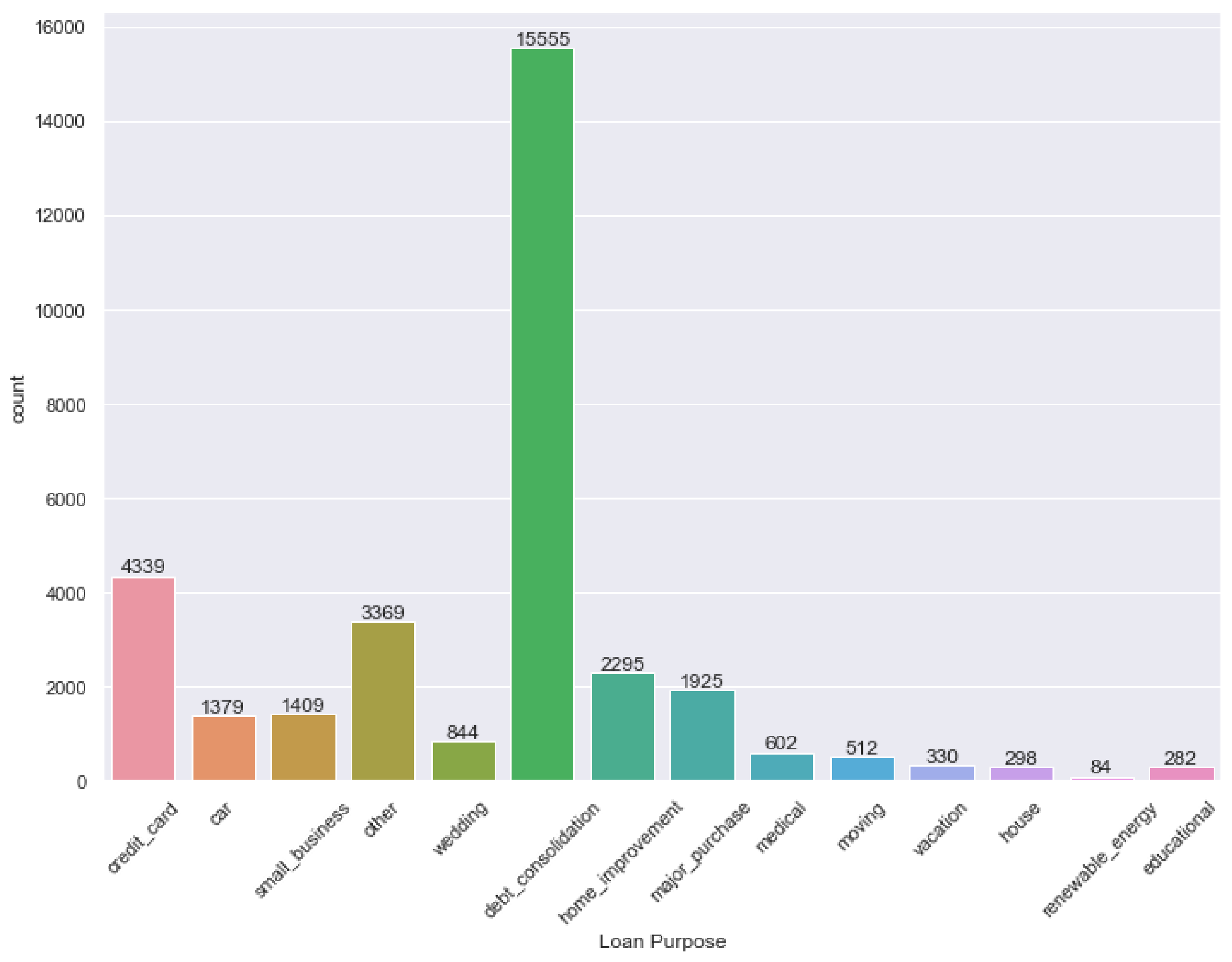


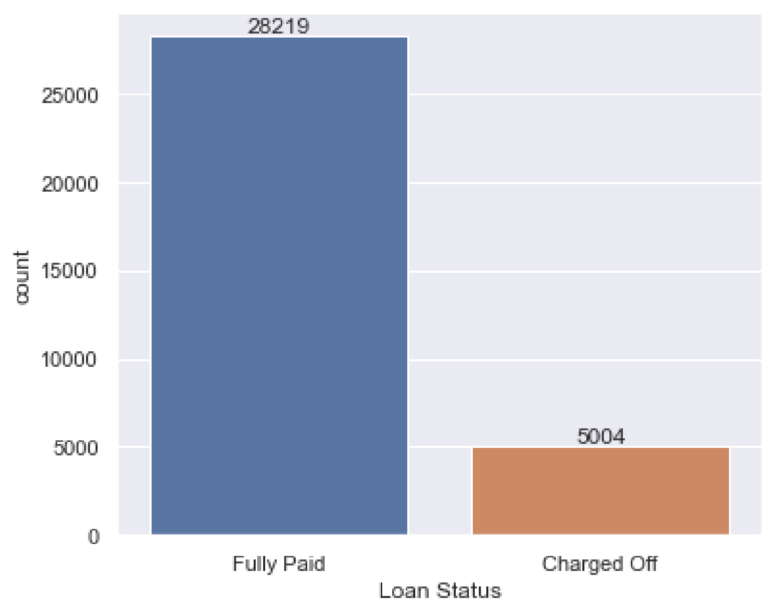
Most of the loans are of Grade A and B, meaning that most of the loans are of high grade

Majority of employees applying for loans have 10 or more years of experience

Most of the borrowers live on rent, followed by living on mortgage

Purpose of loan and Loan status





Most loans are fully paid.

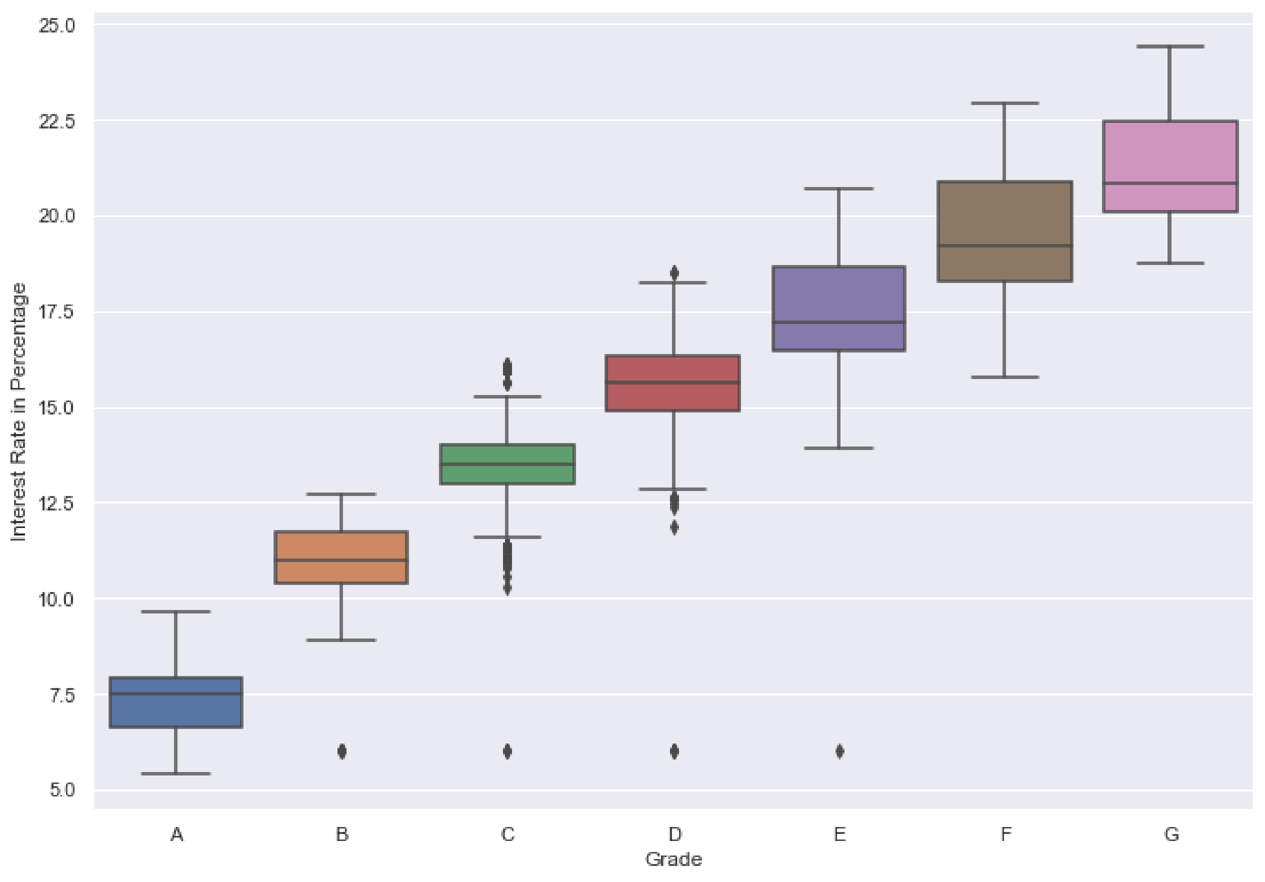
Most loans are sanctioned for the purpose of debt consolidation, followed by credit card. The least is for renewable energy.

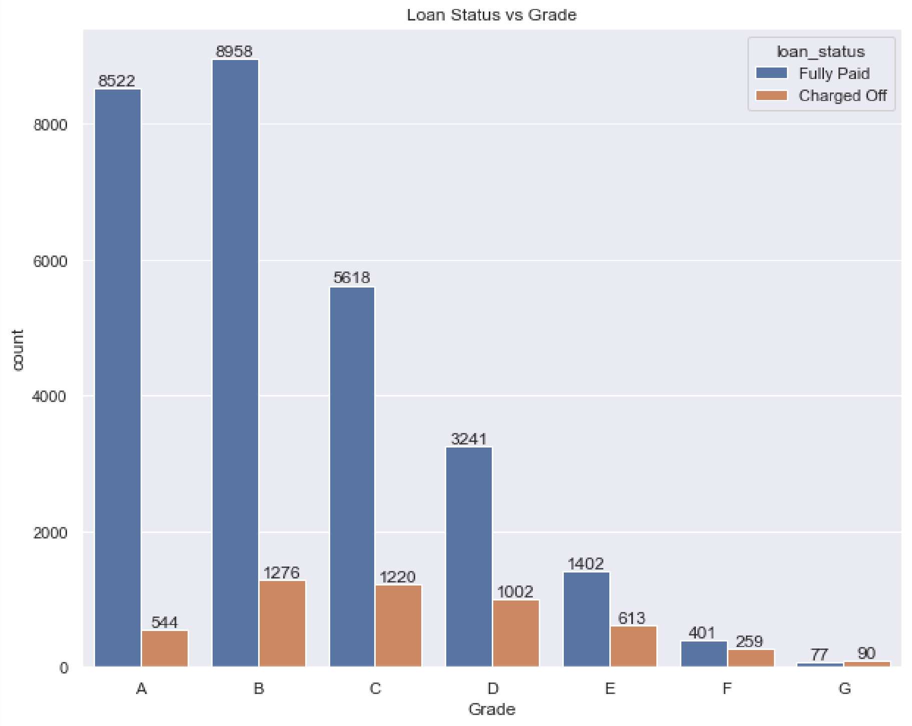
4

Bivariate and Multivariate Analysis

Comparing Loan Grade with Interest rate % and Loan status

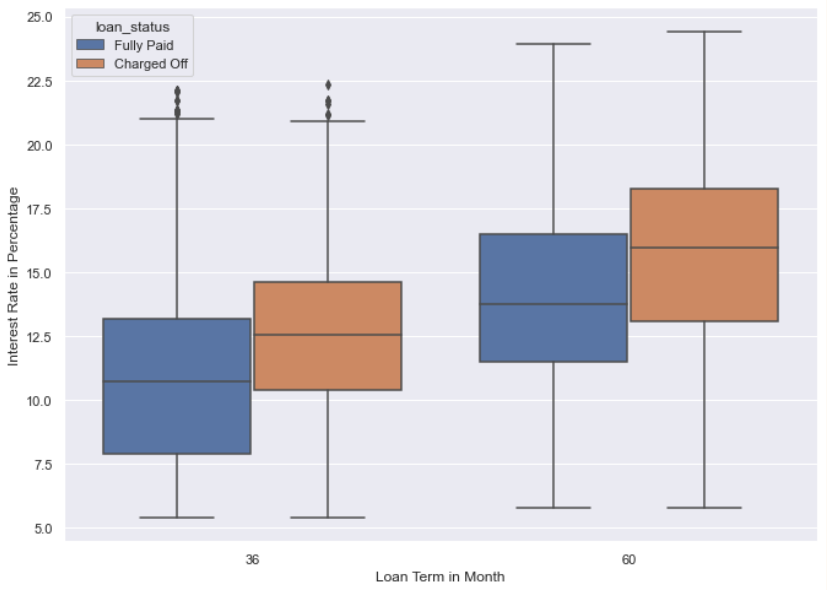
Here, we have plotted bar graph, heat map and subgraph with combination of the important variables identified during univariate analysis, where we saw that the chance of default is changing heavily with the purpose.
In the bivariate analysis, we can see the patterns of defaulting by studying important derived variables.

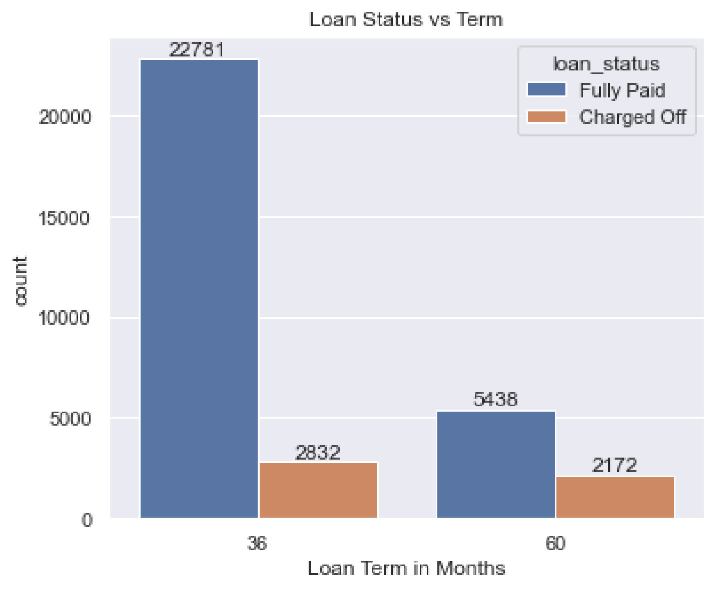




Higher grade loans attract higher interest rate, which have higher tendency of defaulting.

Analysing Loan term with Loan status and Interest rate %



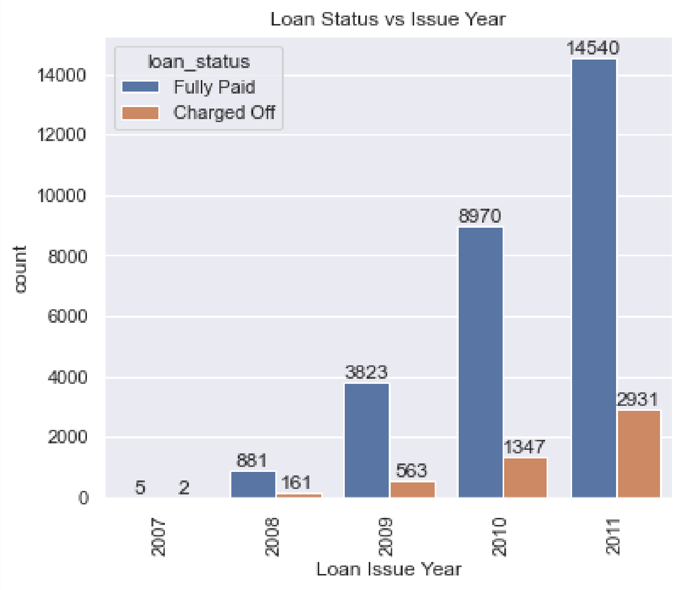


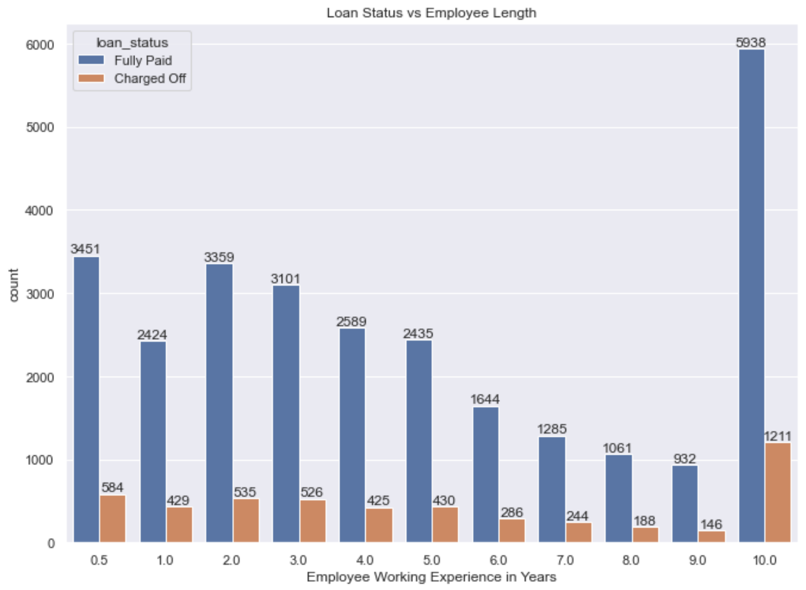
In both cases, the chances of loan default is high if the interest rate is higher

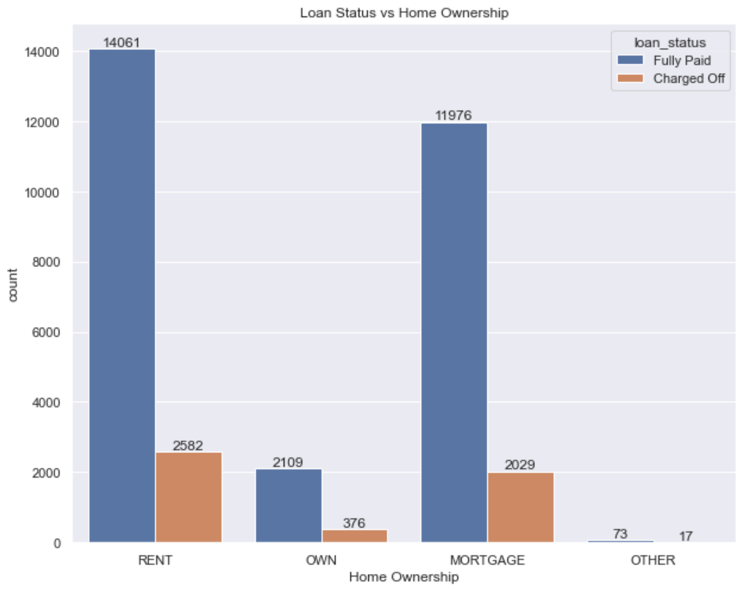
Default rate is 14% on average. A 60-month term attracts 34% default rate whereas 30-month attracts only 13%

Loan amounts are higher for 60-month long term loans

Loan status to Issue date, employment length and home ownership





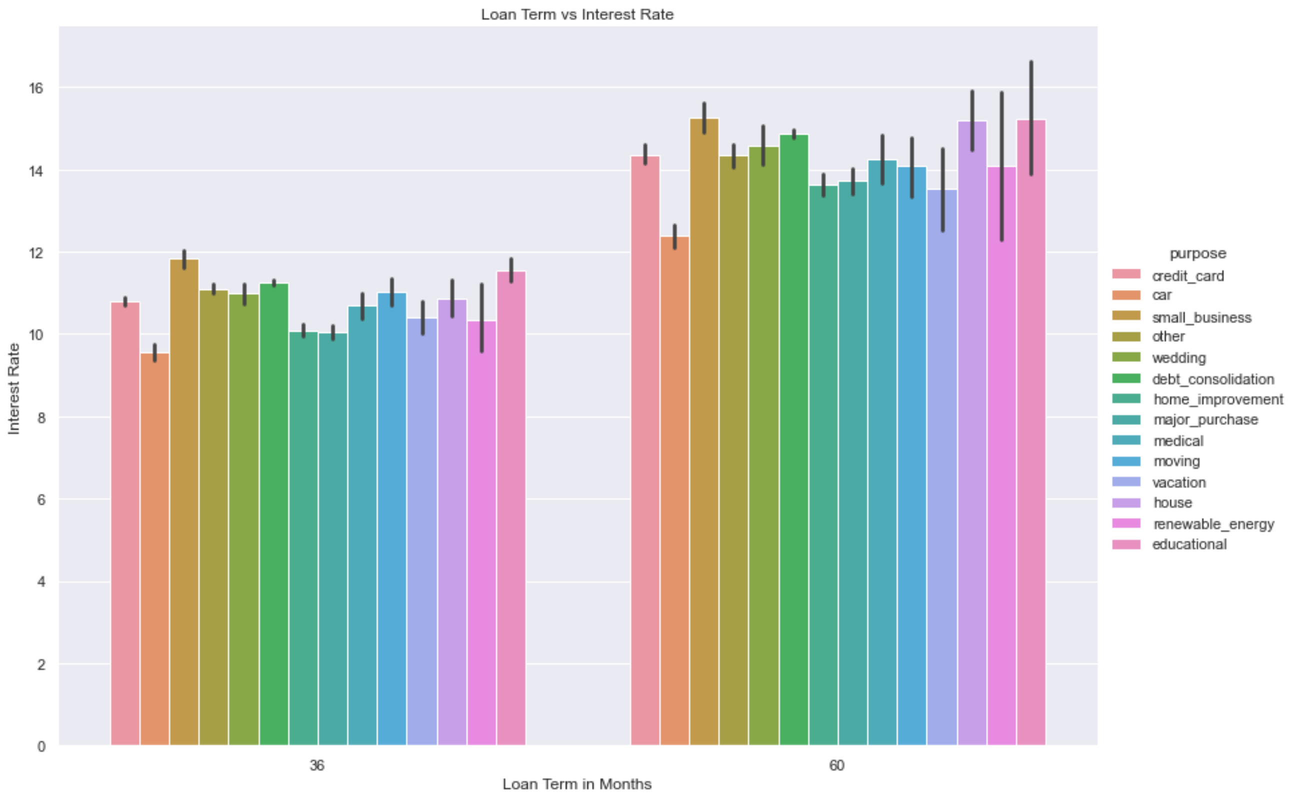


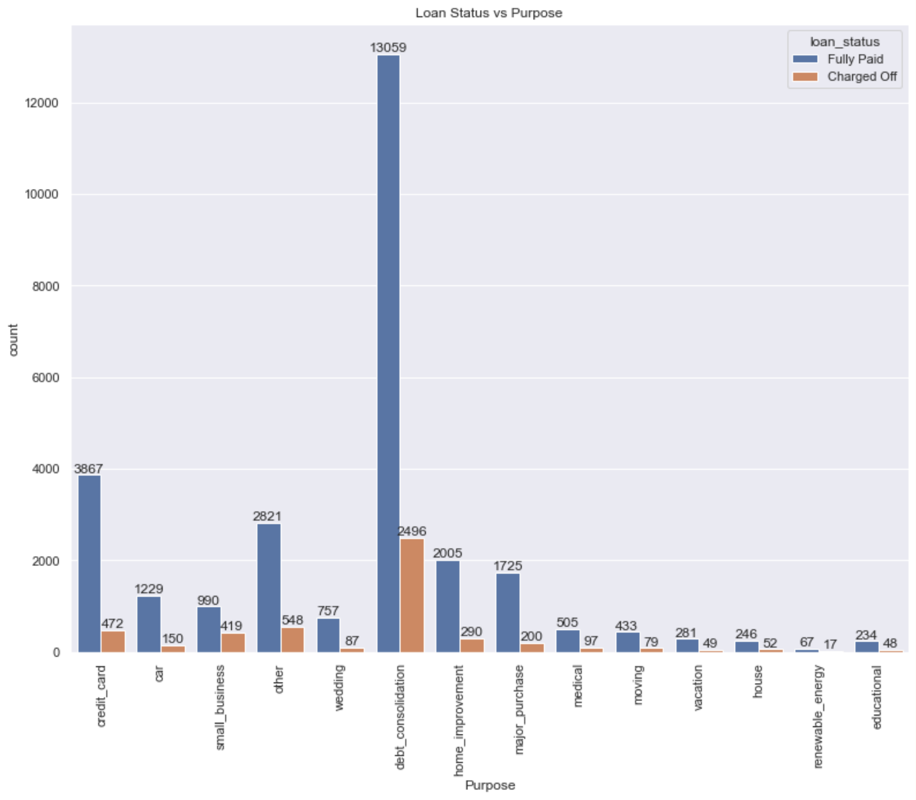
More loans are being granted every year, with default rate being the highest in 2011

While most employees with 10 or more years of experience apply for loans, they also have a higher tendency of defaulting in comparison

Applicants living on rent or mortgage have similar tendency of defaulting

Comparing Loan term with Interest Rate %, Purpose and Loan Status

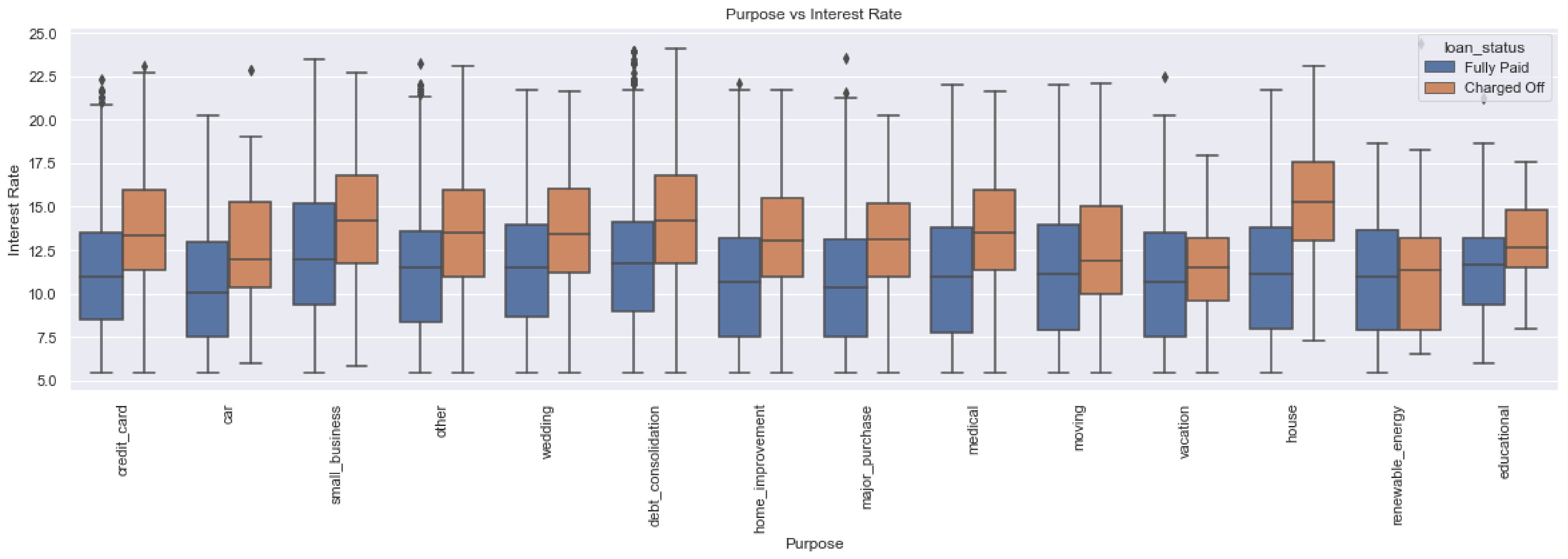




Chance of defaulting are higher for purpose of small business with term being 36 months (about 3 years). Similar chances are observed for other purposes like debt consolidation, education, house and vacation for a 60-month term

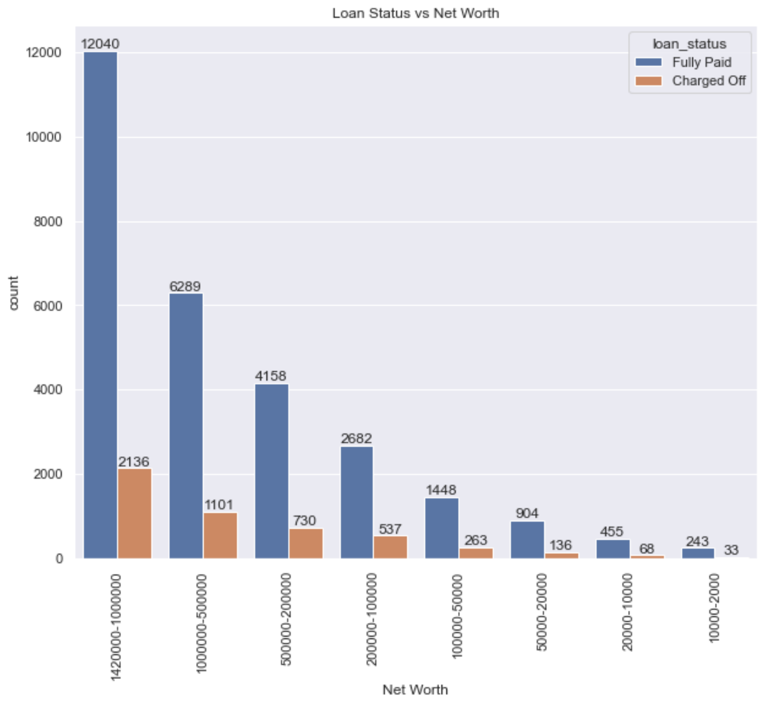
The ratio of defaulters is the highest for small scale business borrowers followed by renewable energy

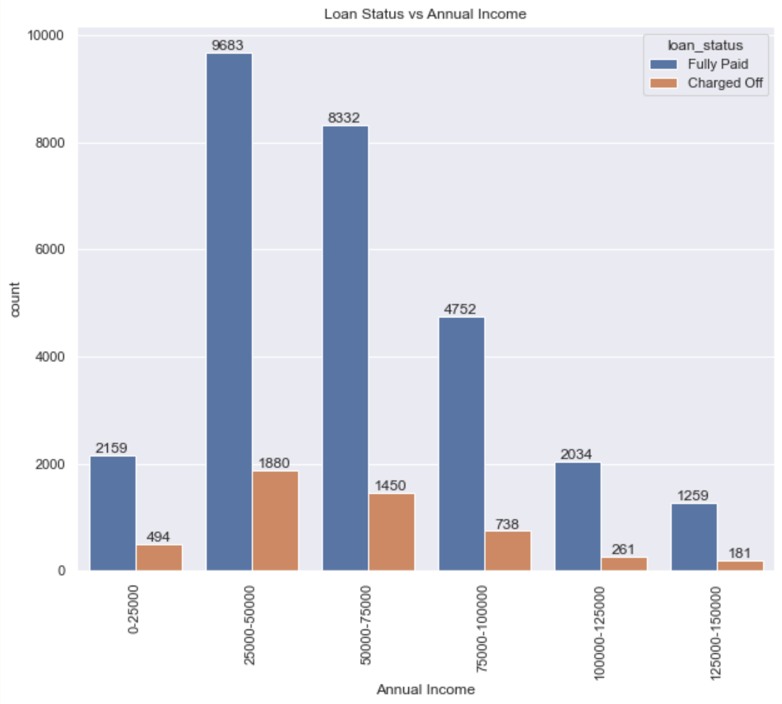
Comparing Interest Rate % with Purpose and Loan Status

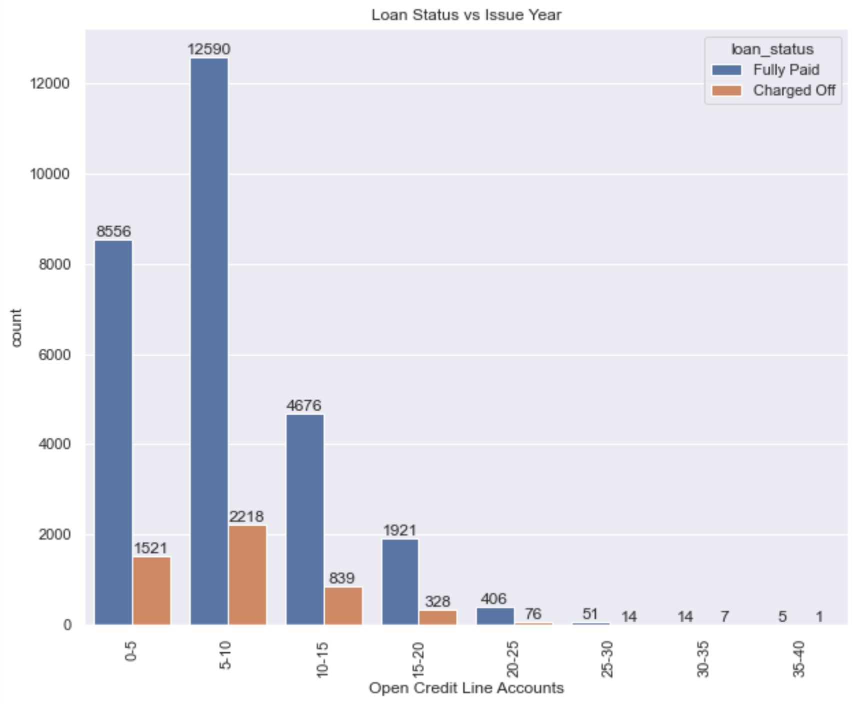


In general, loan applicants with purpose of small business or renewable energy have higher tendency to default. But if we consider interest rate variability, then we find other purposes like house and debt consolidation also showing higher tendency to default.

Analysing Loan Status with Net Worth, Annual Income and Issue Date







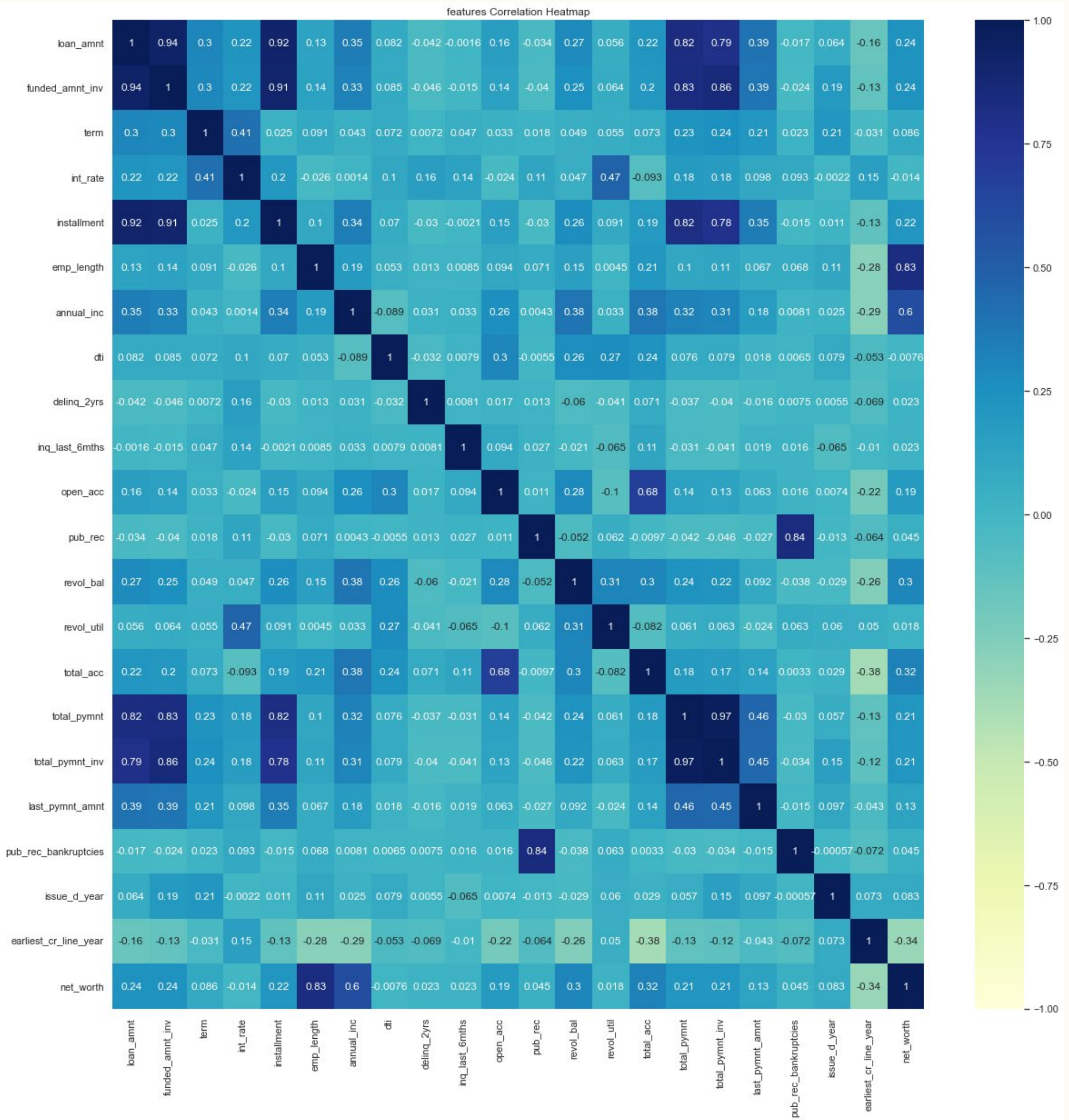
Borrowers with net worth between 100000 and 200000 tend to default more, followed by highest net worth borrowers

Chances of defaulting is inversely proportional to annual income, meaning higher the income, lesser is the chance of defaulting

Borrowers having more credit line open accounts tend to default more.

Correlation Heatmap between all parameters

- Employee working length is highly correlated with net worth.
- Public derogatory records are highly correlated with public bankruptcies records.
- Installments are highly correlated with loan amount.
- Total payment to investors is highly correlated with funded amount y investors.



Bivariate and Multivariate Analysis

Here are the inferences based on the analysis of 5 main drivers of the dataset -

Interest Rate

Higher interest rate loans are mostly deterrent. Therefore, denying such risky individuals might be good approach instead of approving them.

of Open Credit Lines

Borrowers having more open credit lines tends to be more defaulters. So, it is justifiable if such individual's loans are getting handled carefully.

Grade

It is clear from the analysis that high grade loans are having less interest rate and less defaulters. So, it might be good opportunity for the investors to invest on high grades also, lower grades could be avoided.

Loan Amount to Annual Income Ratio

Income plays the key role in repaying the loan. If the income of the individual is more then the chances of defaulters are less. Hence, approving their loans will be in the benefit of the industry.

Purpose of Loan

Loans with purpose small businesses, debt consolidation, renewal energy, are more of defaulters. Keeping that in mind, rejecting such loans might be a good idea in the interest of industry.

5

Recommendations & Insights

Recommendations & Insights

#	Insights
1	Though a greater number of loans are applied for 36 months (about 3 years) and have more defaulters, the ratio of charged off loan is more for 60 months (about 5 years).
2	The chance of default is monotonically increasing with the interest rate i.e., the high the interest rate high the default rate chances. Interest rate is high for lower grades. So, there is possibility of more charged off for them.
3	The ratio of defaulters is maximum for small scale business borrowers; however, the numbers are high for debt consolidation
4	Defaulter chance is inversely proportional to the annual income. High the income less chance of default.
5	Default chances is increasing with the increase in the debt-to-income ratio.
6	Default chances is directly proportional to the loan amount. Also, chances of default is more when the home ownership is 'other'.
7	From analysis it's confirm that chances of default is high when purpose is 'small business' and the term is 36 months (about 3 years). Also, for 60-month term the default chances is high when purposes are debt consolidation, educational, house, other, small business and the vacation purposes.
8	Trend shows that default chances are more when the home ownership status is 'other', and the term is 60-month period. Even when the term is 36-month period, the default chances are the high if the home ownership status is other. But there is only one entry (term = 60-month period and home ownership = other), and very less data points in the 'other' home ownership. So, it is better to avoid the loans when the home ownership is others.
9	In general, educational, house, other and the small business loans are having very high interest rate, debt consolidation with medium high interest rate and renewal energy, house with high interest rate are showing more chances of default .
10	The high-income group for all purpose is relatively safe to lend money. The small business is showing high default chances in all income group except very high-income group. Loan with purpose renewal energy is showing high chance of default with very low income, low income, and high-income bracket.
11	Loan with purpose car, credit card, home improvement, medical and wedding is showing less chances of default among various category and one of the safest purpose to lend. On the other hand, loan with purpose small business, debt consolidation, renewal energy, educational, house and other is showing high chances of default.