Lending Club Case Study

Submission

Name-
Kanika Raheja
Ankush Das

Problem Statement

A consumer finance company which specialises in lending various types of loans to urban customers wants to mitigate the risk of loan defaulters. So the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

To derive the impacting factor we are using the historical data from Lending Club.

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Analysis Approach

UpGrad

Data Understanding

- Retrieving info about the shape of data set
- Segregate categorical and continuous dimensions
- Find percentage of null in each column

Data Cleaning

- Remove behavioral columns and those with high NULL values
- Remove outlier values
- Cleaning the variables of special characters
- Impute missing values

Data Preparation

- Creating derived variables issue_year & issue_mnth
- Creating categorical variables from continuous variables(loan amount, interest rate, DTI, etc.) by binning them

EDA

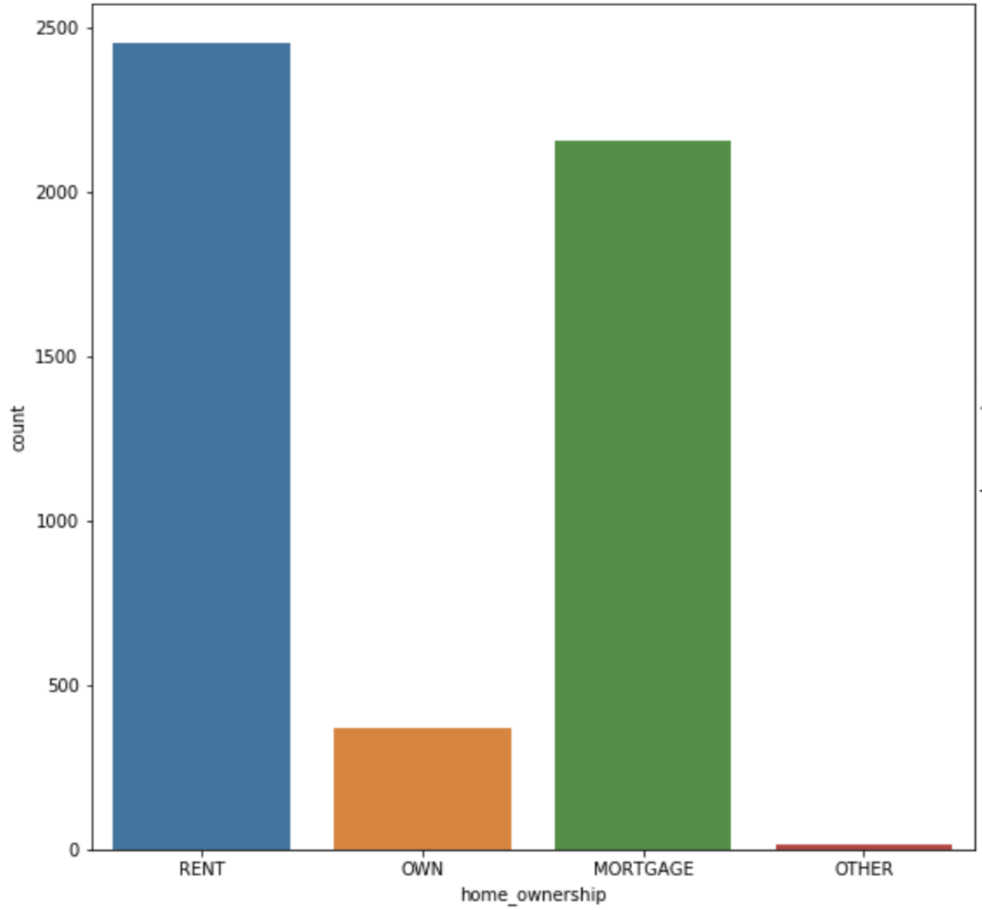
- Univariate Analysis
- Bivariate Analysis

Recommendation

- Outcome of the Analysis and the metrices

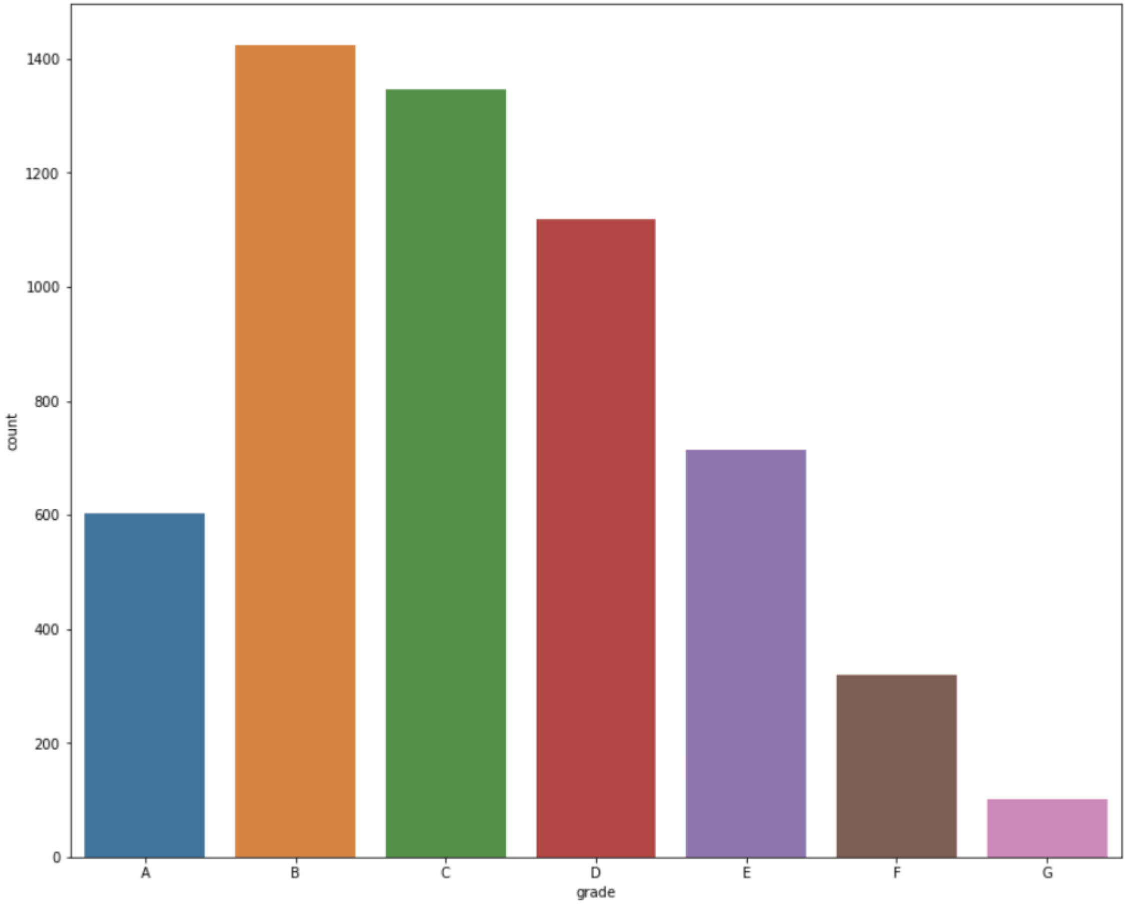
Univariate Analysis – For Charged-Off Loans

Home ownership Status (home_ownership)



For home_ownership, customers with status 'RENT' default most followed by 'MORTGAGE'

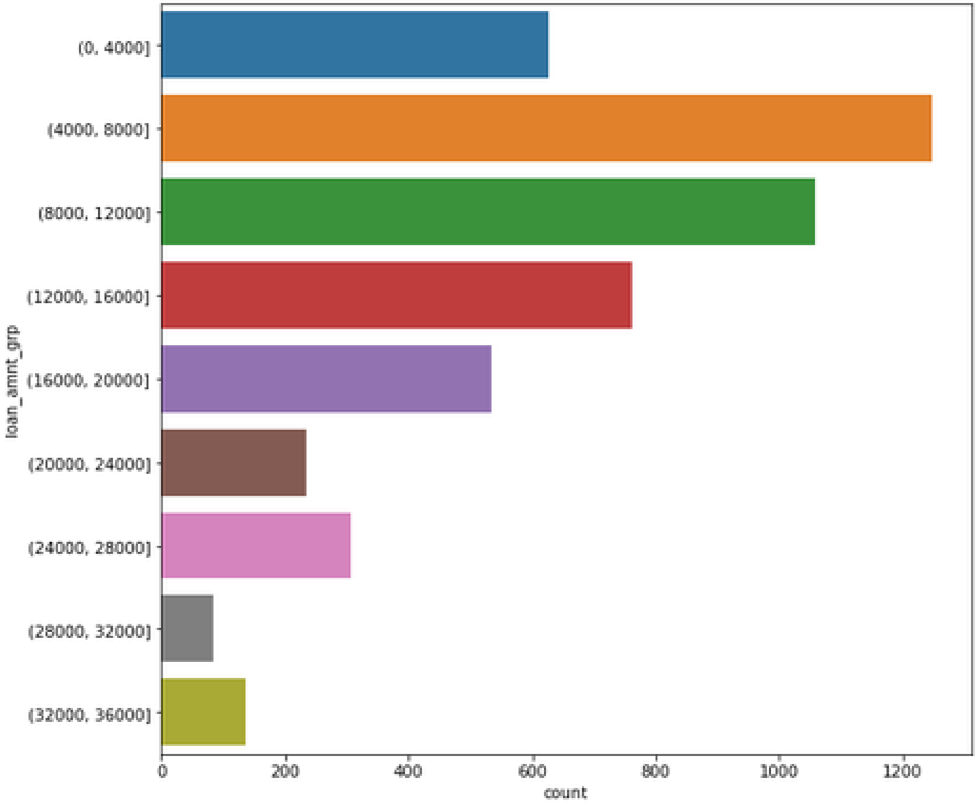
Loan grade (grade)



Approximately 50% of charged off loans belong to loan grade B & C.

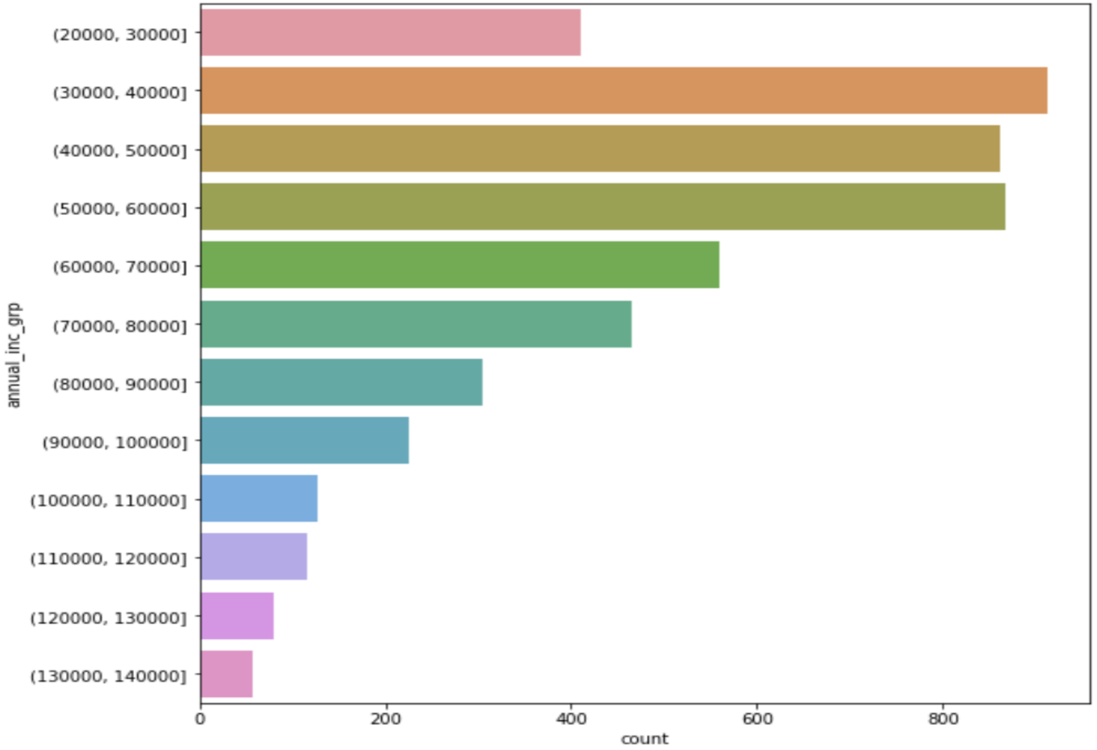
Univariate Analysis – For Charged-Off Loans

Loan amount applied (loan_amnt)



Contrary to common idea, most of the charged off loans values in lower range of loan amount (4000-8000 USD)

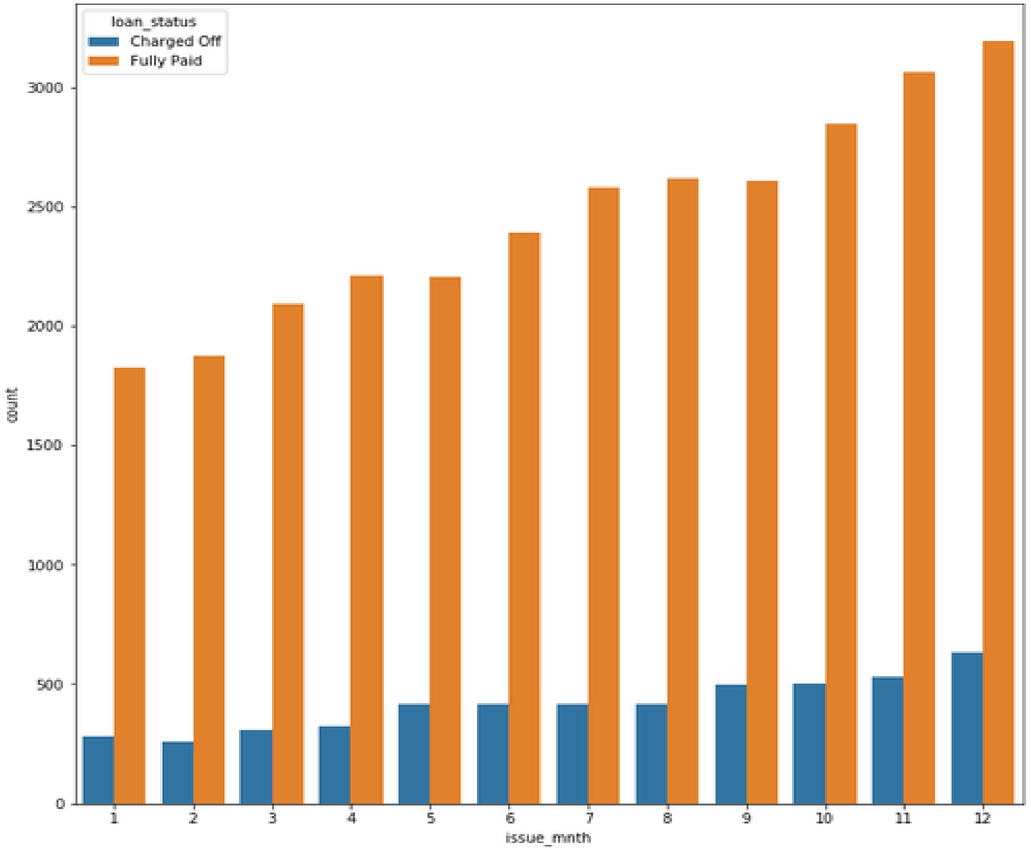
Annual Income (annual_inc)



It is evident from the above graph that customers having annual income of 30K to 60K have the maximum chance of being a defaulter.

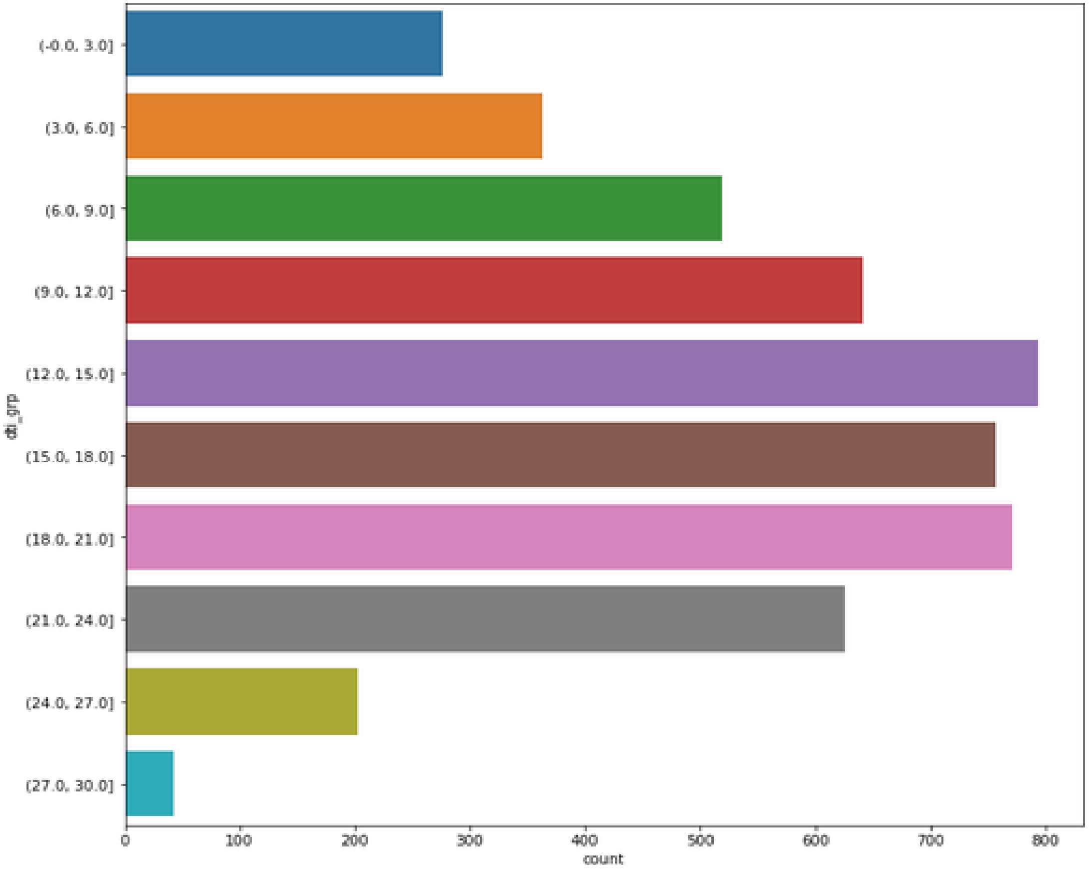
Univariate Analysis – For Charged-Off Loans

Issuing month for Loan



The above graph of historic data clearly shows that maximum number of loans are issued in the month of December

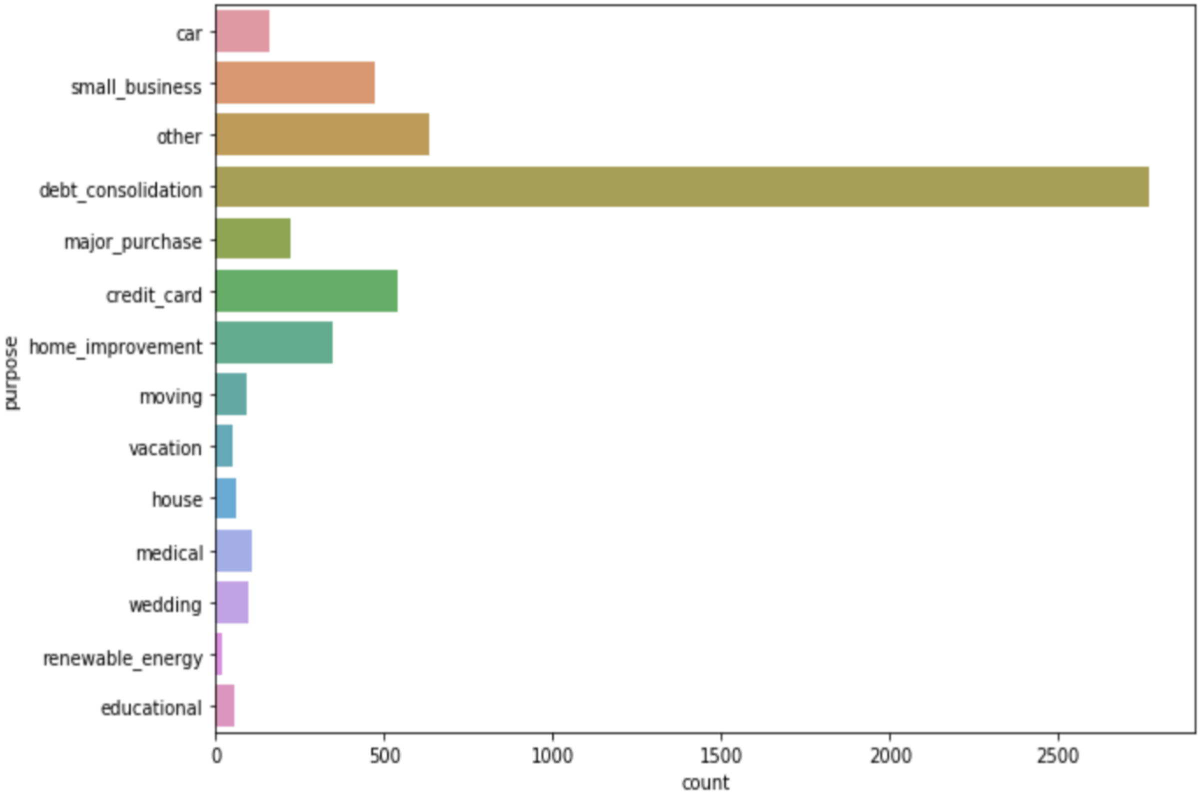
Debt To Income Ratio (DTI)



The DTI index from the above graph shows the DTI in the range of 12 to 15 & 15 to 18 have maximum number of defaults compared to others.

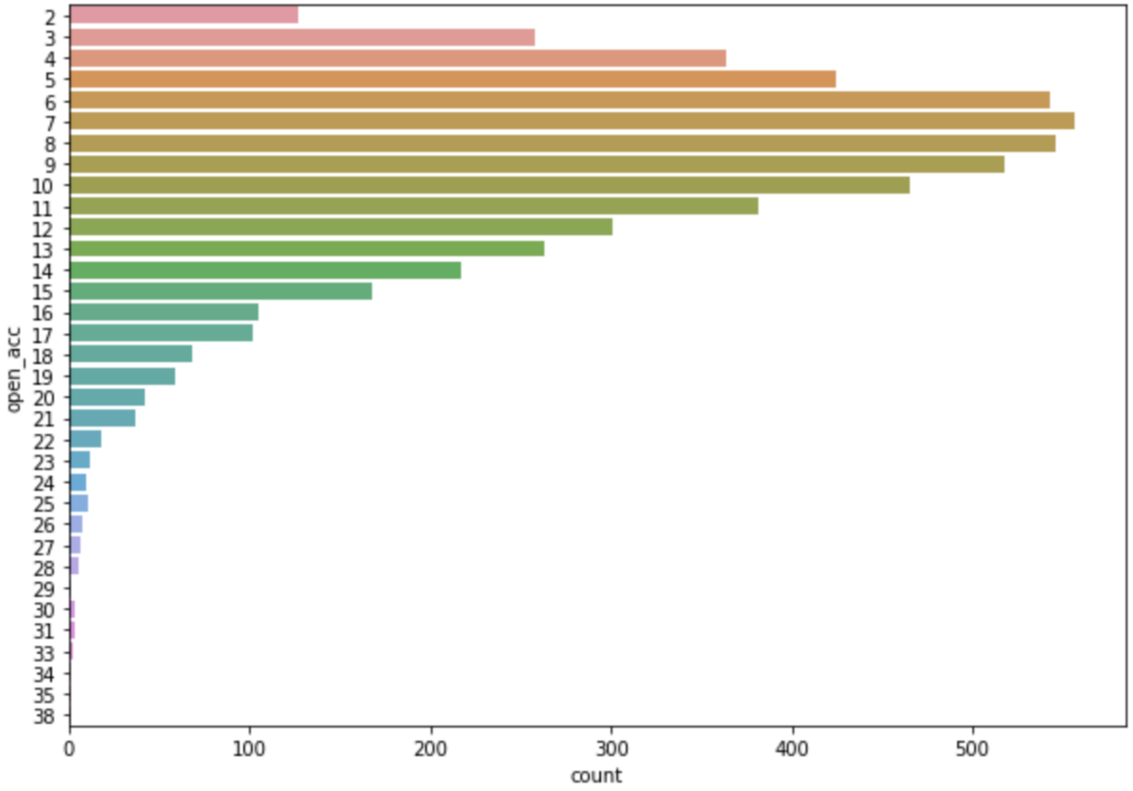
Univariate Analysis – For Charged-Off Loans

Purpose of the loan:



Looking at the number of charged off loans by their purpose, its evident that maximum defaulted loans were taken for debt consolidation.

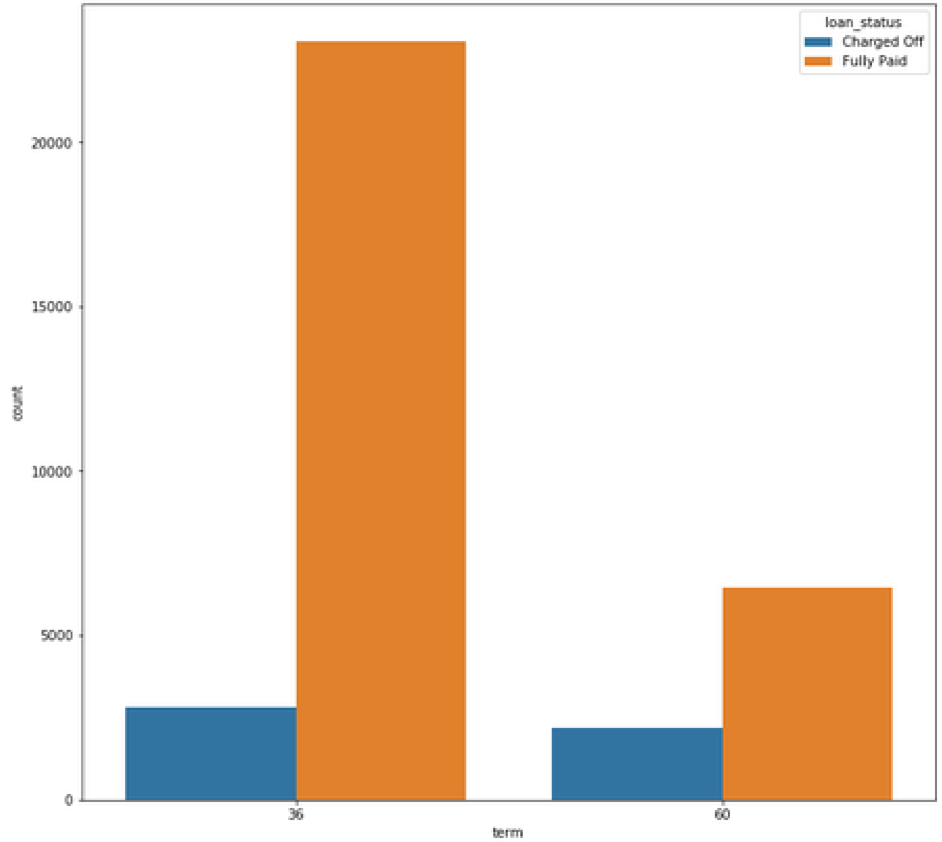
Number of open accounts:



The above graph shows a spike at numbers from 6 to 10. So borrowers who have close to these many must be under extra scrutiny.

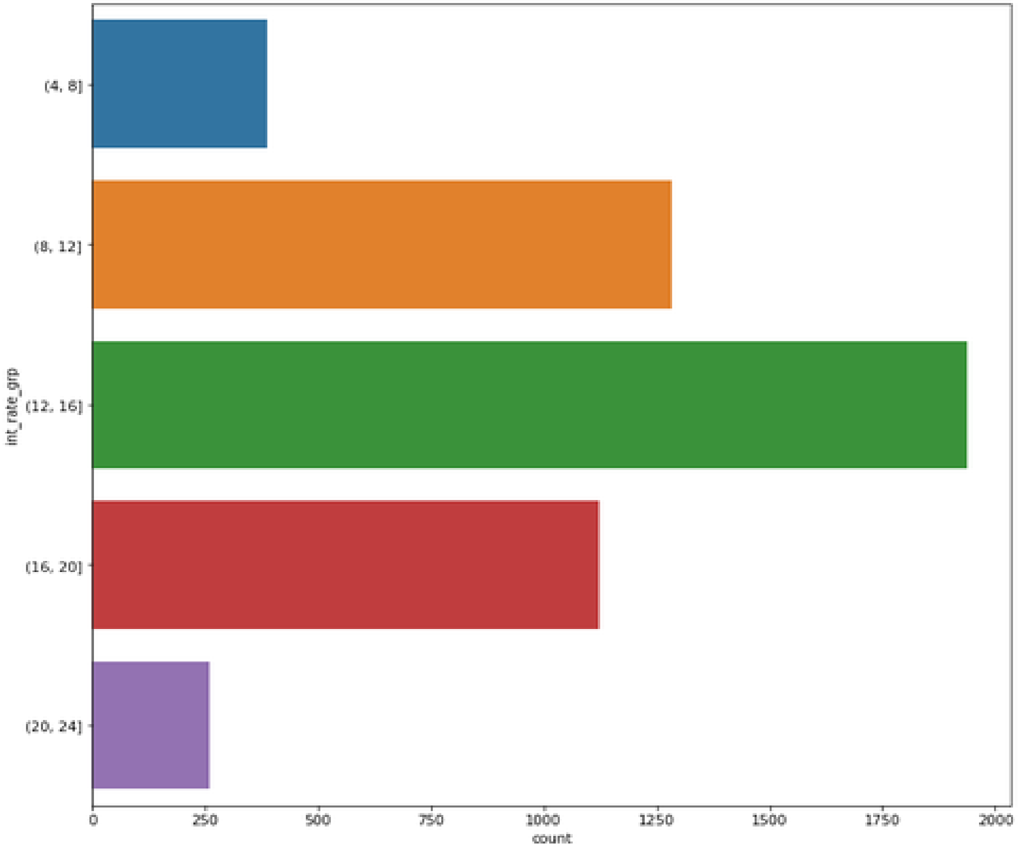
Univariate Analysis – For Charged-Off Loans

Total term of Loan:



The number of customers defaulting with 36 months term loan is more than longer term loans.

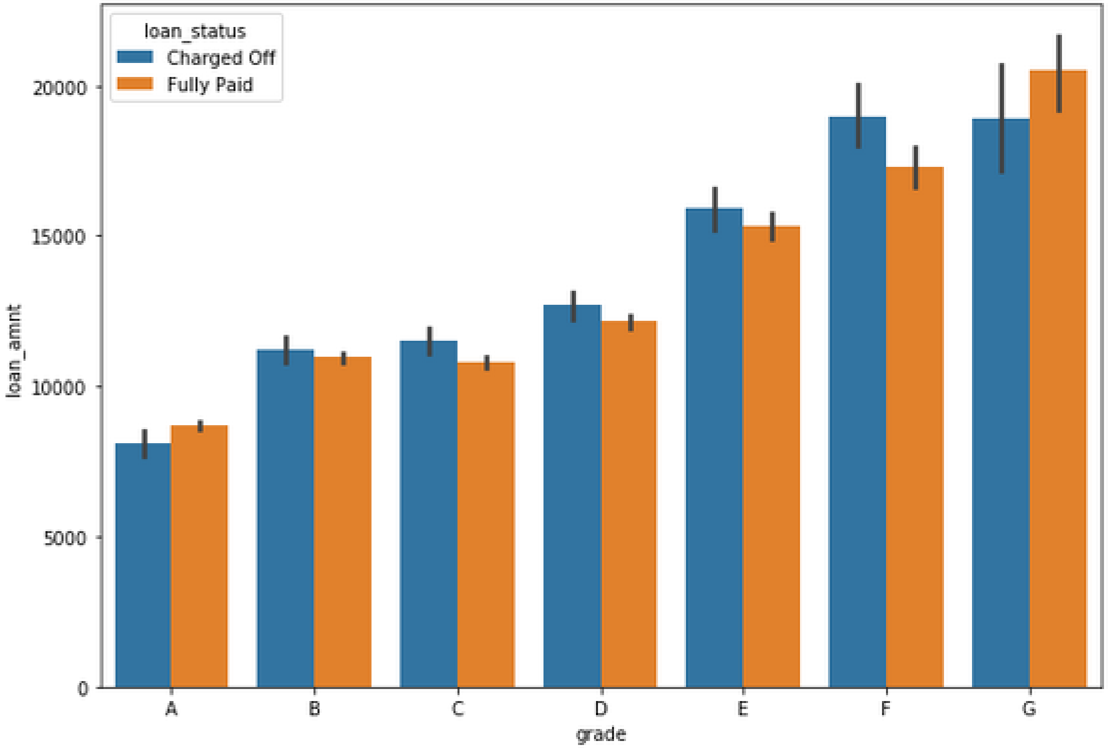
Interest rate at which the loan was given:



After distributing the interest rate in bins its observed that the charged off count is maximum for interest rate in the range of 12% to 16%.

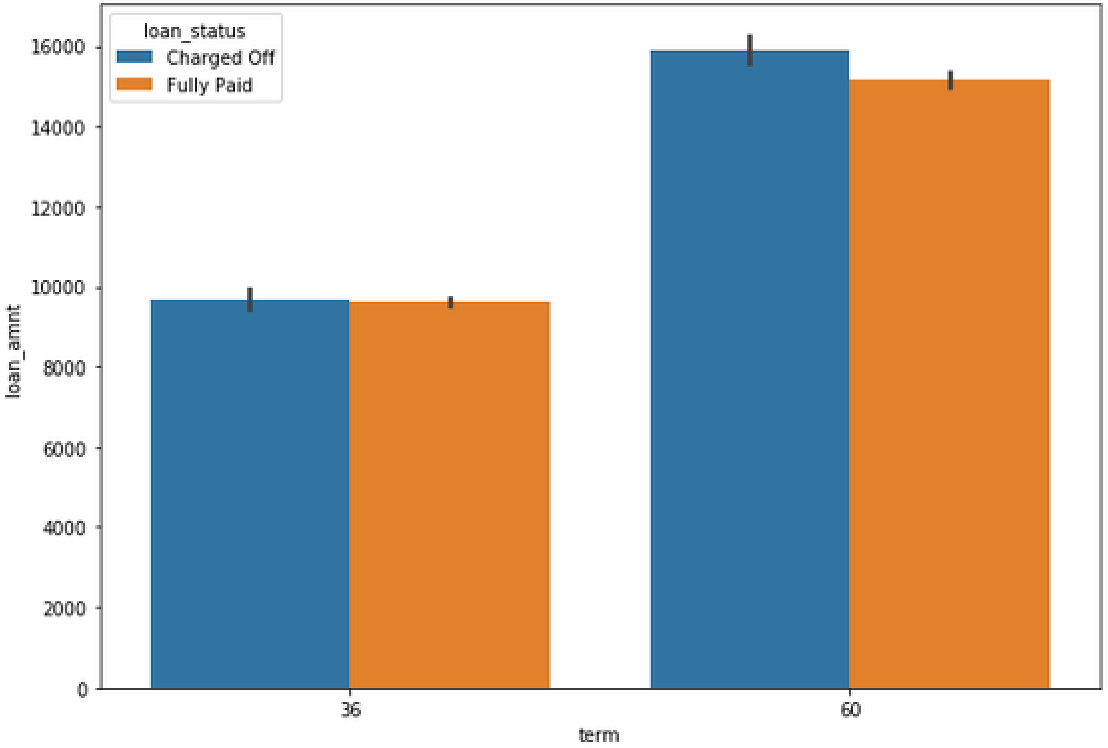
Bivariate Analysis

Loan Grade vs Loan Amount



Large amount Loans (more than 15,000 USD) of Category F & G are more likely to default than loans applied to any other grades.

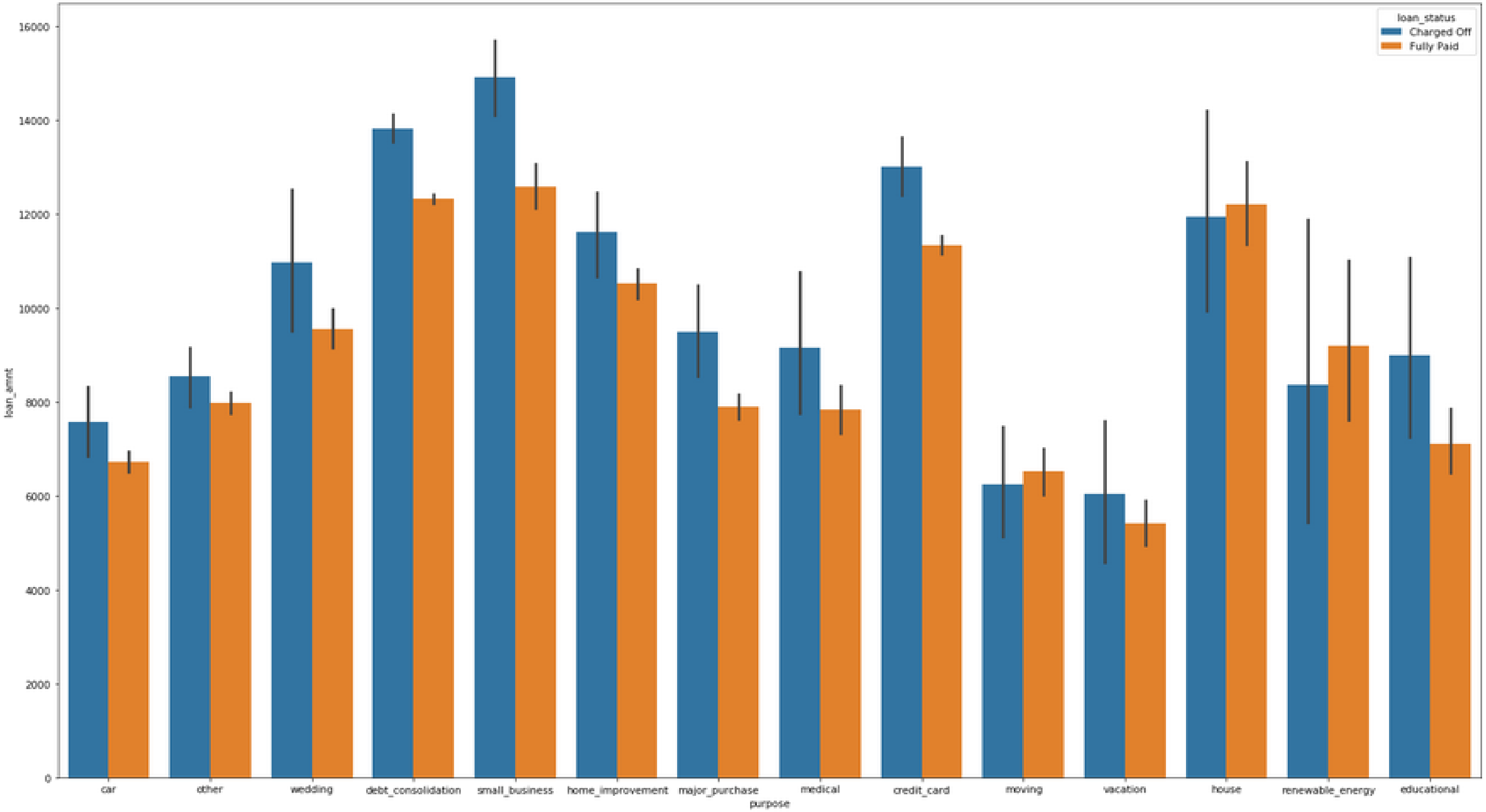
Loan Amount vs Term of Loan



Higher loan amount applied for longer term(60 months) have higher charge-offs or defaults.

Bivariate Analysis

Loan amount vs Purpose of loan

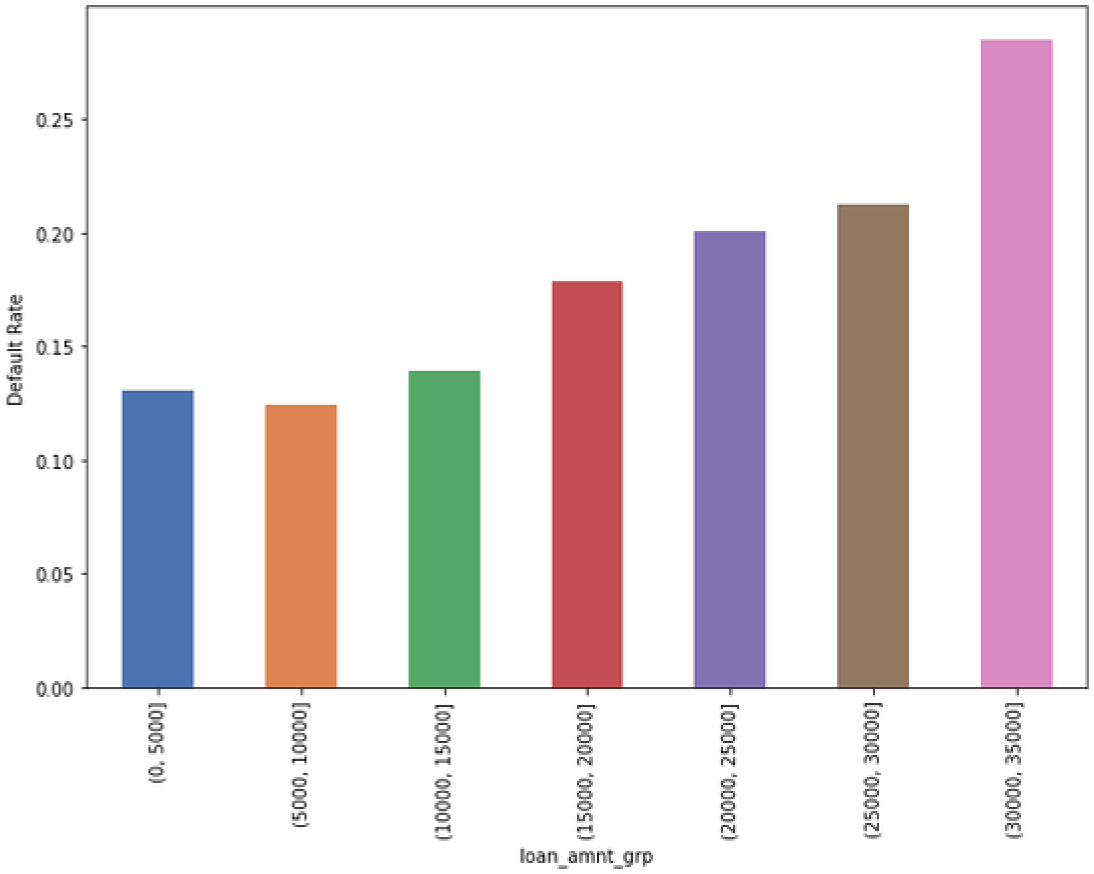


Customers who apply for large amount of loan for small business tend to default more compared to loans applied for other purpose.

Bivariate Analysis

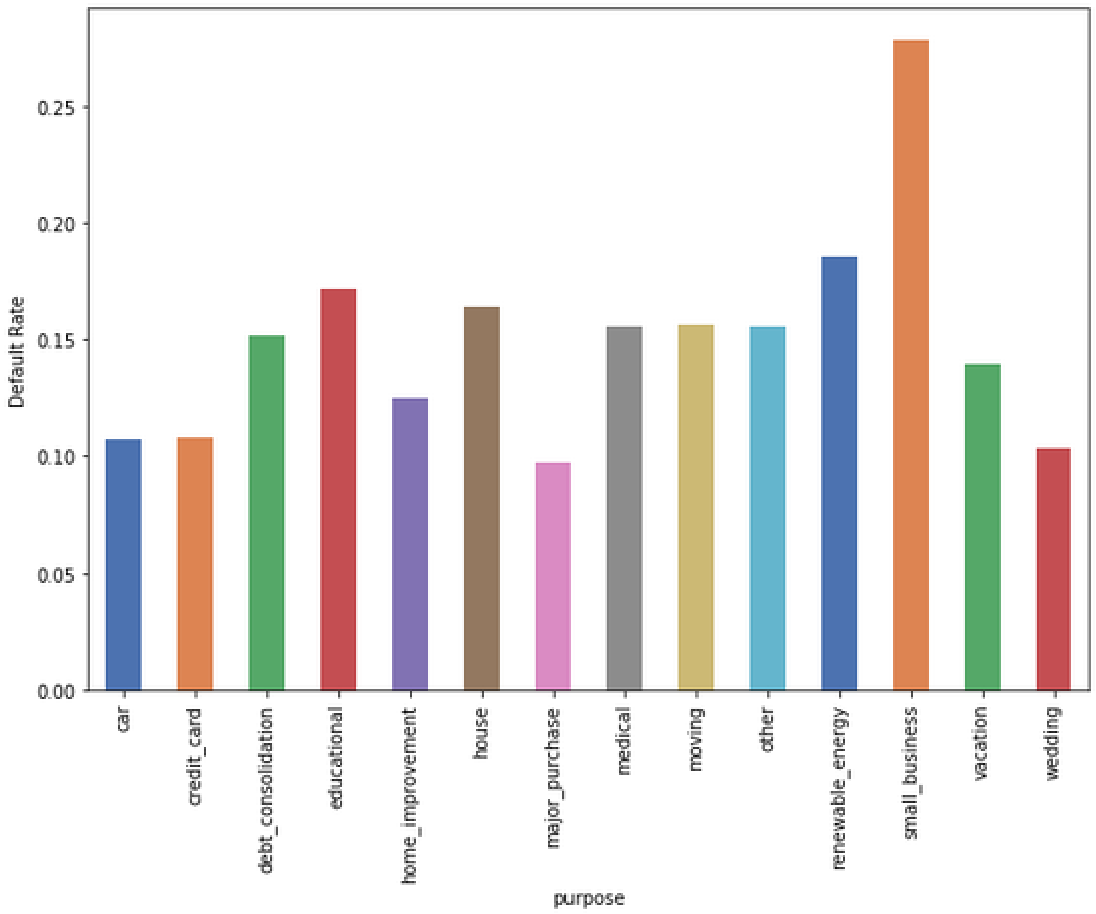
UpGrad

Default Rate vs Loan Amount



Higher loan amounts (more than 30,000 USD) have higher default rate (approx. 28%)

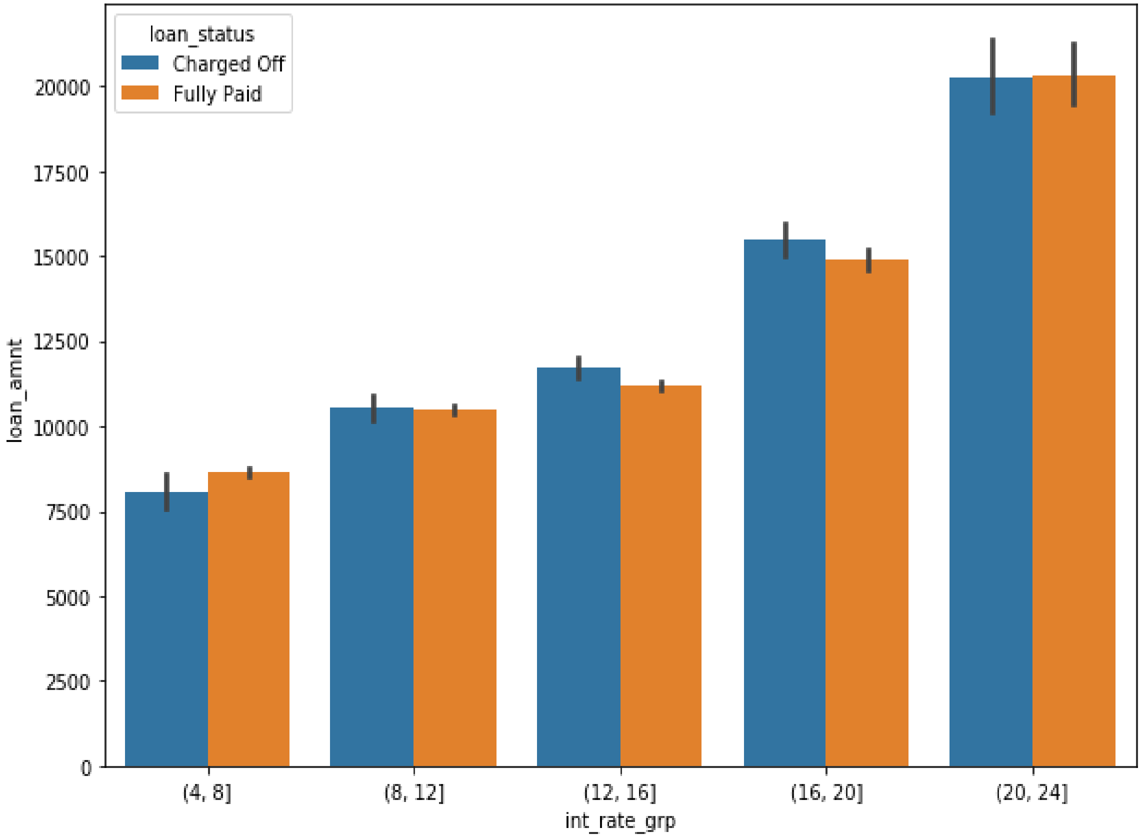
Default Rate vs Purpose of Loan



Loans applied for small business have highest default rate (approx. 28%)

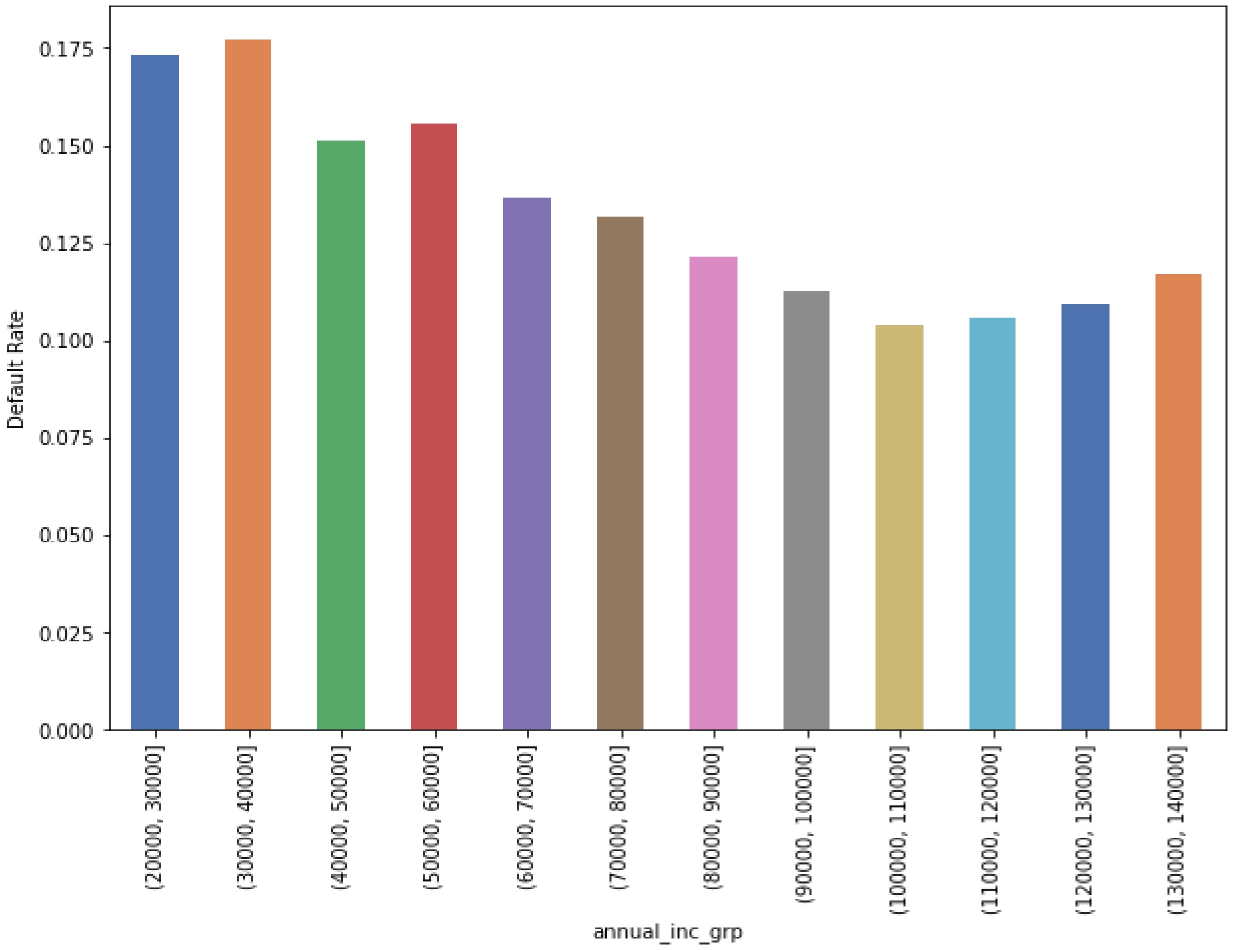
Bivariate Analysis

Interest rate vs Loan amount



Loans with high loan amount issued on a higher interest rate (more than 20%) tend to default more. Loans issued on high interest rates have highest default rate of approx. 40%

Default Rate vs Annual Income

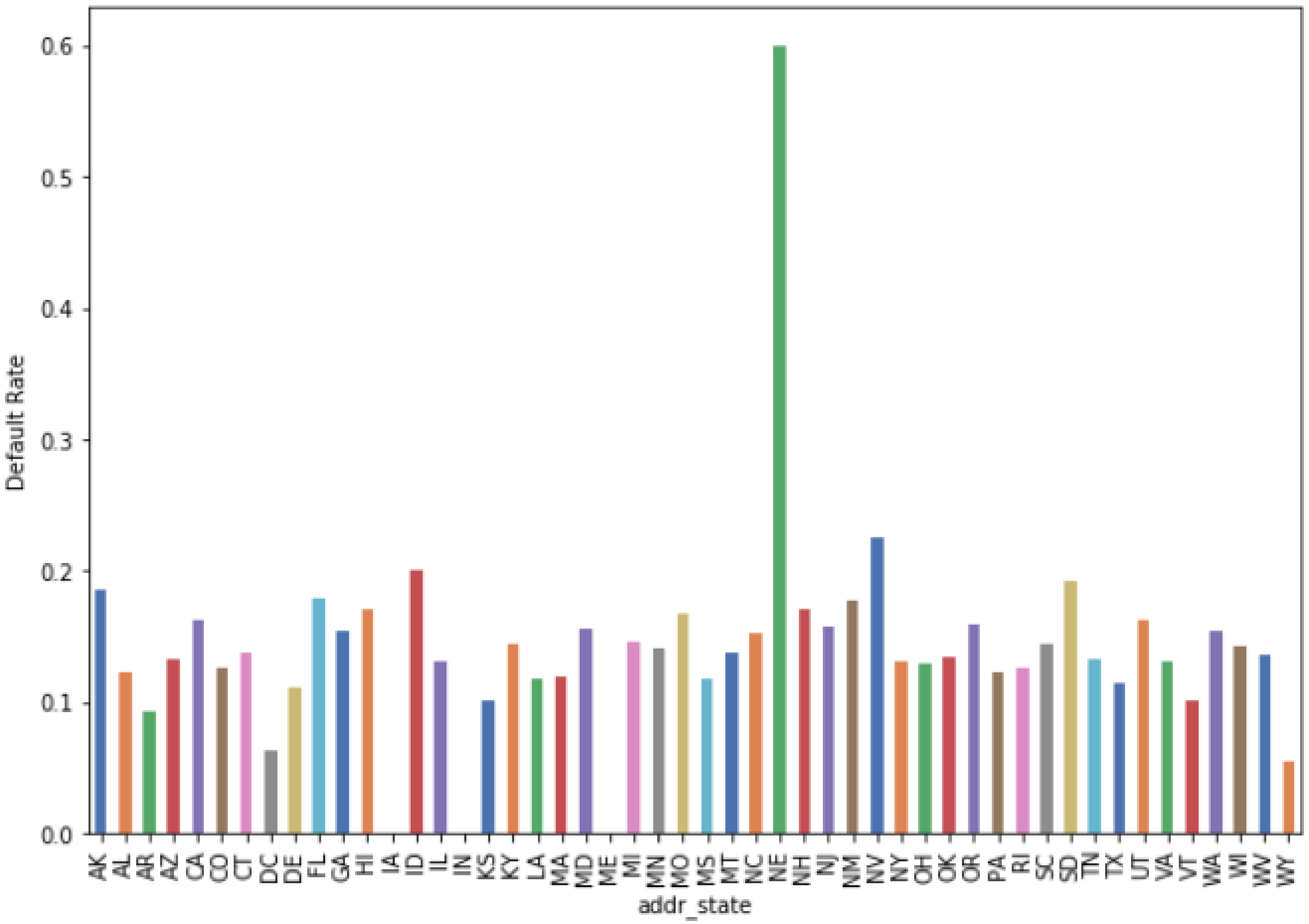


Loans issued to customers with low annual income (less than 40,000 USD) have high default rates as compared to others

Bivariate Analysis

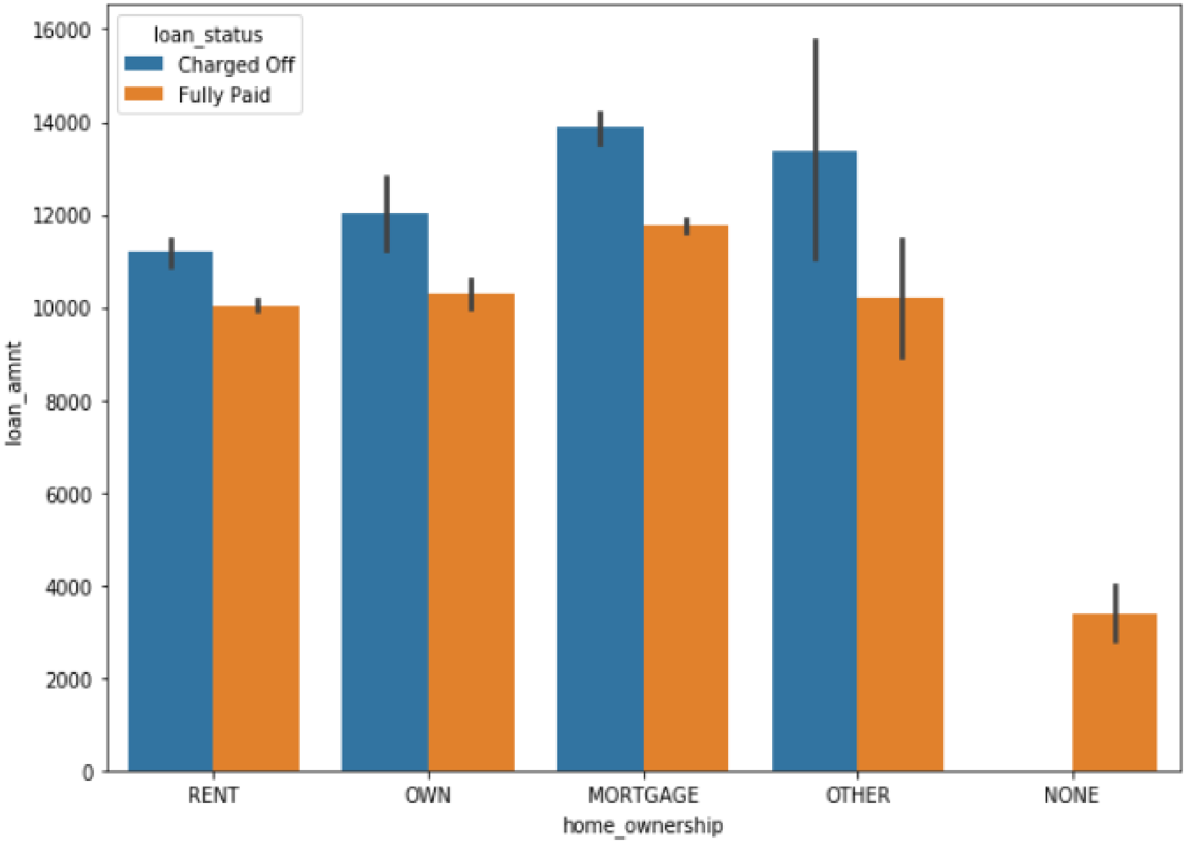
UpGrad

Default Rate vs Applicant State



Customers from Nebraska (NE) approx. 60% followed by Nevada (NV) approx. 20% have highest default rate

Loan Amount vs Home Ownership



Customers who have applied for high loan amount & have a 'Mortgage' on them tend to default more than others.

Recommendations

As per the aforementioned analysis following patterns are strong indicators of customers who are more likely to default and company should be more vigilant while lending to such customers.

- **Purpose**: Customers taking loans for DEBT CONSOLIDATION and taking high amount of loans for SMALL BUSINESS are risky and tend to default more compared to others.
- **Loan Grade**: Customers taking loans of grade F & G with high amount seems to have high default rates as per historic data.
- **Annual Income**: Customers with low annual income (less than 40,000 USD) have high tendency to default.
- **Interest Rate**: Loans with high loan amount when issued at high interest rates, customers tend to default more.
- **Term**: Customers who have loans of high amount issued for longer time period have more tendency to default

THANK YOU