LENDING CLUB CASE STUDY

Presented by:
Srinivas Yalagam
& Satyajit Chaudhuri

Our Team

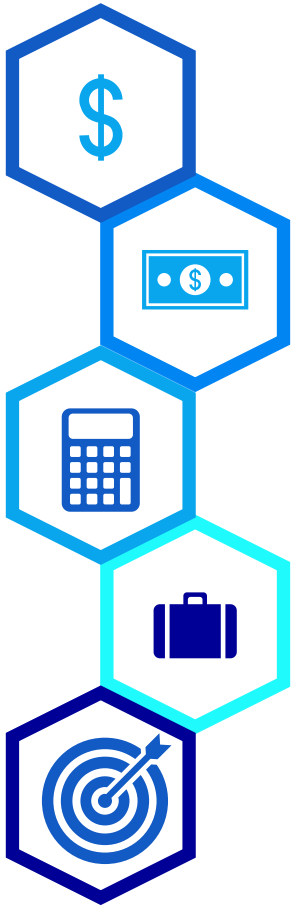
Srinivas Yalagam

(Group Facilitator)

Satyajit Chaudhuri

(Group Member)

Contents




DATA
UNDERSTANDING

Brief Introduction to Lending Club

- Lending Club enabled borrowers to create unsecured personal loan between $1,000 and $40,000.

- Investors were able to search and browse the loan listings on Lending Club website and select loans that they wanted to invest in based on the information supplied about
  - ✓ the borrower,
  - ✓ amount of loan,
  - ✓ loan grade,
  - ✓ and loan purpose.

- Investors made money from the interest on these loans.

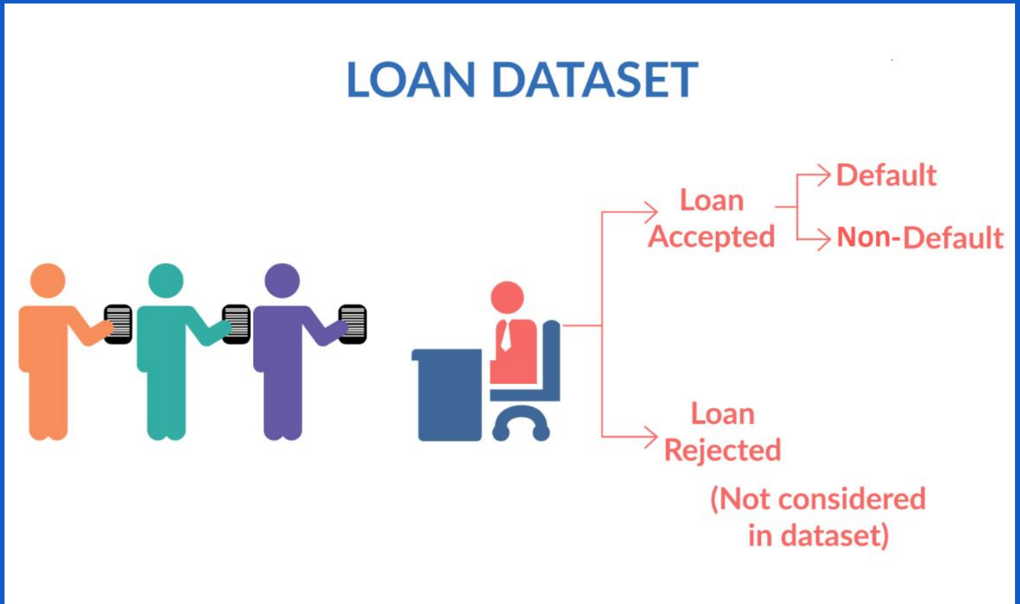
- Lending Club made money by charging borrowers an origination fee and investors a service fee.

Source: Wikipedia

OBJECTIVE OF THE CASE STUDY

- To analyze how consumer attributes and loan attributes influence the tendency of default.
- To identify patterns which indicate if a person is likely to default.
- To provide suggestions and recommendations on actions to be taken against risky applications.

Deep Diving Into The Data



The Dataset Parameters

About the Dataset…

- ❖ The given Data has 39717 rows and 111 columns.
- ❖ The data has 2263364 null values and amongst these 59 columns have null percentage more than 5%. Hence, they would not contribute much to the analysis.
- ❖ The dataset also has 8 columns with only unique values. The columns having unique values won't be required and hence won't contribute much to the analysis.
- ❖ Also, the data dictionary provided can be accessed from the attachment.

Data_Dictionary


DATA
CLEANING

Data Cleaning Methodology

01 **Removing Null Values**

The dataset happens to contain 59 rows where the values for about 95% of the rows is missing and hence it is felt that they won’t be contributing much to the analysis part.

02 **Removing Unique Values**

The dataset also happens to contain many columns which have only unique values and hence they were removed from the analysis point of view.

03 **Outlier Treatment**

After checking the quantile values, the outlier data was eliminated and about 99 percent of the data was kept for the analysis..

04 **Imputing Missing Values**

Some rows which were considered to have important attributes and had few missing values were appropriately imputed with meaningful data.


DATA
MANIPULATION

Data Manipulation Methodology

01

Date Separation

Here we extracted each month and year from issue_d, last_pymnt_d, last_credit_pull_d and earliest_cr_line columns and derive new columns based on year and month values

02

Dropping "Current" Status

Since we are only interested in fully paid and charged off loan status, we can remove the rows having loan status as CURRENT

03

Boolean Values for Loan – Status

Here we created a new data driven metric where we extracted boolean column from loan_status column.

04

Defining Business Driven Metrics

Introduced new metric "installment_to_income_ratio" in order to quantify the income and installment parity.

Data Manipulation Methodology

05

Modify column values for loan data analysis

Employment Length, term, and Home_ownership columns were modified to make the data more relevant.

06

Converting Numeric data to Categorical Bins

Dividied loan_amnt , int_rate, annual_inc, installment_to_income_ratio_percentage and dti into groups

07

Separating Numeric and Categoric Columns

Divided the data frame into numeric and categoric columns.

DATA ANALYSIS

UNIVARIATE ANALYSIS RESULTS

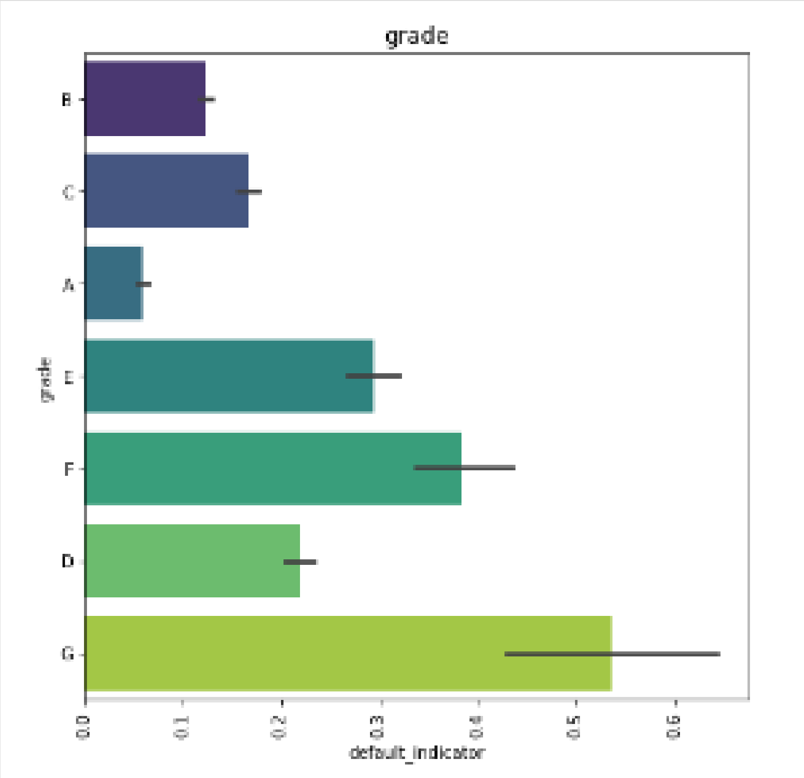


Fig: Grade-wise Chances of Default.

With decreasing grade (with G being lowest), the rate of Write off keeps on increasing.

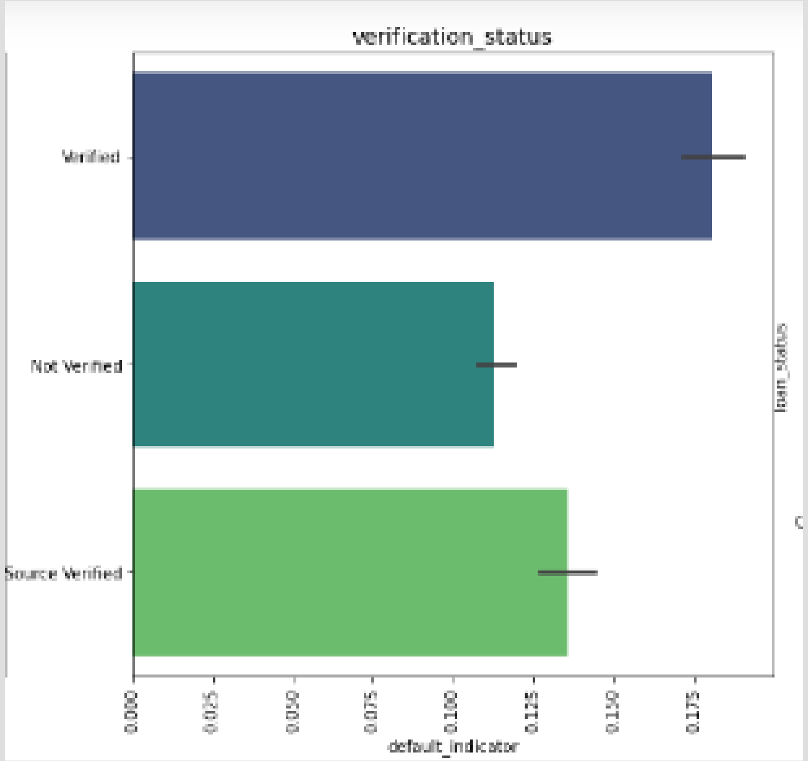


Fig: Verification vs Chances of Default.

With decreasing grade (with G being lowest), the rate of Write off keeps on increasing.

UNIVARIATE ANALYSIS RESULTS

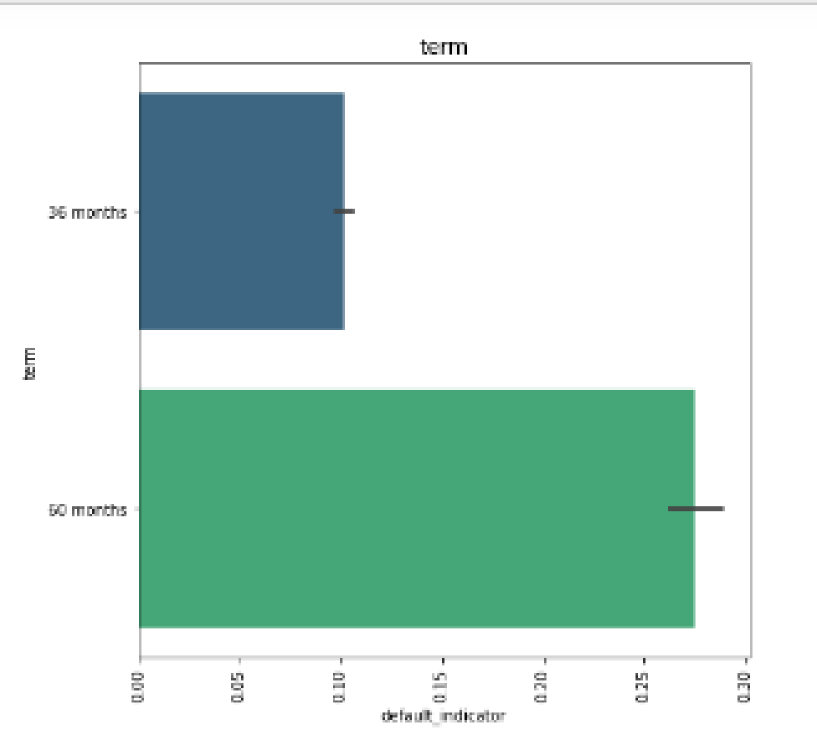


Fig: Term-wise Chances of Default.

With increase in repayment term period, the tendency to get Charged Off increases. In case of 60 months term the rate of default is much more than that off 36 months

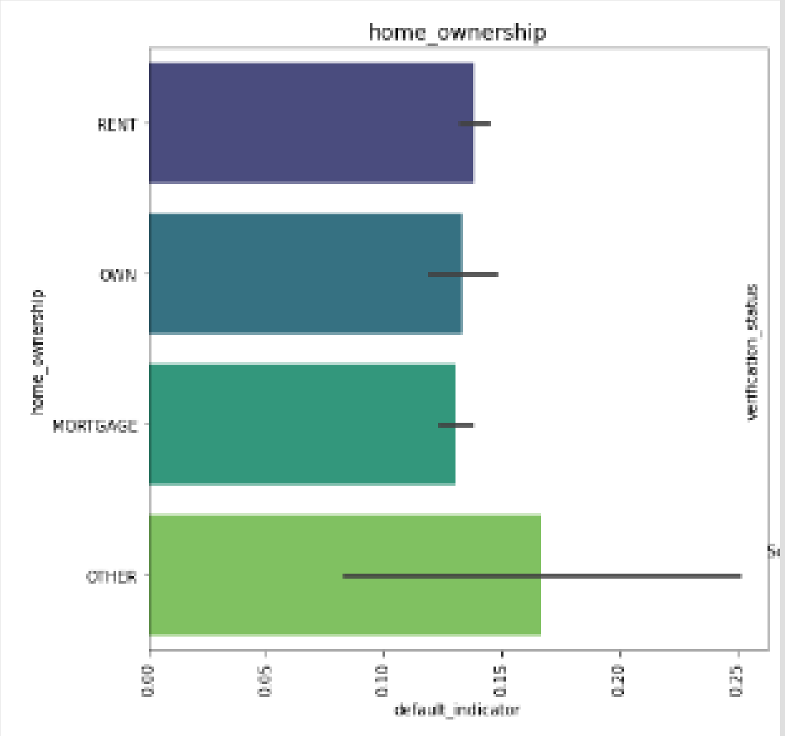


Fig: Home-Ownership vs Chances of Default.

Home-ownership, doesn't impact much on loan status, home ownership status as OTHER is more prone to getting charged off.

UNIVARIATE ANALYSIS RESULTS

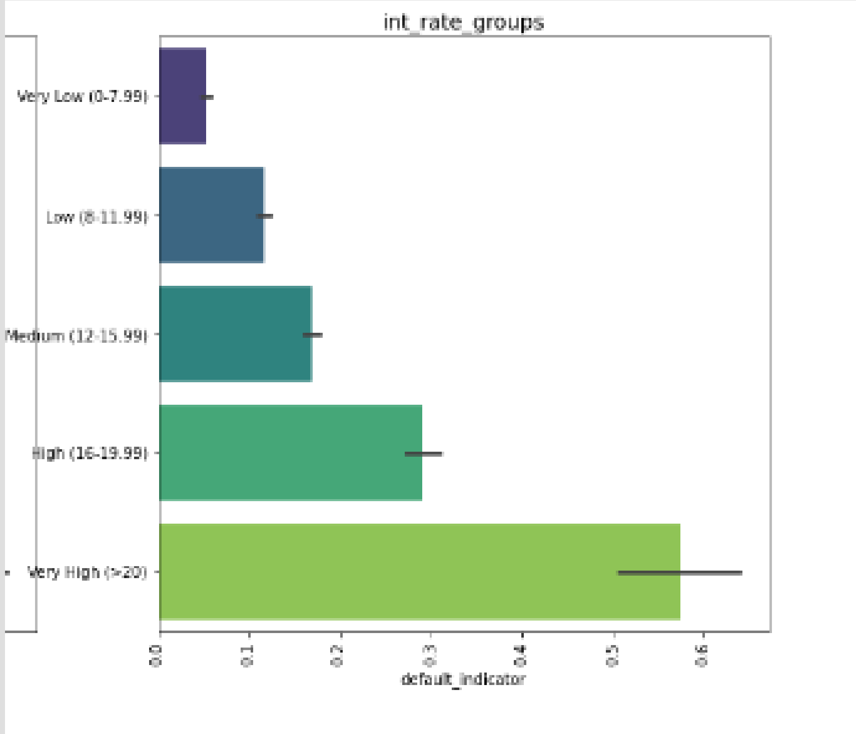


Fig: Interest Rate vs Chances of Default

Higher the Interest Rate, higher is the chance of defaulting.

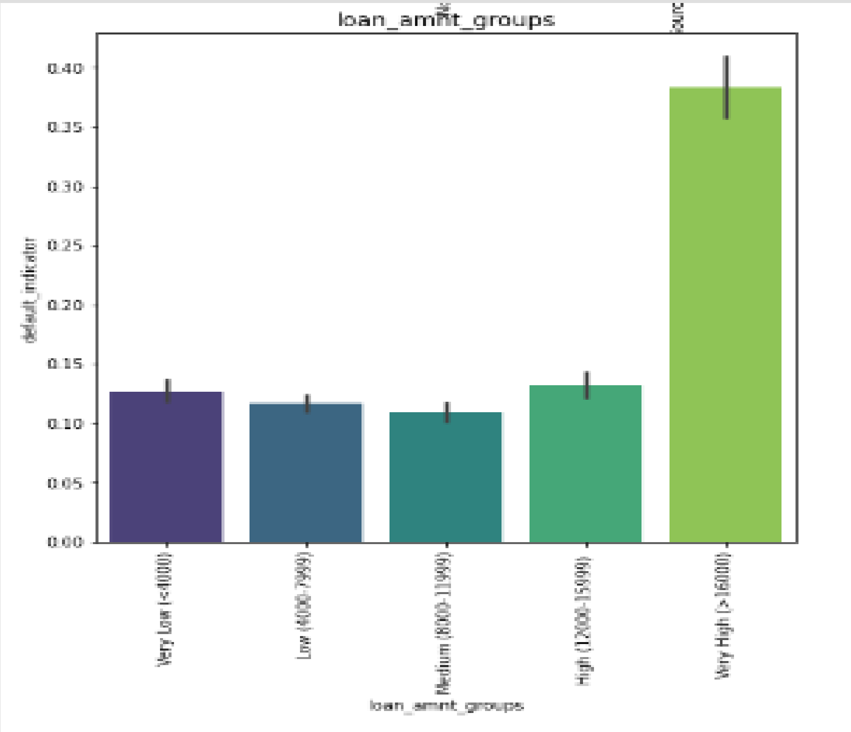


Fig: Home-Ownership vs Chances of Default.

Home-ownership, doesn't impact much on loan status, home ownership status as OTHER is more prone to getting charged off.

UNIVARIATE ANALYSIS RESULTS

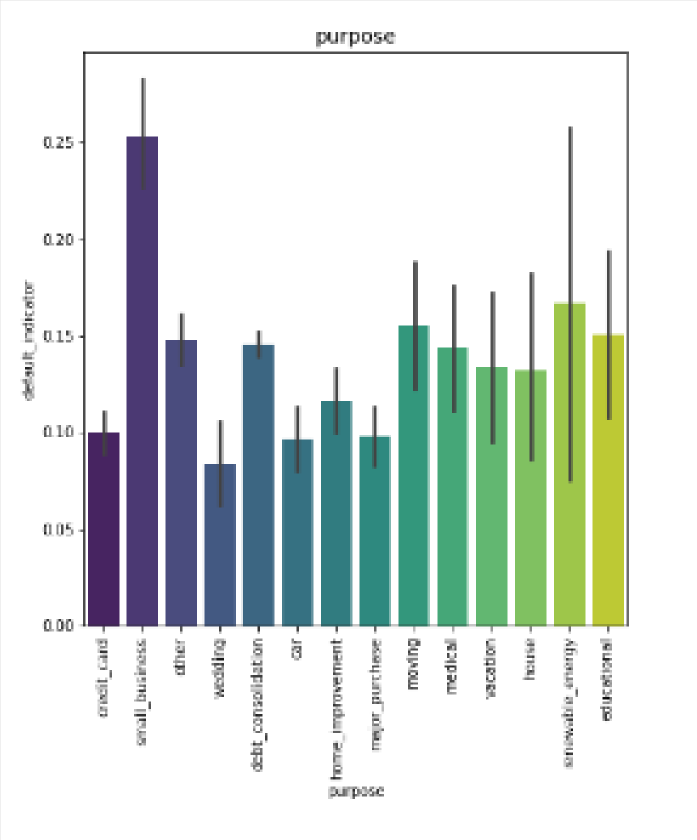


Fig: Purpose vs Chances of Default

If purpose of the loan is small business or renewable energy, higher is the chance of defaulting.

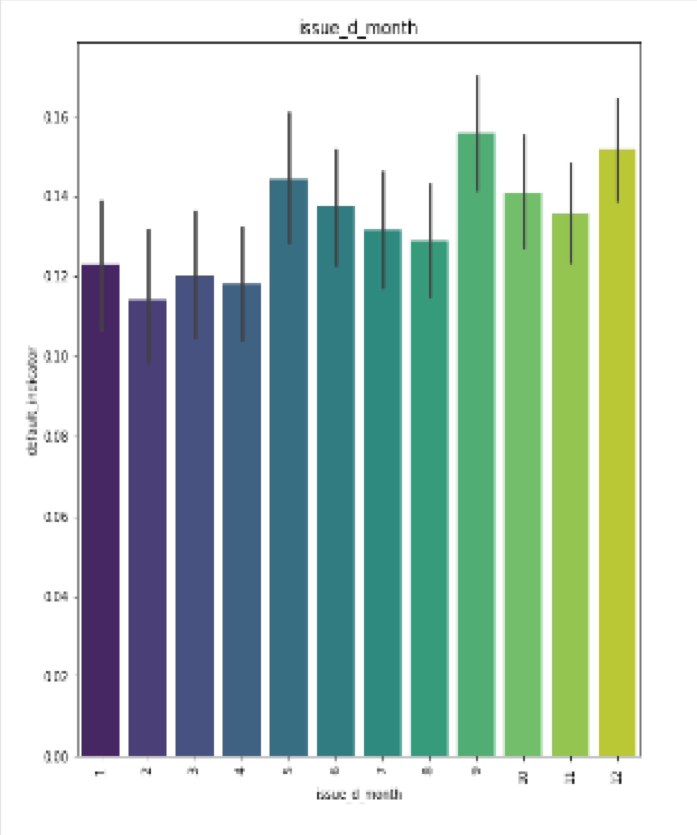


Fig: Issue Month vs Chances of Default.

Interestingly the loans taken in the month of September and December show a much higher chance of getting Charged Off..

Observations from Univariate and Segmented Univariate Analysis

❖The decreasing **Grade** (with G being lowest), the rate of Write off keeps on increasing.

❖Background check **Verification status** doesn't impact much. Surprisingly, the verified person is more prone to getting Charged Off.

❖With increase in repayment **Term** period, the tendency to get Charged Off increases. In case of 60 months term the rate of default is much more than that off 36 months.

❖**Home-ownership**, doesn't impact much on loan status, home ownership status as OTHER is more prone to getting charged off.

❖Higher the **Interest Rate**, higher is the chance of defaulting.

❖Higher default rate is seen in **loan amounts** above 16000.

❖If **purpose** of the loan is small business or renewable energy, higher is the chance of defaulting.

❖Interestingly the loans taken in the **month** of September and December show a much higher chance of getting Charged Off.

BIVARIATE ANALYSIS RESULTS

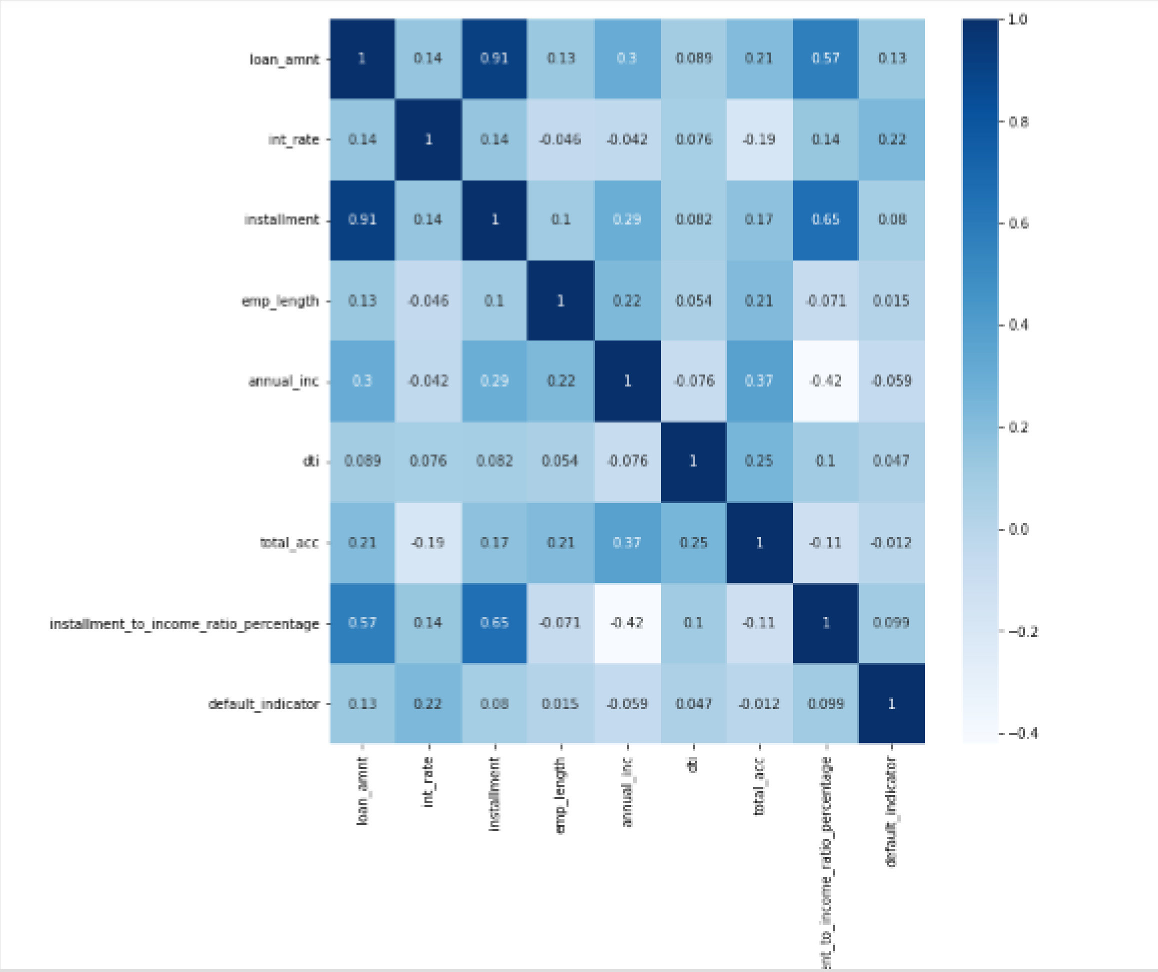


Fig: Correlation Visualization

From this we can see that "Installment - Loan_amount" & "Installment - Annual_Income" are highly correlated..

BIVARIATE ANALYSIS RESULTS

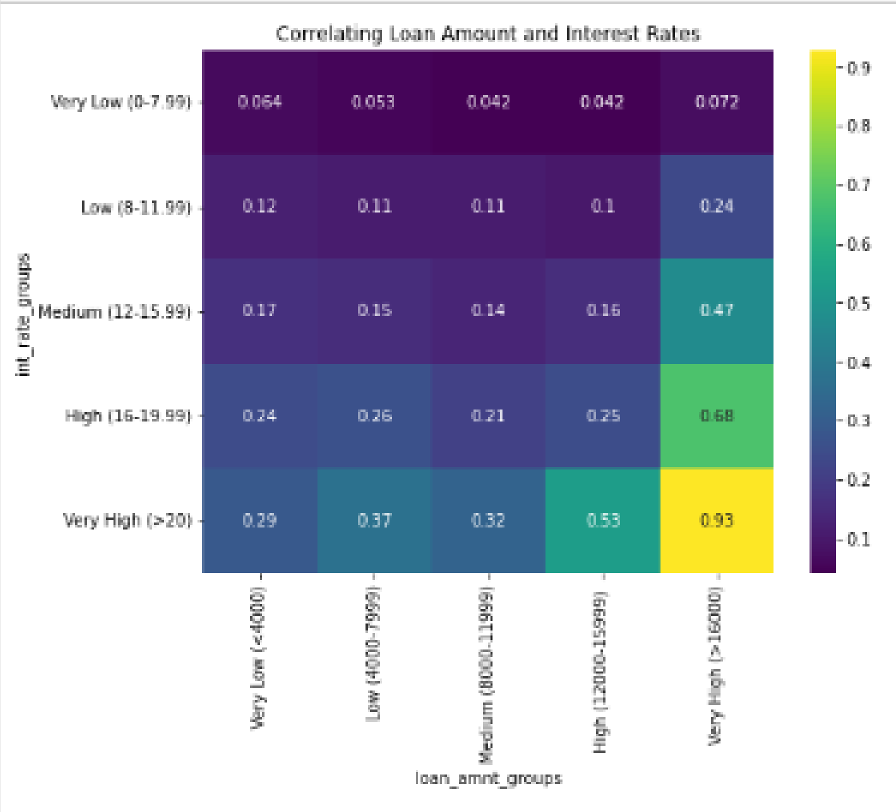


Fig: Loan-Amount vs Interest Rate

From Analysis of Loan Amount with Interest Rates, it is quite evident that very high interest rates come with higher risk of default. And this risk increases for Loan amounts more than 12000.

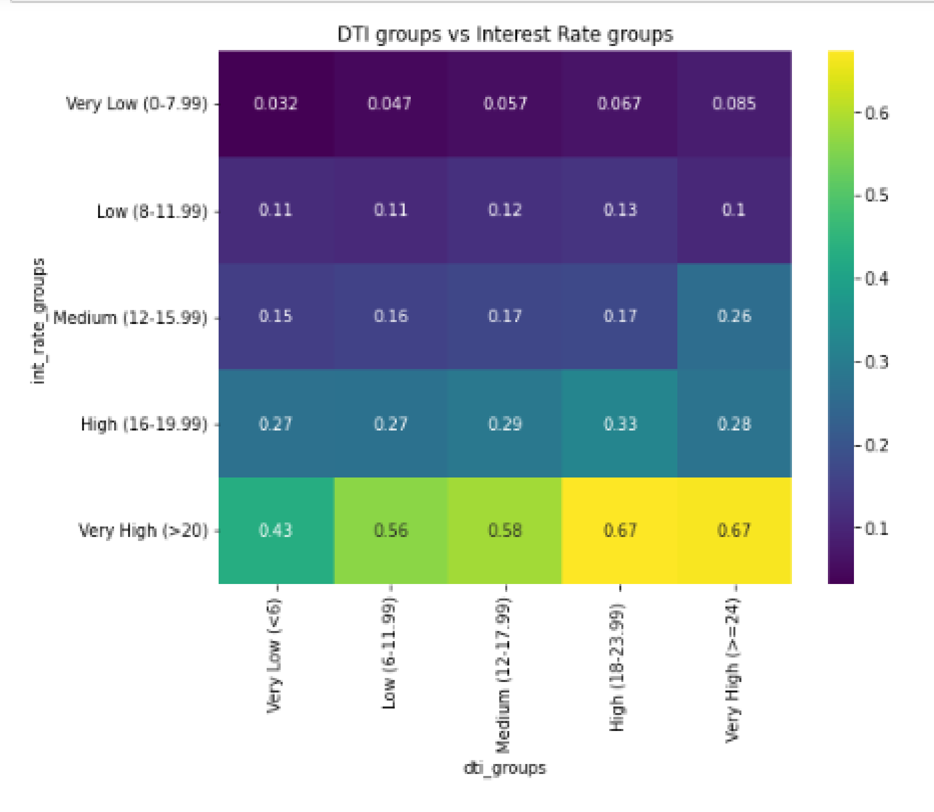


Fig: Issue Month vs Chances of Default.

Interestingly the loans taken in the month of September and December show a much higher chance of getting Charged Off..

BIVARIATE ANALYSIS RESULTS

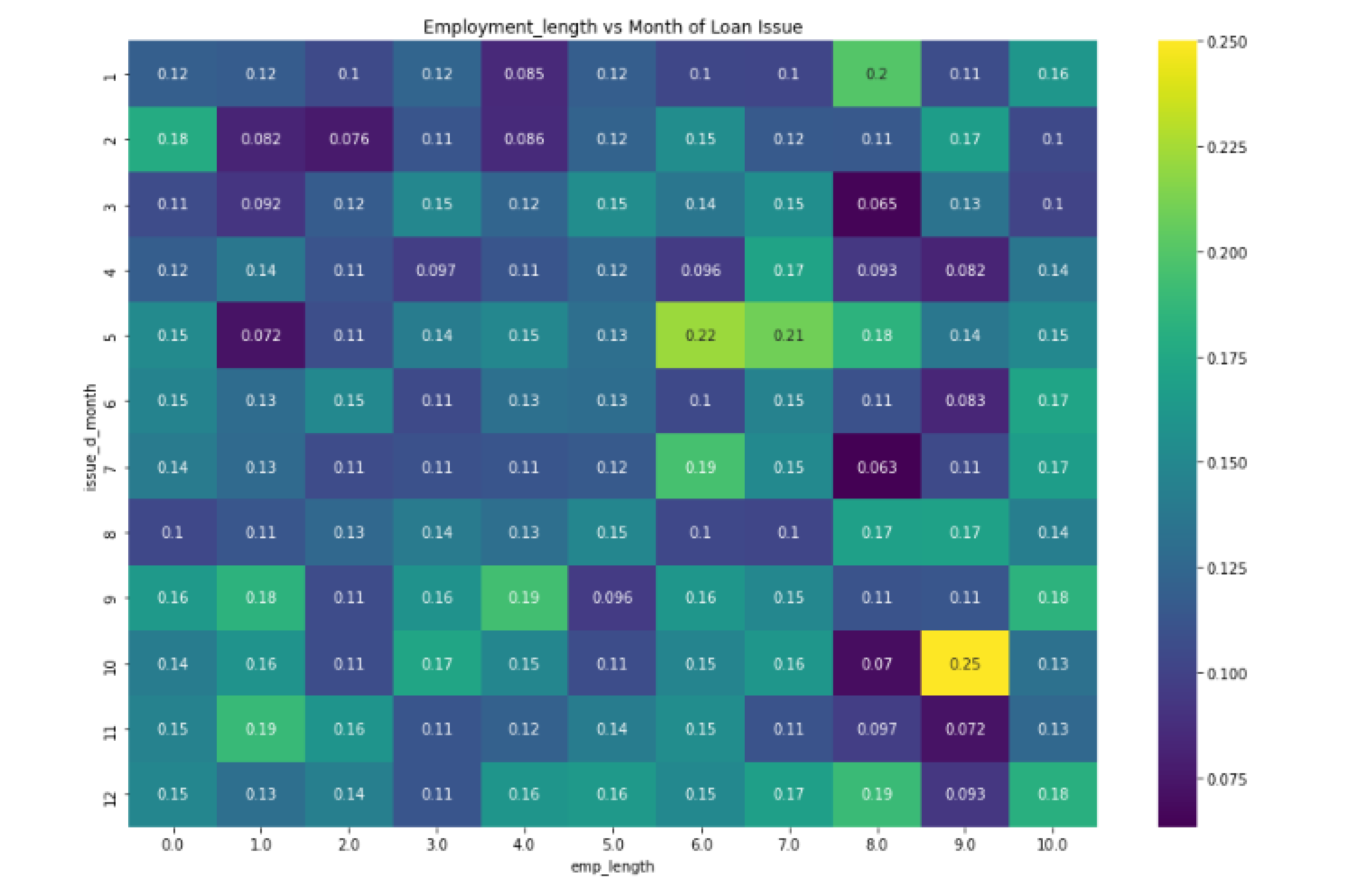


Fig: Employment Length vs Month of Loan

While reading Employment length with Month of Loan, it can be seen that the default rate of May borrowers with six and seven years experience are higher then other months. Employees with 9 years of experience and issue month is October have more chances of getting defaulted.

BIVARIATE ANALYSIS RESULTS

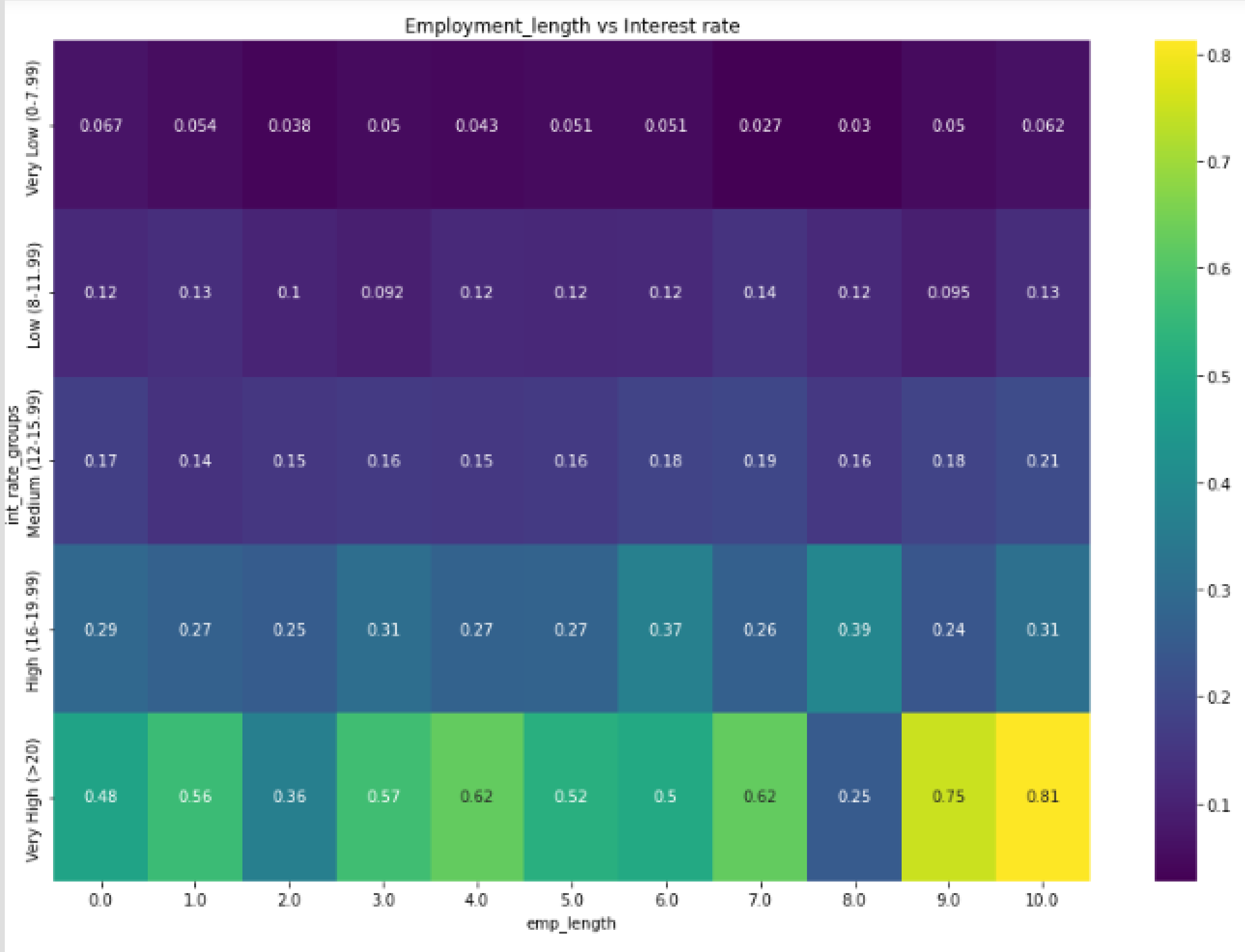


Fig: Employment Length vs Interest Rate

With higher Employment Length and Very High Interest Rate, the tendency to default keeps on increasing. The highest chances of default occurs at employment length above 9 years and Very High Interest Rate

BIVARIATE ANALYSIS RESULTS

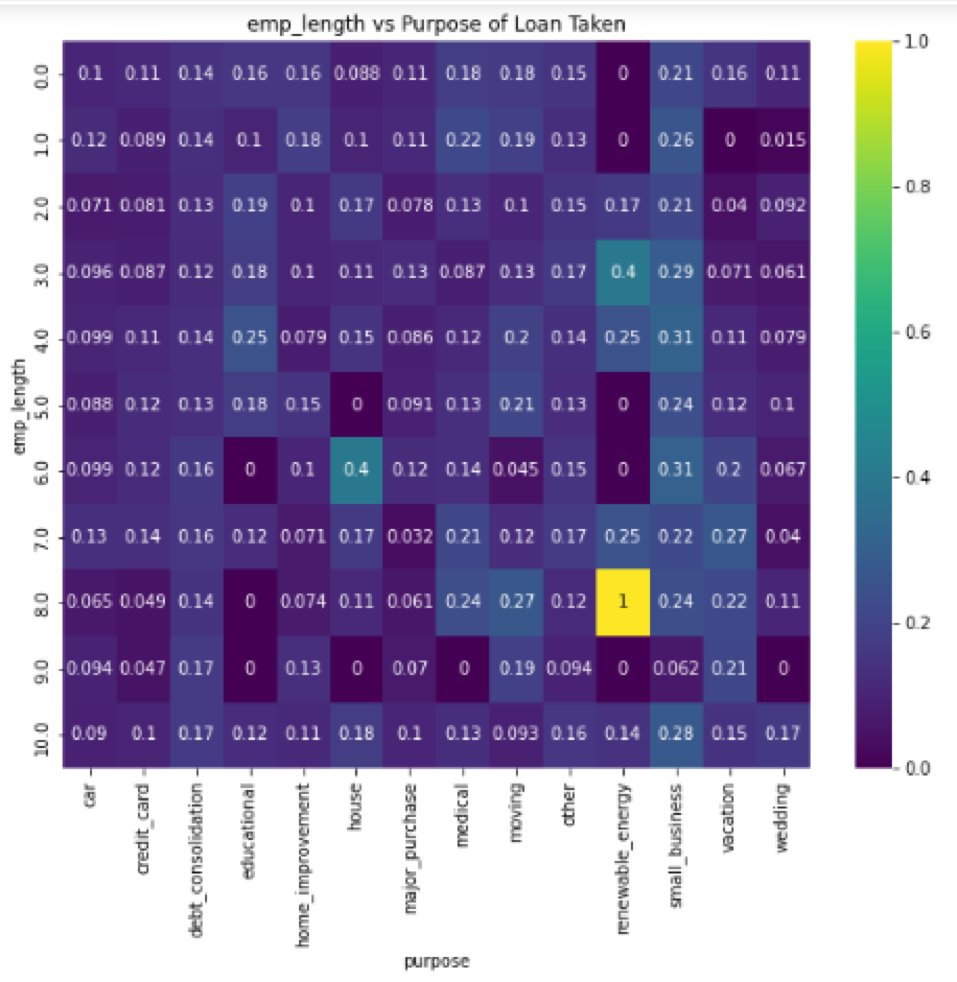


Fig: Employment Length vs Purpose of Loan

While investigating relationship of Employment length and Purpose, it can be noted that the ones with purpose as "renewable energy" and "vacation" are more prone to being "Charged off".

BIVARIATE ANALYSIS RESULTS

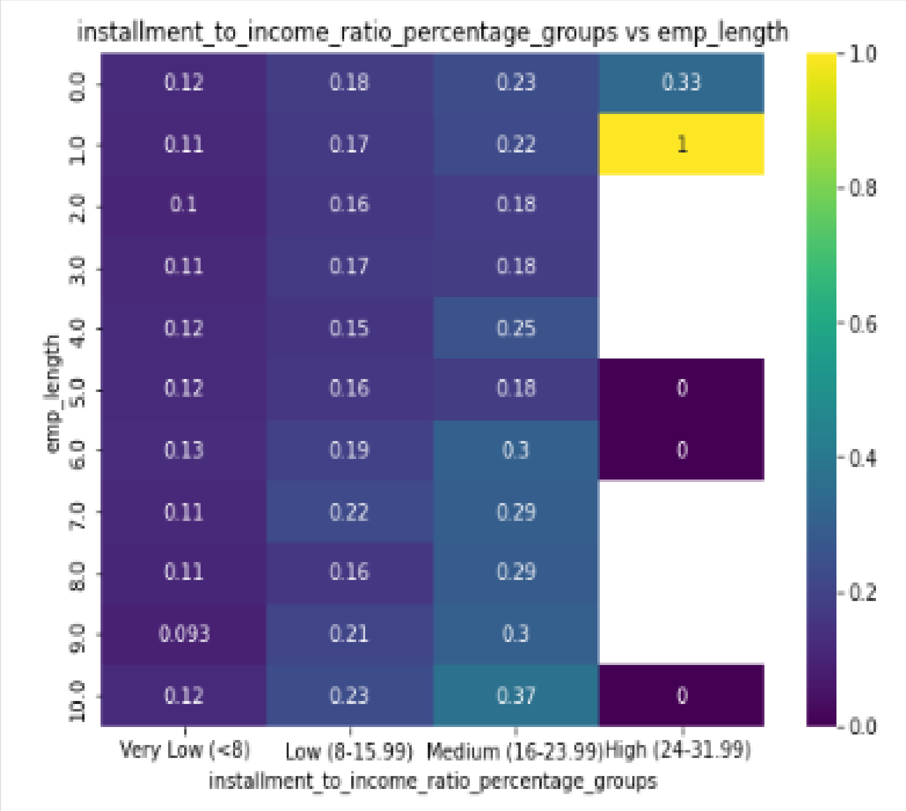


Fig: Installment to Income Ratio vs Employee Length

In comparing employment length to Installment to Income ratio, the ones with very high Installment to Income ratio and less experience has a higher chance of being Charged off.

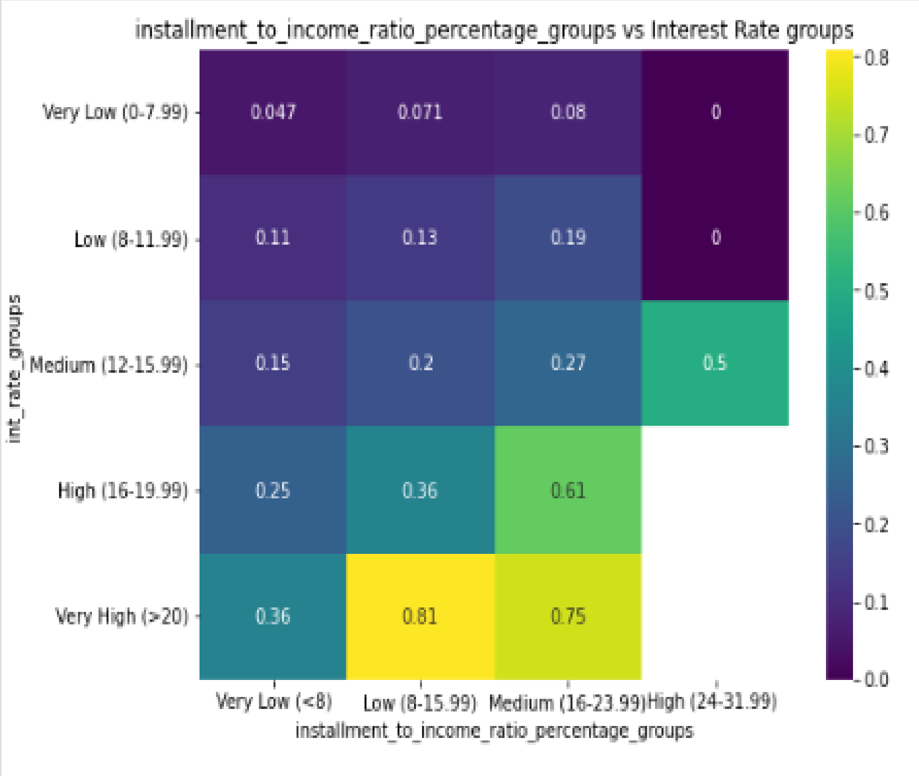


Fig: Installment to Income Ratio vs Interest Rates.

An applicant falling in High Interest Rate Bucket and Low installment_to_income_ratio bucket is more prone to get charged off..

BIVARIATE ANALYSIS RESULTS

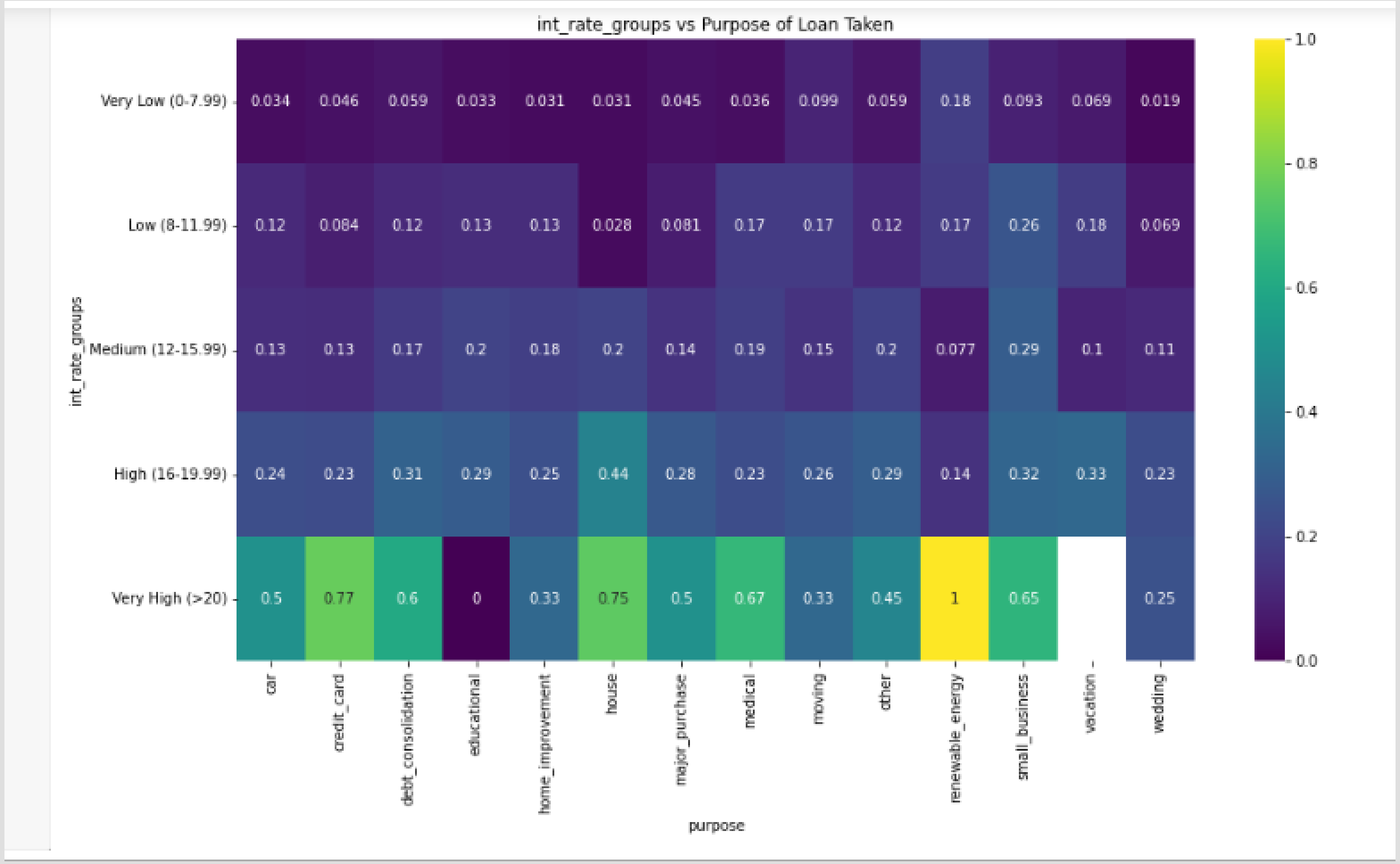


Fig: Interest Rate vs Purpose of Loan

While investigating relationship of interest rates and purpose, it can be noted that the high interest rates cause more defaults and "renewable energy" , "credit card" and "house" are more prone to being "Charged off".

BIVARIATE ANALYSIS RESULTS

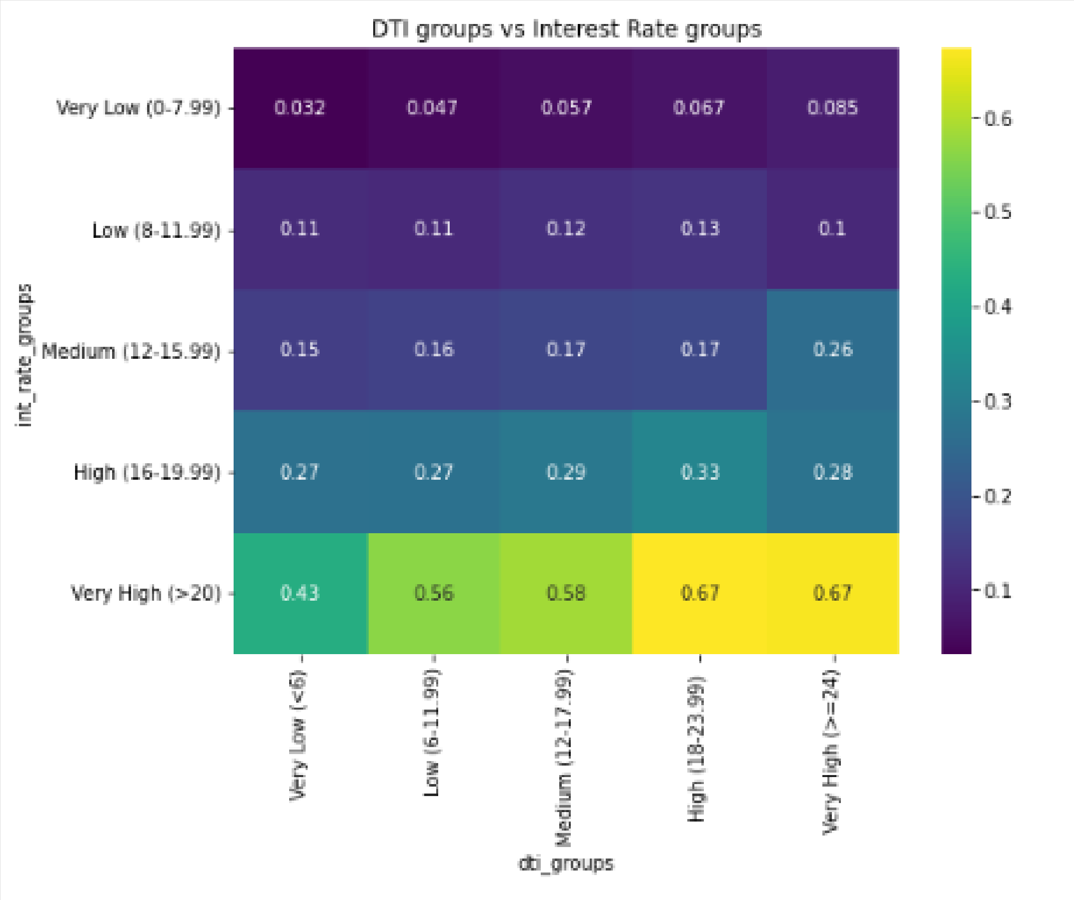


Fig: DTI groups vs Interest Rate groups

While analysis DTI groups with the Interest rate groups, it is notable that the combination of higher DTI with higher interest rates groups are more tend to getting default.

Observations from Bivariate Analysis

- From Analysis of Loan Amount with Interest Rates, it is quite evident that very high interest rates come with higher risk of default.And this risk increases for Loan amounts more than 12000.

- While reading Employment length with Month of Loan, it can be seen that the The default rate of May borrrowers with six and seven years experience are higher then other months. Employees with 9 years of experience and issue month is October have more chances of getting defaulted.

- While investigating relationship of Employment length and Purpose, it can be noted that the ones with purpose as "renewable energy" and "vacation" are more prone to being "Charged off".

- In comparing employment length to Installment to Income ratio, the ones with very high Installment to Income ratio and less experience has a higher chance of being Charged off.

- With higher Employment Length and Very High Interest Rate, the tendency to default keeps on increasing. The highest chances of default occurs at employment length above 9 years and Very High Interest Rate.

- While analysis dti groups with the Interest rate groups, it is notable that the combination of higher dti with higher interest rates groups are more tend to getting default.

- In comparing interest rates with Installment to Income ratio, it can be seen that with increase in interest rate slabs and increasing Installment to Income ratio, the rate of defaulting increases. The risk becomes highest at the combination of very high IR and very high installment to income ratio.

- With decreasing Subgrade and increasing Loan Amount, the tendency to default keeps on increasing. The Grades F & G have a higher tendency to default.

- Again an applicant falling in High Interest Rate Bucket and Low installment_to_income_ratio bucket is more prone to get charged off.

- From the correlation visualization we can see that :

  Installment - Loan_amount & Installment - Annual_Income are highly correlated.

Recommendations for Lending Club Management

- The applicants taking a loan for "Small businesses" or "Renewable Energy" purpose often fail to pay back the loan. But increasing the interest rates may discourage general public from taking up this professions hence a more in depth research into the specific pain areas can be done. Special loan clauses might be designed for these applicants.

- For loan amount asks above 16000, the funded amount must be cautiously checked and a lesser amount might be funded with a higher interest rate. The applicants categorized in Grade F & G and accompanying subgrades should be handled with special care. Considering the other contributing factors the company needs to decide whether to increase the interest rates or decrease the funding.

- Applicants with "OTHER" as home ownership status may be funded lesser than their ask as the tendency to get written off is more in this group.

- The analysis found that Verification Status of the applicant doesn't contribute much to his chance of getting charged off. Hence the loan granting company needs to carefully analyze if the man hours and cost impact incurred due to this verification_status justified for the business.

Recommendations for Lending Club Management

- While analysis dti groups with the Interest rate groups, it is notable that the combination of higher dti with higher interest rates groups are more tend to getting default. In this cases funding may be reduced for such applicants.

- Applicants stating "Vacation" as Purpose and having experience on the higher side might be charged higher interest rates as they are more prone to default.

- While granting the loan, keen attention needs to be paid, to the ratio between the total monthly loan installments and the monthly income of the applicant. The higher the ratio, the lower is the chance of getting the loan "Fully Paid". Hence, funded amount needs to be reduced in-order to keep this ratio low.

THANK YOU

Lending Club Case Study