

Lending Club Case Study

Group Facilitator :

Name: **Harshali Bhoyar**

Email ID: **harshalibhoyar111@gmail.com**

Phone no: **8857069941**

Team Member Detail:

Name: **Rupesh Shukla**

Email ID: **rupesh.shuklak@gmail.com**

Phone no: **9790715871**

**Assignment
Module**

Exploratory Data Analysis (EDA)

Assignment Title

Lending Club Case Study

**Assignment
Objective**

Right lending decisions based on the likelihood of :

1. Applicants likely to default (Reduce credit loss)
2. Creditworthy applicants (Increased business opportunities)

**Benefits to the
Company**

1. Reduced loss by identifying applicants likely to default
2. Improved Risk Assessment
3. Adjusted loans, interest rates based on risk factors
4. Informed decision for loan approval
5. Improved loan lending strategy
6. Industry leadership and competitive advantage

Methodology

PLAN

Understanding the business requirement and objectives of the project



Identifying right tools to be used for the EDA



DO

Data cleaning & manipulation



ACT

Data analysis using the Univariate, Bivariate and using visualization plots



CHECK

Conclusion from the analysis and recommendation

Tools used

Tools Used

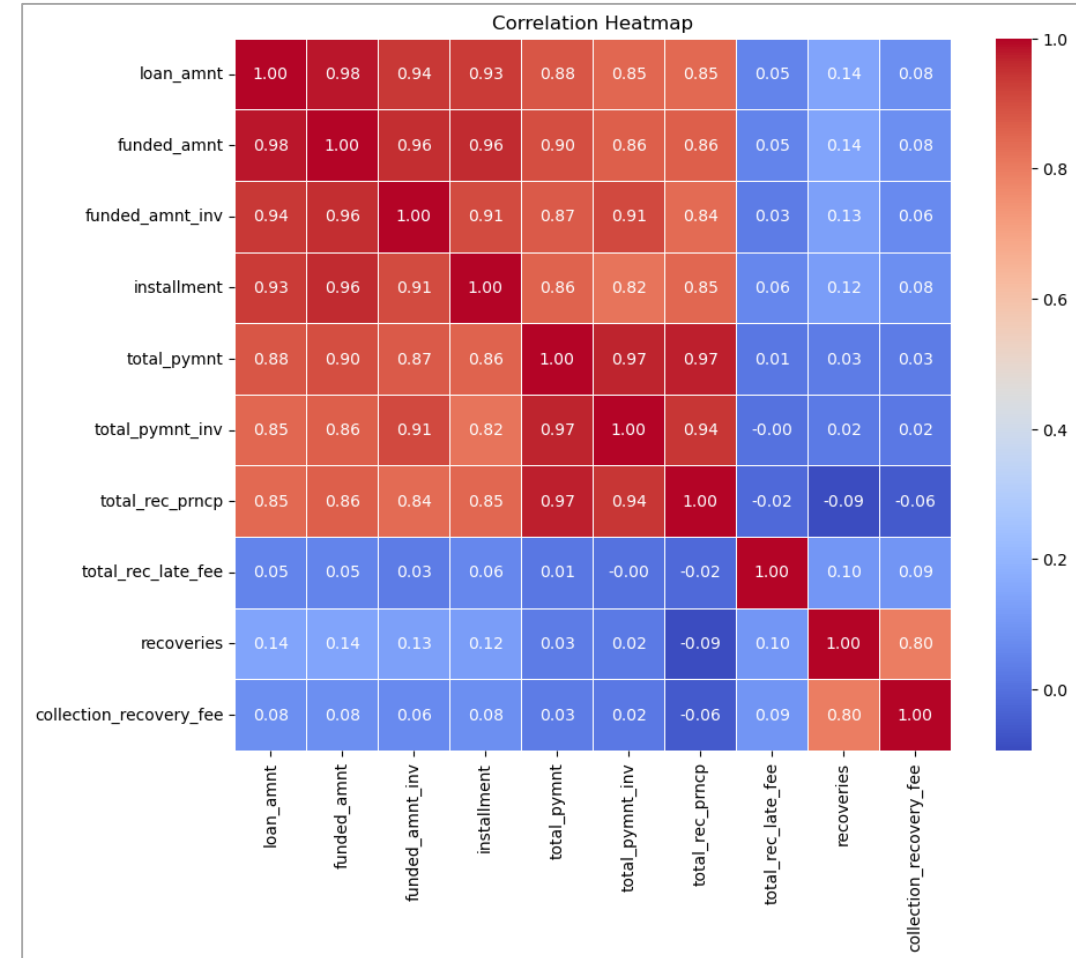
1. Jupyter Notebook
2. Git/Github
3. Excel

Python Libraries

1. Pandas
2. Matplotlib
3. seaborn

Data Cleaning

- Filter Loan Status which is Fully Paid or charged off is the data of our interest
- Identify and drop the columns that are highly co-related
- Identify the number of unique values
- Identify the Columns with missing values
- Identify columns with least significance
 1. Any categorical data with single values won't give us any analytical info
 2. Any categorical data with a huge number of values will be difficult to analyze
 3. Any continuous data with very less unique values tends to lack analytical information



Feature : Revolving utilization

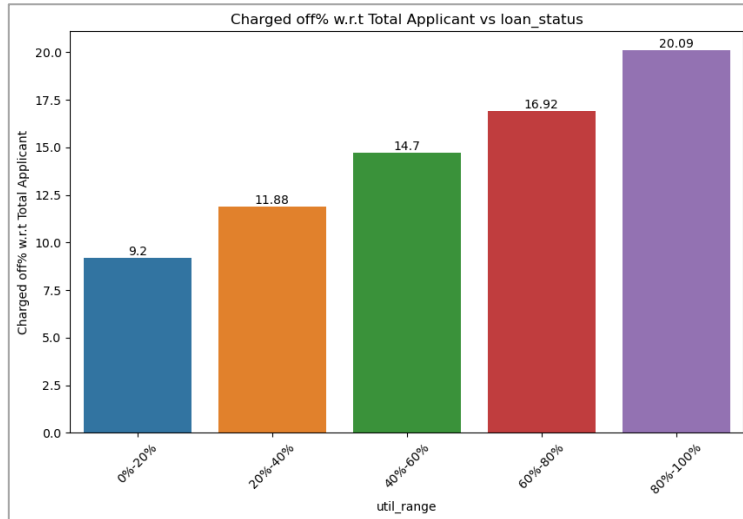


Chart 1: Utilization Range v/s charged off % w.r.t total applicants

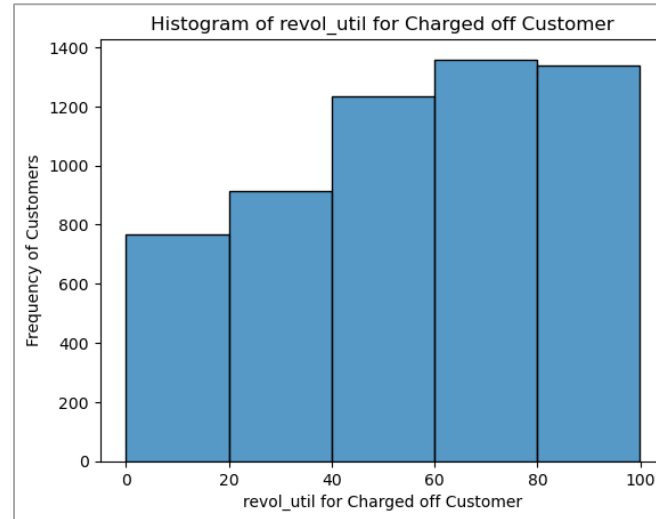
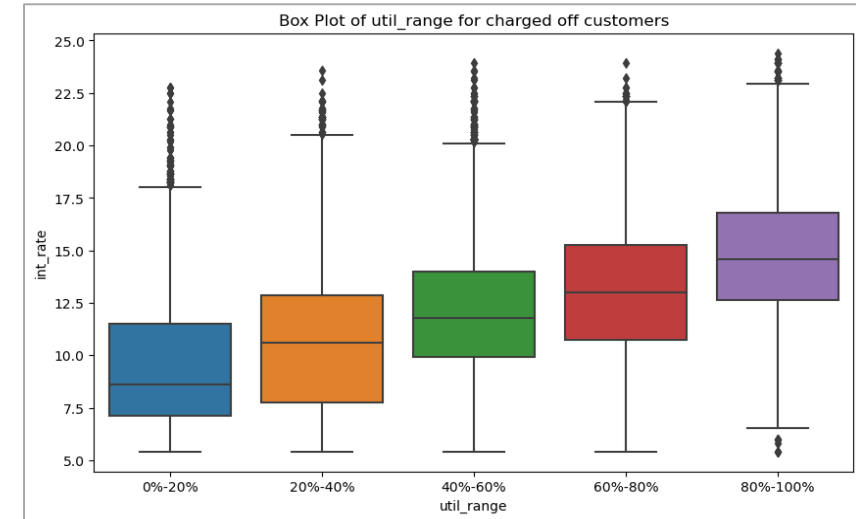


Chart 2 : Frequency of Charged off applicants w.r.t Utilization



Box plot of utilization range w.r.t interest rate for total applicants

1. The charged off % increases with increasing utilization range.
2. The higher the utilization percentage the frequency of getting charged off increases.
3. Thus, we can say that the higher the credit utilization percentage higher the risk
4. The applicants with higher credit utilization have higher interest rates at which the loan is given

Feature : Interest Rate

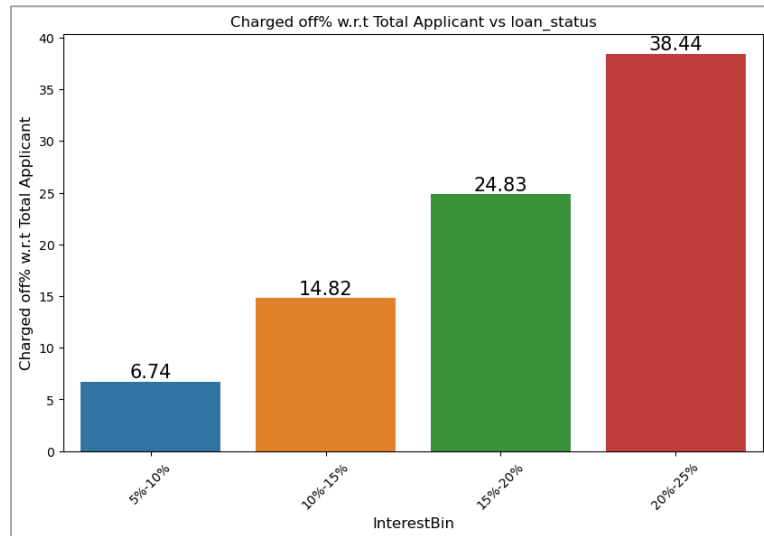


Chart 1: Interest Rate v/s charged off % w.r.t total applicants

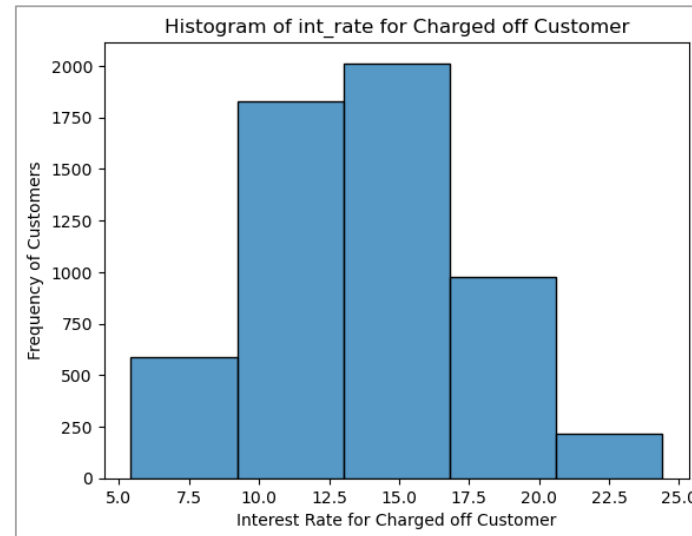


Chart 2 : Frequency of Charged off applicants w.r.t Interest Rate

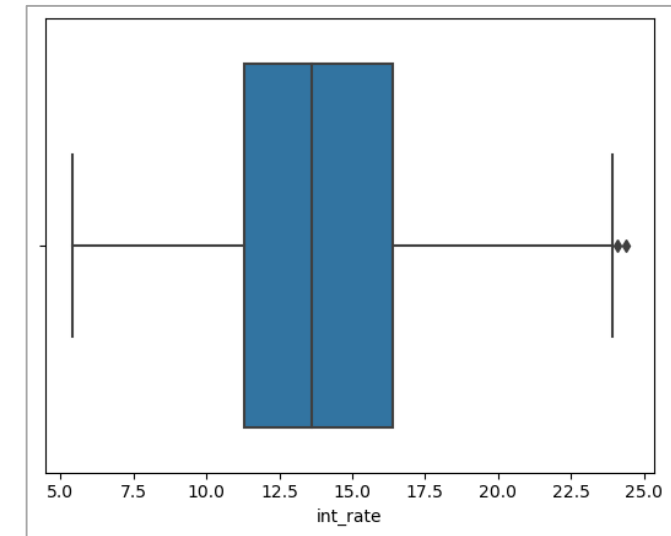


Chart 3 : Box plot of Interest rate

1. The charged-off % increases with increasing interest rates.
2. The higher charged-off frequency is for the range b/w 12.5% – 17.5% interest rate
3. Thus, we can say that the people with higher interest rates have been charged-off more.
4. The most charged off employees are concentrated b/w 11%-16%

Feature : Grade of the applicant

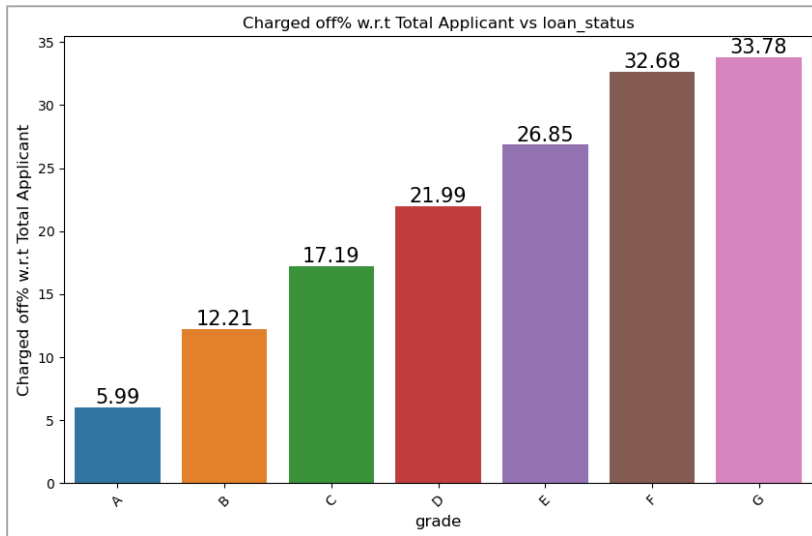


Chart 1: Grade of applicant v/s charged off % w.r.t total applicants

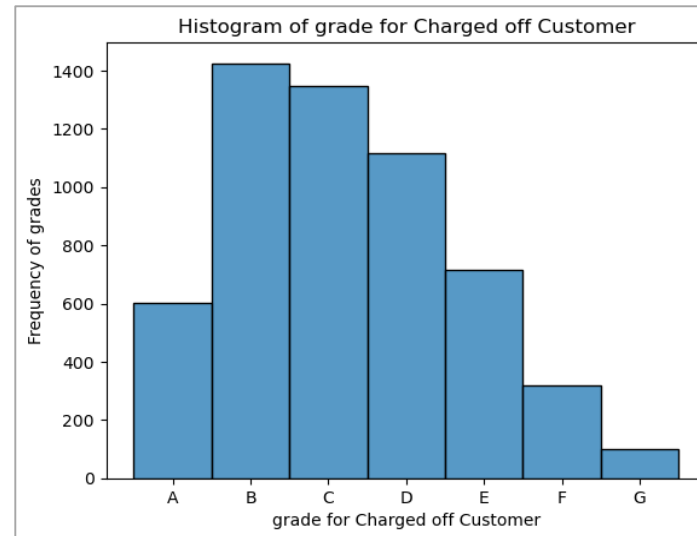


Chart 2 : Frequency of Charged off applicants w.r.t grade

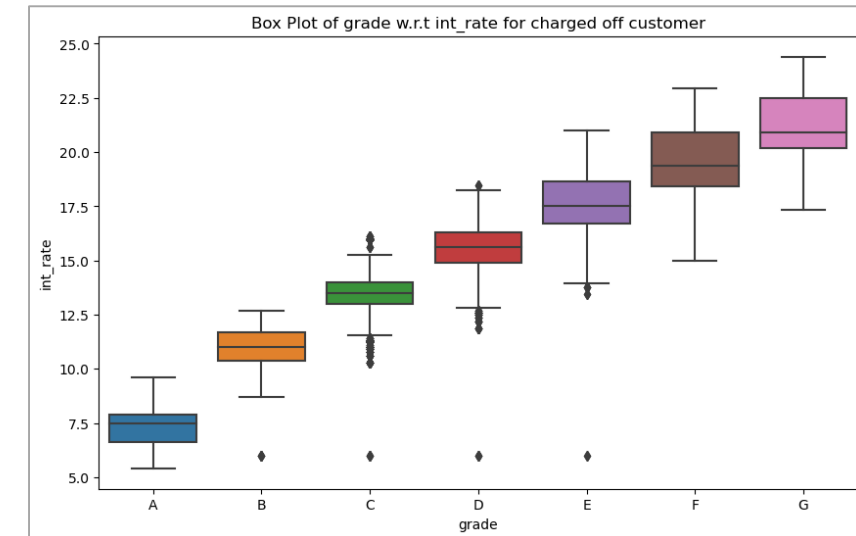


Chart 3: Box plot of Interest rate and the grade of the applicant

1. The higher the grade of the applicant higher the chance of being charged off
2. But the frequency of applicant is the most for grade B, and least for grade G.
3. Thus, we can say that the applicant with higher grades has been charged-off more, but the frequency of such cases are less.
4. The interest rate increases with the increasing grade.
5. There are few outliers too.

Feature : Loan purpose

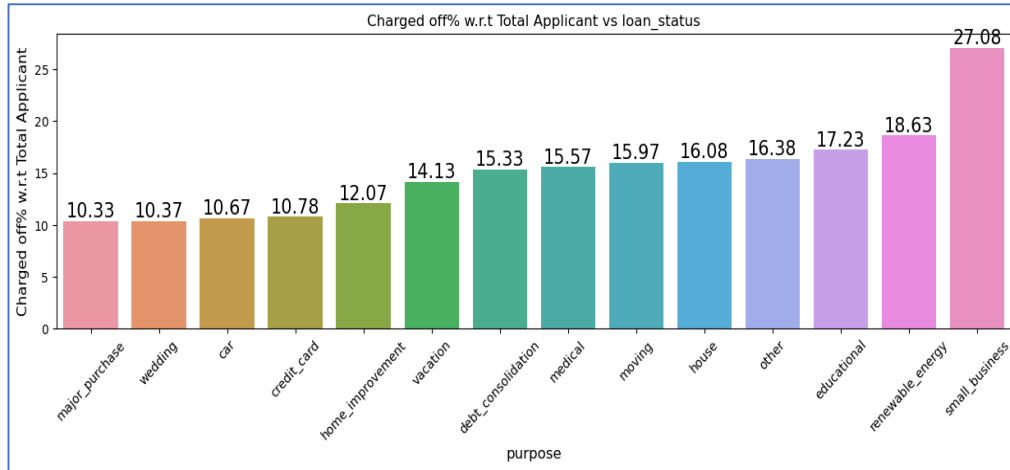


Chart 1: Purpose of loan v/s charged off % w.r.t purpose of the loan

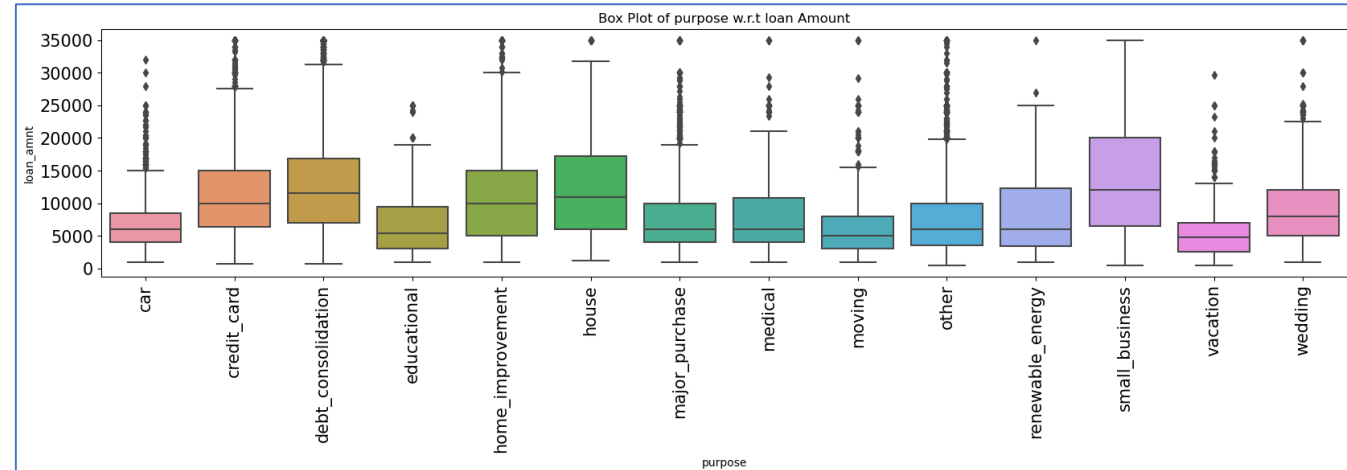


Chart 2: Purpose of loan v/s loan amount taken

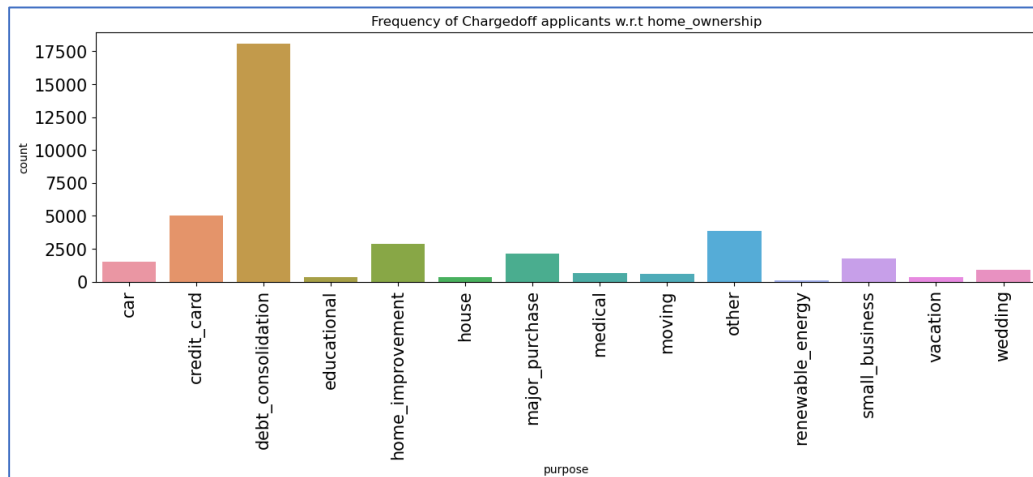


Chart 3: Frequency of applicants w.r.t purpose of the loan

1. The charged-off % is the most for small business
2. It will be risky to give loans for small businesses
3. The median of the interest rate for the 60-month term plan is higher than the 75th percentile of the 36-month plant
4. More people have opted for the 36-month plan thus it is the more obvious choice.

Feature : Region/State

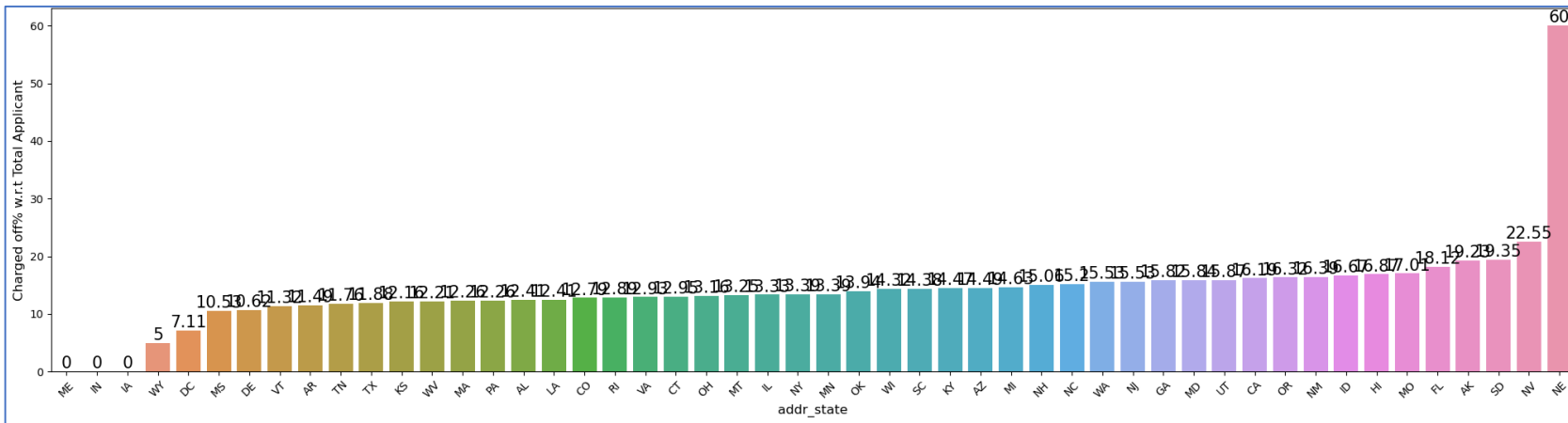


Chart 1: Purpose of loan v/s charged off % w.r.t address of the state

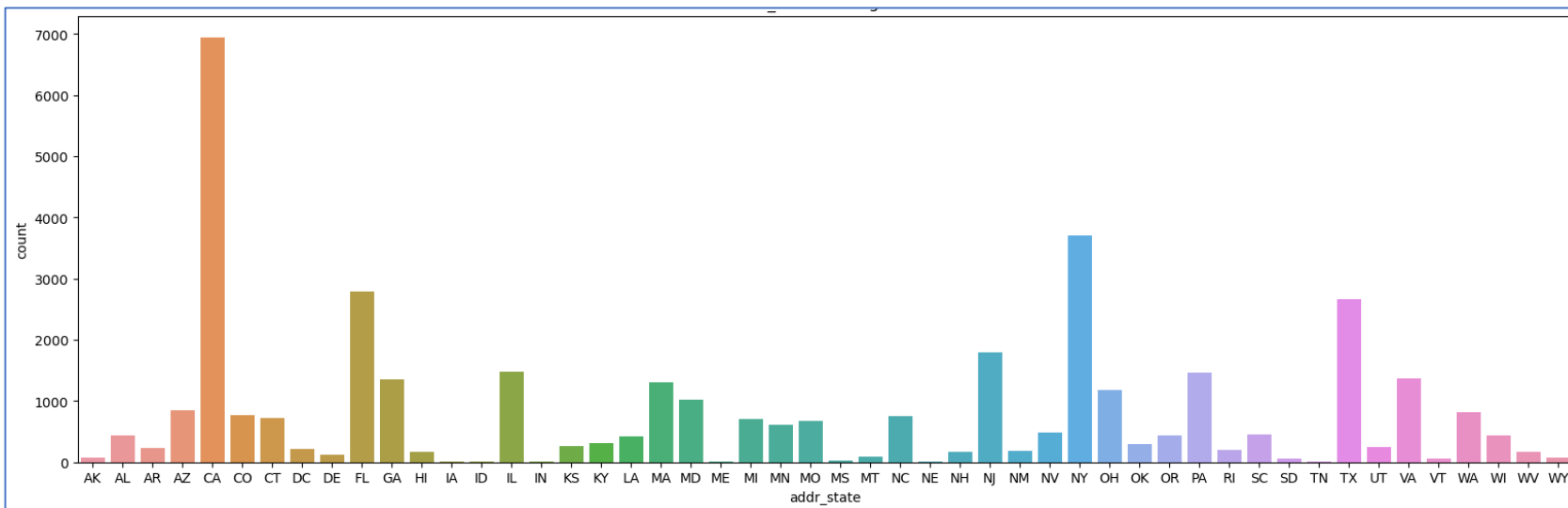


Chart 3: Frequency of applicants w.r.t address of the state

1. From the chart 1, we can see that the charged-off percentage is highest for the address state “NE”.
2. From chart 2, we can see that most applicants stay in the address state “CA”.
3. This will help us filter out the applicants who are likely to be charged off based on the region of residence.

Feature : Loan Home ownership

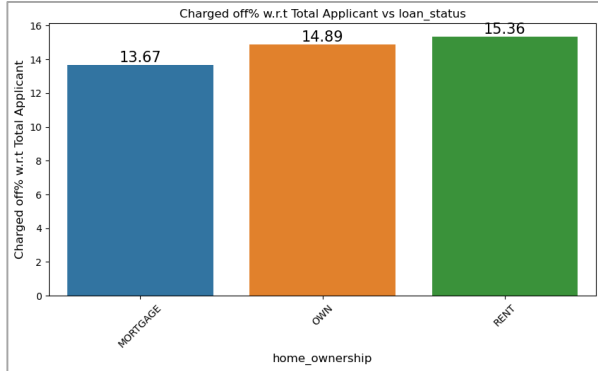


Chart 1: home ownership v/s charged off % w.r.t total applicants

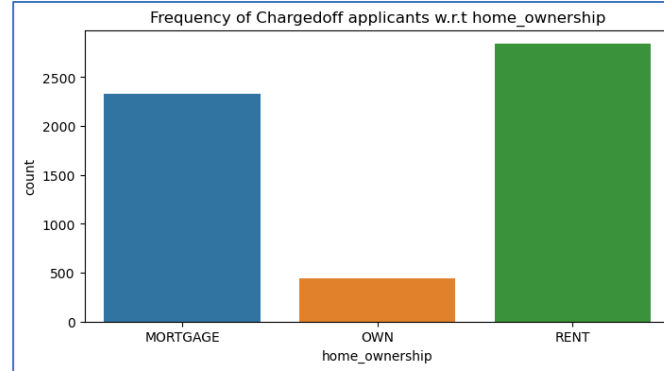
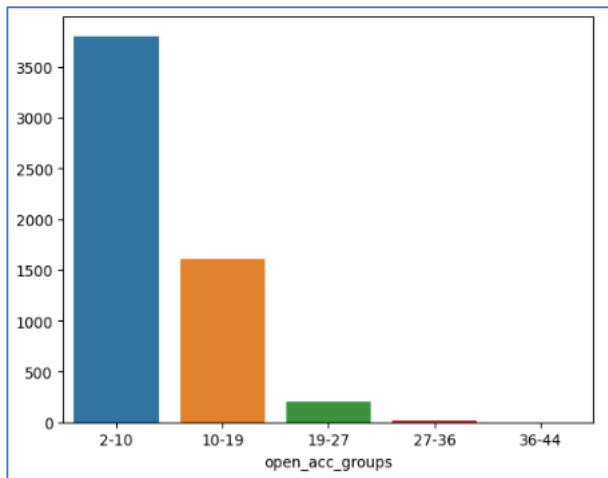


Chart 2: Frequency of applicants w.r.t type of homeownership

1. Applicants with the rented flat have higher charged off% w.r.t total applicants
2. Applicants prefer rented flat as it has the highest frequency

Feature : Open accounts



1. 2 -10 credit lines in the borrower's credit file.
2. Higher the number of open accounts lesser is the charged off count

Charged off Analysis on Year and Month

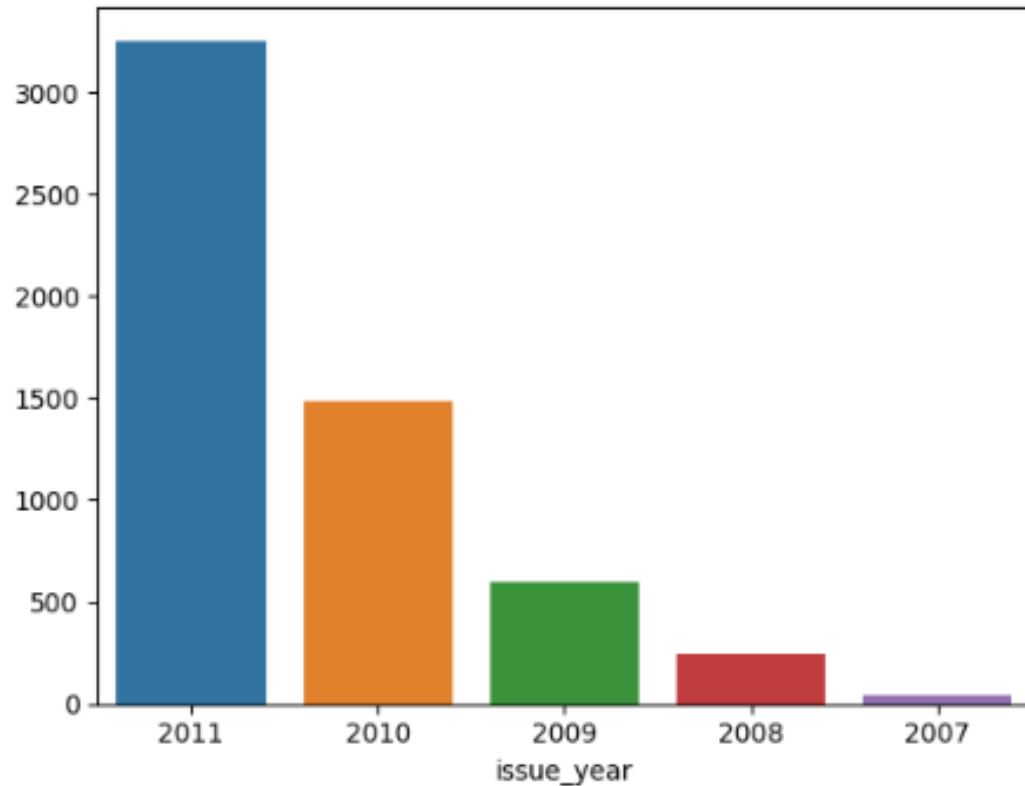


Chart 1: Charge off issue Year v/s charged off % w.r.t total applicants

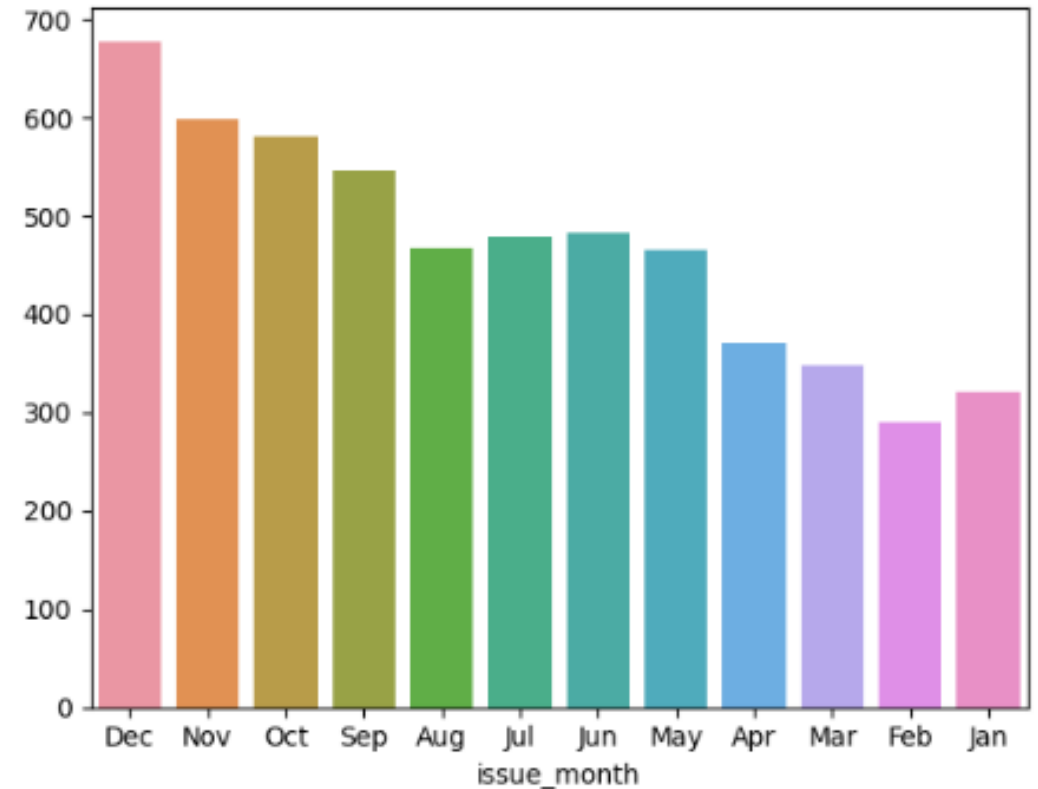


Chart 2: Charge off issue month v/s charged off % w.r.t total applicants

□ While performing Bivariate Analysis we observe :-

- Maximum no. of applicant charge off on specially year 2011
- Even analyze year 2011 charge off case we observe specifically Dec applicant charge off in huge number

Analysis of loan Term Distribution and Loan status Vs Interest rate

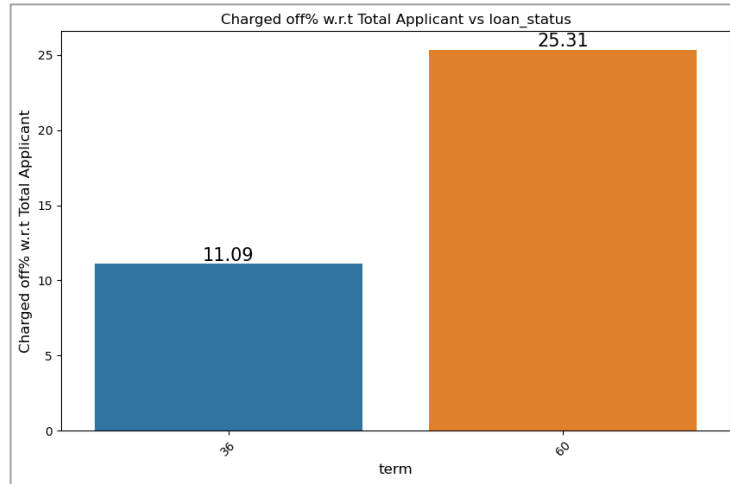


Chart 1: Loan re-payment term v/s charged off % w.r.t total applicants

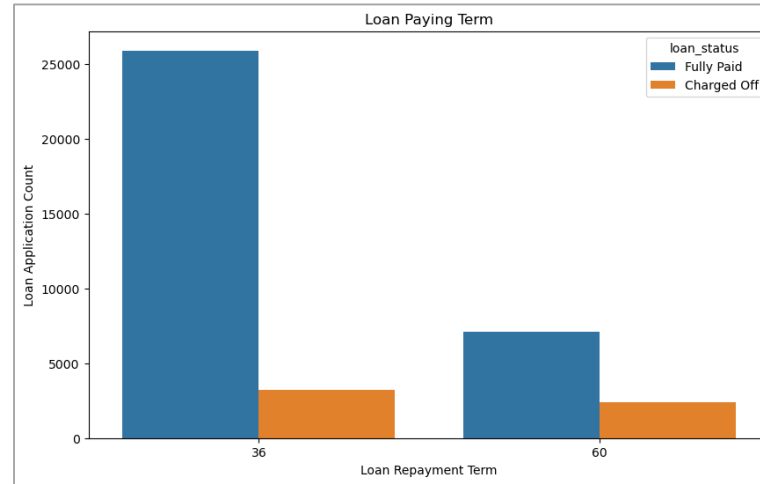


Chart 2 : Frequency of applicants w.r.t payment terms

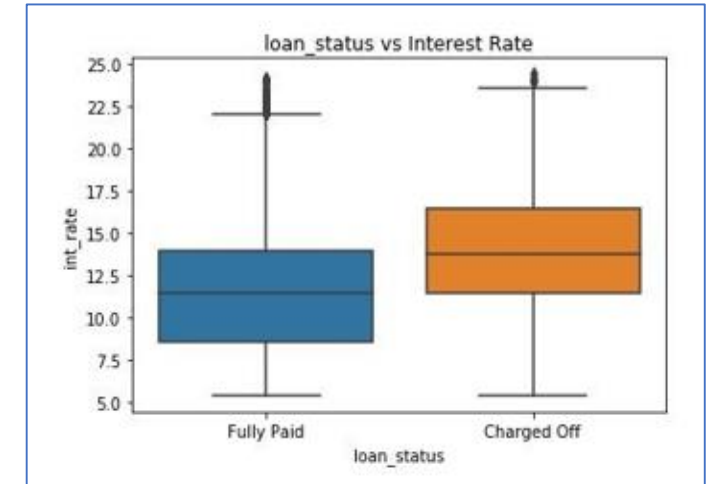


Chart 3 : Loan Status v/s Interest Rate % w.r.t total applicants

1. There are only two loan terms 36 and 60 months. Around 75% borrowers took loans with 36 months term.
2. The charged off borrowers are around 15% and fully paid is around 85% in the given data set
3. When the loan interest rate is high there is high chance of loan getting defaulted.
4. More people have opted for the 36-month plan thus it is the more obvious choice.

Univariate Analysis funded_amount and loan_amount

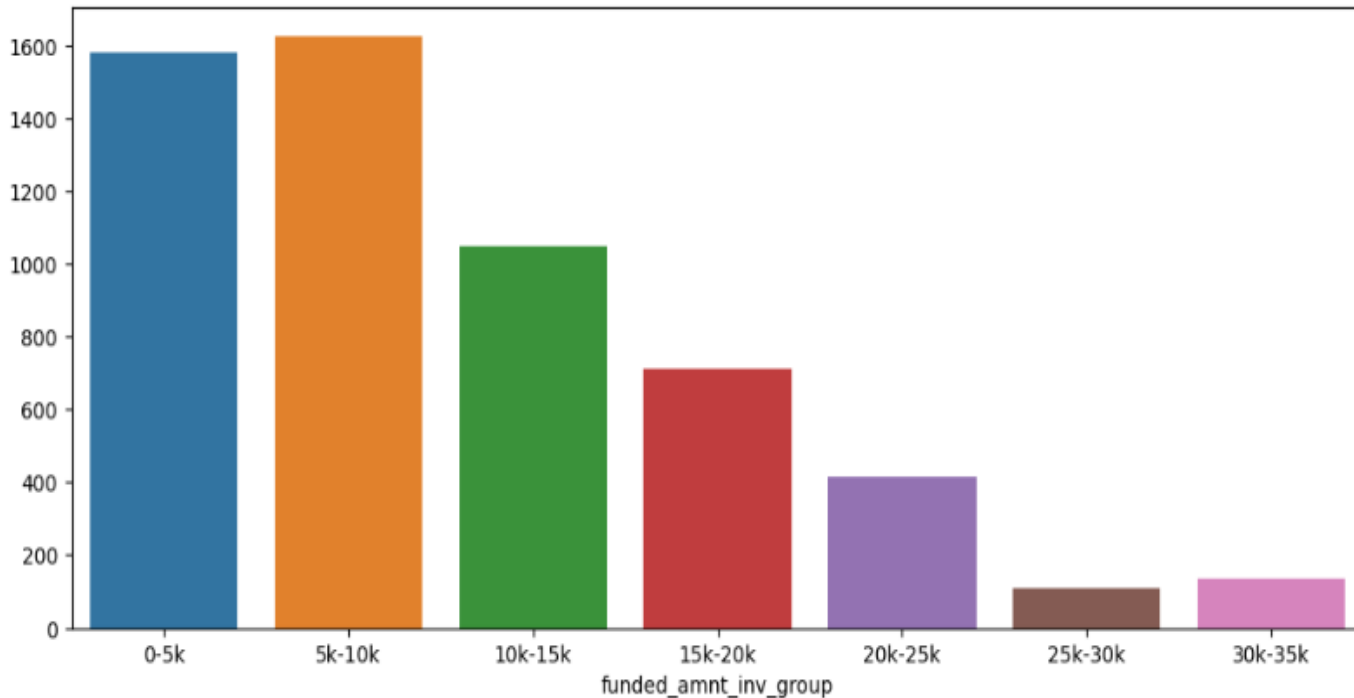


Chart 1: funded Amount v/s charged off % w.r.t total applicants

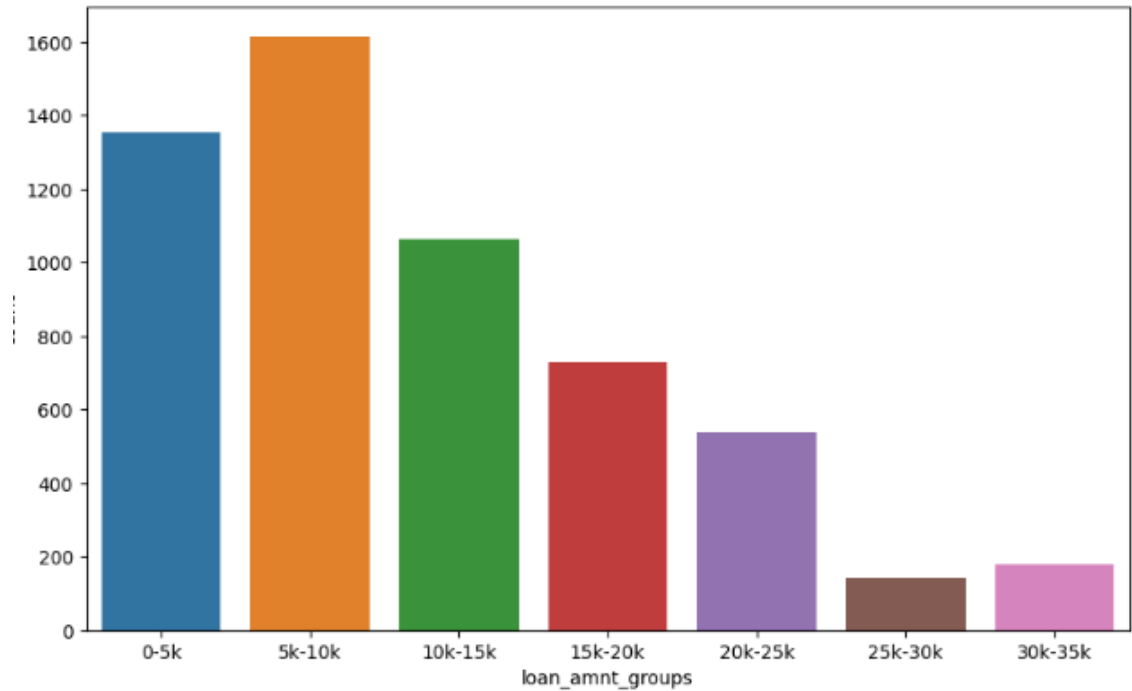


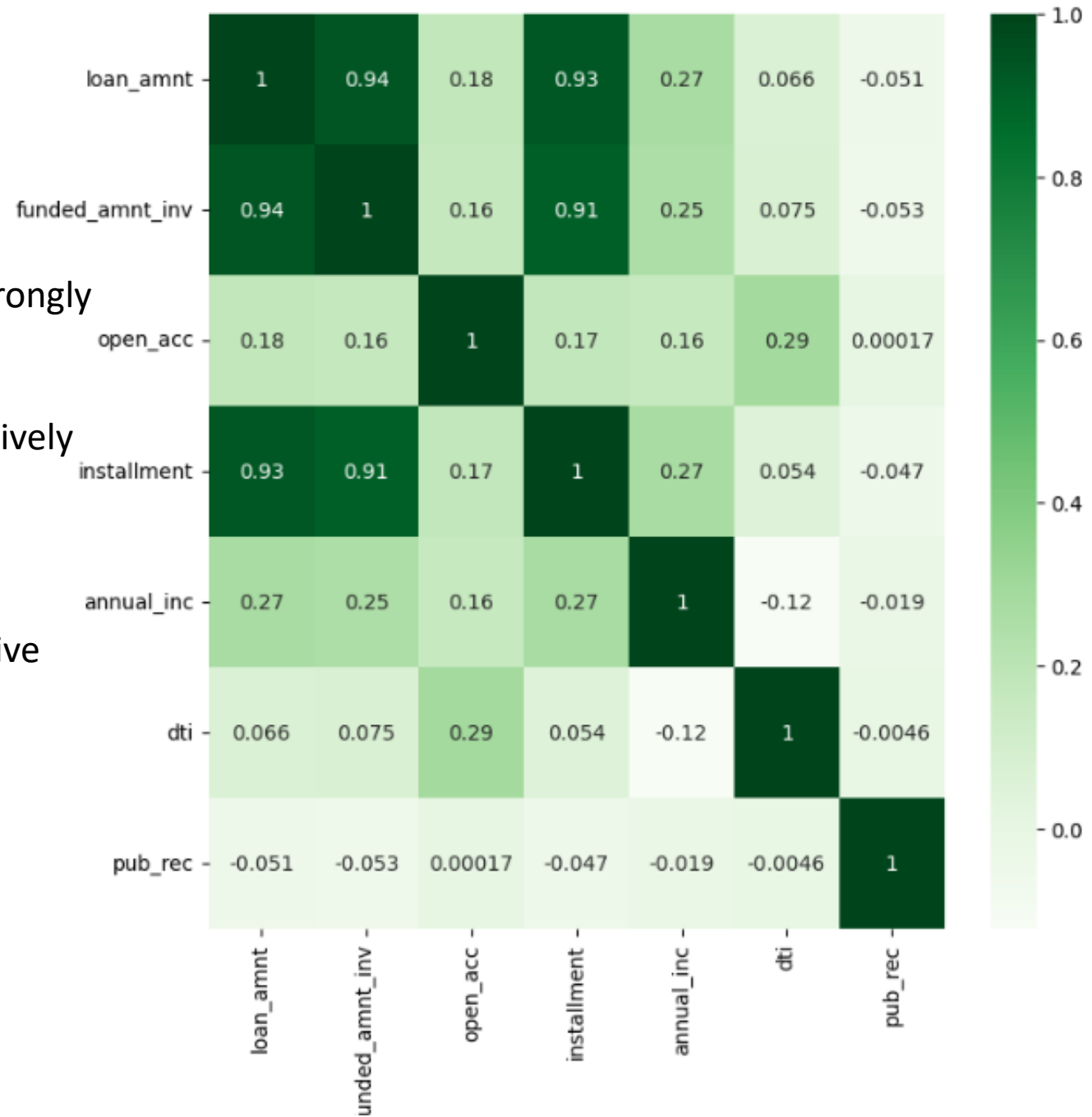
Chart 2: Loan Amount groups v/s charged off % w.r.t total applicants

- Observing charged off vs funded_amnt, less funded amount more chance to charged off by applicant
- Similarly trend is seen in Loan amount

Categorical Value Analysis

❑ Lets analyze data using categorical values :-

- Loan amount, investor amount, funding amount are strongly correlated.
- Annual income with DTI(Debt-to-income ratio) is negatively correlated.
- annual income is low DTI is high & vice versa.
- annual income and employment year correlation positive observe.
- That means income increases with work experience



Conclusions

Feature	Max Frequency	Min Frequency	High Risk	Low Risk
Purpose of loan	Debt Consolidation	Educational	Small Business	Major purchase
Utilization Range	60%-80%	0%-20%	>60%	<40%
Interest Rate	12.5%-17.5%	20%-23%	20%-25%	5%-10%
Grade of applicant	B	G	G	A
Loan Term	36 Months	60 Months	60 Months	36 Months
Region/State	CA	ME	NE	ME
Home Ownership	Rent	Own	Rent	Mortgage
Debt to Income(DTI)	12%-18%	24%-30%	18%-24%	0%-6%

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
- Lending club should reduce to pass the loan of higher utilization percentage applicant
- The charged-off % increases with increasing interest rates. Thus, we can say that the people with higher interest rates have been charged-off more.
- Grades are good metric for detecting defaulters. Lending club should examine more information from borrowers before issuing loans to Low grade (G to A).
- Lending Club should control their number of loan issues to borrowers who are from CA, FL, TX and NY to make profits.
- Small business loans are defaulted more. Lending club should stop/reduce issuing the loans to them.
- Borrowers with Rent home ownership are taking higher loans and defaulting the approved loans. Lending club should restricted number of loans to this category

Recommendations

- Lower credit utilization can be given loans at lower interest rates
- People with DTI in the range 0%-6% can be given the loan
- People in A & B grades are at low risk and have high numbers too
- Credit card and debt consolidation have relatively lower chances of being charged off and are in high frequency, Thus can be targeted.