

Lending Club Case Study

Gaurav Dev Khera

AI/ML Batch - C70

2024-25

<https://github.com/gkhera23/LendingClubCaseStudy>

Table of Contents

[Objective of case study](#)

[Data-set used](#)

[Data Clean-up steps](#)

[Initial insights on loan distributions and status](#)

[Univariate Analysis](#)

[Univariate Analysis on Ordered Categorical Data](#)

[Loan defaults across Grades](#)

[Loan defaults across different Sub-grades](#)

[Distribution of Sub-grades across Grades](#)

[Loan defaults across different terms](#)

[Loan defaults across different employment lengths](#)

[Loan defaults across disbursement year](#)

[Loan defaults across disbursement month](#)

[Analysis from graphs drawn on Ordered Categorical data](#)

[Univariate Analysis on Unordered Categorical Data](#)

[Loan defaults across States](#)

[Loan defaults across Loan Purposes](#)

[Loan defaults across Home Ownership](#)

[Analysis from graphs drawn on Unordered Categorical data](#)

[Univariate Analysis on Quantitative Categorical Data](#)

[Loan defaults across Interest Rate ranges](#)

[Loan defaults across Annual Income Groups](#)

[Loan defaults across Debt-to-Income ratio groups](#)

[Loan defaults across Loan Amount ranges](#)

[Loan defaults across Funded Amount groups](#)

[Analysis from graphs drawn on Quantitative Categorical data groups](#)

[Bivariate and Multivariate Analysis](#)

[Percentage of Loan defaults across Grades](#)

[Percentage of Loan defaults across Purposes](#)

[Percentage of Loan defaults across States](#)

[Percentage of Loan defaults across Income Groups](#)

[Percentage of Loan defaults across DTI ranges](#)

[Percentage of Loan defaults across Employment Length](#)

[Percentage of Loan defaults across Home Ownership](#)

[Percentage of Loan defaults across Loan Amount ranges](#)

[Percentage of Loan defaults across Terms](#)

[Percentage of Loan defaults across Loan Disbursement Year](#)

[Percentage of Loan defaults across Interest Rate ranges](#)

[Conclusions from graphs of Percentage of Loan Defaults vs different factors](#)

[Distribution of Loan defaults across Loan Term](#)

[Distribution of Loan defaults across Loan Term](#)

[Distribution of Loan defaults over Annual Income and Loan Purpose](#)

[Distribution of Loan defaults over Annual Income and Home Ownership](#)

[Distribution of Loan defaults across Annual Income and Interest Rate ranges](#)

[Distribution of Loan defaults across Loan Amounts and Interest Rates](#)

[Distribution of Loan defaults across Loan Amounts and Loan Purpose](#)

[Distribution of Loan defaults across Home Ownerships and Loan Amounts](#)

[Distribution of Loan defaults across Loan Amounts and Disbursement Months](#)

[Distribution of Loan defaults across Loan Amounts and Disbursement Months](#)

[Distribution of Loan defaults across Loan Amounts and Grade](#)

[Distribution of Loan defaults across Interest Rate and Grade](#)

[Distribution of Loan defaults on Loan Amts, Emp Length and Verification Status](#)

[Patterns and inferences drawn from above graphs](#)

[Recommendations for loan approval](#)

Objective of case study

Dataset of last 5 years is given for a consumer finance company, which specializes 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

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

❖ The business objective is to take a decision on a loan application whether to reject or approve based on certain variables.

Data-set used

- A file loan.csv is provided which contains historical data of company.
- It contains information about past credit history of customers.
- Data contains information about past loan applicants and whether they 'defaulted' or not.
- Data is provided for 39717 customers distributed over 5 years
- Dataset contains the information for 111 fields, which contain fields like loan_amnt, term, int_rate, grade, sub_grade, emp_length, home_ownership, annual_inc, verification_status, loan_status, purpose, addr_state, dti and many more.

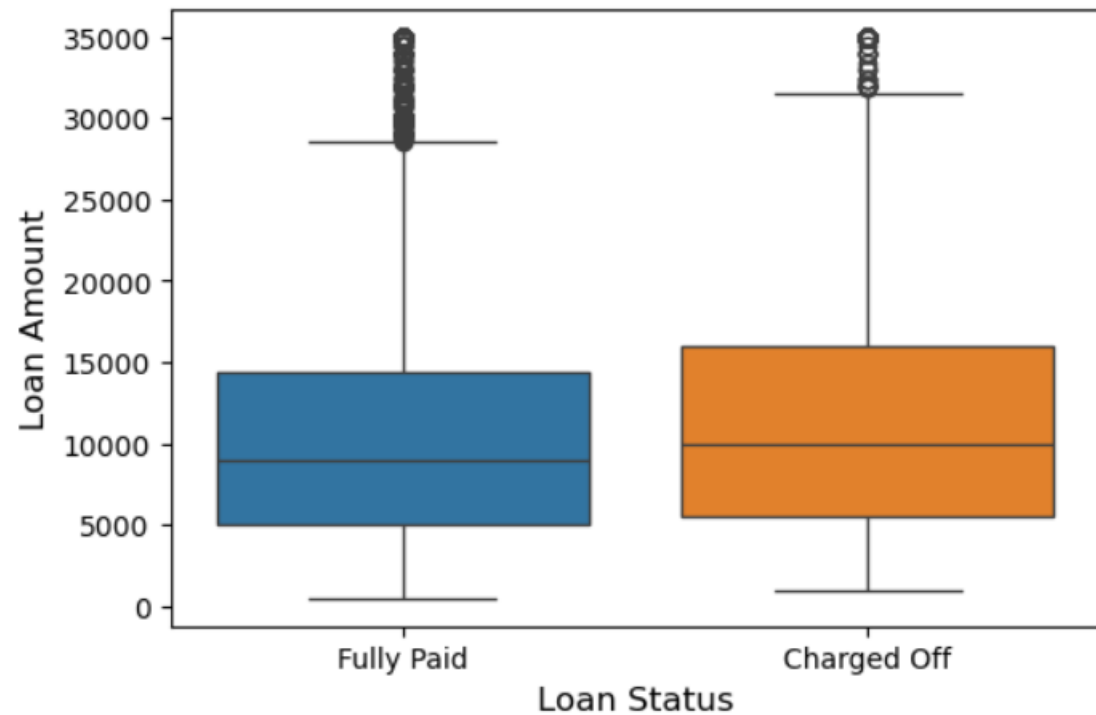
Data Clean-up steps

Following data clean-up steps were taken:

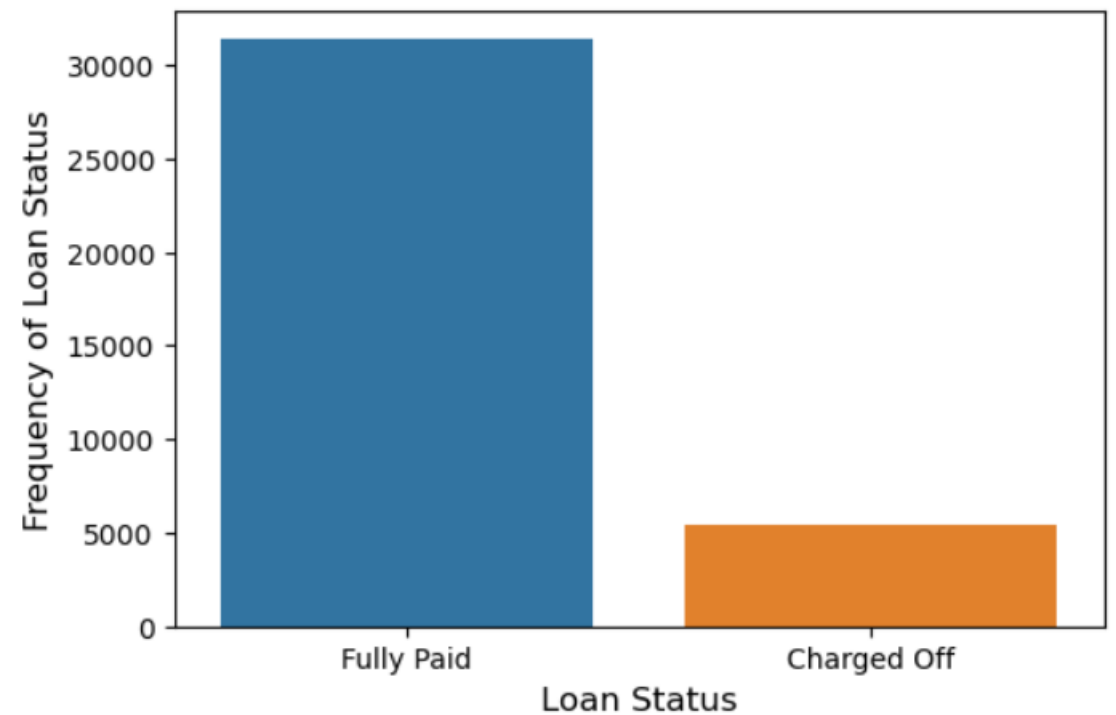
- Import the Data
- Remove Null value columns
- Remove duplicate data
- Remove columns which are not useful for analysis
- Fix Null values in columns
- Correct data types of columns
- Remove outliers using visualizations

Initial insights on loan distributions and status

Loan Status wise distribution of Loan Amount



Distribution of Loan Status



Insights:

- ❑ The distribution of loan amounts is almost similar in both loan statuses - 'Fully Paid' and 'Charged Off'
- ❑ Most of the loans have been fully paid as compared to the loans which were charged off

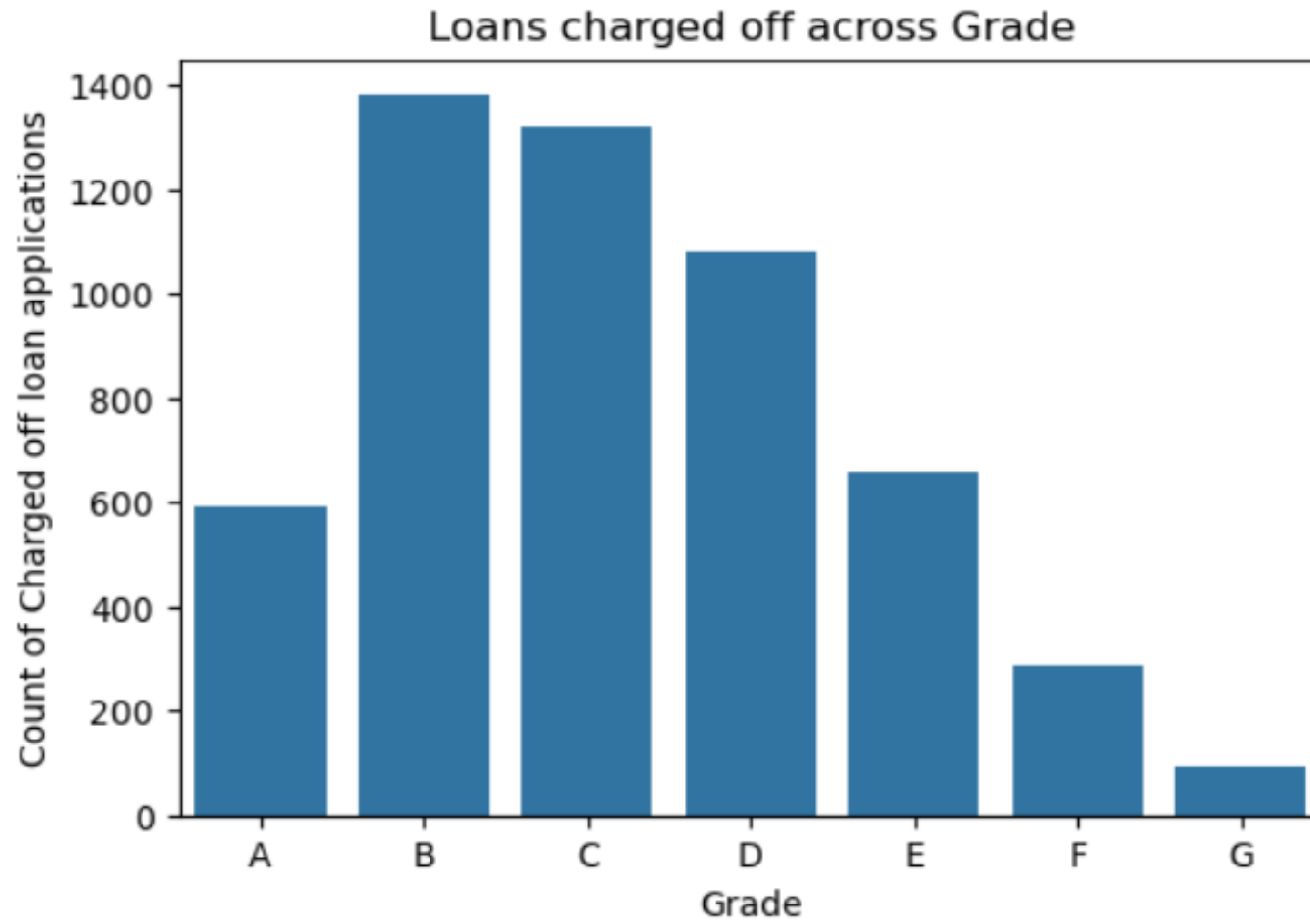
Univariate Analysis

Univariate Analysis on Ordered Categorical Data

Following columns were analyzed for ordered categorical data, to see the impact on defaulted loans:

1. grade - A, B, C, D, E, F, G
2. sub_grade - Sub grades 1, 2, 3, 4 , 5 on each grade
3. term - 36 or 60 months
4. emp_length - Employment length in 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 years
5. issue_y - Year in which loan was issued
6. issue_m - Month in which loan was issued

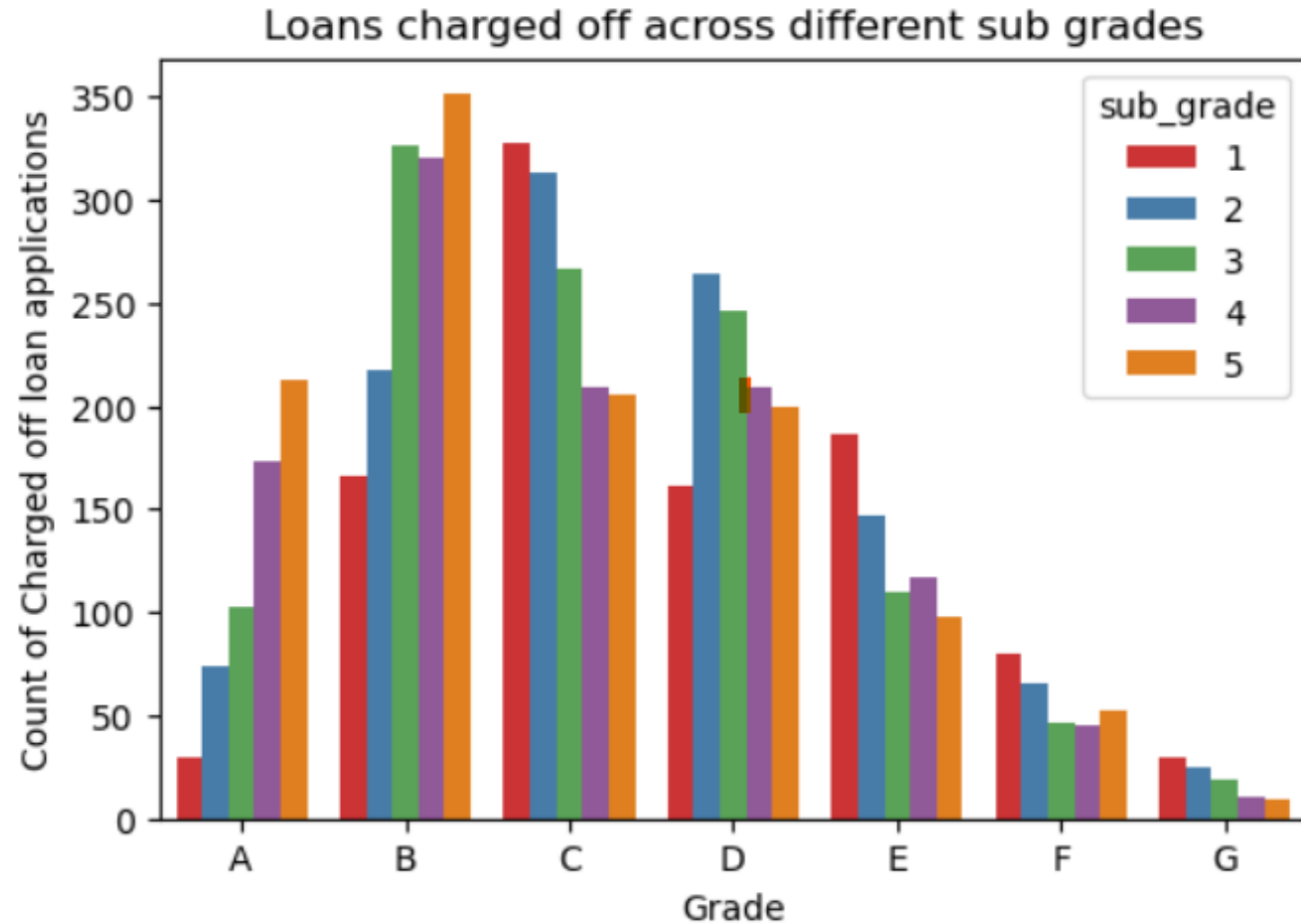
Loan defaults across Grades



Insights:

- Maximum loans given for grades B have defaulted followed by grade C.

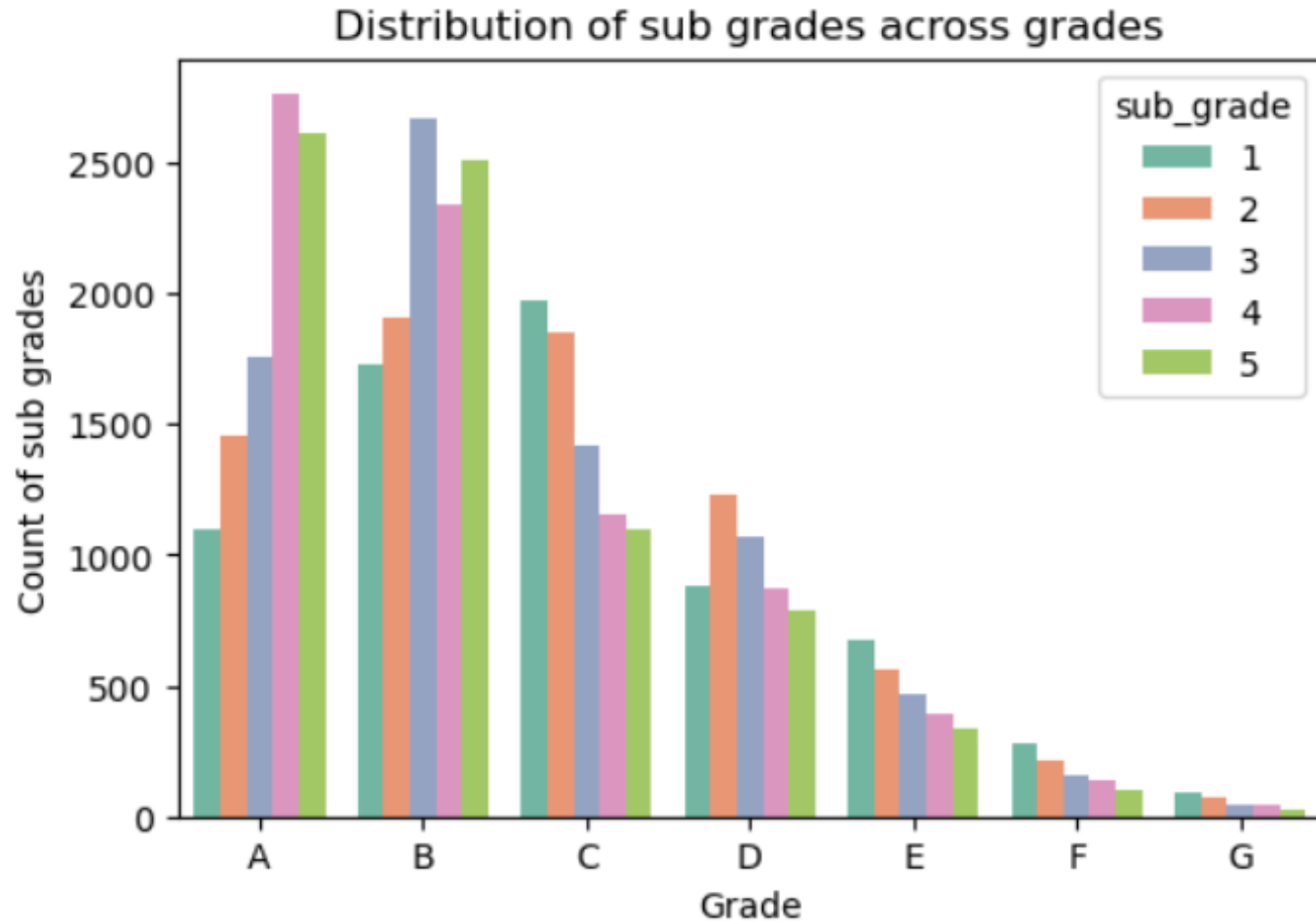
Loan defaults across different Sub-grades



Insights:

- Loans given for sub-grades B3, B4, B5 and C1 have maximum defaults.

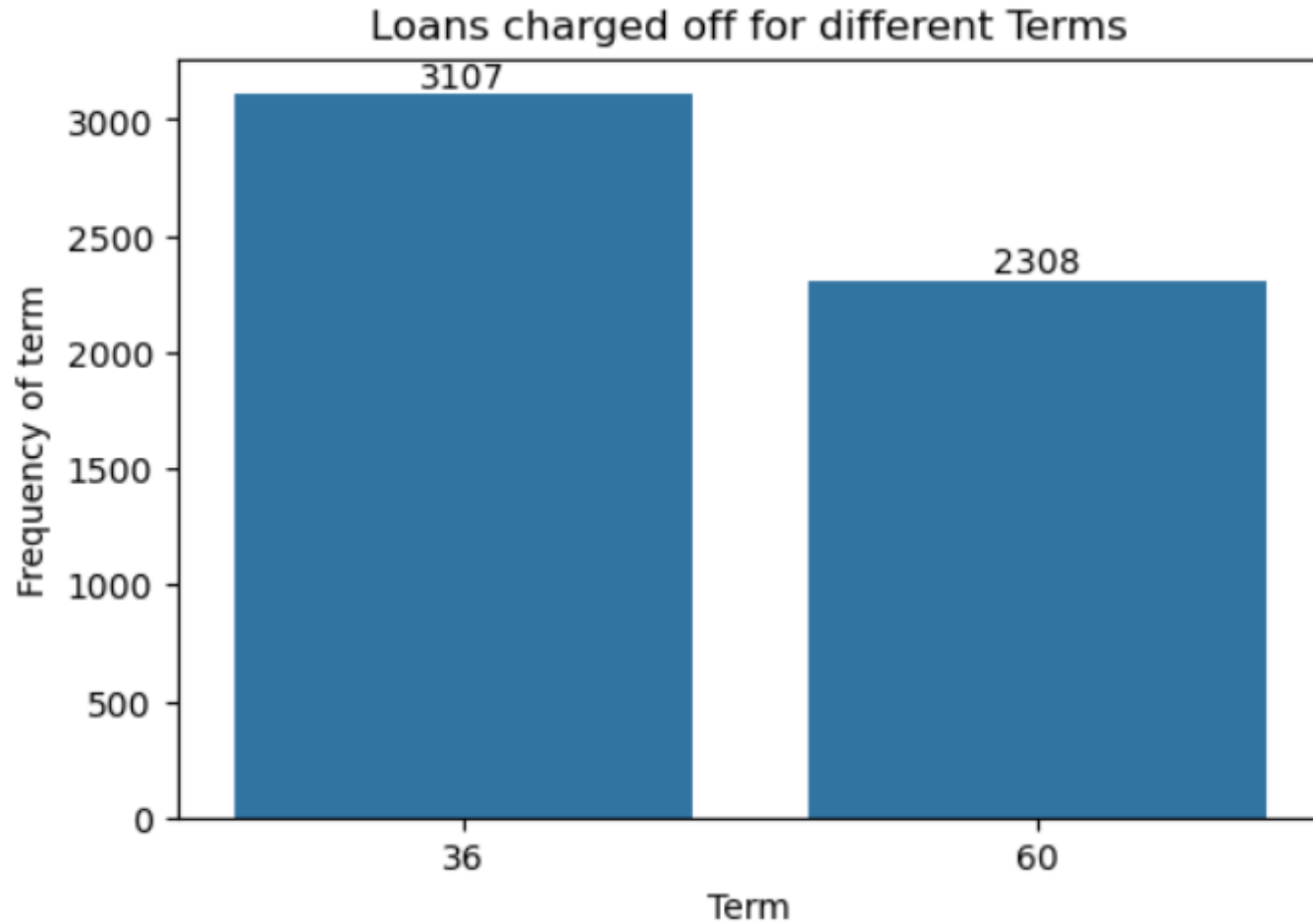
Distribution of Sub-grades across Grades



Insights:

- Sub-grades A and B have maximum number of loans given.
- A4 and B3 have maximum loans across the total.

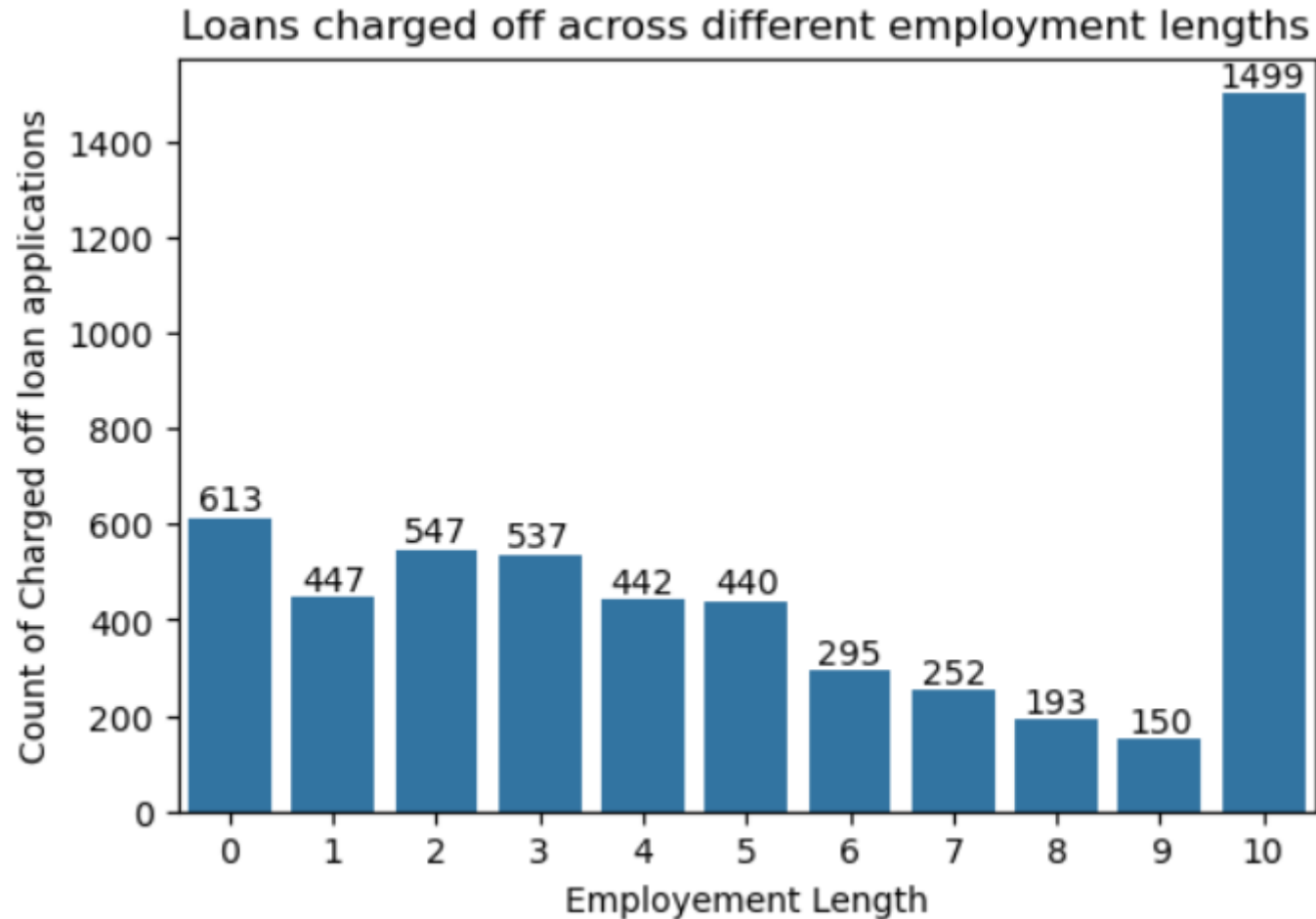
Loan defaults across different terms



Insights:

- Loans given for 36 months have defaulted approx. 25% more than the loans given for 60 months.
- This indicates that the loans given for shorter period might not be secure.

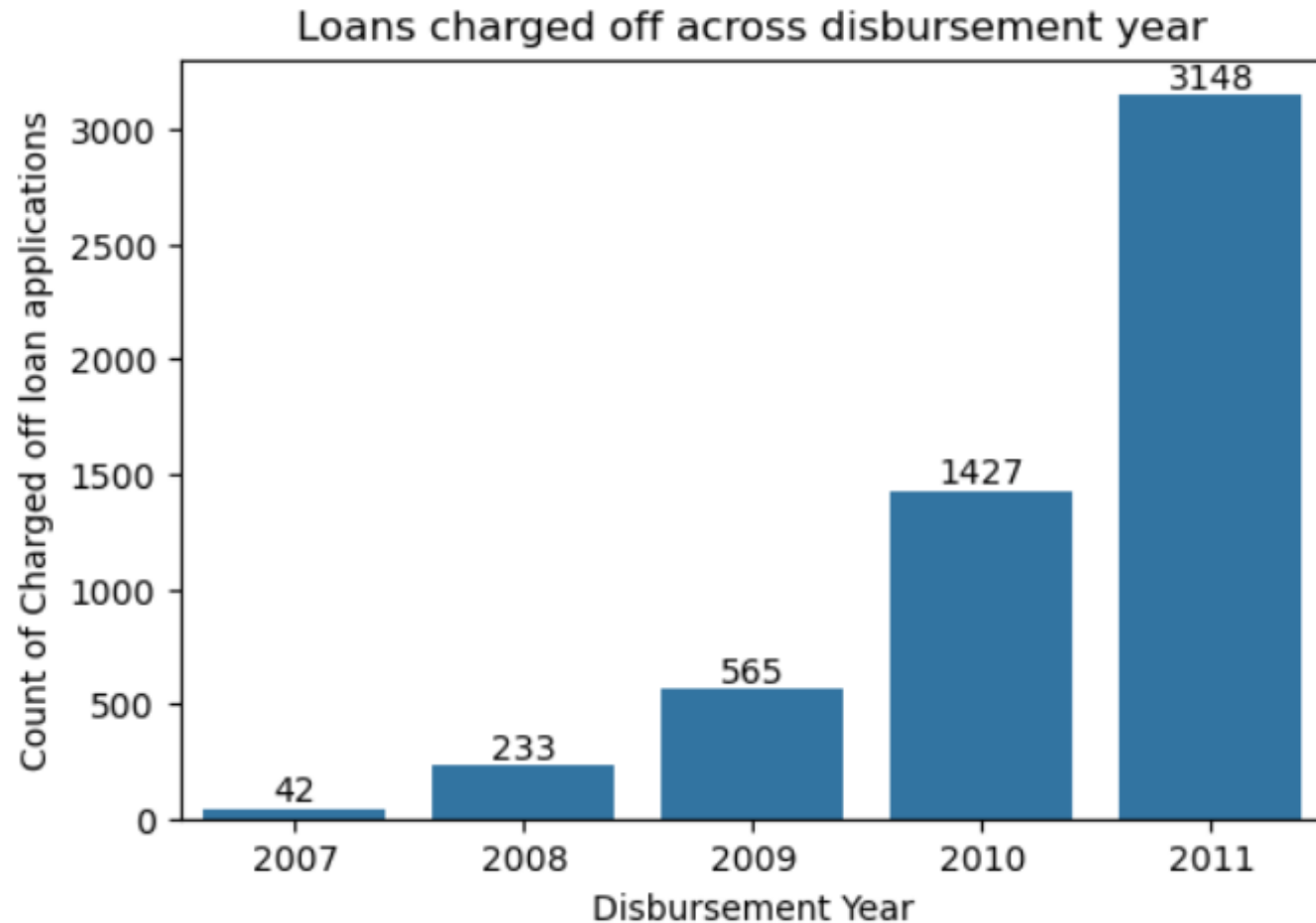
Loan defaults across different employment lengths



Insights:

- Maximum number of loan defaults are by borrowers who have 10 or more years of work experience.
- It shows that the loans given to people who have been in jobs for longer period are still not secure.

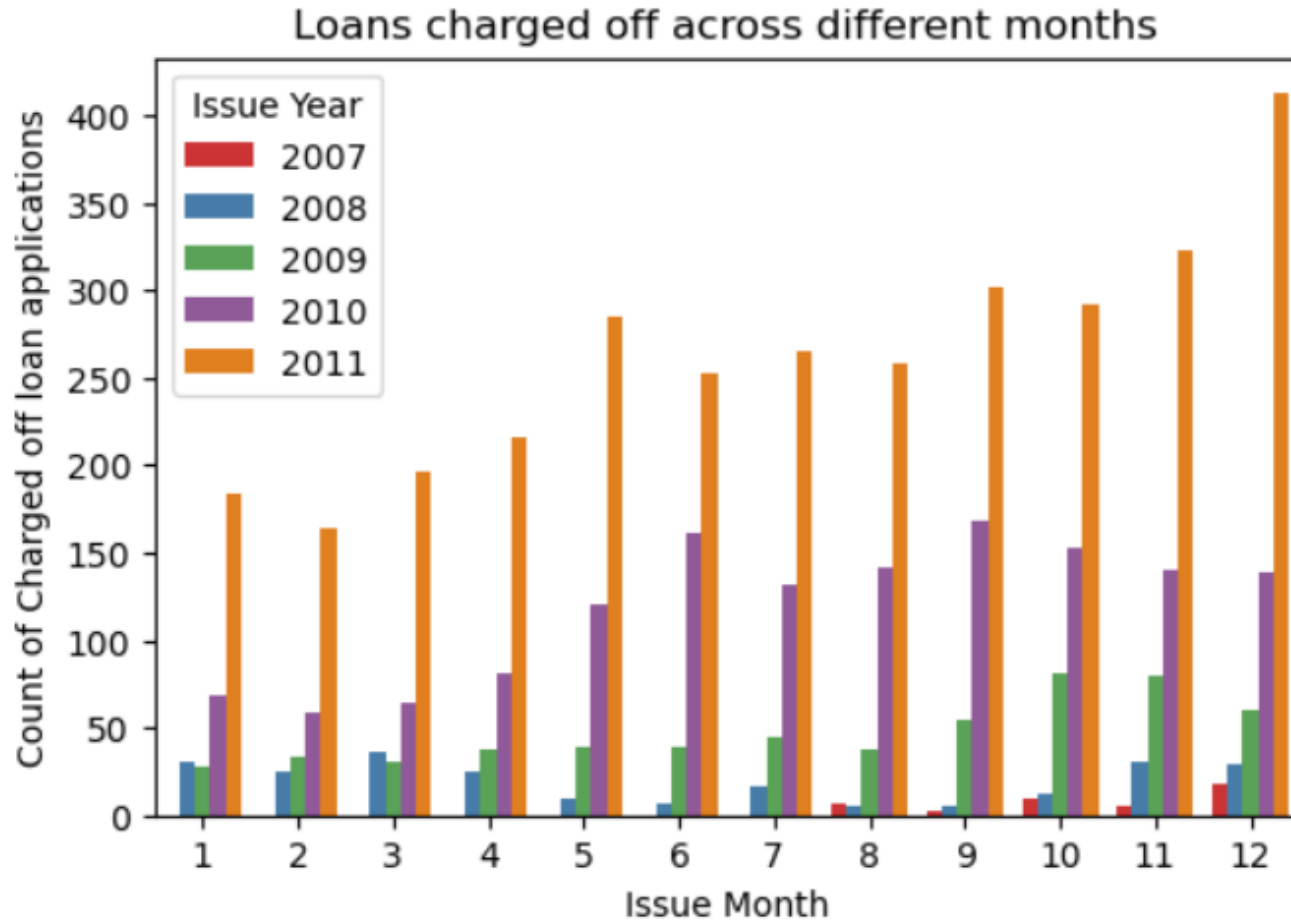
Loan defaults across disbursement year



Insights:

- Number of maximum charged-off loans were given in year 2011.
- This might not have much significance in analysis though, as we don't have any supporting factors, and this could be because of some external factors, like bank policies, change in government policies, overall economic crisis etc.

Loan defaults across disbursement month



Insights:

- Number of maximum charged-off loans were given in December month.
- This factor might again not have much significance in analysis, as we don't have any supporting factors.

Analysis from graphs drawn on Ordered Categorical data

Analysis from the graphs based on Grade, Sub Grade, Term, Employment Length, Year and Month:

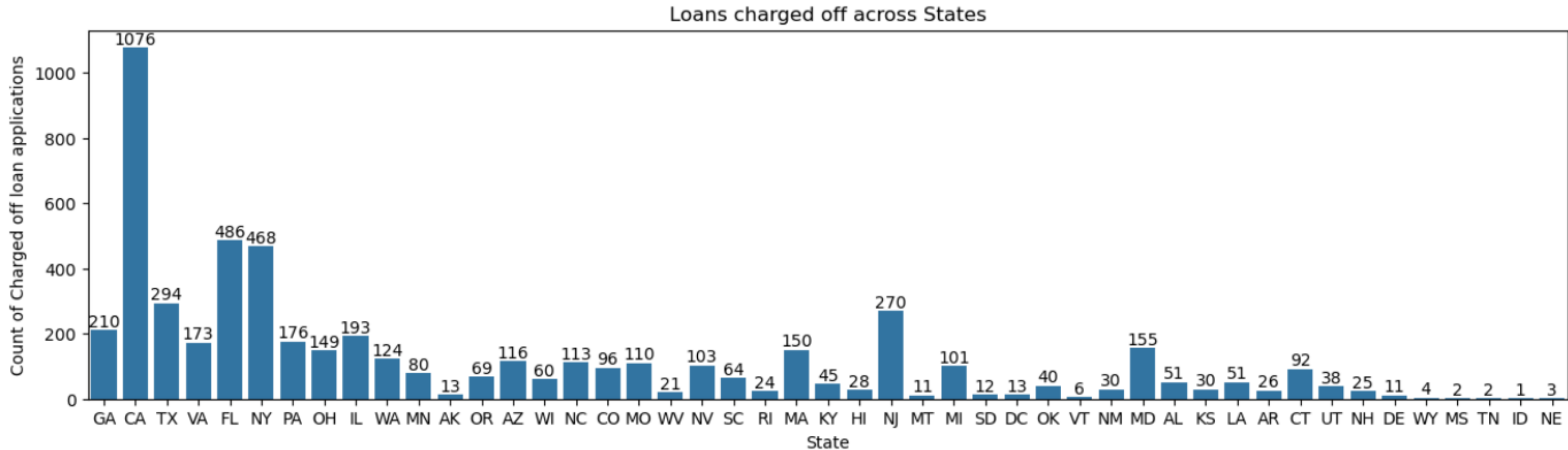
1. Grades B and C have the highest number of 'Charged off' loan applications. Grade D also has a good number of loans 'Charged off'. This indicates that loan applications with B, C and D grade are more likely to default.
2. Borrowers from sub grades B3, B4, B5, C1 and C2 have maximum tendency to default.
3. 57.3% loans which defaulted were of 36 months duration. This indicates that 60 months loan period should be a preference for bank.
4. Applicants employed for more than 10 years defaulted the maximum number of loans. Their ratio was more than double than second highest category of less than 1 year. This shows that long-term employment is not a guarantee for successful loan recovery.
5. Maximum numbers of loans which were defaulted were issued in 2011. This could be because of some external factors or internal technical glitch or for some change of policy by government or bank.

Univariate Analysis on Unordered Categorical Data

Following columns were analyzed for ordered categorical data, to see the impact on defaulted loans:

1. `addr_state` - The state provided by the borrower in the loan application.
2. `purpose` - A category provided by the borrower for the loan request.
3. `home_ownership` - The home ownership status provided by the borrower during registration. Values are: RENT, OWN, MORTGAGE, OTHER.

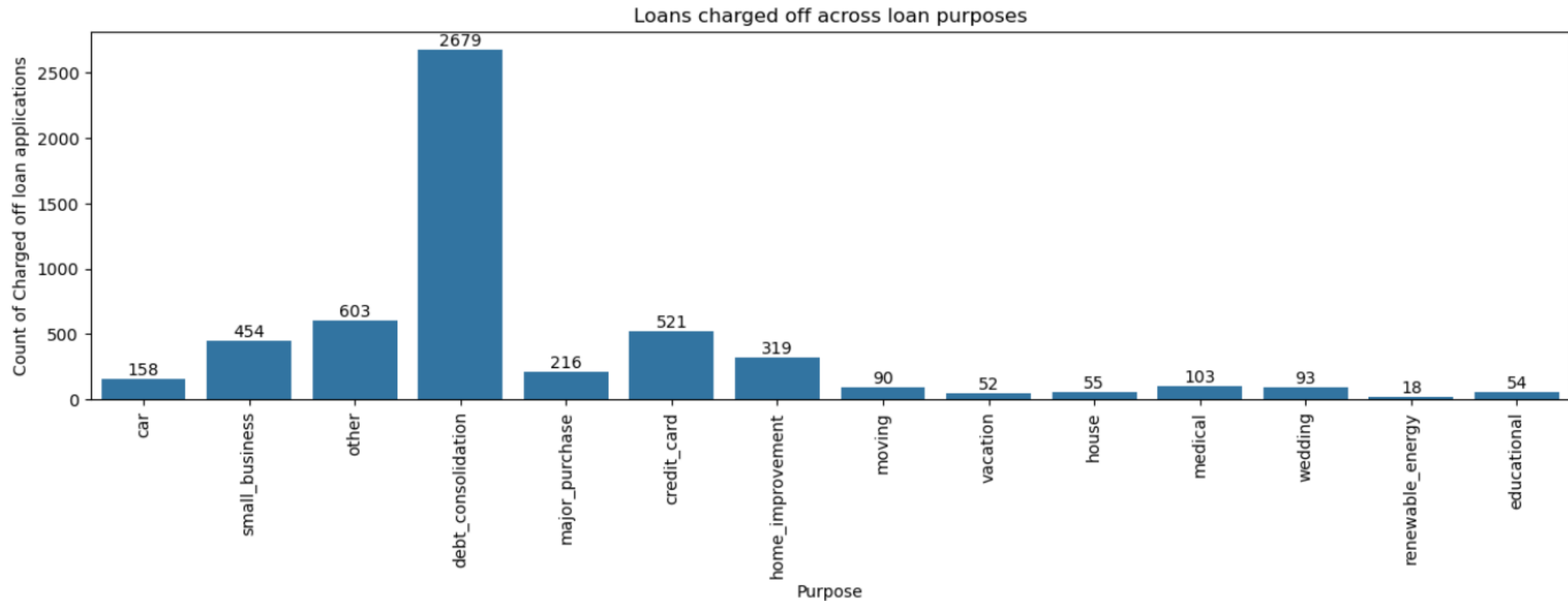
Loan defaults across States



Insights:

- Loans given in states CA, FL and NY have maximum defaults.
- There are few states which have very less loan defaults, probably because of low number of disbursements.

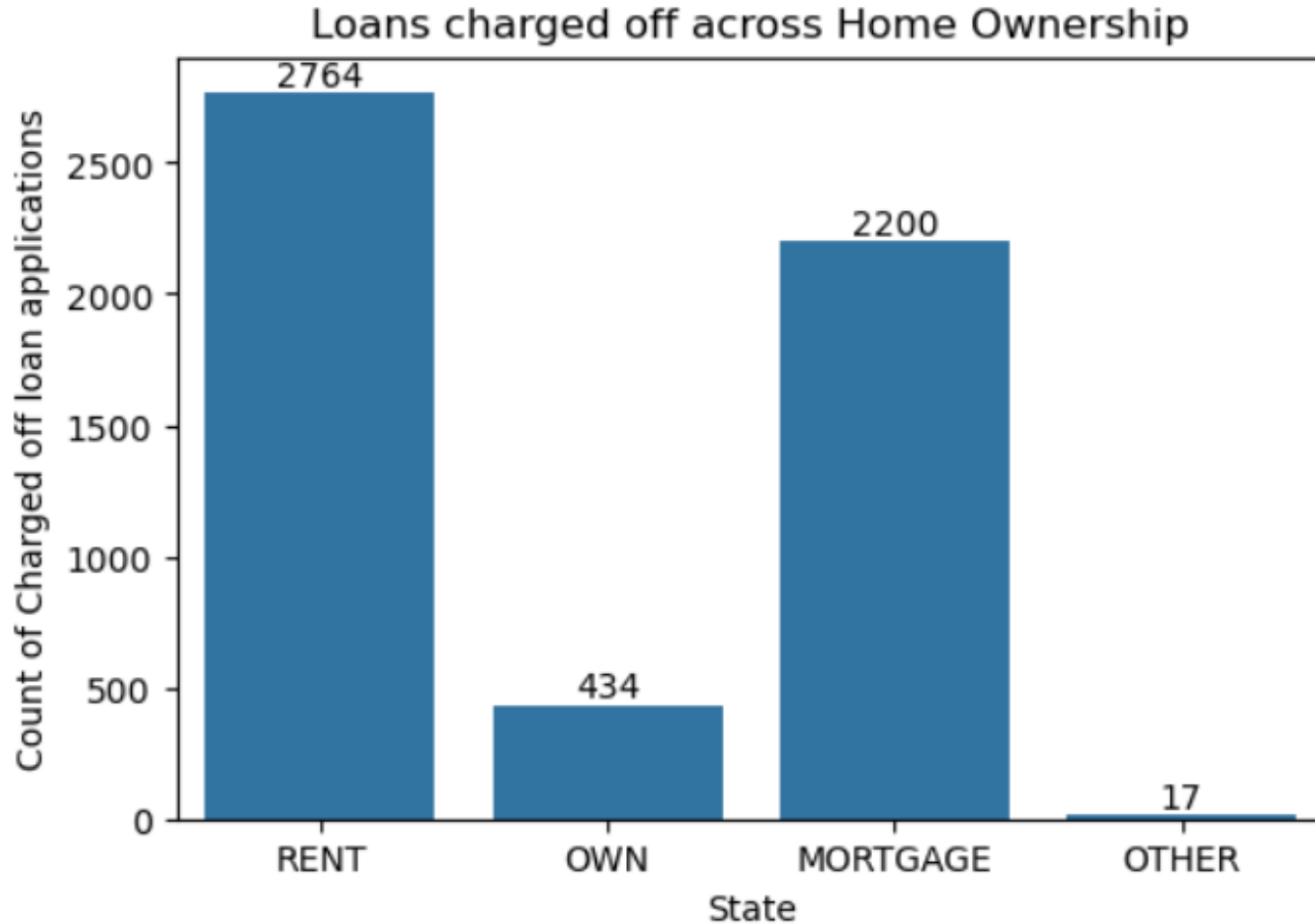
Loan defaults across Loan Purposes



Insights:

- Loans given for purpose of 'Debt Consolidation' have maximum defaults and multi times higher than other purposes.

Loan defaults across Home Ownership



Insights:

- Borrowers living on rent or already have their property mortgaged are more likely to default the loans.

Analysis from graphs drawn on Unordered Categorical data

Analysis from the graphs based on State, Loan Purpose and Home Ownership:

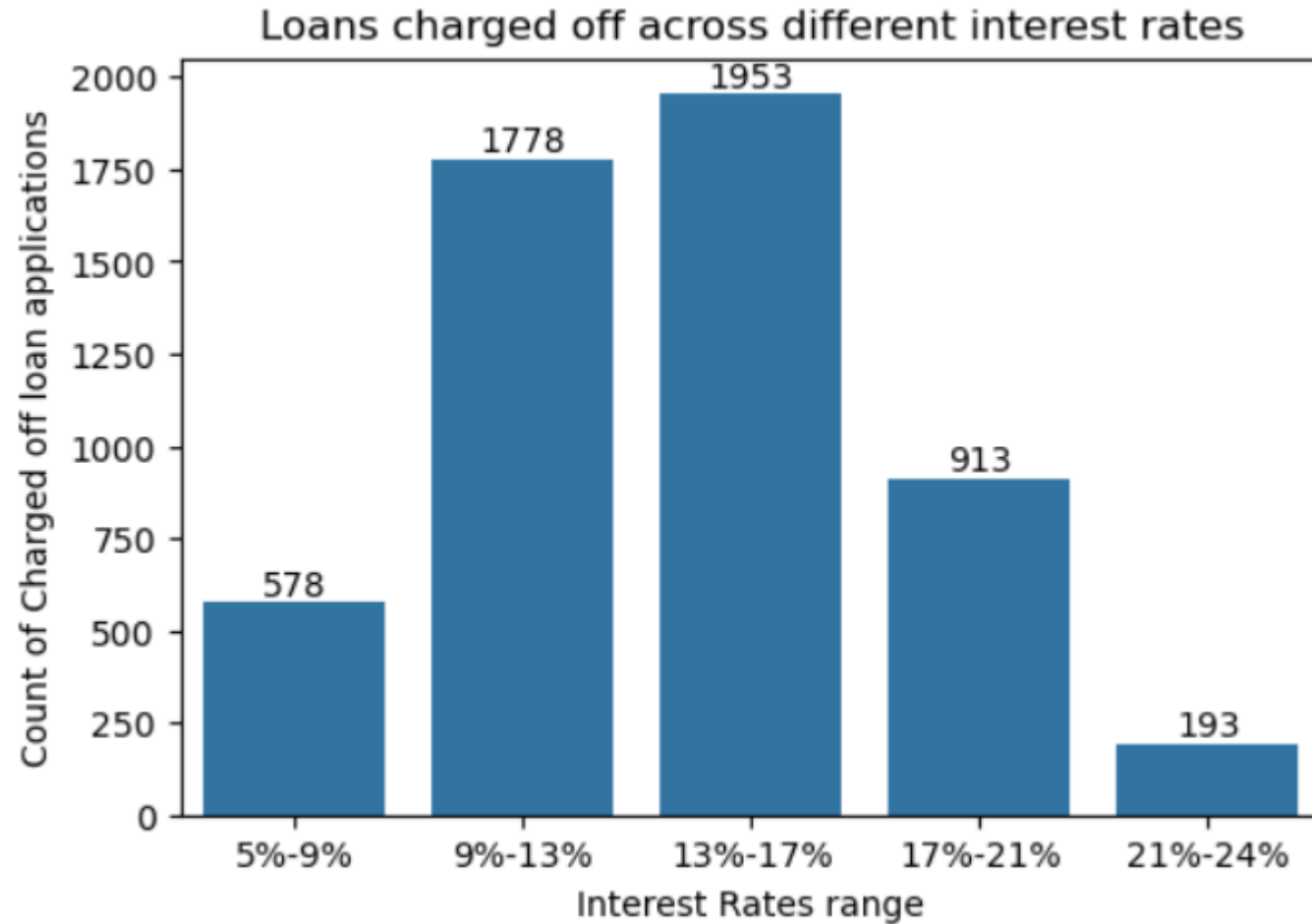
1. Borrowers from states CA, FL, NY are most likely to default the loan.
2. Loans taken for 'Debt Consolidation' are most likely to default.
3. Maximum applicants who defaulted were living either in rented houses or mortgaged ones.

Univariate Analysis on Quantitative Categorical Data

Impact of following fields on defaulted loans was analyzed based on different groups (bins) of these fields

1. `int_rate` - Interest Rate on the loan
2. `annual_inc` - The self-reported annual income provided by the borrower during registration.
3. `dti` - A ratio calculated using the borrower's total monthly debt payments on the total debt obligations divided by the borrower's self-reported monthly income.
4. `loan_amnt` - The listed amount of the loan applied for by the borrower.
5. `funded_amnt` - The total amount committed to that loan at that point in time.

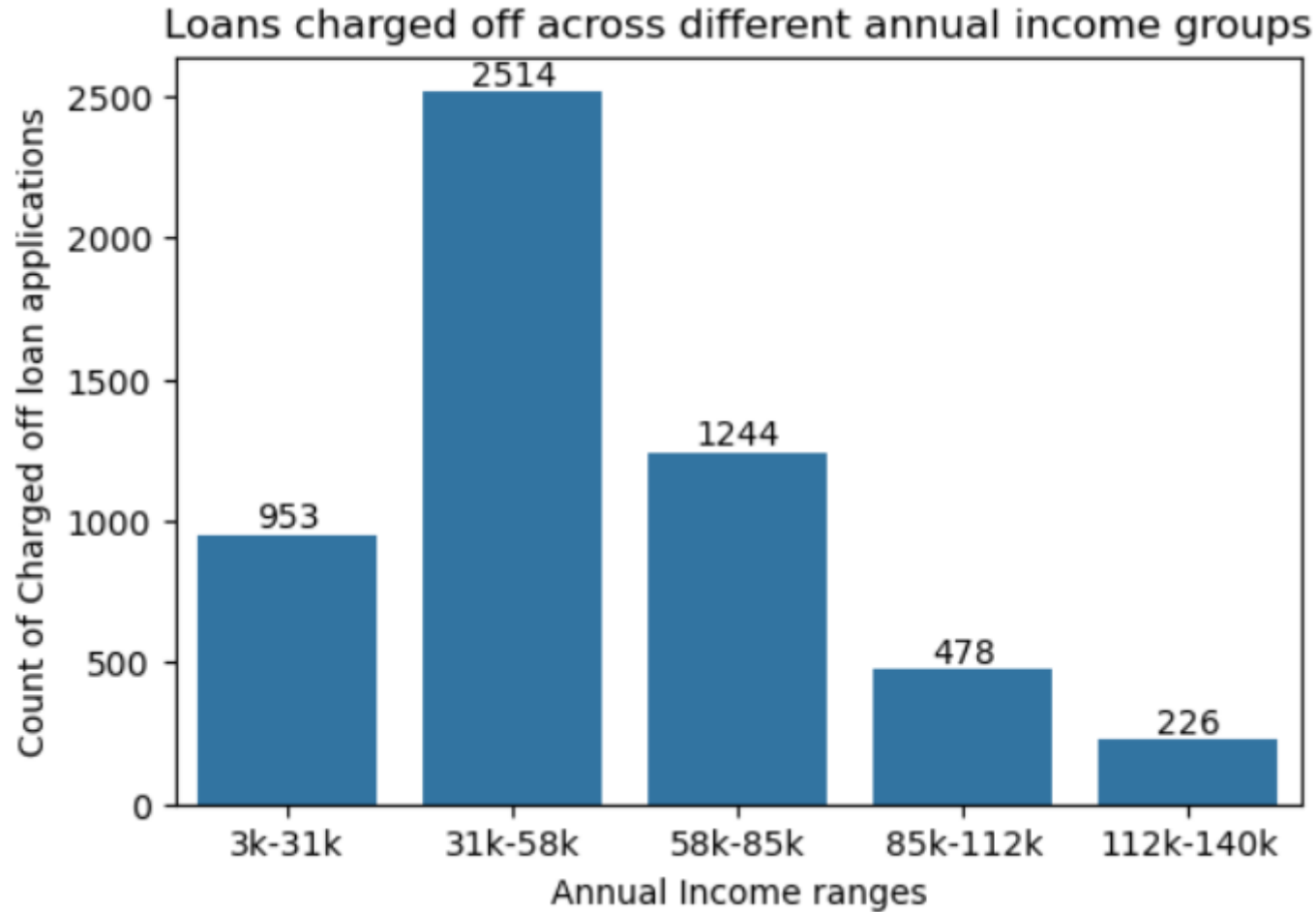
Loan defaults across Interest Rate ranges



Insights:

- Borrowers who got interest rates between 13-17% are more likely to default the loans.

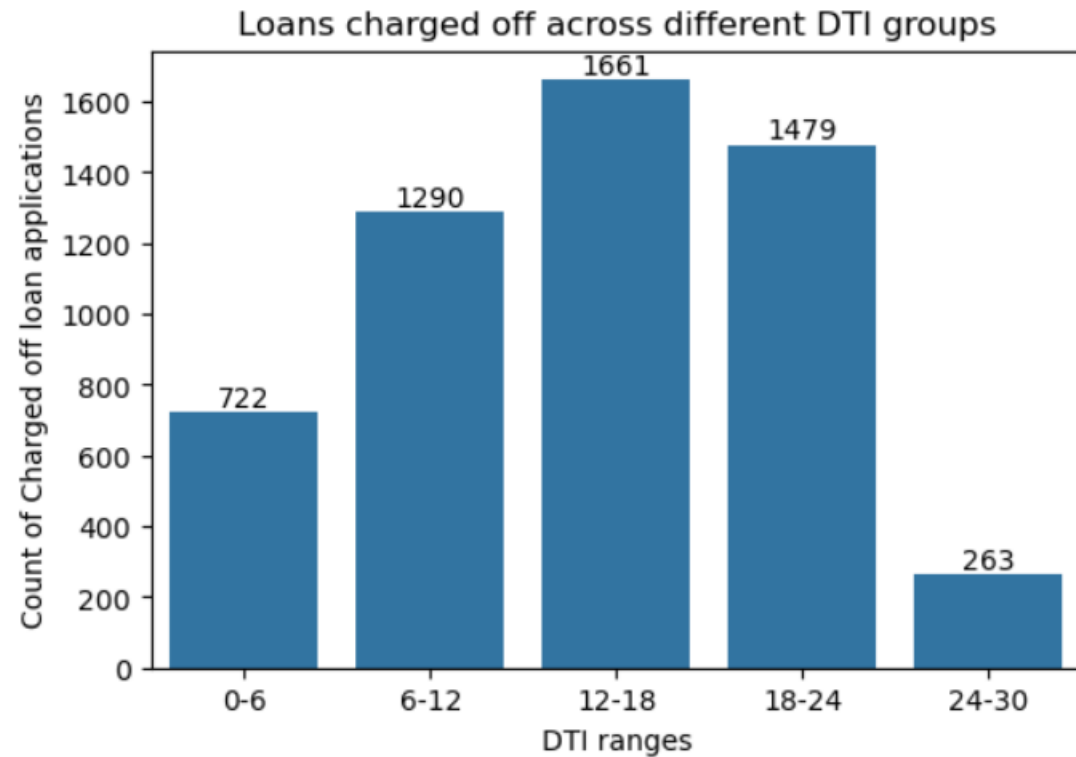
Loan defaults across Annual Income Groups



Insights:

- Borrowers having annual income between \$31-58K are more likely to default the loans.

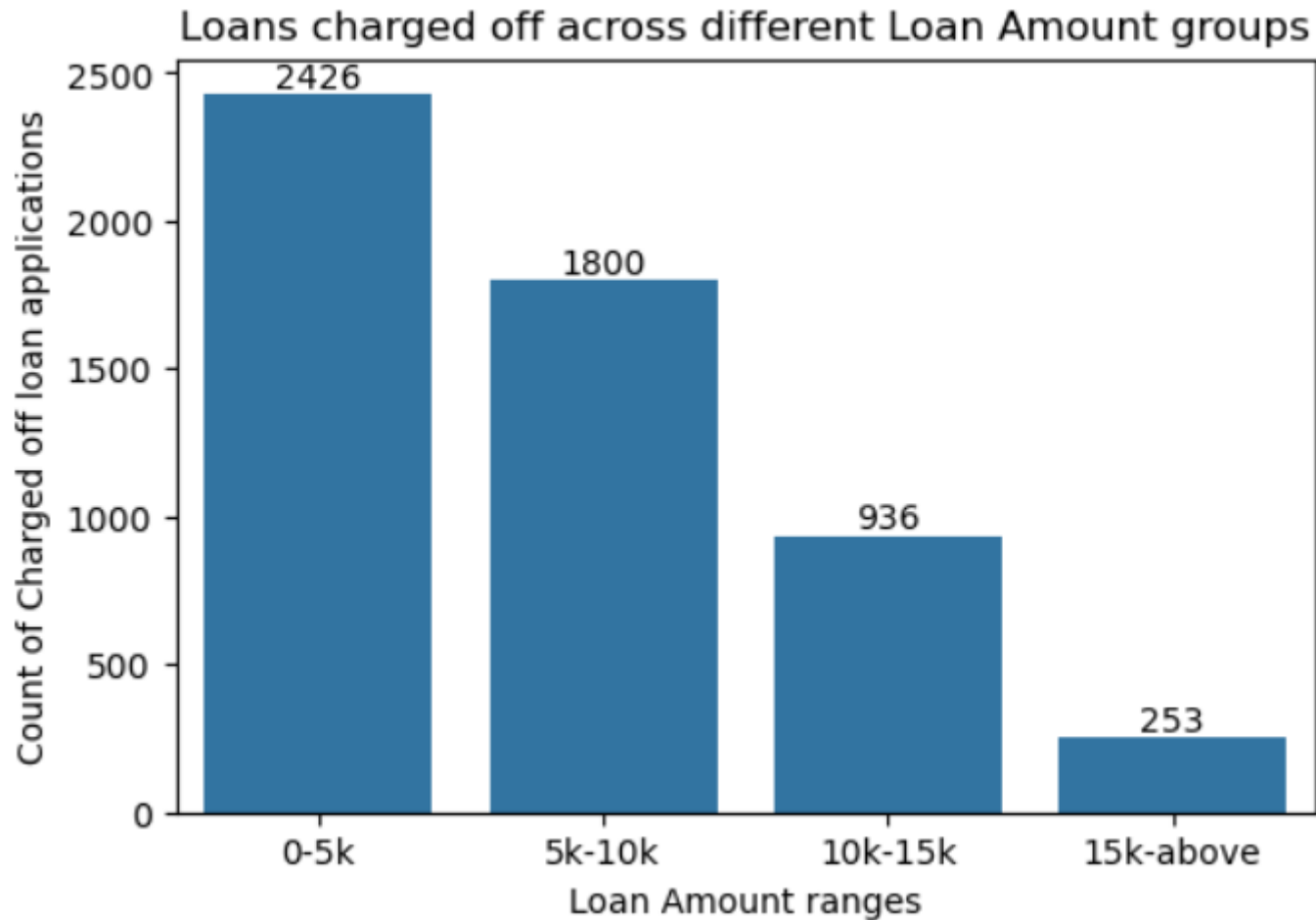
Loan defaults across Debt-to-Income ratio groups



Insights:

- Loans given to people with 12-18 DTI are more likely to default.

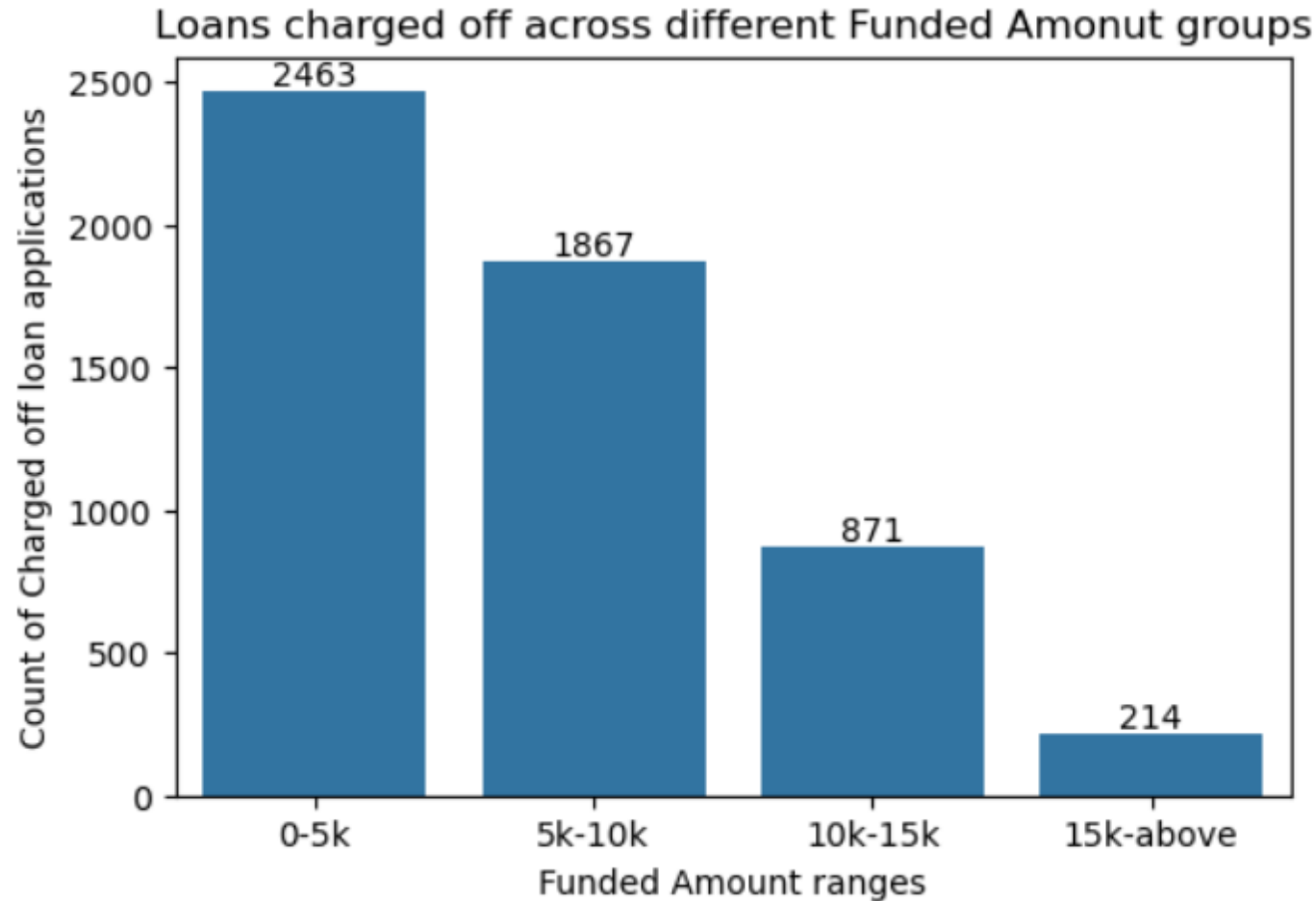
Loan defaults across Loan Amount ranges



Insights:

- Loans given for the amounts less than \$5K are more likely to default than the loans given for higher amounts.
- The graph shows that higher the loan amount, there are lesser chances of default.

Loan defaults across Funded Amount groups



Insights:

- Borrowers who received funded amounts less than \$5K defaulted the maximum loans. The number is closely followed by borrowers who received funded amounts between \$5-10K.
- Higher the number of funded amounts, there are less chances of default.

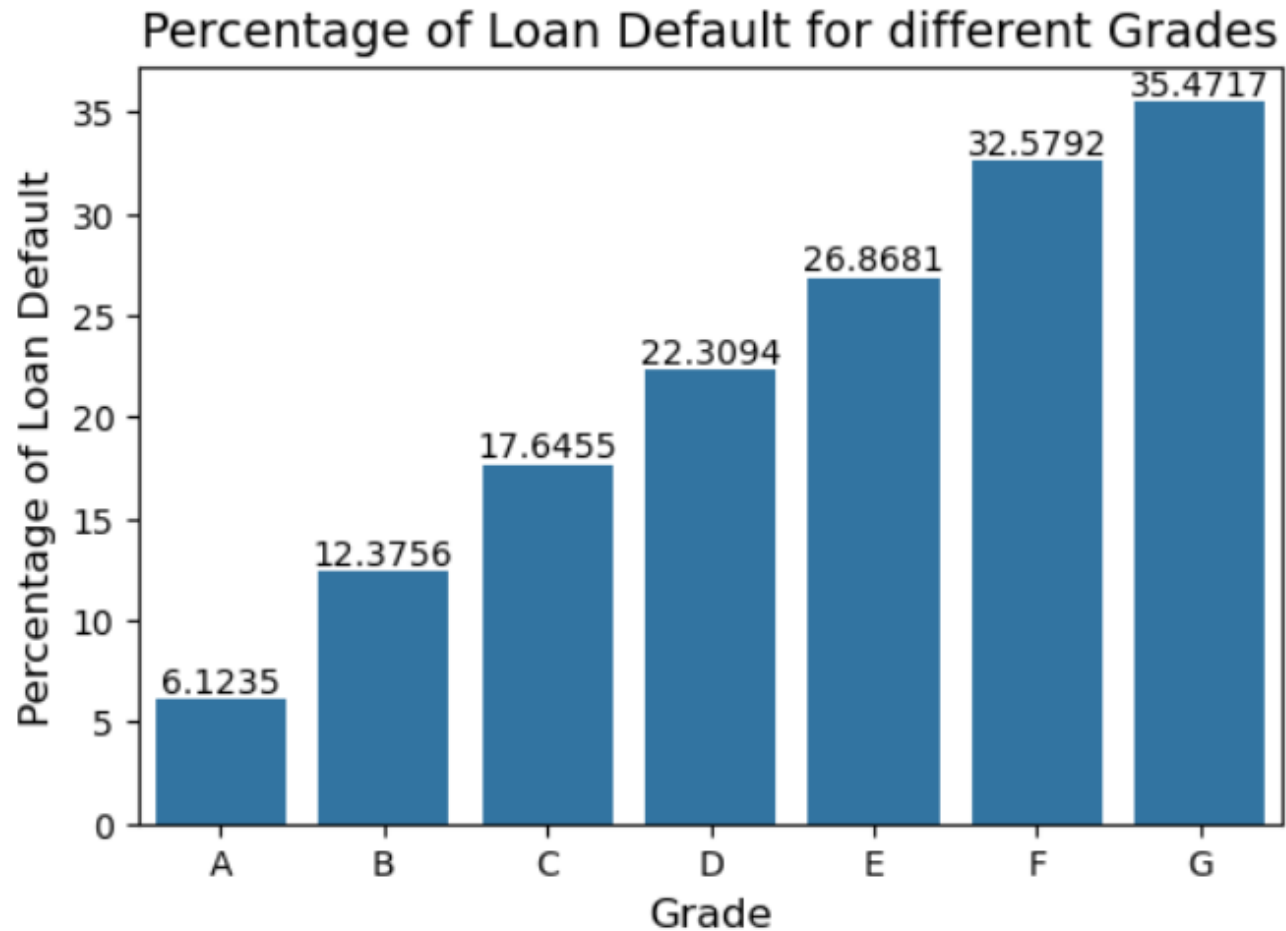
Analysis from graphs drawn on Quantitative Categorical data groups

Analysis from the graphs based on Interest Rates, Annual Income, Debt-to-Income ratio, Loan Amount, Funded Amount:

1. The loans approved within interest rate buckets of 9%-13% and 13%-17% were defaulted. The numbers for both these brackets are more than other three buckets.
2. Most of the defaulted loan borrowers had annual salaries between 31-58K.
3. Most of the defaulted loans had dti between 12-18 and 18-24.
4. Small loans, of below 10k are most likely to default.
5. Loans having funded amounts below 10k are most likely to default.

Bivariate and Multivariate Analysis

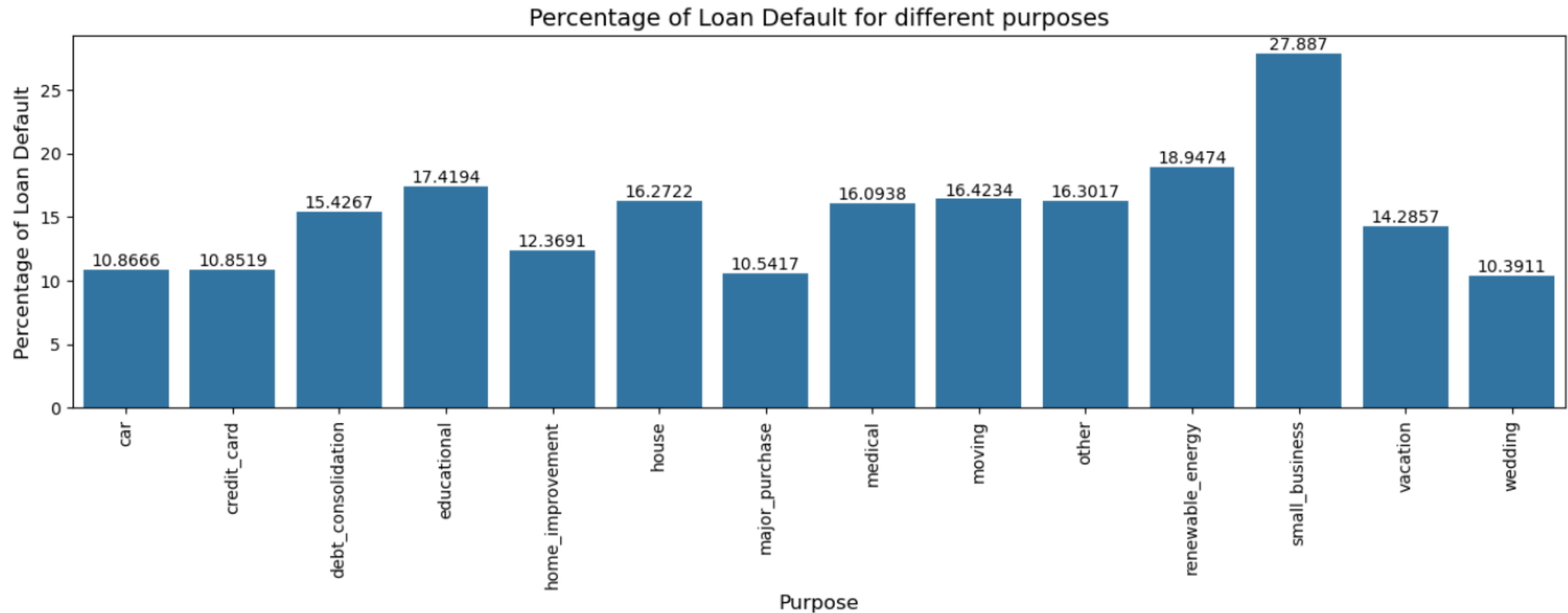
Percentage of Loan defaults across Grades



Insights:

- Percentage of loan defaults increases with Grades in order of A, B, C, D, E, F and G.
- Earlier we saw that the Grade B had maximum number of defaults, but the % of Grade G shows maximum defaults.

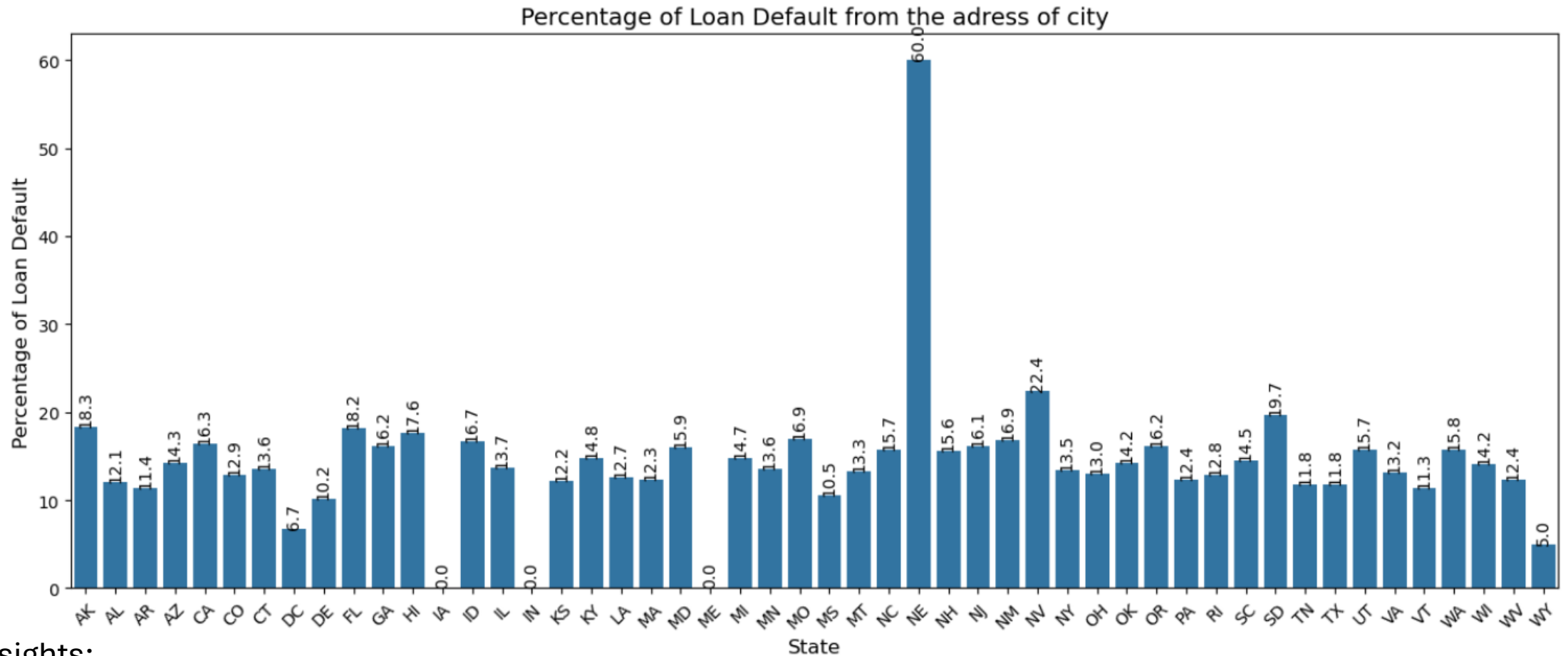
Percentage of Loan defaults across Purposes



Insights:

- Loans given for Small Business have maximum percentage of defaults.
- Earlier we saw that the Debt Consolidation had maximum loan defaults.

Percentage of Loan defaults across States

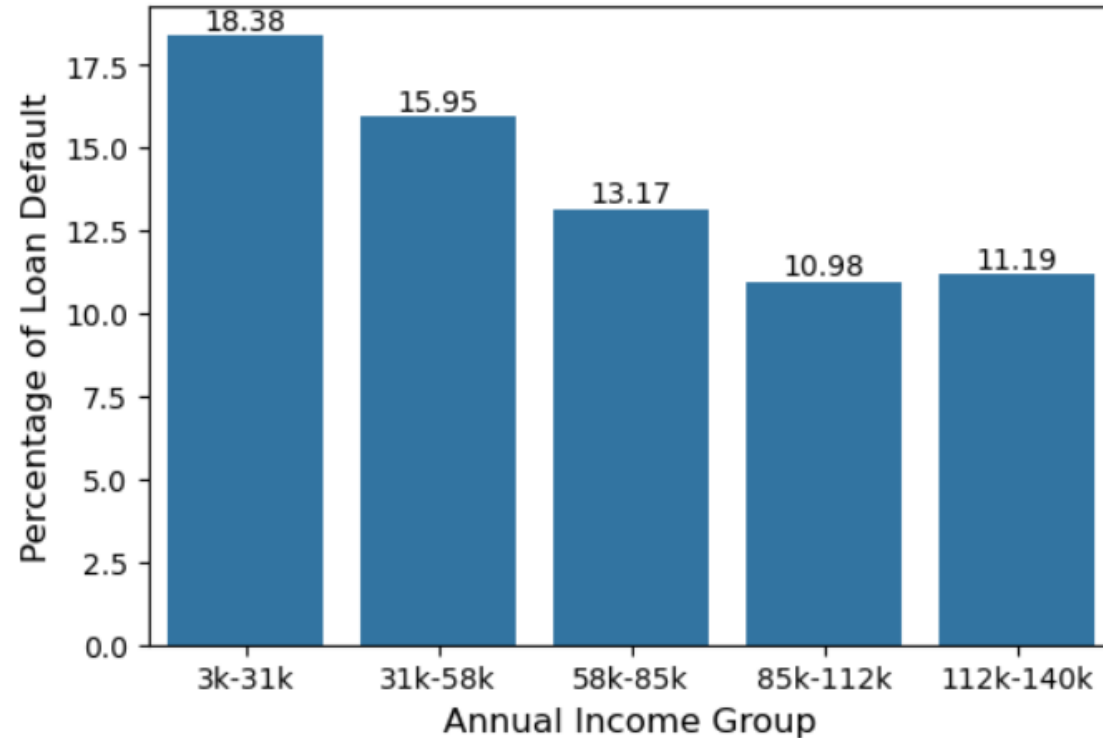


Insights:

- Loans given in State NE had maximum percentage of defaults.
- Earlier we saw that the loans given in States CA, FL and NY had maximum loan defaults.

Percentage of Loan defaults across Income Groups

Percentage of Loan Default based on annual income groups of applications

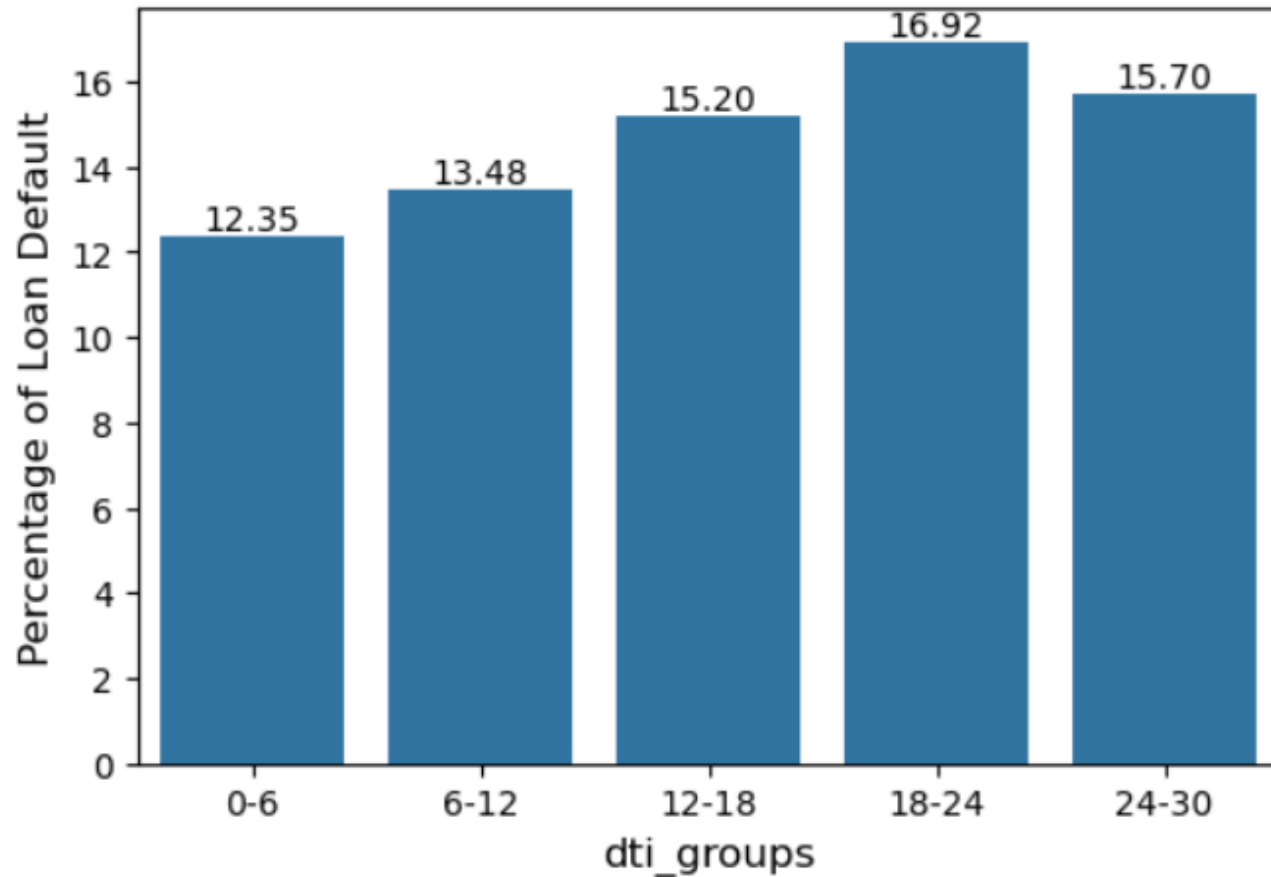


Insights:

- Percentage of loan defaults is maximum with the lower income groups, having income less than \$31K.
- In previous slides we saw that borrowers having income in range between \$31-58K had maximum defaults.

Percentage of Loan defaults across DTI ranges

Percentage of Loan Default based on dti groups of applications

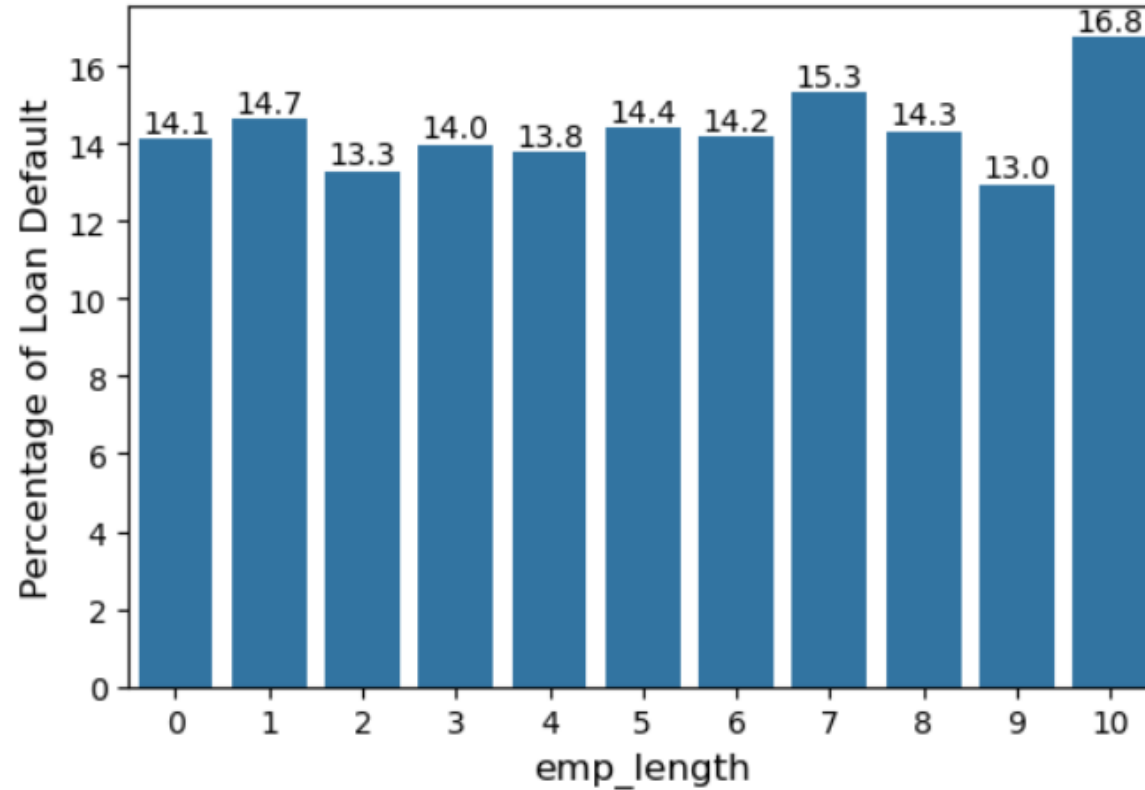


Insights:

- Percentage of loan defaults is maximum when the DTI is in range 18-24.
- In previous slides we noticed that the maximum loan defaults were in DTI range 12-18.

Percentage of Loan defaults across Employment Length

Percentage of Loan Default based on employment length of applications



Insights:

- Borrowers with 10+ years of employment time have maximum percentage of loan defaults.
- This is in-line with the observation we had with the total number of loan defaults w.r.t. employment length.

Percentage of Loan defaults across Home Ownership

Percentage of Loan Default based on home ownership of applications

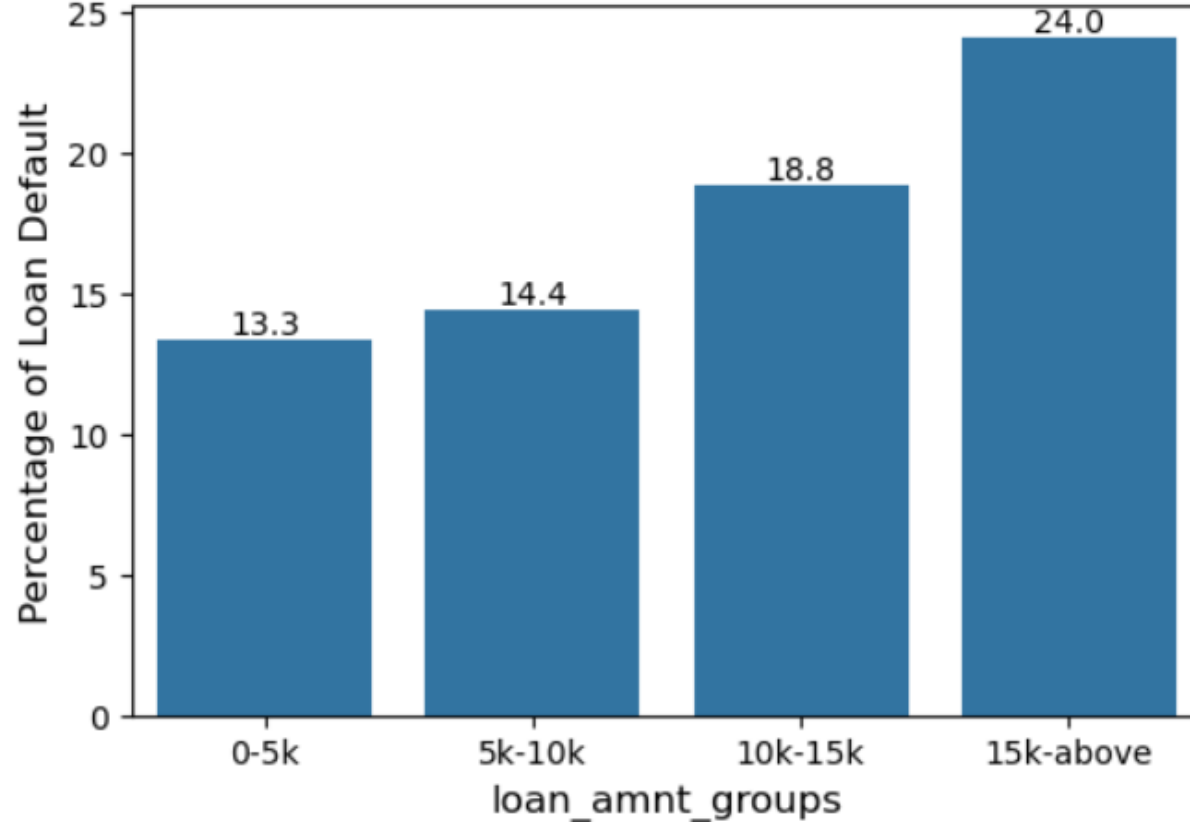


Insights:

- Percentage of loan defaults is maximum for the borrowers who have home ownership listed as 'Other'.
- In previous slides we saw that borrowers having home ownership as Mortgage or Rent had maximum loan defaults.

Percentage of Loan defaults across Loan Amount ranges

Percentage of Loan Default based on loan amount groups of applications

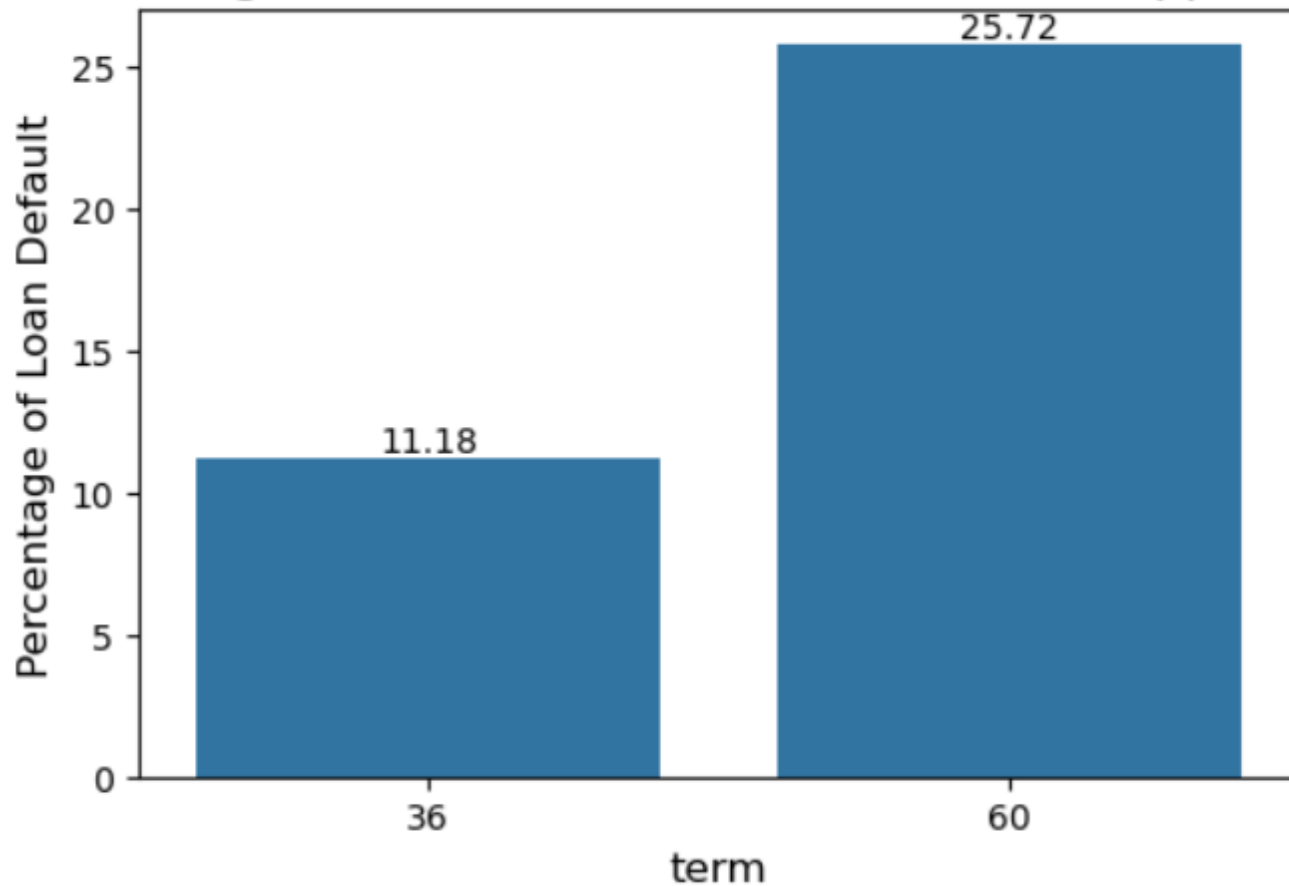


Insights:

- Percentage of loan defaults is maximum for the borrowers who have loans for \$15K and above.
- Earlier we saw that loans which were less than \$5K had maximum loan defaults.

Percentage of Loan defaults across Terms

Percentage of Loan Default based on term of applications

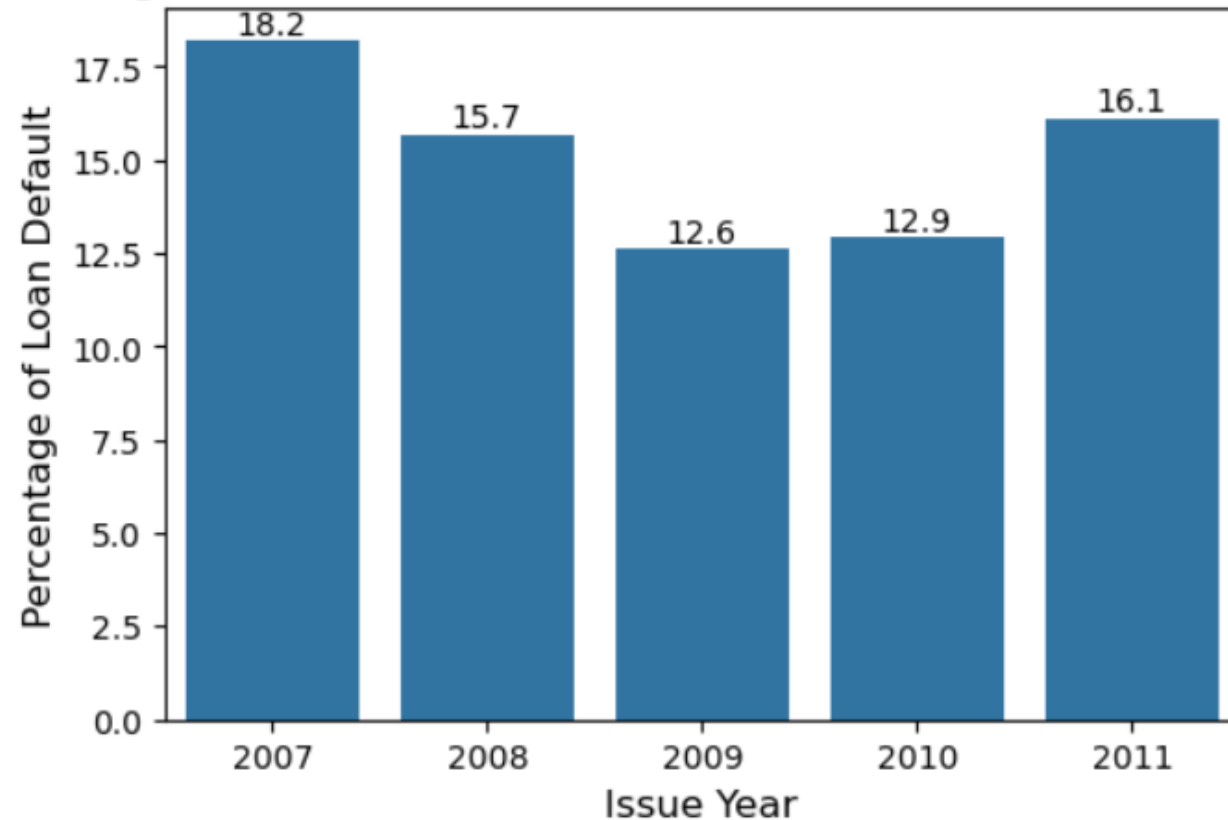


Insights:

- Loans which were given for 60 months term had more than 25% default, while it is only 11% for loans given for a term of 36 months.
- Previously we saw that loans given for 36 months term had maximum number of loan defaults.

Percentage of Loan defaults across Loan Disbursement Year

Percentage of Loan Default based on Loan Issue Year of applications

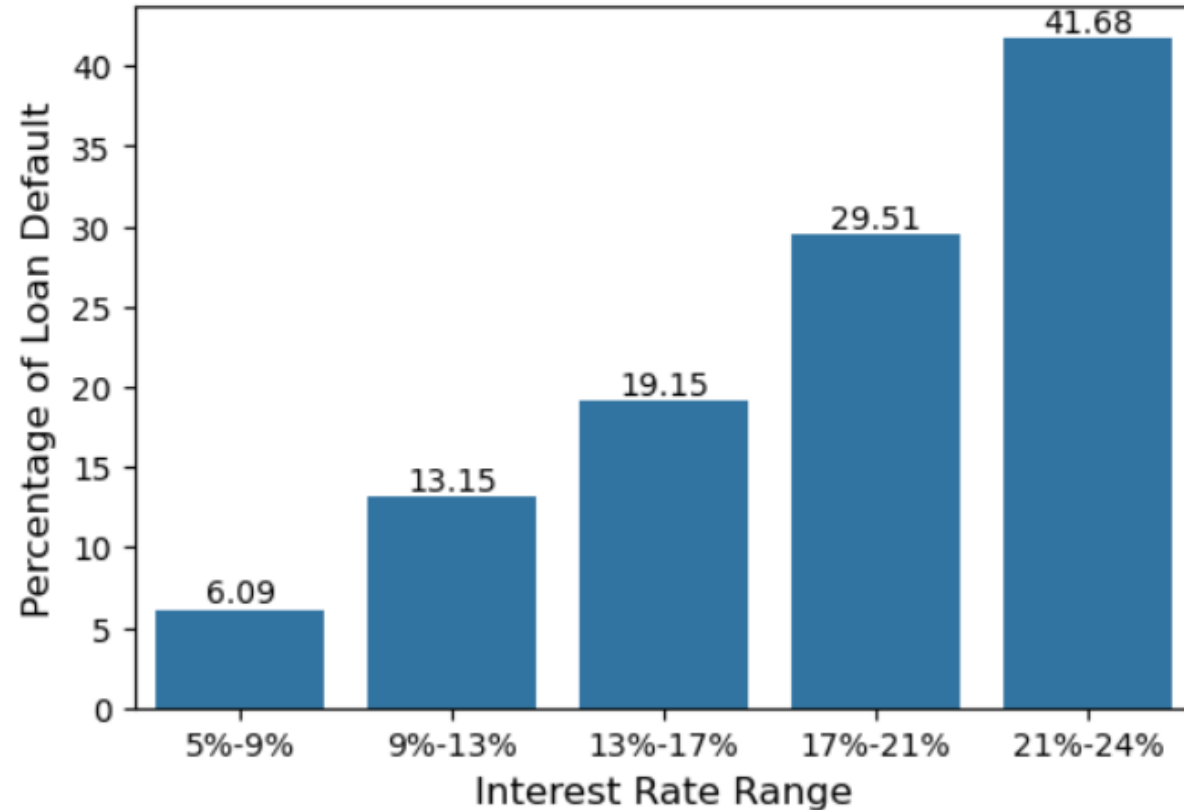


Insights:

- Percentage of loan defaults is maximum for the loans given in 2007.
- In previous slides we saw that loans disbursed in year 2011 had maximum loan defaults.

Percentage of Loan defaults across Interest Rate ranges

Percentage of Loan Default based on Interest Rate Groups of applications



Insights:

- Loans with rate of interest between 21-24% had maximum percentage of loan defaults.
- Earlier we saw that loans between interest rate 13-17% had maximum number of loan defaults.

Conclusions from graphs of Percentage of Loan Defaults vs different factors...

1. The percentage of loan defaults is increasing with grade. More the grade. more are the chances of loan default.

Grade A applicants have the least percentage of loan defaults, while grade G loan applicants have the maximum percentage of loan defaults.

2. Small Business loans have maximum percentage of default.

There are few categories which have almost similar percentage of loan defaults - Renewable Energy, Educational, House, Medical, Moving, Other.

On the other side, categories like Car, Credit Card, Major Purchase and Wedding have least percentage of loan defaults.

3. Loan applicants from the state 'NE' are most likely to default, followed by applicants from state 'NV'.

4. Applicants with annual income less than \$31K are most likely to default, closely followed by applicants with annual income of less than \$58K.

Applicants with annual income above \$85K are less likely to default, but still have a good percentage of ~11% to default.

5. Applicants with high DTI are more likely to default, specially with in range of 18-24.

In-general we can say applicants having DTI above 12 have more defaults than other applicants.

Applicants with DTI less than 6 are the safest to lend the loan.

Conclusions from graphs of Percentage of Loan Defaults vs different factors

6. Loan applicants who have 10+ years of experience, are also likely to default the most. Applicants with 0-1 year of experience are also a high risk category.

7. Loan applicants who have ambiguous home ownership as 'Other' have significantly high risk to default than others.

8. Applicants who received loan amounts of \$15K and above have the maximum defaults percentage, followed by \$10-15K. So, loan amounts above \$10K are the riskier ones.

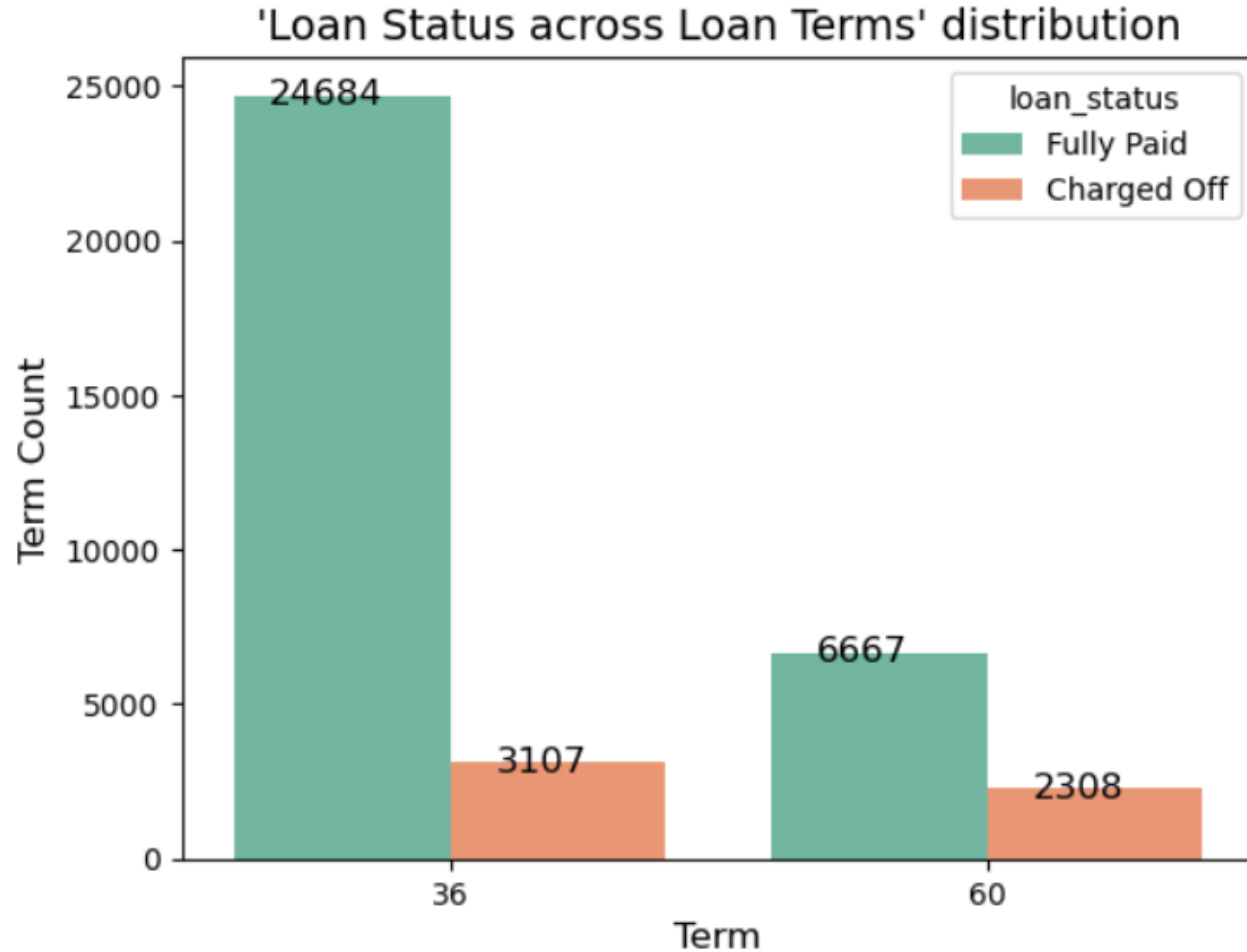
9. Loans with tenure of 60 months have more than double the percentage of default than the loans with tenure of 36 months.

10. The loans issued in year 2007 had the maximum number of defaults, but it reduced significantly for years 2008 and 2009, remained steady in year 2010, but picked the increasing default percentage again in year 2011. Although no conclusive pattern can be identified.

11. Lower the interest rate, lesser are the chances of loan default. The percentage of loan Defaults is increasing with the increase in interest rates.

Loans with interest rates between 21-24% have almost 6 times more default percentage than with loans given at 5-9% rate of interest.

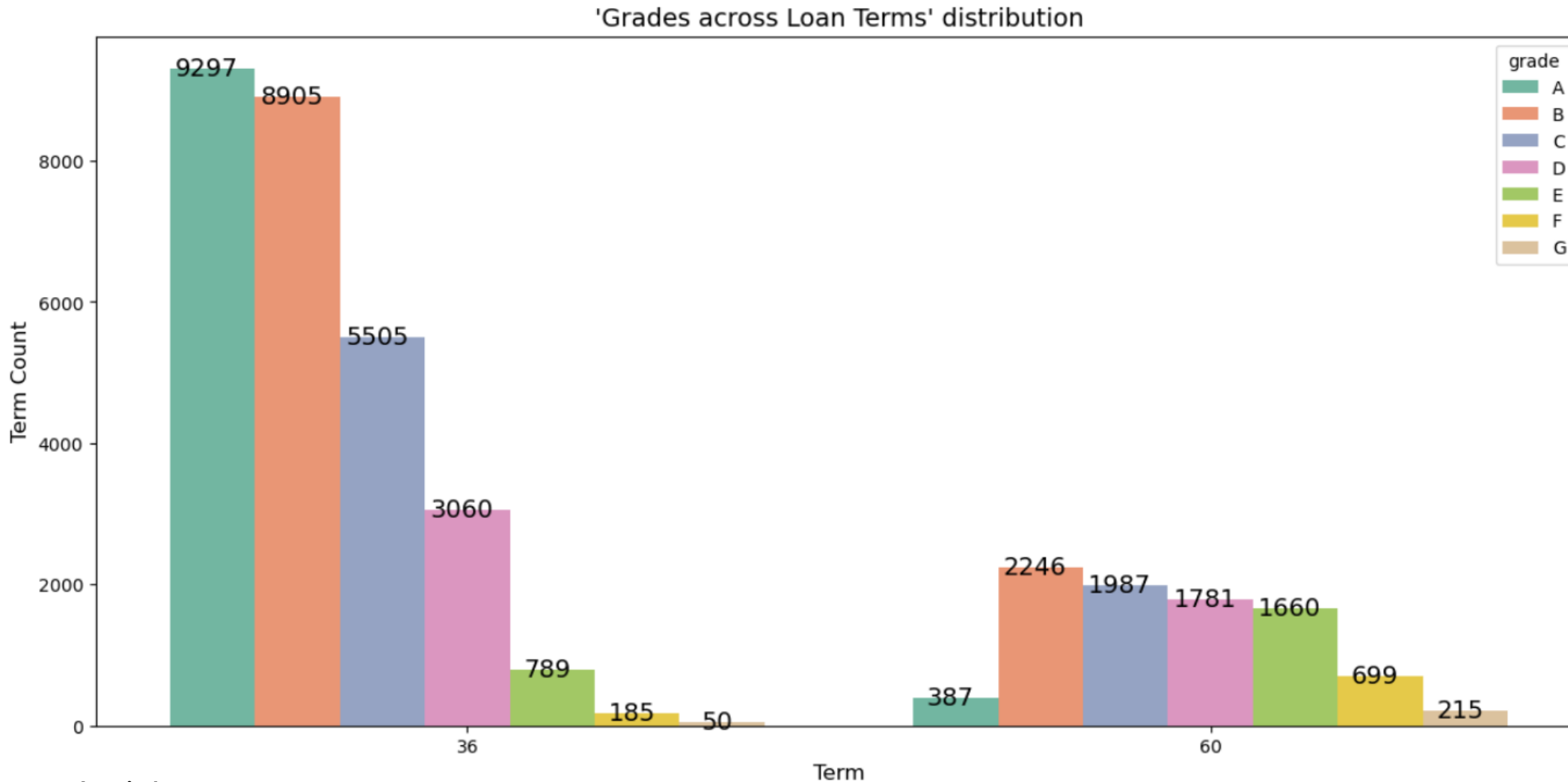
Distribution of Loan defaults across Loan Term



Insights:

- Across 36 months loan term, maximum loans were successfully recovered.
- 60 months loan term loans have a high percentage of loan defaults. This analogy matches with what we saw in previous bivariate analysis on loan terms.

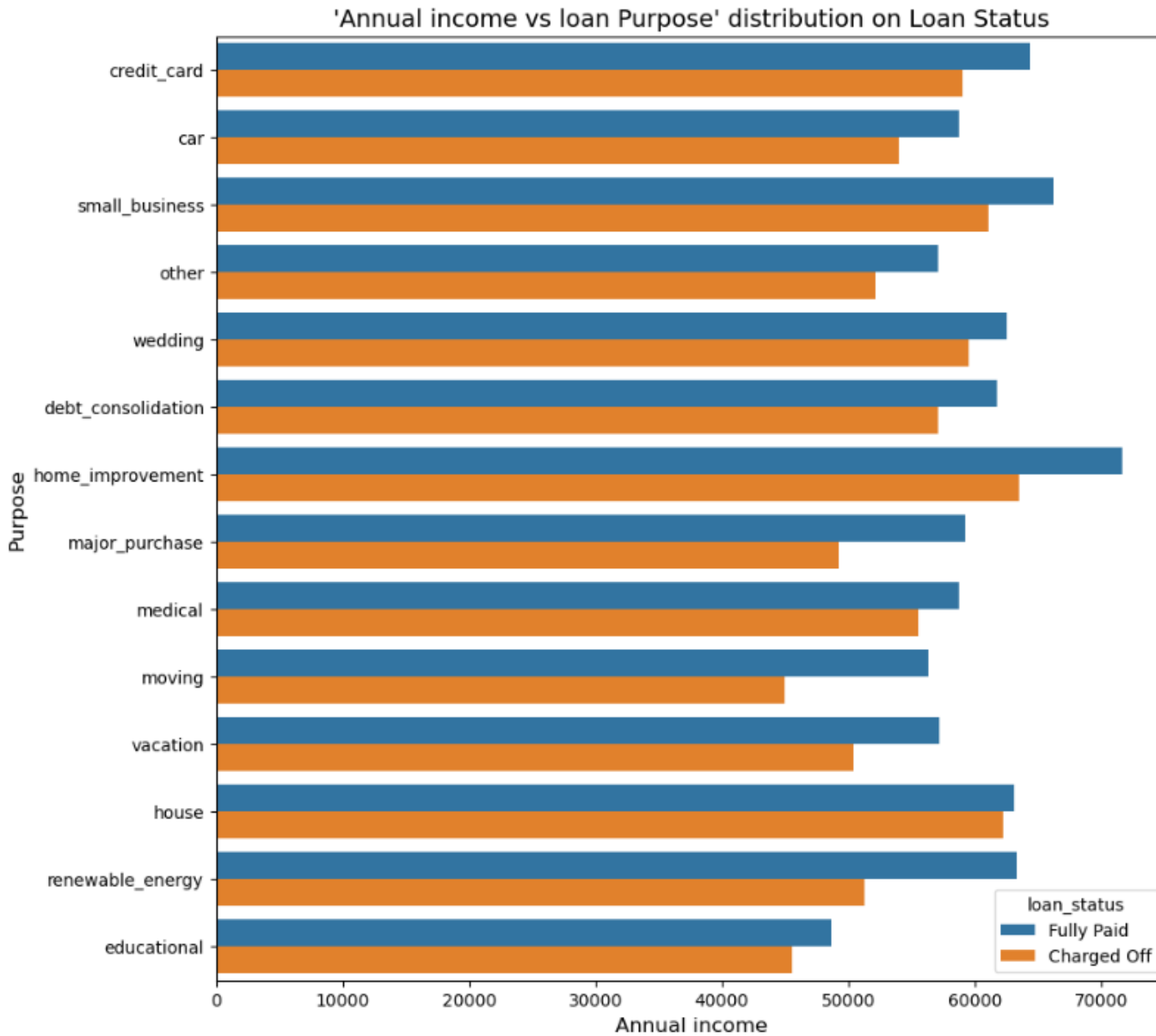
Distribution of Loan defaults across Loan Term



Insights:

- Grades A and B have high chance of loan default for short term loans.

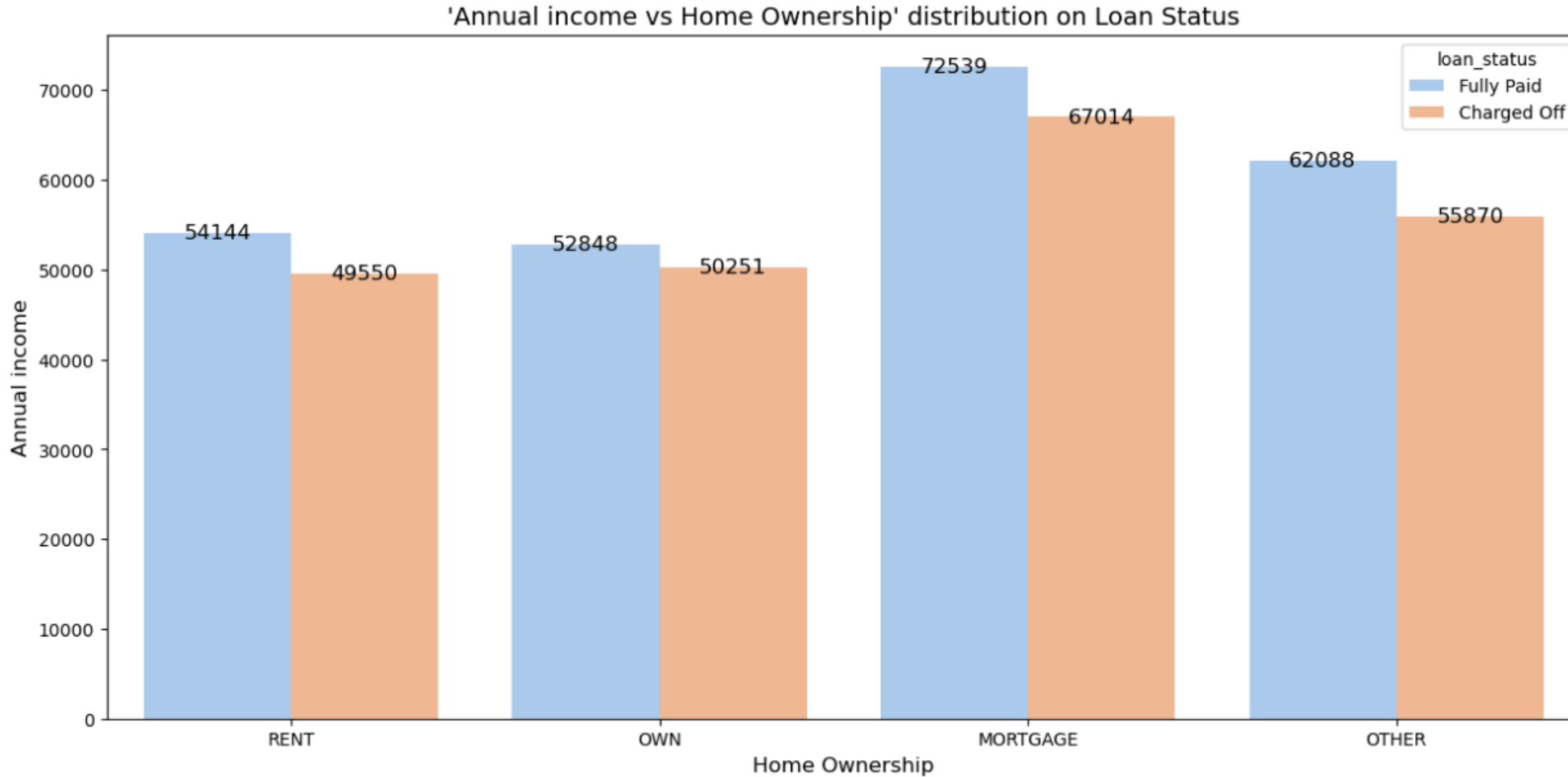
Distribution of Loan defaults over Annual Income and Loan Purpose



Insights:

- Loans given for the purpose of 'Home Improvement' and House, to people having income above \$60K are most likely to default.

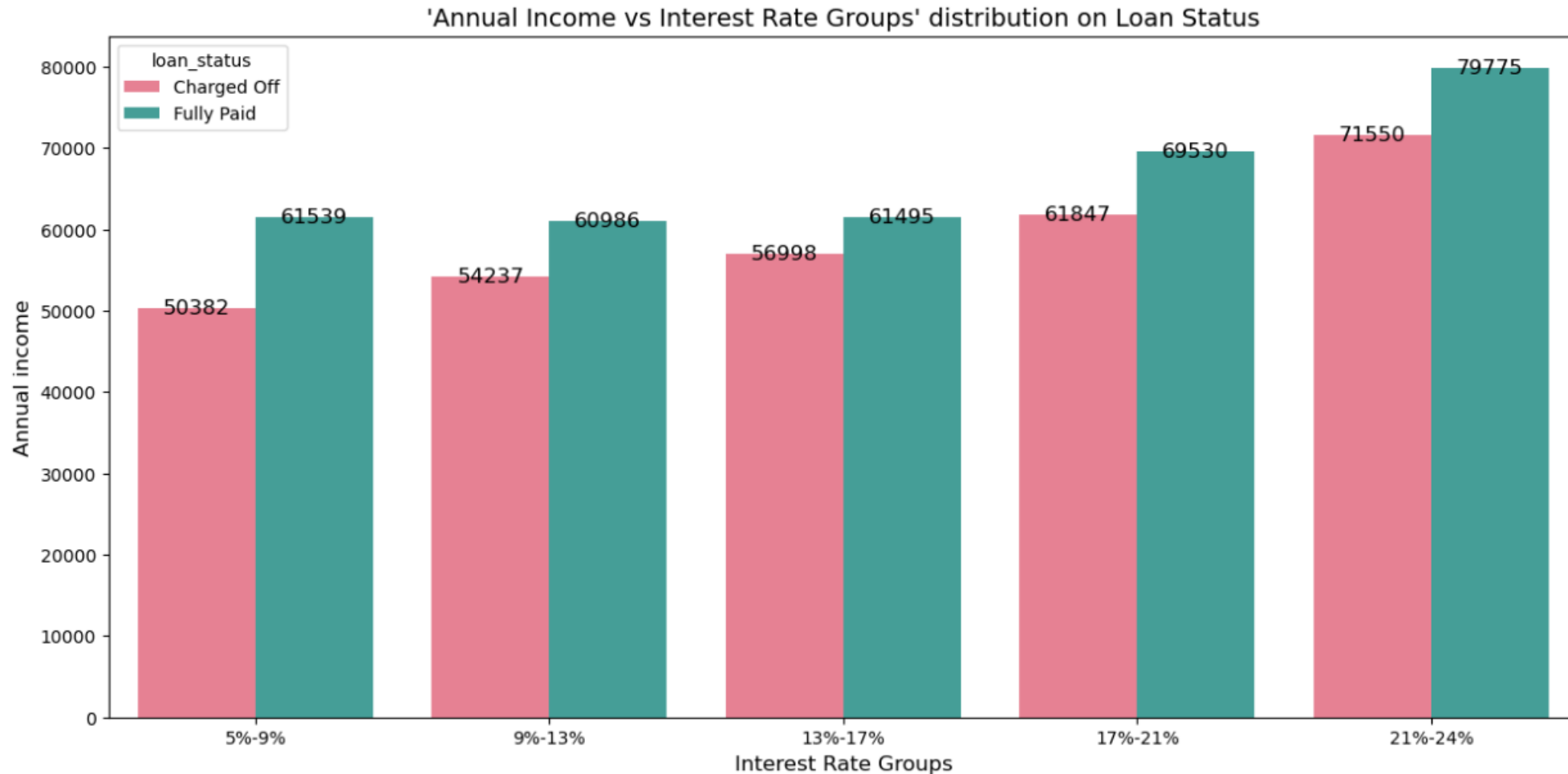
Distribution of Loan defaults over Annual Income and Home Ownership



Insights:

- Borrowers having annual income more than \$60K and already mortgaged their homes are most likely to default.

Distribution of Loan defaults across Annual Income and Interest Rate ranges

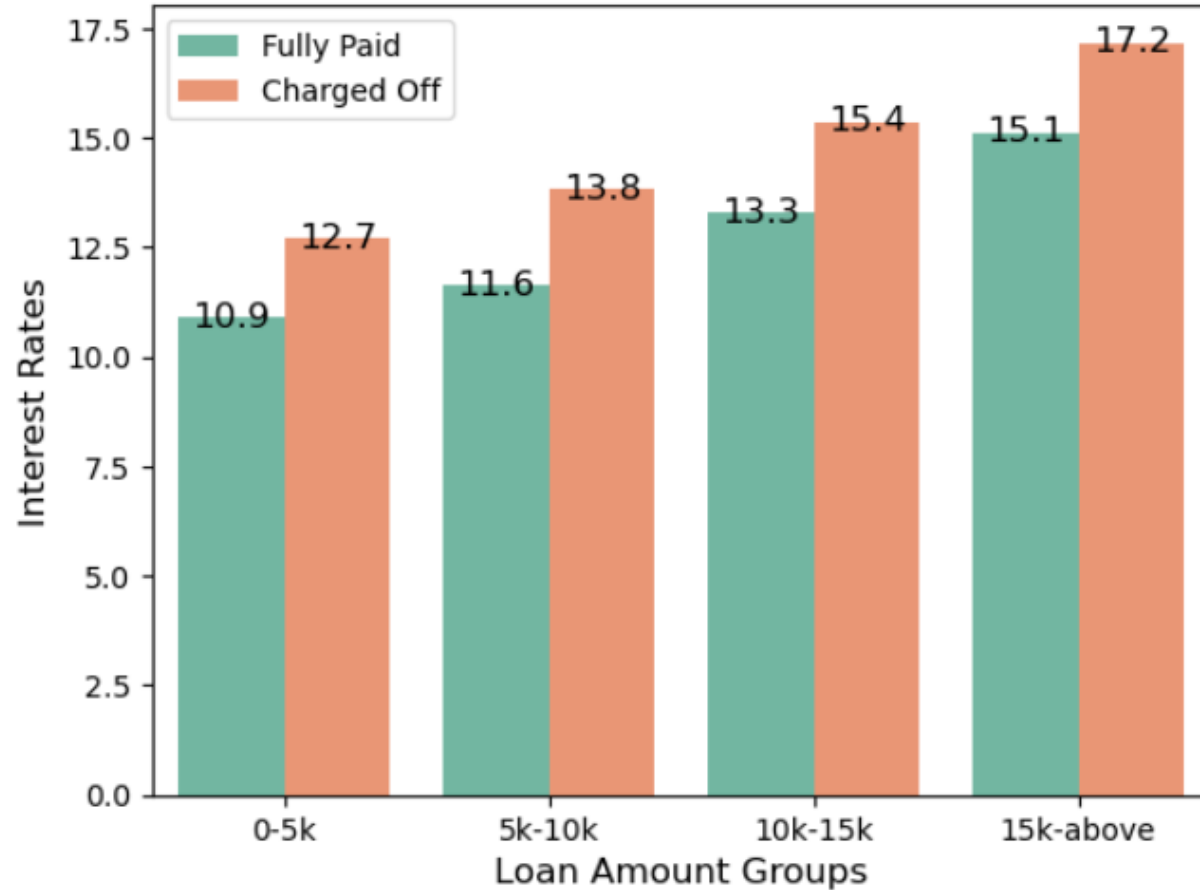


Insights:

- Borrowers with income above \$70K and having loans with interest rate between 21-24% are most likely to default.

Distribution of Loan defaults across Loan Amounts and Interest Rates

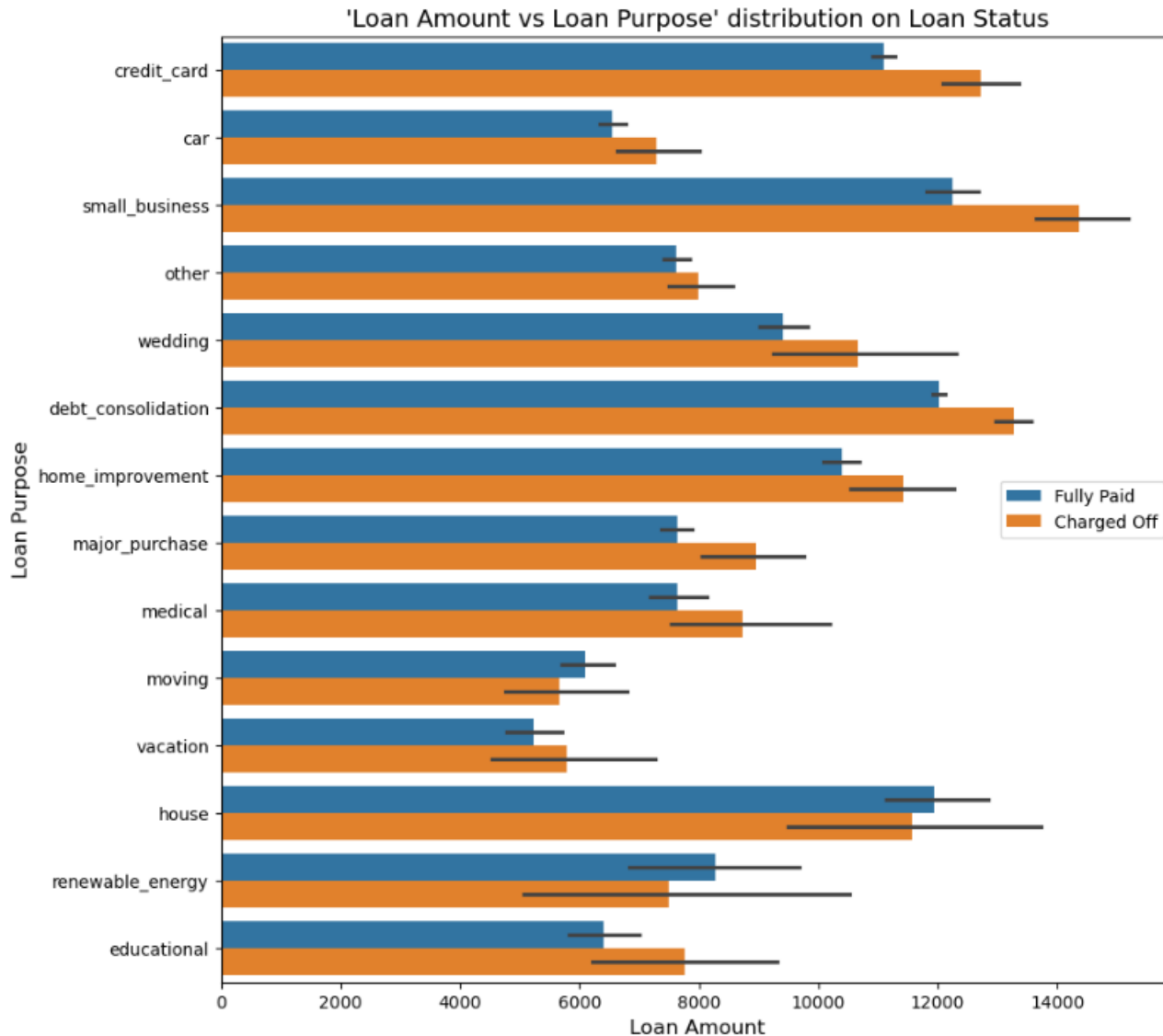
'Loan Amount Groups vs Interest Rate' distribution on Loan Status



Insights:

- Borrowers with loans of amount above \$15K and having loans with interest rate above 15% are most likely to default.

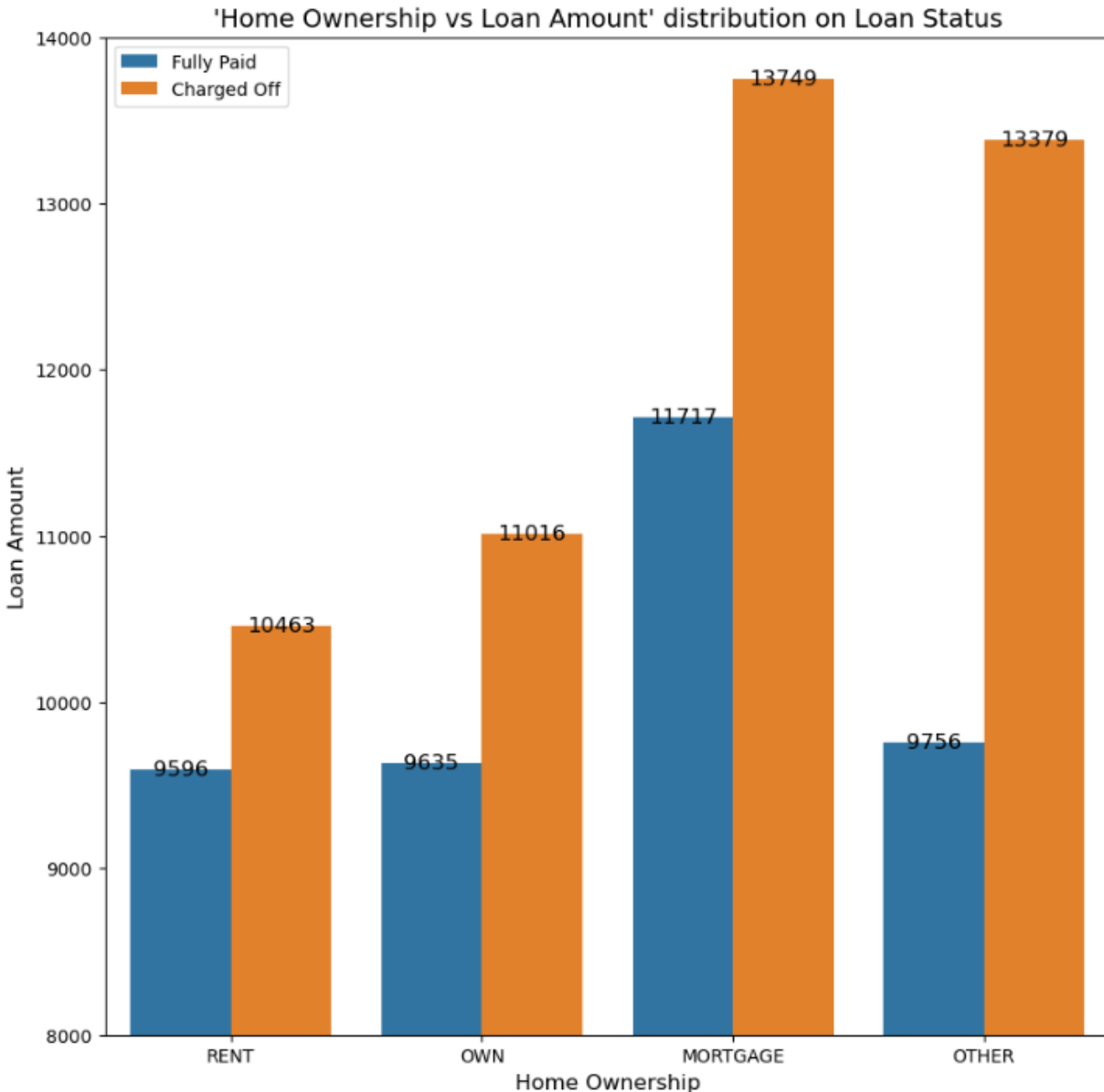
Distribution of Loan defaults across Loan Amounts and Loan Purpose



Insights:

- Loans above loan amount \$12K and with purpose 'Small Business' are most likely to default, closely followed by loans with purpose 'Debt Consolidation'.

Distribution of Loan defaults across Home Ownerships and Loan Amounts

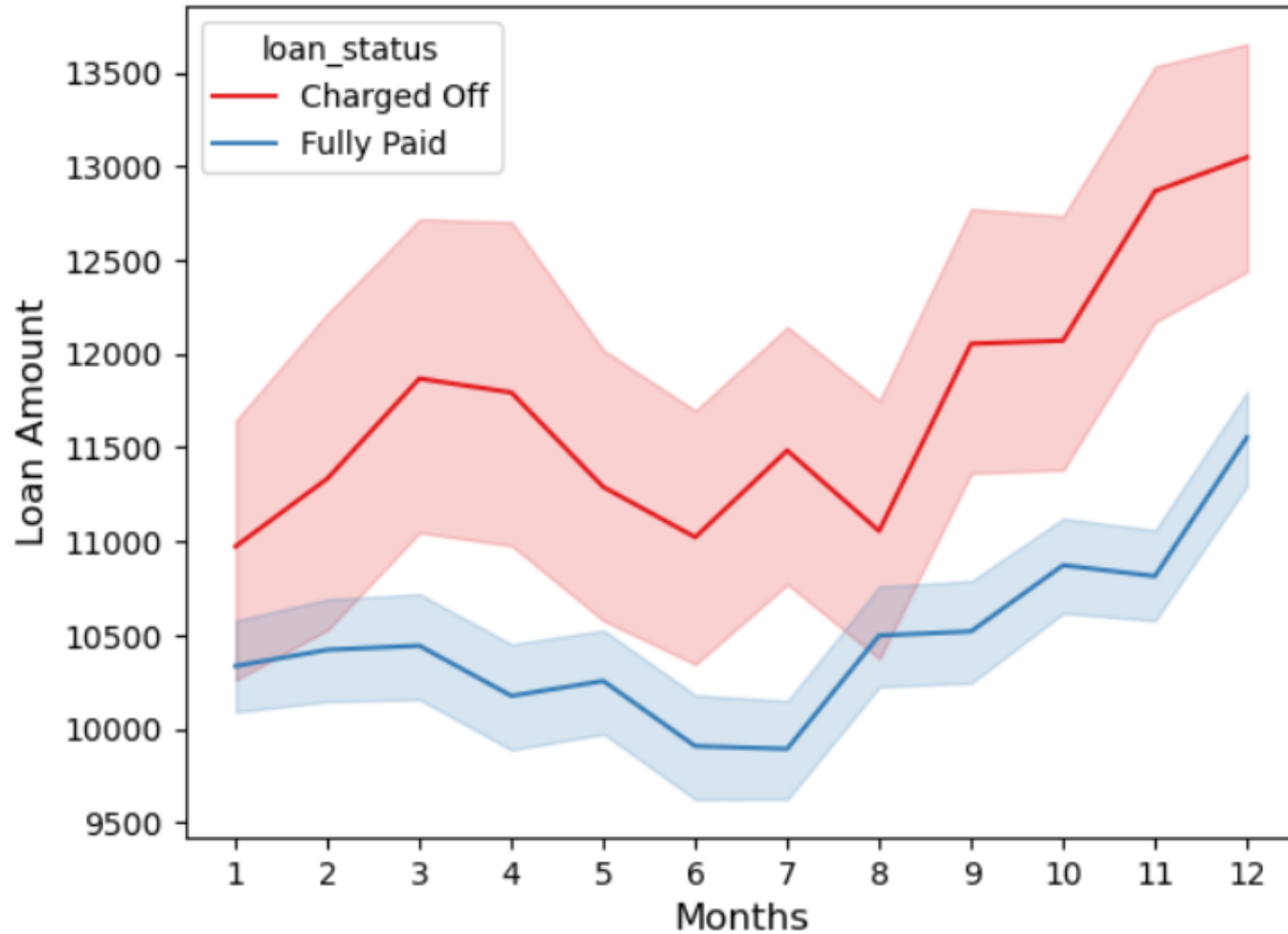


Insights:

- Loans above loan amount \$12K given to borrowers with current Home Ownership as 'Mortgage' or 'Other' are most likely to default.

Distribution of Loan defaults across Loan Amounts and Disbursement Months

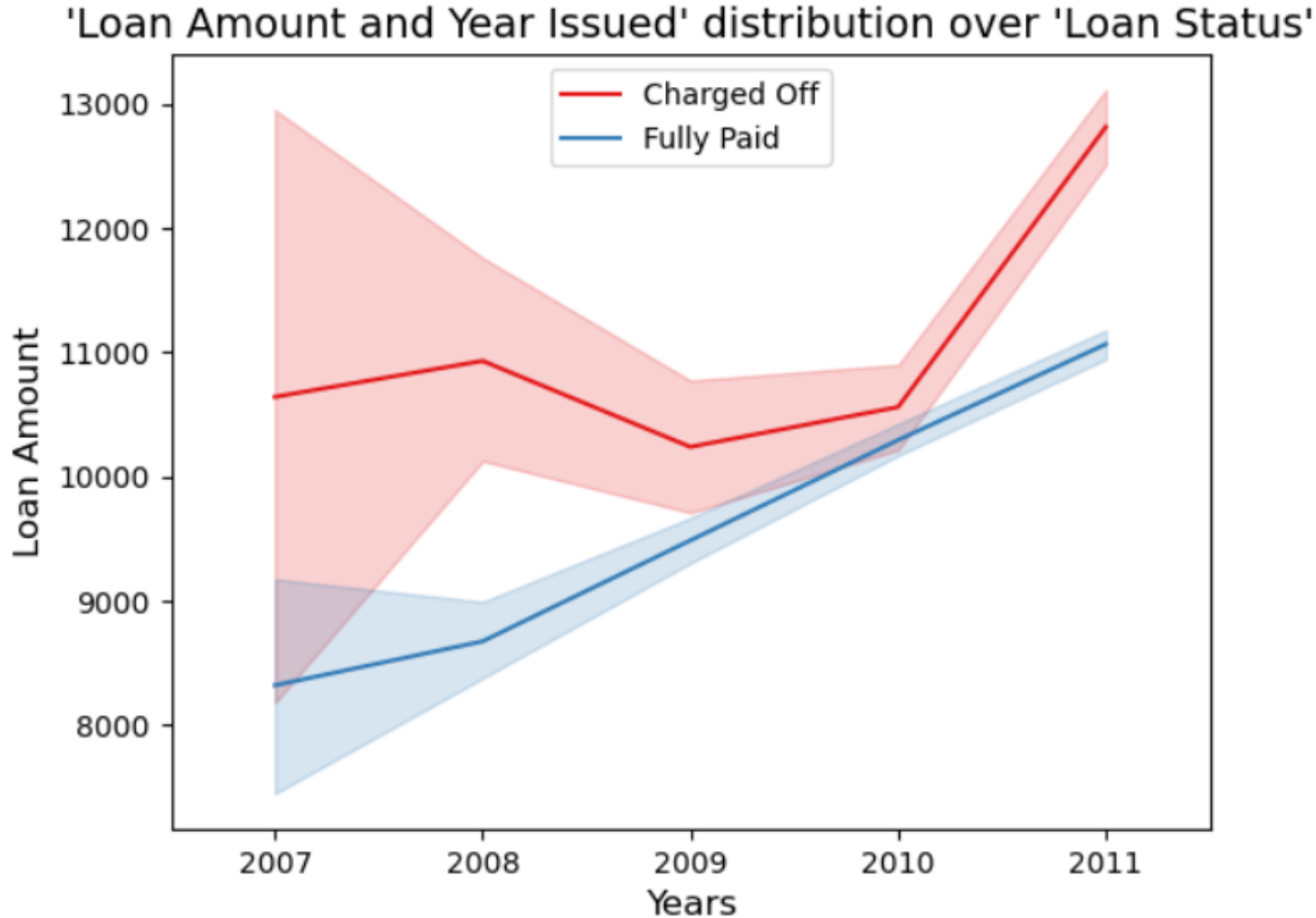
'Loan Amount and Month Issued' distribution over 'Loan Status'



Insights:

- Loans disbursed in November and December are most likely to default.

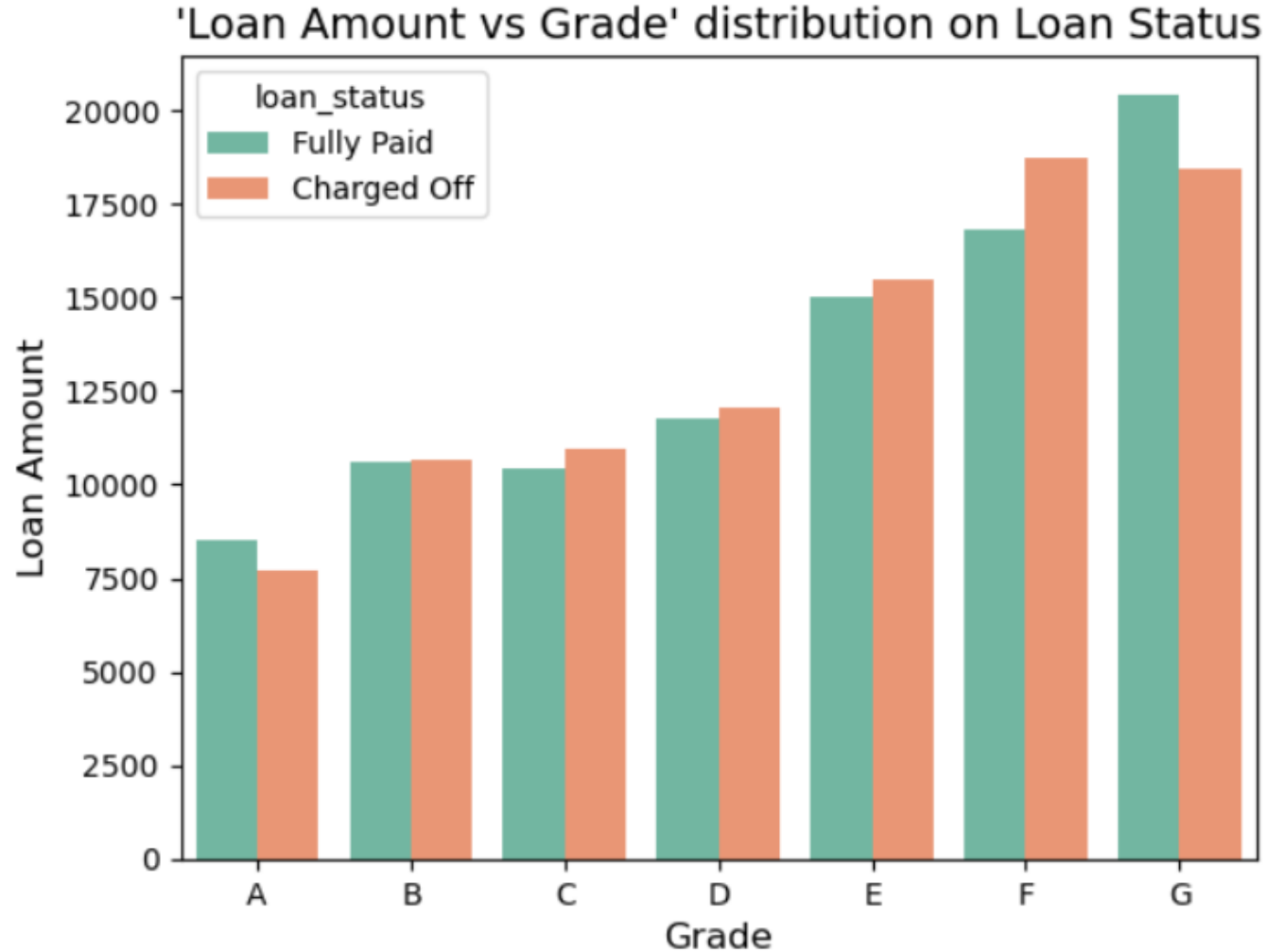
Distribution of Loan defaults across Loan Amounts and Disbursement Months



Insights:

- Loans disbursed in year 2011 defaulted the most.

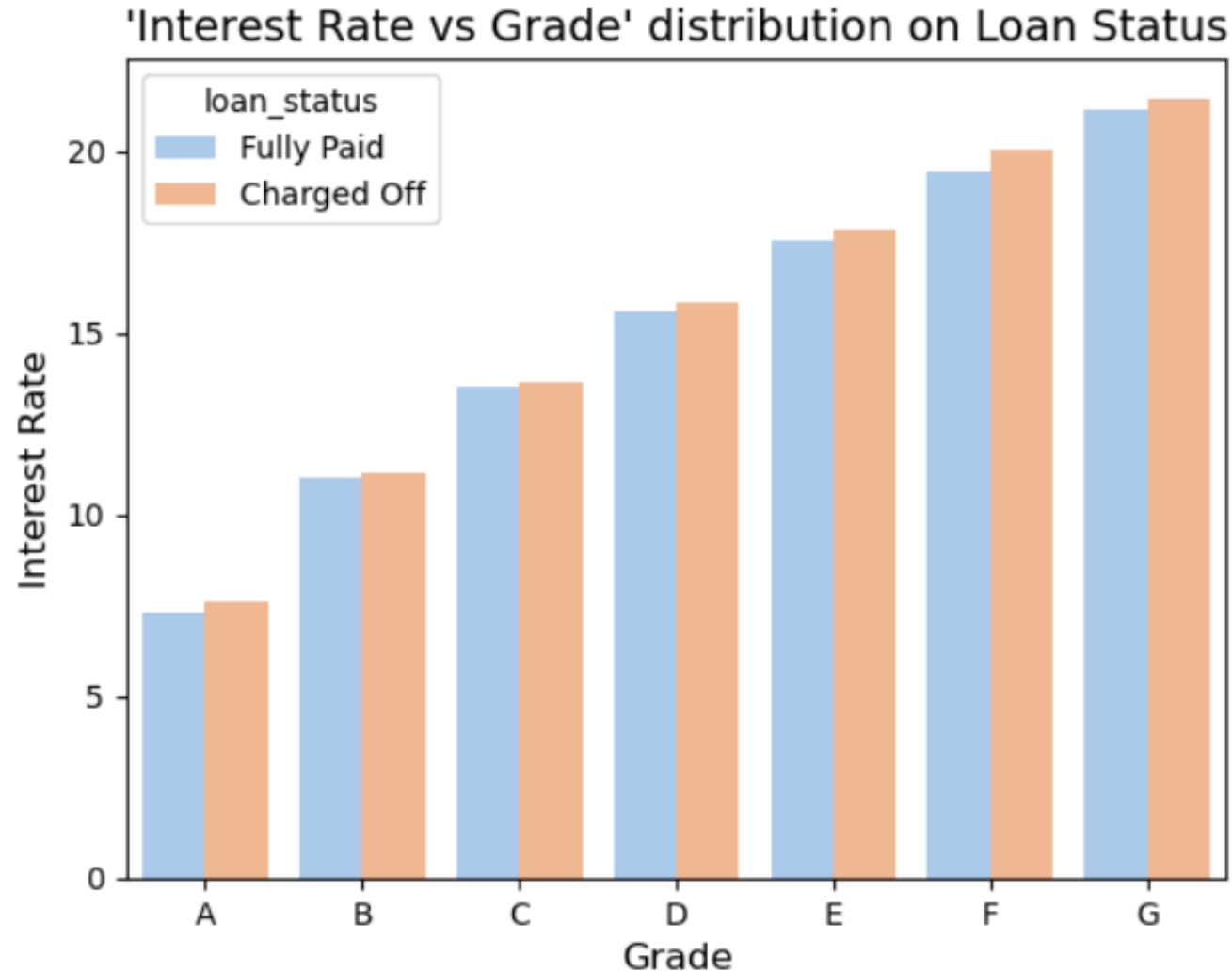
Distribution of Loan defaults across Loan Amounts and Grade



Insights:

- Loans disbursed for Grades F and G, with loan amounts above \$15K are most likely to default.

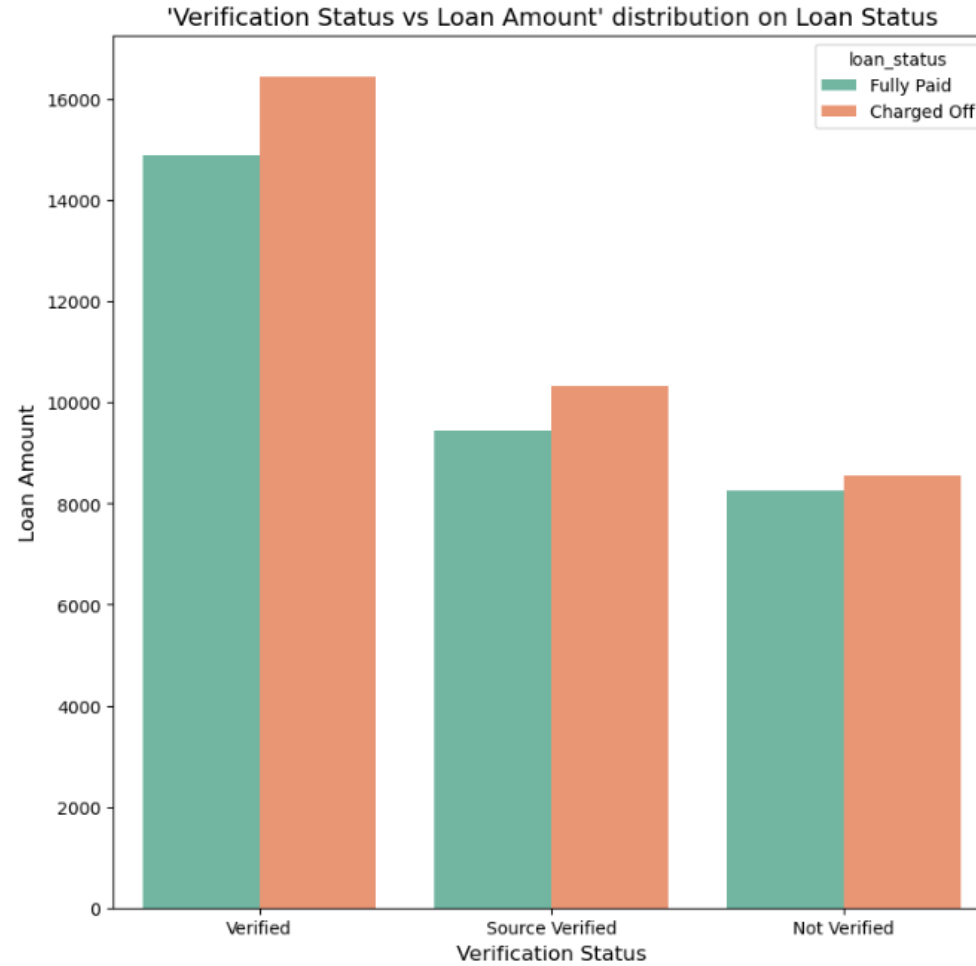
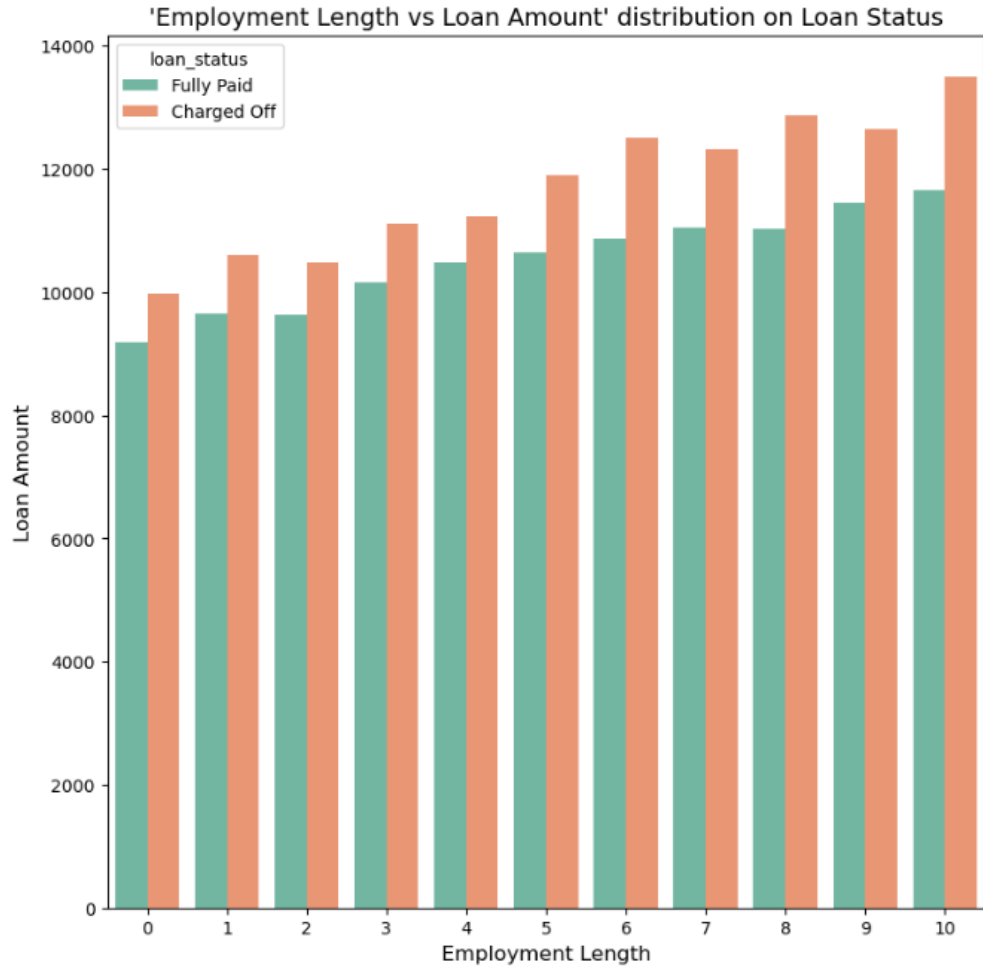
Distribution of Loan defaults across Interest Rate and Grade



Insights:

- Loans given to Grade G with Interest Rate above 20% have maximum defaults.
- Loans disbursed for Grades E and F, with Interest Rates above 15 are also most likely to default.

Distribution of Loan defaults across Loan Amounts, Employment Length and Verification Status



Insights:

- The default rate is maximum for employment length 10+ yrs and loan amount above \$12K.
- For Verified loans with amount above \$14K, the default is maximum.

Patterns and inferences drawn from above graphs

1. Applicants above income of \$60K are taking loans for 'Home Improvement' and 'House' purposes.
2. Applicants with home ownership as 'MORTGAGE' and 'OTHER', and have income of \$60k+ are applying for more loans.
3. Applicants with high income, above \$60K+ get maximum loans in upper interest categories of rates 17-21% and 21-24%.
4. Applicants with loan amounts of \$15k+ are charged interest rates above 15%.
5. Maximum loans are applied for 'Small Business' where loan amount is more than \$12k have not been paid off.
6. Apart from loan purposes House, Renewable Energy and Moving, rest all categories have more loan defaults than paid off.
7. Applicants who already have their homes MORTGAGED, have loans of \$13k+ and defaulted.
8. Loans distributed towards the later months of years, have defaulted more. Although the pattern shows that loans distributed in the same periods are fully paid more as well.
9. Loans defaulted more for grades F and G and for loan amounts \$15k+.
10. Loans defaulted more for people who have employment length as 10+ yrs and their loan amounts are \$12k+.
11. Loans defaulted even when the loan was verified, for loan amount above \$16k, for grade G and interest rate above 20%.

Recommendations for loan approval...

1. The majority of 'Charged off' loan participants lived in rented houses or mortgage.
2. Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k are likely to default.
3. When the grade is either F or G, and loan amount is between \$15-20K, the loan is likely to default.
4. Loan applicants applying loan for 60 months are likely to default more than the one taking loan for 36 months.
5. Small Business category has maximum percentage of loans defaulted.
6. Majority of loan applicants who charged off reported an annual income of less than \$58K.
7. Applicants whose home ownership is 'MORTGAGE' and have income of \$60-70K are likely to default.
8. When the loan is verified, and loan amount is above \$16K likely to default.
9. Loan applicants from the state of Nebraska(NE) have high percentage of default.
10. Borrowers from states NE, NV, NJ have maximum number of loan defaults.

Recommendations for loan approval

11. Loans for grade G and interest rate above 20% are likely to default.
12. The Percentage of loan defaults increase with Grade E, F and G.
13. Applicants taking loan for 'home improvement' and have income of \$60-70K are likely to default.
14. Majority of the loan applicants who defaulted received loan amounts of \$15K or higher.
15. Applicants who had been employed for more than 10 years accounted for the highest number of loan defaults.
16. Applicants with 0-1 year of experience are also a high-risk category.
17. Most of the charged off loan applicants had annual salaries between \$35-60K.
18. Most of the loan participants who defaulted loans have DTI between 13-17.
19. Lower the interest rate, lesser are the chances of loan default. The percentage of loan Defaults is increasing with the increase in interest rates. Loans with interest rates between 21-24% have almost 6 times more default percentage than with loans given at 5-9% rate of interest.
20. Verified loans for amount above \$16k and grade G should have lower interest rate.

Thank You