LENDING CLUB CASE STUDY

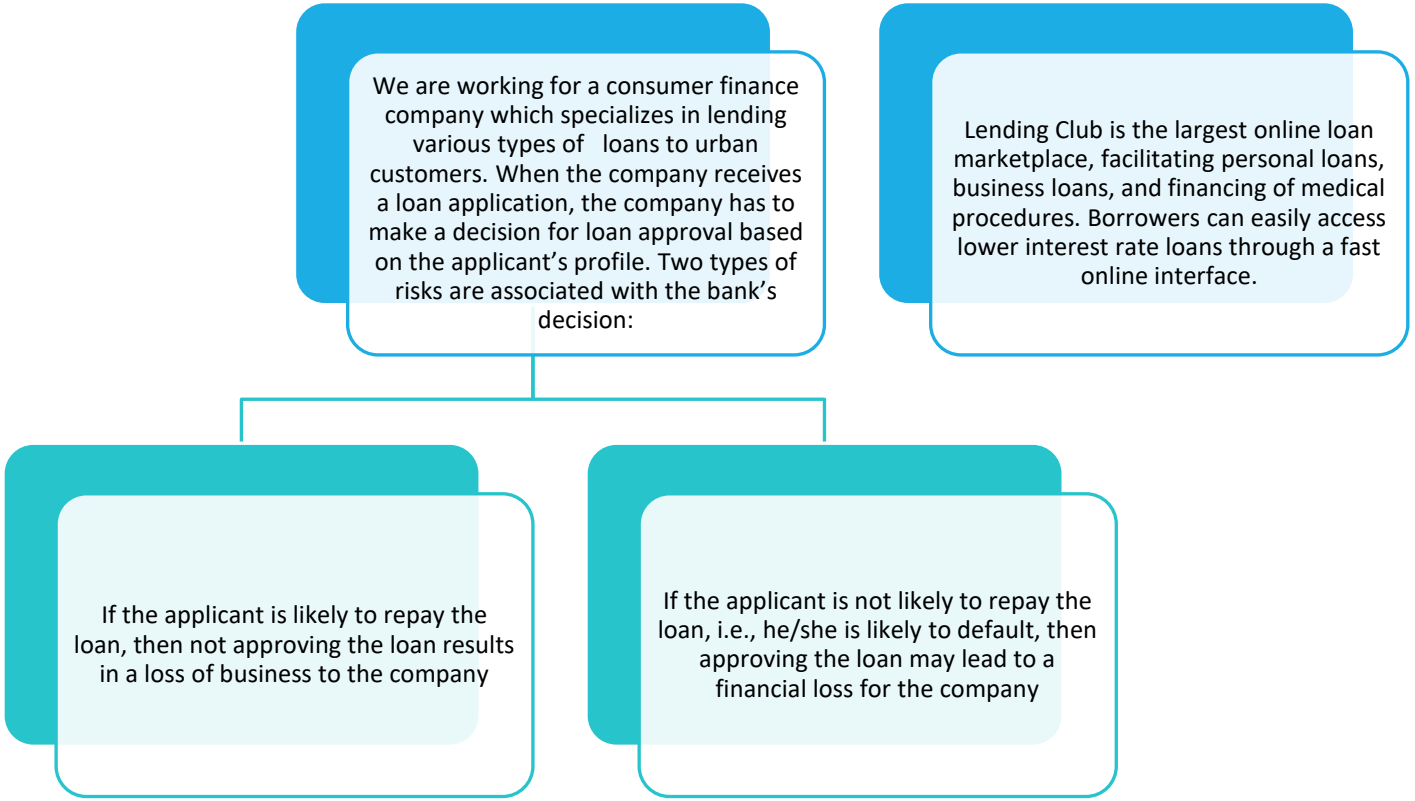
KUMAR GAURAV

YOGESHWARA RANJITH

Objective

This study will help in analyzing how real business problems are solved using Exploratory Data Analysis(EDA). It will also help in understanding risk analytics in banking and financial services and understanding how data is used to minimize the risk of losing money while lending to customers.

Project Information



Business Objective

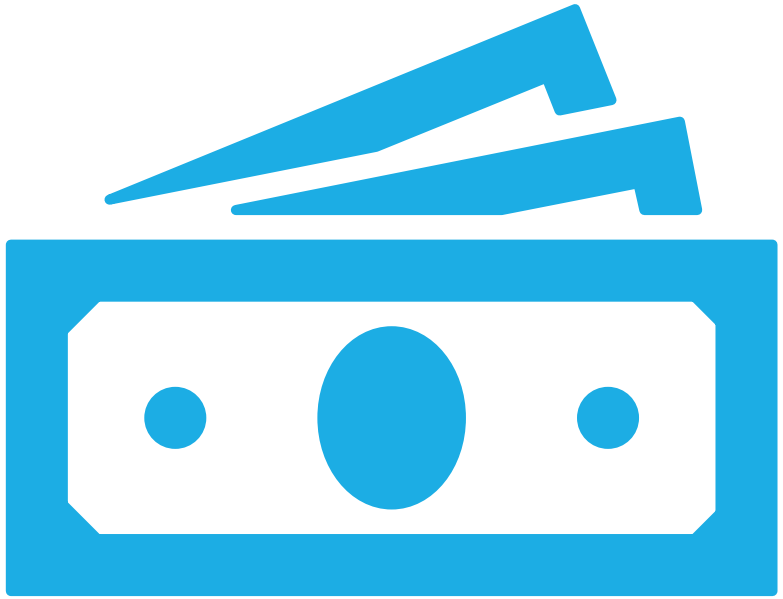
Since lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).In this case, the customers labelled as 'charged-off' are the 'defaulters'. The company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Dataset:

The given dataset is a .csv file containing the loan details of all the loans issued between 2007 to 2011.

Data Analysis and Cleanup

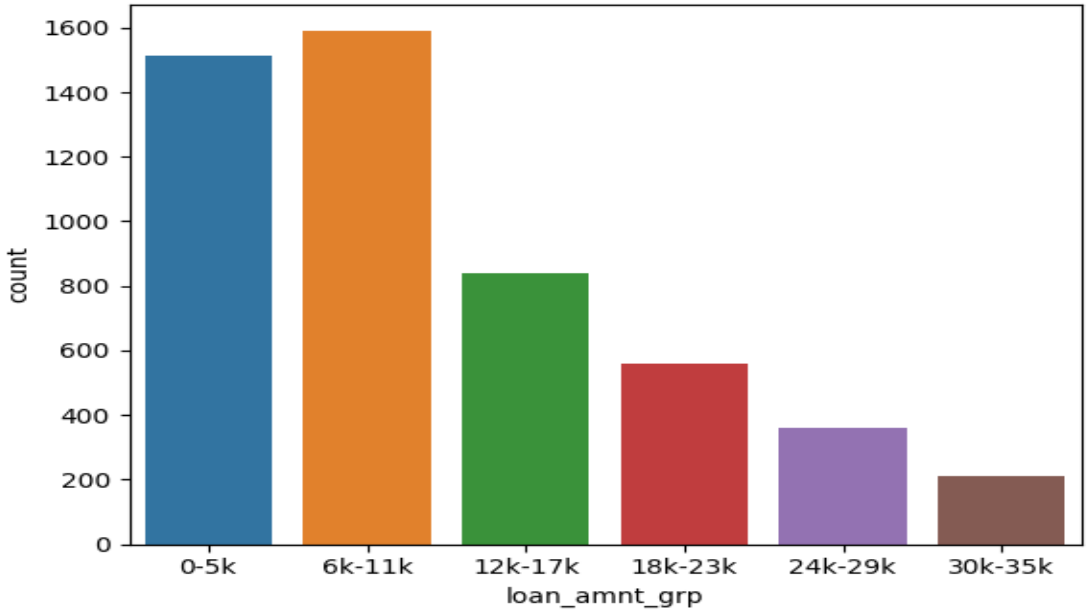
1. Check for missing values and handle them
2. Identify Categorical and numerical variables
3. Visualize univariate distributions
4. Investigate potential outliers
5. Analyze bivariate relationships
6. Visualize multivariate relationships
7. Perform feature engineering
8. Identify pattern and insights



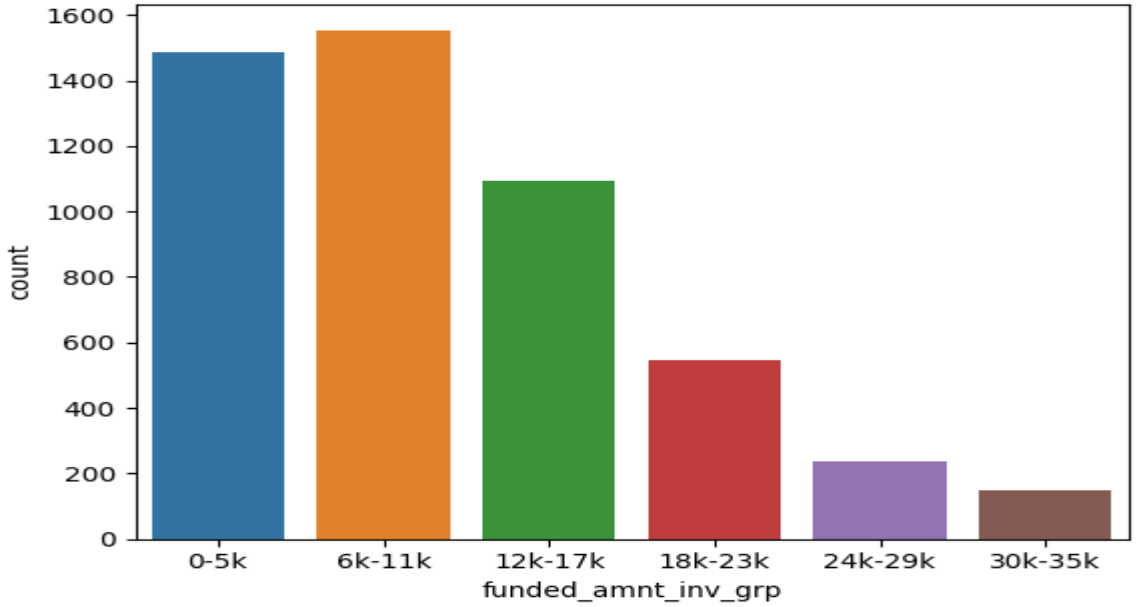
CHARGED OFF Loans Visualization

UNIVARIATE ANALYSIS

Loan Amount Group and Funded Amount Inv Group

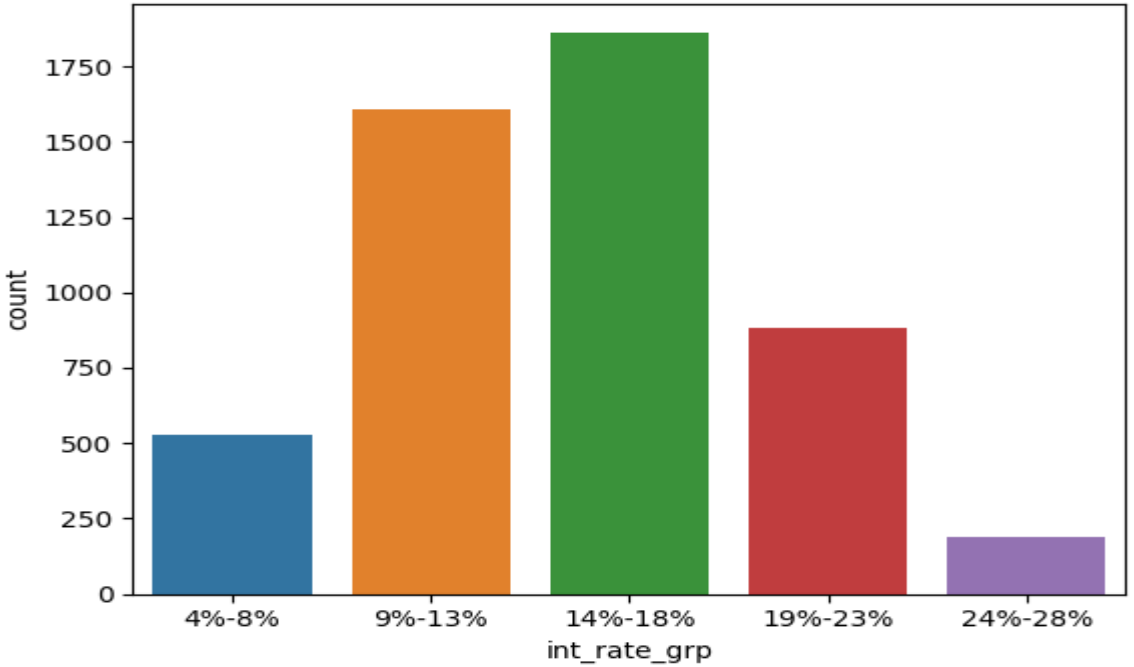


Loan Amount Group: When the loan amount is between 6k to 11k then the loan is most likely to default



Funded Amount Inv Group: When the funding amount by investor is between 6k to 11k then the loan is most likely to default

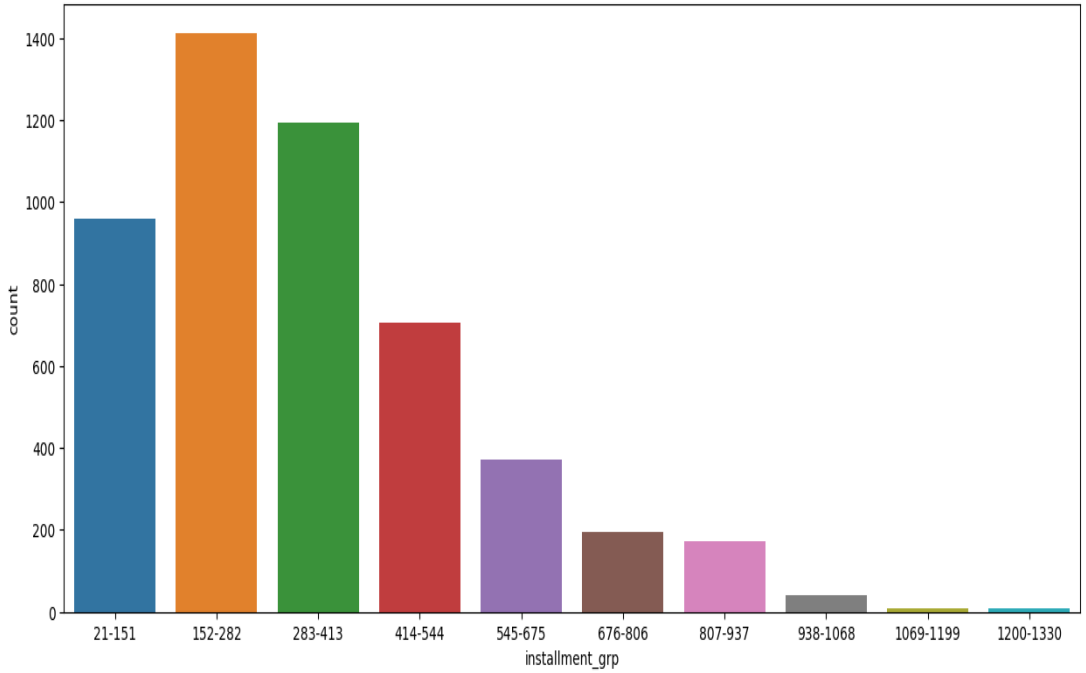
Loan Term and Interest Rate Group

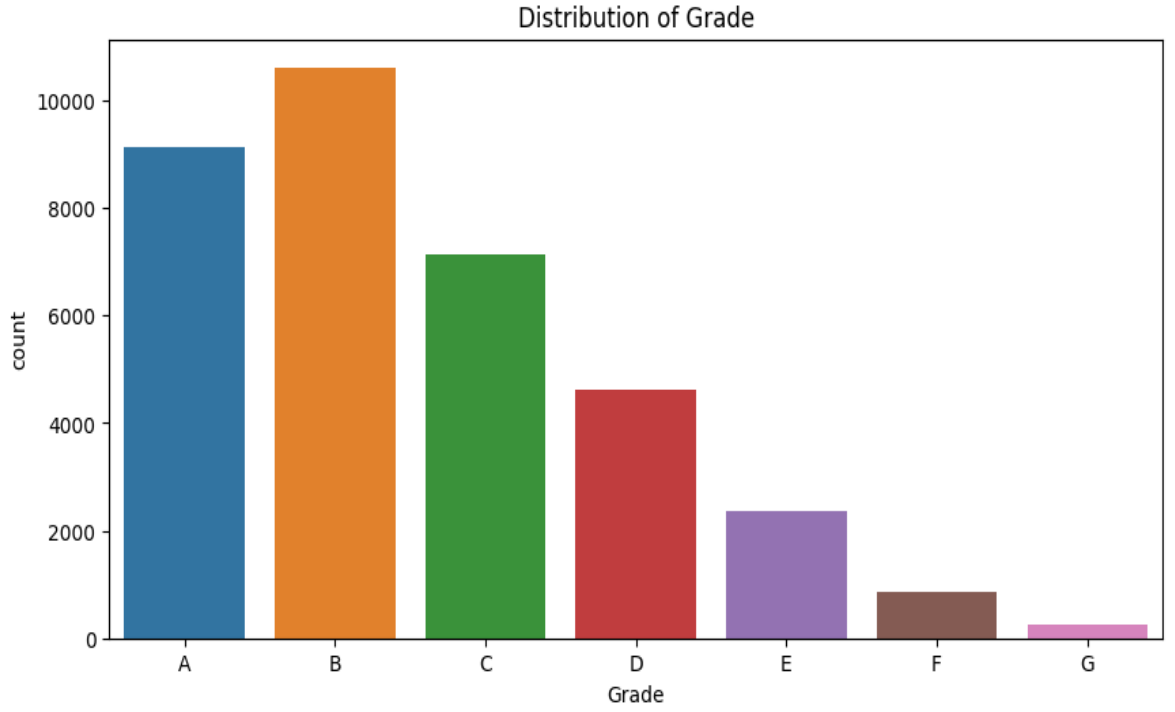


Loan Term: The loans with 36 months term are more likely to default compared to 60 months term

Interest Rate Group: When the interest rate is between 14% to 18% then the loan is most likely to default

Installment Group and Grade

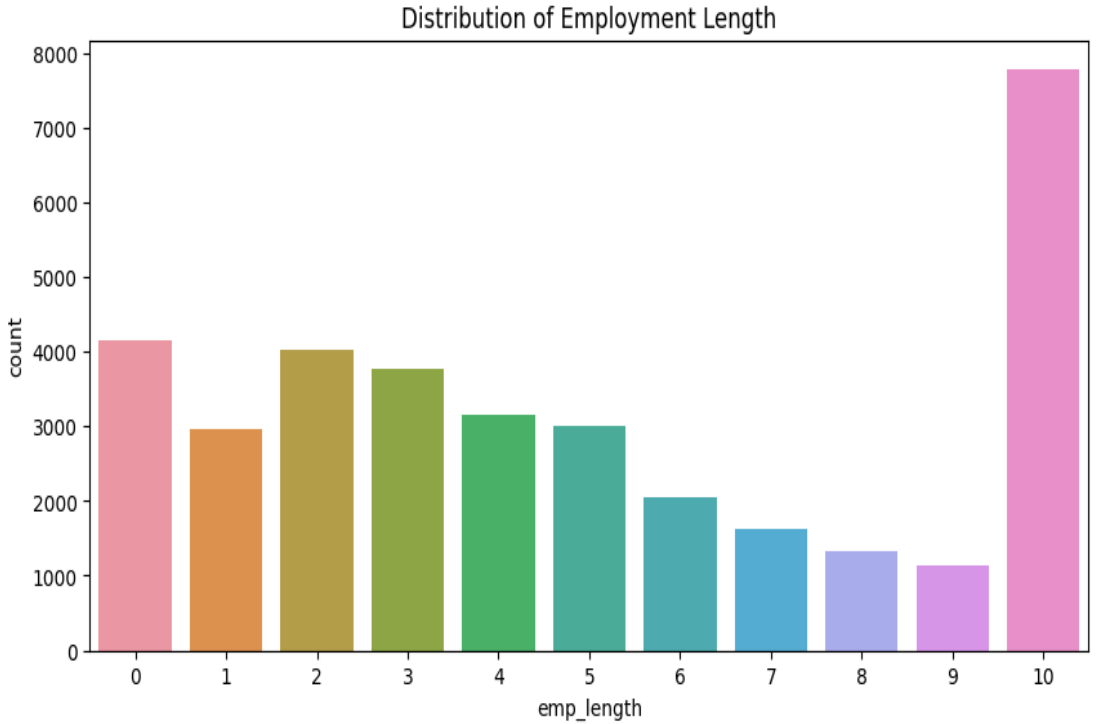


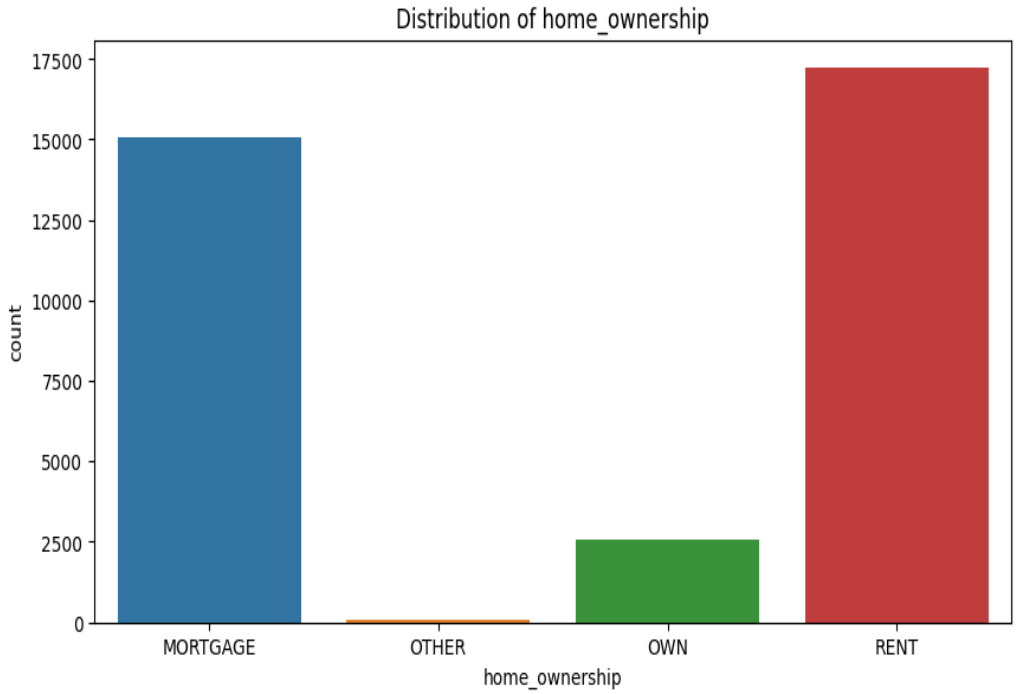


Installment Group: When the installment is between 152 to 282 then the loan is most likely to default

Grade: When the Grade is B then the loan is most likely to default

Employment Length and Home Ownership

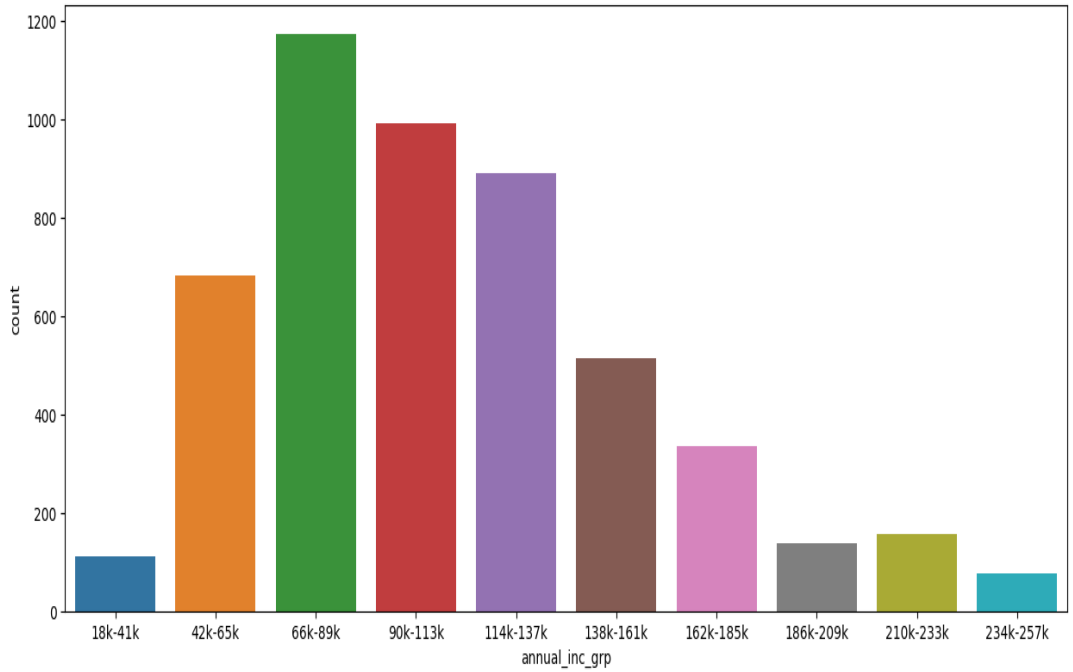


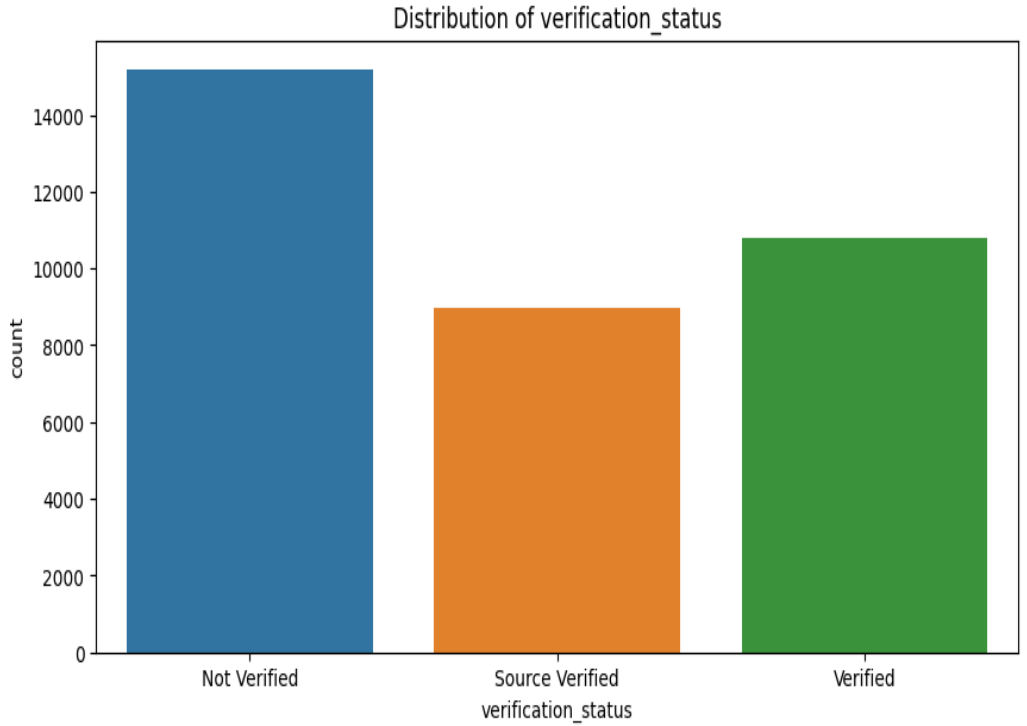


Employment Length: When the Employment length is 10 years then the loan is most likely to default

Home Ownership: When the home ownership is rented then the loan is most likely to default

Annual Income Group and Verification Status

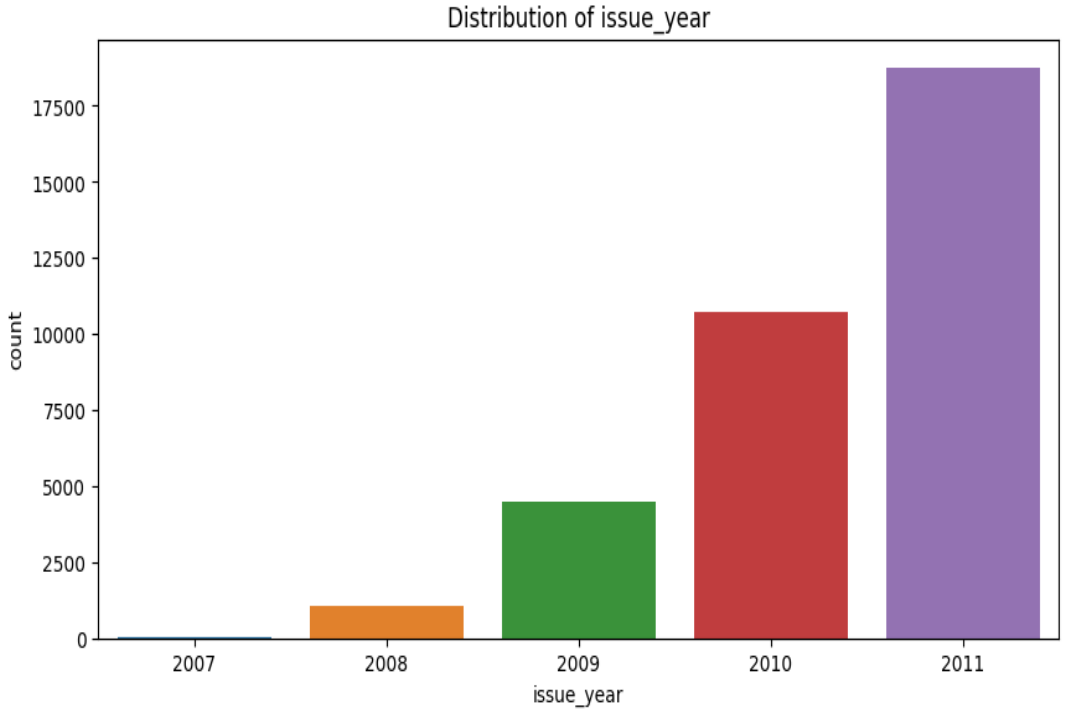




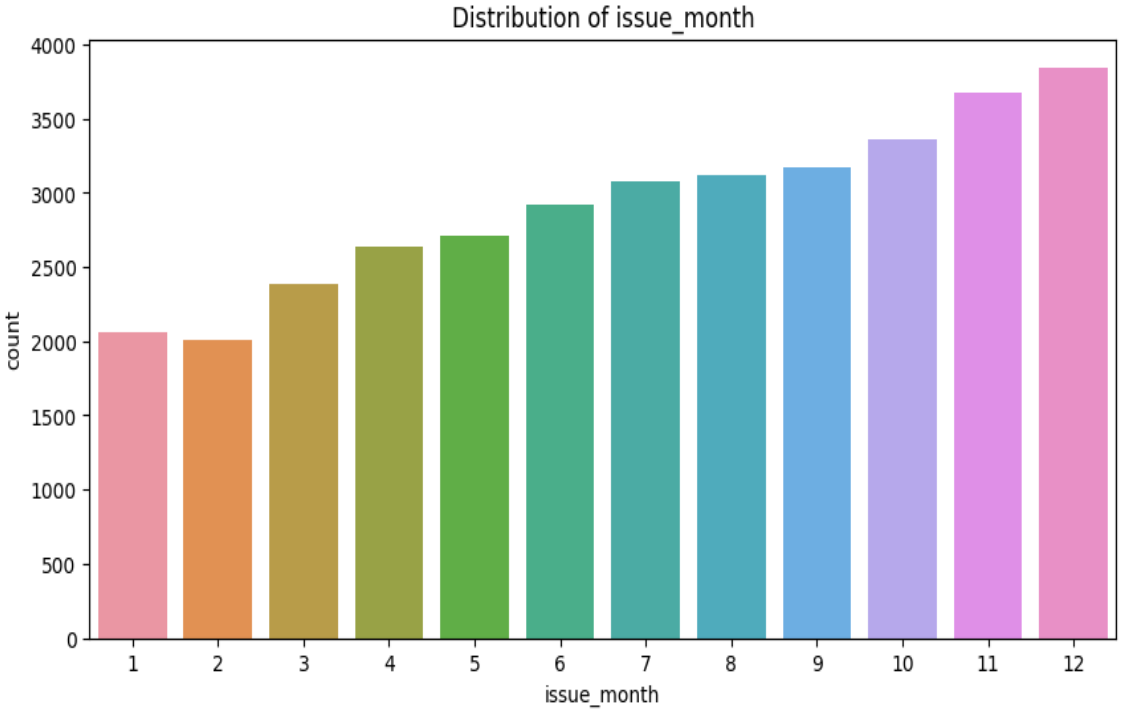
Annual Income Group: When the annual income is between 66k to 89k then the loan is most likely to default

Verification Status: If the income is not verified by LC then the loan is most likely to default

Issue Year and Issue Month

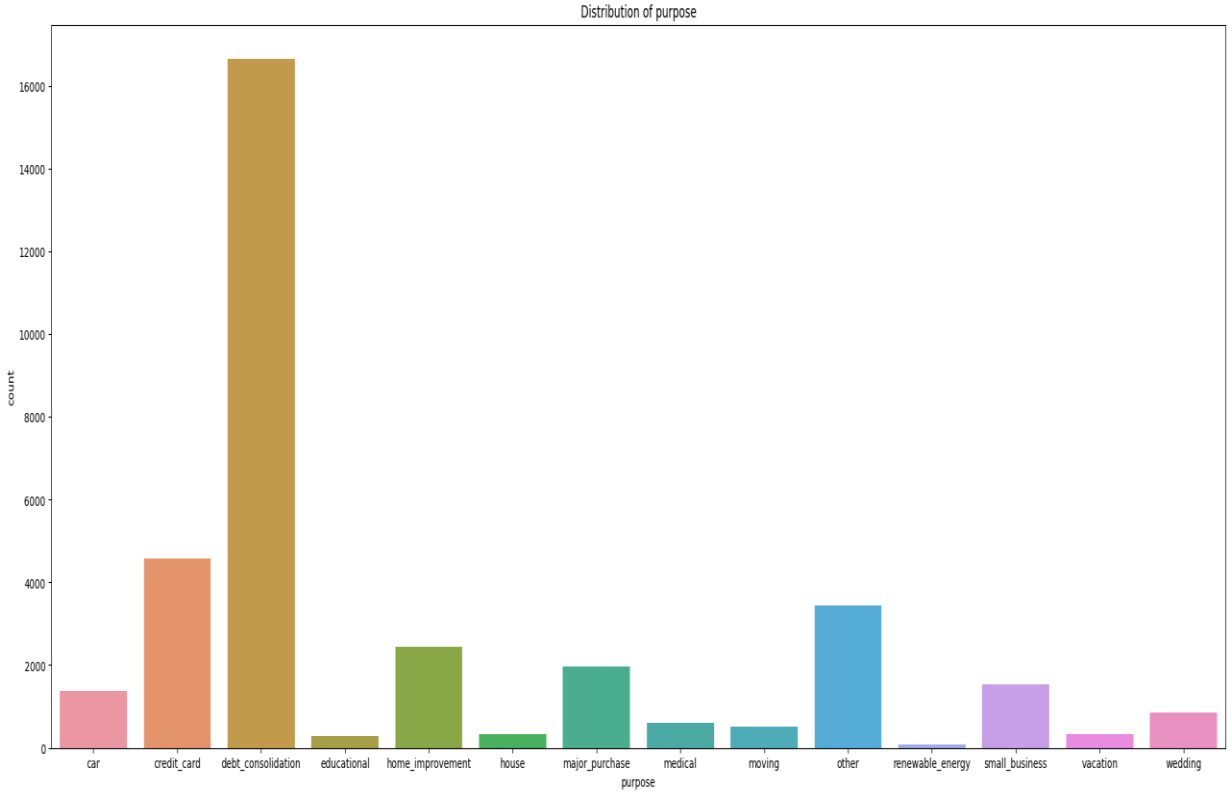


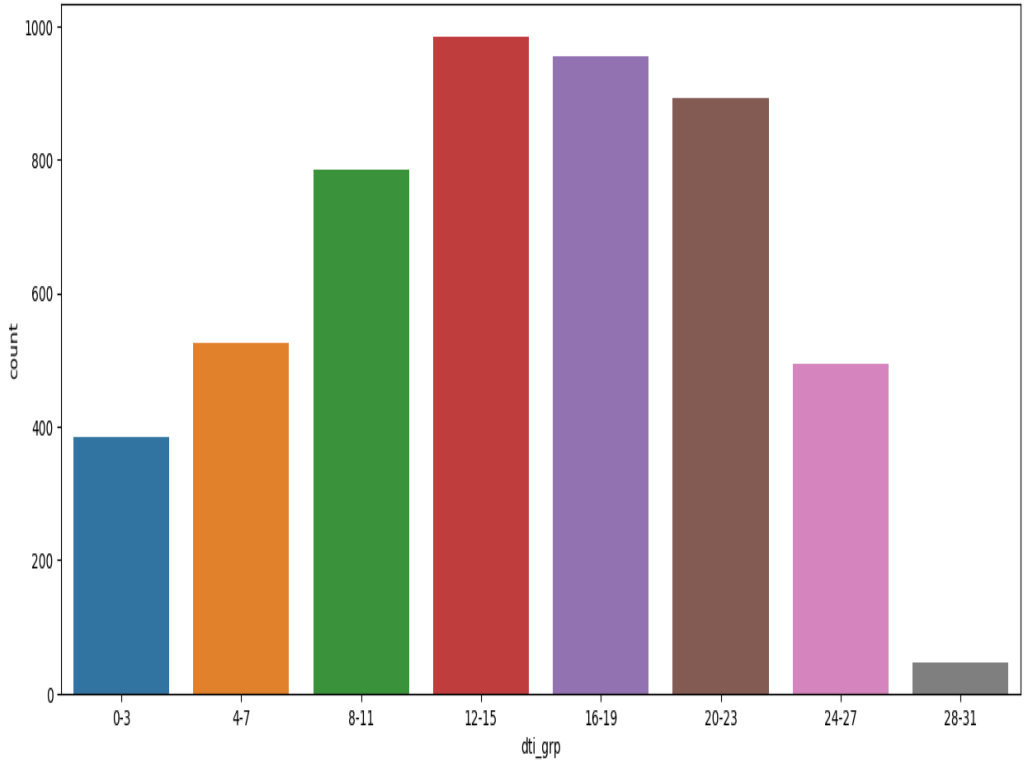
Issue Year: Large number loans defaulted in 2011



Issue Month: Large number of loans defaulted in dec month

Purpose and Dti Group

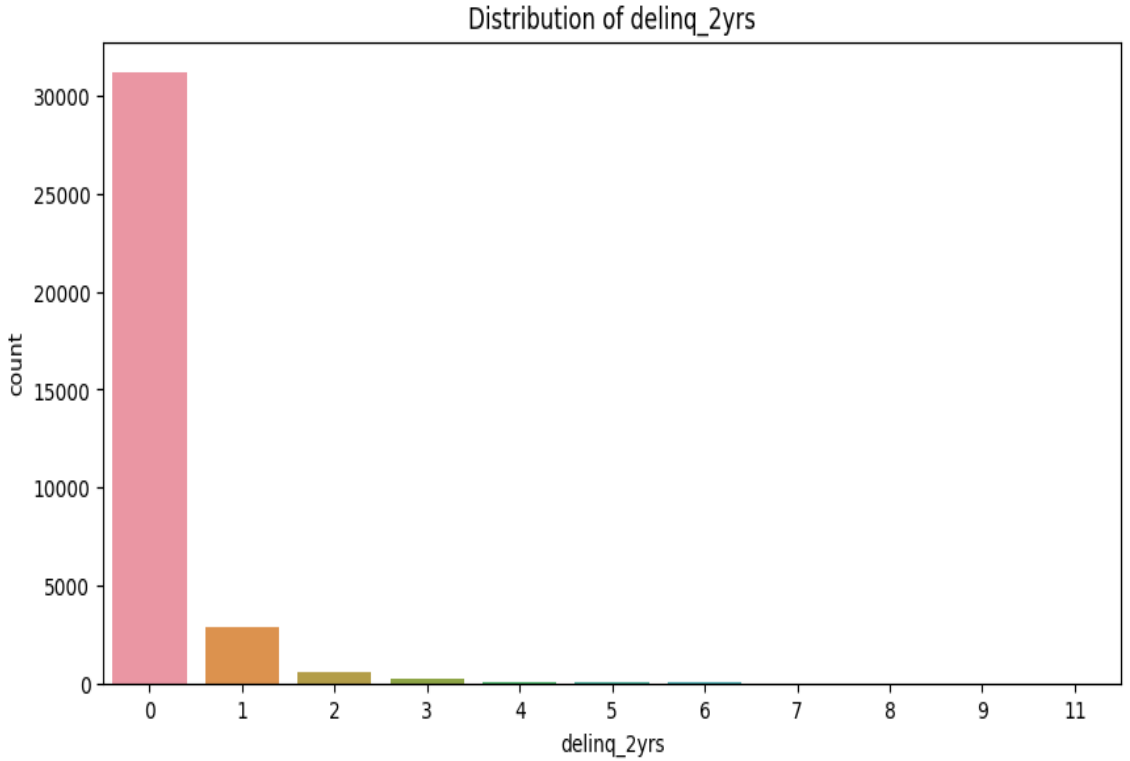


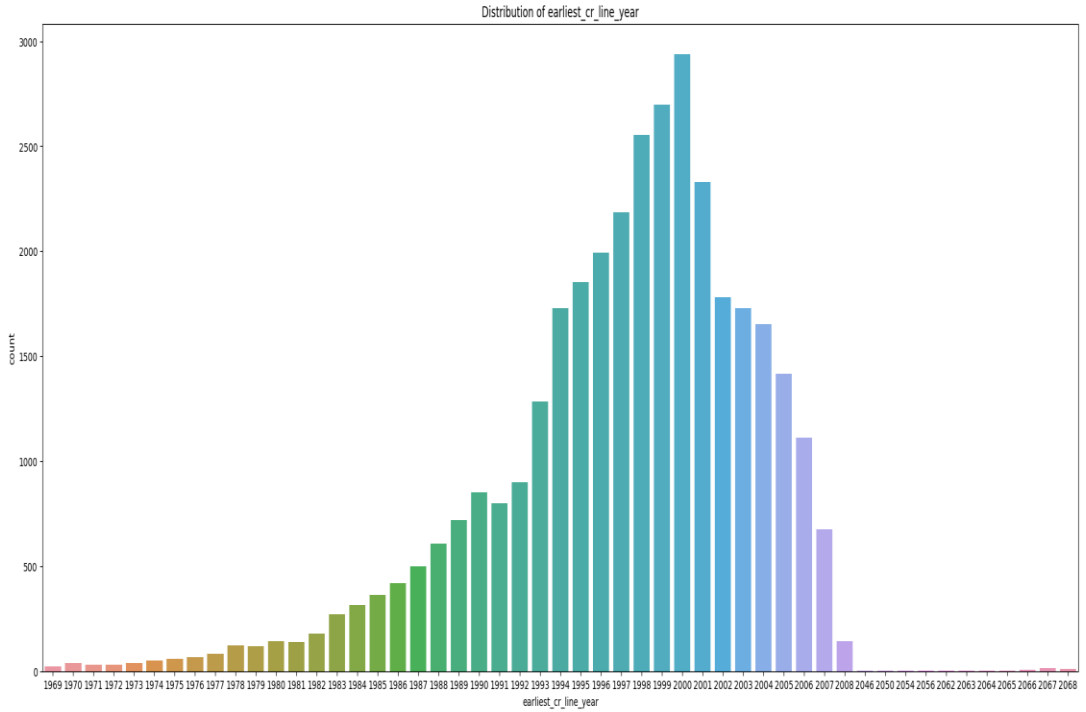


Purpose: When the loan purpose is debt_consolidation then the loan is most likely to default

Dti Group: When dti is between 12-15 the loan is most likely to default

Delinq 2 years and Earliest_CR_Line_Year

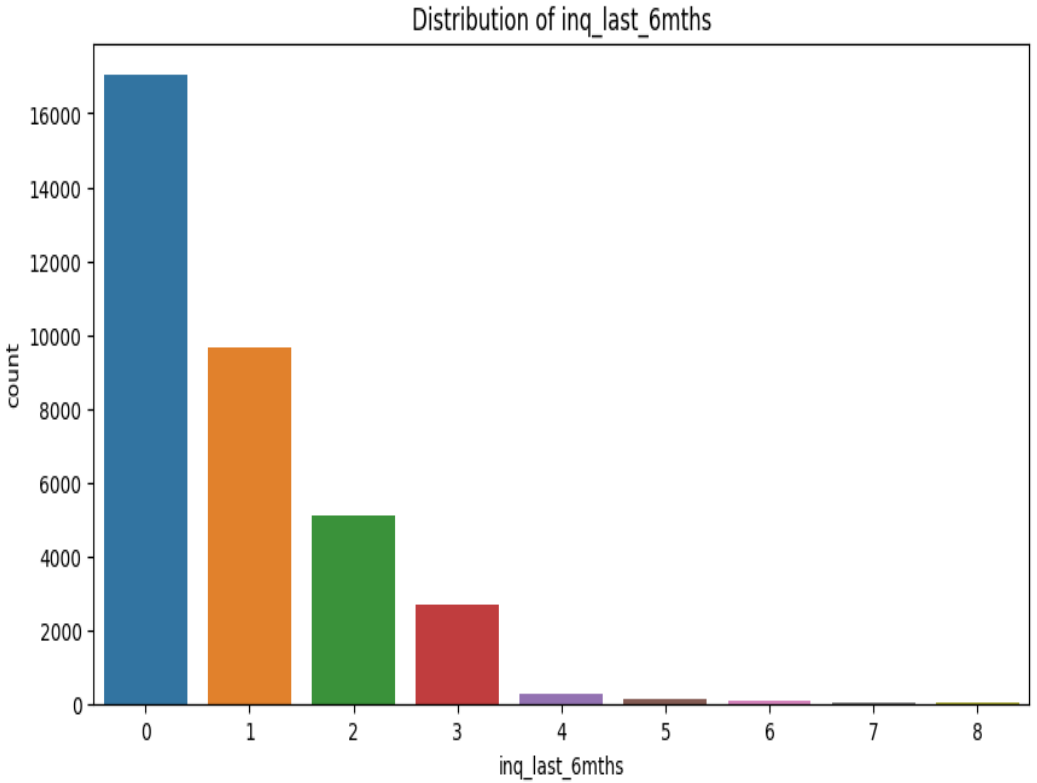


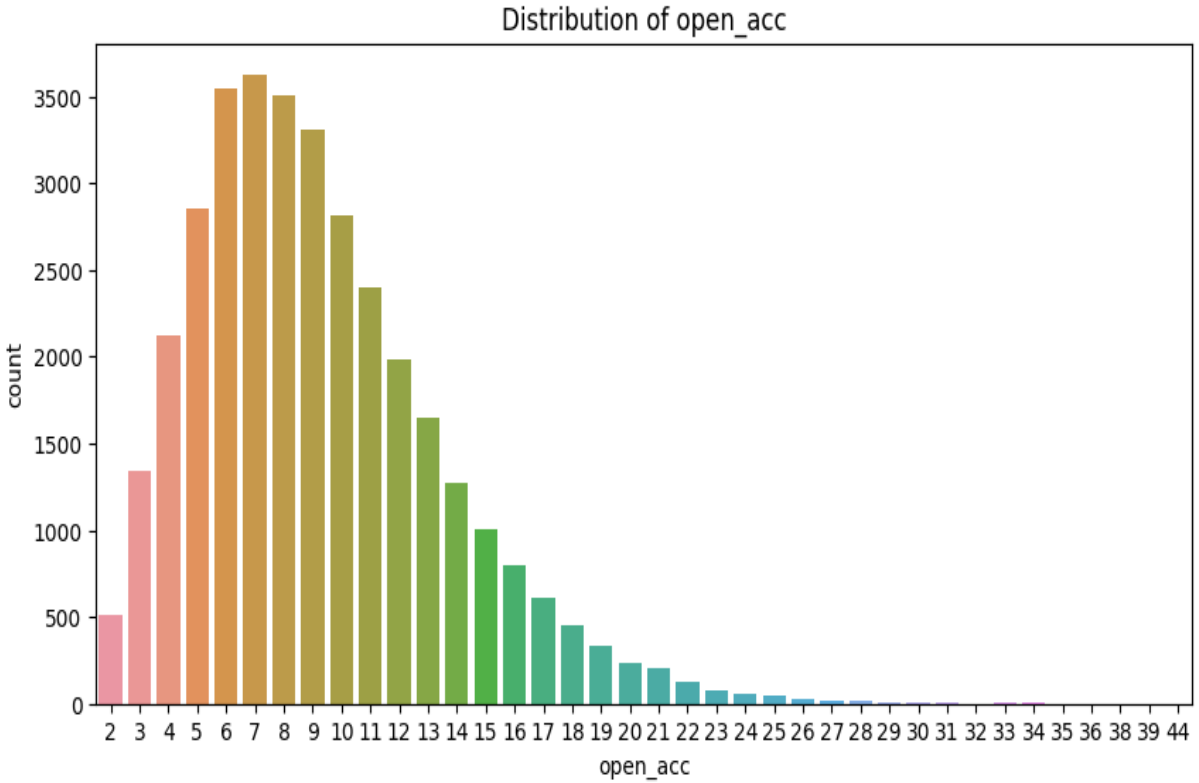


Delinq 2 : For the defaulted loans delinq_2yrs is mostly 0

Earliest_CR_Line_Year: If earliest_cr_line_year is between 1994 to 2004 then the loan is most likely to default

Inquiry Last 6 Months and Open_Acc

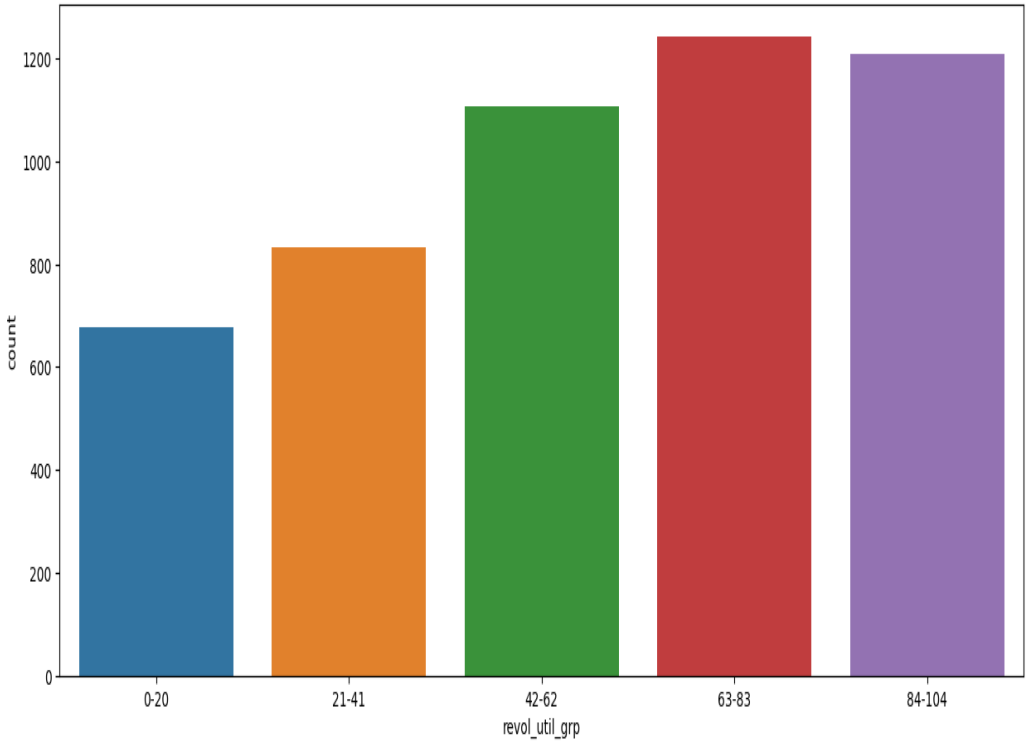


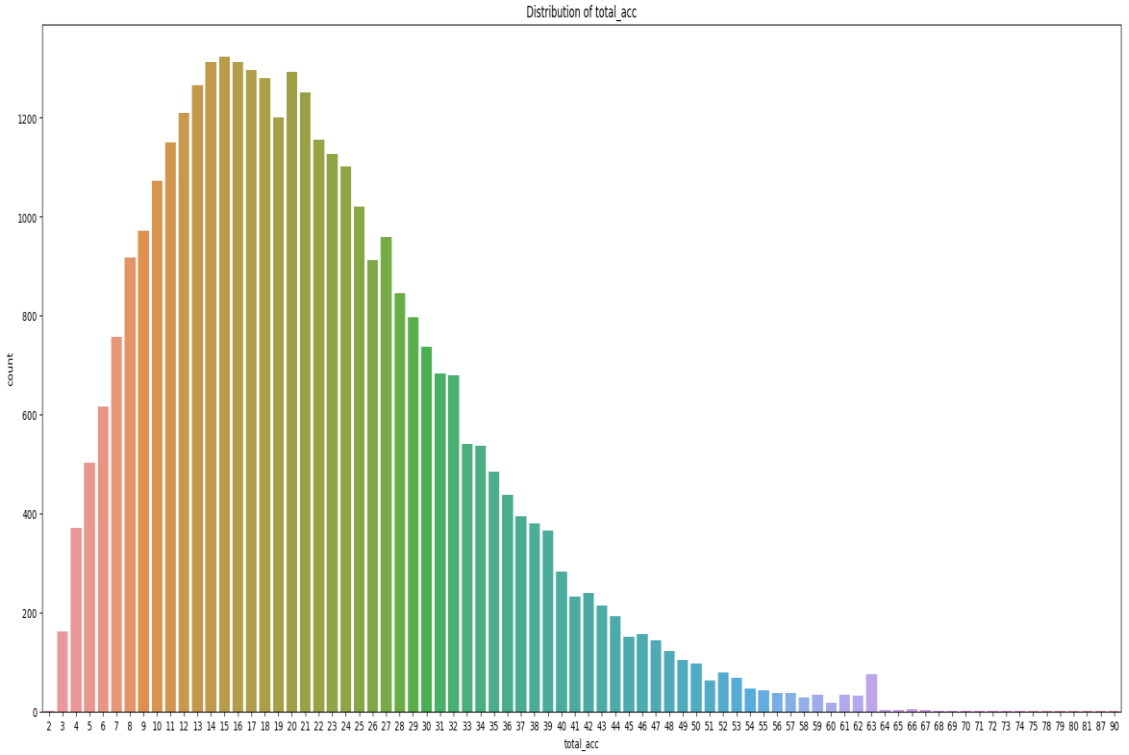


Inquiry Last 6 Months: If inq_last_6mths is 0 then the loan is most likely to default

Open Acc: If the open_acc is between 5-10 then the loan is most likely to default

Revol_Util Group and Total_Acc





Revol_Util Group: If revol_util range is between 63-83, then the loan is most likely to default, but it is not significant

Total_Acc: If the total_acc is between 8-27 then the loan is likely to default

Observations From Univariate Analysis

1.Charged Off (Defaulted) loans are less compared to Fully Paid loans.

2.Applicants with 36 months loan term are more likely to default compared to 60 months loan term

3.Applicants are most likely to default when the intrest rate is 14% to 18%

4.When the installment is between 152 to 282 then the applicants are most likely to default.

5.The applicants are most likely to default when the grade is B

6.When the Employment length of applicants is 10 years then the applicants are most likely to default.

7.Applicants are most likely to default when the home ownership is rented.

8.If the income is not verified by Lending Club then the loan is most likely to default.

Observations From Univariate Analysis

9.Large number loans defaulted in 2011.

10.When the loan purpose is debt_consolidation then the applicants are most likely to default.

11.For the defaulted loans delinq_2yrs is mostly 0.

12.If the open_acc of applicant is between 5-10 then it is most likely to default.

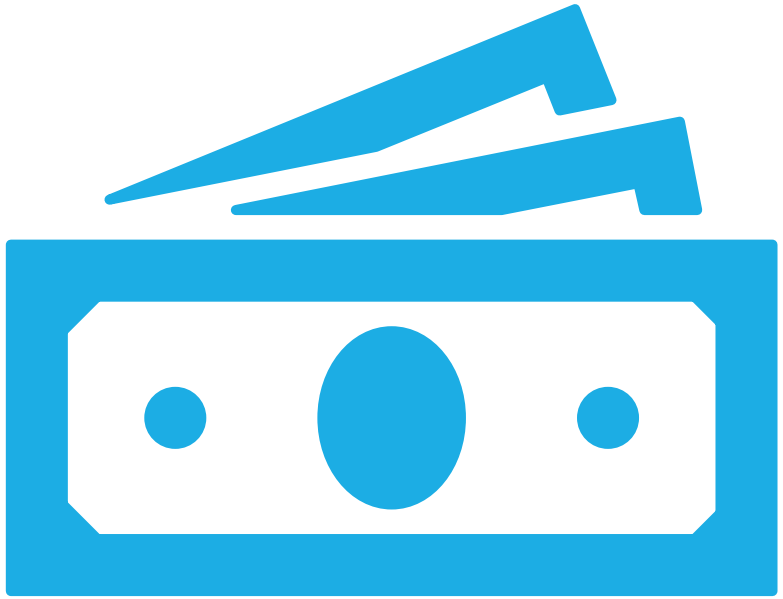
13.When dti is between 12-15 the loan is most likely to default, but this not much significant.

14.If the total_acc is between 8-27 then the loan is likely to default.

15.If inq_last_6mths is 0 then the loan is most likely to default.

16.Applicants with annual income of 66k to 89k are most likely to default.

17.If earliest_cr_line_year is between 1994 to 2004 then the applicants are most likely to default



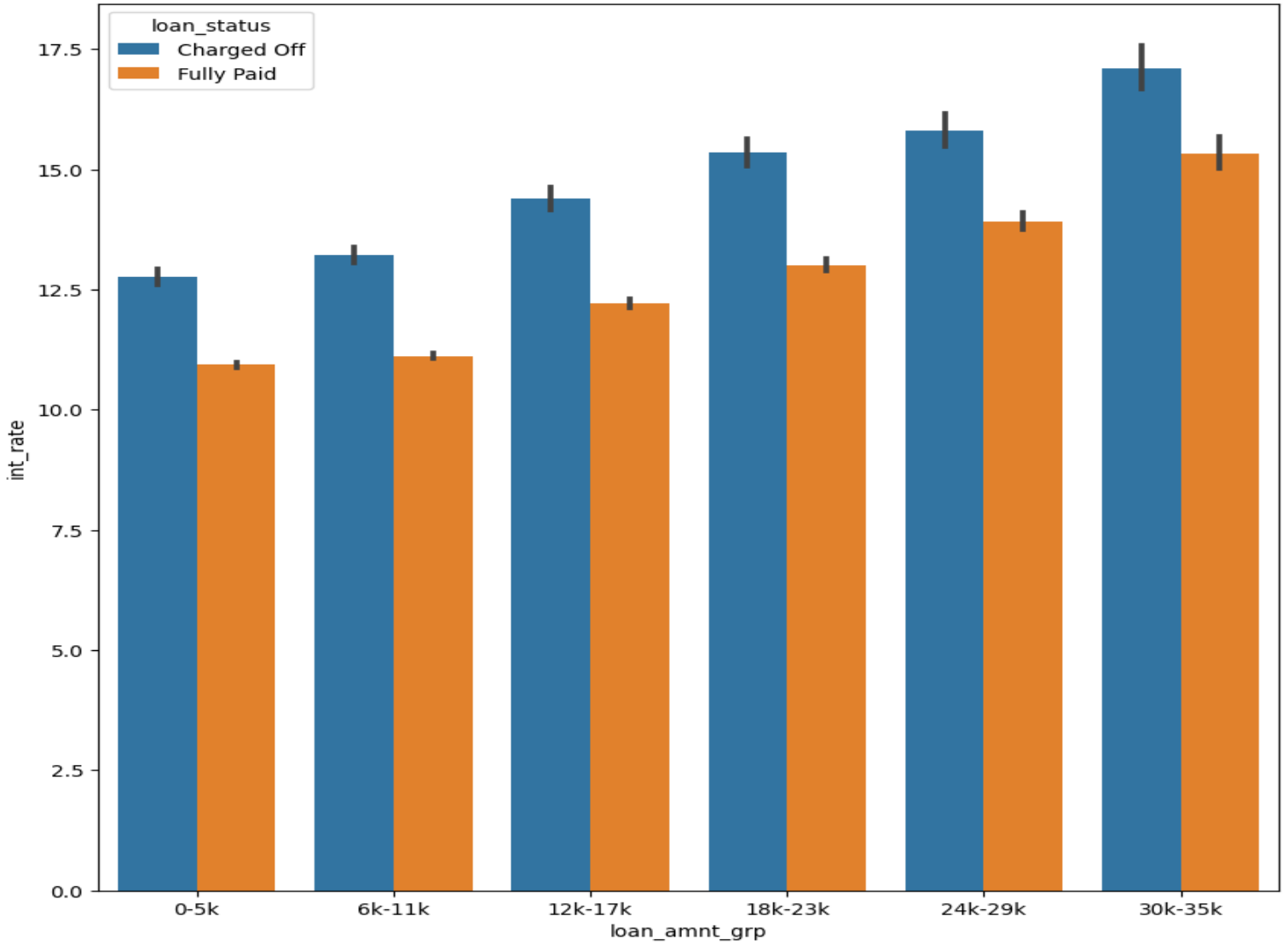
CHARGED OFF Loans Visualization

BIVARIATE ANALYSIS

Loan Amount vs Interest Rate

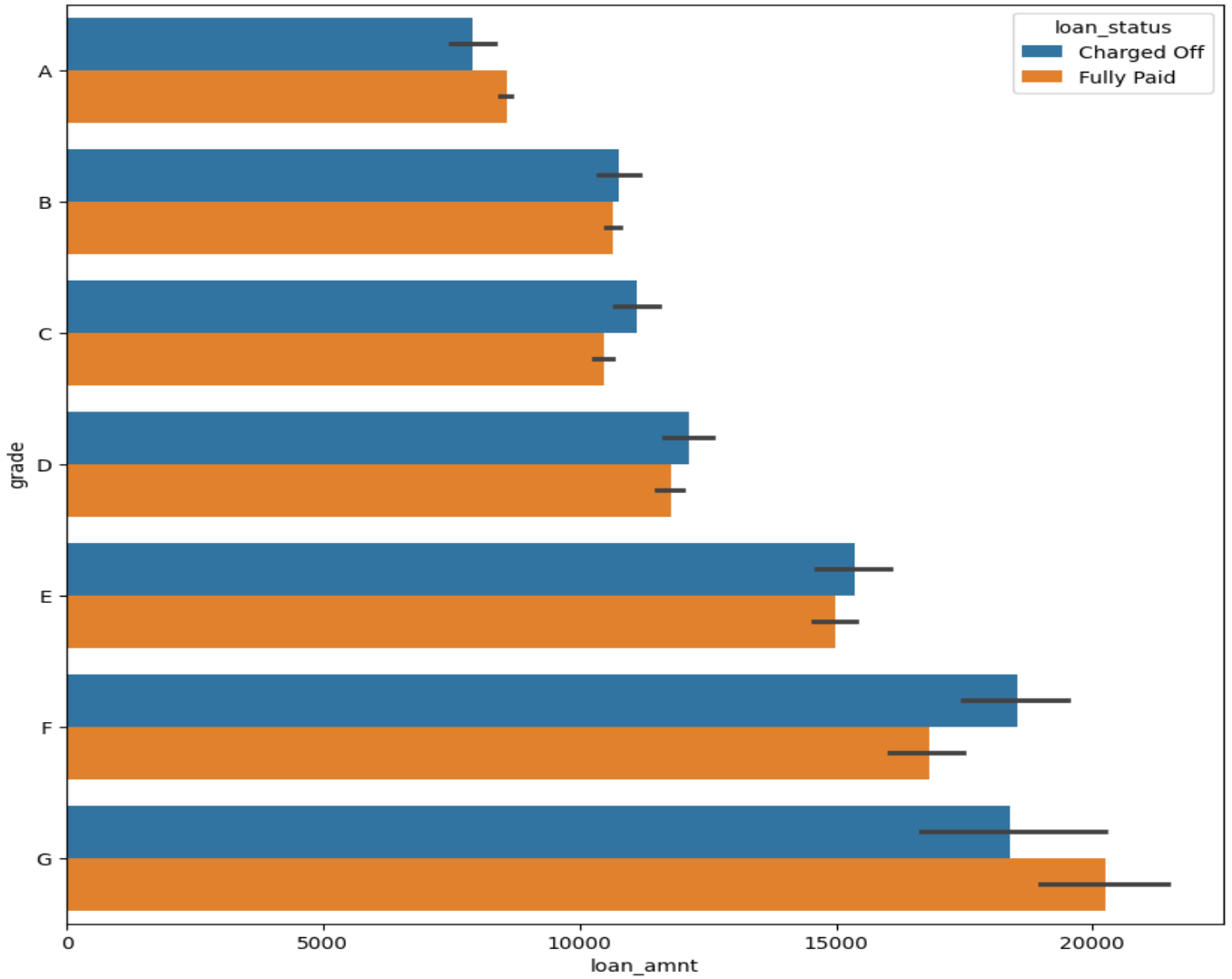
The charged off loans have higher interest rate across all loan amount groups.

So if the interest rate is high there are more chances of loan getting default



Loan Amount vs Grade

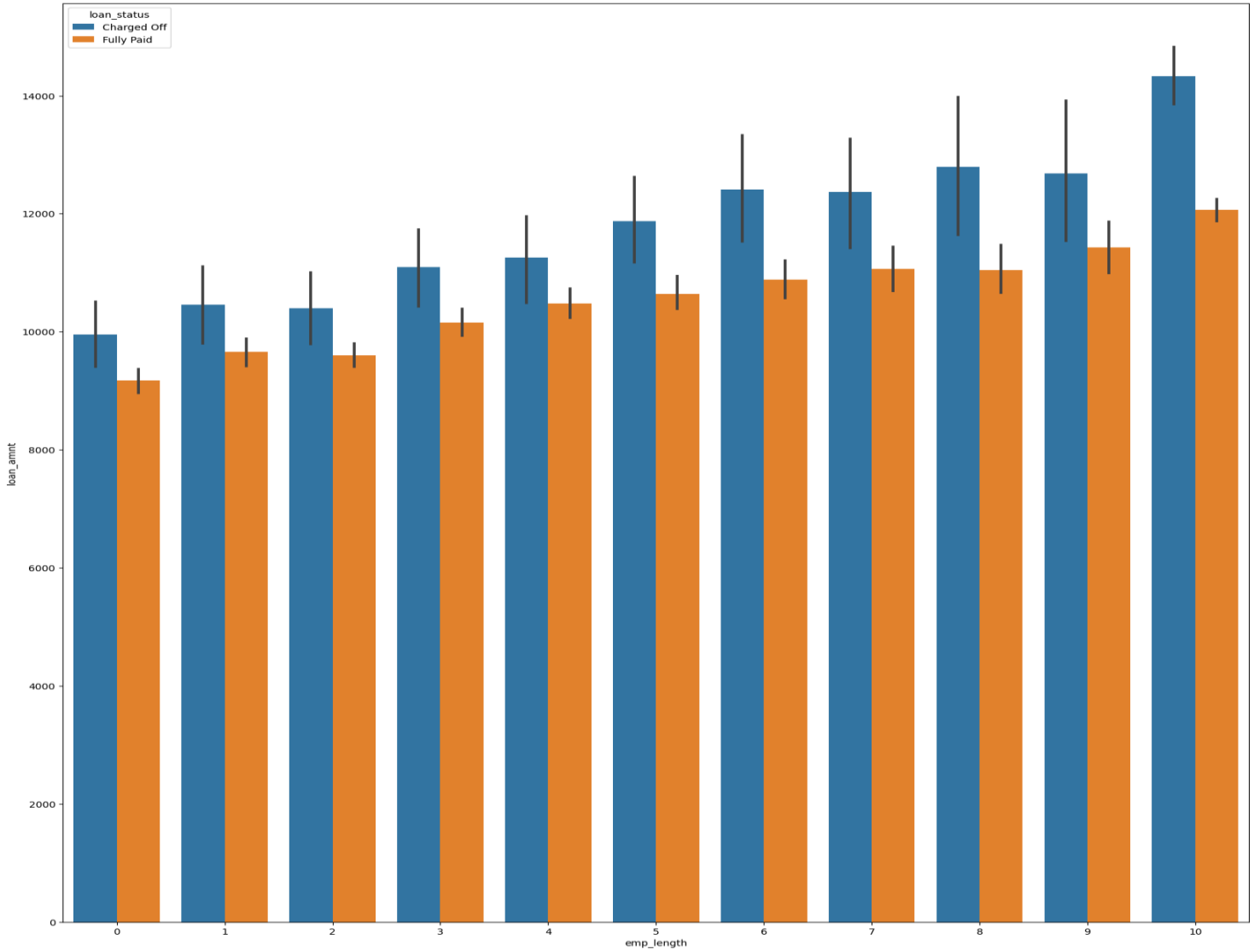
The Loan defaulting is increasing from A to G grade



Loan Amount vs Employment Length

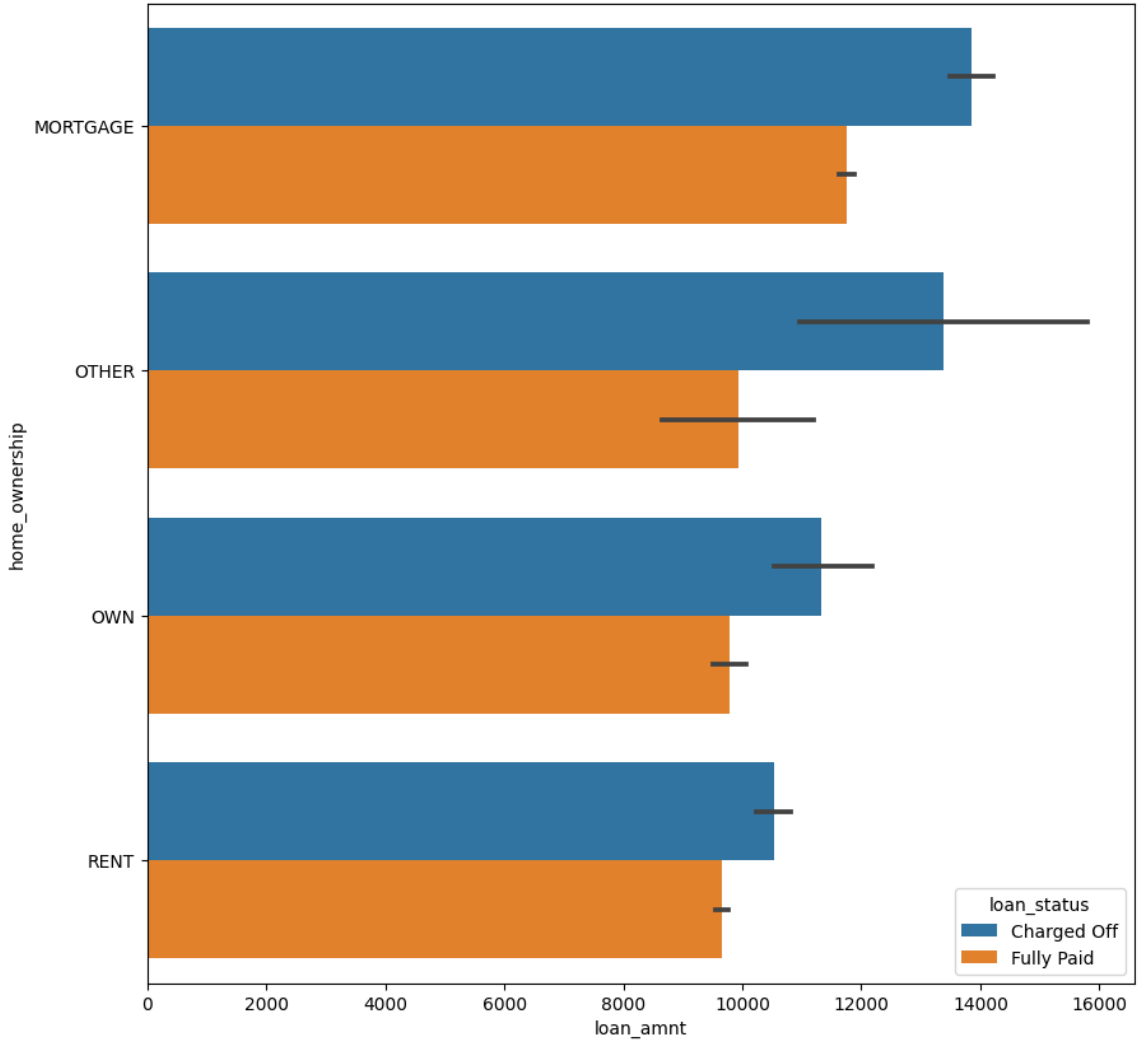
Employees having longer employment length got the high loan amount and also loan was most likely to default

So as the loan amount and employment length is increasing the risk of defaulting is also increasing



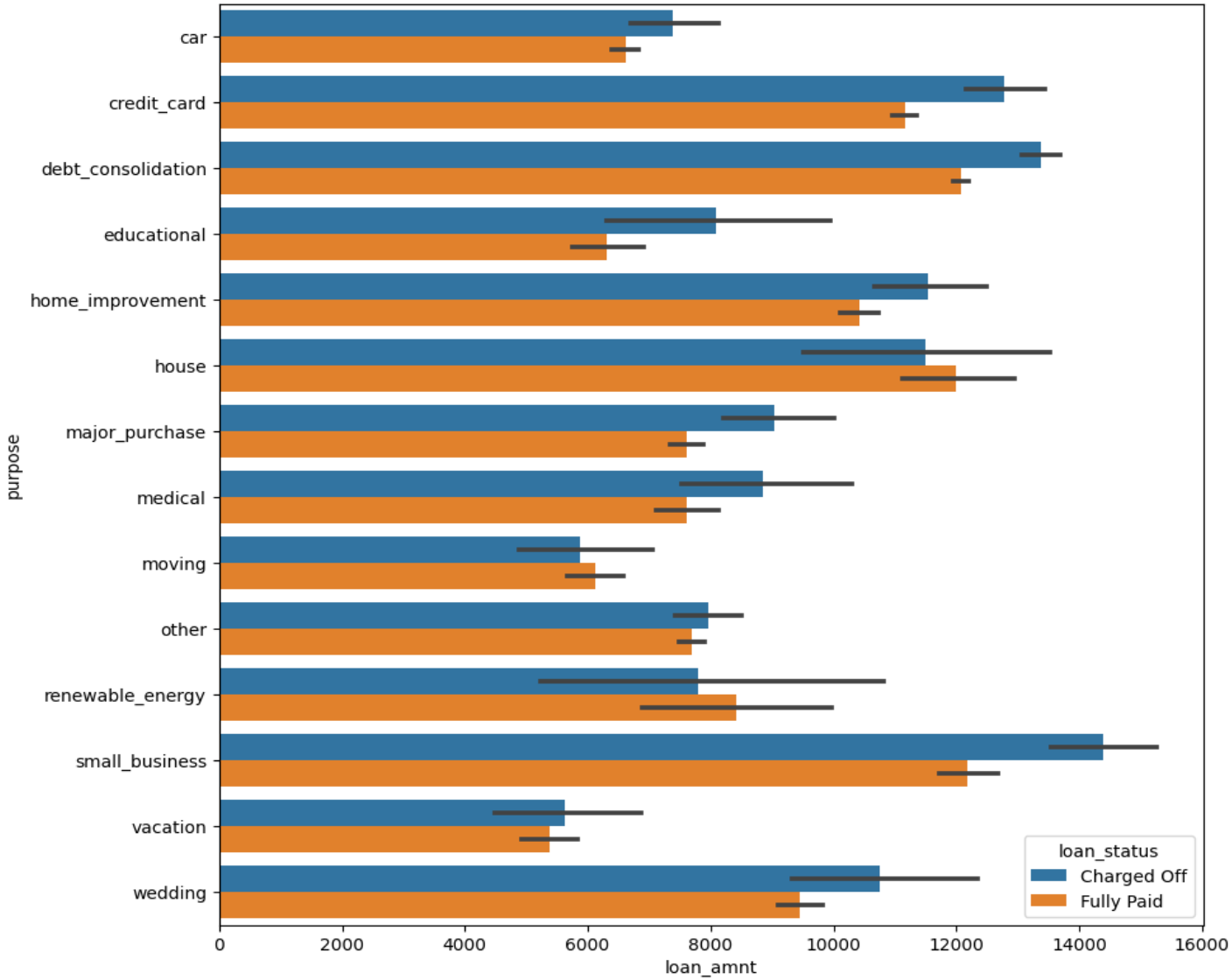
Loan Amount vs Home Ownership

Applicants whose loan amount is high are likely to default.



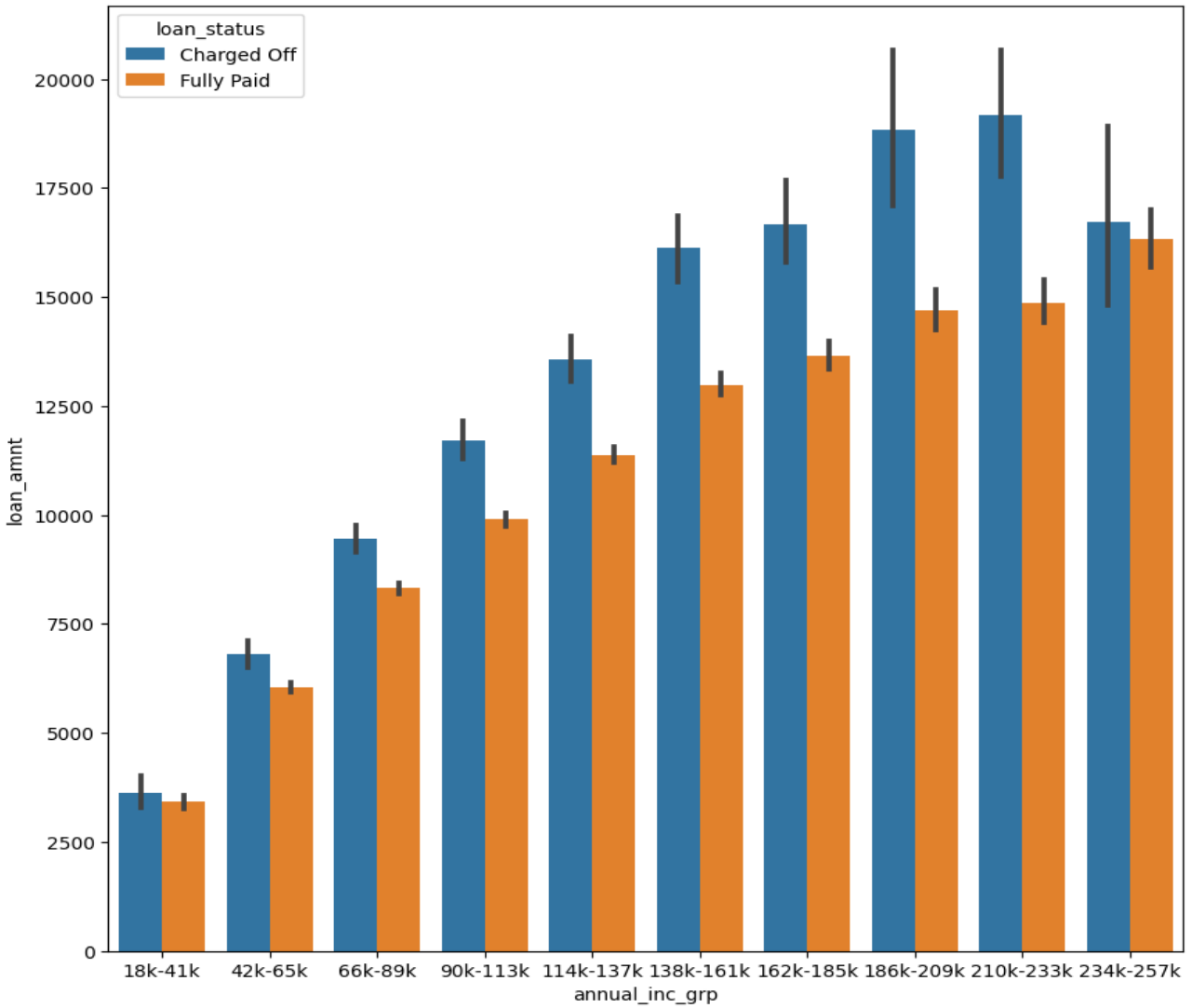
Loan Amount vs Loan Purpose

Applicants taken loan for small business and the loan amount is between 12k to 14.5k, then the applicants are more likely to default



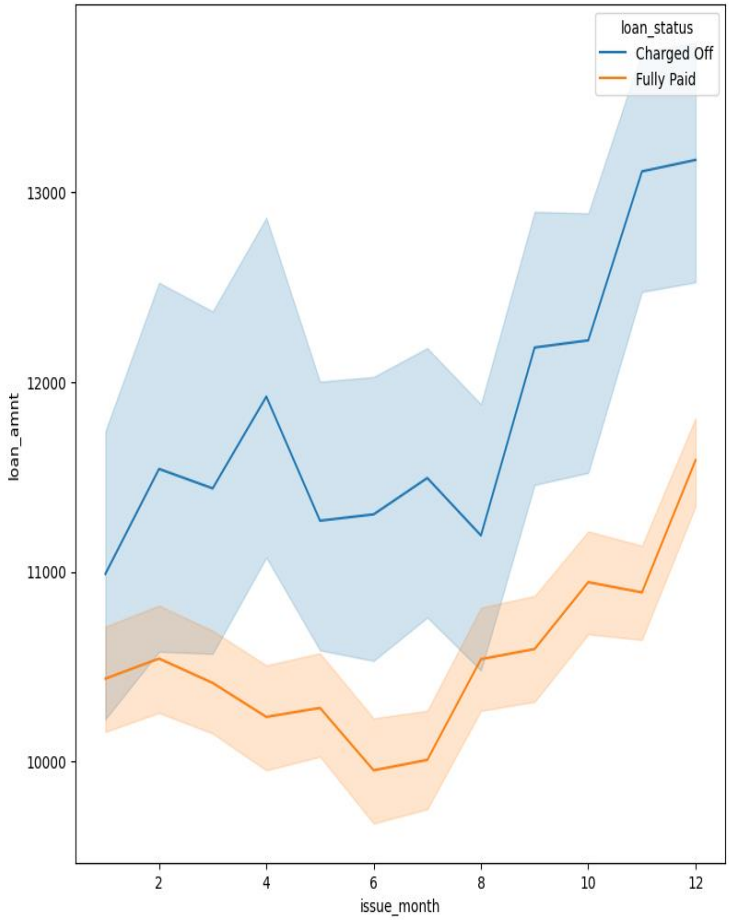
Loan Amount vs Annual Income

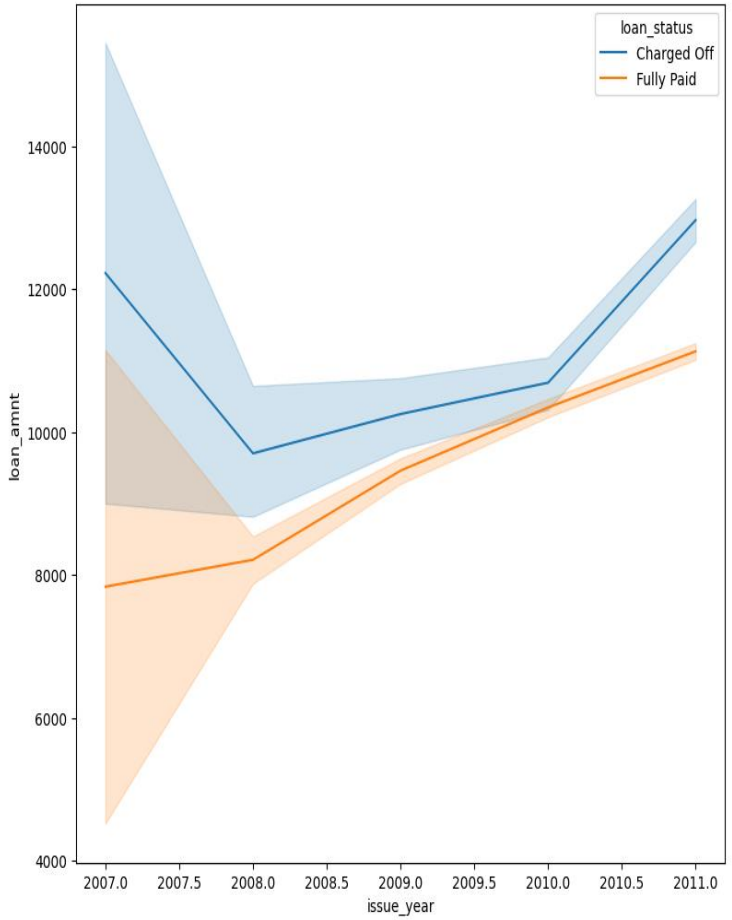
We can observe that loan amount is higher for applicants who defaulted and with annual income between 186k to 233k



Loan Amount vs Issue Month and Issue Year

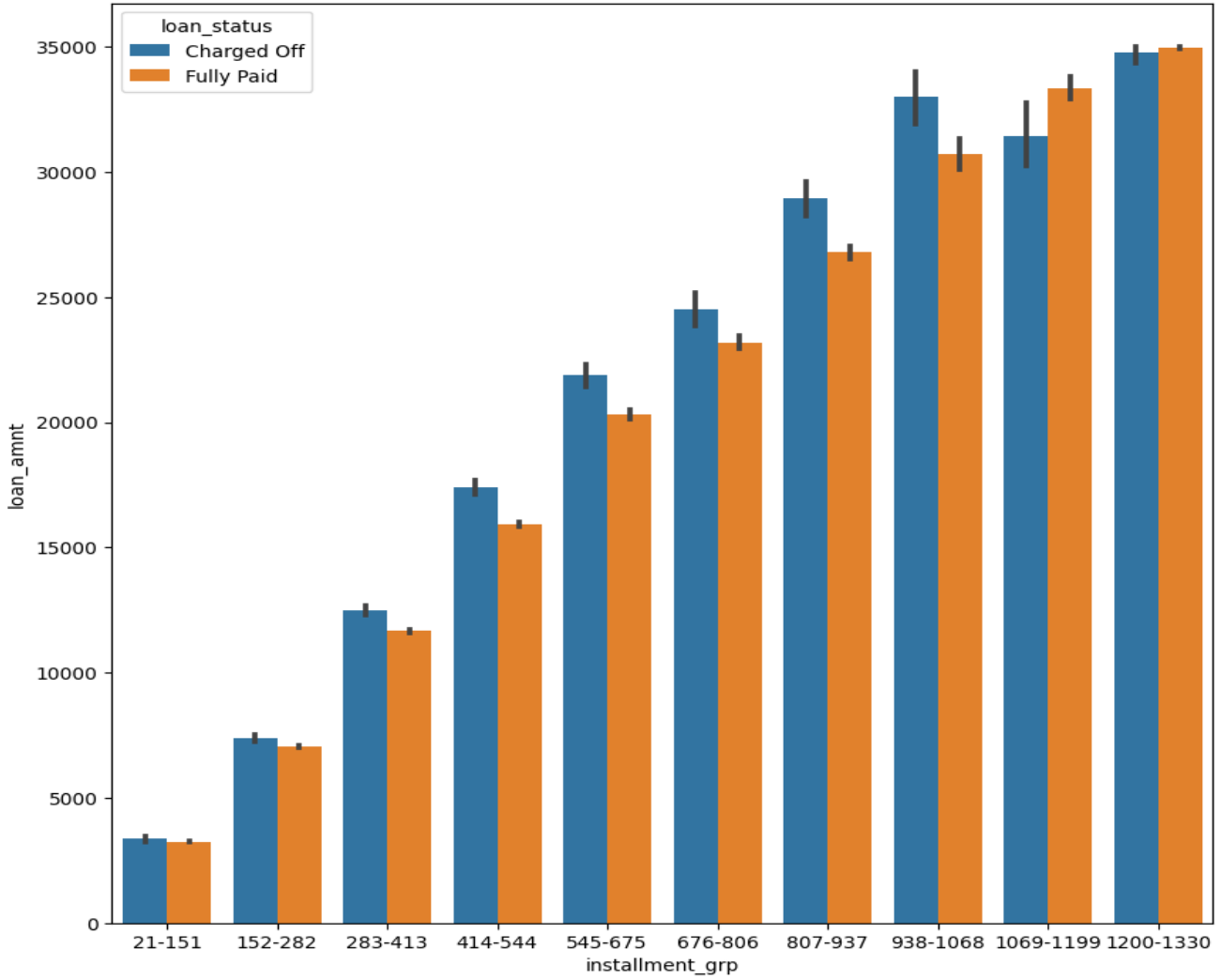
Applicants most likely to default in the months between oct to dec mostly December, applicants defaulted mostly in the year 2011





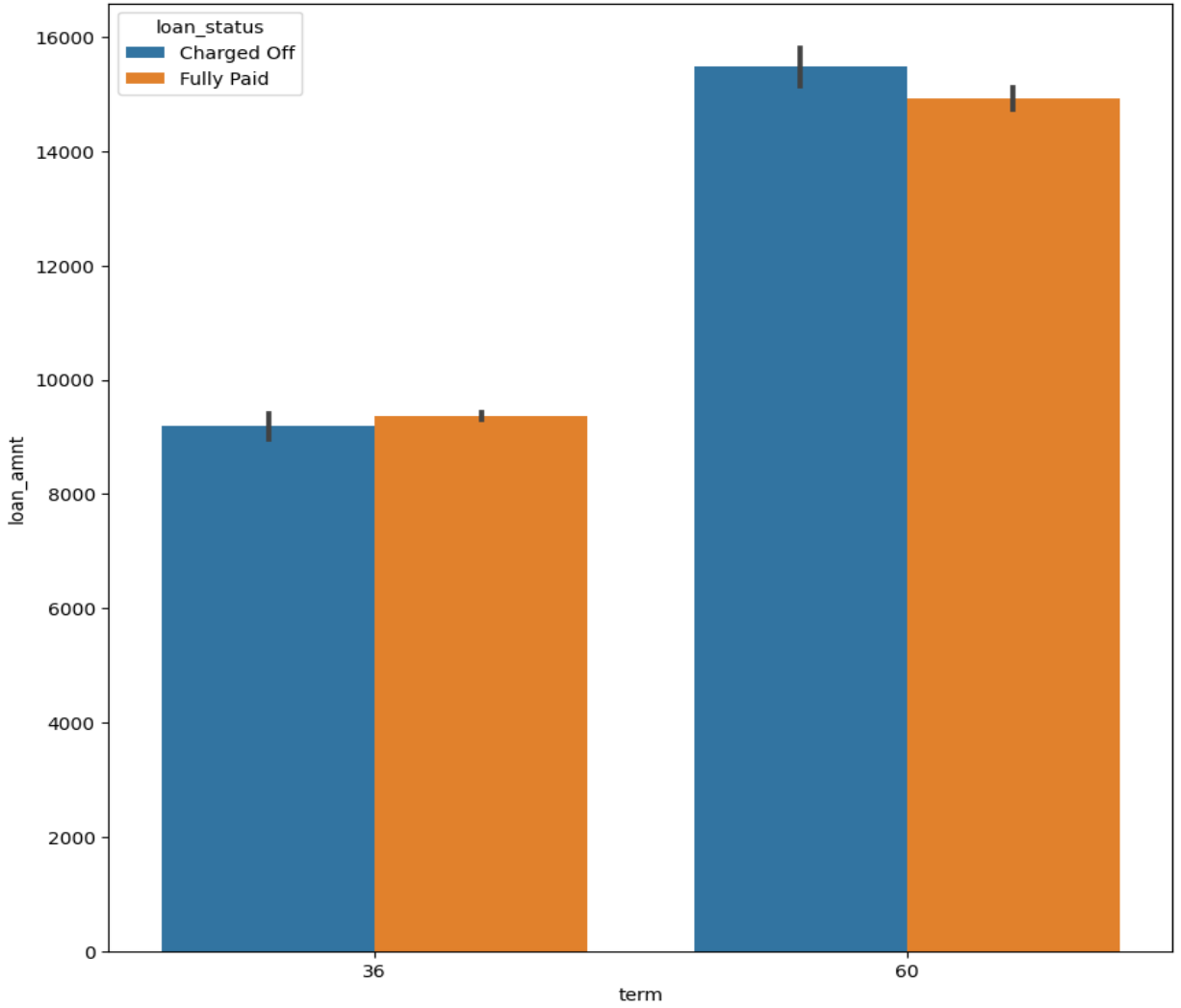
Loan Amount vs Installments Groups

As the installment is increasing the chances of loan defaulting is also increasing



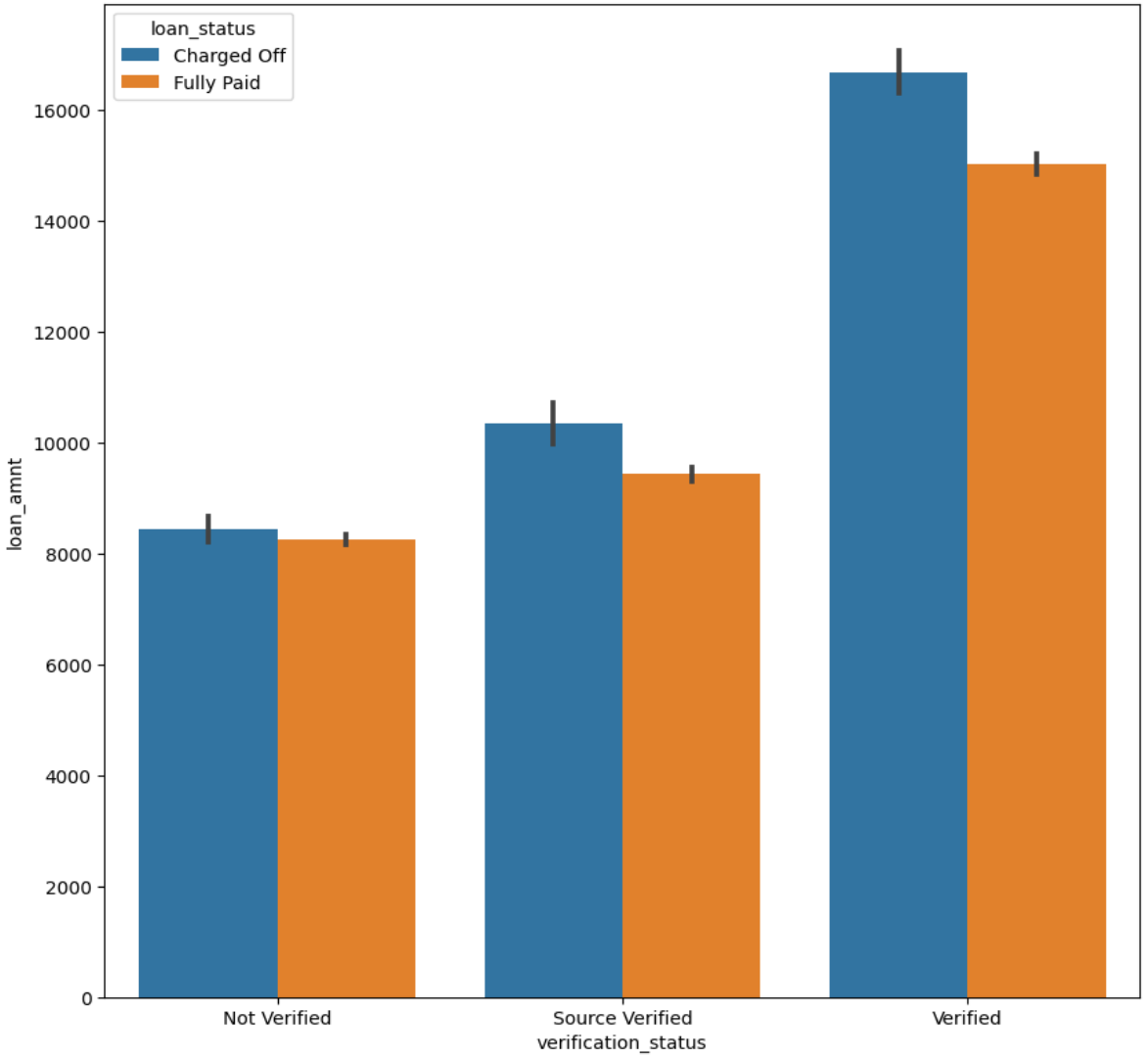
Loan Amount vs Term

Applicants with 60 months loan term are likely to default compared to 36 months term



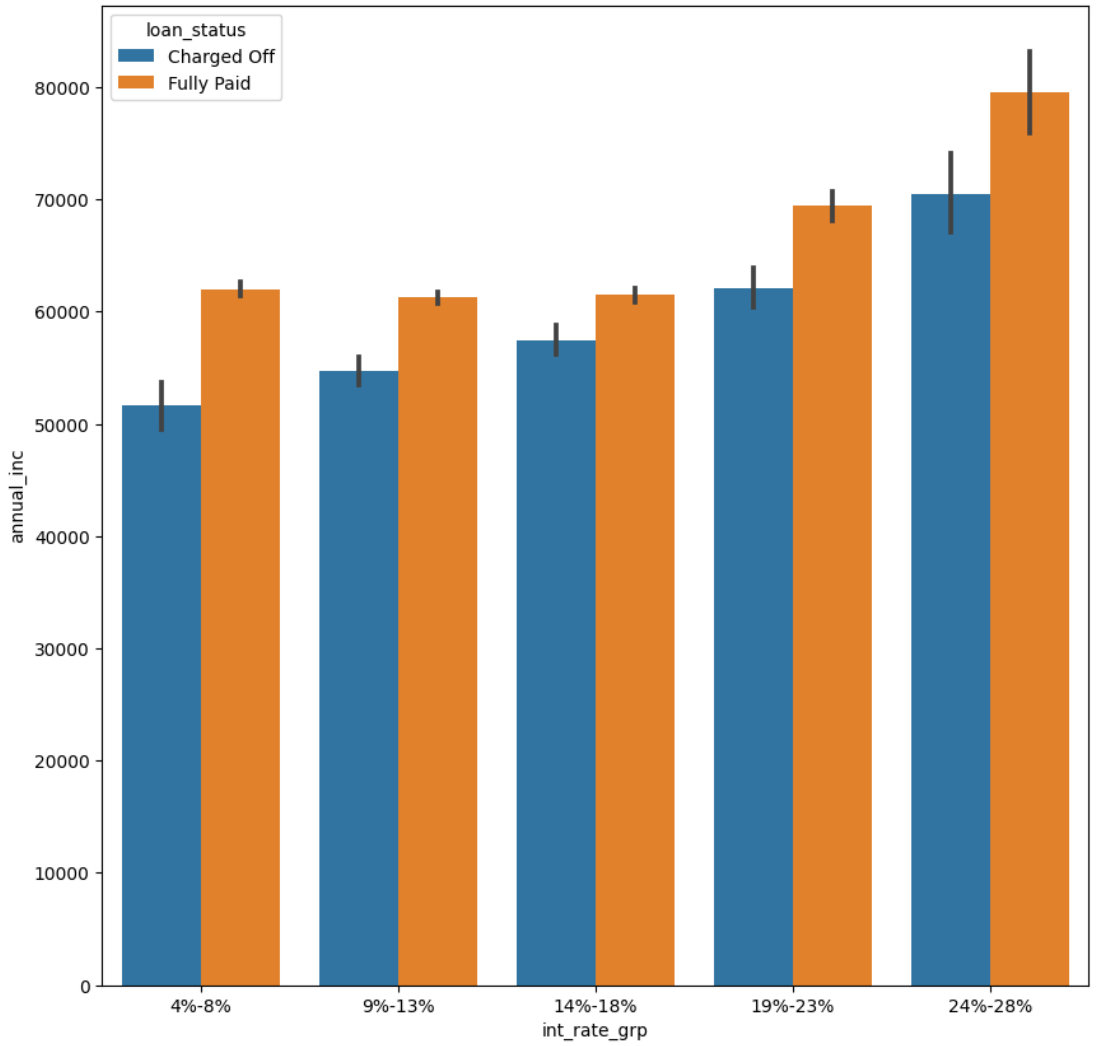
Loan Amount vs Verification Status

When applicants are verified and loan amount is greater than 15k then loan is likely to default



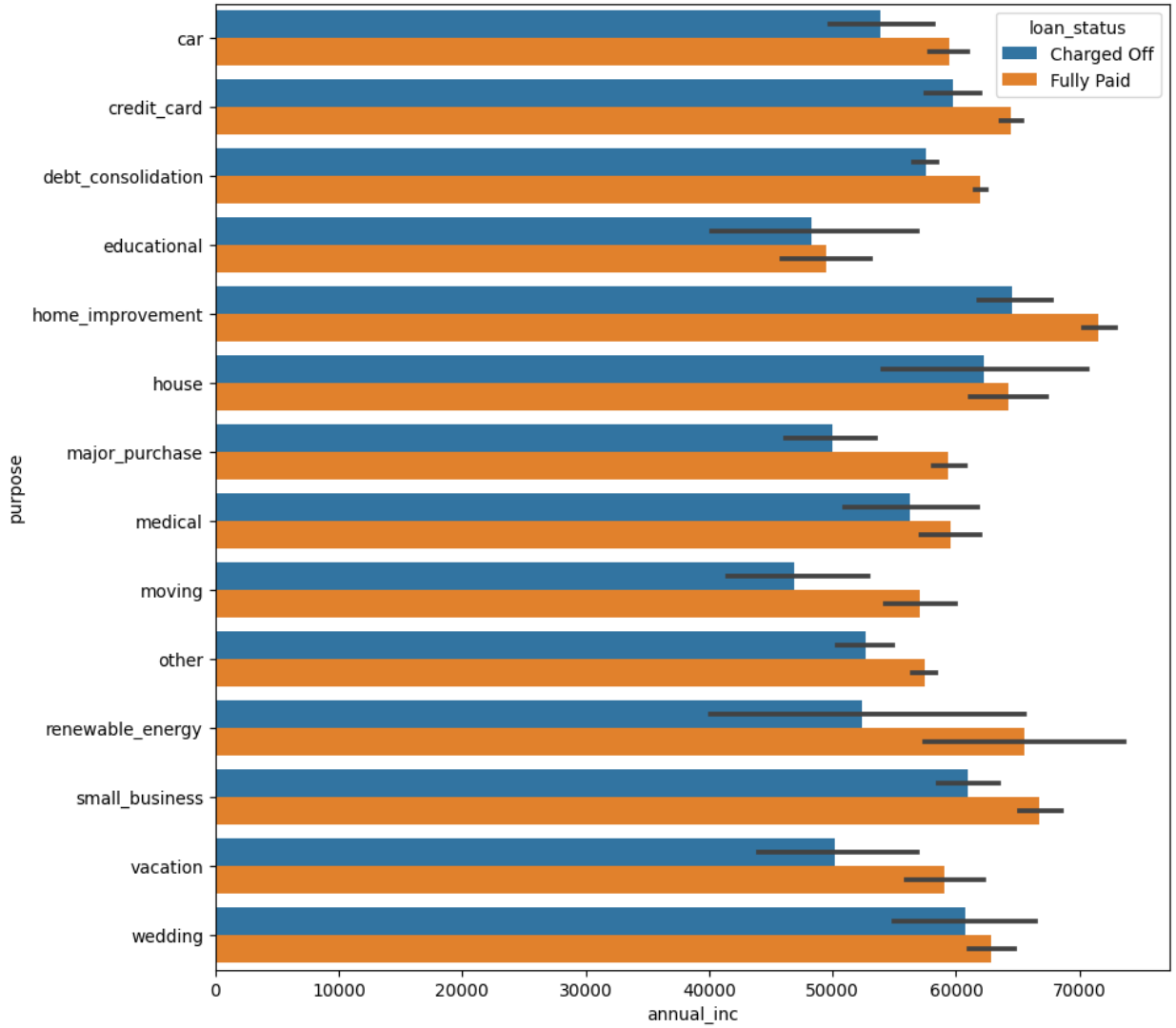
Annual Income vs Interest Rate

If the annual income is high across all interest rate groups, then there is less chances of loan getting default



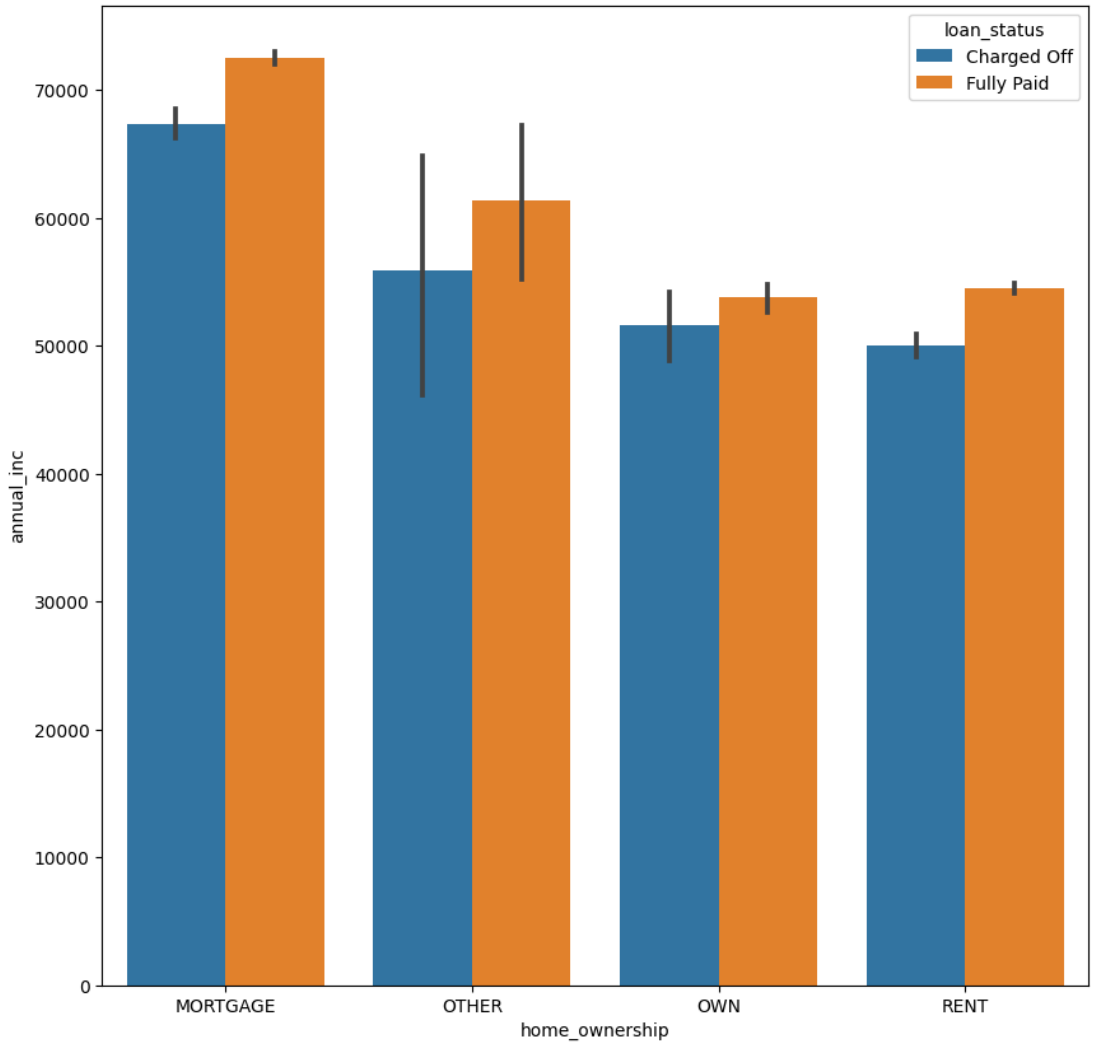
Annual Income vs Loan Purpose

Applicants with highest salary applied loans for home improvement and also they are once likely to default



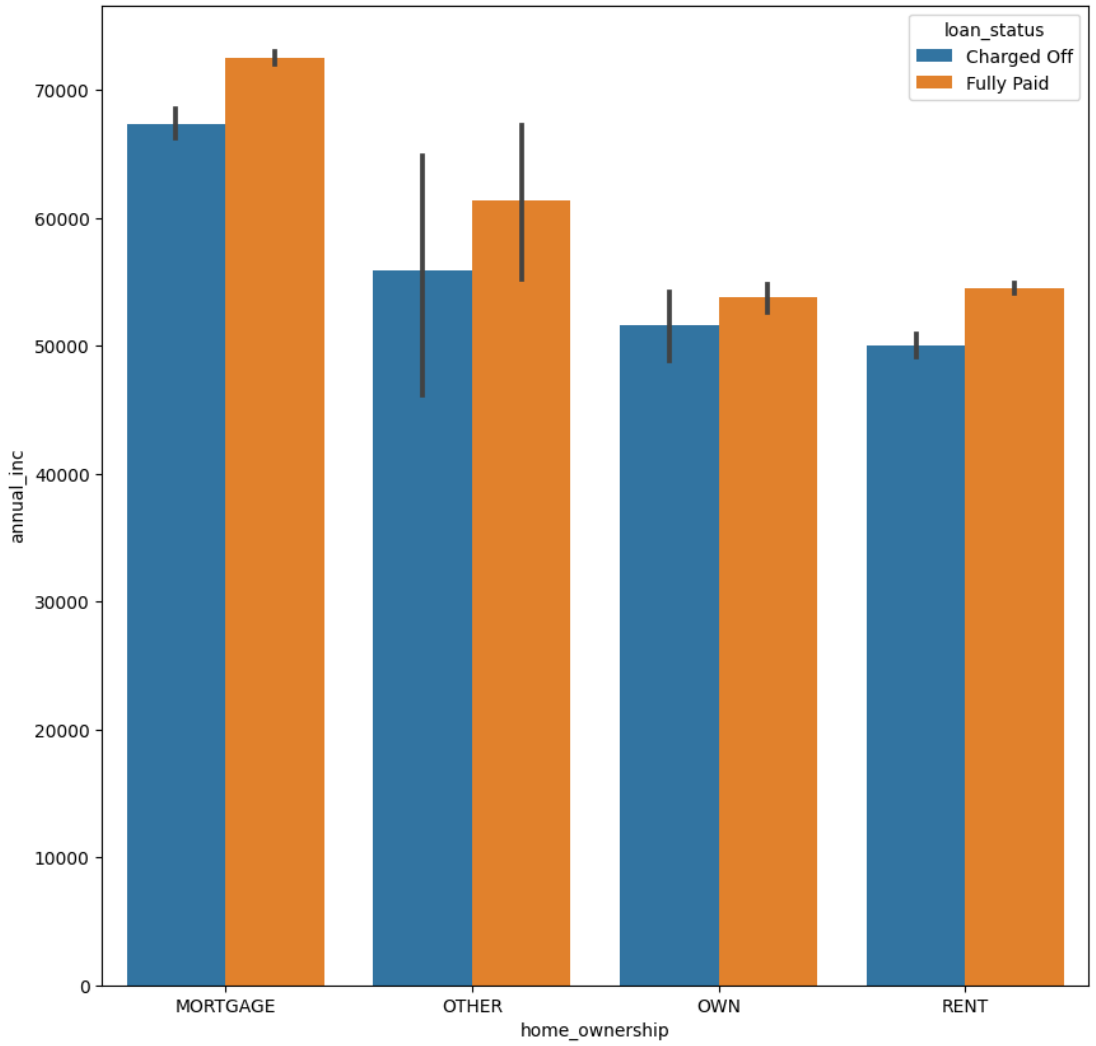
Annual Income vs Home Ownership

If the applicants have high annual income then they are not likely to default



Annual Income vs Home Ownership

If the applicants have high annual income then they are not likely to default



Observations from Bivariate Analysis

1.The charged off loans have higher interest rate across all loan amount groups. So if the interest rate is high there are more chances of loan getting default.

2.The Loan defaulting is increasing from A to G grade.

3.As the loan amount and employment length is increasing the risk of defaulting is also increasing.

4.Applicants whose loan amount is high are likely to default.

5.Applicants taken loan for small business and the loan amount is between 12k to 14.5k, then the applicants are more likey to default.

6.We can observe that loan amount is higher for applicants who defaulted and with annual income between 186k to 233k.

7.Applicants most likely to default in the months between oct to dec mostly december, applicants defaulted mostly in the year 2011

Observations From Bivariate Analysis

8.As the installment is increasing the chances of loan defaulting is also increasing.

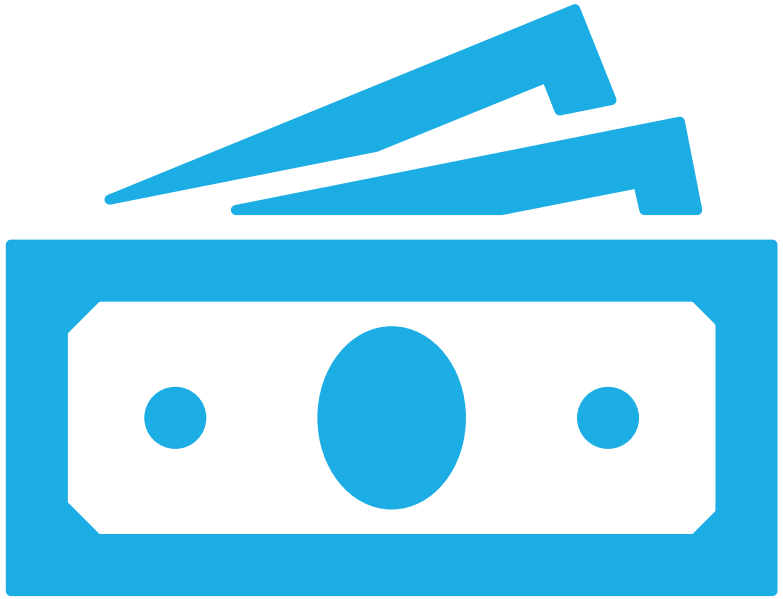
9.Applicants with 60 months loan term are likely to default compared to 36 months term.

10.When applicants are verified and loan amount is greater than 15k then loan is likely to default.

11.If the annual income is high across all interest rate groups then there less chances of loan getting default.

12.Applicants with highest salary applied loans for home improvement and also they are once likely to default.

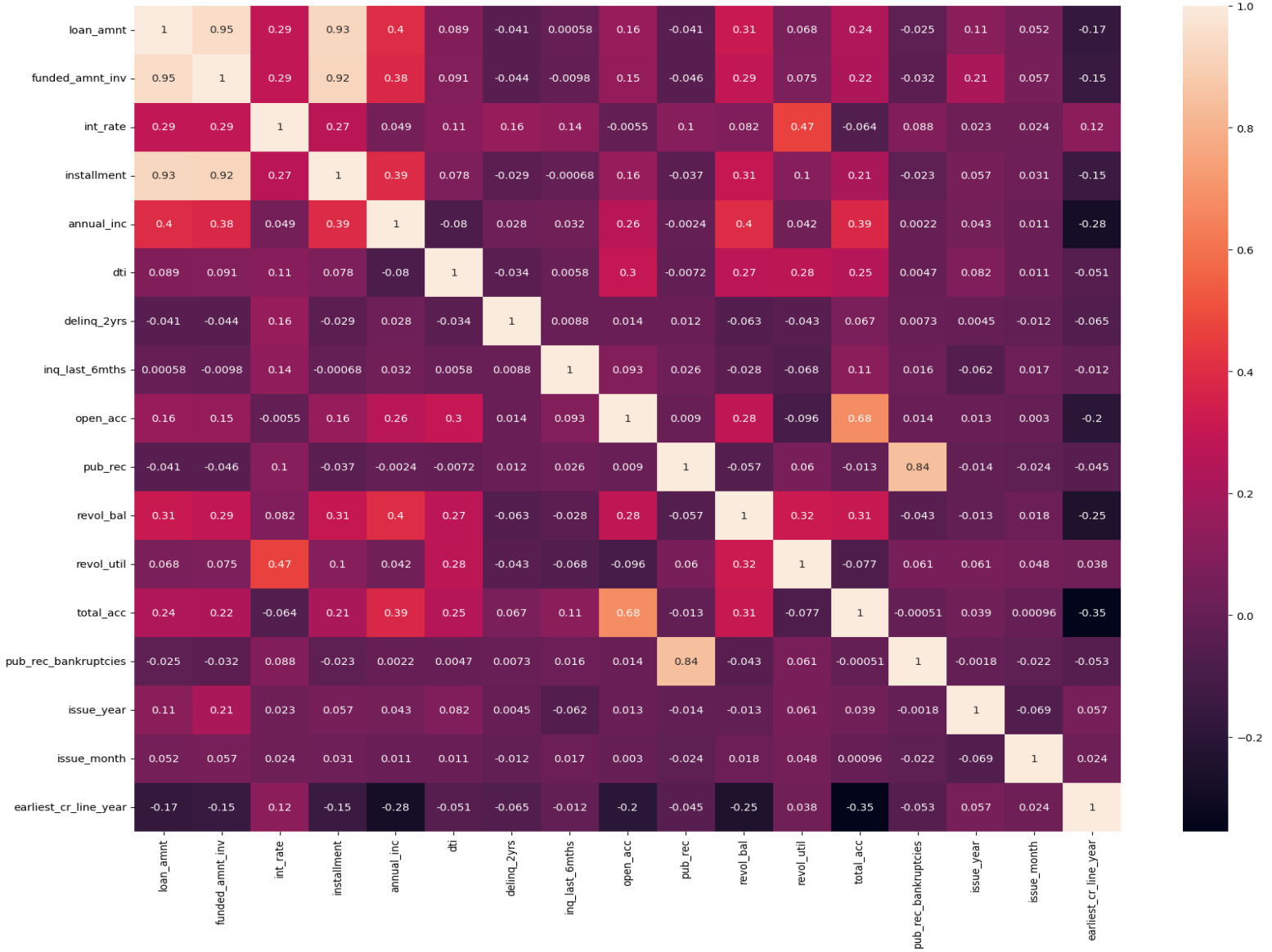
13.If the applicants have high annual income then they are not likely to default.



CHARGED OFF Loans Visualization

MULTIVARIATE ANALYSIS

Loan_amnt, funded_amnt_inv and installment are higly correlated.



Driving Factors

1. Interest rate
2. Loan Amount
3. Installment
4. Grade
5. Term
6. Annual Income
7. Purpose
8. Sub Grade
9. Home Ownership