Loan Defaulter Case Study

BY : MOHAMMAD KAFEE UDDIN &
SHRADDHA

Problem Statement

The company wants to reduce credit loss by identifying risky loan applicants. This involves understanding the driving factors behind loan default.

To address the business objectives outlined, we will conduct an exploratory data analysis (EDA) to understand the driving factors behind loan default. This will involve identifying strong indicators of default by analyzing various variables related to borrowers and loans.

We are going to follow following five steps of EDA in two sections:

1. Data sourcing
2. Data cleaning
3. Univariate analysis
4. Bivariate analysis
5. Derived metrics

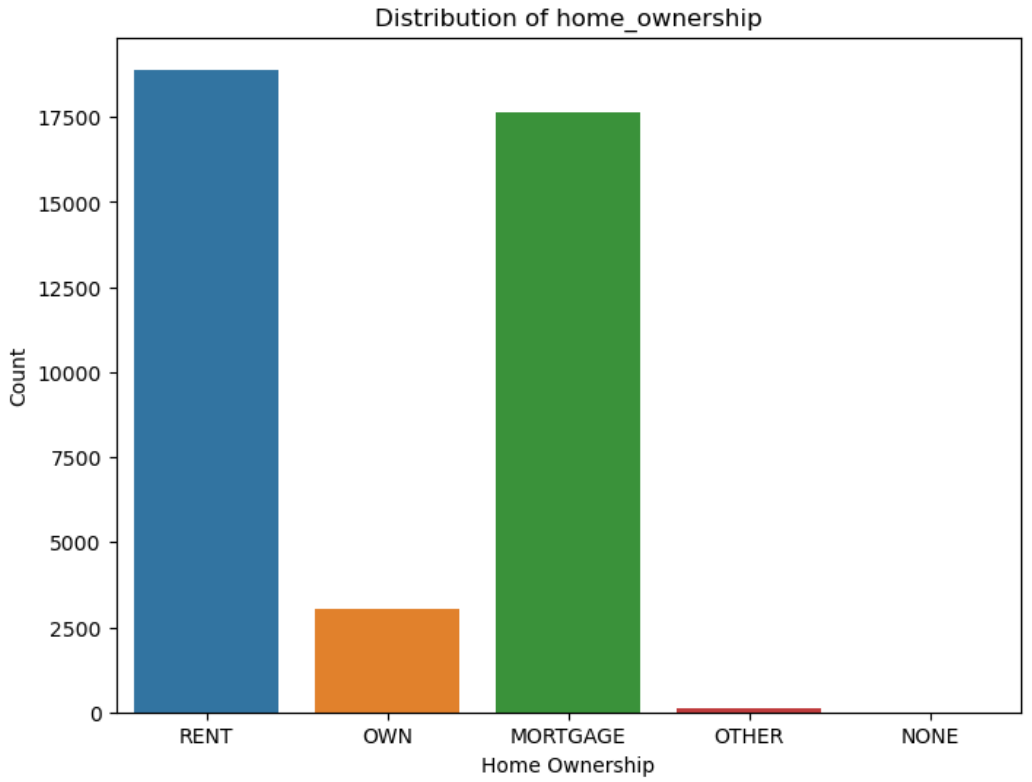
Section 1 - Data Sourcing and Cleaning:

Obtain data related to borrowers, loans, and default status. preprocess the data to handle missing values, outliers, and any inconsistencies.

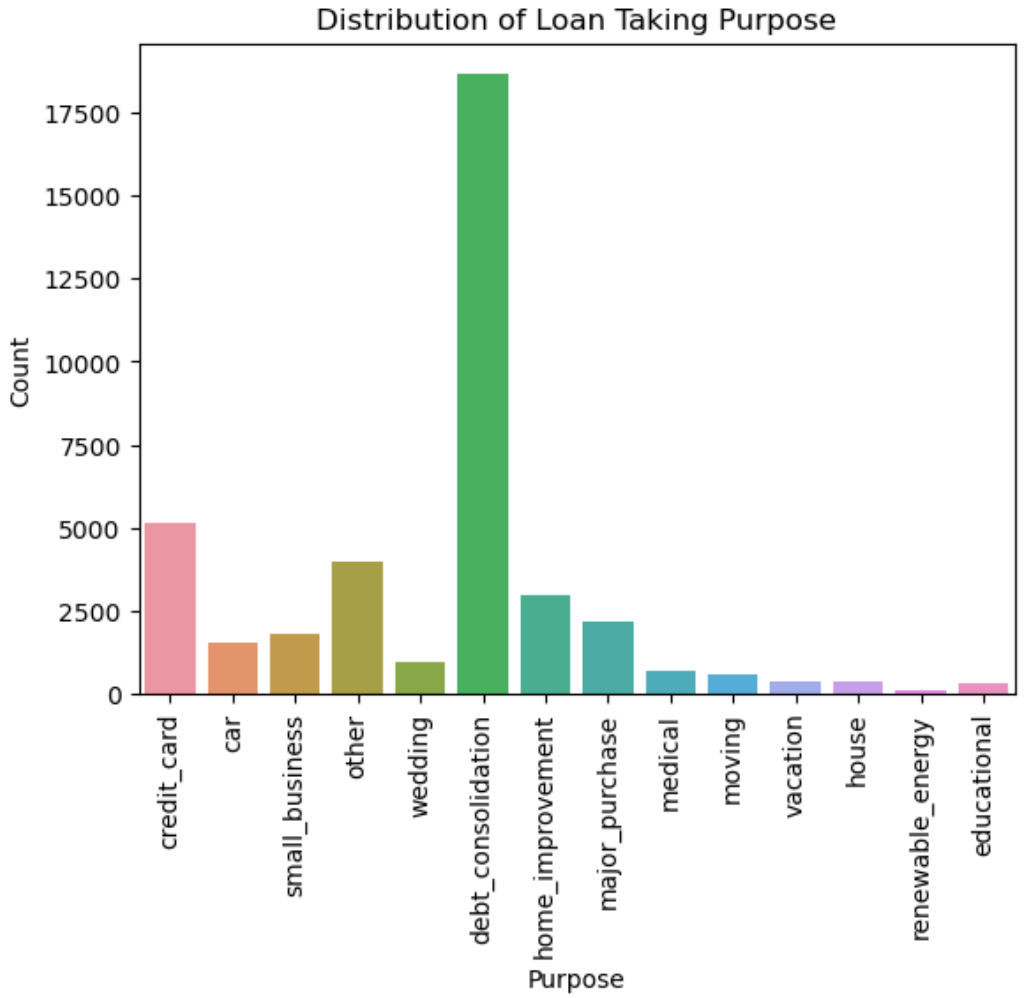
- **Missing value treatment**
- **Handling incorrect data types**

Section 2: Univariate, Bivariate, and Derived metrics analysis

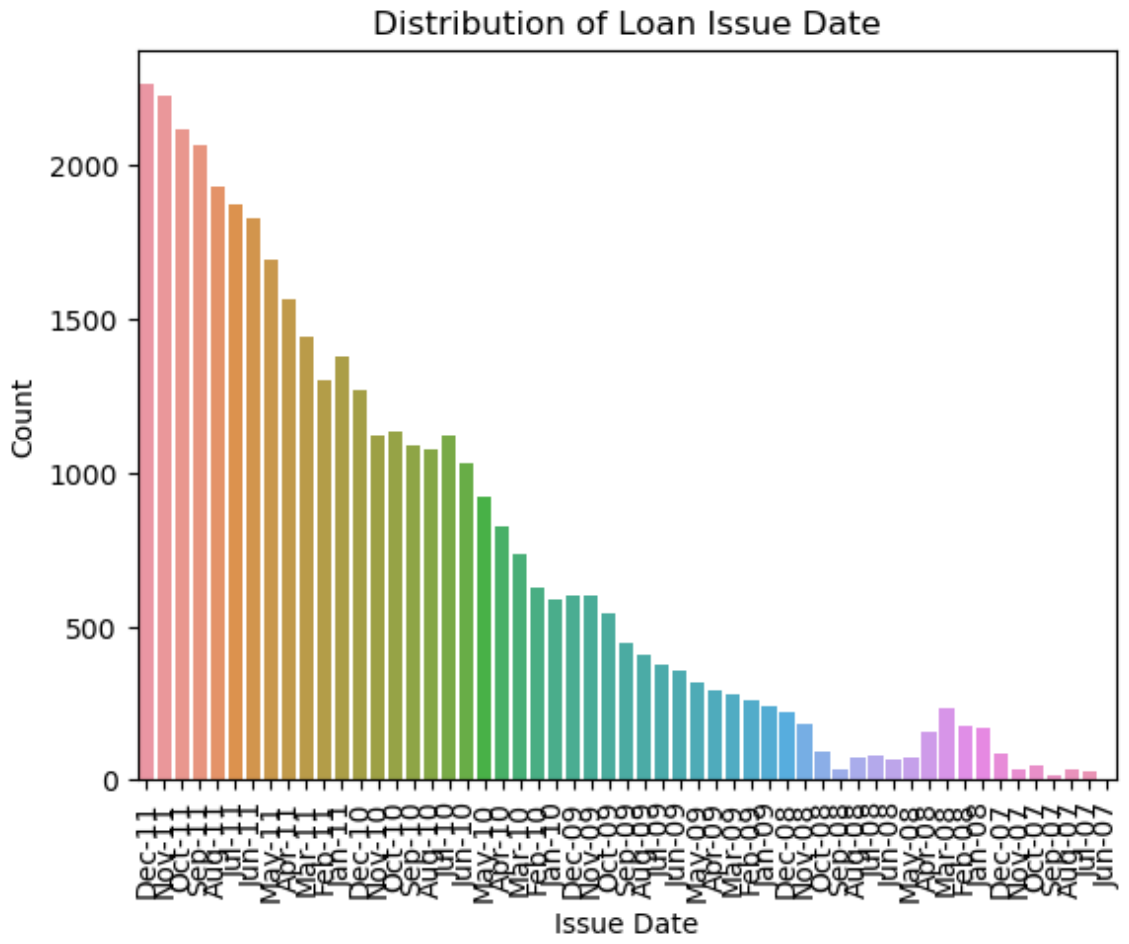
1a.)Univariate Analysis: *From the distribution of home ownership plot, we can understand "rent" and mortgage" categories are mostly provided the loan.*



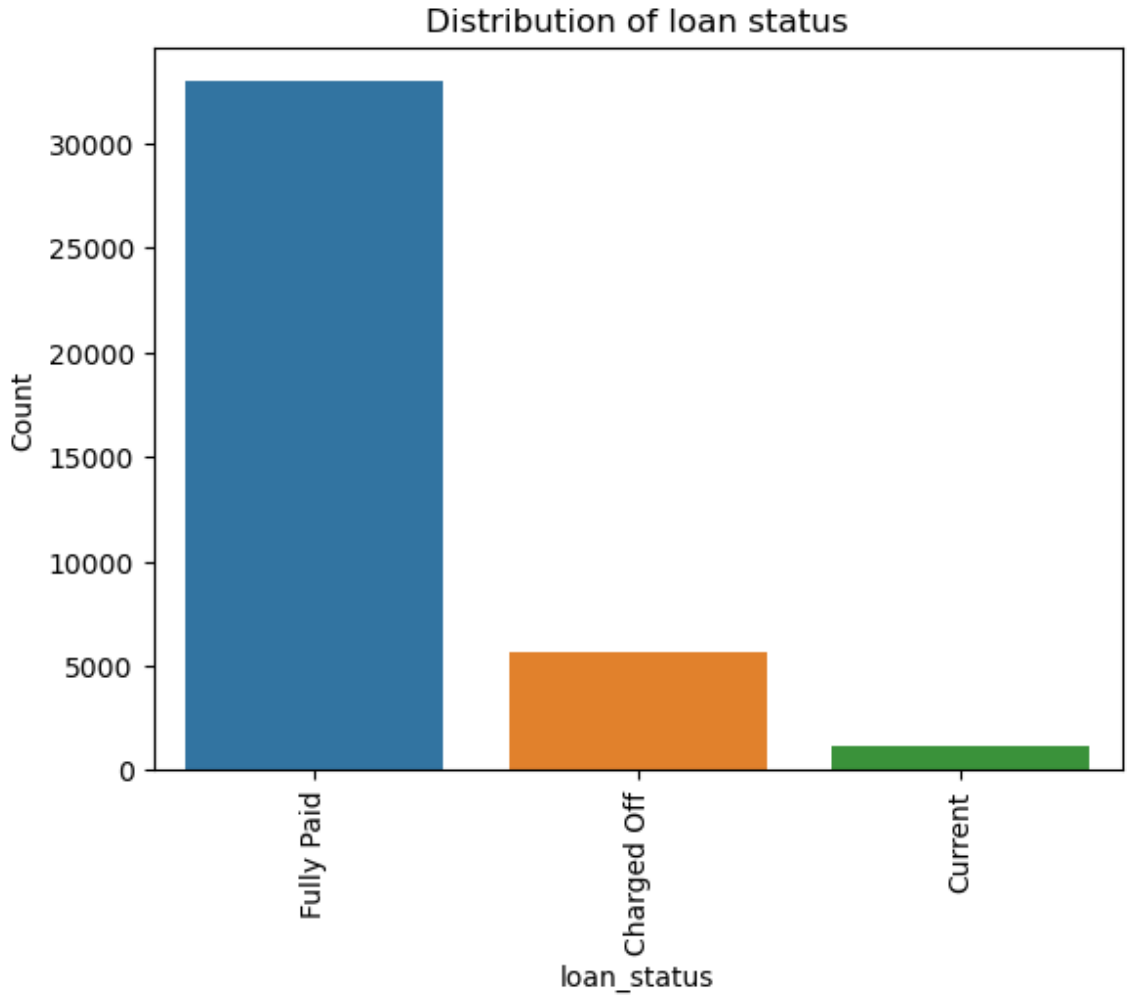
1b.)Univariate analysis: from the distribution of loan purpose plot, we can understand debt consolidation high purpose for loan applying.



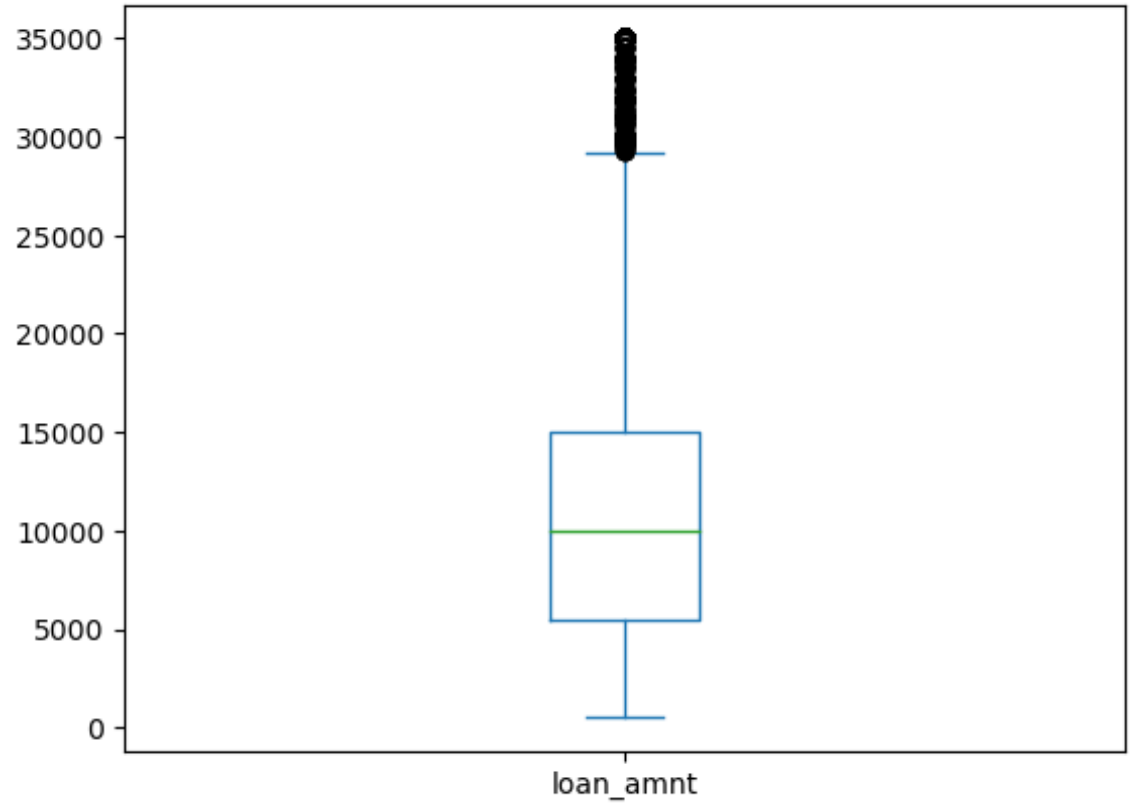
1c.)Univariate analysis: From the distribution of loan issue date plot, we can understand there is increase in applicant receiving by the months.



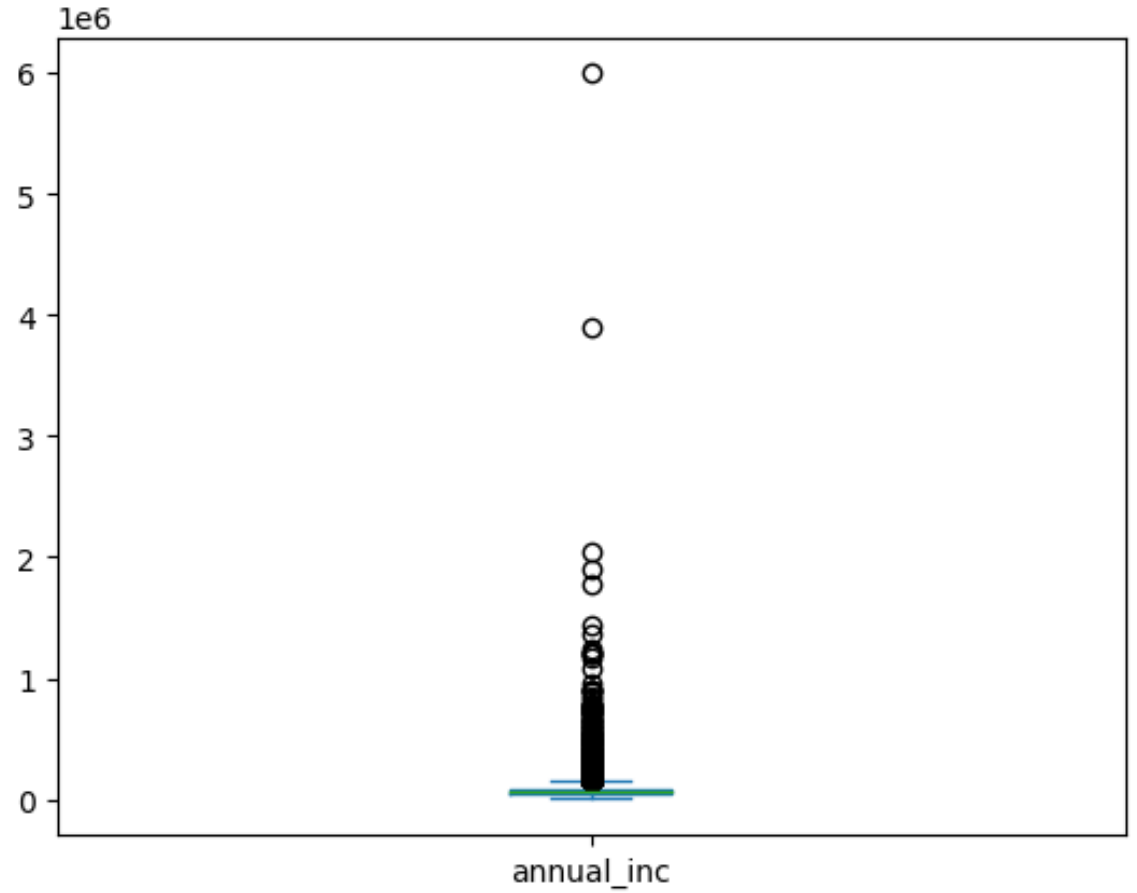
1d.)Univariate analysis: From the distribution of loan status plot, we can understand there are high number of applicate have already paid their loan



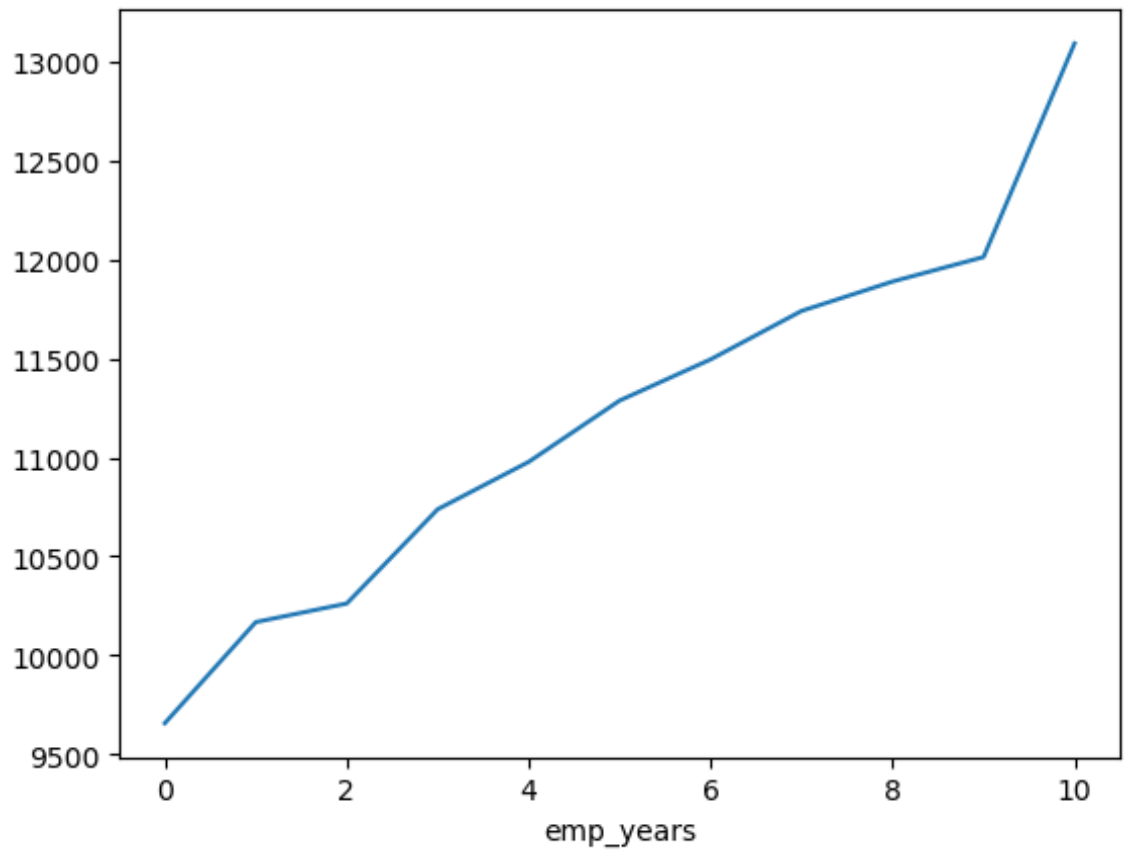
1e.)Univariate analysis: from the spread of loan amount, we can see median is around 10000



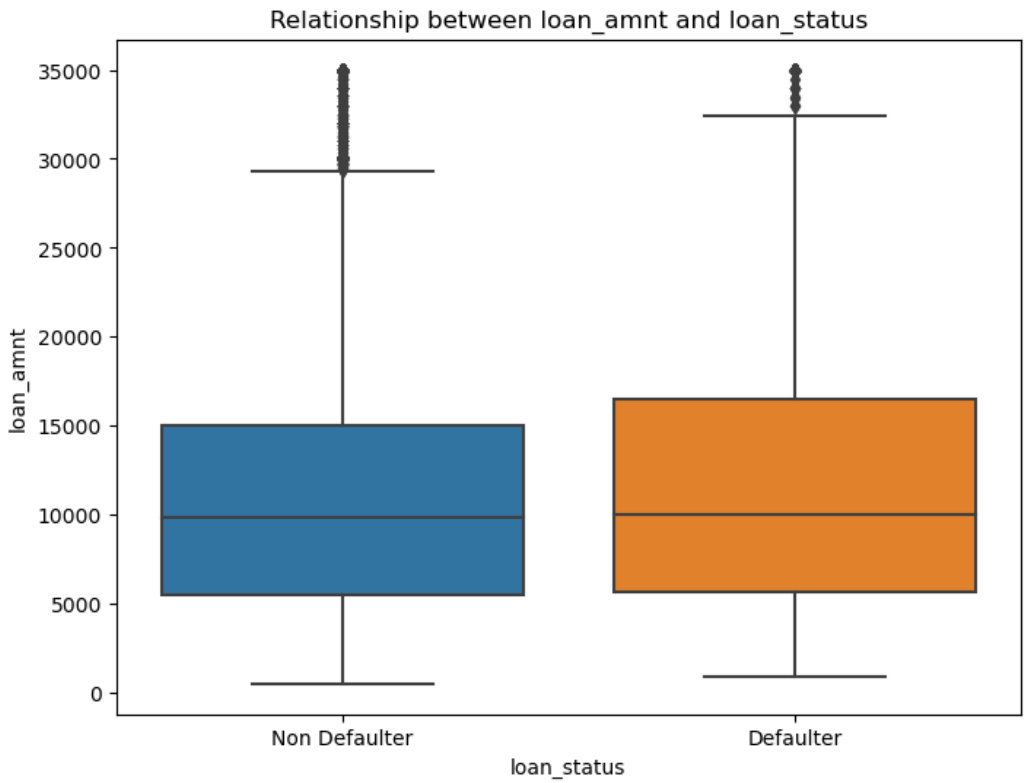
1f.)Univariate analysis: from the spread of annual income, we can see there are two outliers



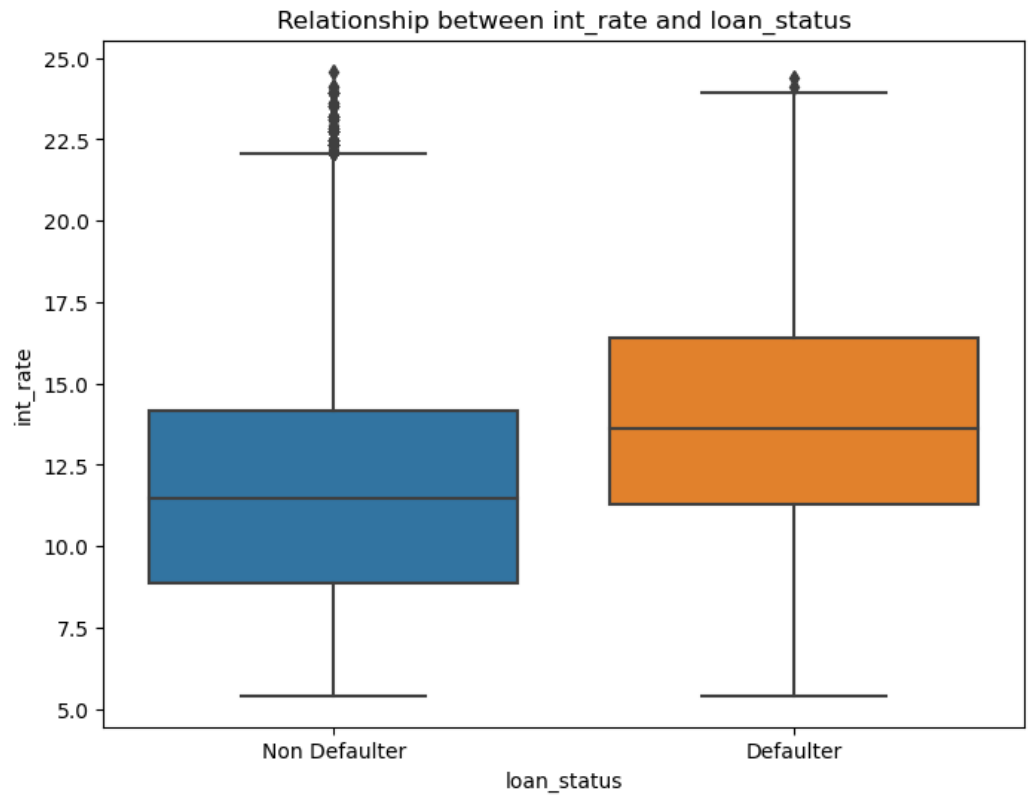
2.Segmented univariates:
From the applicant years of employment, it is clearly visible avg income has increased.



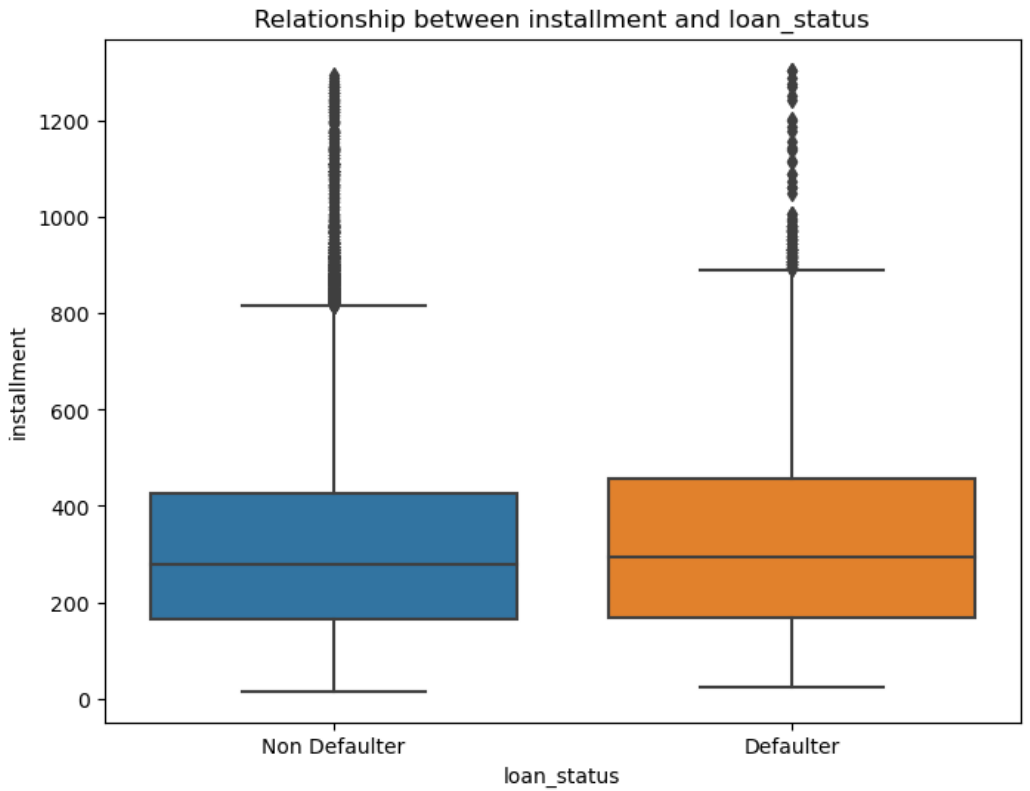
3a.)Bivariate analysis: High loan amount are more defaulter



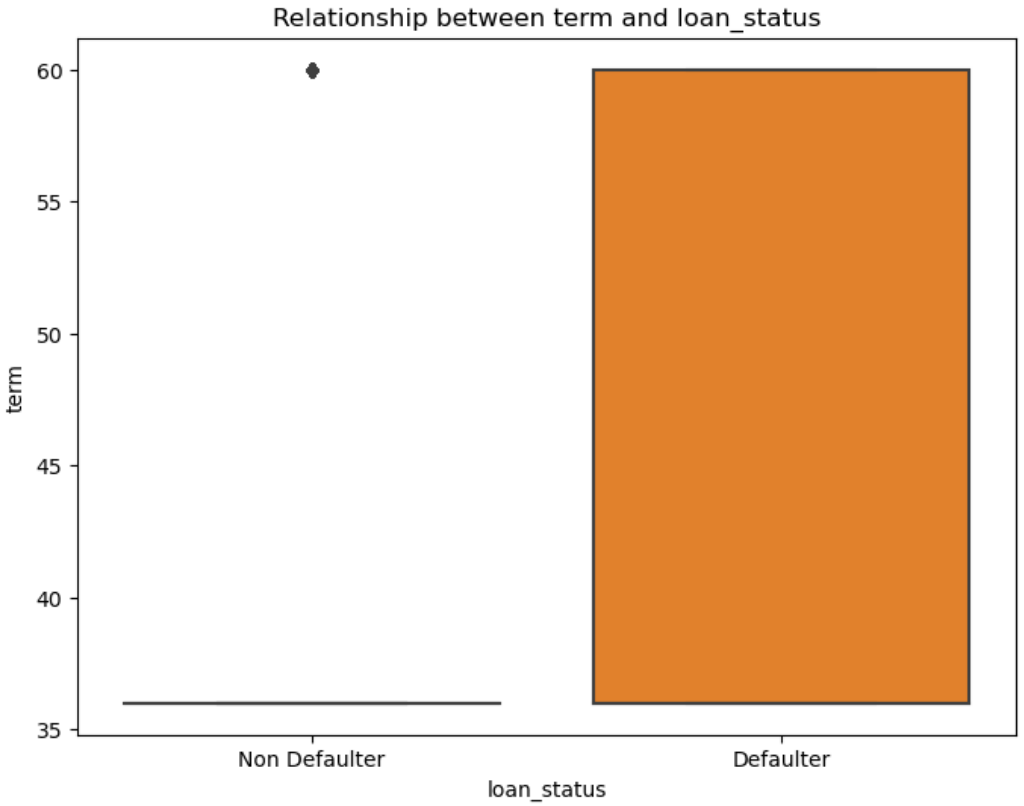
3b.)Bivariate analysis: High interest rate are more defaulter



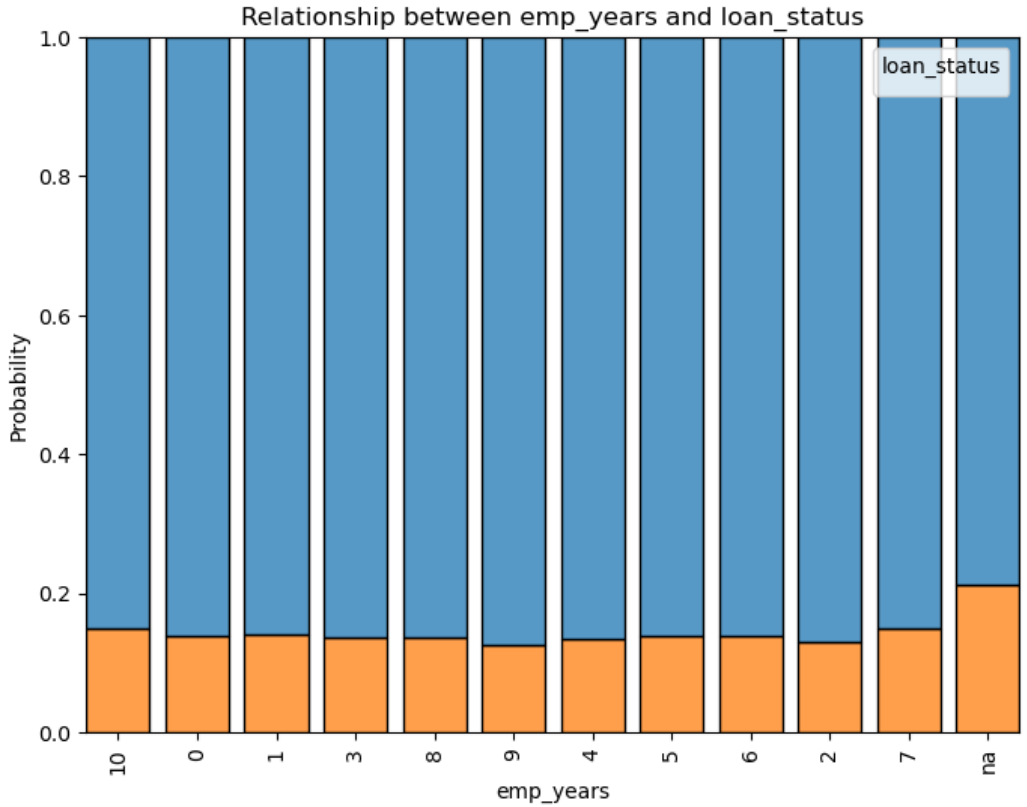
3c.)Bivariate analysis: High installment are more defaulter



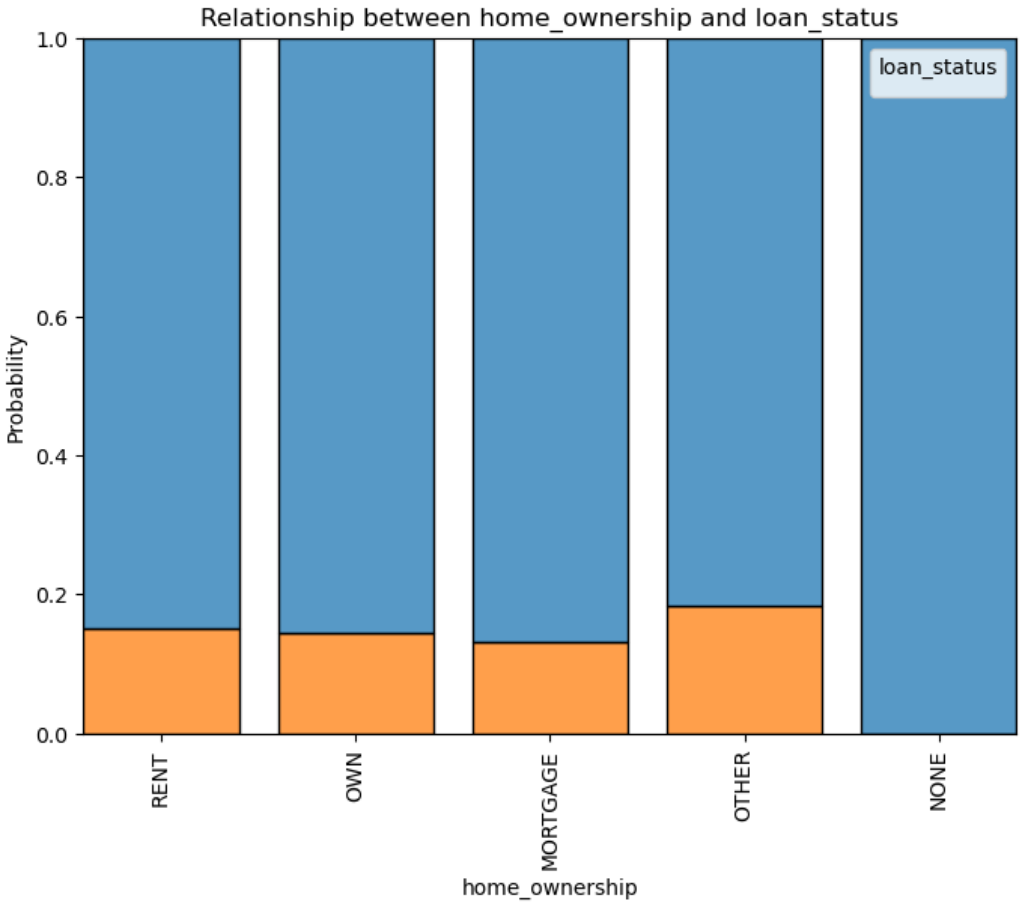
3d.)Bivariate analysis: Higher term are more defaulter



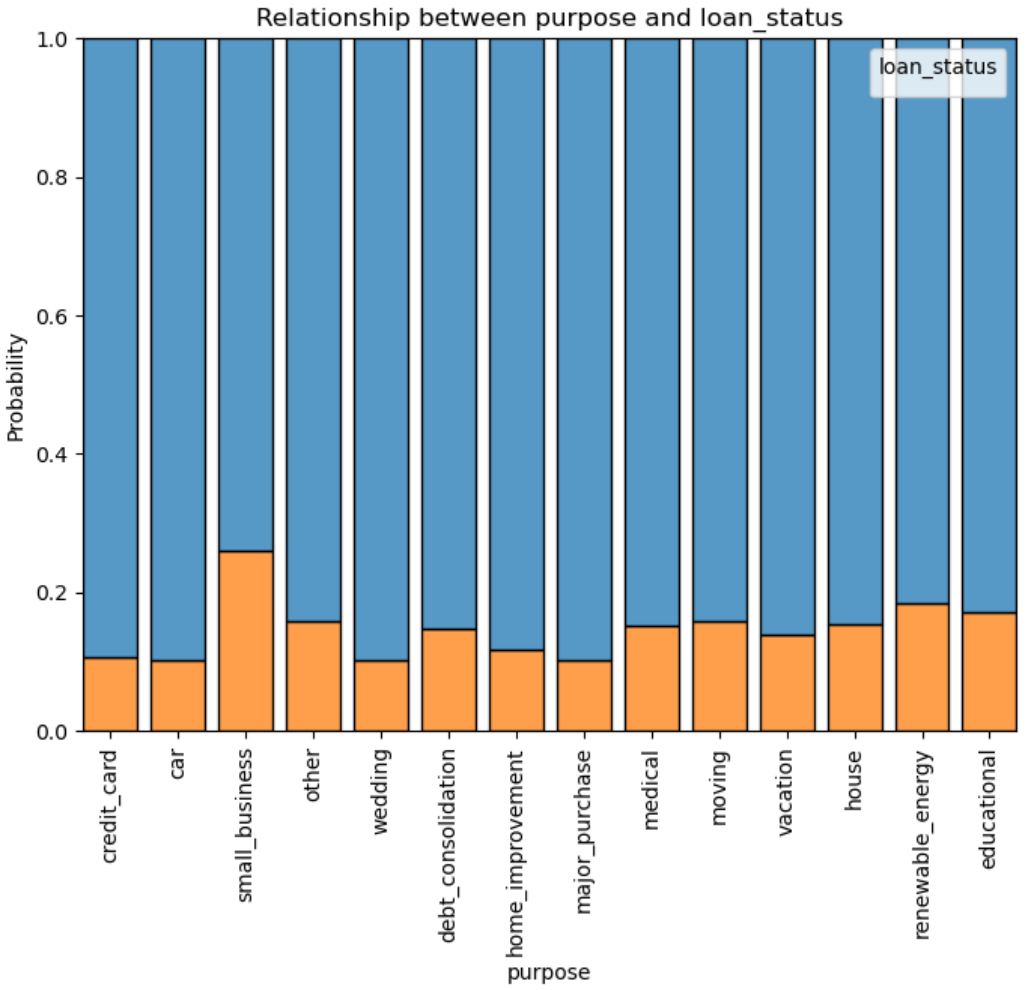
3e.)Bivariate analysis: 9 years employment are less defaulter.



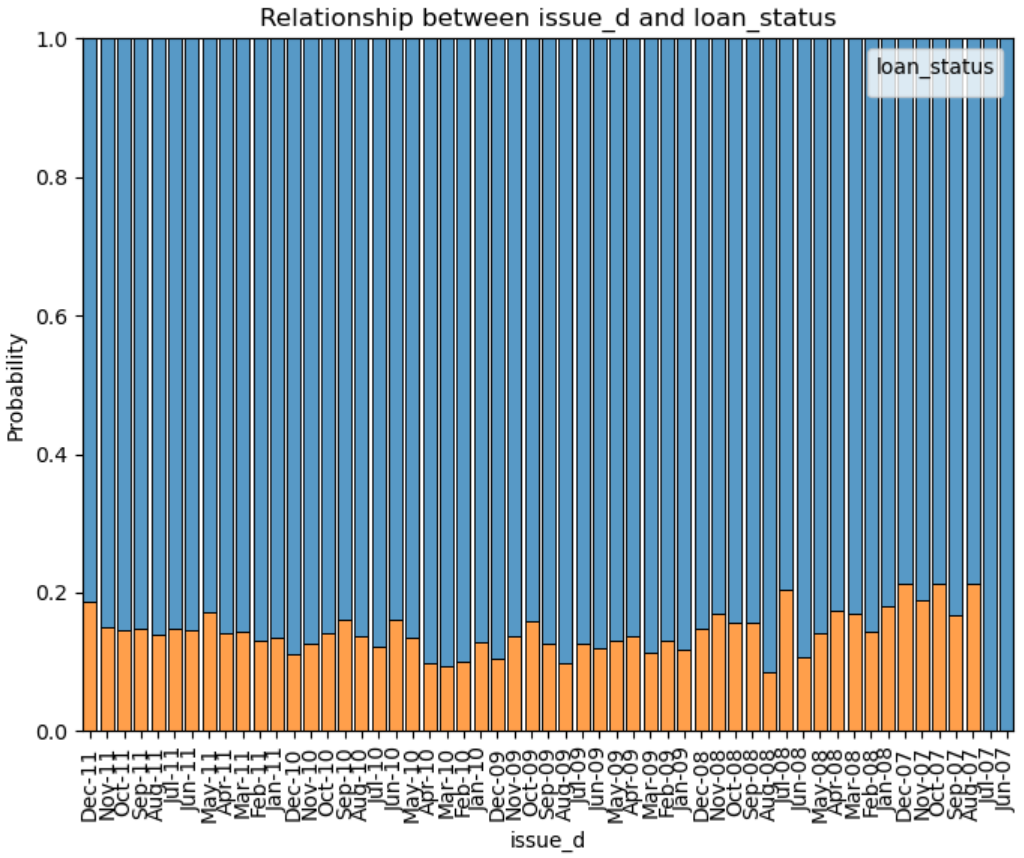
3f.)Bivariate analysis:
Mortgage ownership are less defaulter.



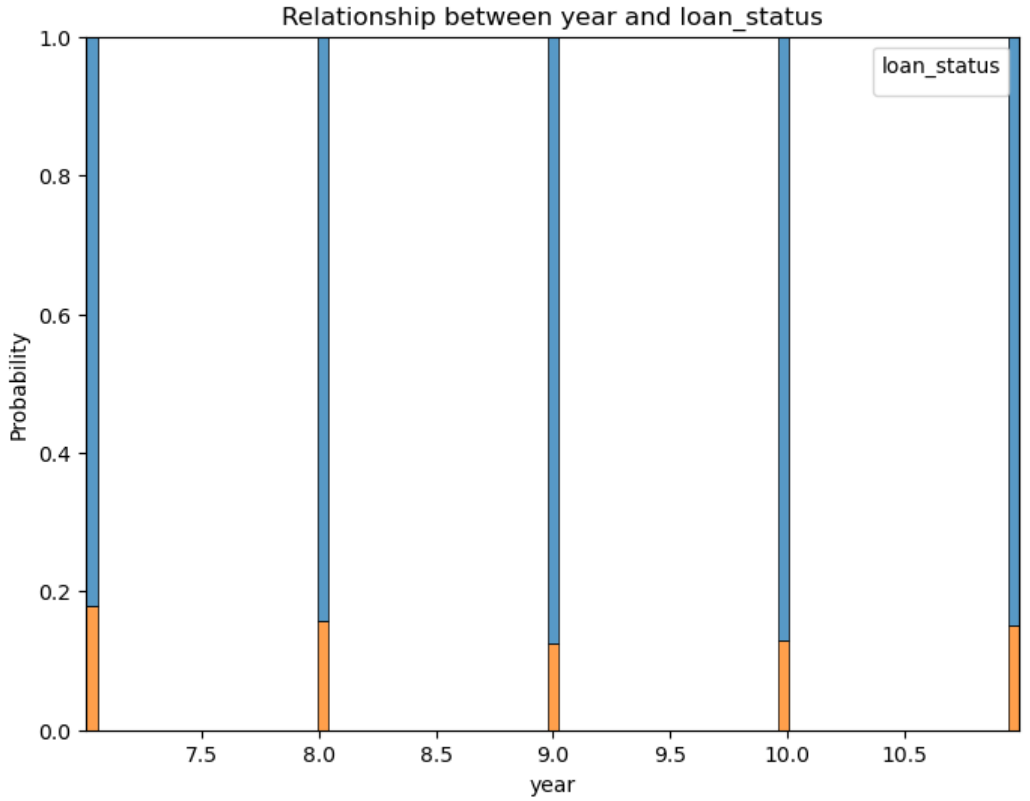
3g.)Bivariate analysis: Small business are more defaulter



3h.)Bivariate analysis: Defaulter trend has changed from higher to lower to higher again with passing months.



4.Derived metrics analysis: In the year 2007 probability of defaulter is higher as compared to other years.



Thank You