



LendingClub Case Study

Group Members:
Gogineni Sreeram Prudhvi & Anusha Saxena
5th October 2022

Introduction

LendingClub is financial services company and a digital marketplace that offers the largest peer to peer lending platform for personal loans, business loans, auto refinancing and financial services for patients.

Background

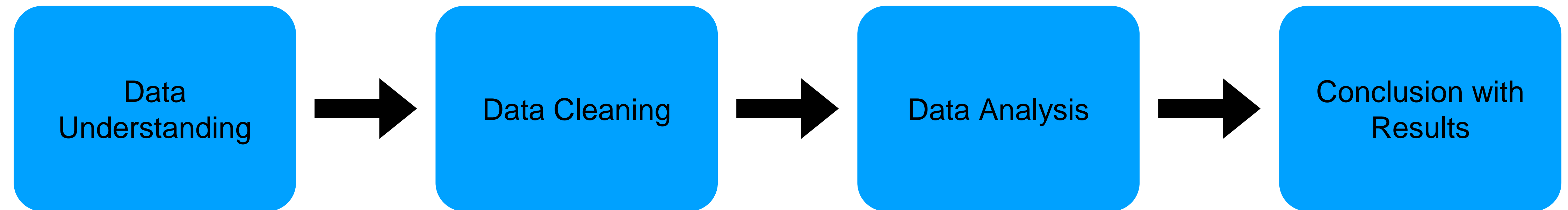
LendingClub has to make decision for loan approval based on the information of past loan applicants, that indicates if a person is likely to default or not. Which would be used to determine the grant or denial of loan or lending at higher interest rates as per the risk associated.

Objective

Perform EDA to analyze the dataset from the past applicant's information. Identify the variables that indicate defaulters in best ways and could be used provide insights for making decisions on approval or rejection or interest rates or amount of loan to be granted.



Analysis Steps



- Studying and validating the data
- Identifying the target variable
- Identify the important columns in dataset that may contribute in analysis

- Cleaning Missing values
- Removing Redundant columns
- Removing rows with all zero
- Convert the values to proper datatype such as float, int etc.

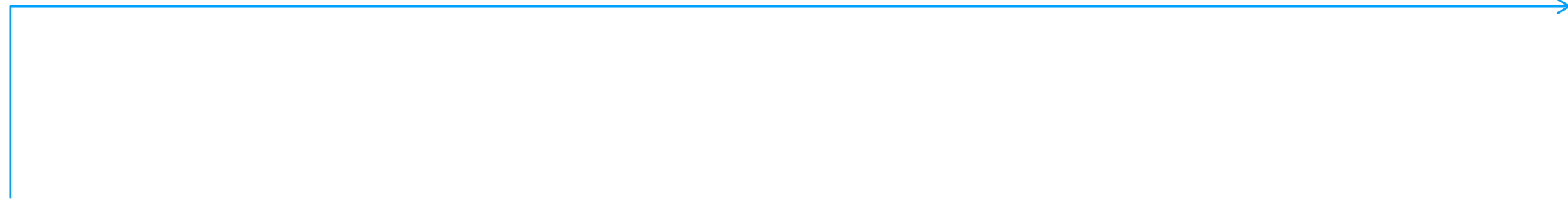
- Univariate Analysis
- Segmented Univariate Analysis

- Insights and Observation

Data Understanding


Types of variables

- Customer information



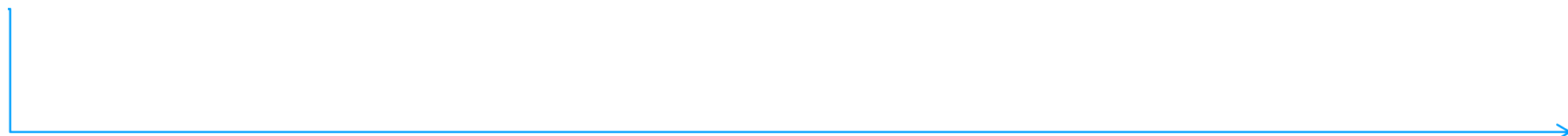
Employment Length
Employment title
Annual Income
Zip Code
Description

- Loan related information



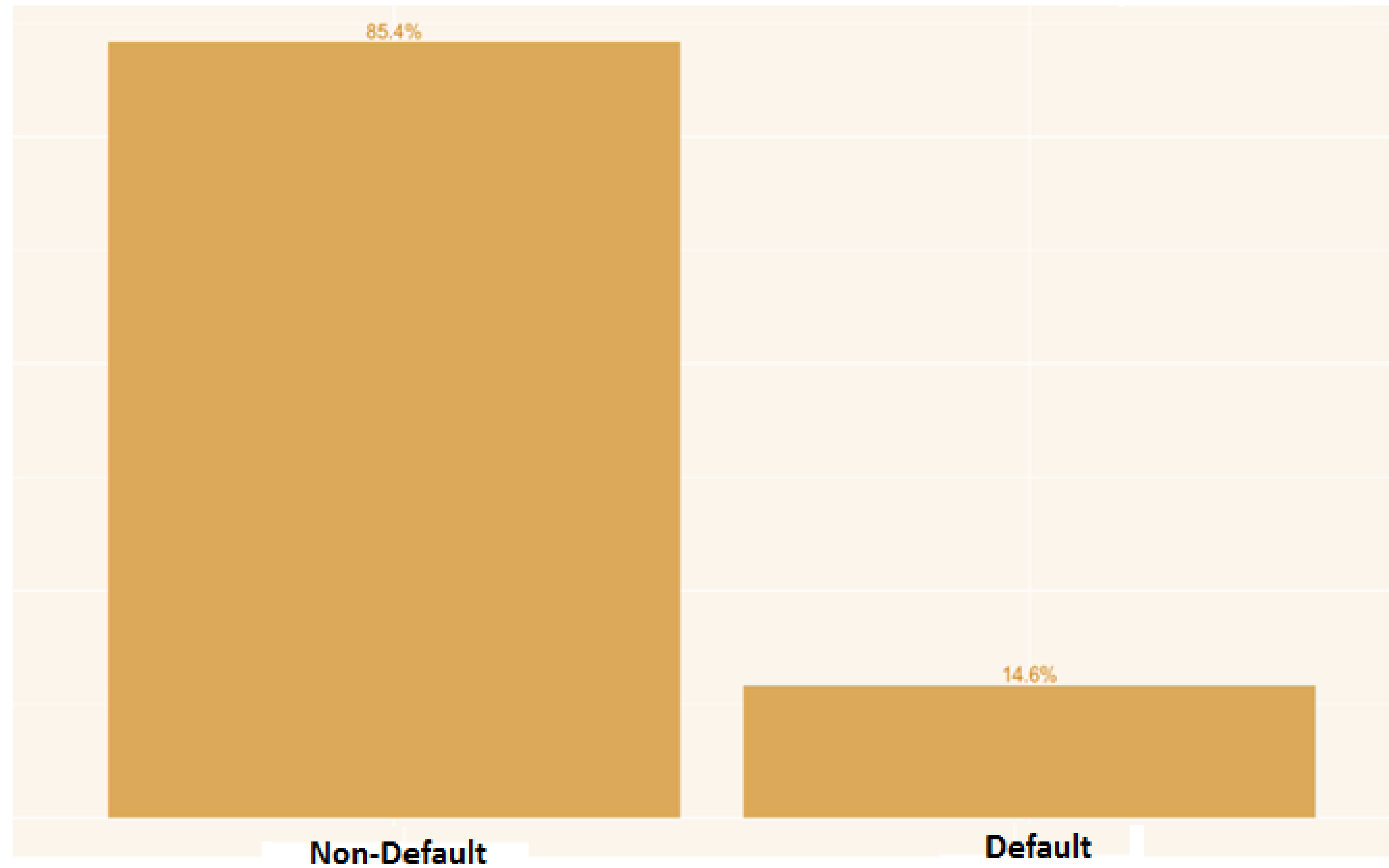
Loan Amount
Funded Amount
Funded Amount Investment
Interest Rate
Loan Status
Laon Grade

- Customer behaviour (If the loan is granted)

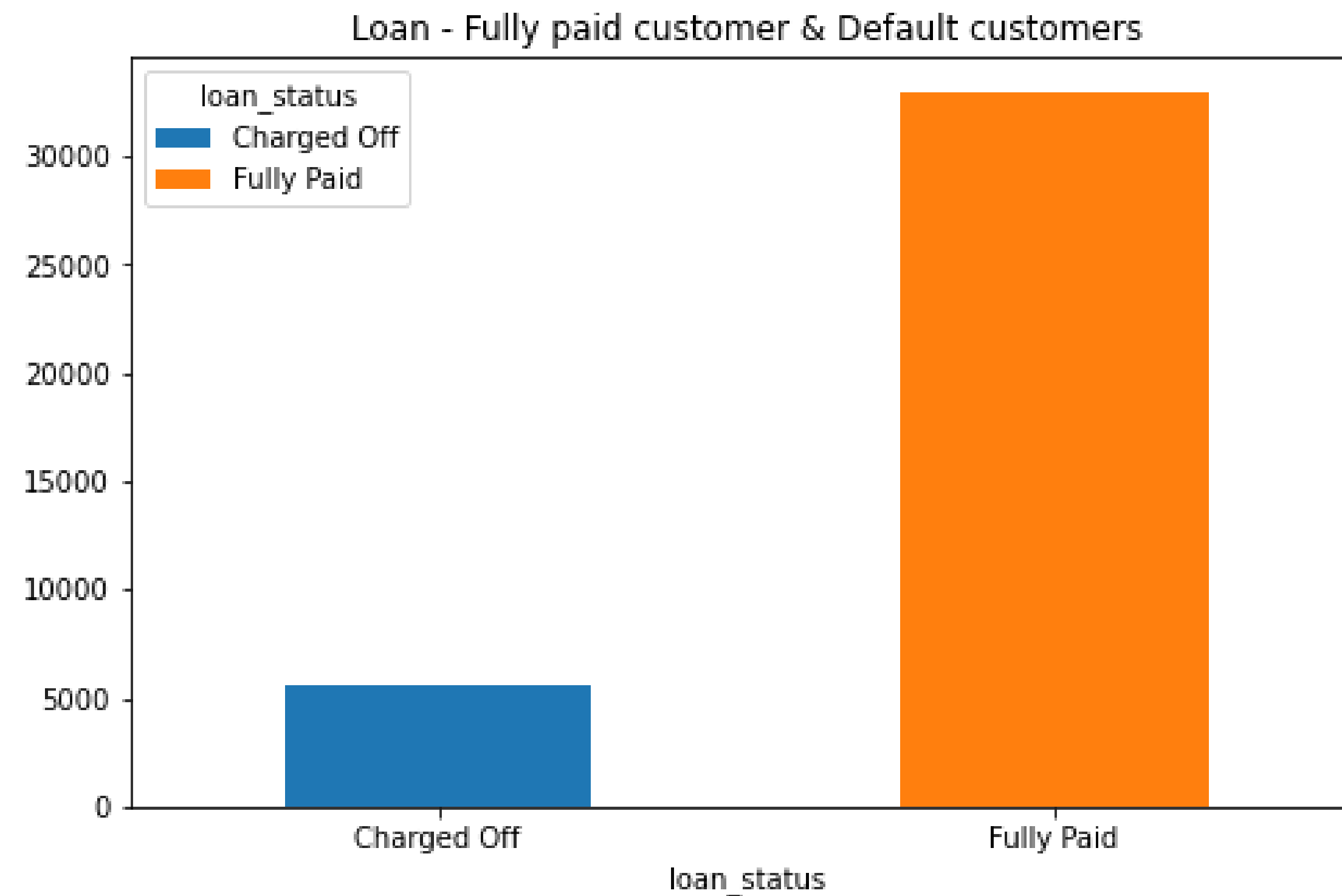


Delinquency year -2
earliest credit line
Revolving balance
Recoveries
Application type
Loan purpose

Data Understanding - Overall Default Rate is 14%



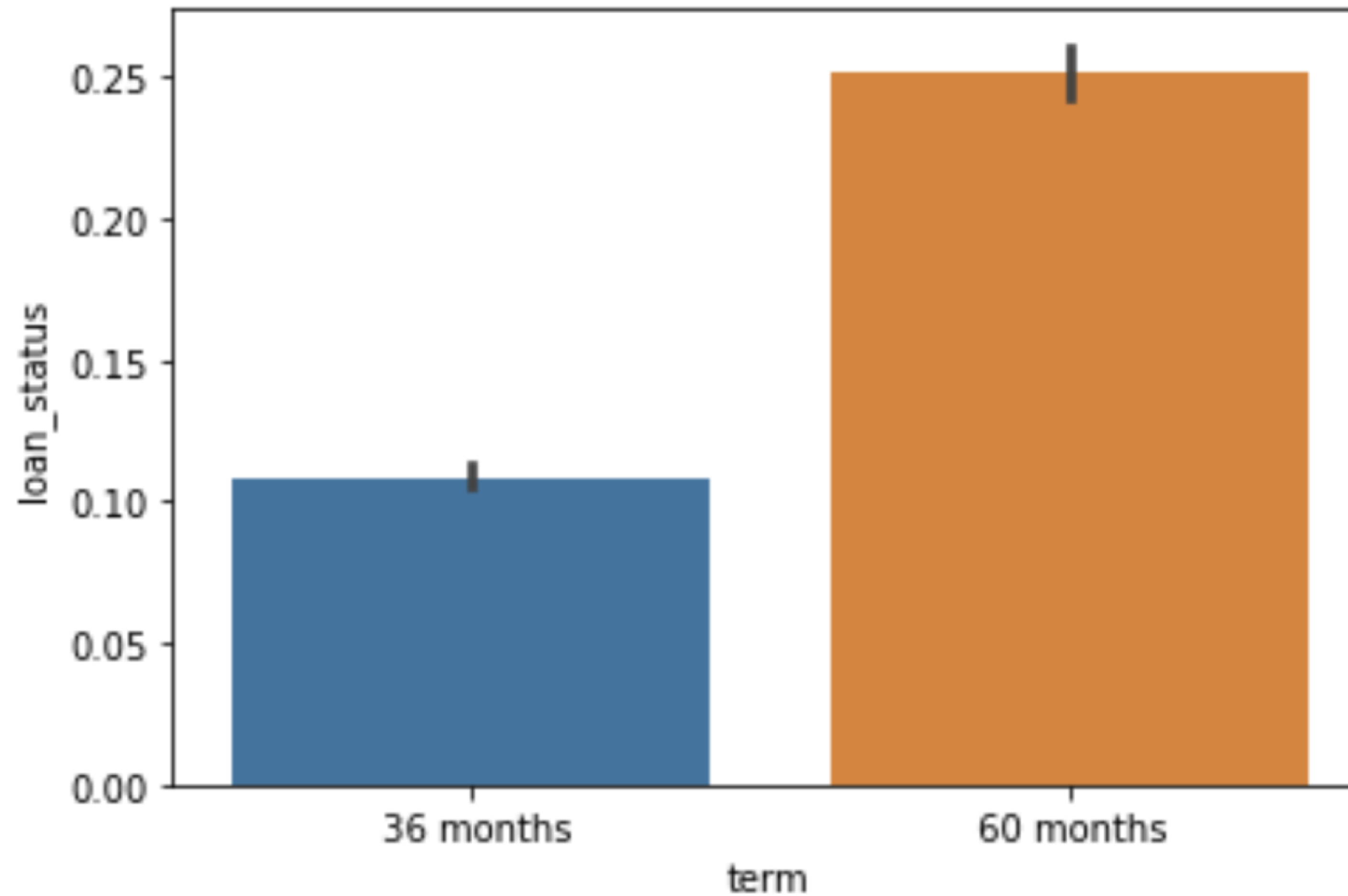
Data Analysis - Loan fully paid & Default customers



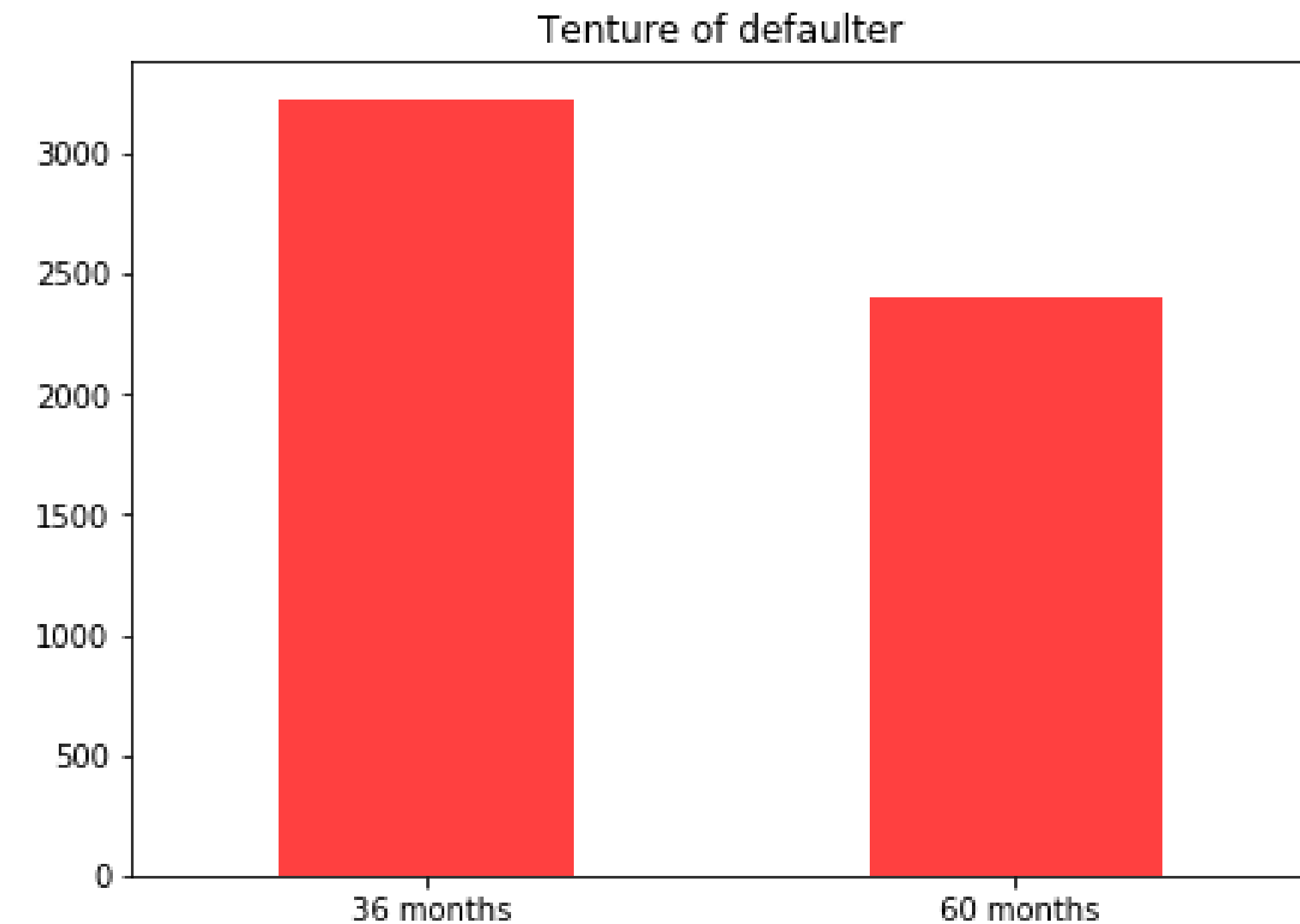
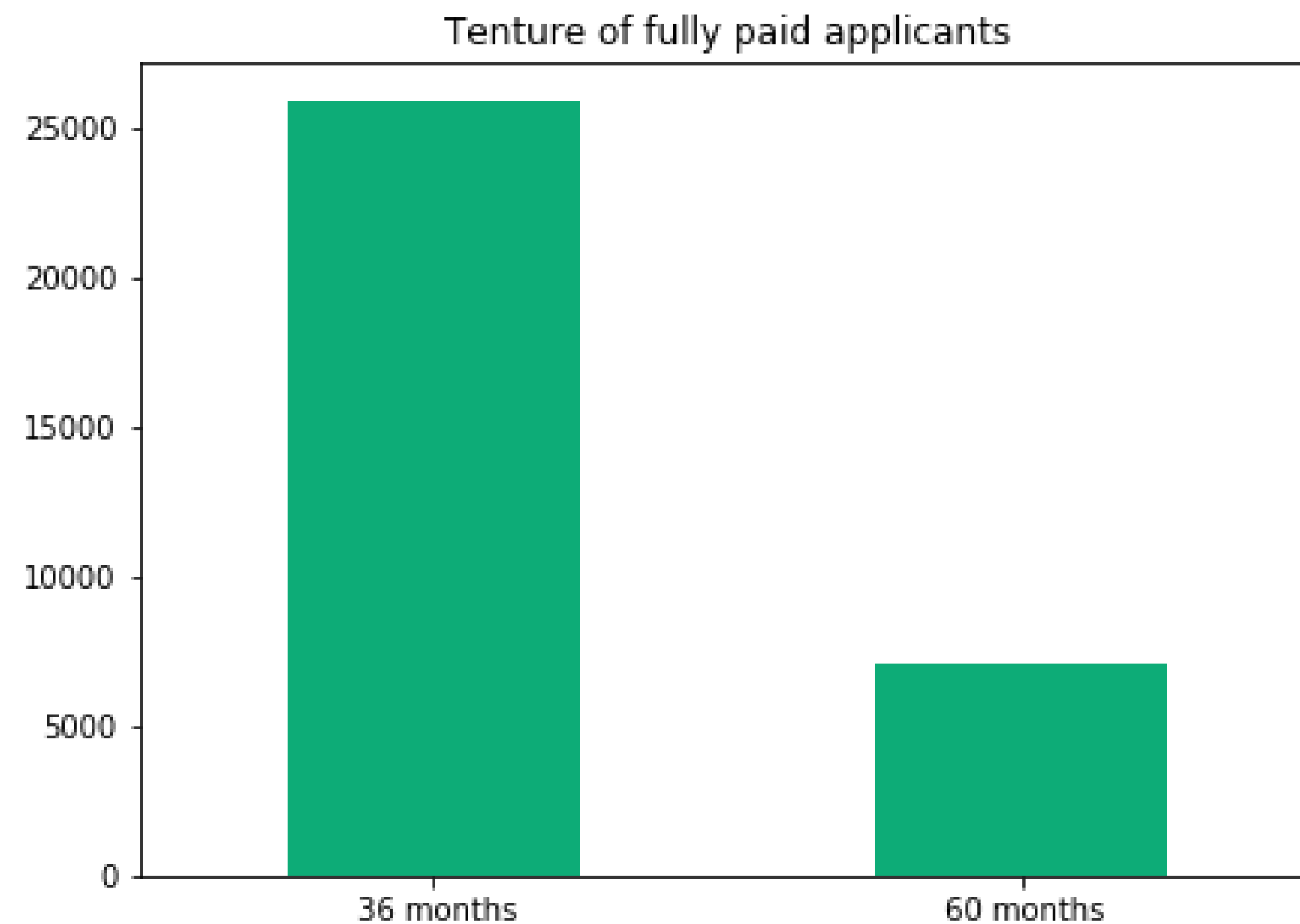
In this dataset we have
32950 – Fully paid customers,
5627 – Default customers

As per our analysis 84% of fully paid customers and 13% of default customers in this dataset

Data Analysis - Loan status by term

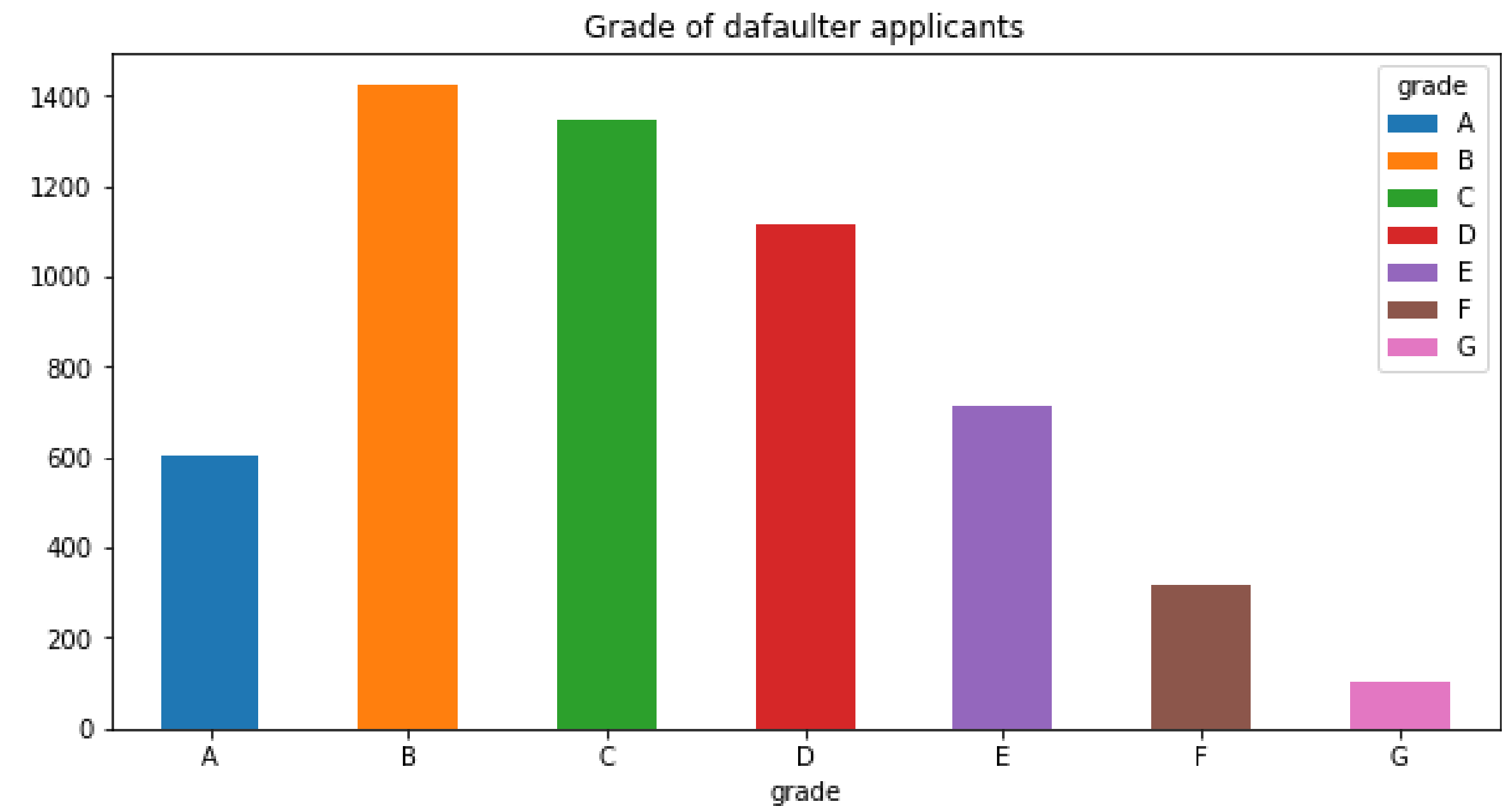
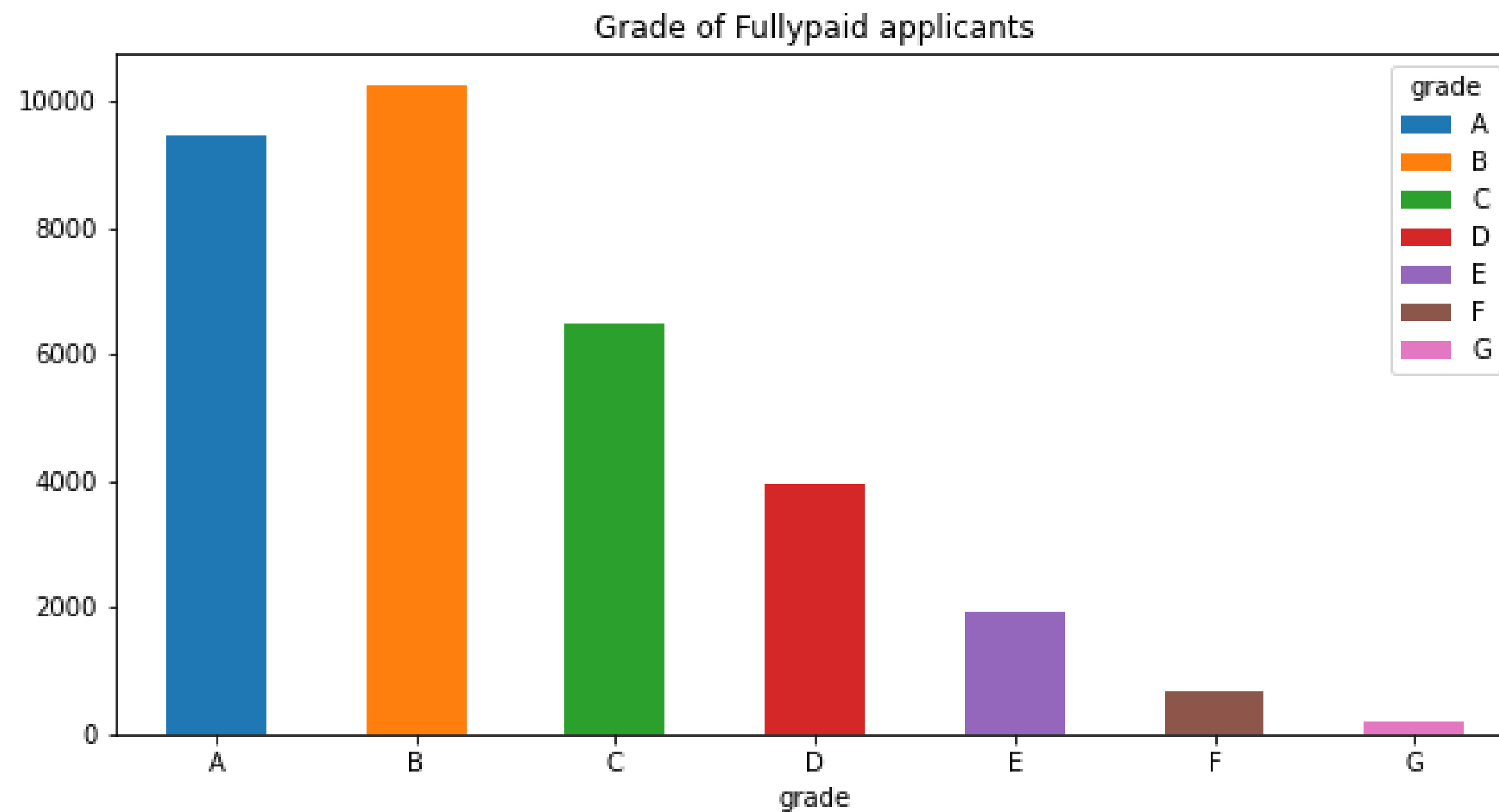


Data Analysis - Analysis based on Loan periods



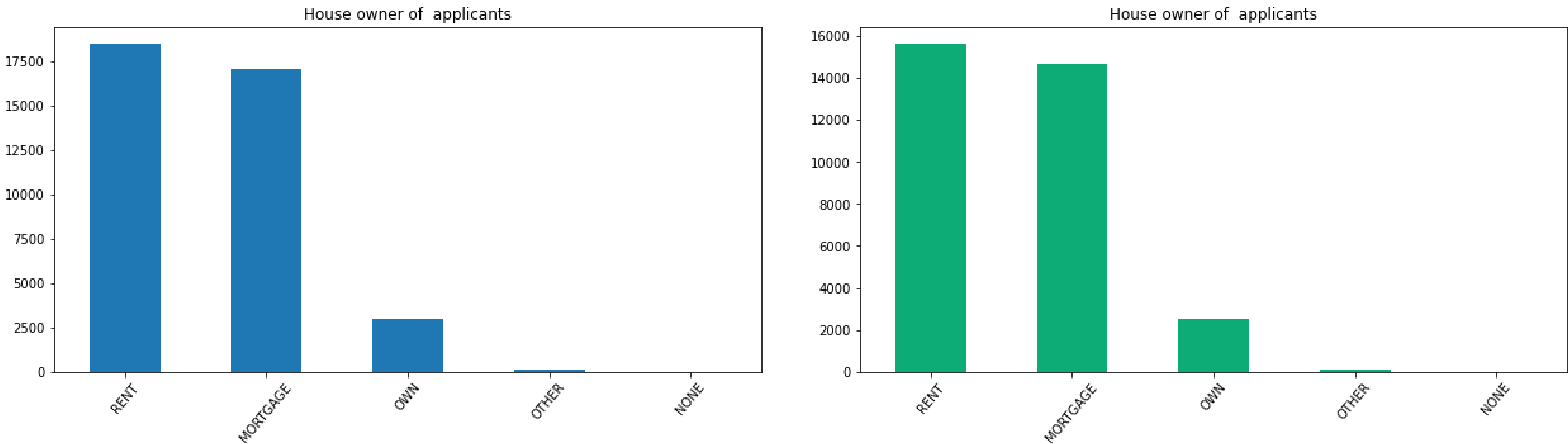
The impact of tenure (36 months and 60 months) of fully paid and default customers. The mostly default customers close to 58% of customers opted for 36 months of loan period. Comparatively 79% of fully paid customers opted for the 36 months of loan terms.

Data Analysis – Grade based analysis



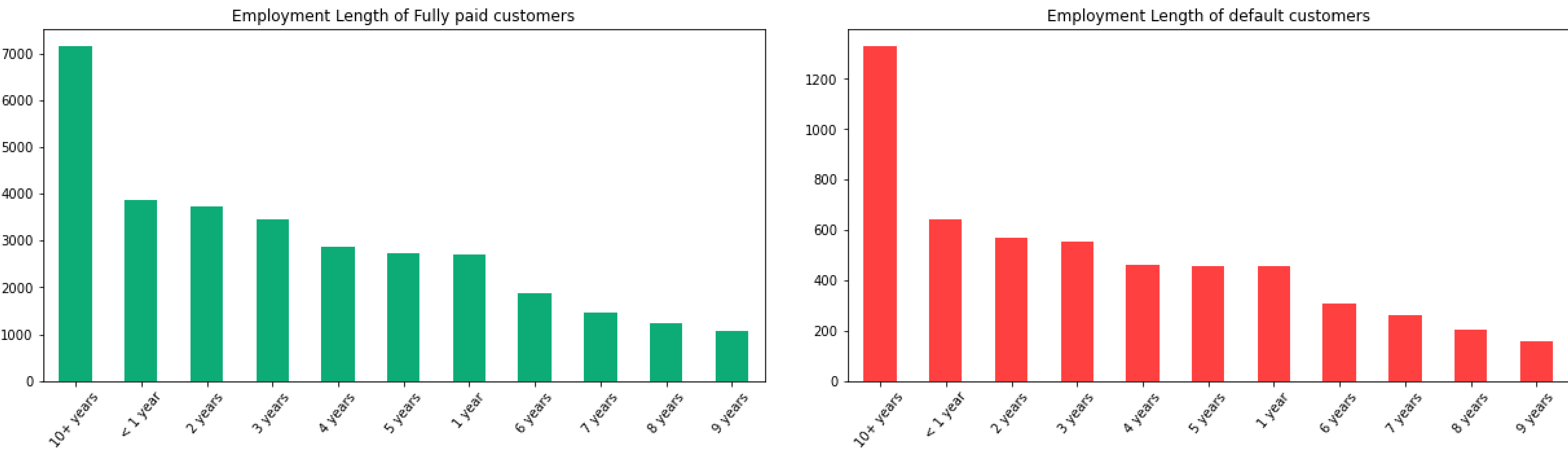
In this grade based analysis most of the defaulters comes under the B & C grade. 49% defaulter are belongs to these two grade. Grade G has low defaulters (1.8%) compared with other grade defaulters.

Data Analysis – Household by customers



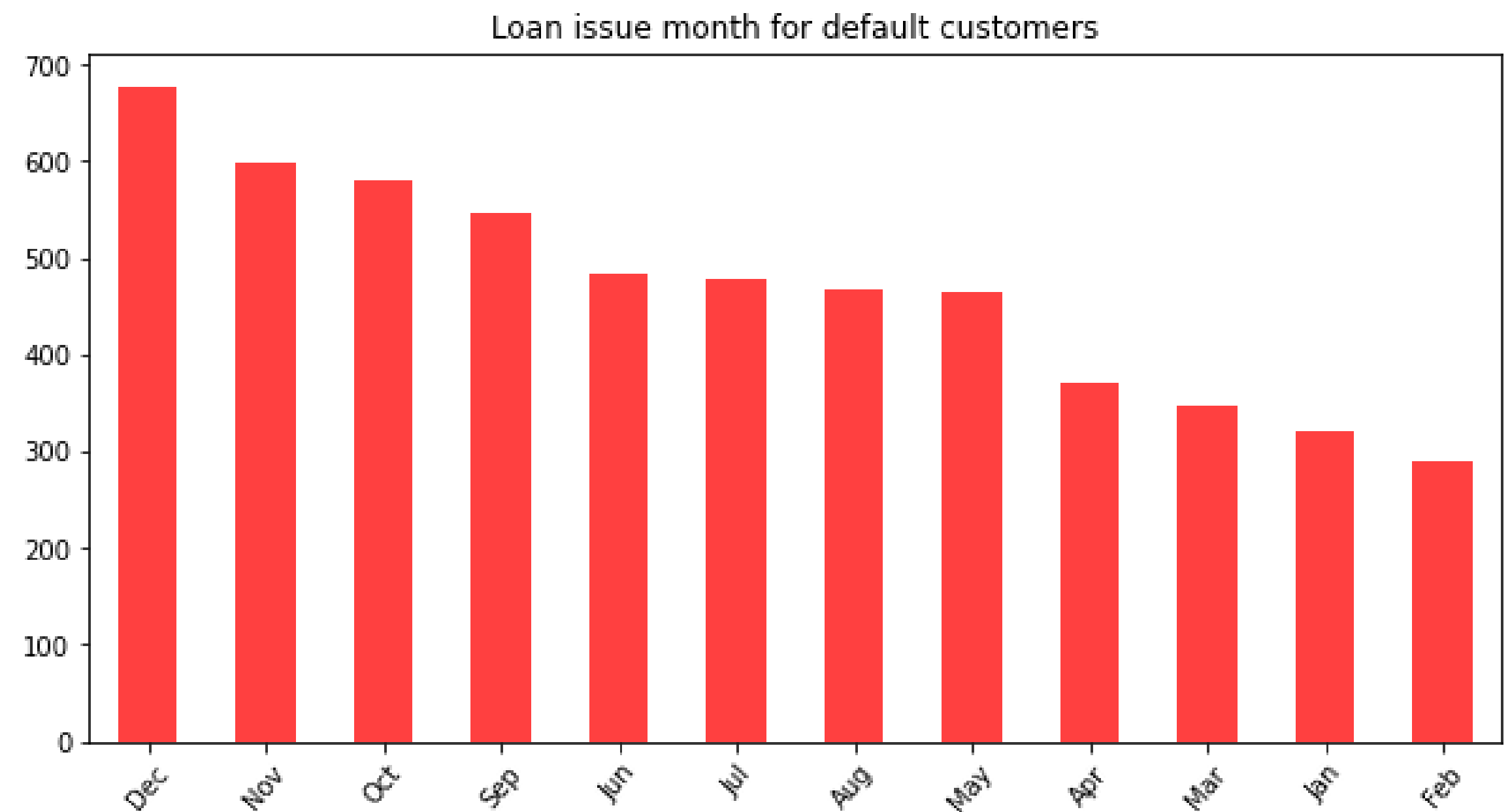
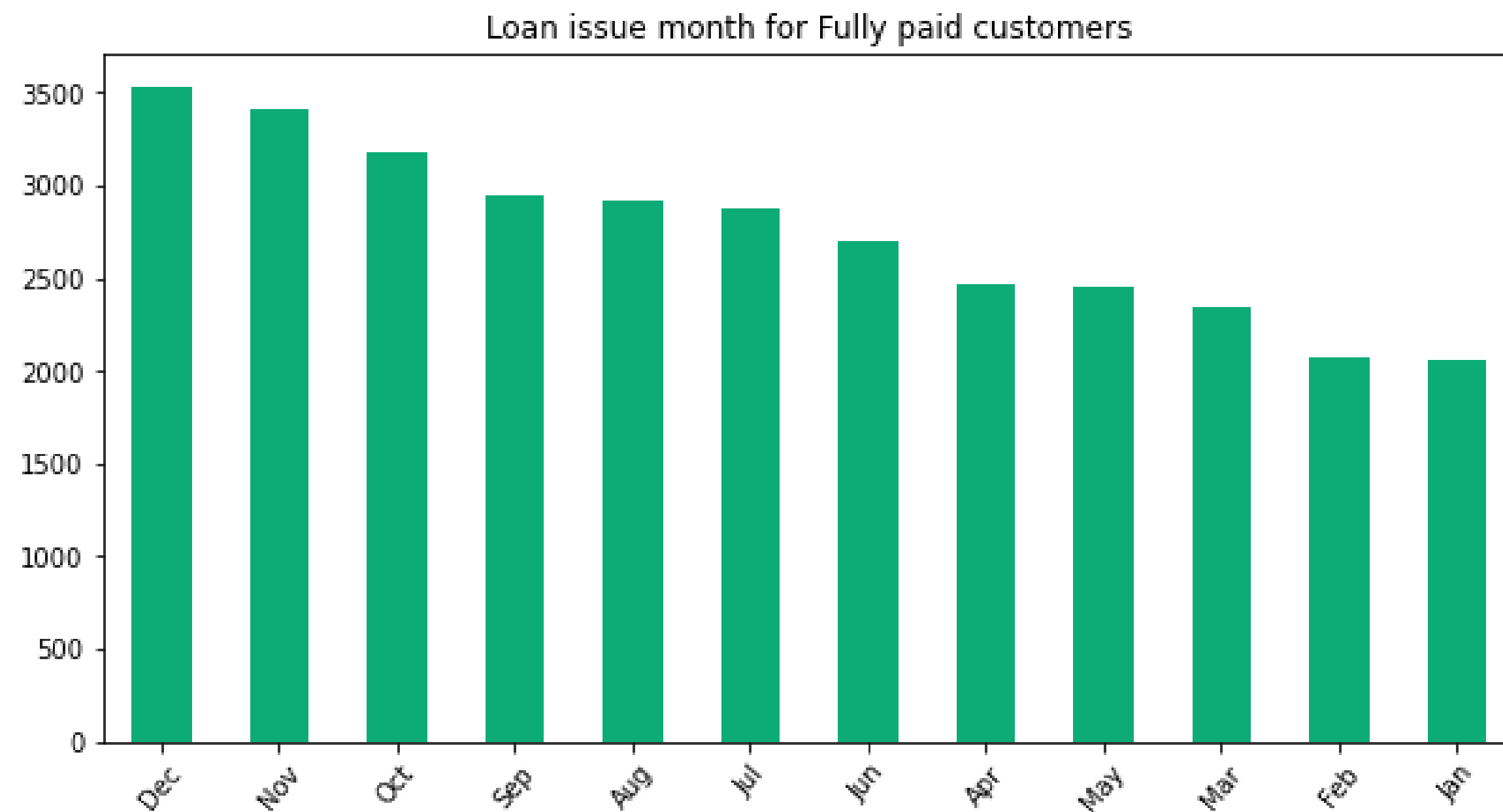
The most of the customers are staying in rented house and mortgage. The most default customers close to 3000 customers (51%) out of 5627 customers are staying in rented house. Who has own house, they’re very less like to be a defaulters (7%) customers.

Data Analysis – Employment length of customers



The employment duration is not having much difference of both type of customers. Both type of customers has equal ratio of employment duration. But who has 1 year and less than employment duration more likely to be defaulters compared with others.

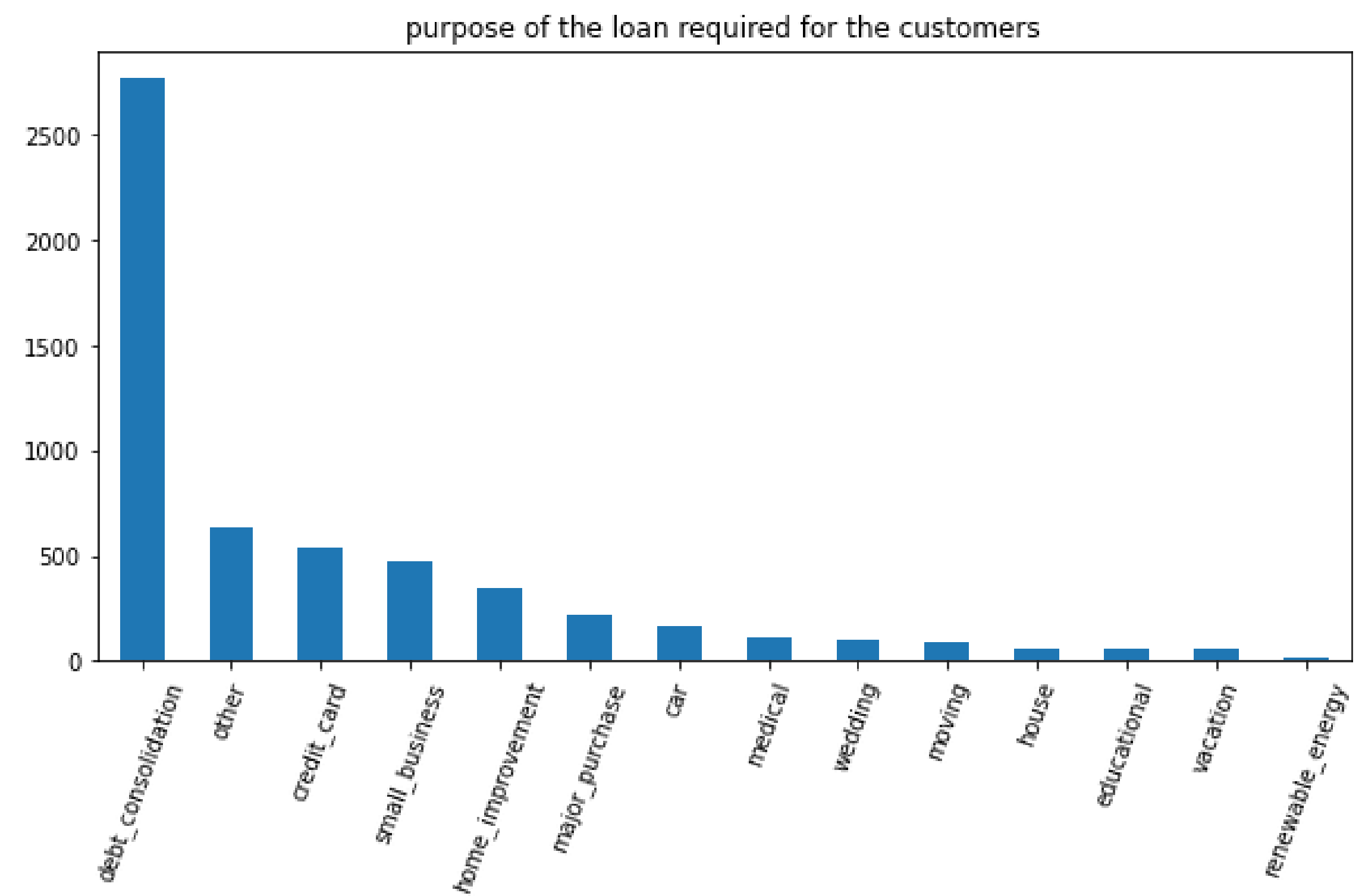
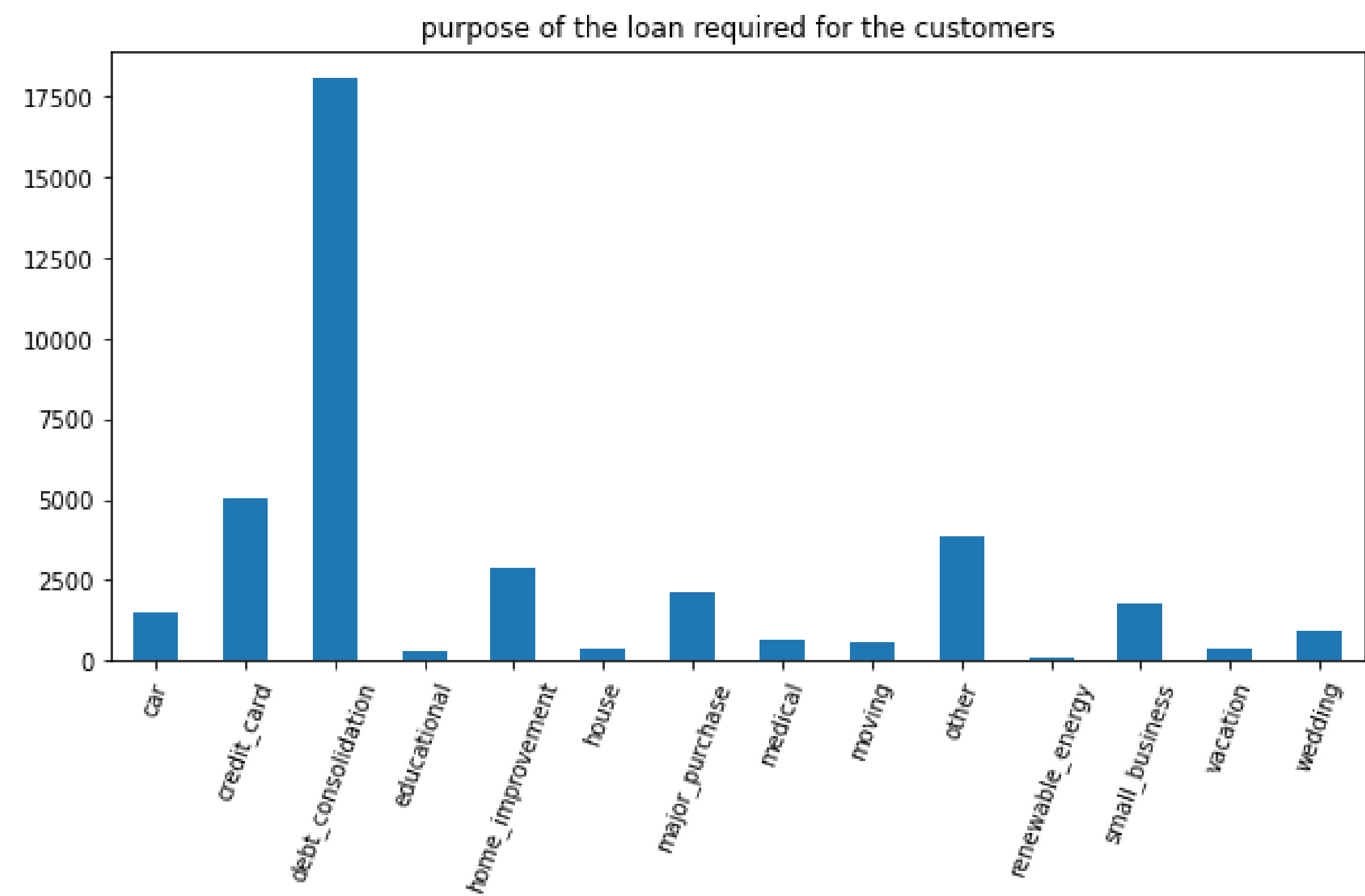
Data Analysis – Loan provide to the customers on particular period



The slight variance in customer who has taken the loan on the particular months. Comparatively customers who has taken loan in the month of June more likely to be defaulter apart from other months. This might be due to season health issue or having festivals etc..

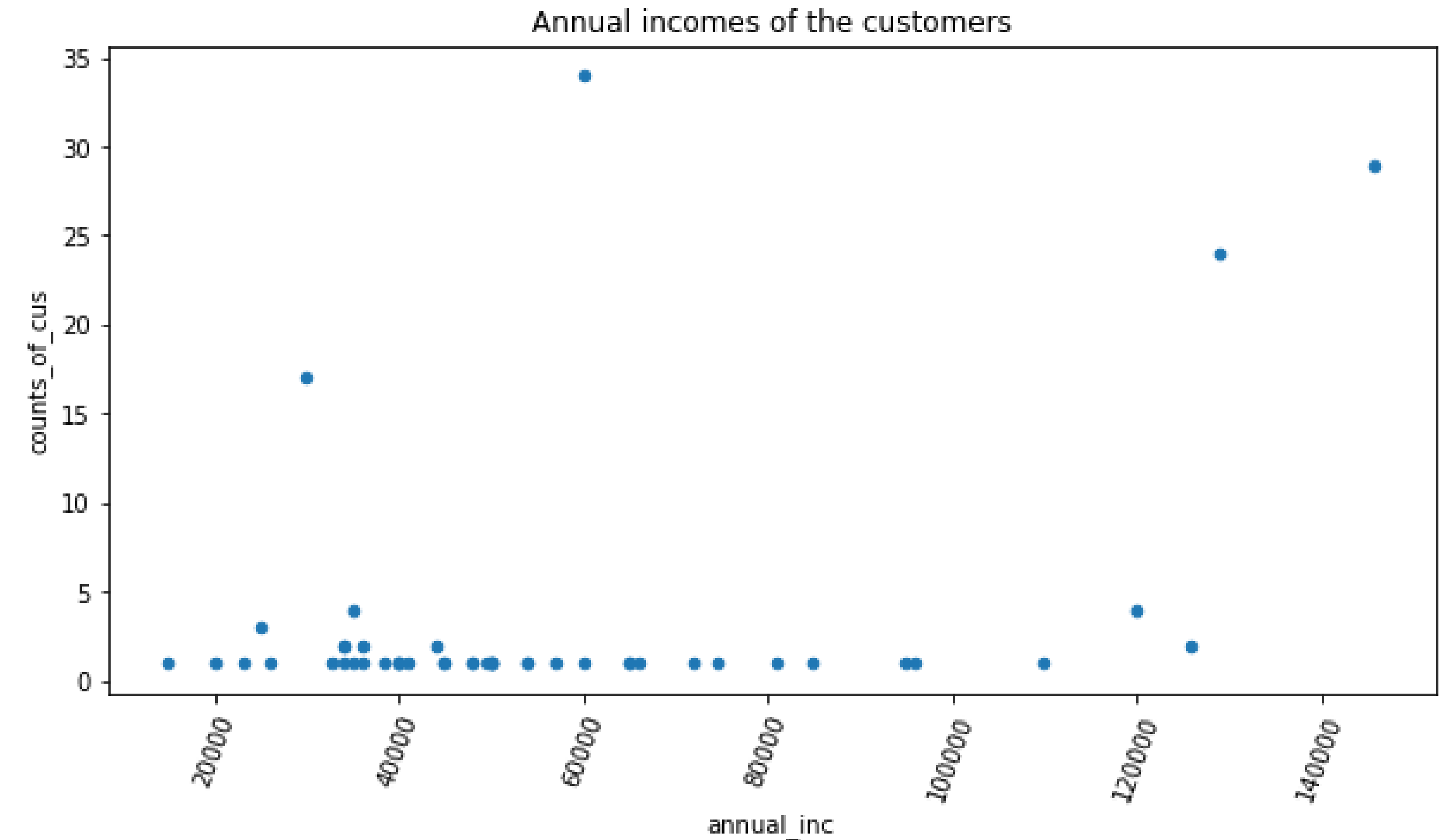
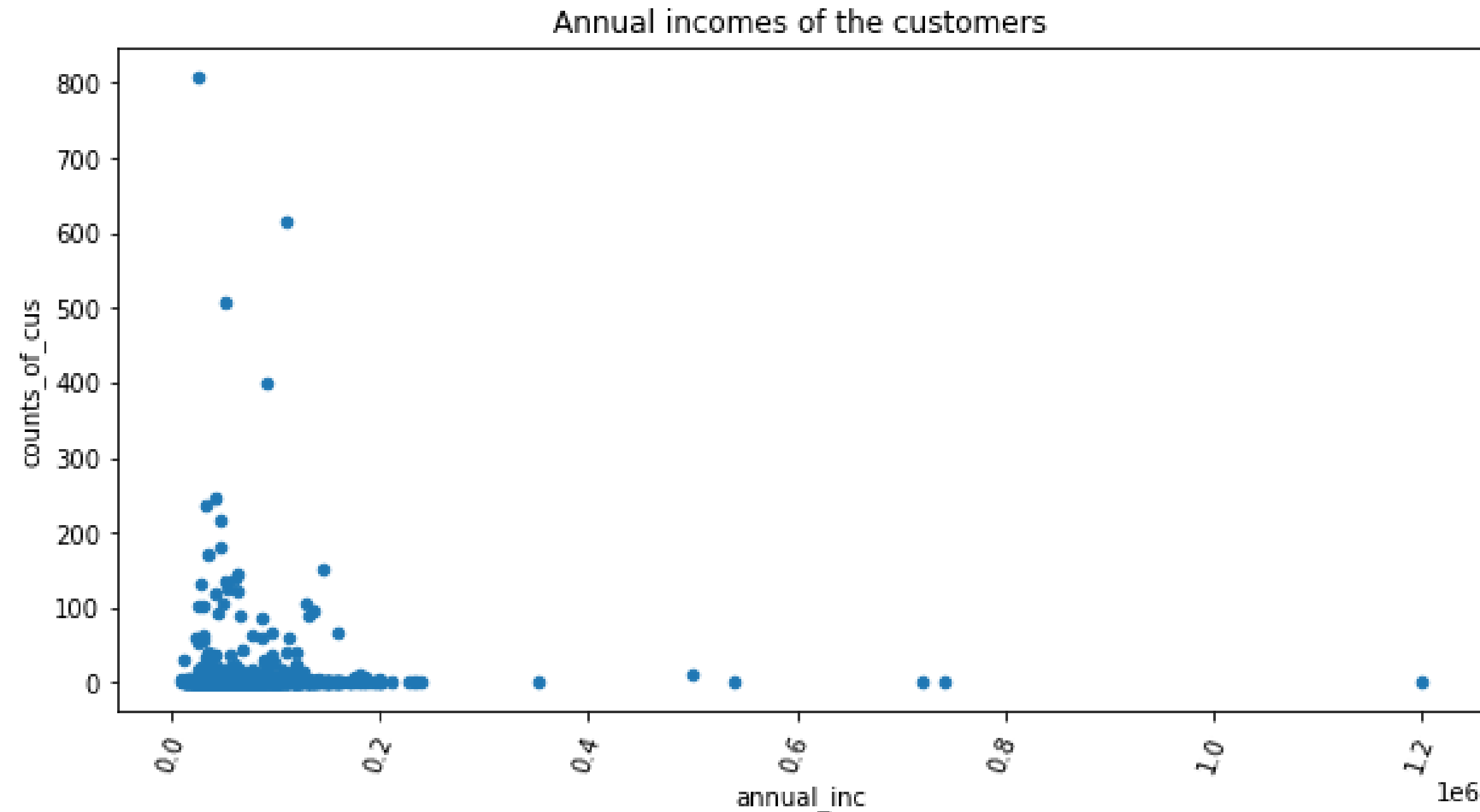
Especially in the month of Dec have festivals like Christmas and New year shopping. Due to this reason customer couldn't pay the EMI for the month. This leads to be defaulter.

Data Analysis – Purpose of the loan



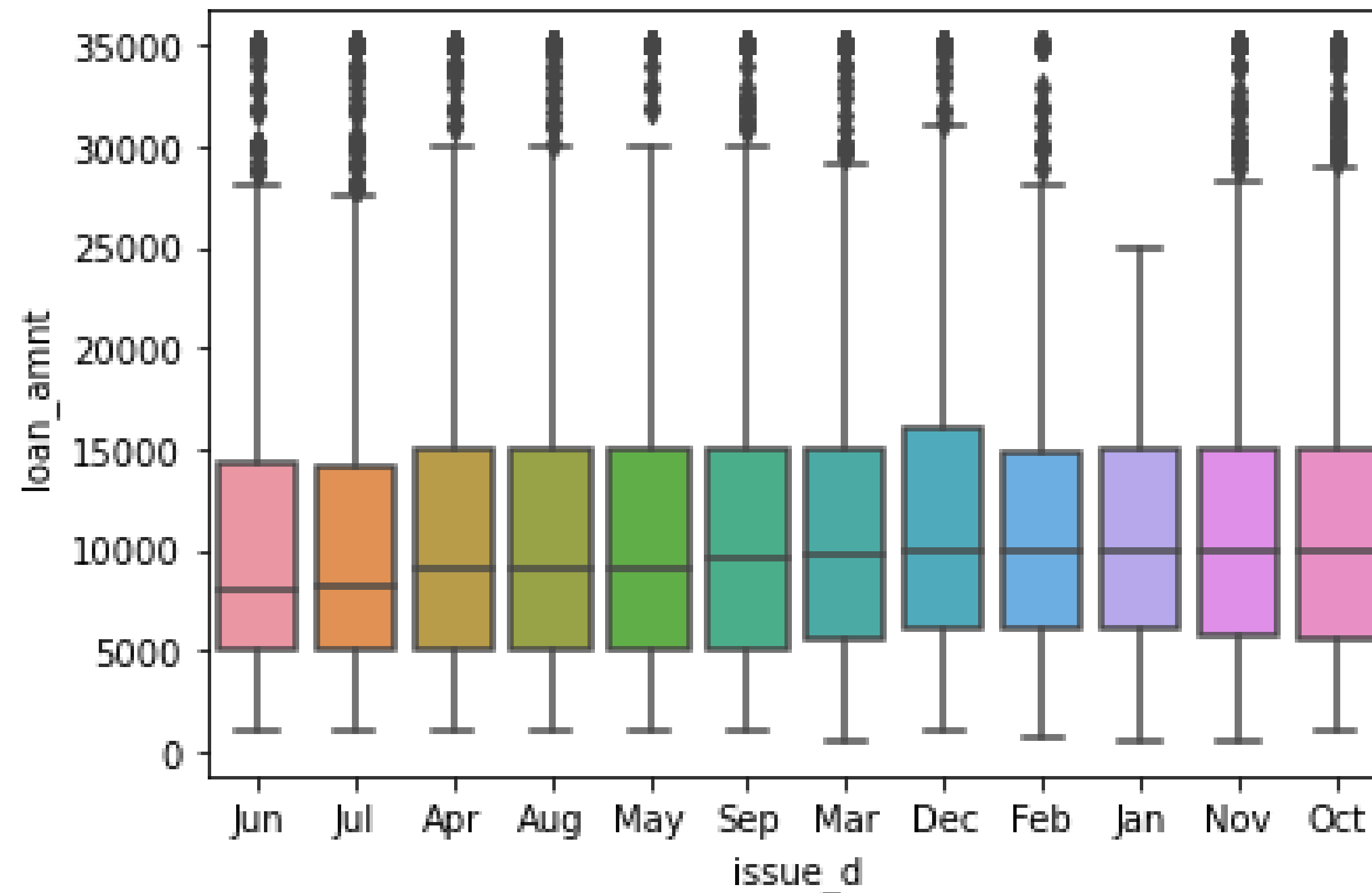
The loan provided to the customers in various category. The customer who has taken loan against debt consolidation more likely to be a defaulters. 49% of default customer who has taken the loan on debt. the customers who taken loan for renewable source has less defaulters (0.34 %).

Data Analysis – Customer annual income base analysis



The customer who has $< 20k$ of annual more like to applied for the loan and almost paid their loan amount. The customer who has close to 40K of annual income has applied loan less but for their category they would've opted more loan amount, but most of them couldn't pay on or before the time. So they're very much likely to defaulters compared with other customers.

Data Analysis – Minimum Loan amount provide particular period



The loan investors provide the average loan amount to customers for the particular period. Based on data we've month of Dec played major role in it and month of Jan likely to be less.

Almost all the month averagely investors given loan close to 10K.

Correlation Analysis



The correlated data - 98% of correlation between loan amount and funded amount.

It’s like the customer request amount and the investor paid amount are correlated.

The loan amount, funded amount and funded amount by investors are mostly correlated. When loan amount increases the funded amount also likely to be increased.

Conclusion

As per this analysis Grade, Term, Bin, interest rate, Year, and Home Ownership are the important variables. The defaulter are more like to be in rented house and takes the more duration of loan period and high amount of loan.

As per our suggestion the investor shouldn't give the requested full amount to the customers even though whoever eligible for higher amount.

The investor should consider the factor household by customer and total income for the customer. Based on these factors, they can provide lesser amount than the requested amount whoever staying apart from own house and whoever taking the higher loan amount. Consider these factor to avoid the risk of business.