Lending Club Case Study

By Anurag Singla / Prakash D

8 Oct 2023

Introduction

- **Lending Industry** - Lending is the growing industry in India every year where Companies provide loan to borrower on attractive interest rates.
- **Problem Statement** - Lending loans to 'risky' applicants is the largest source of financial loss. We need to understand driving factors to such applicants so that companies can take better decision and minimise the loss.
- **Analysis Approach** - Utilizing Exploratory Data Analysis (EDA) to Uncover Key Insights

Initial Observations about Data Set

Missing Values :

- There is no row with all null values
- There are columns in data set with null values
- There are columns which very few non null values and significant null values
- There are columns which single value across all rows

Standard Values issues :

- prefix space should be handled on term column
- % symbol should be removed from interest rate and credit limit consumed
- Employee Length should be modified into numeric value
- Date format should be changed on issue_d, earliest_cr_line, last_pymnt_d etc.
- round the decimal places into 2 digit for total_pymnt column
- Zip code suffix values should be fixed
- Remove outliers from annual income column

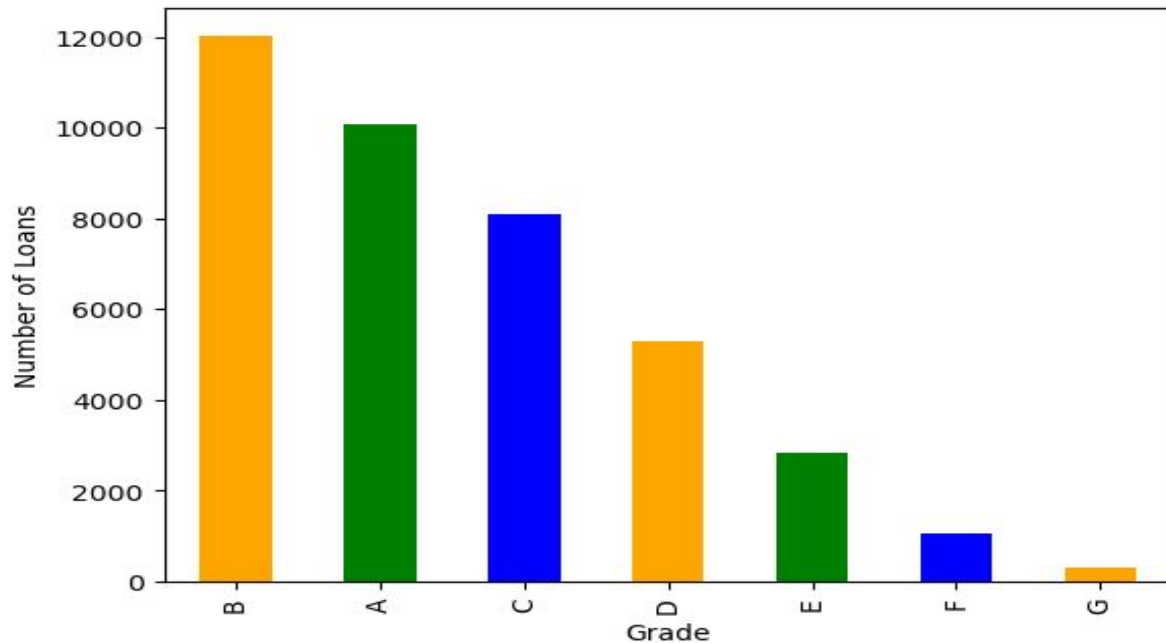**After Data Cleaning and filtering, We are left with 24 columns and 39714 rows.**

Univariate Analysis

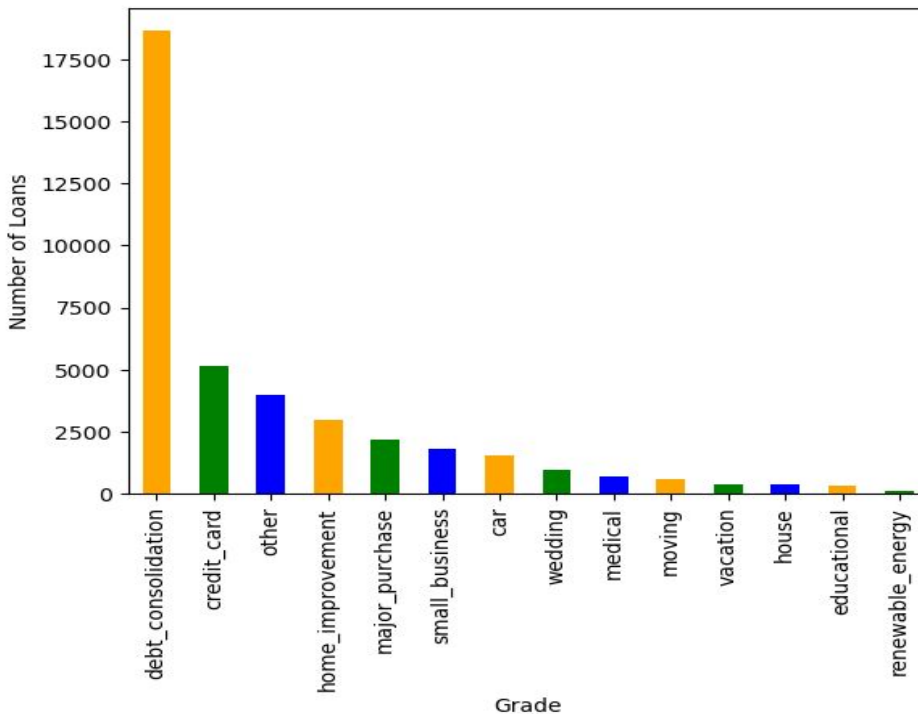**Number of applicants in 3 year tenure is way more than 5 years tenure**
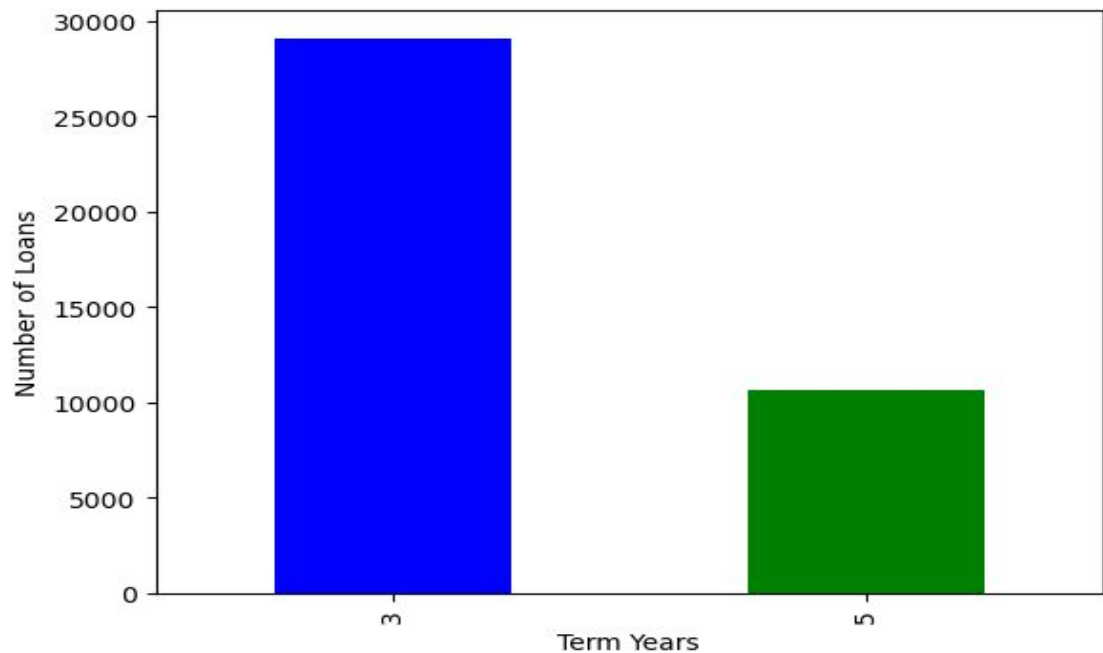
Univariate Analysis
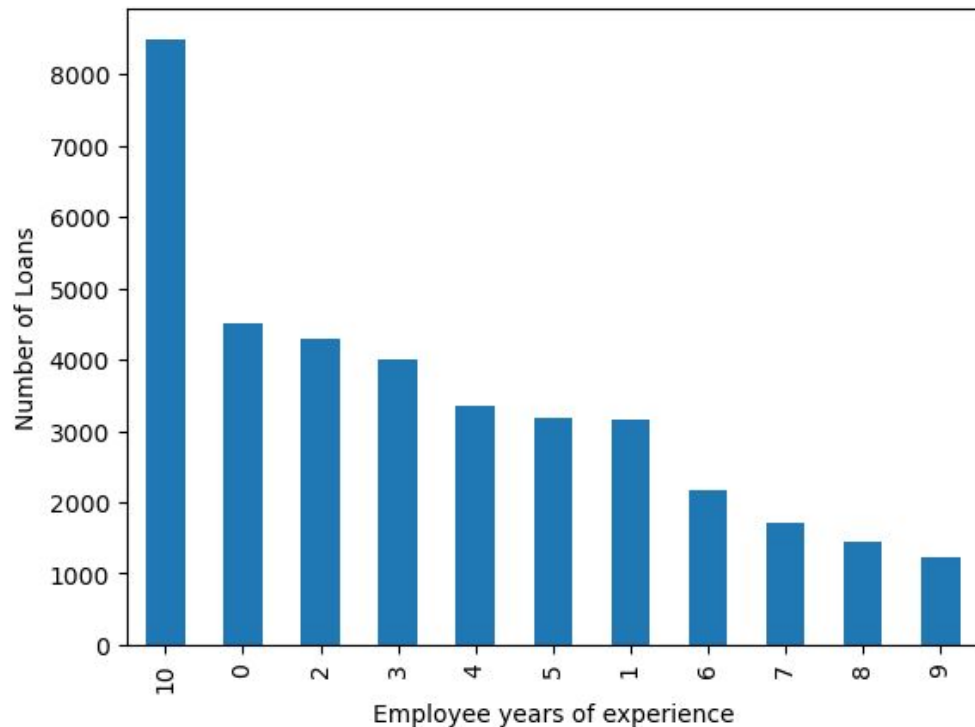
Surprisingly grade B has more volume than A

Univariate Analysis

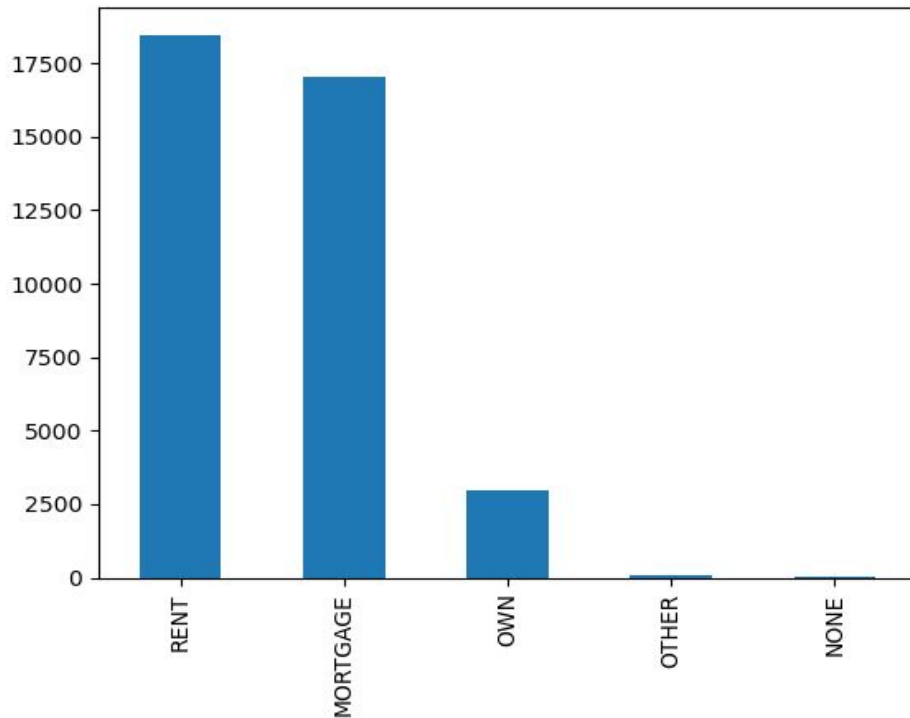**Most common purpose for loan is Debt Consolidation with very high volume**

Univariate Analysis

Most loans are issued to borrowers having more than 10 years of experience
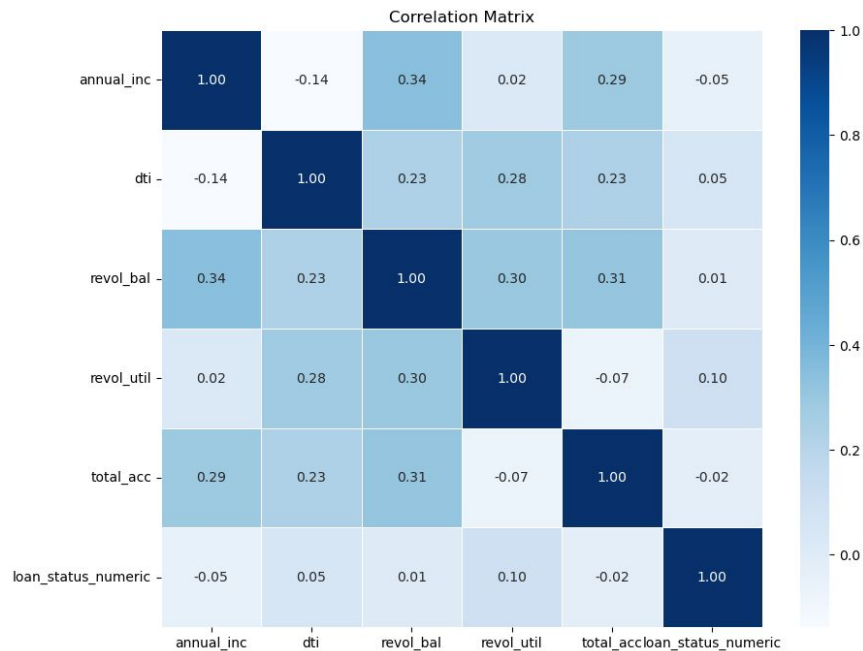
Univariate Analysis

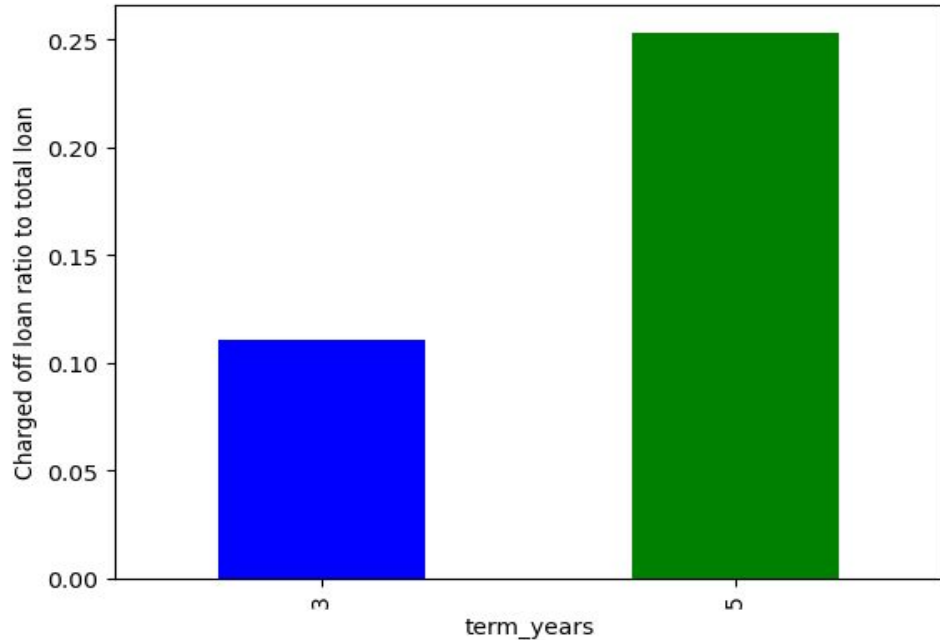**Most loans are issued to borrowers having rented or mortgage property**

Bivariate Analysis

Correlation metrics shows that annual income and total loan accounts have negative correlation with percentage charged off.
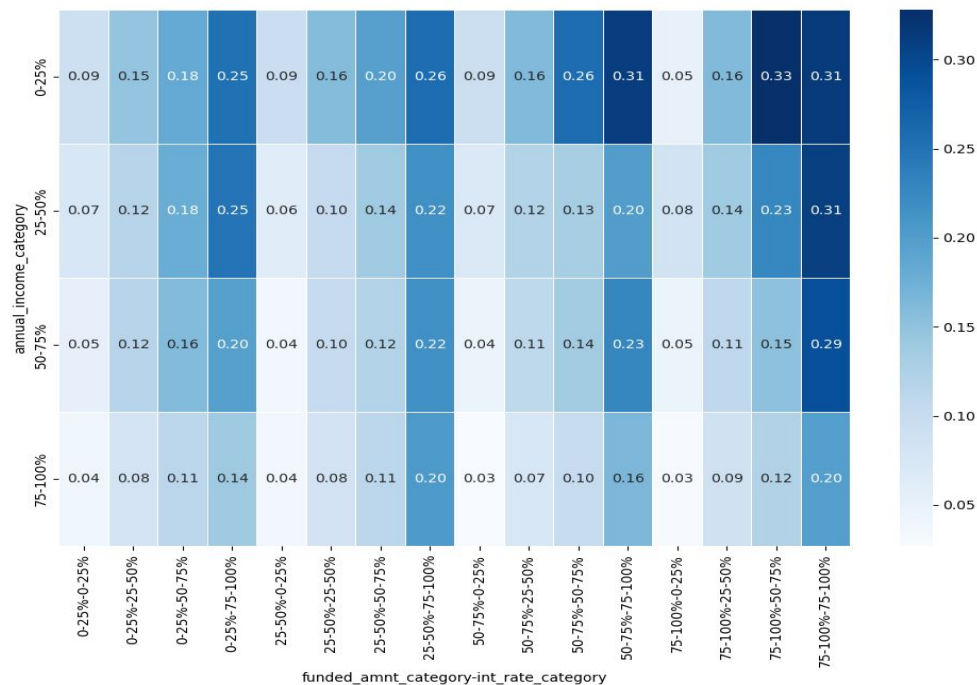
Bivariate Analysis

Loans with 5 year term has more defaulters even when total loans are less in number as compare to 3 years term loans
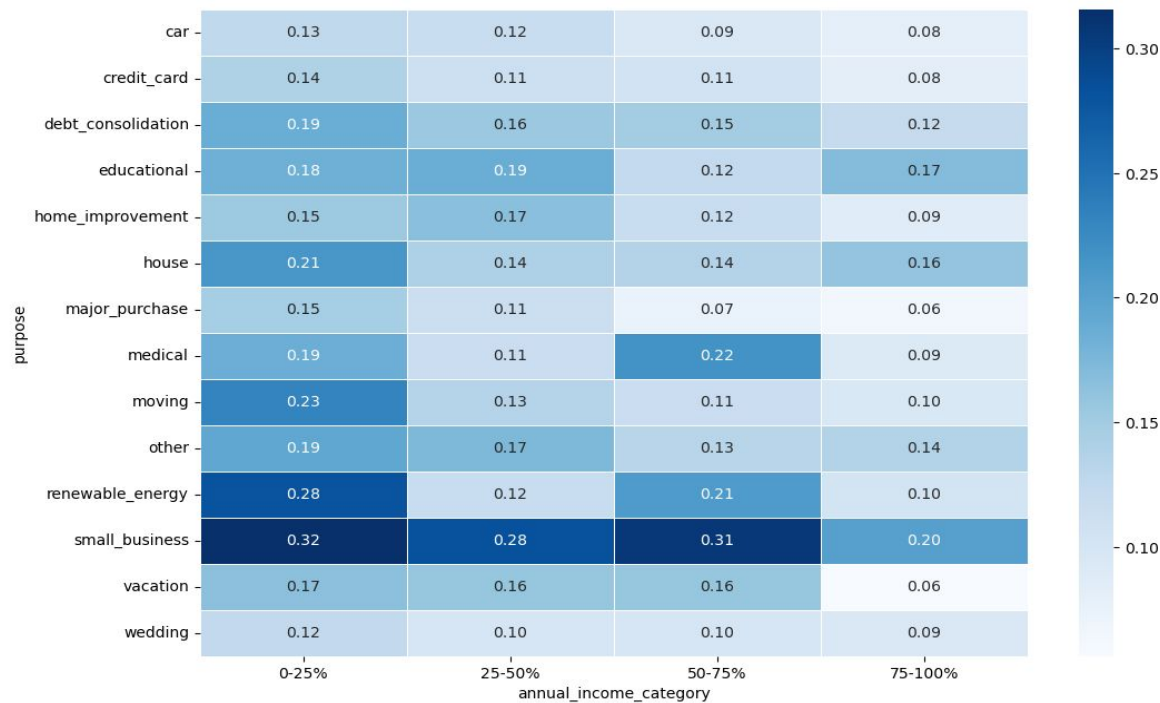
Bivariate Analysis

Lesser annual income with higher funded amount and interest rate tends to have higher defaulters
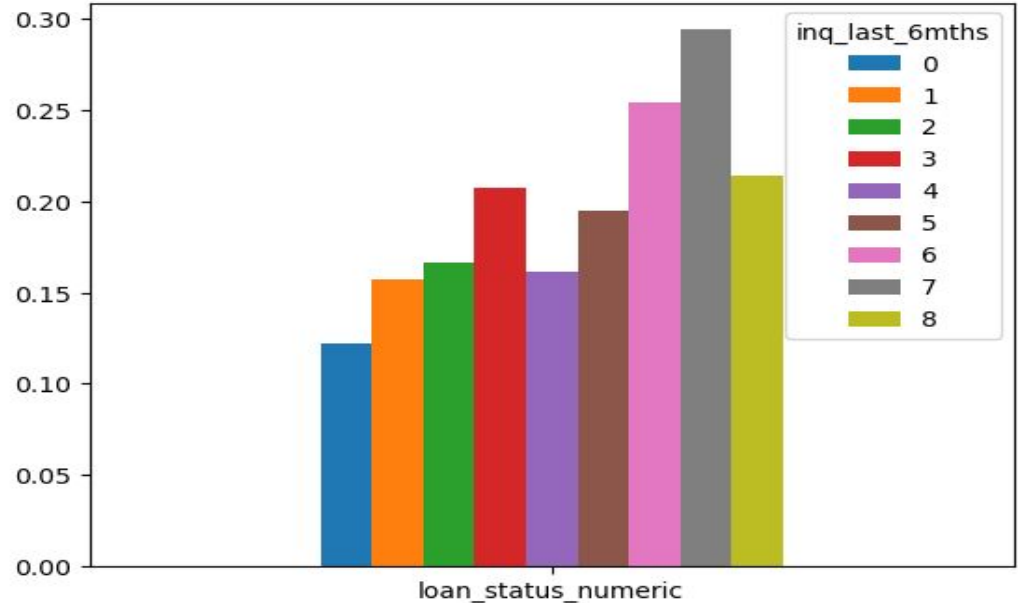
Bivariate Analysis

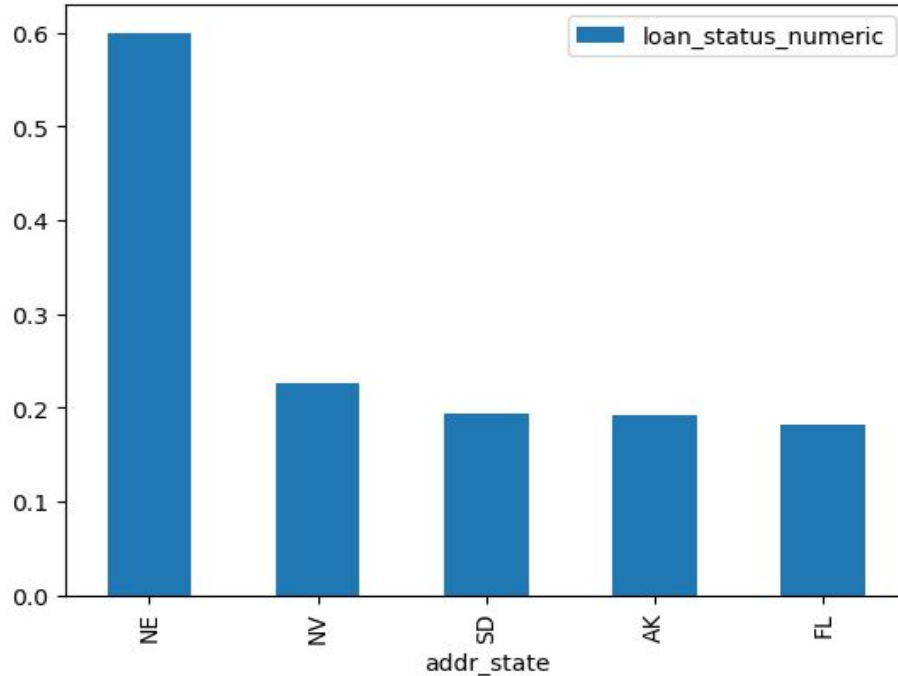## Small business and Renewable energy likely to defaulter

Bivariate Analysis

Higher number of enquires in some purposes like education, small business have very high probability of defaulters.
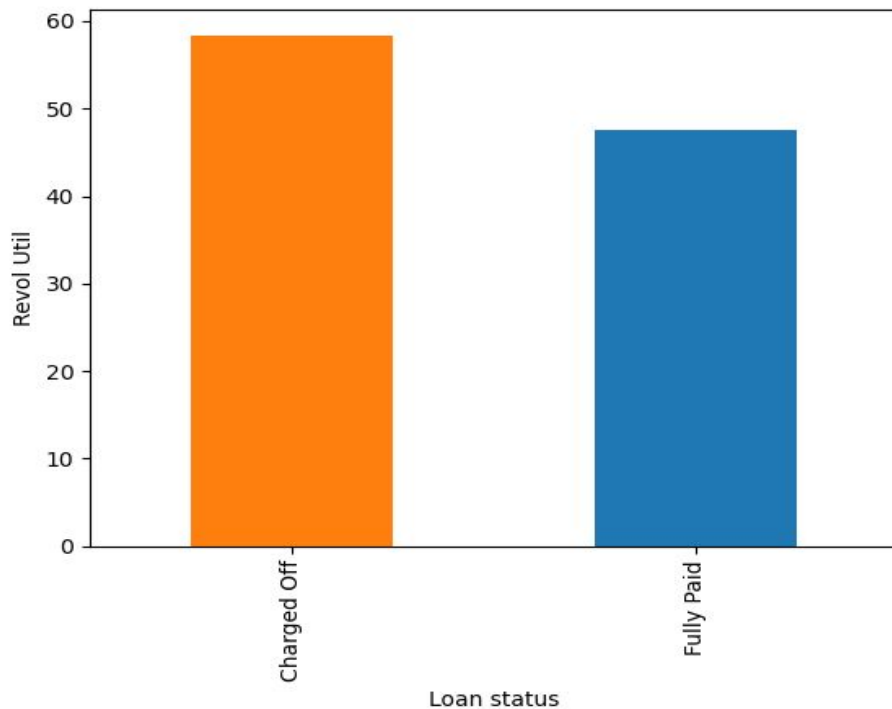
Bivariate Analysis

if address state of borrower is NE, then it is very high chance of defaulter.

Bivariate Analysis

Higher consumption of credit limit shows higher chances of defaulter

Results

- **Purpose -** Small business and Renewable energy likely to defaulter
- **Number of Enquires in last 6 months -** if number of enquires are higher is number, then chances of charged off inceases
- **Interest Rate, Funded Amount, Loan Amount -** in above graph, we can see if annual income is less and interest rate and funded amount is high, then probablity of charged off is higher
- **home_ownership -** Other Category of home_ownership likely to default more
- **Address State -** if address state of borrower is NE, then it is very high chance of defaulter.
- **Term -** Loans with 5 year term has more defaulters even when total loans are less in number as compare to 3 years term loans
- **Credit Limit Consumption** - Higher credit limit consumption shows higher chances of defaulter.