Lender Club Case Study

Vijay Tiwary

Gaurav Bhardwaj

7th June 2022

Problem Statement

Consumer Finance Company's primary business is to lend different type of loans to the customer. For any loan application it receives, there are two types of risks associated with it:-

- Not approving loan of the customers who are very likely to repay the loan is a loss of business opportunity to the company
- Approving loan of the customers who are not likely to repay the loan will result a loss to the company in future

Objectives

- To identify patterns by analyzing past loan application of the both types of customers – defaulters and non defaulters. Use this patterns to indicate if any new loan applications is likely to default or not.
- Exploratory Data Analyses techniques will be used to identify the attributes which influences the tendency of default.
- These attributes will be used to analyze while approving/rejecting any future loan application

Major steps of EDA analysis

- Data understanding
- Data Cleaning
- Data Analysis
- Recommendation Summary

Data Understanding

- The attributes can mainly be divided under two categories
  - Loan attributes – These would be used in our analysis
  - Customer attributes – These attributes like delinq_2yrs, earliest_cr_line etc. are generally available post loan approval and so cannot be made basis for analysis of new loan application. These columns will be removed in the data cleaning phase.

- There are many columns like pymnt_plan, initial_list_status etc. which have just single value across all cells. These attributes will not contribute to the analyses and will be deleted in the cleaning phase.

Data Understanding contd…

- Data where loan status is either Charged Off or Fully Paid is relevant for the analysis
- Out of these attribute – loan_amnt, funded_amnt and we will just use funded_amnt_inv in our analyses as this is the actual amount received by the customer
- Categorical columns -'term', 'grade', 'sub_grade', 'emp_length', 'home_ownership', 'verification_status', 'loan_status', 'purpose', 'pub_rec_bankruptcies', 'issue_month’
- Continuous columns - 'loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'int_rate', 'installment', 'annual_inc', 'dti'

Data Understanding contd…

- Derived columns
  - Issue_month – This attribute will be created from issue month as this is value which would repeat for new customers, and we need to identify if there is any patterns related to this field.
  - monthly_inc_to_inst_ratio – This derived attribute is the ratio between installment and monthly income as same installment amount for a high salary customer and low salary customer would have different impact.

Data Cleaning

The following steps have been performed for cleaning the data

- Removing columns which have blank values for more than 80% of the cells
- Removing customer attributes.
- Removing attributes with single value across all cells.
- Removing free text columns which are not relevant for analysis
- Removing all rows with loan_status as Current
- Removing ‘%’ from int_rate

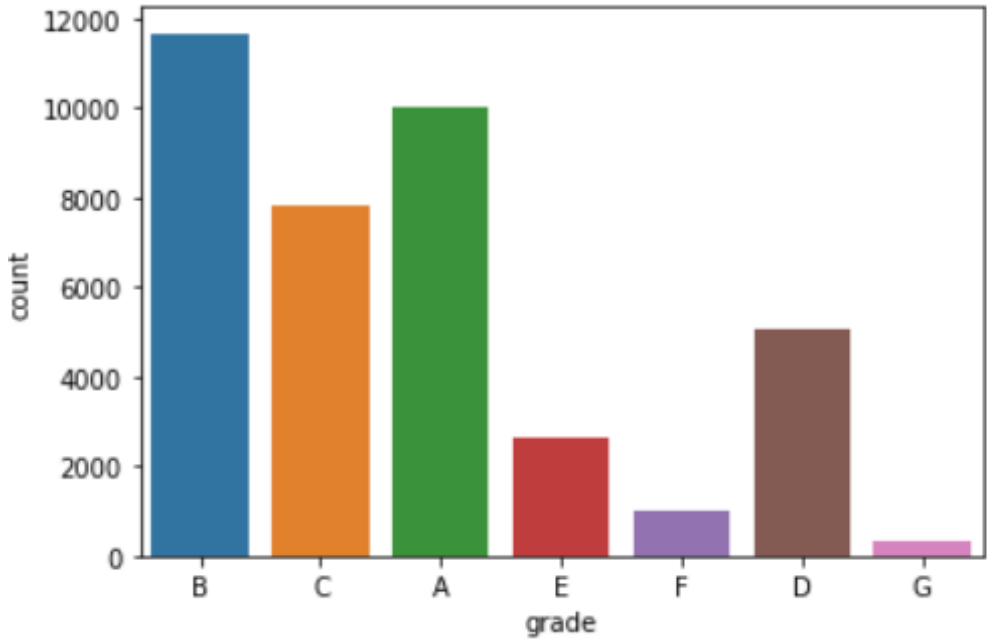
Data Cleaning contd…

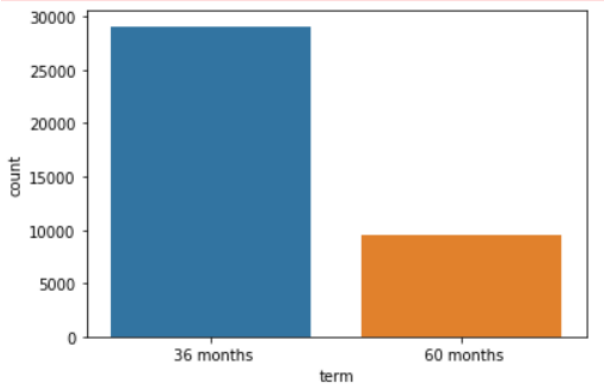
- Emp_length - values <1 to be considered as 1, missing values to be considered as 0 and > 10 to be considered as 10,
- pub_rec_bankruptcies – Missing values to be considered as 0, as the median is 0

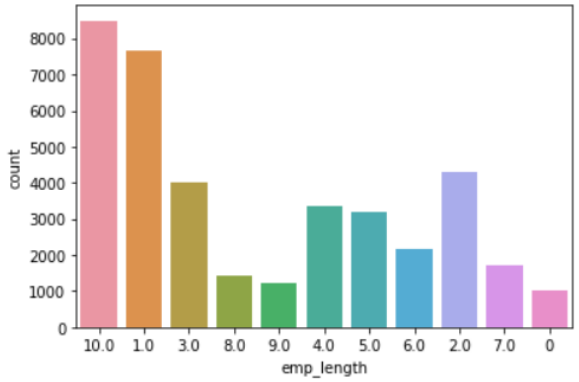
Analyses

Univariate Analyses

Countplot categorical variable to see the distribution

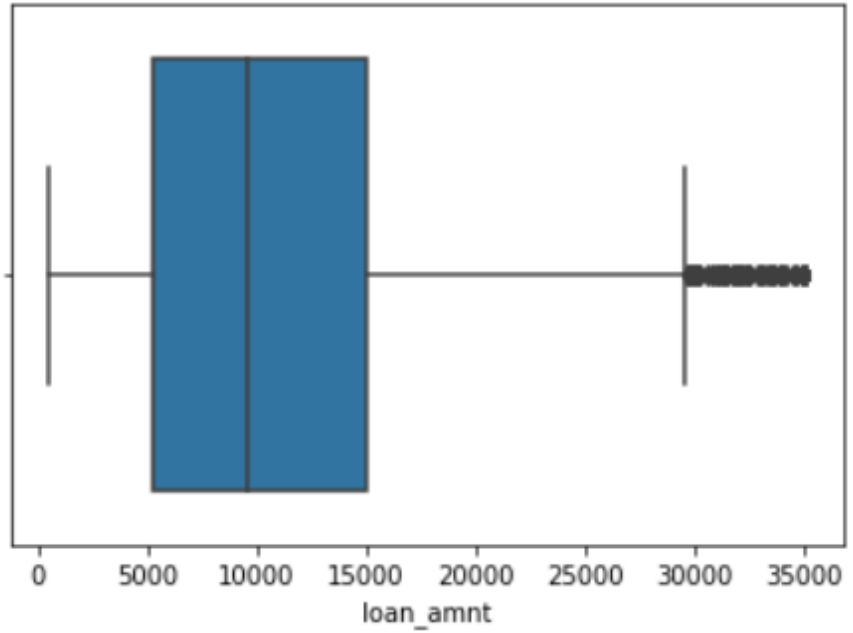


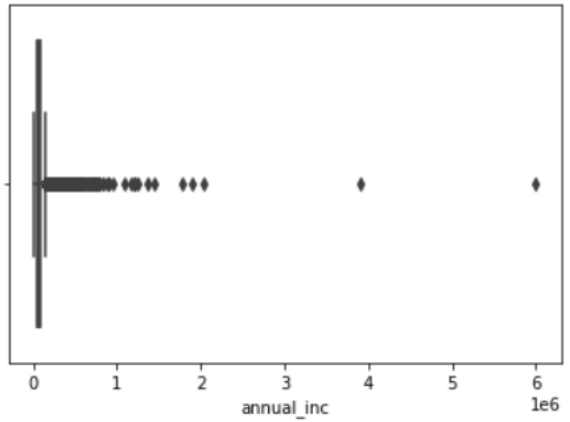


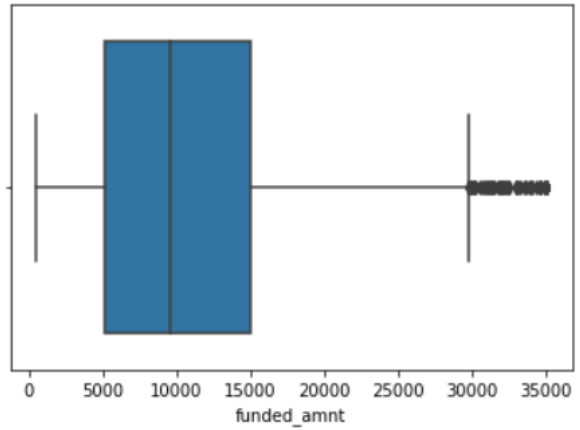


Univariate Analyses

Box plot all the continuous variables to find outliers

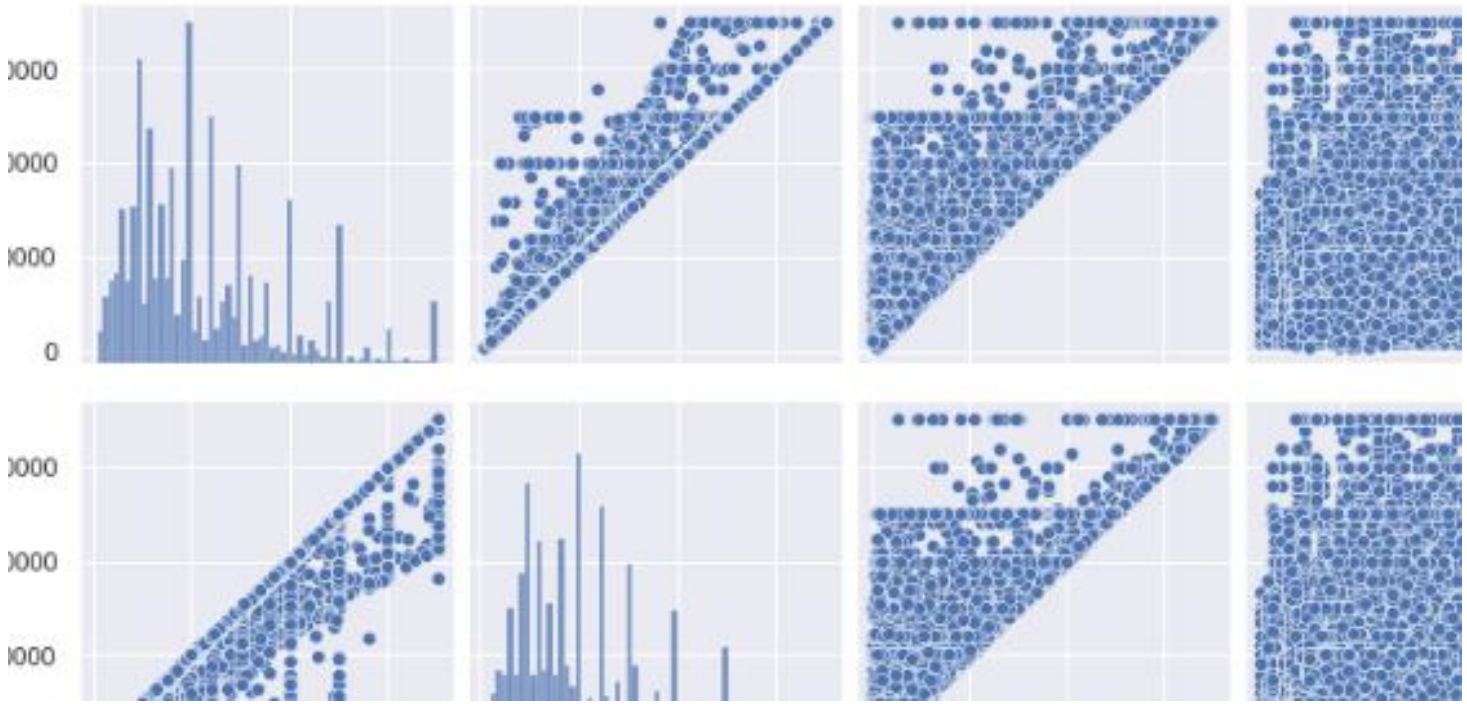






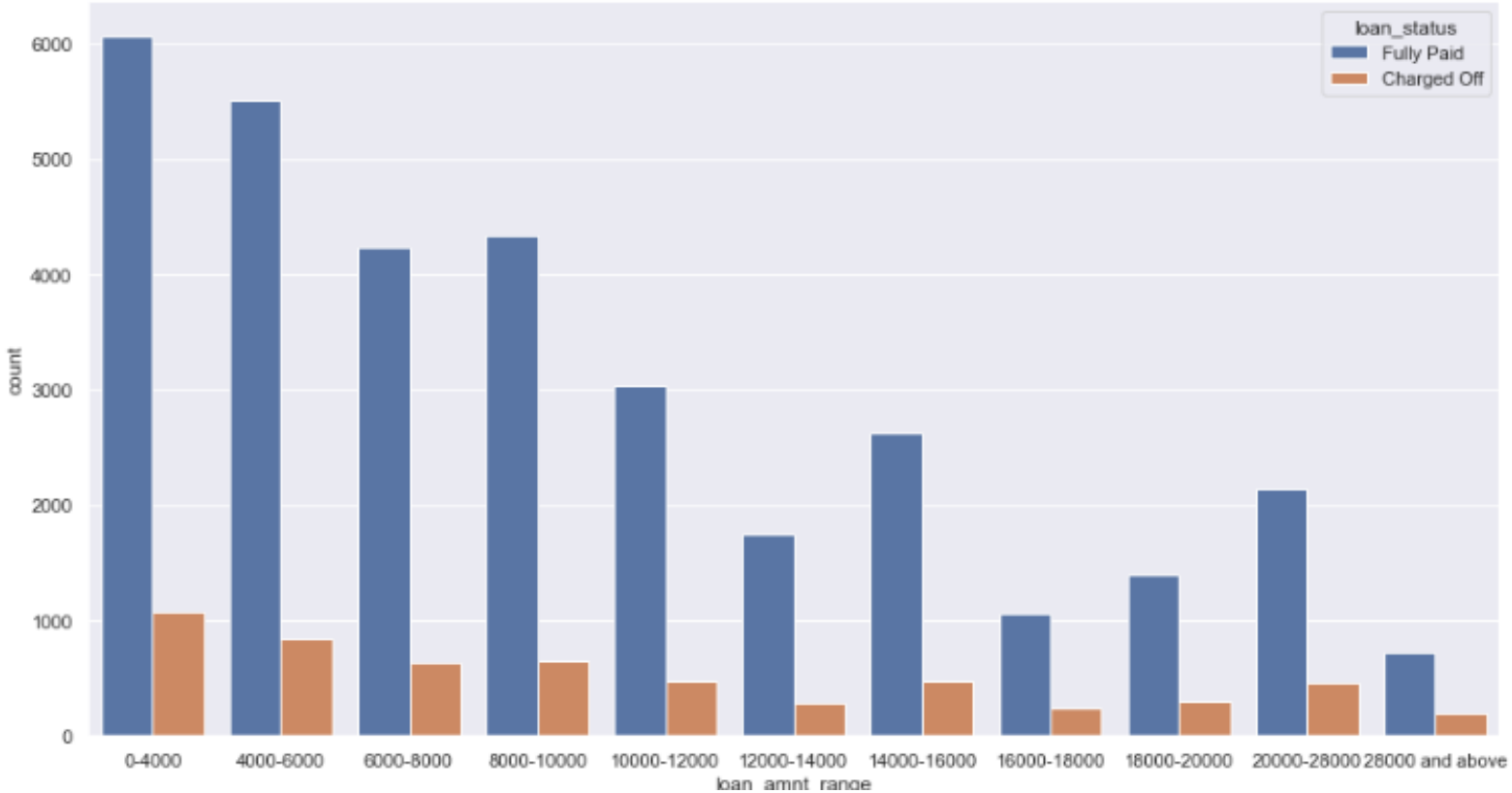
Bivariate Analyses

- Pair plot for the continuous variables doesn’t shows any positive or negative correlation between the variables



Bivariate Analyses

- funded_amnt_inv vs loan_status

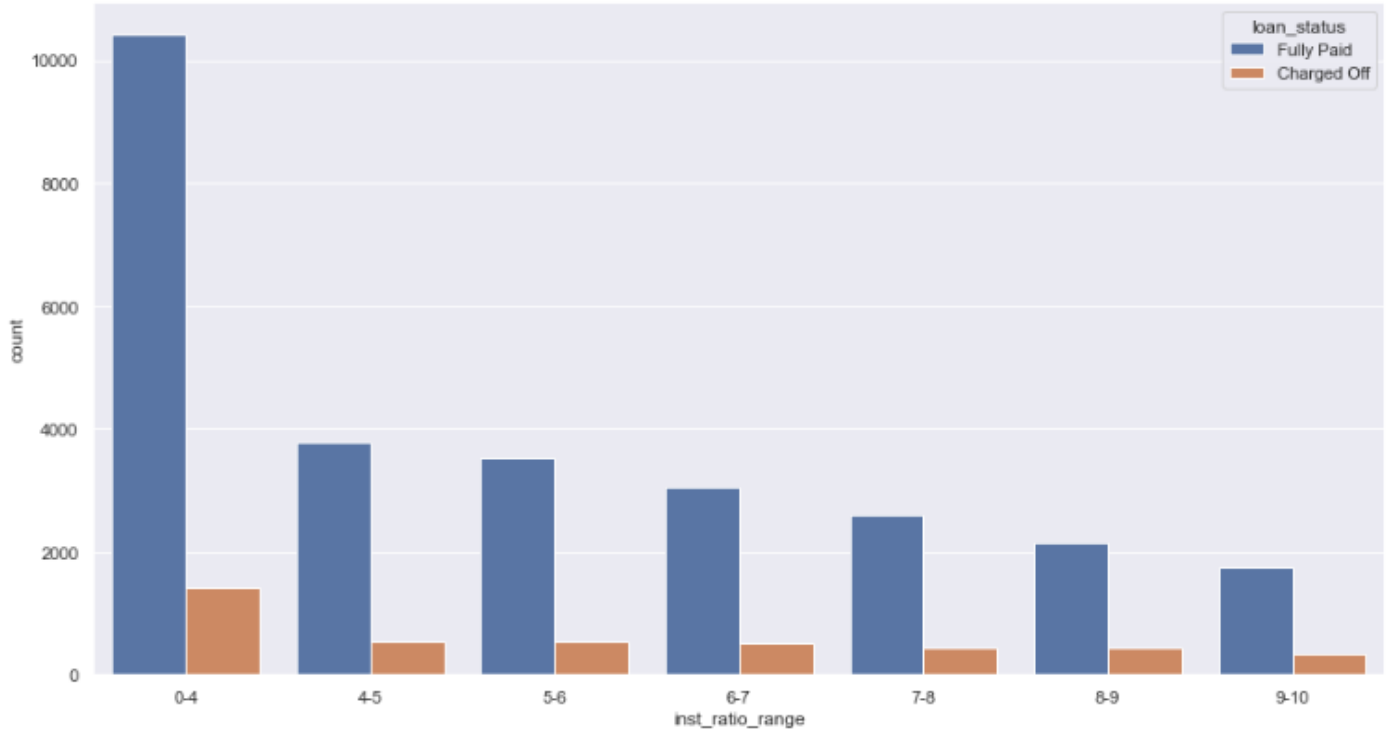



```
Charged Off percentage when funded_amnt_inv < 10000 =  13.724430026133188
Charged Off percentage when funded_amnt_inv is between 10000 and 18000 =  14.801888729213713
Charged Off percentage when funded_amnt_inv > 18000 =  18.316355948491257
```

Higher the loan amount, higher are the chances of default by customer

Bivariate analyses

- monthly_inc_to_inst_ratio vs loan_status



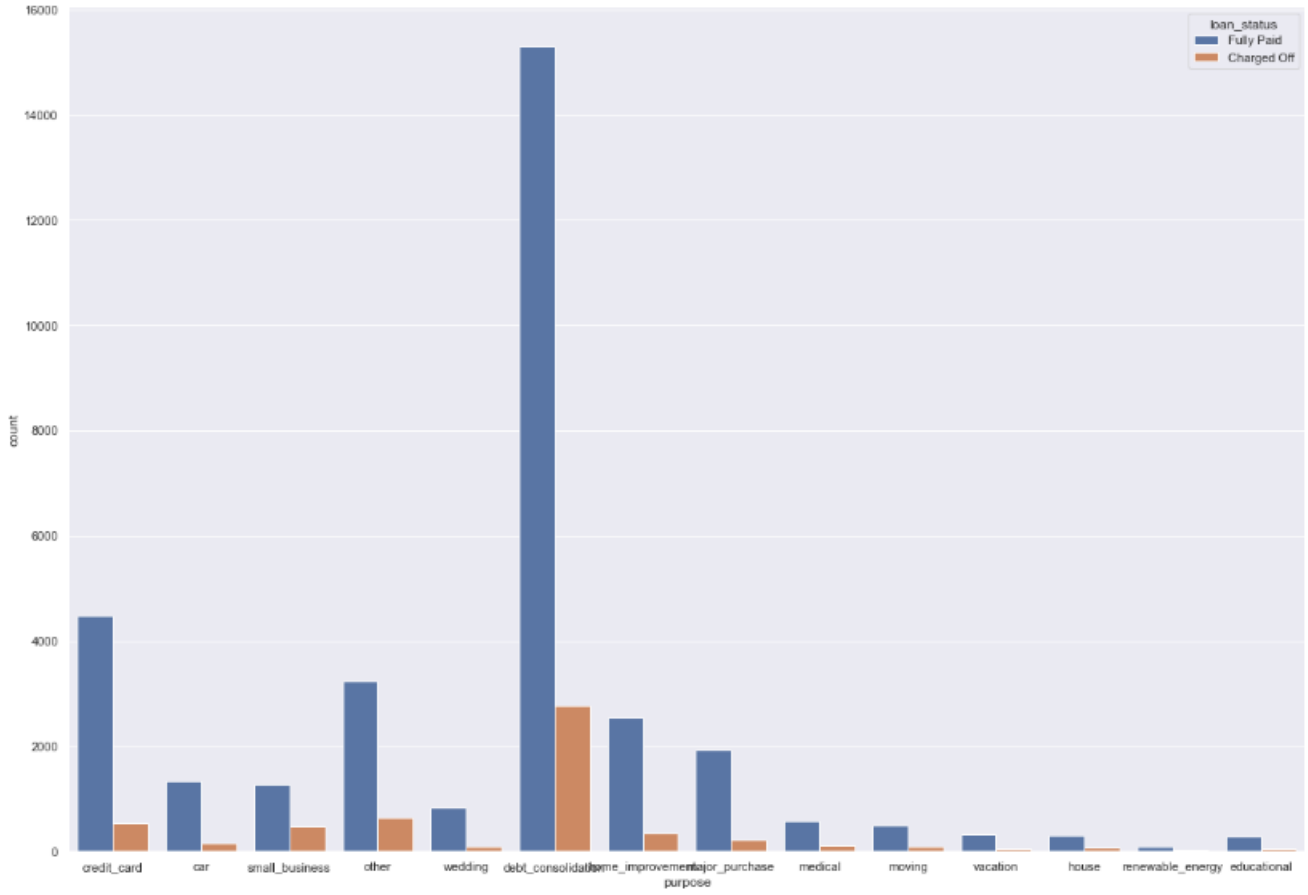

```
Charged Off percentage when monthly_inc_to_inst_ratio < 6 =  12.280875507576509

Charged Off percentage when monthly_inc_to_inst_ratio > 6 =  17.11907740847522
```

Higher the installment amount with respect to monthly salary, higher are the chances of default by customer

Bivariate analysis

- Purpose vs loan_status




```
Charged Off percentage for purpose  credit_card  =  10.781778396658046
Charged Off percentage for purpose  car  =  10.67378252168112
Charged Off percentage for purpose  small_business  =  27.08095781071836
Charged Off percentage for purpose  other  =  16.377749029754206
Charged Off percentage for purpose  wedding  =  10.367170626349893
Charged Off percentage for purpose  debt_consolidation  =  15.325394627527
Charged Off percentage for purpose  home_improvement  =  12.069565217391304
Charged Off percentage for purpose  major_purchase  =  10.325581395348838
Charged Off percentage for purpose  medical  =  15.565345080763581
Charged Off percentage for purpose  moving  =  15.972222222222221
Charged Off percentage for purpose  vacation  =  14.133333333333335
Charged Off percentage for purpose  house  =  16.076294277929154
Charged Off percentage for purpose  renewable_energy  =  18.627450980392158
Charged Off percentage for purpose  educational  =  17.23076923076923
```

The loan for the purpose of small business has a very high probability of getting into Charged Off state when compared to other purposes. The loan for the purpose of - credit_card, car, wedding, major_purchase are more likely to get fully paid

Recommendations

- Higher the proportion of installment to the monthly salary of a customer higher is probability of charged off
- Higher the loan amount higher is the probability higher is probability of charged off
- Loan for the purpose of 'small business' has a very high probability of charged off followed by 'debt consolidation' purpose and 'house' purpose
- loan issued during December has a high probability of charged off compared to Feb and Mar months
- Customers with Number of public record bankruptcies greater than 0, are more likely to get loan Charged Off
- Customer with no employment are more likely to get loan Charged Off
- Higher the dti, higher is the probability of loan getting Charged Off