Lending Club Case Study

Group Member:

1. *Bavani Ennares Maniammai A S (Group_Facilitator)*
2. *Palmari Sivasubramanian*

Lending Club - Business Understanding

- The Lending Club company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface. Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).

- The borrowers who default cause the largest amount of loss to the lenders. 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. Identification of such applicants using EDA is the aim of this case study. These risky loans can be reduced thereby cutting down the amount of credit loss

Problem Solving Methodology – Flow Diagram

Business Understanding

Determine the Lending Club objective clearly

Data Understanding

Define the goals of data analysis based on business understanding of Lending Club

Exploratory Data Analysis Finding Correlation (Heat Map)

Explore and verify data quality Collect Relevant data from loan file and clean the data

EPD Univariate /Segmented Analysis (Continuous and Categorical)

check the default rate across various categorical and continuous(Binning) features

EPD: Bivariate Analysis

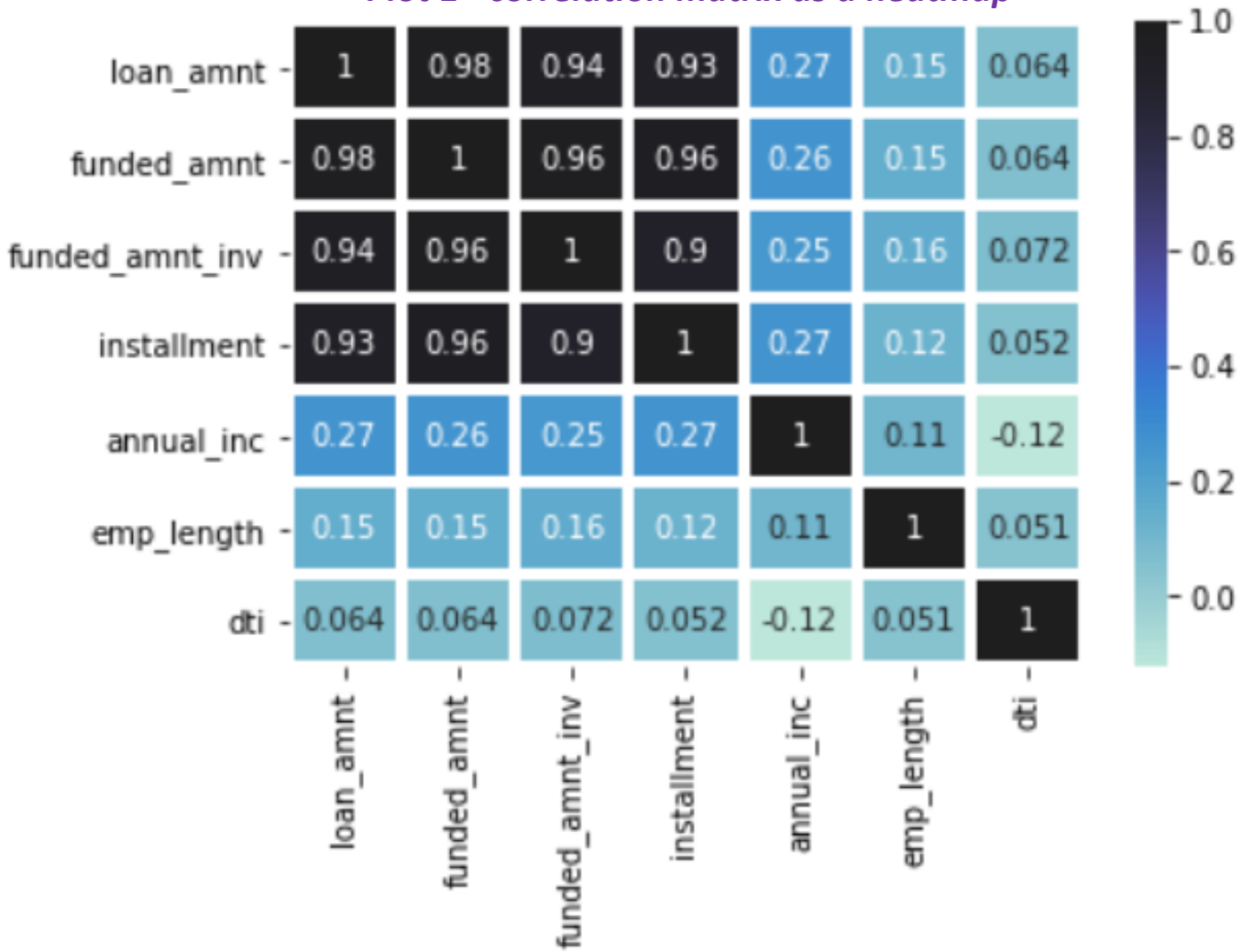
Check two or more variables to understand the default variable

Visualize Results and Recommendation Presentation

Simulate and Analyze the results and Suggest recommendation

Correlations Among columns

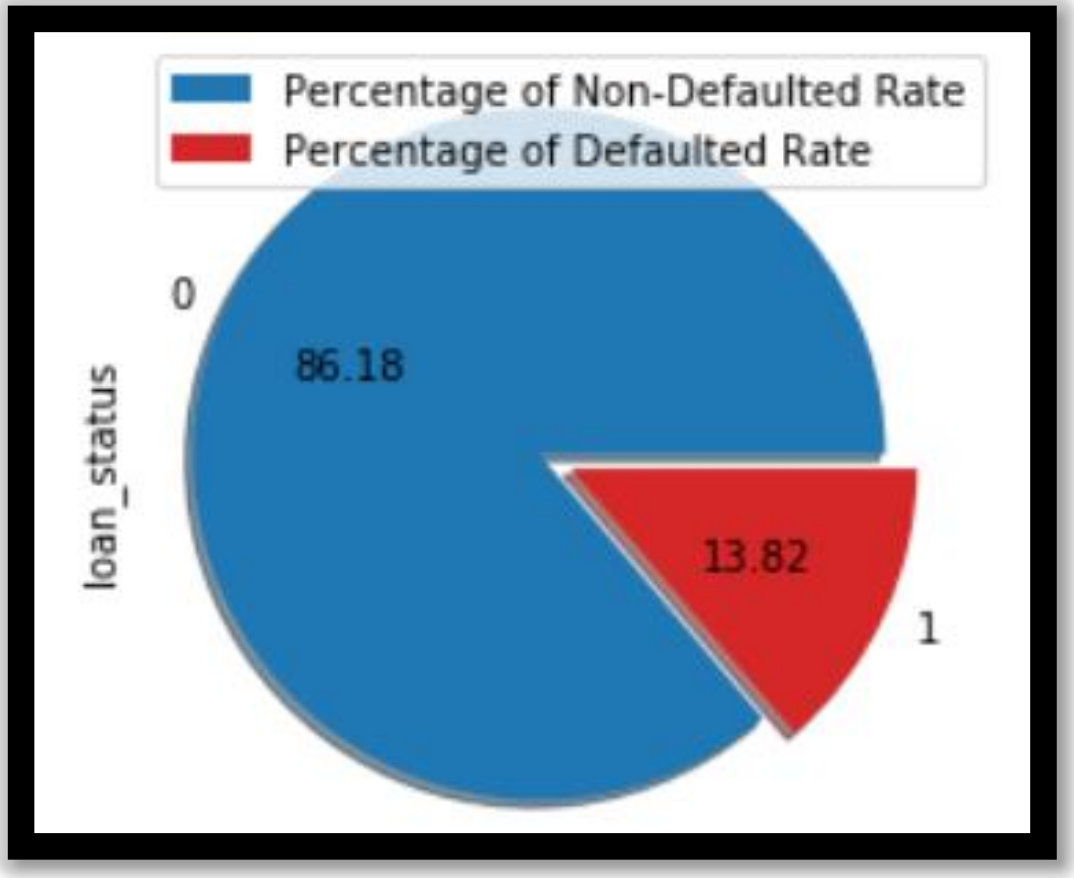
Plot 1 - correlation matrix as a heatmap



- Loan Amount is much correlated with funded amount, funded amount inv and Installment. The correlation is positive.
- annual_inc, emp_length and dti decreases as the loan_amount increases. The correlation in negative.

Target Key Column Analysis - Loan Status

Plot 2 - Overall Predictor of Default Percentage



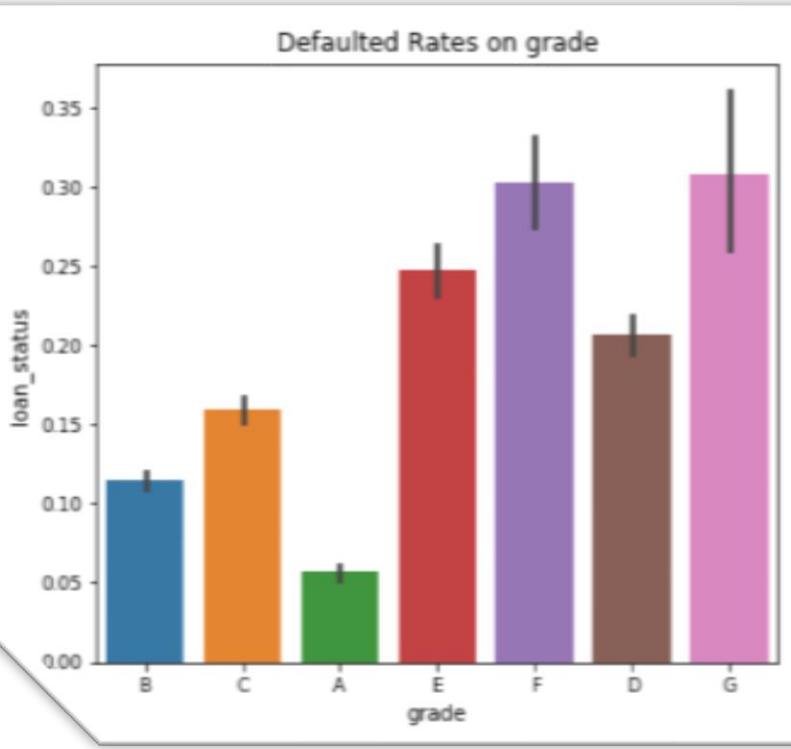
- The ones marked 'current' are neither fully paid int loan status column are not defaulted and the one marked 'fully paid' are completed their loan, tagging the two values as 0 and the one marked as 'Charged-Off' are defaulted so tagging them as 1 to make analysis more crisp.

- ***As we see in the pie chart the overall predictors of default rate is approximately 14 Percentage***

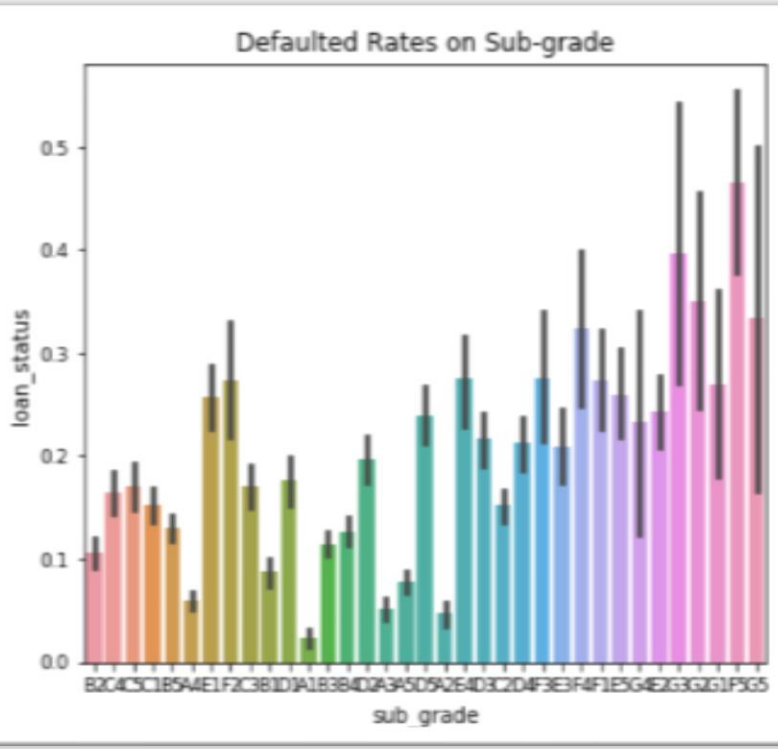
Univariate Analysis - Categorical Variable

Analysis - Default Rates on Grade, Sub-grade, home_ownership

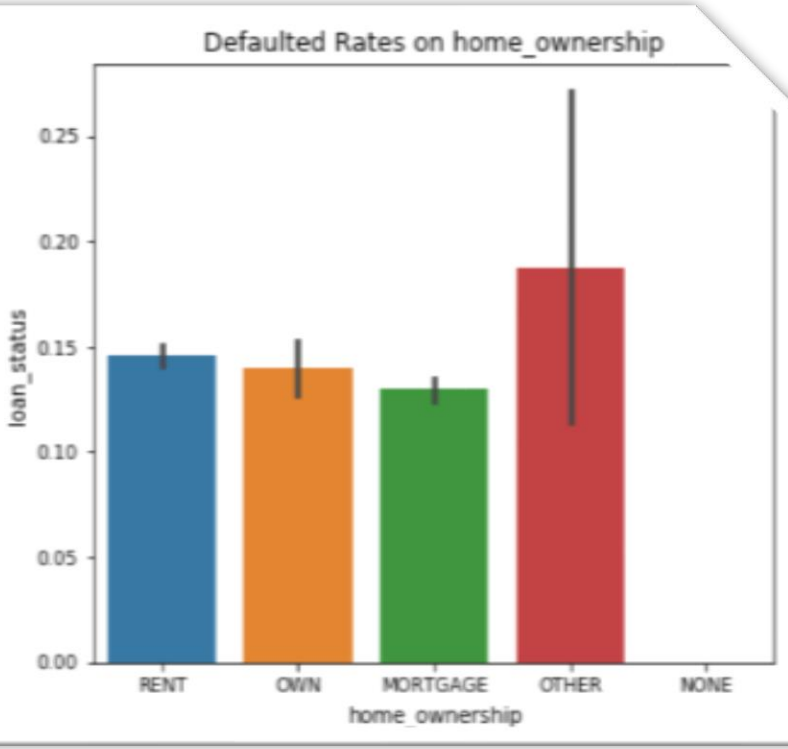
Plot 3



Plot 4



Plot 5

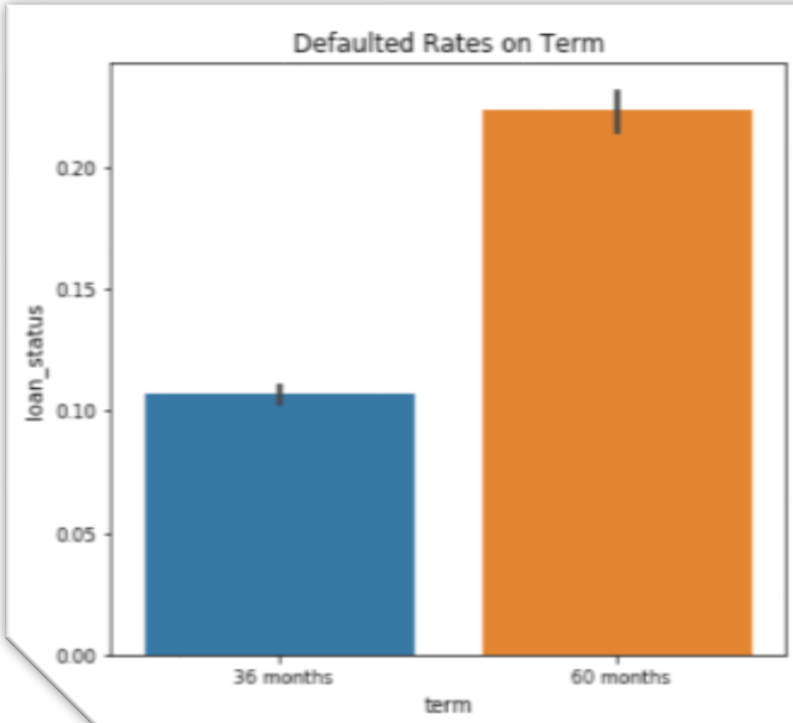


- Plot 3(Defaulted Rates on grade) - Most of the loans fall in upper grade 'G' second,third most are in grade 'F' and 'E' and rest all fall in grade 'D','C','B','A'
- Plot 4(Defaulted Rates on Sub-grade) - Most of the loans fall in upper Sub-grade 'F5'
- Plot 5(Defaulted Rates on home_ownership) - Most of the people have home ownership 'other'

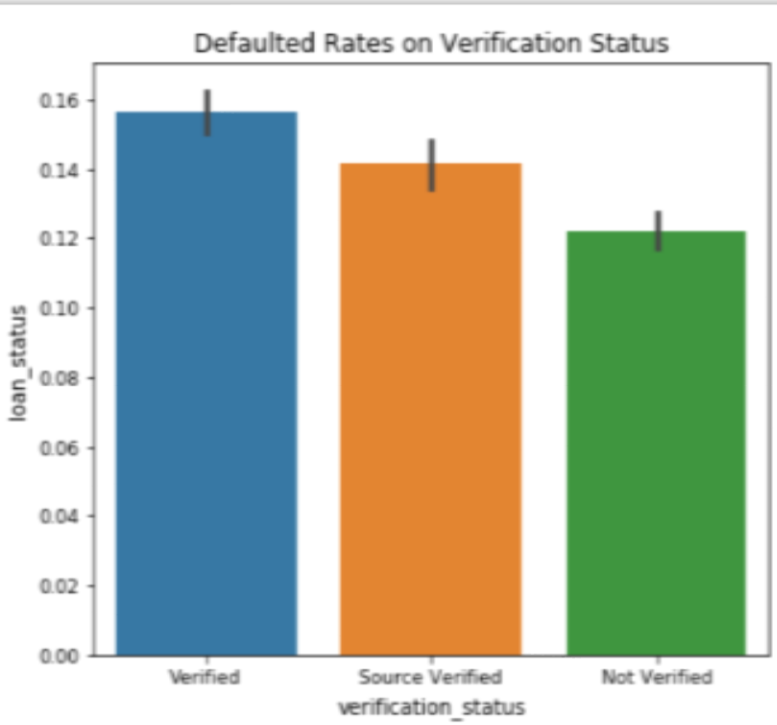
Univariate Analysis - Categorical Variable

Analysis - Default Rates on Term, Verification status, employee length

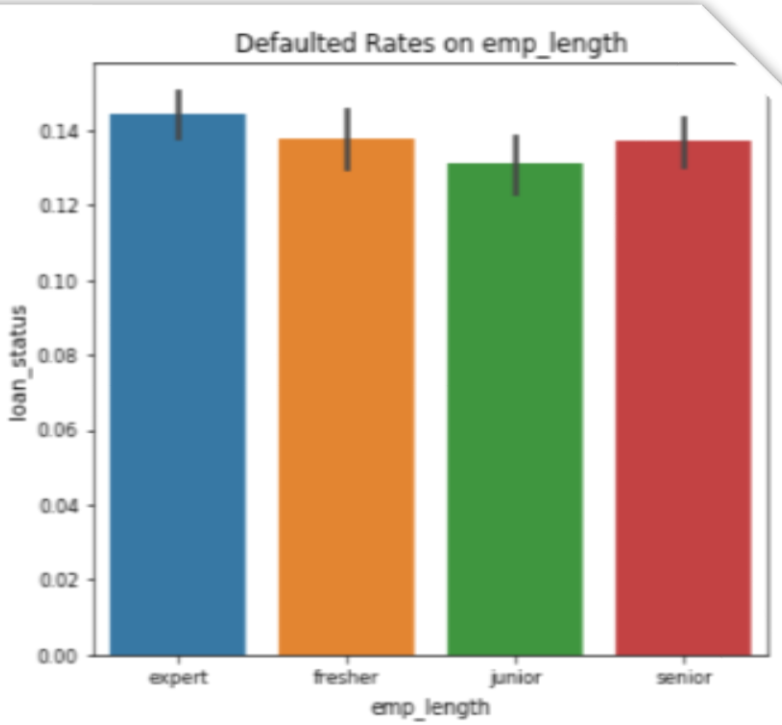
Plot 6



Plot 7



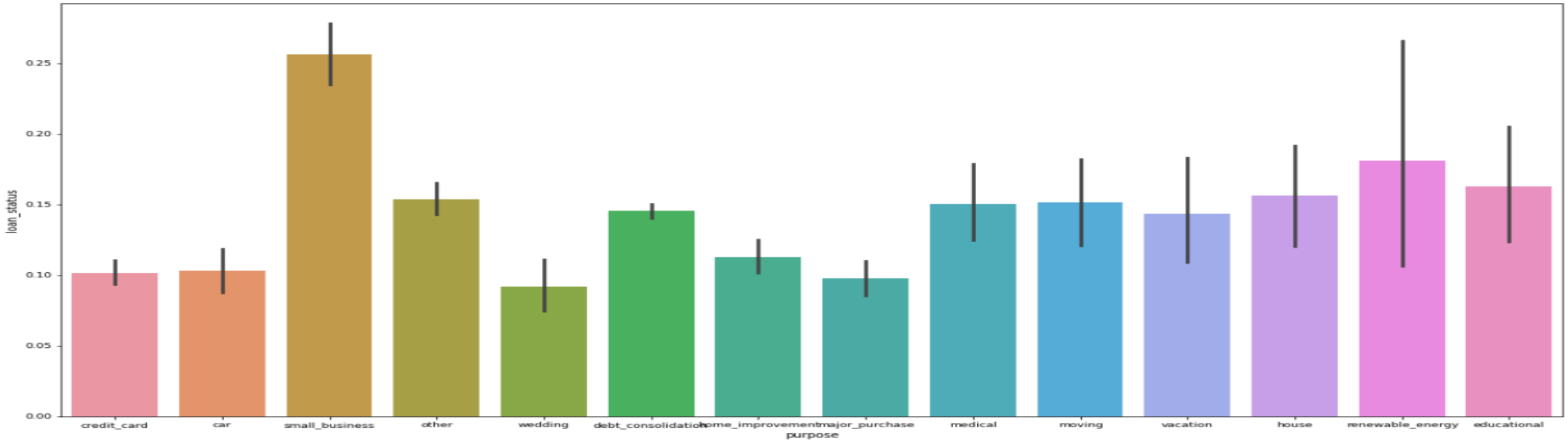
Plot 8



- Plot 6(Defaulted Rates on Term) - Most of the loans happened for 60 months term .Bank for a specific amount that has a specified repayment schedule
- Plot 7(Defaulted Rates on Verification Status) - Majority of the loans defaulted are verified
- Plot 8(Defaulted Rates on Employment length) - Almost similar default happened in Employment length in years. Most of the default happened for the employee having experience of 7, 10years

Univariate Analysis - Categorical Variable

Analysis - Default Rates on Purpose

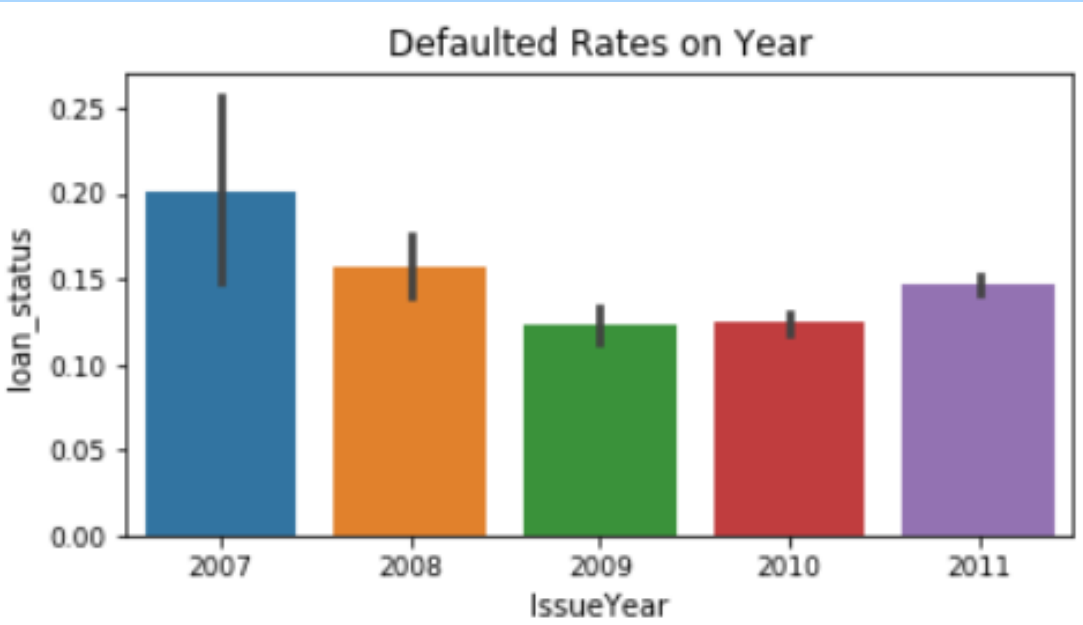


Plot 9 – Default Rates on Purpose

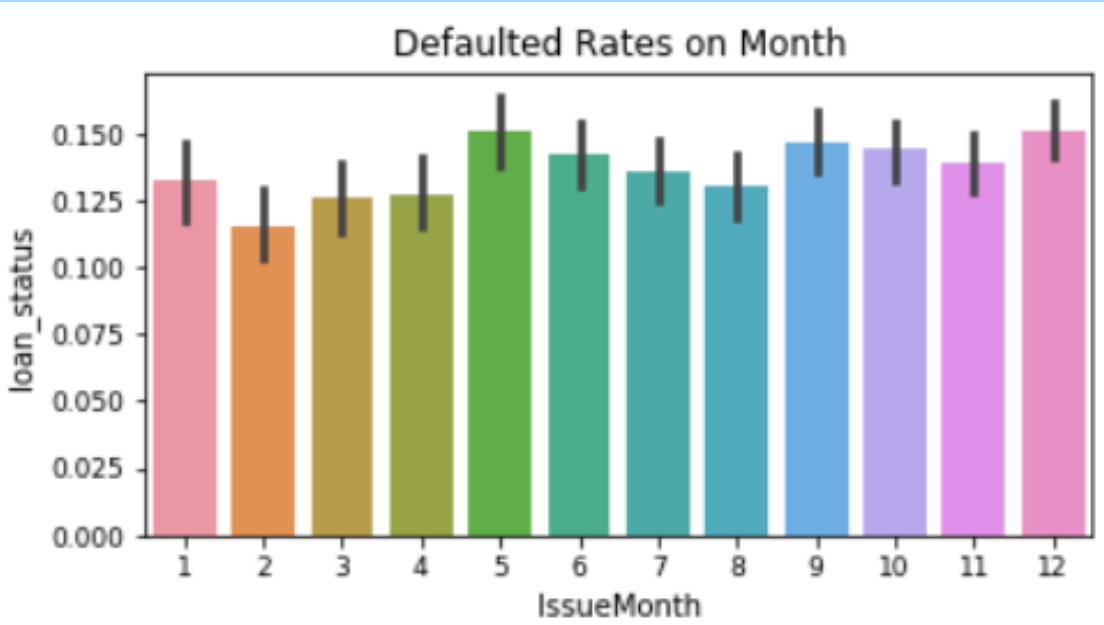
- Most of the loans taken for ***small business***. Borrowing money is expensive for a small company and raises its risk and they are not repaying. Suggesting lending club not to give loan for small business
- Second and third most loans taken for Renewable Energy and Educational

Analysis - Default Rates on Year & Month

Plot 10



Plot 11

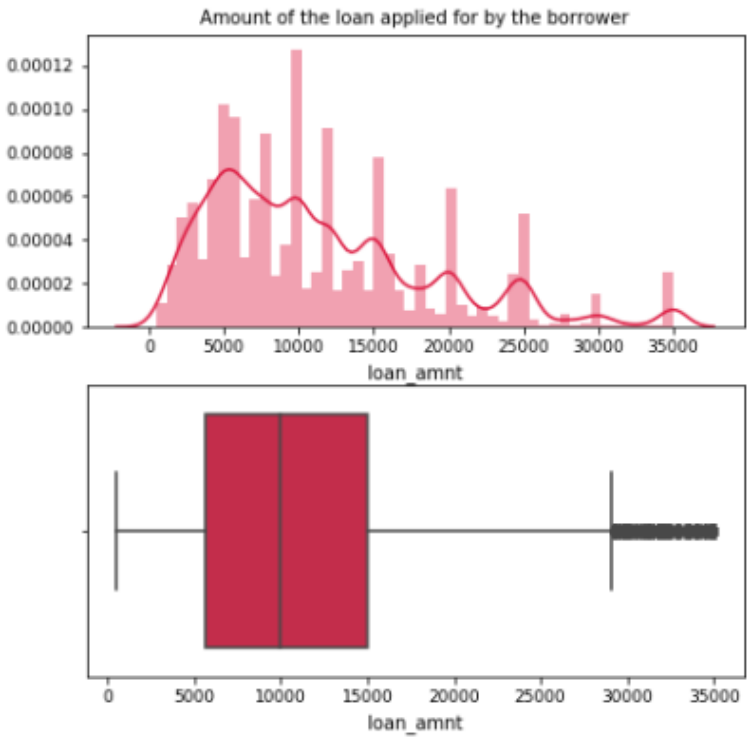


- Most of the loans taken in year 2007 and it got reduced during 2008,2009 and 2010
- Suddenly default rate started increased during 2011
- Most of the loan taken in May and second most December

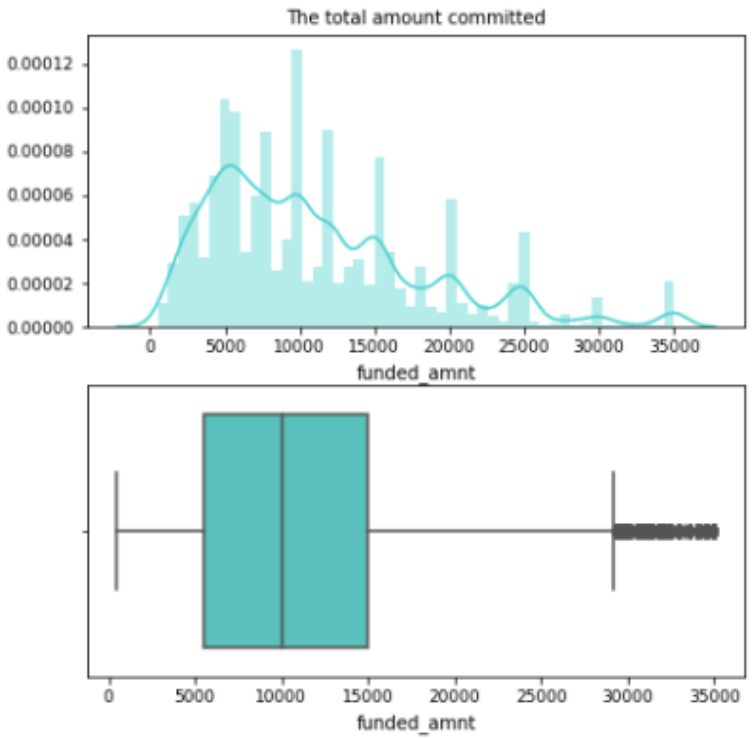
Univariate Analysis - Continuous Variable

Analysis – Loan Amount, Funded Amount, Funded Amount Inv

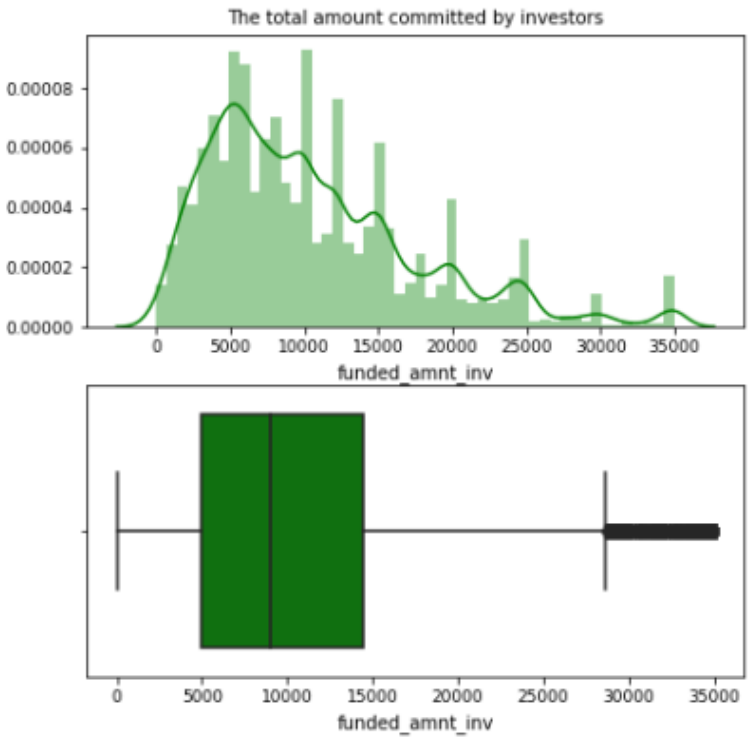
Plot 12



Plot 13



Plot 14

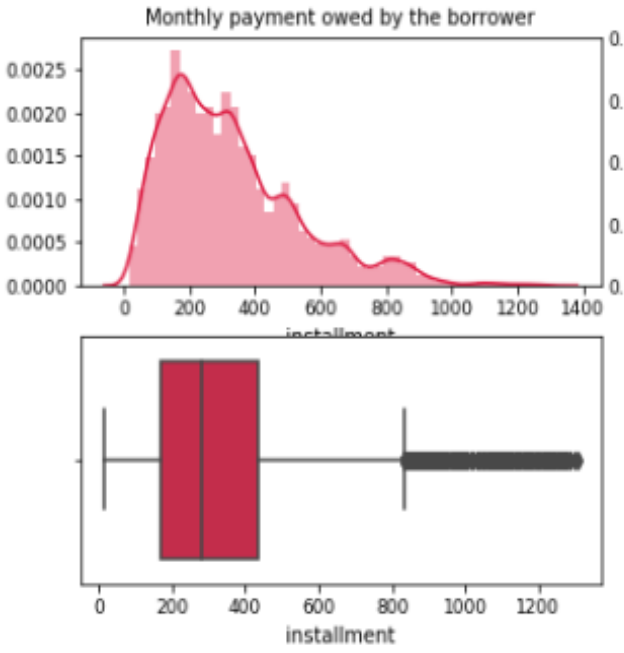


- The amount of the loan applied by the borrower, amount committed to that loan and the loan committed by investors are similarly distributed.
- As shown in plot Maximum amount of loan applied by the borrower were in between 5000 – 15000 USD
- Max amount asked by the borrower were in between 5000 – 15000 USD

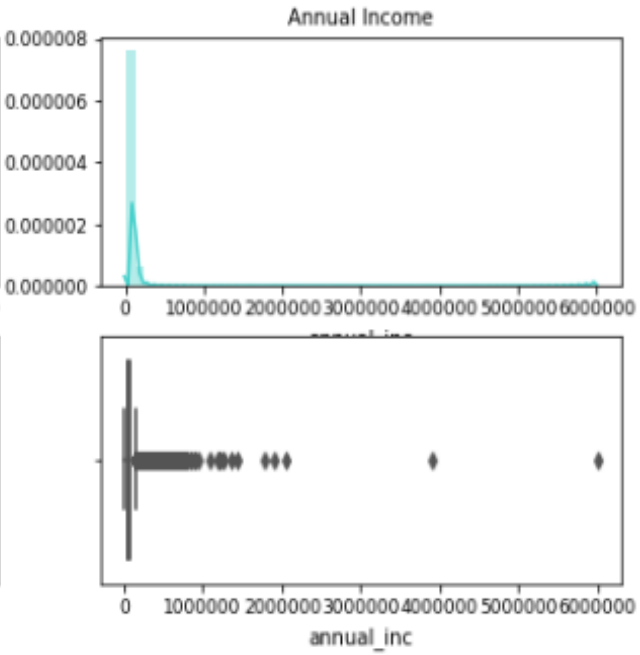
Univariate Analysis - Continuous Variable

Analysis – Installment, Annual Income, DTI, Employee length

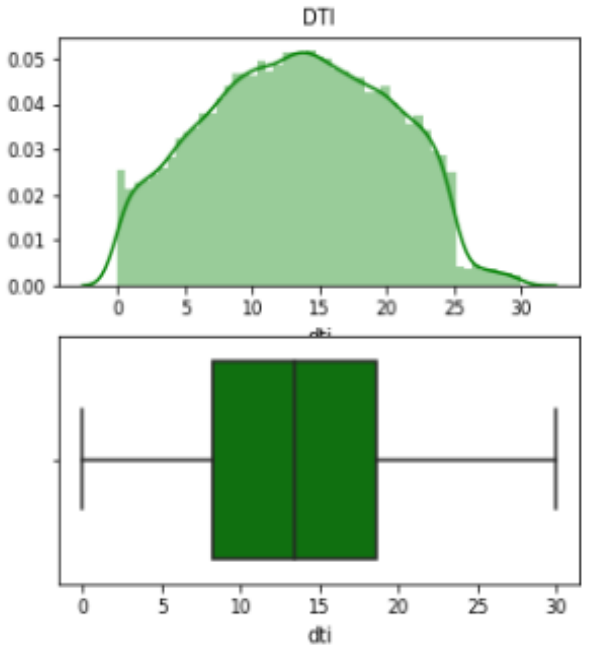
Plot 15



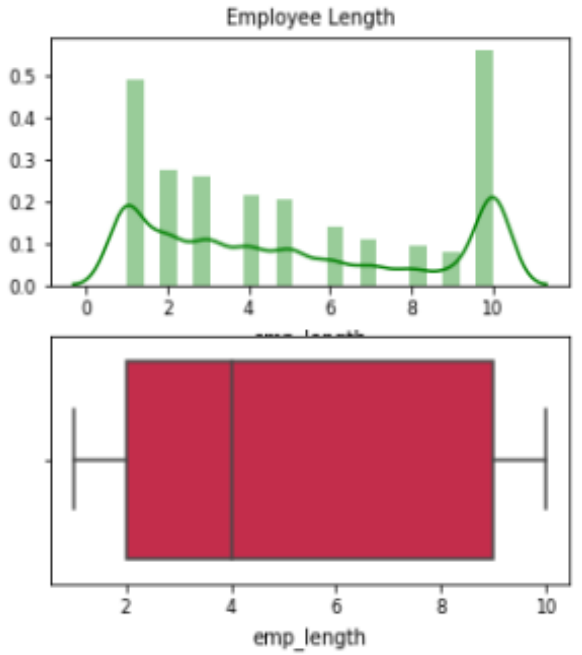
Plot 16



Plot 17



Plot 18

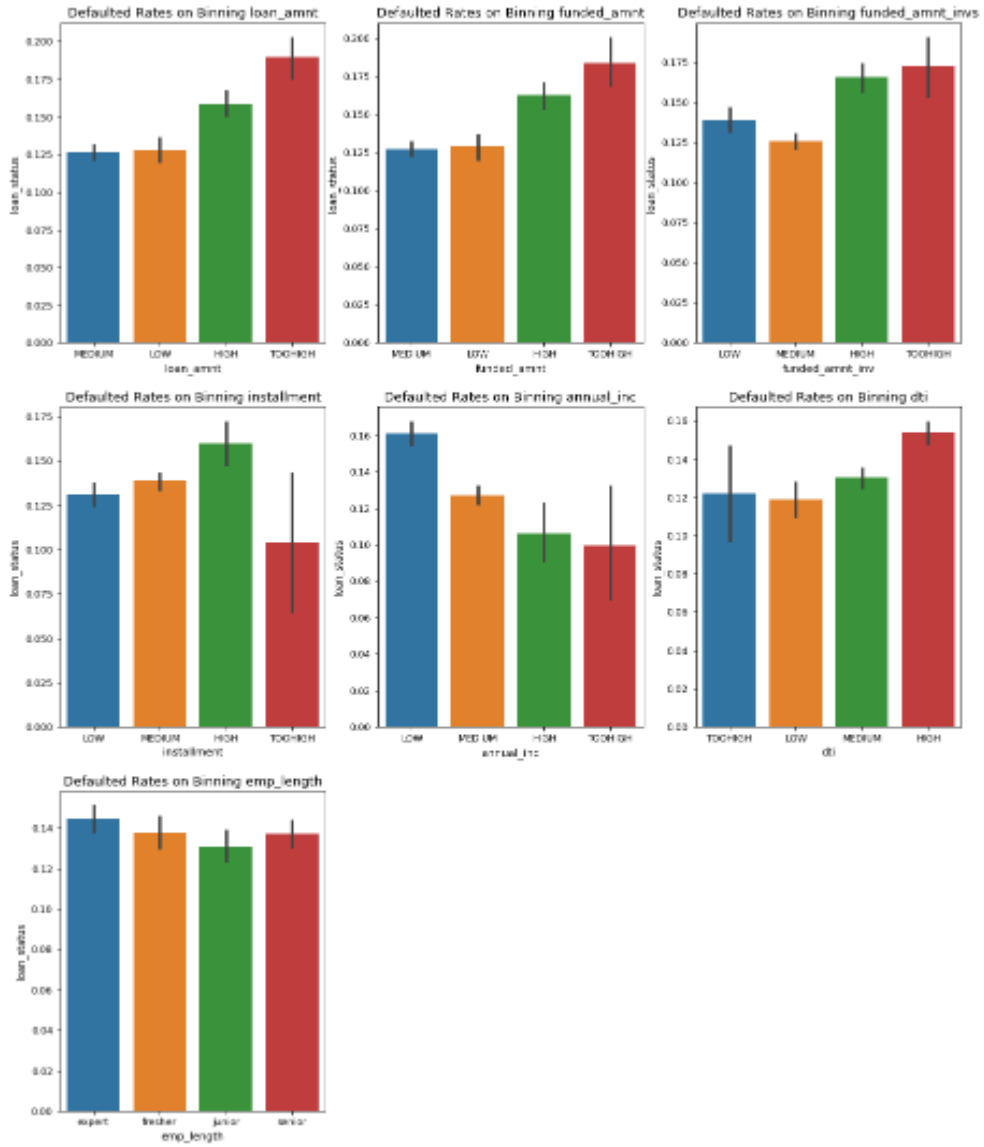


- Maximum monthly payment owed by the borrower if the loan originates is less than 200 and the median is 280.970000.We are not considering here the Maximum because of outliers as shown in box plots
- The median of self-reported annual income provided by the borrower during registration is 59000. We are not considering here the Maximum because of outliers as shown in box plots

Univariate Analysis - Continuous Variable

Analysis – Performed binning to make univariate Analysis

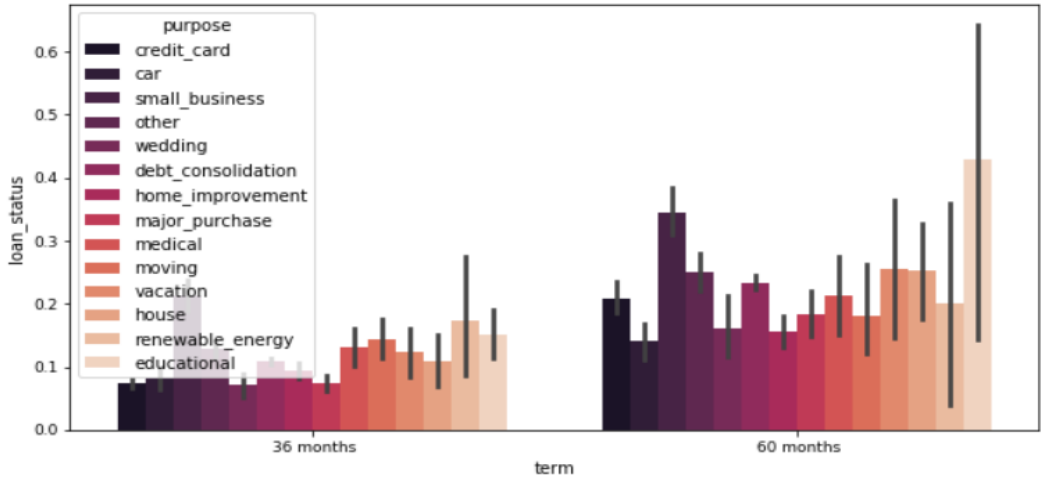
- Most of the defaults happened in Too high loan amount more than 25000 for all the three variables (loan_amnt,funded_amnt,funded_amnt_inv)
- Suggesting lending club to avoid giving money more than 25000
- Borrower whoever getting low income less than 50000 involved in more defaults
- Higher the dti range(15-25) more chances of loan being defaulted
- More Experienced involved in more defaults

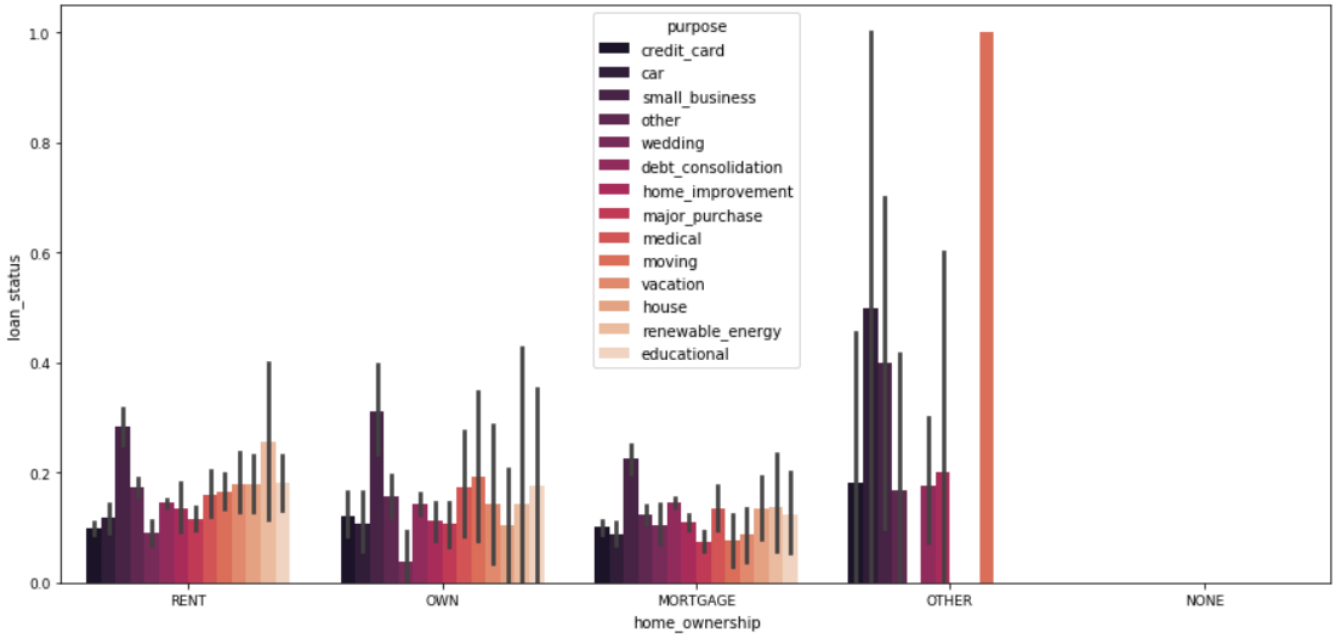


Data Analysis: Bivariate Analysis

Compared two variables to understand the purpose of defaults

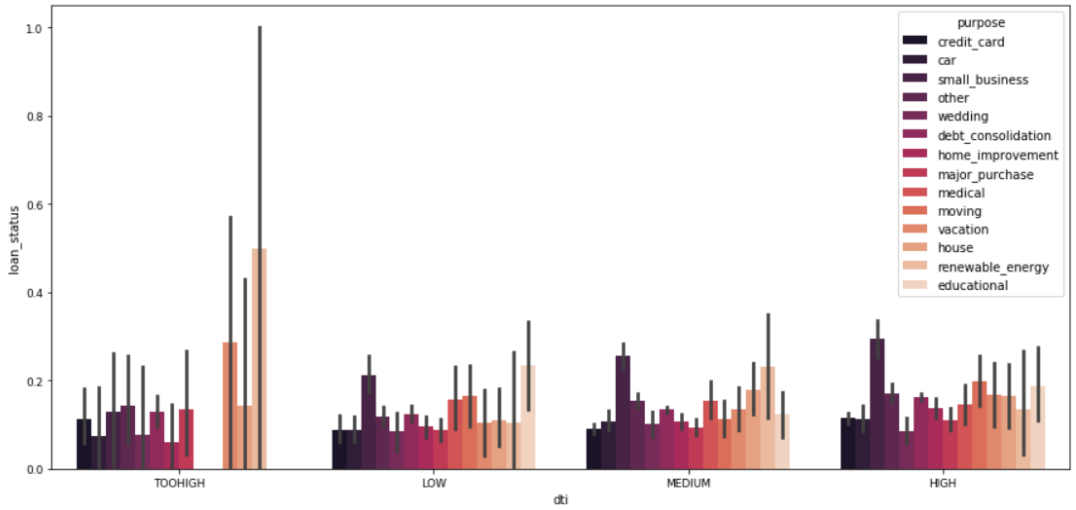
- The number of payments on the loan. Most of the defaults happened in term 60 month for educational purpose and second most defaults loan for small business
- Secondary, In term 36month most of the default happened for Small business purpose
- In both the term commonly the defaults loan happened for small business purpose
- Across all the home_ownership the defaults happened for small business and the second and third most are educational and renewable energy

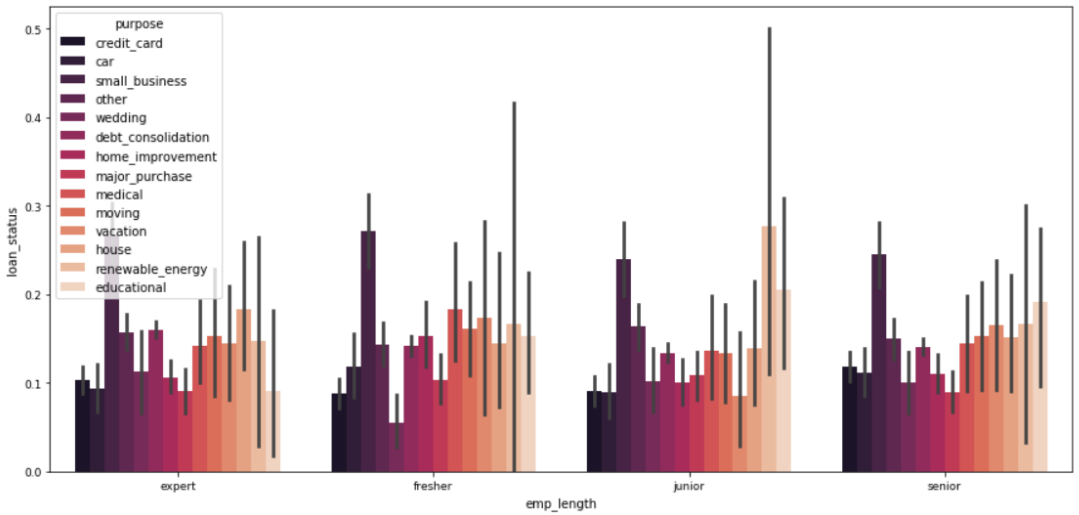


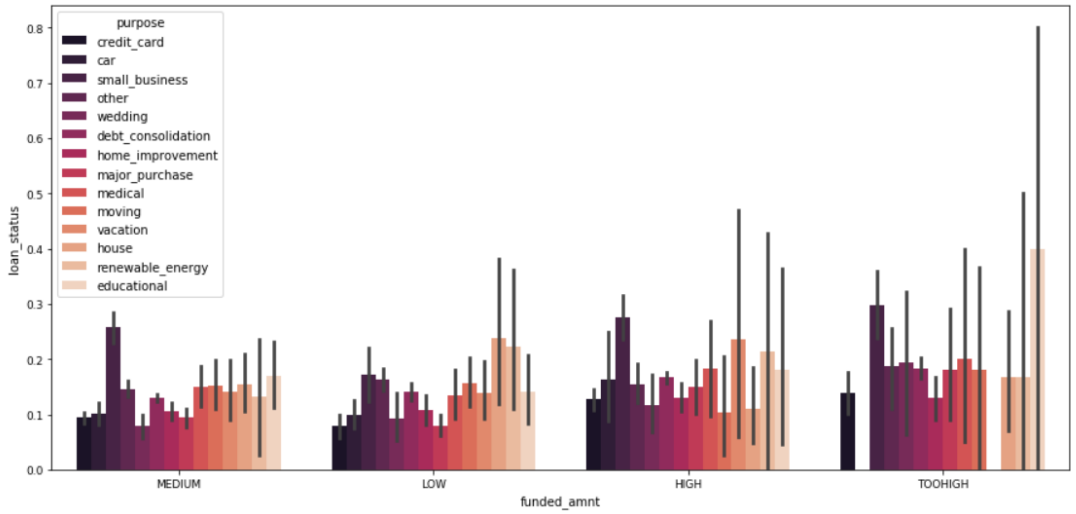


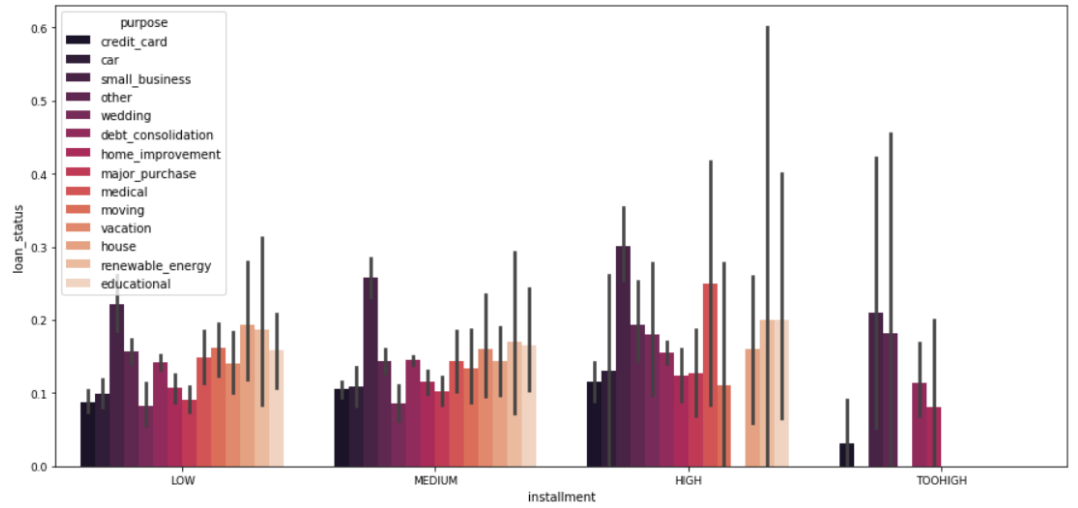
Data Analysis: Bivariate Analysis

Compared two variables to understand the purpose of defaults









- As seen in the above graph observing the purpose of applying loan is similar across all the variables
- Majorly amount of the loan applied by the borrower for small business and education

Conclusion/Recommendations

Conclusion

In Lending Club Case Study based on business objective, defined the goals, examine the large amount of data provided, and even plot statistical models in an attempt to explain what is observed during the entire process.
The overall default rate is 14%

Recommendations to the Investors

Below are the suggestion to investors to emphasize on how they can reduce the chances of finding a likely defaulter.

1. **Suggesting Lending Club to stop approving loans for small business. Borrowing money is expensive for a small company and raises its risk and they are not repaying**
2. **We can reduce the number of loan approvals for Educational and renewable energy**
3. **Stop approving the loans if the annual income is less than 50000 USD**
4. **Reduce or stop approving the loan amount greater than 25000 USD**
5. **Reduce giving loan with people having more employee experience**