Lending Club Case Study

ABSTRACT

Consumer finance 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. 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 utilise this knowledge for its portfolio and risk assessment.

Palla Mohana Silpa

Jonnadula Venkateswarlu

Process to approach the solution

Data Understanding

- Understand the data better by researching the columns andtheir descriptions.
- Do little research to understand the objective better.

Data Cleaning

- Fix missing values, repeated rows, spelling inconsistencies .These issues could make it difficult to analyse data and could lead to errors or irrelevant results.
- Thus, these issues need to be corrected before data is analysed.

Univariate Analysis

- Analysing variables one at a time
- Identifying outliers and dealing with them to get cleaneddata

Bivariate Analysis

- Analysing two variables at a time
- Undestanding the relation between two variables

Derived Metrics

- Creating new variables using the existing ones to get meaningful insights.

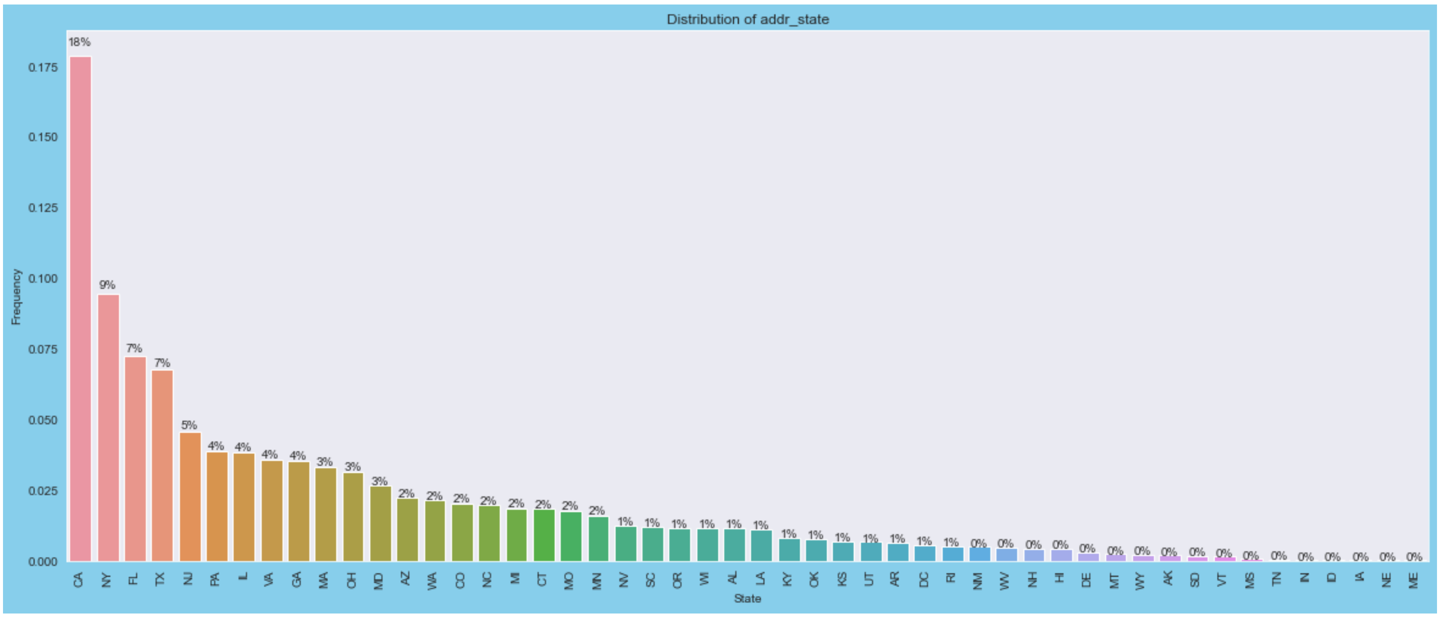
Data Visualization

- Analysing all the variables by plotting and define the relations between them.
- Plot correlation matrix to gain better analysis on the data.

Data Loading & Data Cleaning:

- Drop columns which contains complete null values
- Drop Rows which contains complete null values
- Missing Value check in the remaining columns
- Impute missing values
- Identify the datatypes and correct them
- Extracting the date columns
- Derive new variables which can be used for better analysis

Loan distribution with respect to addr_state

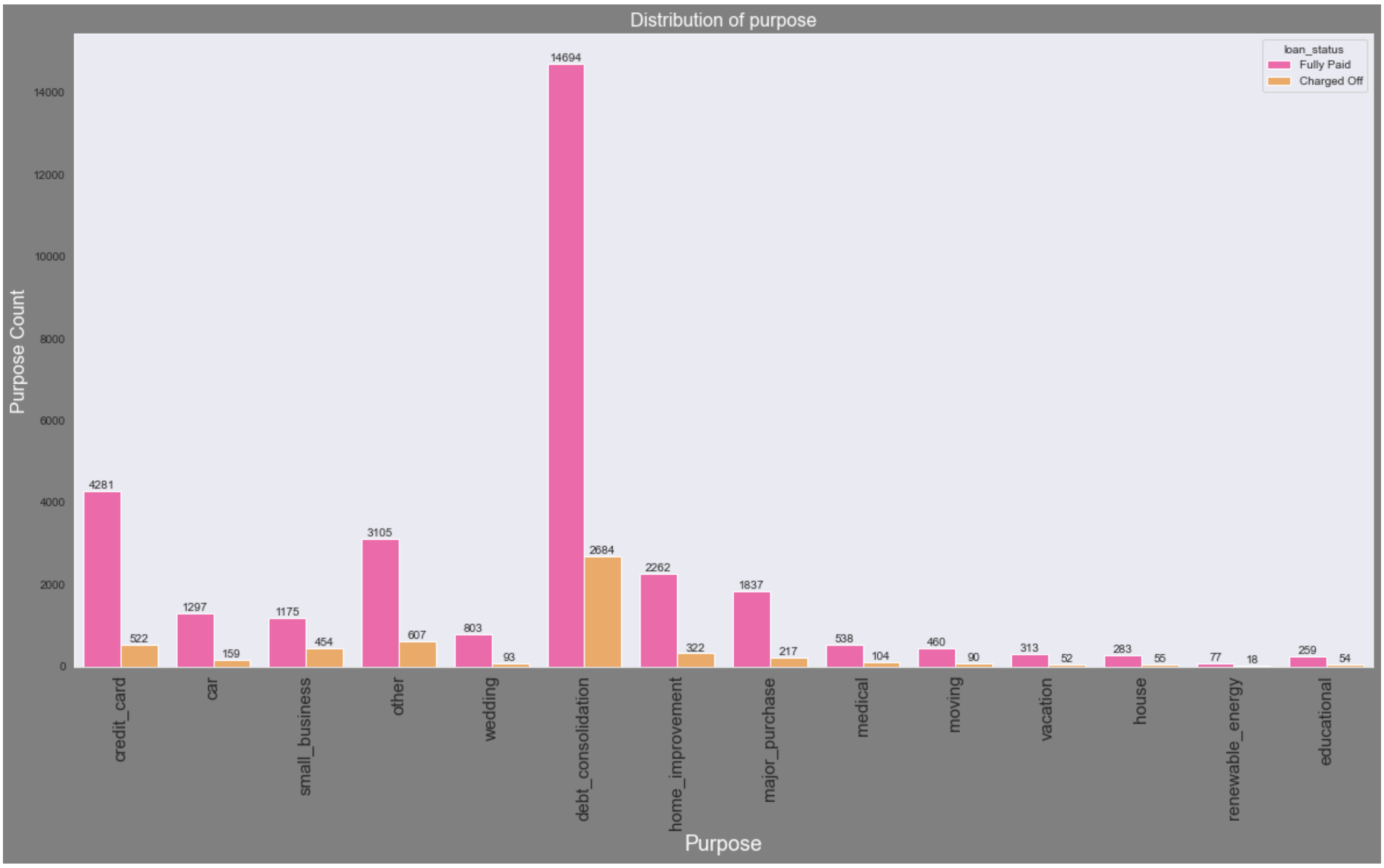


Loans approval distribution is mostly among these states ['CA', 'NY', 'FL', 'TX']

Conclusion:

Company should carefully analyse the address states of the people as we can see in our dataset majority of the defaults are from CA, NY, FL.

Purpose of loan distribution across loan_status

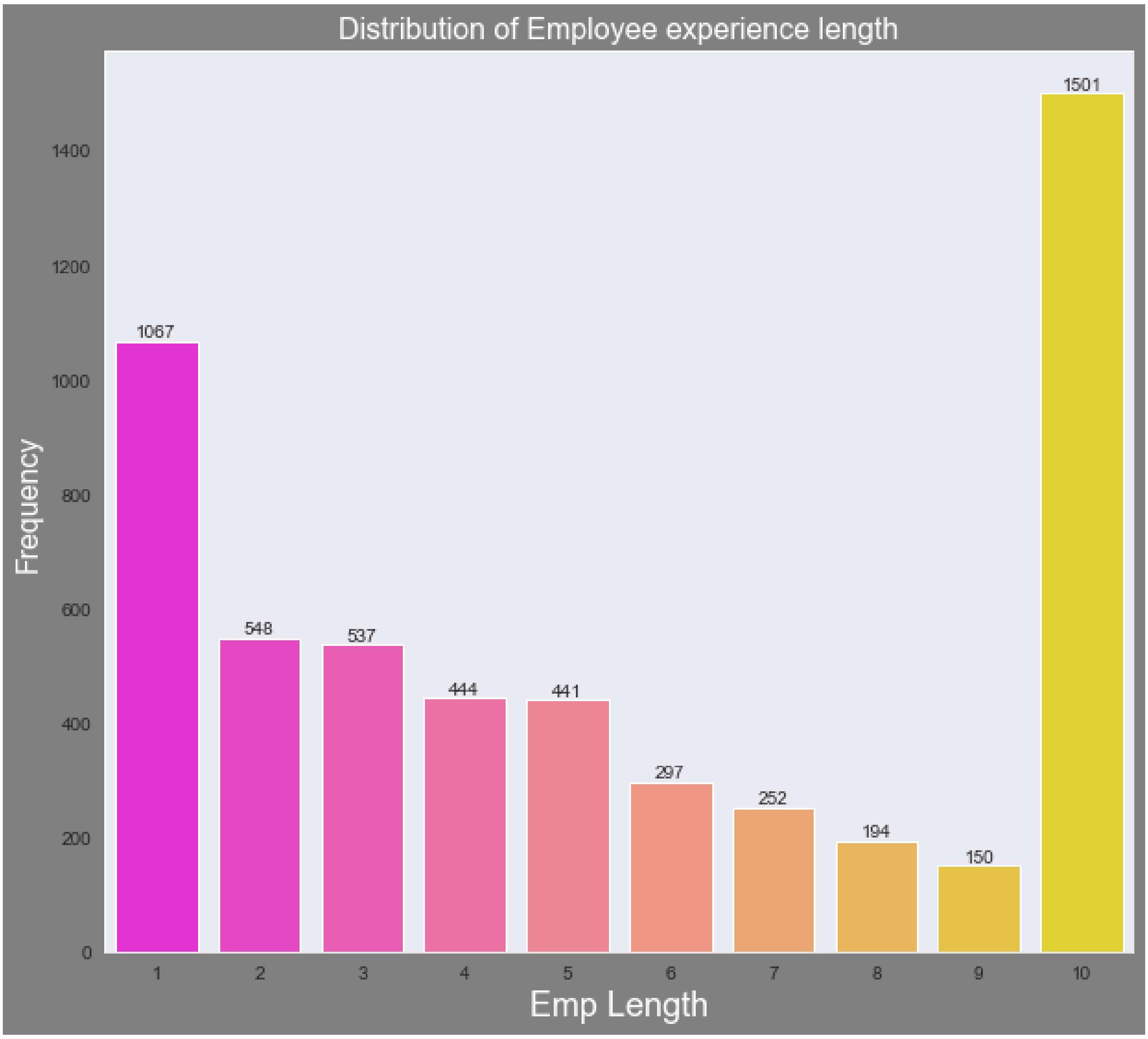


From the chart, it is clear that Employees with 1 and >10 years of working experience are in more defaulters list.

Conclusion:

Customers who took loan with Purpose as DEBT CONSOLIDATION are more likely to get default

Purpose of loan distribution across loan_status

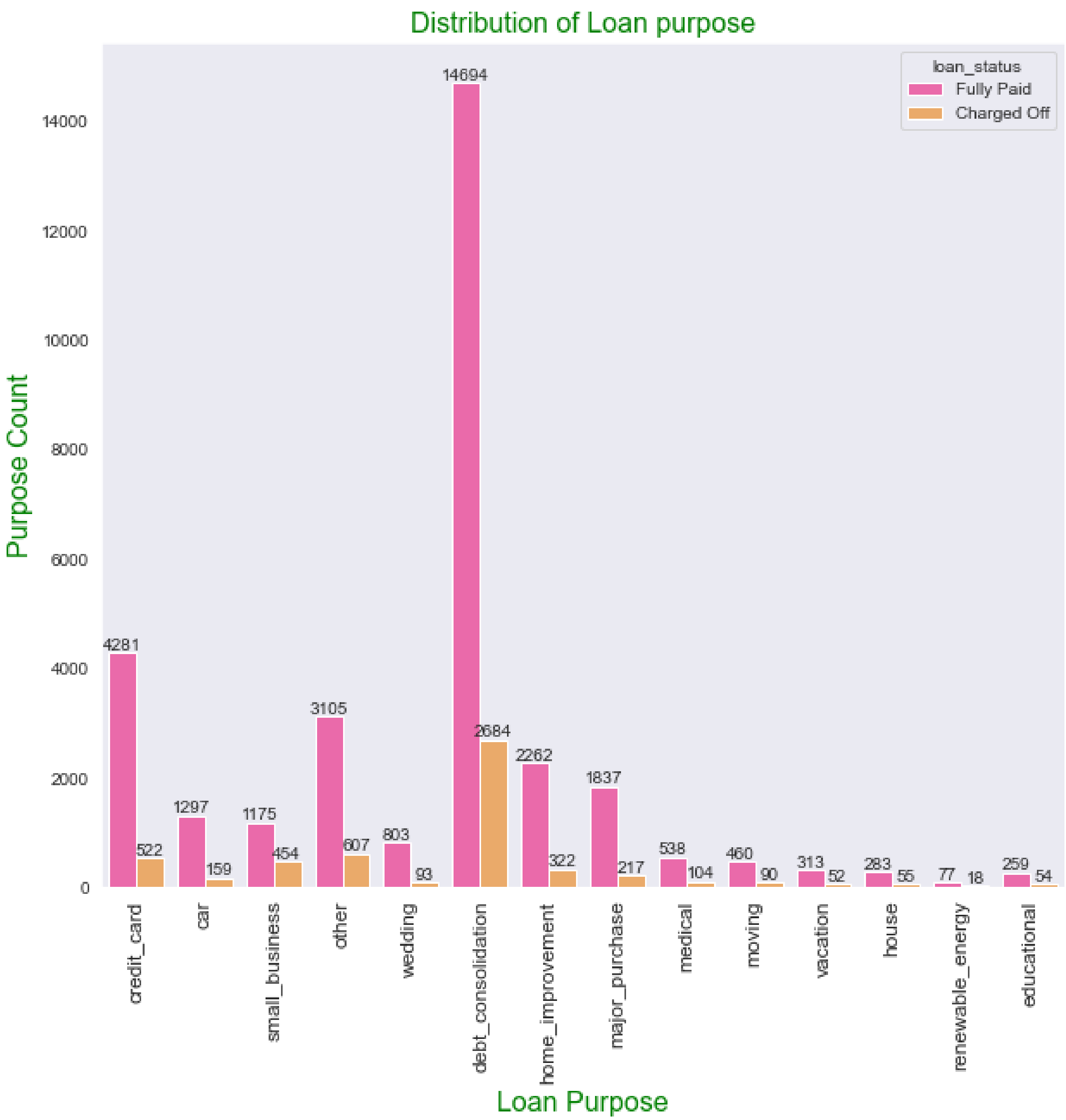


From the above graph Employees with 1 year and 10 and above years have most probable for default, because charged off loan count is high.

Conclusion:

Customers with working experience less than 2 Years and More then 8 years are more likely to get default

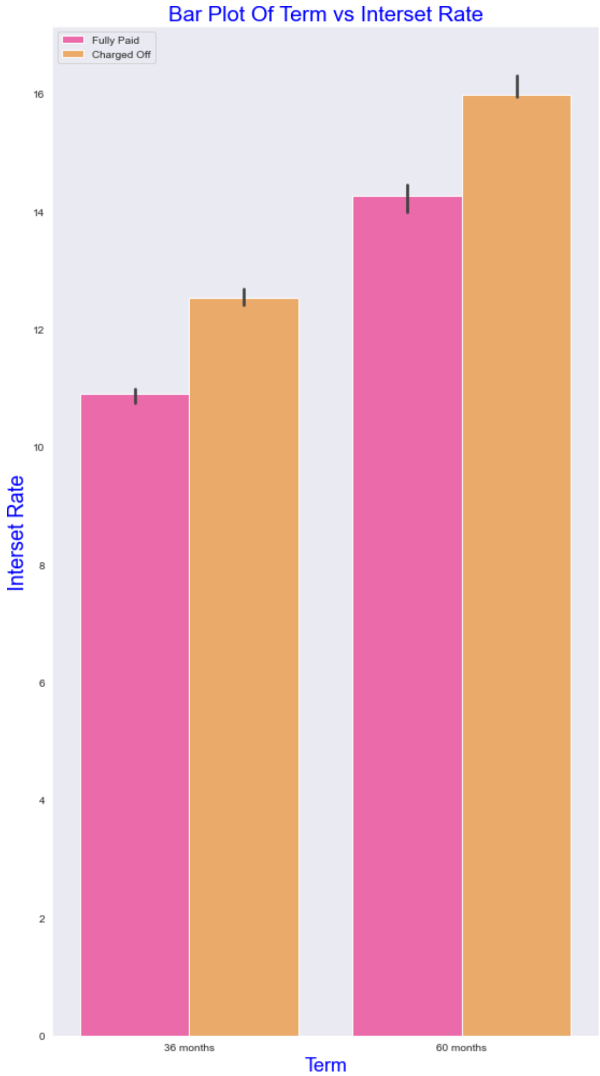
Bar Plot Of Loan Purpose with Loan_status

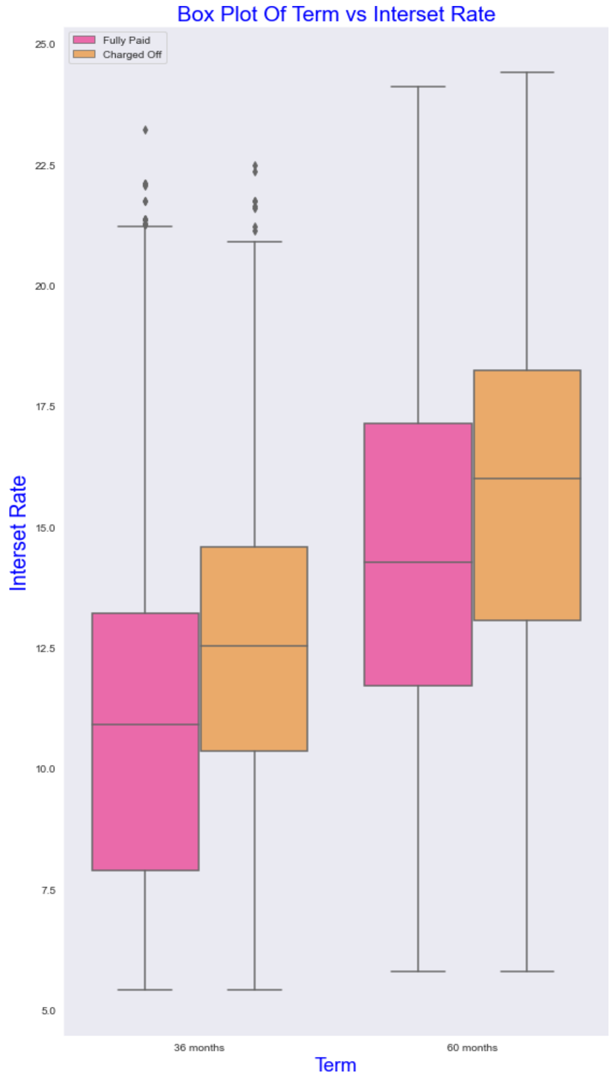


Conclusion:

Customers who took loan with Purpose as DEBT CONSOLIDATION are more likely to get default. Consumer finance company should focus more on DEBT_CONSOLIDATION loan purpose while issuing loans.

Loan Term, Interest Rate variation with Loan Status





Conclusion:

Looks like there are more proportion of borrowers defaulted loan in 60 months term than 36 months. The interest rate is high for 60 months term duration loans which leads users to default the loans. This means the people with loan term of 60 months are more likely to default.

Segmented Univariate Analysis

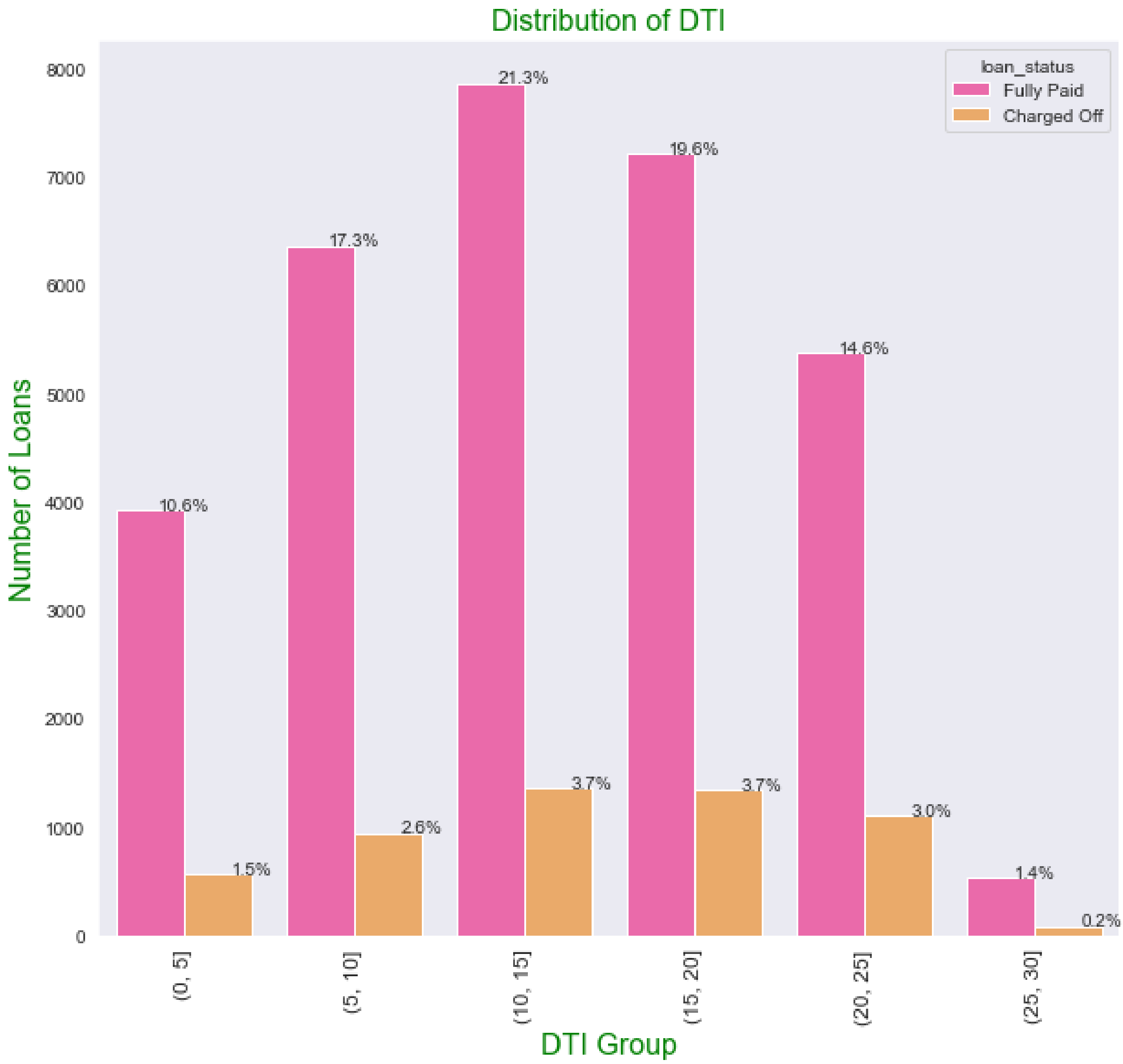
Distribution of Employee Tenure with Loan Status:

From the above graph it is very clear the Charged off loans are high with employees experience less than 2 Years and More then 8 years

Conclusion:
Customers with working experience less than 2 Years and More then 8 years are more likely to get default

Segmented Univariate Analysis

Distribution of DEBT TO INCOME RATIO with Loan Status:



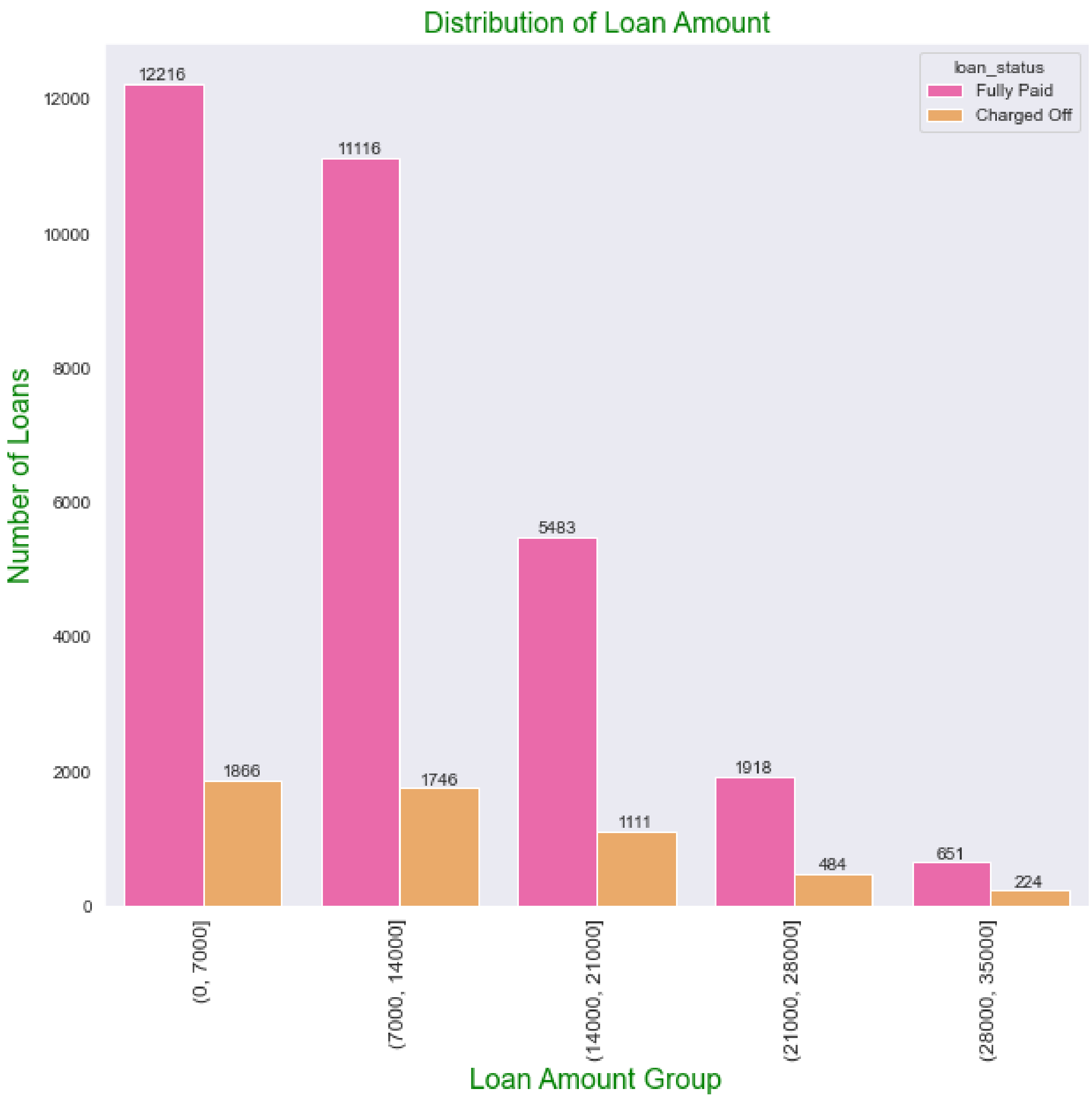
From the chart it is clear that loans with DEBT TO INCOME RATIO within 5 to 25 are more likely to get default

Conclusion:

Customers with DEBT TO INCOME from 5 to 25 are more likely to get default.

Segmented Univariate Analysis

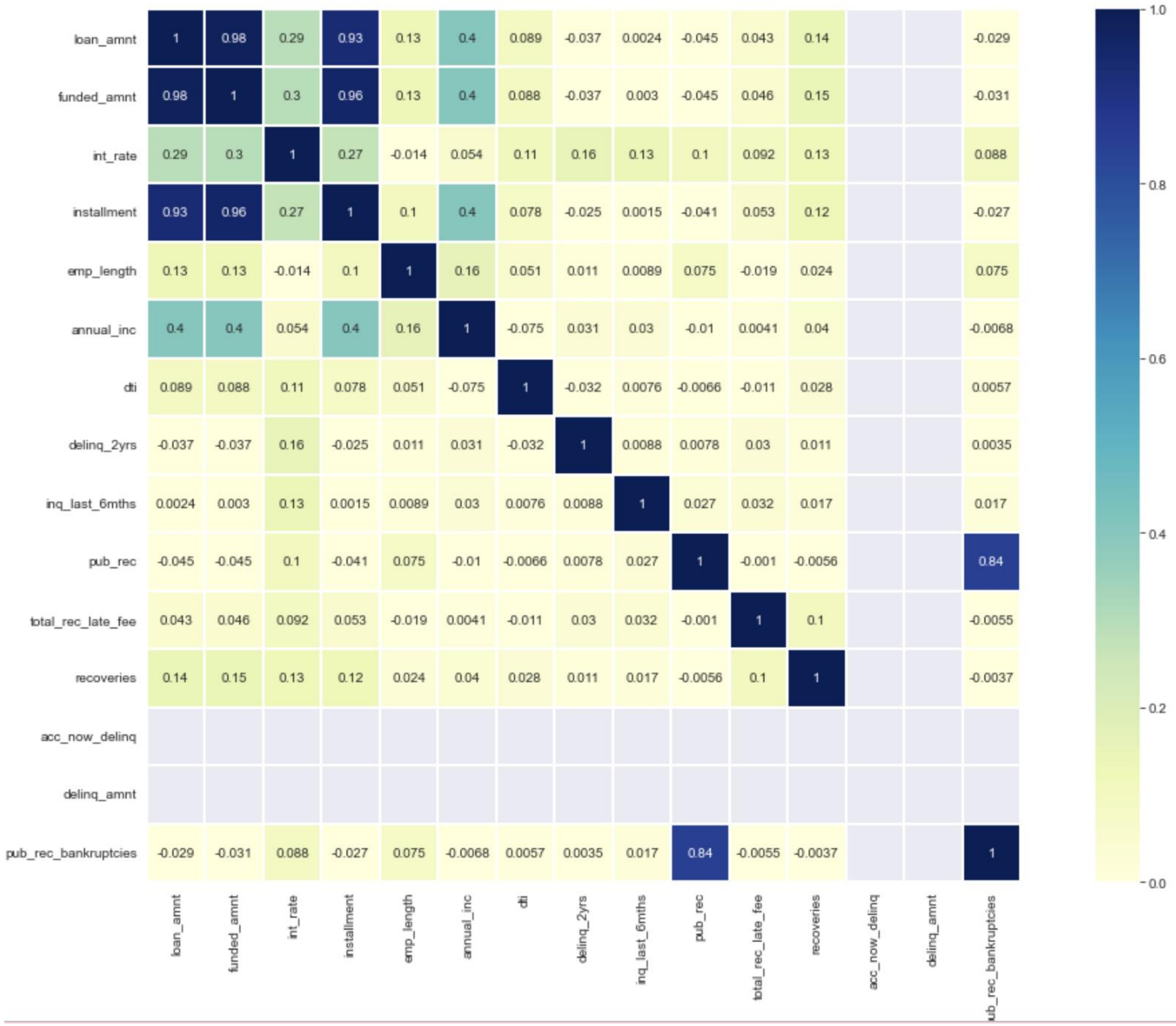
Distribution of Loan Amount Group with Loan Status



Conclusion:

The loans with high interest rate has high percentage of defaults, so high interest rate loans are more likely to get default

Correlation Matrix:



Result from above plot:

1. total_rec_late_fee is negatively correlated with DEBT TO INCOME RATIO and emp_length.

2. Number of derogatory public records (pub_rec) are negatively correlated with instalments.

3. int_rate is negatively correlated with emp_length.

Conclusions:

- Customers who took loan with Purpose as DEBT CONSOLIDATION are more likely to get default.
- Customers with working experience less than 2 Years and More then 8 years are more likely to get default
- The loans with high interest rate has high percentage of defaults, so high interest rate loans are more likely to get default
- Customers who took loan with DEBT TO INCOME RATIO lies in between 5 to 25 are more likely to get default
- Company should carefully analyse the address states of the people as we can see in our dataset majority of the defaults are from CA, NY, FL.
- Consumer finance company should reduce the high interest loans for 60 months tenure because they are prone to loan default.