Lending Club Case Study

Group Members:

Amrita Bais

Archana Kumari

Lending Club Case Study Objectives

Once the loan application is received ,the main goal is Identifying the Loan applicants traits tend to default to payback the loan

Two **types of risks** are associated with the bank's decision:

If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Understanding 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.

Problem Solving approach

Data Understanding /Exploration (Importing the Important Libraries & Data set Overview)

Data cleaning and Manipulation(missing value imputation, outlier treatment and other kinds of data redundancies ,Standardization of Data

## Data Analysis(Deriving the new variables, Univariate Analysis, Bivariate Analysis, Plotting the relevant Graphs)

Observations and inferences based on the Analysis and graphs

There are three types of variables:
1.Customer Demographic Variables
2.Loan characteristic's Variables
3. Customer Behavior Variables
We will focus only on the Loan characteristics Variables and Customer Behavior Variables.

There were columns which were having 100% null values so dropped those columns. Also filtered the null values from the emp_length as the % of null values were very less. In general for treating the missing or null values , we can impute the values based on the type of variables like for numeric we can impute using mean/median and for the categorial we can use mode.

There are 39717 rows and 111 columns in loan dataset.
There are no duplicate columns and there are 54 empty columns.

Loan Status Analysis
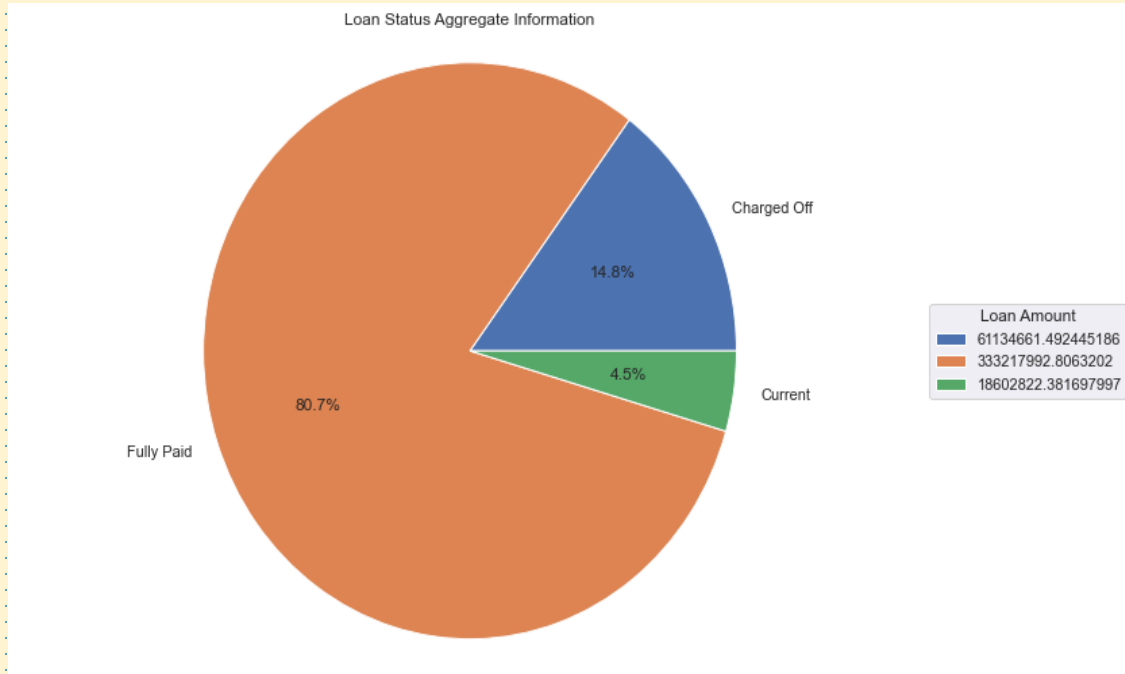
Loan Status is the Target variable in the analysis. As the Current status is not useful in the analysis so it is removed and then analysis continued with Fully-paid and Charged-off.

Employee experience Analysis
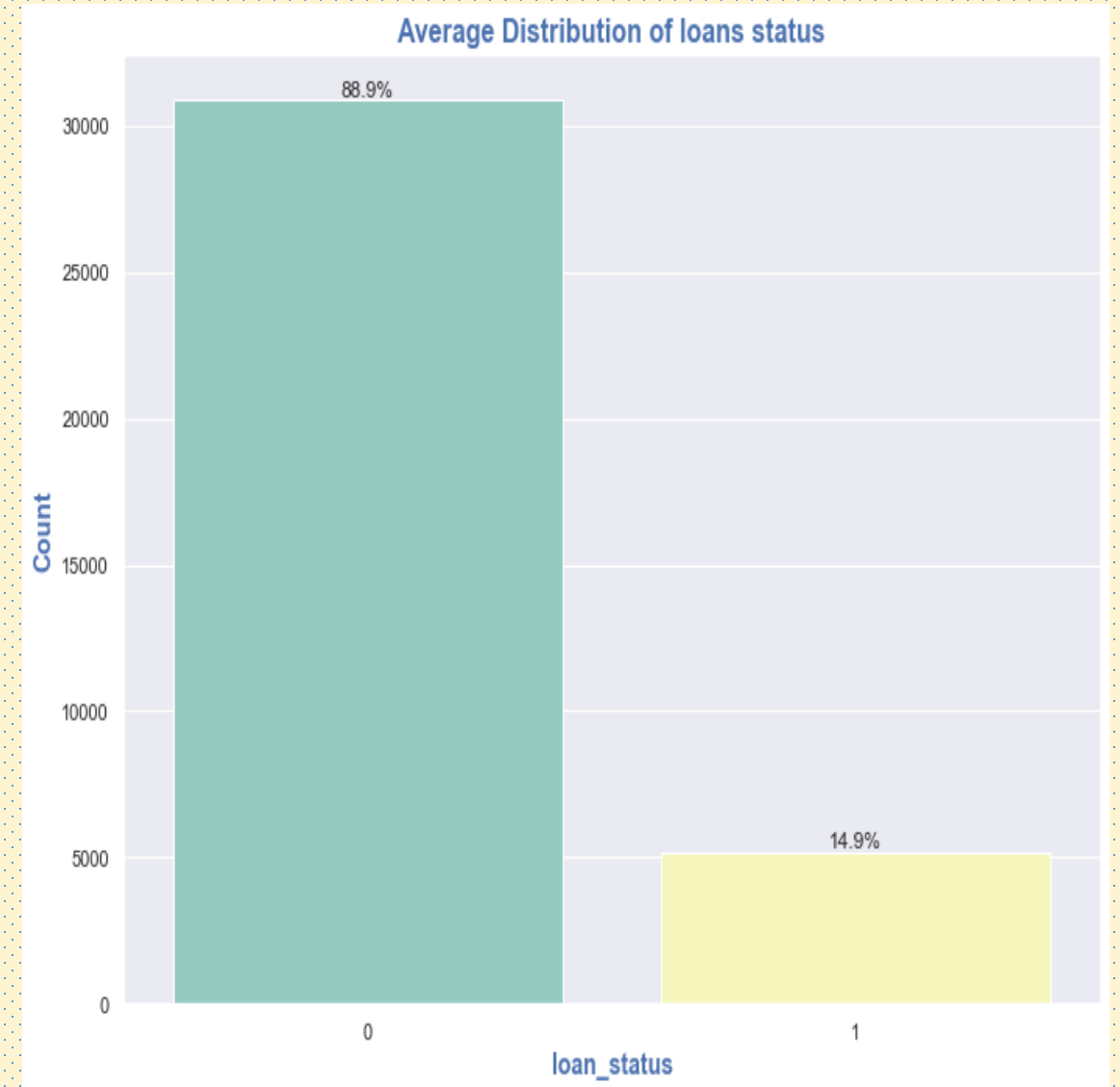
Average Distribution of loans across employee experience
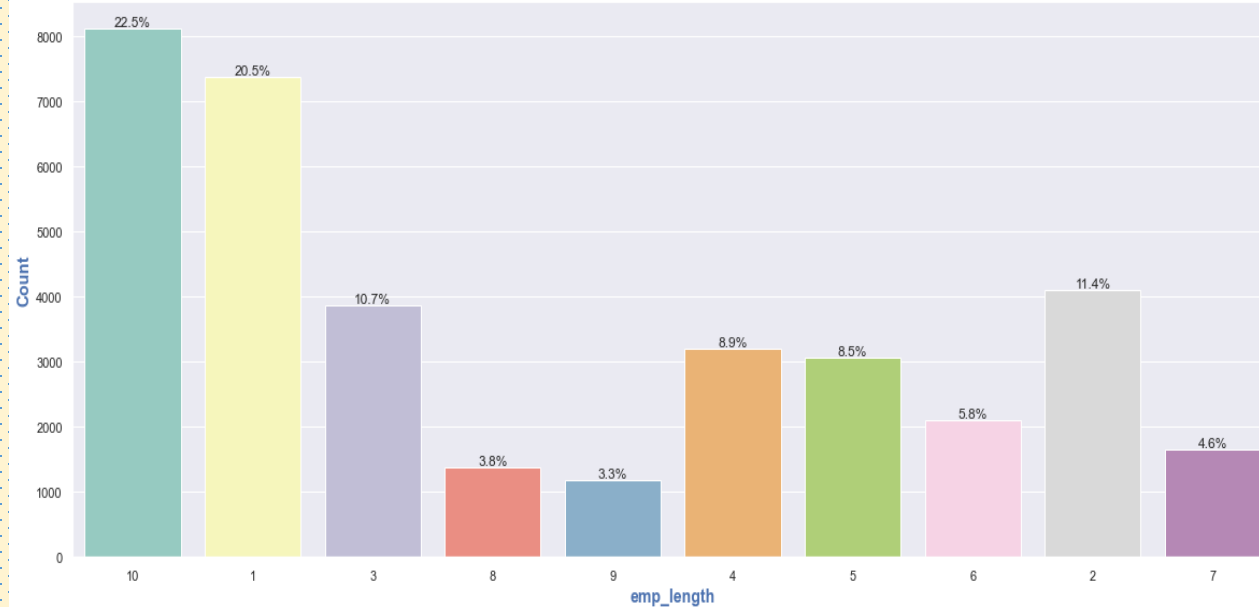
Distribution of Rented ownership on Experience of employee

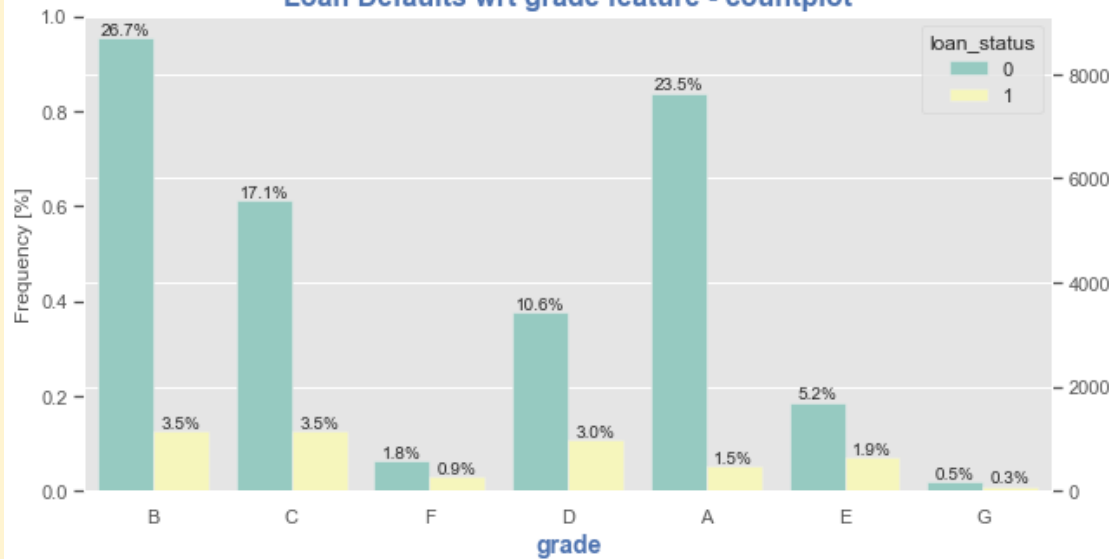Distribution of Verified Source Income on employee experience
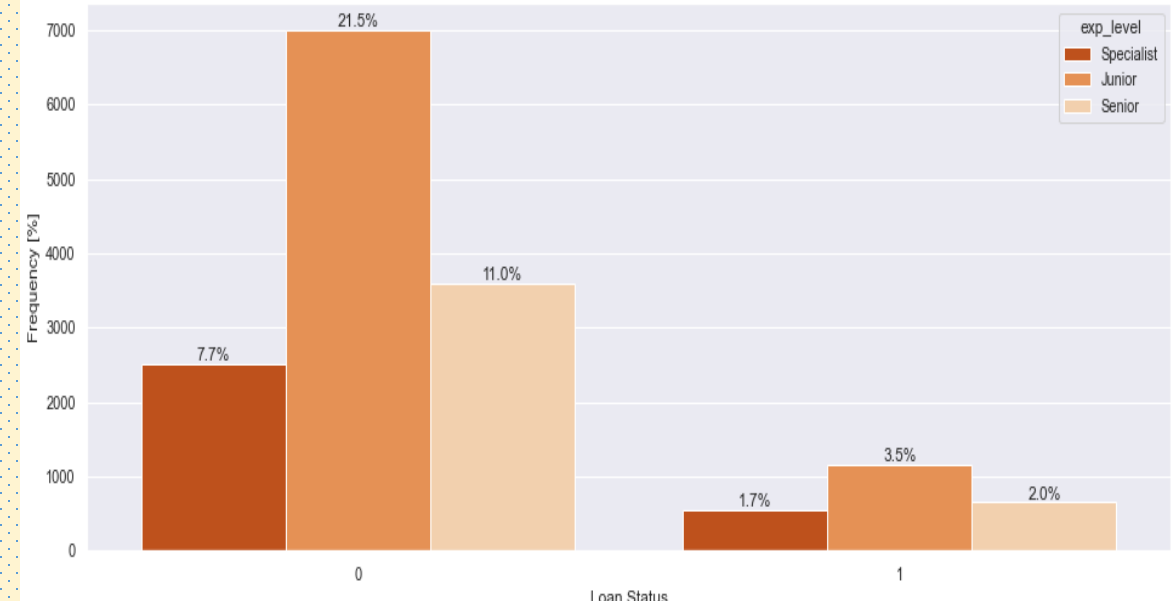
Observations for Univariate and Bivariate:

- Employees having 1 year and 10 years experience are taking loans frequently.
- Employees with the lesser experience are tends to be more default as compared to the other experience level.
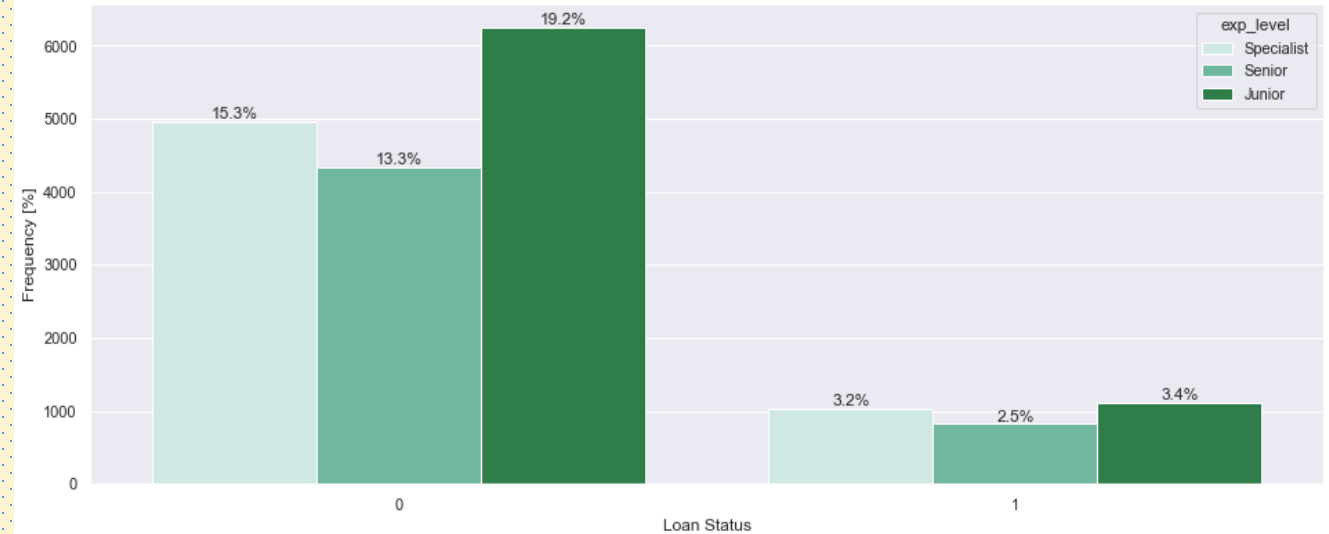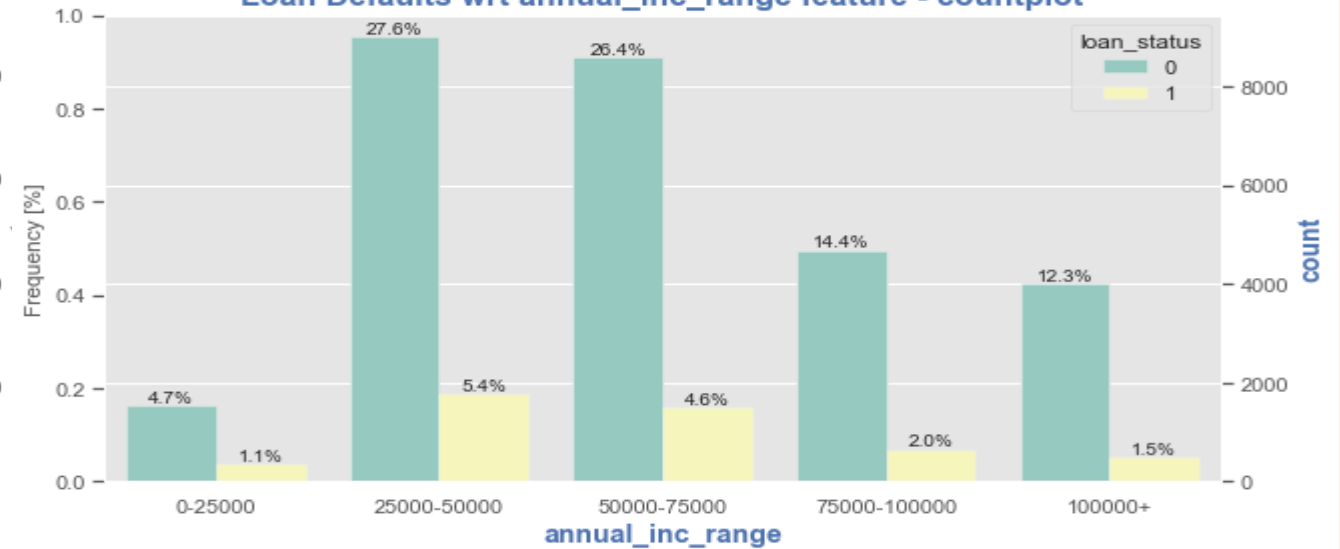
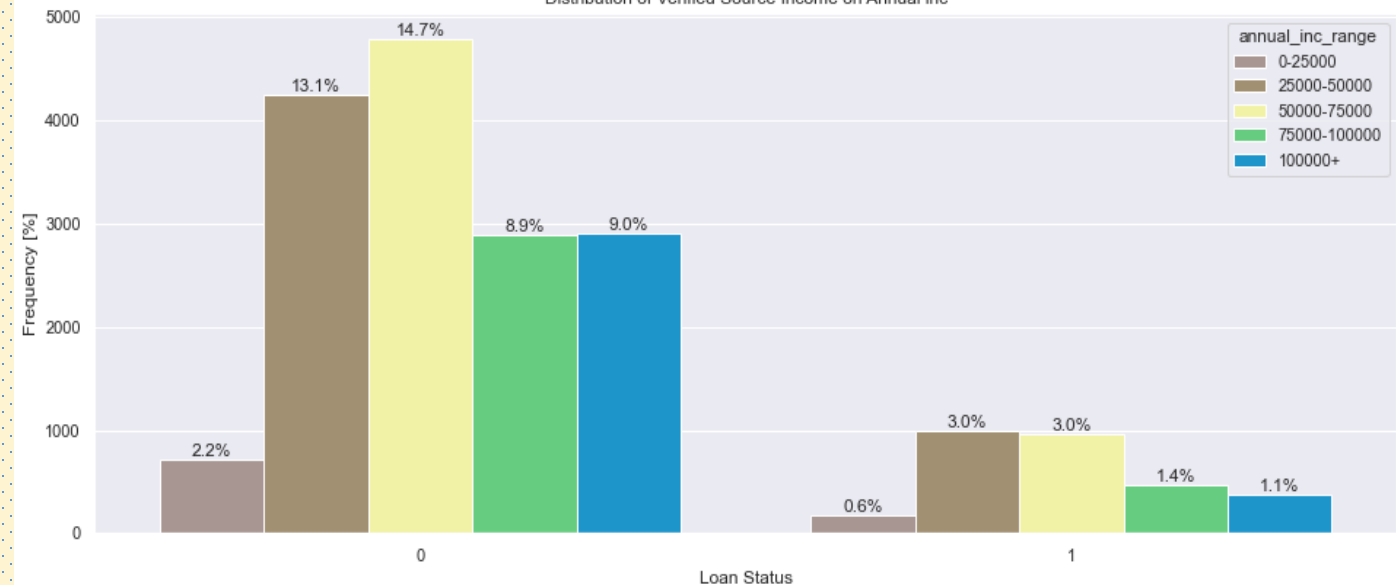Loan Defaults wrt grade feature - countplot

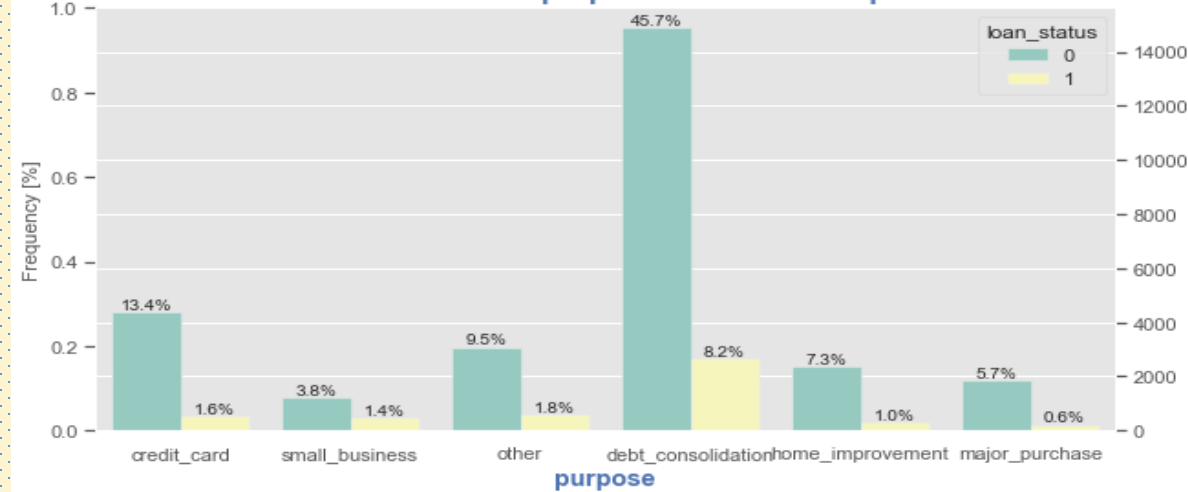Loan Defaults wrt annual_inc_range feature - countplot

Observations for Univariate for Garde ,Annual Inc Range and Bivariate for Annual Income range

- LC assigned grades B,C and D are having max chances for getting default.
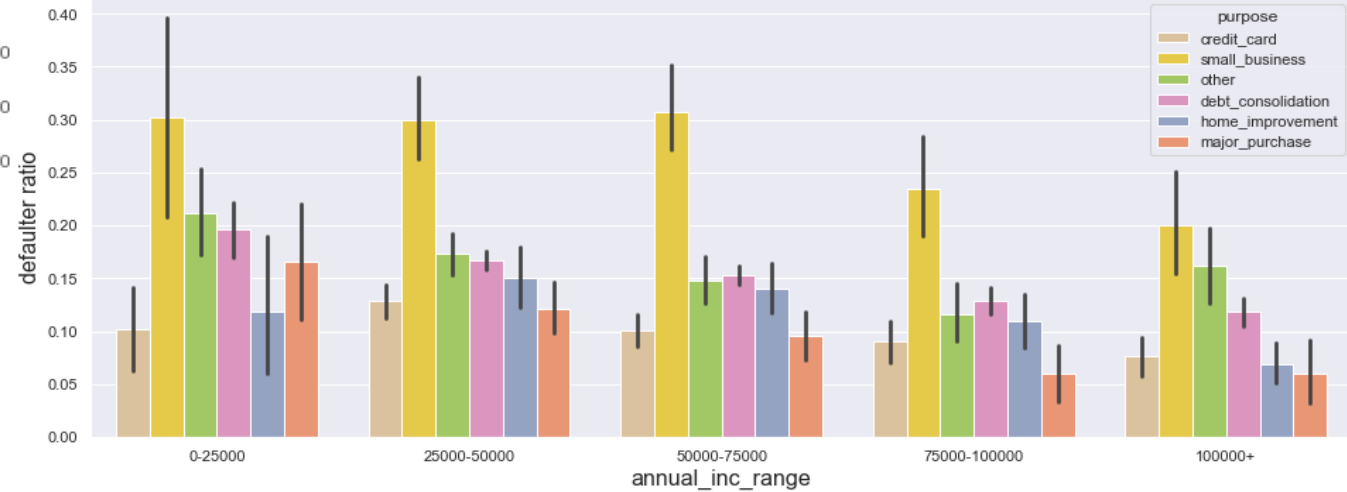- Annual Income with range 25000-75000 are having maximum default rate

Distribution of Verified Source Income on Annual inc

Purpose Analysis

Loan Defaults wrt purpose feature - countplot

Loan Default ratio wrt annual_inc_range feature for hue purpose in the data using countplot

Observations for Univariate and Bivariate:
- Debt consolidation , small business and others are having maximum default rates.
- As we can see straight lines on the plot, default ratio increases for every purpose wrt loan_amnt_range
- From the above graph we can see that the default rate increases with the term and also the max defualt rate is for the small business
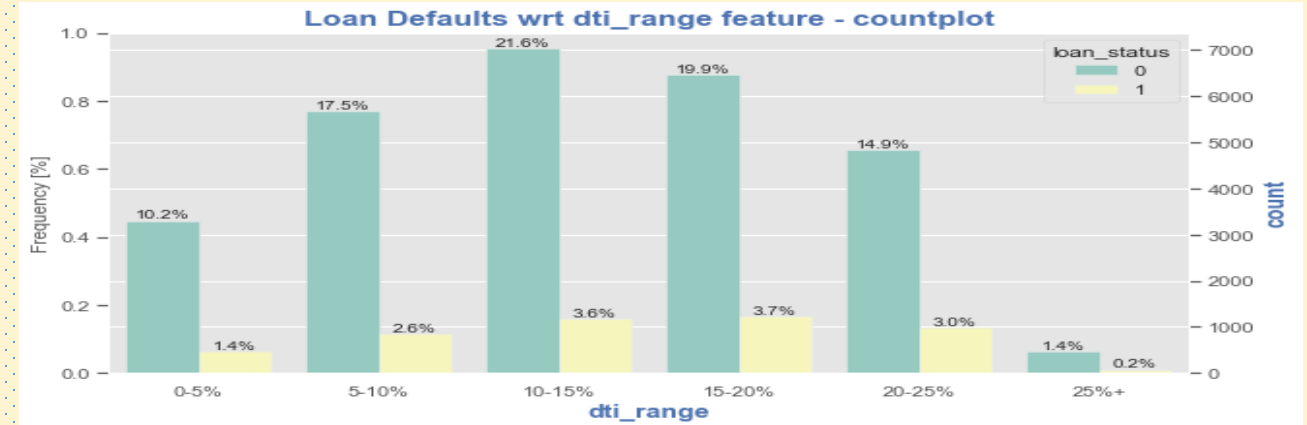

Loan Default ratio wrt loan_amnt_range feature for hue purpose in the data using countplot
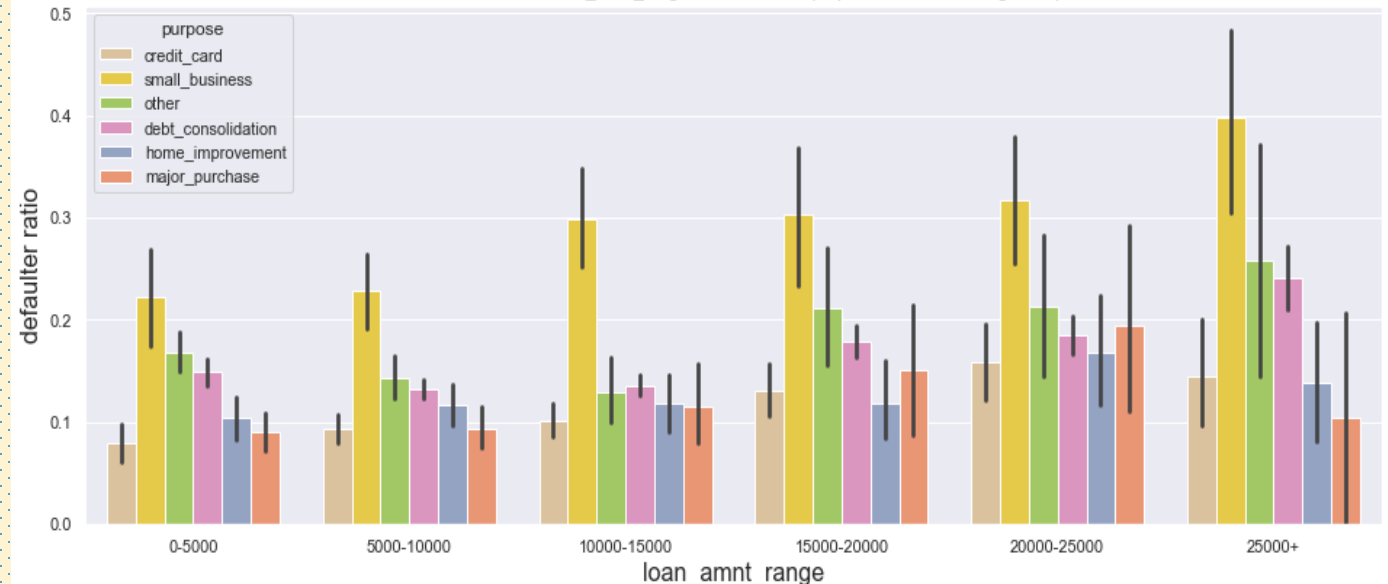
Months Analysis

DTI range

Loan Defaults wrt month feature - countplot

Loan Defaults wrt dti_range feature - countplot
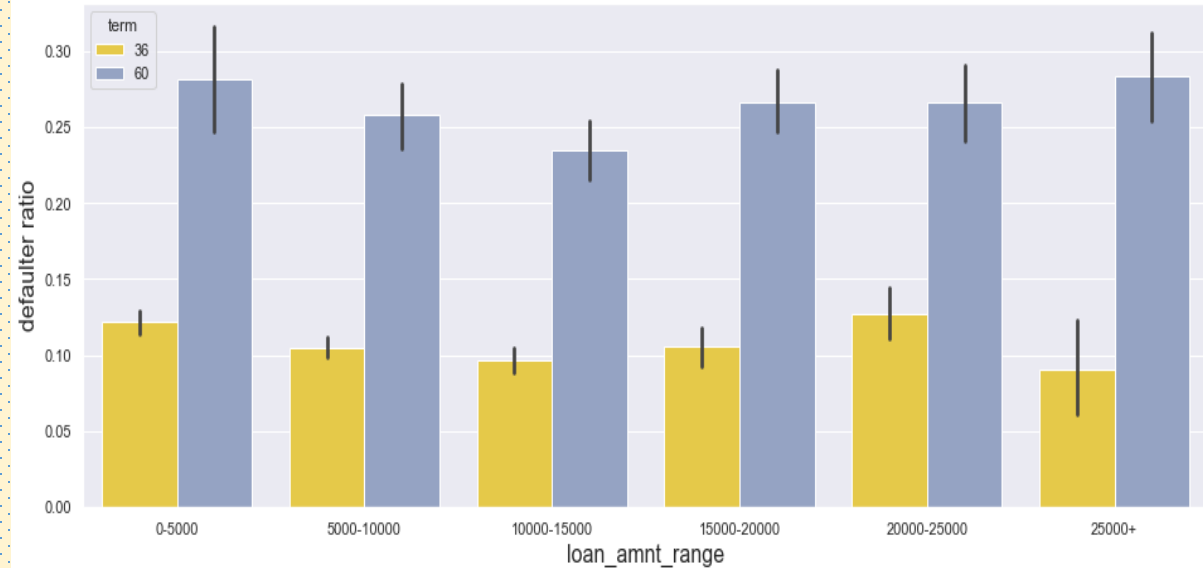
Home-ownership
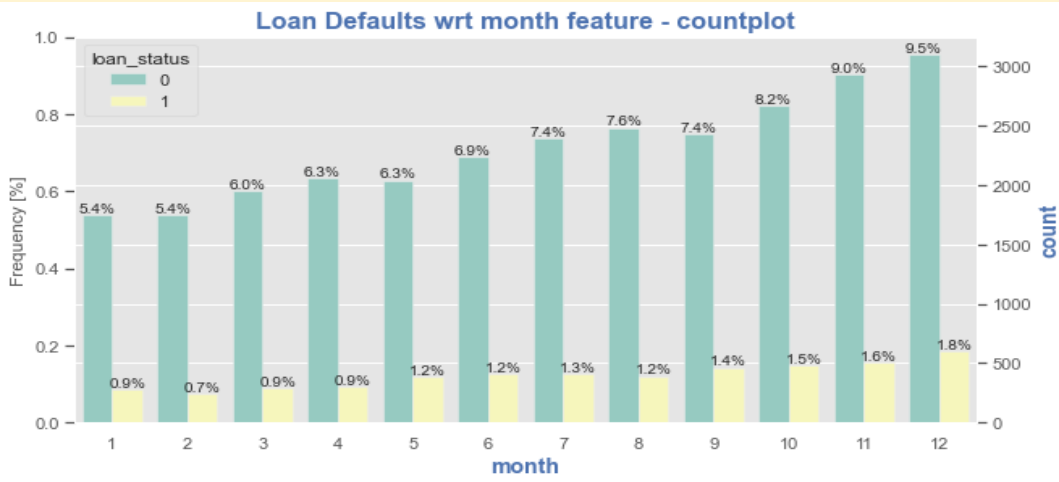
- Observations for Univariate:
- Months like Sept, Oct, Nov and Dec which are at the end of the year are having max default rates.
- DTI Range 10-25% are having max default rates
- People with Rented and Mortgage property are more likely to default.

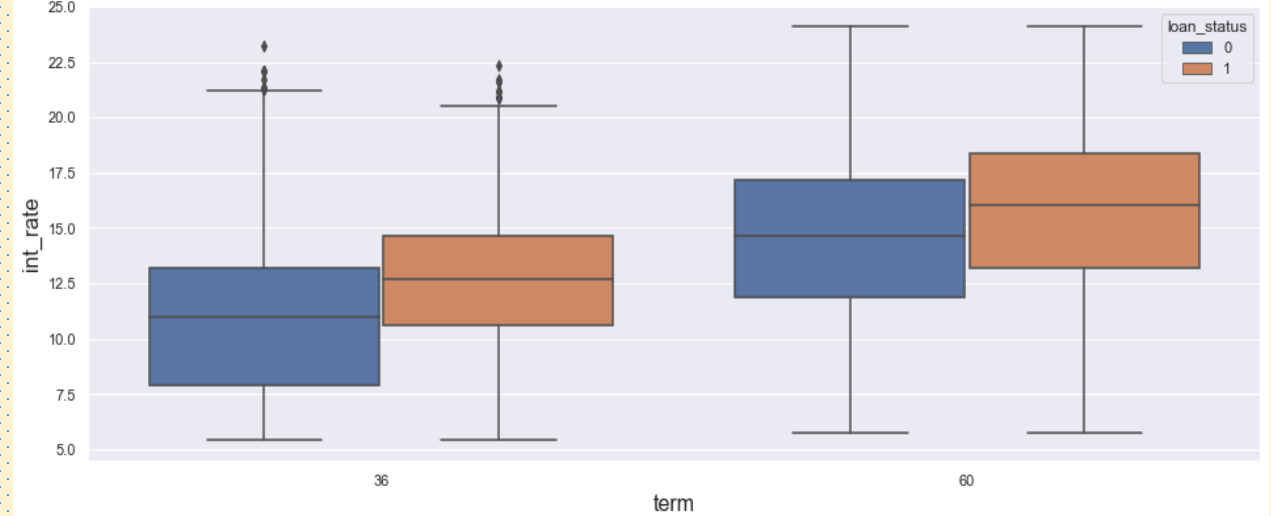Loan Defaults wrt home_ownership feature - countplot
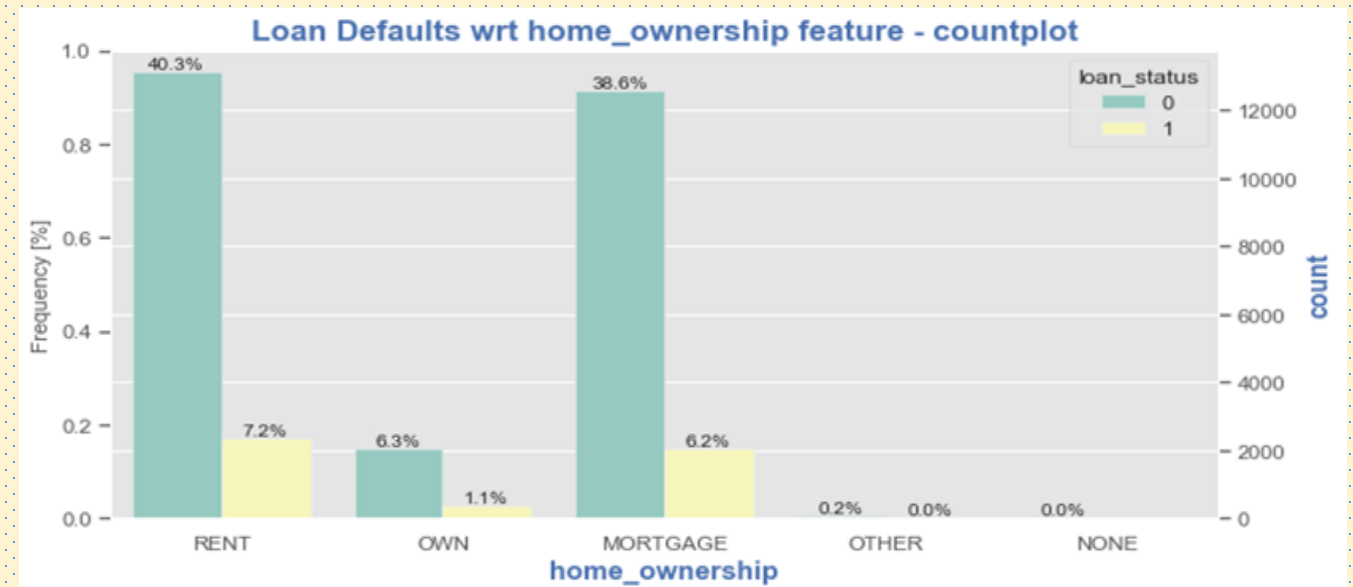
Loan Default ratio wrt loan_amnt_range feature for hue term in the data using countplot

Box plot between term and int_rate for each loan_status

Box plot between grade and int_rate for each loan_status

Observations for Bivariate:
- Default ratio increases for every term wrt loan_amnt_range
- int_rate increases with term on loan and the chances of default also increases.
- int_rate is increasing with every grade and also the defaulters for every grade are having their median near the non-defaulter 75% quantile of int_rate
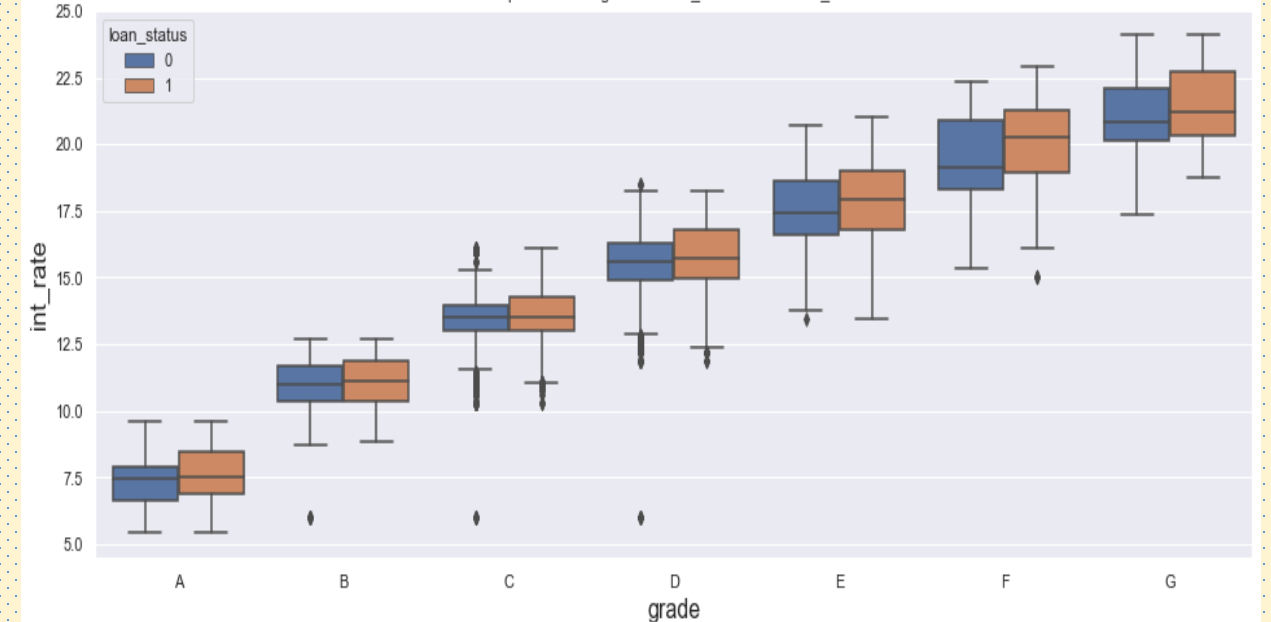
Conclusions

➢ The Variables which are strong indicators for default are: Employee length, Term, Interest rate, DTI, Annual Income, Grades, Home ownership, Purpose.

➢ Employee with the Junior level experience are more likely to default.

➢ People with the DTI range between 10% - 25% are more likely to default.

➢ People with Annual income range 25000-75000 are more likely to default.

➢ People with LC assigned grades as B,C and D are more likely to default.

➢ People with the Rented property and Mortgage property are more likely to default.

➢ we can summarize as for the small business, debt consolidation and others there are max defaulters.

➢ Default rate increasing with the Term and interest rate.

➢ loan amount between the range 5000-10000 are having the maximum default rates while the greater the loan amount lesser the default rate.