Lending Club

Arvind Kumar Patel

Background –Lending Club Case Study

Background

- Lending club is the largest peer-to-peer marketplace connecting borrowers with lenders. Borrowers apply through an online platform where they are assigned an internal score. Lenders decide 1) whether to lend and 2) the terms of loan such as interest rate, monthly instalment, tenure etc.

- Some popular products are credit card loans, debt consolidation loans, house loans, car loans etc.

Business Objective

- To identify variables which are strong indicators of default and potentially use the insights in approval / rejection decision making.

Data Understanding

Shape of data frame	39717 X 111
Null Value Present	Yes
Type of Column	Numerical, object, date-time

Data Understanding

If we observe columns, we can categories columns into 3 categories

Types of variables

- Customer demographic
- Loan related information & characteristics
- Customer behavior (if the loan is granted)

Type	Examples
Borrower's demographic variables	emp_title, emp_length, description, Zip_code
Loan Characteristics	Loan_amount, funded_amount, funded_amount_inv, interest_rate, grade
Customer's behaviour variables	Revol_bal, revol_util, delinq_2yrs mths_since_last_delinq

Data Cleaning

- **Delete columns**: Delete unnecessary columns.

- **Remove outliers**: Remove high and low values that would disproportionately affect the results of your analysis.

- **Missing values:** Treat missing values with appropriate approach.

- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.

- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Data Analysis

- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

- We performed **data cleaning and preparation** on the Loan dataset:
    - Convert columns
    - Created two new columns:
        - ➢ Splitting month and year from "issue_d" column.
        - ➢ Ratio of funded amount and annual income

During **univariate/bi-variate analysis** we have created:

Pie Charts , Heat Maps , Bar charts , Distribution Plots and line graphs to check out the distribution of all the driver variables.

Box plots to detect the Outliers

Performed the Multivariate analysis to understand how different variables interact with each other.

# Loan Status Analysis

Observations:

- 1. Most of the loans are Fully Paid i.e. Upto 85.4%.

- 2. About 14% of loan are having status as defaulters.

Loan Default Rate
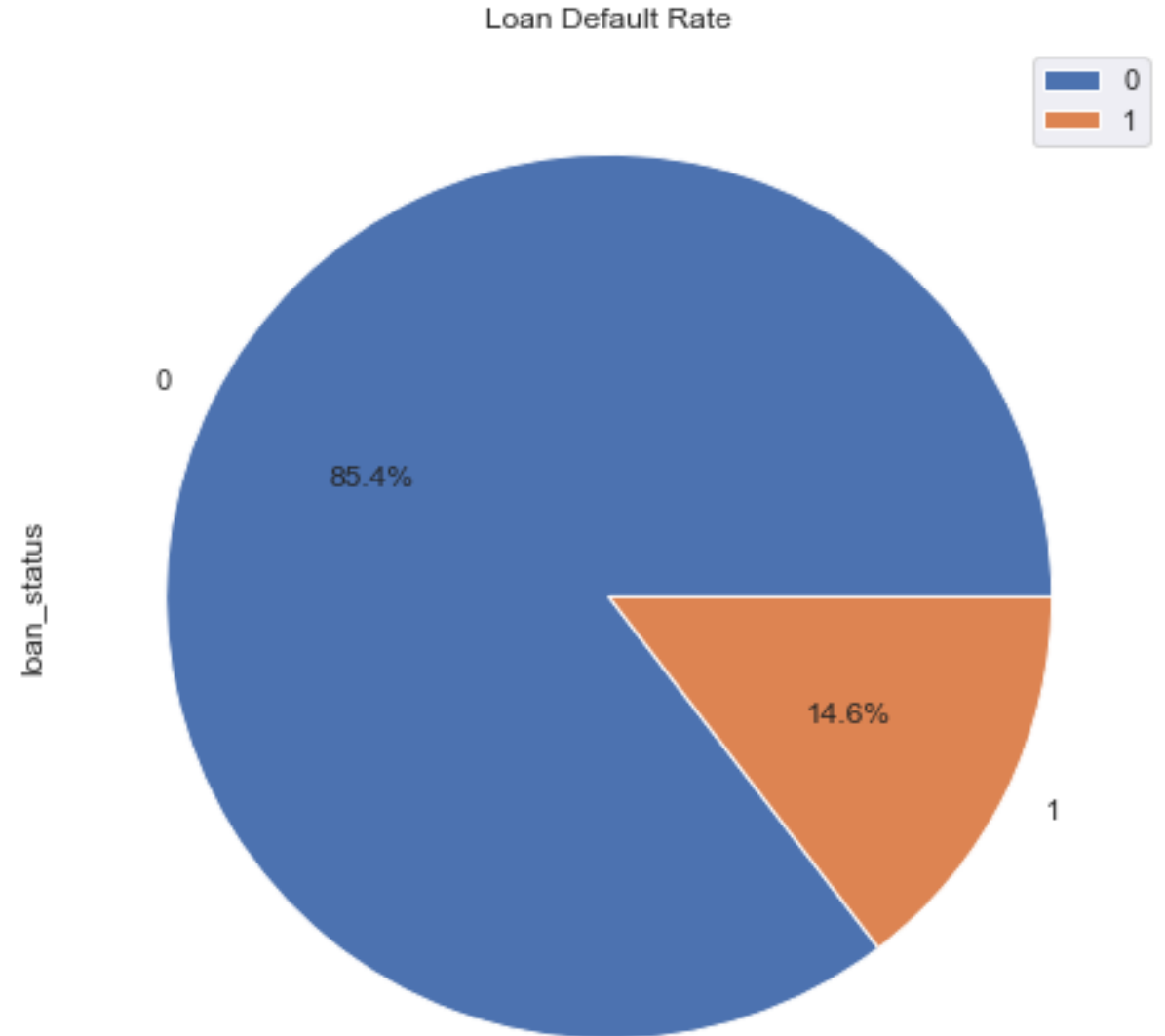
Understanding Correlated Variables for Loan Status

From the Heat Map we can notice that below variables are positively correlated to loan Status. Positive Correlation is increase in one variable tends to cause increase in another variable also.
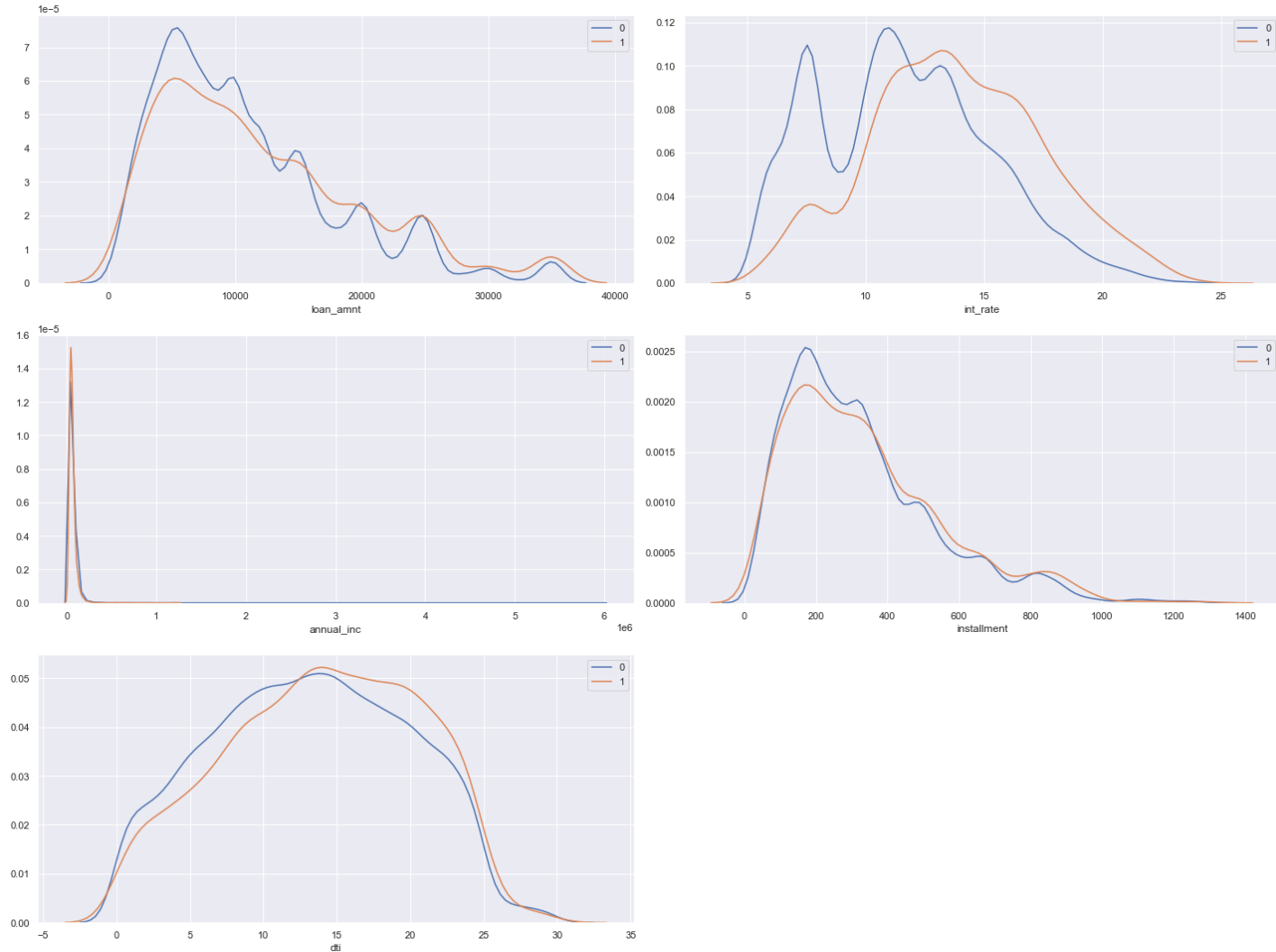
- loan amnt,
- funded amnt,
- funded_amnt_inv
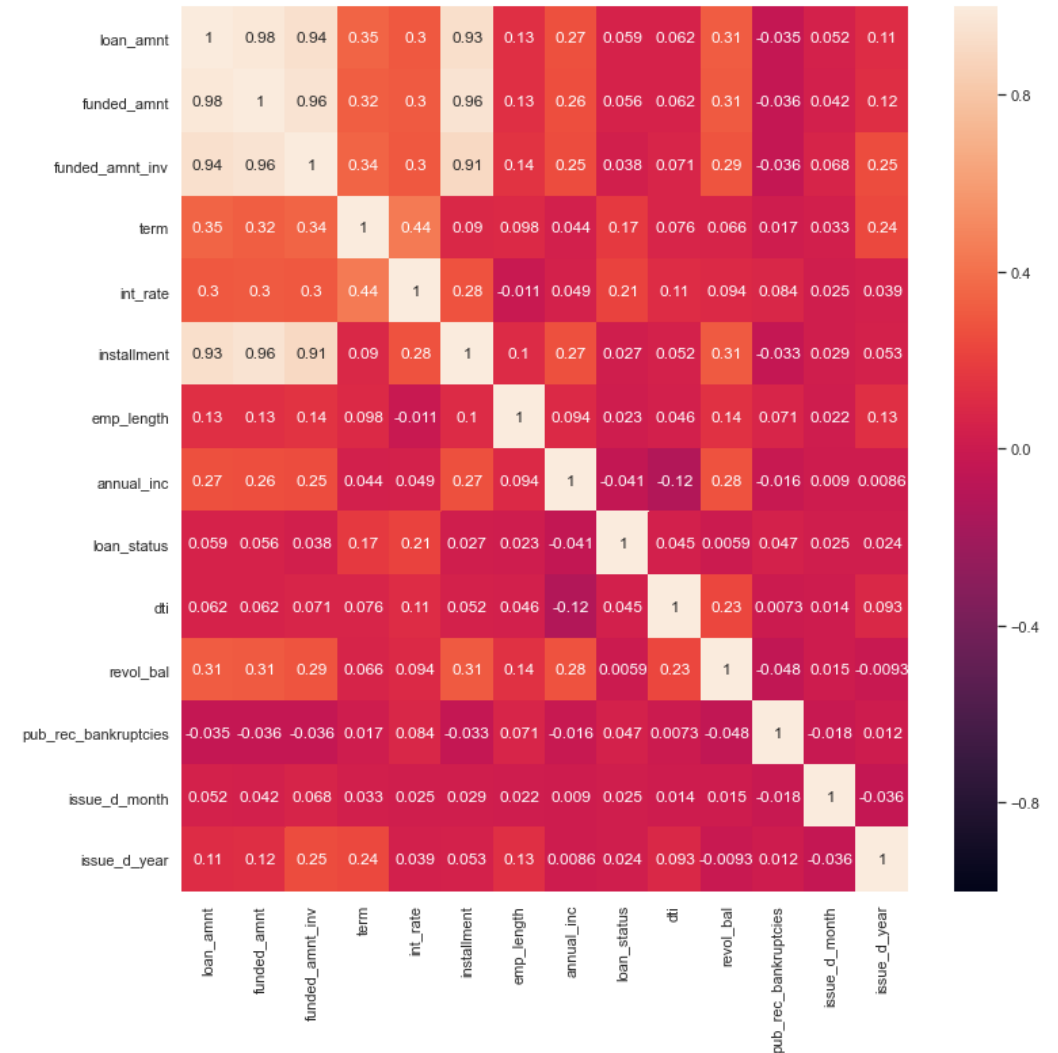
Analysis with Distribution Plots

Observation:

- If we see ***int_rate plot***, We can say those who acquire loan with high interest have higher chances of default.

- Those have high debt to income ratio have higher chances of default loan.

Loan Grade Analysis

Observation:
- Most of the loans have grade of A and B. Therefore stating most of the loans are high graded loans.
- High Quality loans have low interest rate.
- Which depicts, higher interest rate have higher tendency to default the loan.
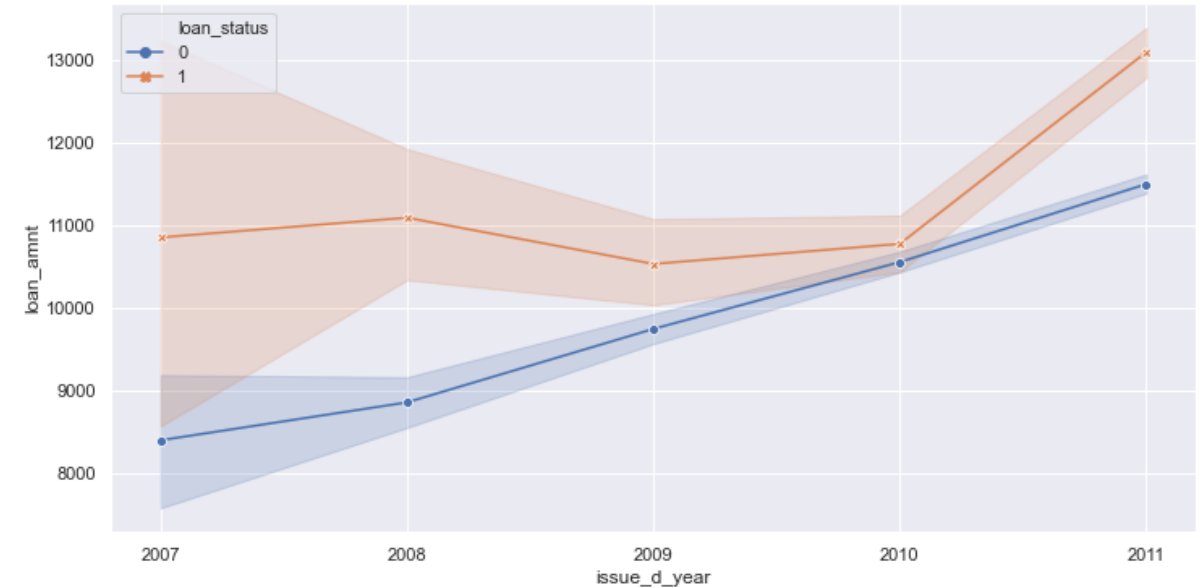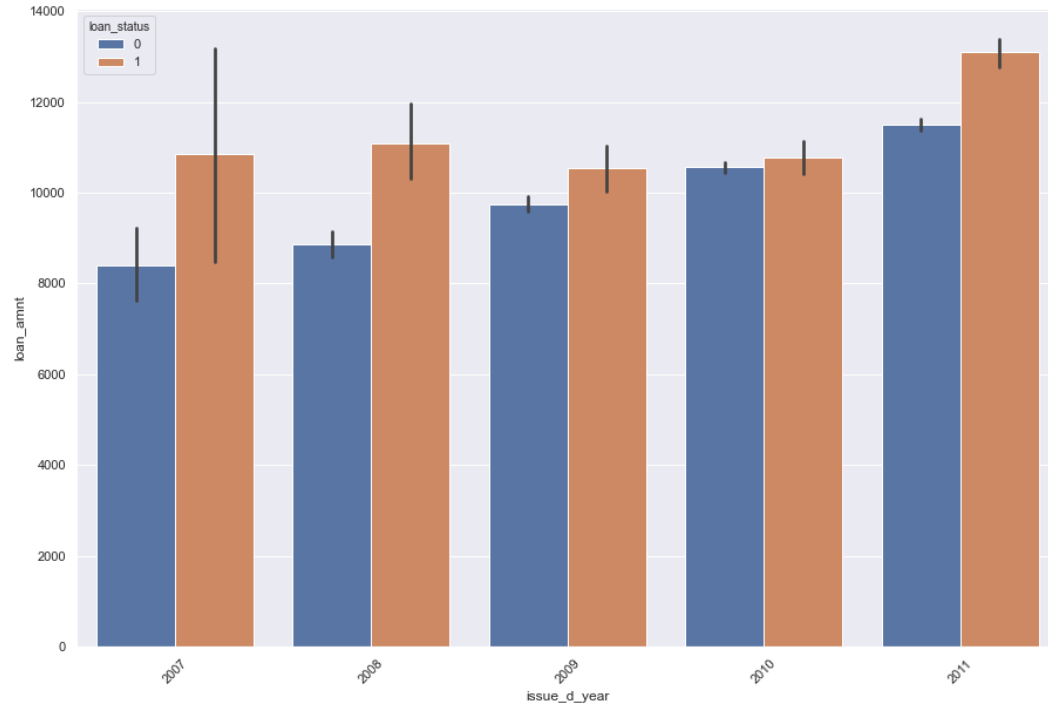

Count of Grade
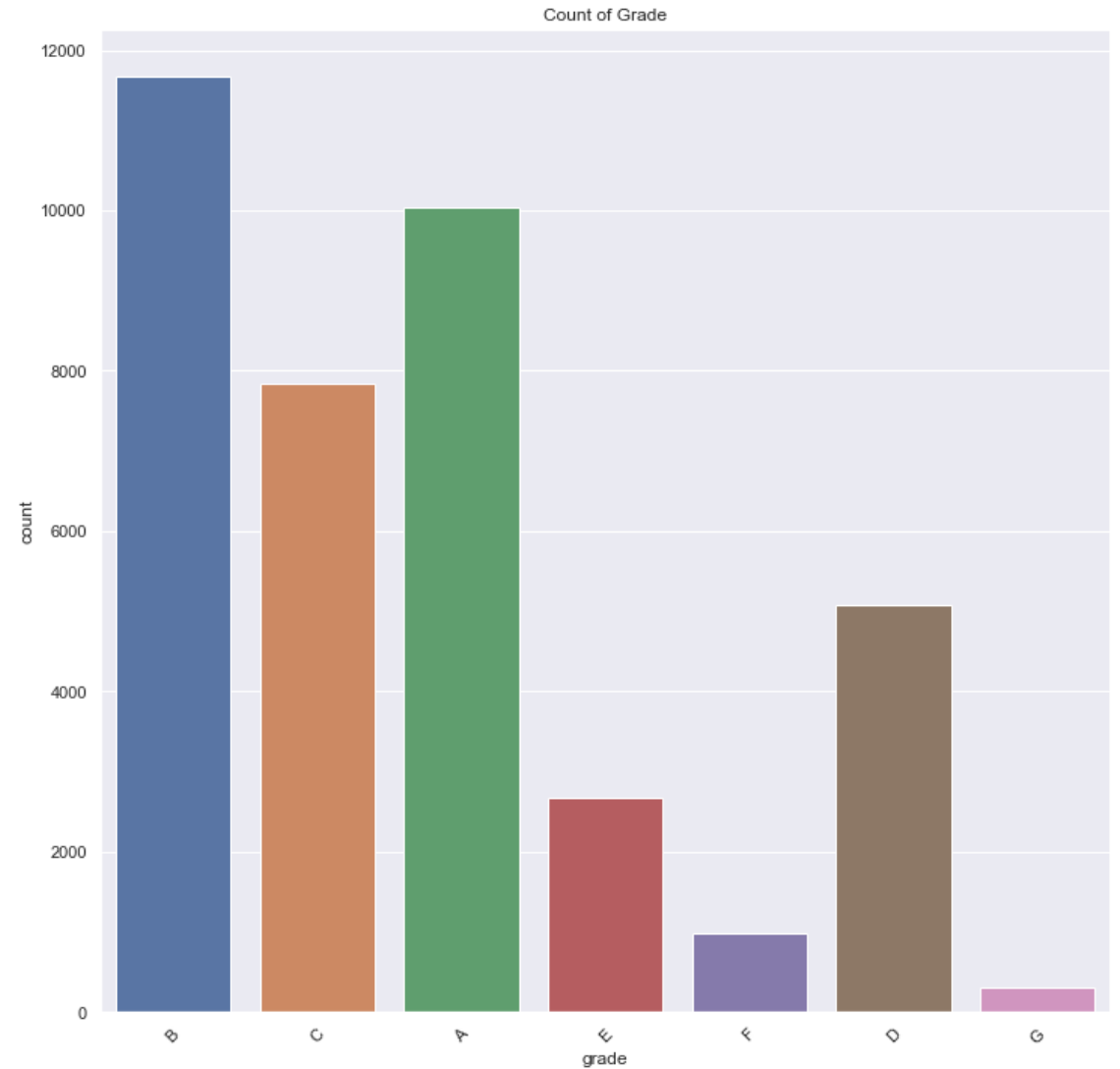
Loan Applicants Work Experience Analysis

Observation:

- Majority of employees applying for the loan have more than 10 years of experience

- Tendency of person to default the loan with 10 years of experience is also high. So, company need to be careful when granting loan.
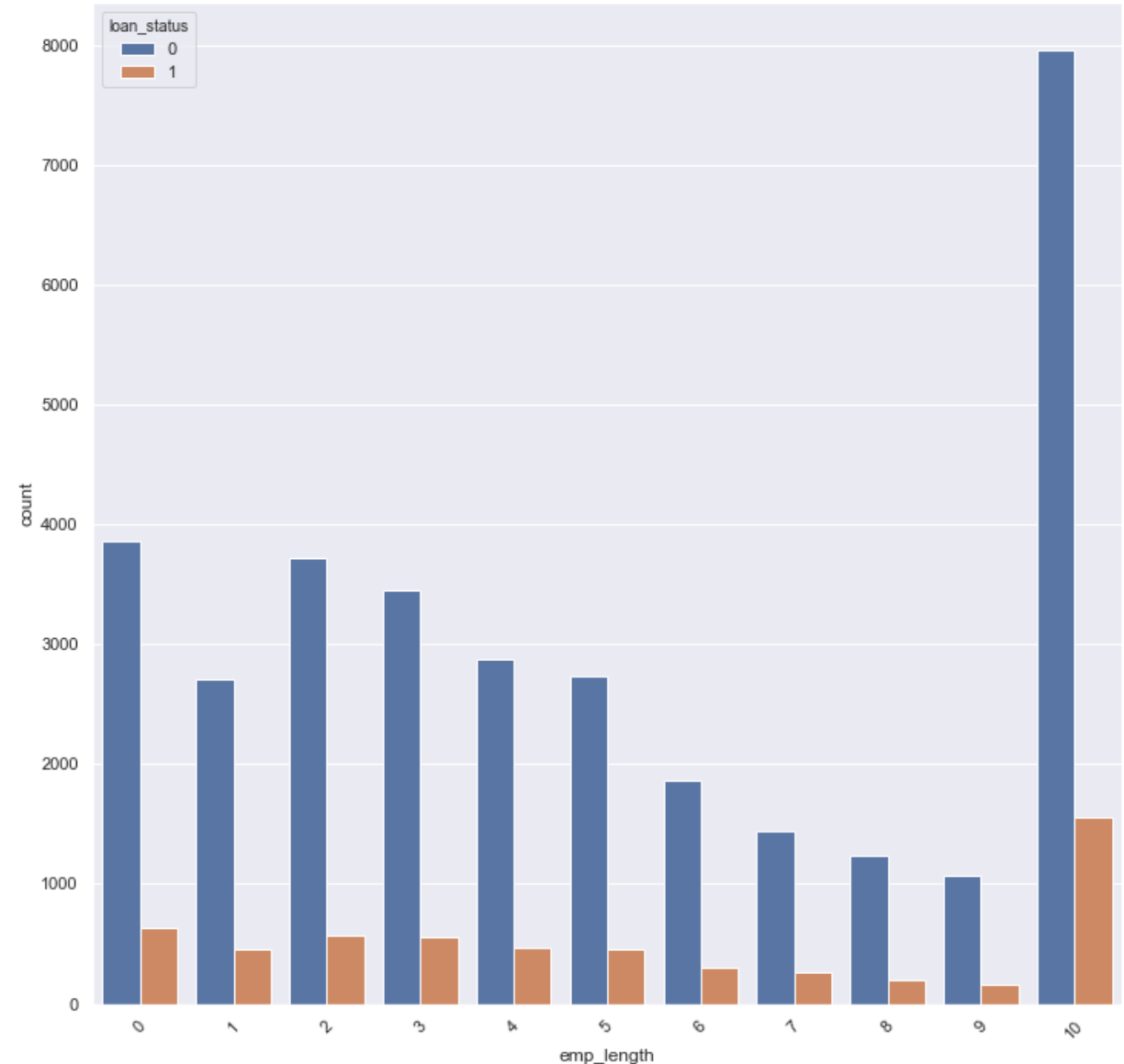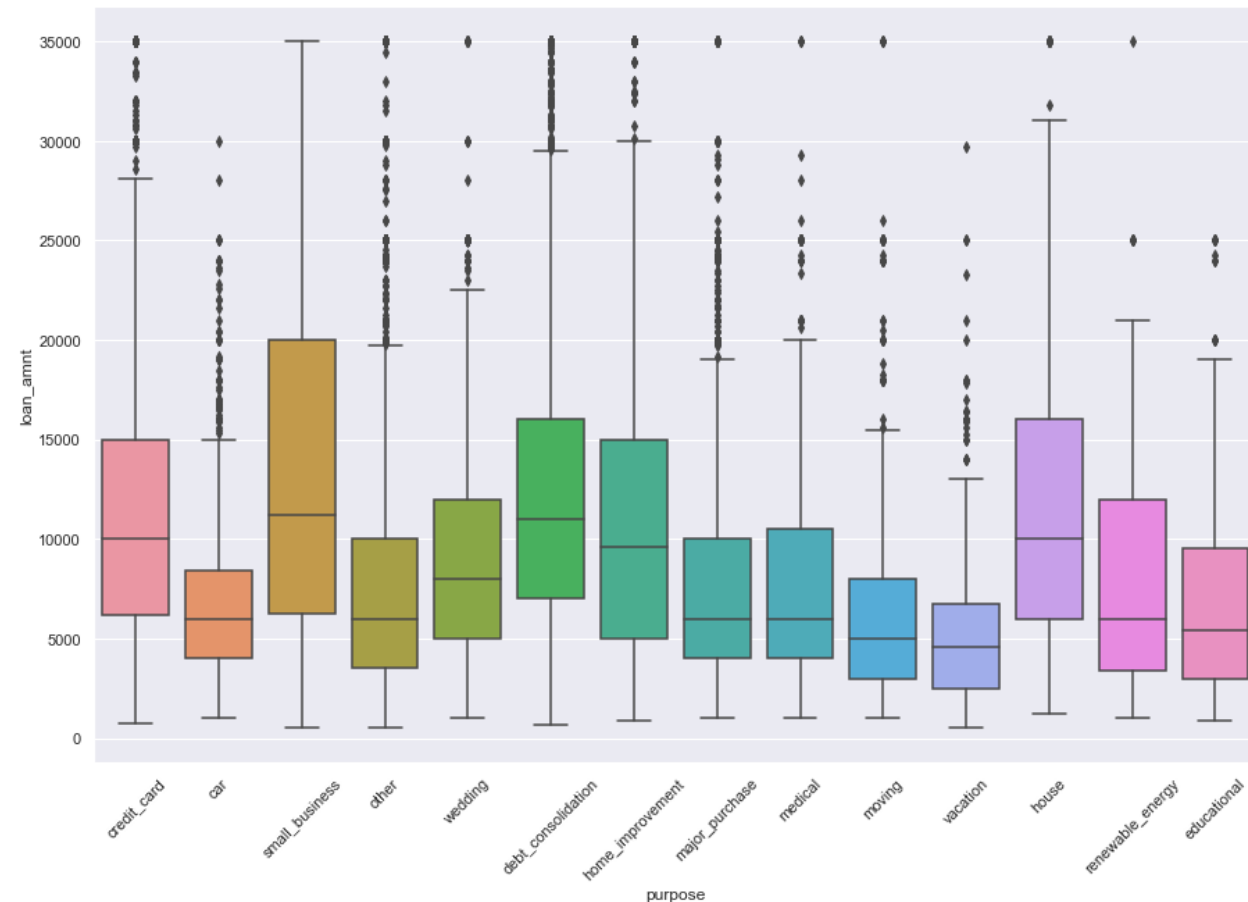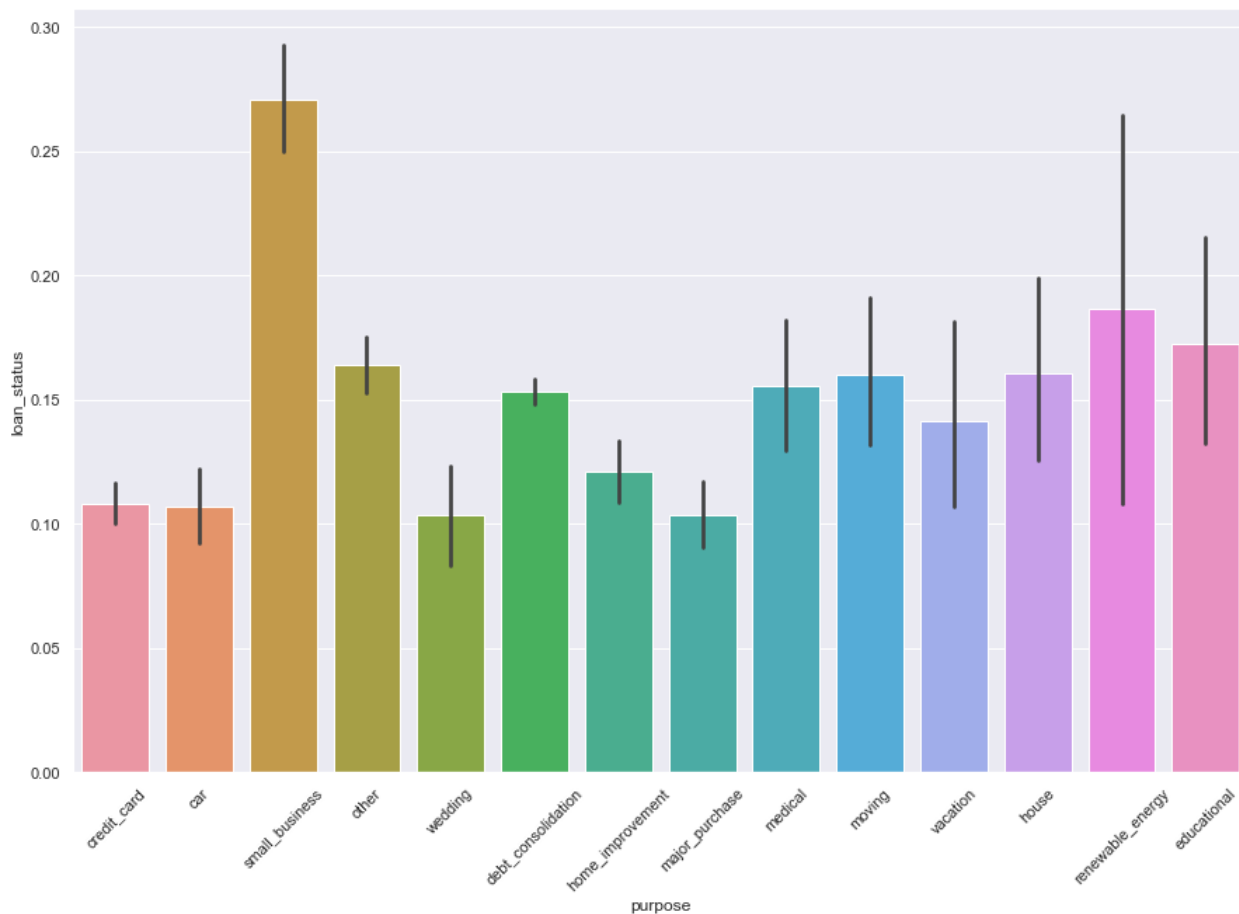
Loan Amount Issued in Time Frame– 2007-2011

Observation:
- Loan Amount issued increased in the year 2011.
- Also Default rate of loan is higher in 2011

Purpose for applying Loan Analysis

Observation:
- From the above two plots, we can conclude that people who are taking loan for small_business are high risk people as they ask for high loan amount and the chances of Default rate is high.

Income Analysis for the loan applicants

Observation:

- If we bin the annual income as among the loan applicants and compare it with the default rate we can notice that less income people have higher rate of defaulting.
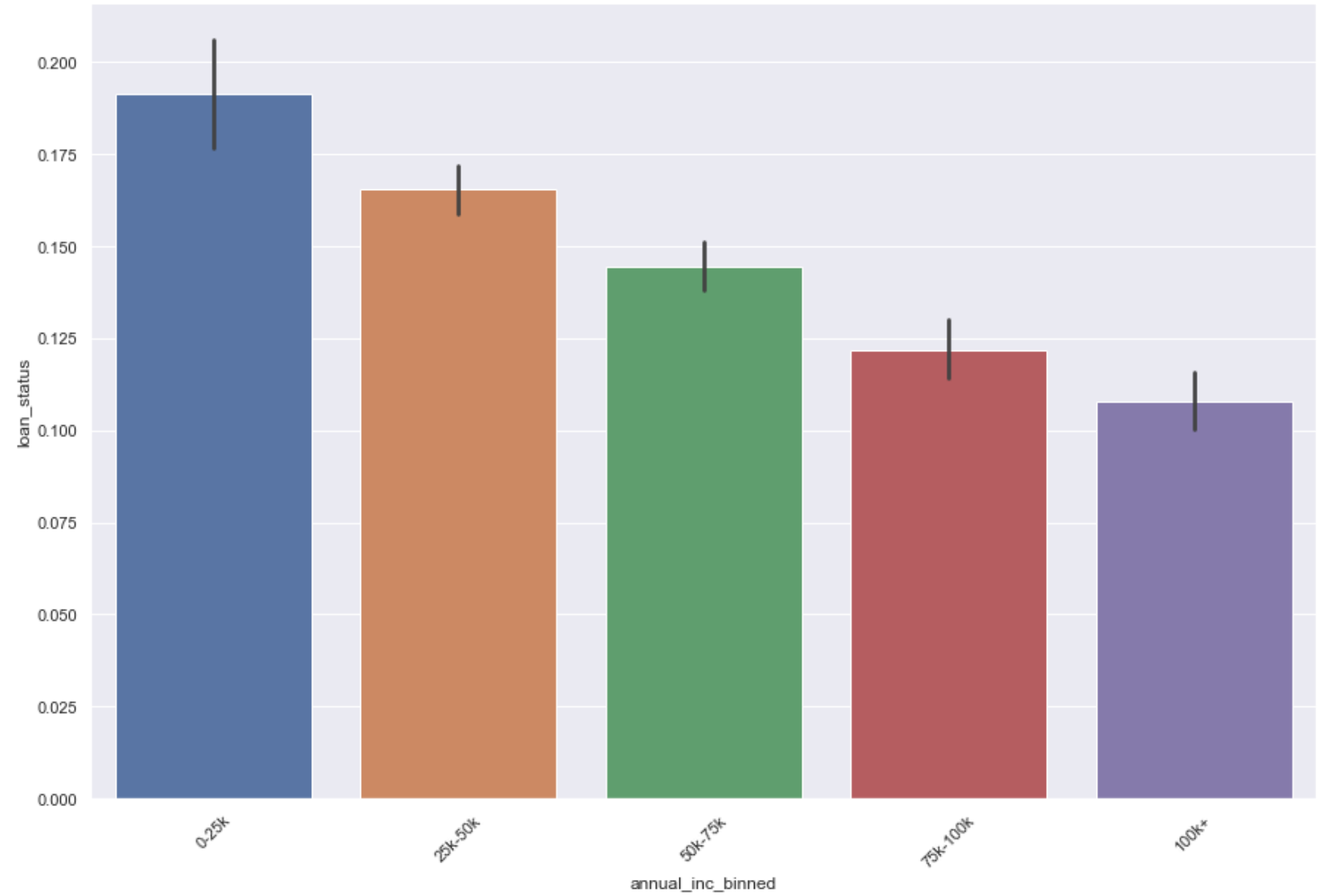
Conclusion and Recommendation

Conclusion

Our Target variable is loan status, which contain information related to who defaulted or not defaulted.

- ✓ Purpose of Loan - Loan amount "small_business' has the highest default rate
- ✓ Term - Loan granted for 60-month term have higher default rate when compared to loan with 36-month term
- ✓ Grade/Interest Rate - Increase in Grade/interest rate results in more default rate.
- ✓ Loan Amount - Higher the loan amount, more is the loan default rate.
- ✓ Annual Income - Lower the borrower's annual income, higher are the chances of default.

Recommendation

Company should cautious when someone is acquiring loan:-

- ✓ if lender taking loan for high term and for small_business purpose.
- ✓ if interest rate or loan amount is high and annual income is low.
- ✓ Basically, all the points given in Conclusion are prime key to default loan, company should be focused more on those.