Lending Club Case Study

Group Members:

1. Megharaj Hothur

2. Shivani Kharra.

Lending Club Case Study

Abstract:

This 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 aim of this case study is to identify risky loan applicants based on several factors.
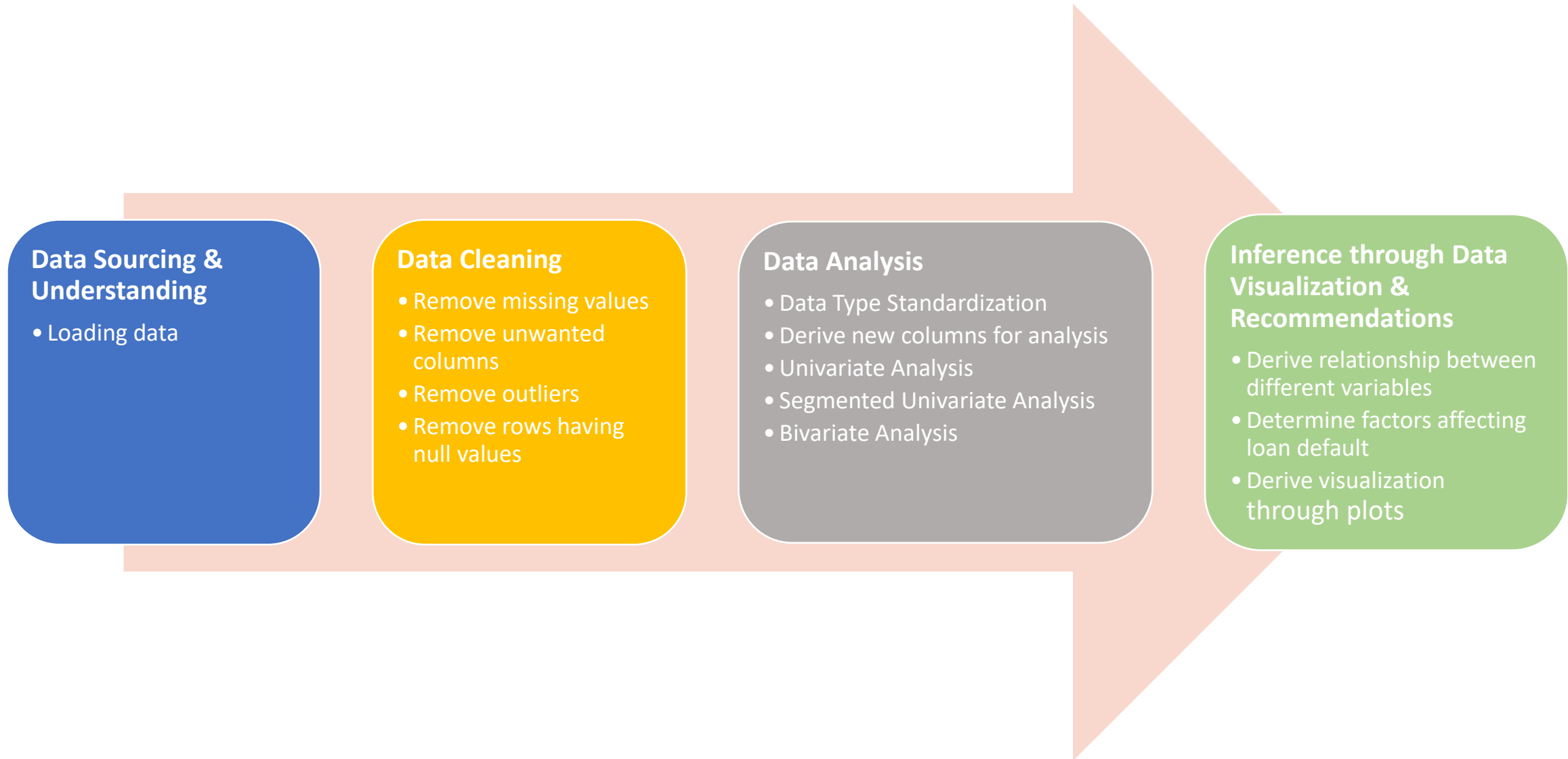
Business Problem Statement

- The lending company is incurring huge cost due to Credit loss
- Analyze dataset containing past loan applicants and their loan information using EDA to understand how customer-centric and loan-centric attributes can impact and influence a loan default

Context & Outcome of EDA

- Lending Club 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.

Exploratory Data Analysis Approach

Data Sourcing & Understanding

• Loading data

Data Cleaning

• Remove missing values
• Remove unwanted columns
• Remove outliers
• Remove rows having null values

Data Analysis

• Data Type Standardization
• Derive new columns for analysis
• Univariate Analysis
• Segmented Univariate Analysis
• Bivariate Analysis

Inference through Data Visualization & Recommendations

• Derive relationship between different variables
• Determine factors affecting loan default
• Derive visualization through plots

Analysis & Assumptions

➢ We are provided with past loan dataset of customers -Loans and Meta data to understand the loans data itself.
➢ After detailed analysis , we identified the following variables that will have influence to loan defaulting for this analysis.

Interest Rate	Loan Grade or Credit Rating	Loan Purpose
Employee Length	DTI (Debt/Income Ratio)	Loan Amount
Annual Income (of borrower)	Verification status of the loan	Term of the loan
Home Ownership (of borrower)	Publicly Recorded Bankruptcies	

➢ The data set contained loans applied from the year 2007 to 2011
➢ only individual type loans were considered for the analysis.
➢ We have excluded loans that are currently in process as they do not provide any significance to the analysis.
➢ Only the loans with statuses of 'Paid Off' and 'Charged Off' are considered for analysis.
➢ A cleaned dataset of 27 columns and 36502 rows are considered for analysis
    ~14% of Charged off loans were present in the data set
    ~86% of Paid off loans were present in the data set
➢ We created few derived columns during the analysis-
    • Status –> 1 – Paid off; 0 – Charged off
    • For Issue_d column; years and months were separated into different derived columns
    • Binning of annual income of borrower , loan amount. Instalment, interest rate are performed
➢ The ratio of the charged off/default loans has been increasing over the years from 2007 to 2011

Loan Status Analysis

➤ Most of the loans from the dataset are Fully Paid

➤ The dataset has ~**14**% default rate with borrower's status changing to Charged-Off.
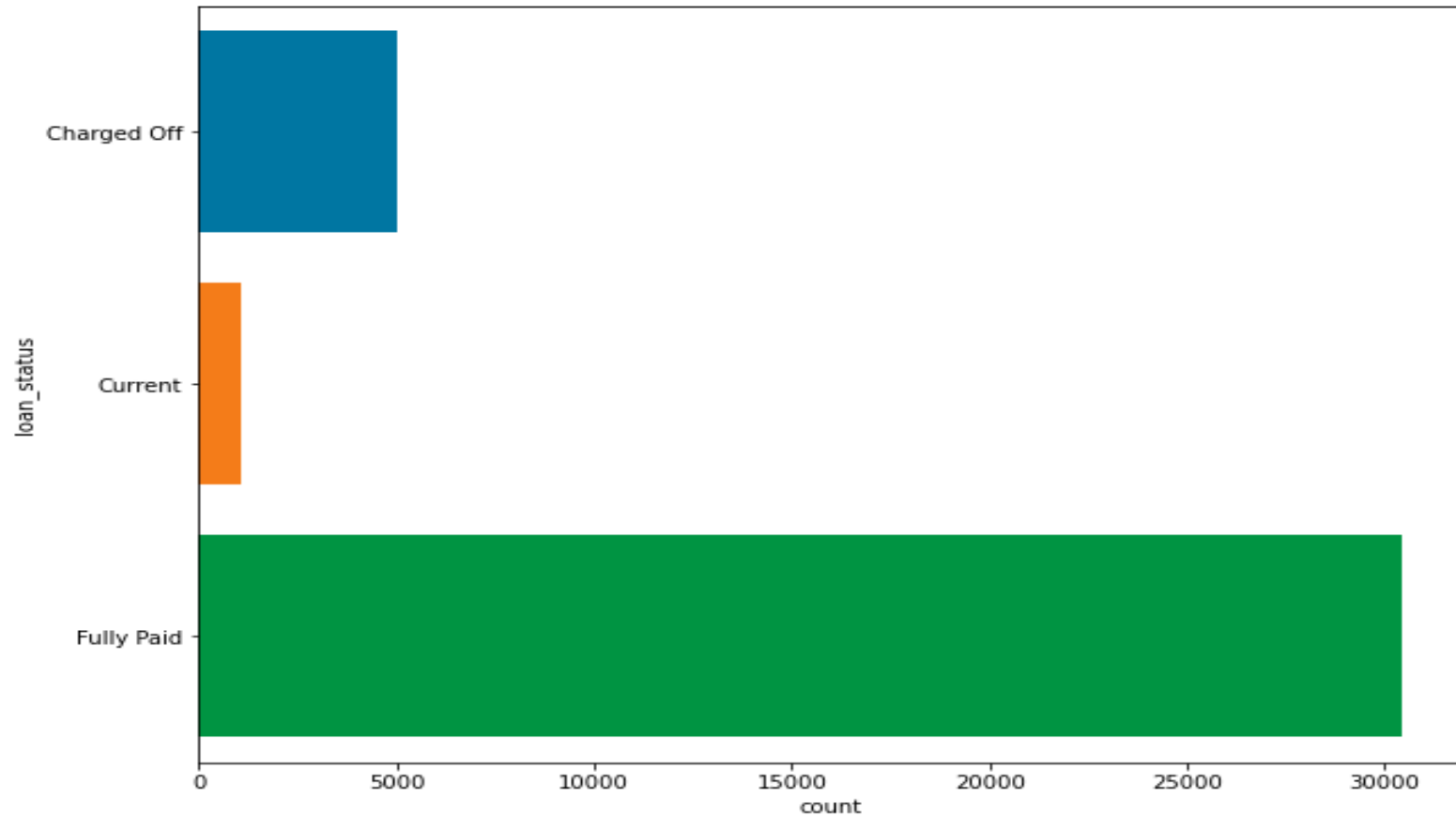
Interest Rate Analysis

➢ **Observations:**

• Interest rate increased for all grades, interest rates dropped to an extent for grades of A and B. The lower grades such as E,F & G have very high interest rates.

• Higher the interest rates, higher are the tendencies for the loan to be charged off.

➢ **Recommendation(s):**

• Lending club should look at less risker loans (grade A and B) as riskier the loan, higher the interest rate and higher the interest rate higher are the chances of defaulting.
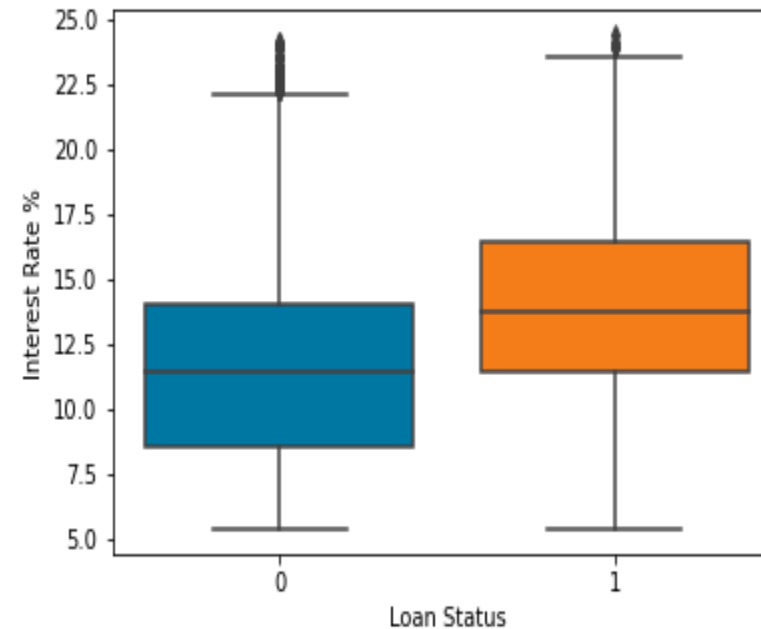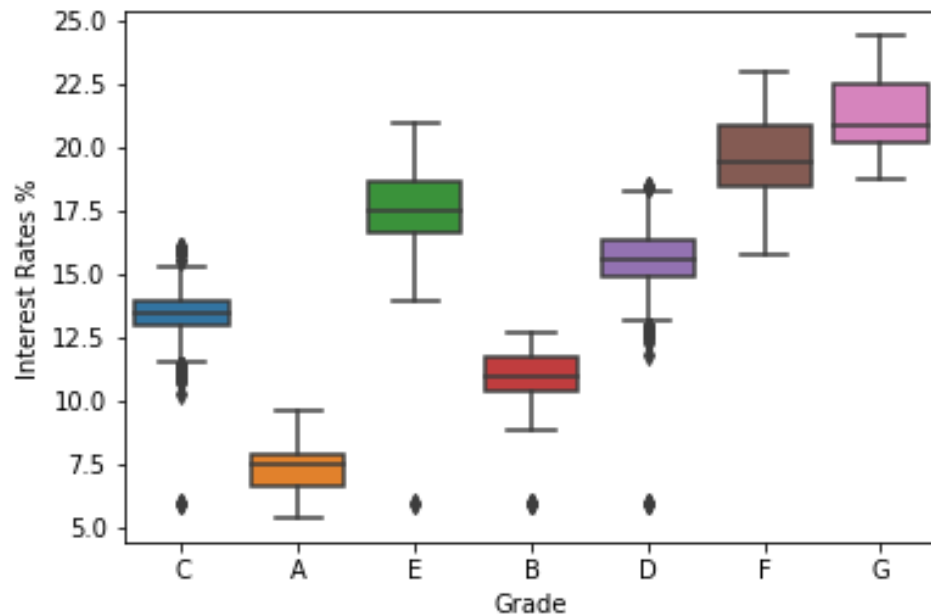
Loan Grade Analysis

➤ **Observations**:

- The lower the grades like E, F and G, higher charged off/default rates.

- Loans belonging to risk category of B and C have the highest number of charged off loans.

➤ **Recommendation(s):** Lending club should avoid providing to loans which belong to grade of E, F and G as default rates are high in those categories.
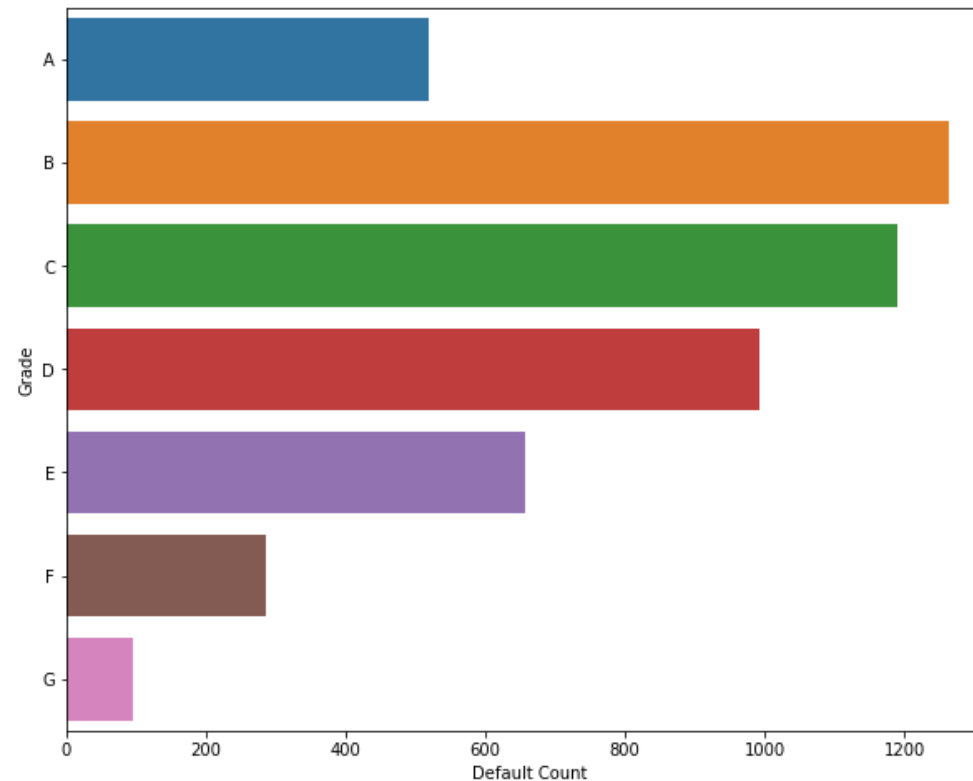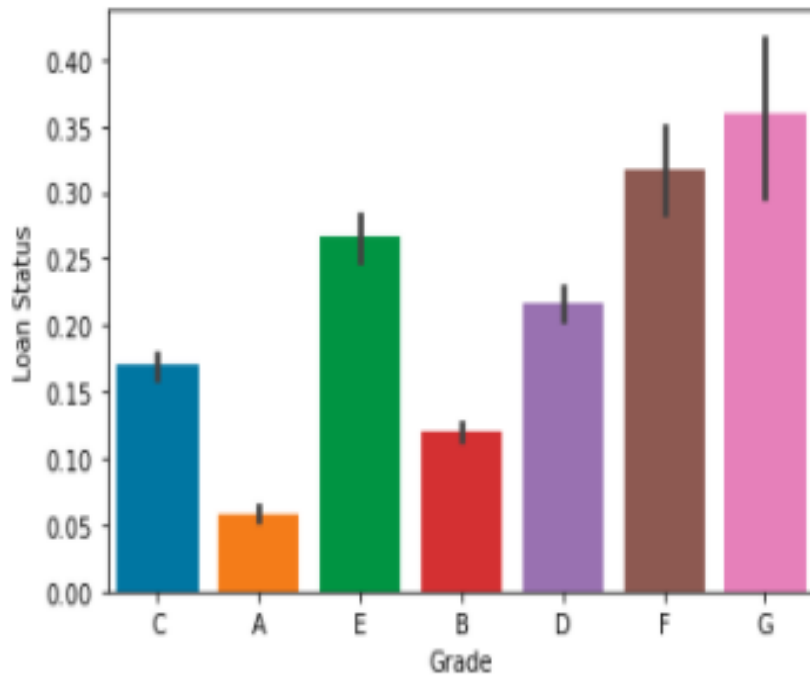
Loan Purpose Analysis

➤ **Observations**: With respect to count, loans with purpose of '*debt consolidation*' have the highest number of charged off loans.

• However, from default rate stand point, loans with purpose of '*small business*', 'renewable energy' and 'educational' have the highest default rate in the chronological order.
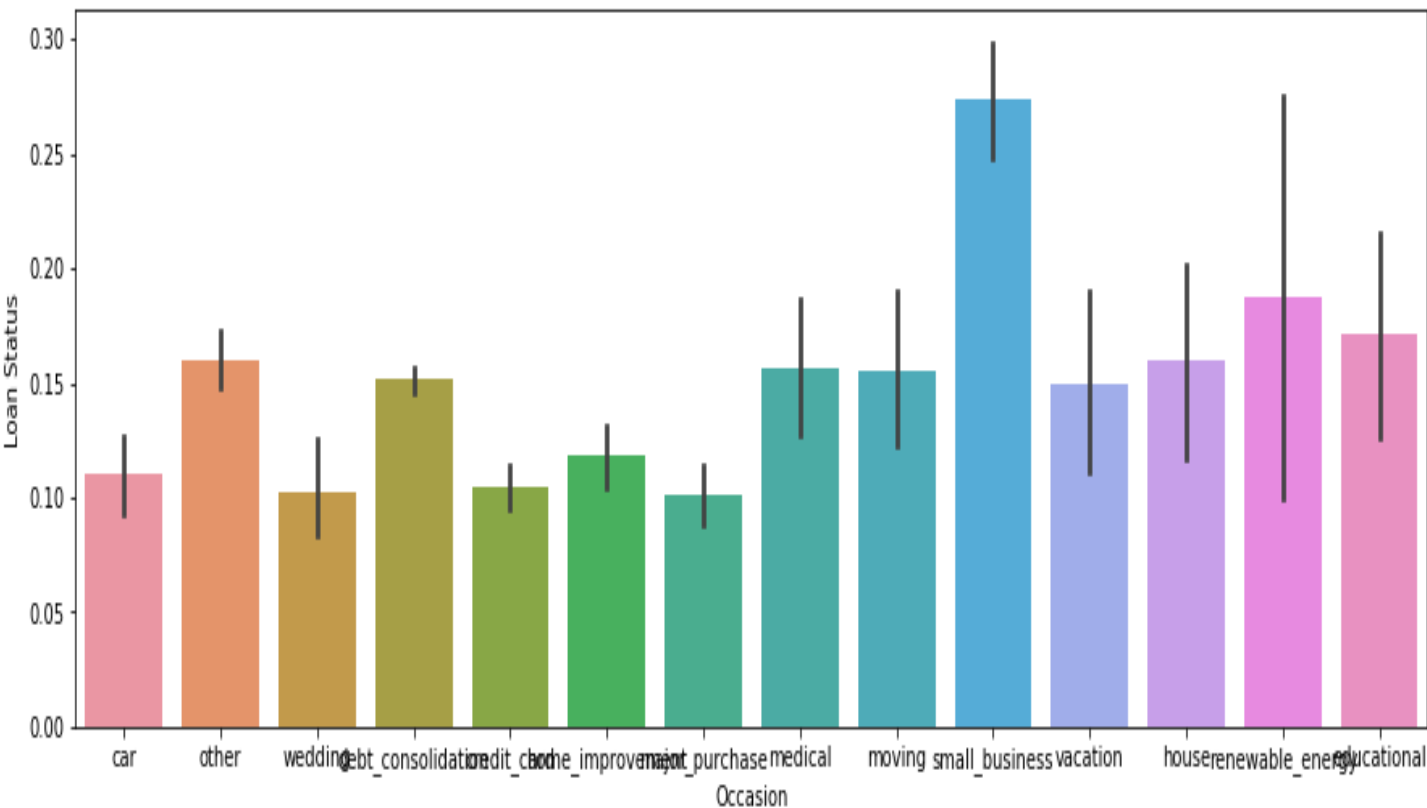
➤ **Recommendation(s)**

• Lending club should be very cautious while approving loans for "Small business" by employing stringent measures and reduce their loans for purposes of '*Small Business*' as 27% of the loans on that purpose are getting defaulted.
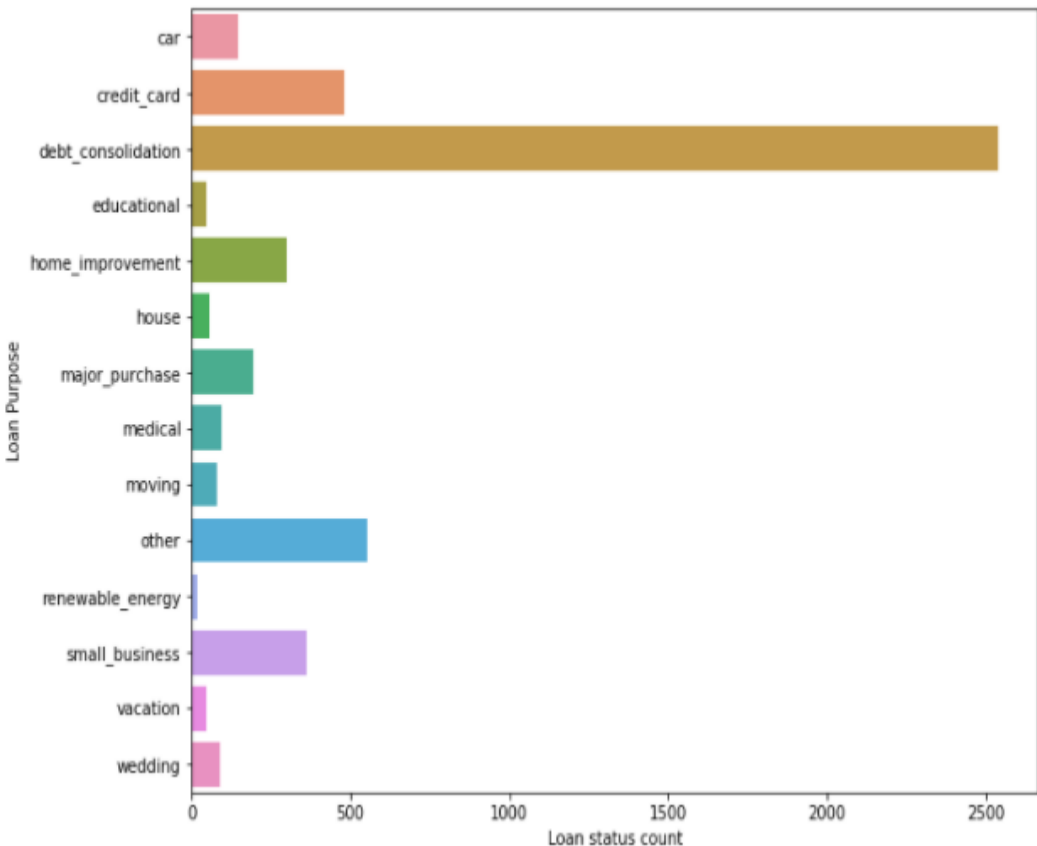
Employee Length Analysis

➤ O**bservation(s):**

• Majority of the customers are employees having 10 years of work experience.

• Higher the employee length, higher is the annual income.

➤ **Recommendation(s):**

• Lending club should be careful while providing loans to people who have higher experience; as higher experience individuals tend to take higher amounts exposing to higher risks of defaulting.

• Lending club should also be careful of providing loans to individuals who have <= 3 years of experience as their salaries will be less and their capacity to pay back will be less comparatively.
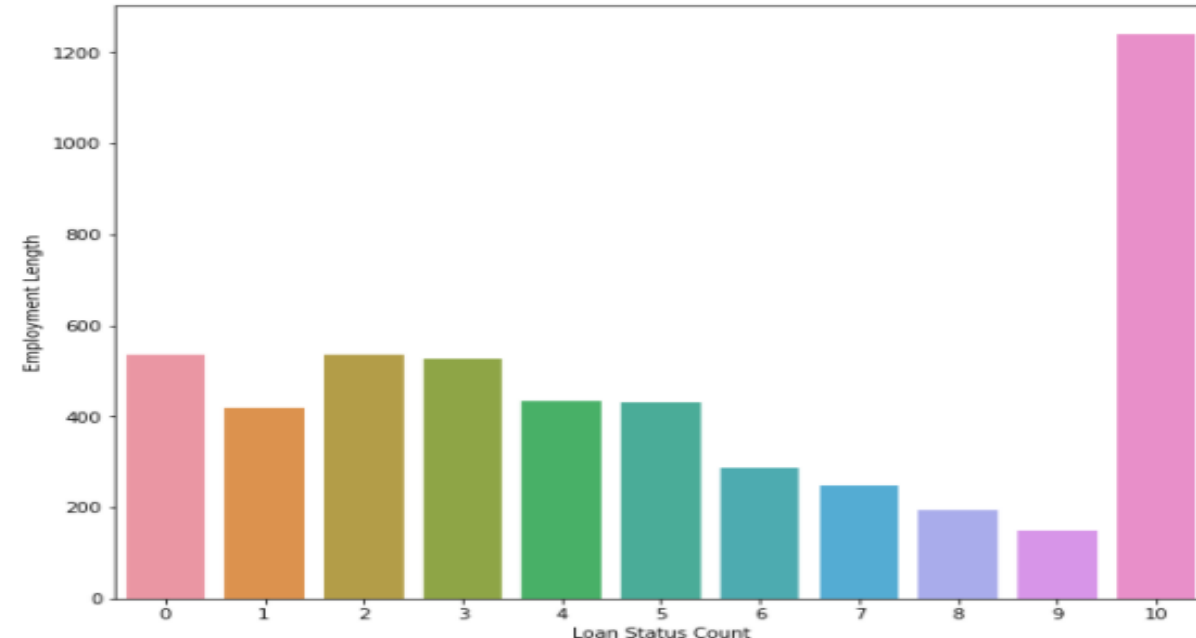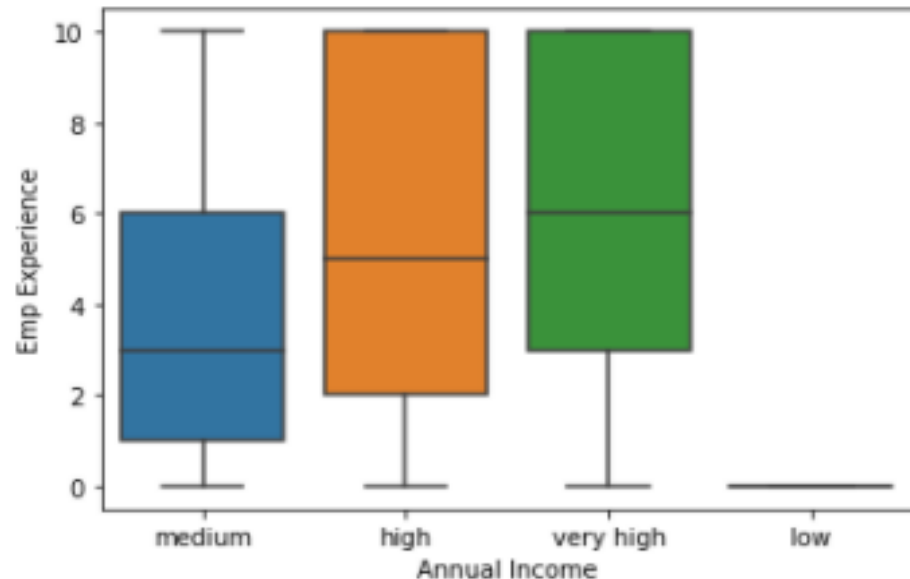
Annual Income Analysis

➢ **Observation(s):**

• Majority of the loans taken by customers have annual income between 40,000 – 80,000 USD.

• Customers with lower annual income (<20,000) are highly susceptible to defaulting.

➢ **Recommendation(s):**

• Lending club should do rigorous due diligence with Lower Annual Income bracket and Medium income bracket individuals as their dti is higher compared with other income brackets. Higher dti means higher risks of defaults as seen in the next slide.
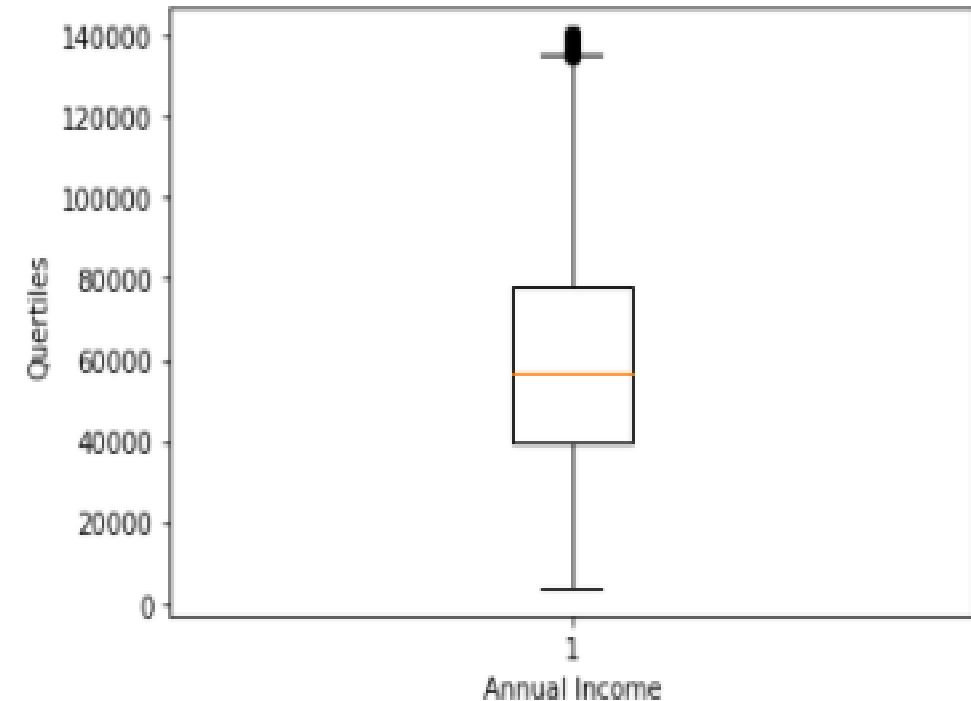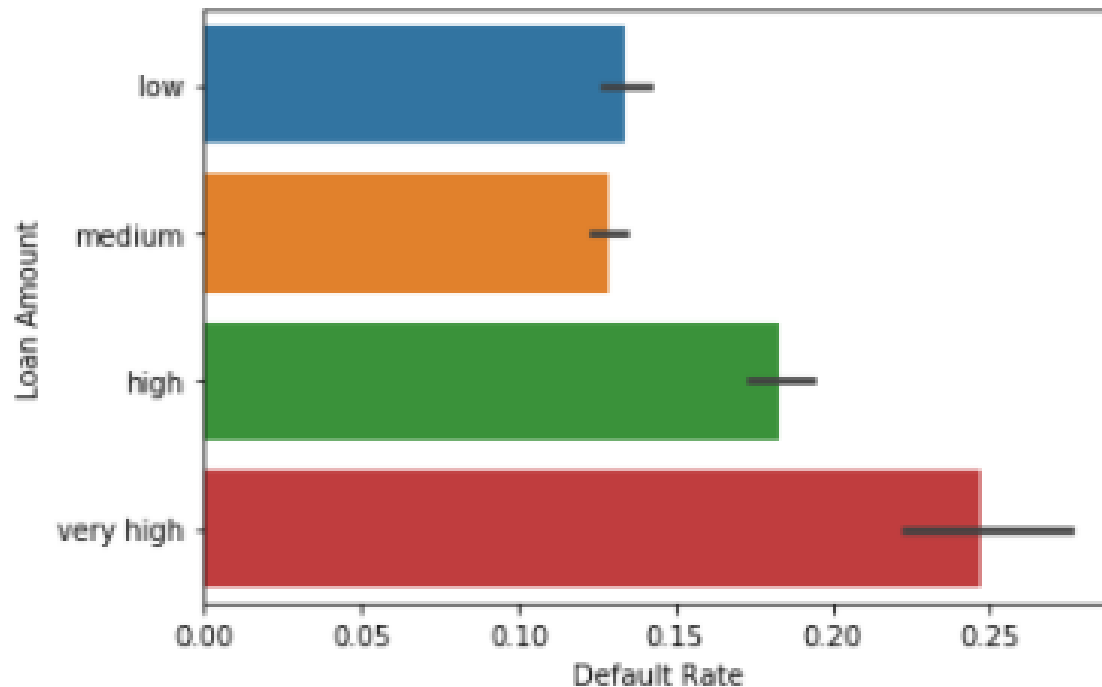
Debt-to-Income Ratio & Loan Amount Analysis

➢ **Observation(s):**

• Higher the debt-to-income ratio, higher are the tendency for loans to get default.

• Higher the loan amount, higher are the default rate. But, from the count perspective, loan amounts that are 15000 and higher are tending towards default.
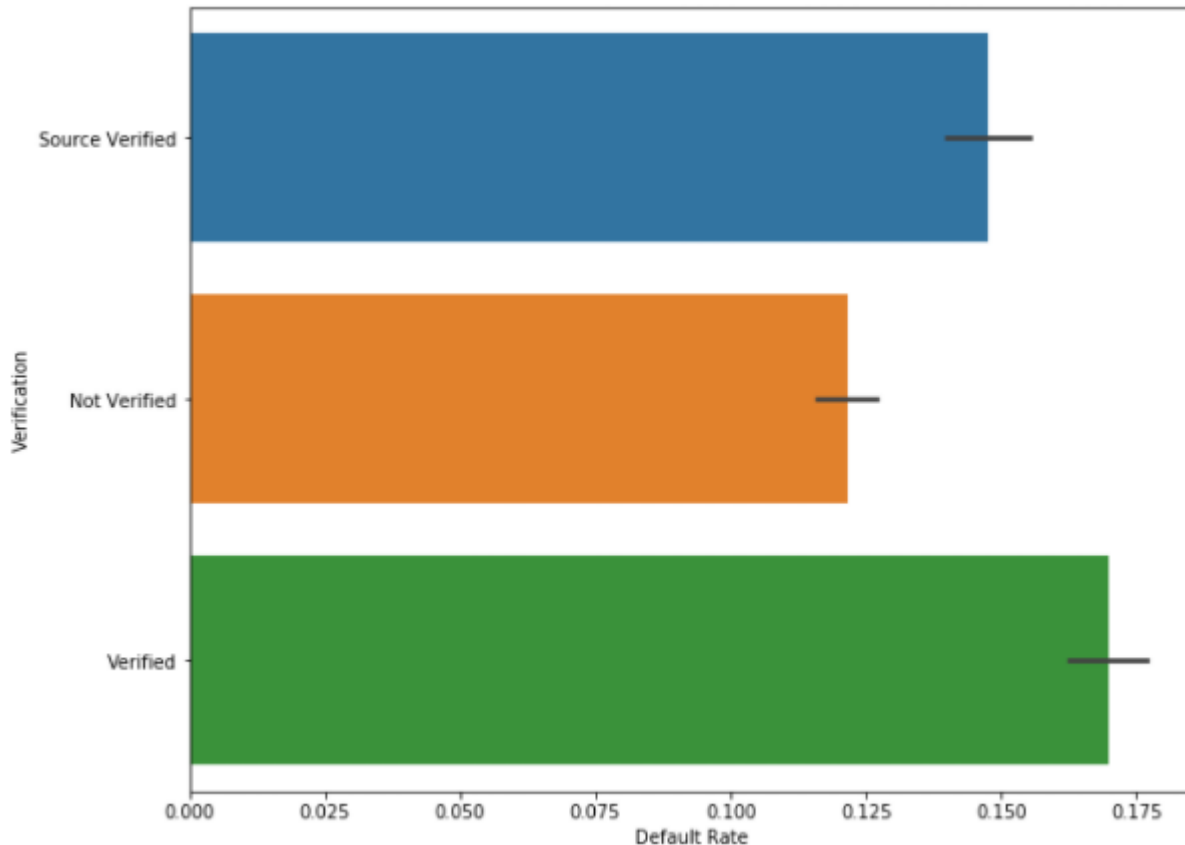
➢ Recommendation(s):

  Lending club should avoid providing loans to individuals with higher DTI as higher the Debt-to-Income ratio; higher are the chances of loans getting charged off.

• LC should consider loans which are less than 10000 and should be careful of approving loans which are more than 20000 since the charged off ratio's are higher.
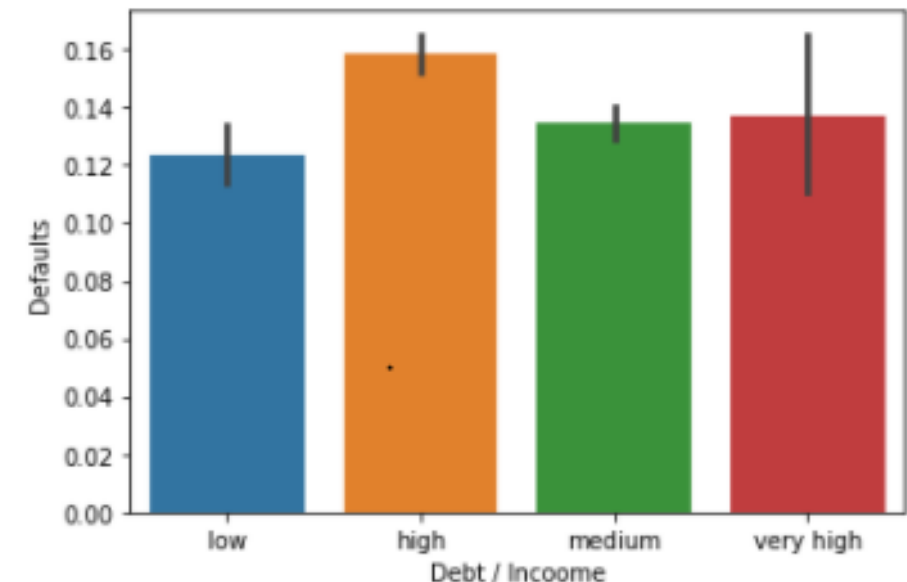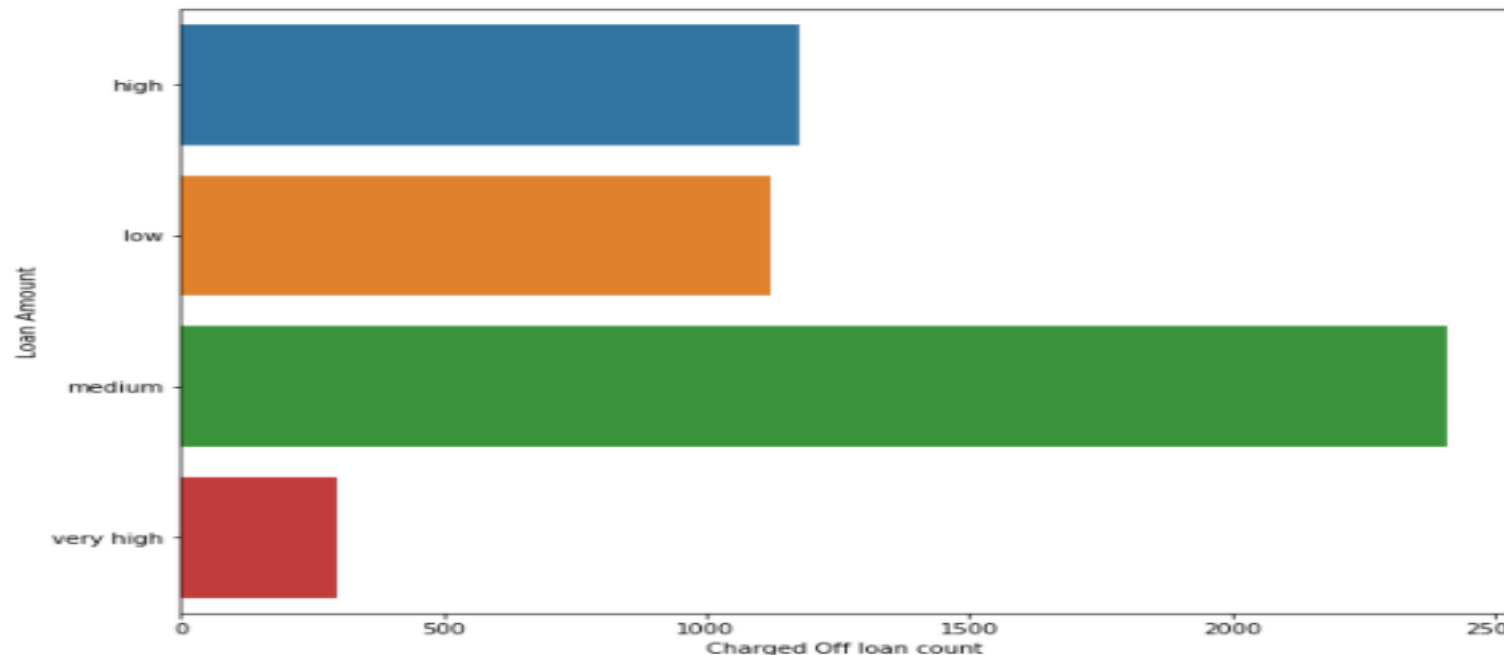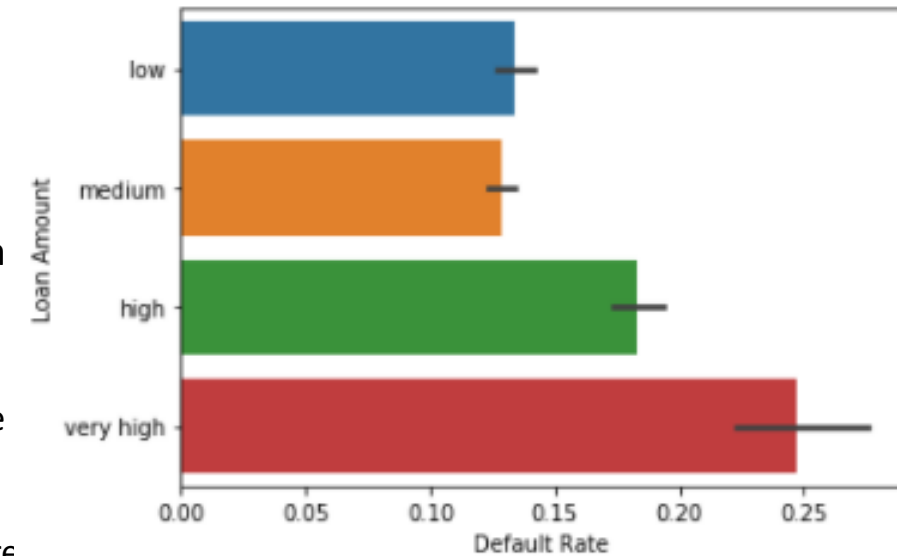
# Verification Status and Term of the loan Analysis

➢ **Observation(s):**

• Loans with verified status have higher default rate of 16.8%.

• Loans with term of 60 months have higher default rate of 25%

➢ **Recommendation(s):**

• Lending club should look at their process of verification to be more stringent. Although verified, their loans are defaulting at higher rate

• Lending club should look at lending more loans with term of 36 months and less.
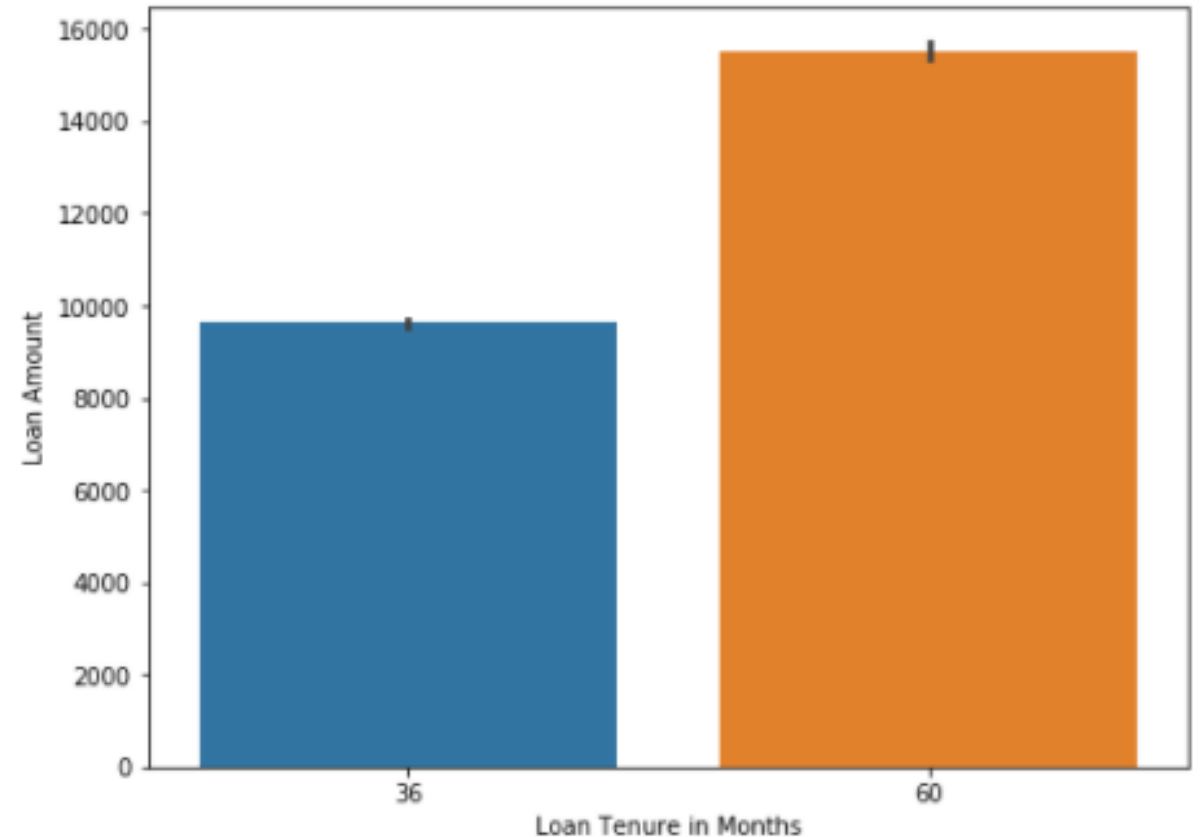
Home Ownership & Publicly Recorded Bankruptcies Analysis

➢ **Observation(s):**

• Borrowers who are renting are defaulting more.

• Borrowers with higher public record of bankruptcies have tendency to default more.

➢ **Recommendation(s)**

• Lending Club should perform more due diligence while providing loans to borrowers who are renting and doesn't own or mortgage.

• Lending Club should avoid lending to individuals who have more publicly recorded bankruptcies.

Loan Defaults against Home ownership
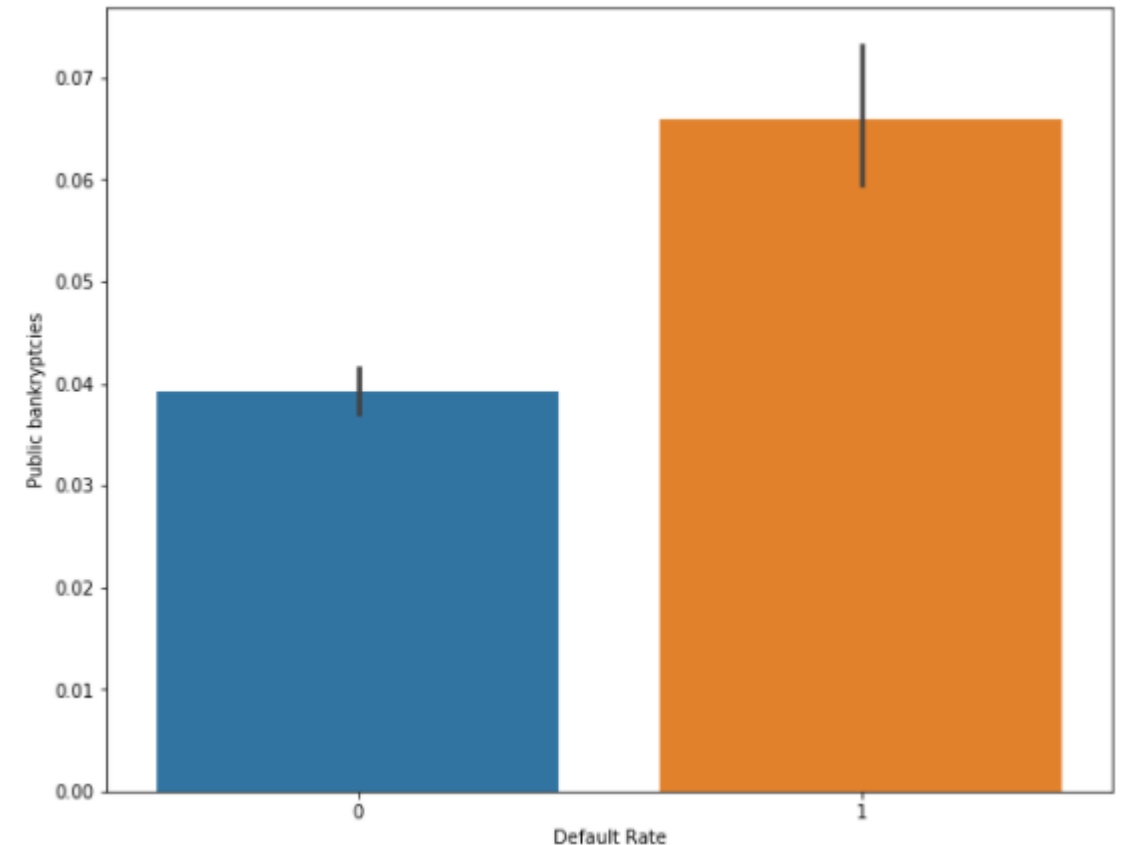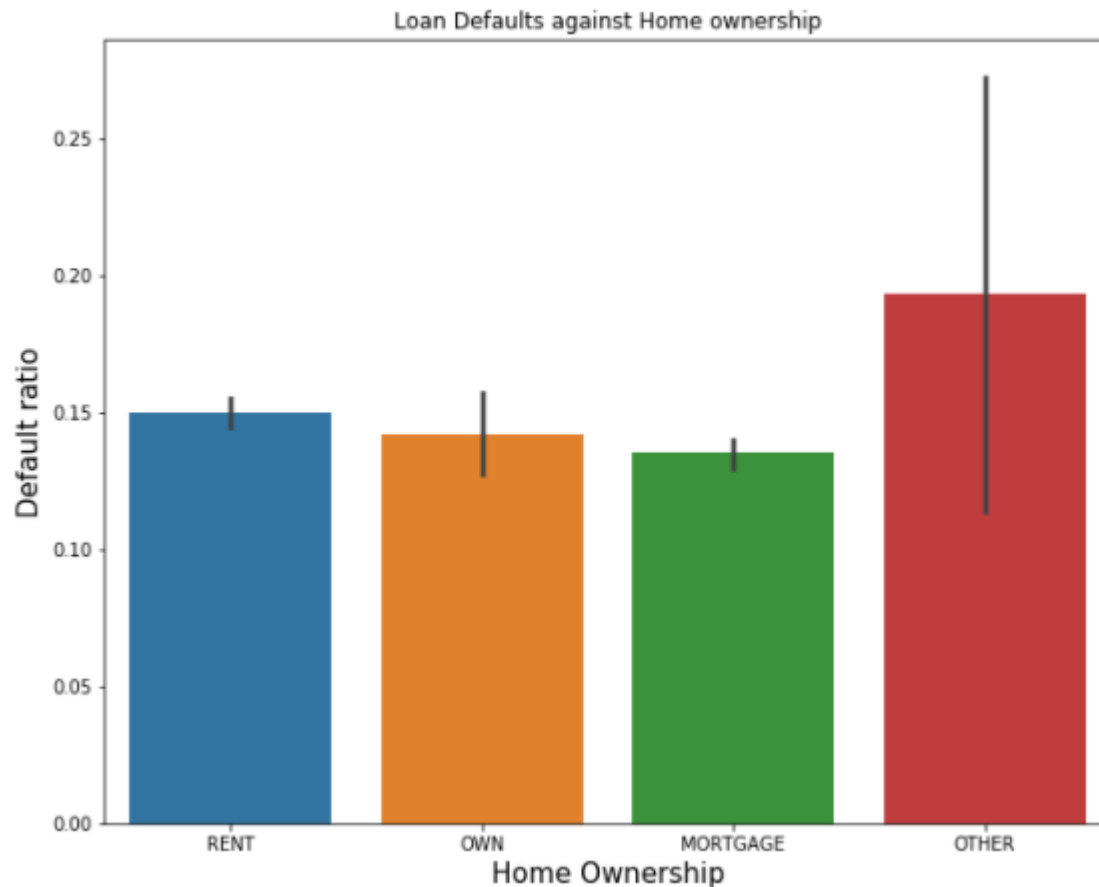
Conclusions

➤ Based on the Exploratory data analysis of Lending Case Study, we can conclude that –

    ➤ Loans having higher interest rate have higher tendency to default. Interest rate is also correlated highly with term, loan amount. Hence, Lending club should check all these factors before approving any loan.

    ➤ Lending Club should be very cautious while approving loans for small businesses & debt consolidation and borrowers living in Rent and who do not own house.

    ➤ Lending club should be very cautious while approving loans with low grade and should also improve their verification process as majority of loans that defaulted in the past were verified internally by Lending Club.