

Lending Club

Case Study Assignment

By Karthik Jayaraman
May 23rd 2023

Lending Club Case Study

Business Understanding

- Consumer Financing:
 - Business or retailer offering customer financing options to its customers using either their own funds or the funds of a lending company or bank.
- Consumer Finance Company:
 - A consumer finance company is a non-bank company that provides financial products to individuals.
- Lending Club Company:
 - LendingClub Corp (LendingClub) is an online marketplace that connects borrowers and investors. The company's product offering for borrowers includes personal loans, patient finance loans, education loans, auto refinancing loans, and small business loans.

Lending Club Case Study

Business Process Understanding - Risks

When Lending company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

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

- Risk 1:
 - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- Risk 2:
 - 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

Lending Club Case Study

Objective & Reasoning

- 1) Understand how consumer attributes and loan attributes influence the tendency of loan payment default.
- 2) Identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan or lending (to risky applicants) at a higher interest rate. Customers labelled as 'charged-off' are the 'defaulters'
- 3) Why?
 - . Lending loans to 'risky' applicants is the largest source of financial loss.
 - . Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.
 - . Company wants to understand the driving factors (or driver variables) behind loan default.
 - . Company can utilize this knowledge for its portfolio and risk assessment.

Lending Club Case Study

Analysis approach

- 1) Understand the Loan Data Dictionary and research on domain knowledge.
- 2) Perform Exploratory Data analysis on Loan Dataset provided using Jupyter Python notebook and essential libraries. (Numpy, Pandas,..)
- 3) Clean, Fix & Standardize dataset. (Also use Inter-Quartile range)
- 4) Eliminate variables which are not useful to determine drivers for defaulters.
- 5) Create Derived metrics & Segmented variables as per need.
- 5) Perform Univariate analysis using Categorical & Numeric variables.
- 6) Perform Bivariate analysis using Categorical & Numeric variables.
- 7) Summarize observations & results.

Lending Club Case Study

Chosen variables for analysis

1) Categorical:

- .term: The number of payments on the loan
- .grade: Lending Club (LC) assigned loan grade
- .emp_length: Employment length in years.
- .home_ownership: Home ownership status provided by the borrower during registration
 - Values: RENT, OWN, MORTGAGE, OTHER
- .verification_status: Indicates if income was verified by LC, not verified, or if the income source was verified
- .issue_d: The month which the loan was funded
- .loan_status: Current status of the loan
- .earliest_cr_line: The month the borrower's earliest reported credit line was opened
- .purpose: A category provided by the borrower for the loan request.
- .addr_state: The state provided by the borrower in the loan application

Lending Club Case Study

Chosen variables for analysis

2) Numeric

.loan_amnt: The listed amount of the loan applied for by the borrower.

.funded_amnt: The total amount committed to that loan at that point in time.

.funded_amnt_inv: The total amount committed by investors for that loan at that point in time.

.int_rate: Interest Rate on the loan

.installment: The monthly payment owed by the borrower if the loan originates.

.annual_inc: The self-reported annual income provided by the borrower during registration.

.dti: A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

.inq_last_6mths: The number of inquiries in past 6 months (excluding auto and mortgage inquiries)

.open_acc: The number of open credit lines in the borrower's credit file.

.pub_rec: Number of derogatory public records

.revol_util: Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.

.total_acc: The total number of credit lines currently in the borrower's credit file

.pub_rec_bankruptcies: Number of public record bankruptcies

Lending Club Case Study

Chosen variables for analysis

3) Derived metrics

.int_rate_grp: Interest Rate bins ('5%-8%','8%-11%','11%-14%','14%-18%','18%-21%','21%-24%')

.annual_inc_grp: Annual income bins (in USD)('3.8k-27.5k','27.5k-51k','51k-74.5k','74.5k-98k','98k-121.5k','121.5k-145k')

.installment_grp: Monthly Loan installment bins (in USD) ('14-231','231-446','446-660','660-875','875-1090','1090-1305')

.loan_amnt_grp: Loan applied bins (in USD)
('466-6.25k','6.25k-12k','12k-17.75k','17.75k-23.5k','23.5k-29.25k','29.25k-35k'

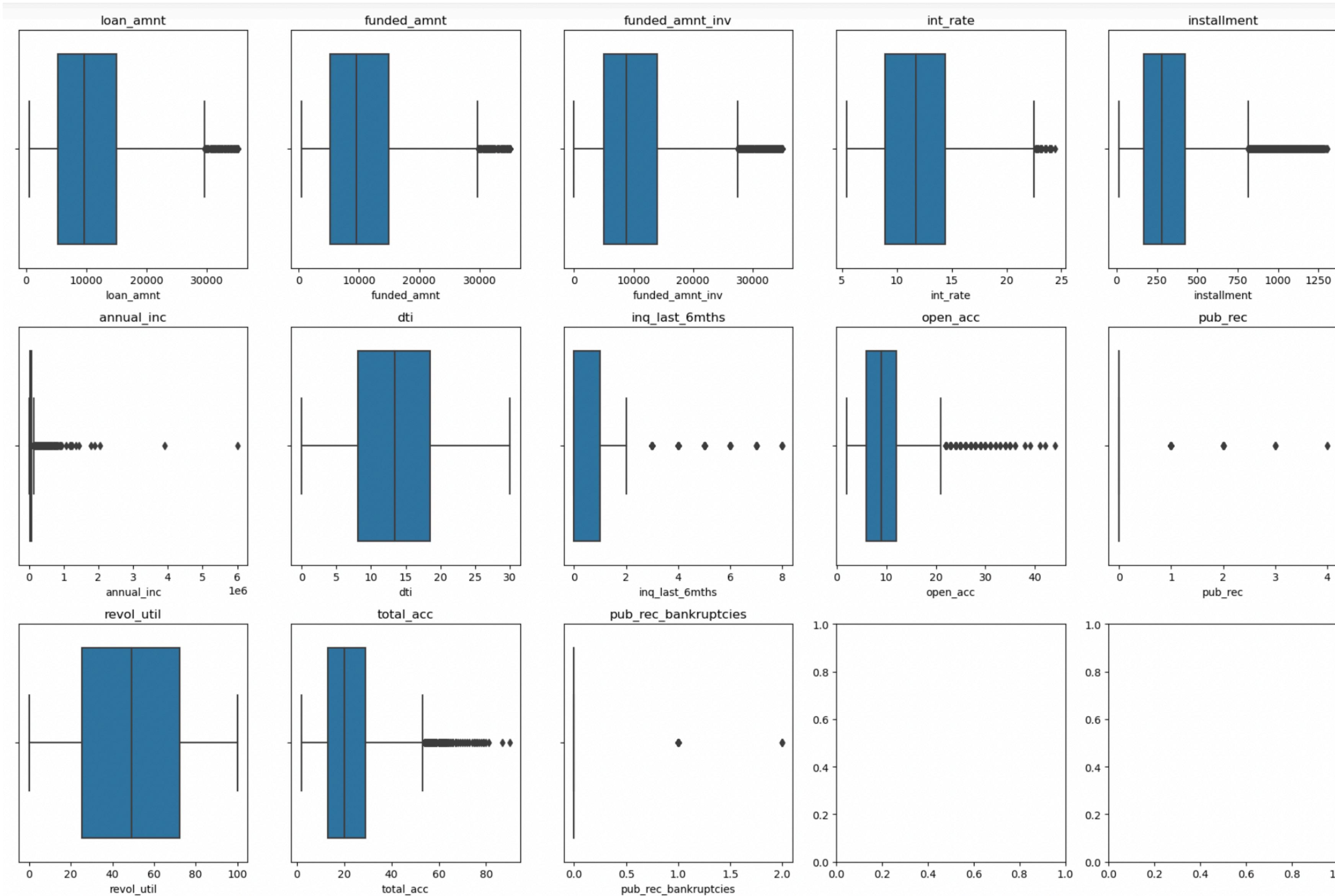
- . Extract issue year & month variables separately.
- . Extract earliest_cr_line year & month variables separately

4) Categorical variables encoded into numeric (for correlation analysis)

. home_ownership_num, grade_num, purpose_num, int_rate_grp_num, annual_inc_grp_num,
installment_grp_num, loan_amnt_grp_num

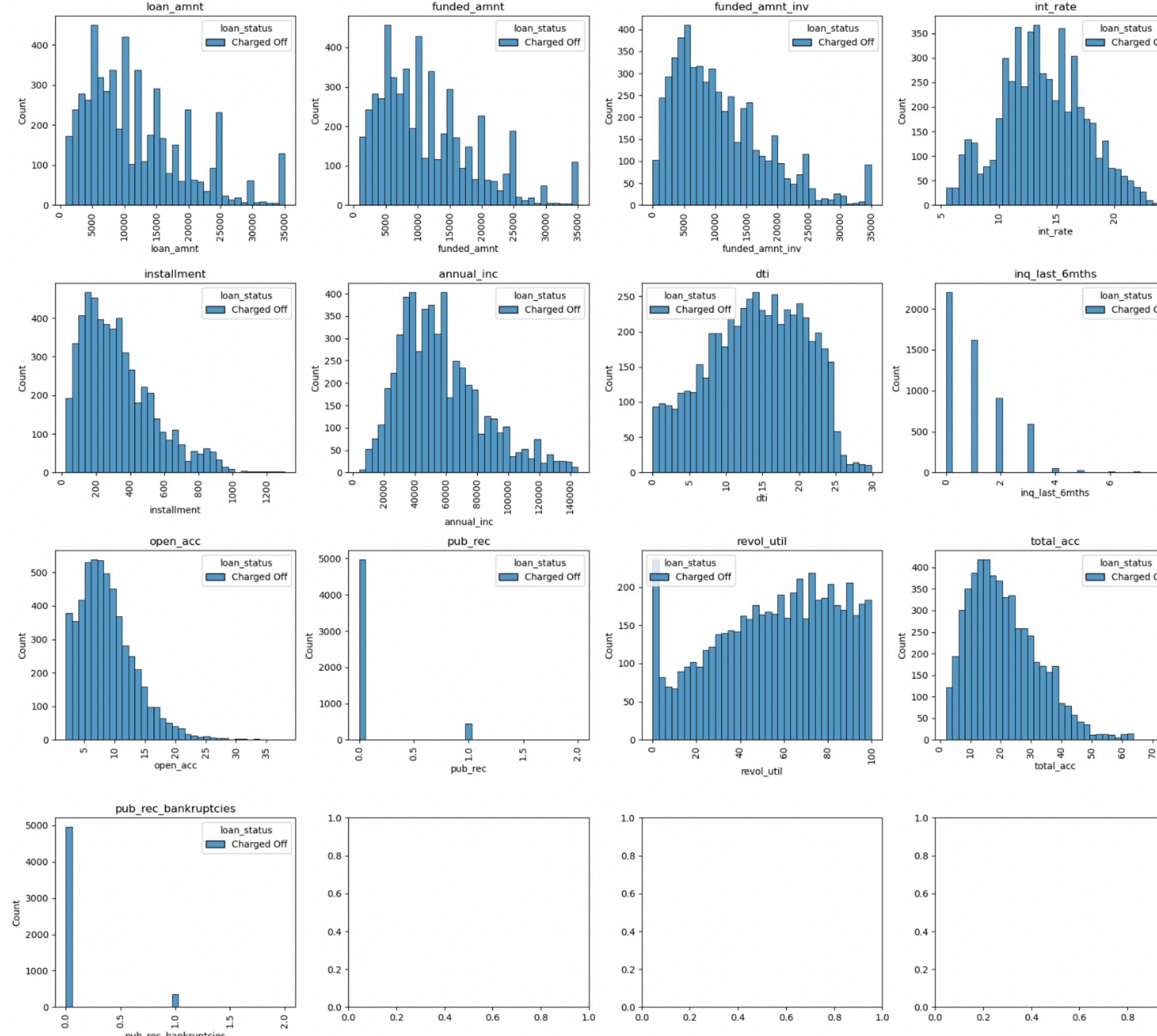
Lending Club Case Study

Outlier detection & removal



Lending Club Case Study

Observation 1: Univariate Analysis (Numeric) on Defaulters

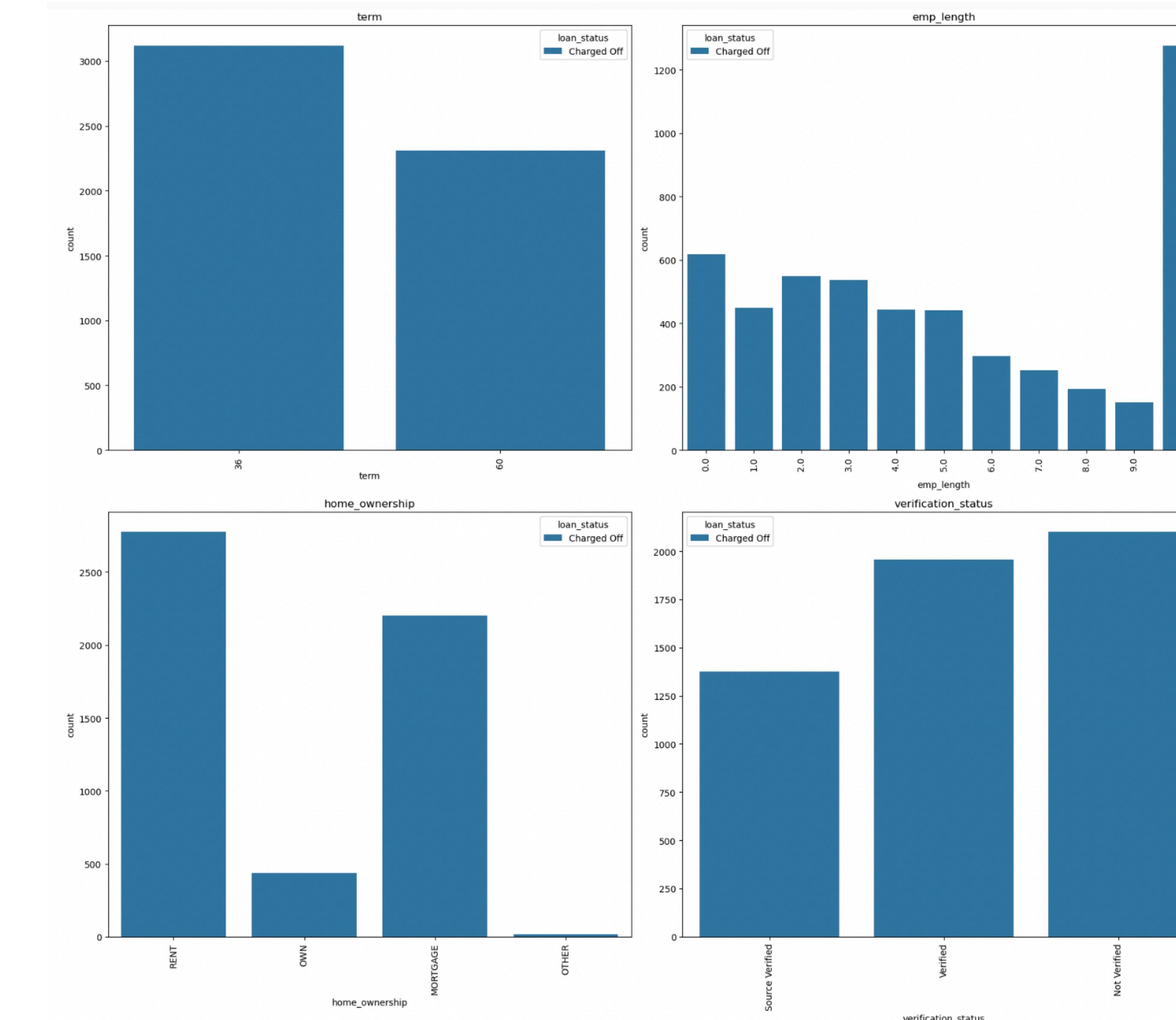


The Number of Loan defaulters are higher in following cases:

- The Funded amount is in range 5000 to 12000
- The Funded amount by investor is in range 3000 to 13000
- The Interest Rate on the loan is between 10% to 16%
- The Monthly installment is between 100 & 400
- The Annual income is between 30000 & 60000
- The DTI (Debt-to-Income) ratio is between 0 & 25

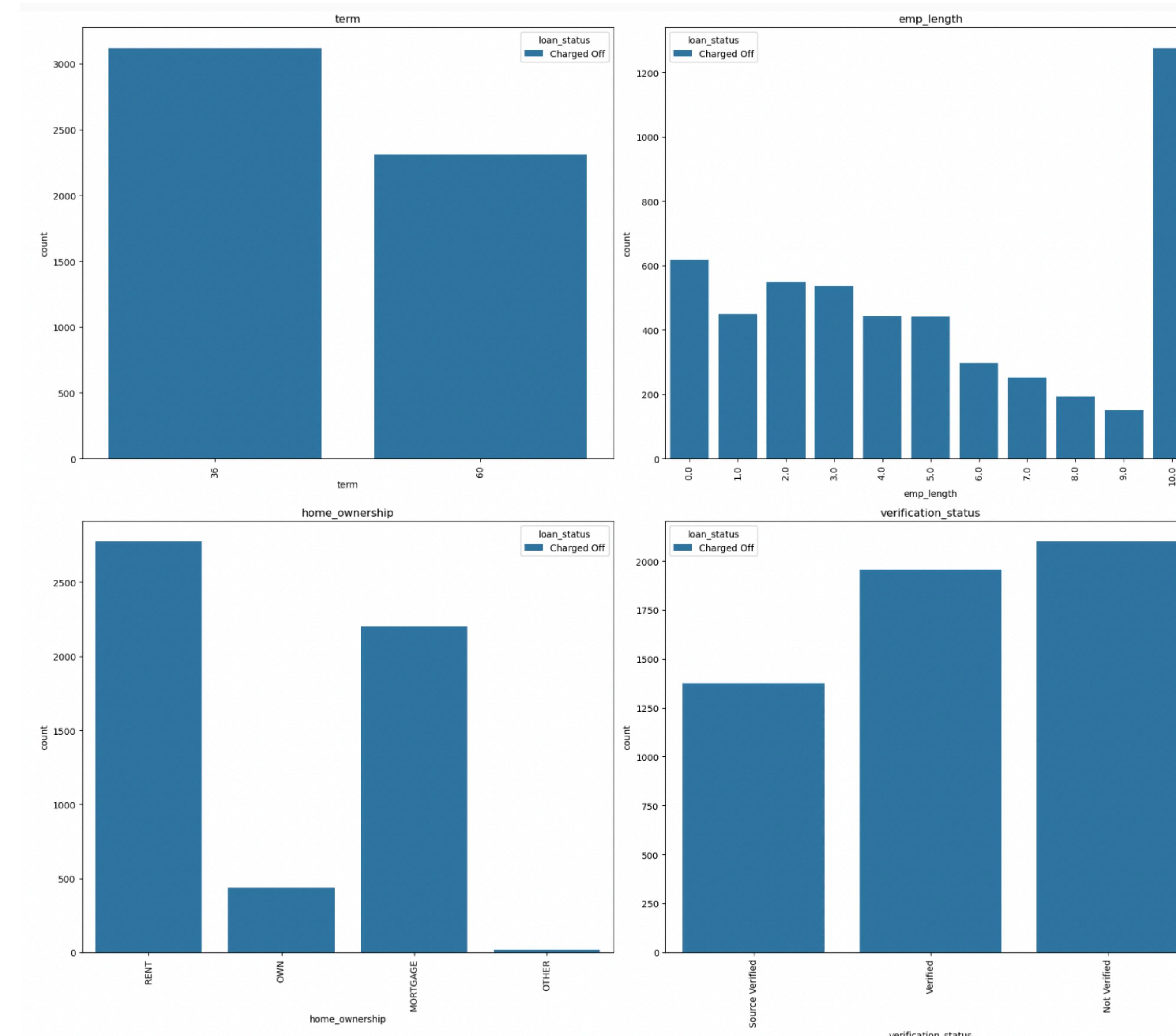
Lending Club Case Study

Observation 2: Univariate Analysis (Numeric) on Defaulters



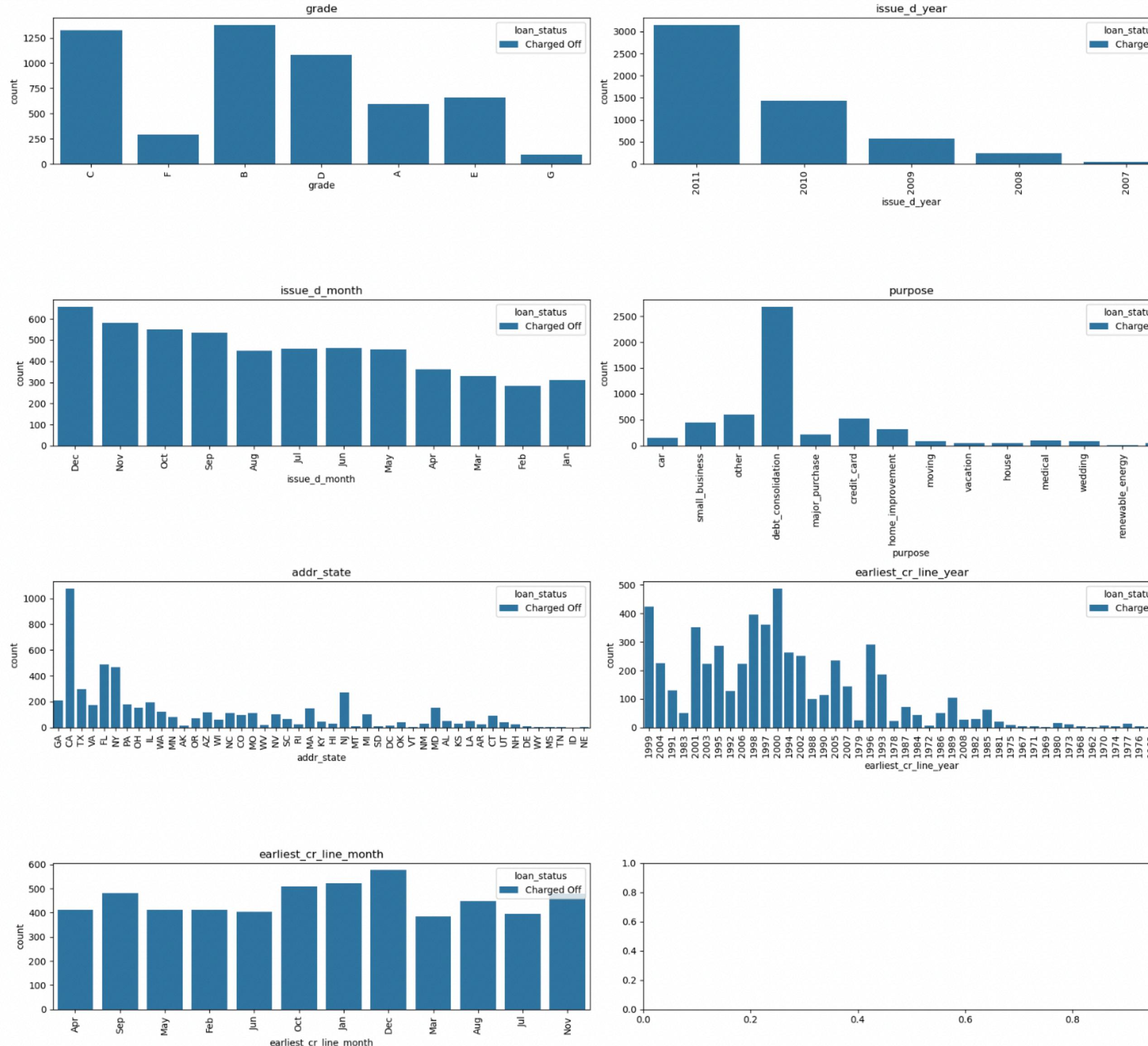
Lending Club Case Study

Observation 2: Univariate Analysis (Categorical) on Defaulters



Lending Club Case Study

Observation 2: Univariate Analysis (Categorical) on Defaulters

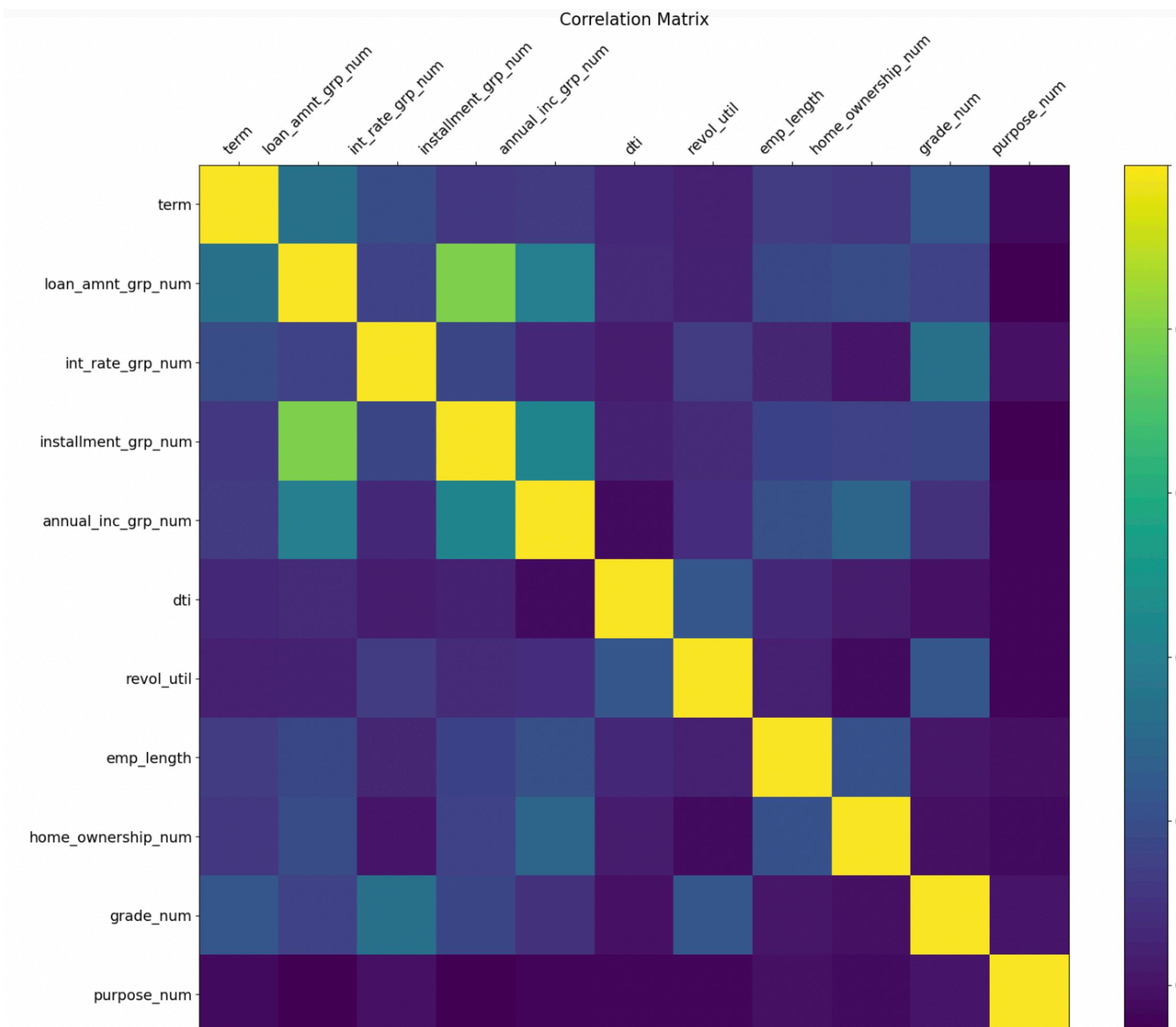


The Number of Loan defaulters are higher in following cases:

- The Loan payment term is 36 months
- The Employment length is 10 years or more
- The home ownership status provided by the borrower during registration is 'RENT'
- The LC assigned loan grade is 'B' or 'C'
- The issue month which the loan was funded was 'Sep', 'Oct', 'Nov' & 'Dec'
- The issue year which the loan was funded was 2011
- The category provided by the borrower for the loan request was 'Debt Consolidation'
- The state provided by the borrower in the loan application was 'California', 'Florida' & 'New York'
- The month the borrower's earliest reported credit line was opened was 'Sep', 'Oct', 'Nov' & 'Dec'
- The year the borrower's earliest reported credit line was opened 1999 or 2000

Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Correlation)



There is strong Positive Correlation between:

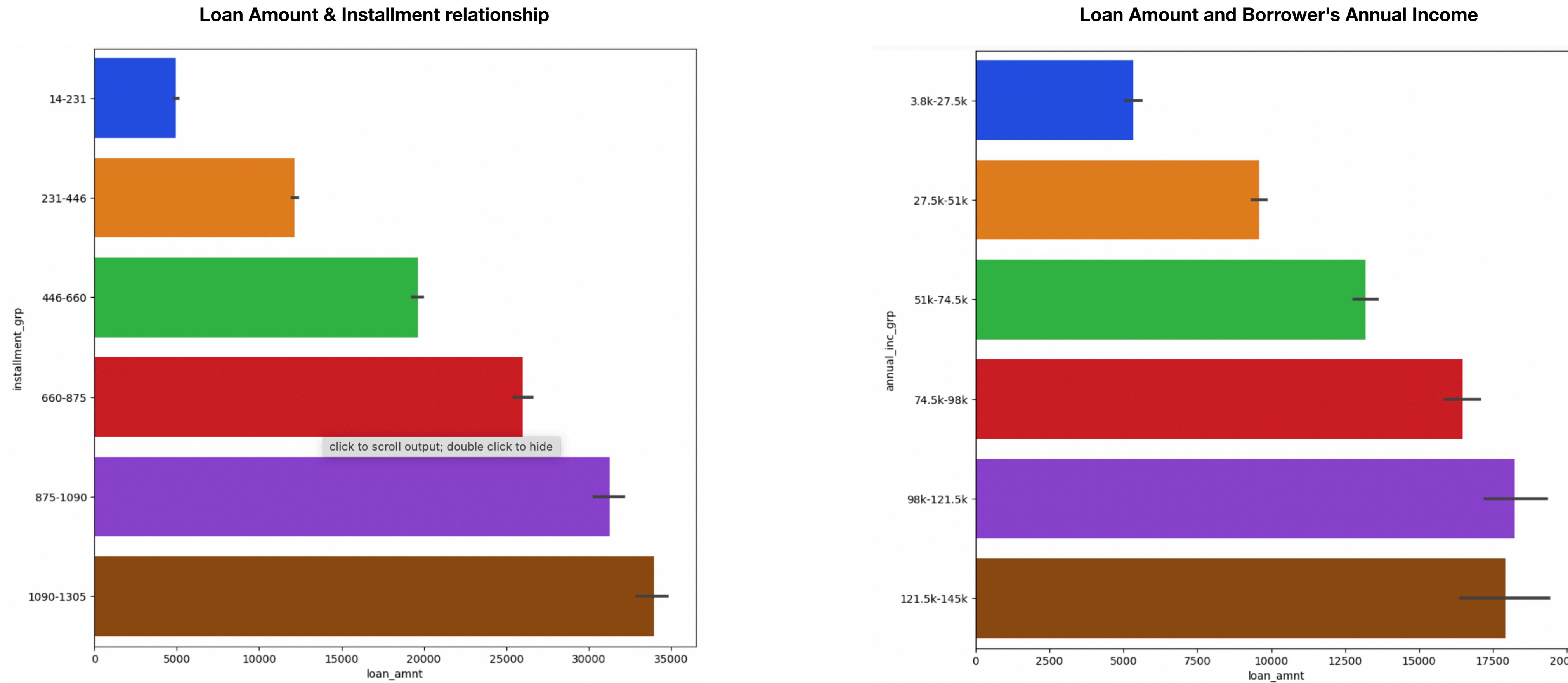
- a. Loan Amount and Installment (monthly payment owed by borrower)
- b. Loan Interest Rate and LC assigned Loan Grade

There is some Positive Correlation between:

- a. Loan Term (number of payments) and Loan Interest Rate
- b. Loan Amount and Borrower's Annual Income
- c. Loan Amount and Loan Interest Rate
- d. Installment and Borrower's Annual Income
- e. DTI (Debt-to-Income ratio) and Loan Amount
- f. Employee Length and Loan Amount
- g. Employee Length and Home ownership category

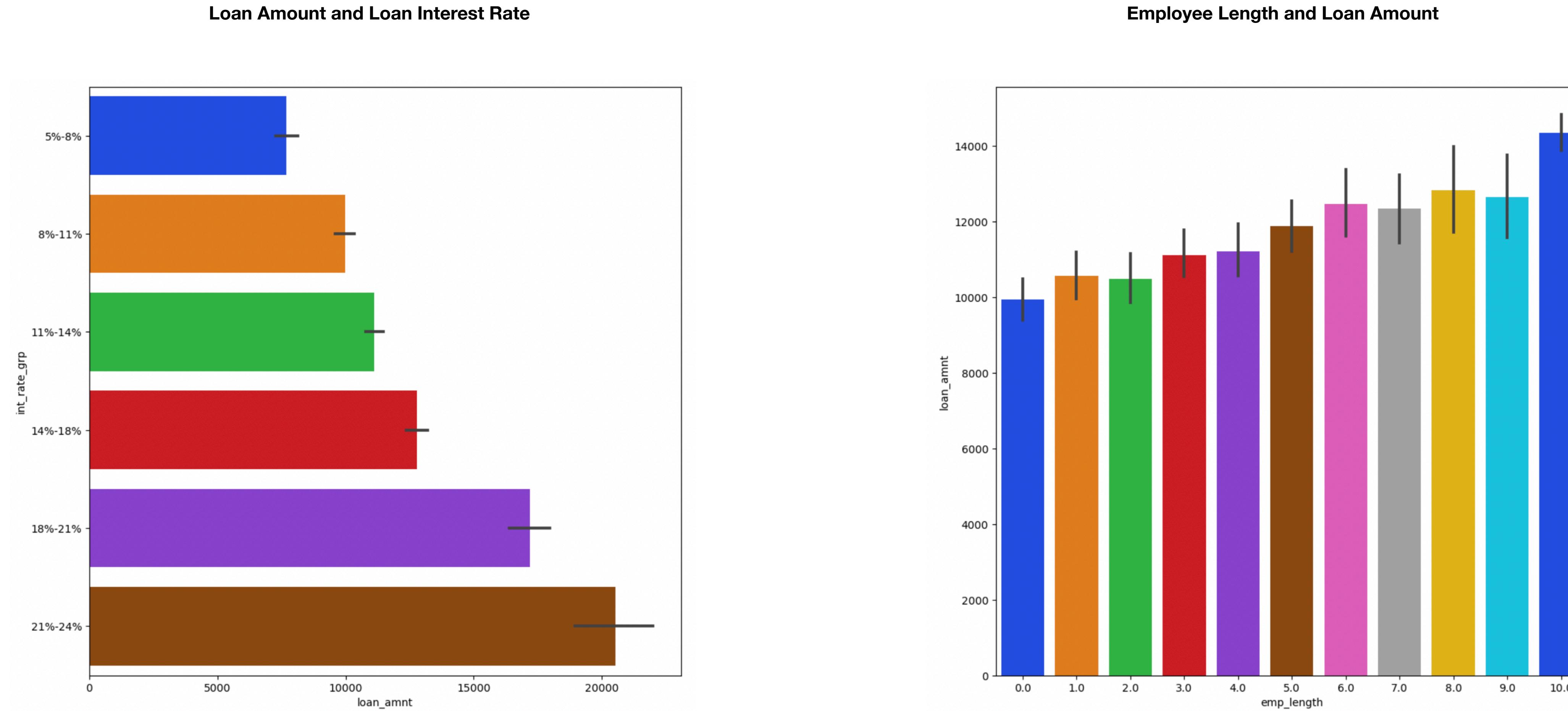
Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



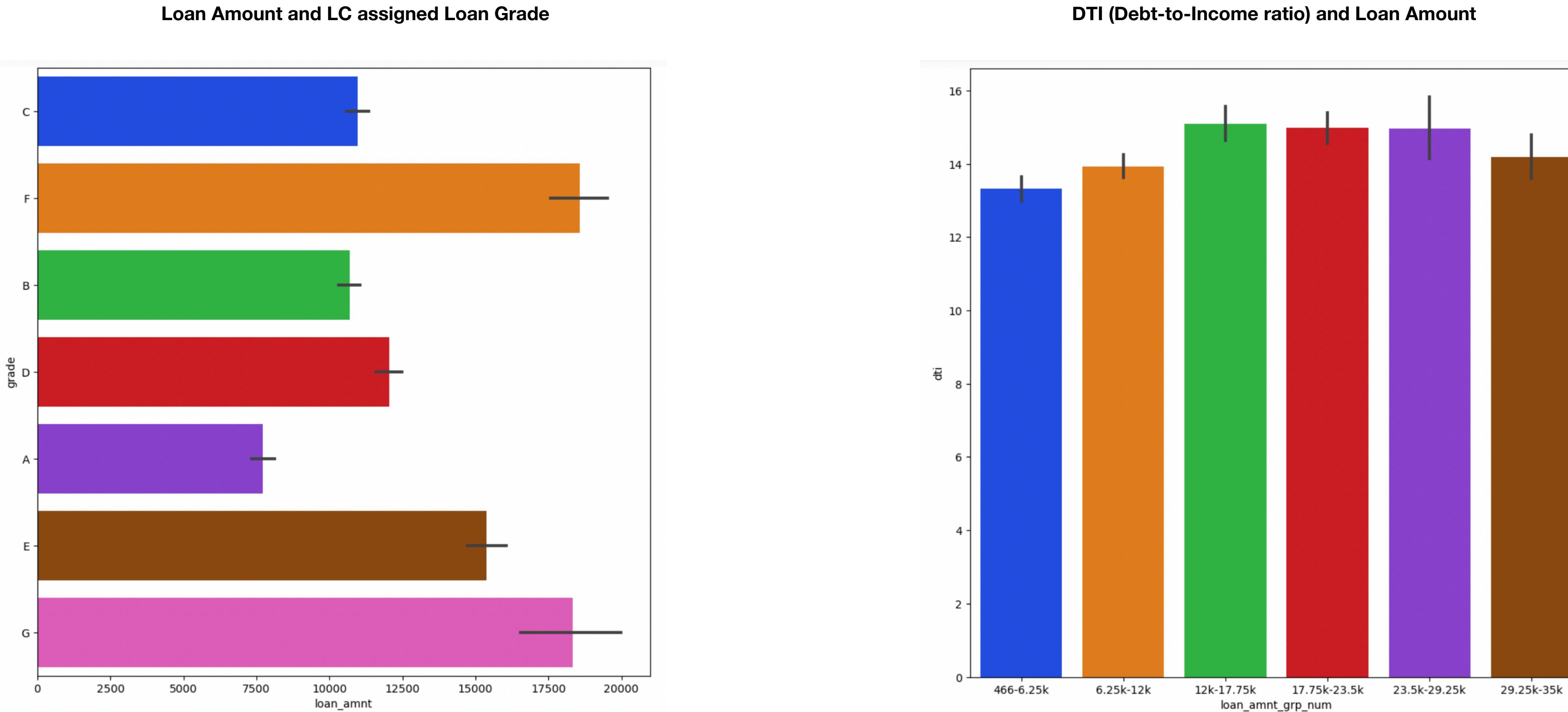
Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



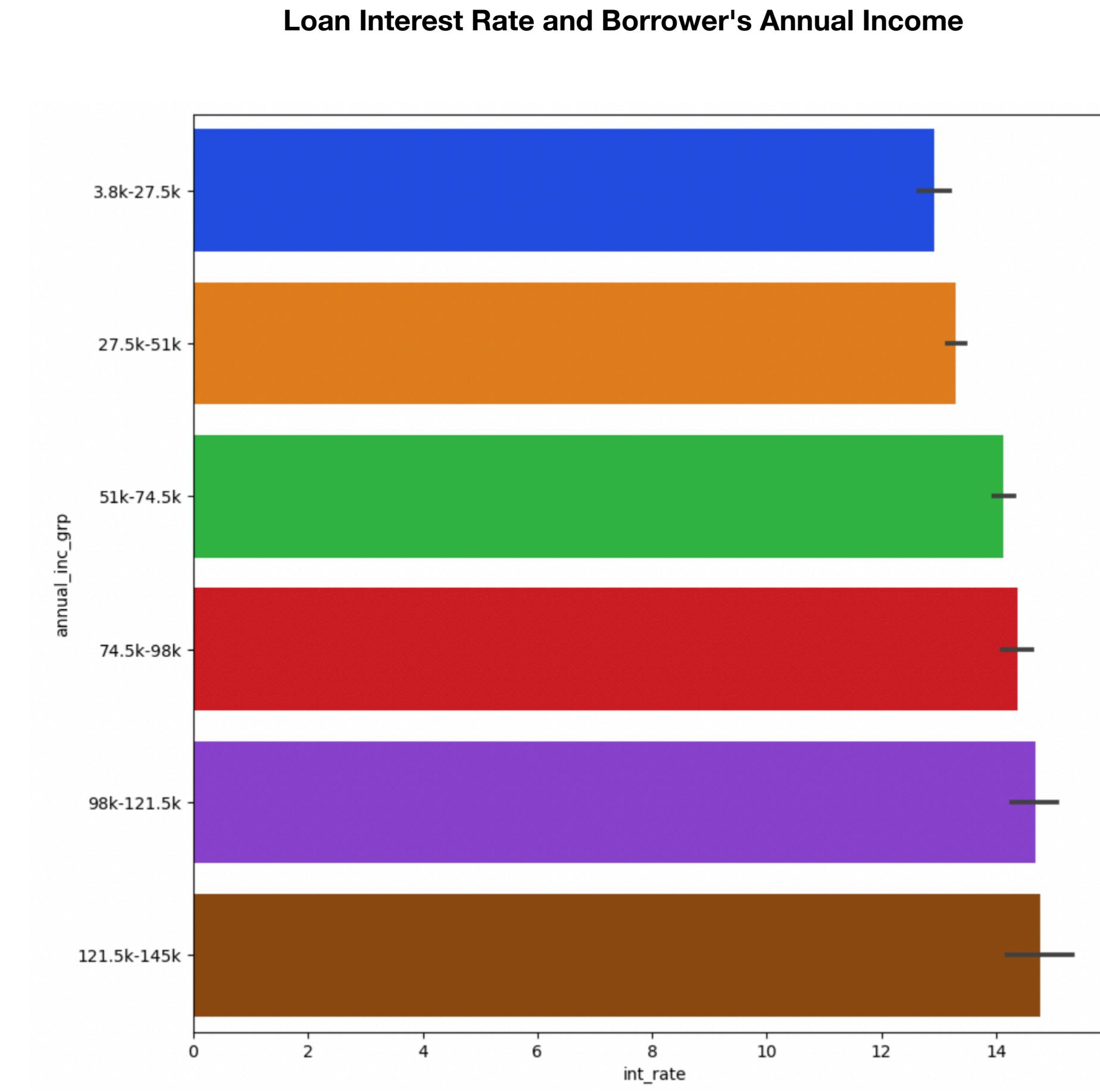
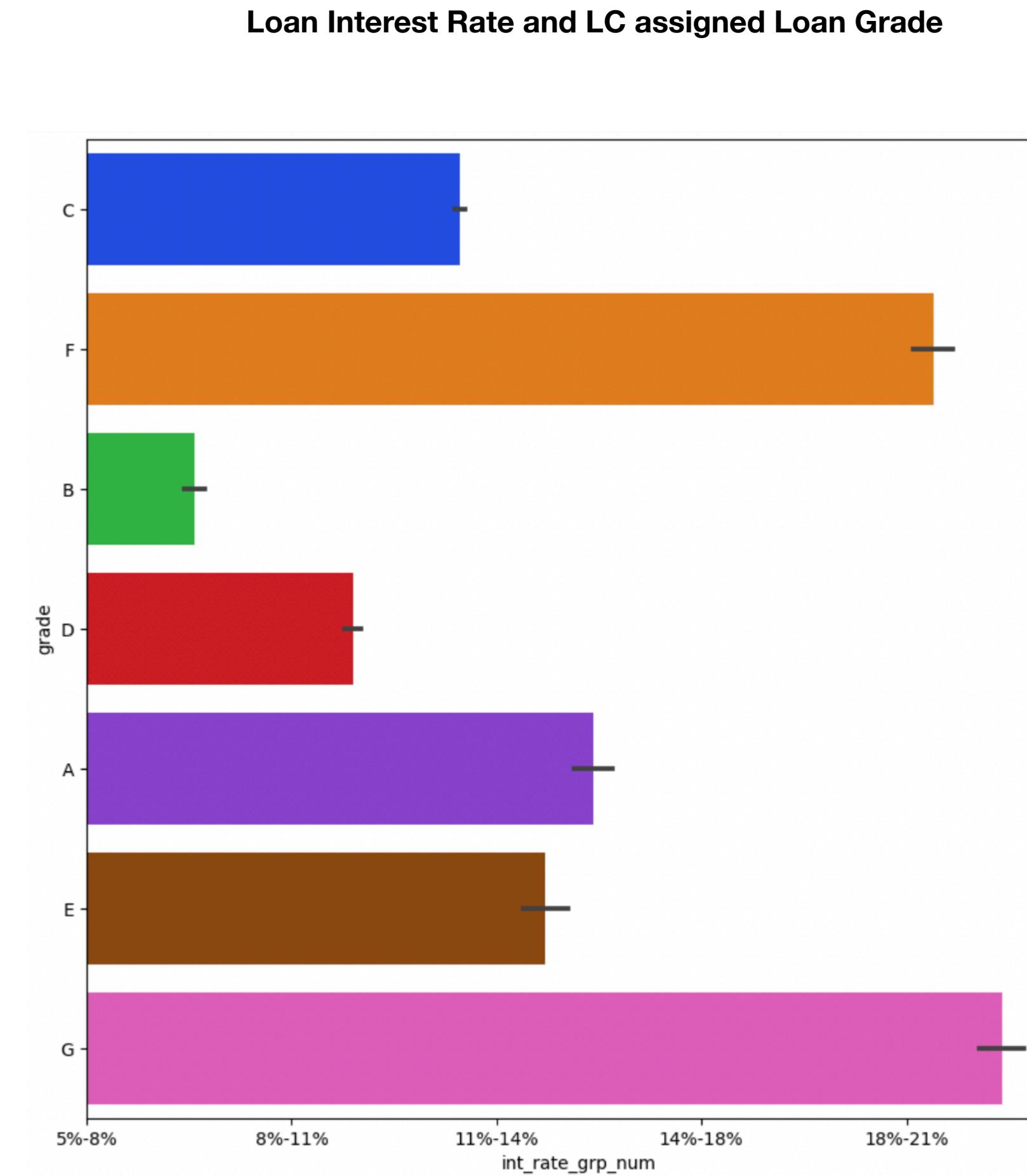
Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



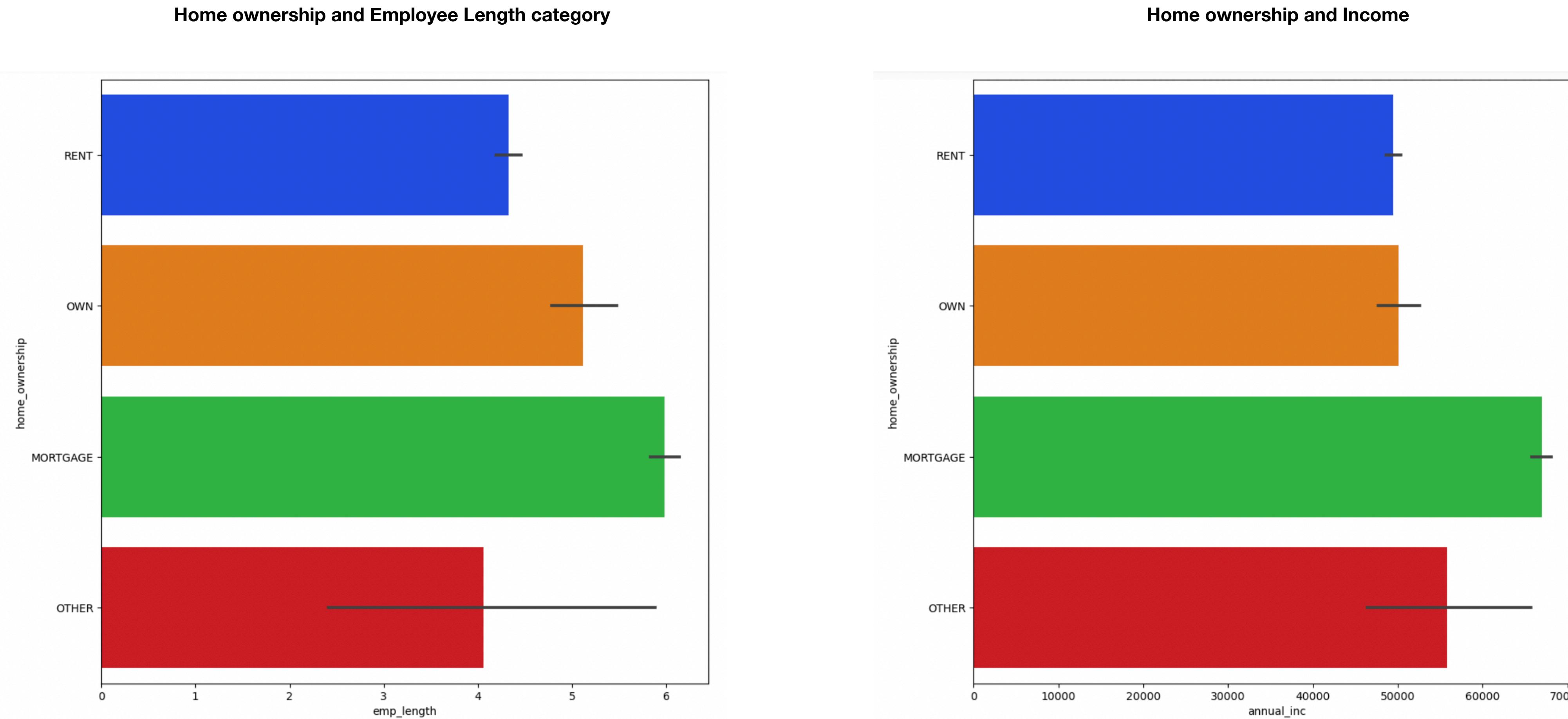
Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



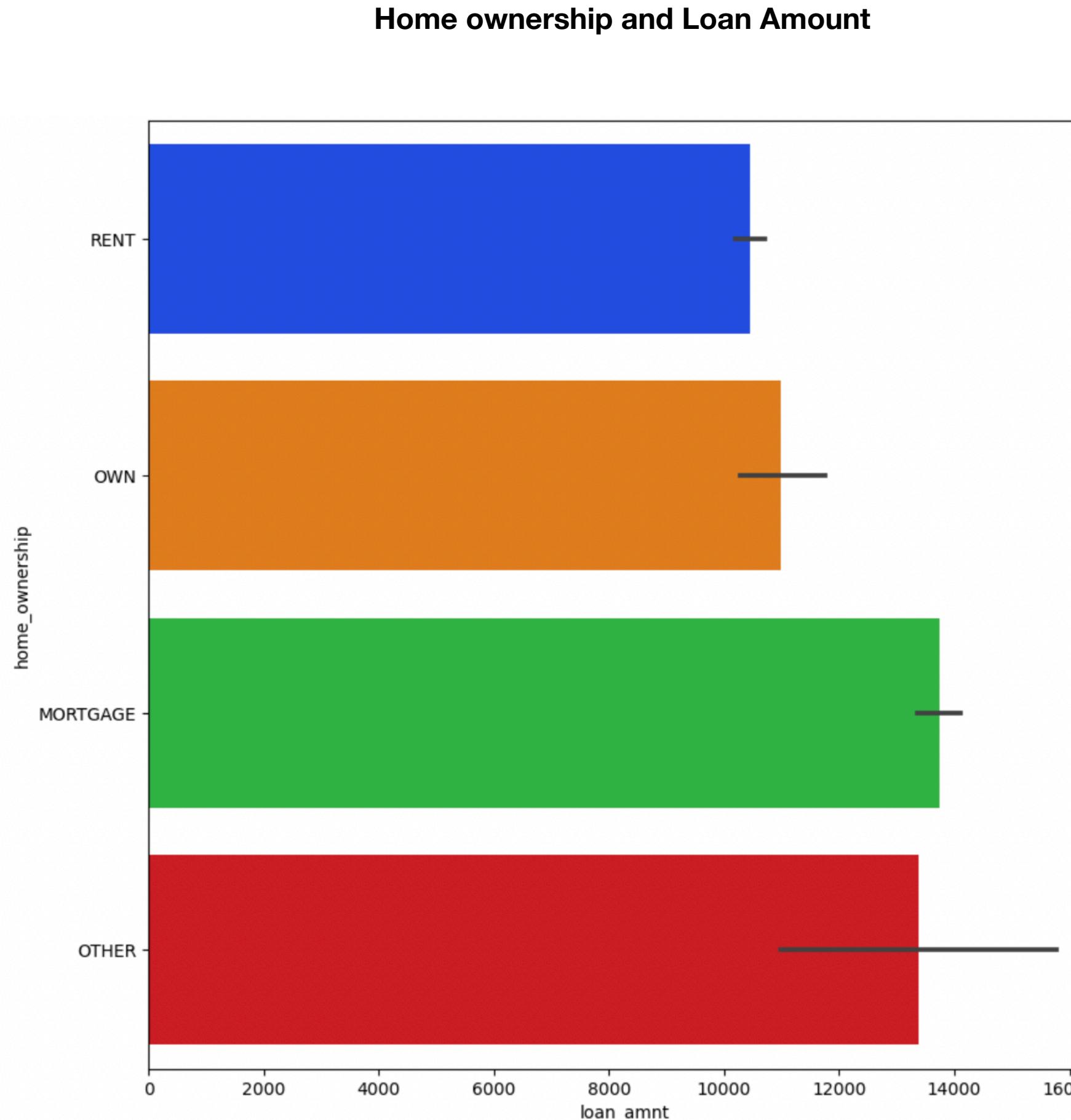
Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



Lending Club Case Study

Observation 3: Bivariate Analysis on Defaulters (Relationship)



When analyzing the Bivariate patterns, the chances of Loan default or charged off occurs in following combinations:

- a. Loan amount is high (around 35000) and Monthly installment is high (1000 ~ 1300)
- b. Loan amount is high (around 20000) and Interest Rate is high (21%-24%)
- c. Loan amount is high (around 14000) and Employee Tenure is high (10+ years)
- d. Loan amount is between 12k to 30k and DTI is > 14
- e. Loan amount > 17000 & Loan grade is F or G
- f. Loan amount > 13000 & Home ownership category is 'MORTGAGE'
- g. Loan grade is F or G and Interest Rate is high (> 18%)
- h. Employee Tenure is high (> 6 years) and Home ownership category is 'MORTGAGE'
- i. Income is high (> 65k) and Home ownership category is 'MORTGAGE'

Lending Club Case Study

Recommendations

- 1) Company to implement the observations to detect potential defaulter or charged off loan applications before approval decision.
- 2) Dedicated detection rule system to be in place with customizable rules to flag off potential defaulters in separate queue.
- 3) Establish stringent loan approval vetting process workflow.
- 4) Dedicated expert team to analyze flagged off potential defaulter loan applications & be sensitive to
 - a. Reject Loan Defaulters with high accuracy
 - b. Approve Genuine applications who may be flagged off as defaulter
- 5) Train detection rule system with AI models and enhance rules database & knowledge with N-1 week dataset.

Thank you