



Lending Club Case Study

Group Members:

- Gaurav Dhuri
- Soma Mukherjee

Abstract:

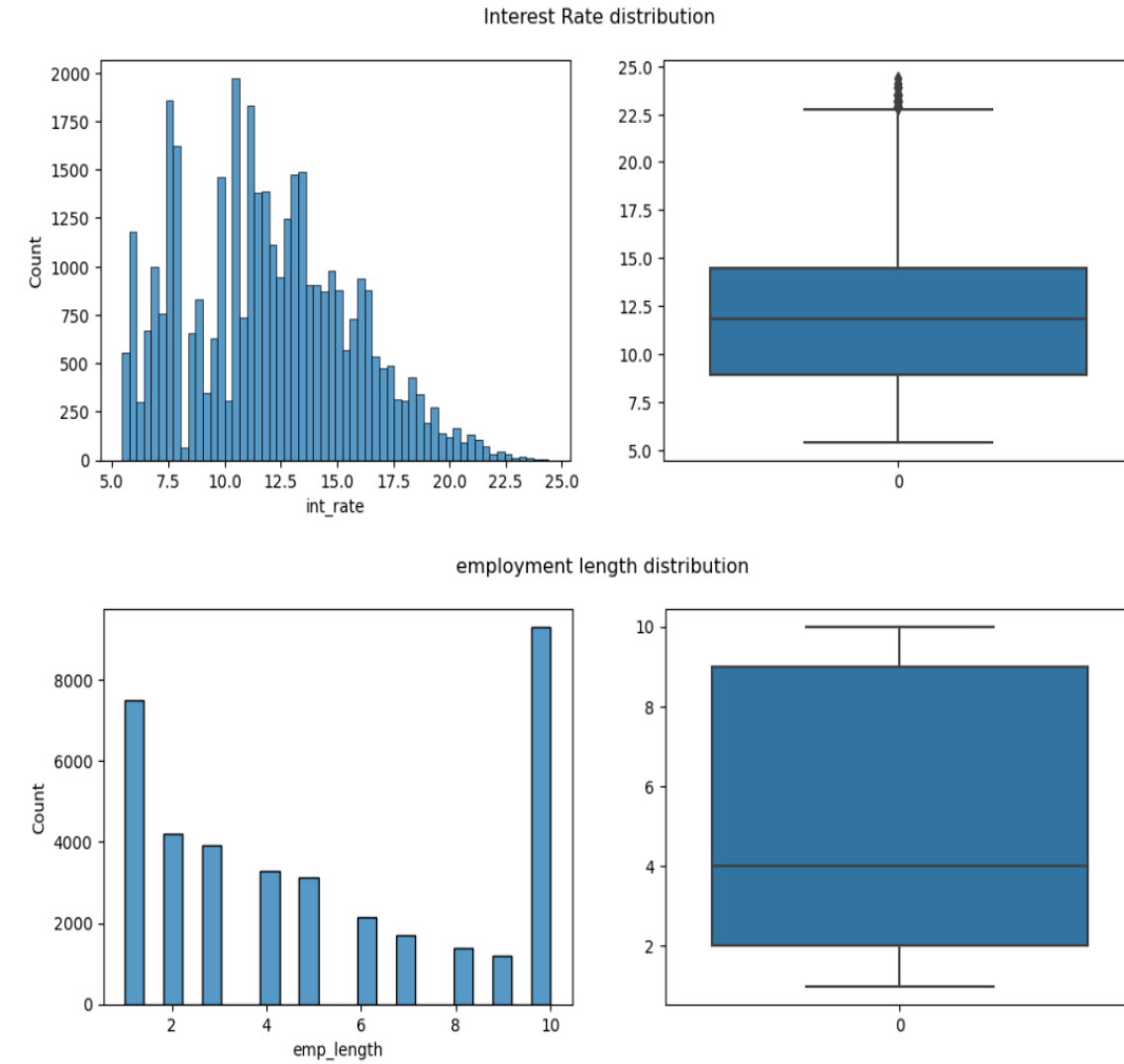
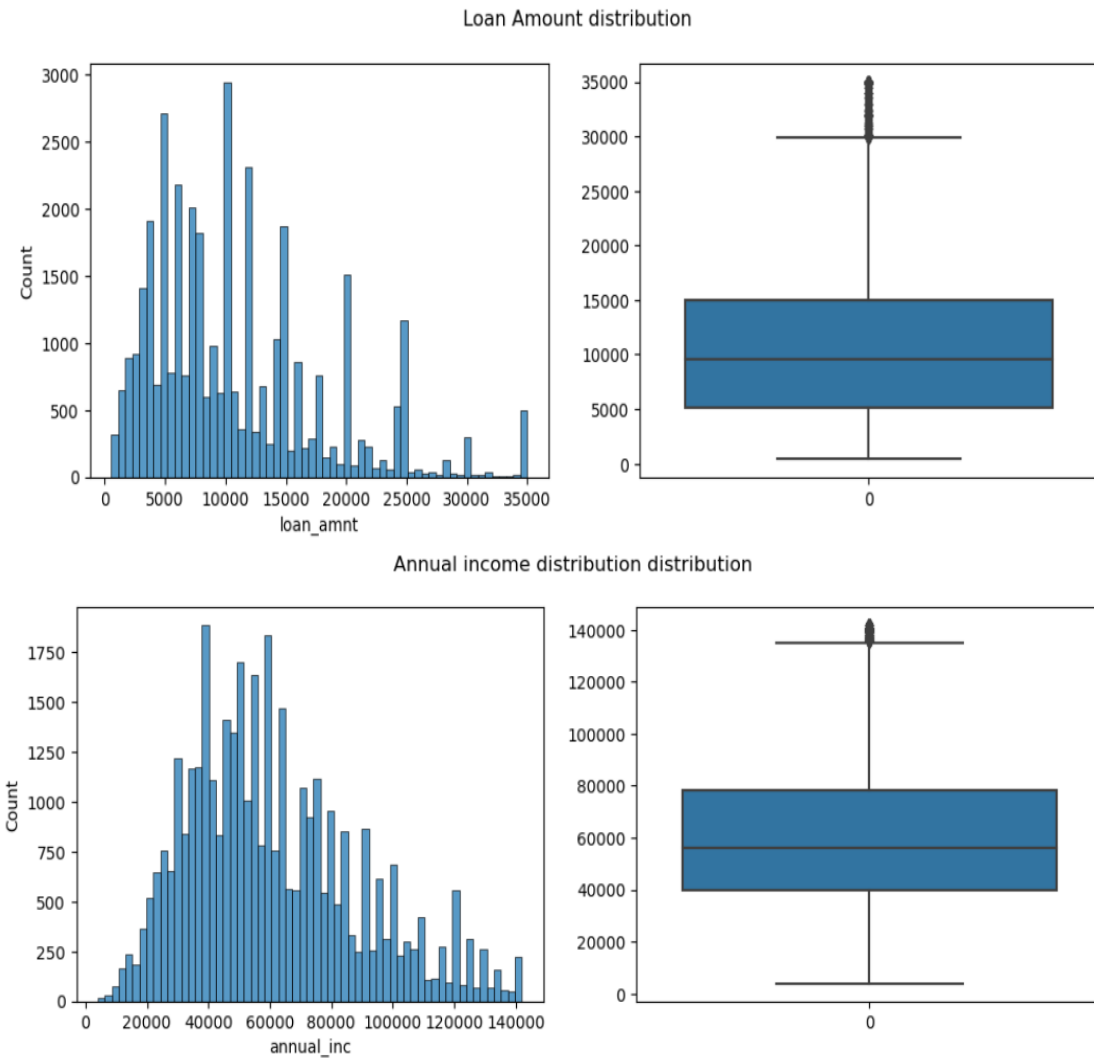
Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

- Borrowers can easily access lower interest rate loans through a fast online interface.
- The main objective of this case study is identification of risky applicants or potential loan defaulters using EDA.
- The company can utilise this knowledge for its portfolio and risk assessment.

Problem Solving Approach:

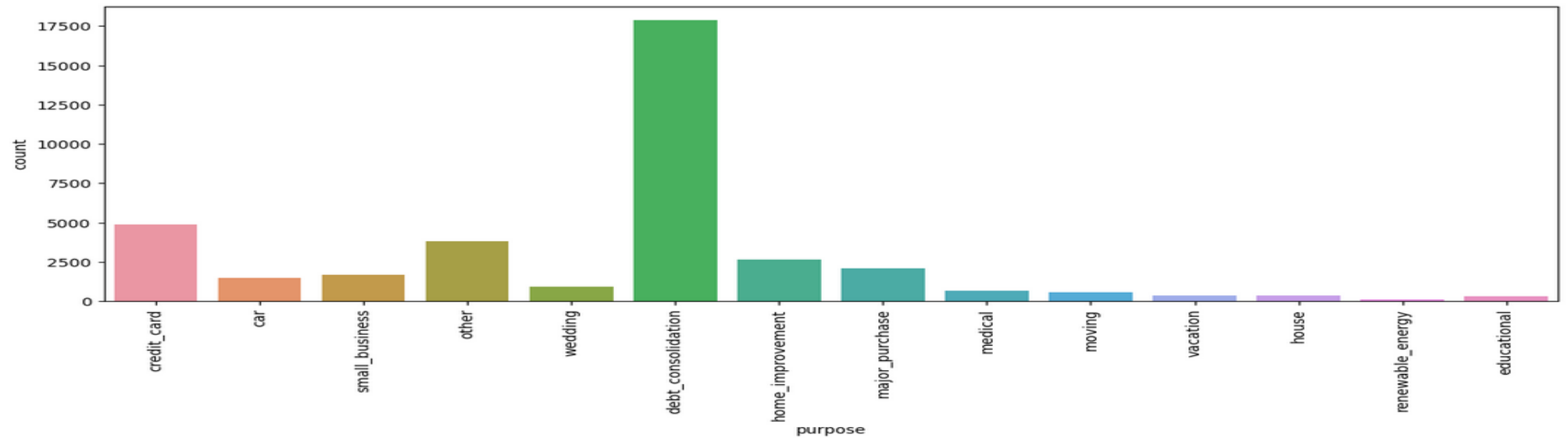
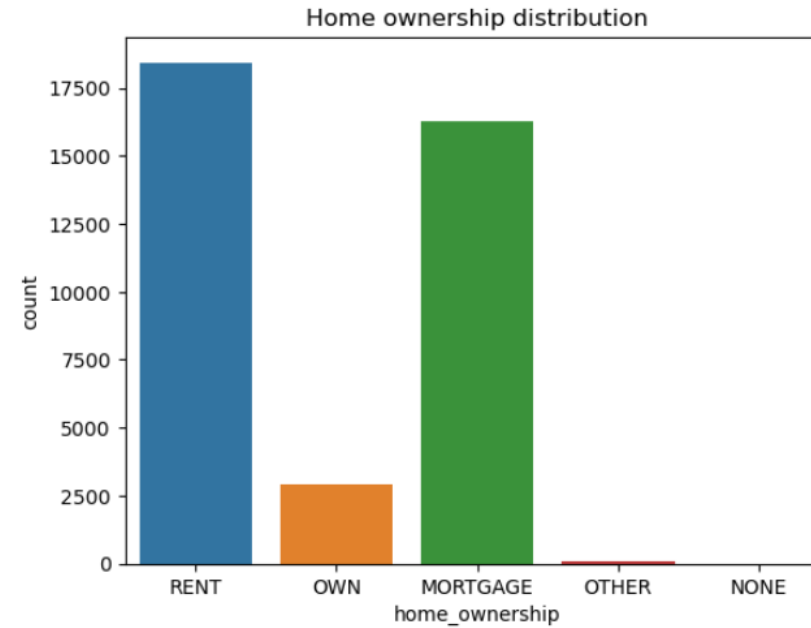
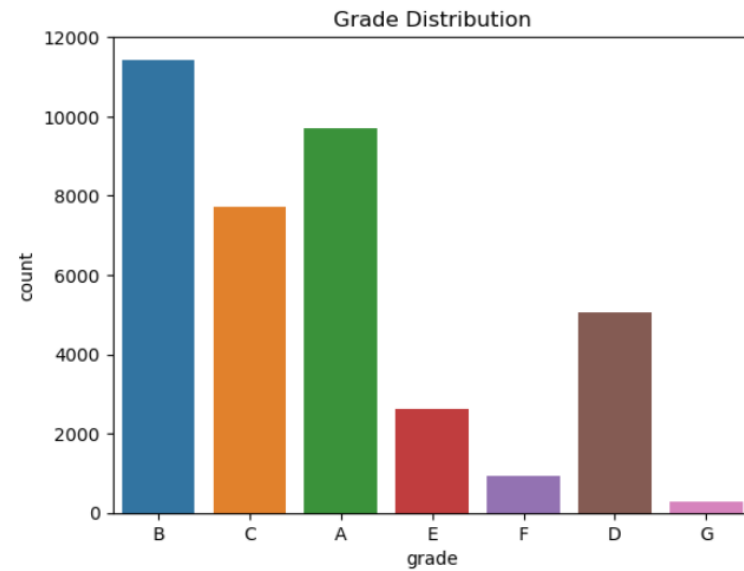
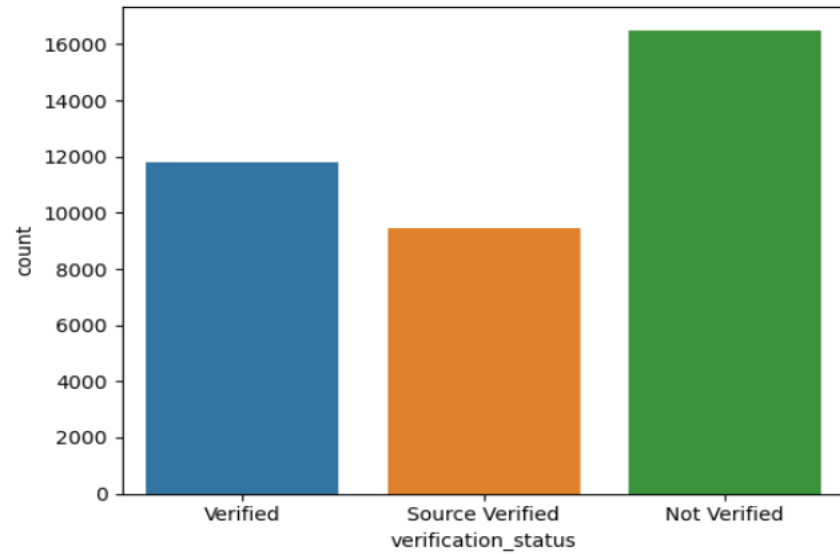
- **Data Understanding :** Understand the dataset and every column
- **Data Cleaning:**
 - Data Inspection: Start by examining the dataset to understand its structure, size, and basic statistics.
 - Handling Missing Values: Identify and handle missing data. You can choose to impute missing values or remove incomplete records.
 - Data Transformation: Convert data types, encode categorical variables, and standardize or normalize numerical features as needed.
- **Data Exploration:**
 - Univariate Analysis: Examine individual variables to understand their distributions and characteristics. Use histograms, bar charts, or summary statistics.
 - Bivariate Analysis: Explore relationships between pairs of variables. Use scatter plots, correlation matrices, and cross-tabulations to identify patterns.
 - Multivariate Analysis: Investigate interactions between three or more variables using techniques like heatmaps, 3D plots, or dimensionality reduction methods.

Univariate Analysis:



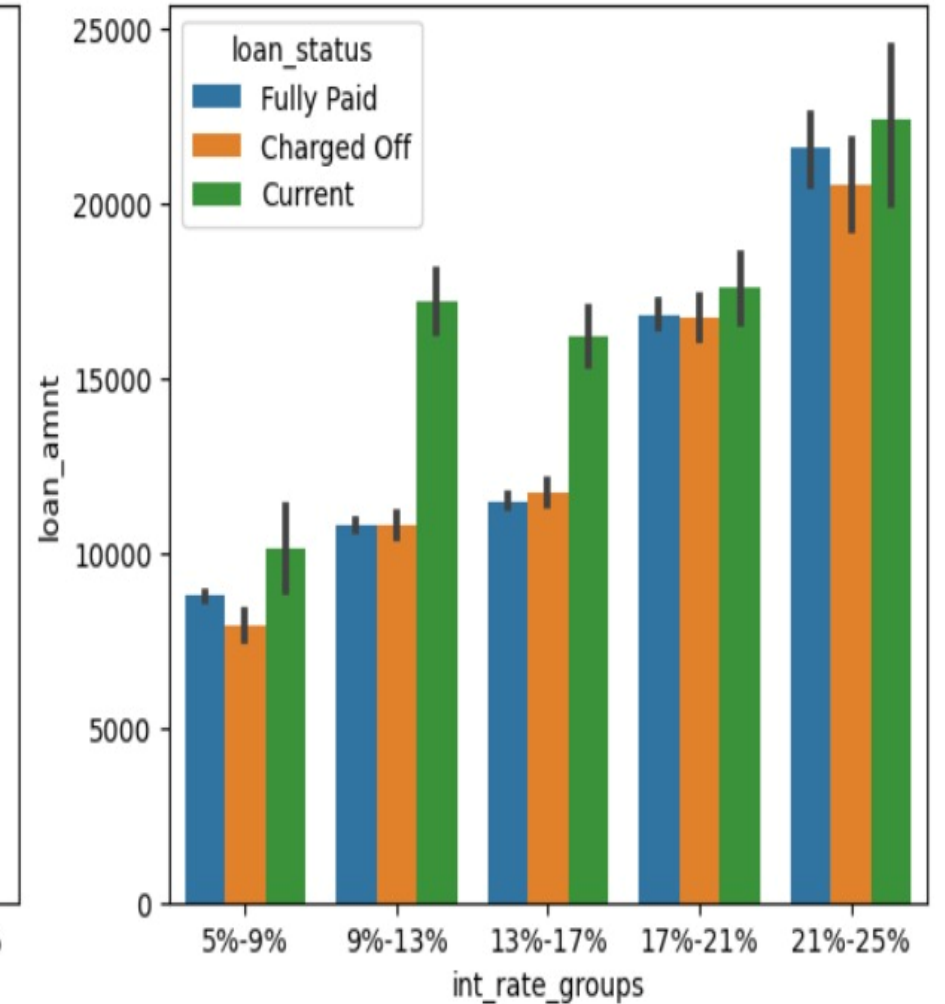
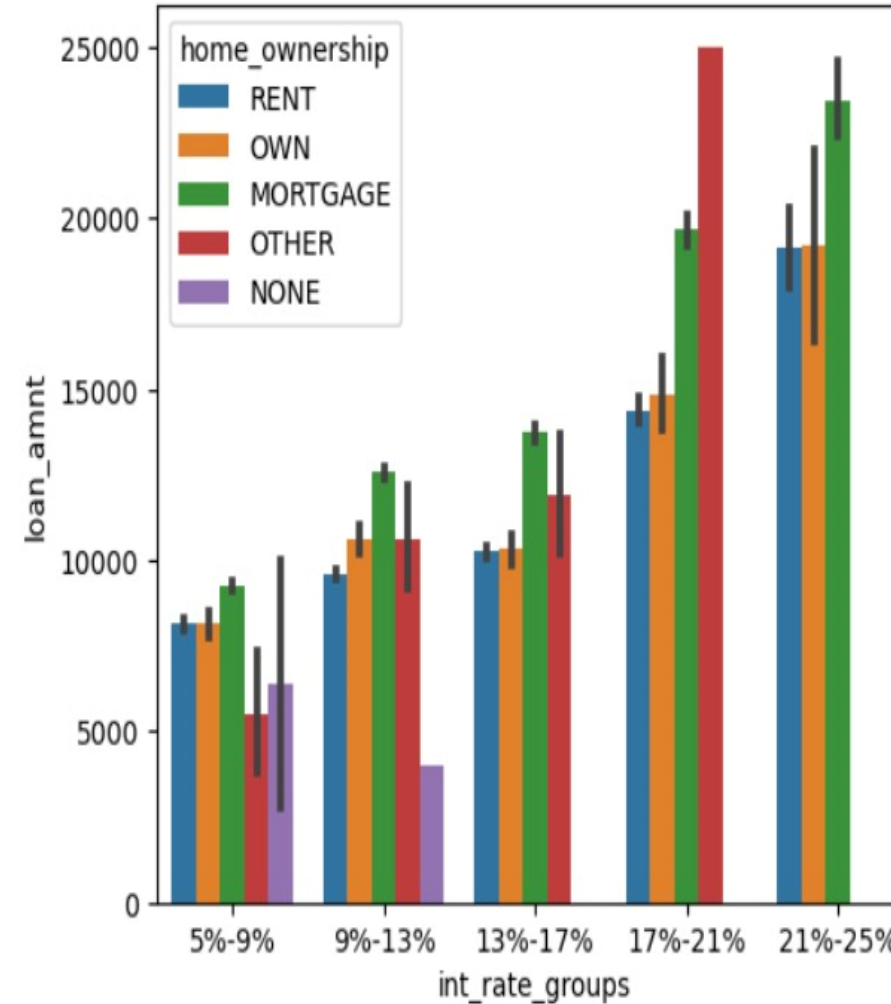
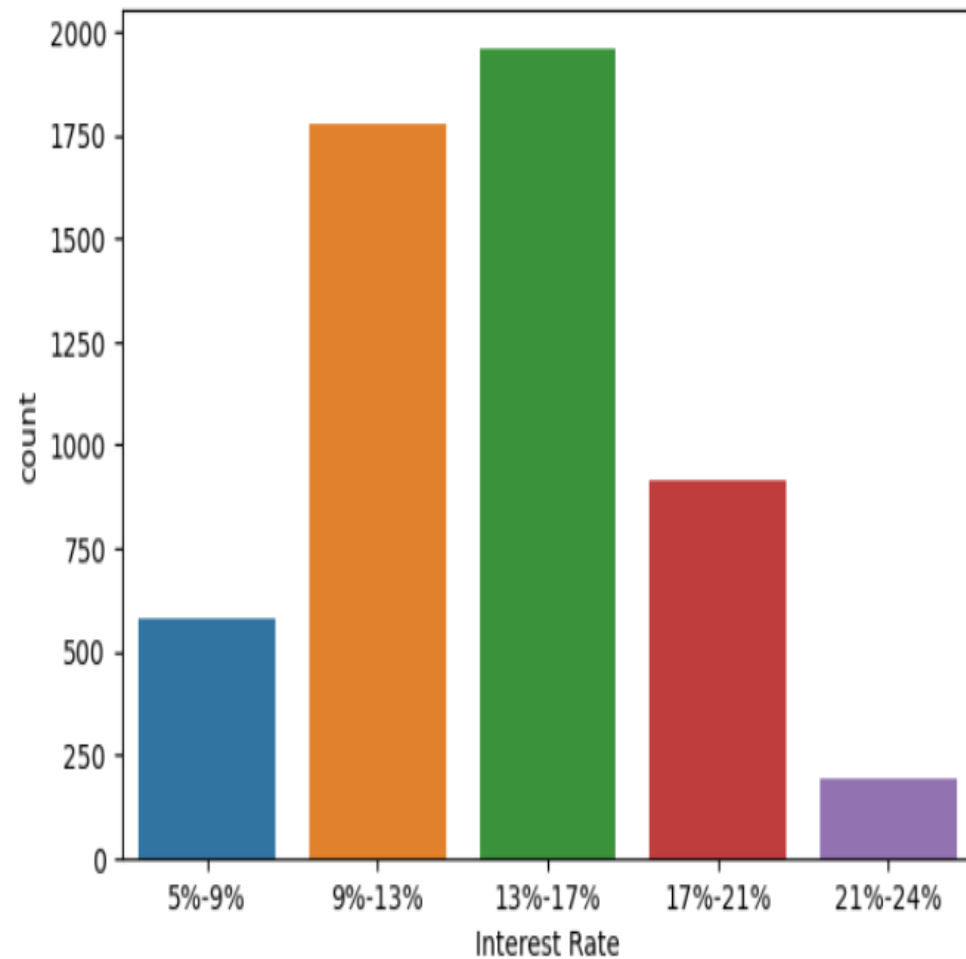
- More number of people took loan amount of 10000. And very few people took a loan amount of more than 30000
- Interest rate mostly lies in the range of 9%-14.5%. There are very few people who took loan at higher interest rate of 22.5%
- Annual income range is mostly found to be within 40000 – 80000
- Most borrowers have 10 years or more years of employment

Univariate Analysis:



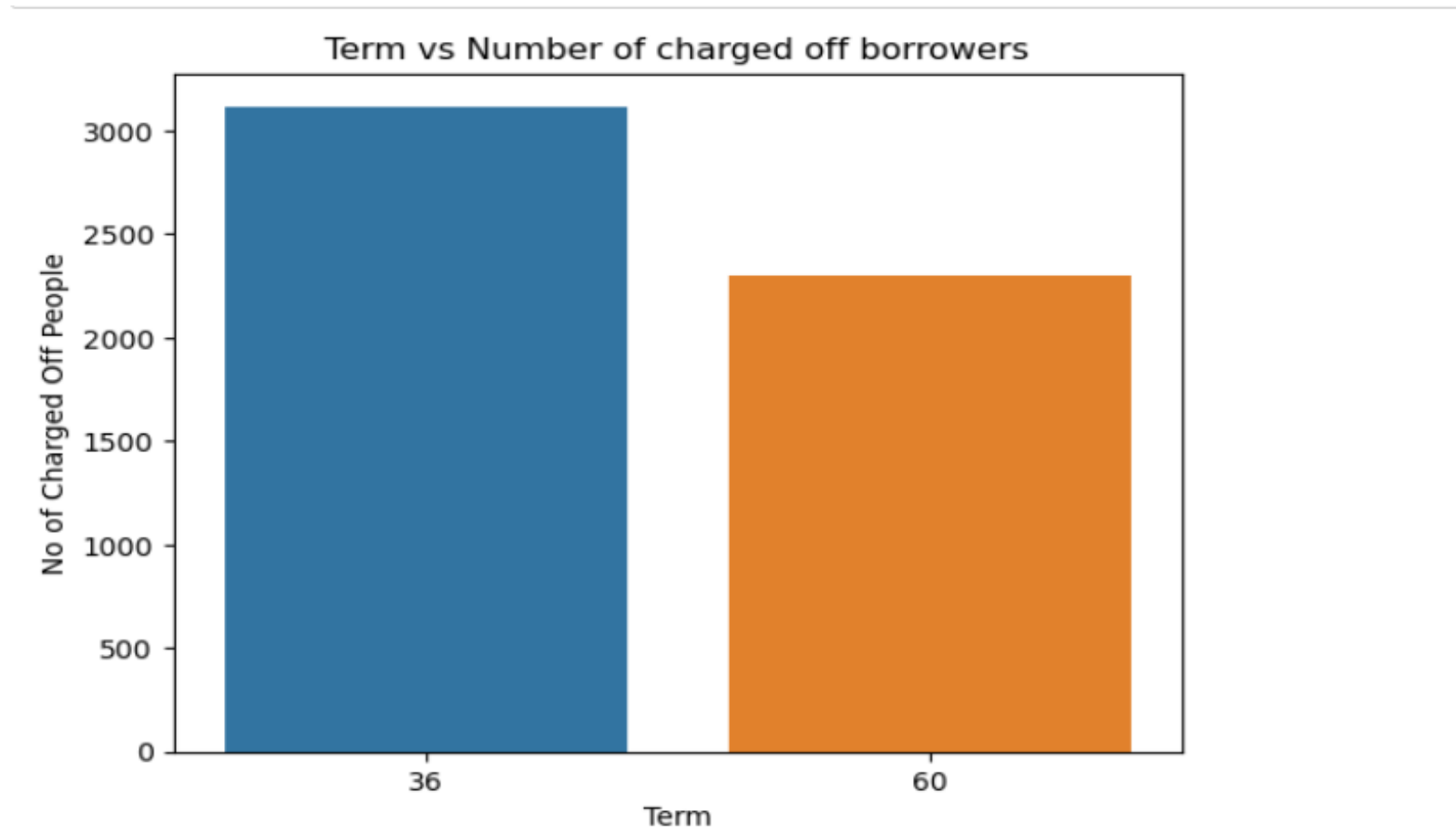
- Not verified borrowers are greater than verified borrowers
- Rent and mortgage borrowers have higher count in home ownership
- More borrowers have taken loans in the grades A and B
- More borrowers have taken loans for the purpose of debt consolidation

Segmented Univariate Analysis:



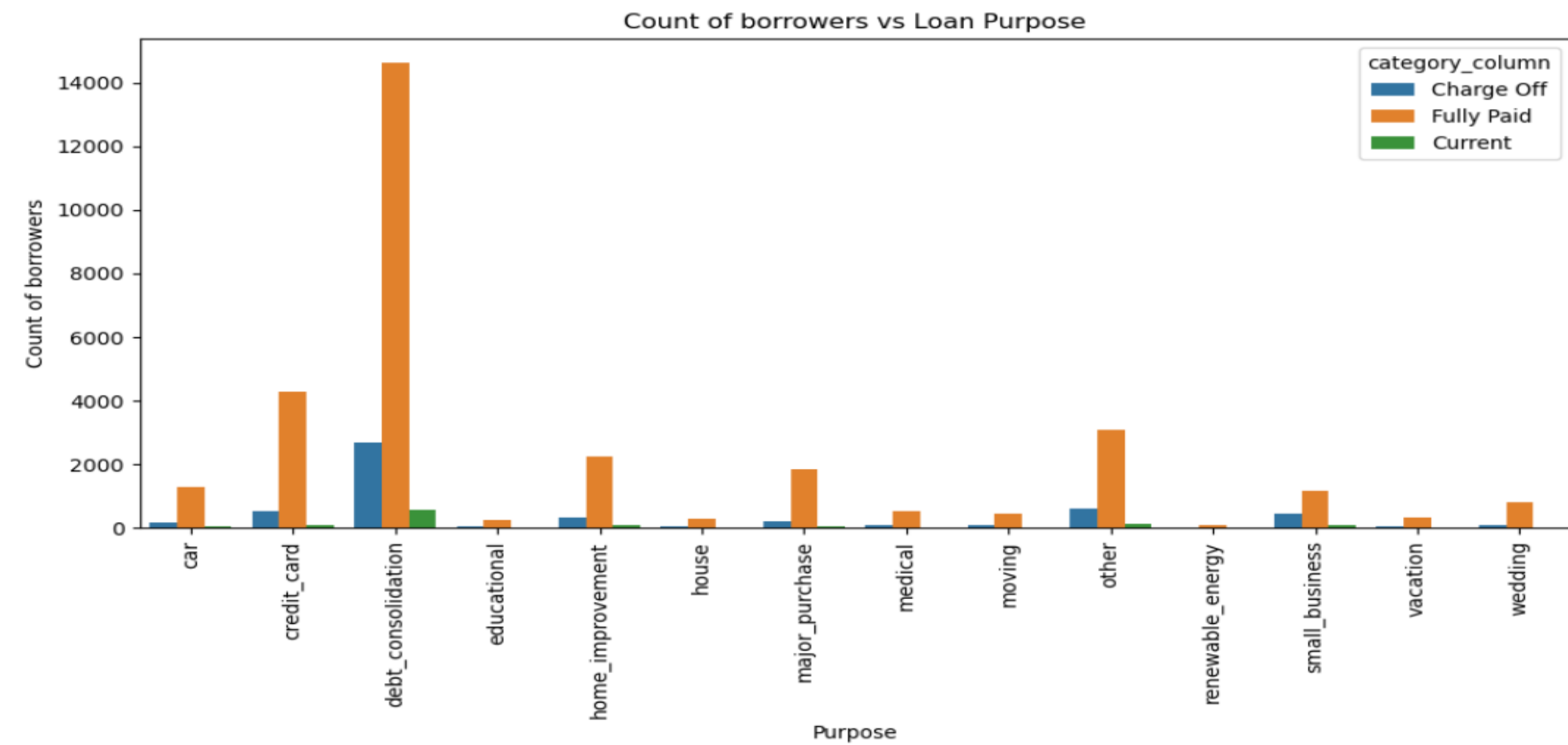
- Maximum number of loan defaulters (charged off borrowers) is in the interest range of 13-17%.
- As the loan amount increases the interest rate increases

Bivariate Analysis:



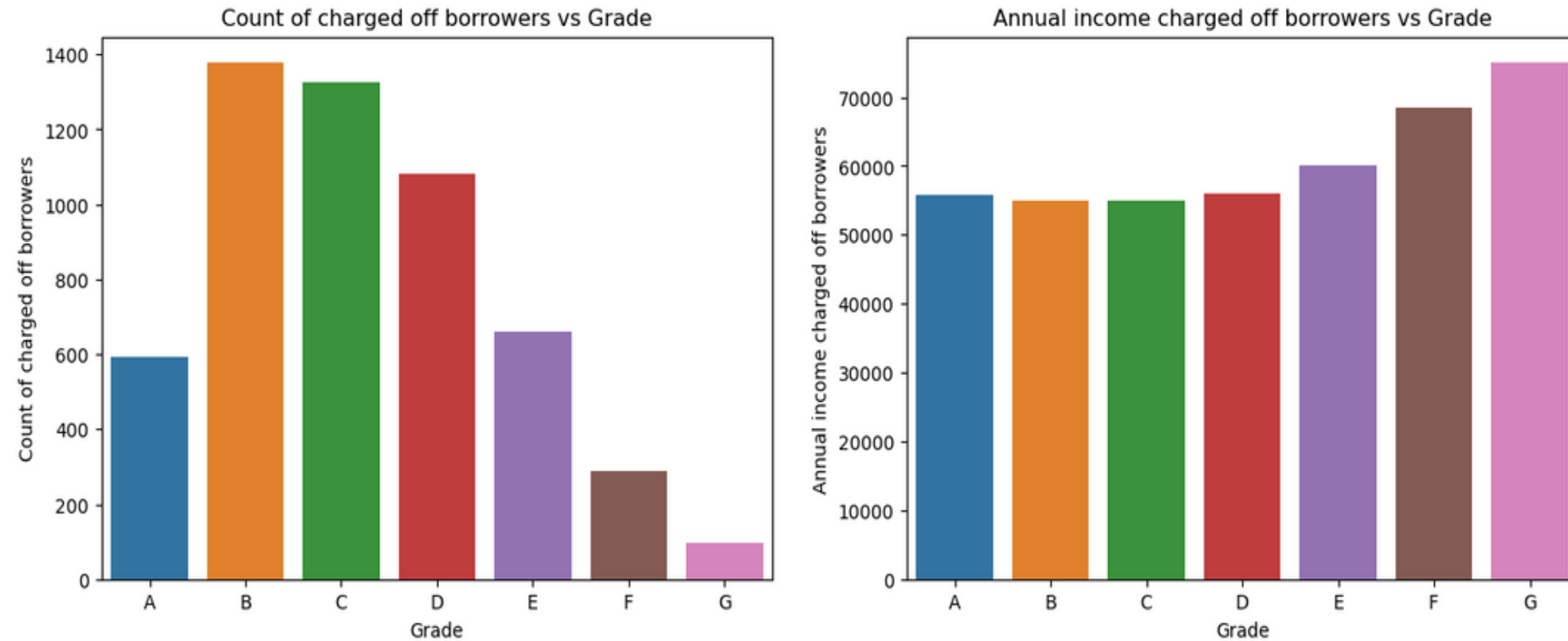
- Number of borrowers who defaulted the loan is more for 36 months than 60 months

Bivariate Analysis:



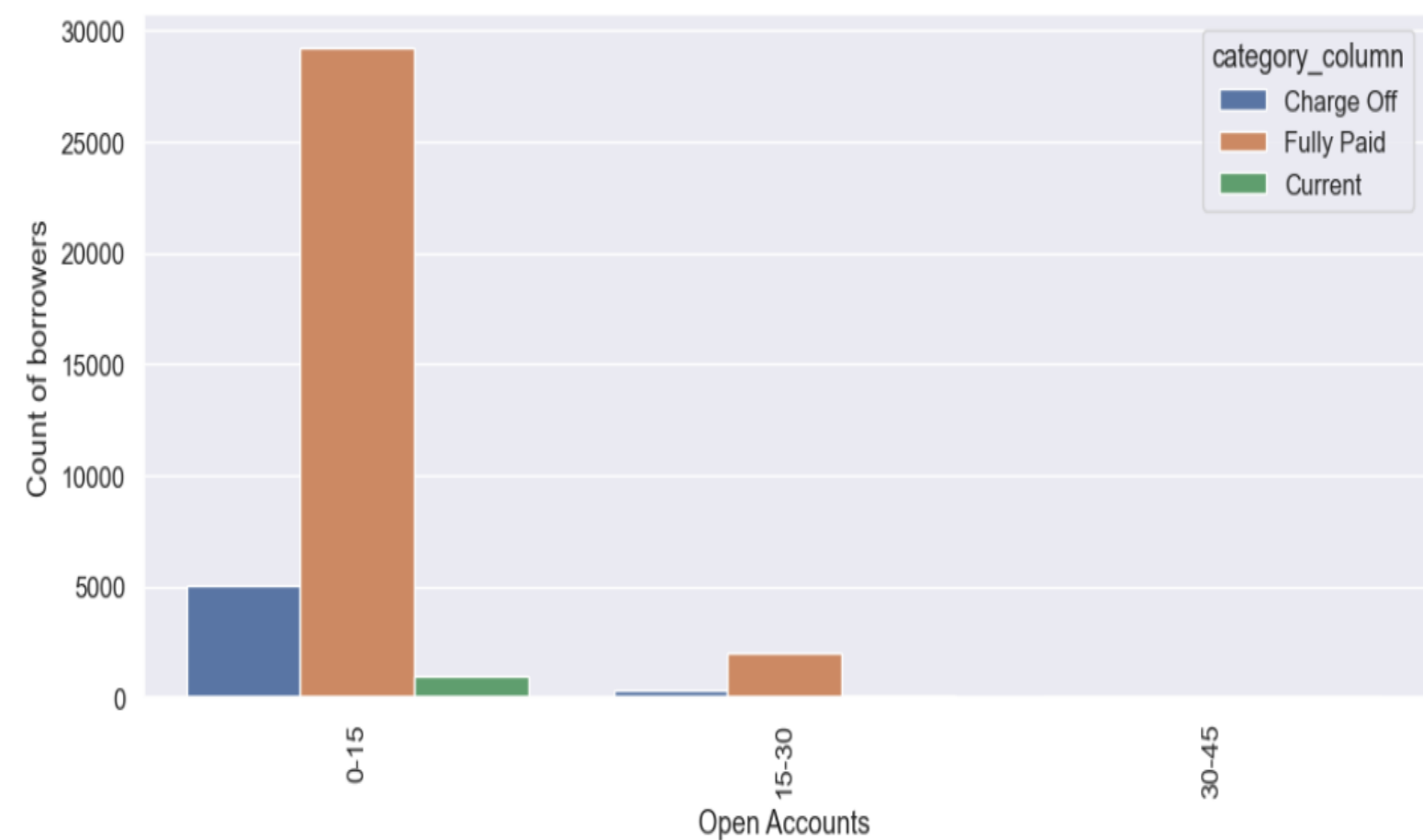
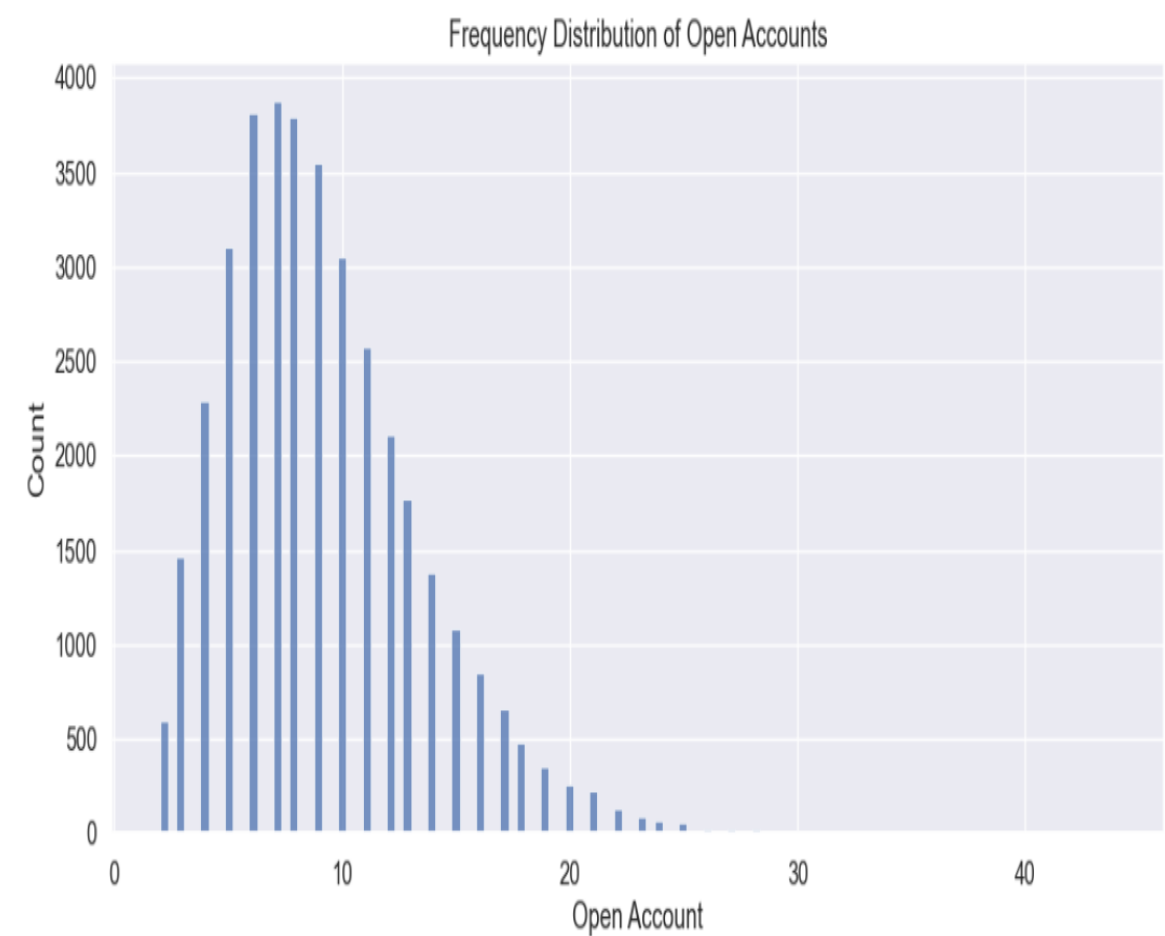
- The charged off borrowers count is highest for loan purpose Debt Consolidation.

Bivariate Analysis:



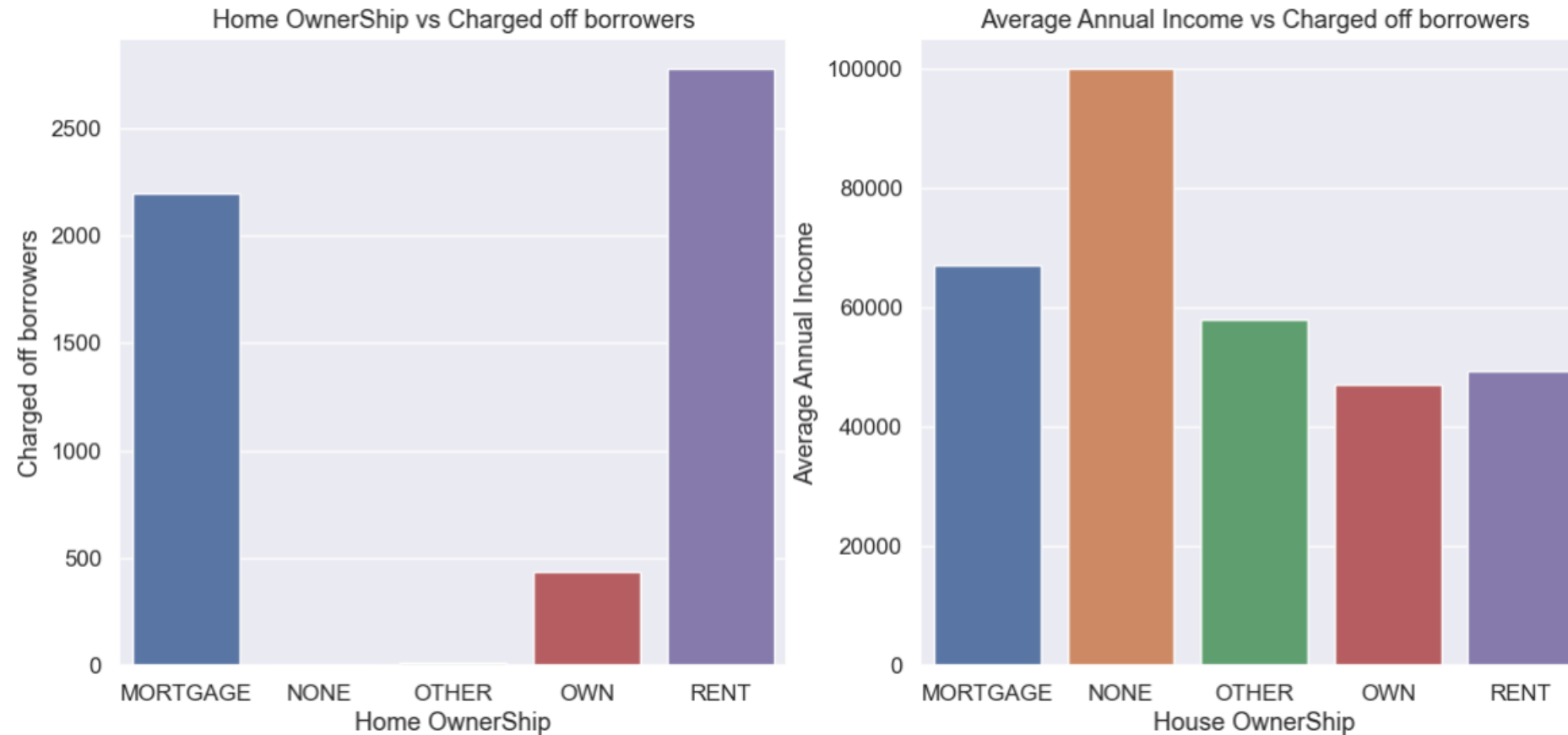
- The count of charged off borrowers is higher for grades B, C, D
- Annual income is lower for grades B, C, D.
- Thus borrowers with lower annual income and grades B, C, D are likely to be loan defaulters.

Bivariate Analysis:



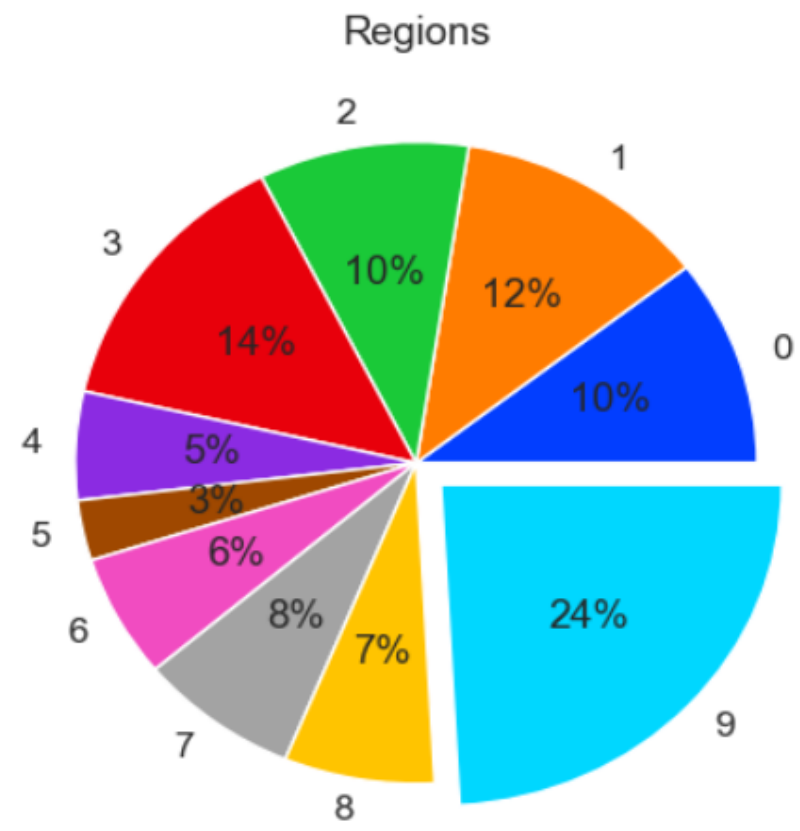
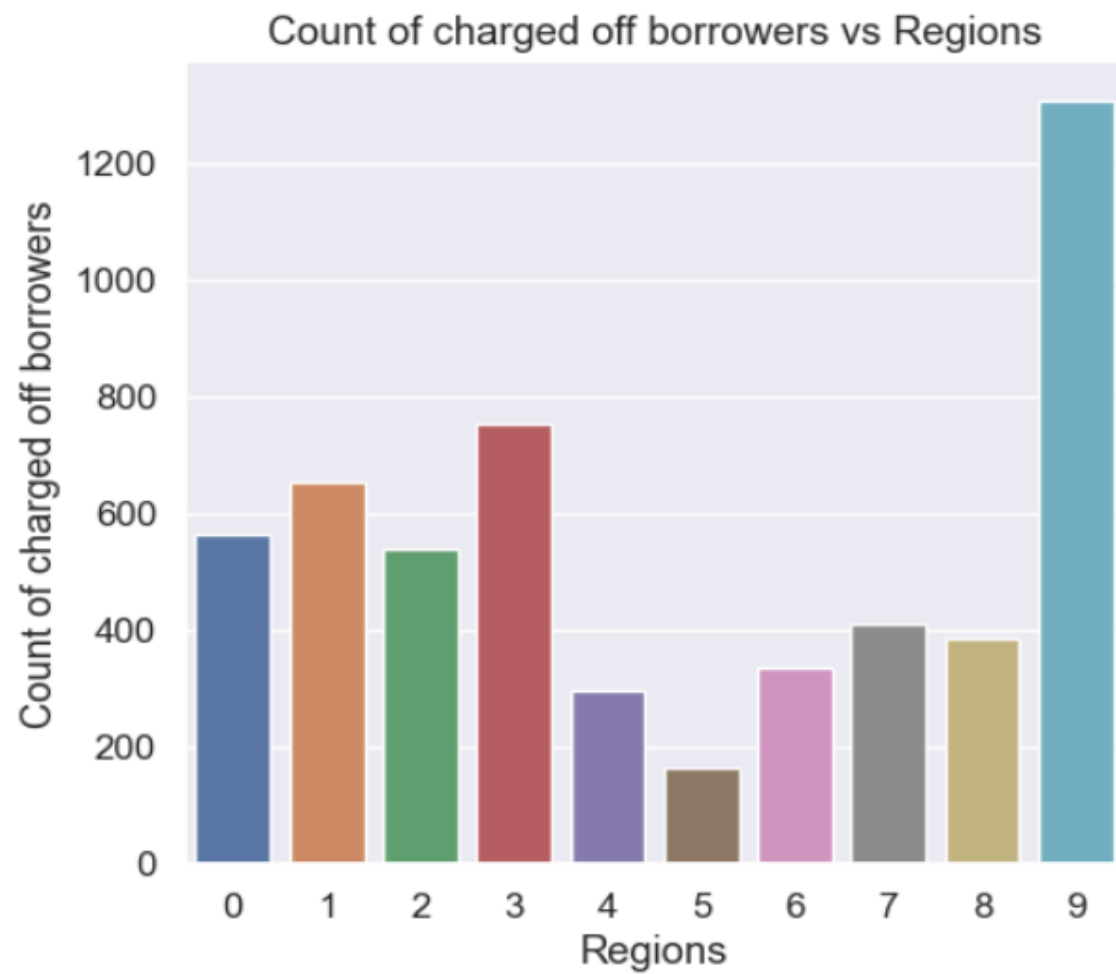
- Shows the frequency distribution of open accounts for all borrowers.
- Second plot shows the distribution for three types of borrowers, charged off, current and fully paid. The charged off borrowers mostly have open accounts in the range of 5-15.

Bivariate Analysis:



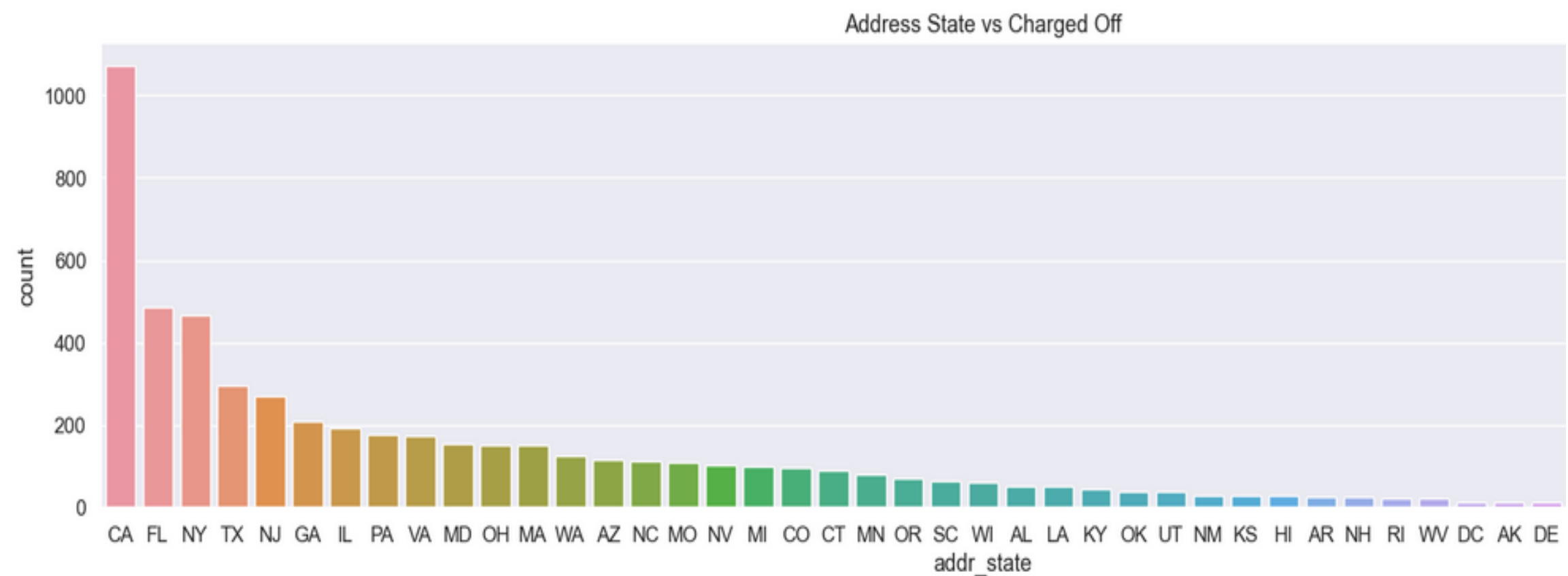
- The count of charged off borrowers with mortgage and rent is highest.
- One of the reasons could be the average annual income which is comparatively lower for mortgage and rented borrowers as we can see in plot 2.

Bivariate Analysis:



- The region 9 has the maximum number of charged off borrowers. Therefore maximum number of loan defaulters is likely to be from region 9.

Bivariate Analysis:



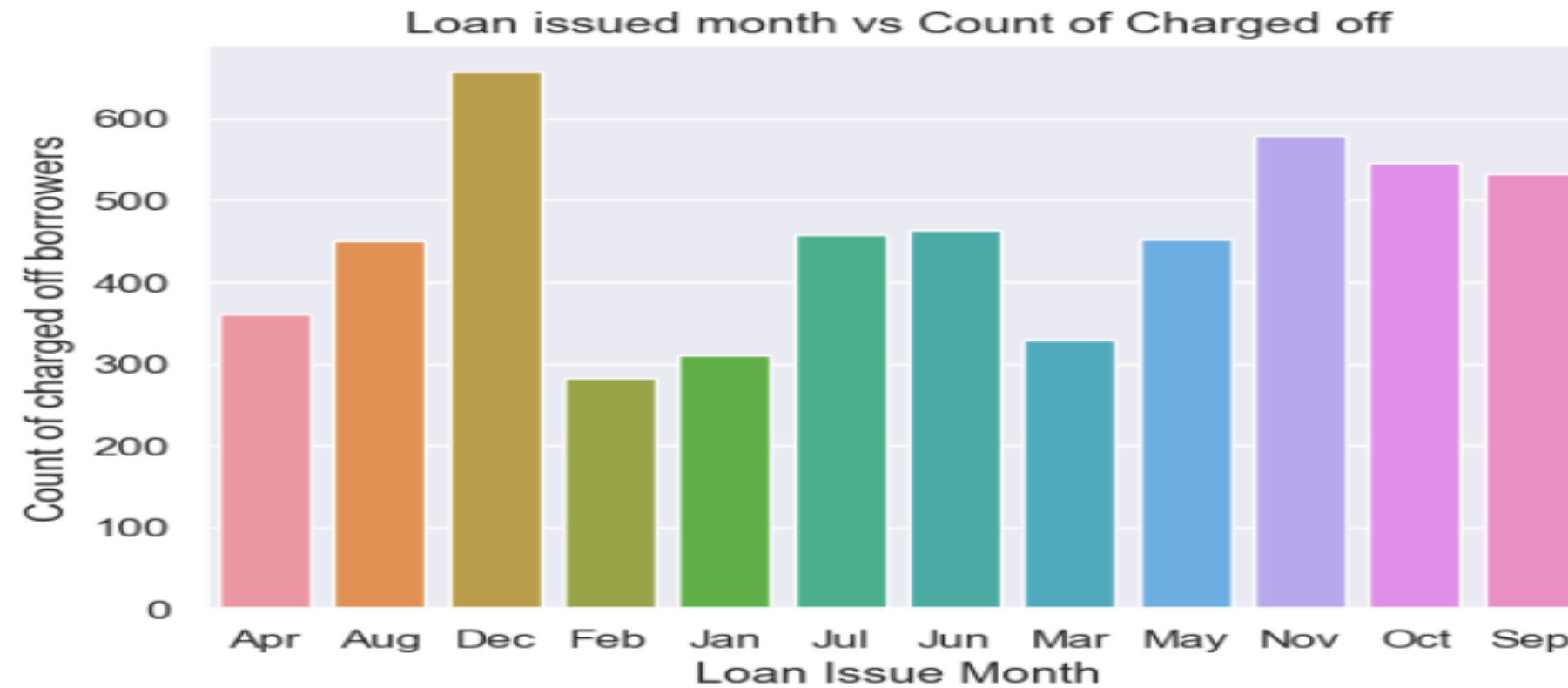
- The maximum number of charged off borrowers are found to be in address state CA – California which also falls under region 9

Bivariate Analysis:



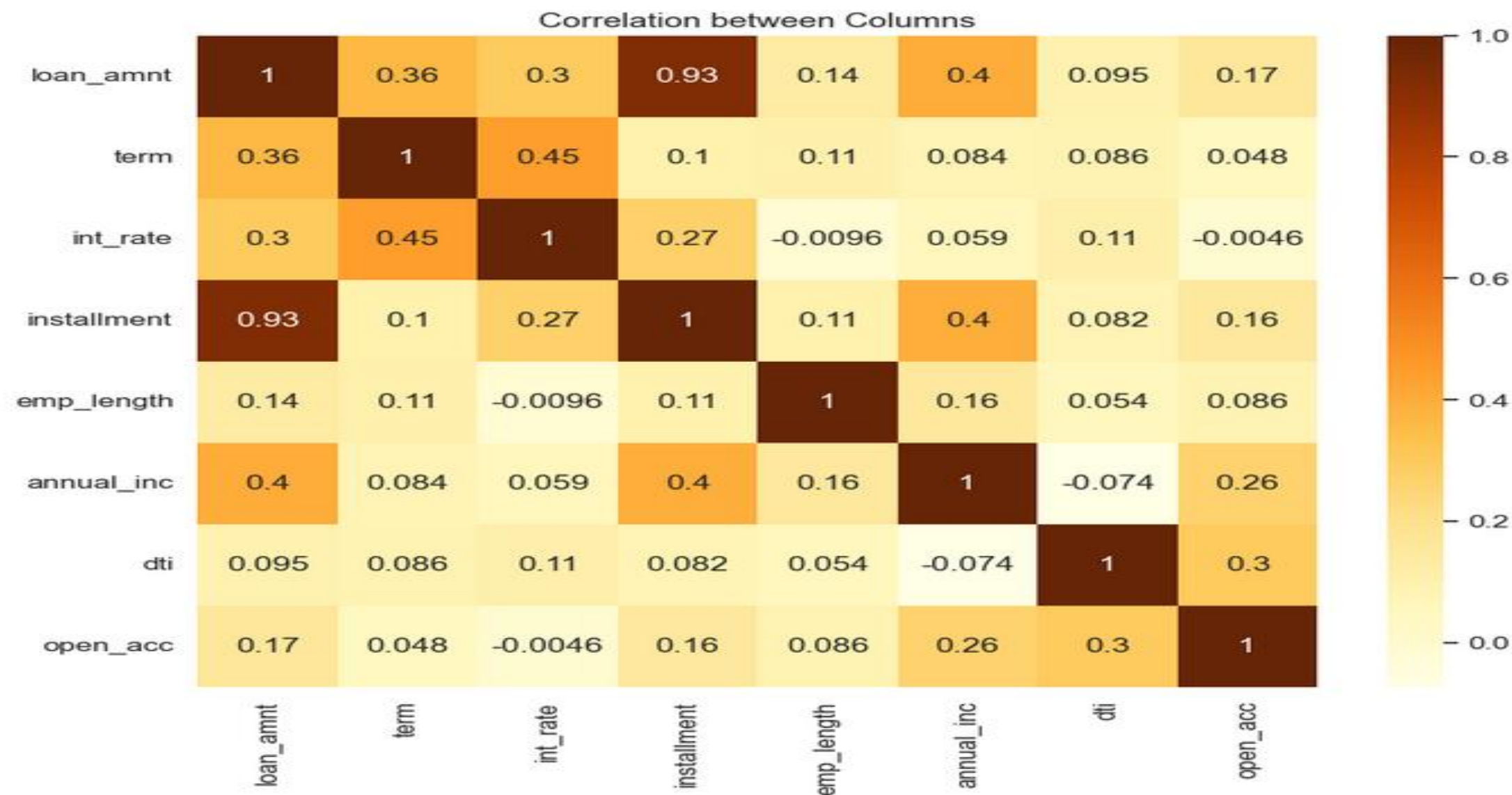
- The 10 years of employment length has higher count of charged off borrowers because it holds the sum of all borrowers above 10 years of employment due to larger distribution of borrowers (10+). Also it was seen that more number of borrowers had taken loans with 10 or more years of employment.

Bivariate Analysis:



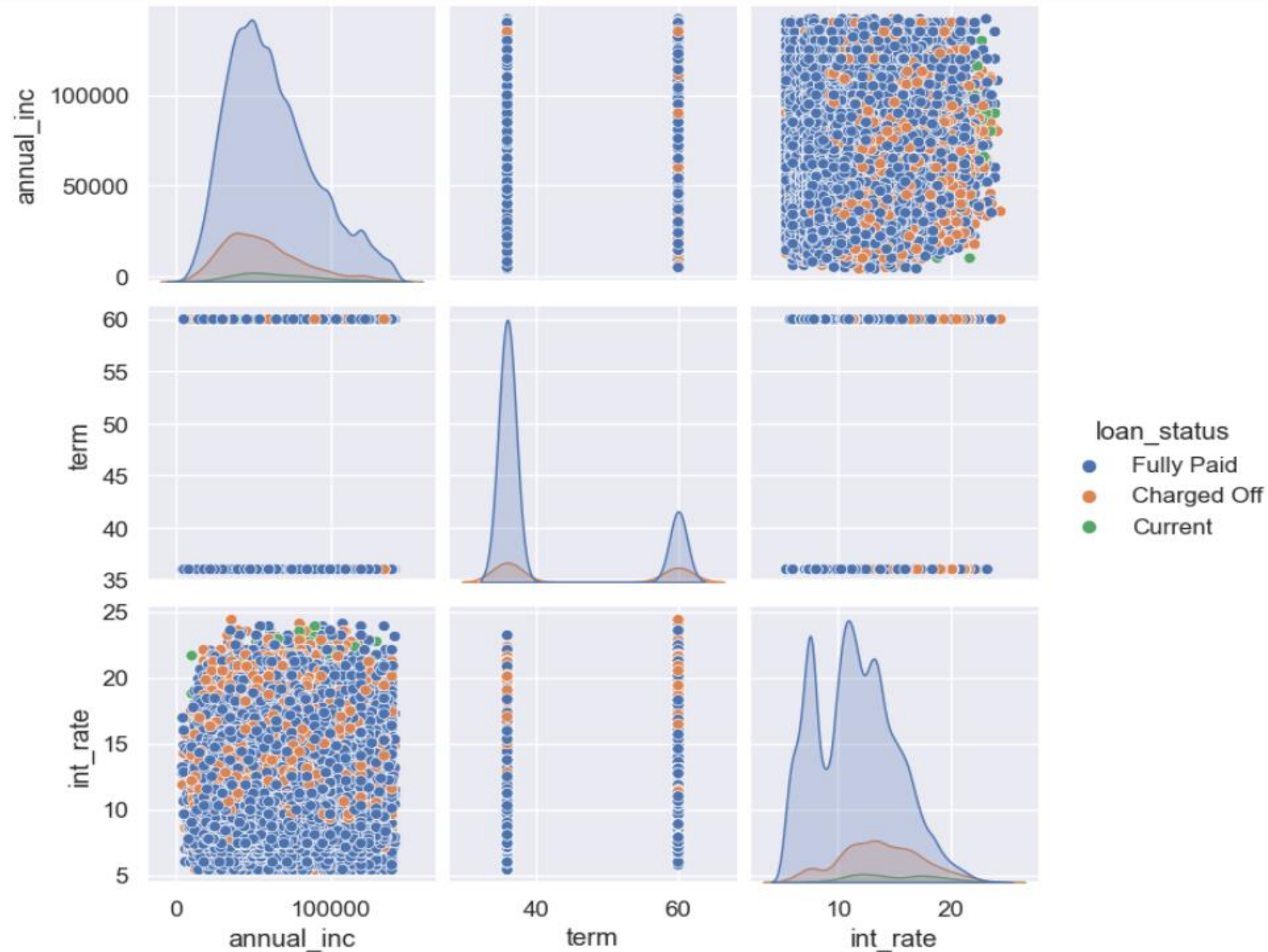
- The maximum number of charged off borrowers are found to be in the last quarter of the year (Oct, Nov, Dec)

Correlation Matrix:



1. Strong positive correlation exists between installment and loan amount. Higher the loan amount more will be the installment.
2. Positive correlation exists between term and interest rate. Higher the loan term , higher the interest rate
3. Positive correlation is observed between annual income and loan amount.

Multivariate Analysis:



- The borrowers with annual income approximately less than 60000 and with higher loan interest rate are more prone to be defaulters.
- For both loan terms (36 and 60 months), if the interest rate is higher, it will lead to more loan defaulters
- The interest rate ranges between 10-20% for loan defaulters.

Conclusion:

- People with lower annual income borrowing the loan with high interest rate for 36 months term, are more prone to be defaulters
- Borrowers with home ownership as mortgage and rent are more likely to be defaulters because the average annual income is relatively lower for these two categories.
- Grades can be a good metric for detecting loan defaulters. The maximum loan defaulters are found to be mostly in grades B, C, D. Therefore lending club should re consider and verify the borrowers who are applying loans for grades B, C and D.
- Region 9 has maximum number of loan defaulters, therefore lending club should examine all the borrowers who are applying loans from this region.
- Address state CA that falls under region 9 also has the maximum number of loan defaulters. Therefore lending club should control their numbers of loans being approved for this address state.
- The maximum number of loan defaulters borrowers have a employment length of 10 or more years.
- The loan defaulters are maximum in the last quarter of the year, so the lending club should examine and control the loans being sanctioned in the last quarter of the year.
- The loan defaulters mostly have open accounts in the range of 5-15. Thus lending club should always verify the loan applicants with more than 5 open accounts.
- Negative correlation is observed between DTI and annual income, which means annual income is inversely proportional to DTI. Therefore lending club should be more cautious in granting loans to borrowers who have higher DTI and lower annual income.