

Lending Club Case Study

Exploratory Data Analysis

Data Understanding

- Data volume
- Columns and their data types
- Numeric vs Categorical Columns

Data volume

- Number of rows are 39,717 rows
- Number of columns are 111 columns
- Number of rows on schema are 115 rows that means there are 115 columns
- Summary: number of columns on data set is not equal to number of columns on schema.

Columns and their data types

- There are 74 columns of float64 type
- There are 13 columns of int64 type
- There are 24 columns of object/string type

Numeric vs Categorical Columns

- Number of numeric columns is 24 columns
- Number of categorical columns is 87 columns

Data Cleaning

- Fixing Rows and Columns
- Treating missing values
- Standardizing values
- Treating invalid values
- Filtering data

Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Correlation b/w features

Univariate Analysis

- Categorical Variables
- Quantitative Variables

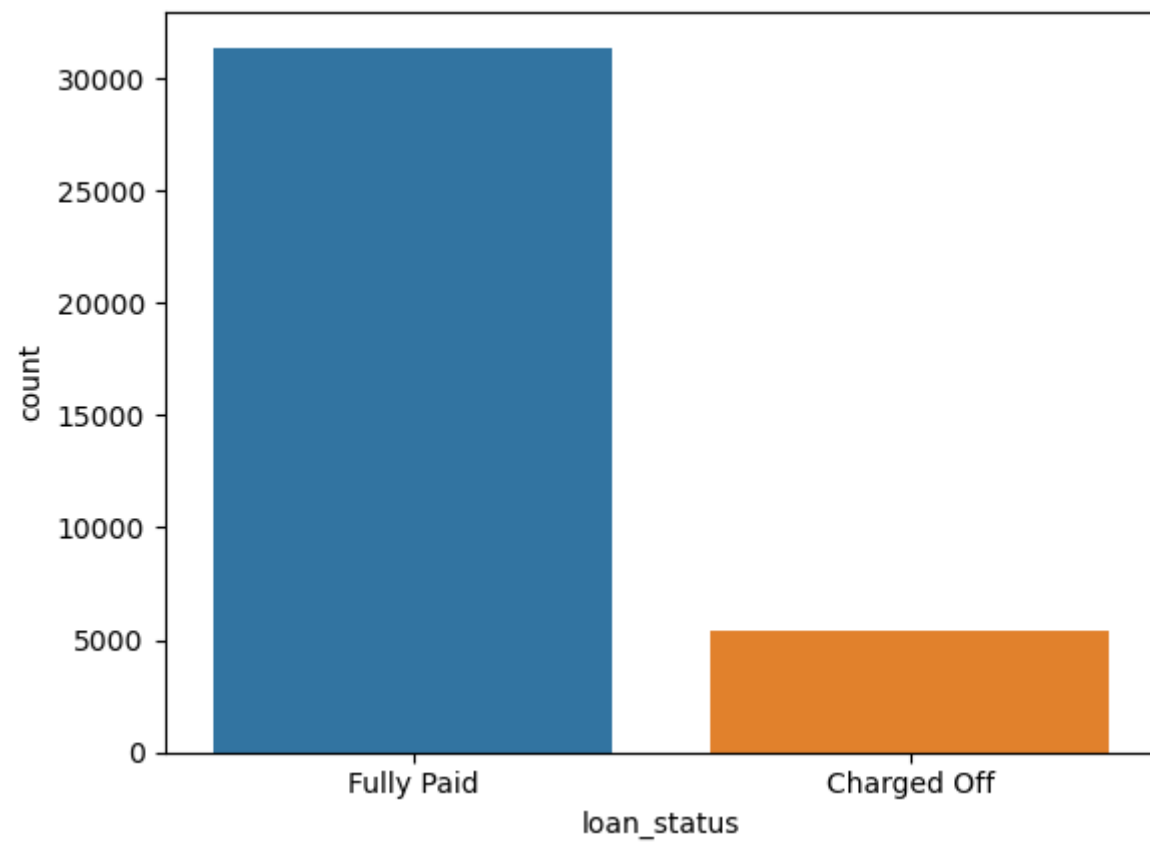
Create new column with defaulted name to mark defaulted applicants

```
df_loan['defaulted'] = df_loan.loan_status.apply(lambda x: 1 if x == 'Charged Off' else 0)
```

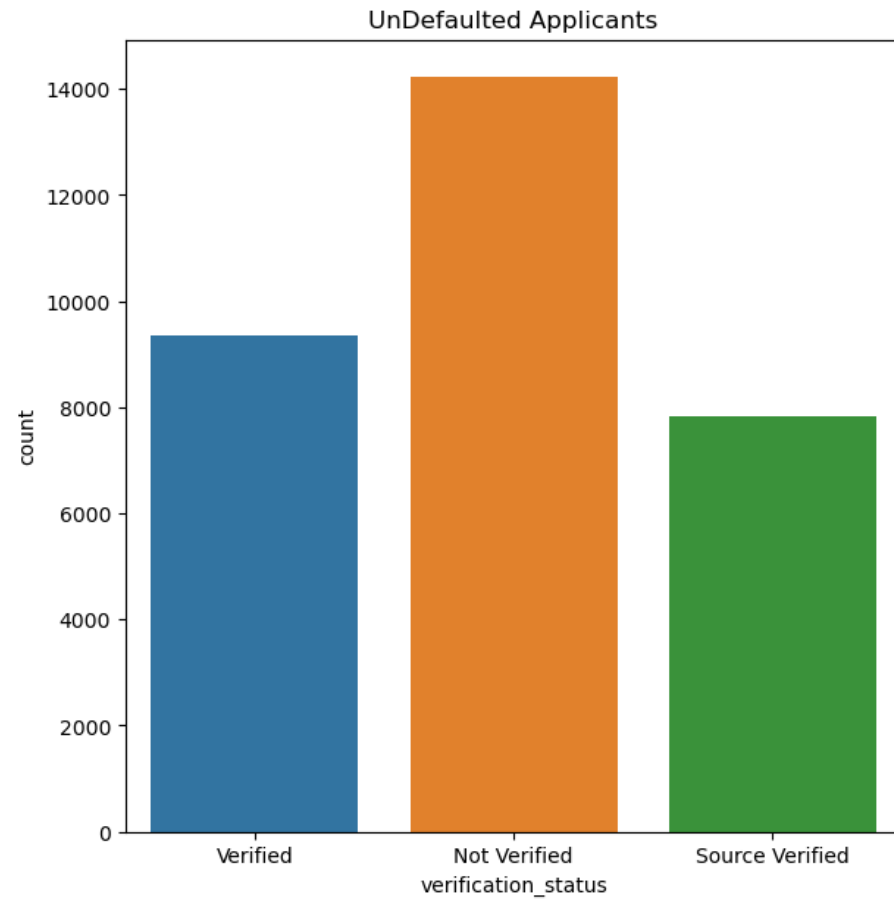
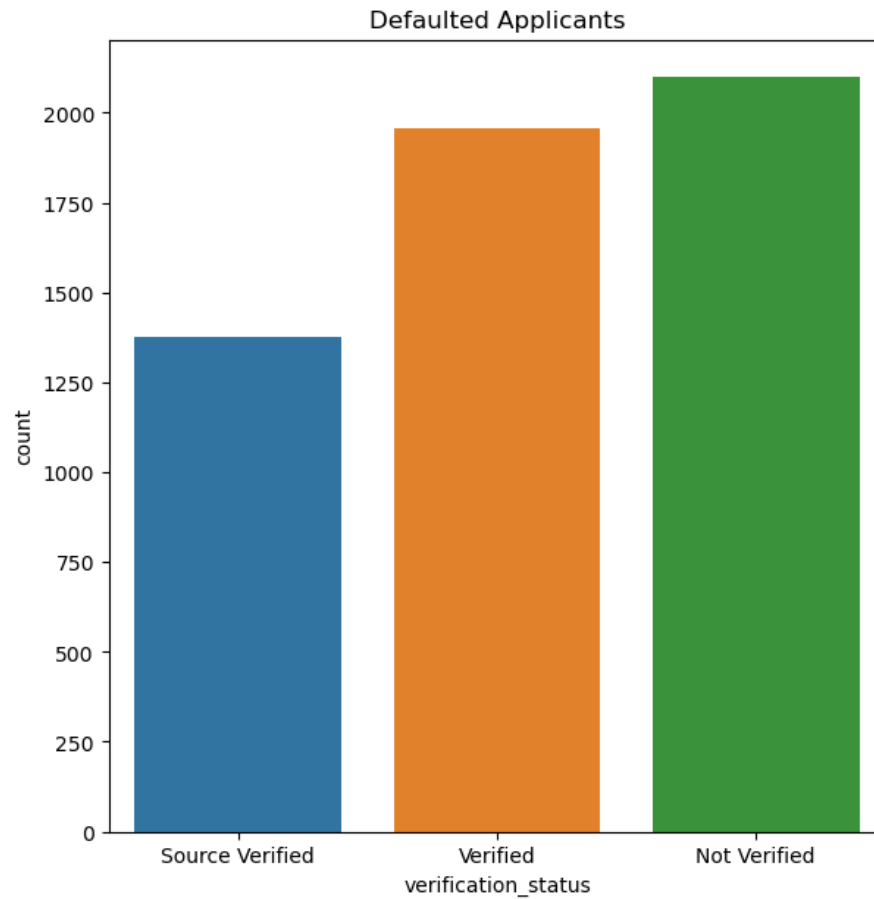
Filter data to 2 data set with defaulted and undefaulted

```
df_analysis_defaulted = df_analysis[df_analysis.defaulted == 1]  
df_analysis_undefaulted = df_analysis[df_analysis.defaulted == 0]
```

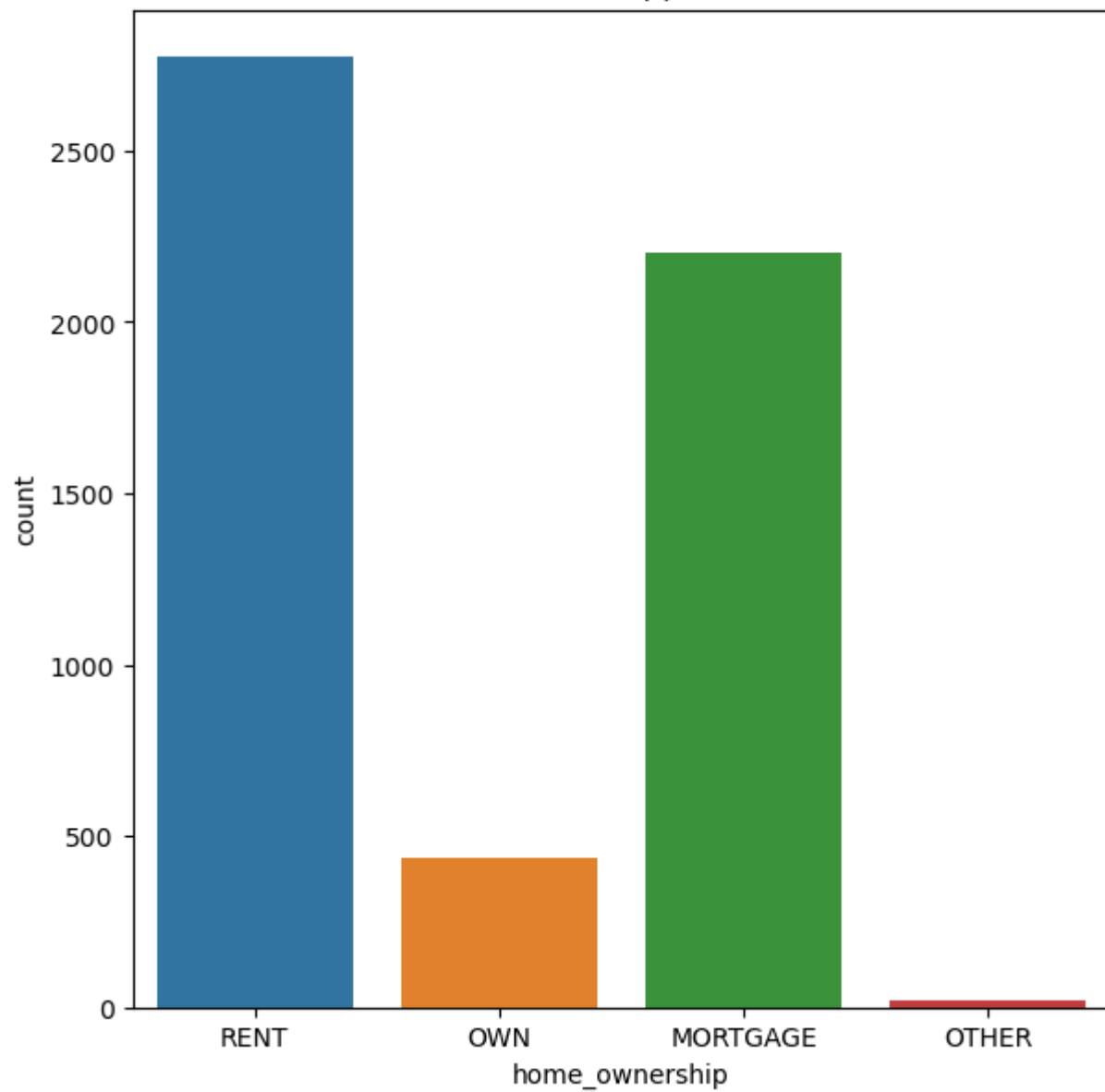

Categorical Variables



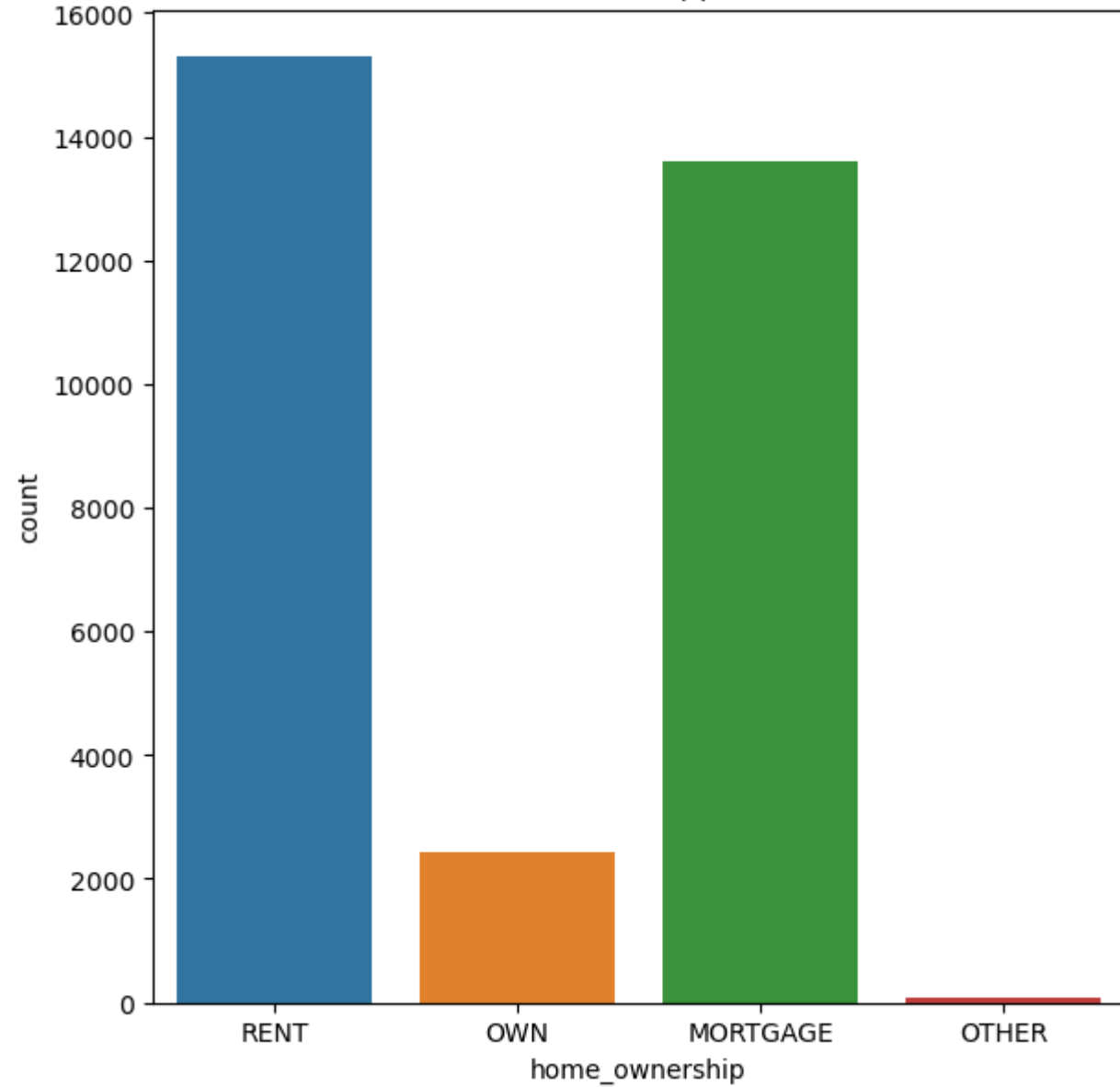
Categorical Variables



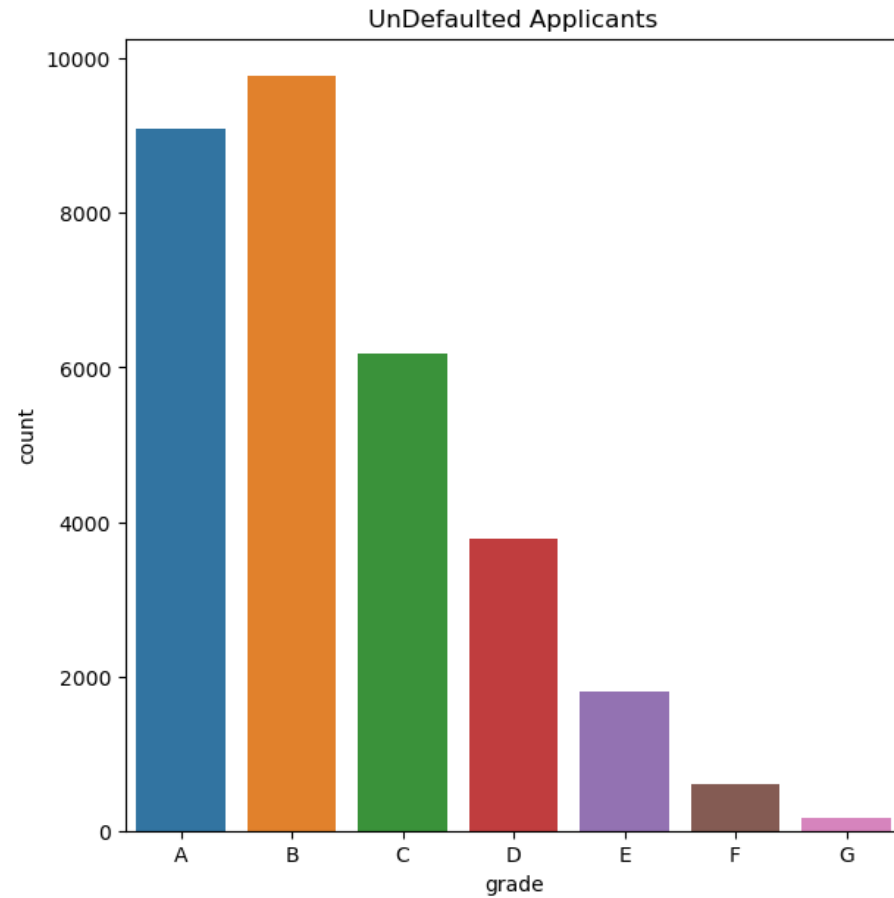
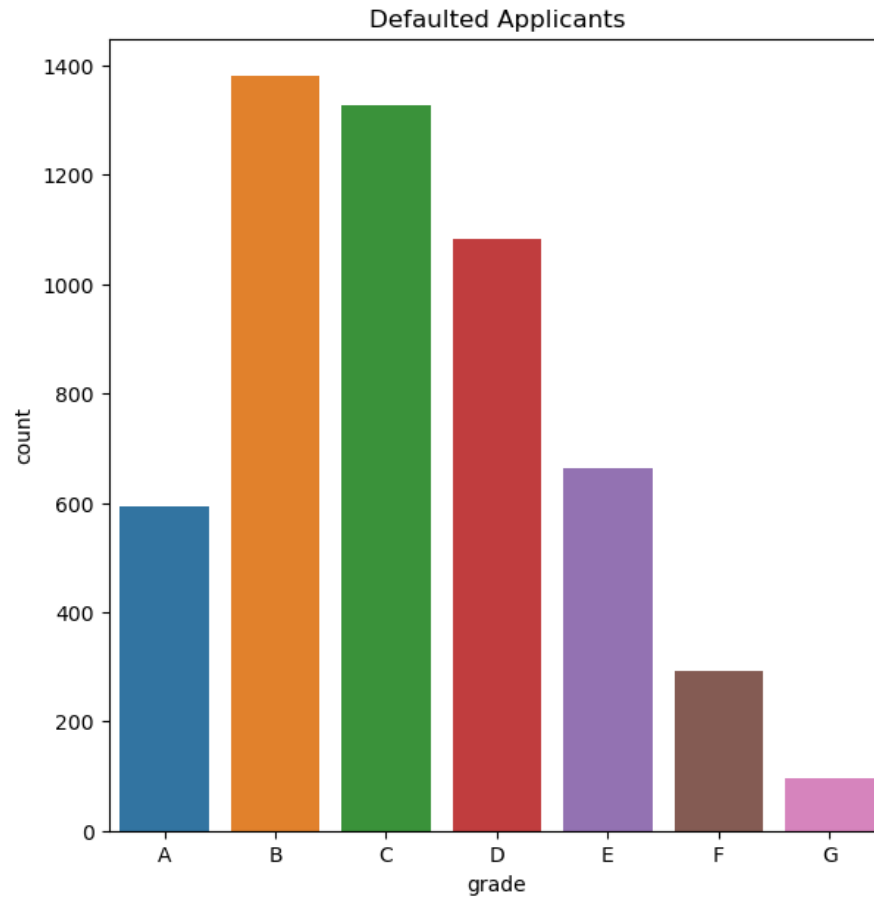
Defaulted Applicants

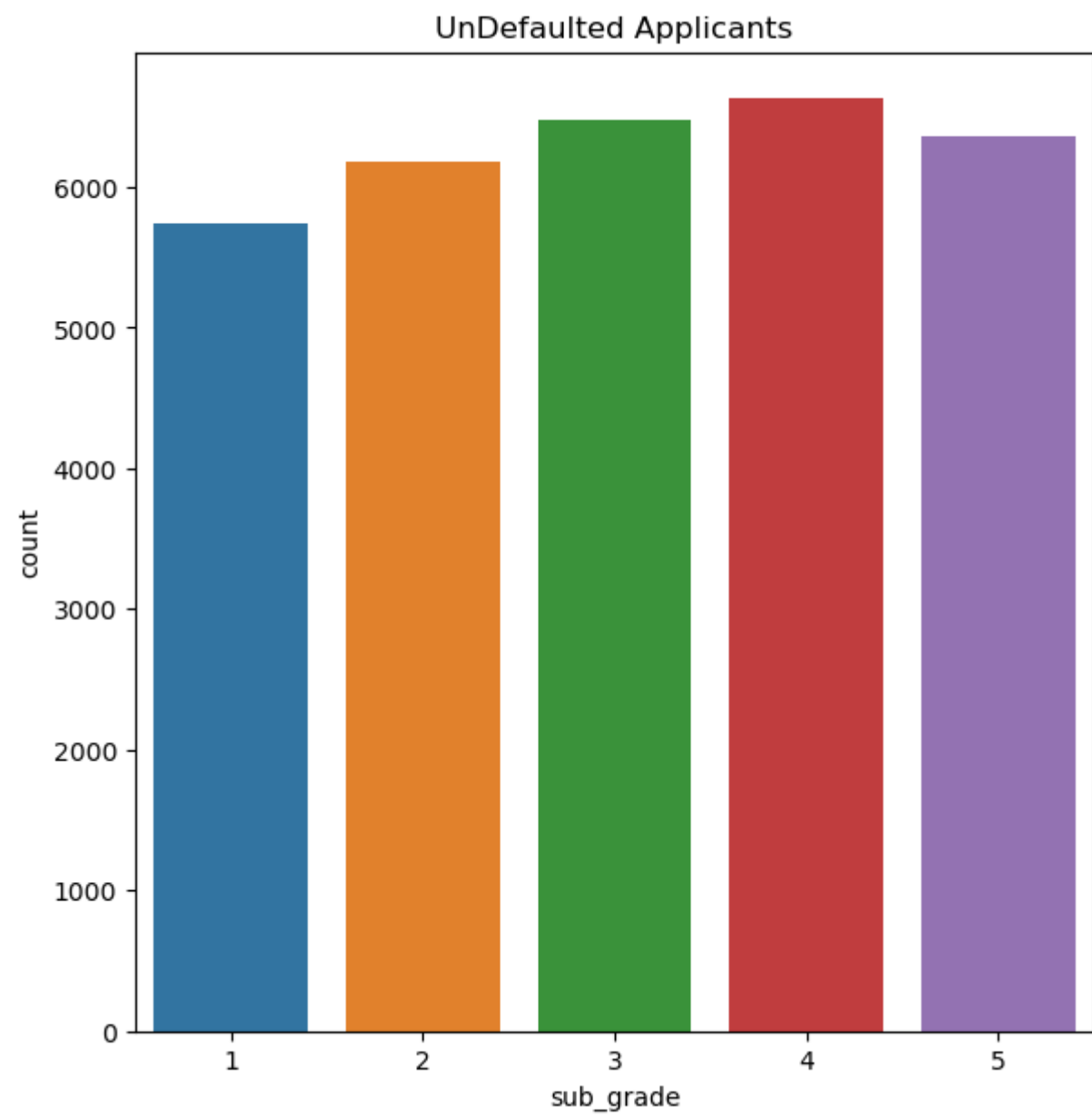
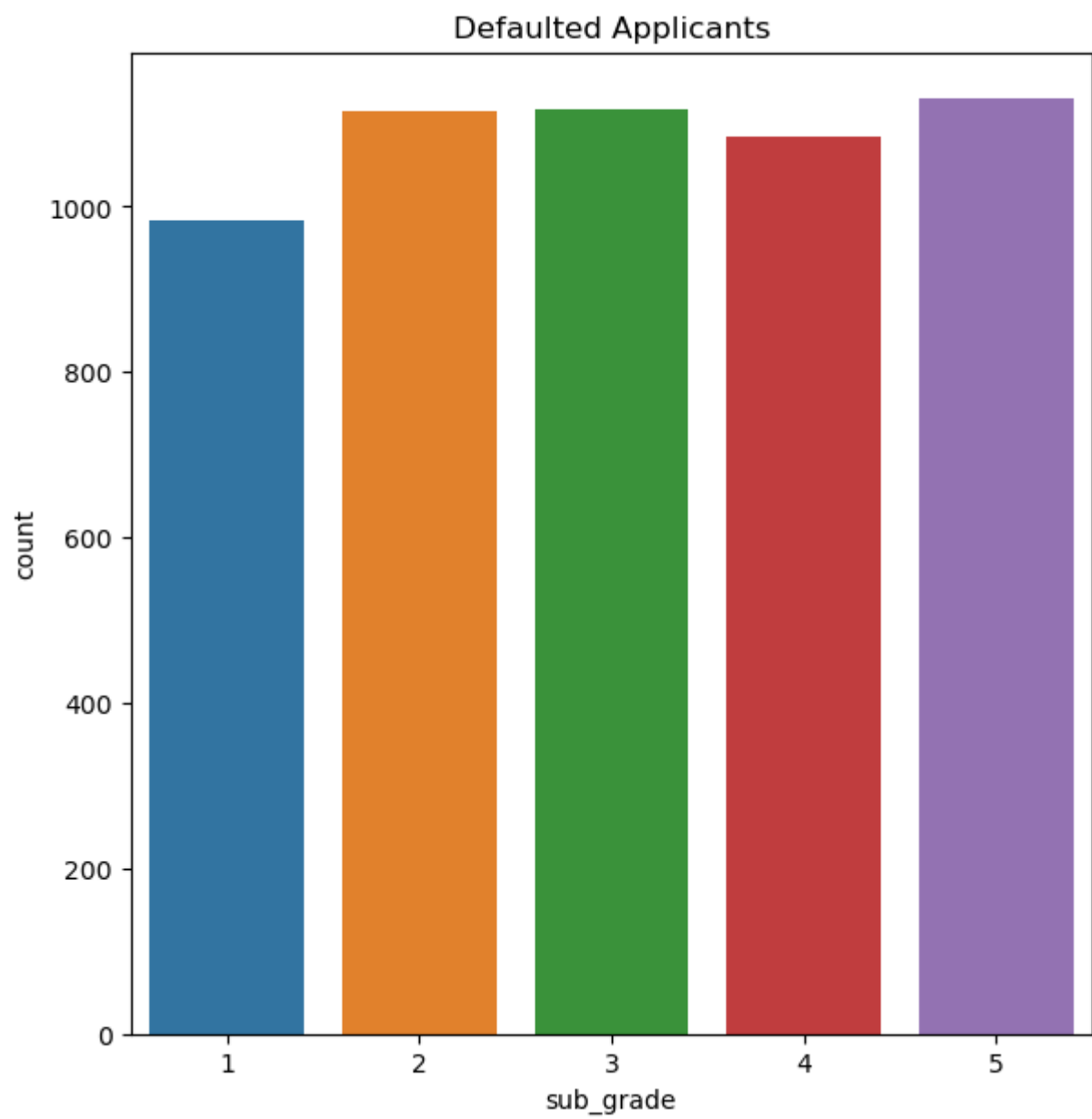


UnDefaulted Applicants

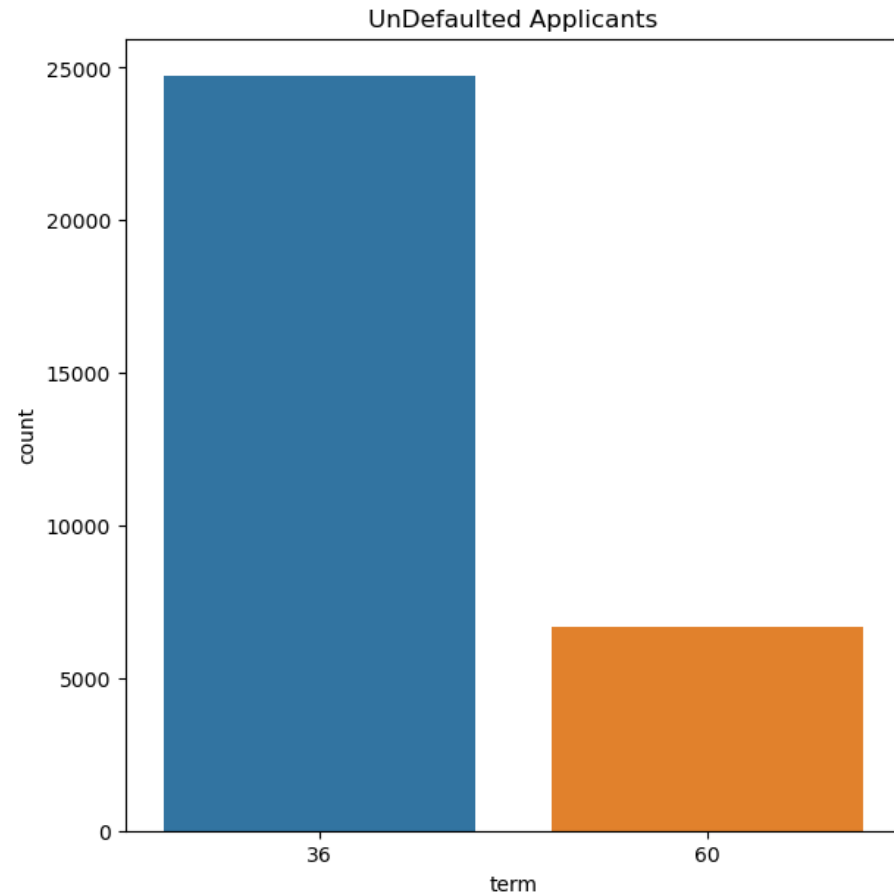
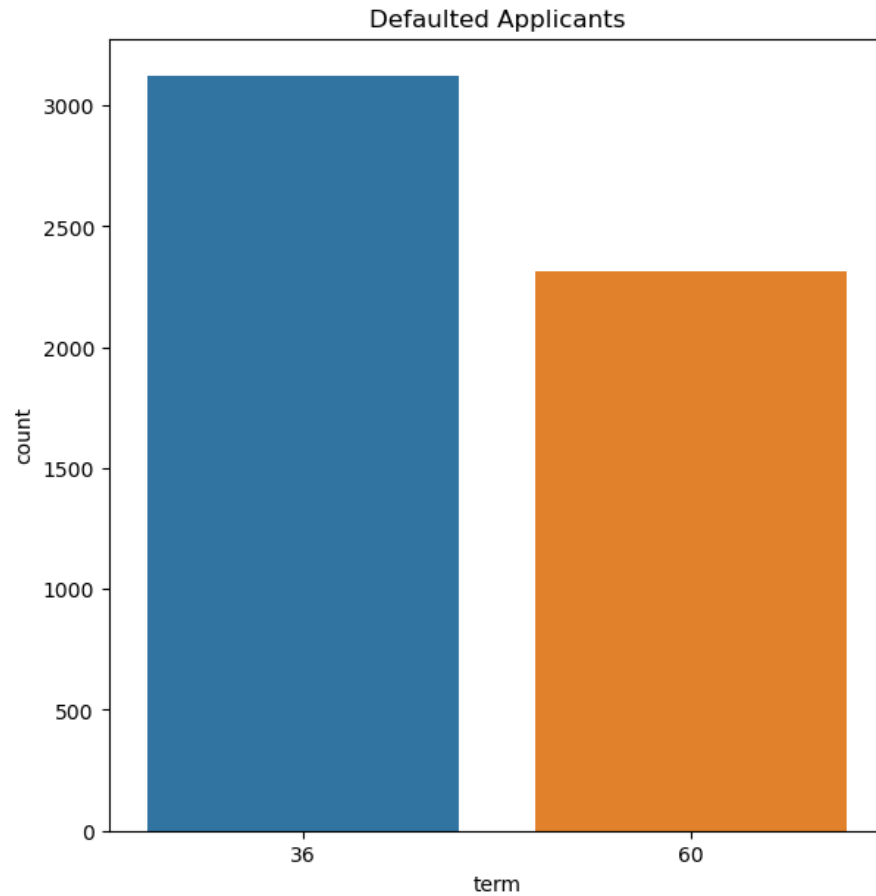


Categorical Variables



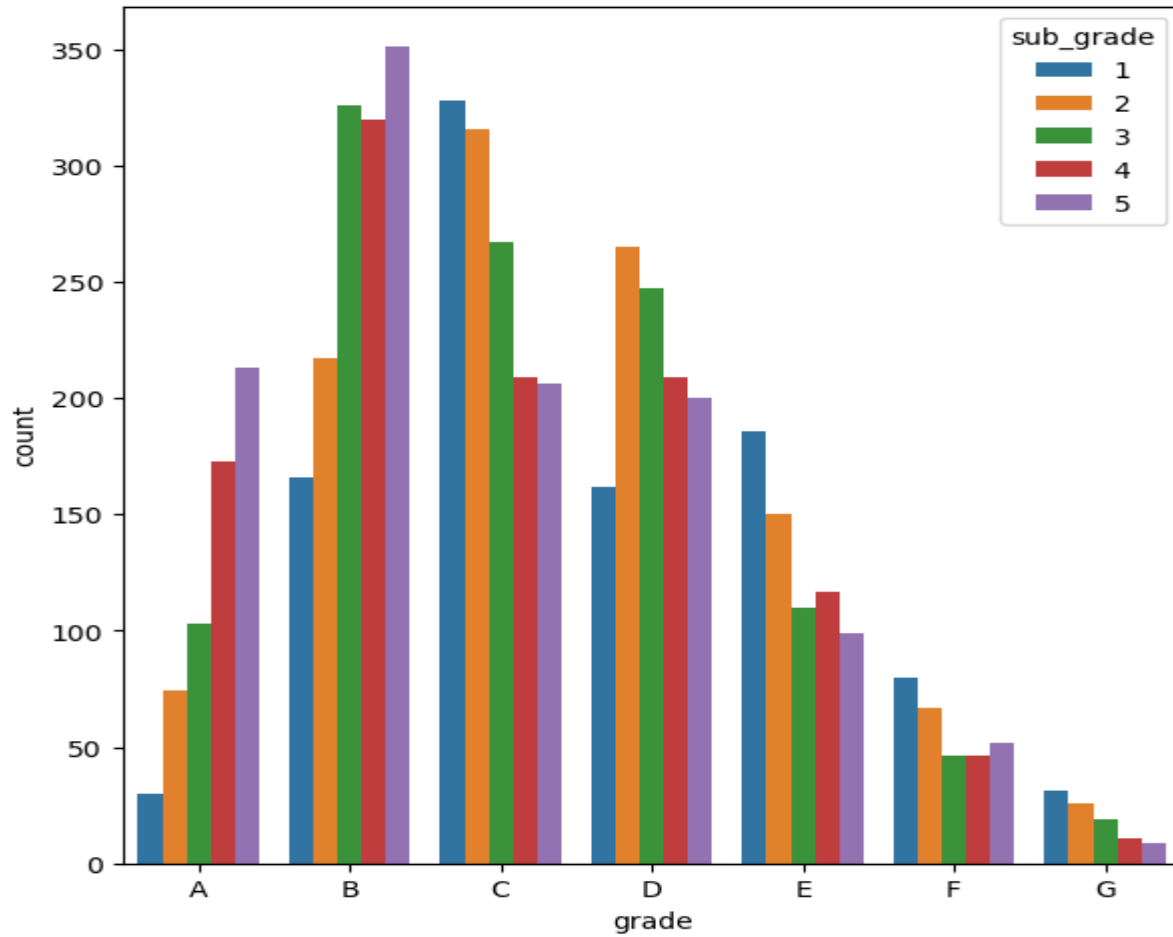


Categorical Variables

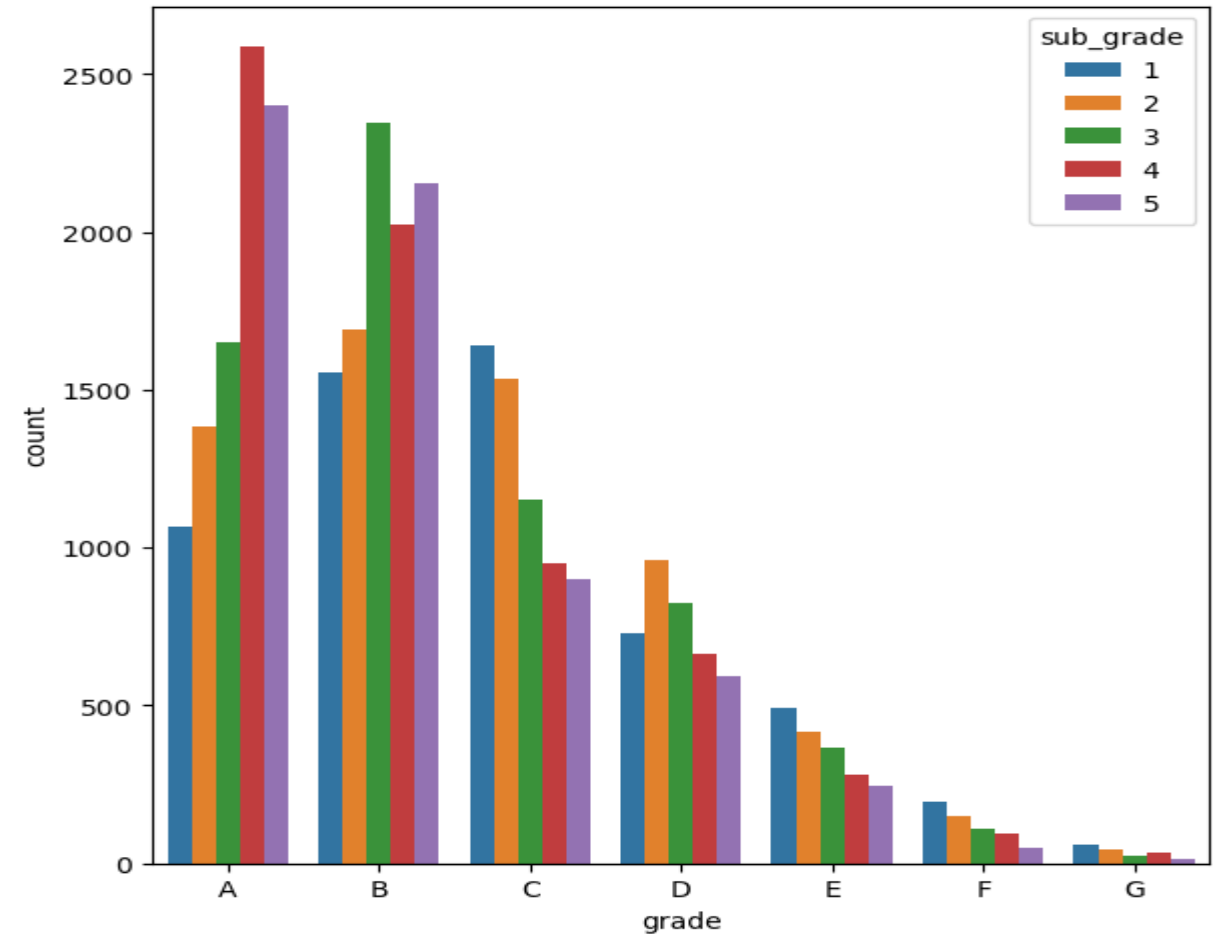


Categorical Variables

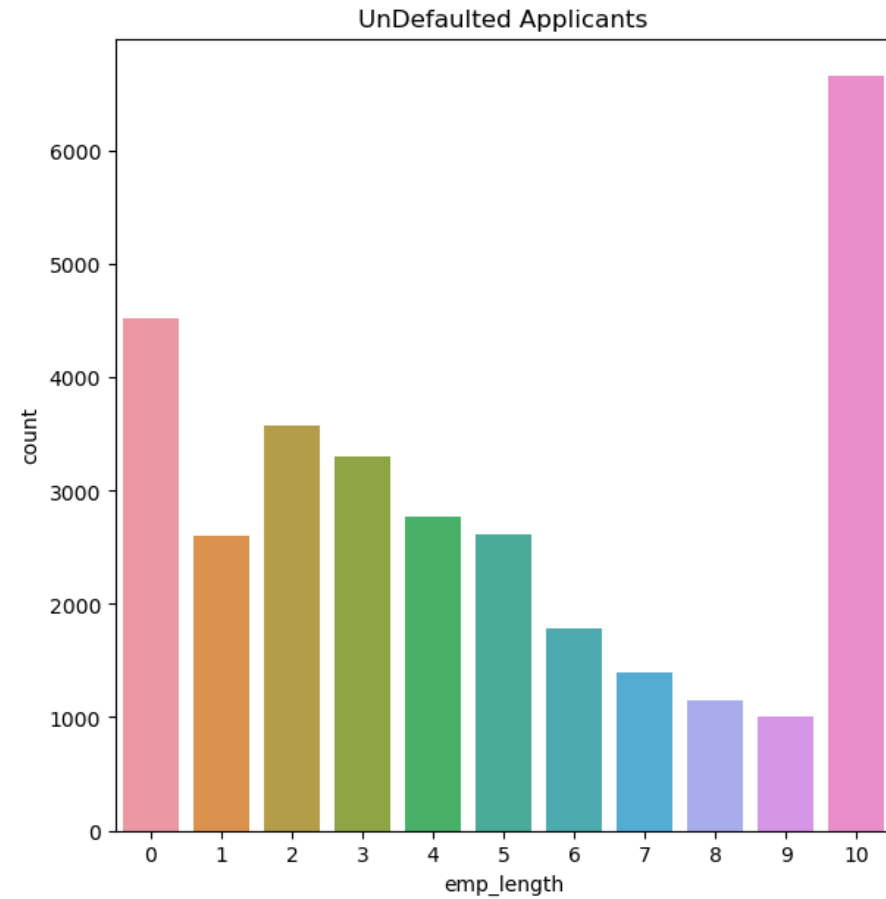
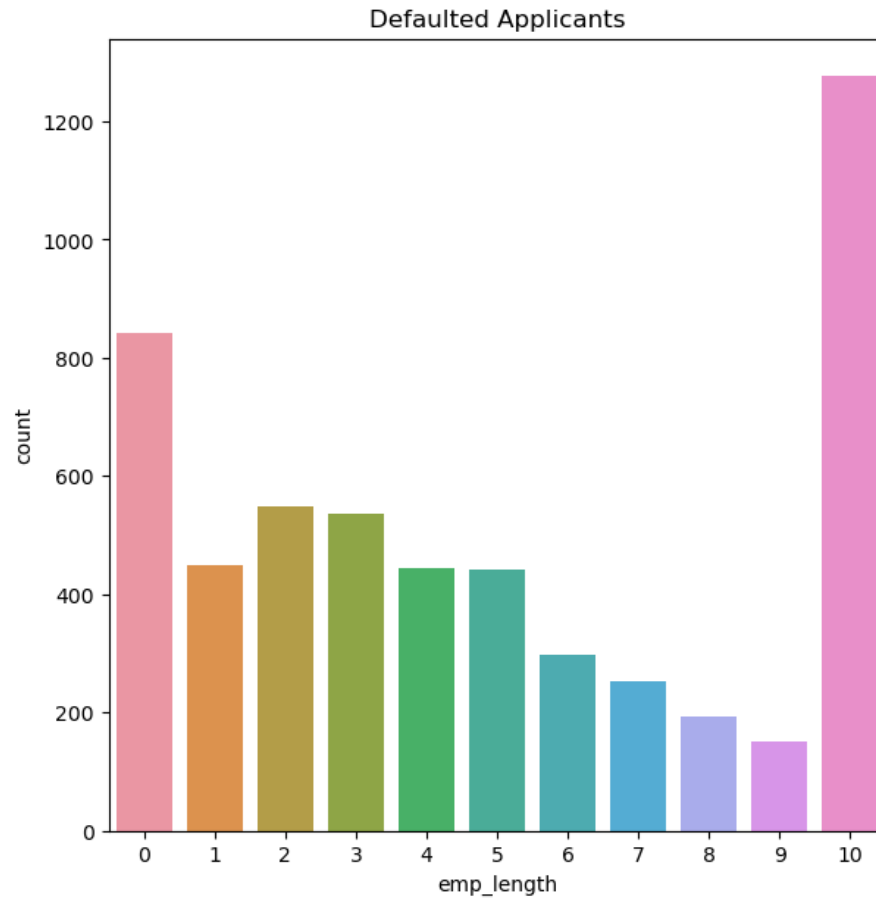
Defaulted Applicants



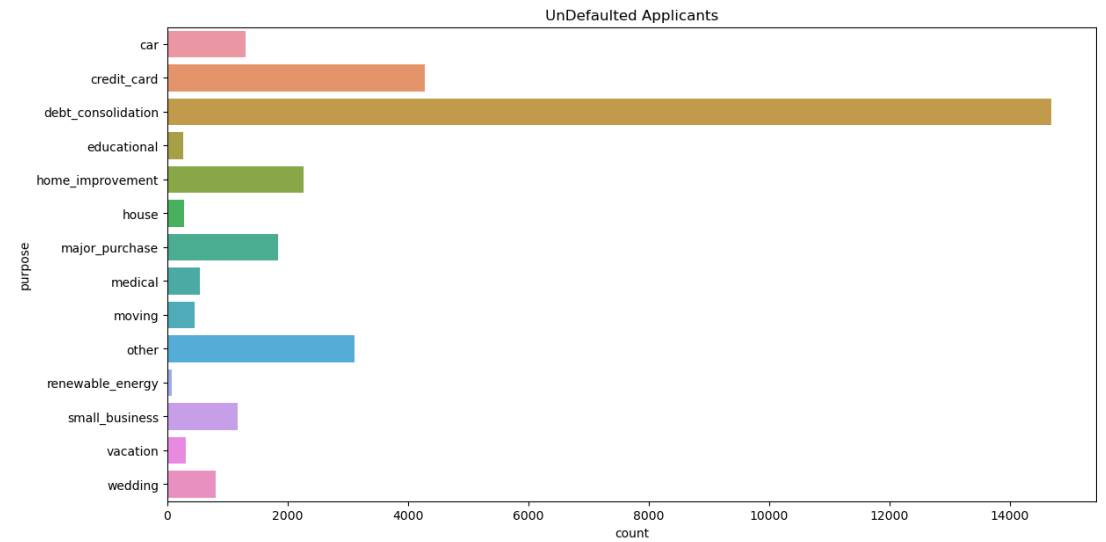
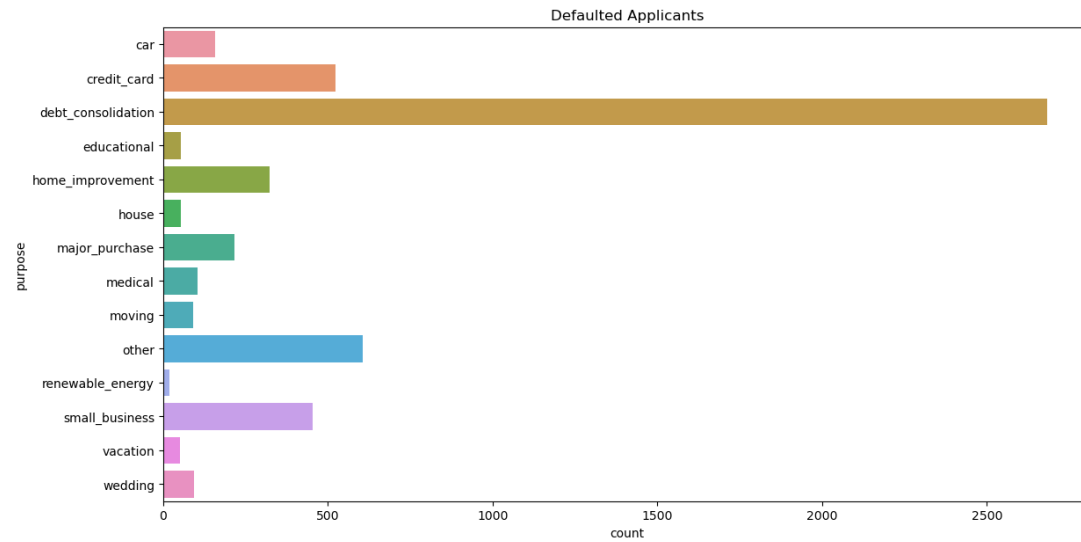
UnDefaulted Applicants



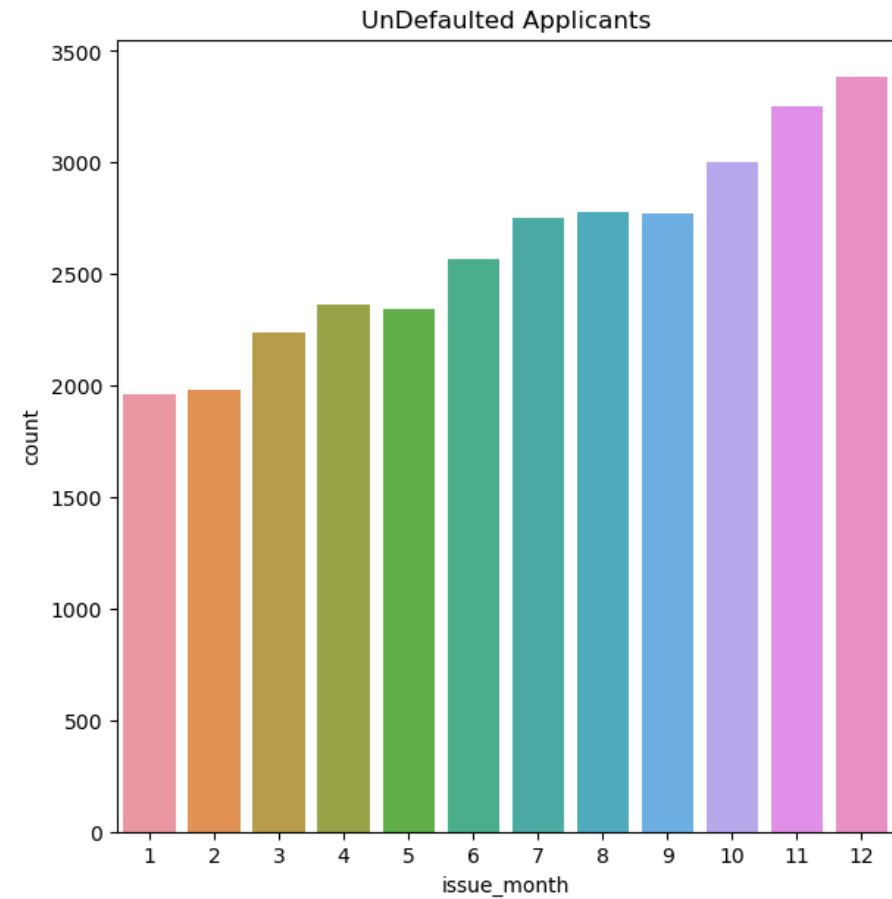
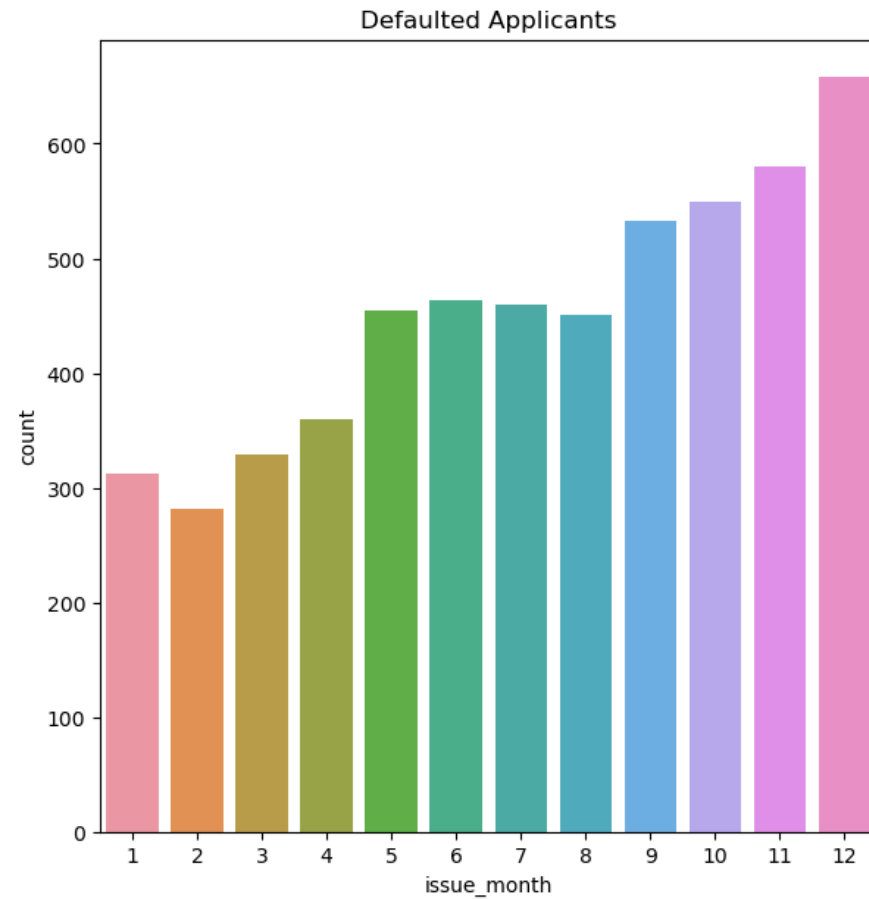
Categorical Variables



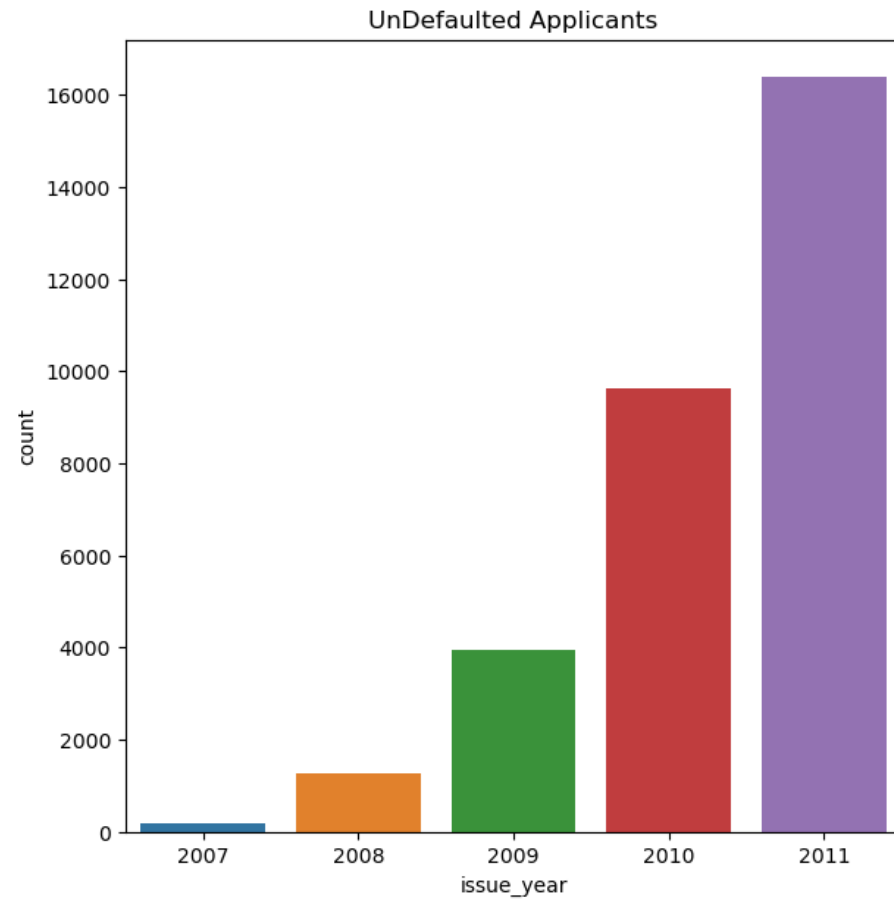
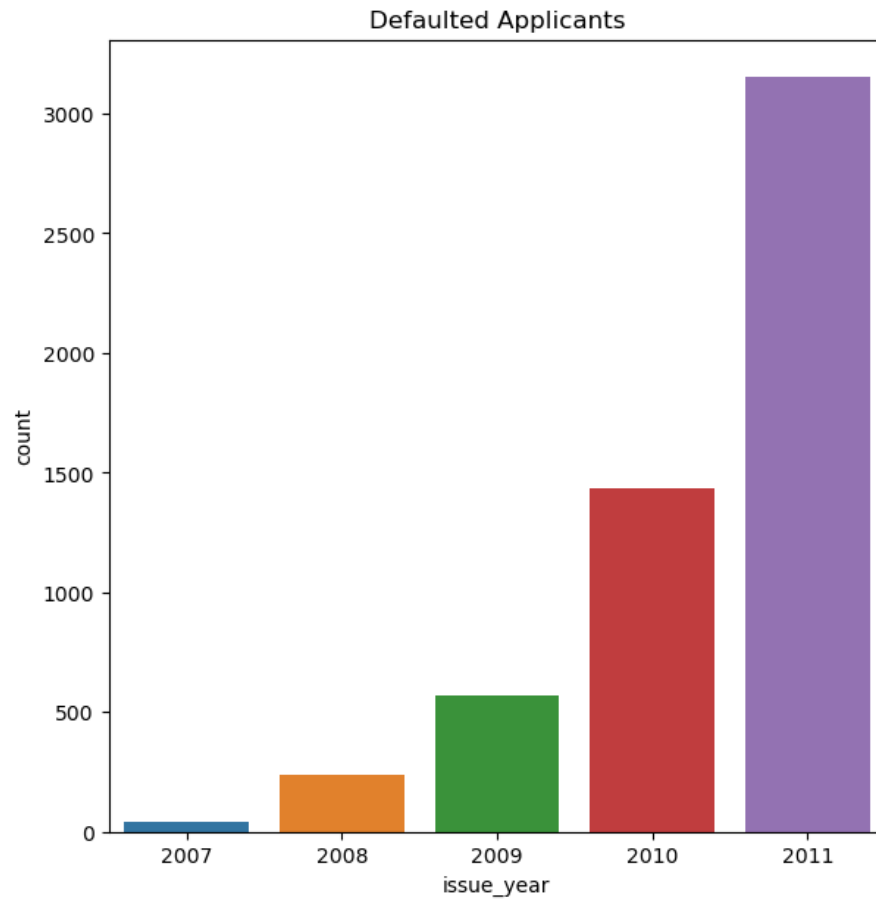
Categorical Variables



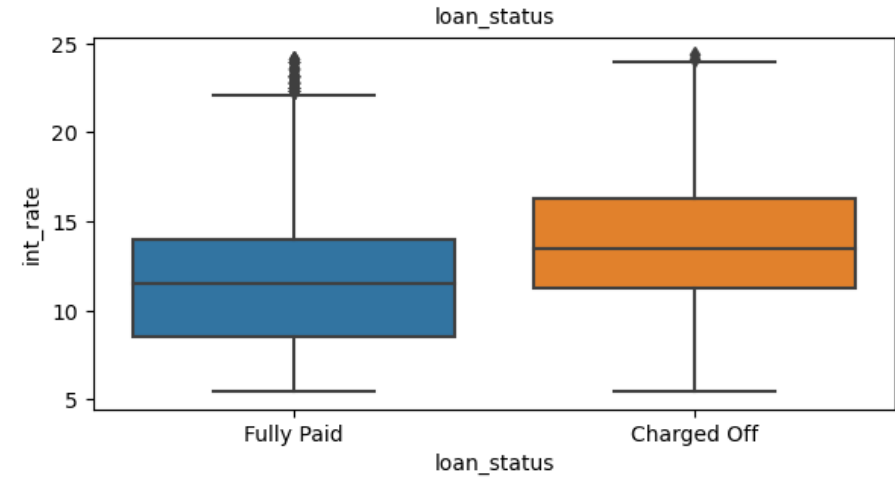
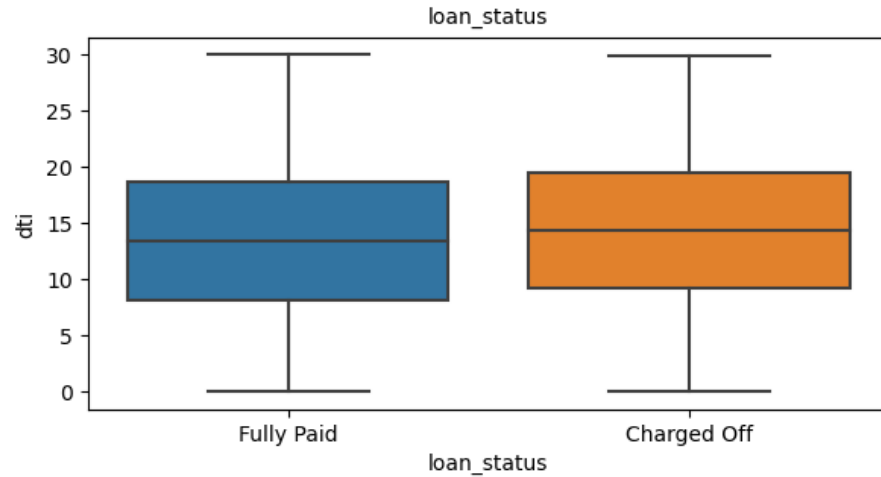
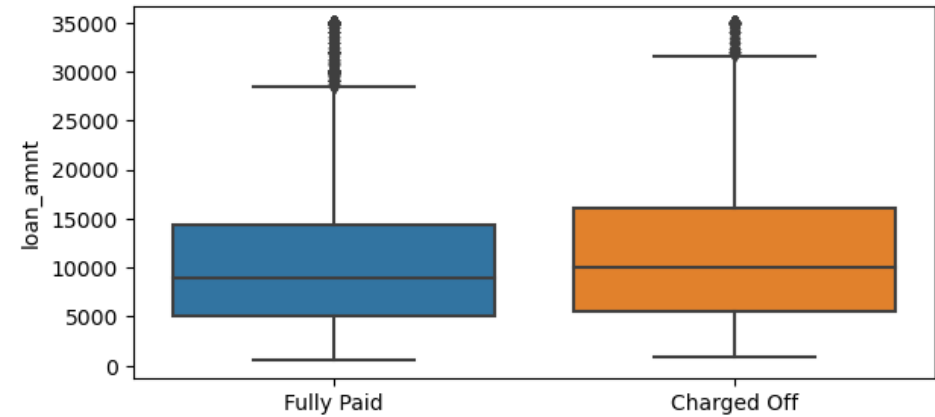
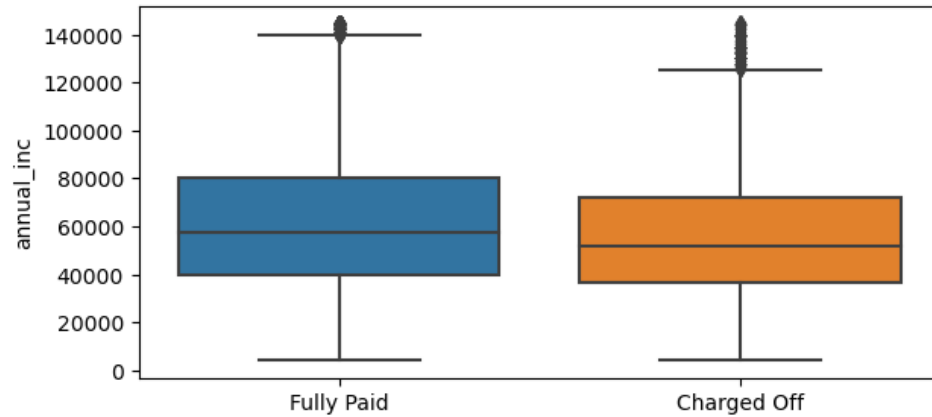
Categorical Variables



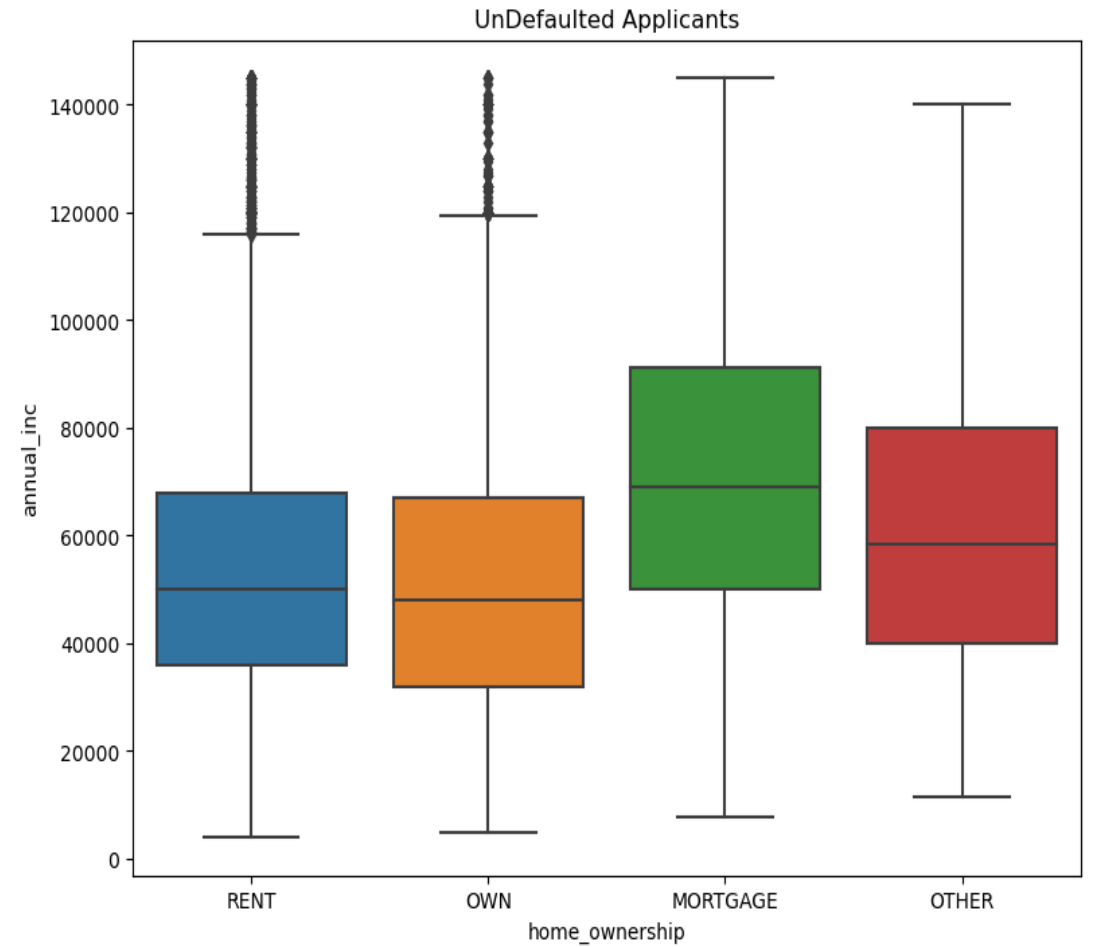
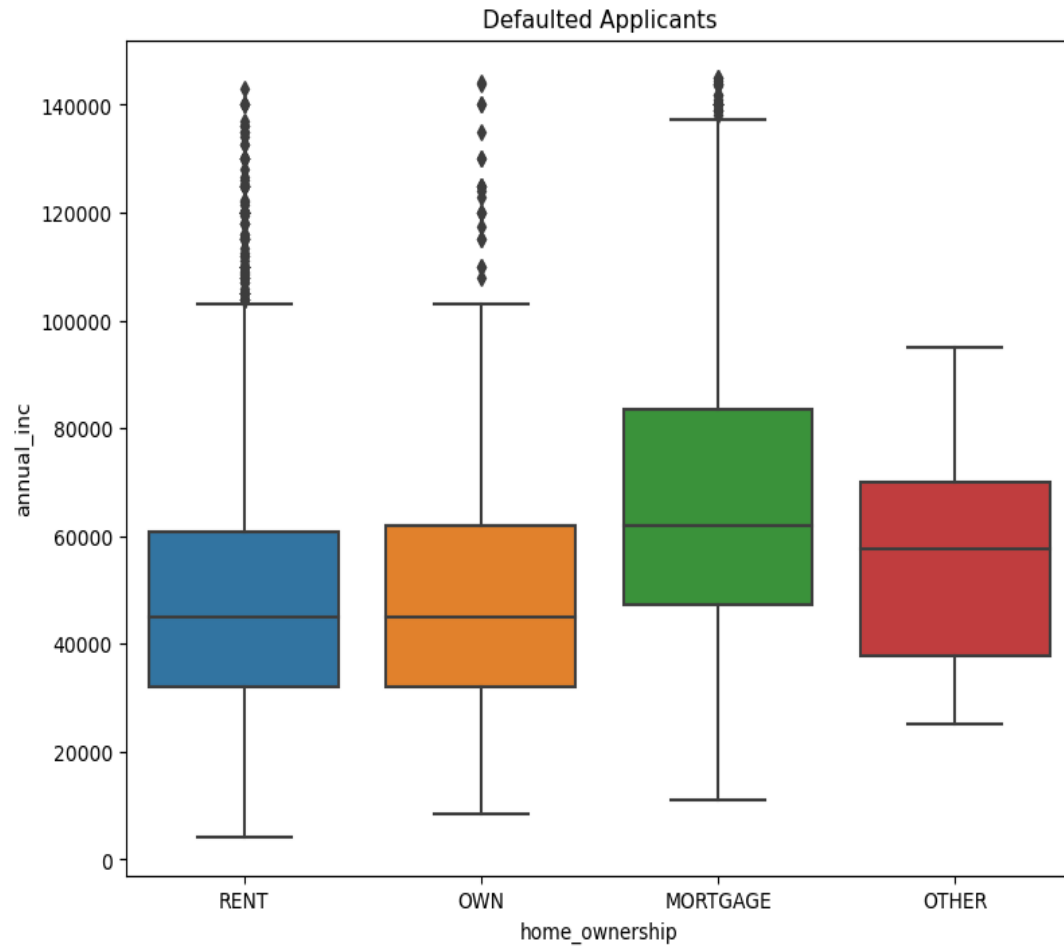
Categorical Variables



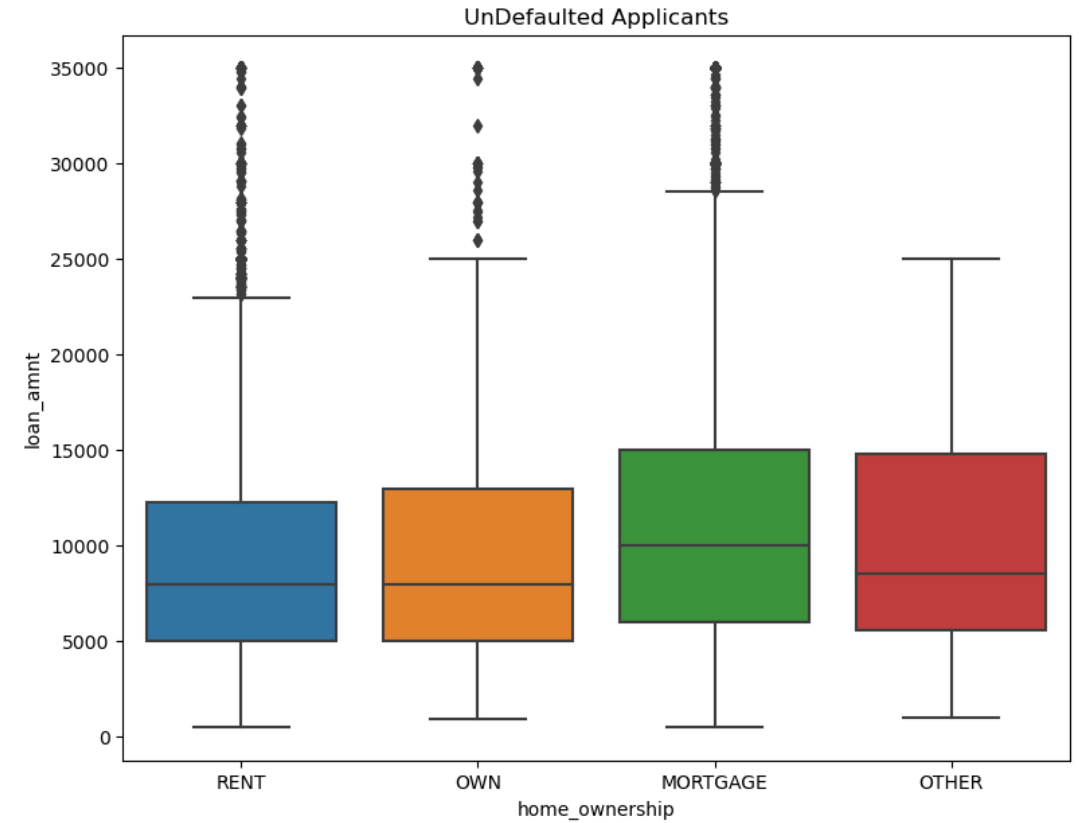
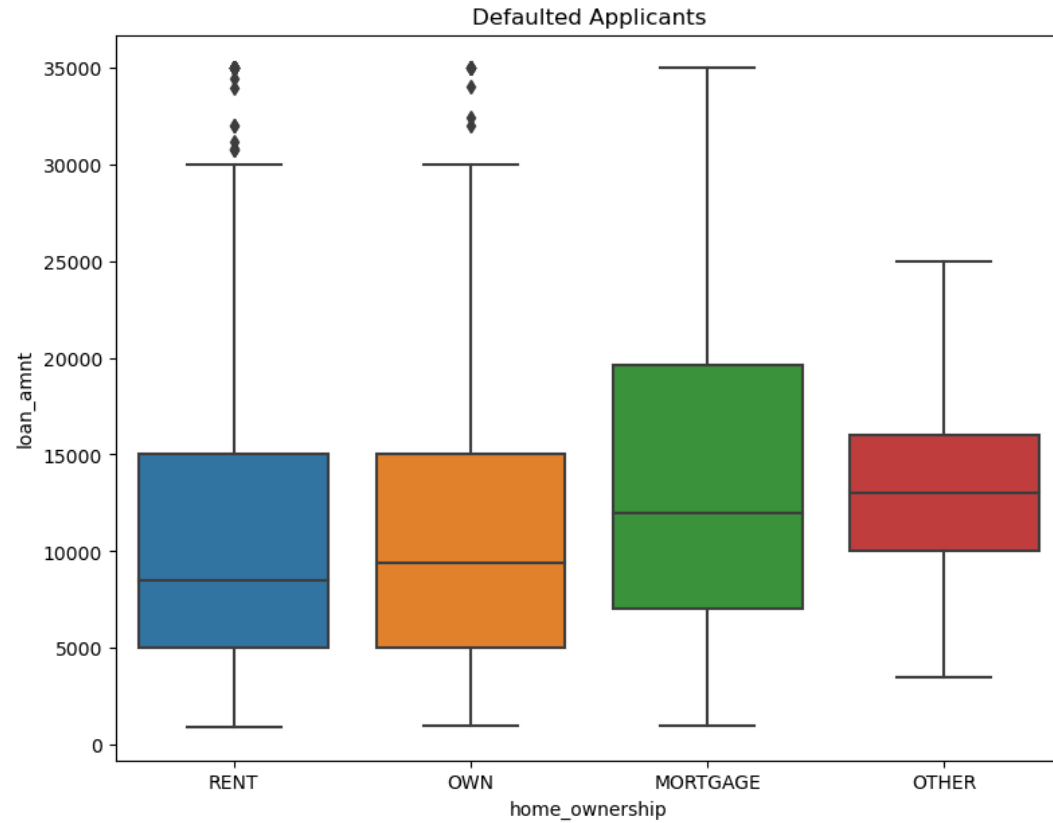
Quantitative Variables



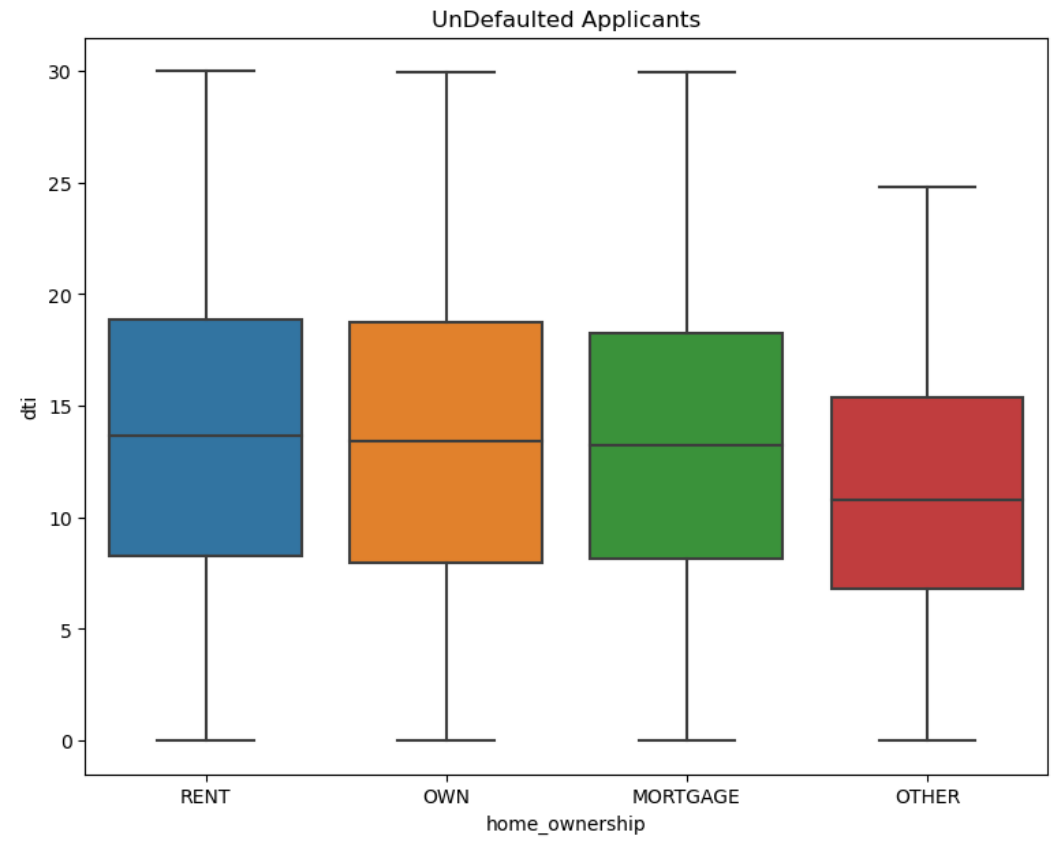
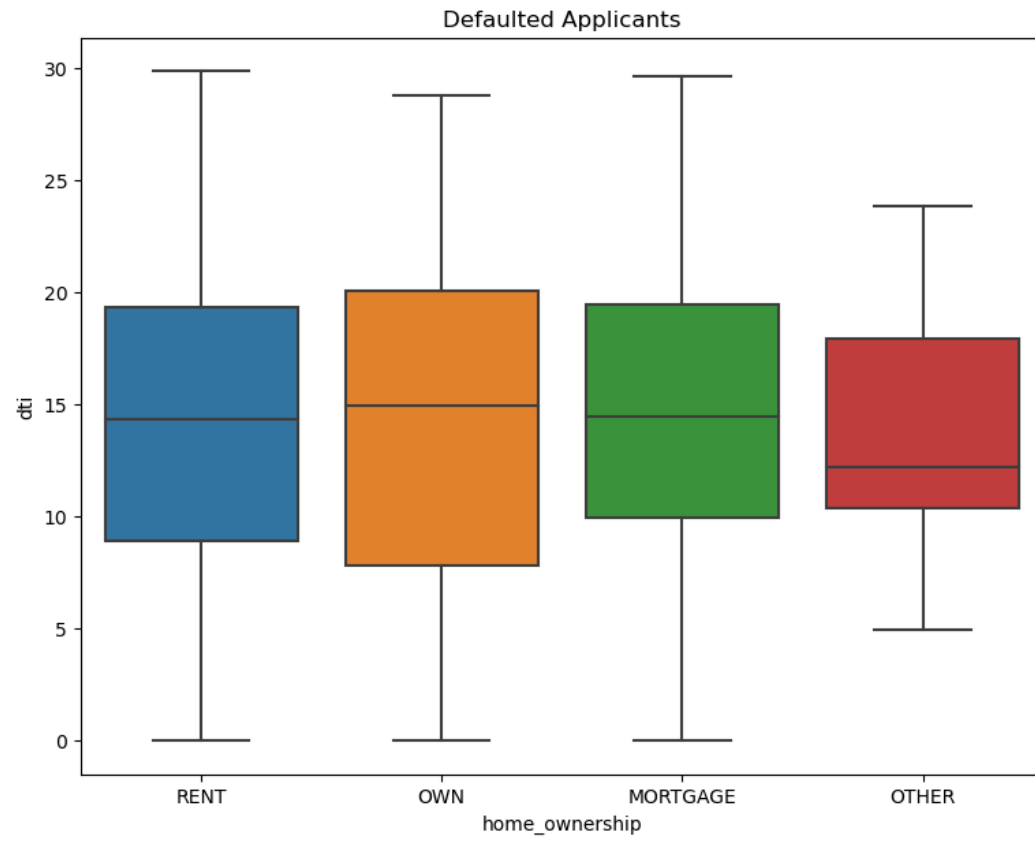
Quantitative Variables



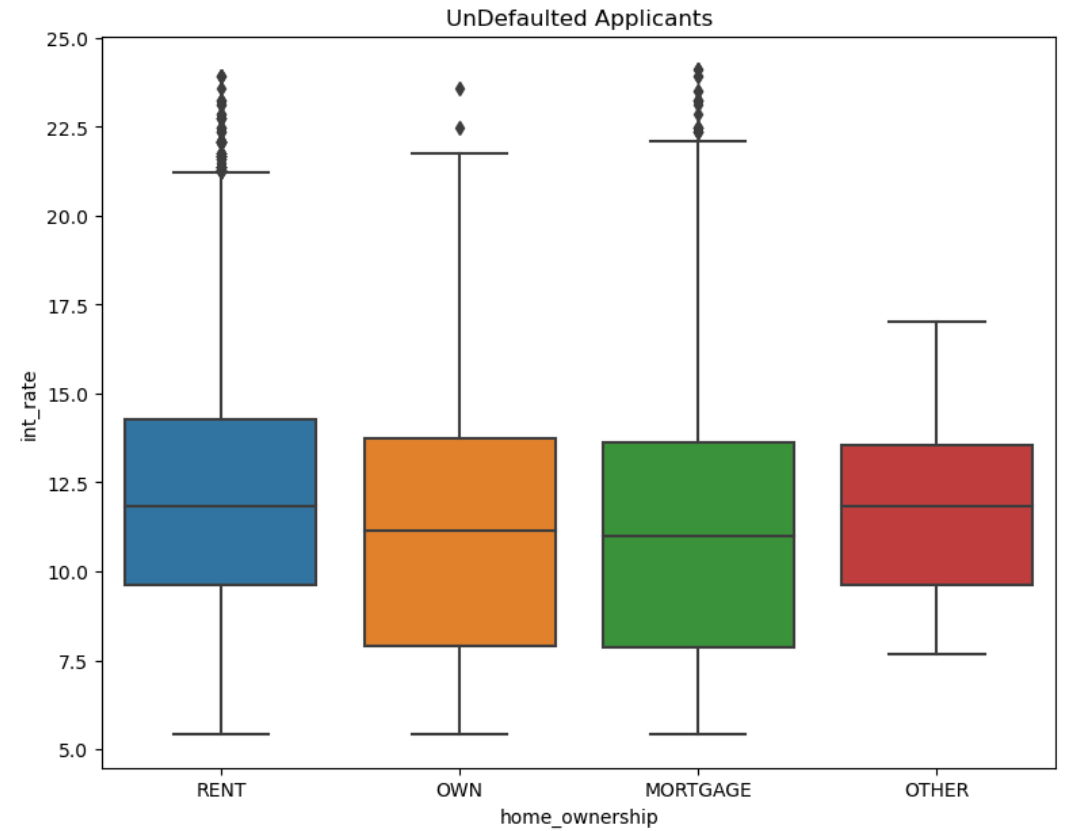
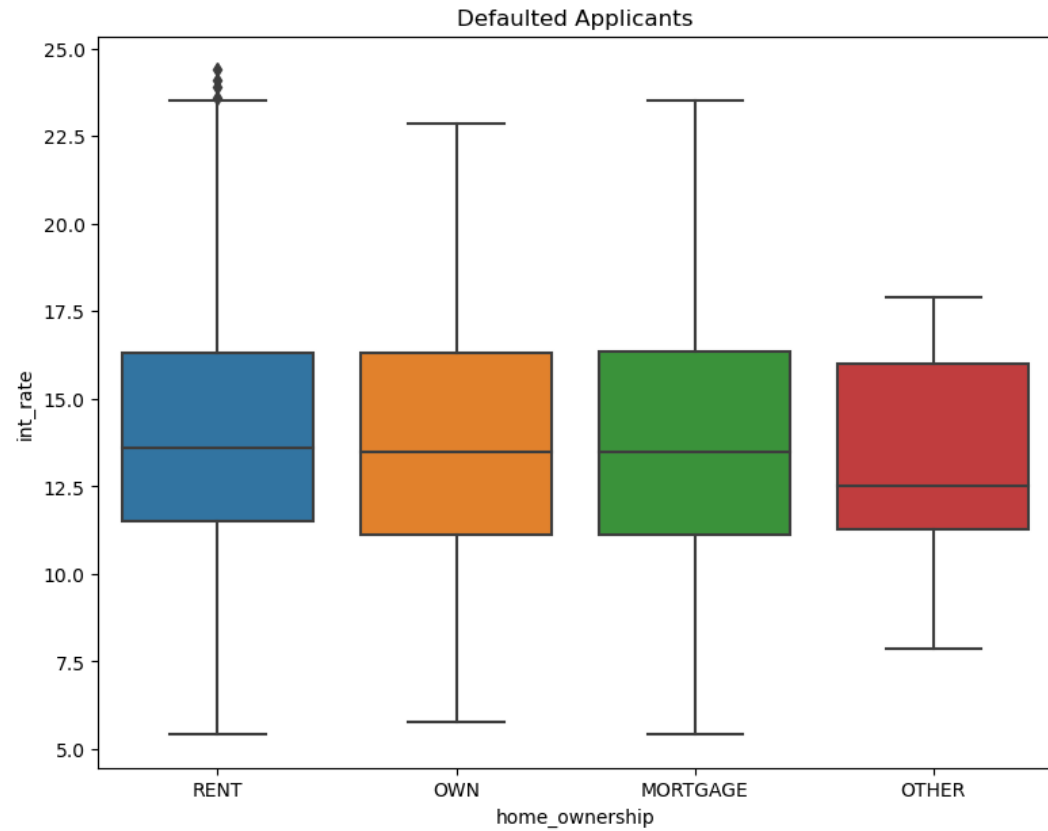
Quantitative Variables



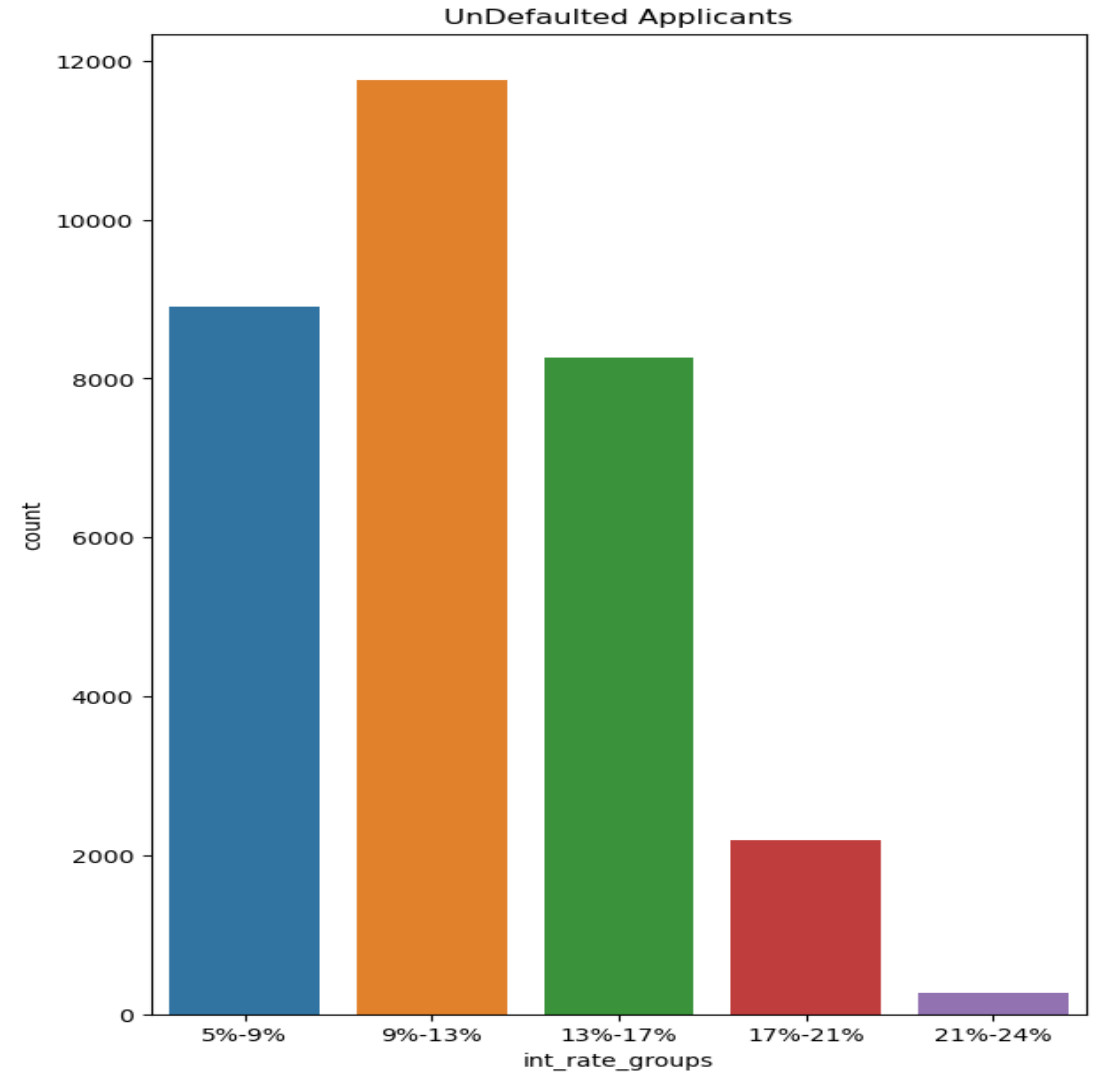
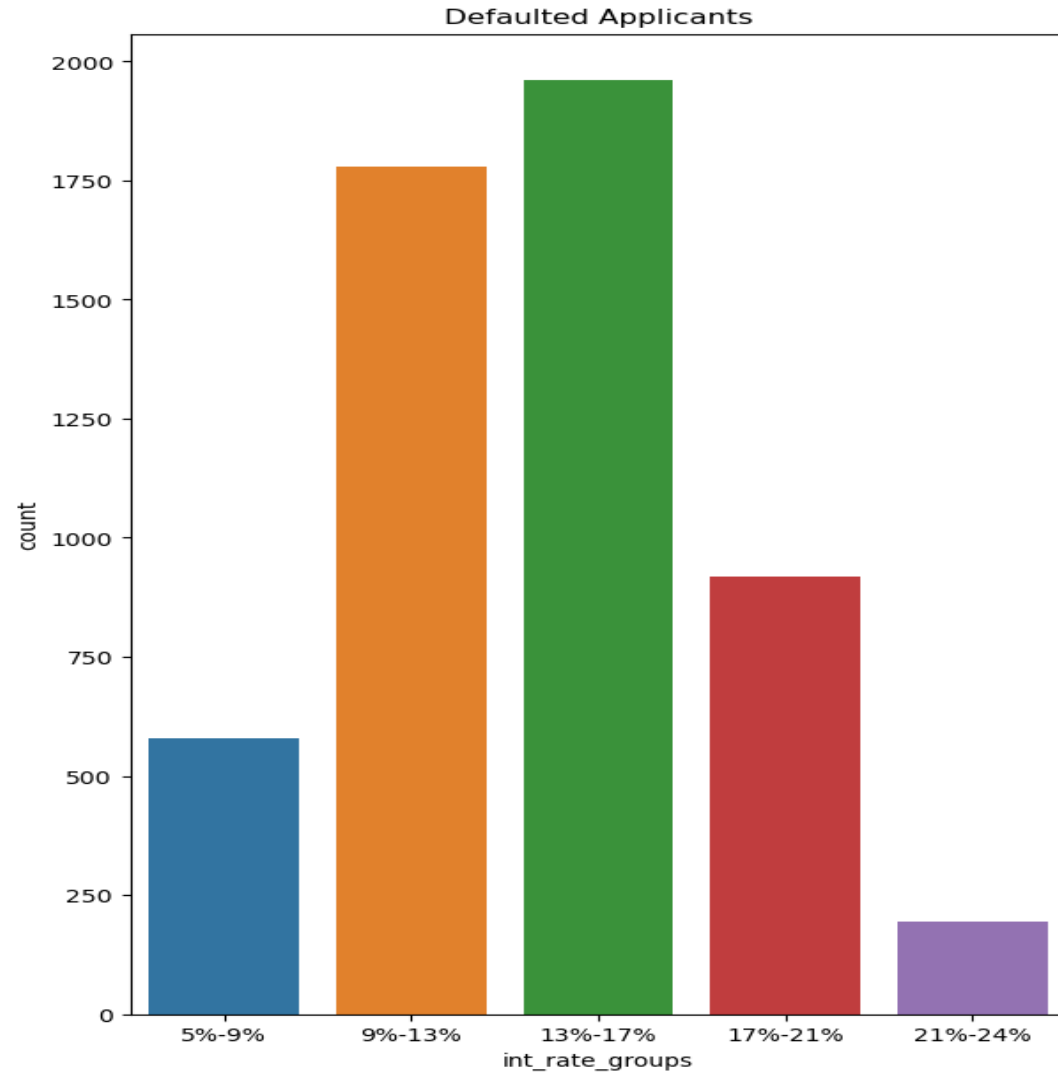
Quantitative Variables



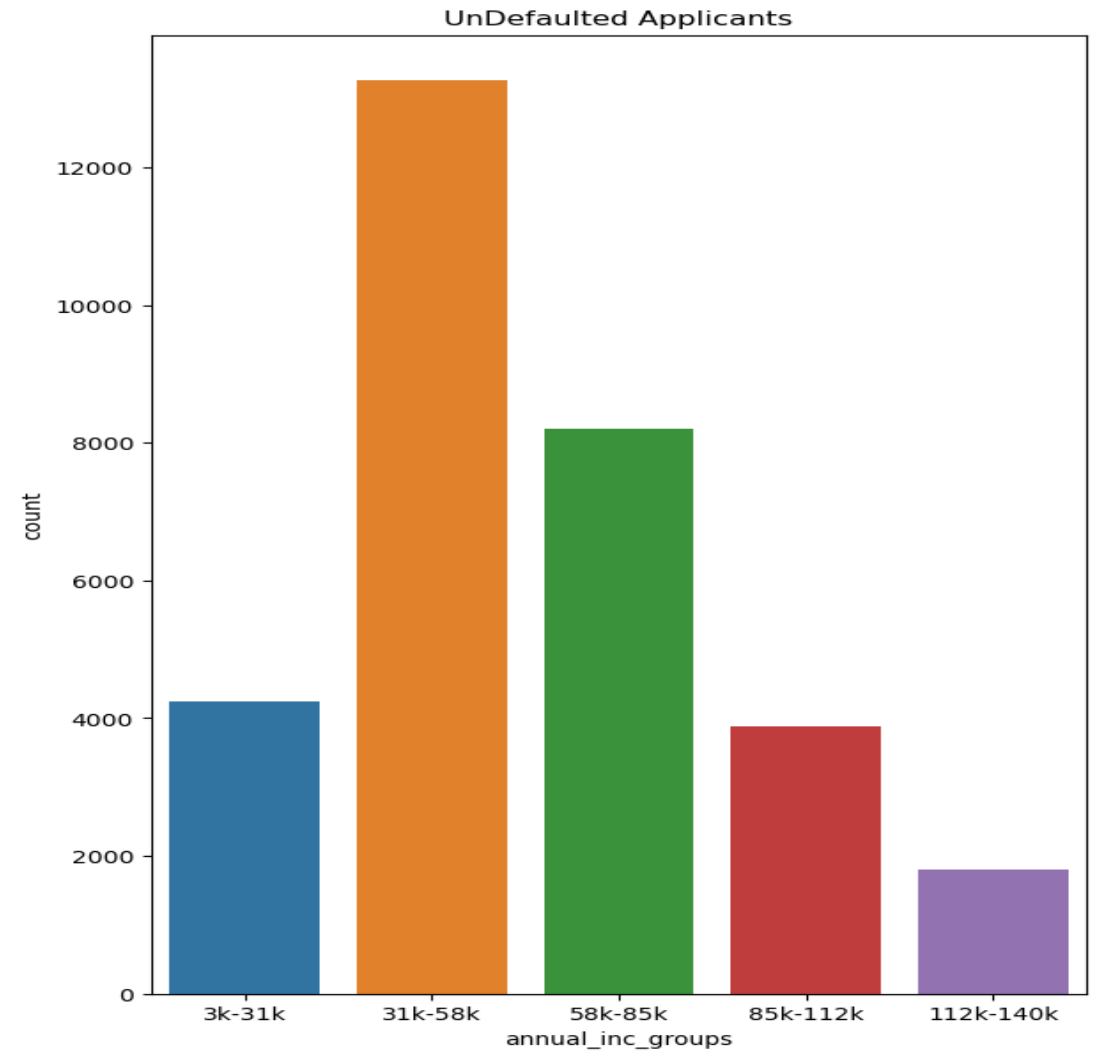
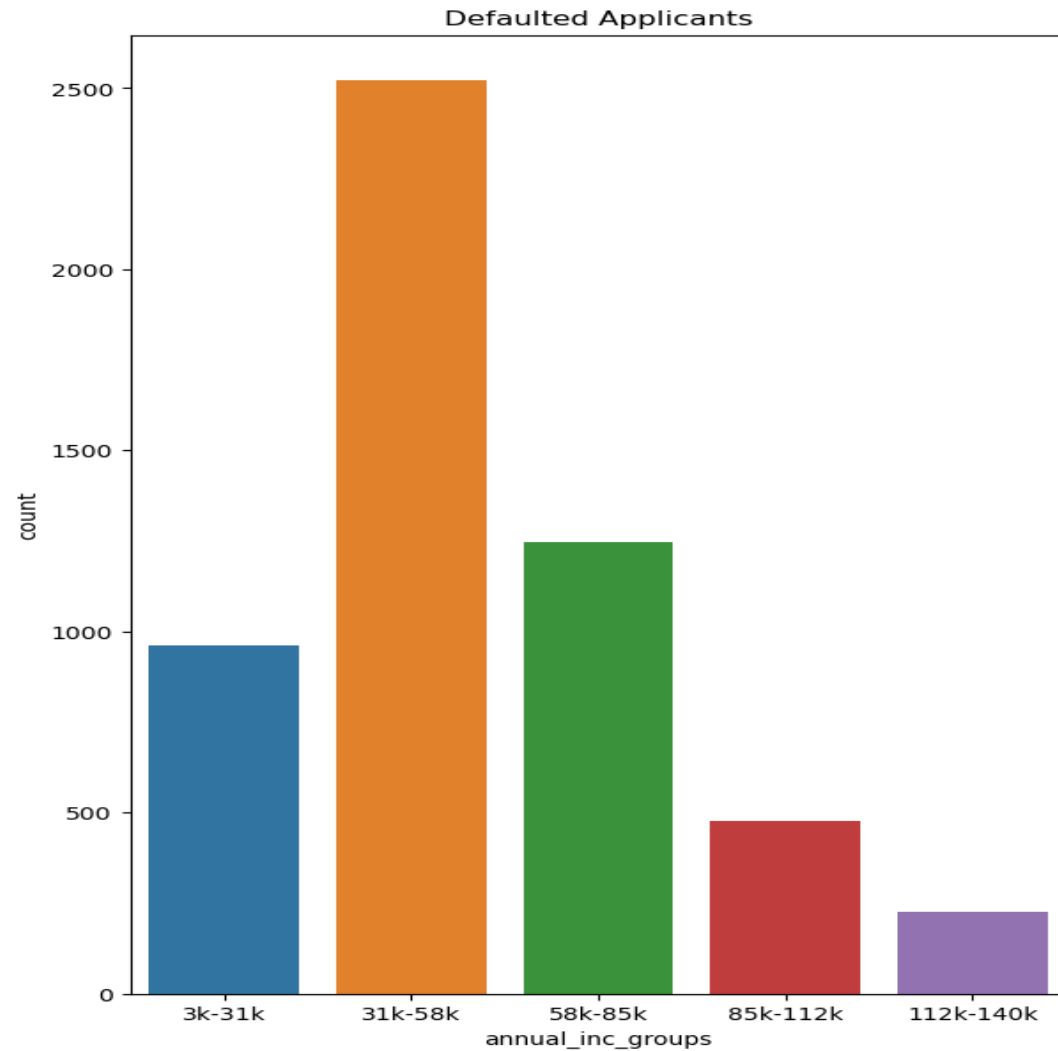
Quantitative Variables



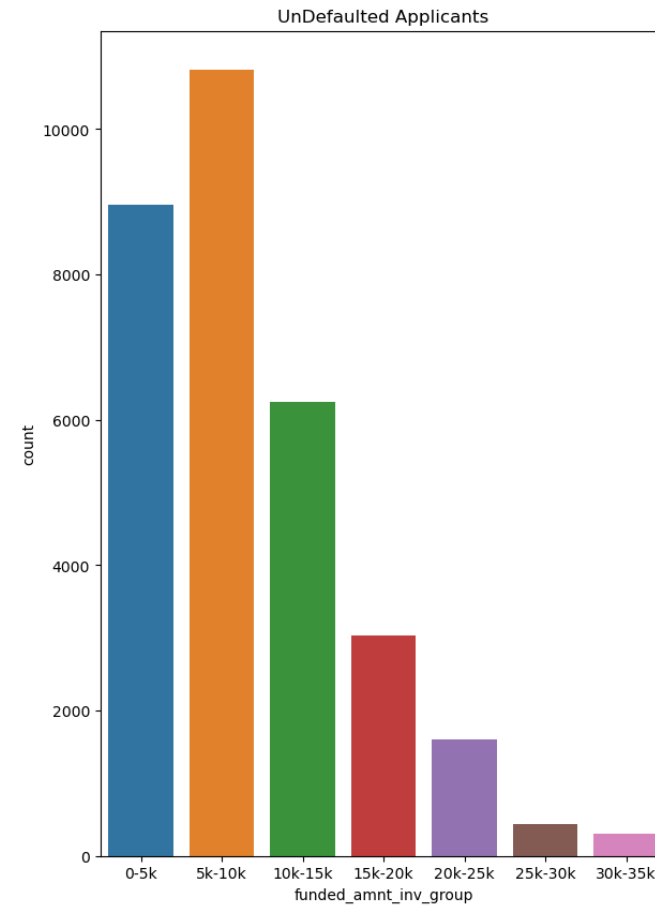
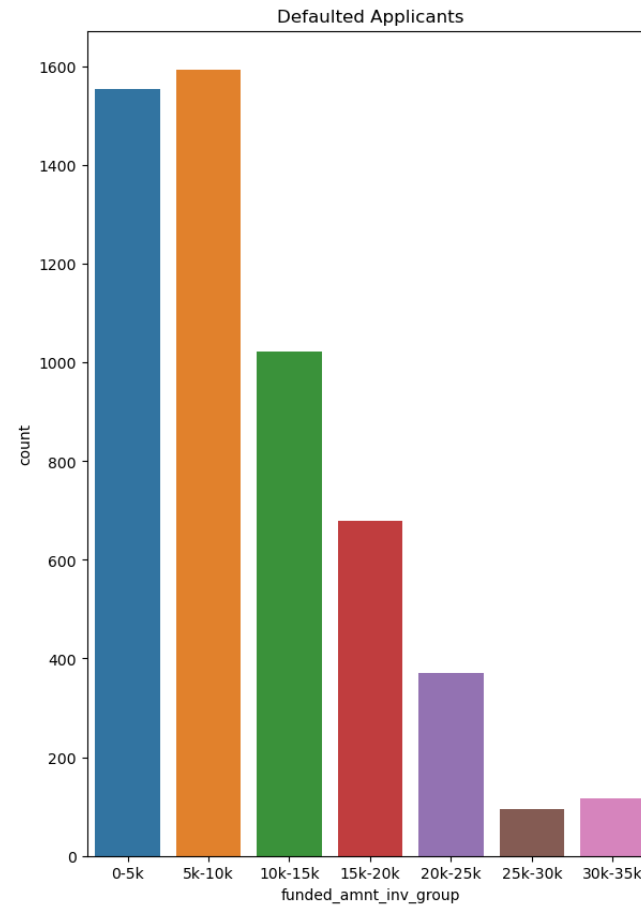
Quantitative Variables



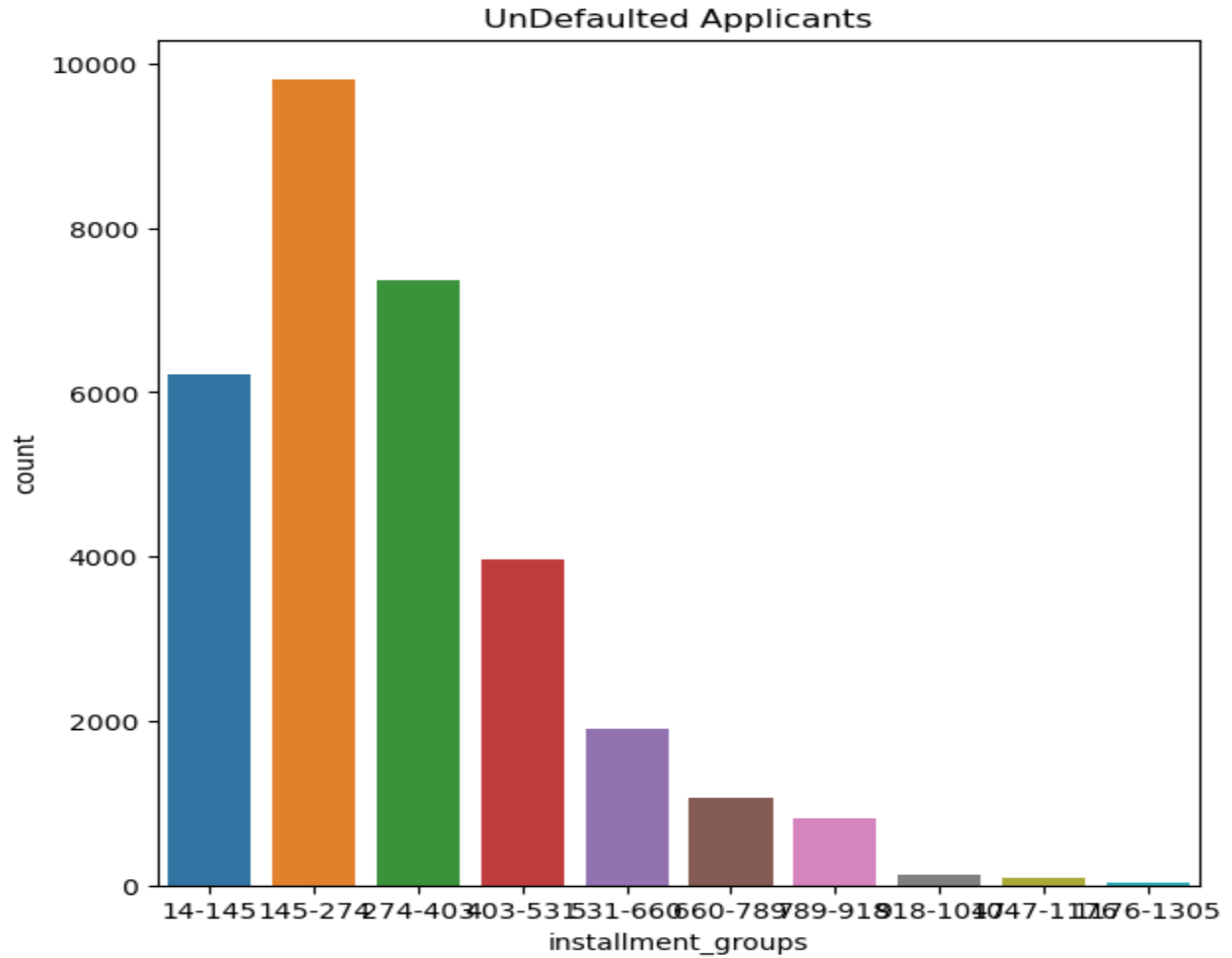
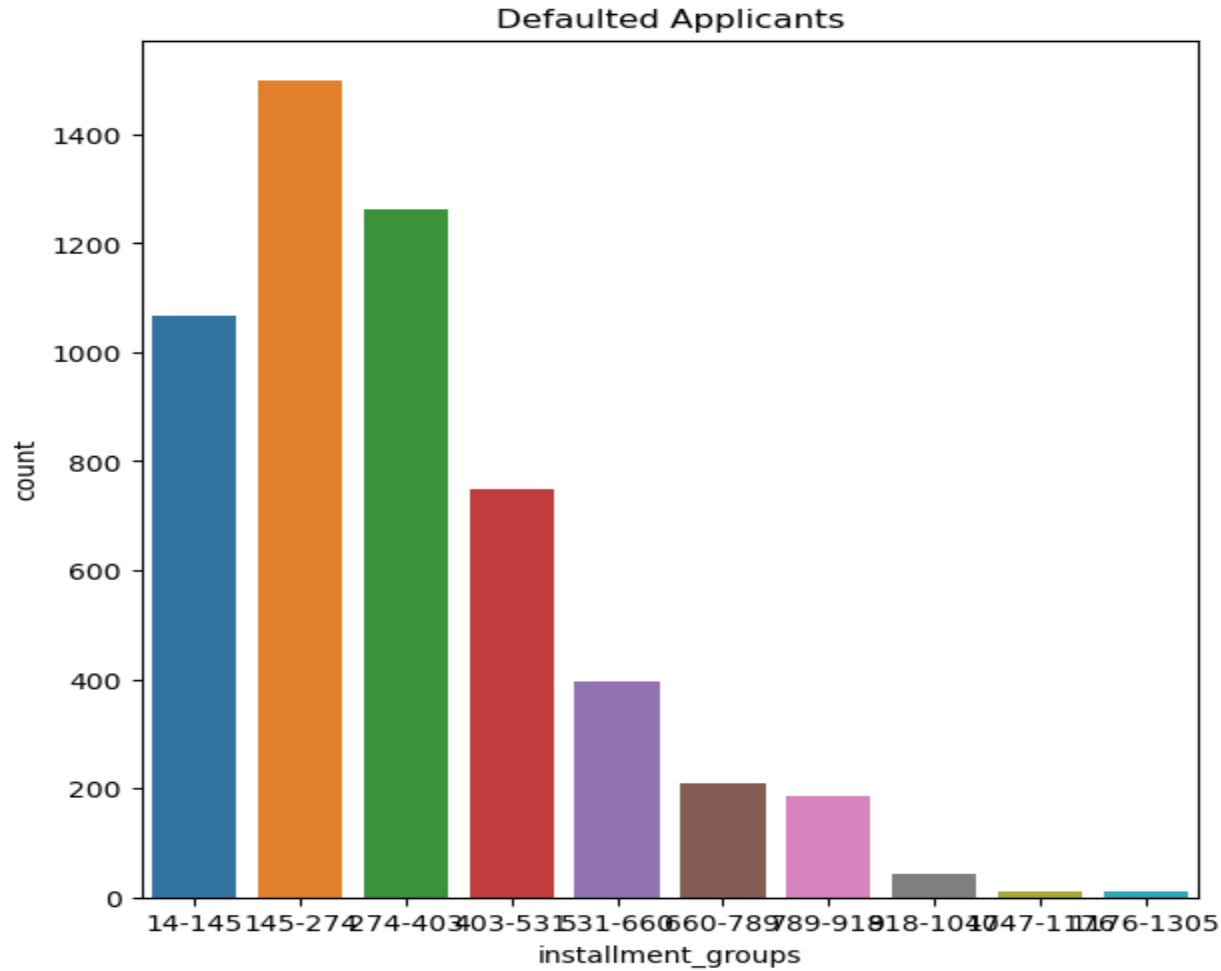
Quantitative Variables



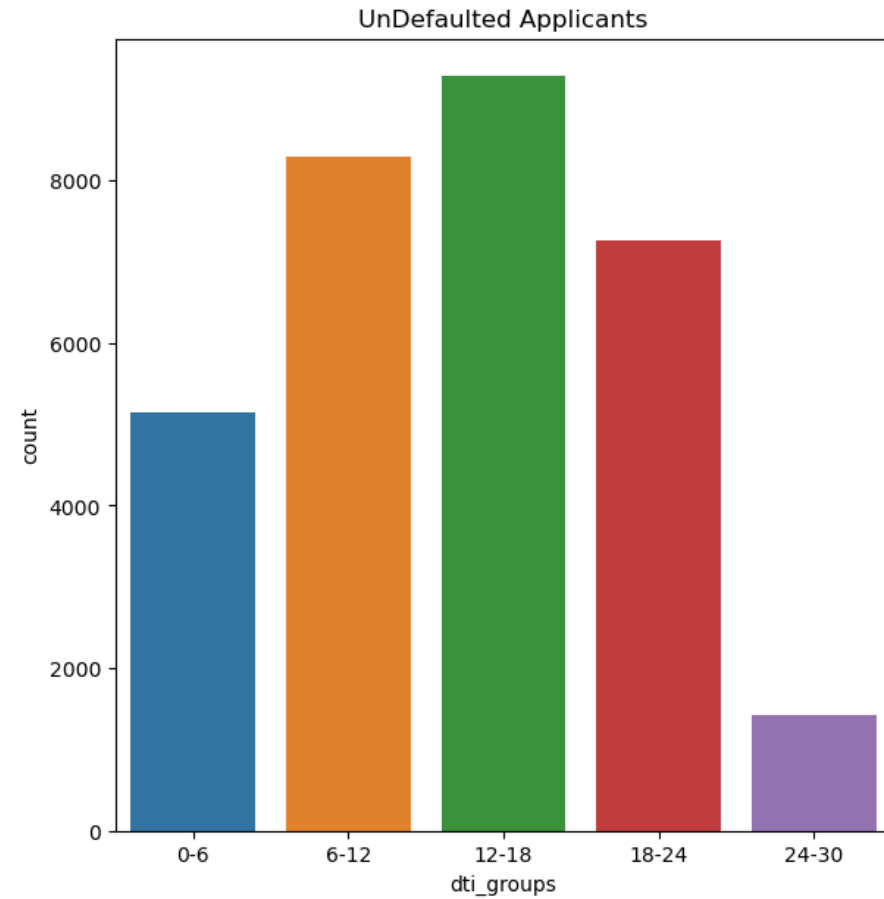
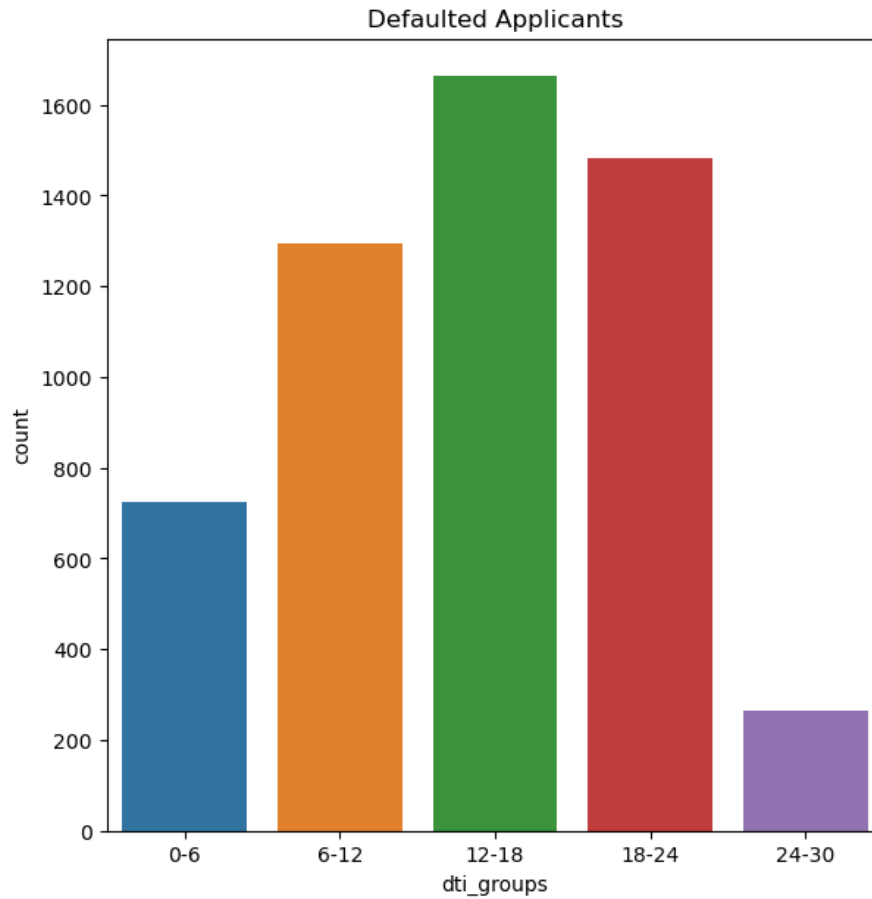
Quantitative Variables



Quantitative Variables



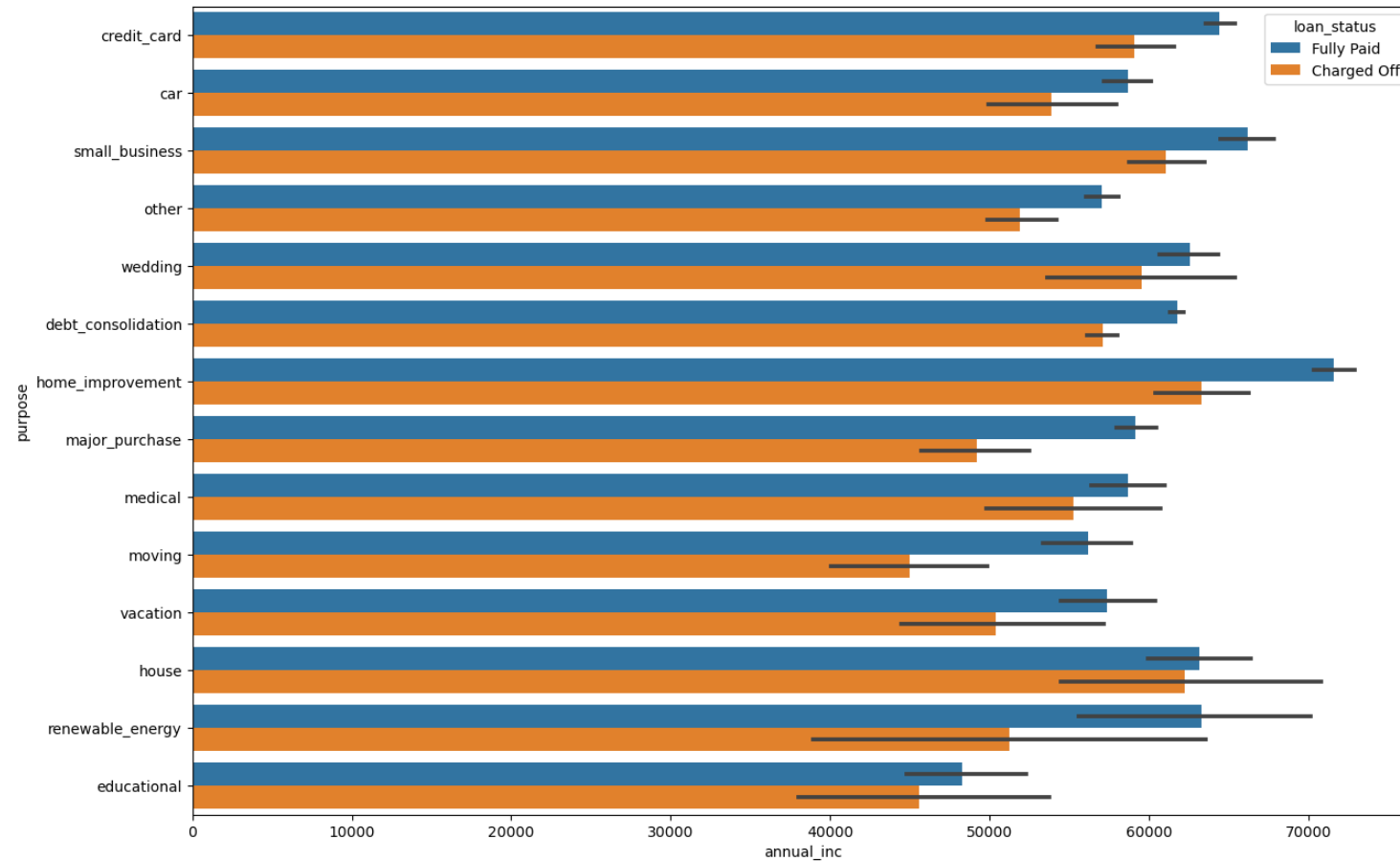
Quantitative Variables



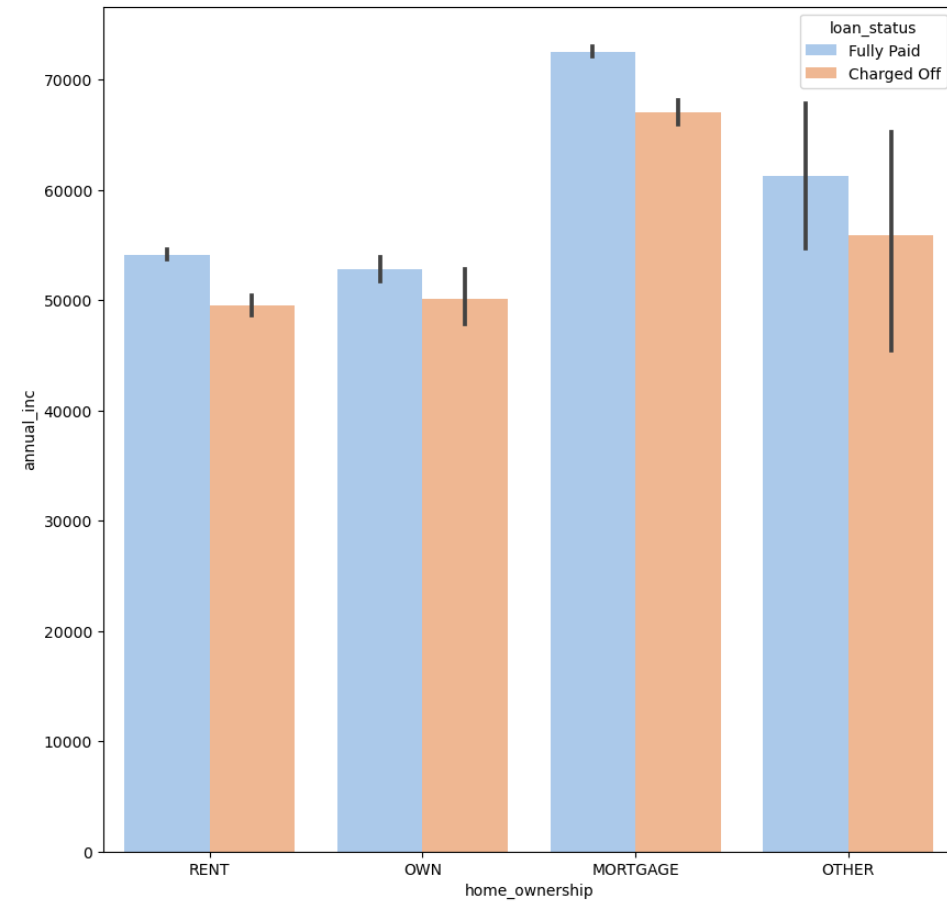
Observations

- **The above analysis with respect to the charged off loans for each variable suggests the following. There is a more probability of defaulting when :**
 - Applicants having house_ownership as 'RENT'
 - Applicants who use the loan to clear other debts
 - Applicants who receive interest at the rate of 13-17%
 - Applicants who have an income of range 31201 - 58402
 - Applicants with employment length of 10
 - When funded amount by investor is between 5000-10000
 - Loan amount is between 5429 - 10357
 - Dti is between 12-18
 - When monthly installments are between 145-274
 - Term of 36 months
 - When the loan status is Not verified
 - When the purpose is 'debt_consolidation'
 - Grade is 'B'
 - And a total grade of 'B5' level.

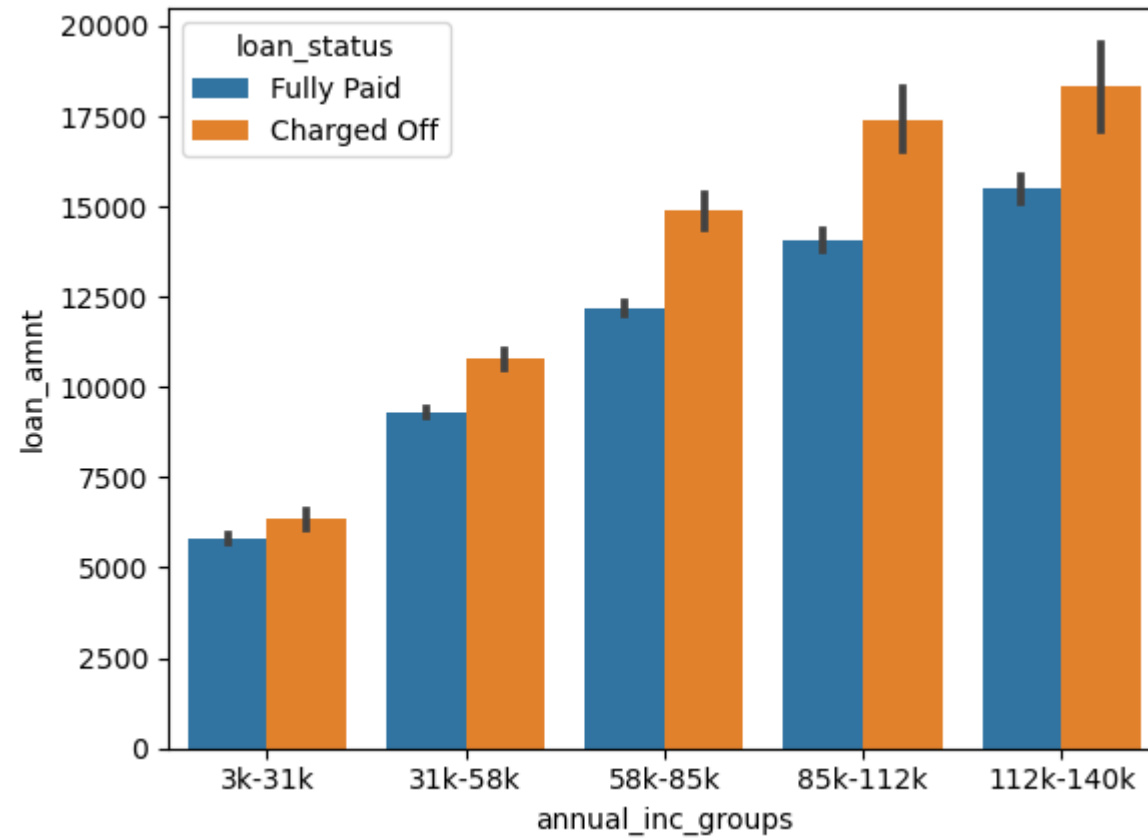
Bivariate Analysis



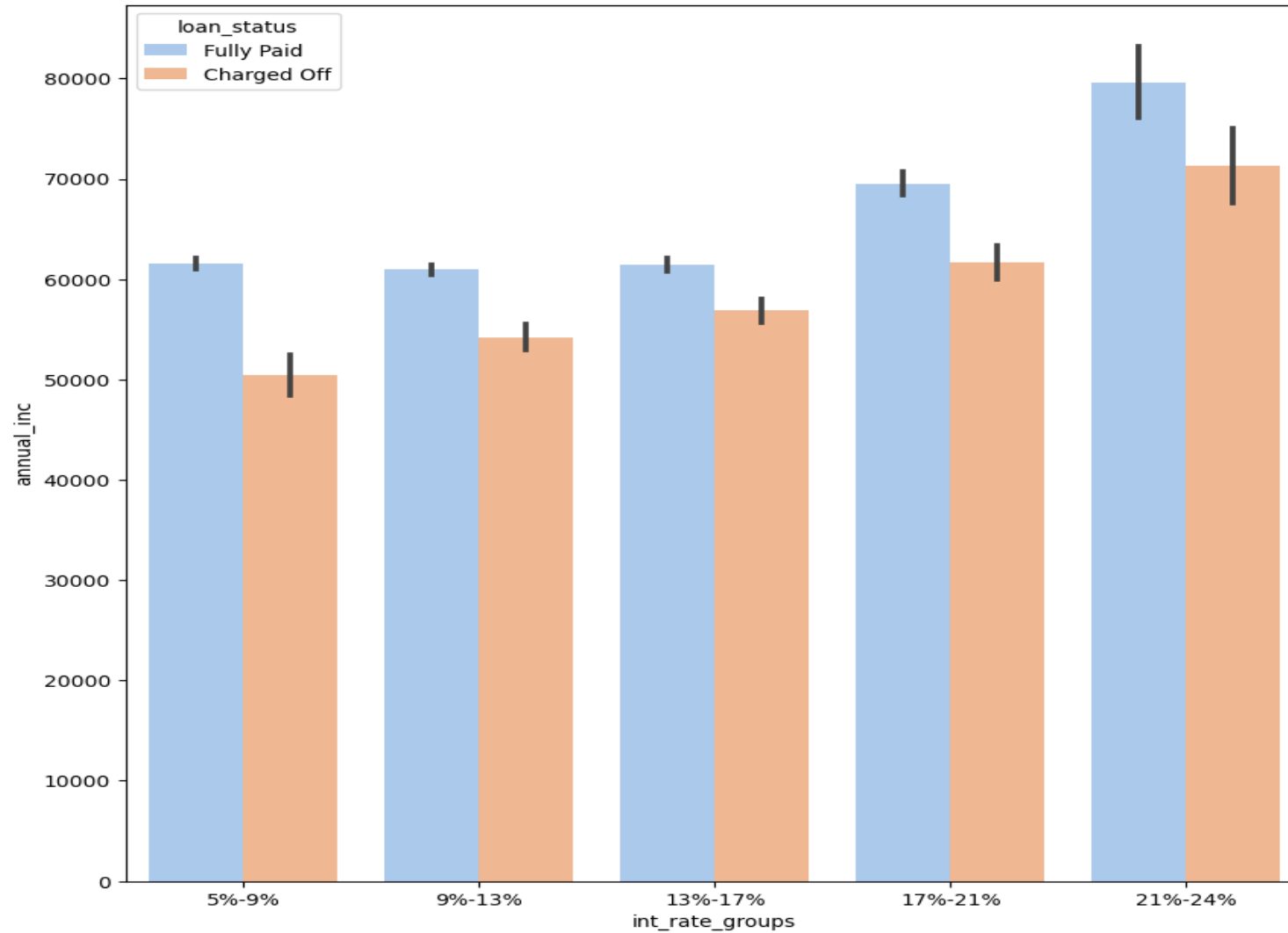
Bivariate Analysis



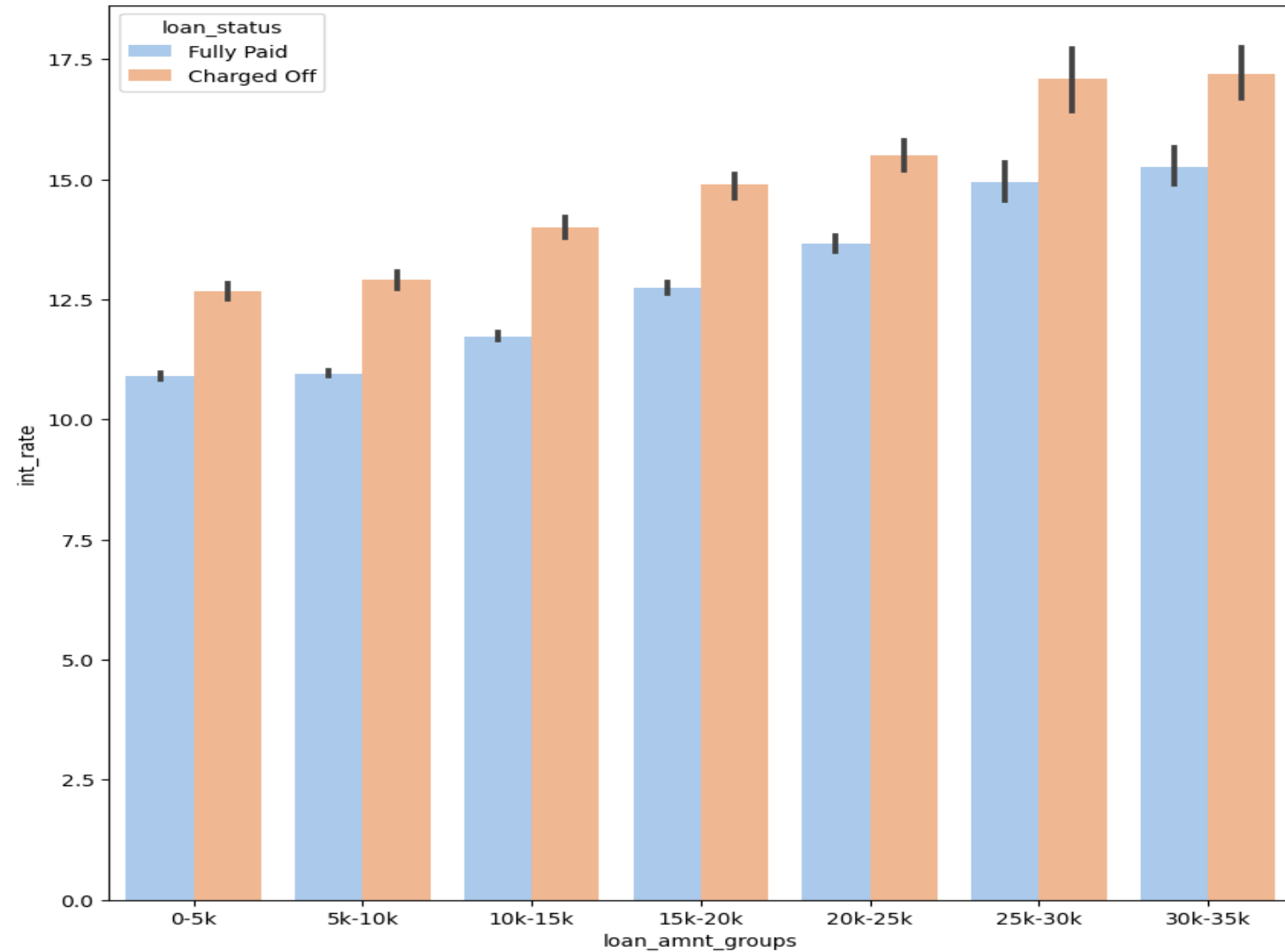
Bivariate Analysis



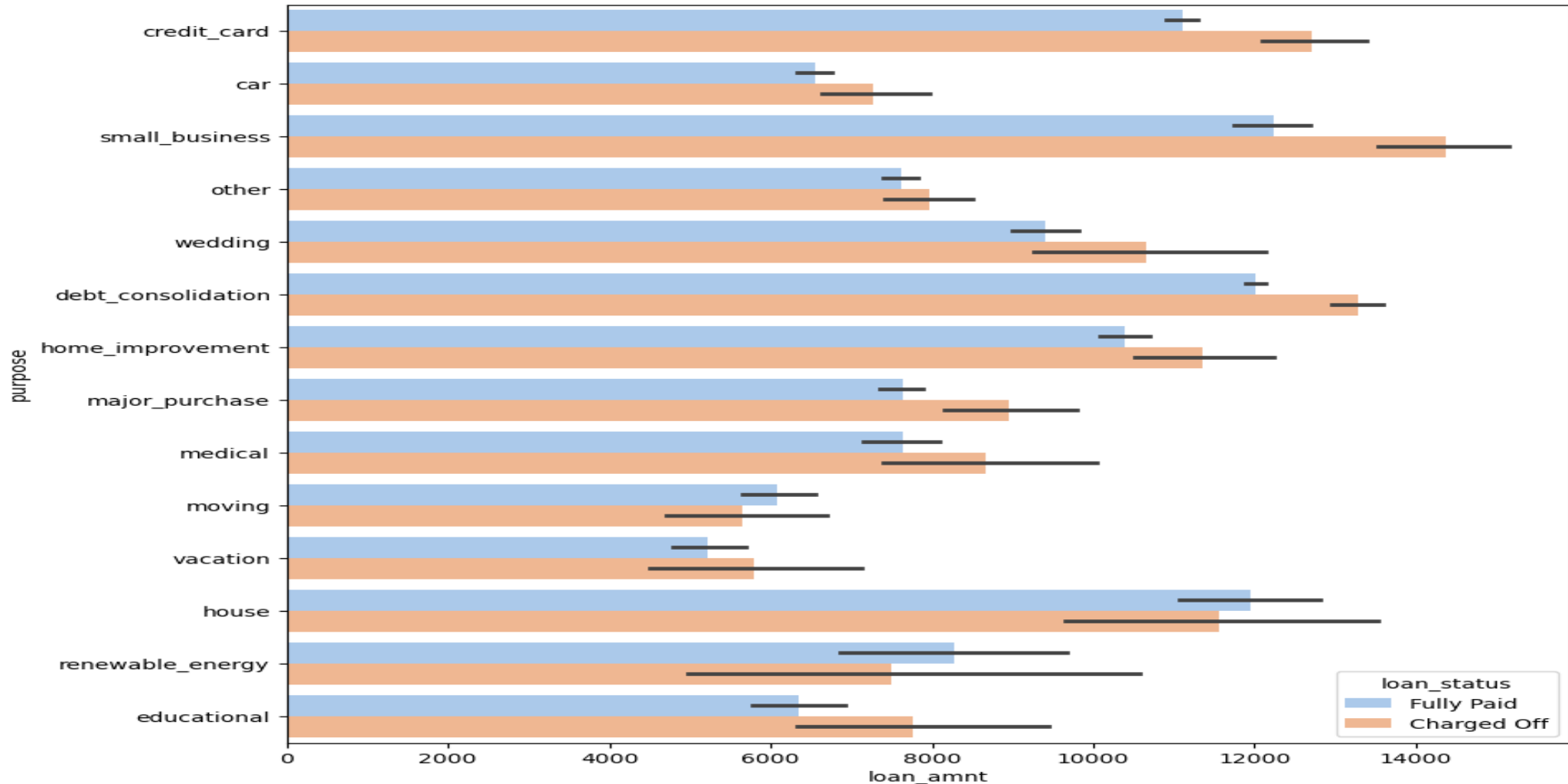
Bivariate Analysis



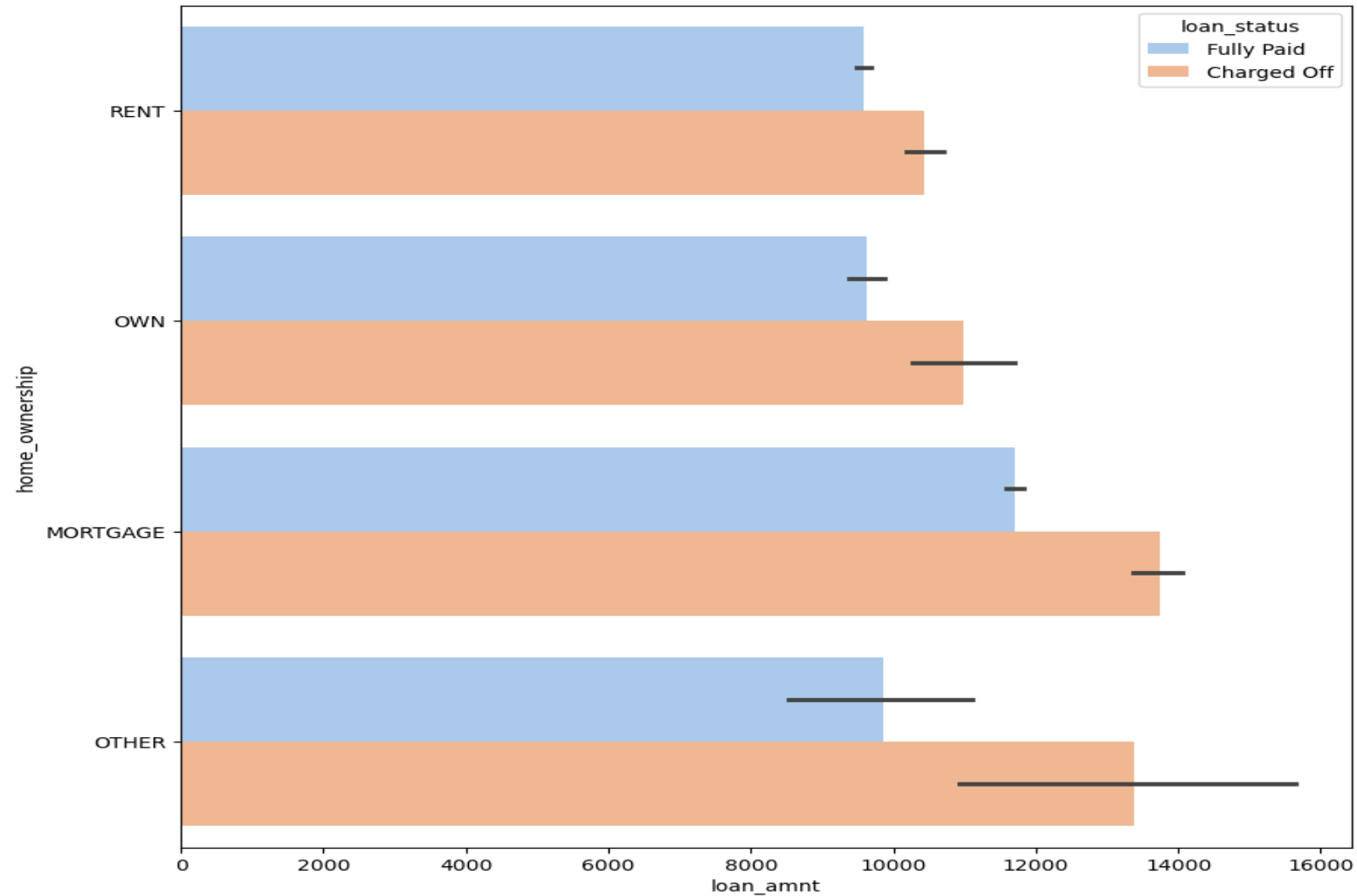
Bivariate Analysis



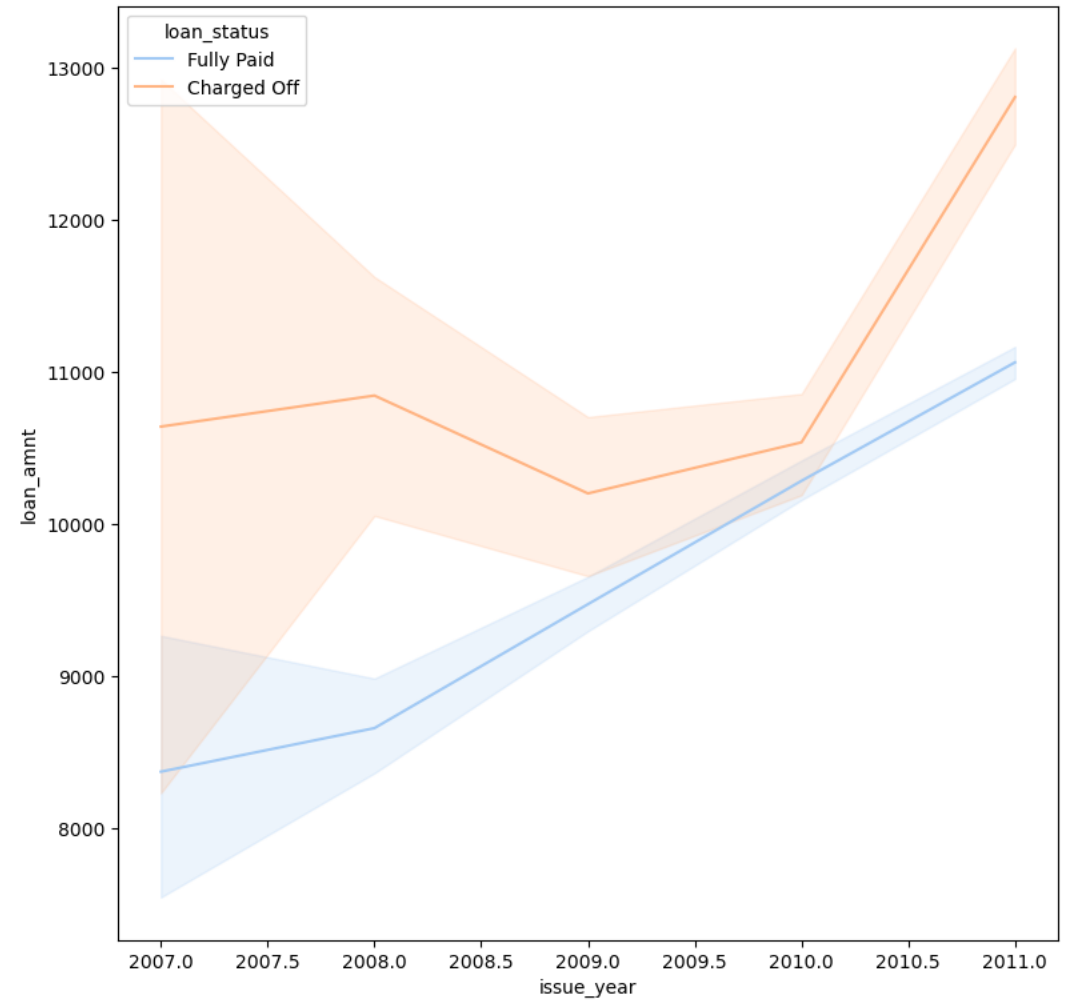
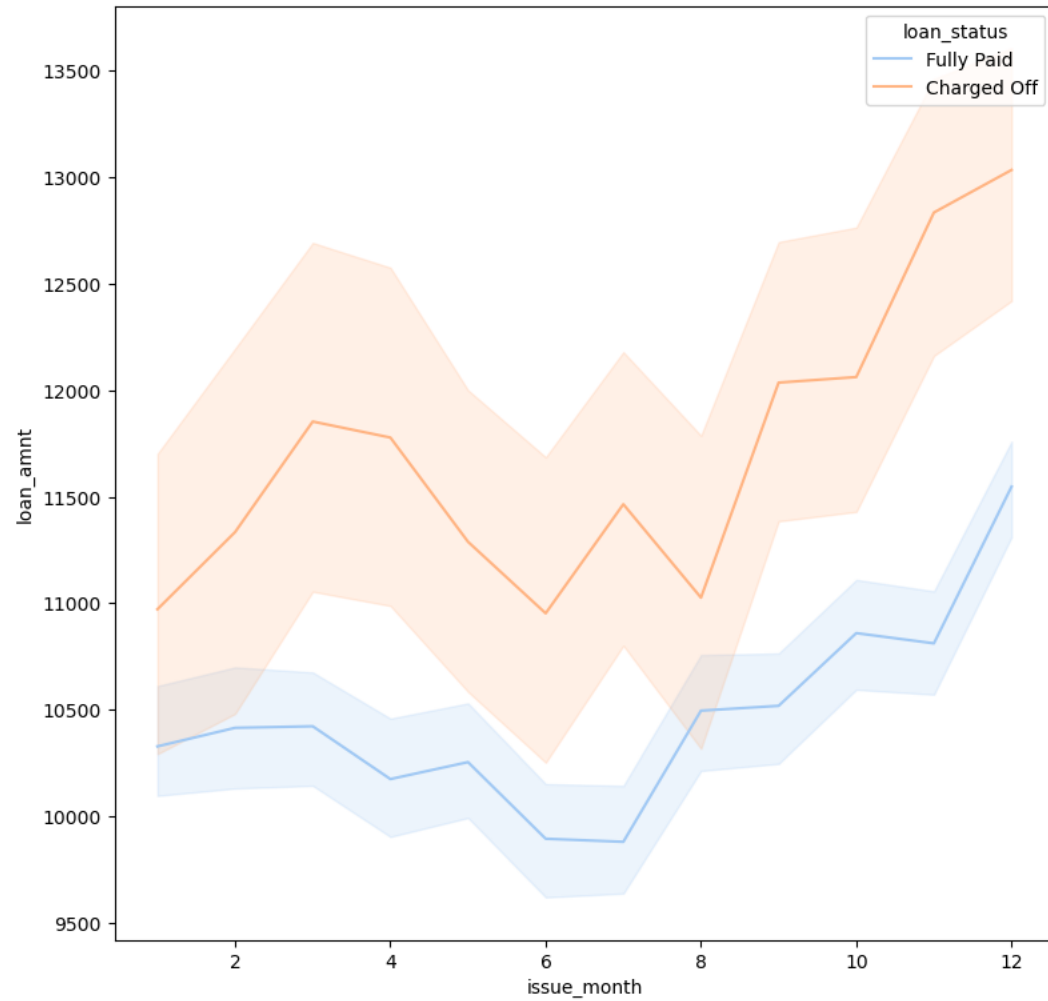
Bivariate Analysis



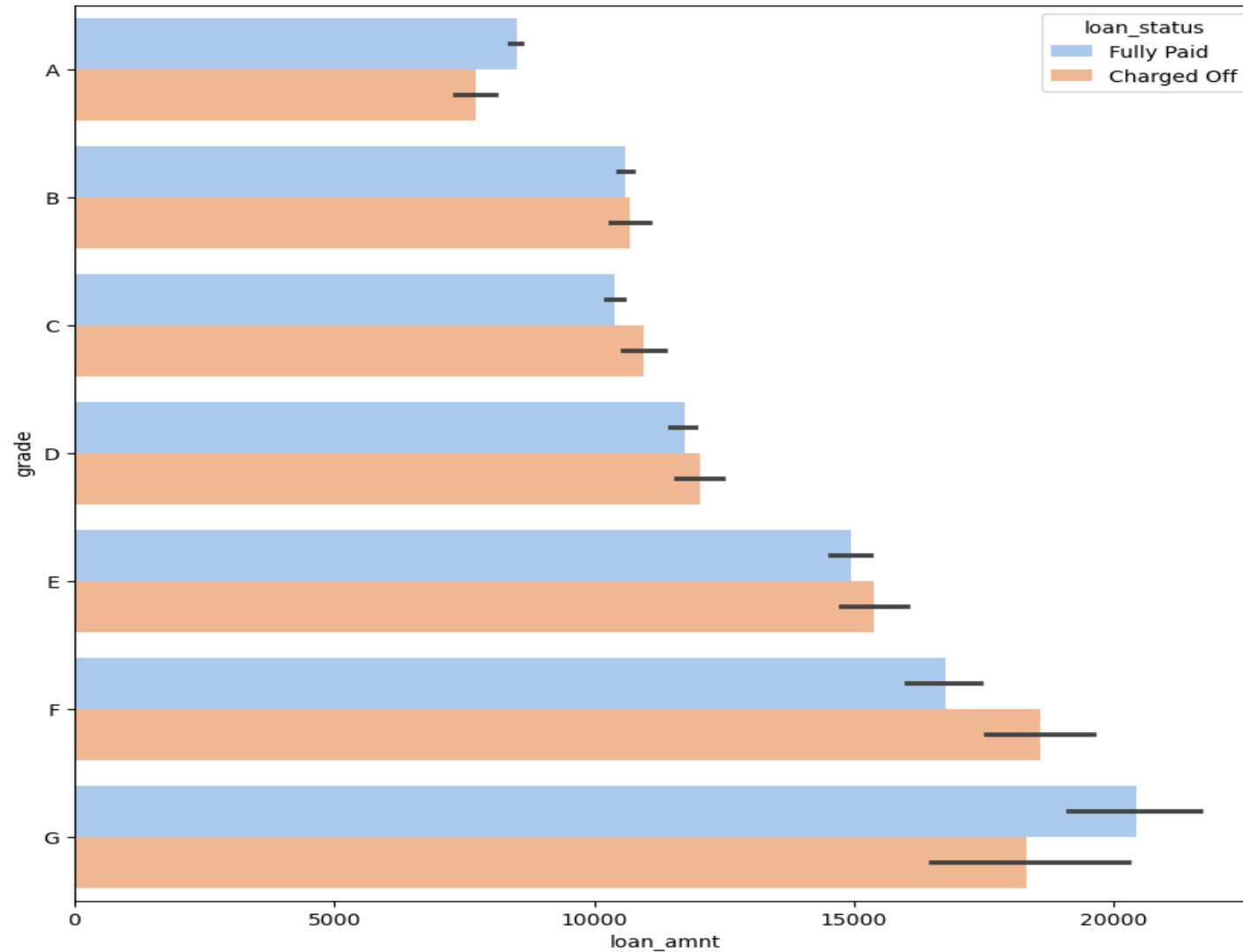
Bivariate Analysis



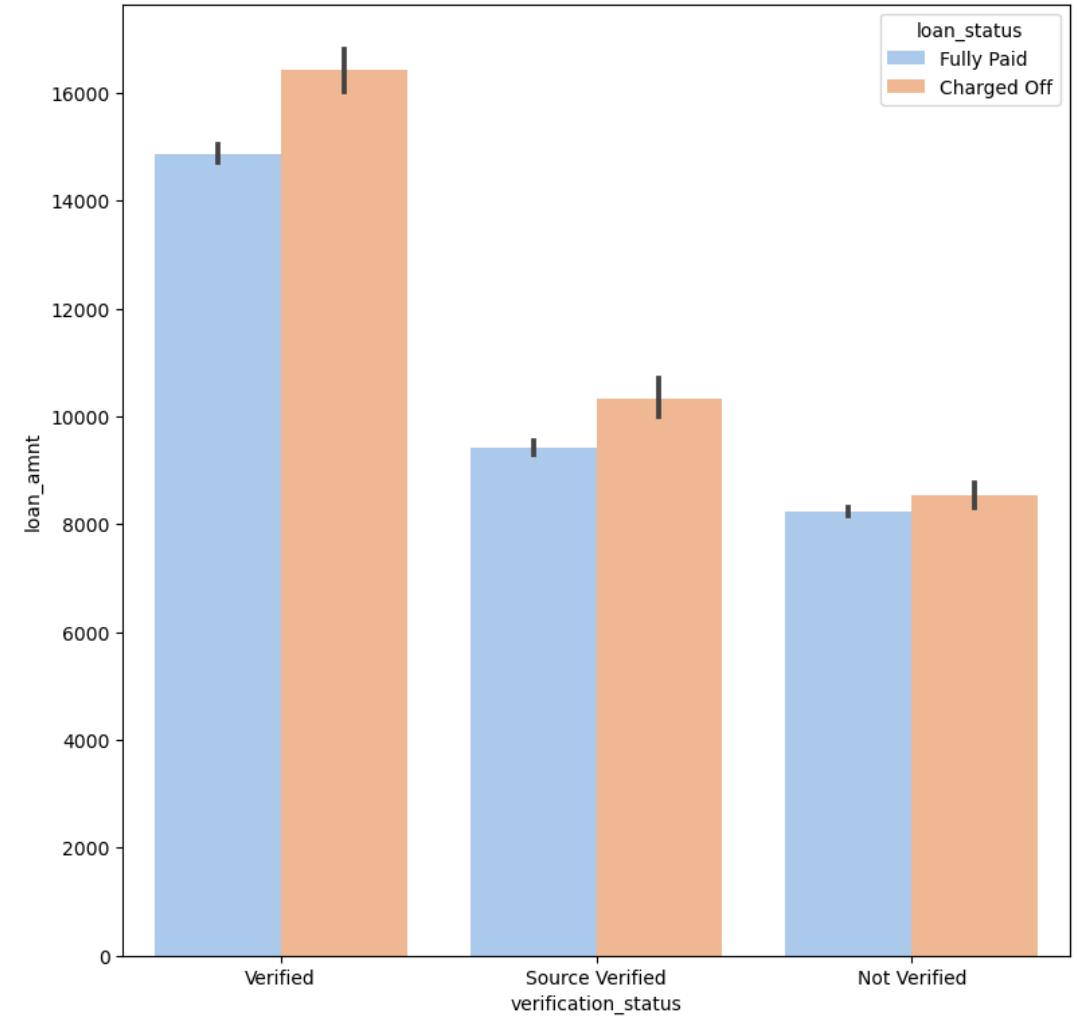
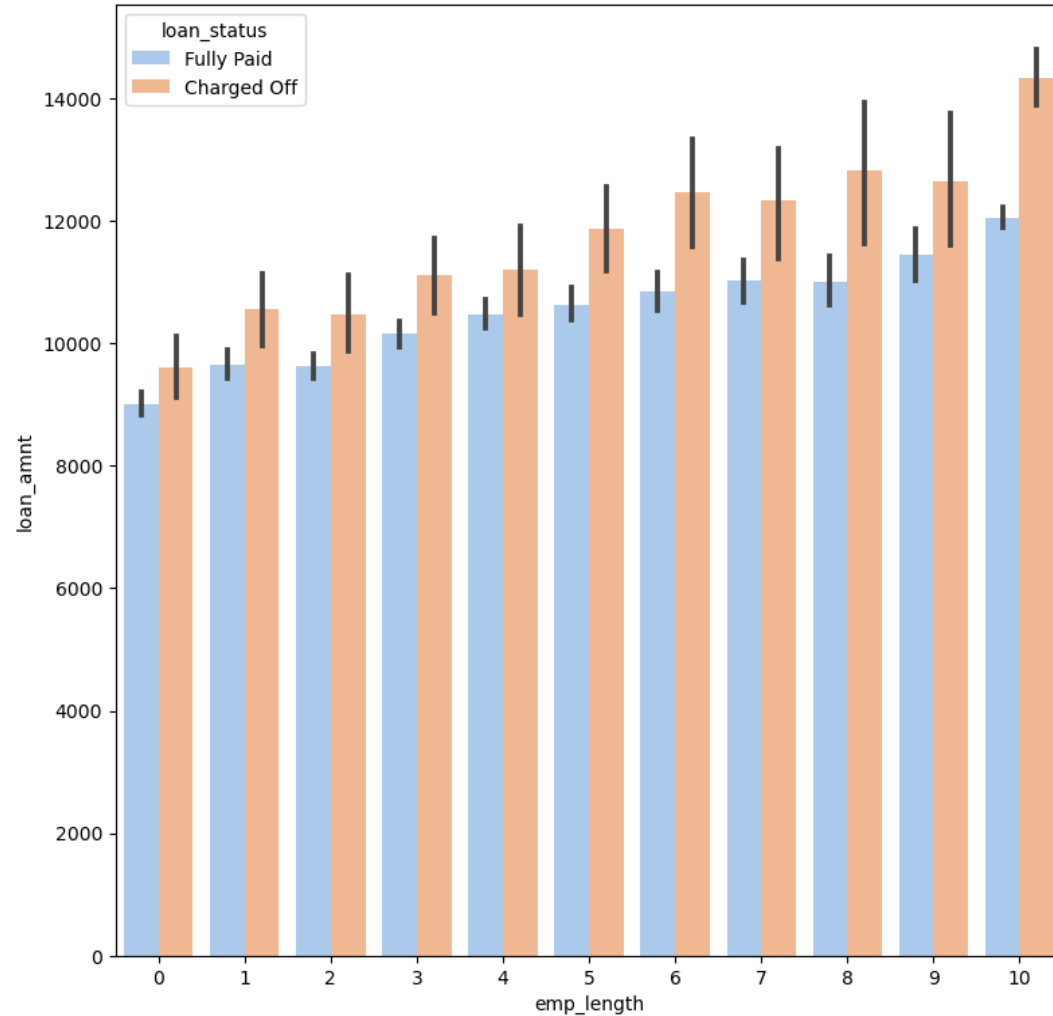
Bivariate Analysis



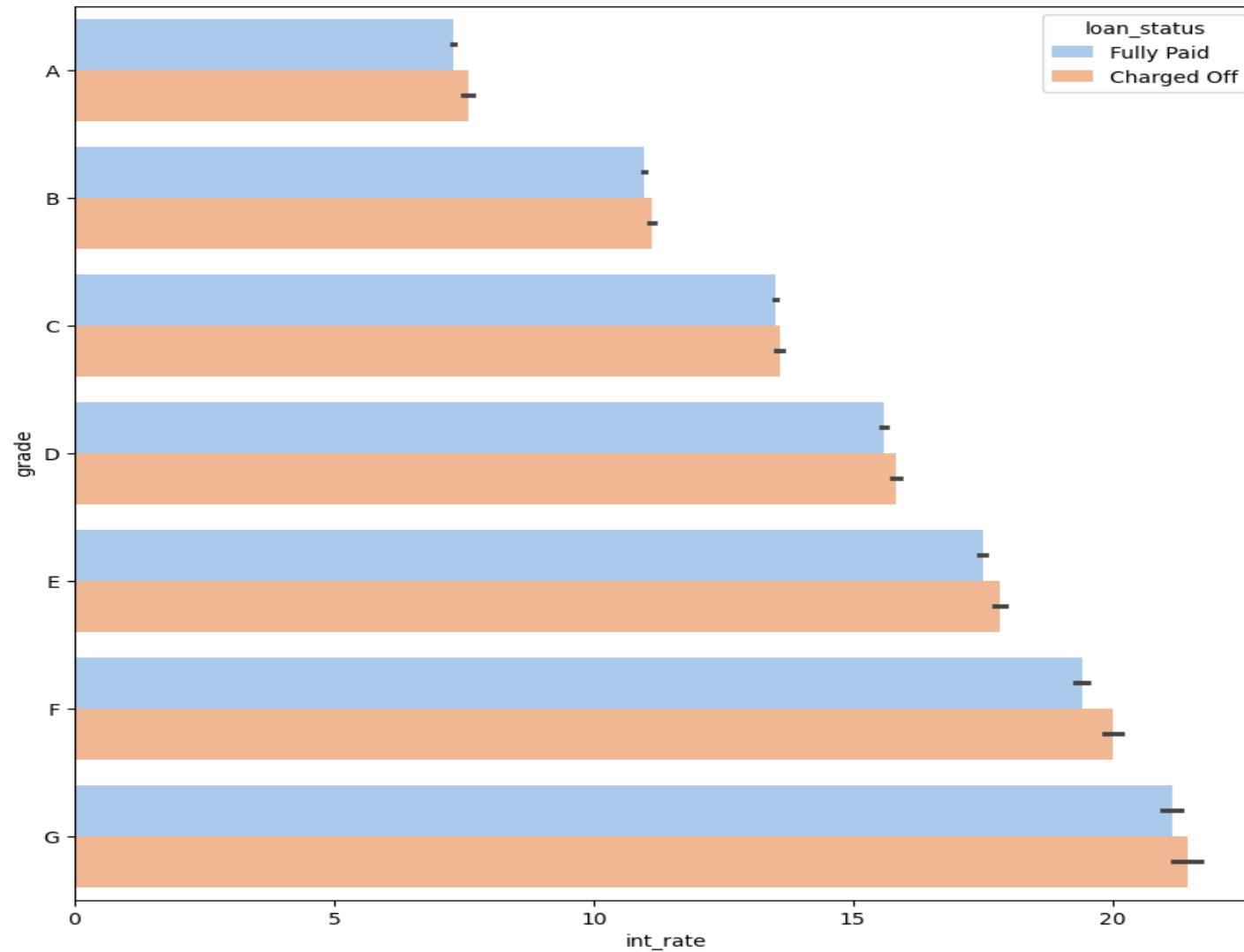
Bivariate Analysis



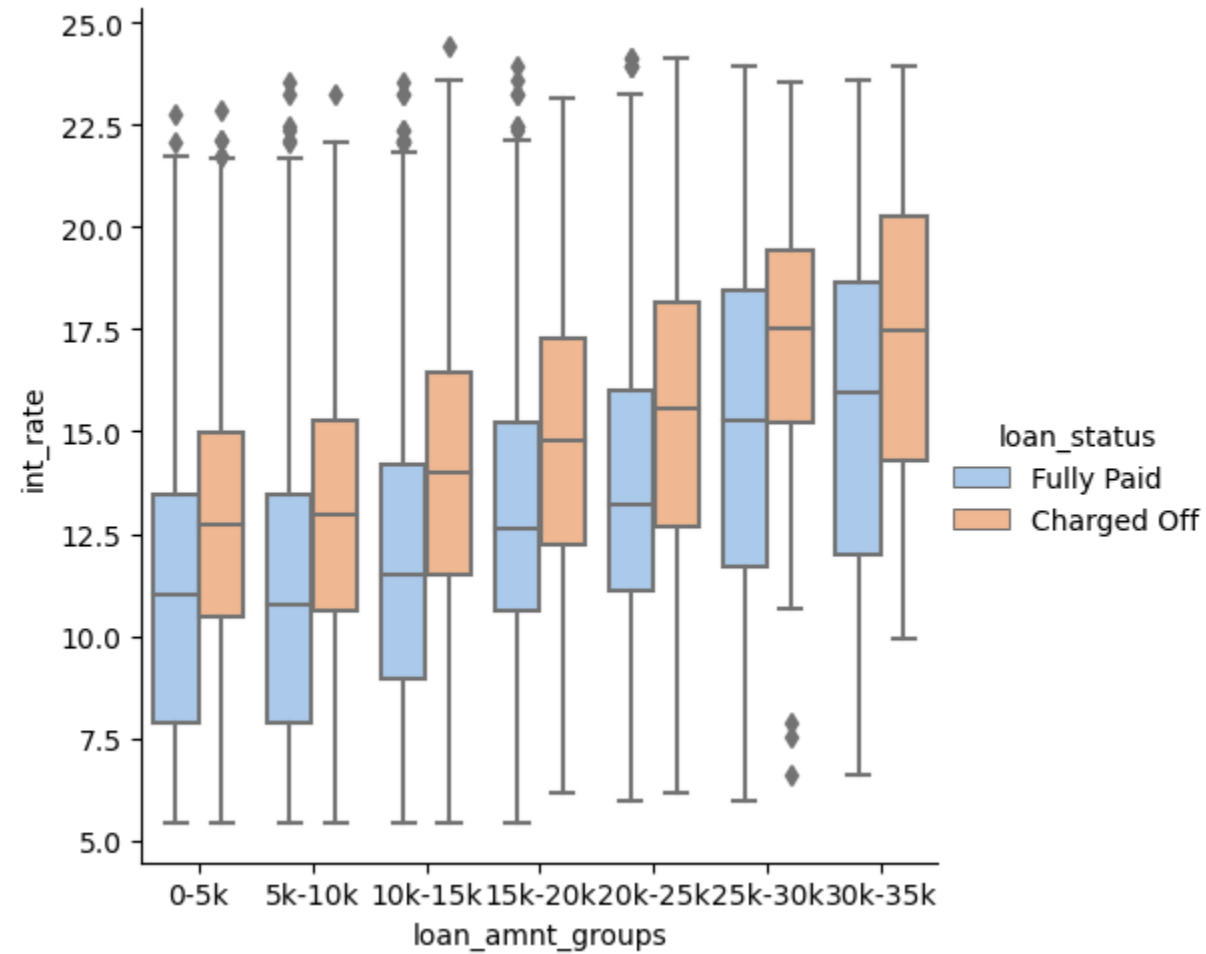
Bivariate Analysis



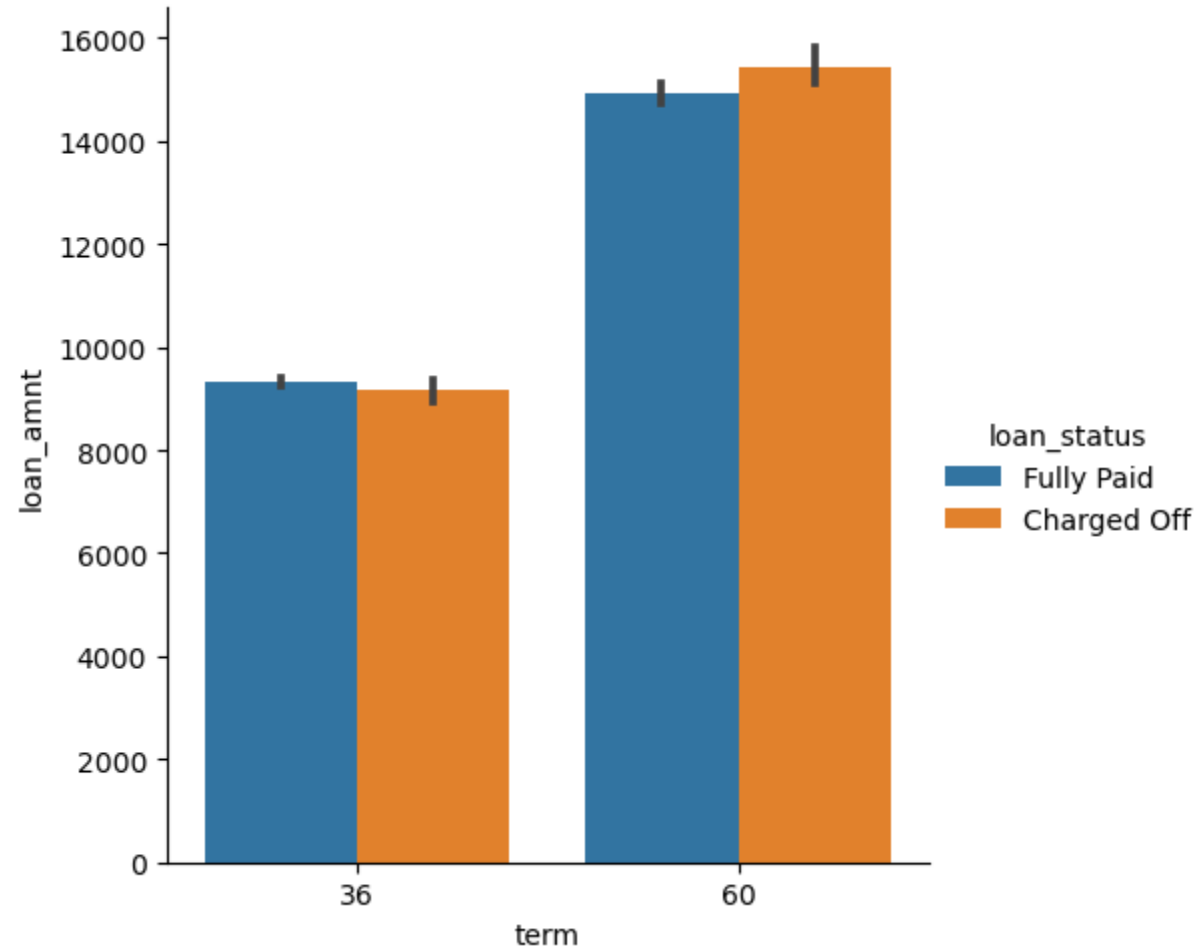
Bivariate Analysis



Bivariate Analysis



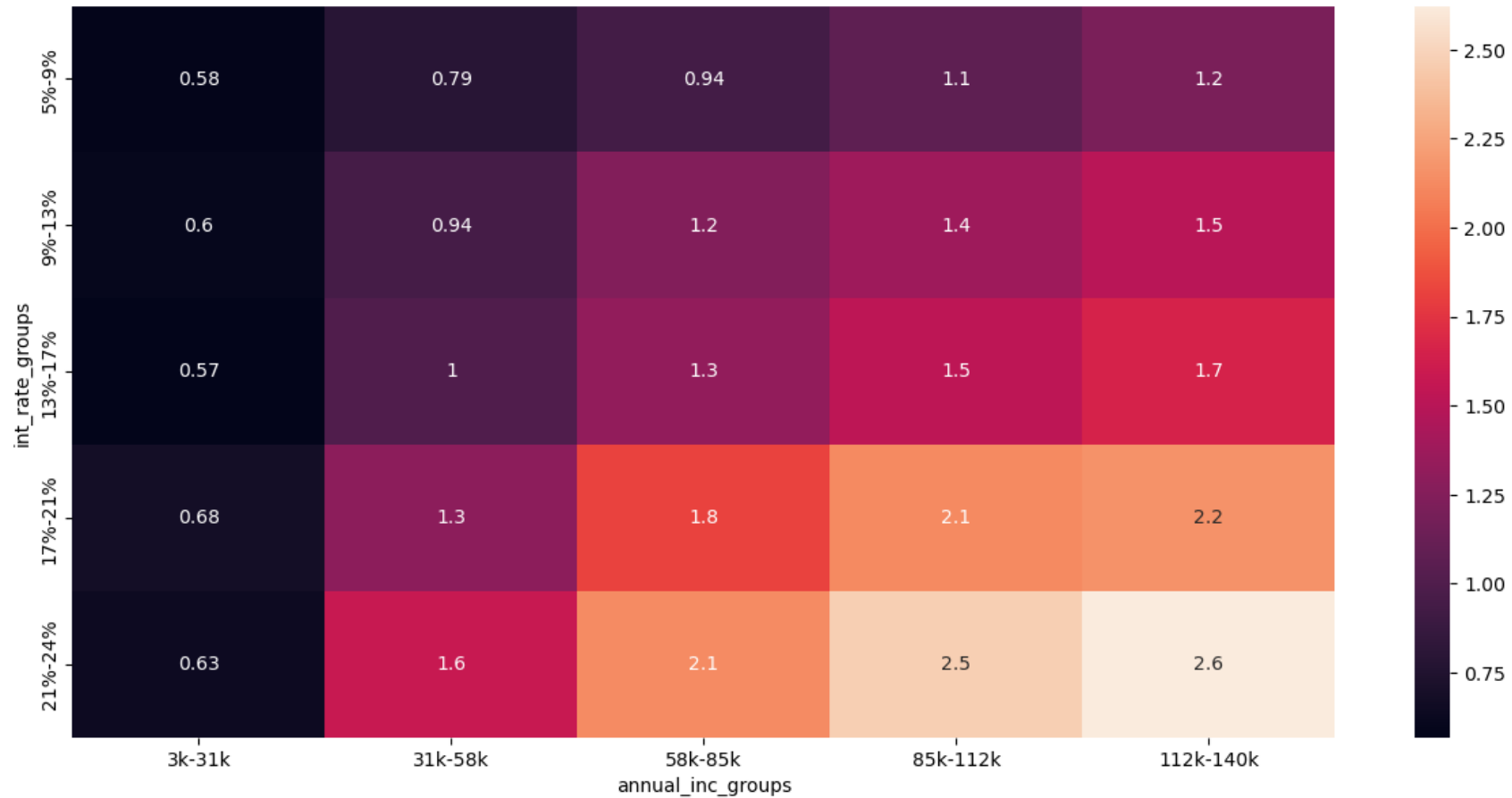
Bivariate Analysis



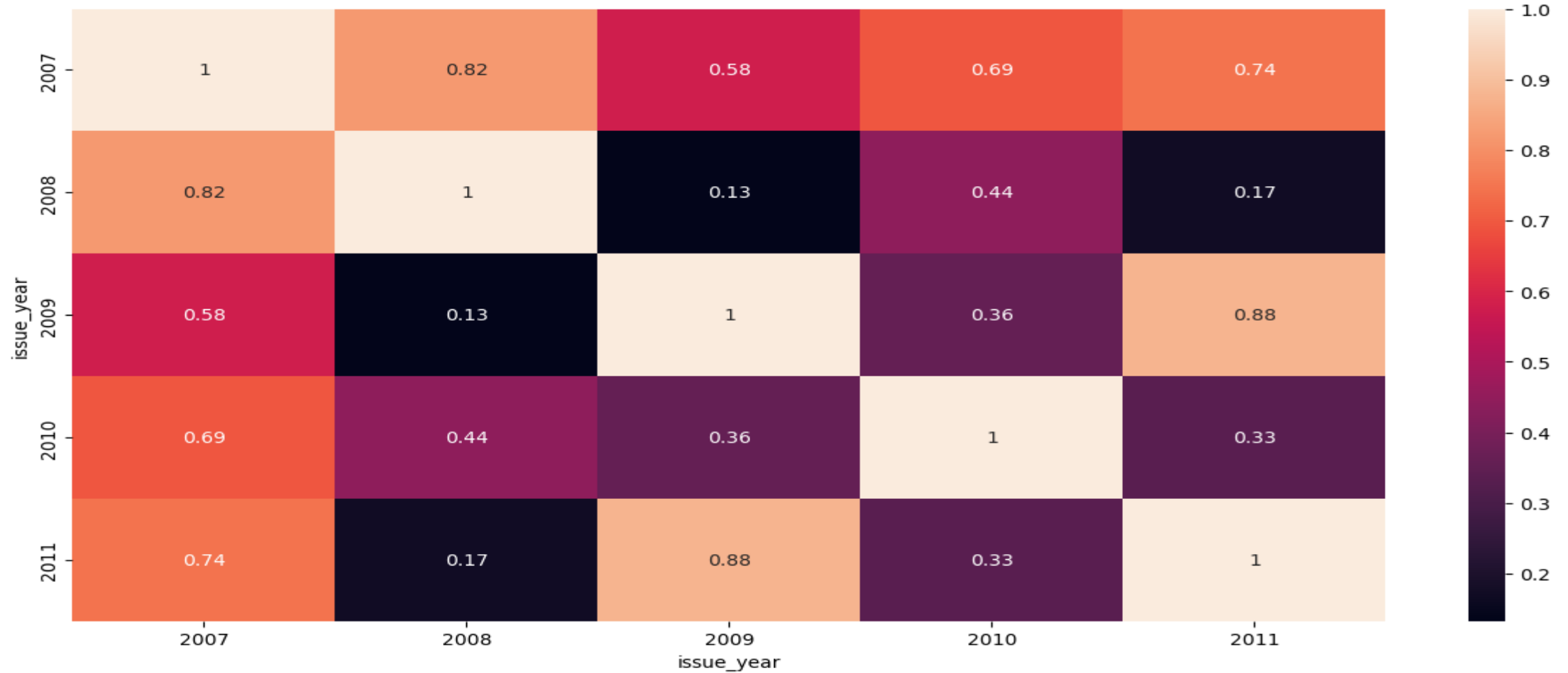
Observations

- **The above analysis with respect to the charged off loans. There is a more probability of defaulting when :**
 - Applicants taking loan for 'home improvement' and have income of 60k -70k
 - Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
 - Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
 - Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %
 - Applicants who have taken a loan for small business and the loan amount is greater than 14k
 - Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k
 - When grade is F and loan amount is between 15k-20k
 - When employment length is 10yrs and loan amount is 12k-14k
 - When the loan is verified and loan amount is above 16k
 - For grade G and interest rate above 20%

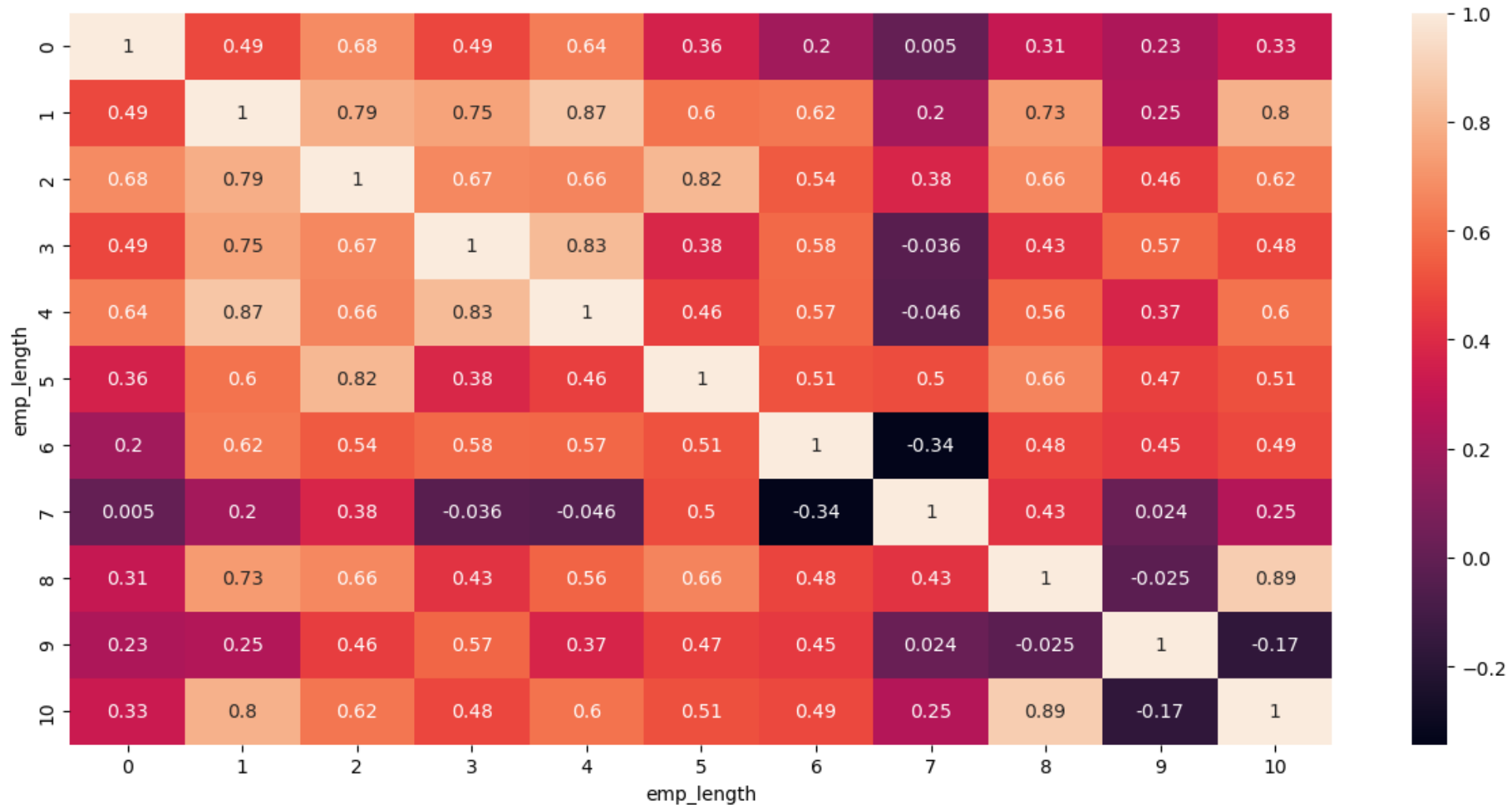
Correlation b/w features



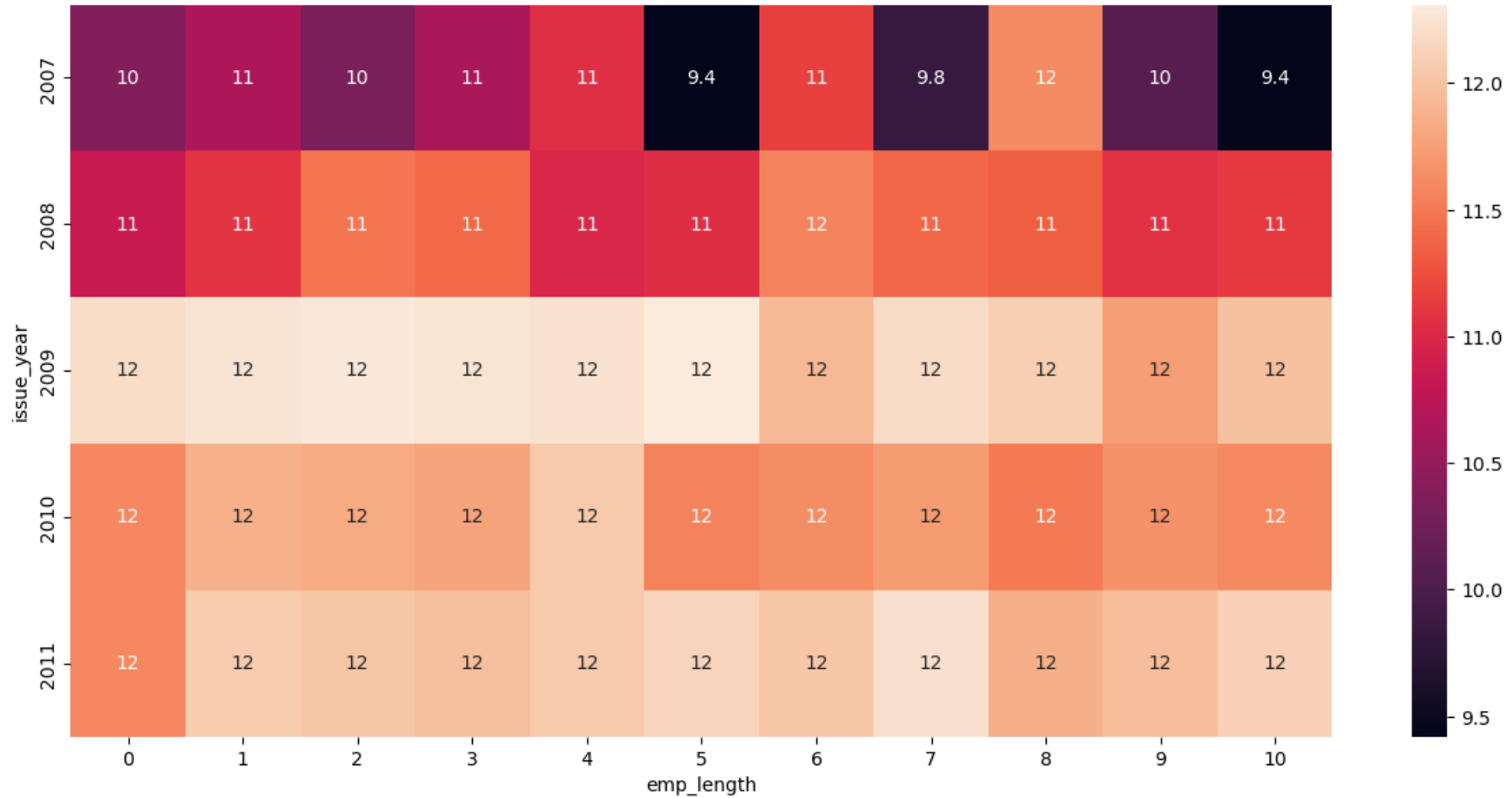
Correlation b/w features



Correlation b/w features



Correlation b/w features



Observations

- **The above analysis is correlation between some variables. There is a more probability of defaulting when :**
 - Applicants are in more dark annual income colors for heat map plt first
 - Applicants were applied the most loan in 2009 and 2011 because 2009 and 2011 are having highest correlation value for heat map second.
 - Applicant were applied the most loan for employed 8 and 10 years are having highest correlation value for heat map third.
 - Applicant were applied the most interest rate in 2009 to 2011 are having highest correlation value for heat map fourth