


Lending Club Case Study

Pavithra Pishu, Vani M
(AIML C53 Batch)



Problem Statement and Approach

- Identify the driving factors/variable behind an applicant defaulting a loan by analysing the given data.
- The analysis takes an Exploratory approach which include Univariate, Bivariate and Multi Variate analysis.
- Python program is used to clean and analyse the data with necessary visualisation

Data Interpretation

- The given dataset has 39717 Rows and 111 Columns out of which :
 - ~54 columns have no or null values.
 - ~10 had single valued columns which did not contribute much to the analysis.
 - ~8 columns were some informative or description columns which again do not contribute to the analysis.
 - ~4 columns had > 90% of missing data Making them good candidates to eliminate these columns
- Segment the data into the necessary buckets to analyse
 - Categorical
 - Continuous
 - Extra

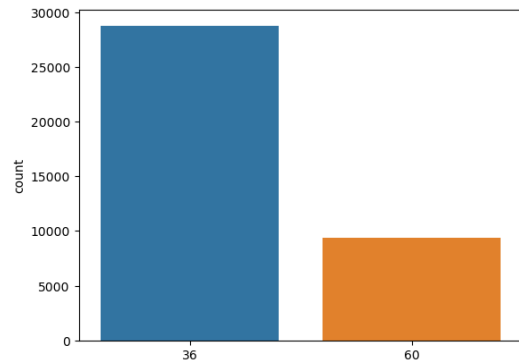
Data Wrangling

- Identify the columns that are missing values i.e., Null Values
- Imputing the Null values with mode values for categorical data
- Identify the different data types across the current columns for analysis
- Delete the columns where the data is not present or cannot be converted into a usable form
- Identify and rectify the outliers
- Convert the data to proper form and format
 - 'emp_length', 'term', 'int_rate' and 'revol_util' is observed to be an object convert to numeric form for better usage

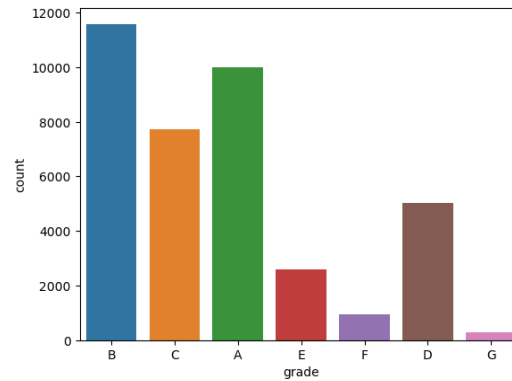
Data Analysis

- The Exploratory Data Analysis was carried out using the below methods.
- **Univariate Analysis:** The term univariate analysis refers to the analysis of one variable.
- **Bivariate Analysis:** An analysis of two variables to determine the relationships between them.
- **Multivariate Analysis:** An analysis where multiple measurements are made on each experimental unit and where the relationships among multivariate measurements and their structure are important

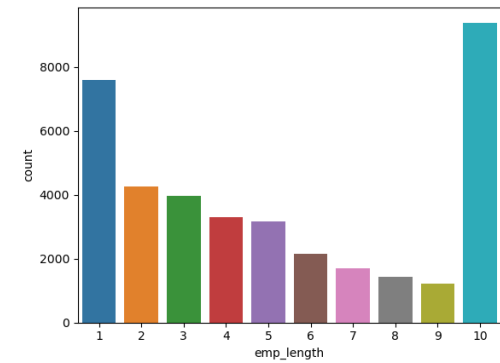
Univariate Analysis



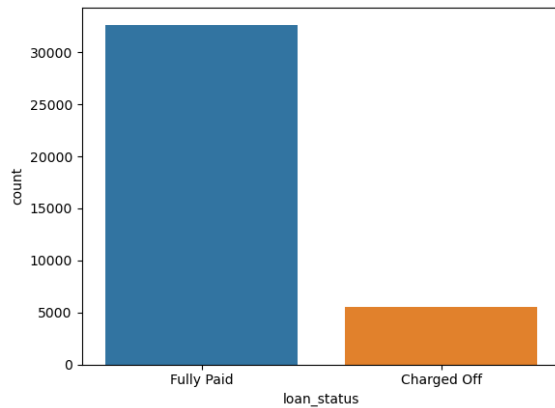
The total loans were more for term of 36 Months



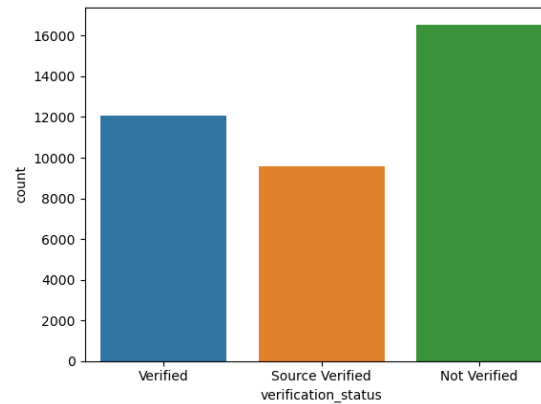
Grade B loans are more in number



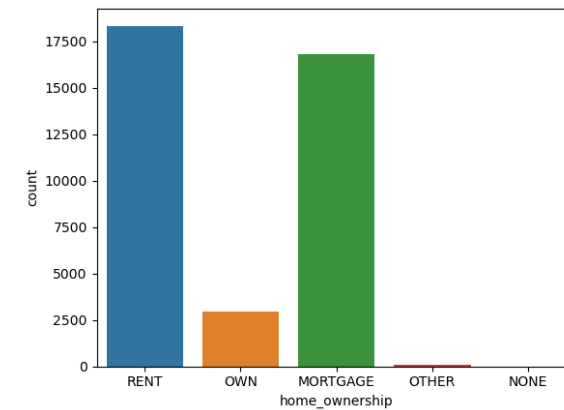
Employee Tenure of 1 yr and 10 yrs are major defaulters



Fully Paid vs Charged Off



Loan Verification Status



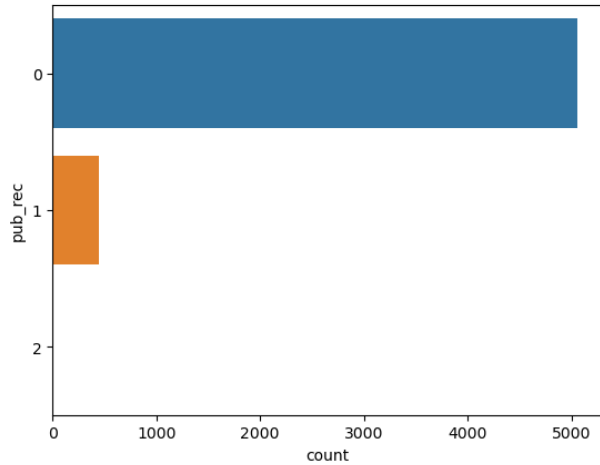
Home Ownership status

Observations: Univariate Analysis

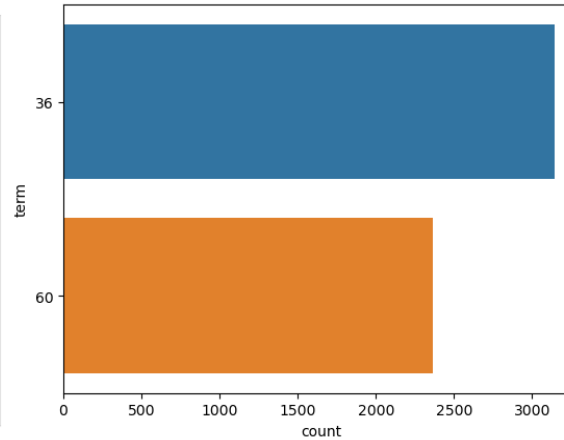
Below are the candidates who fall under the defaulter category:

- Annual_income has some extreme outliers. Rest of them can be managed.
- - 66% of the loans have a term of 36 months
- - 30% loans are B grade
- - 20% candidates have work experience with 10+ years
- - 44% have either a MORTGAGE or RENT
- - 45% have a verified status of Not verified.

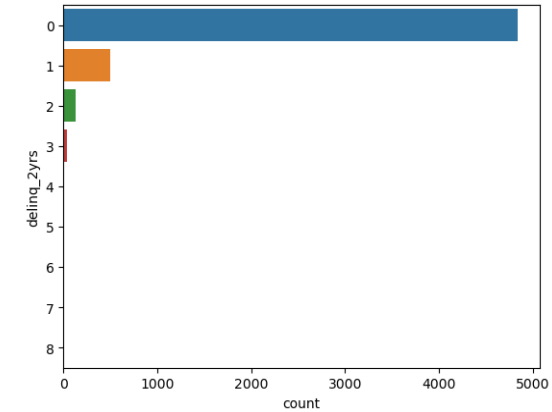
Bivariate Analysis



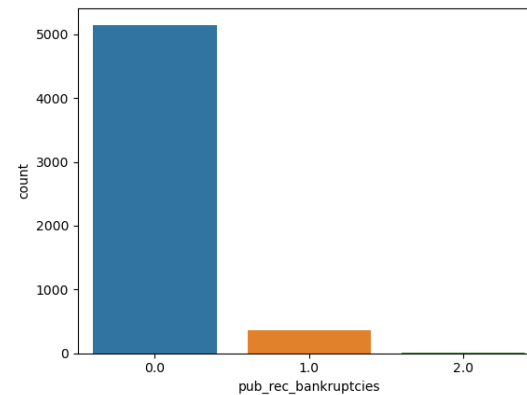
More the Number of derogatory public records
more are the Charged-Offs



The lesser the tenure the more the Charged off
loans



The more the incidences of delinquency
the higher is the Charged-off Loans



More the Number of public record bankruptcies
more are the Charged-Offs

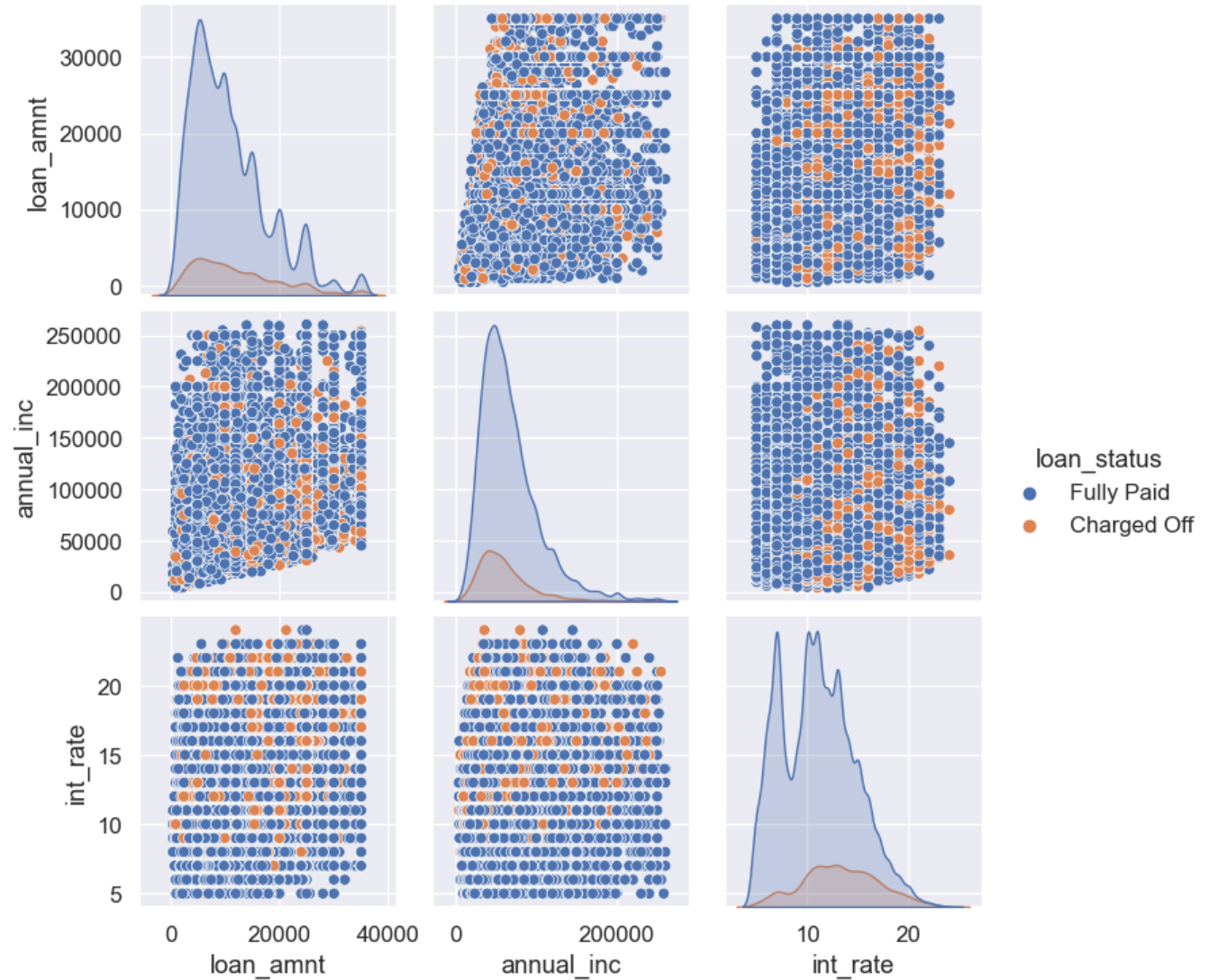
Observations: Bivariate Analysis

Loan defaulters are higher :

- whose 'loan amounts' are between 5000 to 10000.
- who have a 'term' of 36 months
- for the 'int_rate' between 10-14%
- monthly 'installment' are between 145-273
- Grade 'B' and 'B5' loans
- whose employment length is > 10 years
- who 'RENT' a home
- whose annual income is between 30-54k
- when the loan verification status is 'not verified'
- purpose of loan is 'debt consolidation'
- debt to income ratio is between 12-17
- no inquiries in last 6 months (0)
- whose number of open credit lines in their credit file are between 2-10
- whose derogatory public records are 0
- total credit revolving balance is between 0-30
- Revolving line utilization rate is 60-80
- total credit lines 2-20
- principal received to date is 0-5k
- 0 public record bankruptcies
- whose issue year is 2011 and month is December

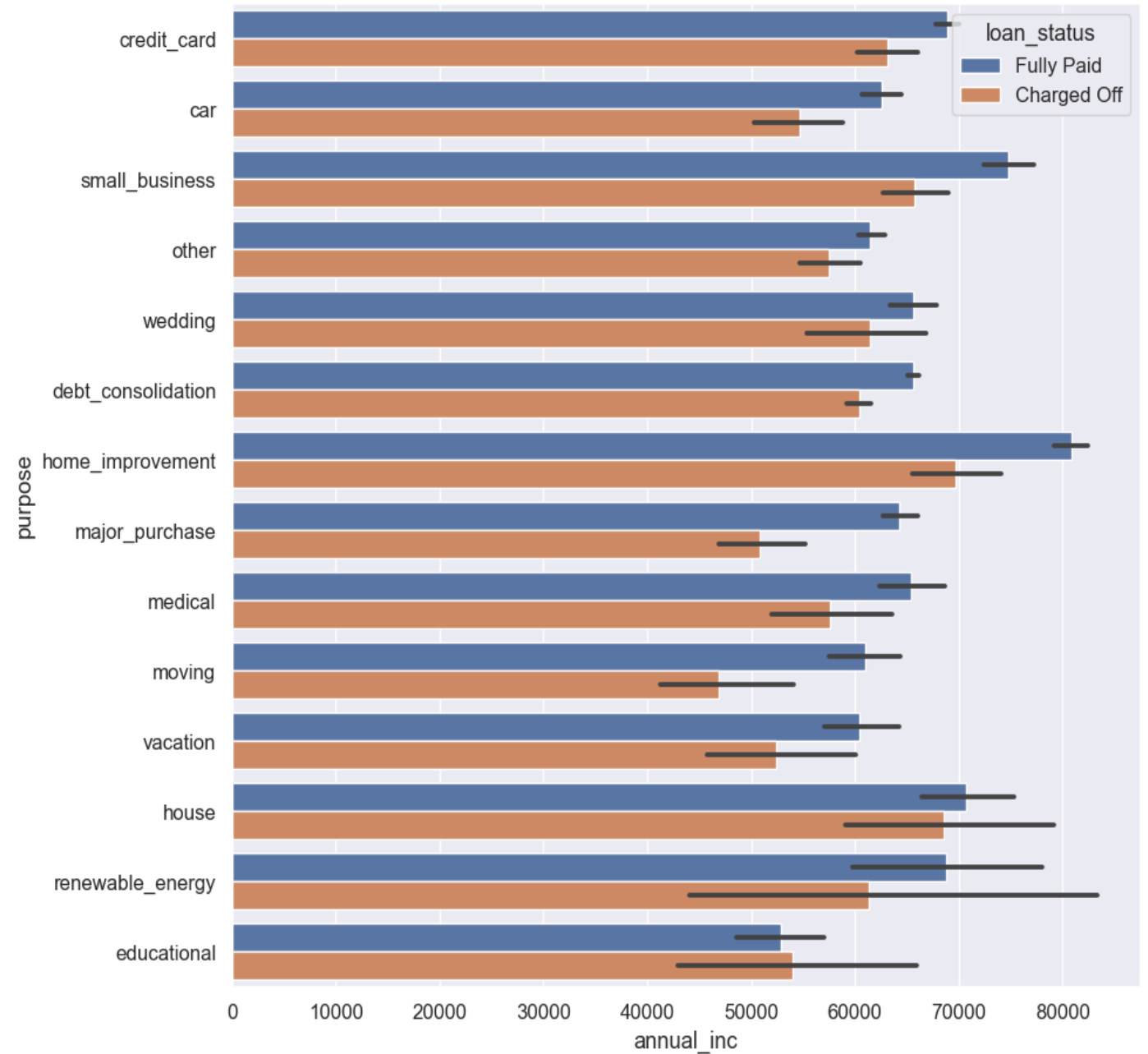
Multivariate Analysis

- pair plot of annual income , loan amount, interest rate



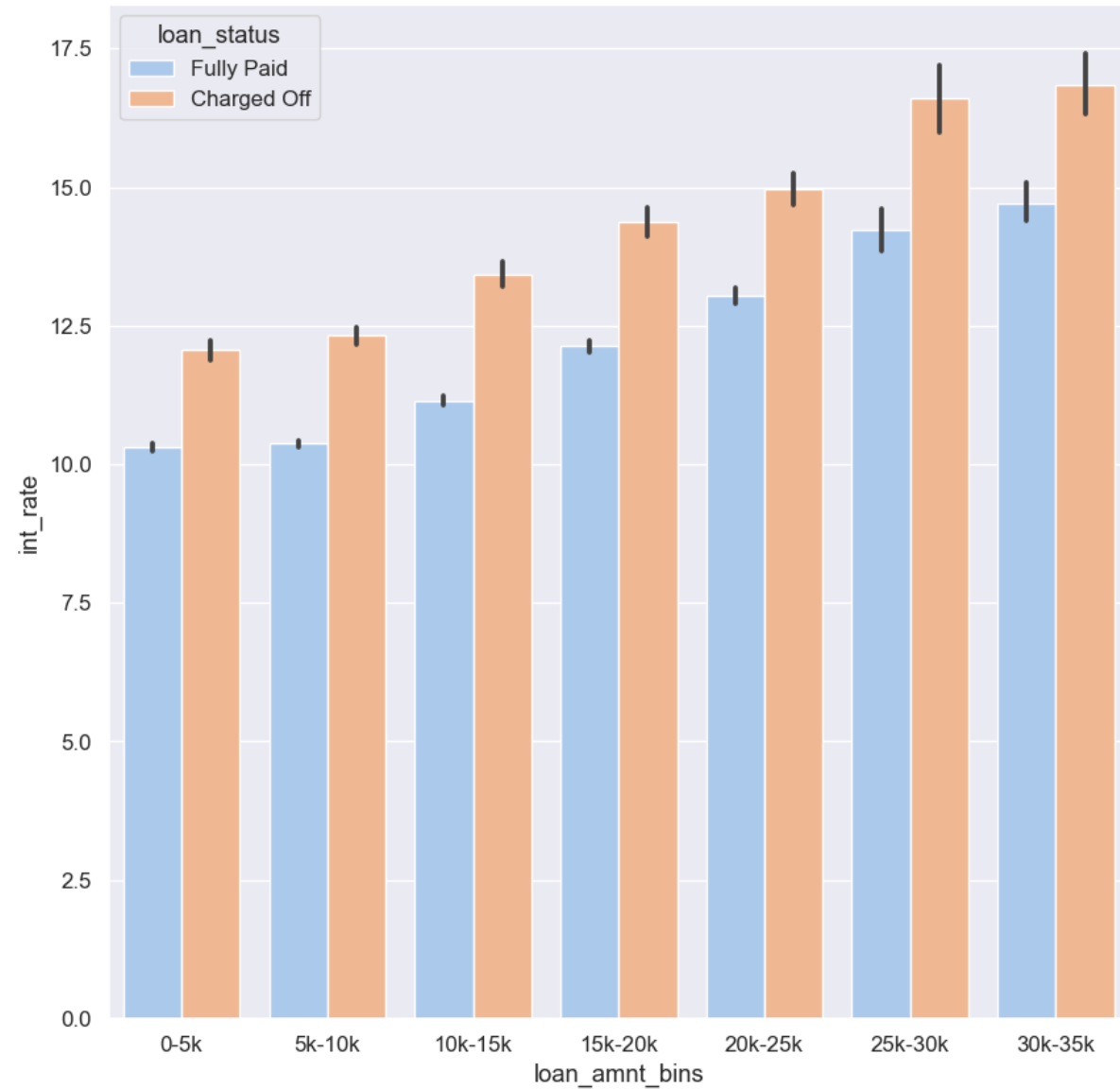
Multivariate Analysis

- Purpose vs Annual inc vs loan status



Multivariate Analysis

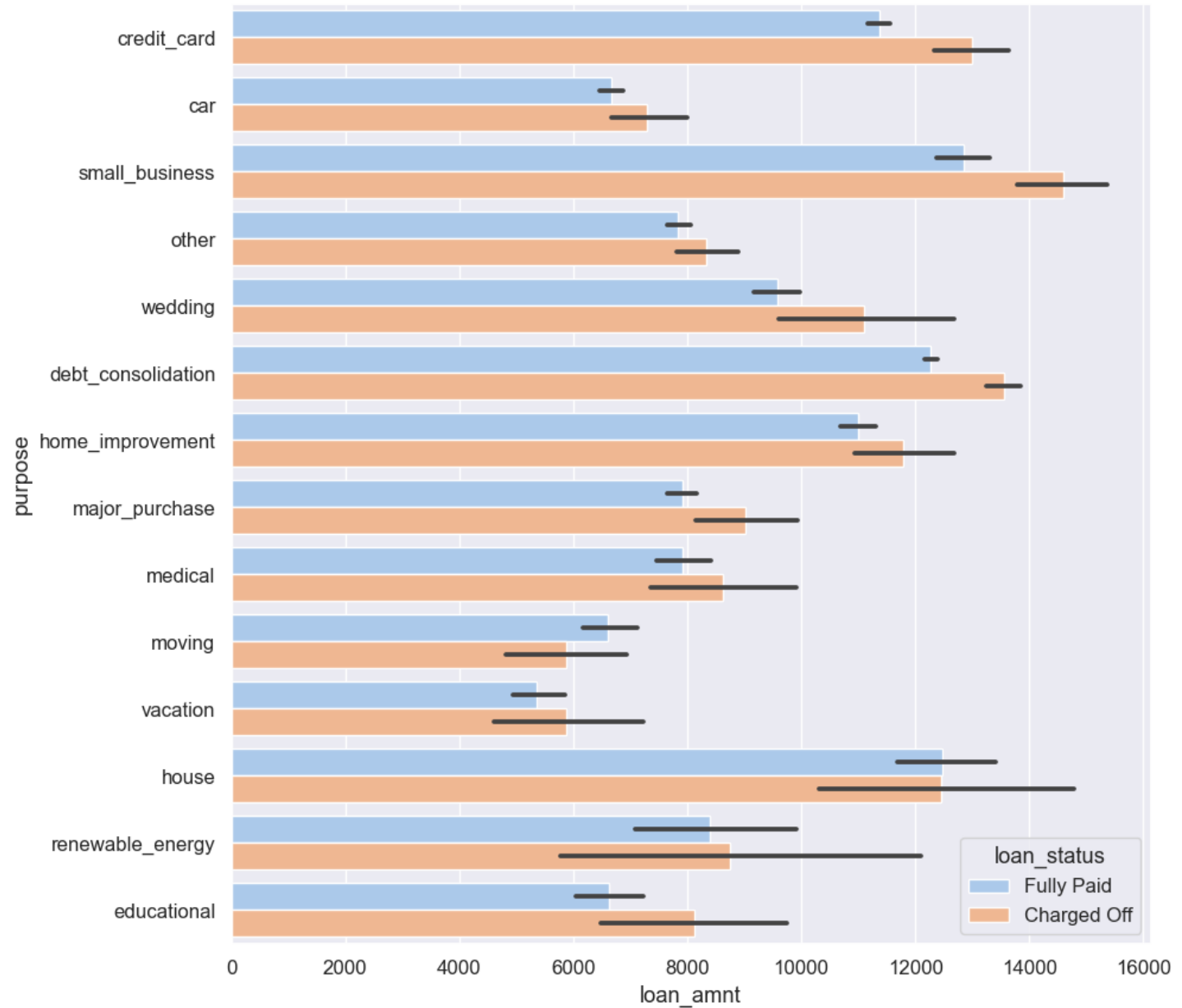
- interest rate vs loan amount vs
loan status





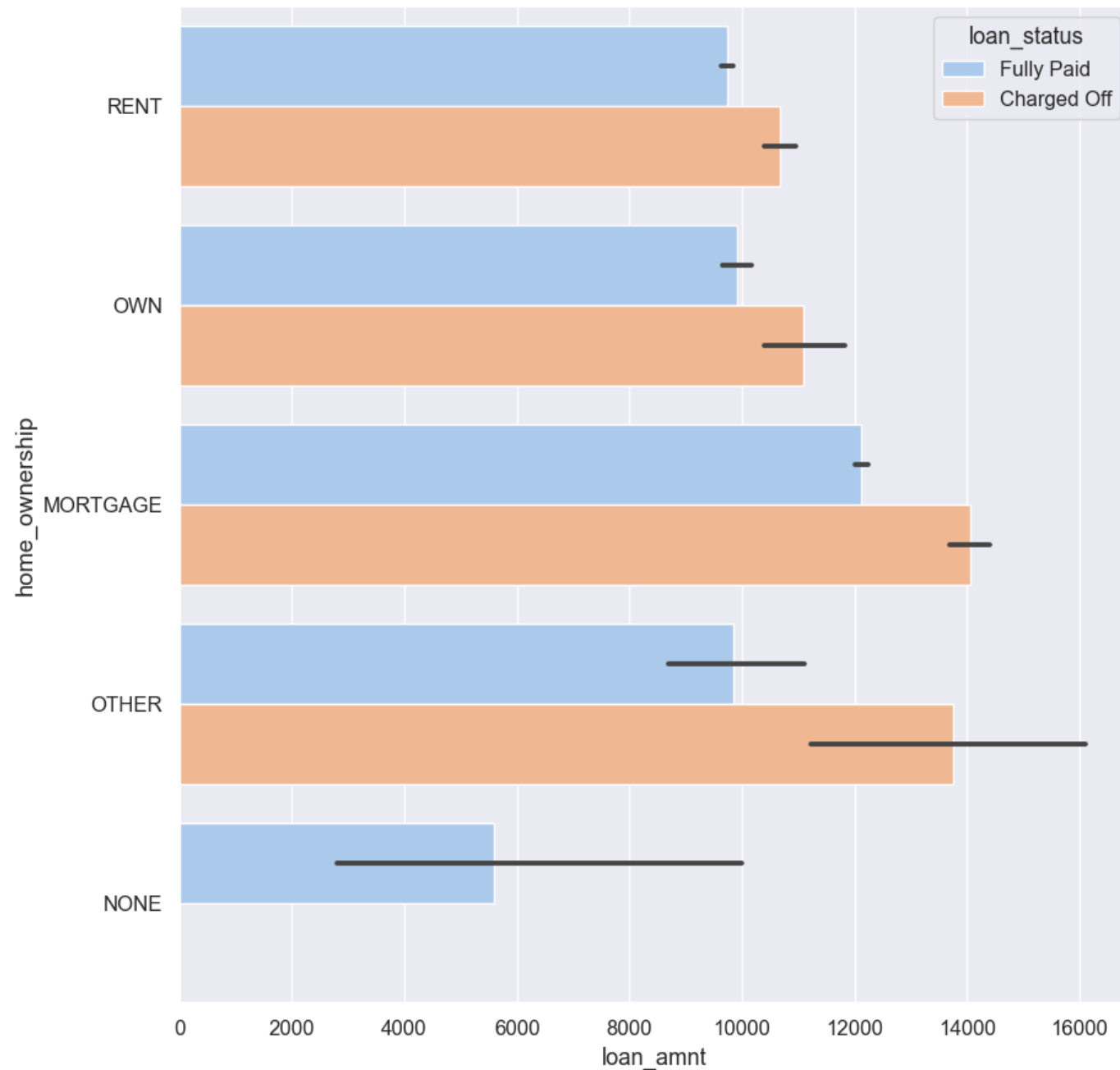
Multivariate Analysis

- purpose vs loan amount vs loan status



Multivariate Analysis

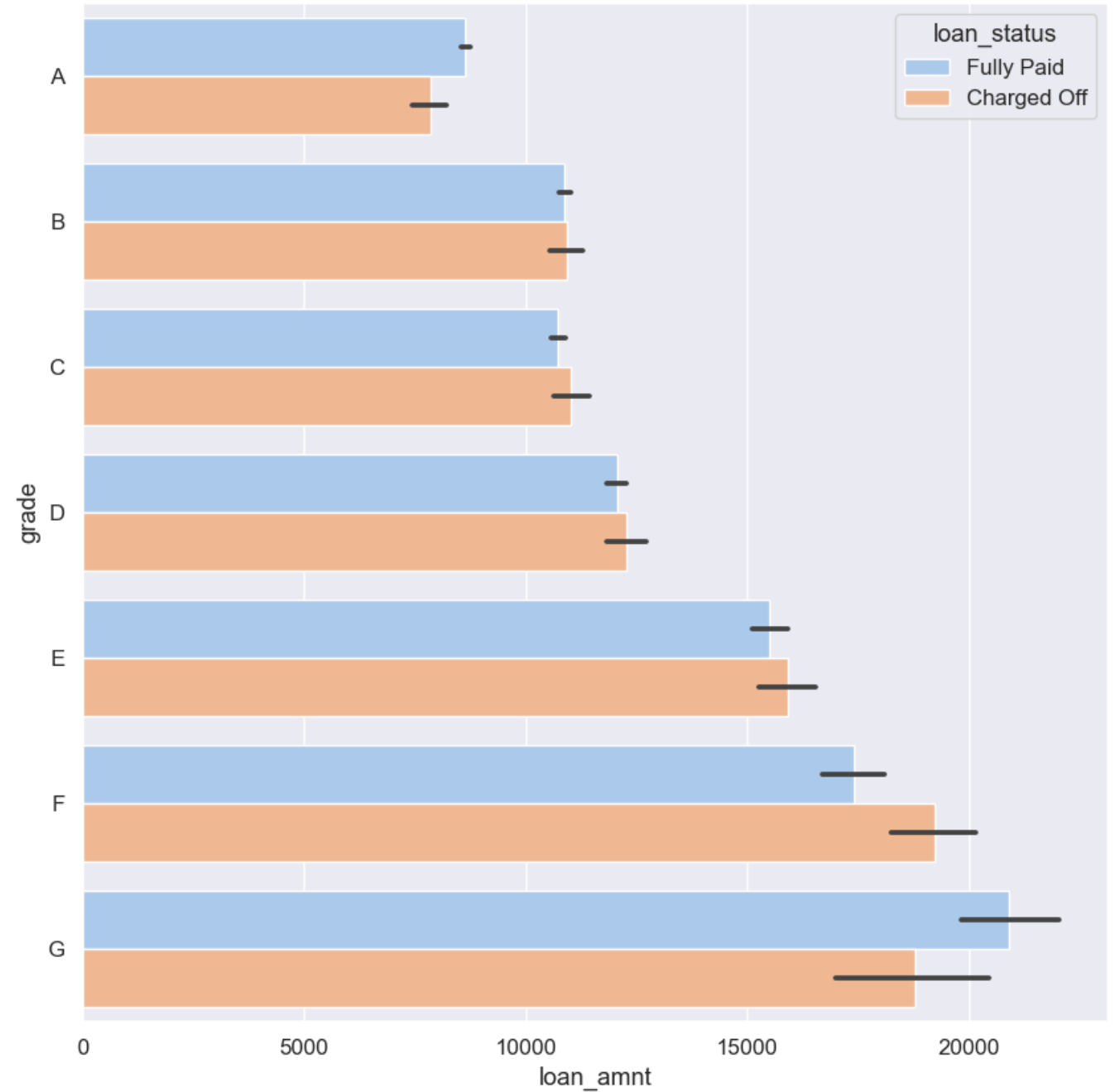
- home ownership vs loan amount vs loan status





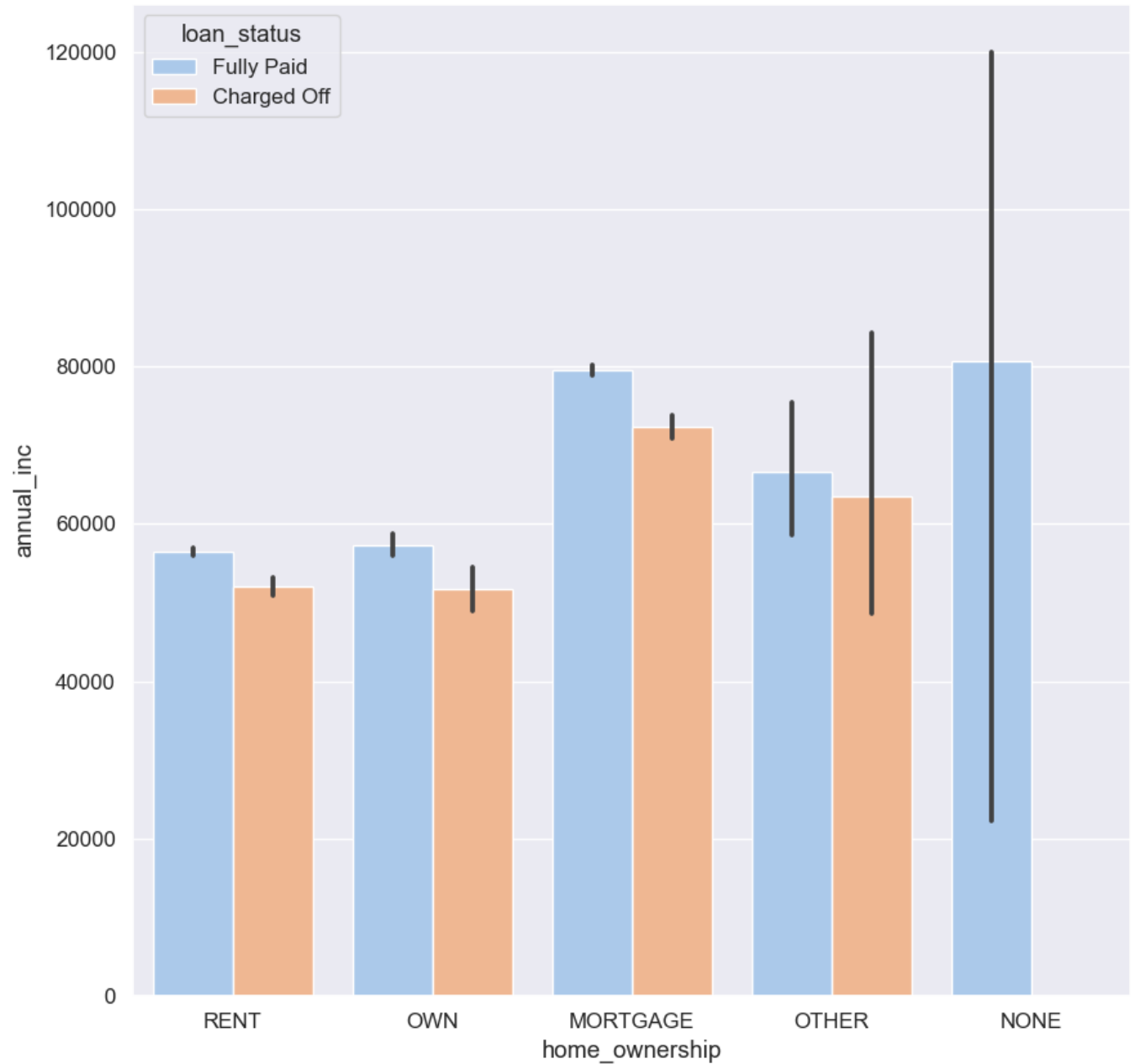
Multivariate Analysis

- grade vs loan amount vs loan status



Multivariate Analysis

- annual income vs home ownership vs loan status

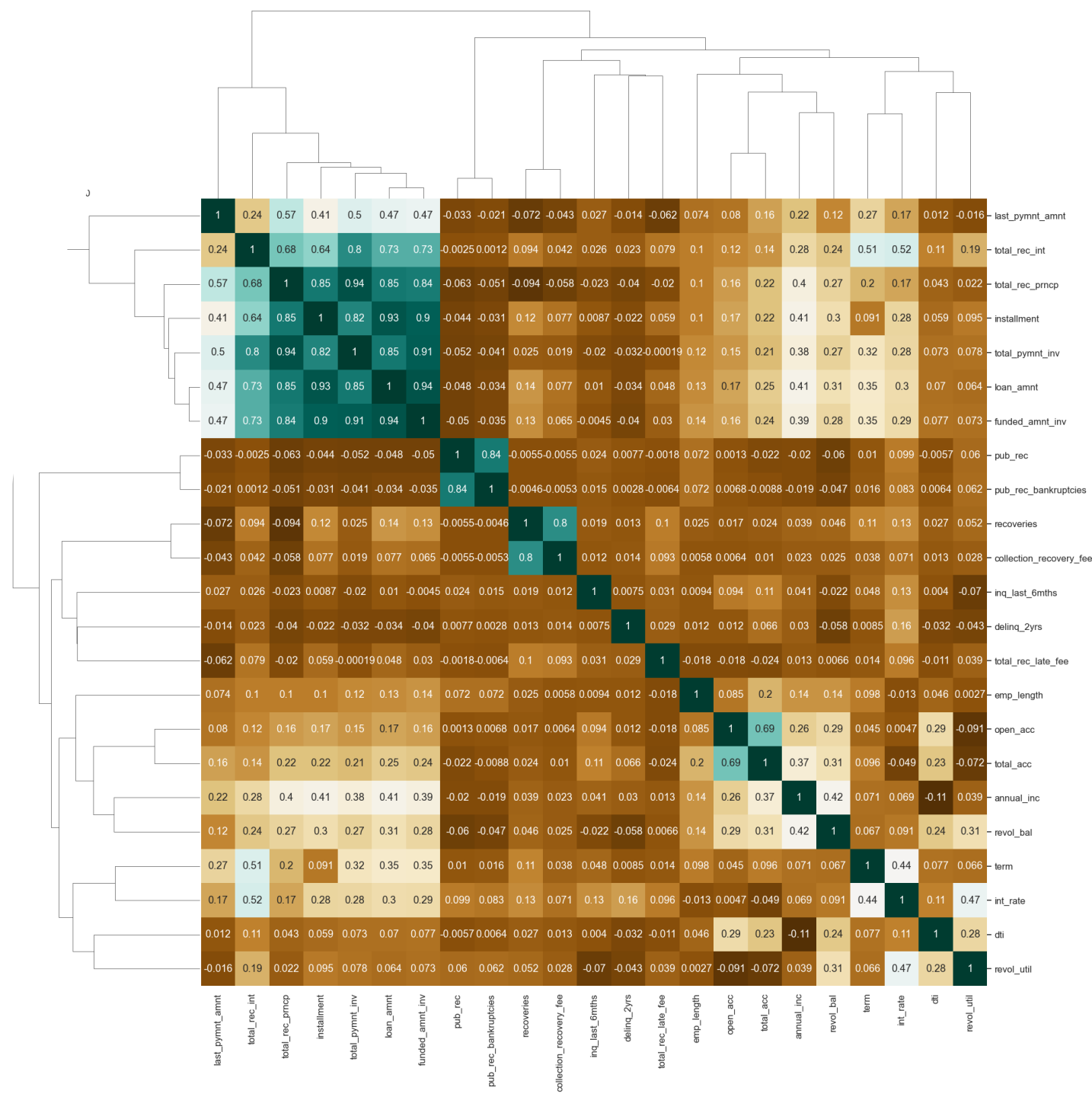


Observations: Multivariate Analysis

- loan amount increases with the increase in annual_income
- interest rate increase with the increase in loan amount
- charged off loans are on the higher interest rates or interest rates are high for charged off loans
- applicants taking loan for 'home' and 'home improvement' have annual income between 60-70k and have a higher probability of defaulting
- applicants who have a MORTGAGE and an annual income of 60-70k are likely for default
- Grade F and applicants and loan amount between 15-20k
- application whose purpose is small business

Correlation Matrix

- Loan Amount, instalment, Payments received to date for portion of total amount funded by investors and The total amount committed by investors for that loan at that point in time are strongly correlated
- Number of derogatory public records and Number of public record bankruptcies are positively correlated
- 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.) and self-reported annual income are negatively correlated



Observations: Correlation Matrix

- loan_amnt, installment, total_payment_inv and funded_amnt_inv are strongly correlated
- pub_rec and pub_rec_bankruptcies are strongly correlated
- dti and annual_inc are negatively correlated



Thank you