Business Objectives

Lending Club Case Study – Defaulter Analysis

- 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.
- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit 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. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
-
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
-
- In other words, the company 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 utilise this knowledge for its portfolio and risk assessment.

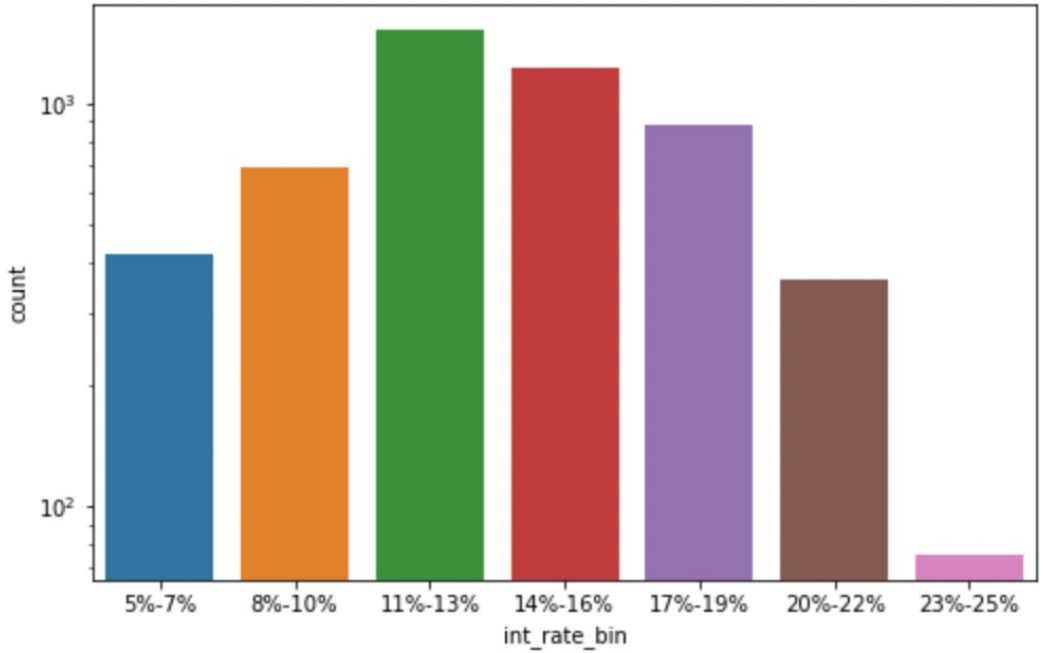
- Puneet Jain (ML C54)

Approach Followed

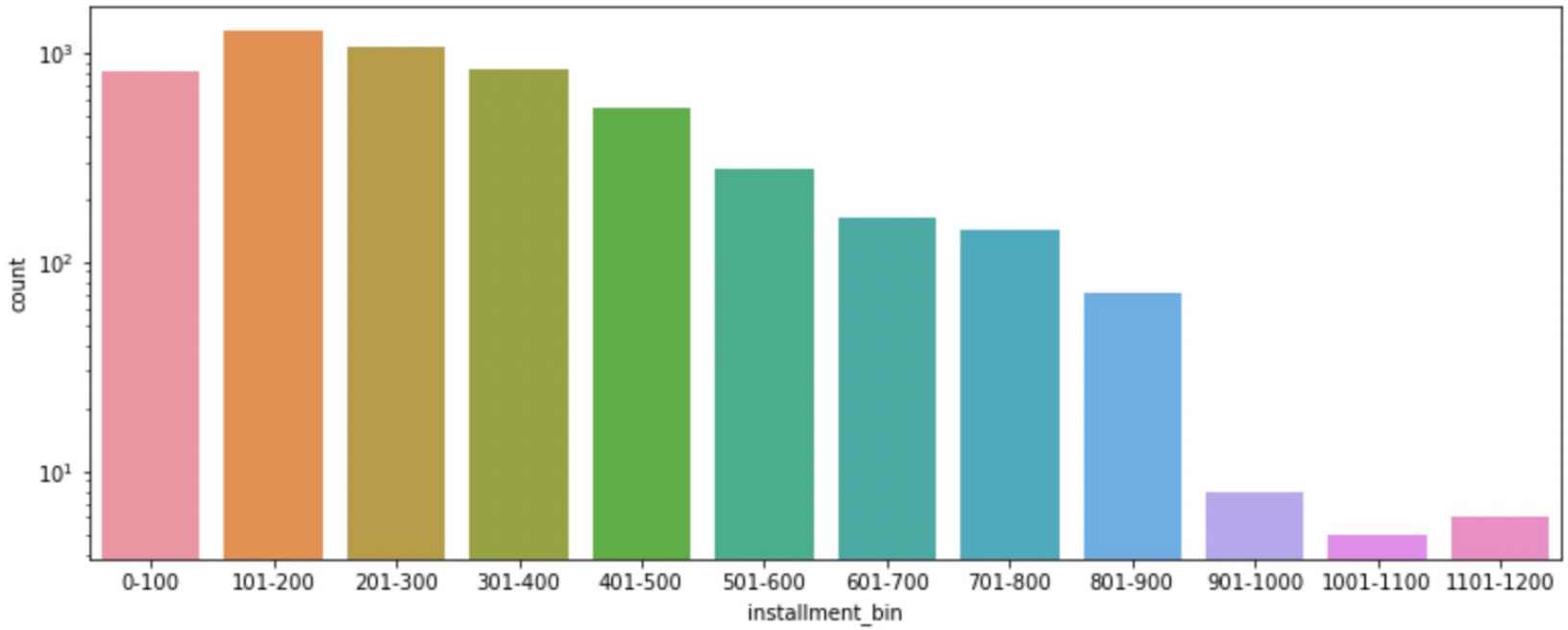
- Check the columns provided and remove the columns with no values populated.
- Remove the columns with only one type of value populated in the entire data set, of if a column hold blank and only one unique value.
- Remove the column which can not add up in the default analysis.
- Remove outliers.
- Filter data for Charged off loans only.
- Update the data with relevant aggregated value for achieve data completeness.
- Remove the column if data is not present in majority of rows.
- Create bin for numerical values to make them categorical.
- Apply univariant analysis.
- Apply bivariant analysis.

Findings – Univariant Analysis

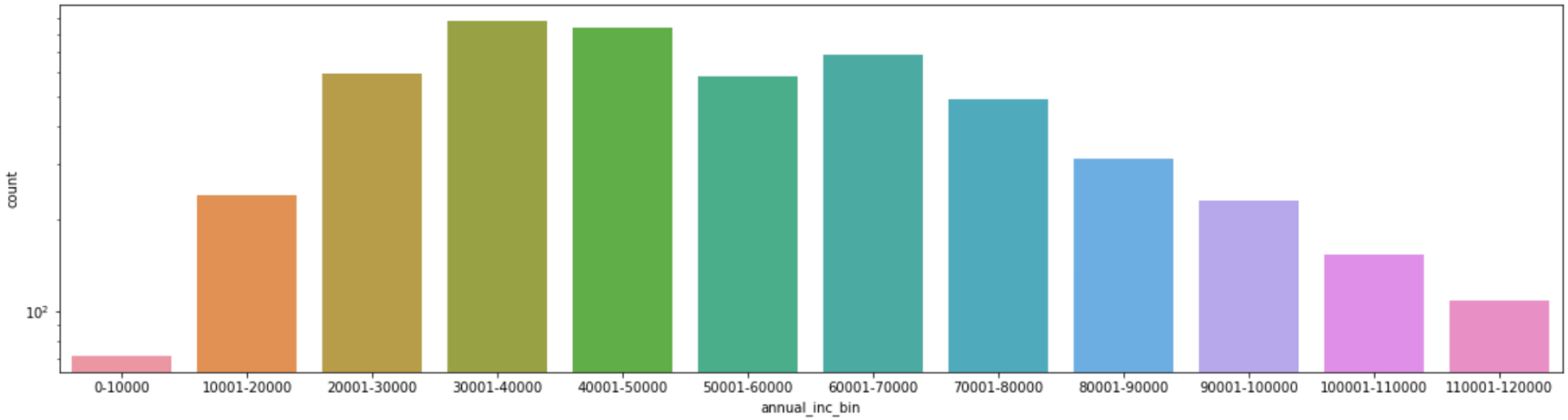
Loans given at 11-13% interest rate have more tendency to default



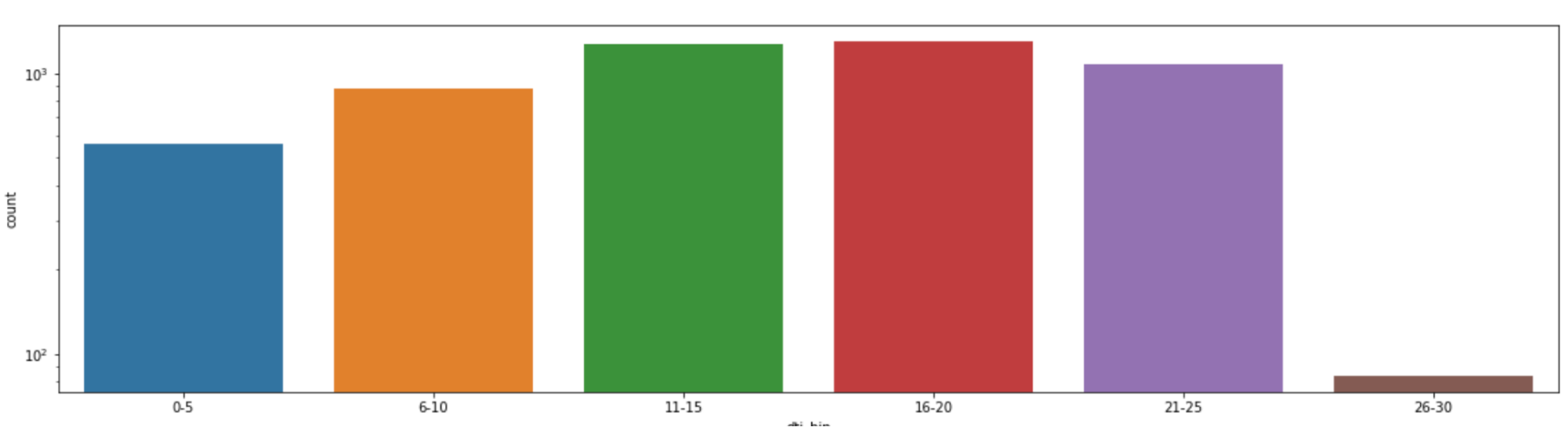
Loans given for 101-200 installment rate have more tendency to default



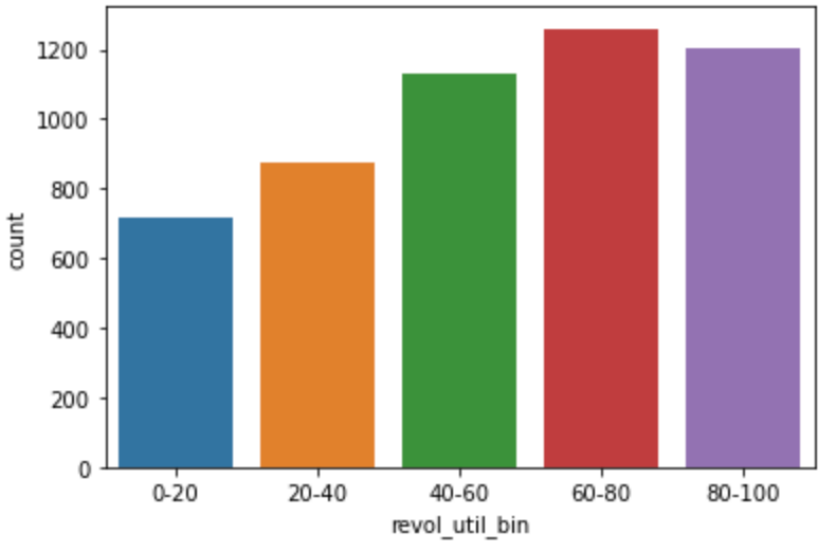
Loans given for 30001-40000 annual_inc have more tendency to default



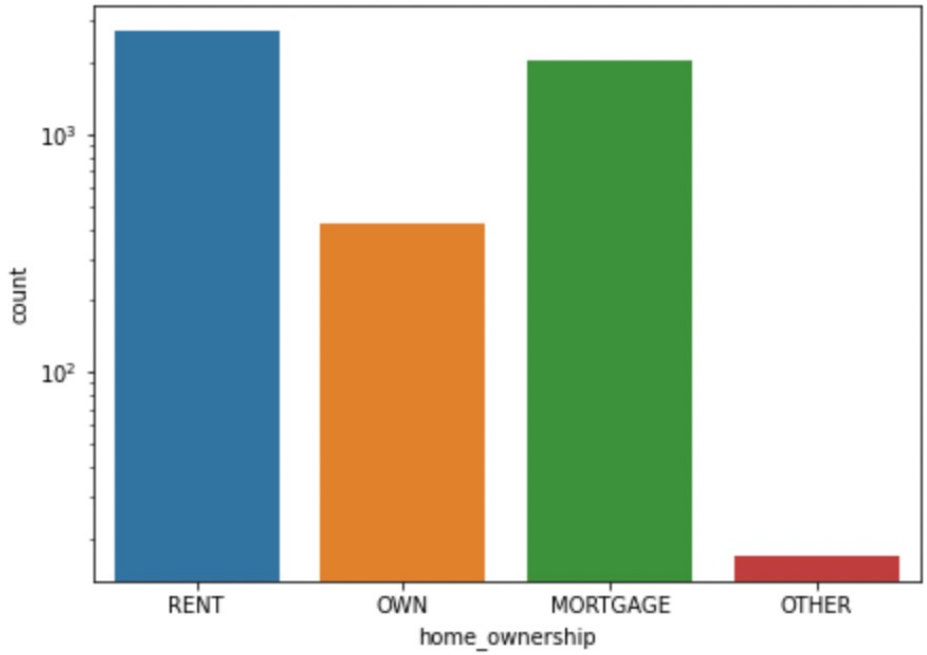
Loans given for 11-20 dti have more tendency to default



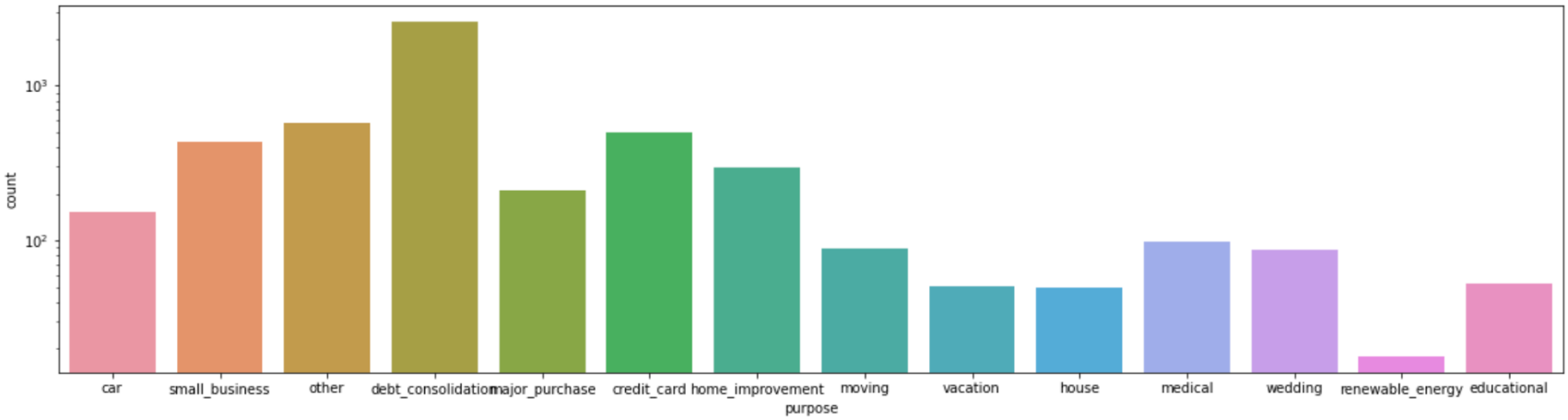
Loans given for 60-80 revol_util have more tendency to default



Loans given to people on rent have more tendency to default

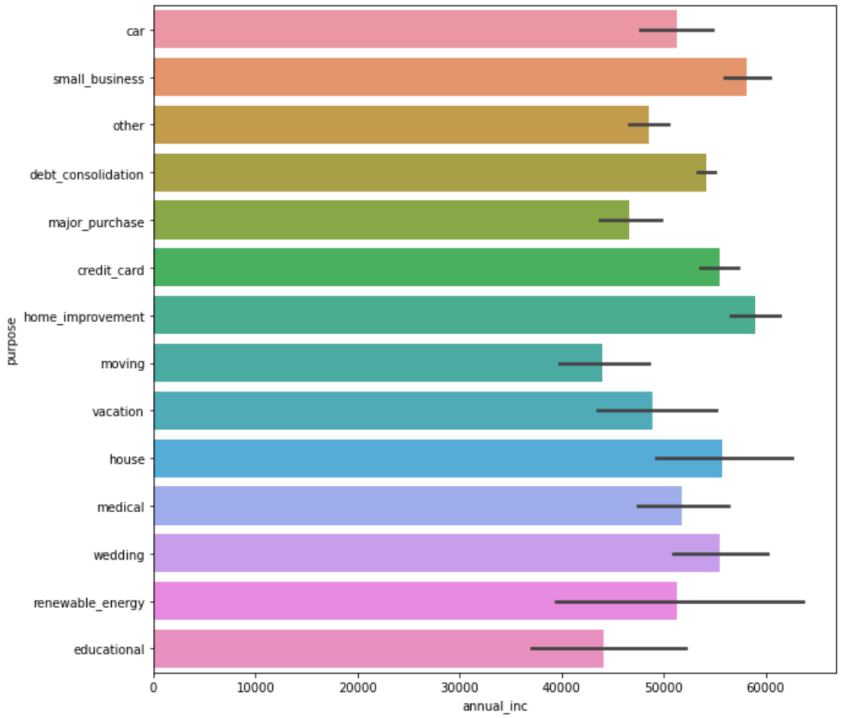


Loans given to people for debt_consolidation have more tendency to default

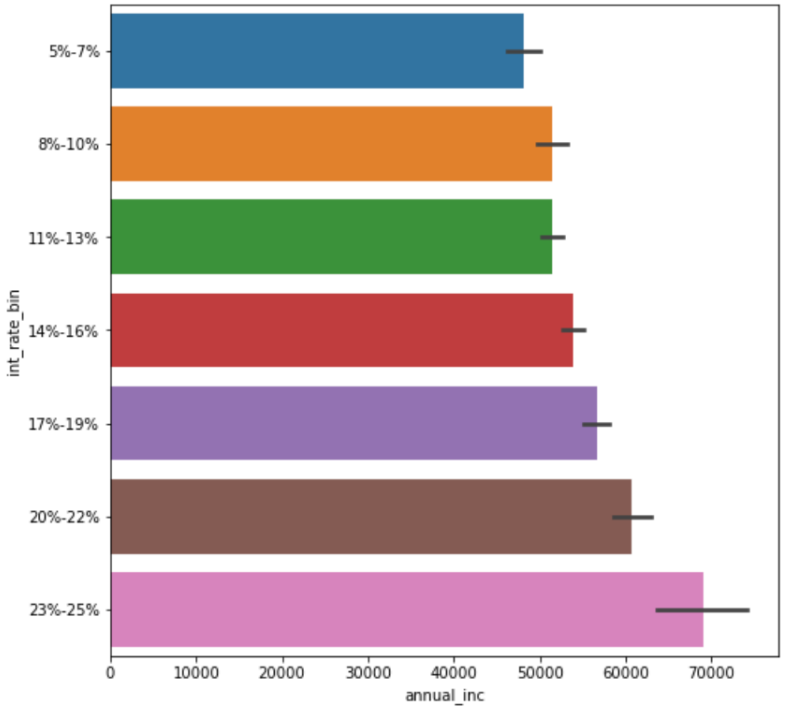


Findings – Bivariant Analysis

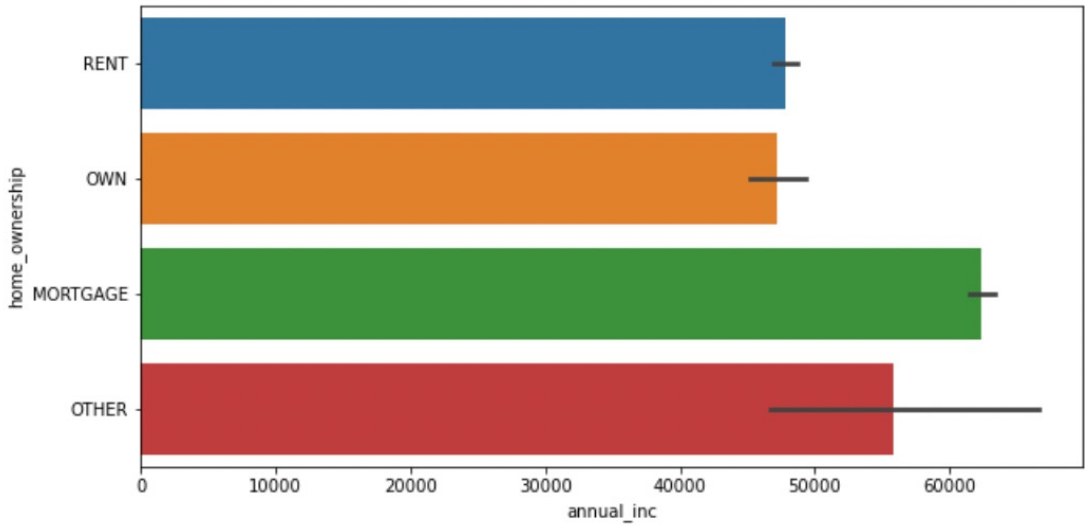
Peopele with salary of 55000 - 60000 and taking loan for home improvement have more tendency to default



People with salary of 65000 - 70000 and given loan at 23-25% intrest have more tendency to default



People with salary greater then 60000 and have home on mortgage have more tendency to default



Higher values loans given on high interest rate have more tendency to default