

# Lending Club Case Study Submission

**Submitted By :**

**Yoshita Dhanwaria**

**Kasi Viswanadh**

# Problem Statement & Key Objectives



## **Business Objective:**

To understand about the drivers of “Default” to quantify the credit risk corresponding to the loan applications. Aiming to minimize the credit loss for the consumer finance company



**Data availability:** Consumer and loan attributes



## **Constraints:**

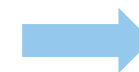
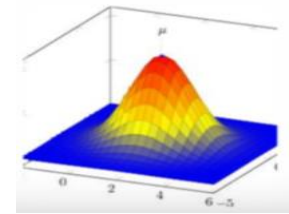
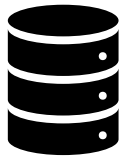
- No transactional information available for the applications that were rejected as lending didn’t happen in such cases
- The driver analysis needs to be done using “EDA”, i.e., without resorting to any modeling techniques



## **Assumptions :**

“Charge off” or “Write off” are synonymous to Default. Usually, when the delinquency past due (DPD) status for applicants move beyond a certain cut-off or threshold, per the regulatory requirements, the lender needs to declare such cases as “Losses” in their P&L accounts which is why it is called writing the loan off the books. The regulatory requirements for writing off loans may differ across portfolios and geographies

# Analytical approach



## Data cleaning

- Data quality checks
- Data formatting and manipulation
- Dimensionality reduction

## Univariate Analysis

- Distributions
- Frequency tables
- Additional variables

## Segmented univariate analysis

- Key metrics across various variable segments
- Derived variables
- Inputs for bivariate analysis

## Bivariate Analysis

- Correlation analysis
- Cross Frequency distribution analysis
- Joint distributions

## Conclusion

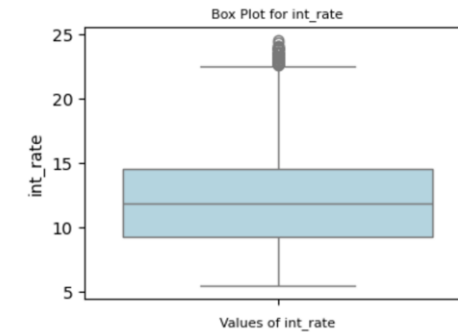
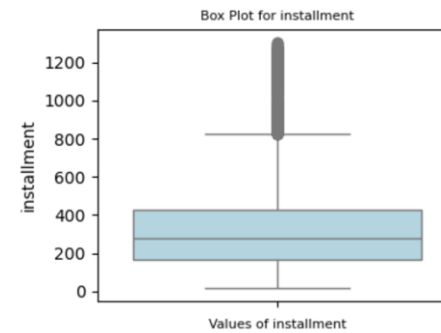
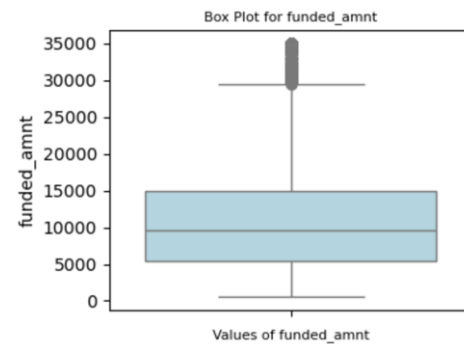
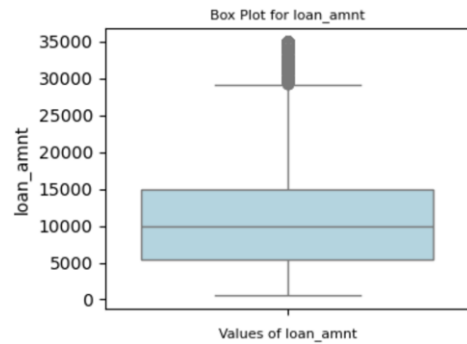
- Combine the results from the analysis
- Actionable insights for the executive team

# Data Cleaning / Preprocessing

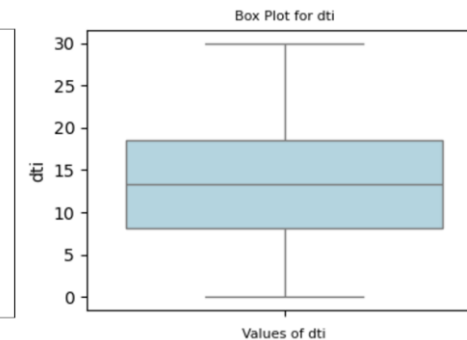
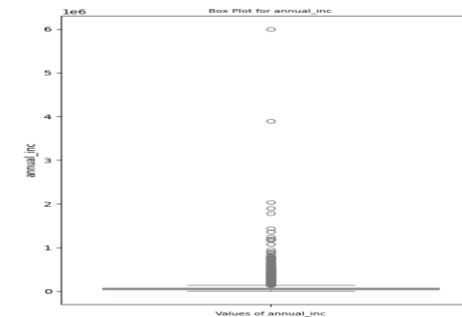
- Dataset has 39,717 observation and 111 variables
- There are 54 columns which having more than 100% NULL
- There are 9 columns which are having only one unique value
- We don't have any duplicated rows in the dataset
- `revol_util` & `int_rate` fields have '%' symbols, and we have treated this variable in code
- `emp_length` clearly filed have "<" & "+" symbols and we have treated this variable in code
- `issue_d`, `earliest_cr_line`, `earliest_cr_line`, `next_pymnt_d`, `last_credit_pull_d`, `last_pymnt_d` are data variable, and we have changed the data type for them to datetime.
- `'emp_title'`, `'desc'`, `'url'`, `'title'`, `'zip_code'` has too many values and won't be useful for data analysis hence we have dropped them.
- `'mths_since_last_record'`, `'next_pymnt_d'` has 90% value as missing hence dropped them



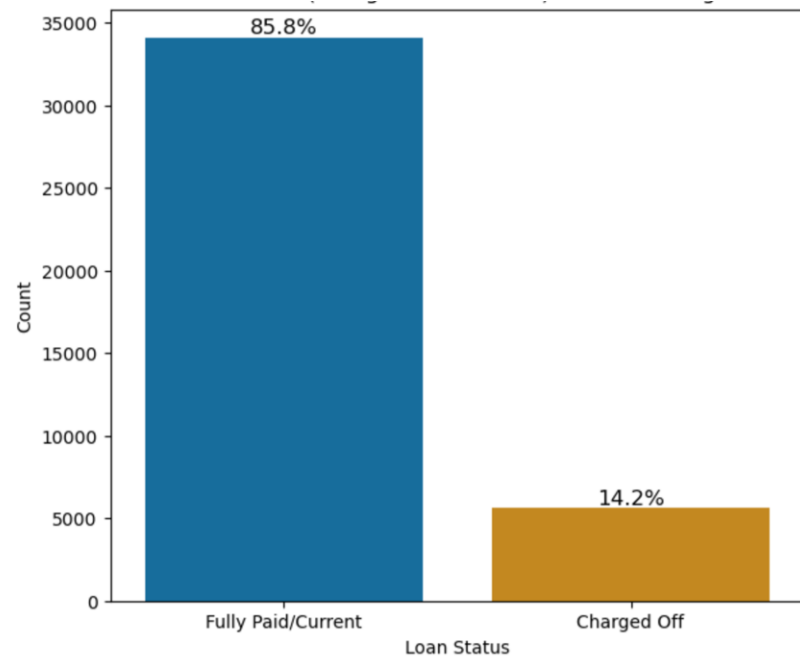
# Quantitative variables Analysis



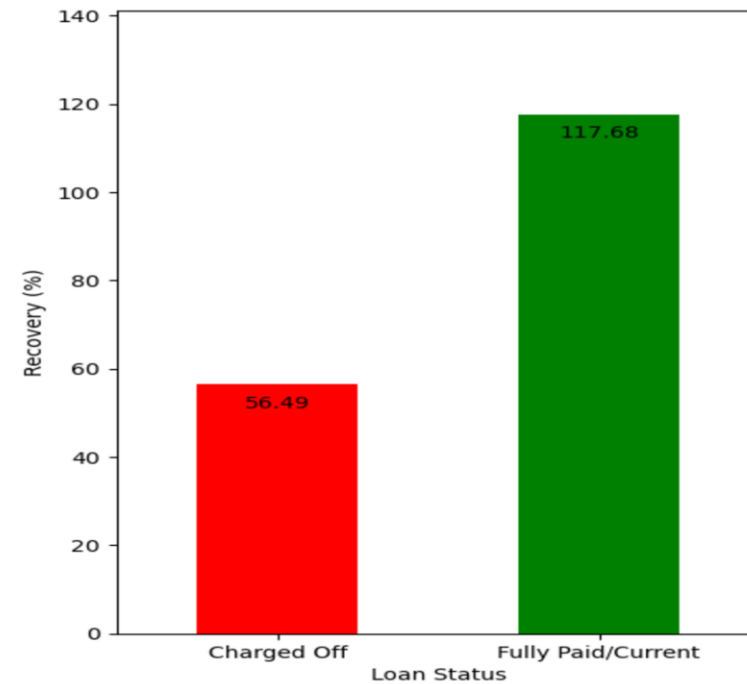
- The annual income of most of the loan applicants is between 40,404 - 82,300
- The loan amount of most of the loan applicants is between 5,500 - 15,000
- The funded amount by investor for most of the loan applicants is between 5,400 - 15,000
- The interest rate on the loan is between 9.25% - 14.59%
- The monthly installment amount on the loan is between 167 - 430
- The debt-to-income ratio is between 8.17 - 18.60



# Defaults & recovery

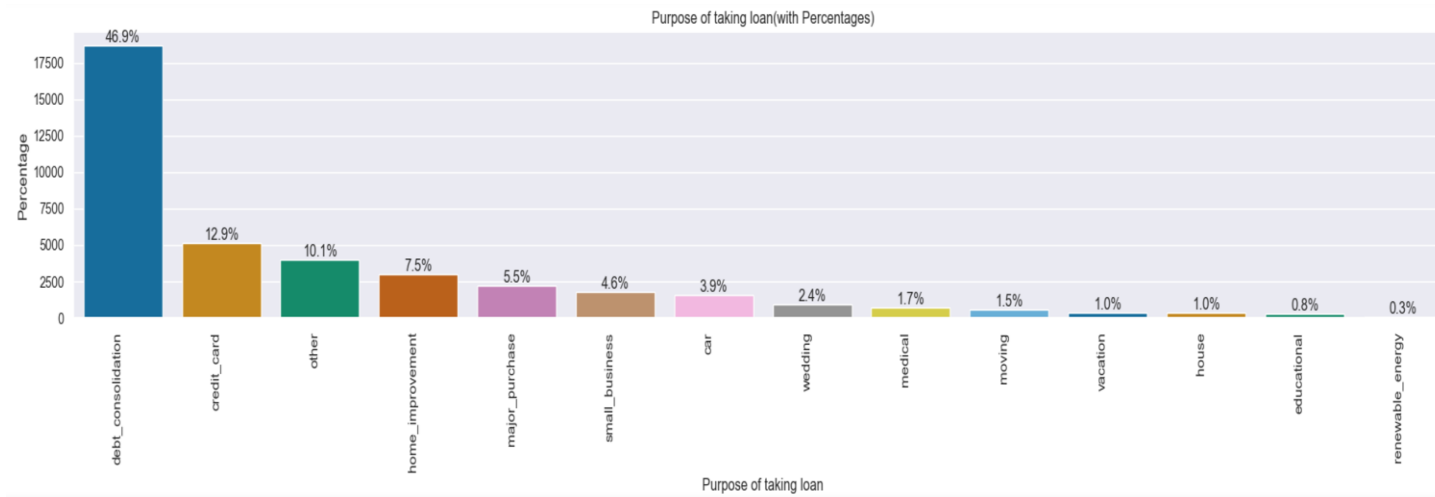


The default rate of the portfolio is ~14%



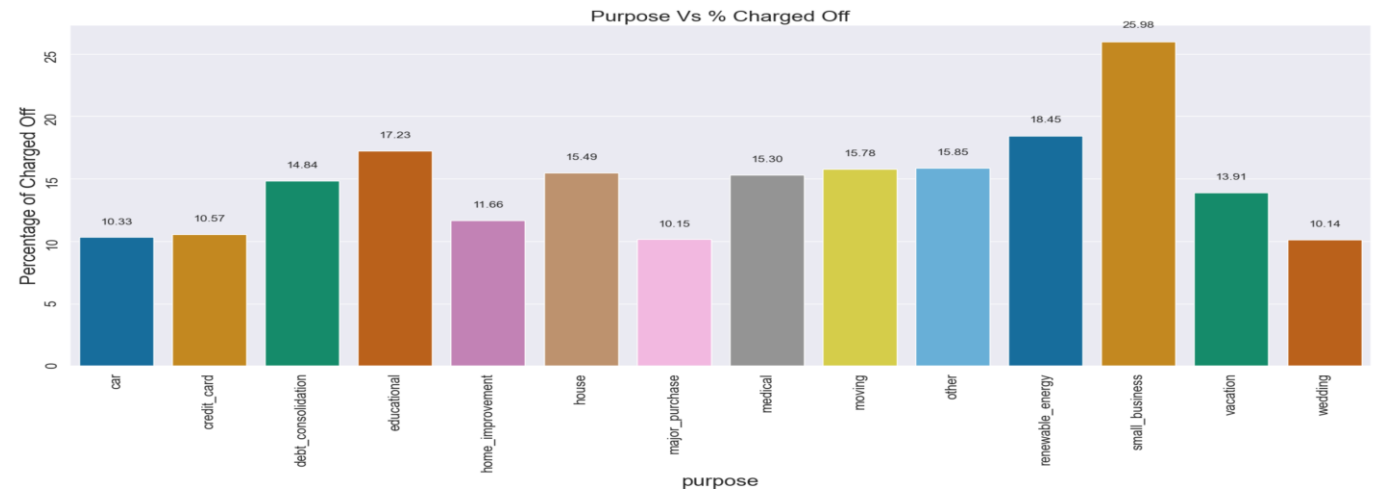
Only ~56% of the loan amount is recovered from the defaulters

# Purpose & riskiness

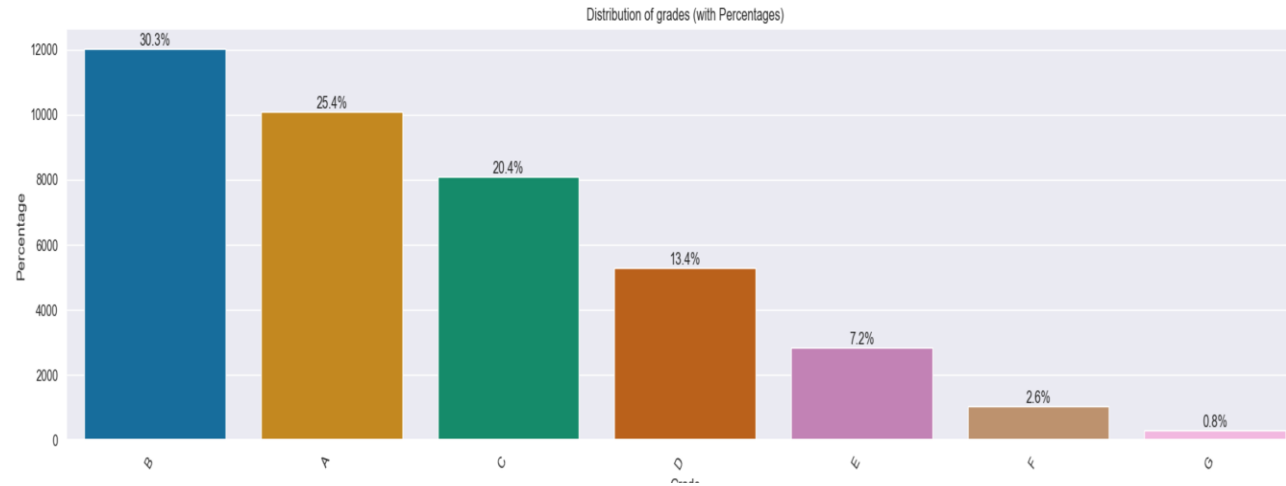


~47% loans' purpose is Debt consolidation followed by Credit Cards at ~13%

Small business loans seem to be the riskiest while wedding loans have least defaults

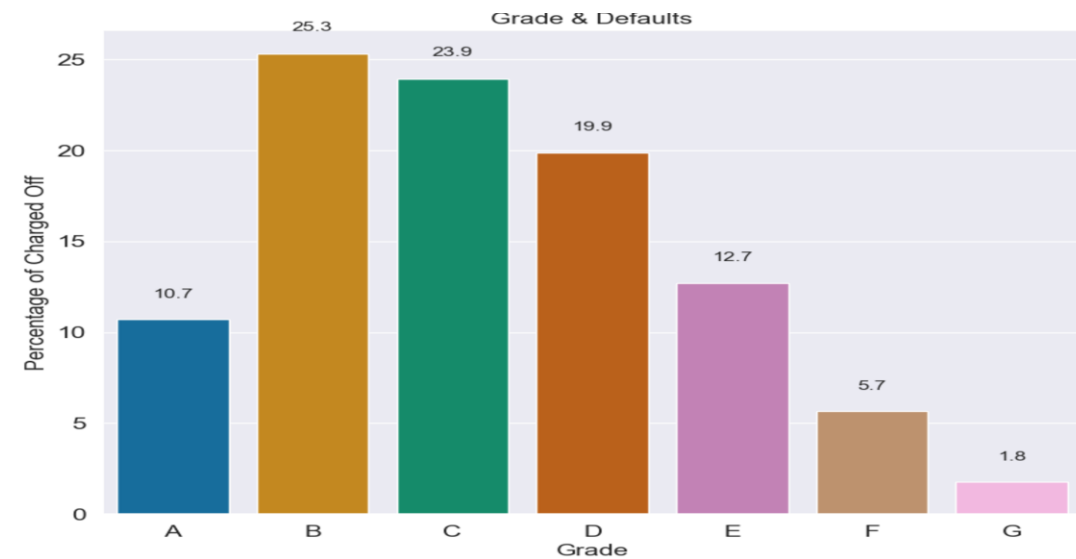


# Grade



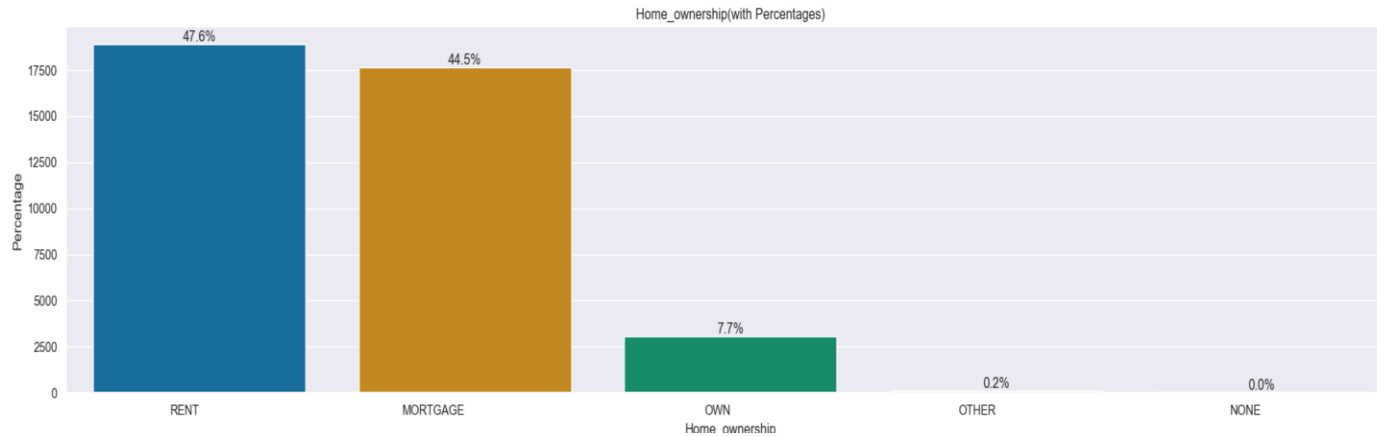
~30% loans belongs to grade B followed by grade A at ~25%

45 % of defaulters belong to grade B & grade C



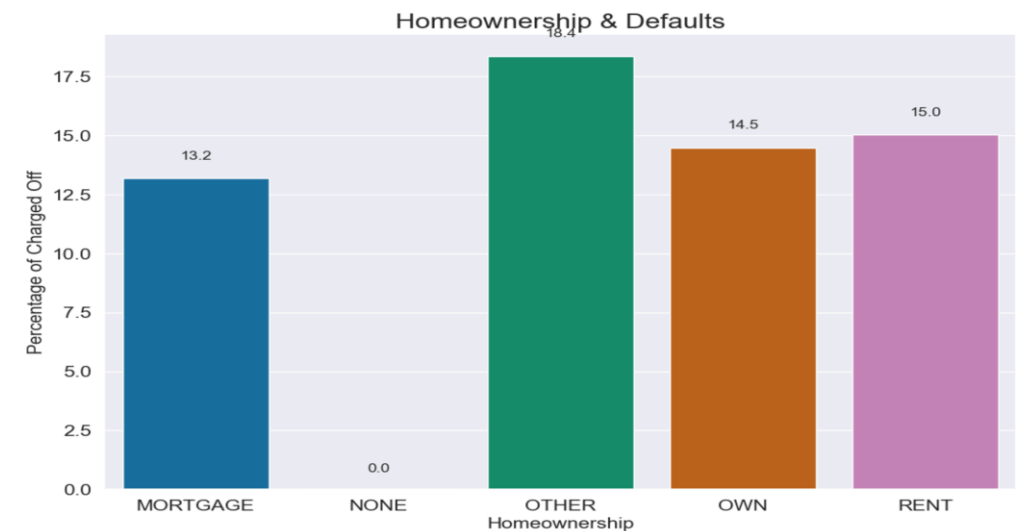


# Homeownership

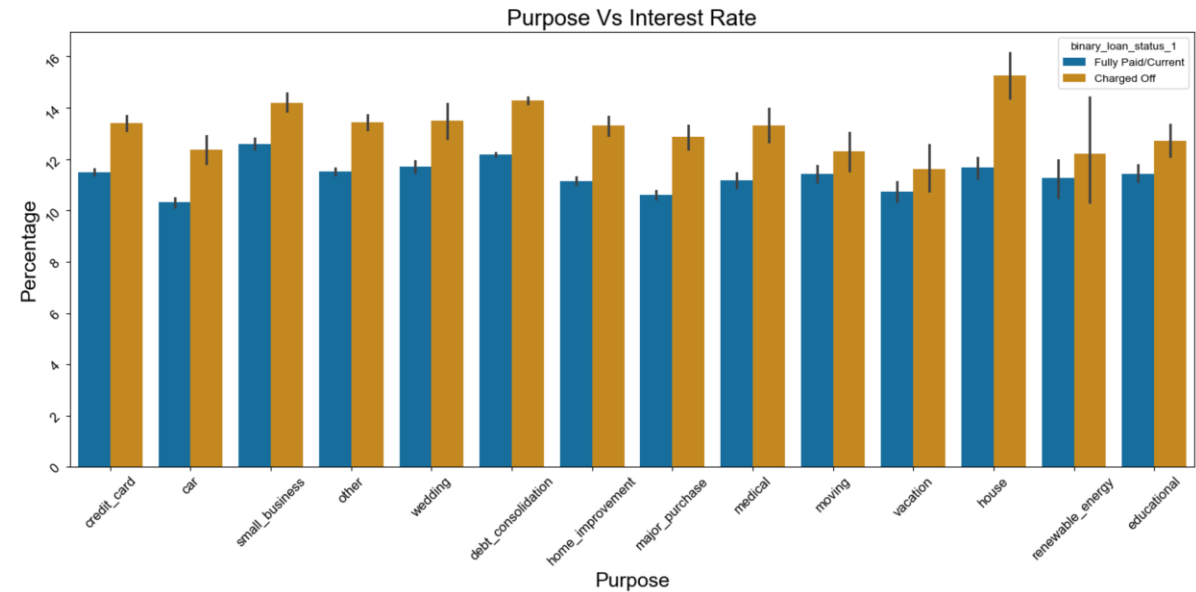
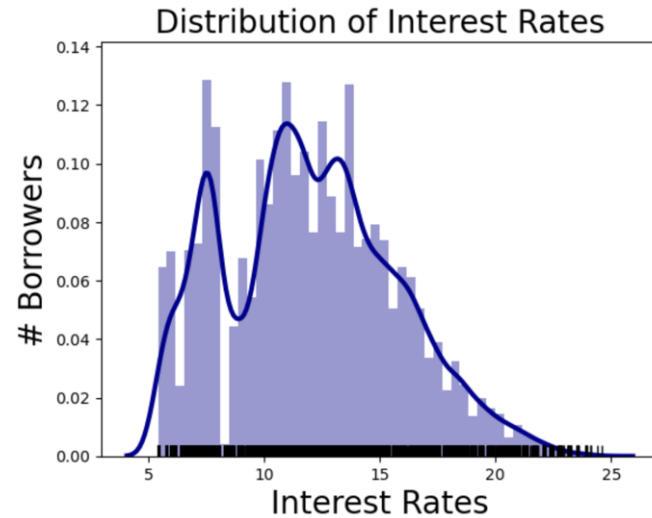


~47% loans belongs to people staying in rented house followed by mortgage at ~45%

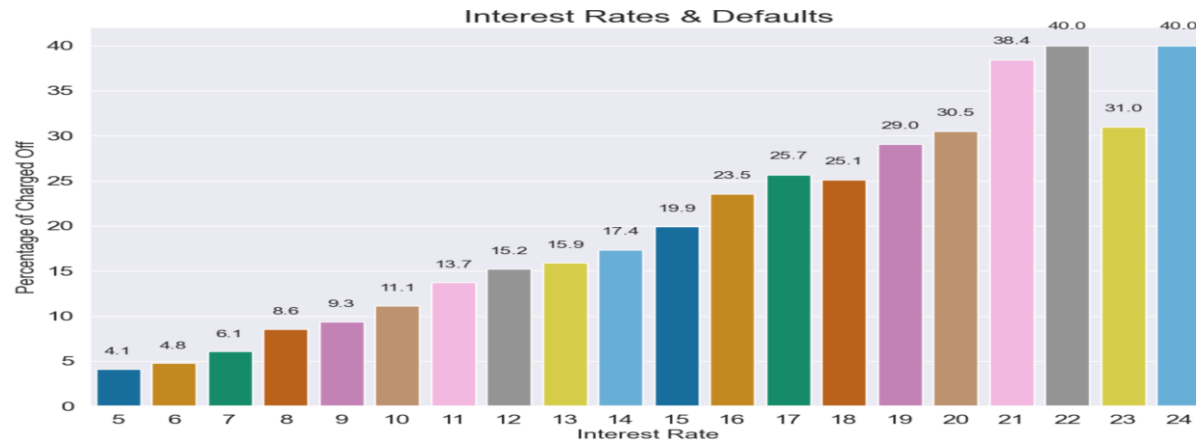
~18 % are defaulters in others category



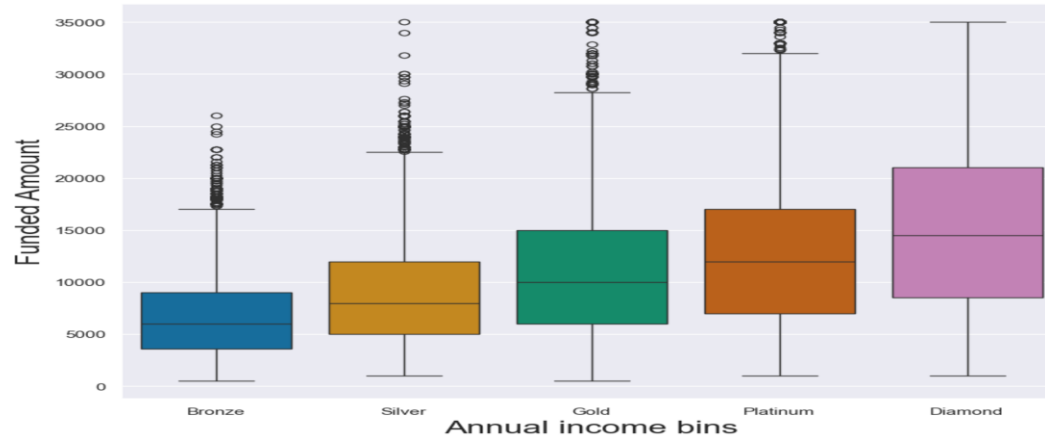
# Interest rates



- The default rates increase significantly as lending rate increases
- This also implies that the lender is assessing the risk correctly which is why a risk premium is being added to the lending rate

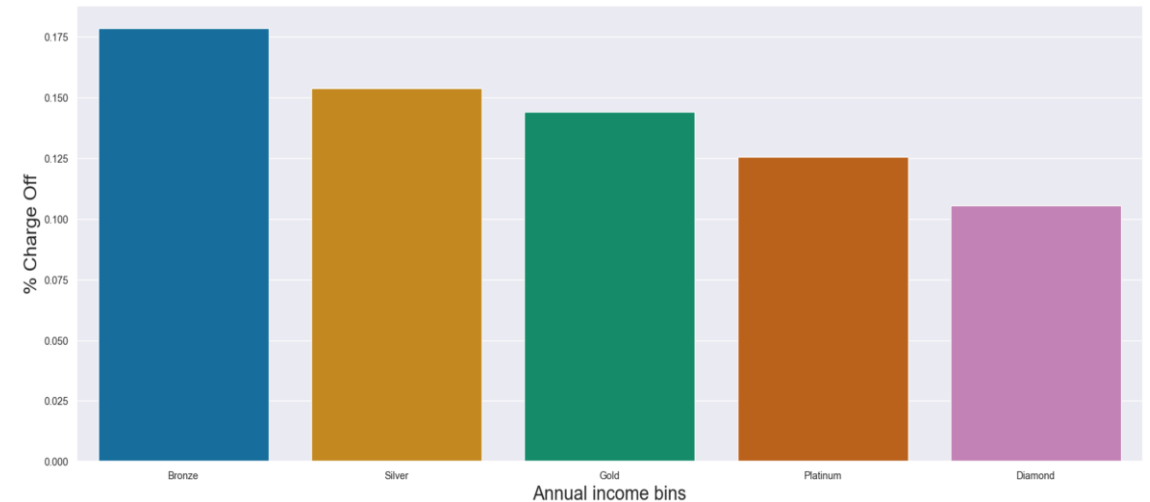


# Annual income buckets and funded amount : Defaults

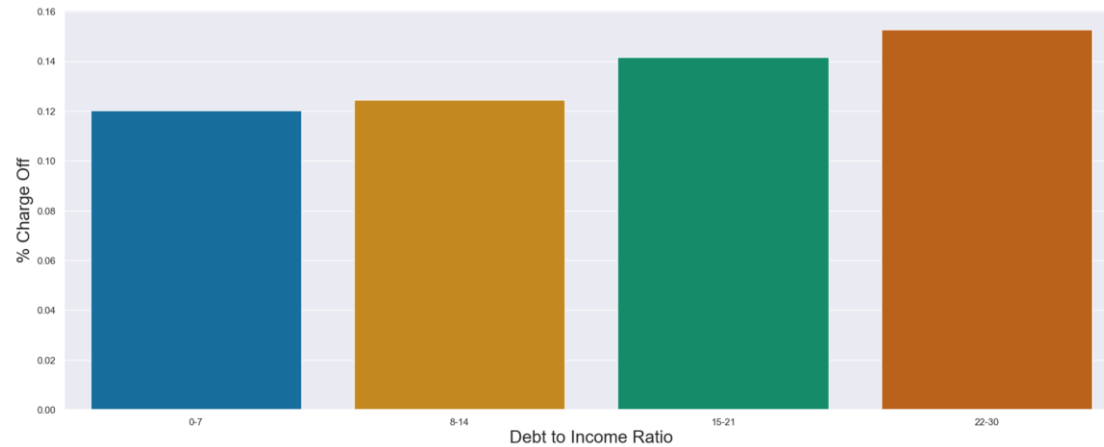


- Assessing the variation in annual income of borrowers, we created homogenous income bins
- The funded loan amount increases with the increase in annual income, probably high-income borrowers are perceived less risky

- Higher income brackets correspond to lesser default rates
- This validates our hypothesis of higher income borrowers being perceived as less risky

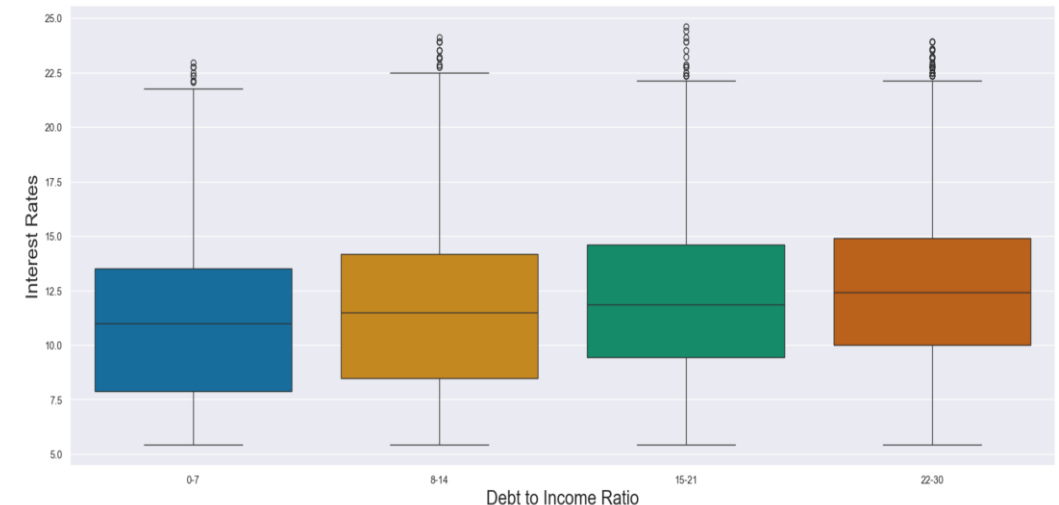


# Debt to Income Ratio & Credit risk

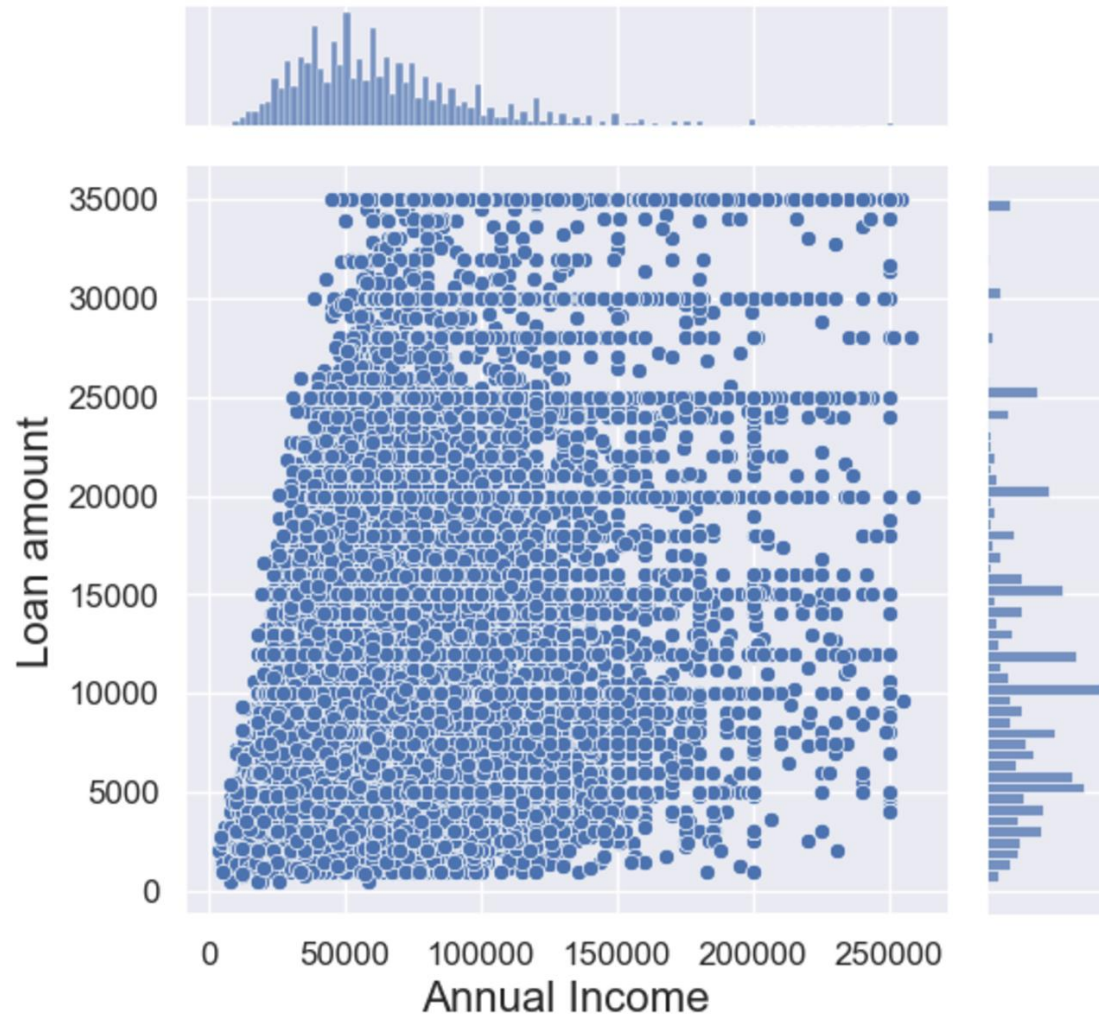


- Debt to Income Ratio is one of the key drivers for credit risk
- A higher Debt to income ratio corresponds to a higher Default Rate, as is evident in the bar plot

- Given that a higher DTI borrower is perceived riskier, we can see that the lending rate increases with an increase in the DTI, the least interest rate being offered to the lowest DTI bucket

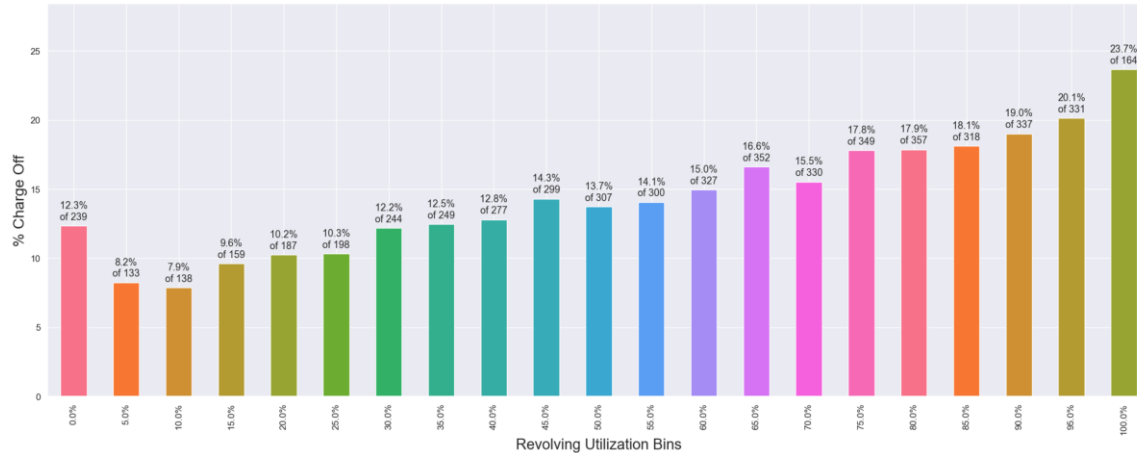


# Joint distribution analysis : Loan amount & Annual income



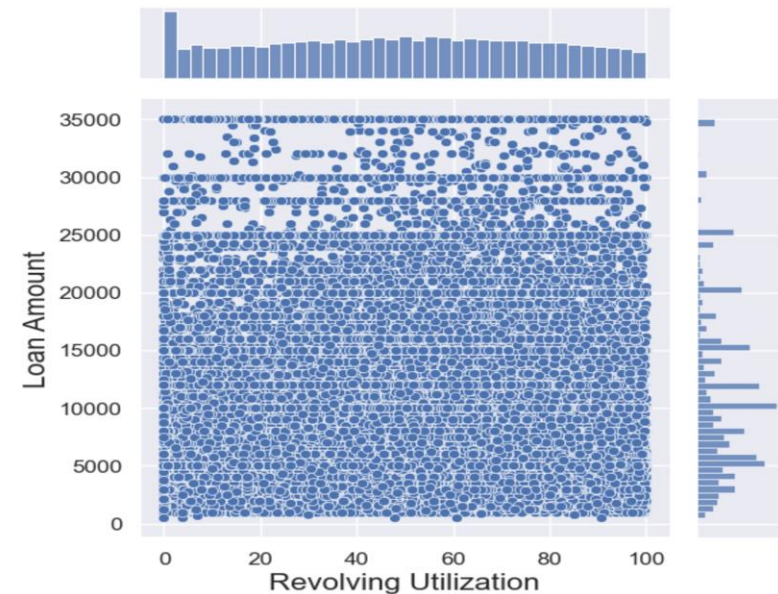
- We notice a few cases where the lender has extended high-value loans to low-income borrowers
- For example, borrowers with annual income of less than 50,000 with approved loans of more than 25,000
- Building on top of the analysis done so far, this increases the inherent risk in the portfolio
- To reduce the risk of the portfolio, such cases should be avoided

# Utilization of revolving line of credit



Higher utilization of revolving line of credit are perceived more risky, as apparent in the higher default rate

We see high-value loans being approved to borrowers with revolving rate higher than 80%, this should be avoided to sanitize the inherent credit risk in the portfolio



# Joint distribution analysis : Loan amount/Interest Rate & Debt to Income Ratio



There are opportunities of reducing the credit risk of the portfolio by avoiding giving high value loans to borrowers with high Debt to Income Ratio. If at all, such borrowers are extended loans, they should be charged a higher interest rate as risk premium



# Recommendations



**STOP**

- Approving high-value loans to borrowers with Revolving Utilization more than 80%
- Approving high-value loans being taken for Small-business purposes



- Giving loans to people staying on loan or having mortgage
- Approving loan amount in excess of 40% of the borrower's annual income



**START**

- Charging higher interest rates from borrowers with high DTI, at least on high value loans
- Approving more loans in less risky propositions like Wedding, car and home improvement