Lending Club Case Study

Group Facilitator - Ankita Darekar
Team Member - Anurag Sharma
Batch: AI&ML Nov’24 C 71

Contents

Problem Statement

A consumer finance company specializes in providing loans to urban customers. Loan approval decisions are critical, as they involve two key risks:

1. Loss of business if reliable applicants are rejected.
2. Financial loss if risky applicants, who are likely to default, are approved.

The company aims to identify patterns in customer and loan attributes that indicate the likelihood of default. This insight will help in making informed decisions, such as denying loans, adjusting loan amounts, or offering loans at higher interest rates to mitigate risk.

Objective:

Using EDA, analyze loan and customer data to uncover driving factors behind loan defaults. This will enable the company to minimize credit loss, improve portfolio management, and optimize risk assessment strategies.

Approach

- Data cleaning & Manipulation
- Data Standardization
- Derived metrics
- Exploratory Data Analysis on Univariate, Bi-varitate, Multivariate.

Data Cleaning & Manipulation

- **Fix Rows/Columns:** Remove empty columns and irrelevant rows.
- **Handle Missing Values:** Drop columns/rows based on missing value percentages.
- **Standardize Data:** Fix dates, numeric fields, and data types.
- **Correct Invalid Values:** Remove unexpected text, symbols, etc.
- **Filter Data:** Refine dataset as needed.
- **Fix Data Types:** Correct mismatched types.
- **Treat Outliers:** Manage anomalies effectively.
- **Dropped columns** containing only a single value.
- **Dropped rows** with more than 65% missing values.
- **Excluded data** for current loans, Fully Paid loan.

Handling Missing Values and Standardising

Impute value:

- As we can see mode is 10+ years and it is almost twice the occurrence of second most frequent value.
- Hence we can assign this value to null values in emp_length
- The missing values are also very low in number imputing it, won't cause any biases

Standardized the value(changing and correcting the data types)

- Removing % from Interest Rate
- Creating numeric from term (Object)
- Object to Int wherever its required
- Funded_amount to int
- Float to Int wherever required

Handling Outliers and Derived Metrics

Outlier Treatment:

- **Annual income:** we have remove outlier ranging more than 95th percentile.

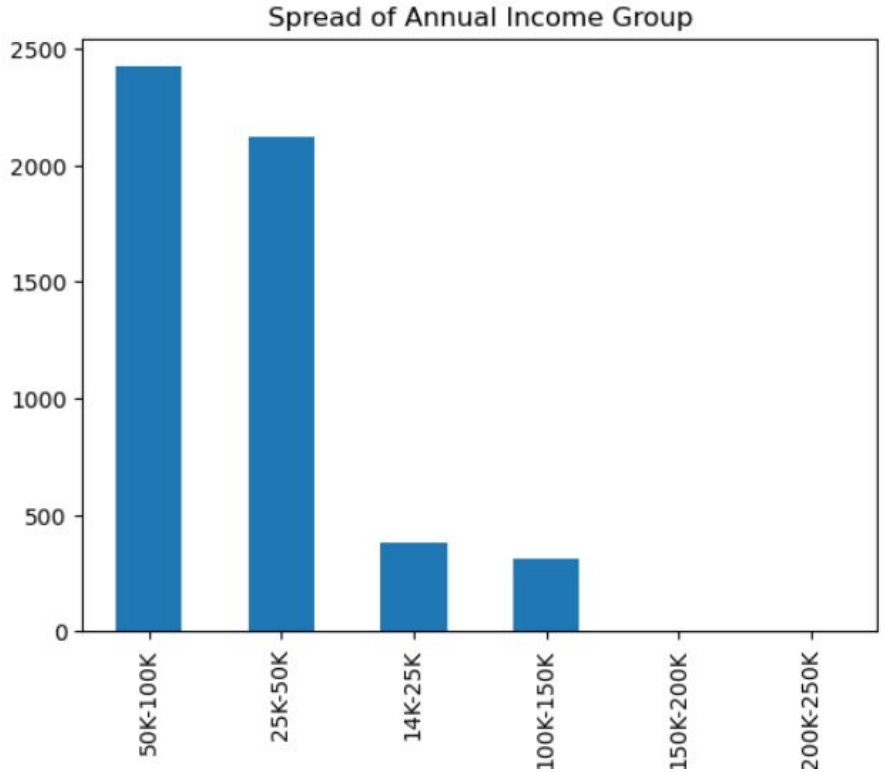
Derived Metrics:

- We have created a derived column based on annual income of customer to identify the group of salaries of customer who have defaulted the most.

Segmented Univariate Analysis

EDA - Annual Income Group Analysis Observations

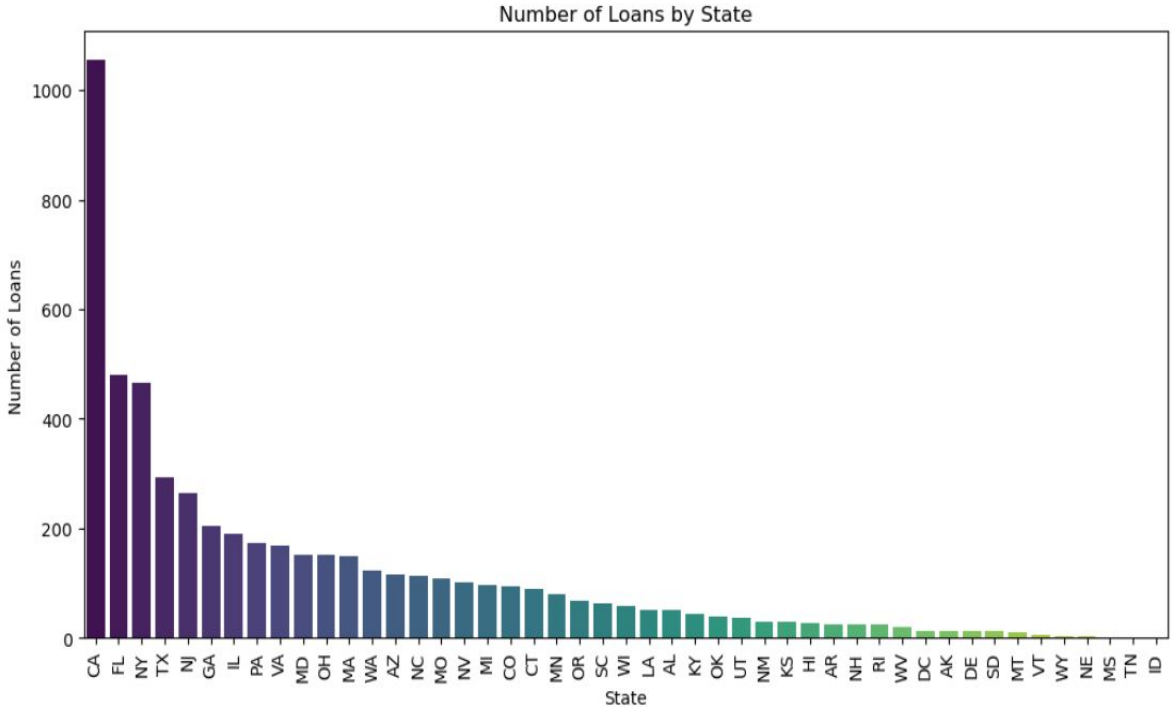
1. From this we can see, The highest loan is issued to the income group 50K-100K USD.
2. We can clearly see if the annual income is high, there are very less number of defaulters.



Univariate Analysis

EDA - Home State of Customer Analysis Observations

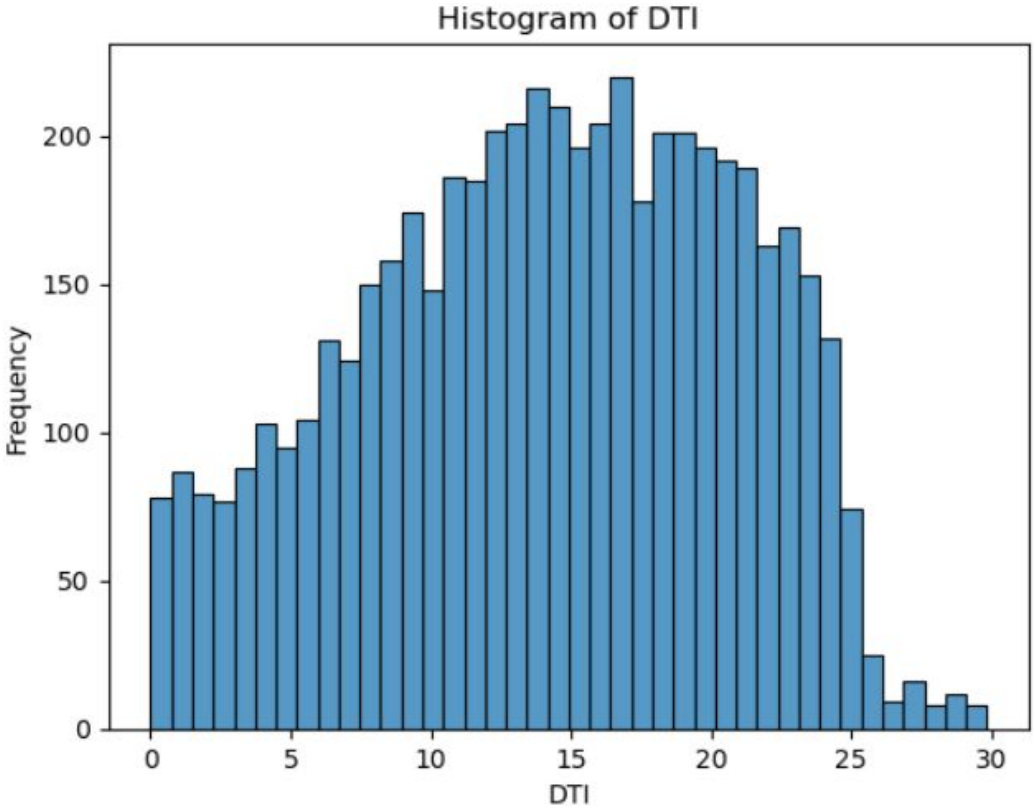
1. **Distribution**: The bar plot shows the distribution of counts across different states.
2. **Highest Count**: California (CA) has the highest count, indicating it might be the state with the most loans/customers.
3. **Lowest Count**: Illinois (IL) has the lowest count among the states listed.



Univariate Analysis

EDA - DTI of Customer Analysis Observations

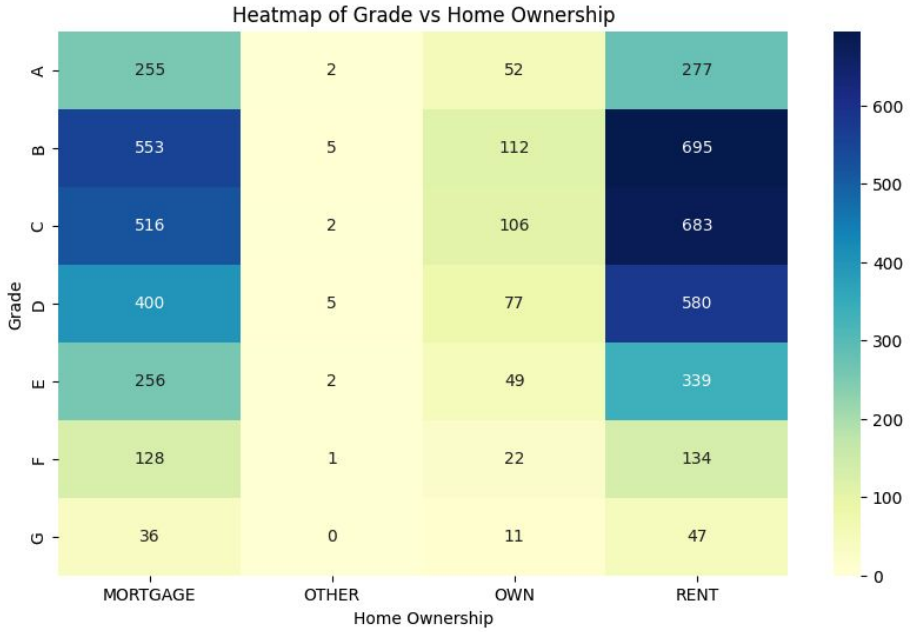
1. From the observation above we can see that maximum number customer have DTI between 10 to 24
2. Customers where DTI is between 25 to 30 have less occurrence of defaulters



Bivariate Analysis

EDA - Grade Vs Home Ownership Analysis Observations

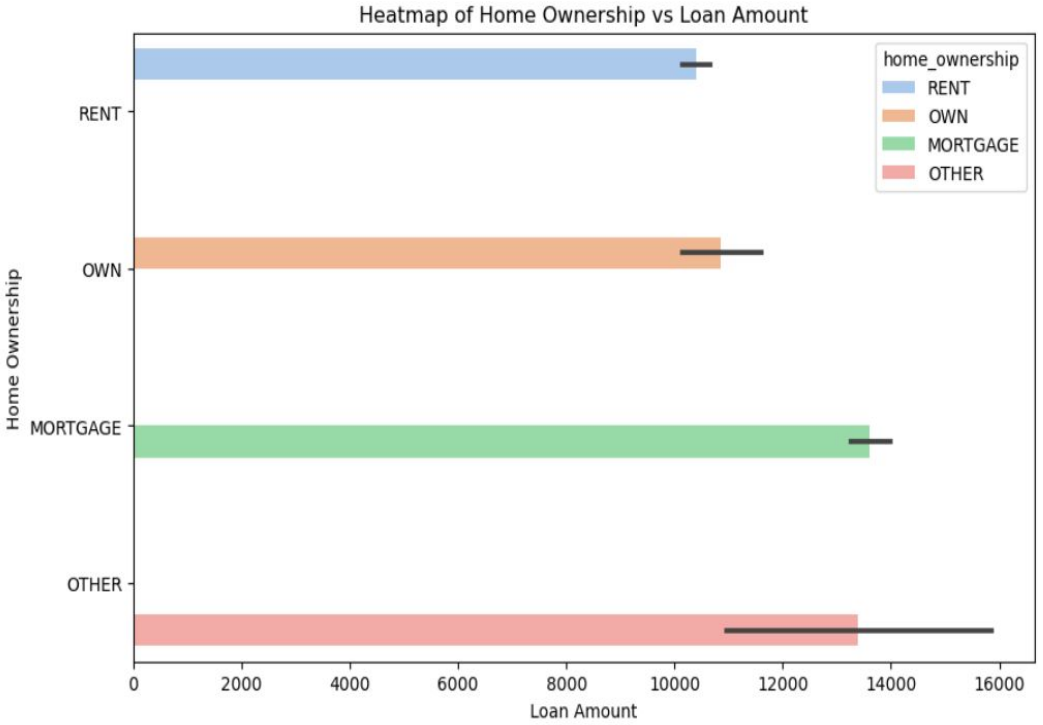
1. The heatmap provides a clear visual representation of the frequency distribution of these two variables, with annotations showing the exact counts.
2. From Graph it is visible that customer who belongs to Grade B and C living on rent and mortgage house have much higher tendency to be defaulted on loan.



Bivariate Analysis

EDA - Home Ownership vs Loan Amount Analysis Observations

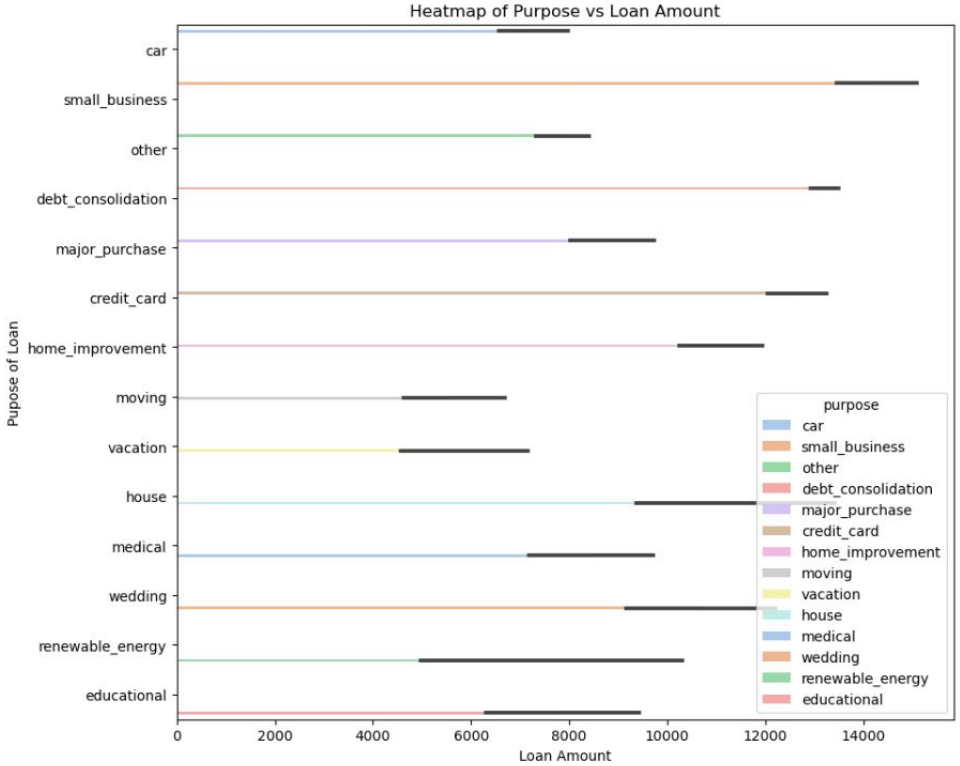
1. From graph it is visible that customer which are having homeownership as Mortgage and Others have significantly more loan amount compared with rent or own category
2. We have already seen then mortgage homeowner are more prone to default on loan, hence this can be consider while deciding the loan amount.



Bivariate Analysis

EDA - Loan Amount Vs Loan Purpose Analysis Observations

1. From Graph we can see that most of the default customer have taken loan on for small business.
2. Apart from small business, debt consolidation is second highest factor for loan application.



Key Findings

1. The highest loan is issued to the income group 50K-100K USD. We can clearly see if the annual income is high, there are very less number of defaulters.
2. California (CA) has the highest count, indicating it might be the state with the most loans/customers. Illinois (IL) has the lowest count among the states listed.
3. The maximum number customer have DTI between 10 to 24. Customers where DTI is between 25 to 30 have less occurrence of defaulters
4. Customer who belongs to Grade B and C living on rent and mortgage house have much higher tendency to be defaulted on loan.
5. Mortgage and Others have significantly more loan amount compared with rent or own category.
6. Most of the default customer have taken loan on for small business. Apart from small business, debt consolidation is second highest factor for loan application.

Recommendations

1. Tighten eligibility criteria for loans in this income bracket to ensure borrowers can repay without overstretching their finances.
2. Since California has the highest number of loan defaulters, it indicates either a high volume of loans or a riskier lending environment in the state.Perform a deeper analysis of the California market, considering factors like unemployment rates, cost of living, and other socio-economic factors that might be contributing to defaults.
3. Consider reducing the allowable DTI ratio for higher-risk borrowers, especially in the 10-24 range, as their default rates are higher.
4. Apply more stringent credit assessments for renters and homeowners with lower credit grades (B and C). This might include higher interest rates or more conservative loan terms.
5. Introduce loan caps or tiered loan products for high-risk customers, especially for homeownership as mortgage, to avoid overextension of borrowers' financial capacities.
6. Implement more rigorous business assessments for small business loans, including viability checks, cash flow analysis, and business plan evaluations to ensure borrowers can repay without going into default.
7. Develop predictive models that monitor borrower behavior and trigger alerts for potential issues, such as missed payments, increases in debt levels, or other signs of financial strain.

Thank You