Lending Club Case Study

by:Maimana Kowatly
15th May 2024

Overview

What is Lending Club?

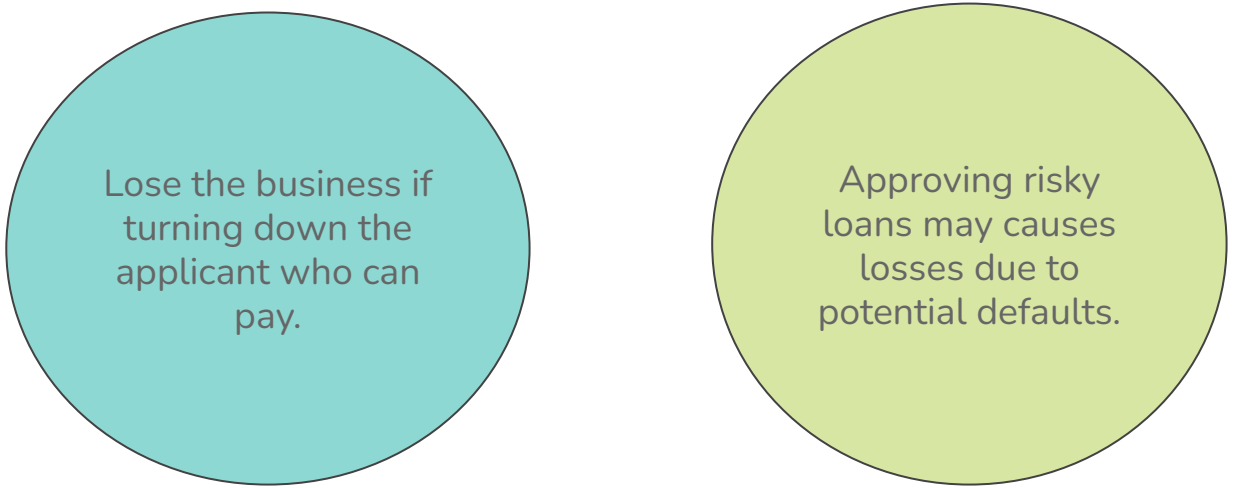
Lending club is peer-to-peer platform the has multiple services:

- operates an online marketplace.
- connecting borrowers who are seeking personal loans,business loans,or financing for various purposes with investors willing to fund those loan.

The platform facilitates borrowing and lending activities through its online interface, allowing individuals and businesses to access loans with relatively low interest rates compared to traditional financial institutions.

Business Objectives

When evaluating a loan application, as a consumer finance company the company faces two distinct risks:



Loans Types

Accepted Loans

Company accept the loan for the applicant

Rejected Loans

Applicant didn’t meet company’s requirements

Accepted Loans

Fully Paid

Payment in progress

Charged-off

Case Study

Project Aim

Find data patterns predicting default to apply ries mitigation strategies. The defaulters are the applicants that is late on their payments due dates.

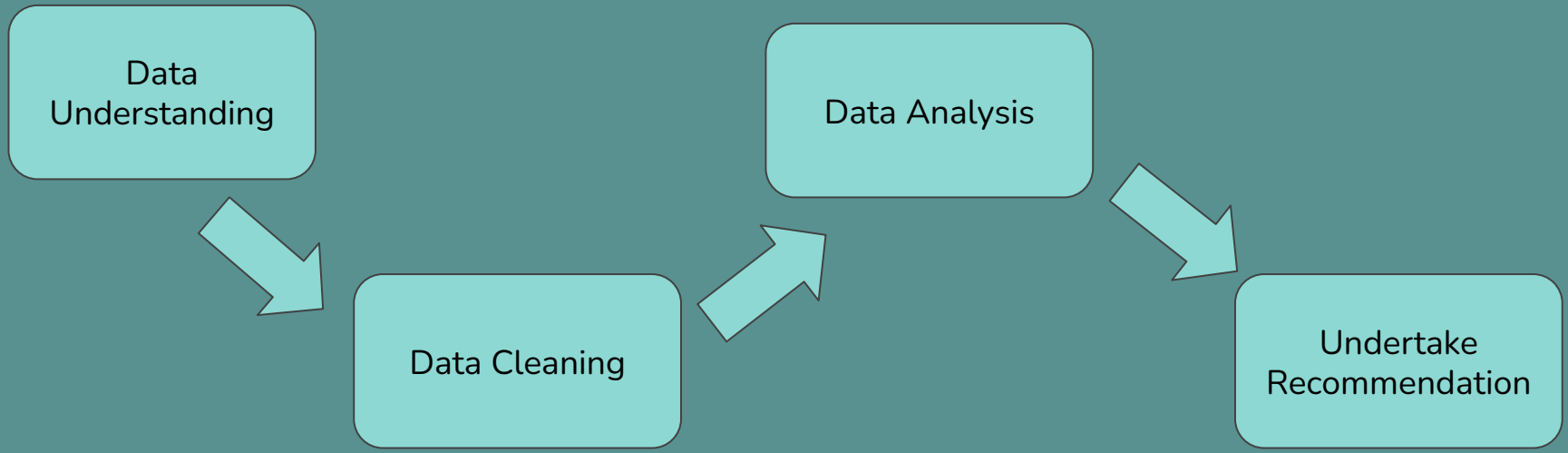
Objectives

Showing how the historical dataset will give the strong indicators and factors that will help the company make it decision.

Techniques

Utilize EDA principles for risk analysis in finance to minimize client risk.

Main Stages of the project:

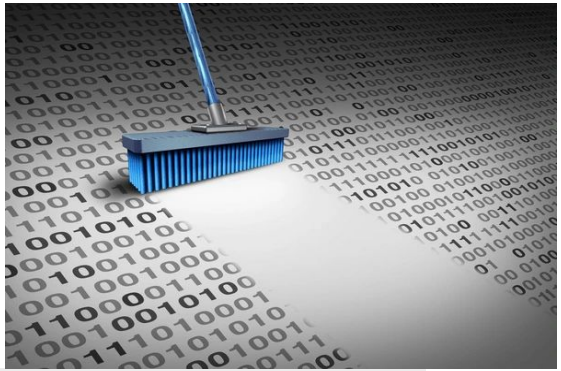


Data understanding

Get to know more about data by using python libraries

- Check dataframes (pandas) [initially dataframe is (39717, 111)]
- Check the number of unique variable in each column

Data Cleaning



Steps done to clean lending club dataset for a better analytics results:

1) Get Number of NA in all the columns
2) Create individual data frames based on loan status
3) Drop all column where all values are missing

   * dataframe after step #3(`14526, 57`) which mean there is significant impact of reducing non important data.

4) Drop all columns where number of unique values is only one.

   *dataframe = `(14526, 48)`

5) Since the id ,member_id is unique values so its data will be irrelevant.
6) Replace all values in column with String concatenation
7) (i.e. 10 years to 10 , 15 months to 15, 10% to 10)

Data Cleaning

8)split the date column to 2 columns (Month,Year)

9)replace data <1 with 0

10) drop all data with loan status = ‘current’ (very important step)

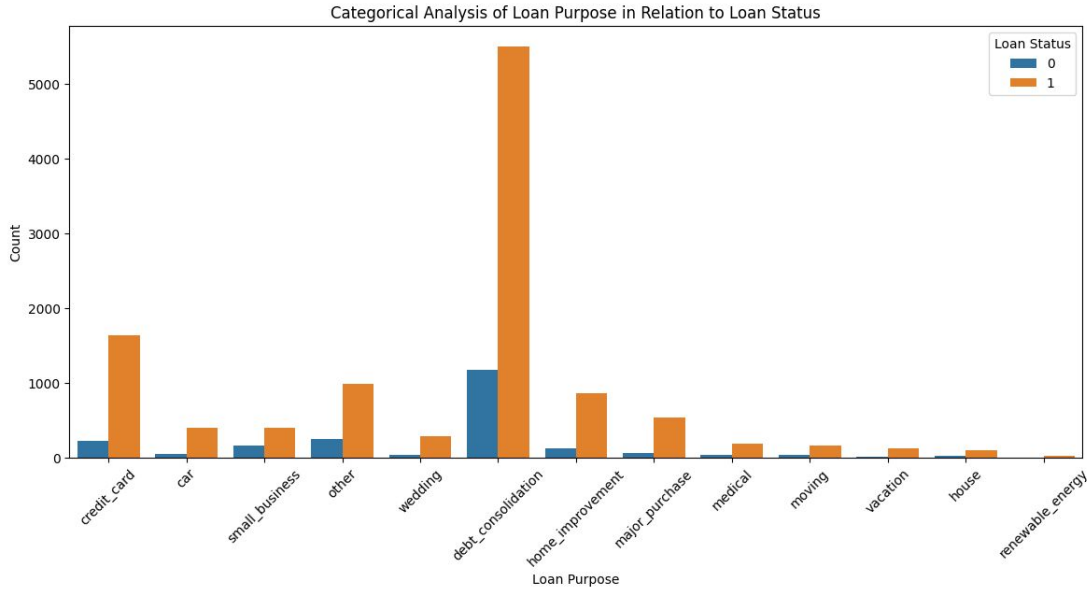
11) convert fully paid to **1**, convert charged-off to **0.**

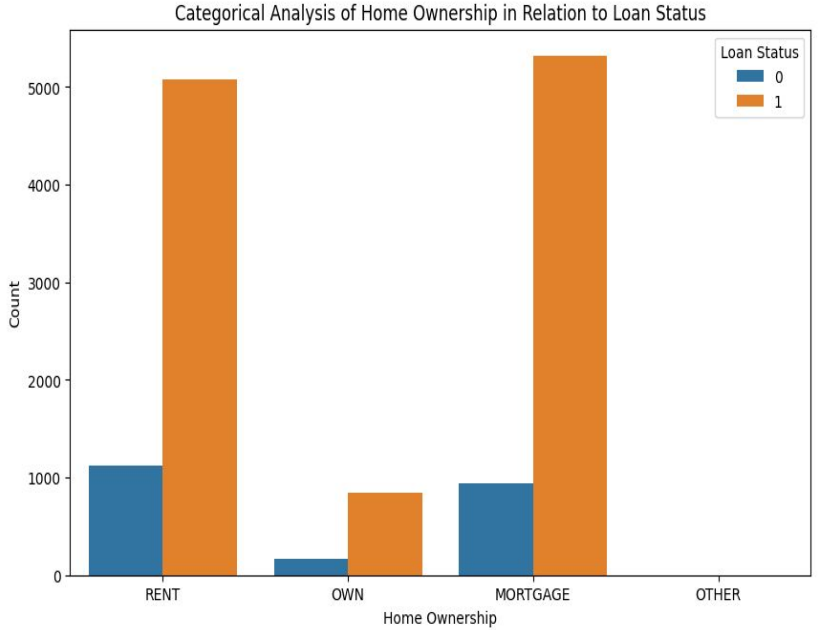
Data Analysis

Now that the data cleaned and prepared for analysis:

HomeOwnership vs loan purposes

Here is a different approach to visualize the categorical analysis of home ownership and loan purpose in the relation to loan status **using bar plots.**





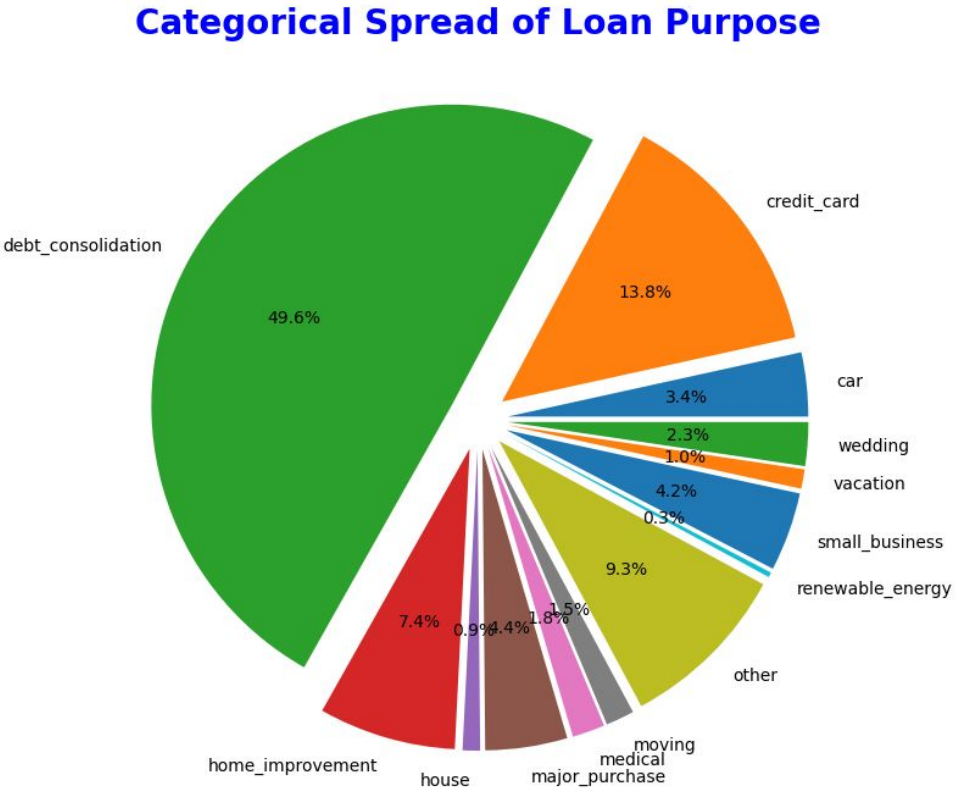
Data Analysis

- Debt_consilidation is the trend

  For the loan purpose

- Mortgage is the trend for the

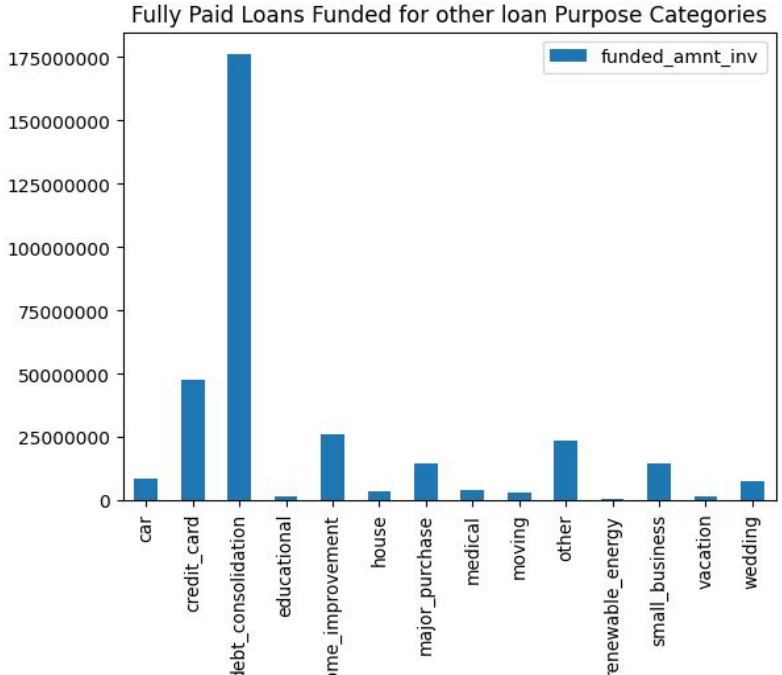
  For home ownership

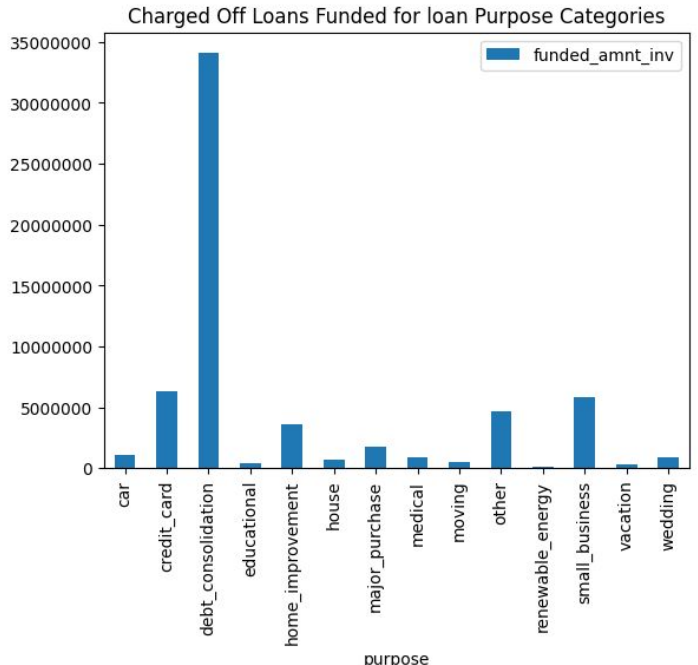


Data Analysis

Funded amount compared with loan purpose category:

Charged-off loans-status (sub dataframe) Vs Fully-paid loan-status (sub dataframe) by purpose:

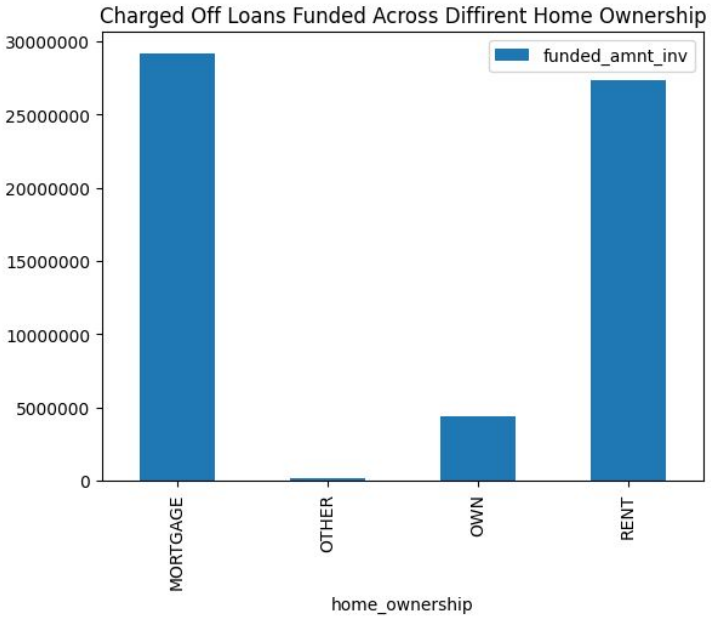


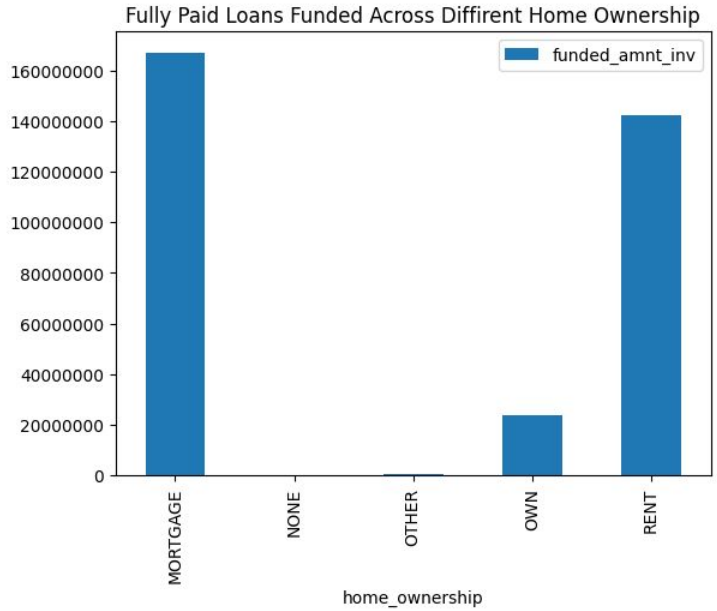


Data Analysis

Funded amount compared with applicants home ownership category:

Charged-off loans-status (sub dataframe) Vs Fully-paid loan-status (sub dataframe) by purpose:

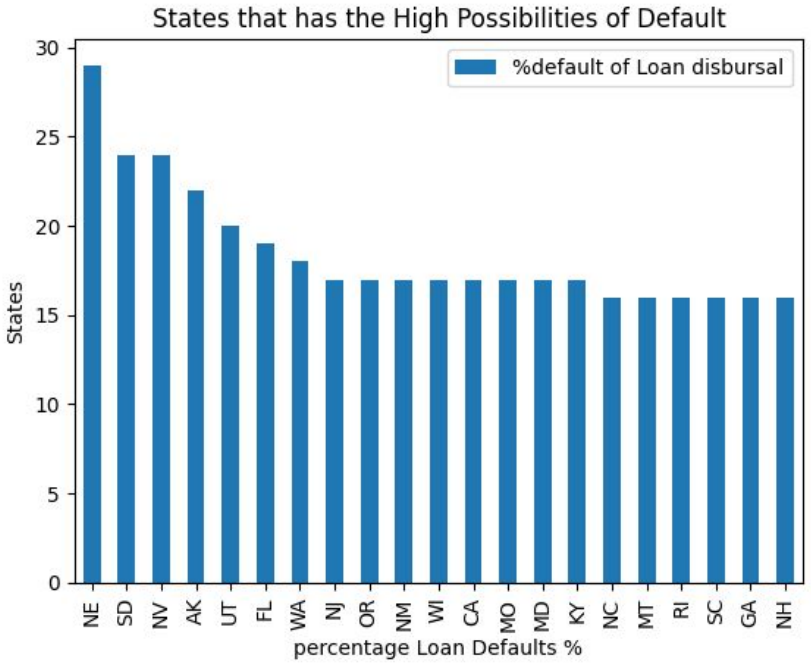




Data Analysis

States

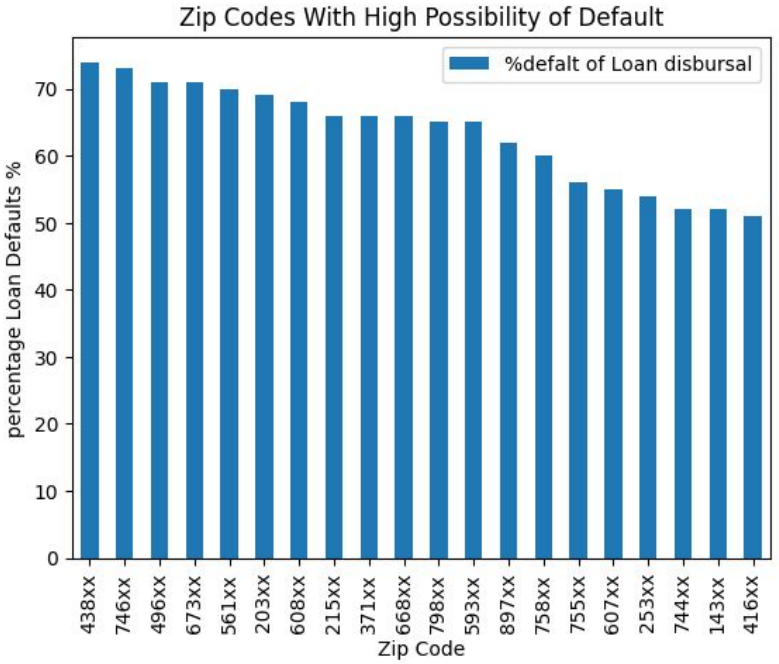
perform an analysis of the impact of state on the percentage of total loan paid.



Data Analysis

Zip codes

perform an analysis of the impact of zip codes on the percentage of total loan paid.



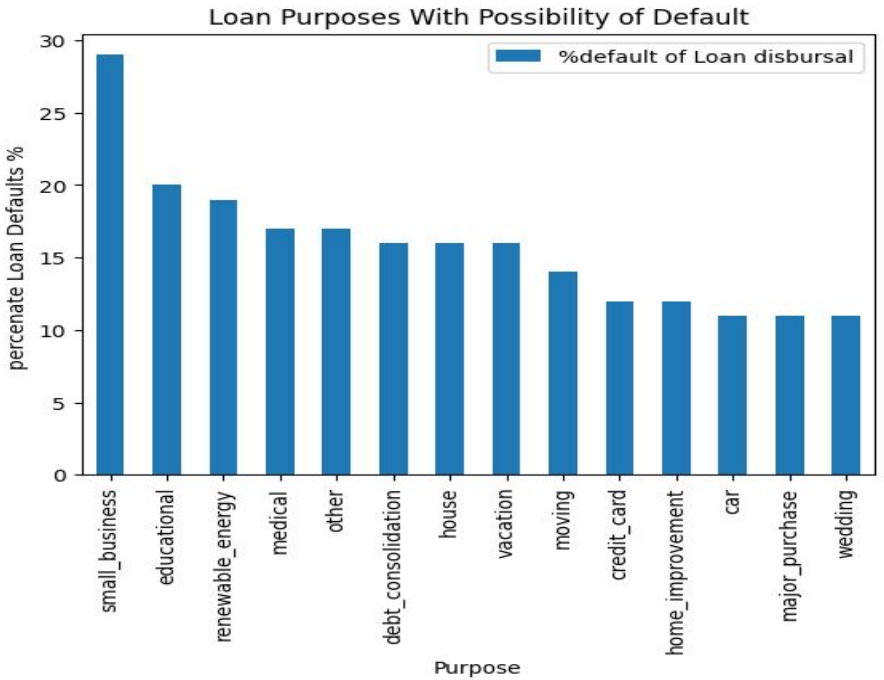
Data Analysis

Analyse impact of purpose on percentage of total loan paid:

The primary goal of this comparison is to analyze and visualize the relationship between loan purposes and the likelihood of loan defaults. By creating a pivot table and calculating the relevant percentages, the analysis identifies which loan purposes are associated with higher default rates. This information can be valuable for lenders and financial analysts to assess risk and make informed decisions regarding loan approval criteria and risk management strategies.

Data Analysis

Loan disbursal refers to the act of releasing the loan funds to the borrower.
When a loan is disbursed, the lender transfers the agreed-upon loan amount to the borrower's account or otherwise provides the funds to the borrower.



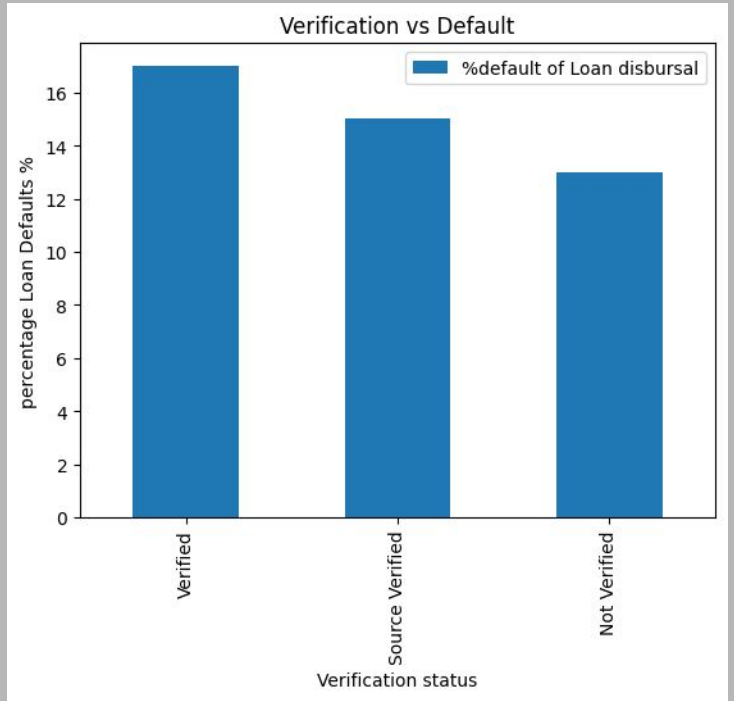
Data Analysis

Analyse impact of Verification Status on percentage of total loan paid:

There is 3 status of verification:

- Source verified
- Verified
- Not verified

No significant variation observed in default against different verification categories



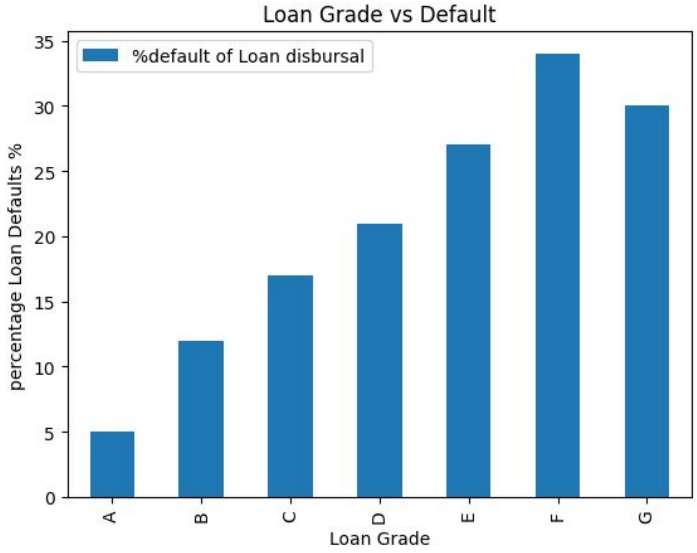
Data Analysis

Analyse impact of Grade on percentage of total loan paid:

Grade varies from **A** to **G**:

Grade **F** is the highest impact on the loan default

There is a direct relationship between lower loan grades and higher default rates.



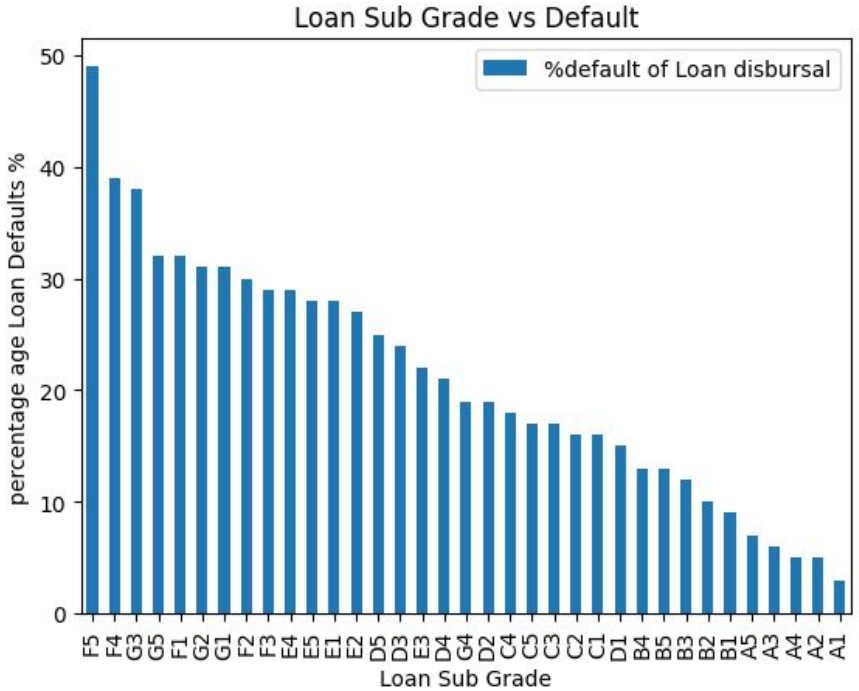
Data Analysis

Analyse impact of Sub-Grade on percentage of total loan paid:

Grade varies from **A1** to **G5**:

Grade **F5** is the highest impact on the loan default

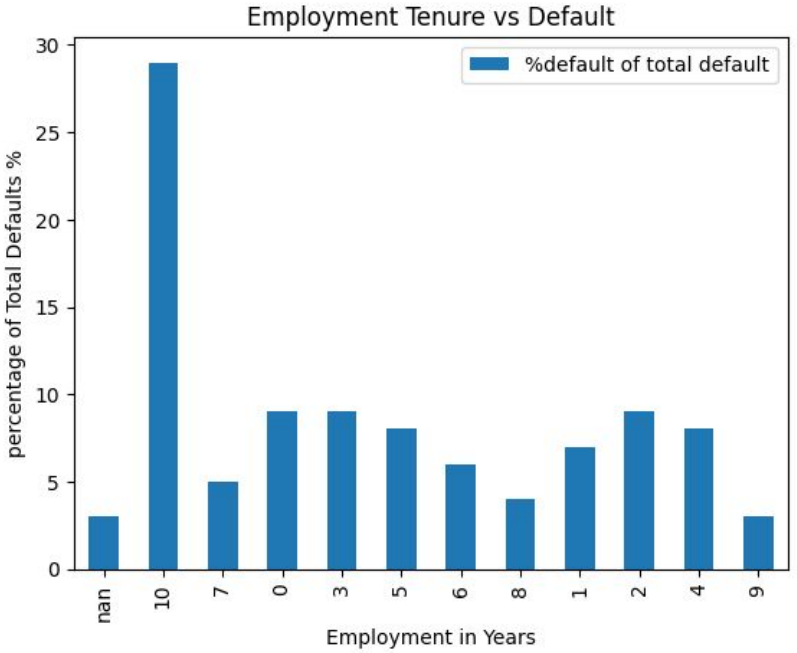
There is a direct relationship between lower loan sub-grades and higher default rates.



Data analysis

Analyse impact of Employee Tenure on percentage of total loan paid:

It is showing that applicants having employment services longer than 10 years are likely to default more



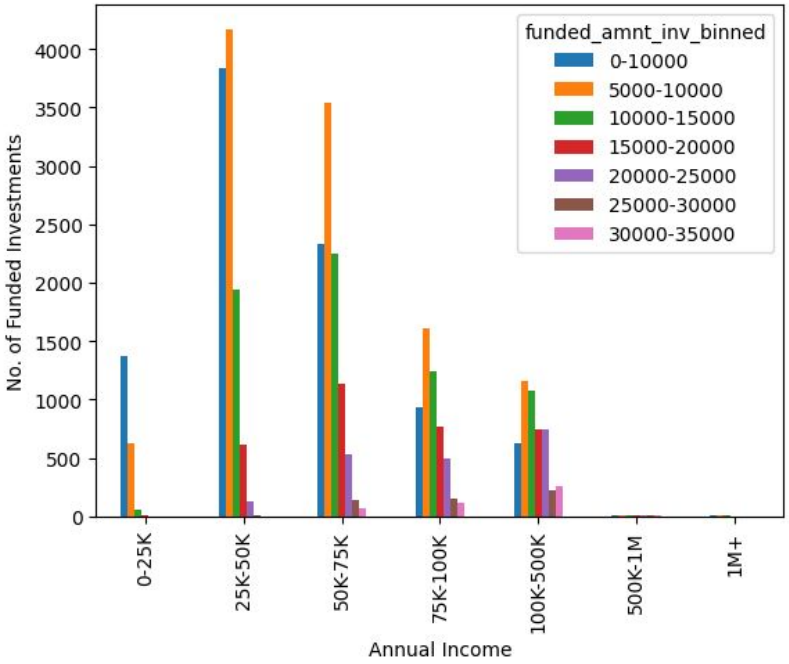
Undertake Binning

When to Undertake Binning

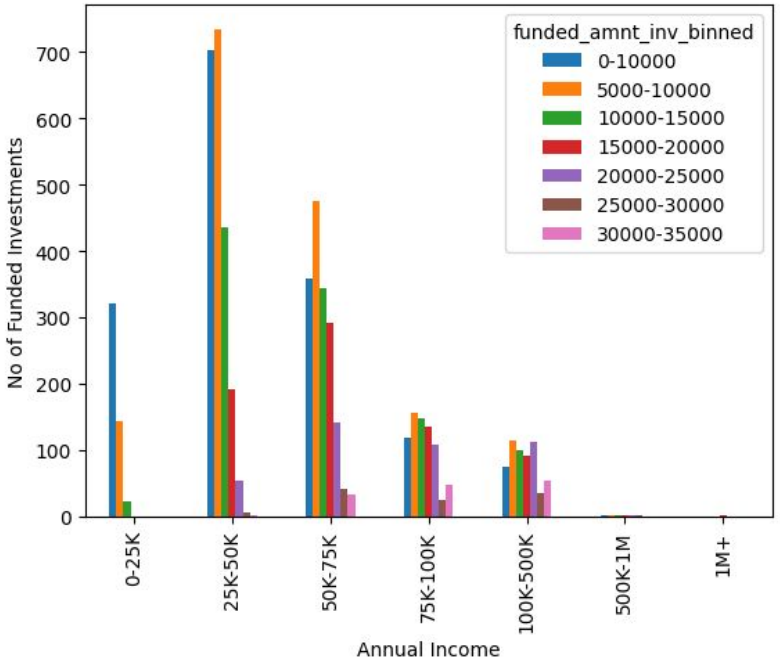
- **Handling Diversity:** When a continuous variable has a wide range of values or too much variability.
- **Simplifying Models**: To reduce the complexity of the data and make models more interpretable.
- **Improving Performance**: In some machine learning algorithms, discrete features might perform better than continuous ones.
- **Visualization:** To create histograms or bar charts for better visual interpretation of data distribution.

Undertake Analysis of binned quantities

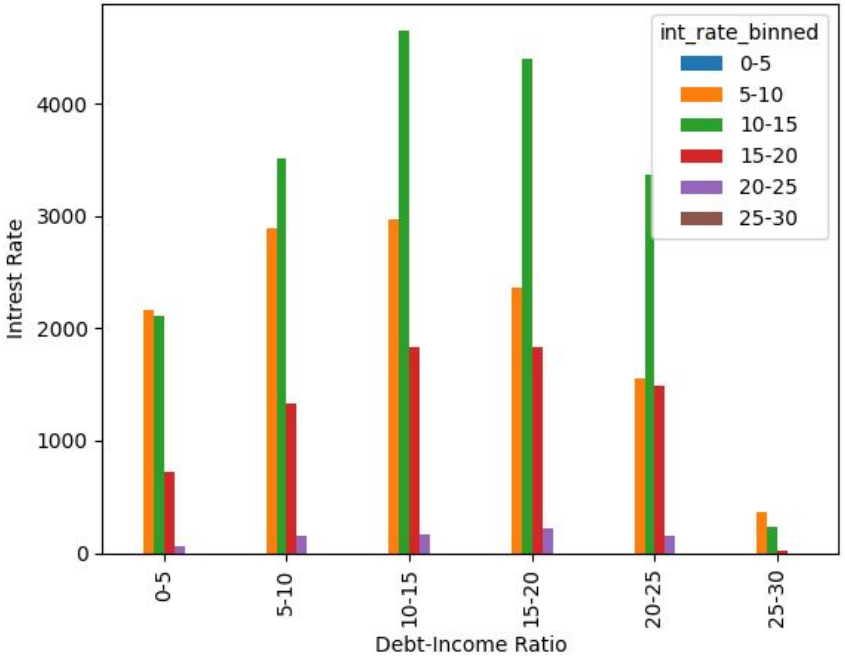
Fully-Paid loan-status

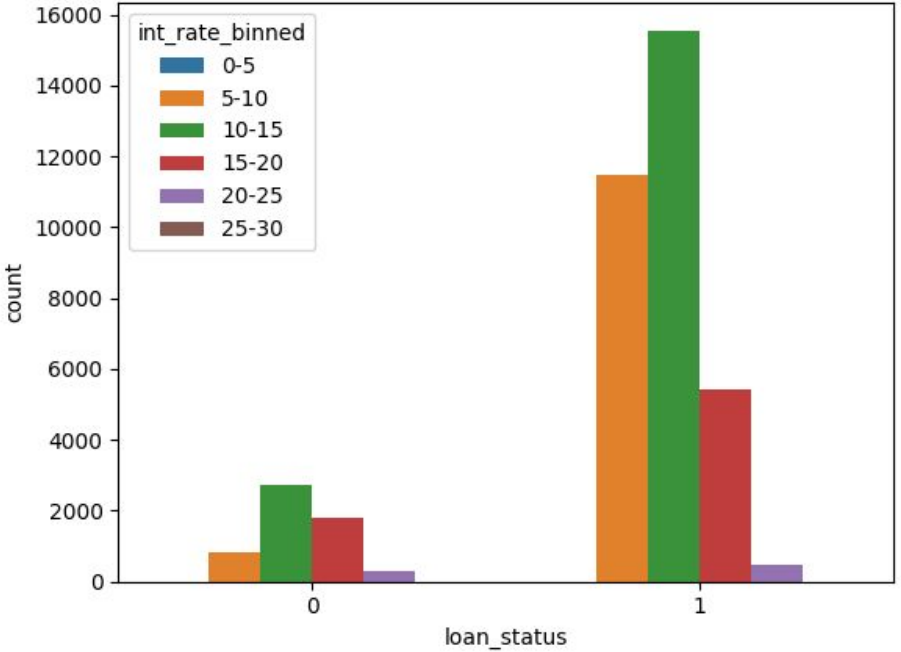


Charged-off loan-status

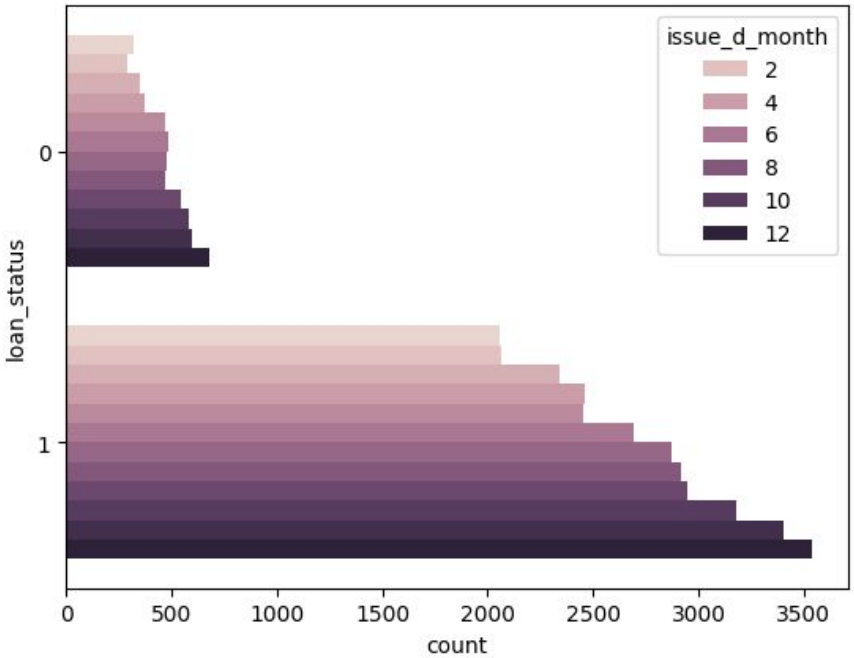


Undertake Assessment of Relationship between Debt-income ratio and interest Rate

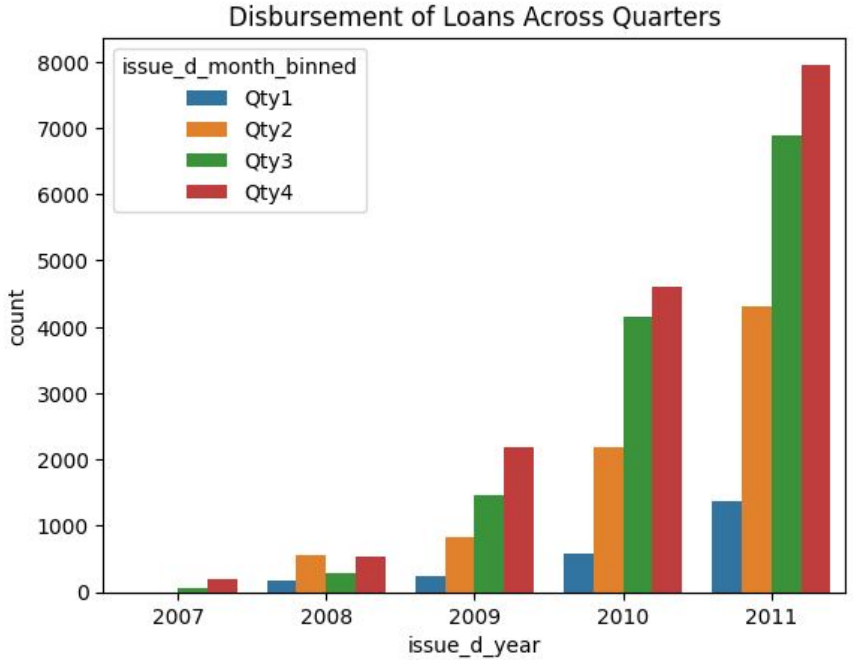




Distribution of Loans Across Quarters

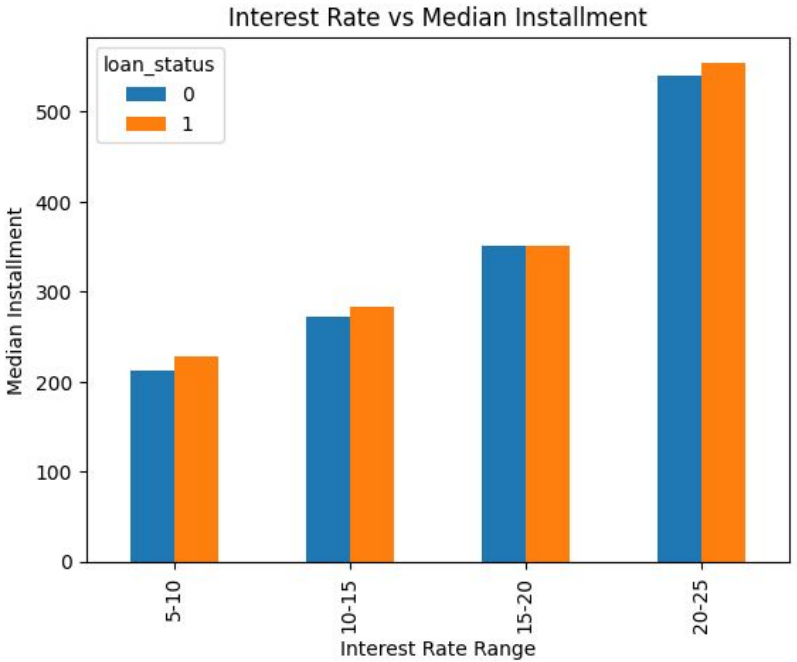


Month View

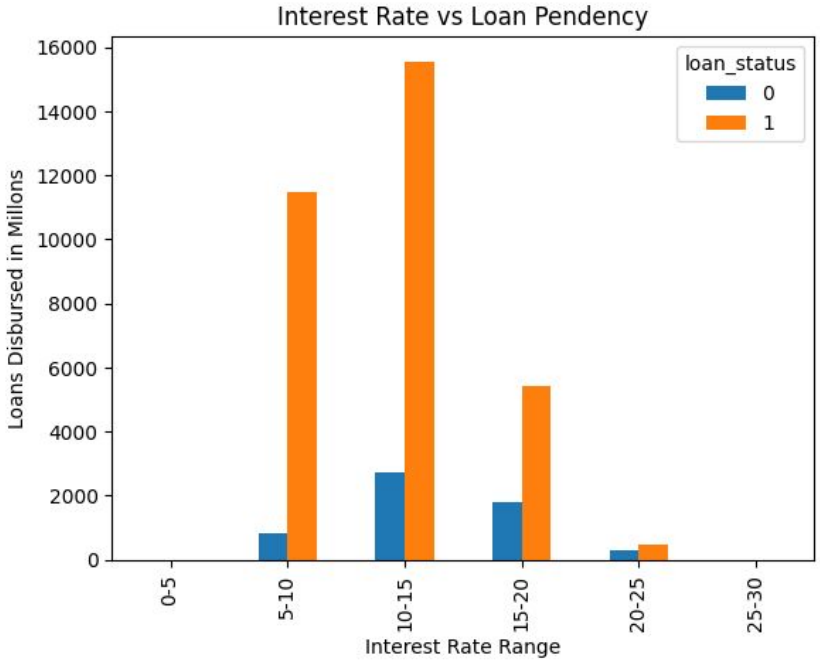


Years Quarters View

Interest Rate Comparison

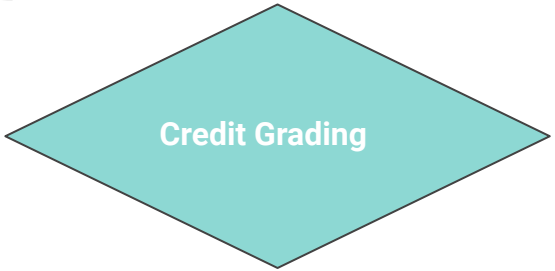


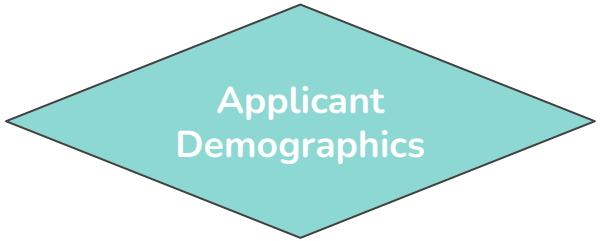
Investors might consider establishing an additional interest bracket below 20%, as there is a notable increase in defaults beyond this threshold.



Lending Club might contemplate offering adjustable interest rates if a loan shows signs of potential default.

Recommendation & Findings





Current systems for loan grading and customer verification appear to be adequate.

Customer demographics serve as a reliable gauge of a customer's repayment capacity.