Lending Club Case Study

Presented By: Shivam & Dipesh

Problem Statement

- Understand the driving factors behind loan defaults.
- Utilizing the above derived data-points and suggest methods on how the company can prevent giving loan to likely defaulter client.

Assumptions:

- It has been assumed that most loans are defaulted as the clients who were sanctioned loans were not capable of re-paying the loan.
- It has been assumed that the company did not properly verify the client's background information before deciding to sanction the loan.
- It has been assumed that the company did not segregate maximum loan limit based on client's annual income.
- It's highly likely that people are taking a longer tenure loan due to their low income and then default repaying it.

Analysis Approach

➢ Understanding the data:

- The “loan.csv” file provided consists of the loan details of clients provided by the company.
- The data consists of brief details related to loan (amount of loan approved, interest rate, loan term, verification status and the on-going status of the loan).
- The data consists of client details (Employment Title, Years of Experience, Purpose of the loan and state of the client).
- The data consists of client’s historic record data (Public records, bankruptcy records, open credit lines, last open credit line, etc.)
- Every loan is identified with an unique ID.
- Similarly, every member is identified with an unique Member ID.

Analysis Approach

➢ Data Cleaning:

- There are many columns in the data-set which have value as “NA” only or are not necessary for the analysis.
- We have deleted the columns from AX except column DB as it gives insights into client’s bankruptcy records.
- We have identified the missing data values as in column K (emp_title) and replaced the missing data with NA.
- There are certain rows which are out of format which have been removed by filtering the column Q (loan_status).
- We have removed the column R-S (pymnt_plan & url) respectively as they are not relevant for this analysis.

Analysis Approach

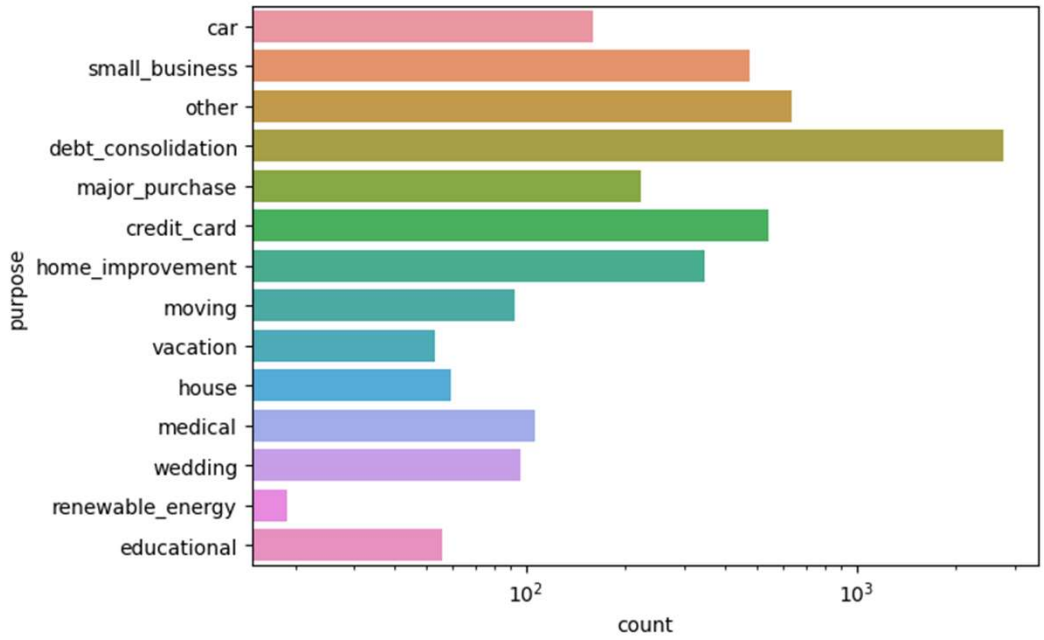
➢ Data Manipulation:

- All numeric columns have been converted into “int” data-type for the analysis.
- All decimal columns have been standardized to 2 places after the decimal point.
- The annual income has been segmented into 5 categories:

1. Upto 25k
2. Upto 50k
3. Upto 75k
4. Upto 1Lac
5. Above 1Lac

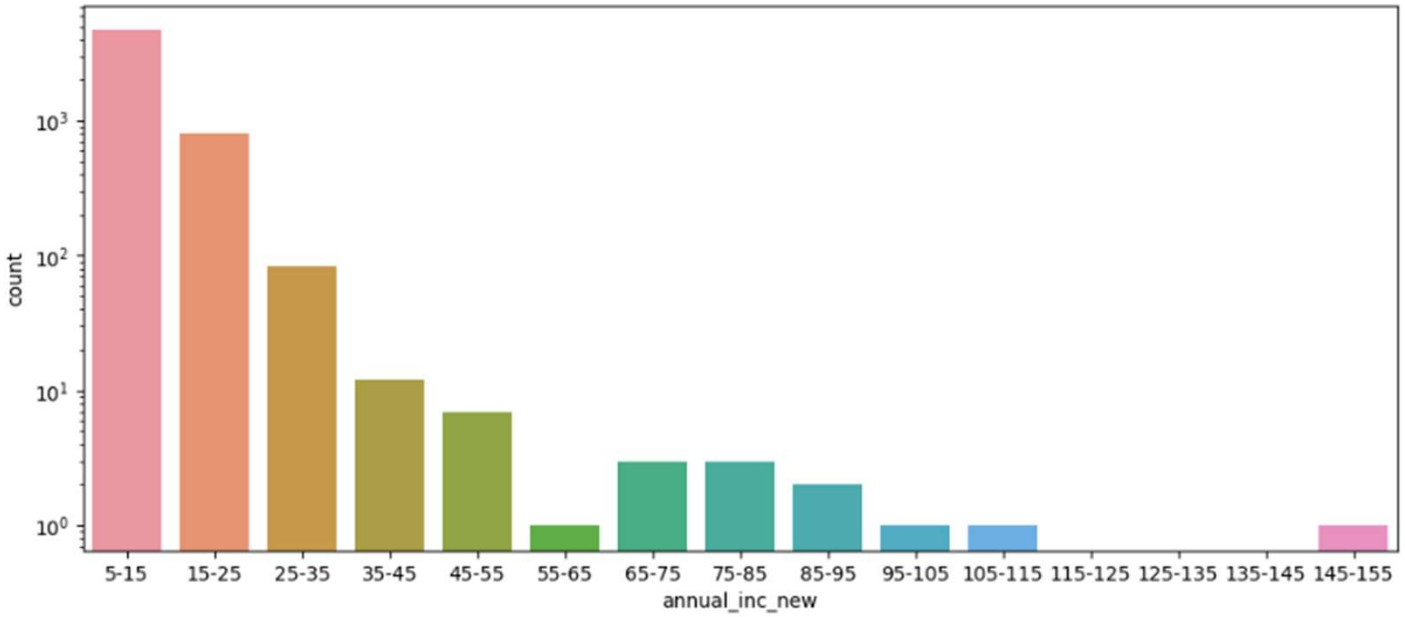
Univariate Analysis

1. Purpose Column: Analyzing this gives us a valuable insight into the data which we can keep in the back of our mind during bivariate analysis. This will help us to draw meaningful conclusions later in the analysis.



Univariate Analysis

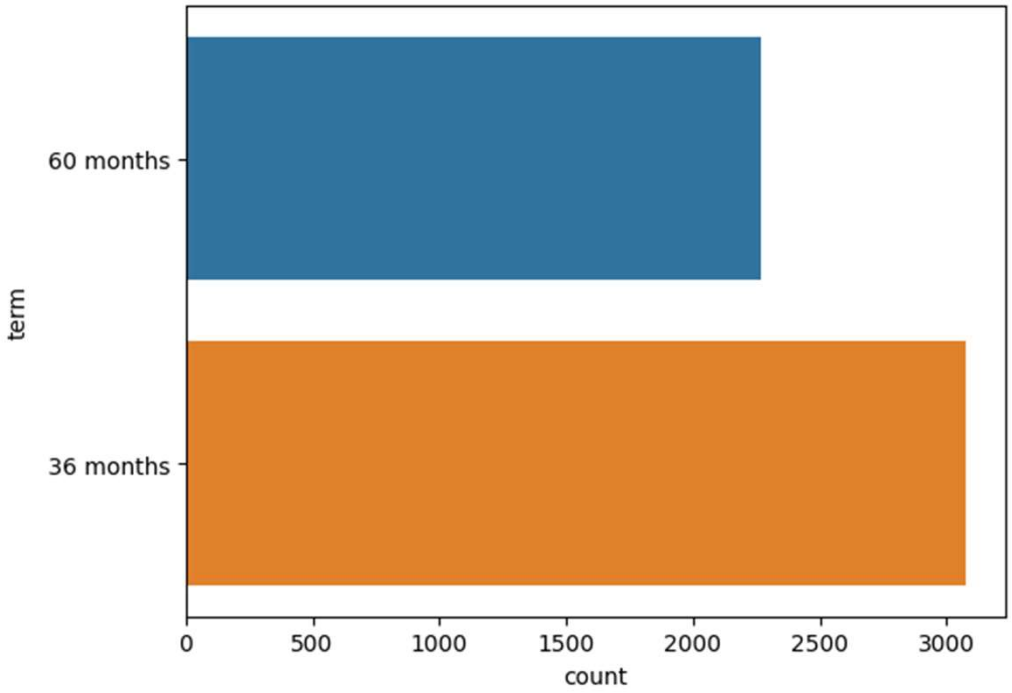
2. **Annual Income Column:** Analyzing this tells us that lower income groups are defaulting the most. Therefore, lower income groups should be charged higher interest rates to minimize credit loss.



Univariate Analysis

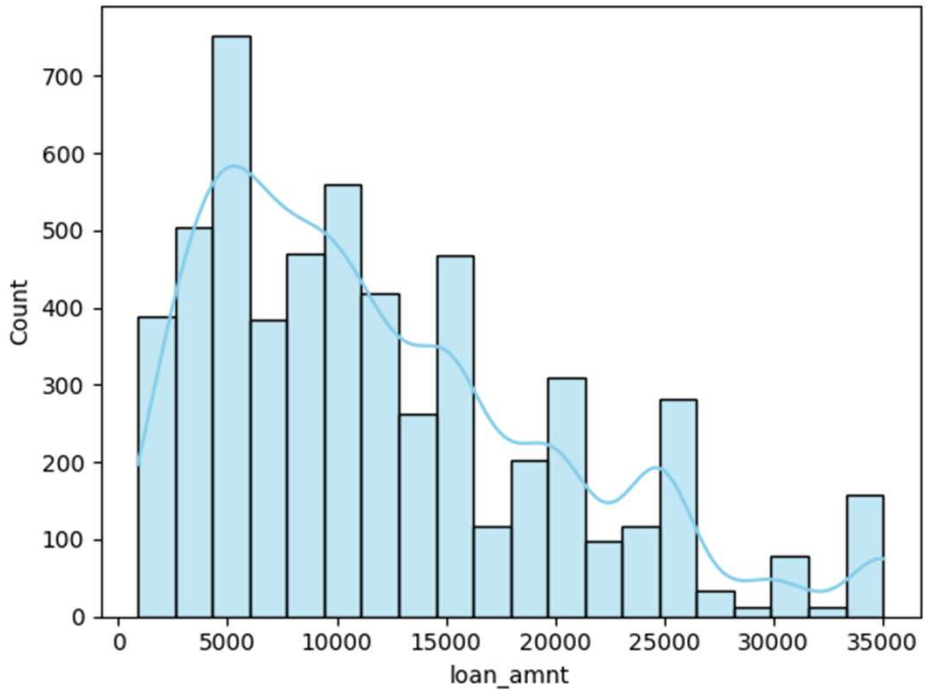
3. **Loan Tenure Column:** This tells us that loans taken for longer duration are more likely to default.

- **Hypothesis:** It's highly likely that people are taking a longer tenure loan due to their low income and then default repaying it.



Univariate Analysis

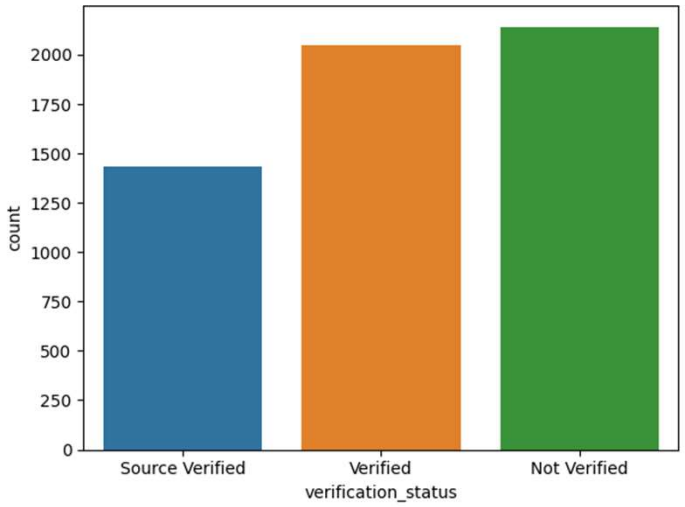
4. **Loan Amount Column:** Studying the spread of data to understand range of loan amount



Univariate Analysis

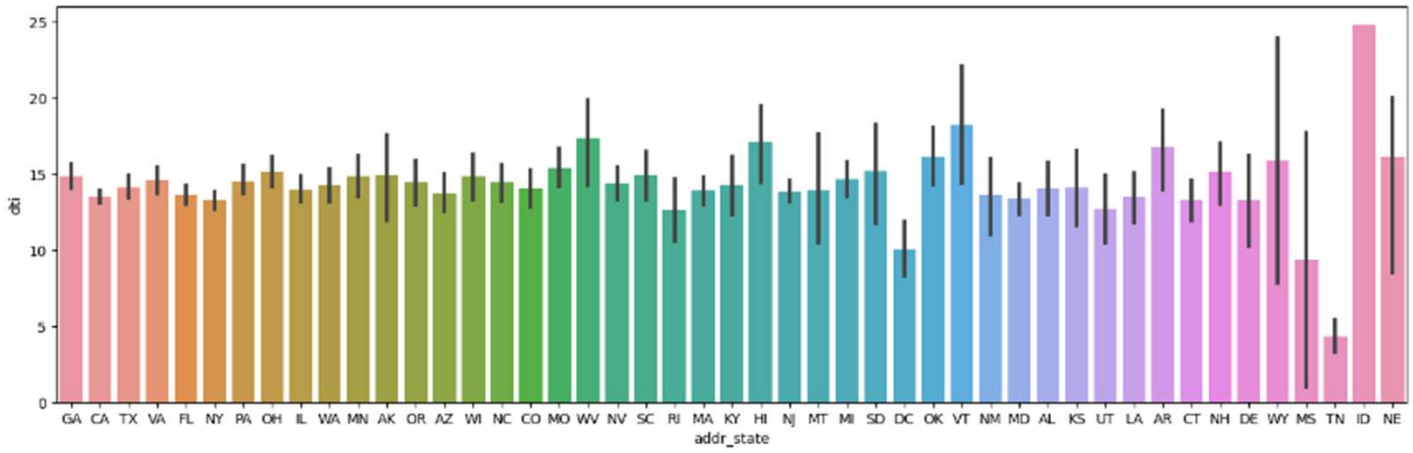
5. **Verification Status Column:** This gives us an insight that most loan defaulters were actually not verified. Therefore, from now onwards, the lending club should make more strict policies and verification processes to ensure that only verified groups can get a loan sanctioned.

- Another possiblity for the large no. of unverified loan takers could be that loan applicants are bribing employees of the lending club to sanction the loan. The club should investigate this possiblity and implement more strict compliance policies if needed.



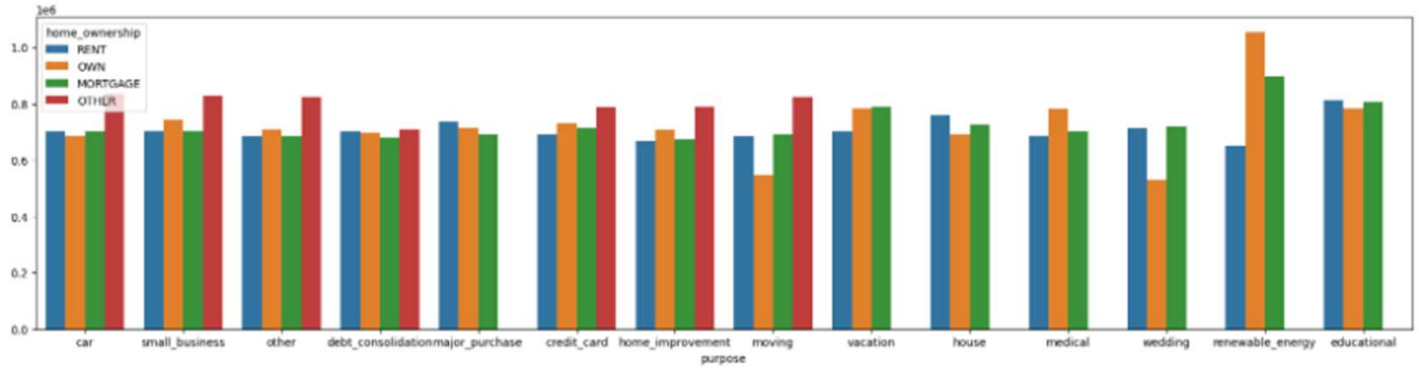
Bivariate Analysis

1. **Comparing dti based on address state:** This tells us that people from 'ID' spend above their means and are more likely to default.



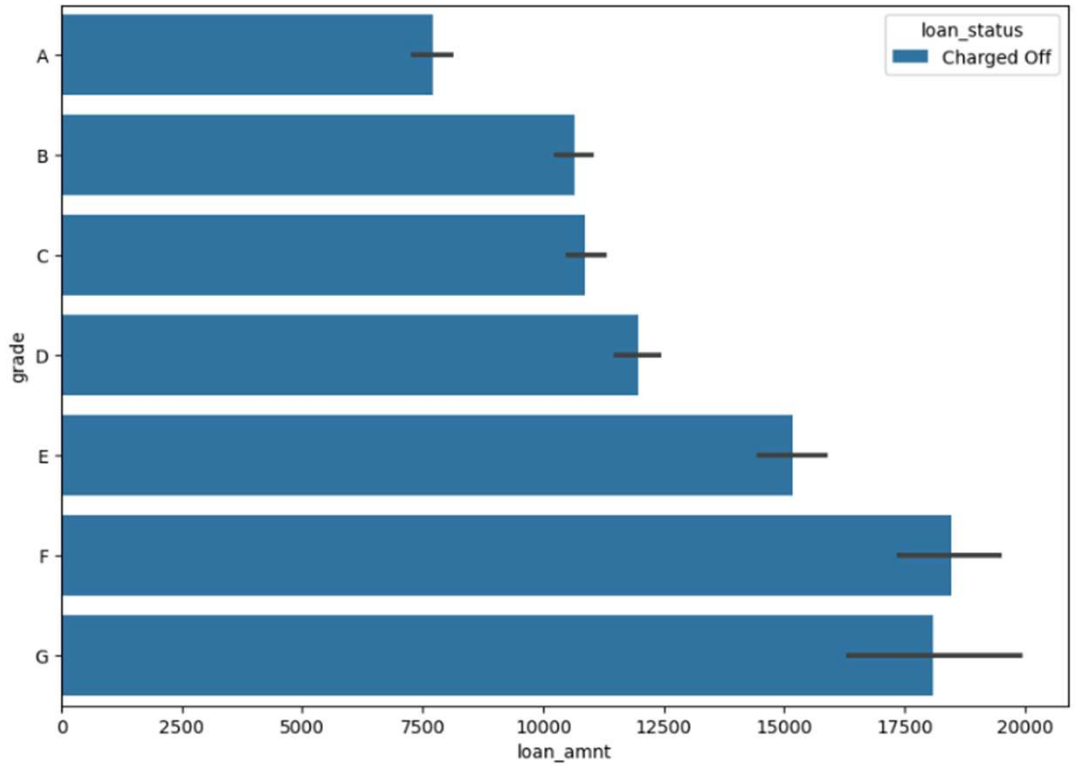
Bivariate Analysis

2. **Comparing home ownership with purpose**



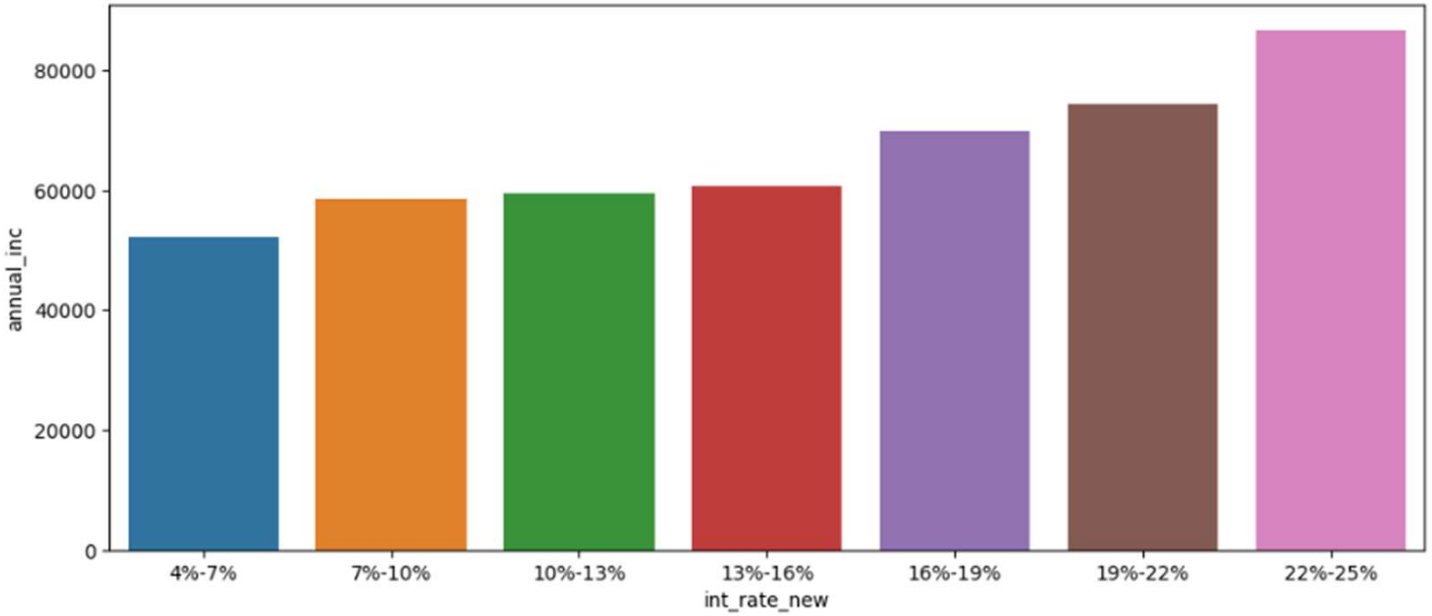
Bivariate Analysis

3. **Comparing Loan Amount with Grade**



Bivariate Analysis

4. **Comparing interest rate & annual income based on grade**



Observations

Below are some observations & actionable insights based on the analysis:

1. Most loan defaulters have an income of less than 55,000. Therefore, more interest should be charged for higher income groups.
2. Borrowers from the 'ID' state spend above their means and are more likely to default. There must be stricter criteria to grant them the loan.
3. Most loans are defaulted by low-income groups. However, interest rates charged to them are the lowest. This is causing more crdit loss.
4. Loans taken for Debt Consolidation are one of the highest.
5. 'F' grade loans have the highest amount of loan defaulters. Set lower loan limits for 'F' grade loans. Restricting the loan amount reduces the potential loss in case of default.

THANK YOU!!