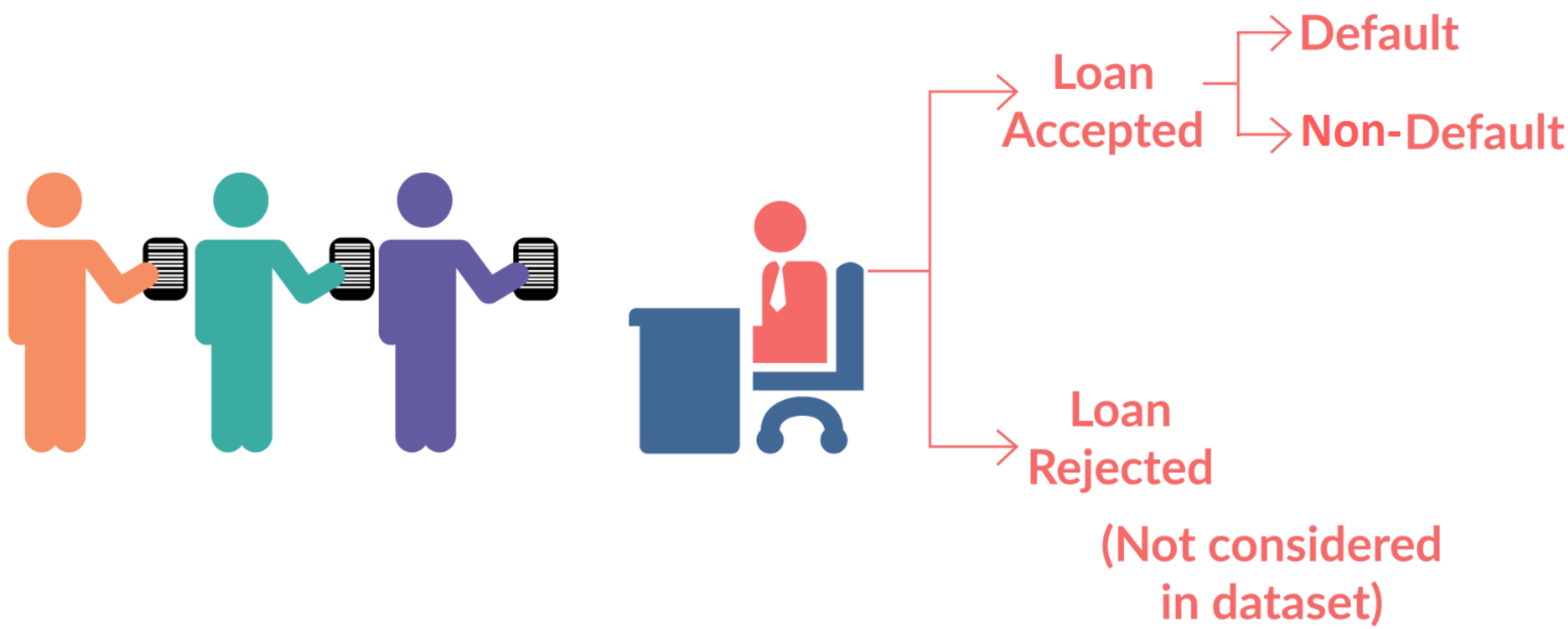
Lending Club Case Study

Assignment Submitted By: Nidhi Pandey

Problem Statement and Objective

The consumer finance company faces two critical risks in the loan approval process: potentially missing out on profitable business by rejecting creditworthy applicants and incurring financial losses by approving loans for applicants likely to default.

The objective of this case study is to analyze data from past loan applicants to identify patterns and factors that indicate the likelihood of loan defaults.



Loan Insights: An Exploratory Data Analysis"

We will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Accepted

- Fully Paid - Loan is completely repaid.
- Current - Applicant is still making payments.
- Charged-off - Applicant has defaulted on the loan.

Rejected: No transactional history is available for the applicant.

Data Preparation

Data Loading:

The dataset, comprising 39717 rows and initially 111 columns, was loaded from a CSV file.

Missing Values:

A comprehensive check for null values revealed several columns with excessive missing data, leading to a reduction in the dataset to 47 columns after cleaning.

Data Cleansing:

Columns with single values and those not relevant to the analysis (e.g., post-loan processing details) were removed.

Null values in critical columns were addressed by filling with the most frequent or appropriate values.

Univariate analysis for Numerical and Categorical Analysis

Numerical Analysis:

Boxplots revealed outliers in annual income. An IQR method was employed to identify and remove these outliers, leading to a clearer distribution.

Categorical Analysis:

Count plots illustrated trends in loan grades, home ownership, verification status, and loan status. Key findings include:

Grade B had the highest loan issuance.

Renters comprised the majority of loan recipients.

A significant portion of loans were not verified.

Bi-Variate Analysis

Annual Income vs. Loan Status: Analysis showed borrowers with lower incomes were more likely to default.

Term and Loan Amount: The analysis indicated that loans with a 60-month term had higher default rates compared to 36-month loans.

Employment Length: Loans were most frequently fully paid by those with over 10 years of employment as well as defaulted.

Interest Rate Analysis: Higher interest rates correlated with higher default rates, especially for lower-grade loans (D, E, F, G).

Pivot Table Insights

A pivot table was created to understand the relationship between various factors like DTI, interest rate, and annual income across different loan grades. Key observations included:

Correlation Analysis: The correlation matrix showed that DTI and revolving utilization have significant impacts on interest rates and are influenced by the number of inquiries and delinquencies.

- **Strong Positive Correlations**:
  - int_rate and revol_util: Borrowers with higher revolving utilization tend to have higher interest rates.
  - int_rate and annual_inc: Borrowers with higher annual incomes might be offered lower interest rates.
  - loan_amnt and annual_inc: Borrowers with higher incomes tend to borrow larger amounts.
- **Strong Negative Correlations**:
  - dti and loan_amnt: Borrowers with higher DTIs might be less likely to borrow large amounts.
  - inq_last_6mths and loan_amnt: Borrowers with more recent inquiries might be less likely to be approved for large loans.

Segmented analysis on Charged Off loans and Grades (D,E,F,G)

The heatmap suggests that charged-off loans are often associated with:

- Higher interest rates
- Higher revolving utilization
- Delinquency history
- Lower income
- Recent inquiries
- Smaller loan amounts

It can be safer to assume that loans from Grade A and B are more stable compared to other grades.

Key Findings

Determinants of Default Risk:

- Annual Income: Lower incomes correlate with higher default rates.
- Loan Term: 60-month loans show higher default rates than 36-month loans.
- Employment Length: Longer employment tenure (10+ years) is associated with higher repayment rates.
- Loan Grade: Higher-grade loans (A, B) have lower default rates compared to lower-grade loans (D, E, F, G).

Economic Indicators:

- High Debt-to-Income (DTI) ratios and revolving credit utilization indicate greater risk.

Interest Rates:

- Charged-off loans tend to have higher median interest rates, reflecting increased repayment burdens.

Predictive Decision-Making:

- The company should refine loan approval criteria based on data-driven insights to better manage risk.

Enhanced Risk Management:

- Pivot tables and correlation analyses highlight interactions between various factors influencing default risk.

Conclusion

In conclusion, this analysis highlights critical factors influencing loan default risk, including annual income, loan term, employment length, and loan grade. By focusing on these key factors, along with borrower credit history, the company can better predict defaults and tailor loan offers to minimize risk.

Refining loan approval criteria and implementing advanced credit scoring models based on these insights will enhance risk management strategies.