



LENDING CLUB CASE STUDY

AVANI KASAT
GOWDA KIRAN GANGADHAR



PROBLEM STATEMENT

Identifying High-Risk Loan Applicants through Exploratory Data Analysis (EDA) for predicting whether loan will be default or not

Our client, the largest online loan marketplace, offers a diverse range of financial products, including personal loans, business loans, and medical procedure financing. While the fast and efficient online platform attracts borrowers seeking lower interest rates, the company faces a substantial challenge in managing credit loss, particularly from loans extended to 'risky' applicants.



BACKGROUND

Like many lending institutions, our client encounters financial losses, termed credit loss, when borrowers fail to repay their loans, leading to default.

In this context, borrowers labeled as 'charged-off' represent defaulters, contributing significantly to credit loss.

The objective of this case study is to employ Exploratory Data Analysis (EDA) to identify key indicators and driving factors associated with loan default among applicants.



SCOPE OF ANALYSIS

The analysis will delve into the company's historical loan data, exploring various types of variables that may serve as indicators of default risk.

These variables could range from demographic information and financial history to loan-specific details.

The focus is on uncovering patterns, trends, and correlations within the dataset that can inform predictive models for identifying potential defaulters



SIGNIFICANCE OF RISK ANALYTICS

Risk analytics plays a pivotal role in this context, as it involves understanding the types of variables that significantly contribute to the likelihood of loan default.

By identifying these key driver variables, the company can enhance its risk assessment procedures, thereby minimizing exposure to high-risk borrowers and mitigating credit loss.



UNIVARIATE ANALYSIS

Univariate analysis involves the examination of a single variable at a time. In the context of risk analytics for loan default prediction, univariate analysis helps us understand

The individual characteristics or attributes of applicants and their potential impact on loan default.

Inferences:

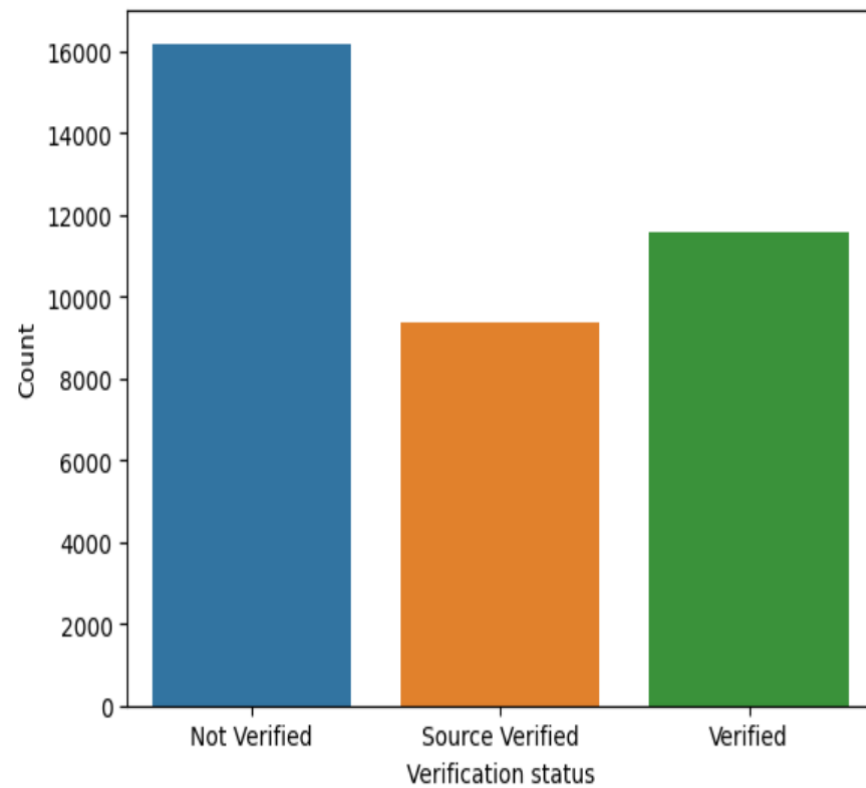
- Loan defaulters are less as compared to people who have fully paid the loan i.e. The number of defaulted loans is 7 times less than the number of fully paid loans.
- More than half of the loan taken has a term of 36 months compared to 60 months
- Most loans have A and B grades.
- The interest rate is more around 6-7.5 and 10-13.



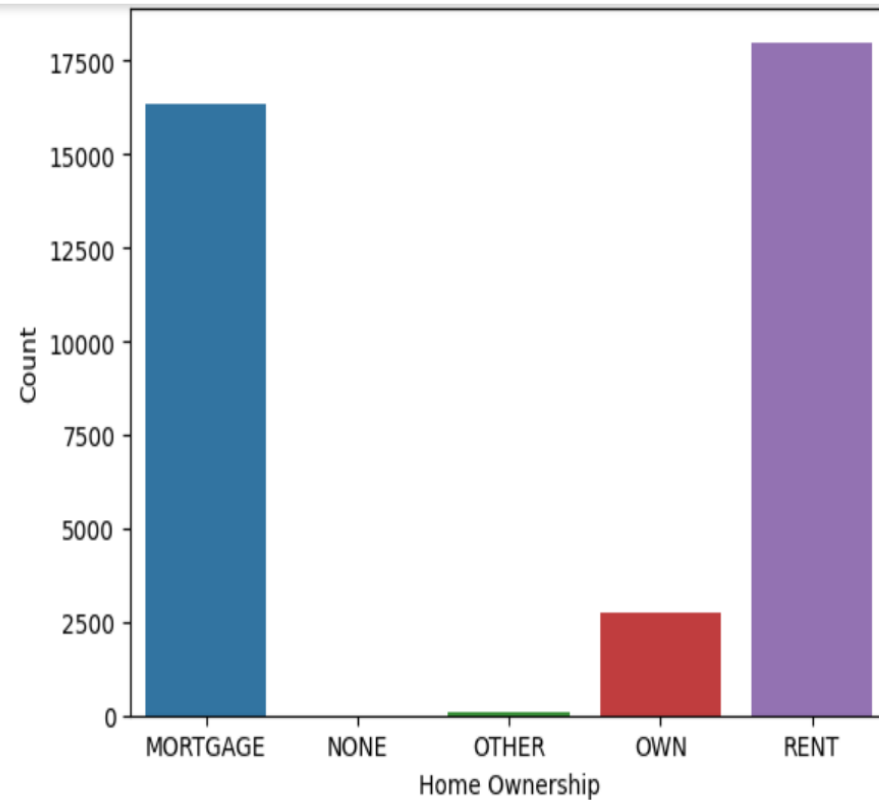
UNIVARIATE ANALYSIS

- Many loans have lower subgrades in grades 'A' and 'B' compared to the rest showing most loans are high-grade loans.
- Most of the people who have loans are having experience of 10 years
- Most people who have taken loans are either living on rent or a mortgage, they don't own their property.
- Only around 50% of people have verified their source by the company
- Many people who took loans have less annual income
- A large percentage of loans are taken for debt consolidation.
- Most of the people who have taken loans are from urban cities as compared to rural cities.
- Many people who have taken loans have high debt.
- Most of the people have 0 bankruptcies record.
- Most of the loans are issued in the last quarter of the year.
- the loan approval rate is increasing with time.

UNIVARIATE ANALYSIS: Some Visualizations from EDA



Distribution of verification status



Distribution of home ownership

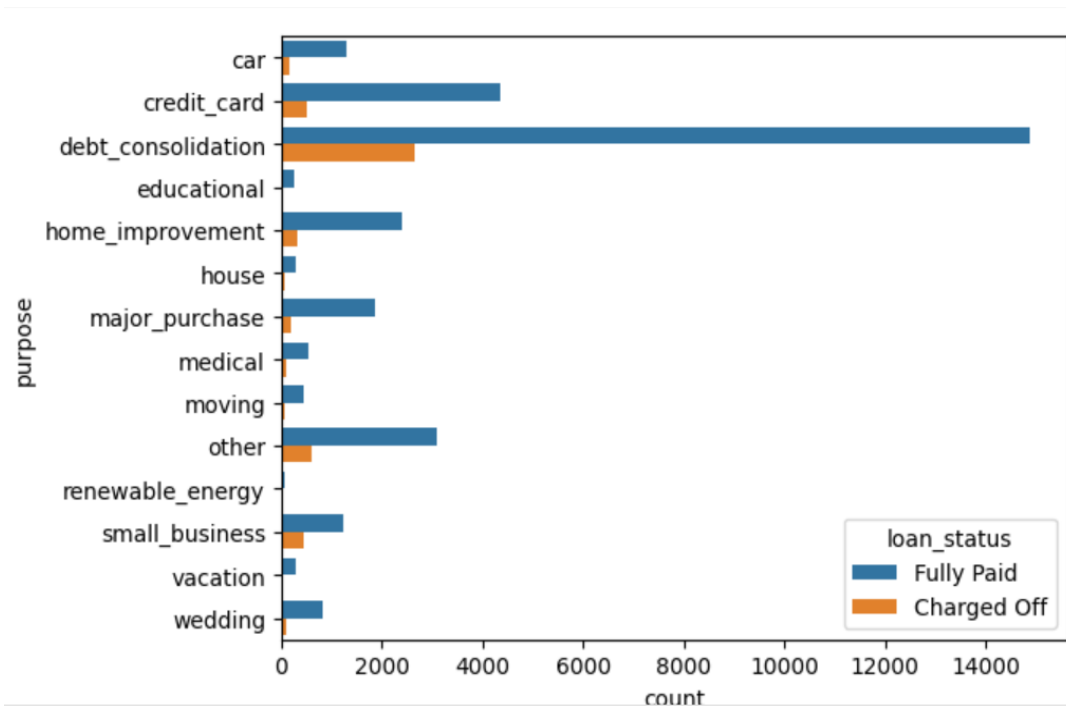


SEGMENTED UNIVARIATE ANALYSIS

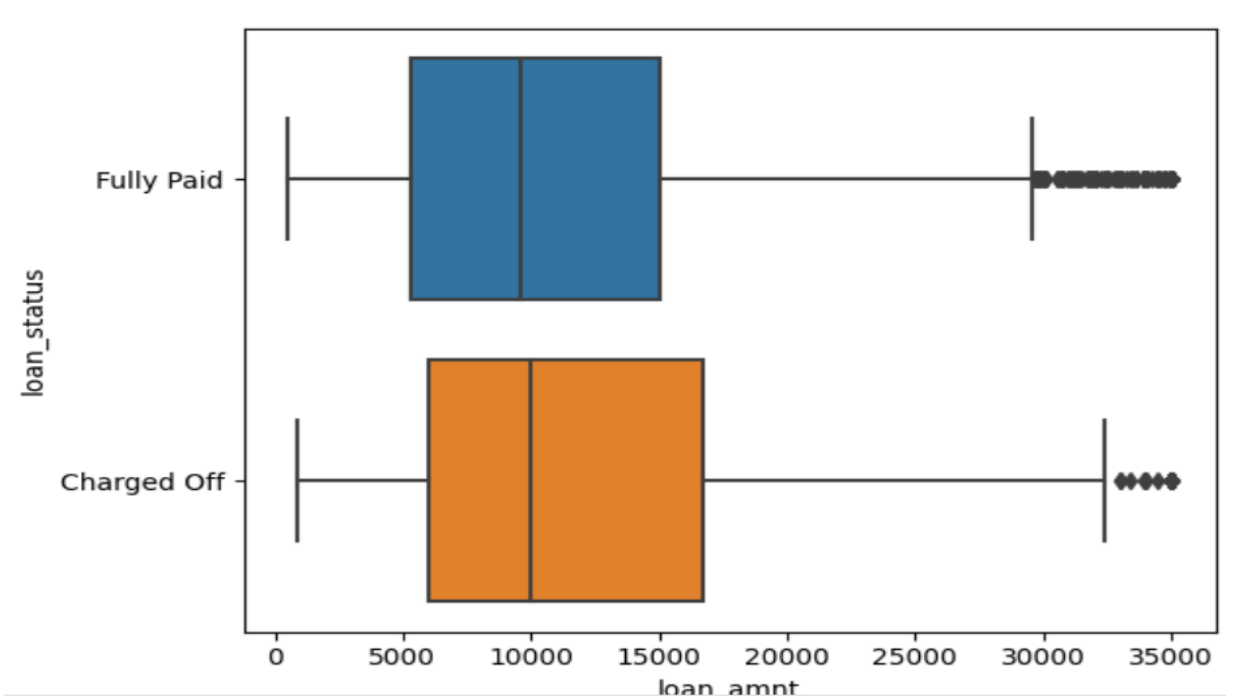
Inferences:

- Debt Consolidation is the most common reason for loan purposes and has the highest number of fully paid loans and defaulted loans.
- The mean and 25% percentile are the same for both but we see a larger 75% percentile in the defaulted loan which indicates a large amount of loan has a higher chance of defaulting.
- The loans in 36-month terms majorly consist of grade A and B loans whereas the loans in 60-month terms mostly consist of grade B, C, and D loans.
- The Defaulted loans are lower for the borrowers who own their property compared to those on mortgage or rent.
- Employees with 10 or more years of experience are likely to default and have a higher chance of fully paying the loan.
- The default loan amount increases with interest rate and shows a decline after a 17.5 % interest rate.

SEGMENTED UNIVARIATE ANALYSIS: Some Visualizations from EDA



Distribution of purpose based on loan_status



Distribution of loan_amnt based on loan_status



BIVARIATE ANALYSIS

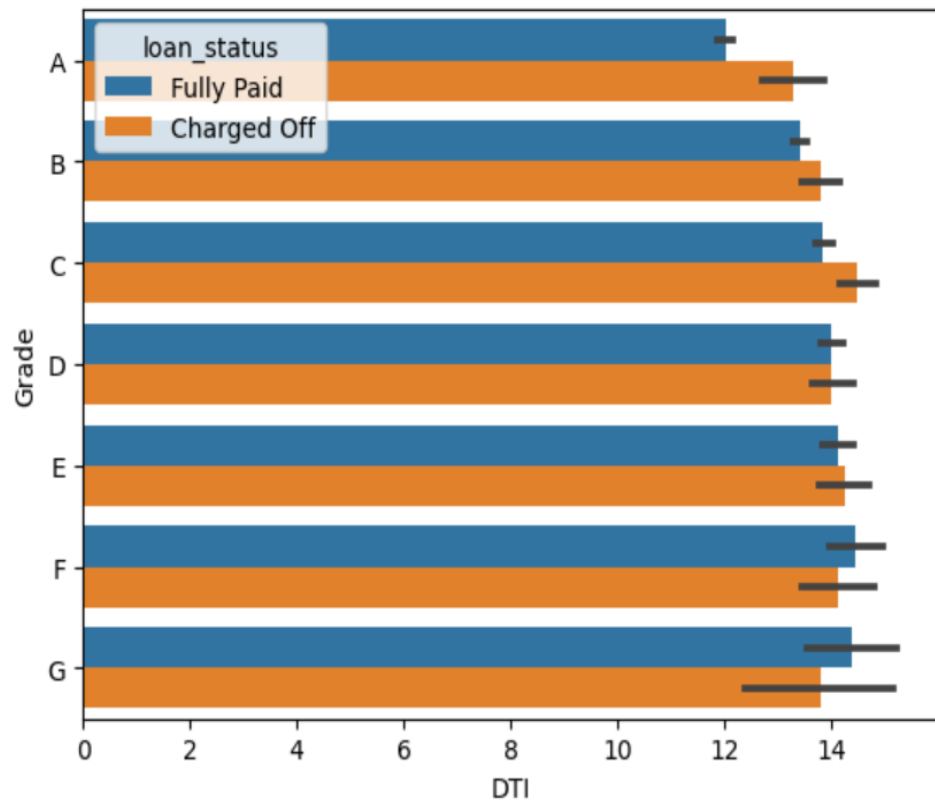
Bivariate analysis involves the examination of two variables simultaneously to understand relationships or correlations between them. In the context of loan default prediction,

Bivariate analysis helps identify how pairs of variables interact and influence the likelihood of default.

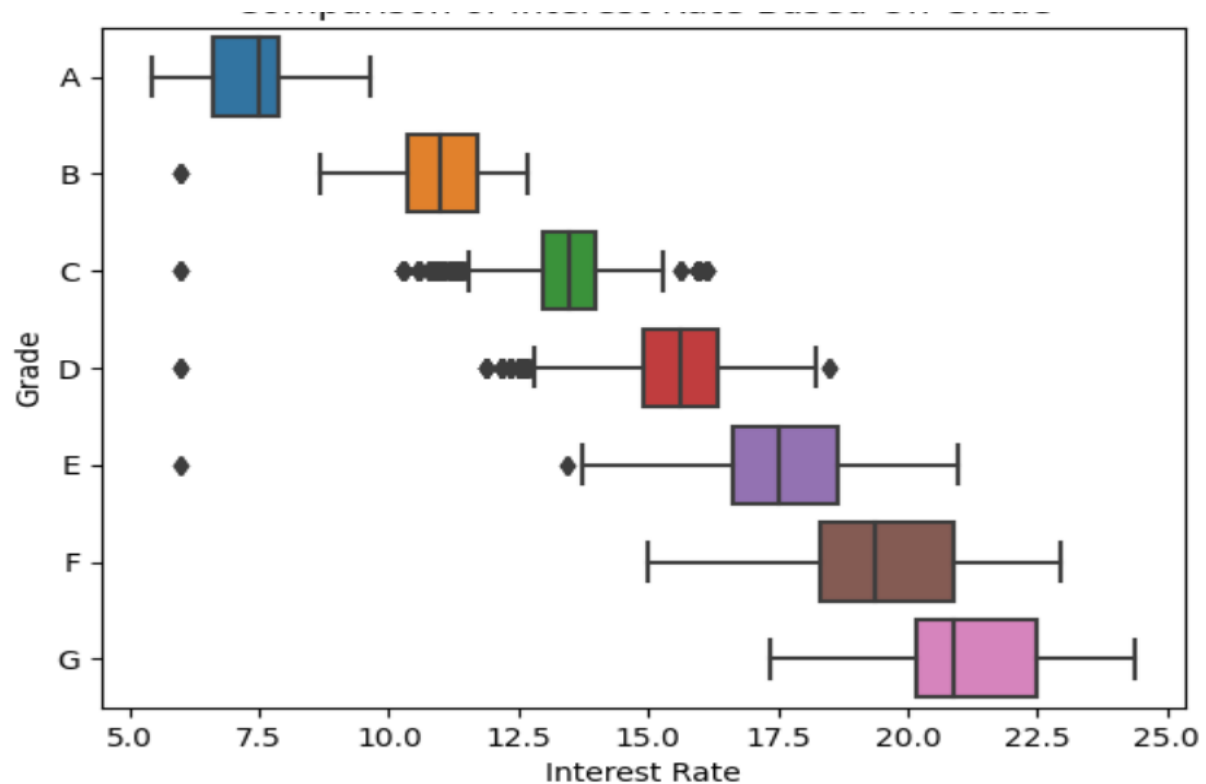
Inferences:

- Grade A which is the lowest risk also has the lowest DTI ratio which we can say that a higher grade has a lower rate of default.
- The people who have taken loans mostly have no record of Public Recorded Bankruptcy and are the safe choice for loan issues.
- As the Grade represents the risk factor we can say interest rate increases with the risk.

BIVARIATE ANALYSIS: Some Visualizations from EDA



Comparison of DTI over grade for loan status



Comparison of interest rate based on grade

RECOMMENDATIONS

Key Predictive Factors for Identifying and Preventing Loan Default and Credit Loss:

- 1)**DTI Ratio:** Leveraging the Debt-to-Income Ratio (DTI) in our risk assessment model is essential, as a high DTI signals potential financial strain, serving as a key predictor for defaults. Setting strategic thresholds and integrating DTI analysis during loan approval enables precise evaluation, minimizing credit loss by gauging applicants' financial stability.
- 2)**Grades:** The grading system is central to assessing borrower creditworthiness; a thorough analysis of historical performance in different grades unveils patterns linked to defaults.
- 3)**Verification Status:** Enhancing the verification process can significantly contribute to reducing defaults. By placing a premium on applications with verified information, we can instill greater confidence in the reliability of borrower data, thereby minimizing the risk of credit loss.
- 4)**Annual income:** Annual income serves as a fundamental indicator of a borrower's financial capacity. Integrating a thorough analysis of annual income into our risk assessment model allows for a more comprehensive understanding of an applicant's ability to meet repayment obligations
- 5)**Pub_rec_bankruptcies:** Examining an applicant's public records, especially bankruptcy history, is paramount in gauging their financial stability and risk of default. Implementing stringent scrutiny and potentially adjusting lending terms for individuals with a history of bankruptcies can be an effective strategy in mitigating credit loss.

RECOMMENDATIONS

Additional considerations for finding loan defaults can include:

- Borrowers residing outside major urban centers such as California, New York, Texas, Florida, etc.
- Borrowers falling within the annual income range of 50,000 to 100,000.
- Borrowers with a history of Public Recorded Bankruptcy.
- Borrowers assigned lower grades (E, F, G), indicating a higher risk profile.
- Borrowers exhibiting significantly elevated Debt-to-Income ratios.
- Borrowers possessing extensive work experience of 10 years or more.

THANK YOU !!