

Lending Club Case Study

By

Rupali Bote

Santosh Raja

Contents

- Problem Statement
- Data Objective
- Data Understanding
- Data Cleaning & Pre-processing
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Summary

Problem Statement

Lending Club, a Consumer Finance marketplace specializing in offering a variety of loans to urban customers, faces a critical challenge in managing its loan approval process. When evaluating loan applications, the company must make sound decisions to minimize financial losses, primarily stemming from loans extended to applicants who are considered “**Risky**”.

These financial losses, referred to as **Credit Losses**, occur when borrowers fail to repay their loans or default. In simpler terms, borrowers labeled as “**Charged-Off**” are the ones responsible for the most significant losses to the company.

The primary objective of this exercise is to assist Lending Club in mitigating credit losses. This challenge arises from two potential scenarios:

- Identifying applicants likely to repay their loans is crucial, as they can generate profits for the company through interest payments. Rejecting such applicants would result in a loss of potential business.
- On the other hand, approving loans for applicants not likely to repay and at risk of default can lead to substantial financial losses for the company.

Objective

The objective is to find applicants at risk of defaulting on loans, enabling a reduction in credit losses. This case study aims to achieve this goal through exploratory data analysis.

Data Dictionary

Lending Club provided us with customer's historical data. This dataset contained information pertaining to the borrower's past credit history and Lending Club loan information. The total dataset consisted of over 39717 records and 111 columns, which was sufficient for our team to conduct analysis. Variables present within the dataset provided an ample amount of information which we could use to identify relationships and gauge their effect upon the success or failure of a borrower fulfilling the terms of their loan agreement.

Data Understanding

1. Loan Amount (loan_amt): Represents the amount of money requested by the borrower as a loan.
2. Grade (grade): Represents a rating assigned to the borrower based on their creditworthiness, indicating the level of risk associated with the loan.
3. Term (term): Duration of the loan, typically expressed in months.
4. Loan Date (issue_d): Date when the loan was issued or approved by the lender.
5. Purpose of Loan (purpose): Indicates the reason for which the borrower is seeking the loan, such as debt consolidation, home improvement, or other purposes.
6. Verification Status (verification_status): Represents whether the borrower's income and other information have been verified by the lender.
7. Interest Rate (int_rate): Represents the annual rate at which the borrower will be charged interest on the loan amount.
8. Installment (installment): Represents the regular monthly payment the borrower needs to make to repay the loan, including both principal and interest.

Data Cleaning & Pre-processing Steps:

1. Loading data from loan CSV
2. Checking for null values in the dataset
3. Checking for unique values
4. Checking for duplicated rows in data
5. Dropping Records & Columns
6. Common Functions
7. Data Conversion
8. Outlier Treatment
9. Imputing values in Columns

Univariate Analysis

Univariate analysis is a statistical method used to analyze and summarize data sets consisting of one variable. It deals with the analysis of a single variable, rather than multiple variables, to understand its distribution, central tendency and dispersion.

It was carried out for both Categorical and Quantitative Variables

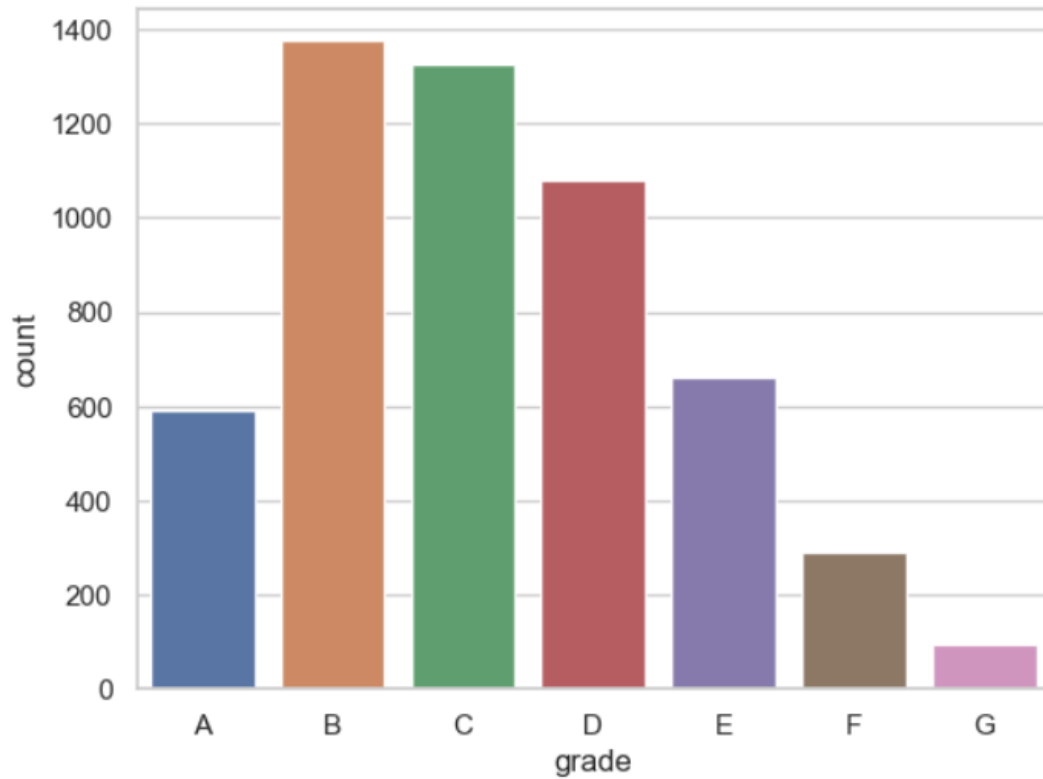
1. Categorical Variables:

- A) Ordered

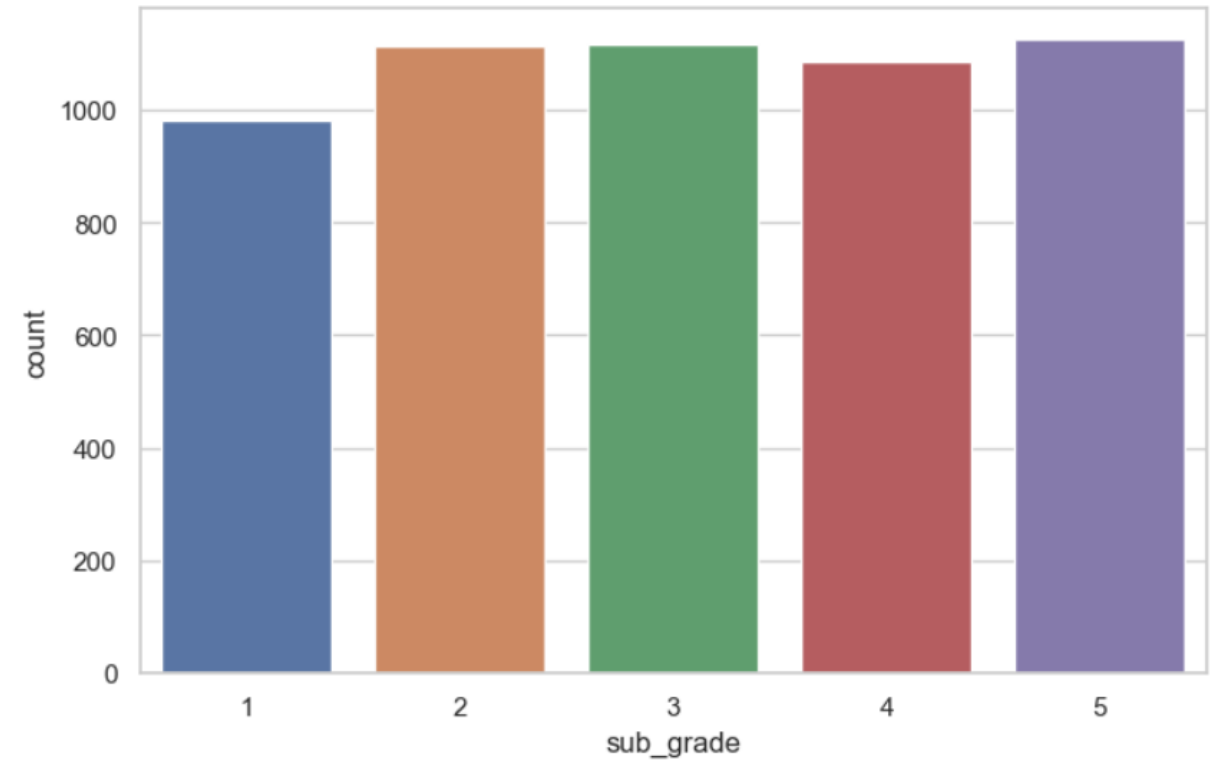
- B) Unordered

2. Quantitative Variables:

Univariate Analysis (Ordered Categorical)

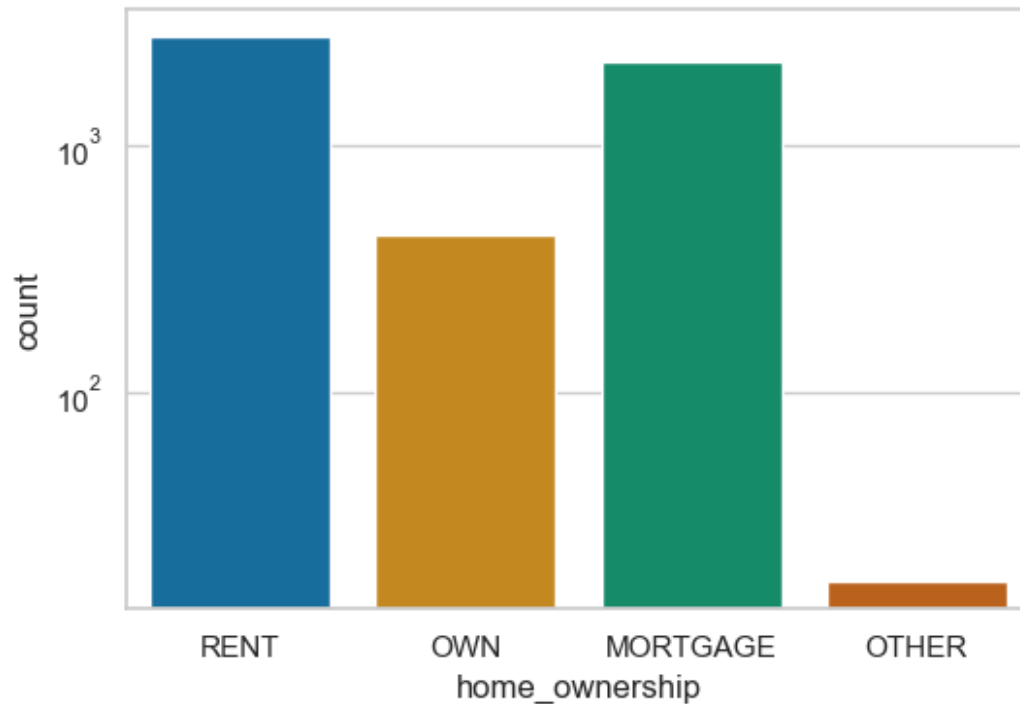


Grade

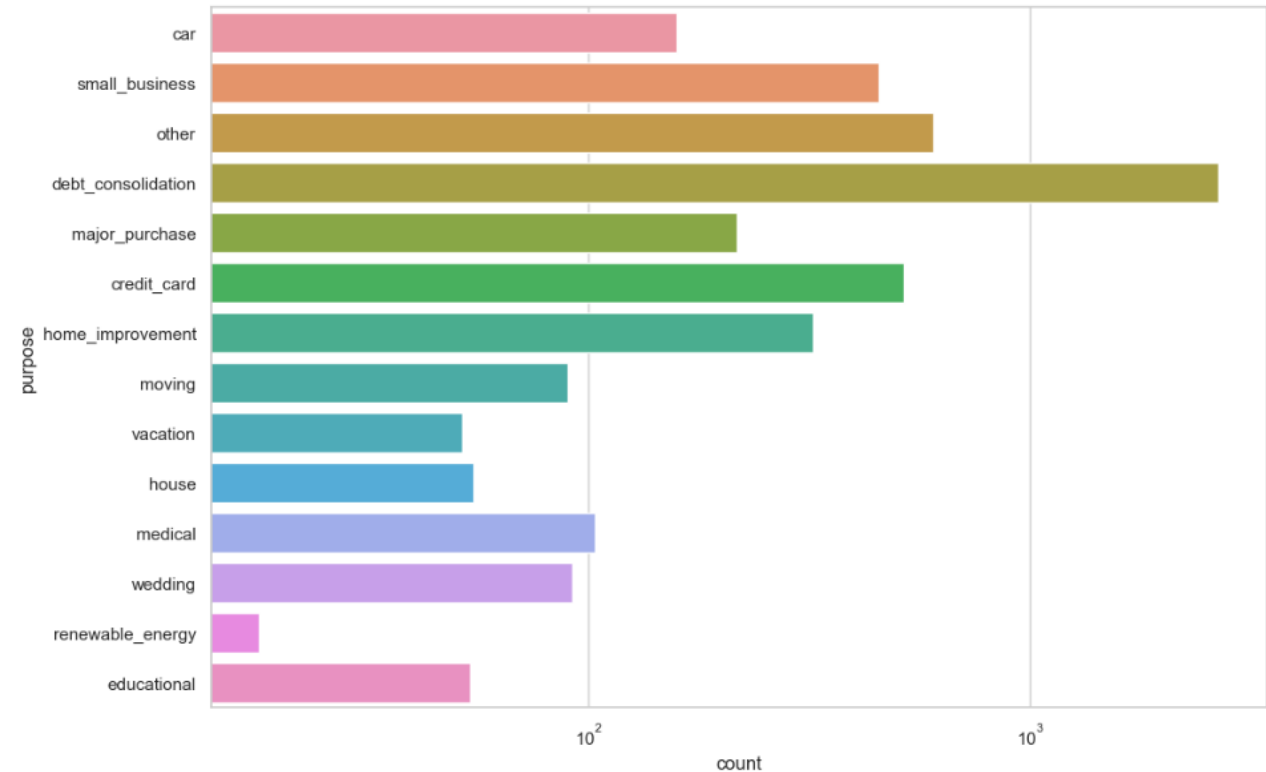


Sub-Grade

Univariate Analysis (Unordered Categorical)

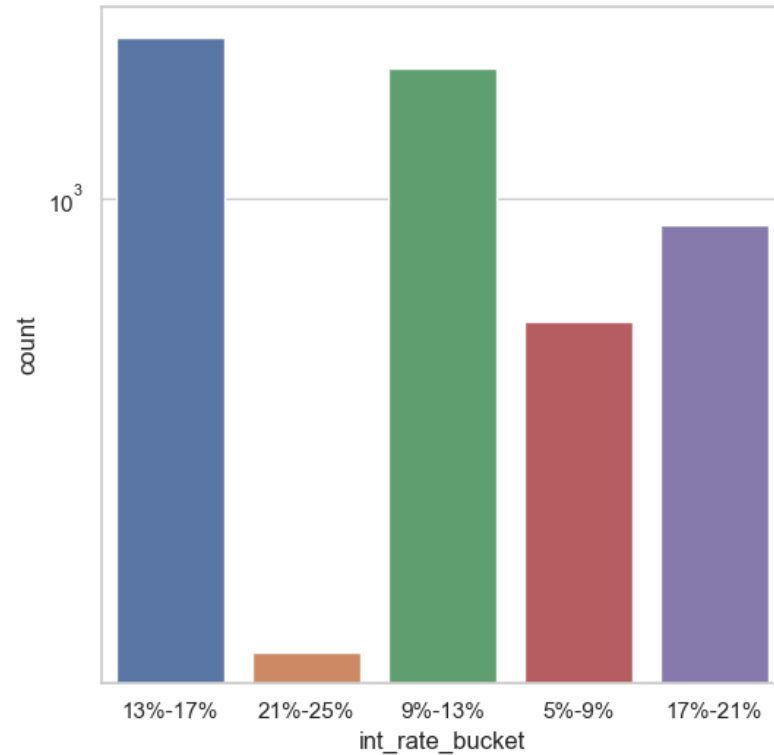


Home Ownership

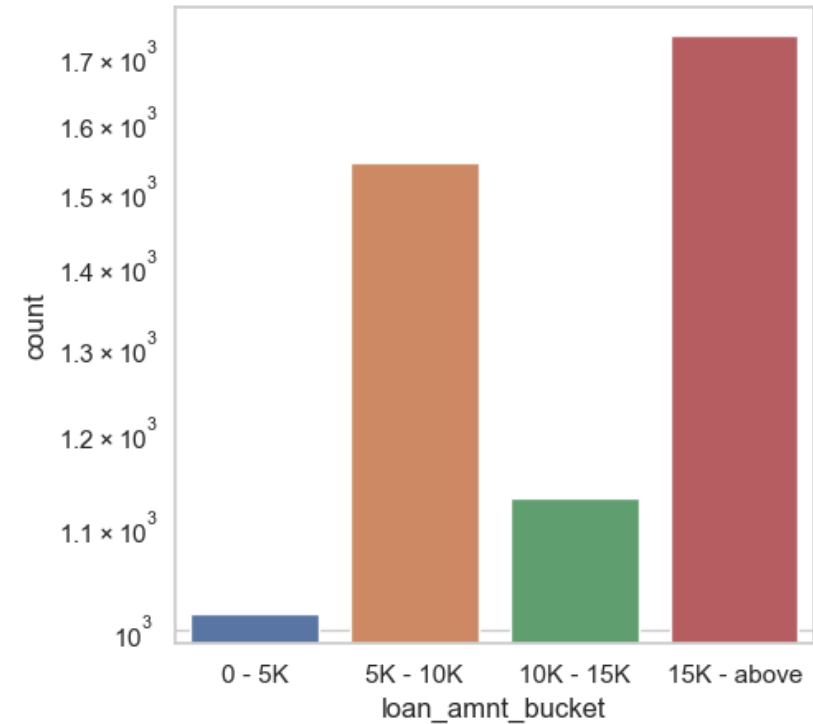


Loan Purpose

Univariate Analysis (Quantitative Variables)



Interest rate



Loan Amount

Observations- Univariant Analysis

Ordered Categorical Variables

- Grade B had the highest number of "Charged off" loan applicants.
- Short-term loans with a duration of 36 months were the most popular among "Charged off" applicants
- Applicants who had been employed for more than 10 years accounted for the highest number of "Charged off" loans.
- The year 2011 recorded the highest number of "Charged off" loan applications
- "Charged off" loans were predominantly taken during the 4th quarter primarily in December.

Observations- Univariant Analysis

Unordered Categorical Variables

- The majority of "Charged off" loan participants who lived in rented houses.
- A significant number of loan participants were loan defaulters, unable to clear their loans.
- California had the highest number of "Charged off" loan applicants,
- Debt consolidation was the primary loan purpose for most "Charged off" loan applicants.

Observations- Univariant Analysis

Quantitative Variables

- Loan applicants who charged off had annual salaries less than 40,000 USD.
- Loan participants who charged off belonged to the interest rate bucket of 13%-17%.
- Loan participants who charged off received loan amounts of 15,000 USD and above.
- Loan participants who charged off received funded amounts of 15,000 USD and above.
- Among loan participants who charged off, loan applicants had very high debt-to-income ratios.
- Among loan participants who charged off, it's observed that the majority of them had monthly installment amounts falling within the range of 160-440 USD.

Bivariate Analysis

Bivariate analysis is a statistical method that involves the simultaneous analysis of two variables (factors). It aims to determine the empirical relationship between them. The analysis can be used to test hypotheses, identify patterns, or explore relationships between the variables.

It was carried out for both Categorical and Quantitative Variables

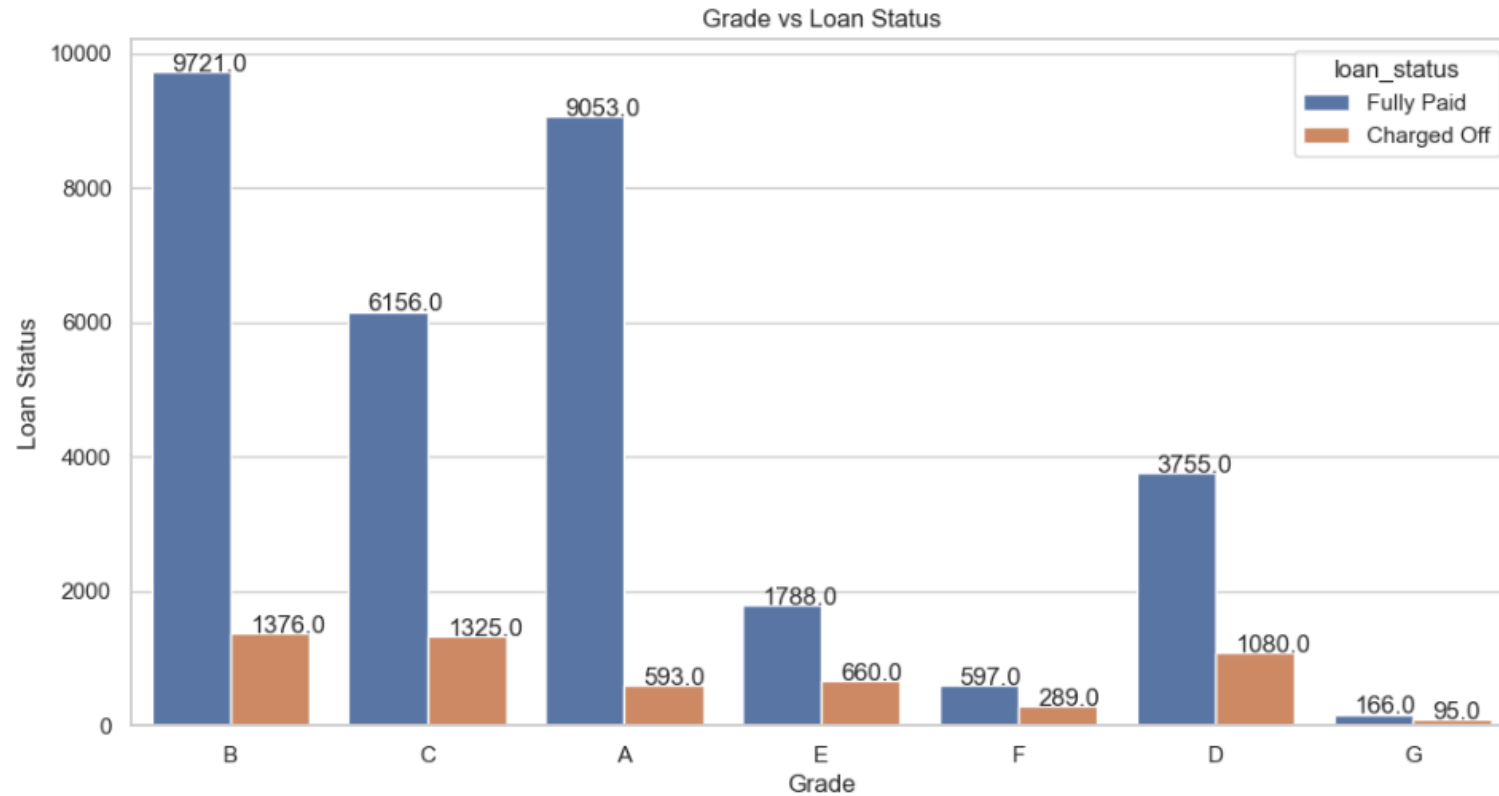
1. Categorical Variables:

- A) Ordered

- B) Unordered

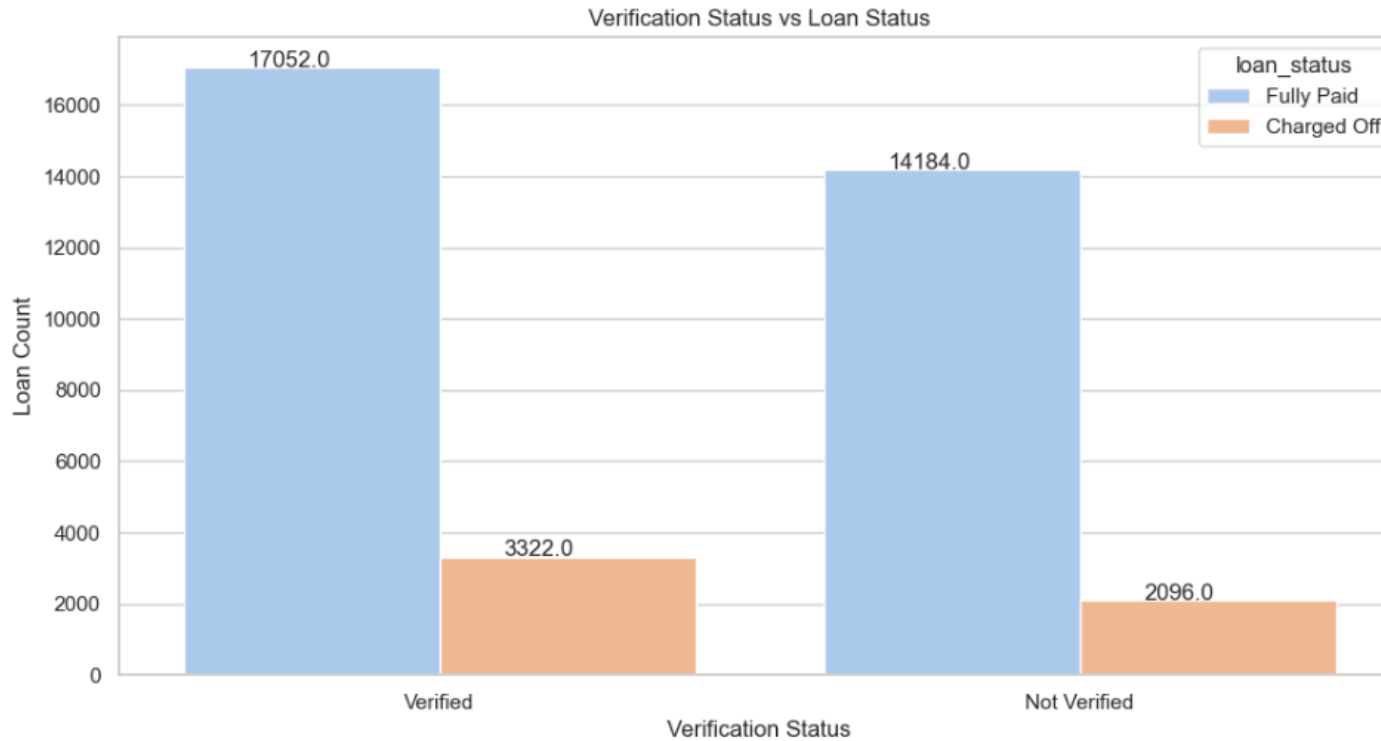
2. Quantitative Variables:

Bivariate Analysis (Ordered Categorical)



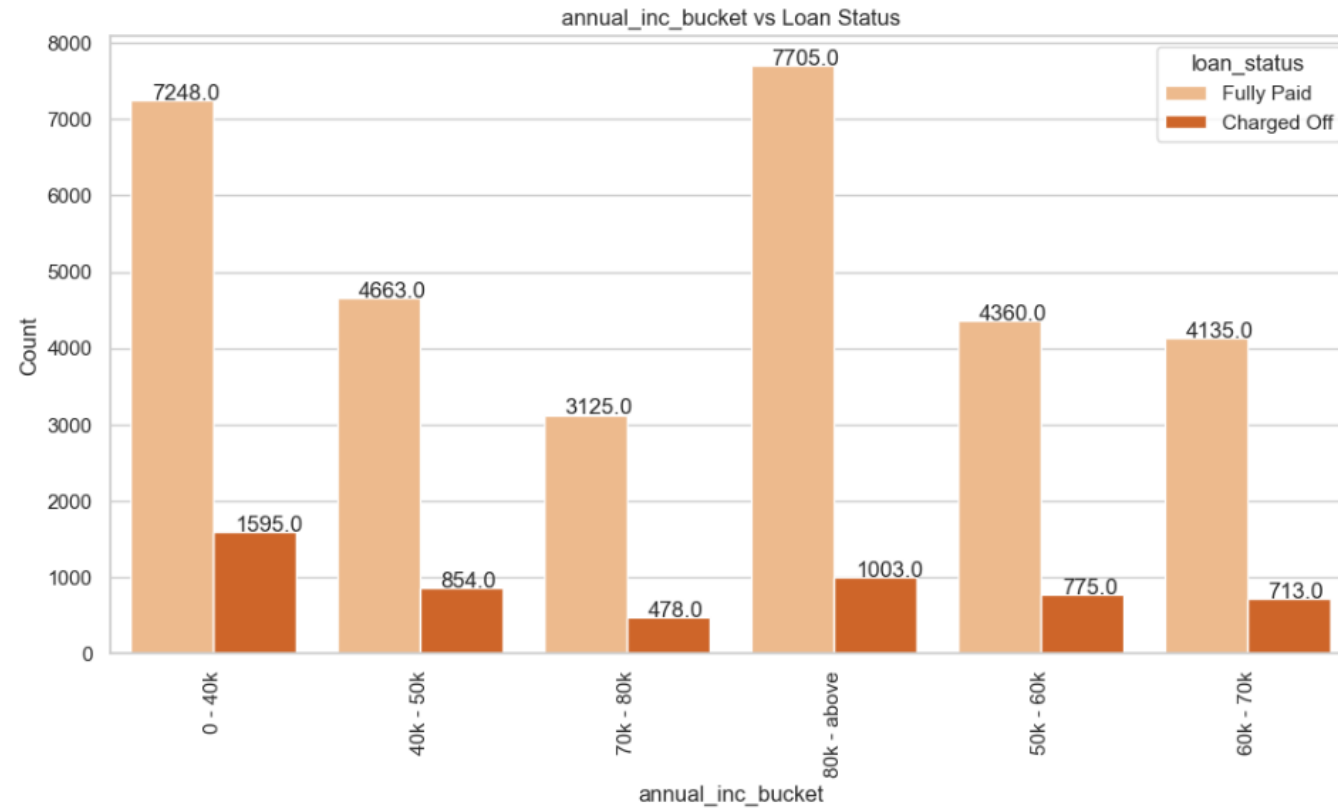
Grade v/s Status of Loan

Bivariate Analysis (Unordered Categorical)



Loan count v/s Verification status

Bivariate Analysis (Quantitative Categorical)



Annual income v/s Loan status

Observations- Bivariate Analysis (Categorical Variables)

Ordered Categorical Variables

- The loan applicants belonging to Grades B, C, and D contribute to most of the "Charged Off" loans.
- Loan applicants belonging to Sub Grades B3, B4, and B5 are more likely to charge off.
- Loan applicants applying for loans with a 60-month term are more likely to default than those taking loans for 36 months.
- Most loan applicants have ten or more years of experience, and they are also the most likely to default.
- The number of loan applicants has steadily increased from 2007 to 2011, indicating a positive trend in the upcoming years.
- December is the most preferred month for taking loans, possibly due to the holiday season.

Observations- Bivariate Analysis (Categorical Variables)

Unordered Categorical Variables

- Debt consolidation is the category where the maximum number of loans are issued, and people have defaulted the most in the same category.
- Loan applicants who live in rented or mortgaged houses are more likely to default.
- Verified loan applicants are defaulting more than those who are not verified.
- Loan applicants from the states of California (CA), Florida (FL), and New York (NY) are most likely to default.

Observations- Bivariate Analysis (Quantitative Variables)

Quantitative Variables

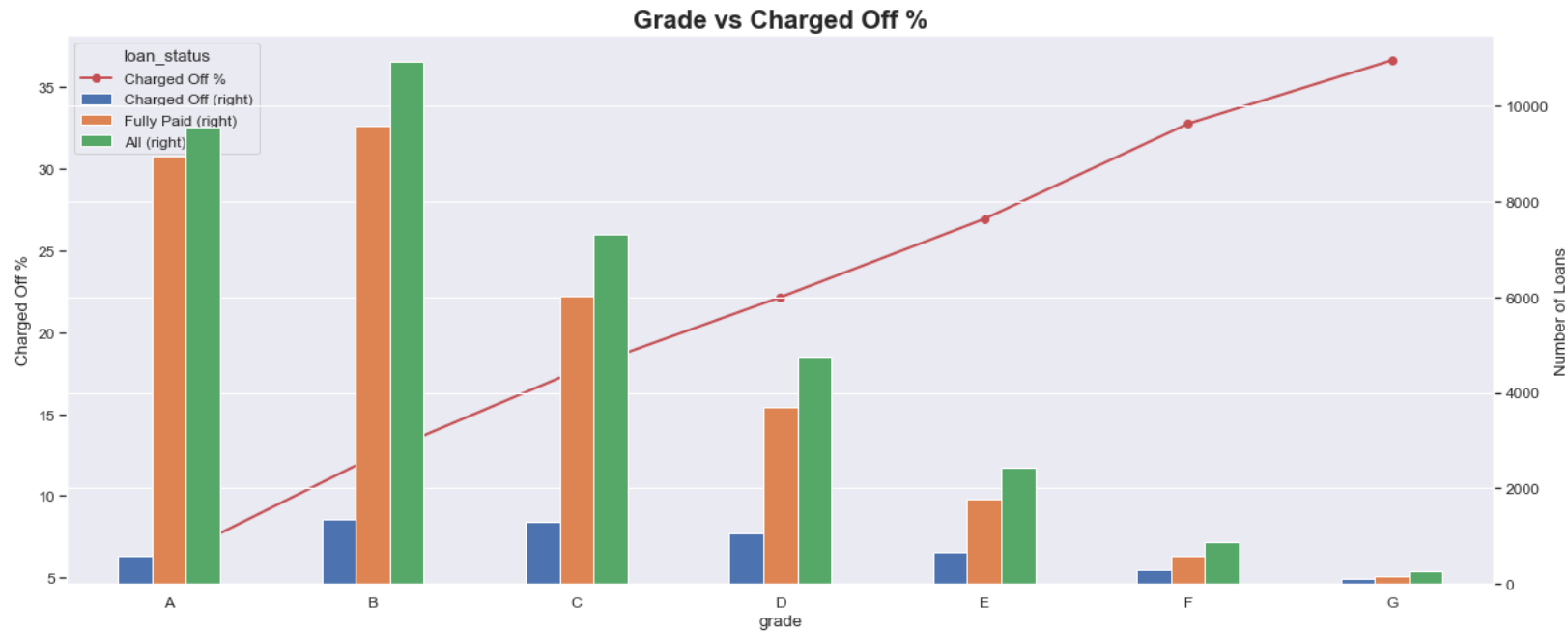
- Loan applicants from the states of California (CA), Florida (FL), and New York (NY) are most likely to default. A majority of the loan applicants who defaulted received loan amounts of \$15,000 or higher.
- The majority of loan applicants who charged off had significantly high Debt-to-Income (DTI) ratios.
- A significant portion of loan applicants who defaulted received loans with interest rates falling within the range of 13% to 17%.
- A majority of the loan applicants who charged off reported an annual income of less than \$40,000.

Multivariate Analysis

Multivariate analysis is a statistical technique used to analyze data that involves more than two variables. Unlike univariate analysis (which deals with one variable) and bivariate analysis (which deals with two variables), multivariate analysis examines the relationships between multiple variables simultaneously.

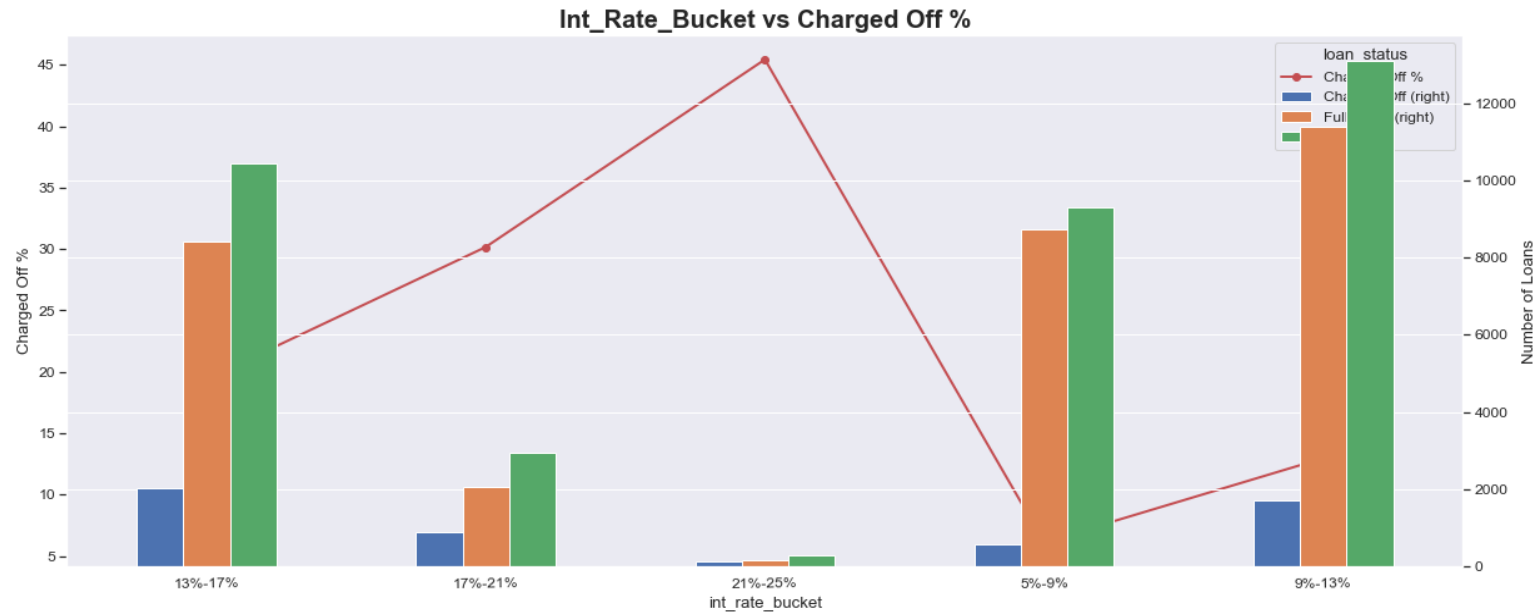
It is widely used in various fields such as economics, social sciences, biology, marketing and environmental science. Multivariate analysis can include different types of variables, such as categorical variables, numerical variables, or a combination of both.

Multivariate Analysis



Grade v/s Percentage of Charged-off Loans

Multivariate Analysis



Buckets of Interest Rate v/s Percentage
of Charged-off Loans

Observations- Multivariate Analysis

- Tendency to default the loan is likely with loan applicants belonging to B, C, D grades.
- Borrowers from sub grade B3, B4 and B5 have maximum tendency to default.
- Loan applicants with 10 years of experience has maximum tendency to default the loan.
- Borrowers from states CA, FL, NJ have maximum tendency to default the loan.
- Borrowers from Rented House Ownership have highest tendency to default the loan.

Summary

To prevent the increasing defaulters lending club should use following points:

- Implement Stricter Criteria for Grades B, C, and D: Consider implementing stricter risk assessment and underwriting criteria for applicants falling into Grades B, C, and D to minimize default risks.
- Focus on Subgrades B3, B4, and B5: Pay special attention to applicants with Subgrades B3, B4, and B5. Consider additional risk mitigation measures or offering lower loan amounts for these subgrades to reduce default rates.
- Evaluate and Limit 60-Month Loans: Evaluate the risk associated with 60-month loans. Consider limiting the maximum term or adjusting interest rates for longer-term loans to decrease the likelihood of defaults.
- Comprehensive Credit Scoring System: Develop a comprehensive credit scoring system that incorporates various risk-related attributes, as experience alone might not be sufficient to gauge creditworthiness.
- Capitalizing on Market Growth: Capitalize on the market's growth trend observed from 2007 to 2011 by maintaining a competitive edge in the industry while ensuring robust risk management practices.
- Anticipate Peak Periods: Anticipate increased loan applications during peak periods such as December and Q4. Ensure efficient processing to meet customer demands during these busy seasons.

Summary

- Careful Evaluation for Debt Consolidation Loans: Carefully evaluate applicants seeking debt consolidation loans, considering potential interest rate adjustments or offering financial counseling services to manage the associated risks.
- Consider Housing Stability: Take housing status into account during the underwriting process to assess housing stability and its impact on the applicant's ability to repay the loan.
- Review Verification Process: Review the verification process to ensure effective assessment of applicant creditworthiness. Consider improvements or adjustments based on the review findings.
- Monitor & Adjust for Regional Risk Trends: Monitor regional risk trends, especially in states like California, Florida, and New York. Adjust lending strategies or rates accordingly in high-risk regions.
- Thorough Assessment for High Loan Amounts: Conduct more thorough assessments for loan amounts of \$15,000 or higher. Consider capping loan amounts for higher-risk applicants to mitigate potential defaults.

Thank You!

