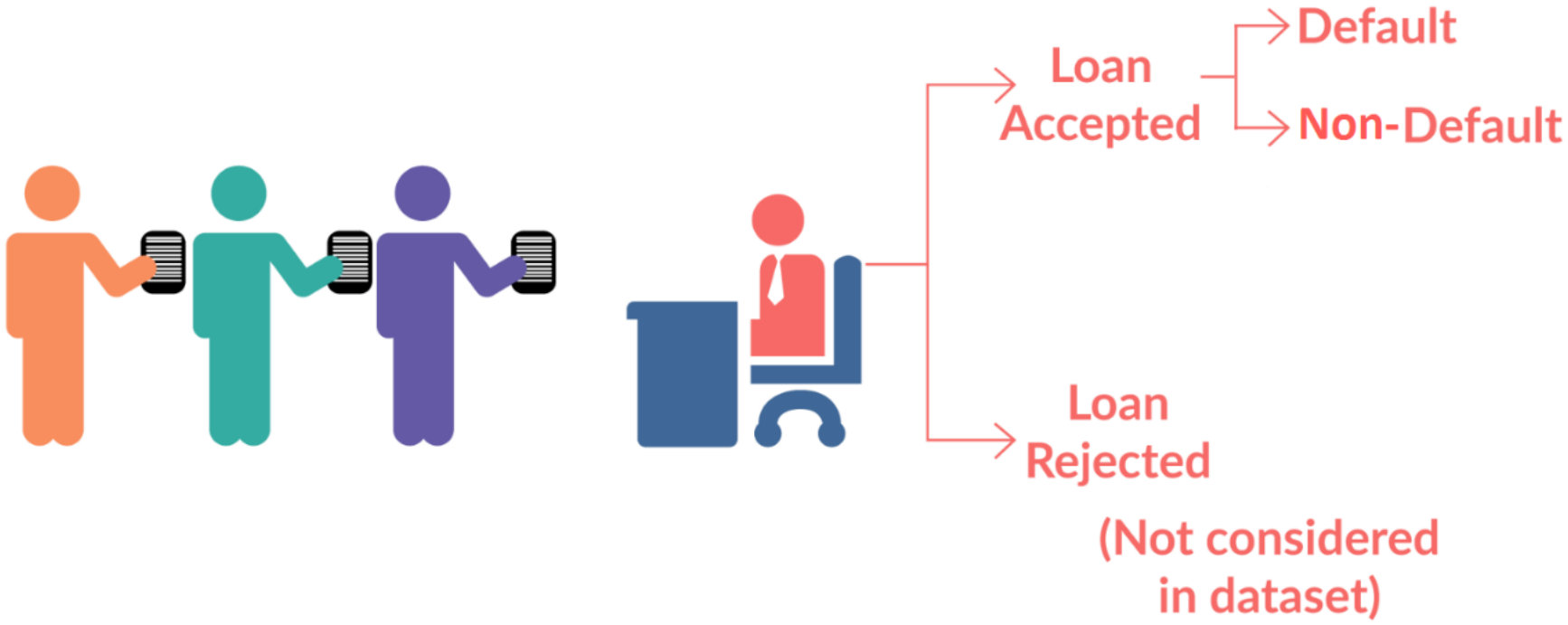
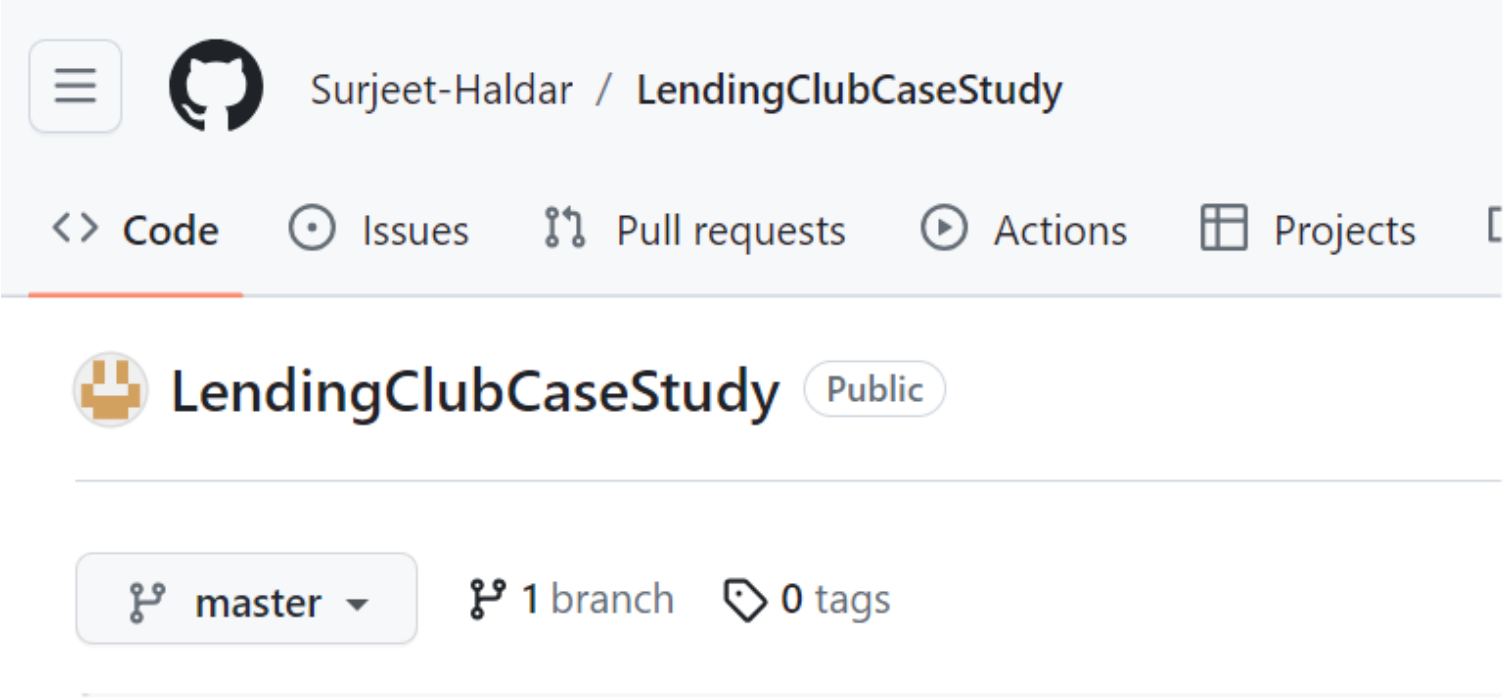
LENDING CLUB CASE STUDY



AGENDA

- Introduction
- Business objective
- Data understanding
- Approach
- Case study analysis
- Technologies used
- Summary



INTRODUCTION

When the company receives a loan application, it needs to make a decision on loan approval based on the applicant's profile. There are two types of risks associated with the company's decision:

1. If the applicant is likely to repay the loan, not approving the loan results in a loss of business.

2. If the applicant is likely to default, approving the loan may lead to a financial loss for the company.

The aim of this case study is to identify patterns in the data that indicate if a person is likely to default on the loan.

BUSINESS OBJECTIVE

- Identify risky loan applicants and reduce the amount of credit loss.

- Understand the driving factors behind loan default to inform portfolio and risk assessment.

DATA UNDERSTANDING

The dataset contains information about past loan applicants and whether they defaulted on the loan or not.

Based on below approach will identify its significance in the context of risk analytics and lending.

- Analysis based on Domain Understanding
- Columns Analysis of the Dataset
- Ignored Rows and Columns because of missing data

APPROACH

- Perform Exploratory Data Analysis (EDA) to understand the data and identify patterns.
- Analyse the distribution of variables, handle missing values, and perform data cleaning if necessary.
- Conduct univariate analysis to understand the distribution of individual variables.
- Perform bivariate analysis to explore the relationship between variables and default status.
- Identify significant variables that strongly indicate default.
- Summarize the important results and insights from the analysis.

CASE STUDY ANALYSIS

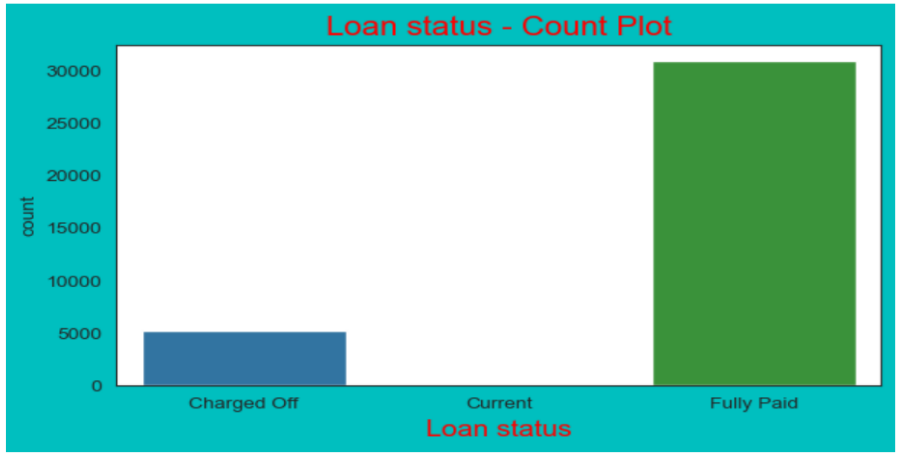
After performing exploratory data analysis (EDA) on the loan dataset, several conclusions and insights can be derived. Here are some key findings approach:

- Univariate Analysis
- Bivariate Analysis

Univariate Analysis Inferences:

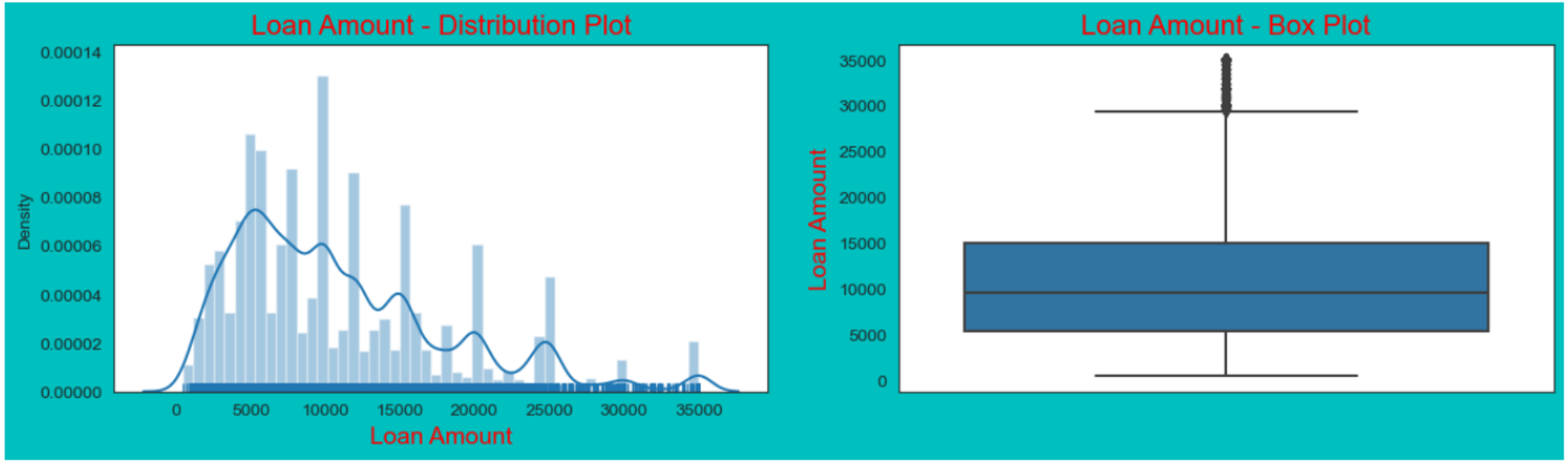
From the univariate analysis of the loan dataset, we can draw the following inferences:

Univariate Analysis on Loan Status



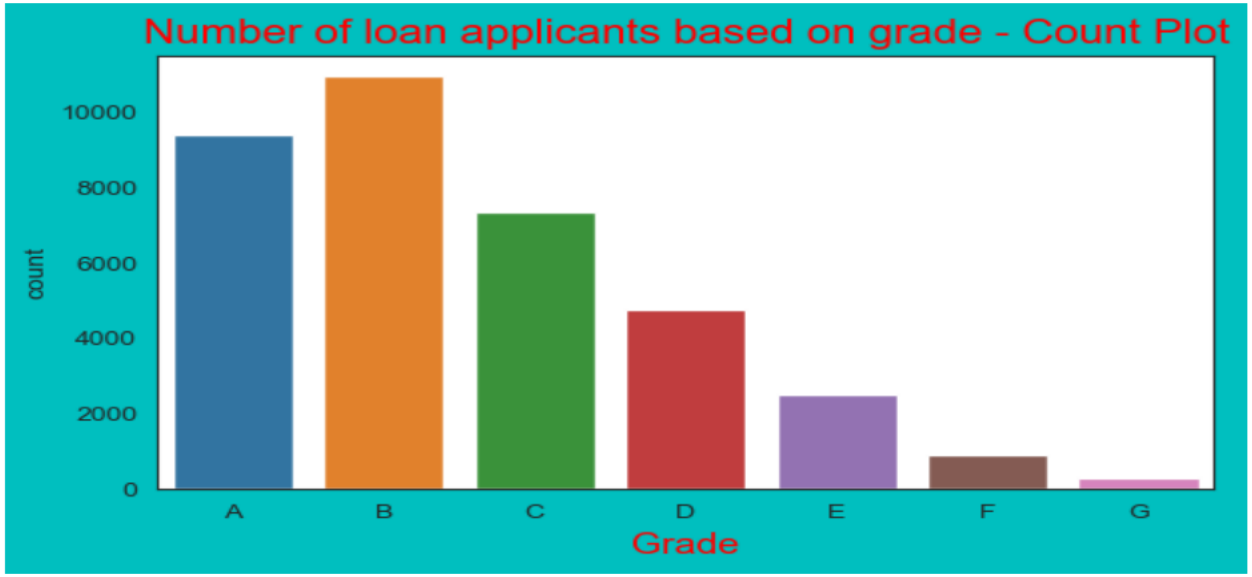
Inferences: The majority of loans in the dataset are fully paid, followed by charged-off loans. This indicates that the company has a relatively low default rate.

Univariate Analysis on Loan Amounts



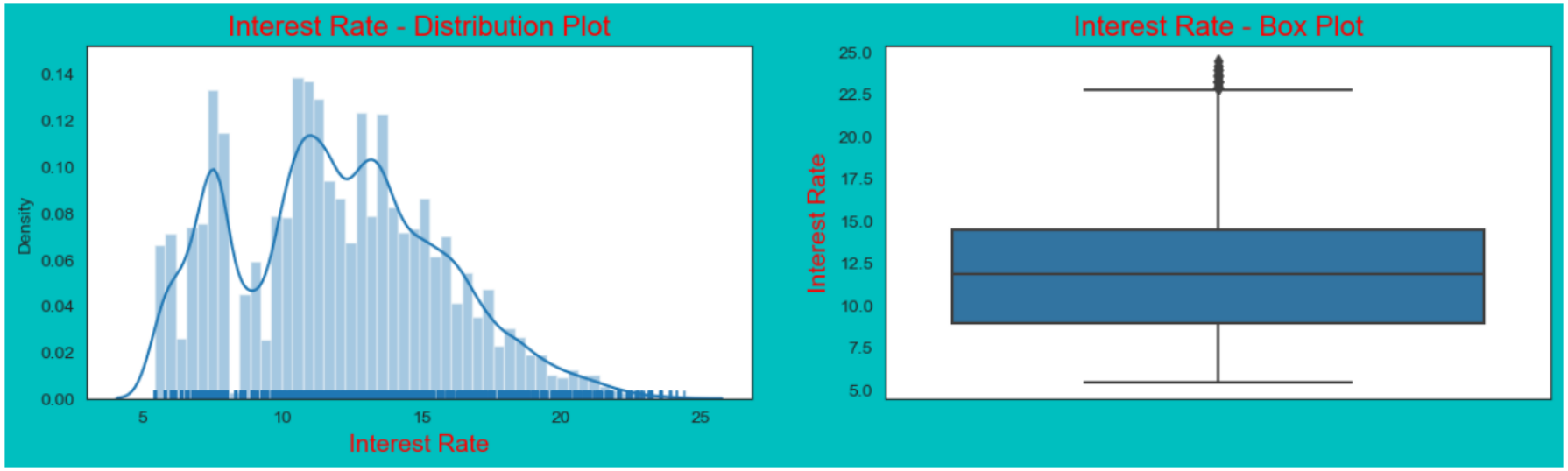
Inferences: The distribution of loan amounts is right-skewed, with a higher concentration of loans in the lower amount range. This suggests that the company primarily provides smaller loans.

Univariate Analysis on Loan Grade



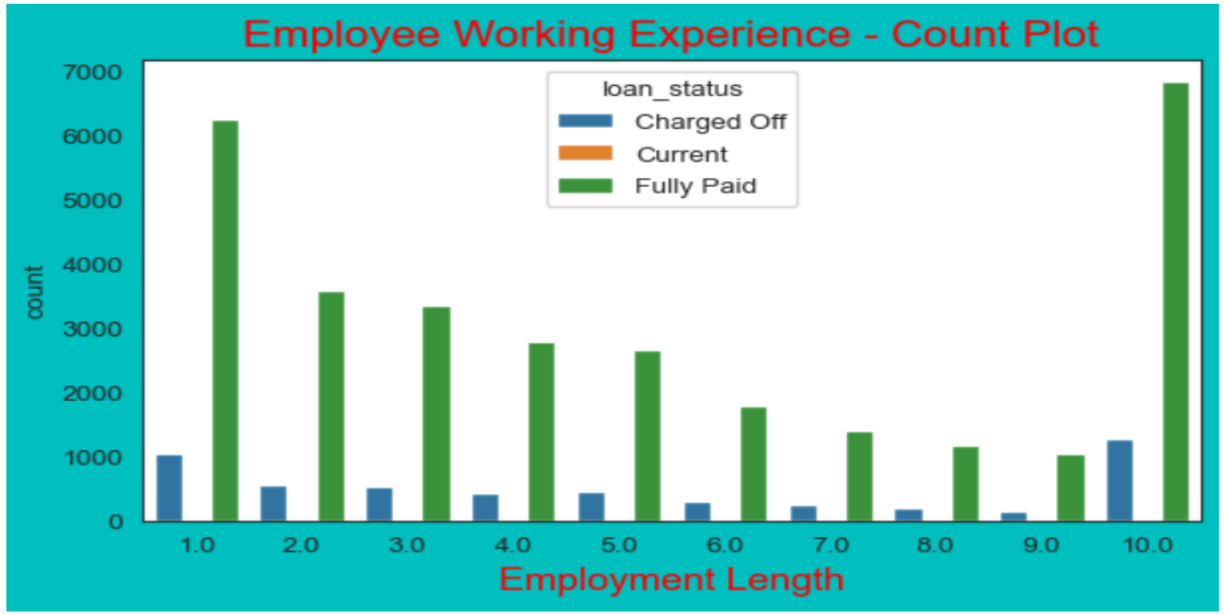
Inferences: The loan grades range from 'A' to 'G', with 'B' being the highest grade representing lower risk. The dataset has a higher number of loans with grades 'A' and 'C', indicating that the company mostly lends to borrowers with moderate creditworthiness.

Univariate Analysis on Interest Rate



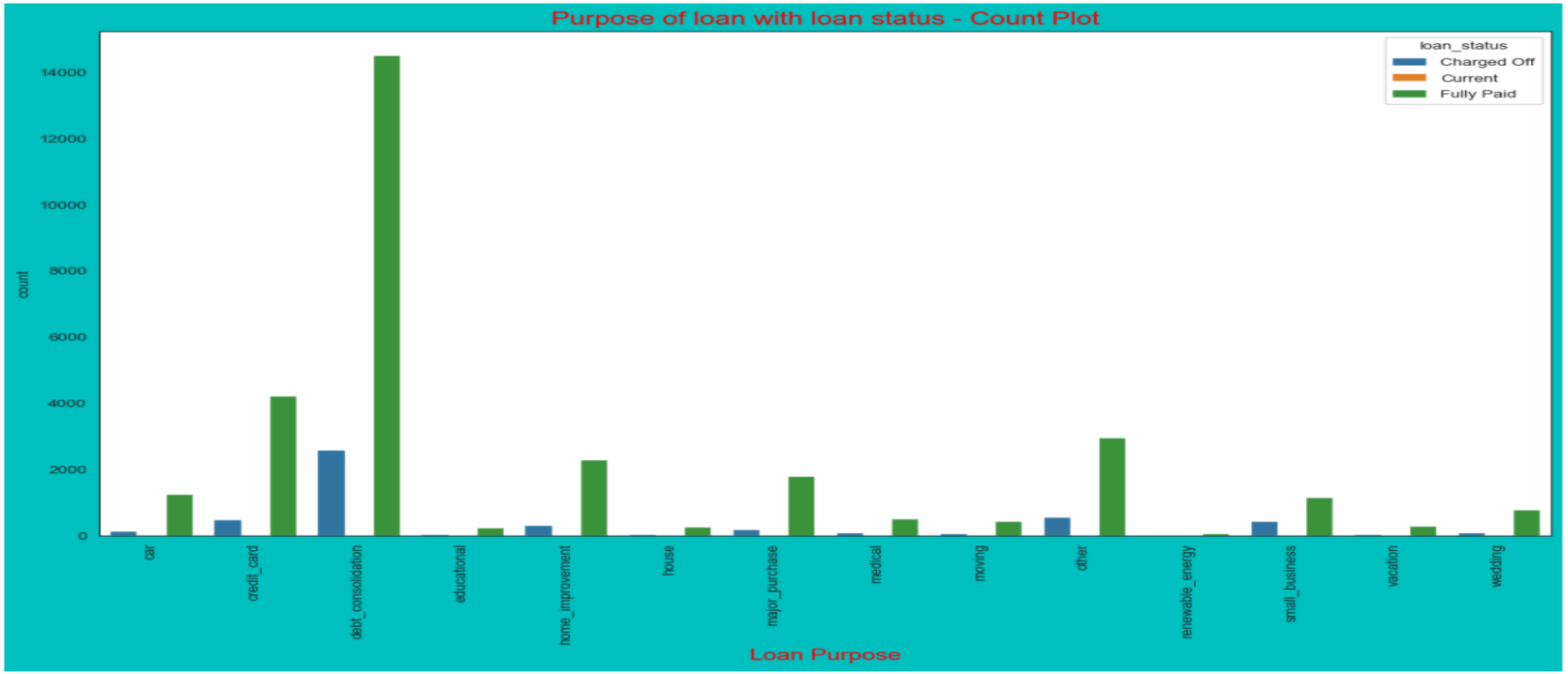
Inferences: The interest rates assigned to loans range from 5% to 25%. The distribution is approximately normally distributed, with a peak around 10% to 15% interest rate. Higher interest rates are typically associated with riskier loans.

Univariate Analysis on Employment Length



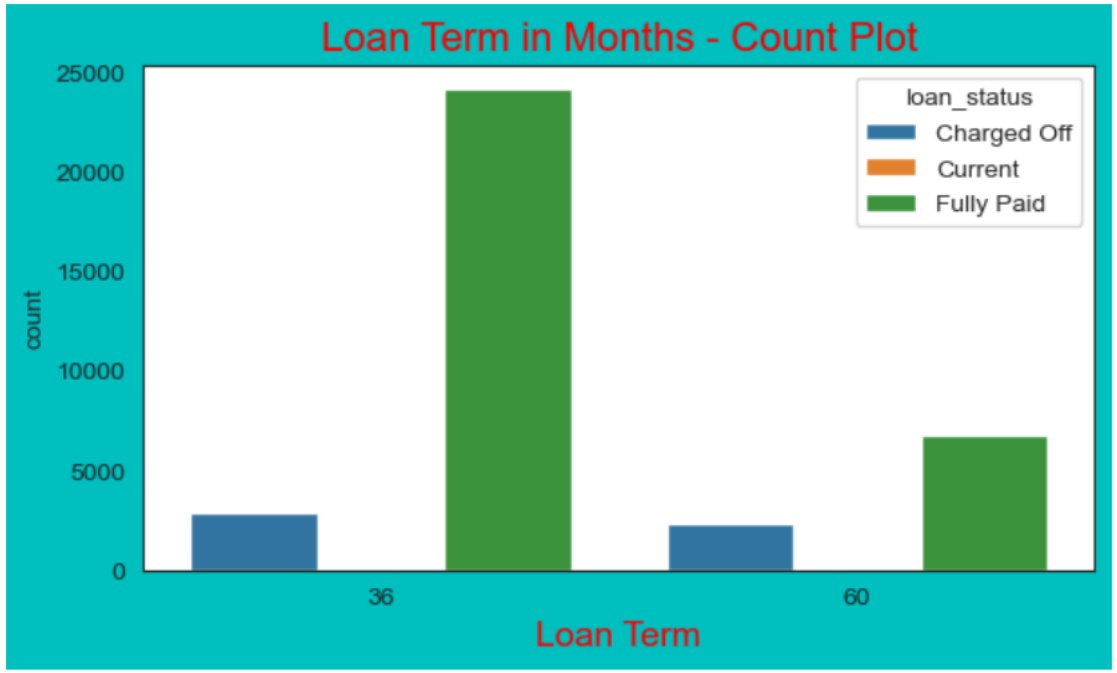
Inferences: The length of employment of borrowers ranges from less than 1 year to over 10 years. The majority of borrowers have an employment length of more than 1 year, suggesting that the company prefers lending to individuals with stable employment.

Univariate Analysis on Loan Purpose



Inferences: Maximum number of loan is lend for the purpose of debt consolidation and it has high chance to be fully paid, it also has maximum number of Charged off compared to other purpose of loan. Less number of loan is granted to Education, Renewable energy.

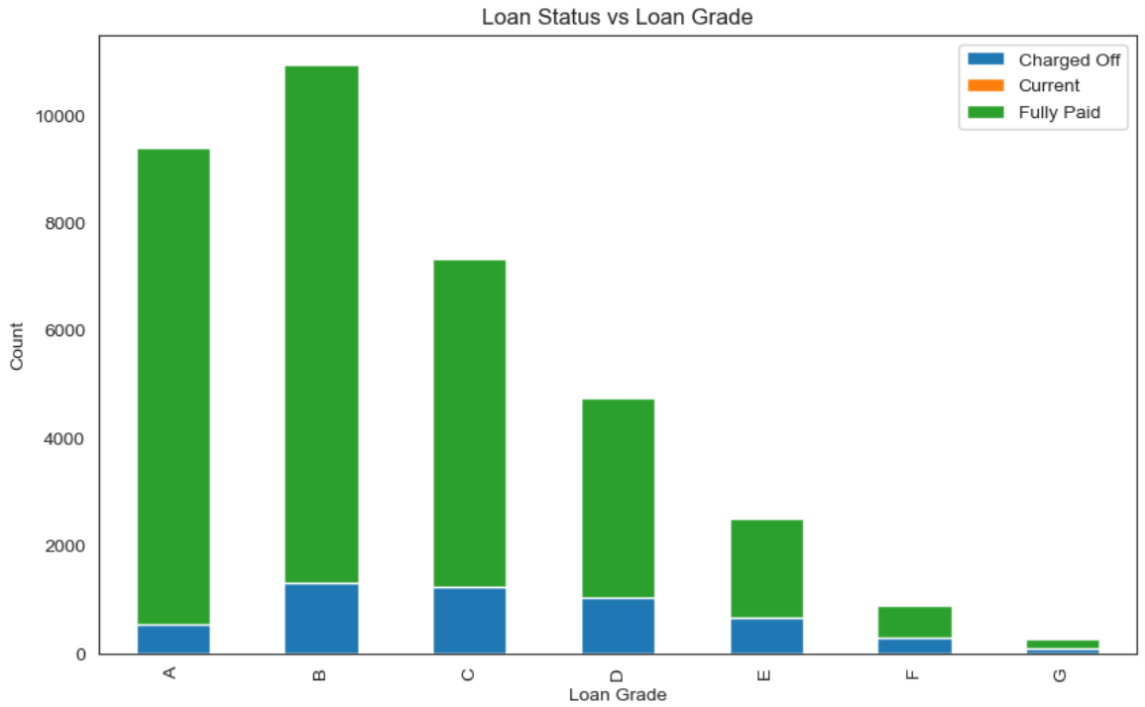
Univariate Analysis on Loan Term



Inferences: Loan is granted to applicants for all purposes either for 36 months or 60 months. debt_consolidation lead high in both the terms followed by credit card. Most of the loan applicants are given 36 months for the loan term.

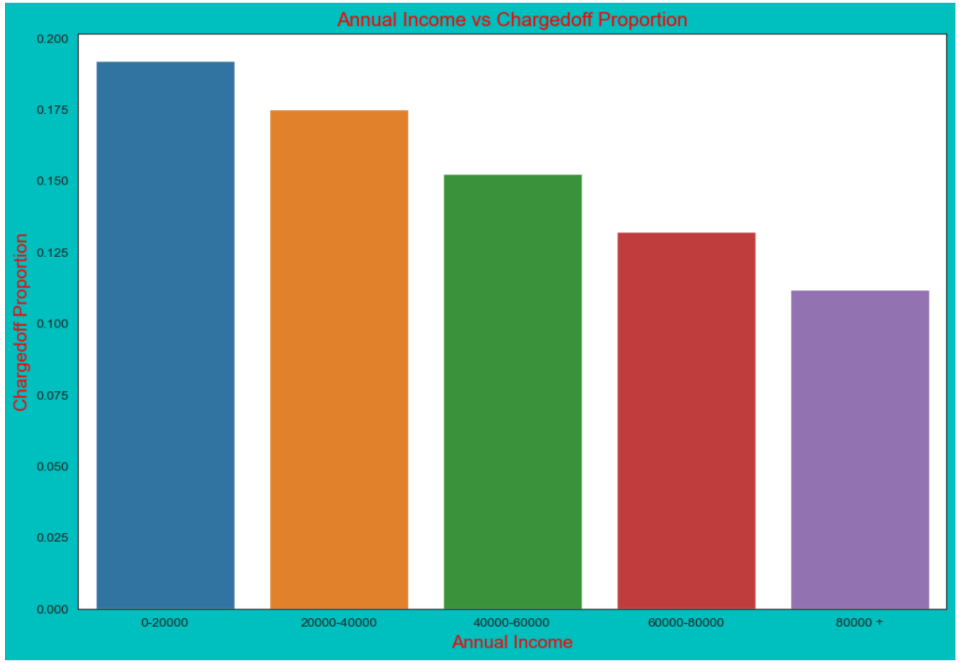
Bivariate Analysis Inferences:

Bivariate Analysis on Loan Status vs Loan Grade



Inferences: There is a clear association between Loan Status and Loan Grade. Higher loan grades (such as A or B) have a higher proportion of fully paid loans, indicating lower default risk. Conversely, lower loan grades (such as E or F) have a higher proportion of charged-off loans, indicating higher default risk.

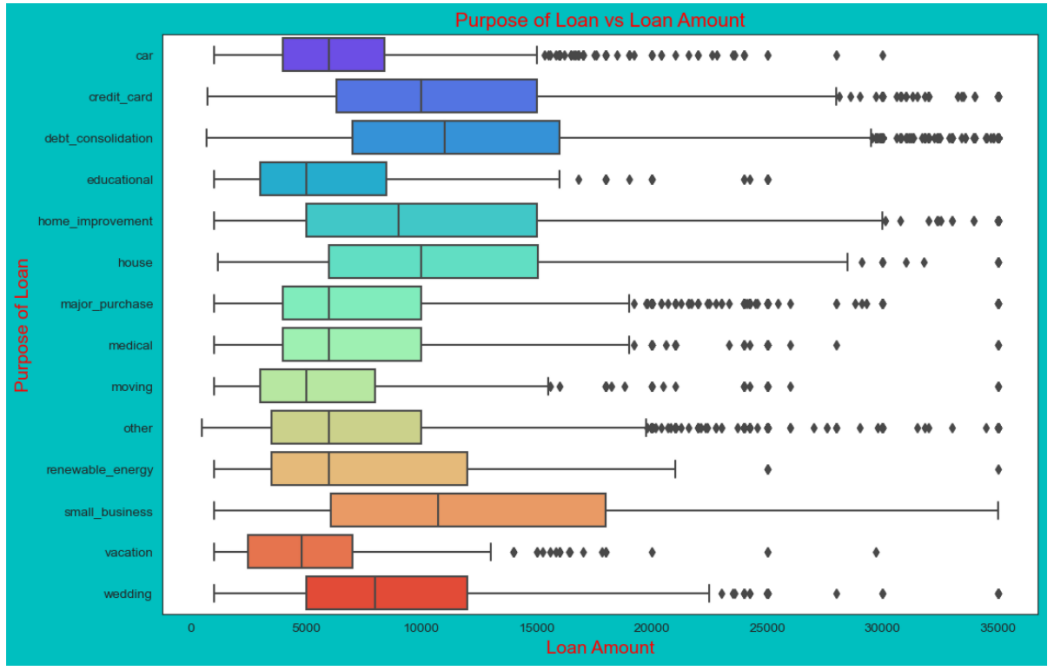
Bivariate Analysis on Annual income vs Charged off Proportion

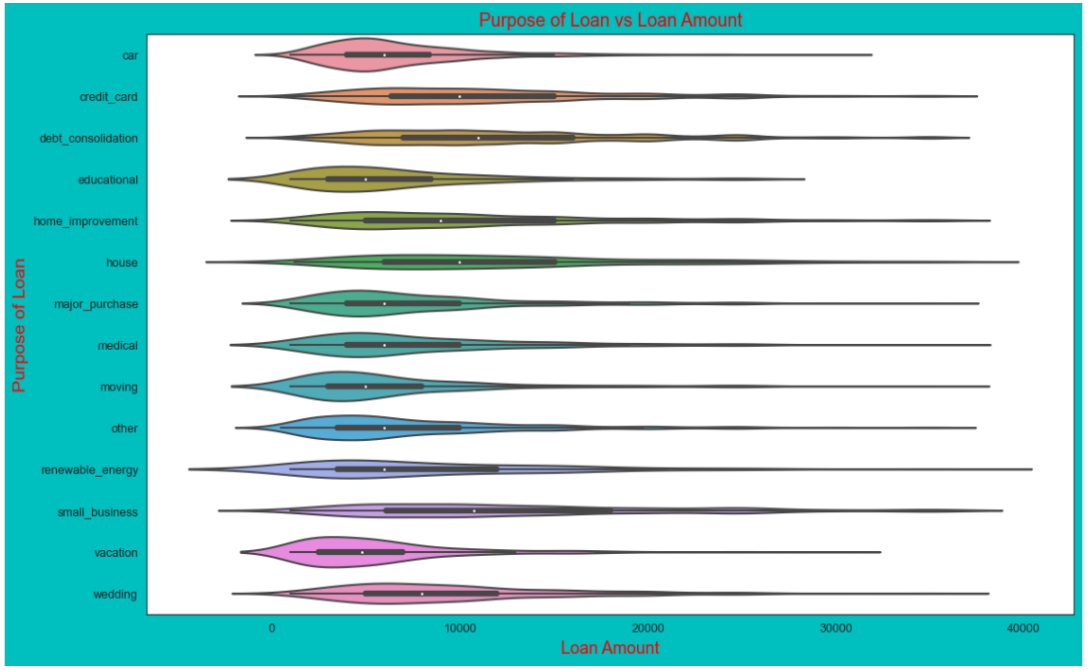


Inferences: Higher Annual Income: Borrowers with higher annual incomes tend to have a lower proportion of charged-off loans. This suggests that higher income individuals have a better ability to repay their loans and are less likely to default.

Income Range Impact: There may be certain income ranges where the proportion of charged-off loans is relatively higher. This indicates that borrowers within these income ranges may face more financial challenges or have higher default risk compared to others.

Bivariate Analysis on Purpose of loan and Loan amount

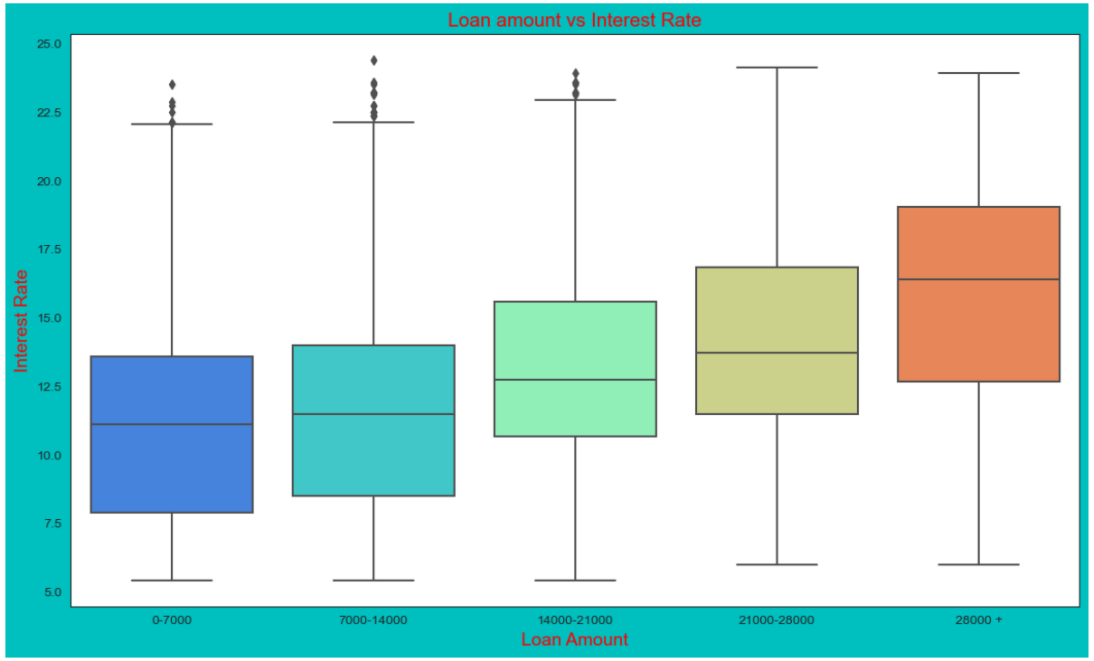




Inferences: Loan Amount Variation: The purpose of the loan has a significant impact on the loan amount. Certain loan purposes tend to have higher average loan amounts, indicating a greater financial need or investment value associated with those purposes.

Major Loan Purposes: The most common loan purposes tend to be debt consolidation, credit card refinancing, home improvement, and major purchases. These purposes often require higher loan amounts due to the nature of the expenses involved.

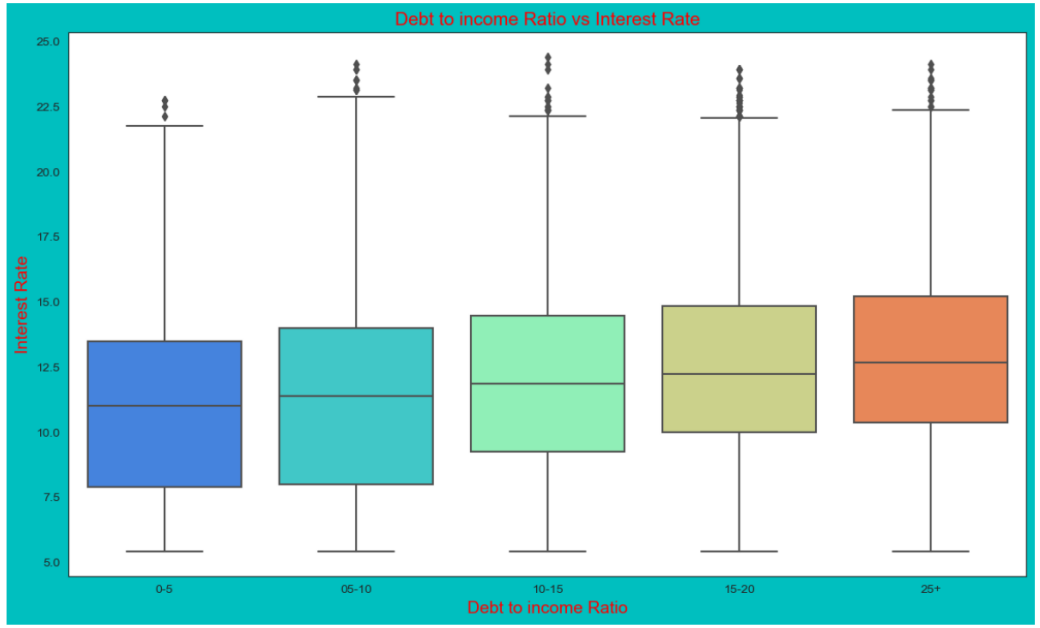
Bivariate Analysis on Loan Amount and Interest Rate



Inferences: Positive Relationship: There is a positive relationship between Loan Amount and Interest Rate. As the loan amount increases, the interest rate tends to be higher. This suggests that lenders may charge higher interest rates for larger loan amounts to compensate for the increased risk associated with larger loans.

Risk Assessment: Loan Amount and Interest Rate can serve as indicators of the borrower's creditworthiness and risk profile. Higher loan amounts may indicate a greater need for funds or higher financial commitments, which could lead to a higher default risk.

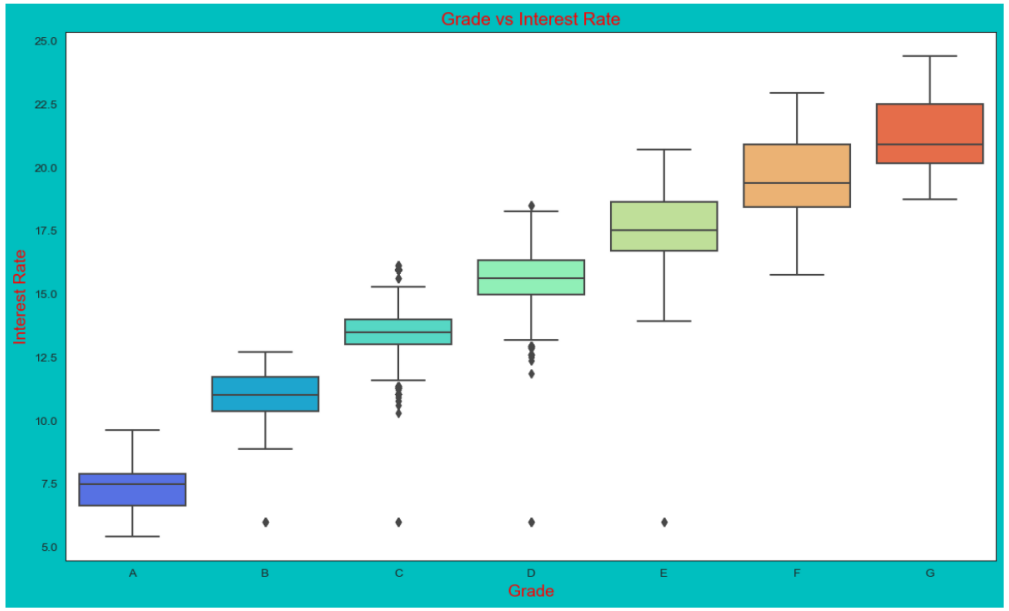
Bivariate Analysis on DTI and Interest Rate



Inferences: Impact on Interest Rate: There is a positive relationship between the DTI ratio and the Interest Rate. As the DTI ratio increases, the Interest Rate tends to be higher. This suggests that borrowers with higher DTI ratios, indicating a higher level of debt relative to income, are perceived as higher risk borrowers by lenders, resulting in higher interest rates.

Risk Assessment: The DTI ratio is an important factor in assessing the borrower's ability to repay the loan. A higher DTI ratio indicates a higher level of debt obligations, which may impact the borrower's financial stability and ability to make timely loan repayments. Lenders may charge higher interest rates to compensate for the increased risk associated with borrowers with higher DTI ratios.

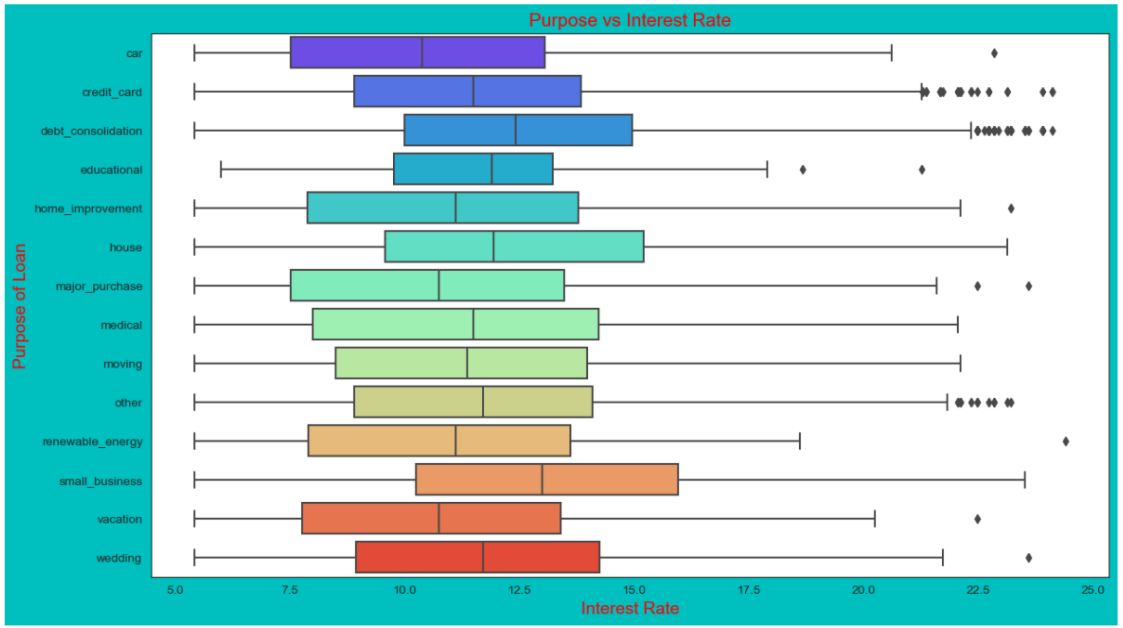
Bivariate Analysis on Grade and Interest Rate



Inferences: Relationship between Grade and Interest Rate: There is a clear relationship between the loan Grade and the Interest Rate. Higher-grade loans (e.g., A, B) tend to have lower Interest Rates, while lower-grade loans (e.g., E, F) have higher Interest Rates. This suggests that the loan Grade is a significant factor considered by lenders in determining the Interest Rate for borrowers.

Creditworthiness and Risk Assessment: The Grade assigned to a loan reflects the borrower's creditworthiness and the level of risk associated with the loan. Higher-grade loans are typically assigned to borrowers with a strong credit history, low default risk, and good repayment capacity. Lenders offer lower Interest Rates to these borrowers as they are considered less likely to default.

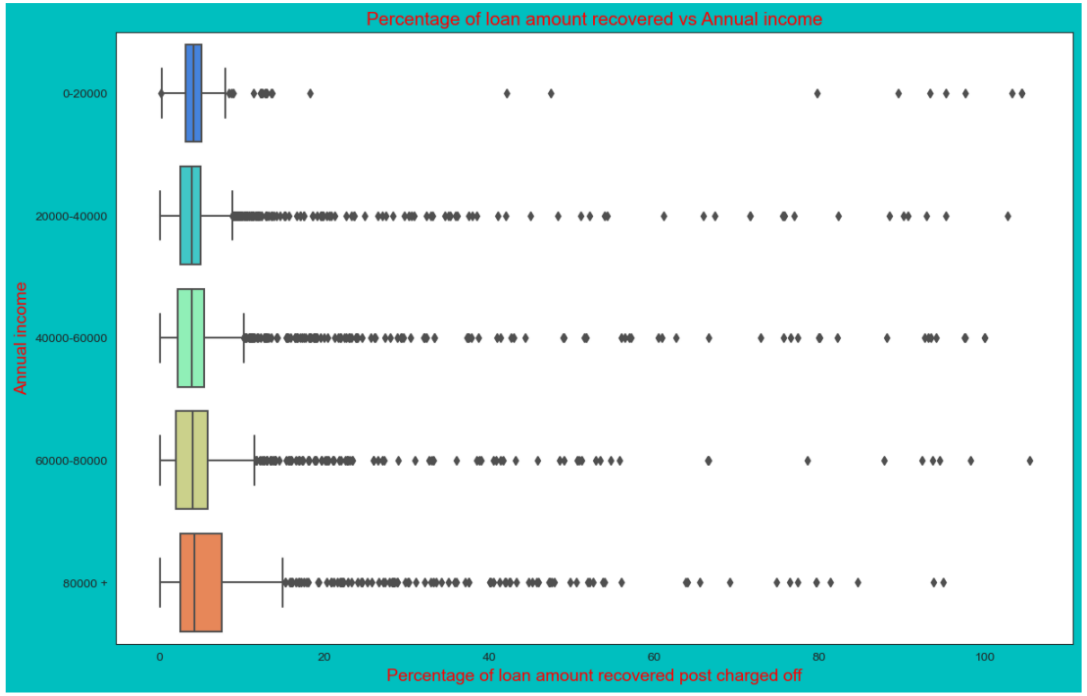
Bivariate Analysis on Purpose and Interest Rate



Inferences: Variation in Interest Rates: The purpose of the loan has an impact on the interest rate charged. Different loan purposes are associated with different interest rates. Some loan purposes may have higher interest rates, indicating higher risk or complexity, while others may have lower interest rates, indicating lower risk or lower financial obligations.

Risk Assessment: The purpose of the loan provides insights into the borrower's intended use of the funds and their financial situation. Lenders assess the risk associated with each loan purpose and adjust the interest rate accordingly. Higher-risk loan purposes may be charged higher interest rates to compensate for the increased likelihood of default.

Bivariate Analysis on Recoveries and Annual income



Inferences: Higher percentage of loan amount is recovered when annual income is high. Plot shows no significant variation but there is slight increase in recovery percentage with increase in annual income.

SUMMARY

•The lending club case study focuses on a consumer finance company that specializes in lending various types of loans to urban customers.
•The company receives loan applications and makes decisions on loan approvals based on the applicant's profile, considering the risk associated with each applicant.
•Two types of risks are involved: approving a loan to an applicant who is likely to default, resulting in a financial loss, and not approving a loan to an applicant who is likely to repay, resulting in a loss of business.
•The goal of the case study is to identify patterns and indicators of loan default using exploratory data analysis (EDA).
•EDA is performed on a loan dataset containing information about past loan applicants and their loan repayment status.
•The dataset includes attributes such as loan grade, annual income, loan amount, interest rate, purpose of the loan, and more.
•Univariate analysis is performed to understand the distribution and characteristics of individual variables.
•Bivariate analysis is performed to understand the relationship between two variables, such as loan status and loan grade, annual income and charged-off proportion, purpose of the loan and loan amount, etc.
•The findings from the analysis can be utilized by the lending company to make informed decisions on loan approvals, interest rates, and risk assessment.
•The case study highlights the importance of EDA in understanding risk analytics in the banking and financial services industry.

TECHNOLOGIES USED

- PYTHON
- NUMPY
- PANDAS
- MATPLOTLIB
- SEABORN
- JUPYPTER NOTEBOOK
- ANDCONDA

THANK YOU

Surjeet Haldar