Lending Club Case Study

Group Members:
Siddhi Shree
Vaishno Jha

Problem Statement

Lending Club specializes in providing a diverse range of loans to urban clientele. Upon receipt of a loan application, the company is tasked with **making a loan approval determination** grounded in the applicant's profile.

Associated Risks

- Not approving the loan for **creditworthy applicants leads to business loss**.
- **Approving high-risk** loans may lead to company losses.

Objective

- The company aims to comprehend the influential factors behind loan defaults, specifically the variables that strongly signal potential defaults.

- This understanding can be harnessed by the company to enhance its portfolio management and risk evaluation strategies.

Understanding the data

Based on our initial observation, we have gathered the following insights and grasp of the data and its dimensions:

- 39717 rows and 111 columns
- Size: 33MB
- Over 60 rows have either single values or only null values.
- Data has variety of data type: integer, float, strings, sentences, links, etc.
- Possible Scenarios: Fully Paid, Current, Charged-Off

Steps of Data Analysis

- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Segmented Analysis
- Derived Analysis

Data Cleaning Steps

- Removed columns with single or null values
- Removed Columns with more than 10K null values
- Removed unnecessary columns
- Reformatting of columns
- Filtering rows based on column
- Filtering Rows with null column values
- Removing outliers

Univariate Analysis: Loan Status
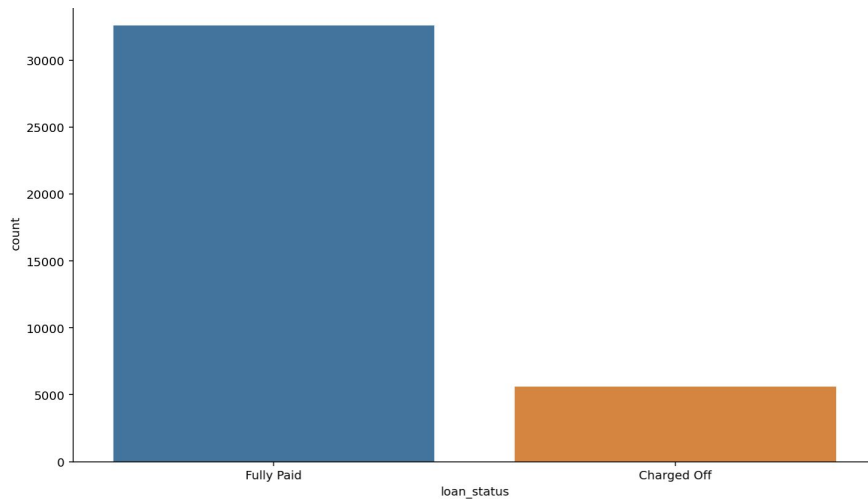
- The number of defaulted loans is very low in comparison to fully paid loans.

Univariate Analysis : Interest Rate

Variation of Interest Rate
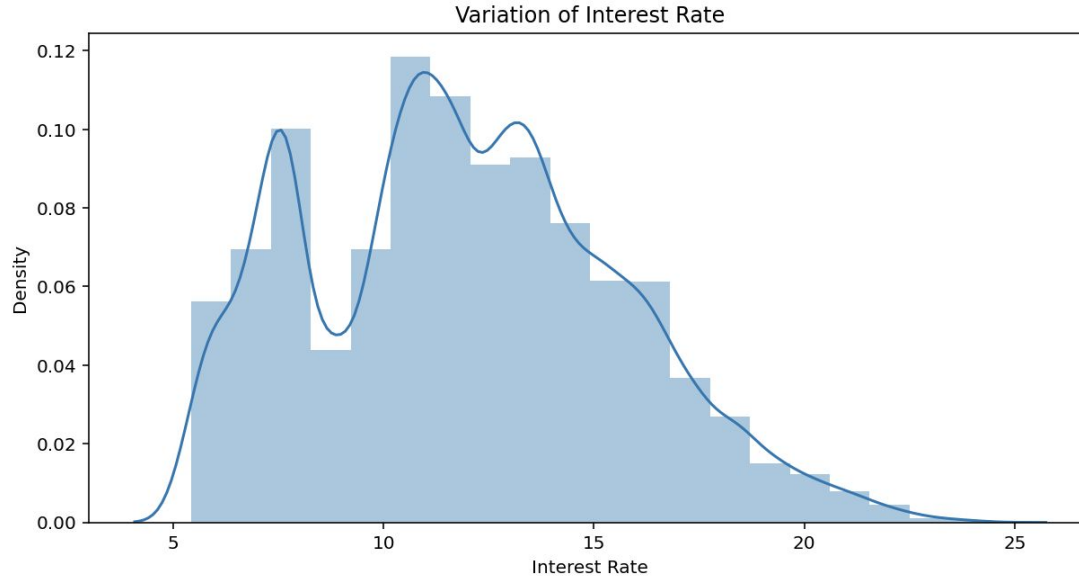
- The interest rate shows more concentration within the ranges of 5-10 and 10-15, with a dip occurring around 10.
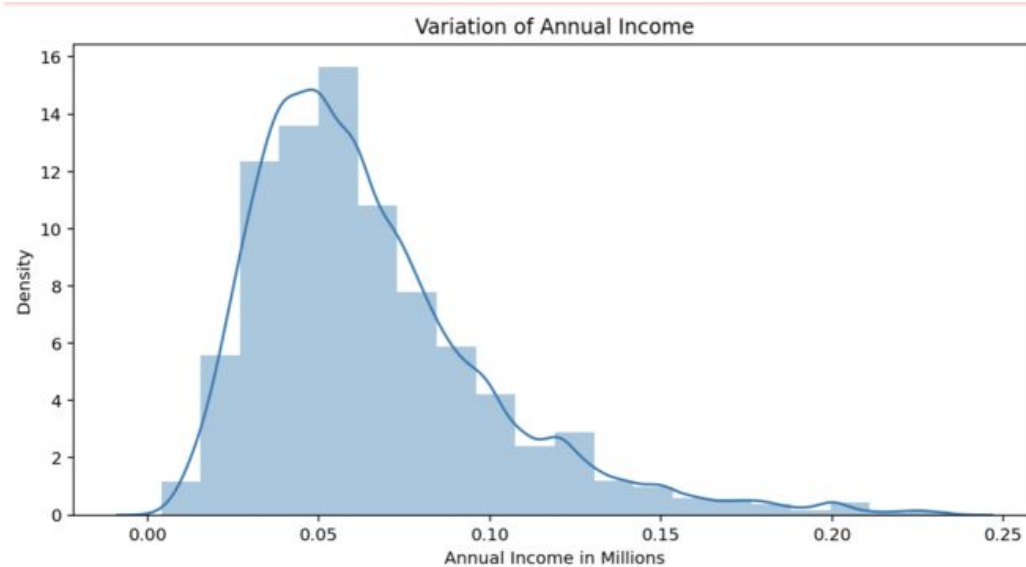
Univariate Analysis: Loan Amount
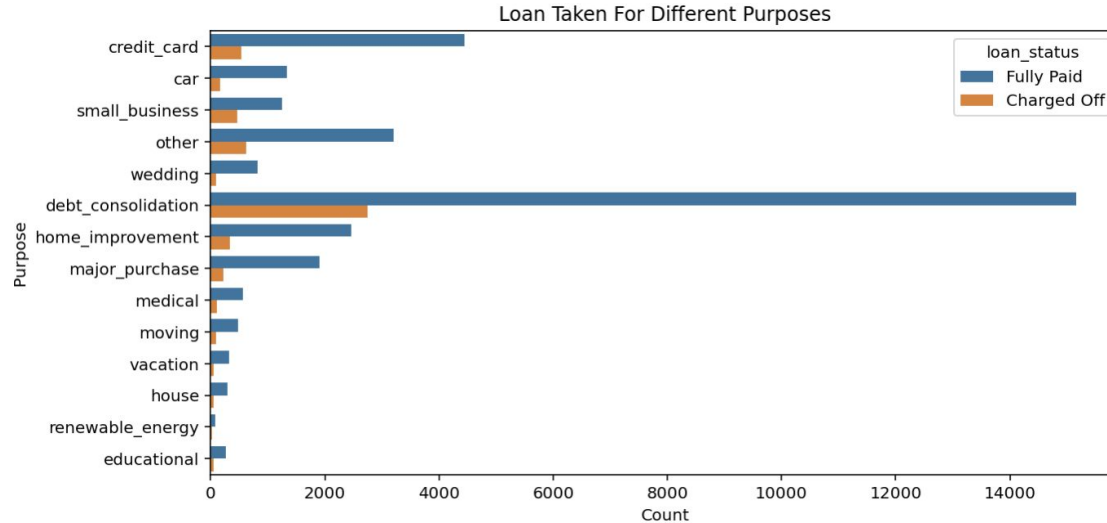
Distribution of Loan Amount

- The loan amount ranges between 500 and 35000, with an average of 9800.

Univariate Analysis: Annual Income

Variation of Annual Income

- Annual income has a skewed distribution, with most borrowers having lower income than the general population.

Segmented Univariate Analysis : Purpose

Loan Taken For Different Purposes

- Debt Consolidation stands as the most prevalent loan purpose, boasting the highest count of both fully paid and defaulted loans.

Segmented Univariate Analysis

- The mean and 25th percentile are identical for both, yet there's a notable increase in the 75th percentile for defaulted loans, suggesting that larger loan amounts have an elevated likelihood of default.
- The probability of defaulting is greater for the 60-month term compared to the 36-month term, while the 36-month term exhibits a higher likelihood of loans being fully paid.
- Borrowers who own their property experience fewer defaulted loans in comparison to those who are on a mortgage or renting.

Bivariate Analysis

Comparison of DTI Based On Grade For Loan Status
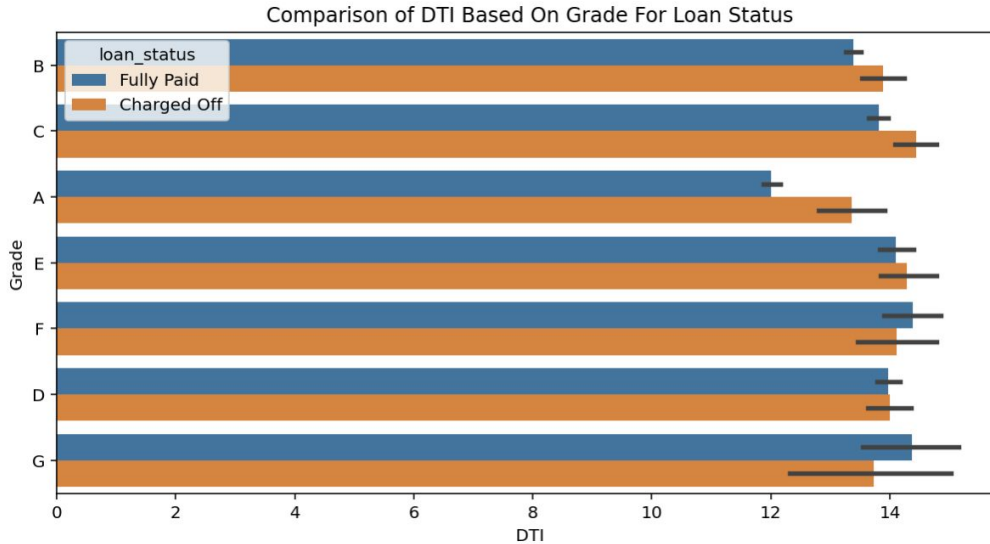
Grade A, representing the lowest risk, exhibits the lowest Debt-to-Income (DTI) ratio, suggesting a correlation where higher grades correspond to lower default rates.

Bivariate Analysis

- The majority of borrowers have no history of Public Recorded Bankruptcy, making them a secure option for loan issuance.
- The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan_amount groups.
- The grade serves as an indicator of risk, implying that the interest rate rises in tandem with the level of risk.