Lending Club Case Study

Group Members:
- Tirth Shah
- Shankar KV

Problem statement

The company wants to find out which factors cause borrowers to default on their loans. By identifying these factors using data analysis, the company can reduce loans to high-risk applicants and minimise financial losses. The company can utilize this knowledge for its portfolio and risk assessment.

Objectives

- ➢ Implement EDA techniques to analyse real-world loan data and gain actionable insights.
- ➢ Identify key factors contributing to loan defaults to reduce business losses.
- ➢ Develop a basic understanding of risk analytics in the banking and financial services industry.
- ➢ Present findings in a business-focused manner to support decision-making.

Approach

Data Overview	Data Clean Up	Find Variables of Interest	Data Visualization	Perform Analysis	Identifying Patterns and Anomalies

Data Overview: Review the dataset to understand its structure, types of variables, and basic statistics.

Data Cleanup: Remove the null fields, irrelevant fields, treat missing data, outliers.

Identify Variables of Interest: Identify key variables such as loan amount, interest rate, term, grade, and loan status.

Data Visualization: Create visualizations such as histograms, box plots, scatter plots, and bar charts to explore distributions, relationships, and trends within the data.

Perform Analysis: Compute correlation matrices and use heatmaps to identify relationships between numerical variables..
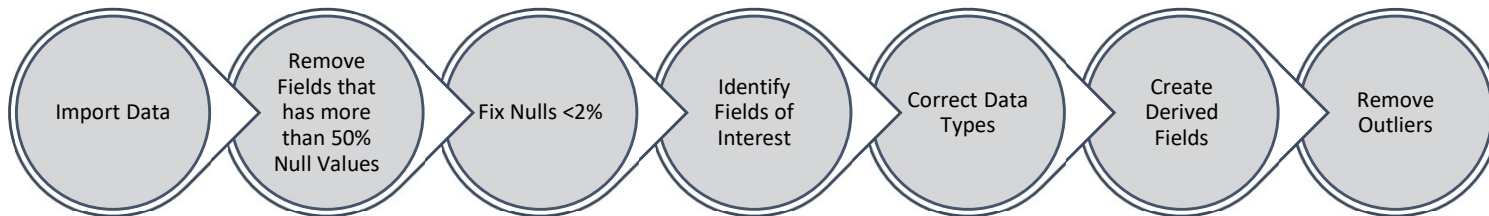
Identifying Patterns and Anomalies: Analyze patterns and detect anomalies or outliers in the data. This includes examining loan performance across different borrower segments (e.g., by grade or purpose) and identifying any unusual patterns or trends that could impact the analysis.

Data Clean Up

The data contains information about past loan applicants and whether they 'defaulted' or not
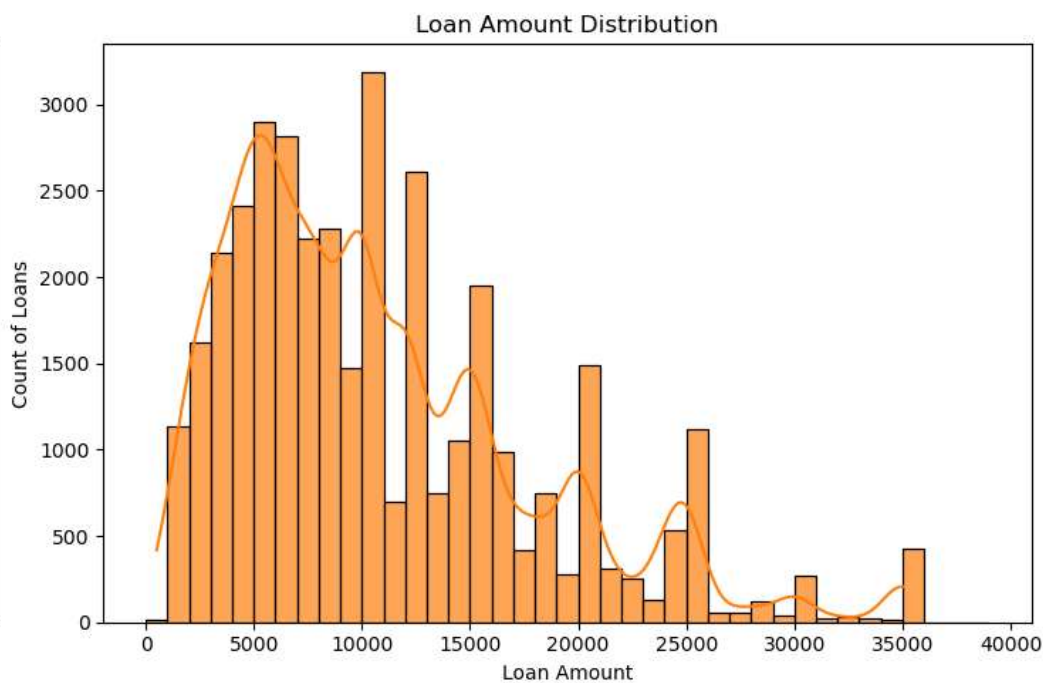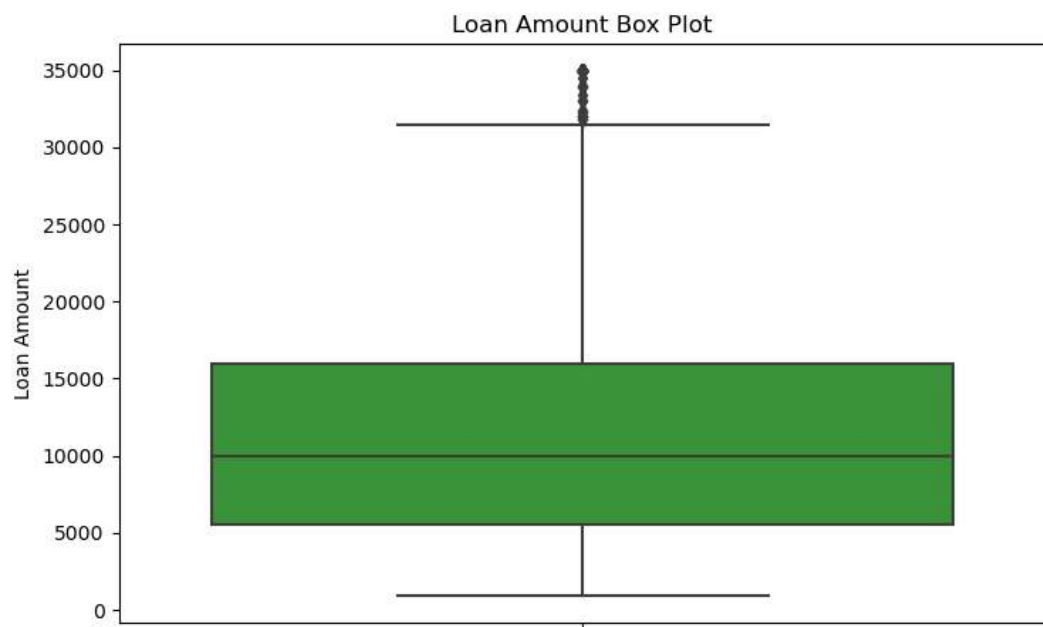
Loan data is for all loans issued through the time period 2007 to 2011.

These are the steps followed in Data Cleanup

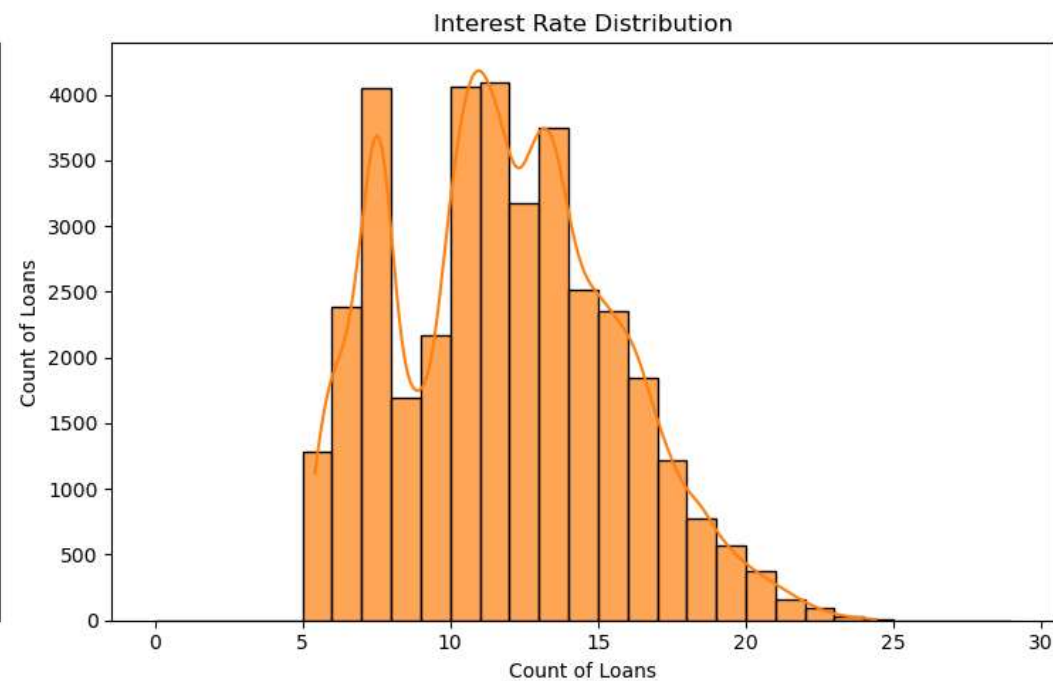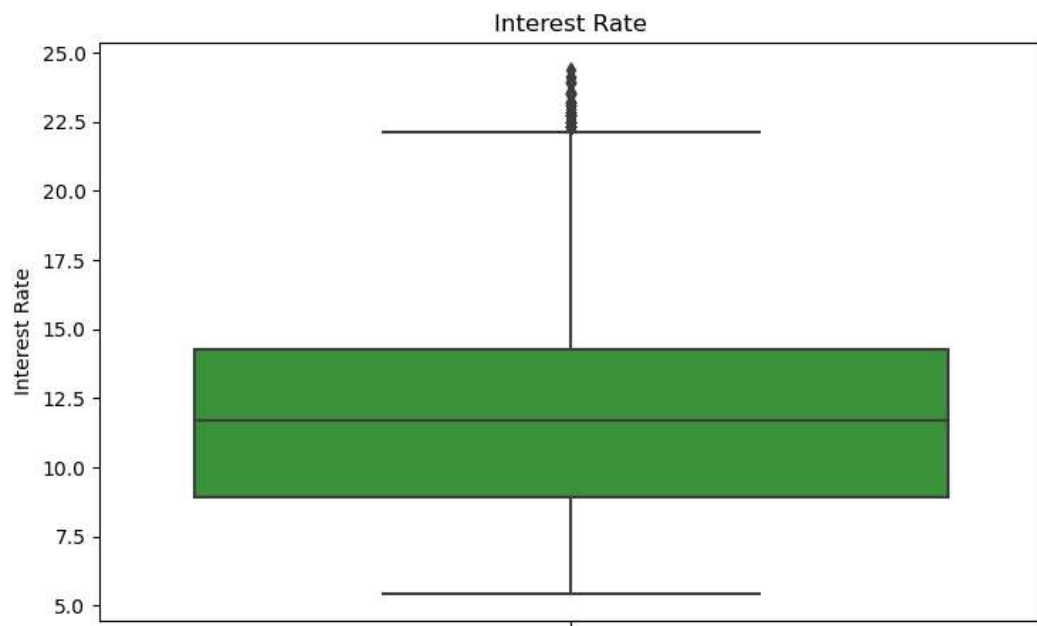Import Data → Remove Fields that has more than 50% Null Values → Fix Nulls <2% → Identify Fields of Interest → Correct Data Types → Create Derived Fields → Remove Outliers

EDA- Univariate

Loan Amount Distribution

Inference- Most of loans amount and funded amount is between 5k-15k, with highest range around 30-35K
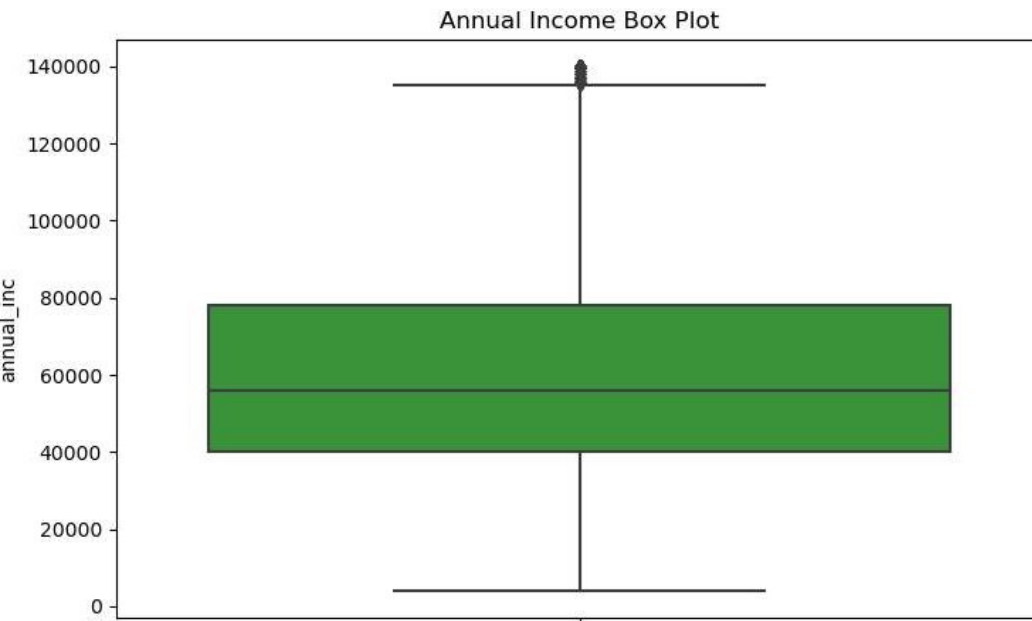
EDA- Univariate

Interest Rate Distribution

Inference: Majority of loans have Interest rate in the rage of 10-15%
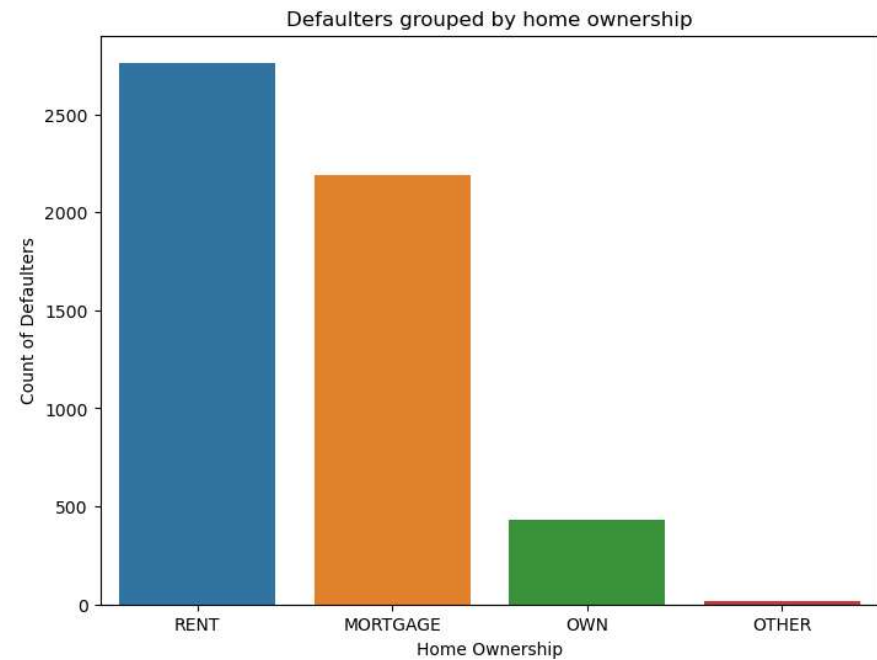
EDA- Univariate

Annual Income Distribution
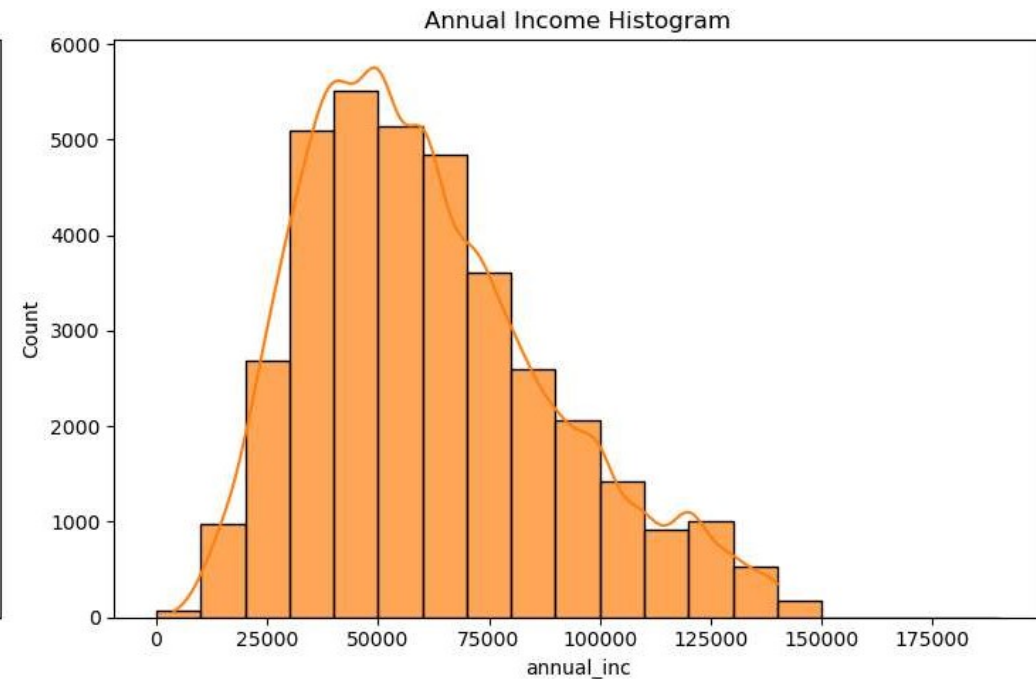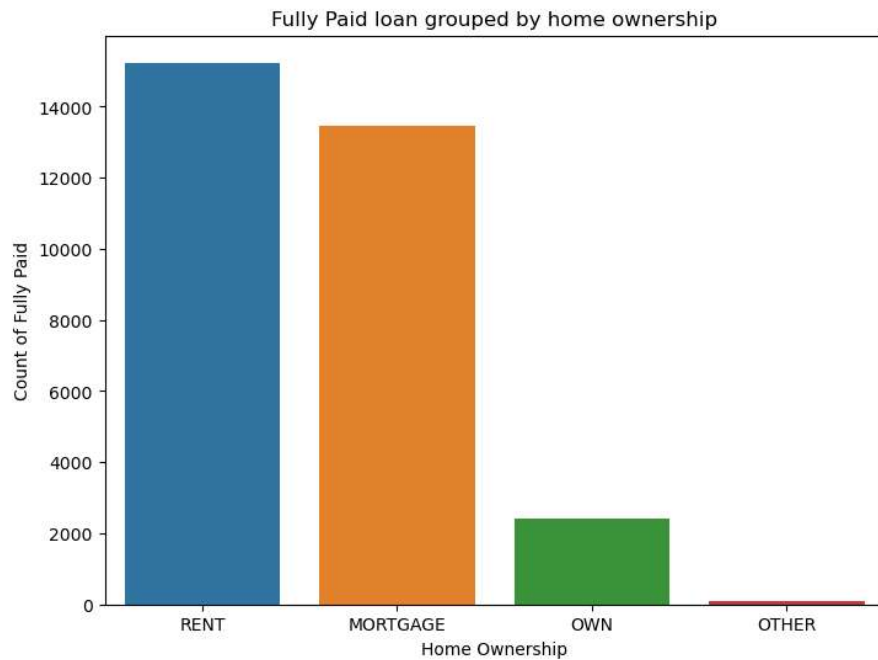

Inference: After removing outliers, it can be seen most of the borrower's annual income range between 40-80K
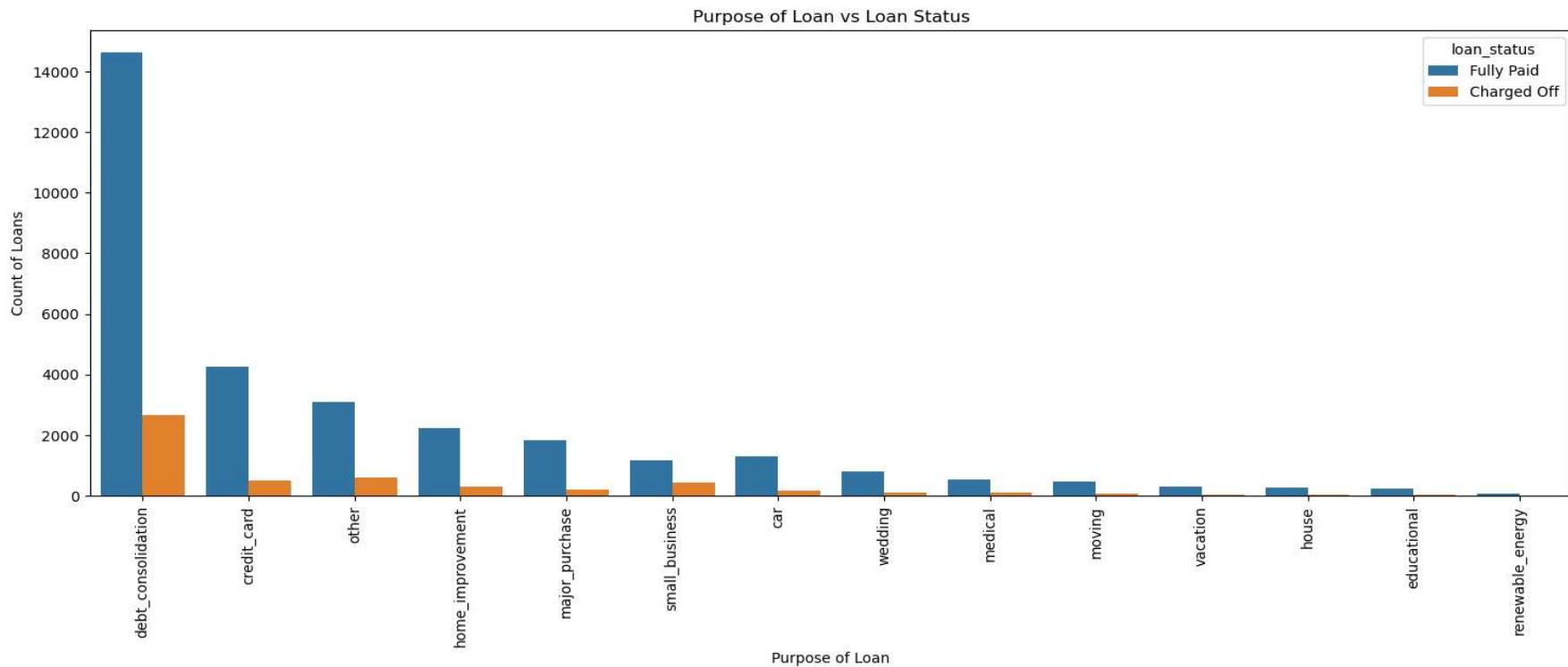
EDA- Segmented/Categorical Variable

Home Ownership

Inference: Very less borrowers have OWN house.
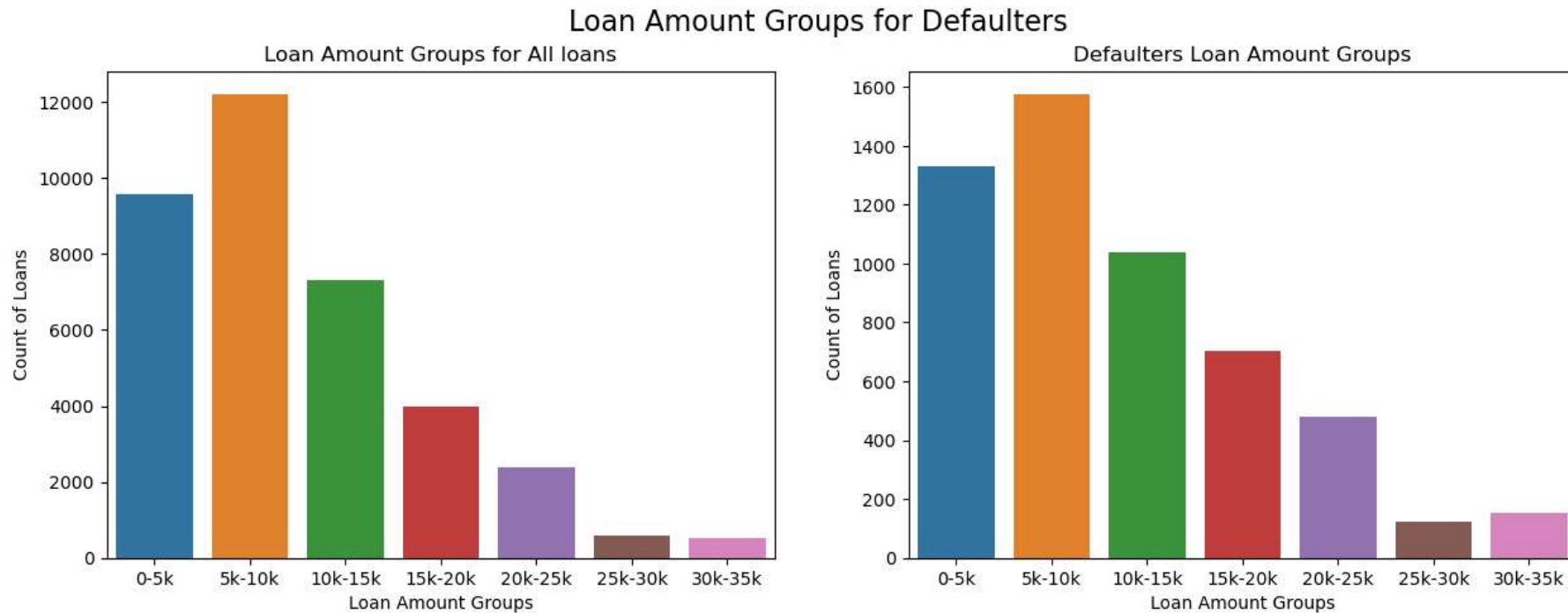
EDA- Segmented/Categorical Variable

Loan Purpose

Inference: Debt consolidation is one of the biggest reason for loans
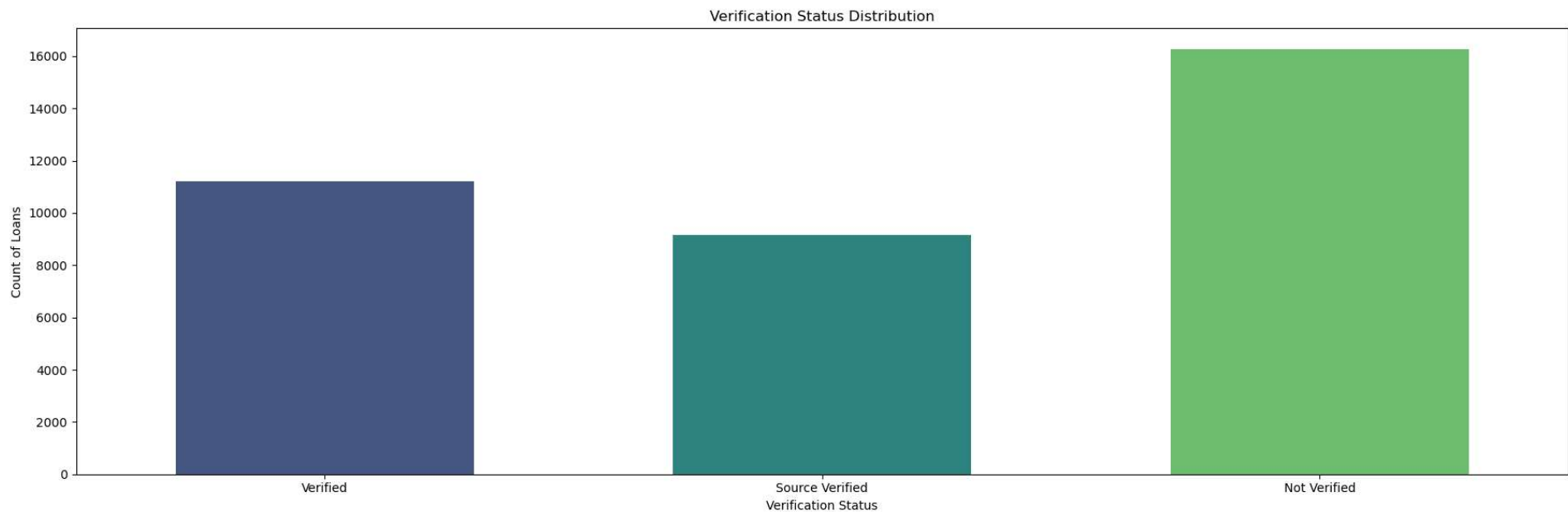
EDA- Segmented/Categorical Variable

Loan Amount Groups grouped by Loan status

Inference: Debt consolidation is one of the biggest reason for loans
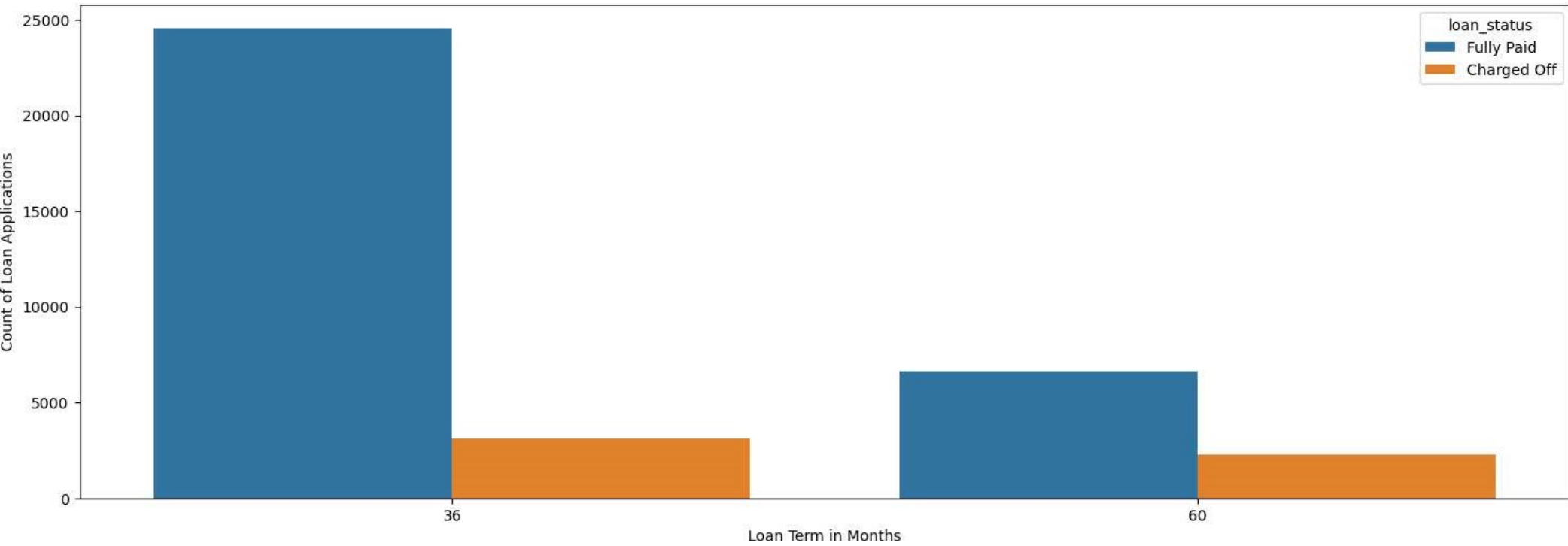
EDA- Segmented/Categorical Variable

Verification Status

Inference: 50% of Loans are not Verified
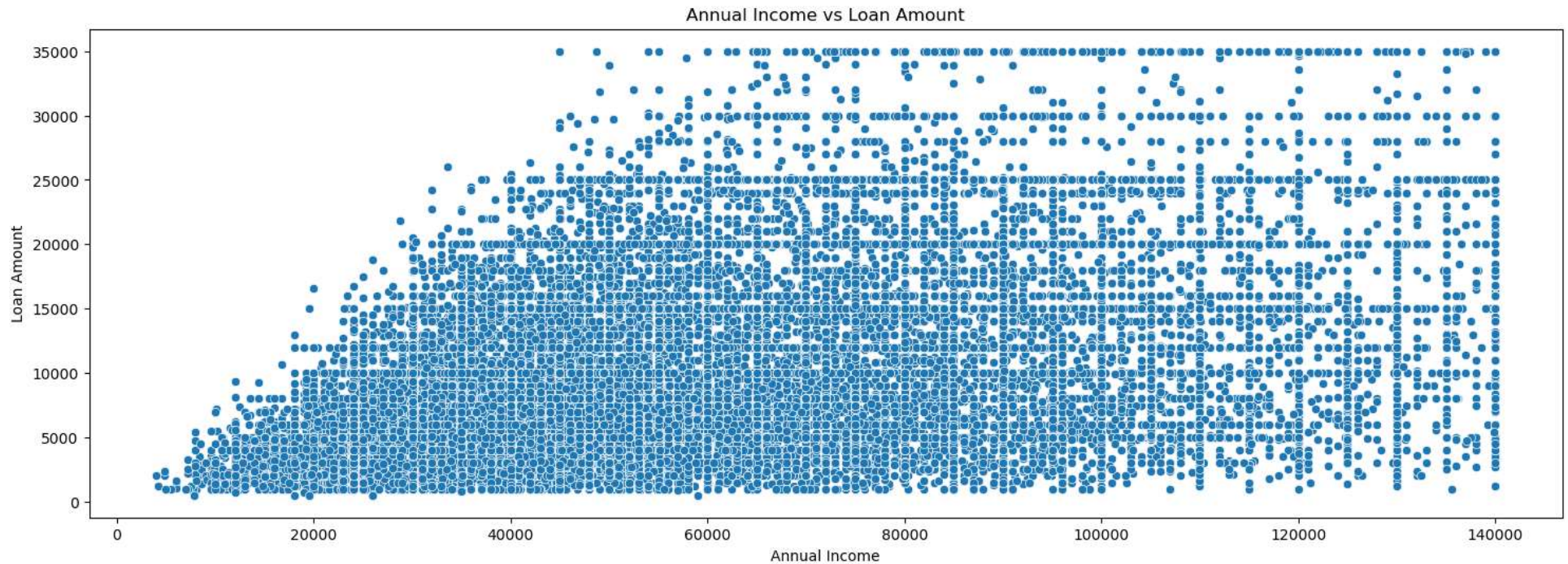
EDA- Segmented/Categorical Variable

Loan Term

Inference: High loan term increases chance of defaulting
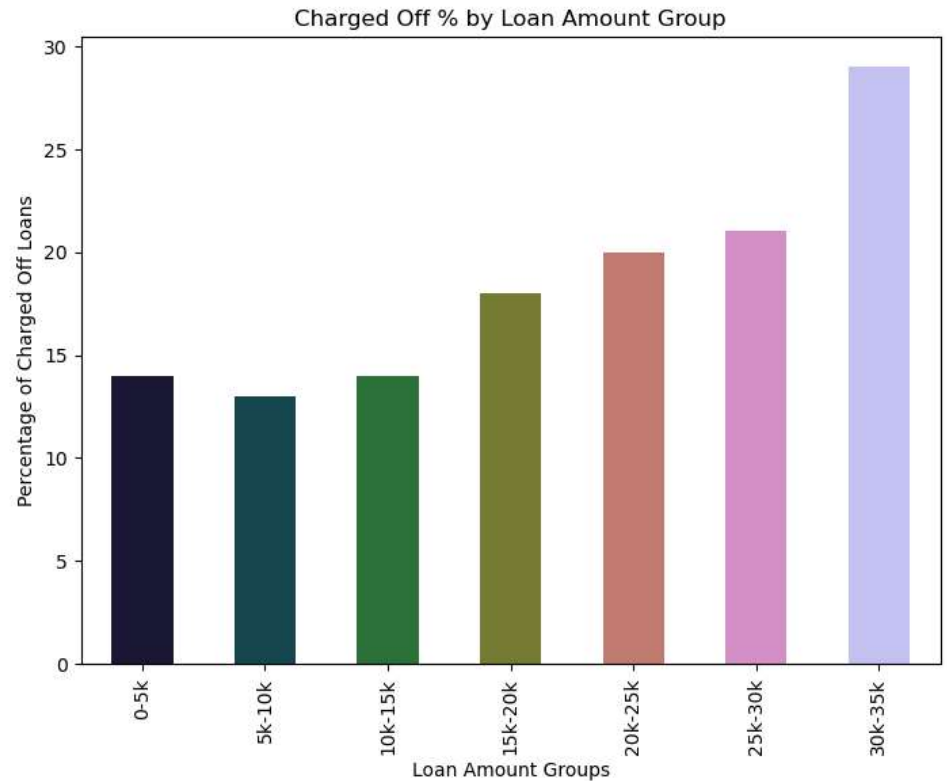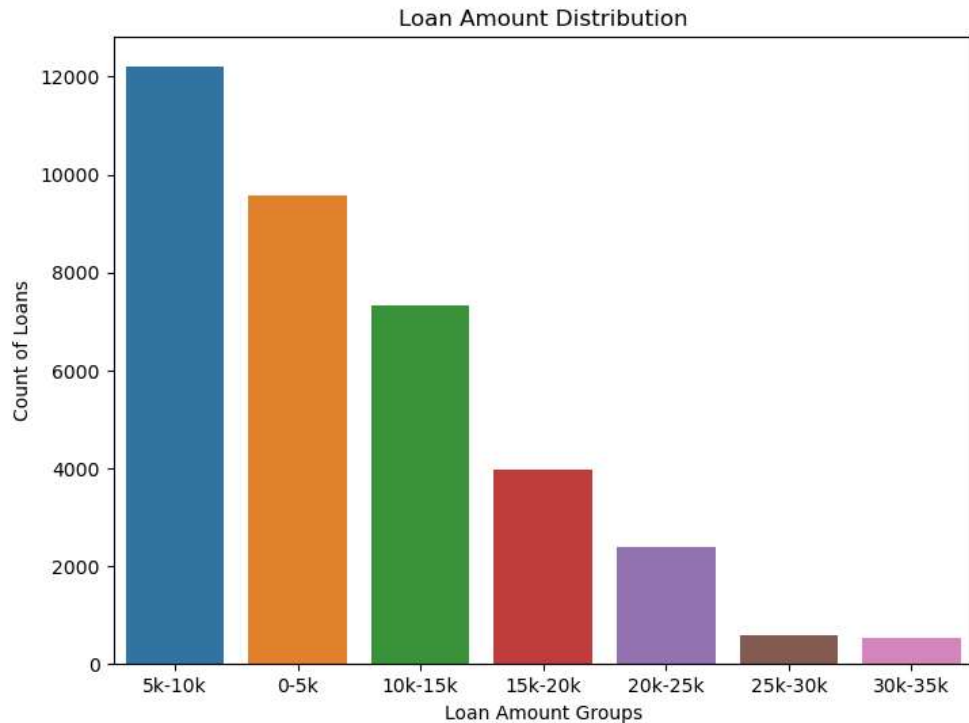
EDA- Bi-Variate Analysis

Annual Income Vs Loan Amount

Annual Income vs Loan Amount

Inference: Annual Income and Loan Amount are weakly correlated. Higher income means defaulting chances are less
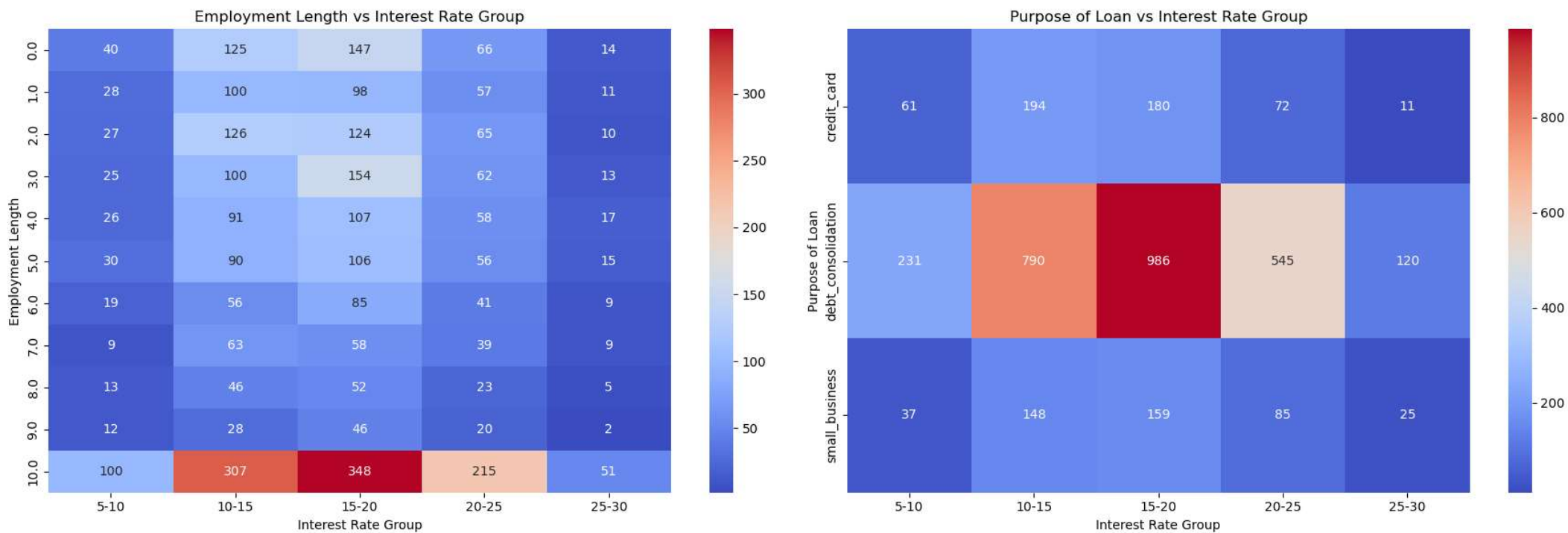
EDA- Bi-Variate Analysis

Finding Loan Amount groups and its Charged off Percentage

Inference: Loan Granted vs Charged off, when compared to Loan Granted, 30% of Loans are charged off in range 30-35K, when compared to overall charged off percentage 16%

EDA- Bi-Variate Analysis

Analyse Employment Duration, Loan Amount, Interest Rate, Purpose, Credit Grade of Defaulters, to identify a pattern.
Only Debt Consolidation, Small Business, Credit Card loan purposes were considered.
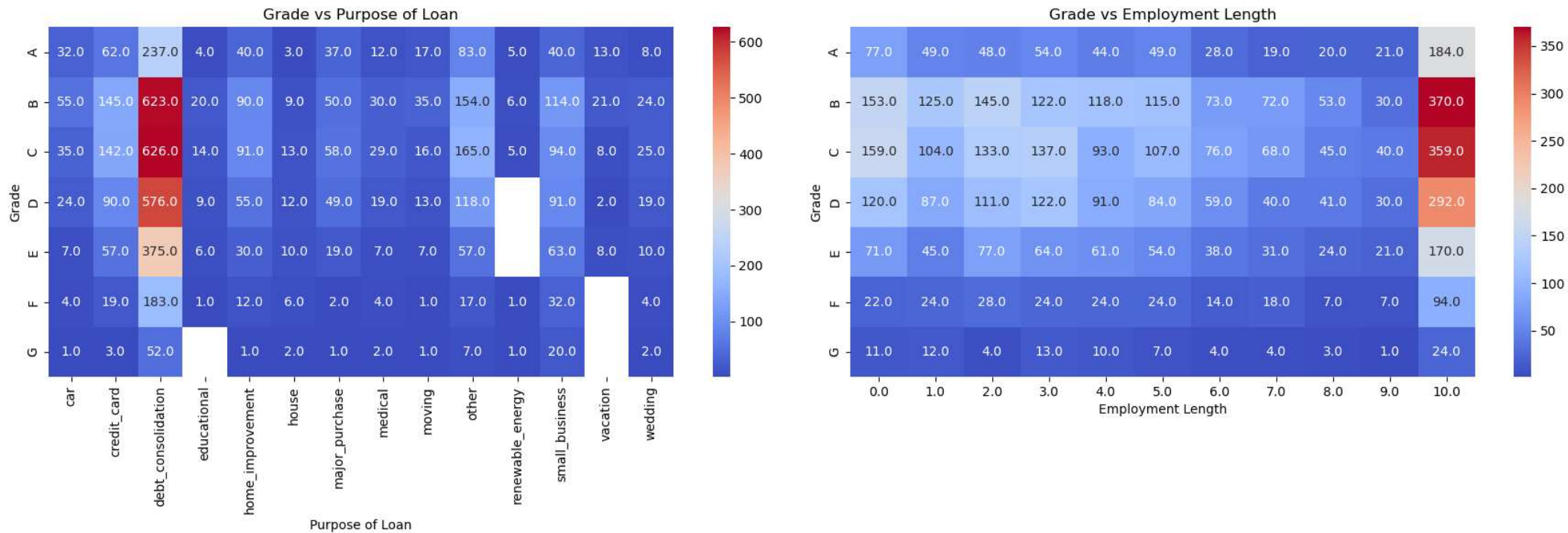

Inference : Debt Consolidation with employment length of 10+ years and interest rate group 10-15% has the highest number of defaulters. *Risky*
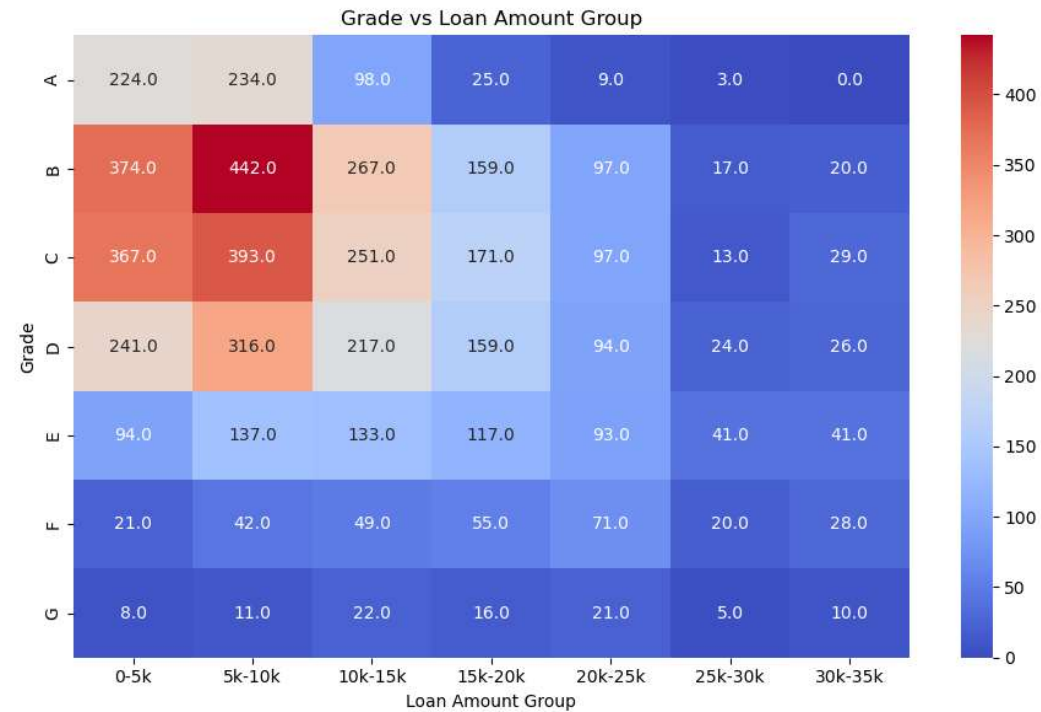
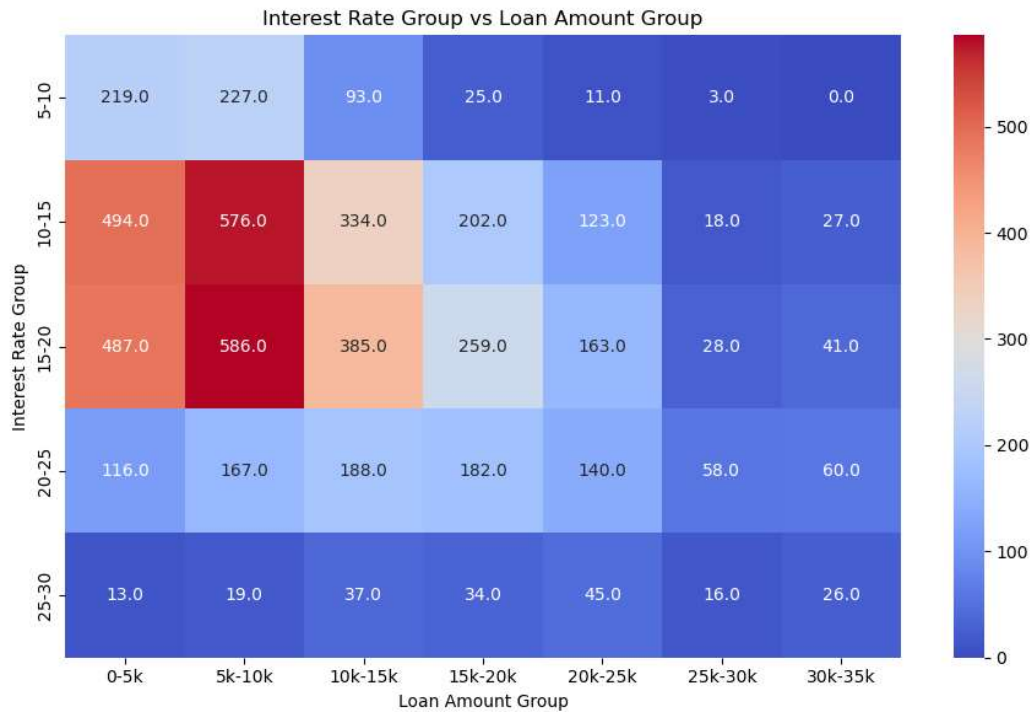# EDA- Bi-Variate Analysis

Analyse loan Grade vs Purpose of loan vs emp duration for defaulters

Inference: Debt Consolidation with Loan grade in B, C D with employment length of 10+ years are the major defaulters.
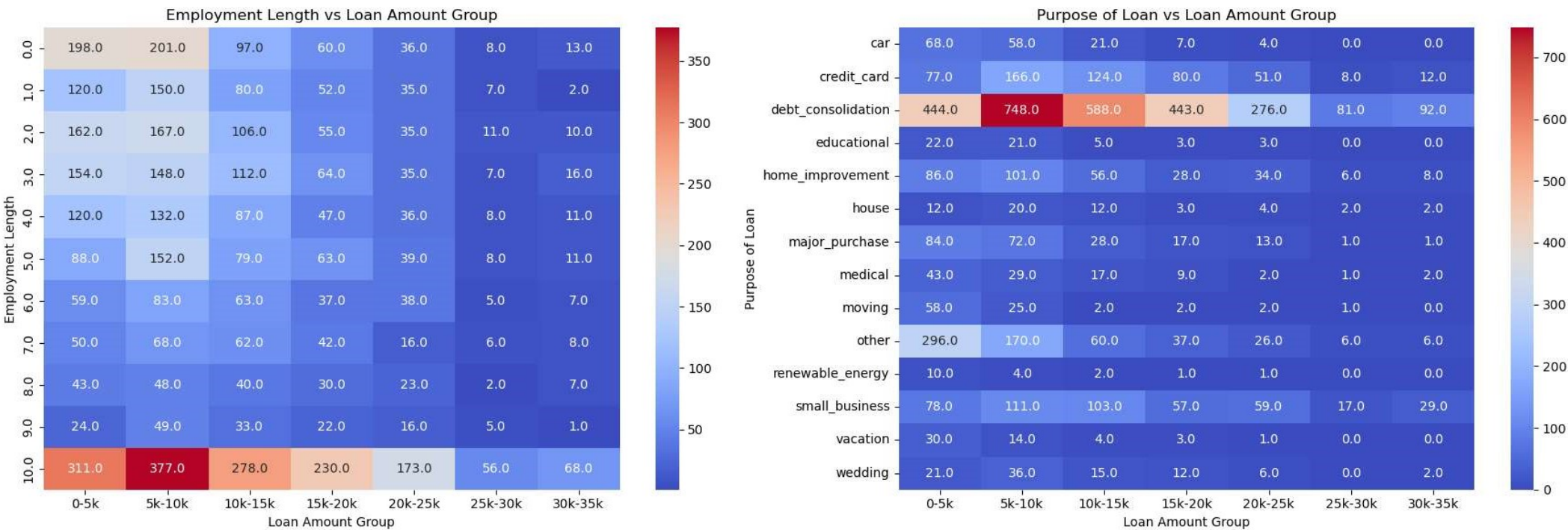
EDA- Bi-Variate Analysis

Analyse loan Grade vs Purpose of loan vs emp duration, Interest Rate

Continued to Next Slide…

EDA- Bi-Variate Analysis

Analyse loan Grade vs Purpose of loan vs emp duration, Interest Rate

Inference: 10+ Years Employment length with interest in the range of 10%-20%, with credit grade of B,C & D with loan value in the rage of 5k-20K, with loan purpose as Debt Consolidation are the riskiest loan

Conculsion

Goal : Understand the driving factors (or driver variables) behind loan default
- Annual Income
- Loan Amount
- Interest Rate
- Loan Purpose
- Employment Duration
- Loan/Credit Grade

Other Findings using EDA
- Loans with 10+ years employment, 10%-20% interest rates, credit grades B, C, D, 5K - 20K amounts, and debt consolidation purpose are riskiest.
- Defaulters' loans are 20%-35% of annual income, with higher median interest rates.
- Nearly 50% of high loan amounts get defaulted; high loan amounts and interest rates are risky.
- Most defaulters rent or have a mortgage, indicating financial burden may lead to defaults.