Lending Club Case Study

UpGrad & IIITB AI-ML Course - C70

Vyshak Srivatsa Raghavendra

OBJECTIVE

How to Minimize Defaulters

The objective of a Lending Club case study typically revolves around analyzing loan data to gain insights into borrower behavior and loan performance. Here are the common goals:

Understand Loan Performance:
- Analyse factors that influence loan defaults and successful repayments.
- Identify patterns in loan performance based on borrower characteristics (e.g., credit score, income, debt-to-income ratio).
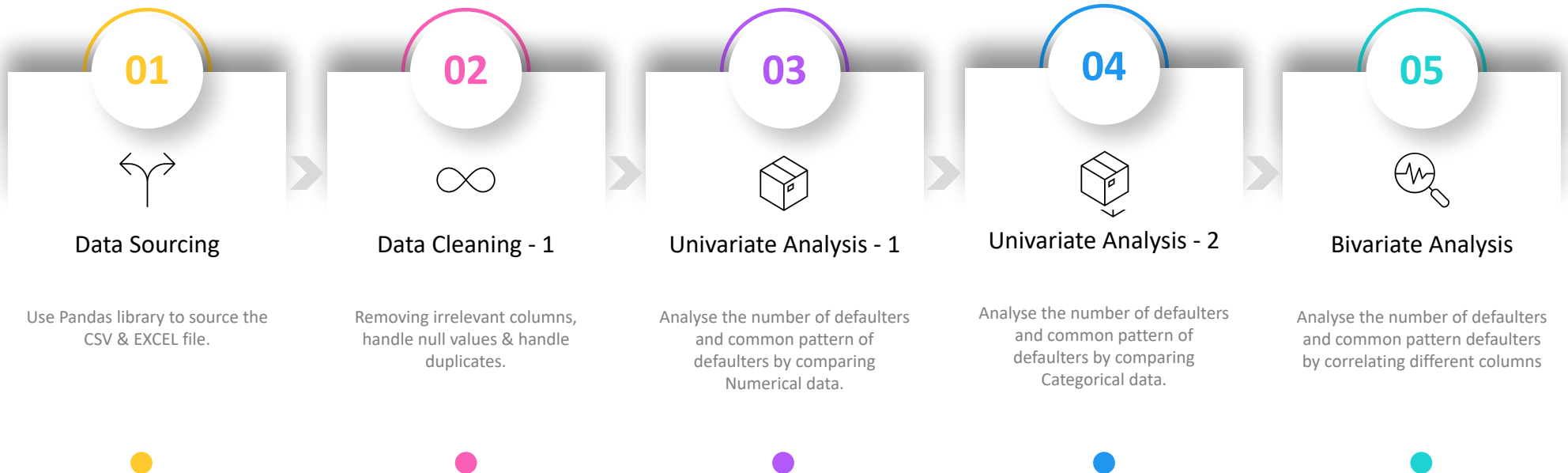
Risk Assessment:
- Develop a risk assessment model to evaluate the likelihood of default based on historical data.
- Provide recommendations for credit scoring improvements.
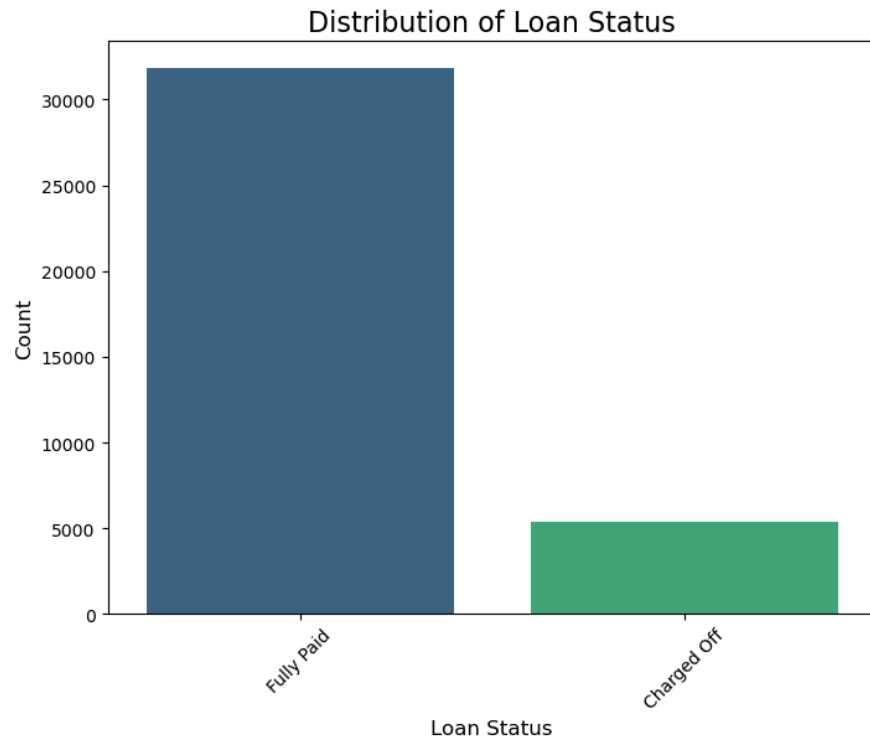
Visualization and Reporting:
- Produce clear visualizations that summarize findings and make them accessible to stakeholders.
- Prepare a detailed report or presentation that summarizes the methodology, findings, and actionable insights.

Methodology

01

Data Sourcing

Use Pandas library to source the CSV & EXCEL file.

02

Data Cleaning - 1

Removing irrelevant columns, handle null values & handle duplicates.

03

Univariate Analysis - 1

Analyse the number of defaulters and common pattern of defaulters by comparing Numerical data.

04

Univariate Analysis - 2

Analyse the number of defaulters and common pattern of defaulters by comparing Categorical data.

05

Bivariate Analysis

Analyse the number of defaulters and common pattern defaulters by correlating different columns

Data Analysis
On Overall Data

Distribution of Loan Status

Around 14% of borrowers are becoming defaulters

Univariate Analysis
On Numerical Data
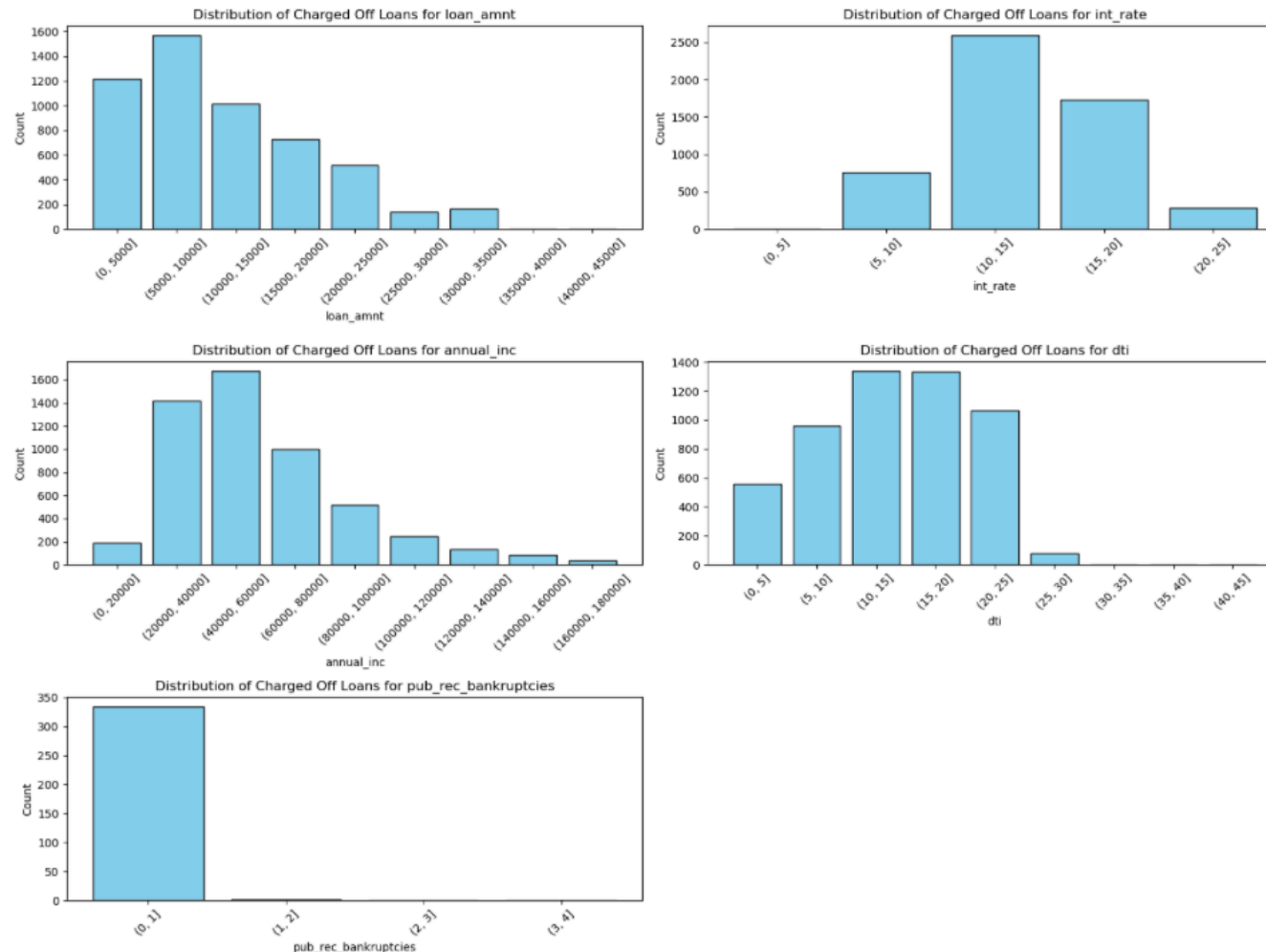
- Majority of loan sanctioned amount is crowded at 5000 likely represent lower middle class people

- Majority of the people who applied loan have very large debt compared to the income registered, concentrated in the 10-15 DTI ratio

- The interest rate is more crowded around 5-10 and 10-15 with a drop near 10.

- Majority of borrowers loan have annual income around 50000

- Majority of the people who applied loan have no record of Public Recorded Bankruptcy.
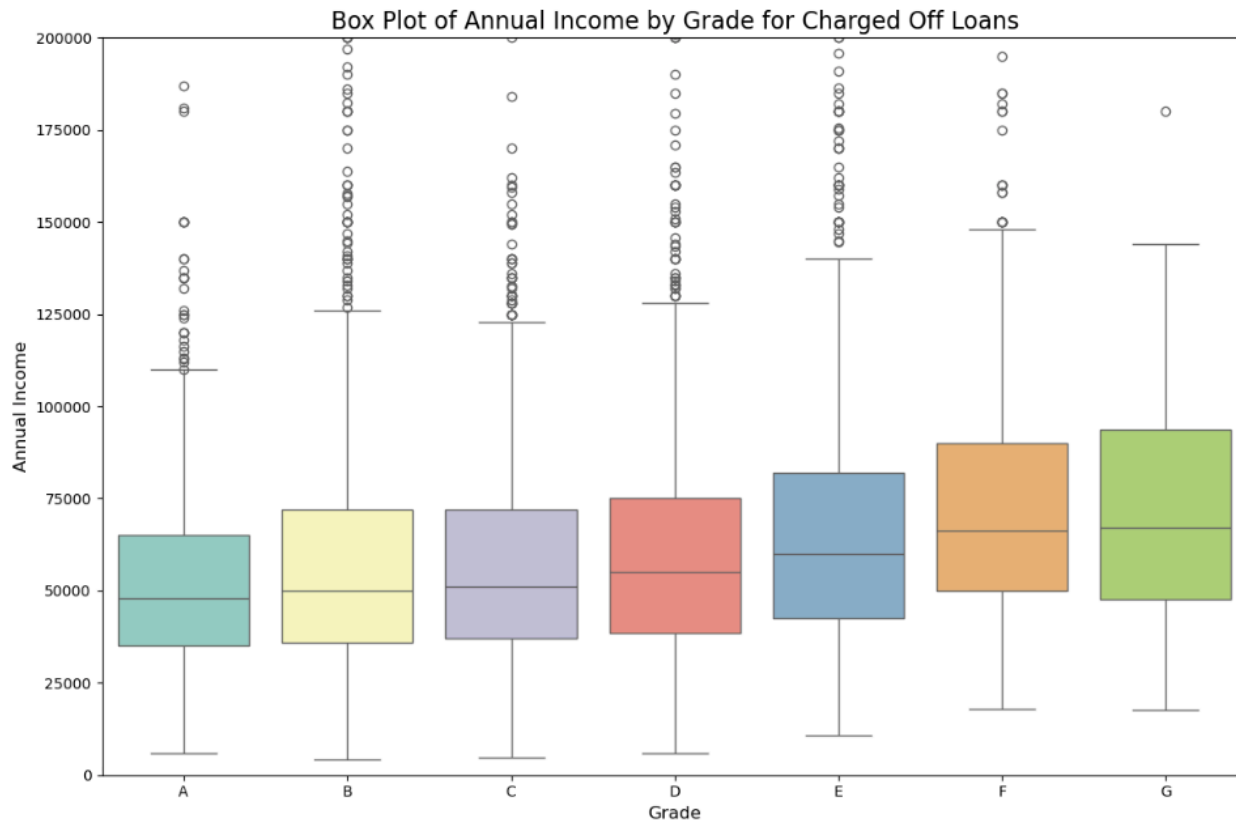
Univariate Analysis
On Categorical Data
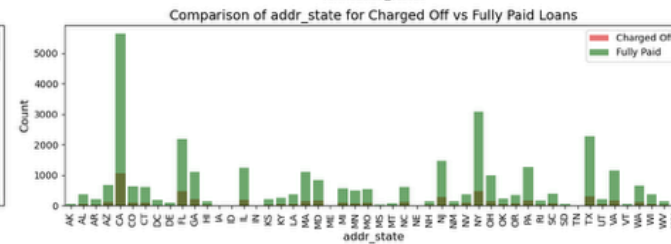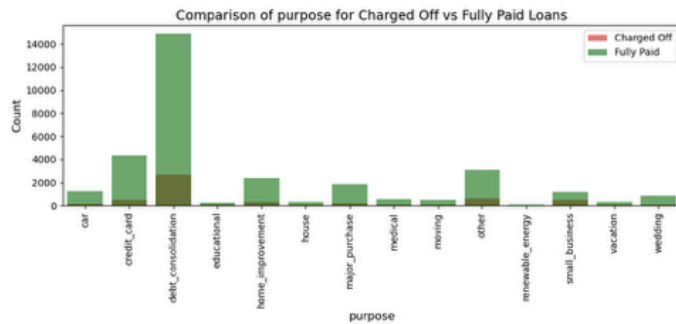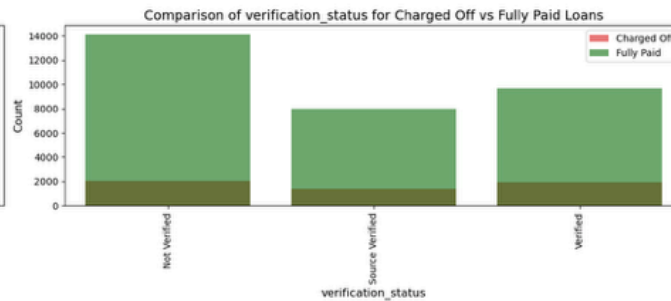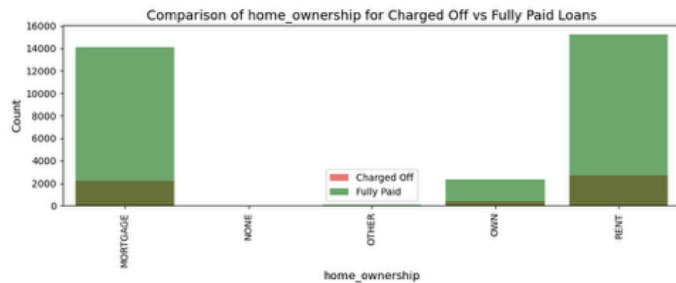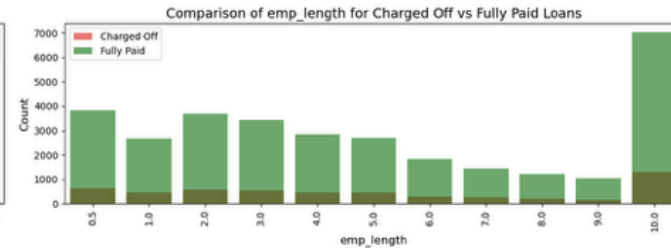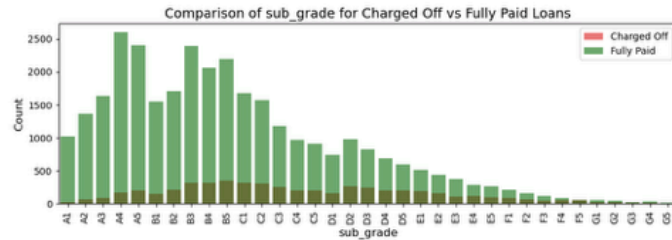
- large amount of loans defaulted are with grade 'B' and 'C' compared to rest.

- Majority of the people who applied loan have working experience greater than 10 years.

- Majority of the people who applied loan are on mortgage or rent.

- Approx 50% of the borrowers are verified by the company or have source verified.

- Large percentage of loans are taken for debt consolidation followed by credit card.

Bivariate Analysis
Focused On Defaulted Data

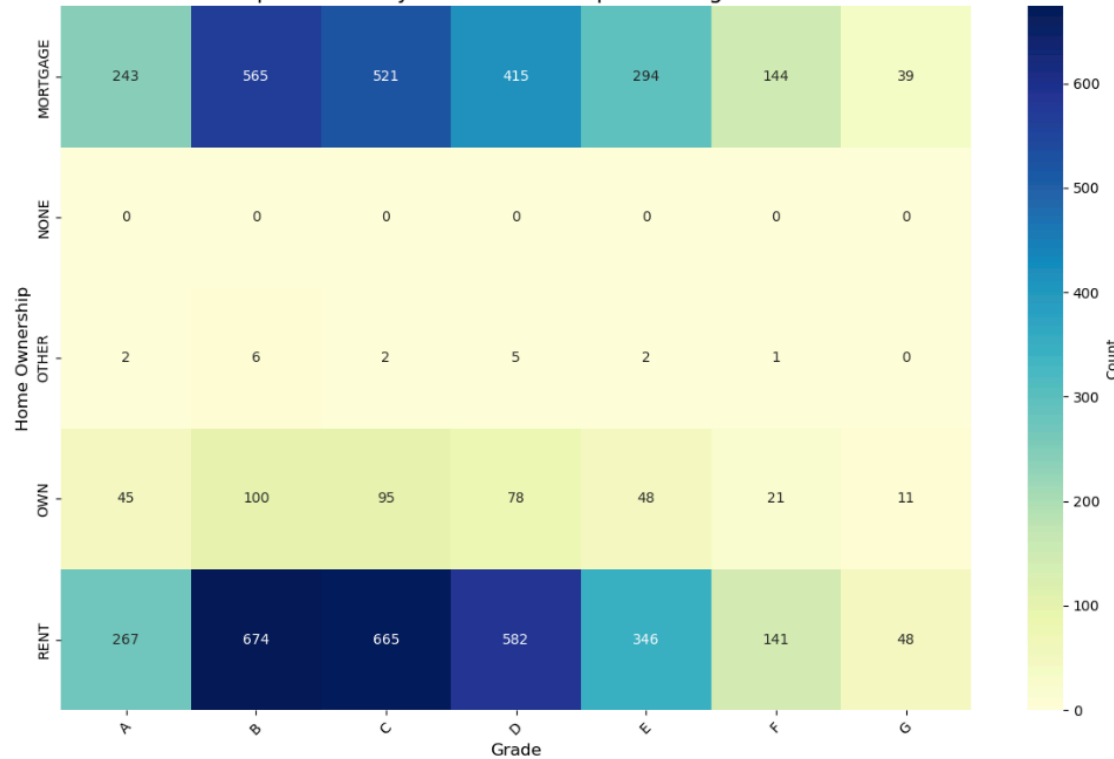Box Plot of Annual Income by Grade for Charged Off Loans

Majority of Defaulters have average annual income ~50000 and belongs to Grade B or C
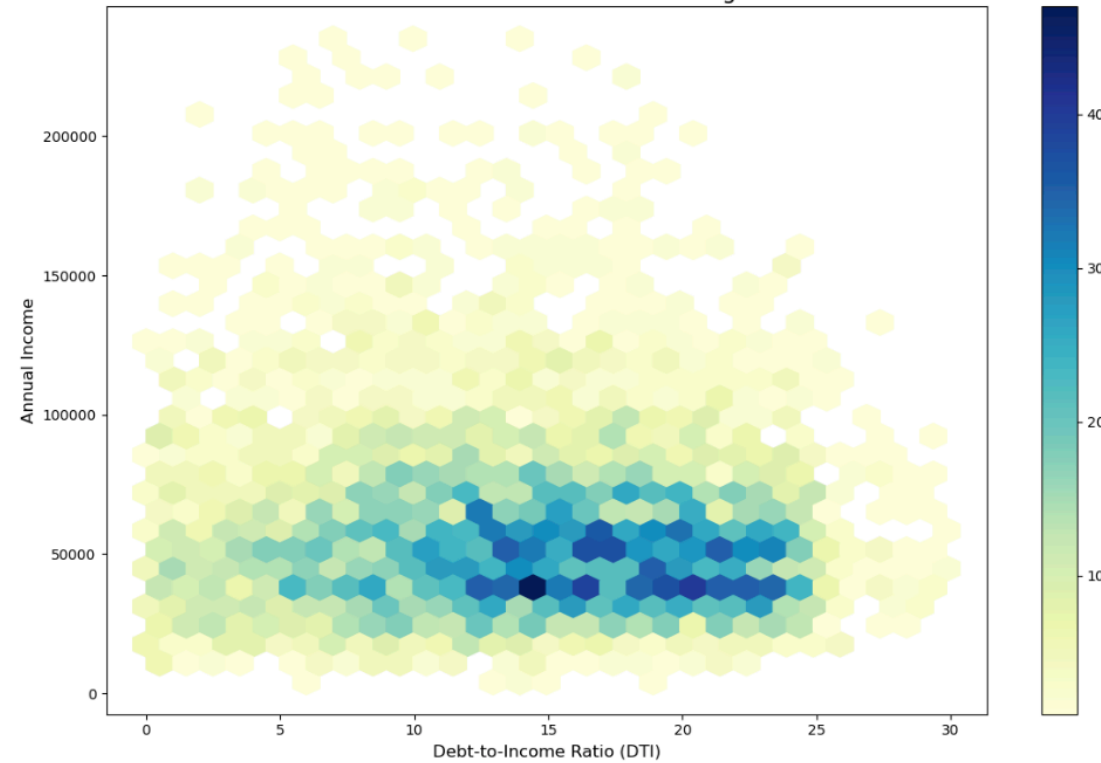
Bivariate Analysis
Focused On Defaulted Data

Heatmap of Grade by Home Ownership for Charged Off Loans

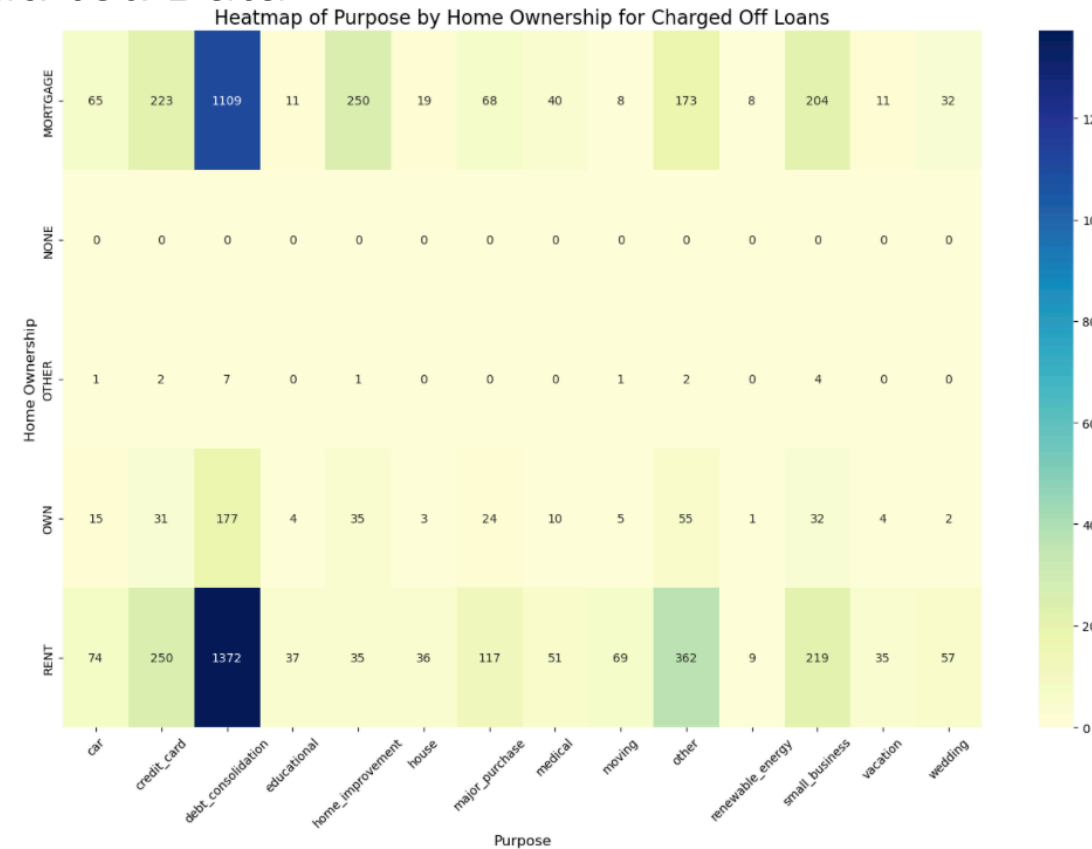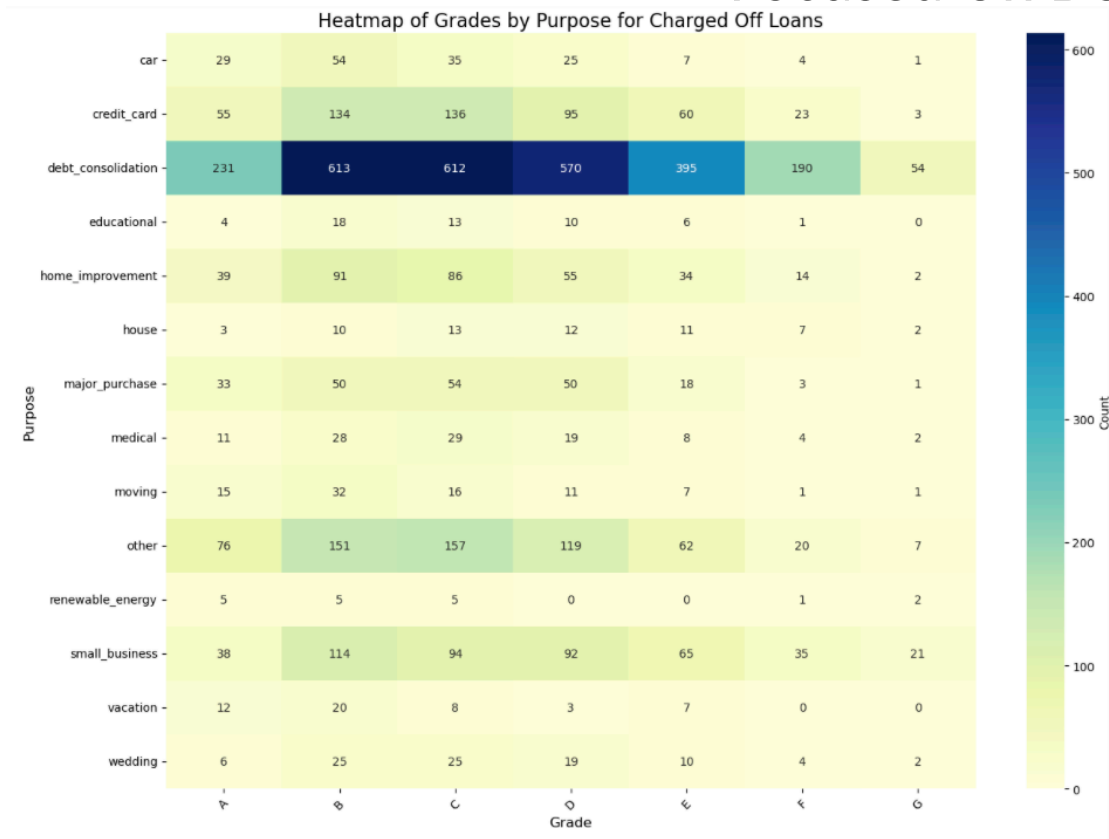Hexbin Plot of DTI vs Annual Income for Charged Off Loans

Majority of Defaulters live in Rented/Mortgage house and show DTI 10 to 20. Rent payment adds additional burden

Bivariate Analysis
Focused On Defaulted Data

Heatmap of Grades by Purpose for Charged Off Loans

Heatmap of Purpose by Home Ownership for Charged Off Loans

Majority of Defaulters belongs to Grade B or C
and purpose of loan is debt consolidation

Data Analysis
Conclusion

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

1. DTI

2. Borrowers having annual income in the less than 50000

3. Borrowers with grades like b,c which indicates high risk.

4. Borrowers with very high debt to Income value.

5. Verification Status

6. If purpose is debt consolidation for people with income < 50000

Create Red Bucket with above combination and Lending company should reject loans who fall in that category or perform stringent analysis before approving

Thank You