Lending Club Case Study

Exploratory Data Analysis

- Bindu Madhav Dadi

Table of Contents

Problem Statement

Data Cleaning

Univariate Analysis

Segmented Univariate Analysis

Bivariate Analysis

Conclusion

Problem Statement

As the largest online loan marketplace, the company provides loans for personal, business, and medical purposes.

Significant financial losses (credit losses) arise from lending to risky applicants.

The company seeks to identify the primary factors that lead to loan defaults to mitigate these credit losses.

Data Cleaning

- **Dropping Columns with All Empty Values**
  - Identify and remove columns that contain no data at all.
  - This helps streamline the dataset by eliminating unnecessary columns, making the analysis more efficient.

- **Dropping Columns that are Insignificant or Have Maximum Null Data**
  - Evaluate columns for their relevance to the problem statement.
  - Remove columns that have minimal impact on loan default predictions or contain a high percentage of missing values.
  - This helps in reducing noise and focusing on meaningful variables.

- **Dropping Columns with Unique Single Values**
  - Identify columns where every entry is the same (e.g., a column where all values are 'N/A' or 'Unknown').
  - Such columns do not provide any useful information for the analysis and can be safely removed.

- **Dropping Rows that are Not Required for the Problem Statement**
  - Filter out rows that do not contribute to the analysis of loan defaults (e.g., loans that are still active and have not yet defaulted or been fully paid off).
  - This ensures that the dataset is focused on the relevant observations for understanding loan defaults.

Univariate Analysis

Observations

- 75% of the loan amount requested is within 15000.

- About 75% of the loan is requested/granted for the term of 36 months.

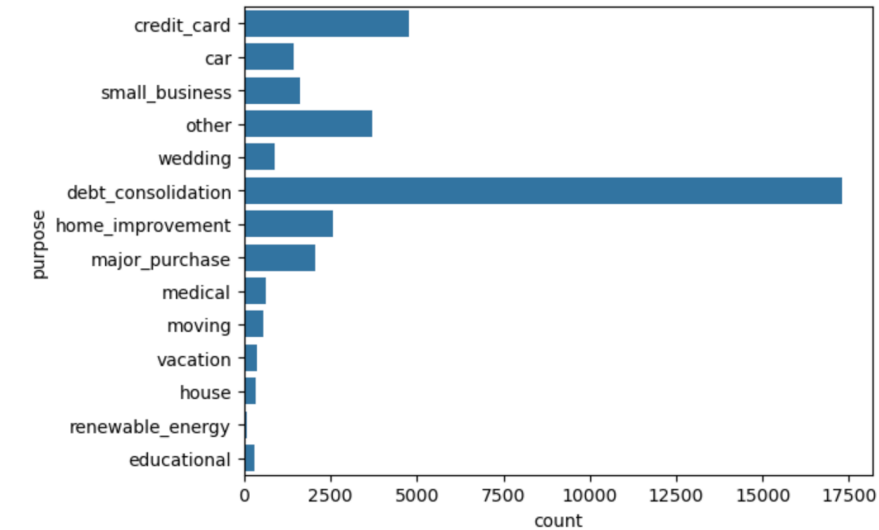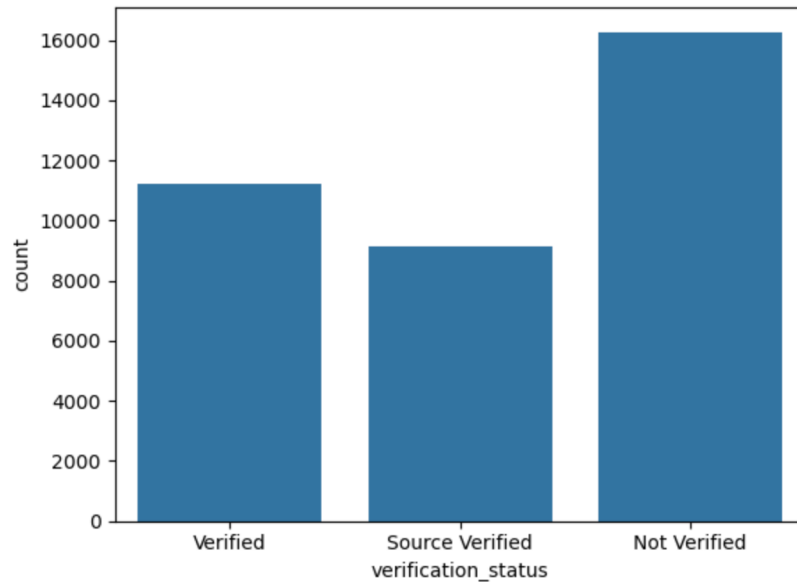- Majority of the loan applicants have employment length of 10+ years
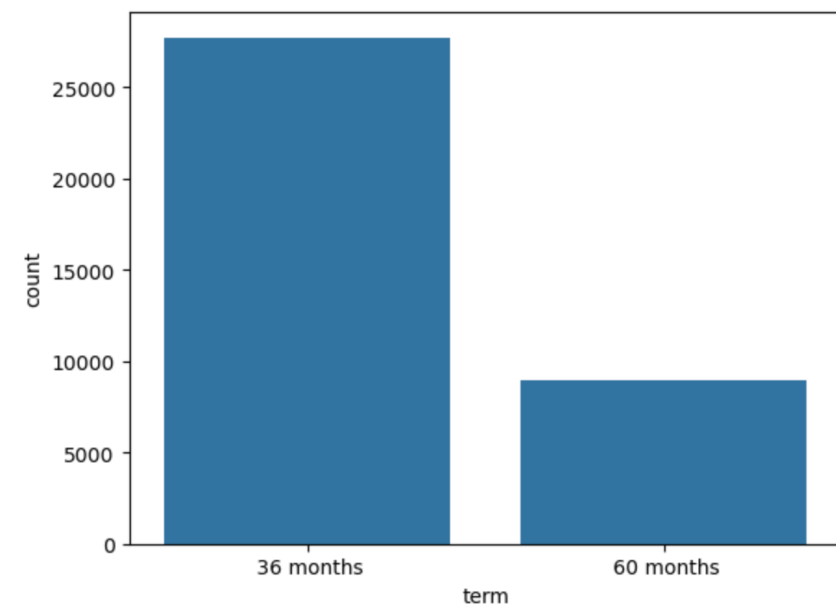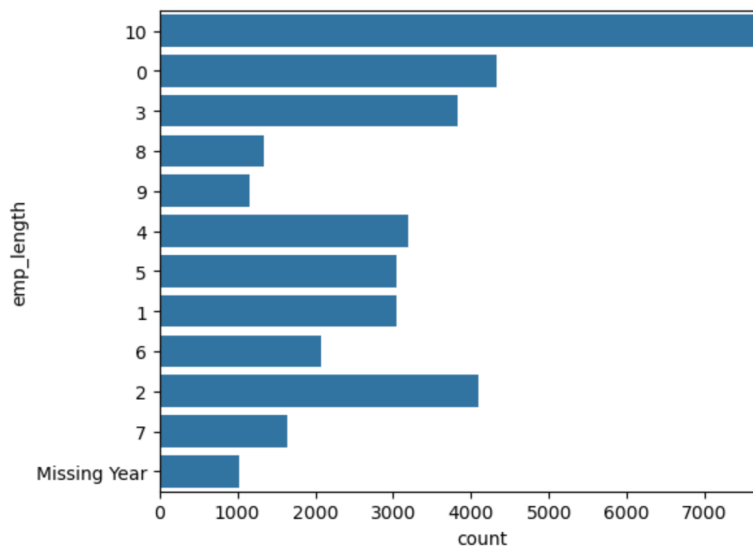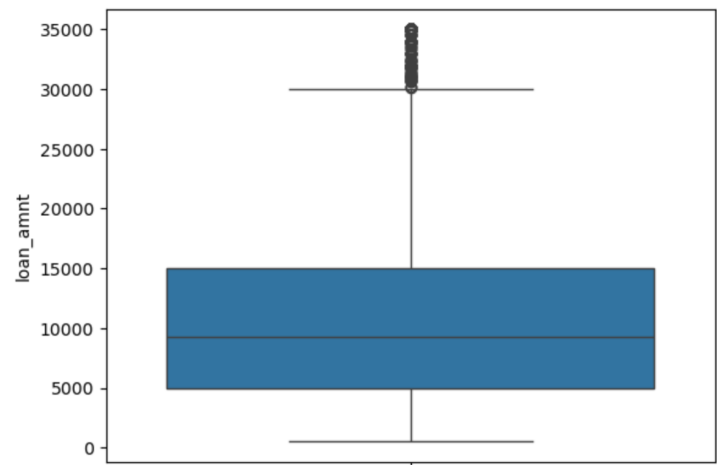
Univariate Analysis
(Home Ownership, Verification Status, Purpose)

Observations

- Approximately 48% of loan applicants have "RENT" as their homeownership status, while 44% have "MORTGAGE".

- About 43% of loan applicants had their income unverified.

Segmented Univariate Analysis

<mark>(Funded Amount, Loan Term, Interest Rate)</mark>

Observations

- Applicants granted higher funded amounts tend to default more frequently than those who received lesser amounts.

- Applicants who opted for a higher term (60 months) tend to default more often than those who chose the 36-month term. This is significant, considering that 75% of the loans were requested/granted for 36 months based on the univariate analysis of the term.

- Applicants with higher interest rates tend to default more frequently than those with lower interest rates.
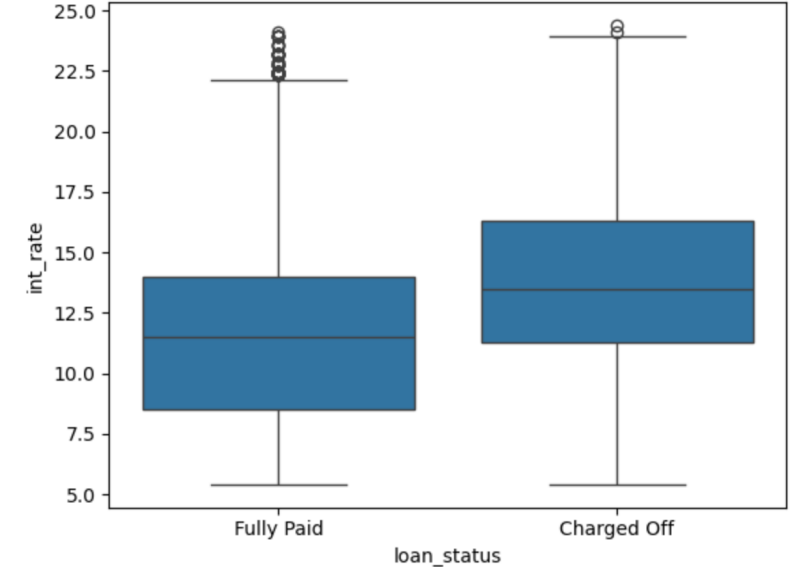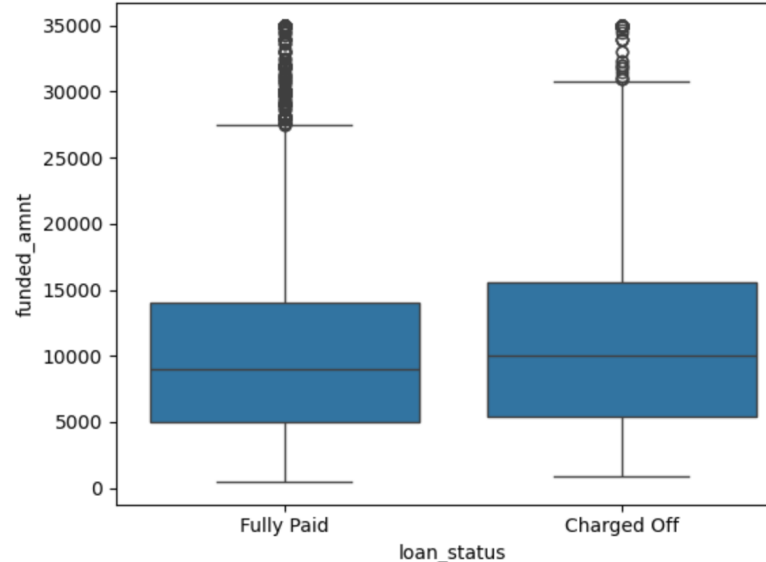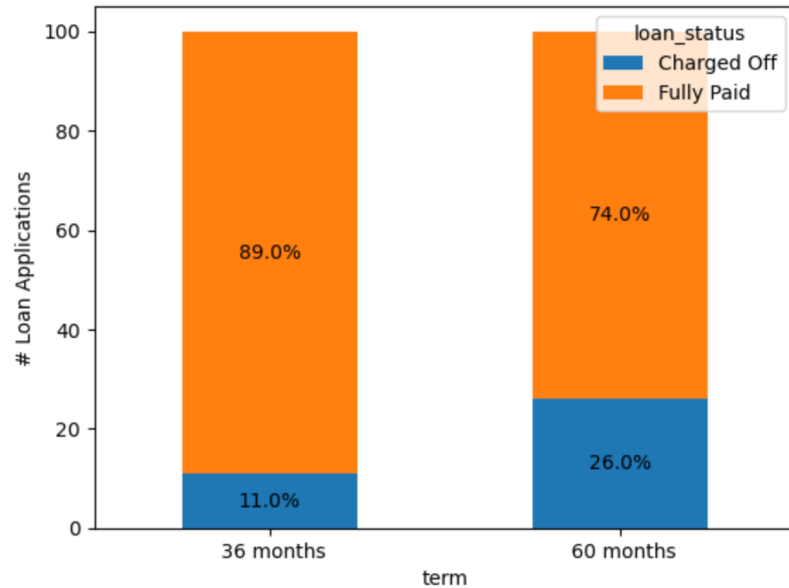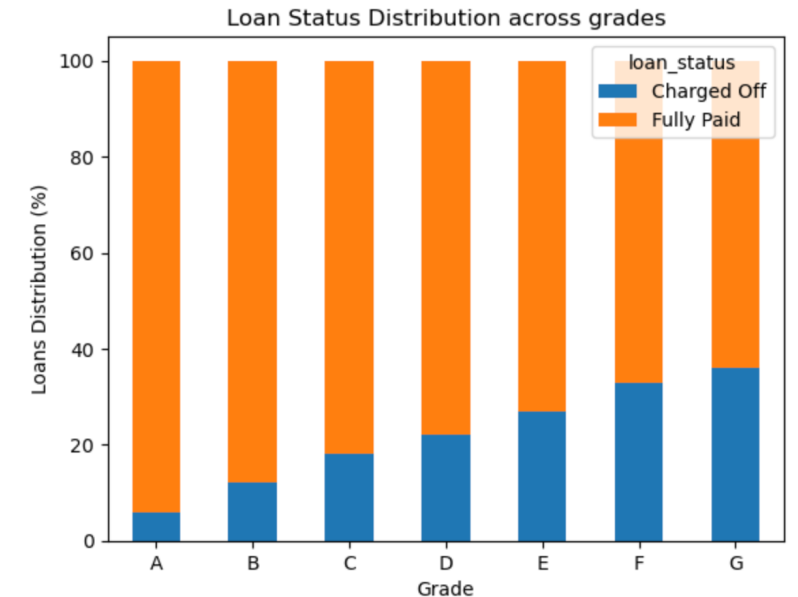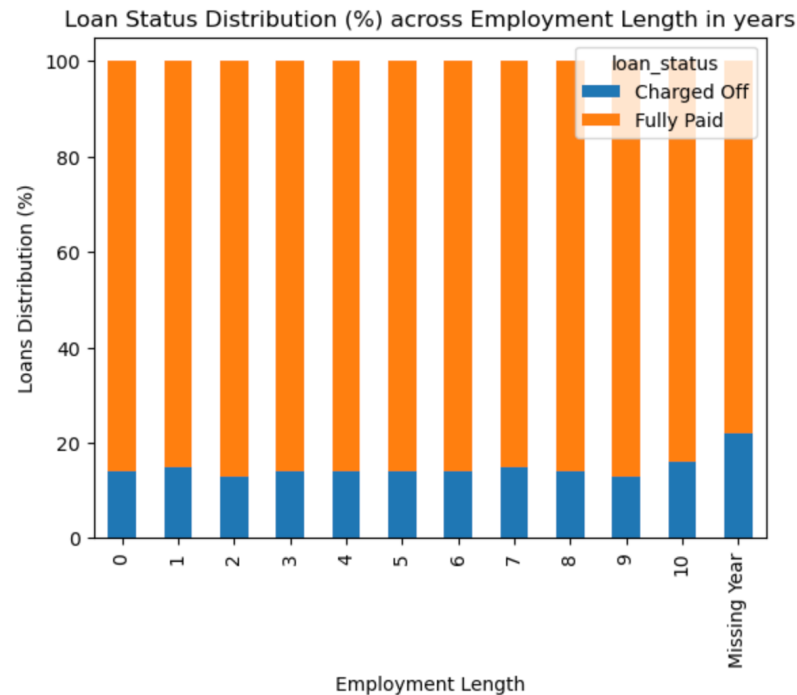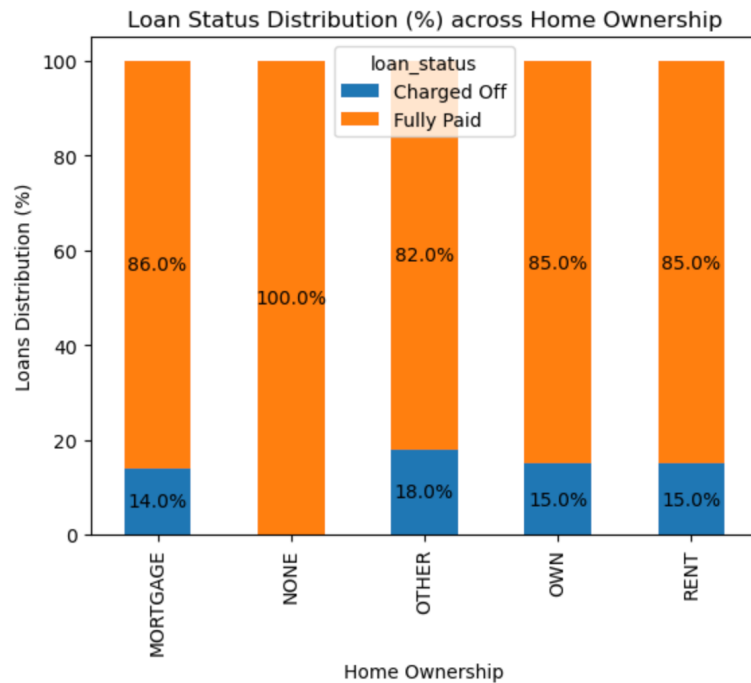
Segmented Univariate Analysis

(Grade, employee length, Home Ownership)

- With grades and subgrades, where higher alphanumeric order grades and subgrades have higher default rates. Business-wise, grades and subgrades are correlated with interest rates, supporting this observation.

- Applicants with an employment length of 10+ years tend to default more frequently than those with shorter employment durations.

- Applicants whose employment length is missing have a higher default rate (22%) than all other groups. This is notable, given that 10+ years is the most common employment length in the distribution based on the univariate analysis.

- Applicants with homeownership status marked as "OTHER" are more likely to default compared to those with "RENT" or "MORTGAGE," despite 92% of applicants falling into the latter two categories.

- Interestingly, there are no defaulters in the "NONE" homeownership category.

Segmented Univariate Analysis

<mark>(Verification status, purpose, Credit revolving)</mark>

- Surprisingly, loan applicants with verified income default at least 2% more than those without verified income, suggesting a need to revisit the verification process. This is significant considering that 43% of applicants' incomes were not verified.

- Applicants with loan purposes labeled as "educational" or "small business" tend to default significantly more than other categories, with a difference ranging from 8% to 17%. This is notable given that "debt consolidation" is the dominant loan purpose in the dataset, accounting for 46% of applicants.

- Applicants with higher credit revolving balances tend to default more often than those with lower credit revolving balances.
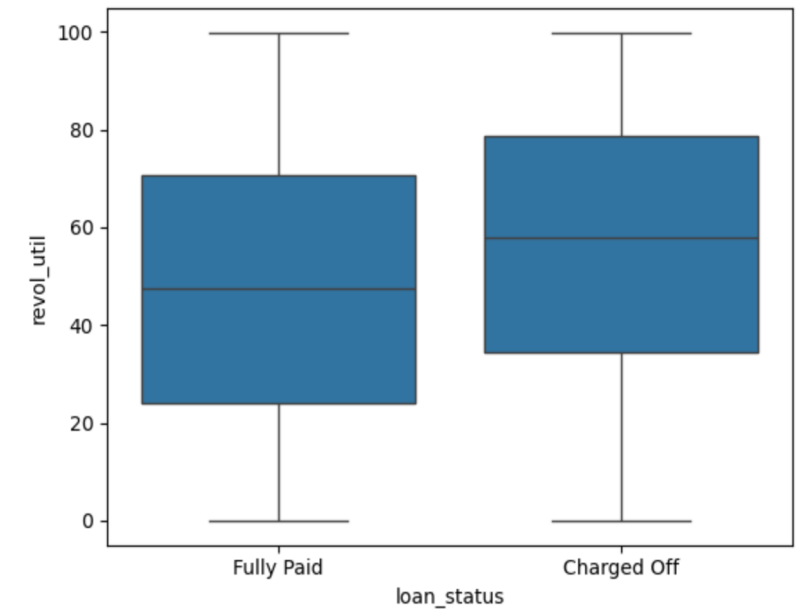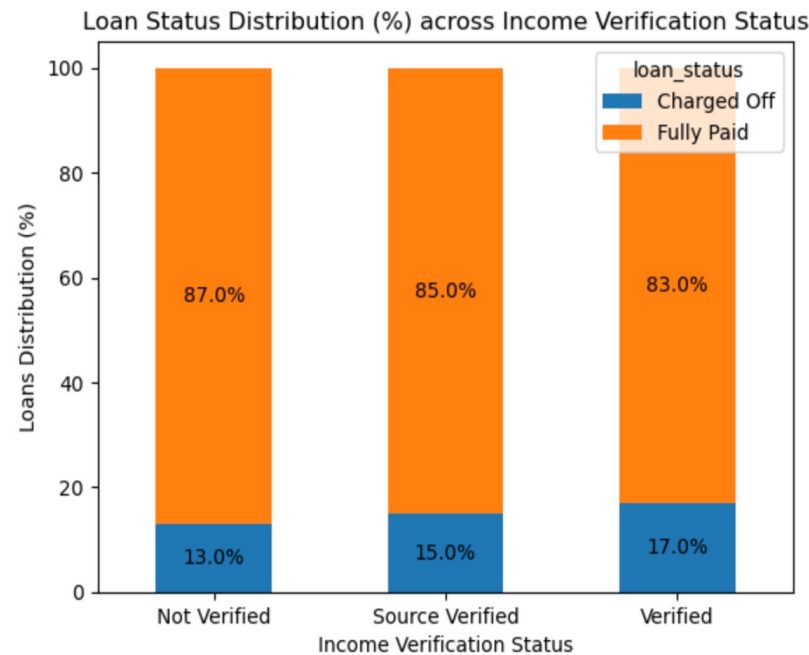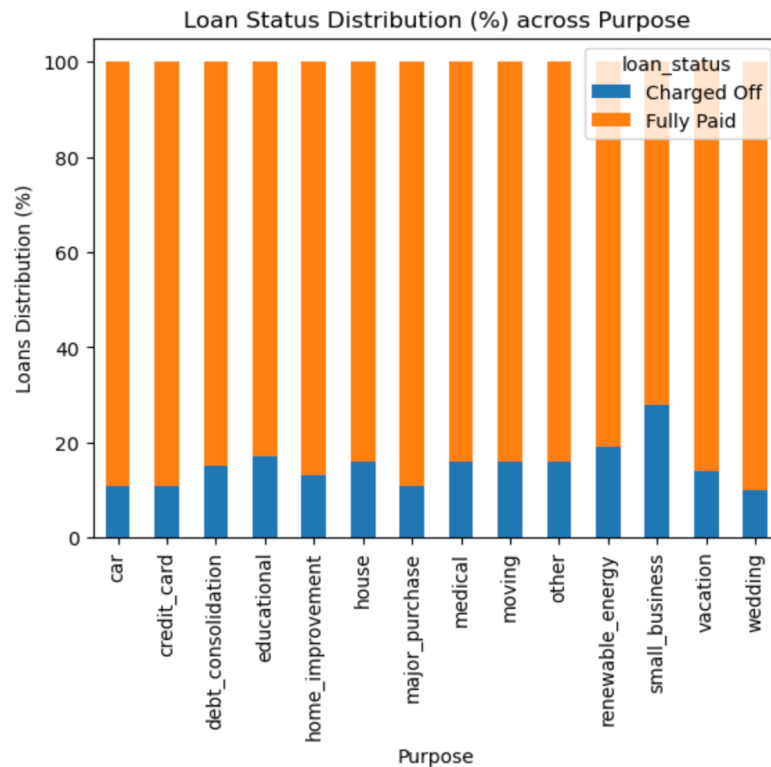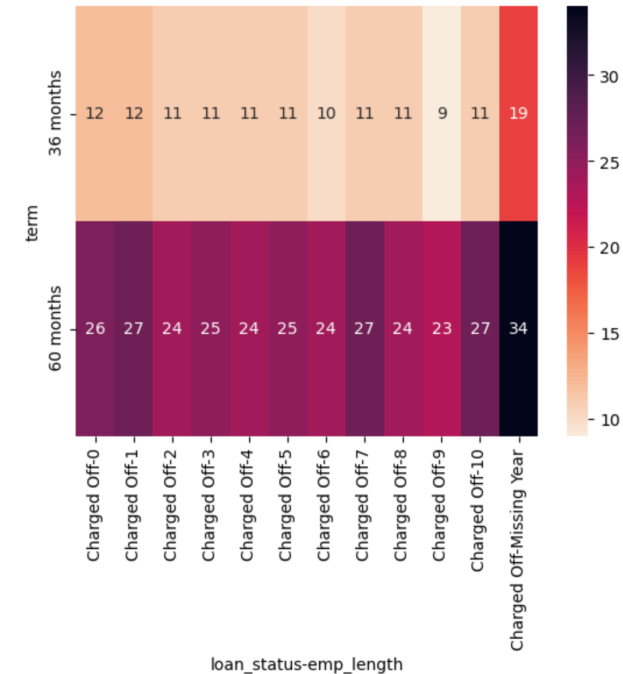
Bivariate Analysis

Loan Amount vs Funded Amount
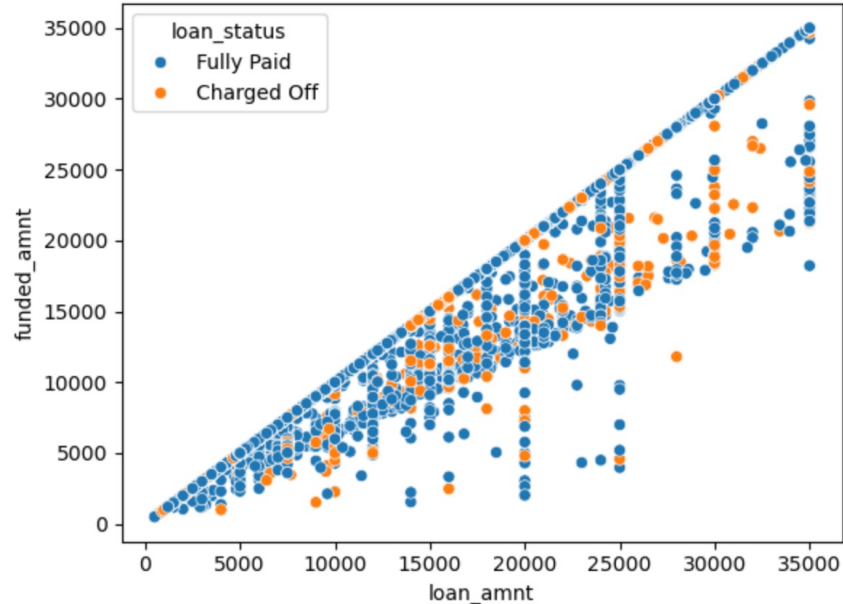
There are very few defaulters when the funded amount is nearly equal to the requested loan amount, indicating a linear relationship.

Term vs Employment Length

Applicants with an employment length of 7 years tend to default more often when the loan term is 60 months. It is also noteworthy that applicants with missing employment length exhibit a similar default pattern for 60-month terms.
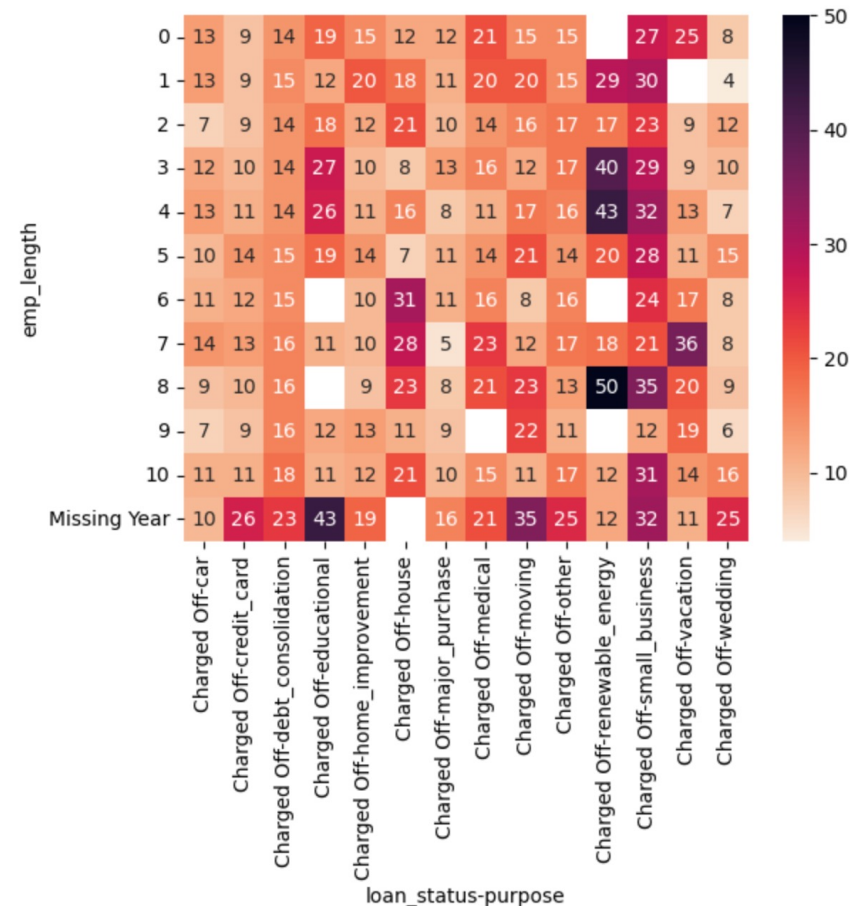
Bivariate Analysis

Loan Applicants for purpose as renewable energy tend to likely default when their employment length is 4 years or 8 years. Also, Loan Applicants with no employment length specified tend to default more for educational purpose

Term vs Credit Revolving Balance
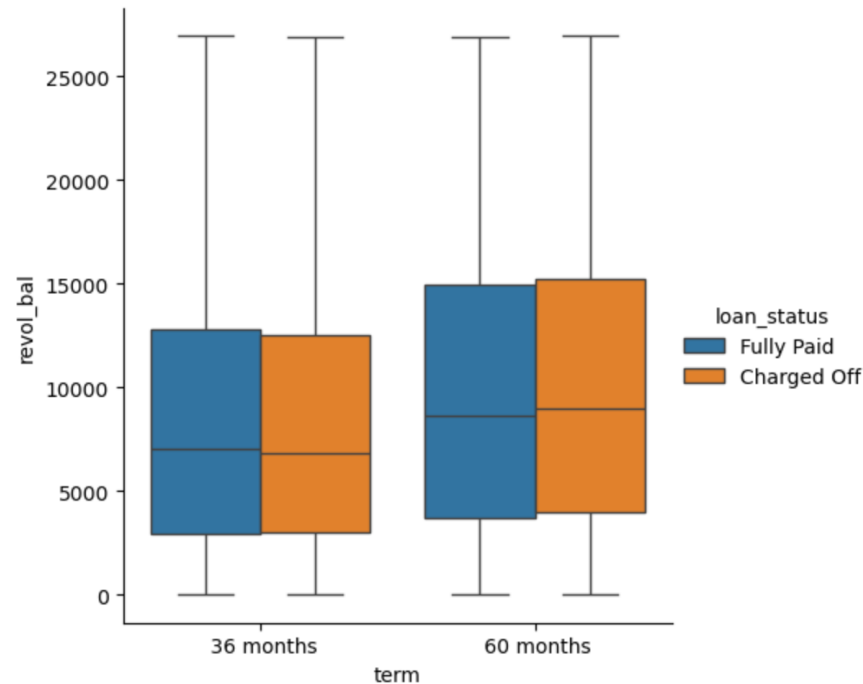
Loan Applicants with higher Total Credit Revolving Balance tend to default likely when the term is 60 months
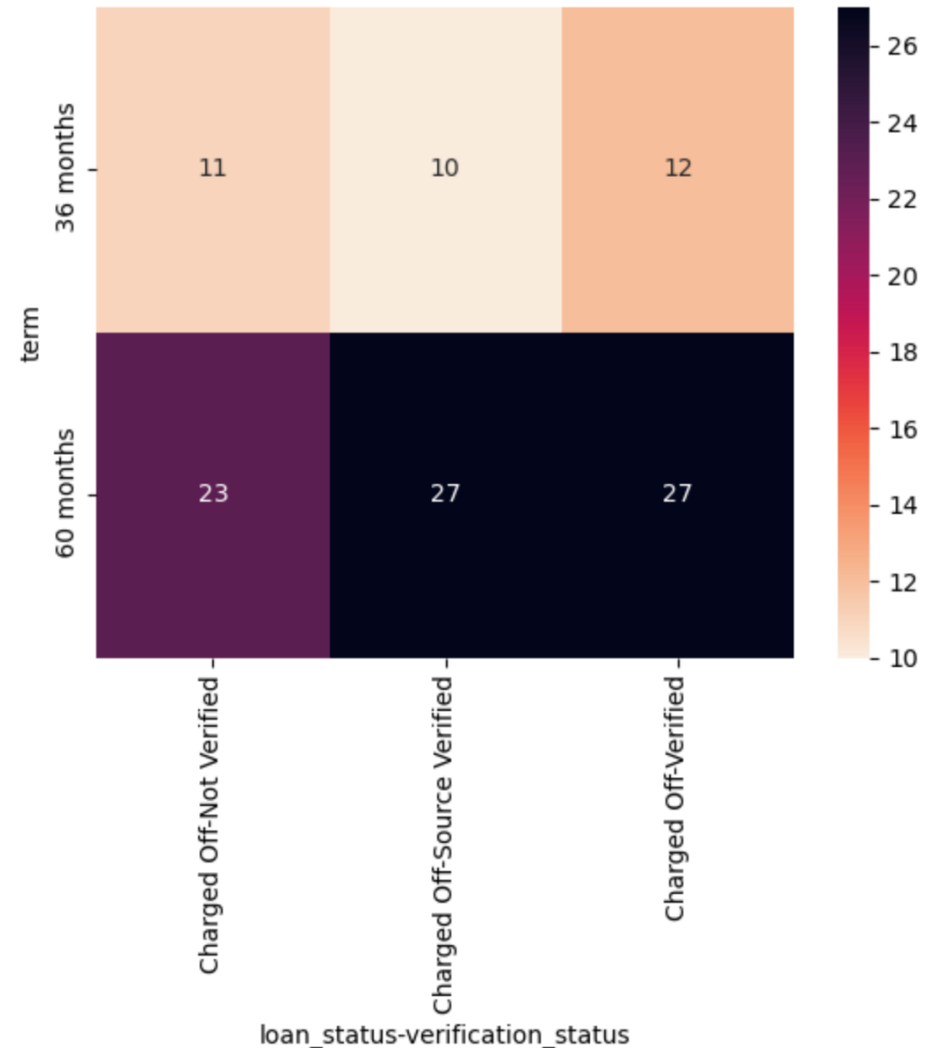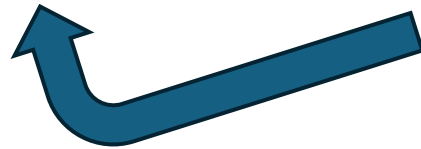
Emp Length vs Purpose

Bivariate Analysis

Term vs Verification Status

Loan Applications with term as 60
months and their income verified and
source verified tend to default more
likely than others

Conclusion

$ 1. Cease approving loans where the amount-to-income ratio exceeds 30%.

🏛 2. Reduce the number of approvals for loans designated for small businesses.

€ 3. Halt approvals for high-value loans when the revolving line utilization rate is greater than 75%.

⚠ 4. Stop approving loans for individuals with a prior bad credit record, or at least refrain from approving high-value loans for them.

💵 5. Implement higher interest rates for loans with a debt-to-income ratio greater than 20%.