Lending Club

Identifying Key Factors for Minimizing Defaults and Enhancing Loan Portfolio Performance

Ashish Chaudhary - 28-Jan-2025

Lending Club Case Study

Background

The background revolves around minimising credit loss for the lending company. By analysing the historical loan data, the company can predict the likelihood of loan defaults.
The project explores the application of EDA in understanding how different applicant and loan attributes impact the likelihood of default, enabling the company to make better, data-driven lending decisions.

Business Objective

The business objective is to identify risky loan applicants who are likely to default, in order to minimise credit loss. By understanding the key factors driving loan defaults, the company can make better lending decisions and reduce financial risks.

Loan Distribution

This slide introduces the **loan amount distribution** of applicants. It provides a **clear overview** of the frequency and spread of loan amounts across applicants.

Key Observations:

1. **Majority of Loan Amounts are in the Lower Range**:

   ○ The **bulk of the loans** in the dataset fall in the **lower loan amount range**, typically between **$2,000 and $20,000**.
   ○ This suggests that most applicants are applying for smaller loans, likely for personal use or short-term financial needs.

2. **Gradual Decrease for Higher Loan Amounts**:

   ○ The distribution **tapers off** as the loan amount increases, indicating that loans above **$20,000** are less frequent.

3. **KDE Curve**:

   ○ The **KDE curve** helps confirm the smooth distribution of loan amounts, showing a peak in the lower range and a gradual decrease in higher amounts.
   ○ There is a **clear drop-off in frequency** as loan amounts increase beyond $20,000.

4. **Outliers**:

   ○ The presence of a few larger loans (e.g., loans greater than $50,000) may suggest **outliers** or instances of high-value loans in the dataset.


Loan amount distribution
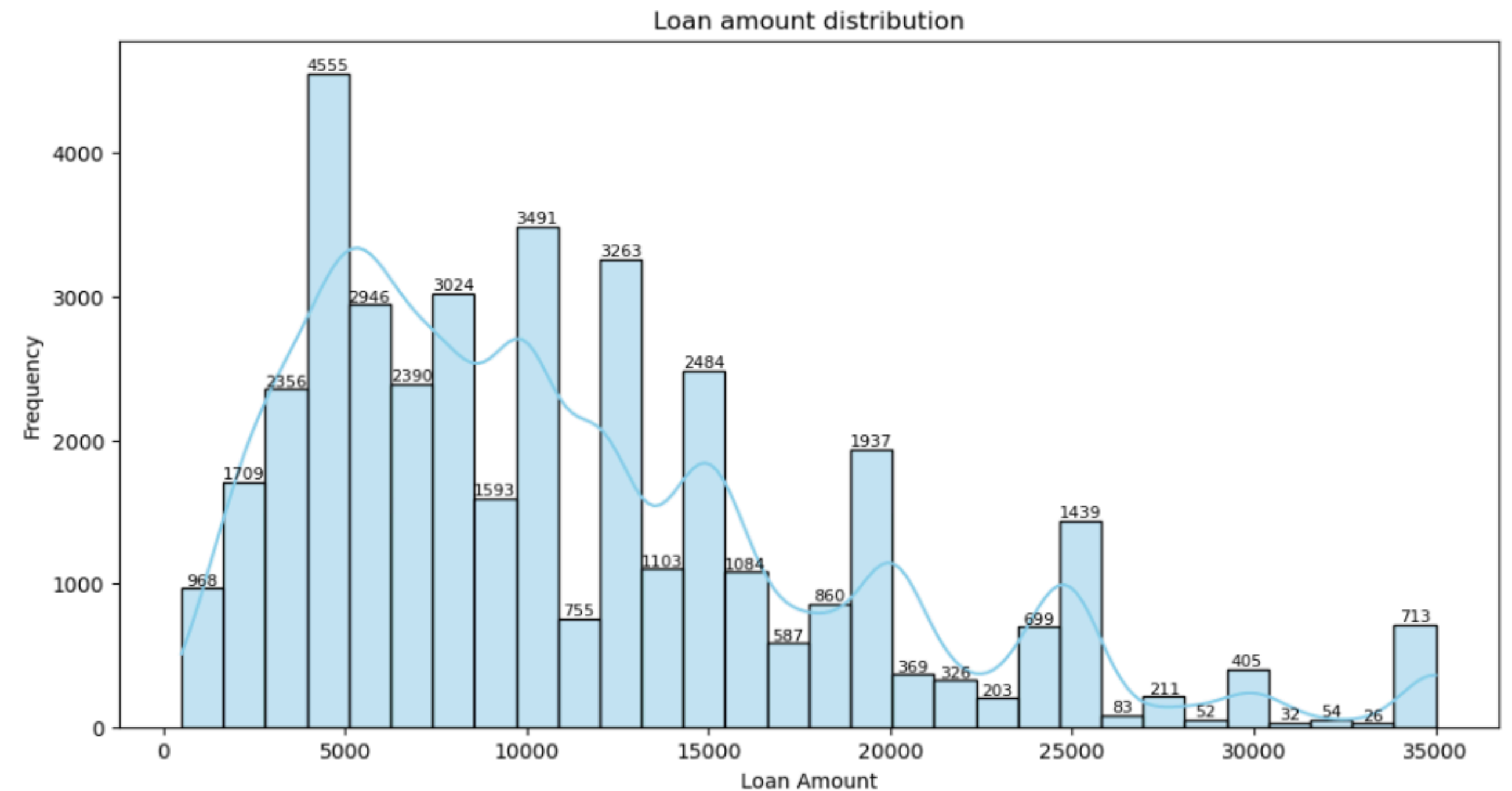
Loan Interest Rate Distribution

The **horizontal bar chart** illustrates the distribution of loan interest rates across different predefined ranges.

Key Insights:

- **Majority of Loans**: The majority of loans fall in the **0%-10%** and **10%-15%** ranges, suggesting that the company's lending is concentrated on borrowers with relatively lower interest rates.
- **Outliers**: There are fewer loans in the **30%+** category, indicating **high-risk loans** are less common but potentially more risky.
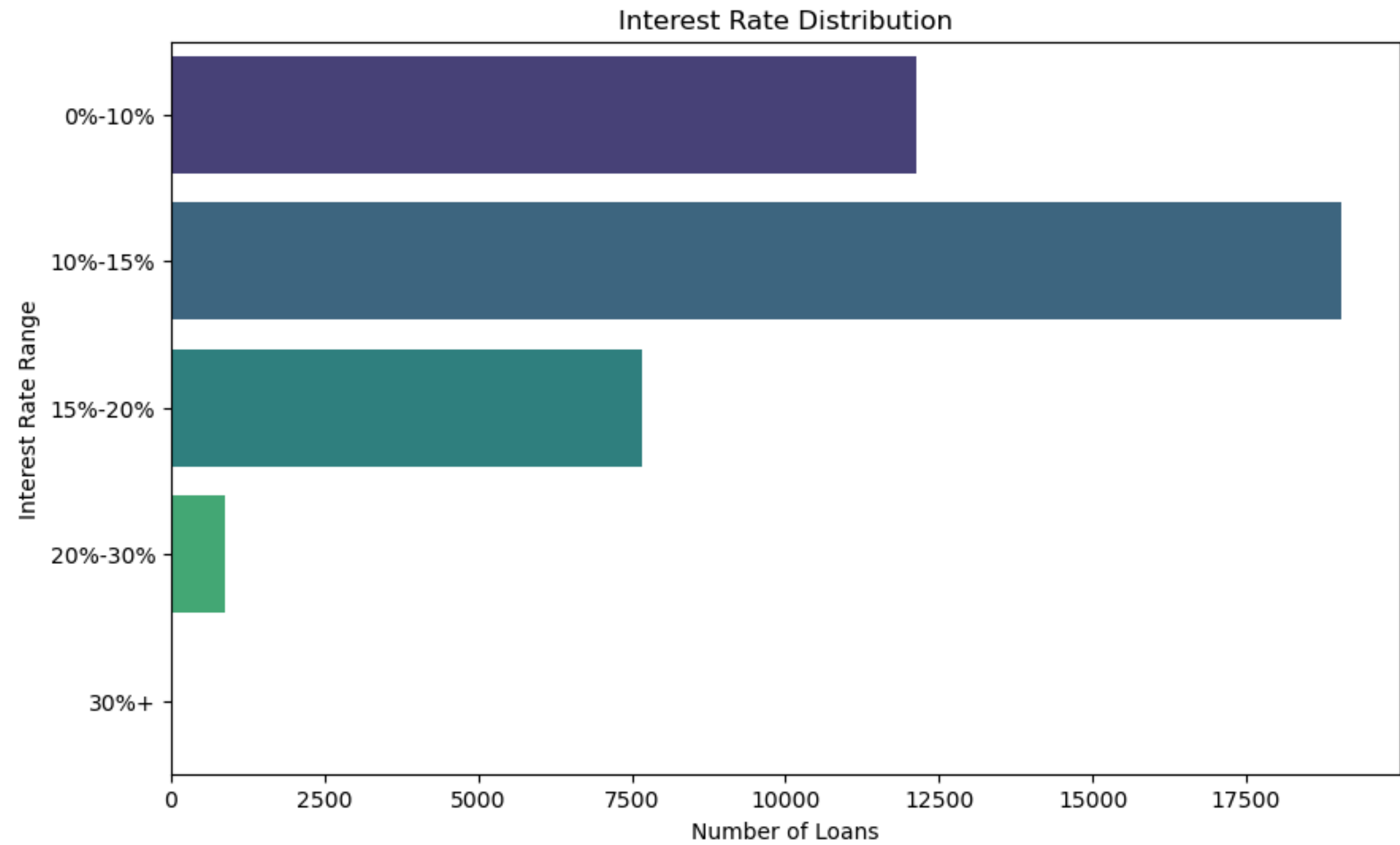
Loan Status Distribution by Term

The **count plot** shows the distribution of **loan status** across different **loan terms (36 months vs. 60 months)**. The chart provides insight into how loan status is related to the length of the loan term.

Key Observations:

1. **Fully Paid Loans**:
   - A higher number of **fully paid loans** are observed in the **36-month term** category. This indicates that borrowers with shorter loan terms are more likely to repay their loans in full within the specified period.
2. **Charged Off Loans**:
   - **Charged-off loans** (i.e., loans where the borrower defaults) are more common in the **60-month term** category. This suggests that borrowers who take longer to repay their loans (5 years) may face more difficulty in maintaining timely payments.
3. **Current Loans**:
   - The **current loans** (loans that are in the process of being repaid) are relatively balanced across both **36-month** and **60-month** terms. This suggests that while both loan term types have loans under repayment, there isn't a significant skew towards either term.
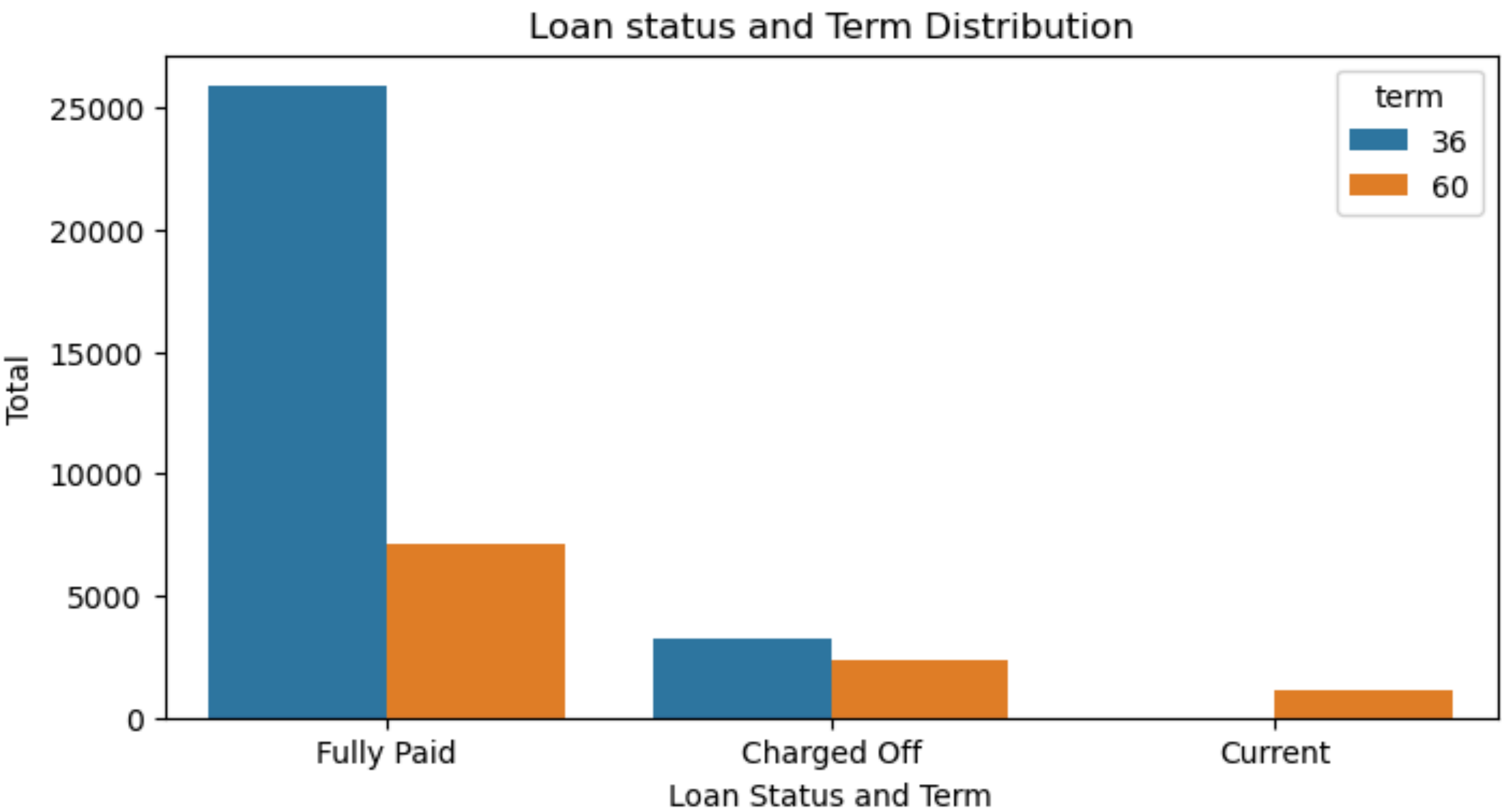
Loan Amount Distribution by Purpose Over Time

The purpose categories are visually represented by different colored areas,
providing insights into how the company's loan distribution has evolved over time.

Key Findings:

1. **Loan Purpose Trends**:

   - **Credit Card** and **Debt Consolidation** loans seem to consistently make up a significant portion
     of the total loan amount, showing their dominance in the loan portfolio over time.
   - **Small Business** loans, while not as large in volume, have seen noticeable growth in certain
     months, indicating a growing demand in this category.
   - **Car** and **Home Improvement** loans display fluctuations but remain relatively stable compared
     to other categories.

2. **Growth Areas**:

   - **Debt Consolidation** loans appear to have experienced significant growth, particularly in the
     later months. This could signal an increasing need for **debt management** solutions.
   - The **Small Business** loan category also shows upward momentum, especially in the most
     recent periods, suggesting a potential growth area for the company.

3. **Seasonal and Time-Based Fluctuations**:

   - There is clear **seasonality** in the loan amounts, with certain months showing more pronounced
     peaks, possibly due to economic conditions or marketing strategies.
   - The **loan distribution** by purpose shows varying **peaks** and **valleys** over time, indicating
     potential periods of increased borrowing activity.
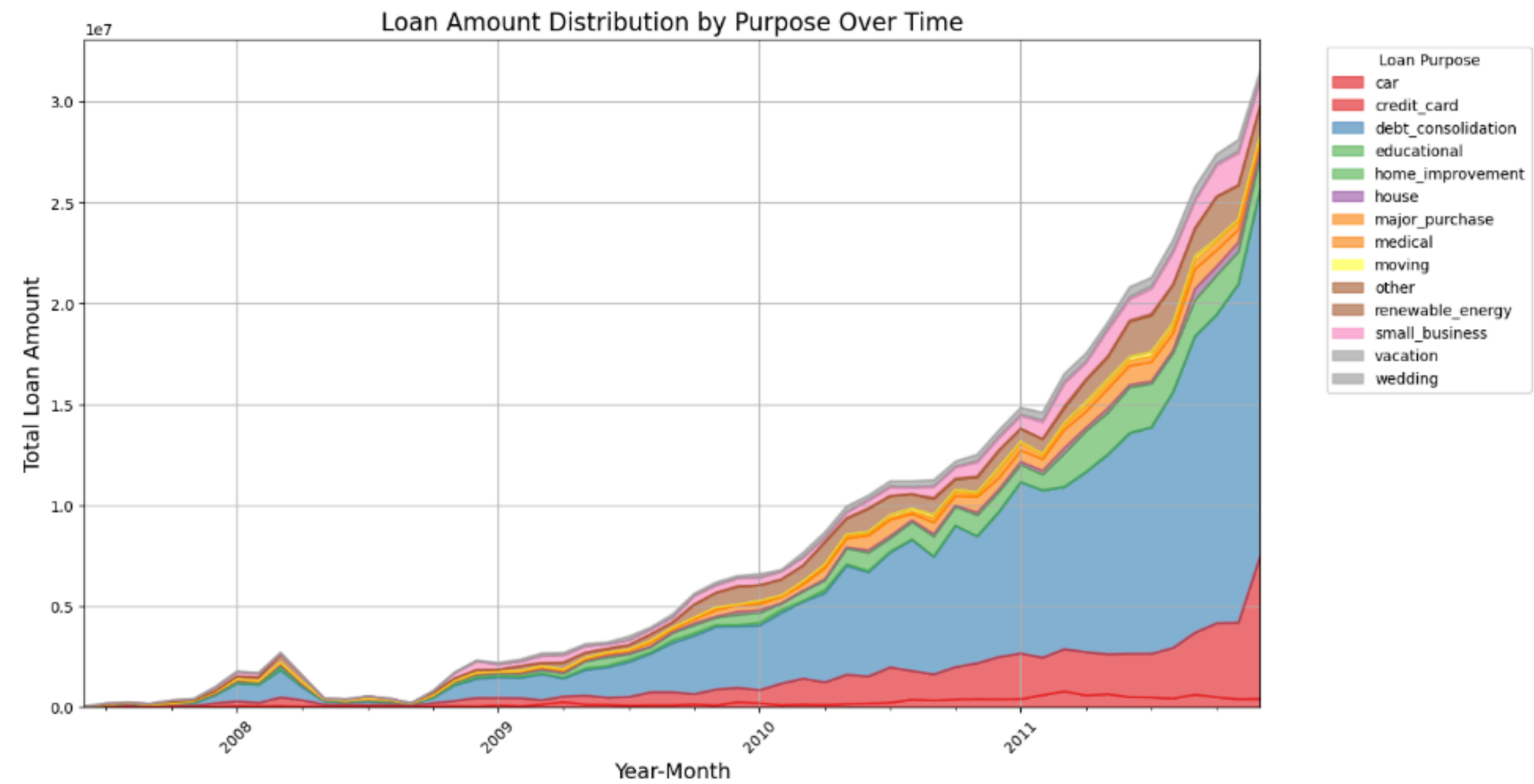
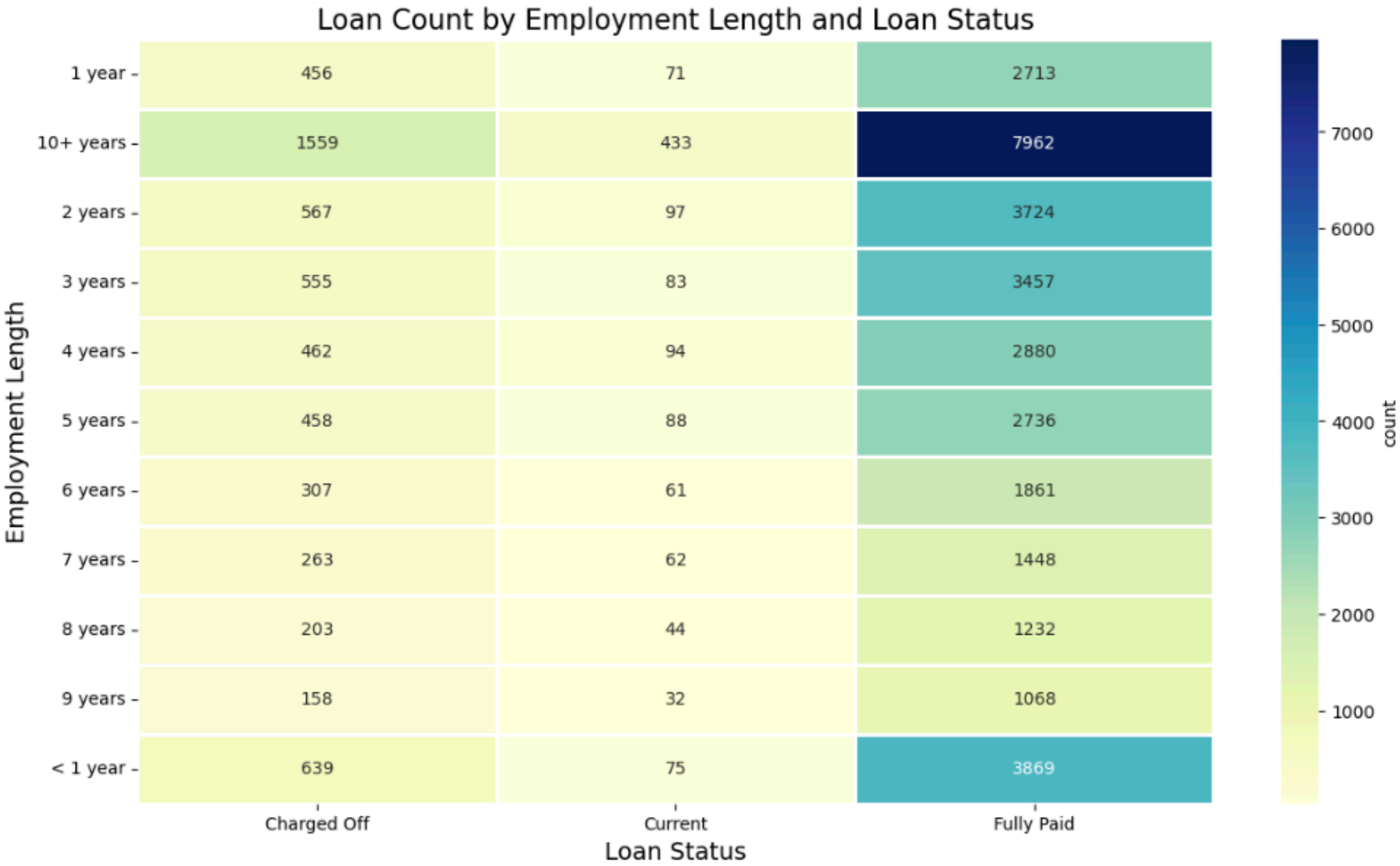# Loan Count by Employment Length and Loan status

The values within the heatmap represent the **number of loans** for each **employment length** and **loan status** combination, providing insight into the correlation between a borrower's employment length and the likelihood of loan repayment.

Key Findings:

1. **Employment Length and Loan Status**:

   ◦ Borrowers with **longer employment lengths** (e.g., **10+ years**) tend to have more **fully paid loans**, indicating stability and reliability.

   ◦ Borrowers with **shorter employment lengths** (e.g., **< 1 year**) exhibit a higher proportion of **charged-off loans**, which suggests they might be higher-risk borrowers.

2. **Loan Status Distribution**:

   ◦ **Fully Paid** loans dominate across **most employment lengths**, particularly for those with **> 1 year** of employment.

   ◦ **Charged Off** loans are more common in the **< 1 year** employment category, indicating that those with less job stability are more prone to loan defaults.

   ◦ The **Current** loans are fairly evenly distributed across different employment lengths, indicating that many loans are still being repaid but are yet to be fully settled.

Key Insights:

• **Employment Length Correlation**: A direct correlation can be observed between **longer employment lengths** and **loan repayment success**, highlighting that **job stability** is a key factor in **loan performance**.

• **Risk Identification**: Borrowers with shorter employment histories (**< 1 year**) are more likely to default, which can help the bank identify **higher-risk borrowers** early on and apply risk mitigation strategies.

Loan Count by Employment Length and Loan Status

Employment Length	Charged Off	Current	Fully Paid
1 year	456	71	2713
10+ years	1559	433	7962
2 years	567	97	3724
3 years	555	83	3457
4 years	462	94	2880
5 years	458	88	2736
6 years	307	61	1861
7 years	263	62	1448
8 years	203	44	1232
9 years	158	32	1068
< 1 year	639	75	3869

# Loan Grade Distribution and contribution to Total Loan Amount

Each bar represents a loan grade, with the percentage of total loans for each grade shown above the respective bars. The **percentage labels** provide a clearer view of how each loan grade contributes to the overall loan portfolio.

Key Findings:

1. **Grade Distribution**:
   - **Grade A** loans represent the **highest proportion** of total loans, indicating that a large number of borrowers are in the lower-risk category.
   - **Grades B and C** follow closely behind, suggesting a substantial share of loans fall into moderate-risk categories.
   - **Grades D, E, and F** represent a **smaller share** of the loan portfolio, with **higher risk** associated with these grades.
2. **Loan Contribution**:
   - **Grade A** loans contribute to a **significant percentage** of the total loan amount, showcasing that the company is predominantly lending to **lower-risk borrowers**.
   - **Grade F** loans contribute to a **small percentage** of the total loans, highlighting the **low volume** but **high risk** associated with this grade.
3. **Percentage Breakdown**:
   - The percentage labels on each bar give a precise understanding of **how much each grade contributes** to the overall loan portfolio, offering a clear view of loan risk exposure.

Key Insights:

- **Dominance of Grade A Loans**: A substantial portion of the loan portfolio is composed of **Grade A** loans, which typically have **lower default rates**. This is a positive indicator of the company's current lending strategy.

- **Risk Management Opportunity in Grades D, E, F**: The **smaller share** of **Grades D, E, and F** represents higher risk, but since they contribute **lesser** to the total loan amount, they might be targeted for **higher interest rates** or **risk-reduction strategies**.

- **Portfolio Composition**: The majority of loans are in the **Grade A to C** categories, which suggests a **low-to-medium risk** portfolio. This distribution is likely a **healthy balance** for the company, with room for growth in slightly higher-risk segments (Grades D and E) if managed appropriately.


Loan Grade distribution
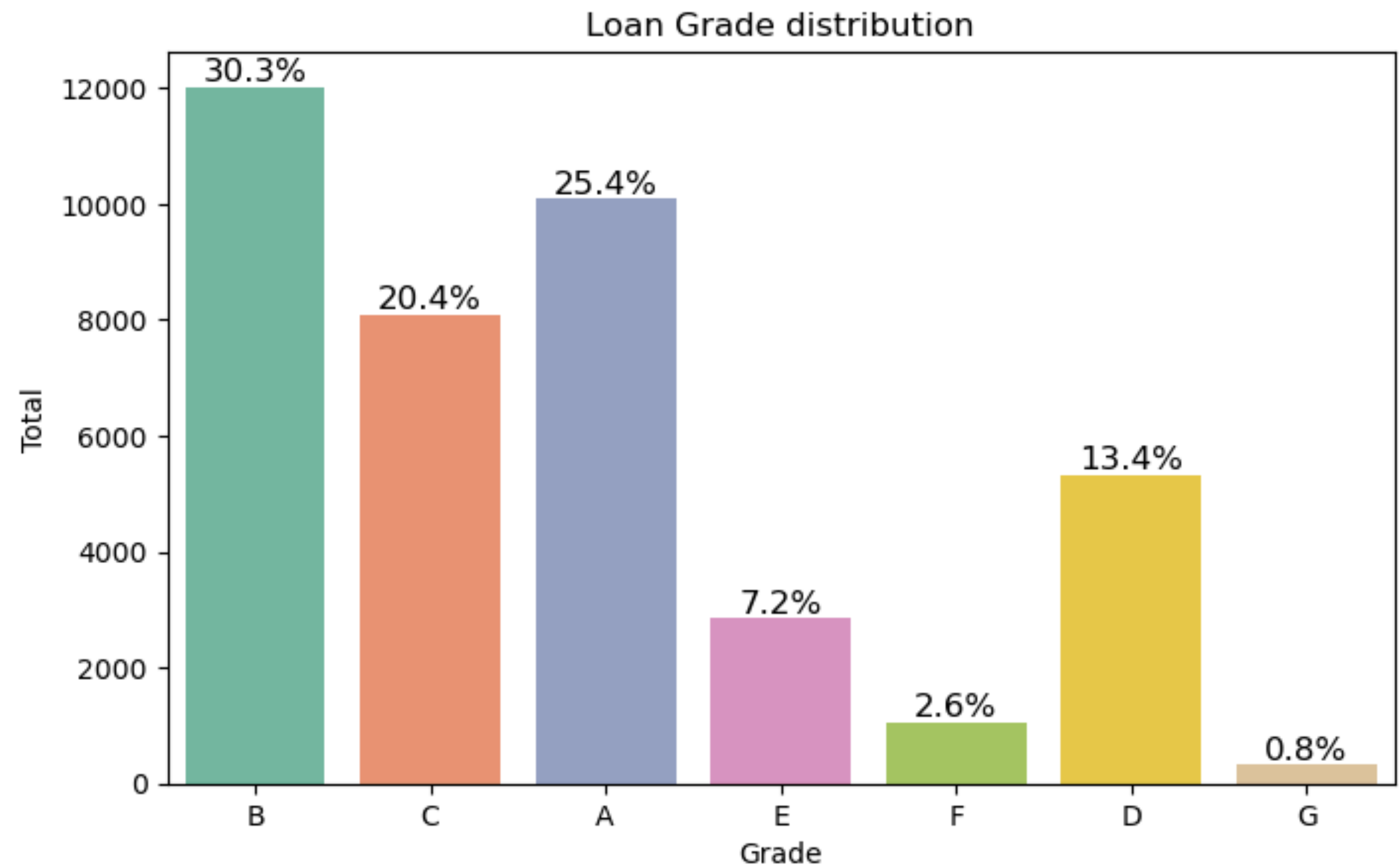
Home Ownership and it's impact on Loan Status

Different loan statuses are shown by **color coding**: **Fully Paid** (green), **Charged Off** (red), and **Current** (blue), making it easy to visualize how home ownership may influence the **likelihood of loan repayment**.

Key Findings:

- **Higher Loan Repayment (Fully Paid)** observed for **Renters** and **Mortgage Holders**.
    - Renters: **Unexpected financial stability** despite not owning a home.
    - Mortgage Holders: **Strong financial responsibility** due to significant existing commitments
- **Higher Risk of Default** among **Others** and **None** categories:
    - Individuals with **no permanent housing** or **temporary situations** are more likely to default
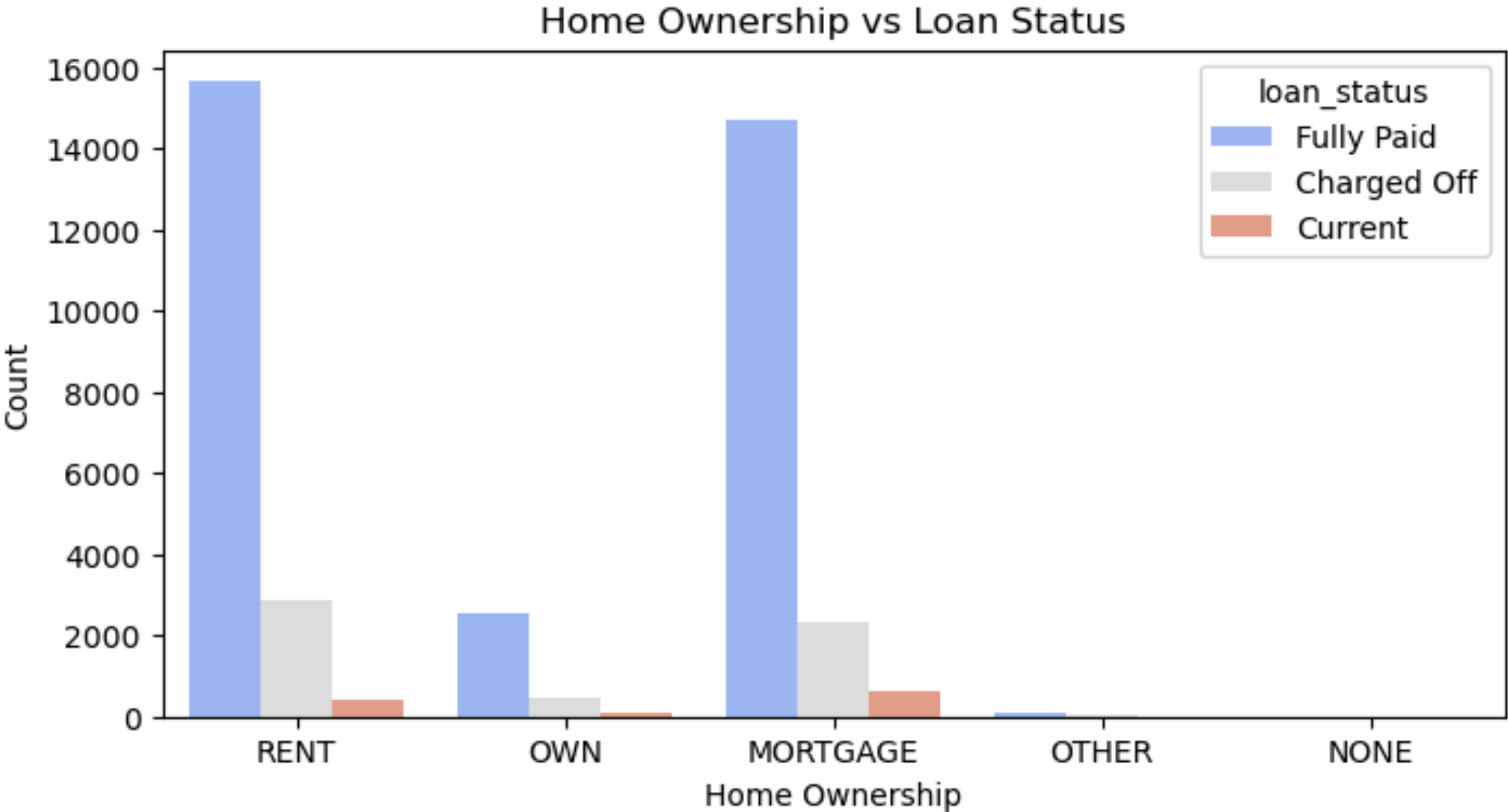
Loan Distribution by Annual Income Ranges

It highlights key patterns, including the presence of **outliers** and the tendency for **higher-income individuals** to receive **larger loans**.

Key Insights:

1. **Higher Loan Amounts with Income:**

   - **Higher income groups** (e.g., **300K-500K**, **500K-1M**) tend to receive **larger loan amounts**.
   - The loan amount increases in line with **annual income**.

2. **Outliers in Lower Income Ranges:**

   - **Outliers** are observed in the **0K-50K** and **50K-100K** income ranges, with some individuals requesting **disproportionately high loans** compared to their income.
   - This highlights the need for **closer scrutiny** in these categories to ensure proper risk assessment.

3. **Outliers in the Higher Income Range:**

   - Another set of **outliers** appear in the **1M-2M** income range, where **loan amounts** are significantly larger than the general trend, which may warrant special handling.

4. **Stable Loan Distribution in Mid-Income Groups:**

   - **100K-150K** and **150K-200K** income ranges show a **more uniform distribution** of loan amounts, suggesting **stable borrowing patterns**.

Loan Amount Distribution by Annual Income Ranges
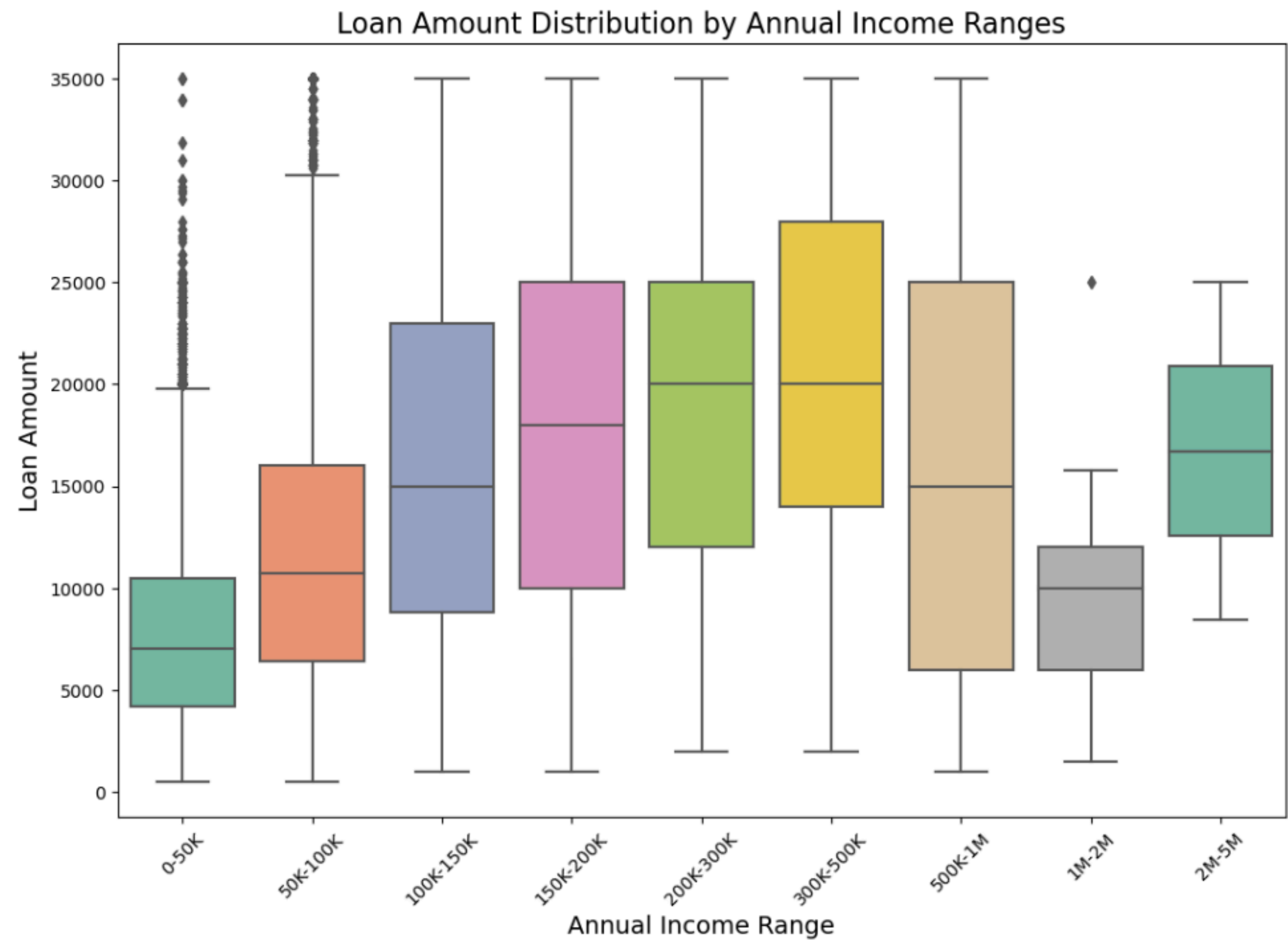
Suggestions

Careful Scrutiny of Low-Income Applicants:

- Applicants in the **0K-50K** and **50K-100K** income ranges showing disproportionate loan requests should undergo **thorough risk assessment** to avoid defaults.

Tailored Loan Amounts Based on Income:

- Use **income ranges** as a guide to offering loan amounts that are proportionate to an applicant's financial capacity. Consider reducing loan sizes for **high-risk applicants** in lower income groups.

Consider Interest Rate as a Risk Indicator:

- **Higher interest rates** often correlate with higher default risks. Therefore, consider reducing the interest rate for **creditworthy borrowers** to ensure timely repayment and minimise defaults.

Favour Applicants with Long Employment Tenure:

- Applicants with **stable employment** (e.g., **10+ years**) should be **prioritised** for loan approval, as they demonstrate financial stability and lower default probabilities.

Monitor Housing Status Carefully:

- **Homeowners** tend to have a lower default risk. Incorporate **housing status** as a **critical factor** in the loan approval process.

Introduce a Loan Cap Based on Income:

- For higher-income applicants (above **1M**), consider implementing **caps on loan amounts** to avoid extreme outliers that could lead to defaults.

Thank you

Ashish Chaudhary