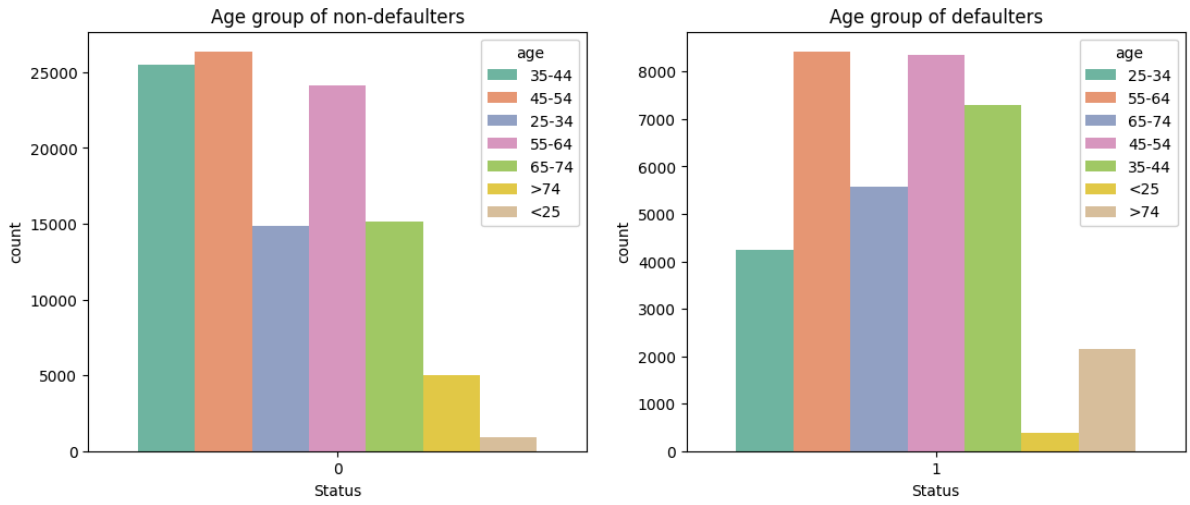
LOAN DEFAULT ANALYSIS

Introduction:

Analyze the factors contributing to loan defaults to improve risk management and loan approval strategies. This project provides an opportunity to enhance risk assessment strategies in lending institutions, leading to more informed decision-making and proactive default prevention measures.

Insights based on Age



Observation:

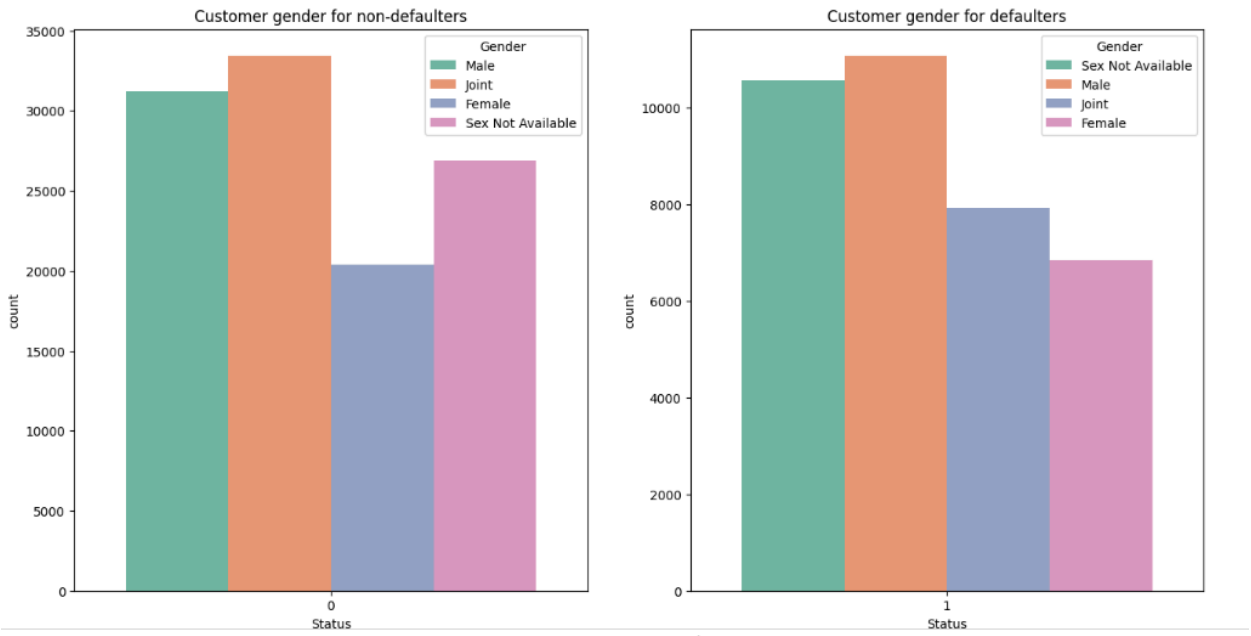
- The **45-54 age group** has the highest count of **Non-Defaulters**.
- The **55-64 age group** has the highest count of **Defaulters**.

Defaulter Count:

- **45-54 Age Group**: 26,357 of total non-defaulters

- **55-64 Age Group**: 8,419 of total defaulters

Insights based on Gender



Observation:

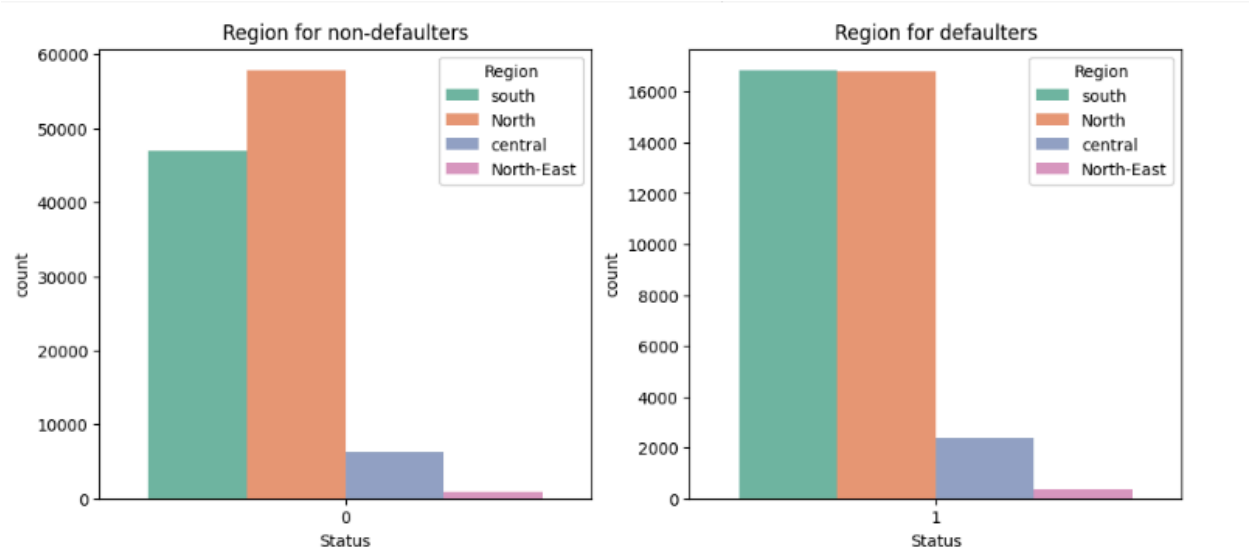
When comparing gender, male clients have a higher number of defaulters compared to female clients.

Male: 11,079 of total defaulters
Female : 6,843 of total defaulters

Insight:

- The data suggests that male borrowers are at a greater risk of defaulting on loans.
- This could be driven by several socioeconomic or behavioral factors, or financial patterns observed more frequently in male borrowers.

Insights based on Region



Observation:

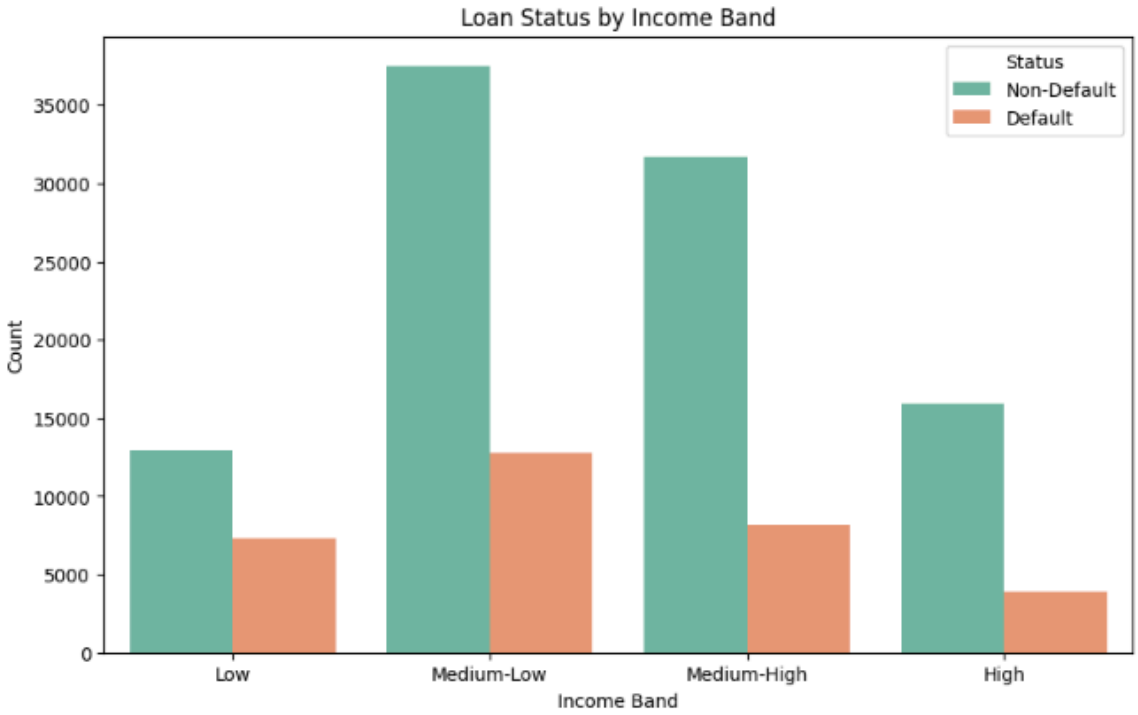
When comparing regions, the **North** and **South** regions have the highest number of both defaulters and non-defaulters.

- **North:** Approximately 16,000 defaulters and over 55,000 non-defaulters.
- **South:** Approximately 16,000 defaulters and over 45,000 non-defaulters.
- **Central:** Significantly lower numbers for both defaulters and non-defaulters (around 5,000 defaulters and fewer than 10,000 non-defaulters).
- **North-East:** The lowest in both categories, with fewer than 1,000 defaulters and non-defaulters.

Insight:

The data suggests that the **North** and **South** regions are at a greater risk of loan defaults due to higher counts of defaulters. This could be influenced by the larger population size, higher loan uptake, or economic conditions.

Insights based on Income Band



Observation:

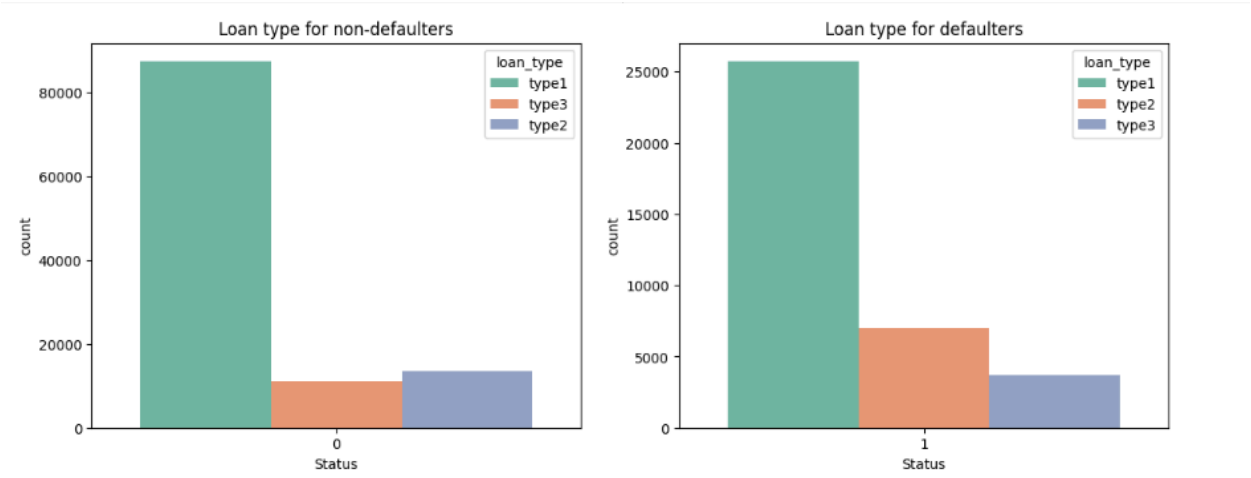
When comparing loan status across income bands:

- In the **Low** income band, the number of non-defaulters is higher than defaulters, with over 15,000 non-defaulters and fewer than 10,000 defaulters.
- The **Medium-Low** income band has the highest number of both non-defaulters (over 35,000) and defaulters (approximately 15,000).
- The **Medium-High** income band has a slightly lower number of non-defaulters (around 30,000), while defaulters in this group are also fewer (around 10,000).
- In the **High** income band, non-defaulters (approximately 15,000) far outnumber defaulters (fewer than 5,000).

Insight:

The data suggests that individuals in the **Medium-Low** income band are at a higher risk of defaulting, as this group has the highest number of defaulters.

Insights based on Loan Type



Observation:

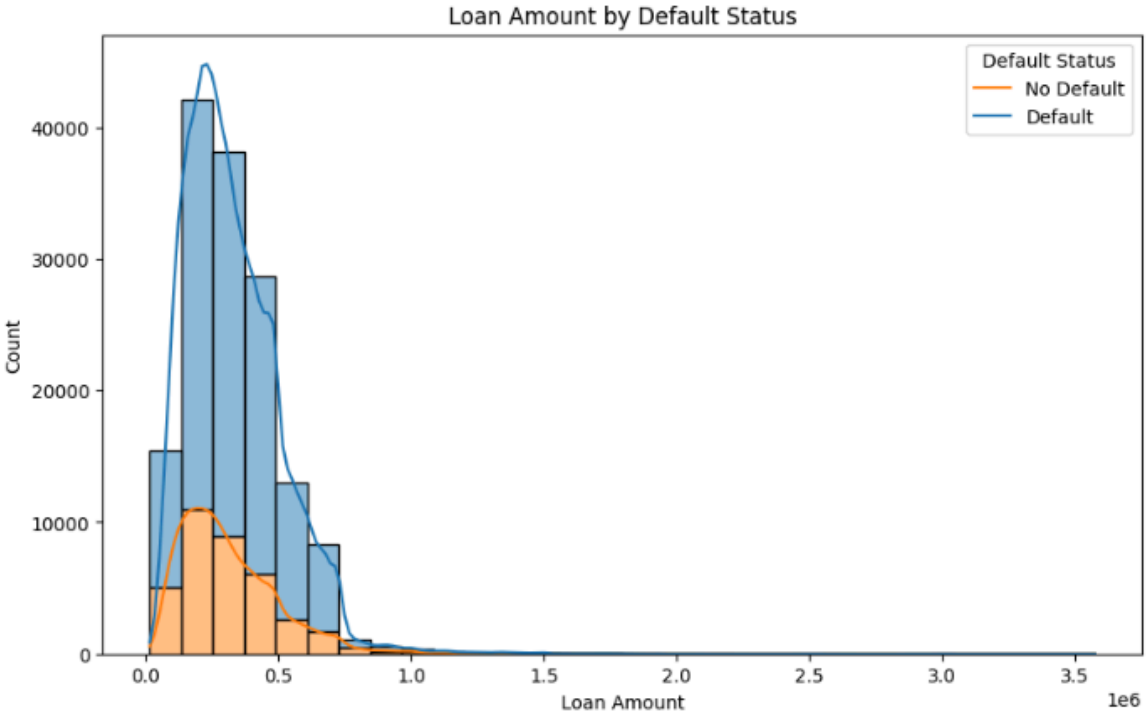
The charts compare the distribution of loan types among non-defaulters and defaulters:

- For **non-defaulters**, **loan type 1** dominates with over 80,000 borrowers, followed by a much smaller number of **loan type 3** and **loan type 2** borrowers (each under 10,000).
- For **defaulters**, **loan type 1** still dominates with around 25,000 borrowers, but both **loan type 2** and **loan type 3** have fewer defaulters (each under 10,000).

Insight:

The data suggests that **loan type 1** is the most common among both defaulters and non-defaulters, but it also has the highest default count. However, **loan type 2** and **loan type 3**, though smaller in number, seem to carry a relatively higher risk of default in comparison to the number of non-defaulters. This could indicate that borrowers of **loan type 1** are more numerous overall, but borrowers of **loan**

types 2 and **3** might face more difficulties in repayment, or these loans may have terms that increase the risk of default.



Observation:

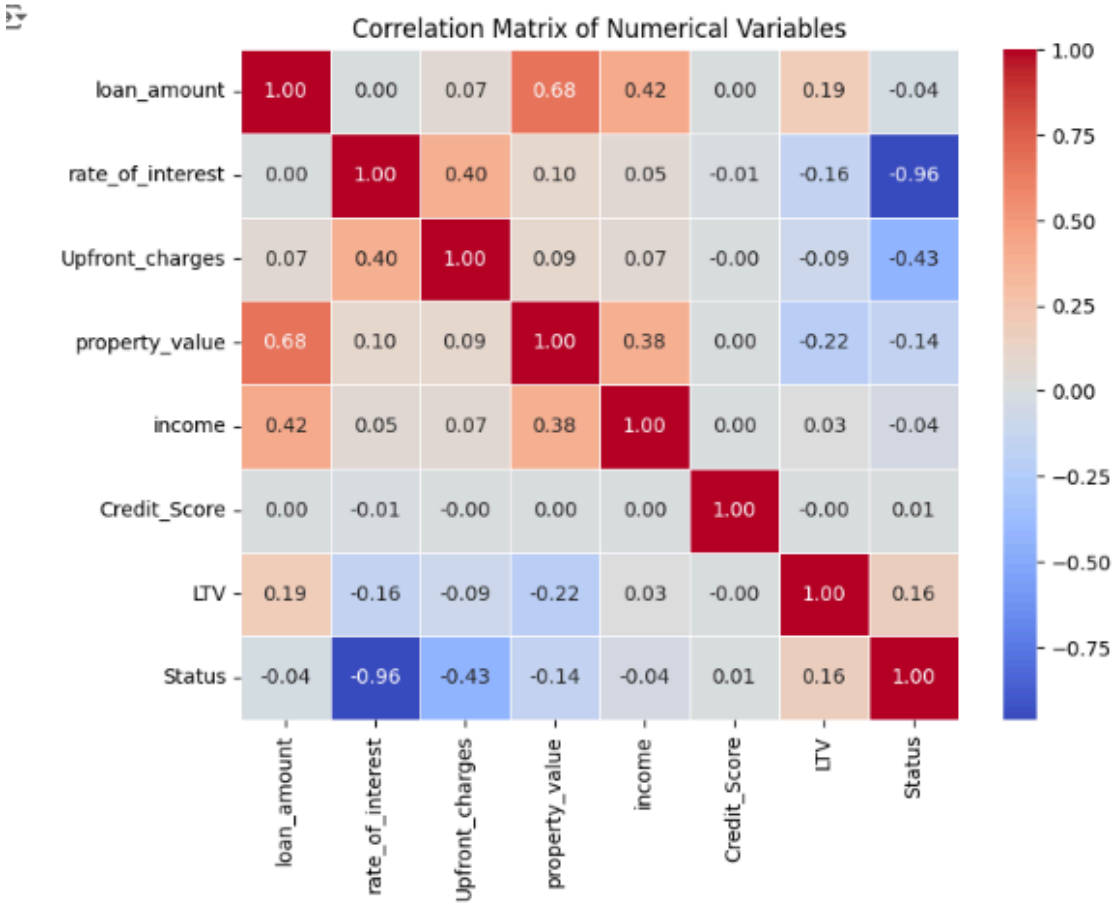
From the plot, we can observe that more default status has a higher loan amount.

- **No Default**: These borrowers tend to have lower loan amounts.
- **Default**: The higher the loan amount it has a higher chance of getting default.

Insight:

The data suggests that higher loan amounts might correlate with an increased likelihood of default. Borrowers with larger loans may face more difficulty in repayment, potentially due to over-leveraging or higher financial burdens. This pattern could help lenders refine their risk assessment models, especially for higher loan amounts.

Correlation across different columns



Observation:

- **Loan Amount** has a positive correlation with **Property Value** (0.68), meaning that larger loans are typically associated with more expensive properties.
- **Rate of Interest** has a strong negative correlation with **Credit Score** (-0.96), showing that lower credit scores are associated with higher interest rates.
- **Upfront Charges** have a moderate negative correlation with **Status** (-0.43), indicating that higher upfront charges may lead to an increased likelihood of default.

- There is a moderate positive correlation (0.16), indicating that higher LTV ratios may be linked to a greater likelihood of loan approval, although the relationship is not very strong.

Recommendations

- **More loans to female customers**: Since fewer female customers default, increase loan opportunities for women to reduce overall risk.
- **Focus on the Central and Northeast regions**: These regions show fewer defaults, so prioritize lending there for better loan performance.
- **Target younger age group (25-44)**: Younger individuals (25-44) should be prioritized as they likely have more earning years ahead and lower default risks.
- **Increase commercial loans**: As commercial loans have a lower default rate, expand offerings in this category to support business growth.
- **Diversify loan types**: Instead of focusing on just one loan type (type1), ensure equal distribution across all loan types to reduce concentration risk.