CREDIT RISK ANALYSIS

Course: Data Science
Student: Saransh Rai
Professor: IIITB

TABLE OF CONTENTS

Objective

The goal of this assignment is to perform Exploratory Data Analysis (EDA) on a credit-related dataset i.e., to uncover patterns, trends, and insights that can help in understanding the factors influencing creditworthiness, loan defaults, or other credit-related outcomes

Key Challenges

Identifying defaulters while ensuring eligible clients are not rejected.

Overall Approach

1. Data Sourcing (Already Provided in the assignment)

- Provided with two data sets i.e., "application_data.csv" and "previous_application.csv"

2. Data Loading and Data Cleaning – [Done All this in Python Notebook]

- Changing Columns names for better understanding
- Checking Null Values.
- Removing Columns with null values above a threshold i.e., 40%.
- Changing Data of Columns in a Standardized format i.e., there were columns where I converted days in age and days of employment to years of employment
- Making box plot for numeric columns to check outliers.
- Fill null-values with mean/median/mode or create a new category if the imputing the values with mean/median/mode create biased approach.
- Null –Values in numeric data are filled with median in case of outliers or mean if there are no outliers.
- Null-Values in 'object' datatype is filled with mode.
- Handling Outliers (via Capping and Flooring etc.)
- Binning of numeric columns in the dataset like (AGE, Annual_Income, Credit_Loan_Amount and Lender_Normalized_Score_External_Source_3 and 2 columns)

Overall Approach

3. Checking Data Imbalance – **Deleting Columns where data is highly imbalanced**

4. Univariate Analysis

5. Correlation – To find relationships between important variables that can help in understanding the risk of loan default.

6. Bivariate Analysis and Multivariate Analysis - To find relationships between important variables that can help in understanding the risk of loan default

7. Merging Two Datasets – i.e., Previous and Application dataset and doing all the analysis like Univariate Analysis, Correlation & Bivariate and Multivariate Analysis

8. Conclusion & Recommendations

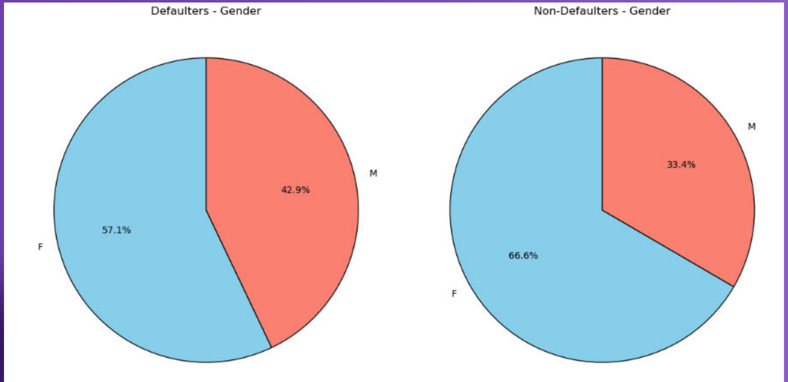
02

EXPLORATORY ANALYSIS

Current Application - Dataset

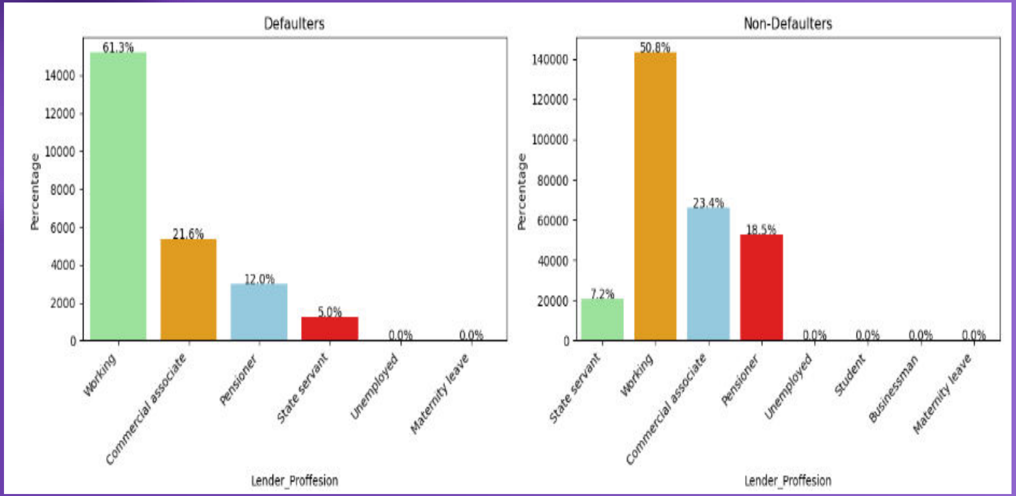
Univariate Analysis

Univariate Analysis - Defaulters and non-defaulters



Possible Interpretations - Gender

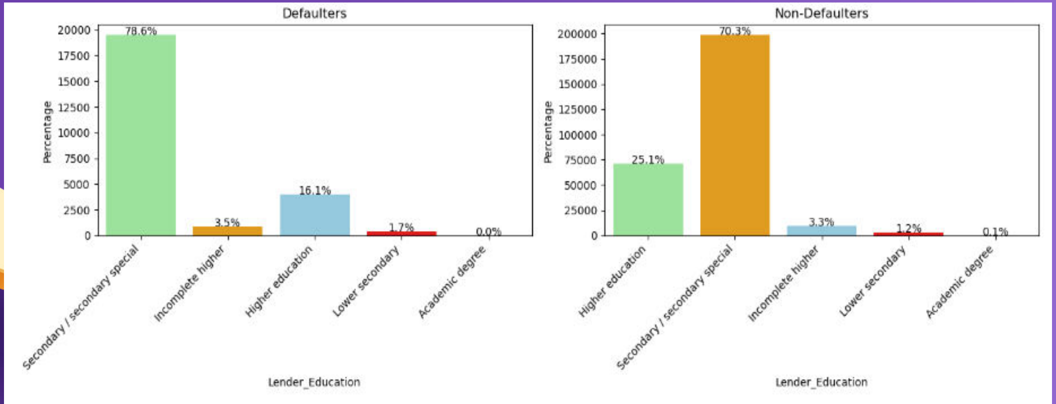
- A higher percentage of females are non-defaulters, suggesting better financial responsibility or repayment discipline.
- Although more females default than males, the overall trend favors females in terms of repayment behavior.
- Males have a relatively higher likelihood of defaulting compared to their presence in the non-defaulters' group.



Possible Interpretations – Lender's Profession

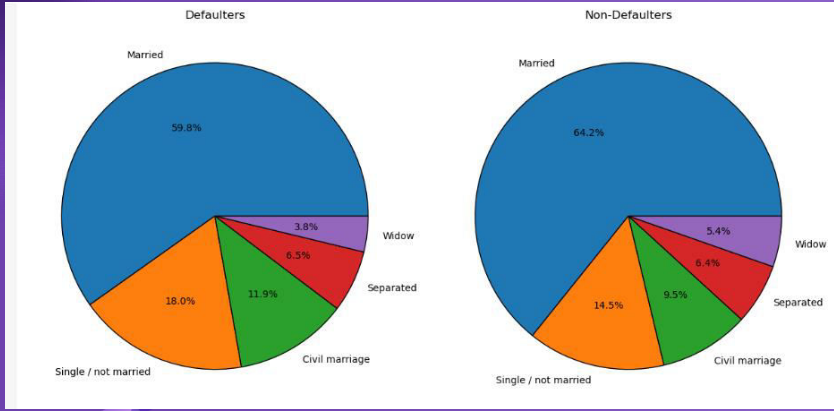
- "Working" professionals have the highest default rate, but they also make up the largest category in non-defaulters.
- "Pensioners" show better financial discipline, as their share among non-defaulters is higher.
- "Commercial Associates" are somewhat balanced but still have a considerable default risk.
- Minority professions like "State Servants" have very little contribution to the dataset.

Univariate Analysis – Defaulters and non-defaulters



Possible Interpretations – Lender's Education

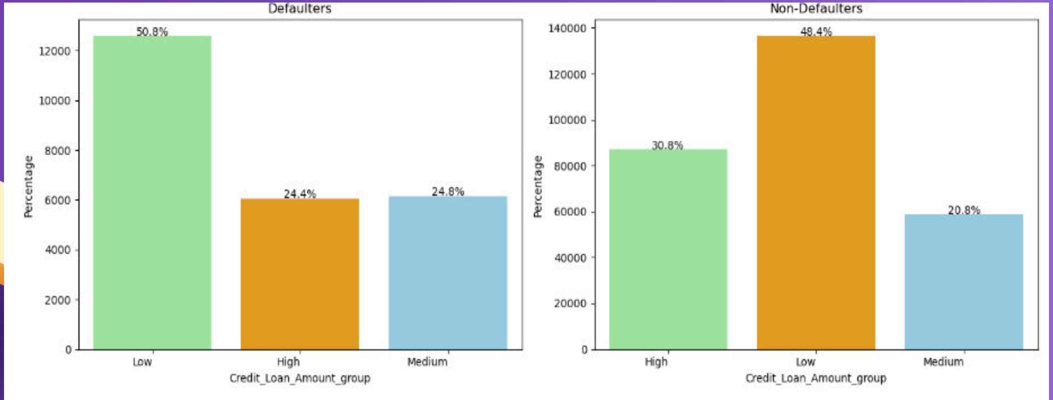
- Higher education reduces default risk – A greater percentage of non-defaulters (25.1%) have higher education compared to defaulters (16.1%).
- Most borrowers, whether defaulters or non-defaulters, have secondary education, but non-defaulters have a slightly lower proportion (70.3%) compared to defaulters (78.6%).
- Individuals with an academic degree have an extremely low default rate (0% among defaulters vs. 0.1% among non-defaulters), suggesting a possible correlation between higher qualifications and better financial responsibility.



Possible Interpretation – Marital Status

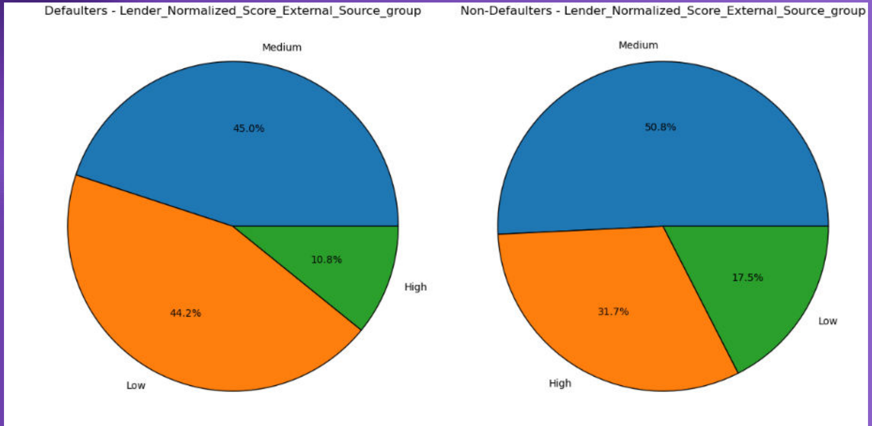
- Married individuals are less likely to default, as seen in their higher proportion among non-defaulters.
- Single and civil marriage individuals have a higher tendency to default, possibly due to financial instability or lower income sources.
- Separated individuals have a similar default rate as non-defaulters, indicating that separation itself is not a major risk factor.

Univariate Analysis - Defaulters and non-defaulters



Possible Interpretations – Credit Amount Group

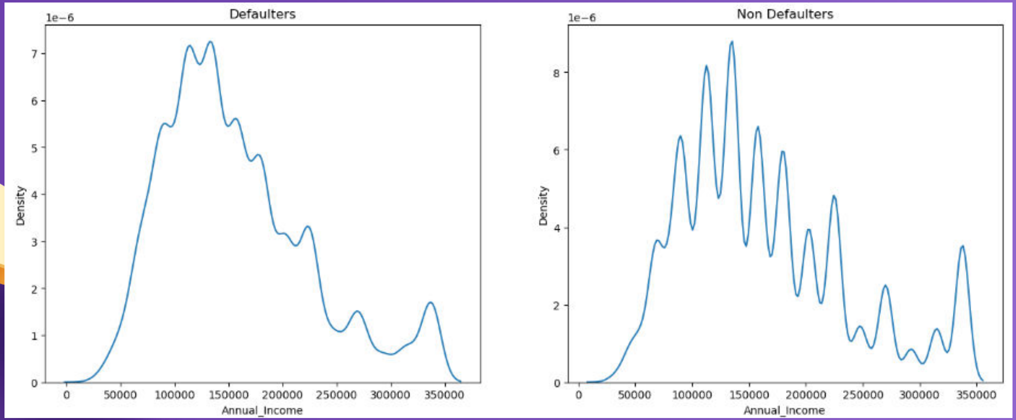
- Small loan borrowers have the highest risk of default.
- High loan borrowers are more likely to repay their loans successfully.
- Medium loan borrowers have an average risk level.

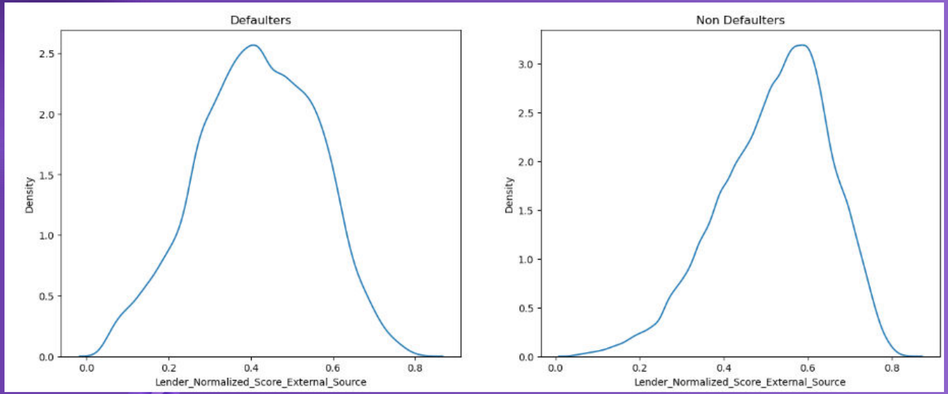


Possible Interpretation – Score of External Source

- Borrowers with low external credit scores are at a higher risk of default.
- A high external credit score is a strong indicator of repayment reliability.
- The medium credit score group has a mixed risk profile, so lenders should assess them carefully with additional factors.

Univariate Analysis (Continuous Variable) – Defaulters and non-defaulters





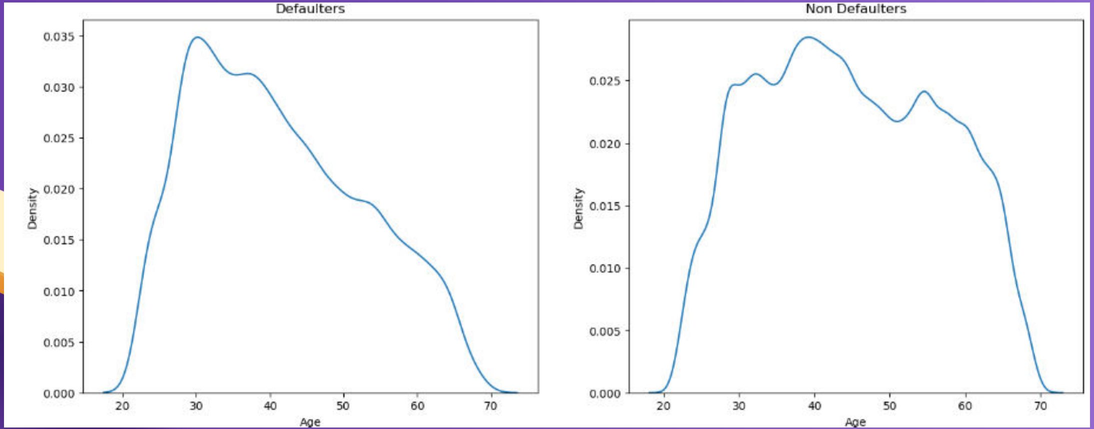
Possible Interpretations – Annual Income

- Lower income increases default risk – Defaulters are more concentrated in lower-income brackets, suggesting financial instability or difficulty in repayment.
- Non-defaulters tend to have a more structured income segmentation, reinforcing the idea that higher income is correlated with lower default rates..

Possible Interpretation – Normalized Credit Score

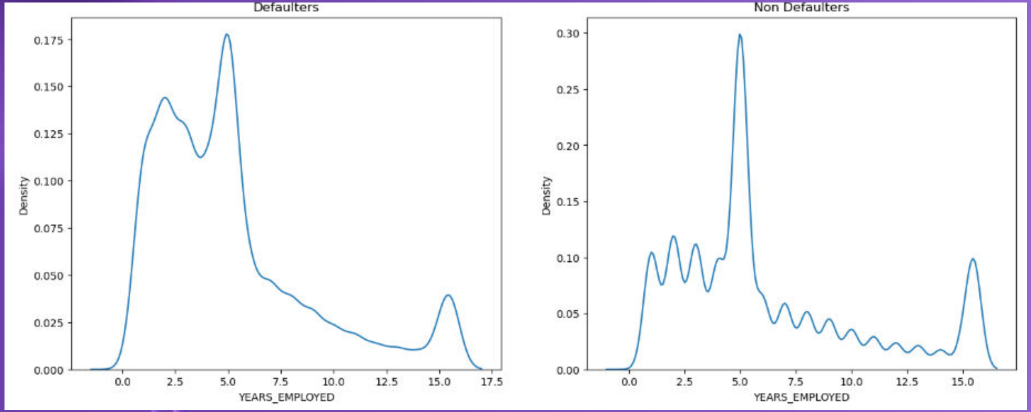
- Higher credit scores correlate with lower default risk, meaning credit score remains a strong predictor for loan performance.

Univariate Analysis (Continuous Variable) – Defaulters and non-defaulters



Possible Interpretations – Age

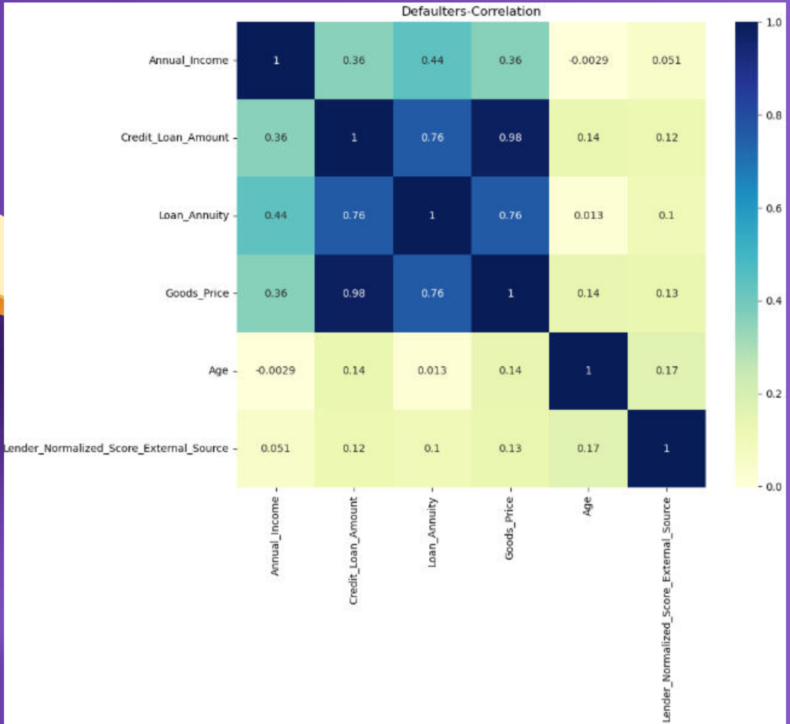
- Younger individuals (20-40 years) are more likely to default, possibly due to lower financial experience, lower income stability, or higher spending habits.
- Older individuals (50+ years) are less likely to default, suggesting they have better financial discipline, accumulated wealth, or stable employment.
- Medium loan borrowers have an average risk level.

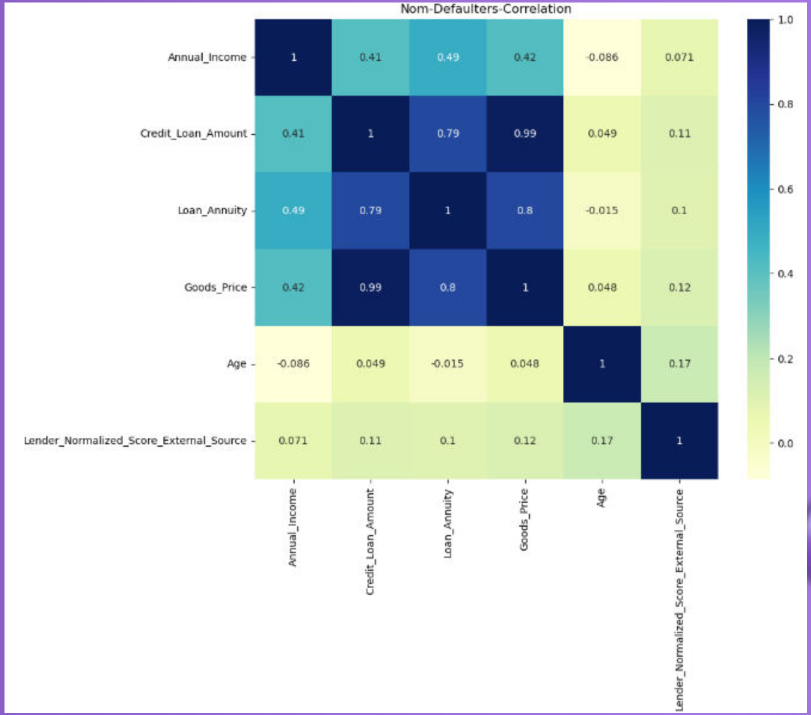


Possible Interpretation – Score of External Source

- Shorter employment tenure (0-5 years) is associated with higher default risk, likely due to job instability or lower income consistency.
- Borrowers with stable jobs (5+ years) tend to be non-defaulters, supporting the idea that employment stability is a key credit risk factor.
- The medium credit score group has a mixed risk profile, so lenders should assess them carefully with additional factors.

Correlation- Defaulters and non-defaulters



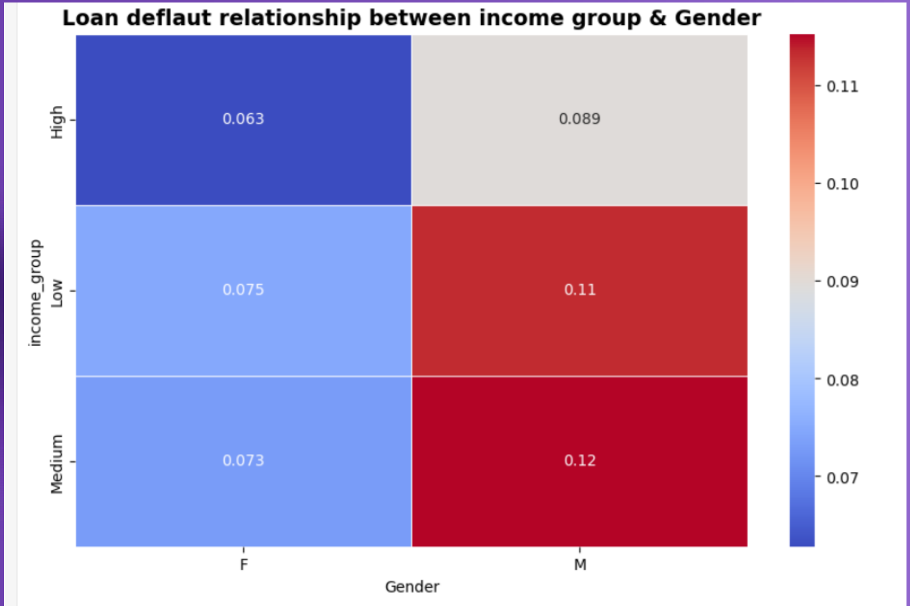


Possible Correlation

We can see that for both defaulters and non defaulters the same pairs of columns are highly corelated.
i.e., Credit_Loan_Amount and Loan_Annuity | Credit_Loan_Amount and Goods_Price | Loan_Annuity and Goods_Price
¶

Bivariate Analysis and Multivariate Analysis

Loan default relationship between income group & Gender



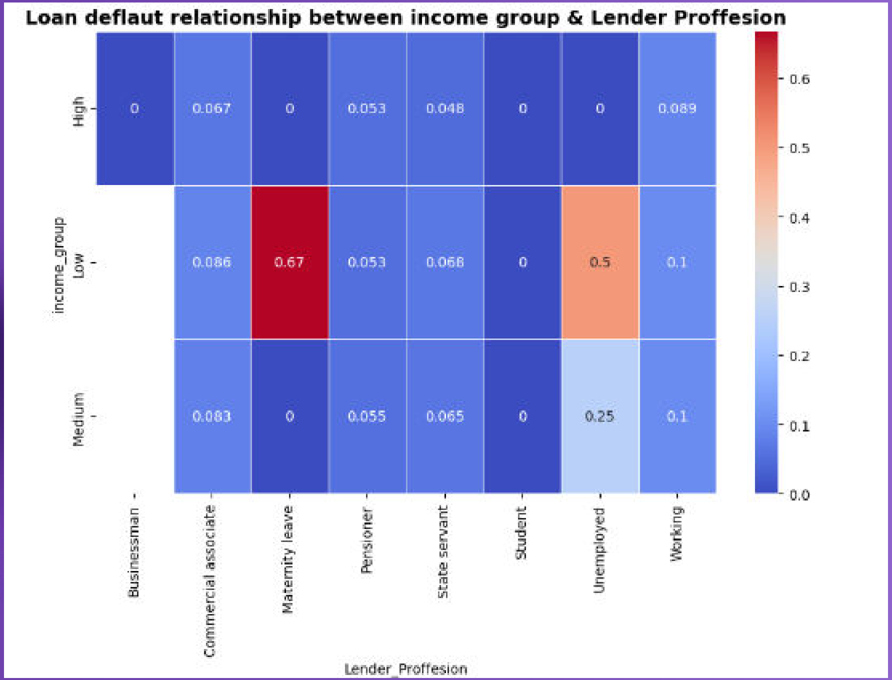
Key Observations

- Males have higher values across all income groups compared to females
- High-income group: Males (0.089) vs. Females (0.063)
- Medium-income group: Males (0.12) vs. Females (0.073)
- Low-income group: Males (0.11) vs. Females (0.075)
- This trend suggests that the metric (possibly loan default rate) is higher for males across all income categories.
- The highest value is observed in the Medium-Income Male category (0.12)
- This indicates that males in the medium-income category have the highest association with the given metric compared to other groups.
- The lowest value is observed in the High-Income Female category (0.063)
- This suggests that females in the high-income group have the lowest value in the dataset, which could mean they are less likely to default or have lower loan annuities.

Possible Interpretations

- **The metric represents loan default rates, males (especially in the medium-income category) might be at higher risk of defaulting.**
- **Females in the high-income group have the lowest observed value, indicating strong financial stability in this segment.**

Loan default relationship between income group & Lender's Profession



Key Observations

Maternity Leave & Loan Default:

- The highest default rate (0.67) is observed for individuals on maternity leave in the low-income group.
- This suggests that individuals on maternity leave in lower income brackets may face financial instability, making them more prone to loan defaults. Unemployed & Loan Default:
- Unemployed individuals in the low-income group also have a significantly high default rate (0.50).
- Medium-income unemployed individuals show a lower default rate (0.25), implying that slightly higher incomes may reduce the risk of default.

Working Professionals & Loan Default:

- The default rates for working professionals across all income groups are relatively low (~0.1), indicating stable financial conditions.

Pensioners & Loan Default:

- Pensioners across all income groups exhibit low default rates (~0.05), suggesting financial security from pension earnings.

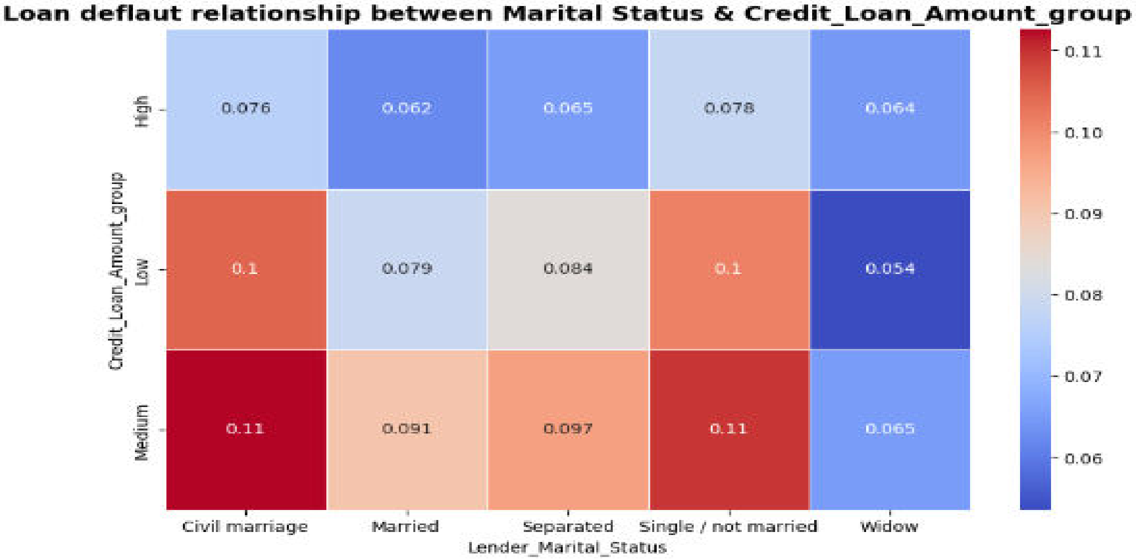
Businessmen & Commercial Associates:

- These groups show some level of defaulting (~0.08) but are not as high-risk as maternity leave or unemployed individuals.

Possible Interpretations

- Lenders should be cautious while approving loans for low-income individuals on maternity leave or unemployed persons as they have the highest default rates.
- Pensioners are relatively low-risk borrowers, making them more reliable loan applicants.
- Income level significantly impacts loan defaults, with higher income groups generally showing lower default rates.

Loan default relationship between Marital Status and Credit Loan Amount group of the loan



Marital Status & Credit_Loan_Amount_group

People who are Single or have done Civil Marriage with medium credit amount are likely to deflault

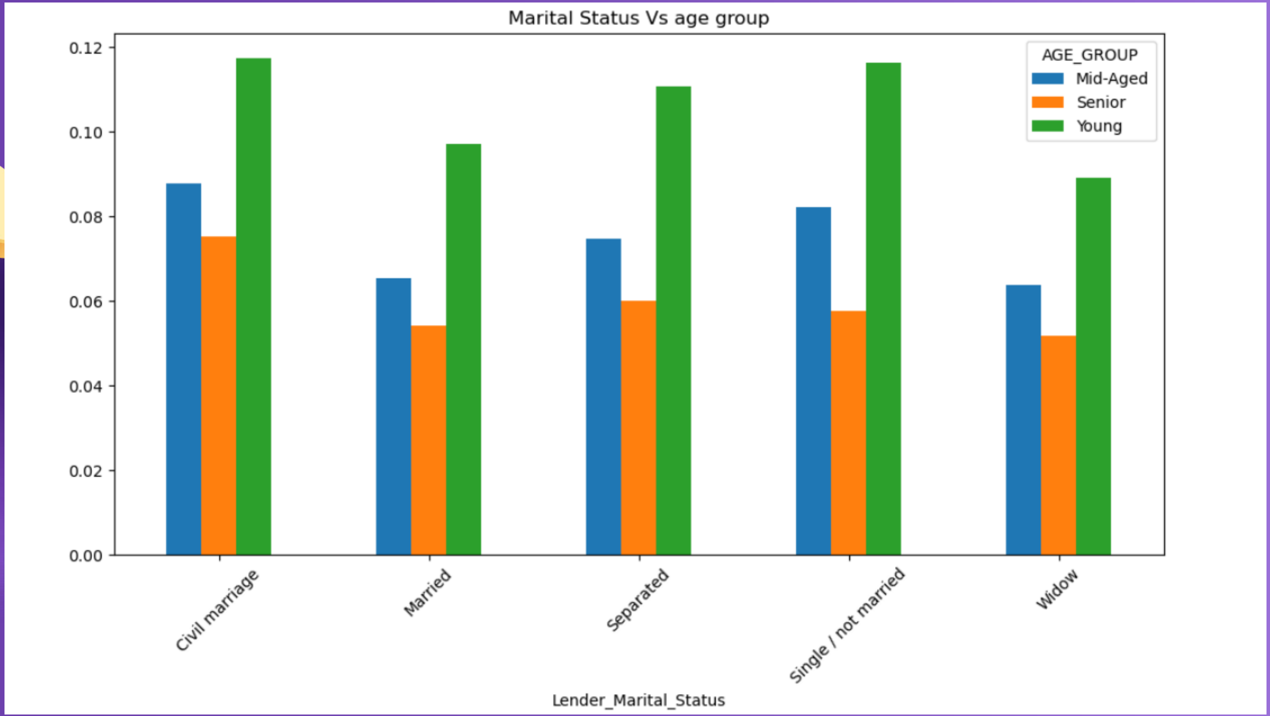
All the high credit amount loans taken by a person irrespective of the marital status are likely to be repaid.

Recommendation:

Prefer lending to high or low credit loans to married, seperated or widow.

Be cautious with Single individuals or indivisuals who have done civil marriage.

Loan default relationship between Marital Status group & Age Group



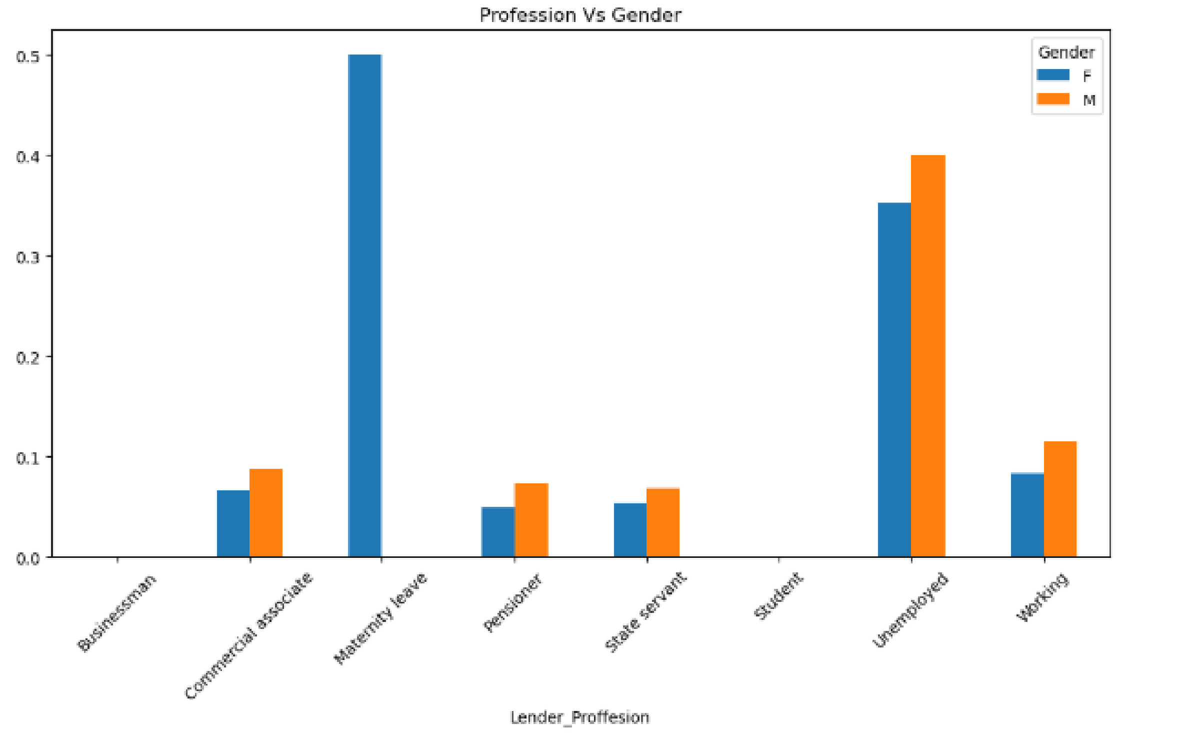
Key Observations

- Young individuals have the highest default rate, especially if they had civil marriage or are single.
- Mid-aged and senior individuals have lower default rates.

Possible Interpretations

- **Prefer lending to mid-aged and senior applicants.**
- **Be cautious with young individuals.**

Loan default relationship between Profession and Gender



Key Observations

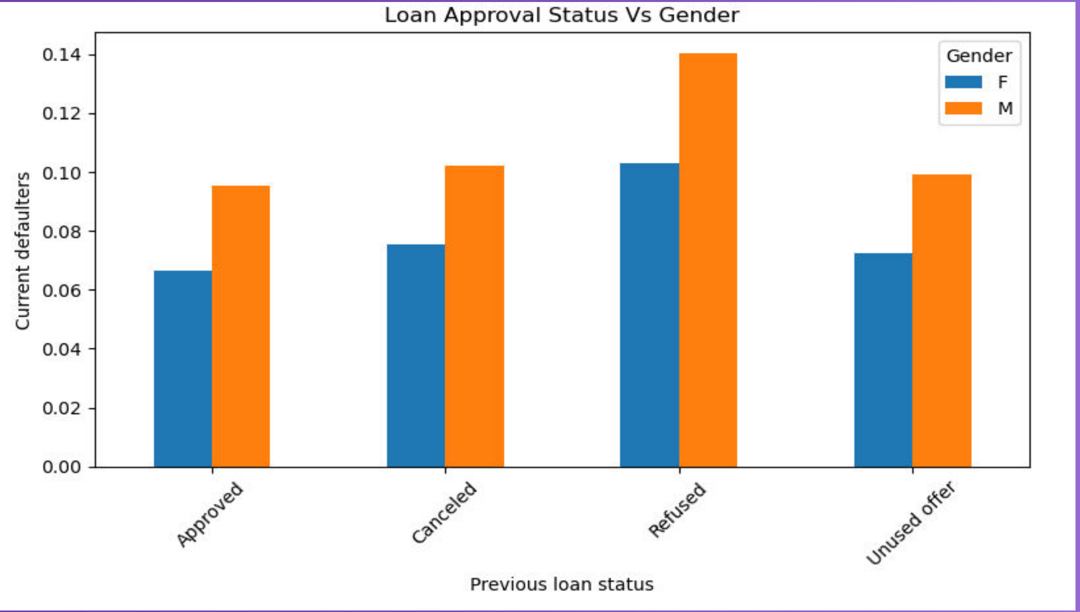
- Maternity leave and unemployed individuals have very high default rates.
- Businessmen and students are more stable.

Possible Interpretations

- **Approve loans for students and businessmen.**
- **Avoid lending to unemployed individuals, and those on maternity leave..**

Relationship between Current and Previous Application - Dataset

Loan default relationship between Loan Approval and Gender



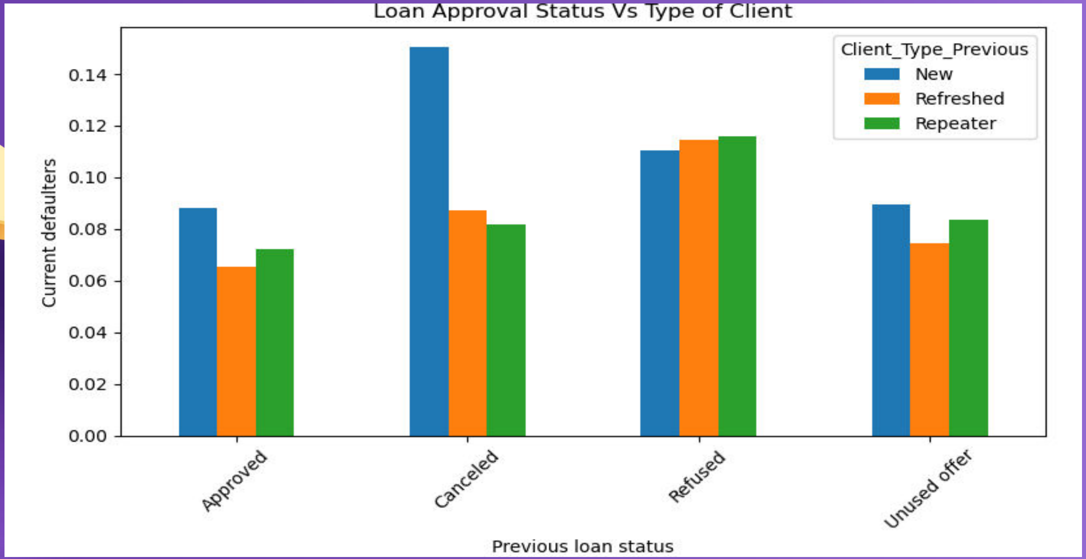
Key Observations

- We see that previously Refused client is more defaulted than previously Approved clients. Also, in all the cases the Males are more defaulted than Females.

Possible Interpretations

- **Avoid or be cautious while approving loans to people who are previously Refused.**
- **Overall, males are the one who are defaulting more than woman. Be cautious while approving loan to them.**

Loan default relationship between Loan Approval and Client's Type



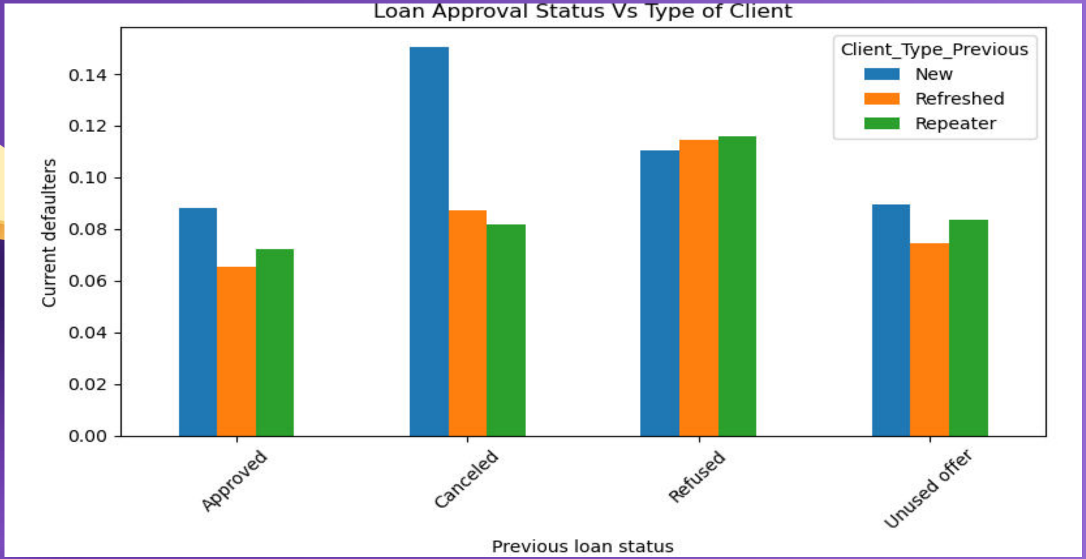
Key Observations

- New Clients Have the Highest Default Rates in Most Categories.
- Approved and Unused Offers Have the Lowest Default Rates.

Possible Interpretations

- **Give Loans to the** clients with a history of Approved or Unused loans, especially those classified as Refreshed or Repeaters..
- New Clients with a Canceled or Refused loan history pose the highest default risk.
- While giving loan Lenders should prioritize Refreshed and Repeater clients with a strong loan history while minimizing exposure to New clients with previous Canceled or Refused loans.

Loan default relationship between previous loan application status and 'Lender Profession'



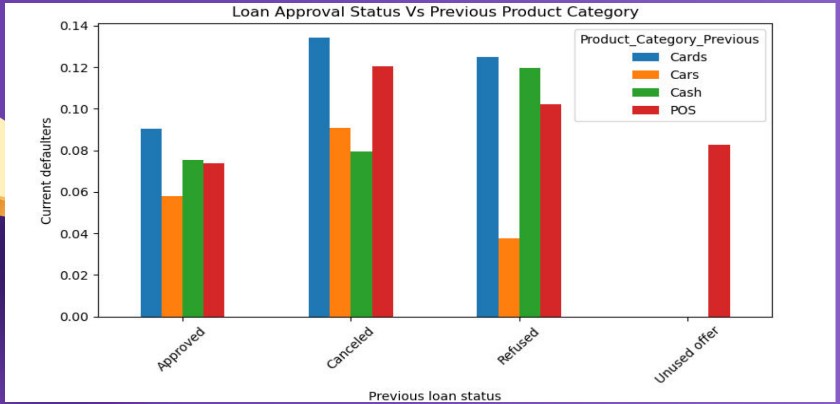
Key Observations

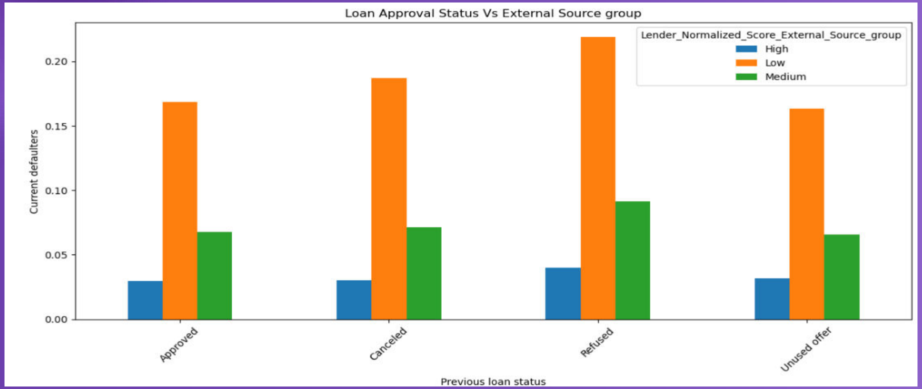
- Maternity leave applicants have the highest default rates.
- Unemployed also show significant risk.
- State servants and pensioners have low default rates.

Recommendations:

- ✅ Give Loans To: Pensioners, State Servants, Working individuals.
- ❌ Avoid Loans To: Maternity leave, Unemployed applicants loans.

Loan default relationship between (previous loan application status Vs Lender Profession) & (previous loan application status Vs External Source group Score)





Key Observations

- ❖ Card users and POS users have higher default rates than others..
- ❖ Previously refused application for cash are deflated at higher rate.
- ❖ Low lender score groups have extremely high default rates.

Recommendations:

- ❖ ✅ Give Loans To: Safer to give loan to the people who are previously approved
- ❖ ❌ Avoid Loans To: Maternity leave, Unemployed applicants. Also, it's a high risk to grant loans for applicants who have poor external score specially who's loans were previously refused, unused and cancelled.

Conclusion

Provide loan to:

- Individual having Higher education! As, Higher education reduces default risk – A greater percentage of non-defaulters have higher education or Academic degrees compared to defaulters
- "Pensioners", "Students", "businessmen" show better financial discipline, as their share among non-defaulters is higher.
- High loan borrowers are more likely to repay their loans successfully.
- Borrowers with high external credit scores is a strong indicator of repayment reliability.
- Approved Clients in the previous applications is a strong indicator of repayment reliability
- Married Clients are likely to repay their loans successfully.
- Females are comparatively Favorable than males.
- Senior or old-aged borrowers are more likely to repay their loans successfully.
- Prefer lending to mid-aged applicants, especially for higher loan amounts.
- Borrowers with stable jobs (5+ years) tend to be non-defaulters, supporting the idea that employment stability is a key credit risk factor.

Avoid providing loan to:

- Males (especially in the medium or low-income category) have a relatively higher likelihood of defaulting .
- Small loan borrowers have the highest risk of default.
- Single and civil marriage individuals have a higher tendency to default, due to lower income sources.
- Borrowers with low external credit scores are at a higher risk of default.
- Young borrowers are more likely to default if they had civil marriage or are single.
- Lenders should be cautious while approving loans for low-income individuals on maternity leave or unemployed persons as they have the highest default rates.
- Be cautious with young individuals in the low-credit category
- Shorter employment tenure (0-5 years) is associated with higher default risk, likely due to job instability or lower income consistency.

Recommendation

- Prioritize higher credit score applicants for approvals or offer them better interest rates.
- Perform deeper risk assessments (e.g., income verification, spending patterns) for medium-score borrowers.
- Apply stricter risk mitigation measures (collateral, co-signers, financial literacy programs) for low-score borrowers.
- Lending institutions should focus on financial literacy and risk assessment for younger borrowers to minimize default rates.
- Lenders may need to impose stricter eligibility criteria for applicants with less than 2 years of employment, or require additional financial guarantees.
- If you're a lender, monitor loans in the range of 200,000 - 500,000 closely, as both defaulters and non-defaulters frequently take loans in this range.

THANKS!

QUESTIONS?