



# RMIT BUSINESS ANALYTICS CHAMPION

## SEASON 5

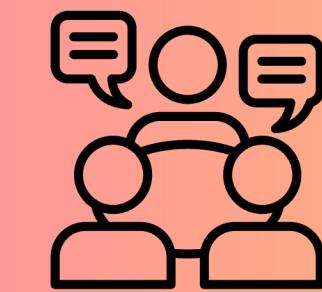
# ROUND 2 CASE STUDY

**TEAM NAME:**

**THE CROWN ANALYSTS**

**TEAM MEMBERS:**

- Ngo Thi My Duyen
- Le Hoang Uyen Thu
- To Gia Bao
- Phan Van Hoang



# CONTENTS



## Data Preparation

- Cleaning data
- Data transformation
- Data integration
- Feature selection



## Exploratory Data Analysis (EDA)

- Detailed description of customers' demographic
- Examine the loans borrowed



## Interpretation & Insights

- Correlation between customer profiles and loan types
- Compare patterns between 2022 and 2023



## Suggestion

- Strategic suggestion to improve loan origination and sales
- Suggestions to increase loan origination and sales

# PROBLEM STATEMENT



- **Impact of COVID-19 on Business:** The pandemic severely affected financial performance and disrupted business operations in 2022 and 2023.
- **Shifting Customer Demographics:** Customer segments and behaviors have changed, requiring a deeper understanding of evolving needs and preferences.
- **Changes in Loan Portfolio:** The characteristics of loan products, including loan types and terms, have been altered, affecting loan origination and sales.

DISBURSEMENT_CHANNEL	LIQUIDITY	INSURANCE_COMPANY	INSURANCE_RATE	DISBURSEMENT_DATE	CONTRACT_NO
1	3	0	3.75	2023-09	571094
2	3	1	3.75	2023-09	569518
1	1	0	3.75	2023-09	569723
1	1	0	3.75	2023-09	569818
1	1	0	3.75	2023-09	569794
1	3	0	3.75	2023-09	810124
1	4	1	3.75	2023-09	858618
1	3	1	3.75		569915
1	2	0	3.75	2023-09	569760

- Total columns:** 17 columns
- Total samples:** 315,684 samples
- Missing data:** Each column has a different rate of missing data, with the highest missing rate at 29.32% (**COMPANY\_ADDRESS\_PROVINCE** column) and the lowest at 14.96% (**WEIGHT** column).

LABOUR_CONTRACT_TYPE	NUMBER_OF_DEPENDANTS	PERMANENT_ADDRESS_PROVINCE	INDUSTRY	JOB	COMPANY_ADDRESS_PROVINCE	CONTRACT_NO
4	3		18	9		571556
2	1	96		0		569818
2	0	48	18	1		569794
4	2	89	18	9		810070
4	3	42	18	1		569905
4	0	48	18	4		569915
4	0	70		6		569760
4	1	51	0	5		569600
4	0	56	18	7		810071
4	1		7	1		569729

# Cleaning data



## File loan\_origin.csv

Save to an excel file customers with:

- Missing values in CONTRACT\_NO
- Missing more than 50% values
- LIQUIDITY equal to 0
- Remove outliers from the non-null LOAN\_AMOUNT group.
- Update HAS\_INSURANCE based on INSURANCE\_RATE

## File demographic.csv

Save to an excel file customers with:

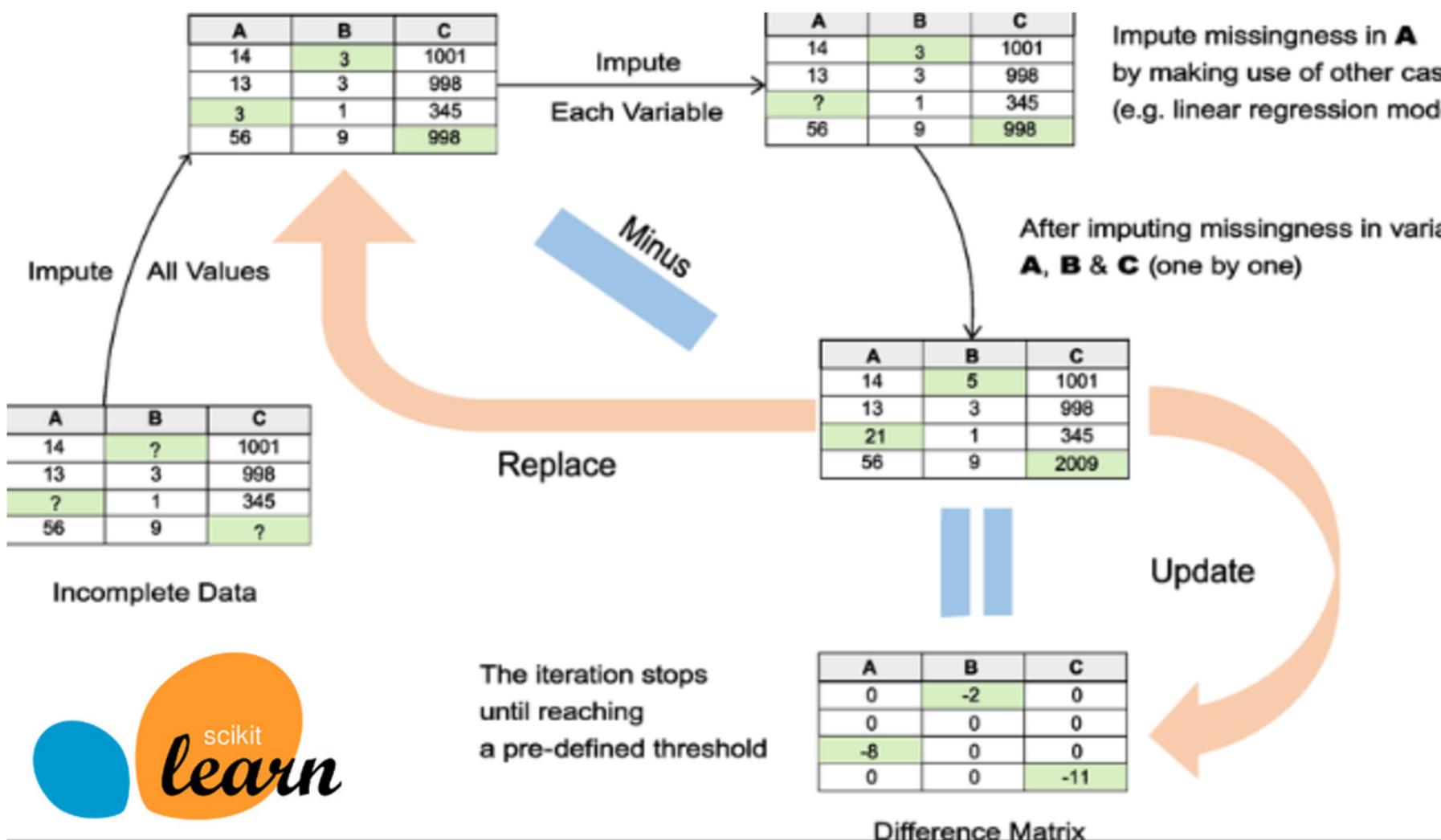
- Missing values in CONTRACT\_NO
- Missing more than 50% values
- CREDITIBILITY equal to 0
- Filter and Save Rows with Low CUSTOMER\_INCOME.
- Update HAS\_INSURANCE based on INSURANCE\_RATE

Calculate and Add Age

Fill Missing Values for HEIGHT and WEIGHT, and calculate BMI

# Iterative Imputing

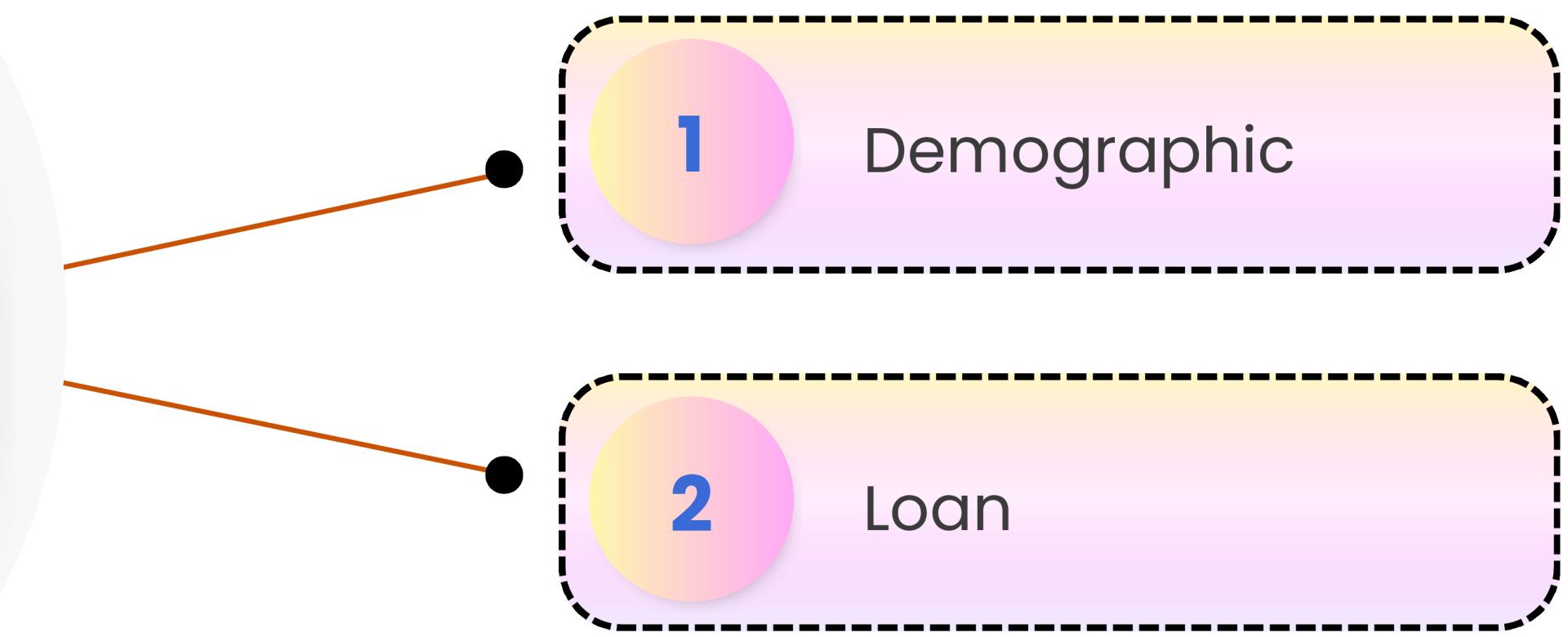
**We used Iterative Imputation combined with machine learning models (XGBoost with GPU and LightGBM) to efficiently handle missing data in specific columns:**



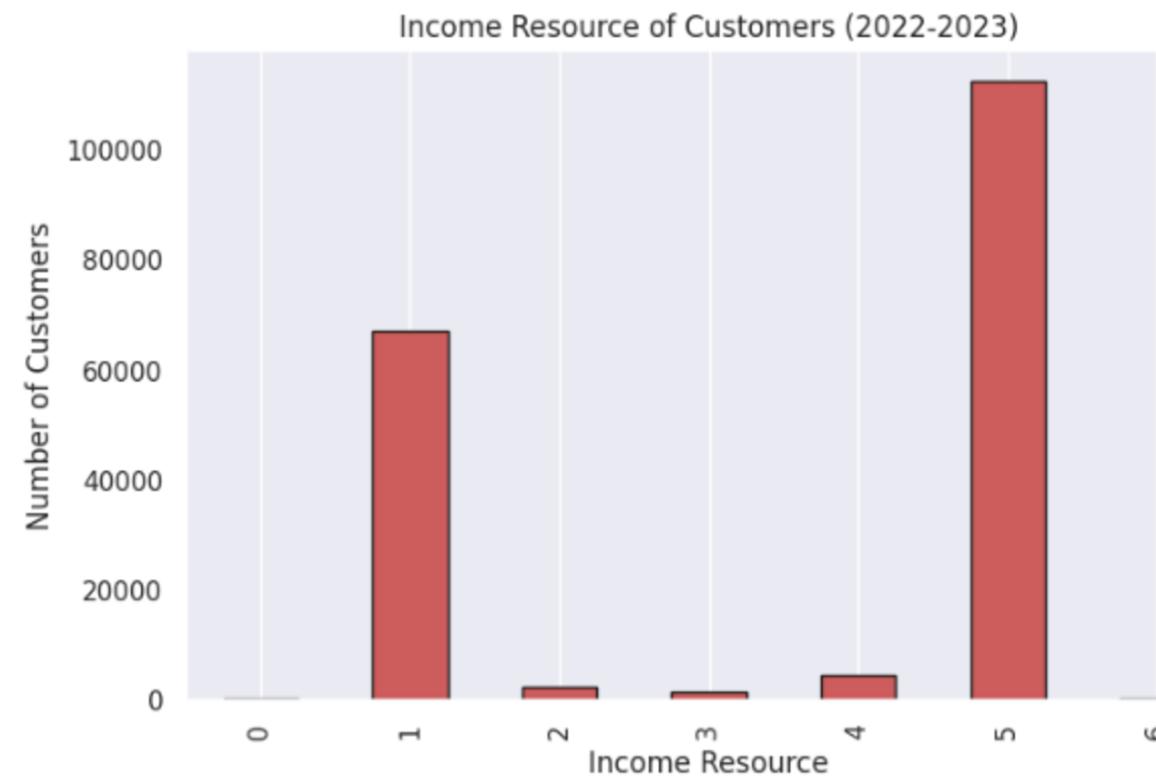
- **Model Setup:** an estimator (XGBoost or LightGBM) was chosen to predict missing values based on other available features.
- **Iterative Imputation:** IterativeImputer was applied to iteratively predict and impute missing values, running multiple iterations to refine predictions.
- **Categorical Handling:** Columns treated as categorical were rounded to integers after imputation.
- **Data Update:** The original dataset was updated with the imputed values.



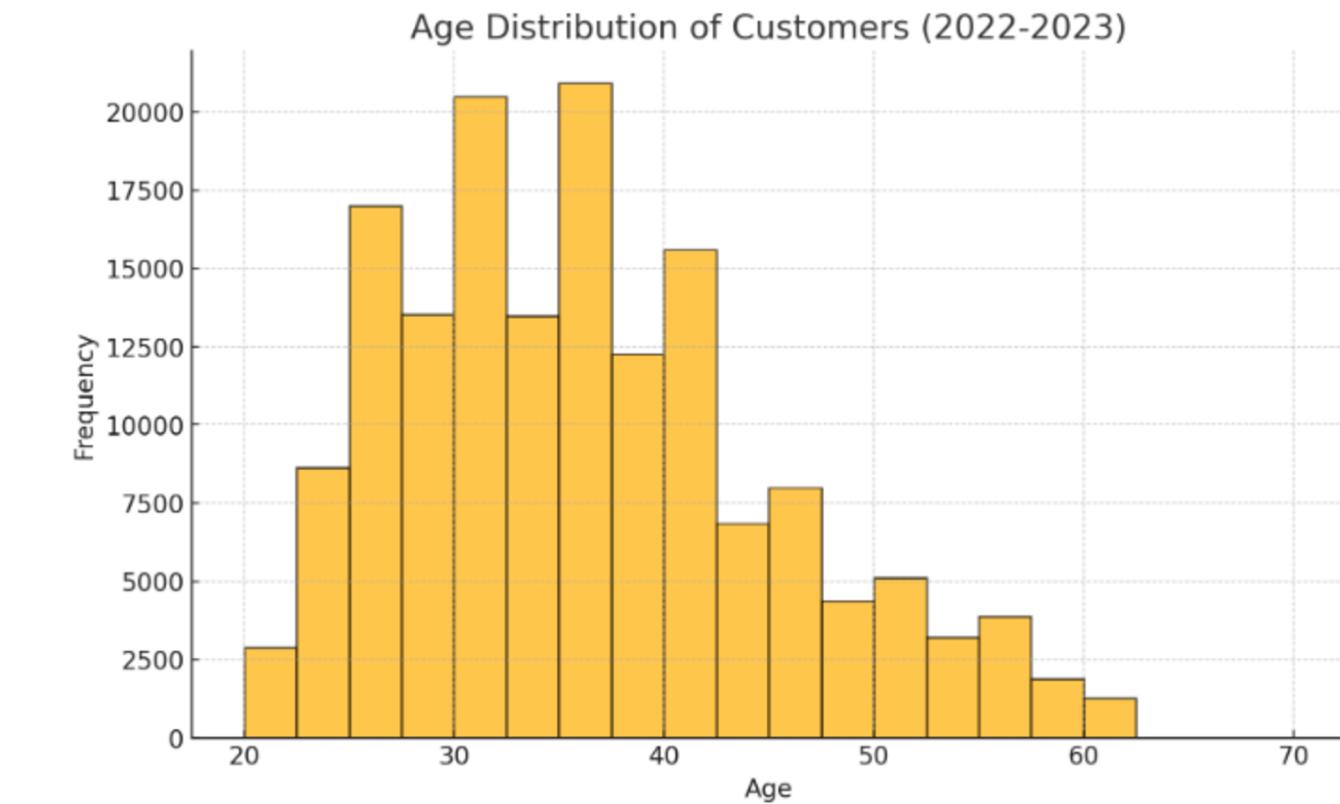
# Exploratory Data Analysis (EDA)



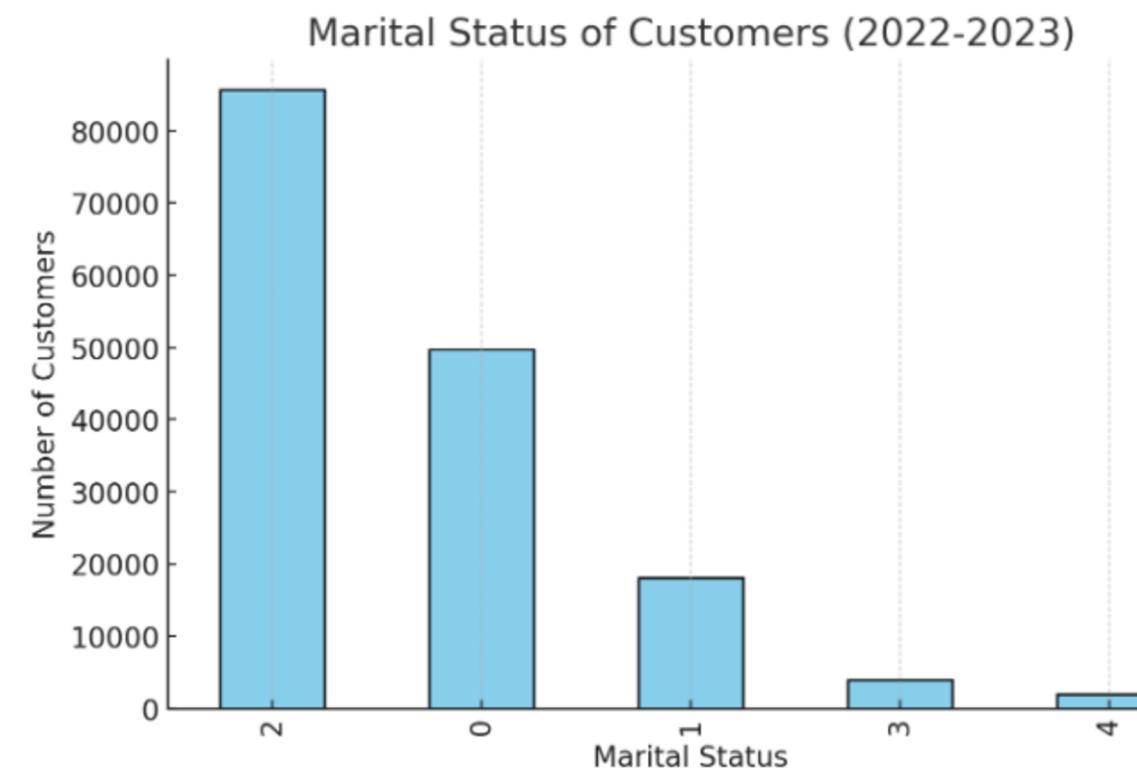
# Demographic



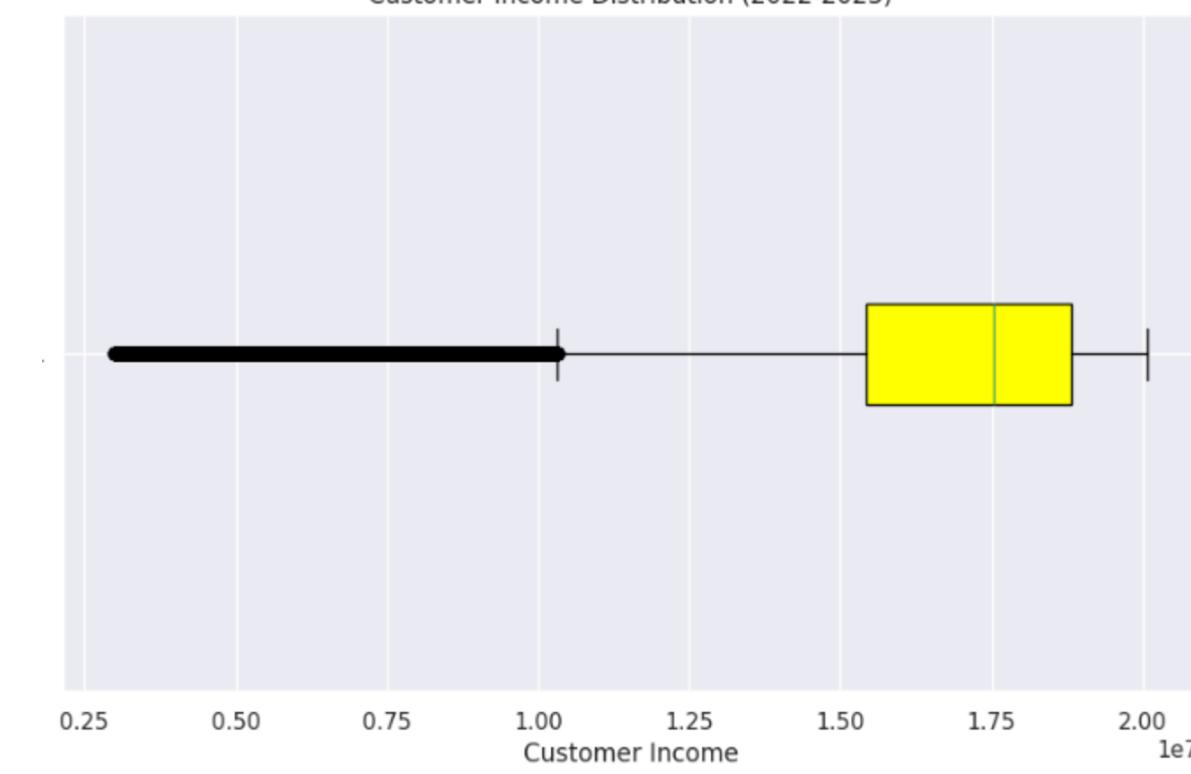
**Income Source:** Customer income sources are categorized, with the majority coming from type 1, 5, while small amount of customers derive their income from other sources like type 2, 3, and 4.



The age of customers ranges widely, with a peak in the 30 to 40 age group, reflecting a core customer segment in the working age demographic.

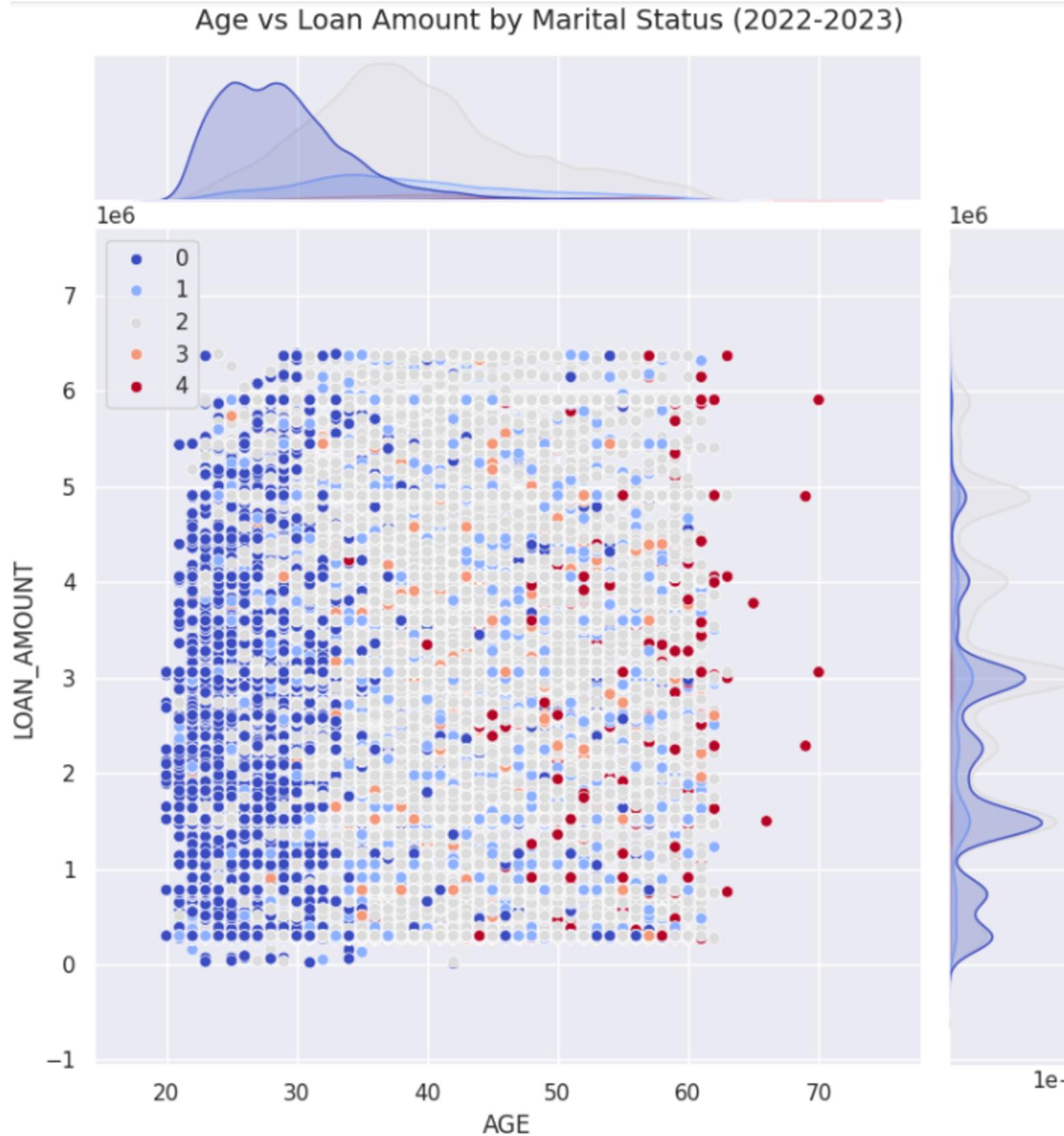


**Marital Status:** The majority of customers group 0, 2, while other groups such as 1, 3, 4 account for a smaller proportion.



A box plot shows that the average income falls within the range of 15 to 18 million VND, with some outliers having higher income levels.

# Age vs Loan Amount by Marital Status



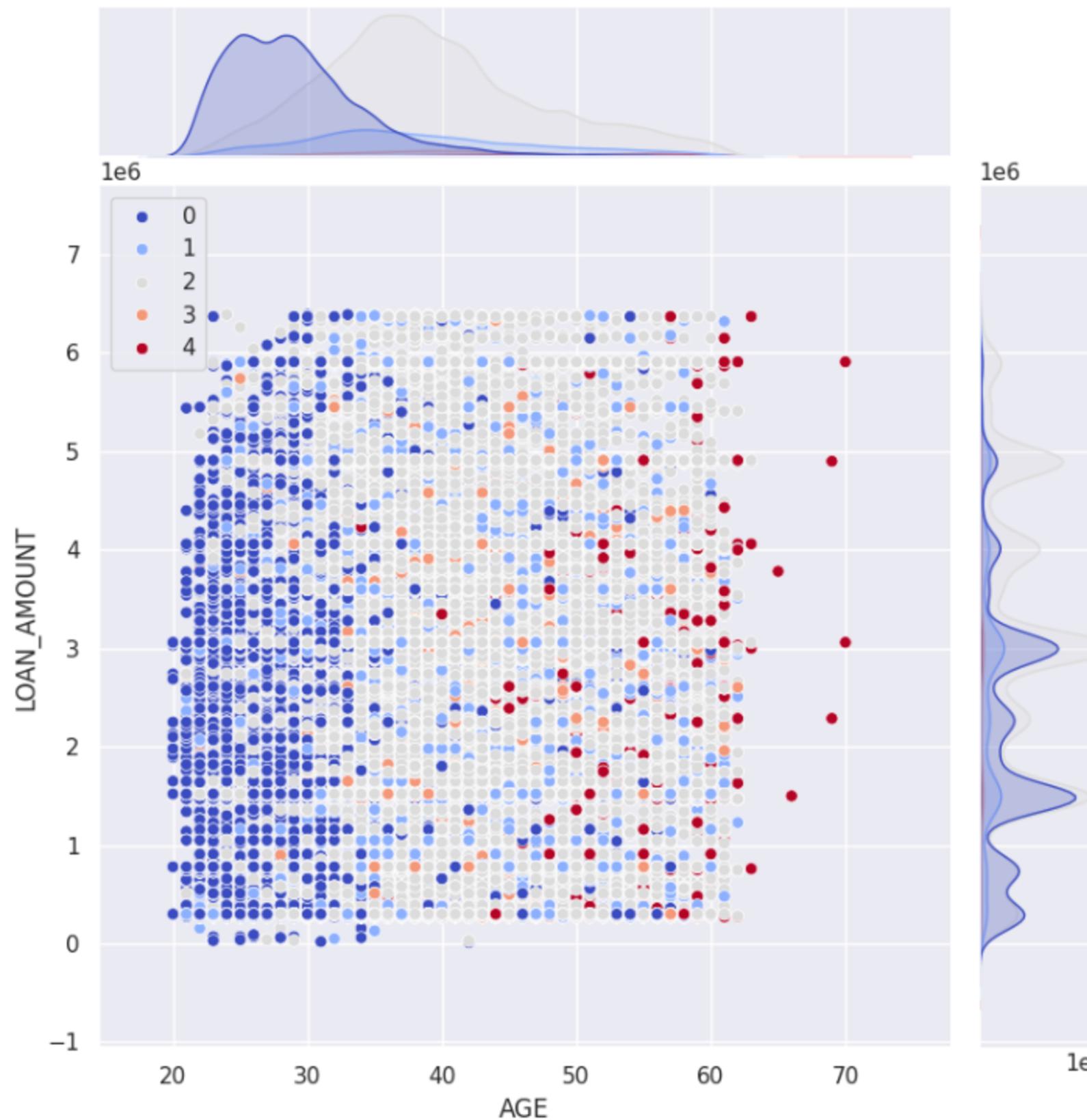
## Marital Status Influence:

- Single (0): This group has a wide range of loan amounts, with a focus on smaller loans.
- Married (1 & 2): Married customers dominate the middle-aged category with larger loan amounts. This might be due to family needs such as buying homes.
- Divorced/Widowed (3 & 4): These categories also show a spread of loan amounts. They might be securing loans for retirement planning or lifestyle stability after life changes.

## Loan Amount Across Age Groups:

- Younger customers (aged 20-30) generally have smaller loan amounts, suggesting that they might be early in their careers or less financially established.
- As age increases (30-50), loan amounts seem to vary widely, potentially indicating more financial stability and diverse needs.
- Older customers (50+) tend to have larger loans, possibly for long-term investments or business purposes.

Age vs Loan Amount by Marital Status (2022-2023)



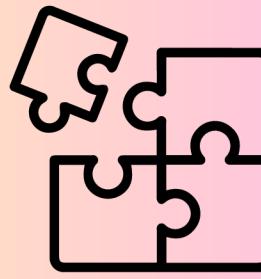
## Age Group Patterns:

- For ages 20-40, there is a higher density of medium to small loans across all marital statuses.
- For ages 40+, a more scattered distribution of higher loan amounts emerges, particularly in the married and divorced/widowed categories.

## General Trends

- Loan size tends to increase with age, likely due to increased financial capacity and stability.
- Marital status might influence loan behavior: married individuals possibly take out loans for joint purposes (family, housing), while single individuals focus on personal needs.

# Detail of loan borrowed



## SUMMARY STATISTIC

	PRODUCT_CATEGORY	LOAN_TERM	MONTH_INTEREST	LOAN_PURPOSE	LOAN_AMOUNT
count	270850.00	270850.00	270850.00	270850.00	270850.00
mean	2.64	24.83	2.01	3.60	2536189.23
std	2.91	8.31	0.60	1.48	1497609.60
min	0.00	0.00	0.00	0.00	10000.00
25%	1.00	18.00	1.70	3.00	1470000.00
50%	2.00	24.00	2.20	3.00	2385645.00
75%	4.00	36.00	2.50	4.00	3540000.00
max	18.00	36.00	3.50	6.00	6390000.00

- **Large sample size:** (270,850), providing reliable information.
- **Average loan term:** is about 25 months, and the \*\*average interest rate\*\* is 2.01% per month.
- **Loan amounts:** vary greatly (from 10,000 to 63,000,000 VND) with an average of about 2.36 million VND, indicating a diverse range of borrowing needs.
- **High variability:** in loan terms and amounts, reflecting the differences in customers' financial capabilities.

A member of

# Detail of loan borrowed

**Table 1: The average LOAN\_AMOUNT for each combination of PRODUCT\_CATEGORY and LOAN\_PURPOSE.**

- Key pattern:** The loan amounts generally increase as the product category index increases, suggesting larger loan amounts for more advanced.
- Trend:** The mean loan amount tends to increase from left to right on x-axis, indicating that higher-numbered product categories tend to have higher average loan amounts.
- Data anomalies:** The Shopping loan purpose shows a large spike in mean loan amount for product category 18, which is higher than the surrounding data points for that loan purpose.

Table 1: Mean Loan Amount by Product Category and Loan Purpose

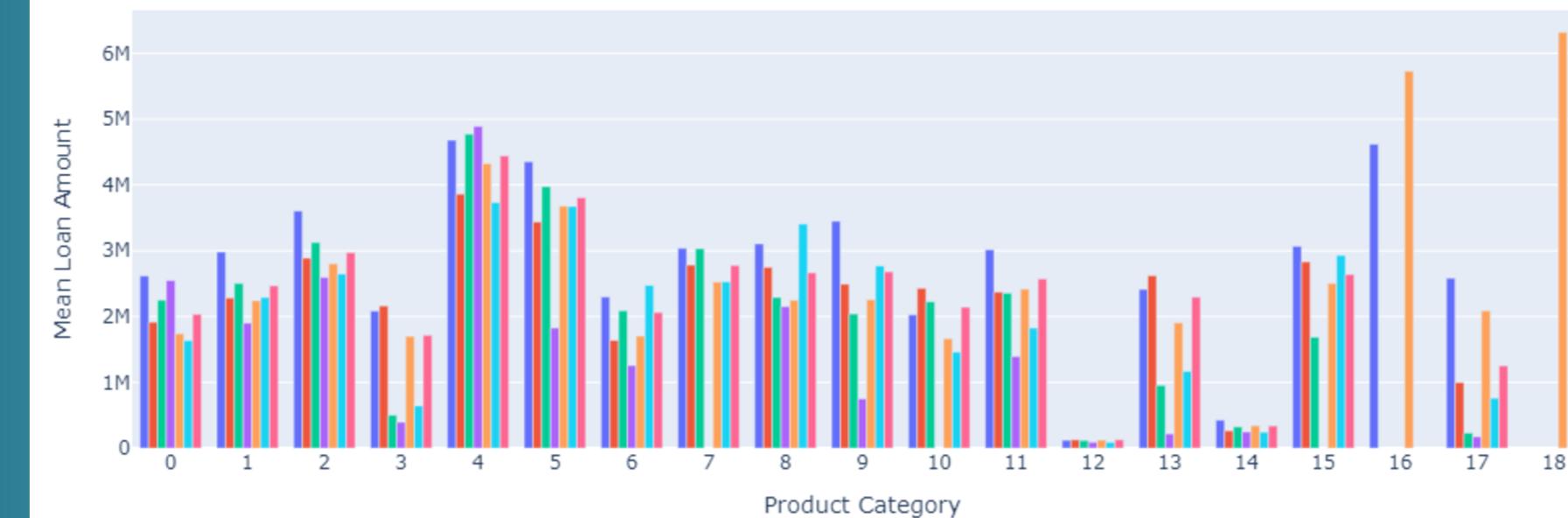
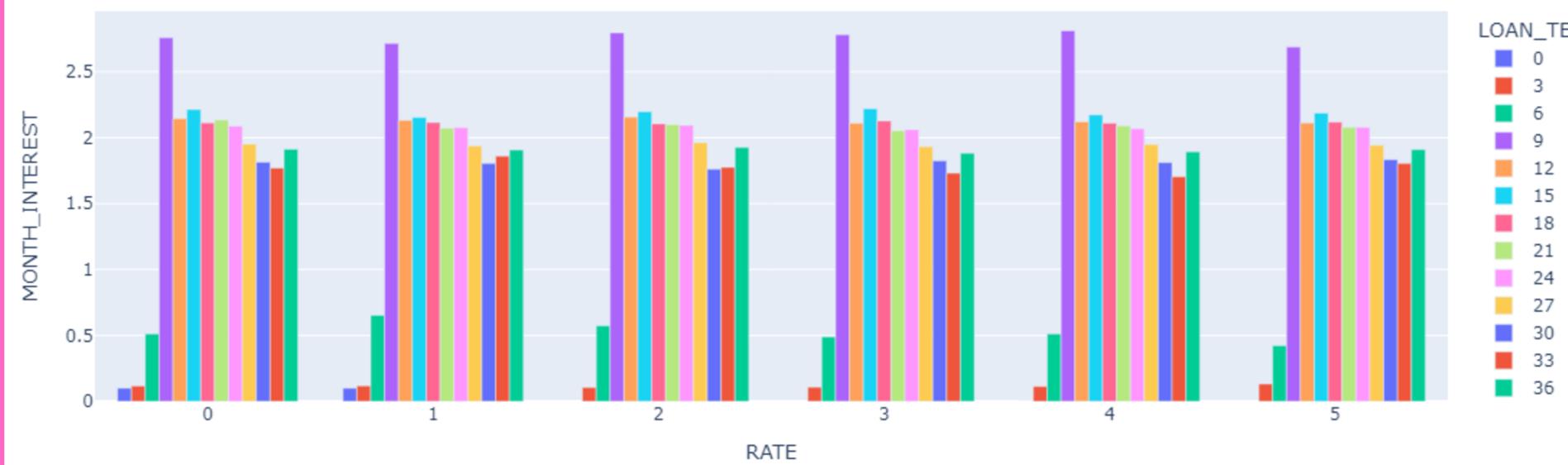


Table 2: Monthly Interest by Rate and Loan Term



**Table 2: the average MONTH\_INTEREST for each combination of RATE and LOAN\_TERM.**

- Key pattern:** There is a clear diagonal band of higher interest rates as the loan term increases, indicating that longer loan terms generally have higher monthly interest rates.
- Trend:** As the interest rate increases from 0 to 5, the monthly interest generally increases across all loan term lengths.
- Data anomalies:** The 27-month loan term has slightly higher monthly interest compared to the 30-month term at some interest rate levels, which is an outlier given the general trend.

# Detail of loan borrowed

Table 3: Monthly Interest by Loan Amount Range and Rate



**Table 3: the average `MONTH\_INTEREST` for each combination of `LOAN\_AMOUNT\_RANGE` and `RATE`**

- **Key pattern:** The graph shows the monthly interest by loan amount range and rate, categorized into four loan amount ranges: Small, Medium, Large, and Very Large.
- **Trend:** The interest rate follows a consistent trend, increasing from the lowest rate to the highest rate within each loan amount range.
- **Data anomalies:** There are no obvious anomalies in the data presented in the graph.

Table 4: Average Loan Amount by Loan Term



**Table 4: Average Loan Amount by Loan Term**

- **Key pattern:** The loan amount increases as the loan term length increases.
- **Trend:** The graph has a smooth, consistent curve, indicating a linear relationship between loan term and average loan amount.
- **Data anomalies:** The graph shows a clear and gradual increase in the average loan amount as the loan term gets longer, without any major deviations or anomalies.



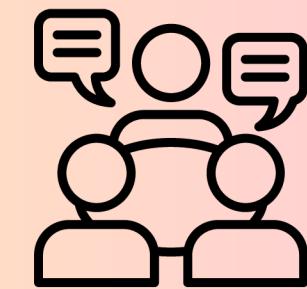
## Issues Observed



## Advantages and Challenges in Loan Structuring



## Compare Patterns between 2022 and 2023



# Interpretation & Insights

"Imagine this finance company as a careful bridge-builder. Each loan is a bridge connecting people to their goals—but some bridges may be less stable, especially when built without a strong foundation of data."

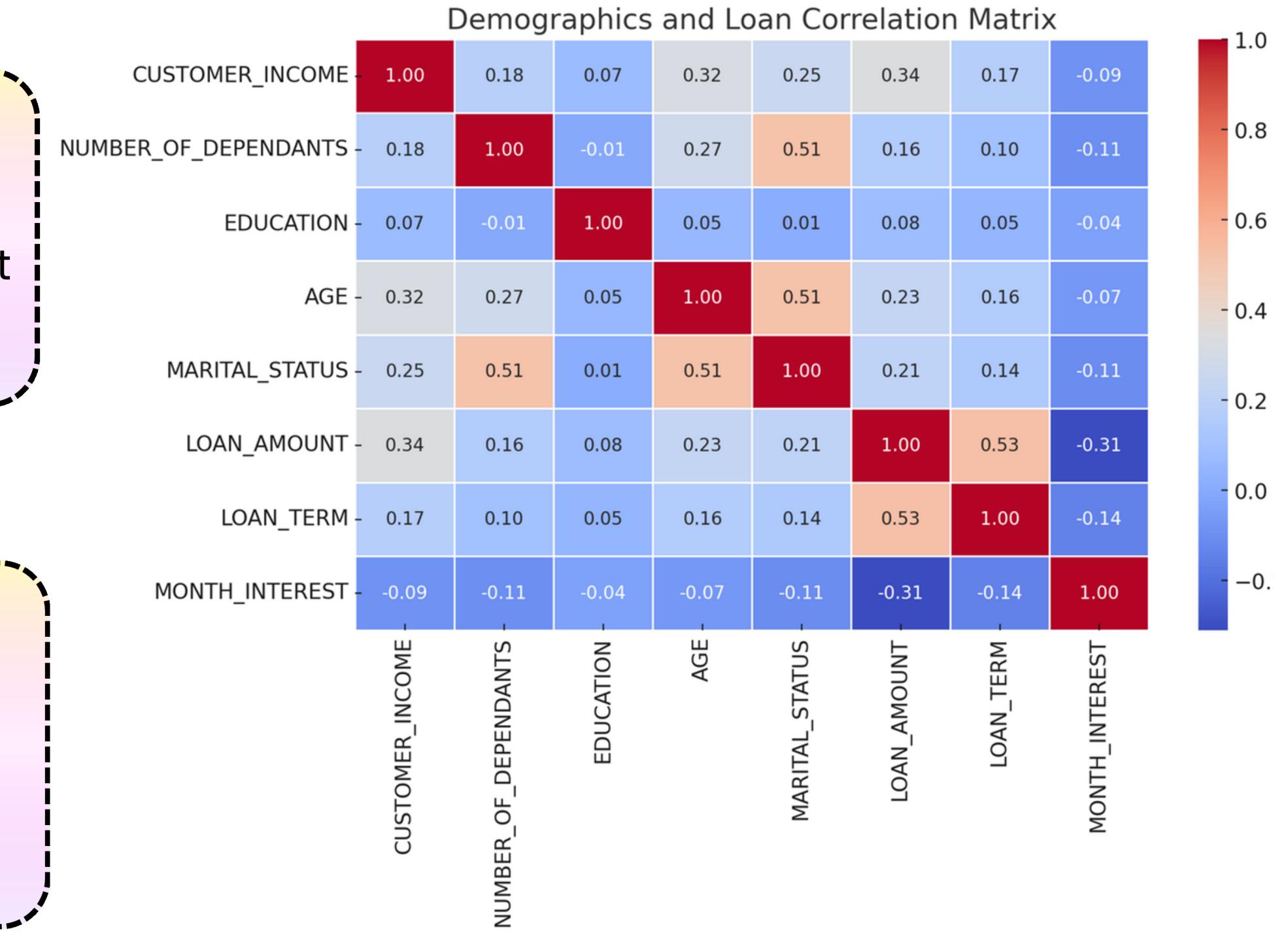
# Advantages and Challenges in Loan Structuring

## Advantage

- Targeted Interest Rates: Adjusted by demographics, aiding risk pricing.
- Broad Demographic Reach: Diverse client base across demographics.

## Challenges

- High Default Risk: Weak income-loan correlation could mean large loans to low-income clients.
- Limited Demographic Use: Weak data integration for tailored loan products.



# Issues observed

## "Risky bridges"

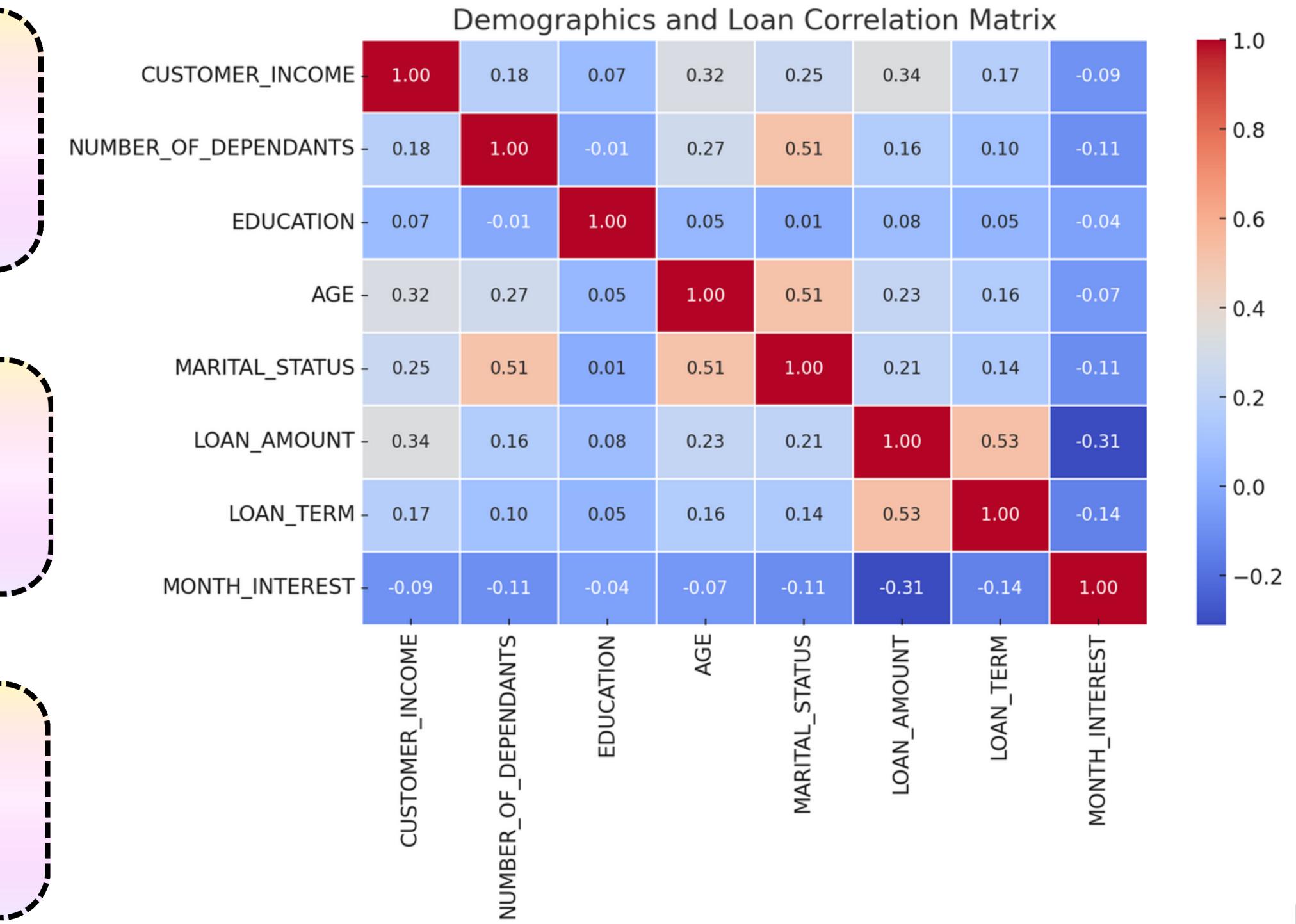
With only a moderate link between income and loan amount, some low-income customers might receive larger loans, increasing their risk of default

## Dependents' Impact

Customers with more dependents often receive longer terms, potentially straining their finances

## Higher Interest Rates for the Young

Lower-income or younger clients face higher monthly interest rates, raising the likelihood of financial strain for this group.



# Strategic Suggestion



## Build Stronger Bridges

Cap loan sizes based on income and dependents, ensuring customers are borrowing within their means



## Harness Data

Regularly use advanced analytics to uncover hidden patterns and evolve loan policies in real-time.



## Design for Diversity

Develop loan products customized for various customer segments, providing manageable terms that align with their specific needs.

# EXECUTIVE SUMMARY

In navigating the challenges posed by the evolving financial landscape, SHBFinance remains committed to supporting low and middle-income individuals through accessible financial solutions. Our analysis provides a clear understanding of SHBFinance's customer profiles, loan characteristics, and the impact of recent market shifts on lending activities. By identifying trends and correlations between customer segments and loan usage:

- **Customer Demographic Shifts:**

- The majority of loan demand shifted to middle-aged and older customers, particularly married individuals seeking larger, long-term loans.

- **Loan Portfolio Evolution:**

- A notable increase in loan amounts was observed among family-oriented and investment-related products.
- Customers with stable income sources, such as those in formal employment or joint family loans, contributed significantly to portfolio stability.

- **Risk and Performance Insights:**

- Married customers demonstrated more consistent repayment behavior, mitigating risk for higher-value loans.
- Single and younger borrowers, while fewer in number, presented higher credit risk, requiring stricter credit policies.

# Reference

1 Immore (2020) Intro to Exploratory data analysis (EDA) in Python.

<https://www.kaggle.com/code/imoore/intro-to-exploratory-data-analysis-eda-in-python>

2 Rob Dabbs (2023, March 17) The importance of Business Insight in Data interpretation.

<https://www.linkedin.com/pulse/importance-business-insight-data-interpretation-rob-dabbs>

3 Kyle Peterdy (?) Loan Structure.

<https://corporatefinanceinstitute.com/resources/commercial-lending/loan-structure/>



"Sincere thanks to everyone for watching. I hope the presentation provided valuable information and inspiration. I look forward to receiving your valuable feedback!"