HOME CREDIT

DATA ANALYSIS

PRESENTATION 2025

"Success is not about one thing. It's about a hundred things 1% better."

Sir Dave Brailsford

039 298 1368

hst.sytan@gmail.com

HOME HOME
CREDIT CREDIT

Agenda

1. INTRODUCTION
2. DATA EVALUATION AND PROCESSING
3. LOAN CONTRACT TYPE ANALYSIS
4. LOAN VALUE ANALYSIS
5. RISK ANALYSIS
6. COMPREHENSIVE ANALYSIS
7. SUMMARY AND CONCLUSION

Introduction

1.1. Objective of the Analysis:

The objective of this analysis is to gain a basic understanding of the operational picture of Home Credit, suggest intuitive ways to monitor performance through visual charts, and extract insights. These insights will then serve as the foundation for applying a SWOT analysis to develop an overall strategy going forward or to inform future planning and development.

Introduction

1.2. Data Used:

The data is sourced from Kaggle:

https://www.kaggle.com/competitions/home-credit-default-risk/data

File used: application_train.csv

Introduction

1.3. Overview of the Data Used

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

HOME CREDIT HOME CREDIT

Introduction

1.3. Overview of the Data Used

With many detailed columns, there are numerous ways to analyze the data. However, we will first categorize the column groups to make tracking and analysis easier. We will also check the missing value ratio for each column. The data is grouped and categorized as shown in the slide.

This analysis will not cover all the data but will be limited to selected key columns with the least amount of missing data for analysis.

TARGET

Loan Info (3 features)

- NAME_CONTRACT_TYPE
- WEEKDAY_APPR_PROCESS_START
- HOUR_APPR_PROCESS_START

Property Info (3 features)

- FLAG_OWN_REALTY
- FLAG_OWN_CAR
  - ↳ OWN_CAR_AGE → 66.0%
- OWN_CAR_AGE → 66.0%

Application Context (3 features)

- NAME_TYPE_SUITE → 0.4%
- DAYS_ID_PUBLISH
- DAYS_REGISTRATION

Social Circle (4 features)

- OBS_30_CNT_SOCIAL_CIRCLE → 0.3%
- DEF_30_CNT_SOCIAL_CIRCLE → 0.3%
- OBS_60_CNT_SOCIAL_CIRCLE → 0.3%
- DEF_60_CNT_SOCIAL_CIRCLE → 0.3%

Credit Bureau (6 features)

- AMT_REQ_CREDIT_BUREAU_HOUR → 13.5%
- AMT_REQ_CREDIT_BUREAU_DAY → 13.5%
- AMT_REQ_CREDIT_BUREAU_WEEK → 13.5%
- AMT_REQ_CREDIT_BUREAU_MON → 13.5%
- AMT_REQ_CREDIT_BUREAU_QRT → 13.5%
- AMT_REQ_CREDIT_BUREAU_YEAR → 13.5%

Financial Info (7 features)

- AMT_INCOME_TOTAL
- AMT_CREDIT
- AMT_ANNUITY → 0.0%
- AMT_GOODS_PRICE → 0.1%
- EXT_SOURCE_1 → 56.4%
- EXT_SOURCE_2 → 0.2%
- EXT_SOURCE_3 → 19.8%

Building/Area Info (50 features)

- APARTMENTS_AVG → 50.7%
- BASEMENTAREA_AVG → 58.5%
- YEARS_BEGINEXPLUATATION_AVG → 48.8%
- YEARS_BUILD_AVG → 66.5%
- COMMONAREA_AVG → 69.9%
- ELEVATORS_AVG → 53.3%
- ENTRANCES_AVG → 50.3%
- FLOORSMAX_AVG → 49.8%
- FLOORSMIN_AVG → 67.8%
- LANDAREA_AVG → 59.4%
- LIVINGAPARTMENTS_AVG → 68.4%
- LIVINGAREA_AVG → 50.2%
- NONLIVINGAPARTMENTS_AVG → 69.4%
- NONLIVINGAREA_AVG → 55.2%
- APARTMENTS_MODE → 50.7%
- BASEMENTAREA_MODE → 58.5%
- YEARS_BEGINEXPLUATATION_MODE → 48.8%
- YEARS_BUILD_MODE → 66.5%
- COMMONAREA_MODE → 69.9%
- ELEVATORS_MODE → 53.3%
- ENTRANCES_MODE → 50.3%
- FLOORSMAX_MODE → 49.8%
- FLOORSMIN_MODE → 67.8%
- LANDAREA_MODE → 59.4%
- LIVINGAPARTMENTS_MODE → 68.4%
- LIVINGAREA_MODE → 50.2%
- NONLIVINGAPARTMENTS_MODE → 69.4%
- NONLIVINGAREA_MODE → 55.2%
- APARTMENTS_MEDI → 50.7%
- BASEMENTAREA_MEDI → 58.5%
- YEARS_BEGINEXPLUATATION_MEDI → 48.8%
- YEARS_BUILD_MEDI → 66.5%
- COMMONAREA_MEDI → 69.9%
- ELEVATORS_MEDI → 53.3%
- ENTRANCES_MEDI → 50.3%
- FLOORSMAX_MEDI → 49.8%
- FLOORSMIN_MEDI → 67.8%
- LANDAREA_MEDI → 59.4%
- LIVINGAPARTMENTS_MEDI → 68.4%
- LIVINGAREA_MEDI → 50.2%
- NONLIVINGAPARTMENTS_MEDI → 69.4%
- NONLIVINGAREA_MEDI → 55.2%
- FONDKAPREMONT_MODE → 68.4%
- HOUSETYPE_MODE → 50.2%
- TOTALAREA_MODE → 48.3%
- WALLSMATERIAL_MODE → 50.8%
- EMERGENCYSTATE_MODE → 47.4%
- HOUSETYPE_MODE → 50.2%
- WALLSMATERIAL_MODE → 50.8%
- TOTALAREA_MODE → 48.3%

Mobile & Email Info (7 features)

- FLAG_CONT_MOBILE
- FLAG_MOBIL
- FLAG_EMP_PHONE
- FLAG_WORK_PHONE
- FLAG_PHONE
- FLAG_EMAIL
- DAYS_LAST_PHONE_CHANGE → 0.0%

Region Info (9 features)

- REGION_POPULATION_RELATIVE
- REGION_RATING_CLIENT
- REGION_RATING_CLIENT_W_CITY
- REG_REGION_NOT_LIVE_REGION
- REG_REGION_NOT_WORK_REGION
- LIVE_REGION_NOT_WORK_REGION
- REG_CITY_NOT_LIVE_CITY
- REG_CITY_NOT_WORK_CITY
- LIVE_CITY_NOT_WORK_CITY

Personal Info (11 features)

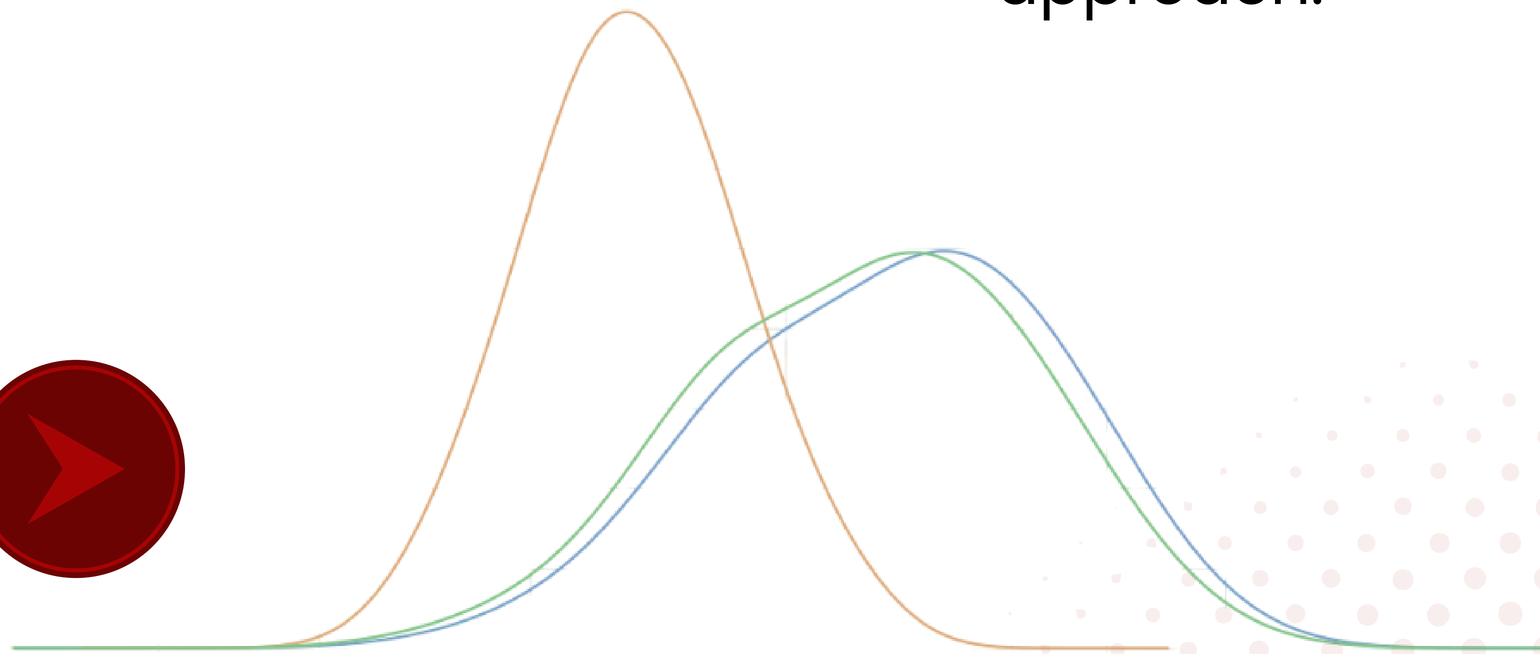
- CODE_GENDER
- DAYS_BIRTH
- NAME_EDUCATION_TYPE
- NAME_FAMILY_STATUS
- NAME_HOUSING_TYPE
- OCCUPATION_TYPE → 31.3%
- CNT_CHILDREN
- CNT_FAM_MEMBERS → 0.0%
- NAME_INCOME_TYPE
- DAYS_EMPLOYED
- ORGANIZATION_TYPE

Documents (20 features)

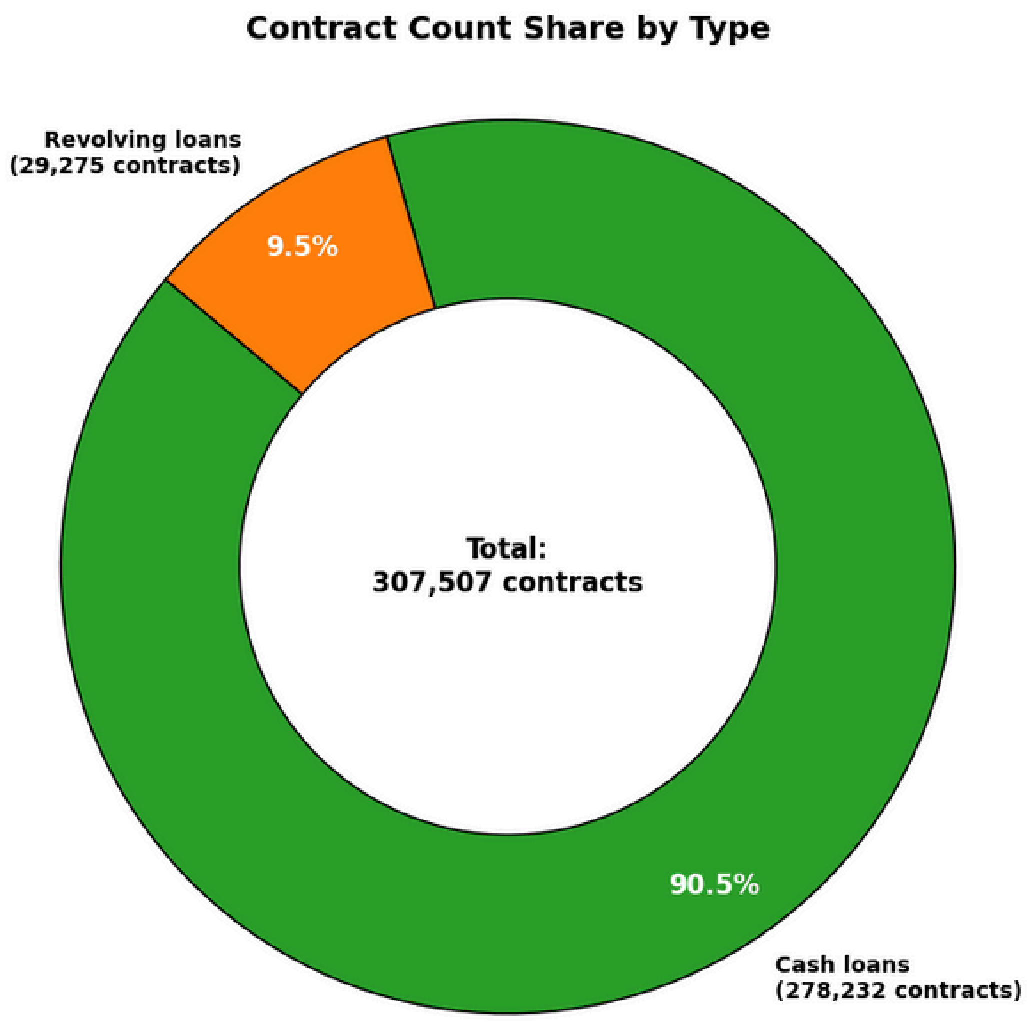
- FLAG_DOCUMENT_2
- FLAG_DOCUMENT_3
- FLAG_DOCUMENT_4
- FLAG_DOCUMENT_5
- FLAG_DOCUMENT_6
- FLAG_DOCUMENT_7
- FLAG_DOCUMENT_8
- FLAG_DOCUMENT_9
- FLAG_DOCUMENT_10
- FLAG_DOCUMENT_11
- FLAG_DOCUMENT_12
- FLAG_DOCUMENT_13
- FLAG_DOCUMENT_14
- FLAG_DOCUMENT_15
- FLAG_DOCUMENT_16
- FLAG_DOCUMENT_17
- FLAG_DOCUMENT_18
- FLAG_DOCUMENT_19
- FLAG_DOCUMENT_20

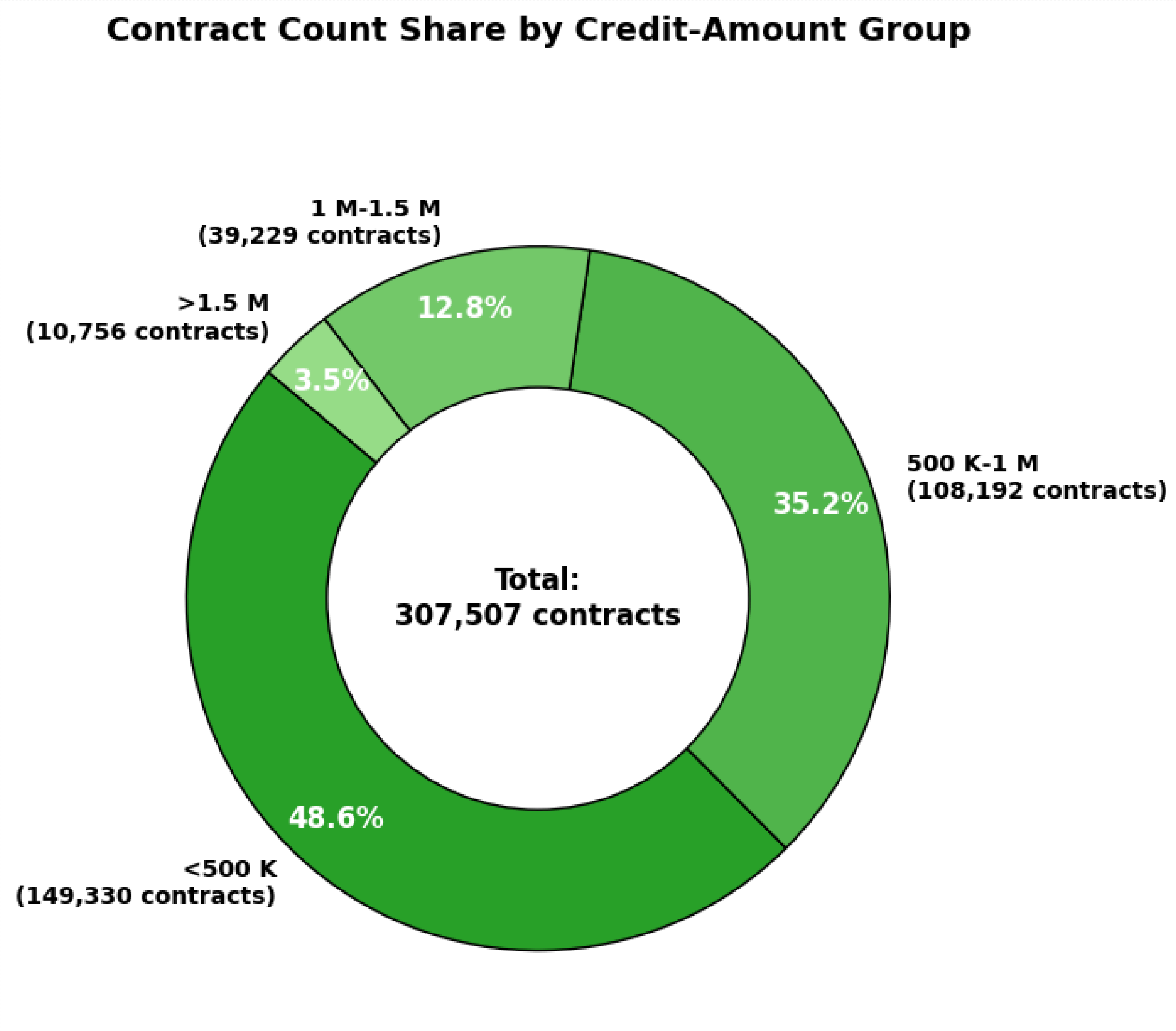
Loan Contract Type Analysis

Loan contract type is a particularly interesting detail. Through this analysis, we can uncover a pattern in user behavior based on distribution. It also offers insights into the current operational approach.



Loan Contract Type Analysis





LOAN CONTRACT STRUCTURE ANALYSIS

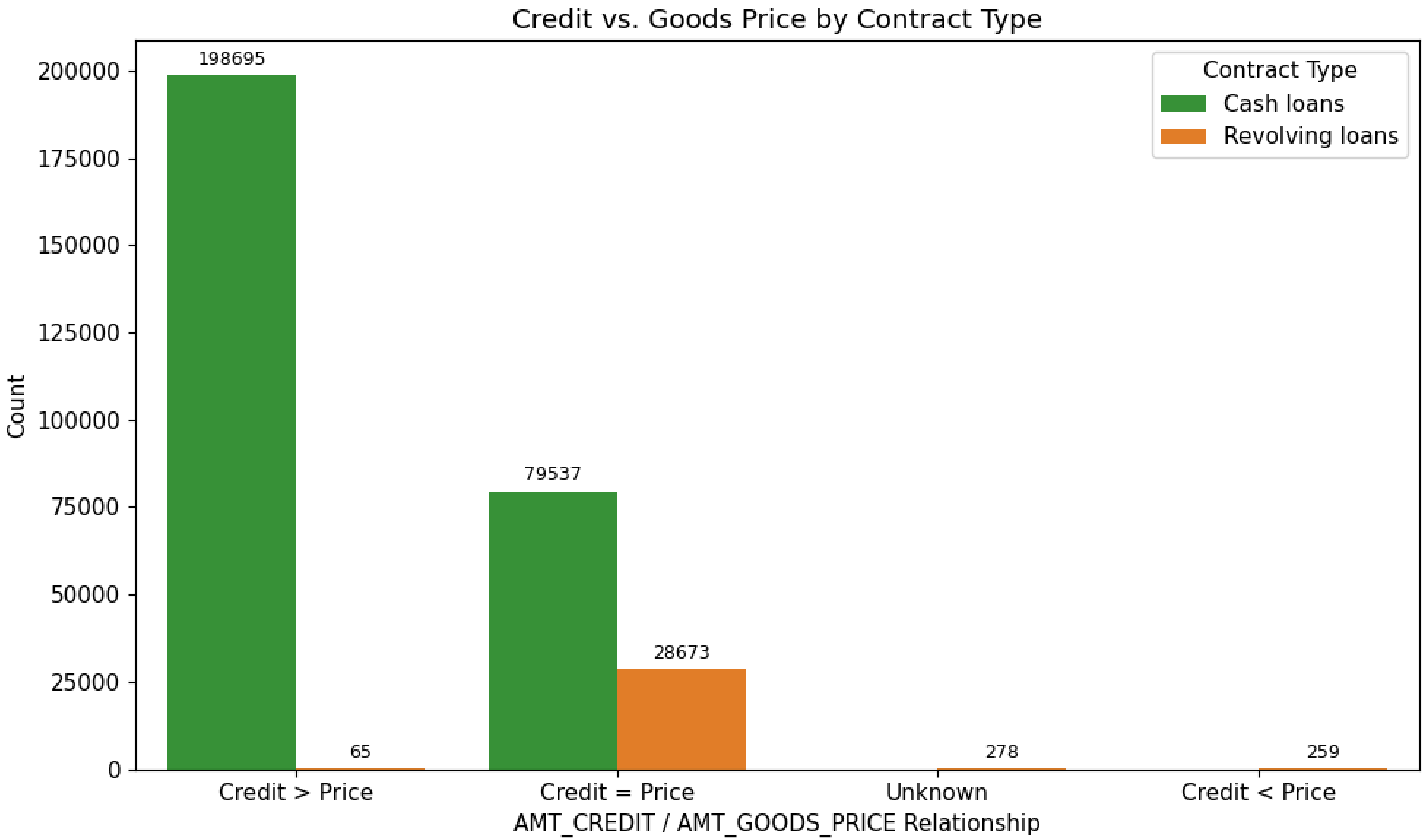
Observations:

- 90.5% of the contracts are Cash loans (standard consumer loans).
- Only 9.5% are Revolving loans (credit lines – typically credit cards or flexible limits).
- Nearly 84% of the loans are under 1 million, broken down as follows:
  - Less than 500K: 48.6%
  - 500K – 1M: 35.2%
  - 1M – 1.5M: 12.8%
- Only 3.5% of loans exceed 1.5 million, indicating a model focused mainly on small consumer loans.
- (Note: 1K = 1,000; 1M = 1,000,000; 1B = 1,000,000,000)

Insight:

- The heavy skew toward Cash loans could be due to:
  - Company policy prioritizing standard consumer lending.
  - Revolving loans accounting for a small share may reflect limited familiarity or adoption among the customer base.
- Although Revolving loans make up a small percentage, they may hold significant future potential and warrant further evaluation from a financial efficiency standpoint.
- The small loan sizes reflect the characteristics of the personal consumer market, aligning with the needs of middle- to lower-income groups.
- Restrictions on large loans may stem from strict underwriting processes or loan caps designed to minimize credit risk.

Loan Contract Type Analysis



ANALYSIS OF LOAN COUNTS BY LOAN AMOUNT VS. PRODUCT PRICE

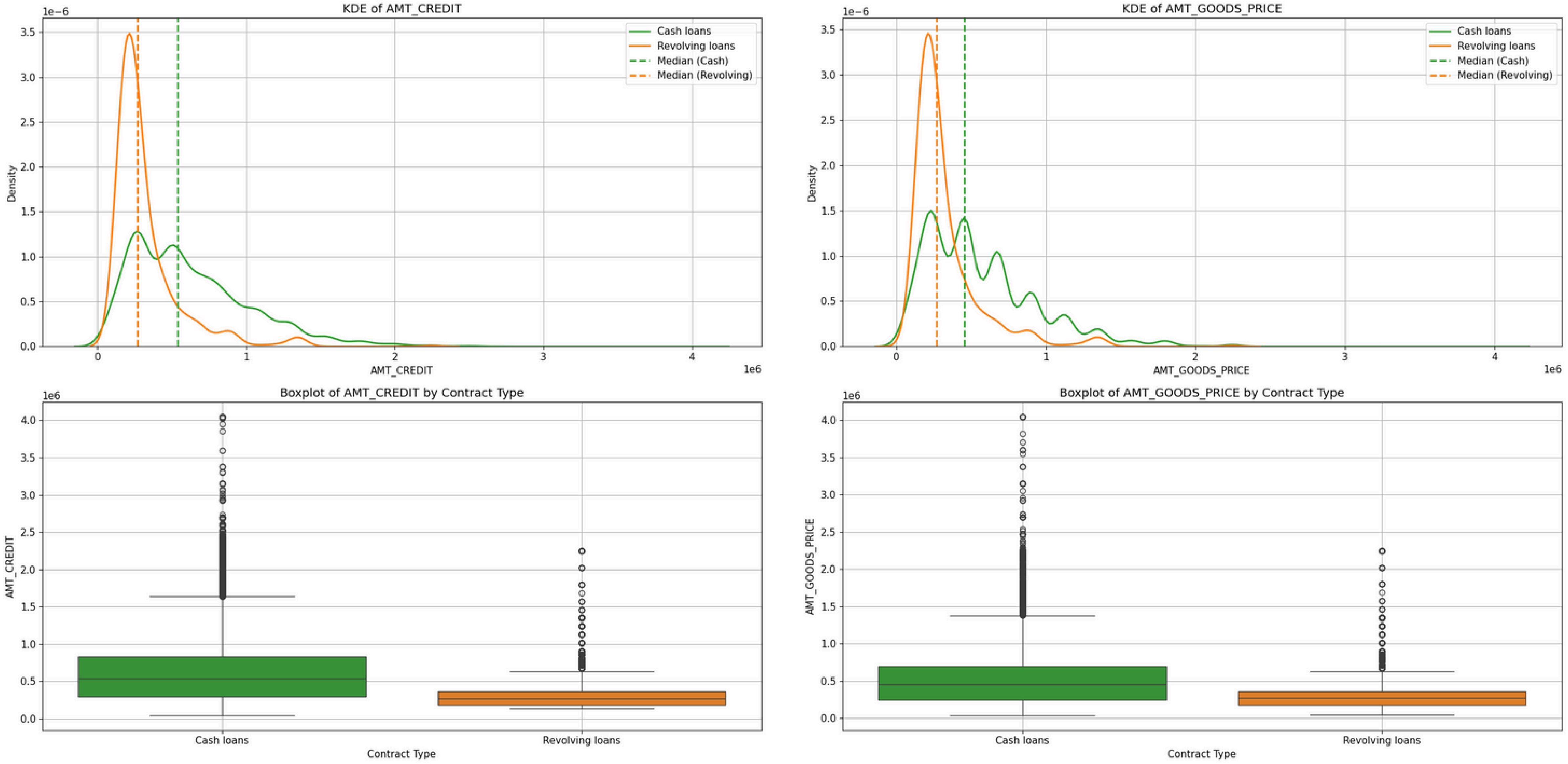
Observations:

- 198,695 Cash loan contracts have a loan amount greater than the item price (Credit > Price) — making up the majority.
- Cash loans typically have structures where Credit = Price or Credit > Price, whereas Revolving loans are almost exclusively in the Credit = Price category.
- A small number of records fall into the "Unknown" or Credit < Price groups (278 Cash, 259 Revolving). These should be reviewed further to assess potential risks.

Insight:

- Loans exceeding the value of the product are common. This suggests that borrowers may be using the extra funds for other spending or to cover additional services tied to the purchase.
- The near absence of Credit > Price cases in Revolving loans indicates stricter control over credit limits for this product type.
- The high rate of Credit = Price in Revolving loans aligns with the nature of small, straightforward transactions.
- There are no pure cash loan transactions in this dataset.
- Cash loans with Credit < Price are virtually nonexistent, which makes sense — no one typically borrows less than the product value unless a down payment is involved. However, the data doesn't clearly reflect this scenario.

Loan Contract Type Analysis



ANALYSIS OF CONTRACTS BY AMT_CREDIT AND AMT_GOODS_PRICE

📊 Chart 1: AMT_CREDIT by Contract Type

Observations:

- The median loan amount for Cash loans is clearly higher than for Revolving loans — indicating that cash loans typically involve larger sums.
- Both types show a strong right-skewed distribution, especially Cash loans.
- The boxplot reveals many outliers in Cash loans, with some exceeding 2 million and even approaching 4 million.

Insight:

- Cash loans are designed for larger financing needs, suitable for higher-consumption behavior or emergency expenses.
- Revolving loans stay small and stable, aligning with typical credit card or installment purchase behavior.
- The high number of outliers in Cash loans warrants risk monitoring: Are these exceptional customers, or signs of lending beyond safe thresholds?

📊 Chart 2: AMT_GOODS_PRICE by Contract Type

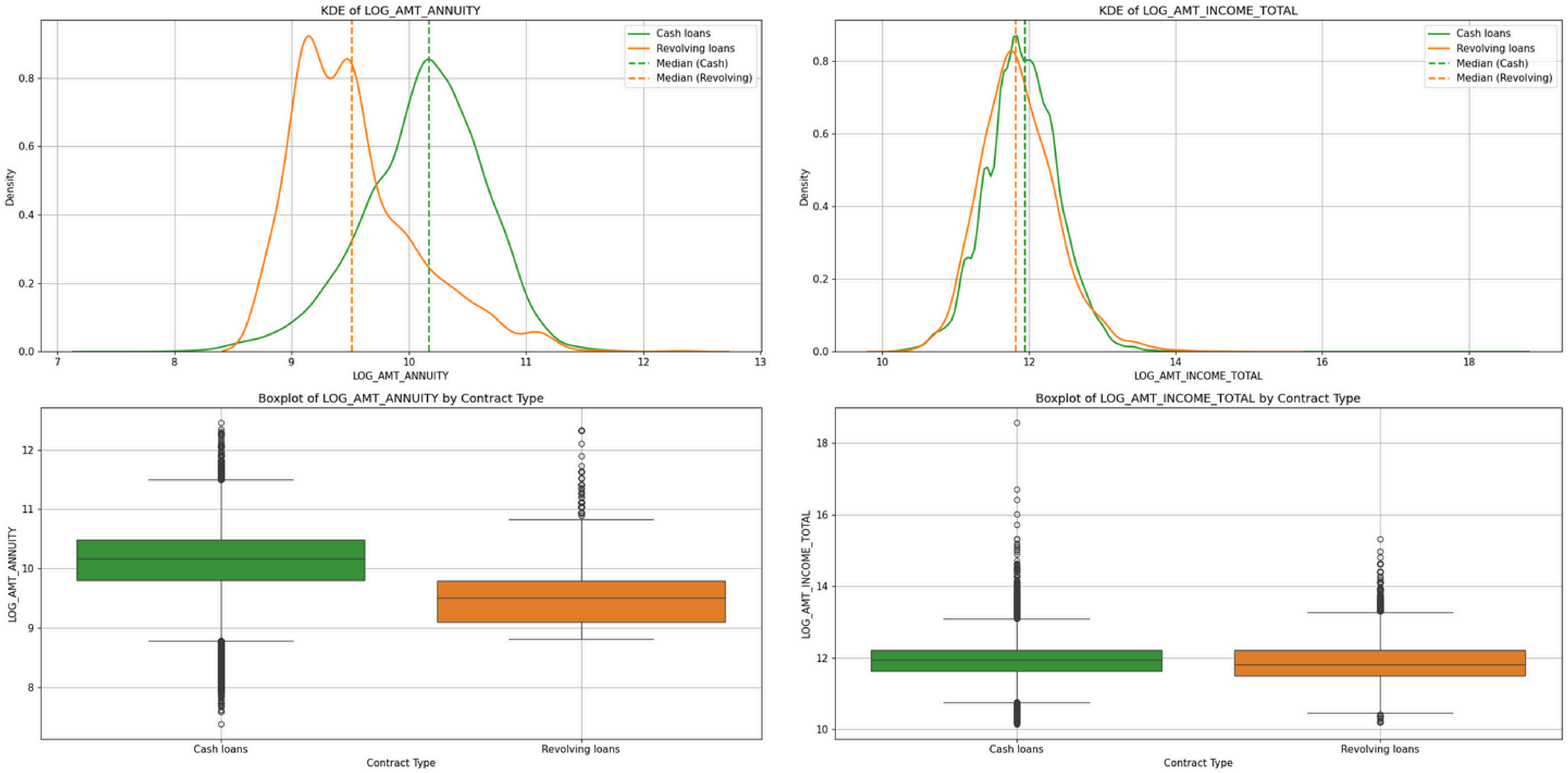
Observations:

- The price of goods closely correlates with AMT_CREDIT, and the trends are similar.
- Median and distribution patterns between the two contract types remain consistent with those seen in the AMT_CREDIT chart.

Insight:

- Loans tied to physical goods remain a central feature.
- Revolving loans involve low-value, low-variance goods — suitable for POS loans (point-of-sale financing).
- Outliers in Cash loans likely reflect high-ticket items like motorcycles, furniture, or premium electronics.

Loan Contract Type Analysis



CONTRACT ANALYSIS BY AMT_ANNUITY AND AMT_INCOME_TOTAL

📊 Chart: LOG(AMT_ANNUITY)

Observations:

- The median annuity for Cash loans is significantly higher than that of Revolving loans.
- Revolving loans show a narrow, sharply peaked distribution — suggesting most repayments are small and consistent.
- Cash loans have a wider spread with many outliers, reflecting their larger size and longer terms.

Insight:

- Revolving loans likely represent small, quick-pay transactions (e.g., POS purchases).
- Cash loans meet larger financing needs with longer terms and higher monthly payments.
- The boxplot confirms that Revolving loans are very "compact" in terms of installment amounts.

📊 Chart: LOG(AMT_INCOME_TOTAL)

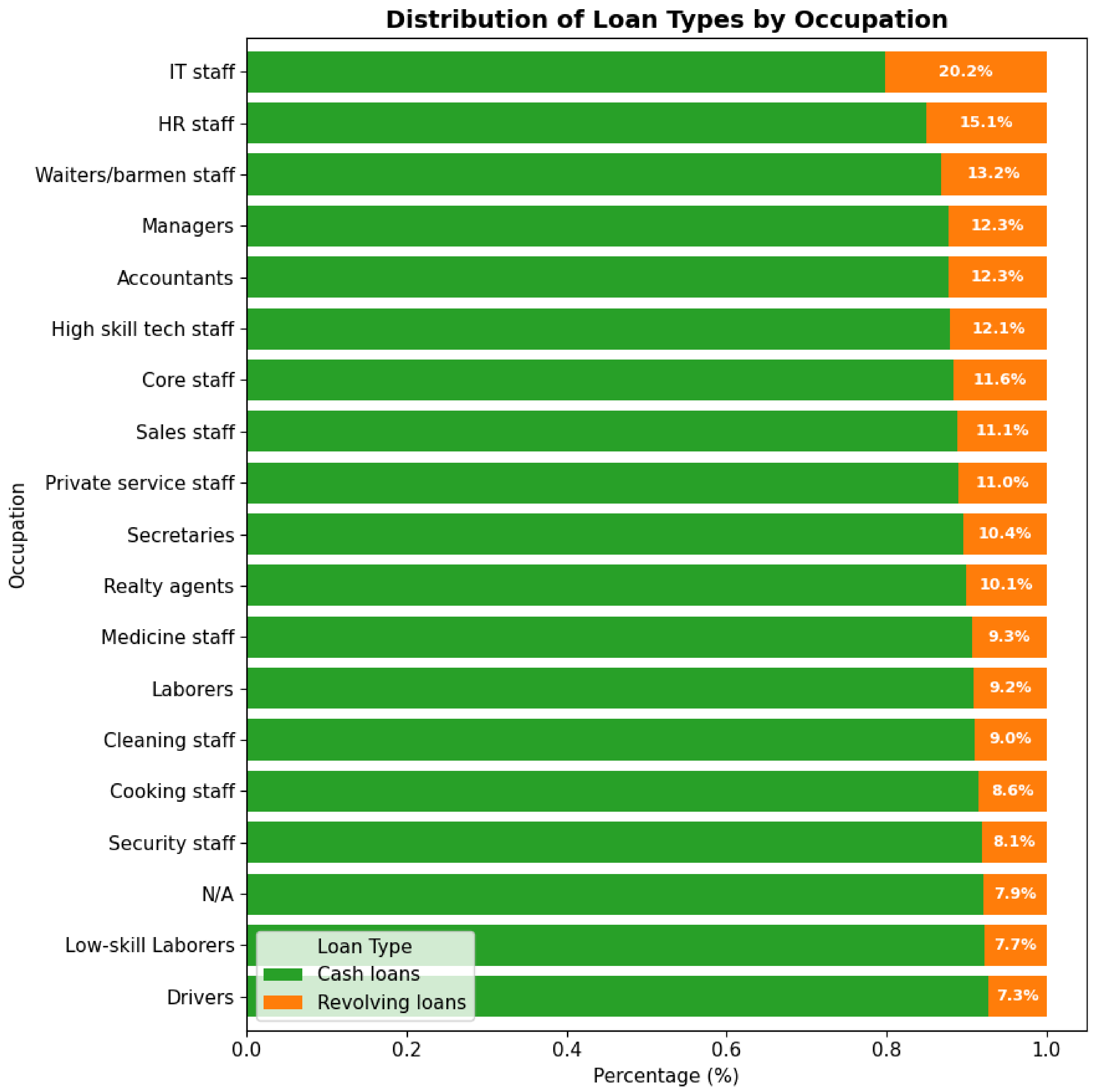
Observations:

- Income distributions for the two groups are quite similar in both median and shape.
- Both contain a small number of high-income outliers (> log 14, or > 1.6 million).

Insight:

- Customer income doesn't appear to be the main factor determining loan type; factors like financial needs and credit history may play a larger role.
- Revolving loans seem more common among customers with stable, average incomes, while Cash loans are spread across all income levels.
- Some high-income customers still choose Cash loans — possibly due to personal habits. This group may represent a premium segment worth targeting with tailored services.

Loan Contract Type Analysis



LOAN CONTRACT ANALYSIS BY OCCUPATION

📊 Chart: Distribution of Loan Types by Occupation

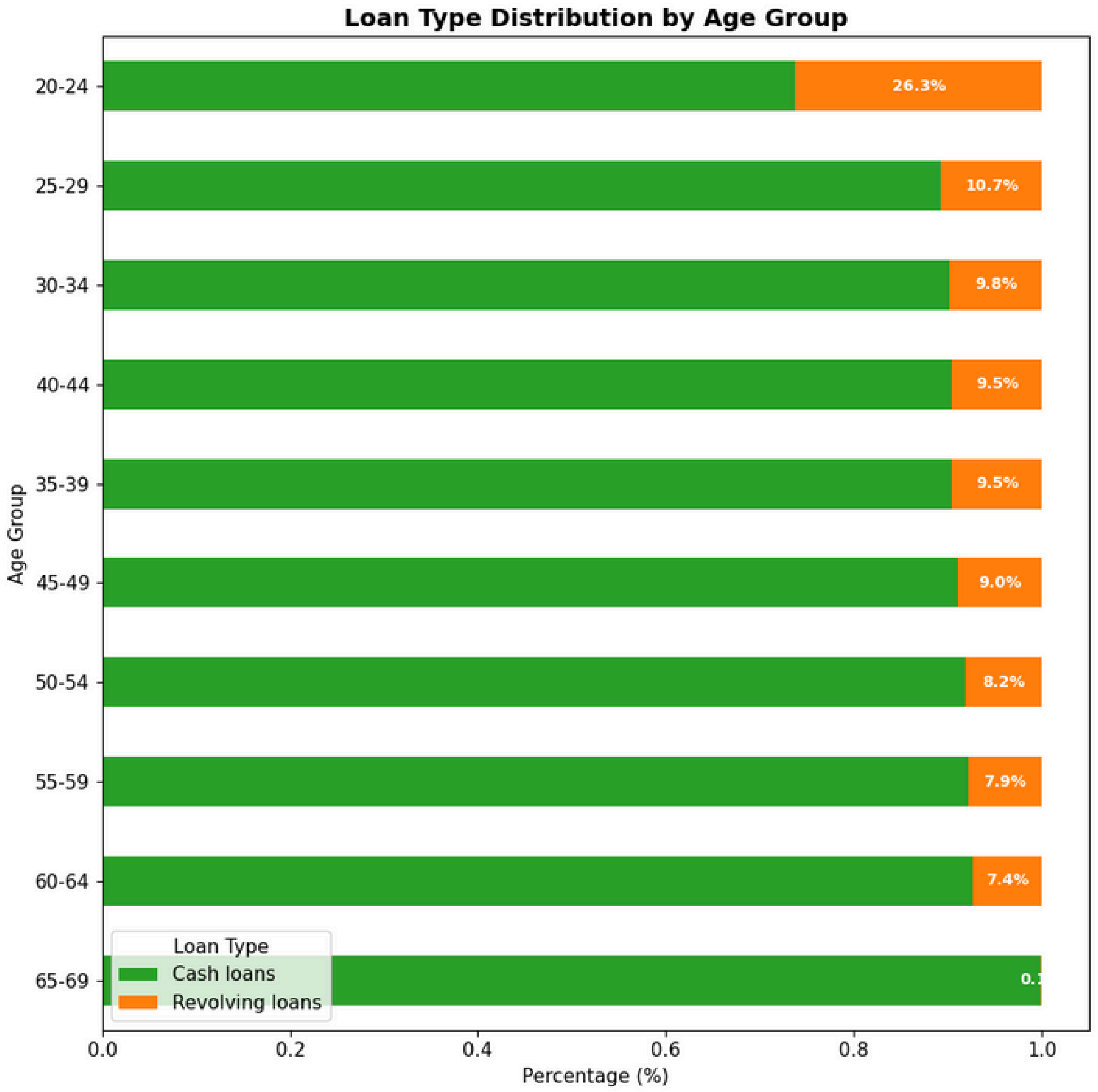
Observations:

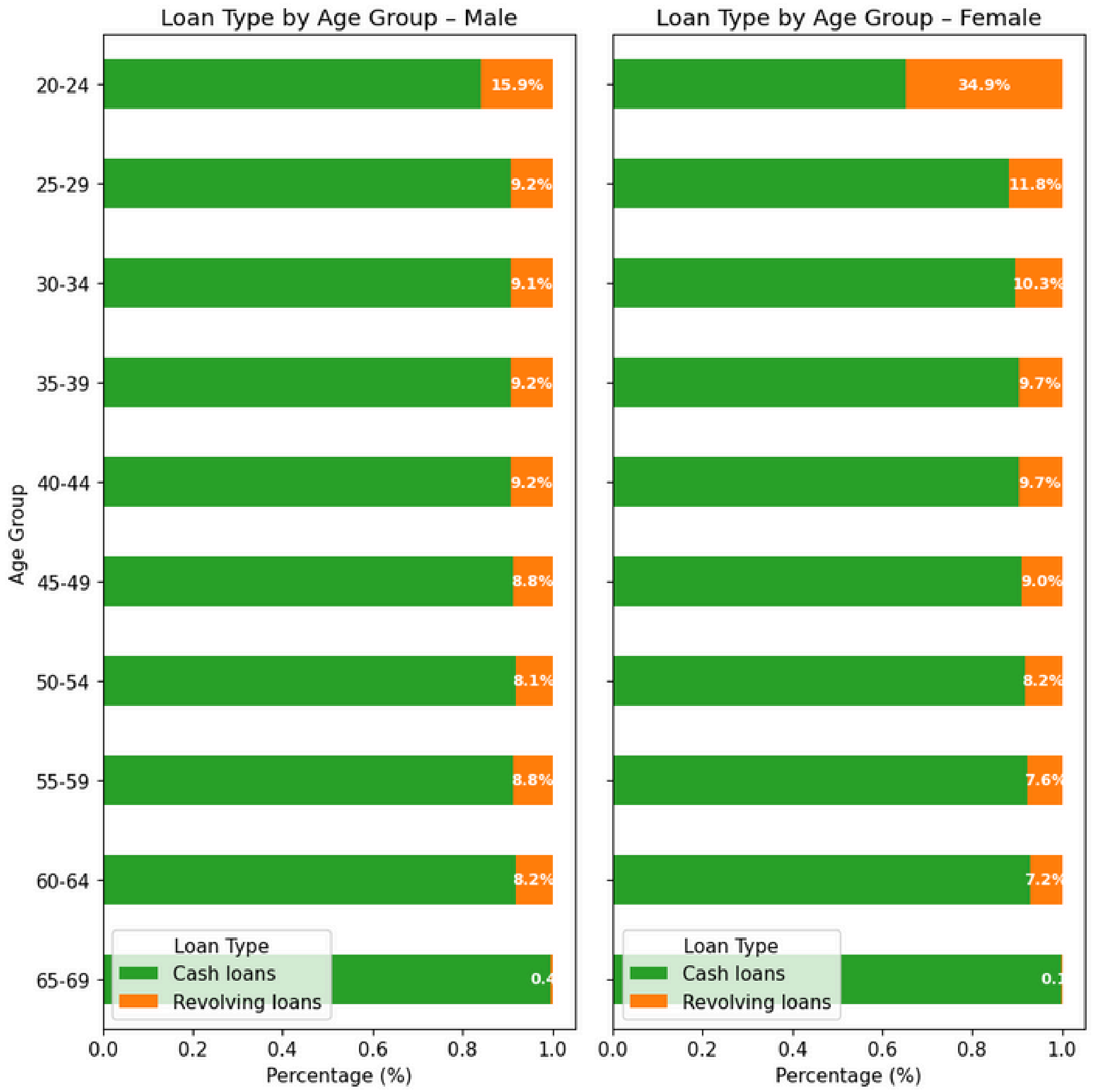
- The highest proportions of Revolving loans (orange) are seen in:
  - IT staff: 20.2%
  - HR staff: 15.1%
  - Waiters/Barmen staff: 13.2%
- Most other occupations have Revolving loan rates around or below 12%.
- Cash loans dominate across all other groups, especially:
  - Drivers: 92.7%
  - Low-skill Laborers: 92.3%
  - Security/Cleaning/Cooking staff: all above ~90%

Insight:

- White-collar professionals tend to use Revolving loans more frequently.
  - Roles like IT, HR, Accountants, Managers, and other high-skill staff have the highest Revolving loan rates.
  - These customers are more tech-savvy, generally have higher incomes, and prefer flexible financial tools like POS loans. They represent a modern customer segment that values speed and convenience.
- Blue-collar workers heavily favor Cash loans.
  - Occupations like Drivers, Low-skill Laborers, and Security staff rarely use Revolving loans, likely due to:
    - Limited access to technology
    - Weaker credit histories
    - Lower chances of being approved for flexible credit limits
- Realty agents and Sales staff show moderate use of Revolving loans (~10%), representing a hybrid segment between labor and professional classes.

Loan Contract Type Analysis





LOAN CONTRACT ANALYSIS BY AGE GROUP AND GENDER WITHIN AGE GROUPS

📊 Chart 1: Loan Type Distribution by Age Group

Observations:

- The 20–24 age group has the highest rate of Revolving loans at 26.3% — significantly higher than all other age groups.
- After age 25, the proportion of Revolving loans steadily declines with age.
- The 65–69 age group almost never uses Revolving loans.

📊 Chart 2: Gender Comparison by Age Group

Observations:

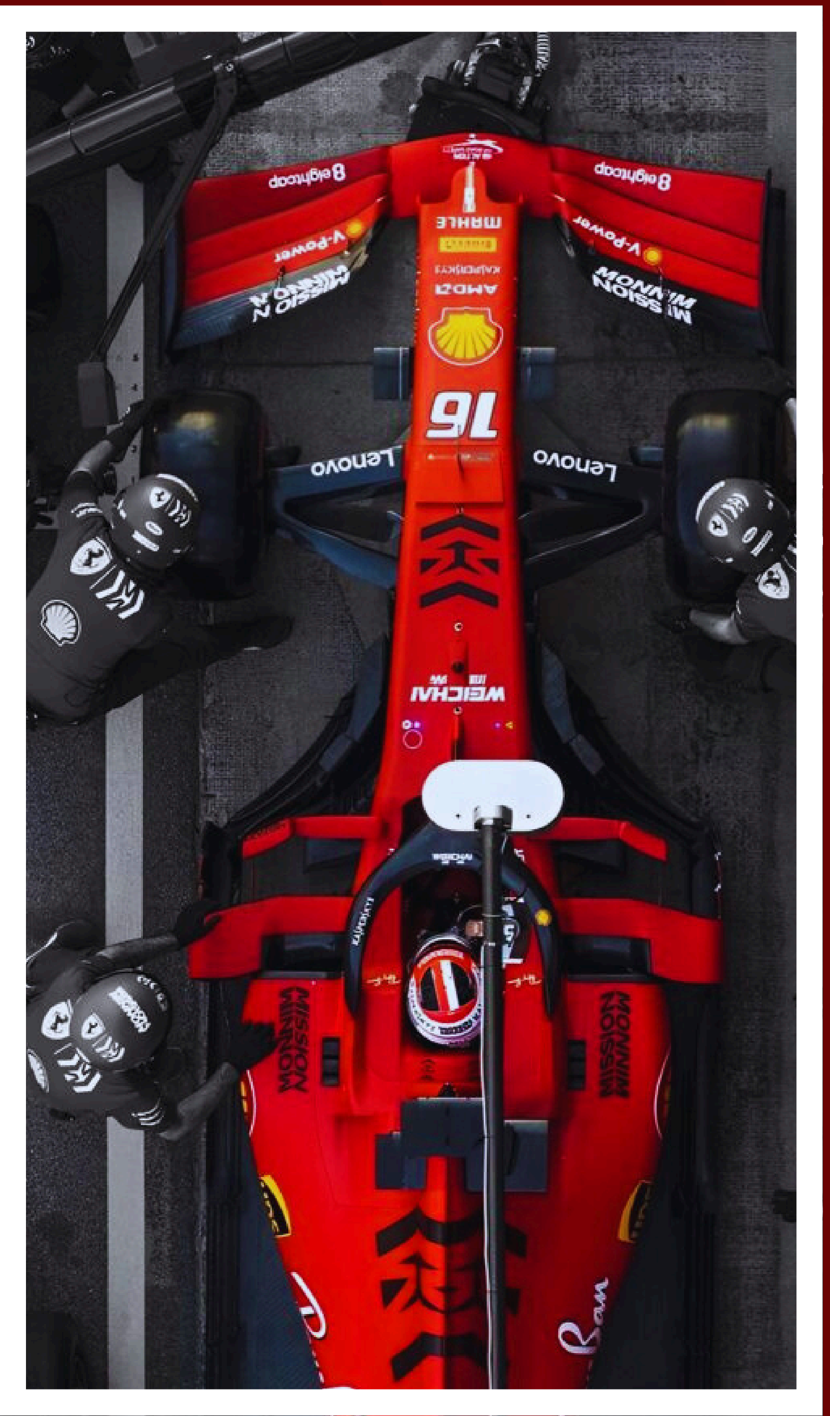
- Among those aged 20–24, women have a Revolving loan rate of 34.9%, while men are at only 15.9% — a stark difference.
- Overall, women tend to use Revolving loans more than men across all age groups, especially in younger brackets (20–29).
- After age 30, the gender gap narrows, but women still lead in Revolving loan usage.

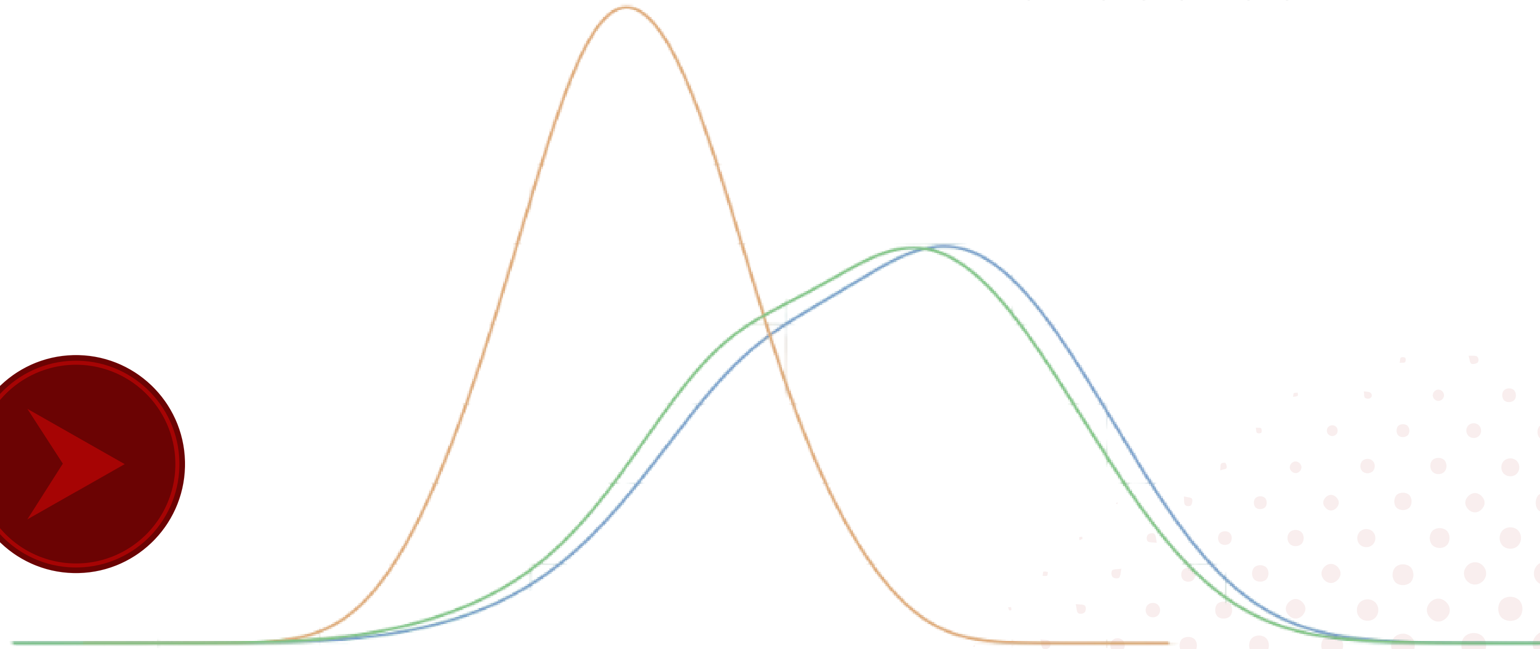
Insight:

- Revolving loans appeal more to younger customers, especially young women. This may be due to:
  - Flexible spending needs
  - Minimal income proof requirements
  - Fast and convenient processing
- Young women may engage more frequently in consumer shopping and POS usage, making this loan type a good fit.
- Cash loans become more favored with age, possibly because older individuals prefer more stable, structured borrowing. They may also be more cautious about flexible credit options.

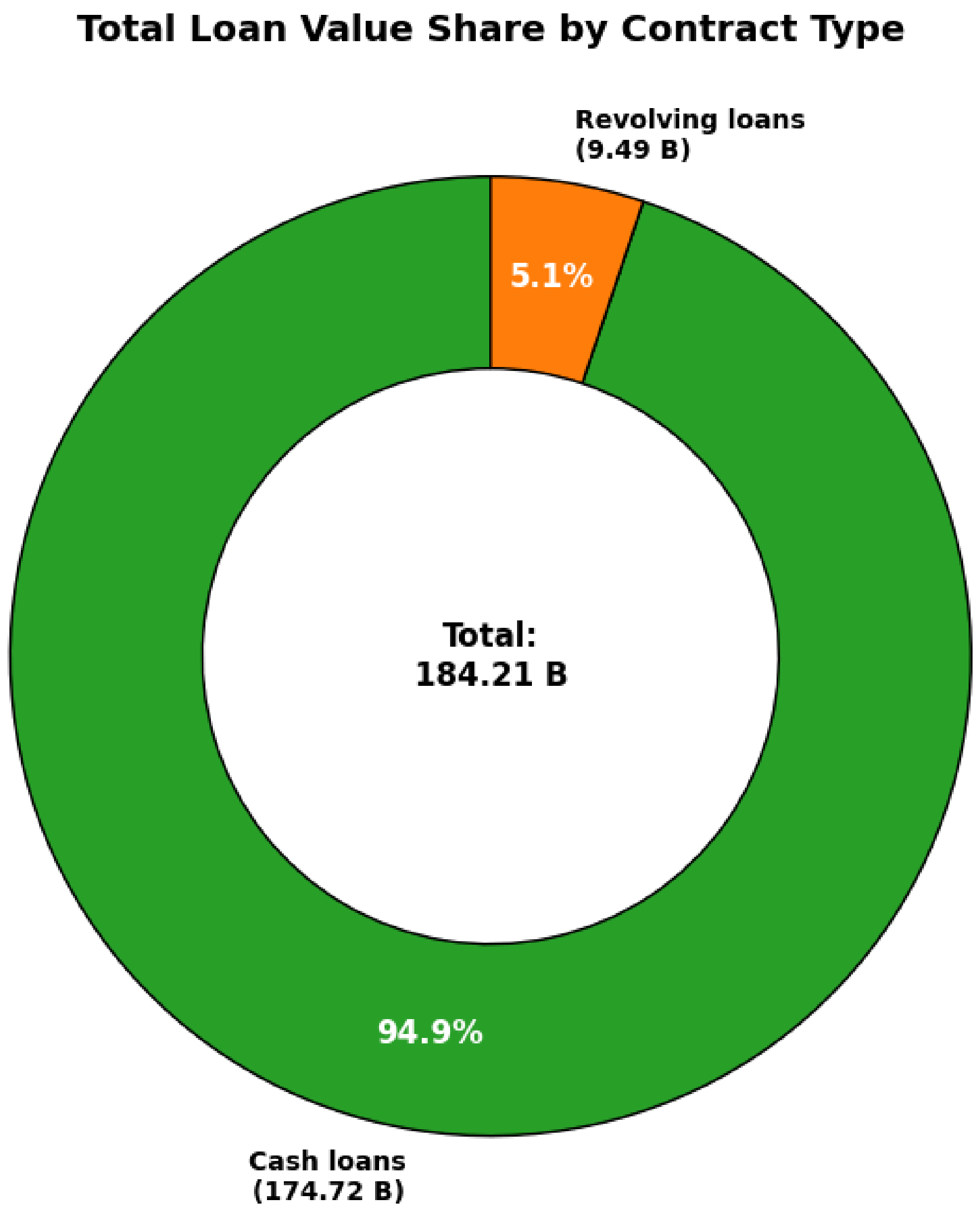
Loan Value Analysis

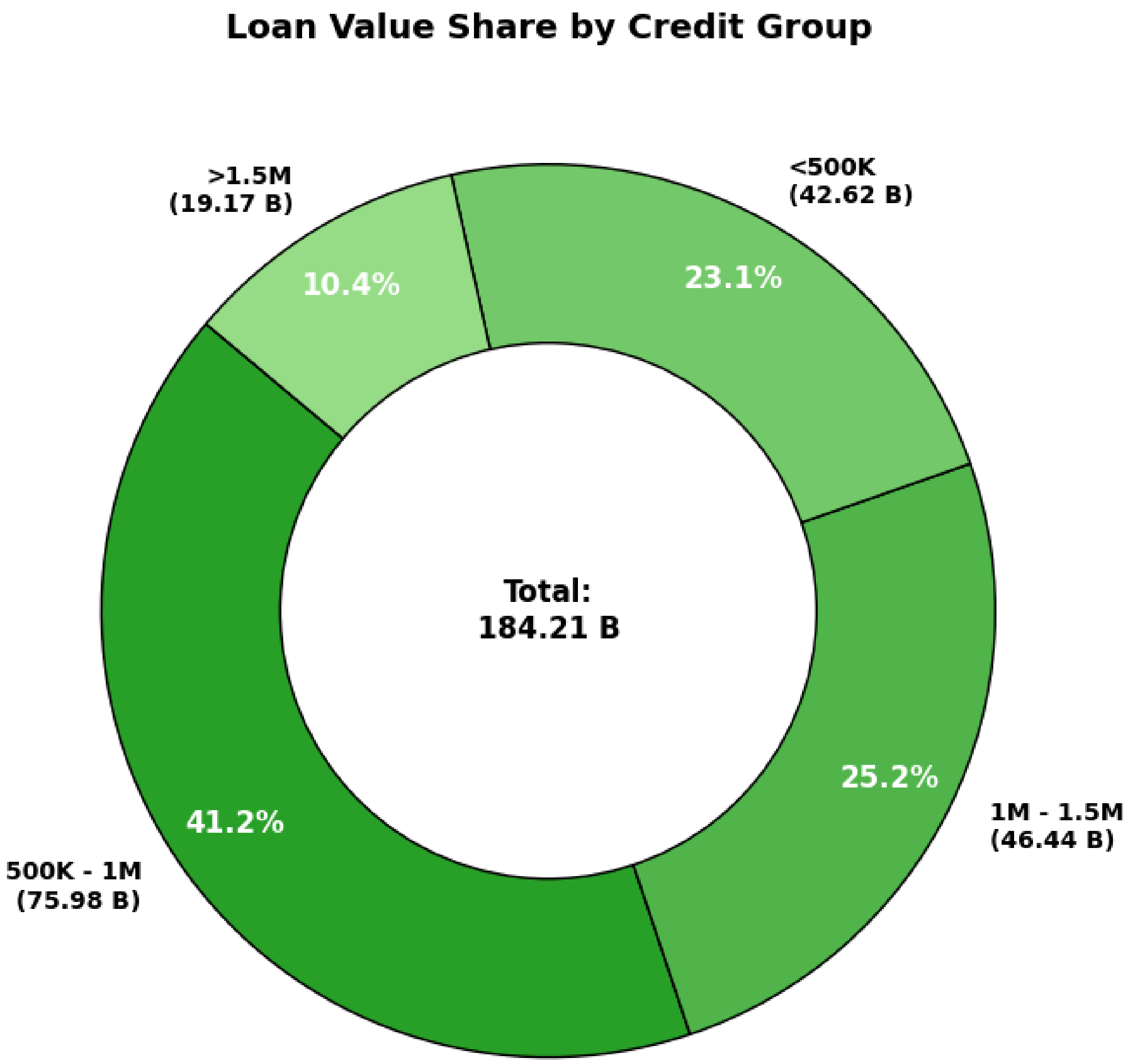
The value of loans provides a wealth of meaningful information. In this analysis, we'll explore the density, distribution, and proportions of current lending flows to understand how capital is being allocated.





Loan Value Analysis





📊 Chart 1: Total Loan Value by Contract Type

Observations:

- Total loan portfolio value: 184.21 billion.
- Cash loans account for 94.9% of the value (174.72 billion).
- Revolving loans make up only 5.1% (9.49 billion).

Insight:

- This aligns with previous insights: Cash loans dominate in both contract volume and value.
- While Revolving loans have a noticeable presence among younger users, IT staff, and women, the loan amounts are small, typically short-term and fast-revolving.
- To assess the financial effectiveness of Revolving loans, a deeper analysis of profit margins, turnover rate, and risk is necessary before considering expansion.
- (Note: 1K = 1,000; 1M = 1,000,000; 1B = 1,000,000,000)

📊 Chart 2: Loan Value by Credit Group

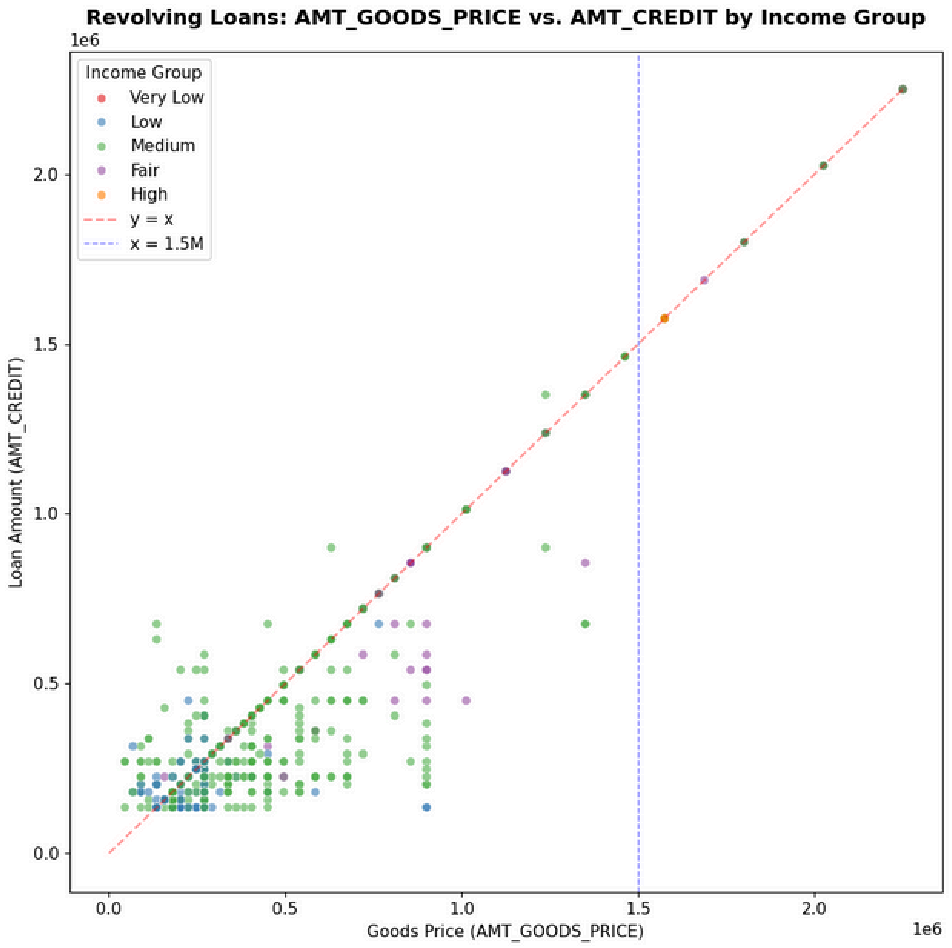
Observations:

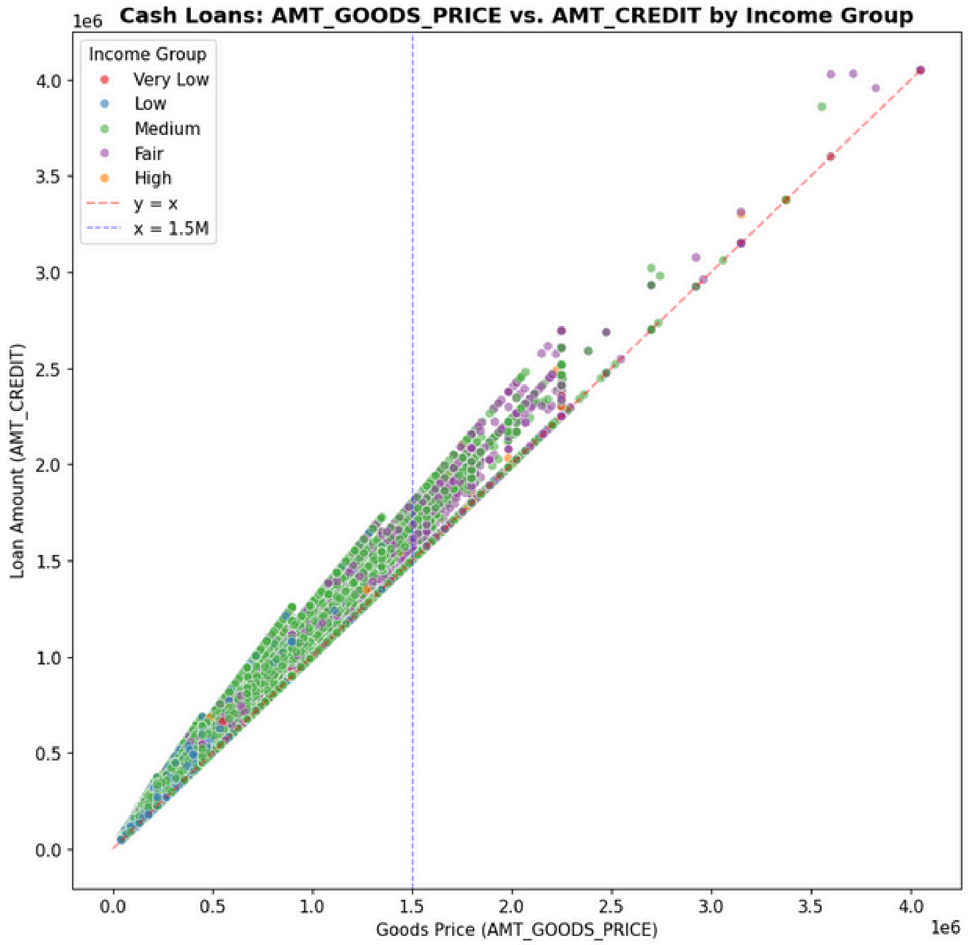
- Loans in the 500K–1M range make up the largest share: 41.2% (75.98 billion).
- Followed by:
  - 1M–1.5M: 25.2% (46.44 billion)
  - <500K: 23.1% (42.62 billion)
  - >1.5M: 10.4% (19.17 billion)

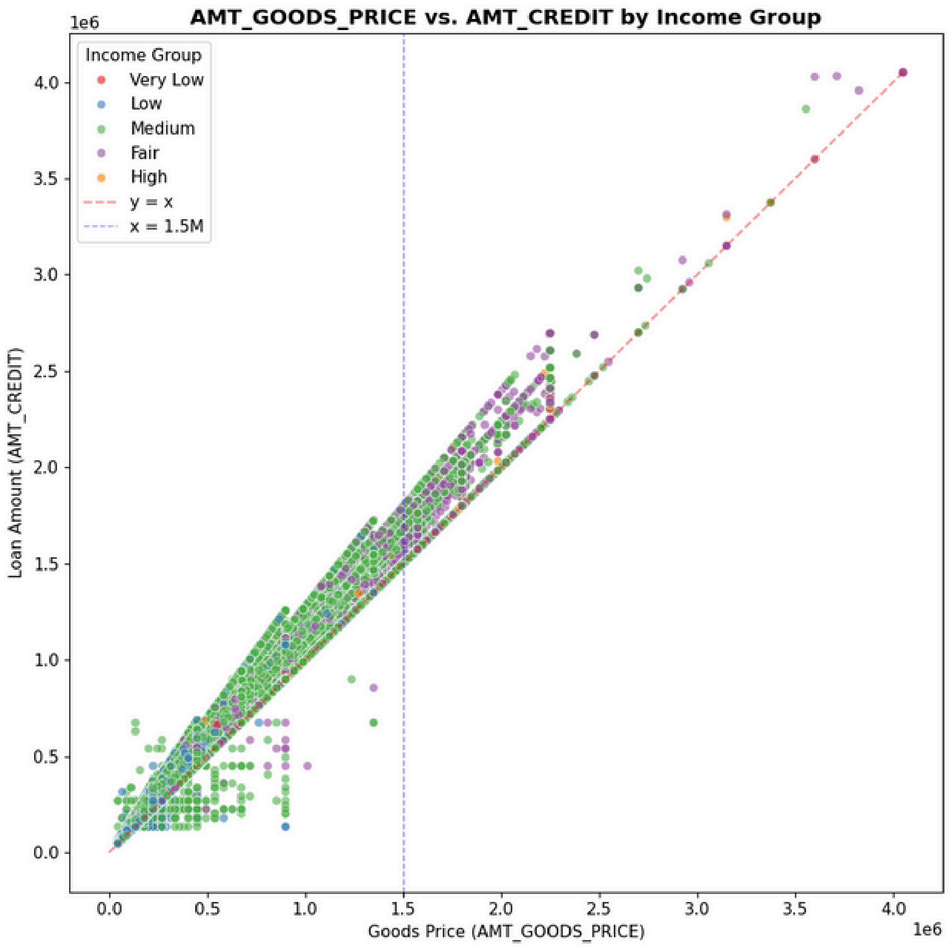
Insight:

- The mid-tier segment (500K–1M) is the backbone of the portfolio, likely consisting of stable, middle-income customers.
- The high-value loan group (>1.5M) represents only 10.4%, indicating a conservative lending strategy — possibly avoiding overexposure or due to limited expansion into this segment.
- The <500K group, while high in volume, contributes less in total value. It's important to evaluate the servicing cost vs. return for this segment to ensure business efficiency.

Loan Value Analysis







📊 Chart 1: Revolving Loans – Correlation Between “AMT_GOODS_PRICE” and “AMT_CREDIT”

Observations:

- Most data points lie close to or just below the y = x line, meaning AMT_CREDIT ≈ AMT_GOODS_PRICE.
- Very few points exceed the 1.5 million threshold (x = 1.5M), consistent with the low credit limit nature of Revolving loans.
- The majority of borrowers fall into the Low and Medium income groups, with very few from the High-income segment.

Insight:

- Revolving loans are typically granted equal to the value of the goods, indicating practical consumer use and tight control.
- High-income customers rarely use this loan type — possibly because they rely on credit cards or have no need for such products. This raises important questions about the product’s positioning, credit limits, and customer needs. Further research is needed into mechanisms, customer behavior, and product relevance in higher income segments.

📊 Chart 2: Cash Loans – Correlation Between “AMT_GOODS_PRICE” and “AMT_CREDIT”

Observations:

- Many data points lie above the y = x line, indicating AMT_CREDIT > AMT_GOODS_PRICE — suggesting that borrowers are receiving more than just the value of the goods, or possibly that additional fees are bundled into the loan. This warrants further investigation.
- Numerous points exceed 1.5M in both goods value and loan amount. Income groups are more diverse, including Medium, Fair, and High-income customers.

Insight:

- Cash loans are more flexible, allowing borrowers to take out more than the cost of the goods, possibly for: Installation fee, insurance, additional consumer spending...
- Higher-income individuals also use Cash loans more frequently, which may indicate a lack of familiarity or trust in Revolving loan products — highlighting a potential education or product positioning gap.

📊 Chart 3: Combined View of All Contract Types

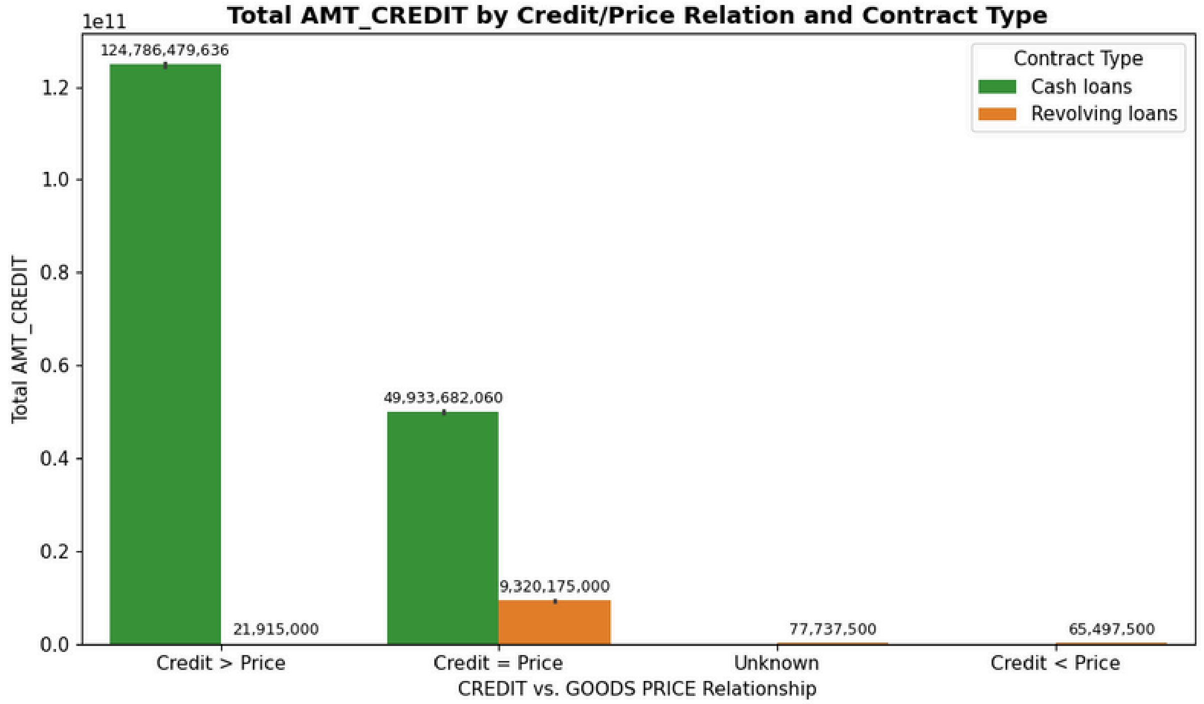
Observations:

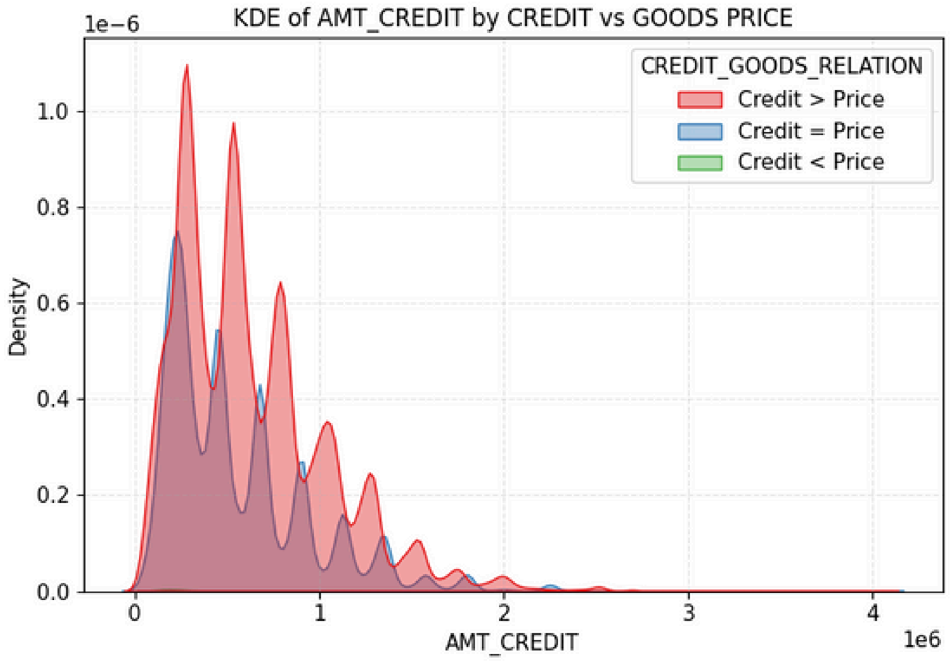
- Overall, the market still shows a strong trend of loan amount ≈ value of goods, but Cash loans create a noticeable cluster above the y = x line, indicating loan amounts greater than the product price.
- There is a clear spread of higher-income groups toward larger loan amounts.

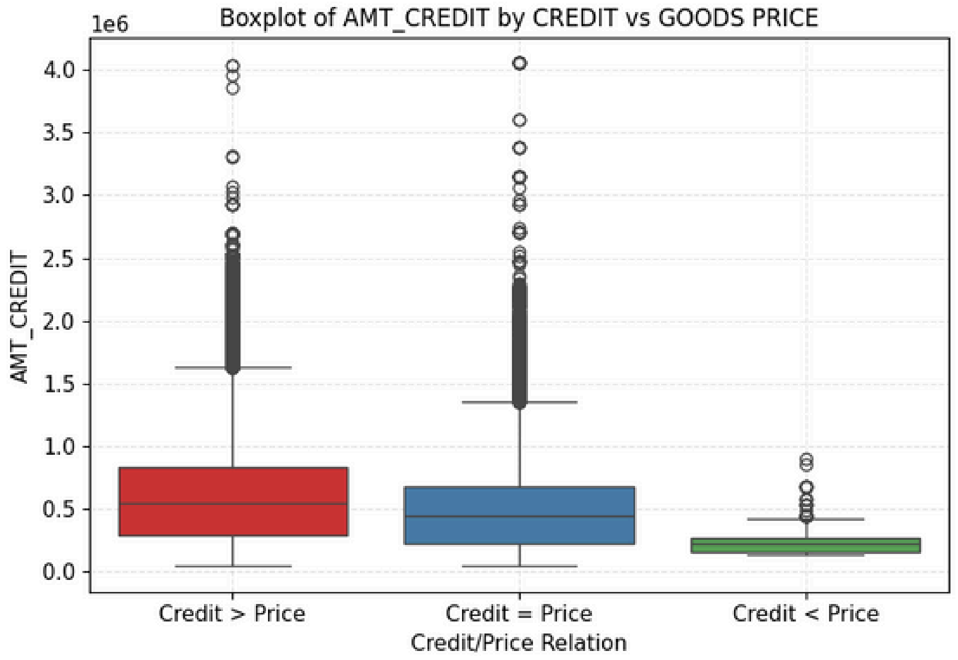
Insight:

- While most loans are tightly tied to product value, Cash loans introduce flexibility, allowing excess borrowing.
- Higher-income customers are actively participating in larger loan segments, suggesting they may be driving higher-value borrowing behavior — a key target group for upselling or premium financial products.

Loan Value Analysis







📊 Bar Chart: Total Loan Value by CREDIT/PRICE Group and Contract Type

Observations:

- Credit > Price: 124.79 billion (Cash Loans) 21.9 million (Revolving Loans)
- Credit = Price: 49.93 billion (Cash Loans) 9.32 billion (Revolving Loans)
- Credit < Price: 65.5 million (Revolving Loans)
- Unknown: 77.7 million (Revolving Loans)

Insight:

- Cash loans overwhelmingly dominate the "Credit > Price" group, accounting for 124.79 billion — this is where the majority of loan capital is concentrated.
- Revolving loans are almost exclusively in the "Credit = Price" group, which aligns with the product's intended structure.
- Loans in the "Credit < Price" and "Unknown" categories have very small values and don't significantly affect totals, but they should be reviewed to clean the data or uncover potential irregularities.

📊 KDE + Boxplot Chart: AMT_CREDIT by CREDIT vs GOODS PRICE

Observations:

KDE:

- The Credit > Price group shows a wider distribution of AMT_CREDIT, with more high-value loans.
- The Credit = Price group is concentrated around the mid-range (~500K–1M).
- The Credit < Price group is very small and low in value, possibly indicating partial upfront payments by buyers.

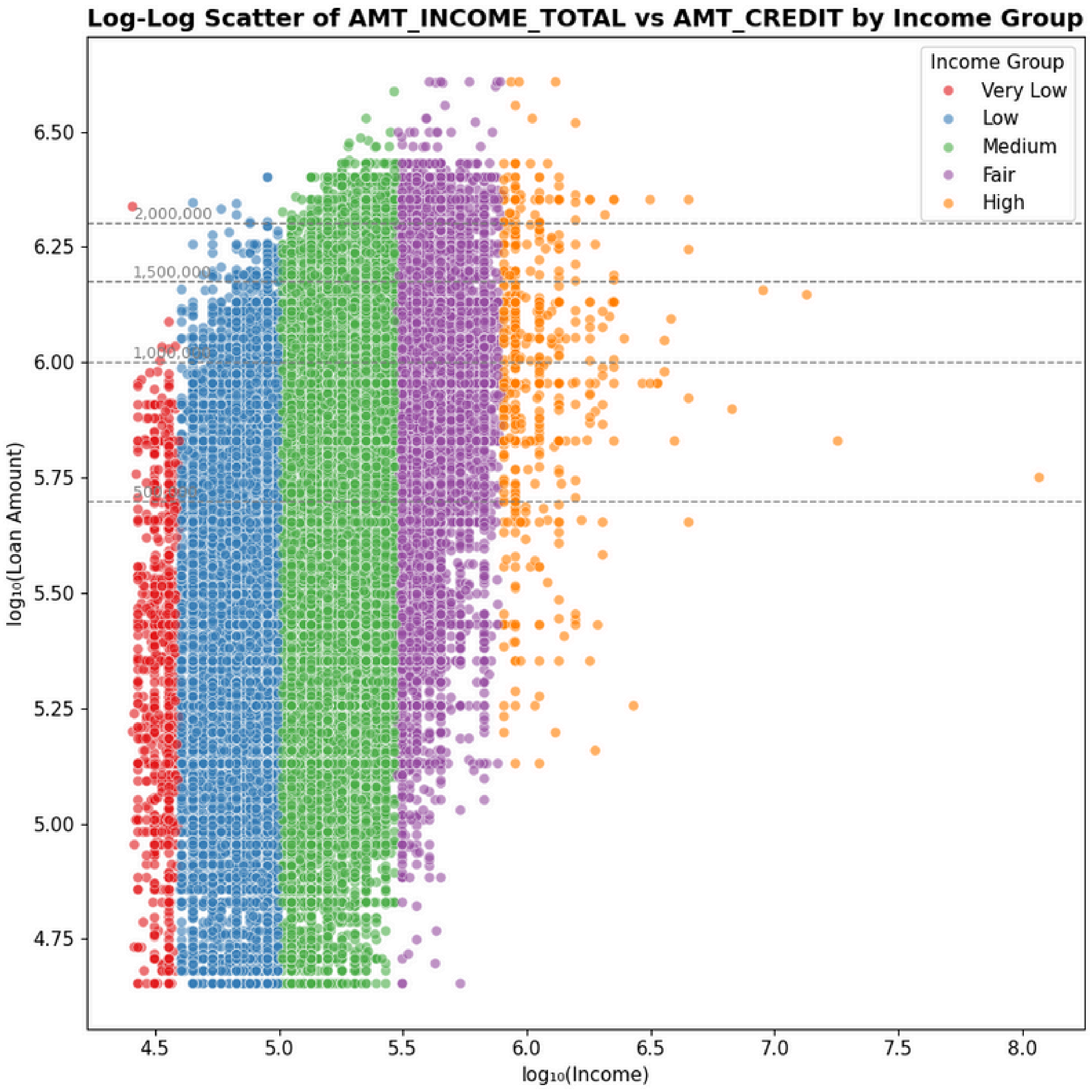
Boxplot:

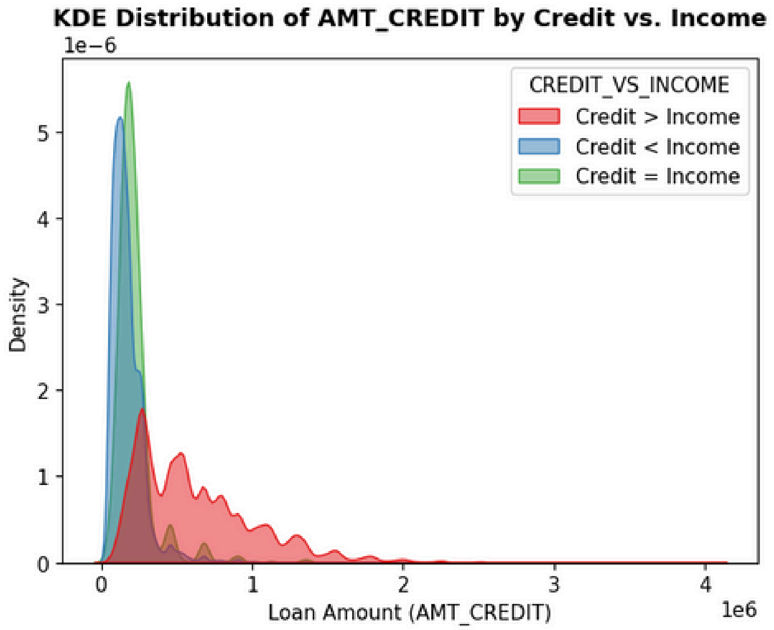
- Credit > Price has the highest median and many outliers, suggesting a pattern of borrowing more than the product value.
- Credit = Price has a moderate spread but remains relatively stable.
- Credit < Price appears unreliable or highly specific.

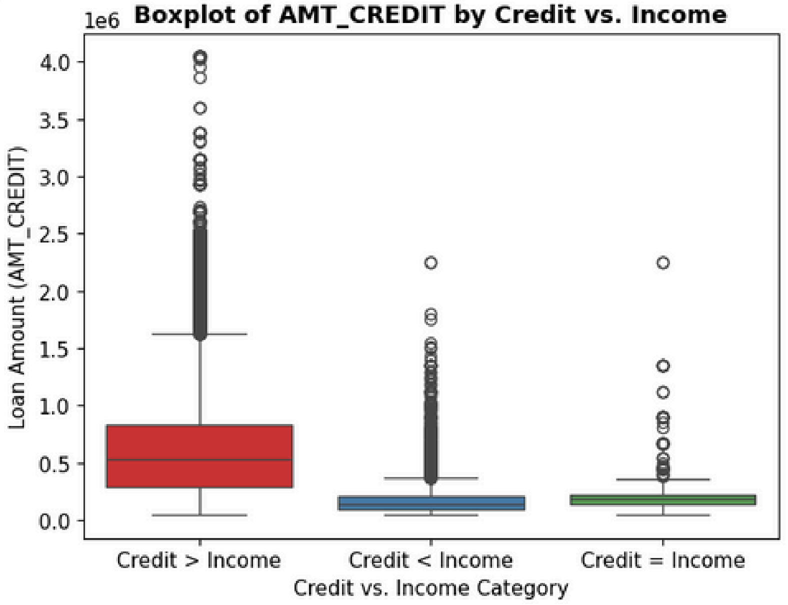
Insight:

- The patterns of Credit = Price and Credit > Price tend to move in parallel across AMT_CREDIT values.
- Credit > Price occurs more frequently than Credit = Price, indicating a strong trend of borrowers seeking extra funds beyond the value of the purchased goods.

Loan Value Analysis







📊 **Chart: Log-Log Scatter – AMT_INCOME_TOTAL vs AMT_CREDIT**

Observations:
- High-income individuals (orange) show up with many data points high on the Y-axis, meaning they take significantly larger loans compared to other groups.
- The Very Low, Low, and Medium income groups are more tightly constrained, usually falling in the "loan < income" zone.

Insight:
- High-income borrowers tend to take out larger loans due to stronger financial capacity, but they don't appear in large numbers, possibly due to credit ceilings that make the offering less attractive. More tailored products may help tap into this segment.
- Mid-income borrowers often take loans that far exceed their income, which could indicate more aggressive credit usage.
- Low-income borrowers take smaller loans, but these can still be risky if the loan amount approaches or exceeds their total income.

📊 **Chart: KDE & Boxplot – AMT_CREDIT Relative to Income**

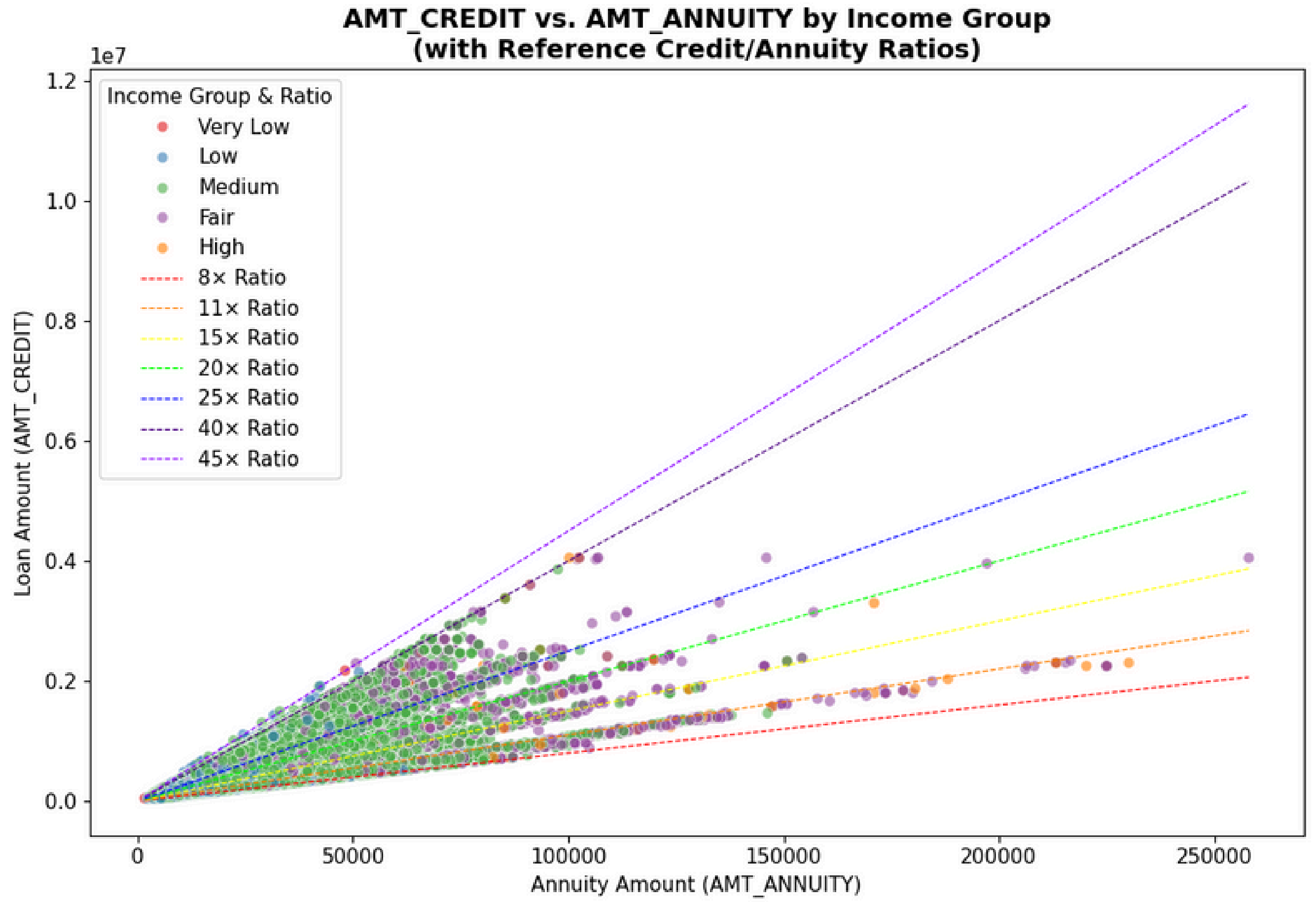
Observations:

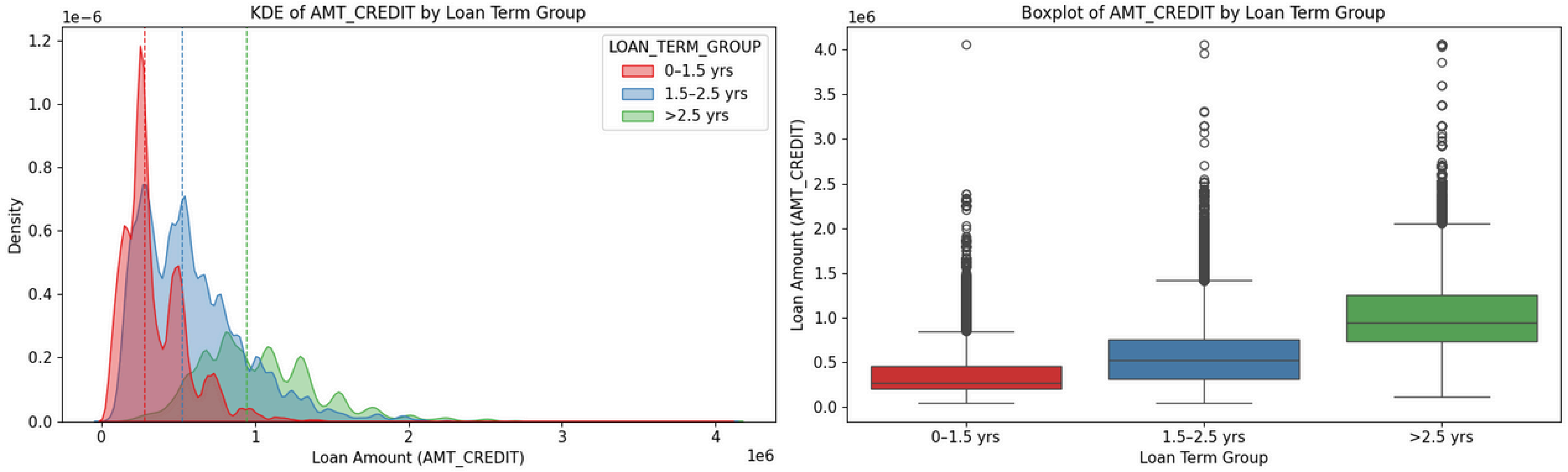
KDE:
- The Credit > Income group has a wide, right-skewed distribution, showing many large loans well above borrower income.
- Credit = Income and Credit < Income groups have tighter, more focused distributions.

Boxplot:
- The Credit > Income group has the highest median and widest spread, posing the greatest credit risk, especially due to outliers if credit history or collateral isn't properly evaluated.

Insight:
- The "Credit > Income" group should be flagged for elevated risk — especially if borrowers are in the low or middle-income range but are borrowing far beyond their means.
- Outliers in all groups should be reviewed carefully to ensure proper risk management and avoid unexpected defaults.

Loan Value Analysis





📊 **Chart: AMT_CREDIT vs. AMT_ANNUITY by Income Group (with Ratio Reference Lines)**

Observations:

- The reference lines such as 8×, 11×, ..., 45× represent CREDIT / ANNUITY ratios, which correspond to assumed loan terms (roughly 8 to 45 months).
- Most data points fall between the 11× and 25× lines, meaning loan terms of about 1 to 2.5 years are the most common.
- High-income borrowers (orange) are spread more evenly and have many data points in the higher ratio zones. Their annuity values also tend to be larger, suggesting shorter loan terms.

Insight:

- High-income customers may be willing to make larger monthly payments over shorter periods. After completing these loans, they could return for new financing — a favorable behavior that improves profitability and risk management.
- Some points exceed the 40× or 45× ratios, which may indicate long-term loans with high repayment pressure. These are risk-prone, especially if the borrower's income drops. This group is mainly composed of low to medium-income individuals — a critical area for focused risk control.

📊 **Chart 2: AMT_CREDIT by Loan Term Group**

Observations:

KDE:

- Short-term loans (0–1.5 years): small loan amounts, very tight distribution.
- Mid-term loans (1.5–2.5 years): broader distribution.
- Long-term loans (>2.5 years): right-skewed with a long tail, many large loan amounts.

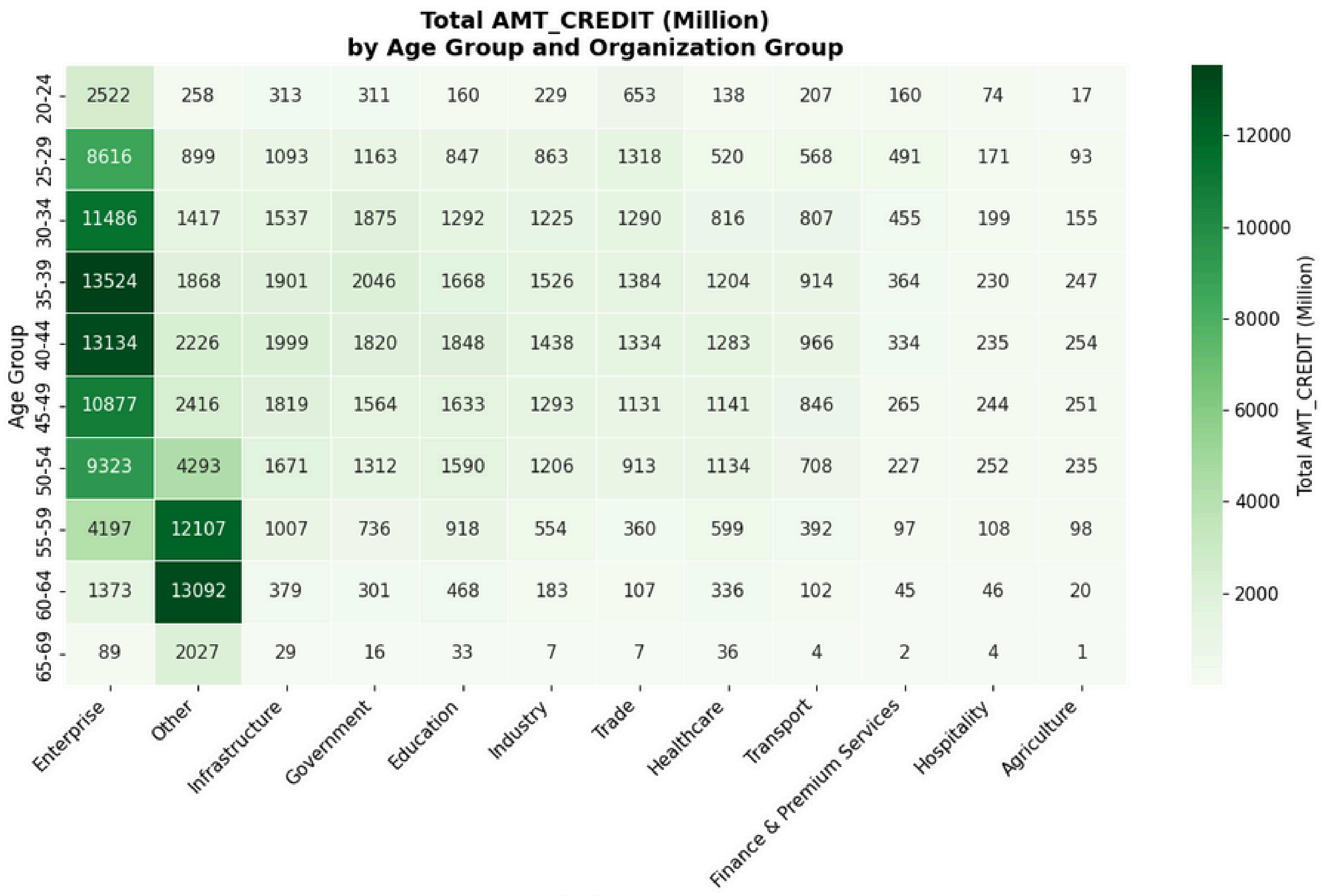
Boxplot:

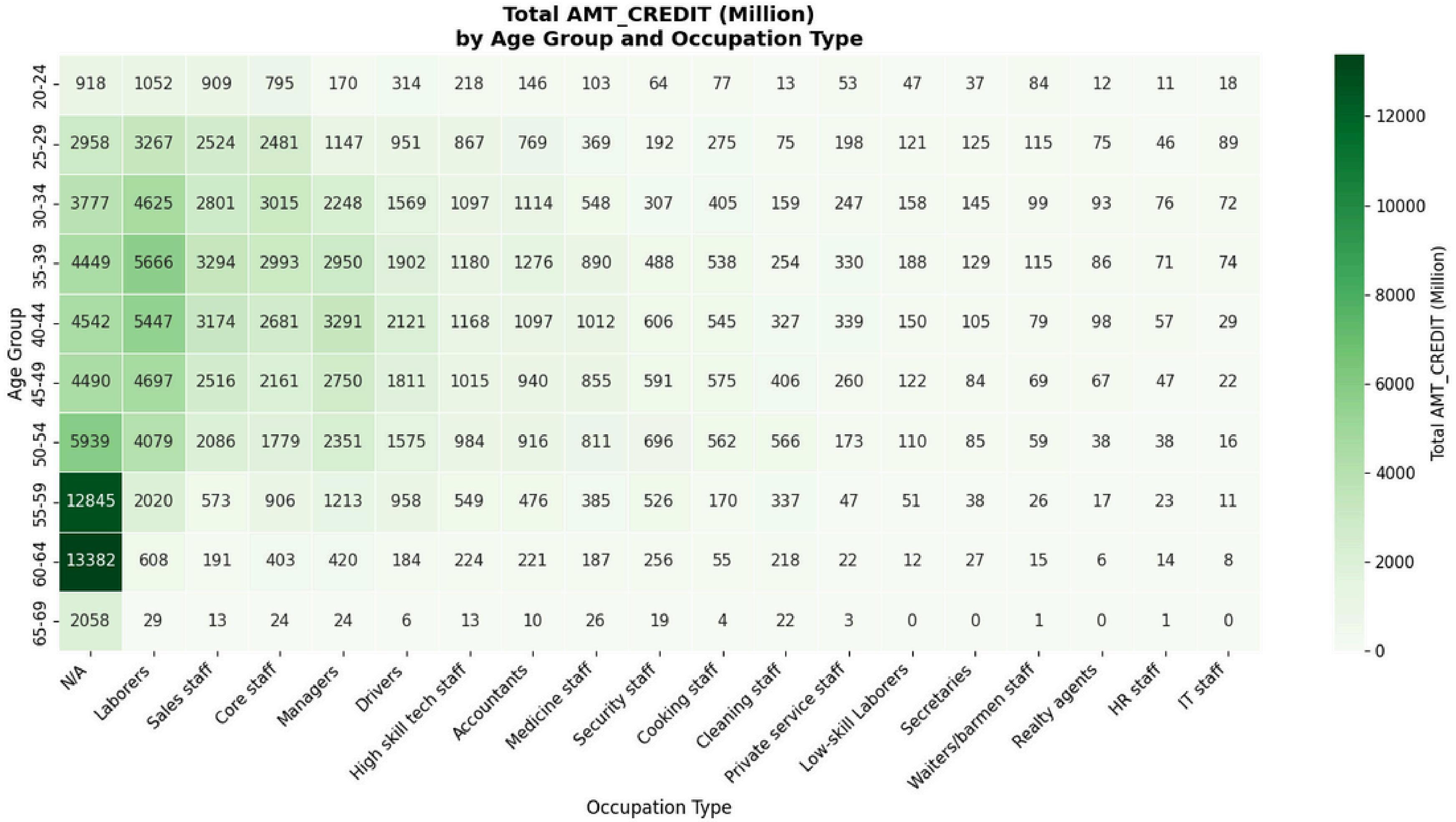
- The longer the loan term, the higher the average AMT_CREDIT — which is logical, but also increases the risk of accumulating debt if not properly monitored.

Insight:

- Most customers borrow within the 1–2.5 year range — long enough to ease monthly payments, but not long enough to generate high credit risk.
- Long-term loans should be carefully assessed based on borrower income, collateral, or job stability.
- The CREDIT/ANNUITY ratio can serve as an indirect indicator of loan term reasonableness and should be considered for use in scoring models.

Loan Value Analysis





📊 Heatmap: Total AMT_CREDIT by Age Group and Organization Type

Observations:

- Enterprise and Other groups contribute the vast majority across nearly all age ranges, especially:
  - Enterprise: strong presence in ages 35–49.
  - Other: spikes sharply in ages 55–64 and 60–64 (12,107 and 13,092 billion), which requires further verification.
- Trade, Healthcare, and Education show more stable distributions, spread evenly across middle-aged groups.

Insight:

- The "Other" group contributes disproportionately among older individuals or retirees — this could indicate incomplete or undefined records, raising concerns about fraud or misclassification risks.
- Enterprise appears to be the most stable organization type, with strong credit activity in the mid-age segment — ideal for targeting with personal loan and insurance packages.
- Sectors like Education, Government, and Healthcare show steady, reliable customer profiles, suggesting potential to develop special loan offers for public sector employees to boost engagement.

📊 Heatmap: Total AMT_CREDIT by Age Group and Occupation

Observations:

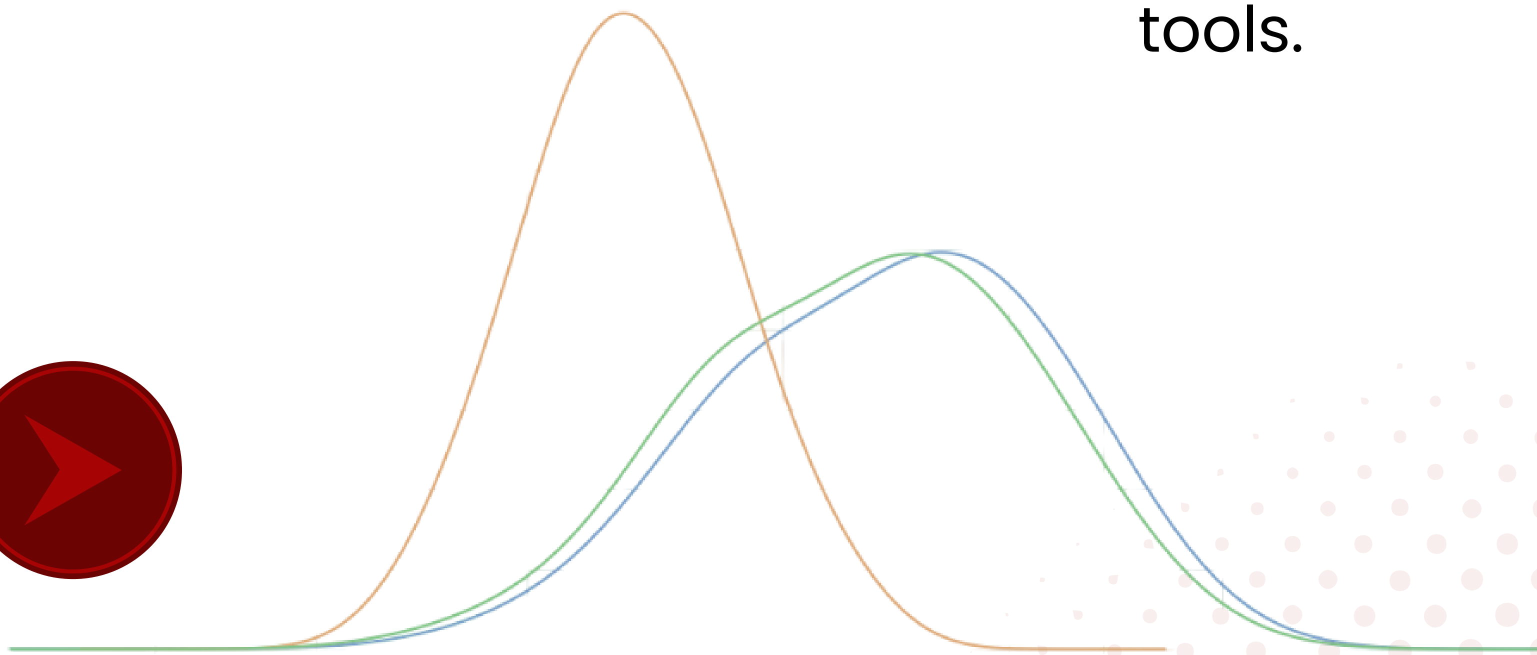
- The highest total credit is found in:
  - Age group 55–64 with occupation marked as N/A (13,382 billion) — possibly retirees, self-employed individuals, or records lacking clear occupational data.
  - Laborers and Sales staff in the 35–49 age range.
- The 30–44 age range is the "golden zone" for total lending across most occupations.

Insight:

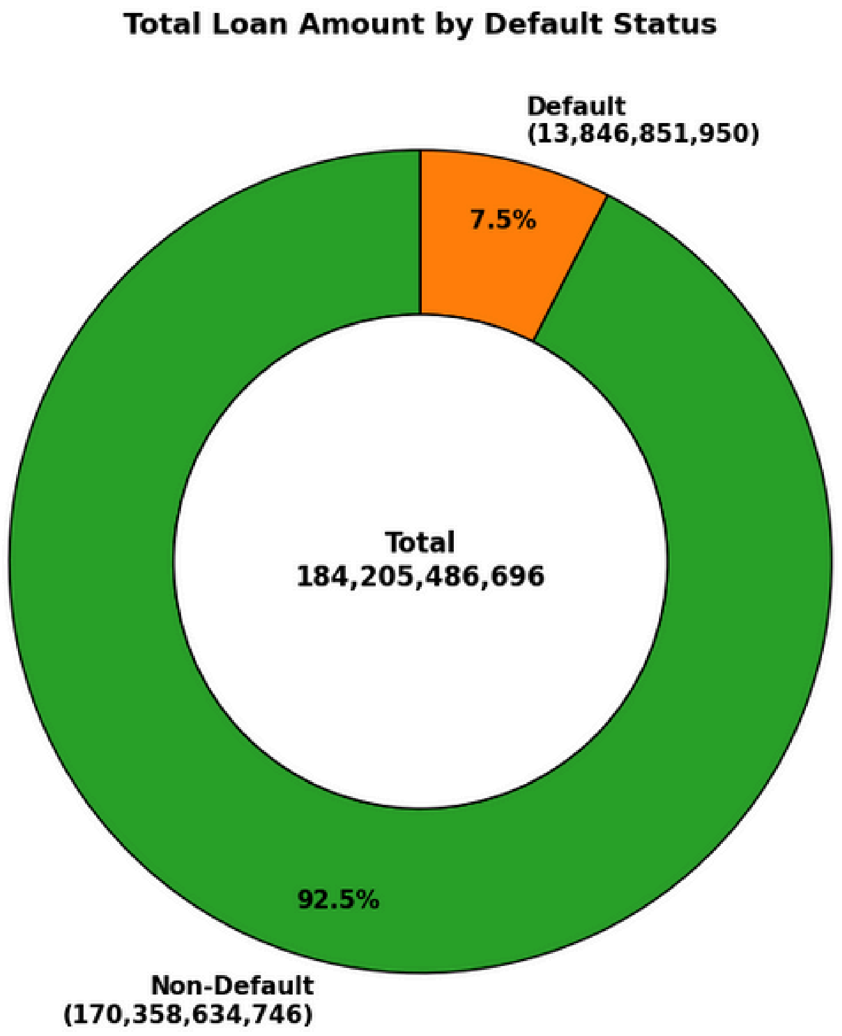
- The 35–44 age group represents the core borrower segment, as they are typically in a stable career phase and have rising financial needs (housing, cars, children, etc.).
- Sales staff and Laborers are two of the most common and high-volume borrower groups — custom loan products for them could boost engagement.
- The high credit total in the N/A category among older individuals (55–64) should be reviewed for potential risks related to unclear or missing occupational data.

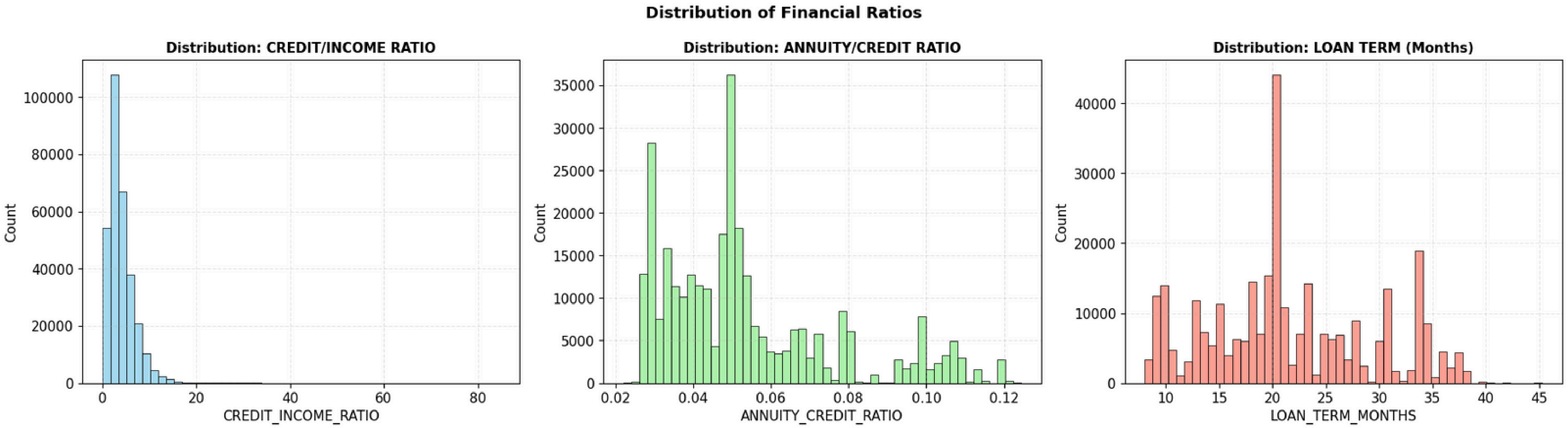
Risk Analysis

Risk Management and Analysis are critical in lending operations. In this section, we'll explore the company's financial health and propose several approaches to support risk management. Key financial ratios will be introduced as measurement tools.



Risk Analysis





📊 CHART: TOTAL LOAN VALUE BY DEFAULT STATUS

Observations:

Total loan portfolio: 184.2 billion

Of that:

- 92.5% (170.36 billion) is non-default
- 7.5% (13.85 billion) is defaulted

Insight:

- A 7.5% default rate by value is fairly reasonable for consumer lending, assuming profit margins are high enough to offset the risk.
- However, given the portfolio size, 13.85 billion in defaulted loans is a significant amount and should be closely monitored — especially to identify which segments are driving the concentration of risk.

📊 CHART: DISTRIBUTION OF FINANCIAL RATIOS

CREDIT/INCOME Ratio:

- Strong right-skewed distribution, with most customers having a ratio between 0–5.
- However, a small group exceeds 10, and even over 20 times their income — this group should be closely monitored for risk.

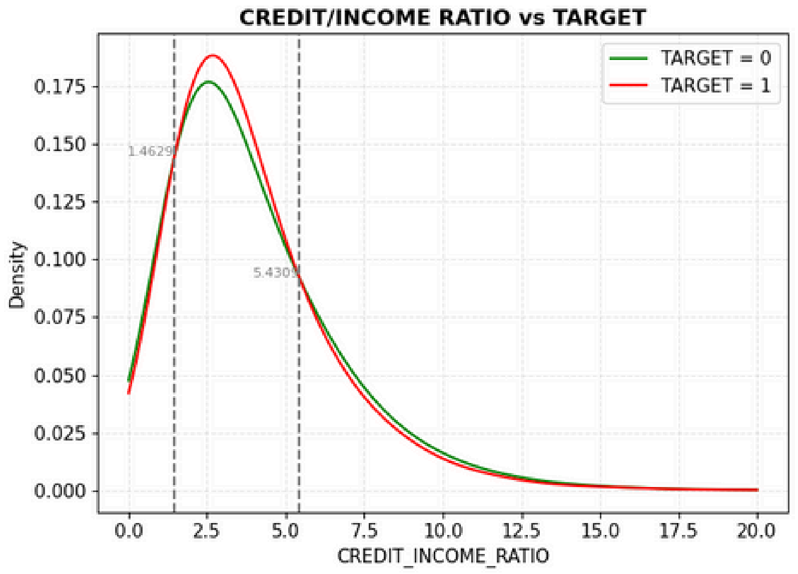
ANNUITY/CREDIT Ratio:

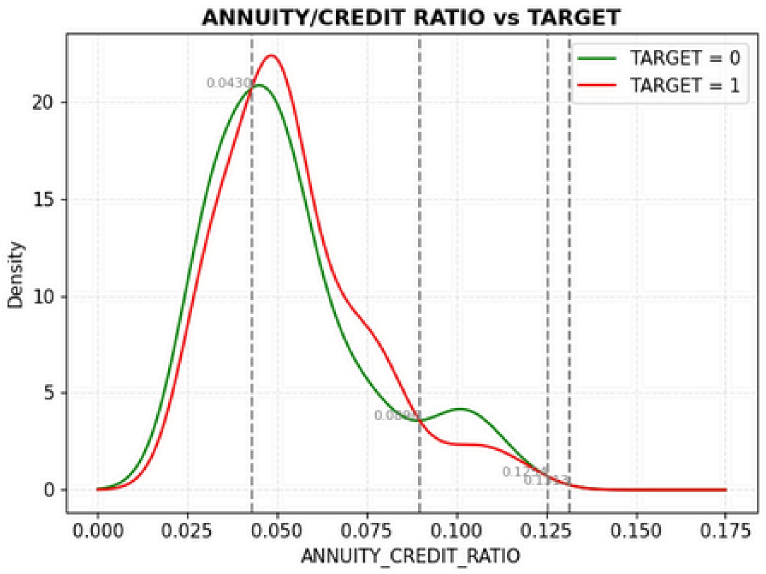
- Ranges from approximately 0.03 to 0.06 (3%–6%), aligning with monthly installment logic (assuming loan terms of 1–3 years).
- The peak of the distribution is around 5%, reflecting a common loan structure.

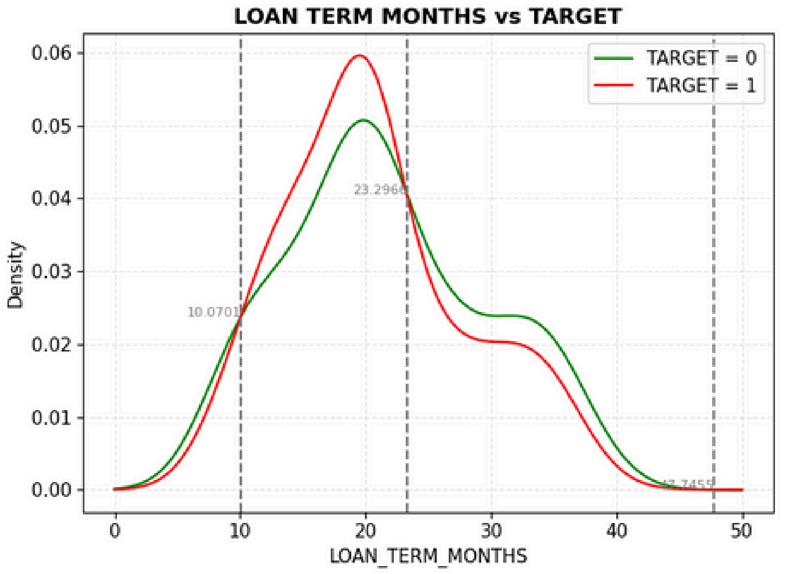
LOAN TERM (Months):

- Scattered distribution, but peaks sharply at 20 months.
- The common range is 12–36 months, confirming insights from previous charts about typical loan durations of 1.5 to 2.5 years.

Risk Analysis








```
📊 CREDIT_INCOME_RATIO segmentation:
🟢 Low:
  - Range            : < 1.4629
  - Default Rate     : 7.08%
  - Customers        : 38,929
  - Default Exposure : 609,145,438
🟡 Medium:
  - Range            : 1.4629 → 5.4309
  - Default Rate     : 8.54%
  - Customers        : 198,078
  - Default Exposure : 8,491,452,574
🔴 High:
  - Range            : > 5.4309
  - Default Rate     : 7.3%
  - Customers        : 70,488
  - Default Exposure : 4,746,253,936
💥 Total Default Exposure: 13,846,851,950
```

```
📊 ANNUITY_CREDIT_RATIO segmentation:
🟢 Low:
  - Range            : < 0.0430
  - Default Rate     : 6.39%
  - Customers        : 114,971
  - Default Exposure : 6,566,045,242
🟡 Medium:
  - Range            : 0.0430 → 0.0896
  - Default Rate     : 9.92%
  - Customers        : 157,902
  - Default Exposure : 6,802,560,878
🔴 High:
  - Range            : > 0.0896
  - Default Rate     : 5.28%
  - Customers        : 34,622
  - Default Exposure : 478,245,830
💥 Total Default Exposure: 13,846,851,950
```

```
📊 LOAN_TERM_MONTHS segmentation:
🟢 Low:
  - Range            : < 10.0701
  - Default Rate     : 5.64%
  - Customers        : 22,771
  - Default Exposure : 313,784,438
🟡 Medium:
  - Range            : 10.0701 → 23.2966
  - Default Rate     : 9.54%
  - Customers        : 169,867
  - Default Exposure : 6,969,879,086
🔴 High:
  - Range            : > 23.2966
  - Default Rate     : 6.39%
  - Customers        : 114,857
  - Default Exposure : 6,563,188,426
💥 Total Default Exposure: 13,846,851,950
```

Note: By fine-tuning the KDE curves, we can quickly identify regions with higher risk indicators based on where the curves intersect. Specifically, when the red line rises above the green line, it signals that area has elevated risk levels compared to surrounding regions.

📊 Summary of 3 Key Risk Indicators & Customer Segments

CREDIT / INCOME RATIO

Insight:

- The medium segment has the largest number of customers and the highest default rate, making it the main contributor to portfolio risk.
- Extra caution is needed when approving large loans or long terms for this group.

ANNUITY / CREDIT RATIO

Insight:

- The medium group shows the highest default rate (9%) and significant exposure, indicating a risk of accumulating debt.
- Set up alerts when ANNUITY/CREDIT falls within the 0.04–0.09 range.

LOAN TERM (Months)

Insight:

- The 10–23 month range is the riskiest, with the highest default rate (9.54%).
- These mid-term loans should be flagged, especially when combined with a low installment ratio (medium ANNUITY/CREDIT) and low income.

🎯 Scoring Strategy

Based on the 3 indicators and customer segments mentioned, a simple scoring logic can be built as follows:

Risk Score Suggestions by Feature:

- CREDIT/INCOME > 5.43 (+2)
- ANNUITY/CREDIT ∈ [0.043–0.0896] (+3)
- LOAN TERM ∈ [10–23 months] (+2)

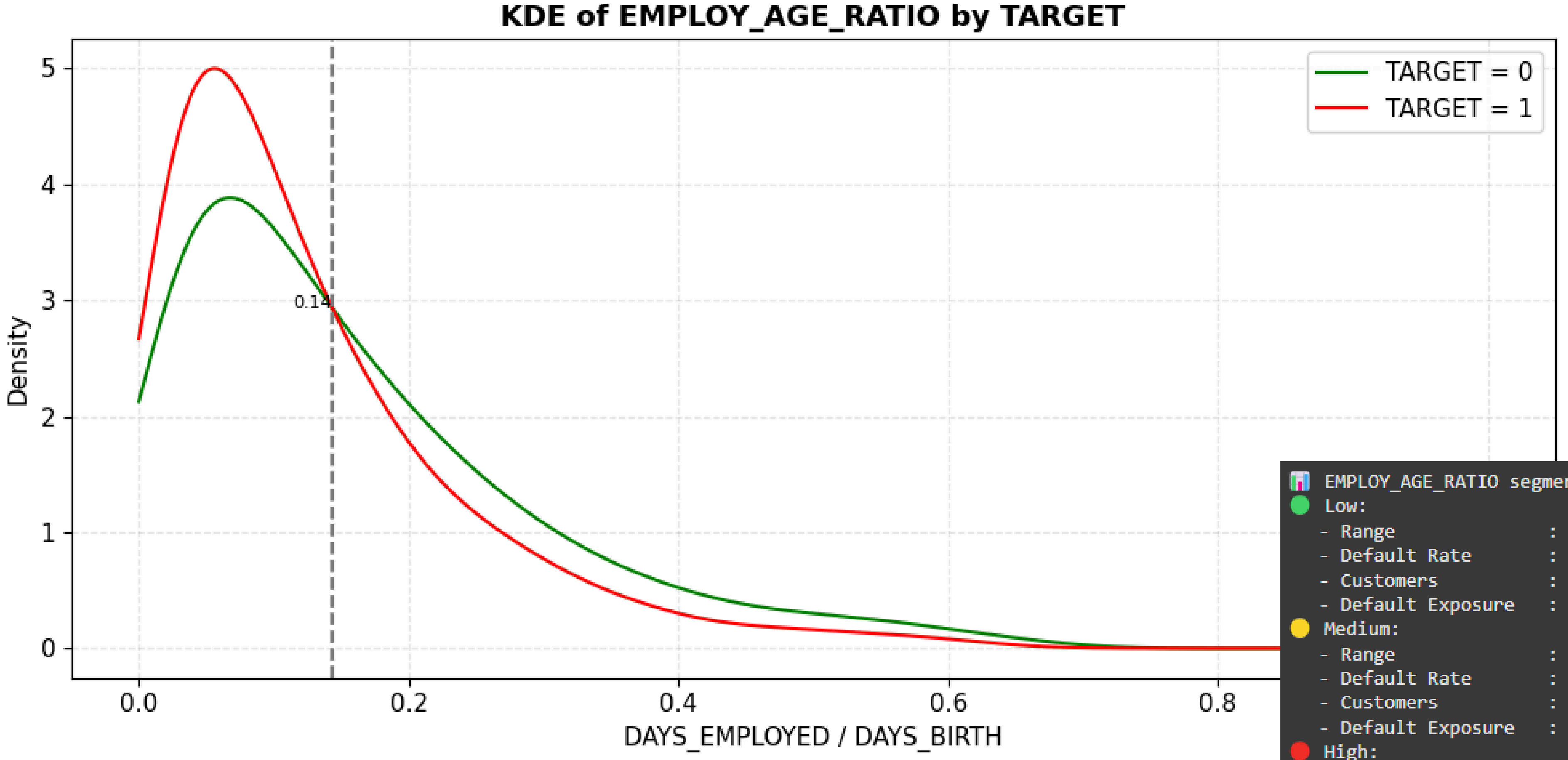
- CREDIT/INCOME < 1.46 (–2)
- ANNUITY/CREDIT > 0.09 (–2)
- LOAN TERM < 10 months (–1)

Then, sum the score and classify:

- ≤ 0: Safe
- 1–3: Medium – further review of occupation/industry/income
- ≥ 4: High risk alert – verify collateral and loan history

Note: The scoring method and weightings can be adjusted depending on the context and changes in borrower behavior and indicators.

Risk Analysis




```
📊 EMPLOY_AGE_RATIO segmentation:
🟢 Low:
  - Range              : < 0.1433
  - Default Rate       : 10.15%
  - Customers          : 144,916
  - Default Exposure   : 7,969,271,067
🟡 Medium:
  - Range              : 0.1433 → 0.6600
  - Default Rate       : 6.67%
  - Customers          : 106,822
  - Default Exposure   : 4,205,907,688
🔴 High:
  - Range              : > 0.6600
  - Default Rate       : 0.76%
  - Customers          : 395
  - Default Exposure   : 2,588,409
💥 Total Default Exposure: 12,177,767,164
```

This method can also be applied to other indicators, not just financial ones.
This chart shows the ratio between the employment start date and date of birth. We can easily identify safe zones and risk zones within customer data, and use this to track shifts in risk patterns over time.

📊 KDE: EMPLOY_AGE_RATIO vs TARGET

Observations:

- A low ratio (< 0.1433) indicates that the borrower has a short work history relative to their age (e.g., recent graduate, unstable career). This group has the highest default rate: 10.15%.
- The medium range (0.1433–0.66) is the most common and carries a lower risk: 6.67%.
- A high ratio (> 0.66) means the individual has been working almost their entire life, with an extremely low risk of default: only 0.76%.

📊 EMPLOY_AGE_RATIO Segmentation

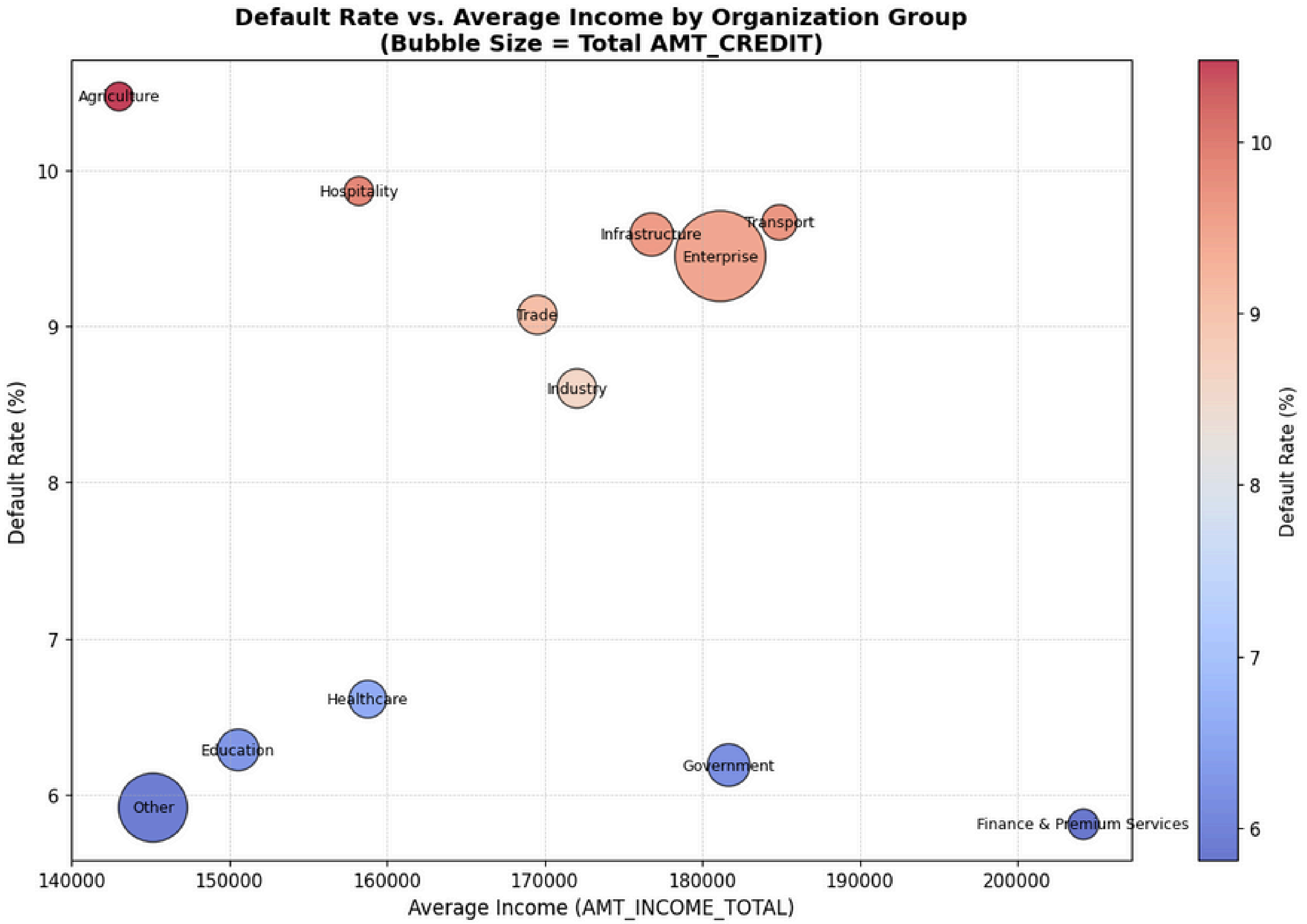
- Customers with a ratio above 0.66 are extremely stable and should be prioritized for approval, possibly even offered special incentives.
- By simply adding this variable to the model, we can optimize and significantly reduce default risk.

SCORING INTEGRATION SUGGESTION:

- If EMPLOY_AGE_RATIO > 0.66, subtract 3 points (very safe).
- If EMPLOY_AGE_RATIO is between 0.1433 and 0.66, no adjustment (stable).
- If EMPLOY_AGE_RATIO < 0.1433, add 3 points (high risk).

Combined with CREDIT/INCOME, ANNUITY/CREDIT, and LOAN TERM, this is a highly effective supporting indicator to avoid defaults — especially from younger customers or those with frequent job changes.

Risk Analysis



📊 Bubble Chart: Default Rate vs. Average Income by Organization Type

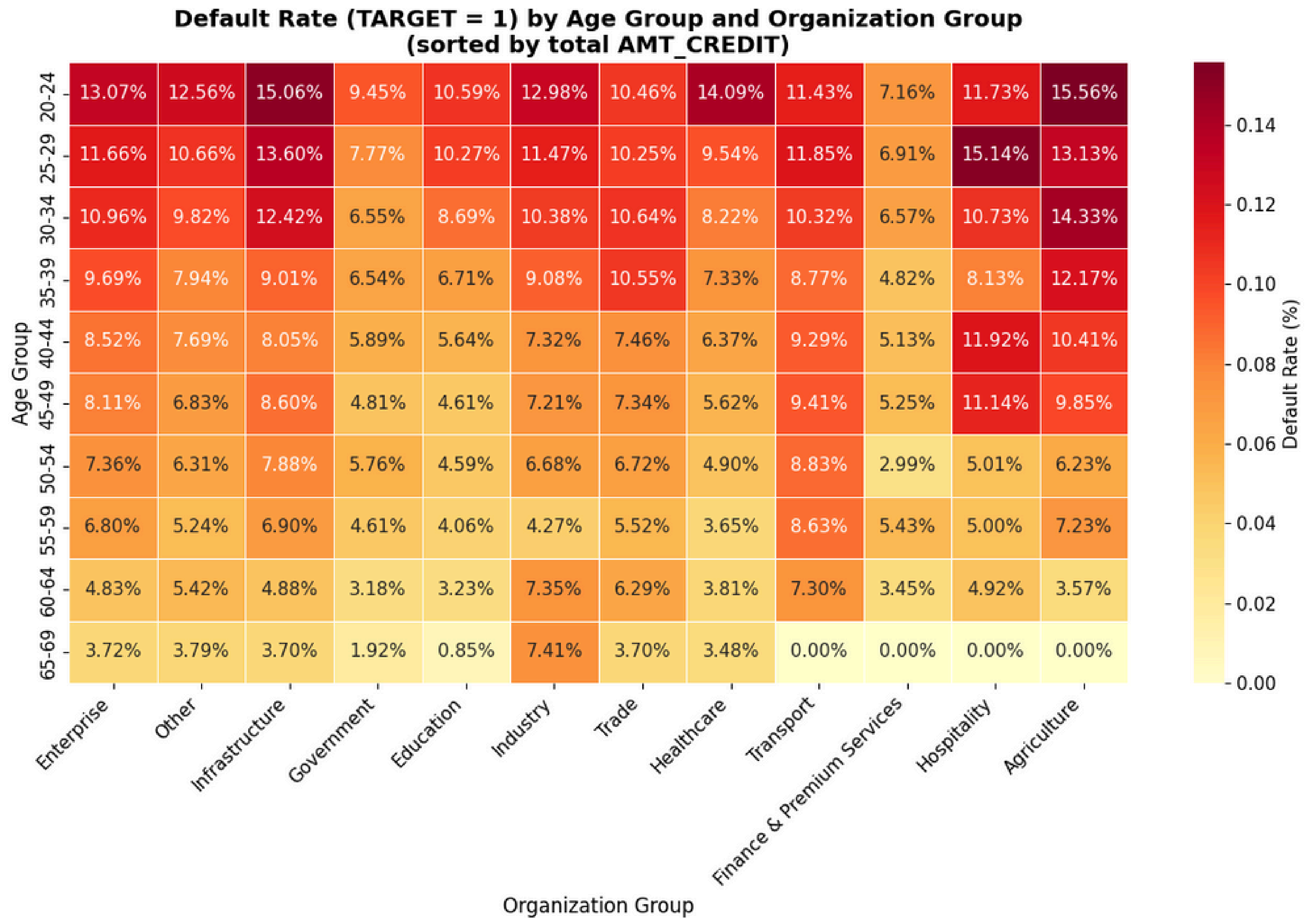
Observations:

- Organizations on the chart are clearly clustered along two main dimensions: default rate and average income.
- High-risk group (default rate 9% and above): Agriculture, Hospitality, Transport, Enterprise, Infrastructure, Trade.
- Low-to-moderate risk group (default rate 6–8%): Industry, Healthcare, Education, Government, Finance & Premium Services, Other.

Insight:

- High-risk, low-loan-value groups:
  - Agriculture and Hospitality stand out with high default rates. Agriculture, in particular, has clearly small loan sizes (small bubble size). These groups require special attention, as even though the loan amounts are small, the default risk is very high.
- Low-risk groups with growth potential:
  - Finance & Premium Services and Government have very low default rates (around 6%). Notably, Finance & Premium Services have both the highest income and the lowest default rate, making this a highly promising group for expanding credit (AMT_CREDIT).
  - Healthcare and Education are also worth considering for expansion due to relatively low risk and stable incomes.
- Saturated groups needing tighter risk control:
  - Enterprise is the largest group in terms of total AMT_CREDIT and carries a relatively high default rate (around 9%). This indicates saturation. Loan approval criteria and credit monitoring should be tightened to prevent further increases in default.
  - Infrastructure and Transport also need attention, as they have high default rates (~9–10%) and significant credit volume.

Risk Analysis



📊 Heatmap Chart: Default Rate by Age Group and Organization Type

General Observations:

- Default rates clearly decline as customer age increases, indicating that age is a key factor in repayment ability.
- Younger age groups (20–34) show the highest default rates, while older groups (55–69) demonstrate much higher stability and very low default rates.

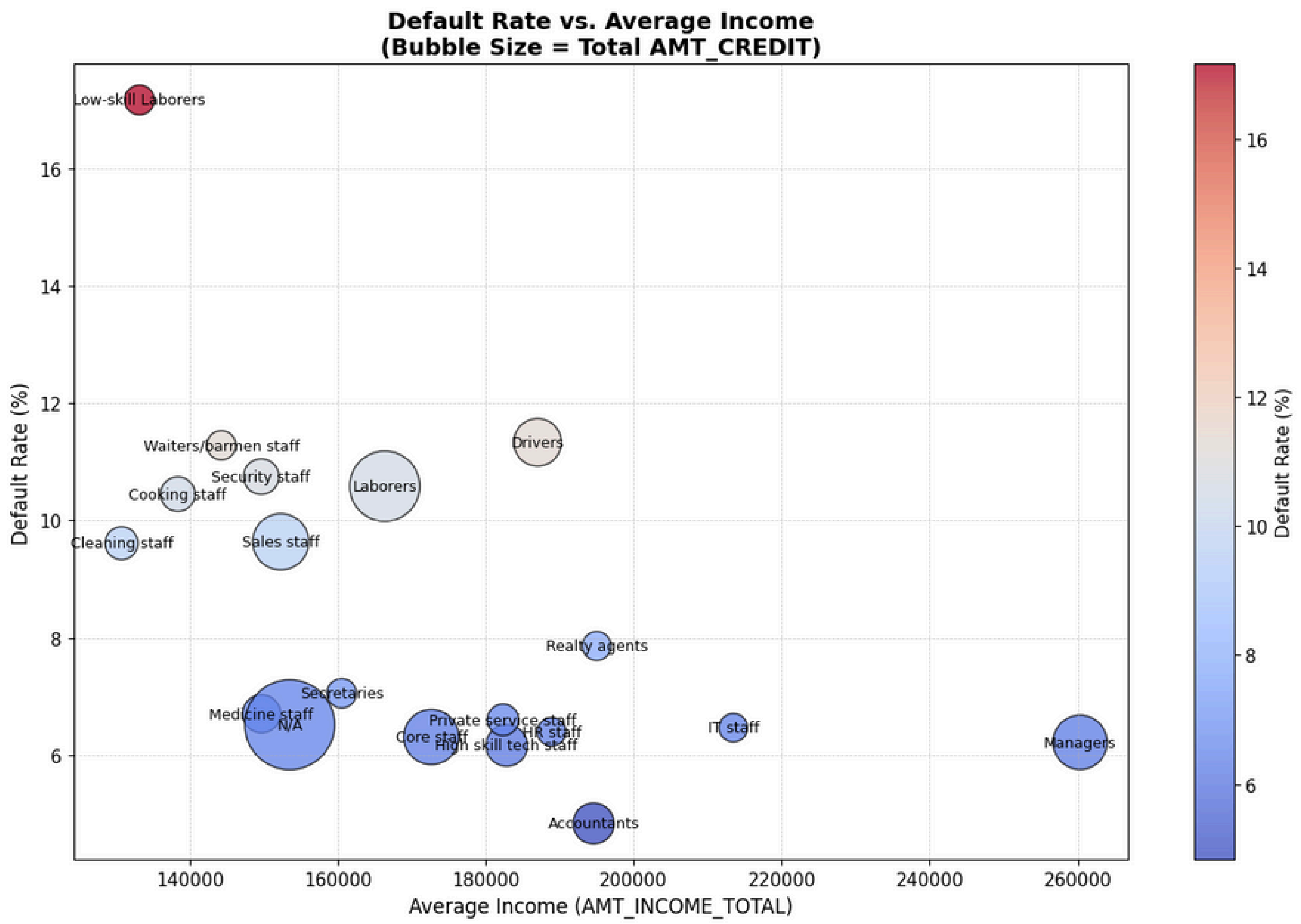
Detailed Analysis by Organization Group:

1. High-Risk Groups:
   - Agriculture and Hospitality continue to show the highest risk, especially among younger customers (20–34 years old).
   - In the youngest age group (20–24), Agriculture reaches a default rate as high as 15.56%, and Hospitality also shows an elevated rate at 11.73%.
   - These two sectors clearly concentrate their risk in the young customer segment.
2. Groups Requiring Tight Monitoring:
   - Enterprise, Infrastructure, Transport, and Trade show default rates above 10% among young borrowers, though slightly lower than Agriculture and Hospitality.
   - Default rates decline significantly after age 40, but these segments still require close risk management.
3. More Stable, Promising Groups:
   - Finance & Premium Services, Government, Healthcare, and Education all show lower default rates, which drop quickly with age, especially stabilizing around age 50 and older (often below 5%).
   - These sectors are strong candidates for growth-focused lending, particularly among middle-aged and older customers.
4. Very Safe and Stable Groups by Age:
   - Government and Finance & Premium Services stand out with near-zero default rates in the 60–69 age group, representing extremely low-risk customers.

Insight:

- Focus risk control efforts on sectors like Agriculture, Hospitality, Enterprise, and Transport, especially for younger customers (20–34 years old).
- This may include stricter loan approval criteria or lower credit limits.
- Accelerate credit growth for sectors like Government, Finance & Premium Services, and Healthcare in the middle-aged and older segments, to tap growth potential while maintaining portfolio stability.
- For high-risk younger groups, consider adding collateral requirements or interest rate adjustments to help mitigate losses.

Risk Analysis



📊 **Bubble Chart: Default Rate vs. Average Income by Occupation**

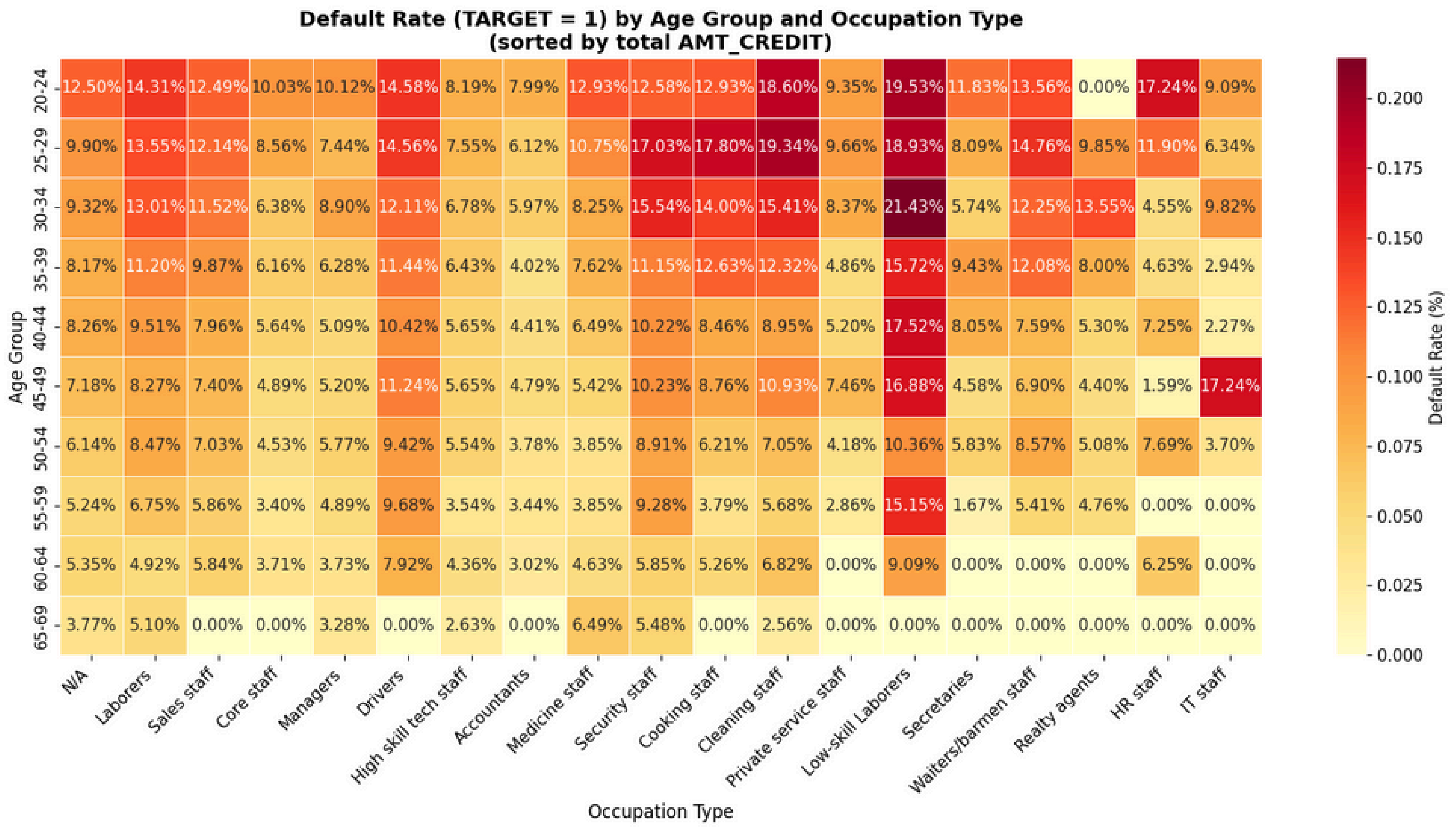
Observations:

- Occupational groups are clearly segmented by default risk and average income level:
  - Low income (left) and high risk (top): mostly manual labor and service jobs.
  - High income (right) and low risk (bottom): mainly professional and senior management roles.

Insight:

- Highest-risk group requiring special attention:
  - Low-skill Laborers stand out sharply with an extremely high default rate (>16%) and very low income. Although their loan amounts (AMT_CREDIT) are small, this group poses the greatest default risk and must be strictly controlled.
- Moderately high-risk groups needing tighter control:
  - Drivers, Waiters/Barmen, Cooking staff, Security staff, Cleaning staff, Laborers all show default rates around 9–12% with low to lower-middle incomes.
  - These groups should be subject to tighter credit policies, including collateral requirements or lower loan limits.
- Low-risk groups with growth potential:
  - Managers, Accountants, IT staff, High-skill tech staff, HR staff, Core staff, and Medicine staff all have high to very high incomes and very low default rates (5–7%).
  - This is a high-potential segment for expanding credit services.
- Moderately low-risk, stable-income group with room for growth:
  - Private service staff, Secretaries, Realty agents have default rates around 7–9% and stable income.
  - While their AMT_CREDIT is still relatively small, this group presents an opportunity for targeted marketing and loan product expansion.

Risk Analysis



📊 Heatmap Chart: Average Default Rate by Occupation and Age Group

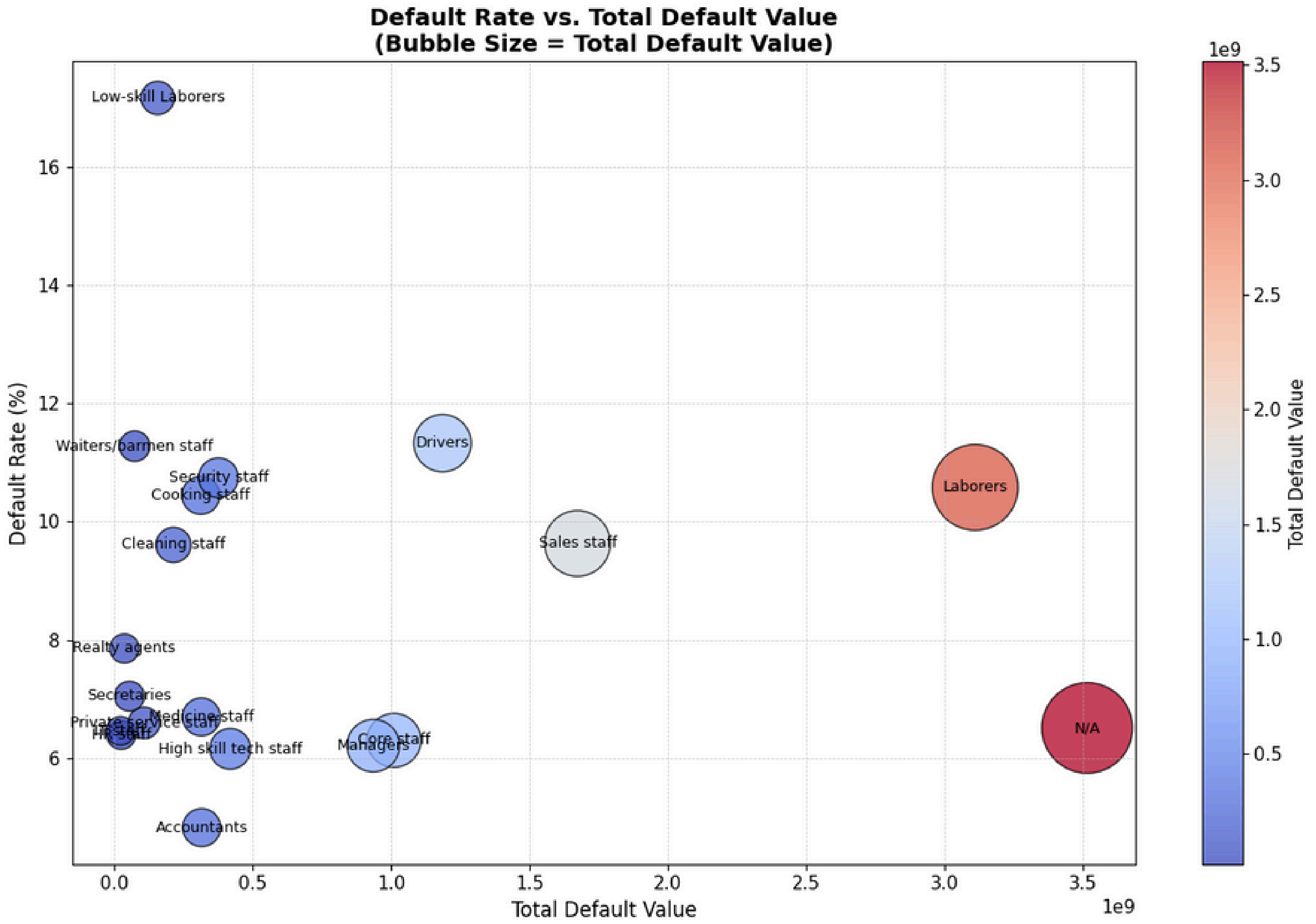
General Observations:

- Overall trend:
  - Default rates decrease significantly with age — the older the customer, the lower the risk.
  - There is a clear risk split among occupational groups, especially among manual labor jobs.
  - Age groups with 0% default rates are often those with few or no loan contracts, not necessarily risk-free.

Insight:

- Extremely high-risk group:
  - Low-skill Laborers and Waiters/Barmen staff show exceptionally high risk, especially among younger customers (ages 20–34).
  - Waiters/Barmen staff (20–24 years old): default rate reaches 19.53%.
  - Low-skill Laborers (25–34 years old): ranges from 17.80% to 21.43%, which is critically high.
  - These groups should be strictly limited or have special risk control policies when issuing credit.
- High-risk groups to watch:
  - Drivers, Cooking staff, Security staff, Laborers, and Sales staff also show high default rates (typically 9–14% in younger age groups).
  - Particularly, Drivers and Cooking staff hit even higher rates (14–19%), requiring tighter risk management.
- Moderate-to-low risk groups with growth potential:
  - Managers, Accountants, IT staff, HR staff, Core staff, High-skill tech staff, and Medicine staff have low default rates, even among younger customers (~5–9%).
  - These professions are consistently stable across age groups, making them ideal targets for safe and effective credit expansion.
- Safe and underutilized customer group:
  - Accountants, Managers, IT staff, HR staff, and Realty agents over the age of 50 show near-zero default rates.
  - This may also reflect an untapped segment with relatively few credit applications. Given their high income and low risk, this is a prime target for loan portfolio growth.

Risk Analysis



📊 **Bubble Chart: Default Value vs. Average Income by Occupation**

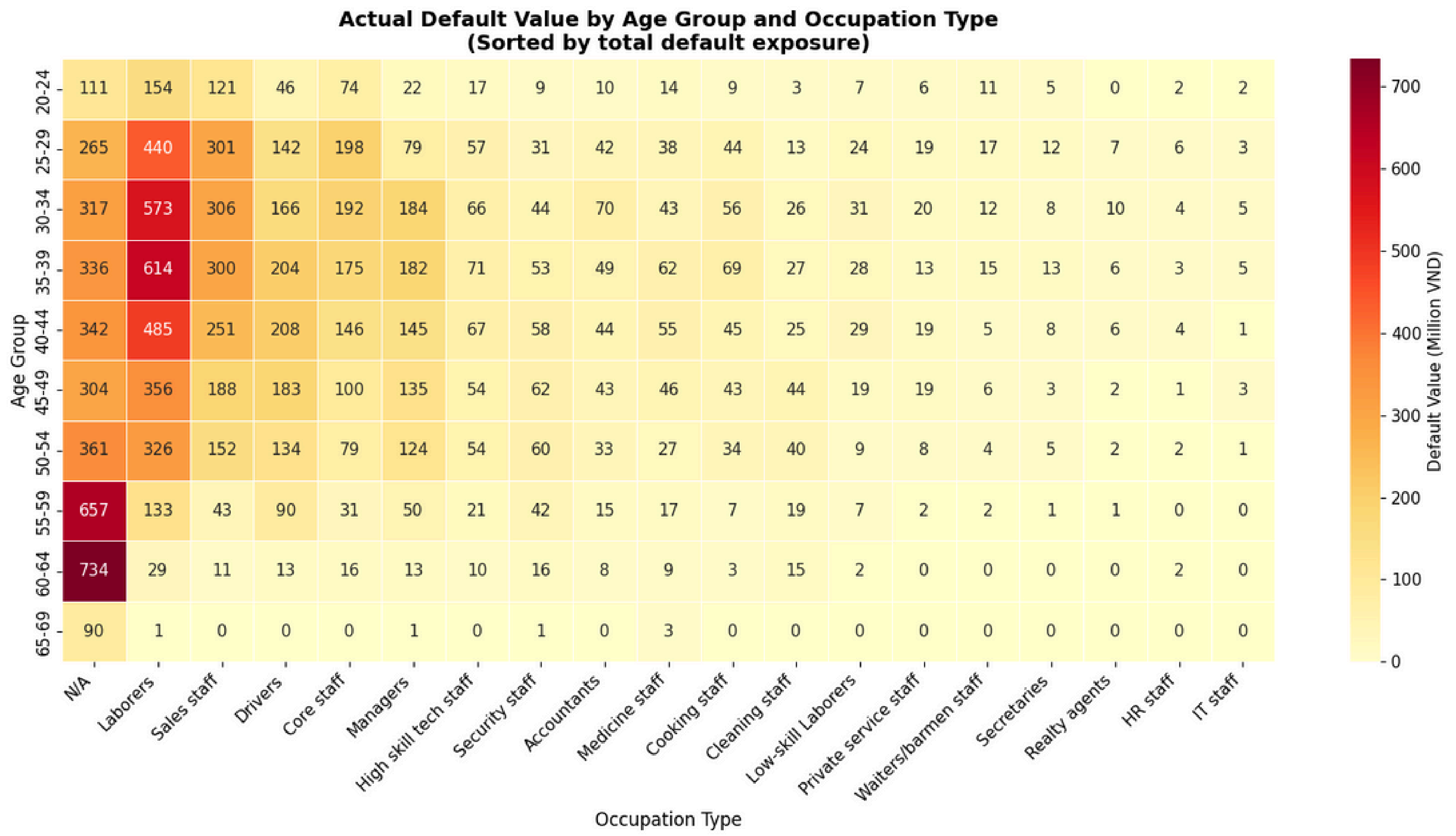
Observations:

- This chart offers a new perspective by showing the total default value alongside the default rate (%), helping identify which occupations contribute most to total losses.

Insight:

- High-impact risk groups by total default value:
  - Laborers:
    - Total default value is very high (~3 billion), with a default rate around 10%. This group causes the largest absolute losses, even though its default rate isn't the highest. Risk management strategy should focus on reducing total exposure, such as tightening loan amounts or requiring stronger collateral.
  - N/A (Unknown Occupation):
    - Has the highest total default value (~3.5 billion), with a default rate around 6–7%. Although the rate is relatively low, the scale is large. This group needs special attention in credit analysis — adding clearer occupation data can help improve control.
- High default rate but low default value groups:
  - Low-skill Laborers:
    - Extremely high default rate (>16%) but low total default value. Despite small amounts, the individual risk is very high, so credit to this group should be limited — especially if loan volume expands. Credit processes may need to be reevaluated for this group.
  - Waiters/Barmen, Cleaning staff, Cooking staff, Security staff:
    - Default rates are high (~10–12%), but total default values are low. These groups require tighter monitoring, particularly when scaling up lending.
- Mid-range default value, moderate-to-high risk groups:
  - Drivers and Sales staff:
    - Default values are significant (1–2 billion), with fairly high default rates (8–12%). These occupations carry both financial impact and risk, and should be subject to stricter credit controls.
- Stable, low-risk groups in both rate and value:
  - Managers, Accountants, IT staff, High-skill tech staff, Medicine staff, HR staff, Core staff:
    - These groups have low to moderate default values and very low default rates (~5–8%). They are stable and reliable segments, well-suited for aggressive credit expansion.

Risk Analysis



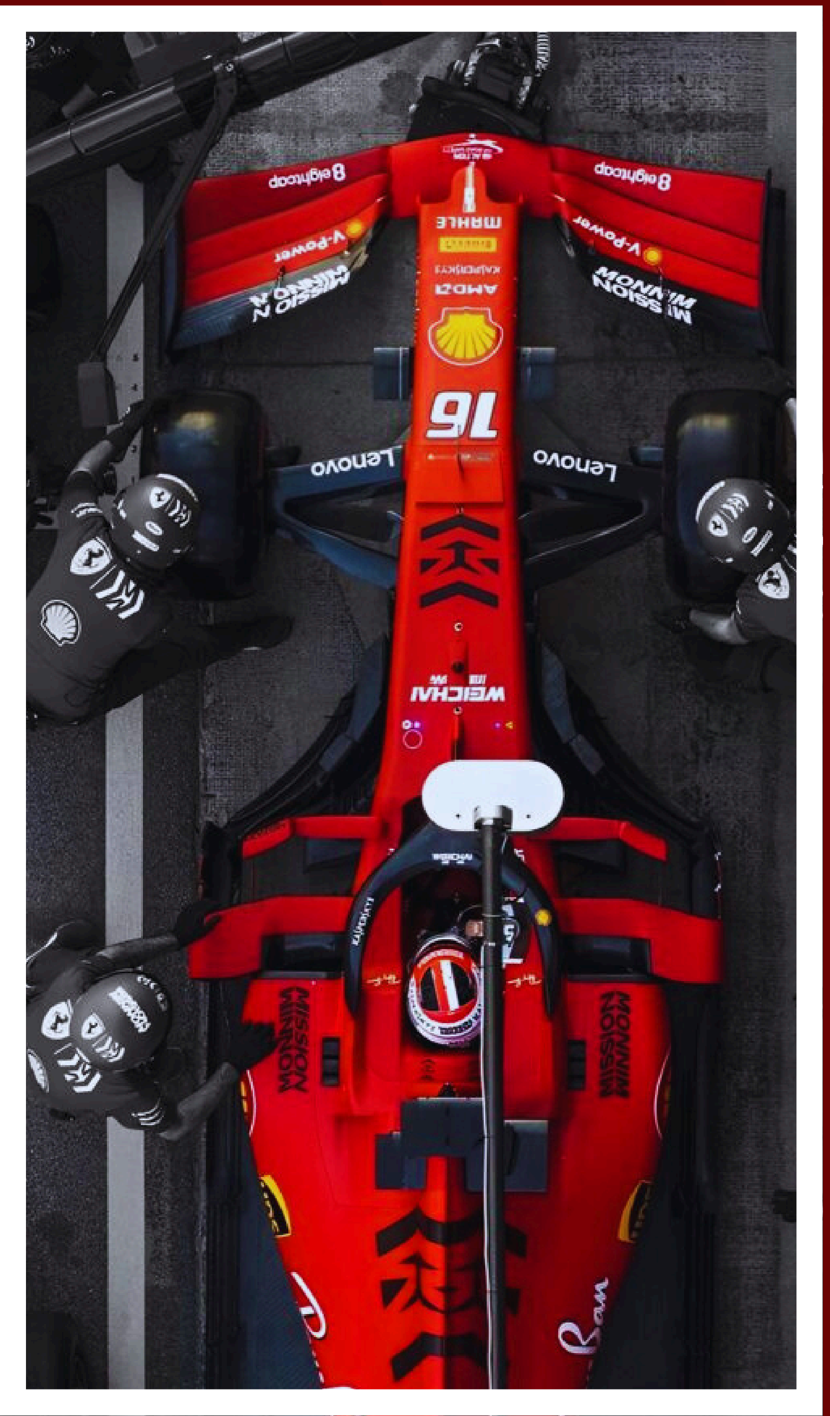
📊 **Heatmap Chart: Default Value vs. Average Income by Occupation**

<u>Observations:</u>

- This chart highlights actual financial loss (monetary value) across different age and occupation groups, allowing for precise identification of which segments cause the greatest financial damage.
- Of course, a group with high financial loss isn't necessarily the worst — other indicators like default rate and total credit volume must also be considered.

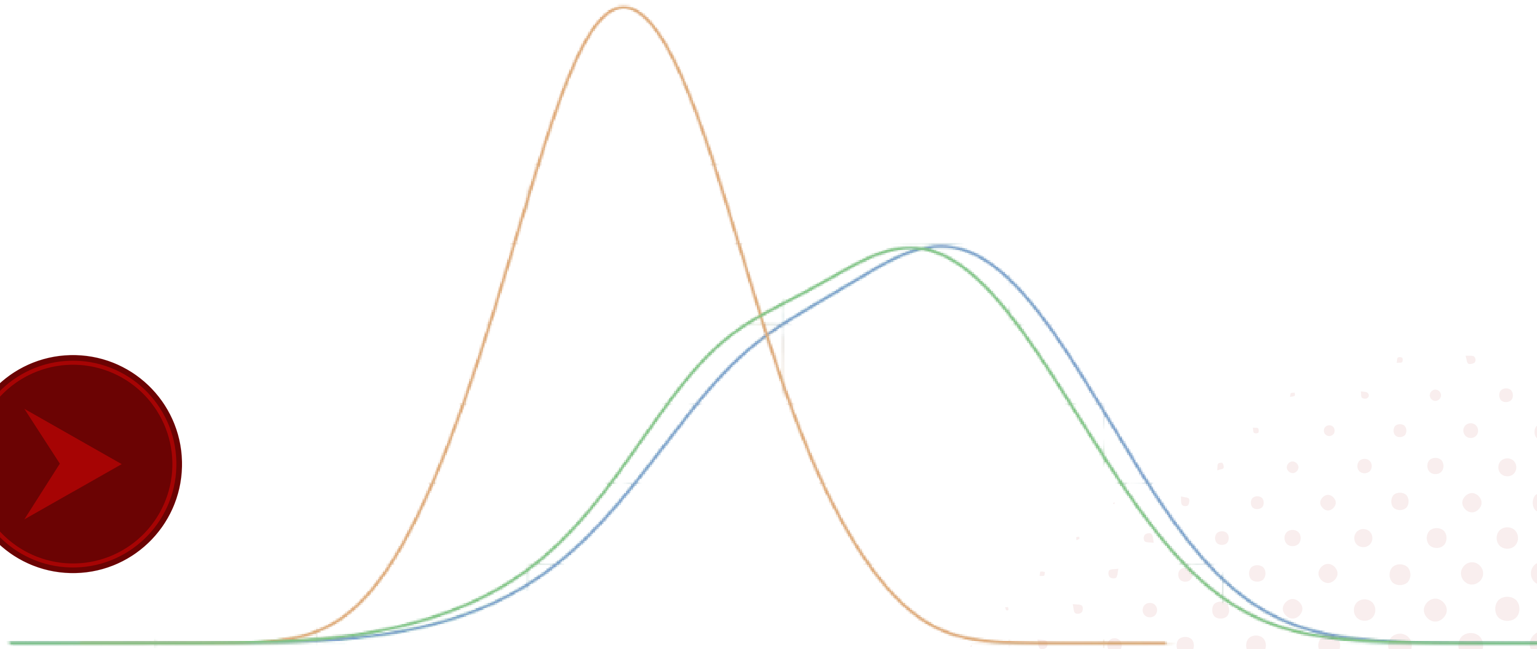
<u>Insight:</u>

- Group causing the highest financial loss:
  - N/A (Unknown Occupation):
    - Stands out with very high losses, especially among older customers (ages 55–64).
    - Age 60–64: 734 million
    - Age 55–59: 657 million
    - The lack of occupational data clearly introduces high risk. It's essential to improve customer information or apply tighter controls to this group.
  - Laborers:
    - Highest losses occur in the 35–39 age group: 614 million
    - In general, this group causes major losses in both young and middle-aged segments (25–44).
    - It requires strict control, and loan expansion should be limited.
- Groups with noteworthy losses:
  - Sales staff:
    - Highest losses in the 40–44 age group (485 million), with clear concentration from ages 25–44.
    - This group handles large credit volumes, so tighter credit management is needed.
  - Drivers:
    - Largest losses in 30–34 age group (198 million).
    - While not extremely high, losses are consistently present across younger age groups and should not be overlooked.
- Low to moderate loss groups:
  - Managers, Accountants, IT staff, High-skill tech staff, Medicine staff, HR staff:
    - Very low default values, typically under 100 million per age group.
    - These are very safe occupations, ideal for credit growth.
- Low-loss groups that still require oversight:
  - Waiters/Barmen, Cleaning, Cooking, Security staff, Low-skill Laborers:
    - Though default values are low, previous charts show high default rates.
    - Thus, these groups still need to be carefully monitored if credit is expanded.

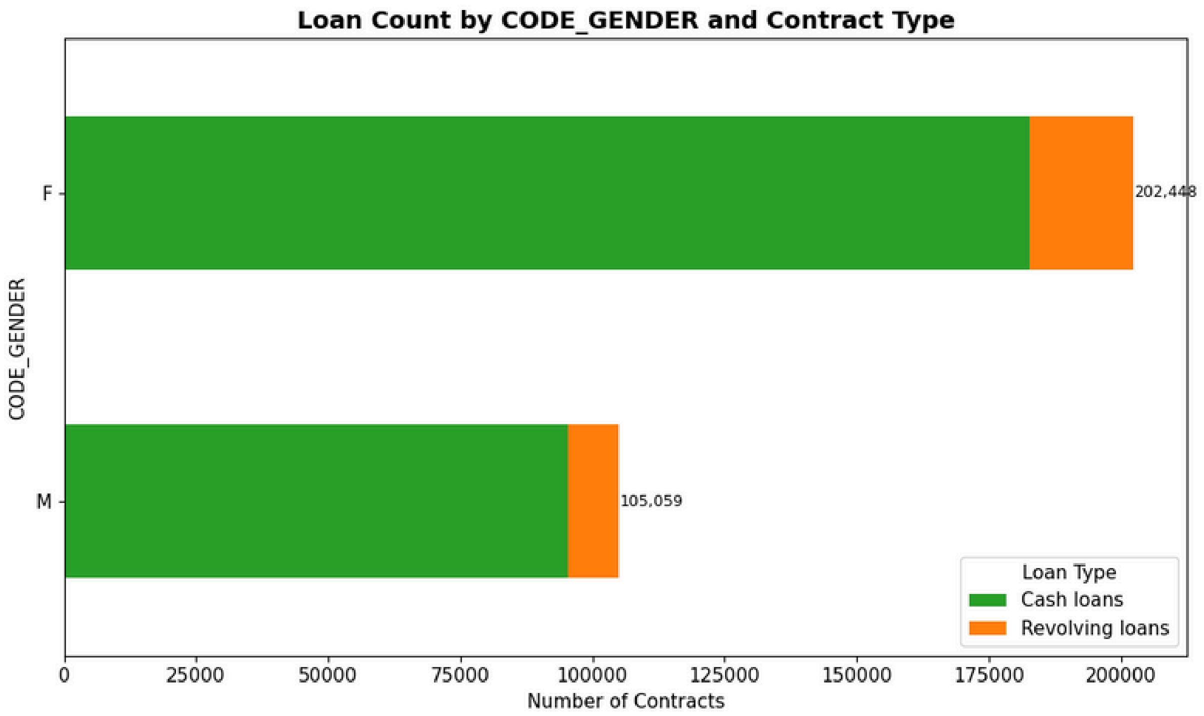


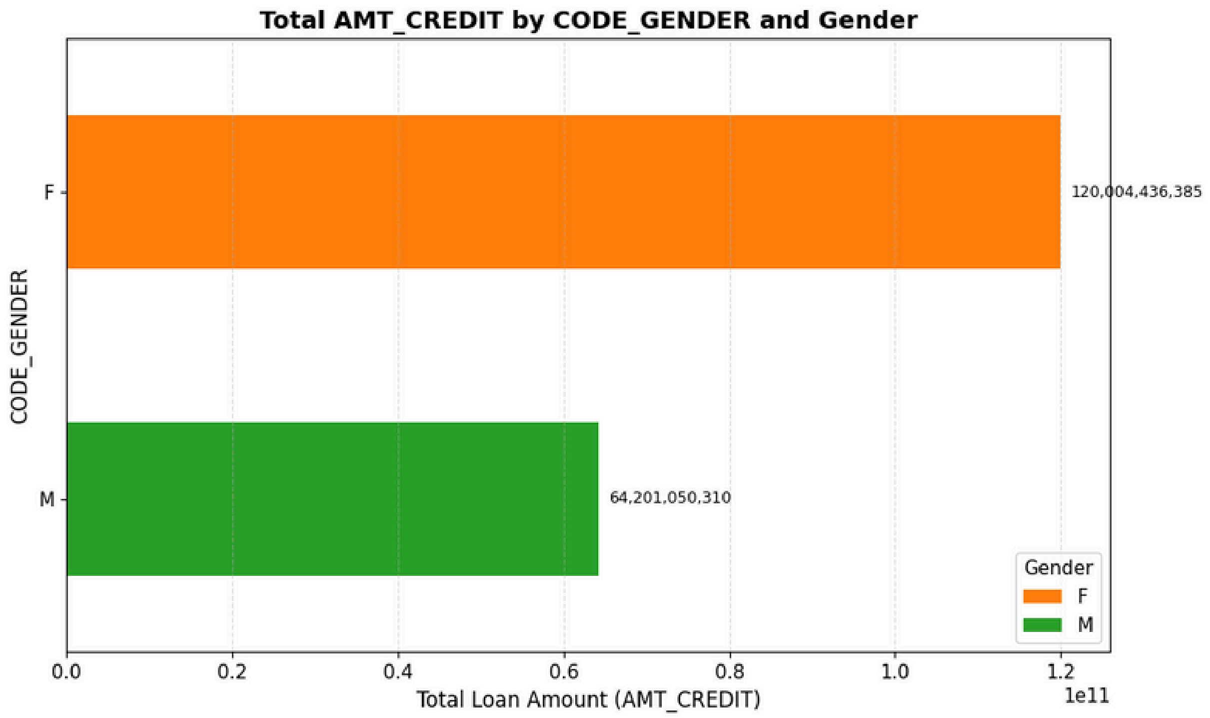
Comprehensive Analysis

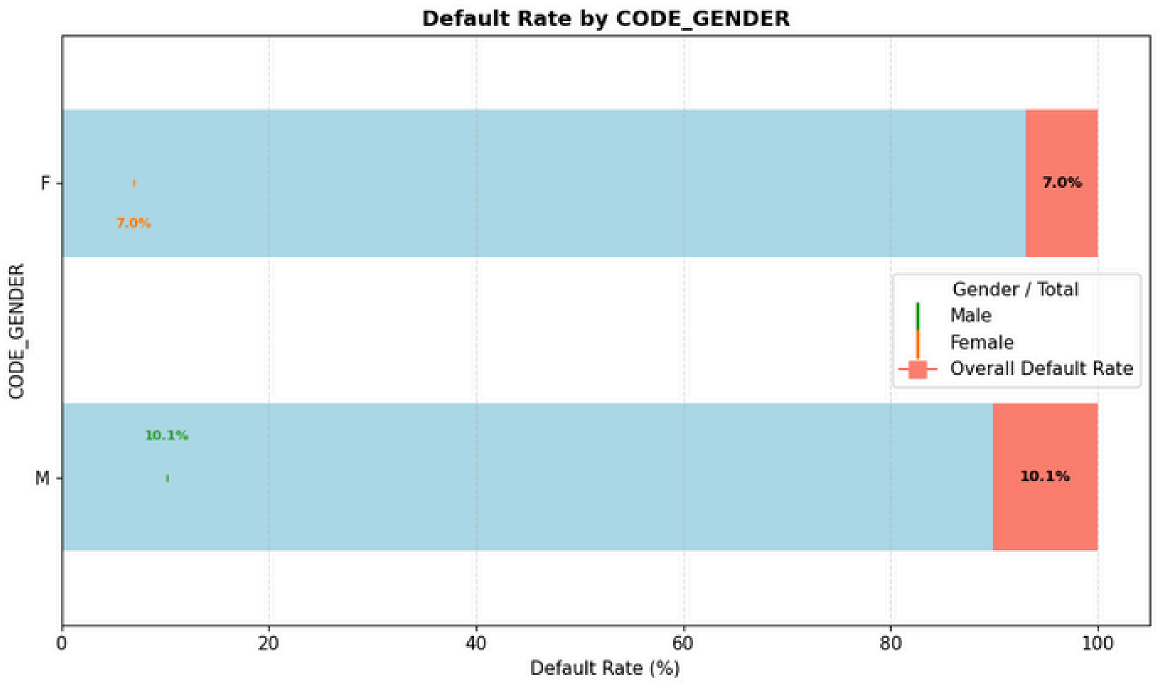
In this analysis, we conduct a bivariate analysis to identify differences in borrowing behavior and activity between male and female customers, along with deeper insights into these patterns.



Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY GENDER

Observations:

Female borrowers have twice as many contracts as male borrowers.

- Female: ~202,000 contracts
- Male: ~105,000 contracts
- Both genders primarily use Cash loans, but Revolving loans are more common among women.

📊 2. TOTAL LOAN VALUE BY GENDER

Observations:

The total loan value for women is approximately 120 billion, nearly double that of men, which is around 64 billion.

📊 3. DEFAULT RATE BY GENDER

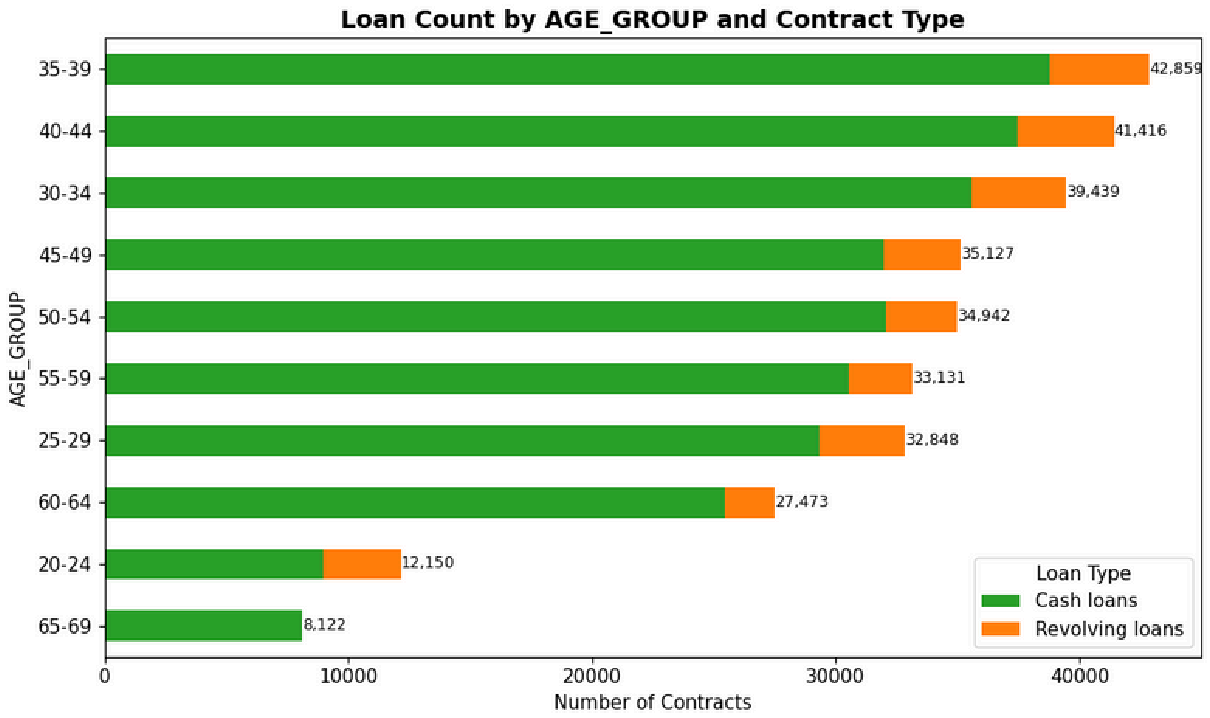
Observations:

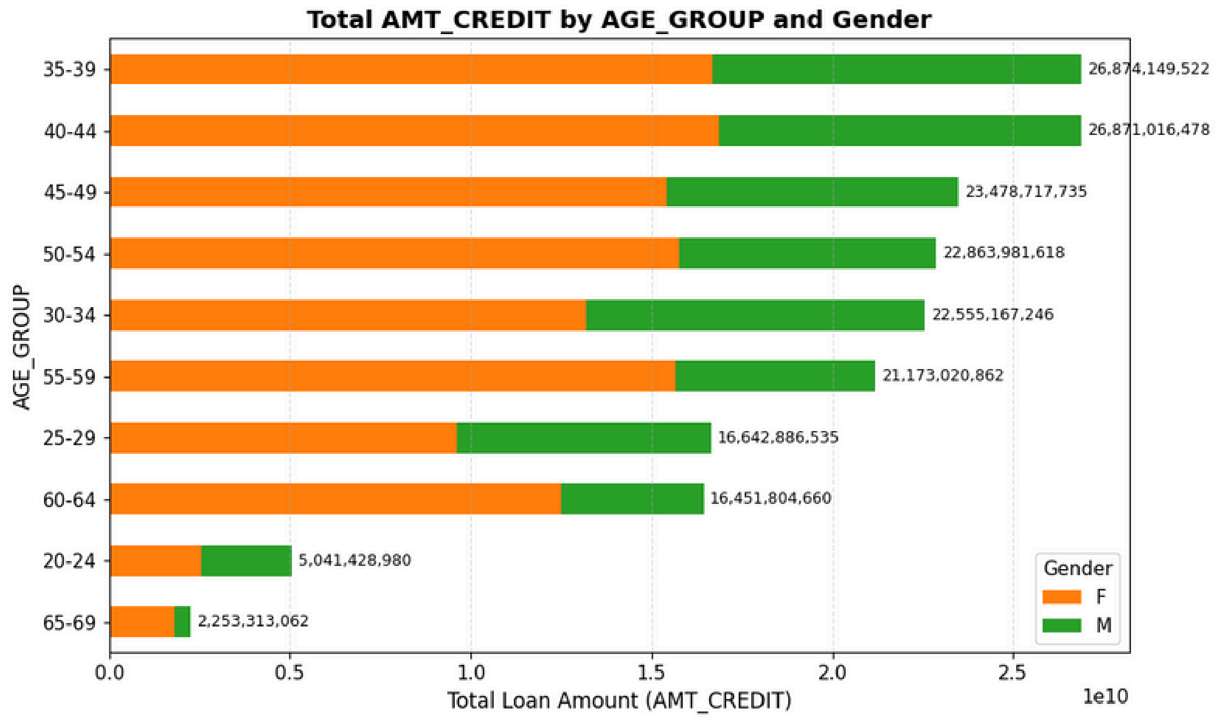
- Gender Default Rate Compared to Average
  - Female (F): 7.0% (Lower than overall)
  - Male (M): 10.1% (Higher than overall)

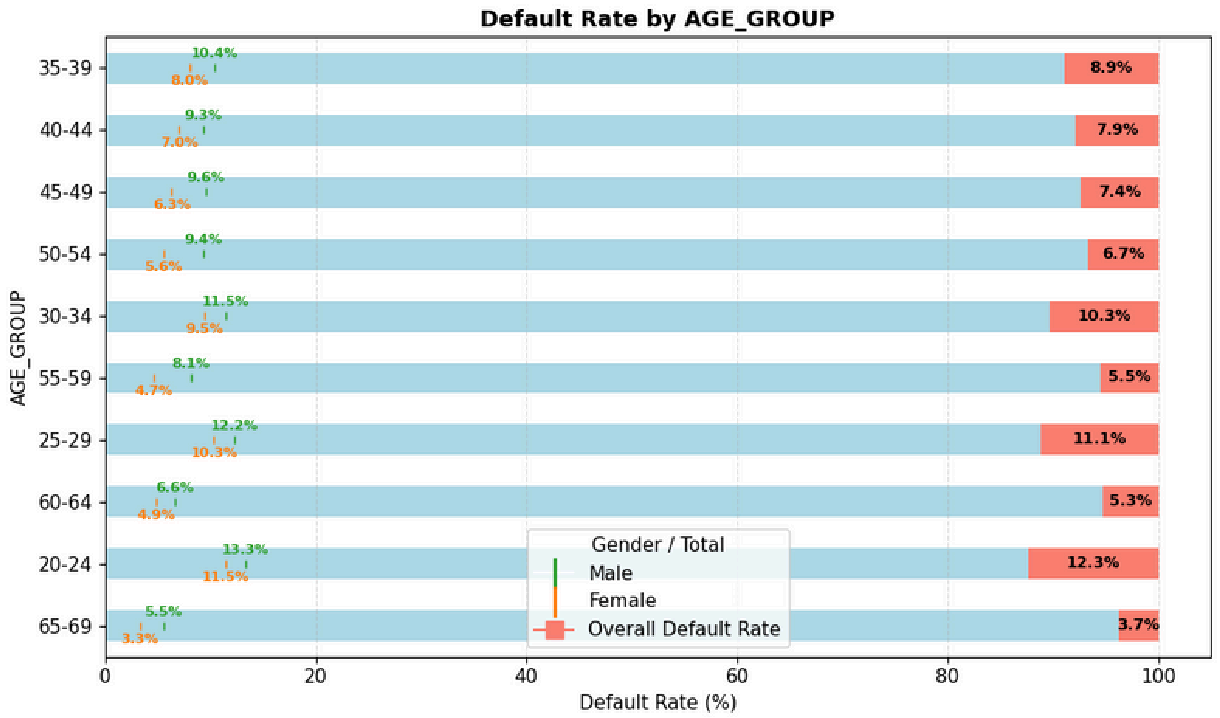
Insight:

- Women dominate in both number of loans and total loan value.
- Female borrowing behavior shows higher use of Revolving loans, indicating more flexible spending habits.
- Men have a significantly higher default rate (10.1%) compared to women (7.0%).
- Strategically, women represent a more stable and promising segment for expanding credit.

Comprehensive Analysis







📊 1. Number of Contracts by Age Group and Loan Type

Observations:

- The highest number of loans comes from the 35–39 and 40–44 age groups, with approximately 42,000 contracts each.
- Next are the 30–34 and 45–54 age groups.
- Loan activity drops sharply in groups under 25 and over 60.
- Revolving loans have a higher share among the 20–29 age group, reflecting a trend toward more flexible spending.

📊 2. Total AMT_CREDIT by Age Group and Gender

Observations:

- The highest total loan values are in the 35–44 age range, with each group contributing around 26–27 billion.
- The 30–54 age range accounts for the majority of the portfolio's total outstanding debt.
- Women borrow more than men across most age groups.

📊 3. Default Rate by Age Group

Observations:

Age Default Rate (%)

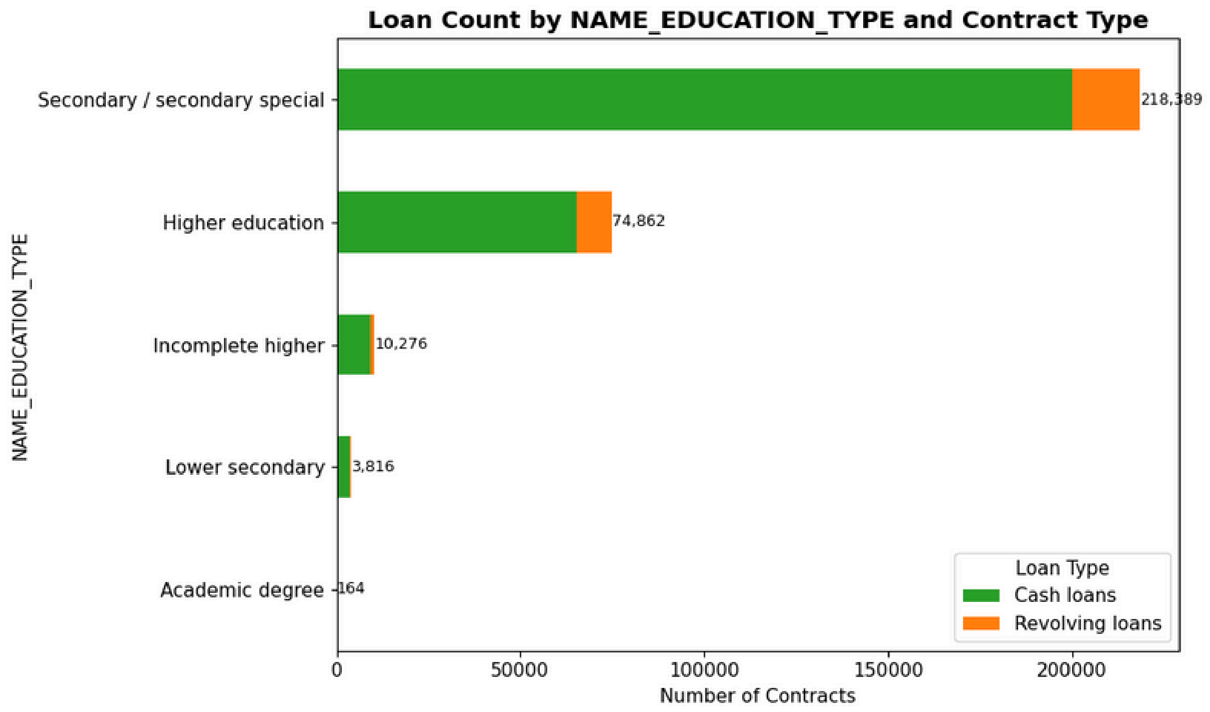
- 20–24: ~12.3% ⚠️ Highest in the portfolio
- 25–29: ~11.1% ⚠️ Very high
- 30–34: ~10.3% ⚠️ Above average
- 35–54: ~6.7–8.9% – Stable
- 55–69: ~3.7–5.5% ✅ Lowest risk

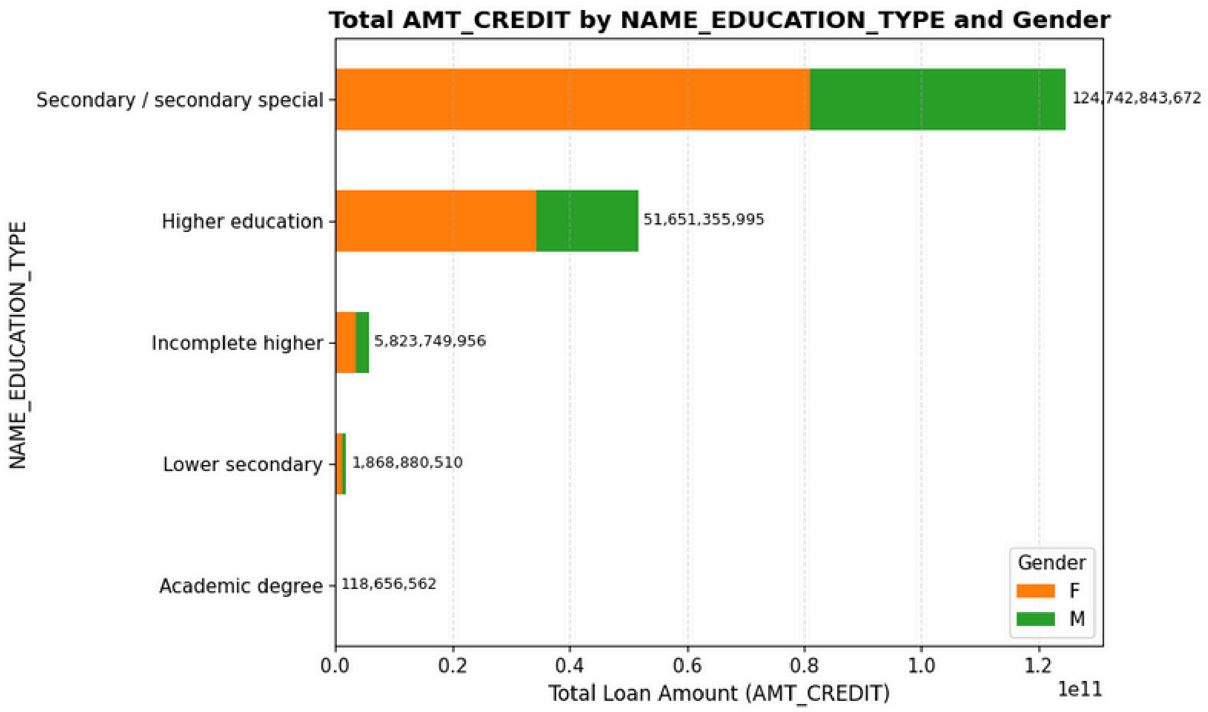
➡️ Across all age groups, women consistently have lower default rates than men.

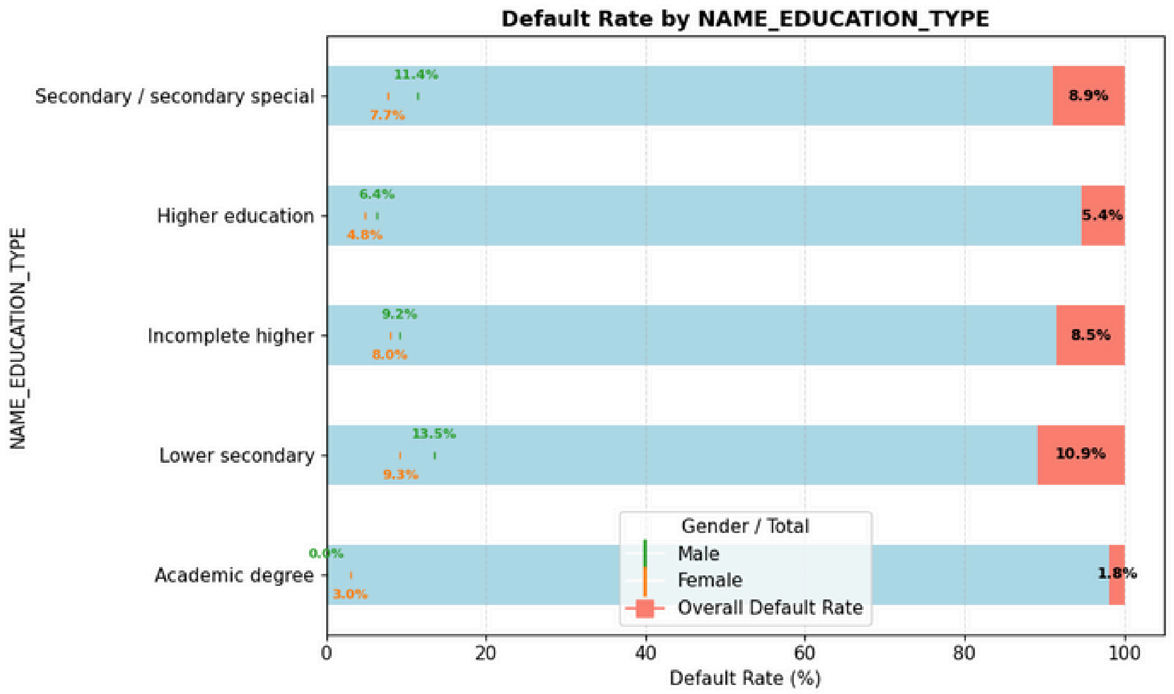
Insight:

- Ages 20–34: High risk (default >10%) ⚠️ Avoid large loans / require strict income verification
- Ages 35–54: Largest share of total loans, moderate risk ✅ Target segment for credit expansion
- Ages 55–69: Low risk but fewer loans ✅ Consider developing tailored products for this demographic
- Women have 2–3% lower default rates across all ages ✅ Prioritize in scoring and consider easing approval criteria

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY EDUCATION LEVEL

<u>Observations:</u>

- The majority of loan contracts come from the following groups:
  - Secondary / Secondary Special: ~218,000 contracts
  - Higher Education: ~74,800 contracts
- All other groups (Incomplete, Lower, Academic) account for only a small proportion.

📊 2. TOTAL LOAN VALUE (AMT_CREDIT) BY EDUCATION LEVEL AND GENDER

<u>Observations</u>:

- Group Total AMT_CREDIT Share
  - Secondary: 124.7 billion ⭐ Highest
  - Higher Education: 51.6 billion– Moderate
  - Incomplete: 5.8 billion – Low
  - Lower / Academic: Less than 2billion – Very low

➡ In both contract volume and total value, the secondary education group overwhelmingly dominates the loan portfolio.

📊 3. DEFAULT RATE BY EDUCATION LEVEL

<u>Observations</u>:

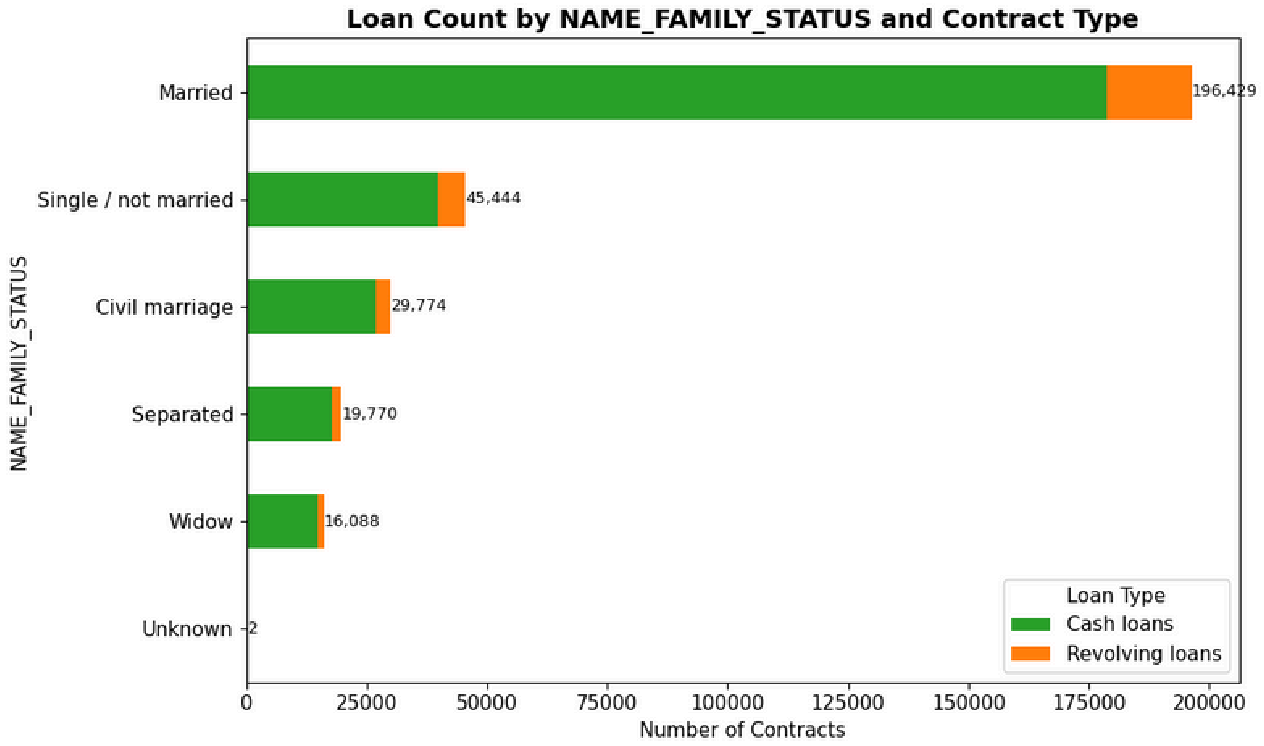
- Lower Secondary: 10.9% ⚠️ Highest
- Secondary / Secondary Special: 8.9% – High
- Incomplete Higher: 8.5% – Moderate
- Higher Education: 5.4% – Low
- Academic Degree: 1.8% ✅ Very safe

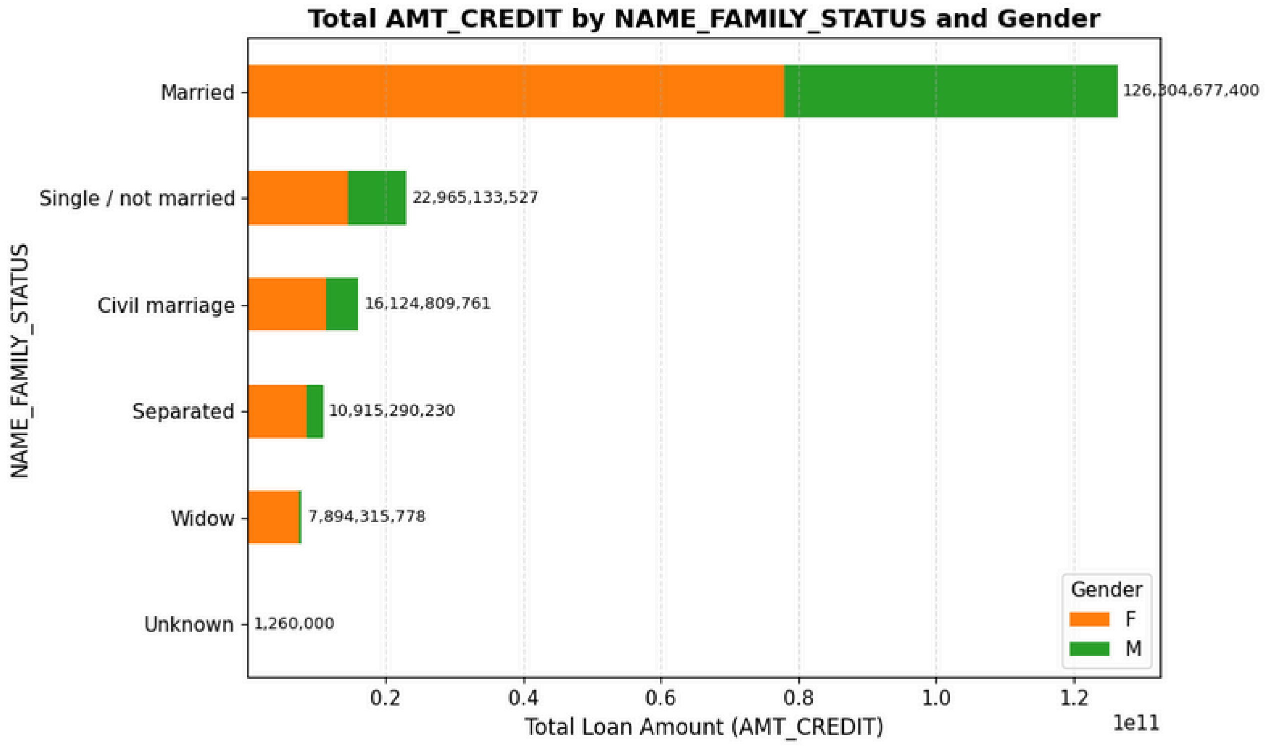
➡ Across all levels, women consistently have lower default rates than men, especially in the Lower Secondary and Secondary groups.

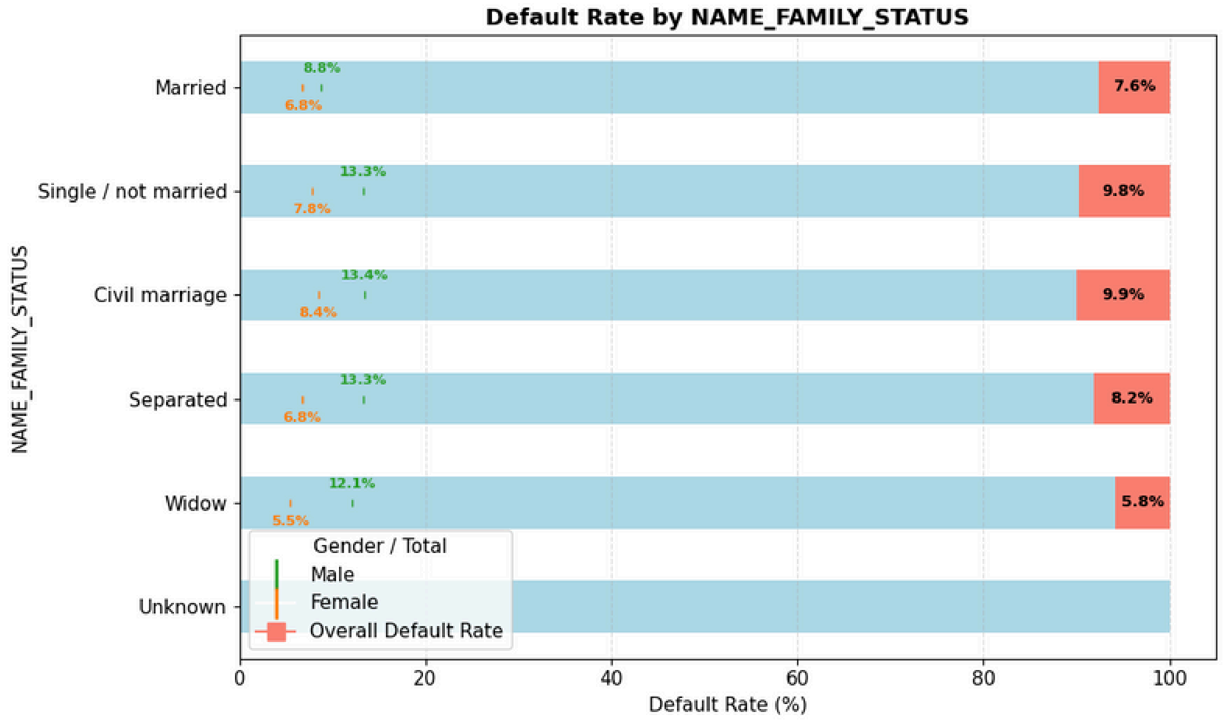
<u>Insight:</u>

- Lower education levels come with higher default risk — loan limits should be restricted, and approval conditions tightened.
- The Higher Education group has low risk and makes up a large share of quality borrowers ✅ Should be prioritized for fast-track approval.
- Academic Degree holders are few but extremely safe ✅ Consider expanding tailored, preferential products for this segment.
- Gender × Education: Women have lower default rates than men across all education levels ✅ Suggest increasing scoring for female borrowers with education levels of Secondary or above.

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY MARITAL STATUS

<u>Observations:</u>

The Married group leads with nearly 200,000 contracts — about 4 times more than the second-largest group.

Other groups in order: Single, Civil marriage, Separated, Widow.

➡ Conclusion: Married individuals make up the clear majority of the loan portfolio.

📊 2. TOTAL LOAN VALUE

<u>Observations:</u>

- Marital Status Total Loan Remarks
  - Married: 126.8 billion – Accounts for ~60–70% of the portfolio
  - Single: 22.9 billion – Significantly lower
  - Civil marriage: 16.1 billion – Moderate
  - Separated: 10.9 billion – Low
  - Widow: 7.9 billion – Low

➡ The total loan value for married individuals is nearly 6 times greater than for singles.

📊 3. DEFAULT RATE BY MARITAL STATUS

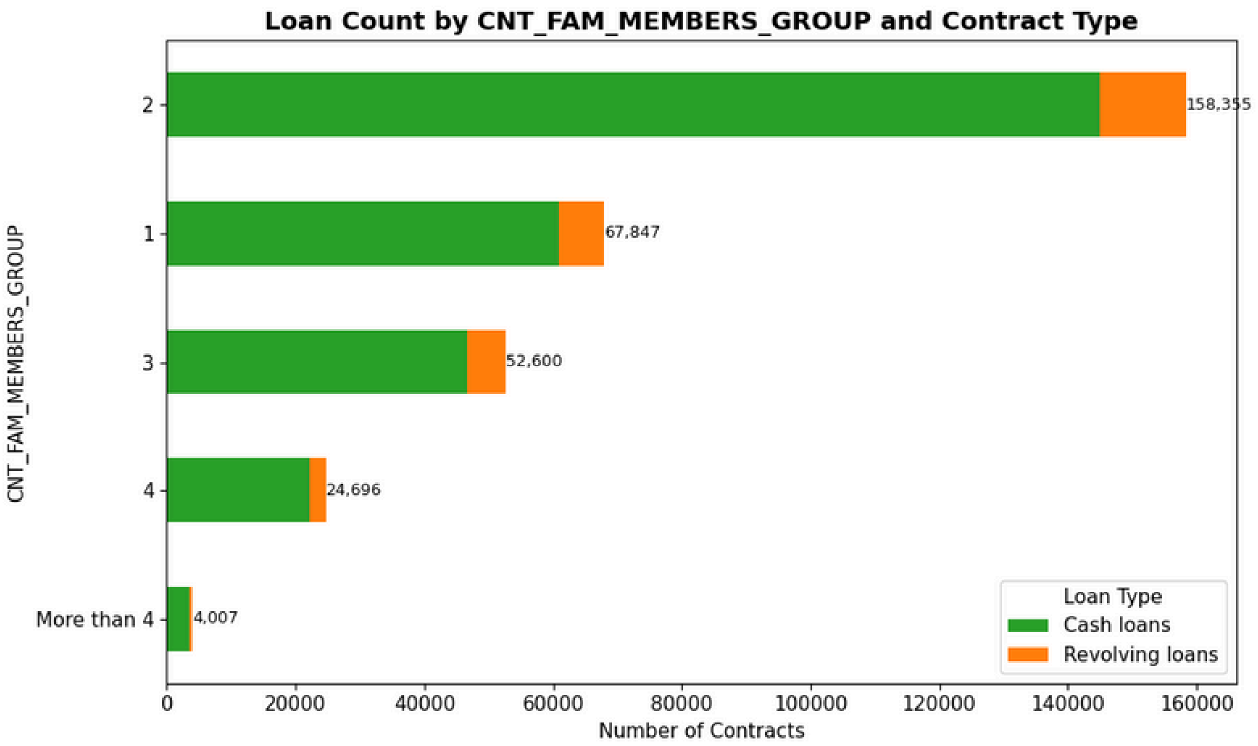
<u>Observations:</u>

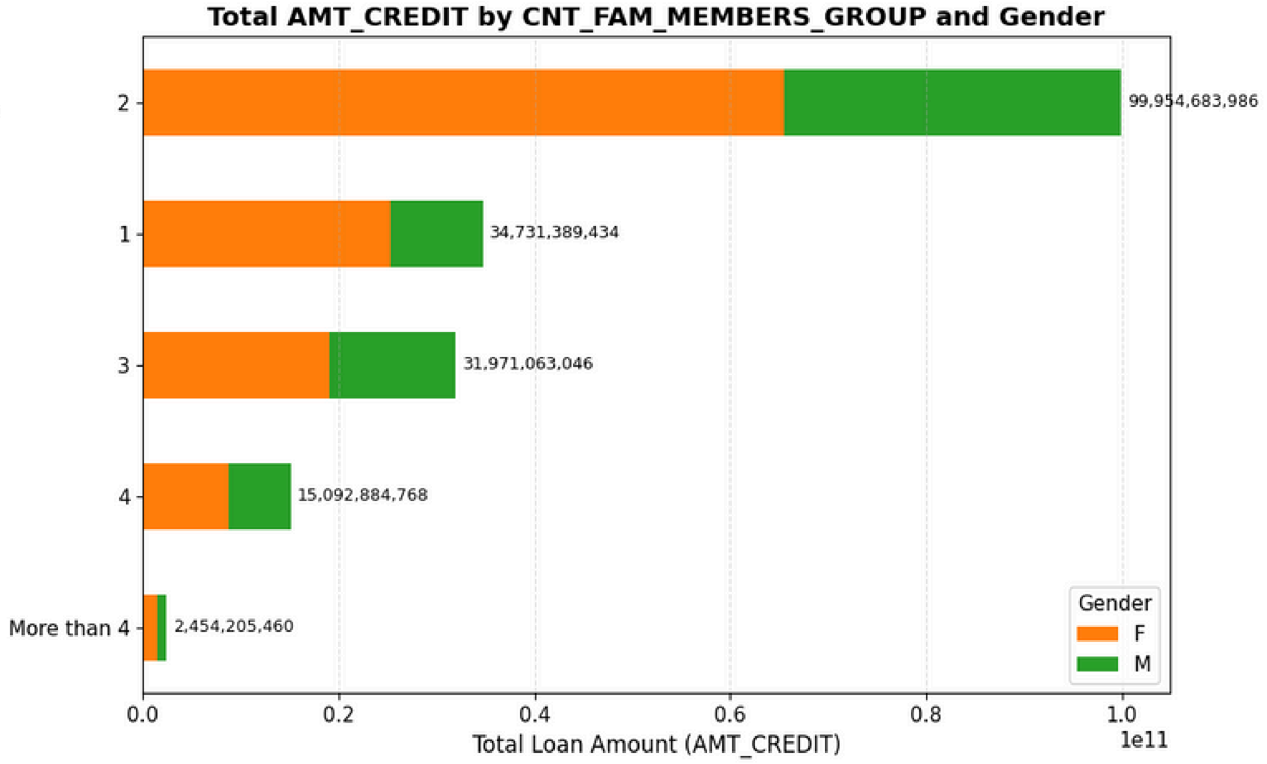
- Married: 7.6% ✅ Lowest among major groups
- Single / Civil marriage / Separated: ~9.8–9.9% ⚠️ Highest
- Widow: 5.8% ✅ Very low, but small in number
- Gender: Men > Women in all groups ✔ Consistent: women have lower default rates

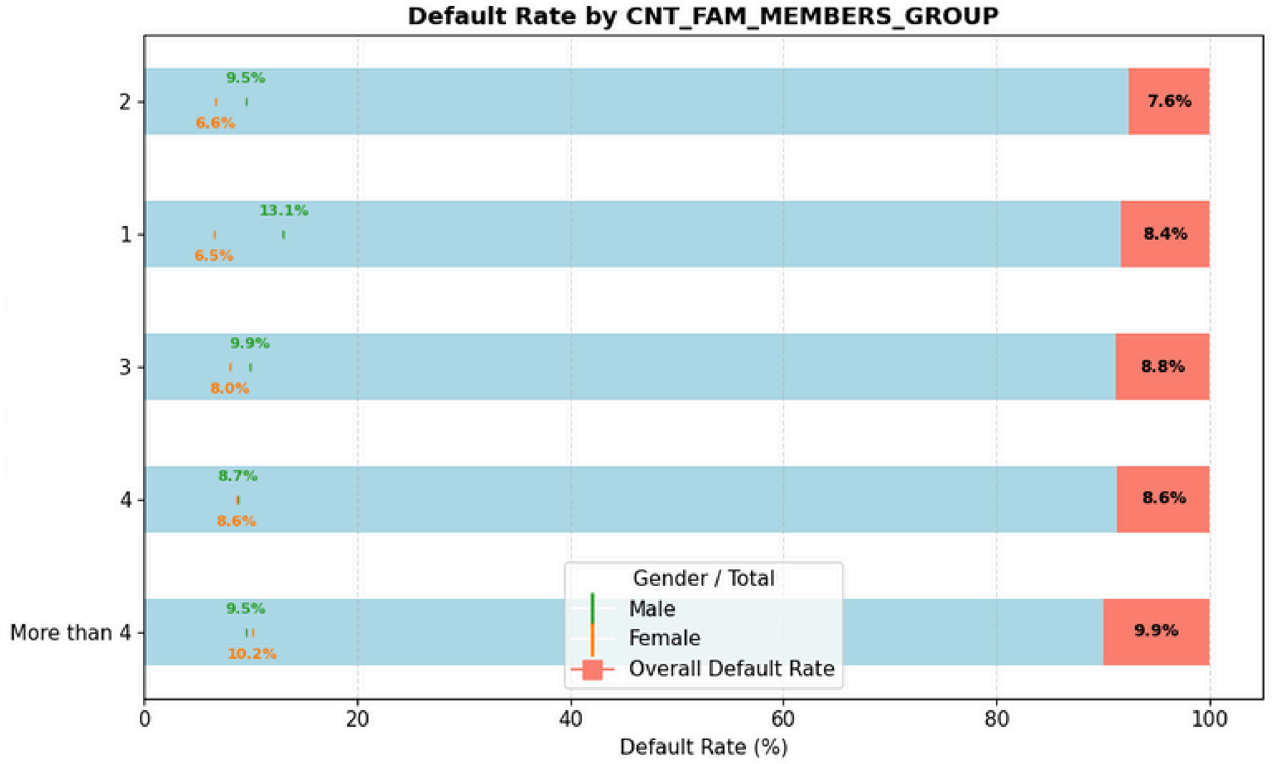
<u>Insight:</u>

- Married individuals borrow more and have lower default rates ✅ Prioritize for credit expansion
- Single / cohabiting / separated individuals have high default rates (~10%) ⚠️ Flag as higher risk, require strict income and credit history checks
- Widows show low risk but have few contracts ✅ Consider developing specialized products (e.g. welfare, medical loans) for this group

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY HOUSEHOLD SIZE

Observations:

- 2-person households have the highest number of loans (~158,000 contracts), far ahead of other groups.
- 1-person households (~68,000 contracts) rank second.
- The number of contracts gradually decreases from 3 → 4 → more than 4 people.

➡ Small households (1–2 people) make up the majority of the loan portfolio.

📊 2. TOTAL LOAN VALUE

Observations:

- 2 people: 99.9 billion ⭐ Dominant
- 1 person: 34.7 billion – Second
- 3 people: 32.0 billion – Close to 1-person group
- 4 people: 15.1 billion – Low
- More than 4 people: 2.5 billion – Very low

➡ Households with 1–3 members are the core customer segment in terms of both loan volume and total value.

📊 3. DEFAULT RATE BY HOUSEHOLD SIZE

Observations:

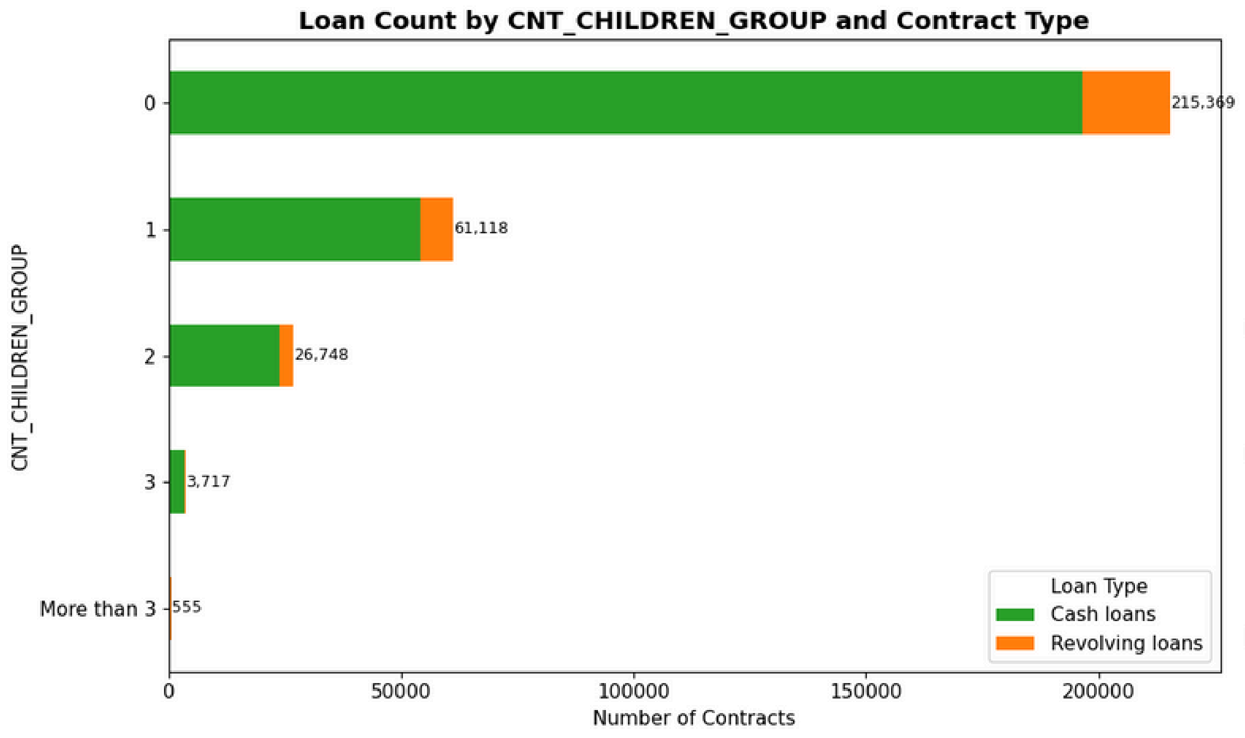
- 1 person: 8.4% (male: 13.1%, female: 6.5%) ⚠️ Single men have the highest risk
- 2 people: 7.6% ✅ Best performance
- 3–4 people: ~8.6–8.8% – Average
- More than 4 people: 9.9% ⚠️ High risk, needs control

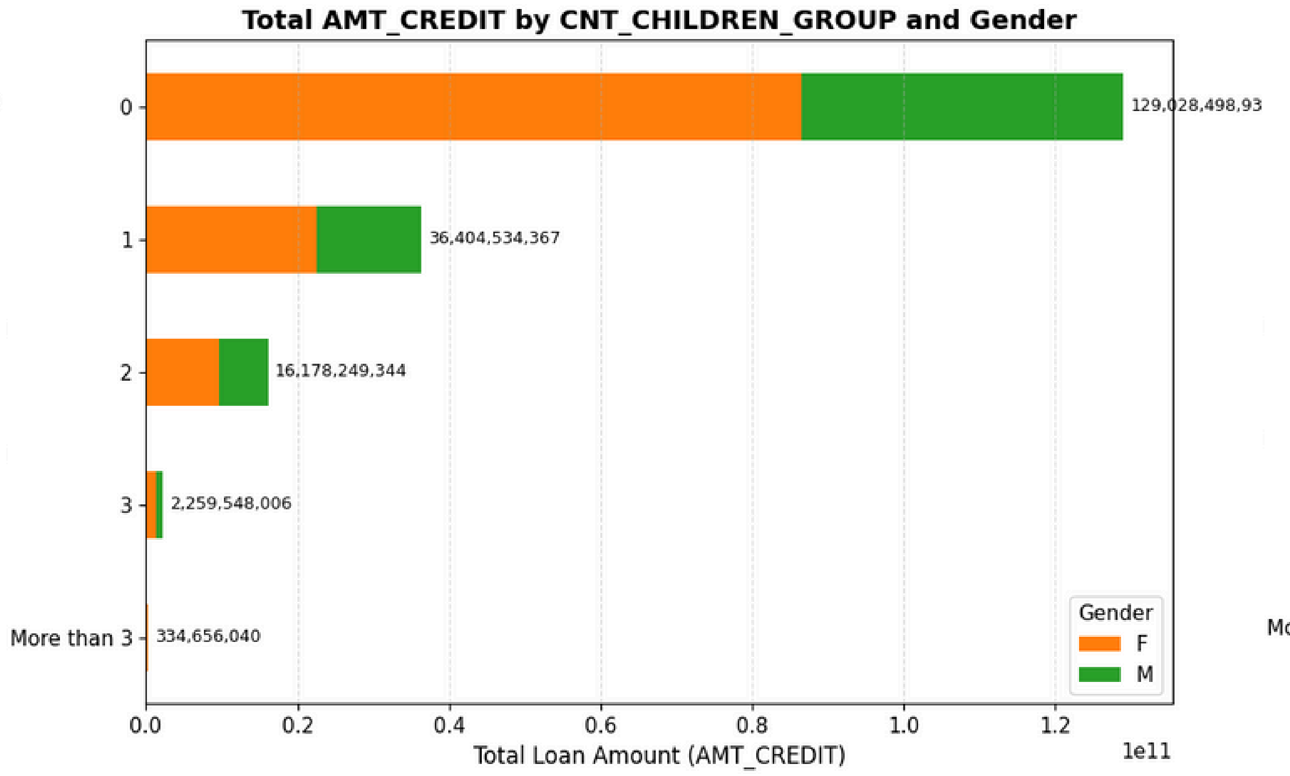
➡ 2-person households are the most stable, while single-person (especially male) and large households carry higher risk.

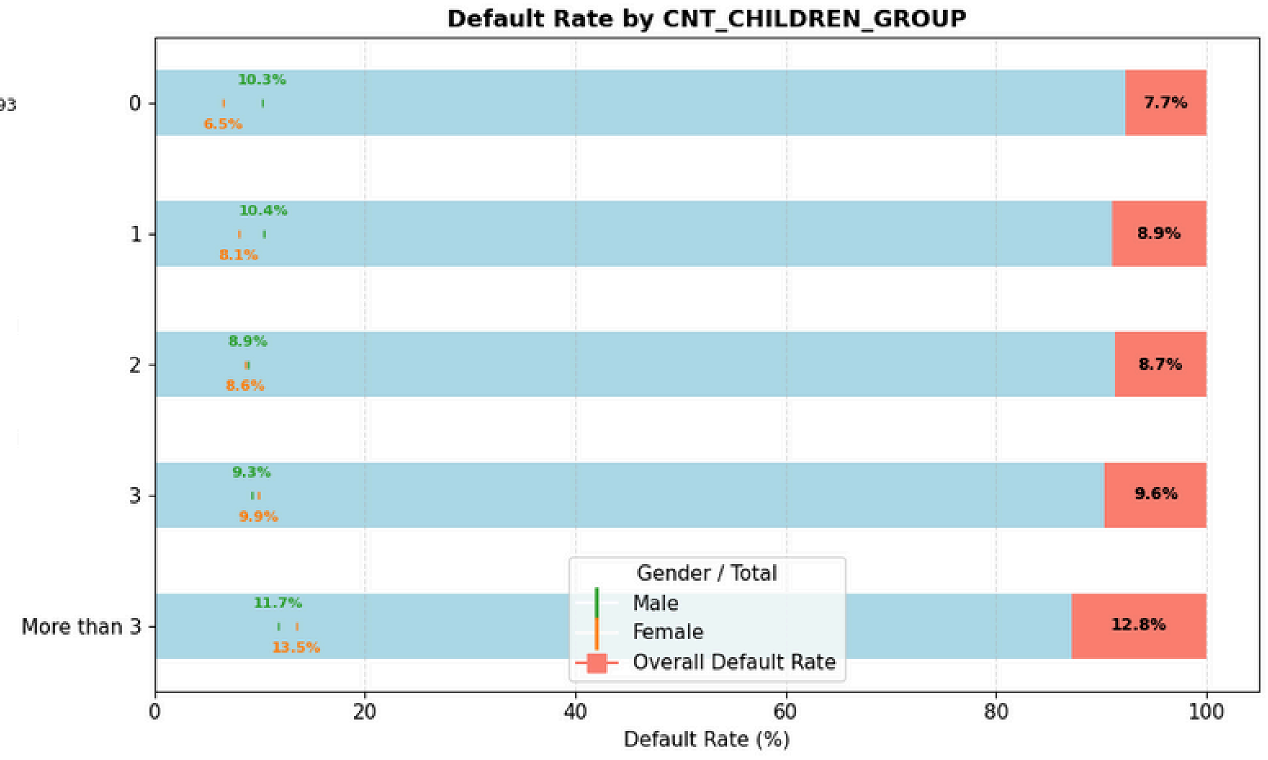
Insight:

- Small families (1–2 people): Core customer group. 2-person households can be prioritized for faster approval and better service.
- Single men (1-member): Highest risk group (13.1%). Requires careful scoring and income verification.
- Large families (>4): High risk and low borrowing volume. Review loan approval processes and strengthen risk management.
- Women consistently have lower default rates than men in every household size group.

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY NUMBER OF CHILDREN

Observations:

- The childless group (0 children) dominates with over 215,000 contracts.
- 1 child: ~61,000 contracts
- All other groups (2–3–>3 children) show a clear downward trend.

➡ Customers without children are the core borrowing group for the lending institution.

📊 2. TOTAL LOAN VALUE

Observations:

- 0 children: 129 billion ⭐ Largest share
- 1 child: 36.4 billion – Second
- 2 children: 16.2 billion – Sharp drop
- 3 – >3 children: Less than 2.5 thousand billion – Very small

➡ Loan value decreases as the number of children increases, which may reflect:
increased financial pressure and reduced repayment capacity, lowering consumer loan demand; or
a strategic choice to avoid targeting large families in order to limit potential risk.

📊 3. DEFAULT RATE BY NUMBER OF CHILDREN

Observations:

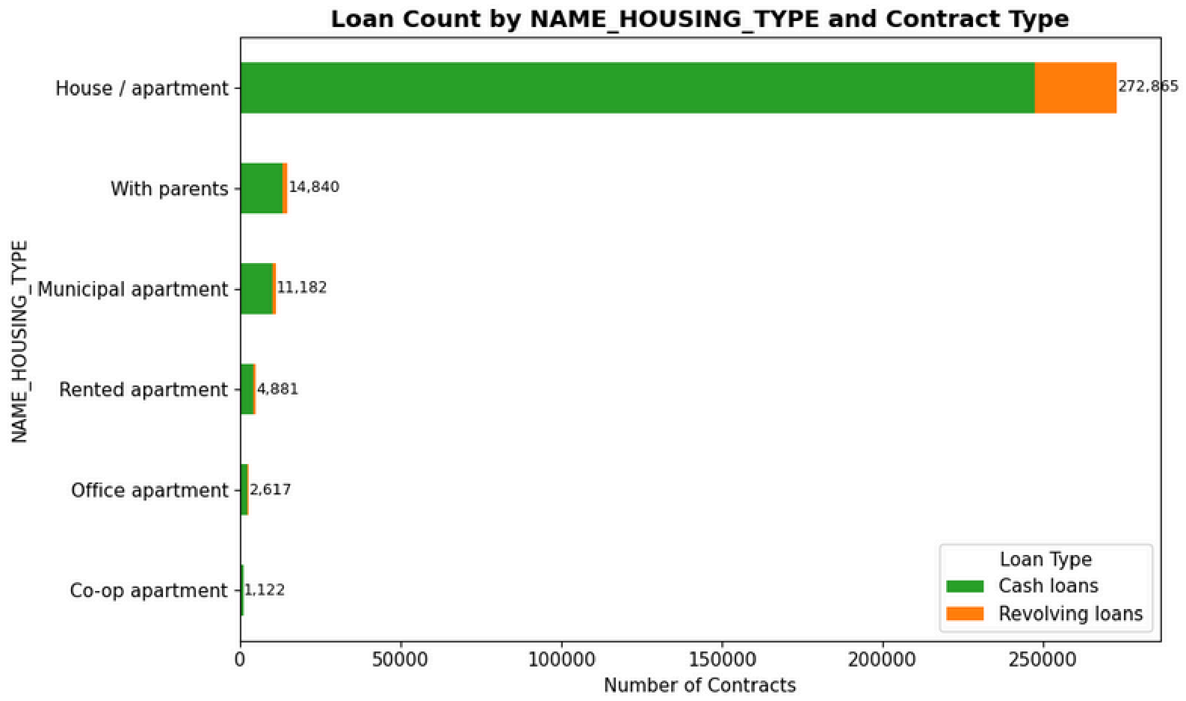
- 0 children: 7.7% ✅ Lowest
- 1–2 children: ~8.7–8.9% – Average
- 3 children: 9.6% – High
- More than 3 children: 12.8% ⚠️ Very high (highest overall)

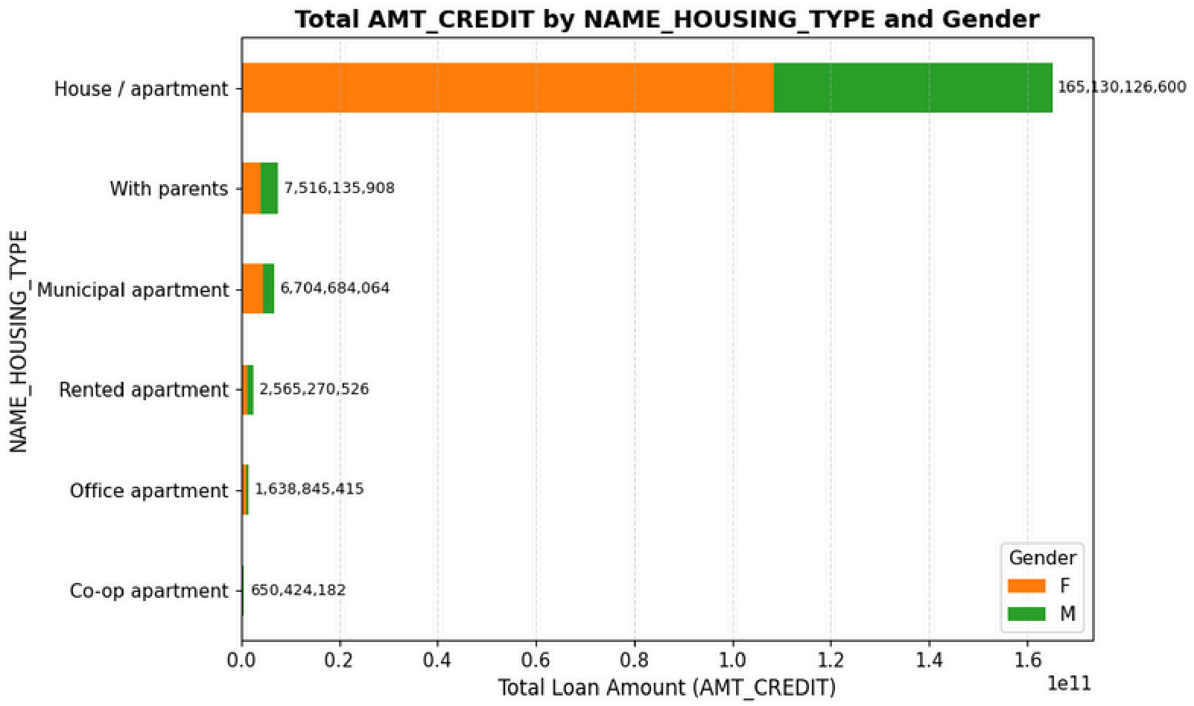
➡ Households with 3 or more children are a high-risk segment, especially among women (>13.5%).

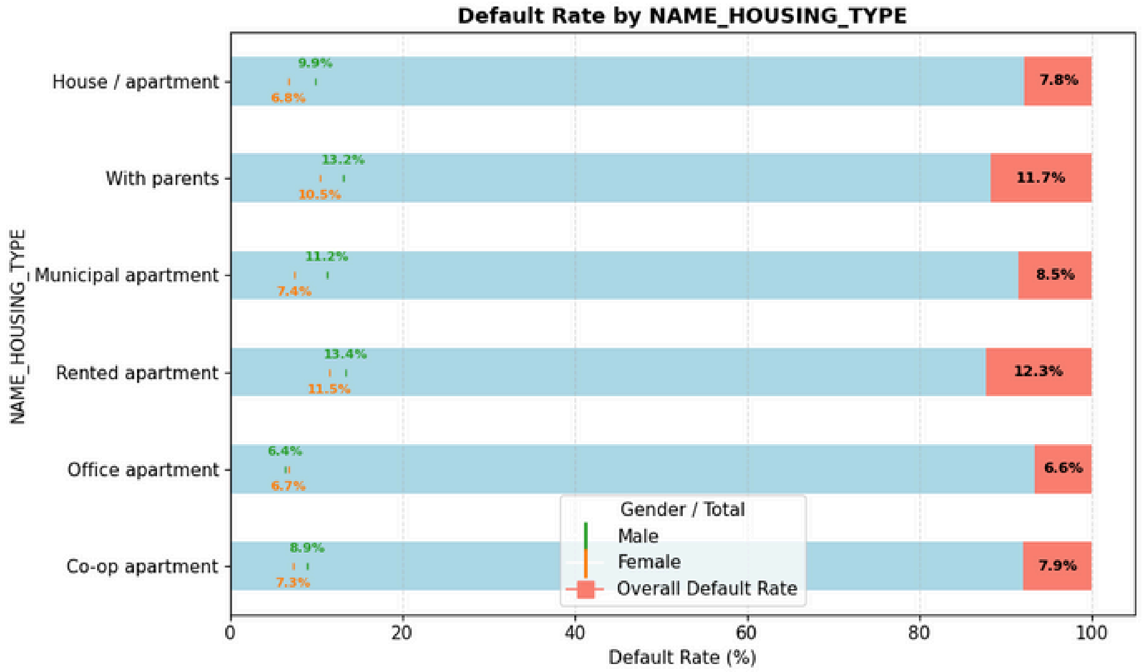
Insight:

- Childless group: Largest segment with the lowest default rate ✅ Ideal for credit expansion
- Larger families: Default rate increases with the number of children ⚠️ Extra caution needed for those with more than 2, especially more than 3
- Women with multiple children: Sharp increase in default rate (13.5%) ⚠️ Require clear income and credit history verification
- Men with 1 child: Default rate remains high (10.4%) ⚠️ Should not be assumed low-risk — needs attention

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY HOUSING TYPE

Observations:

- The House / apartment group accounts for nearly all contracts (~273,000).
- Other groups like With parents, Municipal apartment, Rented, Office, and Co-op make up only a small portion.

➡ Customers with independent housing (house/apartment) are the primary borrowers.

📊 2. TOTAL LOAN VALUE

Observations:

- House / apartment: 165.1 billion ⭐ Represents nearly the entire portfolio
- With parents: 7.5 billion – Low to moderate
- Municipal apartment: 6.7 billion – Low to moderate
- Rented / Office / Co-op: Less than 2.6 billion – Low

➡ In line with the volume, customers living in houses/apartments also account for the highest total outstanding debt.

📊 3. DEFAULT RATE BY HOUSING TYPE

Observations:

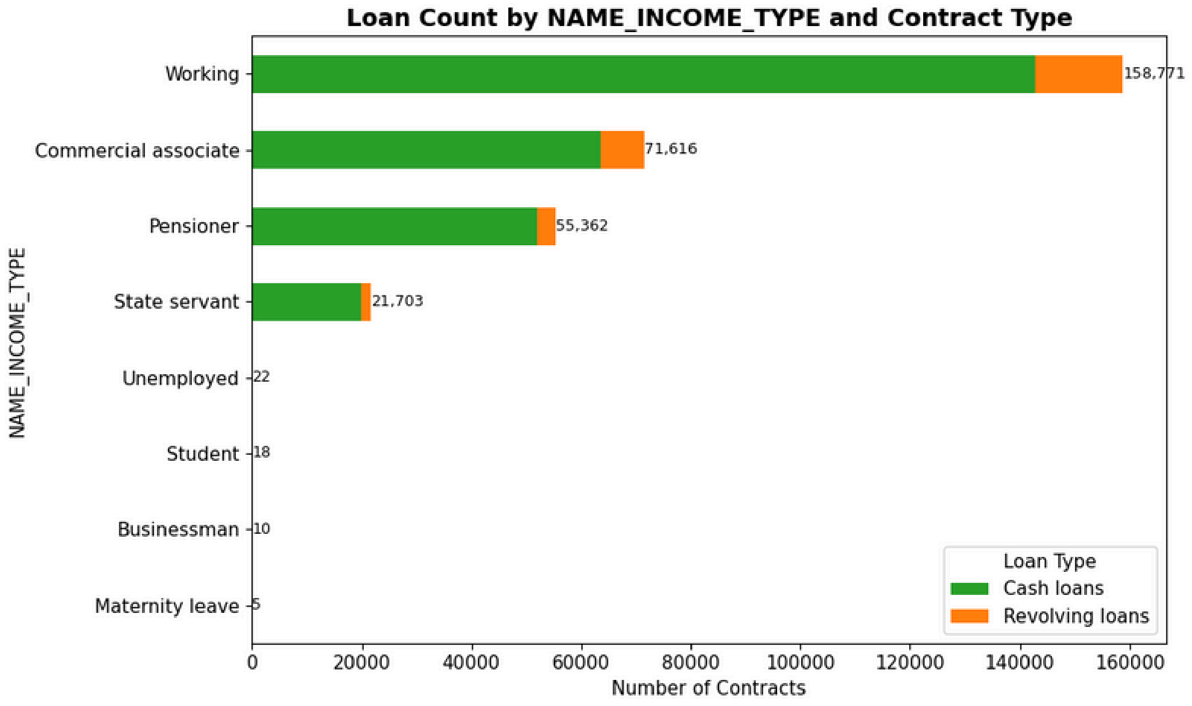
- House/apartment: 7.8% ✅ Low
- With parents: 11.7% ⚠️ High
- Rented apartment: 12.3% ⚠️ Highest
- Municipal apartment: 8.5% – Average
- Office apartment: 6.6% ✅ Lowest
- Co-op: 7.9% – Average

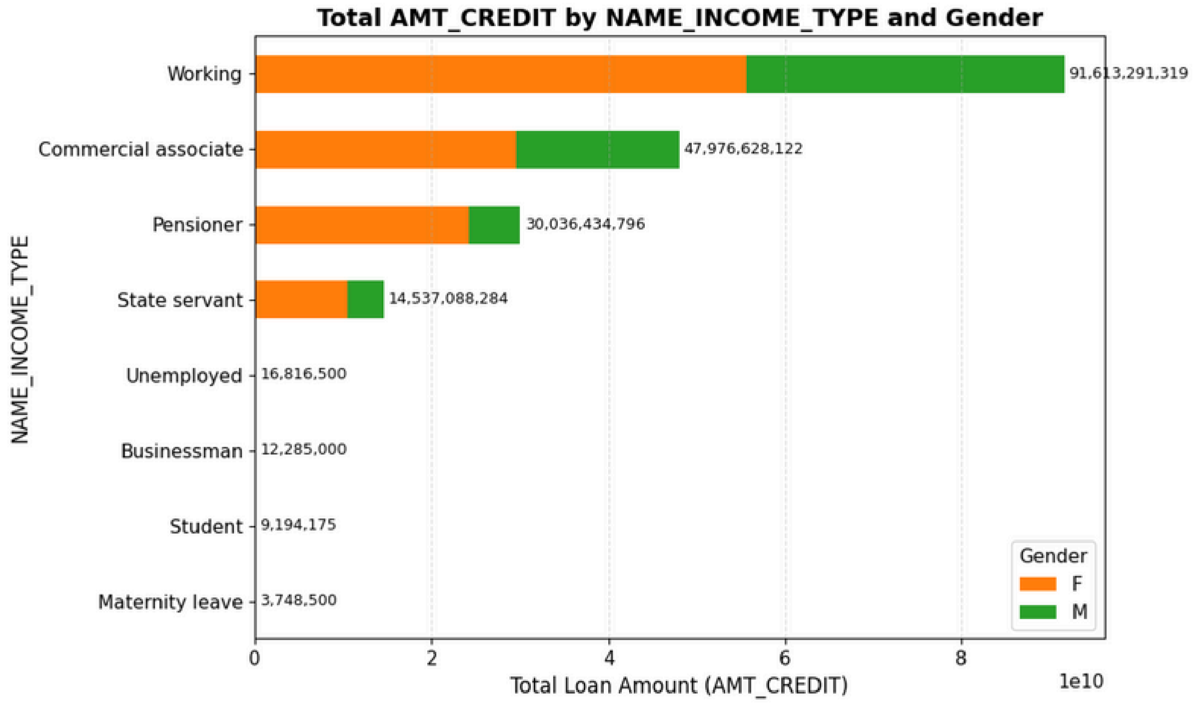
➡ Those renting or living with parents have the highest default rates, especially men (>13%).

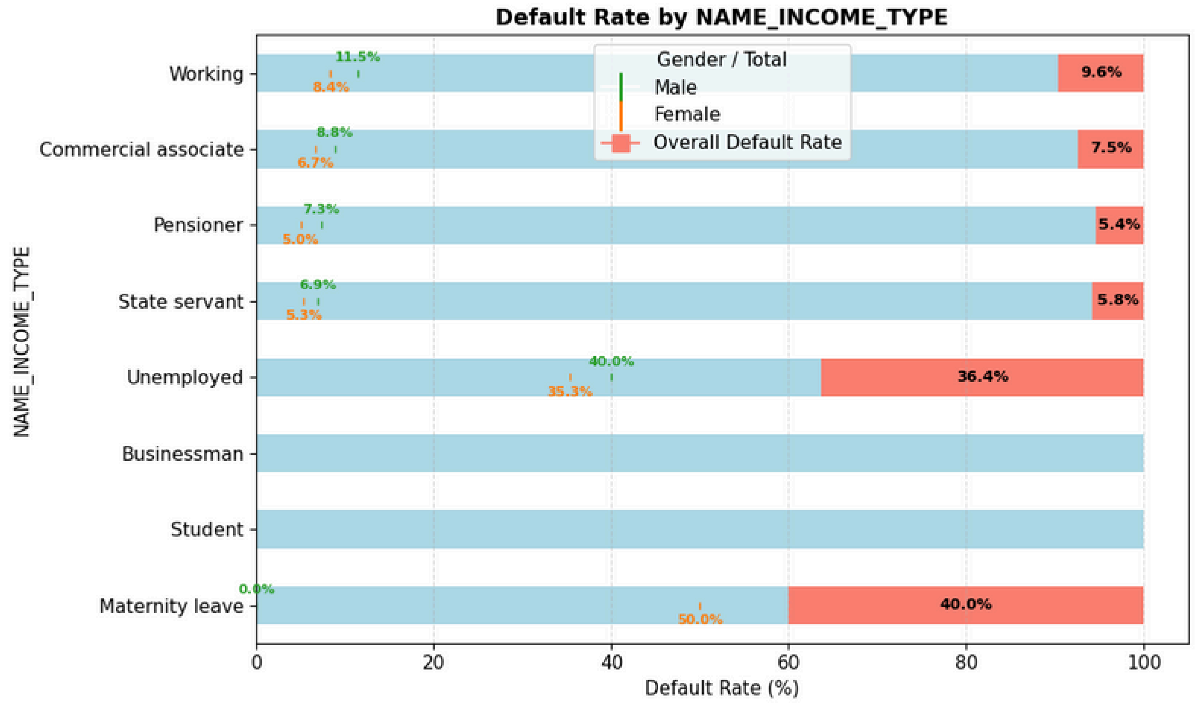
Insight:

- House/apartment: Low default rate and high volume ✅ Prioritize for credit expansion
- Rented / With parents: High default rates (12%–13%) ⚠️ Require strict credit history and income checks
- Office apartment: Low default rate despite few contracts ✅ Consider expanding for independent working professionals
- Gender: Men in rented housing or living with parents show very high default rates ⚠️ Suggest adding risk points or requiring guarantors

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY INCOME TYPE

Observations:

- Working (employed): Largest group (~158,000 contracts)
- Commercial associate (business partner): ~71,600 contracts
- Pensioner (retired): ~55,000 contracts
- Other groups are very small: State servant, Student, Businessman, Unemployed, etc.

➡ Those with stable employment income or business partnerships are the core customer base.

📊 2. TOTAL LOAN VALUE

Observations:

- Working: 91.6 billion ⭐ Largest share
- Commercial associate: 48.0 billion – Upper mid-range
- Pensioner: 30.0 billion – Moderate
- Other groups: Less than 15 billion – Low

➡ Loan size trends align with expectations: stable income groups tend to take larger loans.

.

📊 3. DEFAULT RATE BY INCOME TYPE

Observations:

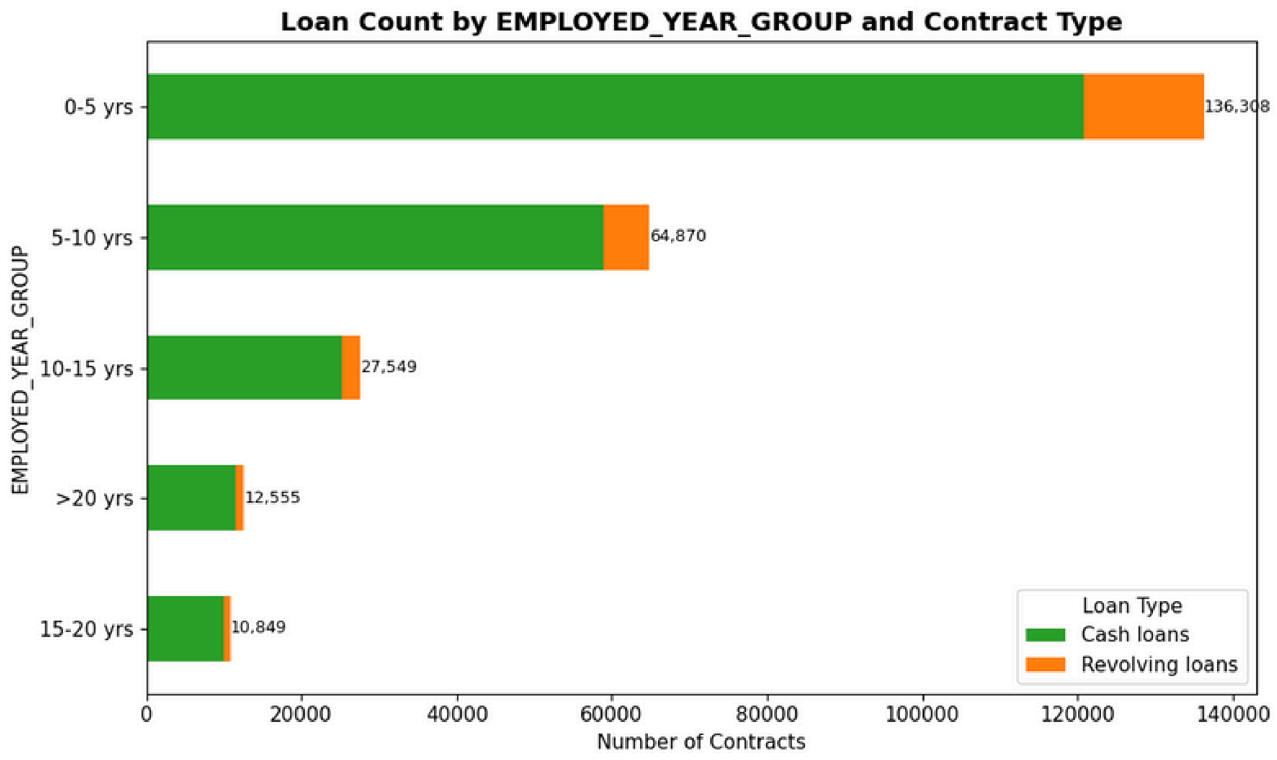
- Working: 9.6% ⚠️ Relatively high
- Commercial associate: 7.5% – Moderate
- Pensioner / State servant: ~5.4–5.8% ✅ Low and stable
- Unemployed: 36.4% ⚠️ Extremely high
- Maternity leave: 40.0% ⚠️ High risk, unstable
- Student / Businessman: Limited data, but high (~35–40%) ⚠️ Warning

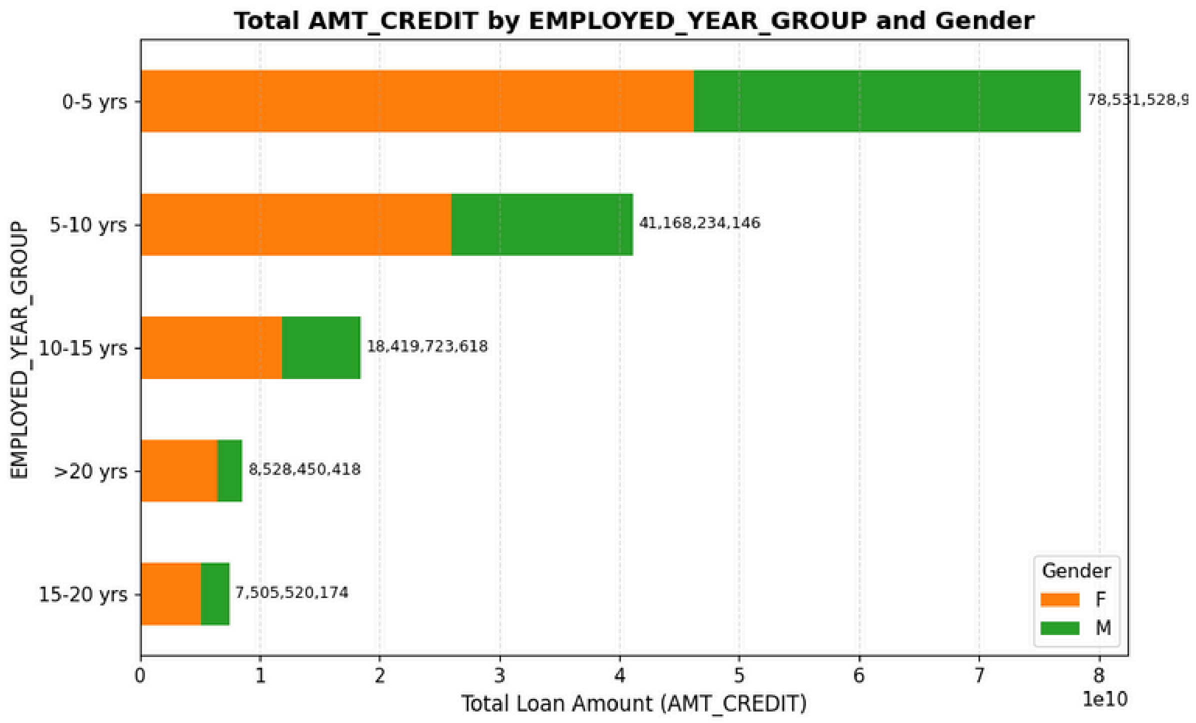
➡ Income source is a strong risk factor — special attention needed for Unemployed, Maternity leave, and Student groups.

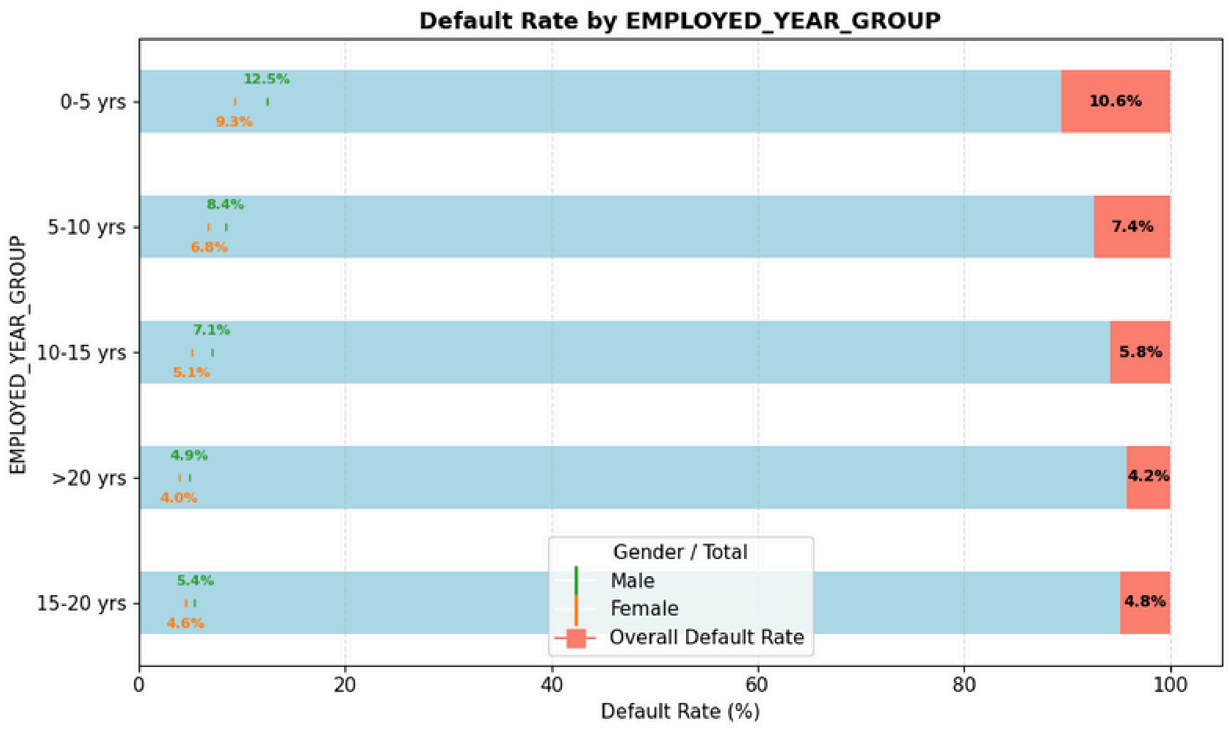
Insight:

- Working: Large volume but high default rate (men ~11.5%) ⚠️ Should be re-segmented by income level and industry
- Pensioners / State servants: Low and stable default ✅ Prioritize for fast approval and preferential interest rates
- Commercial associate: Moderate risk but income volatility is a concern ⚠️ Income sources should be thoroughly verified
- Maternity / Unemployed / Student: ~40% default rate ❌ Credit should be restricted; if granted, co-signers or guarantors should be required

Comprehensive Analysis







📊 1. NUMBER OF LOAN CONTRACTS BY YEARS OF EMPLOYMENT

Observations:

- 0–5 years: ~136K 🔥 Largest share
- 5–10 years: ~64K – Second highest
- 10–15 years: ~27K – Moderate
- 15–20 and >20 years: ~11–12K – Low

➡ Main focus is on newly employed and mid-career customers.

📊 2. TOTAL LOAN VALUE (AMT_CREDIT)

Observations:

- 0–5 years: 78.5 billion ⚠️ Very high, despite limited experience
- 5–10 years: 41.1 billion – High
- 10–15 years: 18.4 billion – Moderate
- >15 years: ~7–8.5 billion – Low

➡ Customers with less experience account for the largest loan amounts, reflecting a strategy focused on younger borrowers — but this also carries potential risk.

📊 3. DEFAULT RATE BY YEARS OF EMPLOYMENT

Observations:

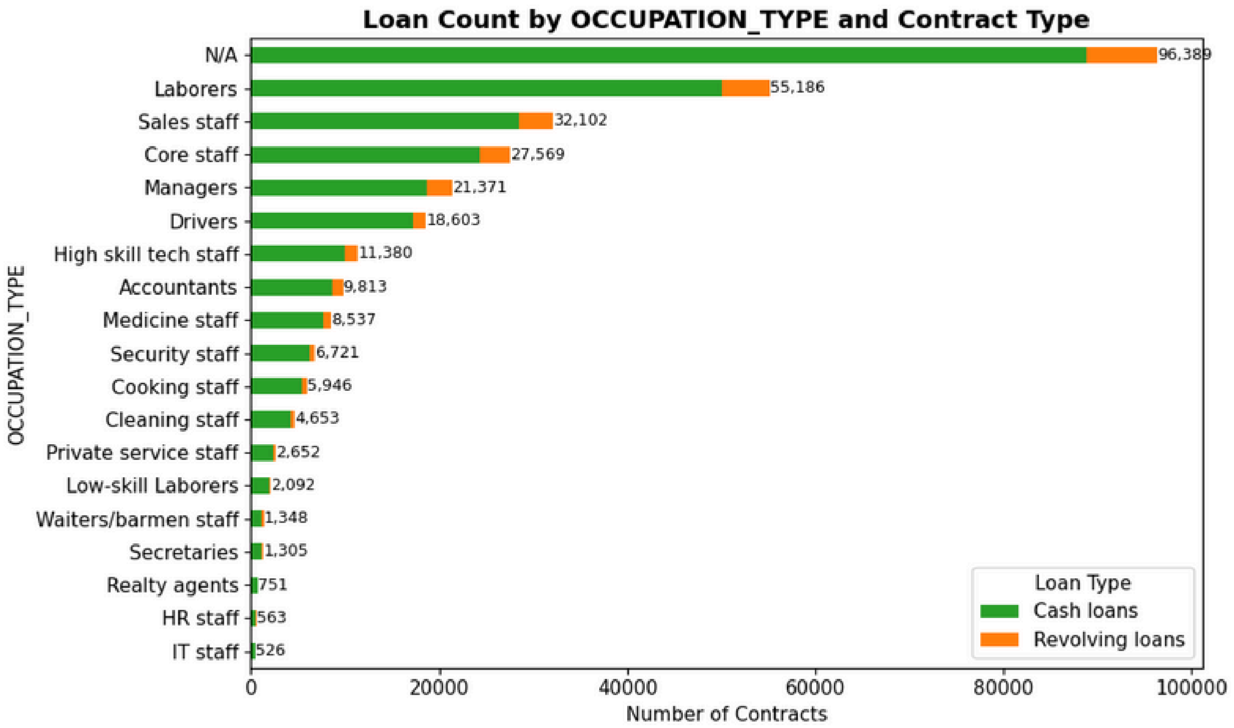
- 0–5 years: 10.6% ⚠️ Very high
- 5–10 years: 7.4% – Moderate
- 10–15 years: 5.8% – Low
- 15–20 years: 4.8% ✅ Low
- >20 years: 4.2% ✅ Lowest

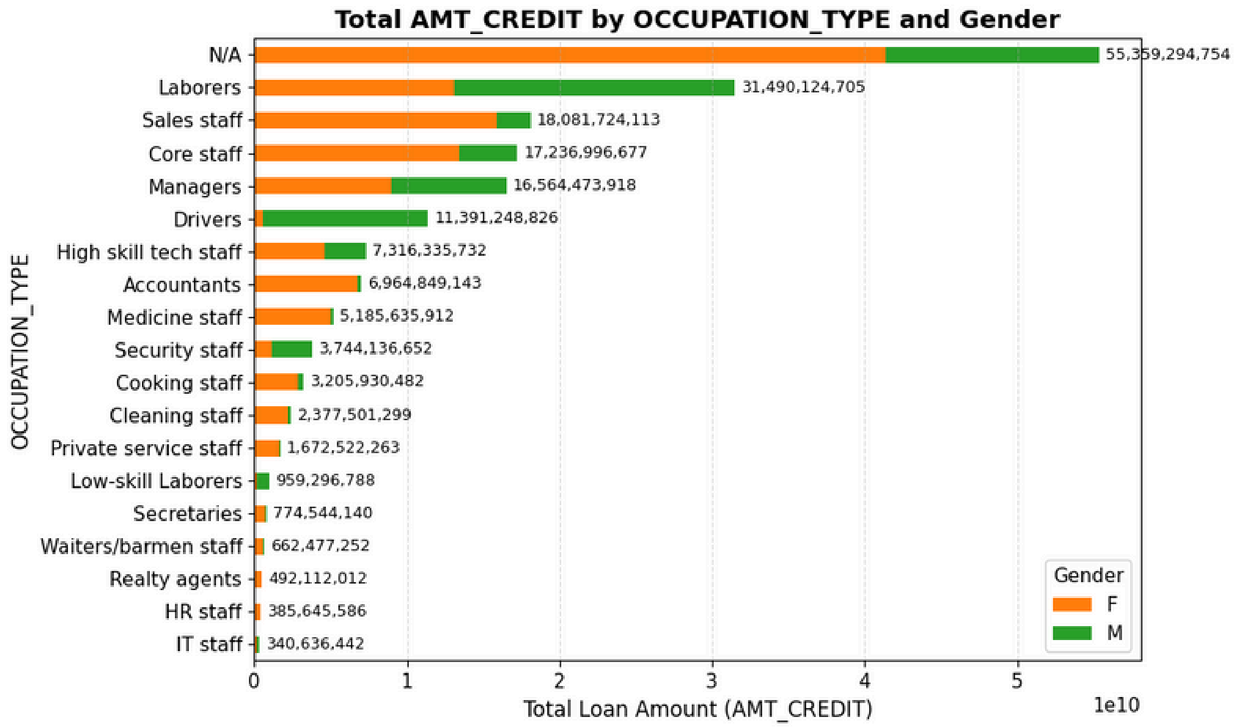
➡ Default rate decreases steadily with years of employment → job stability is a strong risk indicator.

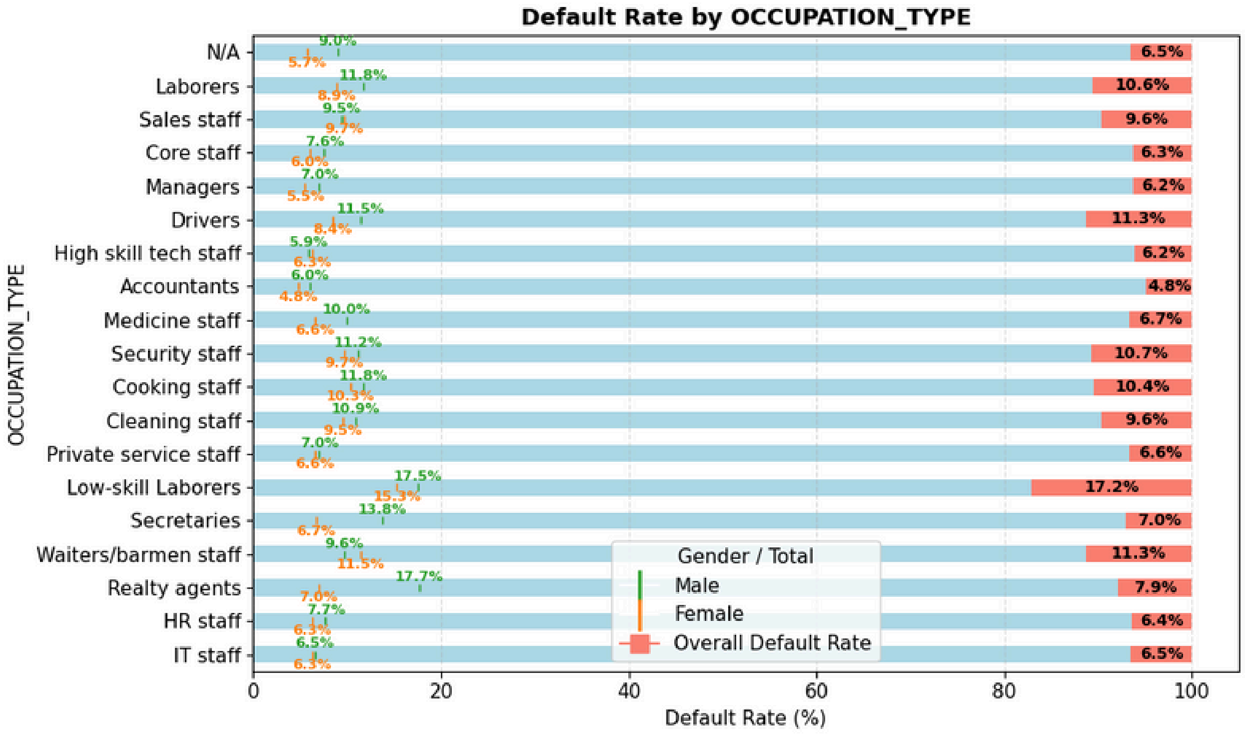
Insight:

- 0–5 years: High loan volume, high risk (12.5% for men) ⚠️ Requires careful income and industry verification
- 5–10 years: Moderate risk, should be further segmented ⚖️ May need additional variables for assessment
- >10 years: Stable default rate under 6% ✅ Should be prioritized for approval
- >20 years: Lowest default rate (4.2%) ⭐ A highly reliable group

Comprehensive Analysis







📊 1. NUMBER OF CONTRACTS BY OCCUPATION

Observations:

- N/A (undisclosed occupation): ~96K ⚠️ Very high
- Laborers: ~55K 🔥 Second highest
- Sales staff: ~32K
- Core staff: ~28K
- Managers: ~21K
- Drivers: ~19K

➡ The majority of contracts are concentrated in N/A and manual labor segments. High-skill and personal service groups are very small (only ~2–11K).

📊 2. TOTAL LOAN AMOUNT BY OCCUPATION

Observations:

- N/A: ~55.4 billion ⚠️ Largest share of loan value
- Laborers: ~31.5 billion 🔥 High
- Sales staff: ~18.1 billion
- Core staff: ~17.2 billion
- Managers: ~16.6 billion
- Drivers: ~11.4 billion

➡ Men dominate total loan volume in manual labor and management, while women lead in office and service sectors.

📊 3. DEFAULT RATE BY OCCUPATION

Observations:

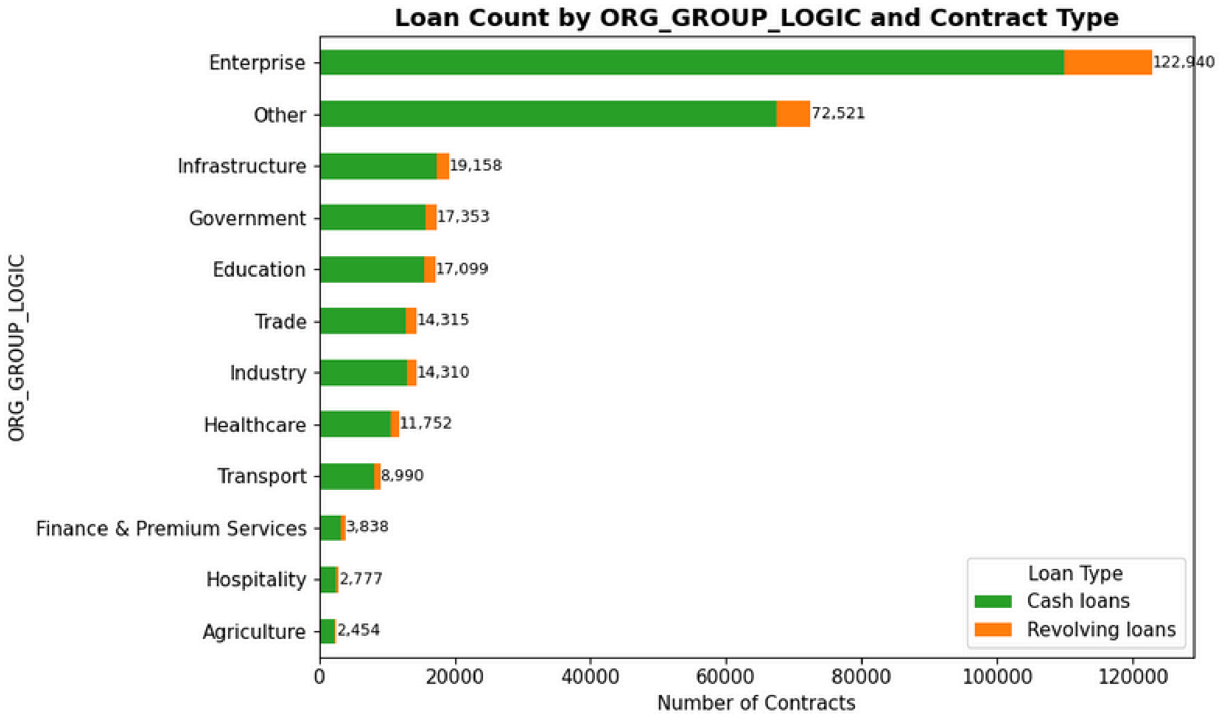
- Low-skill Laborers: 17.2% ⚠️ Very high
- Realty agents: 17.7% ⚠️ Very high
- Drivers, Waiters/Barmen: 11.3% ⚠️ High
- Security staff, Cooking staff, Cleaning staff: ~10–11% – Upper medium
- Sales staff (9.6%), Medicine staff (6.7%), Core staff (6.3%) – Medium
- Accountants, Managers, IT staff: ~4.8–6.5% ✅ Low

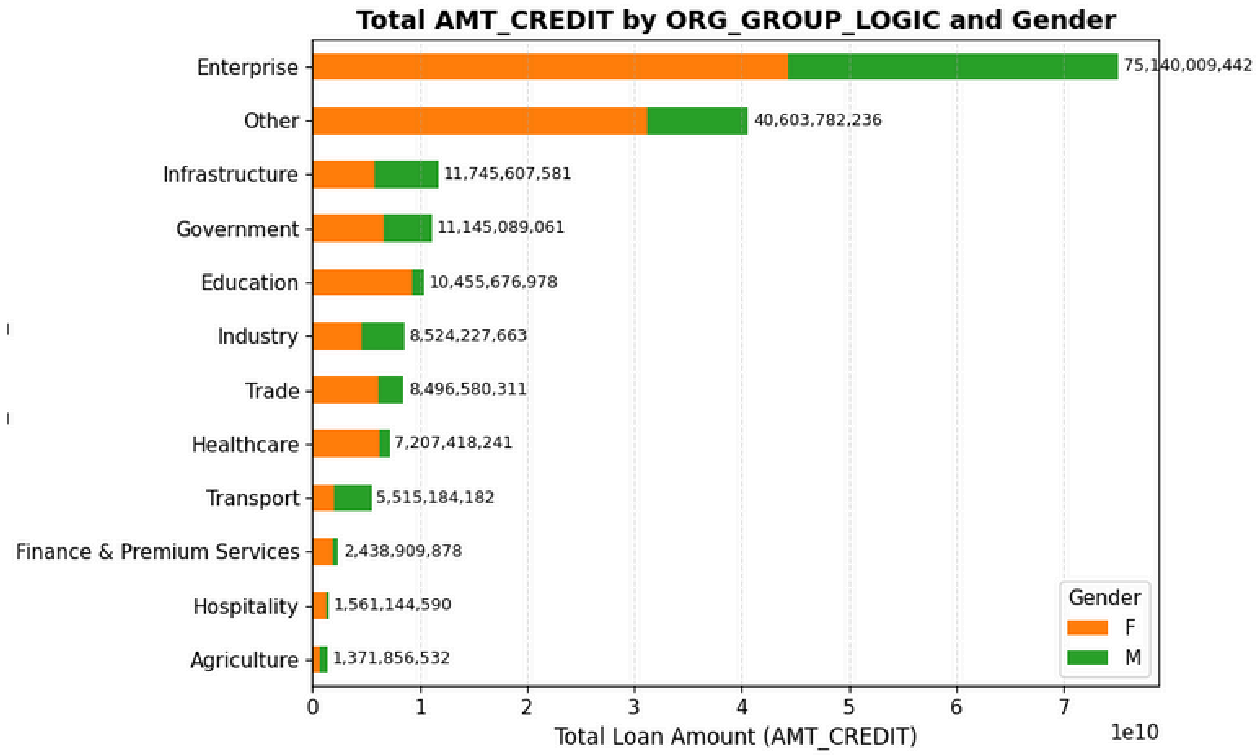
➡ Men in low-skilled or manual labor jobs tend to have higher default rates than women. Women in most job groups have default rates lower than or close to the occupational average.

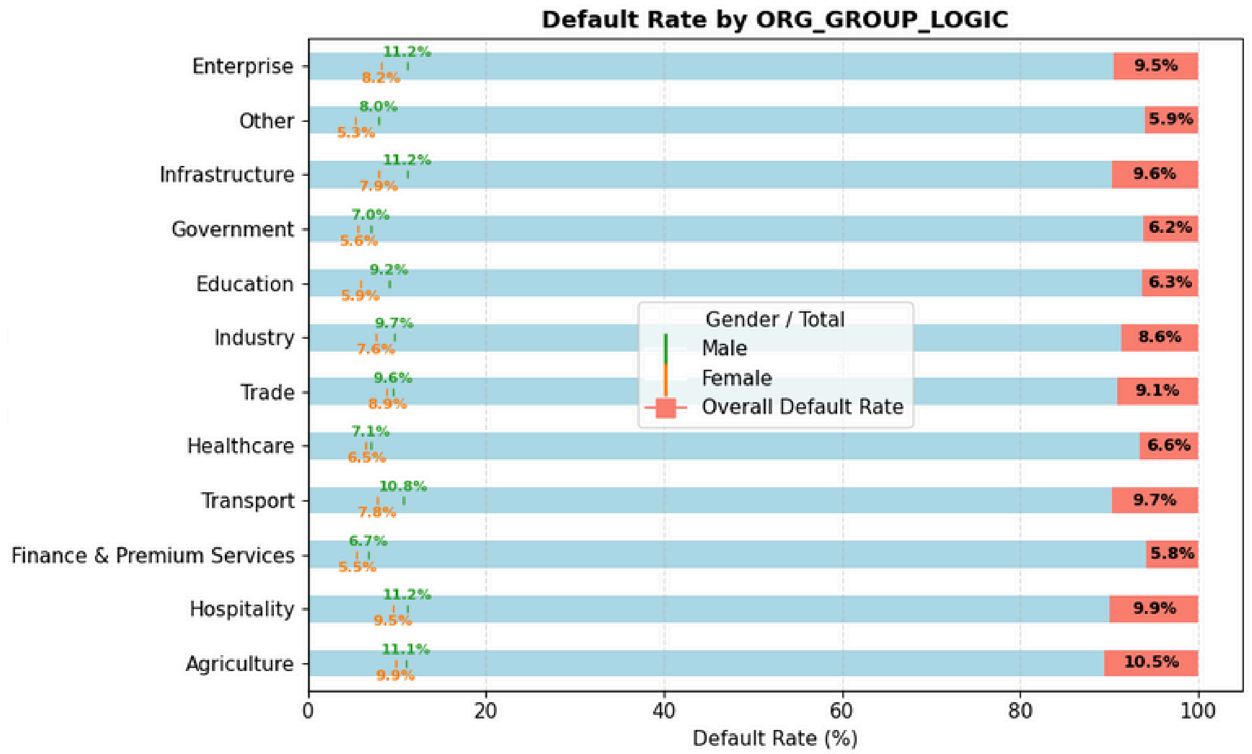
Insight:

- The N/A group holds a large loan volume. These records lacking job information need to be closely reviewed. Many may be retirees.
- Male manual laborers (Laborers, Drivers) borrow heavily but have high default risk ⚠️ → Require stricter verification.
- Female office workers (Sales, Secretaries) have high loan volumes, but default rates are rising slightly.
- High-skill professions (IT, HR, Accountants) show low default rates for both genders — these groups should be prioritized for loan approvals.
- Warning: Both Low-skill Laborers and Realty agents (men and women) have extremely high default rates, exceeding 17%.

Comprehensive Analysis







📊 1. NUMBER OF CONTRACTS BY ORGANIZATION TYPE

Observations:

- Enterprise: ~123K 🔥 Largest group
- Other: ~73K – High
- Infrastructure, Government, Education: ~17–19K
- Trade, Industry: ~14K
- All other groups: <12K

➡ Contracts are mostly concentrated in private enterprises and undefined industry groups.

📊 2. TOTAL LOAN AMOUNT BY ORGANIZATION TYPE AND GENDER

Observations:

- Enterprise: ~75.1 billion 🔥 Dominates
- Other: ~40.6 billion
- Infrastructure, Government, Education: All around 10–11 billion
- Other groups: Below 9 billion

By Gender:

- Enterprise and Other account for more than 60% of total loan value.
- In Other, Government, Education, and Healthcare, women take out more loans than men.

➡ The largest loan value among men comes from those working in private enterprises. In contrast, sectors like education and healthcare skew towards female borrowers.

📊 3. DEFAULT RATE BY ORGANIZATION TYPE

Observations:

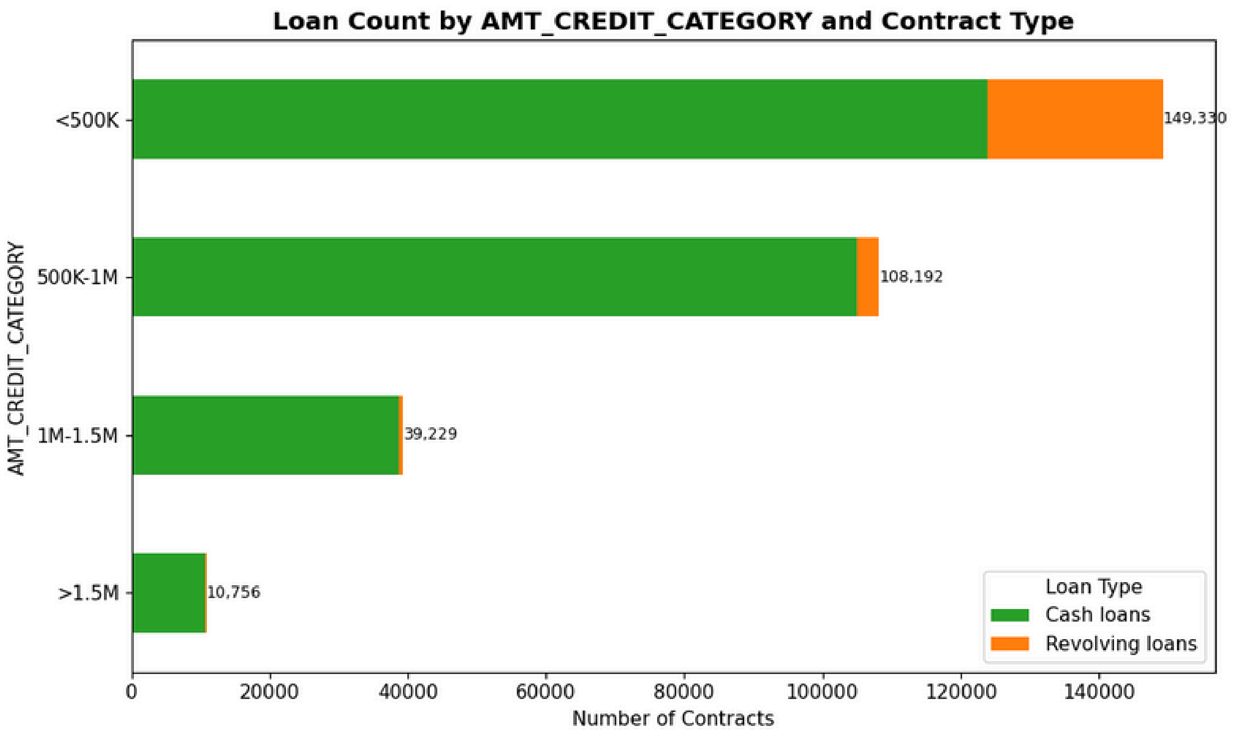
- Agriculture: 10.5%, Hospitality: 9.9%, Transport: 9.7% ⚠️ High risk
- Infrastructure: 9.6%, Enterprise: 9.5% – Medium-high risk
- Trade: 9.1%, Industry: 8.6%
- Remaining groups (Finance, Healthcare, Government, Education): ~5.8–6.6% ✅ Safer segment

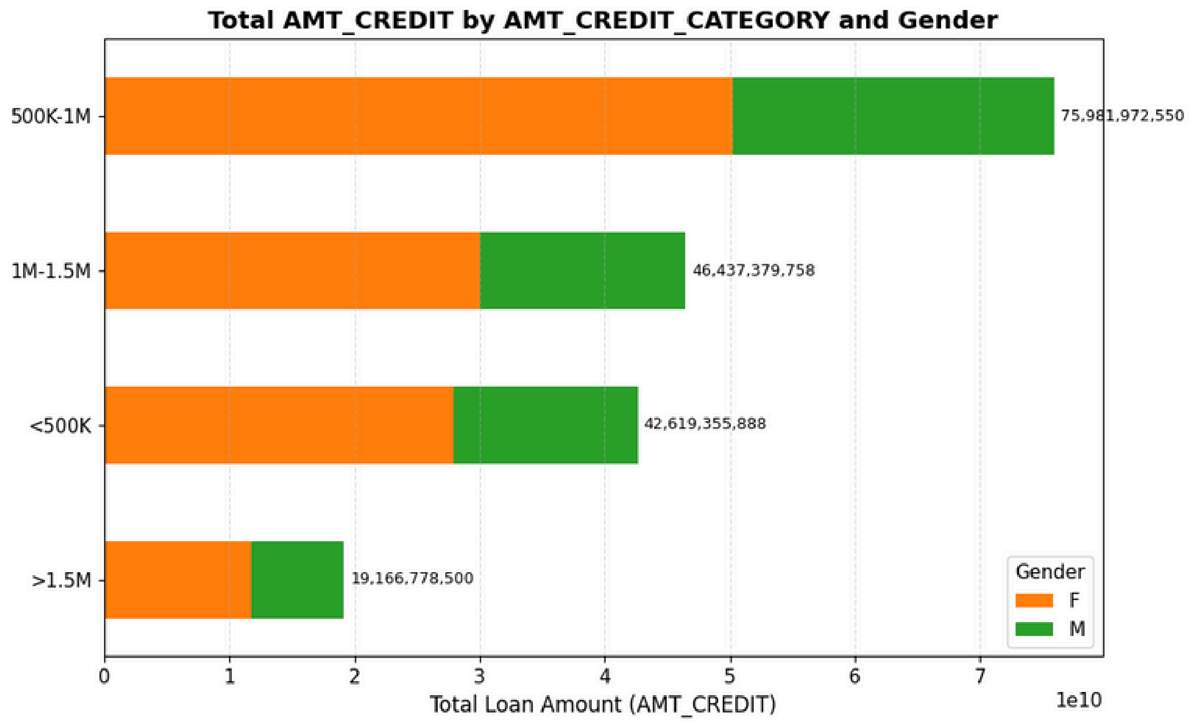
➡ High default risks are concentrated in transportation, agriculture, hospitality, and infrastructure – mainly driven by male borrowers.

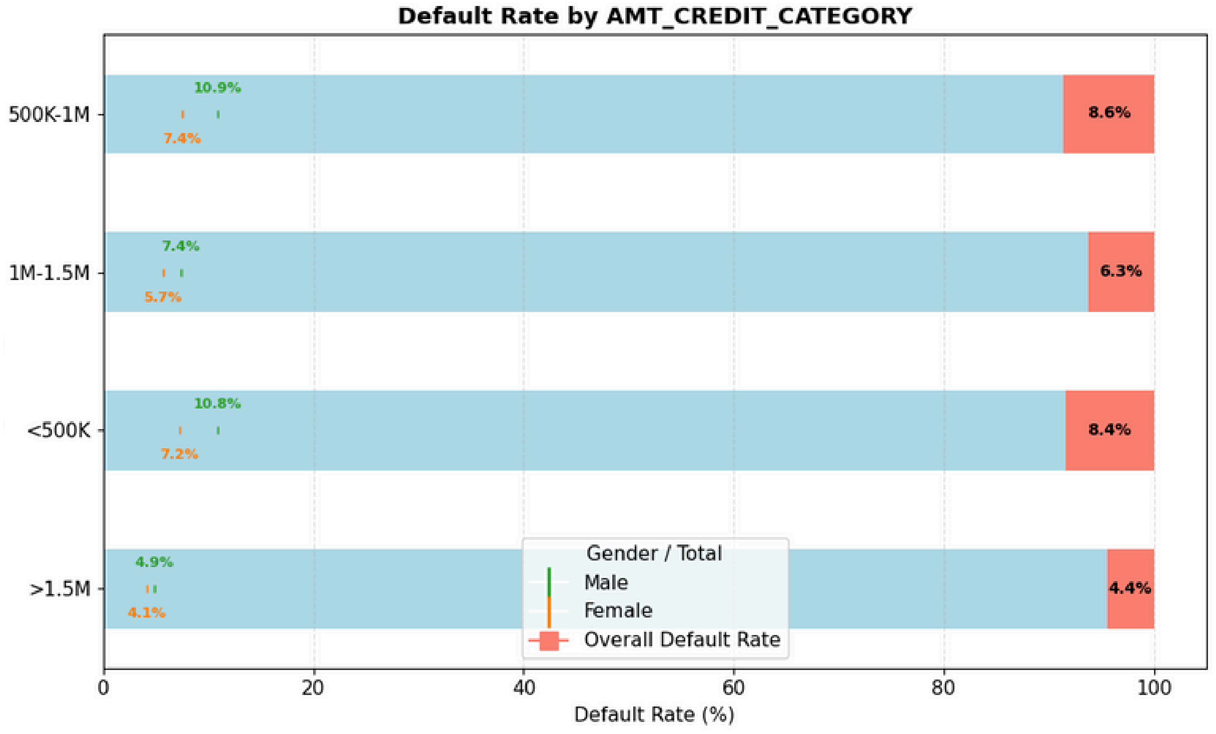
Insight:

- Enterprise and Other dominate in both contract volume and loan amount, with female borrowers as the majority.
- Education and Healthcare are sectors where women lead in contract volume and loan amounts – and show low risk levels.
- Agriculture, Hospitality, Transport, Infrastructure show very high default rates (>9.5%) 🚨 – strict credit assessment needed, especially for male applicants.
- Finance & Premium Services and Healthcare have the lowest default rates (~5.8–6.6%) ✅ – should be prioritized for further credit growth.

Comprehensive Analysis







📊 1. NUMBER OF CONTRACTS BY LOAN AMOUNT RANGE

<u>Observations:</u>

- <500K: ~149K contracts 🔥 Largest segment
- 500K–1M: ~108K contracts – High
- 1M–1.5M: ~39K contracts – Moderate
- 1.5M+: ~11K contracts – Low

➡ Loan contracts are mainly concentrated in small to medium loan sizes.

📊 2. TOTAL LOAN AMOUNT BY LOAN RANGE AND GENDER

<u>Observations:</u>

- 500K–1M: ~76 billion 🔥 Largest total amount
- 1M–1.5M: ~46.4 billion – Moderate
- <500K: ~42.6 billion
- 1.5M+: ~19.2 billion – Low

➡ Clear trend: Women borrow more across all loan size categories.

📊 3. DEFAULT RATE BY LOAN AMOUNT RANGE

<u>Observations:</u>

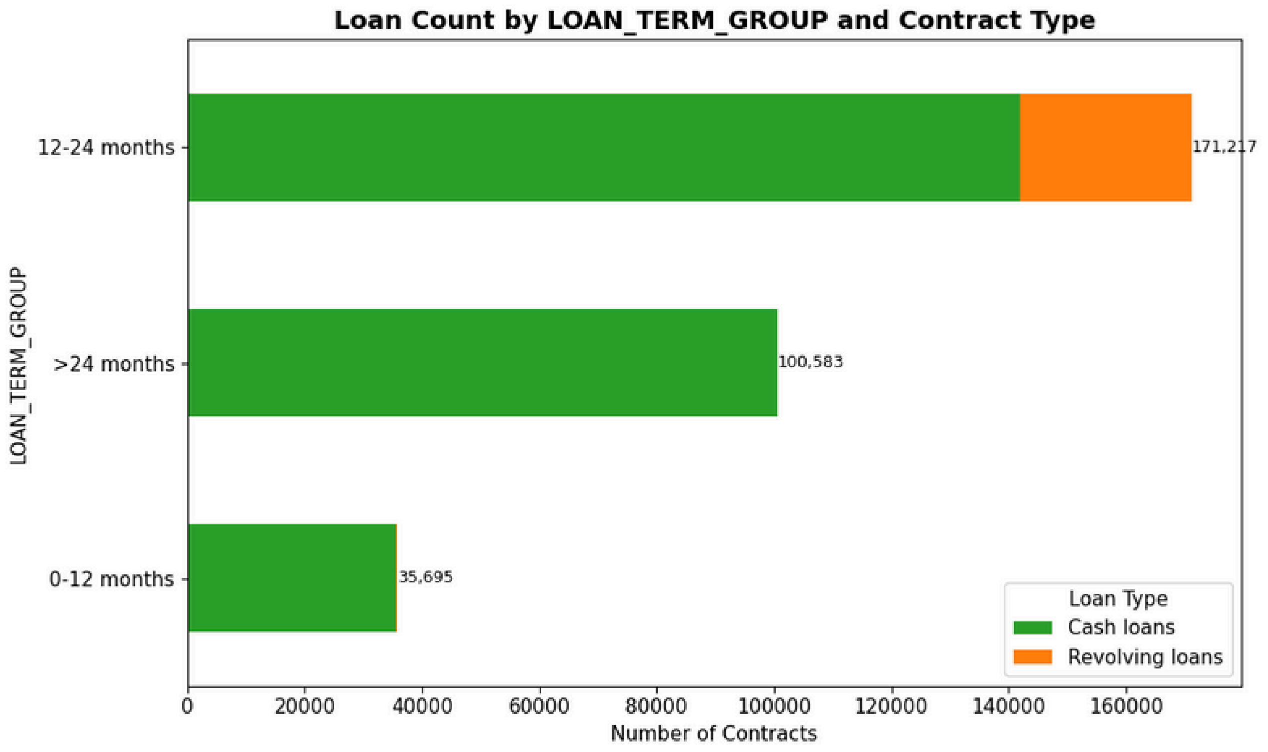
- 500K–1M: 8.6% ⚠️ Highest
- <500K: 8.4% – High
- 1M–1.5M: 6.3% – Medium
- >1.5M: 4.4% ✅ Lowest

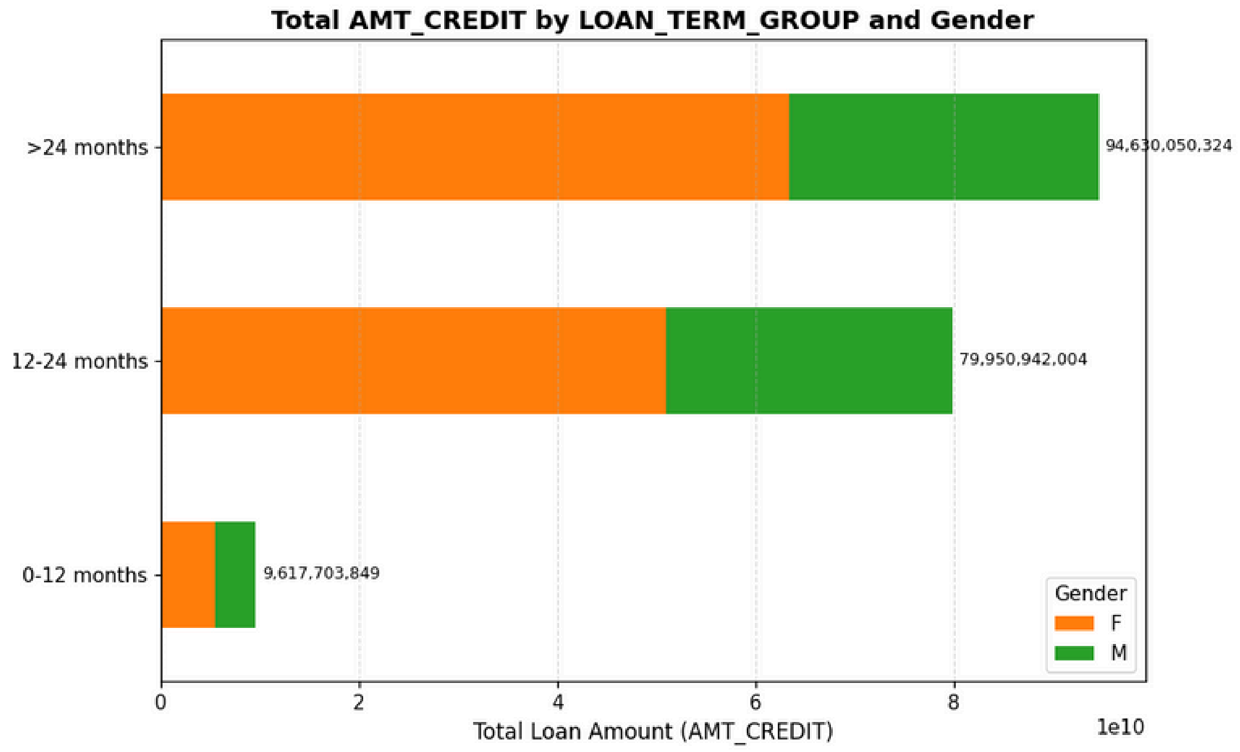
➡ Smaller loans, especially those taken by men, have a significantly higher risk of default. Larger loans (>1.5M) have very low and stable default rates across both genders.

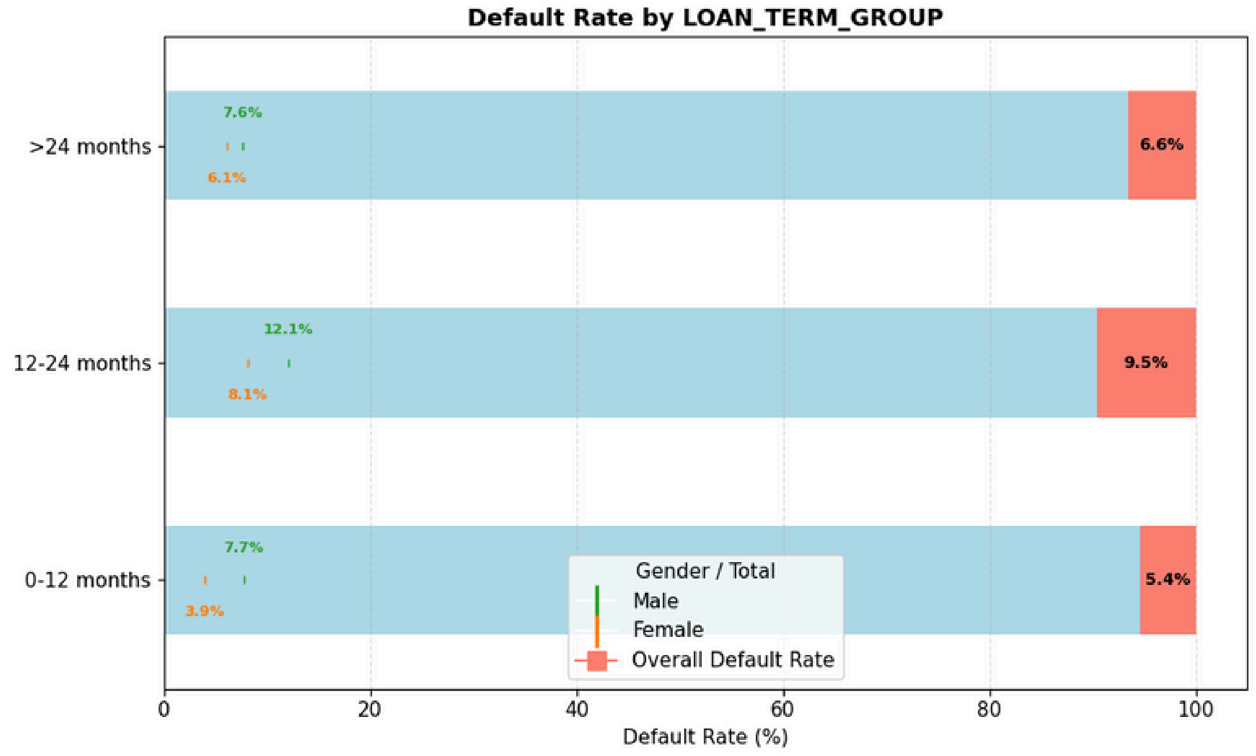
<u>Insight:</u>

- Loans in the 500K–1M and <500K ranges have the highest default rates (~8.6%) 🚨, largely concentrated among male borrowers.
- Loans >1.5M: Despite fewer contracts, they show very strong credit performance (default ~4.4%), suggesting room to expand lending if income documentation is solid.
- Men borrowing small amounts (<1M) show the highest default risk (~10.8–10.9%) ➡ Consider tightening approval criteria for this segment.

Comprehensive Analysis







📊 1. NUMBER OF CONTRACTS BY LOAN TERM

Observations:

- 12–24 months: ~171K contracts 🔥 Largest
- >24 months: ~101K contracts – High
- 0–12 months: ~36K contracts – Low

➡ Loans are mainly concentrated in medium-term (1–2 years). Female borrowers dominate this group. Most revolving loans also fall within the 12–24 month range.

📊 2. TOTAL LOAN AMOUNT BY LOAN TERM

Observations:

- >24 months: ~94.6 billion 🔥 Largest
- 12–24 months: ~80.0 billion – High
- 0–12 months: ~9.6 billion – Low

➡ Even though the number of contracts >24 months is lower, their total loan value is the highest among all terms.

📊 3. DEFAULT RATE BY LOAN TERM

Observations:

- 12–24 months: 9.5% ⚠️ Highest
- >24 months: 6.6% – Moderate
- 0–12 months: 5.4% ✅ Lowest

➡ Loans lasting 12–24 months, especially by male borrowers, show the highest risk of default → repayment capacity of this group should be carefully assessed.

Insight:

- Medium-term loans (12–24 months) account for the largest number of contracts but have the highest default risk (~9.5%), especially among men (12.1%) 🚨.
- Long-term loans (>24 months) represent the largest total loan value and have moderate default risk (~6.6%), mostly taken by men. This is the optimal term for balancing high loan value and acceptable risk.
- Short-term loans (0–12 months) are fewer in number but demonstrate the best credit quality (~5.4%).

Summary and Conclusion

The summary will be compiled using the SWOT analysis framework to identify the most critical and prioritized conclusions.

Summary SWOT

Conclusion

Summary SWOT

Strengths

- Diverse loan types, though mainly concentrated in Cash Loans (95%), which provides an advantage in managing stable cash flow.
- Customers with higher education levels (Higher Education, Academic Degree) show very low risk rates (1.8–5.4%), making them a highly reliable group for expansion.
- Long-term loans (>24 months) represent the highest total credit value (94 billion), yet carry low risk (6.6%) — a solid foundation to scale this segment.
- Married customers account for the largest number of contracts and have relatively low default rates (~7.6%), contributing to portfolio stability.
- Professions such as Managers, Accountants, State Servants, and Pensioners show low risk levels (~4.8–6.3%), helping to optimize profits and reduce risk costs.

Summary SWOT

Weaknesses

- Very high default rates are concentrated among young customers (ages 20–34), especially the 20–24 group (~12.3%).
- Customers in low-skill occupations and with lower education (low-skill laborers, sales staff, drivers) show extremely high default rates (10–17%), driving up risk management costs.
- Short-term loan holders (12–24 months) have the highest risk levels (~9.5%), representing the segment with the most default cases by volume.
- Customers living in rented housing or with parents show high default rates (11.7–12.3%), making it harder to assess their repayment capacity compared to homeowners.
- A large number of loan applications lack occupational data (N/A), which limits accurate risk evaluation, despite forming a significant portion of the dataset.

Summary SWOT

Opportunities

- Expand long-term loan products (over 24 months) and target customers with higher education levels (Higher Education, Academic Degree).
- Grow loan market share among female customers, who show significantly lower default rates compared to males (7.0% vs. 10.1%).
- Develop tailored loan packages for stable professions like Managers, Accountants, State Servants, and Pensioners to ensure long-term portfolio stability.
- Leverage data to personalize interest rates and offer incentives for low-risk, safe customer profiles — such as individuals with over 10 years of work experience, steady income, and smaller households.
- Implement stronger risk prevention measures (financial education, debt counseling) for younger, single, or high-risk segments to minimize future losses.

Summary SWOT

THREATS

- Very high default rates among unemployed individuals and those on maternity leave (36–40%), posing significant risk despite being a small segment.
- Elevated risk levels among low-and middle-income groups, who contribute the highest absolute volume of bad debt.
- Heavy reliance on a single product type (Cash loans); any market shifts or regulatory changes could disrupt the stability of the current portfolio.
- Customers with large households (>4 members) and many children (>3) show higher default rates, indicating financial instability that can be vulnerable to economic fluctuations.
- Incomplete customer data (especially in occupation and organization fields) limits accurate risk prediction and management, leaving room for hidden threats.

Conclusion

STRATEGIC ACTION RECOMMENDATIONS BASED ON SWOT ANALYSIS

- Expand the long-term loan portfolio (loans >24 months) with special incentives to drive sustainable growth.
- Focus on high-quality customer segments: those with high education levels and stable income such as Managers, Accountants, State Servants, Pensioners, and individuals with Higher Education.
- Tighten credit controls and limit approval for high-risk segments, especially young borrowers, low-skill laborers, renters, and unemployed individuals.
- Enhance input data quality by improving the collection of occupation and organization type information to strengthen risk management accuracy.
- Develop and roll out financial education programs targeting young customers and those with limited financial experience to reduce default risk.

THANK YOU

YOUR TIME

Ho Sy Ngoc Tan
HCMC

hst.sytan@gmail.com

039 298 1368