

# Data Visualisation Assignment 1: Analysis of the Bank Loan Defaulter Dataset

Mupparapu Koushik  
IMT2022570  
IIIT-BANGALORE  
Bangalore,India  
Koushik.Mupparapu@iiitb.ac.in

Ananthula Harshith Reddy  
IMT2022023  
IIIT-BANGALORE  
Bangalore,India  
Ananthula.Reddy@iiitb.ac.in

Pathaneni Anirudh  
IMT2022505  
IIIT-BANGALORE  
Bangalore,India  
Pathaneni.Anirudh@iiitb.ac.in

## I. INTRODUCTION

This report analyzes bank loan data, including both current and previous applications, to uncover patterns and trends. Using Tableau visualizations, we explore factors that influence loan approvals, such as client type, loan purpose, and industry. By examining historical loan data, we aim to provide insights into how banks can improve their decision-making processes and better assess risks. These visualizations help identify key factors that affect loan outcomes, offering valuable guidance for optimizing loan approval strategies.

## II. DATASET

This dataset helps to explore various factors affecting loan approval rates and client behavior. We are using a total of 20 columns for our analysis of bank loan applications and previous application data. These columns are:

- 1) Name Contract Type: Type of loan contract (Cash loans, Revolving loans, etc.)
- 2) Name Client Type: Type of client (New or Repeated customer)
- 3) Name Contract Status: Status of the loan application (Accepted or Rejected)
- 4) Name Portfolio: The portfolio of the client
- 5) Amt Annuity: Annual payment amount for the loan
- 6) Amt Credit: Credit amount requested in the loan
- 7) Name Product Type: Type of loan product requested
- 8) Name Cash Loan Purpose: Purpose for which the loan was requested
- 9) Seller Industry: The industry of the seller associated with the loan
- 10) Loan Amount: The amount of loan requested by the applicant
- 11) Code Gender : Gender of the Client (M for Male , F for Female)
- 12) Name Education Type : Level of highest education the client has achieved
- 13) Name Family Status : Family status of the client
- 14) Name income Type : Clients income type (working, student, maternity leave,...)

- 15) Occupation Type : What kind of occupation does the client have

## III. CALCULATED FIELDS

Additionally, we have calculated specific fields to support our analysis:

- Approval Percentage: The percentage of loan applications that were approved
- Loan Amount Percentage: The percentage of the total loan amount that was approved
- Default Percentage: The percentage of borrowers that have failed to repay their loans
- Collateral : Categorizes collateral amounts into ranges: 0 - 500K, 500K - 1M, 1M - 2M, and 2M+
- Age : Categorizes age into ranges: Under 20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60+

These calculated fields help provide a clearer understanding of loan trends and approval patterns within the dataset.

## IV. TASK DIVISION

The project was systematically divided into three key tasks to ensure a comprehensive analysis of loan approval and default trends. Each task was assigned based on specific areas of focus:

We divided our work into three major tasks:

- T1: Factors affecting default rates of loans
- T2: Factors affecting approval status of previous loan applications
- T3: Comparing Trends between Approval Rate and Default Rate

## V. T1: FACTORS AFFECTING DEFAULT RATES OF LOANS

This analysis identifies the factors that contribute to loan default rates across various categorical and financial characteristics, using data visualizations to interpret trends. Factors such as gender, type of loan, education level, family status, income type, occupation, collateral, and age have been examined. The report aims to understand which factors increase the likelihood of default and how different risk factors interact with one another.

### A. Factor 1 - Gender

Females have a 7% default rate, while males have a 10.1% default rate. This suggests that males are more likely to default on loans compared to females. The reason for this could be that males tend to take on more financial risks or larger loans, leading to a higher chance of default. In contrast, women are often seen to be more cautious with financial decisions, which may contribute to their lower default rates.

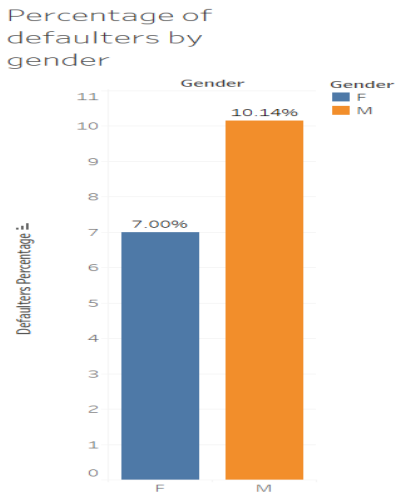


Fig. 1. Percentage of defaulters by gender

#### Marks Used:

- **1D mark:** line

#### Channels Used:

- **Position (X-axis):** Gender (categorical variable).
- **Length (Y-axis):** Default rate (height of the bars).
- **Color:** Distinct colors for each gender.

#### Reason for Choosing Bar Chart:

A bar chart is an effective way to compare default rates between males and females, offering a clear visual distinction between the two categories using color and bar height.

### B. Factor 2 - Type of Loan

Cash loans have a default rate of 8.3%, while revolving loans have a 5.4% default rate. This suggests that Cash loans are more likely to result in default than revolving loans. The reason for this could be that Revolving loans, such as credit cards, typically have structured repayment processes where the credit limit is increased only after proper repayment. In contrast, cash loans often involve lump-sum payments, which can be harder to manage, leading to a higher default rate.

#### Marks Used:

- **1D mark:** line

Percentage of defaulters by Type of Loan

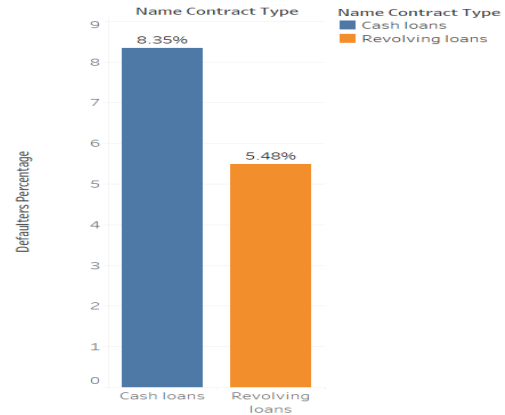


Fig. 2. Percentage of defaulters by Loan Type

#### Channels Used:

- **Position (X-axis):** Loan type (categorical variable).
- **Length (Y-axis):** Default rate (height of the bars).
- **Color:** Different colors for cash loans and revolving loans.

#### Reason for Choosing Bar Chart:

A bar chart makes it easy to compare the default rates of cash loans and revolving loans, emphasizing the difference between the two categories with clear color coding and bar height.

### C. Factor 3 - Education Level

Percentage of defaulters by Education

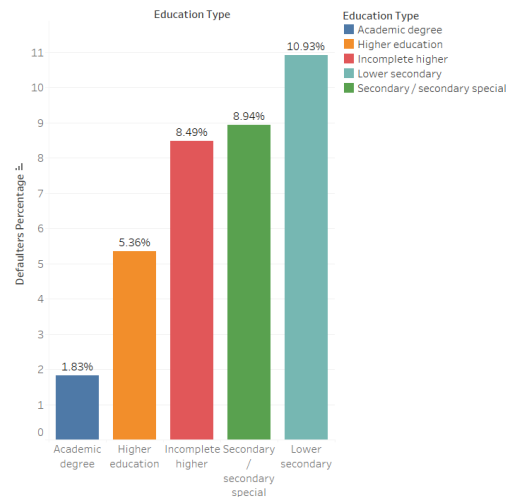


Fig. 3. Percentage of defaulters by Education Level

Lower secondary education has a default rate of 10.9%, secondary education 8.9%, incomplete higher education 5.3%, and academic degrees 1.8%. This suggests that individuals

with lower levels of education are more likely to default compared to those with higher education. The reason for this could be that higher education often leads to better financial literacy, higher-paying jobs, and more stable incomes, which reduce the likelihood of default. Conversely, lower education levels may limit access to such advantages, increasing the risk of default.

*Marks Used::*

- **1D mark:** line

*Channels Used::*

- **Position (X-axis):** Education levels (categorical variable).
- **Length (Y-axis):** Default rate (height of the bars).
- **Color:** Distinct colors for different education levels.

*Reason for Choosing Bar Chart::* The bar chart clearly illustrates how default rates differ across education levels, making it simple to compare categories while using color to differentiate education levels.

#### D. Combined Factors – Gender, Loan Type, Education Level

Tree map of Gender, Loan Type, Education Level by Defaulters Percentage

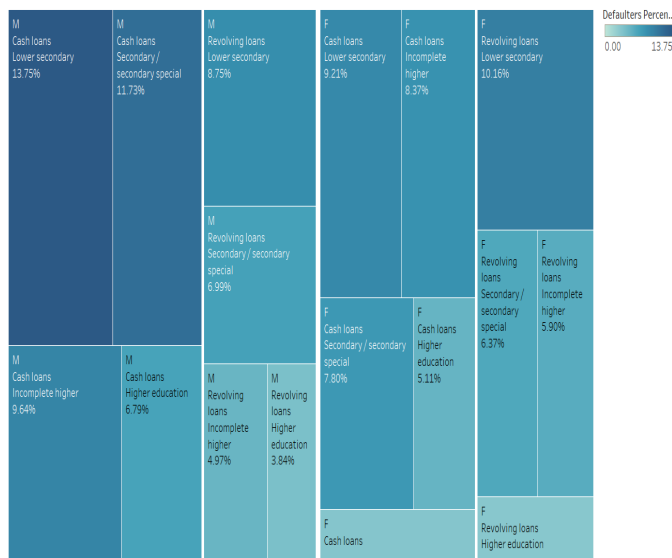


Fig. 4. Tree map of Gender, Loan Type and Education Level

The node combining male, cash loan, and lower secondary education has the highest default rate (13.75%). This suggests that when these factors are combined, the likelihood of default increases significantly, indicating that trends observed individually still hold when considered together. The reason for this could be that males tend to take on more financial risks, cash loans involve lump-sum payments that are harder to manage, and individuals with lower education levels may have less financial literacy and job stability. The combination of these factors creates a compounding effect, where the risk of default is higher when multiple risk factors are present. This means that borrowers with multiple high-risk characteristics, such as being male, opting for cash loans, and having lower education levels, are at a much greater risk of defaulting.

*Marks Used::*

- **2D mark:** area (blocks in the tree map).

*Channels Used::*

- **Position:** Nested blocks representing combinations of gender, loan type, and education level.
- **Area:** Block size represents the default rate.
- **Color saturation:** more saturated colors highlight higher default rate.

*Reason for Choosing Tree Map::* A tree map efficiently displays the interaction between three categorical variables in a compact, organized format. The size and color of blocks provide a visual overview of where the highest default rates occur.

#### E. Factor 4 - Family Status

Percentage of defaulters by Family status

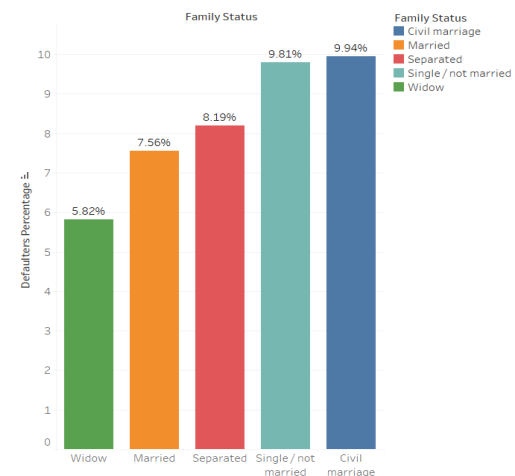


Fig. 5. Percentage of defaulters by Family Status

People in civil marriages have a default rate of 9.94%, single individuals 9.81%, separated individuals 8.19%, married individuals 7.56%, and widows 5.82%. This indicates that individuals in civil marriages or who are single are more likely to default on loans compared to those who are married or widowed. One possible reason is that civil marriages, also known as register marriages, may involve couples who eloped due to familial objections or other personal reasons. These couples might lack the emotional and financial support typically provided by families, making them more vulnerable to financial difficulties. Conversely, traditionally married individuals often benefit from family support and shared financial responsibilities, which can mitigate the risk of default.

Separated individuals may have already made financial adjustments post-separation, leading them to be more cautious with their finances, thus lowering their default risk. Widows, despite the loss of a partner, may have access to financial resources such as life insurance, pensions, or inheritances, which can offer additional financial security. This suggests that while separated and widowed individuals face significant

life changes, their financial safety nets or increased financial awareness can help reduce their risk of default.

#### F. Factor 5 - Income Type

Individuals on maternity leave have a default rate of 40%, and the unemployed have a default rate of 36.36%, both of which are significantly higher compared to working individuals, state servants, or pensioners. This suggests that individuals who are unemployed or on maternity leave are more likely to default on loans. The reason for this is primarily the lack of a stable income. Unemployed individuals have no regular paycheck, making it challenging to manage loan repayments and meet other financial obligations. Similarly, individuals on maternity leave may experience financial strain due to the temporary loss of their regular income. While they may receive some benefits or support, these are often insufficient to cover all expenses, leading to increased difficulty in maintaining loan payments. Consequently, the instability in income for both groups contributes to their higher default rates.

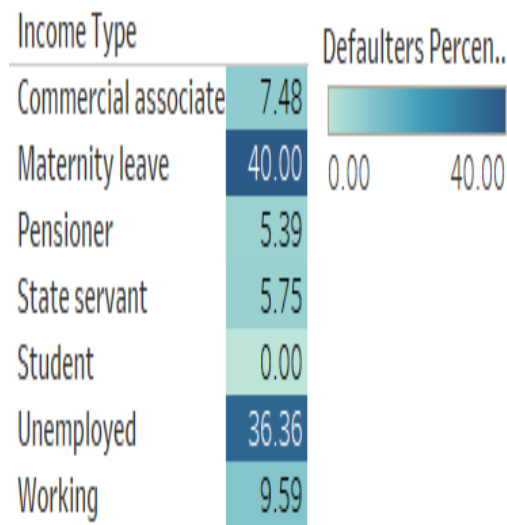


Fig. 6. Percentage of defaulters by Income Type

#### Reason for Choosing Highlight Table:

Color saturation makes it easy to spot which income types have the highest default rates. The simplicity of the table allows for quick identification of the categories most at risk.

#### G. Combined Factors – Family Status, Income Type

The node with unemployed individuals who are married has the highest default rate (60%). This suggests that even though marriage is usually associated with a lower default rate, unemployment can significantly increase the risk of default. The reason for this could be that unemployed individuals lack a stable income, which makes it difficult to manage financial obligations, particularly for married people who may have additional family responsibilities. The financial burden of supporting a household without any income increases the likelihood of default, despite the usual support system that

marriage provides. This means that being unemployed can negate the benefits of marriage in terms of financial stability, making unemployment a critical factor in assessing default risk.

Tree Map of income type and family status

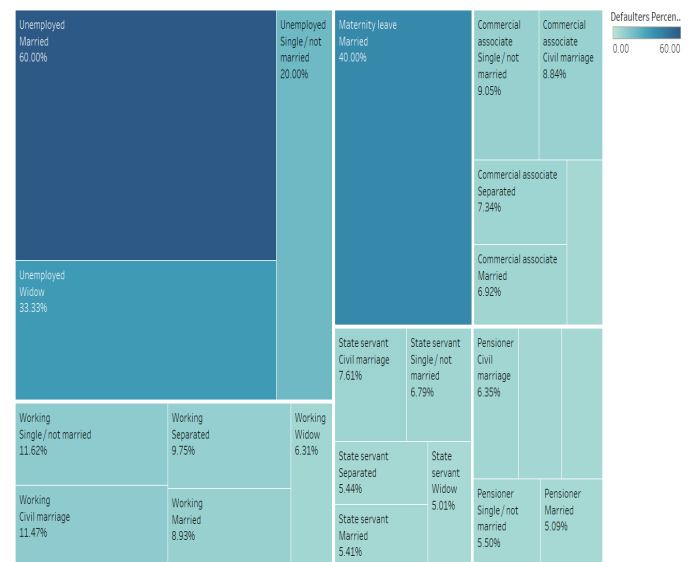


Fig. 7. Tree map of Family Status, Income Type

#### Marks Used::

- **2D mark:** area (blocks in the tree map).

#### Channels Used::

- **Position:** Nested blocks representing combinations of gender, loan type, and education level.
- **Area:** Block size represents the default rate.
- **Color saturation:** more saturated colors highlight higher default rate.

**Reason for Choosing Tree Map::** By using color saturation to emphasize the default percentage, the tree map becomes more intuitive. Color intensity helps identify high-risk combinations of family status and income type.

#### H. Factor 6 - Occupation Type

Low-skill laborers have a default rate of 17.15%, followed by drivers at 11.33% and waiters at 11.28%. This suggests that individuals in low-skill jobs are more likely to default on loans compared to those in higher-skill occupations. The reason for this could be that low-skill jobs typically offer lower wages and less job security, making it harder for individuals to manage loan repayments. With less stable and lower income, these individuals may struggle to meet their financial obligations. In contrast, high-skill jobs generally provide higher and more stable incomes, which can help reduce the risk of default by offering better financial security and the ability to handle unexpected expenses.

#### Marks Used::

- **2D mark:** area (blocks in the tree map).

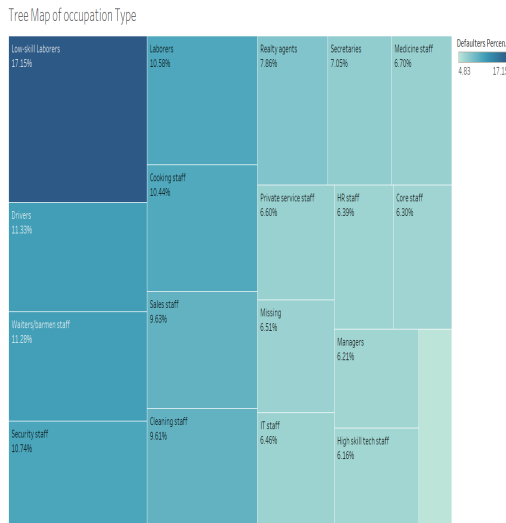


Fig. 8. Percentage of defaulters by Occupation Type

#### Channels Used::

- **Position:** Nested blocks for occupation types.
- **Area:** Block size represents the default rate.
- **Color saturation:** more saturated colors highlight higher default rates.

*Reason for Choosing Tree Map::* A tree map is the ideal solution for presenting a large number of occupation categories in a compact space. It allows the viewer to quickly identify which occupations are associated with higher default rates.

#### I. Factor 7 - Collateral

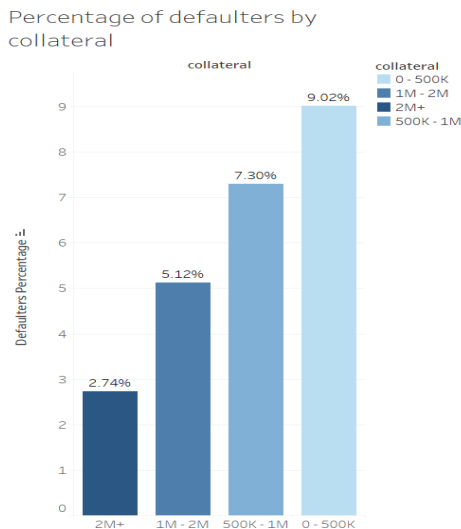


Fig. 9. Percentage of defaulters by Collateral

Borrowers with collateral in the range of 0-500K have a default rate of 9% , those with collateral in the range of 500K-1M have a default rate of 7.3%, borrowers with collateral between 1M-2M have a default rate of 5.12%, and those with

collateral above 2M have the lowest default rate at 2.74%. This suggests that applicants with low collateral are more likely to default on loans than those with high collateral. The reason for this could be that low collateral indicates fewer financial resources or savings, making it more difficult for borrowers to repay their loans if they encounter financial difficulties. Higher collateral serves as a safety net, allowing individuals to cover loan payments in challenging situations thereby reducing the risk of default.

#### Marks Used::

- **1D mark:** line

#### Channels Used::

- **Position (X-axis):** Collateral range (binned categories).
- **Length (Y-axis):** Default rate (height of the bars).
- **Color saturation:** Higher saturation corresponds to higher collateral amounts.

*Reason for Choosing Bar Chart::* A bar chart makes it simple to compare default rates across different collateral ranges. The use of color saturation adds an additional layer of information, visually indicating the impact of higher collateral amounts on default rates.

#### J. Factor 8 - Age

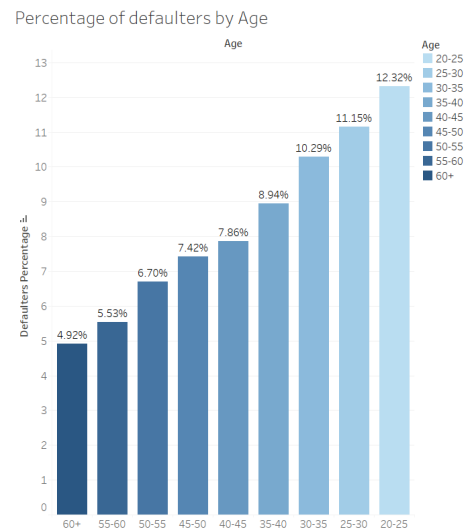


Fig. 10. Percentage of defaulters by Age

Individuals aged 20-25 and 25-30 have the highest default rates at 12.32% and 11.15% respectively .Additionally, the default rate appears to be inversely proportional to age, with default rates consistently decreasing as age increases. This suggests that younger individuals are more likely to default on loans compared to older individuals. The reason for this could be that younger people are often in the early stages of their careers, with less income stability and financial experience, making loan management more difficult. They may also have other financial obligations, such as education loans or personal debts, which can increase their financial burden. In contrast, older individuals are more likely to have established careers,

higher income, and greater financial maturity, all of which contribute to a lower likelihood of defaulting on loans.

*Marks Used::*

- **1D mark:** bar.

*Channels Used::*

- **Position (X-axis):** Age range (binned categories).
- **Length (Y-axis):** Default rate (height of the bars).
- **Color saturation:** Higher saturation corresponds to older age groups.

*Reason for Choosing Bar Chart::* The bar chart effectively communicates default rates across different age ranges. The added use of color saturation highlights the correlation between age and default rates, with higher saturation for older, more stable borrowers.

### K. Combined Factors – Collateral and Age

Bubble chart of collateral and age

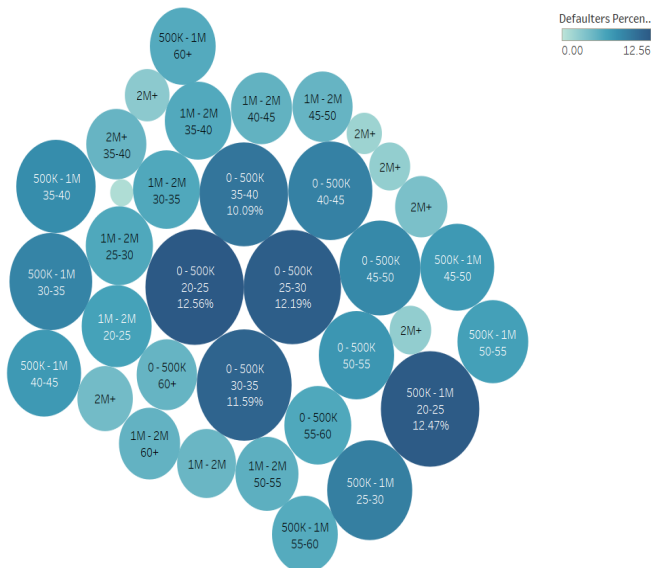


Fig. 11. Bubble chart of Collateral and Age

In analyzing the top four nodes for default rates, we see a pattern where both collateral and age play significant roles:

- 1) Collateral (0-500K), Age (20-25)
- 2) Collateral (500K-1M), Age (20-25)
- 3) Collateral (0-500K), Age (25-30)
- 4) Collateral (0-500K), Age (30-35)

At first glance, low collateral (0-500K) is present in three of the top four nodes, indicating that having less collateral is a significant risk factor. Borrowers with lower collateral tend to default more, likely because they lack the financial security or assets to cover loan payments in the event of financial challenges.

However, a closer look reveals that age is a more dominant risk factor. Specifically, individuals in the 20-30 age range appear in the top three nodes, regardless of whether their collateral is low (0-500K) or higher (500K-1M). Even with a

higher collateral amount (as seen in the second node), individuals aged 20-25 still show high default rates. This highlights that age, particularly in the younger range, outweighs collateral in predicting loan default.

The likely reason for this is that younger individuals typically have less income stability, fewer savings, and less financial experience. They may be in the early stages of their careers, juggling other financial commitments, or facing life transitions that make managing loans more difficult. Despite having higher collateral, younger borrowers may still struggle with financial management, making age a critical risk factor in determining the likelihood of default. Thus, while low collateral amplifies risk, being younger introduces a heightened likelihood of default, even with more substantial collateral.

In conclusion, age—specifically, being in the 20-30 range—emerges as a more significant factor than collateral in predicting default risk. Low collateral does increase the risk of default, but younger individuals face more challenges with income and managing their finances, making them more prone to default, even with larger collateral amounts.

*Marks Used:*

- **0D mark:** point (bubble).

*Channels Used:*

- **Size:** Default rate (larger bubbles represent higher default rates).
- **Color saturation:** Default percentage (higher saturation indicates a higher default rate).

*Reason for Choosing Bubble Chart:*

The bubble chart effectively represents the relationship between two key factors: **age** and **collateral**. By using **bubble size** and **color saturation** to indicate the default rate, the chart makes it easy to visually identify areas where default risk is higher. The saturation of the bubbles helps emphasize which combinations of age and collateral carry the greatest default risk.

## VI. T2: FACTORS AFFECTING APPROVAL RATES OF LOANS

The analysis shows that a significant portion of the total loans received are approved, with an approval rate of 61.6%. Interestingly, the percentage of refused loans is relatively lower at 17.64%, which is even less than the cancelled applications, which account for 19.13%. Additionally, a very small percentage of loans—just 1.6%—fall under unused offers. This suggests that banks are generally inclined to approve loans, provided that applicants present valid reasons and purposes for their requests.

*Marks Used:*

- **1D mark:** Pie Slice.

*Channels Used:*

- **Position (X-axis):** Approval status (categorical variable).
- **Angle:** Proportion of each approval status (representing percentage in the pie chart).
- **Color:** Distinct colors for each approval status.



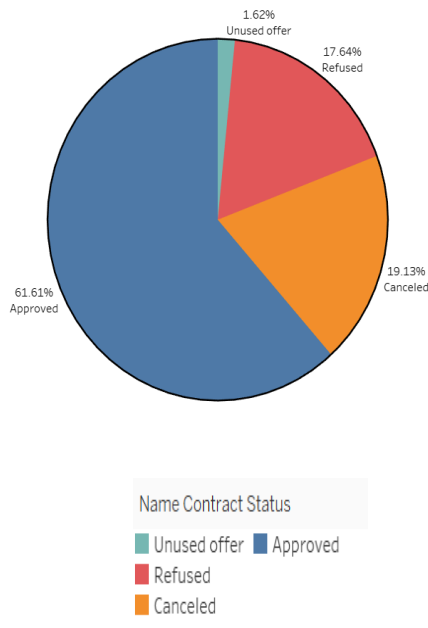


Fig. 12. Approval Status of loan applications

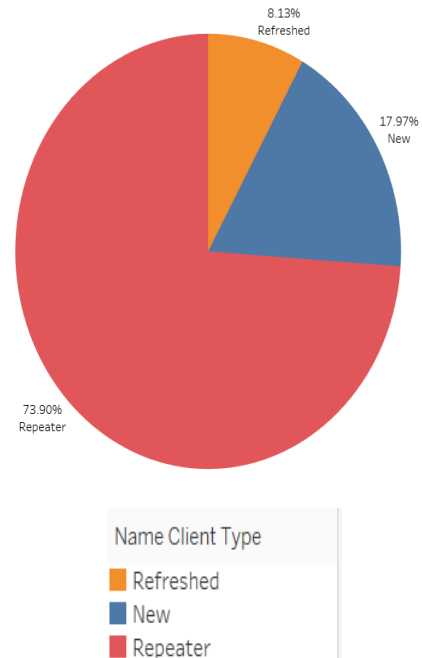


Fig. 13. Distribution of Loan Applications by Client Type

#### Reason for Choosing Pie Chart:

A pie chart was chosen because it visually represents the proportion of each loan approval status clearly and effectively. The angle and color channels make it easy to compare approval, refusal, and other statuses at a glance.

#### A. Factor 1 - Client Type

The pie chart(Fig. 13) shows that the majority of loan applications 73.9% come from repeat clients, while new clients make up 17.97%, and refresher clients account for 8.13%. However, the approval rates tell a different story(Fig. 14). New clients have the highest approval rate at 93.37%, while repeat clients have the lowest at 53.55%. This suggests that banks may prioritize attracting new customers by offering better loan terms and easier approvals. New clients likely have cleaner credit profiles, making them appear less risky. In contrast, repeat clients, who may have more complex credit histories or apply more frequently, face stricter evaluations. Refresher clients, who reapply after some time, fall in between with moderate approval rates.

Overall, banks seem more willing to approve loans for new clients, possibly viewing them as safer and more reliable borrowers.

#### Marks Used:

- **1D mark for Pie Chart:** Pie Slice.
- **1D mark for Bar Chart:** Line.

#### Channels Used for Pie Chart (Client Type vs Count):

- **Angle:** Proportion of each client type (representing the count of loan applications in the pie chart).
- **Color:** Distinct colors for each client type.

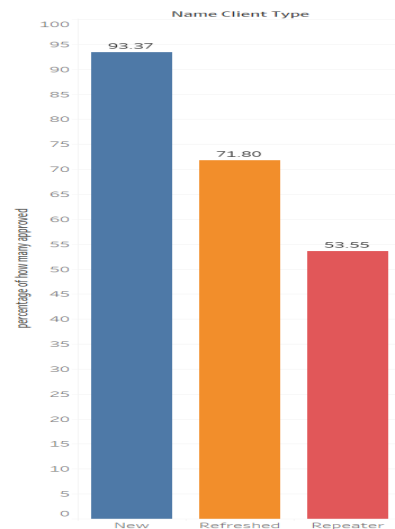


Fig. 14. Percentage of Approval by Client Type

#### Channels Used for Bar Chart (Approval Status by Client Type):

- **Position (X-axis):** Portfolio Type.
- **Height (Y-axis):** Approval Percentage
- **Color:** Distinct colors for each client type (e.g., New, Repeat).

#### Reason for Choosing Charts:

- **Pie Chart (Client Type vs Count):** A pie chart is ideal for representing the proportion of loan applications from different client types, as the entire dataset sums to 100

- **Bar Chart (Approval Percentage by Client Type):** A bar chart was chosen to compare the loan approval percentage across different client types, with the bars colored based on client type. This provides a clear visual comparison of how each client type performs in terms of loan approvals and rejections.

## B. Factor 2 - Portfolio type

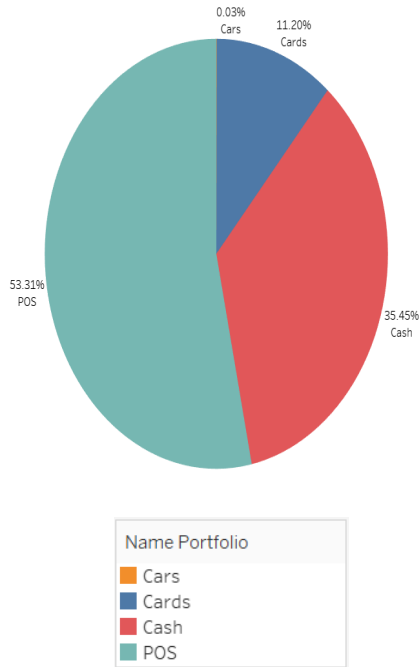


Fig. 15. Distribution of Loan Applications by Portfolio type

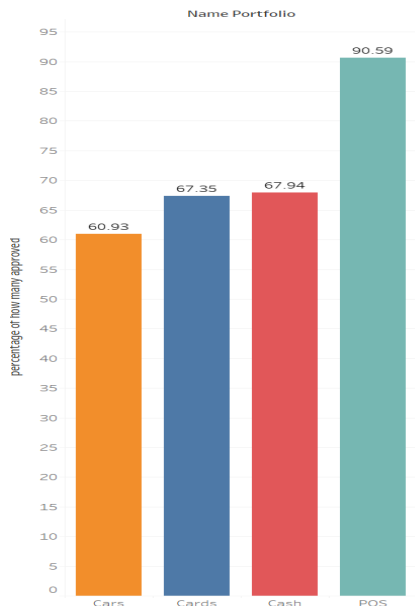


Fig. 16. Percentage of Approval by Portfolio Type

The data reveals that the majority of loan applications are for POS (Point of Sale) loans, which constitute 53.31% of the total applications. These are followed by cash loans at 35.45%, card loans at 11%, and car loans at a minimal 0.03% (Fig. 15). Among these, POS loans boast the highest approval rate at 90%. This high approval rate can be attributed to their common use for consumer goods like electronics and home appliances, which are generally lower-risk purchases. POS loans often involve smaller amounts and come with flexible repayment terms, making them less risky for banks.

In contrast, car loans, despite representing a tiny fraction of total applications, still have a relatively high approval rate of 60%. This is largely because car loans are secured by the vehicle itself, providing collateral that mitigates the bank's risk and thus facilitates higher approval rates.

Cash and card loans, which are unsecured or linked to revolving credit, show approval rates of around 67% (Fig. 16). Cash loans are popular but pose a higher risk due to the lack of collateral, leading banks to be more cautious in their approval process. Similarly, card loans can be riskier due to the potential for high credit utilization and ongoing debt, contributing to their moderate approval rates.

In summary, POS loans are favored by banks due to their lower risk profile and attractive terms for consumers, while car loans benefit from collateral security. On the other hand, cash and card loans face stricter approval criteria due to their unsecured nature and revolving credit risks.

### Marks Used:

- **1D mark for Pie Chart:** Pie Slice.
- **1D mark for Bar Chart:** Line.

### Channels Used for Pie Chart (Portfolio Type vs Count):

- **Angle:** Proportion of each portfolio type (representing the count of loan applications in the pie chart).
- **Color:** Distinct colors for each portfolio type.

### Channels Used for Bar Chart (Approval Status by Portfolio Type, Colored by Client Type):

- **Position (X-axis):** Portfolio Type
- **Height (Y-axis):** Approval Percentage.
- **Color:** Distinct colors for each client type (e.g., New, Repeat).

### Reason for Choosing Charts:

- **Pie Chart (Portfolio Type vs Count):** The pie chart provides a straightforward way to visualize how loan applications are distributed across different portfolio types, giving a clear representation of the share each portfolio holds within the dataset.
- **Bar Chart (Approval Percentage by Portfolio Type):** A bar chart is effective for comparing the approval percentage for different portfolio types. This makes it easy to see the distribution of approval percentage within each portfolio.



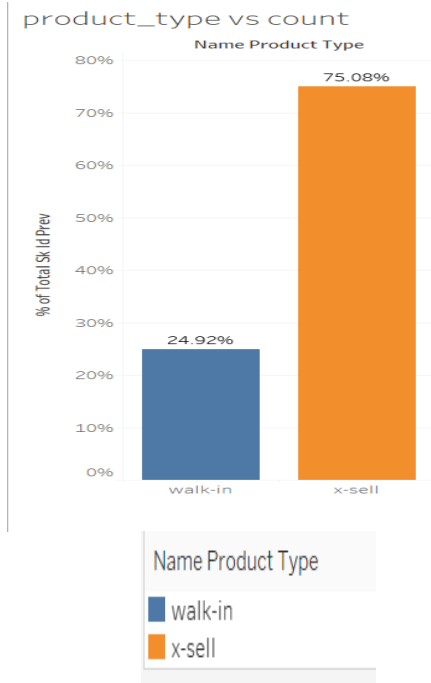


Fig. 17. Distribution of Loan Applications by Product type

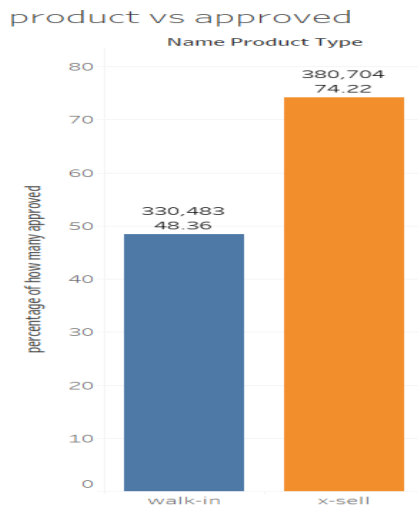


Fig. 18. Percentage of Approval by Product Type

### C. Factor 3 - Product type

Walk-in loans are applied for by customers who approach the bank independently, without any prior relationship, while x-sell loans are offered to existing customers as part of a strategy to provide additional products. The data shows that x-sell loans make up a significant majority of applications at 75.08%, whereas walk-in loans represent 24.92%(Fig. 17). X-sell loans also have a much higher approval rate of 74.22% compared to the 48.36% approval rate for walk-in loans(Fig. 18). Additionally, the average application amount for x-sell loans is 380,704, notably higher than the 330,483 average for

walk-in loans.

The higher approval rate and larger average amount for x-sell loans can be attributed to the established relationship between the bank and the customer, which reduces perceived risk and facilitates higher approval rates. Conversely, walk-in loans involve customers who are less familiar to the bank, leading to more stringent evaluations and lower approval rates. X-sell customers are often in better financial positions, contributing to their ability to apply for and secure larger loan amounts.

*Marks Used:*

- **1D mark for Bar Chart (Product Type vs Count):** Line.
- **1D mark for Bar Chart (Product Type vs Approval Percentage):** Line.

*Channels Used for Bar Chart (Product Type vs Count):*

- **Position (X-axis):** Product type (categorical variable).
- **Height (Y-axis):** Number of loan applications for each product type (count).
- **Color:** Distinct colors for each product type.

*Channels Used for Bar Chart (Product Type vs Approval Percentage):*

- **Position (X-axis):** Product type (categorical variable).
- **Height (Y-axis):** Approval percentage for each product type.
- **Color:** Distinct colors for each product type.

*Reason for Choosing Charts:*

- **Bar Chart (Product Type vs Count):** A bar chart is effective for comparing the count of loan applications across different product types. The bar height visually represents the frequency of loan applications for each type.
- **Bar Chart (Product Type vs Approval Percentage):** This bar chart allows a comparison of approval percentages for different loan product types. It gives a clear indication of which product types are more likely to be approved based on the approval rates.

### D. Factor 4 - Seller Industry

The data shows that consumer electronics receive the highest percentage of loan applications, accounting for 49% of the total, while tourism has the least at 0.06% (Fig. 19). Despite the variations in application volumes, approval rates remain high across most industries, with jewelry having the highest approval rate at 91% (Fig. 20). Even sectors like tourism, which have fewer applications, maintain high approval rates. However, MLM (Multi-Level Marketing) partners are an exception, showing a notably lower approval rate of 67% compared to other industries.

Banks likely apply standardized approval criteria across various industries, resulting in consistent approval rates despite different application volumes. Many industries, except for MLM, probably fall into similar risk categories where banks

seller industry vs count

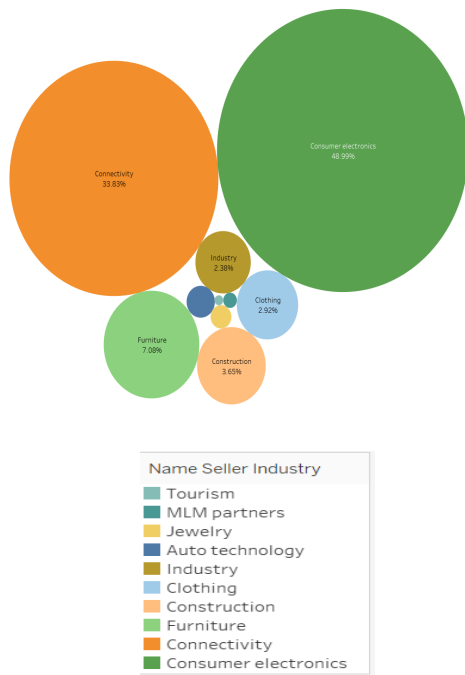


Fig. 19. Distribution of Loan Applications by Seller Industry

seller industry vs approved



Fig. 20. Percentage of Approval by Seller Industry

view them as financially stable and capable of reliable loan repayments. In sectors with fewer applications, the applicants may be more selective and meet higher qualifications, leading to higher approval rates. On the other hand, MLM businesses have a higher risk profile due to unstable revenue models and unclear business structures, which accounts for their lower approval rate.

*Marks Used:*

- **0D mark for Bubble Map:** Point.

*Channels Used for Bubble Map (Seller Industry vs Count):*

- **Size:** Number of loan applications for each seller industry (larger bubbles represent higher counts).
- **Color:** Distinct colors for each seller industry.

*Channels Used for Bubble Map (Seller Industry vs Approval Percentage):*

- **Size:** Approval percentage for each seller industry (larger bubbles represent higher approval percentages).
- **Color:** Distinct colors for each seller industry.

*Reason for Choosing Bubble Maps:*

- **Bubble Map (Seller Industry vs Count):** A bubble map is effective for visualizing the distribution of loan applications across different seller industries. The size of the bubbles indicates the volume of applications, making it easy to identify industries with higher or lower counts.
- **Bubble Map (Seller Industry vs Approval Percentage):** A bubble map is suitable for showing approval percentages for various seller industries. Despite having fewer applications, industries with higher approval rates are highlighted by larger bubbles, allowing for a clear comparison of approval rates across industries.

### E. Factor 5 - Contract Type

The data shows that cash loans make up 44.77% of loan applications, followed closely by consumer loans at 43.61% and revolving loans at 11.62%(Fig. 21). Despite being the most common, cash loans have the lowest approval rate at 42%, likely due to their unsecured nature, which increases the risk of default for banks. In contrast, consumer loans have the highest approval rate of 85.87%(Fig. 22). These loans are often tied to specific purchases and involve lower amounts and less risk, contributing to their higher approval rate. Revolving loans, which allow borrowers to access credit as needed, have a moderate approval rate of 51%. This is because, while they offer flexibility, they also carry the risk of high credit utilization and ongoing debt, leading banks to be more selective in their approval process.

*Marks Used:*

- **1D mark for Pie Chart:** Pie Slice.
- **1D mark for Bar Chart:** Line.

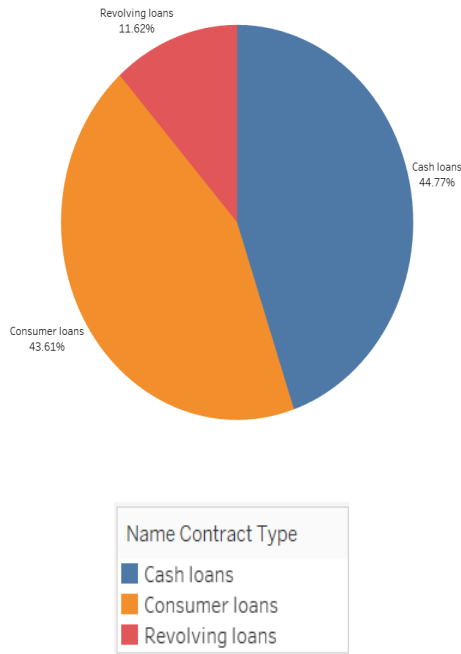


Fig. 21. Distribution of Loan Applications by Contract Type

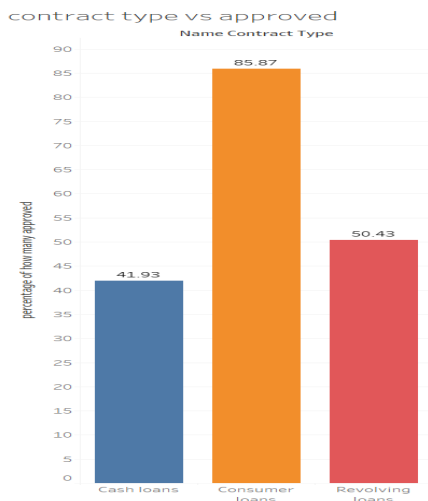


Fig. 22. Percentage of Approval by Contract Type

#### Channels Used for Pie Chart (Contract Type vs Count):

- **Angle:** Proportion of each contract type (representing the count of loan applications in the pie chart).
- **Color:** Distinct colors for each contract type.

#### Channels Used for Bar Chart (Contract Type vs Approved Percentage):

- **Position (X-axis):** Contract type (categorical variable).
- **Height (Y-axis):** Approval percentage for each contract type.
- **Color:** Distinct colors for each contract type.

#### Reason for Choosing Charts:

- **Pie Chart (Contract Type vs Count):** The pie chart is used to represent the proportion of loan applications for different contract types. It provides a clear visualization of the distribution of applications among contract types, with each slice representing the relative count.
- **Bar Chart (Contract Type vs Approved Percentage):** The bar chart effectively compares the approval percentages across different contract types. By visualizing the approval rates as bar heights, it provides a straightforward way to see which contract types have higher or lower approval percentages.

#### F. Factor 6 - Loan Purpose

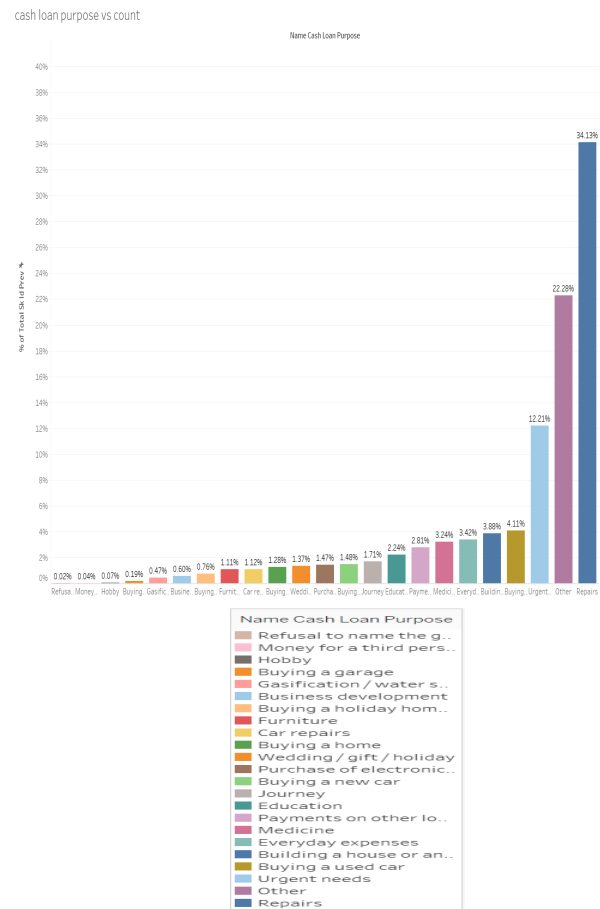


Fig. 23. Distribution of Loan Applications by Loan Purpose

The majority of loan applications are for repairs (34.13%) and urgent needs (12.21%), with urgent needs having a relatively high approval rate of 41.6% (Fig. 22). Loans for a third person, although constituting only 0.04% of applications, have the highest approval rate, while loans for electronic equipment purchases also enjoy a strong approval rate of 55%. Surprisingly, hobby-related loans, though non-essential, have a 35% approval rate, likely because they are often made by regular bank customers with established relationships. The lowest approval rate is for loans where the purpose is not



Fig. 24. Percentage of Approval by Loan Purpose

disclosed (0.02% of applications)(Fig. 23), as this lack of transparency raises concerns for banks. The high approval rates for third-party and electronic equipment loans reflect the trust and perceived stability of these applications, while the low approval rate for undisclosed purposes highlights the increased risk associated with such applications.

#### Marks Used:

- **1D mark for Bar Chart:** Line.
- **0D mark for Bubble Chart:** Point.

#### Channels Used for Bar Chart (Loan Purpose vs Count):

- **Position (X-axis):** Loan purpose (categorical variable).
- **Height (Y-axis):** Number of loan applications for each loan purpose.
- **Color:** Distinct colors for each loan purpose.

#### Channels Used for Bubble Chart (Loan Purpose vs Approval Percentage):

- **Size:** Size of the bubble represents the approval percentage for each purpose.
- **Color:** Distinct colors for each loan purpose.

#### Reason for Choosing Charts:

- **Bar Chart (Loan Purpose vs Count):** The bar chart is used to visualize the number of loan applications for each loan purpose. This is effective for showing purposes with very low counts, which may not be well-represented in a pie chart or bubble chart due to their small sizes.
- **Bubble Chart (Loan Purpose vs Approval Percentage):** The bubble chart is ideal for visualizing the approval percentage for each loan purpose. Even if some purposes have fewer applications, the size of the bubble helps highlight the high approval percentages, making it easier to see which purposes have high approval rates despite lower counts.

### G. Factor 7 - Loan Amount

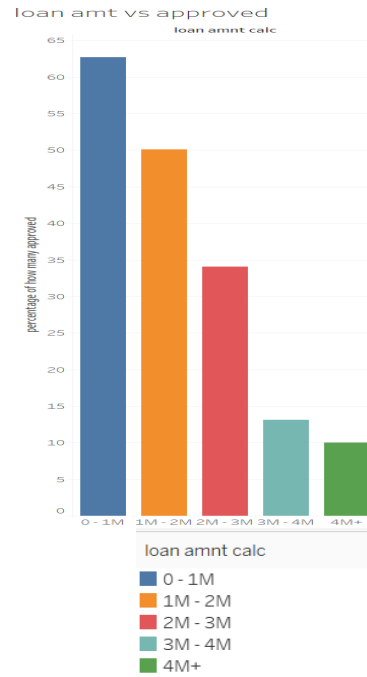


Fig. 25. Distribution of Loan Applications by Loan Purpose

The observed trend, where approval rates are highest for loans with application amounts between 0 and 1 million and decrease for larger amounts, can be attributed to several factors. Larger loan amounts present higher risk for banks, leading to more cautious approval processes. The higher the loan amount, the stricter the creditworthiness standards and collateral requirements, which fewer applicants can meet. Additionally, regulatory constraints and internal policies may limit the size of loans banks are willing to approve, requiring additional approvals for high-value loans. Economic conditions and market volatility also influence banks' willingness to approve larger loans, as they seek to manage risk and avoid potential losses. As a result, approval rates decrease for higher loan amounts due to these combined factors.

#### Marks Used:

- **1D mark for Bar Chart:** Line.

#### Channels Used for Bar Chart (Loan Amount vs Approval Percentage):

- **Position (X-axis):** Loan amount (categorical or continuous variable).
- **Height (Y-axis):** Approval percentage for each loan amount.
- **Color:** Distinct colors for different ranges or categories of loan amounts.

#### Reason for Choosing Bar Chart:

- **Bar Chart (Loan Amount vs Approval Percentage):** A bar chart is suitable for comparing approval percentages

across different loan amounts. The height of the bars clearly illustrates the differences in approval rates, and the use of color helps to distinguish between different loan amount categories or ranges.

H. Combined Factors - Loan Amt and Contract type

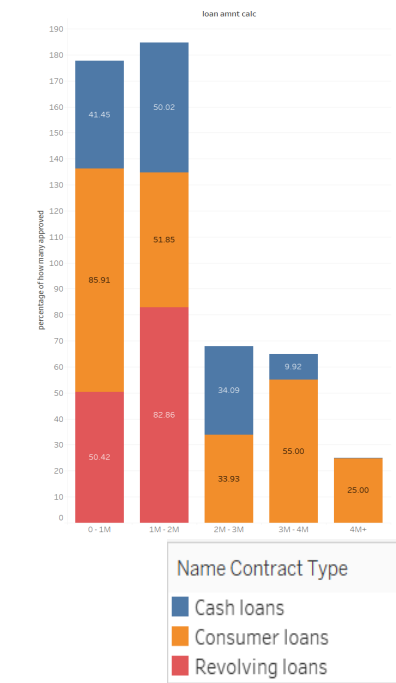


Fig. 26. Distribution of Loan Applications by Loan Purpose

In the stacked bar plot(Fig. 26), for loan amounts in the 0-1 million range, the approval percentages align with the general trend where consumer loans receive higher approval rates. However, in the 1-2 million range, revolving loans exhibit a higher approval percentage, and there are no applications for revolving loans above this amount. This shift could be attributed to the increased risk associated with larger loan amounts, where banks may become more selective. For the 0-1 million range, consumer loans are generally considered lower-risk and more manageable, leading to higher approval rates. In contrast, revolving loans in the 1-2 million category might benefit from being more manageable within this range, leading to higher approval rates. Above this threshold, the lack of revolving loan applications could indicate that banks are less inclined to offer such products for larger amounts due to increased risk and complexity.

Marks Used:

- **1D mark for Stacked Bar Plot:** Line.

Channels Used for Stacked Bar Plot (Loan Amount and Contract Type vs Approval Percentage):

- **Position (X-axis):** Loan amount (categorical variable).
- **Height (Y-axis):** Approval percentage, stacked by contract type.

- **Color:** Distinct colors for different contract types within each loan amount category.

Reason for Choosing Stacked Bar Plot:

- **Stacked Bar Plot (Loan Amount and Contract Type vs Approval Percentage):** A stacked bar plot is effective for visualizing the approval percentages of different contract types across various loan amounts. The stacking of bars allows for a clear comparison of how each contract type contributes to the overall approval percentage within each loan amount category, while the color coding helps to differentiate between contract types.

VII. T3: COMPARING TRENDS BETWEEN APPROVAL RATE AND DEFAULT RATE

This analysis explores key factors that influence loan approval and default rates by comparing historical loan data with current loan applications. The objective is to identify trends and patterns that can help financial institutions refine their loan approval criteria. The focus will be on client types, portfolio categories, seller industries, contract types, loan purposes, and borrower demographics such as gender, education level, family status, occupation, collateral, and age. The goal is to provide insights that can reduce default risks while maintaining appropriate approval rates.

A. Factor 1 - Client and Product:

X-sell has a notably lower default rate (8.06%) compared to walk-in clients, which leads to the expected result of a better approval rate. However, new applicants through the walk-in channel show a significantly higher default rate (13.967%), despite having a higher approval rate. This situation suggests that, while banks are willing to approve walk-in clients, they may be approving riskier clients with less established financial stability. This raises concerns, as approving clients with a higher likelihood of defaulting may undermine the long-term profitability of the loan portfolio.

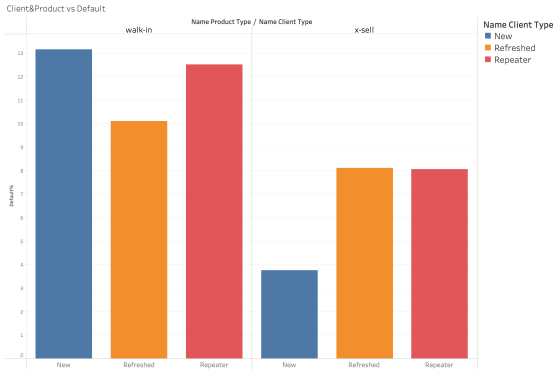


Fig. 27. Product-wise Client Default Percentage

Note that:

- By using different colors for each client, the chart ensures that each client is visually distinct, making it easy to track individual client data across multiple products.

- To efficiently use chart space while representing multiple categories (products) and keeping the comparison within a single view, avoids clutter and allows for better pattern recognition.

The contrasting approval and default rates between X-sell and walk-in clients could stem from the fact that X-sell clients, often solicited through cross-selling efforts, may have stronger financial backgrounds or established relationships with the bank. Walk-in clients, on the other hand, might represent a more varied group with uncertain credit profiles. Hence, refining the approval criteria for walk-in applicants, perhaps by introducing more stringent checks or leveraging advanced risk assessment tools, could help banks maintain a more sustainable approval-to-default ratio.

#### B. Factor 2 - Seller:

The jewelry industry (91.07%) and auto technology sector (90.65%) both show very high approval rates but also relatively high default rates (10.81% and 8.02%, respectively). This indicates that while these industries are favored in terms of loan approvals, they carry a considerable risk of default, which could potentially lead to financial losses for the lender. Conversely, MLM (multi-level marketing) partners exhibit both a lower approval rate (66.71%) and a lower default rate (7.81%), making this sector a more reliable option for loan approvals.

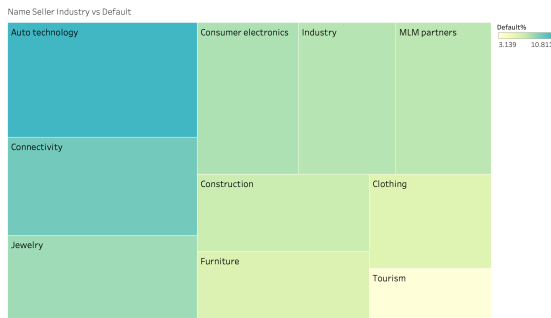


Fig. 28. Sellers vs Default Percentage

The high default rates in jewelry and auto technology can be attributed to the nature of these industries. Jewelry, being a luxury product, is susceptible to fluctuations in consumer demand and economic conditions. Auto technology, often related to high-ticket items, might attract borrowers with less stable financial positions, making repayment more challenging. MLM partners, despite their lower approval rate, may consist of entrepreneurs or individuals with more stable income streams, resulting in a more manageable default rate. It would be prudent for banks to consider shifting more approvals toward lower-risk industries like MLM to enhance portfolio stability.

#### C. Factor 3 - Collateral:

The data shows a positive correlation between collateral and loan performance: as the amount of collateral increases,

the approval rate increases, and the corresponding default rate decreases. This trend is an encouraging sign, as it suggests that higher collateral provides more security to the lender, resulting in a lower risk of default. Collateral acts as a safety net, reducing the lender's exposure to risk in the event of borrower default, and hence, loans with higher collateral tend to be approved more readily.

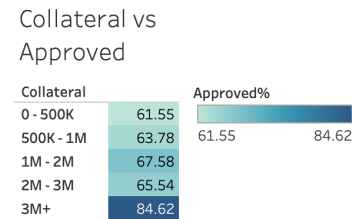


Fig. 29. Collateral vs Approval Percentage

#### D. Factor 4 - Loan Purpose:

Loans for buying a new car exhibit a low approval rate (21.82%) and a low default rate (8.72%), suggesting that this loan purpose is generally associated with lower risk and could be encouraged. In contrast, loans for car repairs and urgent needs show much higher approval rates (45.23% and 41.97%, respectively), but also much higher default rates (19.99% and 15.33%). This highlights the need for stricter approval criteria for these loan purposes, as they seem to attract higher-risk borrowers

#### Loan Purpose vs Default

Name Cash Loan Purpose	Default%
Building a house or an annex	14.73
Business development	13.21
Buying a garage	4.17
Buying a holiday home / land	11.86
Buying a home	10.72
Buying a new car	8.72
Buying a used car	13.23
Car repairs	18.99
Education	10.31
Everyday expenses	9.78
Furniture	12.14
Gasification / water supply	16.27
Hobby	26.92
Journey	11.22
Medicine	12.75
Money for a third person	13.33
Other	12.51
Payments on other loans	16.63
Purchase of electronic equipment	11.23
Refusal to name the goal	12.50
Repairs	12.80
Urgent needs	15.33
Wedding / gift / holiday	10.36

Fig. 30. Highlight table representing Loan Purpose and their Default Percentage

Note that:



- Highlight map has been used as this method condenses complex data (loan purposes and their associated risks) into a clear, visual format. It enables easy comparison across multiple categories at once, helping to identify trends or outliers in default rates without overwhelming the viewer

The discrepancy between car purchase loans and repair or urgent need loans can be attributed to the financial circumstances of the borrowers. Those taking loans for car repairs or urgent needs may be facing immediate financial pressures, which could increase the likelihood of default. In contrast, borrowers taking out loans for new cars may have more stable financial conditions, as buying a car generally requires a larger financial commitment, indicating stronger credit profiles. Banks should consider reassessing their risk criteria for repair and urgent need loans, possibly incorporating more stringent evaluations of the borrower's financial health and repayment capacity.

#### E. Factor 5 - Contract:

Revolving loans show a higher approval rate (50.53%) and a high default rate (10.539%), which suggests that these loans, despite their popularity, pose a significant default risk. The structure of revolving loans, which allows borrowers to continuously borrow and repay, may lead to poor debt management and higher default rates.

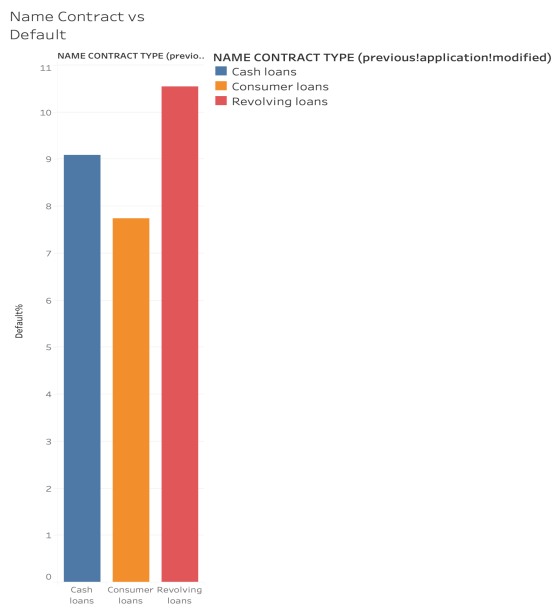


Fig. 31. Bar Chart representing Contract-type-wise Default Percentage

The flexible nature of revolving loans can tempt borrowers to overextend their credit limits, which can eventually lead to financial strain and default. These loans are often approved more easily because of their perceived flexibility and lower immediate financial burden. However, banks should consider tightening their approval processes for revolving loans by incorporating stricter debt-to-income ratios or requiring more robust credit checks. Additionally, educating borrowers on responsible credit use may help reduce the default rate.

#### F. Factor 6 - Education:

Borrowers with lower secondary (64.25%) and secondary/secondary special (62.95%) education levels show high approval rates, but also higher-than-average default rates (10.93% and 8.94%, respectively). This indicates that while these individuals are being approved for loans, they also pose a significant risk of default.



Fig. 32. Tree-map representing Approval Percentage by Education

Lower education levels are often associated with lower earning potential and less job stability, which could explain the higher default rates. Although these individuals are being approved for loans, their financial situations might not be strong enough to support consistent repayment. Banks should consider implementing more comprehensive financial assessments for borrowers with lower education levels to ensure they have the financial capacity to repay loans without defaulting.

Note that:

- A tree map visually represents the proportion of loan approvals for different occupations by size, making it easy to compare the relative approval rates across various occupations at a glance.
- The use of colors in a tree map can highlight occupations with varying approval percentages, helping users quickly identify occupations with high or low approval rates based on color intensity.
- Tree maps also utilize available space efficiently by nesting categories within a compact area

#### G. Factor 7 - Gender and Contract:

Approval rates for male and female borrowers are similar, but the default rate for males (10.14%) is significantly higher than for females (7%). Similarly, revolving loans show comparable approval rates (65.34%) to cash loans (62.62%), but cash loans have a much higher default rate.

The higher default rate among male borrowers could be linked to spending habits or financial responsibilities that differ from those of female borrowers. Revolving loans, though popular due to their flexibility, tend to have better repayment outcomes compared to cash loans, which are often larger and require more immediate repayment. Banks may want to reevaluate the approval criteria for cash loans, ensuring that the borrower's financial stability is thoroughly assessed before approval.

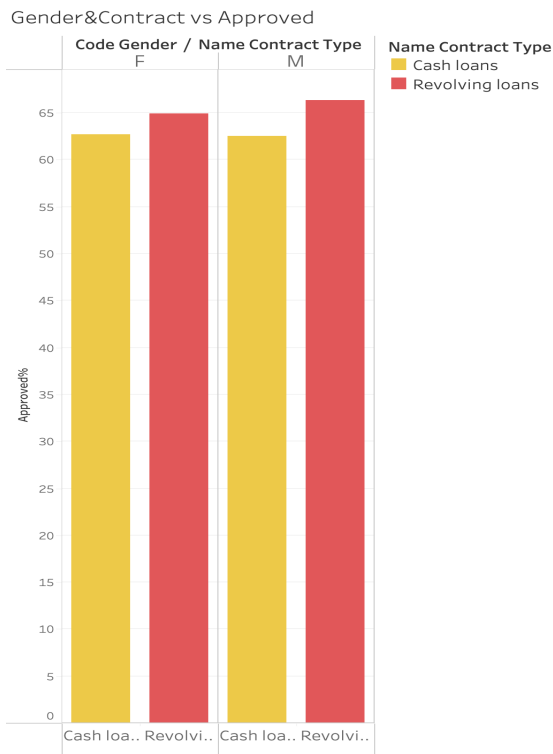


Fig. 33. Stacked Bar chart representing Approval Percentage of Contract-type by Gender

#### H. Factor 8 - Income Type:

Borrowers on maternity leave show an exceptionally high approval rate (70.00%), but also an alarming default rate (40.00%), which is far higher than any other income category. This raises significant concerns about the financial stability of individuals on maternity leave, as they may be approved for loans despite facing temporary income limitations.

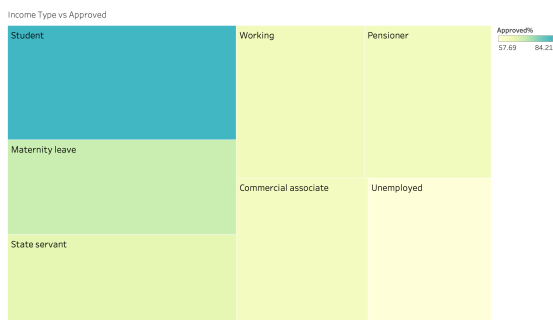


Fig. 34. Tree map to depict Approval Percentage by Income Type

The high default rate among maternity leave borrowers can likely be attributed to reduced income during the leave period, making it difficult for these borrowers to keep up with loan payments. Banks should consider implementing more cautious approval processes for this income category, possibly taking into account the borrower's long-term financial stability and the potential for income recovery post-leave.

#### I. Factor 9 - Age:

Approval rates peak during middle age, although this pattern is not reflected in the default rates. This suggests that while middle-aged borrowers may be viewed as having greater financial stability, older borrowers might exhibit better repayment behavior, despite being perceived as having less earning potential.

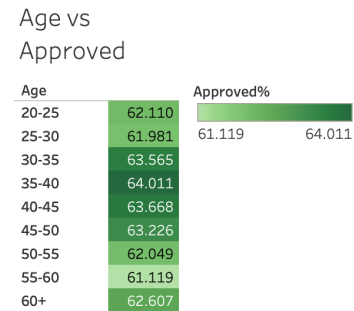


Fig. 35. Age vs Approval Percentage

Older borrowers tend to have more stable financial situations, often due to accumulated savings or reduced financial responsibilities. In contrast, middle-aged borrowers, while enjoying higher approval rates, may face more financial obligations (such as mortgages or education expenses for children), which could contribute to higher default risks. Banks should consider balancing approval criteria to reflect both earning potential and financial stability across different age groups.

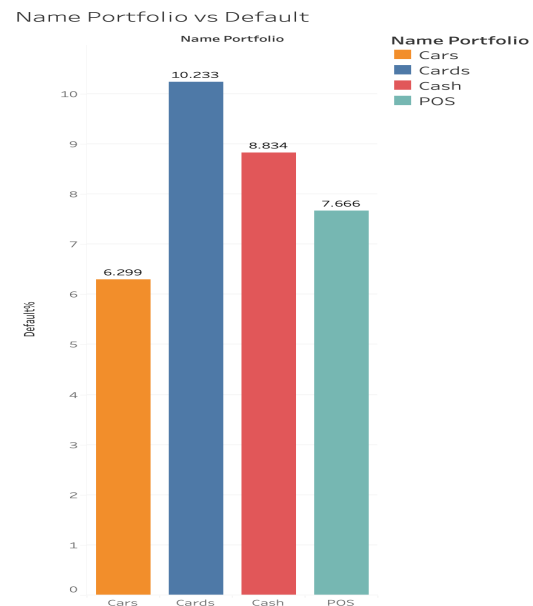


Fig. 36. Defaulter percentage by Portfolios

#### J. Factor 10 - Portfolio:

Loans for credit cards show a very high default rate (10.23%) coupled with a moderate approval rate (67.35%), suggesting that the risk of default is not adequately reflected.

in the approval criteria. Point-of-sale (POS) loans, on the other hand, have an extremely high approval rate, but still show a notable default rate, raising questions about the criteria used for approving these loans.

The revolving nature of credit card debt likely contributes to the higher default rate. Borrowers may find themselves accumulating debt over time, which becomes unmanageable, leading to defaults. POS loans, which are typically issued quickly at the point of sale, might lack the thorough risk evaluation required for sustained financial commitment, contributing to defaults despite the high approval rate. Banks should consider adjusting the approval process for both of these loan portfolios, possibly by implementing stricter credit checks or more comprehensive risk assessments before granting approval.

#### K. Factor 11 - Family Status:

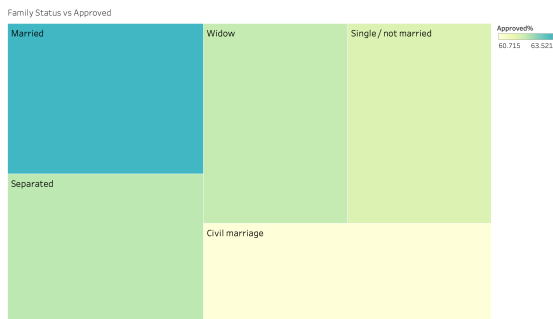


Fig. 37. Tree-map depicting Approval Percentage by Family Status

Borrowers who are civil married or single/not married exhibit higher default rates, which aligns with their lower approval rates. This finding indicates that family status plays a role in determining loan risk. Civil married and single borrowers may face unique financial challenges, such as supporting a household or managing expenses on a single income, which could contribute to their higher likelihood of defaulting on loans.

## REFERENCES

- [1] <https://www.kaggle.com/datasets/amity024/data-sources>

## VIII. AUTHORS CONTRIBUTIONS

#### A. Anirudh Pathaneni

- Conducted data analysis on the dataset, focusing on key factors such as client types, loan portfolios, seller industries, and loan purposes.
- Created visualizations (bar charts, tree maps, highlight tables) to represent trends in approval and default rates across different categories.
- Compared and contrasted approval rates with default rates to assess overall loan performance and identify areas for improvement.
- Contributed to the technical writing by documenting the analysis methodology, results, and interpretations of T3

#### B. Koushik Mupparapu

- Did Data pre-processing on the datasets
- Conducted data analysis on the dataset, focusing on key factors such as gender, type of loan, education level, family status, income type, occupation, collateral, and age
- Created visualizations (bar charts, tree maps, highlight tables) to represent trends in default rates across different categories.
- Contributed to the technical writing by documenting the analysis methodology, results, and interpretations of T1

#### C. Ananthula Harshith Reddy

- Conducted data analysis on the loan application dataset, focusing on key factors such as client types, loan portfolios, product types, and loan purposes etc..
- Created visualizations including bar charts, pie charts, bubble maps, and stacked bar plots to represent trends in loan counts and approval rates across different categories.
- Used multivariate analysis to explore the relationships between loan amounts, contract types, and approval percentages.
- Contributed to the technical writing by documenting the analysis methodology, results, and interpretations of T2