Analysing Financial Risk : Trends and Insights

Prateek Rath
IMT2022017
IIIT Bangalore
prateek.rath@iiitb.ac.in

Mohit Naik
IMT2022076
IIIT Bangalore
mohit.naik@iiitb.ac.in

Anurag Ramaswamy
IMT2022103
IIIT Bangalore
anurag.ramaswamy@iiitb.ac.in

I. Introduction

Through this assignment we have tried to analyse financial risk, and look for insights and trends, by visual exploration on the Financial Risk Dataset[1], which contains comprehensive details about the social, economical and financial conditions of different people. The dataset has the following important columns -

- **CreditScore**: Credit score (300-850)
- **AnnualIncome**: Annual income (USD)
- **LoanAmount**: Requested loan amount (USD)
- **LoanDuration**: Loan duration (years)
- **Age**: Age (years)
- **EmploymentStatus**: Employment status (Employed, Unemployed, Self-Employed)
- **MaritalStatus**: Marital status (Single, Married, Divorced, Widowed)
- **NumberOfDependents**: Number of dependents
- **EducationLevel**: Education level (High School, Associate, Bachelor, Master, Doctorate)
- **HomeOwnershipStatus**: Home ownership status (Own, Rent, Mortgage, Other)
- **MonthlyDebtPayments**: Monthly debt payments (USD)
- **CreditCardUtilizationRate**: Credit card utilization (0-1)
- **NumberOfOpenCreditLines**: Open credit lines
- **NumberOfCreditInquiries**: Credit inquiries in the last 6 months
- **DebtToIncomeRatio**: Debt-to-income ratio (0-1)
- **BankruptcyHistory**: Bankruptcy history (0: No, 1: Yes)
- **LoanPurpose**: Loan purpose (Home, Auto, Education, Debt Consolidation, Other)
- **PreviousLoanDefaults**: Previous loan defaults (0: No, 1: Yes)
- **InterestRate**: Loan interest rate (0.01-0.3)
- **PaymentHistory**: Payment history (years)
- **SavingsAccountBalance**: Savings account balance (USD)
- **CheckingAccountBalance**: Checking account balance (USD)
- **InvestmentAccountBalance**: Investment account balance (USD)
- **RetirementAccountBalance**: Retirement account balance (USD)
- **EmergencyFundBalance**: Emergency fund balance (USD)
- **TotalAssets**: Total assets (USD)
- **TotalLiabilities**: Total liabilities (USD)
- **NetWorth**: Net worth (USD)
- **LengthOfCreditHistory**: Length of credit history (years)
- **RentPayments**: Monthly rent payments (USD)
- **AutoLoanBalance**: Auto loan balance (USD)
- **PersonalLoanBalance**: Personal loan balance (USD)
- **StudentLoanBalance**: Student loan balance (USD)
- **HealthInsuranceStatus**: Health insurance status (Insured, Uninsured)
- **LifeInsuranceStatus**: Life insurance status (Insured, Uninsured)
- **CarInsuranceStatus**: Car insurance status (Insured, Uninsured)
- **HomeInsuranceStatus**: Home insurance status (Insured, Uninsured)
- **EmployerType**: Employer type (Private, Public, Self-Employed, Other)
- **JobTenure**: Job tenure (years)
- **MonthlySavings**: Monthly savings (USD)
- **AnnualBonuses**: Annual bonuses (USD)
- **AnnualExpenses**: Annual expenses (USD)
- **MonthlyHousingCosts**: Monthly housing costs (USD)
- **MonthlyTransportationCosts**: Monthly transportation costs (USD)
- **MonthlyFoodCosts**: Monthly food costs (USD)
- **MonthlyHealthcareCosts**: Monthly healthcare costs (USD)
- **MonthlyEntertainmentCosts**: Monthly entertainment costs (USD)
- **LoanApproved**: Loan approval status (0: No, 1: Yes)

II. Tasks

By visualizing data, we aim to obtain insights on the following three tasks -

1) Identifying high-risk borrowers based on financial statistics
2) Analysing the influence of socioeconomic factors on loan metrics
3) Analysing credit-related factors as key indicators of financial health

III. Assumptions and Preprocessing

The data in the dataset is fully randomly generated, and not sampled or collected. Almost all of the columns are generated based on a normal distribution without considering the effect of other columns. So, the data does not always reflect the trends one would normally see in real life, which is why some of the visualizations won't make much sense. There were some obvious outliers generated as a result of the randomness of the data, which we cleaned via some filtration. Furthermore, the number of data points was very large, so to have less cluttered visualizations, we worked on a small random subset of the data.

IV. Task 1 : Analysing high-risk borrowers based on financial statistics

This task aims to answer the question - *What are the financial characteristics of high-risk borrowers?* High-risk borrowers are individuals considered to have a greater probability of defaulting on loans or failing to meet their financial obligations. In our analysis, we define high-risk borrowers as those whose loan applications were not approved, indicated by a value of 0 in the $LoanApproved$ column. The objective of this task is to create a detailed profile of high-risk borrowers by examining their financial statistics, identifying patterns or common characteristics that may contribute to their risk status. By understanding the financial behavior and background of these individuals, we can gain insights into factors that lenders may consider when assessing loan applications and the overall risk landscape.

We first look at the impact of open credit lines on whether a loan is approved or not approved. This is shown via a line chart (Fig. 1) which shows how the percentage of unapproved loans for all people with a certain number of credit lines changes with the number of open credit lines.

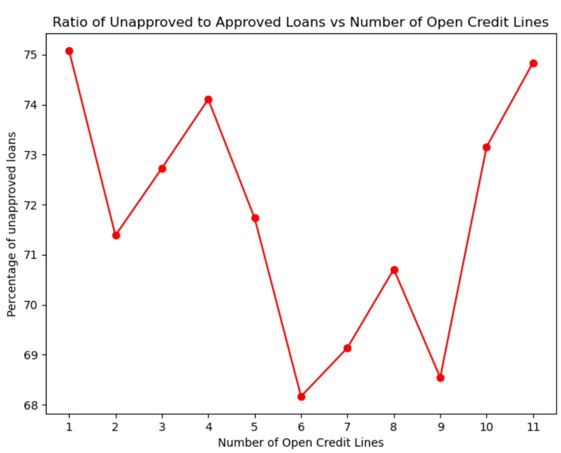

Fig. 1. Effect of open credit lines on loan approval

It is easy to see that there is no clear pattern in the percentage of unapproved loans. One thing to notice is that the percentage rises, then drops, then rises again with the exception of 8 open credit lines. Lenders typically see individuals with lesser credit lines as high-risk because they don't have enough data to assess the credit history of that person, and they may see them as inexperienced with debt and credit management. On the other hand, having too many open credit lines is an indication of the borrower being over-leveraged and too reliant on credit to manage their finances. Thus, people with a moderate number of open credit lines (6-7) seem to have the best chance of getting their loan approved.

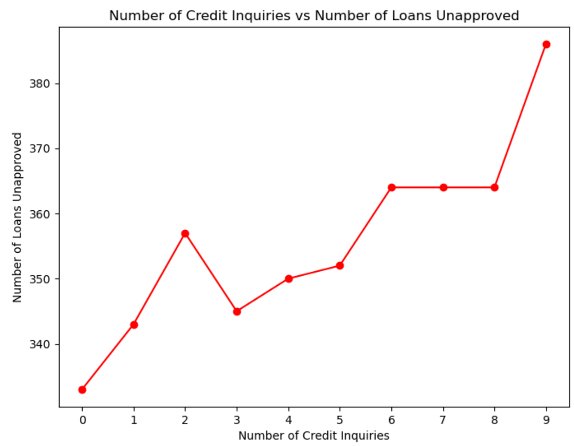

Fig. 2. Effect of credit inquiries on loan approval

Next, we look at the number of credit inquiries and their effect on loan approval. A credit inquiry is a request made by an individual or institution to check someone's credit report, typically done to assess their creditworthiness at the time of applying for a financial service. A number of such inquiries can affect your credit score, and hence your chances of getting a loan. Therefore, we expect the number of unapproved loans to increase with the number of credit inquiries, which is exactly the case as shown in (Fig. 2).

Some of the most important factors in determining loan approval are a person's bankruptcy history, history of previous defaults, and the debt-to-income (DTI) ratio. To understand this impact, we look at the number of loans approved and not approved for each of these 3 reasons via a heatmap (Fig. 3).

No loans have been approved for people who have been bankrupt or defaulted a loan in the past. For people with a high DTI ratio (over 0.75), around 87% loans have been rejected. The inference is that past debt history is a very important factor in determining your current creditworthiness, but it also excludes potentially creditworthy borrowers who are struggling due to past financial difficulties.

Note that we can also look at the visualization in another way - 156 out of 5000 people i.e. 3.12% are bankrupt, 6.44% have defaulted on a previous loan, while almost 18% have more than 75% of their income as existing debt, which tells us that a notable percentage of the population is under financial strain, and it is more prevalent than historical debt

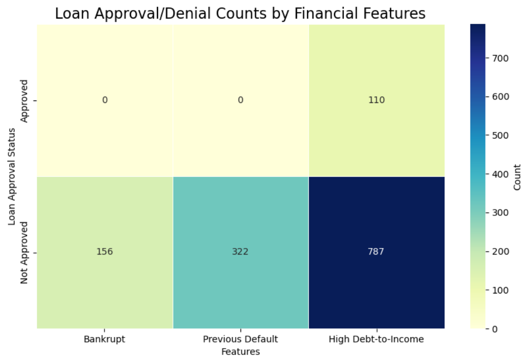

Fig. 3. Effect of previous debt on current loan approval

issues like bankruptcies or defaults.

We also analysed the average monthly costs (Fig. 4) and average account balances (Fig. 5) of people whose loans were not approved to gain insights into their spending capacities and habits.

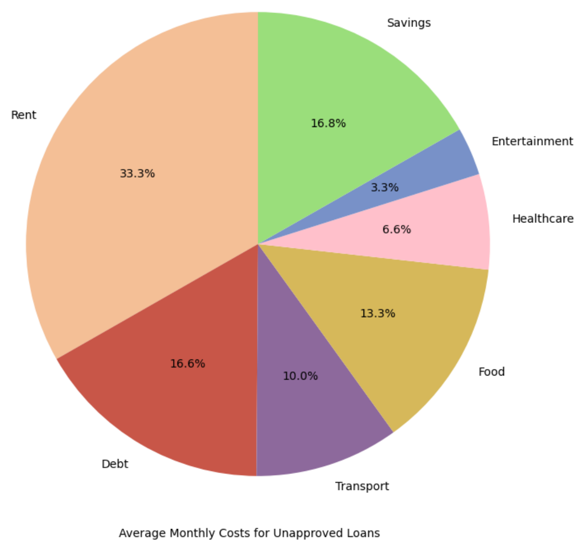

Fig. 4. Breakdown of monthly costs for high-risk borrowers

Almost half of the monthly expenses of the average individual whose loan was rejected is comprised of rent and debt payments. A significant section for savings does denote some financial priority despite loan rejection, but it is clear that investments and savings are compromised, which can lead to long-term financial insecurity. The large portion of debt may also indicate the presence of debt cycles, where a person takes debt to cover existing debt. The presence of financial constraints is also clear from the small amounts going towards healthcare and entertainment.

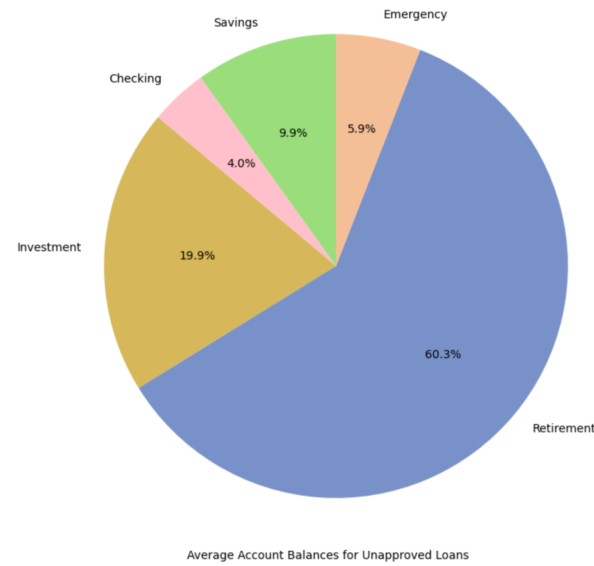

Fig. 5. Breakdown of account balances for high-risk borrowers

For existing account balances, the majority of assets are held in retirement accounts, with a notable portion allocated to investments. While this indicates sound financial planning, it also limits short-term liquidity since retirement accounts are usually not easily accessible without penalties or tax consequences. Additionally, with only 5.9% of assets set aside as emergency funds, this appears to fall short of the generally recommended 3-6 months' worth of living expenses, which could be problematic in case of unexpected financial needs.

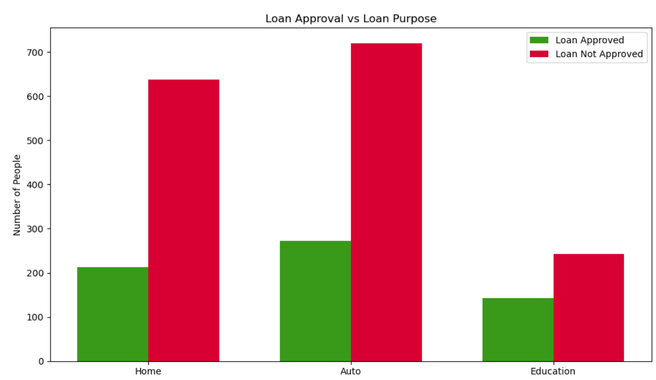

Fig. 6. Loan approval rates for various purposes

To understand what role does the purpose of the loan play in determining its approval, we looked at the number of loans approved and declined for each specific purpose, which is shown as a side-by-side bar chart in (Fig. 6). The first noticeable observation is that the number of unapproved loans is greater than the number of approved loans regardless of the purpose, which is generally true in real life as well. Furthermore, the total amount of loans taken is the highest for automobile loans, followed by home loans. Finally, the percentage of approval of a loan is the highest in education loans, probably because they require less stricter checks and usually have government backing. The future earning potential for education loans also increases the likelihood of repayment and thus lowers risk. On the other hand, automobiles are depreciating assets, which increases the lender's risk if the borrower defaults. Home loans are the least approved in terms of percentages because they are subject to rigorous scrutiny and strict regulations.

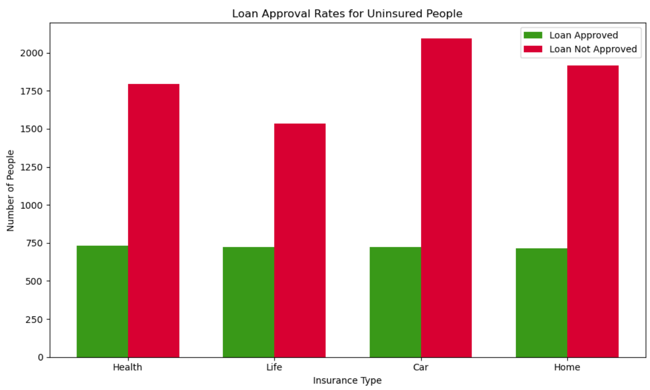

Fig. 7. Effect of insurance on loan approval

(Fig. 7) above shows the number of loans approved and unapproved for individuals who are uninsured, separated by category of the insurance. The numbers are more or less the same due to the data being sampled, but here we can draw insights by looking at the proportion of loans approved. Health and life insurance are typically not needed for most types of loans, except in cases where the borrower is seen as high risk due to high age or some health condition which can potentially increase medical expenses. Life insurance is typically also required in cases where the loan amount is very large, such as in mortgages and house loans. But the higher number of loans unapproved in these cases is most probably not due to the fact that the respective individuals are uninsured. On the other hand, car and home insurance is almost always required for most auto and home loans, respectively, which is why we see a majority of uninsured borrowers being rejected.

The loan balance, which represents the amount of outstanding previous loans, is also an important contributor in determining the approval status of the current loan. In (Fig. 8) we look at the average loan balances for 3 types of loans - Personal, Auto and Student, for both approved and unapproved loans. We see that the average student loan balance is less for approved loans as compared to unapproved loans, because lower loan balance directly contributes to a lower DTI (Debt-To-Income) ratio, which is a key factor lenders consider while lending. However, the same does not seem to apply to auto and

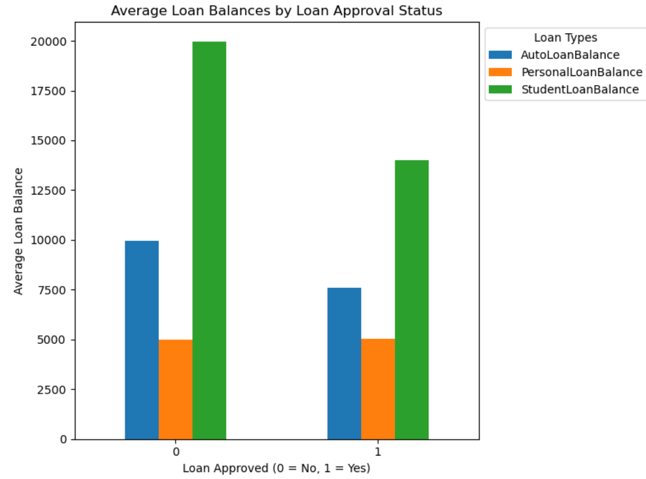

Fig. 8. Effect of previous loan balances on loan approval

personal loans. For auto loans, one possible reason is that the vehicle itself serves as security (collateral), so having a large auto loan balance is not a big problem. In case of personal loans, they can be used for a variety of reasons (consolidating debt, medical bills, vacations), and lenders may take a more holistic view of the borrower's financial profile rather than focusing solely on existing personal loan balances.

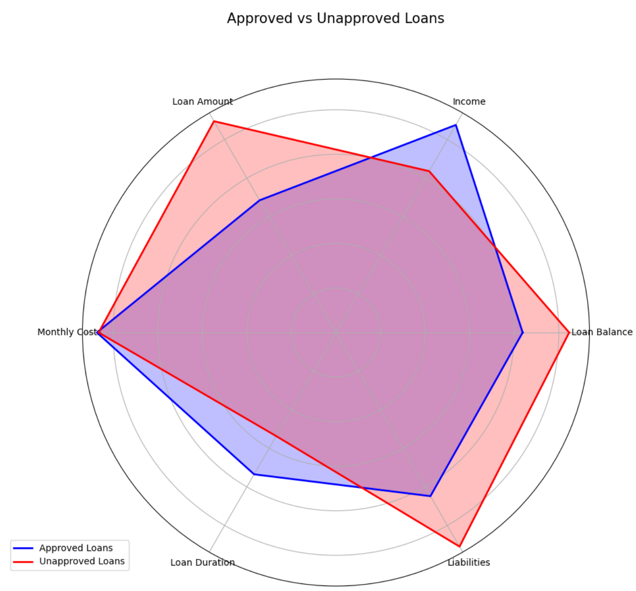

Fig. 9. Overview of multiple features affecting loan approval rates

Finally, to have a comprehensive overview of the major financial factors and their effects on loan approval rates, we look at a radar chart (Fig. 9) which shows the average values for loan amount, monthly costs, loan duration, liabilities, loan balance and income, all scaled between their minimum and maximum, for both unapproved loans (red) and approved loans (blue). From the chart we can infer the following:

1) Unapproved loans tend to have a much higher loan amount.
2) The average income of individuals whose loans were not approved is comparatively less, however the monthly costs are more or less the same.
3) They also have a higher existing loan balance, and more liabilities.
4) The loan duration is higher in case of approved loans, suggesting that loans with a lower time duration i.e. higher monthly payments may appear risky to lenders.

By looking at all these visualizations, following is the financial profile of a typical high-risk individual -

1) Credit-related factors : A low credit score, extremely low or high number of open credit lines, a large number of credit inquiries, high DTI ratio, previous defaults or bankruptcies, and possibly existing loan balances.
2) Finances - Sizeable spending on housing and debt, low liquid balance, higher liabilities, and a large loan amount
3) Miscellaneous - People with uninsured automobiles are at a higher risk of not getting their loan approved

V. Task 2: Analysing the Influence of Socioeconomic Factors on Loan Metrics

The previous section analysed financial and statistical factors that affect loan metrics. In this section, we wish to see the impact of a multiple socio-economic factors on loan metrics. At first sight, this section may seem a bit off-topic, since loans are generally approved and asked for based on statistical and financial data or requirements. However, it is important to know that as social animals, humans are always making decisions based on other metrics. Here we aim to answer the question: *How do various social factors impact loan parameters, and to what degree?*

A. Distribution of Loan Purposes By Age Category

This visualization seeks to understand the kind of people in the dataset. We seek to understand the aspirations of different age categories i.e. for what purpose does each age category want/need to take a loan?

From the stacked bar chart it is clear that people of age 24 and below take the most education loans while people of age between 25 and 40 take the most loans in general for different purposes. This seems to make sense as below 24 years of age people mostly want to invest in their education and complete their degrees. People between ages 25 and 40 are ready to, or have already started families and hence need money for automobiles, houses etc. A small fraction of people pursue higher education in this age group. There are people in the category 41-55 also who take loans for various purposes. The 'other' category here could refer to loans needed for personal expenses, medical loans, vacation loans, medical loans, appliance or furniture loans etc.

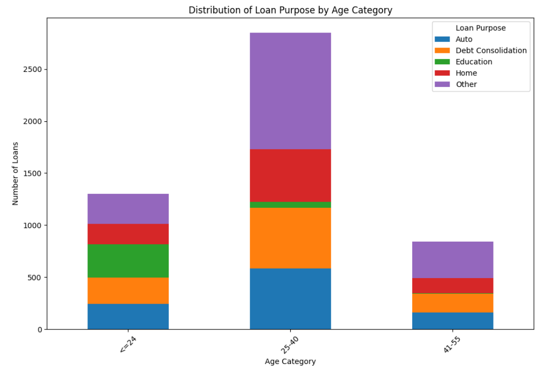

Fig. 10. Distribution of Loan Purpose by Age Category

B. Dependence of House Loan Amount on Number of Dependents

Through this visualization we seek to understand how the number of dependents impacts the Loan Amount that a family/person demands. Is it true that more dependents leave you with lesser savings and hence you need larger loans to buy a house?

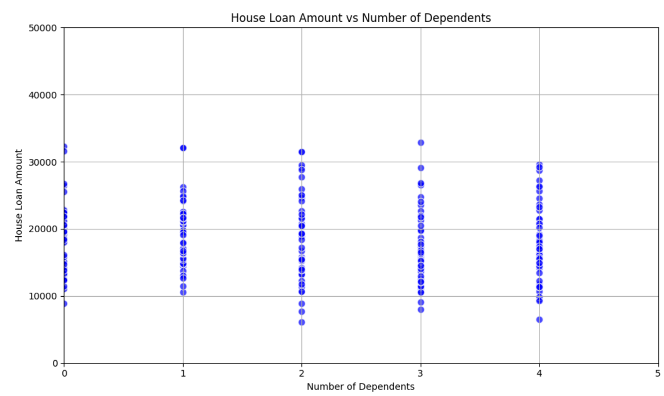

Fig. 11. Amount Spent on House Loans vs Number of Dependents

The above visualisation is a scatter plot. Every point on the plot is indicative of a family with no of dependents = the x coordinate of the point and loan amount = y coordinate of the point. From the visualisation, it seems like irrespective of the number of dependents, the mean expenditure on house loans is roughly 20000. So, through this visualization it seems as though there is no direct correlation between House Loan Amount and Number of Dependents.

C. Loan Amount vs Loan Interest by Employment Status

Through this visualization we seek to answer a few questions. For unemployed people, do lenders charge higher interest rates? Is there a threshold value of Loan Amount below which lenders are impartial towards the unemployed and the employed?

Again we have a scatter plot where each red point represents a loan asked for by an unemployed person while the blue

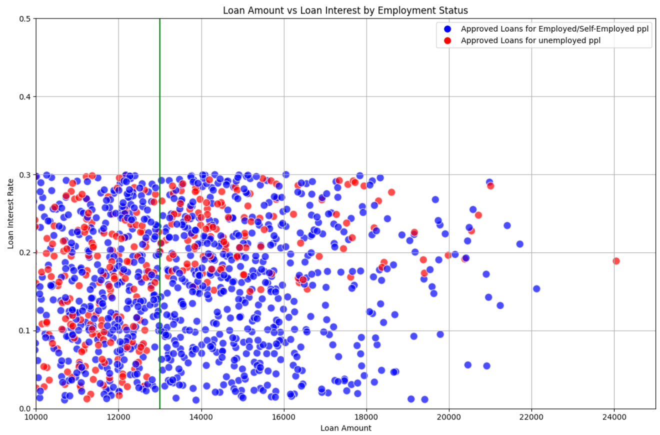

Fig. 12. Loan Amount vs Loan Interest by Employment Status

points are either employed or self-employed people. Notice the green line in the visualization. Beyond the green line, there are very few red dots that have low interest rates. It seems to be true that lenders when lending to unemployed people are taking greater risks and hence charging higher interest to compensate for the same. The threshold seems to be the 13000 mark(green line) above which lenders charge higher interest rates to unemployed people.

D. Loan Approval and its dependence on family structure

Here, we characterize a person's family structure by number of dependents and the person's marital status. We try to see if there is a relation between loan approval and a person's family structure.

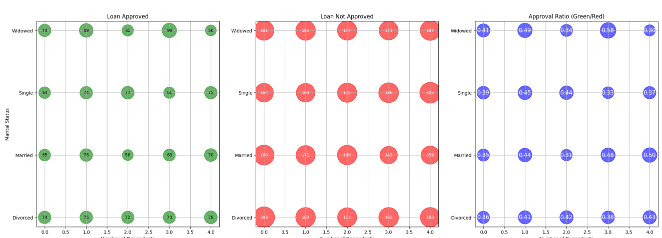

Fig. 13. Loan Approval dependence on Family Structure

The above plots are somewhat special categorical scatter plots where rather than plotting points corresponding to each individual entry in the dataset, we group them and plot one circle for each group. The size of a circle determines the number of entries belonging to the category that the circle corresponds to. By seeing the first two plots, it seems as though some types of families are preferred over others while receiving loan approval. However, upon further inspection, we see that the ratios of approved to unapproved loans is still always close to 0.5 which means that banks don't pay much attention to a person's family structure while approving loans.

E. Distribution of Loan Approvals By Employer Type

A lender might think that public sector jobs are more secure, private sector next and then self-employed followed by unemployed. Here we wish to see if the distribution of Loan Approvals has anything to do with the employer type.

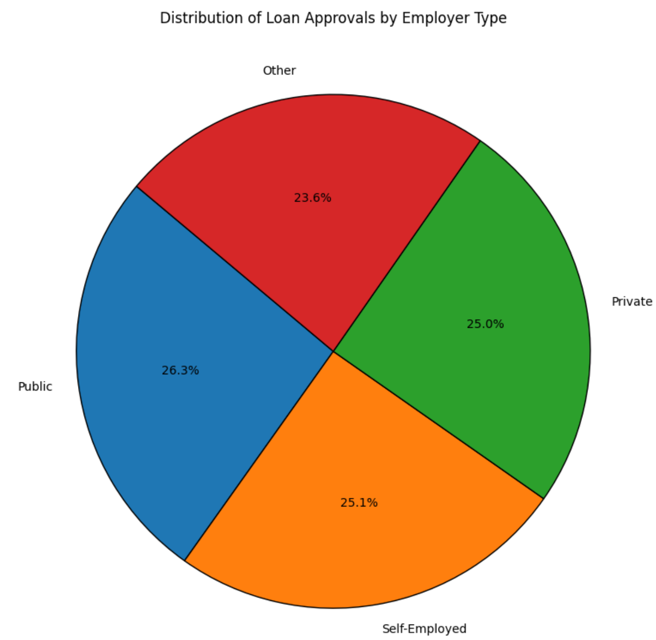

Fig. 14. Distribution of Loan Approvals By Employer Type

The pie chart shows the distribution of loan approval by employer type. Above we see that(in this dataset), loan approval doesn't seem to depend on employer type. About 25% of the loans approved are from each category.

F. Interest Rate vs Loan Duration for higher Loan Amounts

For relatively higher loan amounts, as the loan duration increases, the interest rate charged by lenders typically is higher(a compensation for the amt. of time they are kept away from their money). Turns out this isn't always the case.

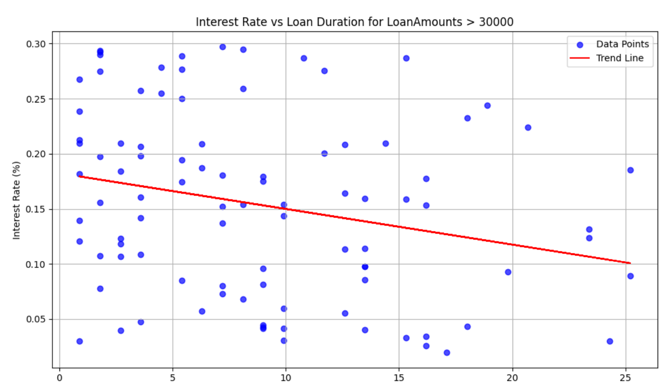

Fig. 15. Interest Rate vs Loan Duration for higher Loan Amounts(>30000)

In the above figure, the trend line is shown in red. It clearly has negative slope. The above plot shows some interesting and unconventional results for bigger loans(>30000). Ideally for short loan duration, lenders charge lower interest rates as their money is going to be returned quickly. On the other hand with loans of longer duration, due to the greater risk and delay associated, lenders charge higher interest rates. The scatter plot above shows the opposite trend line. Perhaps, lenders feel that since short duration loans are mostly paid back, they prefer to exploit this by charging higher interest rates for these whenever possible. On the other hand for riskier loans, they feel decreasing the interest rate may increase the chance of repayment.

G. Loan Approval Status for Less Educated and Unemployed Individuals

Here we try to see what percent of people in the less educated and unemployed category of people actually get their loans approved. We consider those who have completed only a high school level of education and those who have completed an associate level of education to have 'less' education.

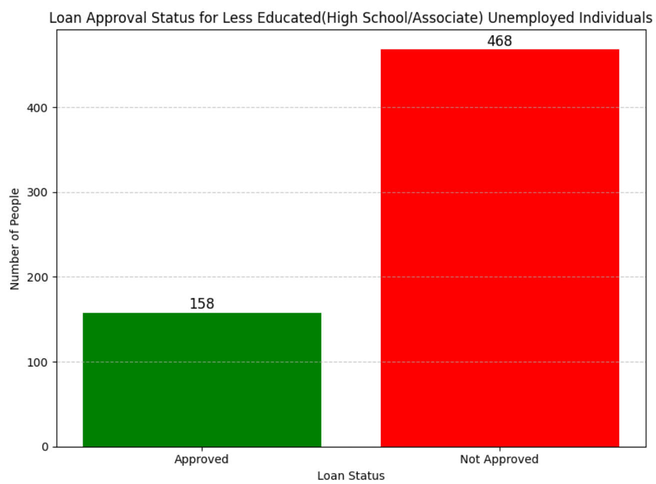

Fig. 16. Loan Approval Status for Less Educated and Unemployed Individuals(High School or Associate level)

From the above bar chart we see that out of the unemployed people, who have completed just high school or an associate level of education only a third of them get their loans approved. Thus education appears to be an important factor in determining loan approval for unemployed individuals.

H. Average Loan Duration By Loan Purpose

Each loan has different duration. This may depend on a large number of factors such as the interest rate,the loan amount and most importantly the purpose of the loan. In this section, we disregard the influence of the first two factors and focus solely on the third. In particular we answer the question: *"Which loans typically take the longest to pay off?"* Education loans, house loans, car loans or something else?

We see that education loans take the longest to pay off(14 years) while all loans taken for other purposes take about the same time(10 years). This makes sense because with education

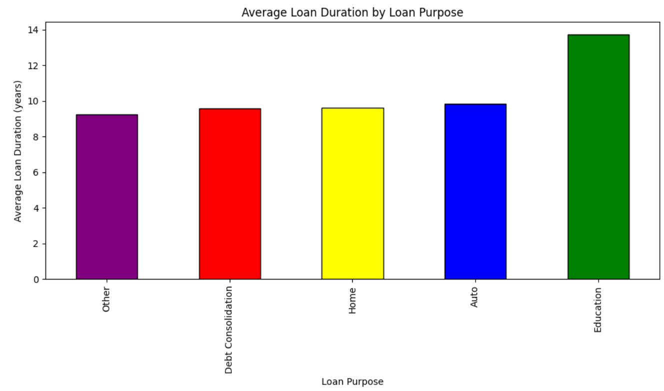

Fig. 17. Average Loan Duration By Loan Purpose

loans, students are yet to start earning and that's exactly what is indicated by the 4 year difference.

VI. Task 3 : Analysing credit-related factors as key indicators of financial health

This analysis investigates the critical question: *How do various credit-related factors impact financial health outcomes?* Through a series of visualizations, we explore how elements such as credit scores, credit card utilization rates, credit inquiries, and bankruptcy history shape financial profiles and influence decisions on loan approvals. The analysis provides a comprehensive overview of trends and patterns, offering insights into the dynamics of financial risk and stability across different demographic groups.

A. Exploring Credit Scores, Utilization Rates, and Loan Approval Trends

The first step in understanding financial health is to explore the distribution of credit scores within the dataset. A histogram (Fig. 18) reveals the concentration of credit scores across various ranges. The histogram shows a higher concentration of individuals in the low credit score range (300-499). Then there's a noticeable drop as we move to the medium credit score range (500-699) followed by another drop for high credit scores(700-850), with the least number of individuals in this category. This distribution is crucial for understanding how the population as a whole fit into different credit tiers.

A line chart (Fig. 19) delves deeper into how credit scores correlate with loan approval rates. As expected, a trend emerges: higher credit scores almost always increase the likelihood of loan approval (the slight decrease in loan approval rate from credit score 600-649 is an outlier). Borrowers with low credit scores (300-499) have a loan approval rate of 0, reinforcing the importance of maintaining a good credit score. This visualization highlights the direct impact of credit scores on access to financial resources, greatly influencing borrowing opportunities for individuals.

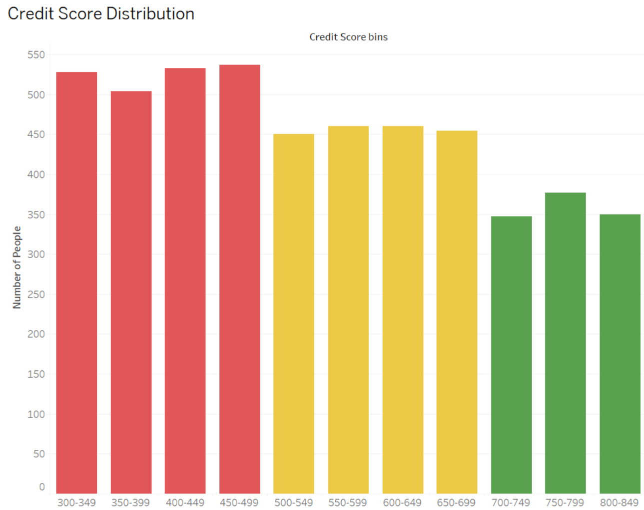

Fig. 18. Distribution of Credit Scores

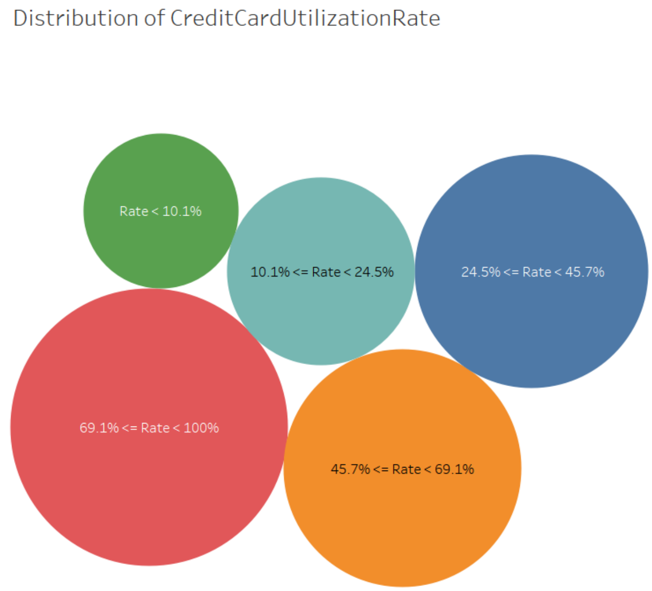

Fig. 20. Distribution of Credit Utilization Rates

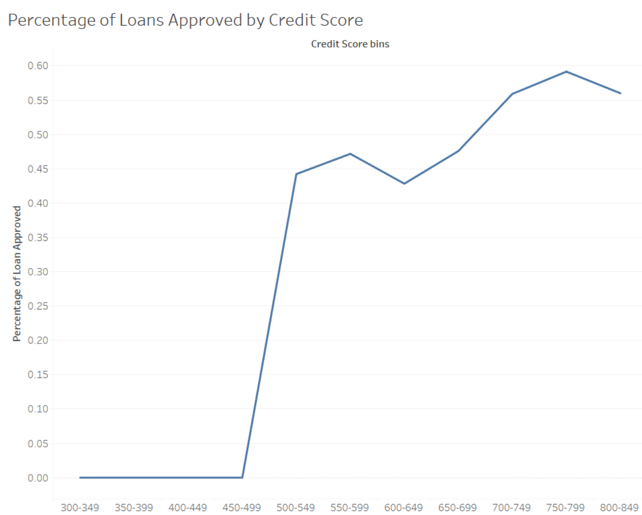

Fig. 19. Percentage of Loan Approved vs Credit Score

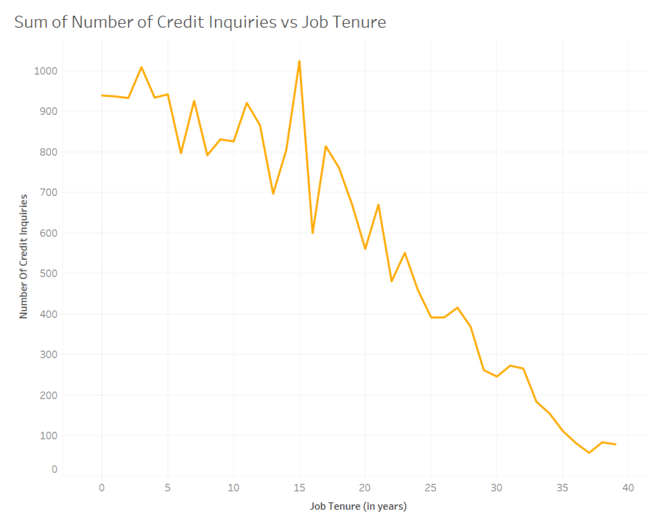

Fig. 21. Sum of Credit Inquiries vs Job Tenure

A key indicator of financial health is credit card utilization rate (also called credit utilization rate), which measures the percentage of available credit a borrower is using. A packed bubble chart (Fig. 20), highlights the number of borrowers with the different utilization rates[3].

We see that this visualization illustrates the somewhat inverse relationship between credit utilization and credit scores. The decreasing bubble sizes for lower utilization rates correspond to higher credit scores, which aligns with the observation from Fig. 18, where we saw fewer individuals with higher credit scores. This connection highlights that those with low credit utilization tend to have higher credit scores, while those with high utilization typically have lower scores.

B. Financial Behavior Across Age and Job Tenure: Debt Payments and Credit Inquiry Patterns

Job stability often correlates with financial stability, making it essential to analyse the relationship between job tenure and the number of credit inquiries. A credit inquiry occurs when a lender, financial institution, or another entity checks an individual's credit report to evaluate their creditworthiness. A line chart (Fig. 21) explores this dynamic, revealing that individuals with longer job tenure tend to have fewer credit inquiries, suggesting greater financial security. Conversely, frequent credit inquiries are more common among those with shorter job tenure, potentially indicating financial uncertainty or difficulty in accessing credit.

Age can be a determining factor in financial behaviour, particularly regarding credit inquiries. The number of credit inquiries is broken down by age group, with additional colouring to indicate employment status via a stacked bar chart (Fig. 22). Interestingly, middle-aged individuals (age groups 36-45 and 46-55) tend to have the most credit inquiries, which reflects their higher financial activity or greater need for credit during this stage of life often related to major financial decisions such as family planning, home loans, auto loans, credit card applications, educational expenses, debt consolidation, or unexpected medical costs. In contrast, younger individuals have fewer inquiries but may still face challenges in securing credit. Although the differences in credit inquiries across employment groups are not significant, it is interesting to note that self-employed individuals consistently have higher numbers of inquiries compared to those who are employed. This may be due to banks perceiving employed individuals as having more stable income and lower risk, leading to greater trust and potentially fewer credit inquiries for employed borrowers.

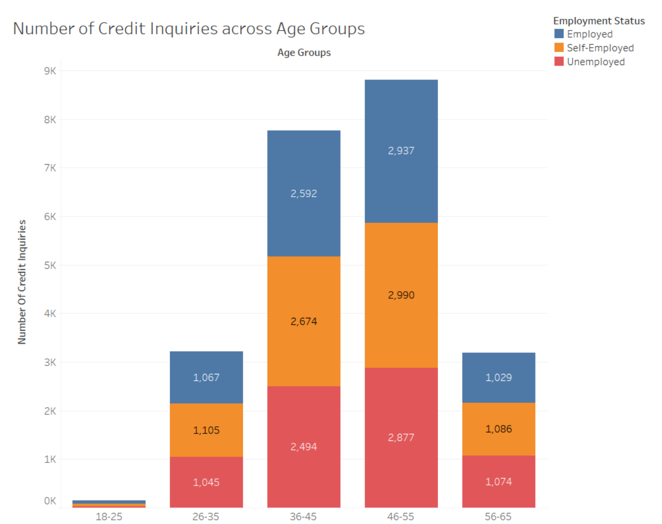

Fig. 22. Number of Credit Inquiries across Age Groups and Employment Status

Building up on the previous visualization, a line plot (Fig. 23) provides an overview of the sum of monthly debt payments across age groups, revealing that younger individuals (age ¡ 35 years) as well as older individuals (aged 55 years and above) generally have lower total debt payments. In contrast, middle-aged people face heavier monthly burdens, reflecting the increased financial responsibilities during their peak earning years. This trend aligns with the patterns observed in Fig. 5, where middle-aged individuals also showed a higher number of credit inquiries, suggesting a correlation between financial obligations and credit-seeking behavior during this life stage.

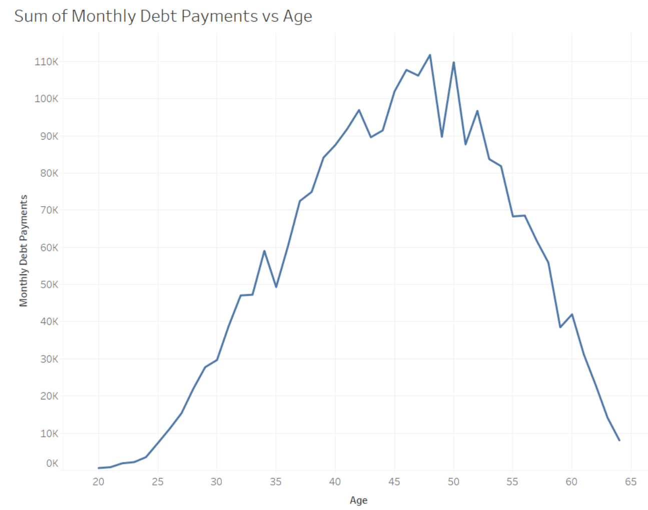

Fig. 23. Sum of Monthly Debt Payments vs Age

C. *Bankruptcy Insights: Analyzing Debt-to-Income Ratios and Age Group Dynamics*

Bankruptcy is a key indicator of financial distress, often triggered by a high debt-to-income (DTI) ratio, that is, a DTI Ratio $>$ 0.35 [2]. A horizontal bar chart (Fig. 24) compares bankruptcy history across 2 DTI groups, showing that people with high DTI ratios are more than twice as likely to have filed for bankruptcy as compared to people with lower DTI ratios. This visualization underscores the critical role of DTI in assessing a person's financial stability and highlights the tipping point where debt obligations become unsustainable.

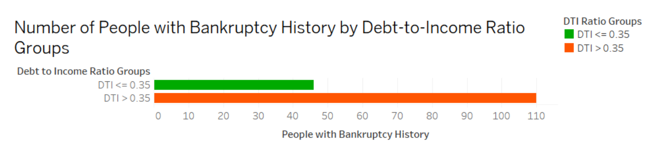

Fig. 24. Number of People with Bankruptcy history in different DTI Ratio groups

Next, a stacked bar chart (Fig. 25) attempts to find trends in bankruptcy across different age groups, with additional coloring to indicate employment status. Notably, there are no recorded bankruptcies in the 18-25 age group, likely because younger individuals have had less time to accumulate significant financial obligations. Once again similar to Fig. 22 and Fig. 23, the graph shows a clear peak in bankruptcy rates among middle-aged borrowers (age groups 36-45 and 46-55), likely due to the increased financial pressure from homeownership, loan repayments, and other significant financial commitments typical of middle age.

In conclusion, this analysis highlights the intricate relationships between various credit-related factors and their impact

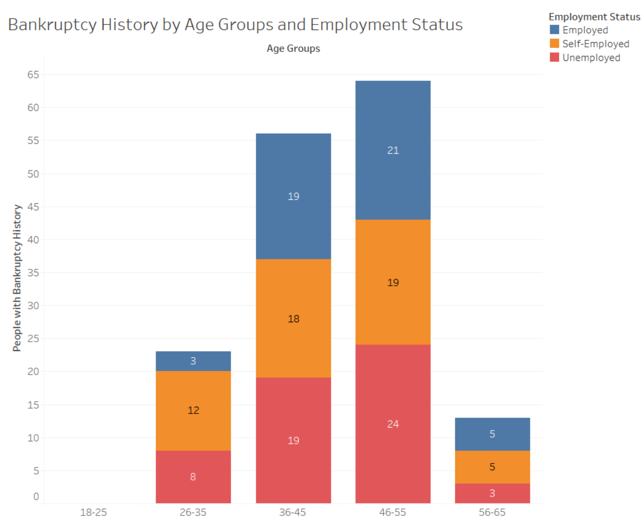

Fig. 25. Bankruptcy history by Age Groups and Employment Status

on financial health. From the distribution of credit scores to the influence of credit card utilization rates, credit inquiries, and bankruptcy history, these findings reveal significant patterns across different demographic groups. The middle-aged individuals emerges as an important group, showing higher financial activity and greater financial burdens, as proved by higher numbers of credit inquiries and monthly debt payments. Similarly, the correlations between job tenure, credit inquiries, and employment status shed light on the dynamics of financial stability, while the debt-to-income ratio and bankruptcy history emphasize the critical role of debt management in maintaining financial well-being. Together, these visualizations offer a detailed picture of how credit dynamics shape an individual's financial well-being.

VII. Visualizations

Following are the visualizations that are used and described in detail in the tasks above:

1) Scatter Plots
2) Pie Charts
3) Radar Charts
4) Heatmaps
5) Bubble Charts
6) Stacked Bar Plots
7) Horizontal Bar Plots
8) Line Plots
9) Bar Plots

In each of the types wherever applicable, we have employed various marks and channels for making the visualization more informative.

VIII. Challenges

The biggest challenge we faced was that the dataset was random. Almost all columns were normally distributed. To make sense out of this data, we had to place checks on each row to ensure that each row was semantically meaningful. For instance job tenure was sometimes greater than age. Similarly, people with extremely high DTI ratios got their loans approved. Such rows had to be removed. Despite applying such filters, we were still left with a huge dataset.This led to inevitable visual clutter in each of our visualisations which made it increasingly difficult to find insights from them. So we took a small random subset of the data, for clearly visualisations from which we could obtain meaningful insights.

IX. Author's Contributions

The tasks were initially discussed over a meeting and we came up with the tasks together. After that, we worked independently on our respective tasks and combined our findings in our report.

Mohit Naik: (Task 1: Identifying high-risk borrowers based on financial statistics)

1) Conducted research to understand what financial factors lead to higher loan rejections.
2) Analysed financial and statistical parameters of data to create appropriate visualizations.
3) Deriving insights and profiling high risk borrowers.

Prateek Rath: (Task 2: Analysing influence of socioeconomic factors on loan metrics)

1) Took note of factors that indicate a person's social and economic status.
2) Analysed factors noted above to create appropriate visualisations.
3) Derived insights and dependencies of loan metrics on social and economic factors.

Anurag Ramaswamy: (Task 3: Analysing credit-related factors as key indicators of financial health)

1) Observed the general distribution trends of credit scores and credit utilization rates and penned down key credit-related factors which impact a person's financial well-being.
2) Plotted different visualizations to analyse relations between the above factors
3) Drew inferences from the above visualizations on bankruptcy, credit inquiries and monthly debt payments across various demographics.

References

[1] Financial Risk Data - Large (Kaggle)
[2] What is a Good Debt-to-Income Ratio? — Wells Fargo
[3] What Is a Credit Utilization Rate? - Experian