Loan Data Analysis:

Insights using SQL

By Yas Jivani

ABSTRACT

This project analyses loan applicant data to derive actionable insights for enhancing the lending process. The primary objectives include understanding demographic patterns, evaluating loan approval trends, examining financial behaviours, and identifying risks. SQL-based techniques were employed to process and analyse the dataset, covering metrics such as loan approval rates, debt-to-income ratios, and correlations between income and other financial variables.

The analysis revealed patterns in income and loan behaviour, with higher-income applicants seeking larger loans and borrowers across different demographics exhibiting varying financial needs. Findings also highlighted the significance of addressing financial stress among borrowers and customizing loan offerings to meet diverse requirements.

The study emphasizes the importance of tailored loan products, pre-loan eligibility tools, and region-specific strategies to improve loan performance. By leveraging these insights, financial institutions can optimize their loan portfolios, reduce risks, and enhance customer satisfaction, ensuring a sustainable and profitable lending model.

Justification for Selecting the Loan Dataset

I chose this project and dataset because of my passion for data science and my aspiration to specialize in the finance domain. Loan data provides a rich and complex dataset that aligns perfectly with my interests, as it encompasses diverse aspects such as demographics, financial behaviours, and risk assessment. This project allowed me to explore how data science can be applied to solve real-world problems in the financial sector, such as optimizing loan portfolios, identifying risks, and enhancing customer satisfaction.

Additionally, the dataset offers an opportunity to work with critical financial metrics like debt-to-income ratios, approval trends, and income correlations. By analysing these factors, I aimed to gain a deeper understanding of customer behaviour and the factors influencing lending decisions. This experience not only strengthened my SQL and data analysis skills but also allowed me to contribute meaningful insights to a domain that plays a vital role in economic growth and individual financial stability. It was the perfect project to combine my technical expertise with my long-term career aspirations in the finance industry.

Tools and Technology used

For this report, SQL and Power BI were the primary tools used for data analysis and visualization.

- **SQL:** SQL was used for querying and extracting data from the dataset. It facilitated the analysis of key metrics, including loan approval rates, debt-to-income ratios, and demographic trends, by performing complex queries to derive meaningful insights.
- **Power BI:** Power BI was utilized to create interactive visualizations and dashboards, presenting the analysis results in an easy-to-understand format. It helped in illustrating patterns, trends, and comparisons across different parameters, making the findings more accessible and engaging.
- **Excel:** Excel was used primarily for reading and initial eyeballing of the data. It served as a tool for quickly reviewing the dataset and identifying any obvious patterns or issues before conducting a more in-depth analysis using SQL and Power BI.

Interpretation of Analysis and Business Recommendations

Distribution of an Applicants by Gender:

- The total pool of loan applicants comprises a nearly equal distribution of males and females, with a slightly higher representation of female applicants.
- **Recommendations:**
    - Since women tend to favour school loans more than men, concentrate on providing them with additional products.
    - Improve the application procedure and examine the reasons for loan denials to ensure equal opportunity for male candidates.

Loan Approval vs Rejection analysis:

- Sixty-four percent of all applicants have their applications approved and receive a bank loan, while the other applicants' applications are denied.
- **Recommendations:**
    - Implementing pre-loan eligibility tools involves offering a system or platform that enables allow applicants to assess their loan eligibility before filing a formal application.

Average Annual Income of Applicants based on the type of Work they do:

- Compared to professionals and business owners, freelancers and salaried workers earn the highest and nearly identical average yearly salaries. All of the groups fall within the same range, and the variances between them are insignificant.
- **Recommendations:**
    - Since the applicants' average salaries are comparable, the bank could provide the same type of loan to everybody, emphasizing on accessibility and ease of application for everyone.
    - Businesspeople may have varied financial needs, even if their incomes are similar. To meet their needs, the bank could issue particular business loans or flexible payback periods.

Average Rate of Interest (ROI) on Loans:

- The rate of interest is the percentage applied per period to the principal amount to determine how much interest is charged or earned on a loan or deposit.
- The average ROI for bank lending is 9.24%. It means that the bank earns 9.24% profit on every dollar it lends over a given time period, usually a year. This reflects how much money the bank earns from its loans. A 9.24% ROI indicates that the bank is making a decent profit from lending, but what that actually means depends on factors such as the cost of the bank's money, the risk of the loans, and the general market scenario.
- **Recommendations:**
  - The bank's 9.24% ROI from lending indicates good earnings, but it is essential to continuously monitor risks, modify loan rates as needed, and be informed of market developments to maintain profitability.

Debt-to-Income Ratio Analysis:

- The DTI ratio (Debt-to-Income ratio) is a measure of how much of your income goes toward debt payments. It is computed by dividing your total monthly debt payments by your monthly income and multiplying by 100 to yield a percentage.
- An average DTI ratio of 60.51% indicates that these individuals spend more than half of their monthly income on debt repayment. This is a large sum, so if they borrow more, they may have difficulty repaying it.
- This could be challenging for the bank because it shows that the customer does not have much money left after paying off their debts. Banks prefer to see a lower percentage, such as 40-45%, because it indicates that the client has adequate money to repay fresh loans with ease.
- So, with a DTI of 60.51%, it's possible that these folks may struggle to repay the loan, and the bank will be hesitant to lend them further money.
- **Recommendations:**
  - Loan limits should be lowered for applicants with high debt-to-income ratios and financial planning services should be encouraged.
  - Provide loan restructuring plans so that debtors can better manage their loans and lower their debt-to-income ratio.

Risk Analysis:

(Based on Credit Score of an applicant and Default rate of loans)

- According to risk evaluation analyses based on applicant credit scores and loan default rates, the majority of applicants have a moderate degree of risk, while some of them are classified as low or high risk, with high-risk applicants being the most troubling to lenders.
- People at the *higher level of risk* are the riskiest for the banks. They may have bad credit or a significant risk of not repaying the loan. Lenders may be hesitant to lend to them or charge much higher interest rates.
- The applicants with *moderate risk* are not too risky However, they are responsible for the risk of failing to repay the loan.
- People with the *lower level of risk* are the safest borrower. They have good credit and are quite likely to repay the loan amount.
- **Recommendations:**
  - Adjust interest rates according to risk to maintain profitability while reducing de faults.
  - Provide high-risk consumers with financial literacy training to help them avoid defaults.

Good Loans vs Bad Loans:

- Simply defined, *good loans* are manageable for the bank and accessible for the borrower, but *bad loans* harm one's financial health. Based on the data, it is good to observe that the majority of total loans are good, with fewer bad loans.
- **Recommendations:**
  - Banks can improve the overall stability of their loan portfolio by enforcing stricter criteria in high-risk areas and implementing efficient strategies in low-risk jurisdictions.

States with the Highest Percentage of Good Loan Applications

- *Nebraska* (NE) and *Maine* (ME) have the highest percentage of good loans, with all applications approved, but they have fewer applicants. *Tennessee* (TN), *Vermont* (VT), and *Hawaii* (HI) also have mostly good loans, but a few loans were bad loans. Overall, most loan applications in these states are approved as good, with only a few bad ones.

- **Recommendations:**
    - *Nebraska* (NE) and *Maine* (ME) have a high percentage of good loans, even though they have fewer applicants. The bank could analyse their lending processes and apply successful strategies from these states to other regions with higher risk.
    - Since states like *Tennessee* (TN), *Vermont* (VT), and *Hawaii* (HI) have mostly good loans, the bank could boost marketing and loan offerings in these areas to take advantage of the lower risk.

States with the Highest Percentage of Bad Loan Applications

- *Iowa* (IA) has the lowest proportion of good loans, with about two-thirds of applicants approved, suggesting higher risk. *Idaho* (ID) and *Delaware* (DE) have better approval rates, but still only about three-quarters of loans are good, indicating moderate risk. The bank may need to apply stricter criteria in these states to reduce the number of bad loans.
- **Recommendations:**
    - In states like *Iowa* (IA), *Idaho* (ID), and *Delaware* (DE) where there are majorly bad loans, the bank should use stricter lending rules or better risk assessments to reduce these loans.
    - Provide borrower education programs in these regions to mitigate risks and improve repayment behaviours.

Gender-Based Distribution of Loan Applications Across Different Purpose:

- Overall, home loans are the most common kind, with the other loan kinds falling into the same range. Home loans are the most preferred by both men and women. Since purchasing a home is a major aim that provides long-term stability, home loans are likely the most popular option for both men and women. Because they have longer repayment periods and lower interest rates than other loans, they may also be perceived as being less expensive.
- However, female applicants' education loans over male applicants. Personal and vehicle loans have approximately equal numbers of male and female applicants.

- **Recommendations:**
  - Since house loans are the most popular among both men and women, the bank should keep affordable rates and flexible terms to attract more applications while pointing out the long-term stability it provides.
  - Female prefer educational loans more than males, so the bank should offer specially developed education loan products with features such as reduced interest rates or additional financial assistance to better satisfy their demands.

Gender Comparison of Loan Approval and Denial Rates

- Females have a slightly greater approval rate than males, although this difference is not significant. Similarly, females show slightly higher rates of denial than males, but the difference is small. Overall, both genders exhibit identical acceptance and denial behaviours, with no variance between them.
- **Recommendations:**
  - The bank should continue its current lending procedures because the acceptance and denial rates for both genders are identical, with just small variances.

Overview of Loan Types: Popularity, Terms, and ROI Comparison

- Home loans have the highest number of borrowers, followed by personal and vehicle loans, with education loans having the fewest borrowers.
- Vehicle loans have the longest repayment terms, just over 10 years, while home, personal, and education loans have slightly shorter terms, all just under 11 years.
- In terms of interest rates, vehicle loans offer the lowest, followed by personal loans, while home loans have the highest rates, with education loans being very similar to home loans.
- Home loans are the most popular, offering long terms and slightly higher interest rates. Vehicle loans, though less common, have the longest terms and the lowest interest rates, making them ideal for larger purchases. Education loans, with fewer borrowers, offer competitive terms and rates, though slightly higher than those for vehicle loans.
- **Recommendations:**
  - To grow the business, focus on home loans since they are the most popular, offering lower rates or better terms. For vehicle loans, emphasize their long

repayment periods and low interest rates to attract customers buying expensive items. Education loans could be promoted more to reach more borrowers, offering competitive rates and better awareness.

Occupation-wise Analysis of Income, Expenses, and Loan Patterns

- Freelancers have the highest income and request the largest loan amounts. They have a slightly higher number of transactions than other occupations.
- Salaried people earn around the same as freelancers and spend nearly the same amount of money each month. Their existing and proposed loan amounts are slightly lower than freelancers', and they conduct fewer transactions on average.
- Professionals make less than freelancers and salaried employees, yet they have similar monthly expenses and loan amounts. They also handle a similar number of transactions per month.
- Business owners have the lowest monthly income and expenses. Their loan amounts and loan requests are similar to those of other occupations, with a somewhat lower transaction frequency.
- Overall, all occupations have similar annual income, with freelancers earning a bit more. Monthly spending and loan amounts are very similar across all groups. Freelancers and professionals have slightly higher transaction frequency than salaried and business persons. The loan amounts requested are identical, with freelancers demanding the most.
- **Recommendations:**
    - Freelancers, with their highest income and loan requests, could be offered tailored loan with higher limits and flexible repayment options to attract this high-earning group. Similarly, salaried persons and professionals with similar incomes but lesser loan requests could be encouraged to borrow more through competitive loan arrangements, such as lower interest rates or longer repayment periods.
    - Although earning slightly less than other, business owners request loans in identical quantities. To encourage growth and attract new company clients, the bank could offer business-specific loans with additional benefits such as reduced interest rates or greater loan limits.
    - Although income and loan requests are similar across groups, the bank should monitor the average level of debt owed by customers. It is crucial to ensure that

the loans offered are accessible and manageable, especially considering that borrowers in all occupations currently have relatively significant loan amounts.

Analysis of Applicant Distribution by Residential Status and Marital Status

- The majority of applicants have their own residences, and they are married.
- People who live on rent are fewer, but they follow a similar pattern, with married applicants being the most common.
- In the United States, the "Other" group often consists of people living in non-traditional housing conditions, such as temporary housing, with family or friends, or in institutions. This category has the fewest applicants, with the majority of them married, while single candidates are the second most common.
- **Recommendations:**
    - Since most applicants are married, the bank could offer family-focused financial plans, including joint income-based loans, flexible terms, and affordable options for both homeowners and renters. Providing special terms like lower down payments for married renters could also boost applications from this group.
    - The bank could provide flexible lending choices with favourable terms for applicants seeking non-traditional or temporary housing, particularly married individuals with steady resources in the smaller "Other" category.

Applicant Distribution by Marital Status and Number of Dependents

Single applicants with 0 to 2 dependents are in a similar range, but there is significant variation when compared to those with 3 dependents, as the latter group has the fewest applicants.

- However, for married candidates, the situation is reversed because they have a family, therefore most of them fall into the category of three dependents rather than zero or one.
- As the number of dependents increases (from 0 to 3), the number of divorced applicants decreases, indicating that the data contains fewer divorced people with more dependents.
- **Recommendations:**

- The bank should focus on offering customized loans for single applicants with 0 to 2 dependents, as this group is larger. Loans for married applicants could be organized to help families with three dependents.
- Additionally, the bank might also need to provide flexible choices for the smaller percentage of divorced applicants who having dependents.

Loan Application Trends Among Different Occupations

- The number of people seeking for secured and unsecured loans is fairly similar across all occupations, with just minor variations. Professionals have the most applicants overall, while Business owners have the fewest. Overall, most people in all job types are applying for both types of loans in almost equal amounts, though there are small differences depending on the occupation.
- **Recommendations:**
  - Since applicants across all occupations are applying for both secured and unsecured loans in nearly equal amounts, the bank should continue to offer both types of loans and ensure that neither type is overlooked.

Education Level and Employment Status of Loan Applicants

- Graduates and high school graduates have similar job patterns. Most of them are employed, which shows they have full-time or part-time jobs. Additionally, many of them are self-employed, running their own businesses or working as freelancers. Postgraduates, on the other hand, have fewer applicants overall, but most of them are employed, with a smaller group working for themselves as self-employed individuals.
- Across all education levels, there are fewer unemployed applicants, which means that fewer people in these groups are without a job. However, Graduates stand out because they have the most applicants overall, meaning more graduates are applying for loans than those with lower or higher education levels.
- **Recommendations:**
  - The bank should ensure loan options are accessible to both employed and unemployed applicants. For those with stable jobs, loans should be flexible and easy to obtain. At the same time, the bank should consider offering special loan

options for the unemployed, with relaxed requirements or alternative ways to prove financial stability, if possible and practical.

Co-Applicant Trends in Secured and Unsecured Loan Applications

- More people apply for secured loans with a co-applicant, meaning they ask someone to join the application. However, for unsecured loans, most people apply on their own without a co-applicant. This suggests that people are more likely to have a co-applicant for secured loans but prefer to apply on their own for unsecured loans.
- **Recommendations:**
    - The bank should promote co-applicant inclusion in unsecured loans to enhance approval rates.
    - By offering co-applicant benefits, such as lower interest rates or higher loan limits, the bank can encourage more applicants to include a co-applicant, making loans more accessible and attractive.

Correlation Analysis of Annual Income with Other Financial Variables

Correlating annual income with monthly expenses, loan amount requested, and credit score is carried out to determine how an individual's income and financial behaviour are related. It also provides insights into how income influences financial decisions, such as spending habits, loan requests, and creditworthiness.

1. **Annual Income and Monthly Expense**
   - **r = 0.01:** There is very weak correlation of 0.01 between Annual Income and Monthly Expense
   - **Interpretation:** There is almost no connection between income and monthly expenses. This implies that a person's monthly spending is not much influenced by their income. It doesn't seem that a person's spending patterns are significantly impacted by their income.
   - **Recommendation:** Given the weak association between income and monthly expenses, the bank should consider additional factors such as lifestyle, personal choices, or financial behaviour when offering loans and other related services.

2. **Annual Income and Loan Amount Requested**
   - **r = 0.85:** There is strong positive correlation of 0.01 between Annual Income and Amount requested by them for Loan.
   - **Interpretation:** There is a high positive correlation between annual income and loan amount requested, indicating a direct relationship between the two. As a person's annual income increases, they also increase the amount of loan they want. People with greater earnings are more likely to ask for larger loans, possibly because they wish to acquire more expensive items such as homes or cars, or because they have the financial means to repay larger debts. This significant association indicates that income is an important factor in deciding how much loan a person is going to apply for.
   - **Recommendation:** Since there is a significant association between income and loan amount requested, the bank should use income as an important factor in deciding loan approvals and customizing loan packages.

3. **Annual Income and Credit Score**
   - **r = 0.65:** There is moderate positive correlation of 0.01 between Annual Income and their Credit score.
   - **Interpretation:** The correlation between annual income and credit score is relatively favourable, showing a relationship, but not a perfect one. People who earn more income typically have higher credit scores, most likely because they can better manage debt and have the financial stability to repay their loans.
   - **Recommendation:** Given the moderate positive correlation between annual income and credit score, bank should recommend to create financial literacy programs for individuals with moderate incomes. Since higher income often leads to better credit scores, these programs can help customers understand how to manage debt, improve their credit scores, and increase their borrowing potential. This will not only empower customers to make smarter financial decisions but also help the bank build stronger, long-term relationships with them, leading to greater lending opportunities.

FINDINGS

The analysis revealed that loan applicants are evenly distributed between genders, with females slightly favouring educational loans and both genders predominantly opting for home loans. Sixty-four percent of applicants receive loan approval, while others face rejection due to unmet criteria. Higher-income applicants tend to request larger loans, demonstrating a strong correlation between income and loan amounts. However, a high average debt-to-income (DTI) ratio of 60.51% poses repayment challenges, potentially limiting additional borrowing opportunities. Risk analysis categorized applicants into low, moderate, and high-risk groups, with high-risk borrowers being the most concerning.

Regionally, states like Nebraska and Maine exhibit higher proportions of good loans, whereas Iowa and similar states face significant lending risks. Home loans are the most popular, while vehicle loans stand out for their low interest rates and long repayment terms. Furthermore, income showed strong positive correlations with loan amounts and moderate correlations with credit scores, highlighting its importance in financial behaviour and lending decisions.

BUSINESS RECOMMENDATION

To address these findings, financial institutions should offer tailored products such as specialized educational loans for female applicants and ensure equal opportunities for male applicants by reviewing loan denial reasons. Pre-loan eligibility tools should be implemented to reduce rejections, while risk-adjusted interest rates and financial literacy programs can help manage high-risk borrowers. Successful lending strategies from low-risk states like Nebraska and Maine should be replicated in higher-risk regions like Iowa, alongside targeted marketing efforts in safer regions.

Customized loan options should cater to business owners and high-income borrowers, while loan restructuring and financial planning services can assist applicants with high DTI ratios. Promoting home loans for stability and vehicle loans for their favourable terms can enhance market penetration, while educational initiatives aimed at improving credit scores can empower moderate-income groups and foster sustainable lending practices.

CONCLUSION

This project provided valuable insights into loan application data, highlighting key trends such as loan approval rates, debt-to-income ratios, and demographic influences. By using SQL and Power BI, we identified the need for tailored loan products and pre-loan eligibility assessments. The findings emphasize the importance of data-driven decision-making to optimize loan portfolios, reduce risks, and improve customer satisfaction. Overall, the project demonstrates how data science can enhance lending practices and drive smarter decisions in the finance sector.