LOAN APPROVAL PROCESS

Introduction

In the dynamic landscape of banking and finance, the optimization of loan approval processes remains a critical challenge for institutions. Inefficient or inaccurate decision-making can result in increased defaults, leading to financial losses, or missed opportunities, thereby limiting growth potential. A robust and efficient loan approval process is essential to balance the risks and benefits, ensuring that deserving borrowers receive timely financial support while minimizing the likelihood of defaults.

This project is undertaken to address these challenges using a data-driven approach. By analyzing a comprehensive dataset of loan applicants(Home Loan, Car Loan, Educational Loan, Personal Loan) , we aim to uncover key factors influencing loan approval decisions and provide actionable insights to optimize the approval process. Our analysis focuses on several primary and secondary objectives to offer a holistic view of the factors at play and to propose improvements to the current system.

Objectives

Primary Objectives

1. **Identify Influential Factors in Loan Approvals:**
   - Examine the dataset to determine which variables have the most significant impact on loan approval decisions. This includes assessing credit scores, income levels, employment status, and other relevant factors.
2. **Enhance the Loan Approval Process:**
   - Develop recommendations to refine and improve the current loan approval workflow. The goal is to make the process more efficient and accurate by incorporating data-driven insights.
3. **Risk Profile Evaluation:**
   - Assess the risk associated with different applicant profiles. This involves identifying high-risk groups and suggesting strategies to mitigate these risks, ensuring better-informed decision-making.

Secondary Objectives

1. **Analyze Demographic Influences:**
   - Investigate how various demographic factors, such as age, gender, and marital status, influence loan approval rates. This analysis aims to uncover any demographic trends or biases in the approval process.
2. **Evaluate Financial Health Metrics:**
   - Study the impact of applicants' financial stability, including their income, monthly expenses, and savings, on loan approval outcomes. Understanding these metrics will provide insights into the financial behaviors that affect loan decisions.
3. **Assess Asset Contributions:**
   - Explore the role of different types of assets, like residential properties, commercial assets, and luxury items, in securing loan approvals. This analysis will highlight the importance of collateral and asset value in the decision-making process.

Data Preprocessing

STEP 1- To facilitate our exploratory data analysis and statistical analysis, we have created several new columns based on existing ones. These new columns categorize the data into meaningful bins, allowing for more detailed insights and comparisons.

1. **Age Group:** We categorized the ages into distinct groups to better understand the distribution and impact of different age brackets.
    - < 20 years
    - 20-30 years
    - 30-40 years
    - 40-50 years
    - 50-60 years
    - >= 60 years
2. **Income Buckets:** Based on annual income, we created bins to differentiate between various income levels.
    - 0-5 Lakhs
    - 5-10 Lakhs
    - 10-15 Lakhs
    - 15-20 Lakhs
3. **Monthly Salary:** Aggregated monthly salary data from the annual income column to facilitate monthly financial analysis.
4. **DTI** (Debit to Income Ratio): DTI which is a financial metric that calculates the percentage of your monthly gross income that goes towards paying off your debts, like rent, mortgage, credit card payments, and car loans; it's a key factor lender consider when evaluating your ability to repay a loan.
5. **Savings Bins:** We categorized the savings account balances into bins for a clearer understanding of savings distribution.
    - 0-1 lakh
    - 1-2 Lakhs
    - 2-3 Lakhs
    - 3-4 Lakhs
    - 4-5 Lakhs
    - >= 5 Lakhs
6. **Investment Bins:** Created bins for investment portfolio values to analyze the impact of investments on loan approvals.
    - 0-2 Lakhs
    - 2-4 Lakhs
    - 4-6 Lakhs
    - 6-8 Lakhs
    - 8-10 Lakhs
7. **Insurance Bins:** Categorized insurance policy values to understand their role in the loan approval process.
    - 0-10 Lakhs
    - 10-20 Lakhs
    - 20-30 Lakhs
    - 30-40 Lakhs

- 40-50 Lakhs

8. **Loan Amount Bins:** The loan amounts were divided into the following categories for detailed analysis.
    - 0-1 crore
    - 1-2 crore
    - 2-3 crore
9. **Term Bins:** The loan terms in years were categorized to understand the impact of loan duration on approvals.
    - 5-10 years
    - 10-15 years
    - 15-20 years
    - 20-25 years
    - >= 25 years
10. **CIBIL Score Bins:** The CIBIL scores were grouped into categories to analyze their influence on loan approval decisions.
    - 600-649: Doubtful Score
    - 650-699: Satisfactory Score
    - 700-749: Good Score
    - 750-900: Excellent Score

STEP 2- The dataset is dumped into python for visualization, EDA and statistical testing.

STEP 3- In data preprocessing and cleaning on a loan application dataset to ensure the data integrity and plausibility of the scenarios within it, We calculate the remaining salary for each applicant after their monthly financial obligations are subtracted from their monthly salary. This new metric helps us understand the financial buffer each applicant has.

- We remove rows where the applicant has a negative remaining salary, is unemployed, and the loan is approved. This is to ensure that applicants who cannot realistically support a loan are not marked as approved in the dataset.
- Additionally, we remove rows where the applicant is unemployed, but the loan is still approved. This helps maintain the credibility of the dataset by ensuring that unemployed applicants are not granted loans.
- After filtering the data, we drop the Remaining_Salary column since it is no longer needed for further analysis.
- Finally, we display the first few rows of the updated DataFrame to verify that our filtering and cleaning operations have been correctly applied.

This process helps in refining the dataset to make sure it only includes realistic and plausible loan approval scenarios, enhancing the quality and reliability of any subsequent analyses or models built using this data.

Data Visualization

After the data cleaning process, the dataset was preprocessed to ensure compatibility with visualization tools. Categorical variables were encoded, and numerical columns were scaled where needed to provide a clear basis for visual analysis.The dataset is now free from anomalies and is properly structured, making it ideal for plotting and exploring relationships between features.

- **Visualizations Generated**:
  - A series of **bar charts**, **box plots**, **pie charts**, and **correlation heatmaps** were generated to gain insights into loan approval trends.
  - Key features analyzed include **CIBIL Score**, **Income Buckets**, **Savings Levels**, **Credit History**, and **Debt-to-Income Ratio (DTI)**, each plotted against **Approval Status**.
  - The visualizations help to identify patterns, such as how higher income, better credit history, or higher savings levels positively correlate with loan approvals.
- **Purpose of Visualizations**:
  - The visualizations were created to better understand how different financial and demographic factors impact the loan approval process.
  - Patterns in approval and rejection rates across categories such as **age group**, **employment type**, and **investment levels** provided valuable insights for further analysis and model building.
- **Link to Statistical Analysis**:
  - The visual insights gathered will form the basis for statistical hypothesis testing in subsequent steps. Hypotheses like whether **higher CIBIL scores significantly impact loan approval rates** or whether **employment type is a strong predictor of loan acceptance** were formulated based on these visual findings.
  - Statistical tests, such as **Chi-Square Tests**, **T-Tests**, and **ANOVA**, will further validate the trends observed in the visualizations to derive actionable conclusions.

Statistical Analysis Summary

Despite conducting various statistical tests, many failed to show significant relationships between the examined factors and loan approval status. However, visual representations indicated patterns that were not always corroborated by statistical tests. Here are some key points from the hypothesis testing:

Personal Loan

- **Chi-square Tests:**
  - Credit History vs. Approval Status: The null hypothesis was rejected, suggesting a significant association between Credit History and Approval Status.
  - Employment Type vs. Approval Status: The null hypothesis was rejected, indicating a significant association between Employment Type and Approval Status.
  - Investment vs. Approval Status: The null hypothesis was rejected, indicating that the level of investment significantly influences Approval Status.
- **T-Test:**
  - Investment Portfolio Value: The null hypothesis was rejected, suggesting a significant difference in mean investment portfolio value between approved and rejected applications.
- **One-Way ANOVA for DTI by Income Buckets:**
  - The null hypothesis was rejected, indicating that there is a significant difference in the mean DTI between different income bucket categories .

Home Loan

- **Chi-square Test:**
  - Employment Type vs. Approval Status: The null hypothesis was rejected, indicating a significant association between Employment Type and Approval Status.
- **One-Way ANOVA for DTI by Income Buckets:**
  - The null hypothesis was rejected, suggesting significant differences in the mean DTI across different income bucket categories.

Car Loan

- **Chi-square Test:**
  - Income vs. Approval Status: The null hypothesis was rejected, indicating a significant association between income and Approval Status.
- **Two-Sample T-Test for Monthly Salary:**
  - The null hypothesis was rejected, showing a significant difference in the mean Monthly Salary between approved and rejected applicants.
- **One-Way ANOVA for DTI by Income Buckets:**
  - The null hypothesis was rejected, indicating significant differences in mean DTI across different income bucket categories.
- **One-Way ANOVA for DTI by CIBIL Score:**
  - The null hypothesis was rejected, suggesting a significant difference in mean DTI across different CIBIL score categories.

Educational Loan

- **Chi-square Test:**
  - Employment Type vs. Approval Status: The null hypothesis was rejected, indicating a significant association between Employment Type and Approval Status.
- **One-Way ANOVA for DTI by Income Buckets:**
  - The null hypothesis was rejected, suggesting significant differences in mean DTI across different income bucket categories.
- **One-Way ANOVA for DTI by CIBIL Score:**
  - The null hypothesis was rejected, indicating significant differences in mean DTI across different CIBIL score categories.

❖ **Discrepancies Between Visual and Statistical Analysis**

- Visual Analysis Shows Potential Patterns: Despite the failure of certain statistical tests to reject the null hypothesis, visual analysis, such as bar charts, pie charts, and box plots, often showed discernible trends and patterns. For example, higher CIBIL scores or lower DTI ratios visibly correlated with higher loan approval rates in the visualizations, whereas the statistical tests did not always confirm these associations.

- Significant Variability Affected Results: Descriptive statistics revealed high variability in features such as DTI and monthly salary. This could have limited the ability of statistical tests to detect **significant relationships, as indicated by the high standard deviations in these features.**

Summary

- The analysis shows that statistical tests such as Chi-Square, T-Test, and ANOVA did not always reveal statistically significant relationships. However, visual representations provided more intuitive insights and suggested underlying patterns.

- Features such as employment type, investment portfolio value, income bucket categories, and CIBIL score demonstrated significant relationships with approval status in specific cases, as indicated by the rejection of the null hypothesis in relevant statistical tests.

Key Insights:

1. **Income Buckets and Loan Type**:

   - Across all loan types, higher income buckets tend to show higher approval rates compared to lower income buckets. This is evident from the approval counts, which are consistently higher for applicants in the higher income ranges.

   - Specifically, for **Home Loans**, there is a trend where applicants earning between **10-20 Lakhs** have the highest number of approvals. This indicates that income is an important factor in determining loan approval chances.

2. **Savings Bins**:
    - Applicants with savings between **1-2 Lakhs** and **3-4 Lakhs** have relatively high approval rates across all loan types.
    - **Home Loans** have the highest approval rate for the savings range of **1L-2L**, indicating that applicants with some savings are considered less risky by lenders.
3. **Credit History**:
    - **Good and Very Good Credit History** categories have the highest acceptance rates across all loan types. Applicants with a **Poor Credit History** still have some approvals, which suggests that credit history, although important, may not be the sole determining factor in loan approval.
    - Applicants with **No Credit History (NA/NH)** have very few applicants, but the majority of them were approved, indicating potential favorable consideration for first-time applicants.
4. **Employment Type**:
    - The **Unemployed** category has **zero approvals** for Home Loans, Personal Loans, and Car Loans, while **Educational Loans** show a few approvals despite unemployment, which might be due to the nature of the loan and the expectation of future income.
    - **Self-employed** and **Salaried** applicants have the highest number of approvals across all types, indicating that employment stability positively impacts loan approval.
5. **CIBIL Score**:
    - **Higher CIBIL scores (750-900)** consistently show higher approval rates across all loan types, with the highest rate for Personal Loans (75%). This suggests that maintaining a high credit score significantly increases the chances of loan approval.
    - **Lower CIBIL scores (600-649)** still see a considerable number of approvals for Car Loans and Educational Loans, indicating that for some loan types, lenders may be willing to take on greater risk.
6. **Age Below 20 Years:**
    - It was observed that 38 applicants under 20 years of age had unusual educational achievements, with 18 having reported a Master's degree and 4 having completed a PhD. This is highly unlikely and raises concerns about the accuracy of the data or potential misrepresentation.
    - Out of these 38 applicants, 22 were approved for loans, which poses a risk. This highlights the need for lenders to carefully verify the accuracy of educational and employment details for younger applicants.

7. **Type of Residence and Property Ownership:**
    - A confusing trend was noted where some applicants reported living in rented residences while owning property, and yet were approved for loans. This could indicate potential risks in understanding the applicants' financial stability and ownership status.
    - Lenders should exercise caution and ensure clarity in verifying applicants' residence and property ownership to mitigate potential risks associated with unclear or contradictory information.

Recommendations:

1. **Emphasize Creditworthiness**: Applicants should focus on improving their credit scores, as a **high CIBIL score (750-900)** clearly correlates with higher approval chances.
2. **Stable Employment**: Lenders heavily favor applicants with stable employment (Salaried or Government jobs). Unemployed applicants have **zero or very few approvals**, indicating the importance of employment in determining loan eligibility.
3. **Savings are a Strong Indicator**: Encourage applicants to have some level of savings (between **1-4 Lakhs**). Savings contribute positively to loan approval rates across all loan types.
4. **Income Impact**: Higher-income individuals (above **10 Lakhs**) have higher chances of approval. Therefore, applicants in lower income buckets may need to explore options like co-signers or better collateral to increase approval likelihood.
5. **Loan Terms and Previous Loan Experience**: Applicants with a **positive loan history** are viewed favorably. Providing guidance on maintaining good repayment behavior in existing loans could improve future loan approval chances.
6. **Thorough Verification for Younger Applicants:** Given the discrepancies in education and employment details for applicants under 20, lenders should implement stricter verification processes for younger individuals to ensure the authenticity of the information provided.