**INTRODUCTION**  
**Objective Of the Project**  
This project aims to develop a predictive model that identifies low-risk customers for loan approvals, pinpointing top-performing customers and providing actionable insights into cities with favourable risk profiles. Through comprehensive data analysis, we will:

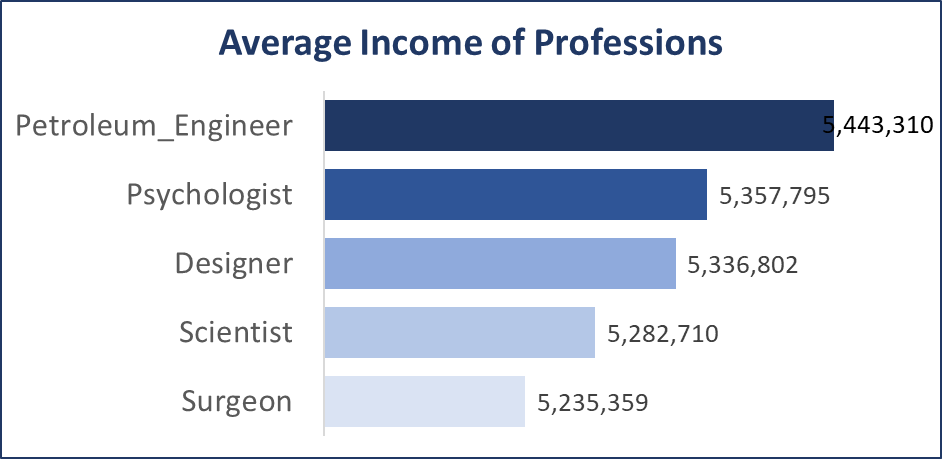
1. Identify low-risk customers: Develop a robust risk assessment framework to flag customers with a high likelihood of loan repayment.  
2. Determine top-performing customers: Analyze customer behavior and characteristics to identify high-value, low-risk borrowers.  
3. Provide city-level insights: Uncover cities with low-risk profiles, enabling targeted marketing and business growth strategies.

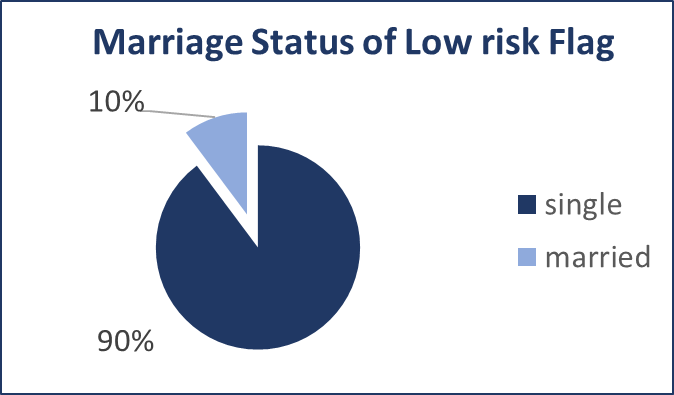
By achieving these objectives, this project will support informed lending decisions, optimize risk management, and drive business growth.  
**Problems Being Addressed**  
1. Risk Assessment: Difficulty in accurately identifying low-risk customers for loan approvals.  
2. Customer Segmentation: Challenge in determining top-performing customers who are likely to repay loans.  
3. Geographic Risk Profiling: Limited insights into cities or regions with favourable risk profiles.  
**Key Datasets and Methodologies**  
Key Datasets  
1. Customer Demographic Data: Age, income, occupation, education level, etc.  
2. Loan Application Data: Loan, repayment terms.  
3. Geographic Data: City, region, zip code, etc.  
Methods:   
Data Cleaning/ preprocessing, Pivot Tables, and Dashboard Creation/visualization  
**STORY OF DATA**  
**Data Source**  
Database for BOD Group Microfinance.  
**Data Collection Process**  
The data was gathered through Customer KYC (know your customer) logs  
**Data Structure**  
Each row represents the customer ID, while each column has different independent and dependent variables, e.g., age, income, experience, married/single, etc.  
Important Features and Their Significance  
● High Risk Flag: Determines customers that pose high risk.  
● Low Risk: Identifies the top-low-risk customers.  
● Profession: Measures the average income that will default or not.  
● Low Risk Status: Provides insight into how status can bring low risk.  
● Region: Provides details about performance in different regions

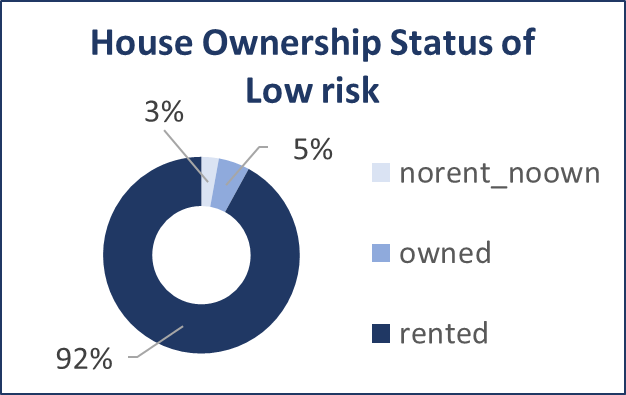
**Data Limitations or Biases**  
1. No Payment History Data: Repayment records, defaults, late payments, etc.  
2. Limited scope: Data may not capture all relevant variables or factors.  
3. Outdated data: Data may not reflect current market trends or customer behavior.  
4. Sample size limitations: A Small sample size may not be representative of the larger population.  
5. Information bias: Data may be biased due to incorrect or incomplete information.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
**DATA SPLITTING AND PREPROCESSING**  
**Data Cleaning**  
● Removed empty rows  
● Checked the date formats and put them in standard form.  
● Converted income from text to numeric data.  
● Converted the data to a standard Excel table  
**Data Transformations**  
● None was done.  
**Data Splitting**  
● The following columns were split and identified as independent variables:  
• Customer ID  
• City  
• State  
• Married/ Single  
• House Ownership  
• Profession  
● The dependent variables:  
• Risk Flag  
• Experience  
• Current Job years  
• Income  
• Age  
• Current House years  
• Car Ownership  
Industry Context  
● Credit Loan Company Data  
● Knowing the industry type of data gives perspective into the analysis to be made and what success means to such a company.  
Stakeholders  
● The Micro-finance Management  
● Credit Bureau Commission  
● Head, Risk and Strategy  
Value to the Industry  
● Insights from this report help identify credit-worthy Customers, that is, customers who are flagged as low risk (0) and will not default on a loan.

**PRE-ANALYSIS**  
Identify Key Trends  
● Fine-tune the best profession that would most likely be considered for a credit facility  
● fine-tune the average income of customers in a profession  
● Fine-tune the age group of customers that would most likely be considered for a credit loan.  
Potential Correlations  
● High revenue categories had consistent order frequency.  
● Cities with high revenue seem to have high-performing salespeople   
● Customers who are Physicians are most likely not to default on a loan   
Initial Insights  
● Customers in the age group ranging from 50–80 years of age are very reliable customers who will adhere to the conditions of their loan facility  
● The best average income of about 5m is for customers who are petroleum engineers  
● Higher discounts lead to increased sales volume but reduced profits.  
● The best performing customer(s) were in low-risk region  
**IN-ANALYSIS**  
Unconfirmed Insights  
● Certain categories had a higher risk.  
● Certain categories had lower risk.

**Recommendations**  
● Our micro-finance institution should open a sub-branch in the city of Vijayanagram to give access to more customers within the city.  
● More focus should be on customers in the age group ranging from 50–80 years of age are very reliable customers who will adhere to the conditions of their loan facility.  
● Build strong relationships with high-potential customers.  
**Analysis Techniques Used in Excel**  
● Pivot Tables  
**POST-ANALYSIS AND INSIGHTS**  
Key Findings  
● Petroleum engineers had an average income.  
● Physicians are the low-risk profession, followed closely by psychologists   
● Vijayanagaram thrives the most as a top city with low risk.   
**DATA VISUALIZATIONS & CHARTS**

The average income of professions shows that a petroleum engineer, compared to a surgeon

Marriage status shows that the singles have a low risk flag

Those living in a rented apartment are considered to be of low-risk status.

**RECOMMENDATIONS AND OBSERVATIONS**  
Actionable Insights  
● Customers who are Physicians may be considered for any loan facilities from the company  
● Customers whose average income is below 5m may be prioritized for loan facilities due to their low income  
● Our micro-finance institution could open a sub-branch in the city of Vijayanagram to give access to more customers within the city.  
● Customers within the age group of 61 to 70 and 41 to 50 years of age are most reliable and may be considered for their loan applications  
● Customers in No-rented or No-owned apartments are very high-risk customers  
● High-risk customers who are likely to default on a loan should not be considered for any loan facility.  
● We may incur more costs to retrieve such a loan from High-risk customers.   
● Low-risk customers should at least meet the categories specified from point 1 to point 5 above.

Optimizations or Business Decisions  
● Our micro-finance institution should open a sub-branch in the city of Vijayanagram to give access to more customers within the city.  
● More focus should be on customers in the age group ranging from 50–80 years of age are very reliable customers who will adhere to the conditions of their loan facility.  
● Build strong relationships with high-potential customers. Investigate weaker sales in February and April and introduce time-limited promotions.  
Unexpected Outcomes  
● Those living in a rented apartment may not necessarily be considered to be of low-risk status.   
● The age group contributes to the low or high risk for loan prediction.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**CONCLUSION**  
Key Learnings  
● Loan prediction is largely influenced by specific categories like age, experience, etc   
● Not all professions can be significantly rated as high or low risk.

**Limitations**  
The data used has limitations.  
**Future Research**  
● Conduct a deeper analysis, including detailed:  
○ Loan Application Data: Loan amount, interest rate, repayment terms, etc.  
○ Payment History Data: Repayment records, defaults, late payments, etc.