



# GUVI

## DATA SCIENCE JOB FAIR-2024

### DATA SCIENTIST AI ML

### Financial Risk Detection



## Introduction:

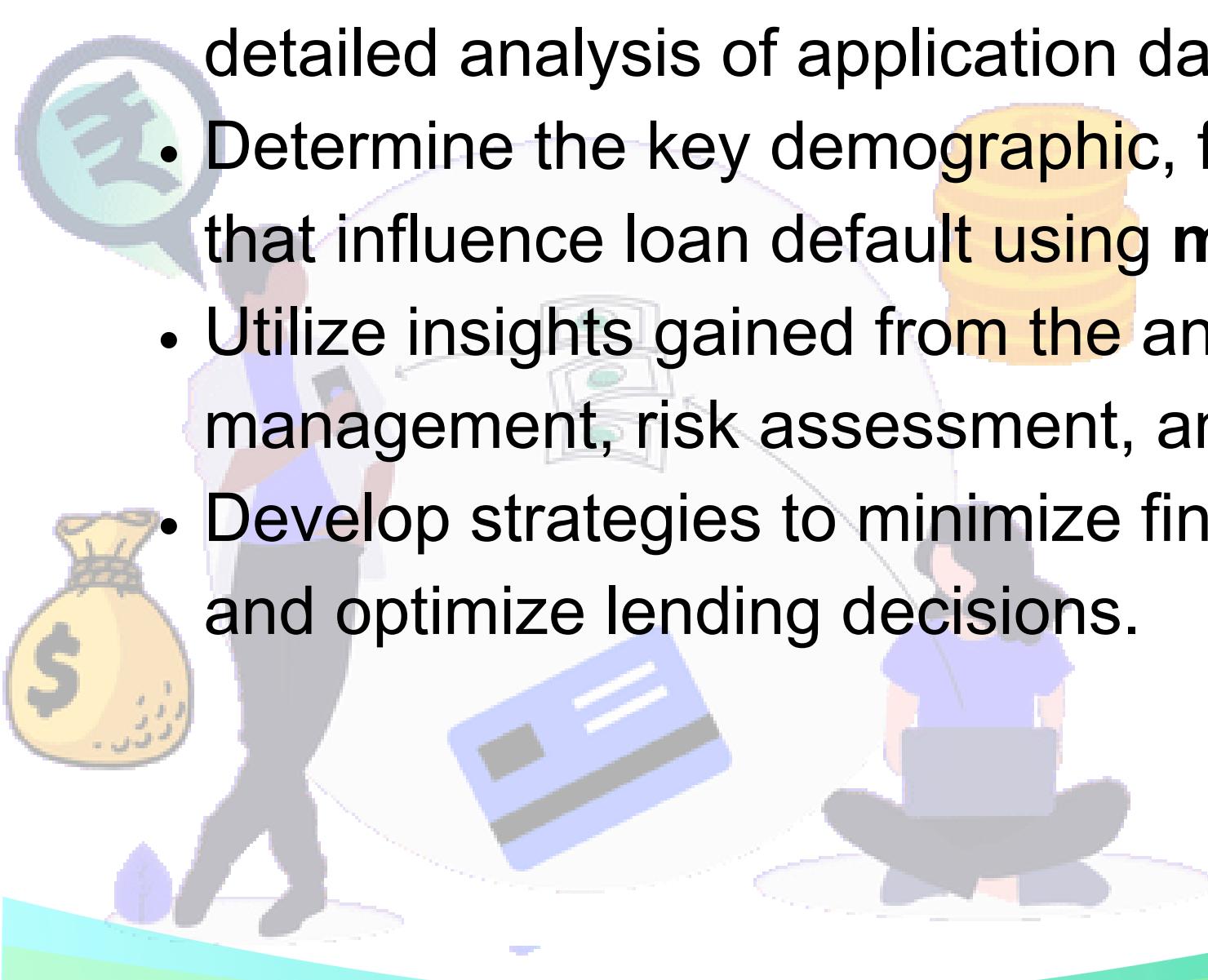
This project aims to leverage **Exploratory Data Analysis (EDA)** and **machine learning** to conduct risk analysis for loan default prediction in the context of a consumer finance company. By analyzing historical loan application data, we will identify patterns and factors that indicate whether a client is likely to default on their loan payments. This analysis will assist the company in minimizing financial losses while ensuring that creditworthy applicants are not unfairly rejected.

## Business Understanding

The lending industry faces significant challenges in assessing creditworthiness, particularly for applicants with limited or no credit history. Loan defaults pose financial risks to lending institutions, making accurate risk assessment crucial. Our primary objective is to use EDA and machine learning techniques to understand the drivers behind loan default and develop strategies to mitigate these risks effectively.

## Business Objectives

- Identify patterns indicating clients' difficulty in repaying loans through detailed analysis of application data.
- Determine the key demographic, financial, and credit-related factors that influence loan default using **machine learning models**.
- Utilize insights gained from the analysis to inform portfolio management, risk assessment, and lending practices.
- Develop strategies to minimize financial losses due to loan defaults and optimize lending decisions.



# Data Understanding

- [application\\_data.csv](#): Contains client information at the time of loan application
- [previous\\_application.csv](#): Provides data on clients' previous loan applications
- [columns\\_description.csv](#): Data dictionary defining the variables and their meanings

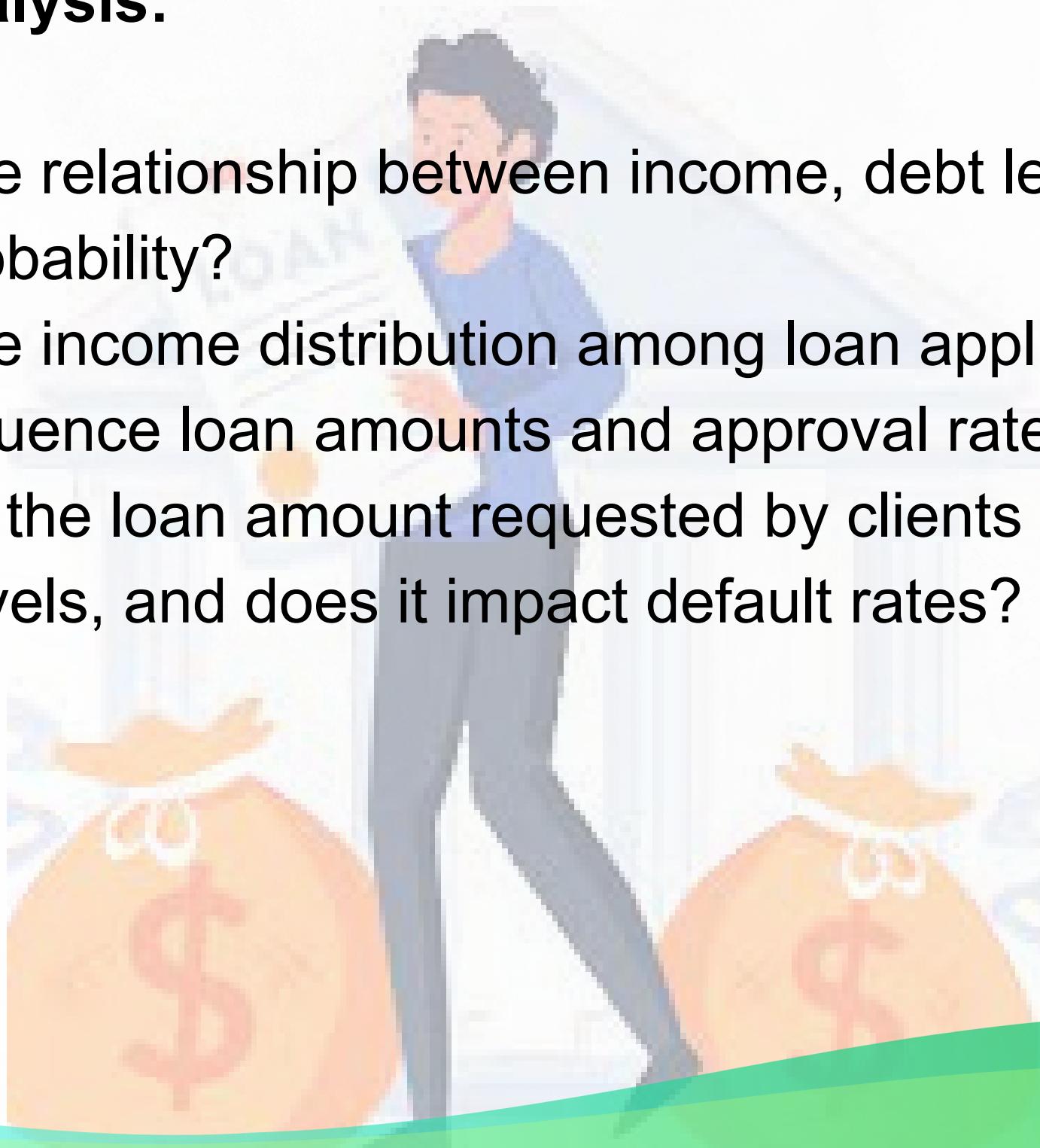
## Exploration ideas

### Client Profile Analysis:

- How do default rates vary across different demographic groups?
- Are there specific educational or employment patterns associated with higher default risk?
- Are there differences in the client profiles between those with payment difficulties and those without?
- Calculate Income to Debt ratio ?

### Financial Analysis:

- What is the relationship between income, debt levels, and loan default probability?
- What is the income distribution among loan applicants, and how does it influence loan amounts and approval rates?
- How does the loan amount requested by clients compare to their income levels, and does it impact default rates?



## Credit History Analysis:

- How do the number of inquiries to the Credit Bureau and past defaults impact the creditworthiness of clients?
- Are there any relationships between the client's credit risk and the reasons for rejection of their previous loan application
- Is the type and length of credit history relevant to default risk assessment?

## Loan Application Decisions Analysis:

- Do specific loan types, amounts, or purposes correlate with higher default risk?
- Is there a difference in approval rates between different types of loans (cash vs. revolving)?
- How do previous application outcomes (approval, refusal) affect future default risk?
- Are there any correlations between the client's credit risk and the weekday or hour they applied for the loan?

## Feature Importance Analysis:

- Which factors have the strongest predictive power for loan default?
- Are there any unexpected or surprising relationships between features and default risk?

## Portfolio and Risk Assessment:

- How can we segment the client base based on risk profiles?
- What strategies can be implemented to mitigate risks for different borrower groups?
- Can we identify segments of clients with higher default risk based on their portfolio characteristics?

- How do macroeconomic factors impact the overall risk profile of the loan portfolio?
- Are there any trends or patterns in the default rates over time, and how do they affect the risk assessment?

## Note:

After completion of all the task you need to create a PowerPoint presentation that should contain the:

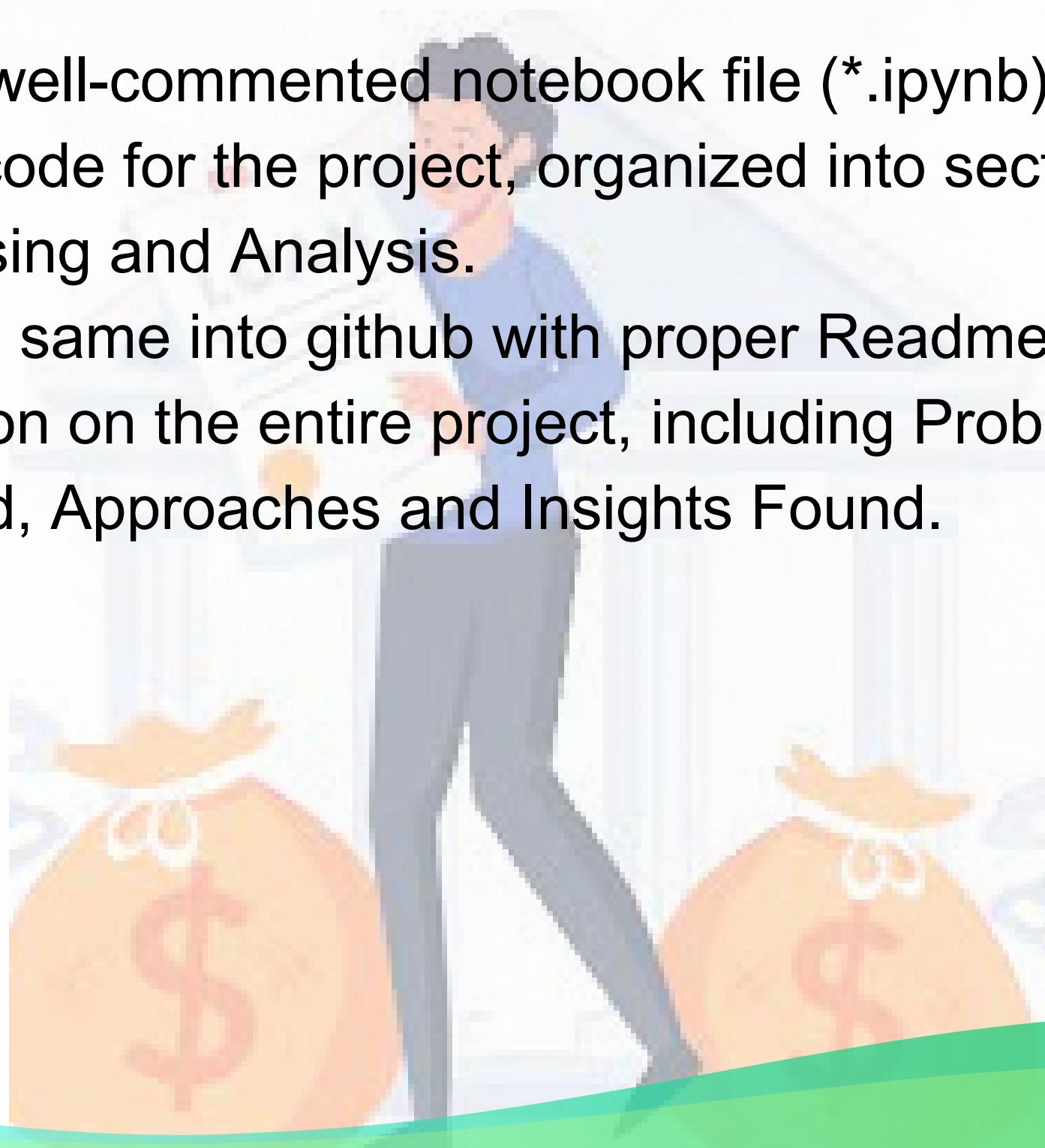
1. Problem Statement
2. Tools Used
3. Approaches
4. EDA Insights
5. Model Building & Evaluation
5. Conclusion/Suggestion

## Disclaimer:

It's important to note that this project is open-ended, and additional analyses can be conducted based on specific requirements or preferences.

## Submission:

- Provide a well-commented notebook file (\*.ipynb) containing the complete code for the project, organized into sections for data preprocessing and Analysis.
- Upload the same into github with proper Readme file.
- Presentation on the entire project, including Problem Statement, Tools Used, Approaches and Insights Found.



## Evaluation metrics:

- Project evaluation will be done in the live session and have to showcase the approaches done to complete the project
- You are supposed to write a code in a modular fashion (in functional blocks)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.(Mandatory)
- You have to keep your GitHub repo public so that anyone can check your code.(Mandatory)
- Proper readme file you have to maintain for any project development(Mandatory)
- Follow the coding standards:  
<https://www.python.org/dev/peps/pep-0008/>
- You should include basic workflow and execution of the entire project in the readme file on GitHub

## GitHub Repo:

The attached reference document will help you use GitHub effectively. - [Link](#)

