# Portfolio Project: Loan Approval Prediction & Credit Risk Analysis

## Problem Statement (Business Context)

You are an AI **Engineer** working in the **Retail Credit Risk Analytics** team of a commercial bank.

The bank receives thousands of **personal loan applications** every month. Each application contains demographic, financial, and credit-behavior information about the applicant. The bank's underwriting team wants to:

1. **Understand what factors influence loan approvals**

2. **Build a predictive model** that estimates the probability of loan approval

3. **Design a decision rule** (approval/decline threshold) based on business risk

4. Ensure the solution is **explainable, fair, and production-ready**

Your task is to use the provided dataset (`csv`) to simulate a **real-world bank credit underwriting project**, covering the full AI lifecycle.

This project is expected to be **portfolio-ready** and suitable for presentation to recruiters, hiring managers, or stakeholders.

## Dataset Overview

The dataset contains historical loan applications with the final decision:

**Target Variable**

- `LoanApproved`

- 1 → Loan Approved

- 0 → Loan Declined

**Features include**

- Applicant demographics

- Income and debt information

- Credit history and risk indicators

- Loan request details

- Application metadata

# Project Objectives

By completing this project, you should demonstrate the ability to:

- Translate a **business problem** into a machine learning task

- Perform **bank-style exploratory data analysis**

- Build **ML pipelines** using `sklearn`

- Communicate results clearly to **non-technical stakeholders**

# Project Structure (Required)

Your submission must follow this structure:

```
loan-approval-ml-project/
|
├── data/
```

```
|      └── Loan.csv
|
├── notebook/
|   ├── ML.ipynb
```

# Task Breakdown & Expectations

## 1. Data Understanding & Cleaning

- Inspect schema, data types, and distributions

- Convert date fields appropriately

- Validate data ranges (e.g., age, credit score)

- Check for duplicates and inconsistencies

- Document all assumptions and cleaning steps

**Deliverable:**
Clear explanation of data quality checks and feature readiness.

## 2. Exploratory Data Analysis (EDA)

Focus on **decision-oriented insights**, not just plots.

**Required analyses**

- Overall loan approval rate

- Approval rate by:

  - Employment status

- ○ Education level

- ○ Home ownership

- ○ Loan purpose

- Impact of:

  - ○ Credit score

  - ○ Debt-to-income ratio

  - ○ Previous defaults / bankruptcy

- Identification of high-risk segments

**Deliverable**

- 8–12 meaningful visualizations

- 5–7 written insights phrased like a business memo

# 3. Modeling & Evaluation

**Minimum models**

- KNN

- Decision Tree

- One tree-based model (Random Forest)

**Requirements**

- Train/test split

- Cross-validation on training data

**Metrics (mandatory)**

- ROC-AUC

- Precision-Recall

- Confusion Matrix

- Classification Matrix

**Deliverable**

- Model comparison table

- Clear justification of final model choice

## 4. Model Interpretability

Explain **why** the model makes decisions.

**Required**

- Partial dependence plots (2–3 key numeric features)

- Coefficient analysis

**Deliverable**

- Ranked feature importance

- Written interpretation in plain English

## Documentation Requirements

**README.md (Mandatory)**

Must include:

- Business problem summary

- Dataset overview

- Approach & methodology

- Key insights

- Final model performance

- How to run the project

## Grading Rubric

| Category | Weight |
|---|---|
| Problem framing & business understanding | 10% |
| Data cleaning & EDA insights | 30% |
| Modeling & evaluation rigor | 30% |
| Interpretability & explainability | 20% |
| Documentation & presentation quality | 10% |
| **Total** | **100%** |