AI & ML	
Problem Statement:	Predicting Loan Default Risk + Building a Gen AI Assistant for Risk Analysts
Instructions	

Part A: Predictive Modeling & ML	**Task A1: Data Exploration** • Identify top predictors of default using EDA techniques. • Visualize trends between default vs. numeric variables. • Use PCA or clustering (e.g., KMeans) to identify natural groupings of risky customers. **Task A2: Model Development** • Build **at least 2 models** (e.g., XGBoost, Random Forest, Logistic Regression). • Evaluate using precision, recall, F1, ROC AUC. Focus on **recall** (for identifying defaulters). • Discuss trade-offs between explainability and performance. **Task A3: Model Interpretation** • Use SHAP or LIME to interpret why a specific customer defaulted. • Explain how the interpretation would affect a real bank's decision-making. **Task A4: Real-World Simulation** • Simulate a mini production use case: take a batch of 10 new customers and predict default probability. • Rank them and recommend 3 for **auto-rejection**, 3 for **manual review**, 4 for **approval.**

Part B: Generative AI Application	**Task B1:** Prompt Engineering
	• Design 3 high-quality prompts for GPT-4:
	    ◦ Summarizing customer risk profile from raw loan features.
	    ◦ Generating an explainable rejection note for a loan.
	    ◦ Creating a simple investment recommendation from a group of loan accounts.
	Task B2: LLM Workflow Design
	• Create a **flow diagram** where GPT supports analysts:
	    ◦ Accepts structured data ➜ Converts to readable profiles ➜ Flags risk with rationale ➜ Outputs summaries.
	• Describe token optimization and latency strategies for real-time use.
	Task B3: Finetuning or Embedding Use Case
	• Choose one:
	    ◦ Option 1: Propose how you'd fine-tune GPT on historical loan memos to generate custom risk reports.
	    ◦ Option 2: Describe how you'd use vector embeddings to power semantic search over customer complaint data or risk logs.

Part C: Communication & Ethic	**Task C1**: Executive Summary
	• Write a 1-page report summarizing:
	    ◦ Key insights
	    ◦ Model strategy
	    ◦ Gen AI use case
	    ◦ Business impact
	Task C2: Risk & Ethics
	• Discuss how bias could creep into models (e.g., location-based decisions).
	• Propose 2 safeguards for ethical Gen AI usage in credit scoring.

Development	
Problem Statement:	Build a chatbot
Instructions: Online platform or GitHub repo with .zip + Readme	
Section A: **Core** **Programming** **& Problem** **Solving**	Create a simple console chatbot that: o Greets the user o Accepts a question about loan terms like "What is EMI?" or "What is tenure?" o Responds with predefined answers (use if-else or dictionaries)
Part B: **Loan Risk** **Calculator**	Using the following logic: risk_score = (missed_repayments * 2) + (loan_amount / collateral_value) + (interest / 2) Write a function that accepts these values and classifies the loan as 'LOW', 'MEDIUM', or 'HIGH' risk based on: o Score < 15 → LOW o Score 15–25 → MEDIUM o Score > 25 → HIGH
Part C: **Business Logic** **Automation**	EMI Risk Tagging (15 Marks) Build a function classify_risk(row) that: o Takes a row and applies: score = (missed_repayments * 2) + (loan_amount / collateral_value) + (interest / 2) o Adds a risk_level column: o LOW (<15), MEDIUM (15–25), HIGH (>25) Use apply() to add this to the full dataset.