Hackathon Problem Brief -Pragati FinCorp Credit Risk Insights Challenge

Background

Pragati FinCorp is a fast-growing Non-Banking Financial Company (NBFC) operating across multiple regions in India. Over the past few years, the company has witnessed an increase in non-performing assets (NPAs), inconsistent approval decisions, and operational bottlenecks in underwriting.

To scale responsibly and minimize credit losses, Pragati FinCorp wants to develop a data-driven understanding of borrower behaviour. Before building a predictive model, the leadership team needs clarity on what factors drive credit risk and how customer, loan, and bureau variables interact.

Objective

Your task is to analyze the provided merged dataset and uncover the patterns, signals, and features that influence loan repayment risk. You are expected to:

- Clean and preprocess the dataset
- Perform exploratory data analysis (EDA)
- Identify dominant features that distinguish potentially risky borrowers from safe borrowers
- Present data-backed business insights that can guide smarter lending decisions

Note:
No machine learning model is required. This challenge is about understanding the business through data.

Dataset Description

You will receive a cleaned and merged dataset containing:

- Customer demographics
- Loan-level details
- Bureau and credit report attributes

All columns are anonymized but structurally representative of real NBFC data.

Deliverables

Each team must submit:

1. A **Jupyter Notebook** showing:
    - Data cleaning steps
    - EDA (graphical + statistical)
    - Feature importance reasoning

    - Outlier & anomaly investigations
    - Final business insights
2. A **short presentation (5 minutes)** summarizing:
    - What drives credit risk
    - What patterns were discovered
    - What actions Pragati FinCorp should take

Key Questions You Should Answer

- Which features correlate strongly with repayment issues?
- What borrower segments show inherently higher risk?
- How do loan amount, EMI burden, credit utilization, and bureau history interact?
- What early warning signs appear before a borrower defaults?

Hints to Help Participants During the Hackathon

How to Identify Outliers

Encourage them to check numeric columns like:

- processing_fee
- loan_amount
- emi
- credit_utilization
- current_balance

Useful techniques:

- Boxplots to visualize extreme values
- Compare *processing_fee (less than 3%)* as a percentage of *loan_amount*
- Look for unrealistic values such as:
    - fee > loan_amount
    - loan_amount 10x typical range
    - tenure extremely short or long

A processing_fee that is unusually high compared to the loan amount often signals either data issues or aggressive pricing for higher-risk customers.

Feature Engineering Hints (Domain-Driven)

Loan-related features

- **LTI (Loan-to-Income Ratio)**
  loan_amount / income
- **Deliquency-ratio**

For Exploratory Data Analysis can use KDE plots