

## **Project Description: Loan Risk Analysis and Data Cleaning**

### **Project Overview**

The objective of this project is to analyze and clean a dataset containing customer loan information to ensure data quality and reliability for subsequent analysis. The dataset includes various attributes related to customer demographics, loan details, and historical loan performance. By performing data cleaning, we aim to remove inconsistencies, handle missing values, and eliminate duplicate records, thus preparing the data for more accurate and insightful analysis.

### **Steps Undertaken**

#### **1-Loading and Initial Inspection:**

- Loaded the dataset into an SQL table for initial inspection.
- Queried the dataset to get an overview of the data.

#### **2-Checking for Missing Values:**

- Identified missing values in critical columns such as 'customer\_id' and 'loan\_amnt'.

#### **3- Identifying Duplicate Records:**

- Used SQL window functions to detect duplicate entries based on customer\_id, 'customer\_income', and 'loan\_amnt'.

#### **4- Inspecting and Cleaning Specific Records:**

- Inspected specific records for anomalies or errors by querying individual customer\_ids.

#### **5- Fixing Data Formats :**

- Corrected the loan\_amnt column format to ensure it is stored as a numeric value.

### **Outcome**

By following these steps, the dataset is now cleaned and ready for further analysis. This cleaned dataset will provide more accurate insights and support reliable data analysis, including loan risk assessment and predictive modeling.

### **Future Steps**

#### **Exploratory Data Analysis (EDA):**

- Perform detailed EDA to uncover trends and patterns in the data.
- Visualize data distributions and relationships between variables.

#### **Reporting and Visualization:**

- Create dashboards and reports to communicate insights and findings.
- Present recommendations based on the analysis.