Data Collection and Preprocessing Phase

Date	20 June 2025
Team ID	SWTID1749791625
Project Title	Smart Lender- Applicant Credibility Prediction for Loan Approval
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	This machine learning project aims to predict loan approval based on applicant information. Using a dataset with features such as gender, marital status, income, and credit history, the objective is to build a model that accurately classifies loan status (approved or denied), enabling efficient and informed decision-making in the lending process.
Data Collection Plan	☐ Search for datasets related to loan approvals, financial information, and applicant details ☐ Prioritize datasets that include diverse demographic information
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle and UCI, two popular platforms for data science

competitions and repositories. The sample data represents a subset of the collected information, encompassing variables such as gender, marital status, income, and loan-related details, suitable for machine learning analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (such as gender and marital status), financial metrics (including income and loan amount), and loan approval outcomes.	https://www.kaggle.com/datasets/rishikeshkonapure/home-loan-approval?select=loan_sanction_train.csv	CSV	15KB	Public
UCI	This dataset focuses on credit card applications and offers a good mix of attributes for analysis	https://archive.ics.uci.edu/dataset/27/credit+approval	CSV	13.6 kB	Public