Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | xxxxxx |
| Project Title | xxxxxx |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

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 | The 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),  facilitating efficient and informed decision-making in the lending  process. |
| Data Collection Plan | Search for data related to travel bookings using train and aeroplane booking.we can get the data in the goole or in the app chart/spreadsheet |
| Raw Data Sources Identified | The raw data sources includes url/spreadsheet/Kaggle datasheet.they can also avail the charts prepaired by the application  Regarding travel bookings. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access**  **Permissions** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset 1 | The dataset  comprises applicant  details (gender,  marital status),  financial metrics  (income, loan  amount), and loan  approval outcom | https://www.kagg  le.com/datasets/ri  shikeshkonapure/  home-loan-appro  val?select=loan\_s  anction\_train.csv | CSV | 15 GB | Public |
| … | … | **…** | … | … | … |
| … | … | … | … | … | … |