**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | July 5, 2024 |
| Team ID | 739838 |
| Project Title | Customer segmentation using Machine Learning |
| 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 project aim is to enable the learner to understand the business use case of how and why to segment the customers. Using a dataset with features such as ID, sex, marital status, Age, Education, Income, Occupation, settlement size, the objective is to build a model that accurately classifies the customers facilitating efficient and informed decision- making in the segmentation process. |
| Data Collection Plan | * Search for the datasets related to the customer purchases history, financial information. * Prioritize datasets with diverse demographic information |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as sex, marital status, income, and customer related details for machine learning analysis. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset | The dataset comprises applicant details(ID, sex, marital status, Age, Education, Income, Occupation, settlement size) and type of customer outcomes. | <https://docs.google.com/spreadsheets/d/1NnUMX3sjJgRRerkJTAXemlfdyo2GiUhgE_m4w-fAhvs/edit?gid=1219451115#gid=1219451115> | CSV | 72 KB | Public |