Data Collection and Preprocessing Phase

Date	18 June 2024
Team ID	739642
Project Title	Customer shopping segmentation using machine learning
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 acuration and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan:

Section	Description
Project Overview	Customer shopping segmentation using machine learning involves collecting transactional, demographic, and behavioral data. After preprocessing and feature engineering, exploratory data analysis (EDA) helps understand patterns. Segmentation techniques like K-Means clustering or hierarchical clustering are then applied to group customers. These segments enable targeted marketing and personalized experiences, optimizing customer engagement and business strategies.
Data Collection Plan	Collect transactional data (purchase history, frequency, monetary value), demographic data (age, gender, location), and behavioral data (browsing patterns, cart abandonment, feedback). Use point-of-sale systems, online transactions, customer surveys, and CRM databases as sources. Integrate all data into a unified database for analysis.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as age,

	Gender, quantity, price, payment method, shopping mall, category.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (invoice no, customer id, quantity, age , gender, invoice date, shopping mall , payment method, category, price.	https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset	CSV	15 kB	Public