Data Collection and Preprocessing Phase

Date	07-07-2024
Team ID	740669
Project Title	Customer shopping segmentation by using machine learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

To segment customer shopping data, gather transaction history, demographic info, and behavioral data from multiple sources like online platforms, loyalty programs, and surveys. Ensure data quality and integrity through meticulous curation. Analyze this data to identify distinct customer segments for targeted marketing and improved decision-making.

Section	Description
Project Overview	The customer shopping segmentation project aims to segment customers based on their shopping behavior and demographic information. Using a dataset with features such as age, gender, purchase history, and spending habits, the objective is to build a model that accurately identifies different customer segments, facilitating targeted marketing and personalized customer experiences
Data Collection Plan	● Search for datasets related to customer shopping behavior, demographic information, and purchase history. ● Prioritize datasets with diverse customer profiles and detailed transaction data.

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, purchase history, and spending patterns.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises gender(female,male), age,category,quantity ,customer_id, price, quantity, payment_ method (credit,debit card),shopping mall.	https://www.kaggle.com/datasets/rishikeshkonapure/home-loan-approval?select=loan_sanction_train.csv	CSV	2Mb	Public