

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	LTVIP2024TMID25001
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 customer segmentation project aims to group customers into distinct segments based on behavioral and demographic characteristics. Using features such as age, income, spending habits, and location, the objective is to develop a model that identifies segments for targeted marketing, personalized offerings, and improved customer retention.
Data Collection Plan	<ul style="list-style-type: none"> - Search for datasets related to customer demographics and purchasing behavior. - Prioritize datasets that include diverse customer profiles, including age, income, occupation, and spending patterns.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle and other data repositories that offer customer behavior insights. The dataset comprises variables such as age, income, spending scores, and region, allowing for in-depth segmentation analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset includes customer demographics, purchase patterns, and behavior for segmentation	Kaggle Customer Segmentation Dataset	CSV	26 MB	Public
UCI	This data focuses on customer purchases, with a variety of behavioral attributes.	UCI Customer Data	CSV	18 KB	Public
UCI	This dataset contains customer transactions and behavioral data, used for segmenting retail customers.	Kaggle CLV Dataset	CSV	67 MB	public