Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	LTVIP2024TMID25001
Project Title	Customer Segmentation Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Customer segmentation dataset	Missing values in the Age,Income,Spending_Score, And gender columns.	Moderate	Use mean/median imputation
Customer segmentation dataset	Categorical data in the gender, Occupation,and Region columns.	Moderate	Encoding has to be done (one-hot or label encoding).
Customer segmentation dataset	Outliners in Income and Spending_Score columns	High	Use outlinerer detection methods (e.g.IQR OR Z-score).

Customer segmentation dataset	Inconsistent formatting in the Age column.	Low	Convert all value to a consistent Numeric format