Data Collection and Preprocessing Phase

Date	27 January 2025
Team ID	SWUID20240011509
Project Title	Restaurant Recommendation System
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 curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to predict restaurant recommendation based on applicant information. Using a dataset with features such as gender, price, reviews, and market trend, the objective is to build a model that accurately classifies status, facilitating efficient and informed decision-making in the selection process.
Data Collection Plan	• Search for datasets related to recommendation,statistical information, and restaurant details. Prioritize datasets with • diverse demographic information.
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 gender,

	price and review related details for machine learning analysis.

Raw Data Sources Report:

Source Name	Description	Location	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises of all details.	File is provided in the git folder	CSV	15kB	Public
UCI	This data concerns restaurants	File is provided in the git folder	CSV	13.6 kB	Public