**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 18 June 2025 |
| Team ID | SWTID1749631993 |
| 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 recommend restaurants to users based on their preference. Using a dataset having features such as restaurant names, locations, cuisines, customer reviews and cost, the objective is to build a model that provides personalized restaurant suggestion, enhancing user experience and helping in making informed dining decisions. |
| Data Collection Plan | Search for datasets that give comprehensive information about restaurant location, ratings, and user reviews. Primary sources include platforms like Kaggle. |
| Raw Data Sources Identified | The raw data source for this project includes dataset obtained from Kaggle, which is a popular public platform offering datasets for machine learning and data science projects.  The selected dataset includes restaurant details such as name, rating, cost, location, and user reviews. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | The dataset comprises of 17 columns and over 51,000 rows across various features like restaurant name, location, cuisines, rate and cost. | [Zomato Bangalore Restaurants](https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants) | CSV | 574.07 MB | Public |