**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 18 June 2025 |
| Team ID | xxxxxx |
| Project Title | Restaurant Recommendation System |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Report**

Data set variables will be statistically analyzed to identify patterns and outliers, with Python

Employed for pre-processing tasks like normalization and feature engineering. Data cleaning will

address missing values and outliers, ensuring quality for subsequent analysis and modeling, and

forming a strong foundation for insights and predictions.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | **Dimensions:**  51717 rows x 17 columns  **Descriptive statistics:** |
| Univariate Analysis |  |
| Bivariate Analysis | **Numerical vs Categorical Analysis:**      **Categorical vs Categorical Analysis:** |
| Multivariate Analysis |  |
| Outliers and Anomalies |  |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering | prepares **reviews\_list** to be used as a textual feature for TF-IDF vectorization |
| Save Processed Data | - |