

Data Collection and Preprocessing Phase

Date	18 June 2025
Team ID	SWTID1749631993
Project Title	Restaurant Recommendation System
Maximum Marks	2 Marks

Data Quality Report:

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
Kaggle Dataset	Missing values in important columns like rate, cost, and reviews_list	Moderate	Dropped rows with any missing values using dropna() to ensure clean and complete records for modeling
Kaggle Dataset	Presence of irrelevant columns such as phone, url, and dish_liked	Low	Dropped these columns using drop() as they don't contribute to recommendation logic
Kaggle Dataset	Duplicate entries for restaurants.	Moderate	Removed duplicate rows using drop_duplicates() to avoid bias in model training.

Kaggle Dataset	Inconsistent data formats in rate (e.g., "4.5/5", "NEW", "-") and cost (e.g., "2,000", "500").	High	Cleaned rate by removing '/5' and converting to float; cleaned cost by replacing commas and converting to float using <code>astype(float)</code>
----------------	--	------	--