


## DataCollectionandPreprocessingPhase

Date	25 July 2025
Name	Shahrukh Dilawar Sanadi
ProjectTitle	RestaurantRecommendationSystem
MaximumMarks	2 Marks

## DataQualityReportTemplate

Data Source	DataQuality Issue	Severity	ResolutionPlan
Dataset (Restaurant reviewsand metadata)	Missing values in fields like restaurantname,location,or ratings	Moderate	Perform data imputation using techniques like mean/mode for numeric values andmostfrequent value for categorical data. Alternatively, removerows with critical missing fields.

Dataset (User reviews)	Duplicate user review entries	Low	 Removed duplicate records using drop_duplicates in pandas or SQL DISTINCT queries. Use datetime parsing libraries (e.g., pandas.to_datetime) to standardize all date/time fields.
Dataset (Timestamp field)	Inconsistent or incorrect timestamp format	Moderate	



Dataset (Restaurant metadata)	Inconsistent formats (e.g., location written in different ways like "NY", "New York")	Moderate	Apply data standardization techniques, using string functions or regex patterns to unify the format.
Dataset (User preferences)	Sparsely data or insufficient user history	High	Implement fallback strategies such as popularity-based or content-based recommendations when user data is lacking.