DataCollectionandPreprocessingPhase

Date	25 July 2025
Name	Rutuja Digamber patil
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, removerowswith critical missing fields.

			Removeduplicate records using drop_duplicates in pandas or SQL DISTINCTqueries.
Dataset (User reviews)	Duplicateuserreviewentries	Low	
Dataset (Timesta mpfield)	Inconsistentorincorrect timestamp format	Moderate	Use datetime parsing libraries (e.g., pandas.to_datet ime) to standardize all date/time fields.

Dataset (Restaura nt metadata)	Inconsistent formats (e.g., locationwrittenindifferentways like "NY", "New York")	Moderate	Apply data standardization techniques, using stringfunctionsor regex patterns to unify the format.
Dataset (User preferenc es)	Sparsedataorinsufficientuser history	High	Implementfallback strategies such as popularity-basedor content-based recommendations when user data is lacking.