Strategies to Mitigate Data Quality Issues

Handling Missing Values

Imputation: Use statistical methods (mean, median, mode) to fill missing values.

Data Collection: Enhance data collection methods to reduce the occurrence of missing values.

Deletion: If the percentage of missing values is too high, consider removing those records.

Ensuring Data Consistency

Standardization: Implement data entry standards and validation rules.

Data Cleaning: Use tools like Python (pandas) or R to clean and format data consistently.

Addressing Duplicates

Deduplication: Use functions to identify and remove duplicate records.

Data Integration: Improve data integration processes to avoid duplication.

Data Validation

Automated Scripts: Create scripts to validate data integrity and accuracy.

Manual Review: Conduct periodic manual reviews to catch any data quality issues.