Glossary terms from module 3

Terms and definitions from Course 3 Module 3

Algorithm: A process or set of rules followed for a specific task

Data aggregation: The process of gathering data and expressing it in a summary form

Data derivation: The process of combining and processing existing data using an algorithm to create new data

Data integrity: The accuracy, completeness, consistency, and trustworthiness of data throughout its life cycle

Deduplication: The process of eliminating a dataset's redundant data

Duplicate data: A record that repeats the information—in whole or in part—of another record

Exact duplicate: A record where all the data repeats another record

Inner join: A join that only returns the rows that match in both tables

Missing data: When the rows you expect to be returned do not appear in the joined table

NULL value: No available value for a field

Outer join: A join that returns both matched and unmatched rows from one or both tables

Partial duplicate: A record where only part of the data repeats another record