ONE PAGE SUMMARY

IT & Software › Data Analytics › Data Quality

Introduction to Data Quality

Data Quality: The Backbone of Reliable Analytics

DEFINITION

> Data quality refers to the accuracy, completeness, and reliability of data used in analysis. High-quality data leads to better decision-making and insights.

KEY CONCEPTS

Accuracy

Data must be correct and free from errors to be useful.

Completeness

All necessary data should be present; missing data can skew results.

Consistency

Data should be consistent across different datasets and sources.

Timeliness

Data must be up-to-date to ensure relevance for current analysis.

Validity

Data should meet the requirements and constraints of the intended analysis.

EXAMPLES

→ A customer database missing phone numbers is incomplete.

→ Using outdated sales data can lead to poor business decisions.

→ Inconsistent formats for dates across datasets can cause confusion.

MEMORY TIPS

★ Remember 'ACT TV' for Accuracy, Completeness, Timeliness, Validity.

★ Think of data quality as a house: strong foundation (accuracy) and complete structure (completeness).

★ Use the acronym 'CATS' for Consistency, Accuracy, Timeliness, and Completeness.

COMMON MISTAKES

✗ Assuming all data from a source is accurate without verification.

✗ Neglecting to update datasets regularly, leading to outdated information.

✗ Overlooking the importance of data validation processes.

QUICK RECAP

Data quality is essential for effective data analytics. Focus on accuracy, completeness, consistency, timeliness, and validity to ensure reliable insights.