Data Quality 101

Data quality measures how well a dataset meets criteria for accuracy, completeness, validity, consistency, uniqueness, timeliness, and fitness for purpose, and it is critical to all data governance initiatives within an organization. Data quality standards ensure that companies are making data-driven decisions to meet their business goals.

> If data issues, such as duplicate data, missing values, outliers, aren't properly addressed, businesses increase their risk for negative business outcomes.

Dimensions of Data Quality

Data quality is evaluated based on a number of dimensions, which can differ based on the source of information.

These dimensions are used to categorize data quality metrics:

- Completeness: This represents the amount of data that is usable or complete.
- Uniqueness: This accounts for the amount of duplicate data in a dataset.
- Validity: This dimension measures how much data matches the required format for any business rules.
- Timeliness: This dimension refers to the readiness of the data within an expected time frame.
- Accuracy: This dimension refers to the correctness of the data values based on the agreed upon "source of truth."
- Consistency: This dimension evaluates data records from two different datasets, allowing organizations to trust the any actionable insights from their analyses.
- Fitness for purpose: Finally, fitness of purpose helps to ensure that the data asset meets a business need.
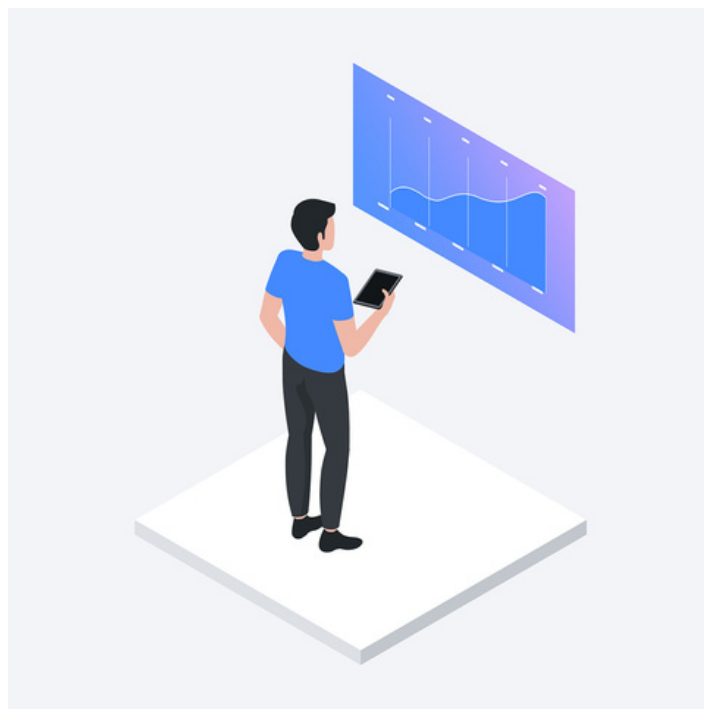
Why is Data Quality important?

Over the last decade, developments within hybrid cloud, artificial intelligence, the Internet of Things (IoT), and edge computing have led to the exponential growth of big data. As a result, the practice of master data management (MDM) has become more complex, requiring more data stewards and rigorous safeguards to ensure good data quality.

Businesses rely on data quality management to support their data analytics initiatives, such as business intelligence dashboards. Without this, there can be devastating consequences, even ethical ones, depending on the industry (e.g. healthcare).

Data quality solutions exist to help companies maximize the use of their data, and they have driven key benefits, such as:

Better business decisions

Improved business processes

Increased customer satisfaction

IBM