

Data Quality

20 December 2022 11:35

By what metrics do we measure data quality?

- **Completeness** - Ensures values are present in all relevant rows or columns (i.e. no missing values in critical fields)
- **Consistency** - Ensures data is stored in the same format or standard when it is stored in multiple locations (e.g. Name is stored using first name and last name across multiple systems)
- **Validity** - Ensures values are aligned to specific rules or business expectations (e.g. Age should be within 1 and 100)
- **Uniqueness** - Ensures values are unique (i.e. no duplicates exist)
- **Timeliness** - Ensures data is present when it is needed (i.e. data needed today is ready today and not weeks later when it has become irrelevant)
- **Accuracy** - Ensures values reflect reality. This can sometimes be difficult to measure without a reliable source. (e.g. a customer's phone number may be valid in terms of the format but inaccurate because it belongs to someone else)