

Ontology of Column Meaning

1. Entity or Attribute Meaning

Specifies what real-world entity, event, or attribute the column represents — i.e., the conceptual object that the data corresponds to.

Guiding Questions:

- What does this column represent in the real world?
- Is it an entity (customer, product), an event (order, visit), or an attribute (price, date, category)?
- What aspect of that entity/event does it describe?

Examples:

- Represents the date when the order was placed.
- Identifies the unique customer associated with each order.
- Shows the city where the customer resides.

2. Value Representation

Describes how the column's values are expressed or formatted — including datatype, units, and standardized conventions.

Guiding Questions:

- What is the unit, format, or convention of the values?
- What standard, scale, or data type do they follow?
- Is there a pattern or schema that defines valid expressions?

Examples:

- Values are in euros.
- Values follow the ISO 8601 date format (YYYY-MM-DD).
- Values are UTC timestamps accurate to the second.

3. Value Semantics

Defines what the possible values mean — including categorical codes, numeric ranges, or domain-specific interpretations — regardless of their syntactic form.

Guiding Questions:

- What does each possible value or range signify?
- Are there special codes or thresholds with specific meaning?
- Do numeric or categorical values represent states, levels, or statuses?

Examples:

- 'M' and 'F' indicate male and female.
- 1 = active, 0 = inactive.
- Values from 1 to 5 represent customer satisfaction levels.

4. Row Context

Specifies how the column relates to the meaning of a single row — i.e., what each value corresponds to given the dataset's granularity.

Guiding Questions:

- What does one row represent (an entity, event, or observation)?
- How should this column's value be understood within that row context?
- Is this value specific to a single instance or an aggregate over multiple instances?

Examples:

- Each row corresponds to one transaction; this value is the total amount of that transaction.
- Each row represents a city-day pair; this value is the average temperature for that day in that city.

5. Aggregation and Derivation

Indicates whether the column value is directly observed or derived from other data, and if derived, specifies the aggregation type and scope.

Guiding Questions:

- Is this value measured directly or computed?
- What operation (sum, average, ratio, etc.) produced it?
- Over what unit or group (per customer, per day, per category) was the aggregation performed?

Examples:

- Sum of all purchase amounts per customer.
- Total sales amount for all transactions on a single day.
- Calculated as $\text{revenue} = \text{price} \times \text{quantity}$.

6. Relational Context

Describes relationships or dependencies between this column and other columns — structural, referential, temporal, or hierarchical.

Guiding Questions:

- Does this column link or refer to another (e.g., a foreign key)?

- Does it depend on, or determine, other columns?
- Does it group, order, or organize other data elements?

Examples:

- Groups transactions belonging to the same region.
- Occurs after order_date for the same transaction.
- Customer_ID links to the Customer table.

7. Temporal Scope

Defines the time span or instant that the column's value refers to — whether it captures a point in time, a duration, or an ongoing state.

Guiding Questions:

- Does the value refer to a specific time, period, or interval?
- Is it a timestamp, a duration, or a status valid within a timeframe?
- How does it align temporally with other attributes in the same row?

Examples:

- Valid from start_date to end_date.
- Represents the timestamp when the record was last updated.
- Shows the number of active users in a given week.