Data Exploration Findings

Understanding `Discount`

Finding 1: Applied Per-Product

Discount is not applied to the whole order, but to individual products within it.

Proof: Query found orders where `MAX(discount) > MIN(discount)` for the same `order_id`.

Finding 2: It's a Percentage

The `discount` column (e.g., 0.5) represents a percentage (50%), not a fixed amount.

Proof: The formula `sales * (1 - discount)` produced logical `sale_after_discounts` values.

Finding 3: Highly Dynamic

Discount for a single product can change over time, even within the same day.

Proof: Grouping by `product_code` and `order_date` showed multiple discounts for the same combination.

Finding 4: Schema Lacks Derived Metrics

The tables do not contain pre-calculated columns for values after promotion (e.g., net sales).

Implication: For efficiency, it's recommended to create a database VIEW or a new summary table containing these calculated metrics for consistent analysis.

Understanding `Sales` & `Quantity`

Finding 1: Data is Consistent

The relationship between `sales` and `quantity` is reliable.

Proof: Allows for consistent calculation of `unit_price = sales / quantity` across the dataset.

Finding 2: Calculating Totals

Total items sold for a product in a time range can be calculated reliably.

Method: Use `SUM(quantity)` grouped by `product_code` and a date range.

The Quest for `Cost`

Hypothesis 1: Find "True Cost" at Break-Even

Idea: If `profit = 0` and `discount = 0`, then `cost = sales`.

Result: DISPROVEN. Query returned 0 records. This implies the business never sells at its exact break-even point.

Hypothesis 2: Calculate "Implied Cost"

We can reverse-engineer the cost from every transaction.

The Master Formula:
`implied_cost = sales * (1 - discount) - profit`

Final Finding: Cost is DYNAMIC

The `implied_cost` for the same product is not stable, even in the same country.

Proof: The `cost_variance` query showed large differences between `MIN(cost)` and `MAX(cost)` for the same product/country group. This suggests cost changes over time.