

Limits of the dataset and Recommendations for the following analysis

No.	Data problems	Recommendation
1	Inconsistent language: English and Vietnamese.	Using 1 language for all data.
2	Unclear header name and different meaning under same name across different sheets.	Naming header column clearly to avoid confusion.
3	Some columns are not used due to NULL/not in standard format/not have statistical meaning/duplicate information.	Standardizing format when inputting raw data. Removing duplicate or redundant columns that do not have statistical meaning.
4	Lacking product/machine information to link 3 fact tables: Order, Distribute, Sales.	Inserting column(s) that show product ordered from Supplier to Master Warehouse and distributed from Master Warehouse to Child Warehouse. ⇒ Required data to analyze product stock.
5	Current data: 1 transaction = 1 product purchased.	Recording quantity for analysis of purchasing more than 1 product per transaction.
6	Lacking Customer dimension table.	Creating 1 dimension table for Customer demographic (age range, gender, education, occupation) at each location type as well as Passersby (if possible).
7	Lacking some information in current dimension tables.	Dimension tables should have more information: <ul style="list-style-type: none"> ● Product: utility, class quality, brand name, flavor, shelf life, packaging, unit price, unit cost. ● Machine: beginning operating date, end operating date, machine model (new, outdated, obsolete), display design. ● Parent Warehouse: population.