. Implementation:

2.1 Data gathering

To analyze global sales performance and customer behavior, we integrated two key data sources (Both from Kaggle):

Transactional Sales Data (CSV):

Structure: 100 records with fields like Transaction ID, Date, Customer ID, Product Category (Beauty, Clothing, Electronics), Quantity, Price per Unit, Country, and City. Scope: Sales across 50+ countries (e.g., USA, Canada, Germany, Japan).

Customer Demographics (JSON):

Structure: 42 records detailing Customer ID, Gender, Age, Country, and City.
Purpose: Link transactional data to customer profiles for segmentation.

2.2 Data preparation

To streamline the integration of our dataset into the data warehouse, we utilized Python for data manipulation and configuration. The dataset was sourced in two formats: JSON and CSV.

During careful examination, critical observations revealed two primary issues:

Duplicated Data:

Redundant rows were identified and rectified using Python to ensure data uniqueness.

Missing/Non-Available Data:

Gaps in the dataset were addressed to improve completeness and reliability.

Additionally, we performed data type standardization to ensure consistency across fields (e.g., converting strings to dates, and numerical values to integers/floats). This preprocessing ensured the dataset was clean, structured, and ready for seamless integration into the data warehouse.

This is our ETL’s python code :
https://drive.google.com/file/d/1_Ib-RebWbXpLS7yC-Qhwq5kyuPvP-9Hq/view?usp= sharing
csv:
https://limewire.com/d/aatUy#fDvT6ms4Sv

2.4 Data Analysis & Visualization Tools:

Metabase for interactive dashboards. Queries: We created queries based on 7 Business insights: this is the link for the full queries in

SQL :
https://docs.google.com/document/d/1AoZ1vvNjMpOJBMXGvhHvNg3rmj2yZ9IoV9Ym1Rg49Uw/edit?usp=sharing