Coding Challenge: Relationship Analysis

Domain: E-Commerce Sales Data

You need to connect multiple datasets, build correct relationships between them, and create meaningful dashboards to analyze customer, product, and sales performance.

Datasets

Customers Table

- CustomerID
- CustomerName
- Gender
- AgeGroup
- Location (City/State)

2. **Products Table**
    - ProductID
    - ProductName
    - Category
    - SubCategory
    - Price
3. **Orders Table**
    - OrderID
    - OrderDate
    - CustomerID (Foreign Key → Customers)
    - ProductID (Foreign Key → Products)
    - Quantity
    - TotalAmount
4. **Returns Table** *(optional to make it harder)*
    - ReturnID

- OrderID (Foreign Key → Orders)
- ReturnReason

Tasks

1. **Import & Model Data**
   - Load all tables into Power BI.
   - Define correct relationships:
     - Customers ↔ Orders (CustomerID)
     - Products ↔ Orders (ProductID)
     - Orders ↔ Returns (OrderID).
   - Ensure *1-to-Many* and *Many-to-1* relationships are properly set.
2. **Data Cleaning/Transformations**
   - Check for missing values (like no return reason).
   - Create calculated columns:
     - *Profit = TotalAmount – (Price × Quantity)*
     - *OrderMonth = MONTH(OrderDate)*
3. **Analysis & Visualizations**
   You should create a dashboard answering:
   - Which **customer segment (AgeGroup/Gender)** spends the most?
   - What is the **relationship between product category and total revenue**?
   - Which cities generate the highest sales?
   - How do **returns** affect sales performance?
   - Which products are frequently returned?
4. **Advanced Challenge**
   - Create a **hierarchical relationship** (Category → SubCategory → Product).
   - Use DAX to calculate **Customer Lifetime Value (CLV)**.
   - Create a **relationship diagram screenshot** to prove they built the model correctly.

Note : Your are provided with 4 csv files which are interlinked identify how they are related to each other and then start the analysis.