**Data Quality Review**

All three datasets Customer, Order, and Shipping have been verified for basic quality:

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Total Records** | **Null Values** | **Summary** |
| Customer | 250 | 0 | Clean and complete |
| Order | 250 | 0 | Clean, has customer linkage and transaction details |
| Shipping | 250 | 0 | Clean, delivery status for each customer |

**Proposed Domain Model Overview**

**Tables:**

* Customer
  + Customer\_ID, First, Last, Age, Country
* Order
  + Order\_ID, Item, Amount, Customer\_ID
* Shipping
  + Shipping\_ID, Customer\_ID, Status

**Relationships:**

* Customer.Customer\_ID = Order.Customer\_ID
* Customer.Customer\_ID = Shipping.Customer\_ID

**Business-Driven Analytical Requirements**

|  |  |  |  |
| --- | --- | --- | --- |
| **Reporting Goal** | **Source Tables** | **Dimensions** | **Measures/Filters** |
| 1. Total amount & country for Pending deliveries | Shipping, Customer, Order | Country | Status = 'Pending', SUM(Amount) |
| 2. Customer-level transactions and product breakdown | Order, Customer | Customer\_ID, Item | COUNT(Order\_ID), SUM(Amount) |
| 3. Most purchased product per country | Order, Customer | Country, Item | MAX(SUM(Amount)) |
| 4. Top product by age segment (<30, ≥30) | Order, Customer | Age Group, Item | MAX(SUM(Amount)) by group |
| 5. Country with minimum transactions and sales | Order, Customer | Country | MIN(COUNT(Order\_ID)), MIN(SUM(Amount)) |

**Technical Specification Story for Data Engineering**

**Story: Enable “Pending Delivery Spend by Country”**

**Objective:**  
Create a dataset showing the total spend for all orders associated with customers whose delivery status is still “Pending”, grouped by customer country.

**Input Tables:**

* Customer
* Order
* Shipping

**Join Logic:**

Customer.Customer\_ID = Order.Customer\_ID

Customer.Customer\_ID = Shipping.Customer\_ID

**Transformations:**

* **Filter:** Shipping.Status = 'Pending'
* **Group By:** Customer.Country
* **Aggregate:** SUM(Order.Amount) as Total\_Spend

**Output Table (Schema):**

|  |  |
| --- | --- |
| **Country** | **Total\_Spend** |
| India | 15,000 |
| USA | 9,000 |

**Testing Instructions (QA):**

1. Confirm all Customer\_IDs in output exist in Shipping with Status = 'Pending'.
2. Ensure join count integrity — no duplicates or missing.
3. Manually verify SUM for a sample country using raw order data.