1. **Verify the accuracy, completeness, and reliability of source data.**

1. There are few customer_ids which are missing in order table, it means there is no order by these customers.
2. There are no missing values in the tables.
3. Important fields are not having missing values. Ex- amount,customer_id, order_id,shipping_id
4. There are no duplicate records for custer, order, shipping tables.
5. Customer table has customer_id, order table has order_id and shipping table has shipping_id as as unique key.

2. **Customer Dataset**
   Necessary Data Components:
   Customer_ID: Unique identifier for each customer (mandatory, integer, non-null).
   First Name: First name of the customer (mandatory, string, non-null).
   Last Name: Last name of the customer (mandatory, string, non-null).
   Age: Age of the customer (mandatory, integer, non-null, range: 0-120).
   Country: Country of the customer (mandatory, string, non-null).

   Data Quality Criteria:
1. **Uniqueness**: Customer_ID should be unique.
2. **Data Type**: Ensure correct data types for each field.
3. **Value Range**: Age should be within a realistic range (0-120).
4. **Consistency**: Names should not contain special characters unless verified.

   Order Dataset
   Necessary Data Components:
1. **Order_ID**: Unique identifier for each order (mandatory, integer, non-null).
2. **Item**: Name of the item purchased (mandatory, string, non-null).
3. **Amount**: Purchase amount (mandatory, float, non-null, non-negative).
4. **Customer_ID**: Identifier linking to the Customer dataset (mandatory, integer, non-null).

   Data Quality Criteria:
1. **Uniqueness**: Order_ID should be unique.
2. **Data Type**: Ensure correct data types for each field.
3. **Value Range**: Amount should be non-negative.
4. **Consistency**: Customer_IDs should exist in the Customer dataset.

Shipping Dataset

Necessary Data Components:

1. **Shipping_ID**: Unique identifier for each shipping record (mandatory, integer, non-null).
2. **Status**: Shipping status (e.g., Pending, Delivered) (mandatory, string, non-null).
3. **Customer_ID**: Identifier linking to the Customer dataset (mandatory, integer, non-null).

Data Quality Criteria:

1. **Uniqueness**: Shipping_ID should be unique.

2. **Data Type**: Ensure correct data types for each field.
3. **Consistency**: Customer_IDs should exist in the Customer dataset.

B. Develop the data models to effectively organise and structure the information and provide a detailed mapping of existing data flows, focussing on the areas of concern.

Entities and Relationships:

Primary Keys and Foreign Keys

- **Customer Table**:

  Primary Key: Customer_ID

- **Order Table**:

  Primary Key: Order_ID

  Foreign Key: Customer_ID (links to Customer_ID in the Customer table)

- **Shipping Table**:

  Primary Key: Shipping_ID

  Foreign Key: Customer_ID (links to Customer_ID in the Customer table)

1. **Customer Entity:**

   Attributes:

   - Customer_ID: Integer, Primary Key
   - First_Name: String
   - Last_Name: String
   - Age: Integer
   - Country: String

   Description: Stores basic information about customers.

2. **Order Entity:**

   Attributes:

   - Order_ID: Integer, Primary Key
   - Item: String
   - Amount: Float
   - Customer_ID: Integer, Foreign Key

   Description: Records details of each order placed by customers.

3. **Shipping Entity:**

Attributes:

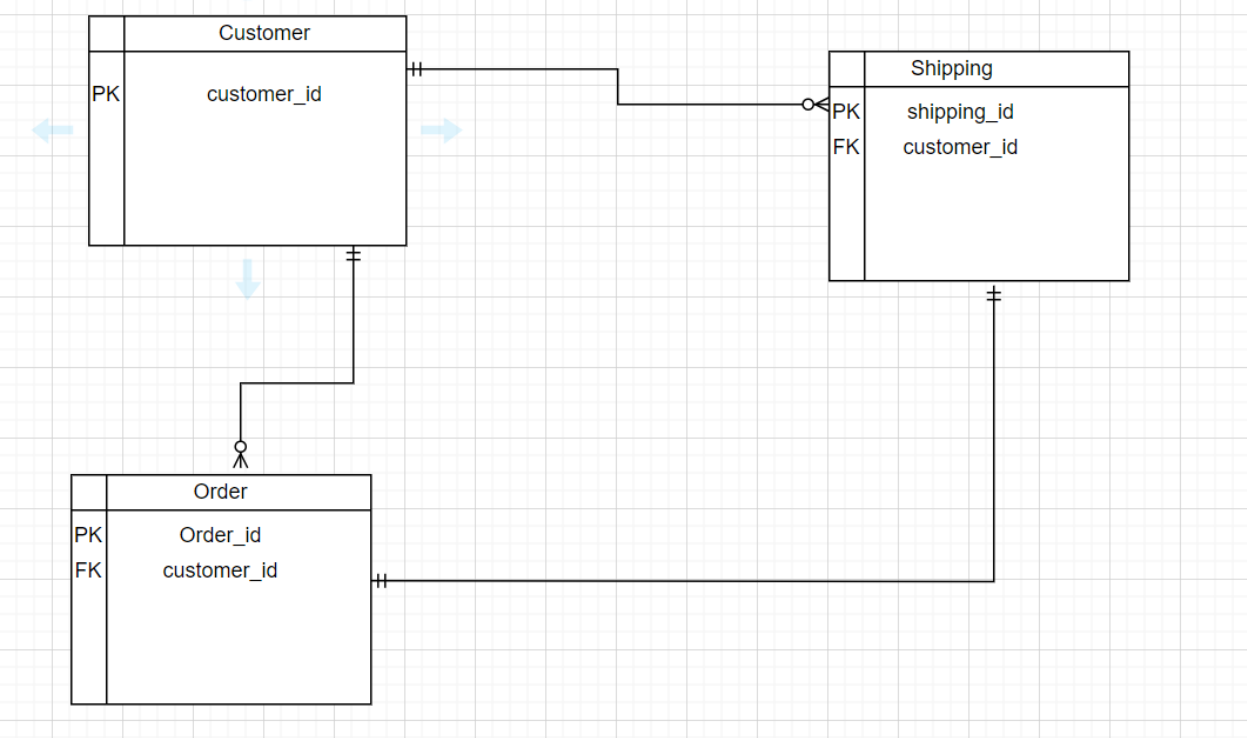
- Shipping_ID: Integer, Primary Key
- Status: String
- Customer_ID: Integer, Foreign Key

Description: Stores shipping status of each order.

2: Relationships

- Each customer can place multiple orders (one-to-many relationship between Customer and Order).
- Each order can have multiple shipping statuses, but this data is less clear. For simplicity, we assume a one-to-one relationship for now.
- Each customer can have 1 or many optional shipping statuses

Data Model:



C. Prepare a story with technical specifications for one part of the data model for a data engineer.

Story: Customer Data Integration

As a Data Engineer
I want to integrate and model customer data
So that I can provide a comprehensive and accurate customer profile to support business analytics and reporting.

Technical Specifications

1. Customer Data Source

Tables Involved:

- Customer
- Order
- Shipping

2. Data Model

Entity: Customer

The Customer entity will be the central part of our data model. It will include personal information, order history, and shipping status. The following attributes will be part of the Customer entity:

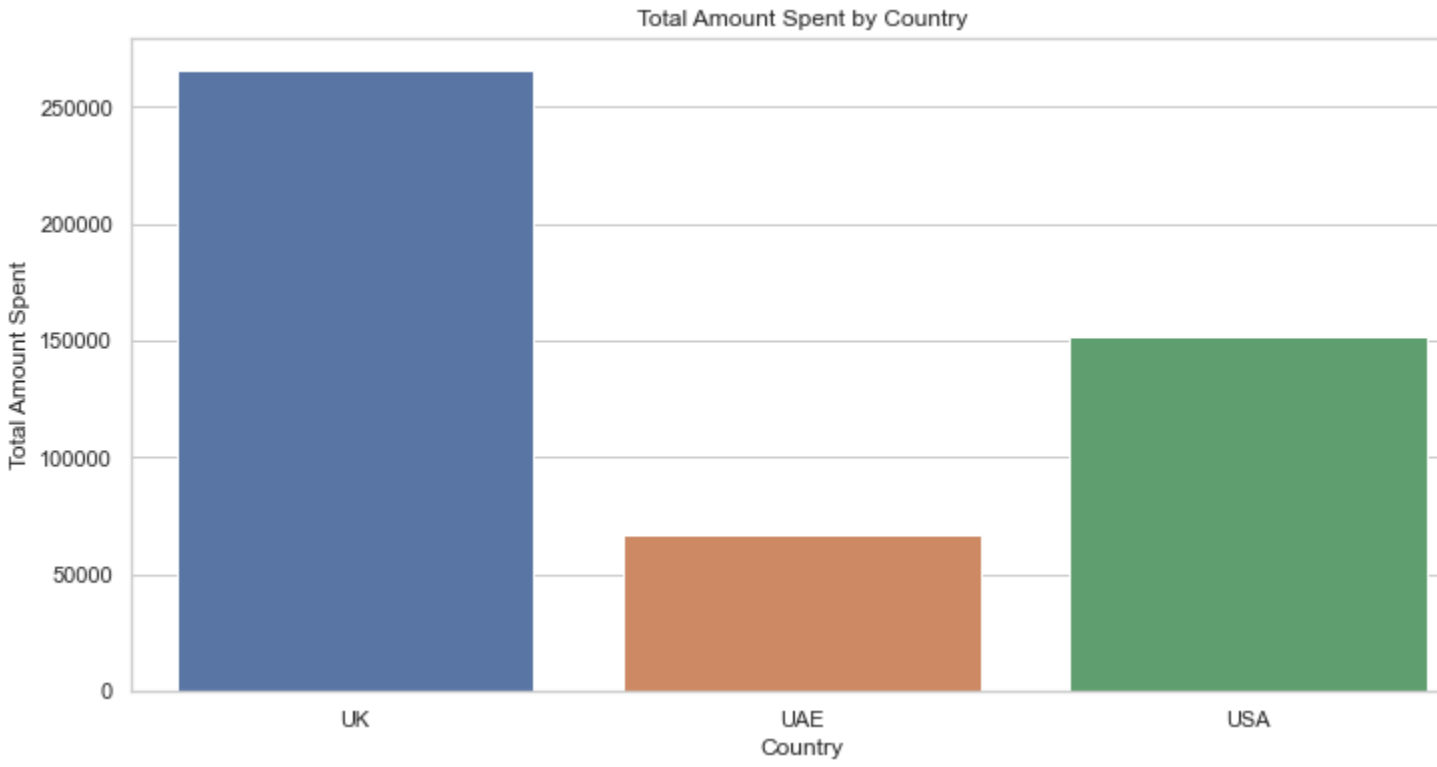
- **Customer_ID**: Unique identifier for each customer.
- **First**: First name of the customer.
- **Last**: Last name of the customer.
- **Age**: Age of the customer.
- **Country**: Country of the customer.
- **Total_Transactions**: Total number of orders placed by the customer.
- **Total_Amount_Spent**: Total amount spent by the customer.
- **Last_Order_Date**: Date of the most recent order.
- **Last_Order_Amount**: Amount of the most recent order.
- **Shipping_Status**: Status of the most recent shipping activity.

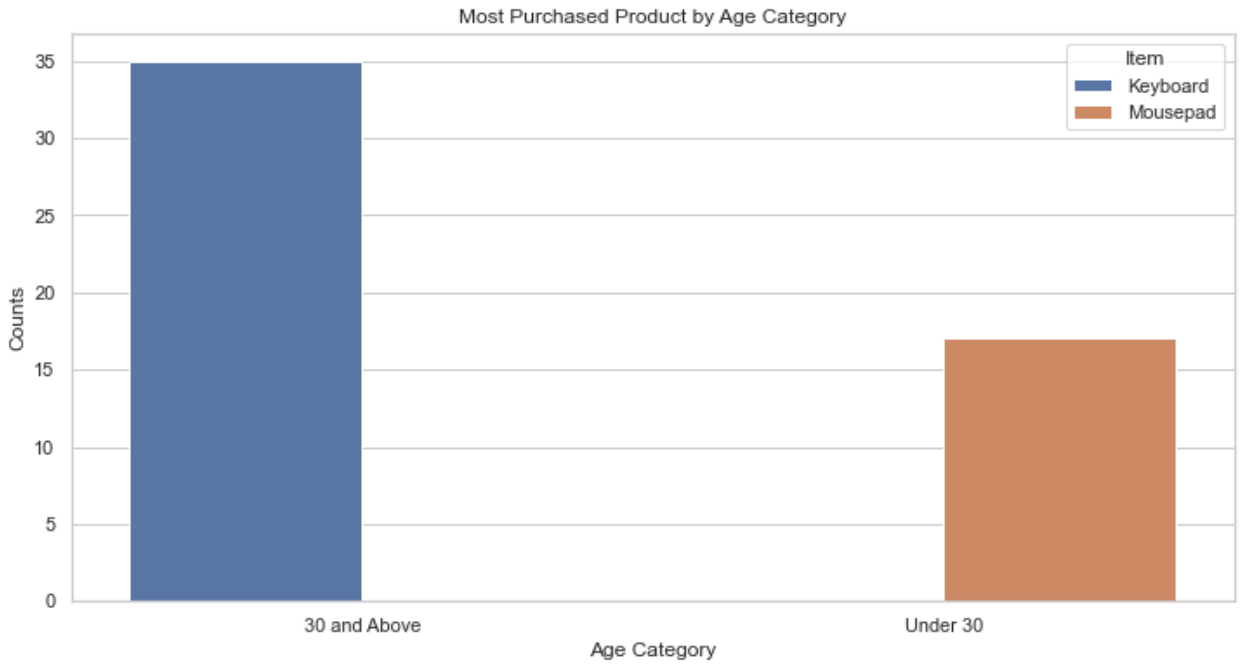
3. Data Flow and Transformation

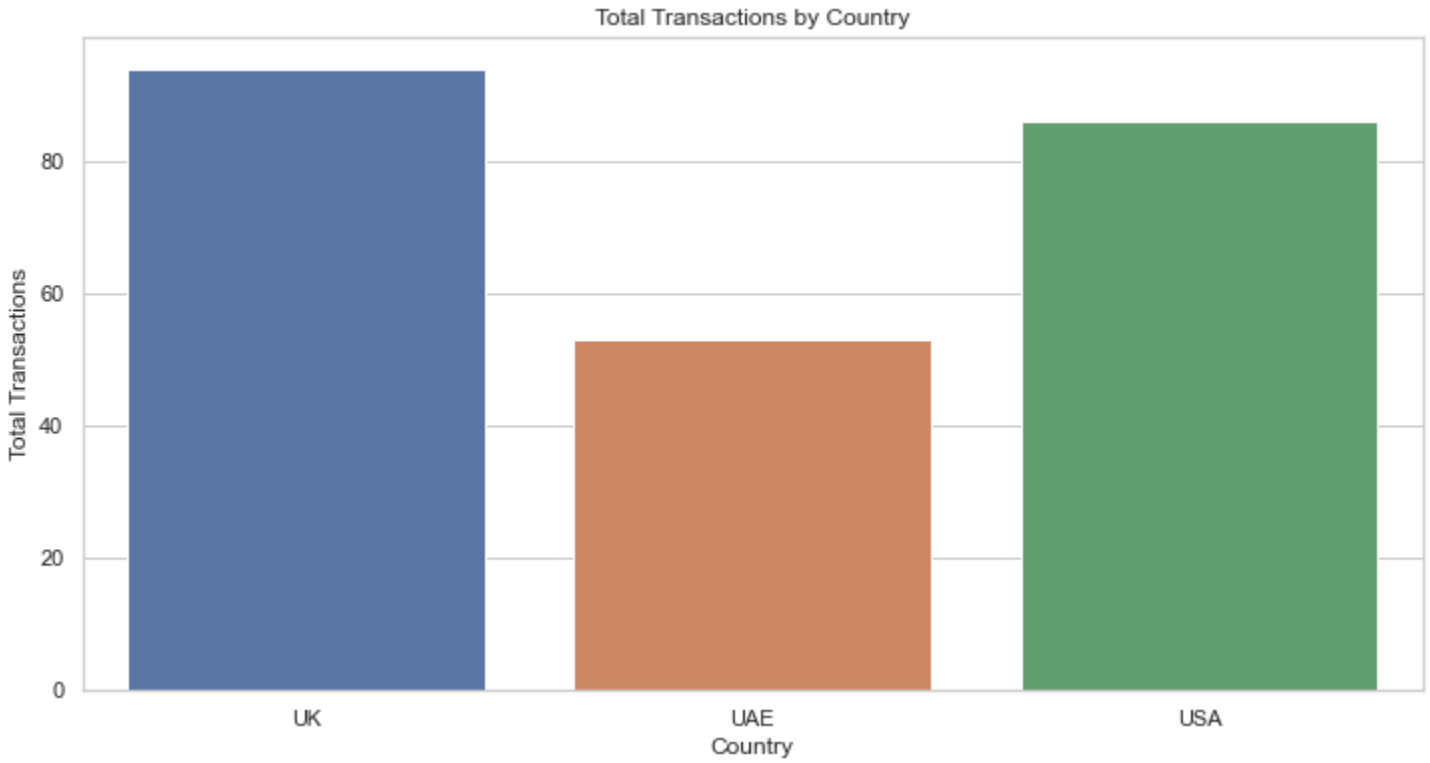
1. **Extract Data**:
    - Extract customer data from the Customer table.
    - Extract order data from the Order table.
    - Extract shipping data from the Shipping table.

2. **Transform Data**:
    - Calculate Total_Transactions for each customer by counting the number of orders in the Order table.
    - Calculate Total_Amount_Spent for each customer by summing the Amount from the Order table.
    - Identify the Last_Order_Date and Last_Order_Amount for each customer from the Order table.
    - Determine the Shipping_Status from the Shipping table by looking up the latest shipping record for each customer.
3. **Load Data**:
    - Load the transformed data into the Customer entity in the data warehouse.

D. Communicate the findings and insights to stakeholders in a visually comprehensive manner.




Most Purchased Product by Age Category
Item
Keyboard
Mousepad
Counts
0
5
10
15
20
25
30
35
30 and Above
Under 30
Age Category


Total Transactions by Country
Total Transactions
0
20
40
60
80
UK
UAE
USA
Country