



Aman Mushirul

Bengaluru

mushirul.aman@gmail.com

+91 9162178254

[Data Expert / Data Solution Consultant@PWC, Ex-Deloitte, VIT University - B.Tech CSE]

Hiring Manager

PEI GROUP - PEDDA

Dear Hiring Manager,

As a Senior Data Expert, I have been responsible for ensuring that data was accurate, complete, and reliable to build a solid foundation for decision-making processes. Below, I will detail how I executed these responsibilities using PEI assessment example, demonstrating my expertise and approach to managing **data extraction, profiling, cleansing, transformation, modeling, and reporting** it for fact based and data driven decision making for business growth.

1. Verified the Accuracy, Completeness, and Reliability of Source Data

During the assessment with PEI sales data project, I first verified the accuracy, completeness, and reliability of the source data. This involved:

- **Accuracy:** I cross-referenced sales data from "**Customers.xls**," "**Orders.csv**," and "**Shippings.json**" to identify discrepancies or anomalies. For example, I ensured that customer, order details matched the shipping records.
- **Completeness:** I ensured all required fields were populated, checking for missing data points, such as incomplete customer information or orders without shipping records.
- **Reliability:** I assessed the consistency of the data by comparing historical records and business expectations. When I noticed significant deviations in sales data that were not aligned with forecasts, I investigated further to detect potential data entry errors.

Through these efforts, I ensured that the data was prepared for analysis and reported the findings to stakeholders, providing transparency about any issues identified and the solutions

2. Outlined Requirements for Anticipated Datasets

Once the data was verified, I outlined the requirements for the anticipated datasets. In this project, I identified the key components necessary for analysis, such as:

- **Required Data Components:** I identified components like customer IDs, product details, sales dates, payment methods, and shipping costs.
- **Data Transformation:** I standardized data from multiple formats by ensuring consistency in fields such as Names and CustomerID. This enabled me to merge the customer, order, and shipping data into a single, coherent dataset.

Close & Apply

New Source

Recent Sources

Enter Data

Data source settings

Manage Parameters

Refresh Preview

Properties

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Sort

Split Column

Group By

Use First Row as Headers

Replace Values

Data Type: Text

Close

New Query

Data Sources

Parameters

Query

Manage Columns

Reduce Rows

Transform

Queries [3]

Customer

Order

Shipping

Table.SelectRows("#Removed Duplicates", each ([First] = "Al!cia" or [First] = "L@rry" or [First] = "N!cole")

123 Customer_ID

ABC First

ABC Last

123 Age

ABC Country

	Valid	100%	Error	0%	Empty	0%		Valid	100%	Error	0%	Empty	0%		Valid	100%	Error	0%	Empty	0%		Valid	100%	Error	0%	Empty	0%		Valid	100%	Error	0%	Empty	0%
1	6	N!cole	Jones	33	USA																													
2	14	N!cole	Lara	77	UK																													
3	109	R0bert	Moore	40	UK																													
4	118	R0bert	Shepherd	28	UK																													
5	162	N!cole	Bennett	51	USA																													
6	171	L@rry	Cole	50	USA																													
7	198	R0bert	Bryan	49	UK																													
8	211	A!licia	Thompson	38	USA																													
9	214	N!cole	Mcintyre	18	UK																													
10	236	A!licia	Jensen	19	USA																													

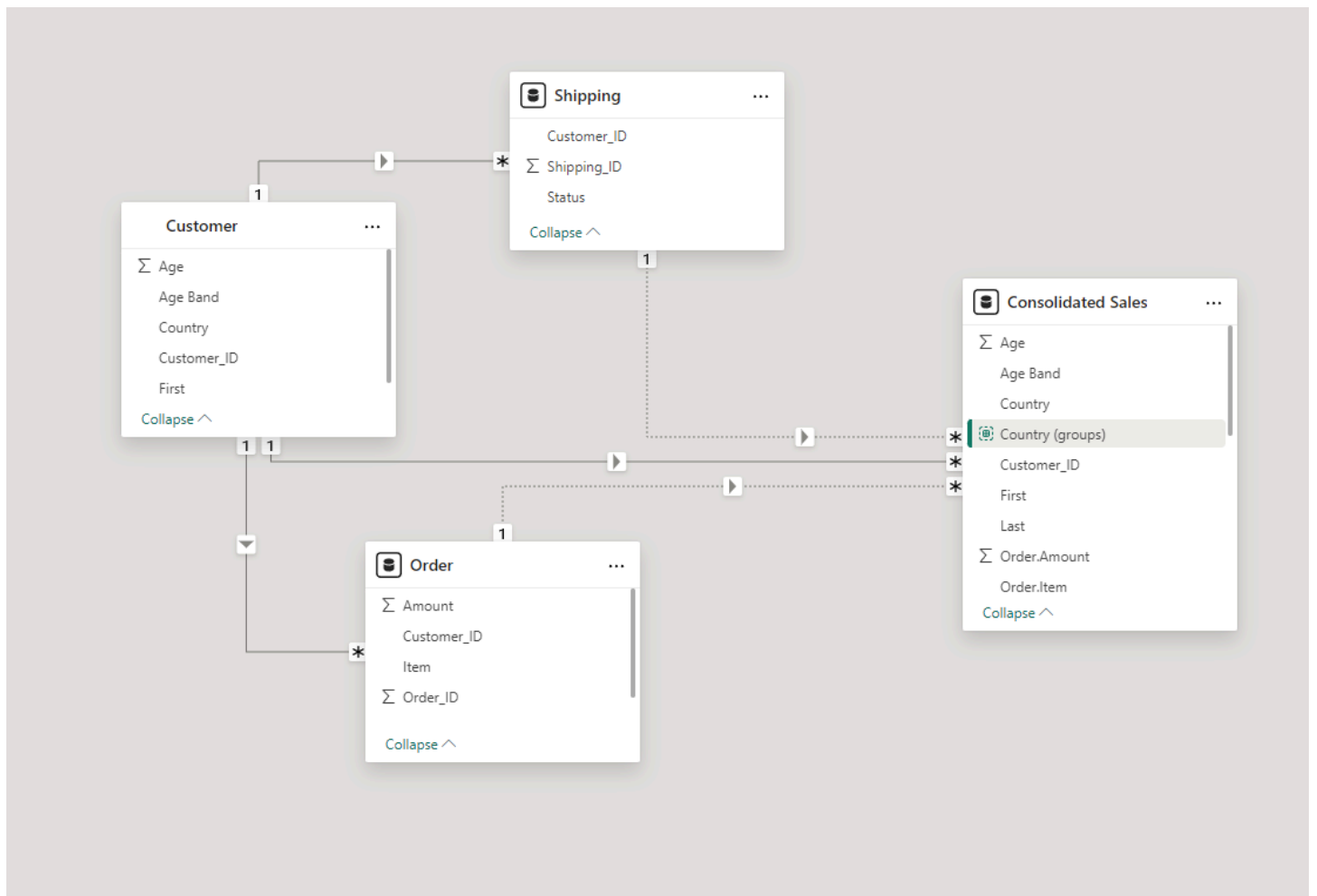
Expectation of the dataset for improved data driven decision making from data producer / business team :

<u>Dataset Name</u>	<u>Purpose</u>	<u>Required Data Components</u>	<u>Usage</u>
Customer Dataset	Analyze customer behavior, spending patterns, and demographics	<ul style="list-style-type: none"> - Customer ID - Customer Name - Country - Age - Gender - Total Spending - Transaction Count - Product Preferences 	Provides insights into customer demographics, spending habits, and segmentation analysis.
Orders Dataset	Monitor sales and order trends, focusing on delivery performance	<ul style="list-style-type: none"> - Order ID - Customer ID - Order Date - Product ID - Quantity - Order Status - Country - Total Amount - Delivery Date 	Track sales trends, highlight pending delivery issues, and optimize order processing.
Product Dataset	Analyze product popularity, sales volume, and demand	<ul style="list-style-type: none"> - Product ID - Product Name - Category - Price - Quantity Sold - Stock Availability 	Helps in inventory management, product demand analysis, and marketing strategies.
Shipping Dataset	Manage and analyze shipping details and performance	<ul style="list-style-type: none"> - Order ID - Shipping ID - Shipping Date - Delivery Date - Shipping Status - Shipping Cost - Delivery Country 	Tracks shipping performance, identifies delays, and optimizes logistics operations.
Transaction Dataset	Capture financial transactions for	<ul style="list-style-type: none"> - Transaction ID - Order ID 	Enables tracking of revenue, payments,

	reporting and analysis	<ul style="list-style-type: none"> - Customer ID - Transaction Date - Payment Method - Transaction Amount - Currency 	transaction volumes, and financial forecasting.
Country Dataset	Analyze country-based metrics, such as transaction volume and sales	<ul style="list-style-type: none"> - Country ID - Country Name - Total Transactions - Total Sales - Average Transaction Value - Pending Orders - Most Purchased Product 	Allows comparison of performance across regions, aiding in market expansion and logistics.
Demographics Dataset	Correlate demographic factors with purchasing behavior	<ul style="list-style-type: none"> - Age Group - Gender - Total Purchases - Product Preferences 	Supports analysis of product preferences by demographic, facilitating targeted marketing.
Sales Performance Dataset	Provide a holistic view of sales performance	<ul style="list-style-type: none"> - Sales Region - Product Category - Total Revenue - Sales Growth - Top Selling Products 	Offers a broad perspective on sales performance, aiding in trend identification and growth strategies.

I presented this outline to stakeholders, clearly detailing how the data would be used in further analysis, making sure both technical and non-technical team members understood the requirements.

3. Developed Data Models



I then developed the data model to structure the data in an optimized format. This included:

- **Entity-Relationship Mapping:** I built relationships between customers, orders, and shipping data, ensuring proper indexing and minimizing redundancy.
- **Data Transformation:** I created new fields, such as “Average Spending per Transaction” or the below measures to enhance the data's analytical capacity.
- **Normalization:** I normalized the data to reduce redundancy and improve consistency, ensuring that future queries would run efficiently.

Developed DAX and Measures Example:

1. Average Spending per Transaction =

DIVIDE(

SUM('Consolidated Sales'[Order.Amount]),

COUNT('Consolidated Sales'[Order.Order_ID])

)

2. Distinct Products = **DISTINCTCOUNT**('Consolidated Sales'[Order.Item])

3. Max Product Purchased by Country =

```
CALCULATE(  
  
    MAXX(  
  
        VALUES('Consolidated Sales'[Order.Item]),  
  
        CALCULATE(SUM('Consolidated Sales'[Order.Quantity]))  
  
    ),  
  
    ALLEXCEPT('Consolidated Sales', 'Consolidated Sales'[Country])  
  
)
```

4. Most Purchased Product Age < 30 =

```
CALCULATE(  
  
    MAXX(  
  
        VALUES('Consolidated Sales'[Order.Item]),  
  
        CALCULATE(SUM('Consolidated Sales'[Order.Quantity]),  
  
            'Consolidated Sales'[Age] < 30  
  
        )  
  
    )  
  
)
```

5. Most Purchased Product Age ≥ 30 =

```
CALCULATE(  
  
    MAXX(  
  

```

VALUES('Consolidated Sales'[Order.Item]),

CALCULATE(**SUM**('Consolidated Sales'[Order.Quantity]),

'Consolidated Sales'[Age] >= 30

)

)

)

6.Total Amount Spent for Pending Deliveries =

CALCULATE(

SUM('Consolidated Sales'[Order.Amount]),

'Consolidated Sales'[Shipping.Status] = "Pending"

)

7.Total Amount Spent per Customer =

SUM('Consolidated Sales'[Order.Amount])

8.Total Orders = **DISTINCTCOUNT**('Consolidated Sales'[Order.Order_ID])

9.Total Quantity Sold per Customer =

SUM('Consolidated Sales'[Order.Quantity])

10.Total Transactions = **DISTINCTCOUNT**('Consolidated Sales'[Order.Order_ID])

11.Total Transactions per Customer = **COUNTROWS**('Consolidated Sales')

Structure

Formatting

Properties

Calculations

1 Total Transactions per Customer = COUNTROWS ('Consolidated Sales')

Customer_ID

First

Last

Age

Country

Age Band

Order.Order_ID

Order.Item

Order.Amount

Order.Quantity

Shipping.Shipping_ID

Shipping.Status

Country (groups)

47

Carolyn

Green

31

USA

30 and Above

126

Harddisk

5000

1

176

Pending

Other

47

Carolyn

Green

31

USA

30 and Above

152

Mouse

300

1

139

Pending

Other

47

Carolyn

Green

31

USA

30 and Above

152

Mouse

300

1

176

Pending

Other

218

Thomas

McLaughlin

29

UK

Below 30

192

Mouse

300

1

100

Delivered

UK

228

Andrea

Velasquez

55

UK

30 and Above

56

Harddisk

5000

1

101

Pending

UK

78

Thomas

Wood

33

UK

30 and Above

47

Webcam

350

1

150

Delivered

UK

78

Thomas

Wood

33

UK

30 and Above

217

Webcam

350

1

150

Delivered

UK

85

Michele

Maxwell

80

UAE

30 and Above

164

Keyboard

400

1

103

Delivered

Other

85

Michele

Maxwell

80

UAE

30 and Above

164

Keyboard

400

1

152

Pending

Other

85

Michele

Maxwell

80

UAE

30 and Above

164

Keyboard

400

1

222

Pending

Other

85

Michele

Maxwell

80

UAE

30 and Above

216

Harddisk

5000

1

103

Delivered

Other

85

Michele

Maxwell

80

UAE

30 and Above

216

Harddisk

5000

1

152

Pending

Other

85

Michele

Maxwell

80

UAE

30 and Above

216

Harddisk

5000

1

222

Pending

Other

194

Nancy

Miller

65

UK

30 and Above

49

DDR RAM

1500

1

181

Pending

UK

50

Charles

Garcia

24

UAE

Below 30

70

Headset

900

1

127

Pending

Other

50

Charles

Garcia

24

UAE

Below 30

70

Headset

900

1

216

Pending

Other

50

Charles

Garcia

24

UAE

Below 30

245

Mousepad

250

1

127

Pending

Other

50

Charles

Garcia

24

UAE

Below 30

245

Mousepad

250

1

216

Pending

Other

70

Juan

Cruz

21

UAE

Below 30

71

Keyboard

400

1

106

Pending

Other

70

Juan

Cruz

21

UAE

Below 30

75

Mousepad

250

1

106

Pending

Other

70

Juan

Cruz

21

UAE

Below 30

131

Keyboard

400

1

106

Pending

Other

70

Juan

Cruz

21

UAE

Below 30

191

Keyboard

400

1

106

Pending

Other

135

Shawn

Johnson

18

UAE

Below 30

247

Webcam

350

1

107

Pending

Other

135

Shawn

Johnson

18

UAE

Below 30

247

Webcam

350

1

243

Delivered

Other

150

Sherry

Parsons

52

UAE

30 and Above

50

Headset

900

1

109

Pending

Other

192

David

Arnold

62

USA

30 and Above

52

Mouse

300

1

125

Delivered

Other

192

David

Arnold

62

USA

30 and Above

52

Mouse

300

1

227

Pending

Other

63

Sarah

Greer

49

UK

30 and Above

138

Mousepad

200

1

112

Pending

UK

63

Sarah

Greer

49

UK

30 and Above

177

Webcam

350

1

112

Pending

UK

63

Sarah

Greer

49

UK

30 and Above

242

Mouse

300

1

112

Pending

UK

191

Joseph

Miller

36

USA

30 and Above

128

Mousepad

200

1

113

Delivered

Other

110

Tiffany

Pearson

77

UAE

30 and Above

248

Mousepad

200

1

115

Delivered

Other

246

Justin

Stewart

73

USA

30 and Above

160

Headset

900

1

117

Pending

Other

Data

Search

Consolidated Sales

Age

Age Band

Country

Country (groups)

Customer_ID

Distinct Products

First

Last

Max Product Purchased by Country

Most Purchased Product Age < 30

Most Purchased Product Age ≥ 30

Order Amount

Order.Item

Order.Order_ID

Order.Quantity

Shipping.Shipping_ID

Shipping Status

Total Amount Spent for Pending Deliveries

Total Amount Spent per Customer

Total Orders

Total Quantity Sold per Customer

Total Transactions

Total Transactions per Customer

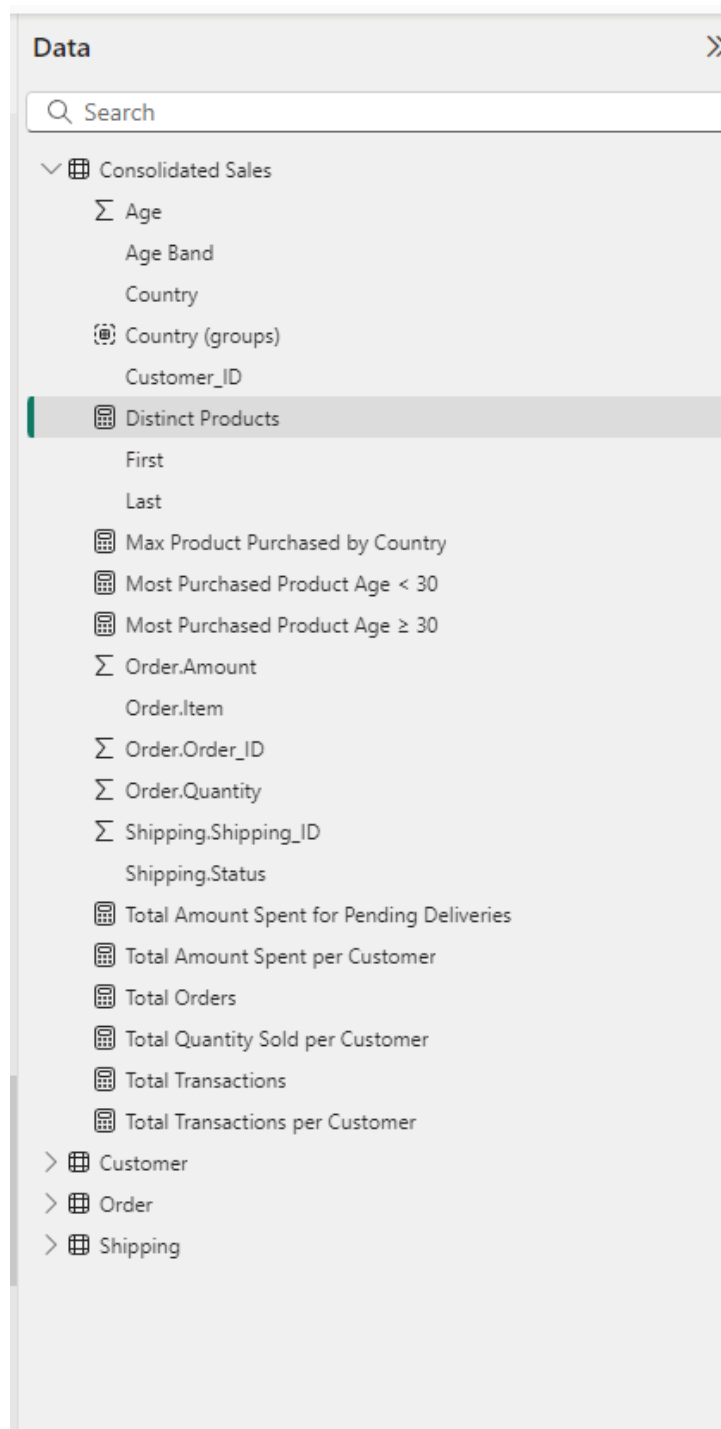
Customer

Order

Shipping

Search

- Consolidated Sales
 - Age
 - Age Band
 - Country
 - Country (groups)
 - Customer_ID
 - Distinct Products
 - First
 - Last
 - Max Product Purchased by Country
 - Most Purchased Product Age < 30
 - Most Purchased Product Age ≥ 30
 - Order_Amount
 - Order_Item
 - Order_Order_ID
 - Order_Quantity
 - Shipping_Shipping_ID
 - Shipping_Status
 - Total Amount Spent for Pending Deliveries
 - Total Amount Spent per Customer
 - Total Orders
 - Total Quantity Sold per Customer
 - Total Transactions
 - Total Transactions per Customer
- Customer
 - Order
 - Shipping



1. Customer Table

<u>Attribute</u>	<u>Data Type</u>	<u>Constraints</u>	<u>Description</u>
Customer_ID	Integer	Primary Key, Not Null	Unique identifier for each customer
Age	Integer	Not Null	Customer's age
Age Band	Text	Not Null	Grouping of customers based on age (e.g., <30, 30+)
Country	Text	Not Null	Customer's country of residence
First	Text	Optional	First name of the customer
Last	Text	Optional	Last name of the customer

Notes:

- Customer_ID uniquely identifies a customer.
- Country can be used for geographic analysis and segmentation.
- Age Band helps in analyzing customer purchases by age group.

2. Order Table

<u>Attribute</u>	<u>Data Type</u>	<u>Constraints</u>	<u>Description</u>
Order_ID	Integer	Primary Key, Not Null	Unique identifier for each order
Customer_ID	Integer	Foreign Key (Customer.Customer_ID), Not Null	Links the order to the corresponding customer
Item	Text	Not Null	Product or service ordered
Amount	Decimal	Not Null	Total amount for the order
Order_Date	Date	Not Null	Date when the order was placed

Notes:

- **Customer_ID** as a Foreign Key creates a relationship with the Customer table.
- Order_ID is a unique identifier for each transaction.
- The Amount represents the total amount spent for each order, which will help in KPIs like "Average Spending per Transaction."
- Order_Date helps in time-based analysis.

3. Shipping Table

<u>Attribute</u>	<u>Data Type</u>	<u>Constraints</u>	<u>Description</u>
Shipping_ID	Integer	Primary Key, Not Null	Unique identifier for each shipping record
Customer_ID	Integer	Foreign Key (Customer.Customer_ID), Not Null	Links the shipping record to the corresponding customer
Status	Text	Not Null	Shipping status (e.g., "Pending", "Completed")
Shipping_Date	Date	Optional	Date when the shipment was made (if applicable)

Notes:

- **Customer_ID** links each shipping record to a customer.
- The Status field is critical for analyzing pending or completed shipments.
- Shipping_Date can help track shipment timelines.

4. Consolidated Sales Table

<u>Attribute</u>	<u>Data Type</u>	<u>Constraints</u>	<u>Description</u>
Customer_ID	Integer	Foreign Key (Customer.Customer_ID), Not Null	Links the record to the corresponding customer
Age	Integer	Not Null	Customer's age (for additional verification)
Age Band	Text	Not Null	Grouping of customers based on age
Country	Text	Not Null	Customer's country of residence
Country (groups)	Text	Not Null	Groups countries based on geography or region for further segmentation
Order_Amount	Decimal	Not Null	Total amount spent by the customer on their order
Order_Item	Text	Not Null	Product or service ordered
Total_Orders	Integer	Computed Field	Number of total transactions made by the customer

Notes:

- This table consolidates all essential information about customer demographics, order amounts, and items purchased, allowing for streamlined querying and reporting.
- **Total_Orders** is a computed field that counts all transactions made by a customer.

Relationships:

1. **Customer ↔ Order (1:n)**

- One customer can place multiple orders.
- Customer_ID acts as a foreign key in the Order table, ensuring that each order is linked to a specific customer.

2. **Customer ↔ Shipping (1:1):**

- Each customer can have only one corresponding shipping record at a time.
- Customer_ID in Shipping is a foreign key that references Customer.Customer_ID.

3. **Order ↔ Shipping (1:n)**

- A single shipping record can link to multiple orders.
- Shipping_ID in the Order table establishes this relationship.

4. **Consolidated Sales ↔ Customer (1:1)**

- Consolidated Sales aggregates customer details and spending, with Customer_ID acting as a foreign key.

Data Model Concerns & Solutions:

- **Data Consistency:** Ensure that all orders have corresponding customer and shipping data, and there are no orphan records.
 - **Solution:** Data integrity constraints and Foreign Key relationships must be enforced in the database.
- **Data Completeness:** Ensure no null values in mandatory fields like Order_Amount, Country, and Status.
 - **Solution:** Implement data validation checks before inserting records into tables.
- **Pending Deliveries:** Analysis of all pending orders and their geographical distribution is important.
 - **Solution:** Use the Status field in the Shipping table to filter pending orders and link them back to customers and countries.

ISSUES IN THE DATAMODEL :

STEP 1 : Profiling

There are Customer_ID from CUSTOMERS which are not present in ORDER and SHIPPING tables resulting in getting NULL records while creating a Golden Consolidation Sheet.

One or more of the datasets could be incomplete, meaning some customers may not have corresponding orders or shipments, resulting in null values during the join.

If a **Customer_ID** exists in the **Customers** table but not in the **Orders** table, it means that customer hasn't placed any orders. This is expected behavior, as some customers may not have ordered yet.

If a **Customer_ID** exists in the **Orders** table but not in the **Shipping** table, it could indicate an issue in the shipping process or that the order hasn't been processed yet.

STEP 2 : Data Cleansing (Getting the data ready for Reporting)

Exclude from Analysis: If these missing customers are minimal or irrelevant, we can filter out these rows in our BI dashboards.

Investigate Source: It's important to investigate why these customers are missing in the first place. Ensure the ETL process correctly maps customers to orders.

Solution for correct reporting : Do inner join on the 3 tables and only take relevant data with not NULL values (Creation of Consolidates Sales table which is the source of Truth for Dashboard visualization and reporting.

4. Mapped Data Flows and Highlighted Areas of Concern

I mapped the data flows, visualizing how the customer, order, and shipping data traveled from their respective sources into the final analysis-ready dataset. This included:

- **Source Integration:** I outlined how the data from multiple sources was combined and transformed to align.
- **Data Transformation Stages:** I visualized how the data was cleaned and standardized, ensuring that stakeholders understood each step.
- **Highlighted Areas of Concern:** I flagged any issues, such as discrepancies between order dates and shipping times, and document these concerns to present to Data Engineering team and Leadership.

This mapping helped ensure that potential bottlenecks were addressed before the final analysis phase.

5. Prepared Technical Specifications for Data Engineers

Part: Shipping & Pending Orders Flow

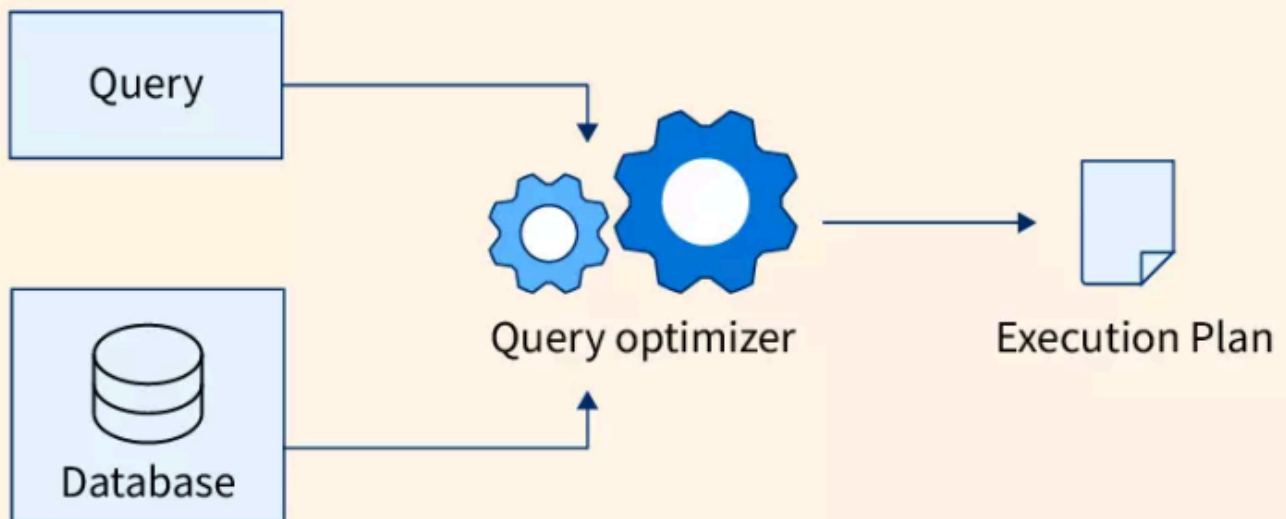
- **Objective:** The goal is to track pending orders efficiently, ensuring that customers whose orders have not yet been shipped are flagged and analyzed geographically.
- **Data Flow:**
 - Use the Shipping.Status column to identify pending deliveries.
 - Join the Shipping table with the Order and Customer tables via Customer_ID to determine the total number of pending orders for each country.
 - Implement logic to compute pending delivery percentages by region.

SQL for Pending Deliveries:

```
SELECT c.Country, COUNT(o.Order_ID) AS Total_Pending_Orders  
  
FROM Order o  
  
JOIN Shipping s ON o.Customer_ID = s.Customer_ID  
  
JOIN Customer c ON o.Customer_ID = c.Customer_ID  
  
WHERE s.Status = 'Pending'  
  
GROUP BY c.Country;
```

For this project, the data engineering team needs to take a look on the below factors. This includes:

- **Indexing for Query Optimization:** I specified which fields needed indexing, such as customer IDs and order dates, to ensure faster querying.
- **Handling Large Datasets:** I outlined strategies for partitioning large datasets, enhancing storage and retrieval speed.
- **Data Validation Rules:** I established validation rules for incoming data, ensuring consistency in future dataset updates.



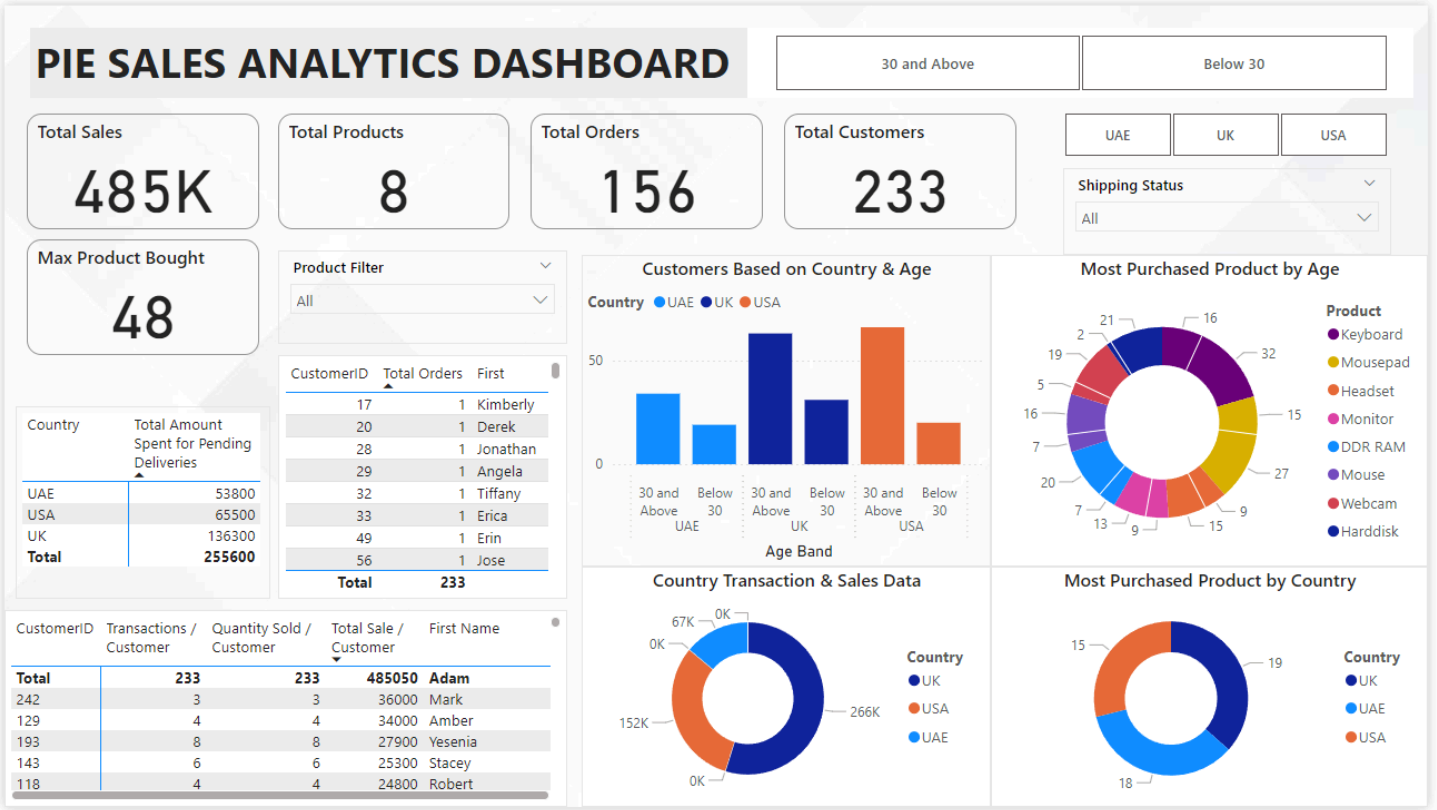
These specifications were documented clearly, enabling the data engineering team to implement the necessary solutions effectively.

6. Communicated Findings and Insights to Stakeholders

These reports included:

- KPIs: I displayed key metrics, such as "Total Sales," "Average Order Value," and "Revenue by Region."
- DAX Measures: I used DAX to develop dynamic insights, such as monthly sales trends and customer segmentation.

- Visual Storytelling: I presented insights in a clear and compelling format, ensuring that stakeholders, including executives, could easily grasp the analysis and make informed decisions.




<u>Insight Category</u>	<u>Description</u>
1. Pending Delivery Status by Country	
Highest Pending Deliveries	Country UK has the highest number of pending deliveries, with a total amount spent of \$136,300 . This indicates potential delays in the supply chain for this region.
Lowest Pending Deliveries	Country UAE has the fewest pending deliveries, with a total amount spent of \$53,800 . This suggests more efficient logistics or lower demand in this area.
2. Customer Spending Patterns	
Top Spending Customer	The top customer, Customer ID 242 (Mark) , has spent \$36,000 across 3 transactions, purchasing 3 units of various products. These customers could be targeted for loyalty programs.
Most Transactions by a Customer	Customer 193 (Yesnia) completed the most transactions, showing a high frequency of purchases but with an average spending lower than Customer 242 (Mark) .
3. Country-Level Product Demand	
Maximum Product Purchased by Country	Product Keyboard is the most purchased in Country UK , accounting for 25% of total sales. This suggests a high demand and opportunity for localized marketing strategies.
4. Age-Based Product Preference	
Under 30 Age Group	Customers under the age of 30 primarily purchased Product Keyboard , accounting for 6.87% of their total purchases. This

product could be marketed more heavily towards younger customers.

Conclusion

In this PEI sales data project, I not only ensured that the data was accurate and well-structured but also developed data models that optimized performance and delivered valuable insights to stakeholders. I was able to deliver a solution that was both technically sound and aligned with business objectives.

As a Data Analyst, my expertise in verifying data quality, structuring data models, and communicating findings allowed me to drive meaningful outcomes for the organization. I ensured that the data solutions were scalable, reliable, and tailored to the company's strategic needs.

**PLEASE CHECK OUT BI
TECHNICAL SPECIFICATION
DOCUMENT FOR DETAILS AND
SCREENSHOTS **