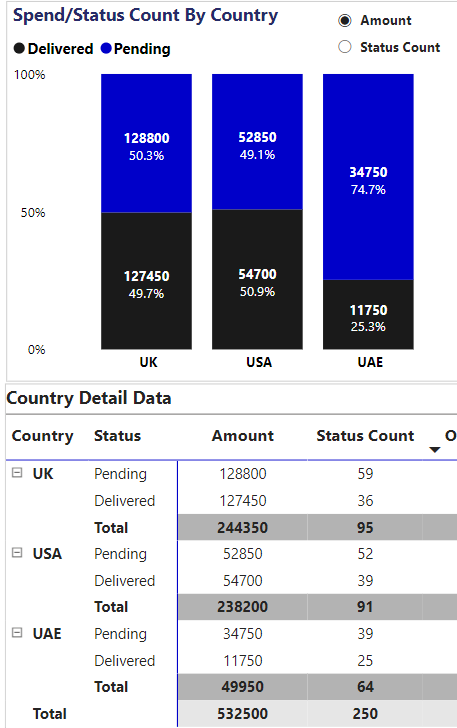
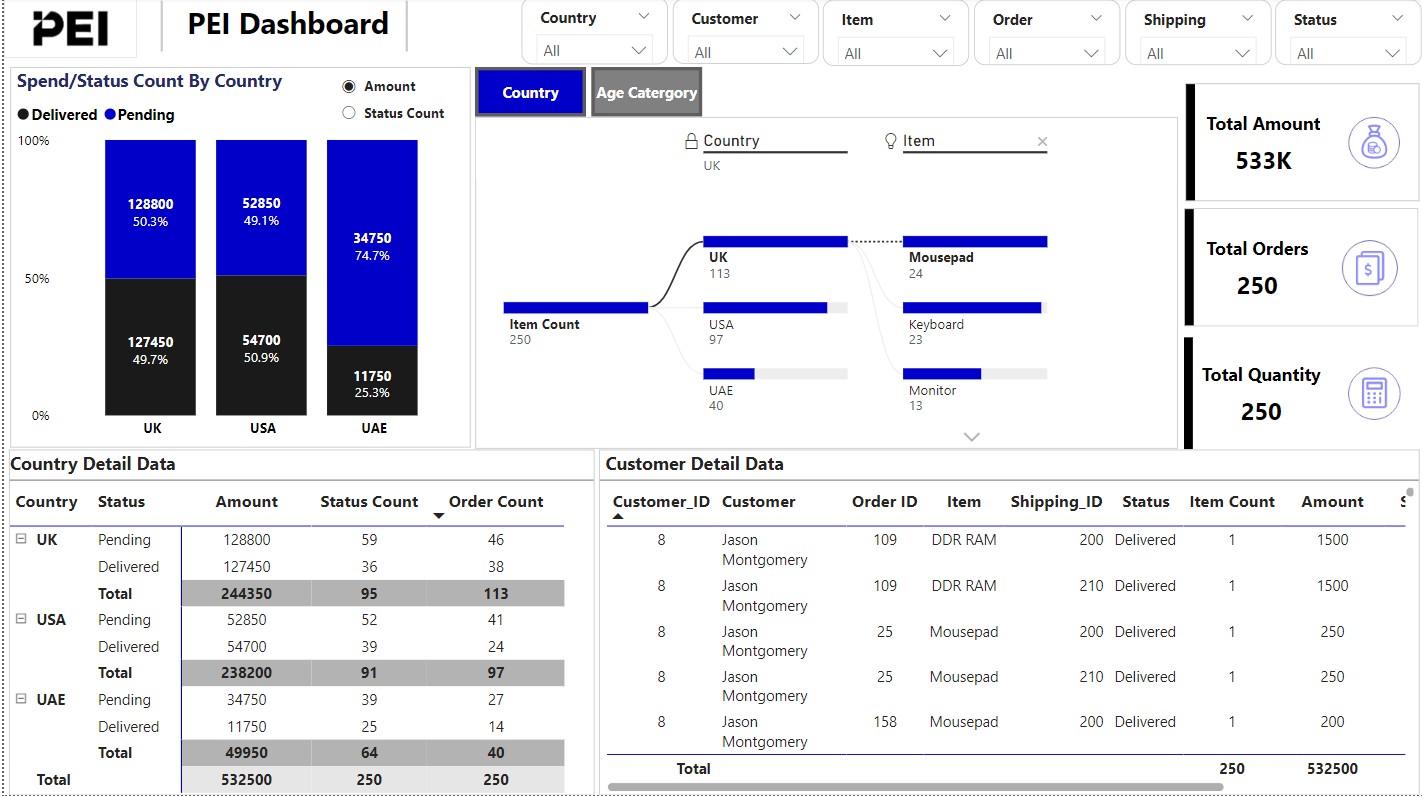


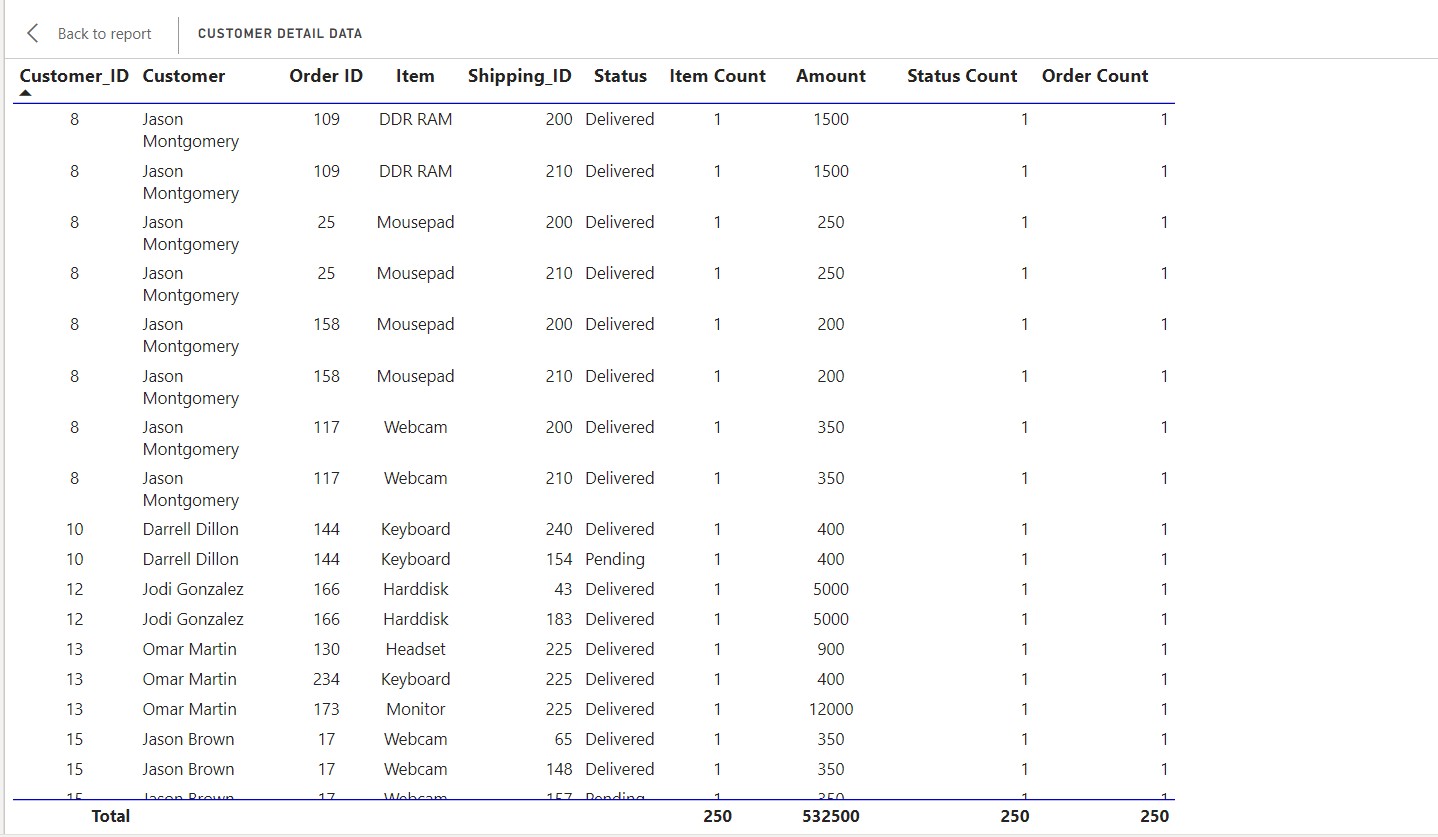
### SHISHIR SHETTY



1. THE TOTAL AMOUNT SPENT AND THE COUNT FOR THE PENDING DELIVERY STATUS FOR EACH COUNTRY
2. Total Amount Spent for each country and Total Amount Spent O verall is displayed in the Country Detail Data with the highlighted value.
3. Count for the Pending Delivery Status and Count for Delivered Delivery Status

Is displayed in the 100% stacked column chart.

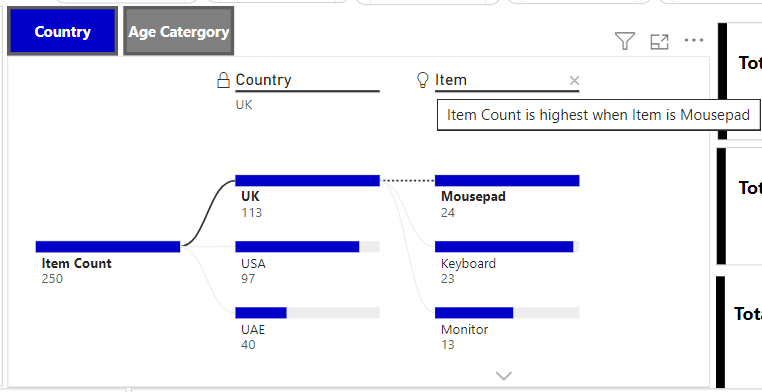
It also displayed the Percentage of Total between the two Status



### THE TOTAL NUMBER OF TRANSACTIONS, TOTAL QUANTITY SOLD, AND TO TAL AMOUNT SPENT FOR EACH CUSTOMER, ALONG WITH THE PRODUCT *DETAILS.*

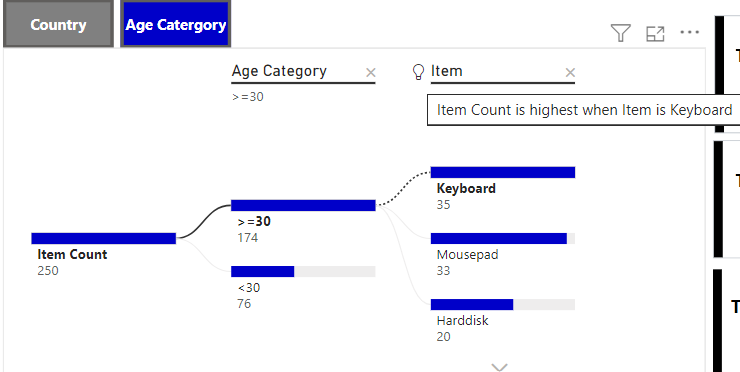
Total Number of Transactions(Order Count) , Total Quantity Sold(Item Count), Total Amount Spent(Amount) is displayed for each customer along with the Product Details in Customer Detail Data Table

## THE MAXIMUM PRODUCT PURCHASED FOR EACH COUNTRY.



Maximum Product purchased for each Country is displayed in the Decomposition Tree / Hierarchy Chart with the information tooltip

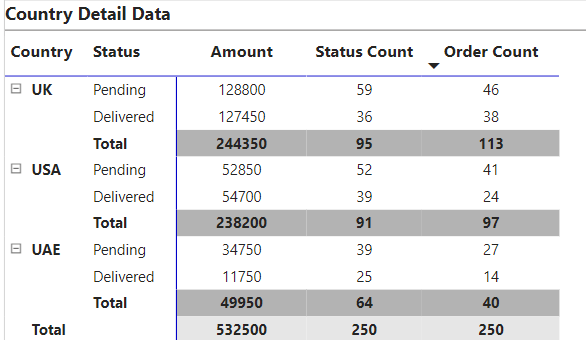
### THE MOST PURCHASED PRODUCT BASED ON THE AGE CATEGORY LESS THAN 30 AND ABOVE



30.

Most purchased product based on the age category is displayed in the Decomposition Tree / Hierarchy Chart with the information tooltip

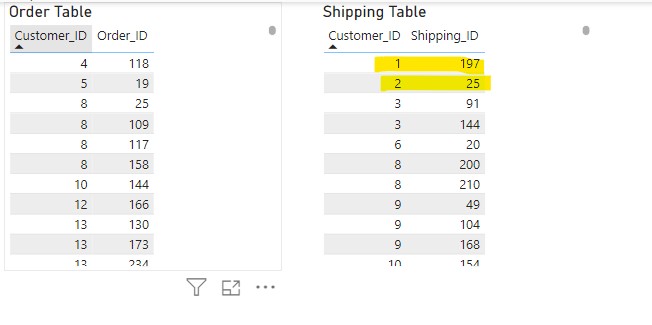
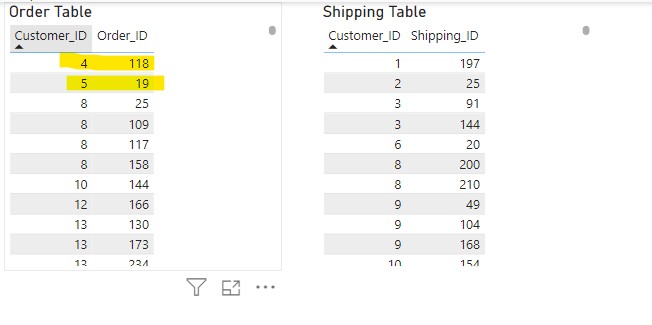
## THE COUNTRY THAT HAD MINIMUM TRANSACTIONS AND SALES AMOUNT.



The Country that had minimum transactions(Order Count) and sales amount(Amount) is displayed in Country Detail Data Table

# AS A DATA ANALYST, YOU ARE REQUIRED TO?

8



* 1. VERIFY THE ACCURACY, COMPLETENESS, AND RELIABILITY OF SOURCE DATA.
  2. BASED ON YOUR FINDINGS, DEFINE AND OUTLINE THE REQUIREMENTS F OR ANTICIPATED DATASETS,

DETAILING THE NECESSARY DATA COMPONENTS.

**Data Integrity Constraints** : Implement constraints to ensure data integrity, such as ensuring that each order and shipping record has a corresponding customer entry in the Customer Dataset.

Concern : There are total 250 customer details in Customer Table.

* + 1. Few customers have placed order(which can be fetched from

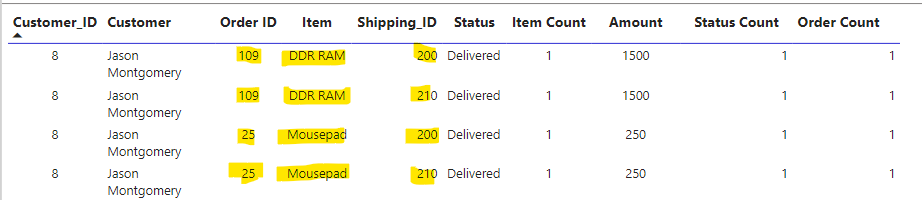
order table Eg: Customer ID = 4.)

But there is no information on their Shipping details in shipping table

* + 1. Shipping Details of few customer is present in Shipping Table (For Eg: Customer ID =1) but there is no

information on what order they placed in Order Table

* + - 1. VERIFY THE ACCURACY, COMPLETENESS, AND RELIABILITY OF SOURCE DATA.



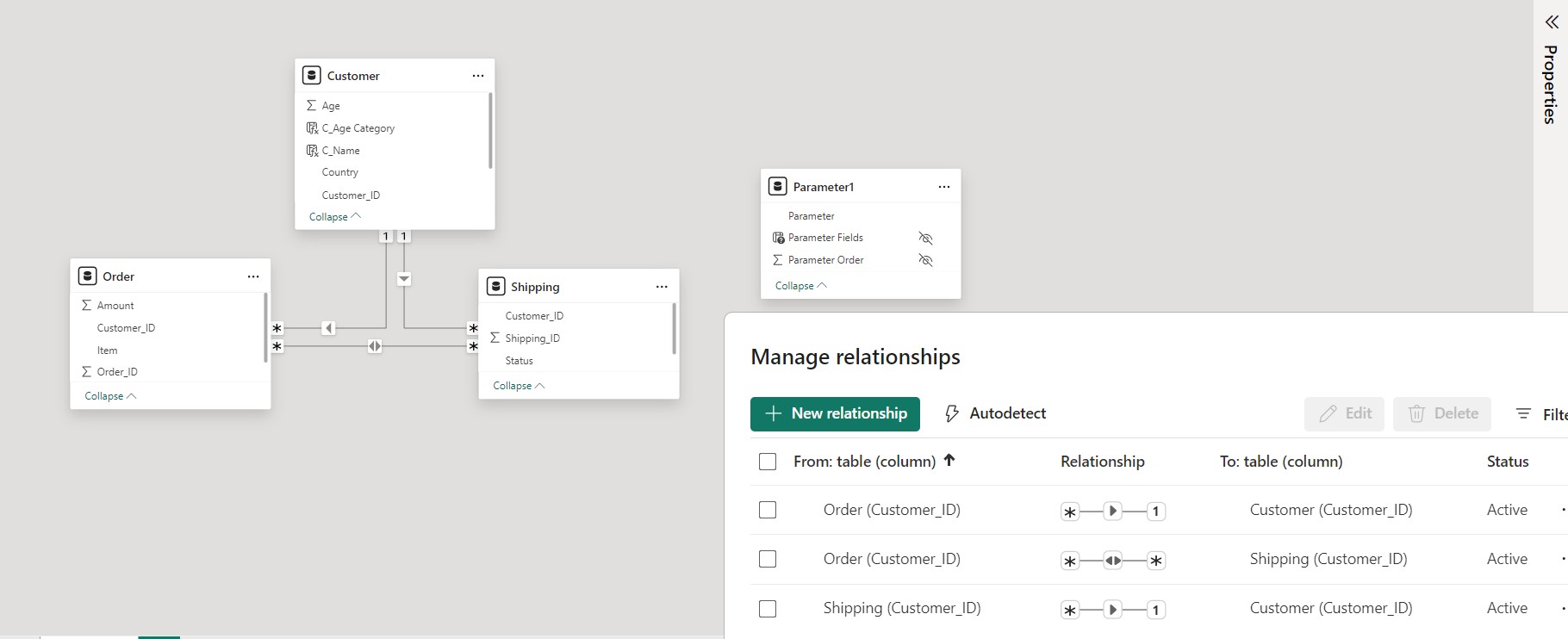
* + - 1. BASED ON YOUR FINDINGS, DEFINE AND OUTLINE THE REQUIREMENTS F OR ANTICIPATED DATASETS, DETAILING THE NECESSARY DATA COMPONENTS.

**Data Integrity Constraints**: Implement constraints to ensure data integrity, such as ensuring that each order and shipping record has a

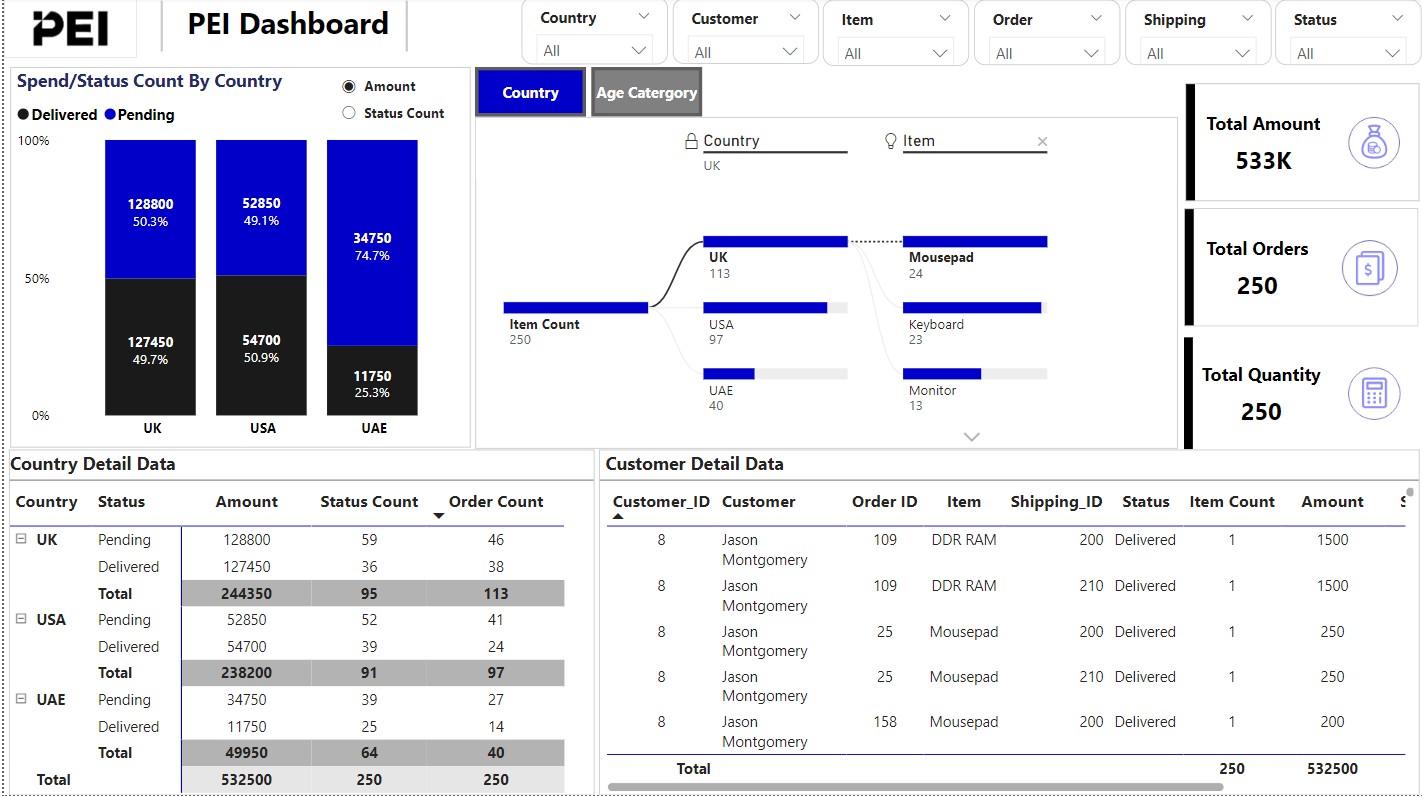
corresponding customer entry in the Customer Dataset.

Concern :

* + - * 1. Few of the customer has ordered same item with same Order ID. However, shipping ID differs. Does that mean every individu al item gets separate Shipping ID or is that an error in capturing the data?
        2. Is the Pending delivery amount received ? because it can be COD
        3. No information on returns hence we cannot exactly calculate sales amount because few items could have been returned
        4. Mousepad is the item with 2 Cost - 200, 250. Why? any coupon discount or different Mouse pad?
        5. There is no information on any date field which is big concern like Order Date, Ship Date etc
      1. 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.



* + - 1. COMMUNICATE THE FINDINGS AND INSIGHTS TO STAKEHOLDERS IN A VI SUALLY COMPREHENSIVE MANNER..



* + - 1. WHAT WILL BE YOUR INSIGHTS TO OTHER PEER TEAMS OF DATA ENGINE ERS, DATA SCIENTISTS AND OTHER TECHNICAL AND NON - TECHNICAL STAKEHOLDERS?

##### Data Integrity and Completeness:

* There are inconsistencies between the Order and Shipping tables, where some customers have placed orders but have no shipping information, and vice versa.
* This indicates potential data integrity issues or gaps in data capture processes. It's essential to investigate and rectify t hese inconsistencies to ensure the completeness and accuracy of the data.

##### Data Relationships:

* The presence of customers with orders but no shipping details, and vice versa, highlights the importance of maintaining prope r data relationships between tables.
* Ensuring referential integrity between the Customer, Order, and Shipping tables is crucial for accurate analysis and reportin g.

##### Handling Duplicate Orders:

* Identifying cases where the same item with the same Order ID has different Shipping IDs suggests potential errors in data cap ture or

processing.

* Data Engineers should investigate whether this is due to individual items receiving separate shipping IDs or if there are iss ues in data recording.
* Clarification on the handling of duplicate orders and shipping IDs is necessary for accurate data interpretation.

5 ) WHAT WILL BE YOUR INSIGHTS TO OTHER PEER TEAMS OF DATA ENGINE ERS, DATA SCIENTISTS AND OTHER TECHNICAL AND NON - TECHNICAL STAKEHOLDERS?

#### Payment and Delivery Status:

* The absence of information on pending delivery amounts and returns impacts the accuracy of sales calculations and inventory management.
* Data Scientists should collaborate with stakeholders to devise strategies for tracking pending delivery amounts, handling ret urns, and updating the relevant datasets accordingly.

#### Product Variations and Discounts:

* The existence of multiple cost values for the same item (e.g., Mousepad) raises questions about potential product variations, discounts, or errors in data entry.
* Data Engineers should investigate and clarify the reasons behind these discrepancies to ensure accurate pricing information f or analysis.

#### Missing Date Information:

* The absence of date fields (e.g., Order Date, Ship Date) is a significant concern as it hinders the ability to perform time -based analysis,

track delivery timelines, and monitor business performance over time.

* Data Engineers should work on incorporating relevant date fields into the datasets to enable comprehensive analysis and repor ting.

#### Collaboration and Data Governance:

* Encourage collaboration between Data Engineers, Data Scientists, and stakeholders to address data quality issues, establish d ata

governance policies, and implement robust data validation processes.

* Regular communication and coordination are essential for maintaining data integrity and driving informed decision -making across the organization. 14

# THANK YOU

SHISHIR SHETTY

15