PEI Assignment – Data Analyst Tasks:

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

* Examined data types, distributions, patterns, and outliers.
* Performed checks to identify anomalies, inconsistencies, or errors.
* Cross-referencing data
* Utilized data cleansing techniques to remove duplicates and build data model

No outliers or data type mismatches identified

1. **Based on your findings, define and outline the requirements for anticipated datasets, detailing the necessary data components.**

Post importing data from various sources and verifying reliability below are tech requirements:

**Data Analysis:** To connect to data sources using Power Bi connectors and analyze the data in all datasets.

**Data Transformations:** To perform transformations required to make it more reliable.

Data transformations used:

1. Removing duplicates
2. Merge
3. Indexes (To create keys)
4. Removing unwanted columns / data
5. Derived column using M query

**Data Modelling:**

* Data modelling to be performed on the datasets to achieve Star Schema
* Data Modelling to be performed in Power Query segregating the tables into different sub tables (Facts and Dimensions)
* Creating the right relationships based on tables created and cross filtration
* Verifying the data model
* To define calculated columns and measures using DAX.

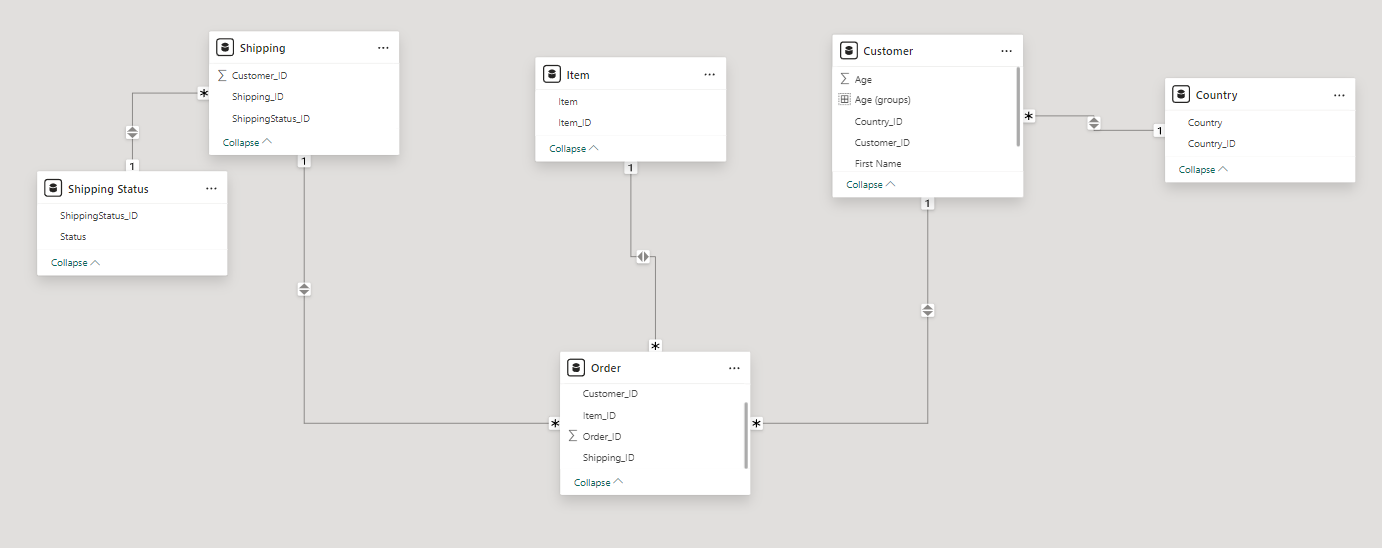
**Data Visualization and Reporting:**

* Create a report design layout
* For each requirement use (charts, graphs and tables, etc.) to present the data in a meaningful way.
* customize them with formatting options

**Data Validation:** Validate the data with Data Sources

1. **Develop the data models to effectively organize and structure the information and provide a detailed mapping of existing data flows, focusing on the areas of concern.**

Below data model has been created and can be referred in the pbix file shared



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

**-**Refer Power BI Report

1. **What will be your insights to other peer teams of Data Engineers, Data Scientists and other technical and non-technical stakeholders?**

**Non-technical stakeholders:**

* Total Amount Spent for all countries is $692,350﻿primarily driven by UK﻿ ( ﻿$322,800﻿ ) followed by USA (﻿$287,900﻿).
* ﻿UAE ﻿has the minimum pending Amount ﻿$68,050﻿ and lowest number of transactions.
* Keyboard and Mouse are the highest selling products in UK.
* Mark Roberts is the highest spending Customer among all who has bought 3 Monitors for $36000.
* Age group 30 and above have purchased Keyboard more frequently

**Data Engineers, Data Scientists and other technical stakeholders:**

* Visually speak able reports with all requirements have been created
* Filters have been provided for Customers, Country, Products & Status  
  where it can help in narrowing down the records.
* Drilldown has been made available for reports in order to get the detailed  
  result set.
* Toggle has been provided to switch between min sales and min transactions.  
   **For Data Engineers:**  
  Can create a Data Mart in the backend with the accurate data model.
* Use ETL Technology to load into backend tables with various sources  
  and maintain the CDC (change data capture) in order to avoid repeated  
  data inserts.
* Retrieving Data from the table with a single connection will provide  
  the faster output and change in data will be reflected in the report.
* Create a schedule to load the data into data mart dimensions and facts.