**PEI : Data\_Analyst\_Task**

**Data Inspection Summary**

Three datasets, `Customers.xls`, `Orders.csv`, and `Shipping.json`, were provided for analysis. Each dataset was loaded into Pandas DataFrames for thorough inspection and validation.

**Customers DataFrame:**

* Uniqueness: Ensured the `Customer\_Id` column is unique.
* Missing Values: The Data Frame was inspected for any missing values in each column.
* Missing Values: Filled missing values in the `Country` column with a default value.
* Age Validation: Verified that no customers have an age greater than 120 or negative values in the `Age` column.

**Orders DataFrame:**

* Duplicates: Checked for duplicates in the `Order\_Id` column, which serves as the primary key.
* Missing Values: The Data Frame was inspected for any missing values in each column.
* Negative Amounts: Ensured there are no negative values in the `Amount` column.

Shipping DataFrame:

* Duplicates: Checked for duplicates in the `Shipping\_Id` column.
* Normalization: Normalized the JSON data into a CSV format using the `json\_normalize` method available in Pandas for easier analysis.

**Cross-Validation:**

* Verified that all `Customer\_Id`s in the `Orders` data exist in the `Customers` data.
* Verified that all `Customer\_Id`s in the `Shipping` data exist in the `Customers` data.

**Observations**:

1. Relationship Between Orders and Shipping:

* There is no direct relationship between the `Orders` and `Shipping` tables, such as a foreign key to define the status of each order. Consequently, it is not possible to determine the exact status of each order directly from the given data.
* Despite this, I calculated the total amount and the count of pending deliveries for each country.

2. Customer Transactions and Product Details:

* While addressing the query about the total number of transactions, total quantity sold, and total amount spent by each customer along with product details, it was noted that the dataset does not contain a `Quantity` field.
* Product details were collected in a list format for each customer.

3. Maximum Product Purchased for Each Country:

* The query did not specify whether to determine the maximum product purchased based on the amount or the order count. For the analysis, the maximum product purchased was derived based on the total amount.

**Recommendations**:

* Establish a direct relationship between the `Orders` and `Shipping` tables for more precise analysis.
* Clarify analysis criteria (amount vs. order count) for better result derivation.

This approach would help us and ensure the data's accuracy and a solid base for further analysis.