Prathamesh Ashok Patil (prathmpatil.11@gmail.com)

Assignment: 1 Data Cleaning and Pre-processing.

Objective: Clean and prepare a raw dataset (with nulls, duplicates, inconsistent formats). Tools: Excel.

Summary:

Used a raw dataset named “Coffee beans sales Ras dataset” from Kaggle.

First step was to clean the data, which was done with help of Power Query Editor (PQE).

For eg. Cleaned the tables Orders, Product, Customers respectively by cleaning empty rows & columns.

Changed some column names for future reference, and for ease to work with while using formulae. Also changed the data types for columns as per the standards of data validation.

In all tables trimmed the columns Customer Name, Customer ID, Email to remove extra space.

Then removed the rows with “null” values.

Also “null” values were replaced with “Not Provided”.

After cleaning established relation between all sheets using Power Pivot.

For e.g.

* orders[Customer ID] → customers[Customer ID]
* orders[Product ID] → products[Product ID]

it forms a relational database.