**Question a)**

Graphical user interface, text, application, email

Description automatically generated

**Question b)**

There are two sheets present:

* Observations
* Products

**Question c)**

The number of non-null rows in the dataset based on each column is as follows:

* Country -> 961732
* City -> 961732
* Obs Date -> 961732
* Product Code -> 961732
* Quantity -> 961732
* Price -> 931766

**Question d)**

Before we eliminated them, there were 65 blank rows present in all columns except for the 'Price' column.

The dropna() function was executed on all columns in the dataset excluding the 'Price' column. **Prior to running the dropna() function, there were 30031 blank rows** in the 'Price' column. After running the function, it now has 29966 blank rows.

**Question e)**

There are 0 empty rows in the Products sheet.

Graphical user interface, text, application

Description automatically generated

**Question f)**

**Chart

Description automatically generated with medium confidence**

The histogram is skewed towards the right. As a result, we will use the median as the imputation method.

**Question g)**

**Graphical user interface, application

Description automatically generated**

The dataset now contains 0 null values.

**Question h)**

The countries with misspellings are as follows:

* Brazil (misspelled as Brazel)
* Indonesia (misspelled as Indonseia)
* Philippines (misspelled as Phillippines)

**Question i)**

Graphical user interface, application

Description automatically generated

We corrected the misspelled country names by mapping them to the correct names. The dataset no longer contains any incorrectly spelled country names.

**Question j)**

There are 14397 duplicate entries.

**Question K)**

The first row of each occurrence with duplicated entries is considered the original. The duplicate rows, excluding the first occurrence, are stored in the "duplicateRowsDF" variable based on all columns. The output is displayed by printing the "duplicateRowsDF" variable.

**The outcome of the executed query is the list of all duplicate entries in the dataset.**

Graphical user interface, text, application, email

Description automatically generated

**Question L)**

**Graphical user interface, text

Description automatically generated**

**Question M)**

**Graphical user interface

Description automatically generated with low confidence**