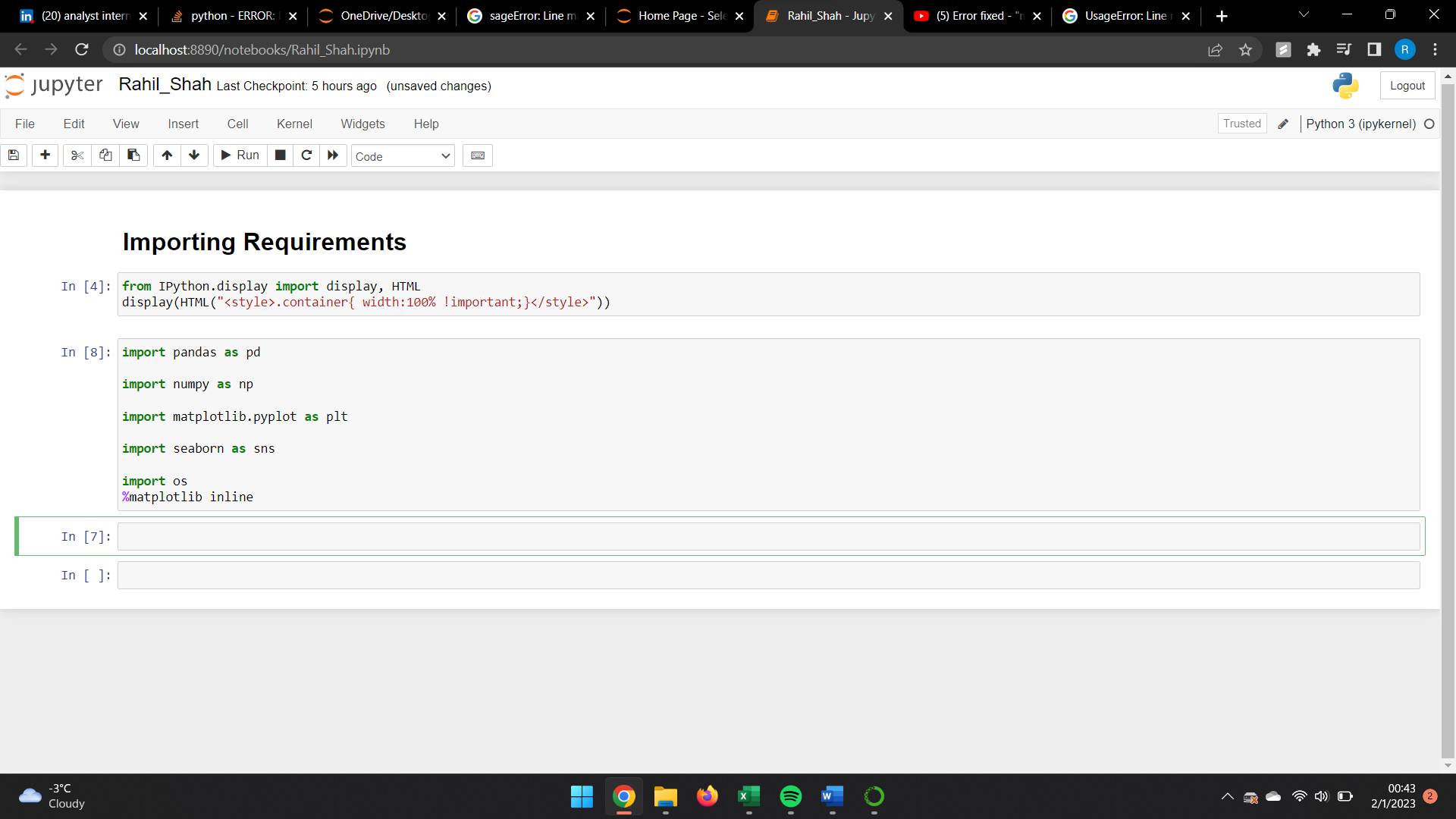
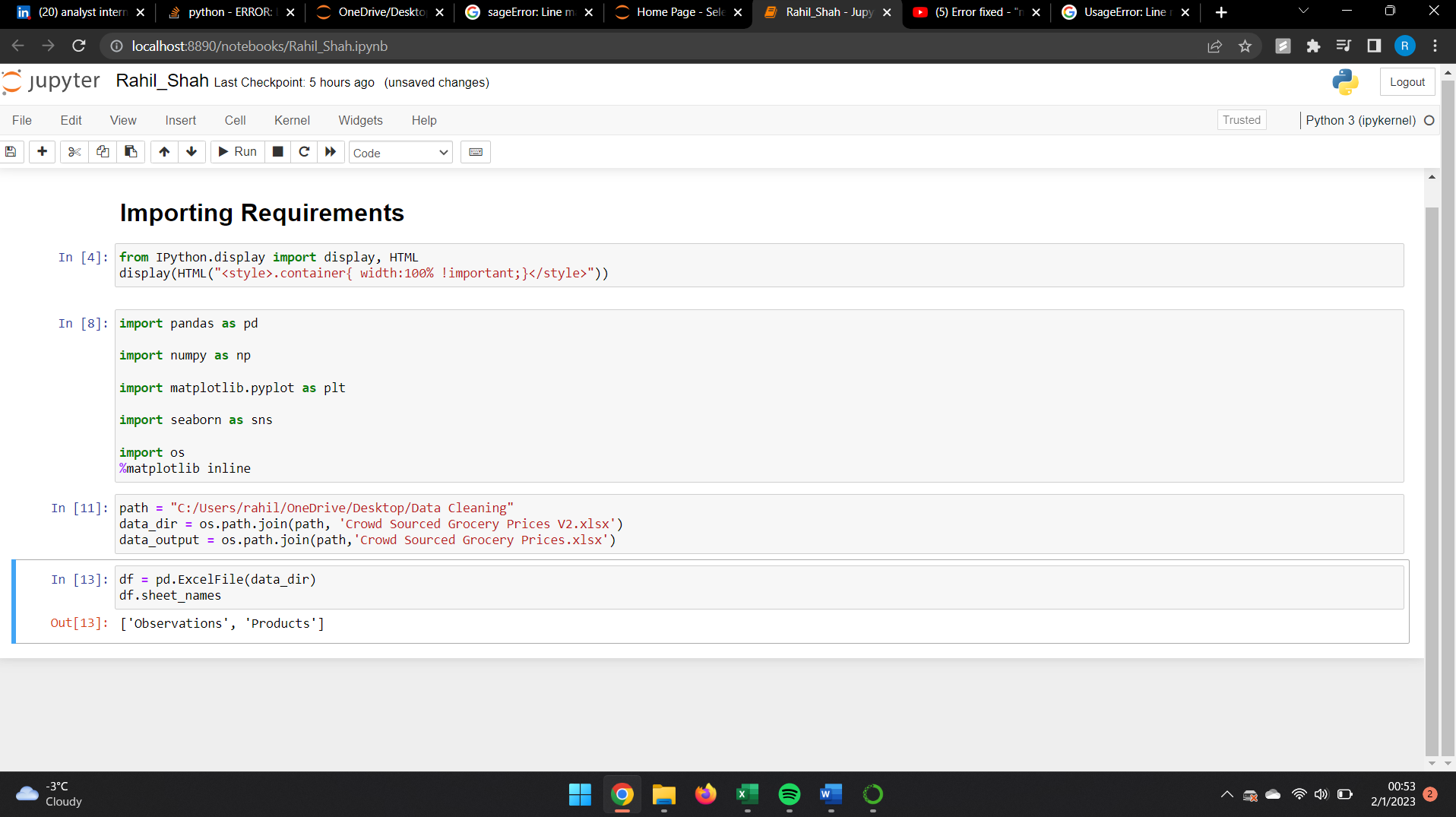
Data Manipulation

**Answer a)**

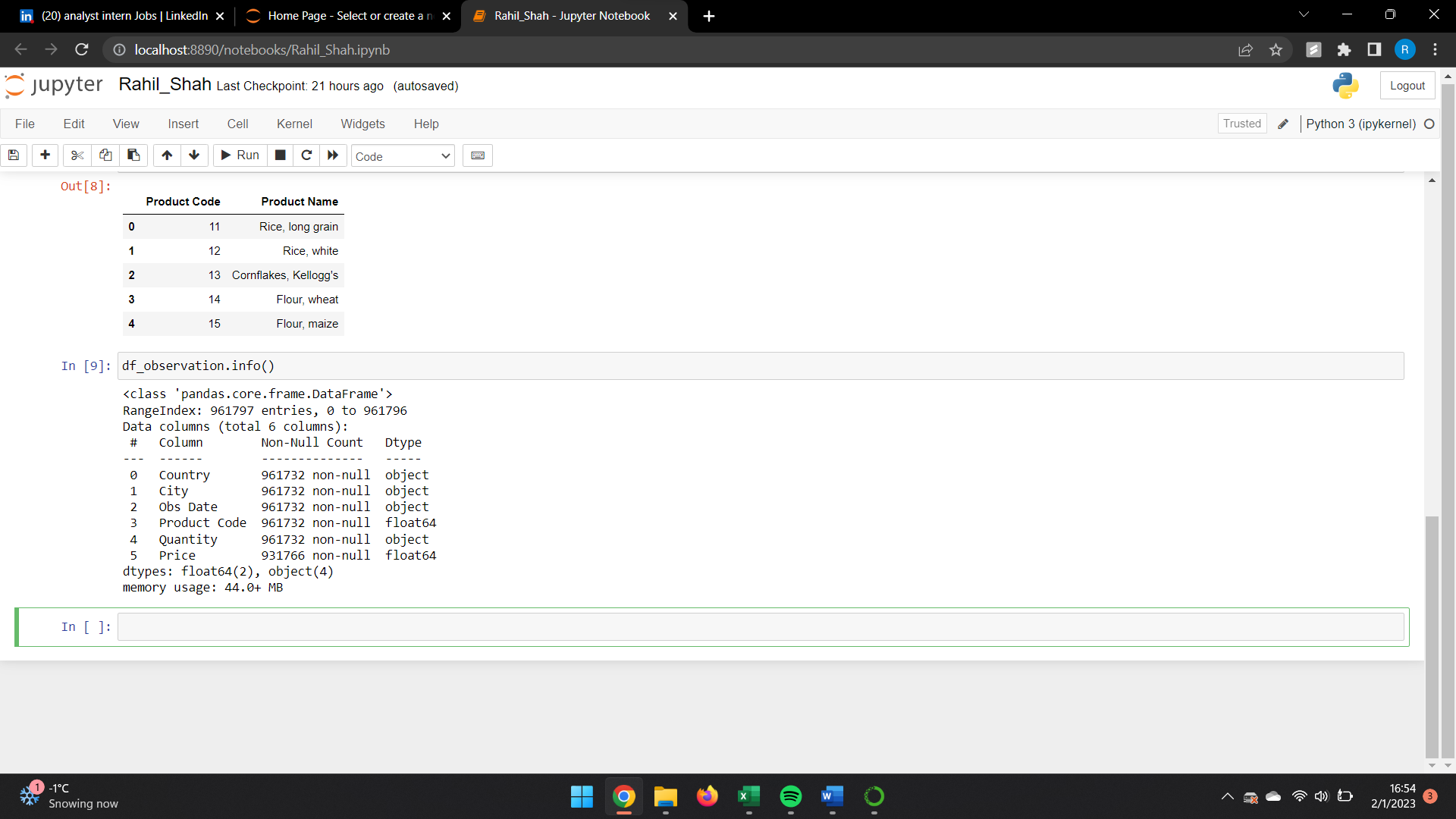
****

**Answer b)**

****

There are two sheets present in the Excel file. Their names are Obeservations and Products.

**Answer c)**

****

Number of non-null rows present in the dataset:

Country: 961732

City: 961732

Obs Date: 961732

Product Code: 961732

Quantity: 961732

Price: 931766

**Answer d)**

**Graphical user interface, text, application

Description automatically generated**

Number of empty rows present in the data before we removed them are 65 (I have not taken price column in consideration)

**Answer e)**

**Graphical user interface, text, application

Description automatically generated**

There 0 null rows present in Products sheet.

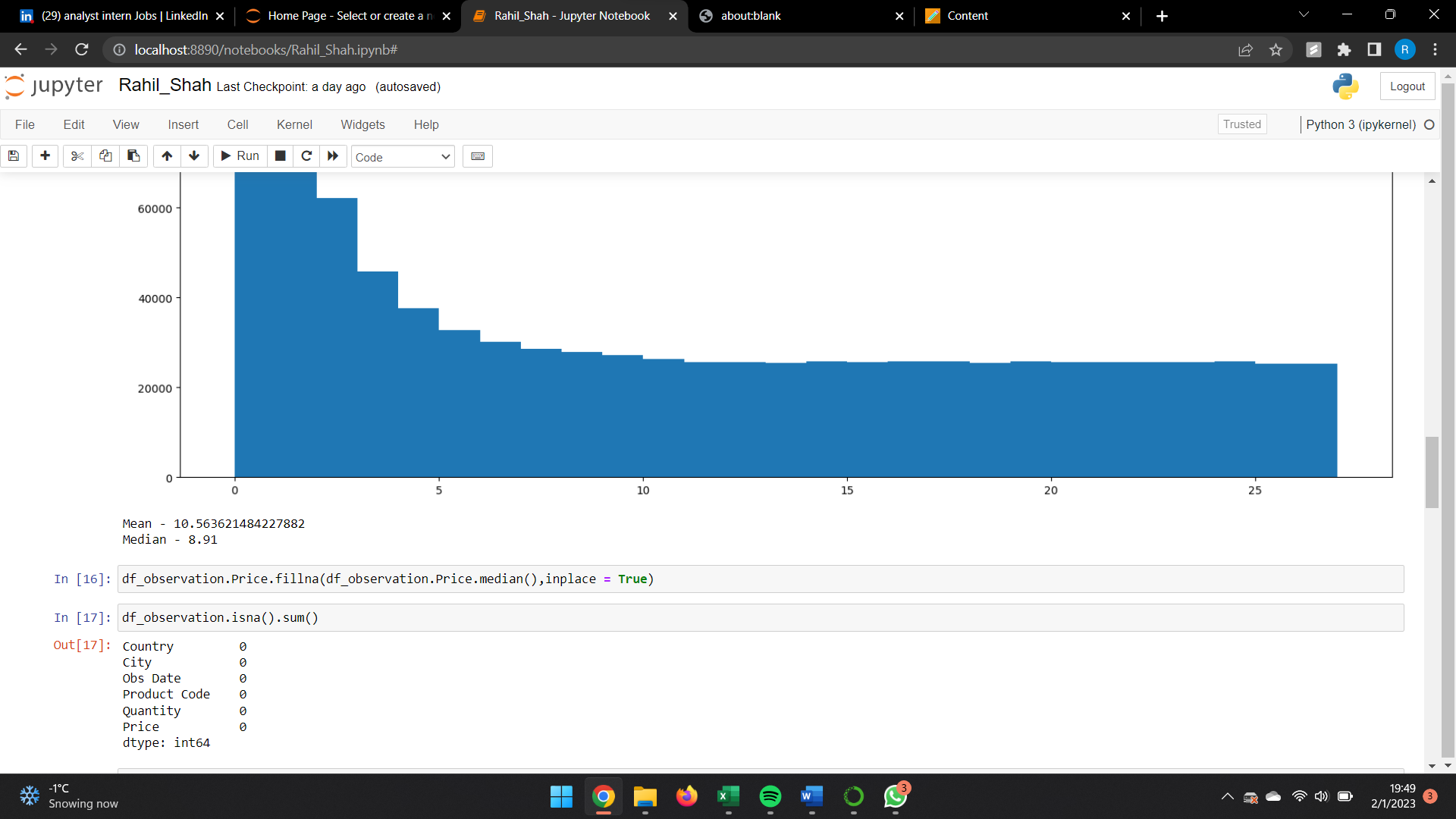
**Answer f)**

Graphical user interface

Description automatically generated with medium confidence

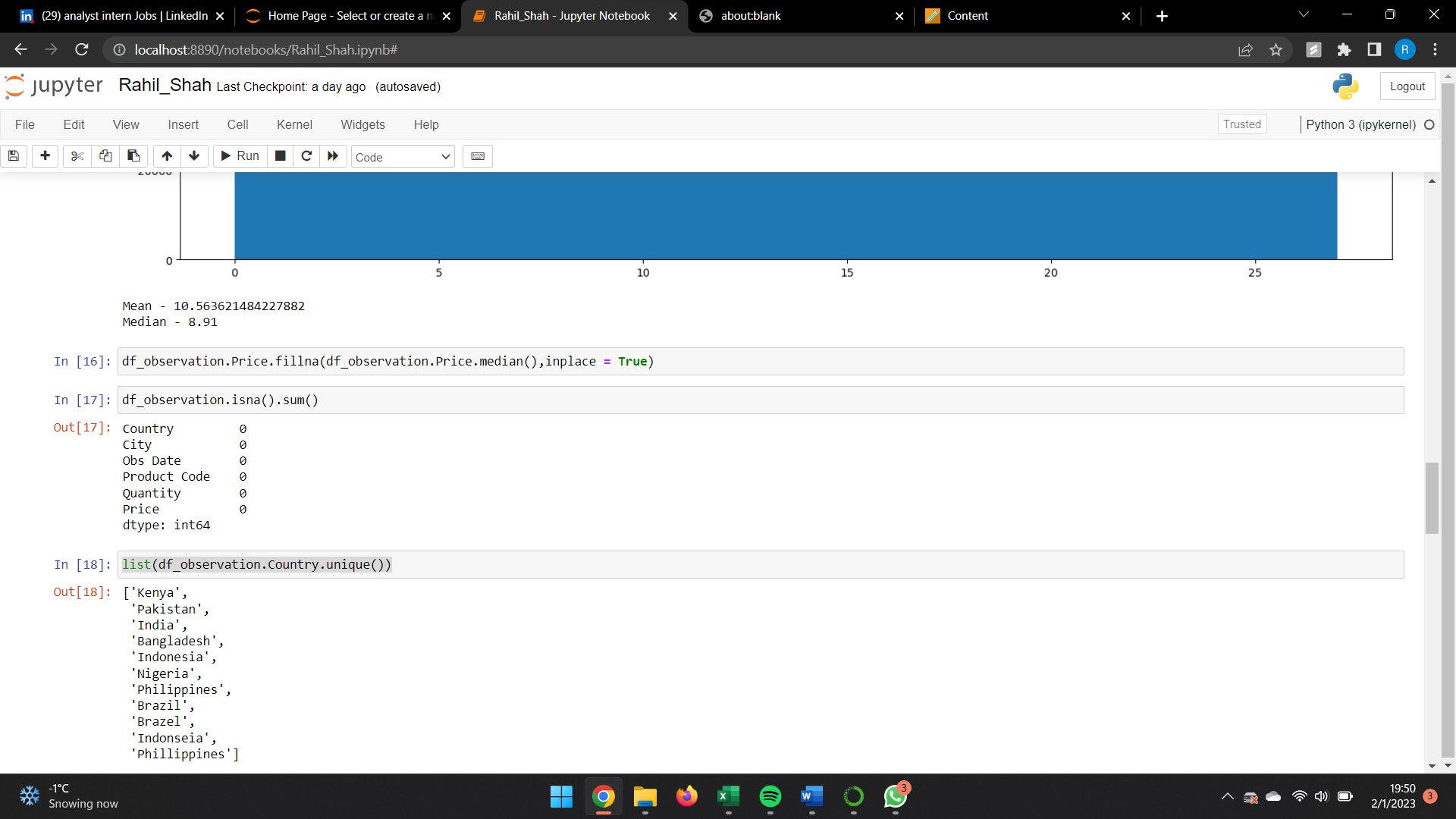
As the histogram is Right skewed hence we must use Median imputation method as median is the middle most value. The method suggest that we must not use mean imputation as our dataset may contain lot of outliers and they can have a significant impact on Mean.

**Answer g)**

****

There are 0 null values in the dataset after imputation of the code using Median Value.

**Answer h)**

****

Countries having duplicates or misspellings are:

1. Indonesia – Indonseia

2. Brazil – Brazel

3. Philippines – Phillippines

**Answer i)**

**A screenshot of a computer

Description automatically generated**

We have corrected the spelling errors by merging the spelling error in misspelled countries.

**Answers j)**

There are 14397 duplicate values in the dataset.

**A screenshot of a computer

Description automatically generated**

**Answer k)**

**A screenshot of a computer

Description automatically generated**

Using this code we can see the all the rows that have duplicate entries except the first entry as it is considered original entry and rest all dataset having same values are the duplicate entries.

**Answer l)**

**A screenshot of a computer

Description automatically generated**

**Answer m)**

A screenshot of a computer

Description automatically generated