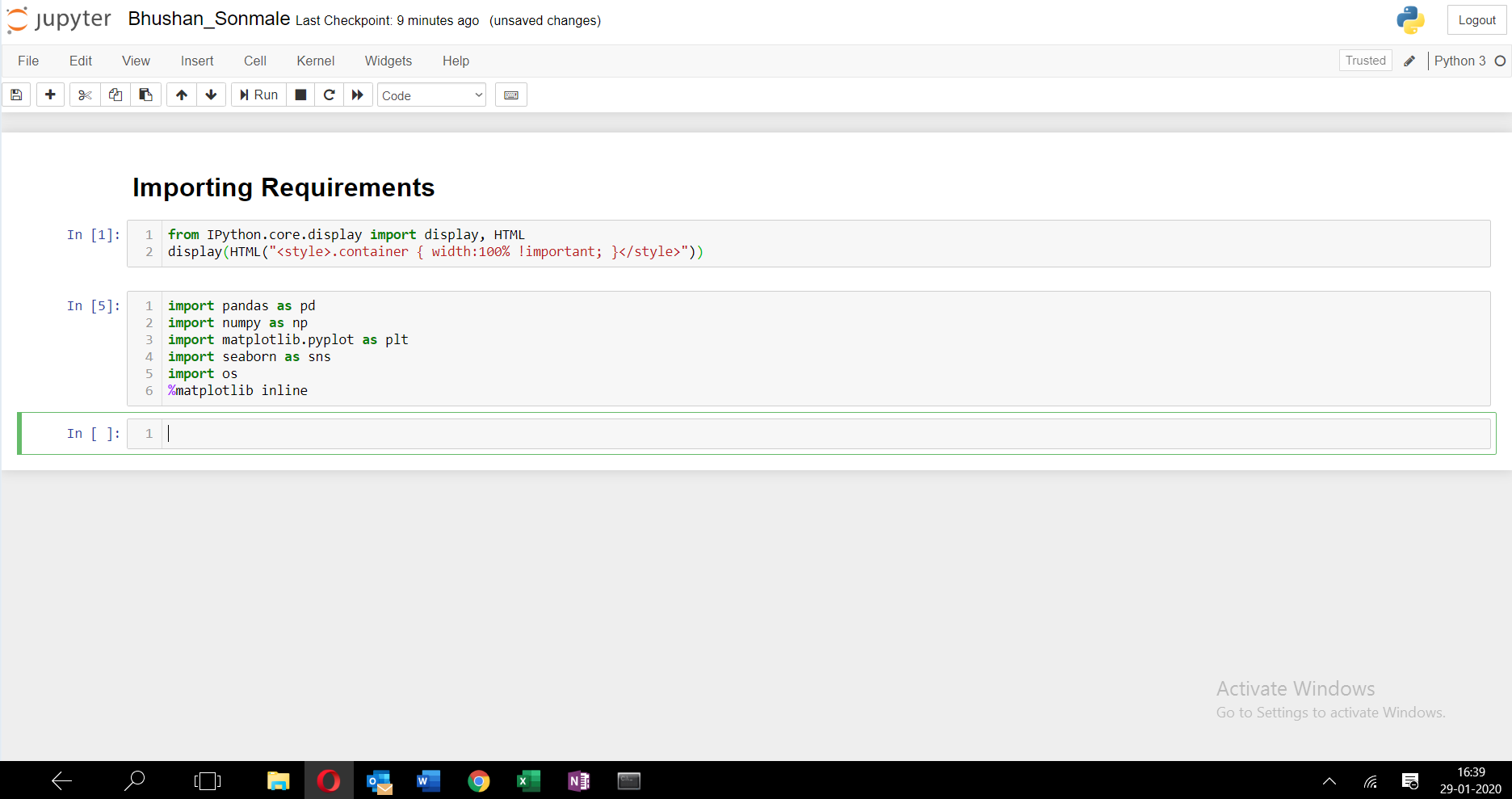
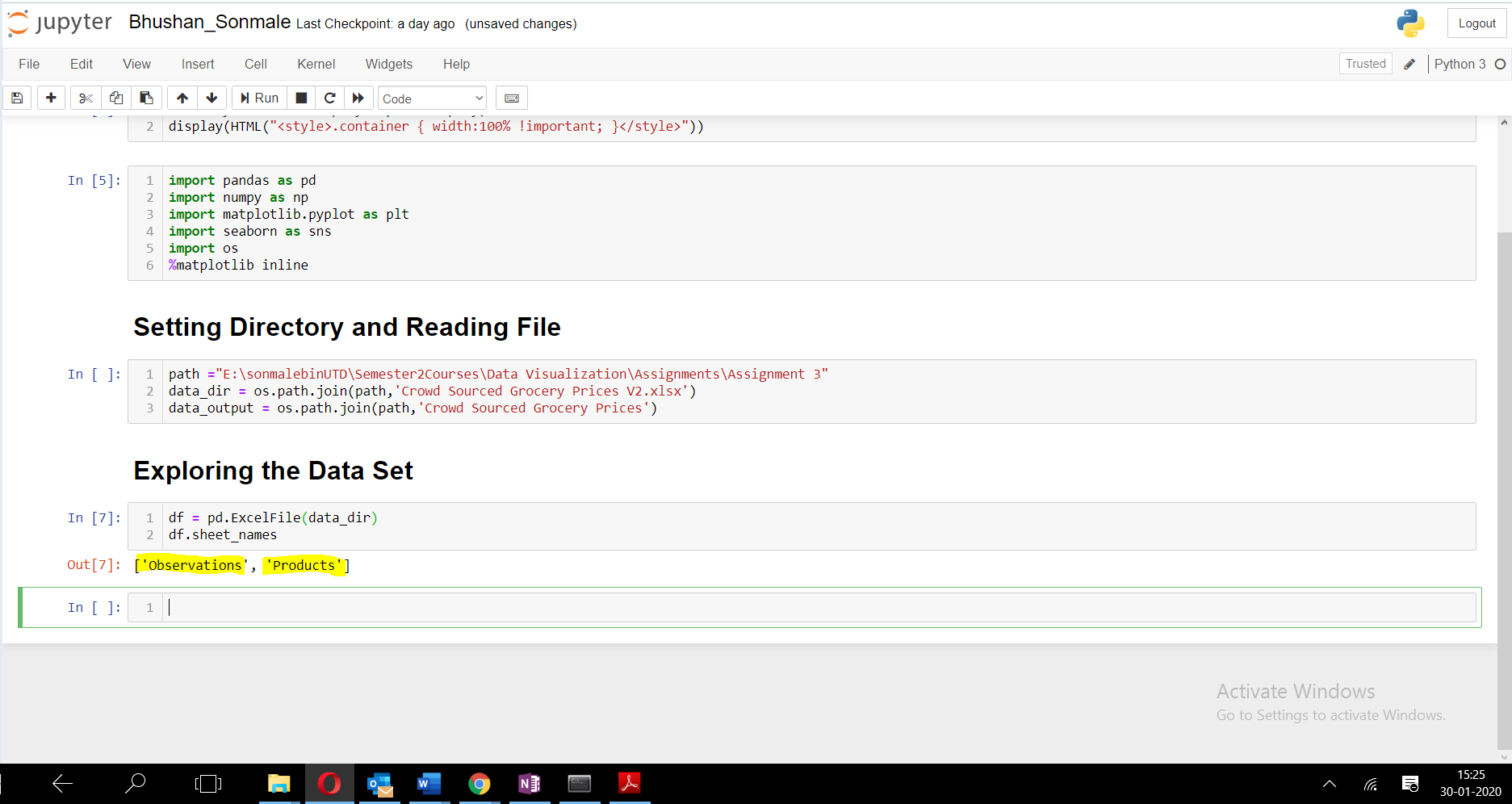
Solution a)





Solution b)

There are two worksheets named: Observations and Products



Solution c)

For all 5 columns except the **Price** column, the total number of non-null rows is 961732.

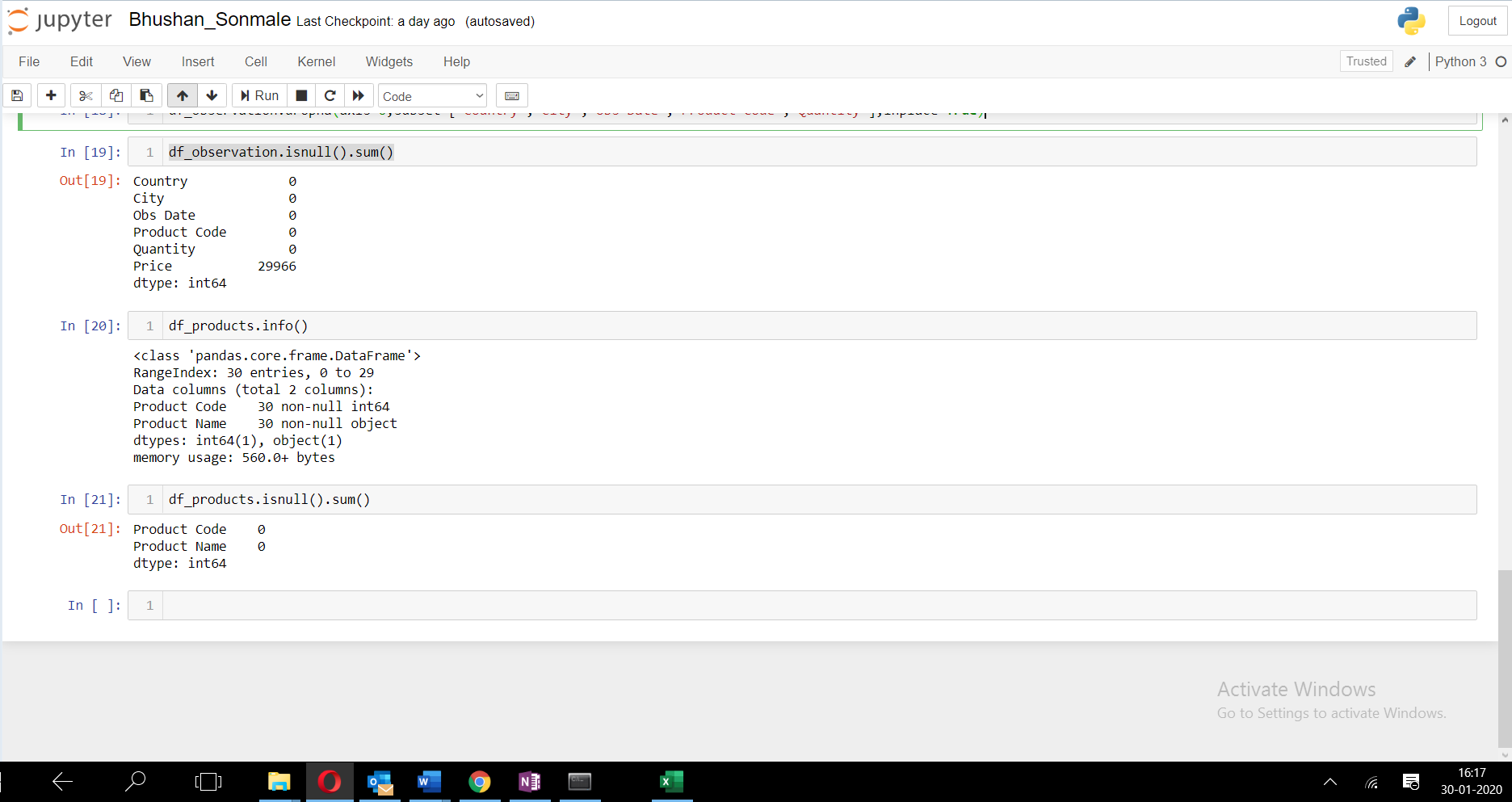
For the Price column, the number of non-null rows is 931766.

Solution d)

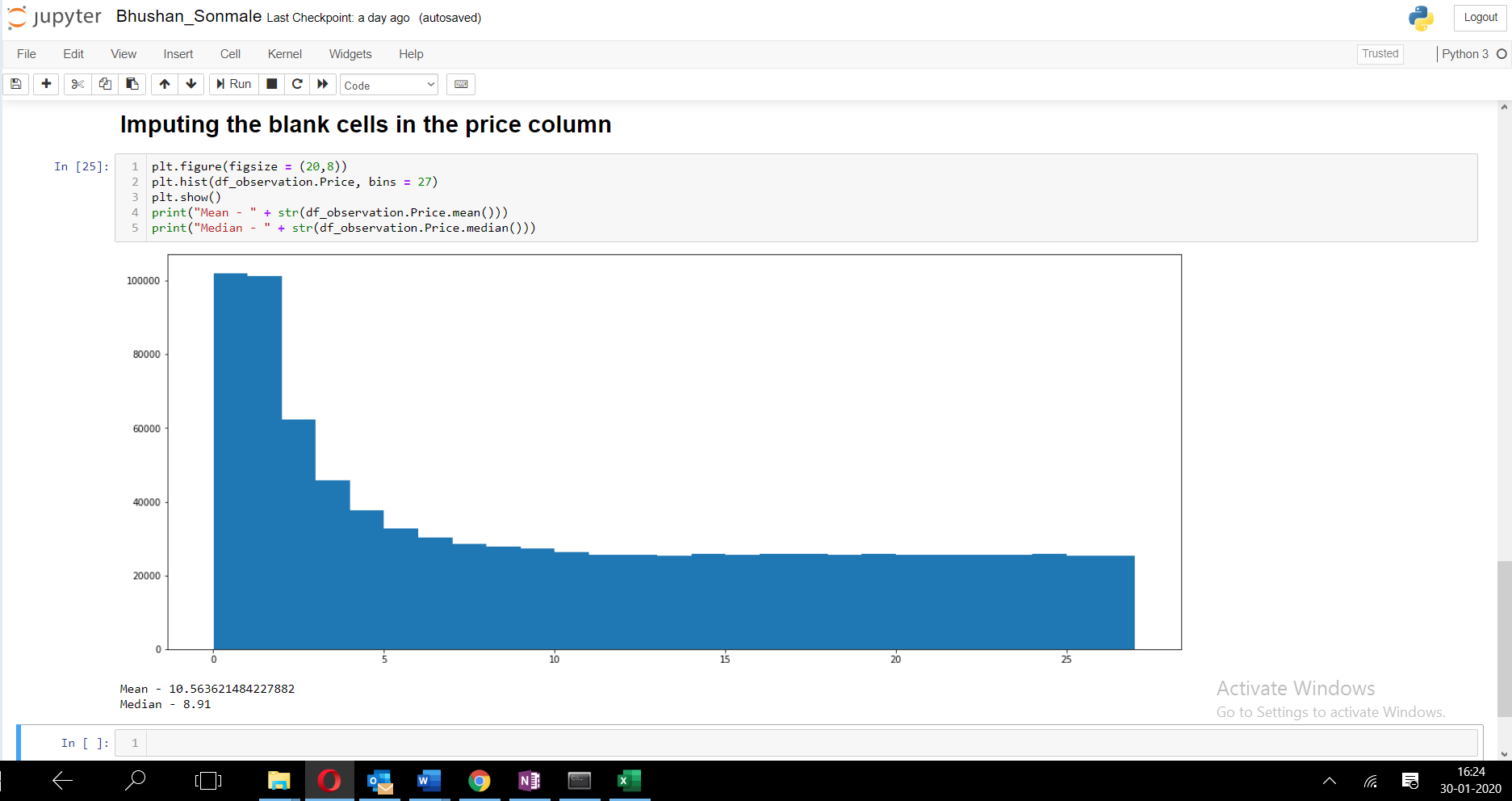
Before dropping, 65 rows were present that were completely empty, while 30031 rows had at least one column empty.

Solution e)

Number of empty rows present in the Products sheet is 0.



Solution f)



Since, the plotted histogram is a **skewed** one, we will adopt the **median** as the method for imputation.

Solution g)

There are 0 null values after adopting the median imputation method.



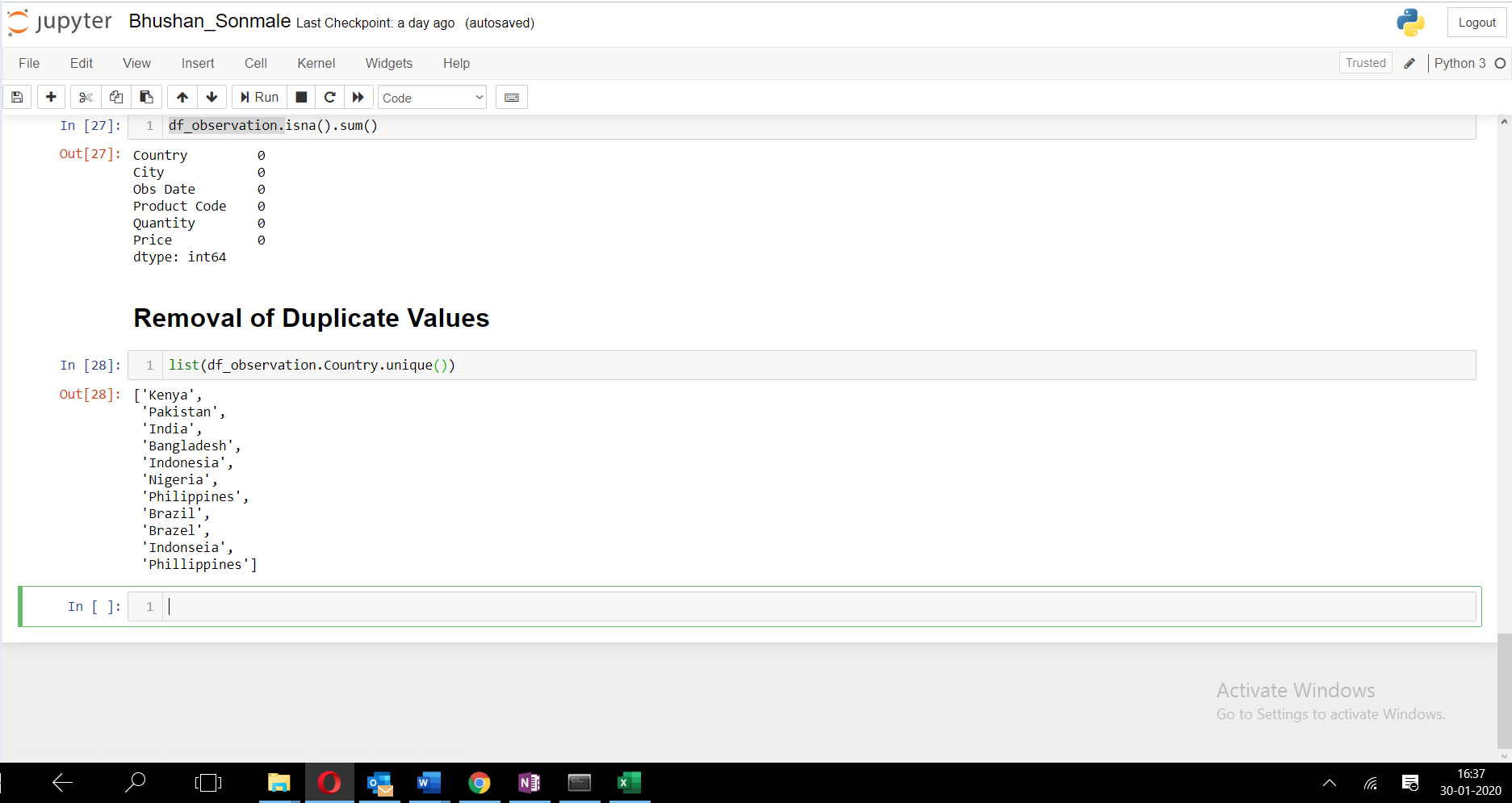
Solution h)

The list of countries is:

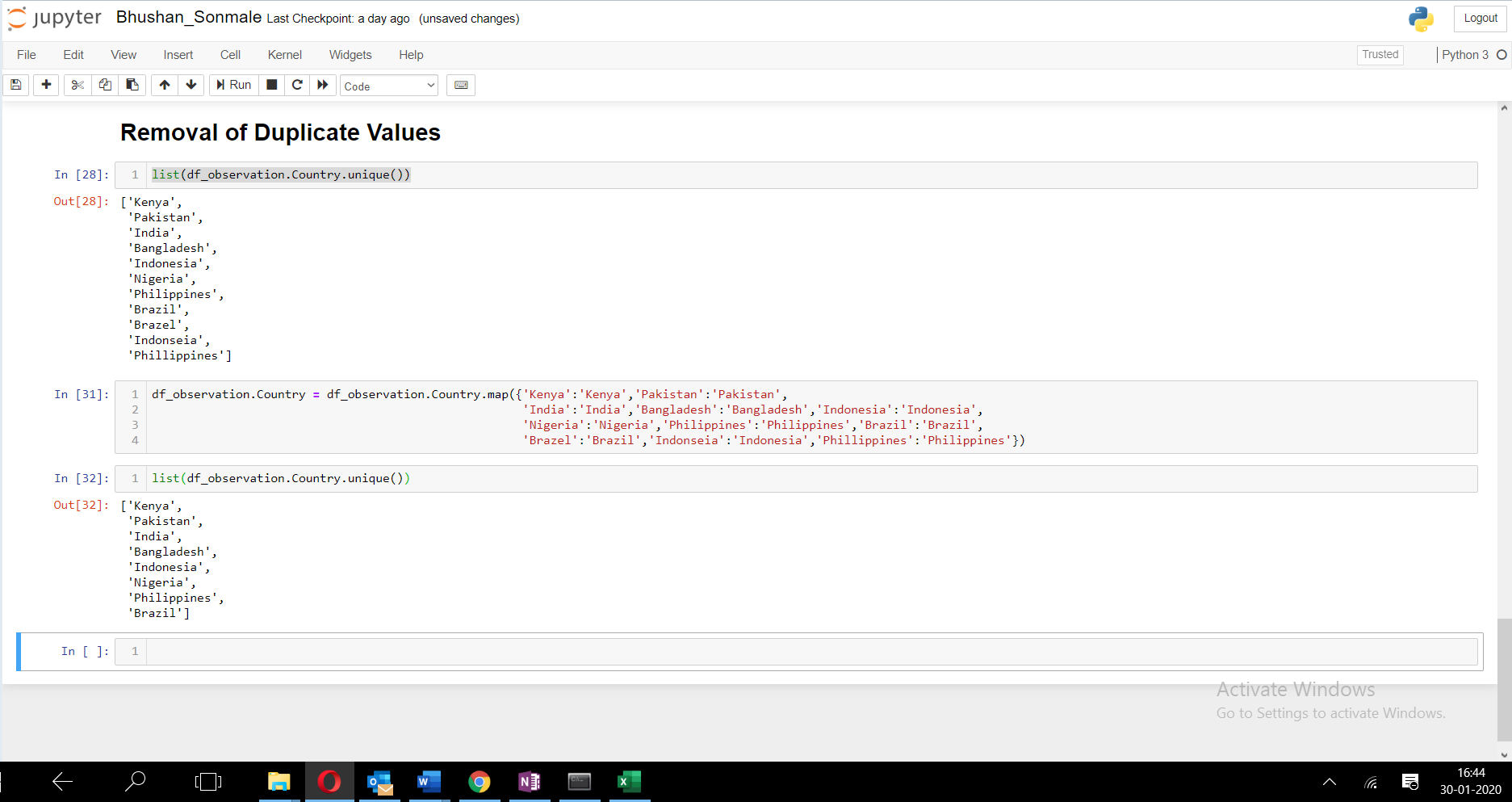
Kenya, Pakistan, India, Bangladesh, Indonesia, Nigeria, Philippines, Brazil, Brazel, Indonseia, Phillippines.

This list consists of 3 spelling errors namely:

**Brazel** instead of Brazil, **Indonseia** instead of Indonesia, **Phillippines** instead of Philippines



Solution i)

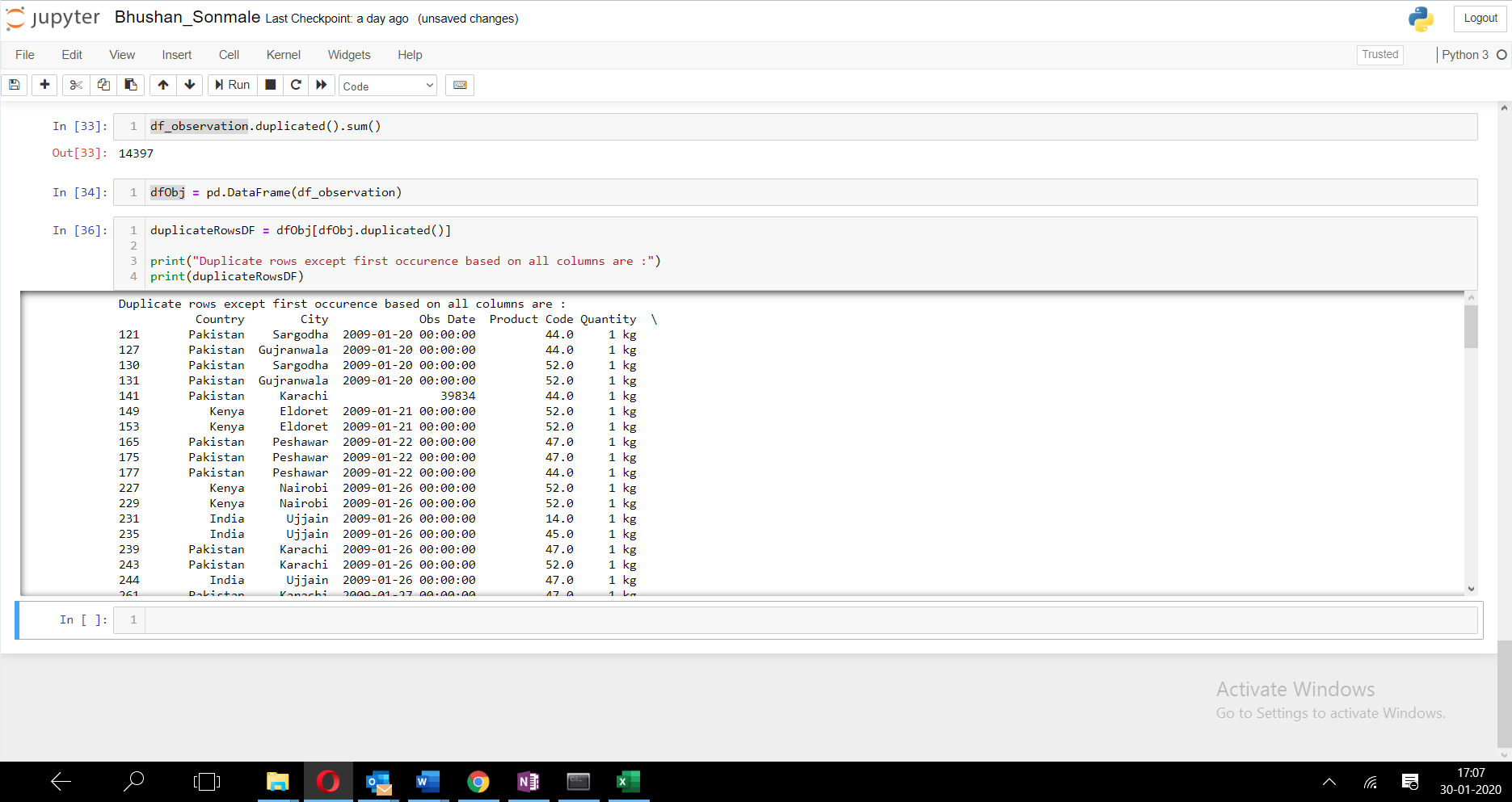


Initially, we had a list of 11 values that were returned by the unique function, while after removal/cleaning, the list came down to 8 unique values. This happened because, in the old list there were 3 values, with incorrect spelling namely: Brazel, Indonseia, Phillippines. On correcting these values, they were added to the list, and since on correction of the values, a duplication was raised which was removed. Eventually resulting in 8 distinct or unique values.

Solution j)

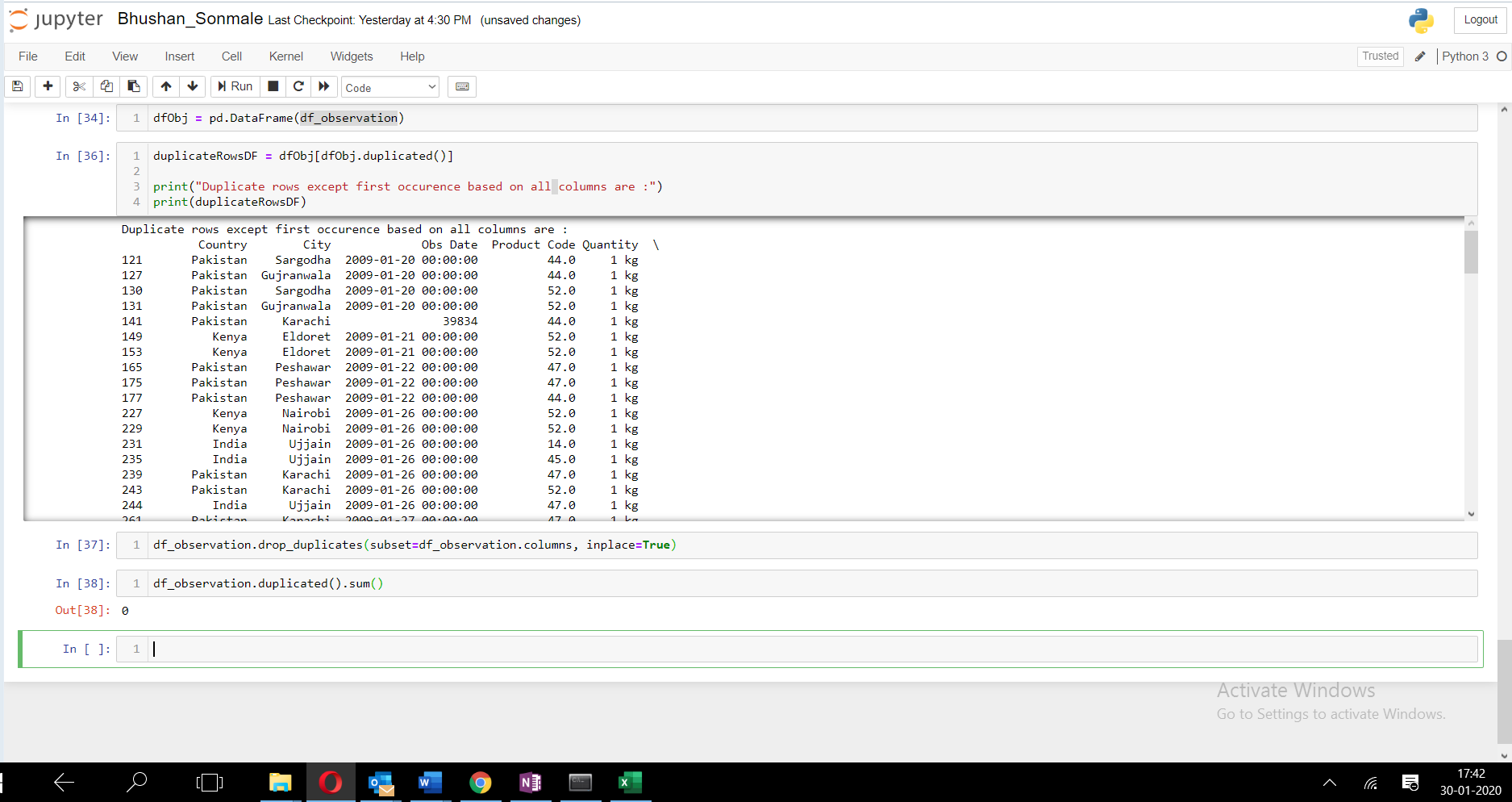
**14397** duplicate entries are present.

Solution k)



The query displays duplicate records except the first record from all columns in the data set. For example: In the Excel, there are 2 records with values Pakistan, Karachi, 39834,44.0,1kg, 8.91 at index positions 141 and 143 and when we check duplicate values only 143rd record number is displayed as the duplicate record excluding the first occurrence at record 141.

Solution l)



After dropping the duplicates, 0 duplicate records are present.

Solution m)

