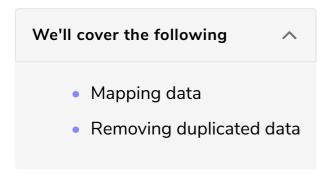
Mapping Data and Finding Duplicates

In this lesson, data mapping will be disussed along with how to find and remove duplicates in data.



Mapping data

As the name suggests, this technique is used to map the values of a Series or a DataFrame. The current values in a Series or a DataFrame are made equivalent to some other values. Then, the pandas map function is used to either replace the mapped values or join them together. The map() function can also be used to fill in the values of new columns.

The following example might make it clear.

Again a DataFrame is defined using the *dictionary* method like in the previous lesson. Different cities are mentioned along with their AQI (air quality index) values in the *dictionary*.

On **line 9**, a dictionary is defined with city names as the key attribute and a country name as their value.

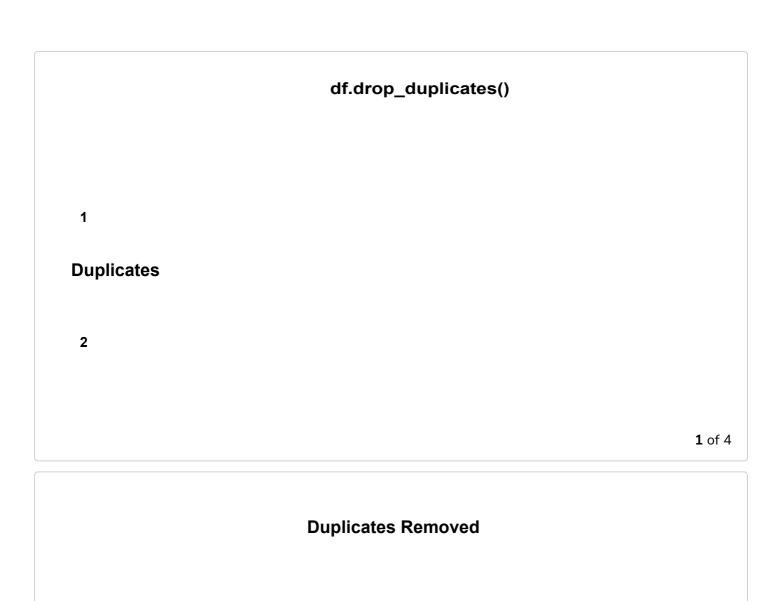
On **line 11**, a new column **Country** is created and assigned values using the **map** function. This function takes the **dict_map** dictionary created on **line 9** as a parameter and automatically assigns the corresponding country values to the cities.

Removing duplicated data

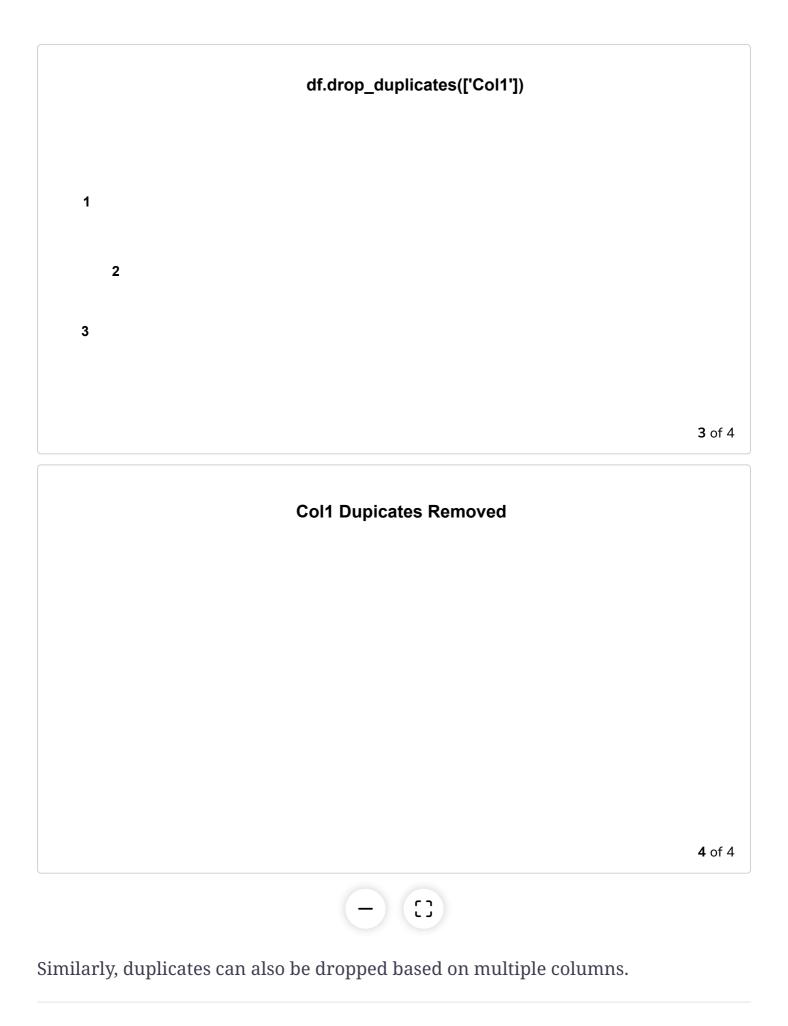
Now, you'll see how to get rid of duplicates in a dataset. The drop_duplcates() function of pandas is used for this purpose.

The drop_duplicates method is used on **line 10** without any parameter. It removes all rows that occur more than once. It only removes the rows where each cell exactly matches any other row of the DataFrame.

The drop_duplicates method is used on line 13. It has a column name passed as a parameter. This method checks for duplicate values in the specified column only, not in the whole row. All the rows with the same column values are removed until only one remains.



of 4



In the next lesson, some more important functions of pandas are explored.