

Springboard Project 5.3 (XML)

Using Mondial data

1. Find the 10 countries with the lowest infant mortality rates.

After parsing the XML file and getting the root, I opened a dictionary and iterated over the XML tree's elements for the names and infant-mortality rates of countries. The try-except command is used in the loop because some countries do not have infant-mortality rates associated with them. Last, I made a dataframe from the dictionary, added appropriate column headings, and displayed the top 10.

2. Find the 10 cities with the largest population.

From the document, I first created a dictionary with city names as keys and populations as values, again using try-except commands. From there, I built a dataframe with appropriate column titles from the dictionary and sorted to find the cities with the 10 highest populations.

3. Find the 10 ethnic groups with the largest overall populations (sum of best/latest estimates over all countries).

Pulling from the root, I created lists instead of dictionaries, as they proved easier to convert to a dataframe. After making one list each for country name, country population, ethnic-group name, and the percent of its country's population that the group makes up, I put the lists into a master list and converted it to a dataframe.

Next, I converted the population variables to numeric data types for mathematical operations and converted the percentages for ethnic groups to decimals for the same reason. With those steps completed, I multiplied the ethnic-group percentages by their respective countries' populations to get a raw count. Finally, I grouped the dataframe by ethnic group while summing their population counts and sorted for the top 10.

4. a) Find the name and country of the longest river.

First, I created lists for name, country, and length of rivers in the root. I then checked the length of the lists for congruency and put them in a master list. After converting that list to a dataframe, the rows and columns were switched, so I added code to transpose the dataframe. Last, I sorted by river length and displayed the longest one.

b) Find the name and country of largest lake.

For this step, I repeated the procedure from part a) using name, country, and size of lakes instead of the properties for rivers.

c) Find the name and country of airport at highest elevation.

For this step, I repeated the procedure from part a) using name, country, and elevation of airports instead of the properties for rivers.