

MySQL Project – World

1. Connect to the “world” database via Python.
2. Identify the tables and the columns within each table.
3. Load the *CountryLanguage* table as a Pandas dataframe and display it. Take a moment to interpret what is contained in this table.
4. Use an appropriate SQL call to generate a dataframe that includes columns for the country name and the country’s region, but only for countries that list “Spanish” as one of the spoken languages. How many are there?
5. **[Challenge]** Make a figure with boxplots where each boxplot indicates the life expectancy for the countries in a specific region. Arrange the boxplots (left to right) in ascending order by average life expectancy.

Hints:

- You’ll want to make a loop over all possible regions. One way to get a list of regions (so you know what to loop over) is with something like `regions = list(set(df['Region']))`. (The `set` command finds the unique values in a list/array/etc.)
- Set up three empty lists (e.g. `avgs = []`) before you start the loop, so you can capture the *average life expectancy*, *region*, and *array of life expectancy values* for each region. Inside the loop, append relevant values to the lists (e.g. `avgs.append(<something>)`).
- Once you’re done with the loop and have fully populated your three lists, you can sort them with something like `z = sorted(zip(list1,list2,list3))`. This will keep the entries in each of the lists synced up but will shuffle them in place so they’re in ascending order according to list1 (in the case of a tie between two values in list1, it will instead sort by list 2, then if needed by list3).