Data analysis exercise: Our World in Data Junior Data Scientist application

Overview

Thank you for helping us evaluate your skills by participating in our data analysis exercise!

The task, in a nutshell, is to take some age-specific death rates, use these to calculate both crude death rates and age-standardized death rates, and then communicate your working process as if you were reporting to your line manager.

You should be able to complete this exercise in two hours or less. If you cannot complete the whole exercise within this time, please submit what you have managed to complete and let us know how far you have gotten.

> 💡 **Please read the instructions below carefully before starting your work.**
>
> If you have any questions, feel free to email us at jobs@ourworldindata.org (the subject line must be “Junior Data Scientist” to ensure proper routing of your email).

The exercise

Below is a table showing the age-specific death rates from chronic obstructive pulmonary disease (COPD) in both the United States and Uganda in 2019. The values are given as deaths per 100,000 people.

Your task is to write a Python script that calculates both the crude death rate and the age-standardized death rate for COPD for all ages in both the United States and Uganda for 2019.

Your answers should represent the whole population of each country, and the value should be given as deaths per 100,000 people, rounded to one decimal place.

📝 **At the beginning of your Python script, write a short paragraph (<300 words) describing your process as if you were reporting to your line manager.**

This should cover the steps you have taken, any assumptions you have made, and a brief description of the reasons for the differences between the different calculated death rates.

Comments in the main body of your code that aid understanding of what the code is doing are *not* included in the 300 words.

Resources to use

The three items in this section are the only ones you should need to complete the task, but you may use others to help aid your understanding of crude and age-standardized death rates.

1. UN World Population Prospects (2022) — Population Estimates 1950-2021
2. WHO Standard Population — Table 1 in 'Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJ, Lozano R, Inoue M (2001). Age standardization of rates: a new WHO standard.'
3. Table of age-specific death rates of COPD:

Age group (years)	Death rate, United States, 2019	Death rate, Uganda, 2019
0-4	0.04	0.40
5-9	0.02	0.17
10-14	0.02	0.07
15-19	0.02	0.23
20-24	0.06	0.38
25-29	0.11	0.40
30-34	0.29	0.75
35-39	0.56	1.11
40-44	1.42	2.04
45-49	4.00	5.51
50-54	14.13	13.26
55-59	37.22	33.25
60-64	66.48	69.62
65-69	108.66	120.78
70-74	213.10	229.88
75-79	333.06	341.06
80-84	491.10	529.31
85+	894.45	710.40

Submitting your solution

There are two elements to submitting your solution:

1. Please attach the Python script or notebook you wrote to your application email.
2. Enter your calculated values of crude and age-standardized death rates for both the United States and Uganda in **this Google Form**.
    a. When filling out the form, please use the same email address you used to send your application email.

Frequently asked questions

How long should I spend on this task?

No more than two hours, out of respect for your time. If you are reaching the limit, please submit whatever you've managed to do.

Can I use Copilot, LLMs, or other AI tools?

Yes. As part of your work with us, we would provide you with GitHub Copilot and access to OpenAI/ChatGPT, so feel free to use these and other tools as you normally do.

When will I hear about my application?

We'll contact you 1 to 2 weeks after the application deadline.