PALINDROME DATA

Data Science Residency Interview Assignment

Congratulations on getting to the next round! This assessment will evaluate your ability to do a bit of coding and data manipulation in Python.

All work and output display must be completed in python. We encourage partial submissions i.e., even if you do not complete all the tasks, we would still like for you to submit what you have managed by the due date.

Deadline: 9am, This Friday 9 April, 2021. We will offer interviews as valid exercises are received.

Instructions & Tasks:

1) Download the dataset here.
2) Calculate and output the following:
   a) What is the total number of people living with HIV (NoPLHIV) in the listed districts according to the *Survey* estimate?
   b) What is the average NoPLHIV of the two estimates used for "*Xhariep*"?
   c) Add a column and populate it with the *number of people **not** living with HIV* for each row.
   d) What is the total NoPLHIV in all the cities (districts with the word "city" or "metro" in the name)?
3) Write the original data (without the caption - originally row 1) with the extra columns as *comma-separated values* (CSV) to a new .csv file.

Deliverable Format: Please submit your single code file in a python format (*.py, .ipynb,* etc.) by committing your notebook of results to your/ a public git repository and emailing us the link.

Special Considerations:

- Please track and include in the email us the amount of time (hr:mins) you spent on the assignment. Also great if this is just documented in the code.
- We do not expect everyone to complete all of the tasks. Please submit even if you do not complete all the tasks; you may proceed to the next round where you will be able to explain your challenges (this is actually more important than getting the right answer!).
- Remember that this is your opportunity to showcase your creativity – we hope that you do not feel limited by our simple instructions. Feel free to add anything of interest, but do note we're looking for prompt responses.