

# Test for Data Scientist - Data Lab Nairobi

*Time: 2.5 hours*

*Directions: Complete all questions within 2.5 hours and return your answers to [slezak@un.org](mailto:slezak@un.org) with a copy to [teran1@un.org](mailto:teran1@un.org) and [hdx@un.org](mailto:hdx@un.org).*

## *Question 1. Using Python, scrape data from the UN Iraq website*

- A. You are requested to extract data from the [UN Iraq website](#). Please create a table in a google spreadsheet that contains the casualty figures from November 2012 to February 2015. Your extraction should be done programmatically (coded) to respond to updates of the website (ie, it should be robust enough to react to changes of the past estimates and to incorporate new estimates (March 2015).
- B. Your resulting dataset should also contain the following metadata fields: source, methodology, date of dataset, location, caveats and comments. In addition, your dataset should have the following tabular structure:

Date	Killed	Injured
YYYY/MM		

- C. Submit your entire code or publish it on Github

## *Question 2. Correct Source Errors*

This [repository](#) contains a broken script that needs to be fixed. The script contains 3 errors:

1. The function fails to run.
2. The columns of the CSV output are improperly organized.
3. Some records fail to be written in the output CSV file.

Please identify the source of the three errors above and fix them, making sure that the script runs successfully and that the output CSV file is correct. Submit the fixed code or publish it on Github.