

**Req21560 – Extended Term Consultant – Data Developer – Practical Test**

April 7, 2023

**Instructions**

Congratulations on being shortlisted for the Data Developer/Consultant position in the Integrity Vice-Presidency (INT) of the World Bank. Please complete the practical test below, which will be reviewed by the INT Selection Advisory Committee (SAC) as part of your application. You may be asked to explain, comment on, or expand on your responses during the interview with the SAC.

Once you complete Sections I and II below, please commit your reply to your GitHub account and notify the SAC at: [mdonolo@worldbank.org](mailto:mdonolo@worldbank.org); [bkesari@worldbank.org](mailto:bkesari@worldbank.org); [tloftus@worldbank.org](mailto:tloftus@worldbank.org) by **Monday, April 10 at 13:00pm** as per your time zone (please indicate your time zone in your notification email).

For Section I, please refer to two files: **titanic.csv** and **ubo.json**. The SAC will run your code in a local IDE to assess code quality. For Section II, the SAC will review your written responses, which should be provided in a .txt file. Please do not share your code over a text document or any other format.

**Assessment**

For Section I, the SAC will review the quality of your code, i.e., its attributes and characteristics that show functionality, clarity, efficiency, testability (i.e., free of bugs or defects), maintainability and replicability, and documentation (i.e., the ease with which your code can be followed/understood). For Section II, the SAC will assess the completeness and quality of the explanations provided. Good luck!

**Practical Test**

Section I

1. **Creating an API Endpoint**

* Use the attached **titanic.csv** file and create an API endpoint ‘http:/localhost:8000/titanic’ using a python framework of your choice, SQLalchemy, and SQLite
* Preferably: Use FastAPI framework
* Note: In the README.md file clearly document which python API framework you are using and run command to test in our local. We expect API to run in port 8000 at ‘‘http:/localhost:8000/titanic’

1. **Parsing**

* Parse the attached **ubo. json** file to generate a tabular data frame. Make sure to explore all levels in the nesting.

Section II

1. Would you store data in databricks as delta or parquet format? Based on your choice, please explain why?
2. Please explain the process of how you would optimize PySpark or SQL code to effectively use databricks spark cluster?