**Dataset Description.**

Two data tables have been extracted from a database into two separate Excel files:

1. "final result"

2. "laboratory investigation

These two tables are linked using a common keys called "**person\_investion \_id**" in "final result" and "**laboratory\_investigation\_id**" in "laboratory investigation".

"**final result**" is a file which contains the final result of a test done on a patient. One of the tests conducted on a patient is the HIV test written as “**HIV**” under column “**test**”.

"**laboratory investigation**" is a file which contains all the tests that were done on a patient, leading to the generation of the final results in the "**final result**" file.

The "**time**" column in the "**laboratory investigation**" file shows the date and time at which each test was performed.

**Question a:**

Write a script (using any language of your choice) that extracts all the people who had an **HIV** test and tested **positive** (the value in result column in “**final result**” is positive (ignore case of positive)) in the "**final result**" file. Retrieve the details of the tests that were done for these patients from the "**laboratory investigation**" file.

The output file must have the following columns:

Table : Output template

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Person\_id | HIV Final result | HIV test result 1 | HIV test result 2 | HIV test result 3 |

- "**person\_id**" is a column that comes from the "**final result**" table.

- "**HIV Final result**" is a column that comes from the "**final result**" table, where the column name is "result".

- "**HIV test result 1**" is a column that is derived from the "**laboratory investigation**" table, containing the result of the first HIV test performed on the patient.

- "**HIV test result 2**" is a column that is derived from the "**laboratory investigation**" table, containing the result of the second HIV test performed on the patient.

- "**HIV test result 3**" is a column that is derived from the "**laboratory investigation**" table, containing the result of the third HIV test performed on the patient.

Note:

- Discard all tests that are after the third test.

- Ensure that all your column names are named as shown in Table 1.

**Question b.**

Write a script that split your output file into 3 files and ensure that your splits are exactly the same as the original output file.