

**Assessment for the position of:Junior Data Engineer**

**Time for Completion: 3 Days**

**GUIDELINES:**

**Introduction**

We would like you to please complete this at home technical assessment first. Once this test is completed, we will review the results and you will be scheduled for a second in-person technical interview and to meet other members of our team.

**Deliverables**

Upload to a Git repository and send us a repository link of your assessment which can be accessed by our team.

PLEASE DO NOT SEND US A ZIP FILE.

**Delivery Date**

Completion of the assessment is 3 days.

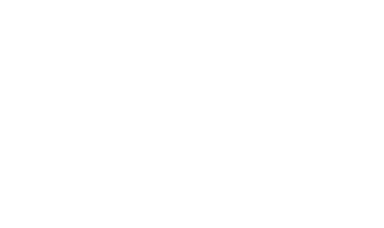
**Submission Process**

Please email to hr@horizontech.biz with the Git repository link and any additional information you feel needs to be shared.

In case of any technical queries related to this assignment, please email at the provided HR email address.

**Task 1:**

The flow of patient registration is as follows:



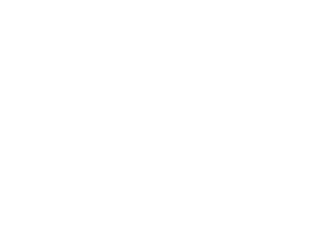
A patient consult the clinic

for fertility issue.

Primary patient

information is registered in

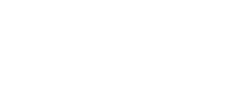
the system or database.



Partner information is

registered into the

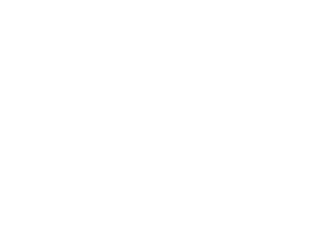
system or database.



Invites partner (if

any) to fill the

information



Booking of

Appointment,

consultation and

treatment

**Index**:

Patients are coupled in form of partner. For example: Patient ID 1 has a partner patient ID 4. Information of both are in Patient\_Info.

It is not necessary that every patient must have a partner.

File “Patient\_Info.xlsx” contains the following fields for both types of patients (primary and partner): a.a. Patient\_ID

1. Patient\_Name
2. Patient\_Gender
3. Patient\_Phone
4. Province
5. Created\_at: When was the record created

File “Patient\_Partner\_Mapping.xlsx” contains the pairing of couples:

1. Patient\_Partner\_ID
2. Patient\_ID
3. Partner\_ID

**SQL related:**

1. Create a MySQL database instance and insert the data from file “Patient\_Info.xlsx” and “Patient\_Partner\_Mapping.xlsx” into the database.

You can create tables as per the requirement or your understanding.

1. Write SQL query to get the following:

* Get patient information with their partner information (Patient Name, Patient Gender, Patient\_Phone, Province, Partner Name, and Partner Gender).
* Get patient information of those patients who do not have a partner.
* How many patients are there from each unique province.
* How many patients are registered on each month?

**Submission file types:**

For part a, Export your SQL or SQL dump.For part b, SQL query file.

**Task 2:**

File “sample physician information.xlsx” contains data scrapped from an online physician repository. All the information is in one single column in the form of a string.

* The task is to get the registration/CPSO number, Physician Name, and Phone number in separate columns and export it to an excel file.
* Get **Area of Specialization** wise count of physicians.

Use Python or R for the above-mentioned tasks.

Note: If any information for a physician is not available, mark it as NULL.

**Example of a record:**

**Input:**

|  |
| --- |
| **Full Information** |
| Ghumman, Simranjit Singh (CPSO#: 111515 Location of Practice: Suite 101464 Boulevard Saint Jean BaptisteMercier QC J6R 2B1 Phone: (450) 699-0539 Area(s) of Specialization:Family Medicine |

**Output:**

|  |  |  |  |
| --- | --- | --- | --- |
| **CPSO Number** | **Name** | **Phone** | **Area of**  **Specialization** |
| **111515** | Ghumman,  Simranjit Singh | (450) 699-  0539 | Family  Medicine |

**Submission file types:**

Py or ipynb file.

**Self-Assessment:**

1. How much do you rate yourself in Python (out of 5):
2. Which libraries have you used in Python and rate yourself in each (out of 5):
3. How much do you rate yourself in SQL queries (out of 5):
4. How much do you rate yourself in understanding the requirements and building ERD (out of 5):
5. How much do you rate yourself in English Communication (out of 5):

Reading: Writing:

Listening:

Speaking:

1. What are your hobbies?
2. What are your expectations from us (in general)?

**Submission file type:**

PDF