



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

**Time for Completion: 5 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 5 days.

Kindly send us with the completed assessment by Date (24/07/2022)

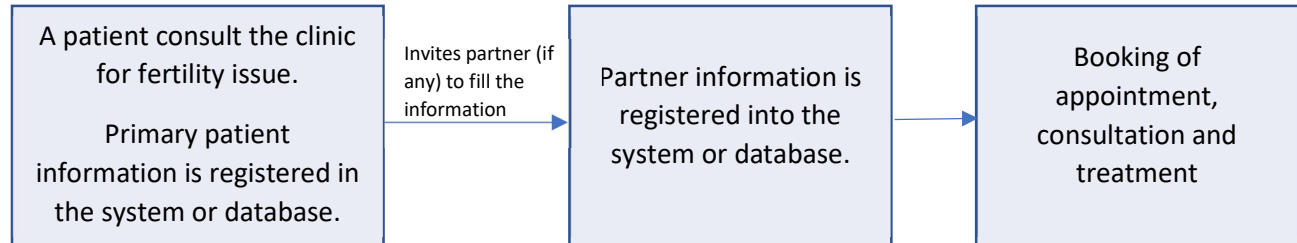
### **Submission Process**

Please email to [hr@horizontech.biz](mailto: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:



## Index:

Patients can be of 2 types:

Primary Patient: Patient who consulted the clinic for fertility treatment.

Patient Partner: Partner of the patient.

Tip: Primary Patient is the patient who is registered first into the system.

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):

- Patient\_ID
- Patient\_Name
- Patient\_Gender
- Patient\_Phone
- Province
- Created\_at: When was the record created

File "Partner\_Partner\_Mapping.xlsx" contains the pairing of couples:

- Patient\_Partner\_ID
  - Patient\_ID
  - Partner\_ID
- Create a MySQL database instance and insert the data from file "Patient\_Info.xlsx" and "Partner\_Partner\_Mapping.xlsx" into the database.  
You can create tables as per the requirement or your understanding.
  - Write SQL query to only get the information (Patient\_ID, Patient\_Name, Patient\_Gender, Patient\_Phone, Province and Created\_at) of primary patients.
  - Using Python, connect and load data from MySQL database you created and execute a "select \* from table" query.

**Submission file types:**

For part a, Export your SQL or SQL dump.

For part b, SQL query file.

For part c, py or ipynb file.

**Task 2:**

**The following tasks need to be done in Python using Pandas.**

- a. Load the data from provided files: 'Patient Info.xlsx' and 'Partner\_Partner\_Mapping.xlsx'.
- b. Join both data frames based on Patient\_ID.
- c. Get patient information of those patients who do not have a partner.
- d. Get patient information of only primary patients.
- e. Create a dataframe/dictionary (whichever you prefer) to show how many patients are from each unique province.
- f. Create a table with the number of referrals each month.

**Submission file types:**

Py or ipynb file.

**Self-Assessment:**

1. How much do you rate yourself in Python (out of 5):

4.5

2. Which libraries have you used in Python and rate yourself in each (out of 5):

Pandas – 4.5

Sqlalchemy – 4

I have many more libraries.

3. How much do you rate yourself in SQL queries (out of 5):

5

4. How much do you rate yourself in understanding the requirements and building ERD (out of 5):

5

5. How much do you rate yourself in English Communication (out of 5):

Reading: 4.5

Writing: 4.5

Listening: 4.5

Speaking: 4.5

6. What are your hobbies?

Watching series, playing badminton, socialize.

7. What are your expectations (in general)?

**Submission file type:**

PDF