X0PA – TECHNICAL ASSESSMENT (DATA SCIENTIST/ ML ENGINEER)

Thank you for applying to the Data Scientist/ ML Engineer role at X0PA AI Pte. Ltd. As part of our interview process, you will be required to complete the two tasks below as part of the technical assessment. Please send your submissions or any queries to: leegang@x0pa.com.

Task 1

Your task is to write the following queries in SQL based on the **Database Schema** below. The database schema is written in PostgreSQL. However, you are free to use any SQL dialect but please specify the SQL dialect you are using.

a) Write a query to print the total number of open positions (open positions = vacancies - # of hired candidates) for each company as of the current date.
b) Write a query to print the industry, company name, and total number of local applications (Singapore candidates that applied to a job in the company) for companies with the highest number of different jobs in each industry. For an example, ABC company has 80 different jobs (the highest in the agriculture industry) and has 1,000 Singapore candidates applying to its jobs. The query would print {'industry: 'Agriculture', 'company':'ABC', 'local_applications':1000} for this record.

Deliverables

- **Code**

  Please send us the SQL script for the two queries above. Please specify the SQL dialect you are using.

Database Schema

```
CREATE TABLE IF NOT EXISTS candidates (
        profile_id CHAR(11) PRIMARY KEY,
        first_name VARCHAR(32),
        last_name VARCHAR(32),
        country VARCHAR(16),
);

CREATE TABLE IF NOT EXISTS company (
        company_id CHAR(11) PRIMARY KEY,
        company_name VARCHAR(64),
        industry VARCHAR(64)
);

CREATE TABLE IF NOT EXISTS jobs (
        job_id CHAR(10) PRIMARY KEY,
        company_id CHAR(11) REFERENCES company (company_id),
        job_title VARCHAR(64),
        job_function VARCHAR(64),
        country VARCHAR(16),
```

```
        vacancies INTEGER NOT NULL
);

CREATE TABLE IF NOT EXISTS application (
        app_id CHAR(11) PRIMARY KEY,
        job_id CHAR(11) REFERENCES jobs (job_id),
        profile_id CHAR(11) REFERENCES candidates (profile_id),
        datetime TIMESTAMP
);

CREATE TABLE IF NOT EXISTS hired (
        app_id CHAR(11) REFERENCES application (app_id),
        datetime TIMESTAMP,
        hired BOOLEAN NOT NULL
);
```

Task 2

Problem Statement

The core of X0PA products often involve APIs around machine learning algorithms and NLP. In real world data especially in human resources, they come in different schemas and it is important to standardize them into one. For this problem, we will be looking into classifying any job titles into job functions. Initially, we tried to train a model that classify job titles into all job functions. However, we found out that the information technology job functions are too general and it is important for us to break them further down into various subclasses. By breaking them down, we are then able to match candidates to jobs to a higher accuracy.

This problem will test your ability to build a basic NLP model based on a given dataset. Your end task will be to:

- develop a model to predict one of the 16 classes (see **Variables Schema**);
- provide justification for model evaluation and report your results; and
- deploy your model in the form of an API endpoint (any API framework will do, but FastAPI is preferred).

You will be assessed on:

- your ability to build a model pipeline and deploy it as an API (50%);
- model accuracy (10%); and
- writing clean, readable code (40%).

You are highly recommended to write your code in Python. Do contact us if you wish to use other languages instead.

Dataset

You will be able to download the data from this link.

The training data ("x0pa_ds_interview_round_2_train.xlsx") consists of 50000 job titles and its respective Y labels while the test data (""x0pa_ds_interview_round_2_test.xlsx"") consists of 20000 job titles.

```
import pandas as pd
train_df = pd.read_excel("x0pa_ds_interview_round_2_train.xlsx", encoding = "utf-8")
test_df = pd.read_excel("x0pa_ds_interview_round_2_train.xlsx", encoding = "utf-8")
```

Variables Schema

Column Name	Description
id	A unique identifier for every job title. This is purely for our reference. Do not use it at all.
Job Title	Job Title scraped from the job description. Do take note that this data is unclean and may consist of unnecessary field. In various job titles, you will be able to see the duration as well. This column is your X label.
Type	Job Function of a job. This column is your Y label. It consists of the following 16 classes. Non-IT, Backend Engineer, Project Management, Product Management, Customer Support, Design, Data Science, Full Stack Engineer, Technical Support, Front End Engineer, Data Analyst, Mobile Application Developer, Database Administration, Cloud architect, Information Security, Network Administration.

Deliverables

- **Code**

  Please send us the script you used for your models and API. Please also let us know the versions of the Python libraries you use and ensure that your file runs as we will test it.

- **Results**

  Submit an updated test file with a new third column called "Type". The type will contain one of the 16 classes.

Code Plagiarism

You are allowed to refer to any resources you want. However, by submitting this response, you have automatically declared that your work is written by your own and not by anyone else.

Concluding the interview

This is the final round of technical interviews. Thank you for the hard work and applying to X0PA AI! After your submission, we will also give some brief comments on your work. In the meantime, stay safe and take care!