Technical test for Data Engineer

***Prerequisites***

Database engine of your choice (preferably PostgresQL) with full access rights.

**Description**

You will need to accomplish the following tasks and revert back as soon as possible. It is recommended not to take more than 4 hours and the results should be mailed back within 24 hours for consideration.

You are provided two datasets representing user location and the city locations. Please reach out to us if you have any questions / clarifications.

**Task 1**

You will need to accomplish the following:

1. Create a database with engine of your choice (preferably postgresql )
2. Import the data into a database
3. Build a query that outputs all user locations along with nearest city and corresponding distance
4. Build a query that outputs all users along with nearest city and corresponding distance

**Task 2**

Generate a histogram of number of users within a given distance to closest city. The end result should roughly resemble below:

|  |  |
| --- | --- |
| **Distance** | **Num Users** |
| 0-5 | xxx |
| 5-10 | xxx |
| 10-15 | xxx |
| 15-20 | xxx |
| 20-25 | xxx |
| 25-30 | xxx |
| 30-35 | xxx |
| >35 | xxx |

**Task 3**

(please answer in plain text)

We intend to use these queries for a reporting dashboard. How do you design the database and query for reporting performance?

**Task 4**

1. Develop a CRUD system with respect to a row of data.
2. Expose query built in step number 2 as an API
3. Expose an API to search by crop or location

**Deliverables**

* SQL Queries with results as CSV files
* Auxiliary comments / readme
* Text document for Task 3
* Code for Task 4
* Steps to deploy code in docker for Task 4 (optional)
* Live deployment of Task 4 on cloud (optional)

**Recommended tools :**

Postgres & Python