Junior applicant task

Imagine you are a data analyst at an e-commerce company. Your task is to set up a MySQL database, insert data into it, extract data, analyze it, and answer the following questions:

- What is the total revenue for each product category in the last year?
- What is the top-selling product for each month in the last year?
- Identify the customers who are within the top 10% in terms of generating revenue in 2022.

You are to create a MySQL database named sales_db with three tables. (You can use another database if you prefer.)

- **customers** with name and country
- **products** with name and category
- **sales** to hold sales transactions that identify the customer, the product, the quantity, the unit price, and the sale date. Timeframe should be the whole year of 2021 and 2022.

Tasks:

- Set up the environment. Use the mysql-connector-python library. (or other similar.)
- Database creation and connection: Create and connect to the MySQL database sales_db.
- Table creation and data insertion: Create the three tables and insert suitable data into them. You can generate this data yourself.
- Data extraction and analysis: Use SQL queries to extract the necessary data from the tables and answer the given questions.

Assignment Steps:

- Connect to your MySQL server and create a new database named sales_db.
- Create the customers, products, and sales tables with the fields specified above.
- Generate or adapt publicly available data to insert into these tables. Be sure to have enough data to make the analysis tasks meaningful.
- Use SQL queries to answer the questions given. For this, you might need to write complex queries using aggregate functions, subqueries, and JOINs. Make sure your SQL queries can run directly on a MySQL server and provide the correct answers.
- Document your findings and explain your process and reasoning.

Deliverables:

- A Python script (.py file) with the code to create the database and tables, insert the data, extract and analyze the data, clearly commented.
- A short report (can be a separate document or included as comments in the Python script) explaining your findings.
- SQL queries you used for data extraction, table creation and data insertion.
- Please submit these files via email.