ASSIGNMENT

SCENARIO:

Company Z is a biotech company that specialize in gene expression profiling for different cancer types. Throughout the year, the company has collected many data, and they want to store these data so that they can query it efficiently. You have been hired by Company Z to design and provide the storage solution and together with APIs implementation to query the data.

Information about Data:

- **Gene Information** which consists information about specific gene, covering the attributes like Gene ID, Gene Name, Chromosome, Start Position, End Position.
- **Expression Data** which consists of Sample ID, Gene ID, Expression Level, Condition (e.g.: cancer type, health)
- **Sample Information** which consists of Sample ID, Patient ID, Collection Date, Condition, Tissue Type.

Retrieval Requirement (for APIs):

1. To retrieve ALL information about specific gene by Gene ID
2. To retrieve expression data for a specific gene across all samples.
3. To retrieve all sample information for a specific patient.

You are given the data samples (in CSVs) in ZIP file in eLearning to start work on the MongoDB design.

YOUR RESPONSIBILITY:

MongoDB:

1. Based on the given sample data, go through the data and try to understand the relationship between the data samples and design an appropriate schema in MongoDB to store the data.
2. You need to produce a schema structure in JSON.
3. Then you implement and store all data in the CSVs into MongoDB according to your design. You should decide and define your own database name and collection name.

FastAPI:

1. Based on the API requirements given above, you need to design and implement the API accordingly in FastAPI.
2. Each requirement should resulted as individual API endpoint.
3. Each API developed must provide appropriate documentation and explanation about the functions of that API endpoint.

DELIVERABLES (MANDATORY):

1. The design of the MongoDB structure in JSON format *(in PDF)*
2. Export of your resulted MongoDB database. *(in .json)*
   You may export using following command in Terminal/Command Prompt:

   ```
   mongoexport -d your_db_name -c your_collection_name -o outputfilename.json
   ```

   Note: For the components in red please define and replace according to your case.
3. Python implementation (how the data from CSV are stored accordingly into MongoDB according to design) *(in .py or ipynb)*
4. Python implementation of the FastAPI. *(in .py)*
5. A README file that describe all the deliverables you upload in elearning. *(in .txt)*