Prompt Engineering Problem Set #3

You are required to use Meta Llama 3 8B model for the exercise below. Load the model as a server and issue the prompts from the client.

In this exercise, we will do a more complete product only using prompt engineering and SQL (or MongoDB) databases.

Product Goals:

We will build an poll analyzer product that uses the dataset from 2019 Indian General Elections. You are provided a CSV file named: details_of_assembly_segment_2019.csv. We will ingest this data in to a SQLite (or MongoDB) database with a suitable field names. Our goal is to receive a natural language statement as a query and using Meta Llama 3, generate a SQL statement. We will build a runtime module that receives the SQL, execute it and return the results. The results should be post processed using an LLM and final output generated. All the input/output should be performed through streamlit powered GUI as a chatbot interface.

Detailed Instructions:

1. Run Meta Llama3 8B model as a server using LMStudio or Ollama. If you are using Google Colab, you can load the LLM using transformers as explained during earlier sessions. Test your code by sending test prompts and evaluating the outputs. Ensure that your have set the inference parameters correctly, e.g. set temperature to 0, context length to 8k.
2. Data Preprocessing:
    a. Remove any unwanted rows/columns
    b. Detect any empty fields and fill them appropriately using Pandas
    c. Rename field names as it is beneficial in 2 aspects: It will be easy to write the SQL code, it also may be understood better by the LLM. A meaningful field name makes it easy for the LLM to understand the semantics better.

3. DB Creation: Create a database using the data obtained from step 2 as above. Name your database as “elections” and table name as “elections_2019”
4. Runtime:
    a. Write a runtime module that has a method “execute_sql”. This should accept a SQL string and execute, return the results.
    b. Write a suitable validation function that validates the LLM generated code and flags any harmful code, e.g. code that may delete records from database.
5. Questions Generation:
    a. Using some few shot examples of dataset records and using any LLM, e.g. ChatGPT, create a set of 25 questions that we will use to evaluate our product.
    b. For this exercise, we will produce only those questions that result in only one SQL statement (or one MongoDB query).
6. SQL Generation:
    a. Develop suitable prompts that receive a natural language question and generate SQL that can run on the runtime.
    b. Using Meta Llama 3, execute sample questions and generate SQL
7. Execute SQL: run the SQL using runtime, consolidate the results
8. Post Process: Write a suitable prompt that accepts the SQL results, formats it in a human consumable form and returns
9. UI development: Build a GUI based chatbot that supports the product we described. You can use streamlit for this purpose.
10. Show the demo, create a directory under the shared folder and upload your work.