QAC 387

Assignment 3: LLM-powered data analysis application

In this assignment, you will build an AI data analysis assistant. The type of assistant you build can be a part of your bigger course project, so you can build something that is tailored toward the goals of your project. However, **the AI assistant should be built using LangChain, OpenAI, and Streamlit**.

Your assistant should be designed to help automate or enhance a component of your project pipeline, such as coding, data cleaning, statistical analysis, visualization, model interpretation, or report generation. The goal is to use this assignment to create something that moves your project along. The agent should be capable of analyzing data or guiding users through other data-related tasks.

The core requirements of your agent are that it:

1. Is implemented using LangChain with an OpenAI LLM
2. Uses Streamlit to provide a user interface for input, interaction, and output display.
3. Accepts user input in the form of a question, task, or instruction related to a dataset or data analysis goal.
4. Executes without error and returns output to the user.
5. Provides a natural language explanation/interpretation of the output.
6. Is evaluated for accuracy.

You should build your application in a virtual environment.

The deliverable for this assignment is a GitHub repository for your project that is shared with the instructor. The repository should include:

1. A README.md file that includes purpose of the application, instructions for using the app, and cautions for using the app.
2. A requirements.txt file with your project dependencies.
3. A *.py file with the code that develops and executes the app via a Streamlit UI
4. The data set used to develop the application.
5. A document that contains a log of the accuracy evaluation you conducted, any issues, and how the issues were resolved.