

Home assignment for Senior Data Scientist position

Objective:

The main objective of this exercise is to design a comprehensive presentation on developing a LLM analytics tool tailored for exploring, analyzing, predicting, and generating insights and recommendations from tabular data. You will be provided with a dataset (attached separately) to serve as the foundation for your tool's development process and analysis.

Deliverables:

- A detailed presentation in PowerPoint, Google Slides, or a similar format.
- A brief report summarizing your approach, methodologies, with possible predictions, insights, and recommendations.
- Architecture diagram of the proposed tool
- (Optional) pseudo code snippets showcasing the implementation of your proposed solutions.

* You don't have to run the model/s and we don't check its correctness

Dataset:

The provided dataset contains 3 tables in the E-commerce domain. It includes a variety of features/columns that the LLM can use to analyze and generate insights from.

Task Details:

Exploratory Data Analysis (EDA):

- Perform an initial exploration of the dataset to understand its characteristics, including distribution of key features, missing values, potential outliers, and correlations between features.
- Present your findings with appropriate visualizations and statistics.

Analytics and Insights Generation:

- Describe how you would use LLMs and related frameworks to analyze the dataset. Focus on how NLP can extract insights, trends, and patterns that might not be immediately apparent.
- Provide examples of insights or recommendations that could be generated from the data.

Generating Recommendations:

- Explain how the model can be used to generate actionable recommendations based on its predictions and insights.
- Include at least one scenario or use case where these recommendations could be applied to achieve a specific goal or solve a problem.

[Optional] Predictive Modeling:

- Propose a model architecture for predicting outcomes based on the dataset. Explain your choice of model, including any variations specific to LLMs that make it suitable for tabular data.
- Discuss how you would train, validate, and test your model, including any metrics you would use to evaluate its performance.
-

Challenges and Solutions:

- Identify potential challenges in designing and implementing an LLM analytics tool for tabular data.
- Offer solutions or approaches to overcome these challenges.

Future Work:

- Discuss potential improvements or next steps in evolving the tool to handle larger datasets, more complex analyses, or real-time data processing.

Evaluation Criteria:

- Clarity and Coherence: Presentation and report are well-organized, with clear explanations of technical concepts and methodologies.
- Creativity and Innovation: Shows originality in approach to data analysis, insight generation, and problem-solving.
- Practicality and Application: Insights and recommendations are actionable and relevant to real-world scenarios.
- Technical Proficiency: Effective use of data science and machine learning tools and techniques.

Submission Guidelines:

- Submit your presentation and report via email to avi.grushka@sisense.com cc: adi.feldman@sisense.com

Thanks in advance and Good luck!