

Task: Integrate LLM for Employee Activity Tracking

Objective: Develop a system where an LLM (Large Language Model) reads an SQL database tracking employee activities and generates natural language summaries based on natural language queries.

Database Fields:

- **Employee ID:** Unique identifier for each employee.
- **Week Number:** (1-10)
- **Number of Meetings:** Integer representing the number of meetings attended.
- **Total Sales (RMB):** Decimal representing the total sales value in RMB.
- **Hours Worked:** Decimal representing the total hours worked.
- **Activities:** Text field describing activities such as preparation for meetings, sales strategies, challenges faced, and solutions implemented.
- **Department:** Text field indicating the employee's department (e.g., Sales, Marketing, Product Development, Finance, IT).
- **Hire Date:** Date indicating when the employee was hired.
- **Email Address:** Employee's email address.
- **Job Title:** Employee's job title (e.g., Sales Manager, Data Analyst, Marketing Specialist).

Requirements:

1. Database Creation and Population:

- Create an SQL database with the fields outlined above.
- Generate synthetic data for 10 employees over a 10-week period.
- Ensure numerical data is realistic and consistent.
- Provide meaningful qualitative entries in the "Activities" field that reflect typical work activities.

2. LLM Integration:

- Implement an LLM that can read the database and understand its structure.
- The LLM should be able to process natural language queries and generate appropriate responses, including:
 - Numerical summaries (e.g., total meetings, total sales, total hours worked).
 - Qualitative summaries (e.g., descriptions of activities, challenges, and solutions).

3. Query Functionality:

- Allow users to input queries in natural language.
- The LLM should be able to understand and respond to a variety of query types including:
 - Point queries (e.g., "What did John do from weeks 3-7?")
 - Aggregation queries (e.g., "How many employees does the company have?")
 - Knowledge & capabilities queries (e.g., "Which employees were hired during a time of industry recession?")
 - Reasoning queries (e.g., "Who are the employees that faced challenges with customer retention?")
 - Match-based queries (e.g., "List all employees who work in the IT department.")
 - Comparison queries (e.g., "Compare the hours worked by 'Wei Zhang' and 'Tao Huang' during week 1.")
 - Ranking queries (e.g., "Who are the top 3 employees by total hours worked during the last 4 weeks?")

4. **Output Format:**

- Ensure the numerical data adds up correctly and provides a precise summary.
- The qualitative data should be summarized in natural language, reflecting the entries in the database.

Example Queries (20):

1. What is the email address of the employee who is the Sales Manager?
2. Which employee in the company works in the Product Development department?
3. What was the sales revenue of 'Wei Zhang' for the week starting on '2024-08-28'?
4. Who are the employees working in the 'Finance' department?
5. Retrieve the total number of meetings attended by 'Na Li' in her weekly updates.
6. Which employees worked more than 40 hours during week 1?
7. How many employees does the company have in total?
8. What is the average hours worked by all employees during week 2?
9. How much total sales revenue has the Sales department generated to date?
10. What is the total sales revenue generated by the company during week 1?

11. Who worked the most hours during the first week of September 2024?
12. Which employee attended the most meetings during week 2?
13. Which employees in the company were hired during a time of industry recession (requires external knowledge)?
14. Who are the employees that faced challenges with customer retention, and what solutions did they propose?
15. Which employees work in roles that likely require data analysis or reporting skills (based on job titles)?
16. List all employees who work in the IT department within the company.
17. Compare the hours worked by 'Wei Zhang' and 'Tao Huang' during week 1.
18. Who are the top 3 employees by total hours worked during the last 4 weeks?
19. Who achieved the highest sales revenue in a single week, and when?
20. What is the total number of hours worked and average sales revenue for employees in the Business Development department?

Directions for Completion:

1. Complete the task, designing and implementing the database and LLM integration as described.
2. **Benchmark 1:** Run the 20 example queries through your system and record the answers generated by the LLM.
3. **System Improvement:** Analyze the results of Benchmark 1, identifying areas where the LLM's performance could be improved (e.g., accuracy, clarity, completeness). Modify your system (database structure, LLM prompts, etc.) to address these weaknesses.
4. **Benchmark 2:** Run the same 20 example queries through your improved system and record the new answers generated by the LLM.
5. **Analysis and Description:** Describe how you designed your system initially, what modifications you made for improvement, and compare the results of Benchmark 1 and Benchmark 2. Analyze how the changes you made impacted the LLM's performance.