

Natural Language Q&A System for Retail Sales Using Google Palm LLM

Introduction:

The retail sector is increasingly leveraging technology to gain insights from large data volumes. Managers often require real-time access to sales data for better decision-making, but traditional reporting systems can be cumbersome and not user-friendly. To bridge this gap, the project developed a Natural Language Q&A system for retail sales, using Google Palm LLM, Langchain framework, and a MySQL database.

This system enables store managers to interact with sales data in natural language, eliminating the need for technical knowledge of SQL queries. The Natural Language Processing (NLP) capabilities of Google Palm LLM allow for intuitive and seamless querying, making it possible for managers to retrieve meaningful insights quickly.

Objectives:

- To build an end-to-end Natural Language Q&A system for a retail store.
 - To enable the system to interpret natural language queries and retrieve the corresponding sales data from a MySQL database.
 - To simplify the interaction between the user and the database by removing the need for SQL query writing.
 - To ensure accurate and efficient responses to user queries through a user-friendly interface.
-

System Architecture:

The Natural Language Q&A system was built using several components to ensure seamless interaction between the user and the sales database. The key elements of the system architecture include:

1. Google Palm LLM:

- A large language model capable of understanding and interpreting complex natural language queries.
- Used to process questions input by the user and break them down into SQL-like commands that can be executed on the database.

2. Langchain Framework:

- A versatile NLP tool that facilitates the interaction between the language model and the backend database.

- Integrates Google Palm LLM with other components to manage the query flow and response generation.
3. **MySQL Database:**
- The retail sales data, including product details, sales transactions, and customer information, is stored in a MySQL database.
 - The system retrieves relevant data from the database based on user queries and presents it in a clear format.
4. **Streamlit User Interface:**
- A web-based user interface built using Streamlit, designed for ease of use.
 - Store managers can input their questions in natural language through the interface.
 - The system then provides real-time, accurate responses based on the sales data.
-

Methodology:

The project was developed using the following approach:

1. **Data Storage:**
- Sales data, including transactions, product information, and other key metrics, were stored in a structured format in the MySQL database.
2. **Language Processing:**
- The Langchain framework connected the natural language queries from the user interface to the backend.
 - Google Palm LLM was employed to understand and translate natural language into database queries (SQL).
3. **User Interface Development:**
- A simple and intuitive UI was built using Streamlit to allow managers to input questions.
 - The UI supported queries such as "What were the total sales for Nike T-shirts last month?" or "Which product had the highest sales last quarter?"
4. **Data Retrieval and Presentation:**
- Based on the query, relevant SQL statements were generated and executed on the MySQL database.
 - The results were fetched from the database and displayed in an easily interpretable format on the UI.
-

Challenges:

- Ensuring accuracy and relevance in the translation of natural language queries to SQL commands.
 - Handling a wide range of questions with variations in phrasing while ensuring the system responds with correct data.
 - Integrating the language model with the database to maintain real-time processing and response times.
-

Results:

The system was successfully developed to:

- **Answer sales-related questions** from store managers using natural language inputs.
 - Provide accurate data insights, including **sales trends, product performance, and periodic sales reports**, without requiring SQL query writing.
 - The **streamlined user interface** significantly reduced the time spent on sales data retrieval and enhanced the decision-making process for store managers.
-

Future Enhancements:

- Expanding the system to handle more complex customer segmentation or predictive analytics queries.
 - Incorporating additional databases or real-time data feeds for a more comprehensive view of sales performance.
 - Improving the natural language understanding capabilities to cover a broader range of questions and formats.
-

Conclusion:

This project demonstrated the potential of using NLP technologies to simplify and enhance the interaction between users and databases. By leveraging Google Palm LLM, Langchain, and Streamlit, the system successfully bridged the gap between technical data storage and everyday sales queries. This approach not only made sales data more accessible to non-technical users but also improved the efficiency and accuracy of business decision-making.

By Ansh Tyagi

MSc Statistics & Data Science