AI-Based E-Commerce SQL Agent

Project Overview

The AI-Based E-Commerce SQL Agent is a Python-powered assistant that allows users to ask natural language questions related to e-commerce data such as sales, ads, and performance. The agent translates these questions into SQL queries, executes them on a SQLite database, and returns human-readable answers, along with optional data visualizations.

This project was developed as part of a GenAI Internship Interview Task at VIT, aiming to demonstrate the integration of LLM-style logic, SQL querying, and graph generation without using actual cloud APIs due to limitations. A rule-based approach simulates an LLM for local execution.

Key Features

- Converts natural questions into SQL (like "What is my total sales?")
- Retrieves and processes data from a SQLite database
- Plots real-time graphs for better visualization of trends
- Works completely offline – no need for real API calls
- Video demo included to show complete interaction
- Easy to extend with real LLM APIs like Gemini or Ollama

Dataset Descriptions

1. AdSales.csv

Column	Description
date	Date of the ad performance
item_id	ID of the product
ad_sales	Revenue generated via advertisements
impressions	Number of times the ad was shown
ad_spend	Amount spent on the ad
clicks	Number of clicks received
units_sold	Number of products sold via the ad

Used for analyzing advertising performance (e.g., ROAS, top spending items).

---

2. TotalSales.csv

Column	Description
date	Date of the sales
item_id	ID of the product sold
total_sales	Total revenue on that date
total_units_ordered	Quantity sold on that date

Used for tracking product sales performance and generating time-series plots.

---

3. Eligibility.csv

Column	Description
item_id	Unique ID of the product
price	Selling price of the product
eligible	Whether the item is eligible for ads

Useful for filtering eligible products for marketing campaigns.

Project Structure

```
AI-Ecommerce-SQL-Agent/
|
├── AdSales.csv            # Ad performance data
├── TotalSales.csv         # Sales data over time
├── Eligibility.csv        # Product eligibility info
├── create_db.py           # Creates SQLite database from CSVs
├── ecommerce.db           # Final database with 3 tables
├── llm_sql_agent.py       # Python agent that answers user questions
├── AI_AGENT.mp4           # Demo video with Q&A + Graphs
├── README.md              # Project summary
```

Sample Questions Handled

- What is my total sales?
- What is the ROAS (Return on Ad Spend)?
- Which item had the highest ad spend?

Each of these returns:

- SQL Query
- Human-readable answer
- Optional graph

How It Works

1. User Input → "Which item had the highest ad spend?"
2. LLM Logic (rule-based) → SQL:
3. SELECT item_id, MAX(ad_spend) FROM ad_sales;
4. Query Execution → SQLite database
5. Answer + Graph (if applicable) → Shown to user

Future Scope

- Integrate with real LLM APIs like Gemini, OpenAI GPT, or Ollama
- Handle complex queries (multi-table joins, filters)
- Add voice interface or web-based UI
- Enable user authentication and dynamic dashboards

THANK YOU