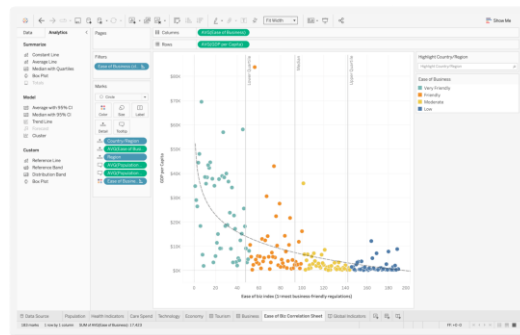
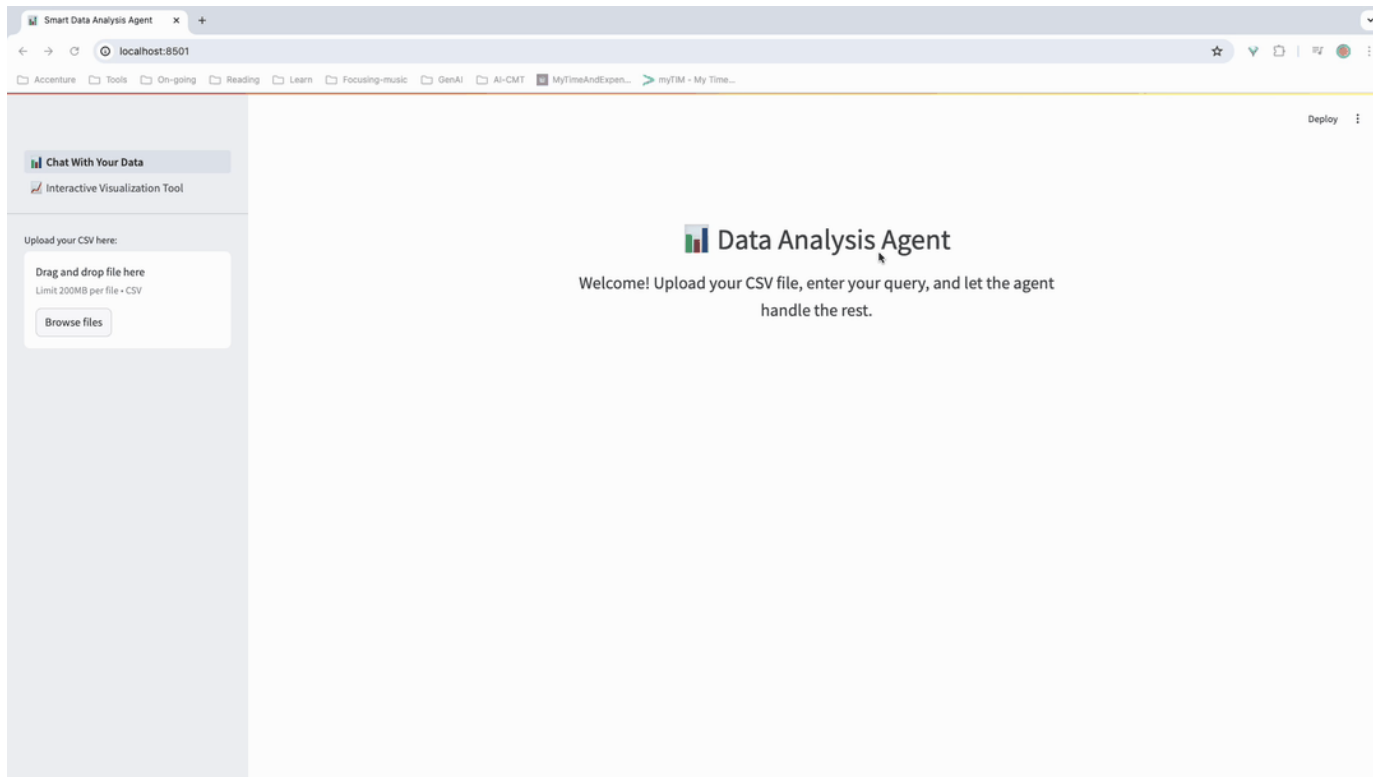


Tạo Công cụ phân tích dữ liệu tùy chỉnh với LLMs

Easy Step-by-Step Guide



1. Introduction to DAGPT - Demo



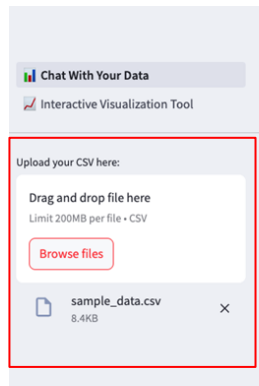
1. Introduction to DAGPT - Overview

- **DAGPT:** Data Analysis GPT là một công cụ phân tích dữ liệu hữu ích sử dụng LLM để hỗ trợ các tác vụ phân tích và thao tác dữ liệu thông qua giao diện hội thoại.
- Sử dụng **Streamlit** để cung cấp ứng dụng Web tương tác nơi người dùng có thể tải lên dữ liệu của họ, đặt câu hỏi, khám phá dữ liệu của bạn một cách tương tác và nhận thông tin chi tiết.
- Công cụ thay thế cho **Chat GPT-4o** và **Tableau** và phù hợp với mọi tác vụ phân tích dữ liệu trong máy cục bộ của bạn.

1. Introduction to DAGPT - Features (1/3)

- **CSV File Upload:** Dễ dàng tải lên dữ liệu CSV của bạn thông qua thanh bên (sidebar).
- **Data Analysis:** Nhập truy vấn về dữ liệu của bạn và nhận phản hồi do LLM cung cấp.
- **Data Visualization:** Tạo và hiển thị các đồ thị dựa trên truy vấn dữ liệu của bạn.
- **Interactive Visualization Tool:** Khám phá dữ liệu của bạn một cách tương tác bằng cách sử dụng nhiều hình ảnh trực quan khác nhau. Tải tập dữ liệu của bạn lên và bắt đầu phân tích bằng chức năng kéo và thả để tạo biểu đồ tùy chỉnh.

1. Introduction to DAGPT - Features (2/3)



CSV File Upload: Easily upload your CSV data through the sidebar.

Data Analysis Agent

Welcome! Upload your CSV file, enter your query, and let the agent handle the rest.

DataFrame Head:

	Store Name	Date	Product	Unit Price	Sales	Revenue	Stock Quantity	Profit	Store
0	7-Eleven	2023/1/8	Bakery	623.95	33	20,590.35	67	8,236.14	Osaki
1	7-Eleven	2023/1/23	Bakery	714.97	40	28,598.8	46	11,439.52	Osaki
2	7-Eleven	2023/2/2	Bakery	174.14	41	7,139.74	79	2,855.896	Osaki
3	7-Eleven	2023/3/8	Bakery	816.88	44	35,942.72	37	14,377.088	Osaki
4	7-Eleven	2023/1/29	Beef Bowl	247.26	9	2,225.34	23	890.136	Osaki

Select Agent Type

AgentType.OPENAI_FUNCTIONS

Enter your query:

plot sale distribution

Run Query

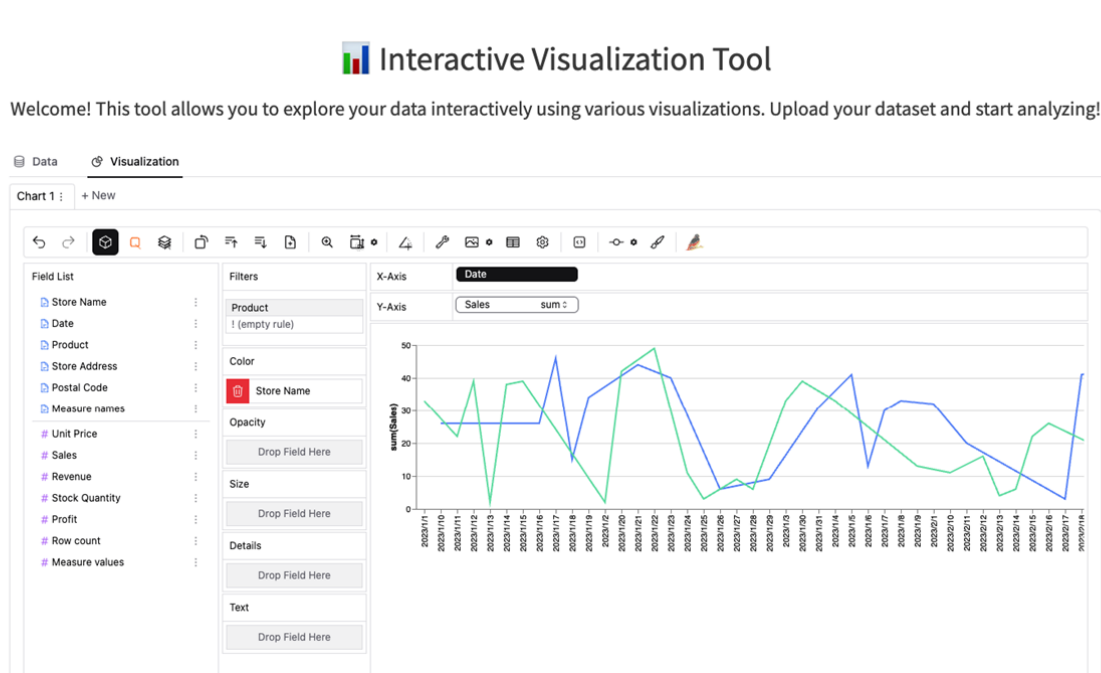
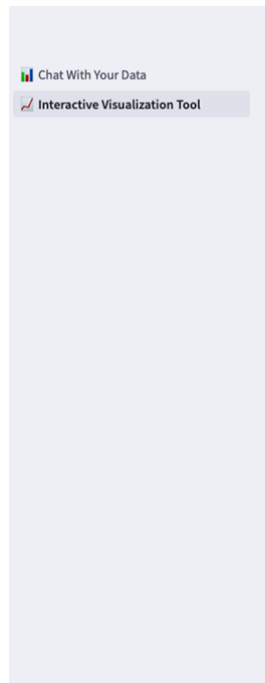
Chat History:



Data Analysis: Enter queries about your data and get responses powered by LLMs.

Data Visualization: Generate and display plots based on your data queries.

1. Introduction to DAGPT - Features (3/3)



Interactive Visualization Tool: Explore your data interactively using various visualizations. Upload csv data and start analyzing with drag-and-drop functionality to create customized charts.

2. Tools and Packages Needed to Build This App

Python 3.9 or Higher:

- Ngôn ngữ lập trình chính được sử dụng để phát triển ứng dụng.

Langchain:

- Một framework để phát triển các ứng dụng được hỗ trợ bởi các mô hình ngôn ngữ. Nó cung cấp các công cụ để kết nối các mô hình ngôn ngữ với các nguồn dữ liệu khác, quản lý tương tác và tạo các ứng dụng AI nâng cao.

Streamlit:

- Một framework mã nguồn mở cho Machine Learning and Data Science projects. Nó cho phép bạn tạo các ứng dụng web tương tác bằng Python.

Pygwalker:

- Một công cụ để tạo trực quan tương tác kéo thả.

OpenAI API (optional):

- Được sử dụng để truy cập các mô hình ngôn ngữ nâng cao để nâng cao khả năng.



3. Tutorial - Steps

Step 1

Basic setup

- Xây dựng cấu trúc dự án
- Thiết lập môi trường ảo
- Cài đặt các thư viện cần thiết
- Thiết lập github để lưu trữ dự án

Step 2

Build “Chat With Your Data” Tool

- Sử dụng streamlit để xây dựng giao diện UI
- Sử dụng bộ công cụ agent toolkits do langchain cung cấp để xây dựng agent phân tích dữ liệu
- Use OpenAI GPT model

Step 3

Build “Interactive Visualization” Tool

- Embed pygwalker vào giao diện để xây dựng công cụ trực quan hóa tương tác

3. Tutorial - Basic setup

Project structure:

- Sử dụng “Cookiecutter Data Science” để xây dựng một cấu trúc dự án được tiêu chuẩn hợp lý.

Setup virtual environment:

```
1 python -m venv .venv
2 source .venv/bin/activate # On Windows use `.venv\Scripts\activate`
```

Install necessary packages: `pip install -r requirements.txt`

```
# requirements.txt
langchain==0.2.6
langchain-community==0.2.6
langchain-openai==0.1.13
streamlit==1.36.0
pygwalker==0.4.9.1
pandas==1.5.3
matplotlib==3.9.0
```

3. Tutorial - Live coding

```
1 def main():
2
3     st.set_page_config(page_title="Smart Data Analysis Tool", page_icon="📊", layout="centered")
4
5     st.header("📊 Smart Data Analysis Tool")
6
7     # Load model
8     llm = load_llm(model_name=MODEL_NAME)
9     logger.info(f"Successfully loaded {MODEL_NAME} !")
10
11     with st.sidebar:
12         uploaded_file = st.file_uploader("Upload your csv file here: ", type="csv")
13
14         if "history" not in st.session_state:
15             st.session_state.history = []
16
17         if uploaded_file is not None:
18             st.session_state.df = pd.read_csv(uploaded_file)
19
20             # display uploaded dataframe
21             display_data(st.session_state.df)
22
23             # Create our agent to help us answer all our question about our data
24             da_agent = create_pandas_dataframe_agent(
25                 llm=llm,
26                 df=st.session_state.df,
27                 allow_dangerous_code=True,
28                 verbose=True,
29                 return_intermediate_steps=True,
30                 agent_type="tool-calling",
31             )
32
33             # Query
34             query = st.text_input("Enter your question: ")
35
36             if st.button("Run Query"):
37                 with st.spinner("Processing..."):
38                     process_query(da_agent, query)
39
40             st.divider()
41             display_chat_history()
42
43
44 if __name__ == "__main__":
45     main()
```

4. Conclusion

- Với khả năng LLMs, **việc viết một ứng dụng thực hiện các nhiệm vụ phân tích dữ liệu phức tạp hàng ngày là rất dễ dàng.**
- Sử dụng **Streamlit** giúp dễ dàng xây dựng giao diện đàm thoại và khả năng hiển thị mạnh mẽ, giúp mọi người có thể truy cập phân tích dữ liệu.
- **Pygwalker** là một công cụ miễn phí và hữu ích thay thế cho Tableau và có thể được nhúng vào ứng dụng cá nhân.
- Bằng cách làm theo hướng dẫn này, **mọi người đều có thể xây dựng công cụ phân tích dữ liệu của riêng mình bằng cách tận dụng sức mạnh của LLMs và Streamlit.**

Phụ lục

How to setup OpenAI API key

Sign Up / Log In to OpenAI:

- Visit the [OpenAI website](#) and create an account or log in if you already have one.

Navigate to API Keys:

- Once logged in, go to the API section in your account dashboard. This is typically found under the "API Keys" tab.

Create a New API Key:

- Click on the "Create new secret key" button. This will generate a new API key for you to use.

Copy the API Key:

- After the key is generated, make sure to copy it and replace your API key in the ``.env`` file in your project.

```
1 OPENAI_API_KEY=your-openai-api-key
```