AI-Assistant Data Science

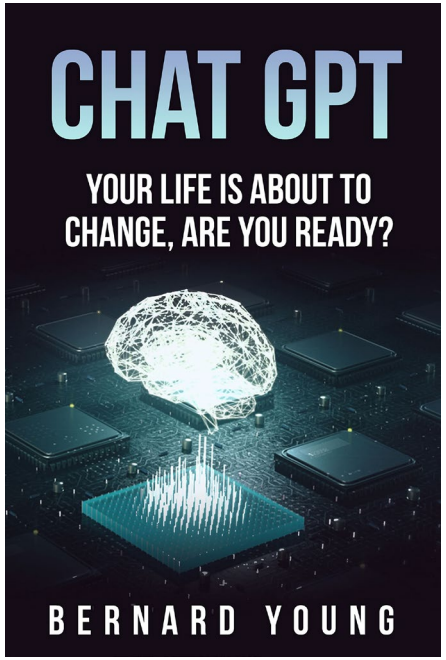
CSC 405/605/705



Do You Have Following Questions?



- Question 1: If I already know Python, Pandas, Visualization, Statistics and Machine Learning Models, can I work as a data scientist today?



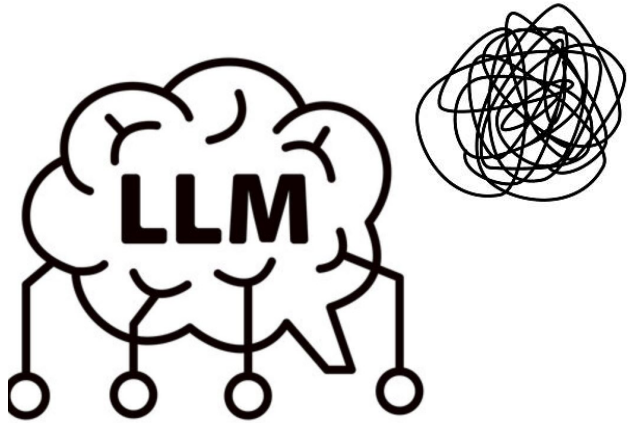
Those skills are still necessary, but they are often not sufficient anymore.

UNC GREENSBORO

Do You Have Following Questions?

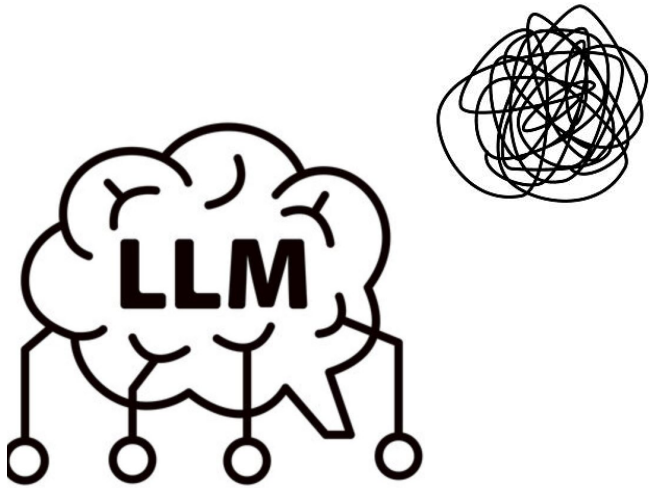


- Question 2: If LLMs can write code and analyze data, why do I need to learn those foundations?



Because LLMs do not understand your business problem automatically.

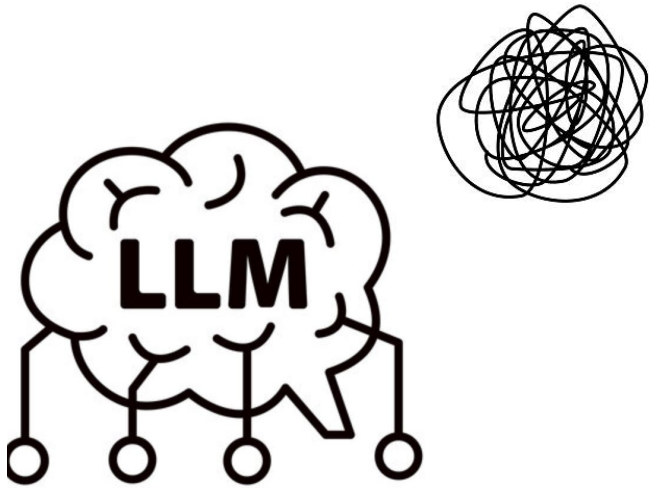
What do LLMs struggle with Data Analysis?



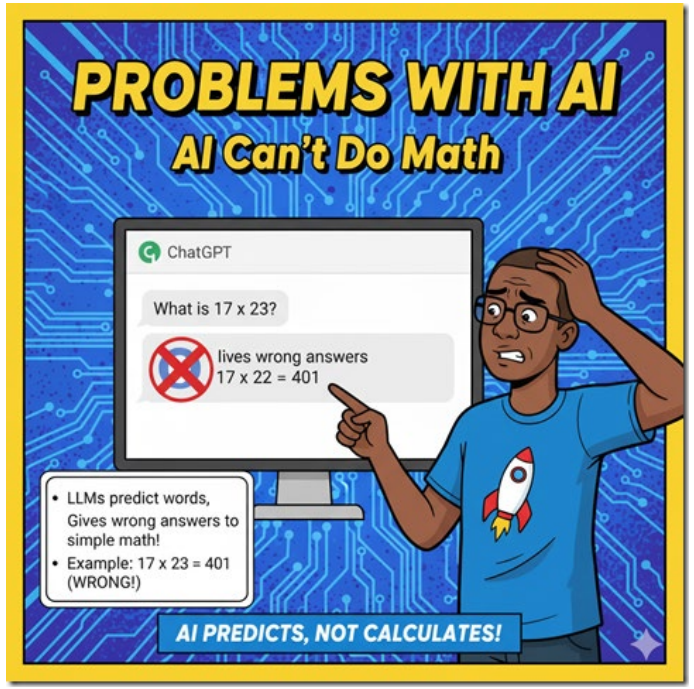
- **AI is powerful but aimless**
  - AI is not alive; it doesn't have desires.
  - AI doesn't know why it's doing something. It does what it's told (and sometimes what it infers).
  - If the context is missing or inconsistent, the AI will make assumptions, and often wrong ones.
  - An LLM doesn't have unlimited memory of your world; it needs structure to keep on track.

UNC GREENSBORO

What do LLMs struggle with Data Analysis?

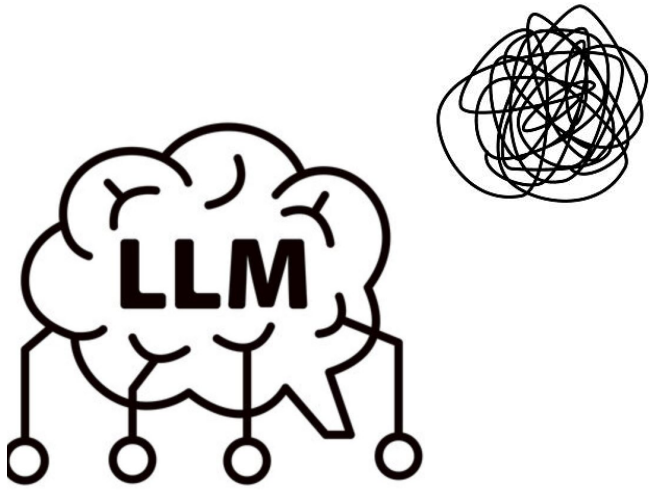


- **AI can’t do math**
  - AI predicts what a correct-looking answer might be. The result can sound perfectly confident and well-formatted but still be entirely wrong.



UNC GREENSBORO

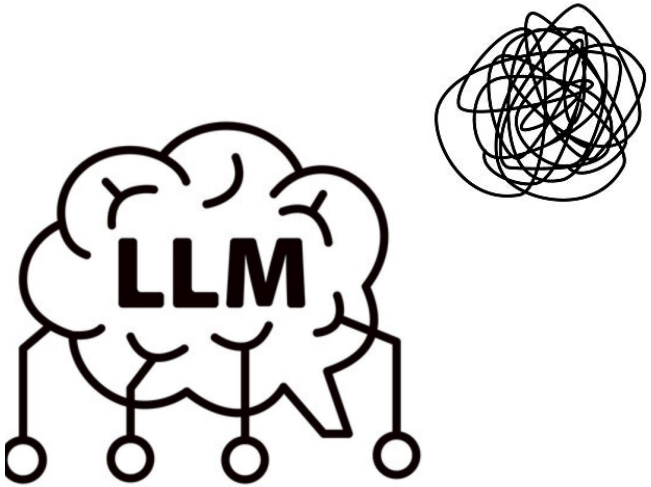
What do LLMs struggle with Data Analysis?



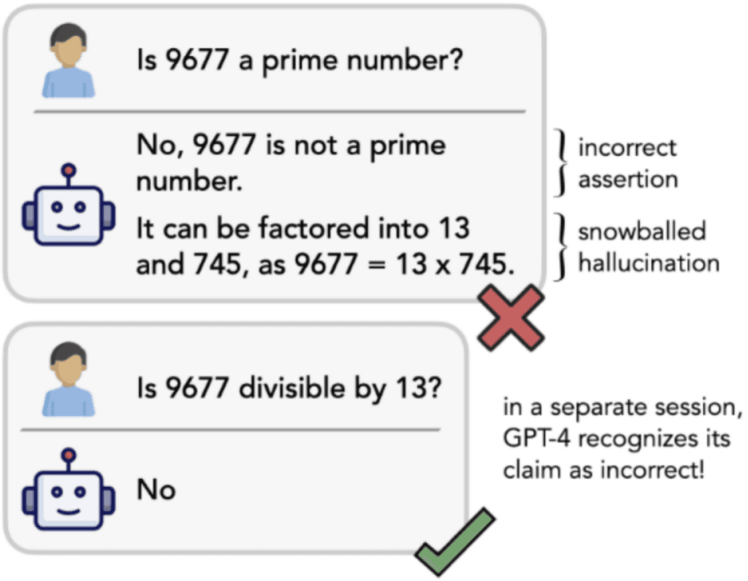
- **AI can’t handle tabular data**
  - LLMs process text, so converting a grid of numbers into text format loses meaning.
- **AI has no real-world experience**
  - They cannot know how to handle missing data or guess missing information or catch errors.

UNC GREENSBORO

What do LLMs struggle with Data Analysis?



- **AI often hallucinate**
  - AI confidently create false information, code, or statistics.
  - AI cannot debug error created by itself. Human Oversight and Verification are required.



Do You Have Following Questions?



- Question 3: Then What is Changed by LLMs?

The role is shifting:

Before

Write every line of analysis code manually.

Now

Guide an AI assistant to generate, test, and refine analyses.

UNC GREENSBORO

Do You Have Following Questions?

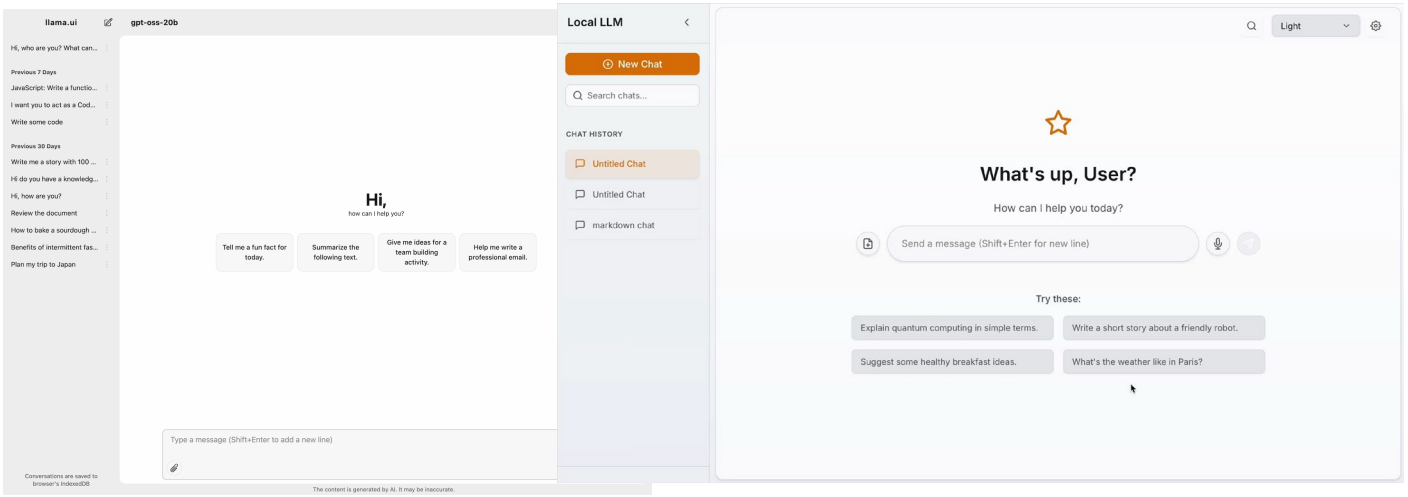


Do You Have Following Questions?



- Question 4: How Do We Control an LLM?

Graphical User Interfaces (GUIs)?

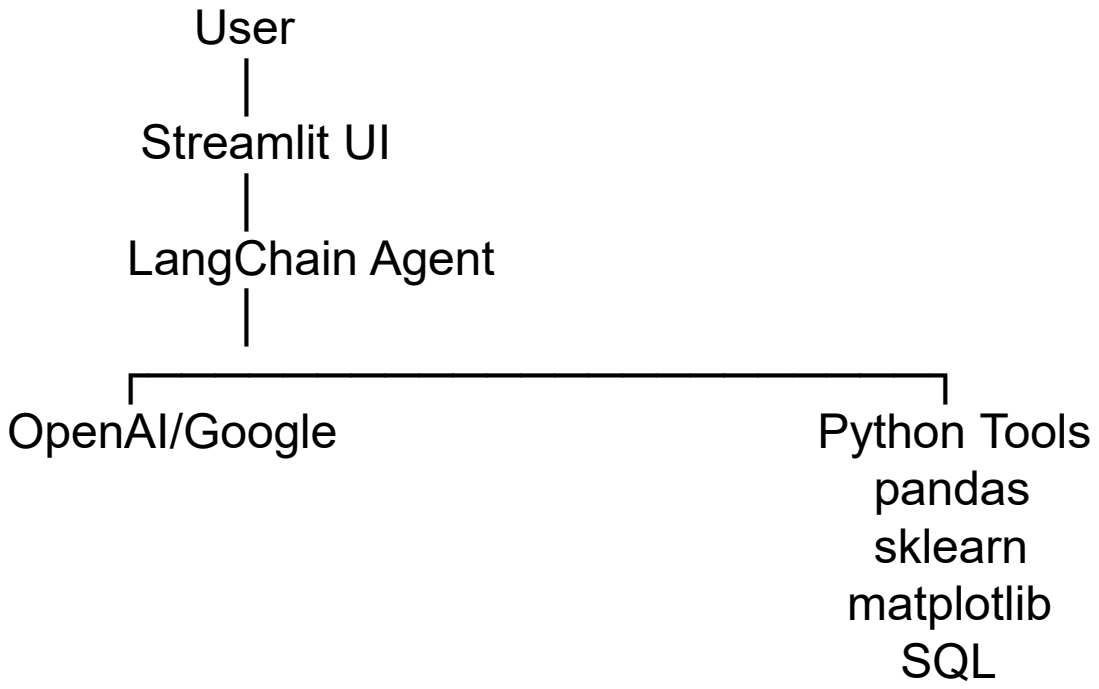


Application Programming Interfaces (APIs)

AI-assisted Data Science Workflow

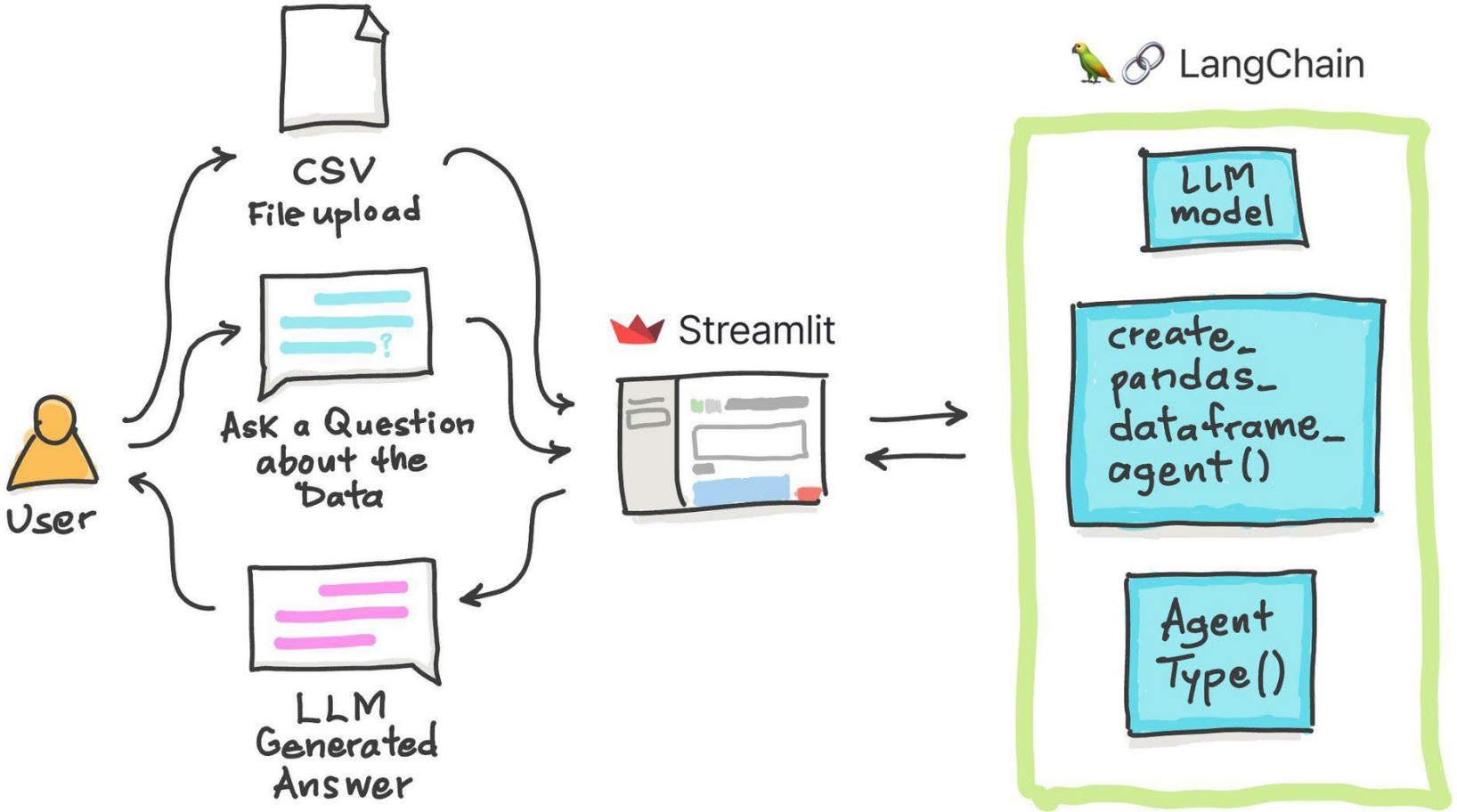
- We will build an AI-assisted workflow

Tool	Purpose
Streamlit	Build an interactive data science application.
LangChain	Connect LLMs with data, tools, and workflows.
Python/statistics	Provide the analytical foundation that guides the LLM.



UNC GREENSBORO

AI-assisted Data Science Workflow



➢ Introduction of Streamlit



UNC GREENSBORO

https://docs.streamlit.io/

What is Streamlit?

- An open-source Python framework for building interactive web applications. Designed for data scientists and ML engineers—no front-end expertise needed.
- Turns Python scripts into web apps with minimal effort.
- Uses a declarative approach—write Python, and Streamlit handles the UI.

Why Use Streamlit?

- **Simplicity**: Create web apps with just a few lines of Python.
- **Interactive Widgets**: Sliders, buttons, file uploads, and more.
- **Rapid Prototyping**: Ideal for showcasing ML models and data visualizations.
- **No Frontend Needed**: No HTML, CSS, or JavaScript required.
- **Deployment Friendly**: Easily share apps via Streamlit Cloud or other platforms.

UNC GREENSBORO

When to Use Streamlit?

- ✅ Building Interactive Data Dashboards
- ✅ Quick Prototyping for ML & AI Models
- ✅ Creating Custom Web Apps for Data Science
- ✅ Showcasing Research & Analytical Reports
- ✅ Internal Tools for Teams & Businesses

UNC GREENSBORO

When to Use Streamlit?

- ✅ Building Interactive Data Dashboards
- ✅ Quick Prototyping for ML & AI Models
- ✅ Creating Custom Web Apps for Data Science
- ✅ **Showcasing Research & Analytical Reports**
- ✅ Internal Tools for Teams & Businesses

UNC GREENSBORO

Install Streamlit

```
TERMINAL

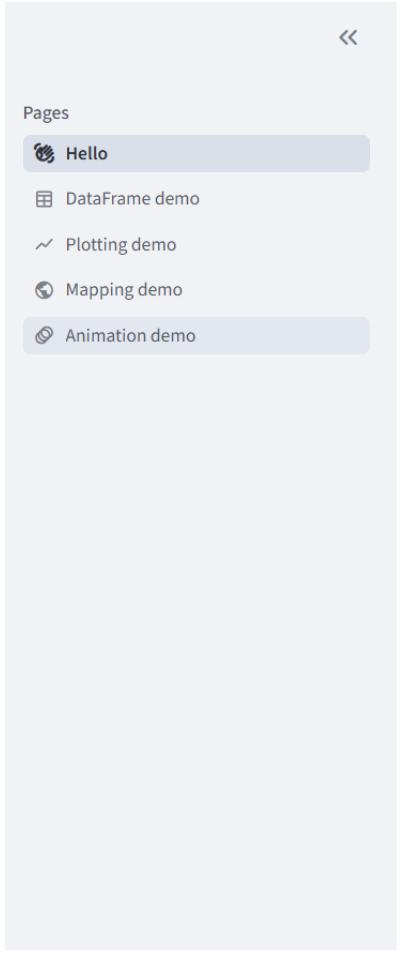
pip install streamlit
streamlit hello
```

When test ‘streamlit hello’, you may encounter:
Error: The term 'streamlit' is not recognized as the name of a cmdlet, function, script file, or operable program. Check the spelling of the name, or if a path was included, verify that the path is correct and try again. At line:1 char:1 + streamlit hello + ~~~~~~~~~~ + CategoryInfo : ObjectNotFound: (streamlit:String) [], CommandNotFoundException + FullyQualifiedErrorId : CommandNotFoundException

python -m streamlit hello

UNC GREENSBORO

Install Streamlit



Welcome to Streamlit! 👋

Streamlit is an open-source app framework built specifically for machine learning and data science projects.

👈 **Select a demo from the sidebar** to see some examples of what Streamlit can do!

Want to learn more?

- Check out streamlit.io
- Jump into our documentation
- Ask a question in our community forums

See more complex demos

- Use a neural net to analyze the Udacity Self-driving Car Image Dataset
- Explore a New York City rideshare dataset

UNC GREENSBORO

Create a Streamlit APP

```
1   import streamlit as st
2
3   st.title("My First Streamlit App")
4
5   st.write("Hello World!")
6
```

UNC GREENSBORO

Create a Streamlit APP

```
PS C:\Users\X_LI12\Documents\data_science\AI-assistant_Data_Science> python -m streamlit run basic_streamlit.py
2026-07-10 15:29:01.992 Uvicorn server started on 0.0.0.0:8501

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.1.62:8501
```

http://localhost:8501/



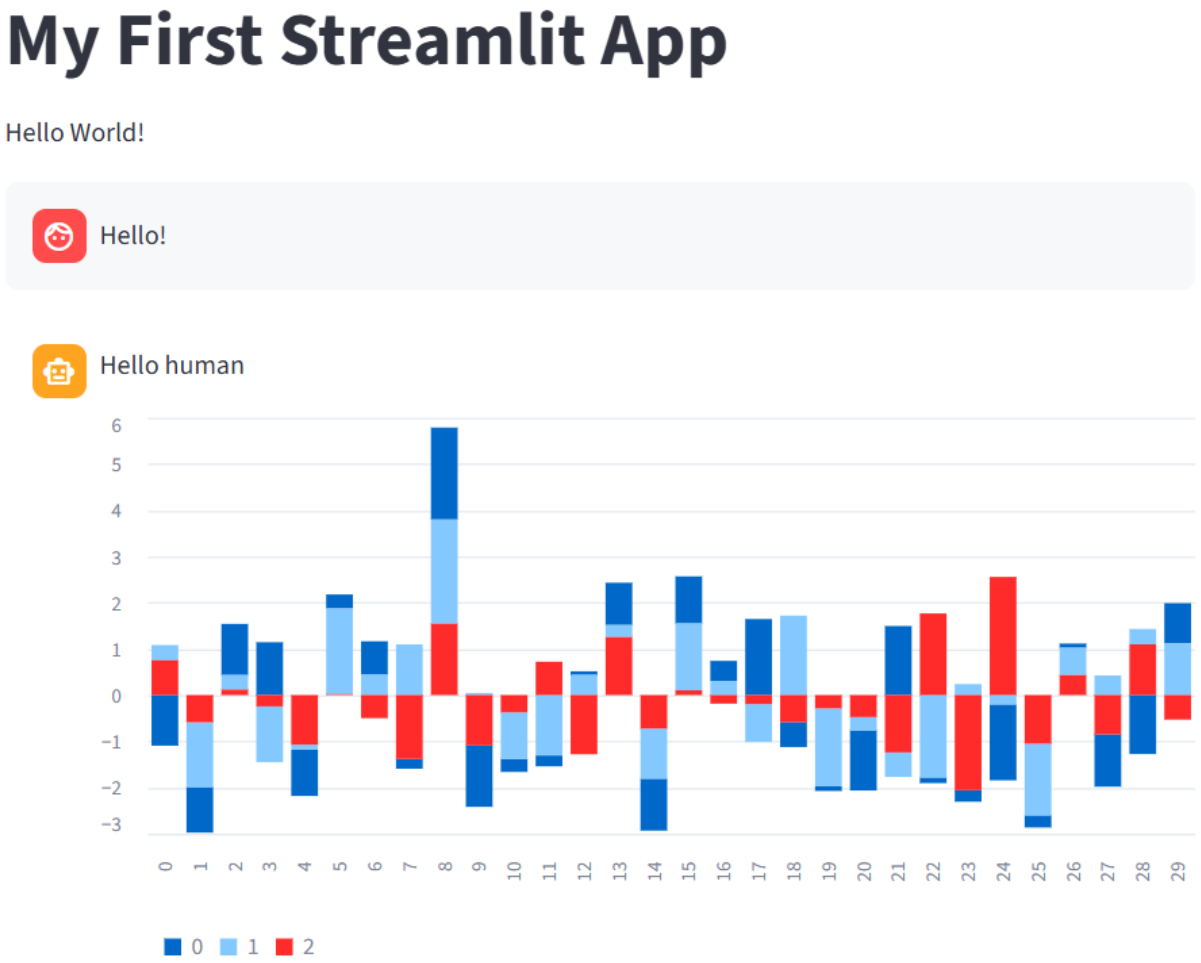
Get User's Input

- st.chat_message

```
with st.chat_message("user"):
    st.write("Hello!")

with st.chat_message("assistant"):
    st.write("Hello human")
    st.bar_chart(np.random.randn(30, 3))
```

https://docs.streamlit.io/develop/tutorials/chat-and-llm-apps/build-conversational-apps



https://docs.streamlit.io/develop/api-reference/charts

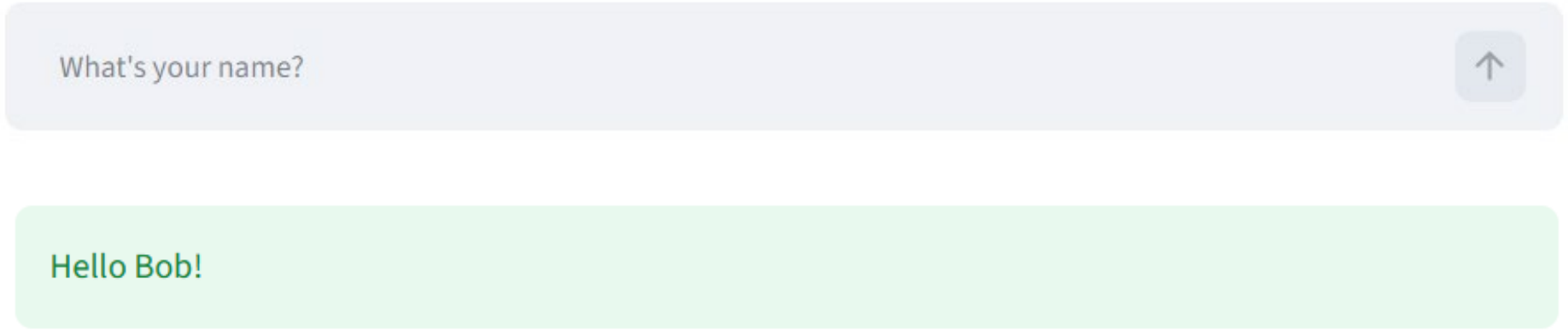
Get User's Input

- st.chat_input

```
name = st.chat_input("What's your name?")

if name:
    st.success(f"Hello {name}!")
```

https://docs.streamlit.io/develop/tutorials/chat-and-llm-apps/build-conversational-apps



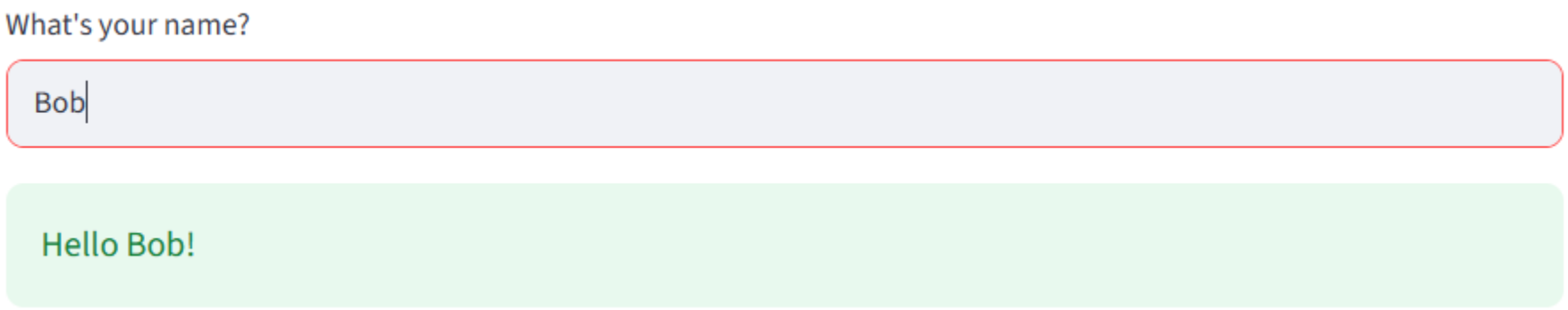
UNC GREENSBORO

Get User's Input

- st.text_input

```
name_text = st.text_input("What's your name?")

if name:
    st.success(f"Hello {name_text}!")
```



UNC GREENSBORO

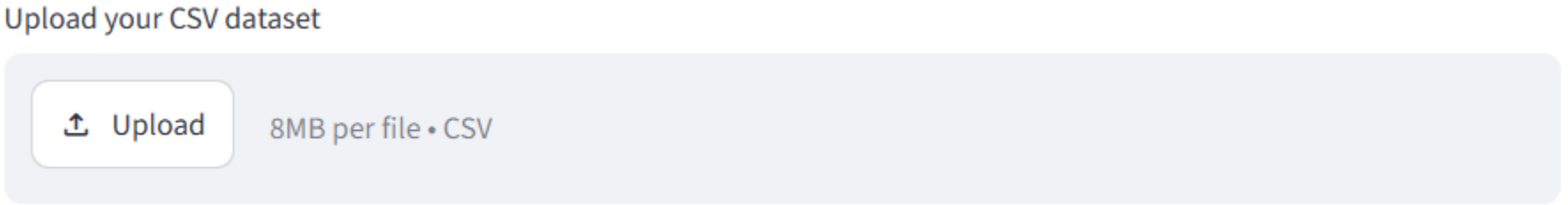
Get User's Input

- Upload files

```python
uploaded_file = st.file_uploader(
    "Upload your CSV dataset",
    max_upload_size=8,
    accept_multiple_files=False,
    type="csv"
)

if uploaded_file is not None:
    df = pd.read_csv(uploaded_file)
```

https://docs.streamlit.io/develop/api-reference/widgets/st.file_uploader



UNC GREENSBORO

Get User's Input

- Upload files

```
uploaded_files = st.file_uploader(
    "Upload your CSV datasets",
    accept_multiple_files=True,
    type="csv"
)

for uploaded_file in uploaded_files:
    df = pd.read_csv(uploaded_file)
    st.write(df.head())
```

https://docs.streamlit.io/develop/api-reference/widgets/st.file_uploader

Upload your CSV datasets

nba.csv 32.1KB

titanic.csv 58.9KB

	Name	Team	Number	Position	Age	Height	Weight	College	Sala
0	Avery Bradley	Boston Celtics	0	PG	25	6-2	180	Texas	7730
1	Jae Crowder	Boston Celtics	99	SF	25	6-6	235	Marquette	6796
2	John Holland	Boston Celtics	30	SG	27	6-5	205	Boston University	N
3	R.J. Hunter	Boston Celtics	28	SG	22	6-5	185	Georgia State	1148
4	Jonas Jerebko	Boston Celtics	8	PF	29	6-10	231	None	5000

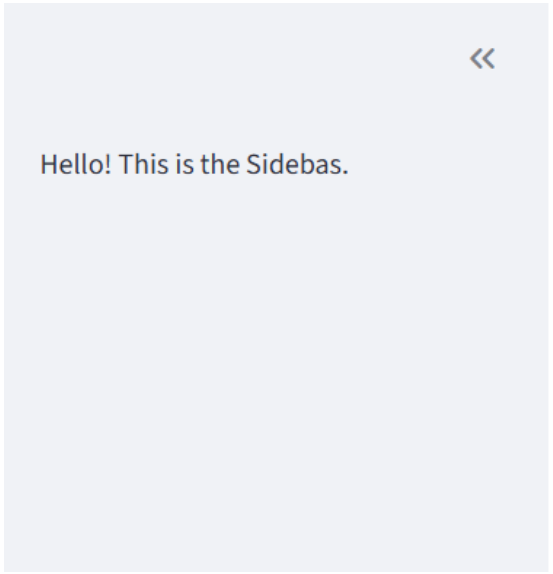
	PassengerId	Survived	Pclass	Name	Sex	Age	S
0	1	0	3	Braund, Mr. Owen Harris	male	22	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	
2	3	1	3	Heikkinen, Miss. Laina	female	26	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	
4	5	0	3	Allen, Mr. William Henry	male	35	

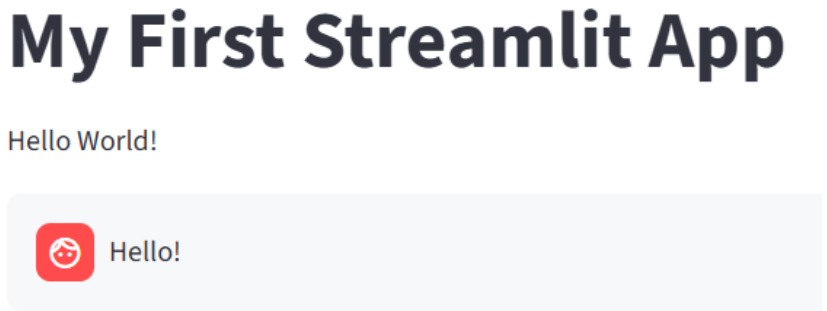
UNC GREENSBORO

Sidebar

```
with st.sidebar:
    st.write('Hello! This is the Sidebas.')
```

https://docs.streamlit.io/develop/api-reference/layout/st.sidebar





UNC GREENSBORO

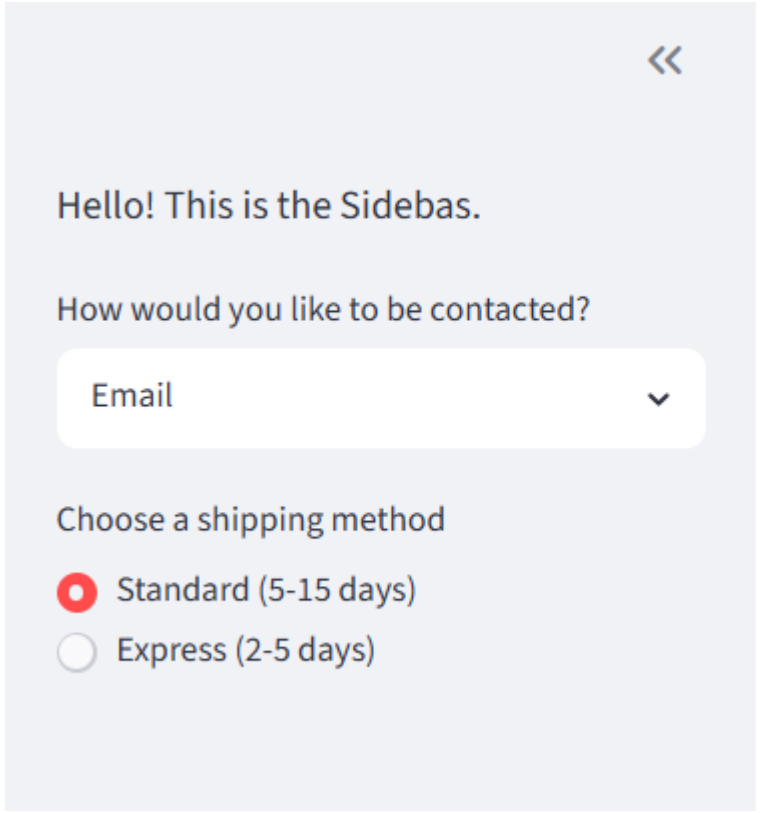
Sidebar

```python
# Using object notation

add_selectbox = st.sidebar.selectbox(
    "How would you like to be contacted?",
    ("Email", "Home phone", "Mobile phone")
)

# Using "with" notation

with st.sidebar:
    add_radio = st.radio(
        "Choose a shipping method",
        ("Standard (5-15 days)", "Express (2-5 days)")
    )
```



UNC GREENSBORO

https://docs.streamlit.io/develop/api-reference/layout/st.sidebar

Button

```python
if st.button("Say hello", type="primary"):
    st.write("Why hello there")
else:
    st.write("Goodbye")

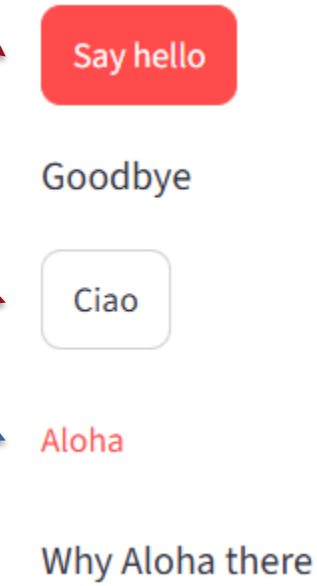
if st.button("Ciao"):
    st.write("Why Ciao there")

if st.button("Aloha", type="tertiary"):
    st.write("Why Aloha there")
```

type *("primary", "secondary", or "tertiary")*
keyword-only

An optional string that specifies the button type. This can be one of the following:

- `"primary"`: The button's background is the app's primary color for additional emphasis.
- `"secondary"` (default): The button's background coordinates with the app's background color for normal emphasis.
- `"tertiary"`: The button is plain text without a border or background for subtlety.



UNC GREENSBORO

https://docs.streamlit.io/develop/api-reference/widgets/st.button