Generative AI

For Data Analysis

GENERAL ASSEMBLY

LEARNING OBJECTIVES

By the end of the lesson, you will be able to...

- ☐ Identify the different capabilities of ChatGPT for Data Analysis
- ☐ Use ChatGPT to perform exploratory data analysis

© 2023 General Assembly

LEARN

How to Use ChatGPT for Data Analysis

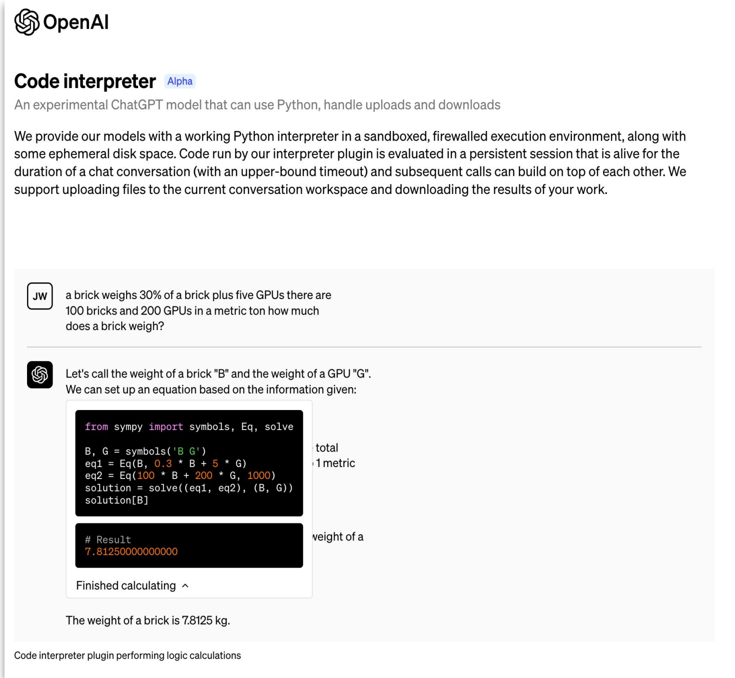
Code Interpreter, now Advanced Data Analysis

In March 2023, OpenAI announced “Code Interpreter”, an experimental ChatGPT model able to run Python code and work on uploaded files.

It was renamed *Advanced Data Analysis* (ADA), and now it lives within the GPT-4 Default chat!

Current supported file types include:

- PDF (.pdf)
- Text (.txt)
- PowerPoint (.ppt)
- Word (.doc)
- Excel (.xlx)
- Comma-separated values (.csv)



OpenAI

Code interpreter Alpha

An experimental ChatGPT model that can use Python, handle uploads and downloads

We provide our models with a working Python interpreter in a sandboxed, firewalled execution environment, along with some ephemeral disk space. Code run by our interpreter plugin is evaluated in a persistent session that is alive for the duration of a chat conversation (with an upper-bound timeout) and subsequent calls can build on top of each other. We support uploading files to the current conversation workspace and downloading the results of your work.

JW: a brick weighs 30% of a brick plus five GPUs there are 100 bricks and 200 GPUs in a metric ton how much does a brick weigh?

Let's call the weight of a brick "B" and the weight of a GPU "G". We can set up an equation based on the information given:

```
from sympy import symbols, Eq, solve

B, G = symbols('B G')
eq1 = Eq(B, 0.3 * B + 5 * G)
eq2 = Eq(100 * B + 200 * G, 1000)
solution = solve((eq1, eq2), (B, G))
solution[B]
```

```
# Result
7.81250000000000
```

total

1 metric

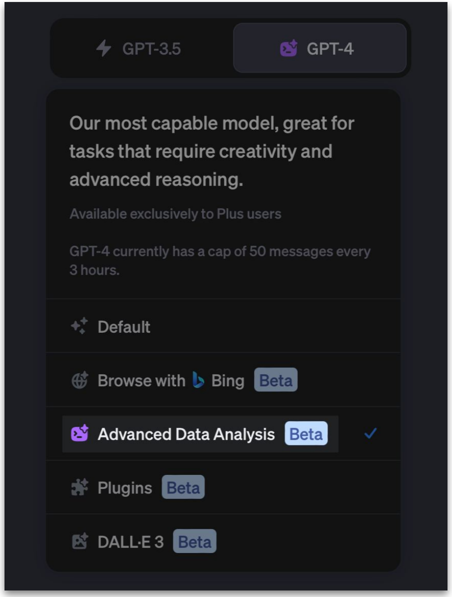
weight of a

Finished calculating

The weight of a brick is 7.8125 kg.

Code interpreter plugin performing logic calculations

From This....

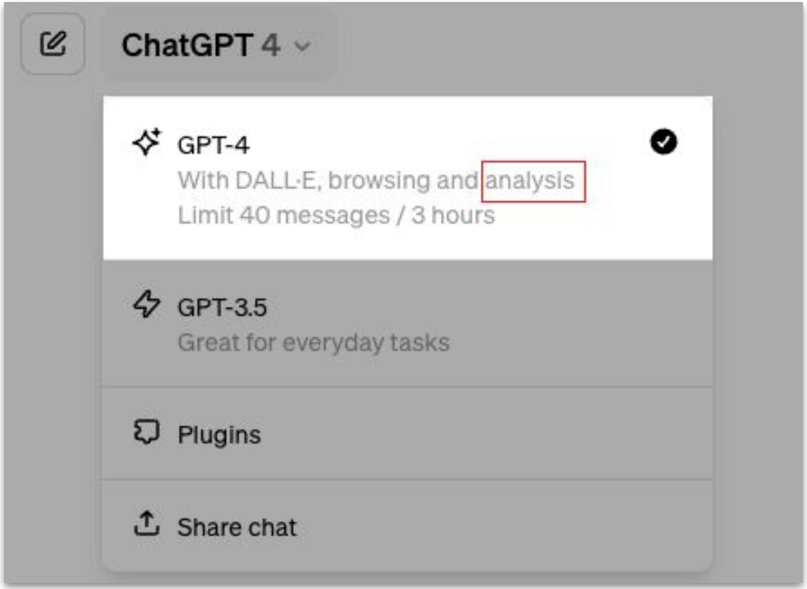


Our most capable model, great for tasks that require creativity and advanced reasoning.

Available exclusively to Plus users

GPT-4 currently has a cap of 50 messages every 3 hours.

To This!



With DALL·E, browsing and analysis

Limit 40 messages / 3 hours

Great for everyday tasks

© 2023 General Assembly

ChatGPT Advanced Data Analysis, Your Coding Intern

- **Comprehensive**: familiar with all public coding documentation until Sep 2021.
- **Speedy Outputs**: produce entire programs in seconds.
- **Adaptive**: can read error messages and self-correct.
- **Over-compliant**: executes tasks as directed, even if ill-advised…
- **Receptive to feedback**: adjusts swiftly to your input.
- **Requires good prompting**: practice makes perfect.

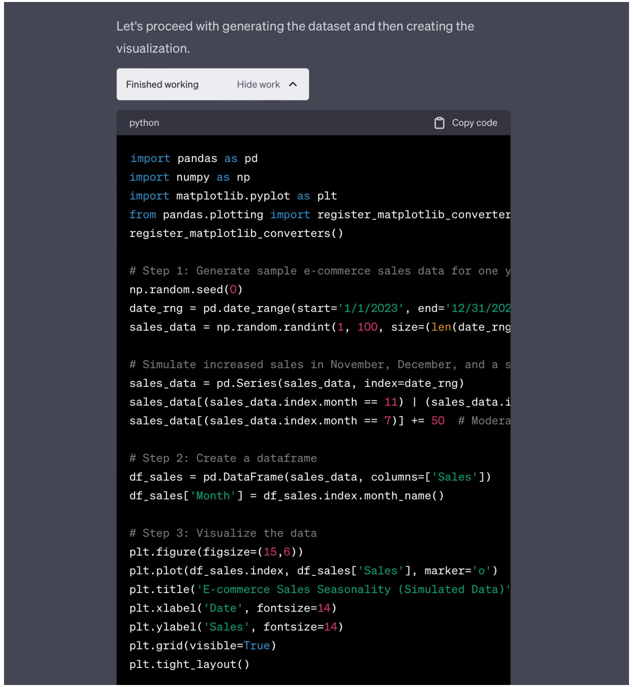
© 2023 General Assembly

GUIDED WALKTHROUGH: Introducing ADA

10 min

Introducing how it works

- ☐ First, let’s visualize without any data
- ☐ Next, let’s take a look at a CSV
- ☐ Let’s write a program that renames invoices. Or at least try…
- ☐ Now, let’s do a find and replace on a PowerPoint



Let's proceed with generating the dataset and then creating the visualization.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pandas.plotting import register_matplotlib_converter
register_matplotlib_converters()

# Step 1: Generate sample e-commerce sales data for one y
np.random.seed(0)
date_rng = pd.date_range(start='1/1/2023', end='12/31/202
sales_data = np.random.randint(1, 100, size=(len(date_rng

# Simulate increased sales in November, December, and a s
sales_data = pd.Series(sales_data, index=date_rng)
sales_data[(sales_data.index.month == 11) | (sales_data.i
sales_data[(sales_data.index.month == 7)] += 50  # Modera

# Step 2: Create a dataframe
df_sales = pd.DataFrame(sales_data, columns=['Sales'])
df_sales['Month'] = df_sales.index.month_name()

# Step 3: Visualize the data
plt.figure(figsize=(15,6))
plt.plot(df_sales.index, df_sales['Sales'], marker='o')
plt.title('E-commerce Sales Seasonality (Simulated Data)'
plt.xlabel('Date', fontsize=14)
plt.ylabel('Sales', fontsize=14)
plt.grid(visible=True)
plt.tight_layout()
```

© 2023 General Assembly

ChatGPT ADA + Python

ChatGPT Analysis is essentially an LLM that can run Python code. So...

What exactly *is* Python?

What is Python?

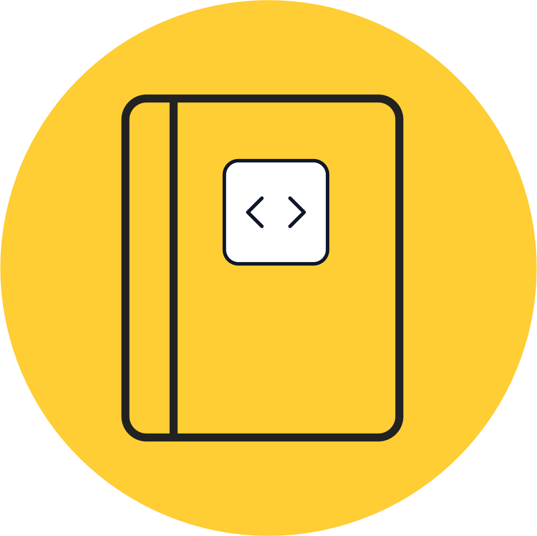
- Python is a high-level, interpreted programming language.
- It's used for web development, data analysis, AI, and more.
- It emphasizes readability and simplicity in its syntax
- It a popular choice for beginners and experienced developers alike.

Source: python.org

© 2023 General Assembly

What are Python Libraries?

- Python libraries are **pre-written code modules**
- They help developers avoid "reinventing the wheel."
- Examples include:
    - **NumPy** for numerical operations
    - **Flask** for web applications
    - **Pandas** for data visualization



© 2023 General Assembly

A Few Use Cases For Python and Python Libraries

Text Analysis	Classification, sentiment analysis
Image Processing	Resize, convert, QR codes, GIFs
Computer Vision	Face detection, OCR
Machine Learning	Build predictive models, segment
Data Visualization	Create charts, maps, diagrams
Data Analysis	Build dashboards, generate insights

© 2023 General Assembly

ADA doesn’t know everything it can do.

© 2023 General Assembly

Libraries ADA Might Be Unaware Of

Text Analysis	TextBlob, Spacy	**Machine Learning**	PyTorch, TensorFlow, scikit-learn
Image Processing	OpenCV, Pillow	**Data Visualization**	seaborn, folium
Computer Vision	pytesseract	**Data Analysis**	Plotly

***Note**: You don't have to know exactly how libraries work or be an expert coder to tell ChatGPT to use a particular library!*

© 2023 General Assembly

PRACTICE

ChatGPT for Data Analysis

ChatGPT for Data Analysis

15 Minutes

Try out Advanced Data Analysis. Feel free to try working with **any of the files** in your **Participant Folder**, including images.

Instructions:

1. First upload the file(s) and ask for some initial help.
2. Remember: ChatGPT is over-compliant BUT receptive to feedback. Refine the prompt/with follow-ups.
3. Share what you've learned - both what you can and can't do - in the meeting chat.

Potential Use Cases:

Try asking for help with one of the following:

- Data analysis with spreadsheets
- Image conversions (including cropping)
- Editing a Python file
- Doing a find and replace on a document
- Building a predictive model!

© 2023 General Assembly

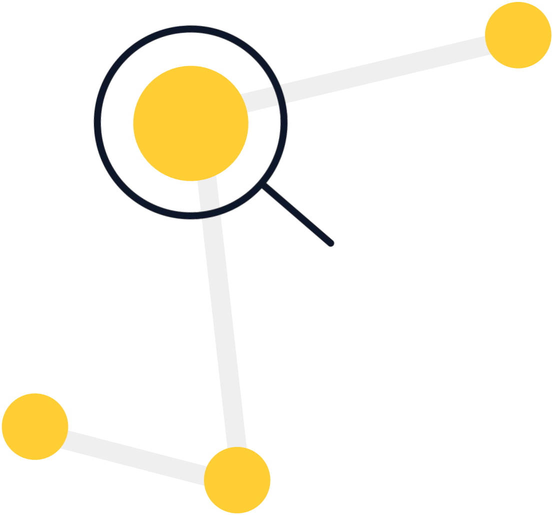
LEARN

Exploratory Data Analysis With Generative AI

What is Exploratory Data Analysis (EDA)?

Exploratory Data Analysis (EDA) is the initial process of analyzing and visualizing data sets to summarize their main characteristics, identify patterns, and inform subsequent analysis.

EDA is essentially just “getting to know” a data set.



© 2023 General Assembly

PRACTICE

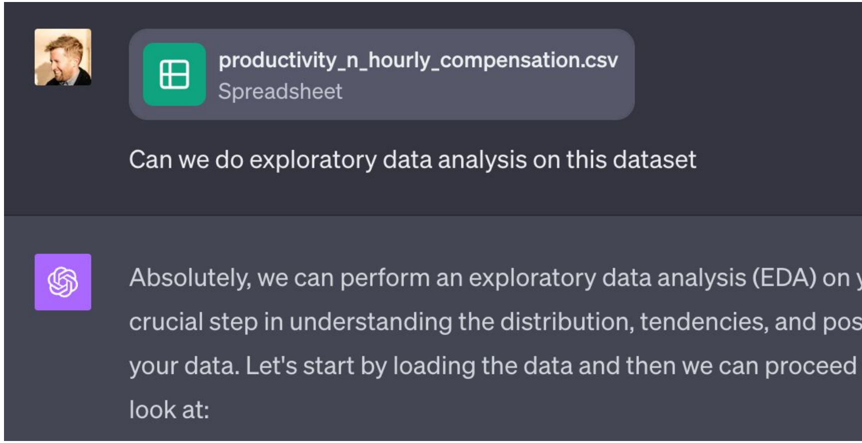
Guided Walkthrough: EDA with Gen AI

GUIDED WALKTHROUGH: EDA with Gen AI

10 min

Productivity & Hourly Compensation Data

- Ask: “*can we do exploratory data analysis on this dataset*”
- Prompt for further exploration.
- Request the top five insights
- Ask for suggestions on what actions policymakers could take based on these insights.



Can we do exploratory data analysis on this dataset

Absolutely, we can perform an exploratory data analysis (EDA) on
crucial step in understanding the distribution, tendencies, and pos
your data. Let's start by loading the data and then we can proceed
look at:

© 2023 General Assembly

LEARN

A Cautionary Tale: ChatGPT for Data Analysis

Hallucinations + Data Analysis

It is **extremely** important to remember that ChatGPT, as well as any other LLM, can have **hallucinations** at any point.

This has some **serious** implications when using ChatGPT to work with data.

Hallucinations + Data Analysis

1. ChatGPT may generate **fake data** that it needs to complete a user request.

2. ChatGPT will often ***hint*** that it is creating fake data, but it is easy to miss.

3. For this reason, it is **critical** that you document the use of ChatGPT when working with data, especially if you are not a data professional yourself!

© 2023 General Assembly

Think About It: Data Analysis 🤔

What are some of the worst things that could happen when using ChatGPT for Data Analysis?

Type your response in the chat.

© 2023 General Assembly

REFLECT

On ChatGPT ADA

Reflection: On ChatGPT ADA

Taking what you learned and what you practiced, answer the following reflection questions.

1. To what extent do you agree that the ability to run code represents “GPT 4.5”?
2. Which capabilities are you keen to try out for yourself, with your own data?
3. How do you plan to apply what you’ve learned in your work?

Tips for Reflection:

Think carefully through the questions, provide specific details and examples, and try to break down your thinking.

© 2023 General Assembly

GENERAL ASSEMBLY