



Simplifying Data Analysis With GitHub Codespaces, Jupyter Notebooks & Open AI

Nitya Narasimhan, PhD
Senior Cloud Advocate, AI
Microsoft



#PyDataNYC | Nov 2023

<https://aka.ms/pydata-workshop-2023>



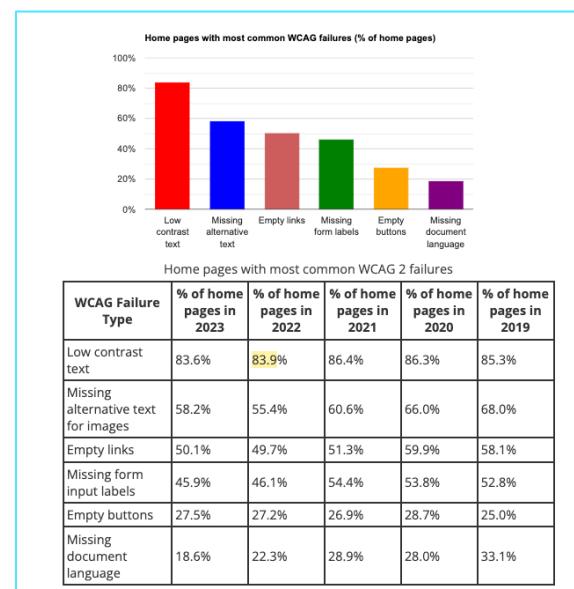
Motivation

“I’m an app developer not a Python expert. How can I learn enough to use it to tackle specific goals in my apps?”

Examples:

- The [USA Computing Olympiad](#) allows Python but lacks parity for resources vs. C/C++, Java. *Can I create self-driven learning resources for Python competitors like my 15yo?.*
- [Accessibility Testing](#) generates rich data (1K+ elements per page, 100 rules, 3-5 checks per rule = ~250-500K data points. *Can I give developers insight at authoring time, so they can fix issues earlier in development cycle?*
- [Information APIs](#) can generate pages of JSON data with dense information. *Can I visualize information in ways that help users get actionable insights?*

The screenshot shows the USACO Guide's Problems (Beta) section. At the top, there are search bars for "Search" and "Algolia". Below that, there are filters for "Easy", "Modules", "Bronze", "Solved", and "Starred". The main area displays a grid of problems: "Where Am I?", "Sleepy Cow Herding", "Shell Game", and "The Bucket List". Each problem card includes its name, difficulty level (Easy or Set), and a "View Solution" link.



```
uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
383 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/c--kA8Nj0Kg-
384 ", "type_of": "article",
385 "id": 1644652,
386 "title": "Want to buildintelligentApps with AKS & Serverless on Azure? Join us to #Learn
387 #DevOps with @mashavayn on the upcoming #LearnToBuild
388 "readable_publish_date": "Oct 24",
389 "slug": "want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
390 "url": "https://dev.to/azure/want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
391 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
392 "social_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-bckB2oX/c--Image
393 uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
394 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
395 "type_of": "article",
396 "id": 1644652,
397 "title": "Want to buildintelligentApps with AKS & Serverless on Azure? Join us to #Learn
398 "#DevOps with @mashavayn on the upcoming #LearnToBuild
399 "readable_publish_date": "Oct 24",
400 "slug": "want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
401 "url": "https://dev.to/azure/want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
402 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
403 uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
404 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
405 "type_of": "article",
406 "id": 1644652,
407 "title": "Want to buildintelligentApps with AKS & Serverless on Azure? Join us to #Learn
408 "#DevOps with @mashavayn on the upcoming #LearnToBuild
409 "readable_publish_date": "Oct 24",
410 "slug": "want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
411 "url": "https://dev.to/azure/want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
412 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
413 uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
414 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
415 "type_of": "article",
416 "id": 1644652,
417 "title": "Want to buildintelligentApps with AKS & Serverless on Azure? Join us to #Learn
418 "#DevOps with @mashavayn on the upcoming #LearnToBuild
419 "readable_publish_date": "Oct 24",
420 "slug": "want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
421 "url": "https://dev.to/azure/want-to-buildintelligentapps-with-aks-serverless-on-azure-join-us-to-buildintelligentapps-with-aks-serverless-on-azure-jo-
422 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-u2j4XJ/c--Image
423 uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
424 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/s-WMdo21B/c--Image
425 uploads.s3.amazonaws.com/uploads/organization/profile_image/512/64ce0b82/73bd-4ca0-8359-2c21513a0
426 "profile_image_90": "https://res.cloudinary.com/practicaldev/image/fetch/s-1C-NuP/c--Image
427 "type_of": "article",
428 "id": 1637344,
429 "title": "BuildintelligentApps Just AskTheExpert!",
430 "description": "Join experts from @ZContainerApps #AzureFunctions, Azure AppService and
431 "readable_publish_date": "Oct 19",
432 "slug": "buildintelligentapps-just-asktheexpert-13d",
433 "url": "https://res.cloudinary.com/practicaldev/image/fetch/s-1C-NuP/c--Image
434 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-1C-NuP/c--Image
435 "type_of": "article",
436 "id": 1637344,
437 "title": "BuildintelligentApps Just AskTheExpert!",
438 "description": "Join experts from @ZContainerApps #AzureFunctions, Azure AppService and
439 "readable_publish_date": "Oct 19",
440 "slug": "buildintelligentapps-just-asktheexpert-13d",
441 "url": "https://dev.to/azure/o-buildintelligentapps-just-asktheexpert-13d",
442 "cover_image": "https://res.cloudinary.com/practicaldev/image/fetch/s-1C-NuP/c--Image
443 "type_of": "article",
444 "id": 1637344,
```

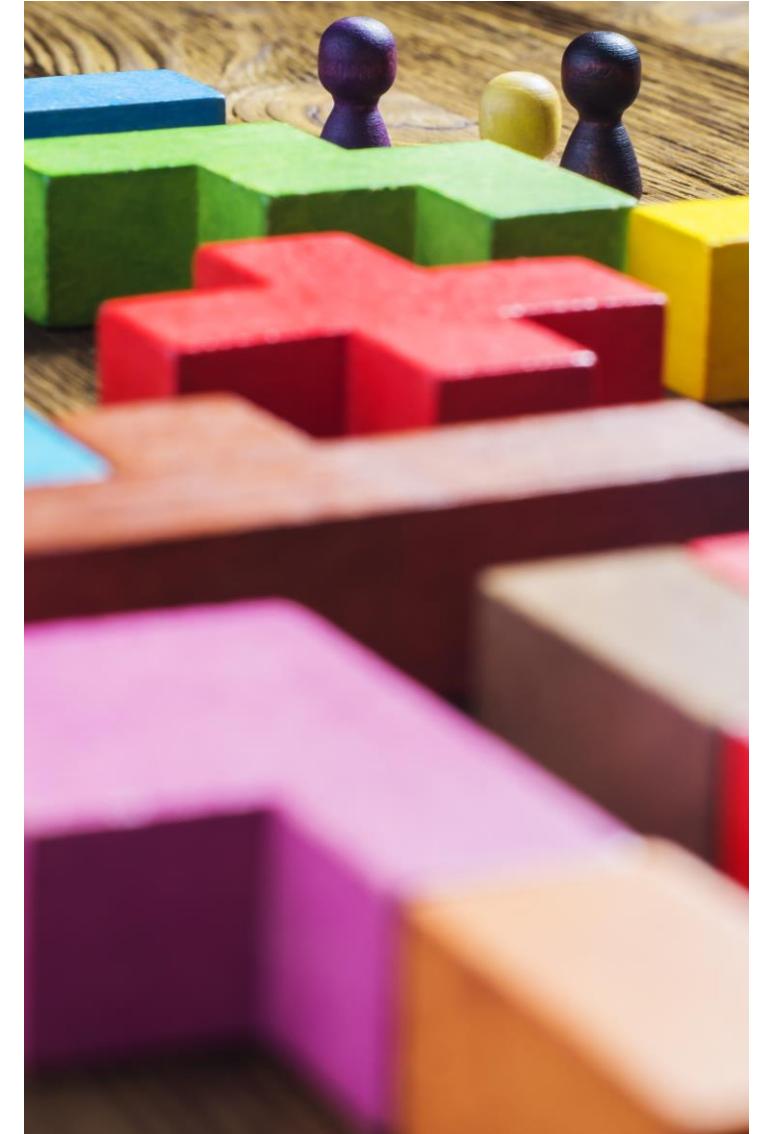
Objectives

Focused Learning – I want to prioritize things that help me get closer to my goal.

"Don't boil the ocean. Tell me how I can optimize my time,"

Transferable Learning – I want to make it easy for others to reproduce or extend my work.

"What tools or environments should I use that make this possible?"



Challenges

Developer Environment – setup can be hard to navigate for beginners.

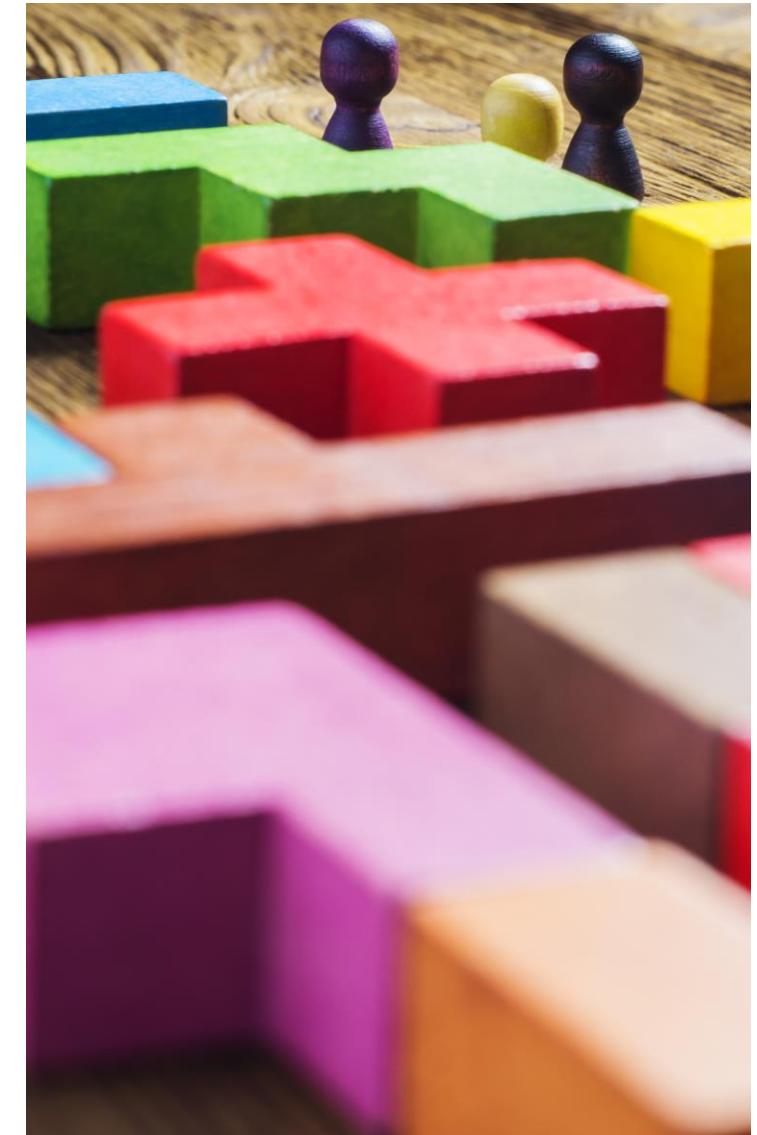
"It works on my machine. Must be some dependency."

Learning Process – can disrupt state of flow when context-switching across tools.

"I went to Google it – and lost time going down a rabbit hole."

Knowledge Gaps – can hamper the intuition that only comes with experience.

"I don't know what I don't know – what if I missed something?"



What I want to cover today

Transferable Learning
with [Jupyter Notebooks](#)

Consistent Dev Environment
with [GitHub Codespaces](#)

Focused Learning (in-context)
with [GitHub Copilot](#)

Building AI-assisted intuition
using [Microsoft LIDA](#)





<https://aka.ms/pydata-for-jsdevs>

What You Need

- GitHub Account
- Github Codespaces (within free quota)
- Kaggle Dataset (or bring your own)
- GitHub Copilot (free trial available)*
- Open AI API Key (paid subscription)*

* require paid accounts – optional exercises if you don't have one.



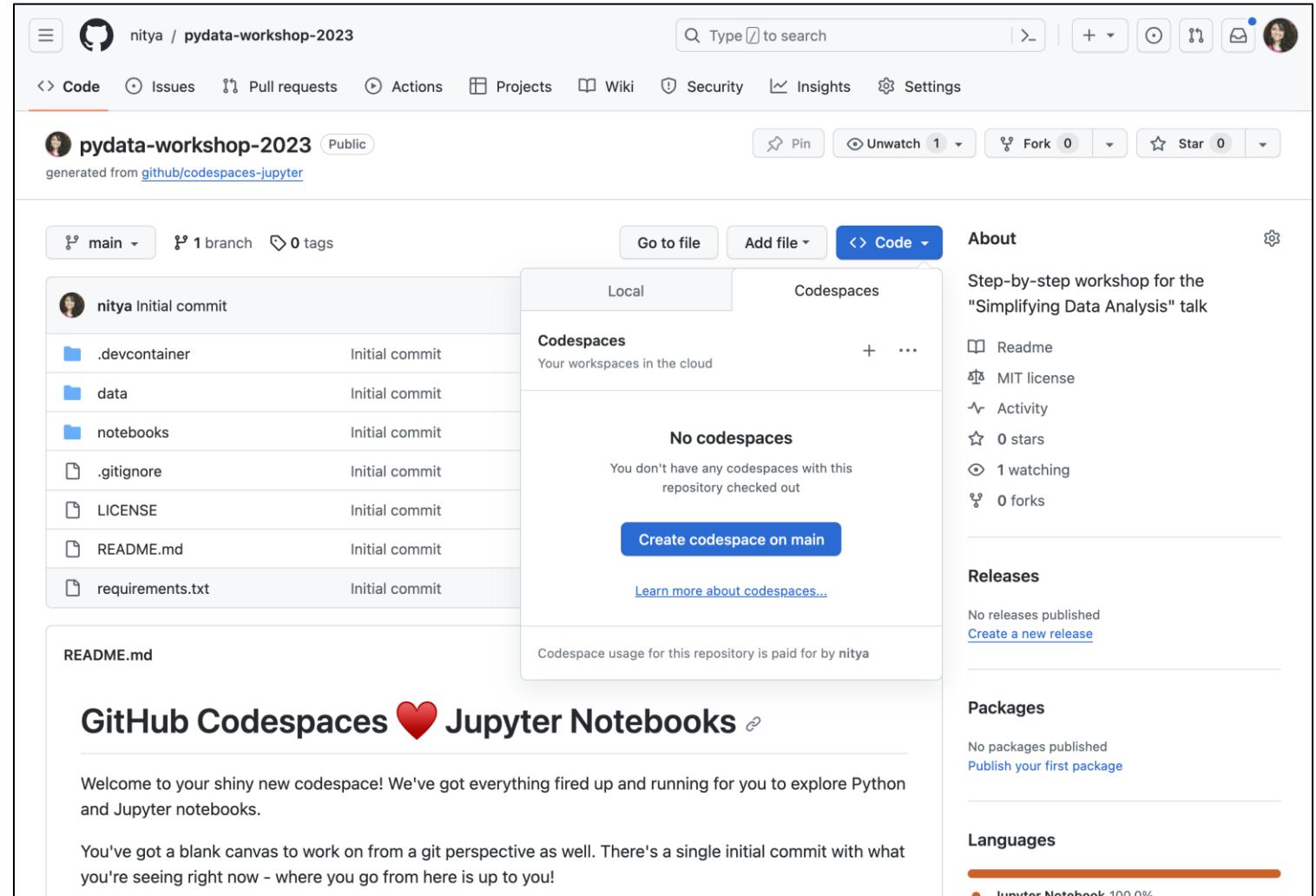
Part 1:

Get A Reproducible Development Environment with GitHub Codespaces + Jupyter Notebooks

Exercise 1: Fork Repo, Launch Codespaces

Based on the official GitHub Codespaces for Jupyter template.

Fork the repo to your own profile, then launch a codespace from the menu

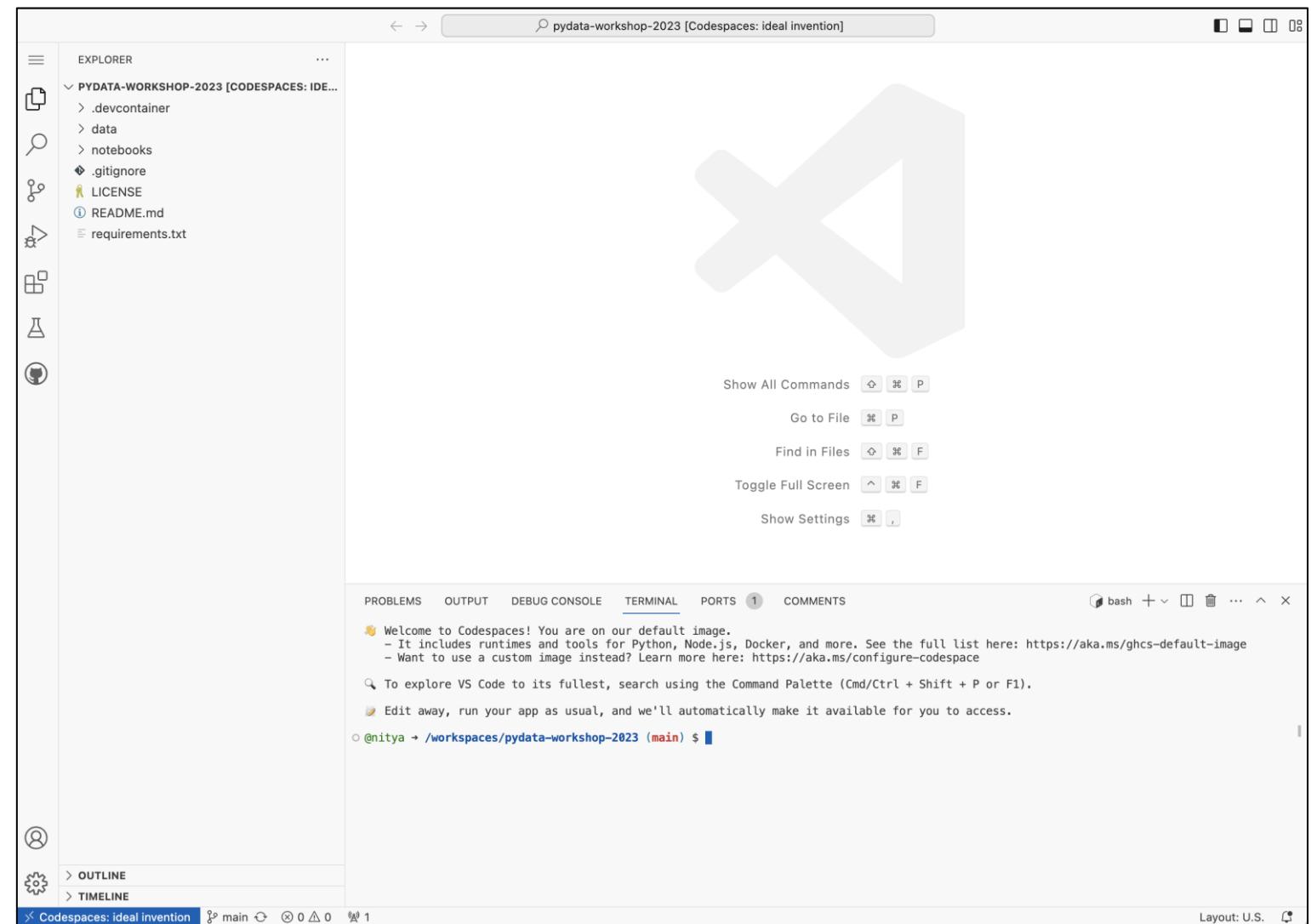


Exercise 1: Fork Repo, Launch Codespaces

Based on the official GitHub [Codespaces for Jupyter](#) template.

Fork the repo to your own profile, then launch a codespace from the menu

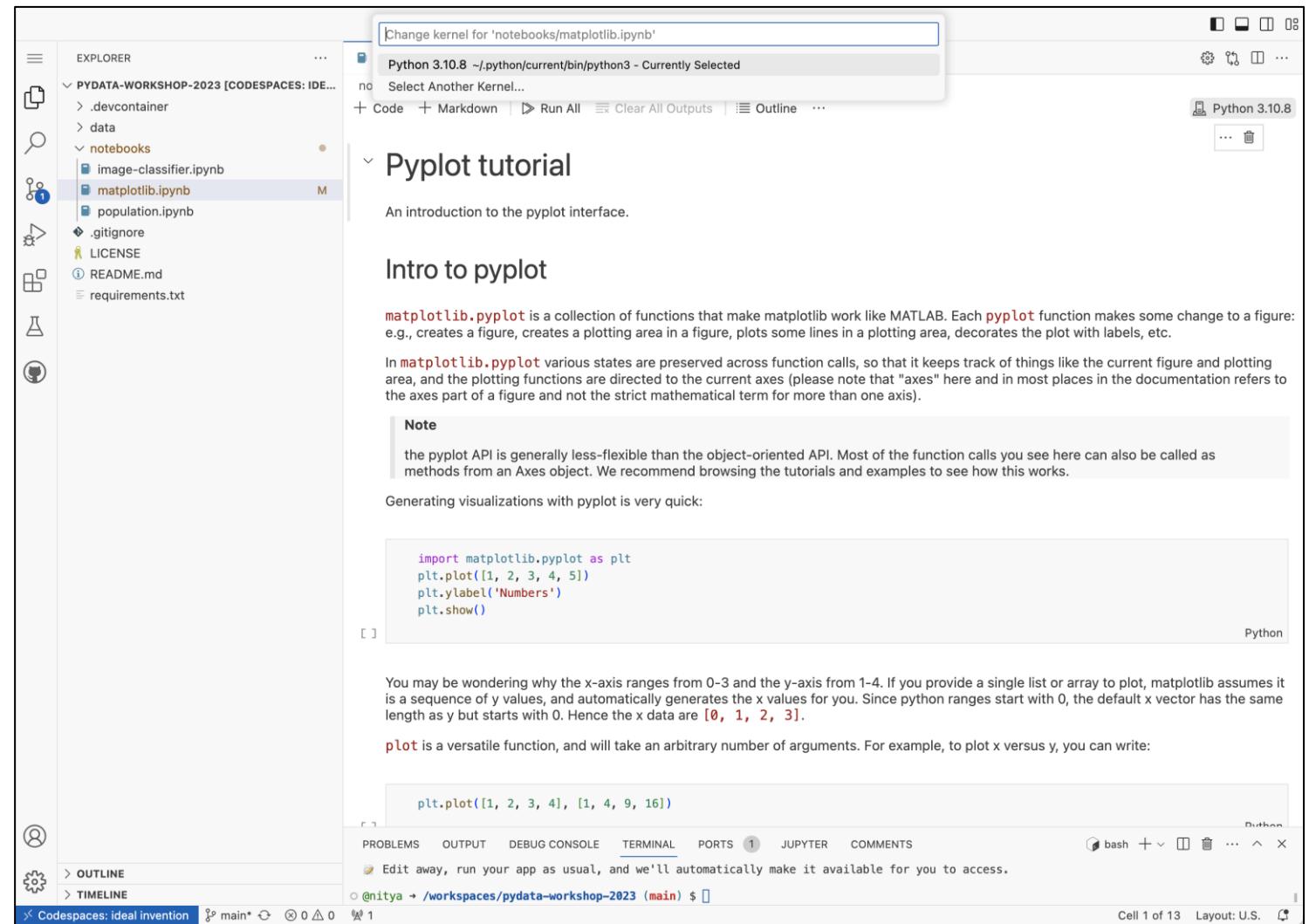
You should see a Visual Studio Code IDE + a dev environment ready to use!



Exercise 2: Let's Run the Default Notebook!

Open the *notebooks*/ folder and look for the *matplotlib.ipynb* file.

If not already set, 'Select Kernel' pick Python 3.10.8 target, 'Clear All Outputs'



Exercise 2: Let's Run the Default Notebook!



Open the *notebooks*/ folder and look for the *matplotlib.ipynb* file.

If not already set, 'Select Kernel' pick Python 3.10.8 target, 'Clear All Outputs'

"Run All". You just setup a Python env with Jupyter runtime with a few clicks.

The screenshot shows a Jupyter Notebook interface within a CodeSpace environment. The left sidebar displays a file tree for a workspace named 'PYDATA-WORKSHOP-2023'. The 'notebooks' folder contains three files: 'image-classifier.ipynb', 'matplotlib.ipynb' (which is currently selected), and 'population.ipynb'. Other files in the root include '.devcontainer', 'data', '.gitignore', 'LICENSE', 'README.md', and 'requirements.txt'. The main area shows a Jupyter notebook cell titled 'matplotlib.ipynb' with the following code:

```
import numpy as np

# evenly sampled time at 200ms intervals
t = np.arange(0., 5., 0.2)

# red dashes, blue squares and green triangles
plt.plot(t, t, 'r--', t, t**2, 'bs', t, t**3, 'g^')
plt.show()
```

The output of this code is a line plot with three data series: a red dashed line, blue square markers connected by a solid line, and green triangle markers connected by a solid line. The x-axis ranges from 0 to 5, and the y-axis ranges from 0 to 100. The plot shows a cubic relationship where the red dashed line is a straight line, the blue squares follow a parabolic curve, and the green triangles follow a cubic curve.

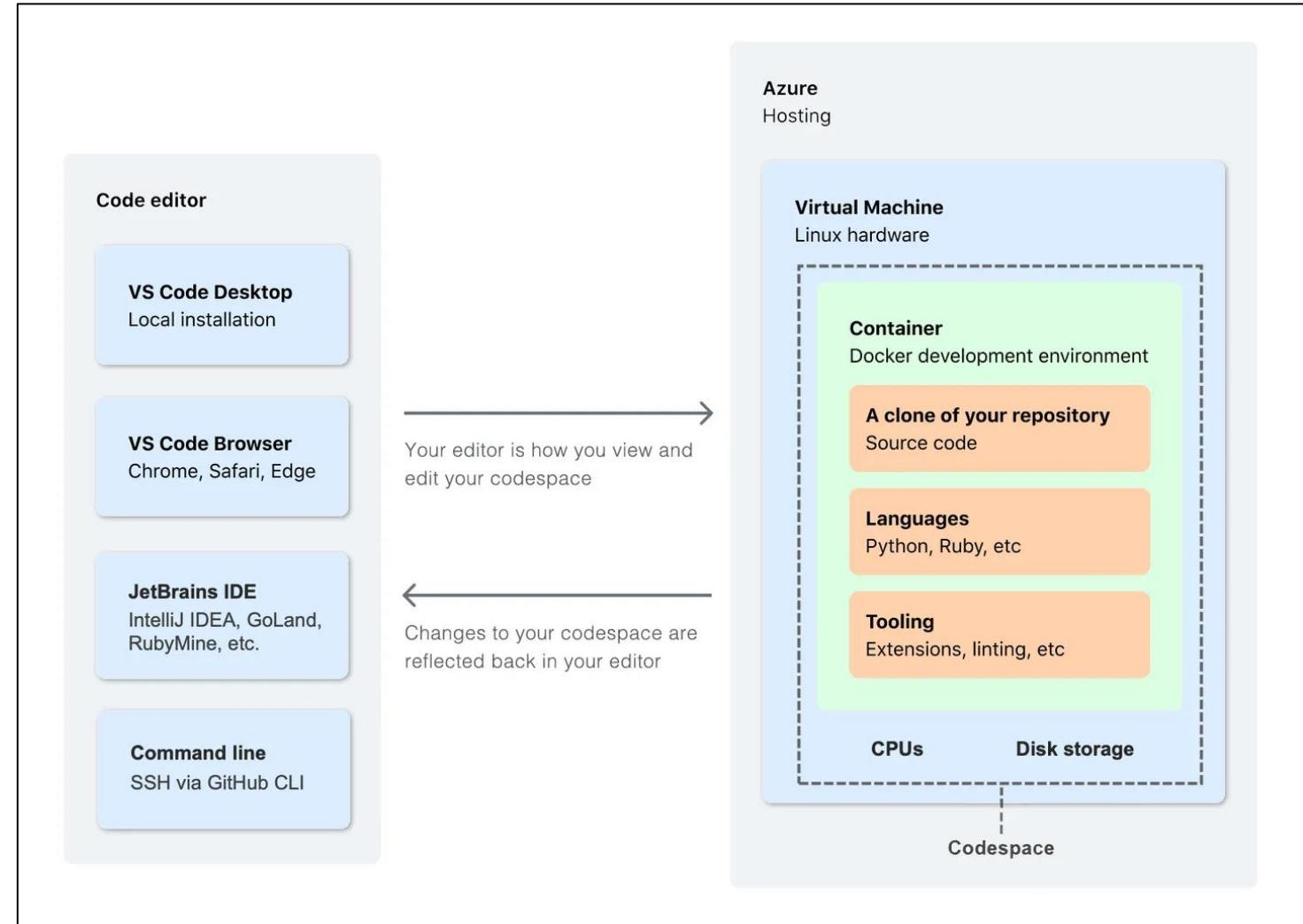
At the bottom of the interface, there are tabs for PROBLEMS, OUTPUT, DEBUG CONSOLE, TERMINAL, PORTS (with a count of 2), JUPYTER, and COMMENTS. The JUPYTER tab is active. Below the tabs, it says '@nitya → /workspaces/pydata-workshop-2023 (main) \$'. The status bar at the bottom right indicates 'Cell 1 of 13 Layout: U.S.'.

Under The Hood: What's GitHub Codespaces?

A GitHub Codespace is a “[Development Container](#)” that runs in a dedicated VM in the Azure Cloud.

You can set a different default editor – but VS Code gets us Extensions!

| Editor preference | |
|----------------------------------|--|
| <input type="radio"/> | Visual Studio Code Connect to the cloud from your local desktop client. Requires Visual Studio Code with the GitHub Codespaces extension. |
| <input checked="" type="radio"/> | Visual Studio Code for the Web Edit and preview changes straight from the browser. |
| <input type="radio"/> | JetBrains Gateway Connect to the cloud from your local desktop client. Requires the JetBrains Gateway with the GitHub Codespaces plugin, and a JetBrains license. |
| <input type="radio"/> | JupyterLab Edit and run notebooks from the browser with JupyterLab. |

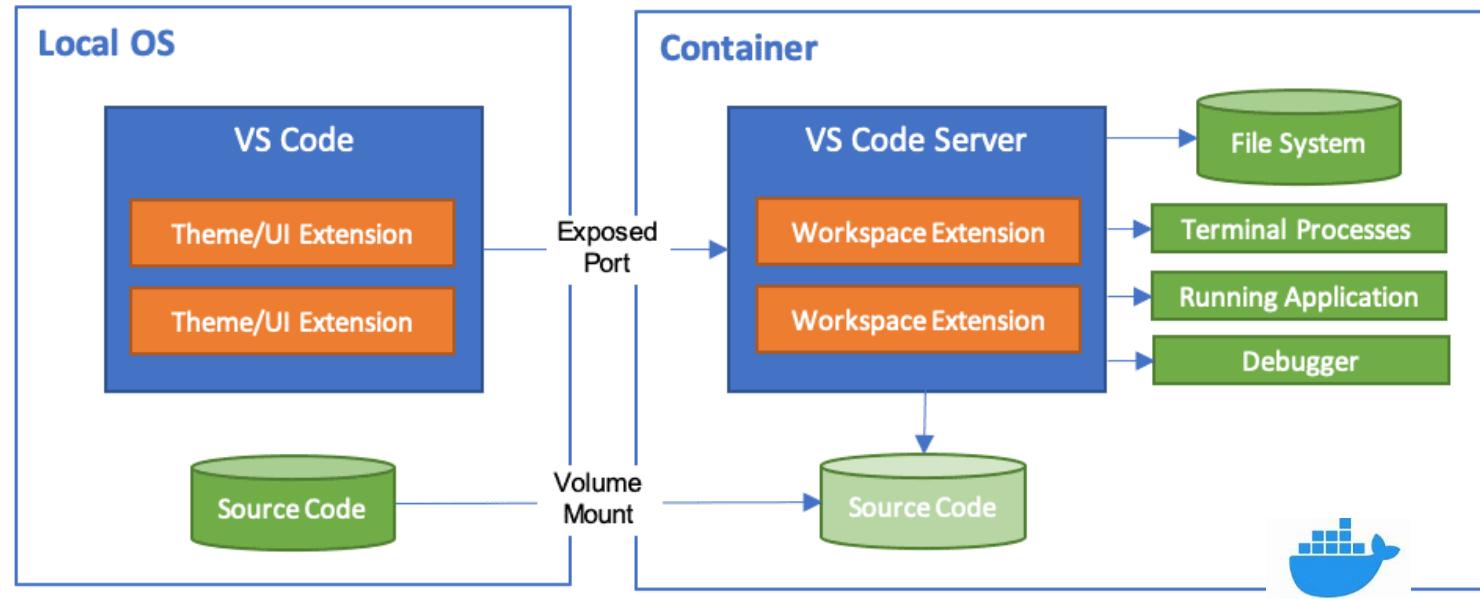


Under The Hood: What's a “Development Container”?

“Configuration as Code”
using *devcontainer.json* –
think version controlled!

Lifecycle hooks – update
env from *requirements.txt*

```
devcontainer.json x
1
2  "image": "mcr.microsoft.com/devcontainers/universal:2",
3  "hostRequirements": {
4    "cpus": 4
5  },
6  "waitFor": "onCreateCommand",
7  "updateContentCommand": "python3 -m pip install -r requirements.txt",
8  "postCreateCommand": "",
9  "customizations": {
10    "codeSpaces": {
11      "openFiles": []
12    },
13    "vscode": {
14      "extensions": [
15        "ms-toolsai.jupyter",
16        "ms-python.python"
17      ]
18    }
19  }
20
21 }
```



Your Visual Studio
Code editor runs in
the host machine
outside the container

Your code (filesystem)
is mounted by the
container – and stays
in sync with changes

A dev container is a Docker
container with a predefined
image that has all your dev
environment dependencies

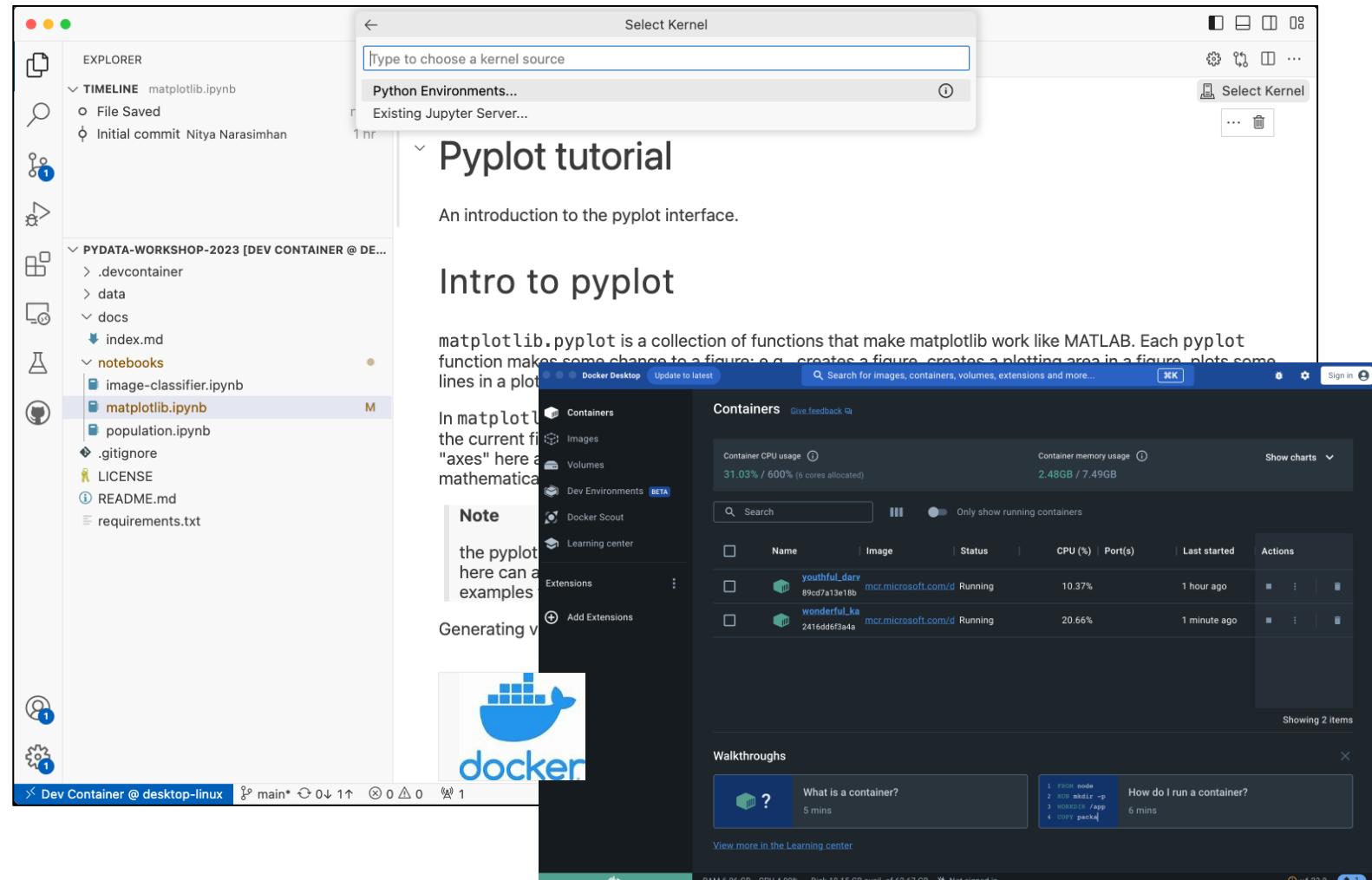
Everyone gets the **same dev environment** instantly

Under The Hood: Run it locally - with Docker Desktop!

Just “Open in Remote Container” – status shows Dev Container (local) or Codespaces (cloud) in blue

No code changes needed to move between the two.
It just works!

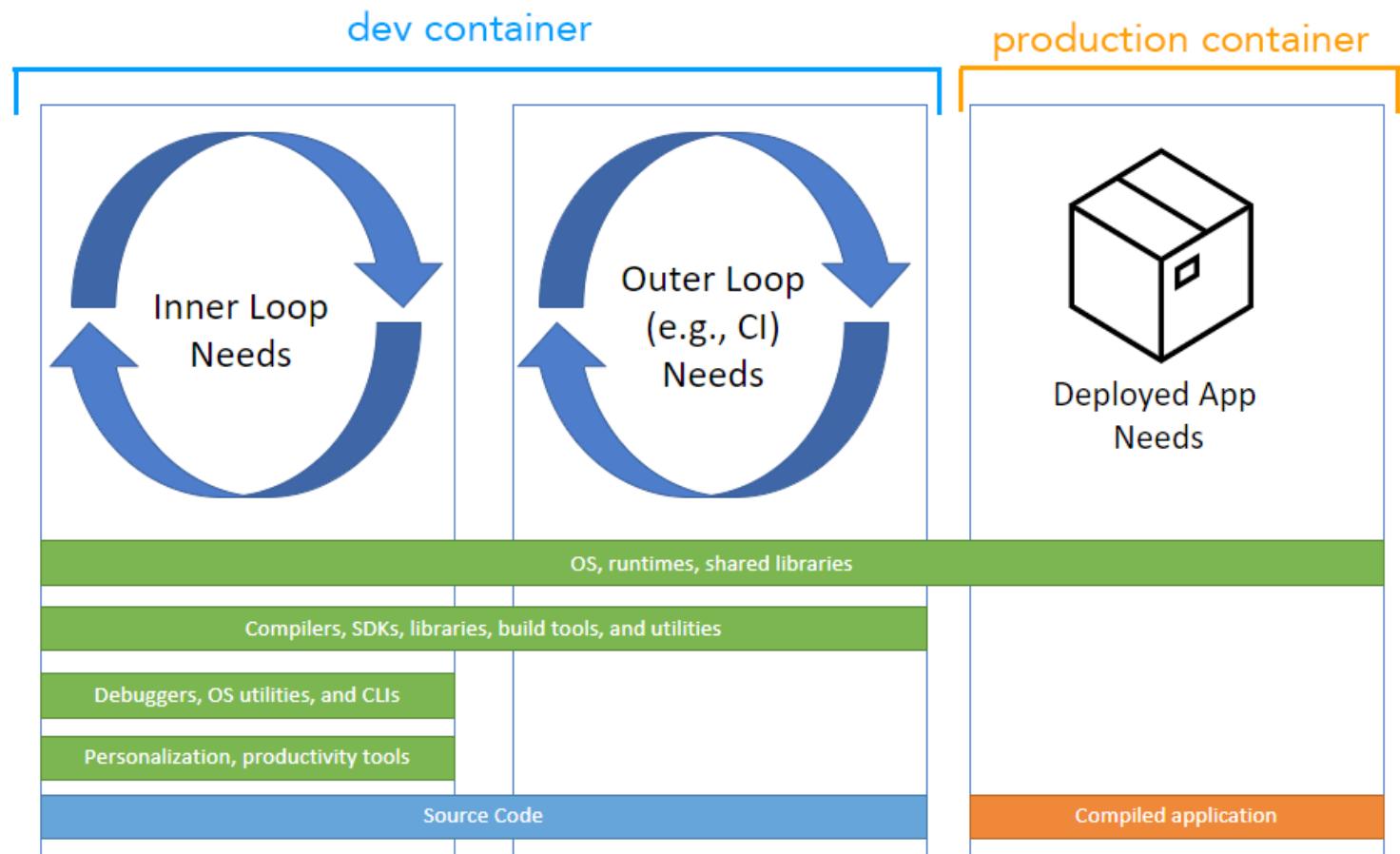
Local usage saves your quota – but has some limitations (e.g, “secrets”)



Under The Hood: Production vs. Dev Containers

As an app developer, I get a **consistent runtime** for the entire cycle from dev to staging to production

And I can debug & test in development (with richer tools) & be confident that deployed experiences will function as expected





Part 2: Get Focused Learning With GitHub Copilot

Exercise 3: Install GitHub Copilot Chat Extension

Add it on demand – or save to *devcontainer.json* for reusability by all.

I've chosen to save this – users will now get this in their env automatically.

The screenshot shows the VS Code interface with the Extensions Marketplace open. The GitHub Copilot Chat extension is selected and highlighted. A context menu is open over the extension, with the option "Add to devcontainer.json" being selected. To the right, the "devcontainer.json" file is displayed in the editor, showing the configuration for the dev container. The "vscode" section includes the "extensions" array with entries for "ms-toolsai.jupyter" and "ms-python.python".

```
1  "image": "mcr.microsoft.com/devcontainers/universal:2",
2  "hostRequirements": {
3      "cpus": 4
4  },
5  "waitFor": "onCreateCommand",
6  "updateContentCommand": "python3 -m pip install -r requirements.txt",
7  "postCreateCommand": "",
8  "customizations": {
9      "codeSpaces": {
10          "openFiles": []
11      },
12      "vscode": {
13          "extensions": [
14              "ms-toolsai.jupyter",
15              "ms-python.python"
16          ]
17      }
18  },
19  "files": [
20      ".vscode/extensions/ms-toolsai.jupyter@0.1.14"
21  ]
22 }
```

Below the editor, the terminal window shows a welcome message for Codespaces and provides instructions for using the environment.

```
Welcome to Codespaces! You are on our default image.
- It includes runtimes and tools for Python, Node.js, Docker, and more. See the full list here: https://aka.ms/ghcs-default-image
- Want to use a custom image instead? Learn more here: https://aka.ms/configure-codespace

To explore VS Code to its fullest, search using the Command Palette (Cmd/Ctrl + Shift + P or F1).
Edit away, run your app as usual, and we'll automatically make it available for you to access.

When using "conda" from outside of GitHub Codespaces, note the Anaconda repository contains restrictions on commercial use that may impact certain organizations. See https://aka.ms/ghcs-conda
```

Exercise 3: Save config, then open chat window



You get the 'chat' icon in your sidebar, with a slide-out chat window ...

And your config file is updated (rebuild now!)

```
(1) devcontainer.json M
1  {
2    "image": "mcr.microsoft.com/devcontainers/universal:2",
3    "hostRequirements": {
4      "cpus": 4
5    },
6    "waitFor": "onCreateCommand",
7    "updateContentCommand": "python3 -m pip install -r requirements.txt",
8    "postCreateCommand": "",
9    "customizations": {
10      "codespaces": {
11        "openFiles": []
12      },
13      "vscode": {
14        "extensions": [
15          "ms-toolsai.jupyter",
16          "ms-python.python",
17          "GitHub.copilot-chat"
18        ]
19      }
20    }
21  }
```

The screenshot shows the VS Code interface with the GitHub Copilot Chat extension installed. The sidebar features a 'CHAT: GITHUB COPILOT' icon. A slide-out chat window from 'GitHub Copilot' is visible, displaying a welcome message and information about its AI-powered capabilities. The main workspace shows the 'devcontainer.json' configuration file with the 'GitHub.copilot-chat' extension listed under 'customizations'. The bottom of the screen shows the terminal output, which includes instructions for using Codespaces and notes about Anaconda restrictions.

CHAT: GITHUB COPILOT

GitHub Copilot

Welcome @nitya, I'm your Copilot and I'm here to help you get things done faster.

I'm powered by AI, so surprises and mistakes are possible. Make sure to verify any generated code or suggestions, and [share feedback](#) so that we can learn and improve. Check out the [Copilot documentation](#) to learn more.

devcontainer.json

Extension: GitHub Copilot Chat

GitHub Copilot Chat v0.10.0 Preview

GitHub [github.com](#) | ⚡ 3,968,058 | ★★★★☆ (74)

AI chat features powered by Copilot

Disable ⚡ Uninstall ⚡ ⚡

Extension is enabled on 'Dev Container @ desktop-linux'

DETAILS FEATURE CONTRIBUTIONS EXTENSION PACK RUNTIME STATUS

GitHub Copilot Chat

GitHub Copilot Chat is a companion extension to [GitHub Copilot](#) that houses preview Chat features:

- **Chat view:** Ask Copilot for help with any task or question in the Chat view, bringing in code from your current files.
- **Inline Chat:** Apply Copilot's AI suggestions directly to your code, staying

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS 3

Welcome to Codespaces! You are on our default image.
- It includes runtimes and tools for Python, Node.js, Docker, and more. See the full list here: <https://aka.ms/ghcs-default-image>
- Want to use a custom image instead? Learn more here: <https://aka.ms/configure-codespace>

To explore VS Code to its fullest, search using the Command Palette (Cmd/Ctrl + Shift + P or F1).

Edit away, run your app as usual, and we'll automatically make it available for you to access.

When using "conda" from outside of GitHub Codespaces, note the Anaconda repository contains restrictions on commercial use that may impact certain organizations. See <https://aka.ms/ghcs-conda>

codespaces → /workspaces/pydata-workshop-2023 (main) \$

Categories

- Programming Languages
- Snippets
- Machine Learning
- Education

Under The Hood: What's GitHub Copilot (Chat)?

- Code explanations
- Code assistance
- Code refinement
- Unit testing
- Code profiling
- Code debugging

Ask questions in natural language – get responses inline, and in context!

About GitHub Copilot

GitHub Copilot is an AI pair programmer that offers autocomplete-style suggestions as you code. You can receive suggestions from GitHub Copilot either by starting to write the code you want to use, or by writing a natural language comment describing what you want the code to do. GitHub Copilot analyzes the context in the file you are editing, as well as related files, and offers suggestions from within your text editor. GitHub Copilot is powered by a generative AI model developed by GitHub, OpenAI, and Microsoft.

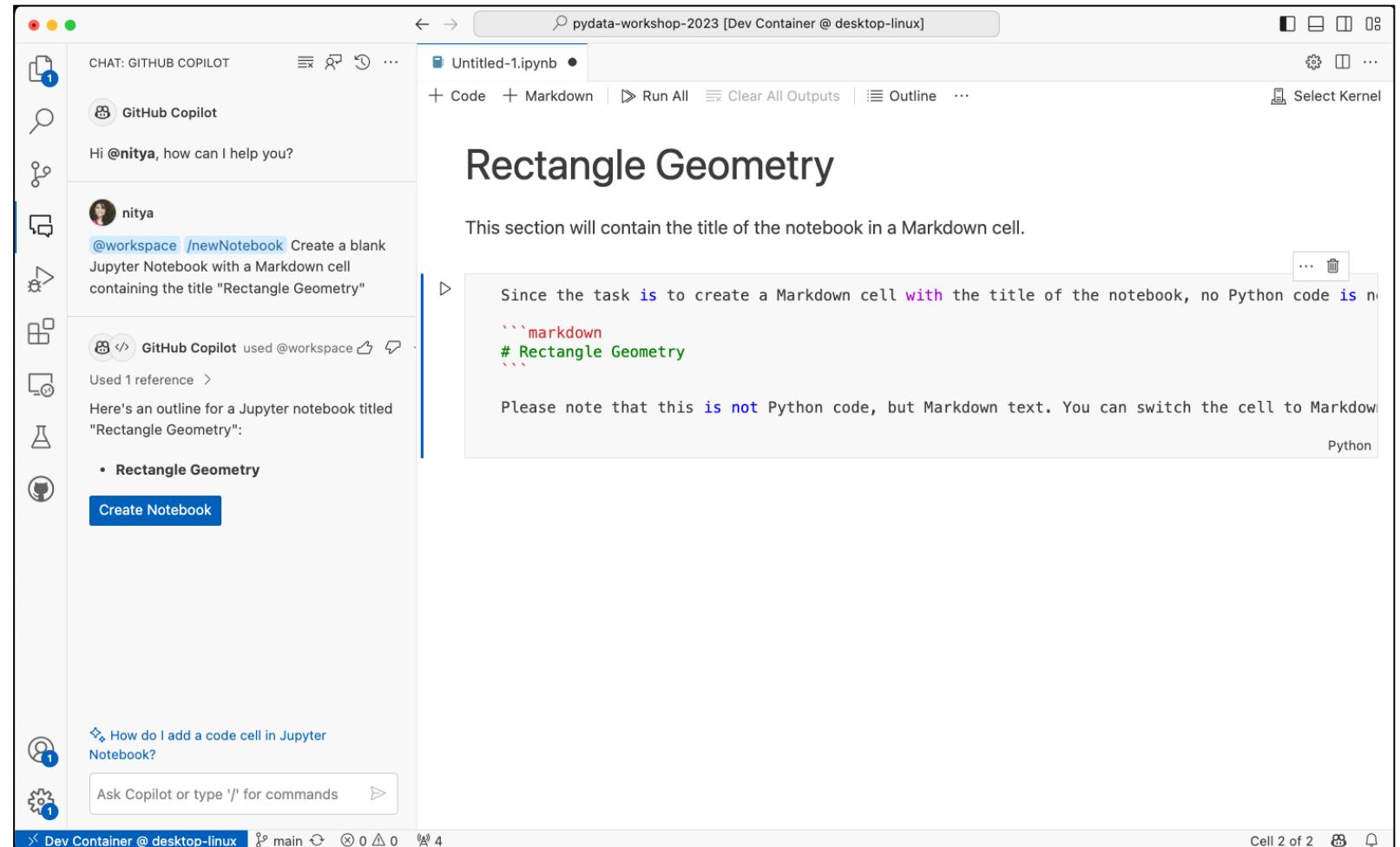
GitHub Copilot is trained on all languages that appear in public repositories. For each language, the quality of suggestions you receive may depend on the volume and diversity of training data for that language. For example, JavaScript is well-represented in public repositories and is one of GitHub Copilot's best supported languages. Languages with less representation in public repositories may produce fewer or less robust suggestions.

GitHub Copilot is available as an extension in Visual Studio Code, Visual Studio, Vim, Neovim, the JetBrains suite of IDEs, and Azure Data Studio. For more information on using GitHub Copilot in Visual Studio Code, Visual Studio, Vim, Neovim, and JetBrains, see "[Getting started with GitHub Copilot](#)." For more information on using GitHub Copilot in Azure Data Studio, see [GitHub Copilot extension](#) in Microsoft Learn.

Exercise 4: Have it create a `/newNotebook` for me

Copilot Chat provides a chat experience that uses the current window and history as context.

It's not perfect – see the Python cell – but it's a quick start to my task.



Exercise 4: Let's setup the notebook with our problem

I have a USACO problem called "[Blocked Billboard](#)" that has a [published solution](#).

Let's first add the problem into our Notebook as Markdown. This becomes a workspace where I can try coding my answers.

The screenshot shows a Jupyter Notebook interface with the following details:

- EXPLORER** panel on the left:
 - TIMELINE**: Shows activity: "edit" (1 min), "Insert Cell" (2 mins), "edit" (3 mins).
 - PYDATA-WORKSHOP-2023 [DEV ...]**:
 - .devcontainer
 - devcontainer.json
 - icon.svg
 - data
 - docs
 - notebooks
 - pydata**:
 - 01-newNotebook.ipynb
 - billboard.in
 - billboard.out
 - .gitignore
 - LICENSE
 - README.md
 - requirements.txt
- Code Editor** panel on the right:
 - 01-newNotebook.ipynb** (active tab)
 - billboard.in**

Rectangle Geometry

Let's look at the [Blocked Billboard](#) problem from USACO 2017 December Contest, Bronze

During long milking sessions, Bessie the cow likes to stare out the window of her barn at two huge rectangular billboards across the street advertising "Farmer Alex's Amazingly Appetizing Alfalfa" and "Farmer Greg's Great Grain". Pictures of these two cow feed products on the billboards look much tastier to Bessie than the grass from her farm. One day, as Bessie is staring out the window, she is alarmed to see a huge rectangular truck parking across the street. The side of the truck has an advertisement for "Farmer Smith's Superb Steaks", which Bessie doesn't quite understand, but she is mostly concerned about the truck potentially blocking the view of her two favorite billboards.

Given the locations of the two billboards and the location of the truck, please calculate the total combined area of both billboards that is still visible. It is possible that the truck obscures neither, both, or only one of the billboards.

INPUT FORMAT (file billboard.in):

The first line of input contains four space-separated integers: $x_1 \ y_1 \ x_2 \ y_2$, where (x_1, y_1) and (x_2, y_2) are the coordinates of the lower-left and upper-right corners of the first billboard in Bessie's 2D field of view.

The next line contains four more integers, similarly specifying the lower-left and upper-right corners of the second billboard. The third and final line of input contains four integers specifying the lower-left and upper-right corners of the truck. All coordinates are in the range -1000 to +1000. The two billboards are guaranteed not to have any positive area of overlap between themselves.

OUTPUT FORMAT

The output should contain a single integer representing the total combined area of the billboards that are still visible.

TERMINAL tab at the bottom shows a bash session in a pydata codespace.

Exercise 4: Copy solution – and run it (provide inputs)

Now we can use Copilot Chat to ask questions to help us understand the code and complexity

 nitya

what is time complexity mean?

 GitHub Copilot ◀ ▶ ⋮

Used 1 reference ▾

 01-newNotebook.ipynb:1-29 pydata

Time complexity is a concept in computer science that describes the amount of time an algorithm takes to run as a function of the size of the input to the program. It is usually expressed using Big O notation, which describes the upper bound of the time complexity in the worst-case scenario. For example, a linear search algorithm has a time complexity of $O(n)$ because the time it takes to find an item (or determine it's not there) grows linearly with the size of the list being searched.

The screenshot shows a Jupyter Notebook interface with the following details:

- Top Bar:** Includes tabs for Code, Markdown, Interrupt, Restart, Clear All Outputs, Go To, Variables, and Python 3.10.8.
- Left Sidebar (EXPLORER):** Shows a file tree with the following structure:
 - TIMELINE:** 01-newNotebook.ipynb
 - edit (now)
 - edit (34 secs)
 - edit (1 min)
 - PYDATA-WORKSHOP-2023 [DEV CONTAINER @ DE...]**
 - .devcontainer
 - devcontainer.json
 - icon.svg
 - > data
 - > docs
 - > notebooks
 - pydata**
 - 01-newNotebook.ipynb (selected)
 - billboard.in
 - billboard.out
 - .gitignore
 - LICENSE
 - README.md
 - requirements.txt
- Code Editor:** Displays the following Python code:

```
1 import sys
2
3 MAX_POS = 2000
4
5
6 def main():
7     sys.stdin = open("billboard.in", "r")
8     sys.stdout = open("billboard.out", "w")
9
10    visible = [[False for _ in range(MAX_POS)] for _ in range(MAX_POS)]
11
12    for i in range(3):
13        x1, y1, x2, y2 = map(int, input().split())
14        x1 += MAX_POS // 2
15        y1 += MAX_POS // 2
16        x2 += MAX_POS // 2
17        y2 += MAX_POS // 2
18        for x in range(x1, x2):
19            for y in range(y1, y2):
20                visible[x][y] = i < 2
21
22    ans = 0
23    for x in range(MAX_POS):
24        for y in range(MAX_POS):
25            ans += visible[x][y]
26    print(ans)
27
28
29 main()
```
- Bottom Navigation:** Includes tabs for PROBLEMS, OUTPUT, DEBUG CONSOLE, TERMINAL, PORTS (5), JUPYTER, and a terminal tab labeled "bash - pydata".
- Bottom Status Bar:** Shows the current file as "01-newNotebook.ipynb" and the status "main*". It also displays the line number (Ln 6, Col 12), tab size (Tab Size: 4), and cell count (Cell 9 of 10).

Exercise 4: Copy solution – and run it (provide inputs)

Notice how I stay in flow
(no external visits) – and
am prompted to learn
more in context ...

- explain this code to me
 - how can I improve the solution
 - can you simplify it using a class

The “[Fast Solution](#)” from USACO agrees on O(1) but uses classes. I can ask about that too..

 nitya

explain this code to me

 GitHub Copilot thumb up thumb down ...

Used 1 reference ▾

 01-newNotebook.ipynb:1-29 pydata

This Python script is designed to solve a problem related to billboards. It reads input from a file named "billboard.in" and writes the output to a file named "billboard.out".

Here's a step-by-step explanation:

1. It first initializes a 2D boolean array `visible` of size `MAX_POS x MAX_POS`, where `MAX_POS` is a constant set to 2000. This array represents a grid where each cell indicates whether a point is visible or not.
2. It then reads three sets of coordinates from the input file. Each set of coordinates represents a rectangle on the grid. The coordinates are adjusted by adding `MAX_POS // 2` to handle negative coordinates.

❖ What is the purpose of the `MAX_POS` constant in the billboard code?

 nitya

What is the purpose of the `MAX_POS` constant in the billboard code?

 GitHub Copilot

Used 1 reference >

The `MAX_POS` constant in the billboard code is used to define the size of the 2D `visible` array. This array represents a grid of points, and `MAX_POS` is the maximum number of points in each dimension of the grid.

The value of `MAX_POS` is also used to adjust the coordinates of the rectangles read from the input. By adding `MAX_POS // 2` to each coordinate, the code ensures that all coordinates are non-negative and fit within the grid. This is necessary because the problem statement may allow for negative coordinates, but array indices in Python must be non-negative.

Exercise 4: I can also explore this more intuitively



This time let's use Copilot inline, so it generates code for us

We don't know "how" to visualize it – but since the goal was to explore our intuition visually (vs. code the solution), the end result really helps!

Can You visualize the billboards and the truck for me

Accept Discard ⚡

Changed 20 lines

```
1 import matplotlib.pyplot as plt
2
3 # billboard 1 coordinates
4 x1, y1, x2, y2 = 1, 2, 3, 5
5
6 # billboard 2 coordinates
7 x3, y3, x4, y4 = 6, 0, 10, 4
8
9 # truck coordinates
10 x5, y5, x6, y6 = 2, 1, 8, 3
11
12 # plot the billboards and the truck
13 fig, ax = plt.subplots()
14 ax.add_patch(plt.Rectangle((x1, y1), x2-x1,
15 ax.add_patch(plt.Rectangle((x3, y3), x4-x3,
16 ax.add_patch(plt.Rectangle((x5, y5), x6-x5,
17 plt.xlim(-10, 20)
18 plt.ylim(-10, 20)
19 plt.show()
```

The figure shows a coordinate system with both x and y axes ranging from -10 to 20. Three rectangles are plotted: a blue rectangle centered at (1, 2) with vertices at (1, 2), (3, 2), (1, 5), and (3, 5); a green rectangle centered at (6, 0) with vertices at (6, 0), (10, 0), (6, 4), and (10, 4); and a red rectangle centered at (2, 1) with vertices at (2, 1), (8, 1), (2, 3), and (8, 3). The rectangles overlap slightly in the middle area.



Part 3:

Explore Data Visualization the traditional way

(I ask what I know about)

Exercise 5: Get a dataset to explore (e.g., from Kaggle)

ARAVIND A S · UPDATED A YEAR AGO

70 New Notebook Download (4 kB) ...

IPL 2022-Dataset

Matchwise data of Tata IPL 2022 (IPL Season 15)

Data Card Code (6) Discussion (0)

About Dataset

Complete Match-wise data of IPL 2022

MAKE AN EDA with this dataset

Content

This dataset contains Matchwise data of IPL matches 2022 (March26 - May29 2022) ,The complete data of group stage matches.

Attribute Information

1. Match Number
2. Date of the match
3. Name of Venue
4. Playing team 1
5. Playing team 2
6. Stage of the tournament

Usability ① 10.00

License Database: Open Database, Cont...

Expected update frequency Daily

Tags Tabular Sports Beginner Cricket pandas



ASHISH JANGRA · UPDATED 2 YEARS AGO

258 New Notebook Download (305 kB) ...

TED Talks

This datasets is having all the ted talks available on Ted.com

Data Card Code (58) Discussion (1)

About Dataset

Context

This dataset is created for beginner students of Data Analysis who can explore the field with real-life data. Using TED talk data will help them to analyze the talks and they can also watch the talks of their favorite author with the help of the dataset as well.

Content

This dataset contains 6 different features of each talk available on TED's website which you can find below

- title - Title of the Talk
- author - Author of Talk
- date - Date when the talk took place
- views - Number of views of the Talk
- likes - Number of likes of the Talk
- link - Link of the talk from ted.com

Acknowledgements

Usability ① 10.00

License CC BY-NC-SA 4.0

Expected update frequency Quarterly

Tags Earth and Nature Education Beginner Data Analytics Data Visualization Exploratory Data Analysis Python



[IPL2022 Dataset](#) - shared with [Open Data Commons License](#)
↗ downloaded Oct 2023 · [Example EDA](#) on Kaggle.

[TED Talks Dataset](#) - shared with [Creative Commons License](#)
↗ downloaded Oct 2023 · [Example EDA](#) on Kaggle.

Exercise 5: Can I have it teach me to visualize it?



nitya

@workspace /newNotebook Create a notebook that explains how to use the seaborn package for data visualizations - use the data/kaggle/IPL-2022.csv file as your input example

GitHub Copilot used @workspace

Used 1 reference >

Here's an outline for a Jupyter notebook that explains how to use the seaborn package for data visualizations using the IPL-2022.csv file:

- Import Required Libraries
- Load the IPL-2022.csv Dataset
- Data Overview
- Visualize the Distribution of Runs
- Visualize the Relationship between Variables
- Visualize Categorical Data

[Create Notebook](#)

Use pandas to load the IPL-2022.csv dataset.

```
1
2 # Load the IPL-2022.csv dataset using pandas
3 ipl_data = pd.read_csv('../data/kaggle/IPL-2022.csv')
4
5 # Display the first 5 rows of the dataset
6 ipl_data.head()
7
```

3] ✓ 0.0s Python

| | match_id | date | venue | team1 | team2 | stage | toss_winner | toss_decision | first_ings_s |
|---|----------|---------------|---|-----------|-----------|-------|-------------|---------------|--------------|
| 0 | 1 | March 26,2022 | Wankhede Stadium, Mumbai | Chennai | Kolkata | Group | Kolkata | Field | |
| 1 | 2 | March 27,2022 | Brabourne Stadium, Mumbai | Delhi | Mumbai | Group | Delhi | Field | |
| 2 | 3 | March 27,2022 | Dr DY Patil Sports Academy, Mumbai | Banglore | Punjab | Group | Punjab | Field | |
| 3 | 4 | March 28,2022 | Wankhede Stadium, Mumbai | Gujarat | Lucknow | Group | Gujarat | Field | |
| 4 | 5 | March 29,2022 | Maharashtra Cricket Association Stadium, Pune | Hyderabad | Rajasthan | Group | Hyderabad | Field | |

Exercise 5: Can I have it fix mistakes it made?

nitya

@workspace /newNotebook Create a notebook that explains how to use the seaborn package for data visualizations - use the data/kaggle/IPL-2022.csv file as your input example

GitHub Copilot used @workspace    

Used 1 reference >

Here's an outline for a Jupyter notebook that explains how to use the seaborn package for data visualizations using the IPL-2022.csv file:

- Import Required Libraries
- Load the IPL-2022.csv Dataset
- Data Overview
- Visualize the Distribution of Runs
- Visualize the Relationship between Variables
- Visualize Categorical Data

[Create Notebook](#)

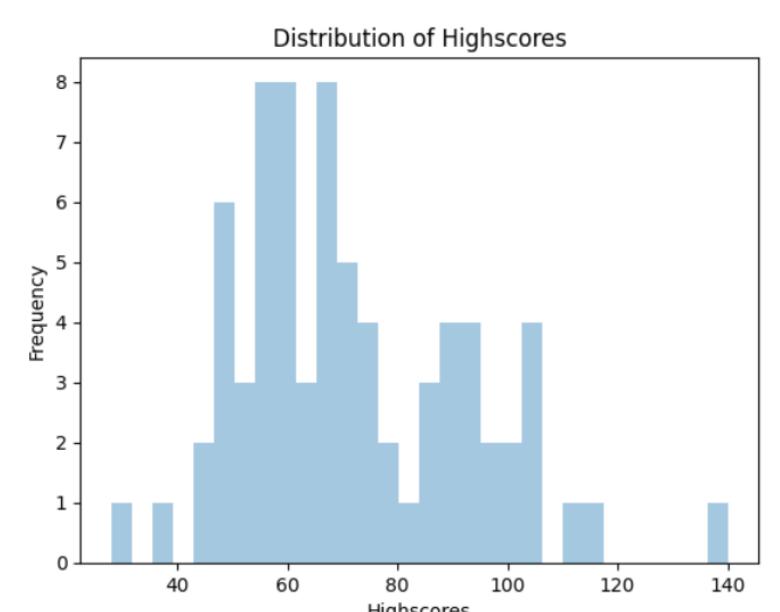
fix this to use highscore column for runs 

Copilot generated code may be incorrect

```
1 # Visualize the Distribution of Runs
2
3 # Check if the 'runs' column exists in the dataset
4 if 'runs' in ipl_data.columns:
5     # Use seaborn's distplot() function to visualize the distribution of runs in the data
6     sns.distplot(ipl_data['runs'], kde=False, bins=20)
7
8     # Set the title for the plot
9     plt.title('Distribution of Runs')
10
11    # Set the x-label for the plot
12    plt.xlabel('Runs')
13
14    # Set the y-label for the plot
15    plt.ylabel('Frequency')
16
17    # Display the plot
18    plt.show()
19
20 else:
21     print("'runs' column does not exist in the dataset.")
```

[5] 0.0s

... 'runs' column does not exist in the dataset.



| Highscore Range | Frequency |
|-----------------|-----------|
| 35-40 | 1 |
| 40-45 | 1 |
| 45-50 | 2 |
| 50-55 | 6 |
| 55-60 | 8 |
| 60-65 | 8 |
| 65-70 | 5 |
| 70-75 | 4 |
| 75-80 | 3 |
| 80-85 | 2 |
| 85-90 | 3 |
| 90-95 | 4 |
| 95-100 | 2 |
| 100-105 | 4 |
| 105-110 | 1 |
| 110-115 | 1 |
| 115-120 | 0 |
| 120-125 | 0 |
| 125-130 | 0 |
| 130-135 | 0 |
| 135-140 | 1 |

Exercise 5: Can it visualize something specific?

nitya

@workspace /newNotebook Create a notebook that explains how to use the seaborn package for data visualizations - use the data/kaggle/IPL-2022.csv file as your input example

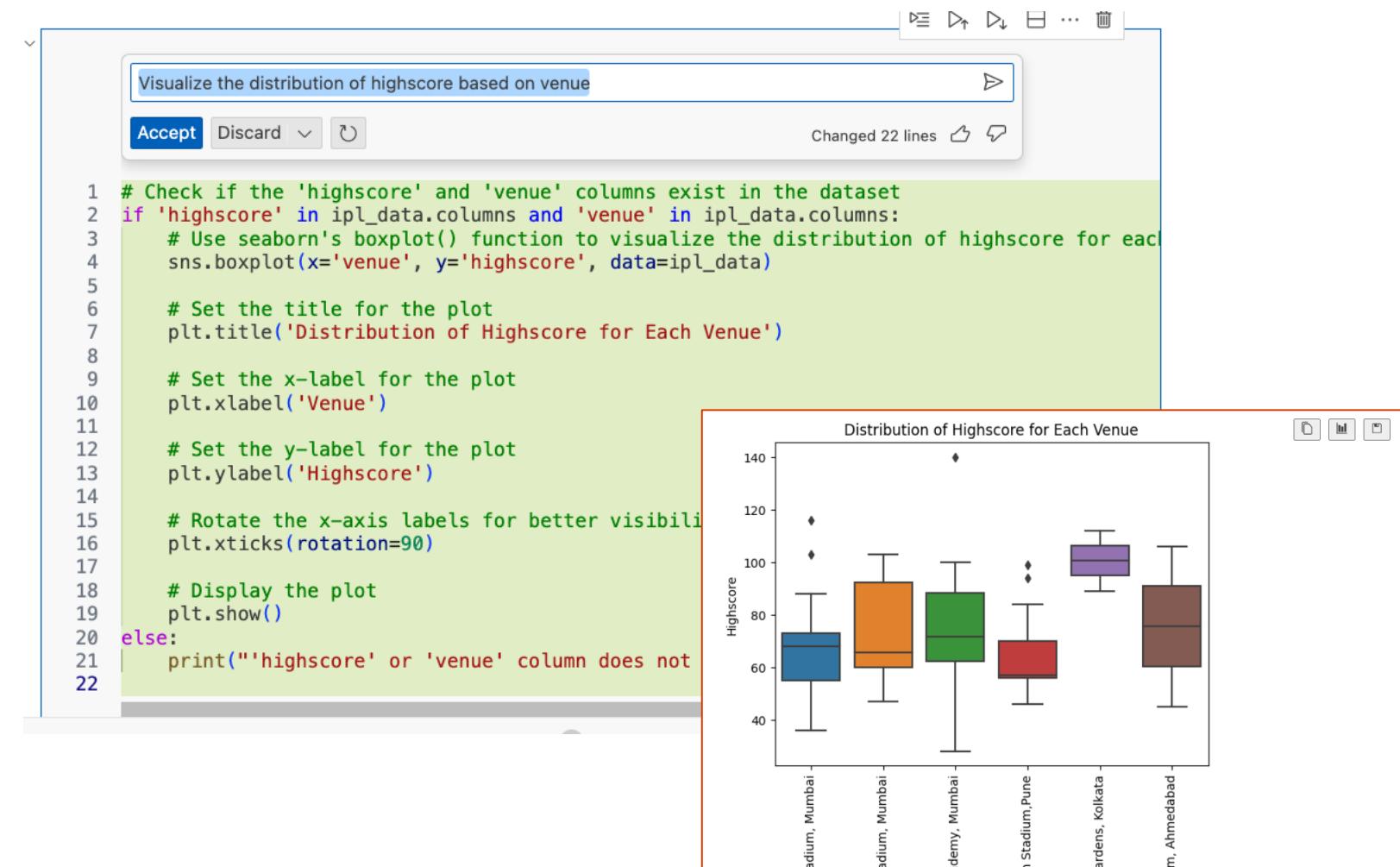
GitHub Copilot used @workspace

Used 1 reference >

Here's an outline for a Jupyter notebook that explains how to use the seaborn package for data visualizations using the IPL-2022.csv file:

- Import Required Libraries
- Load the IPL-2022.csv Dataset
- Data Overview
- Visualize the Distribution of Runs
- Visualize the Relationship between Variables
- Visualize Categorical Data

Create Notebook



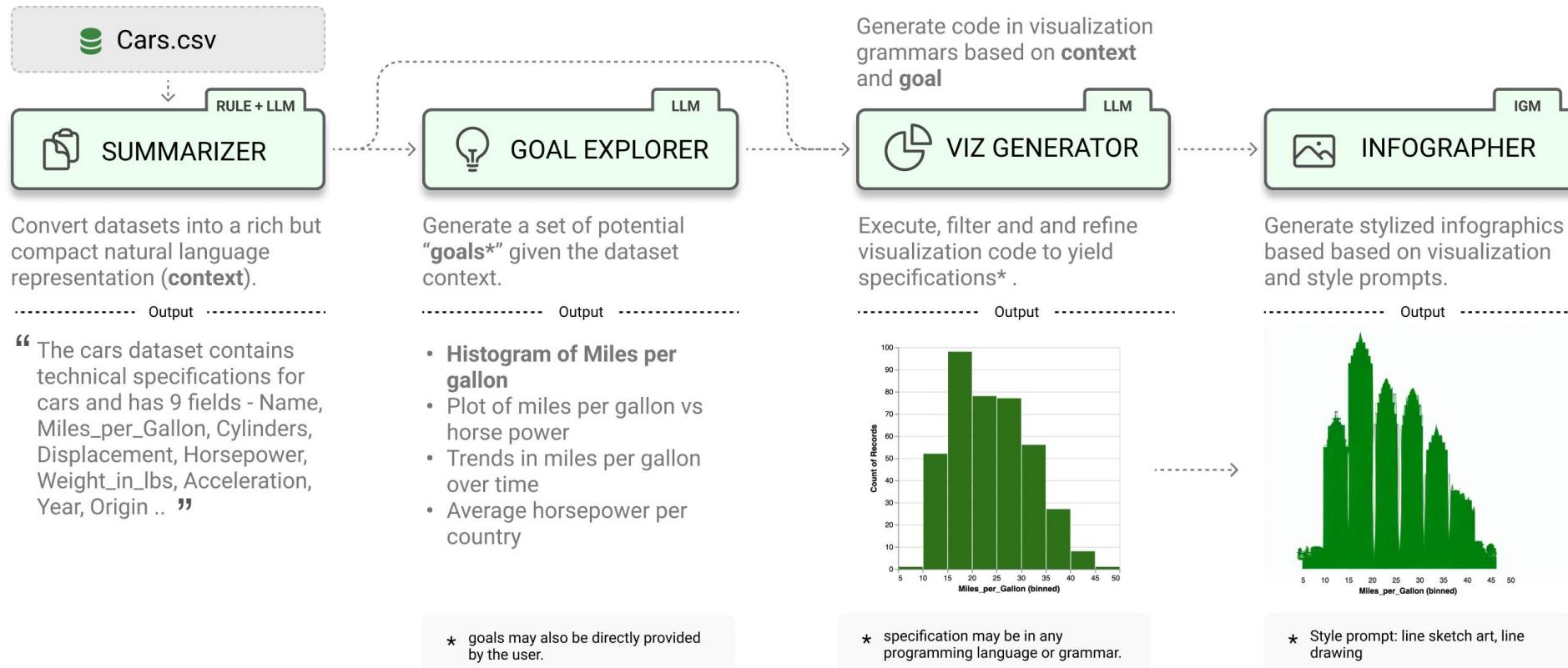


Part 4:

Explore Data Visualization without intuition (Using Open AI – with Microsoft LIDA)

LIDA: Generate Data Visualizations & Infographics

LIDA is a library for generating data visualizations and data-faithful infographics. LIDA is grammar agnostic (will work with any programming language and visualization libraries e.g. matplotlib, seaborn, altair, d3 etc) and works with multiple large language model providers (OpenAI, Azure OpenAI, PaLM, Cohere, Huggingface). Research Paper: <https://arxiv.org/abs/2303.02927>



Open AI API Key: Set it as env var (protect it as secret)



```
# Setup
from lida import Manager, TextGenerationConfig , llm

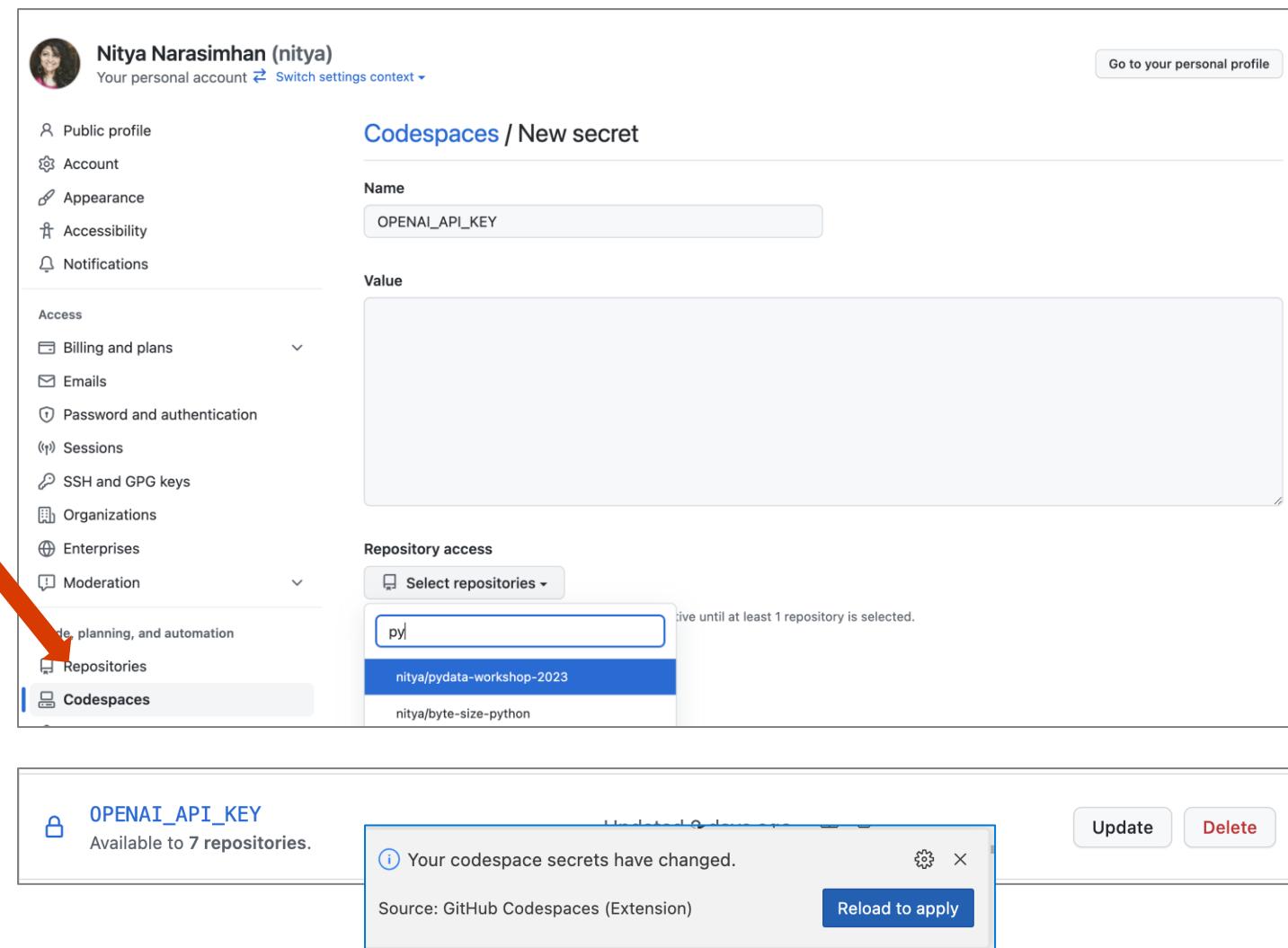
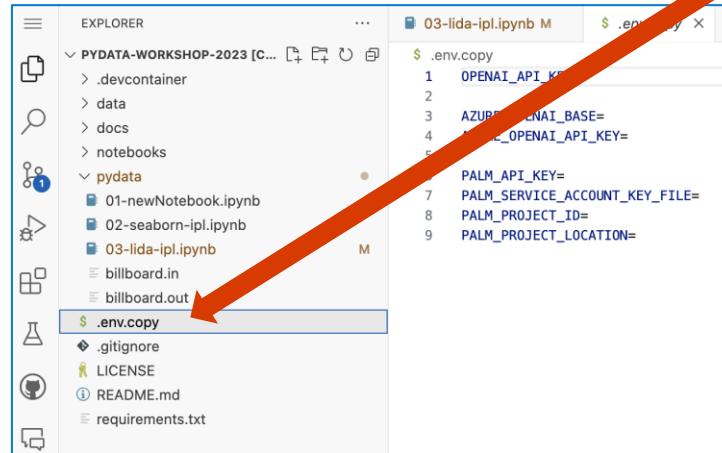
csvfile = "./../data/kaggle/IPL-2022.csv"
lida = Manager(text_gen = llm("openai")) # palm, cohere .
textgen_config = TextGenerationConfig(n=1, temperature=0.5, model="gpt-3.5-turbo-0301", use_cache=True)

[2] 0.0s

ValueError
Traceback (most recent call last)
File "/workspaces/pydata-workshop-2023/pydata/03-lida-ipynb Cell 3 line 5
  2 from lida import Manager, TextGenerationConfig , llm
  3 csvfile = "./../data/kaggle/IPL-2022.csv"
  4 lida = Manager(text_gen = llm("openai")) # palm, cohere .
  5 textgen_config = TextGenerationConfig(n=1, temperature=0.5, model="gpt-3.5-turbo-0301", use_cache=True)

File "/usr/lib/python3.10/site-packages/llms/generators/text/textgen.py:26, in llm(provider, **kwargs)
   23     api_key = api_key or os.environ.get("OPENAI_API_KEY", None)
   24     if api_key is None:
   25         raise ValueError(
   26             "OpenAI API key is not set. Please set the OPENAI_API_KEY environment variable."
   27         )
   28     openai.api_key = api_key
   29
   30 if organization:

ValueError: OpenAI API key is not set. Please set the OPENAI_API_KEY environment variable.
```



The screenshot shows the GitHub Codespaces 'New secret' configuration page for 'Nitya Narasimhan (nitya)'. The 'Name' field is set to 'OPENAI_API_KEY' and the 'Value' field is empty. The 'Access' section includes options for Billing and plans, Emails, Password and authentication, Sessions, SSH and GPG keys, Organizations, Enterprises, and Moderation. A dropdown menu for 'Select repositories' shows 'py' selected, with 'nitya/pydata-workshop-2023' highlighted. A modal at the bottom right indicates that 'Your codespace secrets have changed.' and provides 'Update' and 'Delete' buttons, along with a 'Reload to apply' button.

Exercise 6: Data Summarization with LIDA (foundation)

The screenshot shows a Jupyter Notebook interface within a GitHub repository. The repository name is 'pydata-workshop-2023'. The notebook is titled 'Step 4. Visualize Data Using Traditional Means'.

The code in the notebook is as follows:

```
# Setup
from lida import Manager, TextGenerationConfig , llm

csvfile = "./../data/kaggle/IPL-2022.csv"
lida = Manager(text_gen = llm("openai")) # palm, cohere .
textgen_config = TextGenerationConfig(n=1, temperature=0.5, model="gpt-3.5-turbo-0301", use_cache=True)

[1]    ✓ 1.0s                                         Python
```

```
# Summarize
summary = lida.summarize(csvfile)
summary_data = list(summary.keys())
for keys in summary_data:
    print(keys, ":", summary[keys])

[2]    ✓ 0.0s                                         Python
```

```
... name : IPL-2022.csv
file_name : IPL-2022.csv
dataset_description :
fields : [{"column': 'match_id', 'properties': {'dtype': 'number', 'std': 21, 'min': 1, 'max': 74, 'samples': [5, 64, 19], 'num_uni
field_names : ['match_id', 'date', 'venue', 'team1', 'team2', 'stage', 'toss_winner', 'toss_decision', 'first_ings_score', 'first_
```

Given a dataset, generate a compact summary of that data in a compact natural language representation that serves as context for subsequent tasks. The goal of the summarizer is to produce an dense-but-compact information summary for a given dataset that is useful as grounding context for visualization tasks. The grounding context is defined as one that contains information an analyst would need to understand the dataset and the tasks that can be performed on it.

See paper for details

Exercise 6: Goal Generation with LIDA (build intuition)

```
▶ 
1 # generate 10 goals from the summary with default persona
2 goals = lida.goals(summary, n=10, textgen_config=textgen_config,) # exploratory data analysis
3
4 # create a list of dictionaries containing the goal information
5 import pandas as pd
6 goal_list = []
7 for goal in goals:
8     goal_dict = {'Question': goal.question, 'Visualization': goal.visualization, 'Rationale': goal.rationale}
9     goal_list.append(goal_dict)
10 df = pd.DataFrame(goal_list)
11 display(df)
12
```

[4] ✓ 16.1s Python

| ... | Question | Visualization | Rationale |
|-----|---|----------------------------------|---|
| 0 | What is the distribution of first innings scor... | Histogram of first_ings_score | This visualization will help us understand the... |
| 1 | What is the distribution of second innings sco... | Histogram of second_ings_score | This visualization will help us understand the... |
| 2 | What is the average margin of victory in the m... | Bar chart of margin | This visualization will help us understand the... |
| 3 | Which team has won the most matches in the tou... | Bar chart of match_winner | This visualization will help us understand whi... |
| 4 | What is the distribution of scores for the top... | Histogram of highscore | This visualization will help us understand the... |
| 5 | What is the distribution of best bowling figur... | Histogram of best_bowling_figure | This visualization will help us understand the... |
| 6 | What is the distribution of the number of wick... | Bar chart of first_ings_wkts | This visualization will help us understand the... |
| 7 | What is the distribution of the number of wick... | Bar chart of second_ings_wkts | This visualization will help us understand the... |
| 8 | What is the distribution of player of the matc... | Bar chart of player_of_the_match | This visualization will help us understand the... |
| 9 | What is the distribution of matches played in ... | Bar chart of venue | This visualization will help us understand the... |

Exercise 6: Goal Generation with LIDA (add persona)

```
1 # generate 5 goals from the summary - with the persona is a fan of the Mumbai team
2 goals = lida.goals(summary, n=5, textgen_config=textgen_config, persona="fan of the Mumbai team who wants to see t
3
4 # create a list of dictionaries containing the goal information
5 import pandas as pd
6 goal_list = []
7 for goal in goals:
8     display(goal)
```

[13] ✓ 7.9s

Python

... Goal 0

Question: What is the performance of Mumbai team in the first innings compared to other teams?

Visualization: Bar chart of first_ings_score for Mumbai and other teams

Rationale: This visualization will help us understand how Mumbai team is performing in the first innings compared to other teams. It will give us an idea of whether Mumbai team is scoring more or less in the first innings compared to other teams.

... Goal 1

Question: What is the performance of Mumbai team in the second innings compared to other teams?

Visualization: Bar chart of second_ings_score for Mumbai and other teams

Rationale: This visualization will help us understand how Mumbai team is performing in the second innings compared to other teams. It will give us an idea of whether Mumbai team is scoring more or less in the second innings compared to other teams.

Exercise 6: Data Visualization Generation with LIDA

```
1 # Visualize A Goal
2 charts = lida.visualize(summary=summary, goal=goals[0]) # exploratory data analysis
3 print("Charts length:", len(charts))
4 charts[0]
[5] ✓ 5.5s
...
... Charts length: 1
...


What is the distribution of first innings scores across all matches?



A histogram showing the frequency distribution of first innings scores. The x-axis ranges from 80 to 220, and the y-axis (Frequency) ranges from 0.0 to 20.0. The distribution is roughly bell-shaped, centered around a mean of 171.12, indicated by a vertical red dashed line and a text box.


```

```
1 # Visualize it again - and specify a different library and textgen_config (change te
2 target = goals[2]
3 library = "seaborn"
4 textgen_config = TextGenerationConfig(n=1, temperature=0.2, use_cache=True)
5 charts = lida.visualize(summary=summary, goal=target,library=library,textgen_config=
6 charts[0]
[7] ✓ 4.2s
...


What is the average margin of victory in the matches?



A histogram showing the frequency distribution of margins of victory. The x-axis ranges from 0 to 80, and the y-axis (Margin) ranges from 0 to 80. A horizontal red dashed line is drawn at a margin of 16.97, labeled as the "Mean Margin".


```

Exercise 6: Data Visualization Generation (user query)

```
1 # Use natural language user query instead of pre-formulated goal
2 user_query = "What is the frequency of toss decisions based on team ?"
3 textgen_config = TextGenerationConfig(n=1, temperature=0.2, use_cache=True)
4 charts = lida.visualize(summary=summary, goal=user_query, textgen_config=textgen_config)
5 charts[0]
```

[8] ✓ 5.4s Python

...

What is the frequency of toss decisions based on team ?

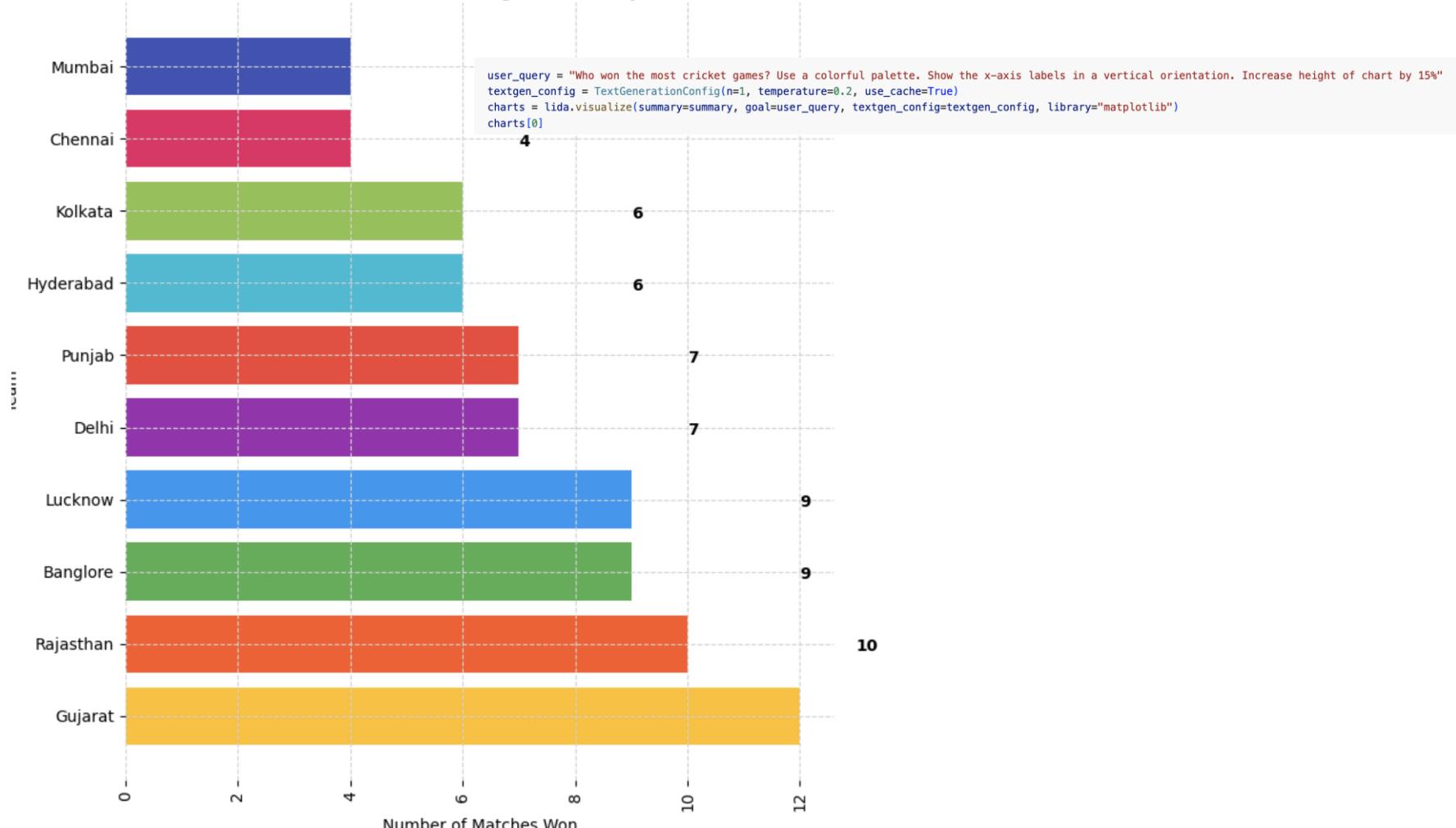
| City | Bat | Field |
|-----------|-----|-------|
| Bangalore | 3 | 13 |
| Chennai | 2 | 10 |
| Delhi | 1 | 11 |
| Gujarat | 3 | 7 |
| Hyderabad | 2 | 5 |
| Kolkata | 1 | 6 |
| Lucknow | 2 | 4 |
| Mumbai | 0 | 3 |
| Punjab | 1 | 0 |

GitHub Copilot
can do this too ..

But you can vary
the parameters
and customize the
base prompt here
programmatically

Exercise 6: Data Visualization Generation (user query)

Who won the most cricket games? Use a colorful palette. Show the x-axis labels in a vertical orientation. Increase height of chart by 15%



GitHub Copilot
can do this too ..

But you can vary
the parameters
and customize the
base prompt here
programmatically

Provides flexibility
for trial-and-error
experiments to
build intuition.

Exercise 6: Data Visualization Explanation with LIDA

```
▶ 1 # Explain visualization
  2 explanation = lida.explain(code=charts[0].code)
  3 for obj in explanation[0]:
  4     display(obj)
  5
[9] ✓ 9.8s
...
... {'section': 'accessibility',
  'code': 'None',
  'explanation': 'The goal of the chart is to show the frequency of toss decisions based on team. It uses a stacked bar chart with the x-axis representing the teams, the y-axis representing the frequency of toss decisions, and the color representing the toss decision (bat or field). The chart uses a white grid style and a Set2 color palette. The legend indicates the two possible toss decisions: bat or field.'}
...
... {'section': 'transformation',
  'code': "data = data[['team1', 'toss_decision']]\\n\n    data = data.groupby(['team1', 'toss_decision']).size().reset_index(name='count')",
  'explanation': "The code filters the data to include only the 'team1' and 'toss_decision' columns. It then groups the data by 'team1' and 'toss_decision' and counts the frequency of each group. The resulting data frame has columns 'team1', 'toss_decision', and 'count'."
}
...
... {'section': 'visualization',
  'code': "sns.set_style('whitegrid')\\nsns.set_palette('Set2')\\nax = sns.barplot(x='team1', y='count',
hue='toss_decision', data=data)\\nax.set(xlabel='Team', ylabel='Frequency')\\nplt.title('What is the frequency of toss
decisions based on team ?', wrap=True)\\nplt.legend(title='Toss Decision', loc='upper right')\\nplt.xticks(rotation=45,
ha='right')\\nreturn plt;",
  'explanation': "The code sets the style to white grid and the color palette to Set2. Then, it creates a bar plot using the seaborn library, with 'team1' on the x-axis, 'count' on the y-axis, and 'toss_decision' as the hue. The resulting chart is a stacked bar chart showing the frequency of toss decisions for each team. The x-axis labels are rotated 45 degrees for better readability. The chart has a title and a legend showing the two possible toss decisions: bat or field."}
```

system_prompt =

You are a helpful assistant highly skilled in providing helpful, structured explanations of visualization of the plot(data: pd.DataFrame) method in the provided code. You divide the code into sections and provide a description of each section and an explanation. The first section should be named "accessibility" and describe the physical appearance of the chart (colors, chart type etc), the goal of the chart, as well the main insights from the chart.

You can explain code across the following 3 dimensions:

- accessibility:** the physical appearance of the chart (colors, chart type etc), the goal of the chart, as well the main insights from the chart.
- transformation:** This should describe the section of the code that applies any kind of data transformation (filtering, aggregation, grouping, null value handling etc)
- visualization:** step by step description of the code that creates or modifies the presented visualization.

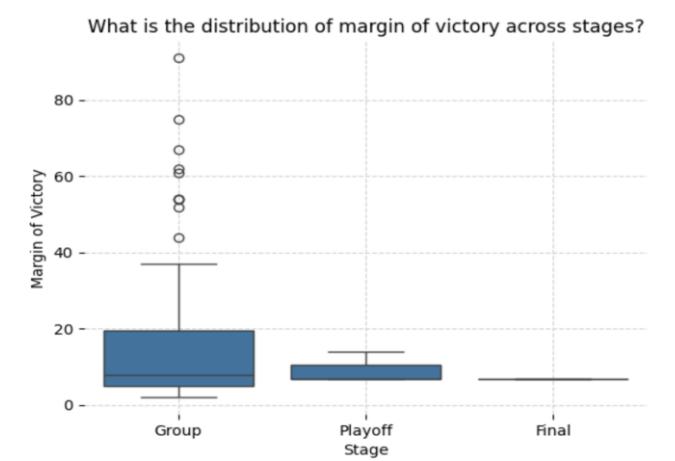
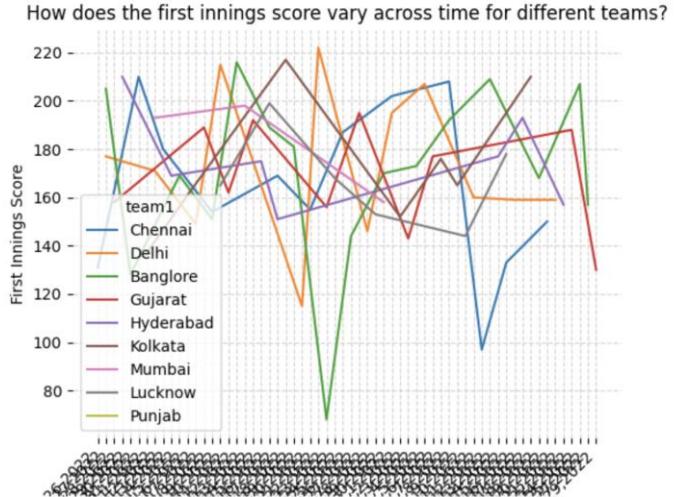
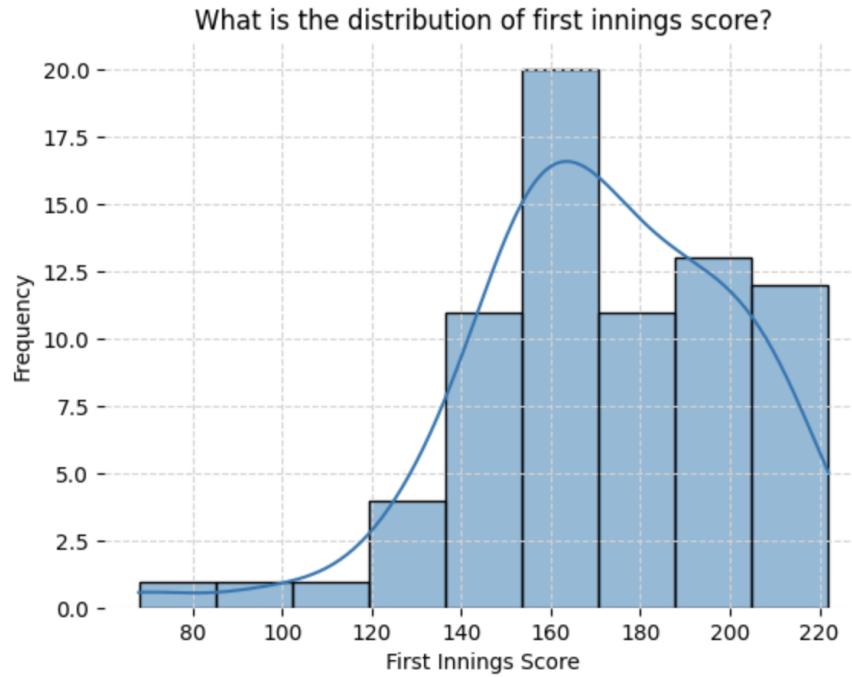
[lida/components/viz/vizexplainer.py](#)

Exercise 6: Visualization Recommendation (for a goal)

```
▷ ▾ # Recommend 3 visualizations
recommendations = lida.recommend(code=charts[0].code, summary=summary, n=3, textgen_config=textgen_config)

for chart in recommendations:
    display(chart)
[11] ✓ 5.3s
...

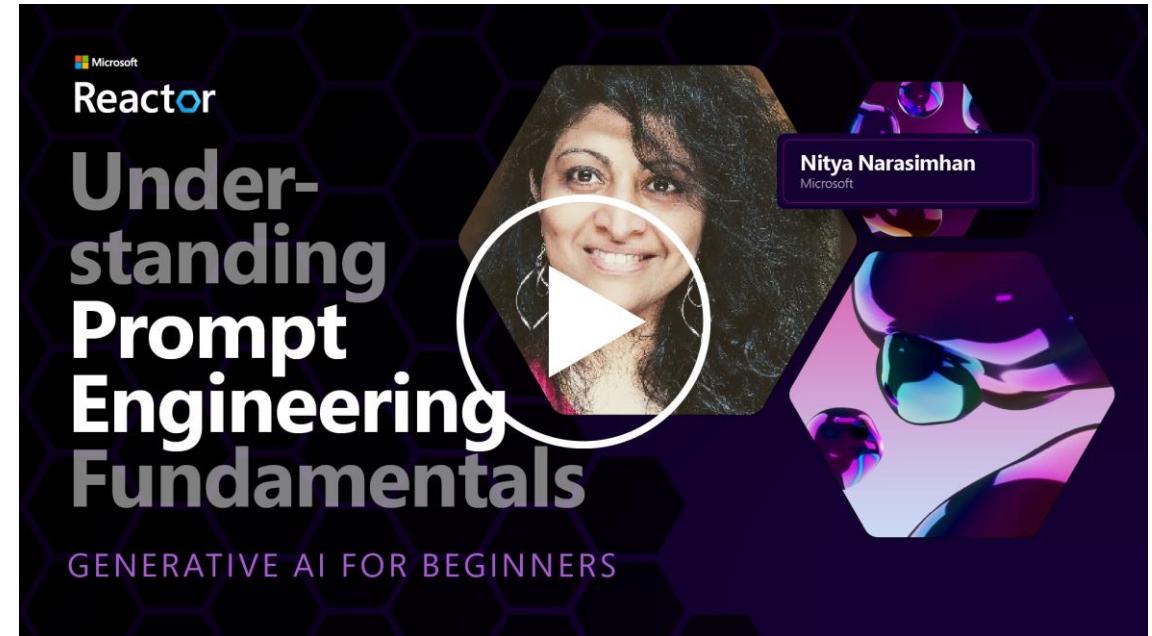
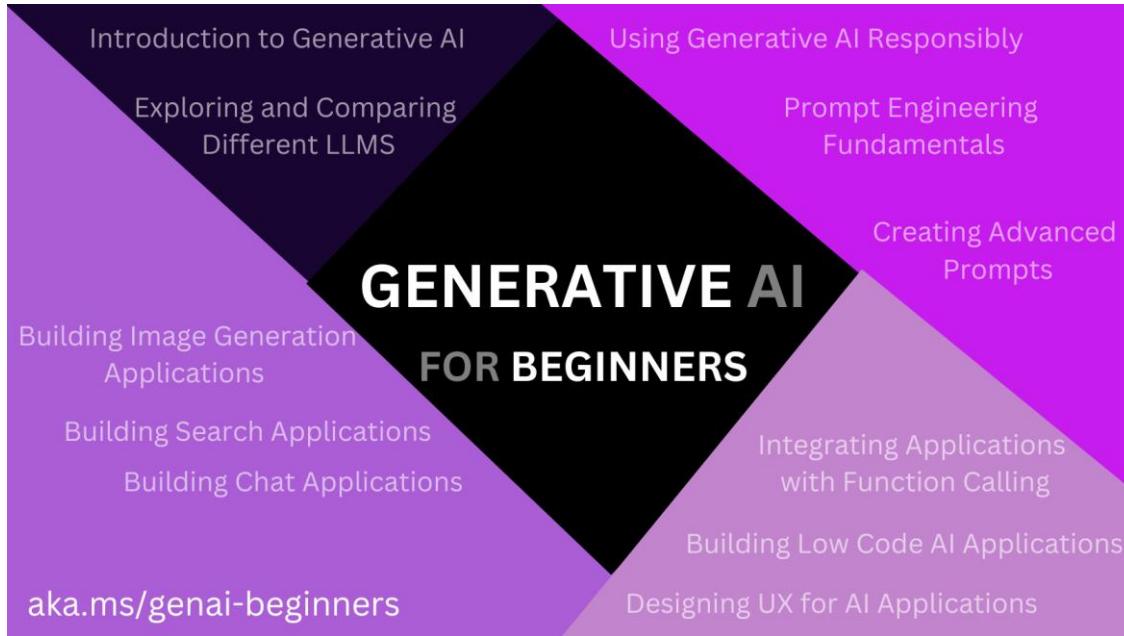
```





Wrap-Up: Where do we go from here?

Generative AI Course – From Principles to Prototypes



12 Lessons to start with (free, open-source)
Jupyter Notebooks (assignments)
Github Codespaces (environment)
OpenAI or Azure OpenAI (API key)
Practice Prompt Engineering

Learn by doing more ..

Learning Resources

The dev container can be used with either GitHub Codespaces (online, in the cloud) or with Docker Desktop (offline, in local device). It is set up for default usage with a Visual Studio Code editor frontend. With this in mind, here are some learning resources for beginners:

1. [Python Developer Roadmap](#) - step-by-step topics guidance for learning
2. [Get Started with Python in VS Code](#) - learn fundamentals.
3. [Data Science in VS Code](#) - skill up on Jupyter Notebooks!
4. [GitHub Codespaces for Machine Learning](#) - build classifiers & more.
5. [OpenAI Cookbook](#) - example code, guides for prompt engineering
6. [Getting Started With GitHub Copilot](#) - understand capabilities
7. [GitHub Copilot: Fly with Python ...](#) - RealPython community article
8. [How to use GitHub Copilot: Prompts, tips, use cases](#) - Github blog
9. [Create Dynamic Data Viz with OpenAI & React](#) - OSS Dashboard
10. [Chat with your CSV..](#) - OSS Demo with Langchain/Streamlit
11. [PandasAI](#) - Analyze data using natural lang (OpenAI, HuggingFace)

Contribute to help others ..



LIDA Project

Organization for sharing projects and resources that build on the open-source microsoft/lida tool (LLMs for Automated Visualization)

Follow

README.md

About LIDA Project

LIDA Project is an organization for support community-facing projects and experiments using [Microsoft/Lida](#). Here is a brief summary of what LIDA is, and its core features.

What is LIDA?

LIDA is a library for generating data visualizations and data-faithful infographics.

LIDA is grammar agnostic (will work with any programming language and visualization libraries e.g. matplotlib, seaborn, altair, d3 etc) and works with multiple large language model providers (OpenAI, Azure OpenAI, PaLM, Cohere, Huggingface).

- 📖 | Read [the paper](#) for technical details on how LIDA is implemented.
- 📈 | See [this notebook](#) for a sample tutorial.
- 🌐 | Check out [this website](#) for project updates.
- ⭐ | Star [this repo](#) if you like the project.
- 🎥 | Watch [this video](#) for a quick walkthrough of LIDA in action.

 View as: Public

You are viewing the README and pinned repositories as a public user.

[Get started with tasks](#) that most successful organizations complete.

Discussions

Set up discussions to engage with your community!

[Turn on discussions](#)

People



[Invite someone](#)



Watch **this** repo for more examples and tutorials for #30Days learning journeys

<https://aka.ms/pydata-for-jsdevs>

Follow **this** org, codespaces template, streamlit application
<https://aka.ms/lida/org>



For the non-Python
Programmer

Simplifying Data Analysis With GitHub Codespaces, Jupyter Notebooks & Open AI

- Transferable Learning with Jupyter Notebooks
- Consistent Dev Environment with GitHub Codespaces
- Focused Learning (in-context) with GitHub Copilot
- AI-assisted intuition using Microsoft LIDA

Thank you

<https://aka.ms/pydata-workshop-2023>

