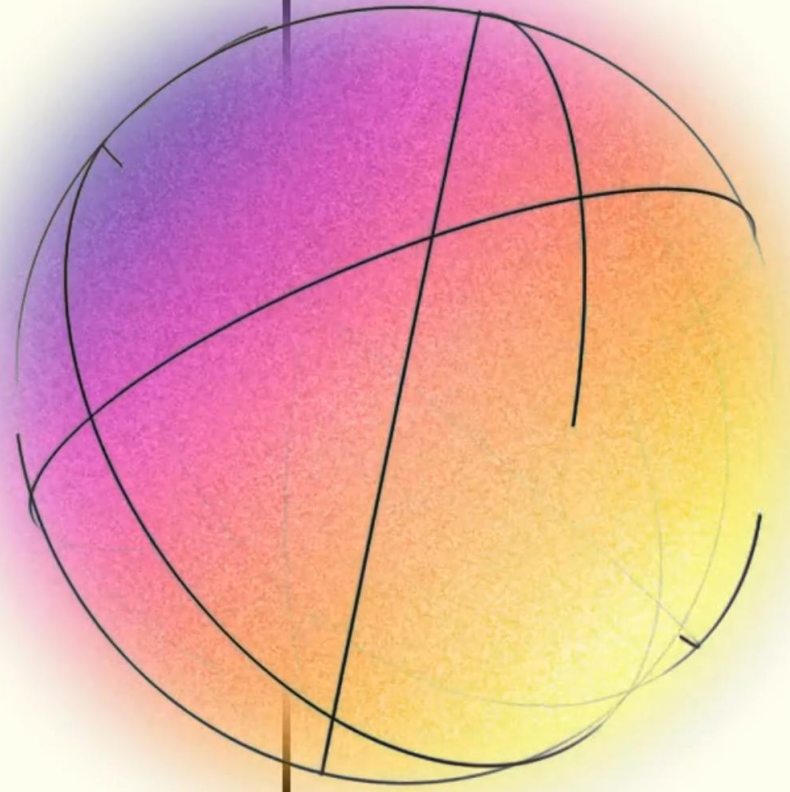


SciPy 2025



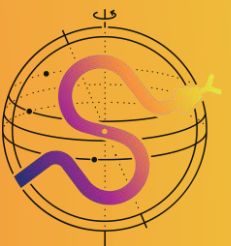
Develop Pythonic spreadsheets: Running Python in and out of the grid

Sarah Kaiser

Sr. Cloud Developer Advocate @ Microsoft

Jim Kitchen

Sr. Software Engineer @ Anaconda



Agenda

01

What is a spreadsheet really?

02

Intro to Python in Excel

03

Advanced Python in Excel

04

Advanced Analysis + Copilot for Python in Excel

05

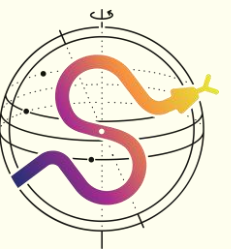
Anaconda Toolbox and Code

06

Using pandas with spreadsheets

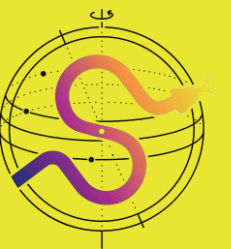
07

Customize spreadsheets with openpyxl

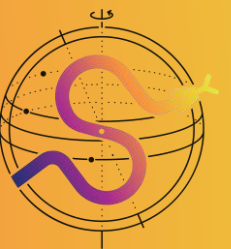


You will learn how to..

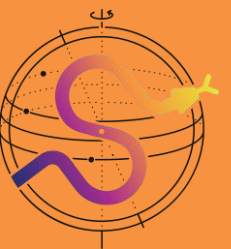
- Load and work with spreadsheets in Python
- Run Python directly in Excel
- Use AI tools like Advanced Analysis in Excel
- Use Anaconda Toolbox and Code in Excel



1. What is a spreadsheet really?



	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa



data.csv

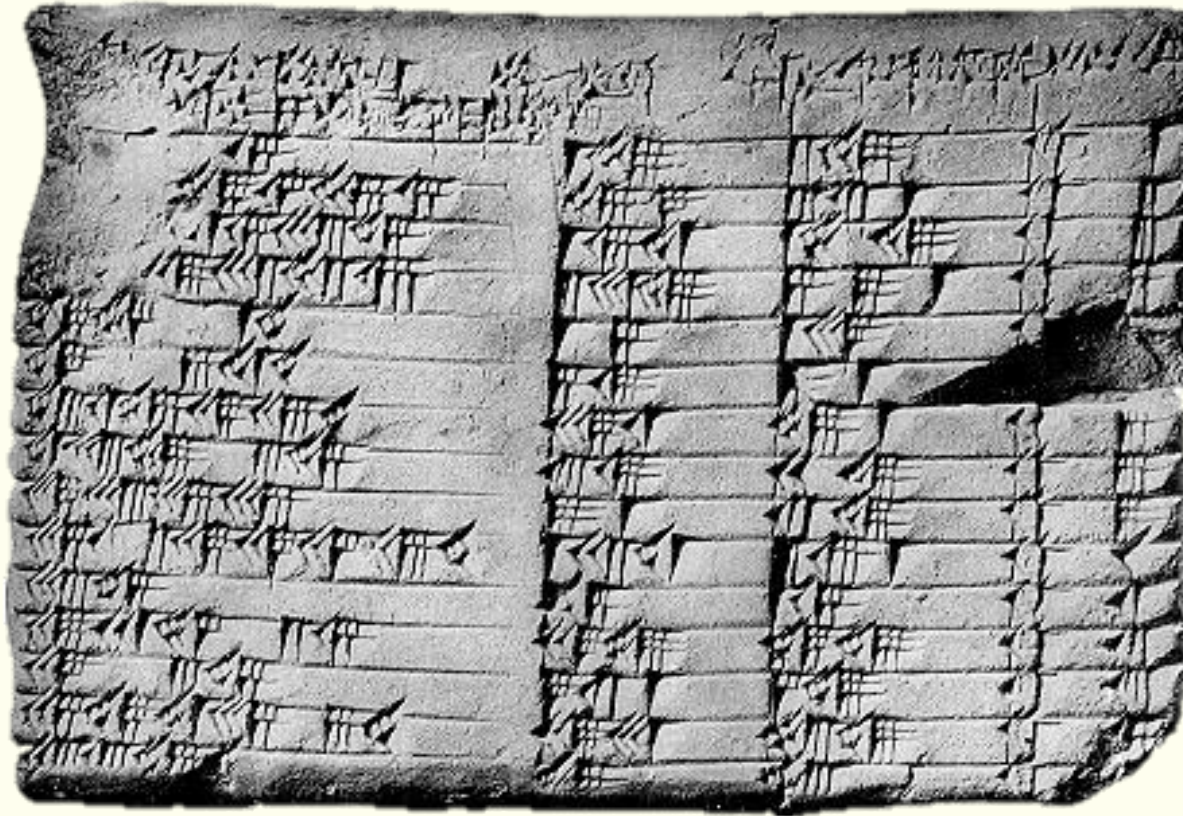


Year, Make, Model, Length

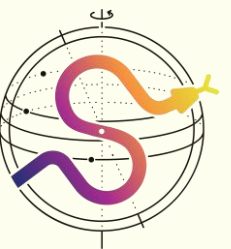
1997, Ford, E350, 2.35

2000, Mercury, Cougar, 2.38

Tab(let)ular data

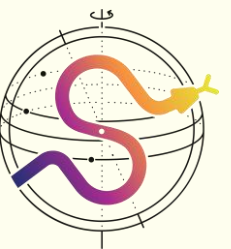


"Plimpton 322". personal.math.ubc.ca. Retrieved 2023-05-20.



Tabular text formats

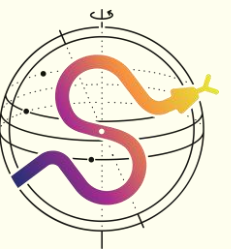
- csv, tsv, or other text delineated files
 - **Lowest common denominator format**
 - Human readable
 - Efficient storage
- Mainly used for storage, need other tools to analyze/plot



Tabular Spreadsheets

- Applications that allow storing and manipulation of tabular data
 - ex. VisiCalc, Lotus 1-2-3, Excel, LibreOffice Calc, PySpread
- Automatic/real-time calculation alongside data
- “What-if” analysis

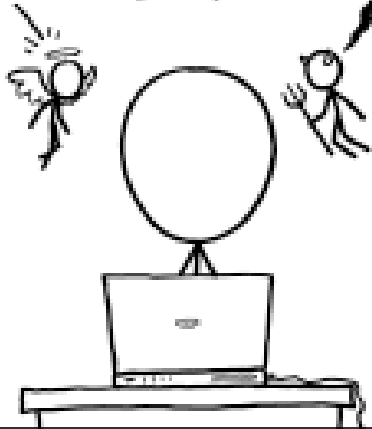
in other words... **Accessible Data Science**



DON'T USE A SPREAD-SHEET! DO IT *RIGHT*.

BUT A SPREADSHEET WOULD BE SO *EASY*.

IN THE LONG RUN YOU'LL REGRET IT!

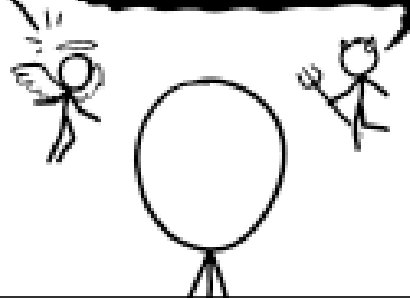


TAKE THE TIME TO WRITE *REAL* CODE.

JUST PASTE THE DATA! TINKER UNTIL IT WORKS!

BUILD A LABYRINTH OF REGEXREPLACE() AND ARRAYFORMULA()!

FEEL THE POWER!



FIGHT THE TEMPTATION!

EVER TRIED QUERY() IN GOOGLE SHEETS? IT LETS YOU TREAT A BLOCK OF CELLS LIKE A DATABASE AND RUN SQL QUERIES ON THEM.



DON'T LISTEN TO—
...WAIT. REALLY?

YES. AND LET ME TELL YOU ABOUT IMPORTHTML()...

OOOH...



<https://xkcd.com/2180/>

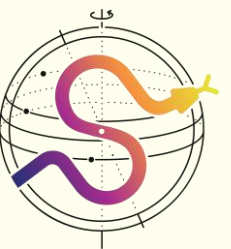


Spreadsheets: Limitations

- Harder to audit/version control
 - Can be harder to collaborate
- Potentially lots of function repetition
- More limited data sizes



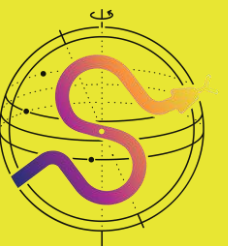
sales-numbers-finalV3_FINAL_sent.xlsx



	A	B	C
1	sepal length	sepal width	species
2	5.1	3.5	setosa
3	4.9	3	setosa
4	4.7	3.2	setosa
5	4.6	3.1	setosa
	sheet1		

Cell reference syntax

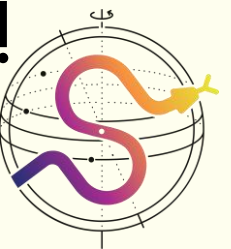
- B2
- A3:C3
- sheet1!C5
- \$C\$1
 - \$ pins a reference, can do column, row or both



Historically, dates are hard...

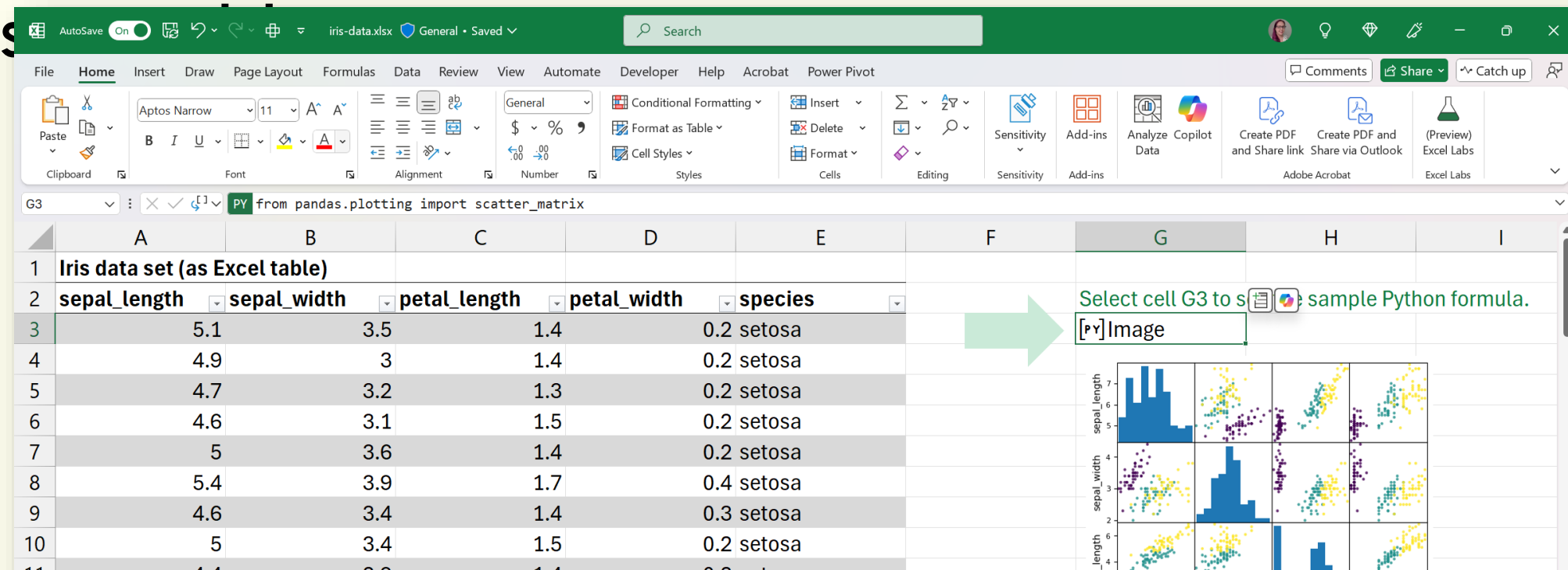
- [Excel incorrectly assumes that the year 1900 is a leap year - Microsoft Learn](#)
 - [Dates and Times — openpyxl documentation](#)
- ISO 8601 is best practice
 - 2025-05-12T23:38:55Z or
20250512T233855Z

tl;dr Be cautious if you are using dates prior to 1900!



Unwrapping an xlsx file

Spreadsheet documents are standardized, change extensions to *.zip and unpack an xlsx/odf

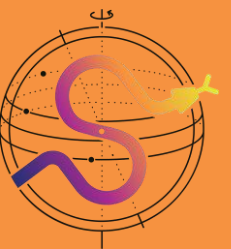


Excel Document Format

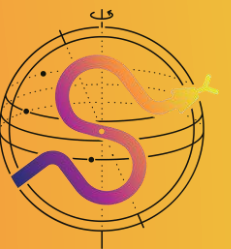
```
\IRIS-DATA-XLSX
├── docMetadata
├── docProps
├── xl
│   ├── drawings
│   │   └── _rels
│   ├── media
│   ├── richData
│   │   └── _rels
│   ├── tables
│   ├── theme
│   ├── worksheets
│   │   └── _rels
│   └── _rels
└── _rels
```

ODS Document Format

```
\IRIS-DATA-ODS
├── Configurations2
│   ├── accelerator
│   ├── floater
│   ├── images
│   │   └── Bitmaps
│   ├── menubar
│   ├── popupmenu
│   ├── progressbar
│   ├── statusbar
│   ├── toolbar
│   └── toolpanel
├── META-INF
├── Pictures
└── Thumbnails
```



2. Python in Excel: Use Python directly in spreadsheets



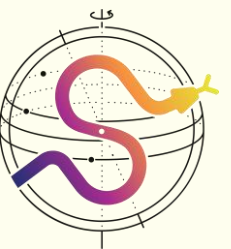
Macros refine data

- Spreadsheets
- Macros
- actions of
- Can have
- VBA or
- was for

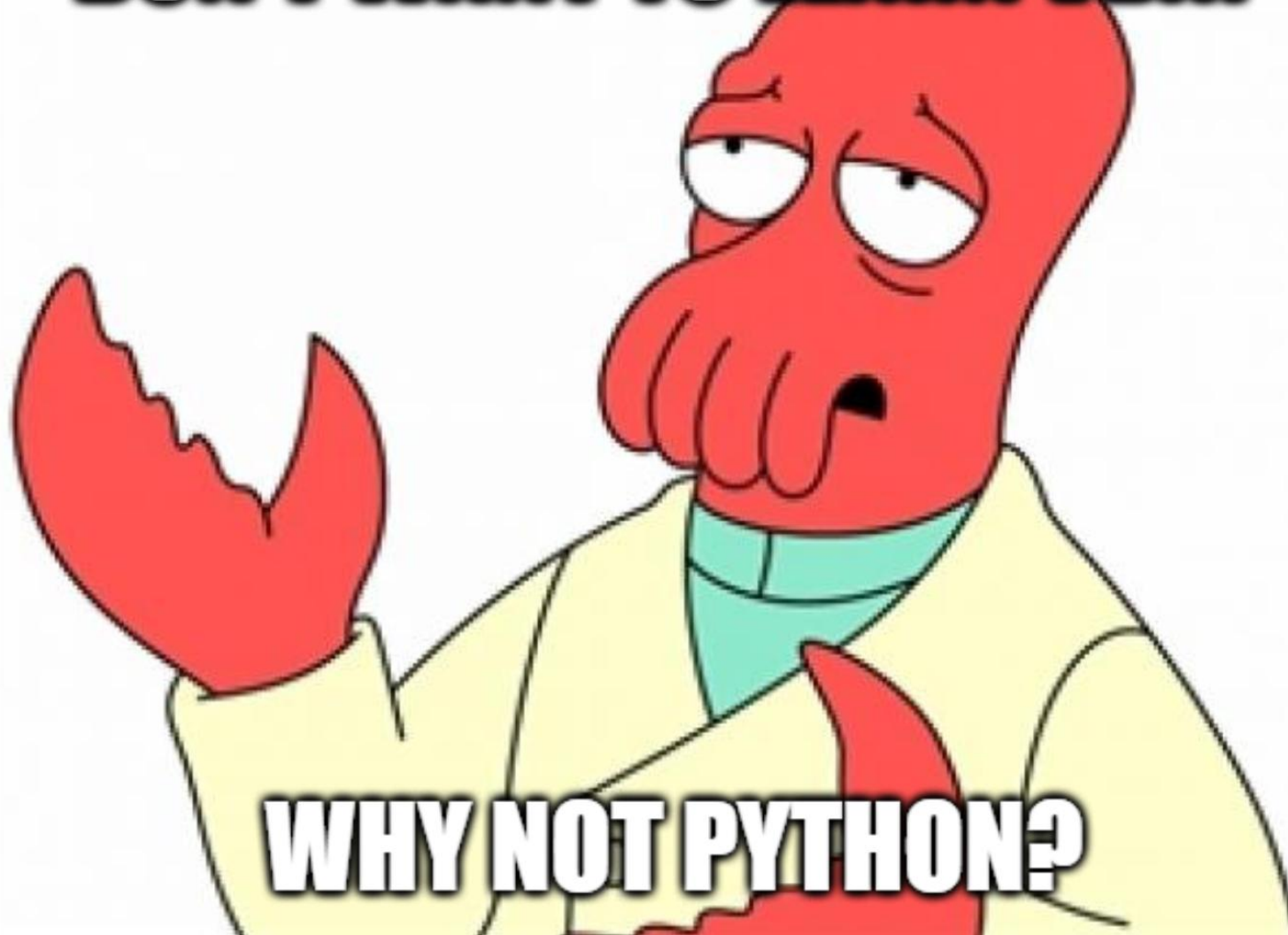


or functions
advanced

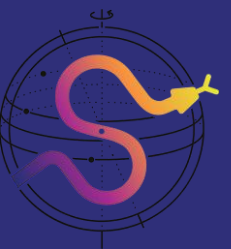
all there



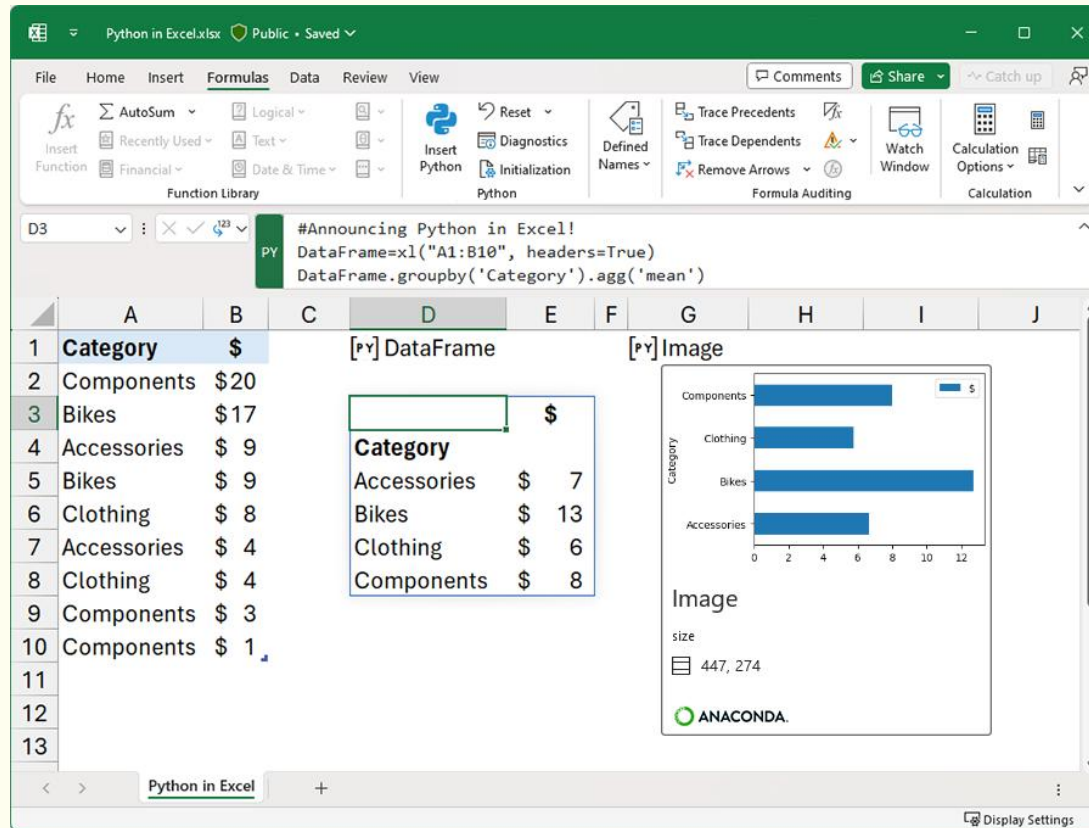
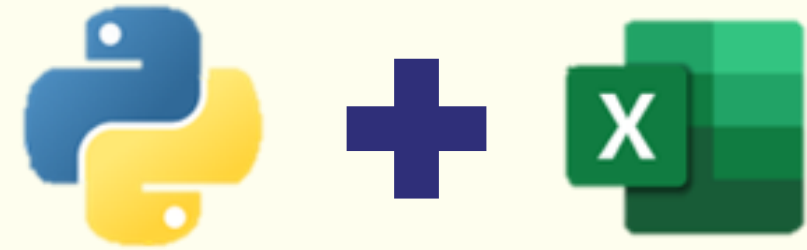
DON'T WANT TO LEARN VBA?



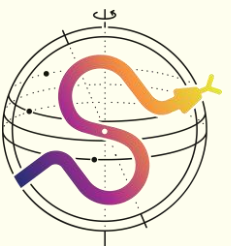
WHY NOT PYTHON?



Python in Excel

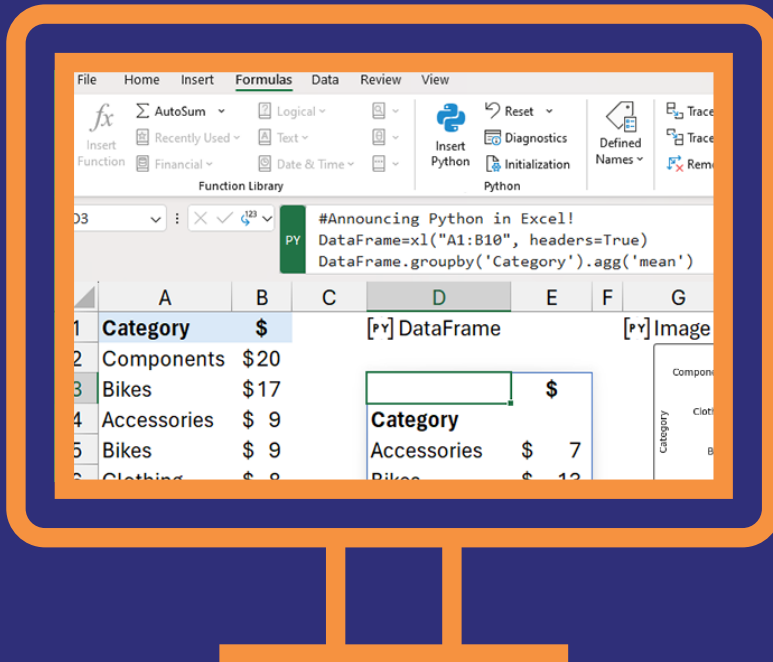


- Newish add-on feature of Excel that allows you run Python directly in the cell
- Add Python to existing workflows and analyses without creating new “projects”

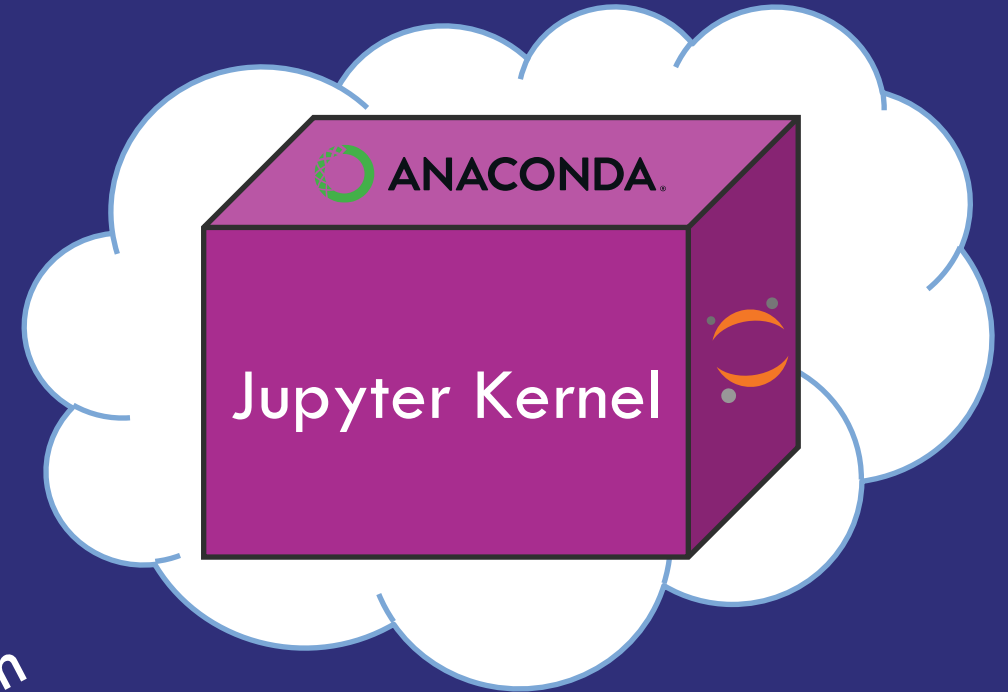


How does it work?

Code cells are turned into a Jupyter notebook cells and run it in a cloud container



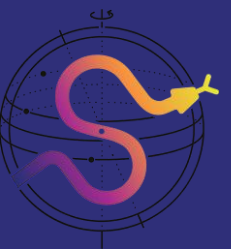
Data is exchanged with container from Excel via x1 package



The container stay online as long as the workbook is open or until a timeout occurs.



TL;DR: 2D Jupyter Notebook running in
an Anaconda distro in the cloud



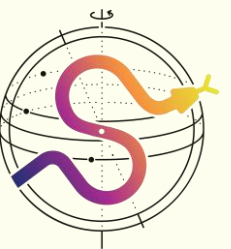
Python in Excel tradeoffs

Advantages

- Highly vetted, secure environments to run code on the cloud
- Interoperates with other 365 experience tools like live editing, share links, and comments
- Copilot integration

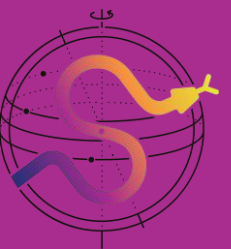
Limitations

- Need network access to use Python and can't access internet in cloud container to get data
- Supports a small subset of the extensive Python package ecosystem
- Current data size limit (~100MB)





Let's look at a demo!



Remote Execution

Python code runs on hypervisor isolated containers built on Azure Container Instances.

Trusted Packaging

The container has Python and a curated set of secured libraries provided by Anaconda.

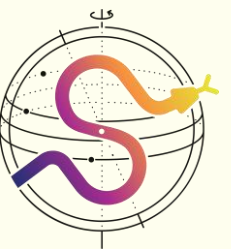
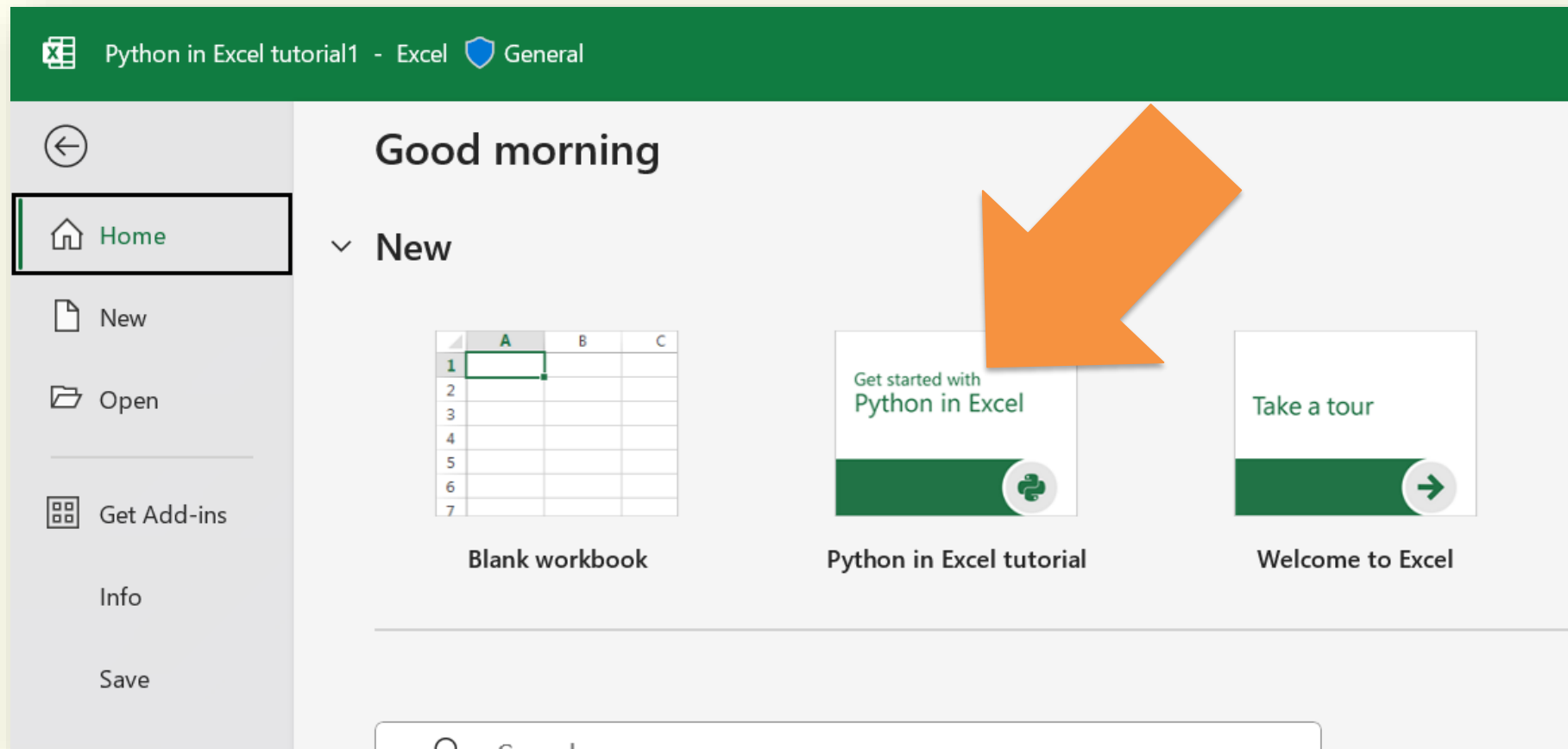
Secure Devices

No local disk access, user token, or network access.

Easy Collaboration

Connected to other tools like OneDrive, Advanced Analysis, sharing links, etc.

Fantastic built-in reference for all things Python in Excel

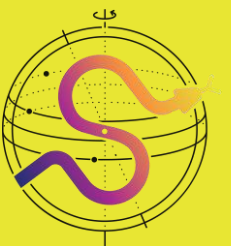


Exercise: Packages in Python in Excel

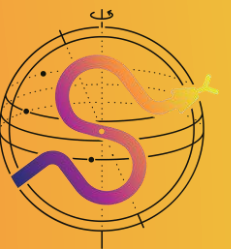
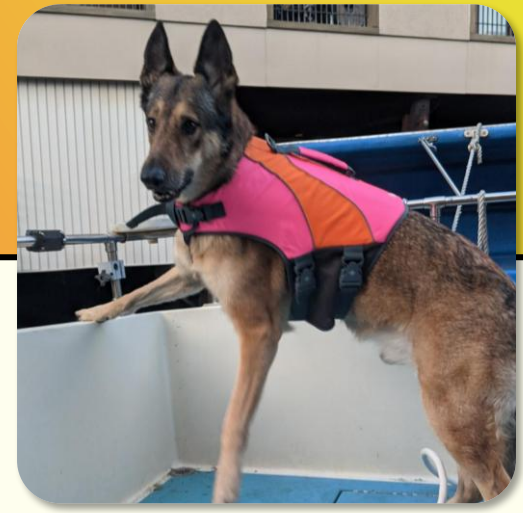
1. In the tutorial repo, download `2_python-in-excel.xlsx` which in Library Index sheet has a list of the packages in the current Python in Excel environment.
2. Pick one of the non-core libraries and in a new sheet create a code sample of using that library in Excel. Checking out the docs for the package is a good start!



aka.ms/scipy-sheets

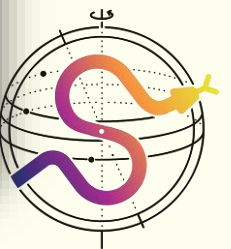
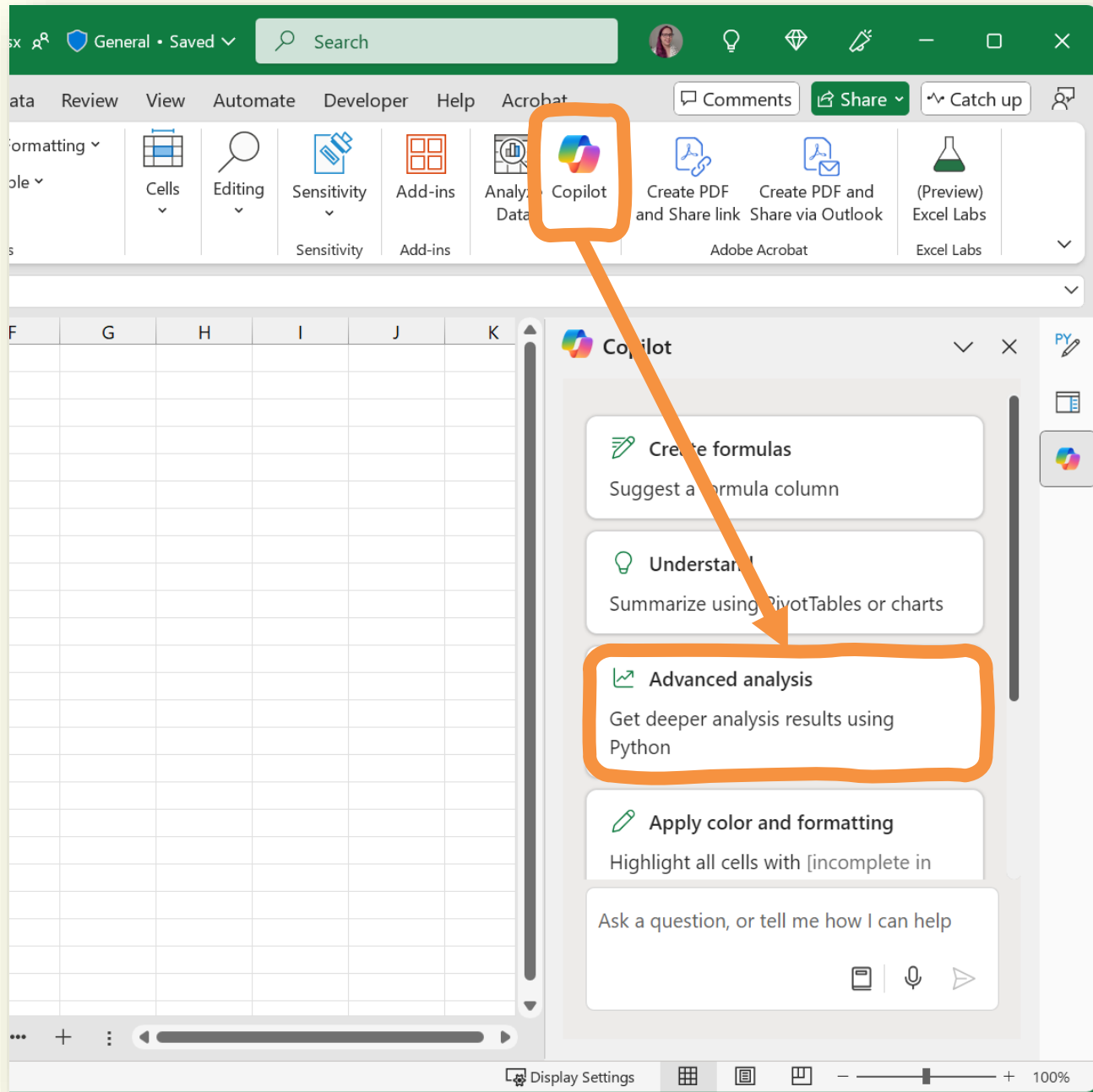


4. Advanced Analysis + Copilot for Python in Excel



Pair program your data analysis

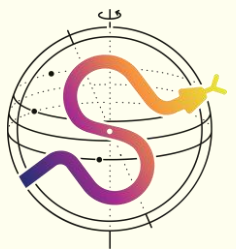
Use **Advanced Analysis**
mode in **Copilot** which can
generate Python code for
you to start with!



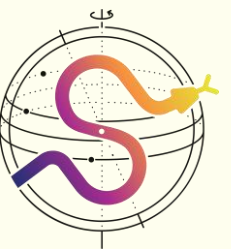
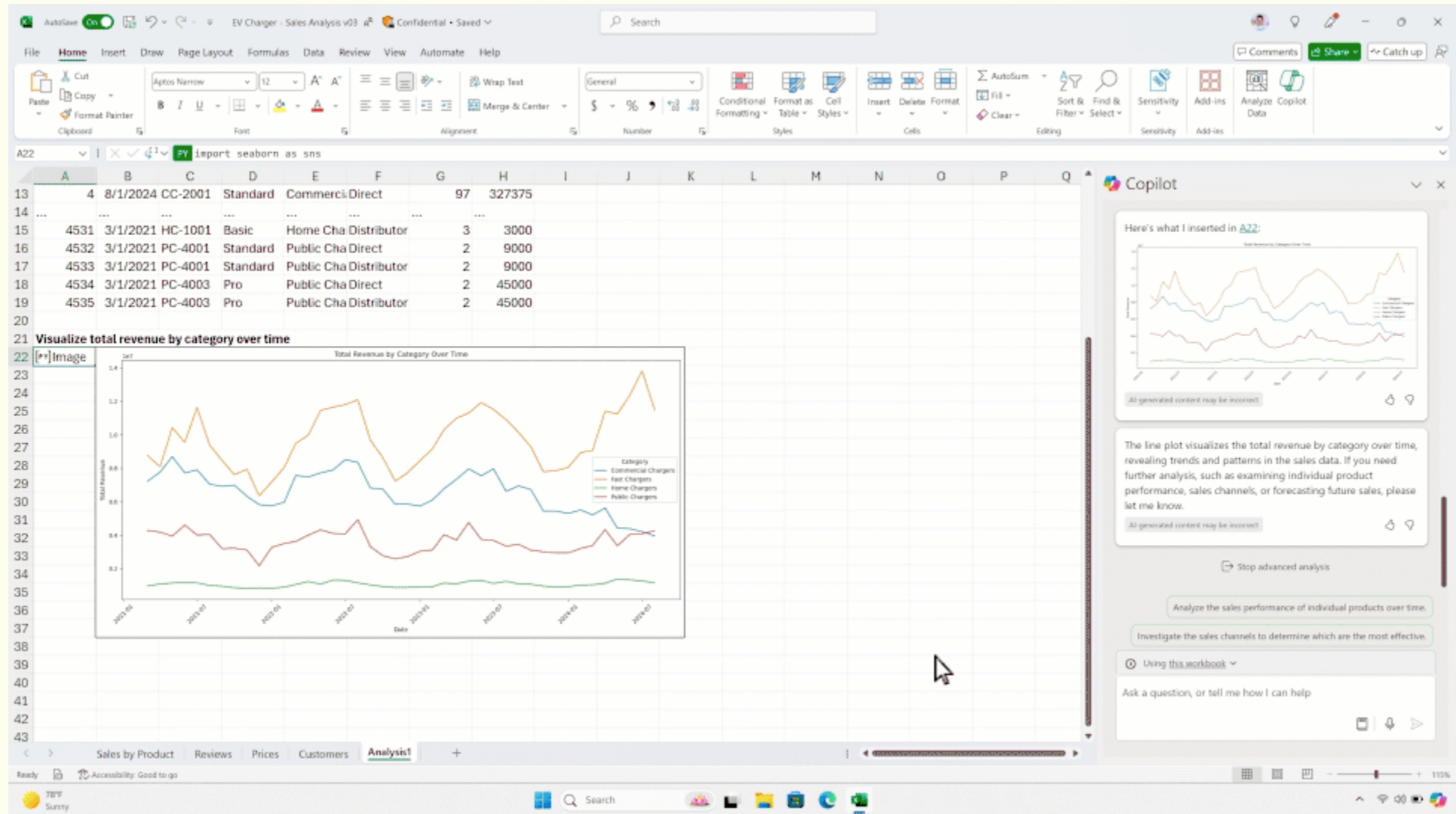
Ask for help starting a project

[illegible]

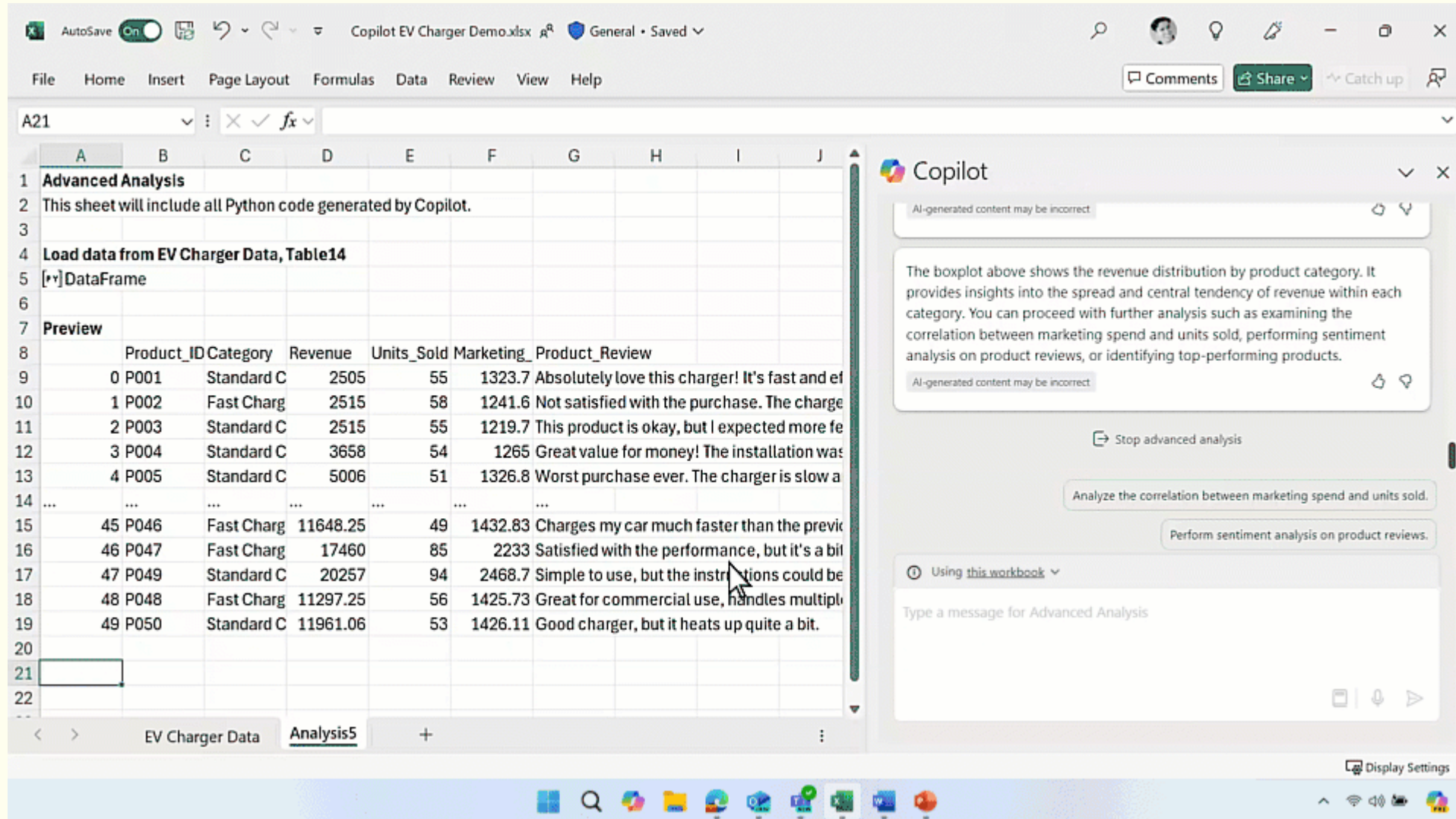
The screenshot shows a chat window with a light gray header containing the Copilot logo and a downward arrow. The chat history includes a user question in a light green bubble: "Can you give me some sample code in Python that runs a monte carlo simulation for a finance application?". The AI response is in a white bubble: "A common Monte Carlo simulation in finance is to estimate the future price of a stock. This involves simulating the stock price over time using random variables to model the uncertainty in price movements. The Geometric Brownian Motion (GBM) model is often used for this purpose. Let's write a Python code snippet to simulate future stock prices using the GBM model." Below the response is a gray warning box: "AI-generated content may be incorrect" and two thumbs-up/down icons. The user's next message is in a white bubble: "Here's what I inserted in A13:". Below this is a gray box with two prompts: "Ask for the data or provide a specific query to analyze." and "Request a specific type of analysis or visualization." The bottom of the screen shows a text input field with the placeholder "Type a message for Advanced analysis" and three icons: a document, a microphone, and a send arrow.



Ask for help forecasting and extending visuals



Ask for help creating visualizations...

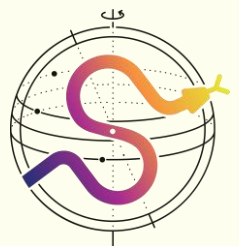


The screenshot displays the Microsoft Excel interface with the Copilot AI assistant integrated on the right side. The Excel window is titled 'Copilot EV Charger Demo.xlsx' and shows a worksheet named 'Analysis5'. The data in the worksheet is as follows:

Product_ID	Category	Revenue	Units_Sold	Marketing_	Product_Review
0 P001	Standard C	2505	55	1323.7	Absolutely love this charger! It's fast and ef
1 P002	Fast Charg	2515	58	1241.6	Not satisfied with the purchase. The charge
2 P003	Standard C	2515	55	1219.7	This product is okay, but I expected more fe
3 P004	Standard C	3658	54	1265	Great value for money! The installation was
4 P005	Standard C	5006	51	1326.8	Worst purchase ever. The charger is slow a
...
45 P046	Fast Charg	11648.25	49	1432.83	Charges my car much faster than the previ
46 P047	Fast Charg	17460	85	2233	Satisfied with the performance, but it's a bit
47 P049	Standard C	20257	94	2468.7	Simple to use, but the instr
48 P048	Fast Charg	11297.25	56	1425.73	Great for commercial use, handles multipl
49 P050	Standard C	11961.06	53	1426.11	Good charger, but it heats up quite a bit.

The Copilot sidebar on the right provides the following content:

- Copilot** header with a warning: 'AI-generated content may be incorrect'.
- A text box explaining: 'The boxplot above shows the revenue distribution by product category. It provides insights into the spread and central tendency of revenue within each category. You can proceed with further analysis such as examining the correlation between marketing spend and units sold, performing sentiment analysis on product reviews, or identifying top-performing products.'
- A button: 'Stop advanced analysis'.
- Two suggested actions: 'Analyze the correlation between marketing spend and units sold.' and 'Perform sentiment analysis on product reviews.'
- A section titled 'Using this workbook' with a text input field: 'Type a message for Advanced Analysis'.



...even with pretty vague directions

Network Analysis.xlsx

File Home Insert Page Layout Formulas Data Review View Help

A8 `PY import networkx as nx`

1 **Advanced Analysis**
2 This sheet will include all Python code generated by Copilot.
3
4 **Load data from Network Data, Table1**
5 `[*] DataFrame`
6
7 **Plot network graphs using nodes and edges data**
8 `[*] Image`

Network 1 Network 2 Network 3

Copilot

Using Python, can you create a set of network graphs using my data?

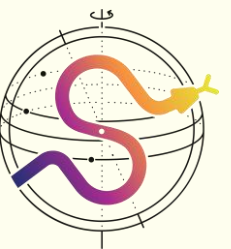
The network graphs have been successfully plotted using the nodes and edges data. You can now proceed with further analysis or visualization. What would you like to do next?

AI-generated content may be incorrect

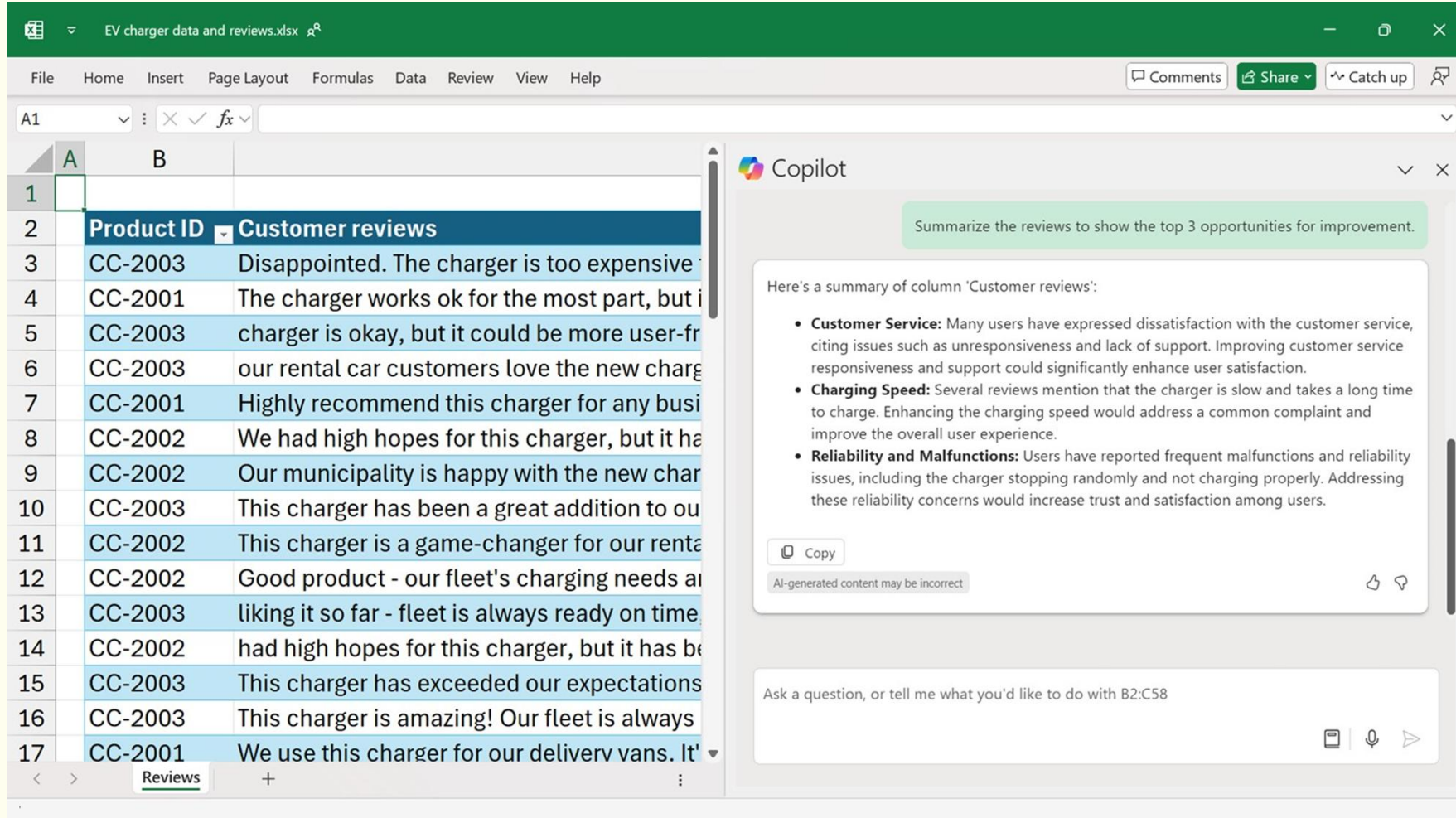
Here's what I inserted

Network 1 Network 2 Network 3

Type a message for Advanced Analysis



Ask for summaries of data



The screenshot shows a Microsoft Excel window titled "EV charger data and reviews.xlsx". The spreadsheet has two columns: "Product ID" and "Customer reviews". The data is as follows:

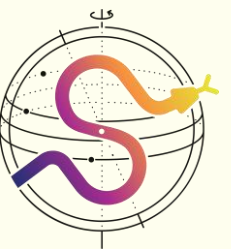
Product ID	Customer reviews
CC-2003	Disappointed. The charger is too expensive
CC-2001	The charger works ok for the most part, but i
CC-2003	charger is okay, but it could be more user-fr
CC-2003	our rental car customers love the new charge
CC-2001	Highly recommend this charger for any busi
CC-2002	We had high hopes for this charger, but it ha
CC-2002	Our municipality is happy with the new char
CC-2003	This charger has been a great addition to ou
CC-2002	This charger is a game-changer for our renta
CC-2002	Good product - our fleet's charging needs a
CC-2003	liking it so far - fleet is always ready on time
CC-2002	had high hopes for this charger, but it has be
CC-2003	This charger has exceeded our expectations
CC-2003	This charger is amazing! Our fleet is always
CC-2001	We use this charger for our delivery vans. It'

The Copilot AI assistant is open on the right side of the Excel window. It has received the prompt: "Summarize the reviews to show the top 3 opportunities for improvement." and has generated the following summary:

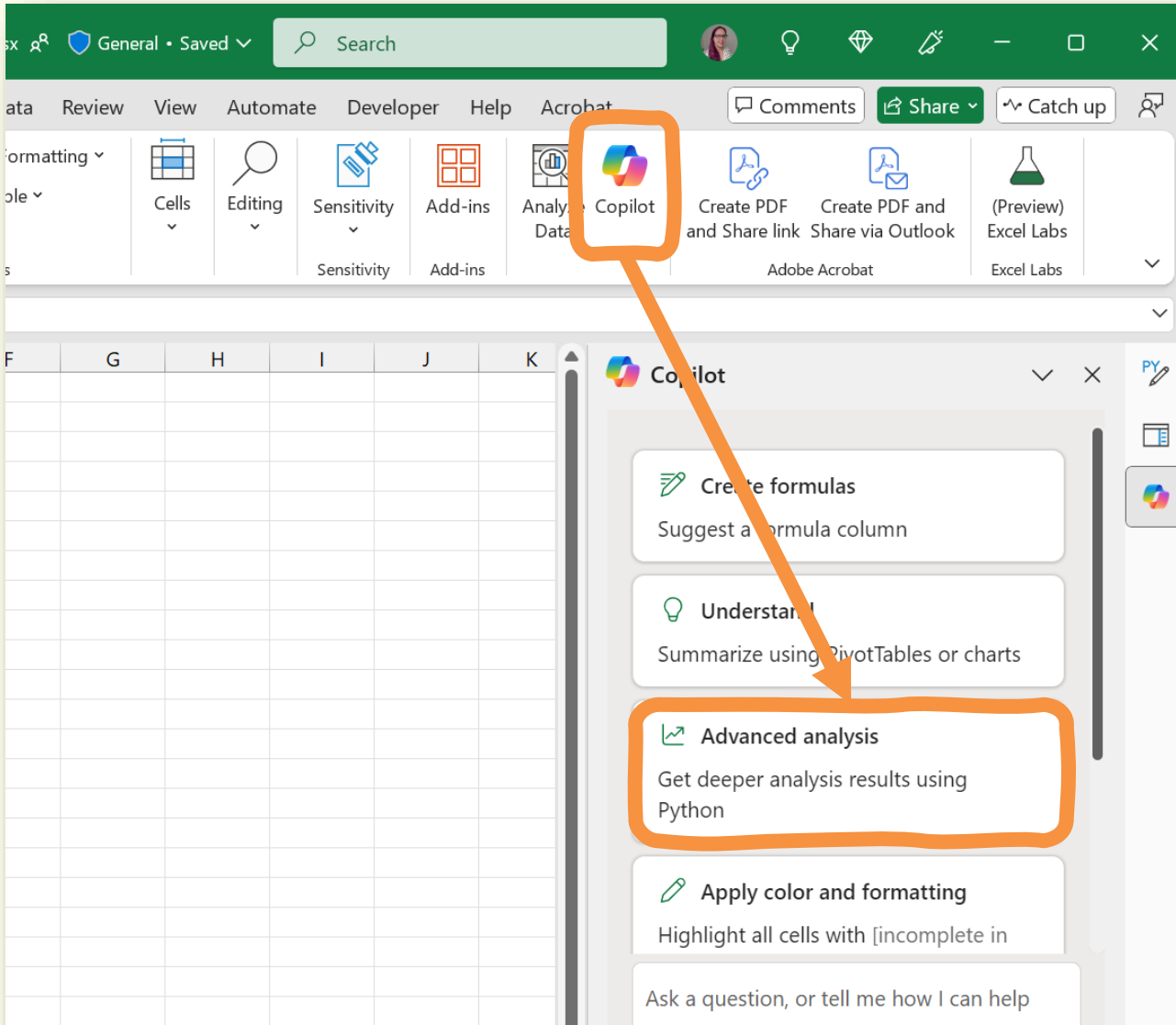
Here's a summary of column 'Customer reviews':

- **Customer Service:** Many users have expressed dissatisfaction with the customer service, citing issues such as unresponsiveness and lack of support. Improving customer service responsiveness and support could significantly enhance user satisfaction.
- **Charging Speed:** Several reviews mention that the charger is slow and takes a long time to charge. Enhancing the charging speed would address a common complaint and improve the overall user experience.
- **Reliability and Malfunctions:** Users have reported frequent malfunctions and reliability issues, including the charger stopping randomly and not charging properly. Addressing these reliability concerns would increase trust and satisfaction among users.

Below the summary is a "Copy" button and a disclaimer: "AI-generated content may be incorrect". At the bottom of the Copilot panel, there is a text input field with the placeholder "Ask a question, or tell me what you'd like to do with B2:C58" and icons for document, voice, and send.

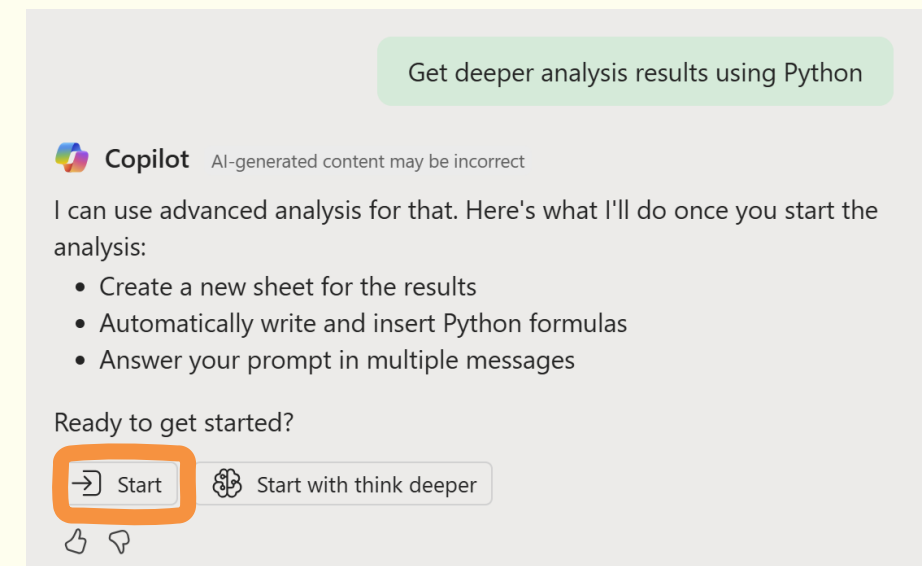


How do I start Advanced Analysis?



Or start with a prompt like:

- "Analyze this data using Python"
- "Run advanced analysis using Python"



The Analysis sheet:

Analysis Sheet

This sheet will include all the Python formulas generated by Copilot.

Load data from Sheet1, OfficeForms.Table

[Py] DataFrame

Preview

	Id	Start time	Completion time	Email	Name	What are you using for AI?	Are you using AI in your work?
0	1	45792.5872569444	45792.5874189815	anonymous	None	Learning/Exploring	Yes, I'm using it as a tool
1	2	45792.5874768519	45792.5913657407	anonymous	None	Data Science	Yes, I'm using it as a tool
2	3	45792.5918981481	45792.5923726852	anonymous	None	Learning/Exploring	Yes, I'm building applications
3	4	45792.5934722222	45792.5938078704	anonymous	None	Data Science	Yes, I'm building applications
4	5	45792.5950347222	45792.5972800926	anonymous	None	Data Science	Yes, I'm building applications
...
233	234	45794.526087963	45794.543900463	anonymous	None	cloud services	A little, I occasionally use it
234	235	anonymous	None	Automation	Yes, I'm building applications
235	236	anonymous	None	Data Science	Yes, I'm using it as a tool
236	237	anonymous	None	Data Science	Yes, I'm using it as a tool
237	238	anonymous	None	Data Science	A little, I occasionally use it

Visualize main categories for Python use, cloud, and AI adoption

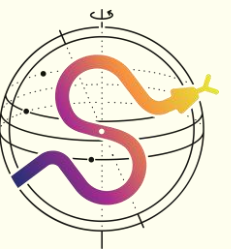
[Py] Image

Python Use Cases

238 rows x 8 columns

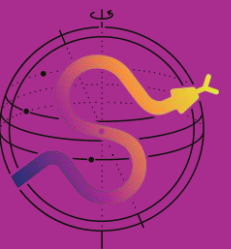
	Id	Start time	Completion time	Email	Name
0	1	45792.5872569444	45792.5874189815	anonymous	0
1	2	45792.5874768519	45792.5913657407	anonymous	0
2	3	45792.5918981481	45792.5923726852	anonymous	0
3	4	45792.5934722222	45792.5938078704	anonymous	0
4	5	45792.5950347222	45792.5972800926	anonymous	0
...
233	234	45794.526087963	45794.543900463	anonymous	0

Message Copilot





Let's look at a demo!

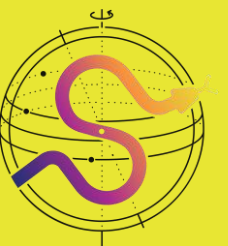


Exercise: Discover your Data

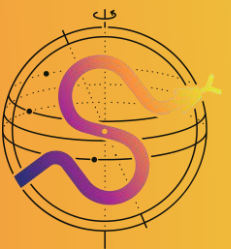
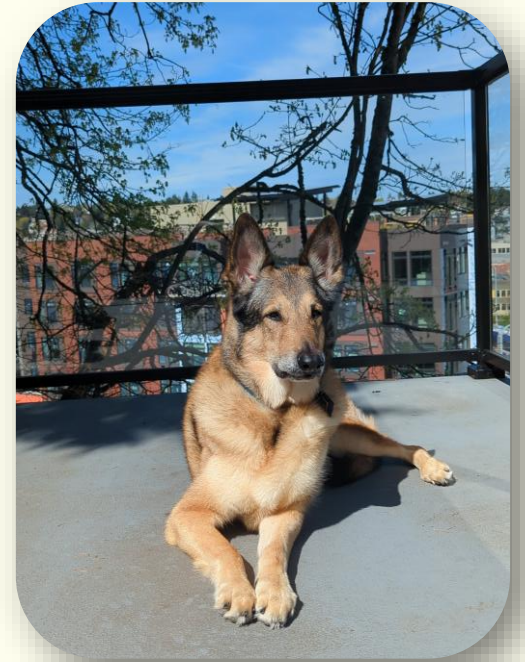
1. Using a data set of your own or one provided in the tutorial repo, use Advanced Analysis to help you make some sort of visualization of the data.
2. Then use the Copilot chat to learn more about the data by trying out some of it's suggested prompts.



aka.ms/scipy-sheets

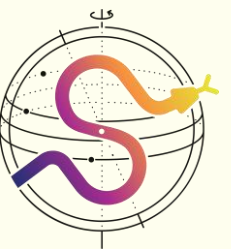


6. Spreadsheets with pandas



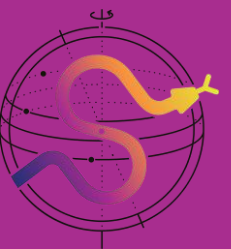
Pandas ↔ Spreadsheet terms

Pandas	Spreadsheet
DataFrame	worksheet
Series	column
Index	row headings
row	row
NaN	empty cell





Let's look at some code!

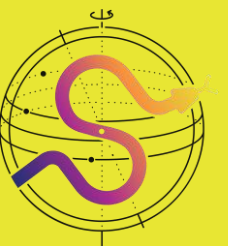


Exercise: Pandas and Sheets

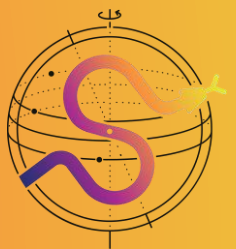
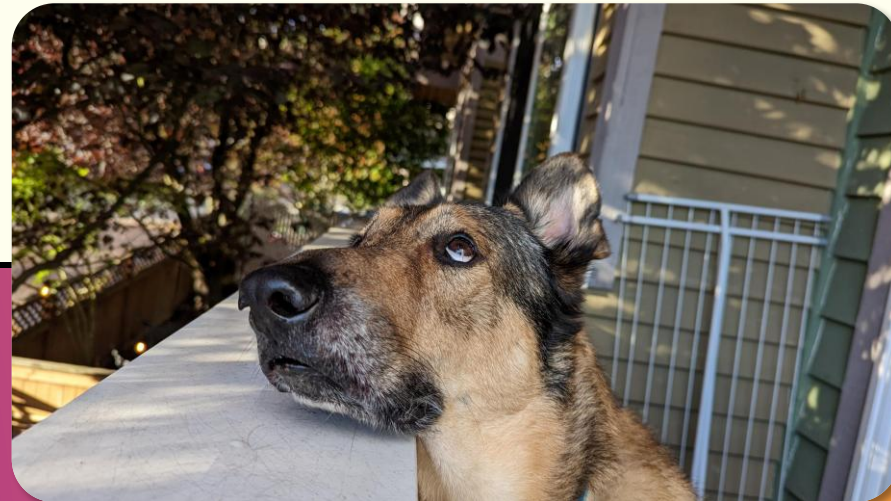
1. Take the `fish.csv` sample data and use pandas to load into Python (Jupyter notebook or script)
2. Clean the data in a way you think would be helpful (maybe parse columns, change value types, etc.). You can also ask Copilot in VS Code for suggestions.
3. Export the cleaned data back to a spreadsheet with pandas.



aka.ms/scipy-sheets

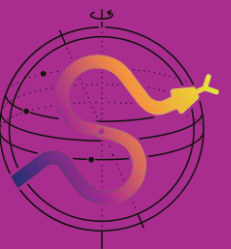


7. Build spreadsheets with openpyxl





Let's look at some code!

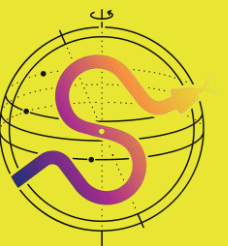


Exercise: Plots

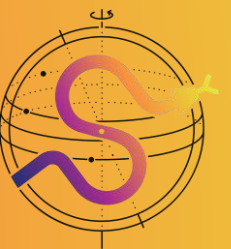
1. Use your spreadsheet viewer of choice to add a plot to the spreadsheet from the pandas exercise.
2. Parse the spreadsheet with `openpyxl` and extract the image of the plot with Python.

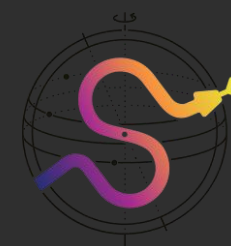
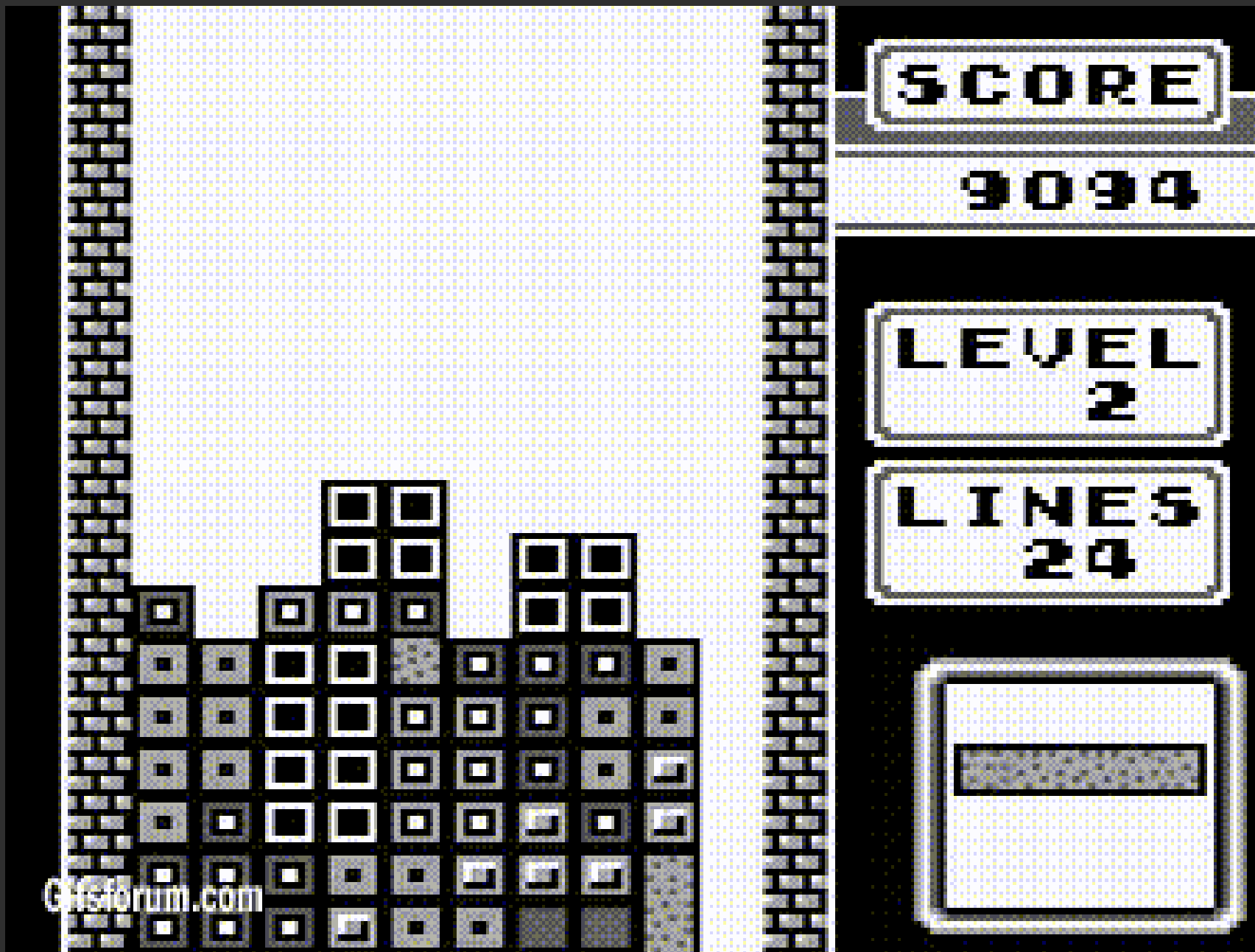


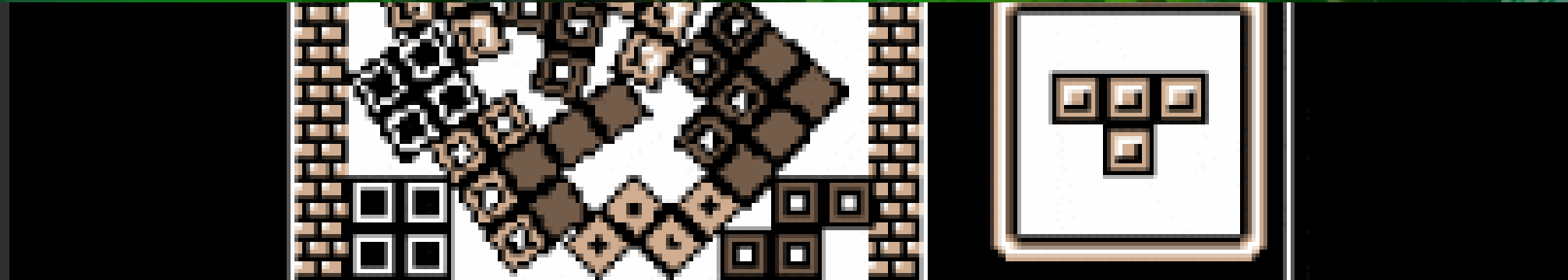
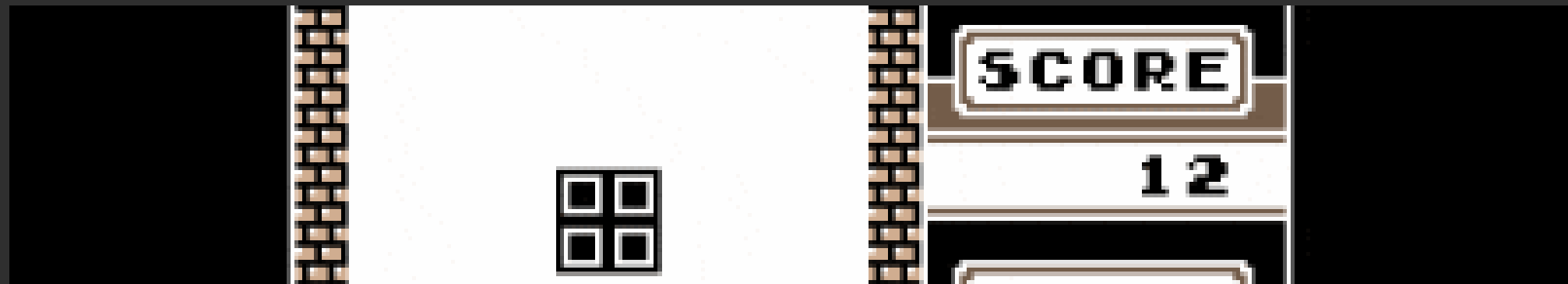
aka.ms/scipy-sheets



=SUM(everything)







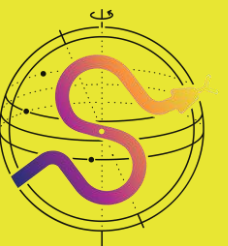
Exercise: Excel E-Sports challenge

In the tutorial repo, download/open the file `excel-esports-tetris.xlsx` for instructions.

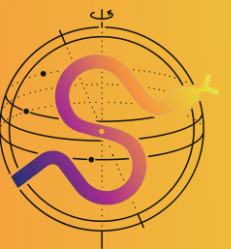
You can either work in Python and export your answers in a spreadsheet or use Python in Excel/Anaconda Code to solve it. Anyone who finishes will get some limited-edition swag!



aka.ms/scipy-sheets

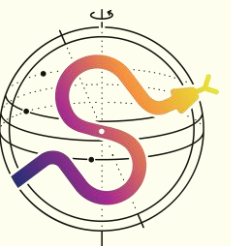


Wrap-up



- Load and work with spreadsheets in Python
- Run Python directly in Excel
- Use AI tools like Advanced Analysis in Excel
- Use Anaconda Toolbox and Code in Excel

What did we just do?



Find Sarah at:

@crazy4pi314 | sckaiser.com

Find Jim at:

@Jim22k on GitHub

Come hangout on our Discord:

<https://aka.ms/azureai/discord>

Ask the PyEx devs questions on
GitHub:

<https://github.com/microsoft/pythontesting/issues>

aka.ms/scipy-sheets



Thank You!

