STRINGFEST
ANALYTICS

Synthetic data in Excel

A walkthrough using Faker in Python in Excel.

Swipe →

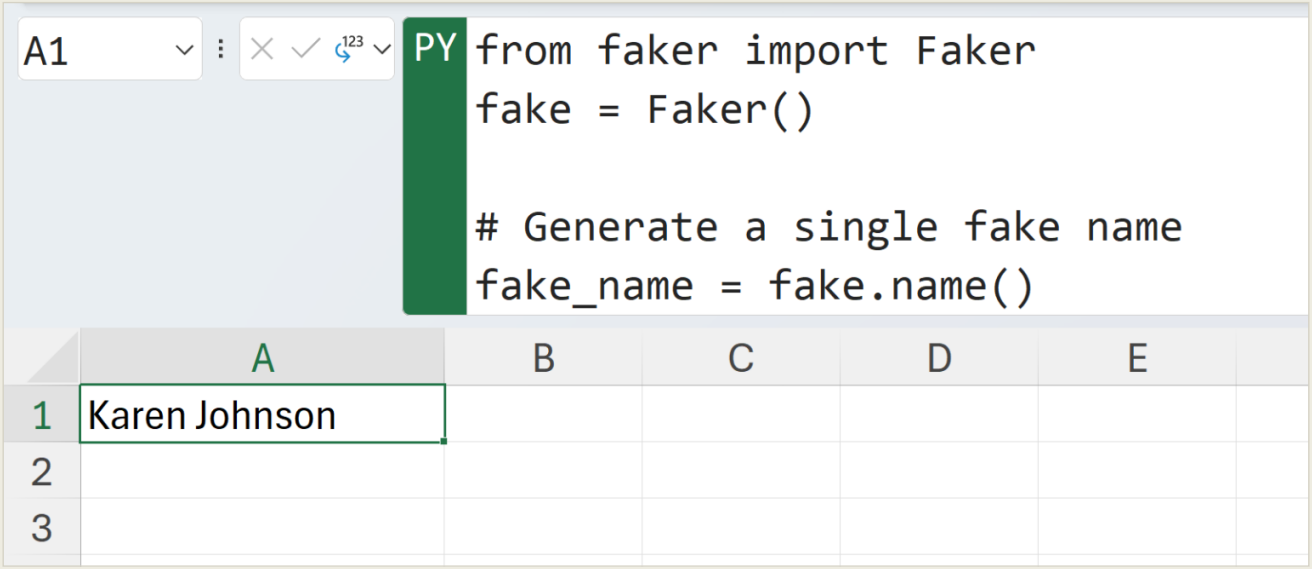
Three situations where this comes up.

1. Building a demo or training file before you have access to the real data.
2. Stress-testing a workbook to see what breaks once row counts get large.
3. Feeding AI tools (Copilot, Claude, anything) without pasting real customer or employee information into a prompt.

Faker, the Python library, solves all three. It ships inside Python in Excel, so you don't need to install anything.

Step 1. Import Faker and create an instance.

The first line brings the Faker class into your session. The second creates an instance you'll call methods on. fake.name() returns one randomly-generated name.

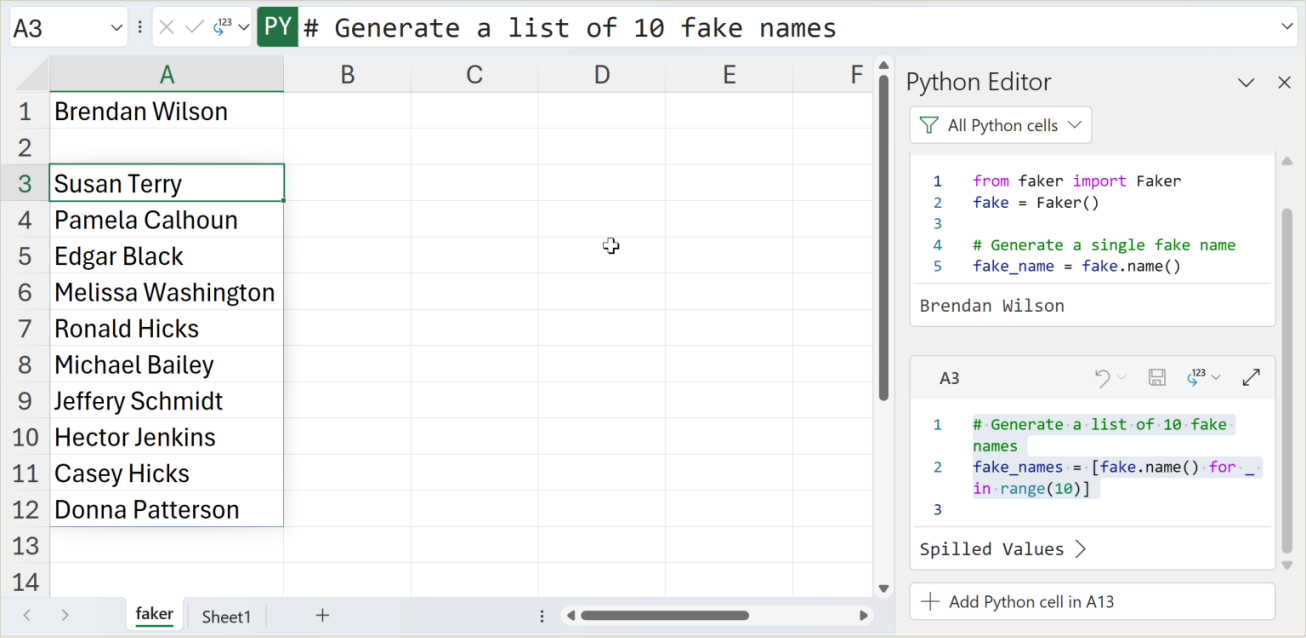


Two lines of setup, then a method call. That's it.

STEP 2 OF 5

Step 2. Generate many names with list comprehension.

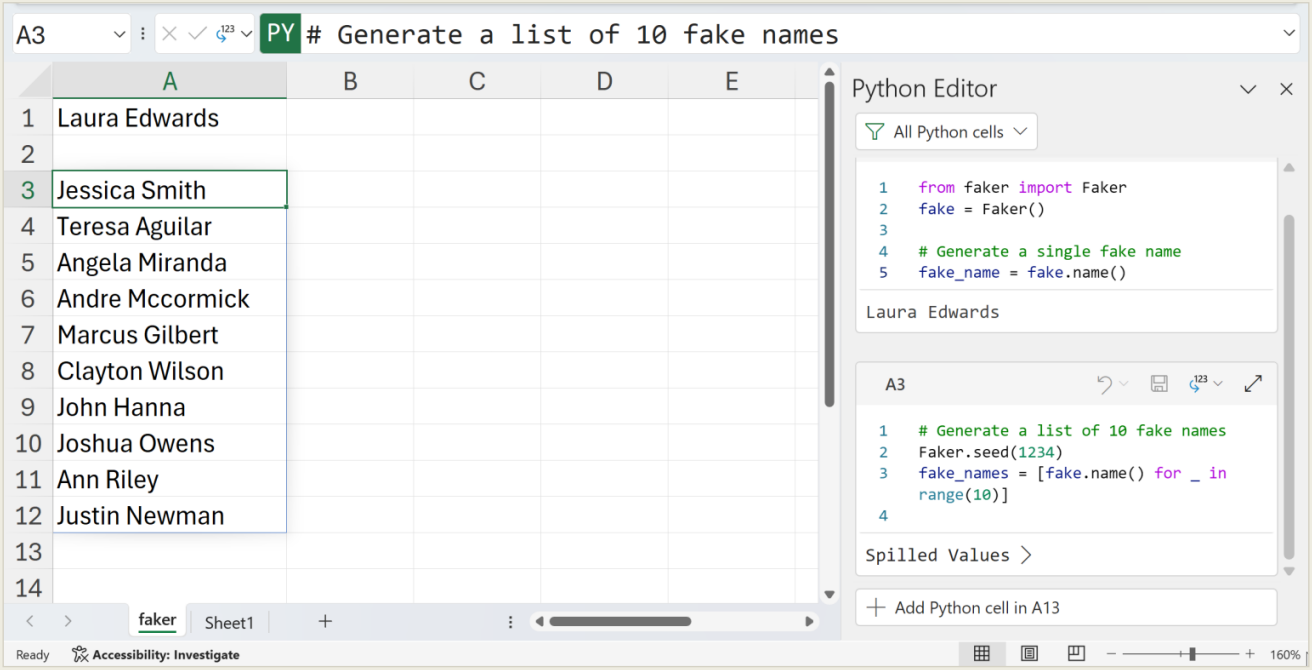
One name is rarely enough. A list comprehension calls fake.name() repeatedly inside a single expression, and range(10) controls how many you want.



Re-run the cell and you'll get ten different names each time.

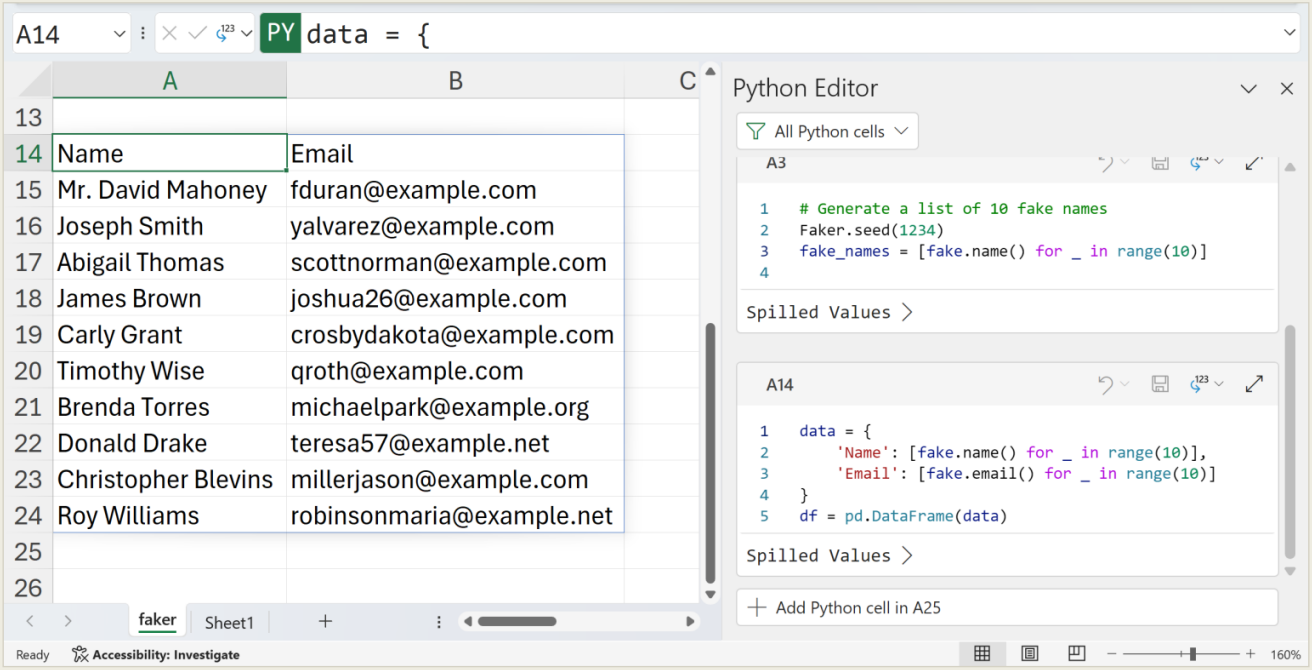
Step 3. Lock the output with a random seed.

Faker.seed(1234) tells Faker to use the same starting point every time. Now the same ten names come back on every run, which matters when you want demos or training files to stay stable across sessions. Pick any integer and stick with it.



Step 4. Combine multiple columns into a DataFrame.

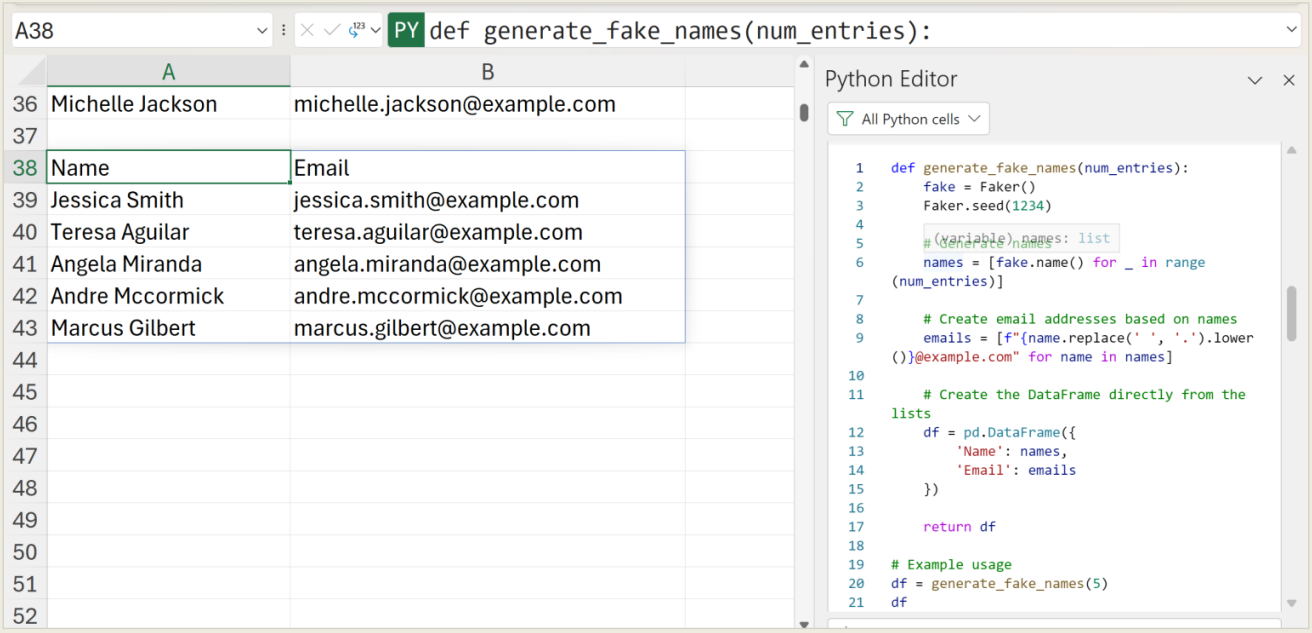
When you need a table instead of a list, build a Python dictionary where each key is a column name and each value is a list of fake data. Pass it to pd.DataFrame() and you get a structured table that spills into Excel like any other Python output.



STEP 5 OF 5

Step 5. Wrap it in a function for reuse.

If you'll regenerate this dataset at different sizes, package the logic into a function. The Faker instance and the seed live inside, so generate_fake_names(5) and generate_fake_names(500) both return reproducible output.



Call it anywhere in the workbook with any row count.

ALTERNATIVE

Or: ask Copilot to write the script for you.

If you'd rather skip writing the Python yourself, Copilot in Excel can generate the script. Describe the columns you want, the row count, any distributions, and a random seed. Copilot returns code you can paste straight into a PY() cell.

EXAMPLE PROMPT

Generate a Python script for a 500-row employee performance dataset. Columns: Employee ID, Department, Performance Score (normal, mean 70, SD 10), Salary (lognormal). Departments: Sales, HR, IT, Marketing, Finance. Set Faker.seed(1234).

WHERE TO GO NEXT

That's the full Faker workflow.

There's more on the blog: setting the locale for international data, when to swap Faker for NumPy on quantitative variables, and a longer Copilot prompt for bigger datasets.

READ THE FULL POST

stringfestanalytics.com/blog