

Order.co



www.fmworldcup.com



Getting Started with Python in Microsoft Excel

Second Edition



Follow Financial Modeling World Cup on LinkedIn

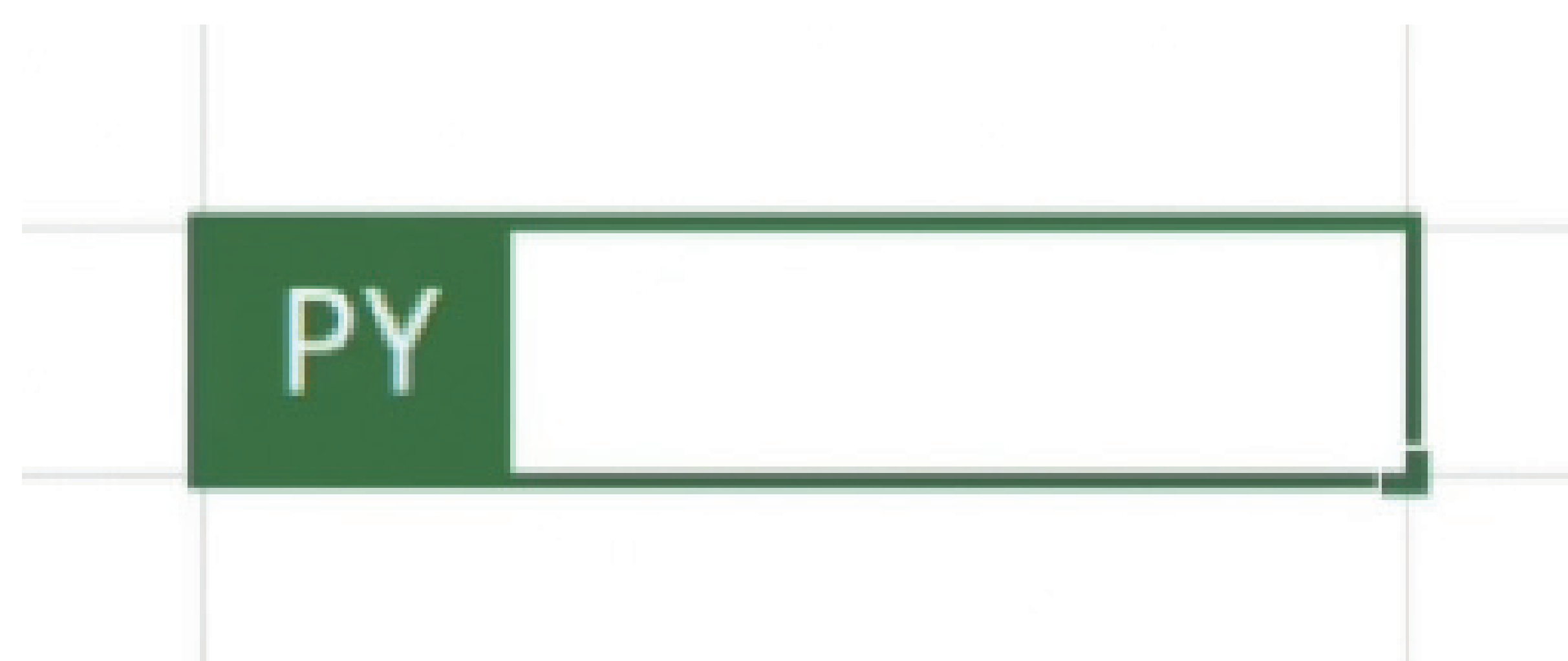


Python availability in Excel introduces a fresh realm of possibilities for data analysis that was once primarily accessible to data scientists and developers. Now, within the comfort of your well-known spreadsheet environment, you have the ability to tap into the capabilities of Python.

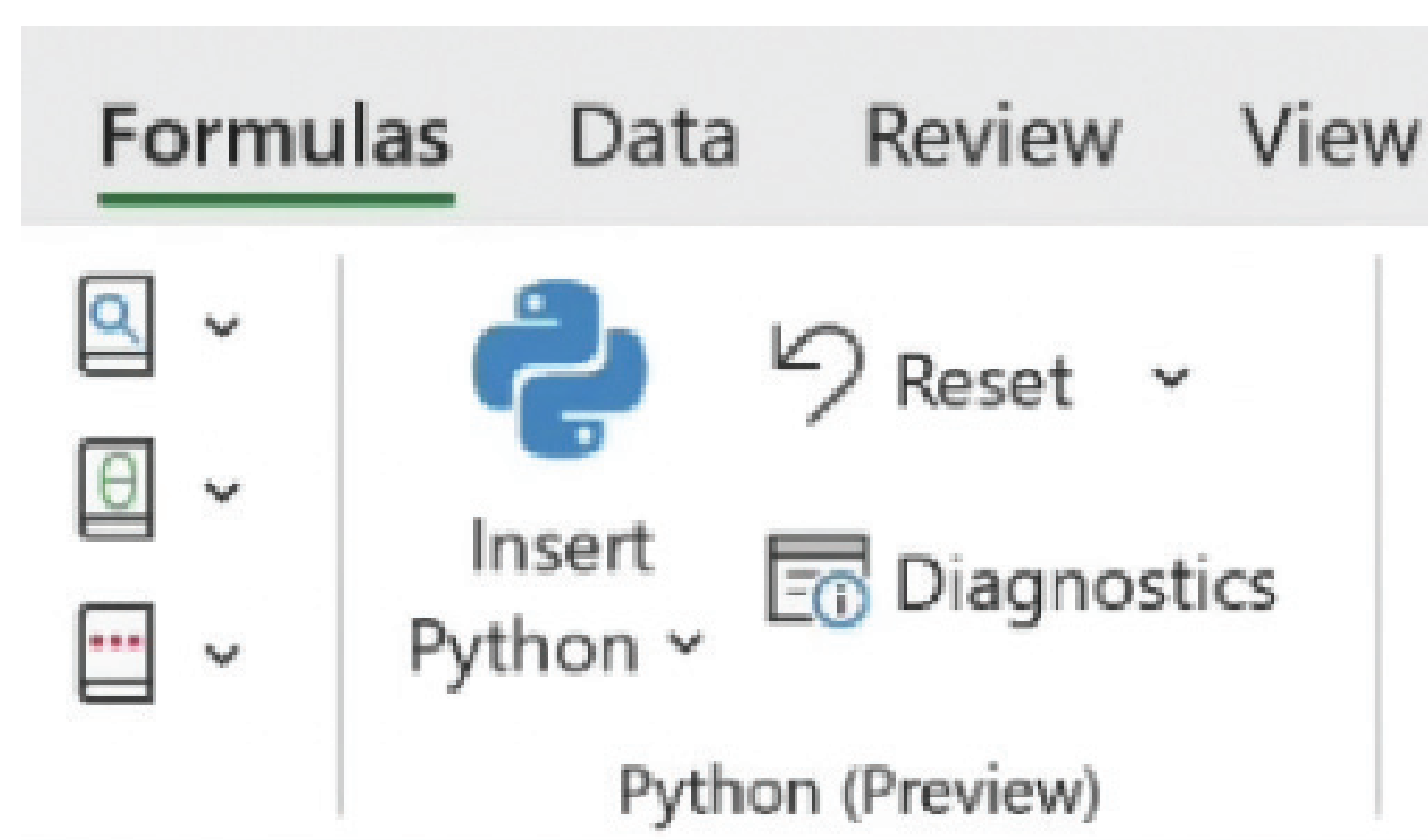
But how to start using Python in Excel and what functionalities does it offer?

Just type =PY()

Could it be simpler? Write =PY() and get started! After the formula the Excel cell will change and you can enter your Python code and see the visualizations or calculations right there in your Excel spreadsheet!



You can also find Python under the formulas section in Excel!



Ok, but what can I actually create with Python in Excel?

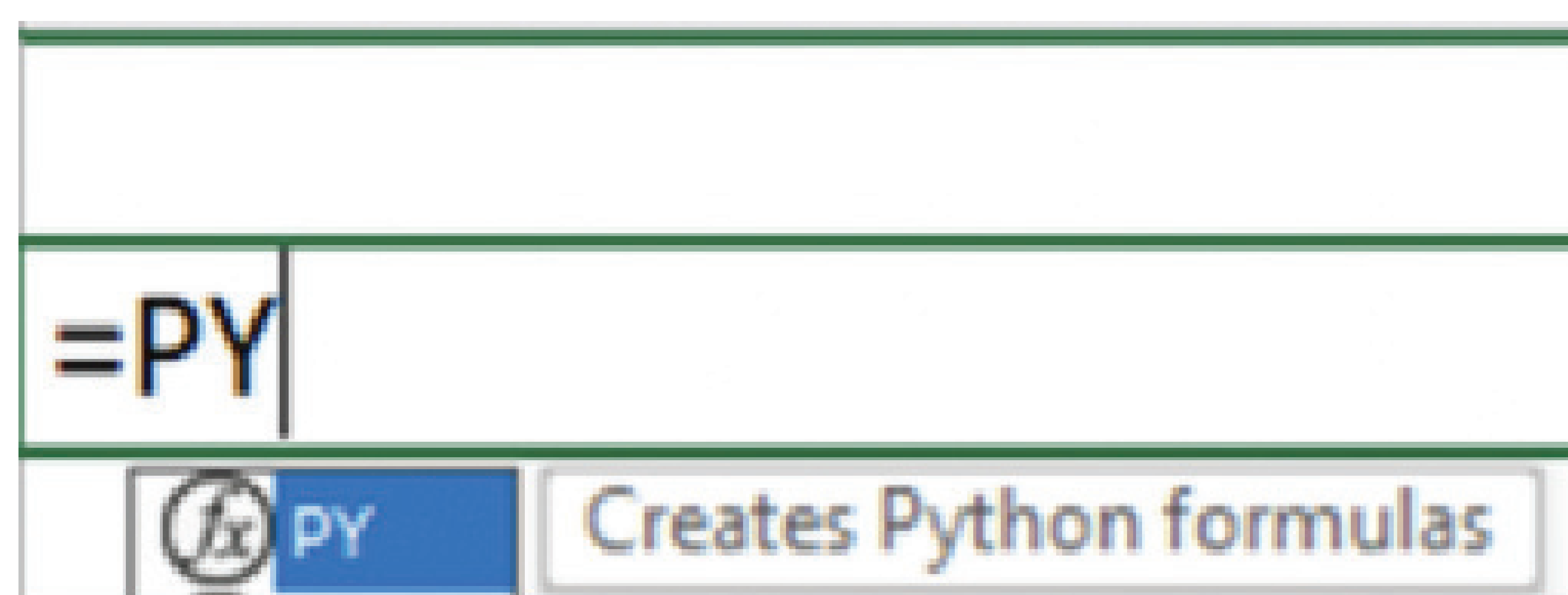
1. Creating a Data Frame

Python code runs in a controlled cloud container with limitations on its capabilities.

Python code lacks internet access and the ability to reach files on your local computer. The Excel workbook defines the boundaries of Python's reach.

With these constraints in mind, it's not surprising that creating a pandas DataFrame, for instance, often serves as the initial step when incorporating Python within Excel.

Constructing a pandas DataFrame from an Excel table is a clear process. The newly introduced PY() function facilitates the construction of Python formulas containing the necessary Python code.



2. Writing the Code

```
PY xl("A2:D98", headers=True)
```

So, here's how the code works:

1 The xl() function is employed to extract data from Excel, generating a pandas DataFrame.

2 The initial parameter indicates the data's location, which in this instance refers to the Excel table within the range A2:D98.

3 The subsequent parameter informs the xl() function about the presence of headers (column names) within the data.

4 You can assign a variable name to your DataFrame like you were using named ranges or LET() function in regular Excel. You can subsequently access the DataFrame in subsequent Python formulas using the same variable name.

Press Ctrl+Enter to execute the Python formula. It might take a moment for the code to work. If your code is error-free, here's what will appear:

 DataFrame

Hover the mouse on the card icon to see a preview of the data within the DataFrame:

SalesOrderNumber	SalesOrderLineNumber	OrderDate
SO43697	1	12/29/2010
SO43698	1	12/29/2010
SO43699	1	12/29/2010
SO43700	1	12/29/2010
SO43701	1	12/29/2010
...
SO75122	1	1/28/2014
SO75122	2	1/28/2014
SO75123	1	1/28/2014
SO75123	2	1/28/2014
SO75123	3	1/28/2014

Source: [Anaconda.com](https://www.anaconda.com/)

The preview shows the first 5 and last 5 rows of the data.

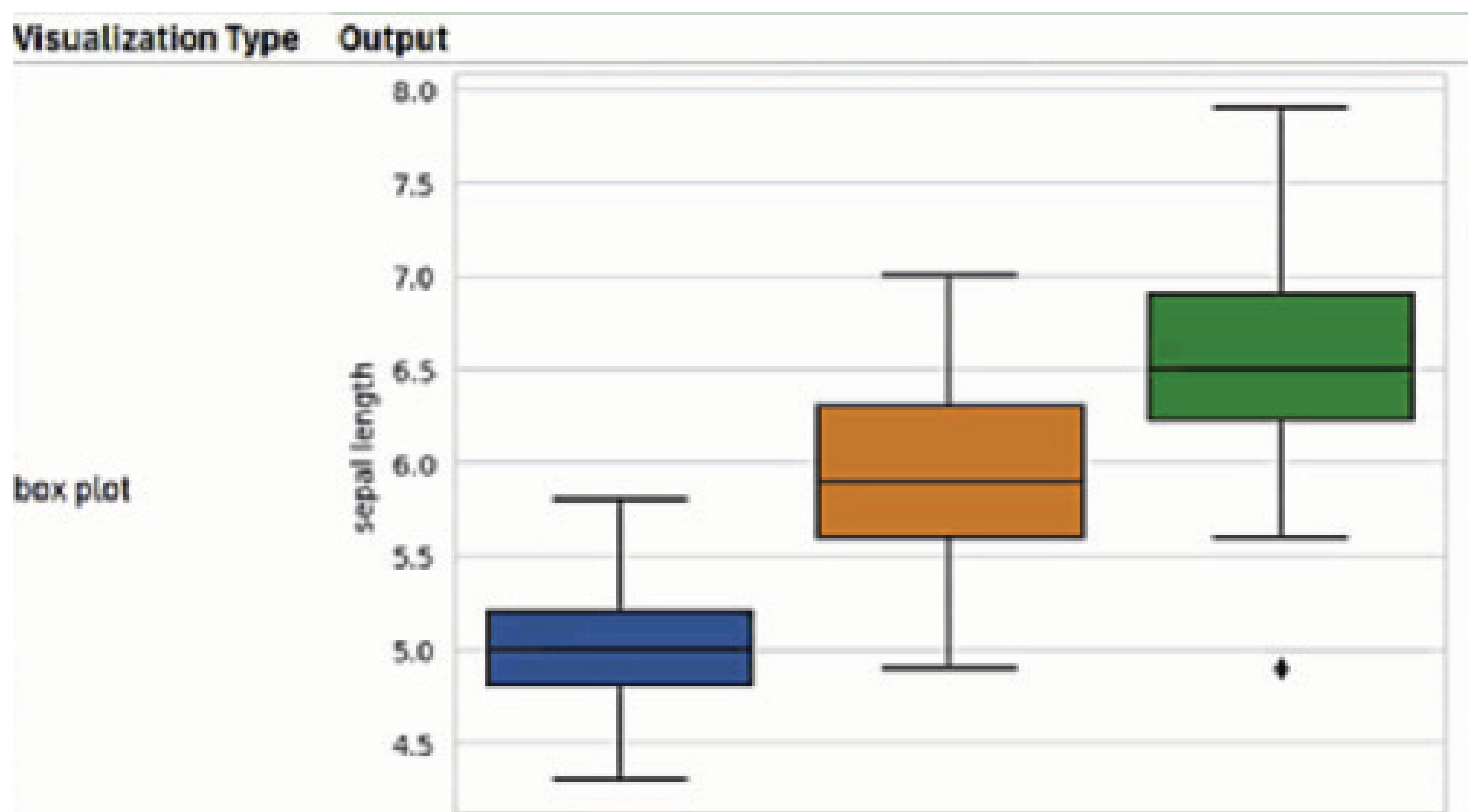
 Follow [Financial Modeling World Cup](#) on [LinkedIn](#)

3. Advanced Data Visualizations

You can also use the powerful visualization libraries of Python, including Matplotlib and Seaborn, directly within your Excel workbook to achieve thorough and insightful data portrayal.

The new Image in Cell feature, specifically, the Create Reference option allows to create an image/shape over the cells that can be resized.

This has the capacity to lead to powerful interactive dashboards, offering a realm of opportunities alongside Python's data refinement capabilities.



Source: Anaconda.com

4. In-Depth Statistical Analysis

Enhance your analysis by harnessing the capabilities of Python's libraries like pandas and statsmodels. Conduct extensive statistical operations seamlessly within your Excel cells. You don't need advanced data science expertise— advanced analytics have become achievable for all. To get descriptive statistics table like the one below, use a formula name.

`describe()`.

Output

	sepal_length	sepal_width	petal_length	petal_width
count	150	150	150	150
mean	5.8433333333	3.054	3.7586666667	1.1986666667
std	0.828066128	0.433594311	1.76442042	0.763160742
min	4.3	2	1	0.1
25%	5.1	2.8	1.5	0.3
50%	5.8	3	4.35	1.3
75%	6.4	3.3	5.1	1.8
max	7.9	4.4	6.9	2.5

Source: [Anaconda.com](https://www.anaconda.com)

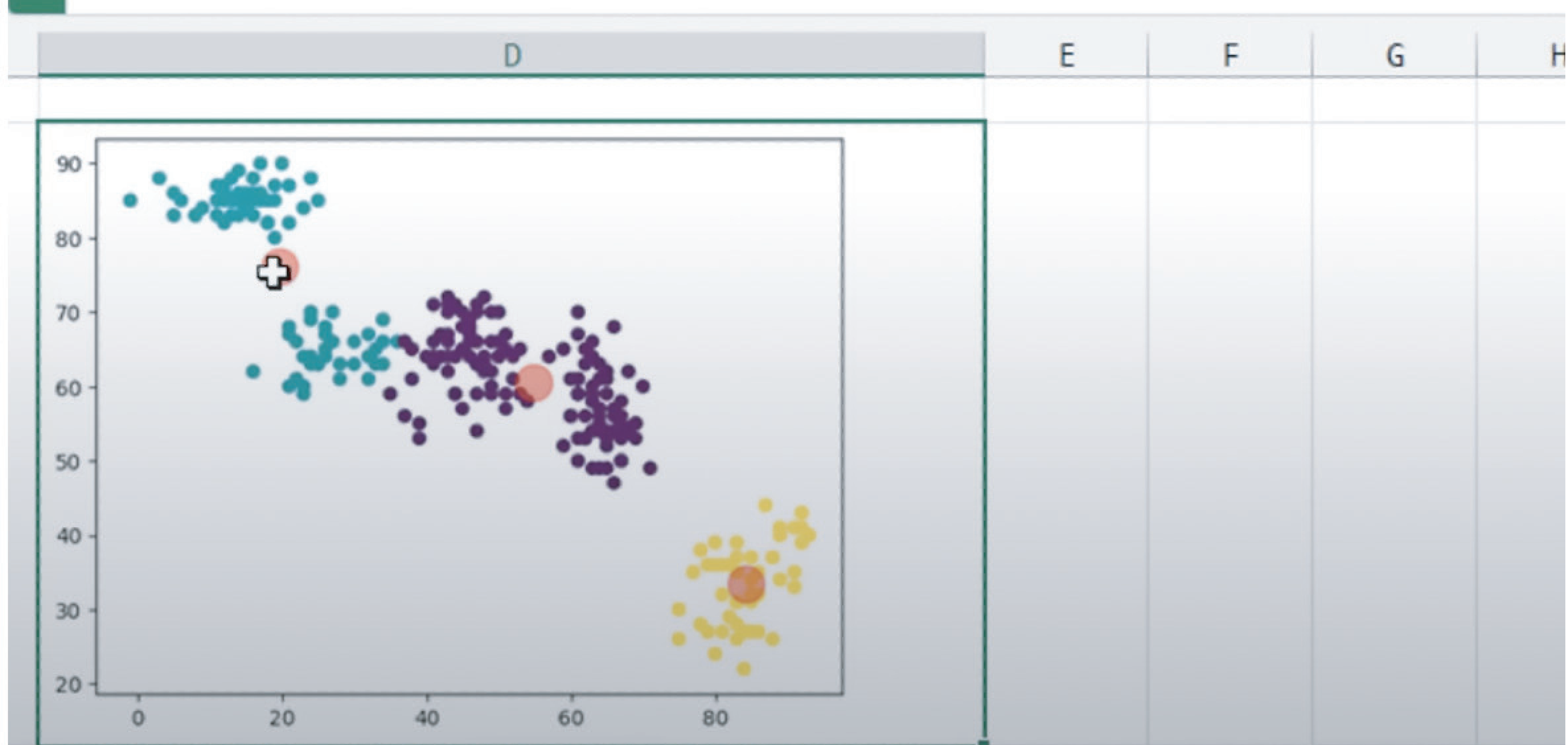
5.K-Means Clustering

You can also easily make clusters with Python in Excel which makes the work for data analysts a lot easier. This previously wasn't possible by default.

```
PY from sklearn.cluster import KMeans

df=xl("A5:B256", headers=True)
df_cluster = df[['x', 'y']]
kmeans = KMeans(n_clusters=3)
kmeans.fit(df_cluster)

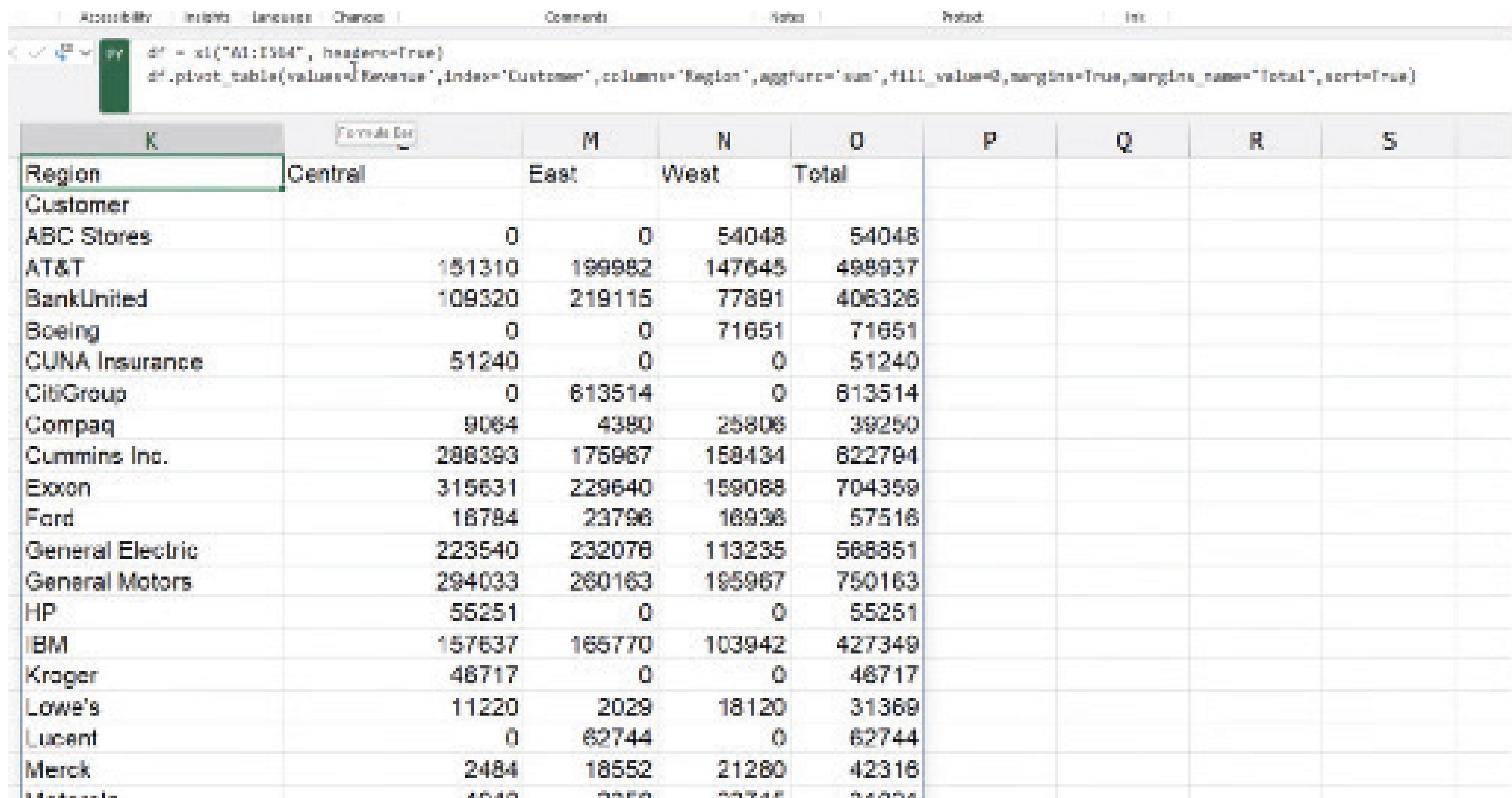
plt.scatter(df_cluster['x'], df_cluster['y'], c=kmeans.labels_, cmap='viridis')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=300, alpha=0.5)
plt.show()
```



Source: Mr. Excel

6. Create Pivot Tables

With Python integration, you can also create an Excel-like Pivot Table in a DataFrame. Compared to the basic Excel Pivot Tables, these ones are easier to use since they do automatic recalculation without refreshing. As well as you can use Date Grouping for some more possibilities within your Pivot Table.



```
df = xl("A1:J34", headers=True)
df.pivot_table(values='Revenue', index='Customer', columns='Region', aggfunc='sum', fill_value=0, margins=True, margins_name='Total', sort=True)
```

Region	Central	East	West	Total
Customer				
ABC Stores	0	0	54048	54048
AT&T	151310	199982	147645	498937
BankUnited	109320	219115	77891	406326
Boeing	0	0	71651	71651
CUNA Insurance	51240	0	0	51240
CitiGroup	0	613514	0	613514
Compaq	9064	4380	25806	39250
Cummins Inc.	288393	175967	158434	622794
Exxon	315631	229640	159088	704359
Ford	16784	23796	16936	57516
General Electric	223540	232076	113235	568851
General Motors	294033	260163	195967	750163
HP	55251	0	0	55251
IBM	157637	165770	103942	427349
Kroger	46717	0	0	46717
Lowe's	11220	2029	18120	31369
Lucent	0	62744	0	62744
Merck	2484	18552	21280	42316

Source: Mr. Excel

7. Bring In External Data

External data can be pulled into Python in Excel workflows using Excel's connectors and Power Query.

Consider that the Python integration is available in preview form only on Windows for Microsoft 365 Insider beta channel members, and only the "Office 365" cloud version receives Python integration.

 Follow [Financial Modeling World Cup](#) on [LinkedIn](#)

8. List Comprehension

List comprehension stands as a replacement for various programming functions such as for loops, lambda, map(), filter(), and reduce(). For many, list comprehension can also prove to be a more straightforward and practical approach to both understand and implement.

List comprehension allows you to apply a single rule to all the values in the test.

```
PY [i for i in test1]
```

Instead of `i` you can also apply functions:

```
PY [3*i+4 for i in test1]
```

Be aware that if you want to apply a square root in a function you have to use `**` instead of `^`

```
PY [i**2 for i in test1]
```

You can also filter the list comprehension:

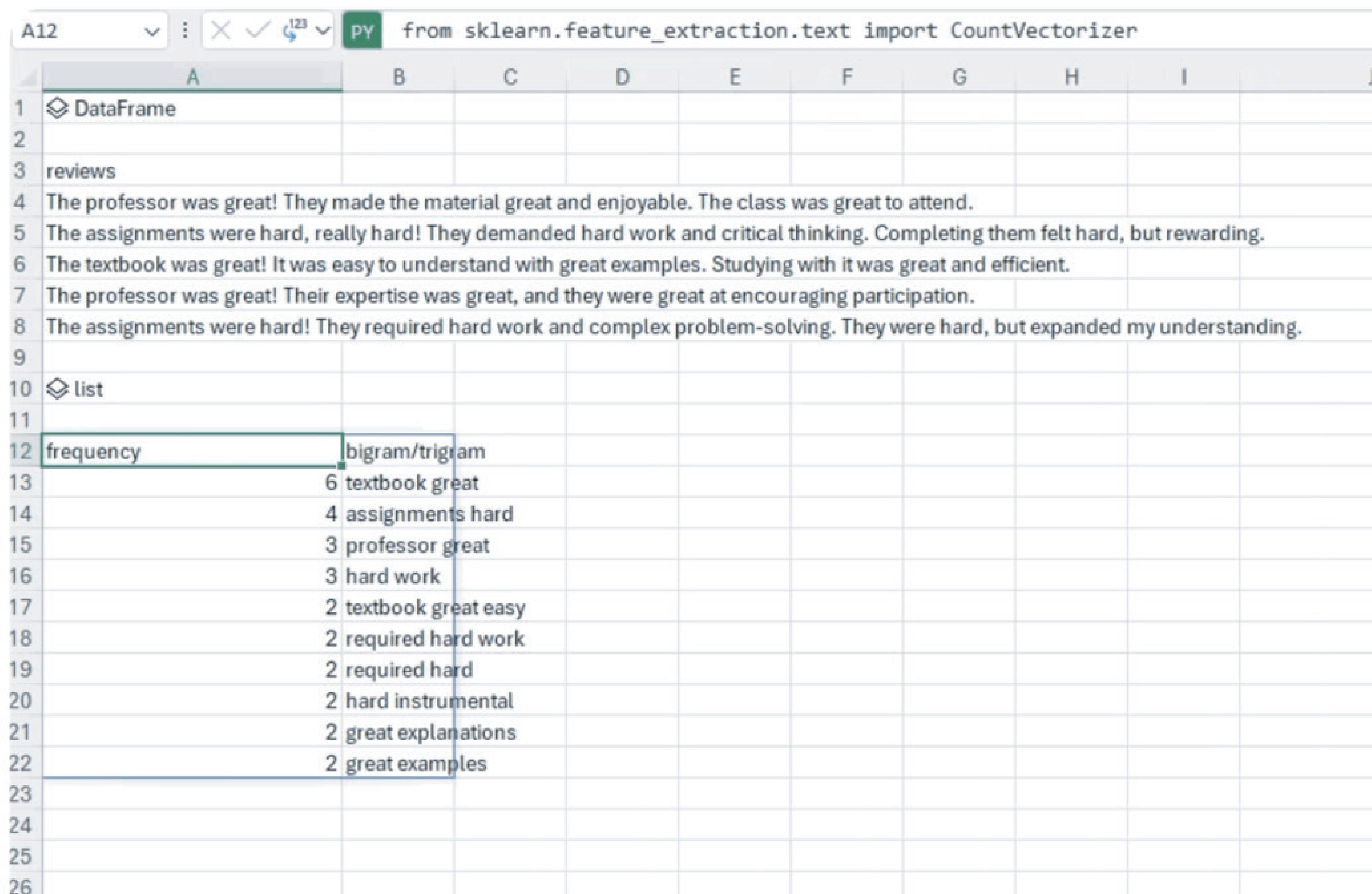
```
PY [i for i in test1 if i%2==0]
```



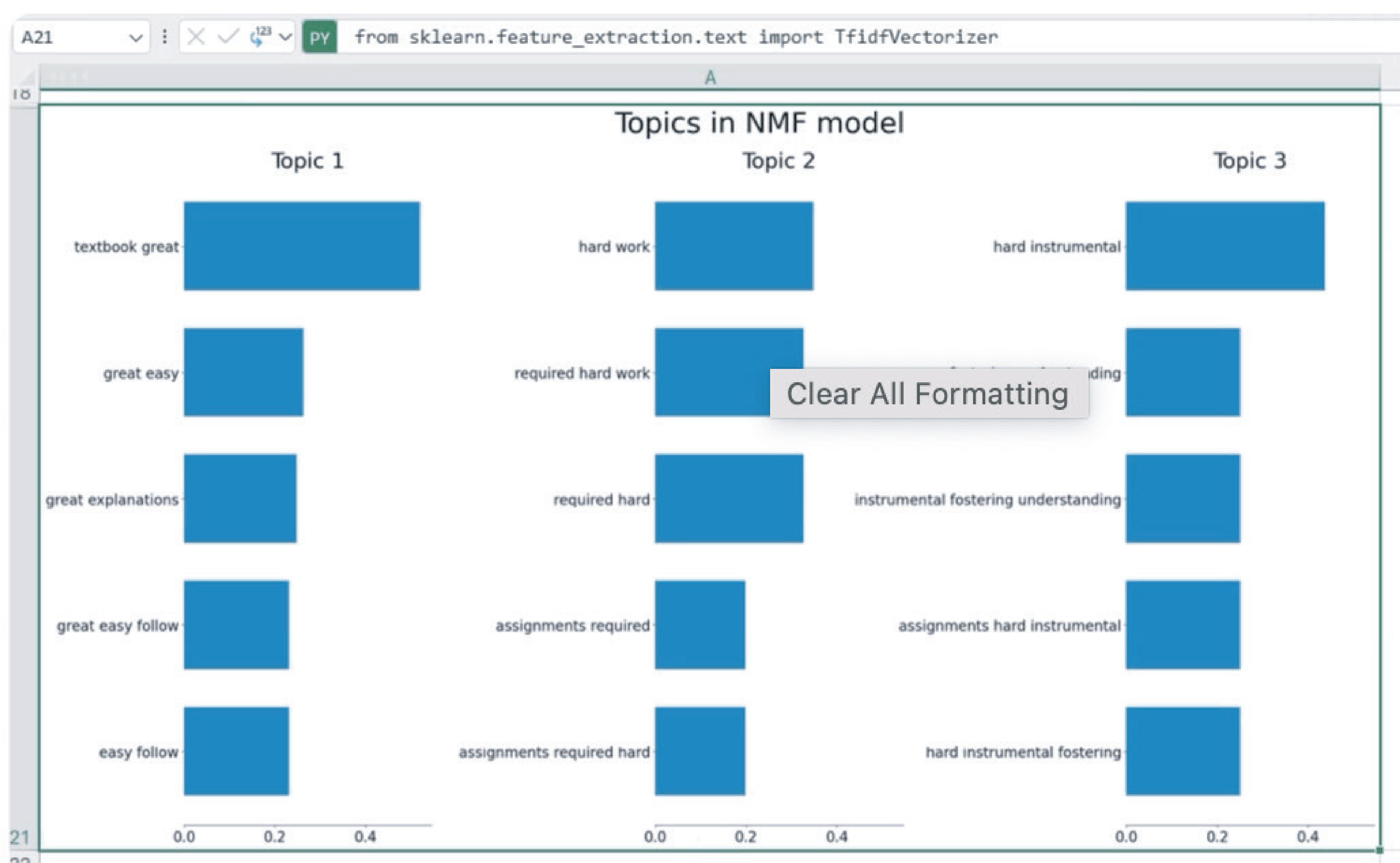
9. Analyze Texts

Text analysis is an essential technique for extracting valuable insights from unstructured text data. With Python integration in Excel, you can use different approaches to analyze texts.

For example, perform N-Gram analysis which is valuable for grasping word connections and spotting frequently recurring phrases.



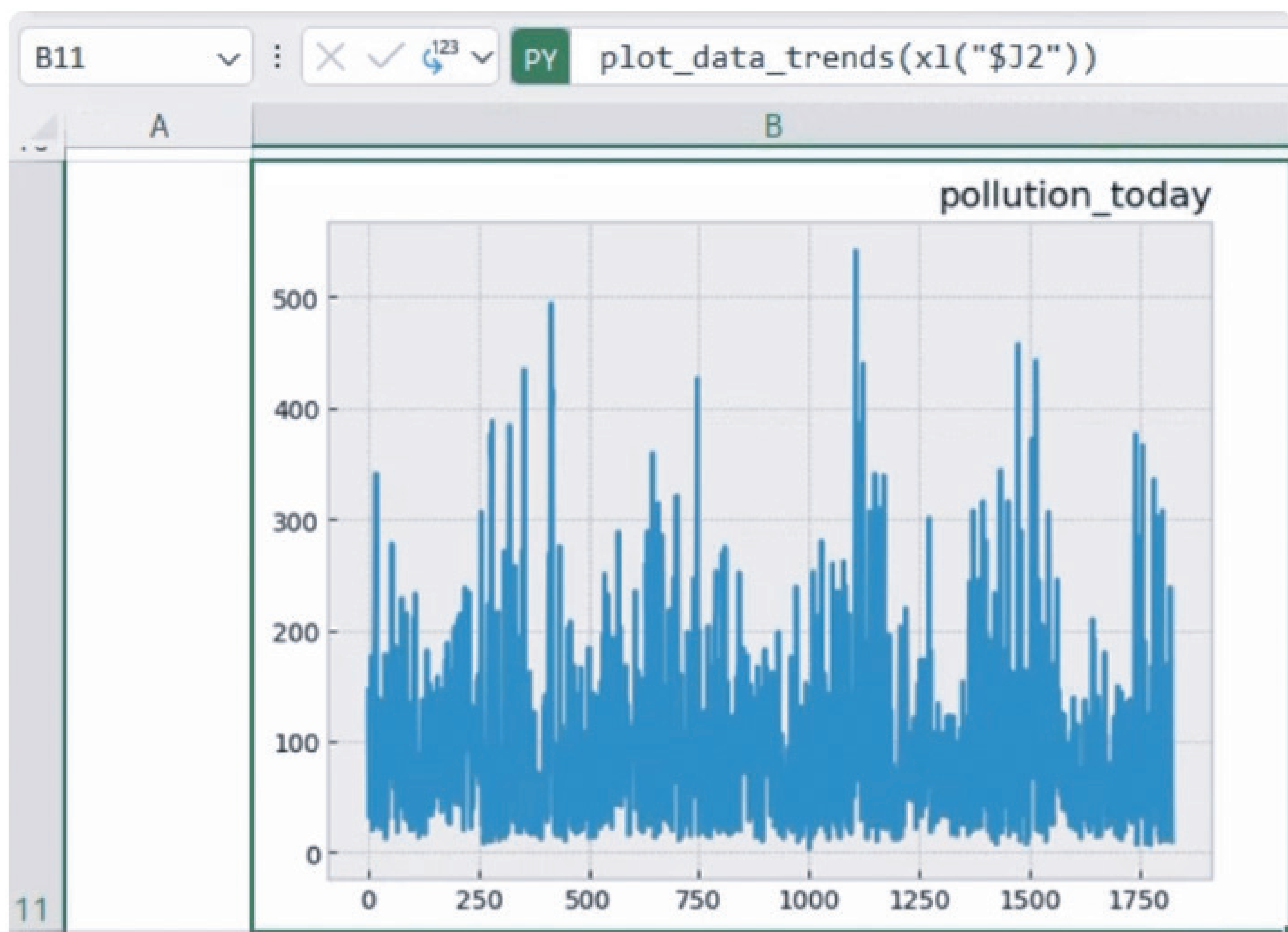
You can also perform topic modeling which is useful for understanding the main topics within a set of texts without actually reading these texts.



This can be especially useful in identifying new trends in a big chunk of search data.

10. Model and Analyze Time Series

Time series is very useful for analysis and forecasting and now it can be done with the new Python in Excel integration enabling a totally new approach to work with time series data.



For example, time series forecasting models like ARIMA or Prophet, can be used to make short-term weather predictions. Python's extensive libraries for data visualization, such as Matplotlib and Seaborn, can be used to visualize historical weather data, model predictions, and evaluation metrics.

11. Generate Code with ChatGPT

Most Excel users don't know Python code, but you can ask ChatGPT to generate the code for you! Try ChatGPT-4 with Code Interpreter turned on.

