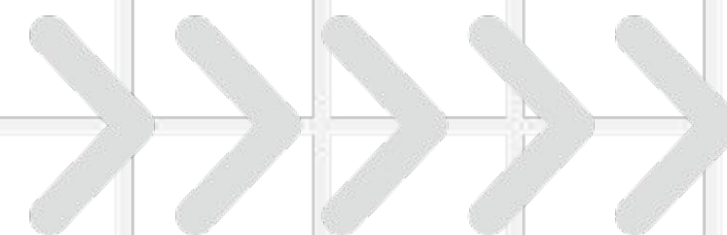


# Week 3



# SCHEDULE

Finish Setup

Review Text Preprocessing

Class Activity

– Break –

Exploratory Data Analysis

Individual/Group Activity

– Lunch –

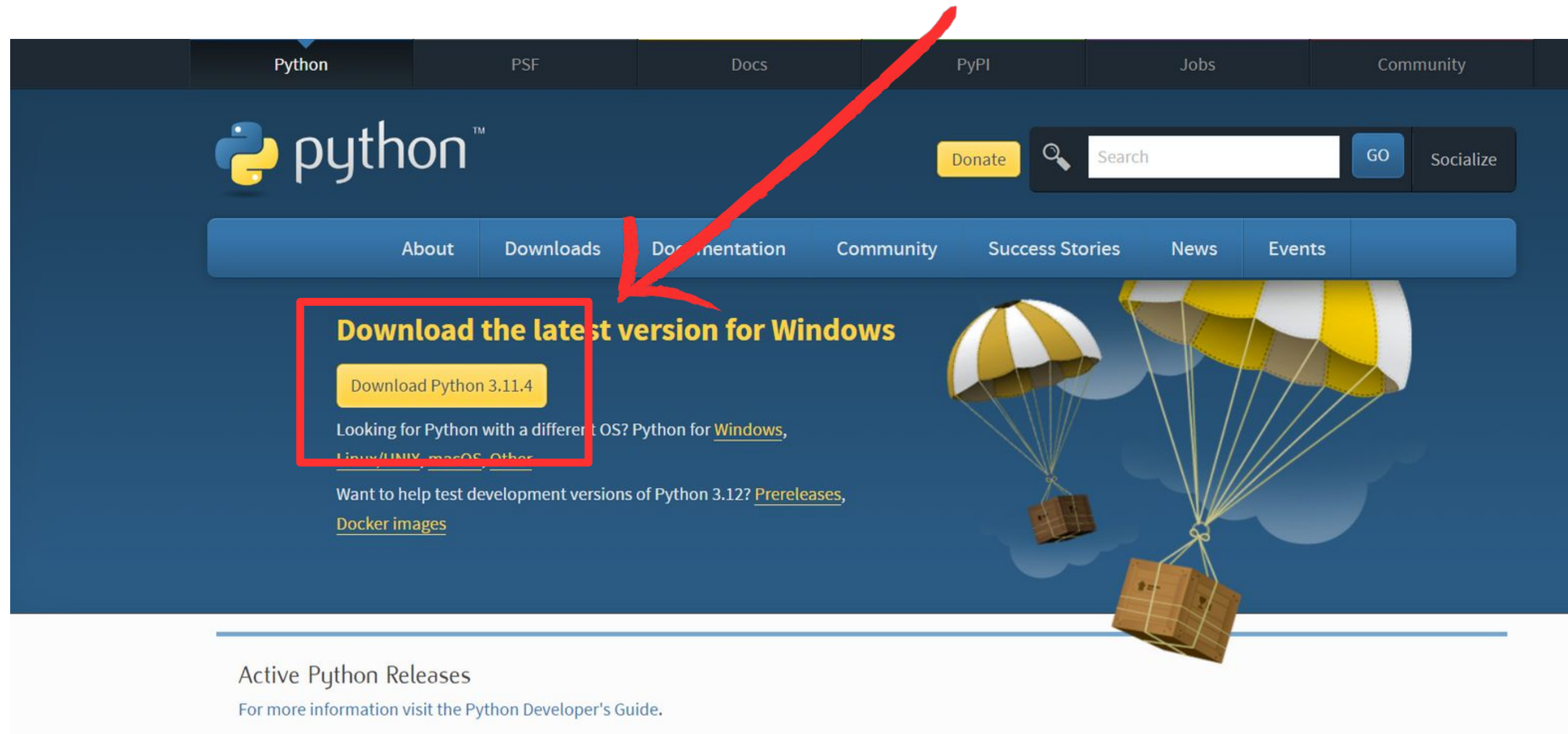
Programming Activity

– End –

# Downloading Python

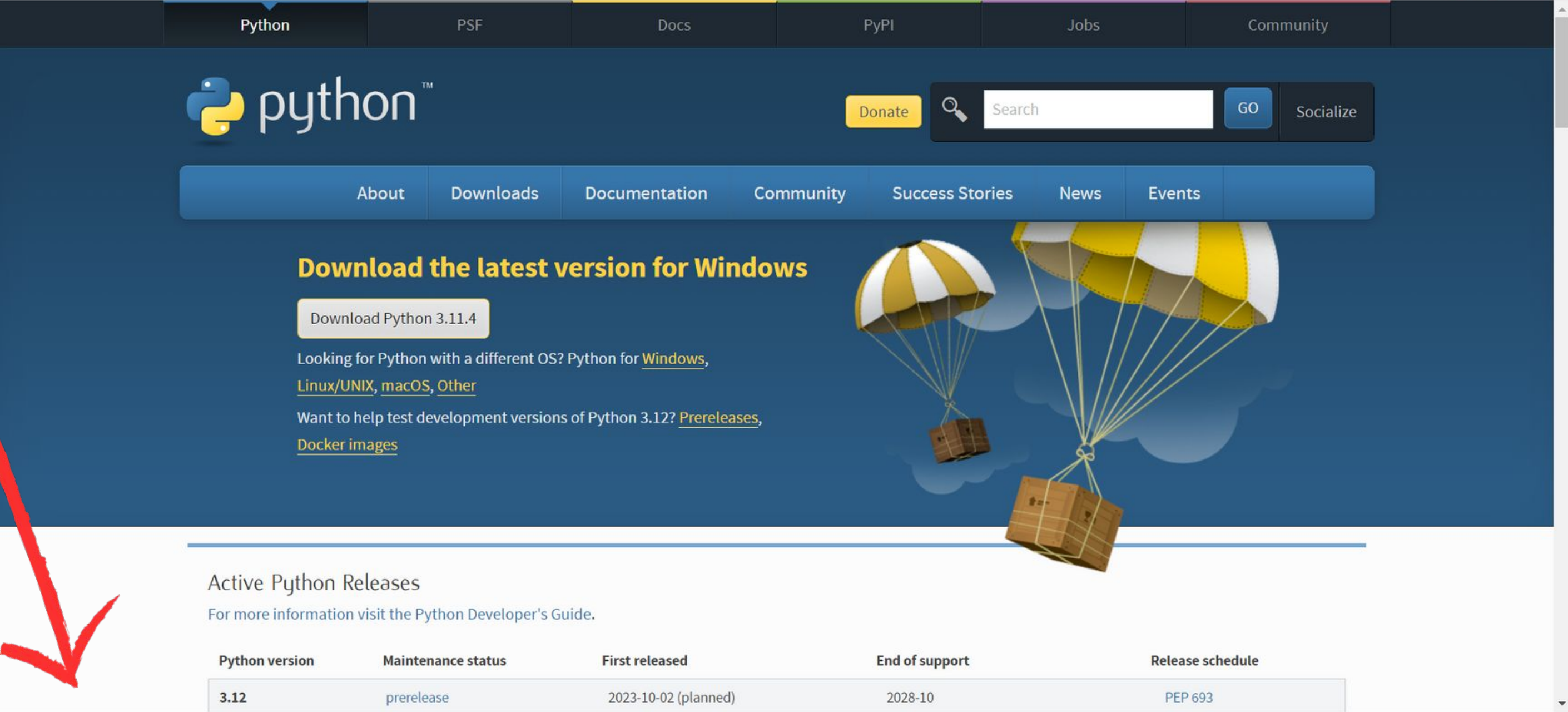
Go to <https://www.python.org/downloads/>

Click Download Python 3.11.4



# Downloading Python

After the .exe file downloaded, double click to open



The screenshot shows the Python.org website. The navigation bar includes links for Python, PSF, Docs, PyPI, Jobs, and Community. The main header features the Python logo, a search bar, and a 'Donate' button. Below the header, there are links for About, Downloads, Documentation, Community, Success Stories, News, and Events. The main content area has a large heading 'Download the latest version for Windows' and a button 'Download Python 3.11.4'. Below this, there are links for other operating systems and development versions. A red arrow points from the text on the left to the 'Download Python 3.11.4' button. At the bottom, there is a table titled 'Active Python Releases'.

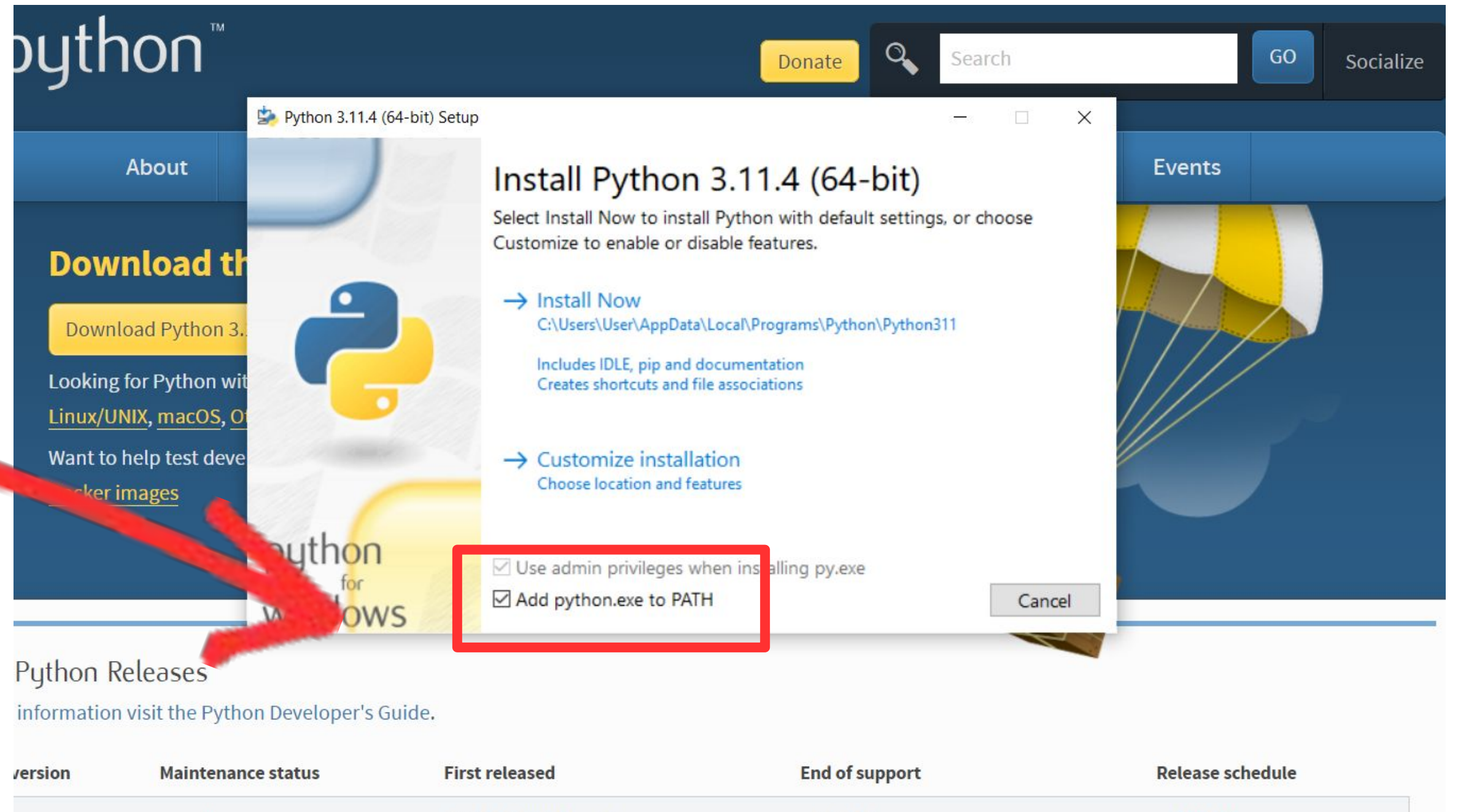
| Python version | Maintenance status | First released       | End of support | Release schedule |
|----------------|--------------------|----------------------|----------------|------------------|
| 3.12           | prerelease         | 2023-10-02 (planned) | 2028-10        | PEP 693          |

At the bottom of the browser window, a taskbar shows a download bar with the file 'python-3.11.4-am....exe' (4.8/24.2 MB, 12 secs left) and a Windows search bar.



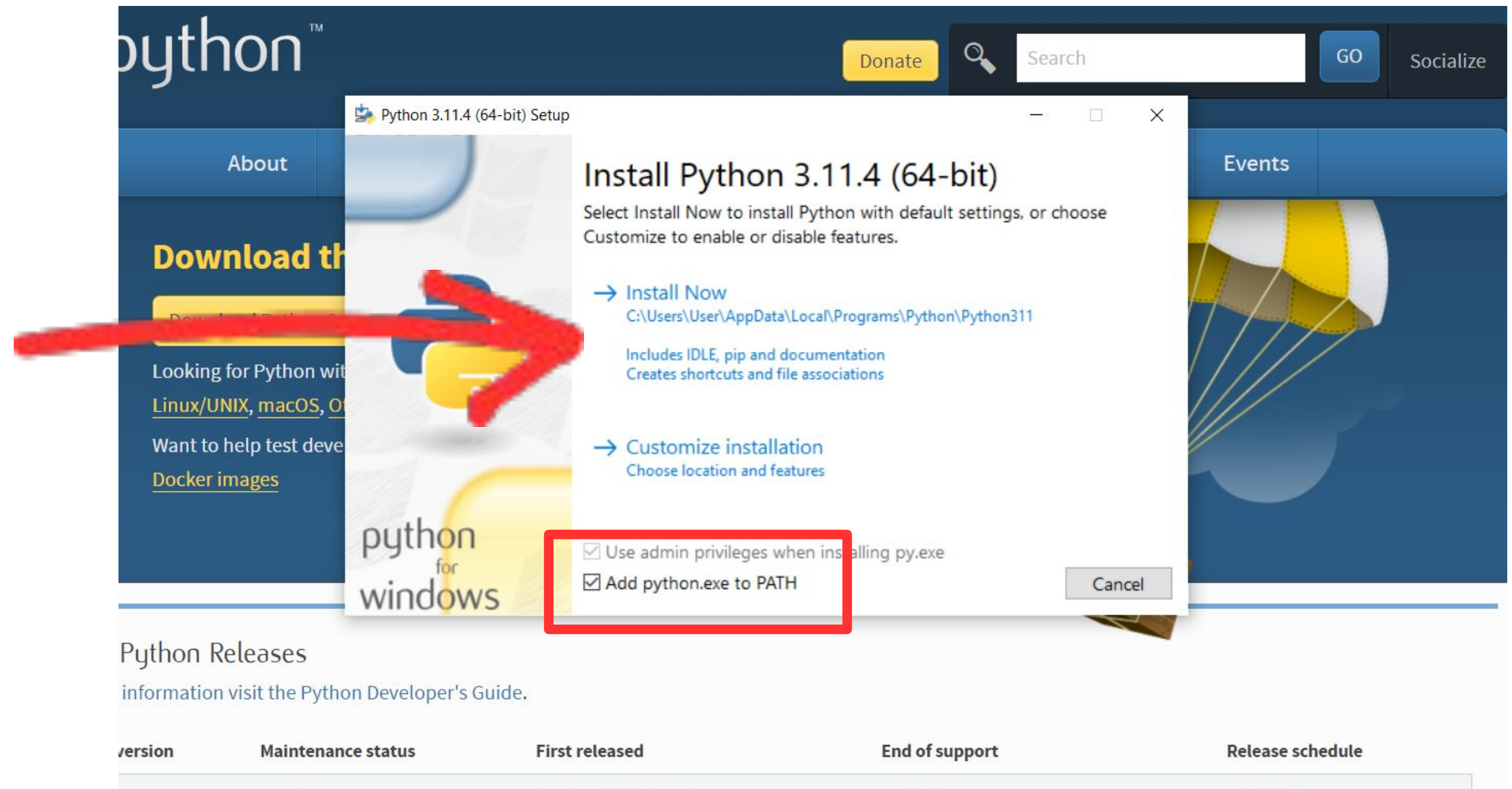
# Downloading Python

Click "Add python.exe to  
PATH"

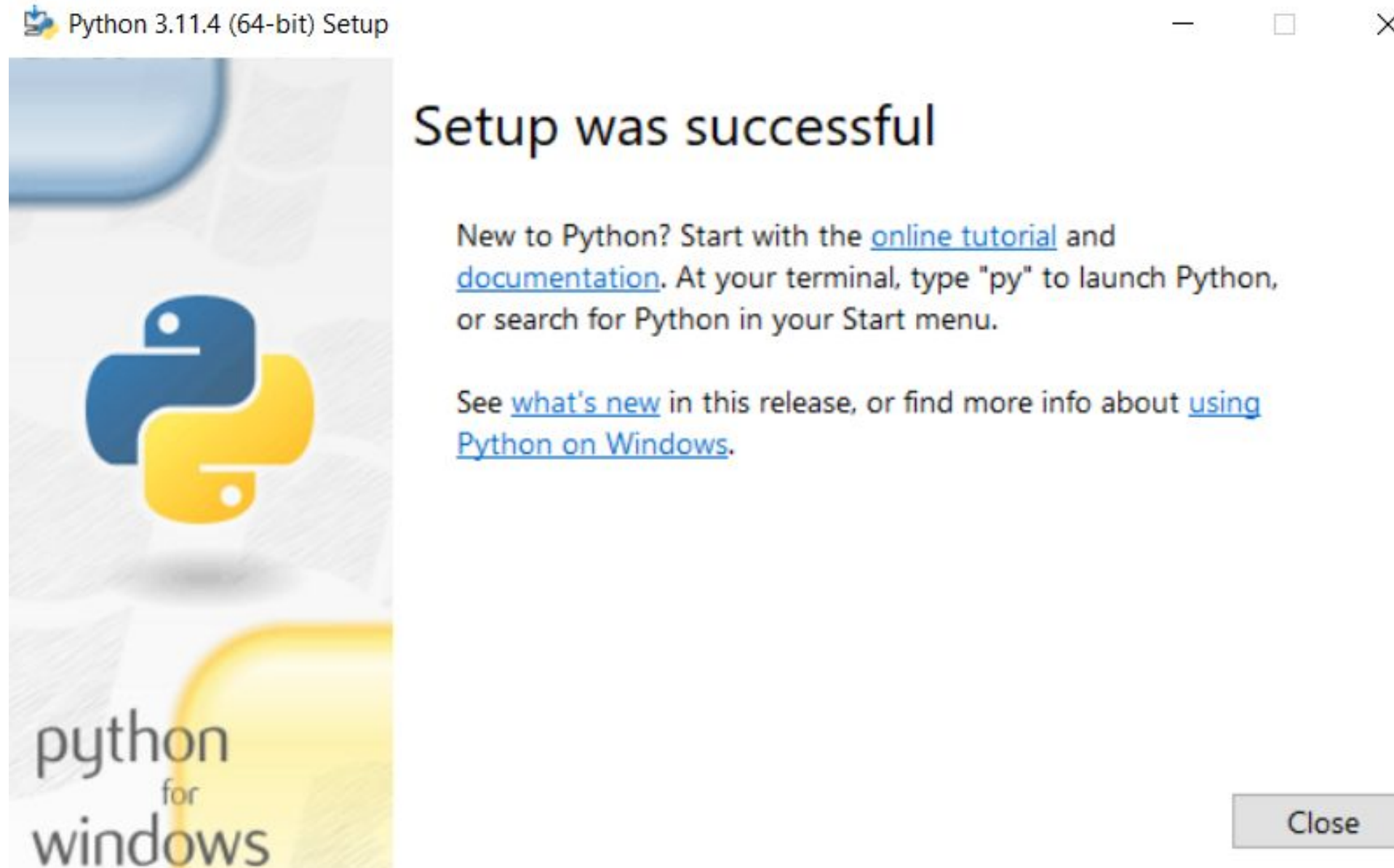


# Downloading Python

Click "Install Now"



# Downloading Python





# Connecting Python to Spyder

Click the Preferences button

Spyder

File Edit Search Source Run Debug Consoles Projects Tools View Help



C:\Users\User\Text\_Mining\untitled9.py

\_Preprocessing\_Stress\_Data.py X Womens\_Clothing\_Text\_Preprocess\_EDA.py X untitled7.py X untitled8.py X untitled9.py\* X

```
1
2  # -*- coding: utf-8 -*-
3  """
4  Text Preprocessing and Exploratory Analysis of Human Stress Dataset
5
6  Created on Sat Jul  1 19:09:56 2023
7
8  @author: User
9  """
10 #%% Import python modules and packages
11
12 # Tools to create a data frame(table)
13 import pandas as pd
14 from pandas import DataFrame
15
16 # Tools for text preprocessing
17 import nltk
18 import re
19
20 # Tools to remove stopwords
21 from nltk.corpus import stopwords
22
23 # Tools for tokenizing
24 from nltk.tokenize.toktok import ToktokTokenizer
25
26 # Tools for lemmatization
27 from nltk.stem import WordNetLemmatizer
28 nltk.download('stopwords')
29 nltk.download('punkt')
30 nltk.download('wordnet')
31
```

C:\Users\User

↓ ↻ 🗑 🔍 ↺

| Name | Type | Size | Value |
|------|------|------|-------|
|------|------|------|-------|

Help Variable Explorer Plots Files

Console 1/A X

Python 3.11.4 (tags/v3.11.4:d2340ef, Jun 7 2023, 05:45:37) [MSC v.1934 64 bit (AMD64)]  
Type "copyright", "credits" or "license" for more information.

IPython -- An enhanced Interactive Python.

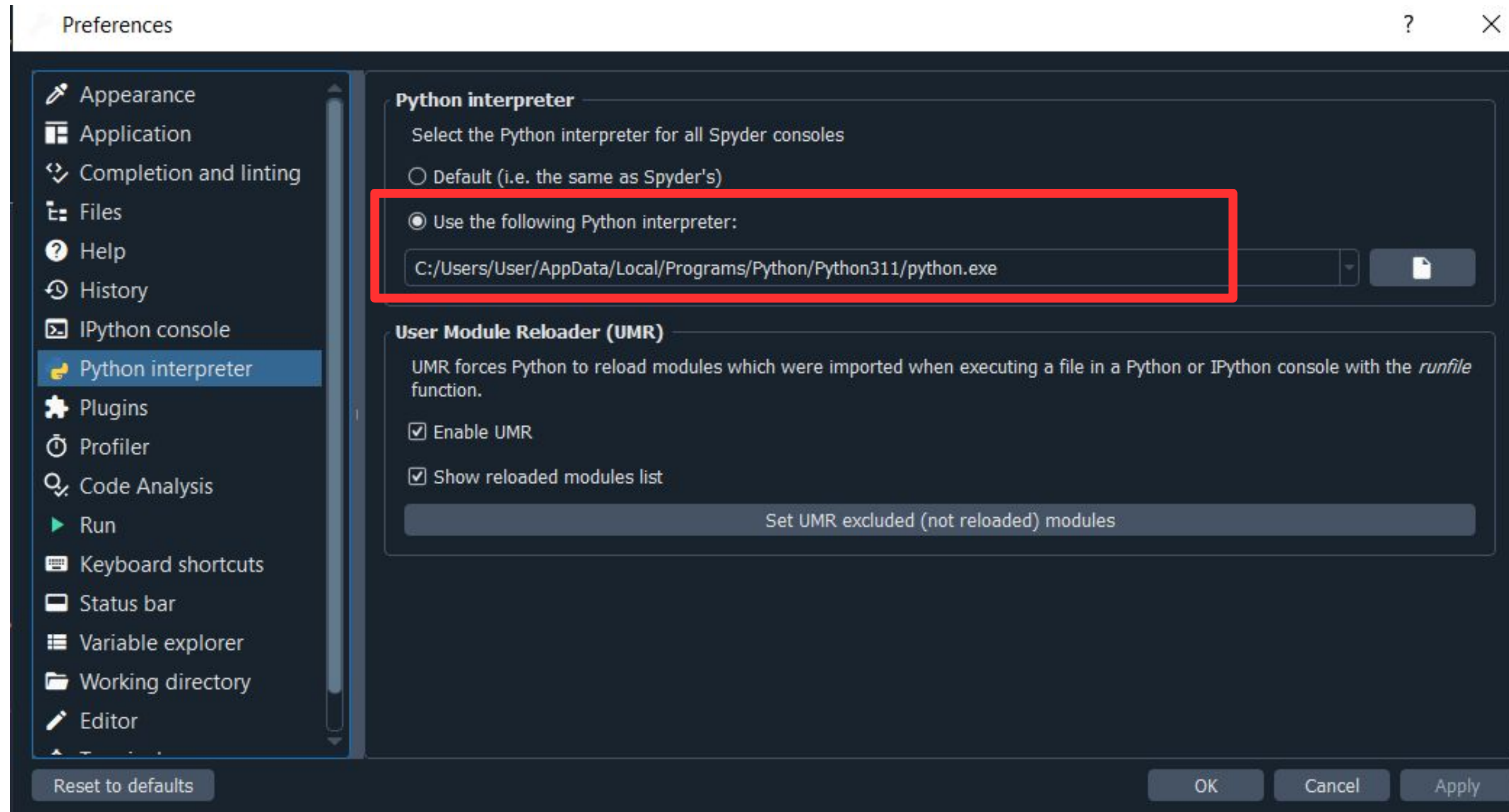
In [1]:



# Connecting Python to Spyder

Click Python Interpreter

Click “USE the following Python Interpreter”





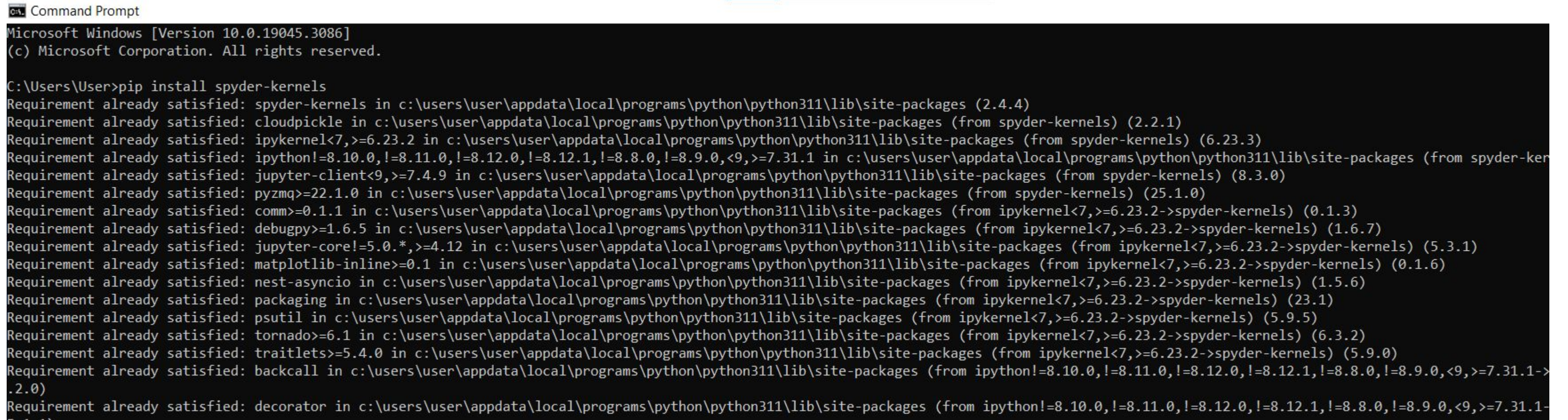
# Connecting the Terminal

Go to “Command Prompt” for Windows or “Terminal” for Mac.

Type... “pip install spyder-kernels”

Then....”pip install -U spyder-terminal”

Restart Spyder



```

C:\Users\User>pip install spyder-kernels
Requirement already satisfied: spyder-kernels in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (2.4.4)
Requirement already satisfied: cloudpickle in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from spyder-kernels) (2.2.1)
Requirement already satisfied: ipykernel<7,>=6.23.2 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from spyder-kernels) (6.23.3)
Requirement already satisfied: ipython!=8.10.0,!8.11.0,!8.12.0,!8.12.1,!8.8.0,!8.9.0,<9,>=7.31.1 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from spyder-kernels) (7.31.1)
Requirement already satisfied: jupyter-client<9,>=7.4.9 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from spyder-kernels) (8.3.0)
Requirement already satisfied: pyzmq>=22.1.0 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from spyder-kernels) (25.1.0)
Requirement already satisfied: comm>=0.1.1 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (0.1.3)
Requirement already satisfied: debugpy>=1.6.5 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (1.6.7)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (5.3.1)
Requirement already satisfied: matplotlib-inline>=0.1 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (0.1.6)
Requirement already satisfied: nest-asyncio in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (1.5.6)
Requirement already satisfied: packaging in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (23.1)
Requirement already satisfied: psutil in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (5.9.5)
Requirement already satisfied: tornado>=6.1 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (6.3.2)
Requirement already satisfied: traitlets>=5.4.0 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipykernel<7,>=6.23.2->spyder-kernels) (5.9.0)
Requirement already satisfied: backcall in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipython!=8.10.0,!8.11.0,!8.12.0,!8.12.1,!8.8.0,!8.9.0,<9,>=7.31.1->ipykernel<7,>=6.23.2->spyder-kernels) (0.2.0)
Requirement already satisfied: decorator in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from ipython!=8.10.0,!8.11.0,!8.12.0,!8.12.1,!8.8.0,!8.9.0,<9,>=7.31.1->ipykernel<7,>=6.23.2->spyder-kernels) (5.1.1)

```

# Python Modules





# NLTK

## Natural Language Toolkit (NLTK)

NLTK is a leading platform for building Python programs to work with **human language data**.

This toolkit is one of the most powerful **NLP** libraries which contains packages to make machines understand human languages and respond in an appropriate manner. Using NLTK we can perform operations such as data cleaning, visualization, and vectorization that will help us in classifying our text.

**NLP** = Natural Language Processing. Machine learning technology that gives the ability for computers to understand human language.

# Install nltk

Windows:

1. Go to **Command Prompt** on your PC
2. Type `pip install nltk`

macOS:

1. Go to **Terminal**
2. Type `pip install nltk`

Command Prompt

```
Microsoft Windows [Version 10.0.19045.3086]  
(c) Microsoft Corporation. All rights reserved.
```

```
C:\Users\User>pip install nltk
```

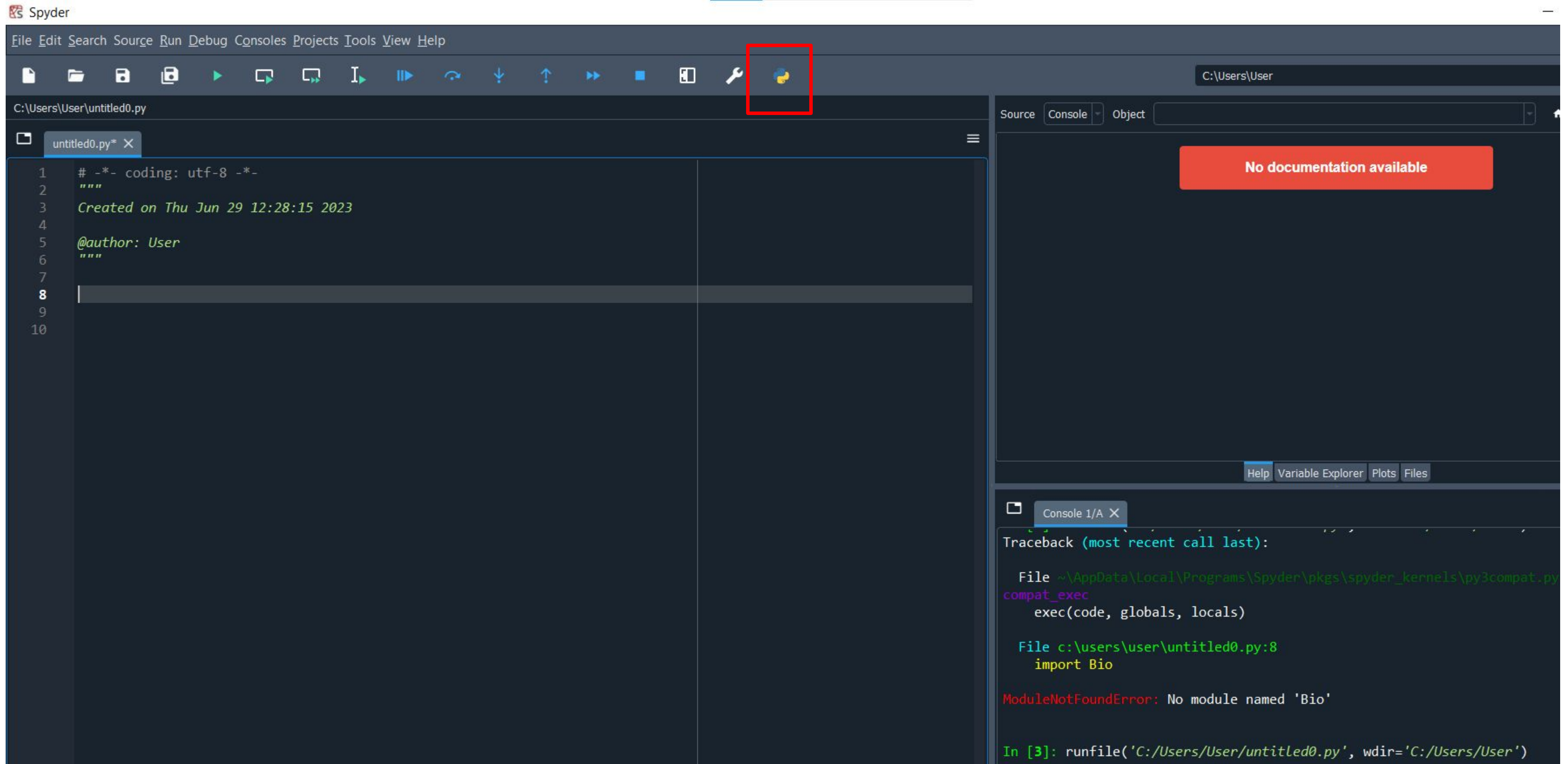
```
Requirement already satisfied: nltk in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (3.8.1)  
Requirement already satisfied: click in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from nltk) (8.1.3)  
Requirement already satisfied: joblib in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from nltk) (1.2.0)  
Requirement already satisfied: regex>=2021.8.3 in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from nltk) (2023.6.3)  
Requirement already satisfied: tqdm in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from nltk) (4.65.0)  
Requirement already satisfied: colorama in c:\users\user\appdata\local\programs\python\python311\lib\site-packages (from click->nltk) (0.4.6)
```

```
C:\Users\User>_
```

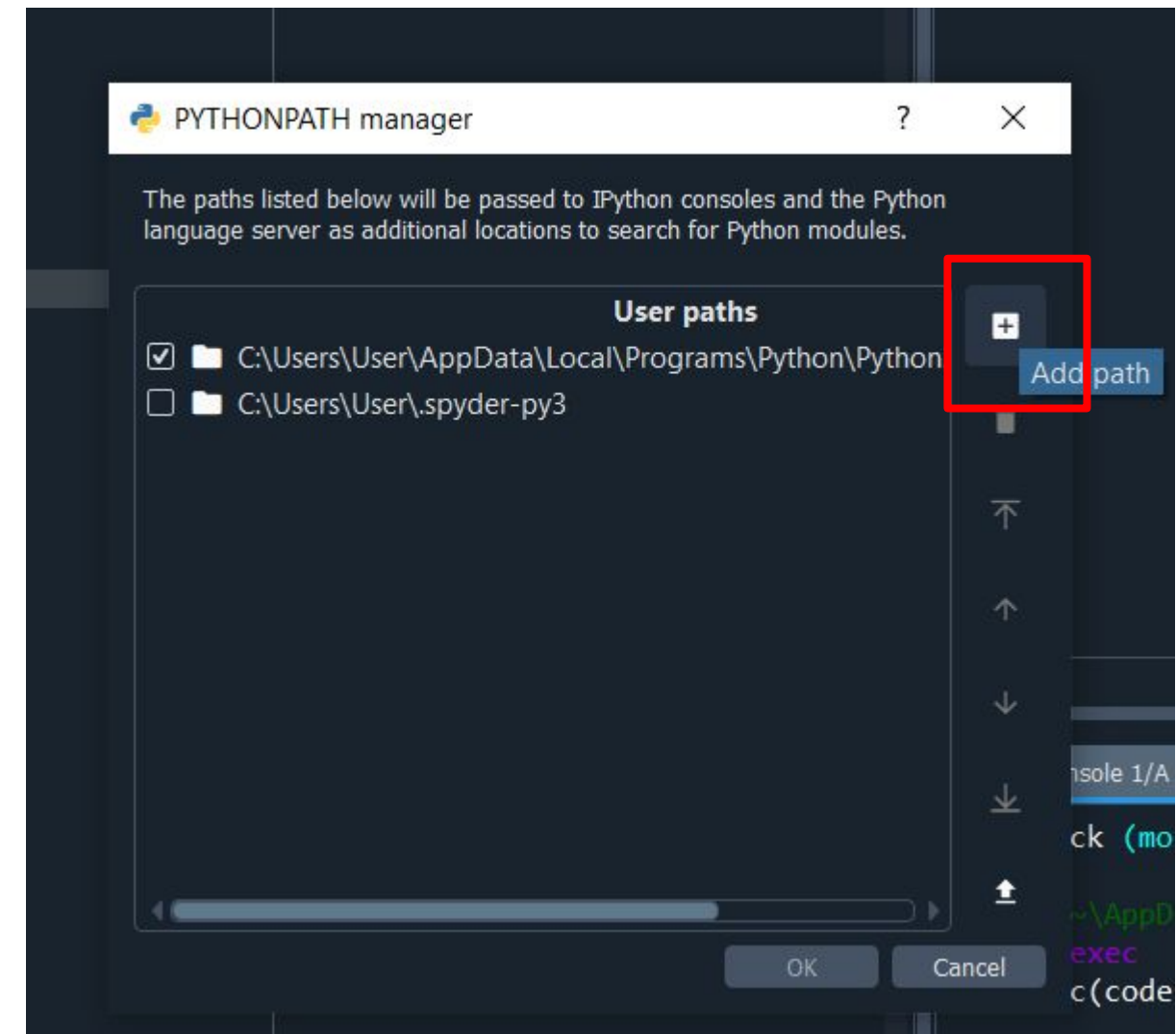
**Copy the path that the package has been downloaded.**



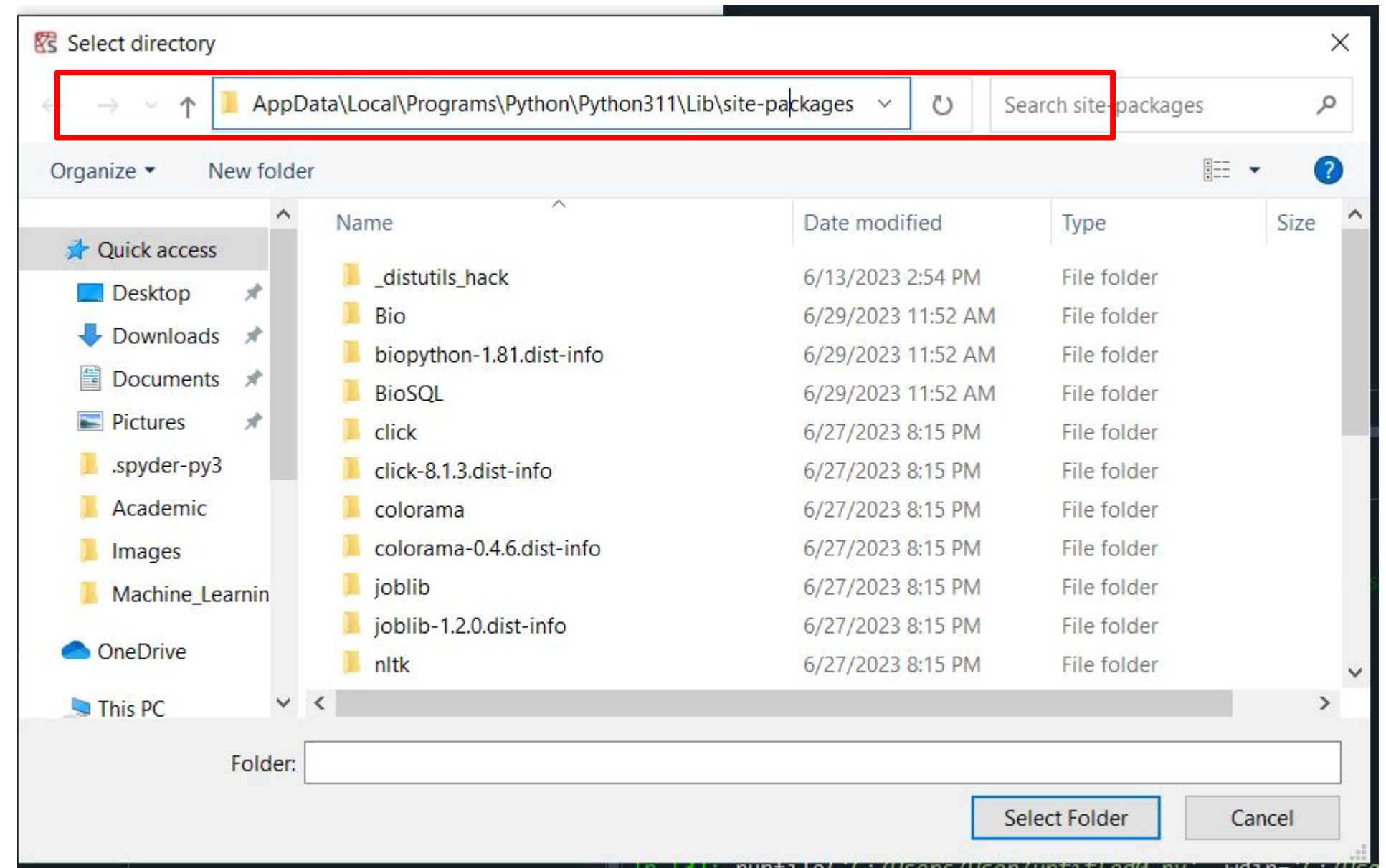
Go to Spyder. Click on the python icon.



This window should pop up. Click the + button to Add path.

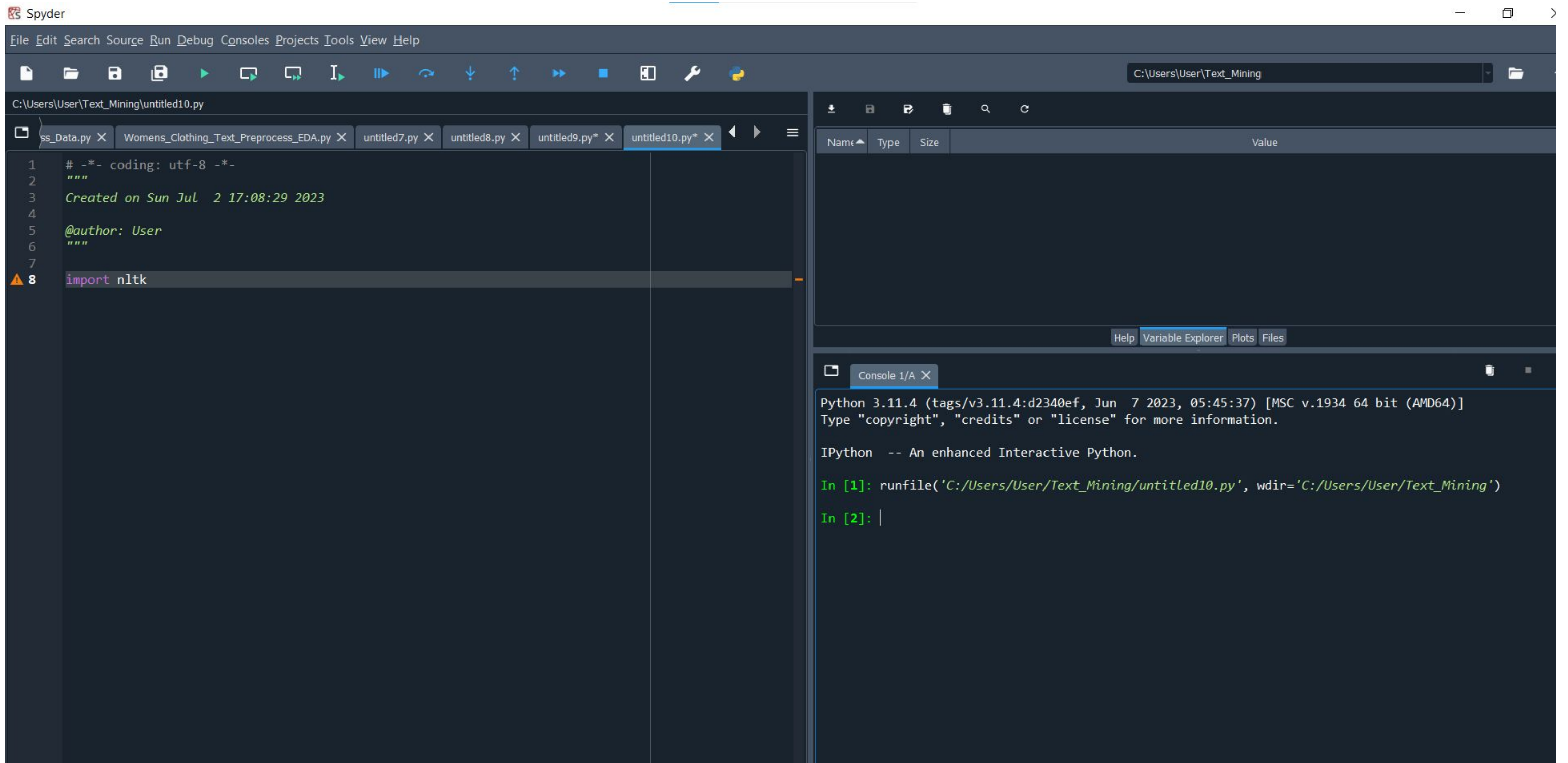


Paste the path on the top box. Then click Select Folder.





To double check that everything works well...



The screenshot displays the Spyder Python IDE interface. The main editor window shows a Python script with the following content:

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Sun Jul 2 17:08:29 2023
4
5 @author: User
6 """
7
8 import nltk
```

The console window at the bottom shows the execution of the script, including the Python version and the IPython prompt:

```
Python 3.11.4 (tags/v3.11.4:d2340ef, Jun 7 2023, 05:45:37) [MSC v.1934 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/User/Text_Mining/untitled10.py', wdir='C:/Users/User/Text_Mining')
In [2]: |
```

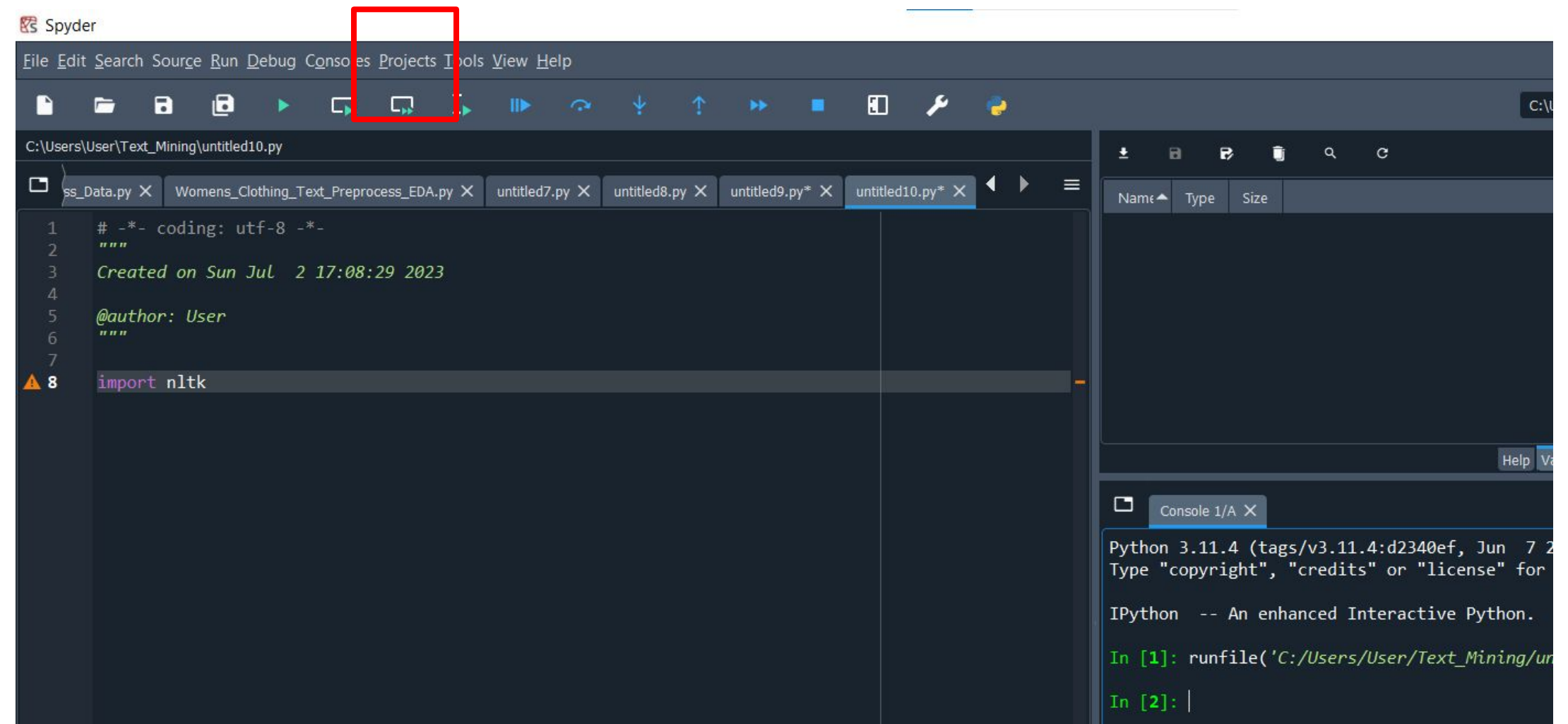
# Creating a project in Spyder

Go to “Projects”

Click New Project

Click New Directory

Name project as “Text\_Mining”



# Review Activities of Text Preprocessing



Go to GitHub and download all these files.

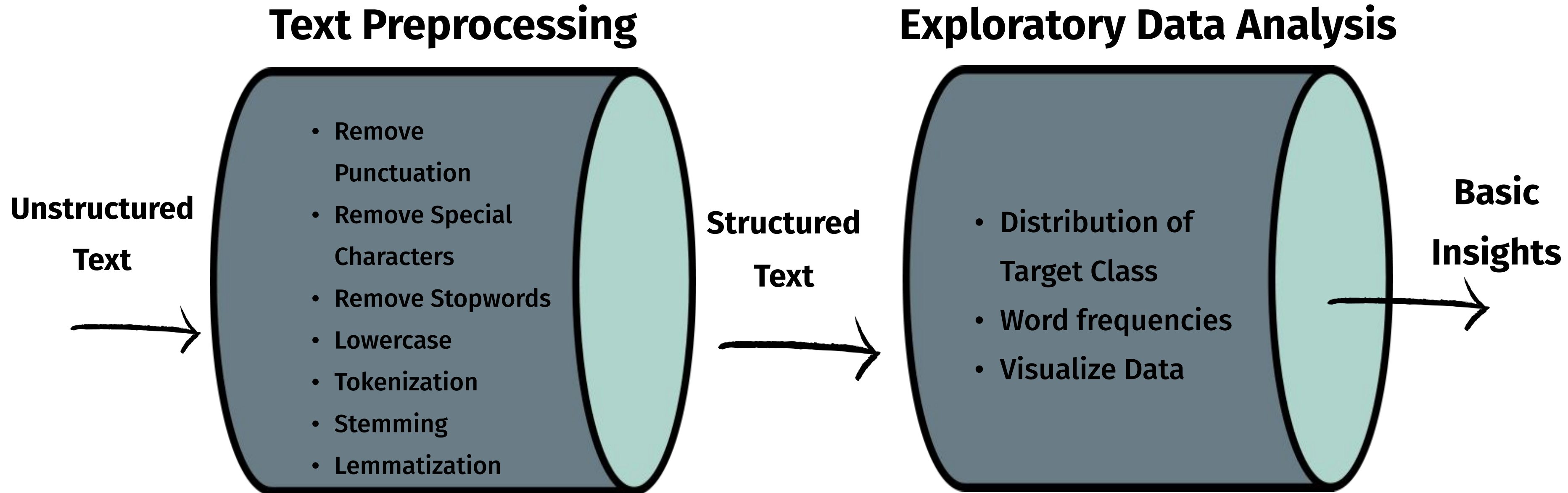
Put files in your Text\_Mining project folder

Open files in Spyder

| GitHub File name                         | Content                               |
|--|---------------------------------------|
| Text_Preprocessing.py                    | Python script with notes and code     |
| Text_Preprocessing_EDA_Stress_Student.py | Practice on text prepressing activity |
| Stress.csv                               | Data set for class activity           |

# Review Activity 2 of Text Preprocessing & EDA

# Building the text mining pipeline



# Wordle Activity