Week 3

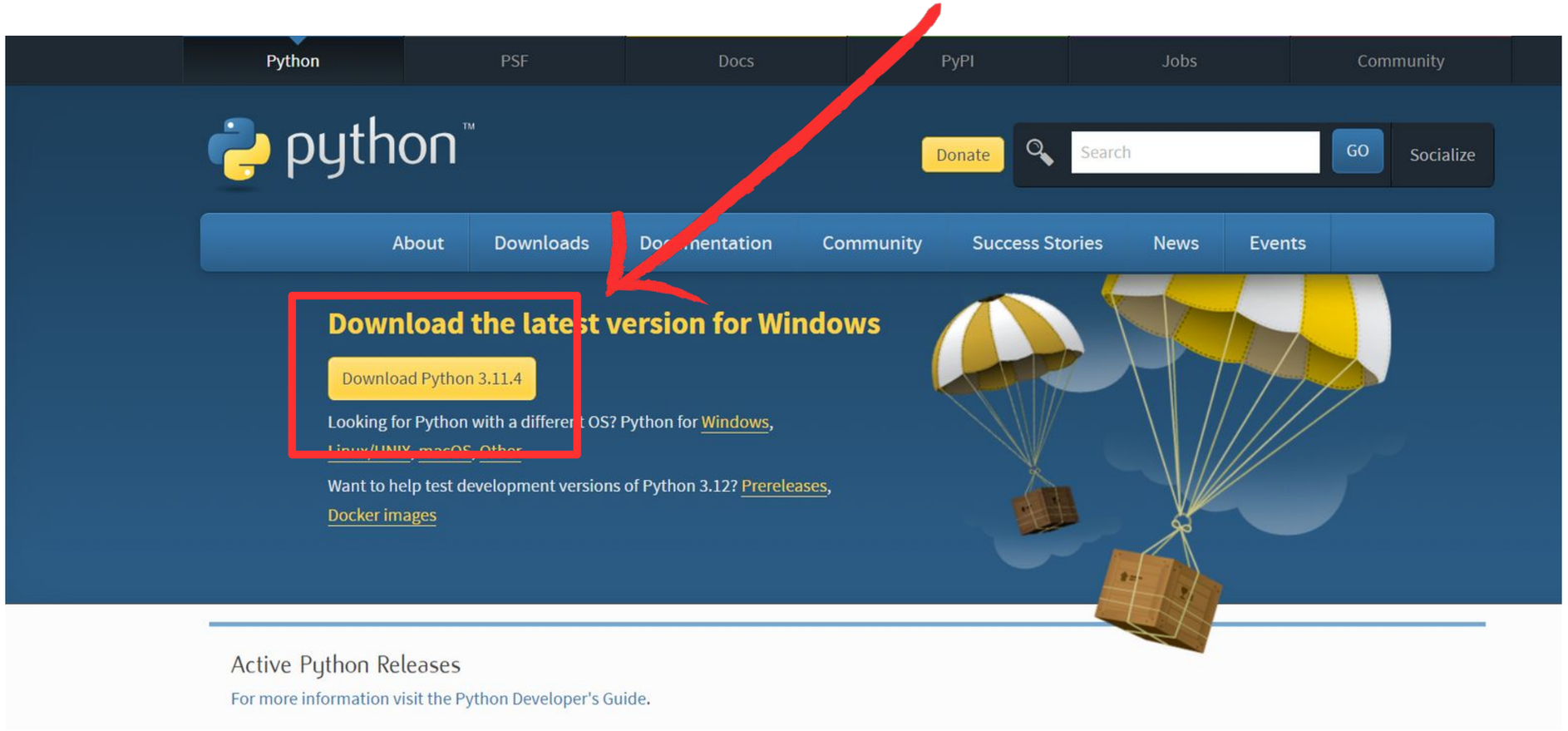
SCHEDULE

Finish Setup
Review Text Preprocessing
Class Activity
– Break –
Exploratory Data Analysis
Individual/Group Activity
– Lunch –
Programming Activity
– End –

Downloading Python

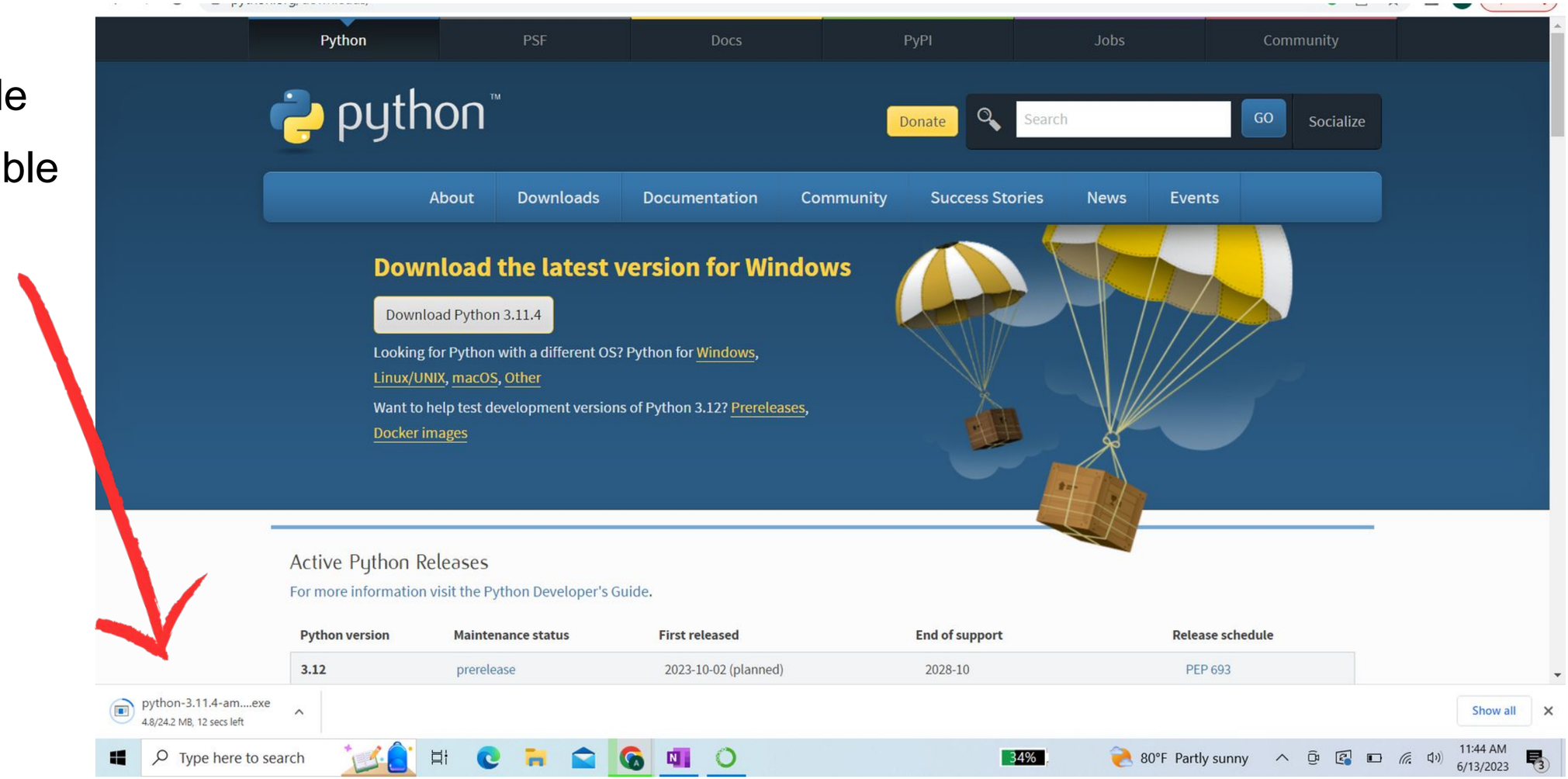
Go to https://www.python.org/downloads/

Click Download Python 3.11.4



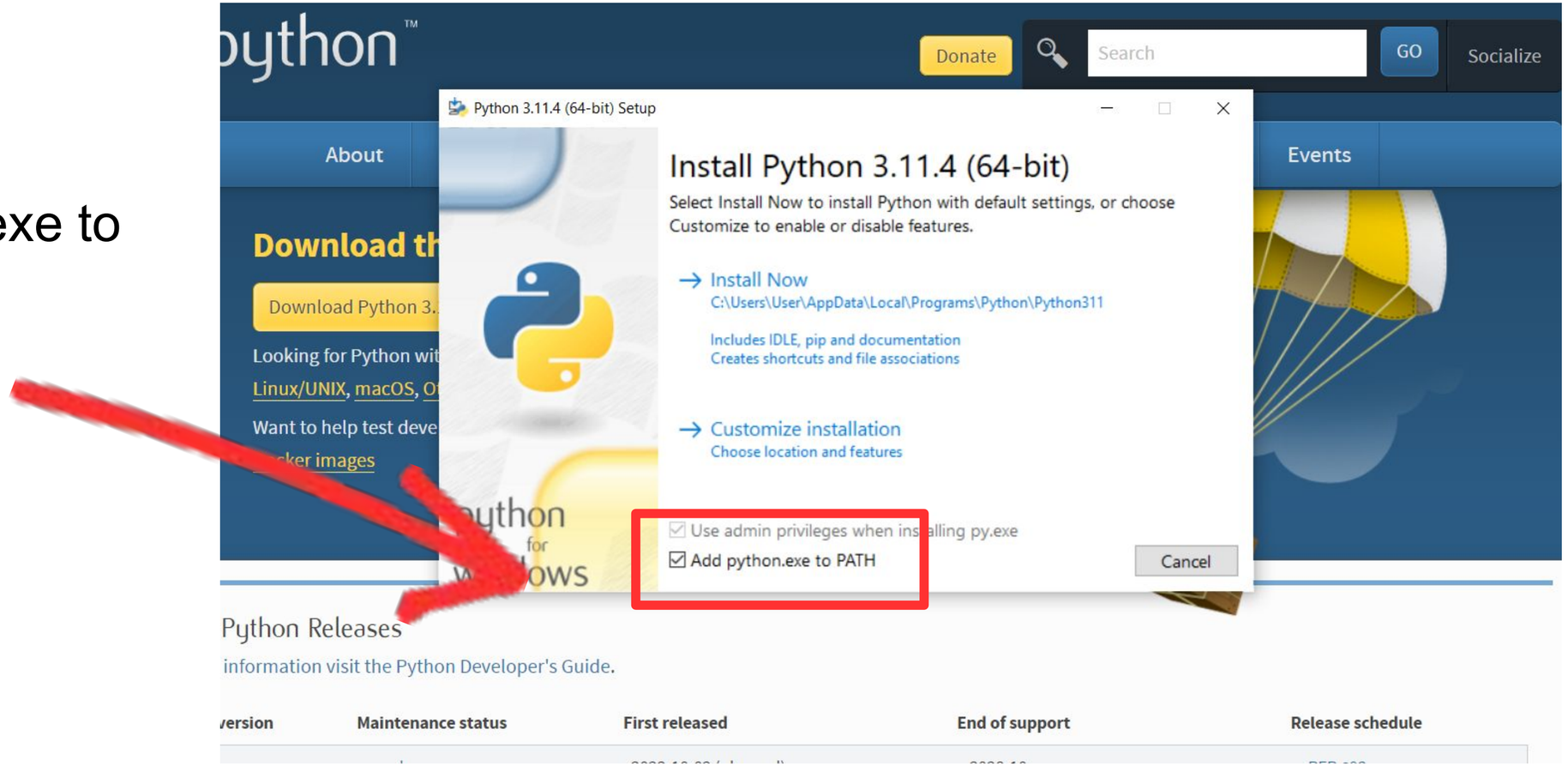
Downloading Python

After the .exe file downloaded, double click to open



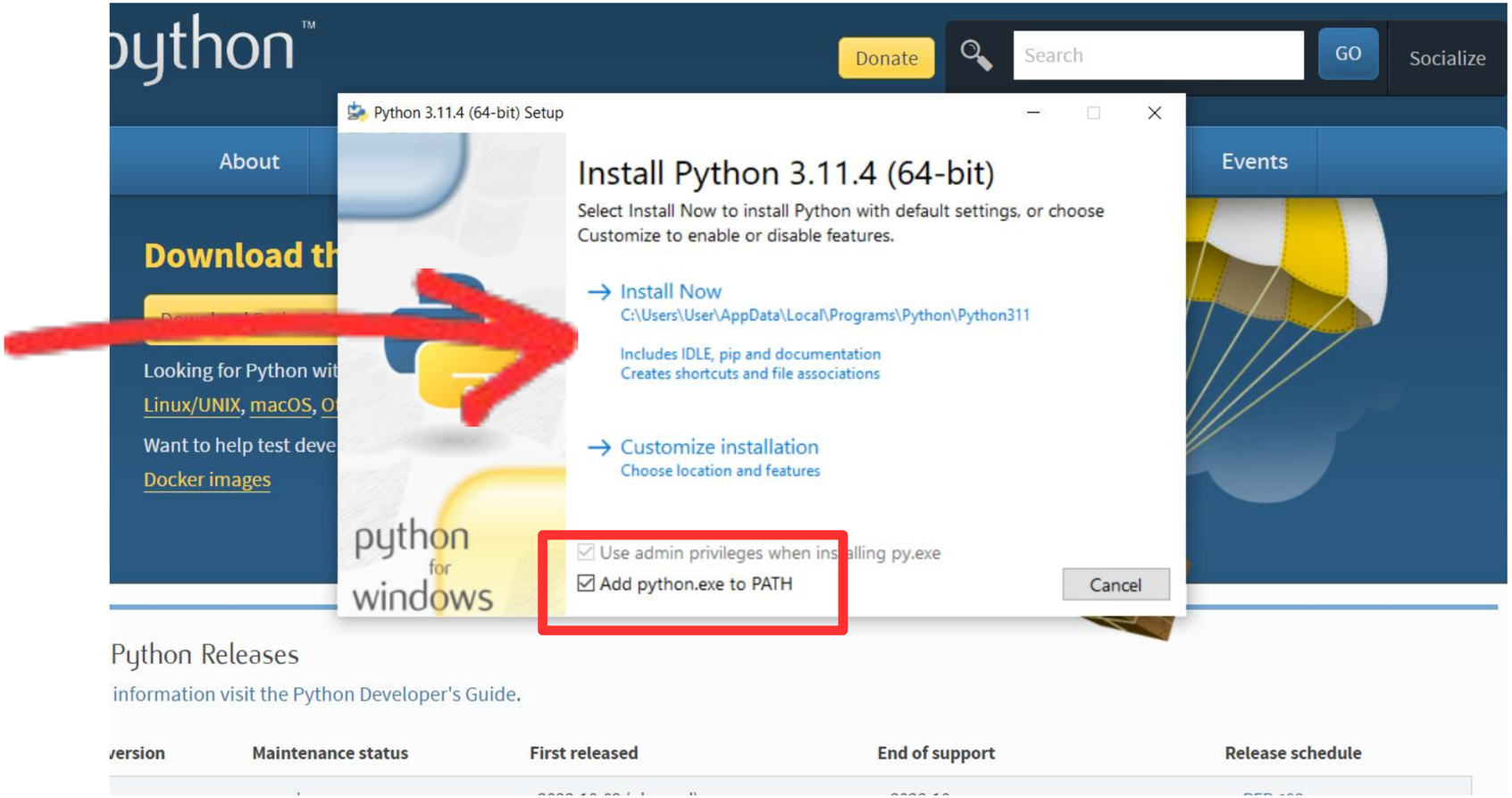
Downloading Python

Click "Add python.exe to PATH"

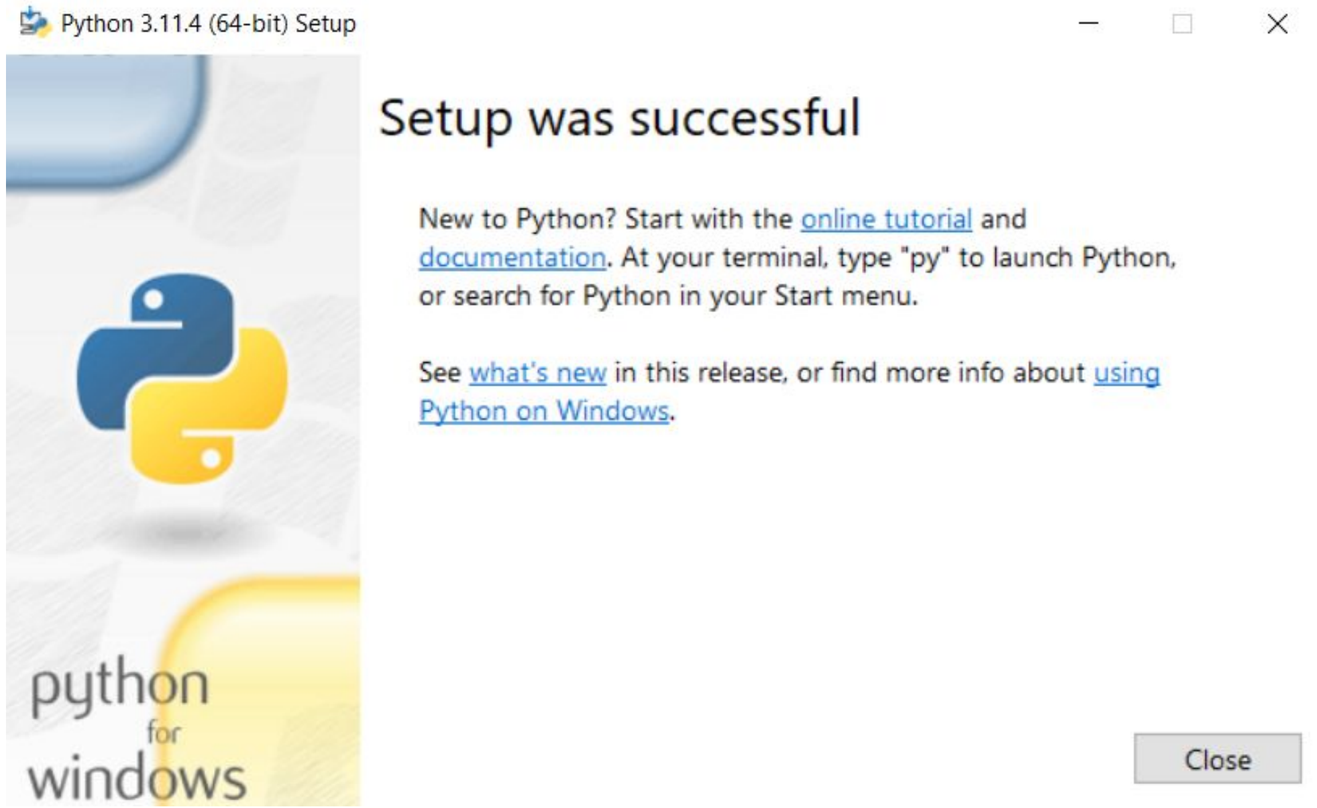


Downloading Python

Click "Install Now"

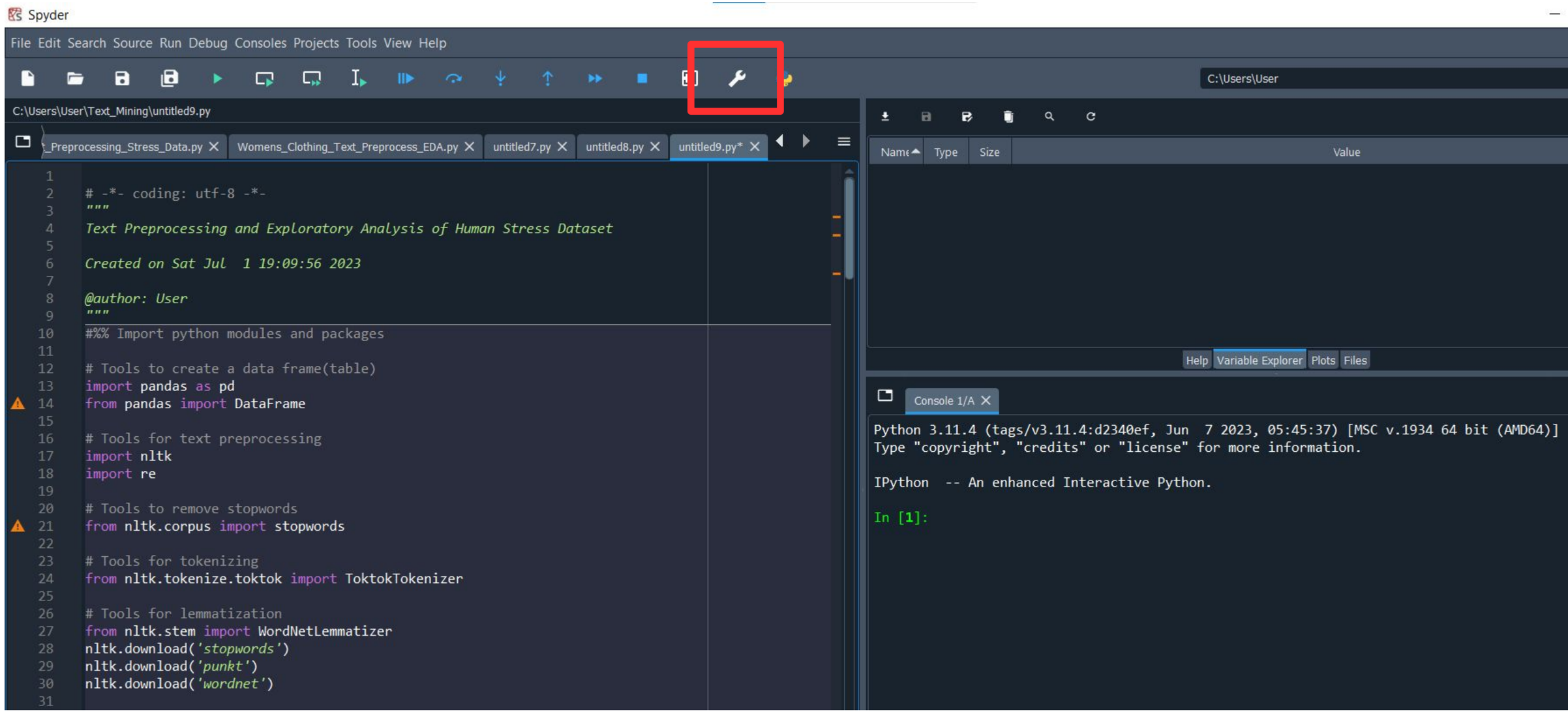


Downloading Python



Connecting Python to Spyder

Click the Preferences button

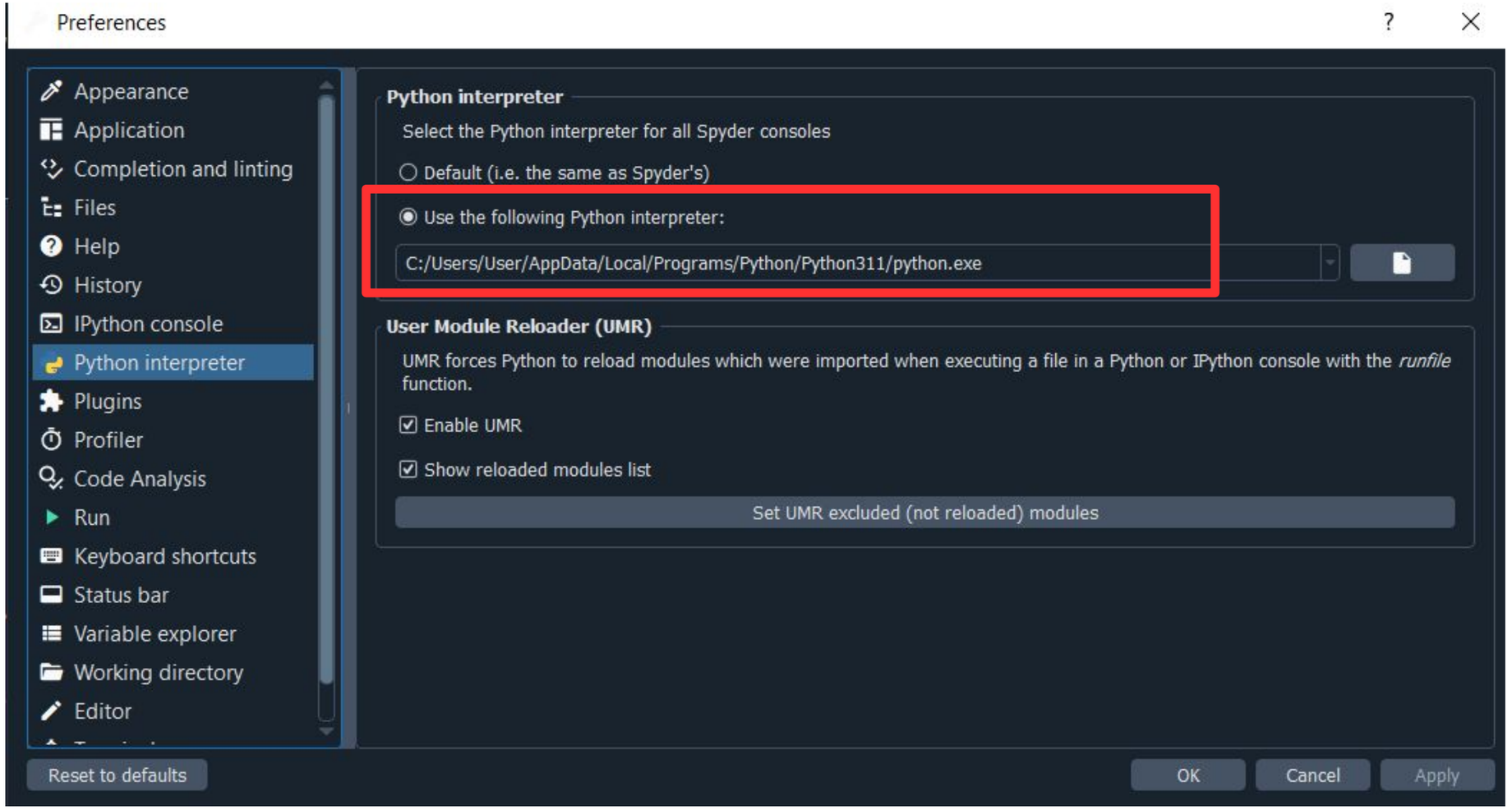



```
# -*- coding: utf-8 -*-
"""
Text Preprocessing and Exploratory Analysis of Human Stress Dataset

Created on Sat Jul  1 19:09:56 2023

@author: User
"""
#%% Import python modules and packages

# Tools to create a data frame(table)
import pandas as pd
from pandas import DataFrame

# Tools for text preprocessing
import nltk
import re

# Tools to remove stopwords
from nltk.corpus import stopwords

# Tools for tokenizing
from nltk.tokenize.toktok import ToktokTokenizer

# Tools for lemmatization
from nltk.stem import WordNetLemmatizer
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')
```

Connecting Python to Spyder

Click Python Interpreter

Click “USe the following Python Interpreter



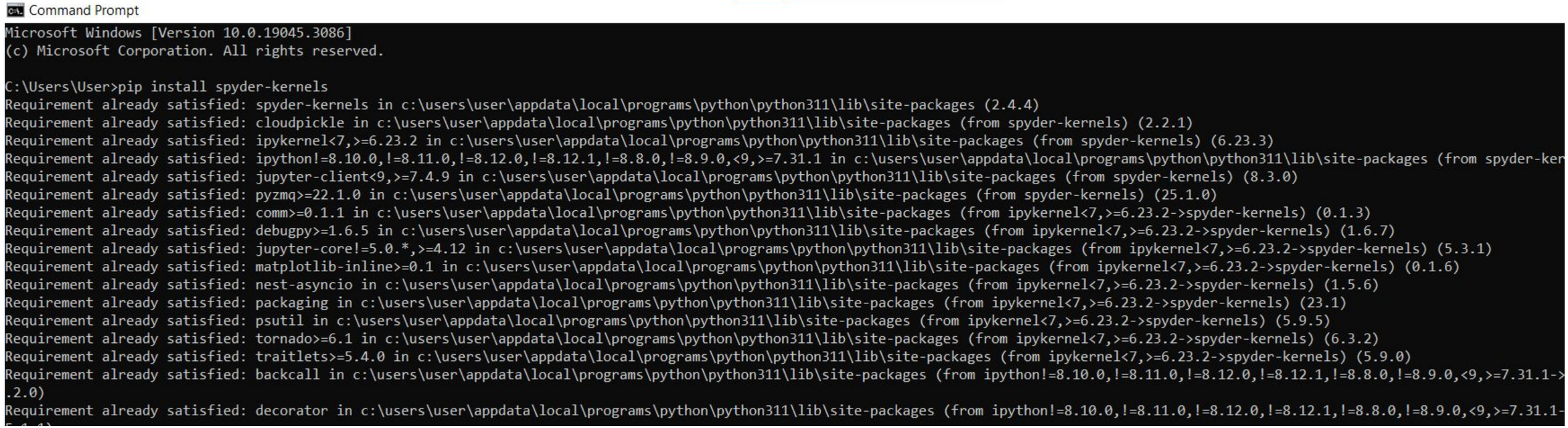
Connecting the Terminal

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

Type… “pip install spyder-kernels”

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

Restart Spyder



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.

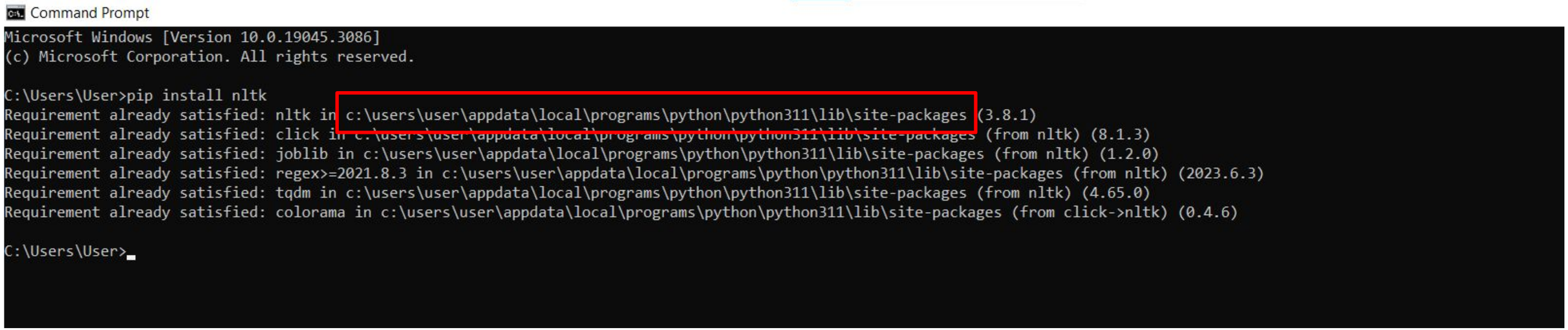
Install nltk

Windows:

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

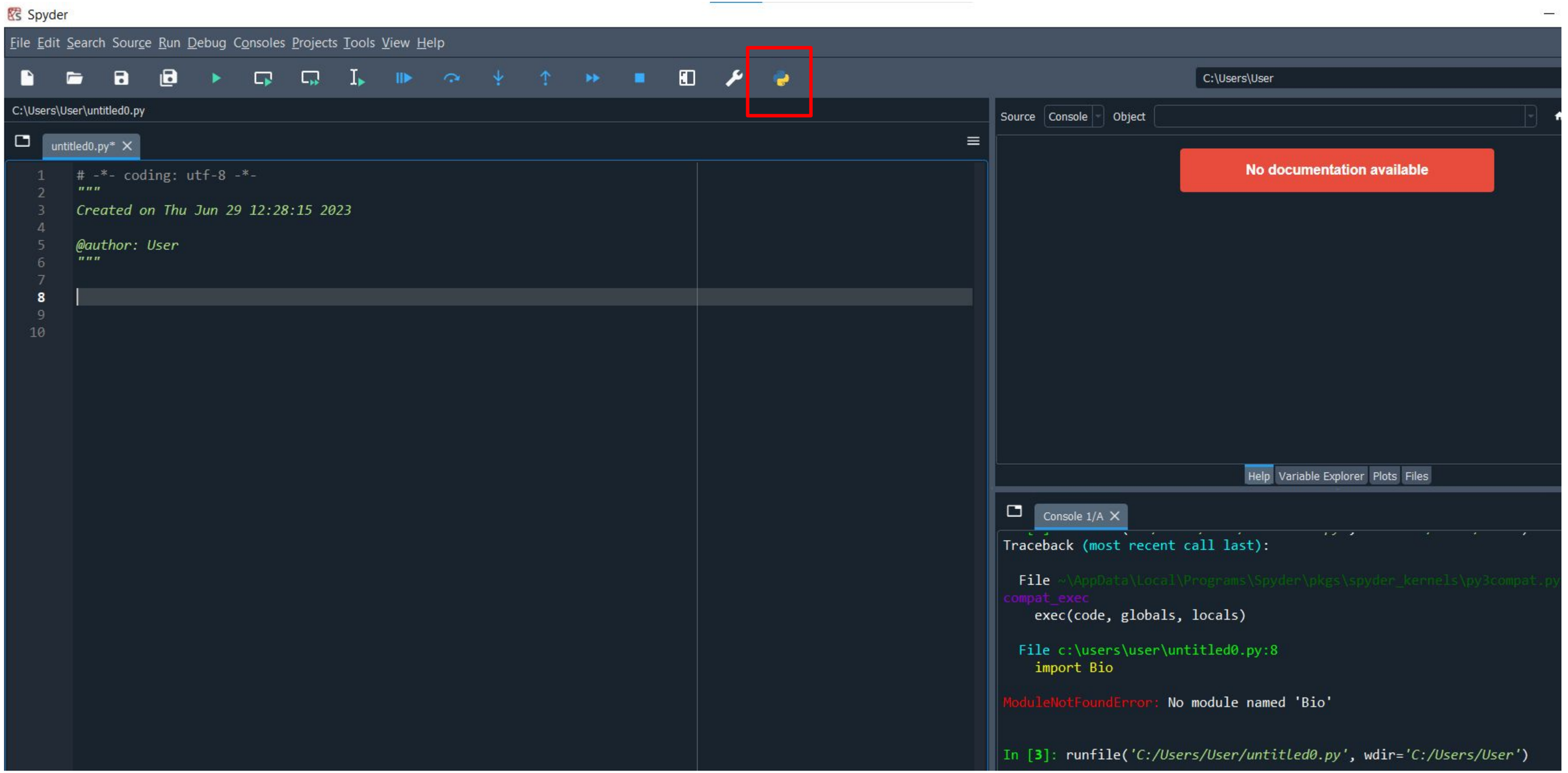
macOS:

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

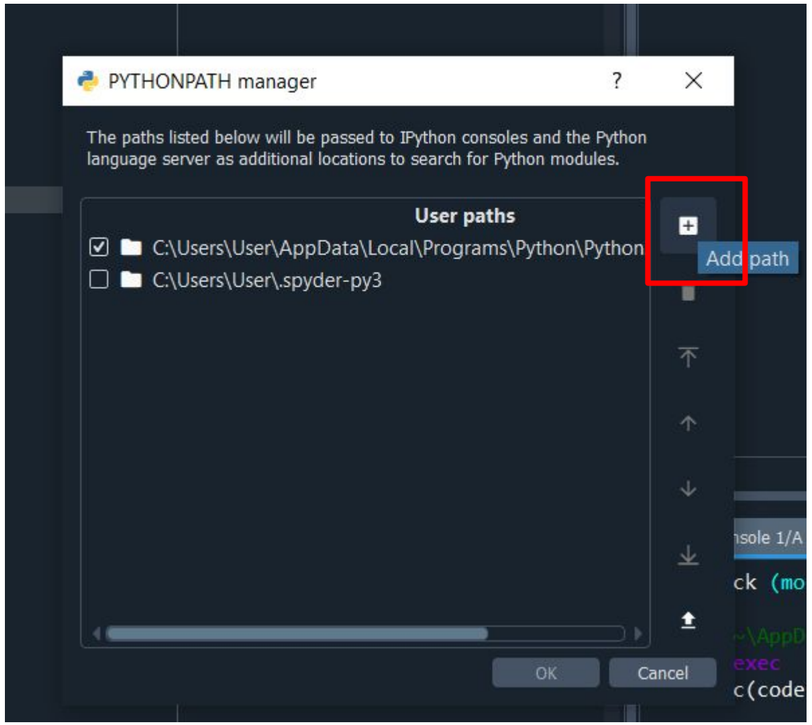


Copy the path that the package has been downloaded.

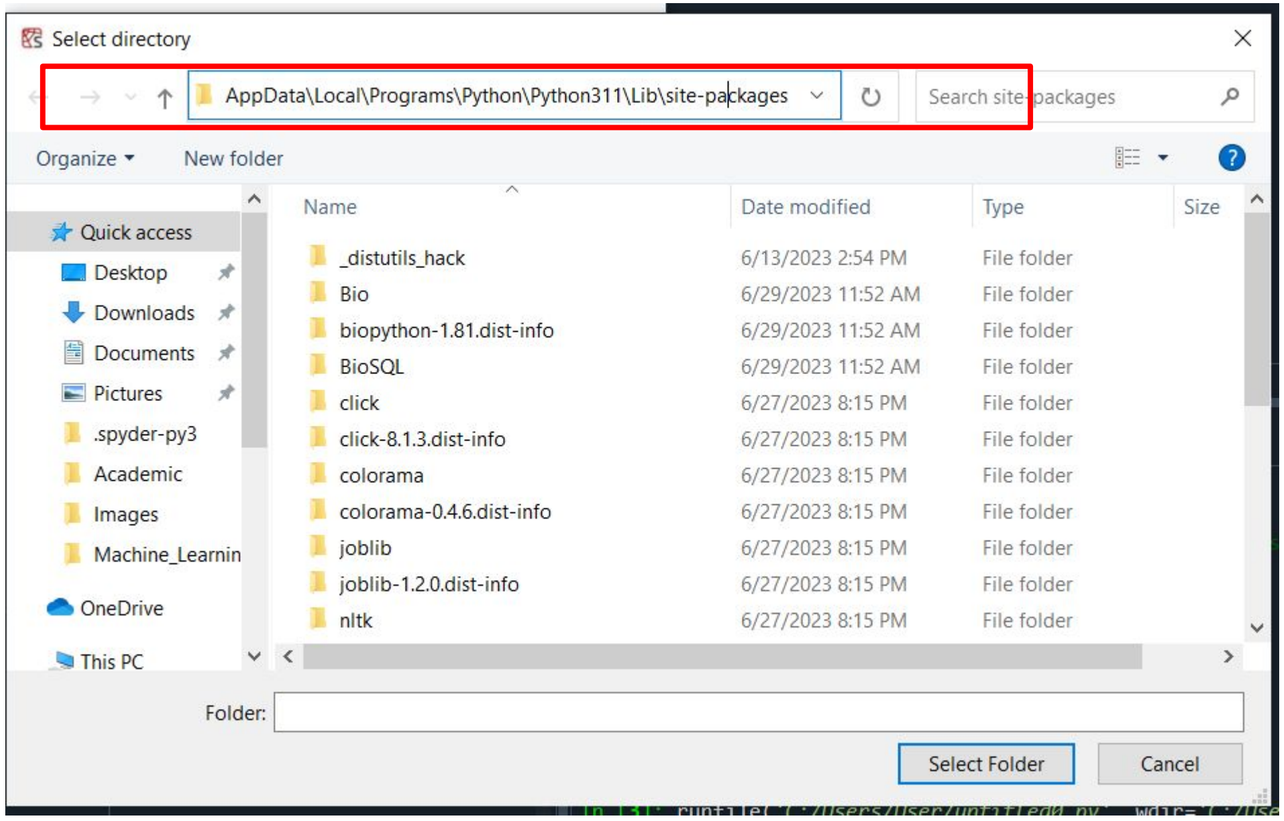
Go to Spyder. Click on the python icon.



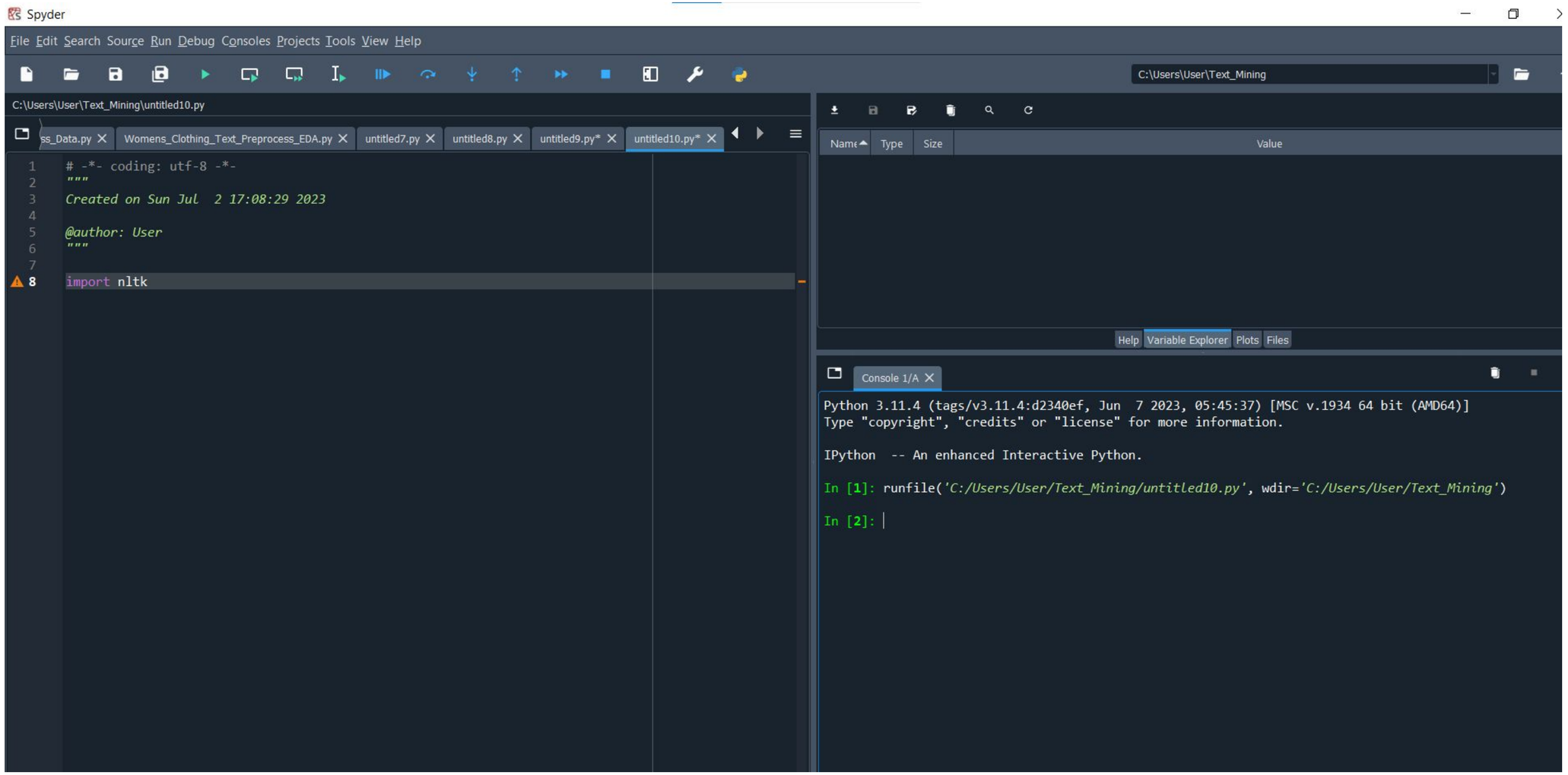
This window should pop up. Click the + button to Add path.



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



To double check that everything works well...



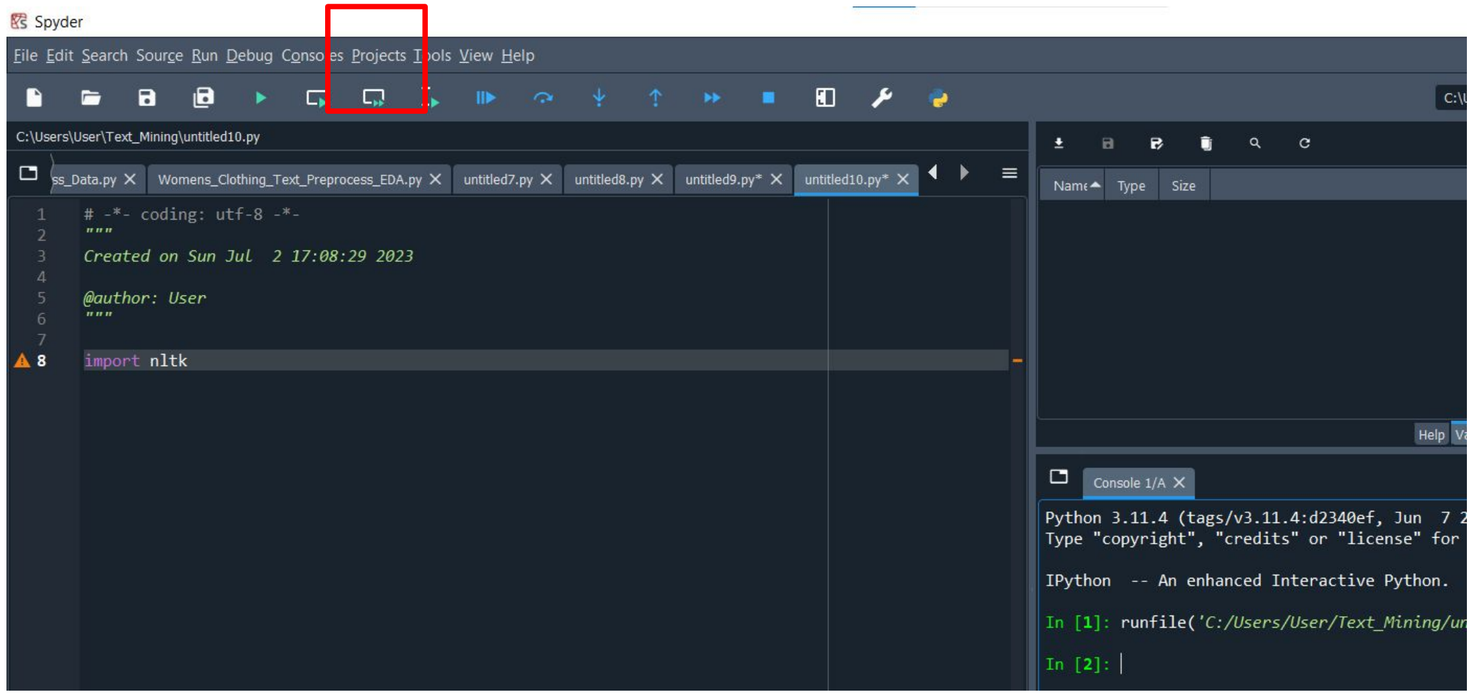
Creating a project in Spyder

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.

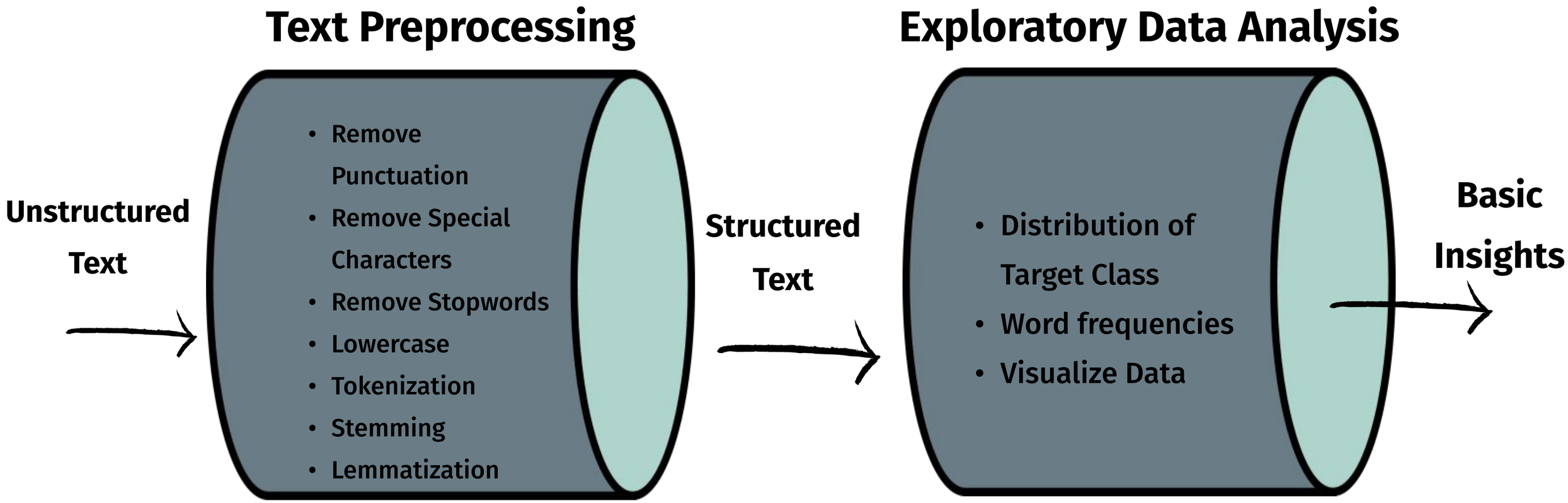
Put files in your Text_Mining project folder

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