Đây là tài liệu được tham khảo từ khóa học "Ứng dụng Phân tích dữ liệu và Trí tuệ nhân tạo với Python" trên web site: https://xmyworkspace.com/learn/course/applied-data-analysis-and-ai-with-python

Tài liệu chỉ dùng để chia sẻ cho các học viên của lớp học do giảng viên Lê Ngọc Thạch trực tiếp giảng dạy.

Liên lạc | Tài trợ

Lê Ngọc Thạch	
(+084) 0908 550 642	facebook.com/ThachLN
Website: https://xmyworkspace.com/learn/sponsor	

Bài 2: Ngôn ngữ lập trình Python

Để thực hành và trải nghiệm các nội dung trong sách này thì các bạn cần làm quen với Ngôn ngữ thống kê hoặc Ngôn ngữ lập trình trong máy tính và vài công cụ phần mềm. Phần này tôi sẽ giới thiệu cho các bạn ngôn ngữ Python vừa đủ để các bạn trải nghiệm các khái niệm về thống kê, về kiểu dữ liệu đã học trong ngày hôm nay.

Biến (variable) và Đối tượng (Object)

Nếu bạn đã học lập trình thì Variable là một cái tên dùng để chỉ một vùng nhớ trong máy tính. Để đơn giản, bạn hãy tưởng tượng cái máy vi tính giống như não người, trong đó có vùng nhớ (memory) để lưu thông tin tạm thời (lúc máy tính đang bật). Một variable được xem như một cái ô nhớ để chứa một giá trị nào đó.

Hình bên dưới là một thiết bị điện tử có trong máy tính của các bạn. Nó là một bản mạch gồm nhiều con chip có thể lưu trữ lại thông tin (bao gồm cả dữ liệu và lệnh) trong lúc máy tính có điện. Mọi người thường gọi ngắn gọn nó là thanh RAM.

Hình 3: Thanh RAM – nơi lưu "Trí nhớ" tạm thời của máy tính

Để các bạn hiểu hơn một chút về việc khai thác bộ nhớ của máy tính thì hãy tưởng tượng làm cách nào mà bạn bắt cái máy tính của bạn nhớ thông tin của một người bạn thân gồm các thông tin như sau:

Tên	Lê Ngọc Thạch
Chiều cao	165 cm
Cân nặng	72.5 kg
Giới tính	Nam
Ngày sinh	29/9/1977
Các chữ số yêu thích	1, 2, 5, 10, 20, 50, 100
Các môn thể thao yêu thích	Bóng bàn, bóng đá, Quần vợt

(Bạn có thể thay bằng thông tin của chính mình cho chính xác hơn nhé!)

Mỗi thông tin ở cột bên trái được gọi là một **biến** (variable). Bạn tưởng tượng là trong thanh RAM ở phần trước có rất nhiều ô nhỏ li ti. Mỗi ô nhỏ như vậy máy tính (*cụ thể các phần mềm mà chúng ta sẽ thực hành ở phần tiếp theo*) được đặt cho một cái tên (name) – gọi là **tên biến** (variable name). Mỗi biến như vậy sẽ có một vùng nhớ khác nhau để chứa thông tin. Để đơn giản cho máy tính thì chúng ta nên sử dụng tên tiếng Anh để đặt cho tên biến.

Tên biến nên gồm các **kí tự chữ cái thường, chữ cái HOA**, **dấu gạch chân** (_) và có thể có kí số (ở giữa hoặc ở cuối tên biến). Để thống nhất cho các bạn khi thực hành thì tôi sử dụng quy trước theo thông lệ chung như sau:

- Tên biến bắt đầu bằng chữ thường.
- Kí tự Hoa và thường được hiểu là 2 kí tự khác nhau. *Ví dụ tên biến là fullName sẽ khác với tên biến là FullName. Tức là có hai vùng nhớ khác nhau để chứa thông tin của 2 biến này.*
- Tên biến phải ngắn gọn và gợi nghĩa.
- Khi tên biến gồm nhiều từ ghép lại (như Full name – 2 từ trong ví dụ trên) thì hãy viết Hoa kí tự của từ tiếp theo.

Để mô tả thông tin trong ví dụ trên thì chúng ta có thể tự quy định tên biến như bảng sau:

Thông tin	**Tên biến**
Họ và Tên	fullName
Chiều cao	height
Cân nặng	weight
Giới tính	sex
Ngày sinh	birthday
Các chữ số yêu thích	favorNumbers
Các môn thể thao yêu thích	favorSports

Trên đây là thông tin của một người, để mô tả thêm một người bạn nữa thì bạn phải làm sao?

Bạn có thể đặt thêm một loạt biến nữa như: fullName1, height1, …Tức là bạn thêm số thứ tự phía sau để có bộ biến mới cho người mới. Tuy nhiên cách này không hay. Giới khoa học máy tính đưa ra khái niệm **Object** để giúp các bạn giải quyết nhu cầu này.

Object là một khái niệm gom nhiều loại thông tin để mô tả một vật, một người hay nói chung là một đối tượng nào đó. Nói cụ thể hơn là Object sẽ chứa trong

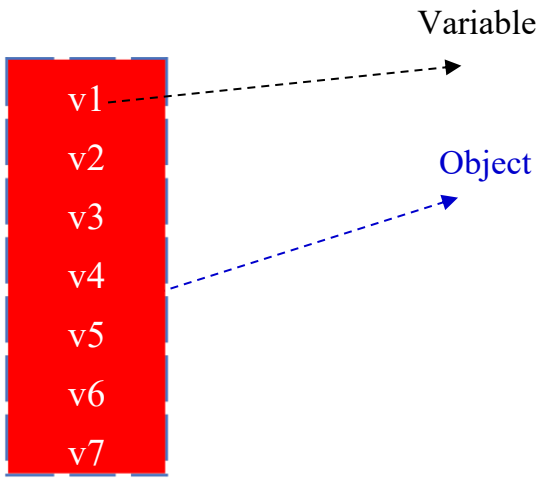
nó nhiều biến. Chúng ta mô tả lại ví dụ trình bày thông tin cho người bạn "Thạch" của chúng ta ở trên dưới dạng một object như sau:

Object: **myFriendThach**	
fullName	Lê Ngọc Thạch
height	165 cm
weight	72.5 kg
sex	Nam
birthday	29/9/1977
favorNumbers	1, 2, 5, 10, 20, 50
favorSports	Bóng bàn, bóng đá, Quần vợt

Trong bảng trên xuất hiện từ **myFriendThach**, đây là một cái tên (name) được máy tính chỉ định (hoặc là **trỏ tới**) vùng nhớ của tất cả các thông tin về bạn Thạch.

Như vậy đến đây bạn biết được khái niệm biến (**variable**) là một cái tên (name) trỏ tới một vùng nhớ chứa thông tin cơ bản nào đó của bạn Thạch (như tên, cân nặng, v.v…). Toàn bộ các biến liên quan đến bạn Thạch được gom lại trong một vùng nhớ (đương nhiên là rộng hơn) gọi lại **Object**.

Hình minh họa bên dưới gồm 7 ô nhớ tương ứng với 7 biến để mô tả thông tin về bạn Thạch (kí hiệu v1 đến v7 tương ứng với fullName…favorSports). Hình chữ nhật màu xanh được bao gởi đường đứt nét được gọi là một vùng nhớ cũng được đặt tên là một đối tượng (Object) với tên là myFriendThach.



Hình 4: Minh họa khái niệm biến (Variable) và đối tượng (Object)

Variable có nghĩa là gì?

📖 Tra tự điển

Nếu tra tự điển Oxford thì variable có thể là danh từ, có thể là tính từ.

☞ Tính từ variable: *able to be changed or adapted* (có thể được thay đổi hoặc điều chỉnh)

☞ Danh từ variable: *an element, feature, or factor that is liable to vary or change* (một yếu tố, một nét đặc trưng, hoặc một nhân tố có khả năng **biến đổi** hoặc **thay đổi**).

Cũng trong Oxford, variable được định nghĩa trong lĩnh vực Computing (điện toán) như sau: *a data item that may take on more than one value during the runtime of a program* (một phần tử dữ liệu có thể mang một hoặc hơn một giá trị trong suốt thời gian thực thi của chương trình).

Như vậy chữ variable có hai nghĩa mà các nhà khoa học máy tính và dịch giả Việt Nam đã dùng từ "biến" đã phản ánh đầy đủ rõ khái niệm "biến" trong máy tính.

Cụ thể là từ **vary** có hàm ý là có thể **biến đổi** thành đối tượng khác. Đối tượng khác ở đây có nghĩa là bản chất thông tin thay đổi hẳn. Chữ **change** có hàm ý là thay đổi giá trị của ô nhớ. Tức là bản chất, loại thông tin không thay đổi, mà chỉ thay đổi về nội dung, về giá trị của chúng.

Ví dụ:

Biến **height** đang có giá trị là 72.5 thì có thể được thay đổi thành một giá trị khác (tùy theo ngữ cảnh, thời gian như là đo lại tại một thời điểm khác) như là 71, 70 (chúng ta hiểu đơn vị là kg). Sự thay đổi này gọi là **change**.

Tuy nhiên, vì lý do nào đó trong ứng dụng phần mềm chúng ta muốn lưu trữ thông tin không phải là chiều cao nữa mà muốn lưu giá trị là một chức vụ cao nhất mà người đó đã từng làm. Tức là height sẽ được lưu giá trị là một **tên của chức vụ** (chứ không là một con số phản ánh chiều cao nữa). Lúc này biến height được biến đổi từ mục đích lưu con số phản ảnh chiều cao thành một tên phản ảnh chức vụ cao nhất. Cái này gọi là **vary** theo nghĩa trong tự điển Oxford.

Sau khi bạn hiểu được khái niệm Variable rồi thì câu hỏi tiếp theo là làm sao thiết lập giá trị cho biến. Cụ thể như thiết lập giá trị cho các ô nhớ từ v1 đến v2 trong hình 4.

Để làm được việc này thì bạn cần học thêm khái niệm gán (assign) trong phần tiếp theo.

Lệnh gán (assign)

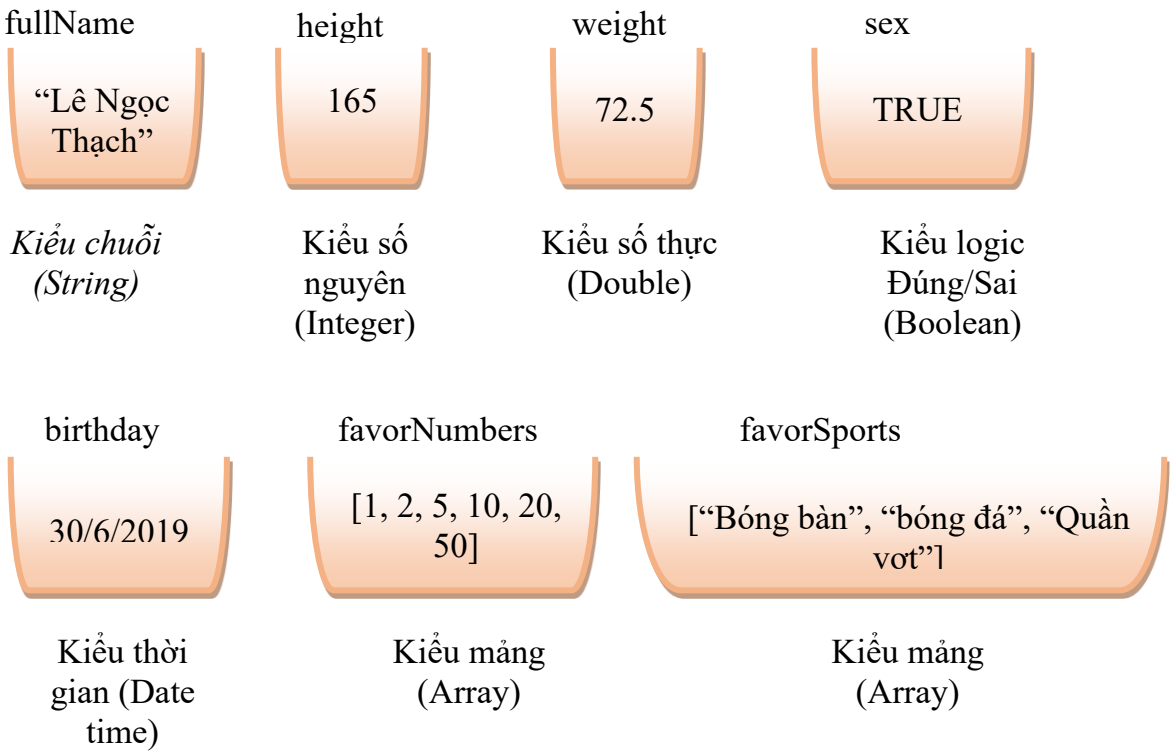
Hình bên dưới minh họa các variable có tên `level`, `score`, `name`, `birthday` tương ứng với các ô nhớ (hãy xem như là một cái thùng) chứa bên trong nó các thông tin tương ứng.

Để thiết lập thông tin (hay còn gọi là dữ liệu) vào biến thì sử dụng phép gán (assign). Python dùng chung dấu bằng (=) để thực hiện phép gán.

Trong Python, phép gán có thể sử dụng dấu mũi tên (gồm dấu bé hơn và dấu trừ: <-

Để thuận tiện thì trong tài liệu này tôi sẽ dùng dấu = làm phép gán.

> Để gán một giá trị cho một biến trong Python và Python thì dùng dấu bằng "=".
>
> Vd:
>
> name = "Thạch"
>
> weight = 70



Hình 5: Minh họa biến (variable)

Code Python minh họa:

```
fullName = 'Lê Ngọc Thạch'
height = 165
weight = 72.5
sex = True
import datetime
birthday = datetime.datetime.strptime('30/6/2019',
'%d/%m/%Y')
```

```
favorNumbers = [1, 2, 5, 10, 20, 50]
favorSports = ['Bóng bàn', 'Bóng đá ', 'Quần vợt']
```

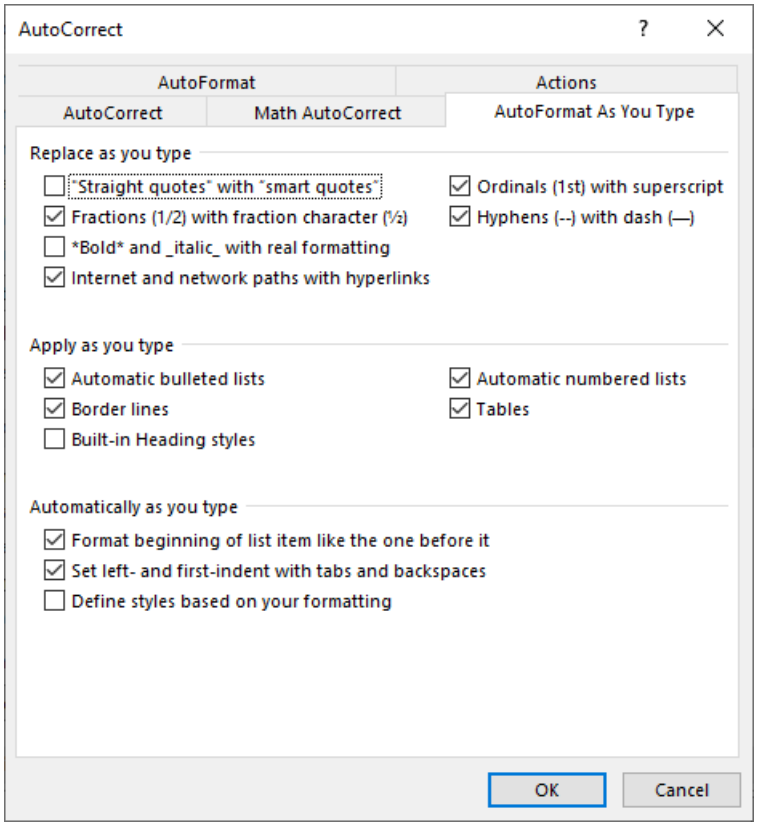
Chú ý:

- Dữ liệu dạng chuỗi thì bắt đầu và kết thúc bởi kí tự dấu nháy đơn hoặc dấu nháy đôi. Thông thường 2 dấu này nằm chung trong một phím phía bên trái phím Enter.

Nhấn phím nháy đơn sẽ nhanh hơn là nhấn phím nháy đôi (phải kèm thêm phím Shift). Vì vậy trong tài liệu này tôi sẽ dùng dấu nháy đơn để bao đóng các dữ liệu dạng chuỗi (string), hay còn gọi là văn bản (text).

Khi chúng ta soạn tài liệu bằng MS Word thì các kí tự nháy đơn và nháy đôi được MS Word thay bằng cách ký tự khác trông đẹp hơn. Điều này sẽ gây ra lỗi nếu chúng ta sao chép mã nguồn từ tài liệu MS Word ra phần mềm thực thi lệnh Python hoặc các phần mềm lập trình nói chung.

Để tắt chức năng thay thế thông minh này trong MS Word, bạn tìm vào chỗ cấu hình chức năng Auto Correct rồi bỏ chọn mục **"Straight quotes" with "smart quotes"** (tùy theo phiên bản của WS Word thì giao diện có thể khác).



- Kiểu luận lý (logical) thì các giá trị [1] đúng/sai là True/False, chỉ viết Hoa kí tự đầu tiên.

- Đối với kiểu dữ liệu ngày tháng thì phức tạp hơn một chút. Việc chúng ta thấy hoặc viết vào máy tính như 30/6/2019 (ngày 30 tháng 6, năm 2019) thì đó là chuỗi các kí tự có ý nghĩa đối với chúng ta. Thật ra máy tính không hiểu nó là giá trị về thời gian.

    - Để Python hiểu được giá trị thời gian thì dùng thư viện datetime (`import datetime`) và viết lệnh như sau:

```
datetime.datetime.strptime('30/6/2019', '%d/%m/%Y')
```

Chức năng của lệnh này là báo cho phần mềm Python biết chuỗi văn bản "30/6/2019" cần phải chuyển sang dạng thời gian với quy định trong tham số định dạng '%d/%m/%Y'.

Định dạng này gồm 3 thành phần ngăn cách nhau bởi dấu xuyệt phải /:

%d cho biết thành phần đâu tiên có nghĩa là ngày (day)

[1] Đối với các bạn học lập trình thì nói chính xác hơn là hằng (constant)

%m sau dấu / đầu tiên là tháng (month)

%Y sau dấu / thứ hai là năm (Year). Chú ý là chữ Y viết hoa nhé.

- Kiểu mảng, còn gọi là dãy trong Python.

  Trong Python sử dụng cú pháp [] để liệt kê các phần tử của mảng cách nhau bởi dấu phẩy. Vd:

```
favorNumbers = [1, 2, 5, 10, 20, 50]
```

Bài 3: Ngôn ngữ Python và phần mềm Anaconda

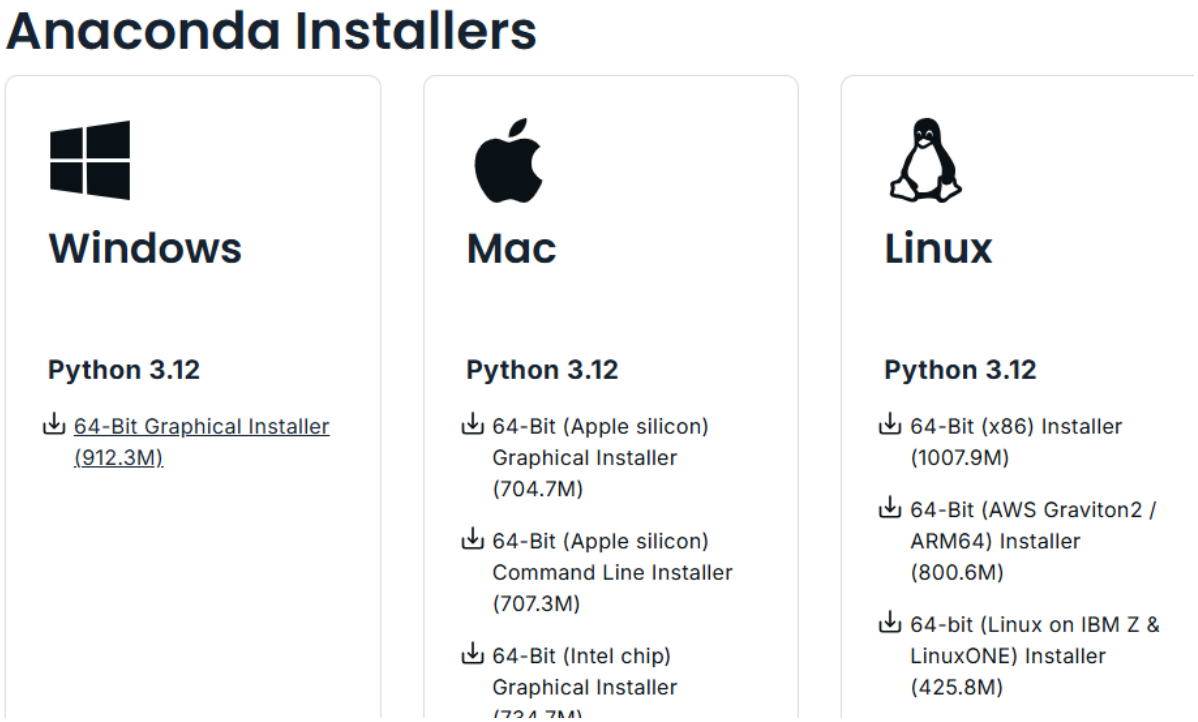
Anaconda

Đối với người mới bắt đầu làm quen với phân tích dữ liệu thì nên cài đặt phần mềm Anaconda tại địa chỉ "https://anaconda.com". Anaconda là bộ quản lý các gói phần mềm (package manager). Trong đó tập trung chủ yếu các gói phần mềm về R và Python. Anaconda miễn phí, dễ sử dụng, có thể chạy được trên các hệ điều hành phổ biến như Windows, Mac, Linux.

Anaconda phù hợp với mọi người để học, thực hiện phân tích dữ liệu, Máy học (Machine learning) bằng ngôn ngữ Python.

Cái đặt Anaconda

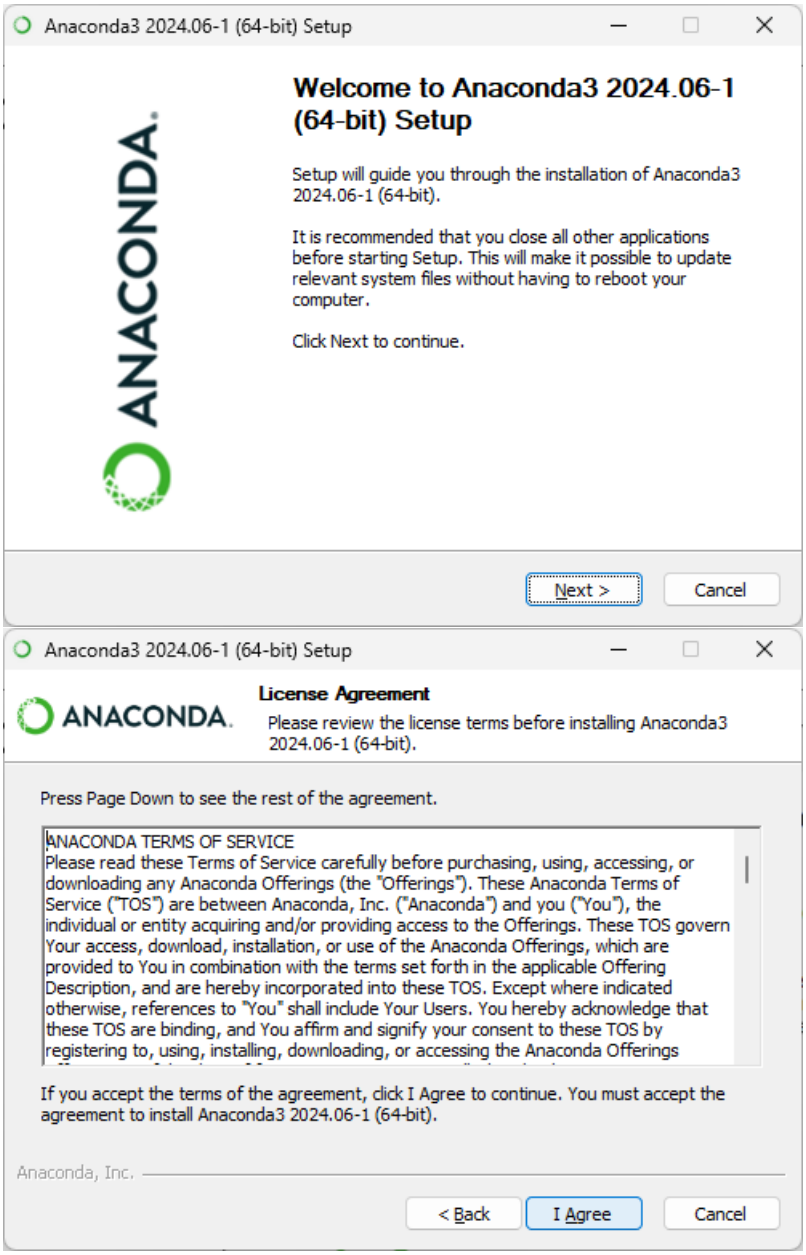
Vào trang "https://www.anaconda.com/download/success", bấm vào nút Download:

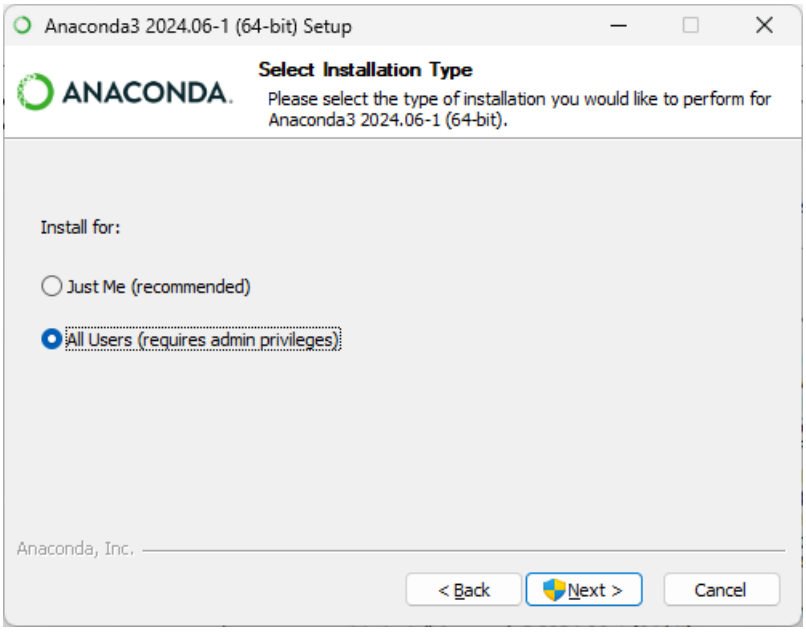


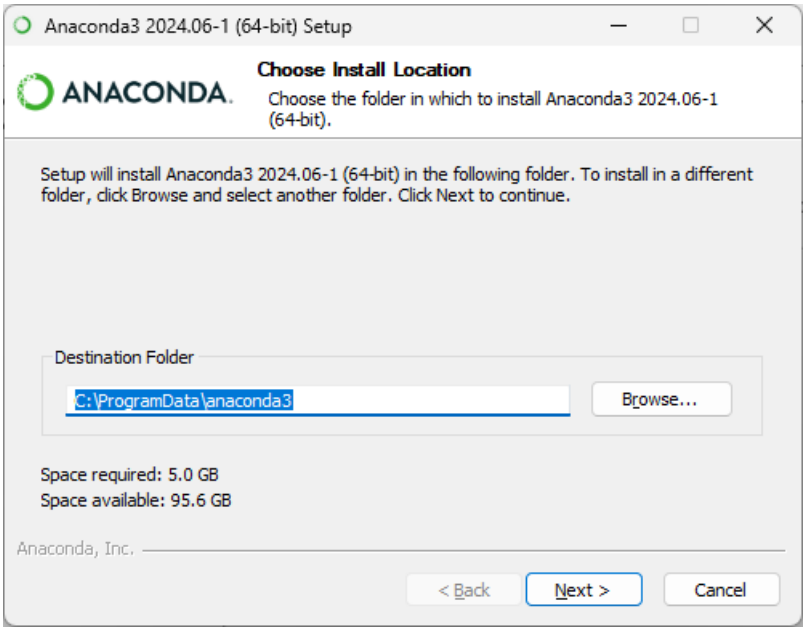
Sau đó bấm mục của gói cái đặt tùy theo máy của bạn. Tại thời điểm viết phần này thì Anconda cung cấp phiên bản mới nhất hỗ trợ Python 3.12.

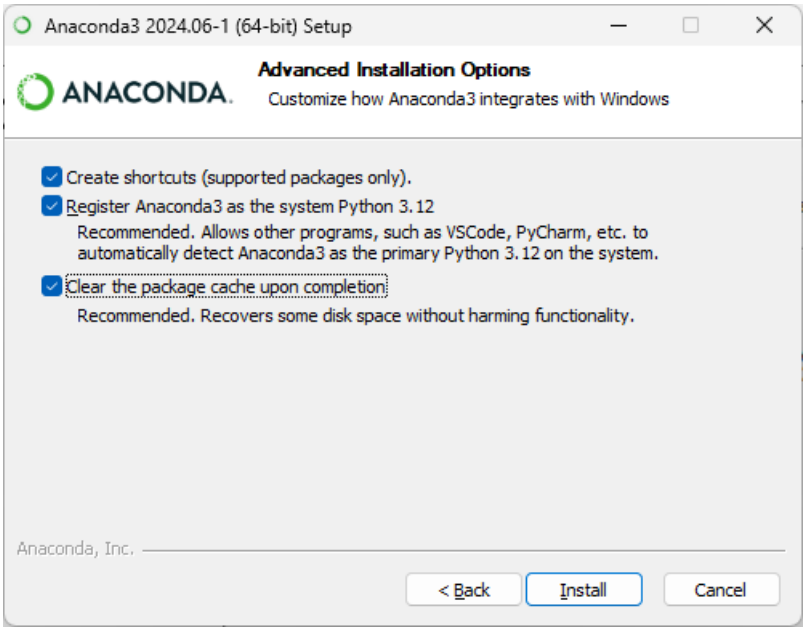
Tôi dùng phiên bản 3.12 "Anaconda3-2024.06-1-Windows-x86_64.exe (934 MB)" do máy tôi dùng Windows 64 bit. Tương tự nếu bạn dùng Window 32 bit, MacOS hoặc Linux thì tìm và vào link tương ứng để tải.

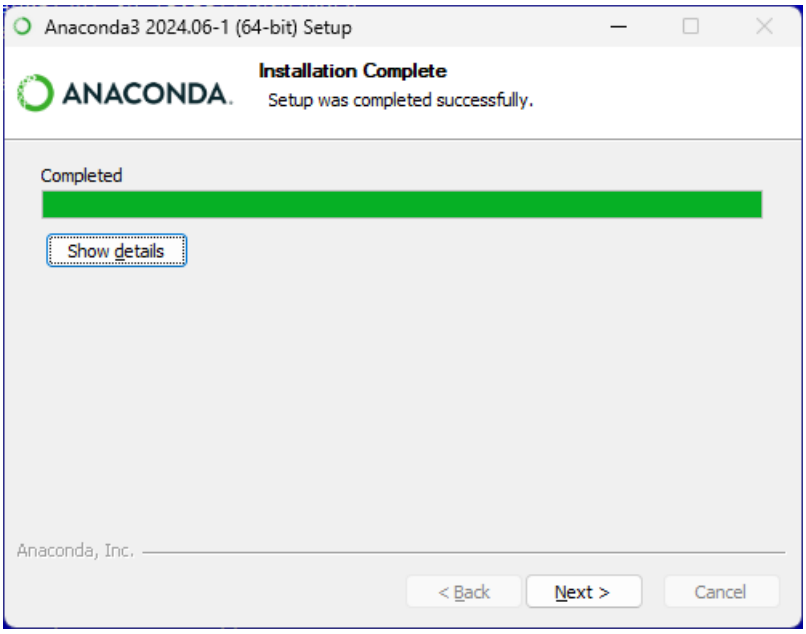
Quá trình cài đặt khá đơn giản. Cơ bản là cứ bấm "Next" và "Agree" rồi làm theo hướng dẫn.


Anaconda3 2024.06-1 (64-bit) Setup
Welcome to Anaconda3 2024.06-1 (64-bit) Setup
Setup will guide you through the installation of Anaconda3 2024.06-1 (64-bit).
It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.
Click Next to continue.
Next >
Cancel
Anaconda3 2024.06-1 (64-bit) Setup
License Agreement
Please review the license terms before installing Anaconda3 2024.06-1 (64-bit).
Press Page Down to see the rest of the agreement.
ANACONDA TERMS OF SERVICE
Please read these Terms of Service carefully before purchasing, using, accessing, or downloading any Anaconda Offerings (the "Offerings"). These Anaconda Terms of Service ("TOS") are between Anaconda, Inc. ("Anaconda") and you ("You"), the individual or entity acquiring and/or providing access to the Offerings. These TOS govern Your access, download, installation, or use of the Anaconda Offerings, which are provided to You in combination with the terms set forth in the applicable Offering Description, and are hereby incorporated into these TOS. Except where indicated otherwise, references to "You" shall include Your Users. You hereby acknowledge that these TOS are binding, and You affirm and signify your consent to these TOS by registering to, using, installing, downloading, or accessing the Anaconda Offerings
If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install Anaconda3 2024.06-1 (64-bit).
Anaconda, Inc.
< Back
I Agree
Cancel

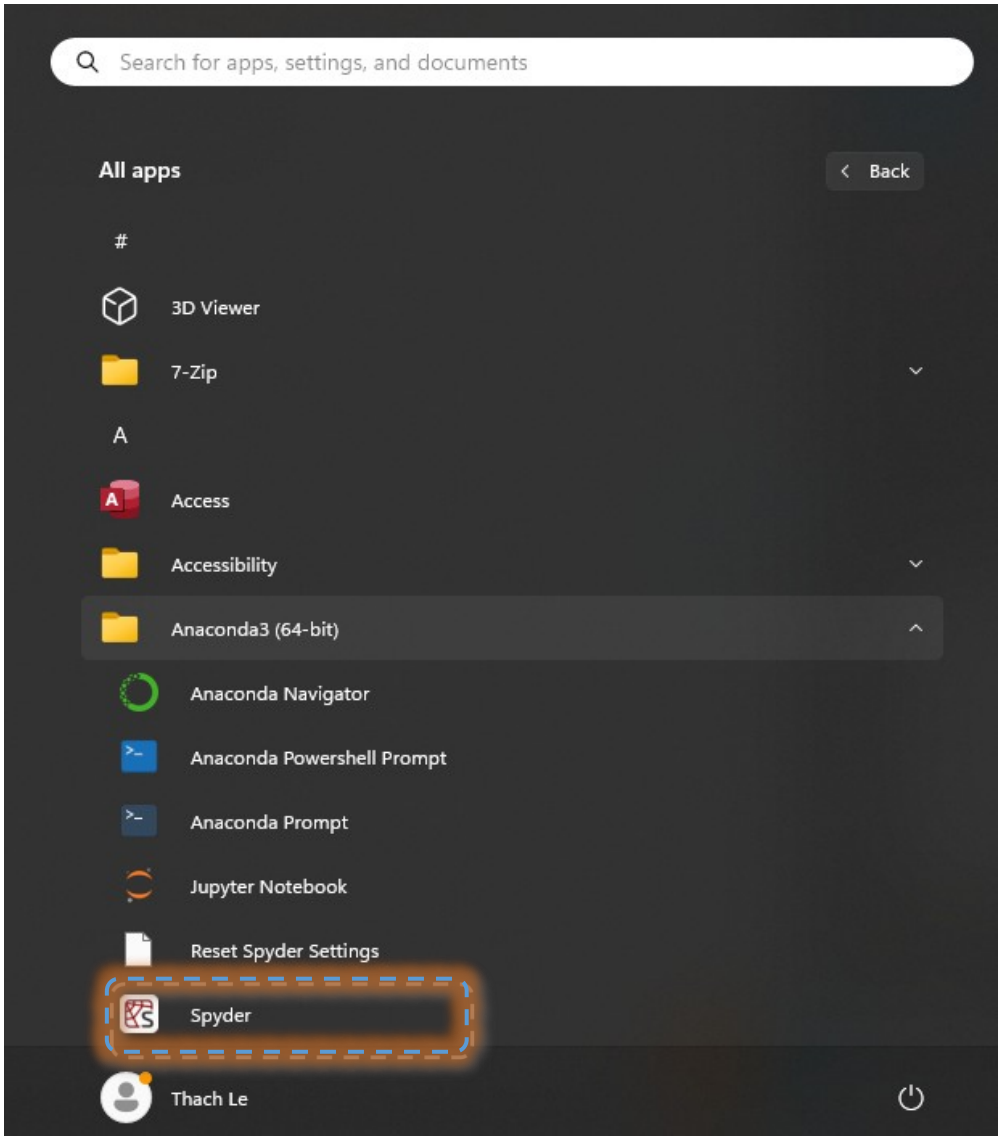

Anaconda3 2024.06-1 (64-bit) Setup
ANACONDA.
Select Installation Type
Please select the type of installation you would like to perform for Anaconda3 2024.06-1 (64-bit).
Install for:
Just Me (recommended)
All Users (requires admin privileges)
Anaconda, Inc.
< Back
Next >
Cancel


Anaconda3 2024.06-1 (64-bit) Setup
ANACONDA.
Choose Install Location
Choose the folder in which to install Anaconda3 2024.06-1 (64-bit).
Setup will install Anaconda3 2024.06-1 (64-bit) in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Destination Folder
C:\ProgramData\anaconda3
Browse...
Space required: 5.0 GB
Space available: 95.6 GB
Anaconda, Inc.
< Back
Next >
Cancel





Sau khi cài xong, vào nút Start của Windows ở góc trái dưới màn hình hoặc bấm phím có hình cửa sổ ⊞ hoặc❖ (tùy bàn phím) bạn sẽ thấy biểu chương trình Spyder (Anaconda 3) như sau:



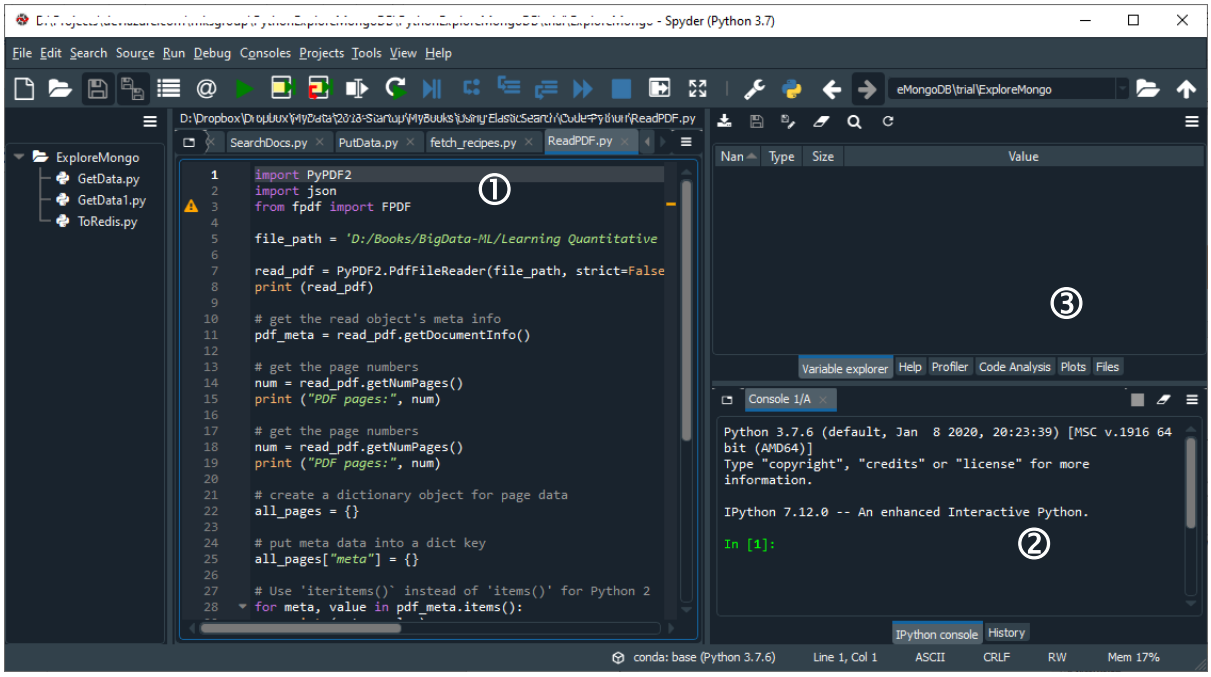
Đến đây bạn đã biết cách tải và cài đặt Anaconda Python 3. Bạn cũng nên thử khởi động Spyder (Anaconda3) và thoát nó, tắt máy tinh đi uống một ly café hoặc trà sữa tùy theo sở thích để tự thưởng cho mình.

Ngôn ngữ lập trình Python

Bạn có thể bắt đầu làm quen với ngôn ngữ Python và thực hành với các gói phần mềm trong bộ Anaconda đã cài đặt trong phần trước.

Sử dụng Spyder

Sau khi cài đặt Anaconda Python, hãy khởi động chương trình Spyder sẽ có giao diện như sau:



Spyder là phần mềm để viết mã lệnh Python được thiết kế bởi các nhà khoa học (scientists), các kỹ sư công nghệ (engineers) và các nhà phân tích dữ liệu (data analysts).

① Phần cửa sổ bên trái giúp bạn viết lệnh Python. Các lệnh này sẽ được lưu vào một file tạm trên máy tính của bạn (ví dụ thư mục trên máy tôi là "C\Users\Le Ngoc Thach"). Tên file untitled0.py có nghĩa là file chưa được đặt tên (untitled) đầu tiên (có thứ tự bắt đầu là 0), phần mở rộng sau dấu chấm là "py" - viết tắt của chữ Python.

② Phần của sổ "Console" ở góc phải dưới là nơi trình bày kết quả của lệnh khi các lệnh được thực thi (execute).

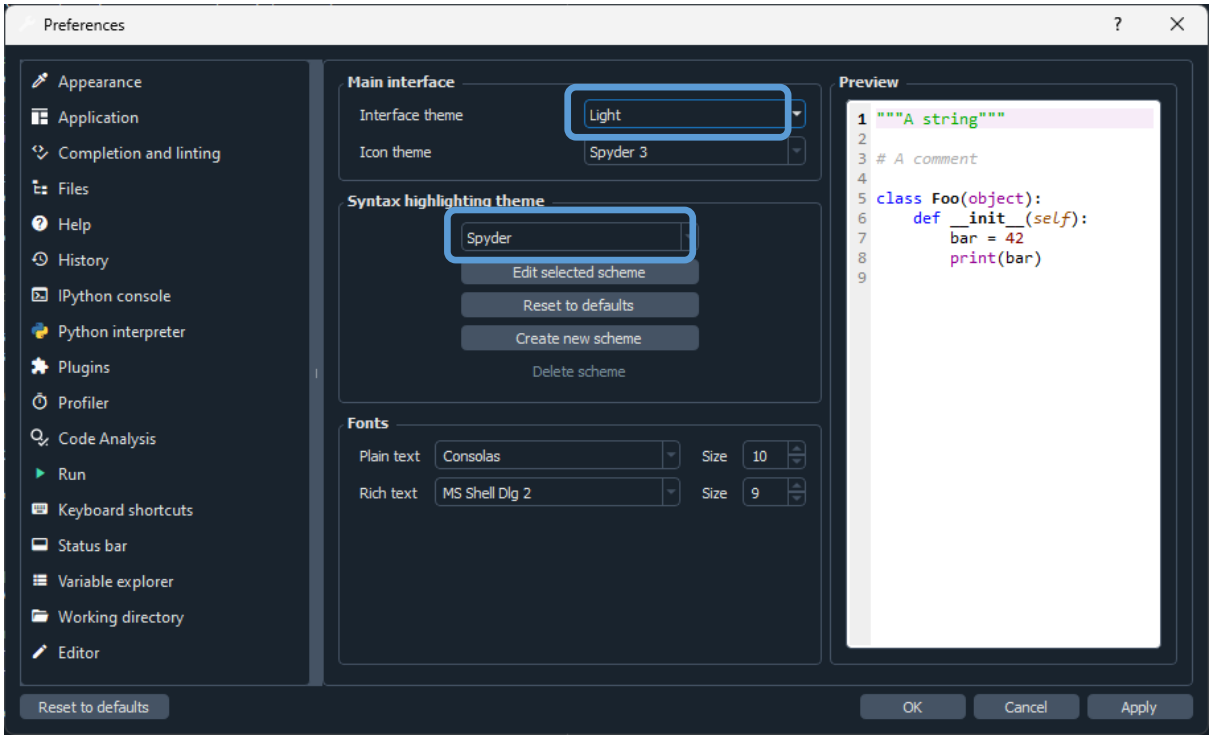
③ Phần cửa số ở góc phải trên có nhiều tab, trong đó 2 tab "Variable explorer" và "Plots". Variable explorer giúp bạn theo dõi các biến mà bạn đã khai báo (declare) trong cửa sổ lệnh bên trái khi các lệnh được thực thi. Plots giúp bạn xem kết quả vẽ biểu đồ.

Đổi theme

Mặc định thì Spyder từ phiên bản 4.x có giao diện đen xì như trên. Nếu bạn không quen thì đổi sang giao diện sáng (light) bằng cách vào menu Tools > Preference, chọn lại:

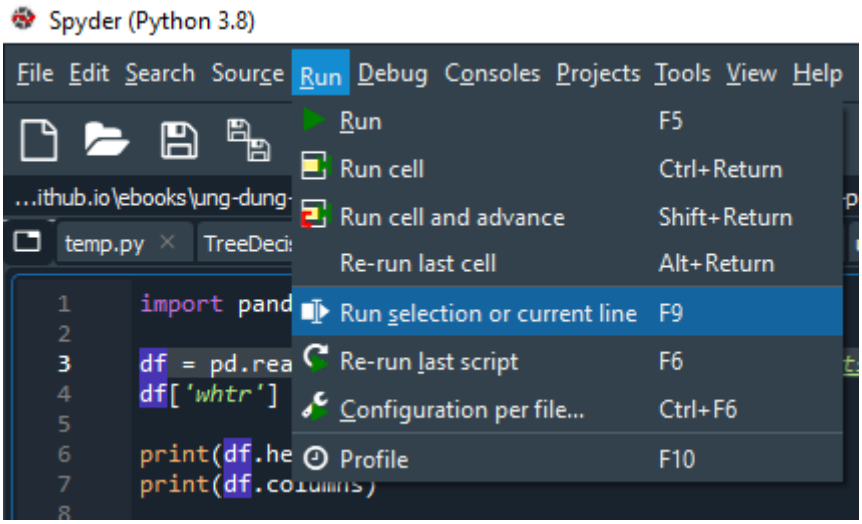
Interface theme: **Light**

Syntax highlighting theme, mục đầu tiên: **Spyder**



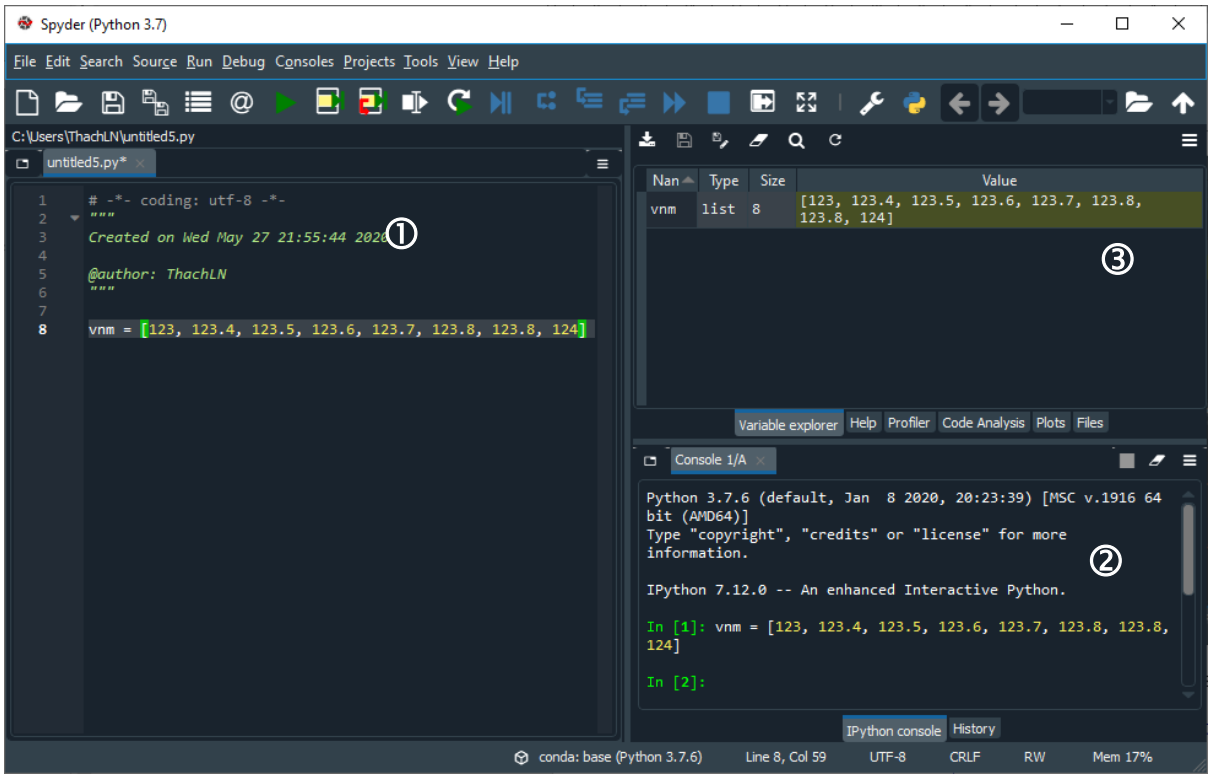
Thực thi lệnh

Chọn dòng lệnh cần thực thi, nhấn phím **F9**. Trường hợp không chọn dòng nào cả thì F9 sẽ thực thi dòng hiện tại có con nháy (cursor), sau đó con nháy sẽ nhảy đến dòng tiếp theo. Như vậy bạn có thể dùng F9 tại dòng đầu tiên của chương trình, vừa chạy từng lệnh vừa quan sát kết quả. Nếu bạn quên phím tắt thì có thể vào menu Run:



Ví dụ trong hình bên dưới khai báo một biến có tên **vnm** được gán (assign) bằng một mảng (array) gồm nhiều giá trị cách nhau bởi dấu phẩy. Cặp dấu móc vuông [] bao đóng array theo qui ước của Python.

```
vnm = [123, 123.4, 123.5, 123.6, 123.7, 123.8, 123.8, 124]
```



Trong cửa sổ bạn bôi dòng lệnh số 8 bằng các cách sau:

1) Dùng chuột bôi từ đầu đến cuối lệnh *bằng cách di chuyển con trỏ chuột đến trước biến vnm, bấm nút trái chuột giữ nguyên nút trái trong lúc di chuyển con chuột sang phải dòng lệnh – hướng di chuyển chuột theo hàng ngang đảm bảo con trỏ chuột lúc nào cũng nằm trên dòng lệnh. Khi con trỏ chuột đến cuối dòng lệnh bạn sẽ thấy dòng lệnh sẽ được bôi màu nền xanh như hình trên.*

2) Dùng phím Shift + Home: khi gõ lệnh xong thì con nháy đang ở cuối dòng lệnh. Bạn chỉ cần nhấn tổ hợp phí Shift + Home (tay trái nhấn và giữ nút Shift, sau đó tay phải nhấn phím Home rồi thả cả 2 tay ra khỏi bàn phím cùng lúc).

3) Dùng phím Shift + End: khi con nháy đang ở bất kỳ chỗ nào trên dòng lệnh, hãy gõ phím Home để đưa chon nháy về vị trí đầu tiên. Sau đó nhấn tổ hợp phí Shift + End (tay trái nhấn và giữ nút Shift, sau đó tay phải nhấn phím End rồi thả cả 2 tay ra khỏi bàn phím cùng lúc).

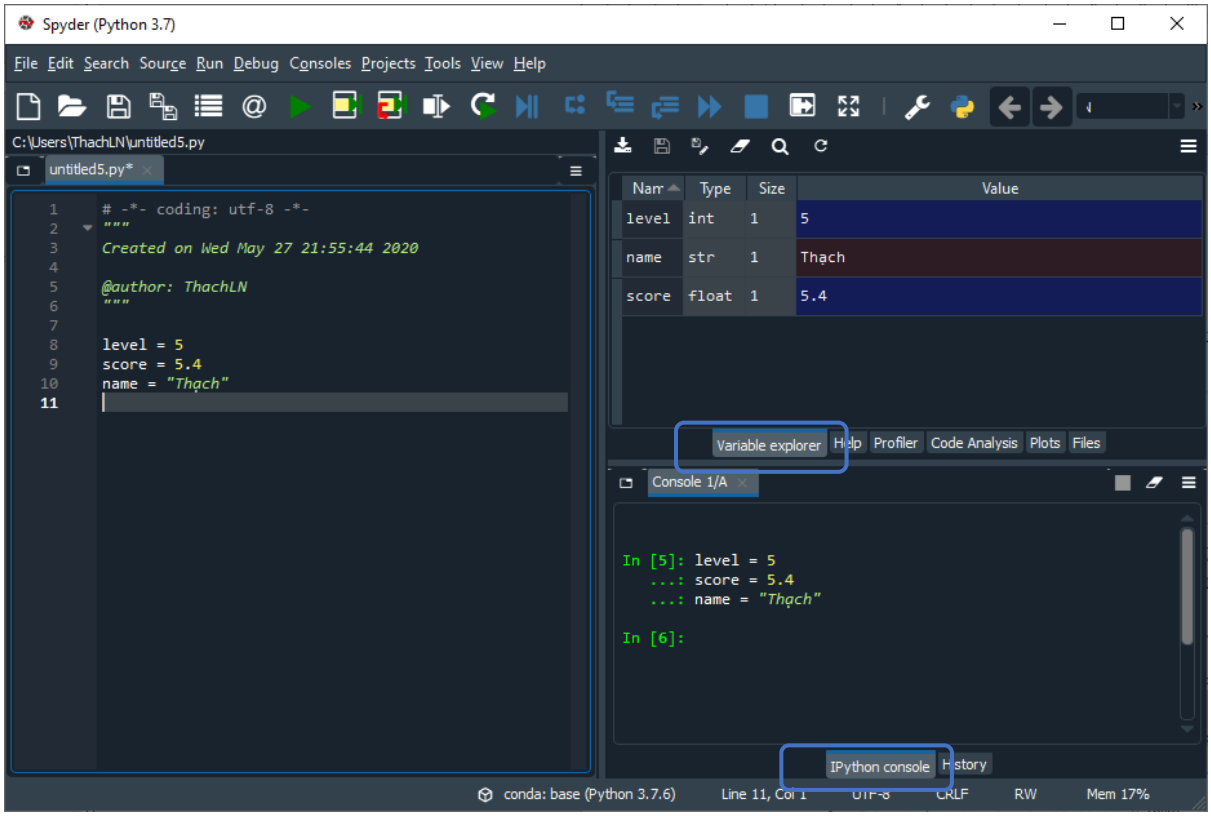
Thực hành phép gán

Hãy khởi động chương trình Spyder, mở file mới bằng cách nhấn `Ctrl + N`. Sau đó gõ 3 lệnh sau:

```
level = 5
```

```
score = 5.4
name = "Thạch"
```

Thực thi 3 dòng lệnh bằng cách bôi cả 3 dòng rồi nhấn phím F9. Quan sát giá trị các biến trong thẻ "Variable explorer" và quan sát các lệnh được thực thi trong của sổ "Console" ở góc phải dưới.



Sử dụng các gói phần mềm

Python cung cấp rất nhiều gói thư viện. Phần mềm Anaconda đã cài sẵn nhiều gói thư viện cơ bản. Khi nào cần sử dụng cần gói thư viện thì dùng lệnh `import` như sau:

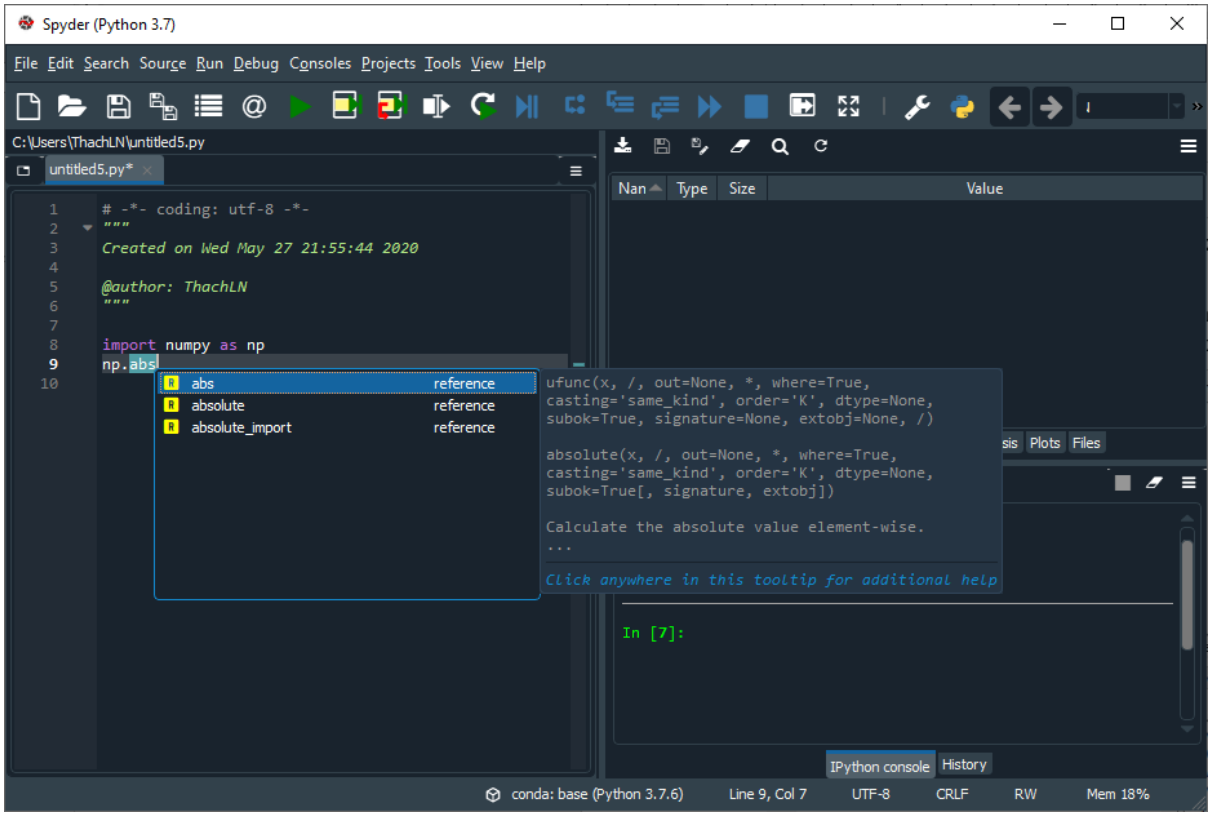
```
import <tên thư viện> as <tên viết tắt>
```

Ví dụ để sử dụng thư viện **numpy** thì sử dụng lệnh

```
import numpy as np
```

Tên viết tắt là do bạn quy định để thuận tiện khi viết lệnh. Dùng tên viết tắt này để cho mã nguồn gọn hơn.

Trong chương trình Spyder khi gõ lệnh **np,** sau đó nhấn Ctrl + Space thì bạn sẽ thấy các hàm của numpy hiển thị ra cho bạn dễ chọn hoặc dễ gõ tiếp.



Gọi hàm

Phần này sẽ giới thiệu các ví dụ về điểm số để bạn làm quen trong Python.

Trong Python thì một số hàm được cung cấp trong thư viện **NumPy**. Vì vậy bạn cần phải thực thi lệnh `import` như sau để bắt đầu sử dụng NumPy:

```
import numpy as np
```

Dùng cú pháp [] để khai báo danh sách điểm. Sau đó gán danh sách cho biến `scores` như sau:

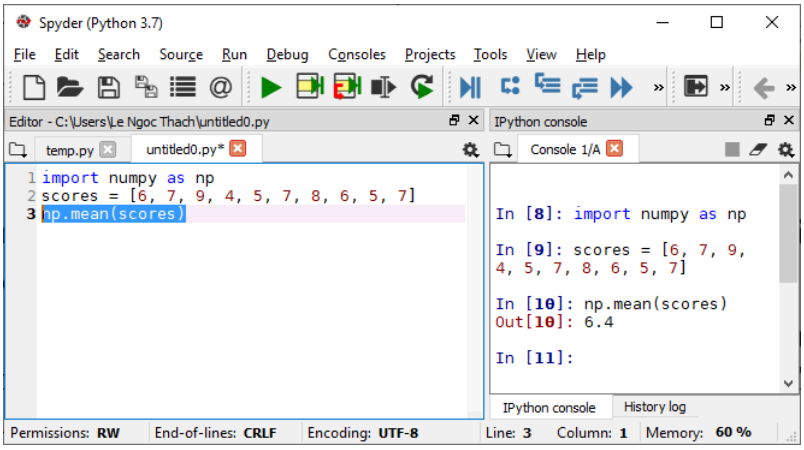
```
scores = [6, 7, 9, 4, 5, 7, 8, 6, 5, 7]
```

- Gọi hàm `mean` của thư viện `numpy` thông qua kí hiệu `np`:

```
np.mean(scores)
```

sẽ cho kết quả: 6.4

Cần nhắc lại một chút là trong lúc soạn thảo lệnh trong Spyder, để thực thi từng dòng lệnh thì bôi chọn từng dòng, nhấn Ctrl + Enter hoặc nhấn F9; trường hợp không bôi đoạn lệnh nào thì F9 sẽ thực thi dòng đang có con nháy. Sau đó theo dõi kết quả trong cửa sổ Console.



- Gọi hàm np.median(x):

```
np.median(scores)
```

sẽ cho kết quả: 6.5

- Gọi hàm np.sort(scores):

```
np.sort(scores)
```

sẽ cho kết quả: array([4, 5, 5, 6, 6, 7, 7, 7, 8, 9])

Phần mềm Spyder in ra kết quả có chữ array() và cặp dấu ngoặc [] để cho chúng ta biết đây là mảng.

- Gọi hàm np.var(x):

```
np.var(scores)
```

sẽ cho kết quả: 2.04

```
np.var(scores, ddof = 1)
```

sẽ cho kết quả: 2.2666666666666666

Trong Python để tính phương sai thì gọi hàm `var` của thư viện NumPy. Nếu không truyền tham số ddof thì coi như bằng 0.

ddof = 0: có nghĩa là tính Phương sai toàn bộ (Population variance)

ddof = 1: có nghĩa là tính Phương sai mẫu (Sample variable).

ddof viết tắt của Delta Degrees of Freedom.

- Gọi hàm np.std(x):

```
np.std(scores, ddof = 1)
```

sẽ cho kết quả: 1.505545305418162

- Gọi hàm np.quantile(a, q) để tính Bách phân vị:

```
np.quantile(scores, 0.5)
np.quantile(scores, 0.25)
np.quantile(scores, 0.75)
```

sẽ cho kết quả tương ứng của Bách phân vị 50%, 25%, 75%: 6.5, 5.25, 7.0

Bài tập thực hành

① Làm quen với Biến và Phép gán.

Trong bài 2 tôi có giới thiệu khái niệm Biến và Phép gán.

Đây là thời điểm để bạn mở Spyder thực hành các lệnh sau:

```
Python
import datetime
fullName = 'Lê Ngọc Thạch'
height = 165
weight = 72.5
sex = True

birthday = datetime.datetime.strptime('30/6/2019', '%d/%m/%Y')
favorNumbers = [1, 2, 5, 10, 20, 50]
favorSports = ['Bóng bàn', 'Bóng đá ', 'Quần vợt']

# Thử xem giá trị của vài biến
birthday
favorNumbers

# Xem phần tử đầu tiên của favorNumbers
favorNumbers[0]

# Đếm số phần tử của biến favorNumbers
```

```
len(favorNumbers)

# Lấy ra phần tử cuối cùng của biến favorNumbers
favorNumbers[len(favorNumbers) - 1]
```

Bạn nên copy từng lệnh hoặc tốt nhất là tự gõ vào Spyder để chạy và quan sát. Sau mỗi lệnh bạn nên gõ lệnh `type` để biết thêm kiểu dữ liệu của biến:

```
type(<tên biến>)
```

Ví dụ:

```
type(fullName)
```

Cho kết quả là: str

str có nghĩa là String (chuỗi)

② Làm quen hàm thống kê

Khảo sát đoạn chương trình sau:

```
import pandas as pd
a = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
s_a = pd.Series(a)
s_a.describe()
```

Kết quả như sau:

```
count    15.000000
mean      8.000000
std       4.472136
min       1.000000
25%       4.500000
50%       8.000000
75%      11.500000
max      15.000000
dtype: float64
```

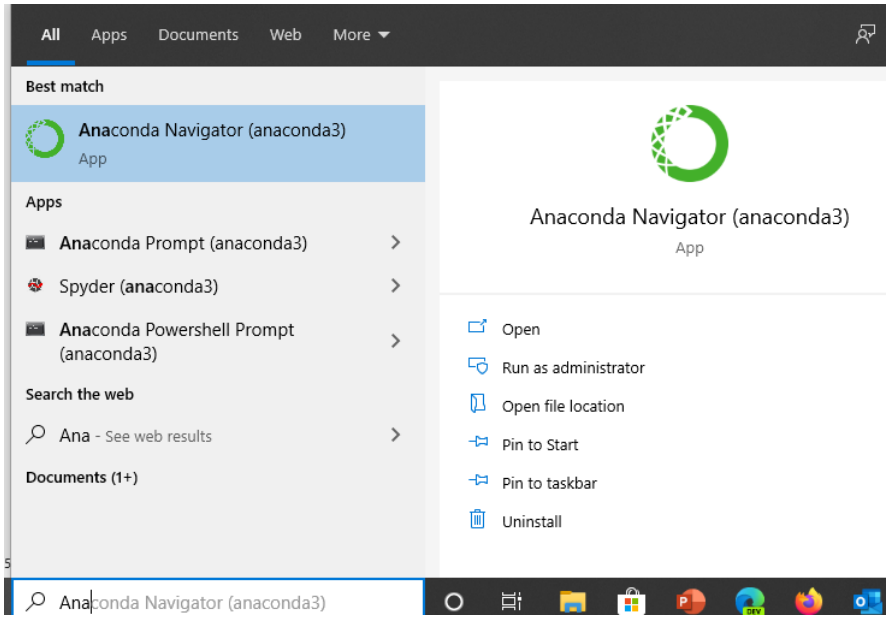
Diễn giải:

- Lệnh đầu tiên là khai báo sử dụng thư viện pandas với định danh là pd.
- Lệnh thứ hai khai báo một dãy số gồm 15 phần tử, mỗi phần tử có giá trị tương ứng từ 1 đến 15.

- Lệnh thứ ba chuyển dãy a thành kiểu dữ liệu gọi là Series – là một cột dữ liệu trong bảng dữ liệu (gọi là Data Frame). Kết quả lưu vào biến s_a.
- Sử dụng hàm .describe() của cột dữ liệu (Series) s_a.
- Kết quả hàm describe() sẽ cung cấp vài thông tin thống kê để mô tả về biến s_a. Cụ thể gồm:
    - count: tổng số phần tử.
    - mean: giá trị trung bình của các phần tử.
    - std: Độ lệch chuẩn (Xem lại mô tả khái niệm Độ lệch chuẩn)
    - min, max: Giá trị nhỏ nhất, Giá trị lớn nhất.
    - 25%: Giá trị bách phân vị 25%. Giá trị mà tại đó chia tập dữ liệu thành 2 phần ¼ và ¾.
    - 50%: Giá trị bách phân vị 50%. Giá trị mà tại đó chia tập dữ liệu thành 2 phần bằng nhau ½ và ½.
    - 75%: Giá trị bách phân vị 75%. Giá trị mà tại đó chia tập dữ liệu thành 2 phần ¾ (Phần các giá trị nhỏ) và ¼ (Phần các giá trị lớn..

Cài đặt thư viện

Một trong các lý do mà ngôn ngữ Python phổ biến nhất tại thời điểm eBook được viết trong lĩnh vực Machine Learning và AI là cộng đồng phát triển rất lớn. Trong đó có rất nhiều thư viện được cung cấp miễn phí. Trong Windows, để cài đặt thư viện Python thì mở cửa sổ của Anaconda Prompt hoặc Anaconda Powershell Prompt bằng cách bấm vào nút Windows Start, gõ chữ Ana thì ra màn hình bên dưới, sau đó bấm vào biểu tượng tương ứng (ví dụ Anaconda Prompt).

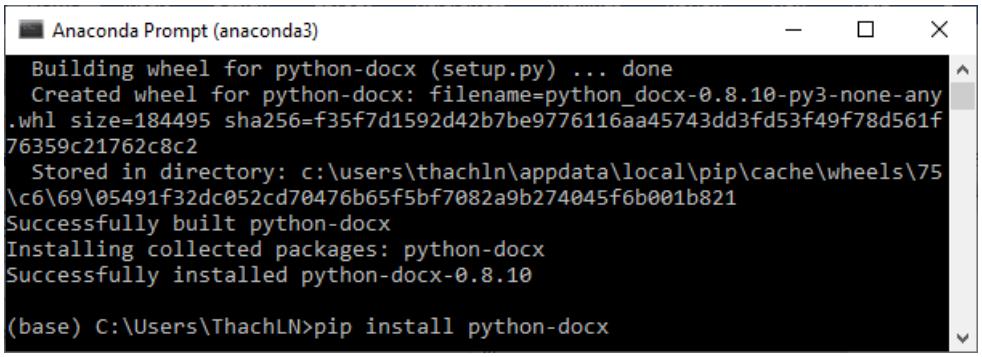


Trong cửa sổ Anaconda Prompt gõ lệnh:

```
pip install <tên thư viện>
```

Ví dụ cài thư viện python-docx để xử lý file .docx của Microsoft Word:

```
pip install python-docx
```

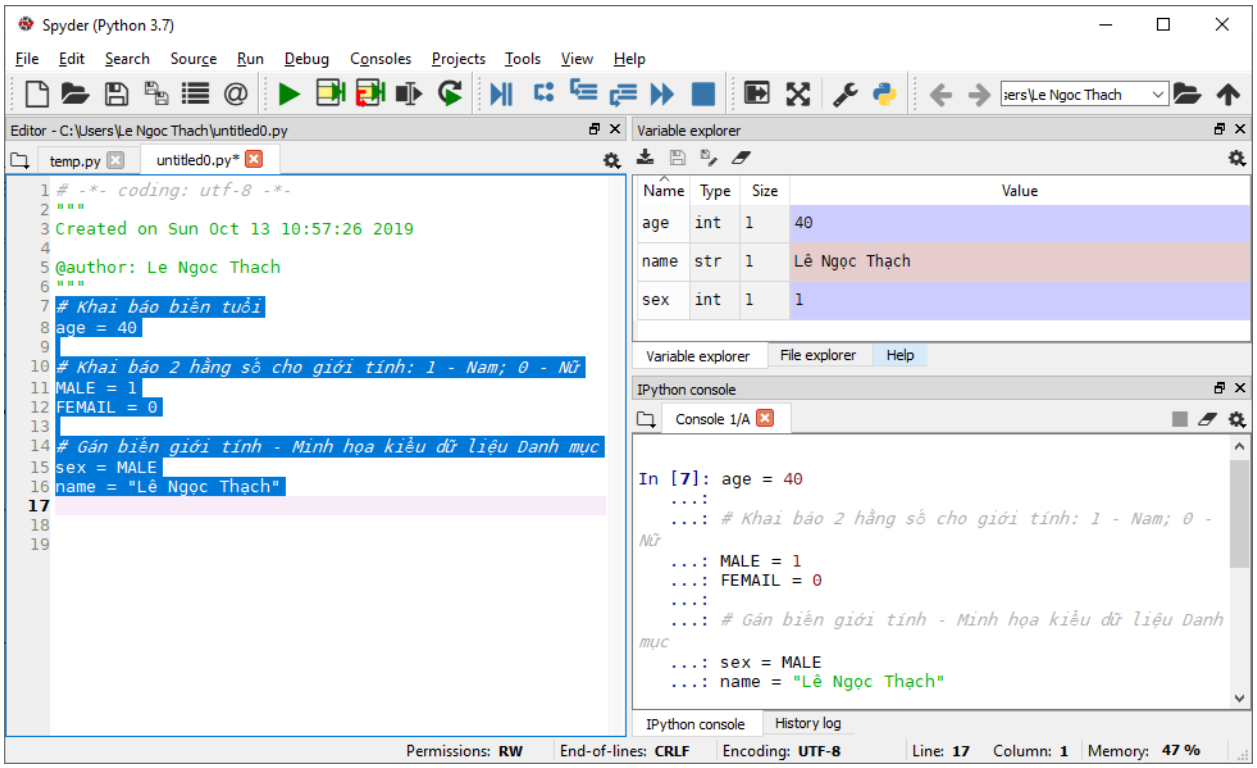


Sử dụng chú thích

Gõ các lệnh sau vào Spyder để chạy thử. Các dòng có dấu thăng # ở phía trước là các dòng chú thích. Spyder sẽ bỏ qua các dòng nay khi thực thi lệnh.

```
# Khai báo biến tuổi
age = 40
# Khai báo 2 hằng số cho giới tính: 1 - Nam; 0 - Nữ
MALE = 1
FEMALE = 0

# Gán biến giới tính - Minh họa kiểu dữ liệu Danh mục
sex = MALE
name = 'Lê Ngọc Thạch'
```



Chú ý là tôi cố tình có lúc dùng cặp nháy đôi, có lúc dùng cặp nháy đơn để bao đóng chuỗi để giúp bạn nhờ là dùng cái nào cũng được, ý nghĩa là như nhau trong Python.

Cập nhật phiên bản mới

Kiểm tra phiên bản mới đã phát hành (release) của Spyder tại website "https://github.com/spyder-ide/spyder/releases".

Trong cửa sổ Anaconda Powershell Prompt thực hiện lệnh:

```
conda install spyder=6.1.3
```

Nếu bạn kiểm tra phiên bản đã phát hành của Spyder lớn hơn 6.1.3 thì sửa lại lệnh trên cho phù hợp.

Liên lạc | Tài trợ

Lê Ngọc Thạch	
(+084) 0908 550 642	facebook.com/ThachLN
Website: https://xmyworkspace.com/learn/sponsor