Python, Jupyter Notebook, Kaggle

Trịnh Tấn Đạt
Khoa CNTT – Đại Học Sài Gòn
Email: trinhtandat@sgu.edu.vn
Website: https://sites.google.com/site/ttdat88/

Cài Đặt

1. Cài đặt Python 3 + IDE Pycharm

2. Cài đặt Python 3 dùng bản phân phối Anaconda (RAM > 4GB)

- **Cài đặt IDE Pycharm**
- Anaconda hỗ trợ IDE Spyder
- Jupyter Notebook

3. Sử dụng Google Colab, Kaggle

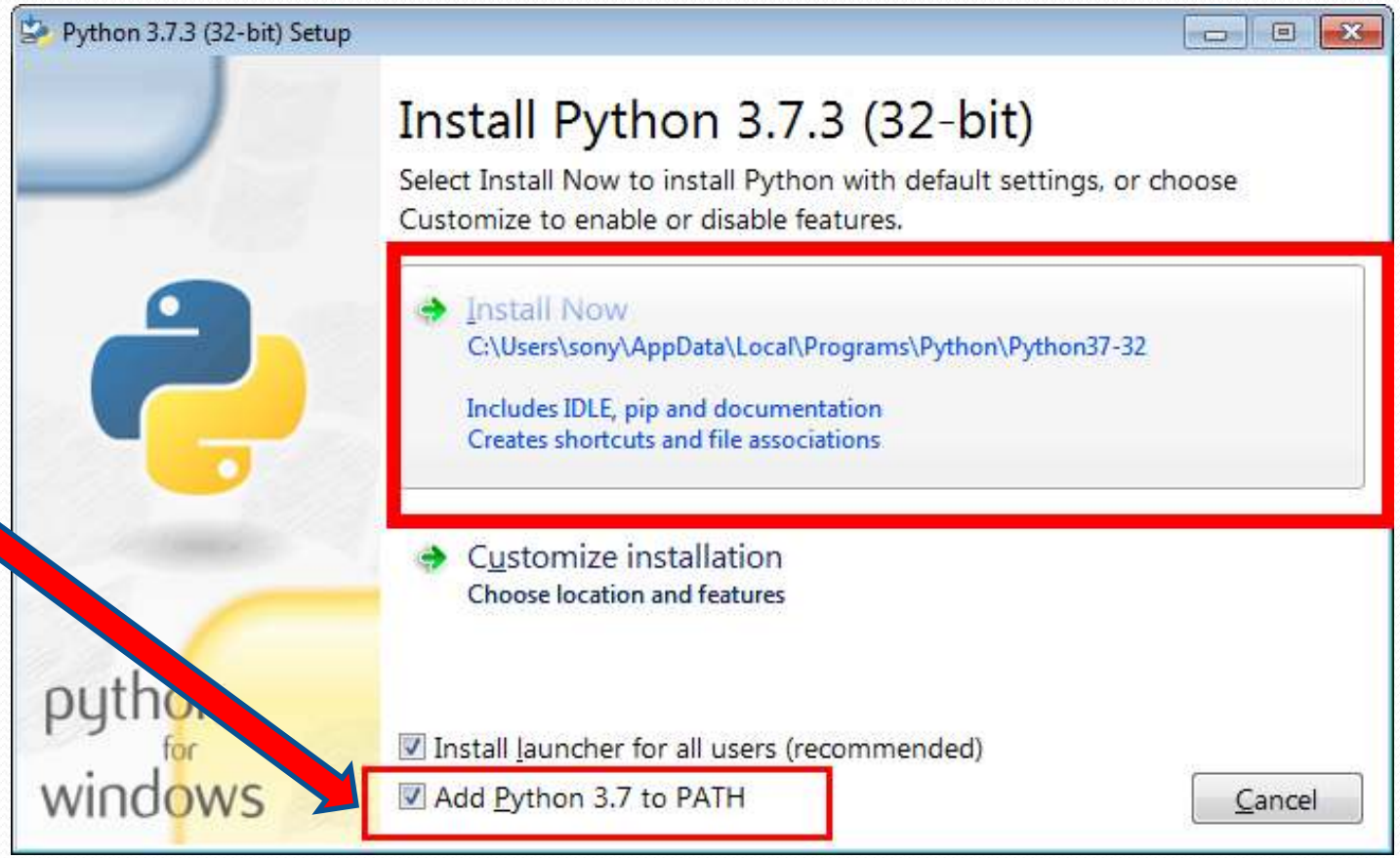
Cài đặt Python 3

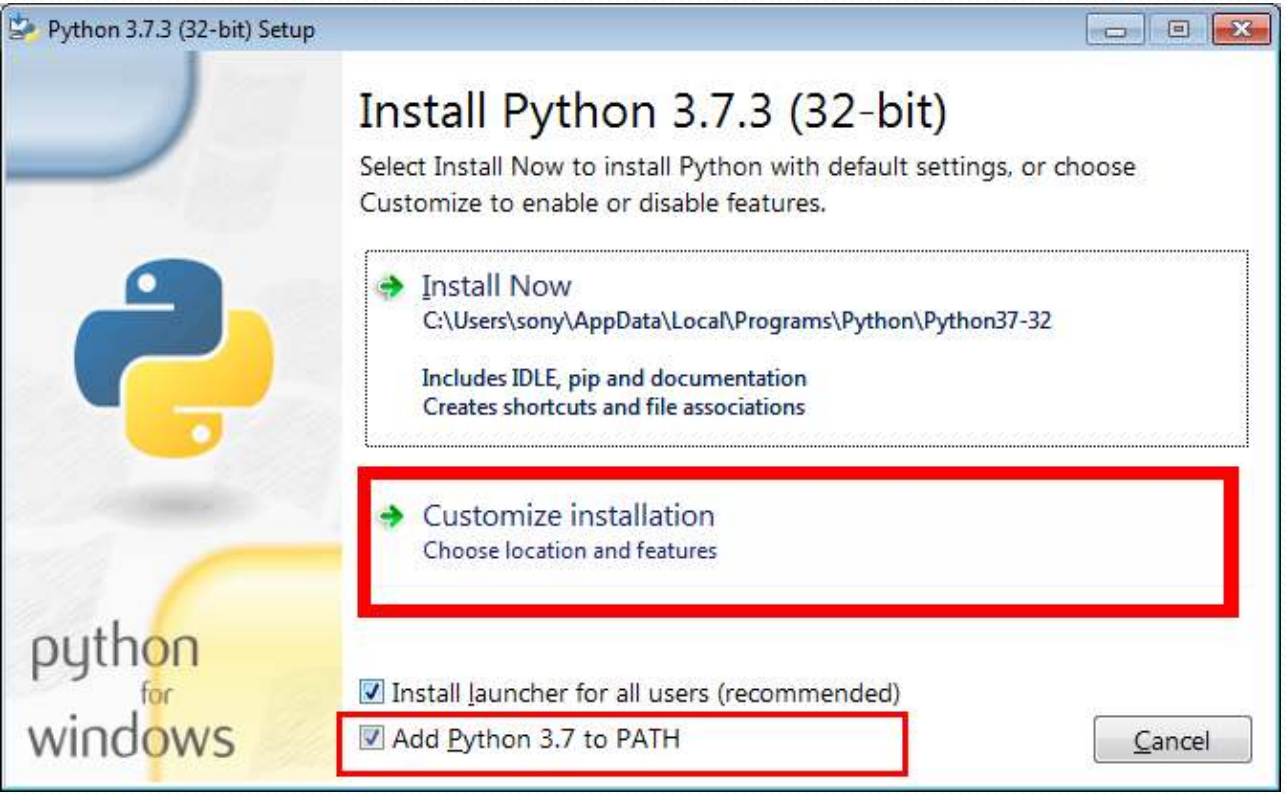
- Bước 1: Để tải và cài đặt Python 3 hãy truy cập vào trang web chính thức của Python: http://www.python.org/downloads/ và chọn phiên bản bạn muốn 3.7.*

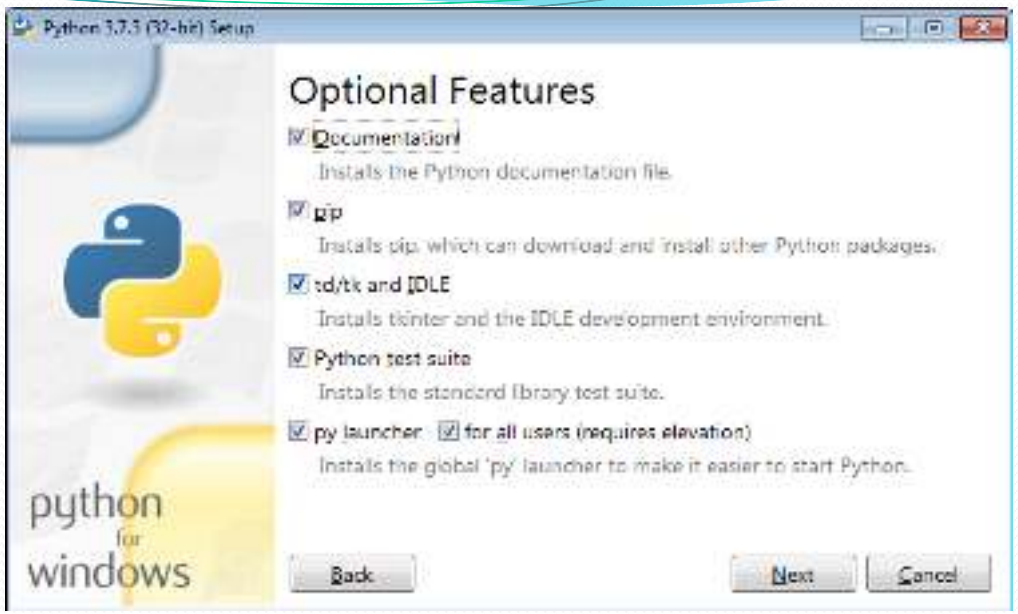
Cài đặt Python 3

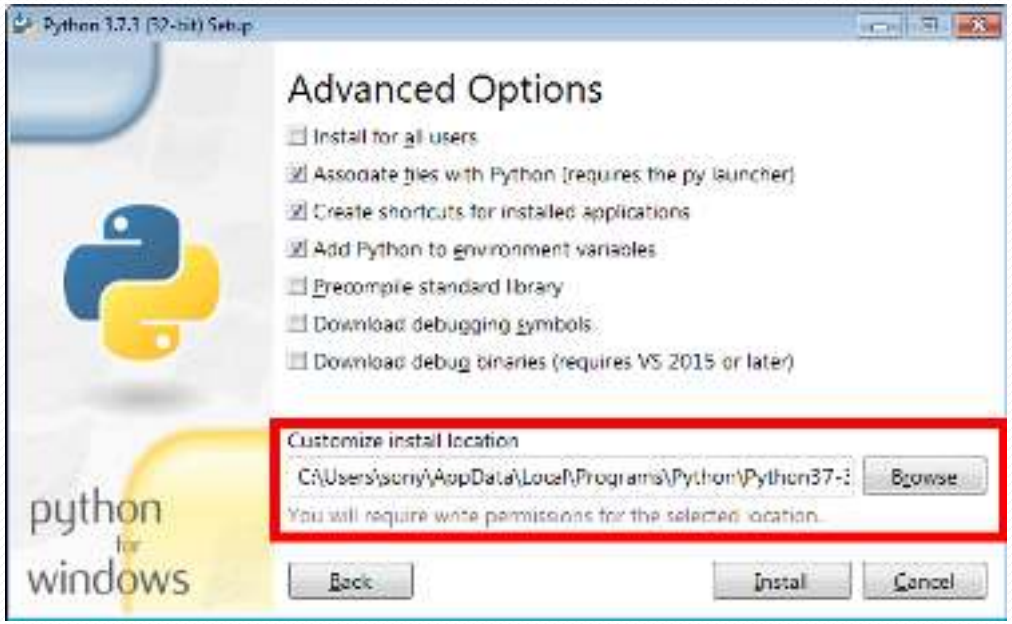
- Bước 2: Sau khi quá trình tải xuống đã hoàn tất, chạy tập tin exe để cài đặt Python. Giờ hãy click vào Install Now.

* Note: Chọn Add Python to PATH

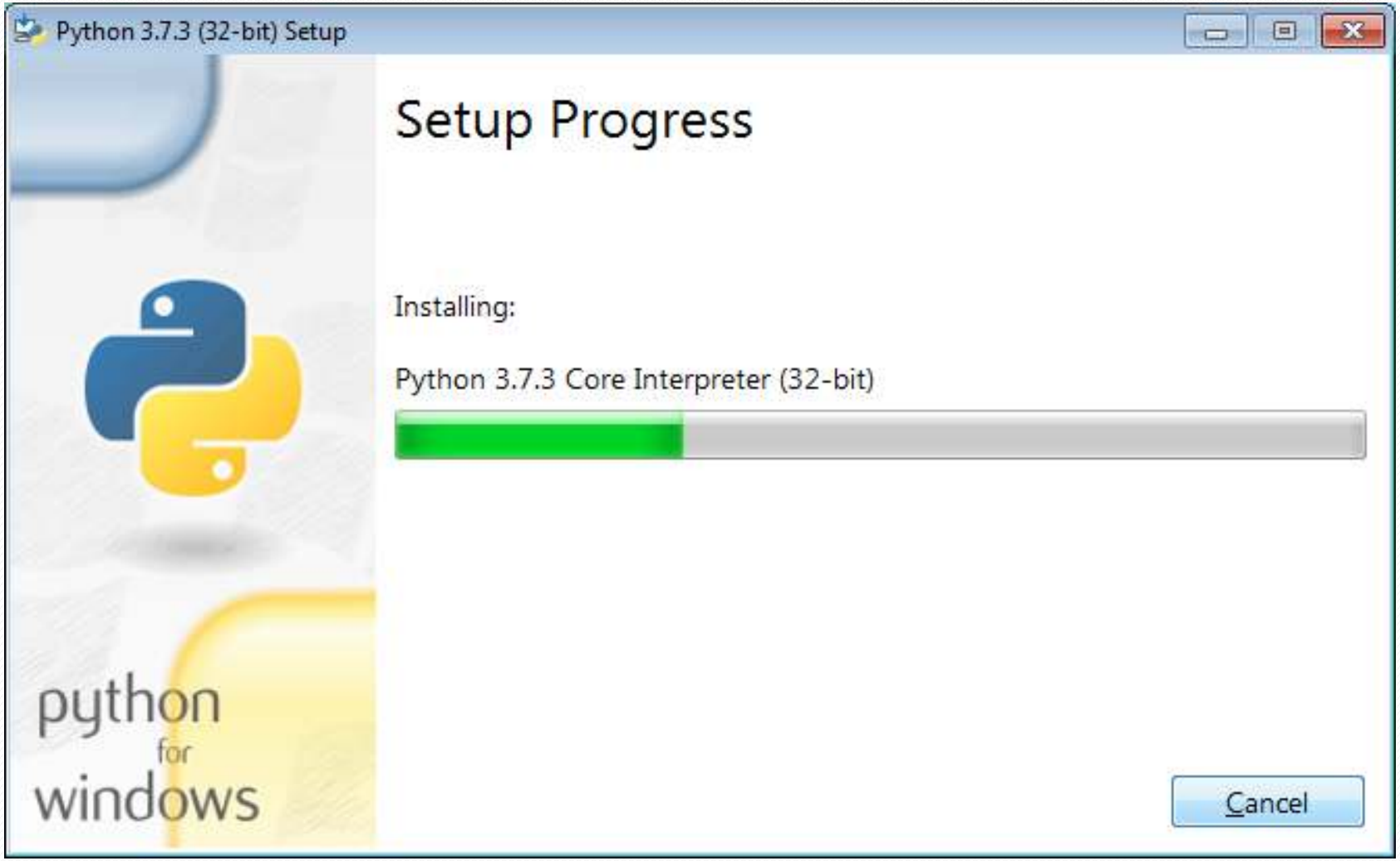



Python 3.7.3 (32-bit) Setup
Install Python 3.7.3 (32-bit)
Select Install Now to install Python with default settings, or choose Customize to enable or disable features.
Install Now
C:\Users\sony\AppData\Local\Programs\Python\Python37-32
Includes IDLE, pip and documentation
Creates shortcuts and file associations
Customize installation
Choose location and features
Install launcher for all users (recommended)
Add Python 3.7 to PATH
Cancel
python
for
windows


Python 3.7.3 (32-bit) Setup
Optional Features
Documentation
Installs the Python documentation file.
pip
Installs pip, which can download and install other Python packages.
tcl/tk and IDLE
Installs tkinter and the IDLE development environment.
Python test suite
Installs the standard library test suite.
py launcher
for all users (requires elevation)
Installs the global 'py' launcher to make it easier to start Python.
Back
Next
Cancel
python
for
windows


Python 3.7.3 (32-bit) Setup
Advanced Options
Install for all users
Associate files with Python (requires the py launcher)
Create shortcuts for installed applications
Add Python to environment variables
Precompile standard library
Download debugging symbols
Download debug binaries (requires VS 2015 or later)
Customize install location
C:\Users\sony\AppData\Local\Programs\Python\Python37-3
Browse
You will require write permissions for the selected location.
Back
Install
Cancel
python
for
windows

Cài đặt Python 3



Cài đặt Python 3

- Run và kiểm tra quá trình cài đặt thành công

1. Mở Command Promt (cmd)
2. Type : python --version hoặc python

Administrator: Command Prompt - python

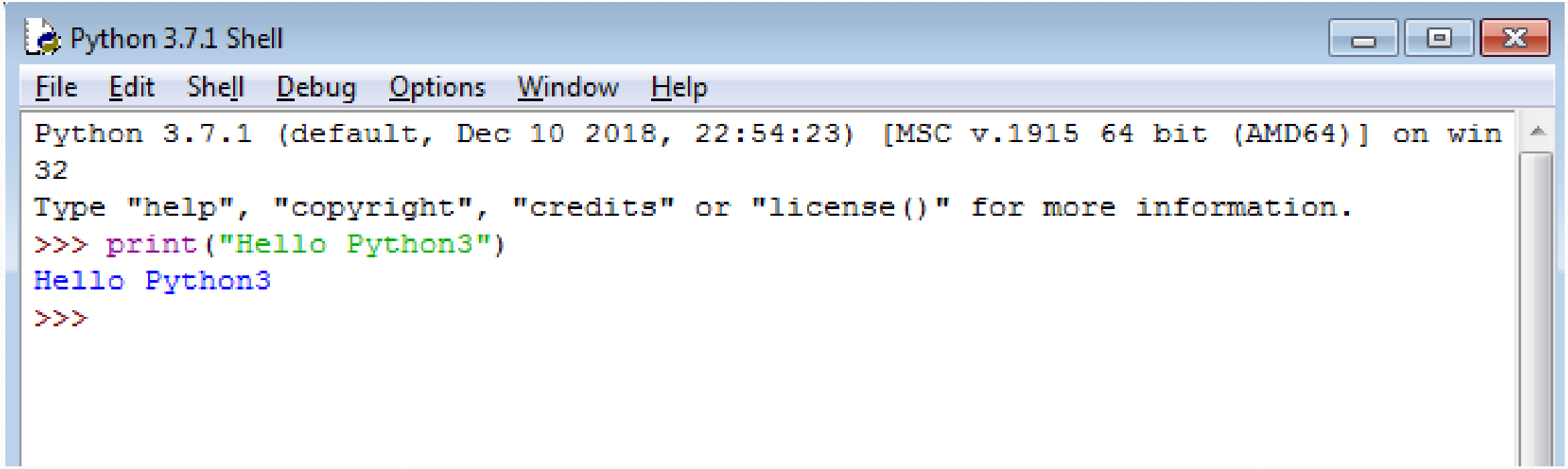
```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\sony>python
Python 3.7.1 (default, Dec 10 2018, 22:54:23) [MSC v.1915 64 bit (AMD64)] :: Ana
conda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello Python3")
Hello Python3
>>> _
```

Cài đặt Python 3

Hoặc dùng Python Shell

- Trên “Start menu” search “ idle” and run file “idle.exe”



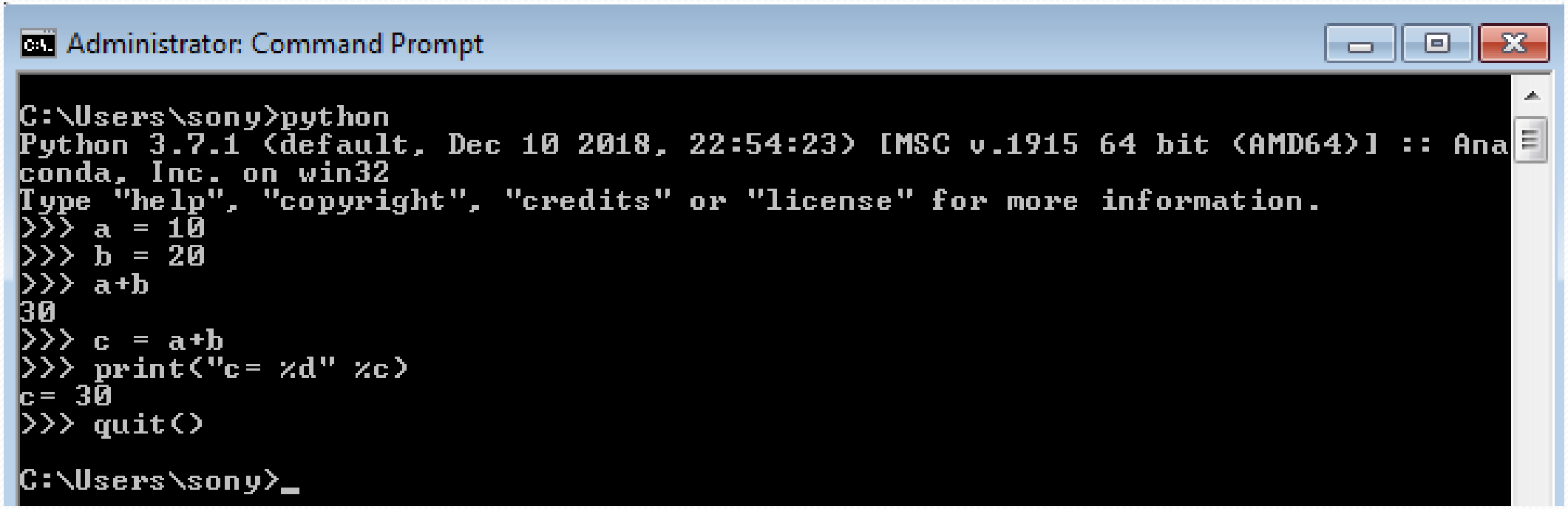
Cài đặt Python 3

- Thực thi: Python có 2 chế độ thực thi
  - Chế độ dòng lệnh: chạy từng lệnh một
  - Chế độ thực thi: chỉ ra chương trình cần thực hiện (file script *.py)
    - Trình dịch python sẽ nạp, dịch và chạy chương trình đó

Cài đặt Python 3

Chế độ dòng lệnh: open cmd -> type python / hoặc open Python Shell

- Lúc này trình thông dịch python sẽ chờ người dùng gõ từng dòng lệnh
- Gõ dòng lệnh nào xong, python chạy liền dòng đó
- Chấm dứt chế độ này bằng cách gõ lệnh: “quit()” hoặc Ctrl+Z



Cài đặt Python 3

Chế độ thực thi trên file script *.py

- Ví dụ: Dùng Notepad (hoặc các text editor) tạo file test.py (lưu file dưới dạng mở rộng *.py)
  - Open Command Promt (cmd)
  - type: python <đường dẫn tới file test.py>
  - Ví dụ: python C:\Users\sony\Desktop\python_co_ban\test\test.py

test.py

```
# -*- coding: utf-8 -*
print("Hello world")
a = 1
b = 2
c = a+b
print("c=%d" %c)
```

```
C:\Users\sony>python C:\Users\sony\Desktop\python_co_ban\test\test.py
Hello world
c=3
```

Hoặc di chuyển đến thư mục chứa file cần thực thi

Type: cd < đường dẫn tới thư mục chứa file >

Ví dụ: cd C:\Users\sony\Desktop\python_co_ban\test

python test.py

```
C:\Users\sony>cd C:\Users\sony\Desktop\python_co_ban\test

C:\Users\sony\Desktop\python_co_ban\test>python test.py
Hello world
c=3
```

Cài đặt Python 3

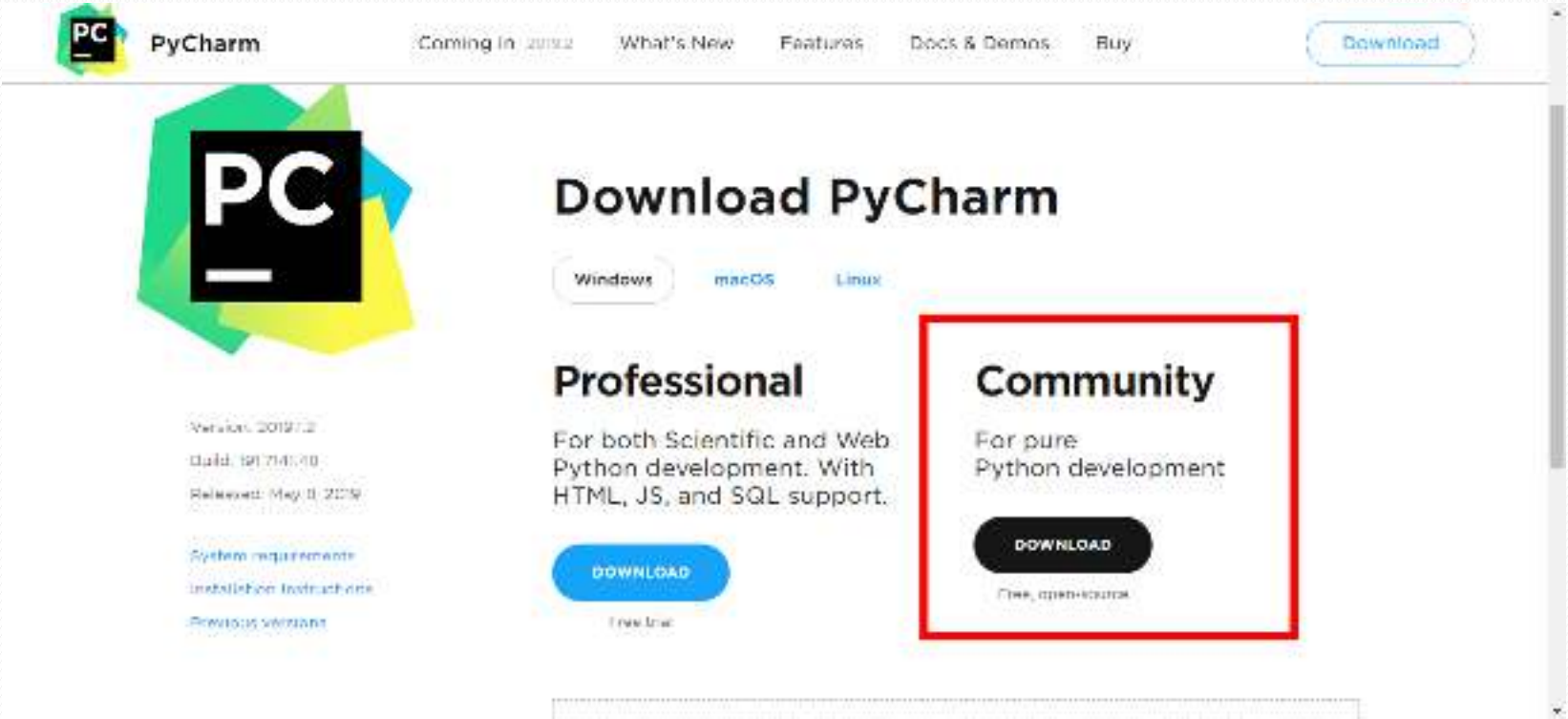
- Hạn chế khi dùng các text editors (notepad, notepad++, wordpad, gedit, vi, ...)
  - Không có tính năng tự động nhắc và sữa lỗi cú pháp (thừa thiếu space/tab , viết sai tên hàm, sai tên thư viện, …)
  - Không có tính năng debug
  - …

- Sử dụng môi trường phát triển tích hợp (Integrated Development Environment-**IDE**)
  - **PyCharm**
  - Spyder
  - PyDev
  - Eclipse
  - …

Cài đặt Python 3

Cách cài đặt Pycharm IDE sau khi đã cài đặt python

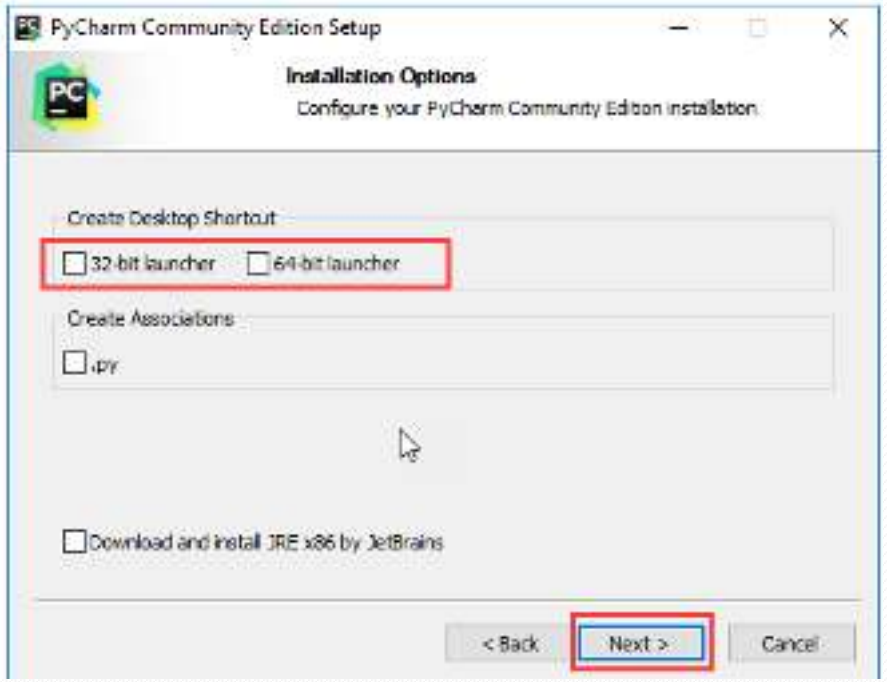
- Bước 1: Để tải PyCharm, hãy truy cập vào trang web: https://www.jetbrains.com/pycharm/download/ sau đó click vào đường dẫn “Download” nằm dưới mục Community:

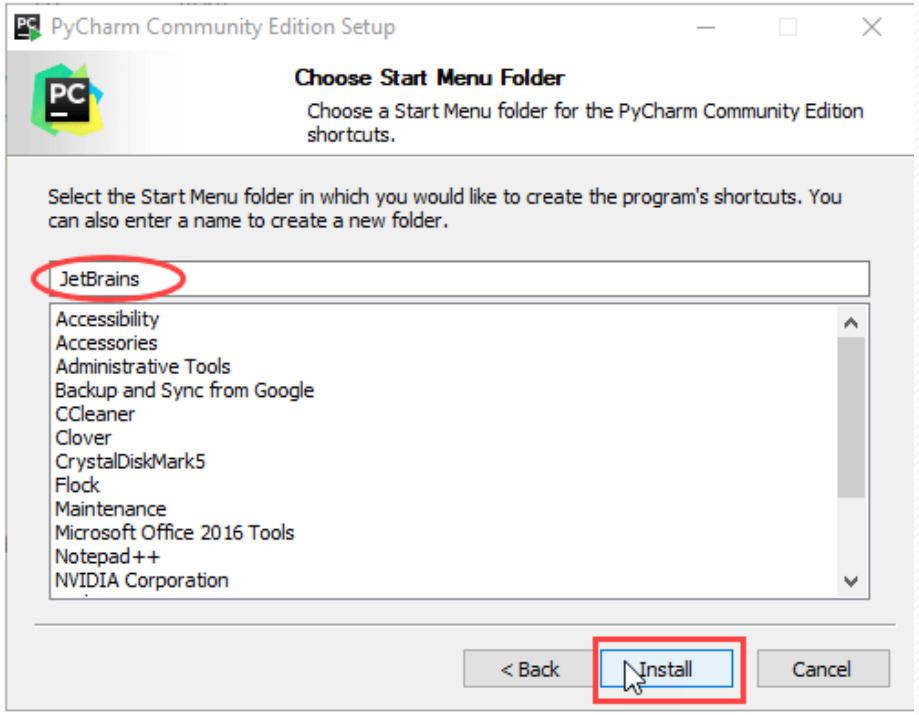


Cài đặt Python 3

Cách cài đặt Pycharm IDE sau khi đã cài đặt python

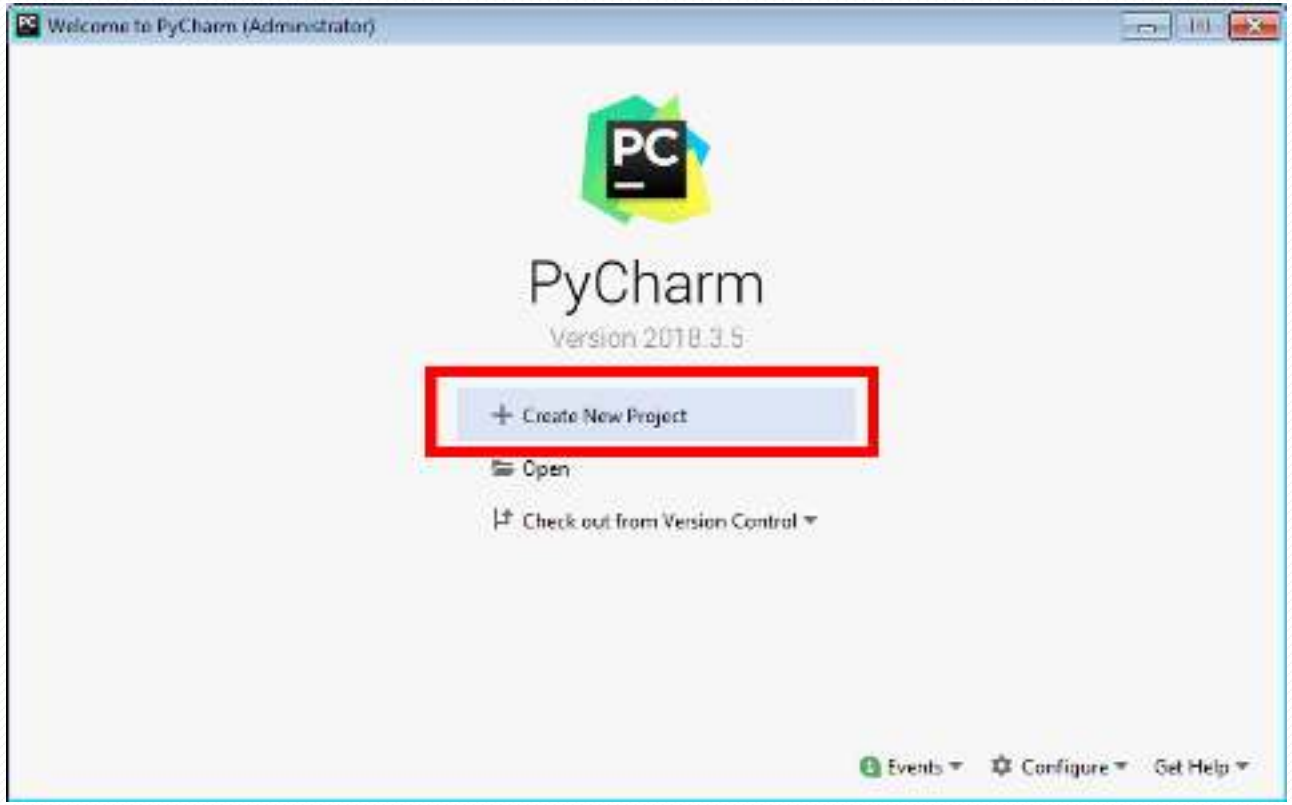
- Bước 2: Khi quá trình download đã hoàn tất, chạy tập tin exe để cài đặt PyCharm.





Cài đặt Python 3

- Tạo một chương trình minh họa dùng pycharm
- B1 : Mở trình soạn thảo PyCharm. Bạn có thể thấy màn hình giới thiệu khi mở PyCharm. Để tạo một dự án mới, bấm vào “Create New Project”.

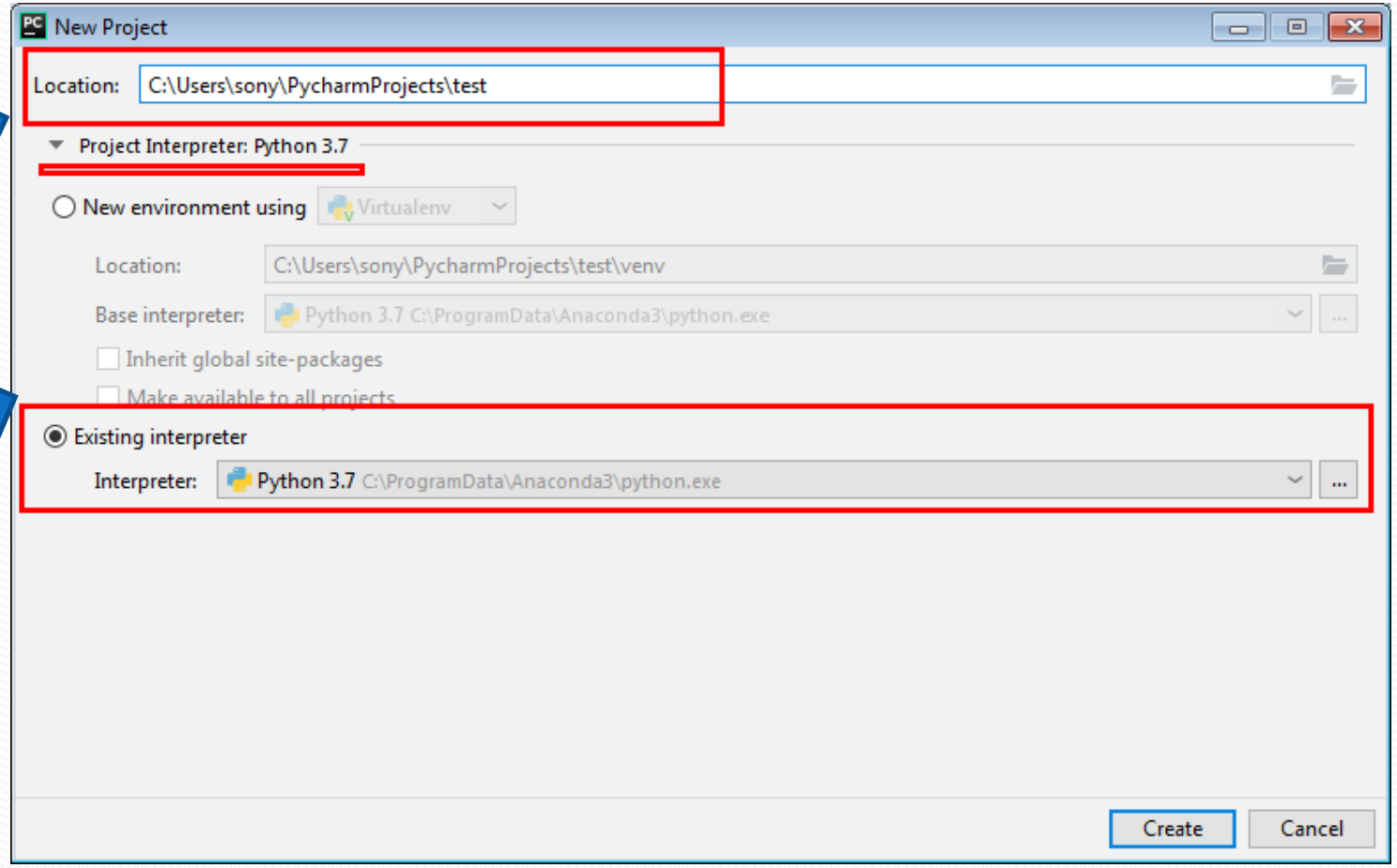


Cài đặt Python 3

- B2: Chọn thư mục và tên của project để lưu (Location)
  - PyCharm sẽ tìm thấy trình thông dịch Python bạn đã cài đặt trước đó.

(đường dẫn đến file python.exe – khi đã cài đặt python)

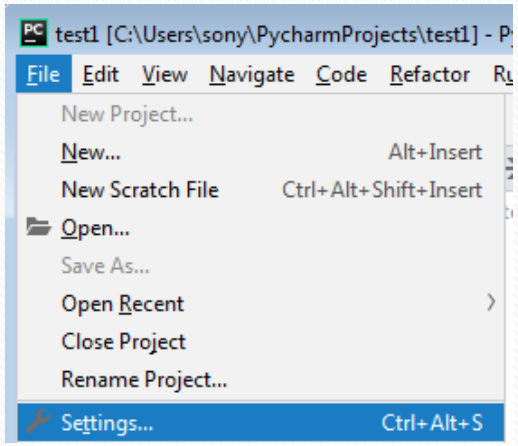
  - Sau đó chọn “Create”

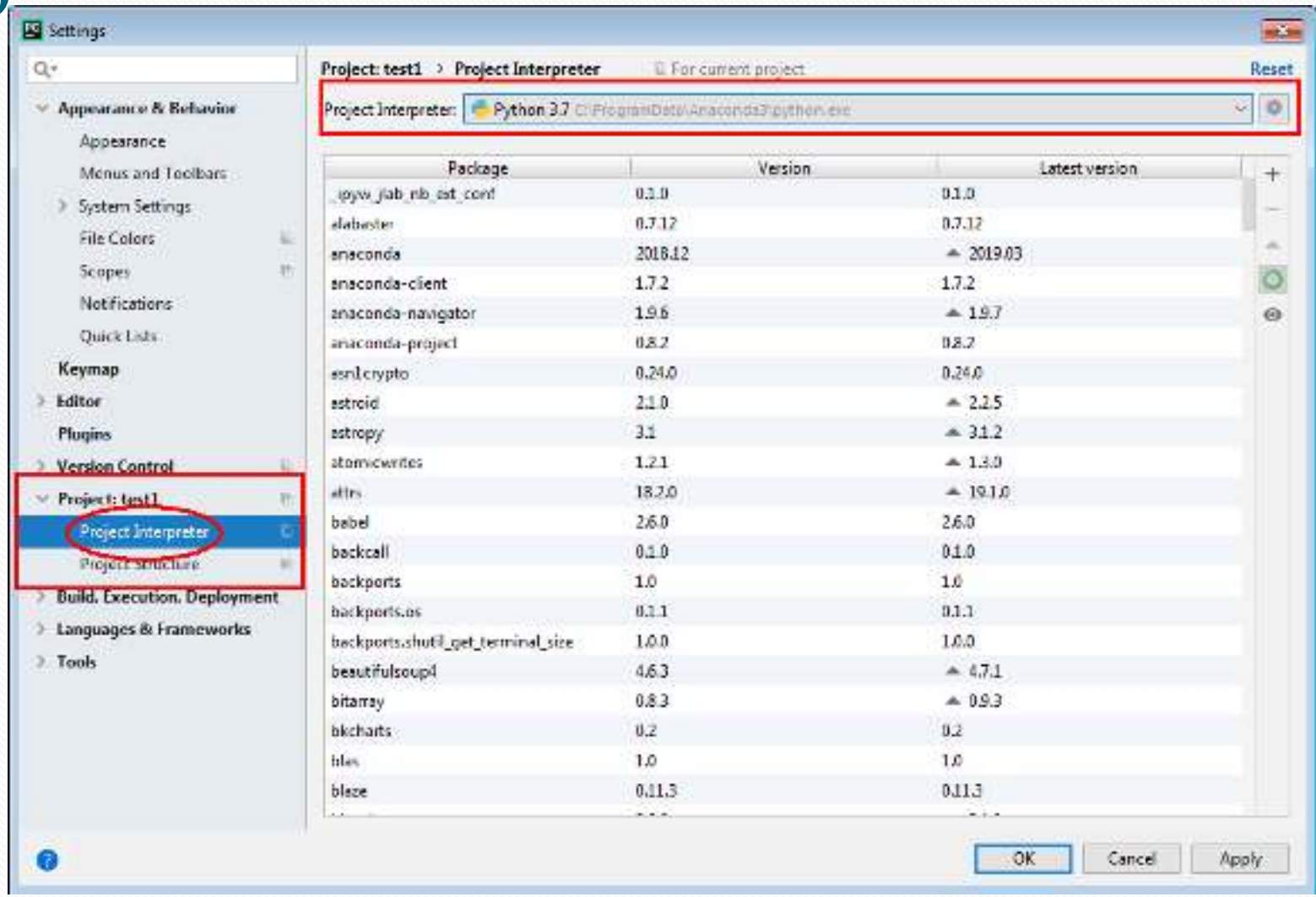


Cài đặt Python 3

(đường dẫn đến file python.exe – trong thư mục cài đặt python)

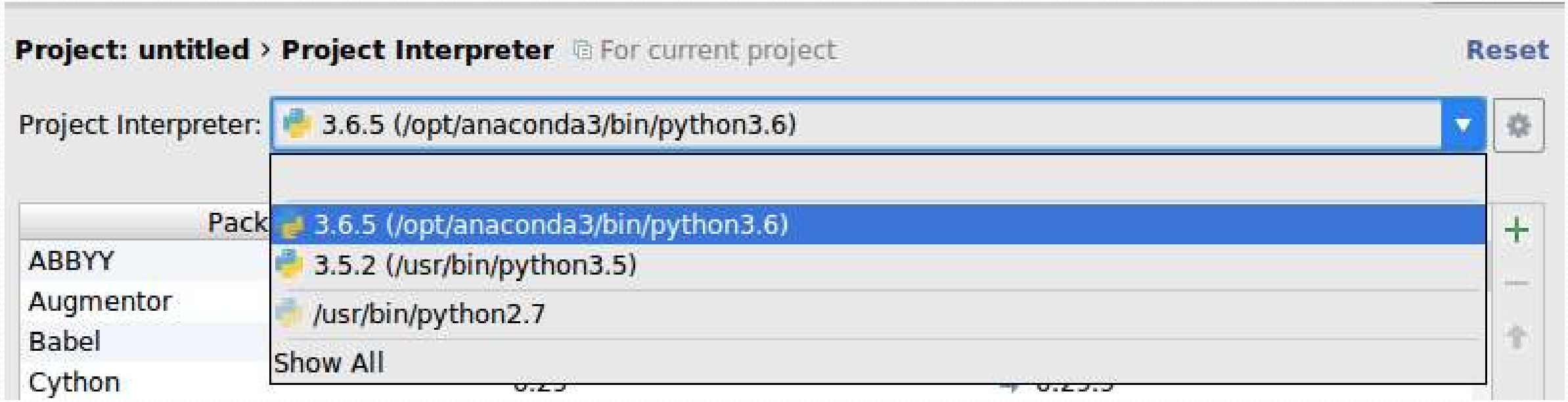
- B3: Load package và update trình thông dịch python cho Pycharm.
  - Chọn File -> Settings…





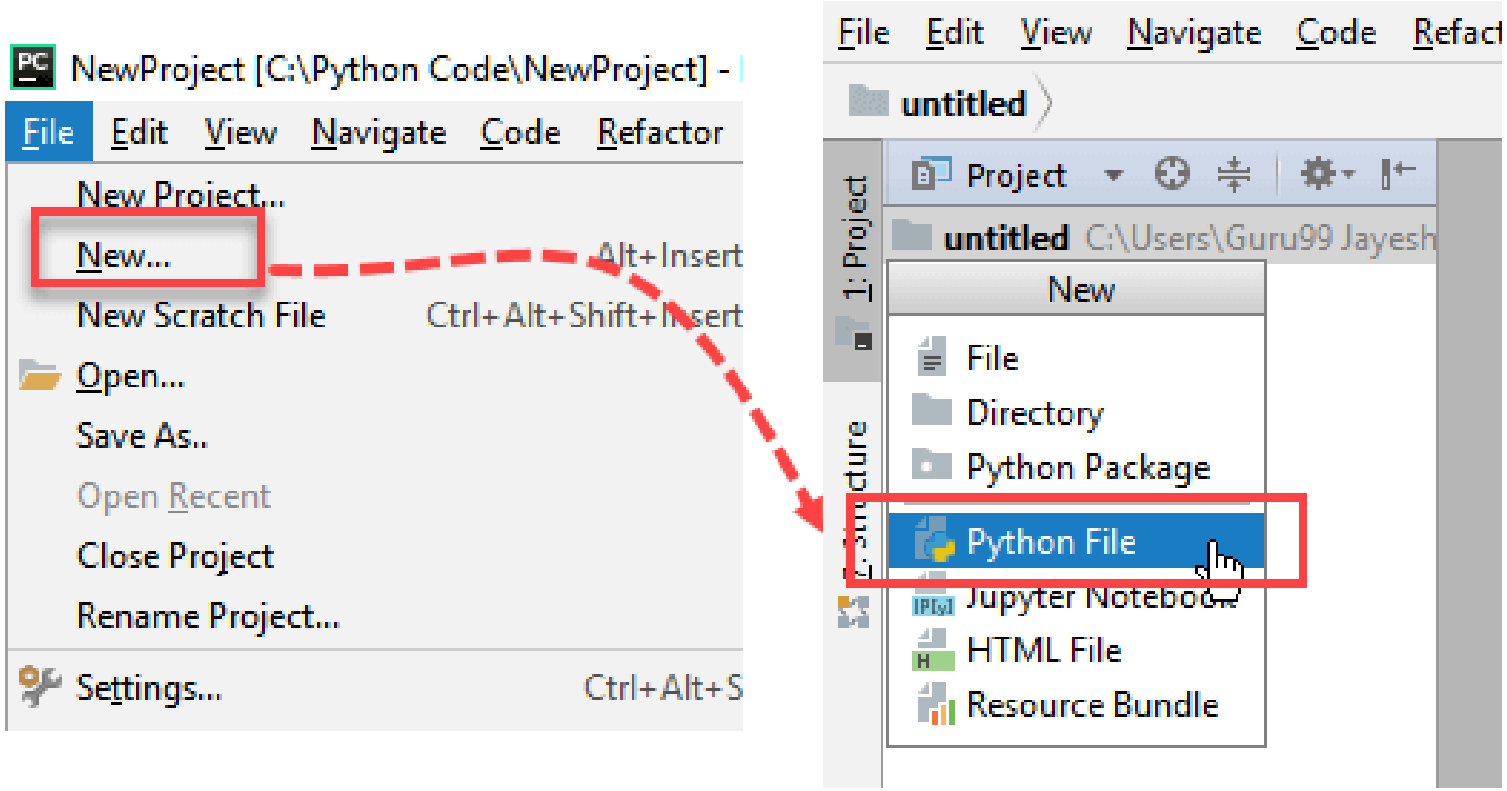
Cài đặt Python 3

- Ví dụ: có thể cài đặt cùng lúc nhiều python version khác nhau và tùy vào project chọn trình thông dịch và môi trường thích hợp



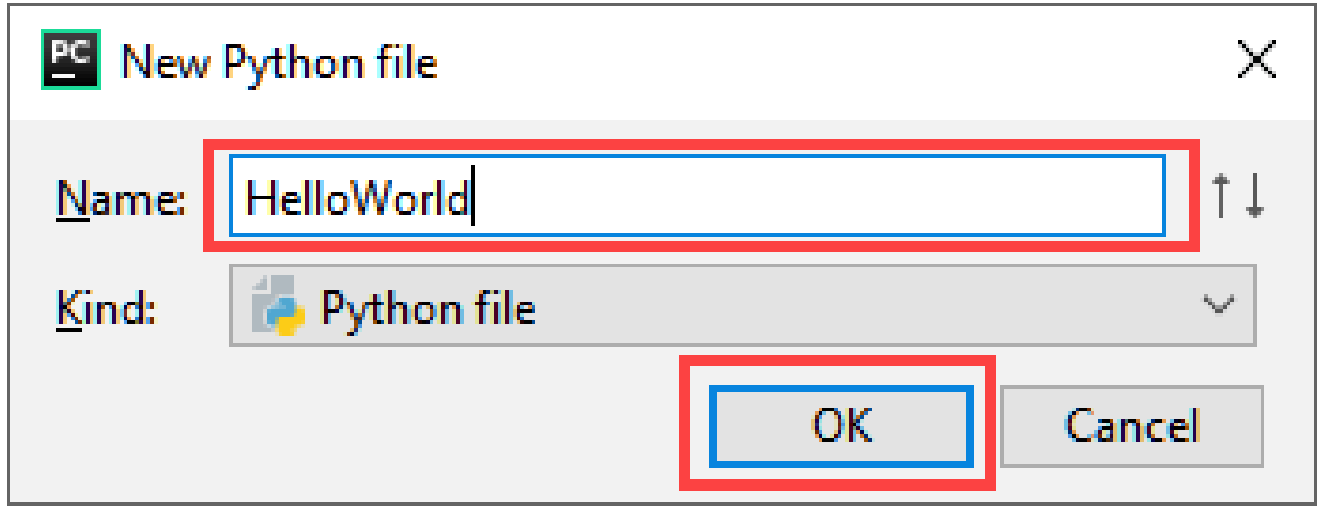
Cài đặt Python 3

- B4: Vào menu “File” và chọn “New”. Tiếp theo, chọn “Python File”.



Cài đặt Python 3

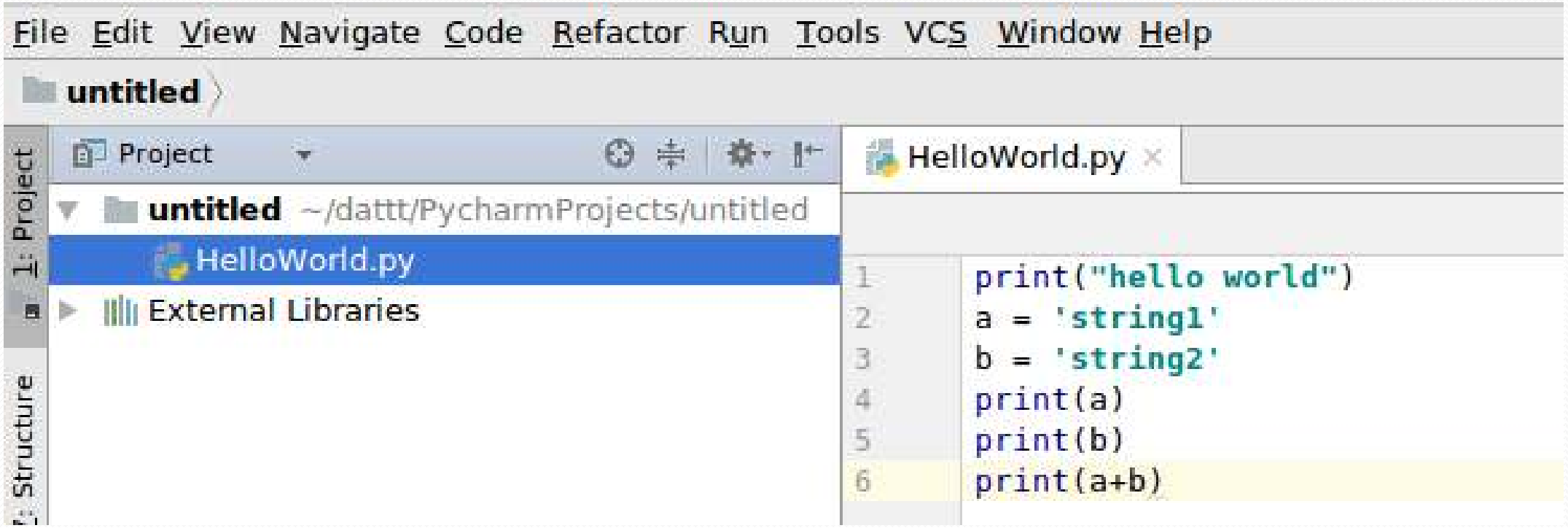
- B5: Một cửa sổ mới sẽ xuất hiện. Giờ hãy nhập tên bạn muốn đặt cho tệp



Cài đặt Python 3

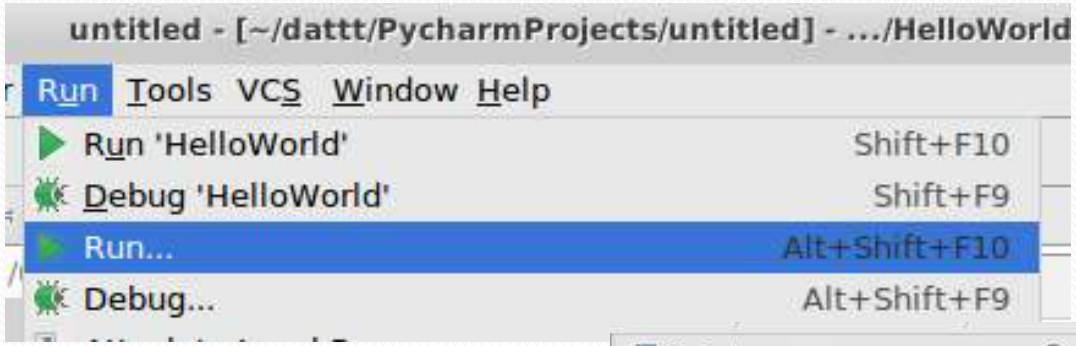
- B6: Viết một chương trình đơn giản

```
print("hello world")
a = 'string1'
b = 'string2'
print(a)
print(b)
print(a+b)
```

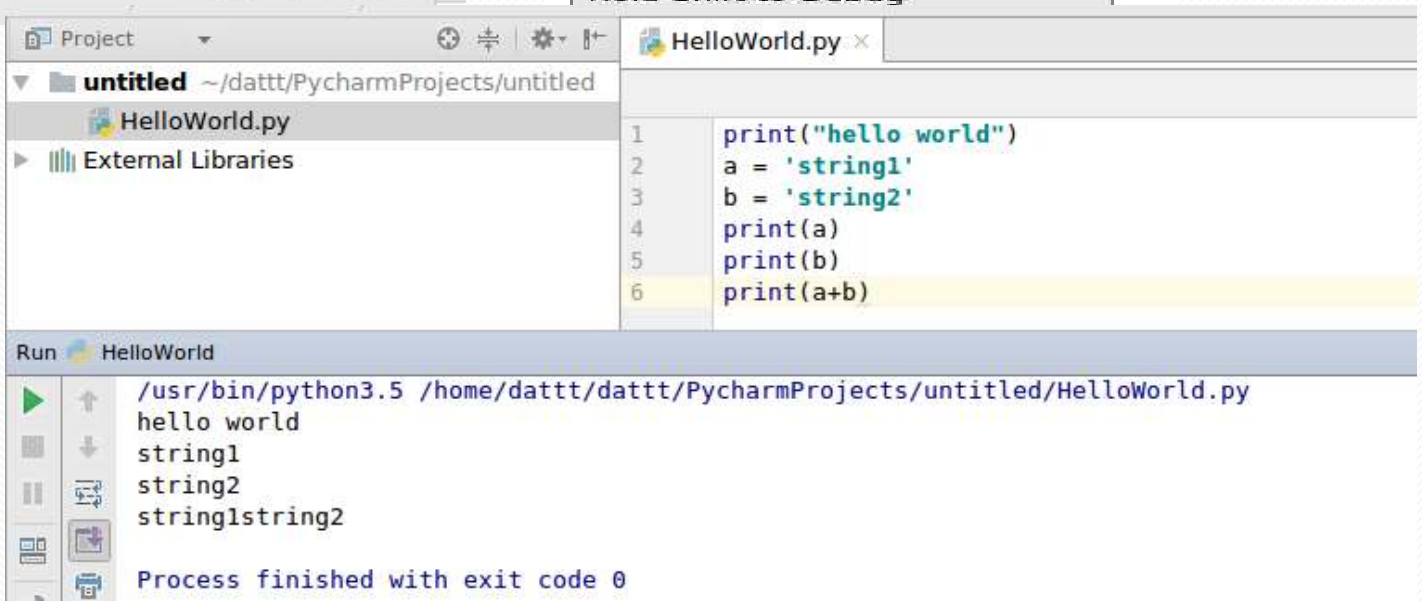


Cài đặt Python 3

- B7 : Vào menu “Run” và chọn “Run” để chạy chương trình của bạn.



- B8: xem kết quả

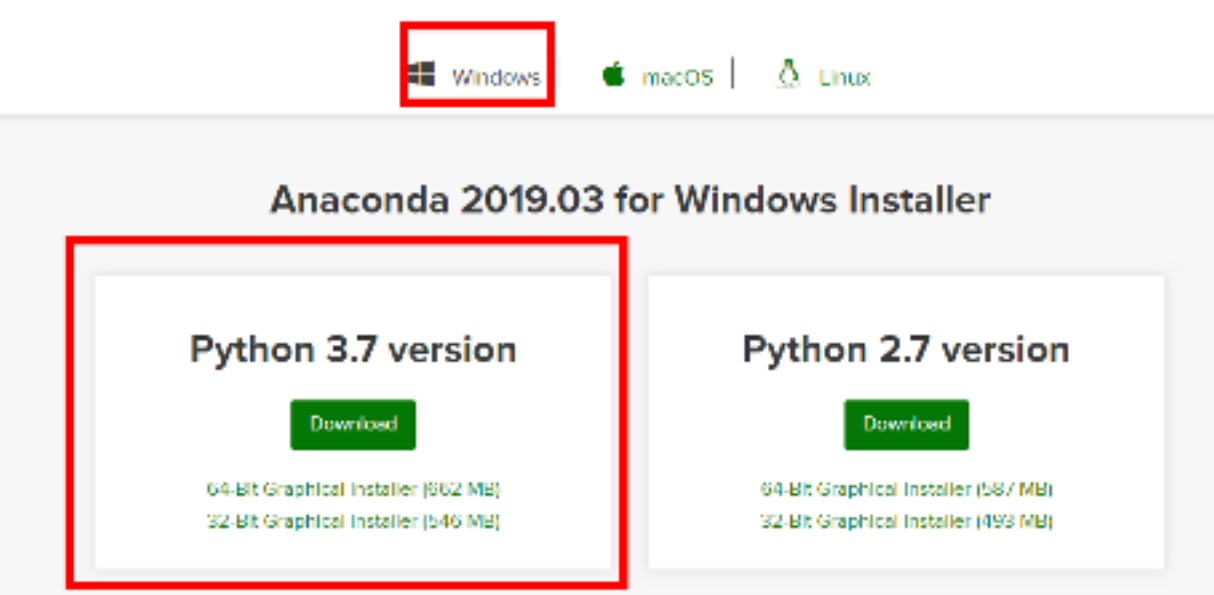


Cài đặt Python 3 dùng bản phân phối Anaconda

- Anaconda là một bản phân phối mã nguồn mở miễn phí của ngôn ngữ lập trình Python (https://www.anaconda.com/distribution/)
  - tích hợp sẵn các phiên bản python.
  - phục vụ cho machine learning, deep learning, xử lý dữ liệu quy mô lớn, phân tích dự đoán và tính toán khoa học.
  - được hơn 6 triệu người sử dụng.
  - Hơn 1,500+ Python/R data science packages

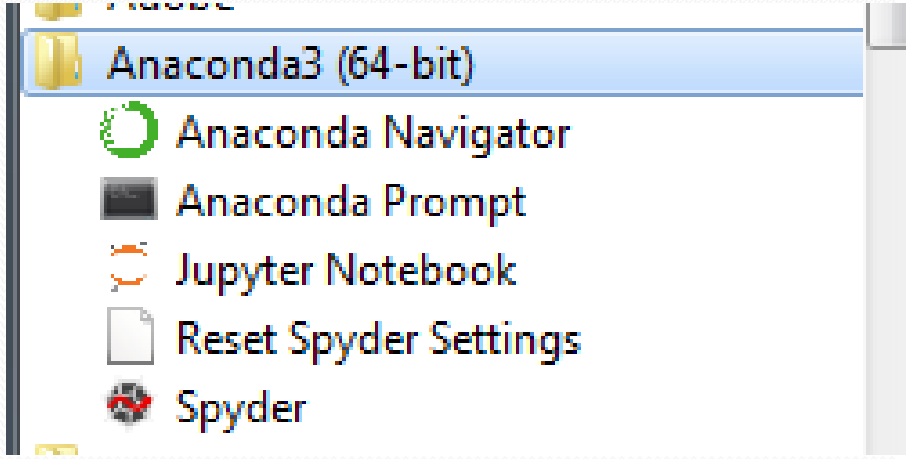
Cài đặt Python 3 dùng bản phân phối Anaconda

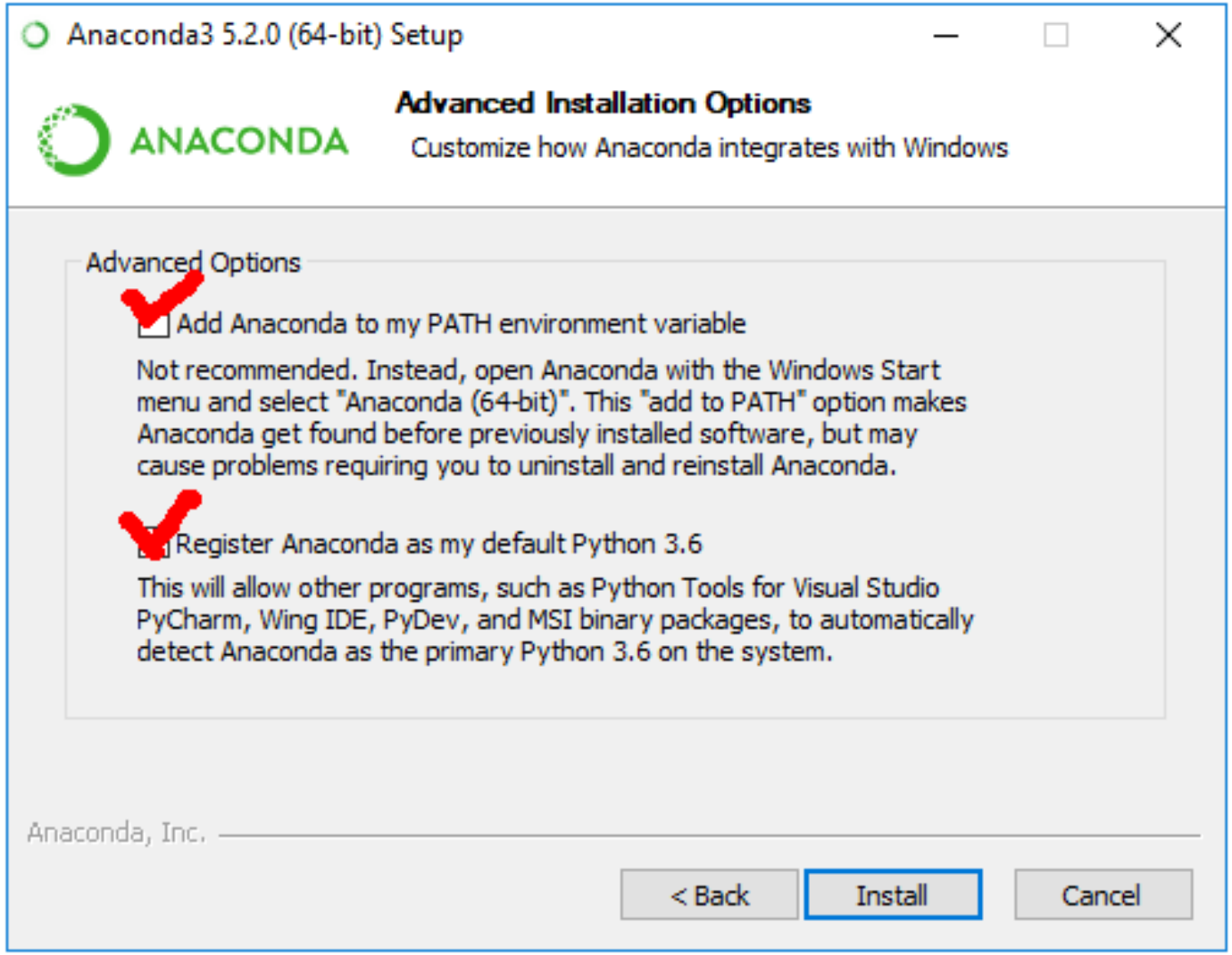
- Cài đặt tương tự như cài đặt python 3
- Download : https://www.anaconda.com/
  - https://www.anaconda.com/distribution/



Cài đặt Python 3 dùng bản phân phối Anaconda

- Lúc cài đặt lưu ý:





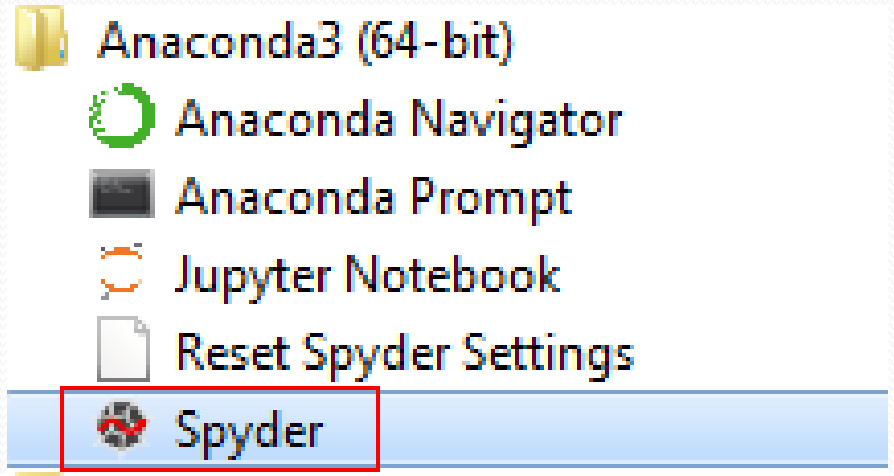
Cài đặt Python 3 dùng bản phân phối Anaconda

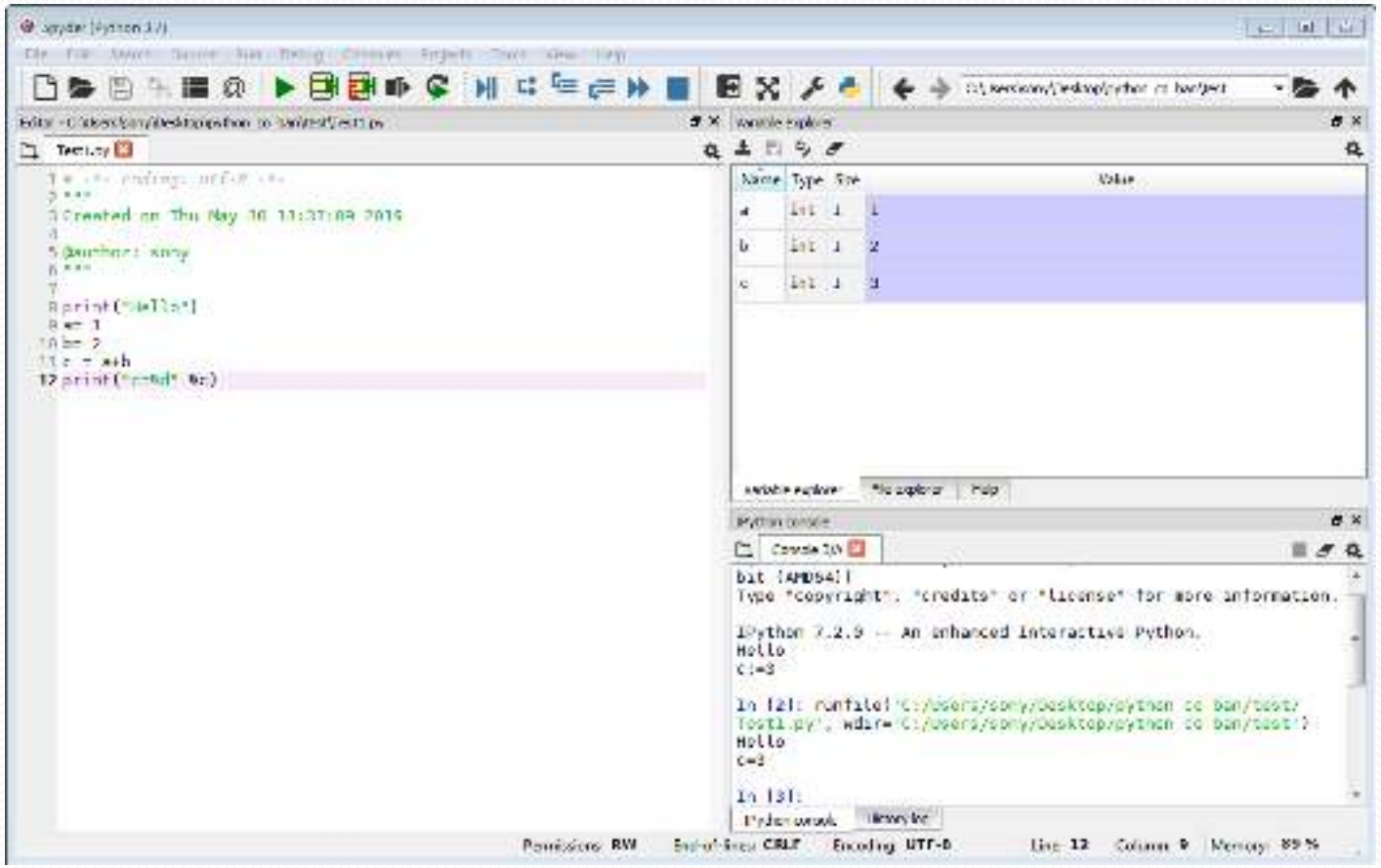
- Kiểm tra cài đặt:
  - Open cmd
  - Type:
    - conda list: nếu mọi thứ ok bạn sẽ thấy danh sách những package đã được cài đặt kèm theo version của chúng
    - python: nếu mọi thứ ok thì bạn sẽ thấy version của python kèm theo Anacoda

- Cài đặt IDE Pycharm dùng Anaconda tương tự như cài đặt cho python 3 (như các slide trước)

Cài đặt Python 3 dùng bản phân phối Anaconda

- Ngoài ra, Anaconda tích hợp sẵn một IDE Spyder



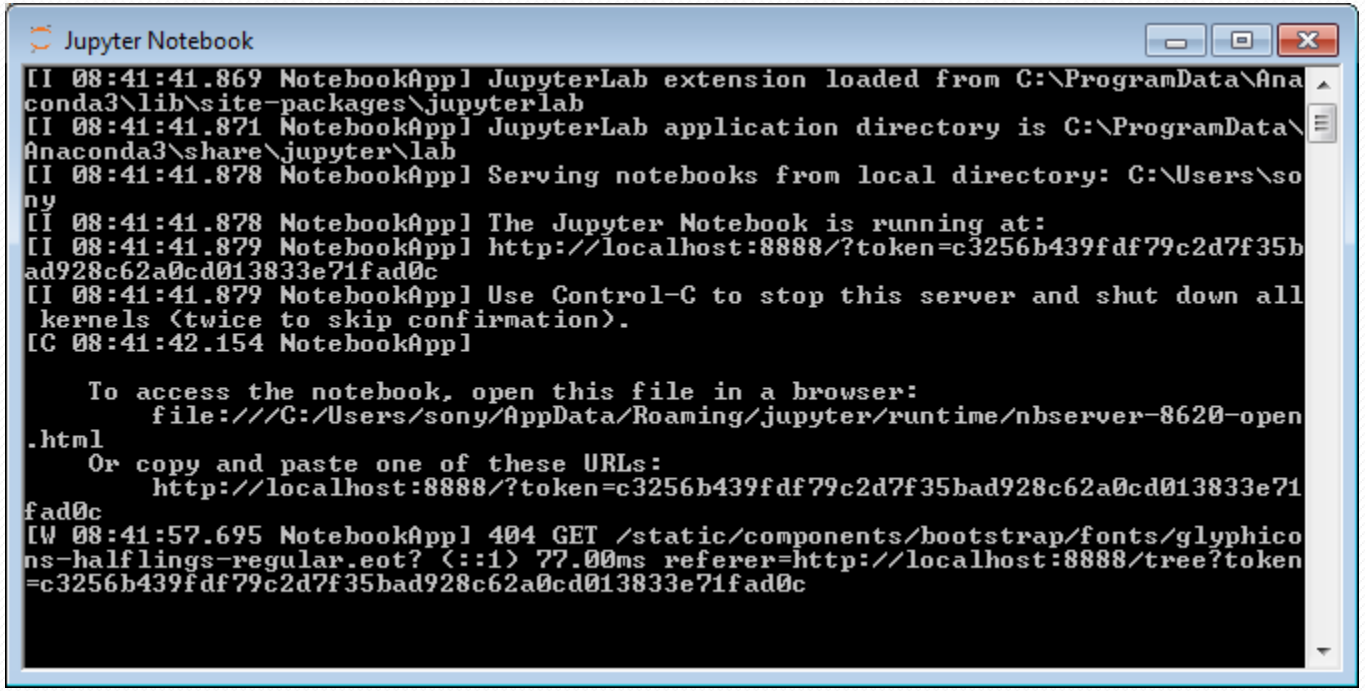


Jupyter Notebook

- Là một ứng dụng web mã nguồn mở cho phép bạn trình bày như một quyển vỡ khoa học và cho tương tác trực tiếp với nó
  - code
  - mô phỏng
  - văn bản diễn giải
- Là một công cụ tuyệt vời để:
  - Học về Python
  - Data processing / transformations
  - Vẽ biểu đồ
  - Machine learning
  - Data science

Jupyter Notebook

- Khởi động: Từ Anaconda Prompt bạn chỉ cần gõ jupyter notebook

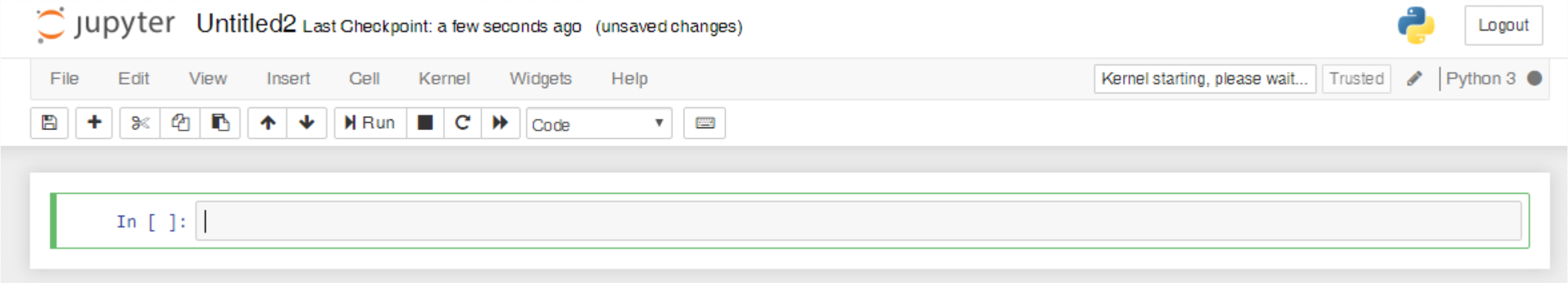


Jupyter Notebook

Jupyter Notebook

- Tạo một notebook python mới

Jupyter Notebook



Jupyter Notebook

- Một notebook bao gồm nhiều cell (ô). Khi tạo mới một notebook, bạn luôn được tạo sẵn một cell rỗng đầu tiên.

```
In [ ]:
```

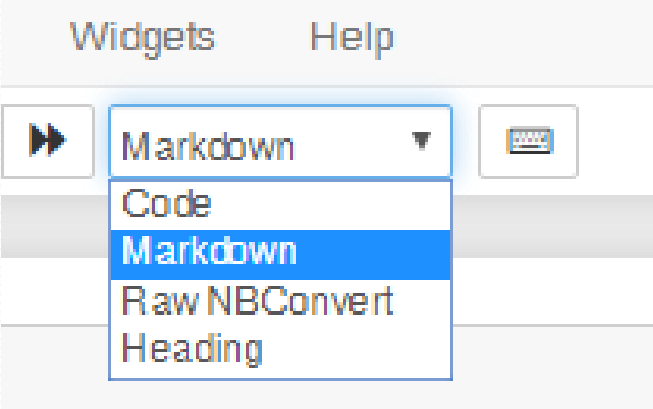
- Cell trên có kiểu là “Code”, điều đó có nghĩa là bạn có thể gõ code Python và thực thi ngay lập tức. Để thực thi code, bạn có thể nhấn nút Run cell hoặc nhấn Ctrl + Enter

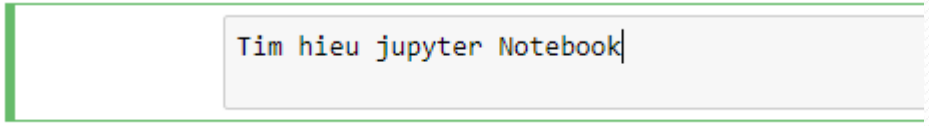
```
In [1]: print("hello world")

        hello world
```

Jupyter Notebook

- Bạn có thể chuyển loại cell từ Code thành Markdown để viết những đoạn văn bản giải thích code của bạn chẳng hạn





nhấn nút Run cell

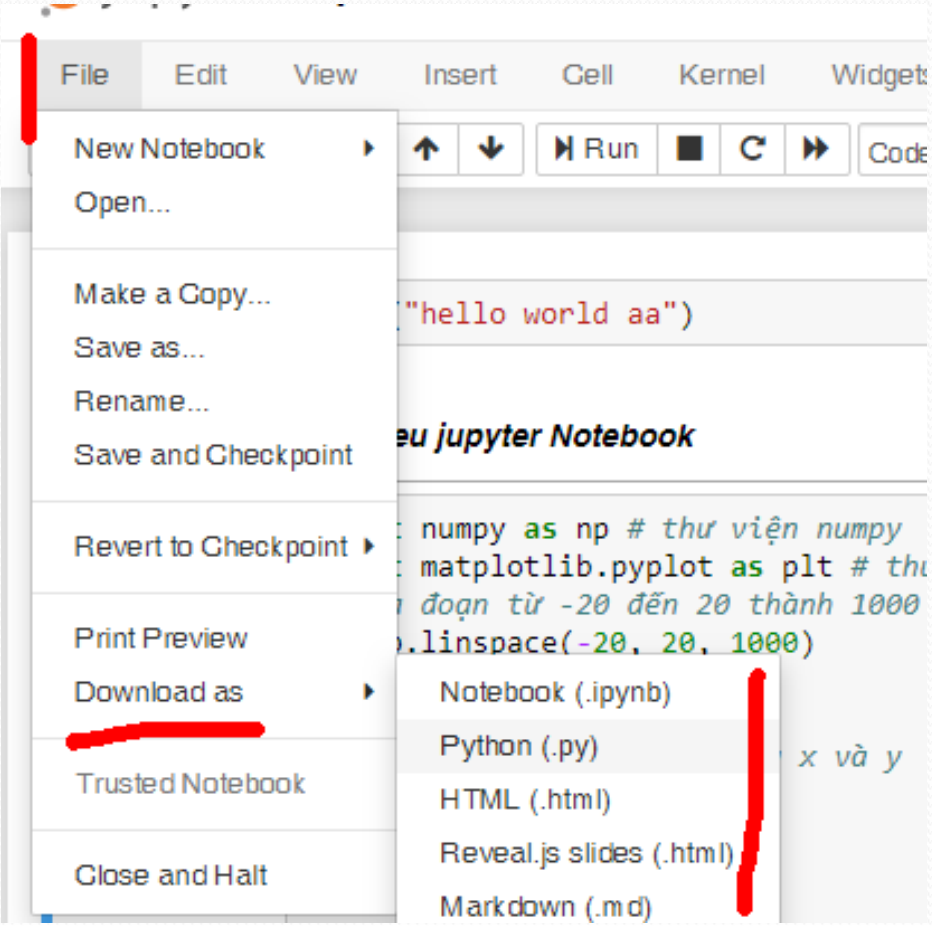
https://en.wikipedia.org/wiki/Markdown

Jupyter Notebook

- Checkpoints: Bằng cách tạo các Checkpoints lưu trạng thái hiện tại của notebook, Jupyter Notebook cho phép bạn có thể quay lại thời điểm tạo Checkpoints để kiểm tra hoặc hoàn tác trước đó.
- Để tạo Checkpoint, chọn *File -> Save and Checkpoint*. Nếu bạn muốn xem lại các Checkpoints trước đó thì chọn *File -> Revert to Checkpoint*.

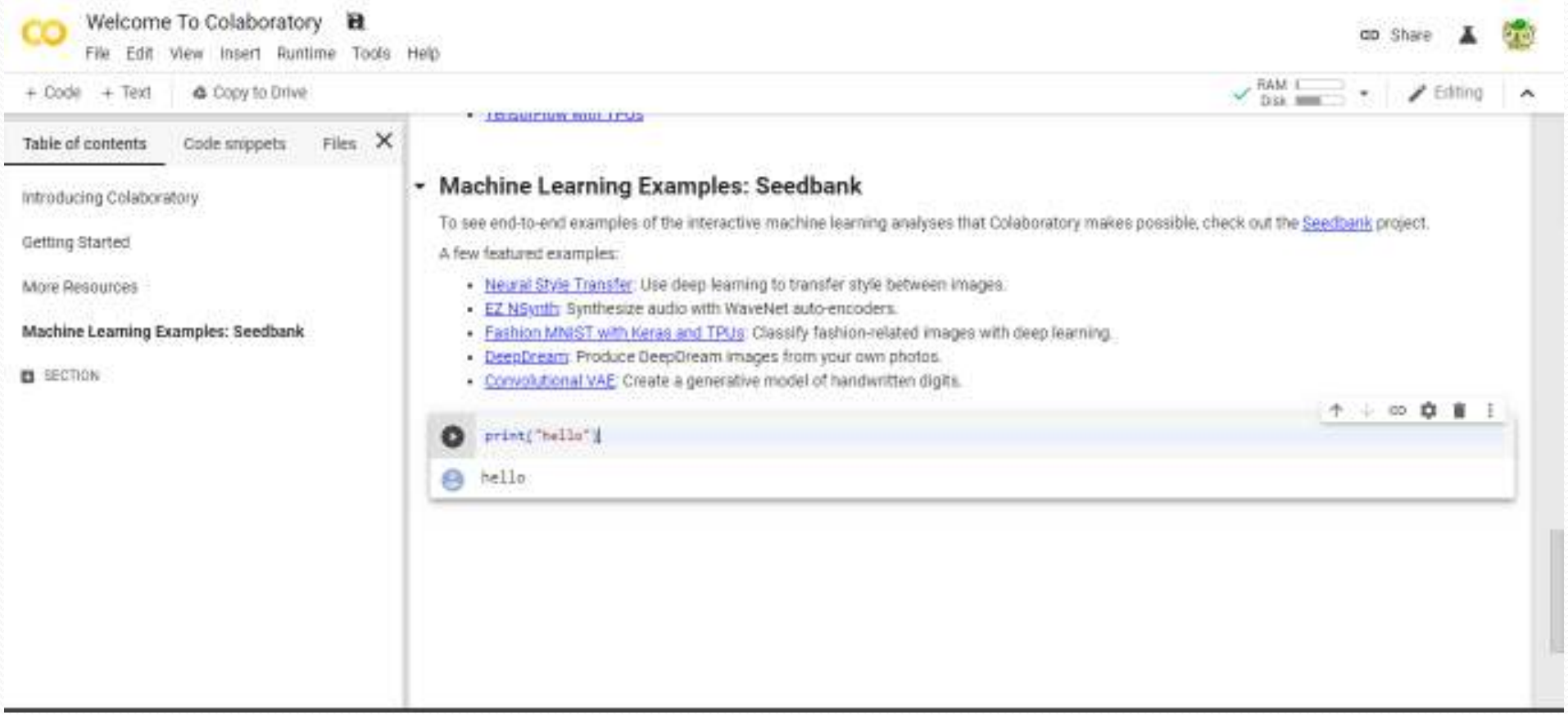
Jupyter Notebook

- **Chức năng Export notebook**



Dịch Vụ Hỗ Trợ Deep Learning Và Machine Learning

- Google Colab https://colab.research.google.com/

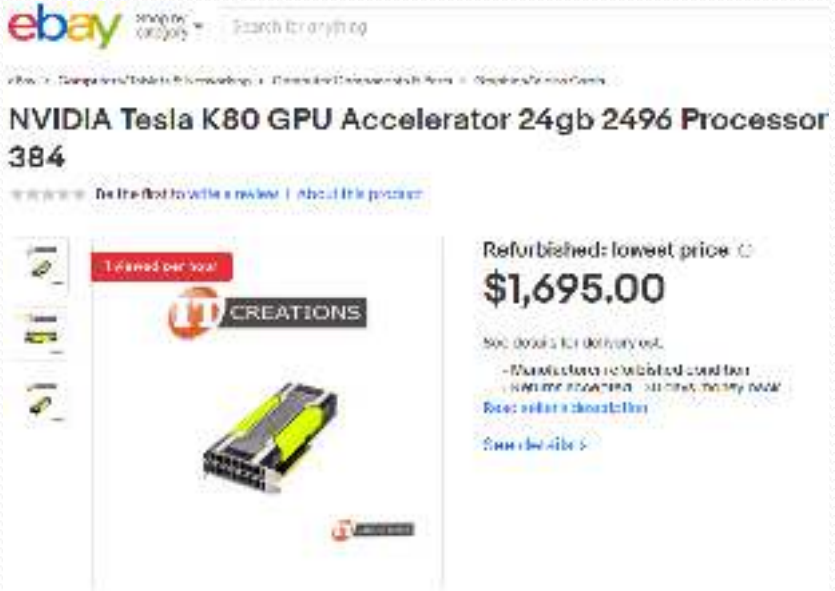


Google Colab

- Google Colab is a free tool for training neural network.
- Run on any browser (Chrome, Internet Explorer, Firefox, etc.)
- Train the model with NVIDIA Tesla K80

https://trituenhantao.io/tutorial/lam-quen-voi-google-colab/

Language program: Python
IDE: Jupyter Notebook
Train on data uploaded to Google Drive (default 15GB)
Maximum time used: 12 hours
Maximum RAM: 13GB



Kaggle

- https://www.kaggle.com/

- Tạo tài khoản

Kaggle Competitions

19 Active Competitions

Severstal: Steel Defect Detection
Can you detect and classify defects in steel?
Featured · Kernels Competition · 3 months to go · manufacturing, image data

$120,000
100 teams

Two Sigma: Using News to Predict Stock Movements
Use news analytics to predict stock price performance
Featured · Kernels Competition · 9 days to go · news agencies, time series, finance, money

$100,000
2,927 teams

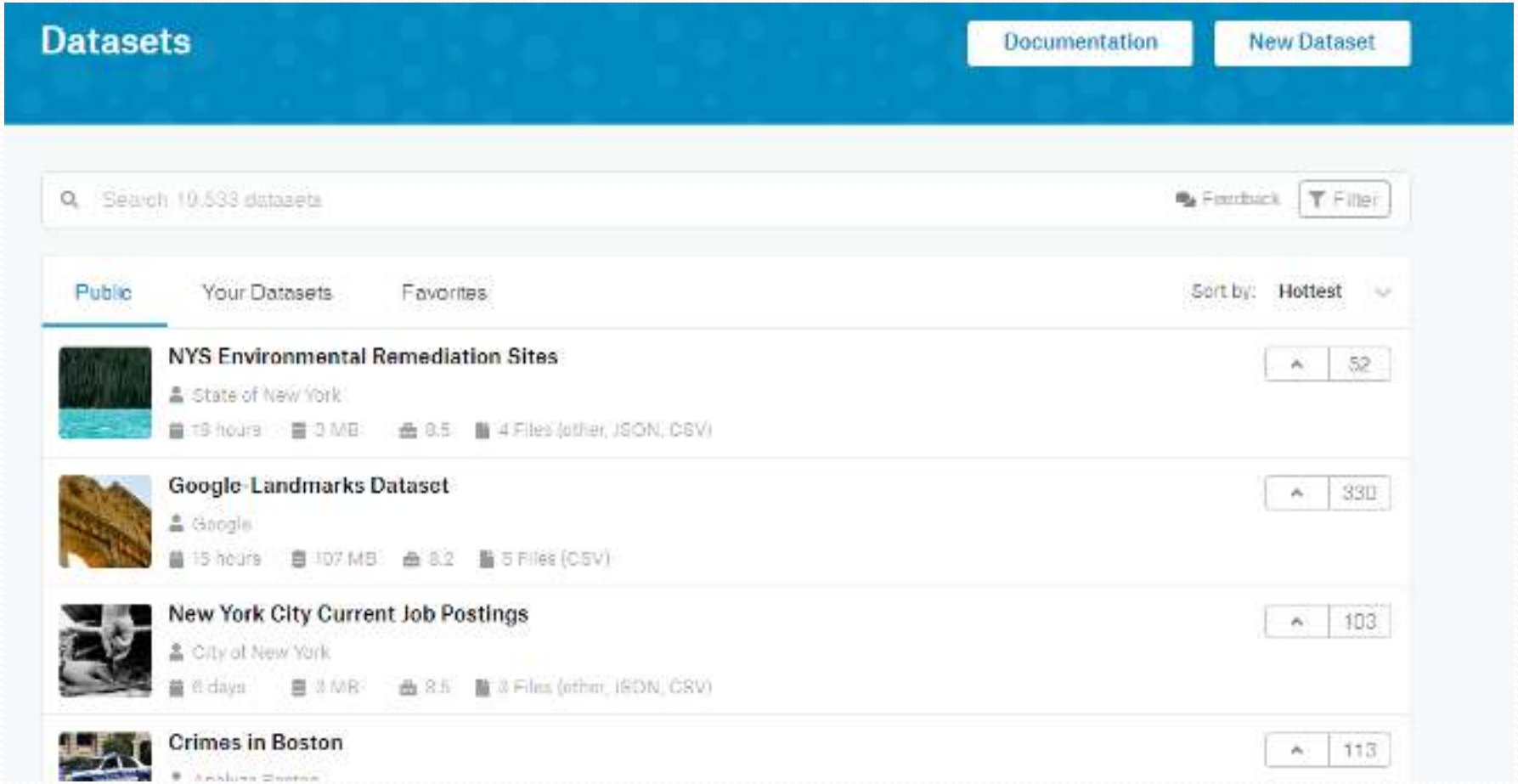
APTOS 2019 Blindness Detection
Detect diabetic retinopathy to stop blindness before it's too late
Featured · Kernels Competition · a month to go · healthcare, medicine, image data, multiclass classi...

$50,000
1,892 teams

SIIM-ACR Pneumothorax Segmentation
Identify Pneumothorax disease in chest x-rays
Featured · a month to go · image data, object segmentation

$30,000
1,182 teams

Kaggle Datasets



Kaggle Kernels

- Kaggle Kernels là một free platform để chạy môi trường Jupyter notebooks trên trình duyệt.
- Những xử lý trên notebooks của bạn được thực hiện trên server, do đó sẽ không ảnh hưởng tới hiệu năng của máy tính.

https://towardsdatascience.com/introduction-to-kaggle-kernels-2ad754ebf77

Kaggle Kernels

Select Kernel Type

Script

```
import numpy as np # linear algebra
import pandas as pd # data processing,

# Input data files are available in th

from subprocess import check_output
print(check_output(["ls", "../input"])

# Any results you write to the current
```

- Python, R, RMarkdown
- Runs all the code, every time
- Ideal for fitting a model and competition submissions
- Shares code for review and RMarkdown reports

Notebook

Introduction

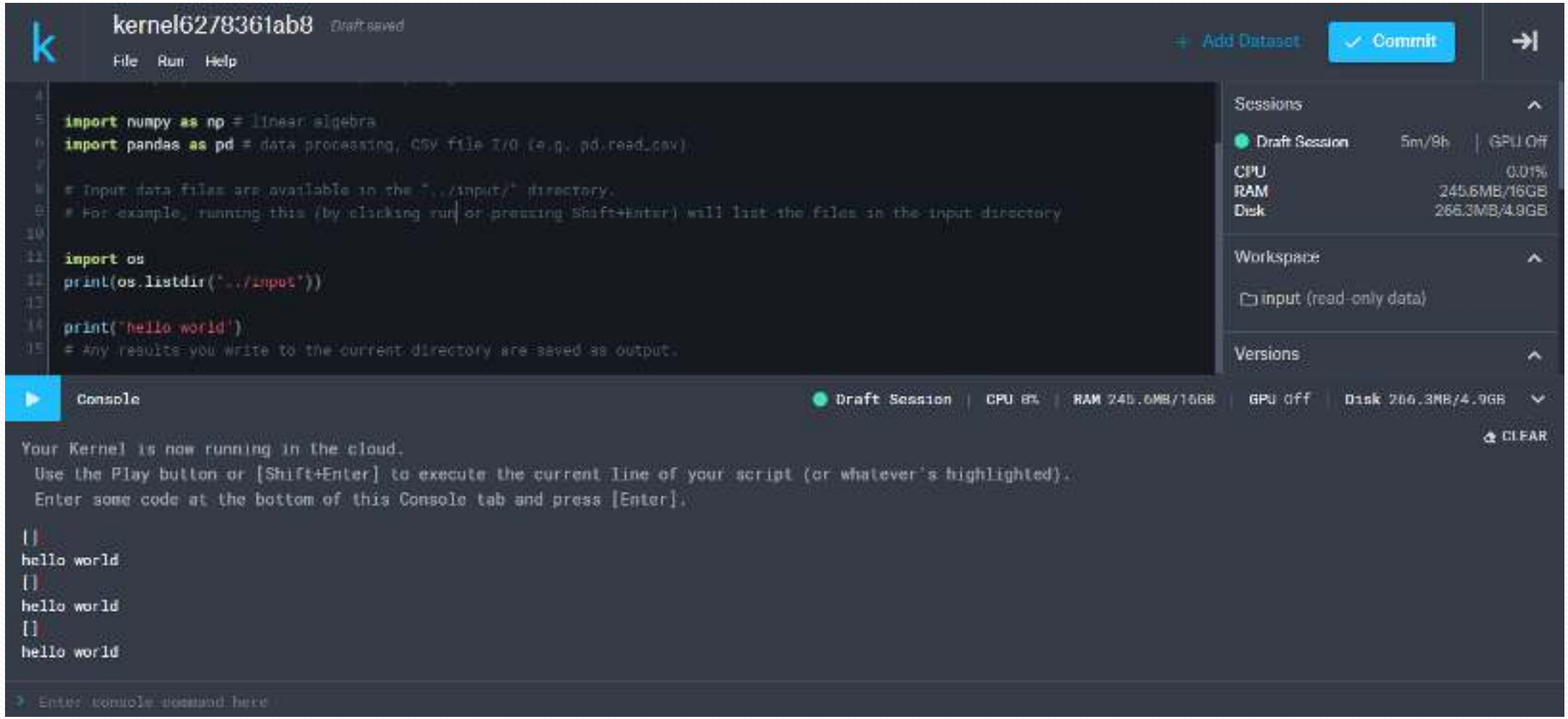
```
# Loading in the training data
train = pd.read_csv("./train.csv")
```

- Jupyter Notebooks in Python or R
- Runs cells of code and Markdown
- Ideal for interactive data exploration and polished analysis
- Shares insights through code & commentary

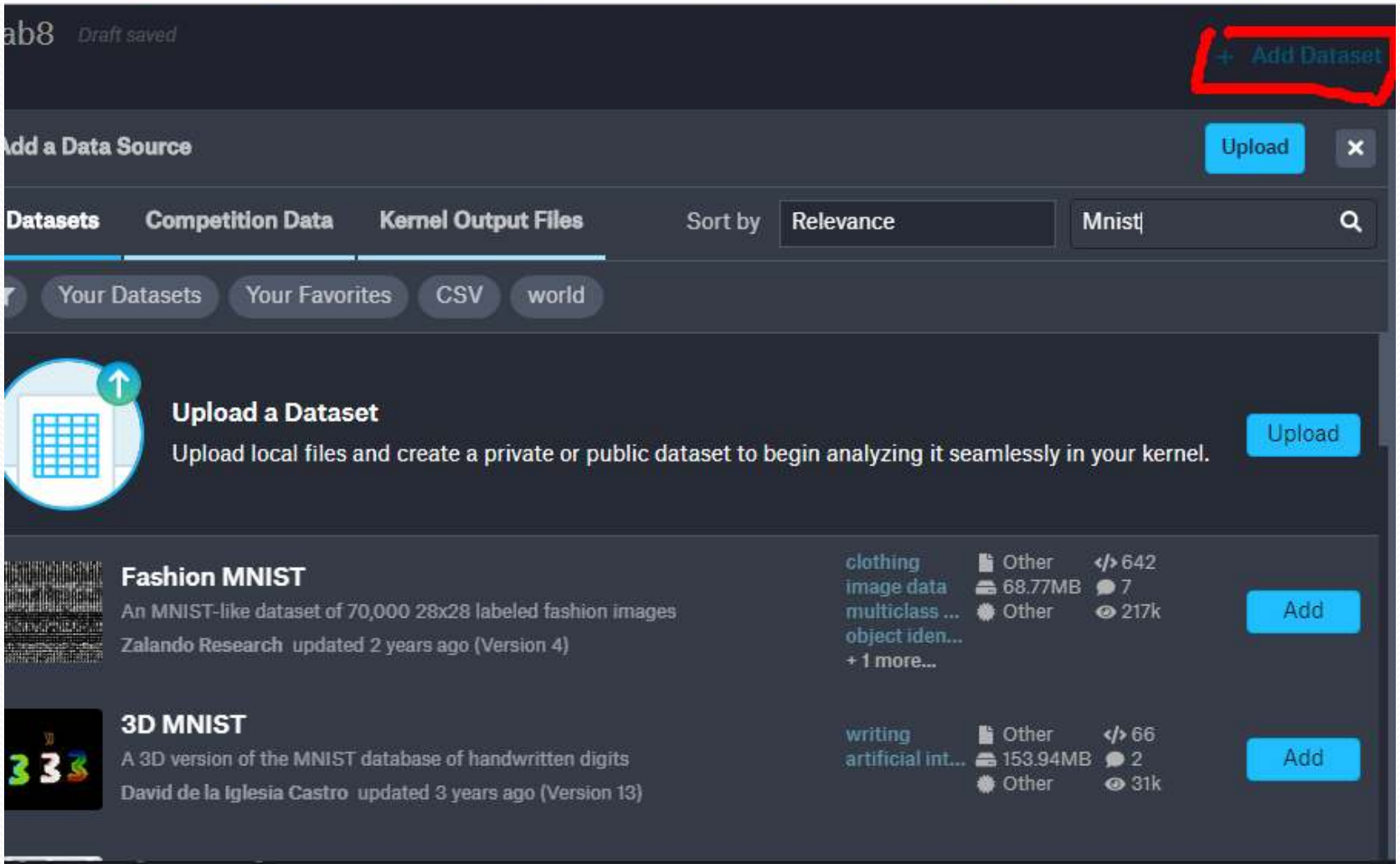
Kaggle Kernels

Old Resources	New Resources
2 CPUs	4 CPUs
8 GB RAM	16 GB RAM
512 MB Disk Space	1 GB Disk Space
20 min execution time	60 min execution time

Kaggle Kernels



Kaggle Kernels



```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
print(os.listdir("../input"))
print("hello world")
# Any results you write to the current directory are saved as output.
import tensorflow as tf
print(tf.__version__)
data_train_file = "../input/fashion-mnist_train.csv"
data_test_file = "../input/fashion-mnist_test.csv“
df_train = pd.read_csv(data_train_file)
df_test = pd.read_csv(data_test_file)
df_train.head()
```

Kaggle Kernels

```
[5 rows x 785 columns]
['fashion-mnist_test.csv', 'fashion-mnist_train.csv', 'train-images-idx3-ubyte', 't10k-images-idx3-ubyte', 't10k-labels-idx1-ubyte', 'train-labels-idx1-ubyte']
hello world
1.14.0
   label  pixel1  pixel2  pixel3  pixel4  pixel5  pixel6  pixel7  pixel8  \
0      2       0       0       0       0       0       0       0       0
1      9       0       0       0       0       0       0       0       0
2      6       0       0       0       0       0       0       0       5
3      0       0       0       0       1       2       0       0       0
4      3       0       0       0       0       0       0       0       0
```