Python I: Installation, Data Structure and Data Types

Day 3

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka

Data Scientists only use Jupyter Notebooks (Local or in Google Colab) to develop Python codes.

A. Benar

B. Salah

Data Scientists only use Jupyter Notebooks (Local or in Google Colab) to develop Python codes.

A. Benar

B. Salah

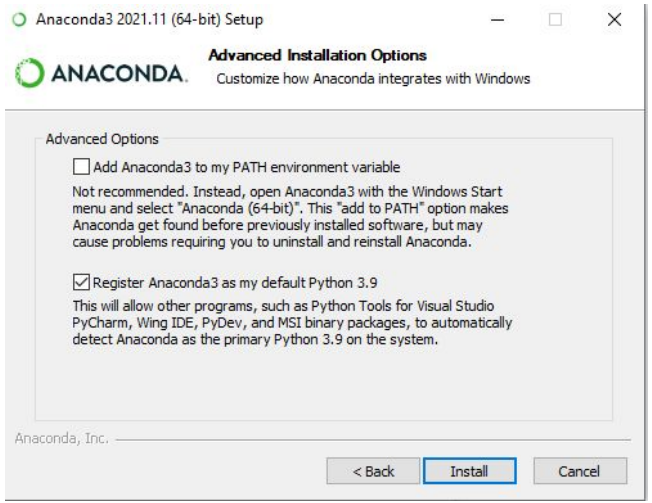
There are two types of developing Python codes. Writing a .py script, and writing a jupyter notebook (.ipnyb).

1. Installing Anaconda and Jupyter Notebook
2. Jupyter Notebook vs Python Script
3. Using IDE (VSCode)
4. Basic Operations in Python
5. Basic Data Types in Python

Installing Anaconda

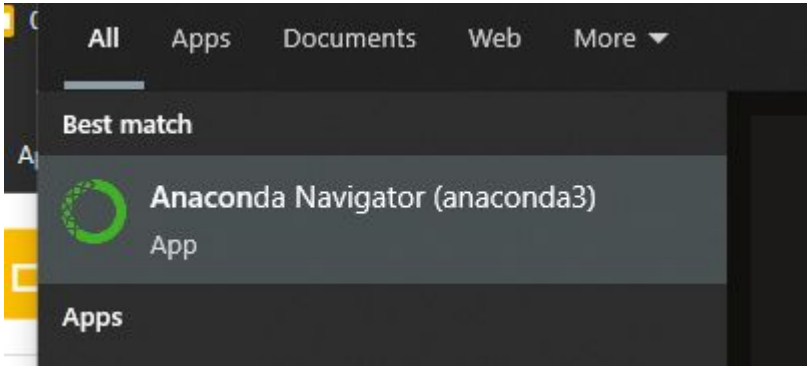
Installing Anaconda

1. Go to https://www.anaconda.com/products/individual
2. Launch the Installer
3. Click next for all the steps
4. Don’t change anything (just leave the default settings) especially on the window to the right
5. Done!
6. For macOS users, please refer to https://docs.anaconda.com/anaconda/install/mac-os/



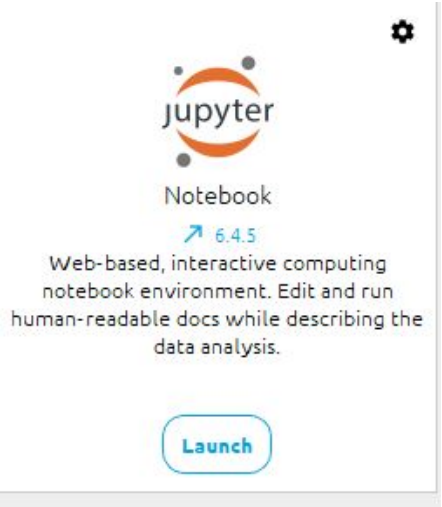
Installing Anaconda

1. Open Anaconda Navigator.



Installing Anaconda

2. Click Jupyter Notebook Launch

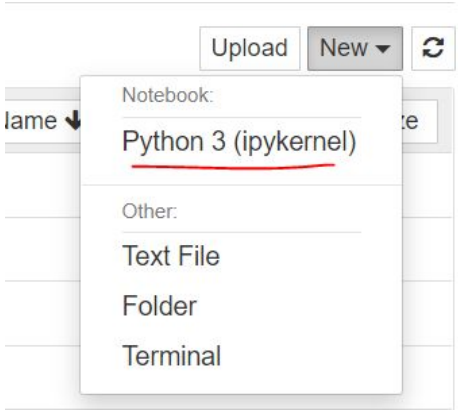


Installing Anaconda

3. Your browser should open a new tab like this:

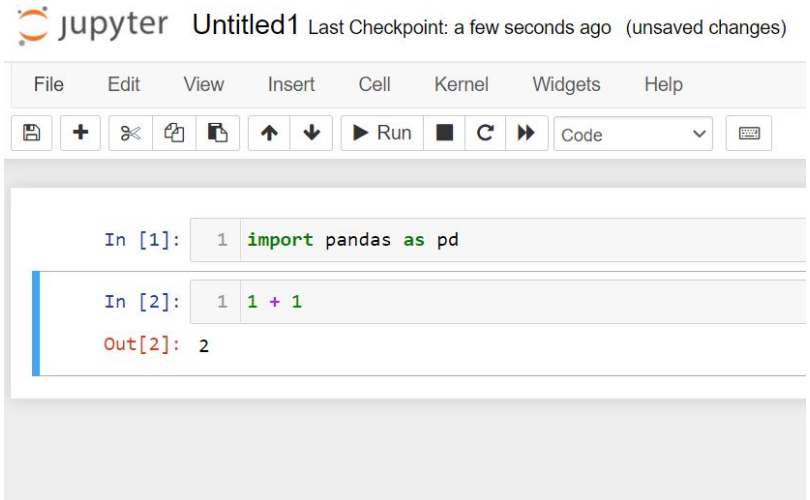
Installing Anaconda

4. We will make sure jupyter notebook has been installed and it can run correctly by creating a new notebook. Go to a folder of your choosing, and select this:



Installing Anaconda

Try to type something like this. If this returns no Error, then jupyter notebook works well in your machine.



Using Jupyter Notebook

Hands On!

Subject Matter Expert will demonstrate basic Jupyter Notebook usage. You can follow it by creating a brand new notebook!

Please get used to writing in jupyter notebook, and get familiar with the keyboard shortcuts. It will definitely make your life easier!

About Python Packages

Developers have created complex Python ‘functions’, ‘classes’, and .py scripts to automate a lot of things in Python. These are often ‘wrapped’ together in a ‘package’.

For example, to find the Quantiles of a Normal Distribution, you don’t need to hard-code the equation by yourself - there’s a package to do that called scipy.

You can make these ‘wrapped .py scripts’ available to use in your projects by importing the appropriate Python packages.

About Python Packages

Common Python packages for Data Scientists:

- **Pandas (dataframe / tabular data processing)**
- **Numpy (mathematical array operations)**
- **Matplotlib and Seaborn (visualization)**
- **Scikit-Learn (machine learning)**

Installing anaconda in your machine should already install these packages.

About Python Packages

Let's try to import a package that has not been installed in our local machine. Let's try to do this. It should return an error. Because we haven't installed the package.

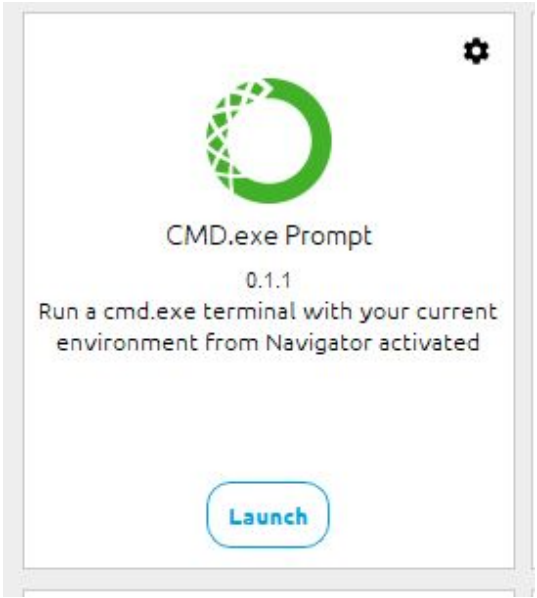
```
1 import lightgbm
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_7784/1544401774.py in <module>
----> 1 import lightgbm

ModuleNotFoundError: No module named 'lightgbm'
```

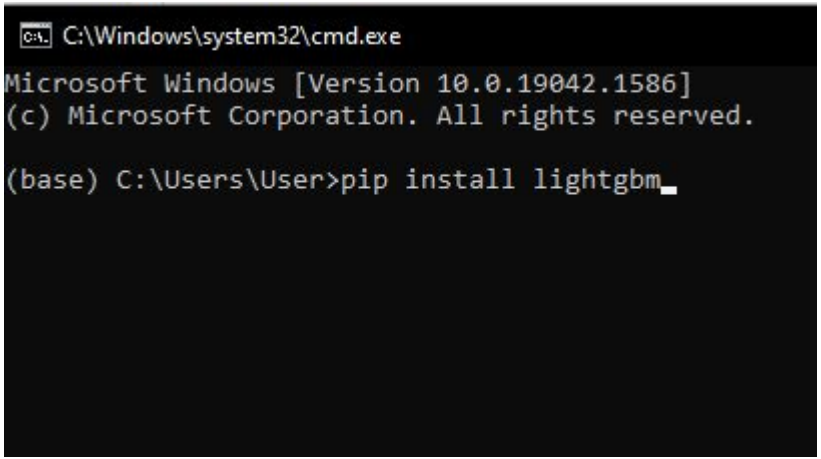
About Python Packages

To install new packages, we need to go to Anaconda Prompt



About Python Packages

Always write ‘pip install *packagename*’ to install a new package. Then, press enter.



About Python Packages

Let’s check, now this cell does not return an error when we run it. This means, this ‘lightgbm’ package is already in our system.

```
In [11]:  1  import lightgbm
```

About Python Packages

Installing Python with anaconda should bring almost all the required package in fundamentals of Data Science.

So first, when you see a new package, try to import it first to see if anaconda has already got it for you.

If not, then you can install it like what we have just done.

Jupyter Notebook vs Script

What is IDE?

Software developers often use something called ‘IDE’ when writing code.

Python has its own standalone IDE called “Pycharm”, and it is closely related to anaconda.

However, the ‘best’ IDE which is most used by a lot of developers (at least in 2022) is called ‘VSCode’.

You can download it here: https://code.visualstudio.com/

IDE (Integrated Development Environment)

What is IDE?

For the sake of the time and to not confuse beginners, we won't dive deep into how to use an IDE.

For now, the Subject Matter Expert will demonstrate what can be done with an IDE (basic .py script development)

Basic Operations in Python

Hands-On

Now, let's jump back to Jupyter Notebook and learn the basic operations in Python!

File that you should open:
day_03_operations_and_data_types.ipnyb

Basic Data Types in Python

Hands-On

Now, let's jump back to Jupyter Notebook and learn the basic Data Types in Python!

File that you should open:
day_03_operations_and_data_types.ipnyb

Terima kasih!

Ada pertanyaan?

Assignment

Assignment

1. Ketika kita ingin menuliskan kode untuk dieksekusi secara berurut dan untuk melakukan automasi pekerjaan, maka yang kita gunakan adalah...:
   a. Jupyter Notebook
   b. JavaScript
   c. CMD.exe
   d. Python Script

Assignment

2. Ketika kita ingin menuliskan kode untuk melakukan eksplorasi data dan pembuatan awal / eksperimentasi model Machine Learning, maka yang kita gunakan adalah...
   a. Jupyter Notebook
   b. JavaScript
   c. CMD.exe
   d. Python Script

Assignment

3. Jika kita ingin menggunakan pandas versi 1.4 di project A, namun ingin menggunakan pandas versi 1.3 di project B, maka yang paling benar dilakukan adalah...
   a. Membuat dua buah folder terpisah untuk masing-masing project
   b. Membuat dua buah virtual environment terpisah untuk masing-masing project
   c. Menginstall 2 versi Python untuk masing-masing project
   d. Membuat project A di Jupyter Notebook dan project B di Google Colab

Assignment

4. File-file .py script yang kita butuhkan untuk sebuah project sudah terkumpul dalam sebuah folder. Sekarang, kita ingin membuat sebuah daftar (‘requirements.txt’) berisi package apa saja yang telah kita ‘import’ dalam satu folder tersebut. Package apa yang mampu membantu kita?
   a. Pip Freeze
   b. Requests
   c. OS (operating system)
   d. Pipreqs