

# Introduction to Python

Data Mining

Prof. Sujee Lee

Department of Systems Management Engineering

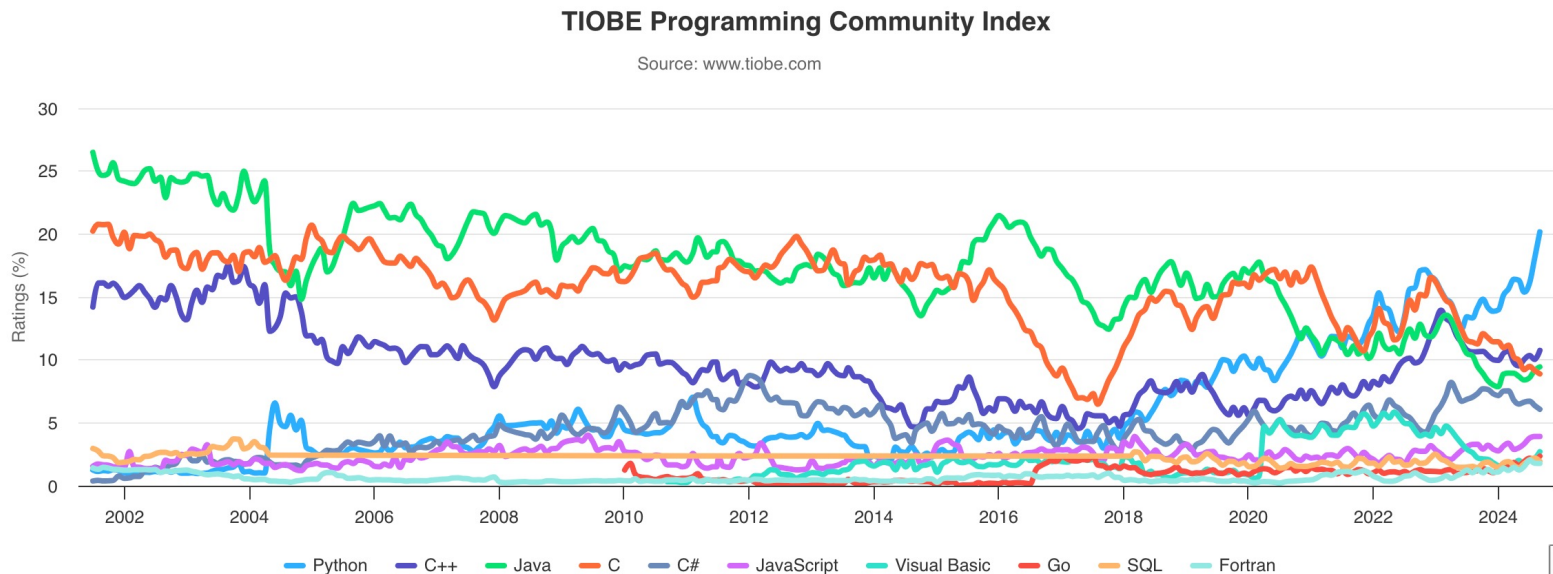
Sungkyunkwan University



# Python

## ■ Why Python?

- Python is a widely used, general purpose programming language.
- Easy to start working with.
- Scientific computation functionality similar to Matlab and Octave.
- Used by major machine learning frameworks such as Scikit-Learn
- Used by major deep learning frameworks such as PyTorch and TensorFlow.



# Python Environment

---

## ■ Installation

- As long as you can use Python and Jupyter notebook, any method can be used!

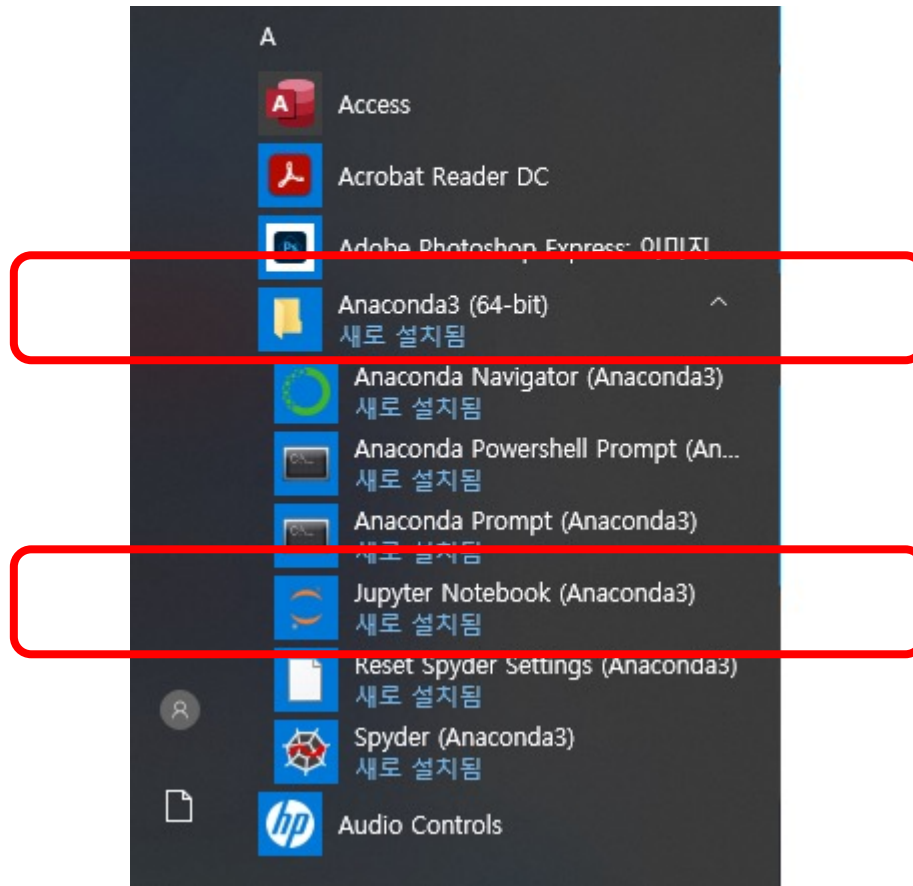
## ■ Recommendation

- Quick Installation : Install Anaconda3
  - <https://www.anaconda.com/products/individual>
- Quick Use : Google Colab
  - <https://colab.research.google.com/>
  - Note : I'll use Jupyter notebook. Shortcut keys may be different between Jupyter and Colab. But, all functionalities are the same (at least for this course)



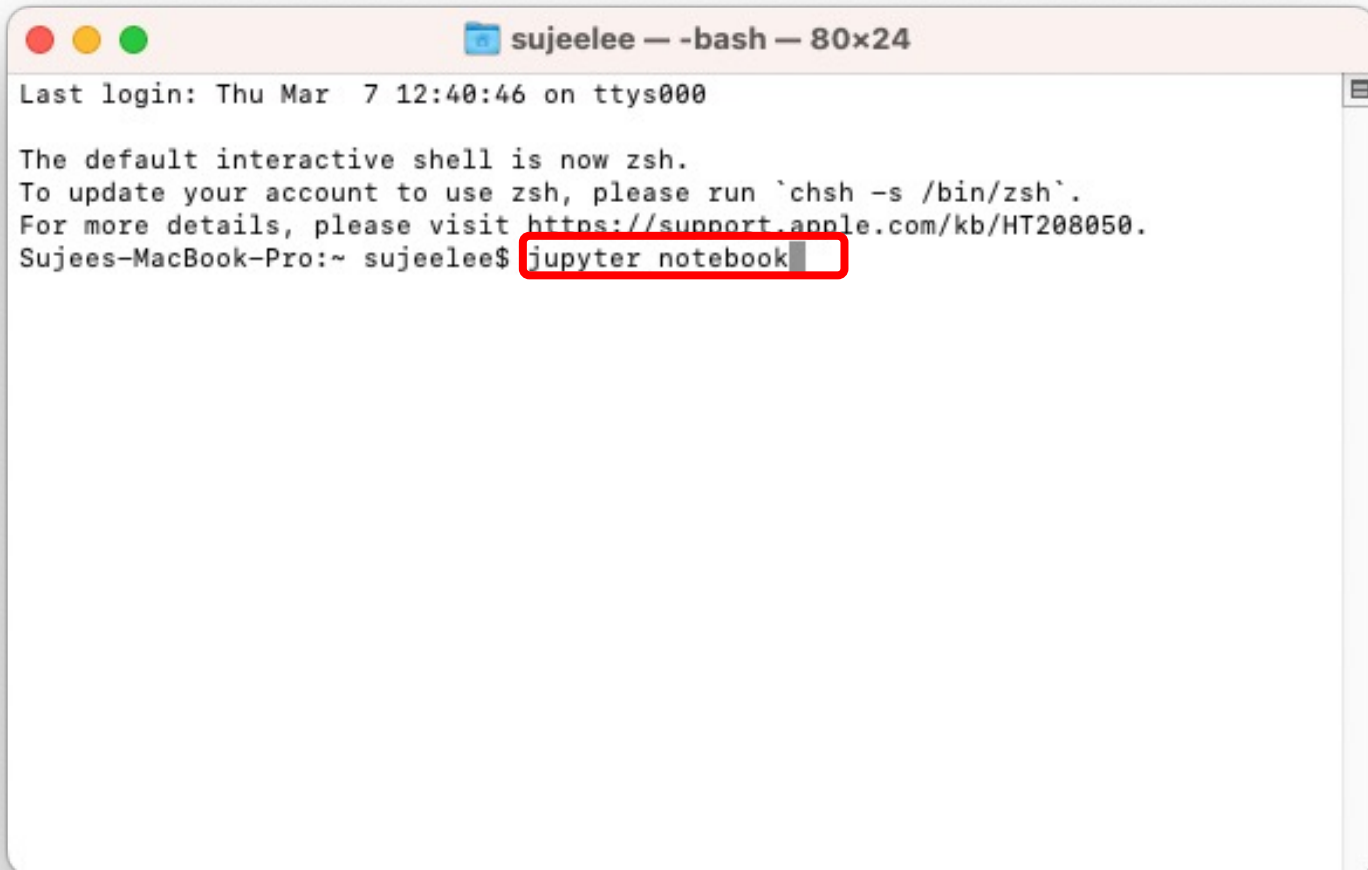
# Python Environment

- After Anaconda3 is installed (Windows OS)



# Python Environment

- After Anaconda3 is installed (Mac / Linux OS)
- In Terminal,

A screenshot of a macOS Terminal window. The title bar at the top reads 'sujeelee — -bash — 80x24'. The terminal content shows the login message 'Last login: Thu Mar 7 12:40:46 on ttys000', followed by system messages about the default shell being zsh and instructions to run 'chsh -s /bin/zsh'. The prompt 'Sujees-MacBook-Pro:~ sujeelee\$' is followed by the command 'jupyter notebook' which is highlighted with a red rectangular box.

```
sujeelee — -bash — 80x24
Last login: Thu Mar 7 12:40:46 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
Sujees-MacBook-Pro:~ sujeelee$ jupyter notebook
```



# Python Environment

## ■ In Google Colab

The screenshot shows the Google Colab web interface. A modal dialog titled '노트 열기' (Open Notebook) is displayed in the center. The dialog has a search bar at the top with the text '노트북 검색' (Search notebooks). Below the search bar is a table with columns: '제목' (Title), '마지막 연 시간' (Last modified time), and '처음 연 시간' (First modified time). The table contains one entry: 'Colaboratory에 오신 것을 환영합니다' (Welcome to Colaboratory), with a last modified time of '오후 2:54' and a first modified time of '오후 2:54'. To the left of the table is a sidebar with options: '예' (Yes), '최근 사용' (Recent), 'Google Drive', 'GitHub', and '업로드' (Upload). The '업로드' option is highlighted with a red rectangle. At the bottom of the dialog, there is a button labeled '+ 새 노트' (New notebook) and a '취소' (Cancel) button. The background shows the Colab interface with a sidebar on the left containing '시작하기' (Get started), '데이터 과학' (Data science), '머신러닝' (Machine learning), and '추가 리소스' (Additional resources). The main area displays 'Colab 시작' (Colab start) and 'Colab이란?' (What is Colab?).



# Requirements

---

## ■ Python Packages for Data Analysis

- **Jupyter Notebook** : Browser-based interactive programming environment
- **NumPy** : Basic array operation and manipulation
- **SciPy** : High-level scientific computing (advanced linear algebra, optimization, statistical distributions, etc.)
- **Pandas** : Data wrangling and analysis (Excel files, CSV files, SQL, etc.)
- **Matplotlib & Seaborn** : Visualization
- **scikit-learn** : Machine learning (on top of NumPy and SciPy)



# Pandas

- **Pandas : An essential library for handling structured data in table or database formats.**
  - Data is represented in a table format using a data structure called DataFrame.
  - Using Pandas, you can create table-like data or load files such as CSV and Excel.
  - It is convenient for exploring and modifying data.

```
In [4]: ## Create a simple dataset
data = {'Name': ["John", "Anna", "Peter", "Linda"],
        'Location': ["New York", "Paris", "Berlin", "London"],
        'Age' : [24, 13, 53, 33]}

df = pd.DataFrame(data)
```

```
In [5]: df
```

```
Out[5]:
```

	Name	Location	Age
0	John	New York	24
1	Anna	Paris	13
2	Peter	Berlin	53
3	Linda	London	33

