



데사 B0002

# 데이터마이닝이해와실습

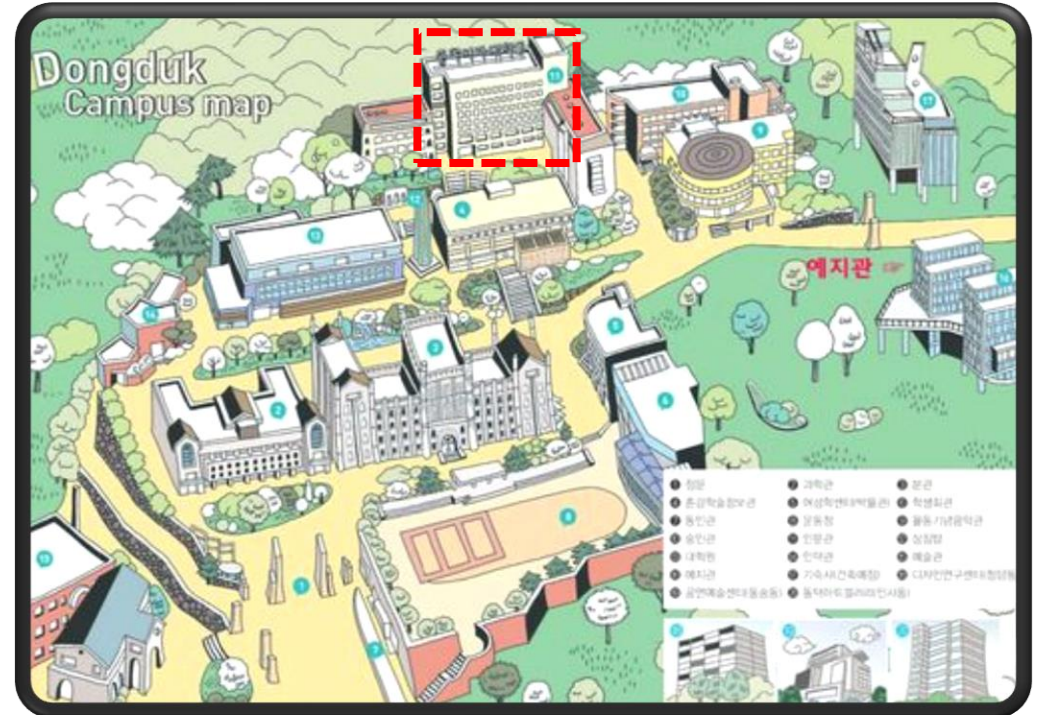
2분반

김 태 완

kimtwan21@dongduk.ac.kr

# 데이터마이닝 이해와 실습

- Time and Location
  - 수업 요일/교시 : 월 5, 6교시
  - 강의실 : 인문관 C0302호
- Instructor : 김 태 완
  - Office : 인문관 B619호 / 예지관 A506호
  - Phone : 02-940-4751
  - E-mail : [kimtwan21@dongduk.ac.kr](mailto:kimtwan21@dongduk.ac.kr)
  - Office hours
    - Smart Class System 쪽지 x
    - E-mail 로 면담 요청



## 교수 소개



동덕여자대학교 데이터사이언스전공 조교수 김 태 완

- 학부: 연세대학교 전기전자공학과 2004 – 2008
- 석/박사: 연세대학교 전기전자공학과 2008 – 2015
- 경력사항: SK텔레콤 2015-2021 (Senior AI Researcher)



Image & Video  
Data Analysis

2015



Machine Learning &  
Data Analysis Platform

2022

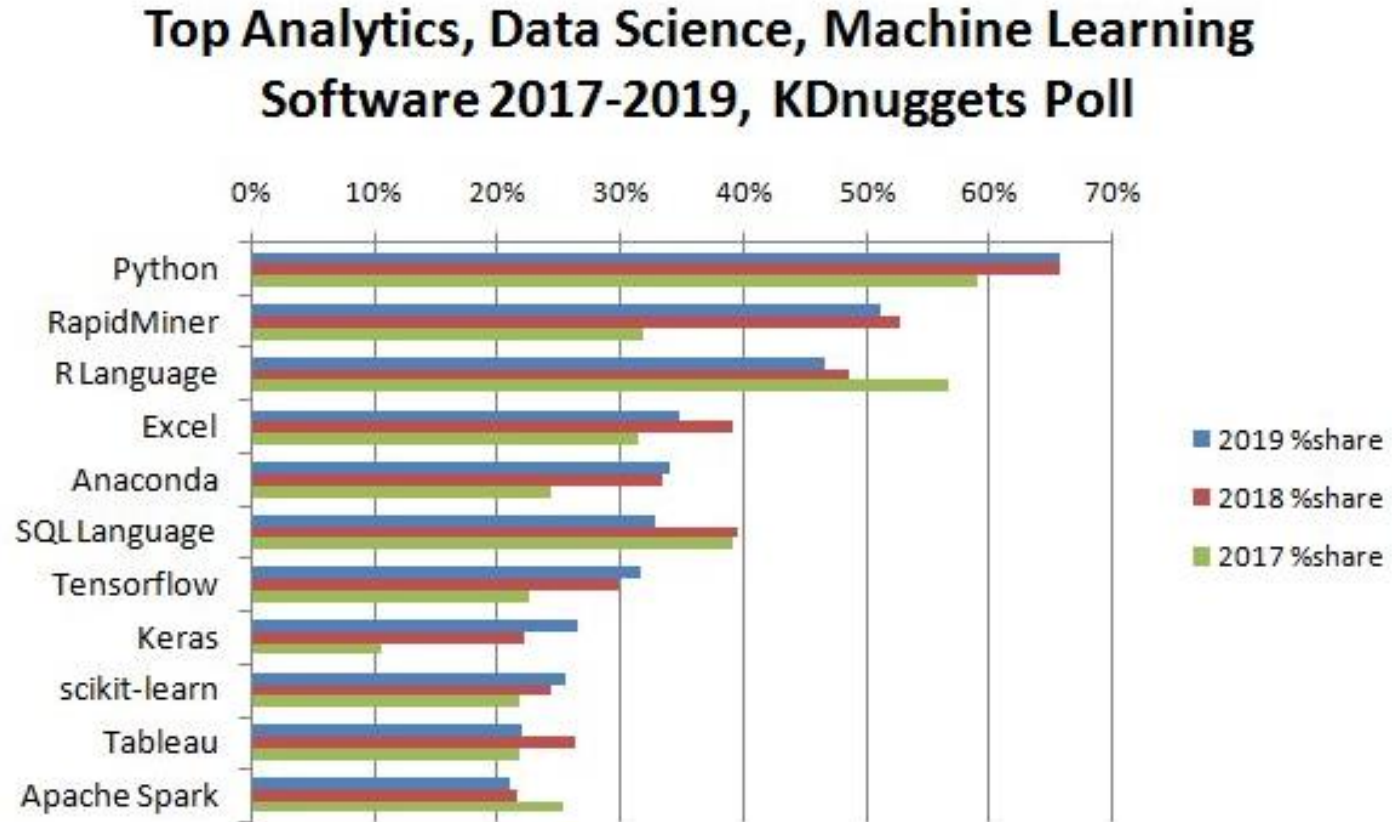


Data Science

# 데이터 분석 도구

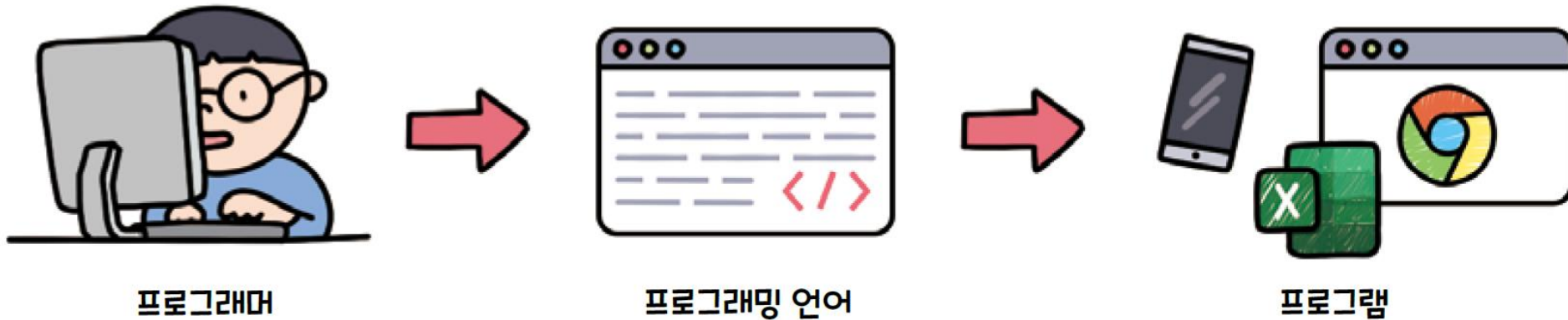
- 데이터 분석가들이 많이 사용하는 데이터 분석 툴

<https://pypl.github.io/PYPL.html>



# 프로그램, 프로그래밍 언어, 프로그래머

- **프로그래밍(Programming)**
  - 프로그램을 만드는 작업
- **프로그래밍 언어(Programming Language)**
  - 사람이 이해하는 말을 언어(Language)라고 부르듯이 컴퓨터가 이해하는 언어를 의미함
- **프로그래머(Programmer)**
  - 프로그래밍 언어를 사용하여 소프트웨어나 앱을 만드는 직업을 가진 사람을 의미함



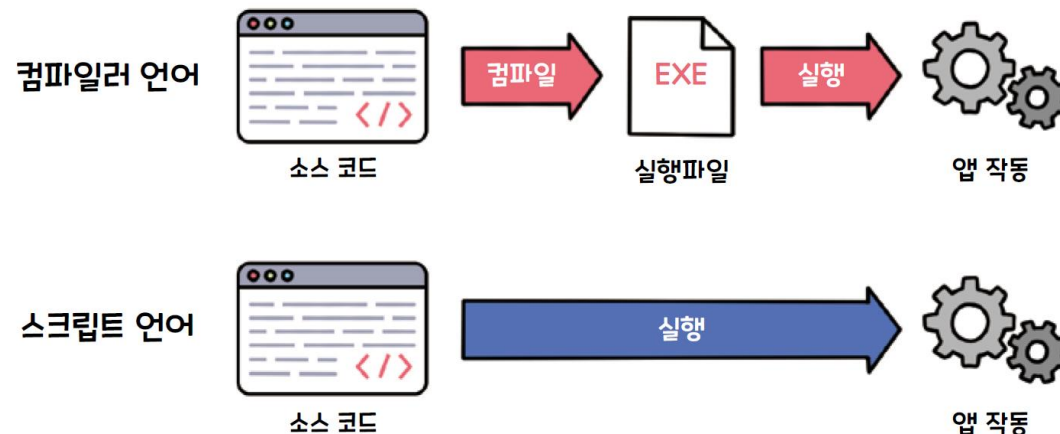
# 프로그래밍 언어의 분류

- 컴파일러(Compiler) 언어

- 컴파일(compile): 소스 코드를 기계어로 번역하는 과정
- 소스 코드를 실행 가능한 기계어로 일괄 번역한 후에, 번역이 완료된 파일(\*.exe, \*.class 등)을 실행하는 언어를 의미함
- 대표적인 컴파일러 언어: C, C++, Java, C# 등

- 스크립트(Script) 언어

- 소스 코드를 한 줄씩 읽어서 실행하는 언어
- 스크립트 언어는 별도의 실행파일이 생성되지 않음
- 대표적인 스크립트 언어: 파이썬, 자바스크립트(JavaScript), 펄(Perl) 등







Now, it's my belief that Python is a lot easier than to teach to students programming and teach them C or C++ or Java at the same time because all the details of the languages are so much harder. Other scripting languages really don't work very well there either.

- Guido van Rossum -

quoteparrot.com

- 귀도 반 로섬이 취미로 만든 프로그래밍 언어
  - C언어로 제작
  - 1991년 최초 발표
  - 2000년 Python2, 2008년 Python3
- Google에서 만들어진 SW 50% 이상이 Python
- Dropbox, Instagram, ...
- 이해하기 쉬워 공동 작업과 유지 보수 편함

# Python 경험하기

---

- 특징
  - 강력한 기능을 무료로 사용할 수 있다
    - 파이썬은 오픈 소스이며, 비용을 지불하지 않고 무료로 사용 가능. 다양한 추가 라이브러리도 무료
  - 읽기 쉽고 사용하기 쉽다
    - 직관적인 코드를 사용해 c나 자바 같은 언어보다 읽기 쉬워 프로그램을 빨리 제작 → 비용 절감 효과 제공

```
if 3 in [1,2,3,4] : print("3이 있습니다")
```

- 다양하고 강력한 외부 라이브러리들이 풍부하다
  - Numpy, Panda 등 매우 유용한 라이브러리 많음
- Python 과 C는 찰떡궁합
  - 상대적으로 쉽지만 느린 Python + 빠른 C 언어 조합하여 최적의 솔루션 생성
- Python은 개발 속도가 빠르다
  - *Life is too short, You need Python.*



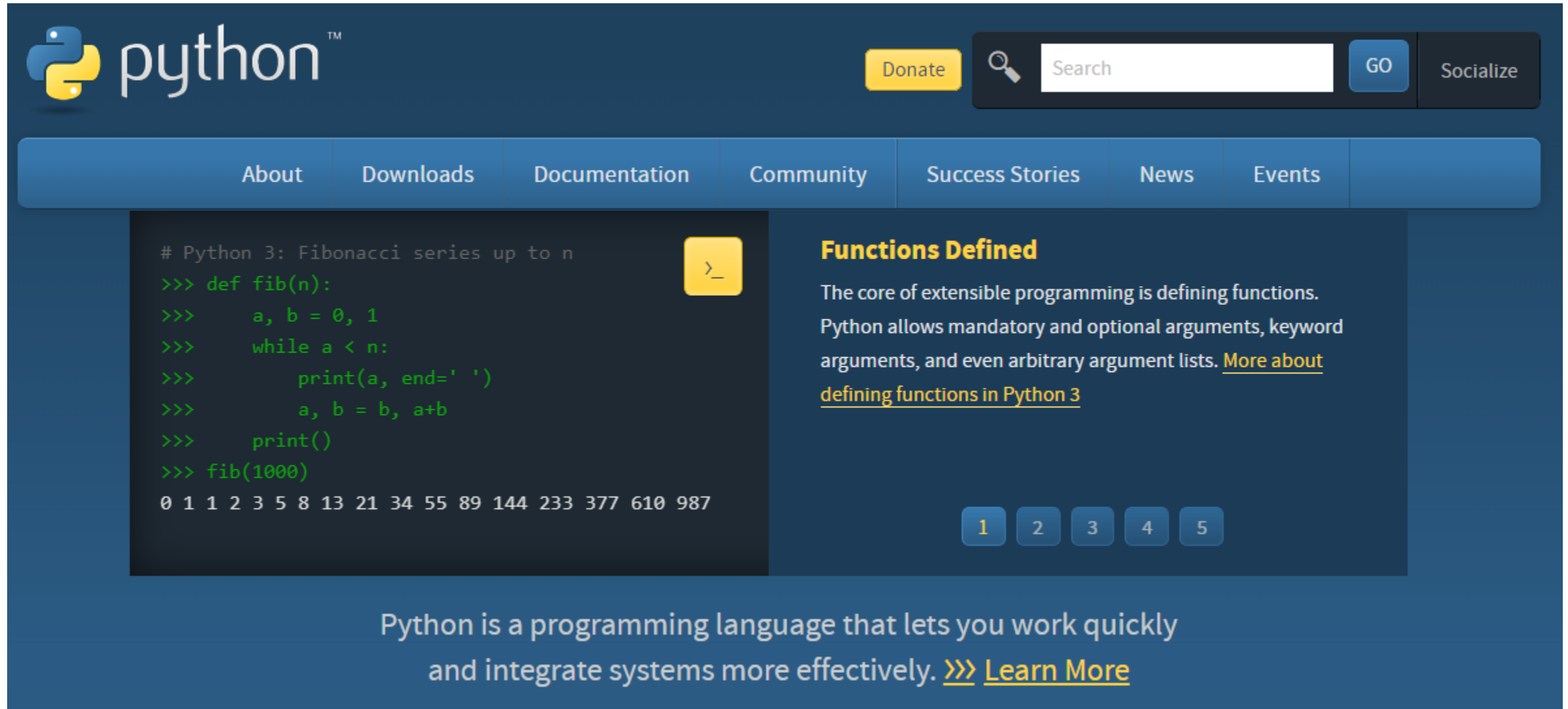
# Python 설치하기

---

- 단점
  - 속도가 느리다 (c에 비해...)
    - 파이썬은 컴파일러 언어가 아닌 스크립트 언어이기 때문에 컴파일러 언어보다 느림
    - 이를 보완하려고 많은 파이썬 패키지를 최적화 시키고 있음
  - 모바일 컴퓨팅 분야에 지원이 약하고 하드웨어 제어 등과 관련된 부분 사용이 어려움
  - 하드웨어 제어 등과 관련된 부분도 사용하기가 쉽지 않음
- 하지만 “데이터사이언스전공” 의 MAIN programming language 는 “파이썬”

# Python 설치하기

- Python 설치하기: <https://www.python.org/>



The image is a screenshot of the Python.org homepage. At the top left is the Python logo, which consists of two interlocking snakes, one blue and one yellow, followed by the word "python" in a lowercase, sans-serif font with a trademark symbol. To the right of the logo is a yellow "Donate" button. Further right is a search bar with a magnifying glass icon, a "GO" button, and a "Socialize" button. Below these elements is a horizontal navigation bar with seven blue buttons labeled "About", "Downloads", "Documentation", "Community", "Success Stories", "News", and "Events". The main content area is divided into two columns. The left column features a dark background with a yellow terminal icon containing a prompt character. Below the icon is a code snippet for a Fibonacci function. The right column has a yellow heading "Functions Defined", followed by a paragraph of text and a link. At the bottom of the right column are five blue buttons numbered 1 through 5. A footer at the very bottom contains a sentence about Python and a link.

python™

Donate

Search GO Socialize

About Downloads Documentation Community Success Stories News Events

```
# Python 3: Fibonacci series up to n
>>> def fib(n):
>>>     a, b = 0, 1
>>>     while a < n:
>>>         print(a, end=' ')
>>>         a, b = b, a+b
>>>     print()
>>> fib(1000)
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987
```

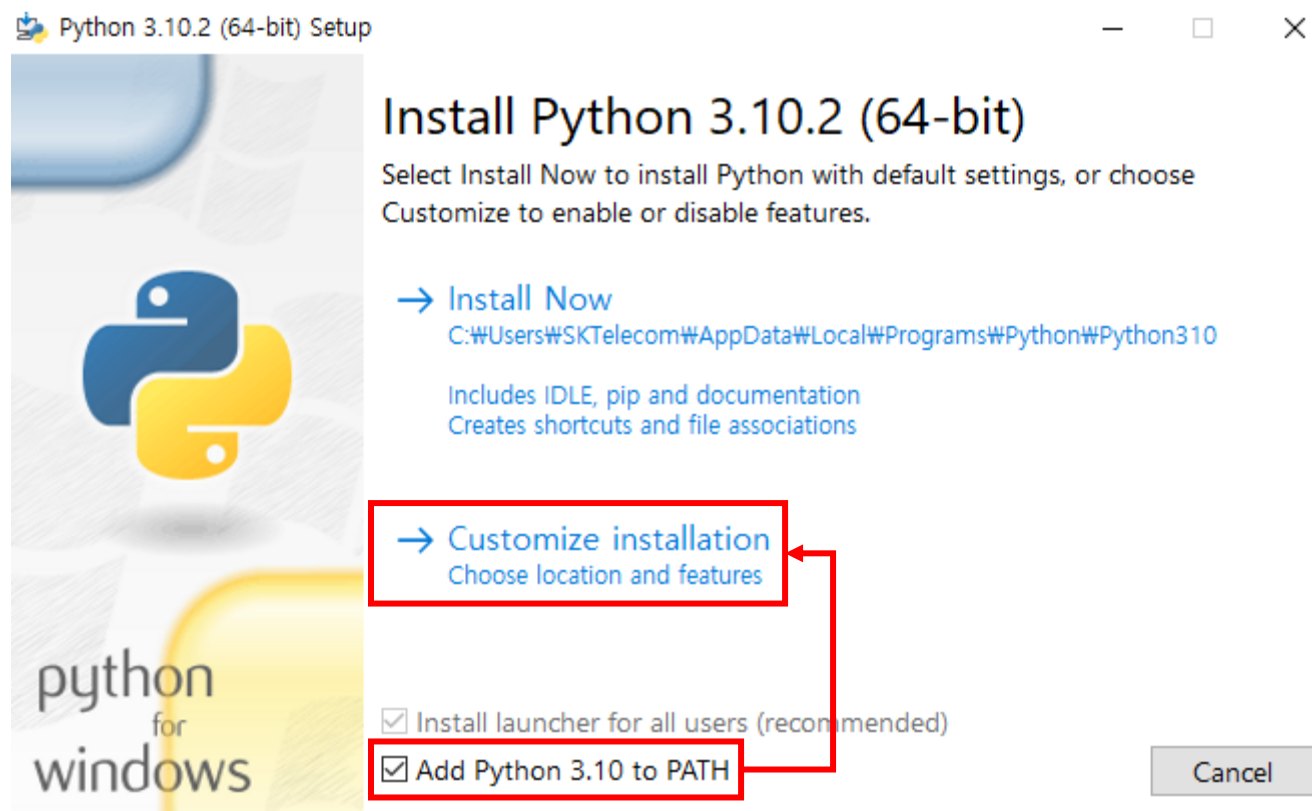
**Functions Defined**

The core of extensible programming is defining functions. Python allows mandatory and optional arguments, keyword arguments, and even arbitrary argument lists. [More about defining functions in Python 3](#)

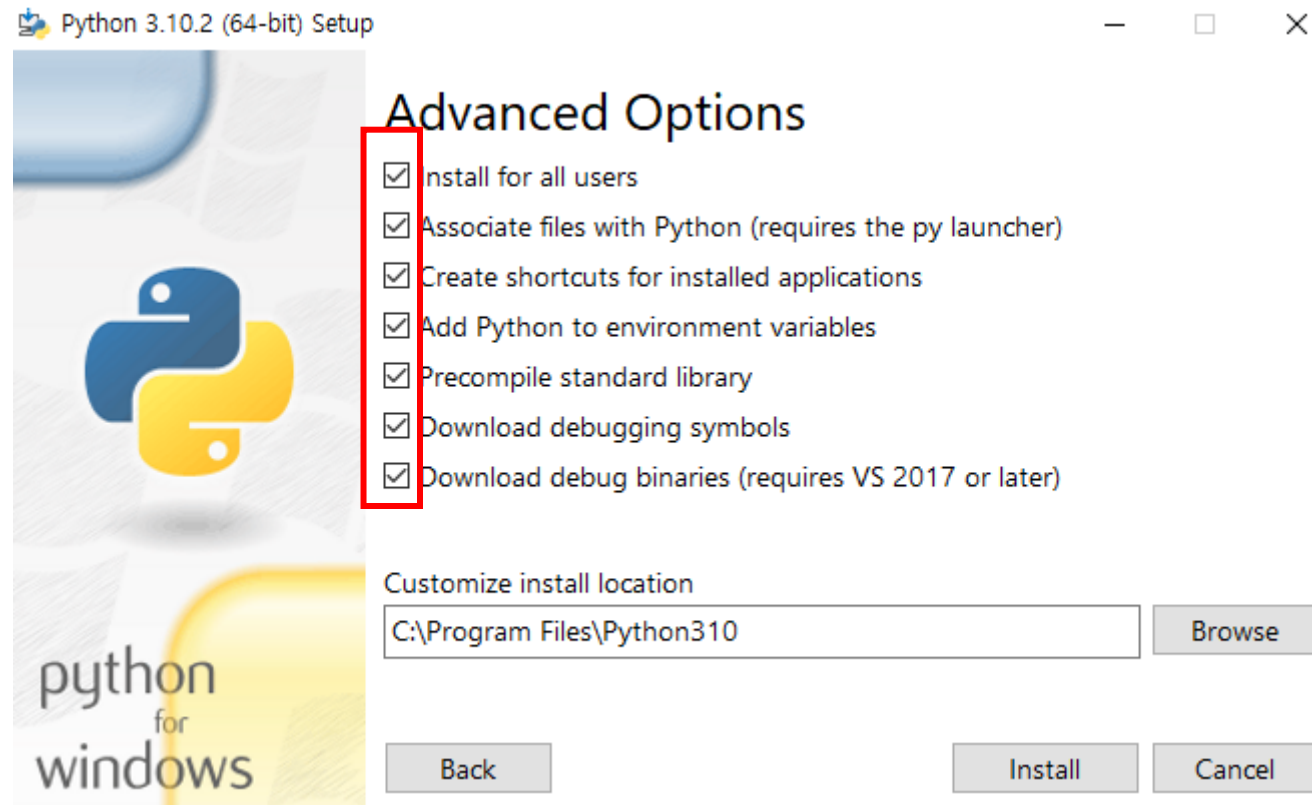
1 2 3 4 5

Python is a programming language that lets you work quickly and integrate systems more effectively. [>>> Learn More](#)

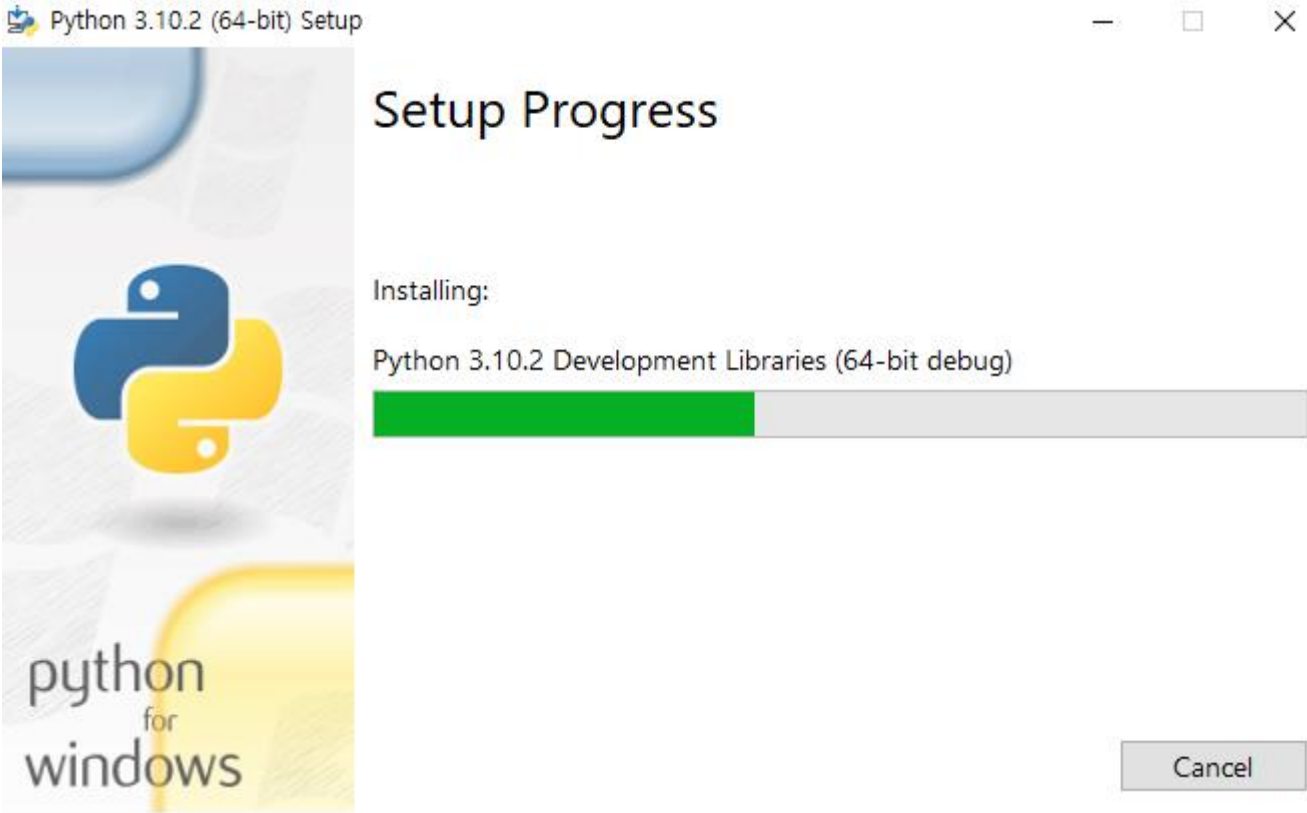
# Python 설치하기



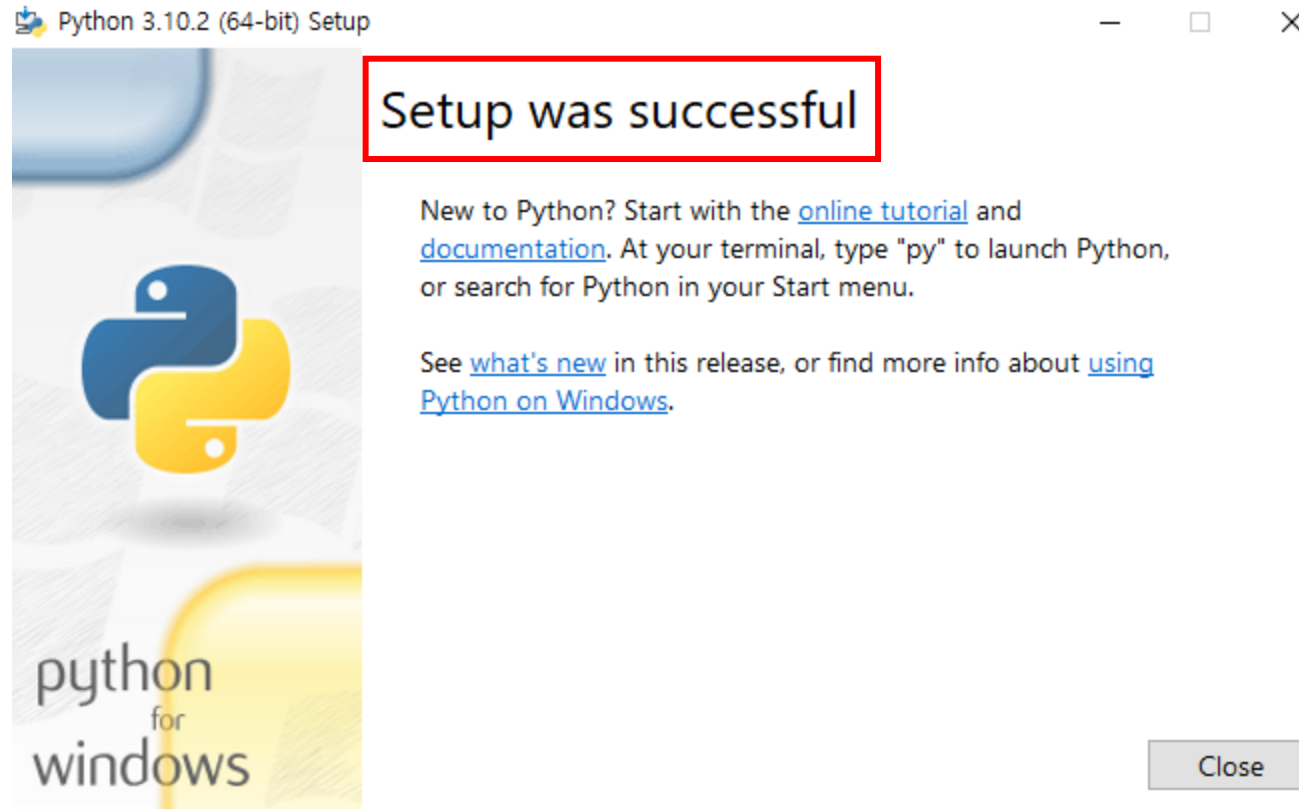
# Python 설치하기



# Python 설치하기



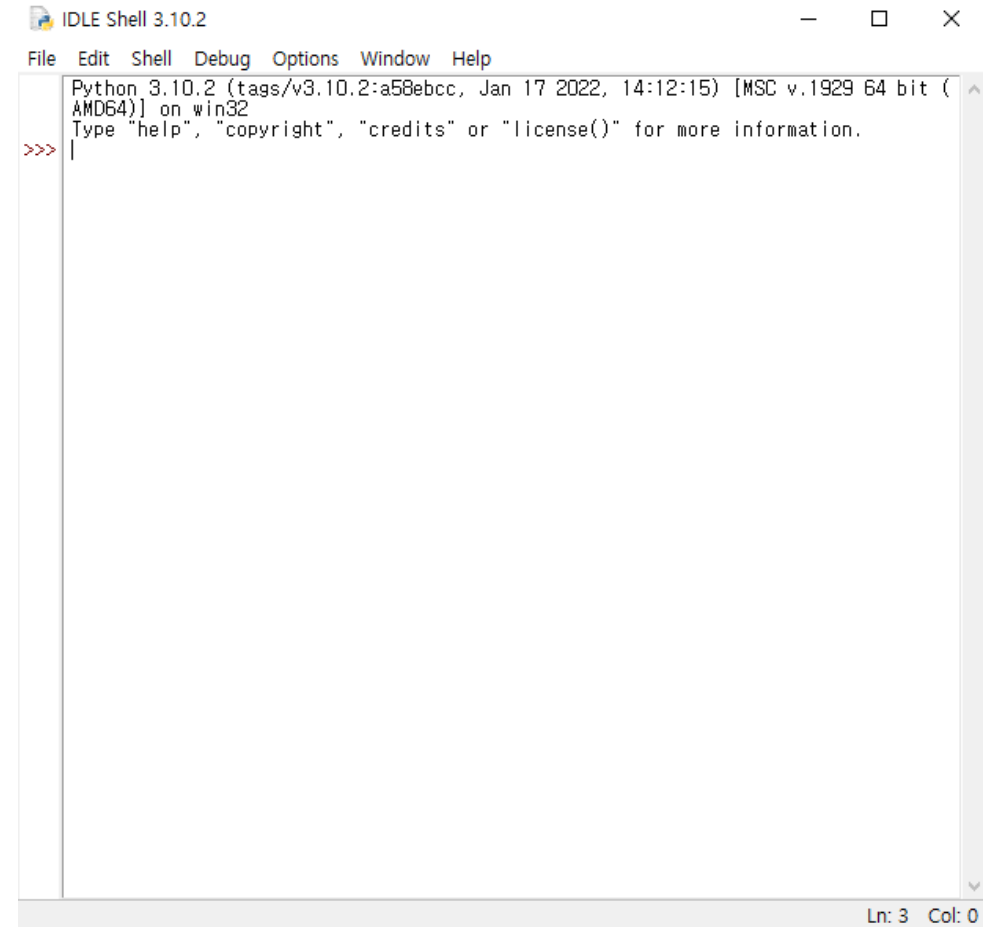
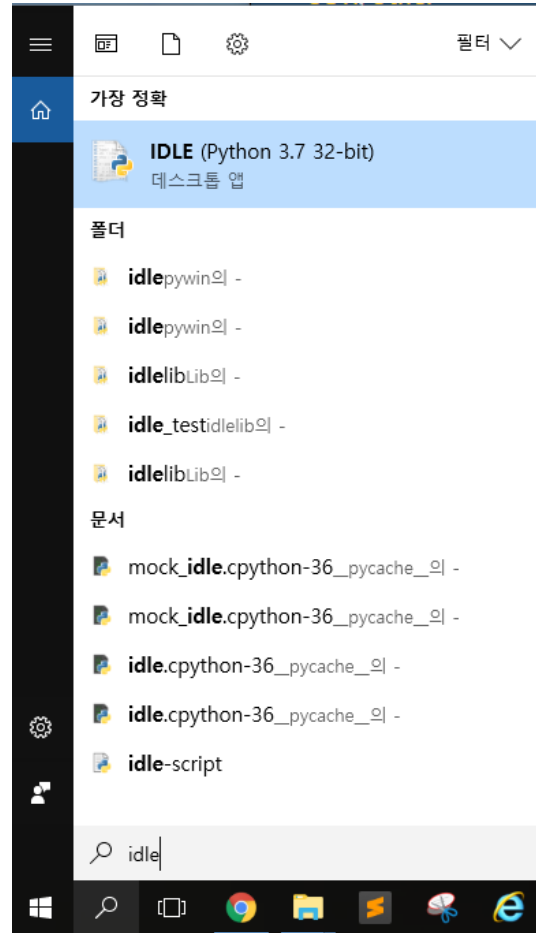
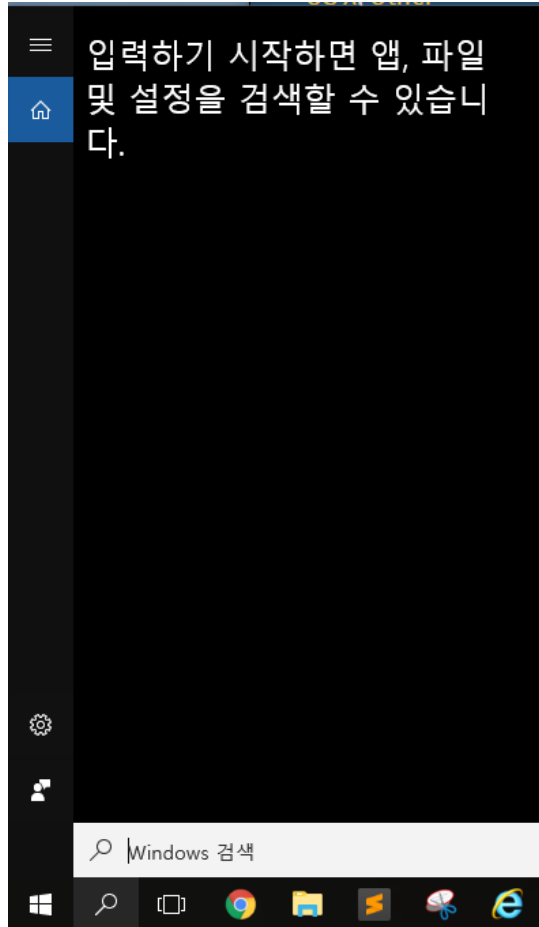
# Python 설치하기





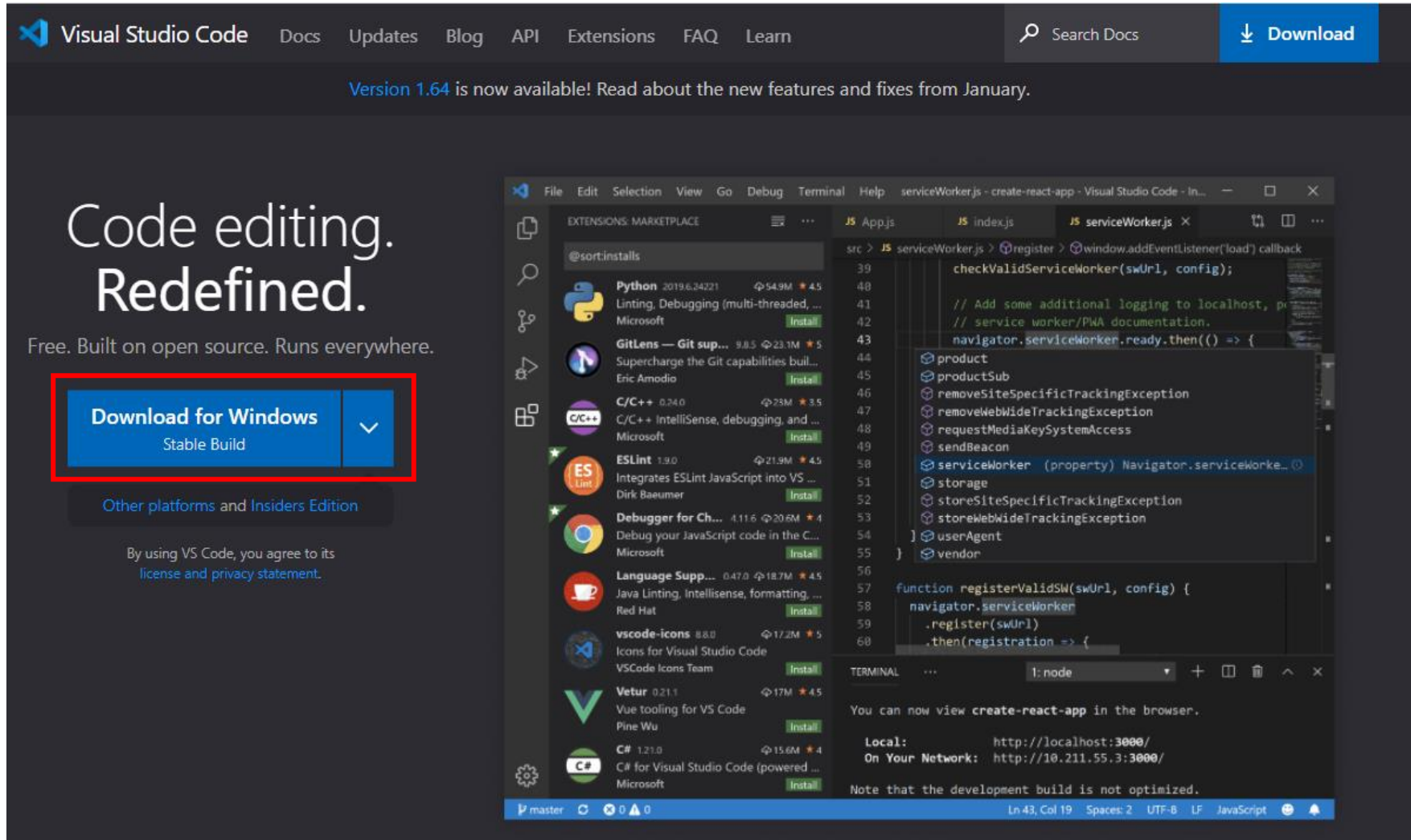
# Python 설치하기

- Python IDLE

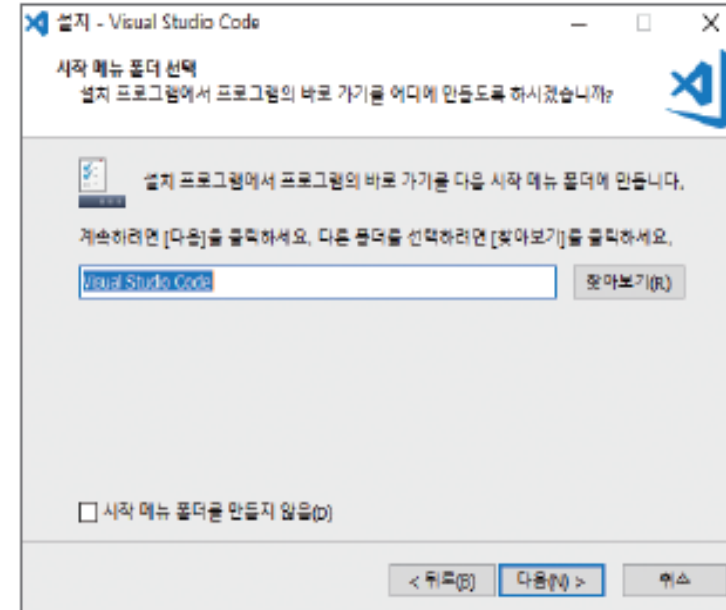
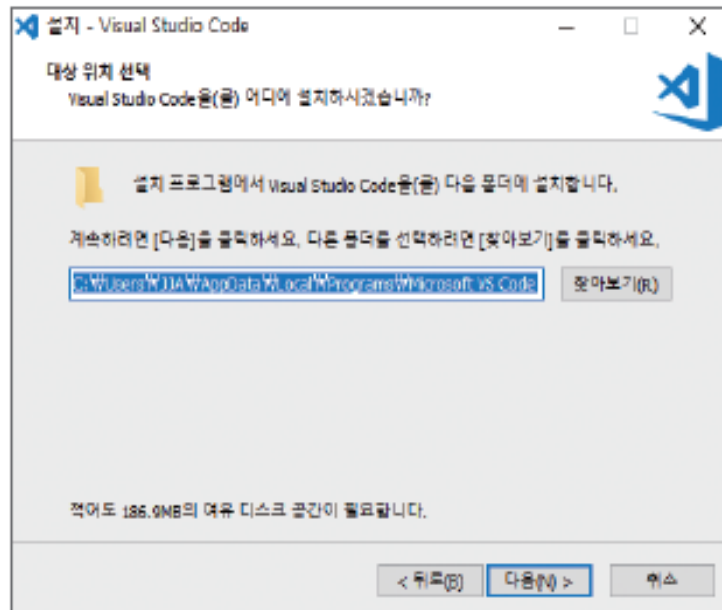
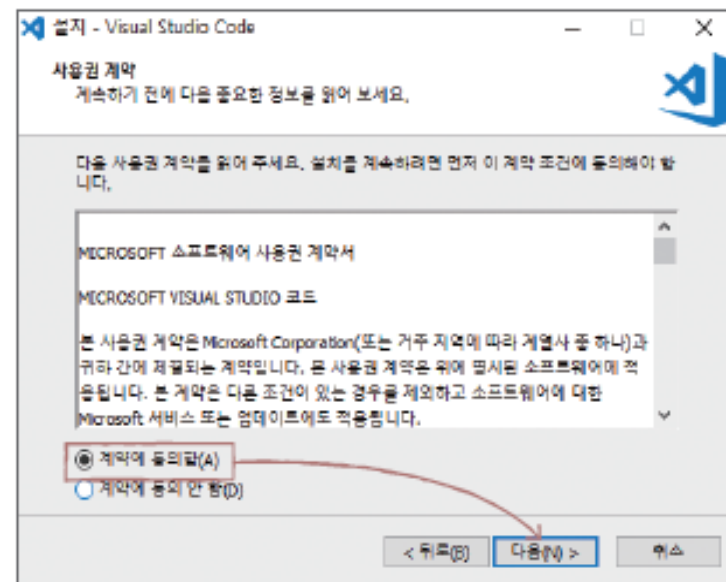
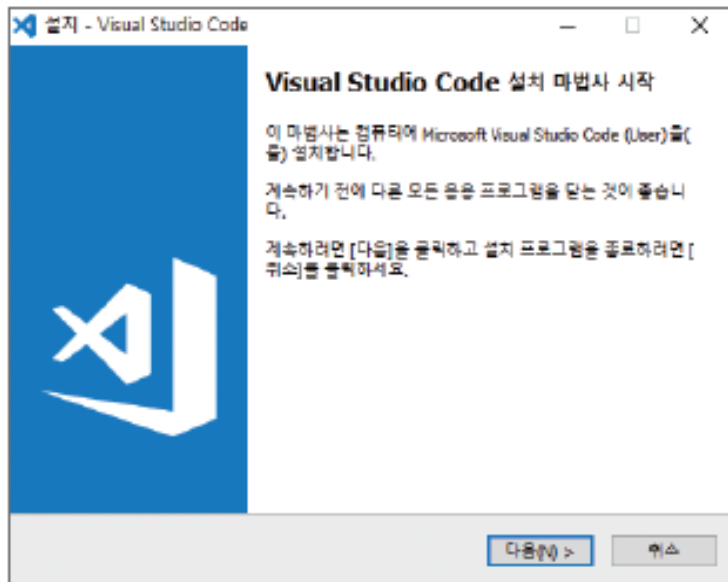


# 텍스트 에디터 사용하기 : 비주얼 스튜디오 코드

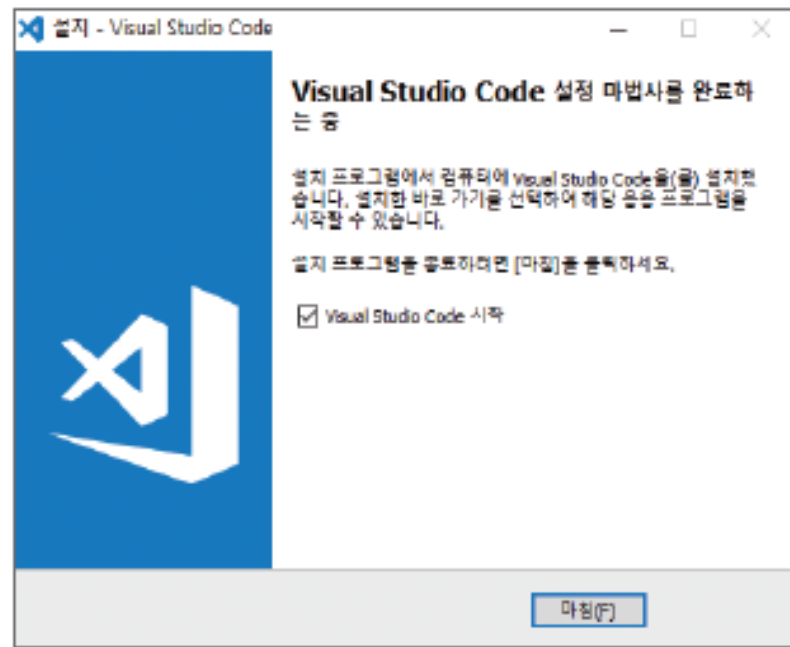
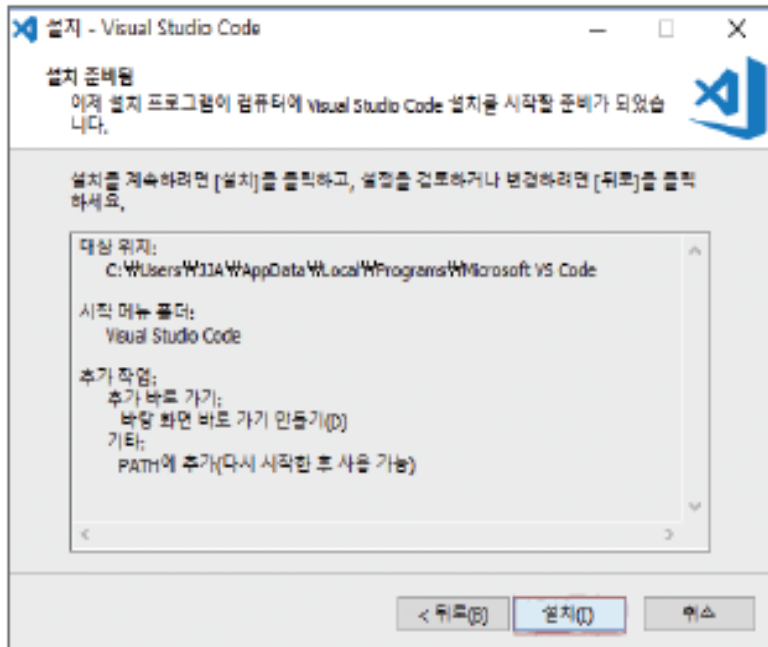
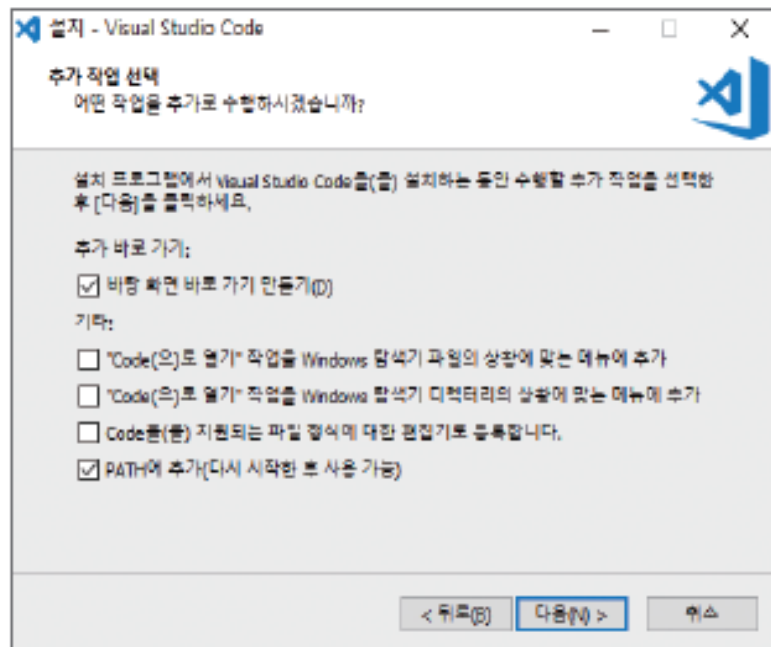
- Visual studio code 설치하기: <https://code.visualstudio.com>
  - [Downloads for Windows] 클릭하여 설치 파일 다운로드



# 텍스트 에디터 사용하기 : 비주얼 스튜디오 코드

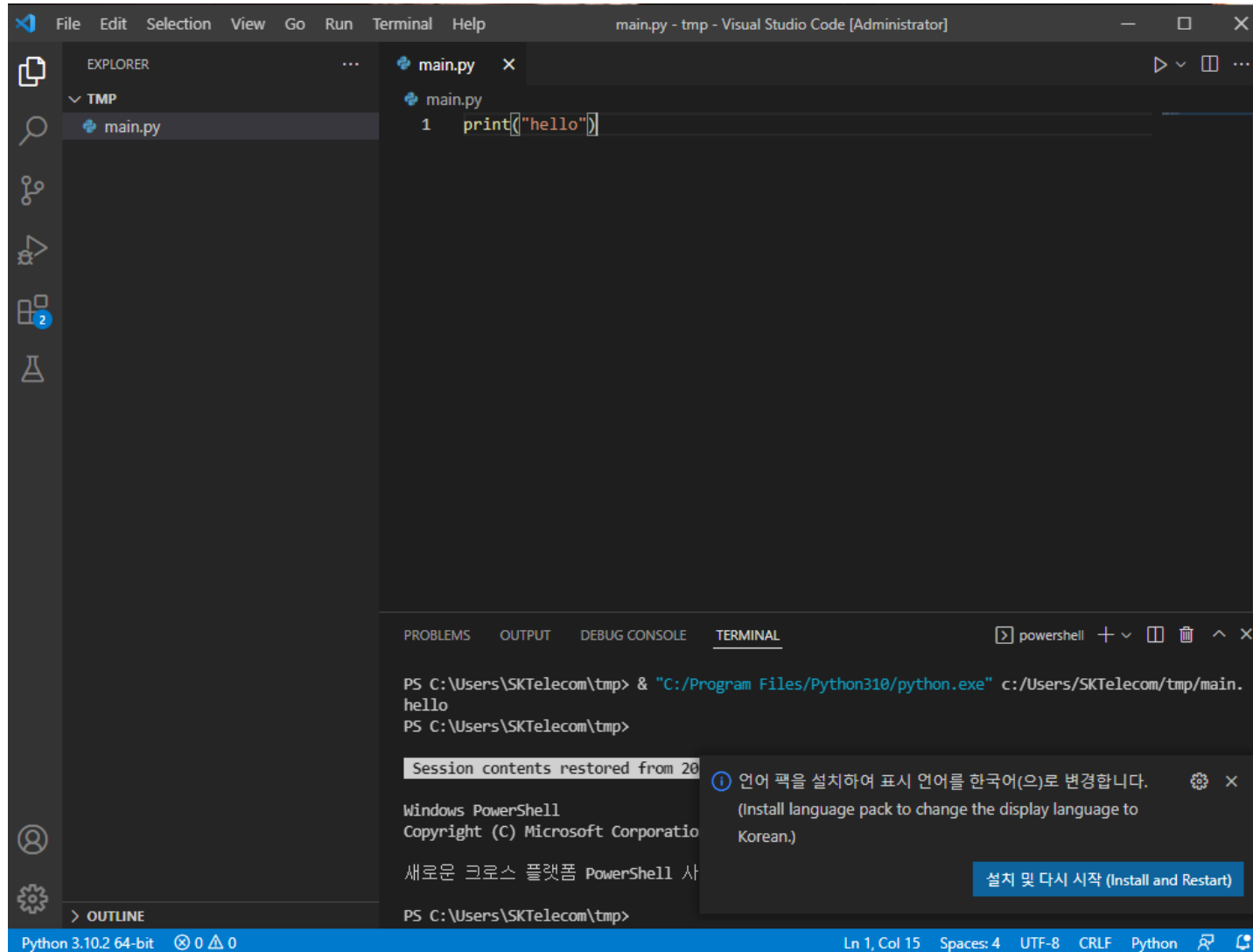


# 텍스트 에디터 사용하기 : 비주얼 스튜디오 코드



# 텍스트 에디터 사용하기 : 비주얼 스튜디오 코드

- 비주얼 스튜디오 코드에서 코드 작성하고 실행하기
  - 시작 화면에서 [파일] - [새 파일] 메뉴 선택 (단축키 [Ctrl] + [N])



## 텍스트 에디터 사용하기 : 비주얼 스튜디오 코드

- [확장] 메뉴에서 [Python] 클릭 – [설치] – [다시 로드]





## 강의 및 교재 소개

---

- 중간 고사 이전 [실습]
  - 파이썬 기본 문법 복습
  - 외부 라이브러리 (numpy, pandas, matplotlib, scikit-learn)
  - 중간고사 : coding test
- 중간 고사 이후 [이론 + 실습]
  - 데이터 마이닝 이론 및 실습
  - 지도학습, 비지도학습 등 이론과 파이썬 간단한 코드 구현
  - 기말고사 : coding test ? paper test? (not yet)

# 강의 소개

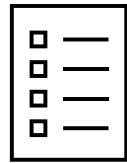
---

- 수업 진행
  - 대면 원칙으로 수업을 진행하며, 학교 규정상 변경될 가능성 있음
- 평가



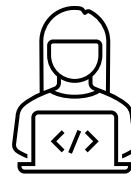
중간고사

35%



기말고사

35%



과제

20%



출석

10%

- Grading
  - A+ 40%
  - B+ 50%
  - C+ 10%
  - F : 중간/기말 불참 또는 3회 결석 (2회 지각 = 1회 결석)
- 수업 진도 및 평가 방법은 불가피한 상황에 의해 변동될 수 있음

## 강의 소개

주차	강의 내용
1	오리엔테이션 / 파이썬 설치
2	if & for
3	function & file
4	numpy
5	특강
6	pandas
7	visualization
8	중간고사
9	데이터 전처리
10	지도학습
11	준지도학습
12	특강
13	비지도학습
14	인공지능
15	기말고사

과제 #1

과제 #2

감사합니다

[kimtwan21@dongduk.ac.kr](mailto:kimtwan21@dongduk.ac.kr)

김 태 완