



# 1. Python 소개

Python을 활용한 분석  
기초 가이드 북





## I . Python 소개

1. Python 이란?
2. Anaconda 소개 및 설치
3. Spyder 소개
4. 사용 패키지 소개
5. 패키지 설치

# 1. Python 이란?

Python(파이썬)은 1991년 프로그래머인 귀도 반 로섬(Guido van Rossum)이 발표한 고급 프로그래밍 언어로, 플랫폼 독립적이며 인터프리터식, 객체지향적, 동적 타이핑(dynamically typed) 대화형 언어입니다.

## ● 언어의 기능

- Python은 언어 자신의 기능은 작게 하여 사용자가 언제나 필요로 하는 최소한의 기능만을 제공하도록 만들어져 있습니다. 이것은 펄의 TIMTOWTDI(there ' s more than one way to do it - ' 문제를 해결하는 방법은 단 한 가지가 아니다 ' ) 철학과는 대조적인 것이며, Python에서는 다른 사용자가 썼더라도 동일한 일을 하는 프로그램은 대체로 모두 비슷한 코드로 수렴합니다.
- 또한, Python에서는 프로그램의 문서화가 매우 중시되고 있어 언어의 기본 기능에 포함되어 있습니다. Python은 원래 교육용으로 설계되었기 때문에 읽기 쉽고, 그래서 효율적인 코드를 되도록 간단하게 쓸 수 있도록 하려는 철학이 구석 구석까지 침투해 있어, Python 커뮤니티에서도 알기 쉬운 코드를 선호하는 경향이 강합니다.



## ● 주요 특징

- 동적 타이핑(dynamic typing)
  - 실행 시간에 자료형을 검사합니다.
- 객체의 멤버에 무제한으로 접근 가능
  - 속성이나 전용의 메서드 혹은 만들어 제한할 수도 있습니다.
- 모듈, 클래스, 객체와 같은 언어의 요소가 내부에서 접근할 수 있고, 리플렉션을 이용한 기술을 쓸 수 있습니다.

## 2. Anaconda 소개 및 설치 (1/3)

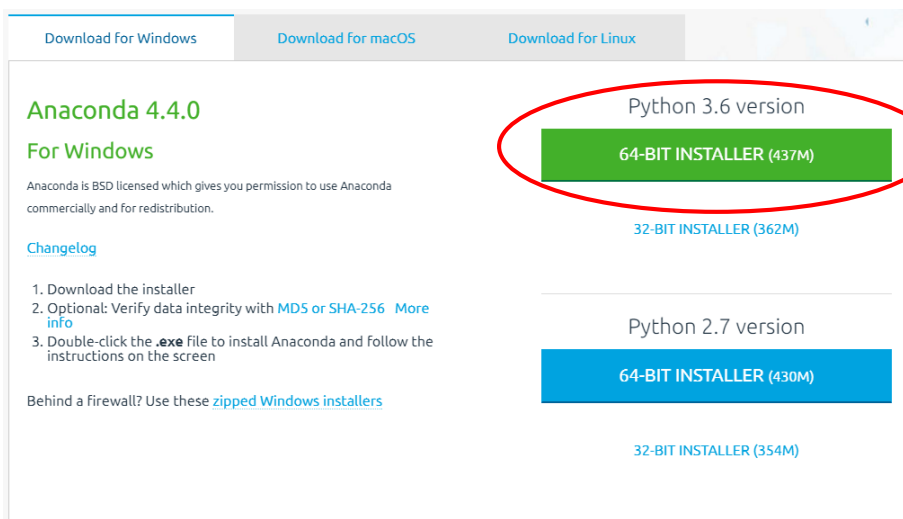
Python을 사용하기 위해 다양한 설치 및 환경설정 방법이 존재하지만, 본 문서에서는 아직 프로그래밍 환경에 익숙하지 않은 독자를 대상으로 합니다. 최근에는 인공지능이나 데이터 분석을 위해 Python을 많이 사용하는데, 주로 Anaconda(아나콘다)를 베이스로 시작합니다.

### ● Anaconda

- Anaconda(아나콘다)는 설치 소프트웨어 가운데 하나로 래드햇에서 만들었으며, 콘솔판과 GUI판을 함께 제공하기 때문에 컴퓨터 환경에 대해 유연하는 이점이 있습니다.
- 또한, Python을 포함해 수학, 과학, 시각화 등에 필요한 거의 모든 패키지들(Numpy, IPython, Matplotlib 등)이 한 번에 설치되기 때문에, 익숙하지 않은 초보자들의 경우 설치의 번거로움이나 각 패키지의 의존성 문제를 줄일 수 있어 추천합니다.

### ● 다운로드

- 홈페이지의 다운로드 사이트(<https://www.continuum.io/downloads>)에서 아래 그림과 같은 녹색 버튼(Python 3.6 version, 64bit)을 눌러 설치파일을 다운받습니다.
- Anaconda를 설치하기 전에 별도로 Python을 설치하지 않도록 하며, Windows 이외의 OS나 32bit 컴퓨터 사용자의 경우에는 그에 맞는 설치 파일을 다운받습니다.



## 2. Anaconda 소개 및 설치 (2/3)

### ● 설치

- 설치 전, 약 2GB 이상의 디스크 여유 공간을 마련합니다.
- 다운받은 설치 파일을 실행시켜 'Next', 'I Agree' 버튼을 눌러 설치를 진행합니다.
- Install Location은 본인이 원하는 위치로 설정 가능합니다.
- Advanced Options의 항목은 체크하지 않고 'Install' 버튼을 눌러 설치를 시작합니다.
- 모든 설치가 완료되면 'Finish' 버튼을 눌러 창을 닫습니다.



## 2. Anaconda 소개 및 설치 (3/3)

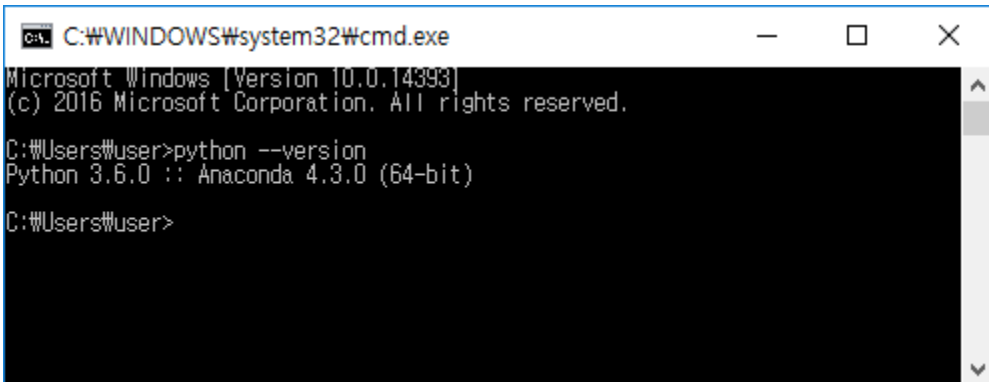
### ● 확인

- 좌측 하단 윈도우창 모양을 마우스 우클릭하여(혹은, 윈도우키 + R을 눌러 cmd를 입력 후 엔터하여) 명령 프롬프트 창을 열고, 아래의 명령어를 입력해 정상적으로 설치가 됐는지 확인합니다.

*# 명령 프롬프트 창에서 Python과 Anaconda의 버전을 확인.*

```
python --version
```

- 설치가 정상적으로 진행됐다면, 아래와 같이 Python과 Anaconda의 버전을 출력합니다.
- 2017.07 기준, Python 3.6.0 :: Anaconda 4.3.0 (64-bit)



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\User>python --version
Python 3.6.0 :: Anaconda 4.3.0 (64-bit)

C:\Users\User>
```

## 3. Spyder 소개

Spyder(스파이더, Scientific PYthon Development EnviRonment)는 Anaconda 설치 시, 함께 설치되는 프로그램들 중 하나로 Python을 프로그래밍 하는데 필요한 도구들을 제공하는 무료 통합개발환경(IDE)입니다.

PyCharm, VIM, VING IDE, Eclipse 등 다양한 개발환경이 있지만, 본 문서에서는 Spyder를 이용해 설명과 실습을 진행합니다.

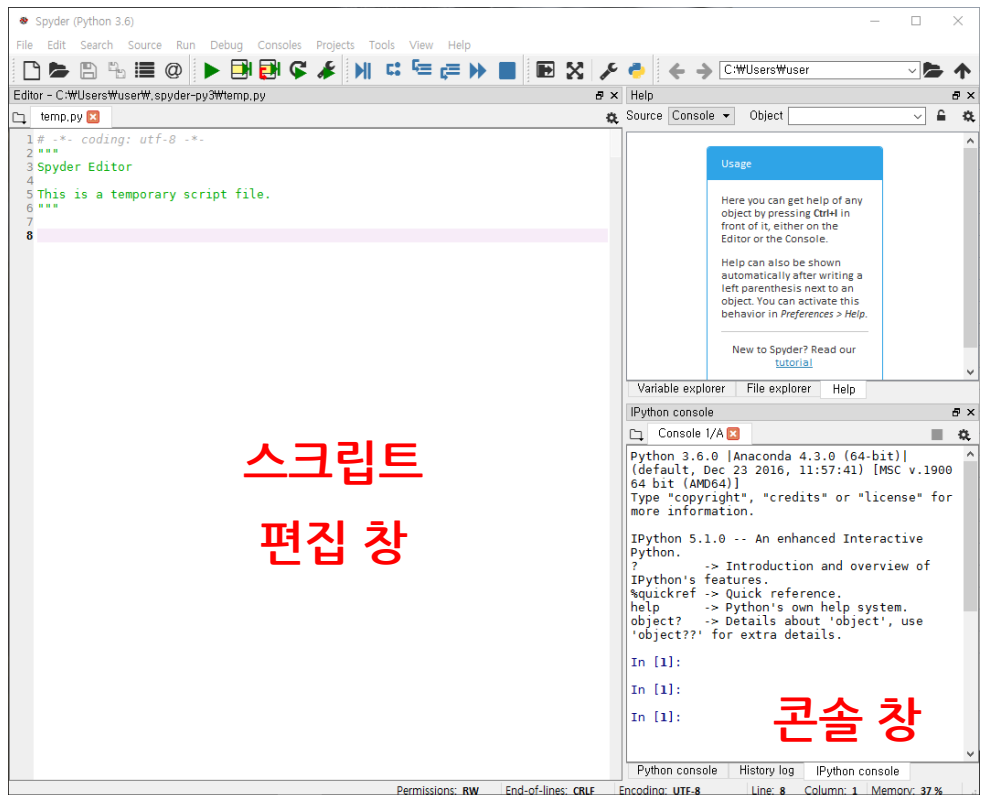
### ● Spyder 실행

- 바탕화면 혹은 시작메뉴에서 아래와 같은 모양의 단축아이콘을 더블 클릭해 Spyder를 실행시킵니다.



◀ 단축아이콘

▼ Spyder 실행 화면



## 4. 사용 패키지 소개

---

Python을 기반으로 분석 실습을 진행하기 위해 사용하는 패키지들을 먼저 살펴봅니다.

### ● NumPy

- NumPy는 조밀한 데이터 버퍼에서 저장하고 처리하는 효과적인 인터페이스를 제공합니다. NumPy 배열은 파이썬의 내장 타입인 list와 비슷하지만 배열의 규모가 커질수록 데이터 저장 및 처리에 효율적으로 기능합니다. 여러 데이터 분석 패키지가 기능하기 위해선 선행적으로 설치되어 있어야 하는 기본 분석 패키지이며, Anaconda 스택을 설치했을 경우 기본적으로 Numpy가 설치되어 있습니다.

### ● Pandas

- Pandas는 데이터 분석에서 가장 자주 쓰이는 패키지로서 일련의 객체를 담을 수 있는 1차원 배열(Series)과 행과 열 레이블이 부착된 다차원 배열(DataFrame)을 자료구조로 제공합니다. 그리고 관계형 데이터베이스 연산, 시계열 기능 등의 데이터 분석 도구를 포함하고 있습니다. Pandas는 NumPy를 기반으로 만들어졌기 때문에 Pandas 설치 전 NumPy가 설치되어 있어야 합니다.

### ● Matplotlib

- Matplotlib는 데이터 시각화 기본 패키지로서 Anaconda 내부에서 기본적으로 제공되고 있습니다. 시계열, 지형도, 3D 도표 등의 상위 도식을 제공합니다. NumPy, Pandas 등 여러 패키지와 호환이 가능하지만, 영문 기반 library 이기에 한글 폰트 사용이 필요합니다.

### ● StatsModels

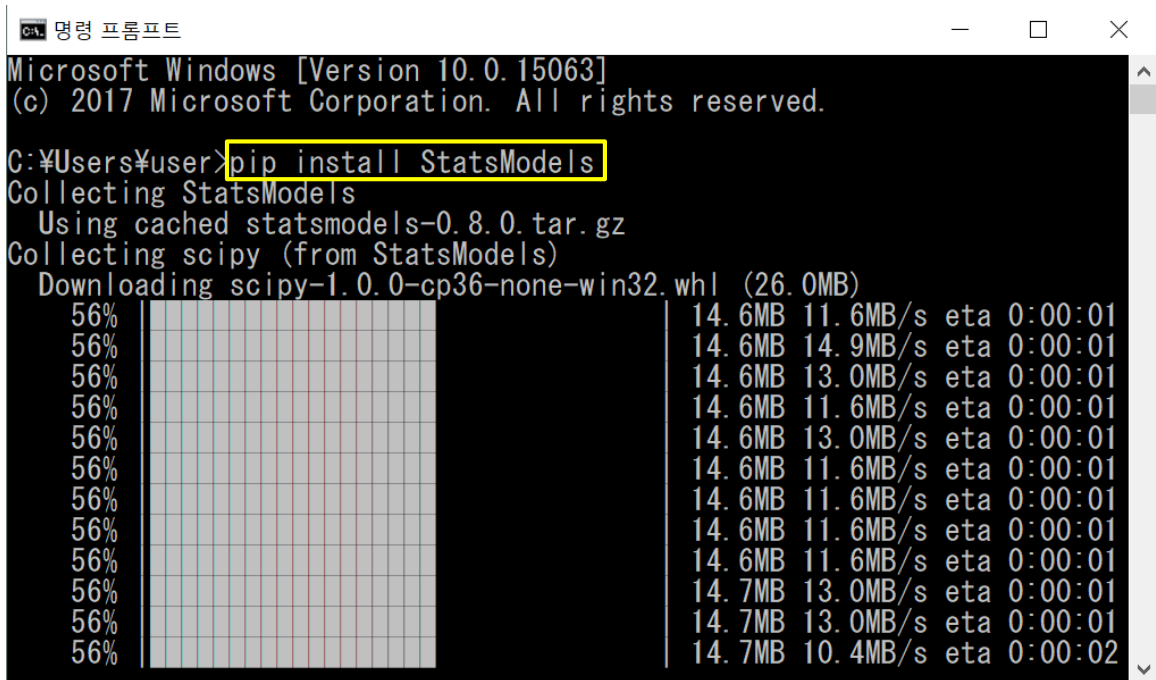
- StatsModels는 기초 통계, 상관 분석, 회귀 분석 등 다양한 통계 분석 기능을 제공하는 패키지입니다. 기초 통계의 t검정,  $\chi^2$ 검정, F검정 등 각종 검정, 커널 밀도 추정 등의 데이터 분석 시 필요한 통계 분석을 수행할 수 있습니다. 그 외에 분산 분석(ANOVA), 회귀 분석, 로지스틱 회귀 분석, 시계열 분석 등의 통계 모형을 만들 수 있습니다.



## 5. 패키지 설치 (1/2)

명령 프롬프트(cmd)창에서 다음 과정을 통해 설치할 수 있습니다.

- 명령 프롬프트(cmd) 창을 통한 설치

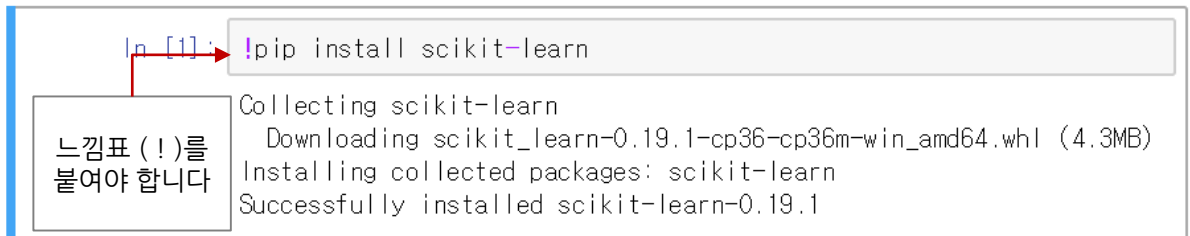


```

Microsoft Windows [Version 10.0.15063]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\User>pip install StatsModels
Collecting StatsModels
  Using cached statsmodels-0.8.0.tar.gz
Collecting scipy (from StatsModels)
  Downloading scipy-1.0.0-cp36-none-win32.whl (26.0MB)
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 14.9MB/s eta 0:00:01
  56% |#####| 14.6MB 13.0MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 13.0MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.6MB 11.6MB/s eta 0:00:01
  56% |#####| 14.7MB 13.0MB/s eta 0:00:01
  56% |#####| 14.7MB 13.0MB/s eta 0:00:01
  56% |#####| 14.7MB 10.4MB/s eta 0:00:02
  
```

- Jupyter을 통한 설치



```

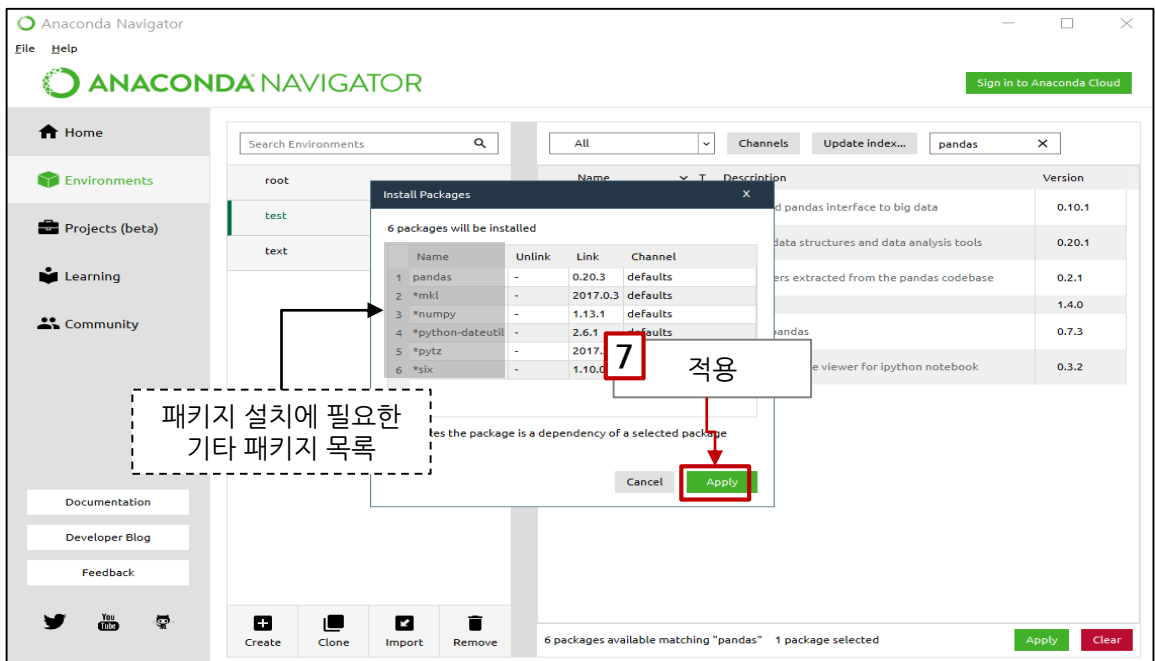
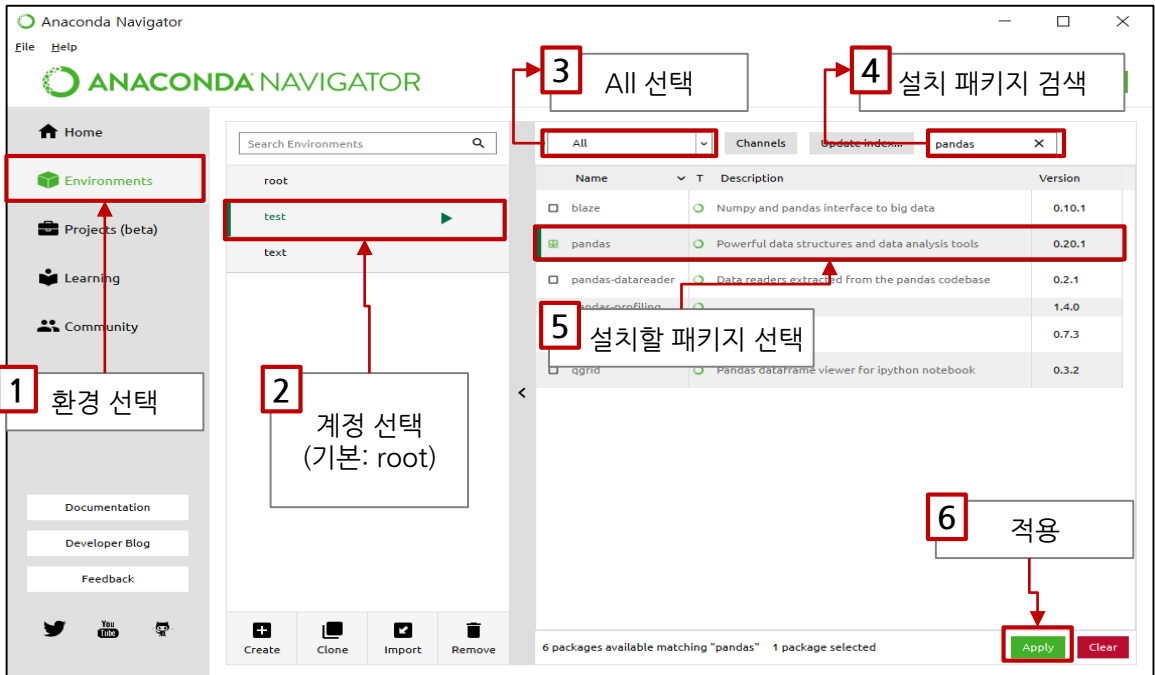
In [1]: !pip install scikit-learn
Collecting scikit-learn
  Downloading scikit_learn-0.19.1-cp36-cp36m-win_amd64.whl (4.3MB)
Installing collected packages: scikit-learn
Successfully installed scikit-learn-0.19.1
  
```

느낌표 (!)를  
붙여야 합니다

## 5. 패키지 설치 (2/2)

ANACONDA에서 다음 과정을 통해 설치할 수 있습니다.

### ● ANACONDA NAVIGATOR을 통한 설치





본 문서의 내용은 기상청의 기상기후 빅데이터 분석(<http://bd.kma.go.kr>)의  
분석 플랫폼 활용을 위한 Python 분석 기초 교육 자료입니다.