



머신러닝을 활용한 빅데이터 분석

강사 - 하예지



BIG DATA

[illegible]

금융상품 선택의 가장 명쾌한 기준
뱅크샐러드

강의 목적

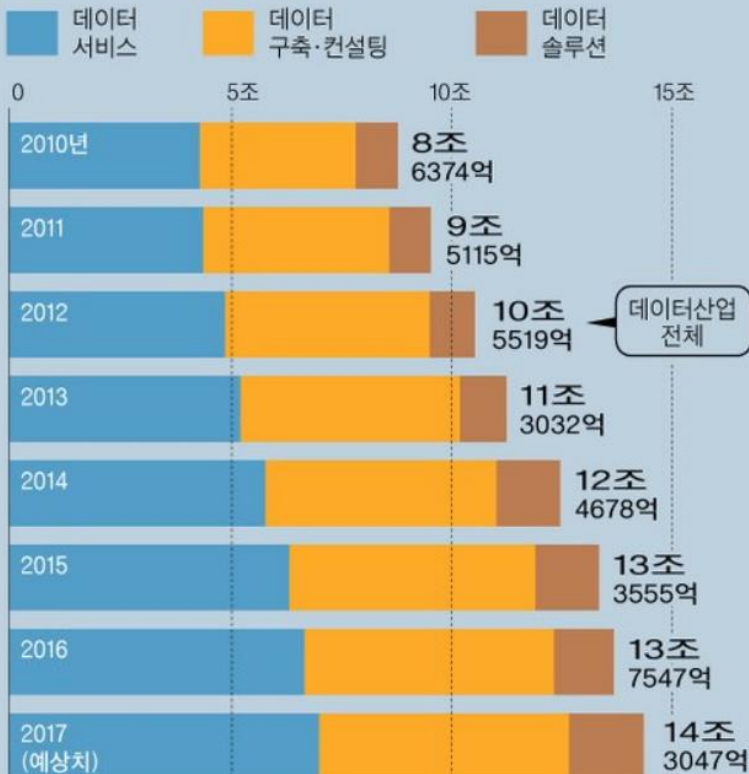
데이터 사이언티스트 전성시대

데이터가 핵심, 데이터 전문가들의 역할이 중요

삼성·SK 등 인재 영입 전쟁

IT업체 뿐 제조·의료 등 다양한 산업에서 활약

국내 데이터산업 시장 규모 단위:원



국내 데이터 직무별 인력 부족률

단위:% ※2017년 기준



강의 목적



- 분석 기법에 대한 전문성 및 기술성
- 빅데이터 처리 기술 관련 스킬
- IT산업 관련 지식

- 대인 관계, 팀워크, 협업, 리더십
- 의사소통, 인간관계 형성, 협업 능력
- 빅데이터의 가치를 발견하는 통찰력
- 논리적인 설득력, 전달력

기술적인 지식 + 데이터 관찰능력 = 인사이트 발굴

커리큘럼 – Part 1

(1주 ~ 4주)

- 파이썬 기초
 - 데이터타입
 - 기본 자료형(리스트, 딕셔너리, 셋, 튜플)
 - 인덱싱/슬라이싱
 - 제어문(조건문, 반복문)
 - 함수



커리큘럼 – Part 1

(1주 ~ 4주)

- Numpy
 - 배열 생성하기
 - 속성을 사용한 배열 활용
 - 배열 처리하기(복사/수정/추가/삭제/결합/분리)
 - 배열 연산(산술연산, 비교연산, 집계함수)
 - 인덱싱/슬라이싱
 - 정렬
 - 데이터 타입



커리큘럼 – Part 1

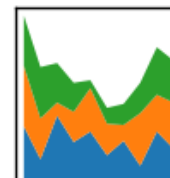
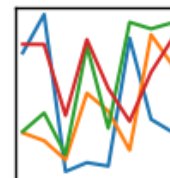
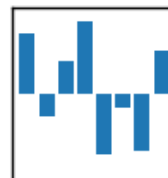
(1주 ~ 4주)

- Pandas

- 데이터 타입
- 외부 데이터 사용하기(excel, csv)
- pandas 연산(산술연산)
- 계층 색인
- 데이터 처리하기(수정/추가/삭제/결합/분리)
- 집계(pivot, groupby)
- 통계
- 정렬

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



커리큘럼 – Part 1

(1주 ~ 4주)

- Matplotlib
 - 라인 그래프
 - 원형 차트
 - 막대 그래프
 - 히스토그램
 - 박스 차트
 - 산포도
 - 히트맵

matplotlib

커리큘럼 – Part 2

(5주 ~ 8주)

- 데이터 분석 입문
 - 데이터 이해
 - 데이터 전처리
 - EDA
 - 모델 선정
 - 학습/평가 데이터 분리
 - 분석 수행
 - 분석 결과 검증



커리큘럼 – Part 2

(5주 ~ 8주)

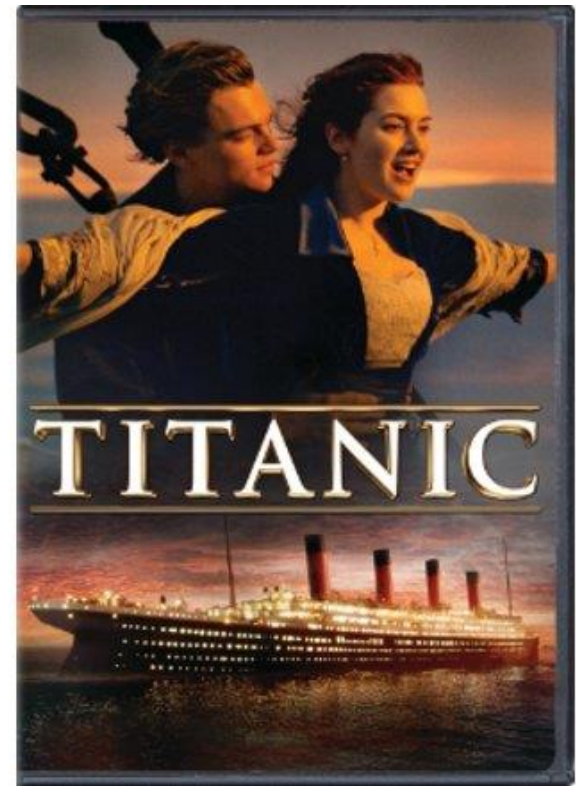
- Scikit-learn 라이브러리 기반 머신러닝
 - 사용데이터 : sklearn datasets(IRIS, wine-quality, boston house price 등)
 - 지도 학습 모델
 - 비지도 학습 모델
 - 학습/평가 데이터 다루기
 - 모델 검증 방법



커리큘럼 – Part 2

(5주 ~ 8주)

- 머신러닝을 활용한 데이터 분석 실습
 - 사용데이터 : 타이타닉 데이터
 - 데이터 전처리
 - EDA(데이터 시각화)
 - 생존 예측 분석
 - 모델 검증



커리큘럼 – Part 2

(5주 ~ 8주)

- 머신러닝을 활용한 데이터 분석 최종 프로젝트
 - 실제 보험사의 고객 데이터를 바탕으로 보험 사기자를 예측하는 머신러닝 분석을 수행
 - 분석 프로세스 산출물 기반의 포트폴리오 콘텐츠 작성





고급

• 데이터 분석

심화

• 데이터 처리

기본

• 프로그래밍



자료형
제어문

The image features a light gray background with abstract geometric shapes in the corners. The top-left corner contains several overlapping triangles in shades of blue, green, and red. The bottom-right corner contains several overlapping triangles in shades of gray.

프로그래밍 기초



실습 환경



Anaconda

- 파이썬 배포판으로 데이터 분석에 유용한 오픈소스 패키지들을 모아 놓은 개발 플랫폼
- 데이터 가공 및 분석에 필요한 다양한 패키지 포함
- 패키지 추가설치 및 업데이트도 용이

ex) scikit-learn, matplotlib, numpy, pandas, statsmodel

- 설치

<https://www.anaconda.com/download/>

Anaconda

설치하기

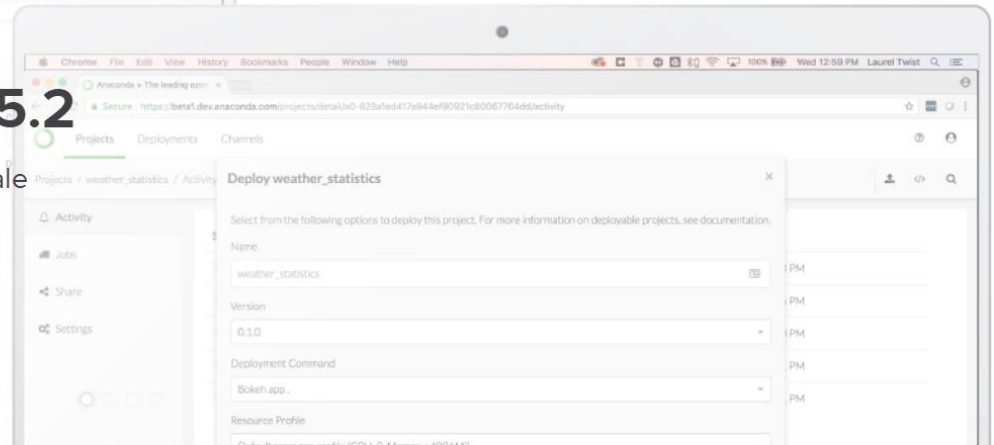


Announcing Anaconda Enterprise 5.2

An AI Enablement Platform for Teams at Scale

[Learn More](#)

Documentation Blog Contact
What is Anaconda? Products Support Resources About [Downloads](#)



Anaconda



[Documentation](#) [Blog](#) [Contact](#) [Q](#)

[What is Anaconda?](#) [Products](#) [Support](#) [Resources](#) [About](#)

[Downloads](#)

High-Performance Distribution


Easily install 1,000+ [data science packages](#)

Package Management

Manage packages, dependencies and environments with [conda](#)

Portal to Data Science

Uncover insights in your data and create interactive visualizations

 **Windows**


 **macOS**

 **Linux**

Anaconda 5.2 For Windows Installer


Python 3.6 version *

 **Download**

[64-Bit Graphical Installer \(631 MB\)](#) 
[32-Bit Graphical Installer \(506 MB\)](#)

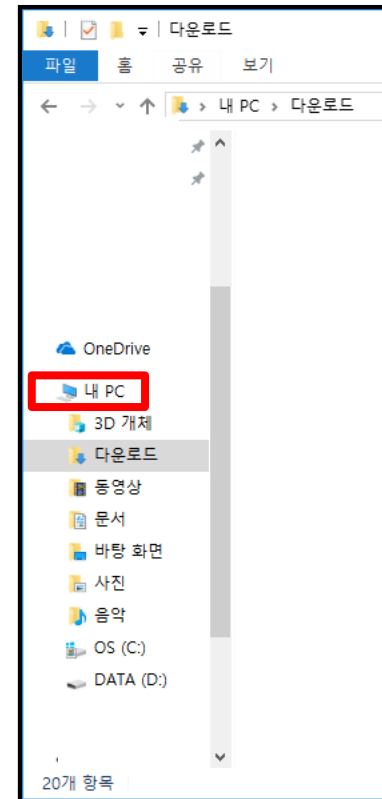
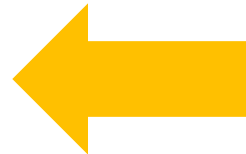
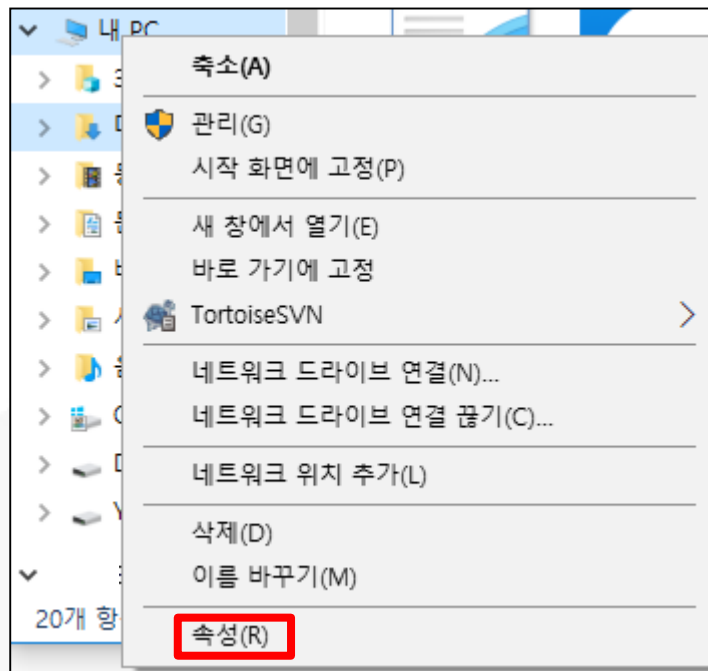
Python 2.7 version *

 **Download**

[64-Bit Graphical Installer \(564 MB\)](#) 
[32-Bit Graphical Installer \(443 MB\)](#)

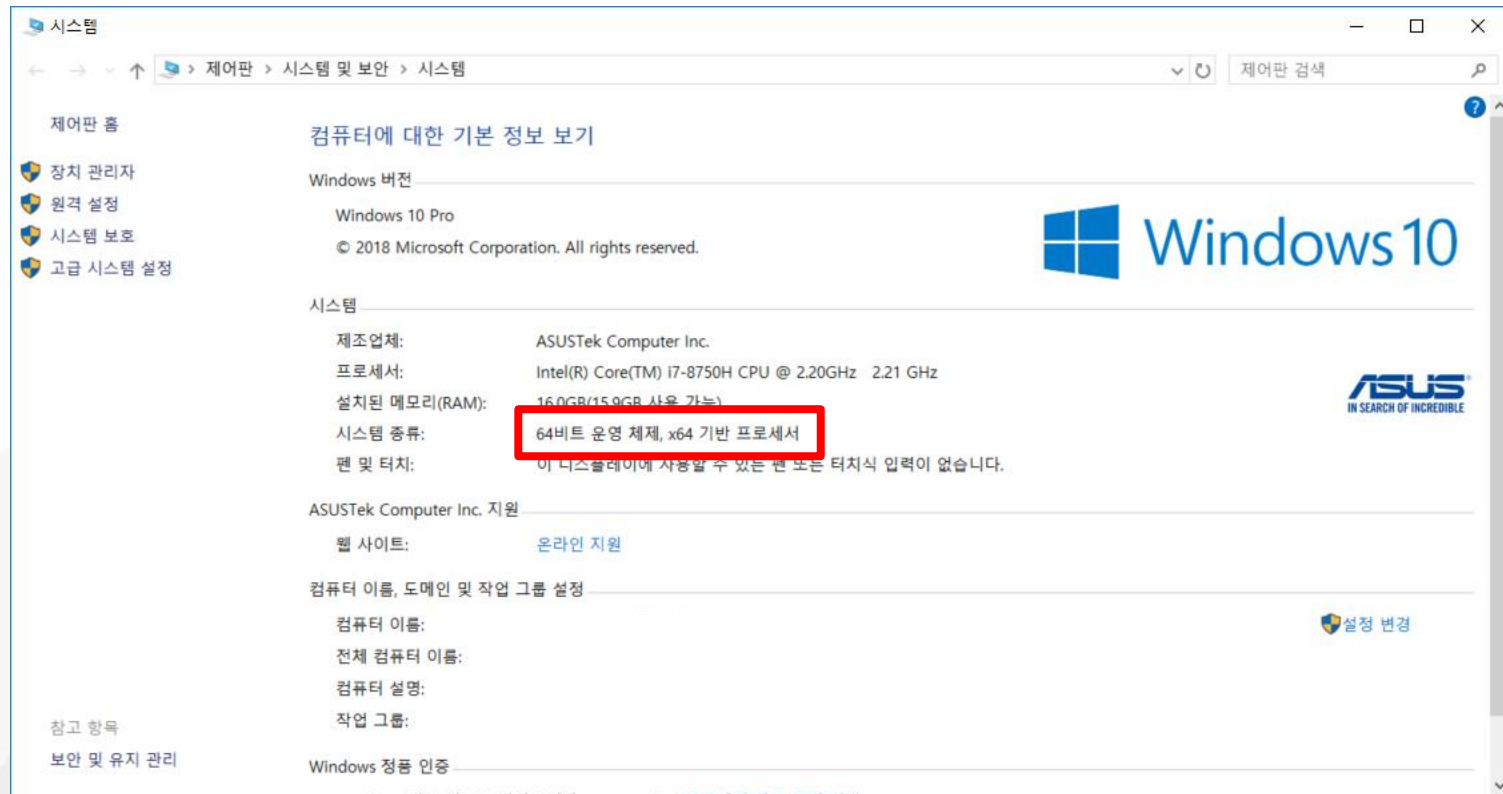
Anaconda

*설치환경 확인하기 > 내PC > 속성 > 시스템

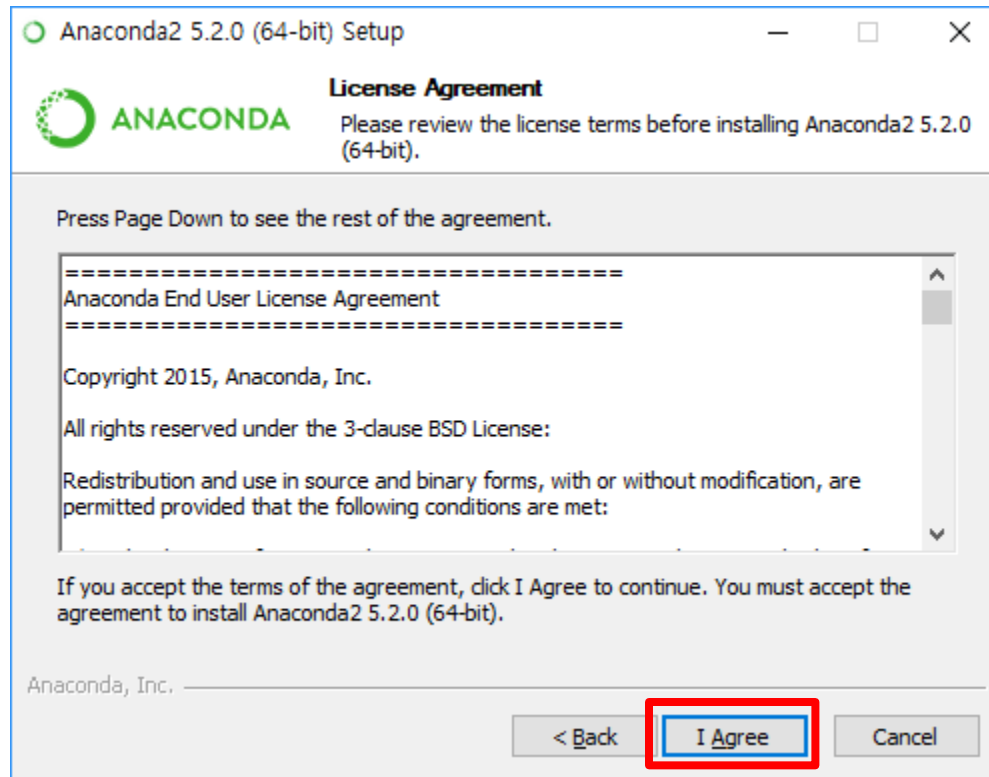


Anaconda

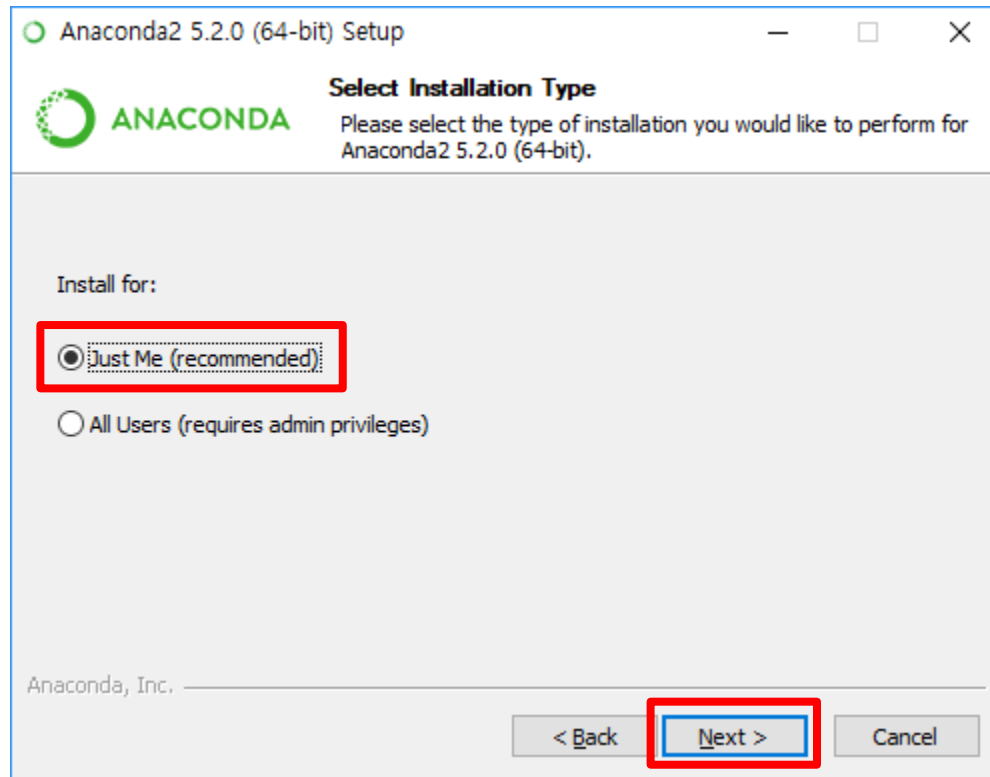
*설치환경 확인하기 > 내PC > 속성 > 시스템



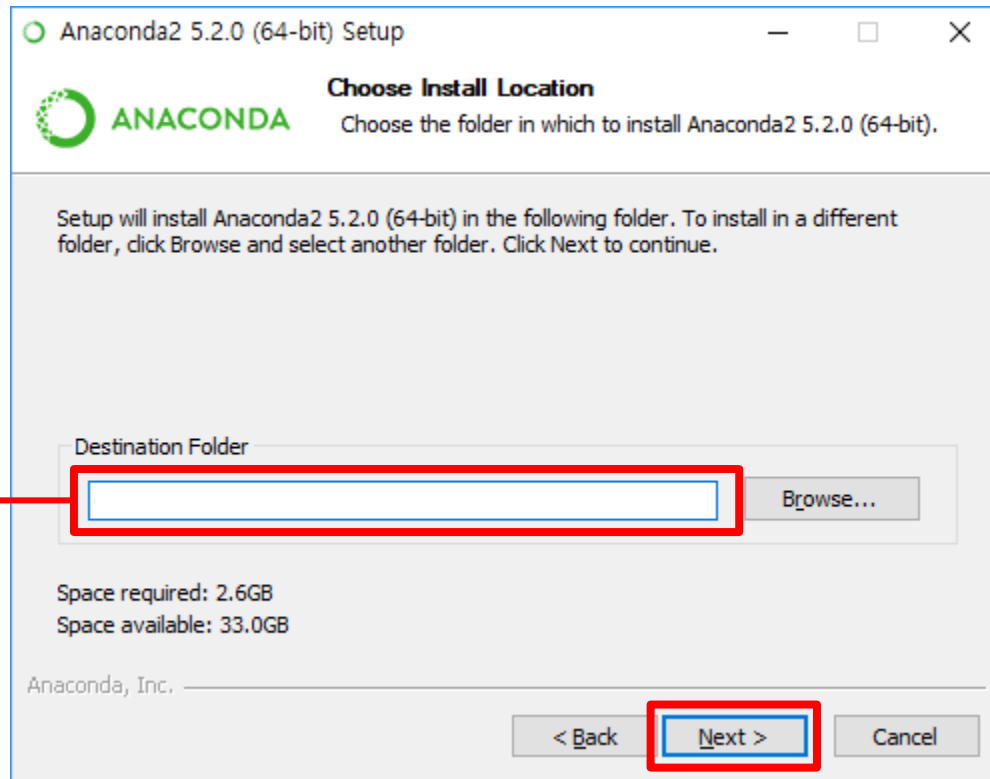
Anaconda



Anaconda

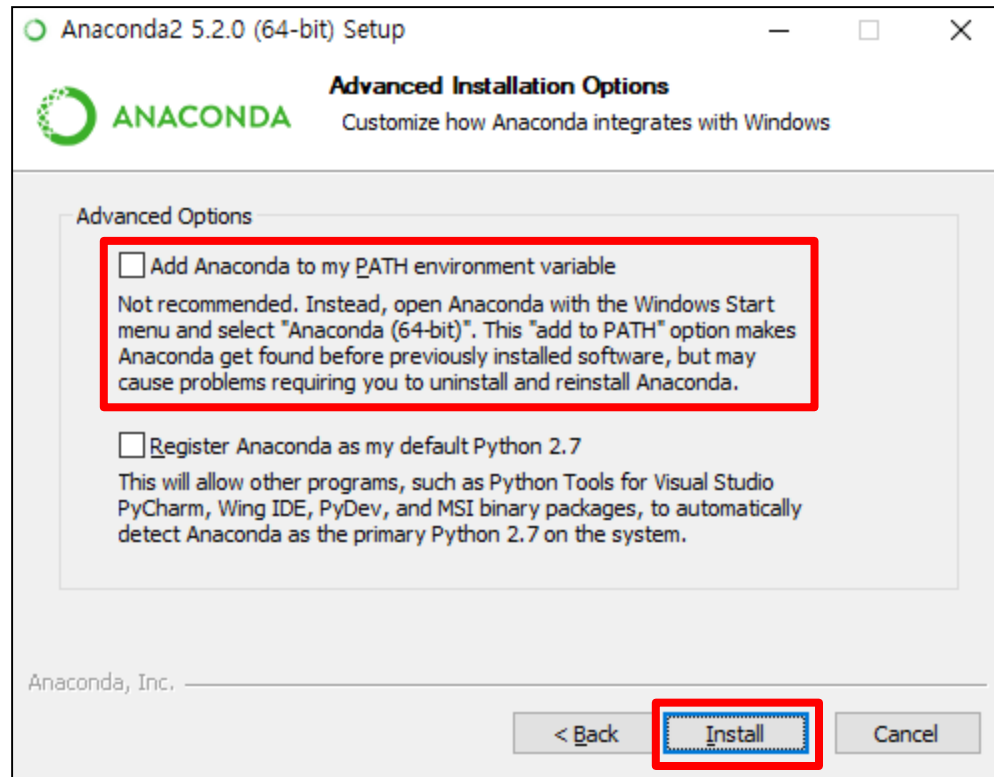


Anaconda



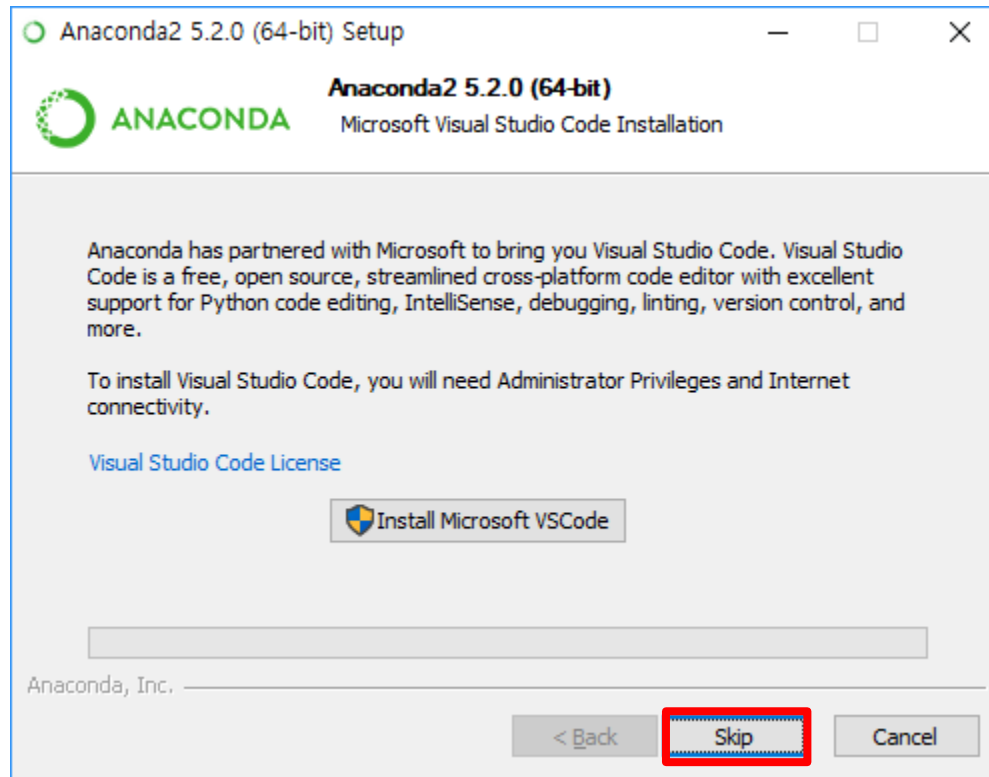
※ 자유롭게 원하는 폴더를 설정할 수 있으나, 기본적으로 설정되는 경로를 권장하고 사용자이름 및 폴더이름은 영어만 있는 것이 좋다.

Anaconda



※ 아나콘다만 설치해서 사용할 경우 두 번째 박스를 선택해도 되지만 파이참, 이클립스 등 다른 IDLE에서 파이썬을 사용하는 환경이라면 첫 번째 박스만 선택하는 것이 좋다.

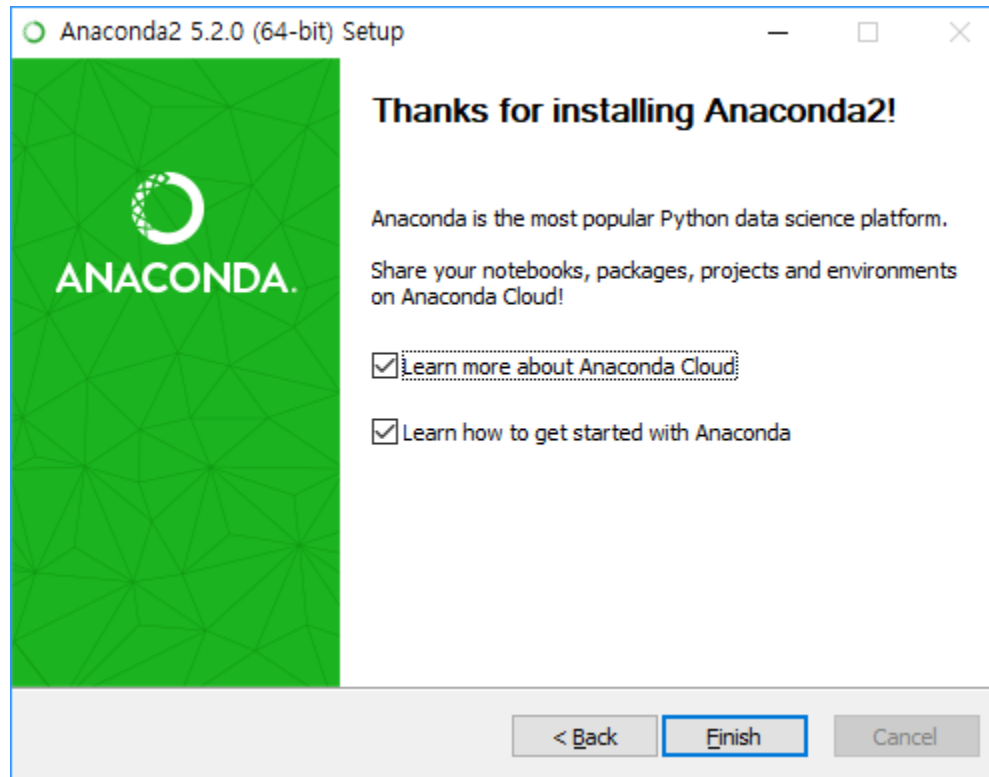
Anaconda



※ 아나콘다를 설치하면 기본적으로 Jupyter Notebook, Spyder 등의 파이썬 IDLE이 제공되지만 Visual Studio Code를 추가적으로 설치할 수 있다. 필요 없다면 **skip**, 필요하다면 **Install Microsoft VSCode**를 클릭한다.

Anaconda

설치완료





Anaconda

* 아나콘다 설치 확인하기 (검색 > cmd)

> conda --version

```
C:\Users\YEJl>conda --version  
conda 4.5.10
```

* 파이썬 확인하기

> python

```
C:\Users\YEJl>python  
Python 3.6.5 [Anaconda, Inc.] (default, Mar 29 2018, 13:32:41) [MSC v.1900 64 bit (AMD64)] on win32  
Type "help", "copyright", "credits" or "license" for more information.  
>>>
```

Anaconda

* 주피터 노트북 앱 환경설정

1) 주피터 노트북 폴더 생성

C:\Users\사용자\Downloads\python

- 1) 사용자 계정 아래 .jupyter 폴더
- 2) 파일이 없다면 config 파일 생성(아나콘다 프롬프트)

jupyter notebook --generate-config

3) jupyter_notebook_config.py 파일에서 경로변경

```
## The directory to use for notebooks and kernels.  
c.NotebookApp.notebook_dir = '원하는 경로 입력'
```