

# PYTHON for 데이터 분석

2019. 03



1. **Python 학습환경 세팅**
  - Python 3.7 with anaconda
  - Anaconda 가상환경 설정
  - Jupyter Notebook 설정
2. **Jupyter Notebook 주요기능**
3. **Python 학습 개요**
  - 예제를 통한 Python 기본 문법 및 주요 패키지 학습
  - Dementia Classification Data Summary
4. **Python Basics**
5. **Major Packages**

# Python 학습 환경 세팅

## □ Python 3.7 with anaconda

### ■ Anaconda?

1. Python 기반의 데이터 분석에 필요한 오픈소스를 모아 놓은 개발 플랫폼
2. 가상환경 및 패키지 관리 제공

### ■ Install Python

#### 1. Download URL

- 1) <https://www.anaconda.com/distribution>

#### 2. OS 선택( Windows, macOS, Linux)

#### 3. Python 3.7 Version Download

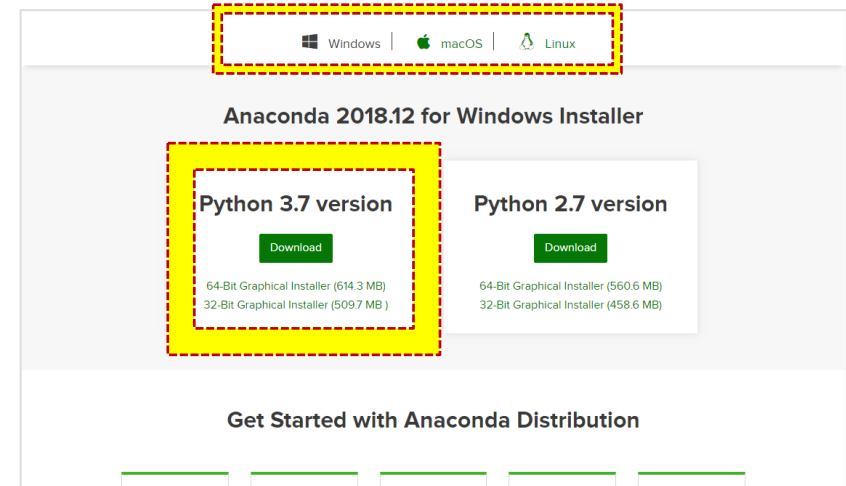
- 1) 64-Bit or 32-Bit System 중에 해당하는 버전 클릭
- 2) Windows System 정보 확인 방법
  - 제어판 -> 시스템

#### 4. default 설정으로 설치

### ■ 설치된 버전 확인

#### 1. Windows Start menu -> 'Anaconda Prompt' 실행

- 1) Command Prompt에서 아래 입력 후 실행
  - (base) c:\WUsers\Wmoodle> `python --version`
  - Python 3.7.1



## □ Python Anaconda

- Anaconda 버전 확인 및 업데이트 : 'Anaconda Prompt' 실행 후
  - 버전 확인
    - ✓ `conda --version`
  - 아나콘다 업데이트
    - ✓ `conda update conda`
  - 패키지 리스트 확인
    - ✓ `conda list`
  - 패키지 설치 : 사용 할 가상환경 활성화 후 진행
    - ✓ `conda install 패키지명`

## □ Python Anaconda 가상환경

### ▪ Anaconda 가상환경 기본

- `conda env list`

- ✓ `(base) C:\Users\Wmoodlee`

### ▪ Anaconda 가상환경 생성 및 관리

- 가상환경 리스트 확인

- ✓ `conda env list`

- ✓ `conda info --envs`

- 가상환경 생성

- ✓ `conda create --n 가상환경명 설치할패키지`

- `conda create --n tutorial python=3.5`

- 가상환경 활성화

- ✓ `conda activate tutorial`

- 가상환경 비활성화

- ✓ `conda deactivate`

- 가상환경 삭제

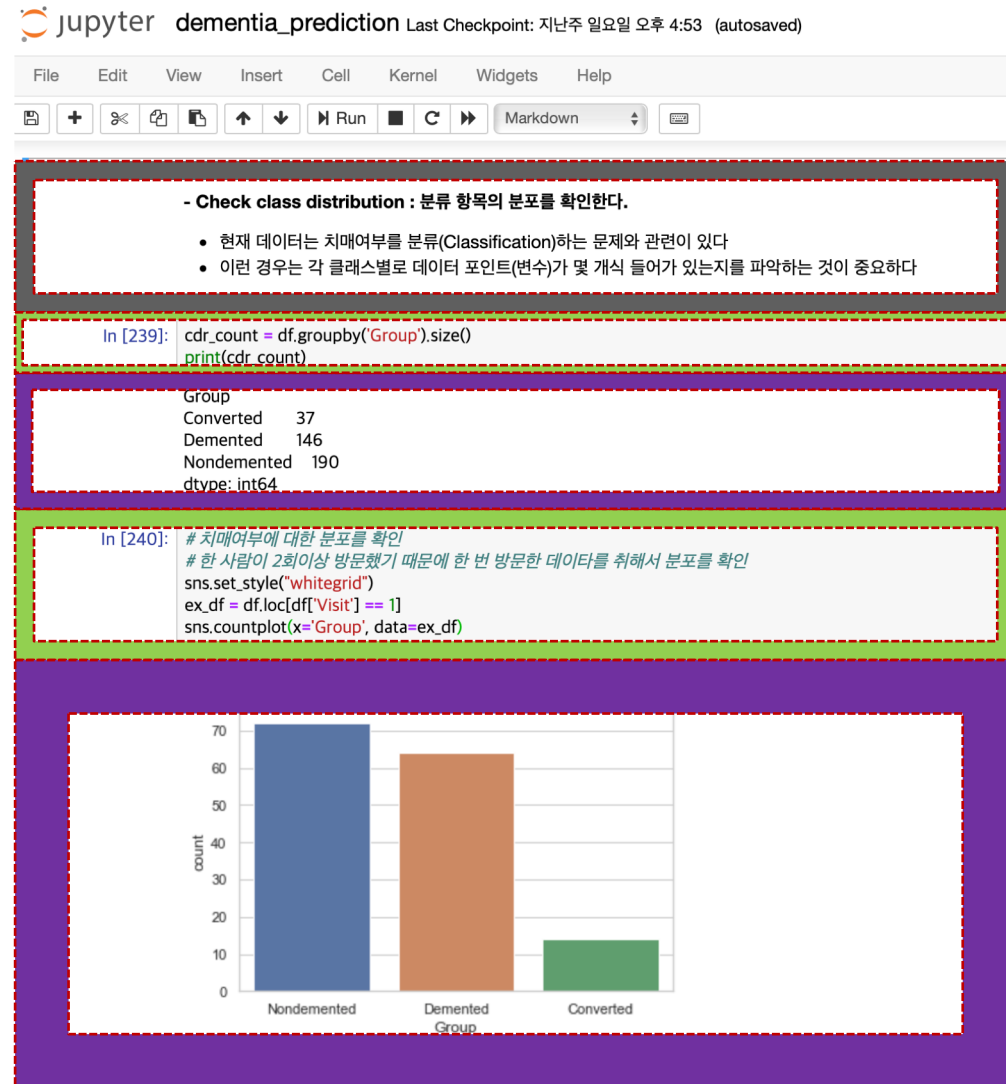
- ✓ `conda env remove --n tutorial`

## □ Jupyter Notebook

- Jupyter Notebook?
  - Interactive coding and visualization of output
- Jupyter Notebook 실행하는 3가지 방법
  - In Windows Start menu -> Anaconda3(64-bit)
    - ✓ 'Jupyter Notebook' 실행
    - ✓ 'Anaconda Prompt' 실행
      - Command Prompt에서 'jupyter notebook' 입력 후 실행
    - ✓ 'Anaconda Navigator' 실행
      - 'Jupyter Notebook' launch 클릭

# Jupyter Notebook 주요기능

- Notebook 생성
  - 우측 'New' ->를 클릭하면 노트북 생성
  - 왼쪽 'File'->'New Notebook'->'Python 3'
- Notebook 열기
  - 왼쪽 'File'->'Open'을 클릭하고 파일선택
- 주요 기능
  - Code
    - ✓ 단위 실행이 가능한 셀들로 구성
    - ✓ 인터랙티브 개발
    - ✓ 메뉴에서 '+'버튼을 선택하여 추가
    - ✓ 실행 결과를 바로 보여줌 : output
  - Markdown
    - ✓ 코드에 대한 설명
    - ✓ #, ##, ###로 폰트 크기 설정
- In[ ] : Ipython shell prompt
  - In[239]



# (부록)Python 학습 환경 세팅

## □ Jupyter Notebook 설정

### ■ Jupyter Notebook 실행 시 시작 디렉토리 설정하는 법

#### 1. 환경 설정 파일 생성

- 1) Anaconda Prompt 실행
- 2) Command Prompt에서 아래 입력 후 실행
  - `jupyter notebook --generate-config`
- 3) 실행 결과에 해당 설정 파일이 생성된 경로가 표시됨
  - Writing default config to: c:\Users\moodlelee\jupyter\jupyter\_notebook\_config.py

#### 2. 경로 설정

- 1) "jupyter\_notebook\_config.py" 파일을 텍스트 편집기로 열기
- 2) 찾기 기능(Ctrl +f)으로 "notebook\_dir"를 찾기
- 3) 맨 앞줄에 "#"로 주석처리 되어 있으므로 "#"삭제
- 4) `c.NotebookApp.notebook_dir = "`에서 `"` 안에 원하는 시작 디렉토리 입력
  - `c.NotebookApp.notebook_dir = 'c:\Users\moodlelee\work\python'`



# (부록) Jupyter Notebook Shortcuts

## ❑ Command mode, Edit mode 공통

- Shift + Enter run the current cell, select below
- Ctrl + Enter run selected cells
- Alt + Enter run the current cell, insert below
- Ctrl + S save and checkpoint

# (부록) Jupyter Keyboard Shortcuts

## Command Mode

- **Enter** take you into edit mode
- **H** Keyboard shortcuts
- **Shift + Space** scroll notebook up
- **Space** scroll notebook down
- **Shift + Up** extend selected cells above
- **Shift + Down** extend selected cells below
- **A** insert cell above
- **B** insert cell below

## Edit Mode

- **Esc** take you into command mode
- **Tab** code completion or indent
- **Ctrl + ]** indent
- **Ctrl + [** dedent
- **Shift + Tab** tooltip
- **Ctrl + A** select all
- **Ctrl + Shift + P** open the command palette

참고: <https://towardsdatascience.com/jupyter-notebook-shortcuts-bf0101a98330>

# Python 학습 개요

## □ 예제를 통한 Python 기본 문법 및 주요 패키지 학습

### ▪ 예제 시연

#### • Dementia Classification : Compare Classifier

- ✓ Dementia Classification.ipynb 예제를 통한 학습(파일로 공유 예정)
- ✓ 참고 : <https://www.kaggle.com/deepak525/dementia-classification-compare-classifiers>

### ▪ Python Basics

- Variables and Strings
- Lists, Tuples, Dictionaries, Sets
- If Statements, While loops, Switch
- Functions, Classes, Inheritance
- Files and Exceptions

### ▪ Popular Packages

- NumPy Basics : numerical computation
- Pandas Basics : numerical computation
- Matplotlib : data visualization
- Seaborn : data visualization
- Scikit-Learn : machine learning

# Python 학습 개요

## □ 첫 번째 예제인 'Dementia Classification Data' 설명

### ▪ Summary

- Longitudinal MRI 데이터를 사용할 것입니다.
- 데이터 세트는 60 세에서 96 세 사이의 150 명의 피험자의 Longitudinal MRI 데이터로 구성됩니다.
- 각 피험자는 2 회 이상 방문하여 총 373 회의 촬영 세션을 위해 1 년 이상 분리되었습니다
- 각 피험자는 적어도 한 번 스캔했습니다.
- 각 과목별로 단일 스캔 세션에서 얻은 3 ~ 4 개의 개별 T1 가중 MRI 스캔이 포함됩니다.
- 모두 오른 손잡이
- 연구 대상자 중 72 명은 'Nondemented'로 분류되었다.
- 대상자 중 64 명은 초기 방문시 'demented'로 분류되어 연구 전체에 걸쳐 그대로 유지되었습니다.
- 14 명의 피험자는 초기 방문 당시 'Nondemented'로 분류되어 나중에 방문했을 때 'Demented'로 특징 지어졌습니다. 이들은 'Converted'카테고리에 속합니다.

### ▪ Dataset

- Total Rows and Columns (Rows,Columns) : (373, 15)

# Python 학습 개요

## □ 'Dementia Classification Data'의 변수 의미 해석

### ▪ Basic Info : 기본 정보

- Subject ID : 피검자 ID
- MRI ID : MRI 번호
- Group : (Converted / Demented / Nondemented)
- Visit : 방문회수
- MR Delay : ???

### ▪ Demographics Info : 인구 통계 정보

- M/F : 성별
- Hand : 손잡이(모든 피검자는 오른 손잡이므로 열을 제거)
- Age : 나이
- EDUC : 교육기간(년)
- SES : 사회적 위치에 대한 Hollingshead Index에 의해 평가된 <사회경제적 지위> : 1(가장 높은 상태) ~ 5(가장 낮은 상태)로 분류

### ▪ Clinical Info : 임상정보

- MMSE : 간이정신상태검사 점수( 0 ~ 30) : 심한 ( $\leq 9$  점), 중등도 (10-18 점), 경미한 (19-23 점) 인지 장애
- CDR : 임상 치매 등급 (0 = 치매 없음, 0.5 = 매우 약한 AD, 1 = 경증 AD, 2 = 중등도 AD)

### ▪ Derived anatomic volumes : 유도된 해부학적 부피

- eTIV : 추정된 총 두개 내 용적, mm<sup>3</sup>
- nWBV : 자동화 된 조직 세분화 프로세스에 의해 회색 또는 흰색 물질로 분류 된 아틀라스 마스크 이미지의 모든 보셀의 백분율로 표현 된 표준화 된 전체 뇌량
- ASF : 아틀라스 스케일링 계수 (단위 없음). 네이티브 - 공간 뇌 및 두개골을 아틀라스 타겟으로 변환시키는 계산 된 스케일링 계수 (즉 변환 매트릭스의 결정자)

## ❑ 'Dementia Classification' Data Types

### ▪ Data types

1. Subject ID object
2. MRI ID object
3. Group object
4. Visit int64
5. MR Delay int64
6. M/F object
7. Hand object
8. Age int64
9. EDUC int64
10. SES float64
11. MMSE float64
12. CDR float64
13. eTIV int64
14. nWBV float64
15. ASF float64

# Python 학습 개요

## □ 'Dementia Classification' 시연

