

# 파이썬 머신러닝 판다스 데이터분석

## Lecture (3)



Dr. Heesuk Kim

Part 0. 개발환경 준비

**Part 1. 판다스 입문**

Part 2. 데이터 입출력

Part 3. 데이터 살펴보기

Part 4. 시각화 도구

Part 5. 데이터 사전처리

Part 6. 데이터프레임의 다양한 응용

Part 7. 머신러닝 데이터 분석



# Part 1. 판다스 입문

---

- 1. 데이터과학자가 판다스를 배우는 이유
- 2. 판다스 자료구조
  - 2-1. 시리즈
  - 2-2. 데이터프레임
- 3. 인덱스 활용
- 4. 산술 연산
  - 4-1. 시리즈 연산
  - 4-2. 데이터프레임 연산

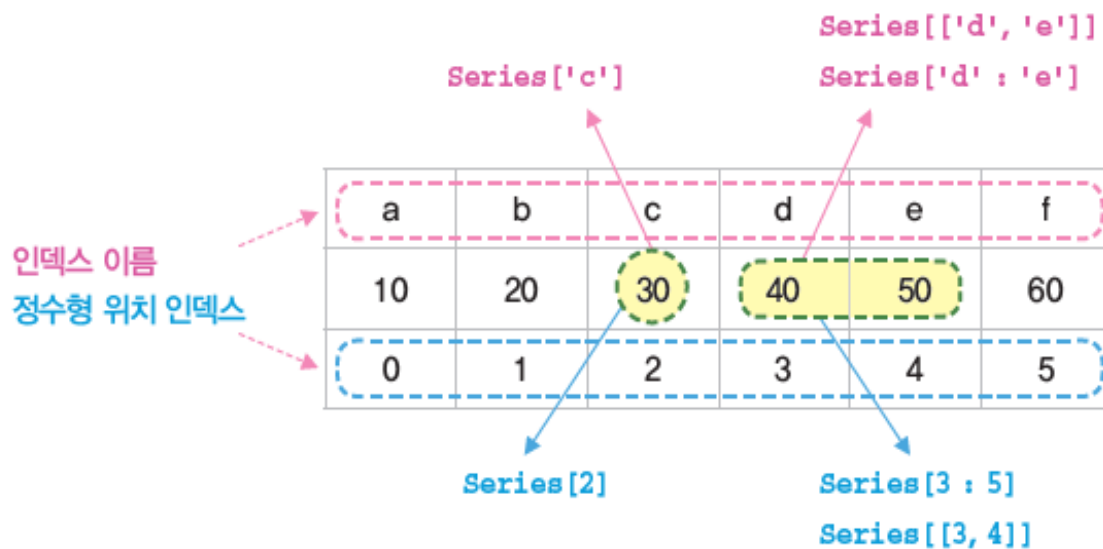


# Part 1. 판다스 입문

## 2-1. 시리즈

### • 원소 선택

- 1) 인덱스를 이용하여, 시리즈의 원소를 선택.
- 2) 하나의 원소를 선택하거나, 여러 원소를 한꺼번에 선택 가능.
- 3) 인덱스 범위를 지정하여 여러 개의 원소 선택 가능.
- 4) 인덱스의 유형에 따라 사용법이 조금 다르다.
  - 정수형 인덱스: 대괄호([ ]) 안에 숫자 입력. (0부터 시작)
  - 인덱스 이름(라벨): 대괄호([ ]) 안에 이름과 함께 따옴표를 입력. 큰 따옴표(" "), 작은 따옴표(' ') 모두 사용.



[그림 1-5] 시리즈 원소 선택

# Part 1. 판다스 입문

## 2-1. 시리즈

### 예제 1-3

#### ① 시리즈 만들기

파이썬 튜플을 시리즈로 변환한다. 정수형 위치 인덱스 대신에 인덱스 이름을 따로 지정할 수 있다. Series() 함수의 index 옵션에 인덱스 이름을 리스트 형태로 직접 전달하는 방식이다.

#### 〈예제 1-3〉 시리즈 원소 선택

(File: example/part1/1.3\_series\_element.py)

```
1 # -*- coding: utf-8 -*-
2
3 import pandas as pd
4
5 # 튜플을 시리즈로 변환(인덱스 옵션 지정)
6 tup_data = ('영인', '2010-05-01', '여', True)
7 sr = pd.Series(tup_data, index=['이름', '생년월일', '성별', '학생여부'])
8 print(sr)
```

#### 〈실행 결과〉 코드 1~8라인을 부분 실행

이름	영인
생년월일	2010-05-01
성별	여
학생여부	True
dtype:	object



# Part 1. 판다스 입문

## 2-1. 시리즈

### 예제 1-3

#### ② 원소 1개 선택

인덱스를 이용하여 원소를 선택할 때는 대괄호([ ]) 안에 인덱스를 입력한다. 시리즈의 첫 번째 데이터를 선택하기 위해 정수형 위치 인덱스(0)와 첫 번째 위치에 있는 인덱스 라벨('이름')을 입력한다.

〈예제 1-3〉 시리즈 원소 선택

(File: example/part1/1.3\_series\_element.py(이어서 계속))

```
11 # 원소를 1개 선택
12 print(sr[0])          # sr의 1번째 원소를 선택 (정수형 위치 인덱스)
13 print(sr['이름'])     # '이름' 라벨을 가진 원소를 선택 (인덱스 이름)
```

〈실행 결과〉 코드 11~13라인을 부분 실행

```
영인
영인
```



# Part 1. 판다스 입문

## 2-1. 시리즈

### • 원소 선택 (계속)

#### ③ 여러 개의 원소를 선택(인덱스 리스트 활용)

```
6 tup_data = ('영인', '2010-05-01', '여', True)
```

인덱스 리스트를 활용하는 방법이다.

여러 개의 인덱스를 리스트 형태로 대괄호([ ]) 안에 입력하면, 짝을 이루는 원소 데이터를 모두 반환한다.

정수형 위치 인덱스는 0부터 시작하기 때문에 2번째 인덱스 이름인 '생년월일'은 정수형 인덱스 1을 사용하고, 3번째 인덱스 이름인 '성별'은 정수형 인덱스 2를 사용한다.

〈예제 1-3〉 시리즈 원소 선택

(File: example/part1/1.3\_series\_element.py(이어서 계속))

생략

```
16 # 여러 개의 원소를 선택 (인덱스 리스트 활용)
17 print(sr[[1, 2]])
18 print('\n')
19 print(sr[['생년월일', '성별']])
```

〈실행 결과〉 코드 16~19라인을 부분 실행

```
생년월일    2010-05-01
성별         여
dtype: object
```

```
생년월일    2010-05-01
성별         여
dtype: object
```



# Part 1. 판다스 입문

## 2-1. 시리즈

### • 원소 선택 (계속)

#### ④ 여러 개의 원소를 선택(인덱스 범위 지정)

```
6 tup_data = ('영인', '2010-05-01', '여', True)
```

인덱스 범위를 지정하여 선택하는 방법이다.

22번 라인의 `sr[1 : 2]` 에서 정수형 위치 인덱스를 사용할 때는, 범위의 끝(2)이 포함되지 않는다('성별' 불포함).

그러나, 21번 라인의 `sr['생년월일' : '성별']` 과 같이, 인덱스 이름을 사용하면 범위의 끝('성별')이 포함된다.

〈예제 1-3〉 시리즈 원소 선택

(File : example/part1/1.3\_series\_element.py(이어서 계속))

~ ~~~ 생략 ~~~

```
22 # 여러 개의 원소를 선택 (인덱스 범위 지정)
23 print(sr[1 : 2])
24 print('\n')
25 print(sr['생년월일' : '성별'])
```

〈실행 결과〉 코드 22~25라인을 부분 실행

```
생년월일    2010-05-01
dtype: object
```

```
생년월일    2010-05-01
성별         여
dtype: object
```

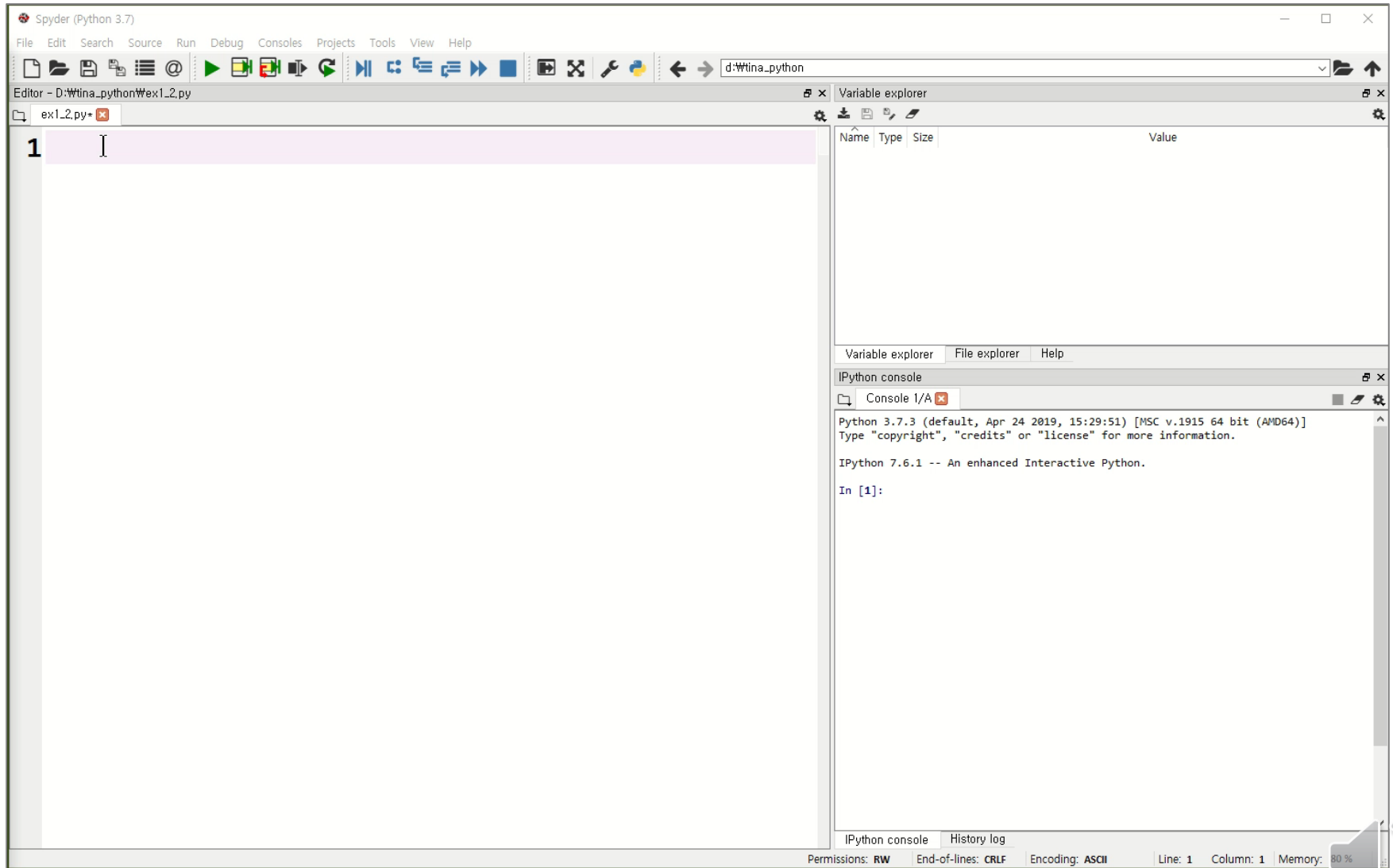




# Part 1. 판다스 입문

## 2-1. 시리즈

예제 1-3



# Part 1. 판다스 입문

## 2-1. 시리즈

- 딕셔너리 (Dictionary) - 중괄호 { } 사용
  - ✓ key 와 value 를 1 : 1 로 대응시킨 형태이다.
  - ✓ 하나의 key 에는 하나의 value 만 대응된다.
  - ✓ key 값은 변하지 않고, value 값은 변경이 가능하다.
  - ✓ 예 : `dict_data = { 'a':1, 'b': 2, 'c': 3 }`
- 리스트 (List) - 대괄호 [ ] 사용
  - ✓ 원소들이 연속적으로 저장되는 형태의 자료형이다.
  - ✓ 저장되는 요소들은 모두 같은 자료형일 필요는 없다.
  - ✓ 예 : `list_data = ['2019-01-02', 3.14, 'ABC', 100, True]`
- 튜플 (Tuple) - 소괄호 ( ) 사용
  - ✓ 여러 데이터를 동시에 저장할 수 있는 자료형으로서, 가장 단순한 자료형이다.
  - ✓ 예 : `tup_data = ('영인', '2010-05-01', '여', True)`



# Homework

1. 아나콘다 기반의 스파이더의 **작업 디렉토리 이름**을 **본인의 영어이름**으로 셋팅 후 화면 캡처 ( 3주차 1번 강의안 25 페이지 참고)
  2. **예제 1-1** 을 스파이더에서 실행 후 화면 캡처
  3. **예제 1-2** 을 스파이더에서 실행 후 화면 캡처
  4. **예제 1-3** 을 스파이더에서 실행 후 화면 캡처
- 위의 화면 캡처 내용 4개를 hwp 파일에 넣어서 제출하세요.
  - 파일명 : AI융합과\_과제(1)\_학번\_성명.hwp

과제 제출 기간 : 3월 26일(월)까지

과제 파일 제출 메일 주소 : **idistina@daum.net**

메일 제목 : **AI융합과\_과제(1)\_학번\_성명**

메일 내용 : hwp 파일 첨부

# Any Questions?

Thank you.

