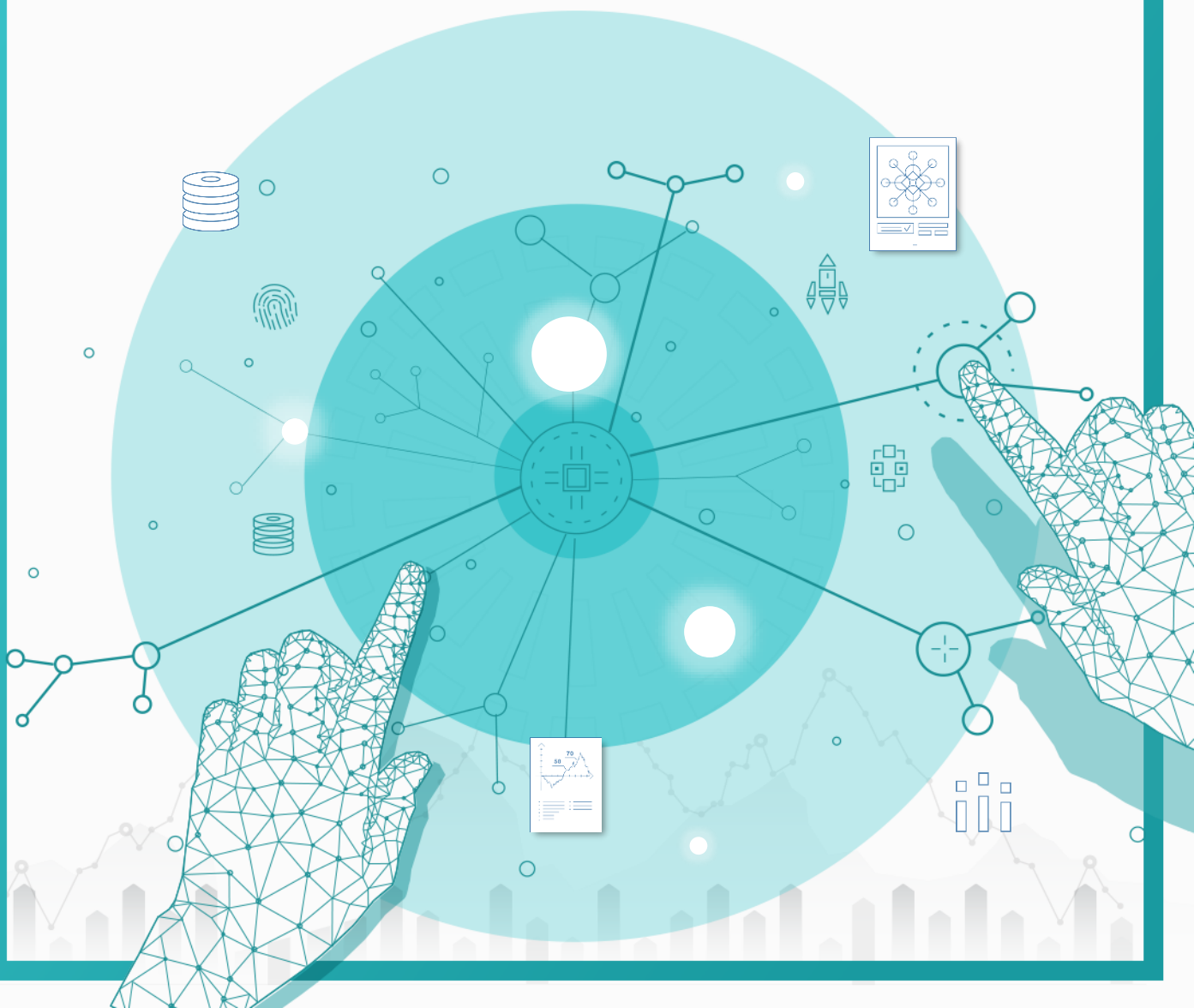




한국기술교육대학교
온라인평생교육원

「파이썬 라이브러리로 하는 데이터 분석과 시각화」

파이썬을 활용한 데이터 분석



파이썬을 활용한 데이터 분석

학습 목표

1. 데이터 분석을 위한 파이썬 필수 문법을 나열할 수 있다.
2. 파이썬의 기본 문법을 활용하여 데이터를 분석하는 방법을 설명할 수 있다.

학습 내용

1. 데이터 분석을 위한 파이썬의 필수 문법
2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1. 데이터 분석을 위한 파이썬의 필수 문법

1) 파이썬의 기본 자료형

(1) 파이썬의 특징

➡ 파이썬 언어는 데이터를 분석하는데 편리한 점이 있음

- ① 데이터 분석에 편리한 다양한 모듈
- ② 고수준의 내장 자료형, 동적 자료형 제공 등

▪ 동적인 자료형 결정

<pre>a = 1 print(type(a))</pre> <p><class 'int'></p>	<pre>b = 1.0 print(type(b))</pre> <p><class 'float'></p>	<pre>c = 'A' print(type(c))</pre> <p><class 'str'></p>
--	--	--

▪ 자동 메모리 관리

- 개발자가 직접 메모리에 신경 쓰며 코딩할 필요가 없음
- 파이썬이 내부적으로 자동으로 메모리를 관리해 주며, 자료형의 범위 또한 신경 쓸 필요가 없음

▪ 고수준의 기본 자료형, 함수 제공

- 문자열, 리스트, 딕셔너리 등 사용하기 편리한 자료형과 함수들을 제공

1. 데이터 분석을 위한 파이썬의 필수 문법

1) 파이썬의 기본 자료형

(1) 파이썬의 특징

- 일관되고 간결한 코드 작성
 - 데이터 분석을 하다 보면 코드가 점점 길어지고 복잡해지는데 들여 쓰기로 코딩 스타일을 일관되게 유지할 수 있음
 - 체이닝 등의 기능으로 짧고 간결하게 코드를 작성할 수 있음

```
a = ['0,1,2',3,4,5]
print(a[0].split(',')[1])
```

1

```
a = 1
if a > 0:
    if a > 2:
        if a > 3:
            print(1)
        else:
            print(2)
    elif a == 0:
        print(3)
    else:
        print(4)
```

1. 데이터 분석을 위한 파이썬의 필수 문법

1) 파이썬의 기본 자료형

(2) 파이썬 기본 자료형 종류

- 수치형 자료형
 - 숫자와 관련된 자료형
 - 정수(int), 실수(float), 복소수(complex) 자료형이 있음
 - 단일 객체로 길이가 없고 숫자들은 모두 하나의 값을 가짐

```
a = 4
b = 2.1
c = 1 + 2j
print(type(a))
print(type(b))
print(type(c))
```

```
<class 'int'>
<class 'float'>
<class 'complex'>
```

```
print(len(a))
```

```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-8-60806578a051> in <module>
----> 1 print(len(a))

TypeError: object of type 'int' has no len()
```

1. 데이터 분석을 위한 파이썬의 필수 문법

1) 파이썬의 기본 자료형

(2) 파이썬 기본 자료형 종류

▪ 문자열 자료형

- 문자를 나열한 자료형으로 여러 개의 문자를 순서대로 저장한 형태
- 값을 변경할 수 없으나, 다양한 함수 사용과 인덱싱, 슬라이싱 연산 가능

```
a = 'abcdef'
print(type(a))

<class 'str'>

print(a[0])
print(a[1:4])

a
bcd
```

```
a[4] = 'z'
```

TypeError Traceback (most recent call last)
<ipython-input-11-8b88da36af11> in <module>
----> 1 a[4] = 'z'

TypeError: 'str' object does not support item assignment

```
a.|
capitalize
casefold
center
count
encode
endswith
expandtabs
find
format
format_map
```

1. 데이터 분석을 위한 파이썬의 필수 문법

1) 파이썬의 기본 자료형

(2) 파이썬 기본 자료형 종류

■ 리스트 자료형

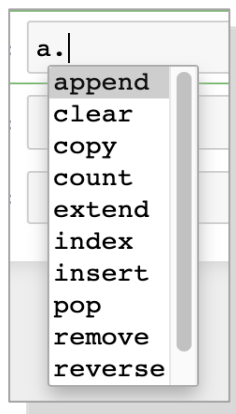
- 임의의 객체를 여러 순서대로 저장할 수 있는 배열 형태의 자료형
- 값을 변경할 수 있으며 다양한 함수를 활용하여 값을 추가, 수정, 삭제, 정렬하는 등의 작업을 할 수 있음

```
a = [1, 'a', [2, 'b'], 3]
print(type(a))

<class 'list'>

a.append('5')
print(a)

[1, 'a', [2, 'b'], 3, '5']
```



■ 딕셔너리 자료형

- 키와 값으로 이루어진 자료형
- 순서가 없어 인덱싱은 불가능하지만 키로 값을 참조할 수 있음

```
a = {1: ('a', 'b'),
      'k': [1, 2, 3],
      ('a', 'b'): 3.5}
print(type(a))

<class 'dict'>

print(a[('a', 'b')])

3.5
```

1. 데이터 분석을 위한 파이썬의 필수 문법

2) 파이썬의 모듈

- 파이썬에서 다양한 라이브러리를 제공하며 표준 모듈 외에도 pip 명령어를 활용하여 쉽게 설치, 사용할 수 있음

(1) 표준 모듈

- 파이썬을 설치하면 별도 과정 없이 바로 import 명령어로 호출하여 사용할 수 있는 모듈
- 해당 링크에서 자세한 목록 및 사용 방법 확인 가능
(<https://docs.python.org/ko/3/library/index.html>)

```
import math  
print(math.sqrt(9))  
3.0
```

```
import os  
print(os.getcwd())  
/Users/zoostar/OneDrive -
```

```
import random  
print(random.randint(1,100))  
47
```

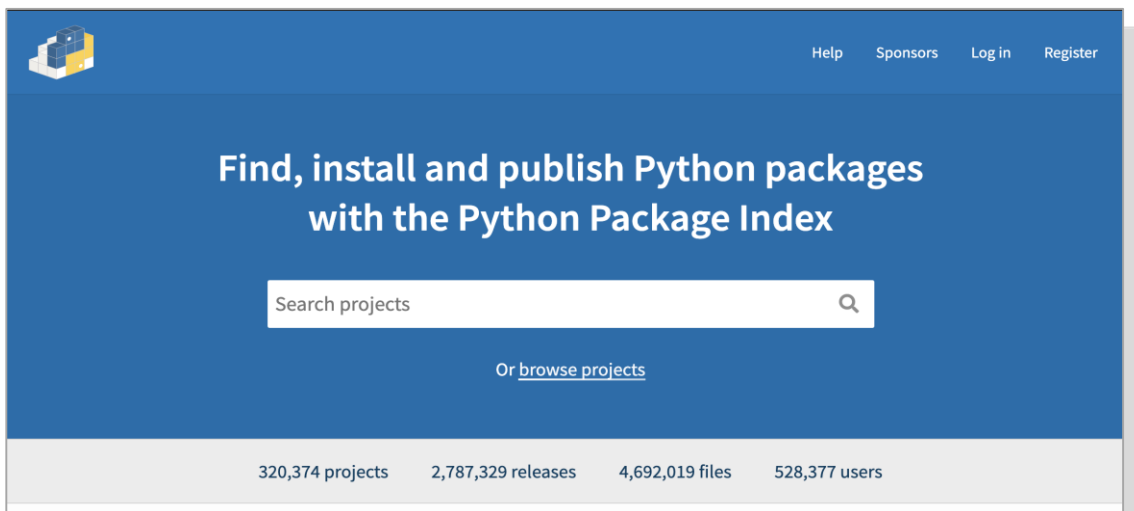

1. 데이터 분석을 위한 파이썬의 필수 문법

2) 파이썬의 모듈

(2) 외부 모듈

- pip 명령어로 설치한 뒤 사용할 수 있는 모듈
- 다양한 라이브러리를 제공함
 - ① 인공지능 관련: Tensorflow, Keras, Pytorch
 - ② 데이터 분석: Pandas, Numpy 등
- 명령 프롬프트(터미널) 혹은 아나콘다 네비게이터 프로그램에서 설치할 수 있음

```
zoostar@~$pip install 모듈명
```



1. 데이터 분석을 위한 파이썬의 필수 문법

3) 파이썬의 제어문

- 프로그램의 실행을 제어할 수 있는 문법으로 특정 조건 및 특정 객체에 따라 프로그램의 실행을 결정하거나 반복할 수 있음

(1) 조건문

- if, elif, else 문법을 활용하여 조건에 따라 코드 실행을 제어할 수 있음
- 문장의 시작과 끝, 포함 관계를 정의하기 위해 콜론과 들여 쓰기를 사용

```
a = 0
if a > 0:
    print(1)
print(2)
```

2

```
a = 0
if a > 2:
    print(0)
elif a > 1:
    print(1)
elif a > 0:
    print(2)
else:
    print(3)
```

3

(2) for 반복문

- 반복 가능한 객체를 한 번씩 순회하며 해당 코드를 반복

```
for a,b in [(1,2),(3,4)]:
    print(a,b)
```

1 2
3 4

```
for i in range(1,6,2):
    print(i)
```

1
3
5

1. 데이터 분석을 위한 파이썬의 필수 문법

3) 파이썬의 제어문

(3) while 반복문

- 조건을 만족할 때까지 코드를 반복할 수 있음
- 특정한 경우, break 보조 제어문으로 강제로 반복에서 빠져나올 수 있음
- 조건식 설정이 잘못된다면 무한 루프에 빠질 수 있음

```
count = 0
while count < 5:
    count += 1
    print(count)
```

```
1
2
3
4
5
```

```
count = 0
while True:
    count += 1
    print(count)
    if count == 4:
        break
```

```
1
2
3
4
```

2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1) 파이썬 기본 문법을 활용한 데이터 분석 방법

- 홈페이지, 데이터베이스, 외부 파일 등에서 수집한 데이터들은 수집 당시 형식은 모두 다르지만, 결국 파이썬의 자료형으로 읽어와 대부분 숫자, 문자 형태의 텍스트 데이터로 이루어져 있음

➡ 파이썬의 기본 문법들을 활용하여 텍스트 데이터 내에서 불필요한 데이터를 제거하거나 패턴을 찾아 필요한 데이터만 분리하는 등의 데이터 분석을 진행할 수 있음

(1) 텍스트 데이터 분석 예시

- 패턴을 찾아 불필요한 데이터를 지우고 필요한 데이터만 분리

예) 오류 로그 데이터 내 로그 메시지 분석

```
1 [2000-01-01 12:00:00] Data Analysis Error - Message:
2 java.lang.IllegalStateException: Illegal access: this web application instance has been stopped already. Could not load []. The
   following stack trace is thrown for debugging purposes as well as to attempt to terminate the thread which caused the illegal
   access.
3   at org.apache.catalina.loader.WebappClassLoaderBase.checkStateForResourceLoading(WebappClassLoaderBase.java:1352)
4   at org.apache.catalina.loader.WebappClassLoaderBase.getResource(WebappClassLoaderBase.java:1028)
5   at com.mysql.jdbc.AbandonedConnectionCleanupThread.checkContextClassLoaders(AbandonedConnectionCleanupThread.java:90)
6   at com.mysql.jdbc.AbandonedConnectionCleanupThread.run(AbandonedConnectionCleanupThread.java:63)
7   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
8   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
9   at java.lang.Thread.run(Thread.java:748)
10 [2000-01-01 10:00:52] Artifact manuscriptlink:war exploded: Artifact is deployed successfully
11 [2000-01-02 10:18:06] Artifact manuscriptlink:war exploded: Deploy took 22,363 milliseconds
```

데이터
분석결과

```
print(error_date)

[2000-01-01 12:00:00]

print(error_msg)

java.lang.IllegalStateException: Illegal access: this web application instance has been stopped already. Could not lo
ad []. The following stack trace is thrown for debugging purposes as well as to attempt to terminate the thread which
caused the illegal access.
    at org.apache.catalina.loader.WebappClassLoaderBase.checkStateForResourceLoading(WebappClassLoaderBase.java:1
352)
    at org.apache.catalina.loader.WebappClassLoaderBase.getResource(WebappClassLoaderBase.java:1028)
    at com.mysql.jdbc.AbandonedConnectionCleanupThread.checkContextClassLoaders(AbandonedConnectionCleanupThread.
java:90)
    at com.mysql.jdbc.AbandonedConnectionCleanupThread.run(AbandonedConnectionCleanupThread.java:63)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
    at java.lang.Thread.run(Thread.java:748)
```

2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1) 파이썬 기본 문법을 활용한 데이터 분석 방법

(2) TXT 파일 데이터

- 파이썬의 `open()` 함수를 활용하여 TXT 파일 형식의 데이터를 읽을 수 있음
 - r: 읽기 모드
 - w: 쓰기 모드

```
f = open('2000_01_01_error_log.txt', 'r')
data = f.read()
print(data)
f.close()
```

```
[2000-01-01 12:00:00] Data Analysis Error - Message:
java.lang.IllegalStateException: Illegal access: this web applic
s to attempt to terminate the thread which caused the illegal
at org.apache.catalina.loader.WebappClassLoaderBase.c
```

2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1) 파이썬 기본 문법을 활용한 데이터 분석 방법

(3) CSV 파일 데이터

- 파이썬의 open() 함수를 활용하여 TXT 파일 형식의 데이터를 읽을 수 있음
- 단, CSV 형식의 특징인 콤마로 구분된 값들을 읽는 것이 아닌 단순 문자열 형태로 데이터를 읽음

```
f = open('data.csv','r')
data = f.read()
print(data)
```

```
name,phone,age,number
홍길동,02-111-1111,11,0000
이길동,041-123-1234,22,1234
박길동,031-000-0000,34,4938
```

```
print(type(data))
print(data[:14])
```

```
<class 'str'>
name,phone,age
```

2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1) 파이썬 기본 문법을 활용한 데이터 분석 방법

(3) CSV 파일 데이터

- 파이썬의 표준 모듈 중 CSV 모듈을 활용하여 CSV 파일 형식 그대로 (콤마로 구분된 상태) 읽을 수 있음

```
import csv
f = open('data.csv', 'r')
data = csv.reader(f)
print(type(data))
```

```
<class '_csv.reader'>
```

```
for i in data:
    print(i)
```

```
['name', 'phone', 'age', 'number']
['홍길동', '02-111-1111', '11', '0000']
['이길동', '041-123-1234', '22', '1234']
['박길동', '031-000-0000', '34', '4938']
```

2. 파이썬 기본 문법을 활용한 데이터 분석 방법의 이해

1) 파이썬 기본 문법을 활용한 데이터 분석 방법

(4) JSON 형식 데이터

- 파이썬의 JSON 모듈을 활용하여 JSON 형식의 데이터를 파이썬의 객체로 변환하여 활용할 수 있음

```
import json
a = [1, 2, 3, {'4': 5, '6': 7}]

b = json.dumps(a)
print(b)

[1, 2, 3, {"4": 5, "6": 7}]

c = json.loads(b)
print(c)

[1, 2, 3, {'4': 5, '6': 7}]
```

→ dumps 함수:
파이썬 → JSON 변환

→ loads 함수:
JSON → 파이썬 변환