

# Python 데이터 분석하기

2019 경북고등학교

**들어가기 전에...**

**데이터 분석을 왜 할까요?**

**믿을 만한 데이터는 어디에 있을까요?**

**데이터 분석 결과를 어떻게 보여주면 효과적일까요?**

# 1. 수업 자료는 여기에 있어요!

[https://github.com/swKyungbock/2019\\_DataProject/](https://github.com/swKyungbock/2019_DataProject/)

# 2. 레퍼런스를 참고하며 프로그래밍 하는 습관!

## 2-1. python 라이브러리 레퍼런스

<https://docs.python.org/3/library/index.html>

## 2-2. pandas 라이브러리 레퍼런스

<https://pandas.pydata.org/pandas-docs/stable/reference/index.html>

## 2-3. matplotlib

[https://matplotlib.org/api/pyplot\\_summary.html](https://matplotlib.org/api/pyplot_summary.html)

### 3. 참고하면 좋은 책 (도서관에 있어요!)

- 너도 데이터 가지고 놀 수 있어
- Do it! 판다스 입문
- 모두의 데이터 분석 with 파이썬

차시	날짜	주 제	학 습 내 용
1	1/6	Python으로 csv파일 다루기	파일처리 [project]매년 내 생일에 서울 기온은 어땠을까?
2	1/7	데이터의 시각화	matplotlib 라이브러리로 그래프 그리기 [project]매년 내 생일에 서울 기온 그래프 그리기
3	1/8	numpy 라이브러리 활용 pandas 라이브러리 활용	데이터 분석 라이브러리 활용하기 [project]서울시 청소년 스트레스 분석
4	1/9	인구 구조 시각화하기 지하철 데이터 시각화하기	공공데이터 분석 프로젝트 [project] 우리동네 인구구조 분석 [project] 지하철 데이터 분석
5	1/10	데이터 분석 팀 프로젝트 주제 선정 및 개발	공공데이터 분석 프로젝트 [project] 나만의 프로젝트 구상 및 개발
6	1/13	팀 프로젝트 완성 및 발표	09:00~10:30 프로젝트 발표 준비 및 완성 10:30~11:30 프로젝트 발표

## 수업의 최종 목표!!

- 각자 흥미있는 주제로 **데이터 분석 프로젝트 완성**하기!

# 내가 관심 있는 주제에 관한

공개된 데이터가 있는지 관심을 가지고 살펴보세요!

# **플튀 금지**(feat. 이로운 선생님)

## [폴더 정리]

### 1. jupyter notebook 기본 폴더

**C:\Users\User**

### 2. C:\Users\User에 자기폴더(예:2020KskDataProject)를 만드세요.

**C:\Users\User\2020KskDataProject**

### 3. 2에서 만든 폴더 아래 data폴더를 만드세요.

**C:\Users\User\2020KskDataProject\data**

*day1*

# Python으로 CSV 파일 다루기

2019 경북고등학교



**#1. File Input/Output의 개요**

**#2. CSV 파일 다루기**

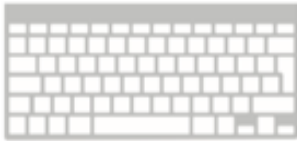
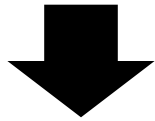
**#3. [project] 내 생일과 관련된 날씨 기온 알아보기**

# #1. File Input / Output 개요

## [표준 입출력]

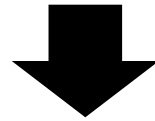
지금까지 여러분이 프로그래밍을 하며 경험했던 **입력**과 **출력**은 무엇인가요?

**input( )**



표준입력장치

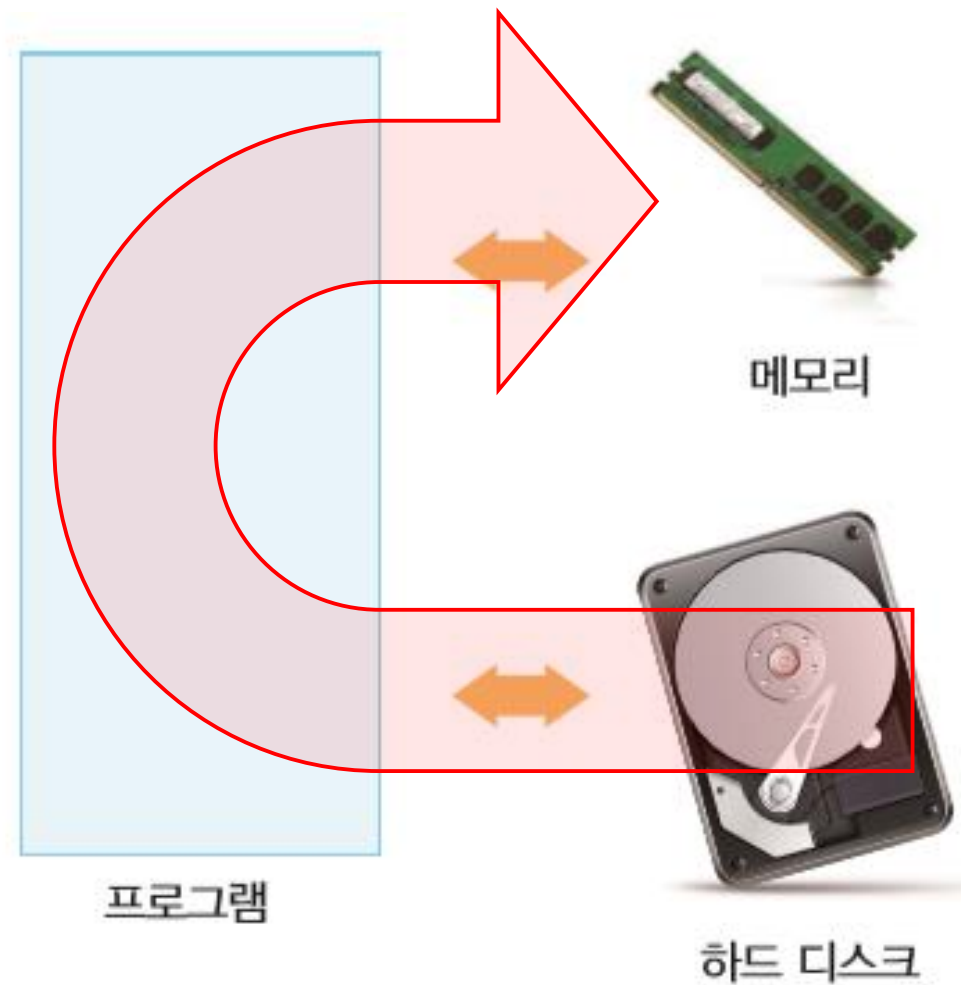
**print( )**



표준출력장치

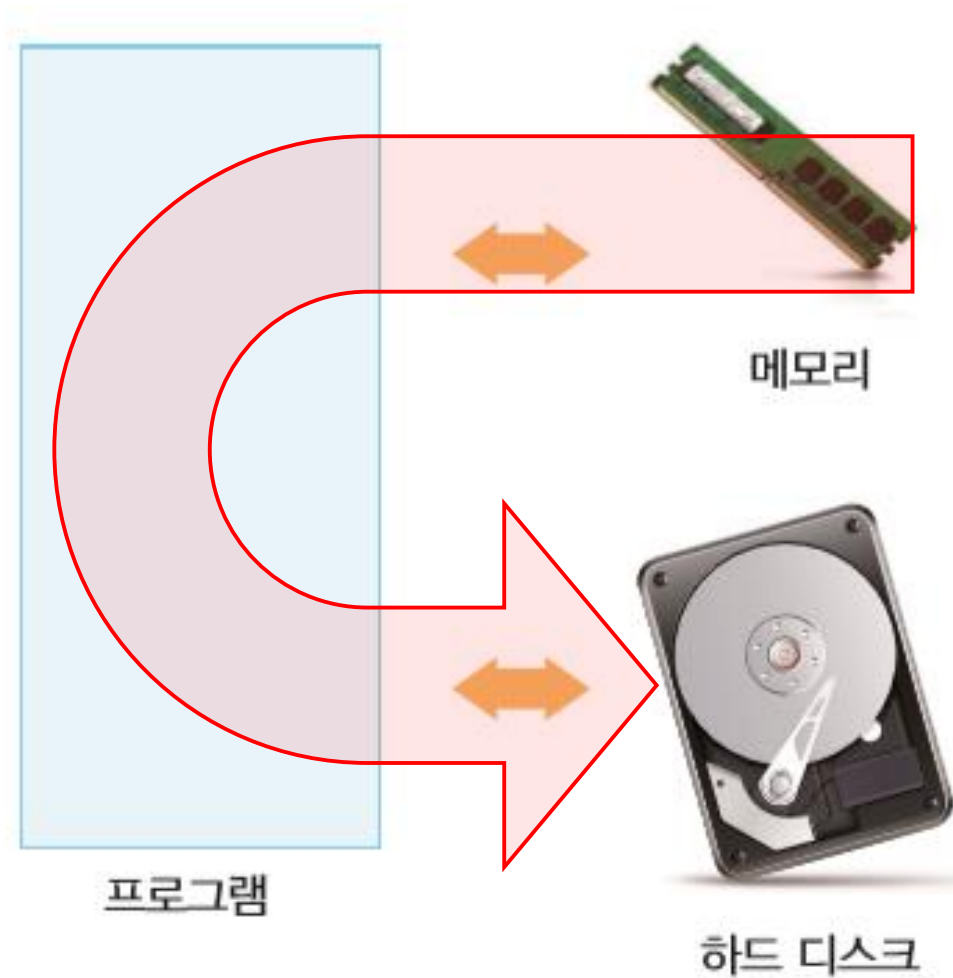
## [파일 입출력]

이제 우리가 하려는 것 : **파일 입력**과 **파일 출력**



## [파일 입출력]

이제 우리가 하려는 것 : **파일 입력**과 **파일 출력**



## [파일 입출력 알고리즘]

파일의 내용을 **입력**하거나 **출력**하려면 어떻게 해야 할지 절차를 생각해 볼까요?

1. 파일을 연다.
2. 파일의 내용을 읽거나 파일에 새로운 내용을 쓴다.
3. 필요한 데이터 처리
4. 파일을 닫는다.

## [파일 입출력 알고리즘]

파일의 내용을 입력하거나 출력하려면 어떻게 해야 할지 절차를 생각해 볼까요?

1. 연다.

open

2. 읽거나 쓴다.

read / write

3. 필요한 데이터 처리

4. 닫는다.

close

## [파일 입력 절차 - 1단계 : 파일 열기 / open]

# **open** -> read -> close

파일접근변수명 = **open**("파일명", "**모드**")

**f=open("test.txt", "r")** # test.txt파일을 읽기 모드로 연다



# [파일 입력 절차 - 1단계 : 파일 열기 / open]

# 파일 열기 모드

파일 모드	모드	설명
"r"	<b>Read</b>	파일을 읽기만 할 때 사용 / 파일이 없으면 Error
"r+"	Read or Write	파일을 읽거나 쓸 때 사용 / 파일이 없으면 Error
"w"	<b>Write</b>	파일에 내용을 쓸 때 사용 / 파일이 없으면 새로 만든다.
"a"	<b>Append</b>	파일의 마지막에 새로운 내용을 추가할 때 사용 / 파일이 없으면 새로 만든다

## [파일 입력 절차 - 2단계 : 파일의 내용 읽기 / read]

# open -> **read** -> 데이터 처리 -> close

파일접근변수명.**read**( )

**data=f.read( )** #파일의 내용을 읽어서 data에 저장

## [파일 입력 절차 - 3단계 : 데이터 처리]

# open -> read -> 데이터 처리 -> close

필요한 처리를 수행

## [파일 입력 절차 - 4단계 : 파일 닫기 / close]

# open -> read -> 데이터 처리 -> **close**

변수명.**close**( )

**f.close( )** # 파일을 닫는다

## [파일 입력 절차]

# open -> read -> 데이터 처리 -> close

```
f=open('readTest.txt','r') #파일열기 : 읽기모드
```

```
data=f.read() #read : 파일 입력
```

```
print(data) #데이터 처리
```

```
f.close() #파일 닫기
```

## [파일 출력 절차]

# open -> 데이터 처리 -> write -> close

```
f=open('writeTest.txt','w') #파일열기 : 쓰기 모드
```

```
data=['1,2,3,4,5'] #데이터 처리
```

```
data.sort(reverse=True)
```



```
max=data[0]
```

```
f.write('max='+str(max)) #write : 파일 출력
```

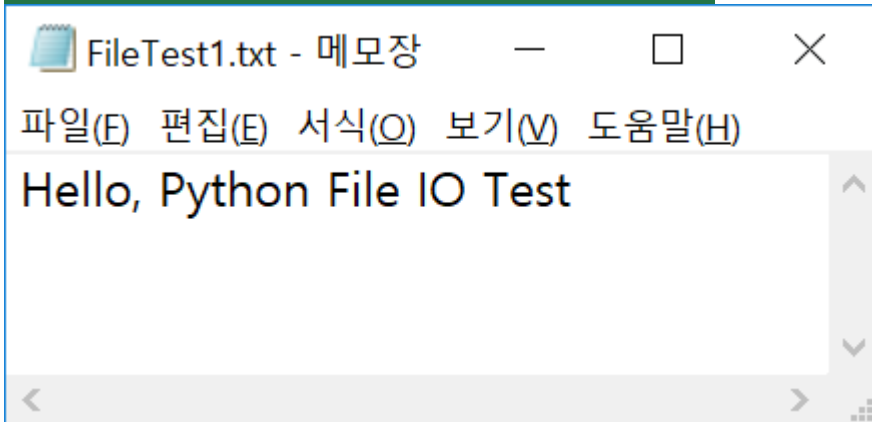
```
f.close() #파일닫기
```

## [실습1]

**data** 폴더에서 **FileTest1.txt** 파일의 데이터를 읽어서 **화면에 출력**해 볼까요?

폴더구조  1일차\_csv파일다루기.ipynb  
 data

### FileTest1.txt 파일 내용



### 화면 출력

파일의 내용 출력  
Hello, Python File IO Test

## [실습1-알고리즘 생각하기]

0106FileIO 폴더에서 FileTest1.txt 파일의 데이터를 읽어서 화면에 출력해 볼까요?

폴더구조

- 0106FileIO
  - 1일차.ipynb

1. FileTest1.txt파일을 읽기 모드로 열기
2. FileTest1.txt파일의 내용을 읽기 : 변수 data에 저장
3. 읽은 내용을 화면에 출력하기
4. 파일 닫기



## [실습1-알고리즘 생각하기]

1. FileTest1.txt파일을 읽기 모드로 열기
2. FileTest1.txt파일의 내용을 읽기 : 변수 data에 저장
3. 읽은 내용을 화면에 출력하기
4. 파일 닫기

```
f=open(' ./data/FileTest1.txt', 'r')  
data=f.read()  
print("파일의 내용 출력")  
print(data)  
f.close()
```

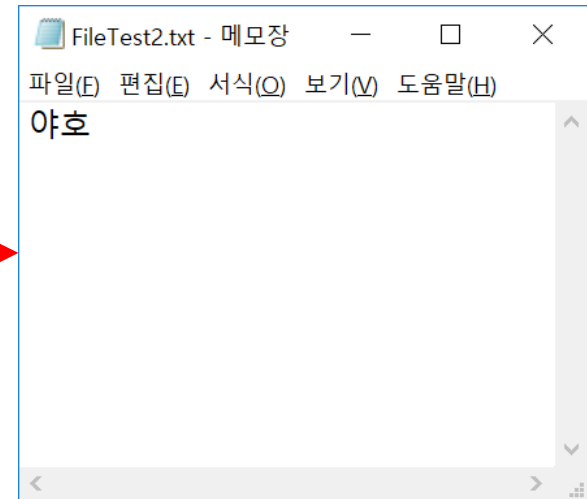
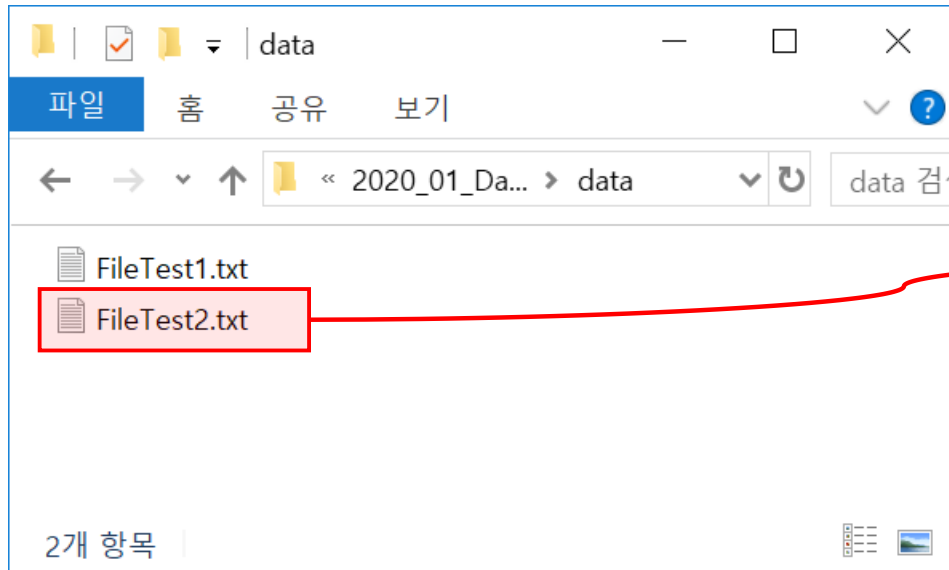
## [실습2]

키보드로 입력 받은 데이터를 dataWFileTest2.txt 파일에 써 볼까요?

폴더구조

- 1일차\_csv파일다루기.ipynb
- data

파일에 쓸 내용 입력 : 야호



## [실습2-알고리즘 생각하기]

**키보드로 입력** 받은 데이터를 FileTest2.txt **파일에 써 볼까요?**

1. FileTest2.txt파일을 쓰기 모드로 열기
2. 키보드로 데이터를 입력받기 : 변수 data에 저장
3. 입력받은 내용을 파일에 출력하기
4. 파일 닫기

## [실습2-알고리즘 생각하기]

키보드로 입력 받은 데이터를 FileTest2.txt 파일에 써 볼까요?

1. FileTest2.txt파일을 쓰기 모드로 열기
2. 키보드로 데이터를 입력받기 : 변수 data에 저장
3. 입력받은 내용을 파일에 출력하기
4. 파일 닫기

```
f=open( './data/FileTest2.txt', 'w' )  
data=input("파일에 쓸 내용 입력 : ")  
f.write(data)  
f.close()
```

## #2. CSV 파일 다루기

## [CSV 파일 : Comma-Separated Values ]

# 각각의 데이터를 쉼표(,)로 분리된 파일

# import csv

# csv.reader(csvFile) : csvFile에서 읽기

# csv.writer(csvFile) : csvFile에 쓰기

# [실습1] classLib.csv 파일을 읽은 후 화면에 내용 출력하기

## classLib.csv 파일 내용

classLib.csv - Excel									
파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 입력하세요... 로그인 공유									
	A	B	C	D	E	F	G	H	I
1	연번	분야	도서명	저자	출판사	발행일	대출여부	대출학생	번호
2	1	문학예술	만이	토어 세이	논장	2017.7.25.			
3	2	문학예술	인생詩선	손나라	트로이목마	2017.8.30.			
4	3	인문학	어린이 대	이만열/이	창비	2017.7.25.			
5	4	인문학	눈, 새로운	김용희 외	궁리	2017.7.28.			
6	5	사회과학	종횡무진	윤덕노	깊은나무	2017.8.31.			
7	6	사회과학	실험실에서	김벌리 베	초록개구리	2017.8.23.			
8	7	자연과학	기억한다						
9	8	실용일반	10대, 나의						
10	9	유아아동	새끼표범						
11	10	유아아동	플로팅 아						
12									
13									

화면출력

```
[ '연번', '분야', '도서명', '저자', '출판사', '발행일', '대출여부', '대출학생 번호' ]
[ '1', '문학예술', '만이', '토어 세이들러 글, 조원희 그림/권자심', '논장', '2017.7.25.', '', '' ]
[ '2', '문학예술', '인생詩선', '손나라', '트로이목마', '2017.8.30.', '', '' ]
[ '3', '인문학', '어린이 대학: 역사', '이만열/이광희?글, 이주희 그림', '창비', '2017.7.25.', '', '' ]
[ '4', '인문학', '눈, 새로운 발견', '김용희 외', '궁리', '2017.7.28.', '', '' ]
[ '5', '사회과학', '종횡무진 밥상견문록', '윤덕노', '깊은나무', '2017.8.31.', '', '' ]
[ '6', '사회과학', '실험실에서 만든 햄버거는 무슨 맛일까?', '김벌리 베네스/김아람', '초록개구리', '2017.8.23.', '', '' ]
[ '7', '자연과학', '기억한다는 것', '이현수 글, 김진화 그림', '너머학교', '2017.7.25.', '', '' ]
[ '8', '실용일반', '10대, 나의 발견', '윤주옥 외', '글담출판', '2017.8.15.', '', '' ]
[ '9', '유아아동', '새끼표범', '강무홍 글, 오승민 그림', '한울림어린이', '2017.7.24.', '', '' ]
[ '10', '유아아동', '플로팅 아일랜드', '김려령 글, 이주미 그림', '비룡소', '2017.8. 1.', '', '' ]
```

# [실습1] classLib.csv 파일을 읽은 후 화면에 내용 출력하기

## 알고리즘 생각하기

1. csv모듈 импорт
2. classLib.csv파일을 읽기 모드로 열기
3. 파일의 내용 읽어와 data변수에 저장하기
4. 화면에 내용(data) 출력하기
5. 파일 닫기



# [실습1] classLib.csv 파일을 읽은 후 화면에 내용 출력하기

## 알고리즘 생각하기

- |                              |  |
|------------------------------|--|
| 1. csv모듈 импорт              | <code>import csv</code>                        |
| 2. classLib.csv파일을 읽기 모드로 열기 | <code>f=open('./data/classLib.csv','r')</code> |
| 3. 파일의 내용 읽어와 rData변수에 저장하기  | <code>rData=csv.reader(f)</code>               |
| 4. 화면에 내용(data) 출력하기         | <code>print(rdata)</code>                      |
| 5. 파일 닫기                     | <code>f.close( )</code>                        |

## [실습1] classLib.csv 파일을 읽은 후 화면에 내용 출력하기

### 문제점 발견

```
import csv
f=open('./data/classLib.csv','r')
rData=csv.reader(f)
print(rData)
f.close()
```

```
<_csv.reader object at 0x0000029E5D4DA2B8>
```

→ Iterator 타입인 reader 객체를 리턴

## [실습1] classLib.csv 파일을 읽은 후 화면에 내용 출력하기

### 문제점 해결

```
import csv
f=open('./data/classLib.csv','r')
rData=csv.reader(f)
print(rData)
for row in rData :
    print(row)
f.close()
```

loop문으로 한 라인씩 출력

## [실습2] classLib.csv에서 연번과 도서명만 출력하기

연번	도서명
1	말이
2	인생詩選
3	어린이 대학: 역사
4	눈, 새로운 발견
5	종횡무진 밥상견문록
6	실험실에서 만든 햄버거는 무슨 맛일까?
7	기억한다는 것
8	10대, 나의 발견
9	새끼표범
10	플로팅 아일랜드

## [실습2] classLib.csv에서 연번과 도서명만 출력하기

```
[ '연번', '분야', '도서명', '저자', '출판사', '발행일', '대출여부', '대출학성',
[ '1', '문학예술', '말이', '토어 세이들러 글, 조원희 그림/권자심', '논장',
[ '2', '문학예술', '인생詩선', '손나라', '트로이목마', '2017.8.30.', '', '' ]
[ '3', '인문학', '어린이 대학: 역사', '이만열/이광희?글, 이주희 그림', '창비
[ '4', '인문학', '눈, 새로운 발견', '김윤희 외', '궁리', '2017.7.28.', '', ''
[ '5', '사회과학', '종횡무진 밥상견문록', '윤덕노', '깊은나무', '2017.8.31.'
[ '6', '사회과학', '실험실에서 만든 햄버거는 무슨 맛일까?', '김벌리 베네스/김
[ '7', '자연과학', '기억한다는 것', '이현수 글, 김진화 그림', '너머학교', '2
[ '8', '실용일반', '10대, 나의 발견', '윤주옥 외', '글담출판', '2017.8.15.',
[ '9', '유아아동', '새끼 표범', '강무홍 글, 오승민 그림', '한울림어린이', '20
[ '10', '유아아동', '플로팅 아일랜드', '김려령 글, 이주미 그림', '비룡소', ''
```

**[Hint] 데이터 타입을 잘 살펴보자!**

## [실습2] classLib.csv에서 연번과 도서명만 출력하기

[0]

[2]

['연번', '분야', '도서명', '저자', '출판사', '발행일', '대출여부', '대출학생']  
[1, '문학예술', '말이', '토어 세이들러 글, 조원희 그림/권자심', '논장', '2017.8.30.', '', '']

```
import csv
```

```
f=open('./data/classLib.csv','r')
```

```
rData=csv.reader(f)
```

```
for row in rData :
```

```
    print(row[0]+"\\t"+row[2])
```

```
f.close()
```

리', '2017.7.28.', '', '']  
'깊은나무', '2017.8.31.'  
일까?', '김벌리 베네스/김  
화 그림', '너머학교', '2  
글담출판', '2017.8.15.',  
림', '한울림어린이', '20  
|주미 그림', '비룡소', ''

## [실습3] classLib.csv파일에 데이터 추가하기]

- ('문학예술', '언어의온도', '이기주', '말글터', '2016.8.19') 추가

### 알고리즘 생각하기

1. csv모듈 임포트
2. classLib.csv파일을 쓰기 모드로 열기
3. data변수에 추가할 데이터 적기
4. data변수를 파일에 쓰기
5. 파일 닫기

## [실습3] classLib.csv파일에 데이터 추가하기

- ('문학예술', '언어의온도', '이기주', '말글터', '2016.8.19') 추가

### 알고리즘 생각하기

1. csv모듈 임포트

```
import csv
```

2. classLib.csv파일을 추가 모드로 열기

```
f=open('./data/classLib.csv','a')
```

3. data변수에 추가할 데이터 적기

```
data=['문화예술','언어의온도', ... ]
```

4. data변수를 파일에 쓰기

```
w=csv.writer(f)
```

```
w.writerow(data)
```

5. 파일 닫기

```
f.close( )
```



## [실습3] classLib.csv파일에 데이터 추가하기

- ('문학예술', '언어의온도', '이기주', '말글터', '2016.8.19') 추가

7	6	사회과학	실험실에서	김벌리 베	초록개구리	2017.8.23.			
8	7	자연과학	기억한다는	이현수 글,	너머학교	2017.7.25.			
9	8	실용일반	10대, 나의	윤주옥 외	글담출판	2017.8.15.			
10	9	유아아동	새끼표범	강무홍 글,	한울림어린이	2017.7.24.			
11	10	유아아동	플로팅 아	김려령 글,	비룡소	2017.8. 1.			
12		문학예술	언어의온도	이기주	말글터	2016.8.19			
13									

classLib (+)

준비

100 %

생각할 꺼리 : 연번을 자동으로 붙여보자.

## [실습3] classLib.csv파일에 데이터 추가하기

classLib.csv - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 입력하세요... 로그인 공유

A1 : 연번

	A	B	C	D	E	F	G	H	I
1	연번	분야	도서명	저자	출판사	발행일	대출여부	대출학생	번호
2	1	문학예술	말이	토어 세이	논장	2017.7.25.			
3	2	문학예술	인생詩선	손나라	트로이목마	2017.8.30.			
4	3	인문학	어린이 대	이만열/이	창비	2017.7.25.			
5	4	인문학	눈, 새로운	김용희 외	궁리	2017.7.28.			
6	5	사회과학	종횡무진	윤덕노	깊은나무	2017.8.31.			
7	6	사회과학	실험실에서	김벌리 베	초록개구리	2017.8.23.			
8	7	자연과학	기억한다는	이현수 글,	너머학교	2017.7.25.			
9	8	실용일반	10대, 나의	윤주옥 외	글담출판	2017.8.15.			
10	9	유아아동	새끼표범	강무홍 글,	한울림어린이	2017.7.24.			
11	10	유아아동	플로팅 아	김려령 글,	비룡소	2017.8. 1.			
12	11	문화예술	언어의온도	이기주	말글터	2016.8.19			
13									

classLib

준비

100 %

## [실습3] classLib.csv파일에 데이터 추가하기

```
import csv
f=open('./data/classLib.csv','r+',newline='')
rData=csv.reader(f)
w=csv.writer(f)
header=next(rData) #첫째줄 처리
rData=list(rData)
cnt=len(rData)
newData=[cnt+1, '문학예술', '언어의온도', '이기주', '말글터', '2016.8.19']
rData.append(newData)
w.writerow(newData)
for row in rData:
    row[0]=int(row[0])
    print(row)
f.close()
```

### #3. [project] 매년 내 생일에 서울 기온은 어땠을까?

# [공공 데이터 가져오기 : 서울의 기온]

기상 자료 개방 포털 : <https://data.kma.go.kr/stcs/grnd/grndTaList.do?pgmNo=70>

## 기온분석

그래프

분포도

### ■ 자료설명

지점별로 평균기온과 최저기온, 최고기온의 시계열 분석을 확인합니다.  
일자료, 월자료, 연자료에 대해 각각 조회할 수 있습니다.

\* '지역/지점'의 '지역'은 전국 및 광역 단위의 평균 제공(1973년~)

(전국 및 광역별 평균에 사용된 지점은 전국 평균산출에 사용되는 45개 지점이며, 제주도는 제주시와 서귀포시 자료임)

태그

기온, 기후통계

■ 자료구분 일 ▾

■ 자료형태 기본 ▾

■ 기간 20020101 ~ 20200101

■ 지역/지점 서울

선택

> 검색

CSV 다운로드

Excel 다운로드

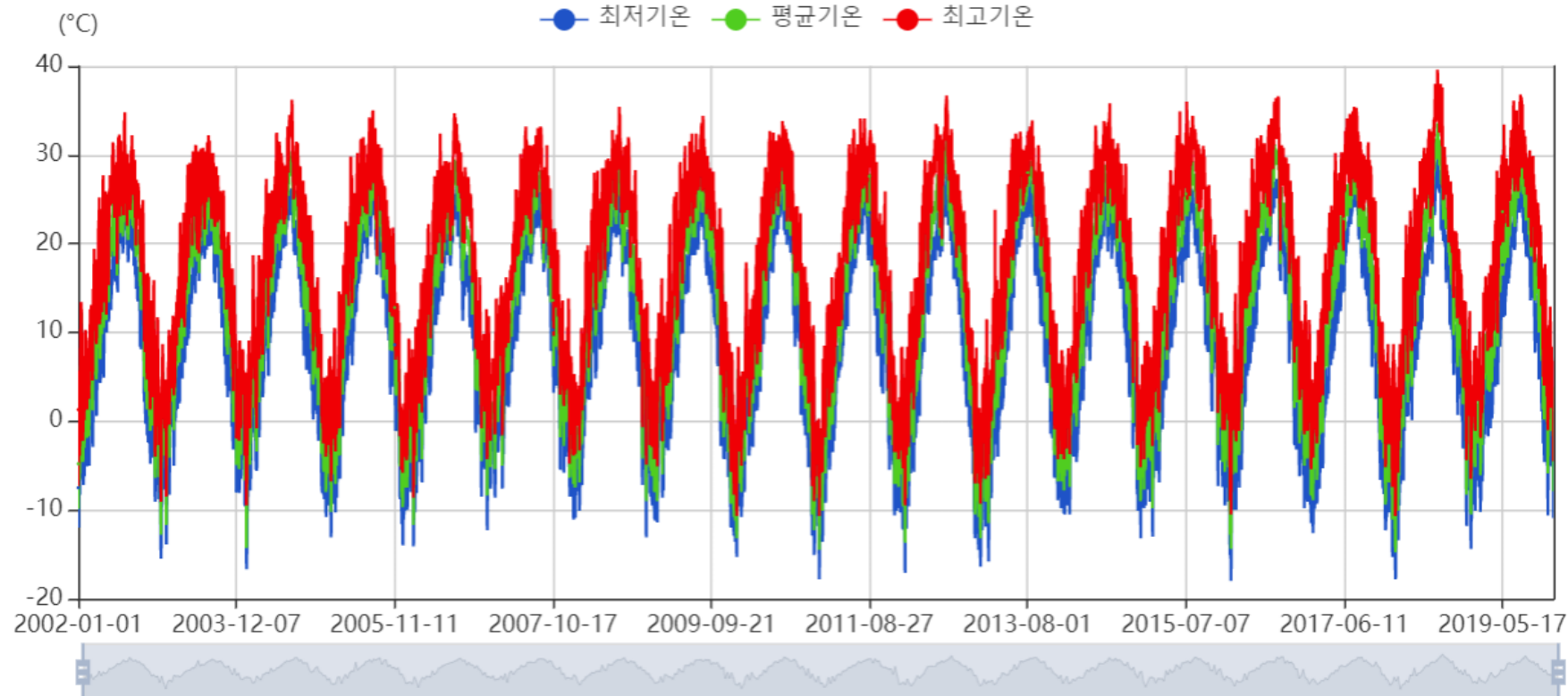
# [공공 데이터 가져오기 : 서울의 기온]

기상 자료 개방 포털 : <https://data.kma.go.kr/stcs/grnd/grndTaList.do?pgmNo=70>

CSV 다운로드

Excel 다운로드

기온분석 기본 서울(108) 일자료 기간 : 20020101 ~ 20200105



# [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날

2002년 01월 01일~현재 날짜(2019년 01월 05일)까지의 기온을 **seoul\_temp.csv**로 저장

seoul\_temp.csv파일을 이용하여

"**서울이 가장 추웠던 날 : 0000-00-00 기온 : 00.0**" 출력해 볼까요?

# [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날

## #1. 데이터 수집 : temp.csv(기상자료이용)

	A	B	C	D	E	
1	기온분석					
2	[검색조건]					
3	자료구분 : 일					
4	자료형태 : 기본					
5	지역/지점 : 서울					
6	기간 : 20020101~20200105					
7						
8	날짜	지점	평균기온(°	최저기온(°	최고기온(°C)	
9	2002-01-01	108	-4.8	-7.5	1.3	
10	2002-01-02	108	-9.9	-11.4	-7.3	
11	2002-01-03	108	-6.5	-12	-2	
12	2002-01-04	108	1.5	-3.4	5.7	
13	2002-01-05	108	-0.7	-3	1.9	
14	2002-01-06	108	0	-5.4	5.2	



# [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날

## #2. 데이터 정제 : 불필요한 자료 삭제 후 seoul\_temp.csv저장

	A	B	C	D	
1	날짜	평균기온(°C)	최저기온(°C)	최고기온(°C)	
2	2002-01-01	-4.8	-7.5	1.3	
3	2002-01-02	-9.9	-11.4	-7.3	
4	2002-01-03	-6.5	-12	-2	
5	2002-01-04	1.5	-3.4	5.7	
6	2002-01-05	-0.7	-3	1.9	
7	2002-01-06	0	-5.4	5.2	
8	2002-01-07	-2.5	-6.9	3.7	
9	2002-01-08	-6.6	-8	-3.9	
10	2002-01-09	-3.5	-8.3	1	
11	2002-01-10	2.2	-1.1	5.7	

## [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날

### #3. 프로그래밍 step1 : 파일 읽어와서 출력해 보기

```
import csv
f=open( './data/seoul_temp.csv', 'r')
rData=csv.reader(f)
for row in rData :
    print(row)
f.close()
```

# [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날 최소값을 찾는 알고리즘을 구현하여 최저기온 찾기

-9.5

-9.1

-10.4

-12.5

-9.8

## [실습1]서울의 기온 데이터 분석 맛보기 : 가장 추웠던 날

```
import csv
f=open( './data/seoul_temp.csv', 'r' )
rData=csv.reader(f)
minTemp=0.0
header=next(rData)
for row in rData :
    if(minTemp>float(row[2])) :
        minTemp=float(row[2])
        minDate=row[0]
print("서울이 가장 추웠던 날 : "+minDate+" 기온 : "+str(minTemp))
f.close()
```

서울이 가장 추웠던 날 : 2016-01-24 기온 : -18.0

## [project완성]매년 내 생일의 평균 기온은 어땠을까?

내가 태어난 해부터 2019년까지의 내 생일의 평균 기온 자료를 출력하고  
평균 기온들의 평균을 구하여 출력해 보세요!

매년 내 생일의 평균 기온 출력

1살	: 2002-11-24	5.7
2살	: 2003-11-24	6.4
3살	: 2004-11-24	9.8
4살	: 2005-11-24	8.4
5살	: 2006-11-24	7.3
6살	: 2007-11-24	6.7
7살	: 2008-11-24	8.3
8살	: 2009-11-24	8.6
9살	: 2010-11-24	6.3
10살	: 2011-11-24	-1.6
11살	: 2012-11-24	1.7
12살	: 2013-11-24	7.3
13살	: 2014-11-24	9.8
14살	: 2015-11-24	5.6
15살	: 2016-11-24	-2.5
16살	: 2017-11-24	-0.4
17살	: 2018-11-24	2.6
18살	: 2019-11-24	9.8

내 생일의 평균 기온들의 평균 값 : 5.54

**내일은,**

**데이터의 시각화 : matplotlib 라이브러리로 그래프 그리기**