

서울의 강수량 데이터 분석하기



소프트웨어학부
진혜진



1. 2020년 서울 강수량 데이터 불러오기

2. 파이썬에서 데이터 출력하기

3. 헤더 저장하기

4. 비가 가장 많이 온 날은 언제였을까?

서울 지역의
2020년 1월 1일부터 3월 31일까지의
강수량을 살펴보자.

■ 기상 자료 개방 포털

■ <https://data.kma.go.kr>



기상자료개방포털

data.kma.go.kr/cmmn/main.do

기상청 기상자료개방포털 보다 나은 정보

국가기후데이터센터 소개 | +가 -가

기상청 날씨데이터 서비스

기상자료개방포털

'관측'을 검색하세요

기상자료개방포털이란?	데이터	기후통계분석	간행물
 <p>평년값</p> <ul style="list-style-type: none">-우리나라 기후평년값-세계기후평년값	<p>통계분석</p> <ul style="list-style-type: none">-조건별통계-기온분석-강수량분석-다중지점통계-전운량 계급별일수-강수 계급별일수-바람장미	<p>기상현상일수</p> <ul style="list-style-type: none">-강수일수-황사일수-폭염일수-열대야일수-24절기-순위값-장마	

1. 2020년 서울 강수량 데이터 불러오기



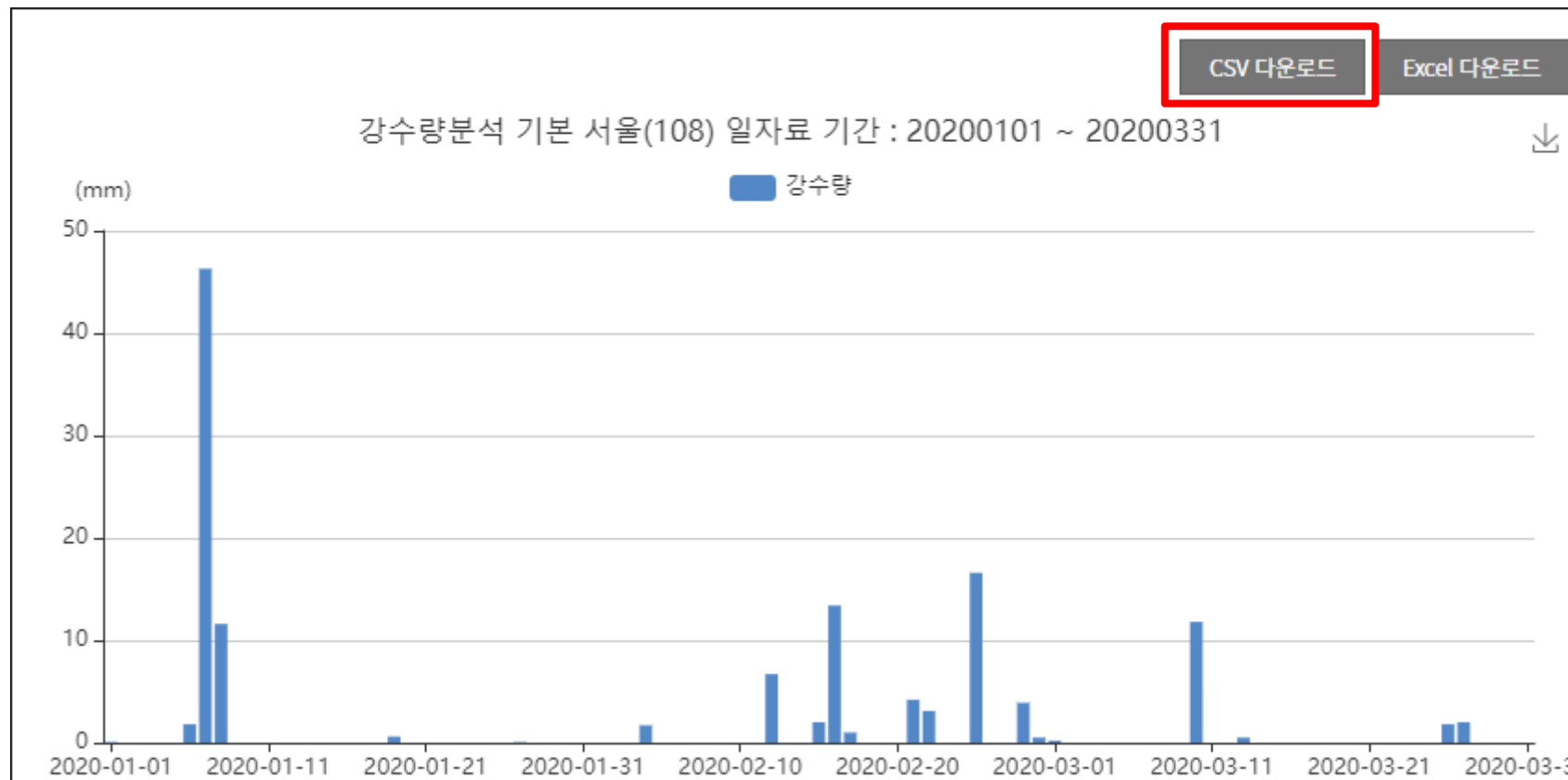
▪ 자료구분 일 ▾

▪ 자료형태 기본 ▾

▪ 기간 20200101 ~ 20200331

▪ 지역/지점 서울 선택

> 검색



1. 2020년 서울 강수량 데이터 불러오기



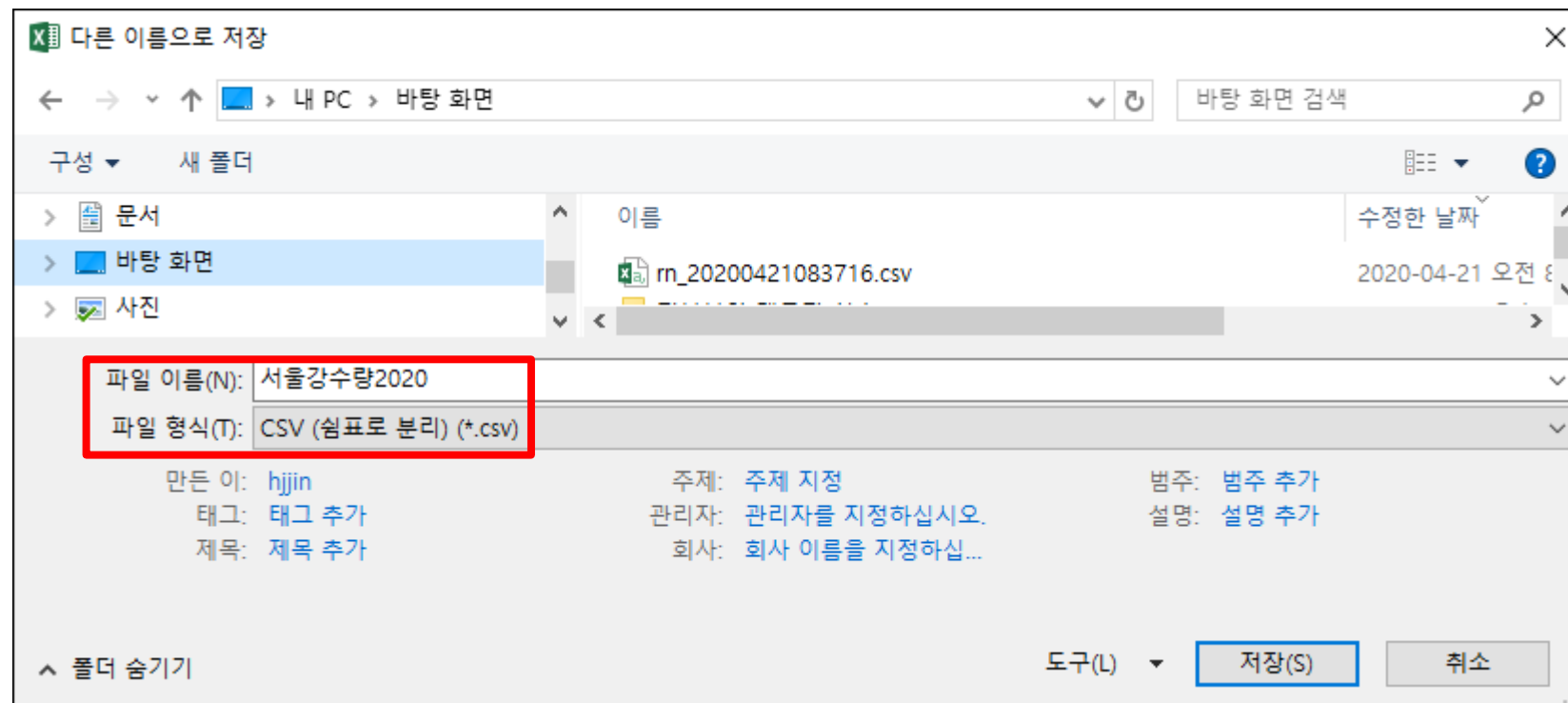
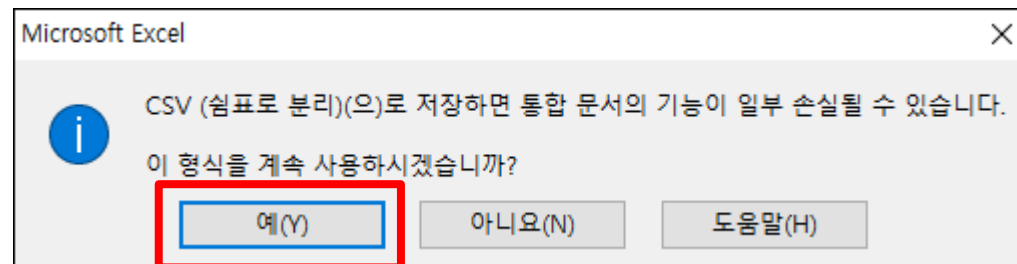
rn_202004...

	A	B	C	D
1	날짜	지점	강수량(mm)	
2	2020-01-01	108	0.1	
3	2020-01-02	108		
4	2020-01-03	108		
5	2020-01-04	108		
6	2020-01-05	108		
7	2020-01-06	108	1.8	
8	2020-01-07	108	46.3	
9	2020-01-08	108	11.6	
10	2020-01-09	108		
11	2020-01-10	108		
12	2020-01-11	108		
13	2020-01-12	108		
14	2020-01-13	108		
15	2020-01-14	108		
16	2020-01-15	108		
17	2020-01-16	108		
18	2020-01-17	108		
19	2020-01-18	108		
20	2020-01-19	108	0.6	
21	2020-01-20	108		

22	2020-01-21	108		
23	2020-01-22	108		
24	2020-01-23	108		
25	2020-01-24	108		
26	2020-01-25	108		
27	2020-01-26	108		
28	2020-01-27	108	0.1	
29	2020-01-28	108	0	
30	2020-01-29	108		
31	2020-01-30	108		
32	2020-01-31	108	0	
33	2020-02-01	108		
34	2020-02-02	108		
35	2020-02-03	108	0	
36	2020-02-04	108	1.7	
37	2020-02-05	108		
38	2020-02-06	108		
39	2020-02-07	108	0	
40	2020-02-08	108	0	
41	2020-02-09	108	0	
42	2020-02-10	108	0	

2020-03-05	108		
2020-03-06	108		
2020-03-07	108		
2020-03-08	108		
2020-03-09	108		
2020-03-10	108	11.8	
2020-03-11	108		
2020-03-12	108		
2020-03-13	108	0.5	
2020-03-14	108		
2020-03-15	108		
2020-03-16	108		
2020-03-17	108		
2020-03-18	108		
2020-03-19	108	0	
2020-03-20	108		
2020-03-21	108		
2020-03-22	108		
2020-03-23	108		
2020-03-24	108		
2020-03-25	108		
2020-03-26	108	1.8	
2020-03-27	108	2	
2020-03-28	108		
2020-03-29	108		
2020-03-30	108		
2020-03-31	108		

1. 2020년 서울 강수량 데이터 불러오기



■ 파일 처리 순서

■ 1. 해당 모듈 불러오기

- `import csv`

■ 2. 파일 열기

- `파일이름 = open('파일명.csv')`

■ 3. 파일 읽기

- `데이터세트이름 = csv.reader(파일이름)`

■ 4. 파일 처리

- `for 변수명 in 데이터세트이름:`
 반복실행문장

■ 5. 파일 닫기

- `파일이름.close()`

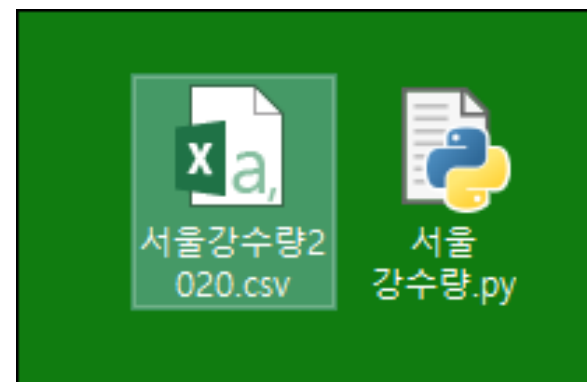

```
import csv

seoul = open('서울강수량2020.csv')

rainfall = csv.reader(seoul)

for i in rainfall:
    print(i)

seoul.close()
```



['날짜', '지점', '강수량(mm)']

['2020-01-01', '108', '0.1']

['2020-01-02', '108', '']

['2020-01-03', '108', '']

['2020-01-04', '108', '']

['2020-01-05', '108', '']

['2020-01-06', '108', '1.8']

['2020-01-07', '108', '46.3']

['2020-01-08', '108', '11.6']

['2020-01-09', '108', '']

['2020-01-10', '108', '']

['2020-01-11', '108', '']

['2020-01-12', '108', '']

['2020-01-13', '108', '']

['2020-01-14', '108', '']

['2020-01-15', '108', '']

['2020-01-16', '108', '']

['2020-01-17', '108', '']

['2020-01-18', '108', '']

['2020-01-19', '108', '0.6']

['2020-01-20', '108', '']

['2020-01-21', '108', '']

['2020-01-22', '108', '']

['2020-01-23', '108', '']

['2020-01-24', '108', '']

['2020-01-25', '108', '']

['2020-01-26', '108', '']

['2020-01-27', '108', '0.1']

['2020-01-28', '108', '0']

['2020-01-29', '108', '']

['2020-01-30', '108', '']

['2020-01-31', '108', '0']

['2020-02-01', '108', '']

['2020-02-02', '108', '']

['2020-02-03', '108', '0']

['2020-02-04', '108', '1.7']

['2020-02-05', '108', '']

['2020-02-06', '108', '']

['2020-02-07', '108', '0']

['2020-02-08', '108', '0']

['2020-02-09', '108', '0']

['2020-02-10', '108', '0']

['2020-02-11', '108', '']

['2020-02-12', '108', '6.7']

['2020-02-13', '108', '']

['2020-02-14', '108', '']

['2020-02-15', '108', '2']

['2020-02-16', '108', '13.4']

['2020-02-17', '108', '1']

['2020-02-18', '108', '']

['2020-02-19', '108', '']

['2020-02-20', '108', '']

['2020-03-06', '108', '']

['2020-03-07', '108', '']

['2020-03-08', '108', '']

['2020-03-09', '108', '']

['2020-03-10', '108', '11.8']

['2020-03-11', '108', '']

['2020-03-12', '108', '']

['2020-03-13', '108', '0.5']

['2020-03-14', '108', '']

['2020-03-15', '108', '']

['2020-03-16', '108', '']

['2020-03-17', '108', '']

['2020-03-18', '108', '']

['2020-03-19', '108', '0']

['2020-03-20', '108', '']

['2020-03-21', '108', '']

['2020-03-22', '108', '']

['2020-03-23', '108', '']

['2020-03-24', '108', '']

['2020-03-25', '108', '']

['2020-03-26', '108', '1.8']

['2020-03-27', '108', '2']

['2020-03-28', '108', '']

['2020-03-29', '108', '']

['2020-03-30', '108', '']

['2020-03-31', '108', '']

■ 헤더

- 데이터 파일에서 여러 가지 값들이 어떤 의미를 갖는지 표시한 행
- 데이터의 첫 번째 줄에 위치해 두 번째 줄부터 나타나는 데이터의 속성을 설명함
- 각 열의 데이터가 어떤 의미를 갖는지 알 수 있기 때문에 중요한 데이터임
- 헤더를 별도로 저장하려면 next() 함수 사용
 - 형식 : next(데이터세트이름)
 - next(rainfall)

```
import csv
```

```
seoul = open('서울강수량2020.csv')
```

```
rainfall = csv.reader(seoul)
```

```
print(next(rainfall))
```

```
seoul.close()
```

```
===== RESTART: C:/Use  
['날짜', '지점', '강수량(mm)']
```

■ next()

- 첫 번째 데이터 행을 읽어오면서 데이터의 탐색 위치를 다음 행으로 이동시키는 명령임
- 두 번째 행부터 출력되는 것을 확인할 수 있음

```
import csv

seoul = open('서울강수량2020.csv')

rainfall = csv.reader(seoul)

next(rainfall)

for i in rainfall:
    print(i)

seoul.close()
```

```
===== RESTART: C:/US
['2020-01-01', '108', '0.1']
['2020-01-02', '108', '']
['2020-01-03', '108', '']
['2020-01-04', '108', '']
['2020-01-05', '108', '']
['2020-01-06', '108', '1.8']
['2020-01-07', '108', '46.3']
['2020-01-08', '108', '14.1']
```