

Chapter 11

데이터 분석 기초

CSV 파일이란?

➤ CSV : Comma-Separated Values

month_temp.csv

지점,일시,평균기온(°C),최저기온(°C),최고기온(°C)

119,2019-10-01,22,15.7,27.4

119,2019-10-02,21.9,20.4,23.8

119,2019-10-03,22.8,19.9,27.8

119,2019-10-04,21.9,17.8,26.9

119,2019-10-05,18.9,15.7,22

...

예제 11-1. csv 파일 읽기

ex11.ipynb

```
import csv

f = open('month_temp.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

for line in lines:
    print(line)

f.close()
```

:: 실행 결과

```
['지점', '일시', '평균기온(°C)', '최저기온(°C)', '최고기온(°C)']
['119', '2019-10-01', '22', '15.7', '27.4']
['119', '2019-10-02', '21.9', '20.4', '23.8']
['119', '2019-10-03', '22.8', '19.9', '27.8']
['119', '2019-10-04', '21.9', '17.8', '26.9']
...
```

예제 11-2. csv 파일 데이터 출력하기

ex11.ipynb

```
import csv

f = open('month_temp.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

for line in lines:
    for x in range(len(line)):
        print(line[x])

f.close()
```

:: 실행 결과

```
지점
일시
평균기온(°C)
최저기온(°C)
최고기온(°C)
119
2019-10-01
...
```

예제 11-3. 특정일자 데이터 출력

ex11.ipynb

```
import csv

f = open('month_temp.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

for line in lines:
    if '2019-10-20' in line:
        print(line)

f.close()
```

:: 실행 결과

```
['119', '2019-10-20', '15.1', '10.8', '21.7']
```

예제 11-4. 데이터 헤더 추출

ex11.ipynb

```
import csv

f = open('month_temp.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

header = next(lines)

print(header)

f.close()
```

:: 실행 결과

['지점', '일시', '평균기온(°C)', '최저기온(°C)', '최고기온(°C)']

예제 11-5. csv 일교차 구하기

ex11.ipynb

```
import csv
f = open('month_temp.csv', 'r', encoding='utf-8')
lines = csv.reader(f)
print('%s %s %s %s' % ('일자', '최저기온', '최고기온', '일교차'))
next(lines)      # 헤더를 건너뛴다.
for line in lines:
    diff = float(line[4]) - float(line[3])
    print('%s %.1f %.1f %.1f' % (line[1], float(line[3]), float(line[4]), diff))
f.close()
```

예제 11-5. csv 일교차 구하기(계속)

:: 실행 결과

일자 최저기온 최고기온 일교차

2019-10-01 15.7 27.4 11.7

2019-10-02 20.4 23.8 3.4

2019-10-03 19.9 27.8 7.9

2019-10-04 17.8 26.9 9.1

2019-10-05 15.7 22.0 6.3

...

예제 11-6. csv 파일 쓰기

ex11.ipynb

```
import csv

file1 = open('month_temp.csv', 'r', encoding='utf-8')
file2 = open('month_temp2.csv', 'w', encoding='utf-8',
newline='')
lines = csv.reader(file1)
wr = csv.writer(file2)

wr.writerow(['일자', '최저기온', '최고기온', '일교차'])
next(lines)
```

```
for line in lines:
    diff = float(line[4]) - float(line[3])
    diff = format(diff, '.1f')
    wr.writerow([line[1], float(line[3]), float(line[4]),
diff])

print('파일 쓰기 완료!')
file1.close()
file2.close()
```

예제 11-6. csv 파일 쓰기(계속)

:: 실행 결과

파일 쓰기 완료!

Q11-1. 데이터 사이에 '/' 삽입하기

:: 실행 결과

```
지점/일시/평균기온(°C)/최저기온(°C)/최고기온(°C)/  
119/2019-10-01/22/15.7/27.4/  
119/2019-10-02/21.9/20.4/23.8/  
119/2019-10-03/22.8/19.9/27.8/  
119/2019-10-04/21.9/17.8/26.9/  
...
```

```
import csv  
  
f = open('month_temp.csv', ❶_____, encoding='utf-8')  
lines = csv.reader(f)  
  
for line in lines:  
    for ❷_____ in line:  
        print('%s' % x, ❸_____)  
    print()  
f.close()
```

Q11-2. 10일간 평균 기온 구하기

:: 실행 결과

10일간 평균 기온 : 61.0

```
import csv
f = open('month_temp.csv', 'r', encoding='utf-8')
lines = csv.reader(f)
next(lines)      # 헤더를 건너뛴다.
sum = 0
for line in lines:
    if (int(❶_____ ) <= 10) :
        sum += float(❷_____)
avg = ❸_____/3
print('10일간 평균 기온 : %.1f' % ❹_____)
f.close()
```

예제 11-7. '경주시'의 '신대원약국' 주소

ex11.ipynb

```
import csv

f = open('pharm_2019.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

header = next(lines)

for line in lines:

    if line[1] == '경주시' and line[0] == '신대원약국':

        print(line[0], line[1], line[2], sep='/')

f.close()
```

:: 실행 결과

신대원약국/경주시/경상북도 경주시 화랑로
53 (성건동)

예제 11-8. '수지'의 '로얄스포츠' 건물 약국

ex11.ipynb

```
import csv

f = open('pharm_2019.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

header = next(lines)

for line in lines:

    if ('수지' in line[2]) and ('로얄스포츠' in line[2]):

        print(line[2], line[0], line[3], sep='/')

f.close()
```

예제 11-8. '수지'의 '로얄스포츠' 건물 약국(계속)

:: 실행 결과

경기도 용인시 수지구 풍덕천로 119 118호 (풍덕천동, 로얄스포츠센터)/희망약국/20170216

경기도 용인시 수지구 풍덕천로 119 119호 (풍덕천동, 로얄스포츠센터)/로얄약국/20170717

경기도 용인시 수지구 풍덕천로 119 305호 (풍덕천동, 로얄스포츠센터)/이화옵티마약국
/20170314

경기도 용인시 수지구 풍덕천로 119 208호 (풍덕천동, 로얄스포츠센터)/수지윤약국/20170103

예제 11-9. 최근 5년 이내 개설 약국 수

ex11.ipynb

```
import csv

f = open('pharm_2019.csv', 'r', encoding='utf-8')
lines = csv.reader(f)
header = next(lines)

city = '원주시'

total = 0
recent = 0
```

```
for line in lines:
    if line[1] == city :
        total += 1
        if int(line[3]) > 20140901 :
            recent += 1

print('%s의 약국 수 : %d개' % (city, total))
print('5년 이내 개설된 약국 수 : %d개' % recent)
```


예제 11-9. 최근 5년 이내 개설 약국 수(계속)

:: 실행 결과

원주시의 약국 수 : 1092개
5년 이내 개설된 약국 수 : 55개

예제 11-10. 2010년에 개설된 약국 목록

ex11.ipynb

```
import csv
```

```
f = open('pharm_2019.csv', 'r', encoding='utf-8')
```

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

```
header = next(lines)
```

```
print('번호', header[1], header[0], header[4],
```

```
header[5], sep=',')
```

```
number = 1
```

```
for line in lines:
```

```
    if ('경상북도' in line[2]) and
```

```
    (int(line[3])>=20100101 and int(line[3])<=20101231):
```

```
        print(number, '.', sep='', end=' ')
```

```
        print(line[1], line[0], line[4], line[5], sep='/')
```

```
        number += 1
```

```
f.close()
```

예제 11-10. 2010년에 개설된 약국 목록(계속)

:: 실행 결과

번호,지역,약국명,경도,위도

1. 포항북구/영천온누리약국/129.3836663/36.0824353
2. 포항북구/늘사랑약국/129.3754548/36.0597963
3. 구미시/형곡비타민약국/128.3398692/36.1143363
4. 영덕군/열린약국/129.4431662/36.5071346
5. 포항북구/아이원약국/

...

예제 11-11. 오름차순(가나다순) 정렬

ex11.ipynb

```
import csv

f = open('pharm_2019.csv', 'r', encoding='utf-8')
lines = csv.reader(f)

header = next(lines)
print('번호', header[0], header[2], sep=',')

data = []
```

```
for line in lines:
    if line[1] == '용인수지구' :
        tmp = '%s/%s' % (line[0], line[2])
        data.append(tmp)

data.sort()

number = 1

for x in data :
    print('%d. %s' % (number, x))
    number += 1

f.close()
```

예제 11-11. 오름차순(가나다순) 정렬(계속)

:: 실행 결과

번호,약국명,주소

1. 5층우리약국/경기도 용인시 수지구 풍덕천로 149 501호 (풍덕천동, 현대그린프라자)
2. LG(엘지)약국/경기도 용인시 수지구 성북2로 84 105호 (성북동, 엘지빌리지1차상가)
3. 가람약국/경기도 용인시 수지구 만현로 88 (상현동, 상현종합상가201-2호)
4. 광고 기분좋은 약국/경기도 용인시 수지구 광고마을로 54 에스비타운 110호 (상현동)
5. 광고수약국/경기도 용인시 수지구 광고중앙로 320 (상현동)

...

제주도 기상 데이터 구조

Jeju_2019.csv

지점,지점명,일시,최저기온(°C),최고기온(°C),일강수량(mm),평균 상대습도(%)

184,제주,2019-01-01,4.8,7,0,61.1

184,제주,2019-01-02,4.5,6.1,,60.9

184,제주,2019-01-03,3.5,8,,60.1

184,제주,2019-01-04,2.6,10,1.6,77.5

184,제주,2019-01-05,5.2,9,0,66.8

184,제주,2019-01-06,4.5,7.7,,64.4

...

예제 11-12. 서귀포 최저기온 출력

ex11.ipynb

```
import csv

f = open('jeju_2019.csv', 'r', encoding='utf-8')
lines = csv.reader(f)

header = next(lines)
print(header[1], header[2], header[3])
```

```
for line in lines:
    if line[1] == '서귀포' and line[2][5:7] == '01':
        print(line[1], line[2], line[3])

f.close()
```

예제 11-12. 서귀포 최저기온 출력(계속)

:: 실행 결과

지점명 일시 최저기온(°C)

서귀포 2019-01-01 3.8

서귀포 2019-01-02 3.6

서귀포 2019-01-03 2.6

서귀포 2019-01-04 6.6

서귀포 2019-01-05 7.2

서귀포 2019-01-06 7.2

...

예제 11-13. 서귀포 최저기온 평균

ex11.ipynb

```
import csv

f = open('jeju_2019.csv', 'r', encoding='utf-8')

lines = csv.reader(f)
header = next(lines)

month = 1  # 기준 월

sum = 0

num_day = 0
```

```
for line in lines:

    if line[1] == '서귀포' and int(line[2][5:7]) == month :

        sum += float(line[3])

        num_day += 1

avg = sum/num_day

print('%d월 일수 : %d' % (month, num_day))

print('%d월 최저기온 평균 : %.1f' % (month, avg))

f.close()
```

예제 11-13. 서귀포 최저기온 평균(계속)

:: 실행 결과

1월 일수 : 31

1월 최저기온 평균 : 4.6

예제 11-14. 지역별 8월 중 최고기온

ex11.ipynb

```
import csv
```

```
f = open('jeju_2019.csv', 'r', encoding='utf-8')
```

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

```
header = next(lines)
```

```
month = 8 # 기준 월
```

```
max_jeju = -1000
```

```
max_sungsan = -1000
```

```
max_gosan = -1000
```

```
max_suguipo = -1000
```

예제 11-14. 지역별 8월 중 최고기온(계속)

```
for line in lines:
    if line[1] == '제주' and int(line[2][5:7]) == month :
        if float(line[4]) > max_jeju :
            max_jeju = float(line[4])

    if line[1] == '고산' and int(line[2][5:7]) == month :
        if float(line[4]) > max_gosan :
            max_gosan = float(line[4])
```

```
if line[1] == '성산' and int(line[2][5:7]) == month :
    if float(line[4]) > max_sungsan :
        max_sungsan = float(line[4])

if line[1] == '서귀포' and int(line[2][5:7]) == month :
    if float(line[4]) > max_suguipo :
        max_suguipo = float(line[4])
```

예제 11-14. 지역별 8월 중 최고기온(계속)

```
print('%d월 제주 최고기온 : %.1f' % (month, max_jeju))
print('%d월 고산 최고기온 : %.1f' % (month, max_gosan))
print('%d월 성산 최고기온 : %.1f' % (month, max_sungsan))
print('%d월 서귀포 최고기온 : %.1f' % (month, max_suguipo))

f.close()
```

예제 11-14. 지역별 8월 중 최고기온(계속)

:: 실행 결과

8월 제주 최고기온 : 34.7
8월 고산 최고기온 : 32.9
8월 성산 최고기온 : 32.5
8월 서귀포 최고기온 : 32.2

예제 11-15. 가장 비가 많이 오는 월

ex11.ipynb

```
import csv

f = open('jeju_2019.csv', 'r', encoding='utf-8')

lines = csv.reader(f)

header = next(lines)

sum_rain = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

for line in lines:
    if not line[5]:
        line[5] = 0
```

```
if int(line[2][5:7]) == 1 : # 1월 이면
    sum_rain[0] += float(line[5])
if int(line[2][5:7]) == 2 : # 2월 이면
    sum_rain[1] += float(line[5])
if int(line[2][5:7]) == 3 : # 3월 이면
    sum_rain[2] += float(line[5])
if int(line[2][5:7]) == 4 : # 4월 이면
    sum_rain[3] += float(line[5])
if int(line[2][5:7]) == 5 : # 5월 이면
    sum_rain[4] += float(line[5])
```

예제 11-15. 가장 비가 많이 오는 월(계속)

```
if int(line[2][5:7]) == 6 : # 6월 이면
    sum_rain[5] += float(line[5])
if int(line[2][5:7]) == 7 : # 7월 이면
    sum_rain[6] += float(line[5])
if int(line[2][5:7]) == 8 : # 8월 이면
    sum_rain[7] += float(line[5])
if int(line[2][5:7]) == 9 : # 9월 이면
    sum_rain[8] += float(line[5])
```

```
if int(line[2][5:7]) == 10 : # 10월 이면
    sum_rain[9] += float(line[5])
if int(line[2][5:7]) == 11 : # 11월 이면
    sum_rain[10] += float(line[5])
if int(line[2][5:7]) == 12 : # 12월 이면
    sum_rain[11] += float(line[5])
```


예제 11-15. 가장 비가 많이 오는 월(계속)

```
max_month_rain = max(sum_rain)
max_month = sum_rain.index(max_month_rain) + 1
print('(1) 최대 강수 월과 강수량 : %d월, %.1f mm\n' % (max_month, max_month_rain))
print('(2) 월별 강수량')

for i in range(1, 13):
    print('%d월 : %.1f mm' % (i, sum_rain[i-1]))

f.close()
```

예제 11-15. 가장 비가 많이 오는 월(계속)

:: 실행 결과

(1) 최대 강수 월과 강수량 : 9월, 2018.9 mm

(2) 월별 강수량

1월 : 79.5 mm

2월 : 187.1 mm

3월 : 299.8 mm

4월 : 296.0 mm

...

11월 : 111.1 mm

12월 : 320.6 mm

예제 11-16. 고산 7월 최저/최고 습도

ex11.ipynb

```
import csv

f = open('jeju_2019.csv', 'r', encoding='utf-8')

lines = csv.reader(f)
header = next(lines)

month = 7 # 기준 월
min_humidity = 1000
max_humidity = -1000
```

```
for line in lines:
    if line[1] == '고산' and int(line[2][5:7]) == month :
        if float(line[6]) < min_humidity :
            min_humidity = float(line[6])

        if float(line[6]) > max_humidity :
            max_humidity = float(line[6])
```

예제 11-16. 고산 7월 최저/최고 습도(계속)

```
print('%d월 최저 습도 : %.1f' % (month, min_humidity))  
print('%d월 최대 습도 : %.1f' % (month, max_humidity))  
  
f.close()
```

예제 11-16. 고산 7월 최저/최고 습도(계속)

:: 실행 결과

7월 최저 습도 : 78.5
7월 최대 습도 : 100.0

예제 11-17. 강수량 최대 지역

ex11.ipynb

```
import csv
f = open('jeju_2019.csv', 'r', encoding='utf-8')
lines = csv.reader(f)
header = next(lines)
total_rain = [0, 0, 0, 0]

for line in lines:
    if not line[5]:
        line[5] = 0
```

```
if line[1] == '제주': # '제주' 지역
    total_rain[0] += float(line[5])
if line[1] == '고산': # '고산' 지역
    total_rain[1] += float(line[5])
if line[1] == '성산': # '성산' 지역
    total_rain[2] += float(line[5])
if line[1] == '서귀포': # '서귀포' 지역
    total_rain[3] += float(line[5])
```

예제 11-17. 강수량 최대 지역(계속)

```
max_year_rain = max(total_rain)
if total_rain.index(max_year_rain) == 0 :
    max_area = '제주'
if total_rain.index(max_year_rain) == 1 :
    max_area = '고산'
if total_rain.index(max_year_rain) == 2 :
    max_area = '성산'
if total_rain.index(max_year_rain) == 3 :
    max_area = '서귀포'
```

```
print('(1) 연 강수 최대 지역: %s\n' % max_area)
print('(2) 지역별 강수량')
print('제주 : %.1f mm' % total_rain[0])
print('고산 : %.1f mm' % total_rain[1])
print('성산 : %.1f mm' % total_rain[2])
print('서귀포 : %.1f mm' % total_rain[3])

f.close()
```

예제 11-17. 강수량 최대 지역(계속)

:: 실행 결과

(1) 연 강수 최대 지역: 성산

(2) 지역별 강수량

제주 : 1979.9 mm

고산 : 1560.9 mm

성산 : 2658.1 mm

서귀포 : 2210.3 mm