

[12장 과제]

맑은 고딕 12 ☞ 학번 : 20203103 이름 : 임정민

실습 12-1. (p309 본문 참조) csv 모듈을 이용하여 'd:/data/weather.csv' 파일에 있는 데이터 중 처음부터 10개의 데이터만 출력하시오.

[결과 화면]

```
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1> & "I
```

12-1. 20203103 임정민

```
['일시', '평균기온', '최대풍속', '평균풍속']
['2010-08-01', '28.7', '8.3', '3.4']
['2010-08-02', '25.2', '8.7', '3.8']
['2010-08-03', '22.1', '6.3', '2.9']
['2010-08-04', '25.3', '6.6', '4.2']
['2010-08-05', '27.2', '9.1', '5.6']
['2010-08-06', '26.8', '9.8', '8']
['2010-08-07', '27.5', '9.1', '5']
['2010-08-08', '26.6', '5.9', '4']
['2010-08-09', '26.9', '5.1', '3.1']
['2010-08-10', '25.6', '10.2', '5.5']
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1> █
```

[프로그램 소스]

```
print('\n12-1. 20203103 임정민 \n')
import csv
f = open('bin/weather.csv') ##D 파티션이 없는 관계로 임시 폴더에 넣어 실행하였습니다.
data = csv.reader(f)
i=11
for row in data:
    print(row)
    i=i-1
    if(i==0) :
        break
f.close()
```

실습 12-2. (p310 본문 변경) csv 모듈을 이용하여 'd:/data/weather.csv' 파일에서 지난 10년간 울릉도의 최소 온도를 구하시오.

[결과 화면]

```
12-2. 20102000 Hong홍길동
지난 10년간 울릉도의 최소 온도는 -9.0 도
>>> |
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1> & "

12-2. 20203103 임정민

지난 10년간 울릉도의 최소 온도는 -9.0 도
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1> █
```

[프로그램 소스]

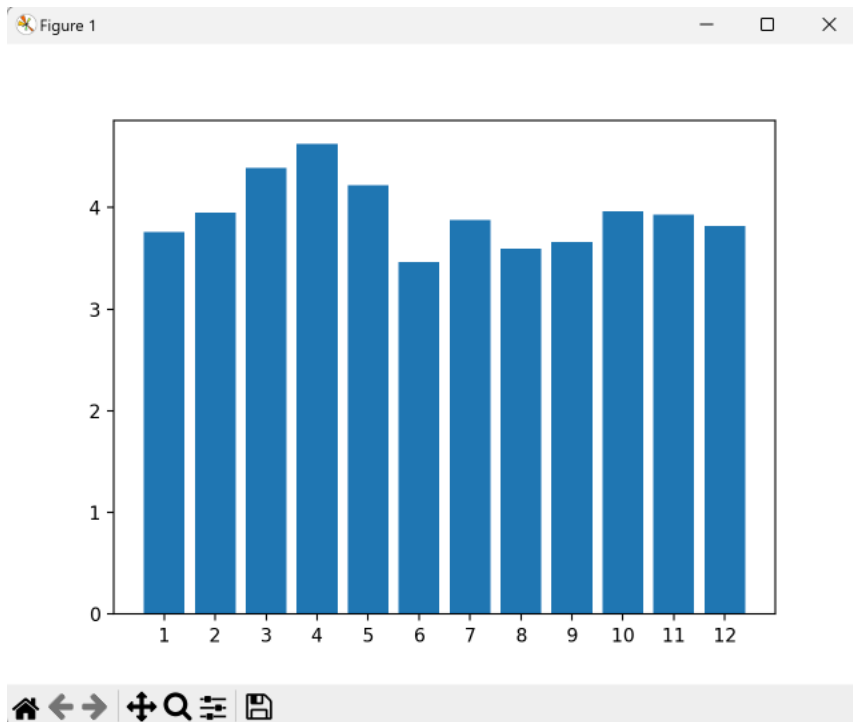
```
print('\n12-2. 20203103 임정민 \n')
```

```
import csv
f = open('bin/weather.csv')
data = csv.reader(f)
header = next(data)

min_temp = 0.0
for row in data:
    if row[1] == '' :
        temp = 100
    else :
        temp = float(row[1])
        if min_temp > temp :
            min_temp = temp
print('지난 10 년간 울릉도의 최소 온도는 ', min_temp, '도')
```

실습 12-3. (p311 LAB12-1 변경) 10년간 울릉도의 월별 평균 풍속을 막대그래프로 나타내시오.

[결과 화면]



[프로그램 소스]

```
import csv
import matplotlib.pyplot as plt
print('\n12-3. 20203103 임정민 \n')

f = open('bin/weather.csv')
data = csv.reader(f)
header = next(data)
monthly_wind = [ 0 for x in range(12) ]
days_counted = [ 0 for x in range(12) ]
for row in data:
    month = int(row[0][5:7])
    if row[3] != '':
        wind = float(row[3])
        monthly_wind[month-1] += wind
        days_counted[month-1] += 1
for i in range(12) :
    monthly_wind[i] /= days_counted[i]

print(monthly_wind)
plt.bar(range(len(monthly_wind)), monthly_wind)
plt.xticks(range(0,12), range(1,13))
plt.show()
f.close()
```

실습 12-4. (p315, p318 본문 참조) countries.csv에 있는 데이터 파일로부터 처음 행의 개수 5개, 마지막 행의 개수 5개, 3~5행까지 3개 출력하기

[결과 화면]

12-4. 20203103 임정민

	country	area	capital	population
KR	Korea	98480	Seoul	51780579
US	USA	9629091	Washington	331002825
JP	Japan	377835	Tokyo	125960000
CN	China	9596960	Beijing	1439323688
RU	Russia	17100000	Moscow	146748600

	country	area	capital	population
KR	Korea	98480	Seoul	51780579
US	USA	9629091	Washington	331002825
JP	Japan	377835	Tokyo	125960000
CN	China	9596960	Beijing	1439323688
RU	Russia	17100000	Moscow	146748600

	country	area	capital	population
CN	China	9596960	Beijing	1439323688
RU	Russia	17100000	Moscow	146748600

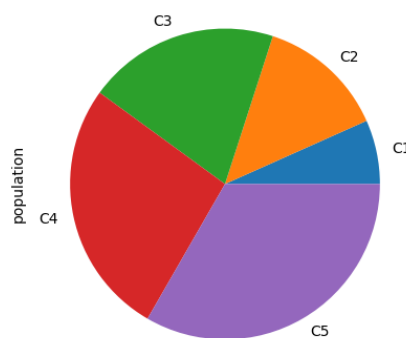
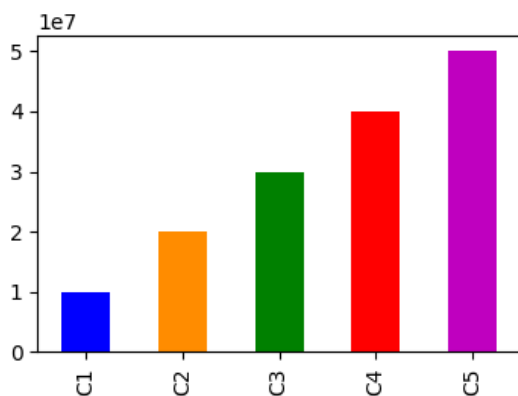
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1>

[프로그램 소스]

```
print('\n12-4. 20203103 임정민 \n')
import pandas as pd
df = pd.read_csv('bin\countries.csv', index_col = 0)
print(df.head())
print('\n', df.tail())
print('\n', df[3:6]) # 3, 4, 5
```

실습 12-5. (p.317 본문 참조) C1, C2, C3, C4, C5 국가의 인구를 [결과 화면]처럼 막대그래프와 원그래프로 나타내시오.

[결과 화면]



[프로그램 소스]

```
print('\n12-5. 20203103 임정민 \n')
import pandas as pd
import matplotlib.pyplot as plt
countries_df = pd.read_csv('bin\countries.csv', index_col = 0)
```

```
countries_df[5:10]['population'].plot(kind='bar', color=('b', 'darkorange', 'g',
'r', 'm'))
#countries_df[5:10]['population'].plot(kind='pie')
plt.show()
```

실습 12-6. (p.319 본문 참조) 인구밀도를 나타내는 열(density)을 추가하고, 처음부터 7개의 데이터를 출력 하시오.

[결과 화면]

```
12-6. 20203103임정민
```

	country	area	capital	population	density
KR	Korea	98480	Seoul	51780579	525.797918
US	USA	9629091	Washington	331002825	34.375293
JP	Japan	377835	Tokyo	125960000	333.373033
CN	China	9596960	Beijing	1439323688	149.977044
RU	Russia	17100000	Moscow	146748600	8.581789

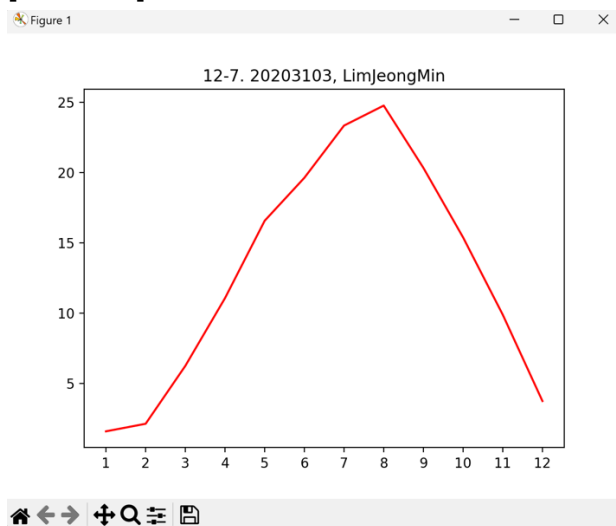
```
PS Microsoft.PowerShell.Core\FileSystem::\mac\Home\Documents\codes\python3.8\test1>
```

[프로그램 소스]

```
print('\n12-6. 20203103 임정민 \n')
import pandas as pd
import matplotlib.pyplot as plt
countries_df = pd.read_csv('bin\countries.csv', index_col = 0)
countries_df['density'] = countries_df['population'] / countries_df['area']
print(countries_df[0:7])
```

실습 12-7. (p322 LAB12-2 응용) 10년간 울릉도의 월별 평균기온을 실선그래프로 나타내시오.

[결과 화면]



[프로그램 소스]

```
print('\n12-7. 20203103 임정민 \n')
import pandas as pd
import matplotlib.pyplot as plt
weather = pd.read_csv('bin\weather.csv', encoding='CP949')
monthly = [ None for x in range(12) ]
```

```

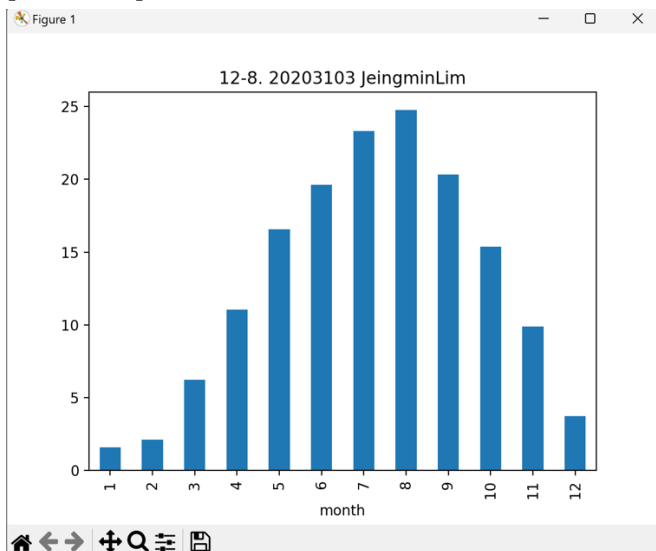
monthly_temp = [ 0 for x in range(12) ]
weather['month'] = pd.DatetimeIndex(weather['일시']).month

for i in range(12) :
    monthly[i] = weather[ weather['month'] == i + 1 ]
    # old version 에서 가능 # monthly_temp[i] = monthly[i].mean()['평균기온']
    monthly_temp[i] = monthly[i]['평균기온'].mean()
plt.title('12-7. 20203103, LimJeongMin')
plt.plot(monthly_temp, 'red')
plt.xticks(range(0,12), range(1,13))
plt.show()

```

실습 12-8. (p324 LAB12-3 응용) group by를 이용하여 10년간 울릉도의 월별 평균기온을 막대그래프로 나타내시오.

[결과 화면]



[프로그램 소스]

```

print('\n12-8. 20203103 임정민 \n')
import pandas as pd
import matplotlib.pyplot as plt
import datetime as dt
weather = pd.read_csv('bin\weather.csv', encoding='CP949')
weather['month'] = pd.DatetimeIndex(weather['일시']).month
# old version 에서 가능 # means = weather.groupby('month').mean()
means = weather[['평균기온', 'month']].groupby('month').mean()
means['평균기온'].plot(kind='bar')
plt.title('12-8. 20203103 JeingminLim')
plt.show()

```

실습 12-9. (p326-327 이용) [결과 화면]과 같이 다음의 내용을 출력하시오.

(1) 결손값 출력하기 (2) 결손값을 0으로 채운 후 출력하기 (3) 결손 값을 평균으로 채운 후 출력하기.

[결과 화면]

```
      평균기온  최대풍속  평균풍속
일시
2012-02-11  -0.7    NaN    NaN
2012-02-12   0.4    NaN    NaN
2012-02-13   4.0    NaN    NaN
2015-03-22  10.1   11.6    NaN
2015-04-01   7.3   12.1    NaN
2019-04-18  15.7   11.7    NaN

평균기온  -0.7
최대풍속   0.0
평균풍속   0.0
Name: 2012-02-11, dtype: float64

평균기온  -0.700000
최대풍속   3.936441
평균풍속   3.936441
Name: 2012-02-11, dtype: float64
PS Microsoft.PowerShell.Core\FileSystem::\\mac\Home\Documents\codes\python3.8\test1>
```

[프로그램 소스]

```
print('\n12-9. 20203103 임정민 \n')
import pandas as pd
dir = 'bin\weather.csv'
weather = pd.read_csv(dir, index_col = 0, encoding='CP949')
missing_data = weather [ weather['평균풍속'].isna() ]
print(missing_data)

weather = pd.read_csv(dir, index_col = 0, encoding='CP949')
weather.fillna(0, inplace = True)
print()
print(weather.loc['2012-02-11'])

weather = pd.read_csv(dir, index_col = 0, encoding='CP949')
weather.fillna( weather['평균풍속'].mean(), inplace = True)
print()
print(weather.loc['2012-02-11'])
```

실습 12-10. (본문 p.334 응용) countries.csv 데이터를 국토면적(area) 순으로 큰 값 5개 국가의 내용만 출력하시오

[결과 화면]

```
12-10. 20203103임정민

   country  area  capital  population
RU  Russia 17100000  Moscow  146748600
US   USA   9629091  Washington 331002825
CN  China  9596960   Beijing  1439323688
JP  Japan  377835   Tokyo    125960000
KR  Korea   98480   Seoul    51780579
PS Microsoft.PowerShell.Core\FileSystem::\\mac\Home\Docum
```

[프로그램 소스]

```
print('\n12-10. 20203103 임정민 \n')
import pandas as pd
import matplotlib.pyplot as plt
countries_df = pd.read_csv('bin\countries.csv', index_col = 0)
sorted = countries_df.sort_values('area', ascending=False)
print(sorted[0:5])
```