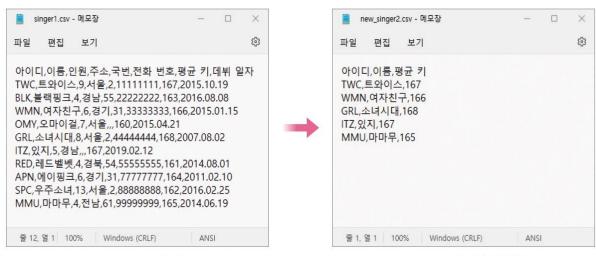
CSV 파일 처리

내용

- CSV 파일의 구조에 대해서 이해한다.
- CSV 텍스트 파일에 내용을 읽거나 쓰는 기능을 배운다.
- CSV 라이브러리를 이용해서 읽거나 쓰는 기능을 배운다.
- GUI 환경을 활용해서 CSV 파일을 엑셀처럼 표현하는 방식을 알아본다.

Section 01 이 장에서 만들 프로그램

- ┗ [프로그램 1] CSV 파일 복사
 - 파이썬의 기본적인 파일 처리 방법을 사용해서 CSV 파일을 다른 CSV 파일로 복사하는 프로그램



CSV 파일

CSV 파일(변경)

Section 01 이 장에서 만들 프로그램

- ┗ [프로그램 2] CSV 파일 출력
 - CSV 파일을 읽어서 엑셀처럼 GUI로 출력하는 프로그램
 - 기능 구현은 csv 라이브러리를 활용하고, GUI 화면은 tkinter를 사용



CSV 파일

GUI 화면 출력

CSV 파일 개념

- CSV 파일은 Comma Separated Values의 약자로 '쉼표로 구분된 값'
- 숫자나 문자열로 구성된 표를 단순한 텍스트 형식으로 저장한 것
- 텍스트 파일처럼 메모장에서 생성, 읽기, 쓰기가 가능

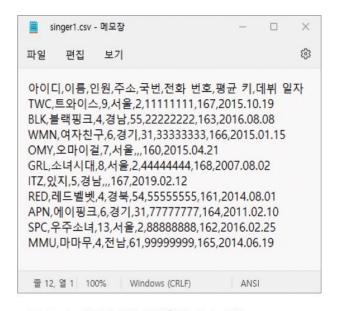


그림 6-1 메모장에서 생성한 CSV 파일

CSV 파일 개념

- [그림 6-1]의 CSV 파일은 [표 6-1]과 같은 형태의 데이터를 입력한 것
- 첫 행은 각 열의 값의 이름을 표현한 헤더(Header), 두 번째 행부터는 데이터 값

표 6-1 CSV로 생성할 표 데이터

이이디	이름	인원	주소	국번	전화 번호	평균 키	데뷔 일자
TWC	트와이스	9	서울	02	11111111	167	2015,10,19
BLK	블랙핑크	4	경남	055	2222222	163	2016.08.08
WMN	여자친구	6	경기	031	33333333	166	2015.01.15
OMY	오마이걸	7	서울			160	2015,04,21
GRL	소녀시대	8	서울	02	4444444	168	2007,08,02
ITZ	있지	5	경남			167	2019,02,12
RED	레드벨벳	4	경북	054	5555555	161	2014,08,01
APN	에이핑크	6	경기	031	77777777	164	2011,02,10
SPC	우주소녀	13	서울	02	88888888	162	2016,02,25
MMU	마마무	4	전남	061	99999999	165	2014,06,19

- CSV 파일 개념

- CSV 파일을 메모장에서 직접 생성하거나 편집하는 것은 불편하고, 쉼표로 구분되어 있기 때문에 가독성이 떨어짐
- CSV를 직접 입력하여 생성할 때는 엑셀(Excel)을 사용하는 것이 편리함

CSV 파일 생성

- 엑셀을 실행해서 [표 6-1]을 다음과 같이 입력
- 국번 열의 경우 앞에 숫자 0이 들어가야 하므로 작은따옴표를 입력해야 한다는 점을 주의
 (엑셀에서는 기본적으로 02를 입력해도 2로 저장되기 때문)

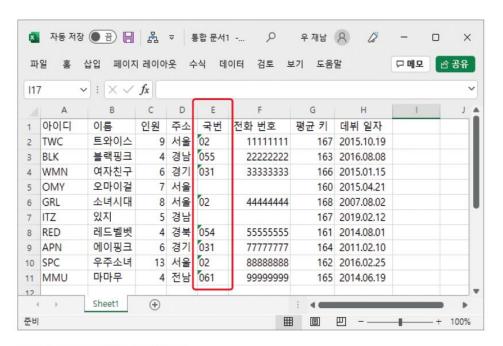


그림 6-2 엑셀에서 데이터 입력

CSV 파일 생성

■ 입력을 완료하면 파일 형식을 CSV로 선택한 후 C:\CookAnalysis\CSV\singer1.csv로 저장 (폴더가 없다면 폴더를 먼저 생성)

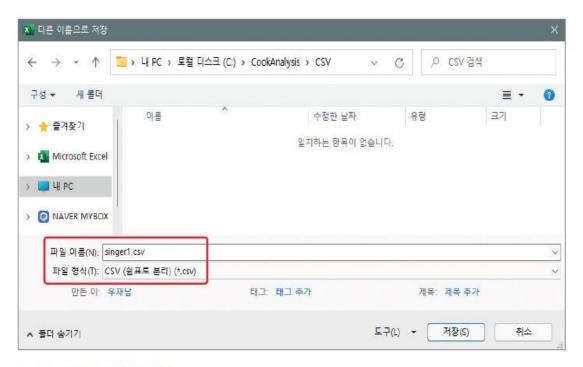
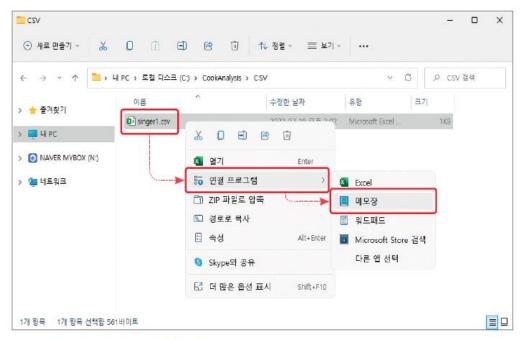


그림 6-3 CSV 파일로 저장

CSV 파일 생성

- 저장된 파일을 메모장에서 읽을 수 있음
- [파일 탐색기]에서 C:\CookAnalysis\CSV\singer1.csv 파일을 선택
- 마우스 오른쪽 버튼을 클릭 -> [연결 프로그램]-[메모장]을 선택





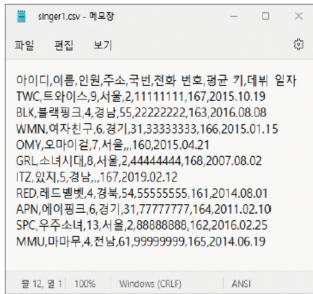


그림 6-5 메모장에서 열린 singer1.csv 파일

CSV 파일 처리

5장에서 학습한 텍스트 파일을 처리하는 방식으로 CSV 파일도 처리할 수 있음

```
Code06-01.py
     inFp = open("C:/CookAnalysis/CSV/singer1.csv", "r")
02
03
     inStr = inFp.readline()
     print(inStr, end = "")
04
05
06
     inStr = inFp.readline()
                                     실행 결과
07
     print(inStr, end = "")
                                    아이디, 이름, 인원, 주소, 국번, 전화 번호, 평균 키, 데뷔 일자
80
     inFp.close()
                                    TWC, 트와이스, 9, 서울, 02, 11111111, 167, 2015.10.19
09
```

CSV 파일 처리

with 예약어를 사용하면 Code06-01.py 9행의 파일을 닫는 과정이 생략되므로 편리함

```
Code06-02.py
    with open("C:/CookAnalysis/CSV/singer1.csv", "r") as inFp:
       inStr = inFp.readline()
       print(inStr, end = "")
       inStr = inFp.readline()
       print(inStr, end = "")
```

CSV 파일 처리

- CSV는 한 줄을 모두 처리하는 것이 아니라 각 데이터를 처리해야 의미가 있음
- 즉, TWC, 트와이스, 9 ... 등을 리스트에 분리해서 저장해야 이후 계산이 가능함

CSV 파일 처리

 CSV 파일의 헤더를 별도로 먼저 읽어서 처리한 후 나머지 모든 행을 리스트로 저장하고 각 항목을 분리해서 출력하는 코드

```
Code06-03.py
          def printList(pList) :
01
02
             for data in pList:
03
                print(data, end='\thickstr')
04
             print()
05
06
          with open("C:/CookAnalysis/CSV/singer1.csv", "r") as inFp:
07
             header = inFp.readline()
80
             header = header.strip()
             header list = header.split(',')
09
10
             printList(header_list)
11
             for inStr in inFp:
                inStr = inStr.strip()
13
                row_list = inStr.split(',')
                printList(row list)
14
```

CSV 파일 처리

 CSV 파일의 헤더를 별도로 먼저 읽어서 처리한 후 나머지 모든 행을 리스트로 저장하고 각 항목을 분리해서 출력하는 코드

실행 결과	ł						
아이디	이름	인원	주소	국번	전화 번호	평균 키	데뷔 일자
TWC	트와이스	9	서울	02	11111111	167	2015.10.19
BLK	블랙핑크	4	경남	055	2222222	163	2016.08.08
WMN	여자친구	6	경기	031	33333333	166	2015.01.15
OMY	오마이걸	7	서울			160	2015.04.21
GRL	소녀시대	8	서울	02	4444444	168	2007.08.02
ITZ	있지	5	경남			167	2019.02.12
RED	레드벨벳	4	경북	054	5555555	161	2014.08.01
APN	에이핑크	6	경기	031	7777777	164	2011.02.10
SPC	우주소녀	13	서울	02	8888888	162	2016.02.25
MMU	마마무	4	전남	061	99999999	165	2014.06.19

CSV 파일 처리

■ 문자열에서 일부 문자를 변경하는 코드 (replace 함수를 사용)

```
myStr = "2025-5-5"
myStr.replace('-', '#')
'2025#5#5'
```

- format()을 이용하여 정수 데이터 형식을 소수점이 들어간 문자열 형식으로 변경할 수 있음
- 두 번째 행은 "{순번:형식} {순번:형식}".format(값, 값, ...) 형식
- 형식은 정수인 경우엔 d, 실수인 경우엔 f를 사용
- 0.0f는 소수점 아래 자릿수가 0이라는 의미
- 0.2f는 소수점 아래 자릿수를 2자리까지 설정하라는 의미

```
num1, num2 = 1234.555, 100
"{0:0.0f} {1:0.2f}".format(num1, num2)
```

CSV 파일 처리

- 리스트를 문자열로 변경하려면 '구분자'.join()을 사용
- 리스트의 내용을 해시(#)로 구분해서 묶어줌

```
myList = ['파이썬', '데이터', '분석']
'#'.join(myList)
'#'.join(myList)
```

- 숫자 리스트의 경우 join을 하기 전에 모두 문자열로 변경해야 함
- map(함수, 리스트)로 리스트에 함수가 한 번에 적용되는 방식을 사용하면 편리함
- map()은 myList의 각 항목마다 str() 함수를 한꺼번에 적용

```
myList = [100, 200, 300]
list(map(str, myList))
```

CSV 파일 처리

• join()과 map()을 모두 사용하여 숫자 리스트를 문자열로 묶을 수 있음

```
myList = [2025, 8, 8]
'/'.join(map(str, myList))
'2025/8/8'
```

CSV 파일 처리

- CSV 파일을 다른 파일로 복사하는 코드
- 단, 날짜의 형식을 "연.월.일"에서 "연/월/일"로 변경
- 평균 키도 소수점 아래 2자리까지 자릿수를 변경

CSV 파일 처리

```
Code06-04.py
    with open("C:/CookAnalysis/CSV/singer1.csv", "r") as inFp:
01
        with open("C:/CookAnalysis/CSV/new_singer1.csv", "w") as outFp:
02
            header = inFp.readline()
03
            header = header.strip()
04
05
            header_list = header.split(',')
            header_str = ','.join(map(str, header_list))
06
            outFp.write(header_str + '\n')
07
08
            for inStr in inFp:
09
                inStr = inStr.strip()
                row_list = inStr.split(',')
10
11
                row_list[-1] = row_list[-1].replace('.', '/')
                height_str = "{0:.2f}".format(int(row_list[-2]))
12
                row_list[-2] = height_str
13
                row_str = ','.join(map(str, row_list))
14
                outFp.write(row_str + '\n')
15
16
17
    print('Save. OK~')
```

┖SV 파일 처리

실행 결과

아이디,이름,인원,주소,국번,전화 번호,평균 키,데뷔 일자 TWC,트와이스,9,서울,02,11111111,167.00,2015/10/19 BLK,블랙핑크,4,경남,055,22222222,163.00,2016/08/08 WMN,여자친구,6,경기,031,33333333,166.00,2015/01/15 OMY,오마이걸,7,서울,,,160.00,2015/04/21 GRL,소녀시대,8,서울,02,44444444,168.00,2007/08/02 ITZ,있지,5,경남,,,167.00,2019/02/12 RED,레드벨벳,4,경북,054,55555555,161.00,2014/08/01 APN,에이핑크,6,경기,031,77777777,164.00,2011/02/10 SPC,우주소녀,13,서울,02,88888888,162.00,2016/02/25 MMU,마마무,4,전남,061,99999999,165.00,2014/06/19

[프로그램 1] 완성

```
아이디,이름,평균 키
Code06-05.py
                                                                       TWC, 트와이스, 167
01
    with open("C:/CookAnalysis/CSV/singer1.csv", "r") as inFp :
                                                                       WMN,여자친구,166
02
        with open("C:/CookAnalysis/CSV/new_singer2.csv", "w") as out
             header = inFp.readline()
03
                                                                       GRL,소녀시대,168
04
             header = header.strip()
             header_list= header.split(',')
                                                                       ITZ,있지,167
05
             idx1 = header_list.index('아이디')
06
                                                                       MMU,마마무,165
07
             idx2 = header_list.index('이름')
08
             idx3 = header_list.index('평균 키')
             header_list = [header_list[idx1], header_list[idx2], header_list[idx3]]
09
             header_str = ','.join(map(str, header_list))[
10
             outFp.write(header_str + '\n')
11
12
             for inStr in inFp:
                 inStr = inStr.strip()
13
14
                 row_list = inStr.split(',')
                 if int(row_list[idx3]) >= 165 :
15
                     row_list = [row_list[idx1], row_list[idx2], row_list[idx3]]
16
                     row_str = ','.join(map(str, row_list))
17
18
                     outFp.write(row_str + '\n')
19
20
     print('Save. OK~')
```

실행 결과

[프로그램 1] 완성

SELF STUDY 6-1

Code06-05.py를 수정하여 이름, 국번, 전화 번호만 new_singer3.csv로 저장해보자. 단, 전화 번호가 없는 행은 제외하고 저장한다.

실행 결과

이름,국번,전화 번호

트와이스,02,11111111

블랙핑크,055,22222222

여자친구,031,33333333

소녀시대,02,4444444

레드벨벳,054,55555555

에이핑크,031,77777777

우주소녀,02,88888888

마마무,061,99999999

- 콤마가 포함된 데이터

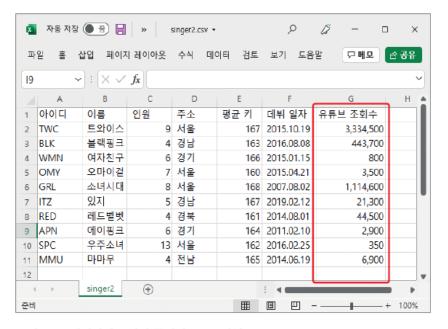


그림 6-6 데이터에 콤마가 들어간 CSV 파일

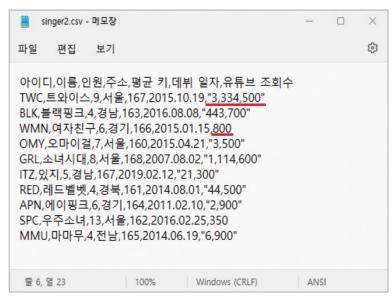


그림 6-7 데이터에 콤마가 들어간 CSV를 메모장에서 확인

- 콤마가 포함된 데이터

singer2.csv 파일을 읽어서 이름(첫 번째 열)과 유튜브 조회수(여섯 번째 열)를 만 단위까지
 만 출력

```
Code06-06.py
      with open("C:/CookAnalysis/CSV/singer2.csv", "r") as inFp :
01
02
          header = inFp.readline()
03
          header = header.strip()
          header_list = header.split(',')
04
          print(header_list[1], header_list[6])
05
          for inStr in inFp:
06
              inStr = inStr.strip()
07
08
              row_list = inStr.split(',')
09
              youtube = int(row_list[6])
10
              youtube = int(youtube/10000)
              print(row_list[1], str(youtube)+"만")
11
```

```
실행결과
이름 유튜브 조회수
Traceback (most recent call last):
File "C:/CookAnalysis/New Code06-08.py", line 9, in <module>
  youtube = int(row_list[6])
ValueError: invalid literal for int() with base 10: '"3'
```

- 콤마가 포함된 데이터

 오류가 발생한 이유는 8행에서 콤마로 분류하여 row_list[6]에는 큰따옴표가 앞에 붙은 "3 값만 들어 있고, row_list[7]에는 334가, row_list[8]에는 500" 값이 저장되어 있기 때문

csv 라이브러리

- csv 라이브러리는 오픈한 파일을 CSV 전용 리더로 변환하면 됨
- 또한 next() 함수를 사용하면 바로 리스트로 반환됨

csv 라이브러리

Code06-06.py의 오류를 해결한 Code06-07.py

```
Code06-07.py
01
      import csv
02
      with open("C:/CookAnalysis/CSV/singer2.csv", "r") as inFp :
03
          csvReader = csv.reader(inFp)
04
05
          header_list = next(csvReader)
          print(header_list[1],header_list[6])
06
                                                               실행 결과
          for row list in csvReader:
07
              youtube = int(row_list[6].replace(',',''))
08
                                                               이름 유튜브 조회수
              youtube = int(youtube/10000)
09
                                                               트와이스 333만
              print(row_list[1], str(youtube)+"만")
10
                                                               블랙핑크 44만
                                                               여자친구 0만
                                                               오마이걸 0만
                                                               소녀시대 111만
                                                               있지 2만
                                                               레드벨벳 4만
                                                               에이핑크 0만
                                                               우주소녀 0만
                                                               마마무 0만
```

CSV 파일 합치기

- 같은 형식의 데이터가 있는 여러 개의 파일을 하나로 계산해야 하는 경우도 있음
- 다음은 singerA.csv와 singerB.csv 파일을 합쳐서 singerSum.csv 파일로 만드는 코드

┖SV 파일 합치기

```
Code06-08.pv
01
    import csv
02
    with open("C:/CookAnalysis/CSV/singerA.csv", "r") as inFpA:
03
        with open("C:/CookAnalysis/CSV/singerB.csv", "r") as inFpB:
04
05
            with open("C:/CookAnalysis/CSV/singerSum.csv", "w", newline=") as outFp:
                csvReaderA = csv.reader(inFpA)
06
07
                csvReaderB = csv.reader(inFpB)
08
                csvWriter = csv.writer(outFp)
09
                header_list = next(csvReaderA)
                header_list = next(csvReaderB)
10
11
                csvWriter.writerow(header_list)
12
13
                for row list in csvReaderA:
14
                     csvWriter.writerow(row_list)
15
                for row_list in csvReaderB:
16
                     csvWriter.writerow(row_list)
17
18
    print('Save. OK~')
```

┖SV 파일 합치기

실행 결과

아이디,이름,인원,주소,국번,전화 번호,평균 키,데뷔 일자 ITZ,있지,5,경남,,,167,2019.02.12 RED,레드벨벳,4,경북,054,55555555,161,2014.08.01 APN,에이핑크,6,경기,031,7777777,164,2011.02.10 SPC,우주소녀,13,서울,02,88888888,162,2016.02.25 MMU,마마무,4,전남,061,99999999,165,2014.06.19 TWC,트와이스,9,서울,02,11111111,167,2015.10.19 BLK,블랙핑크,4,경남,055,22222222,163,2016.08.08 WMN,여자친구,6,경기,031,33333333,166,2015.01.15 OMY,오마이걸,7,서울,,,160,2015.04.21 GRL,소녀시대,8,서울,02,44444444,168,2007.08.02

SELF STUDY 6-2

Code06-08.py를 수정하여 singerA.csv와 singer.B.csv 중 평균 키가 165 이상인 가수 그룹으로 구성된 singer 165.csv 파일을 생성하자.

- CSV 파일의 GUI 출력
 - 4행 3열 워크시트 형식을 출력하는 코드

```
Code06-09.py
     from tkinter import *
01
02
    ## 함수 선언 부분 ##
03
04
    def makeEmptySheet(r, w) :
05
         retList = []
        for i in range(0, r):
06
07
             tmpList = []
08
            for k in range(0, w):
                 ent = Entry(window, text='', width=10)
09
                 ent.grid(row=i, column=k)
10
                 tmpList.append(ent)
11
12
             retList.append(tmpList)
13
         return retList
14
15
    ## 전역 변수 부분 ##
16
     csvList = [ [ '제목1', '제목2' , '제목3'],
17
                 [ 111, 222, 333],
18
                 [ 444, 555, 666],
                 [ 777, 888, 999]]
19
```

- CSV 파일의 GUI 출력
 - 4행 3열 워크시트 형식을 출력하는 코드

```
Code06-09.py
      row Num, co \mathbb{N}Num = 4, 3
20
21
      workSheet = []
22
23
      ## 메인 코드 부분 ##
24
      w indow = Tk()
25
      workSheet = makeEmptySheet(rowNum, co Num)
26
27
      for i in range(0, rowNum):
28
           for k in range (0, co \num):
29
               workSheet[i][k].insert(0, csvList[i][k])
30
31
      window .main loop ()

∅ tk

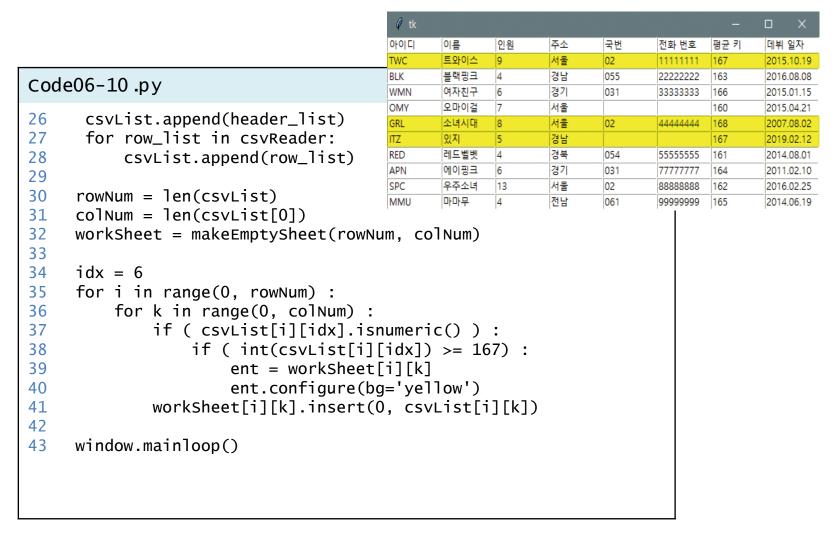
                                                              제목1
                                                      제목2
                                                               제목3
                                            111
                                                      222
                                                               333
                                            444
                                                      555
                                                               666
                                            777
                                                               999
                                                      888
```

┗ [프로그램 2] 완성

- singer1.csv를 읽어서 GUI 환경으로 출력
- 평균 키가 167 이상인 가수 그룹의 행 색상도 변경

```
Code06-10.py
     from tkinter import *
01
02
    import csv
    ## 함수 선언 부분 ##
03
     def makeEmptySheet(r, w) :
04
     ~~~ Code06-09.py의 5~13행과 동일 ~~~
14
    ## 전역 변수 부분 ##
15
16
    csvList = []
     rowNum, colNum = 0, 0
17
18
     workSheet = []
19
20
     ## 메인 코드 부분 ##
21
     window = Tk()
22
23
     with open("C:/CookAnalysis/CSV/singer1.csv", "r") as inFp:
24
         csvReader = csv.reader(inFp)
25
         header_list = next(csvReader)
```

┗ [프로그램 2] 완성



┗ [프로그램 2] 완성

