

## کار با فایل‌های CSV

به بازگشت خوش آمدید! بیایید درباره کار با فایل‌های CSV در پایتون بحث کنیم. فایل با پسوند CSV یک فایل با جداکننده‌های کاما (,) است. همه فایل‌های CSV متن ساده هستند و شامل کاراکترهای الفبایی عددی هستند و داده‌ها را در قالب یک جدول سازماندهی می‌کنند. فایل‌های CSV را با فایل‌های اکسل اشتباه نگیرید، هرچند که فایل‌های CSV به شکل بسیار مشابهی با فایل‌های اکسل قالب‌بندی می‌شوند، اما دارای نوع داده برای مقادیر خود نیستند، همه آنها رشته‌ها با بدون قلم یا رنگ هستند. همچنین، آنها هیچ ورک‌شیتی مانند یک فایل اکسل ندارند. در پایتون چندین کتابخانه برای کار با فایل‌های اکسل وجود دارد، می‌توانید آنها را [اینجا](http://www.python-excel.org/) (<http://www.python-excel.org/>) و [اینجا](https://www.xlwings.org/) (<https://www.xlwings.org/>) بررسی کنید.

فایل‌های با قالب CSV عموماً برای تبادل داده استفاده می‌شوند، معمولاً زمانی که میزان داده بسیار زیادی وجود دارد و بین برنامه‌ها متفاوت تبادل می‌شود. برنامه‌های پایگاه داده، نرم‌افزارهای تحلیلی و سایر برنامه‌ها که مقدار زیادی اطلاعات (مانند مخاطبین و داده‌های مشتری) را ذخیره می‌کنند، عموماً از فرمت CSV پشتیبانی می‌کنند.

بیایید بررسی کنیم که چگونه می‌توانیم یک فایل CSV را با کتابخانه csv در پایتون باز کنیم.

## محل دفترچه نوت بوک

دستور **pwd** را درون یک سلول دفترچه نوت بوک اجرا کنید تا متوجه شوید دفترچه نوت بوک شما در کجا قرار دارد.

In [1]:

```
1 pwd
```

Out[1]:

```
'C:\\Users\\babak\\Python For Everyone\\17-PDFs-and-Spreadsheets'
```

## خواندن فایل‌های CSV

In [2]:

```
1 import csv
```

وقتی مسیر فایل را انتقال می‌دهید، مطمئن شوید که در صورت وجود، پسوند را نیز درج کنید و می‌توانید با استفاده از کلید تب (Tab) نام فایل را به طور خودکار تکمیل کنید. اگر نتوانید از تکمیل خودکار با کلید تب استفاده کنید، این یک نشانگر خوب است که فایل شما در همان محل قرار ندارد. همیشه می‌توانید مسیر کامل فایل را تایپ کنید (از نظر قالب‌بندی شبیه به خروجی **pwd** خواهد بود).

In [3]:

```
1 data = open('example.csv')
```

In [4]:

```
1 data
```

Out[4]:

```
<_io.TextIOWrapper name='example.csv' mode='r' encoding='cp1252'>
```

## Encoding

بسیاری اوقات فایل‌های CSV ممکن است شامل کاراکترهایی باشند که نمی‌توانید با پایتون استاندارد آنها را تفسیر کنید، این می‌تواند مانند نماد @ یا حتی کاراکترهای خارجی باشد. بیا یک مثال از این نوع خطا را مشاهده کنیم. (این امر بسیار رایج است، بنابراین مهم است آن را بررسی کنیم.)  
<https://stackoverflow.com/questions/9233027/unicodedecodeerror-charmap-codec-cant-decode-byte-x->  
[\(\(in-position-y-character](#)

In [5]:

```
1 csv_data = csv.reader(data)
```

تلاش برای تبدیل آن به لیست باعث بروز خطا می‌شود، توجه کنید که در خط خطا، خطای **can't decode** وجود دارد، این یک نشانه است که مشکلی در رمزگذاری وجود دارد!

In [6]:

```
1 data_lines = list(csv_data)
```

```
-----  
-  
UnicodeDecodeError                                Traceback (most recent call last)  
t)  
Cell In[6], line 1  
----> 1 data_lines = list(csv_data)  
  
File C:\ProgramData\Anaconda3\lib\encodings\cp1252.py:23, in IncrementalDecoder.decode(self, input, final)  
    22 def decode(self, input, final=False):  
--> 23     return codecs.charmap_decode(input,self.errors,decoding_table)  
[0]  
  
UnicodeDecodeError: 'charmap' codec can't decode byte 0x8d in position 181  
0: character maps to <undefined>
```

بیا یک بار استفاده از رمزگذاری "utf-8" آن را بخوانیم.

In [7]:

```
1 data = open('example.csv', encoding= 'utf-8')
2 csv_data = csv.reader(data)
3 data_lines = list(csv_data)
```

In [8]:

```
1 # Looks like it worked!
2 data_lines[:3]
```

Out[8]:

```
[['id', 'first_name', 'last_name', 'email', 'gender', 'ip_address', 'city'],
 ['1',
  'Joseph',
  'Zaniolini',
  'jzaniolini0@simplemachines.org',
  'Male',
  '163.168.68.132',
  'Pedro Leopoldo'],
 ['2',
  'Freida',
  'Drillingcourt',
  'fdrillingcourt1@umich.edu',
  'Female',
  '97.212.102.79',
  'Buri']]
```

لطفاً به یاد داشته باشید که مورد اول در لیست، خط سربرگ است که شامل اطلاعات درباره نمایندگی هر ستون است. بیاپید چاپ خروجی را به شکلی کمی تنظیم کنیم:

In [9]:

```
1 for line in data_lines[:5]:
2     print(line)
```

```
['id', 'first_name', 'last_name', 'email', 'gender', 'ip_address', 'city']
['1', 'Joseph', 'Zaniolini', 'jzaniolini0@simplemachines.org', 'Male', '163.168.68.132', 'Pedro Leopoldo']
['2', 'Freida', 'Drillingcourt', 'fdrillingcourt1@umich.edu', 'Female', '97.212.102.79', 'Buri']
['3', 'Nanni', 'Herity', 'nherity2@statcounter.com', 'Female', '145.151.178.98', 'Claver']
['4', 'Orazio', 'Frayling', 'ofrayling3@economist.com', 'Male', '25.199.143.143', 'Kungur']
```

تصور کنید می‌خواهیم لیستی از تمام ایمیل‌ها داشته باشیم. به عنوان نمایشی، از آنجایی که تعداد آیت‌ها 1000 مورد به علاوه سربرگ است، تنها چند ردیف را نمایش می‌دهیم.

In [10]:

```
1 len(data_lines)
```

Out[10]:

1001

In [11]:

```
1 all_emails = []
2 for line in data_lines[1:15]:
3     all_emails.append(line[3])
```

In [12]:

```
1 print(all_emails)
```

```
['jzaniolini0@simplemachines.org', 'fdrillingcourt1@umich.edu', 'nherity2@statcounter.com', 'ofrayling3@economist.com', 'jmurrison4@cbslocal.com', 'lgamet5@list-manage.com', 'dhowatt6@amazon.com', 'kherion7@amazon.com', 'chedworth8@china.com.cn', 'hgasquoine9@google.ru', 'ftarra@shareasale.com', 'abathb@umn.edu', 'lchastangc@goo.gl', 'cceried@yale.edu']
```

اگر می‌خواستیم لیستی از نام و نام خانوادگی کامل داشته باشیم، چه میشد؟

In [13]:

```
1 full_names = []
2
3 for line in data_lines[1:15]:
4     full_names.append(line[1] + ' ' + line[2])
```

In [14]:

```
1 print(full_names)
```

```
['Joseph Zaniolini', 'Freida Drillingcourt', 'Nanni Herity', 'Orazio Frayling', 'Julianne Murrison', 'Lucy Gamet', 'Dyana Howatt', 'Kassey Herion', 'Chrissy Hedworth', 'Hyatt Gasquoine', 'Felicdad Tarr', 'Andrew Bath', 'Luca Chastang', 'Car Cerie']
```

## نوشتن در فایل‌های CSV

ما همچنین می‌توانیم فایل‌های CSV بنویسیم، ساخته شده یا به فایل‌های موجود اضافه کنیم.

### فایل جدید

این عمل همچنین هر فایل موجود با همین نام را برای شما بازنویسی می‌کند، پس در این مورد دقت کنید!

In [15]:

```
1 # newline controls how universal newlines works (it only applies to text
2 # mode). It can be None, '', '\n', '\r', and '\r\n'.
3 file_to_output = open('to_save_file.csv', 'w', newline='')
```

In [16]:

```
1 csv_writer = csv.writer(file_to_output, delimiter = ',')
```

In [17]:

```
1 csv_writer.writerow(['a', 'b', 'c'])
```

Out[17]:

7

In [18]:

```
1 csv_writer.writerows([['1', '2', '3'], ['4', '5', '6']])
```

In [19]:

```
1 file_to_output.close()
```

---

فایل موجود

In [21]:

```
1 f = open('to_save_file.csv', 'a', newline='')
```

In [22]:

```
1 csv_writer = csv.writer(f)
```

In [23]:

```
1 csv_writer.writerow(['7', '8', '9'])
```

Out[23]:

7

In [24]:

```
1 f.close()
```

این همه بود برای مباحث اصولی! اگر فکر می‌کنید که برای اغلب کارهای خود با فایل‌های CSV سر و کار خواهید داشت،

