

# تمرین پازل PDF و صفحه‌های گسترده

بیا یاد مهارت‌های شما را آزمایش کنیم. برای این تمرین، به دو فایل نیاز خواهید داشت و باید وظایف زیر را حل کنید:

- وظیفه اول: با استفاده از پایتون، لینک Google Drive را از فایل csv استخراج کنید. (راهنما: این در طول قطر از بالا سمت چپ تا پایین سمت راست قرار دارد).
- وظیفه دوم: PDF را از لینک Google Drive دانلود کنید (ما قبلاً آن را برای شما دانلود کرده‌ایم، فقط در صورت عدم دانلود از Google Drive) و شماره تلفن موجود در سند را پیدا کنید. توجه: روش‌های مختلفی برای قالب بندی شماره تلفن وجود دارد!

## وظیفه اول: گرفتن لینک از Google Drive از فایل csv.

In [1]:

```
1 import csv
```

In [13]:

```
1 # The correct result is shown below, if you can't download from Google Drive,  
2 # we added the PDF file to the Exercise_Files folder already  
3 data = open('Exercise_Files/find_the_link.csv', encoding='utf-8')  
4 csv_data = csv.reader(data)  
5 data_lines = list(csv_data)
```

## روش اول

In [14]:

```
1 link_list = []  
2  
3 for row_num, data in enumerate(data_lines):  
4     link_list.append(data[row_num])
```

In [15]:

```
1 ''.join(link_list)
```

Out[15]:

```
'https://drive.google.com/open?id=1G6SEgg018UB4_4xsAJJ5TdzhmXipz4M'
```

In [16]:

```
1 link_str = ''
2
3 for row_num, data in enumerate(data_lines):
4     link_str += data[row_num]
5
6 link_str
```

Out[16]:

```
'https://drive.google.com/open?id=1G6SEgg018UB4_4xsAJJ5TdzhmXipz4M'
```

## وظیفه دوم: پیدا کردن شماره تلفن موجود در سند PDF.

In [ ]:

```
1 # You should get this phone number
2 # 505 503 4455
```

In [17]:

```
1 import PyPDF2
```

In [18]:

```
1 f = open('Exercise_Files/Find_the_Phone_Number.pdf', 'rb')
```

In [19]:

```
1 pdf = PyPDF2.PdfReader(f)
```

In [20]:

```
1 len(pdf.pages)
```

Out[20]:

17

In [21]:

```
1 import re
```

In [22]:

```
1 pattern = r'\d{3}.\d{3}.\d{4}'
```

In [23]:

```
1 for n in range(len(pdf.pages)):
2     page = pdf.pages[n]
3     page_text = page.extract_text()
4     match = re.search(pattern, page_text)
5
6     if match:
7         print(match.group())
```

505.503.4455