



پایتون برای علم داده

سعید محققى

04: Working with Data

خواندن فایل CSV

- روش 1 (خواندن به صورت فایل متنی)

```
with open("weather_data.csv") as data_file:  
    data = data_file.readlines()
```

نوع: List

هر خط در یک آیتم (شامل \n)

- روش 2 (استفاده از csv)

```
import csv  
with open("weather_data.csv") as data_file:  
    data = csv.reader(data_file)
```

نوع: _csv.reader

هر خط در یک آیتم (بدون \n)

- روش 3 (استفاده از pandas)

```
import pandas  
data = pandas.read_csv("weather_data.csv")
```

نوع: DataFrame

دارای توابع متعدد برای پردازش داده

عملیات بر روی DataFrame در Pandas

weather_data.csv

day	temp	condition
Monday	12	Sunny
Tuesday	14	Rain
Wednesday	15	Rain
Thursday	14	Cloudy
Friday	21	Sunny
Saturday	22	Sunny
Sunday	24	Sunny

```
import pandas as pd
data = pd.read_csv("weather_data.csv")
data.head()      /      data.head(10)
data.shape()     /      data.dtypes()
data.describe() /      data.to_dict()
data["temp"]     /      data["temp"].to_list()
data["temp"].mean()
data[data["condition"] == "Sunny"]
data[data["temp"] == data["temp"].max()]
```

■ پروژه 2: آنالیز و پردازش داده در فایل csv (csv processing)

1. خواندن فایل csv (و xlsx) و دسترسی به داده ها

2. استفاده از پکیج pandas

3. تغییر داده‌ها (تبدیل واحدهای قد و وزن)

4. محاسبه BMI هر فرد

5. ذخیره نتایج در فایل csv، xlsx، json و html

What will we learn? (Keywords)

- Read / Write CSV Files (+ xlsx)
- Pandas Package
- Loops (For)

- حلقه روی آیتم‌های یک مجموعه

```
for <var> in <iterable>:  
    <statement(s)>
```

```
languages = ['Swift', 'Python', 'Go', 'JavaScript']  
  
# access items of a list using for loop  
for language in languages:  
    print(language)
```

■ ساختار `range(start, stop, step)`

- `x = range(6)`
- `x = range(3, 6)`
- `x = range(3, 20, 2)`

■ حلقه روی Range

- ```
for n in range(1, 100):
 print(n)
```

### ■ استفاده از Enumerate

```
words = ['a', 'b', 'c', 'd']
for i, word in enumerate(words):
 print(i+1, word)
```

```
1 a
2 b
3 c
4 d
```

### ■ استفاده از Iterator

```
nums = [1, 2, 3, 4]
for num in nums:
 print(num)
```

```
1
2
3
4
```

### ■ استفاده از Generator

```
for i in range(5):
 print(i)
```

```
0
1
2
3
4
```

این دستورها را امتحان کنید (List Comprehension):

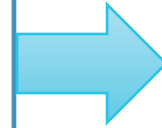
- `[n for n in range(1,21)]`
- `[n/2 for n in range(1,21)]`

## ایجاد / ذخیره DataFrame در Pandas

```
data_dict = {"age": [30, 29],
 "height": ['165', '170']}
```

```
df = pd.DataFrame(data_dict)
```

```
df = pd.DataFrame([(30, 165),
 (29, 170)],
 columns=('age', 'height'))
```



```
>>> df
 age height
0 30 165
1 29 170
```

```
>>> df.values
array([[30, 165],
 [29, 170]])
```

```
df.to_csv("my_data.csv")
df.to_excel("my_data.xlsx")
df.to_json("my_data.json", indent=4)
df.to_html("my_data.html")
```



## پکیج های Pandas و Matplotlib

<https://www.w3schools.com/python/pandas/default.asp>

آشنایی با Pandas



<https://pandas.pydata.org/>

[https://www.w3schools.com/python/matplotlib\\_intro.asp](https://www.w3schools.com/python/matplotlib_intro.asp)

آشنایی با Matplotlib



<https://matplotlib.org/>

<https://matplotlib.org/stable/gallery/>

مثال های Matplotlib

## پکیج های Numpy و Scipy

<https://www.w3schools.com/python/numpy/default.asp>

آشنایی با Numpy

**NumPy** 

<https://numpy.org/>



**SciPy**

<https://www.w3schools.com/python/scipy/index.php>

آشنایی با Scipy

<http://scipy.org/>