

Pandas

Prof.Dr. Bahadır AKTUĞ
JFM212 Python ile Mühendislik Uygulamaları

**Kaynakça bölümünde verilen kaynaklardan derlenmiştir.*

Pandas

- ▶ Pandas, Python ile verilerin kolayca işlenmesi için verilerin işlenmesi için veri yapıları sunan eklentidir.
- ▶ Pandas, Numpy dayalı olarak geliştirilmiştir.
- ▶ Pandas ile birlikte, istenirse SciPy ve Matplotlib de kullanılabilir.
- ▶ Pandas, iki temel veri yapısı sunar:
 - ▶ Seri (Series)
 - ▶ Veri Çerçevesi (DataFrame)
- ▶ Seriler, tek boyutlu diziler şeklinde verilerin tutulması ve üzerlerinde işlem yapılmasını sağlar.
- ▶ Buna karşın, Veri Çerçeveleri tablo şeklinde çok sütunlu veriler için kullanılır.

Seri (Series)

- ▶ Serilerin 1 boyutlu Numpy dizilerine benzemekle birlikte, çok daha esnek ve kuvvetli yanları bulunmaktadır.
- ▶ Seriler tek boyutlu tanımlansalar dahi, indisleri bulunmaktadır.

```
import pandas as pd
a = pd.Series([5, 54, 12, 22, 15, 87])
print(a)
0    5
1   54
2   12
3   22
4   15
5   87
dtype: int64
```

Seri (Series)

```
import pandas as pd
a = pd.Series([5, 54, 12, 22, 15, 87])
print(a.index)
Int64Index([0, 1, 2, 3, 4, 5], dtype='int64')dtype: int64
print(a.values)
[ 5 54 12 22 15 87]
```

► Serilerin indisleri ayrıca verilebilir.

```
import pandas as pd
indisler = ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk']
degerler = [5, 54, 12, 22, 15, 87]
a = pd.Series(degerler,index=indisler)
print(a)
```

```
Ali      5
Ahmet    54
Mehmet   12
Veli     22
Arda     15
Doruk    87
dtype: int64
```

Seri (Series)

- Serilerin istenilen sayı ve sıradaki elemanları indisleri yardımıyla gösterilebilir

```
import pandas as pd
indisler = ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk']
degerler = [5, 54, 12, 22, 15, 87]
a = pd.Series(degerler,index=indisler)
print(a[['Mehmet', 'Arda', 'Ali']])
```

```
Ali      5
Ahmet    54
Mehmet   12
Veli     22
Arda     15
Doruk    87
dtype: int64
```

Seri (Series)

- ▶ Seriler üzerinde doğrudan aritmetik işlemler yapılabilir.
- ▶ Seriler Numpy komutlarına girdi olarak verilebilir.

```
import pandas as pd
indisler = ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk']
degerler = [5, 54, 12, 22, 15, 87]
a = pd.Series(degerler,index=indisler)
print(5*a+2)
print(np.sqrt(a))
```

Ali	27	Ali	2.236068
Ahmet	272	Ahmet	7.348469
Mehmet	62	Mehmet	3.464102
Veli	112	Veli	4.690416
Arda	77	Arda	3.872983
Doruk	437	Doruk	9.327379
dtype: int64		dtype: float64	

Seri (Series)

- Seriler üzerinde doğrudan filtreleme de yapılabilir.

```
import pandas as pd
indisler = ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk']
degerler = [5, 54, 12, 22, 15, 87]
a = pd.Series(degerler,index=indisler)
print(a[a > 50])
```

```
Ahmet    54
Doruk    87
dtype: int64
```

DataFrame

- ▶ DataFrame'ler, birden fazla seriden oluşan tablolar olarak düşünülebilir.
- ▶ Dataframe'lerin her bir sütunu bir seri'dir.

```
import pandas as pd
ogrenciler = {
    'isim' : ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk'],
    'not' : [5, 54, 12, 22, 15, 87],
    'yas' : [21, 24, 18, 28, 20, 22]}
a = pd.DataFrame(ogrenciler)
print(a)
```

```
0  Ali  5  21
1  Ahmet 54  24
2  Mehmet 12  18
3  Veli  22  28
4  Arda  15  20
5  Doruk 87  22
```


DataFrame

- ▶ DataFrame'ler sütunlar birer seri olduğundan doğrudan herhangi bir sütun üzerinde işlem yapılabilir

```
import pandas as pd
ogrenciler = {
    'isim' : ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk'],
    'not' : [5, 54, 12, 22, 15, 87],
    'yas' : [21, 24, 18, 28, 20, 22]}
a = pd.DataFrame(ogrenciler)
print(a['not']*1.5)
print(a['yas'].sum())
```

```
0    7.5
1   81.0
2   18.0
3   33.0
4   22.5
5  130.5
133
```

DataFrame

- ▶ DataFrame sütunları içinde veri olmasa da önceden tanımlanabilir. Bu durumda, veri olmayan sütunlar "NaN" olacaktır.

```
import pandas as pd
ogrenciler = {
    'isim' : ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk'],
    'not' : [5, 54, 12, 22, 15, 87],
    'yas' : [21, 24, 18, 28, 20, 22]}
a = pd.DataFrame(ogrenciler,columns=['isim','not','yas','giris'])
print(a)
```

```
isim not yas giris
0  Ali  5  21  NaN
1  Ahmet  54  24  NaN
2  Mehmet  12  18  NaN
3  Veli  22  28  NaN
4  Arda  15  20  NaN
5  Doruk  87  22  NaN
```

DataFrame

- DataFrame sütunları üzerindeki işlemler dışında doğrudan tablo için de işlemler yapılabilir.

```
import pandas as pd
ogrenciler = {
    'isim' : ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk'],
    'not' : [5, 54, 12, 22, 15, 87],
    'yas' : [21, 24, 18, 28, 20, 22]}
a = pd.DataFrame(ogrenciler)
print(a)
print(a.T)
```

	isim	not	yas		0	1	2	3	4	5
0	Ali	5	21	isim	Ali	Ahmet	Mehmet	Veli	Arda	Doruk
1	Ahmet	54	24	not	5	54	12	22	15	87
2	Mehmet	12	18	yas	21	24	18	28	20	22
3	Veli	22	28							
4	Arda	15	20							
5	Doruk	87	22							

Örnek:

- ▶ Aşağıdaki verileri kullanarak öğrencilerin yarıyol notunu bularak ortalama sütununa yazınız (%25 arasınav, %25 ödev ve %50 final şeklinde).

isim	arasinav	odev	final
Ali	5	68	55
Ahmet	54	44	64
Mehmet	12	87	77
Veli	22	74	88
Arda	15	68	100
Doruk	87	70	90

Çözüm:

- Aşağıdaki verileri kullanarak öğrencilerin yarıyol notunu bularak ortalama sütununa yazınız (%25 arasınav, %25 ödev ve %50 final şeklinde).

```
import pandas as pd
ogrenciler = {
    'isim' : ['Ali','Ahmet','Mehmet','Veli','Arda','Doruk'],
    'arasinav' : [5, 54, 12, 22, 15, 87],
    'odev' : [68, 44, 87, 74, 68, 70],
    'final' : [55, 64, 77, 88, 100, 90]}
a = pd.DataFrame(ogrenciler,columns=['isim','arasinav','odev','final','ortalama'])
a['ortalama'] = a['arasinav']*0.25+a['odev']*0.25+a['final']*0.50
print(a)
```

	isim	arasinav	odev	final	ortalama
0	Ali	5	68	55	45.75
1	Ahmet	54	44	64	56.50
2	Mehmet	12	87	77	63.25
3	Veli	22	74	88	68.00
4	Arda	15	68	100	70.75
5	Doruk	87	70	90	84.25

► Kaynakça

- 1 Wentworth, P., Elkner, J., Downey, A.B., Meyers, C. (2014). *How to Think Like a Computer Scientist: Learning with Python* (3rd edition).
- 2 Pilgrim, M. (2014). *Dive into Python 3* by. Free online version: DiveIntoPython3.org ISBN: 978-1430224150.
- 3 Summerfield, M. (2014) *Programming in Python 3 2nd ed (PIP3)* :- Addison Wesley ISBN: 0-321-68056-1.
- 4 Jones E, Oliphant E, Peterson P, et al. *SciPy: Open Source Scientific Tools for Python*, 2001-, <http://www.scipy.org/>.
- 5 Millman, K.J., Aivazis, M. (2011). *Python for Scientists and Engineers, Computing in Science & Engineering*, 13, 9-12.
- 6 John D. Hunter (2007). *Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering*, 9, 90-95.
- 7 Travis E. Oliphant (2007). *Python for Scientific Computing, Computing in Science & Engineering*, 9, 10-20.
- 8 Goodrich, M.T., Tamassia, R., Goldwasser, M.H. (2013). *Data Structures and Algorithms in Python*, Wiley.
- 9 <http://www.diveintopython.net/>
- 10 <https://docs.python.org/3/tutorial/>
- 11 <http://www.python-course.eu>
- 12 <https://developers.google.com/edu/python/>
- 13 <http://learnpythonthehardway.org/book/>