

# Pandas

Pandas datalarda analiz ve manipulation yapabilmek için yazılmış ücretsiz bir kütüphanedir. Özellikle sayısal tablolar ve zaman serileri için işlemler sunar. Pandas çoğunlukla DataFrameler biçimindeki ML için kullanılır.

Pandas excel gibi dosyalardaki verilerin içe aktarılmasına izin verir. Groupby, join, merge, melt gibi işlemlerin yanı sıra boşlukları doldurma ve veri temizleme gibi özellikleri de vardır.

```
serr = pd.Series([11,21,13,41,15,61])  
serr
```

```
0    11  
1    21  
2    13  
3    41  
4    15  
5    61  
dtype: int64
```

Pandas int, string, float, objects gibi her türlü veriyi labellayabilen tek boyutlu bir dizidir. Diziler sadece tek bir liste içerebiliyorken DataFrameler birden fazla liste içerebilir.

## Data Frame Basics-1 (Attributes)

DataFrame, verilerin satırlar ve sütunlar olarak hizalandığı iki boyutlu bir veri yapısıdır. Veriler, satırlar ve sütunlardan oluşur. Boyutu değişebilir, potansiyel olarak heterojen tablo verisidir.

### Series

Tek boyutlu serilerdir.

#### sort()

Verileri sıralamak için kullanılır.

**sort\_index, sort\_values,**

#### isin()

İçerisine yazılan değerin elimizdeki datadaki varlığını eşleştirme ile bool sonuç olarak döndüren bir methoddur.

#### keys()

Datadaki verilerin start,stop ve step degerlerini veren bir methoddur.

### **values**

Datadaki tipine bagli olarak serileri array olarak dondurur.

### **items()**

Bu method indexleri ve valuelari iterable bir tuple olarak dondurur.

### **read\_csv()**

```
dataframe=pd.read_csv("csv name.csv")
```

Klasorumuzdeki belgeyi okumaya yarayan bir methoddur.

### **head()**

Bu method konuma gore ilk n degerdeki satirlari donderir.Dafault degeri 5'tir.

### **tail()**

Bu method konuma gore son n degerdeki satirlari donderir.Dafault degeri 5'tir.

### **sample()**

Bir object axisinden rastgele ogeler donduren bir methoddur.icerisindeki deger kadar dondurur.

### **shape**

DataFrame'in boyutlarini tanimlar bunu dondurur.

## **Data Frame Basics-2 (Index & Selecting)**

DaraFrame verilerin iki boyutlu olarak satirlar ve sutunlar seklinde hizalanan veri yapisidir.

### **index()**

Tum pandas objeleri icin axis etiketleri saglayan objecttir.

### **columns**

Datanin sutunlarini getirir.

### **shape**

DataFrame'in boyutlarini dondurur.

## size

Bu objedeki elementlerin sayisini int olarak dondurur.

## ndim

Array boyutlarini veya axis sayisini dondurur.

## reset\_index()

DataFrame indexlerini sifirlar ve default degerleri kullanir. DataFrame birden fazla indexe sahipse MultiIndex ise bu yontem bir veya birden fazla duzeyi kaldirabilir.

## set\_index()

Mevcut olan sutundaki degerleri kullanarak DataFrame'e yeni index atar.

	W	X	Y	Z	newidx
A	1	6	7	8	CA
B	4	8	5	0	NY
C	5	8	1	3	WY
D	8	3	3	2	OR
E	8	3	7	0	CO

```
df.set_index('newidx')
```

	W	X	Y	Z
newidx				
CA	1	6	7	8
NY	4	8	5	0
WY	5	8	1	3
OR	8	3	3	2
CO	8	3	7	0

## ["col"]

Sutunlari bu sirayla secmek icin parantez icinde belirtmek gerekir. DataFrame'de bir sutun yoksa bir istisna beklenmedik durum olusur. 'W' basligindaki degerler olur.

## iloc[]

Pozisyona gore gormek icin indexlerle secim yapilir. Ilk girilen deger (:) indexine gore getirir. ikinci deger ise (:) getirilen datanin icinde nelerin gosterilmesi icin baslik

sayisini gosterir.Eger iki deger arasina virgul konulursa getirilen datanın icindeki degeri direkt getirir.

**loc['',']**

Ilk girilen degerin datadaki ikinci girilen basliktaki degerini dondurur.

### **Conditional Indexing**

Parantez icinde belirtilen basliktaki yerde kosula uyan degerleri dondurur.

`df[df["Y"]<6]` Y basliginda 6 dan kucuk degerler

**[ "Baslik" ] == "Deger"**

Bu baslikta buna esit bir deger var mi diye bakar sonra bool degerler donderir.

**df[df["Baslik"] == "Deger"]**

Sadece bu kurala uyan degeri donderir.

**isin[]**

Bu method, data icindeki belli sartlardaki verileri aldigimizda bunlar icinde belli sartlar belirlemek icin kullanilir.Orn;Belli bir seviye uzerinde maas alan kisilerden su mesleklerde calisanlari istedigimizde meslekleri bu method icinde belirtebiliriz.

## **Data Frame Basics-3 (Properties)**

**info()**

Bu method dizin tipi,sutunlar,bos olmayan degerler ve bellek kullanimi dahil DataFrame hakkında bir cok bilgi yazdirir.

**describe()**

eries veya DataFrame'in sayisal deęerlerinin yzdelik, ortalama ve std gibi bazı istatistiksel verilerini hesaplamak iin kullanılır.Bir veri kmesinin daęılımının merkezi eęilimini, daęılımını ve řeklini zetleyenleri ieren tanımlayıcı istatistikler oluřturur.

**shape**

DataFrame'in boyutlarini yazdirir

**len()**

Bir String'teki uzunluğu (karakter sayısını) döndürür. Sözlükler, listeler veya demetler için giriş sayısını verir.

**col.value\_counts()**

Benzersiz değerlerin sayısını içeren bir seri döndürür.

**mean()**

İstenen eksen üzerindeki değerlerin ortalamasını döndürür.

**col.sum()**

İstenen eksen üzerindeki değerlerin toplamını döndürür.

**col.unique()**

Unique'ler görünüm sırasına göre getirilir.Sıralama yapmaz.Benzersiz elemanları getirir.

**isnull().sum()**

Null değerleri tespit eder.Her sütundaki toplam değer sayısını döndürür.

**drop()**

Satırları veya sütunları label adlarını,karşılık gelen axis'ini veya sütun adlarını ya da satırların indexlerini belirterek kaldırmaya yarar.