

Pandas2

Groupby()

Groupby methodu genellikle math islemleri ile kullanilir.mean,standart deviation,max-min,count,sum gibi.

Groupby islemleri orjinal nesne uzerinden bazi islemleri de icerir.Splitting the Object(Nesneyi Bölme),
applying a function(İşlev uygulama),combining the results(Sonuçları birleştirme)

Aggregation

Bir ozet istatistik hesaplar.Bir veya birden cok sutuna islemler yapmak icin kullanilir.Girilen anahtar valuelardaki unique degerlere gore yeniden bir gruplandirma yapar.

```
df.groupby("batch").aggregate([np.min,np.median,np.max])
```

	var1			var2		
	amin	median	amax	amin	median	amax
batch						
A	10	16	22	100	181	262
B	11	17	23	111	182	253
C	33	66	99	333	651	969

Transform()

Gruba ozgu bazi islemleri gerceklestirir.Alt DataFrame'in sonuclarini orjinal DataFrame'ne gore getirir.Her zaman orjinal DataFrame uzunlugunda yeni bir Series dondurur.

filter()

Bazi kosullara bagli olarak verileri atar.Bu alt Dataframe'e gelen verilere bir filtre uygular.Uygulanan kosula bagli olarak orjinal verideki verilerden uyanlari getirir.Bu durumda orjinal DataFrame'den gelen bir kucuk surum elde edilir.

apply()

Yalniz bir isleve izin verir.Toplu sonuclar elde edilir.

Pivot Table & Stack

Özellikler arasındaki ilişki, iki değişkenli analiz yoluyla elde edilebilir. Sayısal ve diğer kategorik özelliklerle ilişkilerini çıkarmak için, groupby ve apply fonksiyonları kullanılır. Pivot tables ve Stack/Unstack fonksiyonları da bu konuda oldukça kullanışlıdır.

Pivot Table

Orjinal DataFrame'den categorical feature'ler kullanılarak index, satır, sütunlar ve veriler kullanılarak yeni bir tablo oluşturmaktır.

```
df = sns.load_dataset("titanic")
df.head()
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

```
df.pivot_table(values="age", index="sex", columns="class")
```

class	First	Second	Third
sex			
female	34.611765	28.722973	21.750000
male	41.281386	30.740707	26.507589

Stack/Unstack

Bir DataFrame'de, en içteki sütun dizini en içteki satır dizini olur. Unstacking ters işlemidir. Yeniden eski haline getirir.

Baslıklardan içteki başlığı index değeri olarak atar. (level =) içerisine yazdığımız başlık değerini index olarak atar.

Useful Methods

```
d={"var1":[1, 2, 3, 4, 5],  
  "var2":[111, 222, 333, 444, 555]}
```

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

	var1	var2
0	1	111
1	2	222
2	3	333
3	4	444
4	5	555

apply()

DataFrame'in bir axis'i boyunca bir işlemi uygular. Verilen axis'teki Seriesler fonksiyondan geçer.

```
df.apply(np.mean)
```

```
var1      3.0  
var2     333.0  
dtype: float64
```

applymap()

```
df.applymap(lambda x:x*5)
```

	var1	var2
0	5	555
1	10	1110
2	15	1665
3	20	2220
4	25	2775

map()

```
df.var1.map({3:"A"})
```

```
0    NaN  
1    NaN  
2     A  
3    NaN  
4    NaN  
Name: var1, dtype: object
```

unique()

Bir Seriesdeki benzersiz değerlerin işlem dizisi, gözlemlenen sırayla döndürülür.

nunique()

Benzersiz değerlerin sayısını verir

value_counts()

Dizin olarak benzersiz değerler ve değerleri olarak frekanslar içeren bir Seri döndürür, sayım azalan düzende sıralanır

sort_values()

Bir DataFrame'i bir sütuna göre sıralama. Sütun adıyla bir 'by' bağımsız değişkenini kabul eder.