Lab 1

2022年7月11日 星期一 14:07

两个库：numpy, pandas

- Series

s = Series (np. random. randn (10))

随机生成一个 " 1-D " array

- Indexing

s[[0, 2, 4]], 选择 1, 3, 5 这三个元素

s[4]: 选择第 5 号元素

s[1:4]: 选择第 2-4 号元素

(切片切前不切后)

- Axis label

s1 = s.copy( )

s1.index = ["item 0", "item 1", "item 2", "item 3" ...]

(改变索引为名称, 从 int 64 → strings)

- Reindex function

s1. reindex (["item 0", "item 1", "item 2", ...])

- 逆序切片

s1 [9: 6: -1]

- len & np.size

print len (s1)   print (np.size(s1))

- Arithmetic operations (Series)

s1 *2   s1+ s2

- String index & Integer Index

s3 = Series (np. arange(3), index=[0, 1, 2])

s4 = Series ([4, 5, 6], index = ["0", "1", "2"]

{ print (s3. index)
  print (s4. index)

- Values

print (s7[1]. values)

['e', 'g'] (输出其中的元素)

- Handling missing data

seriesWithNAs. isnull ( )

能够让数值变成 Bool 值

- Data Frames

1. >> df1 = DataFrame ([[1, 2, 3, 4, 5], [6, 7, 8, 9, 10]], columns = ["a", "b", "c", "d", "e"]

>> df1

>>

	a	b	c	d	e
0	1	2	3	4	5
1	6	7	8	9	10

df2 = DataFrame ({"Name": ["A", "B", "C", "D"], "Age": [25, 26, 23, 40]})

	Name	Age
0	A	25
1	B	26
2	C	23
3	D	40

- Reading in Data.

gplay = pd. read_csv(" ... ")

gplay

- Selecting columns

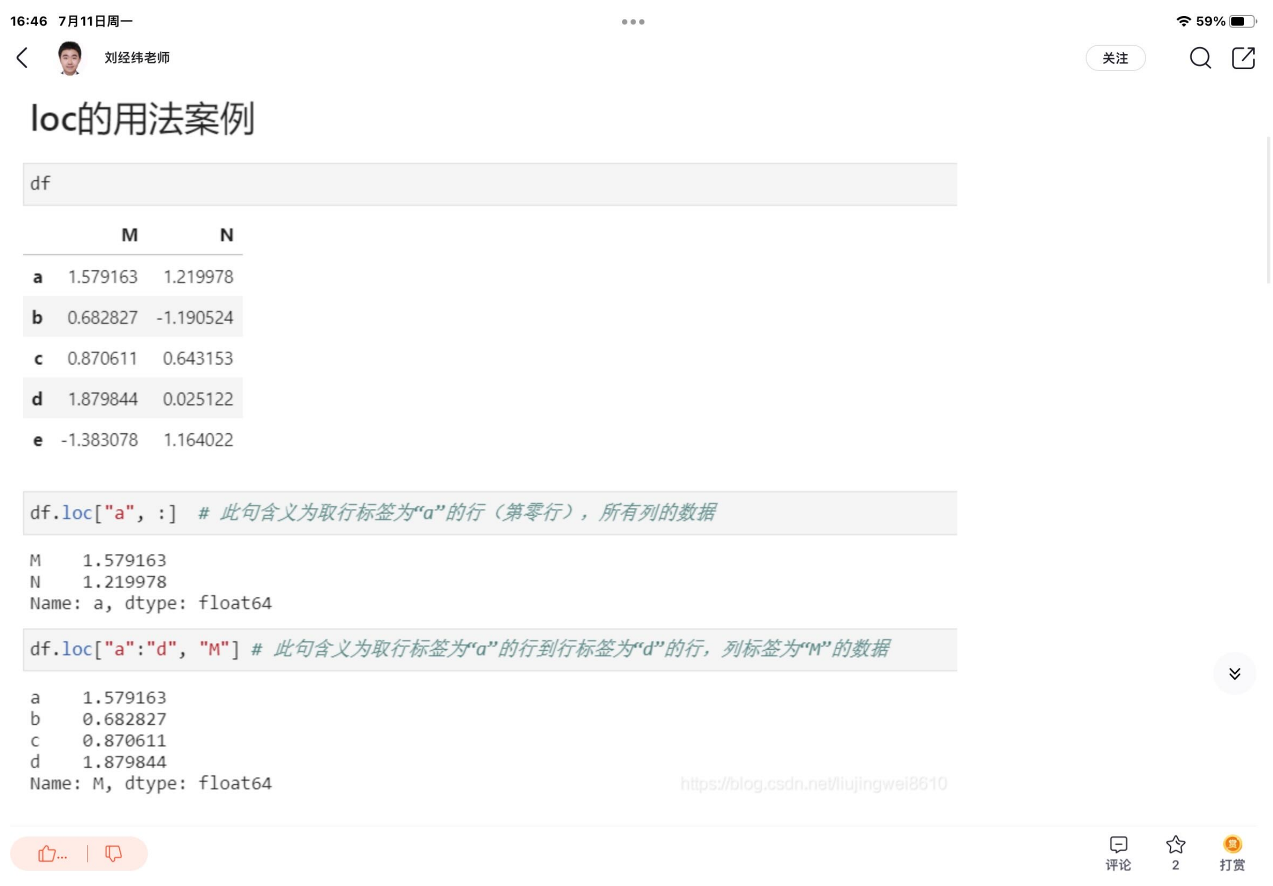
1. print (gplay. columns)   // 打印出所有元素

print (len (gplay.columns))   // 打印出长度

gplay [["Rating"]]   // 仅输出 "Rating" 那一列的

- Selecting Rows

gplay. loc [0] / gplay. iloc



loc的用法案例

```
df
```

	M	N
a	1.579163	1.219978
b	0.682827	-1.190524
c	0.870611	0.643153
d	1.879844	0.025122
e	-1.383078	1.164022

```
df.loc["a", :]  # 此句含义为取行标签为"a"的行（第零行），所有列的数据
```

```
M    1.579163
N    1.219978
Name: a, dtype: float64
```

```
df.loc["a":"d", "M"] # 此句含义为取行标签为"a"的行到行标签为"d"的行，列标签为"M"的数据
```

```
a    1.579163
b    0.682827
c    0.870611
d    1.879844
Name: M, dtype: float64
```

https://blog.csdn.net/liujingwei8610

print (gplay.duplicated ( )) : 重复例为1, 非重为0

gplay. dropduplicates ( ) 去除重复列,

Index of Dataframe

gplay.set_index ("App")

gplay1. loc ["Coloring... "]

Descriptive Statistics

1. gplay. info( )   // 获取更多信息 (关于列表)
2. gplay. describe( )   // 生成数字类的描述类信息
3. gplay. describe (include = "all")   // 涵盖全部的标签

Filtering & Visualization

1. gplay["Category"]. value_counts ( )
2. gplay [" ... "] . value_counts( ). plot ( )   // 曲线图
3. gplay [" ... "] . value_counts ( ), plot. bar( )   // 柱状图
4. ~ ~ ~ ~ . plot (kind = "bar", figsize = (20, 10), fontsize = 20)
5.