

Pandas Series 單維度資料

Created by 戴嘉熙 【小孫學堂】

什麼是Pandas Series

20

= 0

20

索引		內容	
	Α	В	
1	0	Apple	
2	1	Samsung	
3	2	Mi	
4	3	Sony	

建立Pandas Series物件

import pandas as pd

物件名稱 = pandas.Series(資料串列)

```
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
print(phone)

Apple
```

```
1 Samsung
2 Mi
3 Sony
dtype: object
```

20

20

5 0

20

= 3

= 0 = 0

= 3

= 0

= 0

5 0

建立物件索引值

物件.index = ["key1", "key2", "key3"]

p3

p4

Μi

Sony

dtype: object

取得Pandas Series資料

```
1 phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"],
2 index=["p1", "p2", "p3", "p4"])
3 print(phone[1]) # 依據資料順序取值(從0開始計算)
4 print(phone["p3"]) # 依據資料索引值取值

p1 Apple
p2 Samsung
```

合併Pandas Series資料

物件1 = 物件2.append(物件3)

```
1 phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
2 data = pd.Series(["Htc", "Oppo"]) # 新增的資料
3 combined = phone.append(data)
4 print(combined)
```

```
Apple
        Samsung
             Μi
            Sony
            Htc
           Oppo
   dtype: object
```

ignore_index=True or False

是否忽略原始索引值

```
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
    data = pd.Series(["Htc", "Oppo"]) # 新增的資料
    combined = phone.append(data, ignore_index=True)
    print(combined)
                                 Apple
0
       Apple
                         1
                              Samsung
     Samsung
                         2
                                    Μi
          Μi
                         3
                                  Sony
        Sony
                         4
                                   Htc
0
         Htc
                                  Oppo
        0ppo
                         dtype: object
       object
dtype:
```

修改Pandas Series資料

```
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"],
                      index=["p1", "p2", "p3", "p4"])
    phone["p3"] = "Oppo"
    phone[1] = "HTC"
    print(phone)
     Apple
р1
p2
     HTC
p3
  Орро
p4
    Sony
dtype: object
```

Pandas Series字串運算

```
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
print(phone.size) #取得資料筆數
print(phone.str.upper()) # 將字串資料轉換為大寫
print(phone.str.lower()) # 將字串資料轉換為小寫
print(phone.str.contains("Sa")) # 搜尋是否包含特定字串
print(phone.str.cat(sep=";")) # 利用自訂分隔符號連接字串
print(phone.str.replace("Samsung", "Oppo")) # 將Samsung取代為Oppo
```

```
Apple; Samsung; Mi; Sony
                                     False
        APPLE
                               0
                       apple
                                                 Apple
      SAMSUNG
                               1
                                     True
                     samsung
                                                  Oppo
           ΜI
                               2
                                     False
                          mi
                                                    Μi
                               3
                                     False
         SONY
                        sony
                                                  Sony
dtype: object
               dtype: object
                               dtype: bool
                                             dtype: object
```

Pandas Series數值運算

```
1 numbers = pd.Series([22, 5, 10, 12, 6, 30])
2 print(numbers.max()) # 執行結果30
3 print(numbers.min()) #執行結果5
4 print(numbers.sum()) #執行結果85
5 print(numbers.mean()) # 執行結果14.1666666
6 print(numbers.nlargest(2)) # 最大的2個數值
7 print(numbers.nsmallest(2)) # 最小的2個數值
```

```
30
5
85
14.166666666666666
```

= 0

= a

```
5 30
0 22
dtype: int64
1 5
4 6
dtype: int64
```



Pandas DataFrame

雙維度資料

Created by 戴嘉熙 【小孫學堂】

什麼是Pandas DataFrame

20

= 0

S m

索引 内容1 内容2 内容3

	Α	В	С	D
1		name	math	chinese
2	0	Mike	80	63
3	1	Sherry	75	90
4	2	Cindy	93	85
5	3	John	86	70

建立Pandas DataFrame物件(字典) import pandas as pd

物件名稱 = pandas. DataFrame(資料串列)

```
name math chinese
0 Mike 80 63
1 Sherry 75 90
2 Cindy 93 85
3 John 86 70
```

2 0

= 0

= 0

= 0

= 0

= a

< 0</p>

建立Pandas DataFrame物件(清單)

20

20

S 10

20

23

= 0

23

= 0

= 0

= 31 = 3

 ≤ 0

50

```
grades = [
        ["Mike", 80, 63],
        ["Sherry", 75, 90],
        ["Cindy", 93, 85],
        ["John", 86, 70]
    new_df = pd.DataFrame(grades)
    print(new df)
        0
     Mike 80
               63
0
   Sherry 75 90
2
   Cindy 93 85
3
     John 86
               70
```

建立物件索引值與欄位名稱

建立物件索引值與欄位名稱 物件.index = ["key1", "key2", "key3"] 物件.columns = ["key1", "key2", "key3"]

```
grades = {
      "name": ["Mike", "Sherry", "Cindy", "John"],
3
      "math": [80, 75, 93, 86],
      "chinese": [63, 90, 85, 70]
5
6 df = pd.DataFrame(grades)
 |df.index = ["s1", "s2", "s3", "s4"] #自訂索引值
  df.columns = ["student_name", "math_score", "chinese_score"] #自訂欄位名稱
  print(df)
```

```
student name math score chinese score
           Mike
                          80
                                          63
s1
         Sherry
                                          90
s2
                          75
s3
         Cindy
                                          85
                          93
           John
                          86
                                          70
s4
```

物件.head(筆數)#順序取得前幾筆

```
grades = {
       "name": ["Mike", "Sherry", "Cindy", "John"],
       "math": [80, 75, 93, 86],
                                     原來的df
       "chinese": [63, 90, 85, 70]
                                               math chinese
                                          name
                                          Mike
                                                80
                                                        63
   df = pd.DataFrame(grades)
                                        Sherry
                                                75
                                                        90
   print("原來的df")
                                         Cindy
                                                        85
                                                93
   print(df)
                                          John
                                                86
                                                        70
   print("--
                                      取得最前面的兩筆資料
   new df = df.head(2)
                                               math chinese
                                          name
   print("取得最前面的兩筆資料")
                                          Mike
                                                80
                                                        63
   print(new_df)
12
                                        Sherry
                                                75
                                                        90
```

物件.tail(筆數)#逆序取得前幾筆

```
grades = {
        "name": ["Mike", "Sherry", "Cindy", "John"],
        "math": [80, 75, 93, 86],
        "chinese": [63, 90, 85, 70]
                                     原來的df
                                               math chinese
                                          name
                                          Mike
                                                 80
                                                        63
    df = pd.DataFrame(grades)
                                        Sherry
                                                75
                                                        90
    print("原來的df")
                                         Cindy
                                                        85
                                                 93
    print(df)
                                          John
                                                 86
                                                        70
    print("--
    new_df = df.tail(2)
                                      取得最後面的2筆資料
    print("取得最後面的2筆資料")
                                              math chinese
                                         name
                                        Cindy 93
                                                       85
=12
    print(new df)
                                         John
                                                86
                                                        70
```

20

500

-

5 0

20

20

= 3

= 0

= 3

20
20
20

= 3

S

物件["key"]

```
1 grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得單一欄位資料(型別為Series)")
8 print(df["name"])
```

```
取得單一欄位資料(型別為Series)
```

```
0 Mike
1 Sherry
2 Cindy
3 John
Name: name, dtype: object
```

物件[["key"]]

```
1 grades = {
2     "name": ["Mike", "Sherry", "Cindy", "John"],
3     "math": [80, 75, 93, 86],
4     "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得單一欄位資料(型別為DataFrame)")
8 print(df[["name"]])
```

```
取得單一欄位資料(型別為DataFrame)
name
0 Mike
```

1 Sherry

20

50

20

20

= 0

2 0

= 0

= 0

= 3 = 3

S 0

==

- 2 Cindy
- 3 John

物件[["key1", "key2"]]

```
1 grades = {
2     "name": ["Mike", "Sherry", "Cindy", "John"],
3     "math": [80, 75, 93, 86],
4     "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得多欄位資料(型別為DataFrame)")
8 print(df[["name", "chinese"]])
```

取得多欄位資料(型別為DataFrame)

name chinese
0 Mike 63
1 Sherry 90
2 Cindy 85
3 John 70

811

20

20 20 20

= 11

= 0

= 0

物件[起始索引值:結尾索引值]

```
1 grades = {
2     "name": ["Mike", "Sherry", "Cindy", "John"],
3     "math": [80, 75, 93, 86],
4     "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得索引值0~2的資料")
8 print(df[0:3])
```

取得索引值0~2的資料

```
name math chinese
0 Mike 80 63
1 Sherry 75 90
2 Cindy 93 85
```

20

= 0

< B

= 0

= 0

= 3

20 20 物件.at[key:"key"] 物件.iat[key:欄位索引值]

```
grades = {
       "name": ["Mike", "Sherry", "Cindy", "John"],
       "math": [80, 75, 93, 86],
       "chinese": [63, 90, 85, 70]
   df = pd.DataFrame(grades)
   print("利用at()方法取得索引值為1的math欄位資料")
   print(df.at[1, "math"])
   print("利用iat()方法取得索引值為1的第一個欄位資料")
   print(df.iat[1, 0])
10
```

利用at()方法取得索引值為1的math欄位資料 75 利用iat()方法取得索引值為1的第一個欄位資料 Sherry

物件.loc[[key1,key2],["key1", "key2"]]

```
1 grades = {
2    "name": ["Mike", "Sherry", "Cindy", "John"],
3    "math": [80, 75, 93, 86],
4    "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得資料索引值為1和3的name及chinese欄位資料集")
8 print(df.loc[[1, 3], ["name", "chinese"]])
```

取得資料索引值為1和3的name及chinese欄位資料集 name chinese

70

1 Sherry 90

John

物件.iloc[[key1,key2],[key1, key2]]

```
1 grades = {
2    "name": ["Mike", "Sherry", "Cindy", "John"],
3    "math": [80, 75, 93, 86],
4    "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 print("取得資料索引值為1和3的第一個及第三個欄位資料集")
8 print(df.iloc[[1, 3], [0, 2]])
```

取得資料索引值為1和3的第一個及第三個欄位資料集 name chinese 1 Sherry 90 3 John 70

新增DataFrame資料1

物件.insert(key,column='key',value=[值1,值2])

```
1 grades = {
2    "name": ["Mike", "Sherry", "Cindy", "John"],
3    "math": [80, 75, 93, 86],
4    "chinese": [63, 90, 85, 70]
5 }
6 df = pd.DataFrame(grades)
7 df.insert(2, column="engilsh", value=[88, 72, 74, 98])
8 print("在第三欄的地方新增一個欄位資料")
9 print(df)
```

在第三欄的地方新增一個欄位資料

20

50

= 0

= 0

= 3

==

¥ 0

	name	math	engilsh	chinese
0	Mike	80	88	63
1	Sherry	75	72	90
2	Cindy	93	74	85
3	John	86	98	70

新增DataFrame資料2

1 $grades = {$

物件.append({'key':'value'}, ignore_index=True)

```
"name": ["Mike", "Sherry", "Cindy", "John"],
3 "math": [80, 75, 93, 86],
   "chinese": [63, 90, 85, 70]
5
   df = pd.DataFrame(grades)
   new_df = df.append({
                          新增一筆資料
       "name": "Henry",
                                           chinese
                                     math
                               name
       "math": 60,
                               Mike
                                                63
                          0
                                       80
       "chinese": 62
                             Sherry 75
                                                90
   }, ignore index=True)
                          2
                              Cindy
                                       93
                                                85
   print("新增一筆資料")
                               John
                                       86
                                                70
   print(new df)
                          4
                              Henry
                                                62
                                       60
```

合併DataFrame資料

pd.concat([物件1,物件2], ignore_index=True)

```
grades = {
   "name": ["Mike", "Sherry", "Cindy", "John"],
                                               原來的df
   "math": [80, 75, 93, 86],
                                                        math
                                                            chinese
                                                   name
   "chinese": [63, 90, 85, 70]
                                                   Mike
                                                         80
                                                                 63
                                                 Sherry 75
                                                                 90
   df1 = pd.DataFrame(grades)
                                                  Cindy 93
                                                                 85
   print("原來的df")
                                                   John 86
                                                                 70
   print(df1)
                                               合併df來新增資料
   print("----")
                                                        math chinese
                                                   name
   df2 = pd.DataFrame({
10
                                                   Mike
                                                         80
                                                                 63
       "name": ["Henry"],
11
                                                 Sherry 75
                                                                 90
       "math": [60],
                                                  Cindy 93
                                                                 85
                                                   John
                                                        86
                                                                 70
       "chinese": [62]
                                                  Henry
                                                         60
                                                                 62
   })
14
15
   new_df = pd.concat([df1, df2], ignore_index=True)
   print("合併df來新增資料")
16
   print(new_df)
```

修改DataFrame資料

```
grades = {
   "name": ["Mike", "Sherry", "Cindy", "John"],
   "math": [80, 75, 93, 86],
   "chinese": [63, 90, 85, 70]
                                原來的df
                                             chinese
5
                                         math
                                    name
                                    Mike
                                           80
                                                   63
   df = pd.DataFrame(grades)
   print("原來的df")
                                  Sherry 75
                                                  90
                                   Cindy 93
                                                   85
   print(df)
                                    John
                                          86
                                                   70
   print("----")
   #修改索引值為1的math欄位資料
                                修改後的df
   df.at[1, "math"] = 100
11
                                        math chinese
                                   name
   #修改索引值為1的第一個欄位資料
12
                                   Mike
                                          80
                                                  63
   df.iat[1, 0] = "Larry"
13
                                         100
                                  Larry
                                                  90
   print("修改後的df")
14
                                  Cindy 93
                                                  85
   print(df)
                                   John
                                          86
                                                  70
```

物件.drop(["key"], axis=1)

```
grades = {
   "name": ["Mike", "Sherry", "Cindy", "John"],
   "math": [80, 75, 93, 86],
                                        原來的df
   "chinese": [63, 90, 85, 70]
                                                   chinese
                                           name
                                               math
                                           Mike
                                                 80
                                                        63
                                          Sherry
                                                 75
                                                        90
   df = pd.DataFrame(grades)
                                          Cindy
                                                 93
                                                        85
   print("原來的df")
                                           John
                                                 86
                                                        70
   print(df)
                                        刪除math欄位
   print("----")
                                           name
                                               chinese
   new_df = df.drop(["math"],axis=1)
10
                                       0
                                           Mike
                                                   63
                                          Sherry
                                                   90
   print("刪除math欄位")
                                          Cindy
                                                   85
   print(new df)
12
                                           John
                                                   70
```

物件.drop([key1,key2], axis=0)

```
grades = {
  "name": ["Mike", "Sherry", "Cindy", "John"],
   "math": [80, 75, 93, 86],
   "chinese": [63, 90, 85, 70]
                                  原來的df
                                           math
                                               chinese
                                      name
   df = pd.DataFrame(grades)
                                      Mike
                                             80
                                                    63
   print("原來的df")
                                    Sherry 75
                                                    90
                                     Cindy
   print(df)
                                             93
                                                    85
                                      John
                                             86
                                                    70
   print("----")
   new_df = df.drop([0, 3], axis=0)
10
                                  刪除第一筆及第四筆資料
   print("刪除第一筆及第四筆資料")
11
                                           math chinese
                                      name
   print(new df)
12
                                    Sherry
                                            75
                                                    90
                                     Cindy 93
                                                    85
```

物件.dropna()

```
grades = {
   "name": ["Mike", "Sherry", None, "John"],
   "math": [80, None, 93, 86],
   "chinese": [63, 90, 85, 70]
                                 原來的df
                                           math chinese
                                     name
   df = pd.DataFrame(grades)
                                     Mike 80.0
                                                    63
   print("原來的df")
                                   Sherry NaN
                                                    90
                                     None 93.0
                                                    85
   print(df)
                                     John 86.0
                                                    70
   new_df = df.dropna()
                                 刪除空值後的df
   print("刪除空值後的df")
10
                                    name math chinese
   print(new_df)
                                   Mike 80.0
                                                   63
                                    John 86.0
                                                   70
```

物件.drop_duplicates()

```
grades = {
   "name": ["Sherry", "Sherry", "Alan", "John"],
   "math": [90, 90, 93, 86],
   "chinese": [77, 77, 85, 70]
                                 原來的df
 5
                                           math
                                                 chinese
                                      name
   df = pd.DataFrame(grades)
                                   Sherry
                                             90
                                                      77
   print("原來的df")
                                    Sherry 90
                                                      77
   print(df)
                                     Alan
                                             93
                                                      85
   new_df = df.drop_duplicates()
                                      John
                                             86
                                                      70
   print("刪除重複值後的df")
                                 刪除重複值後的df
10
   print(new df)
                                           math
                                                 chinese
                                      name
                                    Sherry
                                             90
                                 0
                                                      77
                                      Alan
                                             93
                                                      85
                                      John
                                             86
                                                      70
```

DataFrame篩選1

物件[物件["key"]>值]

```
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("篩選math大於80的資料集")
print(df[df["math"] > 80])
```

```
原來的df
         math
              chinese
    name
    Mike
0
           80
                   63
  Sherry
           75
                   90
   Cindy
           93
                   85
    John
           86
                   70
篩選math大於80的資料集
        math chinese
   name
  Cindy
          93
                  85
3
   John
          86
                  70
```

DataFrame篩選2

S 0

20

物件[物件["key"].isin(["key"])]

```
grades = {
   | "name": ["Mike", "Sherry", "Cindy", "John"],
   "math": [80, 75, 93, 86],
   "chinese": [63, 90, 85, 70]
                                        原來的df
                                                  math chinese
   df = pd.DataFrame(grades)
                                            name
   print("原來的df")
                                            Mike
                                                    80
                                                             63
   print(df)
                                          Sherry 75
                                                             90
   print("篩選name欄位包含John的資料集")
                                           Cindy
                                                    93
                                                             85
   print(df[df["name"].isin(["John"])])
                                            John
                                                    86
                                                             70
                                        篩選name欄位包含John的資料集
5 0
                                                math chinese
                                          name
                                          John
                                        3
                                                  86
                                                           70
2 10
```

DataFrame排序1

物件.sort_index(ascending=0 or 1)

```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
                                              原來的df
                                                              chinese
                                                        math
                                                   name
    "chinese": [63, 90, 85, 70]
                                              s3
                                                   Mike
                                                          80
                                                                  63
 5
                                                  Sherry
                                                         75
                                                                  90
                                              s1
    df = pd.DataFrame(grades)
                                                  Cindy
                                                          93
                                                                  85
                                              s4
    df.index = ["s3", "s1", "s4", "s2"]
                                              s2
                                                   John
                                                          86
                                                                  70
                                              遞增排序
    print("原來的df")
                                                              chinese
                                                        math
                                                   name
    print(df)
                                              s1
                                                  Sherry
                                                          75
                                                                  90
    new df = df.sort index(ascending=1)
                                                          86
                                                   John
                                                                  70
10
                                                   Mike
                                                          80
                                                                  63
                                              s3
    print("遞增排序")
11
                                              s4
                                                  Cindy
                                                          93
                                                                  85
    print(new df)
12
                                              遞減排序
    new_df = df.sort_index(ascending=0)
13
                                                              chinese
                                                        math
                                                   name
                                                  Cindv
                                              s4
                                                          93
                                                                  85
    print("遞減排序")
                                              s3
                                                   Mike
                                                          80
                                                                  63
    print(new df)
15
                                                   John
                                                          86
                                              s2
                                                                  70
                                              s1
                                                  Sherry
                                                          75
                                                                  90
```

DataFrame排序2

物件.sort_values(["key"],ascending=0 or 1)

```
原來的df
                                                            math chinese
   grades = {
                                                      name
                                                      Mike
   "name": ["Mike", "Sherry", "Cindy", "John"],
                                                s3
                                                              80
                                                                       63
   "math": [80, 75, 93, 86],
                                                s1
                                                    Sherry
                                                              75
                                                                       90
   "chinese": [63, 90, 85, 70]
                                                     Cindy
                                                s4
                                                              93
                                                                       85
5
                                                      John
                                                              86
                                                                       70
                                                s2
   df = pd.DataFrame(grades)
                                                遞增排序
   df.index = ["s3", "s1", "s4", "s2"]
                                                            math chinese
                                                      name
   print("原來的df")
                                                    Sherry
                                                s1
                                                              75
                                                                       90
   print(df)
                                                s3
                                                      Mike
                                                              80
                                                                       63
10
   new df = df.sort values(["math"], ascending=1)
                                                s2
                                                      John
                                                              86
                                                                       70
   print("遞增排序")
                                                     Cindy
                                                                       85
                                                s4
                                                              93
   print(new_df)
                                                遞減排序
   new_df = df.sort_values(["math"], ascending=0)
                                                            math chinese
                                                      name
   print("遞減排序")
14
                                                     Cindy
   print(new_df)
                                                              93
                                                                       85
                                                s4
                                                s2
                                                      John
                                                              86
                                                                       70
                                                      Mike
                                                                       63
                                                s3
                                                              80
                                                    Sherry
                                                                       90
                                                s1
                                                              75
```



Pandas read_excel 讀取Excel檔案

Created by 戴嘉熙 【小孫學堂】

讀取與匯出Excel

import pandas as pd

物件名稱 = pd.read_excel("檔案路徑")

pd.to_excel("檔案路徑")

修正中文對齊問題

20

20

= 0

= 0

= 0

5 0

pd.set_option('display.unicode.ambiguous_as_wide', True) pd.set_option('display.unicode.east_asian_width', True)

工作表讀取

import pandas as pd

物件名稱 = pd.read_excel('檔案路徑 ',sheet_name='工作表名稱')

讀取所有工作表:

sheet_name=None

讀取多個:

= 3

= 0

sheet_name=['工作表1',' 工作表2'...]

欄、列讀取

import pandas as pd

物件名稱 = pd.read_excel('檔案路徑', usecols='欄位名稱')

欄位名稱可替換為索引值或欄位編號

usecols=[1,2,3]或usecols='A,B,C'

取範圍使用:

限制筆數:nrows=列數量