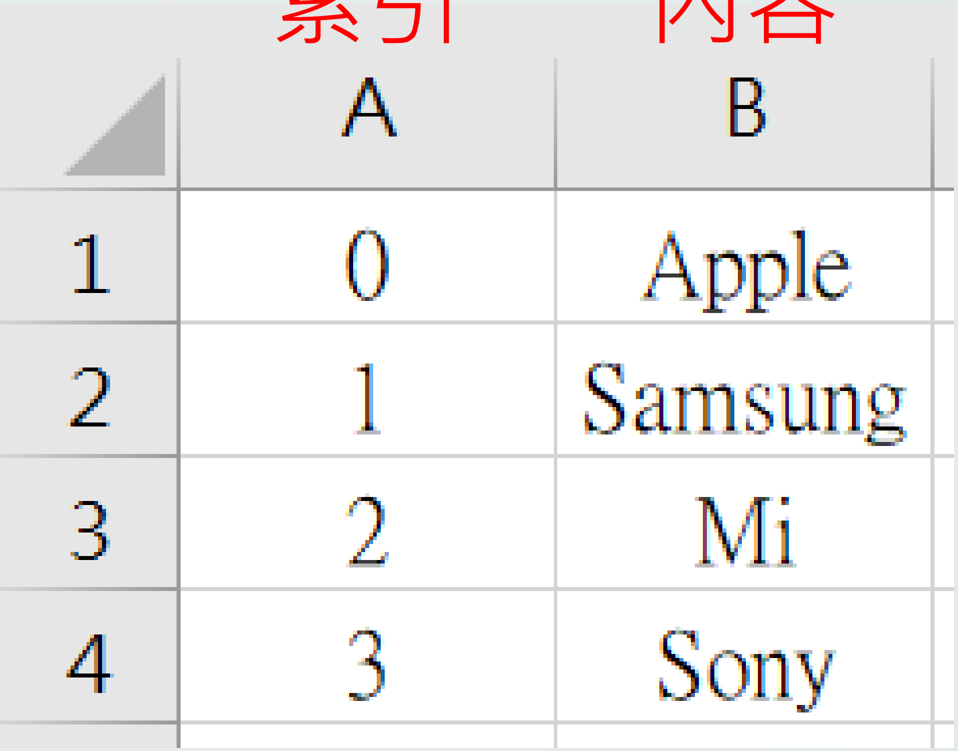
STLectures
小孫學堂

Pandas Series

單維度資料

Created by 戴嘉熙 【小孫學堂】

什麼是Pandas Series



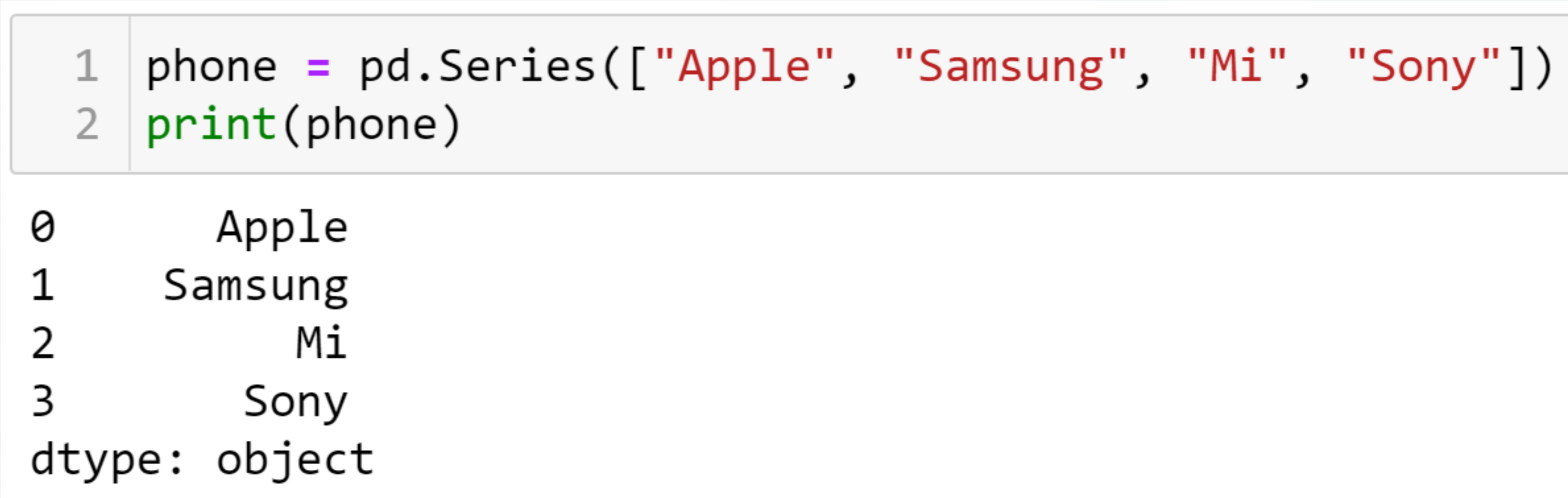
索引 內容

	A	B
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)

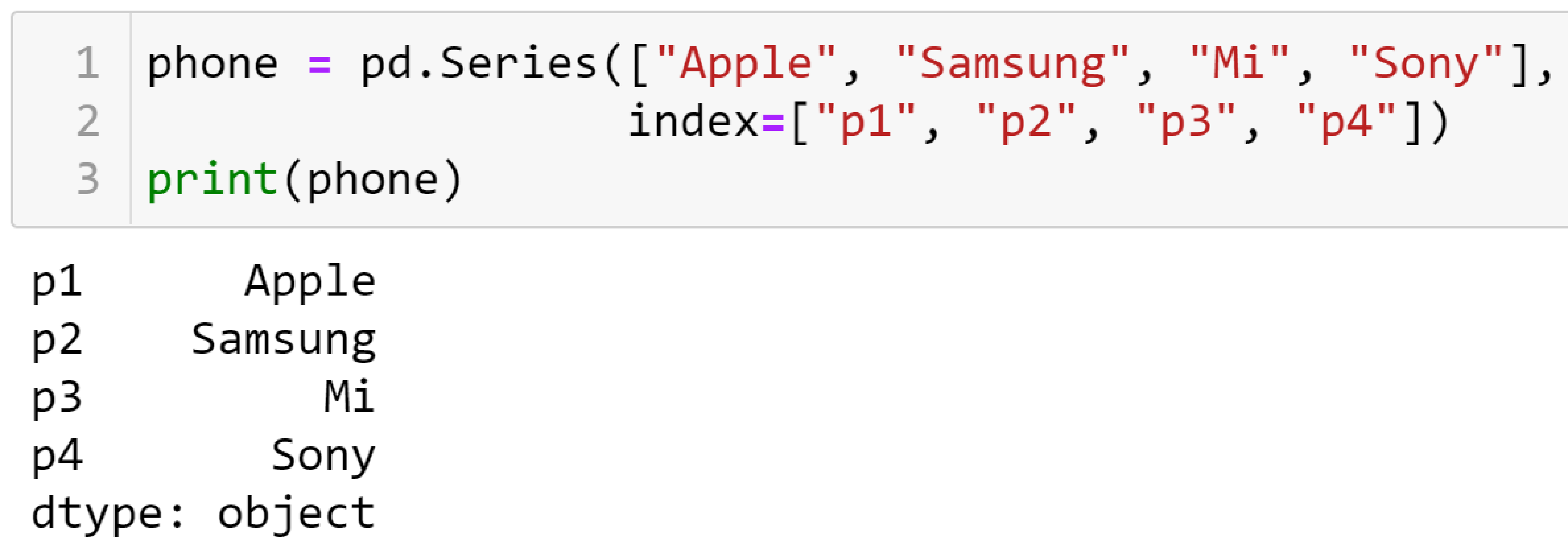


```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
print(phone)
```

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

建立物件索引值

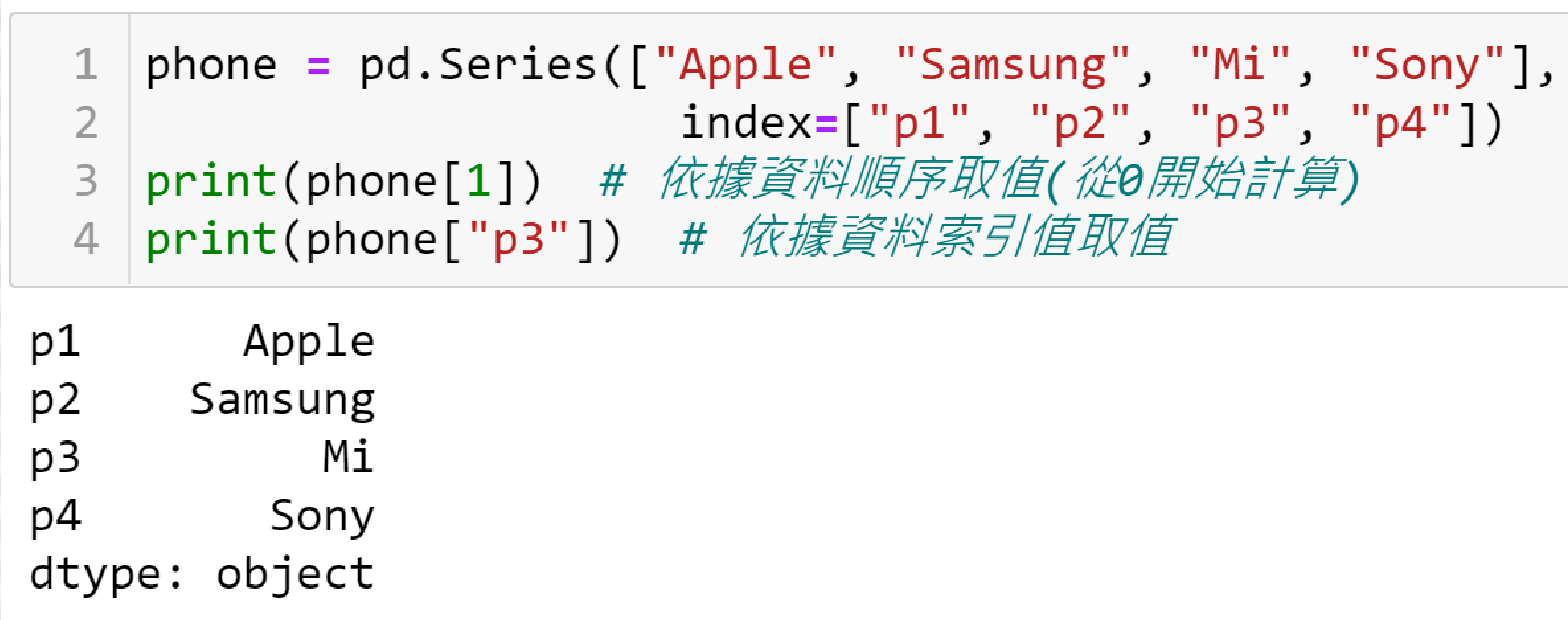
物件.index = [“key1”, “key2”, “key3”]



```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"],
                  index=["p1", "p2", "p3", "p4"])
print(phone)
```

```
p1        Apple
p2      Samsung
p3           Mi
p4         Sony
dtype: object
```

取得Pandas Series資料

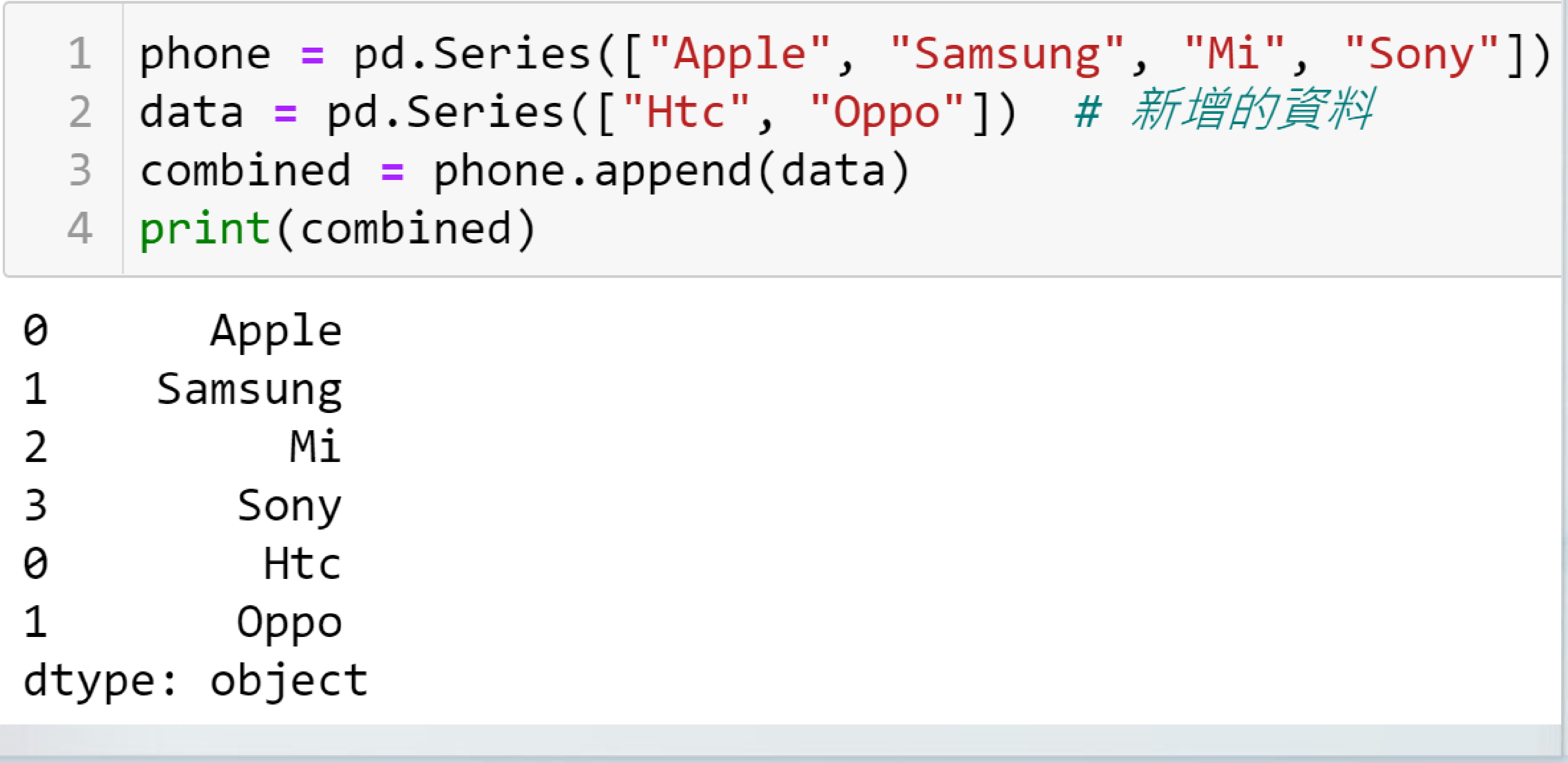


```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"],
                                    index=["p1", "p2", "p3", "p4"])
print(phone[1])  # 依據資料順序取值(從0開始計算)
print(phone["p3"])  # 依據資料索引值取值
```

```
p1        Apple
p2      Samsung
p3           Mi
p4         Sony
dtype: object
```

合併Pandas Series資料

物件1 = 物件2.append(物件3)

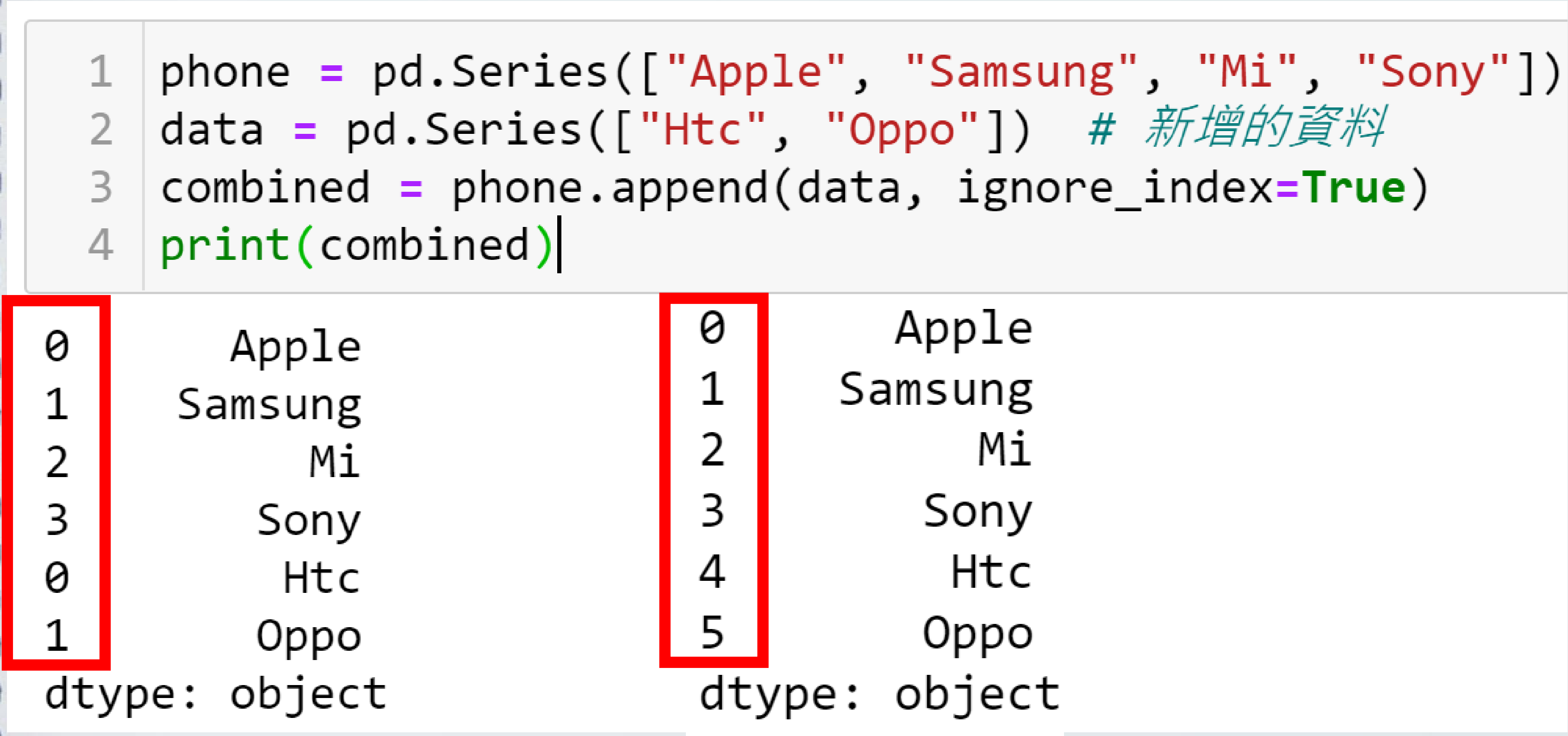


```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
data = pd.Series(["Htc", "Oppo"])  # 新增的資料
combined = phone.append(data)
print(combined)
```

```
0      Apple
1    Samsung
2         Mi
3       Sony
0        Htc
1       Oppo
dtype: object
```

ignore_index=True or False

是否忽略原始索引值

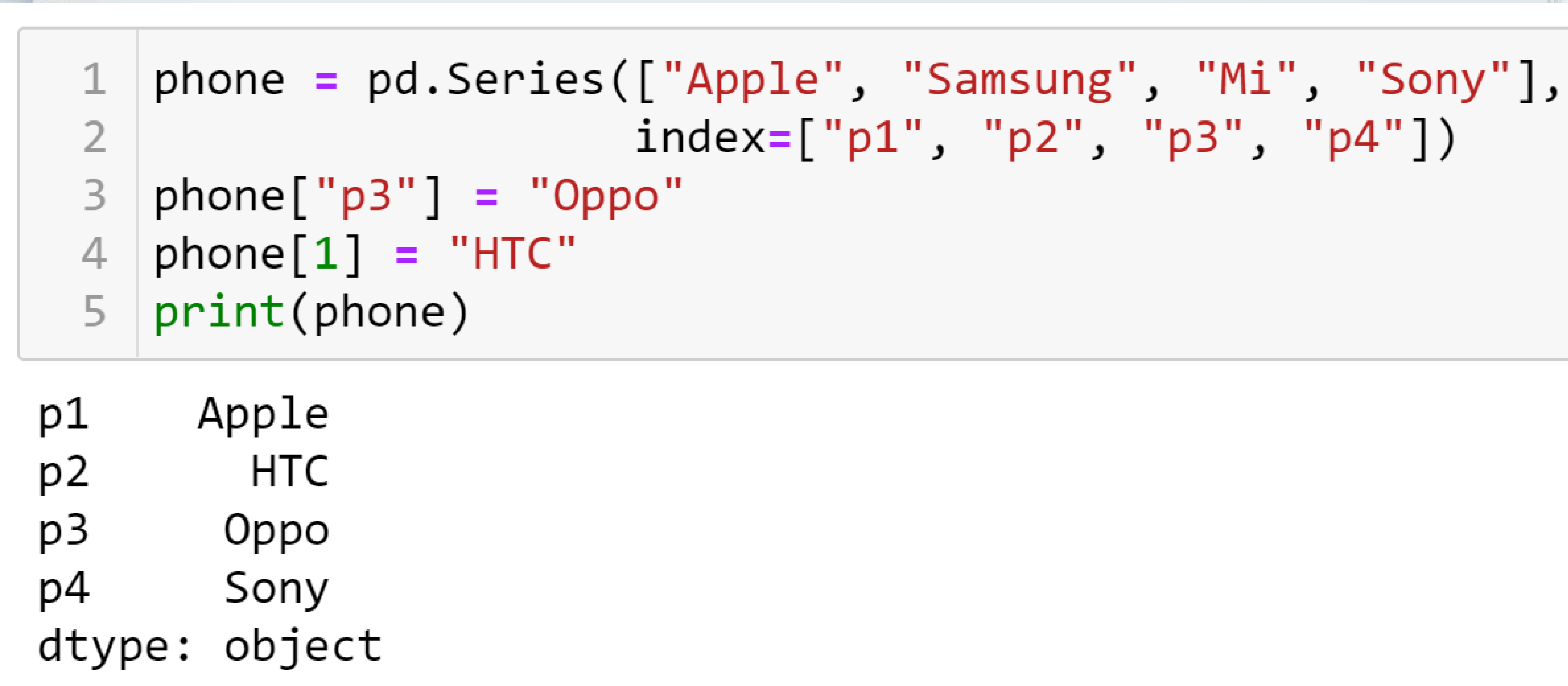


```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
data = pd.Series(["Htc", "Oppo"])  # 新增的資料
combined = phone.append(data, ignore_index=True)
print(combined)
```

```
0      Apple
1    Samsung
2         Mi
3       Sony
0        Htc
1       Oppo
dtype: object
```

```
0      Apple
1    Samsung
2         Mi
3       Sony
4        Htc
5       Oppo
dtype: object
```

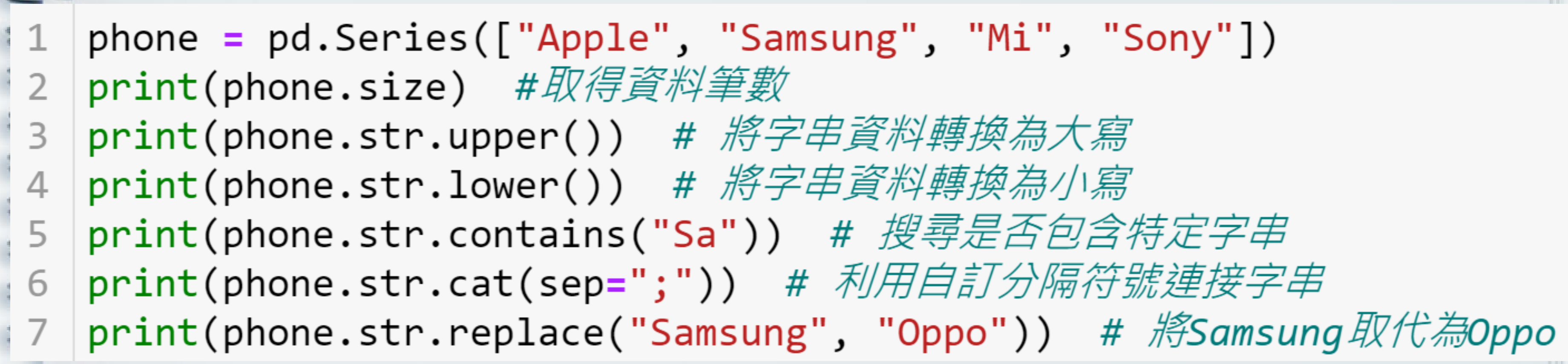
修改Pandas Series資料



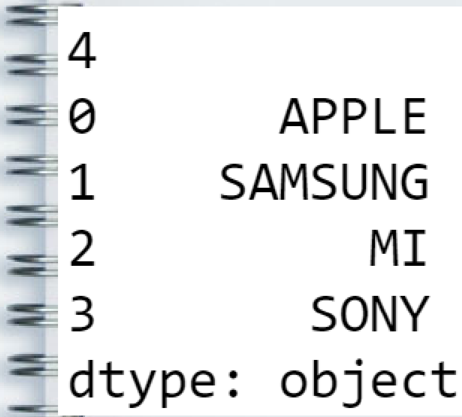
```python
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"],
                  index=["p1", "p2", "p3", "p4"])
phone["p3"] = "Oppo"
phone[1] = "HTC"
print(phone)
```

```
p1    Apple
p2      HTC
p3     Oppo
p4     Sony
dtype: object
```

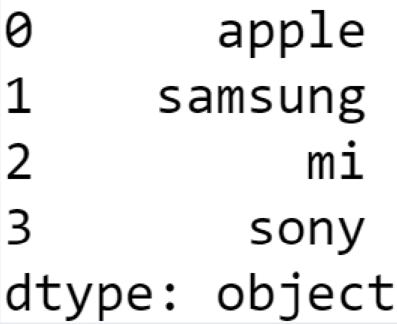
Pandas Series字串運算



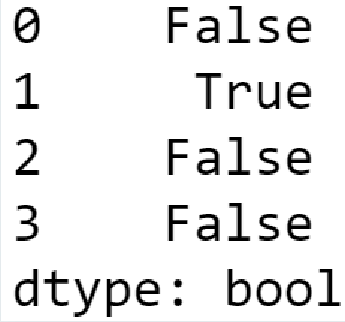
```
phone = pd.Series(["Apple", "Samsung", "Mi", "Sony"])
print(phone.size)  #取得資料筆數
print(phone.str.upper())  # 將字串資料轉換為大寫
print(phone.str.lower())  # 將字串資料轉換為小寫
print(phone.str.contains("Sa"))  # 搜尋是否包含特定字串
print(phone.str.cat(sep=";"))  # 利用自訂分隔符號連接字串
print(phone.str.replace("Samsung", "Oppo"))  #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



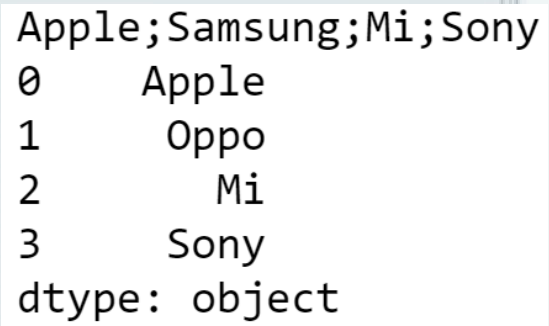
```
4
0      APPLE
1    SAMSUNG
2         MI
3       SONY
dtype: object
```



```
0      apple
1    samsung
2         mi
3       sony
dtype: object
```

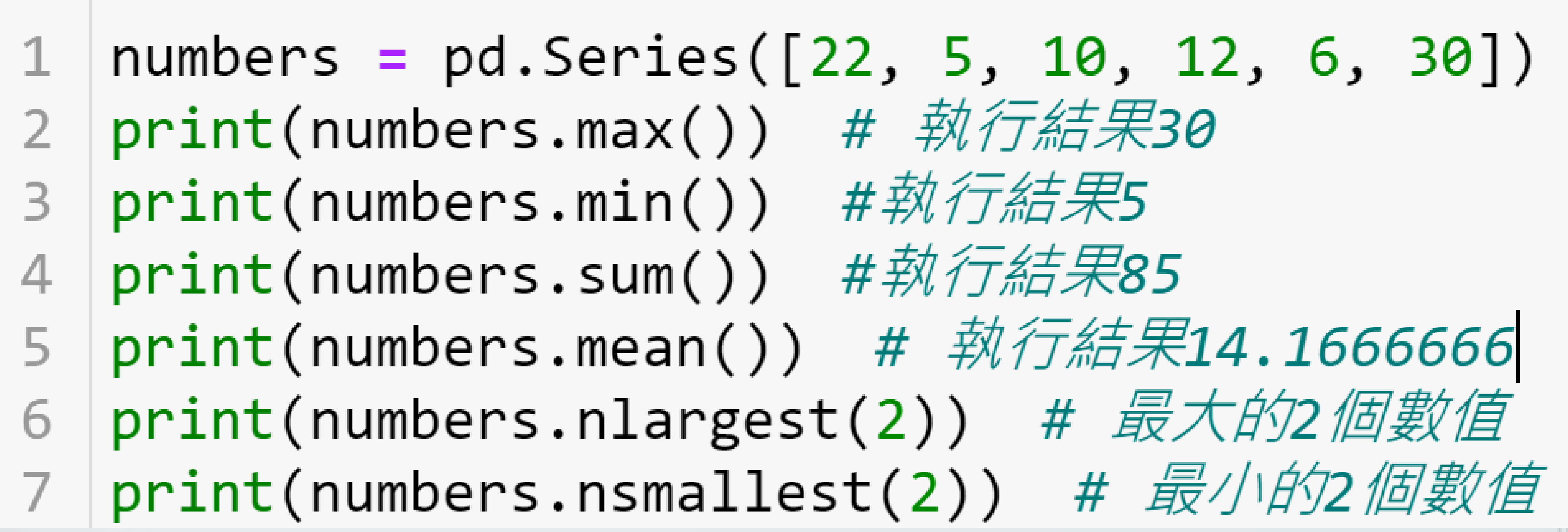


```
0    False
1     True
2    False
3    False
dtype: bool
```

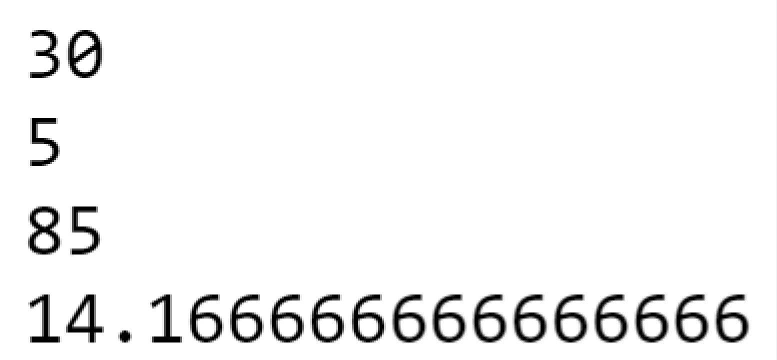


```
Apple;Samsung;Mi;Sony
0    Apple
1     Oppo
2       Mi
3     Sony
dtype: object
```

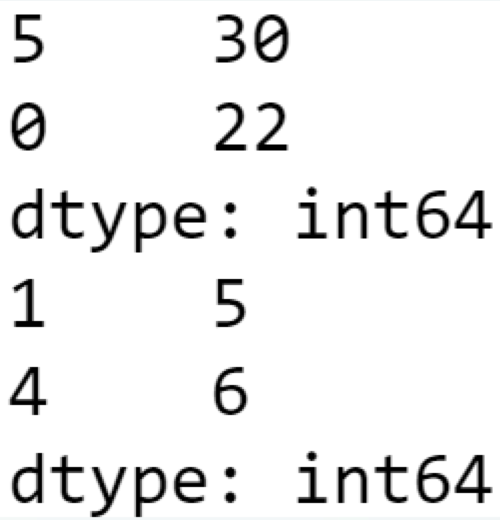
Pandas Series數值運算



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



```
30
5
85
14.166666666666666
```



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

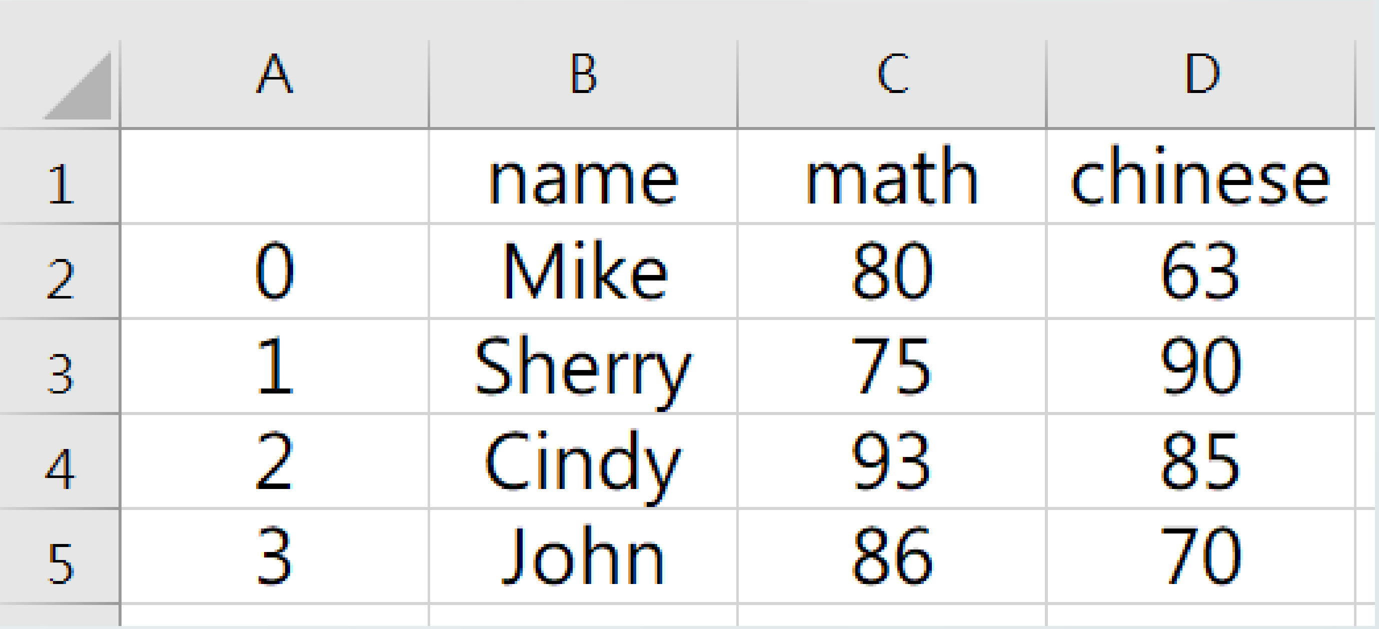
STLECTURES
小孫學堂

Pandas DataFrame
雙維度資料

Created by 戴嘉熙 【小孫學堂】

什麼是Pandas DataFrame

索引　內容1　內容2　內容3

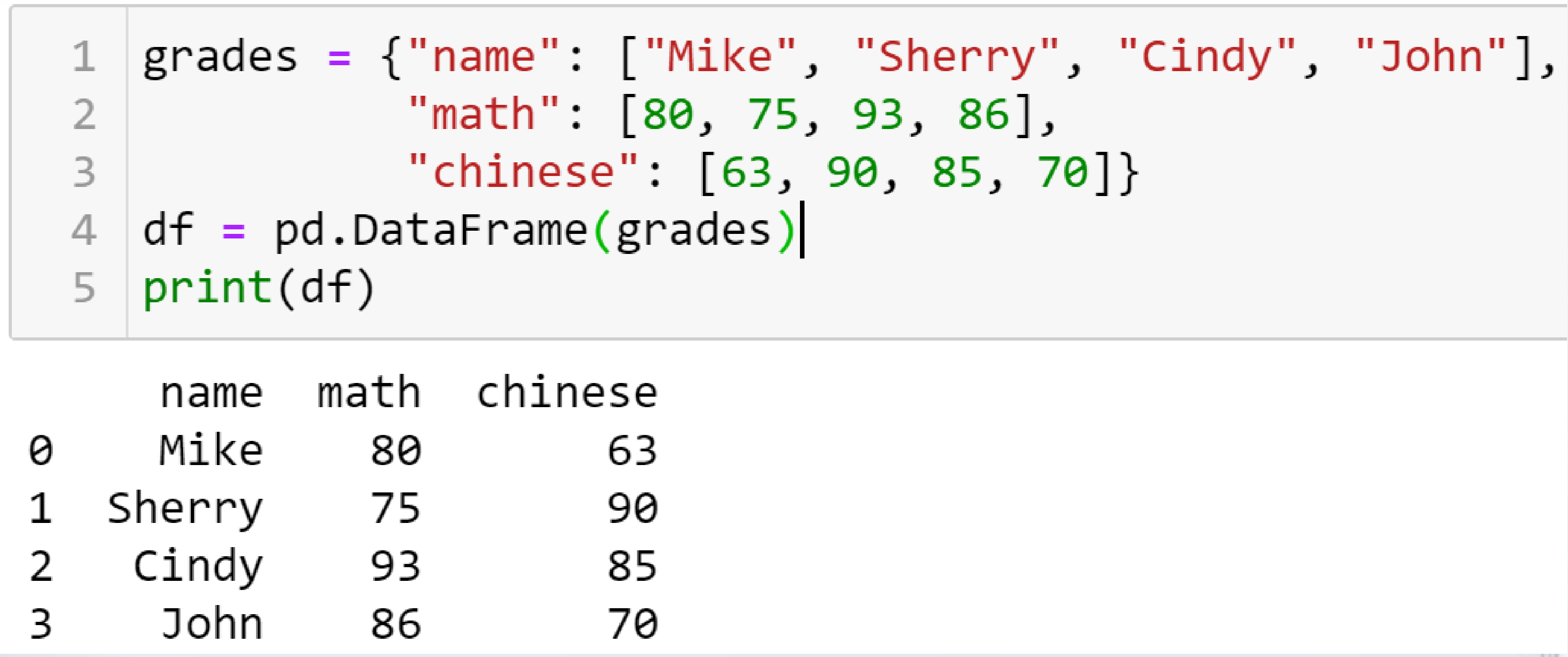


	A	B	C	D
1		name	math	chinese
2	0	Mike	80	63
3	1	Sherry	75	90
4	2	Cindy	93	85
5	3	John	86	70

建立Pandas DataFrame物件(字典)

import pandas as pd

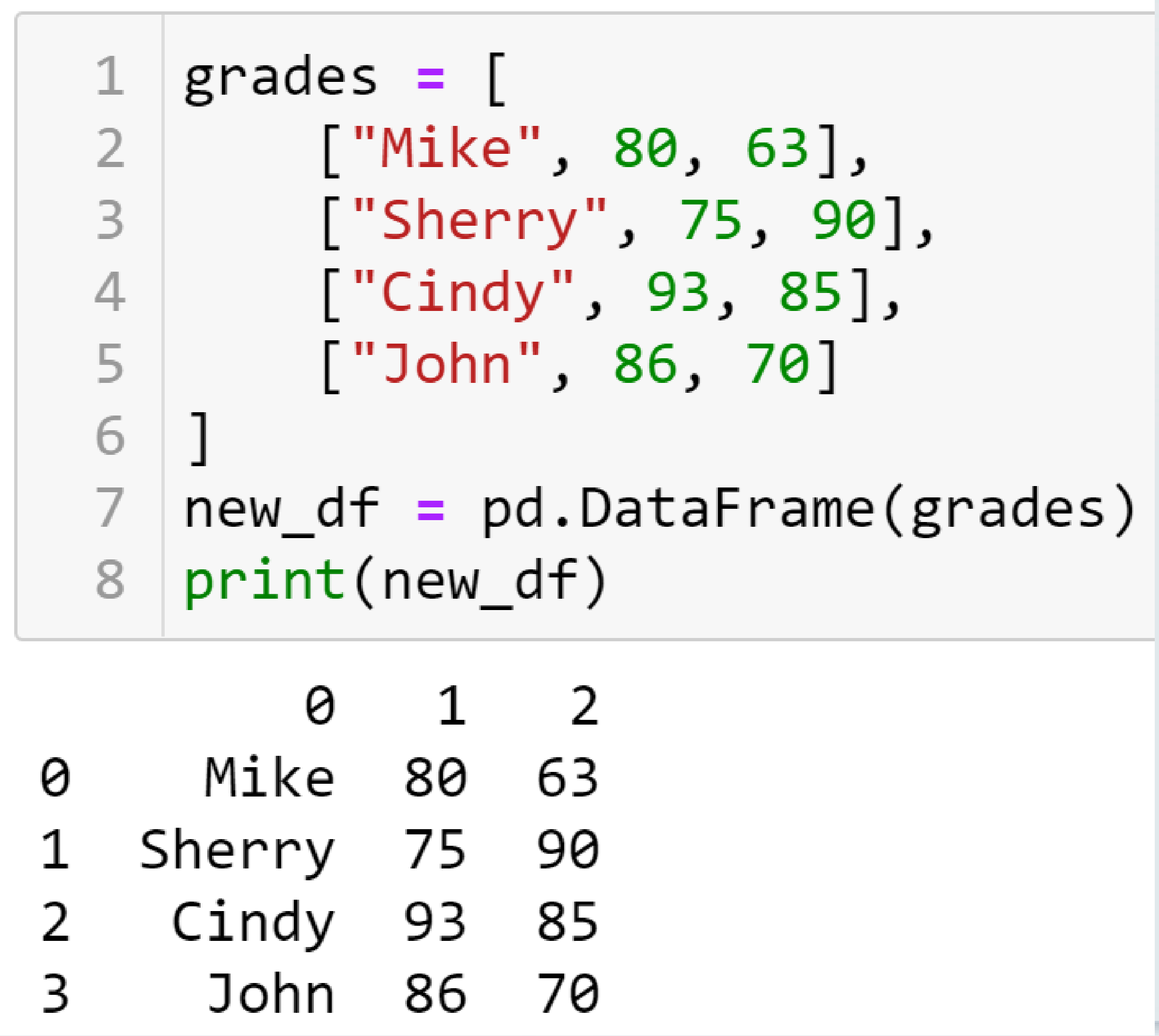
物件名稱 = pandas. DataFrame(資料串列)



```
grades = {"name": ["Mike", "Sherry", "Cindy", "John"],
          "math": [80, 75, 93, 86],
          "chinese": [63, 90, 85, 70]}
df = pd.DataFrame(grades)
print(df)
```

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

建立Pandas DataFrame物件(清單)



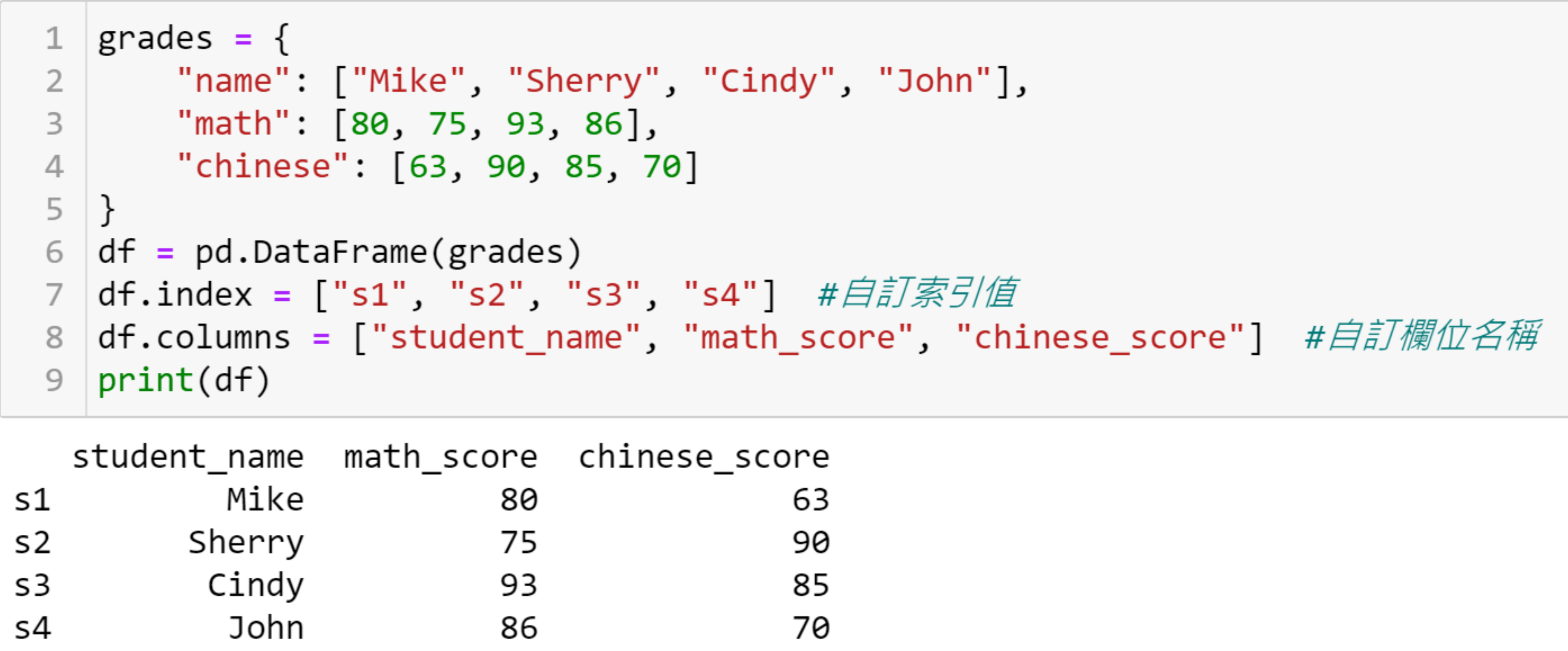
```python
grades = [
    ["Mike", 80, 63],
    ["Sherry", 75, 90],
    ["Cindy", 93, 85],
    ["John", 86, 70]
]
new_df = pd.DataFrame(grades)
print(new_df)
```

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

建立物件索引值與欄位名稱

物件.index = [“key1”, “key2”, “key3”]

物件.columns = [“key1”, “key2”, “key3”]

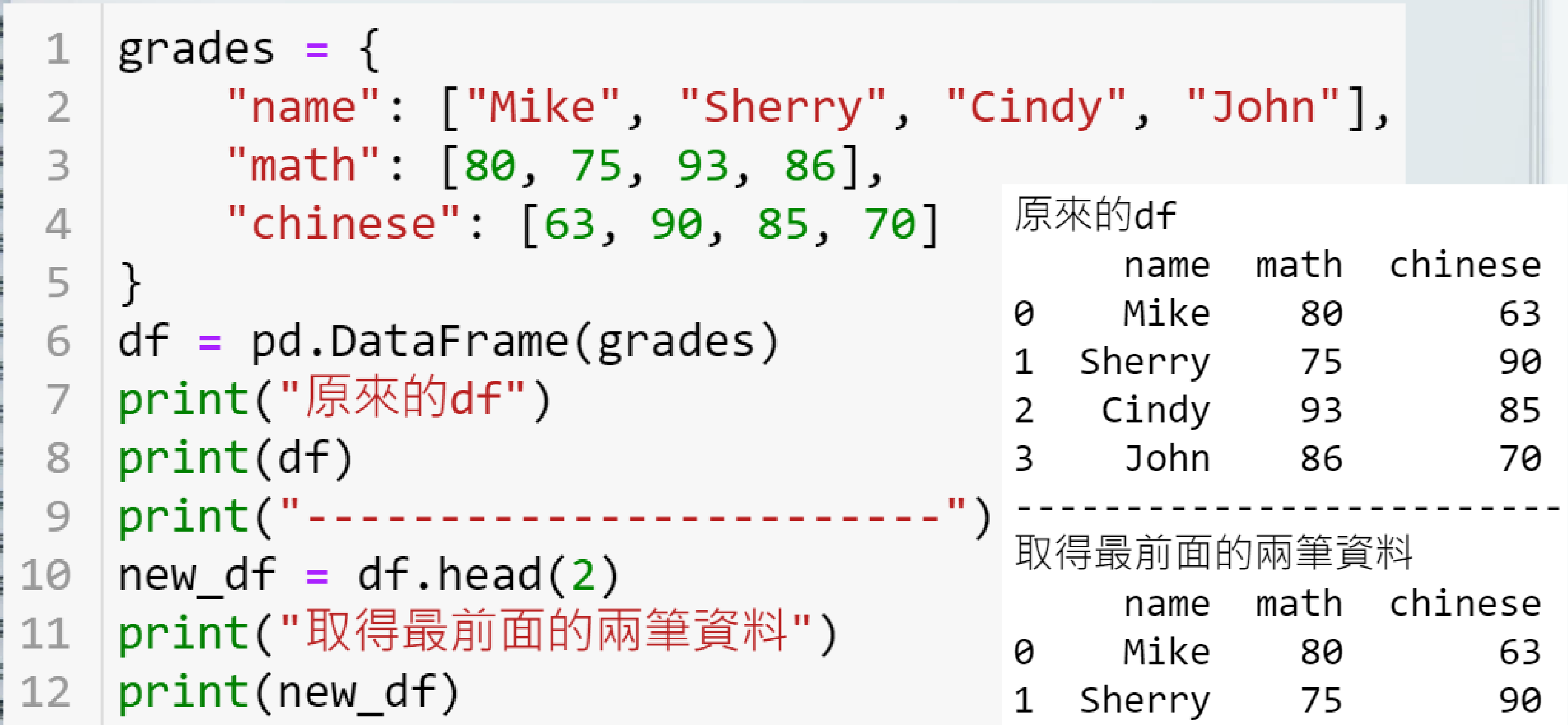


```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
df.index = ["s1", "s2", "s3", "s4"]  #自訂索引值
df.columns = ["student_name", "math_score", "chinese_score"]  #自訂欄位名稱
print(df)
```

```
   student_name  math_score  chinese_score
s1         Mike          80             63
s2       Sherry          75             90
s3        Cindy          93             85
s4         John          86             70
```

取得DataFrame資料1

物件.head(筆數)#順序取得前幾筆

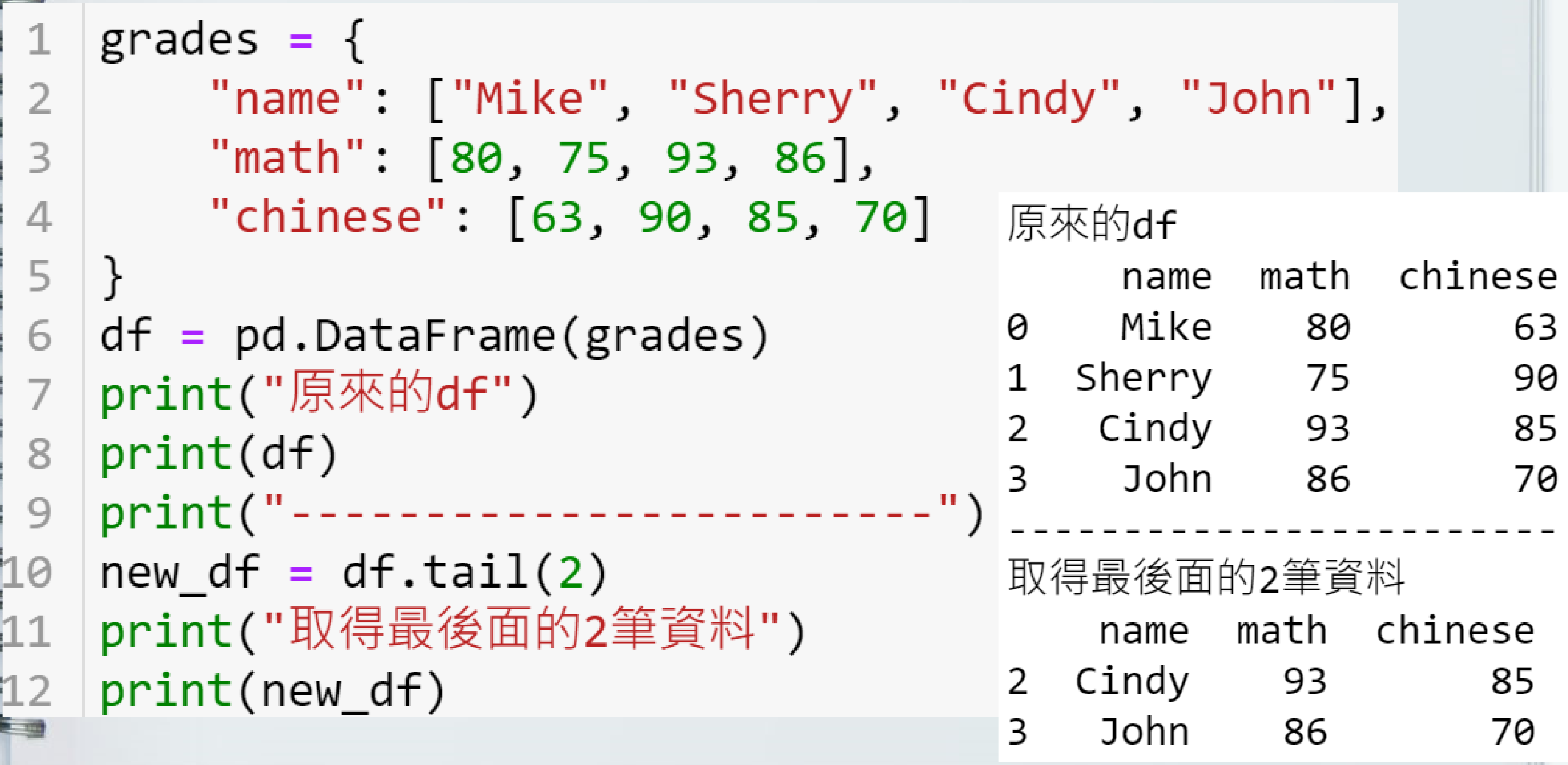


```python
grades = {
     "name": ["Mike", "Sherry", "Cindy", "John"],
     "math": [80, 75, 93, 86],
     "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("------------------------")
new_df = df.head(2)
print("取得最前面的兩筆資料")
print(new_df)
```

```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
------------------------
取得最前面的兩筆資料
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
```

取得DataFrame資料2

物件.tail(筆數)#逆序取得前幾筆

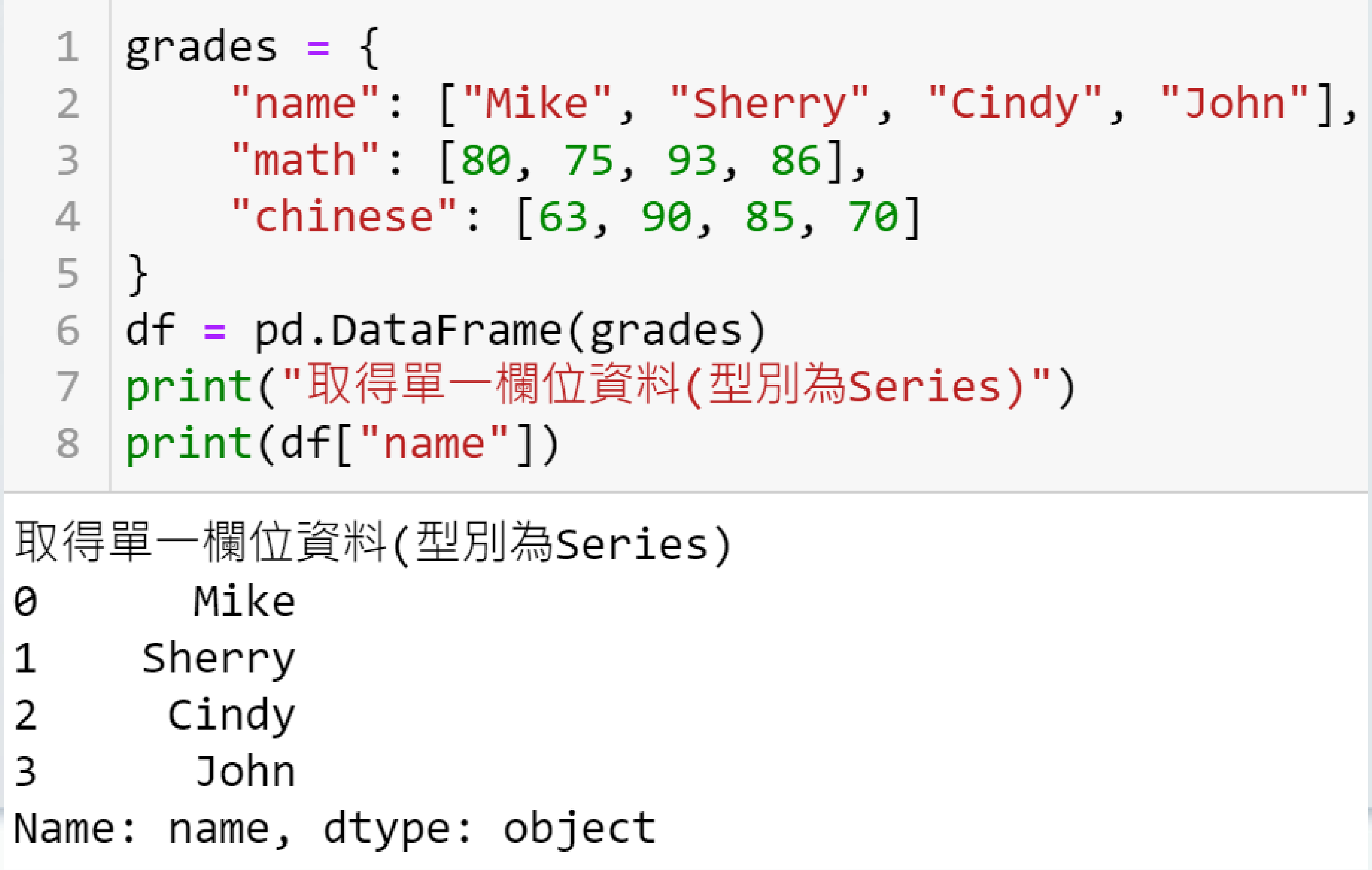


```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("------------------------")
new_df = df.tail(2)
print("取得最後面的2筆資料")
print(new_df)
```

```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
------------------------
取得最後面的2筆資料
    name  math  chinese
2  Cindy    93       85
3   John    86       70
```

取得DataFrame資料3

物件[“key”]

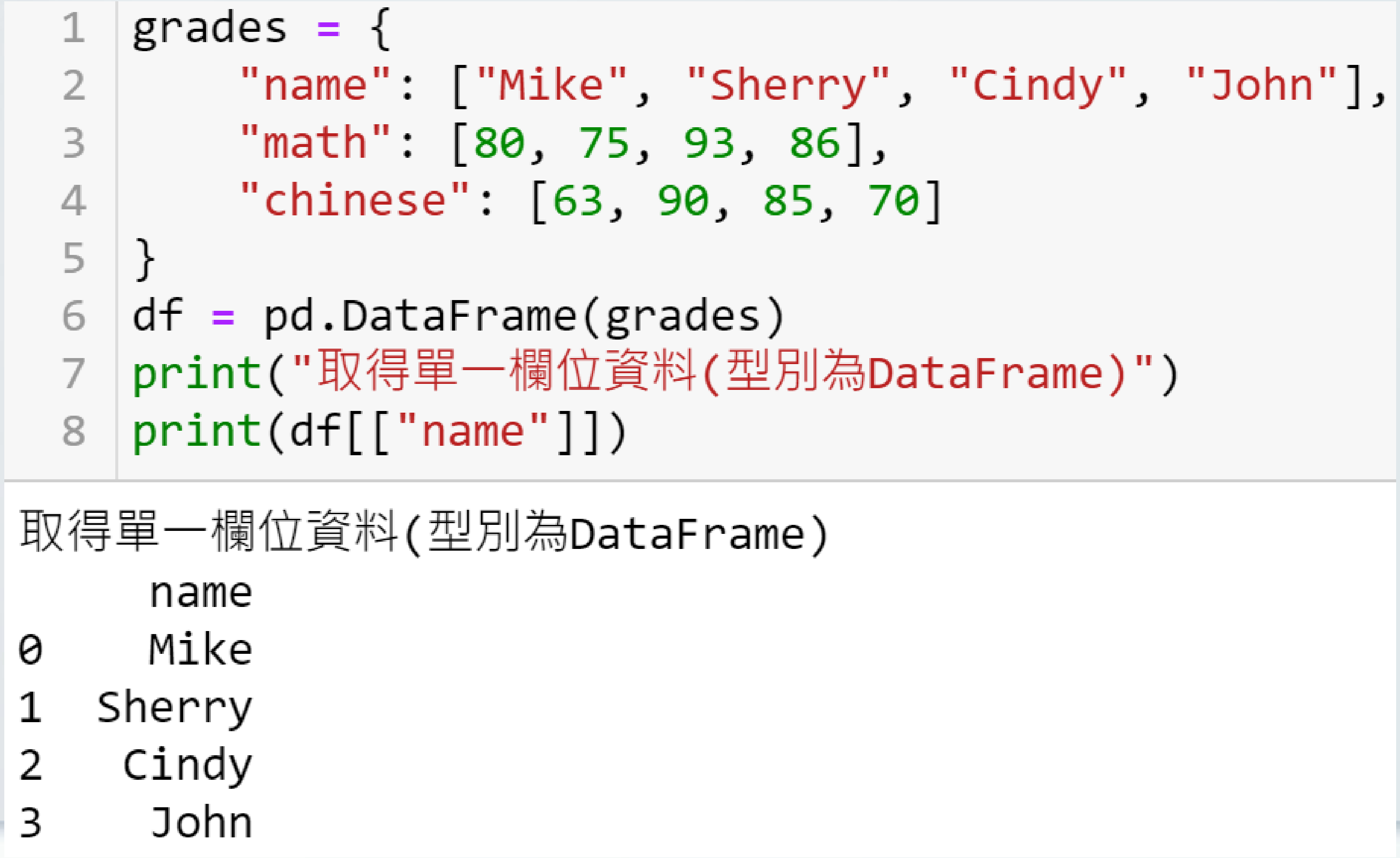


```python
grades = {
     "name": ["Mike", "Sherry", "Cindy", "John"],
     "math": [80, 75, 93, 86],
     "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得單一欄位資料(型別為Series)")
print(df["name"])
```

```
取得單一欄位資料(型別為Series)
0      Mike
1    Sherry
2     Cindy
3      John
Name: name, dtype: object
```

取得DataFrame資料4

物件[[“key”]]

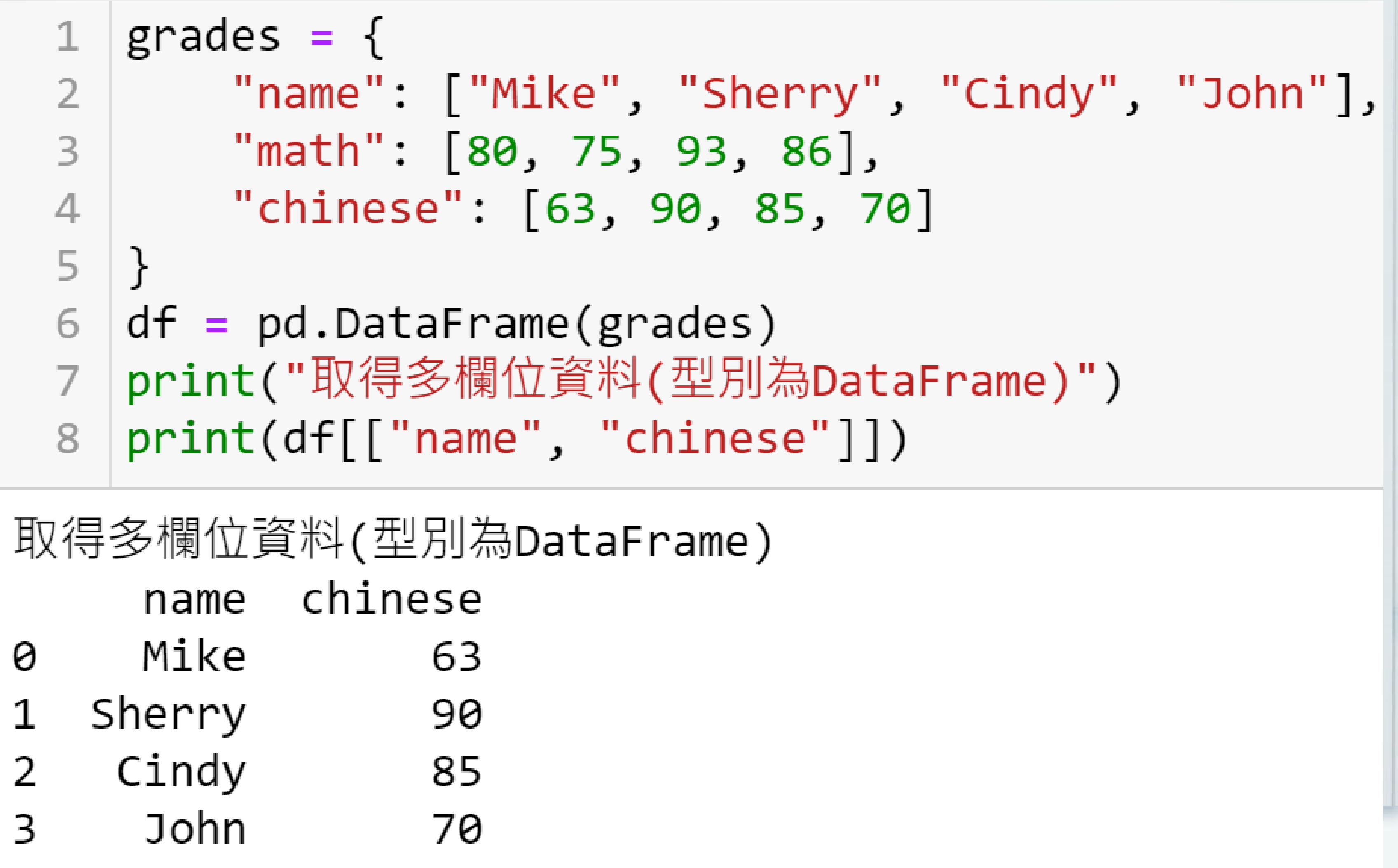


```python
grades = {
     "name": ["Mike", "Sherry", "Cindy", "John"],
     "math": [80, 75, 93, 86],
     "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得單一欄位資料(型別為DataFrame)")
print(df[["name"]])
```

```
取得單一欄位資料(型別為DataFrame)
     name
0    Mike
1  Sherry
2   Cindy
3    John
```

取得DataFrame資料5

物件[[“key1”, “key2”]]

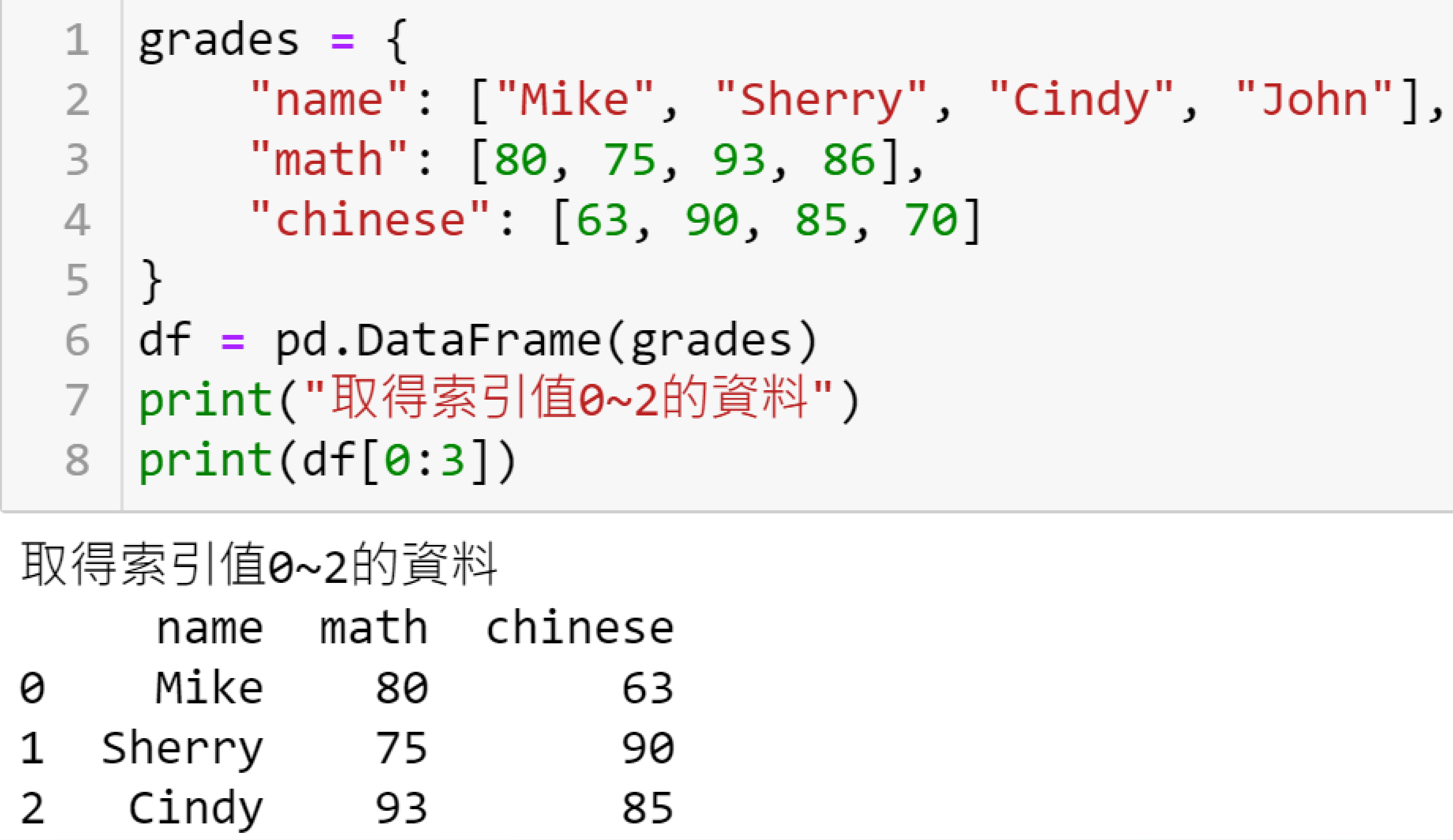


```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得多欄位資料(型別為DataFrame)")
print(df[["name", "chinese"]])
```

```
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
```

取得DataFrame資料6

物件[起始索引值:結尾索引值]

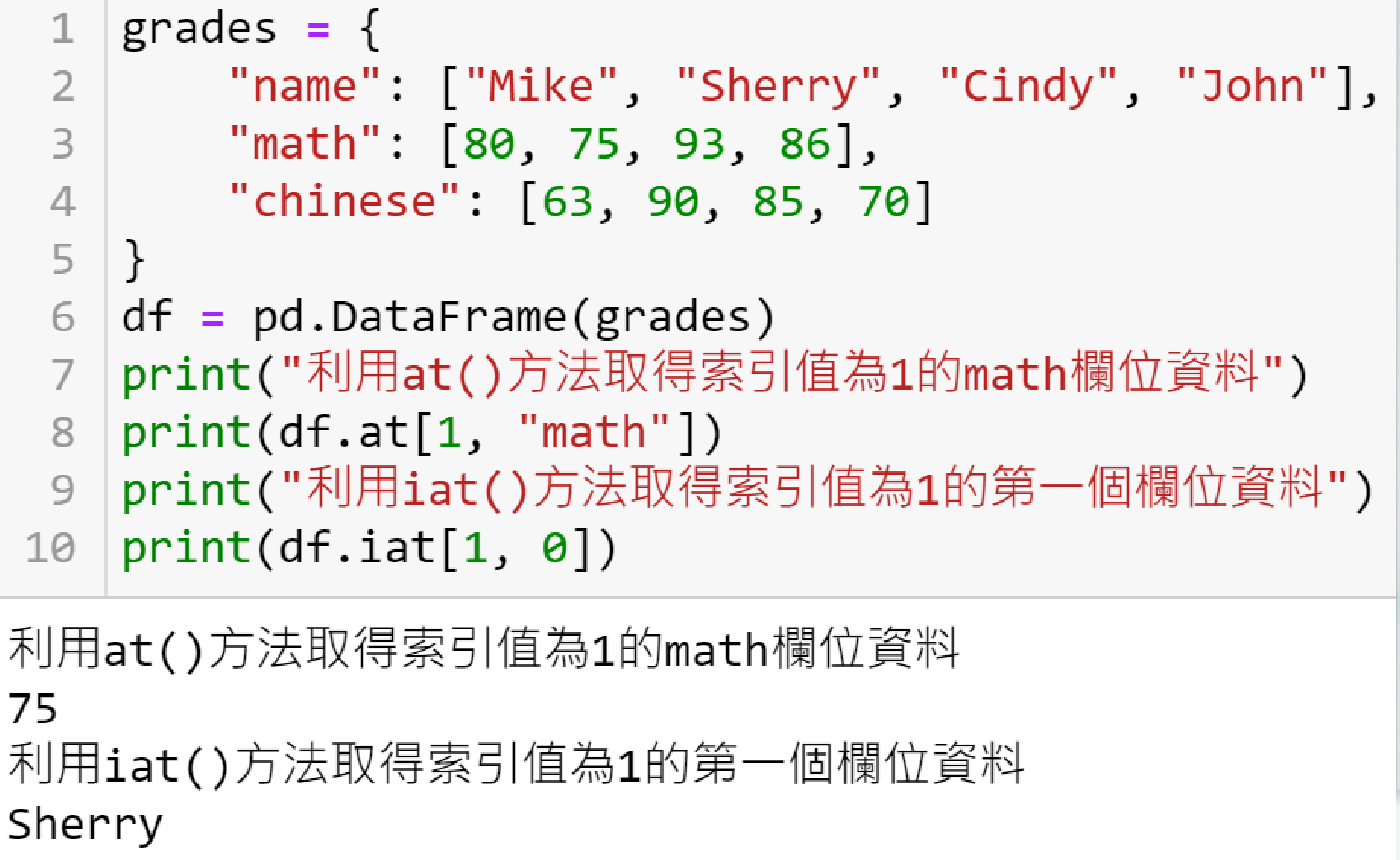


```python
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得索引值0~2的資料")
print(df[0:3])
```

```
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
```

取得DataFrame資料7

物件.at[key:”key”]   物件.iat[key:欄位索引值]

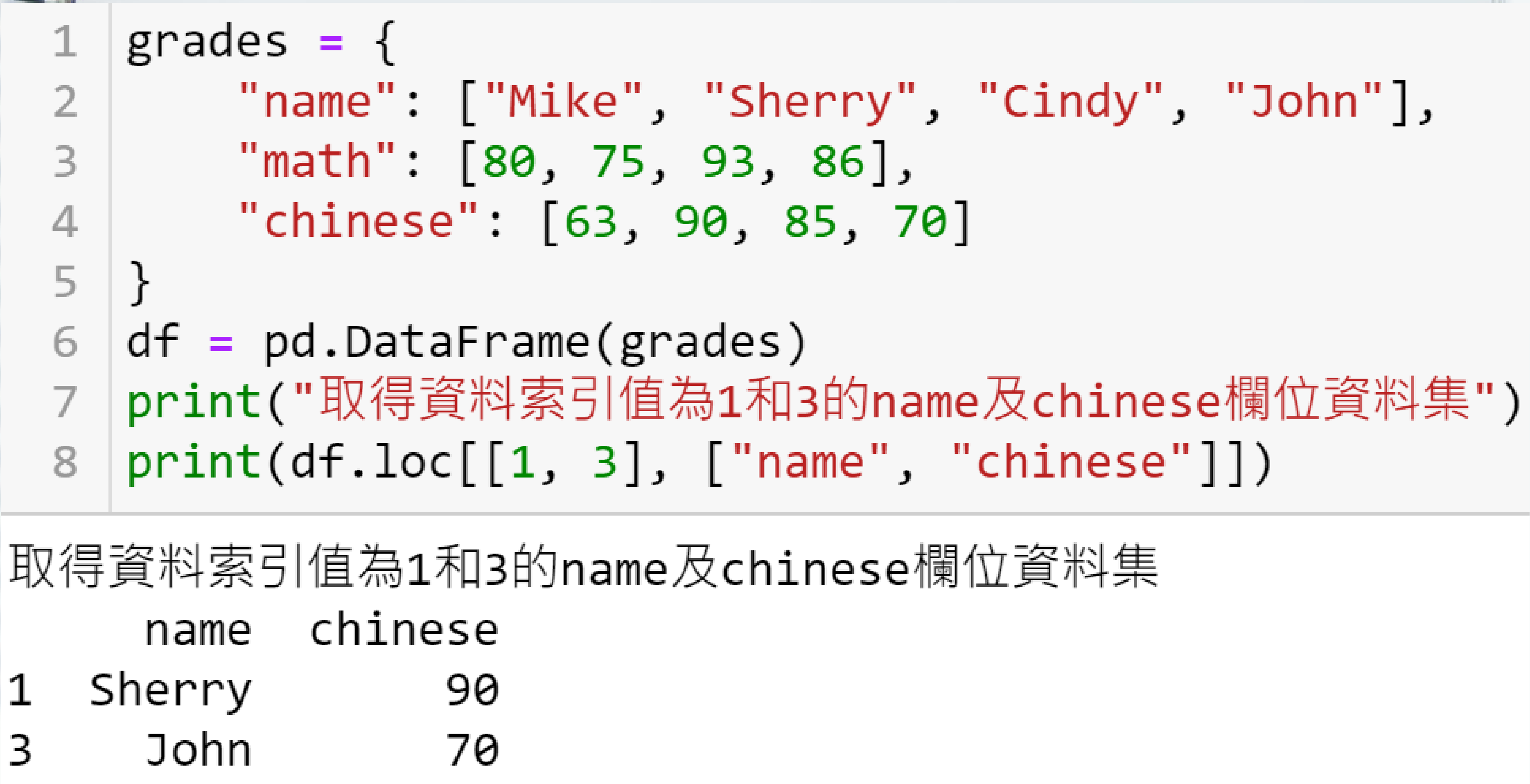


```python
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("利用at()方法取得索引值為1的math欄位資料")
print(df.at[1, "math"])
print("利用iat()方法取得索引值為1的第一個欄位資料")
print(df.iat[1, 0])
```

```
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
```

取得DataFrame資料8

物件.loc[[key1,key2],[“key1”, “key2”]]

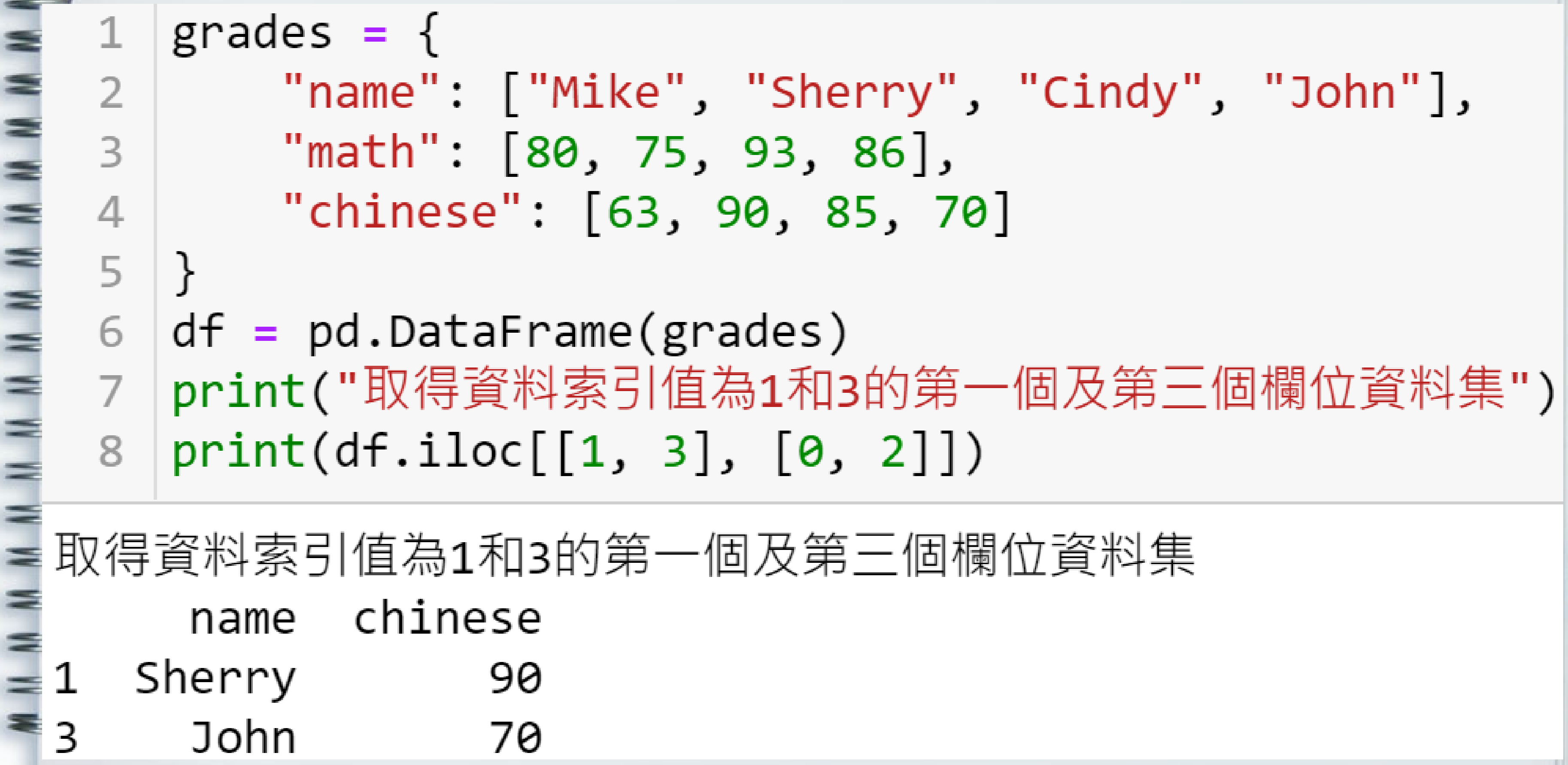


```python
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得資料索引值為1和3的name及chinese欄位資料集")
print(df.loc[[1, 3], ["name", "chinese"]])
```

```
取得資料索引值為1和3的name及chinese欄位資料集
     name  chinese
1  Sherry       90
3    John       70
```

取得DataFrame資料9

物件.iloc[[key1,key2],[key1, key2]]

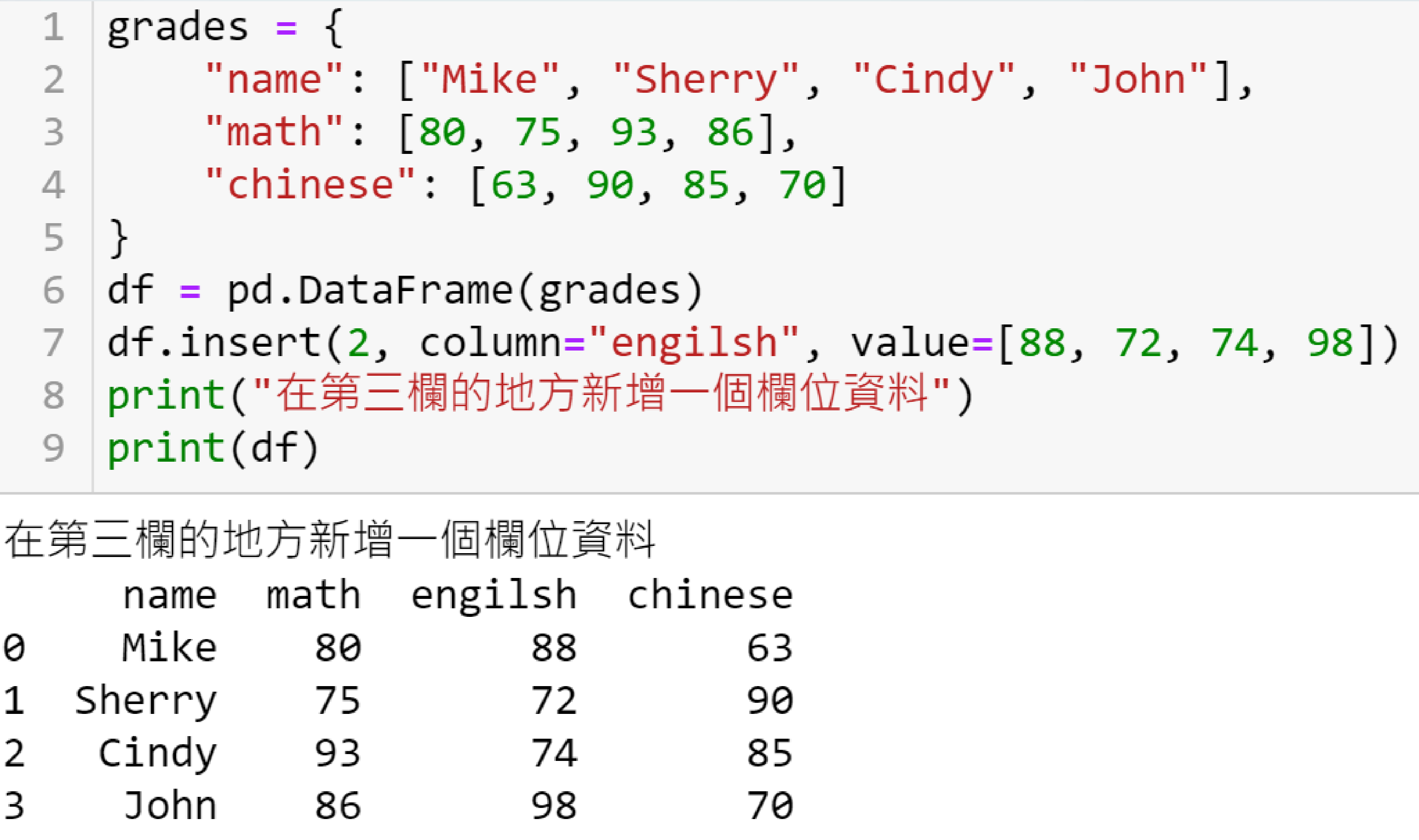


```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("取得資料索引值為1和3的第一個及第三個欄位資料集")
print(df.iloc[[1, 3], [0, 2]])
```

```
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])

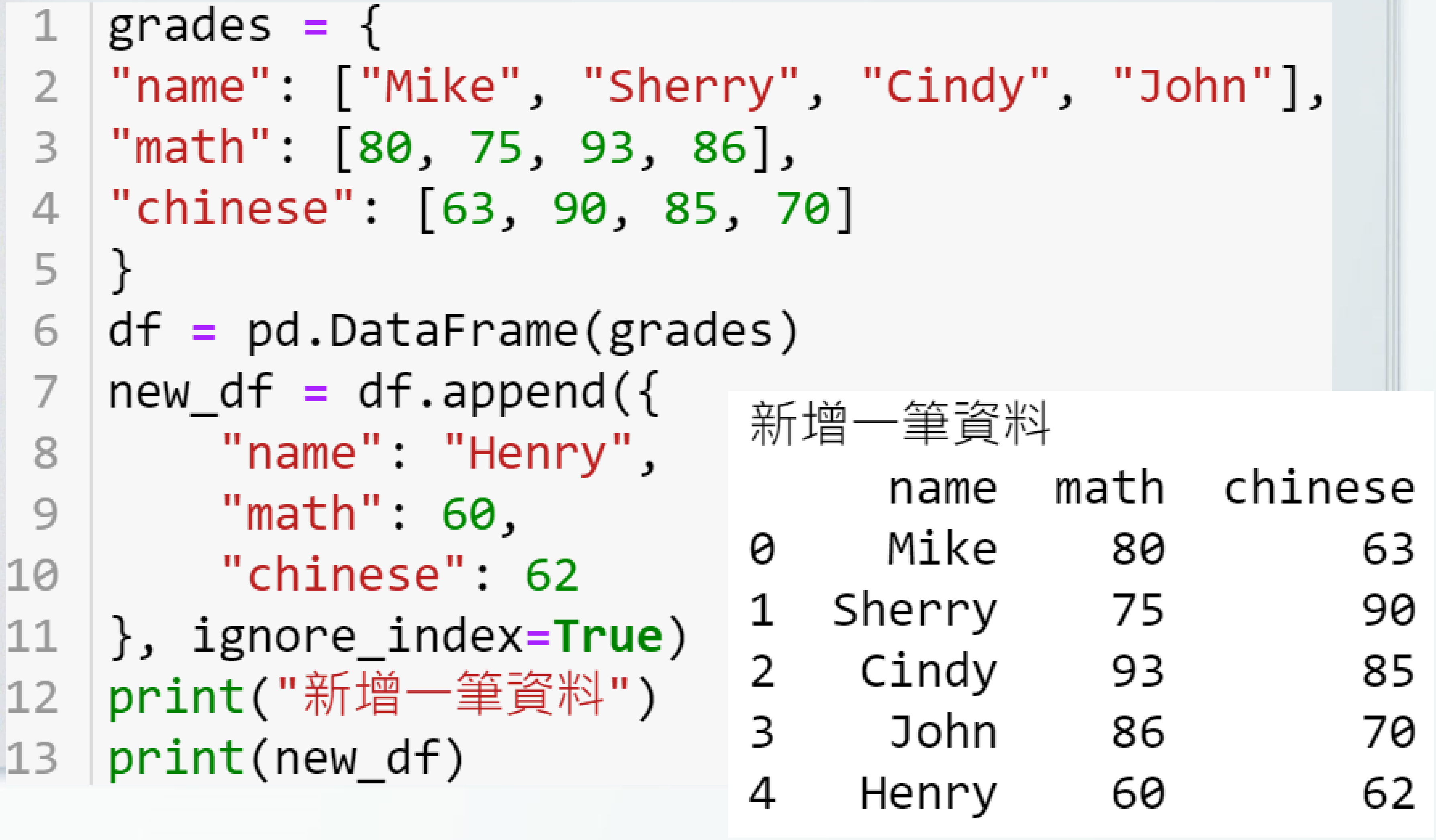


```
grades = {
    "name": ["Mike", "Sherry", "Cindy", "John"],
    "math": [80, 75, 93, 86],
    "chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
df.insert(2, column="engilsh", value=[88, 72, 74, 98])
print("在第三欄的地方新增一個欄位資料")
print(df)
```

```
在第三欄的地方新增一個欄位資料
     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

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

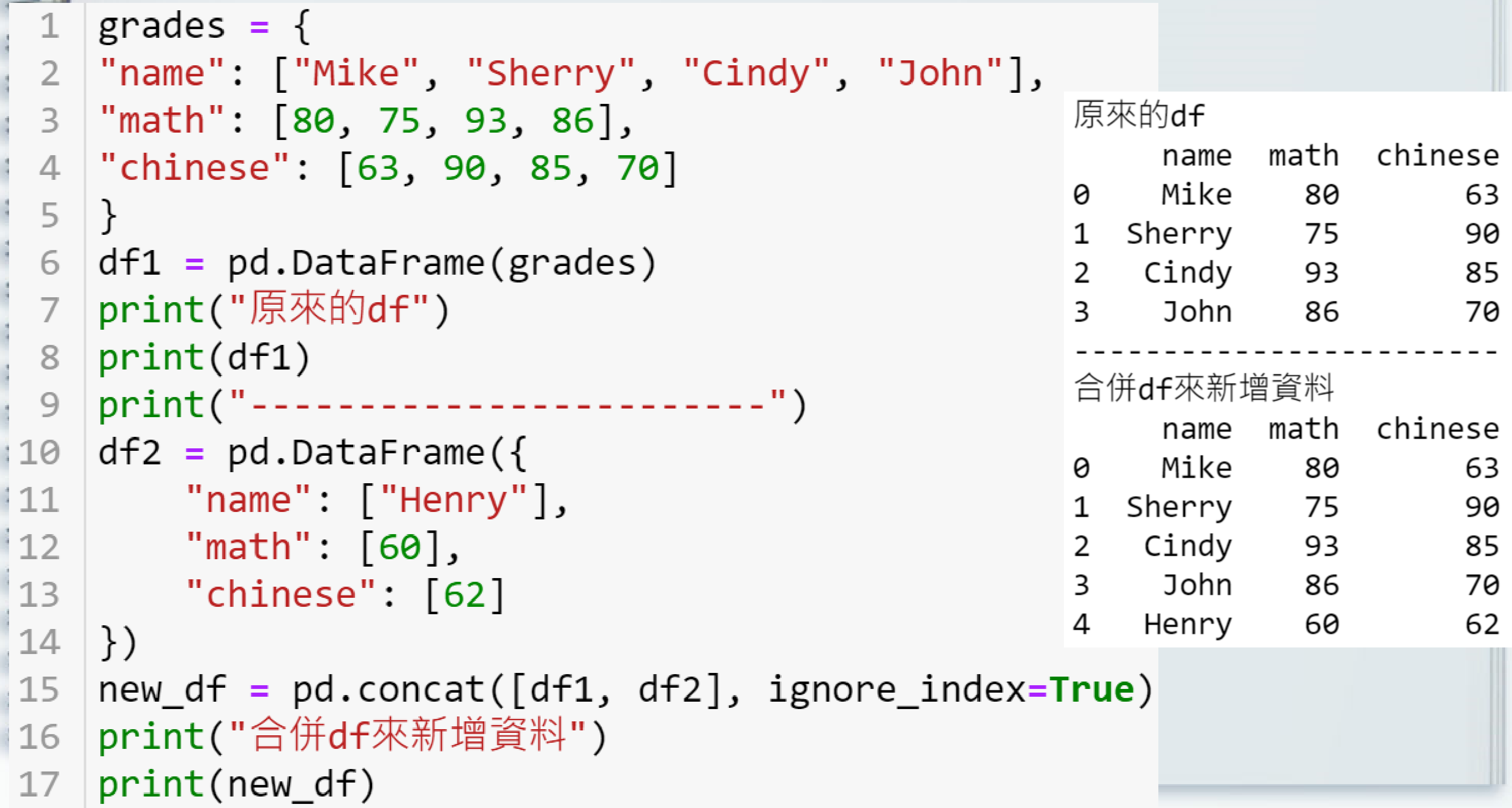


```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
new_df = df.append({
        "name": "Henry",
        "math": 60,
        "chinese": 62
}, ignore_index=True)
print("新增一筆資料")
print(new_df)
```

```
新增一筆資料
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
4   Henry    60       62
```

合併DataFrame資料

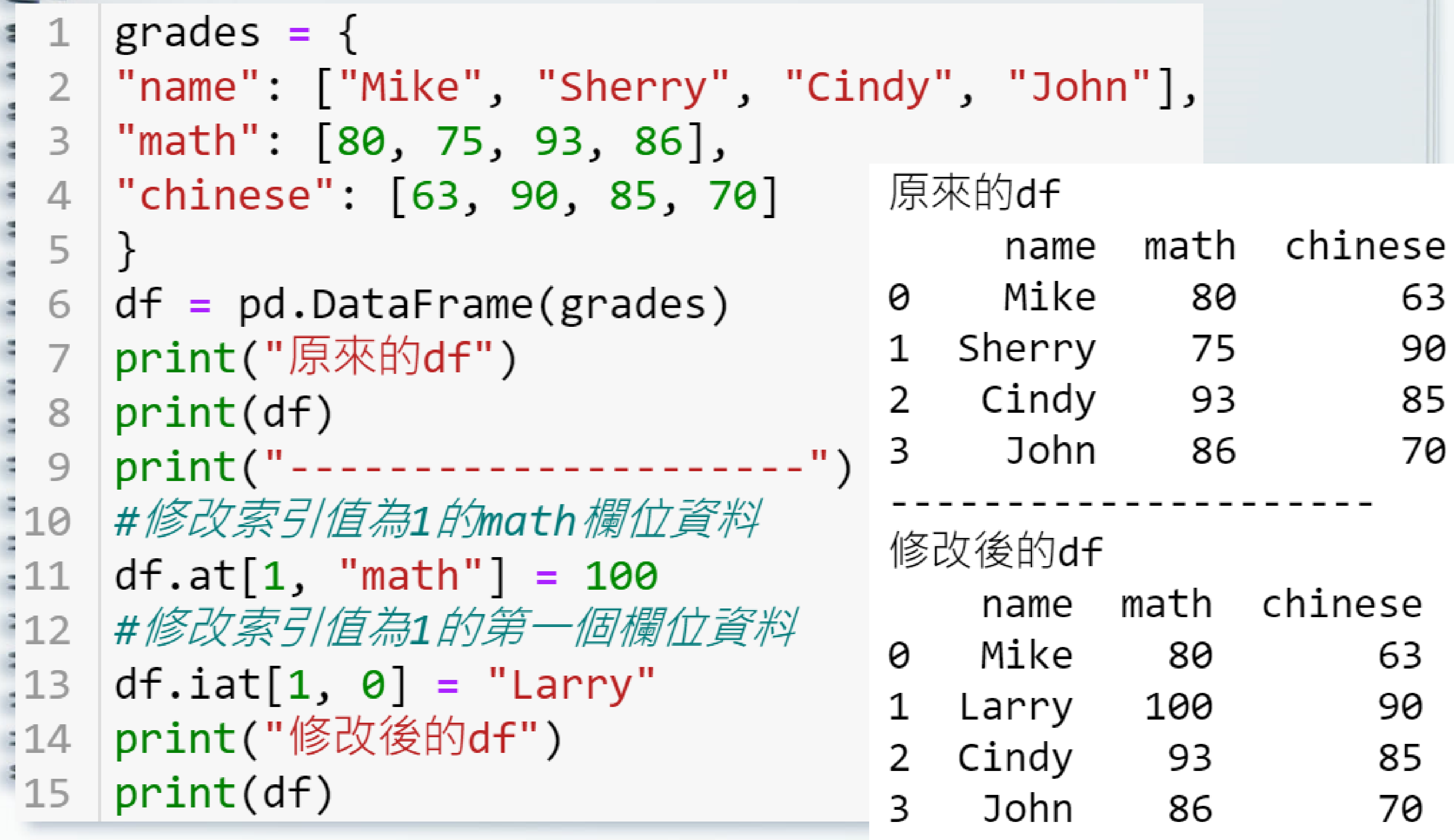
pd.concat([物件1,物件2], ignore_index=True)



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

```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
------------------------
合併df來新增資料
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
4   Henry    60       62
```

修改DataFrame資料

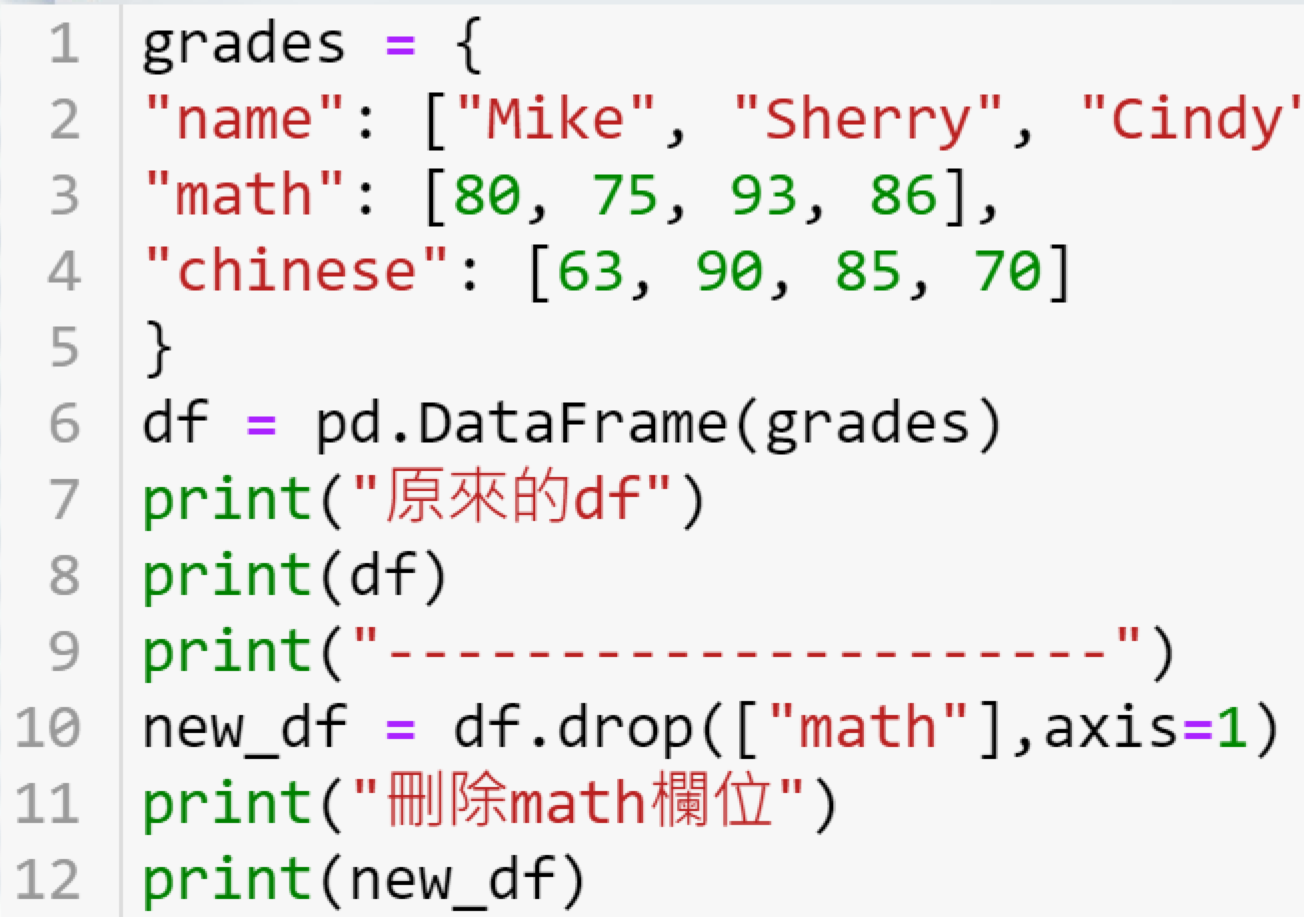


```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("----------------------")
#修改索引值為1的math欄位資料
df.at[1, "math"] = 100
#修改索引值為1的第一個欄位資料
df.iat[1, 0] = "Larry"
print("修改後的df")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

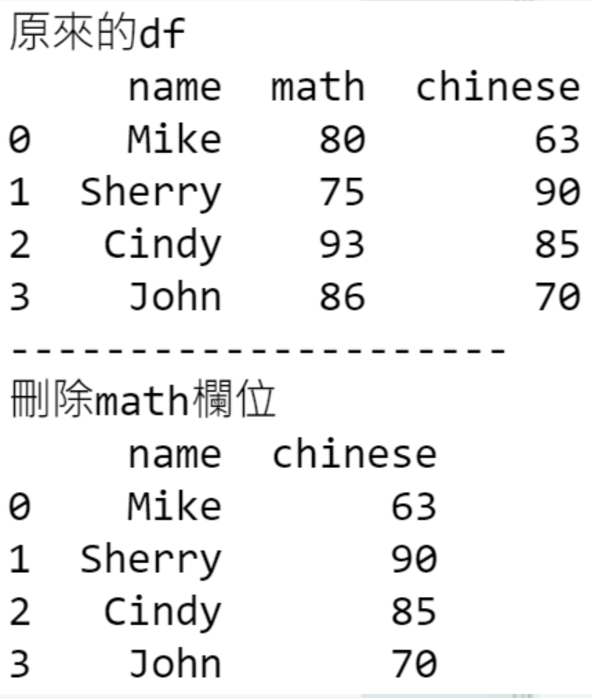
```
原來的df
      name  math  chinese
0     Mike    80       63
1   Sherry    75       90
2    Cindy    93       85
3     John    86       70
----------------------
修改後的df
     name  math  chinese
0    Mike    80       63
1   Larry   100       90
2   Cindy    93       85
3    John    86       70
```

刪除DataFrame資料1

物件.drop([“key”], axis=1)



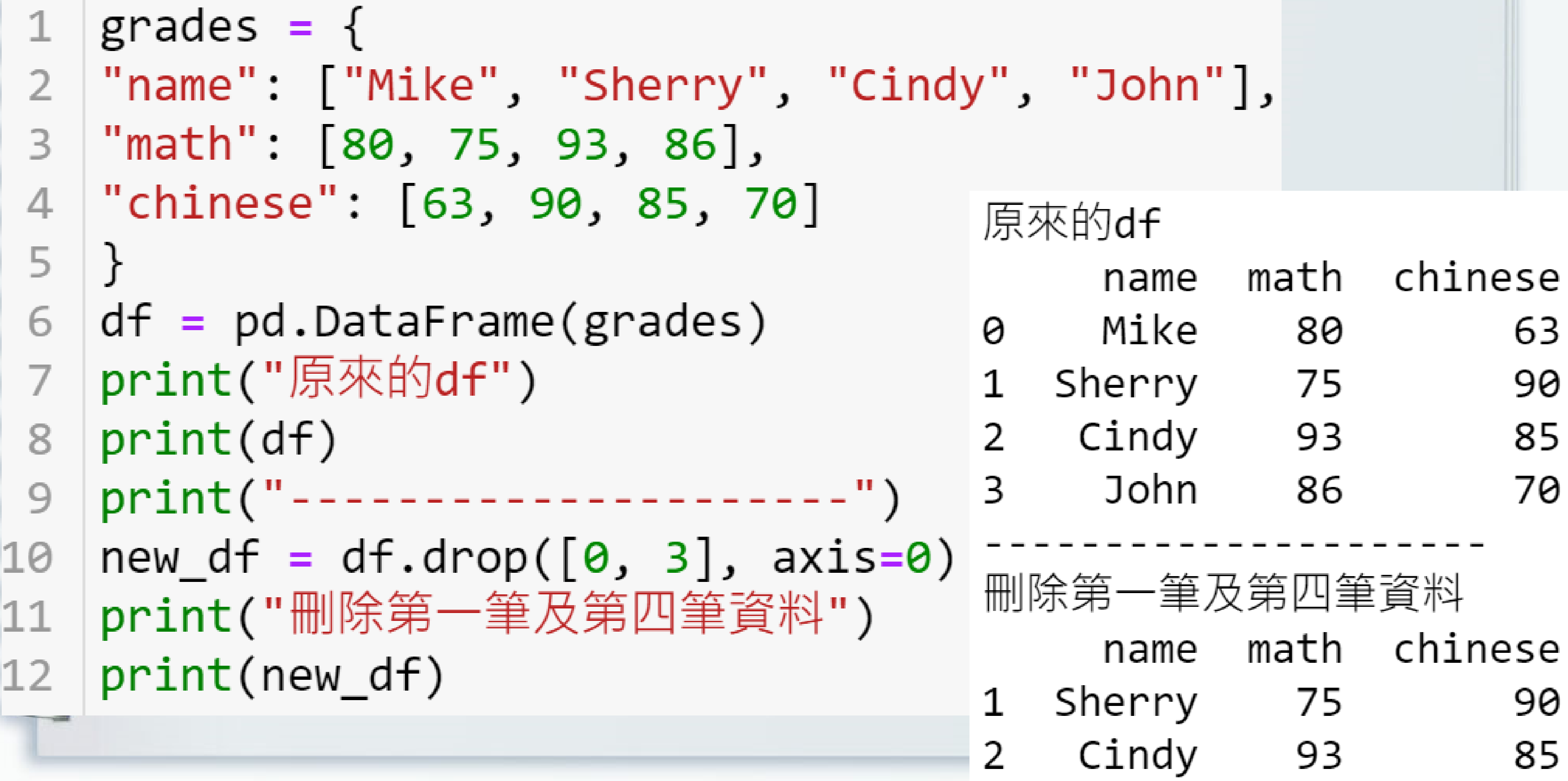
```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("----------------------")
new_df = df.drop(["math"],axis=1)
print("刪除math欄位")
print(new_df)
```



```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
----------------------
刪除math欄位
     name  chinese
0    Mike       63
1  Sherry       90
2   Cindy       85
3    John       70
```

刪除DataFrame資料2

物件.drop([key1,key2], axis=0)

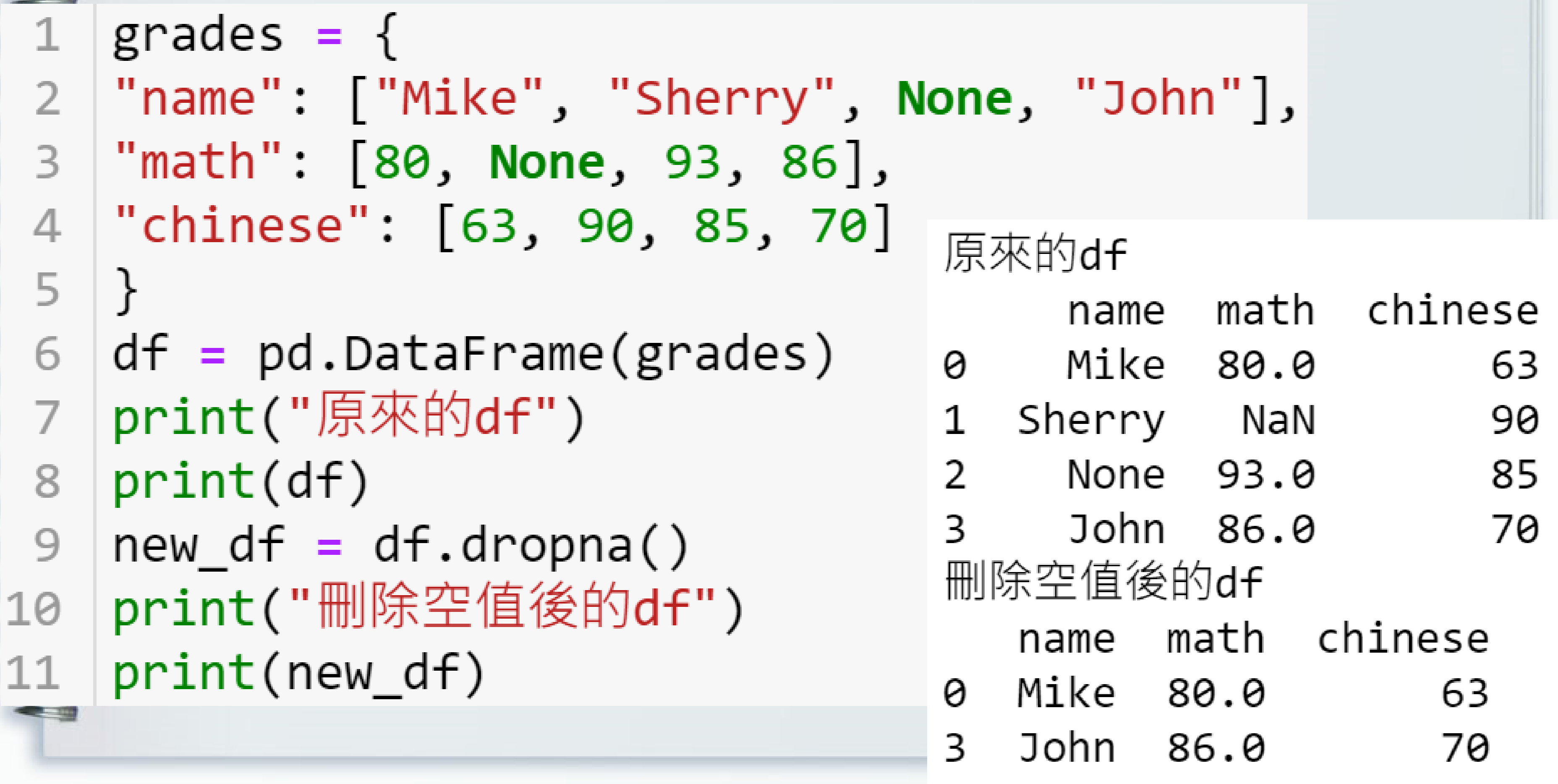


```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("-----------------------")
new_df = df.drop([0, 3], axis=0)
print("刪除第一筆及第四筆資料")
print(new_df)
```

```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
-----------------------
刪除第一筆及第四筆資料
     name  math  chinese
1  Sherry    75       90
2   Cindy    93       85
```

刪除DataFrame資料3

物件.dropna()

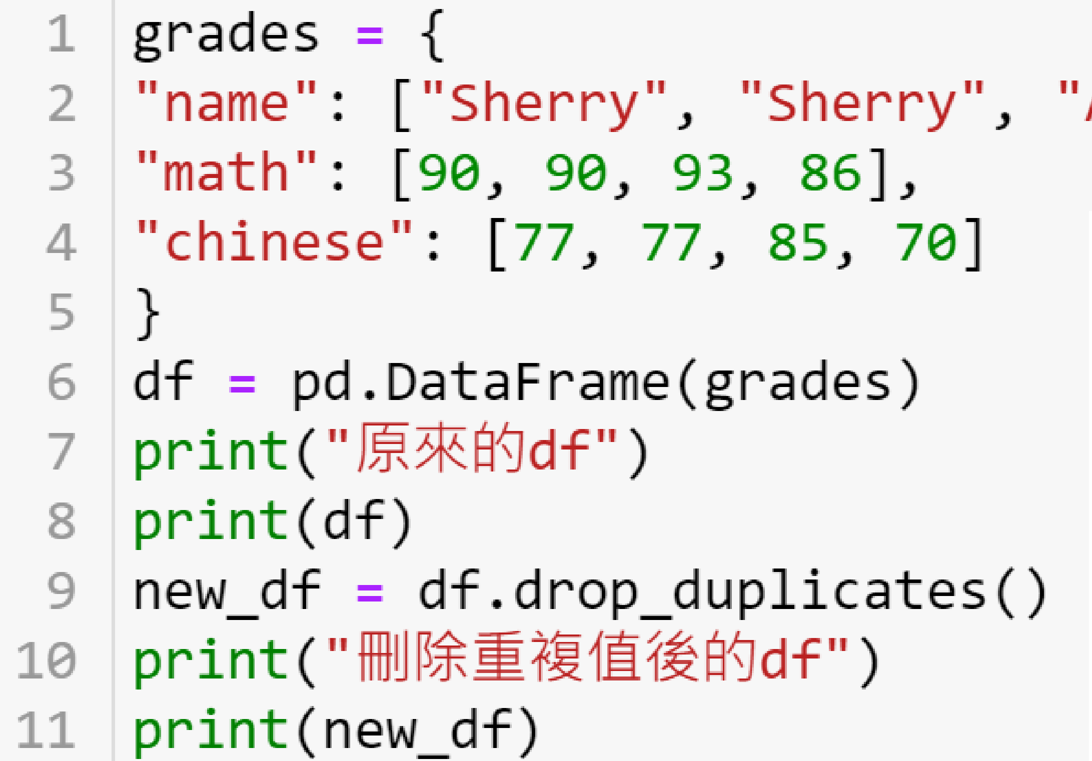


```python
grades = {
"name": ["Mike", "Sherry", None, "John"],
"math": [80, None, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
new_df = df.dropna()
print("刪除空值後的df")
print(new_df)
```

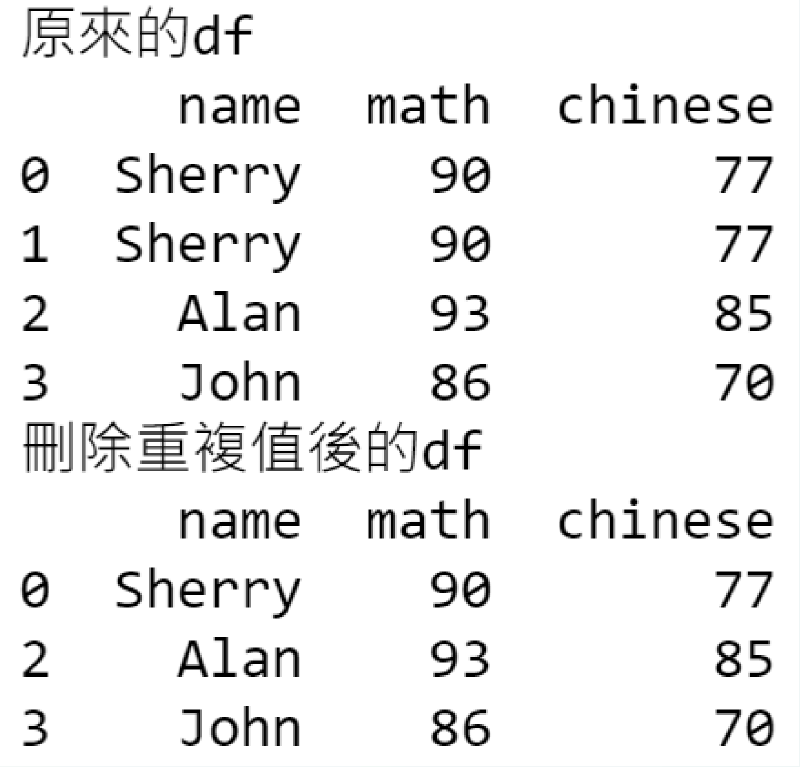
```
原來的df
      name  math  chinese
0     Mike  80.0       63
1   Sherry   NaN       90
2     None  93.0       85
3     John  86.0       70
刪除空值後的df
   name  math  chinese
0  Mike  80.0       63
3  John  86.0       70
```

刪除DataFrame資料4

物件.drop_duplicates()



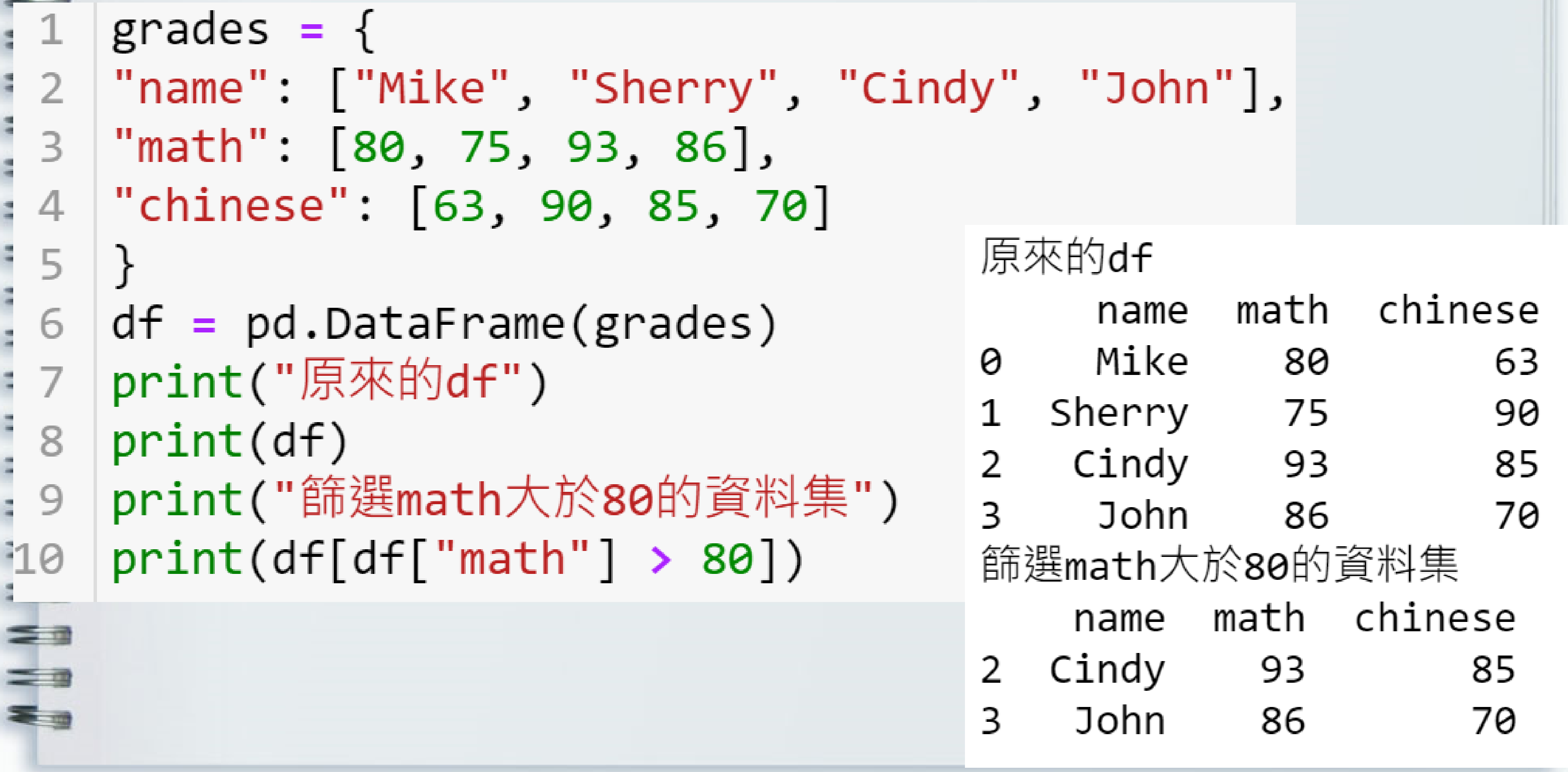
```python
grades = {
"name": ["Sherry", "Sherry", "Alan", "John"],
"math": [90, 90, 93, 86],
"chinese": [77, 77, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
new_df = df.drop_duplicates()
print("刪除重複值後的df")
print(new_df)
```



```
原來的df
     name  math  chinese
0  Sherry    90       77
1  Sherry    90       77
2    Alan    93       85
3    John    86       70
刪除重複值後的df
     name  math  chinese
0  Sherry    90       77
2    Alan    93       85
3    John    86       70
```

DataFrame篩選1

物件[物件[“key”]>值]

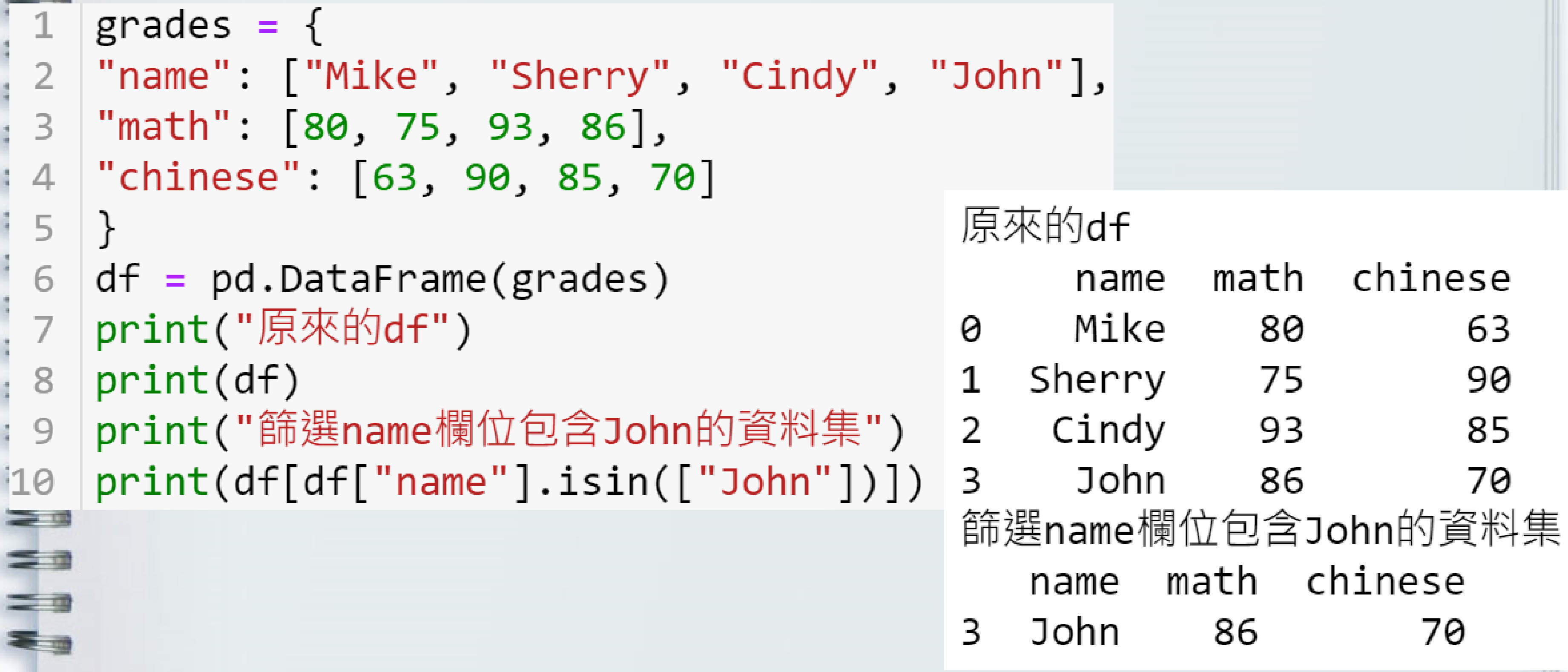


```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("篩選math大於80的資料集")
print(df[df["math"] > 80])
```

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

DataFrame篩選2

物件[物件[“key”].isin([“key”])]

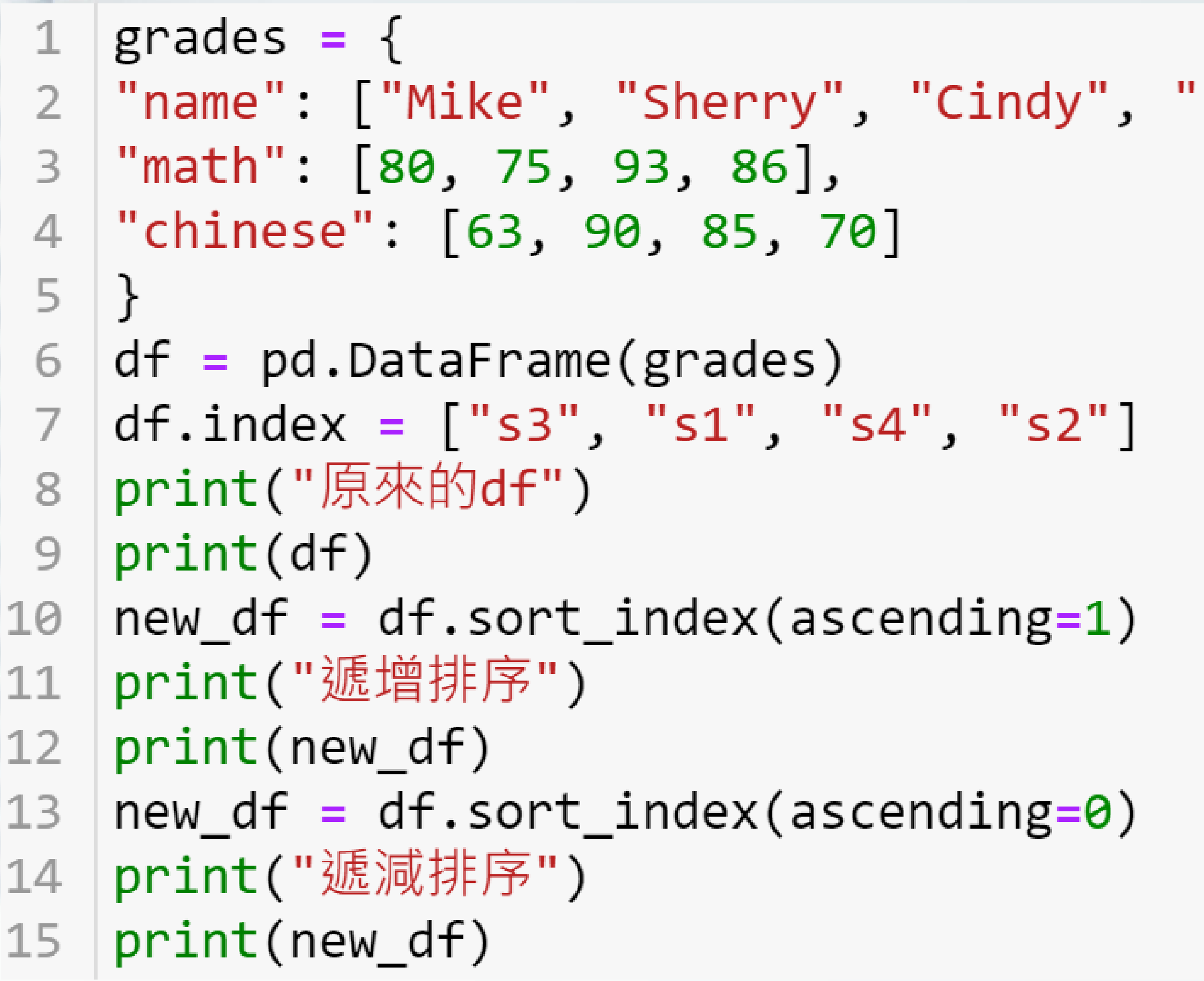


```
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
print("原來的df")
print(df)
print("篩選name欄位包含John的資料集")
print(df[df["name"].isin(["John"])])
```

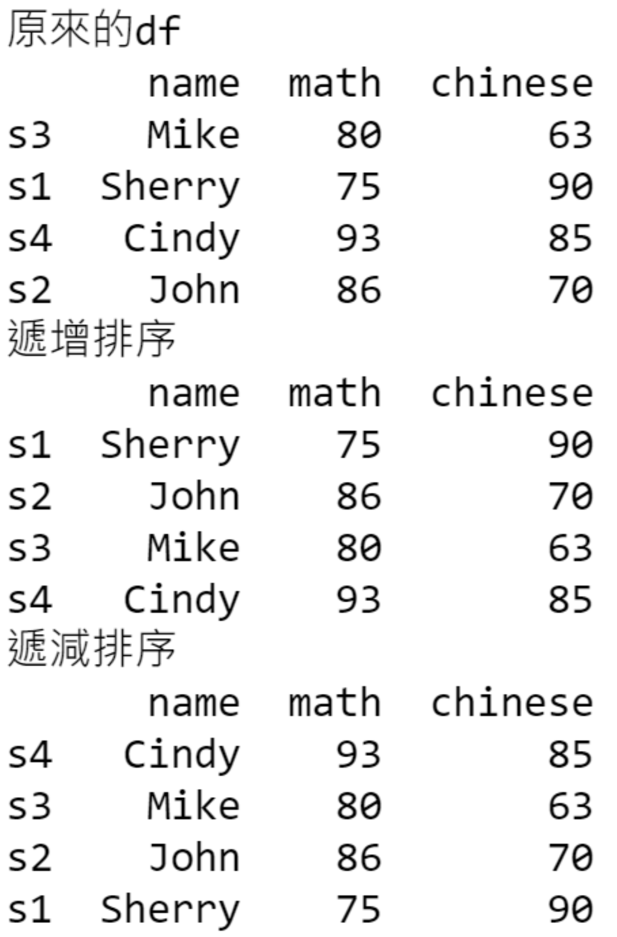
```
原來的df
     name  math  chinese
0    Mike    80       63
1  Sherry    75       90
2   Cindy    93       85
3    John    86       70
篩選name欄位包含John的資料集
   name  math  chinese
3  John    86       70
```

DataFrame排序1

物件.sort_index(ascending=0 or 1)



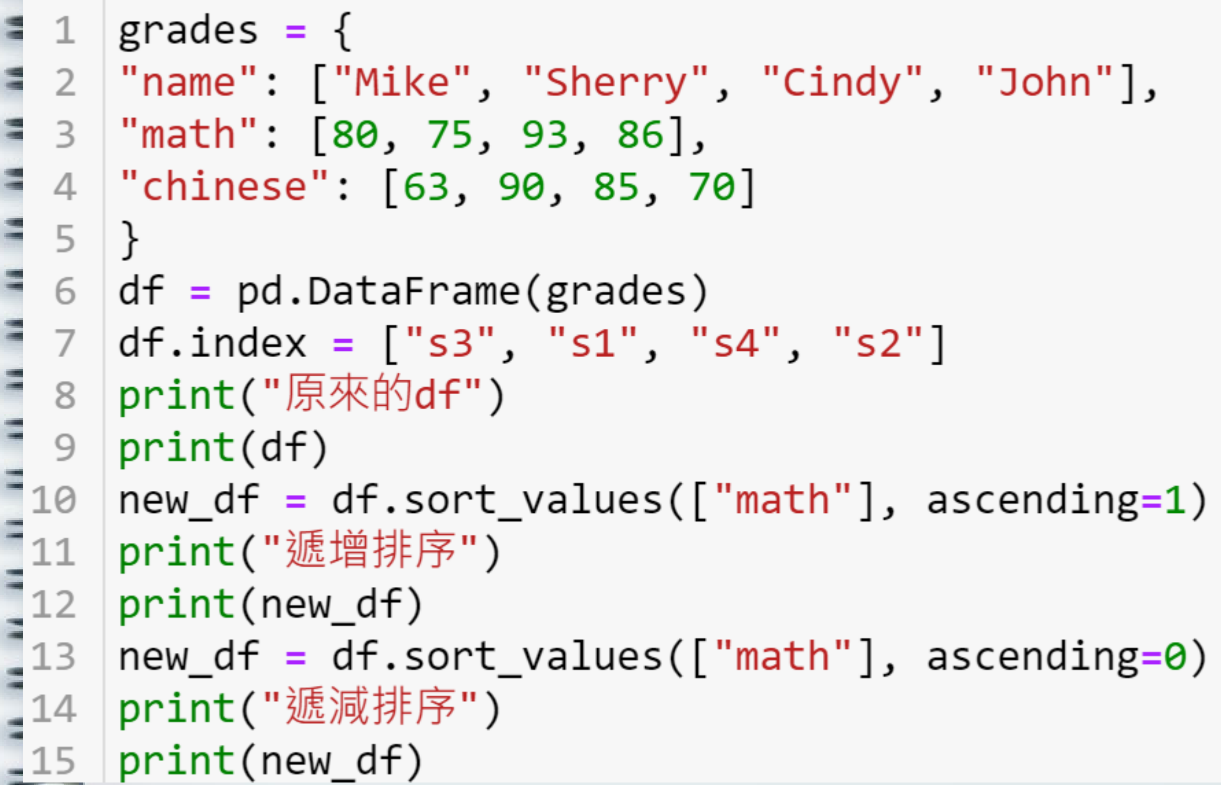
```python
grades = {
"name": ["Mike", "Sherry", "Cindy", "John"],
"math": [80, 75, 93, 86],
"chinese": [63, 90, 85, 70]
}
df = pd.DataFrame(grades)
df.index = ["s3", "s1", "s4", "s2"]
print("原來的df")
print(df)
new_df = df.sort_index(ascending=1)
print("遞增排序")
print(new_df)
new_df = df.sort_index(ascending=0)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



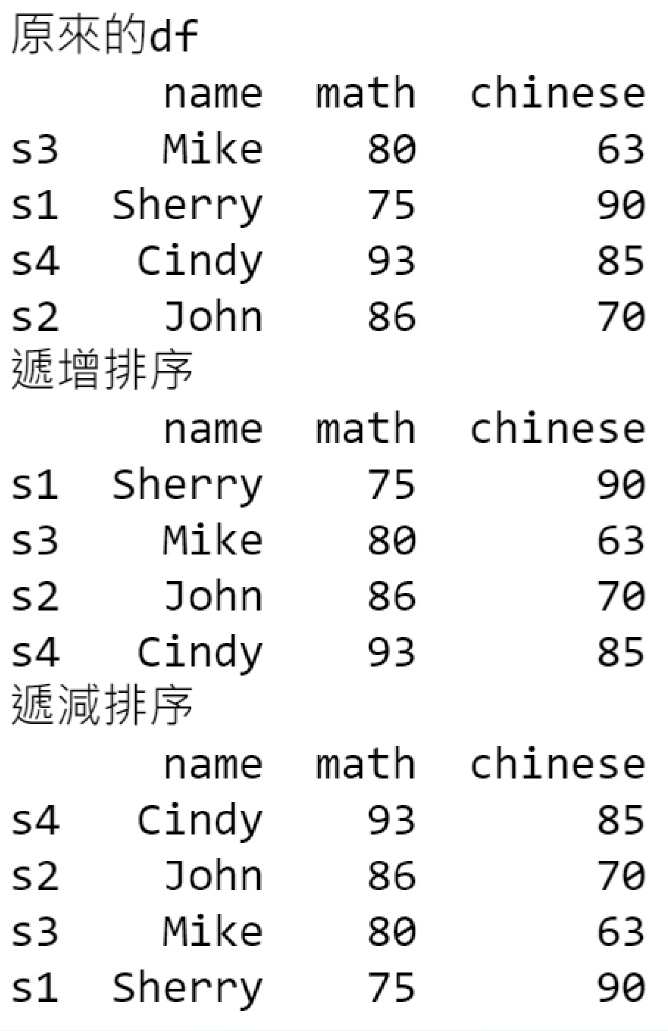
```
原來的df
      name  math  chinese
s3    Mike    80       63
s1  Sherry    75       90
s4   Cindy    93       85
s2    John    86       70
遞增排序
      name  math  chinese
s1  Sherry    75       90
s2    John    86       70
s3    Mike    80       63
s4   Cindy    93       85
遞減排序
      name  math  chinese
s4   Cindy    93       85
s3    Mike    80       63
s2    John    86       70
s1  Sherry    75       90
```

DataFrame排序2

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



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



```
原來的df
      name  math  chinese
s3    Mike    80       63
s1  Sherry    75       90
s4   Cindy    93       85
s2    John    86       70
遞增排序
      name  math  chinese
s1  Sherry    75       90
s3    Mike    80       63
s2    John    86       70
s4   Cindy    93       85
遞減排序
      name  math  chinese
s4   Cindy    93       85
s2    John    86       70
s3    Mike    80       63
s1  Sherry    75       90
```

STLECTUREs
小孫學堂

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

```
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：

```
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