

Basic Python - File

Hoàng-Nguyên Vũ

1. Mô tả:

- **Đọc file** là thao tác lấy dữ liệu từ file vào chương trình Python để xử lý. **Ghi file** là thao tác lưu dữ liệu từ chương trình Python vào file. Có nhiều cách để đọc và ghi file trong Python:

- **Cách 1: Sử dụng hàm open():** Hàm open() được sử dụng để mở file. Sau khi mở file, bạn có thể sử dụng các phương thức khác để đọc hoặc ghi dữ liệu vào file, cuối cùng ta cần phải đóng file lại sau khi thực thi xong các quá trình trên. Ví dụ:

```
1 # Read File
2 f = open("test.txt", "r")
3 data = f.read()
4 f.close()
5
6 print(data)
7
8 # Write File
9 f = open("test.txt", "w")
10 f.write("I Love AI Vietnam")
11 f.close()
```

- **Cách 2: Sử dụng câu lệnh with:** Câu lệnh with giúp bạn đảm bảo rằng file được đóng đúng cách sau khi sử dụng. Ví dụ:

```
1 # Read File
2 with open("test.txt", "r") as f:
3     data = f.read()
4 print(data)
5
6 # Write File
7 with open("test.txt", "w") as f:
8     f.write("I Love AI Vietnam")
```

2. Bài tập:

- **Câu 1:** Tạo List có tên `lst_data` gồm các số từ 1 đến 10, sau đó ghi toàn bộ list trên vào file có tên: `data.txt` với nội dung là 1 chuỗi số từ list trên nối với nhau bằng dấu -
- **Câu 2:** Đọc file `data.txt` vừa tạo ở câu 1 và lưu vào List mới có tên `lst_filter` gồm các số chia hết cho 3

Output:

- Câu 1: 1-2-3-4-5-6-7-8-9-10

- Câu 2: [3,6,9]

Basic Python - File

Hoàng-Nguyên Vũ

1. Mô tả:

Pandas là một thư viện mã nguồn mở được viết bằng Python, chuyên dụng cho việc phân tích dữ liệu và thao tác dữ liệu. Nó cung cấp một loạt các công cụ mạnh mẽ để giúp bạn:

- Đọc và ghi dữ liệu từ nhiều nguồn khác nhau, bao gồm tệp CSV, Excel, cơ sở dữ liệu SQL, v.v.
- Làm sạch và chuẩn bị dữ liệu cho phân tích.
- Thực hiện các phép toán thống kê trên dữ liệu.
- Tạo biểu đồ và hình ảnh để trực quan hóa dữ liệu.

2. Bài tập: Dữ liệu dataset McDonald

- **Phân tích dữ liệu thực đơn của McDonald** Hãy đọc file *menu.csv* có trong dataset phía trên và lưu vào biến *data* với thư viện Pandas. Sau đó sử dụng hàm *describe()*, *info()*, *head()*, *tail()* để thống kê tập dataset. **Lưu ý:** Lấy 5 phần tử đầu tiên và 10 phần tử cuối cùng khi dùng hàm *head* và *tail*

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)
count	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000
mean	368.269231	127.096154	14.165385	21.815385	6.007692	29.965385
std	240.269886	127.875914	14.205998	21.885199	5.321873	26.639209
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	210.000000	20.000000	2.375000	3.750000	1.000000	4.750000
50%	340.000000	100.000000	11.000000	17.000000	5.000000	24.000000
75%	500.000000	200.000000	22.250000	35.000000	10.000000	48.000000
max	1880.000000	1060.000000	118.000000	182.000000	20.000000	102.000000

Hình 1: Kết quả của hàm *describe()*

```
1 # Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 24 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Category                             260 non-null    object
1   Item                                 260 non-null    object
2   Serving Size                         260 non-null    object
3   Calories                             260 non-null    int64
4   Calories from Fat                   260 non-null    int64
5   Total Fat                           260 non-null    float64
6   Total Fat (% Daily Value)           260 non-null    int64
```

Hình 2: Kết quả của hàm *info()*