

THỰC HÀNH LAB 03:

LÀM SẠCH DỮ LIỆU CƠ BẢN

Vấn đề 1: Tiến hành tải dữ liệu vào chương trình ứng dụng Python và giải quyết vấn đề “Missing header in the csv file”

```
# Vấn đề 1: Xử lý thiếu header khi đọc file CSV
file_path = "patient_heart_rate.csv"
column_names = ["Id", "Name", "Age", "Weight", "m0006",
                "m0612", "m1218", "f0006", "f0612", "f1218"]
df = pd.read_csv(file_path, names=column_names)

# Hiển thị dữ liệu sau khi thêm header
print("Dữ liệu sau khi thêm header:")
print(df.head())
```

Dữ liệu sau khi thêm header:

	Id	Name	Age	Weight	m0006	m0612	m1218	f0006	f0612	f1218
0	1	Mickéy Mousé	56.0	70kgs	72	69	71	-	-	-
1	2	Donald Duck	34.0	154.89lbs	-	-	-	85	84	76
2	3	Mini Mouse	16.0	NaN	-	-	-	65	69	72
3	4	Scrooge McDuck	NaN	78kgs	78	79	72	-	-	-
4	5	Pink Panther	54.0	198.658lbs	-	-	-	69	NaN	75

Vấn đề 2: Xử lý vấn đề một cột lưu hỗn hợp nhiều dữ liệu, ở đây là cột “Name” chứa bao gồm “Firstname” và “Lastname”, giải pháp là ta sẽ tách ra làm 2 cột

```
# Vấn đề 2: Tách cột Name thành Firstname và Lastname
df[['Firstname', 'Lastname']] = df['Name'].str.split(expand=True, n=1)

# Hiển thị dữ liệu sau khi tách tên
print("Dữ liệu sau khi tách Firstname và Lastname:")
print(df.head())
```

Dữ liệu sau khi tách Firstname và Lastname:

	Id	Name	Age	Weight	m0006	m0612	m1218	f0006	f0612	f1218	Firstname	Lastname
0	1	Mickéy Mousé	56.0	70kgs	72	69	71	-	-	-	Mickéy	Mousé
1	2	Donald Duck	34.0	154.89lbs	-	-	-	85	84	76	Donald	Duck
2	3	Mini Mouse	16.0	NaN	-	-	-	65	69	72	Mini	Mouse
3	4	Scrooge McDuck	NaN	78kgs	78	79	72	-	-	-	Scrooge	McDuck
4	5	Pink Panther	54.0	198.658lbs	-	-	-	69	NaN	75	Pink	Panther