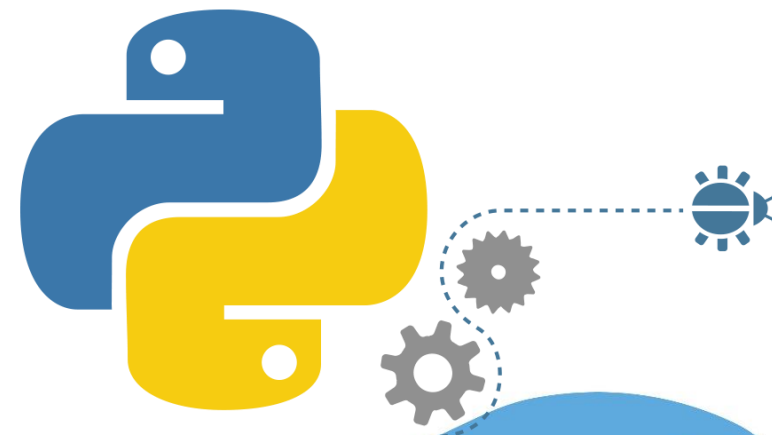


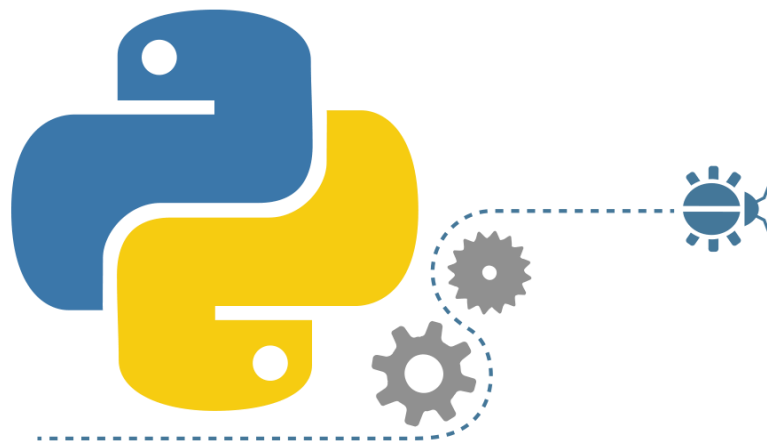


Môn Học **KỸ THUẬT LẬP TRÌNH** **VỚI PYTHON**

GV: Ths. Trần Duy Thanh
thanhtd@uel.edu.vn



XỬ LÝ TẬP TIN





Mục tiêu bài học

- Hiểu được lý do vì sao phải lưu và đọc tập tin
- Phân biệt được các loại tập tin thông dụng
- Ghi được tập tin
- Đọc được tập tin



Nội dung bài học

7.1. Vì sao phải lưu tập tin

7.2. Các loại tập tin thông dụng hiện nay

7.3. Xử lý Text File

7.4. Xử lý XML file

7.5. Xử lý JSON file

7.6. Xử lý CSV file

7.7. Xử lý Excel file

7.1. Vì sao phải lưu tập tin

- ❖ Các dữ liệu ta đang thao tác là trên thanh RAM → mất dữ liệu khi tắt phần mềm, tắt máy
- ❖ Nên cần phải lưu trữ vào thiết bị lưu trữ vĩnh cửu rồi nạp trở lại bộ nhớ.

7.1. Vì sao phải lưu tập tin





7.2. Các loại tập tin thông dụng hiện nay

- ❖ Có rất nhiều loại: Text file, XML, JSON, CSV, Excel...



7.3. Xử lý Text File

❖ Cách ghi tập tin

Dùng hàm:

open('myfile.txt', 'w') mở tập tin để ghi mới

Dùng hàm:

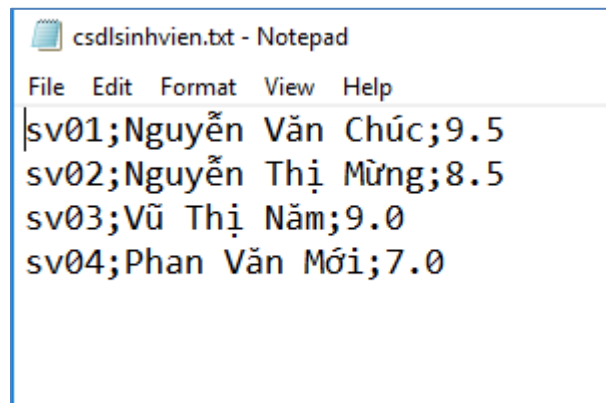
open('myfile.txt', 'a') mở tập tin để ghi nối đuôi



7.3. Xử lý Text File

Coding mẫu:

```
1 def luuFile():
2     file=open('csdlsinhvien.txt', 'w',encoding='utf-8')
3     file.writelines("sv01;Nguyễn Văn Chúc;9.5\n")
4     file.writelines("sv02;Nguyễn Thị Mừng;8.5\n")
5     file.writelines("sv03;Vũ Thị Năm;9.0\n")
6     file.writelines("sv04;Phan Văn Mới;7.0")
7     file.close()
8 luuFile()
```



The screenshot shows a Notepad window titled "csdlsinhvien.txt - Notepad". The menu bar includes "File", "Edit", "Format", "View", and "Help". The text area contains the following content:

```
sv01;Nguyễn Văn Chúc;9.5
sv02;Nguyễn Thị Mừng;8.5
sv03;Vũ Thị Năm;9.0
sv04;Phan Văn Mới;7.0
```



7.3. Xử lý Text File

❖ Cách đọc tập tin

```
def docFile():  
    file = open('csdlsinhvien.txt', 'r', encoding='utf-8')  
    for line in file:  
        print(line.strip())  
    file.close()  
  
docFile()
```



7.4. Xử lý XML file

XML có thể dùng DOM hoặc SAX để đọc. Ví dụ file sau có tên “employee.xml”

```
<?xml version="1.0" encoding="UTF-8" ?>
<employees>
  <employee>
    <id>1</id>
    <name>Trần Duy Thanh</name>
  </employee>
  <employee>
    <id>2</id>
    <name>Lê Hoàng Sử</name>
  </employee>
  <employee>
    <id>3</id>
    <name>Hồ Trung Thành</name>
  </employee>
</employees>
```



7.4. Xử lý XML file

```
from xml.dom.minidom import parse
import xml.dom.minidom
```

Mở file xml bằng minidom parser

```
DOMTree = xml.dom.minidom.parse("employees.xml")
```

```
collection = DOMTree.documentElement
```

Lấy tất cả tag là employee

```
employees = collection.getElementsByTagName("employee")
```

Duyệt vòng lặp để lấy toàn bộ dữ liệu ra

```
for employee in employees:
```

```
    tag_id = employee.getElementsByTagName('id')[0]
```

```
    id=tag_id.childNodes[0].data
```

```
    tag_name = employee.getElementsByTagName('name')[0]
```

```
    name=tag_name.childNodes[0].data
```

```
    print(id, '\t', name)
```



```
DOMXML x
C:\Python39\python.exe D:/Python/FileFactory/DOMXML.py
1   Trần Duy Thanh
2   Lê Hoàng Sử
3   Hồ Trung Thành

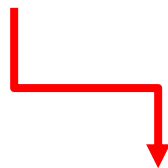
Process finished with exit code 0
```



7.5. Xử lý JSON file

```
import json
jsonString = '{ "ma": "nv1", "age": 50, "ten": "Trần Duy Thanh" }'
dataObject = json.loads(jsonString)
print(dataObject)
print("Mã=", dataObject["ma"])
print("Tên=", dataObject["age"])
print("Tuổi=", dataObject["ten"])
```

Chuyển đổi String
Json qua Python
Object



Python JsonFactory ×

↑
↓
🔍
-

{'ma': 'nv1', 'age': 50, 'ten': 'Trần Duy Thanh'}
Mã= nv1
Tên= 50
Tuổi= Trần Duy Thanh



7.5. Xử lý JSON file

```
import json
```

```
pythonObject = {  
    "ten": "Trần Duy Thanh",  
    "tuoi": 50,  
    "ma": "nv1"  
}
```

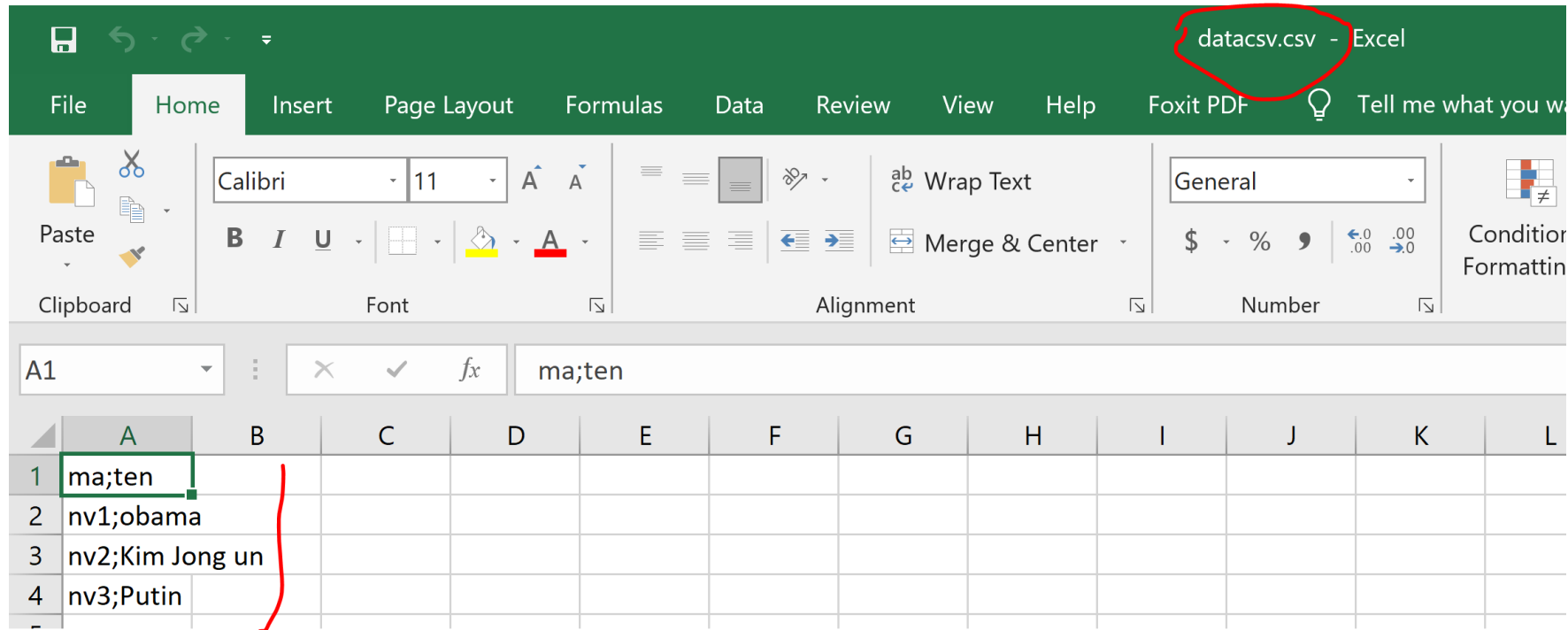
```
jsonString = json.dumps(pythonObject)
```

```
print(jsonString)
```

Chuyển đổi
Python Object
qua Json String

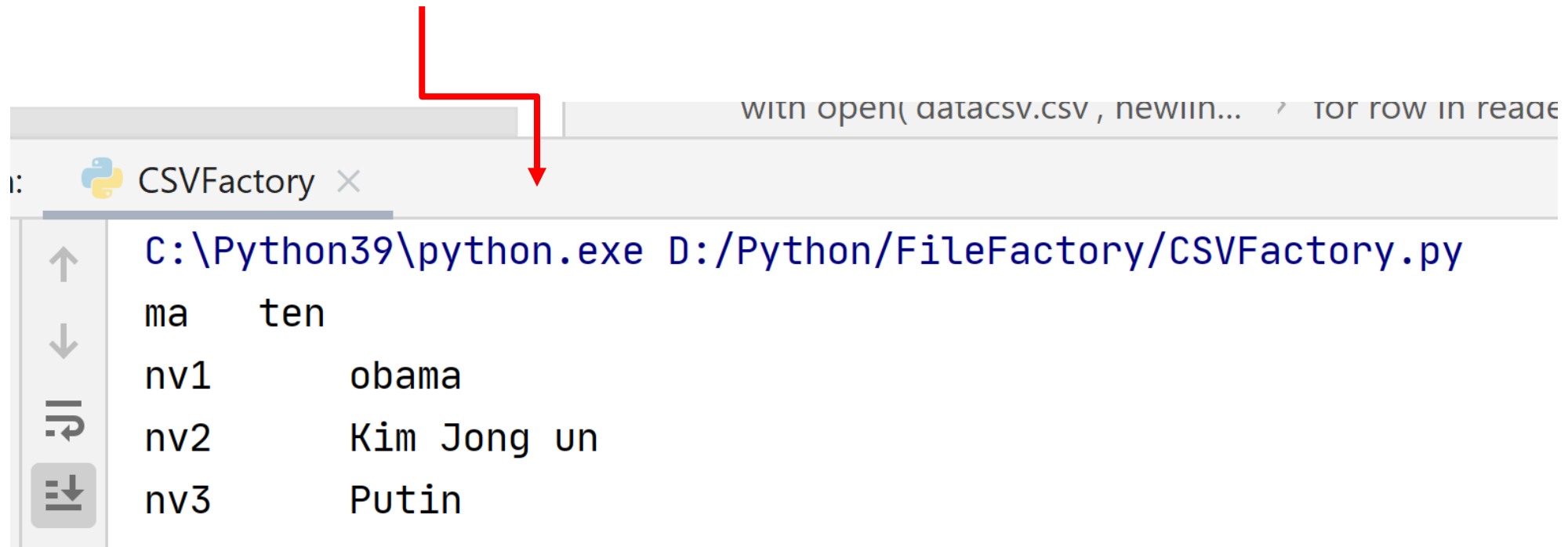
7.6. Xử lý CSV file

Giả sử ta có csv file có định dạng dưới đây, mỗi dòng ngăn cách ;



7.6. Xử lý CSV file

```
import csv
with open('datacsv.csv', newline='') as f:
    reader = csv.reader(f, delimiter=';', quoting=csv.QUOTE_NONE)
    for row in reader:
        print(row[0], "\t", row[1])
```

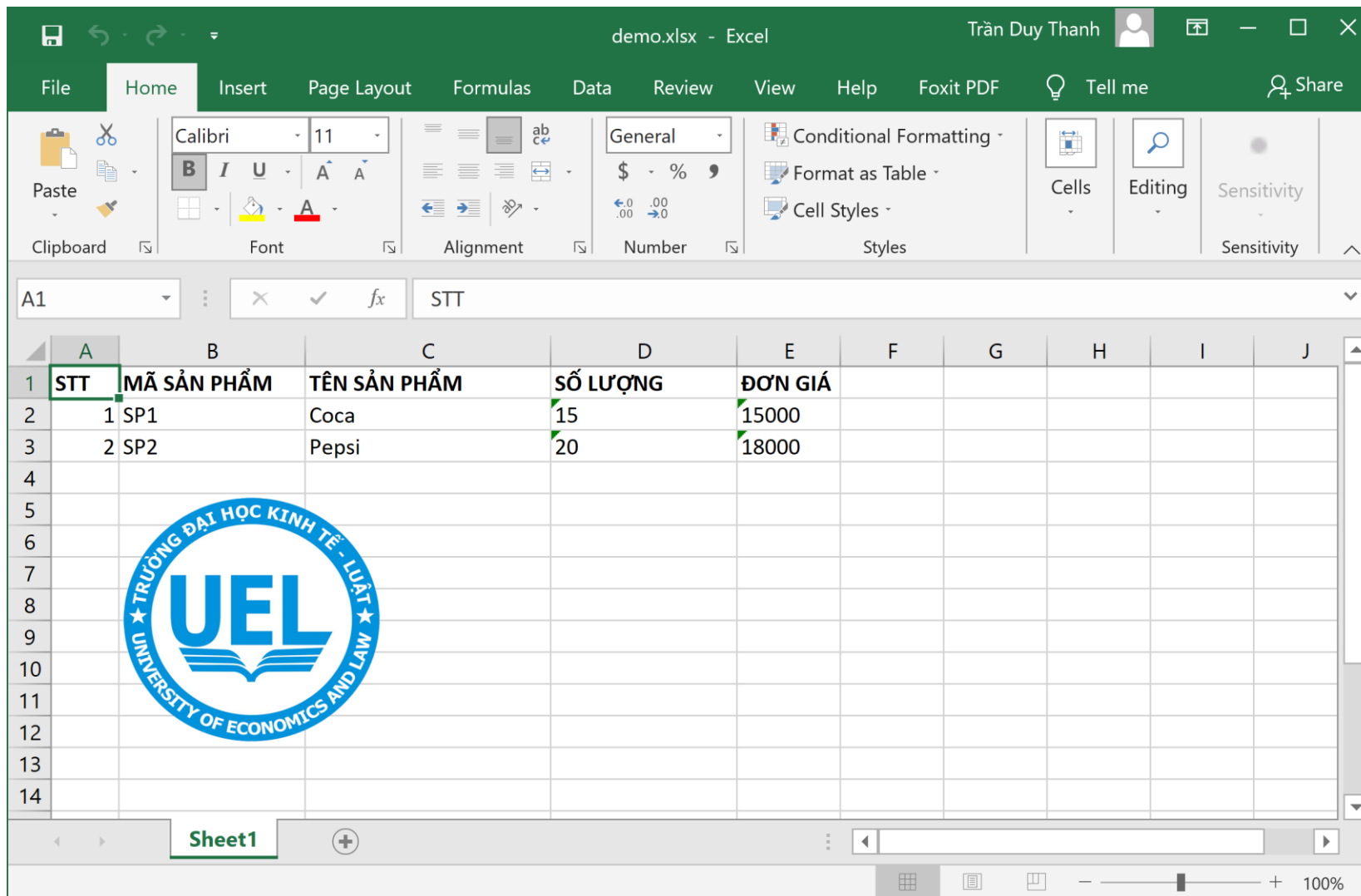


The screenshot shows a Python terminal window titled "CSVFactory". The command prompt is "C:\Python39\python.exe". The command being executed is "D:/Python/FileFactory/CSVFactory.py". The output of the script is a table with two columns: "ma" and "ten". The data rows are:

ma	ten
nv1	obama
nv2	Kim Jong un
nv3	Putin

7.7. Xử lý Excel file

Dùng Python để tạo ra file Excel có định dạng dưới đây



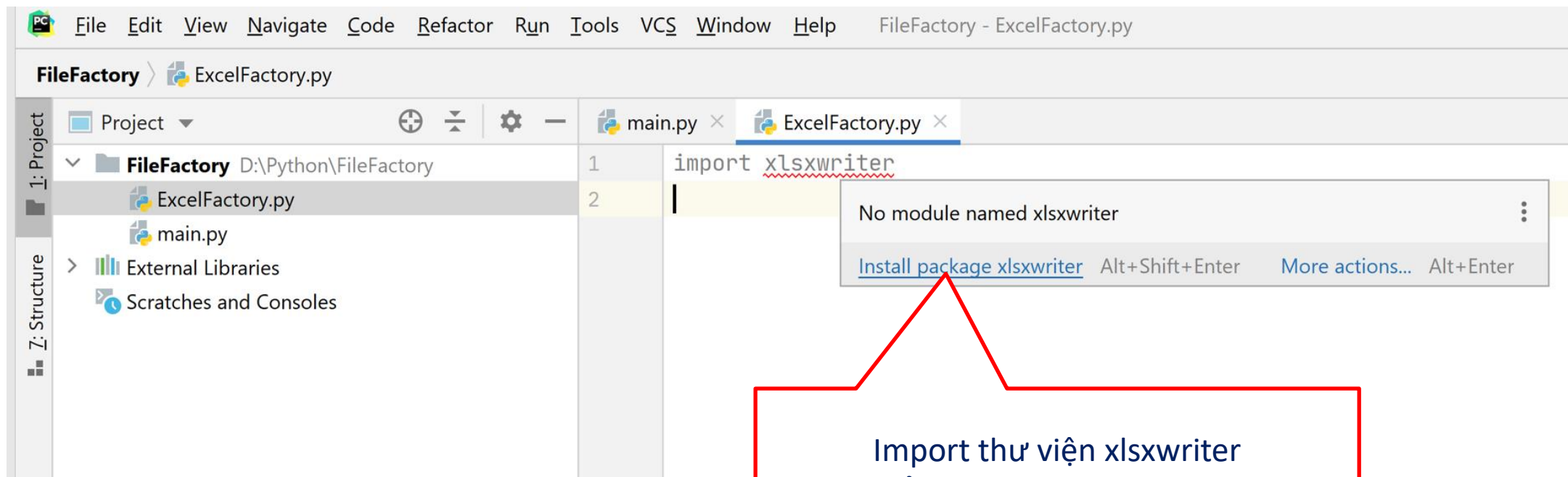
The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

STT	MÃ SẢN PHẨM	TÊN SẢN PHẨM	SỐ LƯỢNG	ĐƠN GIÁ
1	SP1	Coca	15	15000
2	SP2	Pepsi	20	18000

7.7. Xử lý Excel file

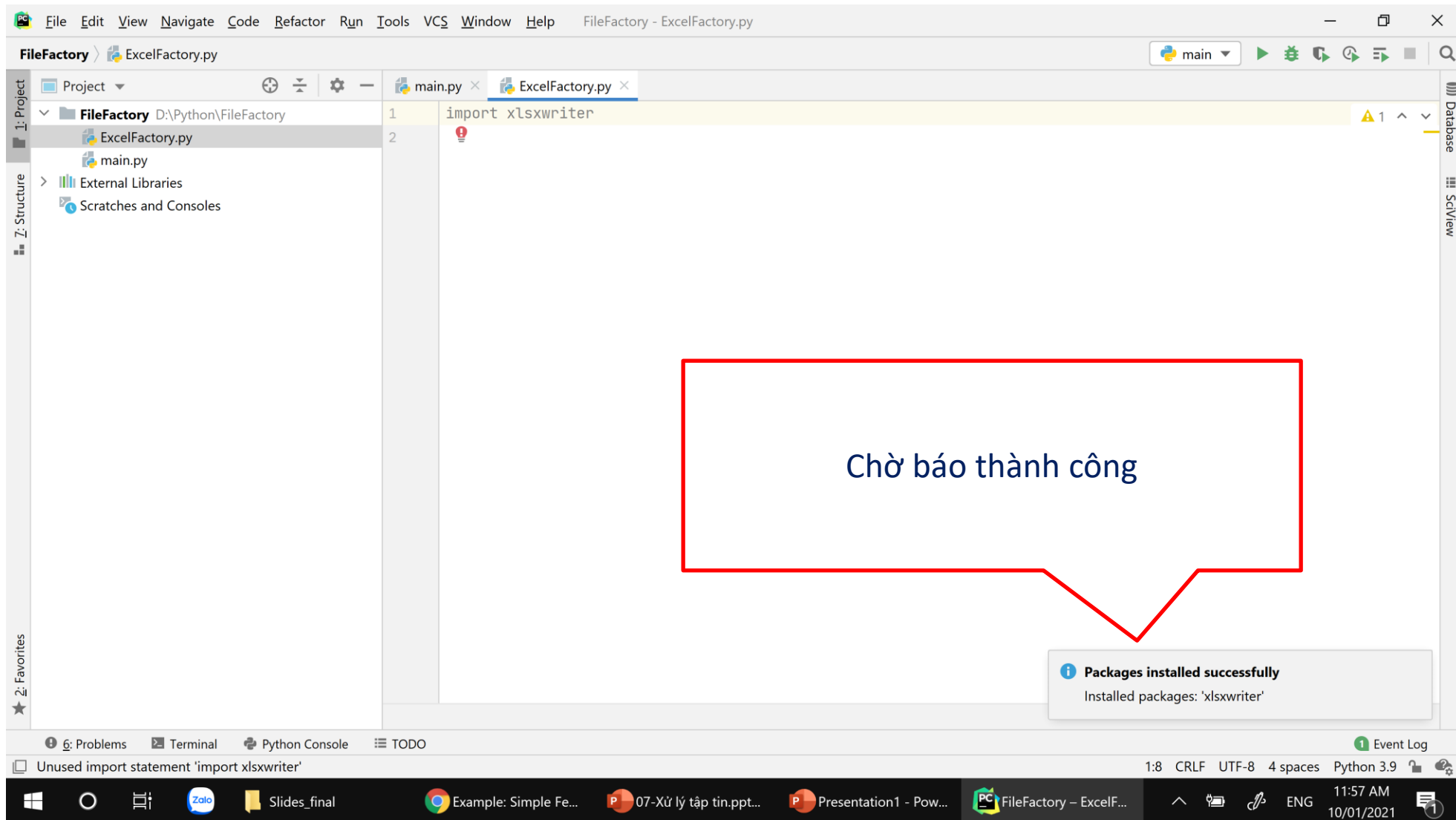
<https://xlsxwriter.readthedocs.io/> dùng thư viện này để tạo file Excel

```
import xlsxwriter
```



Import thư viện xlsxwriter
Di chuyển chuột tới dòng lệnh import
→ chọn Install package xlsxwriter

7.7. Xử lý Excel file



7.7. Xử lý Excel file

```
import xlsxwriter
```

```
# Tạo một file excel cùng 1 sheet
```

```
workbook = xlsxwriter.Workbook('demo.xlsx')  
worksheet = workbook.add_worksheet()
```

```
# thiết lập các cột cho file
```

```
worksheet.set_column('A:A', 5)  
worksheet.set_column('B:B', 15)  
worksheet.set_column('C:C', 20)  
worksheet.set_column('D:D', 15)  
worksheet.set_column('E:E', 15)
```

```
# định dạng tiêu đề cột in đậm
```

```
bold = workbook.add_format({'bold': True})
```

```
# thêm dòng tiêu đề và định dạng in đậm
```

```
worksheet.write('A1', 'STT', bold)  
worksheet.write('B1', 'MÃ SẢN PHẨM', bold)  
worksheet.write('C1', 'TÊN SẢN PHẨM', bold)  
worksheet.write('D1', 'SỐ LƯỢNG', bold)  
worksheet.write('E1', 'ĐƠN GIÁ', bold)
```

```
#thêm một dòng dữ liệu
```

```
worksheet.write('A2', 1)  
worksheet.write('B2', 'SP1')  
worksheet.write('C2', 'Coca')  
worksheet.write('D2', '15')  
worksheet.write('E2', '15000')
```

```
#thêm một dòng dữ liệu
```

```
worksheet.write('A3', 2)  
worksheet.write('B3', 'SP2')  
worksheet.write('C3', 'Pepsi')  
worksheet.write('D3', '20')  
worksheet.write('E3', '18000')
```

```
#Chèn Logo vào
```

```
worksheet.insert_image('B5', 'logo_UEL.png')
```

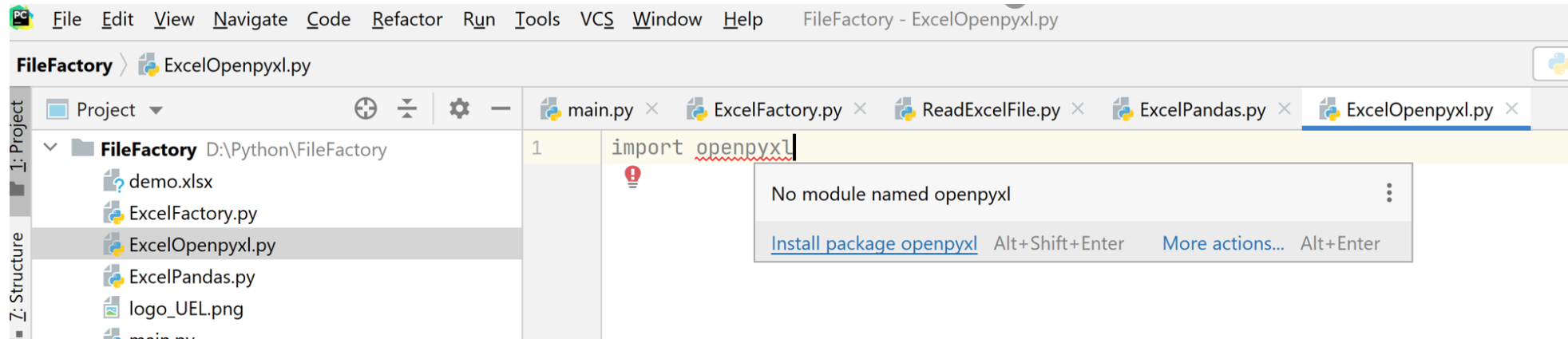
```
workbook.close()
```

Chạy phần mềm và vào thư mục phần mềm xem file Excel sẽ có kết quả như mong muốn

7.7. Xử lý Excel file

Đọc Ghi và đọc file Excel có rất nhiều thư viện, có thể kết hợp.

Ta thử dùng thư viện **openpyxl** để đọc Excel file (<https://openpyxl.readthedocs.io/en/stable/>), có thể dùng thư viện này để tạo file Excel luôn.



7.7. Xử lý Excel file

❖ Coding đọc dữ liệu trong file Excel

```
from openpyxl import load_workbook
wb = load_workbook('demo.xlsx')
print (wb.sheetnames)
ws = wb[wb.sheetnames[0]]
for row in ws.values:
    for value in row:
        print(value, "\t", end="")
    print("")
```

7.7. Xử lý Excel file

The screenshot shows an IDE window titled "FileFactory - ExcelOpenpyxl.py". The left sidebar displays a project structure for "FileFactory" located at "D:\Python\FileFactory", containing files like "demo.xlsx", "ExcelFactory.py", "ExcelOpenpyxl.py", "ExcelPandas.py", "logo_UEL.png", "main.py", and "ReadExcelFile.py". The main editor area shows the code in "ExcelOpenpyxl.py":

```
1 from openpyxl import load_workbook
2 wb = load_workbook('demo.xlsx')
3 print(wb.sheetnames)
4 ws = wb[wb.sheetnames[0]]
5 for row in ws.values:
6     for value in row:
7         print(value, "\t", end='')
8     print("")
```

A red arrow points from the `print("")` statement on line 8 to the console output. The console window at the bottom shows the execution of the script:

```
Run: ExcelOpenpyxl x
C:\Python39\python.exe -i /Python/FileFactory/ExcelOpenpyxl.py
['Sheet1']
```

STT	MÃ SẢN PHẨM	TÊN SẢN PHẨM	SỐ LƯỢNG	ĐƠN GIÁ
1	SP1	Coca	15	15000
2	SP2	Pepsi	20	18000



THANK YOU

028 37244555 www.uel.edu.vn

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KINH TẾ - LUẬT

Số 669, đường Quốc lộ 1, khu phố 3, phường Linh Xuân,
quận Thủ Đức, Thành phố Hồ Chí Minh.