**TRƯỜNG CAO ĐẲNG FPT POLYTECHNIC**

**🕮**

**Text

Description automatically generated**

**LAB 2**

MÔN HỌC : Lập Trình với Pandas

MÃ MÔN : DAT 205

Báo Cáo Bài Tập môn Học

**Sinh viên thực hiện :**  *Trần Gia Lạc*

**MSSV :** *Ps31064*

**Giảng viên hướng dẫn :** Văn Công Khanh

**Mục Lục**

[MỤC TIÊU: 3](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191195)

[PHẦN I 3](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191196)

[Bài 1 (2 điểm) 3](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191197)

[Bài 2 (2 điểm) 5](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191198)

[PHẦN II 6](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191199)

[Bài 3 (2 điểm) 6](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191200)

[Bài 4 (2 điểm) 7](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191201)

[Bài 5: 8](file:///H:\Kho%20dữ%20liệu\PJ\PS31064_Trần%20Gia%20Lạc_%20DP18301_lab7.docx#_Toc157191202)

# MỤC TIÊU:

# PHẦN I

## Bài 1 (2 điểm)

- Tải dữ liệu từ tệp "lab2.csv" vào một DataFrame và gán cho biến df.

- In ra 5 hàng đầu tiên của DataFrame df để khám phá dữ liệu.

- Sử dụng phương thức info() để hiển thị thông tin về cấu trúc của DataFrame df, bao gồm số lượng hàng và cột, và kiểu dữ liệu của các cột.

**Bài làm:**

A screenshot of a computer program

Description automatically generated

**pd.read\_csv()**: đọc file csv

**pd.DataFrame()**: tạo DataFrame lấy từ file csv

**head()**: lấy dòng tính từ đầu của dữ liệu

**info()**: hiện thông tin cấu trúc của dữ liệu ví dụ như số dòng, số cột, có bao nhiêu kiểu dữ liệu, sử dụng bộ nhớ,…

## Bài 2 (2 điểm)

- Tính toán giá trị trung bình và độ lệch chuẩn của cột "age".

- Đếm số lượng giá trị thiếu trong mỗi cột của DataFrame df.

- Liệt kê tất cả các giá trị duy nhất trong cột "gender".

**Bài làm:**

A screenshot of a computer

Description automatically generated

**Mean()**: tính giá trị trung bình

**Std()**: tính độ lệch chuẩn

**Unique()**: trả về các giá trị duy nhất trong dữ liệu

# PHẦN II

## Bài 3 (2 điểm)

- Sử dụng thuộc tính shape để hiển thị kích thước của DataFrame df.

- Sử dụng phương thức describe() để tạo báo cáo mô tả tổng quan về các cột chứa dữ liệu số trong DataFrame df.

- Sử dụng phương thức loc[] để truy cập dữ liệu của hàng thứ 10 trong DataFrame df.

- Sử dụng phương thức iloc[] để truy cập dữ liệu của hàng thứ 5 và cột "name" trong DataFrame df.

**Bài làm:**

A screenshot of a computer

Description automatically generated

**shape()**: trả về kích thước của dữ liệu (hàng, cột)

**describe()**: trả về bảng mô tả tổng quan về dữ liệu với các hàm cơ bản và có thể hiện thêm 1 số hàm cần thiết nếu muốn

**loc[9]**: trả về hàng 10 đã cho trước

**iloc[‘4’][‘name’]**: trả về dữ liệu hàng thứ 3 của cột name

## Bài 4 (2 điểm)

**Bài làm:**