**TRƯỜNG CAO ĐẲNG FPT POLYTECHNIC**

**🕮**

**Text

Description automatically generated**

**ASSIGNMENT**

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**

[Mô tả: 3](#_Toc167704686)

[Giai đoạn I: 3](#_Toc167704687)

[- Nạp dữ liệu từ csv, kiểm tra thông tin cơ bản. 3](#_Toc167704688)

[- Hiển thị một số dữ liệu ở đầu và cuối 3](#_Toc167704689)

[- Xác định cột không cần thiết, thực hiện loại bỏ để làm sạch. 4](#_Toc167704690)

[- Chuyển đổi kiểu dữ liệu ở các cột cho phù hợp (ví dụ cột chứa dữ liệu ngày tháng) 4](#_Toc167704691)

[Y1\_02: 5](#_Toc167704692)

[- Tính tổng số ca nhiễm, tử vong, và khỏi bệnh trên toàn thế giới. 5](#_Toc167704693)

[- Tính tỷ lệ tử vong trên số ca nhiễm 5](#_Toc167704694)

[- Lọc dữ liệu cho một quốc gia cụ thể 5](#_Toc167704695)

[- Phân loại các quốc gia thành các nhóm dựa trên số lượng ca nhiễm 5](#_Toc167704696)

# Mô tả:

Dữ liệu được sử dụng trong dự án này được thu thập từ các nguồn tin cậy. Nguồn dữ liệu này cung cấp thông tin về số ca nhiễm, tử vong và hồi phục từ COVID-19 trên toàn thế giới, cũng như chi tiết về các trường hợp tại từng quốc gia và khu vực cụ thể. Mục tiêu của dự án là phân tích và hiểu rõ hơn về xu hướng, biến động và dự báo về dịch bệnh COVID-19. Chúng tôi cũng muốn tạo ra các báo cáo và biểu đồ trực quan để giúp cộng đồng và quản lý dịch bệnh có cái nhìn tổng quan và chi tiết hơn về tình hình dịch bệnh. Trong giai đoạn 1, mục tiêu chính là làm quen với dữ liệu và thư viện Pandas, khám phá và làm sạch dữ liệu, và thực hiện các phân tích cơ bản. Các thao tác như đọc dữ liệu từ file, xử lý dữ liệu thời gian, thực hiện merge/join giữa các DataFrame, tính toán số liệu thống kê cơ bản như tổng số ca nhiễm, tử vong, và hồi phục sẽ được thực hiện Trong giai đoạn 2, tập trung vào các phân tích sâu hơn và phức tạp hơn. Một số mục tiêu chính bao gồm phân tích Time Series để hiểu xu hướng và biến động của dịch bệnh theo thời gian, dự báo số ca nhiễm trong tương lai sử dụng các mô hình dự báo Time Series, và phân tích sự ảnh hưởng của các biện pháp kiểm soát và vắc xin đối với sự lan rộng của virus.

# Giai đoạn I:

Y1\_01

* Nạp dữ liệu từ csv, kiểm tra thông tin cơ bản.

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

* Hiển thị một số dữ liệu ở đầu và cuối

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

* Xác định cột không cần thiết, thực hiện loại bỏ để làm sạch.

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

* Chuyển đổi kiểu dữ liệu ở các cột cho phù hợp (ví dụ cột chứa dữ liệu ngày tháng)

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

## Y1\_02:

* Tính tổng số ca nhiễm, tử vong, và khỏi bệnh trên toàn thế giới.

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

* Tính tỷ lệ tử vong trên số ca nhiễm

**Bài làm:**

**A black screen with text

Description automatically generated**

* Lọc dữ liệu cho một quốc gia cụ thể

**Bài làm:**

**A screenshot of a computer

Description automatically generated**

* Phân loại các quốc gia thành các nhóm dựa trên số lượng ca nhiễm

**Bài làm:**

A screenshot of a computer

Description automatically generated