**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

Ảnh có chứa đồng hồ đo, ký hiệu

Mô tả được tạo tự động

**NGUYỄN NGỌC UYÊN TRANG**

**MSSV: 19127074**

**BÁO CÁO LAB 2**

**THỐNG KÊ VÀ TRỰC QUAN HÓA DỮ LIỆU**

**|Giáo viên hướng dẫn|**

**Lê Ngọc Thành**

**Thành phố Hồ Chí Minh – 2022**

CHƯƠNG 1. Mức độ hoàn thành tổng thể của mỗi yêu cầu

|  |  |
| --- | --- |
| Yêu cầu | Mức độ hoàn thành |
| Thu thập số liệu thống kê từng ngày từ trang Worldmeter | 100% |
| Sử dụng nhận xét, code/thuật toán để thể hiện thống kê, trực quan các mối quan hệ giữa các trường dữ liệu | 100% |

Chương 2. Mô tả dữ liệu

1. **Nguồn gốc dữ liệu**

* Dữ liệu được lấy từ trang web của Tổ chức Worldometer (www.worldometers.info)
* Dữ liệu được thu thập từ ngày 07/03/2022 đến ngày 12/03/2022

1. **Ý nghĩa của các trường dữ liệu**

* Country, Other: Quốc gia hoặc các châu lục
* Total Cases: Tổng số ca nhiễm Covid 19 được phát hiện tính tới ngày thu thập
* New Cases: Số ca nhiễm mới được ghi nhận trong ngày
* Total Deaths: Tổng số ca tử vong tích lũy trong số các trường hợp được phát hiện tính tới ngày thu thập
* New Deaths: Số ca tử vong được ghi nhận trong ngày
* Total Recovered: Tổng số ca hồi phục tích lũy trong số các trường hợp được phát hiện tính tới ngày thu thập
* New Recoverd: Số ca hồi phục được ghi nhận trong ngày
* Active Cases: Số ca đang được điều trị. Active Cases = Total Cases - Total Deaths - Total Recoverd
* Serious, Critical: Số ca nhiễm nghiêm trọng hoặc trở nặng tích lũy trong số các trường hợp được phát hiện tính tới ngày thu thập
* TotCases/1M pop: Tỉ lệ người nhiễm Covid 19 so với dân số quốc gia. TotCases/1M pop = (Total Cases / Popular) \* 1000
* Deaths/1M pop: Tỉ lệ người tử vong do Covid 19 so với dân số quốc gia. Deaths/1M pop = (Total Deaths / Popular) \* 1000
* Total Tests: Tổng số ca nghi nhiễm Covid 19 tích lũy trong số các trường hợp được phát hiện tính tới ngày thu thập
* Tests/1M pop: Tỉ lệ các ca nghi nhiễm Covid 19 so với dân số quốc gia. Tests/1M pop = (Total Tests / Popular) \* 1000
* Population: Dân số của một quốc gia

Chương 3. Trực Quan Hóa Dữ liệu

1. **Biểu đồ đường**

* Đầu tiên, ta sử dụng biểu đồ đường (line chart) để so sánh sự gia tăng của dịch bệnh Covid 19 ở các châu lục nói riêng và cả Thế giới nói chung.
* Để có các nhìn tổng quan về dịch bệnh, ta sẽ chọn 3 trường Total Cases, Total Deaths và Total Recoverd để trực quan hóa.
* Biểu đồ đường sẽ cho ta thấy cái nhìn trực quan về tổng số ca nhiễm bệnh, tổng số ca tử vong và tổng số ca hồi phục trong thời gian 6 ngày, từ ngày 07/03/2022 đến ngày 12/03/2022.

Ảnh có chứa bàn

Mô tả được tạo tự động

*H1. Biểu đồ so sánh sự gia tăng số ca nhiễm bệnh*

Ảnh có chứa bàn

Mô tả được tạo tự động

*H2. Biểu đồ so sánh sự gia tăng số ca tử vong*

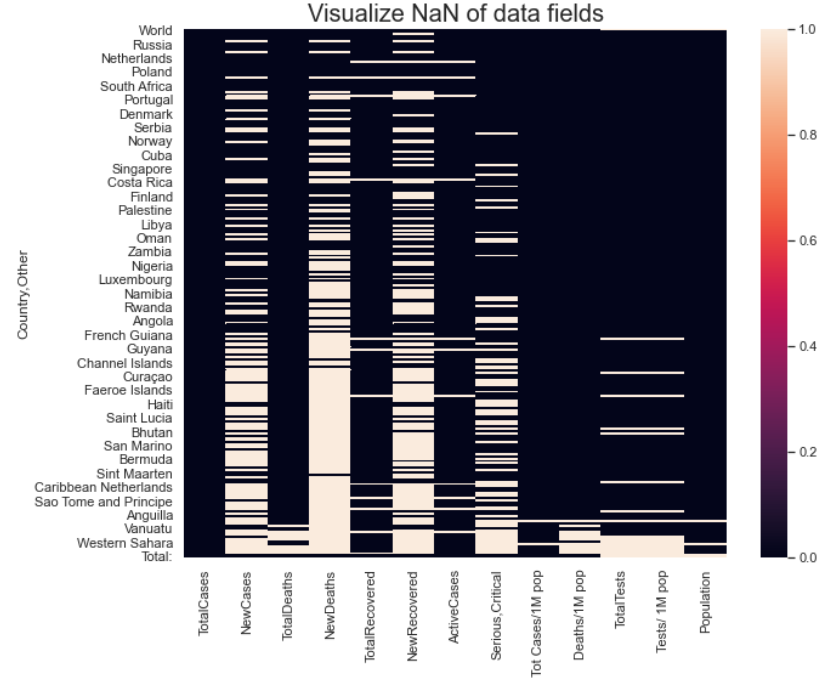
Ảnh có chứa bàn

Mô tả được tạo tự động

*H3. Biểu đồ so sánh sự gia tăng số ca hồi phục*

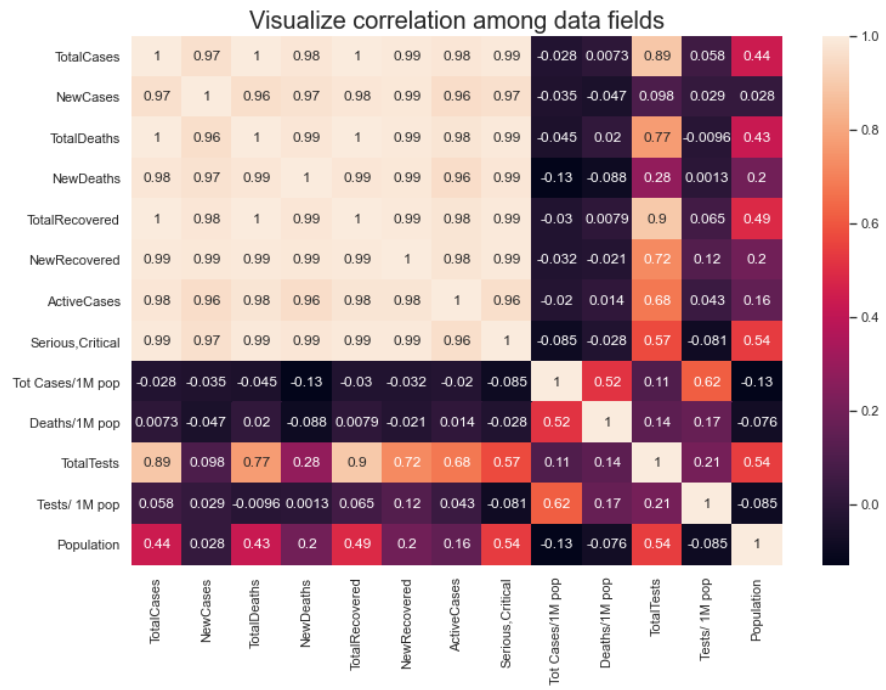
* Trên đây là ba biểu đồ đường so sánh sự gia tăng của 1 trường dữ liệu nào đó giữa Thế giới và các lục địa, và giữa các lục địa với nhau. Trục thẳng đứng là số liệu của trường dữ liệu TotalCases/TotalDeaths/TotalRecovered. Trục nằm ngang là các ngày thu thập dữ liệu.
* Dựa vào 3 biểu đồ, ta có thể thấy số ca nhiễm bệnh và số ca hồi phục của cả thế giới nói chung và của các châu lục nói riêng có sự gia tăng rõ rệt, còn số ca tử vong dường như không thay đổi. Ta có thể lý giải điều này là do hiện nay xuất hiện nhiều biến chủng mới, có khả năng lây lan nhanh hơn, tuy nhiên chinh phủ các nước đã cho người dân tiêm vaccine ít nhất là 2 mũi nên tỉ lệ hồi phục cao hơn. Ta sẽ sử dụng tập dữ liệu của ngày thu thập gần nhất, tức ngày 12/03/2022, để trực quan hóa sự diễn biến dịch bệnh Covid 19 của cả Thế giới cũng như các châu lục.

1. **Bản đồ nhiệt**
   1. **Trực quan hóa giá trị NaN của các trường dữ liệu**



*H4. Biểu đồ trực quan hóa giá trị NaN trong tập dữ liệu*

* Trên đây là bản đồ nhiệt trực quan hóa các giá trị NaN của các trường dữ liệu trong tập dữ liệu tất cả các quốc gia trên thế giới vào ngày gần nhất (12/03/2022). Trục thẳng đứng là tên của các quốc gia. Trục nằm ngang là các trường dữ liệu.
* Nhìn vào bản đồ, ta có thể thấy trường NewDeaths có số lượng giá trị NaN nhiều nhất. Ta có thể lý giải điều này là do các nước đã thực hiện tiêm vaccine cho người dân, vì vậy số ca tử vong của các quốc gia đã giảm đáng kể, cùng với đó, số ca nhiễm mới (NewCases) cũng giảm rõ rệt.
* Các giá trị NaN xuất hiện nhiều nhất ở các quốc gia châu Phi. Điều này cho thấy dịch Covid 19 không diễn biến mạnh ở các quốc gia châu Phi.
  1. **Trực quan hóa sự tương quan giữa các trường dữ liệu**

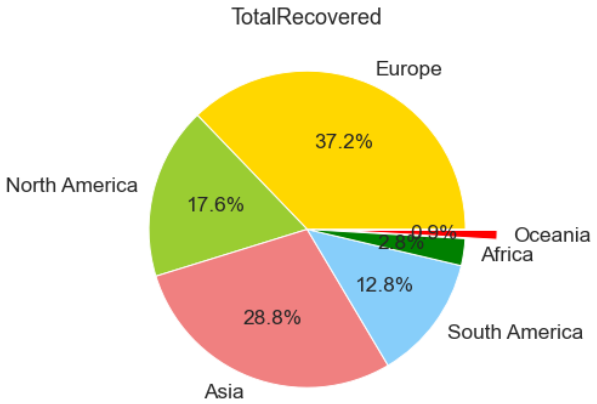
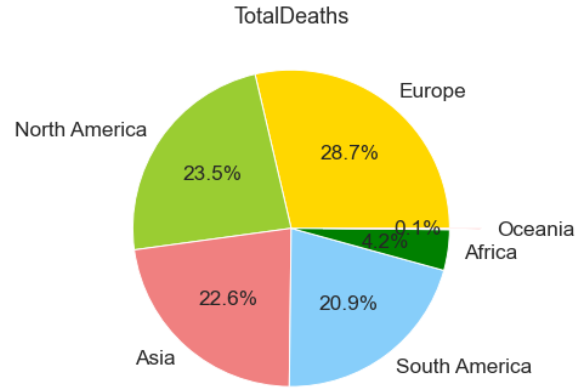
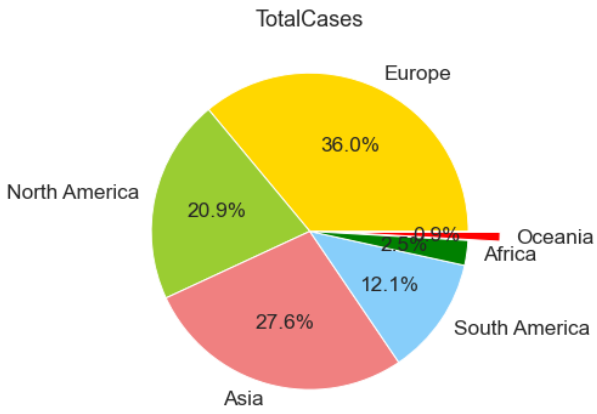


*H5. Biểu đồ thể hiện sự tương quan giữa các trường dữ liệu*

* Trên đây là bản đồ nhiệt thể hiện sự tương quan giữa các trường dữ liệu của tập dữ liệu tất cả các quốc gia trên thế giới vào ngày thu thập gần nhất (12/03/2022).

Từ biểu đồ ta có thể thấy trường TotalCases có sự tương quan với các trường NewCases, TotalDeaths, NewDeaths, TotalRecovered, NewRecovered, ActiveCases và Serious, Critical. Điều này hoàn toàn hợp lý.

1. **Biểu đồ tròn**



*H6. Biểu đồ so sánh số liệu giữa các châu lục về 1 số trường nhất định*

* Trên đây là ba biểu đồ so sánh số ca nhiễm bệnh (TotalCases), số ca tử vong (TotalDeaths) và số ca hồi phục (TotalRecovered) giữa các châu lục.
* Từ 3 biểu đồ trên ta có thể thấy, số ca nhiễm Covid 19 (TotalCases), số ca tử vong (TotalDeaths) và số ca hồi phục (TotalRecovered) ở các quốc gia châu Âu là nhiều hơn cả. Các quốc gia thuộc châu Đại dương có số ca nhiễm bệnh (TotalCases), số ca tử vong (TotalDeaths) và số ca hồi phục (TotalRecovered) ít nhất.

1. **Biểu đồ cột chồng**

* Dưới đây là các biểu đồ cột chồng so sánh số liệu giữa các quốc gia trong cùng 1 châu lục về 1 số trường được chỉ định. Trục thẳng đứng là tên các quốc gia. Trục nằm ngang là số liệu thu thập được.

Ảnh có chứa bàn

Mô tả được tạo tự động

*H7. Biểu đồ so sánh số liệu giữa các quốc gia châu Phi về 1 số trường nhất định*

Ảnh có chứa bàn

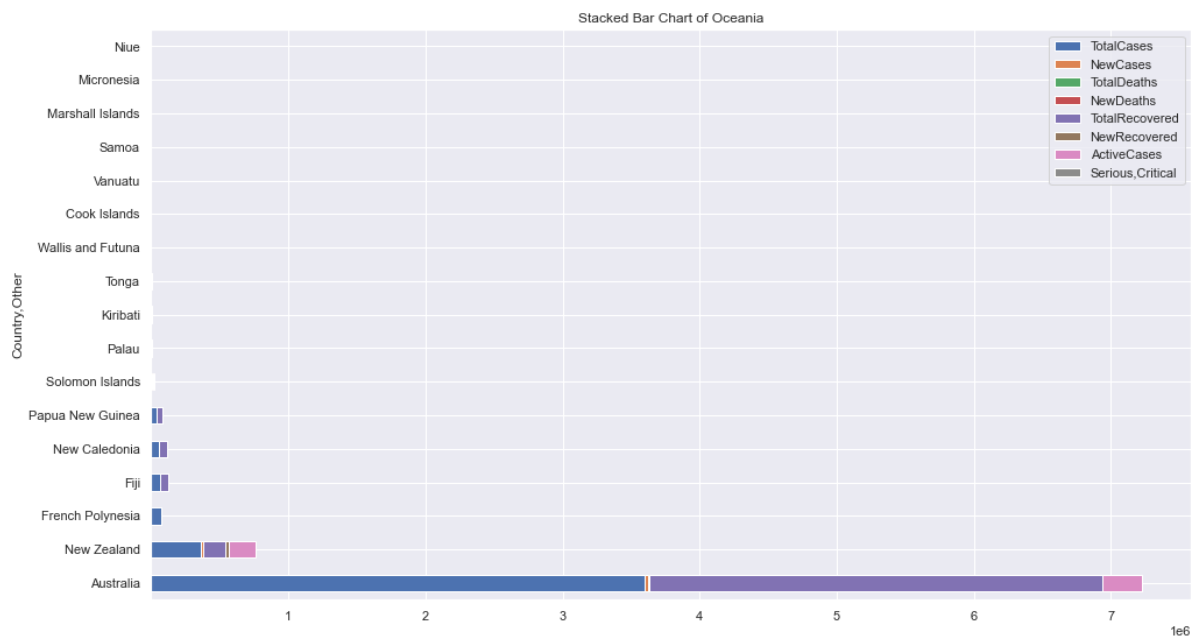
Mô tả được tạo tự động

*H8. Biểu đồ so sánh số liệu giữa các quốc gia châu Á về 1 số trường nhất định*

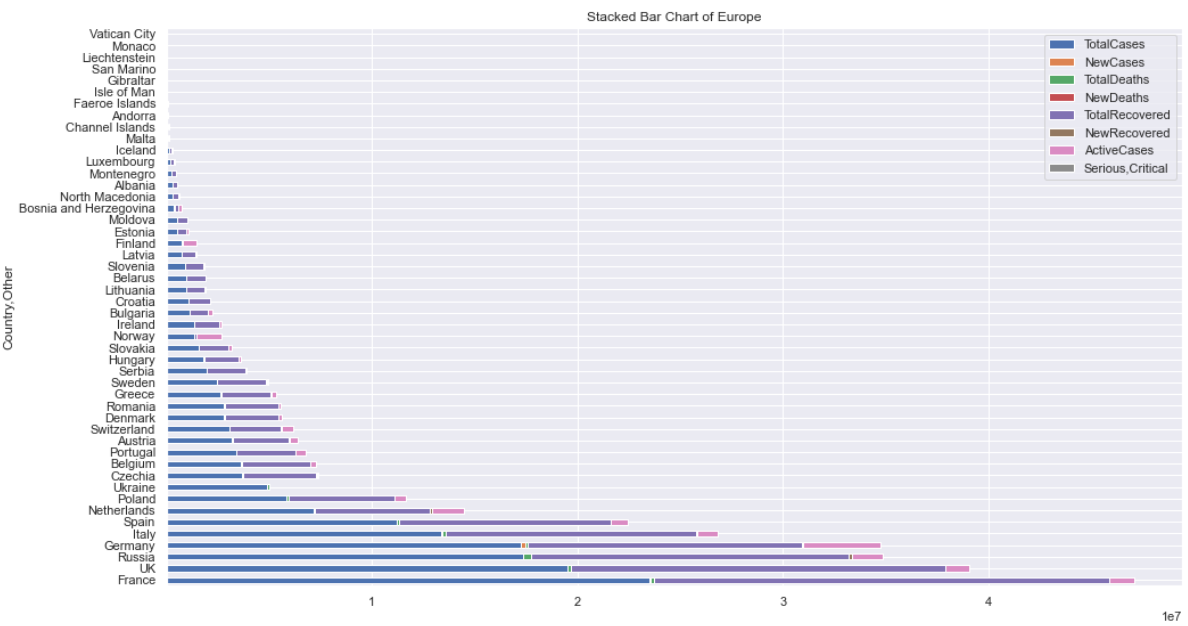
Ảnh có chứa bàn

Mô tả được tạo tự động

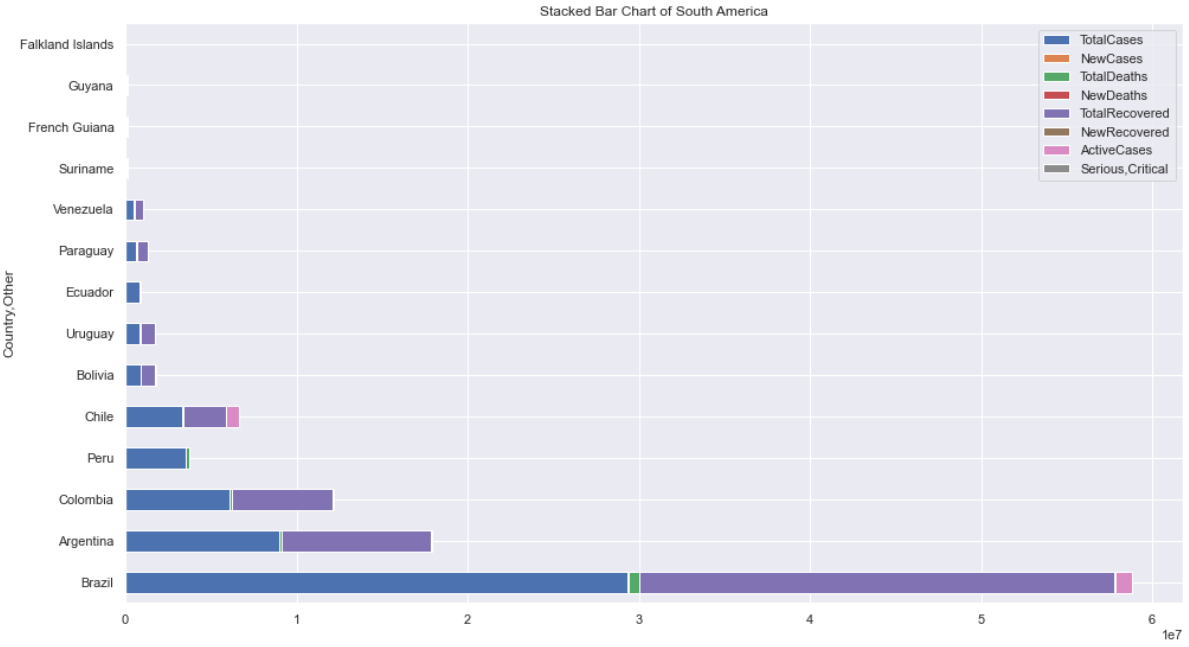
*H9. Biểu đồ so sánh số liệu giữa các quốc gia Bắc Mỹ về 1 số trường nhất định*



*H10. Biểu đồ so sánh số liệu giữa các quốc gia châu Úc về 1 số trường nhất định*



*H11. Biểu đồ so sánh số liệu giữa các quốc gia châu Âu về 1 số trường nhất định*



*H12. Biểu đồ so sánh số liệu giữa các quốc gia Nam Mỹ về 1 số trường nhất định*

* Nhìn chung, dịch bệnh Covid 19 có sự phân bố không đồng đều giữa các quốc gia trong cùng 1 châu lục. Cụ thể, các quốc gia có nền kinh tế phát triển hơn cả trong 1 châu lục sẽ có tình hình dịch bệnh Covid 19 diễn biễn phức tạp hơn.