



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

LAB 1: MỐI QUAN HỆ CỦA DỮ LIỆU
MÔN: TRỰC QUAN HÓA DỮ LIỆU

Sinh viên thực hiện:

Trần Thanh Tùng – 18120258
Trần Hữu Chí Bảo – 18120288
Vòng Cảnh Chi – 18120293
Cao Tất Cường – 18120296
Hà Văn Duy – 18120339

MỤC LỤC

MỤC LỤC	2
PHẦN I: MỨC ĐỘ HOÀN THÀNH	3
1. Mức độ hoàn thành của mỗi yêu cầu:.....	3
2. Mức độ hoàn thành của mỗi thành viên:	3
PHẦN III: THU THẬP DỮ LIỆU	4
PHẦN IV: TRỰC QUAN HÓA DỮ LIỆU.....	5
Xét thuộc tính TotalCase	5
Xét thuộc tính TotalCase với TotalDeaths:	6
Xét thuộc tính TotalCases với TotalTest	8
Xét thuộc tính TotalDeaths, TotalRecovered, ActiveCases	9
Xét thuộc tính TotalCases với Population:	10
Xét thuộc tính TotalScases vs Serious,Critical.....	11
Xét thuộc tính NewCases:.....	12
Phân tích tình hình dịch bệnh giữa các khu vực:.....	13
TÀI LIỆU THAM KHẢO	14

PHẦN I: MỨC ĐỘ HOÀN THÀNH

1. Mức độ hoàn thành của mỗi yêu cầu:

Yêu cầu	Mức độ hoàn thành	Ghi chú
1. Thu thập và tiền xử lý dữ liệu (5%)	100%	Parse dữ liệu và đưa về dạng số liệu phù hợp,
2. Chọn lựa, giải thích, trực quan các trường và các mối quan hệ giữa chúng (50%)	100%	Chọn ra được nhiều trường dữ liệu có liên hệ với nhau và trực quan hóa bằng các biểu đồ thích hợp.
3. Rút ra ý nghĩa hợp lý sau mỗi dữ liệu được trực quan (20%)	100%	Giải thích, đưa ra kết luận và đánh giá phù hợp với tính chất của trường dữ liệu.
4. Xem xét trên nhiều quan hệ, nhiều góc nhìn khác nhau (10%)	100%	Đưa ra kết luận về tính gắn bó, liên quan trực tiếp của các trường dữ liệu.
5. Báo cáo trình bày bố cục và định dạng hợp lý, rõ ràng (15%)	100%	Báo cáo ngắn gọn, rõ ràng.

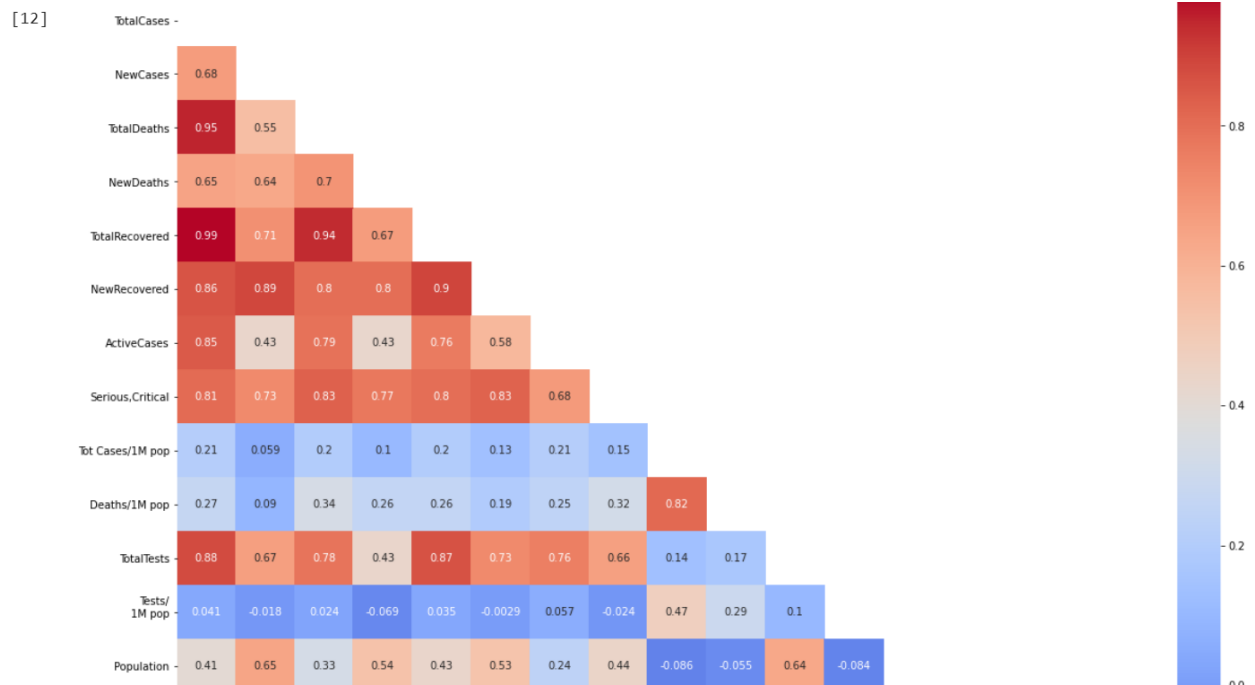
2. Mức độ hoàn thành của mỗi thành viên:

MSSV	Họ tên	Công việc	Hoàn thành
18120258	Trần Thanh Tùng	Thu thập dữ liệu, phân tích, viết báo cáo chi tiết.	100%
18120288	Trần Hữu Chí Bảo	Phân tích dữ liệu, rút ra nhận xét, kết luận về mối quan hệ giữa các trường dữ liệu.	100%
18120293	Vòng Cảnh Chi		100%
18120296	Cao Tất Cường	Phân tích dữ liệu và trực quan hóa dữ liệu bằng các biểu đồ thích hợp.	100%
18120339	Hà Văn Duy		100%

PHẦN III: THU THẬP DỮ LIỆU

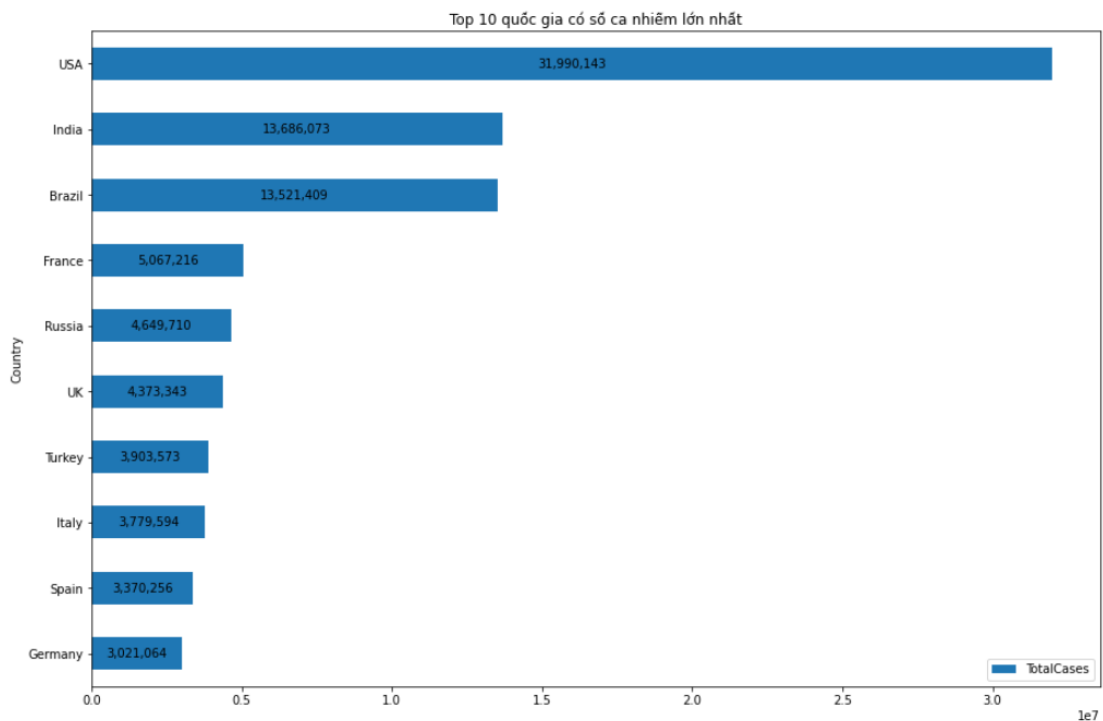
- Thu thập dữ liệu từ trang <https://www.worldometers.info/coronavirus/#nav-yesterday>
- Sử dụng thư viện BeautifulSoup để parseHTML của trang web.

PHẦN IV: TRỰC QUAN HÓA DỮ LIỆU



Hình 1. Biểu đồ heatmap để xem độ tương quan giữa các thuộc tính.

Xét thuộc tính TotalCase:

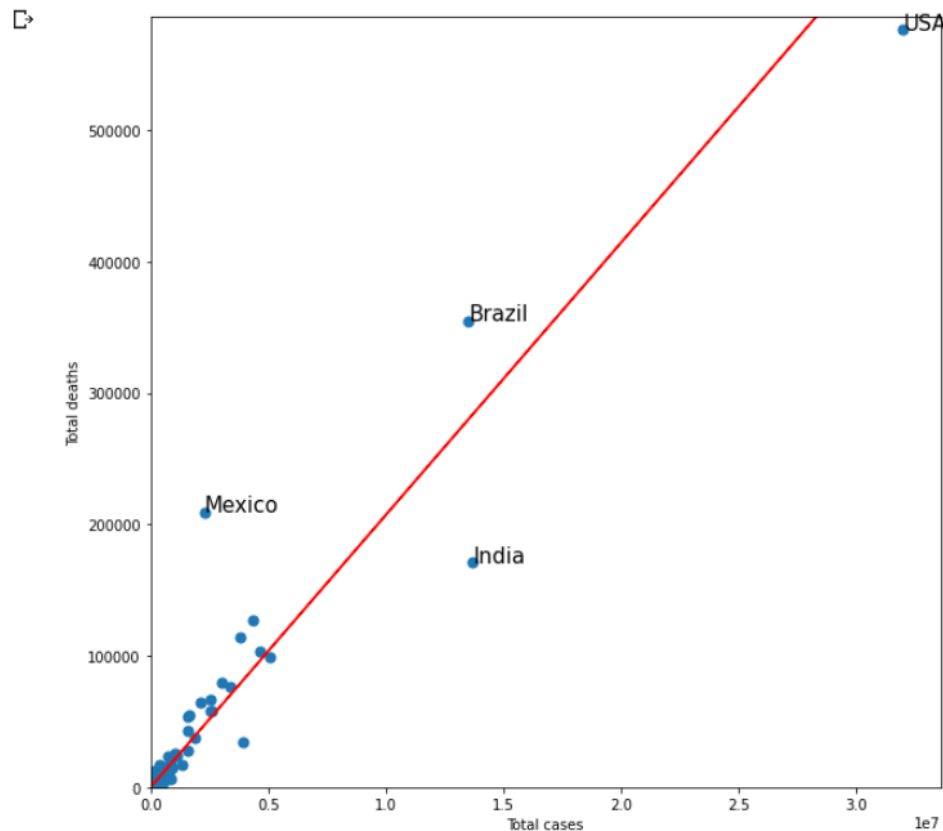


Hình 2. Biểu đồ bar chart thể hiện top 10 quốc gia có số ca nhiễm lớn nhất thế giới.

- **Nhận xét:**

- Biểu đồ trên cho ta cái nhìn tổng quan về tình hình dịch bệnh nổi bật đã diễn ra tại các nước từ khi bùng phát dịch thông qua tổng số ca nhiễm.
- Sử dụng biểu đồ bar chart dạng nằm ngang giúp so sánh tổng số ca nhiễm ở các đất nước khác nhau.
- Nhìn vào biểu đồ, tính đến ngày 9/4/2021 ta thấy nước Mỹ vẫn có số ca nhiễm lớn nhất trên thế giới và gấp hơn 2 lần so với nước có số ca nhiễm nhiều thứ hai.

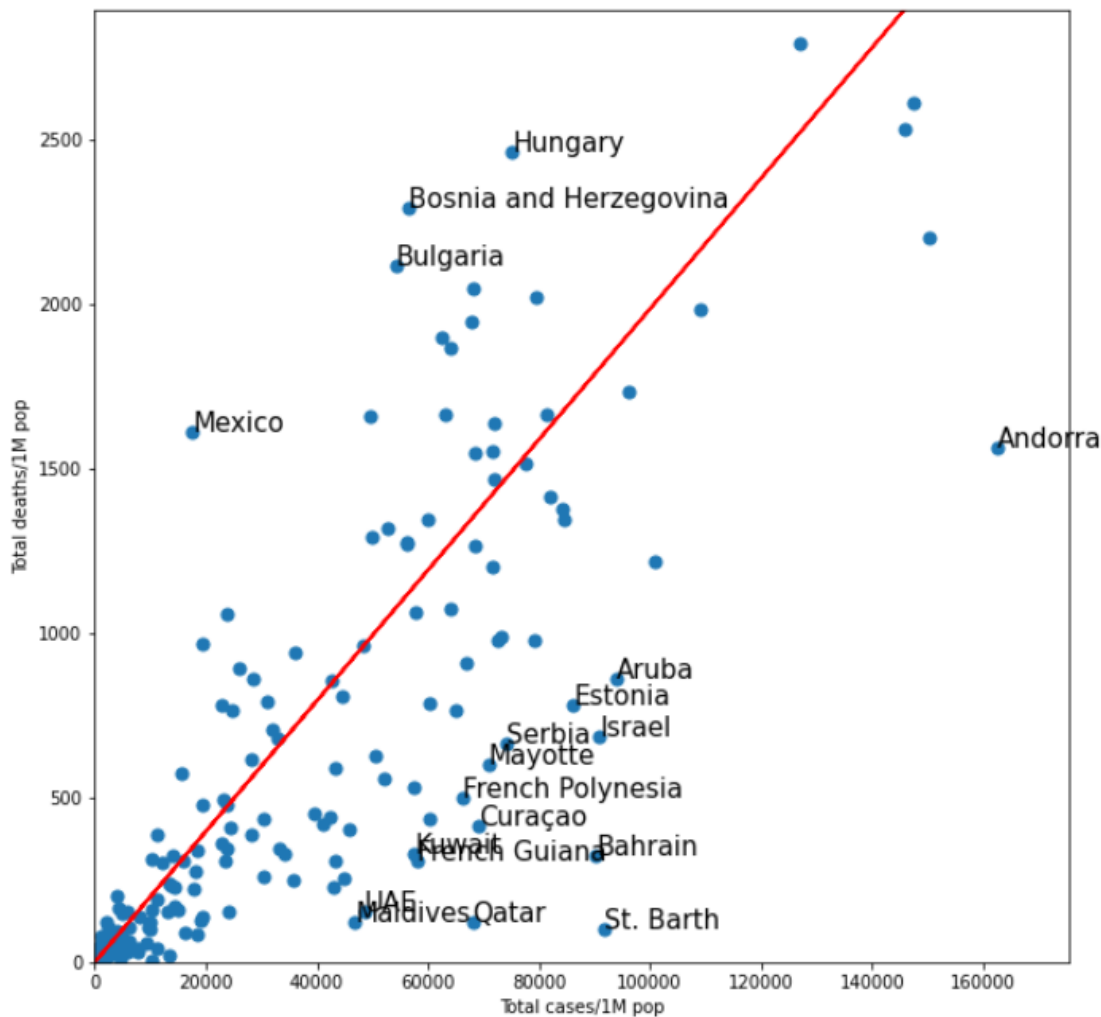
Xét thuộc tính TotalCase với TotalDeaths:



Hình 3. Biểu đồ thể hiện số lượng người chết trong tổng số người bị nhiễm.

- **Nhận xét:**

- Đường màu đỏ thể hiện mức độ tử vong trung bình so với tổng ca nhiễm của các nước trên thế giới.
- Các đất nước nằm phía dưới đường màu đỏ có hệ thống y tế tốt khi có tỉ lệ tử vong so với số ca nhiễm thấp hơn mức trung bình của thế giới.
 - > Vì vậy, những nước khác cần học hỏi hay áp dụng những biện pháp y tế, cách phòng chống dịch covid19 và chữa trị cho người dân nước mình để giảm số ca tử vong xuống thấp nhất có thể.

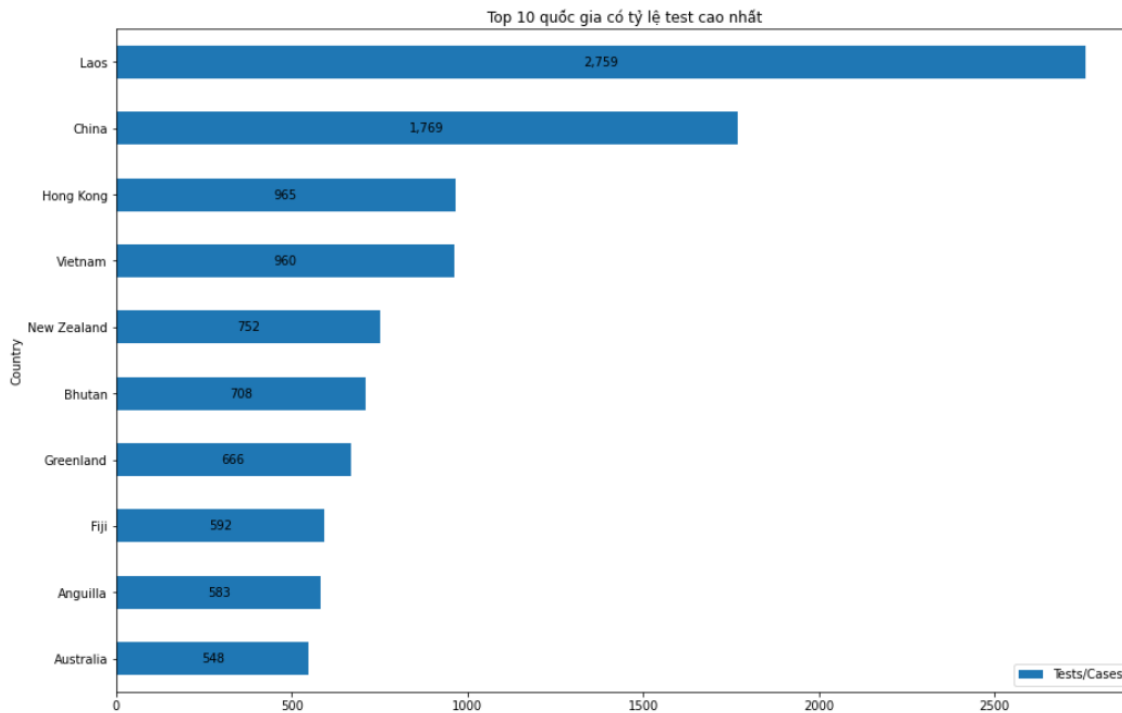


Hình 4. Biểu đồ biểu diễn số lượng người chết so với tổng số ca nhiễm trên 1 triệu dân.

- **Nhận xét:**

- Đường màu đỏ thể hiện mức độ tử vong trung bình so với tổng ca nhiễm trên 1 triệu dân của các nước trên thế giới.
- Các đất nước nằm phía dưới đường màu đỏ có hệ thống y tế tốt khi có tỉ lệ tử vong so với số ca nhiễm thấp hơn mức trung bình của thế giới. Đặc biệt, có một số nước ở Tây Á có số lượng người tử vong rất thấp so với trung bình của toàn thế giới ở cùng mức số lượng người nhiễm.
- Điều đó chứng tỏ, các nước ở Tây Á có đội ngũ y bác sĩ rất tốt và cách điều trị hiệu quả khi dịch bệnh bùng phát.

Xét thuộc tính TotalCases với TotalTest

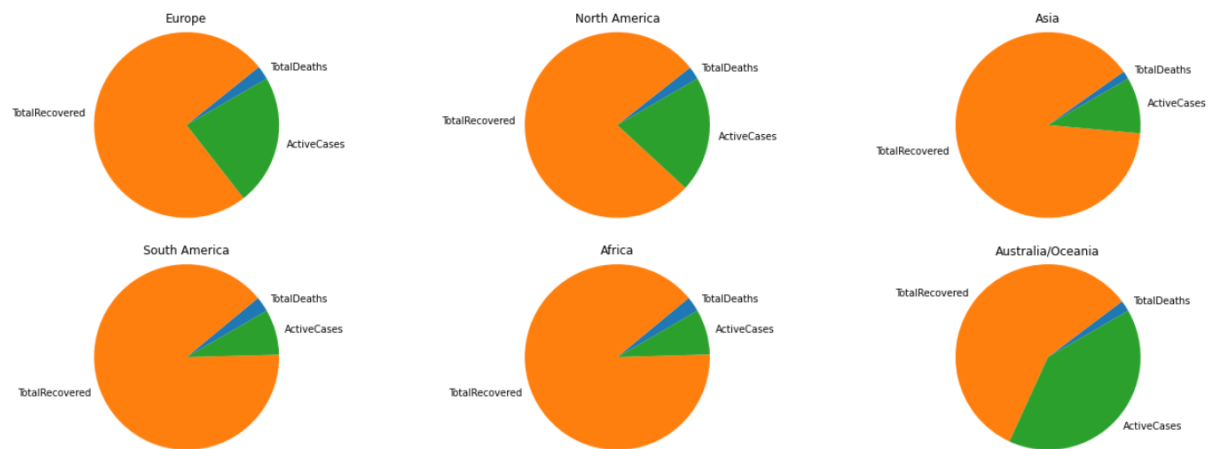


Hình 6. Biểu đồ biểu diễn tỷ lệ số số lượng người đã được xét nghiệm trên số người bị nhiễm.

- **Nhận xét:**

- Sử dụng biểu đồ bar chart dạng nằm ngang giúp so sánh tỉ lệ xét nghiệm trên tổng số ca nhiễm.
- Đây là top 10 quốc gia có tỉ lệ xét nghiệm cao nhất thế giới (tỉ lệ số xét nghiệm trên số ca nhiễm) => Đây là các nước tiêu biểu trong phòng chống dịch bệnh, chủ yếu là các nước châu Á, châu Úc. Việc áp dụng hay học hỏi mô hình chống dịch của những nước này có thể mang lại hiệu quả phòng dịch cao ở những nước, khu vực mà dịch bệnh đang bùng nổ mạnh mẽ.

Xét thuộc tính TotalDeaths, TotalRecovered, ActiveCases

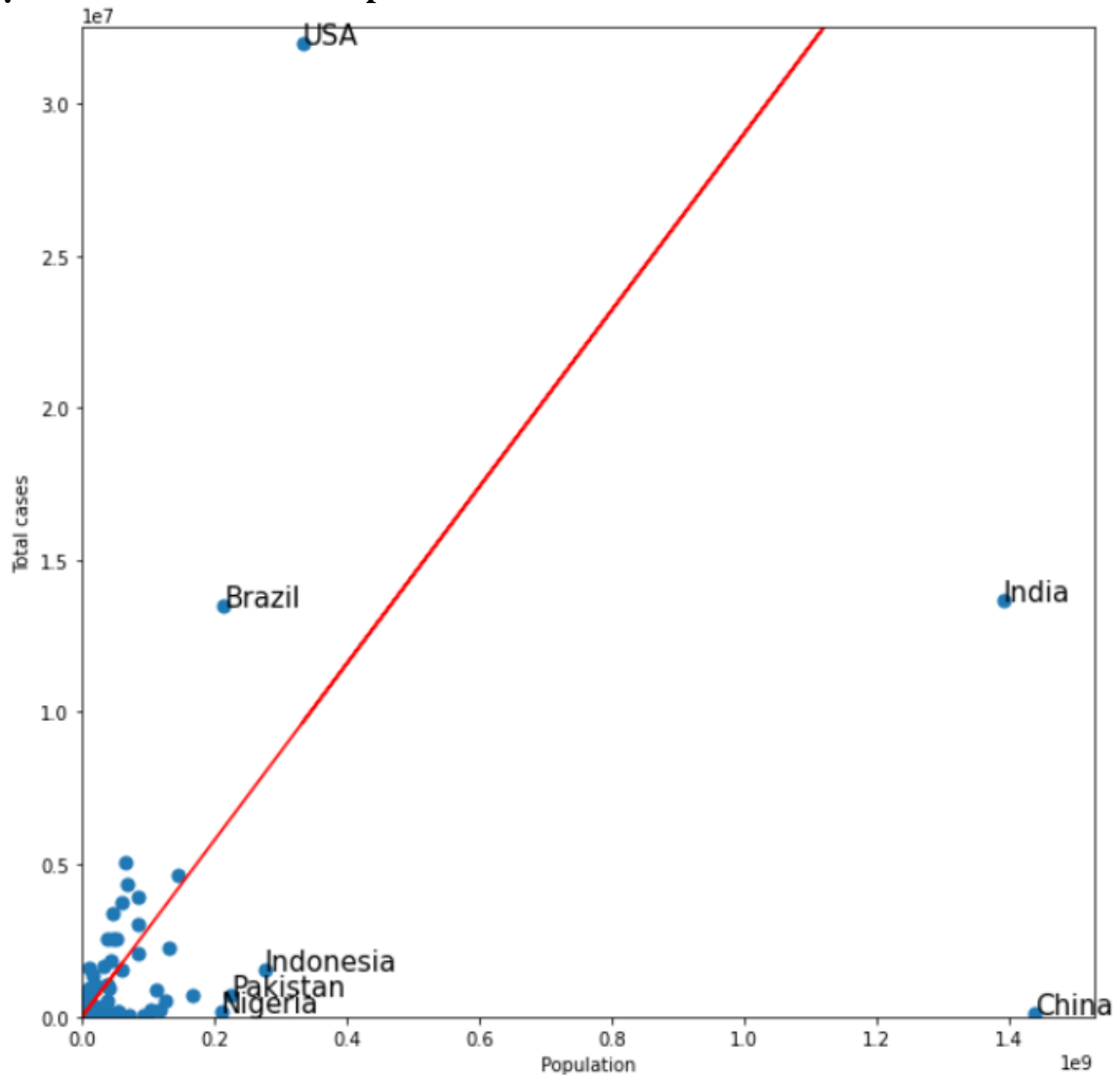


Hình 7. Biểu đồ thể hiện tỷ lệ phần trăm số người tử vong, người đã phục hồi, người đang bị nhiễm ở các châu lục khác nhau.

- **Nhận xét:**

- Mỗi biểu đồ tròn trên thể hiện tỉ lệ số ca nhiễm, số ca đã tử vong và số ca đã chữa khỏi ở 1 khu vực cụ thể trên thế giới.
- Ta nhận thấy rằng tỉ lệ số ca nhiễm ở các khu vực như châu Âu, Bắc Mỹ hay châu Úc đang ở mức cao. Việc chia ra từng biểu đồ tròn riêng biệt như trên giúp ta có thể đánh giá chính xác tình hình dịch bệnh ở từng khu vực cụ thể.

Xét thuộc tính TotalCases với Population:

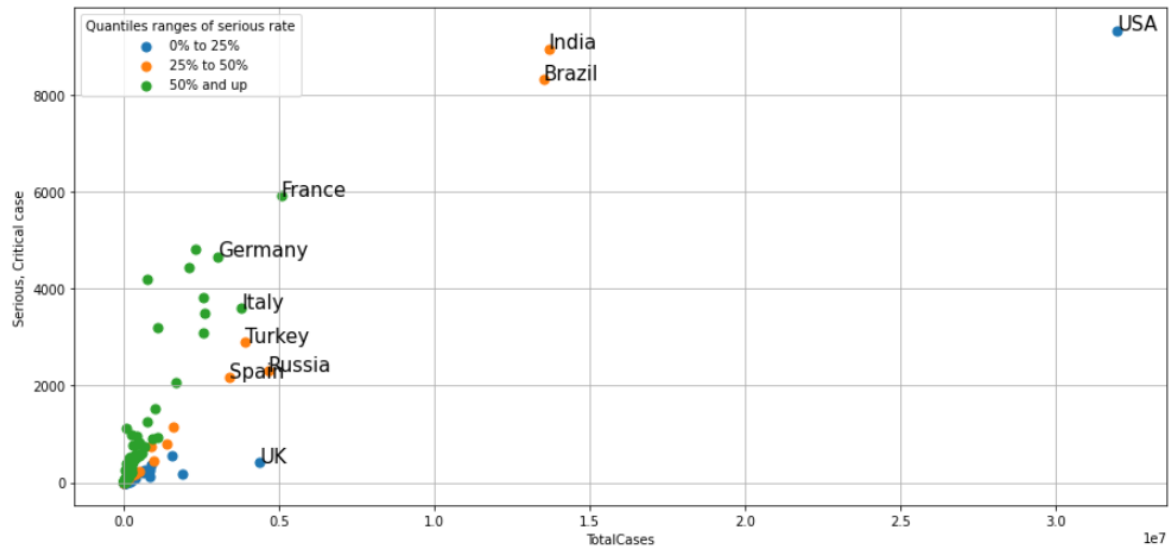


Hình 8. Biểu đồ thể hiện số lượng ca nhiễm so với dân số.

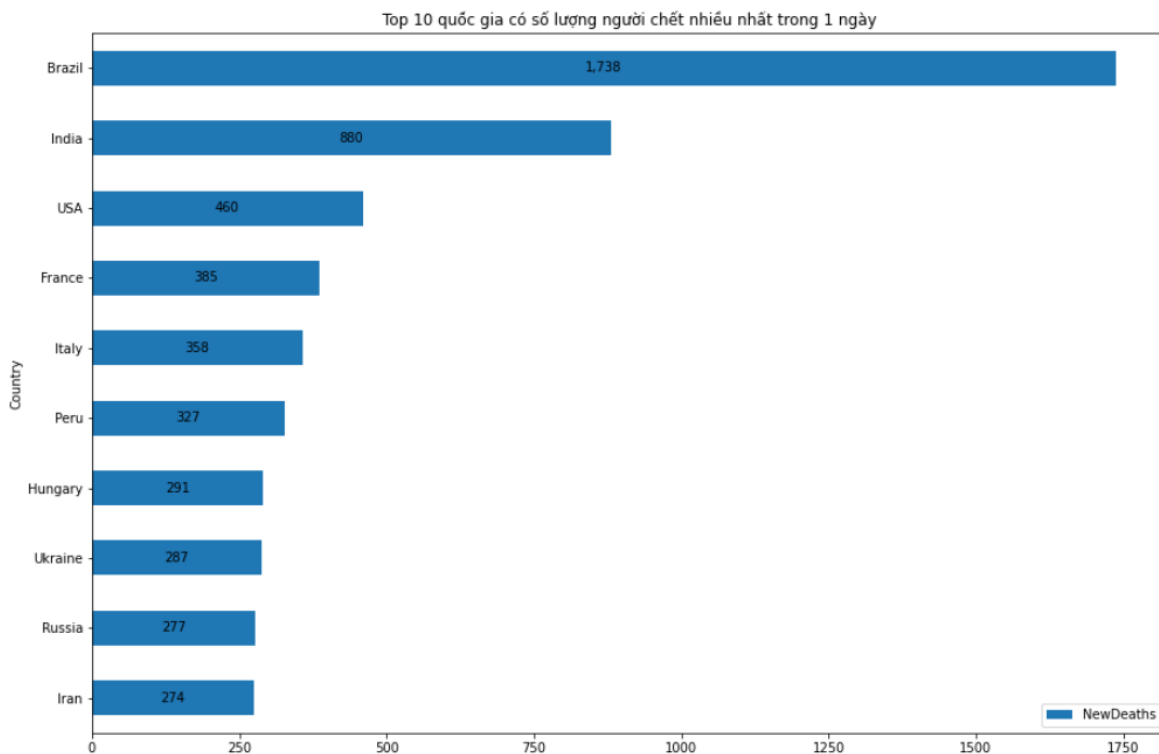
- **Nhận xét:**

- Đây là mối quan hệ giữa số ca nhiễm với dân số của các nước.
- Ta hãy cùng nhau xem xét liệu với quy mô dân số vô cùng lớn của một số nước, dịch bệnh liệu có tăng nhanh và rất khó để kiểm soát hay không (tức là quy mô dân số càng lớn, dịch bệnh sẽ lây lan càng nhanh chóng hay không).
- Đường màu đỏ thể hiện mức trung bình của thế giới về mối liên hệ này. Những nước nằm dưới và càng xa đường này là những nước có dân số lớn tuy nhiên đã có công tác chống dịch vô cùng hiệu quả.
- Vậy là quy mô dân số không phải là yếu tố quyết định và ảnh hưởng chính đến tình hình dịch bệnh ở các nước mà còn có các yếu tố quan trọng khác như là công tác phòng dịch, ý thức người dân, trình độ y tế, kinh tế,...

Xét thuộc tính TotalScases vs Serious,Critical



- Nhận xét:
 - Ở đây, ta xem xét tỉ lệ số ca nghiêm trọng với tổng số ca nhiễm của các nước. Chọn biểu đồ scatter để thể hiện được sự phân bố dữ liệu giữa các nước.
 - Biểu đồ trên thể hiện được tỉ lệ các ca nghiêm trọng của các nước trên thế giới và tổng số ca nhiễm của nước đó. Ta có 3 mức độ được thể hiện ở biểu đồ trên là: tỉ lệ các ca nghiêm trọng ở mức thấp ($< 25\%$), trung bình ($25 \rightarrow 50\%$) và cao (trên 50%).
 - Dựa theo biểu đồ trên, ta có thể đánh giá chính xác được tỉ lệ các ca nghiêm trọng ở các nước để có thể đánh giá được mức độ nghiêm trọng của dịch bệnh để có biện pháp ứng phó phù hợp (ta không thể chỉ nhìn vào số ca đang nguy kịch để cho rằng quốc gia đó đang khủng hoảng trong việc điều trị cho bệnh nhân). Ví dụ như Mỹ, nước có số ca nguy kịch cao nhất hiện nay nhưng nếu xét về tỉ lệ thì mức ca nguy kịch của Mỹ chỉ ở mức thấp so với các nước khác; những nước nhỏ khác tuy có số ca nhiễm thấp nhưng biểu đồ trên lại cho thấy tỉ lệ các ca nguy kịch của nước họ đang ở mức cao,...
 - Nhờ sự đánh giá này mà các nước có thể đánh giá được mức độ nghiêm trọng của dịch bệnh nước họ và có các biện pháp chữa trị phù hợp cho bệnh nhân cũng như kêu gọi sự giúp đỡ của các tổ chức trên thế giới.

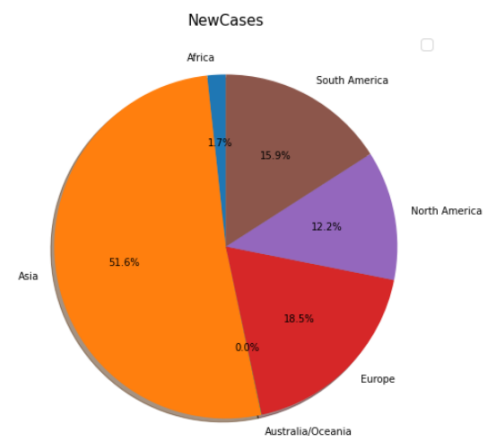
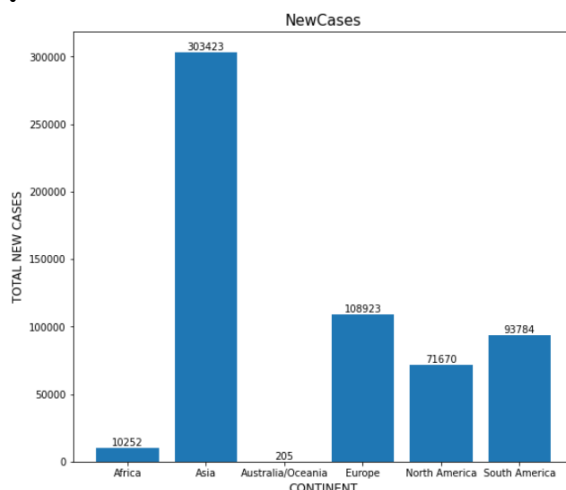


- Hình 9. Biểu đồ thể hiện top 10 quốc gia có số lượng người tử vong nhiều nhất trong 1 ngày.

- **Nhận xét:**

- Ta nhận thấy số lượng người đang mắc bệnh nặng có mối tương quan với số lượng ca tử vong trong ngày tiếp theo.
- Càng nhiều người mắc bệnh nặng thì số lượng tử vong trong ngày tiếp theo và các ngày tiếp sẽ càng nhiều.

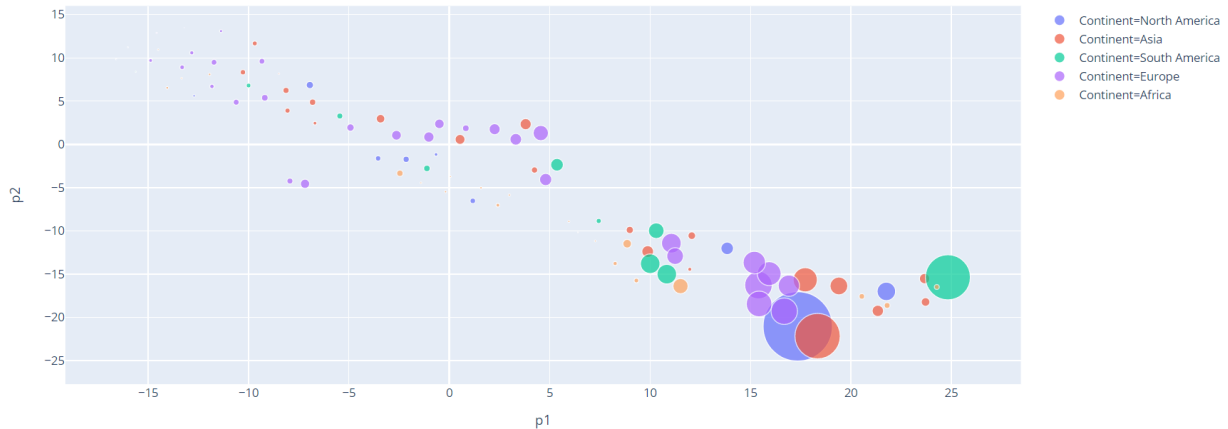
Xét thuộc tính NewCases:



- **Nhận xét:**

- Hai biểu đồ trên thể hiện tỉ lệ số ca nhiễm mới ở các khu vực trên thế giới. Quan sát thấy, dịch bệnh ở châu Á đang diễn biến hết sức phức tạp với số ca nhiễm mới rất nhiều và chiếm hơn 1/2 so với thế giới.
- Biểu đồ trên cho ta cái nhìn tổng quan về tình hình dịch bệnh ở các khu vực trên thế giới để từ đó các chuyên gia có những phân tích về mức độ lây lan dịch ở từng khu vực cụ thể, để rồi các nước sớm biết được các nguy cơ của nước mình để chủ động phòng tránh trước.

Phân tích tình hình dịch bệnh giữa các khu vực:



- **Nhận xét: Dựa theo biểu đồ trên ta thấy được:**
 - Nhóm các nước có số ca mắc cao có xu hướng tách rời nhóm có số ca mắc thấp.
 - Giữa các châu lục, vùng lãnh thổ có tình hình dịch bệnh không khác biệt, điều đó có thể giúp ta cơ bản loại bỏ được giả thuyết virus được tạo ra để đánh vào một khu vực cụ thể nào đó.

TÀI LIỆU THAM KHẢO

- www.python-graph-gallery.com ngày 20 tháng 4 năm 2021
- <https://seaborn.pydata.org> ngày 20 tháng 4 năm 2021
- <https://matplotlib.org> ngày 22 tháng 4 năm 2021
- <https://www.w3schools.com> ngày 22 tháng 4 năm 2021