



Phân tích dữ liệu thông minh

Báo cáo đồ án giữa kì

Phân tích những yếu tố ảnh hưởng tới
tình hình dịch bệnh COVID-19 ở Mỹ

Students

21127027 - Đinh Hoàng Duy

21127406 - Trần Đình Quang

21127657 - Nguyễn Khánh Nhân

21127666 - Trần Thuận Phát

Lecturers

Bùi Tiến Lên

Nguyễn Ngọc Đức

Trương Tấn Khoa

Contents

1 - Giới thiệu vấn đề	3
2 - Chi tiết về bộ dữ liệu	4
2.1 COVID Racial Data Tracker	4
2.2 COVID-19 US State Policy Database	4
3 - Tổng quan tình hình COVID-19 tại Mỹ	5
4 - Phân tích yếu tố sắc tộc ảnh hưởng như nào đến tình hình COVID-19 ở Mỹ	8
4.1 Đặt giả thuyết:	8
4.2 Phương pháp tiền xử lý dữ liệu	9
4.3 Mô tả tập dữ liệu cho giả thuyết thứ nhất:	9
4.4 Trực quan hình ảnh để làm rõ giả thuyết thứ nhất:	10
4.4.1 Sự phân bố dữ liệu về số ca nhiễm dương tính và số ca tử vong của 2 nhóm sắc tộc POCs và White:	10
4.4.2 Số lượng trung bình về số ca nhiễm dương tính và số ca tử vong của 2 nhóm sắc tộc POCs và White	11
4.4.3 Sự phân bố số ca nhiễm dương tính và số ca tử vong qua các tháng trong 2 năm 2020 và 2021	12
4.5 Mô phỏng tập dữ liệu trên qua phương thức bootstrap để làm rõ giả thuyết thứ nhất	14
4.6 Kết luận giả thuyết thứ nhất	15
4.7 Trực quan hình ảnh để làm rõ giả thuyết thứ hai	16
4.7.1 Sự phân bố dữ liệu về số ca nhiễm và tử vong của từng nhóm dân tộc theo thời gian	16
4.7.2 Phân tích chỉ số People of Color Death Rate (POCDR) cho các nhóm sắc tộc ở các bang khác nhau:	17
4.8 Áp dụng phương pháp gom nhóm các chủng tộc theo tổng số ca nhiễm và tử vong làm rõ giả thuyết thứ hai	18
4.9 Kết luận giả thuyết thứ hai	18
5 - Phân tích sự ảnh hưởng của các chính sách xã hội của các bang đến tình hình COVID-19 tại Mỹ	18
5.1 Mô tả dữ liệu và lấy ra bang có số lượng ca nhiễm cao nhất để kiểm định giả thuyết	19

5.2	Tiền xử lý dữ liệu	19
5.3	Trực quan hóa dữ liệu	20
5.4	Áp dụng phương pháp hồi quy tuyến tính	22
5.5	Kết luận giả thuyết	23
6	- Phân tích sự ảnh hưởng của yếu tố con người đến tình hình COVID-19 tại Mỹ	24
6.1	Tiền xử lý dữ liệu	24
6.2	Mô tả dữ liệu	24
6.3	Trực quan hóa dữ liệu	25
6.4	Áp dụng phương pháp hồi quy tuyến tính	27
6.5	Sử dụng ma trận hệ số tương quan <i>Pearson</i>	28
6.6	Kết luận giả thuyết	29
7	- Kết luận	30

1 - Giới thiệu vấn đề

- Đại dịch bệnh Coronavirus Disease 2019 (COVID-19) đã gây ra hậu quả nghiêm trọng trên khắp nước Mỹ và trên toàn thế giới. Các ca nhiễm đầu tiên của virus này được phát hiện tại ‘Vũ Hán, Trung Quốc’, và hơn **1,5 triệu người** trên toàn thế giới đã tử vong vì căn bệnh này. Tỷ lệ lây nhiễm COVID-19 ngày càng gia tăng tại Mỹ, cùng với tỷ lệ tử vong cao, đã thúc đẩy các nhà lãnh đạo phải xây dựng các biện pháp can thiệp không dựa trên được phẩm (NPIs) nhằm kiểm chế sự lây lan của dịch bệnh. Các biện pháp NPIs này bao gồm cách ly và kiểm dịch, giãn cách xã hội, khẩu trang và vệ sinh tay. Tại một quốc gia như Mỹ, nơi tập trung đa dạng chủng tộc và quyền con người được coi trọng cao, thì việc thực thi các chính sách y tế hạn chế để kiểm soát sự lây lan của virus đã trở thành một nhiệm vụ thách thức. Điều này có thể được quy cho sự đa dạng về các thiên hướng chính trị, tình trạng kinh tế-xã hội.

- Ngoài việc áp dụng các NPIs, một số chính sách khác cũng đã được thực thi để kiểm chế sự lây lan của virus [6]. Các chính sách này bao gồm đóng cửa các nhà hàng và quán bar, học tập trực tuyến tại các trường học, phong tỏa, giới nghiêm và khuyến cáo hạn chế di chuyển. Theo chỉ đạo của chính phủ liên bang, những chính sách này đã được thực thi dựa trên quyền quyết định của chính quyền tiểu bang. Các nghiên cứu trước đây đã khảo sát hiệu quả của một số chính sách này đối với tỷ lệ lây nhiễm và số ca tử vong.

- Tại Mỹ, người da màu (POC), đặc biệt là người da đen, đang tử vong với tỷ lệ cao hơn so với người da trắng. Lý do cho tỷ lệ cao hơn này là do các người da màu thường làm các nghề nghiệp thiết yếu như nhân viên siêu thị bán lẻ, nhân viên giao thông công cộng, nhân viên chăm sóc sức khỏe và nhân viên vệ sinh. Những loại công việc này không luôn tuân thủ các biện pháp giãn cách xã hội, làm tăng khả năng lây nhiễm cho những người lao động. Một tác động khác đến tỷ lệ tử vong tương đối là thiếu bảo hiểm y tế trong cộng đồng POC. Việc thiếu bảo hiểm làm giảm khả năng những người bị bệnh đi bệnh viện do lo ngại không thể trang trải được chi phí chăm sóc. Do POC, đặc biệt là người da đen, có nhiều khả năng mắc các bệnh nền như tiểu đường, cao huyết áp và bệnh tim, nguy cơ tử vong do COVID-19 cao hơn đáng kể so với mức trung bình quốc gia.

- Đóng góp chính của bài phân tích này là định lượng sự chênh lệch về chủng tộc của COVID-19 đối với các ca dương tính và tử vong trên toàn Mỹ và lãnh thổ, cả riêng lẻ và tổng thể. Điều này bao gồm một cuộc điều tra về các chính sách đã được các tiểu bang thực hiện.

Mục đích của bài phân tích: Mỹ là một trong những quốc gia chịu ảnh hưởng nặng nề bởi dịch bệnh COVID-19 nguyên nhân có thể đến từ nhiều yếu tố khác nhau như chính sách, chủng tộc, vị trí địa lý, v.v chính bởi vì đây là một đất nước lớn, đa dạng và phức tạp cả về xã hội, chính trị lẫn kinh tế. Bài phân tích này sẽ tập trung vào việc phân tích các yếu tố ảnh hưởng đến dịch bệnh COVID-19 tại Mỹ, qua đó đưa ra những kết luận, nhận định về những yếu tố ảnh hưởng lớn đến dịch bệnh này. Cuối cùng bài phân tích này sẽ kiểm định 4 giả thuyết chính liên quan đến dịch bệnh COVID-19 tại Mỹ:

- Giả thuyết thứ nhất: điều tra xem liệu số ca nhiễm và tỷ lệ tử vong của nhóm dân tộc có màu da (POC) có cao hơn nhóm dân tộc Da Trắng hay không?
- Giả thuyết thứ hai: nghiên cứu xem liệu có sự khác biệt đáng kể về số ca nhiễm xác nhận và tỷ lệ tử vong giữa các nhóm dân tộc trên khắp Hoa Kỳ và lãnh thổ hay không?
- Giả thuyết thứ ba: điều tra xem liệu sự kiểm soát của đảng chính trị và các chính sách chính phủ có ảnh hưởng đến tỷ lệ ca nhiễm và tử vong giữa các nhóm dân tộc hay không?
- Giả thuyết thứ tư: điều tra xem liệu mật độ dân cư và các yếu tố con người khác của mỗi bang có phải là yếu tố ảnh hưởng lớn đến số ca nhiễm và tỷ lệ tử vong không?

2 - Chi tiết về bộ dữ liệu

- Bài phân tích này sẽ kết hợp sử dụng 2 tập dữ liệu chính.

2.1 COVID Racial Data Tracker

- **CRDT_Data.csv**: hay còn được gọi là COVID Racial Data Tracker. Tập dữ liệu Theo dõi Dữ liệu Chứng tộc COVID (CRDT) được chia thành bốn danh mục:

- Số ca nhiễm bao gồm các ca nhiễm COVID-19 được xác nhận và có khả năng nhiễm.
- Số ca tử vong bao gồm các ca tử vong do COVID-19 được xác nhận và có khả năng xảy ra.
- Số ca nhập viện bao gồm những người đã được nhập viện do COVID-19.
- Số xét nghiệm bao gồm tổng số xét nghiệm (cả dương tính và âm tính) được thực hiện để chẩn đoán COVID-19.

- Mỗi danh mục đều có thông tin về ngày và bang, và 14 chủng tộc được phân chia thành 12 nhóm dân tộc: người da trắng, người da đen, người Latinh, người Á châu, người Da đỏ và người Bản địa Alaska (AIAN), người Hawaii bản địa và người Đảo Thái Bình Dương (NHPI), Đa chủng tộc, Khác, Không rõ, Gốc Tây Ban Nha, Không phải Gốc Tây Ban Nha và Không rõ Gốc.

- Bài phân tích chỉ này tập trung vào **6** nhóm dân tộc đầu tiên vì nền tảng sắc tộc của họ đã được xác nhận đầy đủ, và dữ liệu về số ca nhiễm và số ca tử vong của từng nhóm dân tộc được đánh giá. Các nhóm dân tộc khác, số ca nhập viện và số xét nghiệm đã bị bỏ qua vì nghiên cứu này tập trung vào số ca nhiễm và số ca tử vong thay vì tổng số ca nhập viện và xét nghiệm.

- Dữ liệu ghi nhận số ca dương tính, tử vong, nhập viện và test được thực hiện tại từng bang của Mỹ, từ ngày 12/4/2020 đến ngày 7/3/2021. Có tổng cộng 5320 dòng và 54 cột trong tập dữ liệu này.

- Dù có tận 54 cột nhưng thực chất chỉ có 6 nhóm cột chính cần quan tâm là **Date, State, Cases, Deaths, Hosp và Tests**.

Cột	Ý nghĩa
Date	Ngày ghi nhận
State	Tên tiểu bang
Cases_*	Tổng số ca dương tính ghi nhận trong ngày cho từng nhóm người thuộc các chủng tộc khác nhau. * ở đây có thể là White, Black,...
Deaths_*	Tổng số ca chết ghi nhận trong ngày cho từng nhóm người thuộc các chủng tộc khác nhau. * ở đây có thể là White, Black,...
Hosp_*	Tổng số ca nhập viện ghi nhận trong ngày cho từng nhóm người thuộc các chủng tộc khác nhau. * ở đây có thể là White, Black,...
Tests_*	Tổng số ca test ghi nhận trong ngày cho từng nhóm người thuộc các chủng tộc khác nhau. * ở đây có thể là White, Black,...

2.2 COVID-19 US State Policy Database

- **Policies_Social_COVID_19.xlsx**: Tập dữ liệu này được thu thập từ COVID-19 US State Policy Database, chứa thông tin về các chính sách xã hội liên quan đến COVID-19 đã

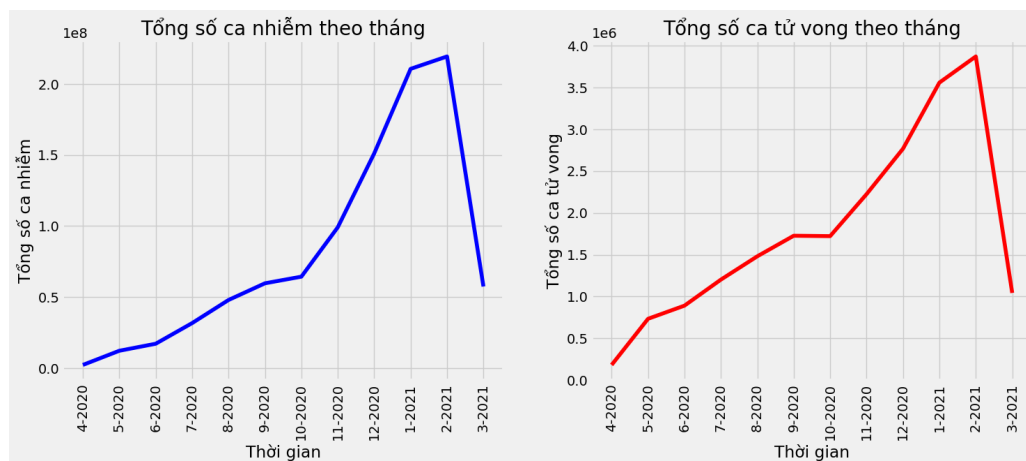
được các tiểu bang thực hiện. Các chính sách này bao gồm đóng cửa các nhà hàng và quán bar, học tập trực tuyến tại các trường học, phong tỏa, giới nghiêm và khuyến cáo hạn chế di chuyển. Cụ thể, tập dữ liệu này chứa thông tin về ngày thực thi, ngày kết thúc các chính sách do tất cả các tiểu bang thực hiện trong giai đoạn xảy ra COVID-19. Cuối cùng còn có thông tin về tình hình xã hội của các tiểu bang này như mật độ dân số, tỉ lệ người vô gia cư, số lượng các chuyên gia dinh dưỡng ở mỗi bang.

- Bộ dữ liệu có tổng cộng 50 dòng và 15 cột:

Cột	Ý nghĩa
STATE	Tên tiểu bang
POSTCODE	Mã bưu chính của tiểu bang
FIPS	Mã số FIPS của tiểu bang
FM_ALL	Ngày thực thi chính sách bắt buộc đeo khẩu trang ở nơi công cộng
FM_END	Ngày kết thúc chính sách bắt buộc đeo khẩu trang ở nơi công cộng
CLREST	Ngày thực thi chính sách đóng cửa nhà hàng và chỉ cho mua đem về
ENDREST	Ngày cho phép nhà hàng phục vụ mở cửa trở lại
CLMOVIE	Ngày thực thi chính sách đóng cửa rạp chiếu phim
END_MOV	Ngày cho phép rạp chiếu phim mở cửa trở lại
CLOSEBAR	Ngày thực thi chính sách đóng cửa quán bar
END_BRS	Ngày cho phép quán bar mở cửa trở lại
POPDEN21	Mật độ dân số của bang năm 2021
HMLS21	Số lượng người vô gia cư của bang năm 2021
MH21	Số lượng các chuyên gia dinh dưỡng trung bình có trong 100,000 người ở mỗi bang năm 2021

3 - Tổng quan tình hình COVID-19 tại Mỹ

- Trước khi đi sâu hơn vào việc phân tích các yếu tố ảnh hưởng đến dịch bệnh COVID-19 tại Mỹ, chúng ta sẽ xem xét tổng quan về tình hình dịch bệnh tại Mỹ từ tháng 4/2020 - 3/2021.



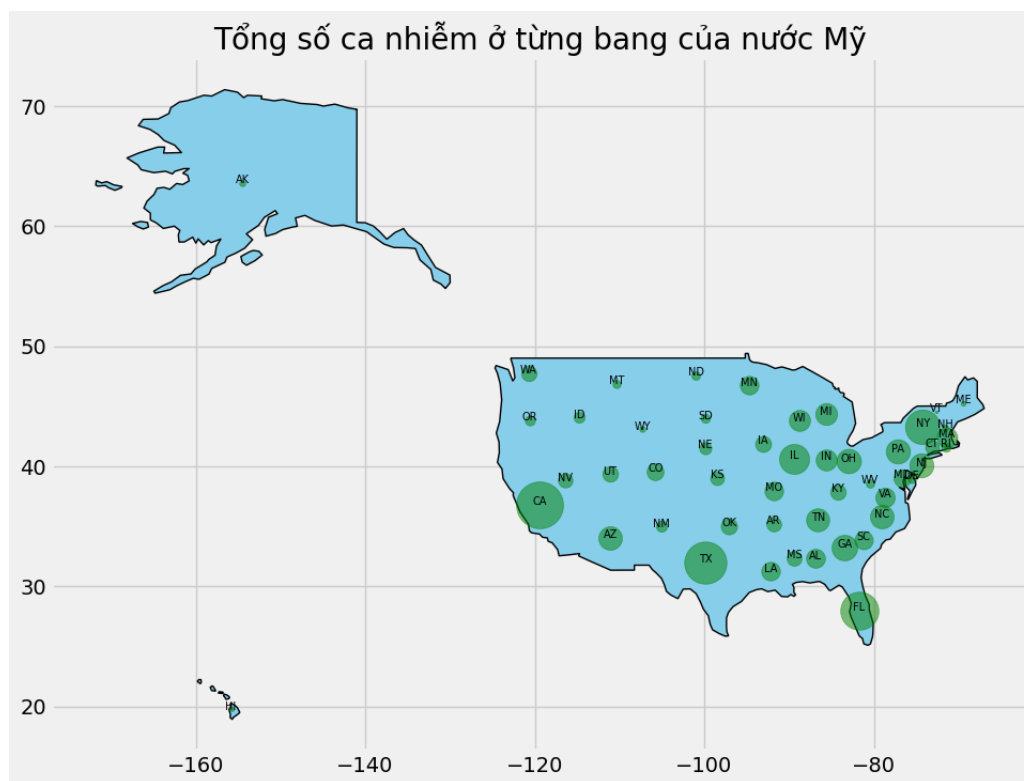
Hình 1. Tổng số ca nhiễm và tử vong theo thời gian

- Nhận xét:

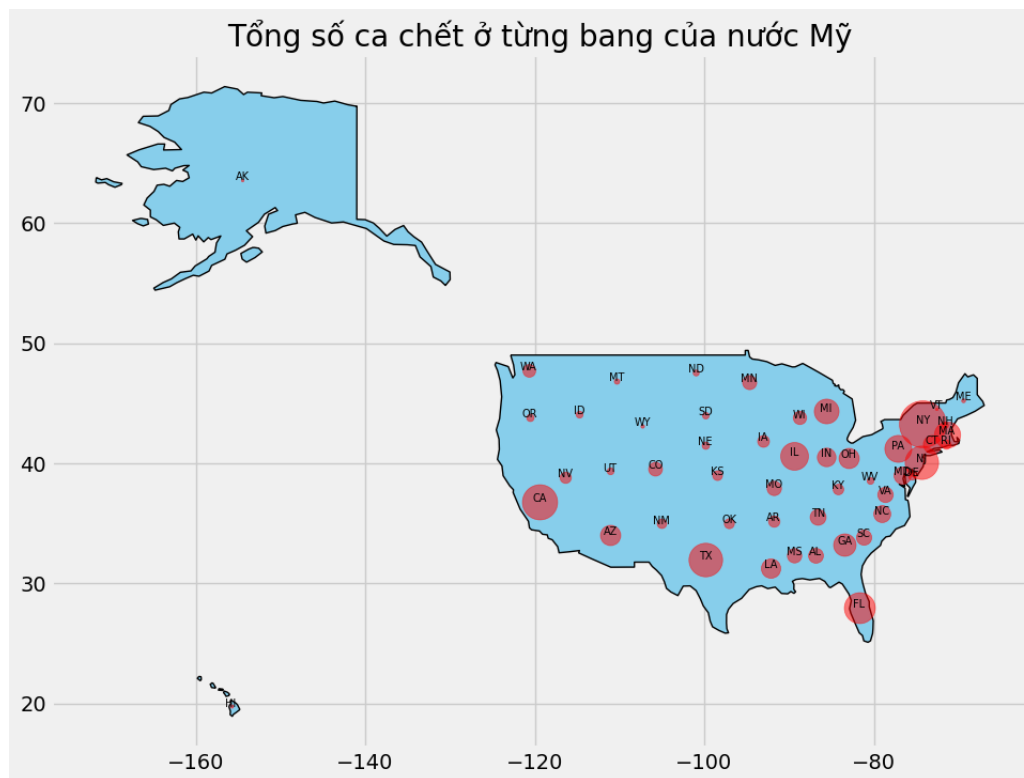
- Quan sát 2 biểu đồ trên, ta thấy rằng số ca nhiễm và số ca tử vong có điểm tương đồng rất lớn với nhau (nếu không muốn nói là y hệt nhau), cụ thể là đều tăng mạnh từ tháng 4/2020 đến sang năm 2021, đạt đỉnh điểm vào tháng 2 năm 2021, tuy nhiên sau đó lại giảm rất mạnh ở tháng 3/2021 (giảm đi gấp 4 lần so với tháng 2/2021). Điều này cho

thấy rằng dịch bệnh COVID-19 đã gây ra hậu quả nghiêm trọng tại Mỹ trong giai đoạn từ tháng 4/2020 đến tháng 2/2021, sau đó đã giảm đi rất nhiều chỉ trong vòng 1 tháng cho thấy việc kiểm soát dịch bệnh tại Mỹ bằng nhiều biện pháp khác nhau đã được áp dụng một cách nghiêm ngặt và hiệu quả.

- Một điều đáng chú ý là Mỹ có thể làm rất tốt trong việc kiểm soát dịch bệnh, cụ thể là số ca nhiễm và tử vong giảm đi rất nhanh chỉ trong vòng 1 tháng, tuy nhiên họ lại thực hiện việc này quá muộn, mãi đến tận 1 năm sau khi dịch bệnh bùng phát mới thực hiện các biện pháp phòng chống dịch bệnh một cách nghiêm ngặt và hiệu quả.
- Cuối cùng là sự giảm đi rất nhanh của số ca nhiễm và tử vong vào tháng 3/2021 cho thấy rằng việc kiểm soát dịch bệnh tại Mỹ đã đạt được kết quả tốt, tuy nhiên nguyên nhân đằng sau sự giảm này là do việc tiêm vắc xin COVID-19 đã được triển khai rộng rãi tại Mỹ, giúp giảm nguy cơ lây nhiễm và tử vong do COVID-19. Theo Cơ quan Quản lý Thực phẩm và Dược phẩm Hoa Kỳ (FDA) [2] thì mũi tiêm vắc xin đầu tiên đã được phê duyệt vào ngày 11/12/2020, và sau đó đã được triển khai rộng rãi tại Mỹ, giúp giảm nguy cơ lây nhiễm và tử vong do COVID-19. Vì vậy rất có thể rằng sự giảm mạnh của số ca nhiễm và tử vong vào tháng 3/2021 là do việc tiêm vắc xin COVID-19 đã được triển khai rộng rãi tại Mỹ từ tháng 12/2020.



Hình 2. Quy mô ca nhiễm ở từng bang



Hình 3. Quy mô ca tử vong ở từng bang

- Nhân xét:

- Quan sát 2 bản đồ trên ta đều thấy được toàn bộ các bang của nước Mỹ đều bị ảnh hưởng bởi ‘COVID-19’. Số ca nhiễm và tử vong của từng bang đều tỉ lệ thuận với nhau. Tức là nếu một bang có tỉ lệ người nhiễm bệnh cao thì đồng nghĩa với đó là số người tử vong cũng cao theo.
- Về sự phân bố của dịch bệnh thì có một điều thú vị là các bang nằm ở phía đông nước Mỹ như ‘New York’, ‘New Jersey’, ‘Massachusetts’,... có số ca nhiễm và tử vong cao hơn so với các bang nằm ở phía tây và trung tâm như ‘Washington’, ‘Nevada’, ‘New Mexico’. Ngoại trừ một số bang lớn ở phía tây và trung tâm như ‘California’ và ‘Texas’ thì vẫn bị ảnh hưởng nặng bởi ‘COVID-19’. Điều này có thể là do các bang ở phía đông nước Mỹ có mật độ dân số cao hơn so với các bang ở phía tây, từ đó dẫn đến việc lây nhiễm dễ dàng hơn. Ngoài ra, nếu quan sát bản đồ các bang ở nước Mỹ ở bên dưới, ta dễ dàng thấy được khu vực phía tây nước Mỹ có rất nhiều bang có kích thước nhỏ và nằm chen chúc với nhau, từ đó dẫn đến việc lây nhiễm dễ dàng hơn so với khu vực phía đông, nơi mà có diện tích rộng rãi hơn và vị trí các bang cũng cách xa nhau, thuận lợi cho việc cách ly và kiểm soát dịch bệnh.
- Điều thú vị cuối cùng là 2 quần đảo lớn của Mỹ là ‘Alaska’ và ‘Hawaii’ lại bất ngờ bị ảnh hưởng rất ít bởi dịch bệnh ‘COVID-19’. Điều này có thể vị trí địa lý của 2 quần đảo này nằm cách biệt với các bang khác ở đất liền, từ đó dễ dàng kiểm soát dịch bệnh hơn. Cụ thể theo cuộc phỏng vấn một bác sĩ ở ‘Alaska’ được thực hiện bởi tờ báo địa phương Alaska [7], lí do chính giúp ‘Alaska’ ít bị ảnh hưởng bởi COVID-19 là:
 - Đây là khu vực tương đối xa xôi, thời tiết khắc nghiệt, không tập trung dân cư nhiều, từ đó dễ dàng kiểm soát dịch bệnh hơn.
 - Alaska có nhiều thời gian hơn để chuẩn bị và tích trữ các xét nghiệm, nguồn cung cấp và điều trị như ‘Remdesivir’ trước khi số ca nhiễm bùng phát.

- Sự cách ly địa lý của Alaska cho phép nó siết chặt biên giới dễ dàng hơn so với các tiểu bang khác. Các biện pháp hạn chế du lịch quyết liệt ngay từ đầu đã giúp hạn chế sự lây lan của virus.
- Alaska không có các thành phố và không gian công cộng đông đúc như nhiều tiểu bang khác, điều này làm giảm sự lây truyền.



Hình 4. Bản đồ các bang ở Mỹ

4 - Phân tích yếu tố sắc tộc ảnh hưởng như nào đến tình hình COVID-19 ở Mỹ

4.1 Đặt giả thuyết:

Trong đại dịch COVID-19 ở những năm 2020 và 2021 ở Hoa Kỳ ta có thể thấy những bài báo uy tín đã được công bố trên internet với tựa đề như là "THE COLOR OF CORONAVIRUS: COVID-19 DEATHS BY RACE AND ETHNICITY IN THE U.S." được đăng trên APM research lab và bài báo "Study: Covid's racial disparities made some white people less vigilant about the virus" ở trang NBC news. Cả 2 bài báo này đều cho biết thông tin của các nhóm sắc tộc trong đại dịch COVID-19 và đặc biệt đó không phải là vấn đề của nhóm người da trắng. Ngoài ra, theo kết quả nghiên cứu cho thấy người da trắng chiếm 76% tổng dân số Hoa Kỳ (USA) và vùng lãnh thổ dựa trên dữ liệu điều tra dân số Hoa Kỳ năm 2020 và nạn phân biệt chủng tộc ở Hoa Kỳ vẫn xuất hiện rất nhiều ở thời điểm đó. Từ đó, ta đặt ra 2 giả thuyết rằng:

- Liệu nhóm người Mỹ gốc Phi và những người da màu khác (gồm chung là nhóm POCs) có số ca dương tính và tỉ lệ tử vong cao hơn người da trắng (White) ở USA hay không?

- Liệu có sự khác biệt đáng kể về số ca nhiễm xác nhận và tỷ lệ tử vong giữa các nhóm chủng tộc trên khắp Hoa Kỳ và lãnh thổ hay không?

4.2 Phương pháp tiền xử lý dữ liệu

- Trong tập dữ liệu **CRDT_DATA.csv** trên sẽ có các nhóm sắc tộc như là Black, Latinx, Asian, AIAN, NHPI, Multiracial và White. Vì vậy chúng tôi sẽ tiền xử lý dữ liệu bằng cách lọc rỗng và tính tổng. Cuối cùng, chúng tôi thu được 2 cột tổng hợp là POCs và nhóm da trắng phục vụ cho giả thuyết thứ nhất. Còn giả thuyết thứ hai, chúng tôi sẽ xét riêng từng sắc tộc. Trong đó:

- **POCs_Cases**: Tổng số ca nhiễm của các nhóm người da màu (Black, Latinx, Asian, AIAN, NHPI, Multiracial)..
- **White_Cases**: Số ca nhiễm của nhóm da trắng.
- **POCs_Deaths**: Tổng số ca tử vong của các nhóm người da màu (Black, Latinx, Asian, AIAN, NHPI, Multiracial).
- **White_Deaths**: Số ca tử vong của nhóm da trắng.

4.3 Mô tả tập dữ liệu cho giả thuyết thứ nhất:

- Dùng phương pháp mô tả dữ liệu cho trường hợp số ca nhiễm dương tính, chúng tôi thu được bảng sau:

	POCs_Cases	White_Cases
count	5230	5230
mean	46062	61801
std	126137	101531
min	0	0
25%	1363	2857
50%	11198	19007
75%	43829	69537
max	1831250	664121

Bảng 1. Mô tả dữ liệu cho trường hợp số ca nhiễm dương tính của POCs và nhóm da trắng

- Ta có thể thấy ở bảng 1 cả 2 nhóm đều có cùng số lượng mẫu là 5230 cho thấy đảm bảo tính cân bằng dữ liệu. Mặt khác, nhóm người da trắng có số ca nhiễm trung bình cao hơn đáng kể so với nhóm POCs. Tuy nhiên độ lệch chuẩn của nhóm POCs cao hơn rất nhiều cho thấy chứa nhiều giá trị nhiễu. Điều đó cho thấy rằng số ca nhiễm trong nhóm này phân tán rộng và có sự khác biệt lớn giữa các giá trị. Đặc biệt là giá trị lớn nhất (max) của nhóm POCs rất cao so với nhóm da trắng, chỉ ra rằng có thể có một số lượng nhỏ khu vực hoặc thời gian có số ca nhiễm rất cao trong nhóm POCs. Mặt khác, các giá trị ở các phân vị (25%, 50%, 75%) của nhóm POCs đều thấp hơn so với nhóm da trắng, cho thấy rằng ở mọi mức độ phân phối đều và rộng, nhóm da trắng đều có số ca nhiễm cao hơn và dày đặc hơn. Qua đó, ta có thể thấy rằng số ca nhiễm ở nhóm da trắng có xu hướng cao hơn và phân tán rộng hơn so với nhóm POCs. Điều này có thể phản ánh số ca dương tính của nhóm POCs thấp hơn người da trắng.

- Dùng phương pháp mô tả dữ liệu cho trường hợp số ca tử vong, chúng tôi thu được bảng sau:

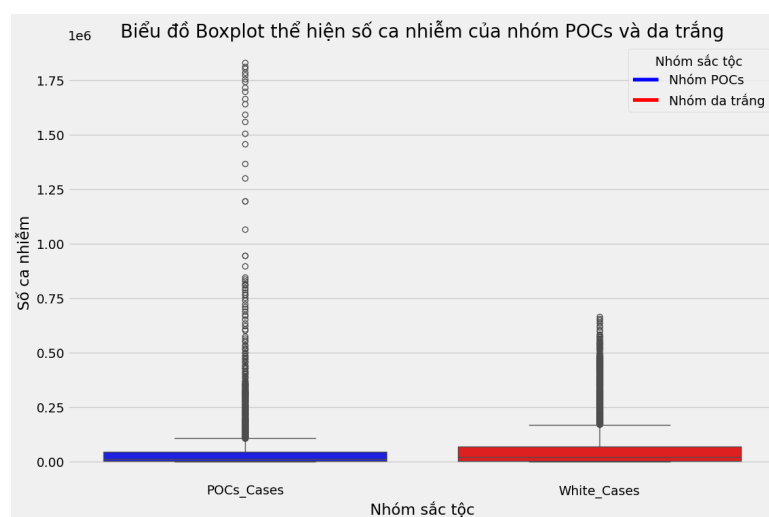
	POCs_Cases	White_Cases
count	5320	5320
mean	1525	2080
std	3463	3090
min	0	0
25%	26	90
50%	286	766
75%	1399	2832
max	34288	19497

Bảng 2. Mô tả dữ liệu cho trường hợp số ca tử vong của POCs và nhóm da trắng

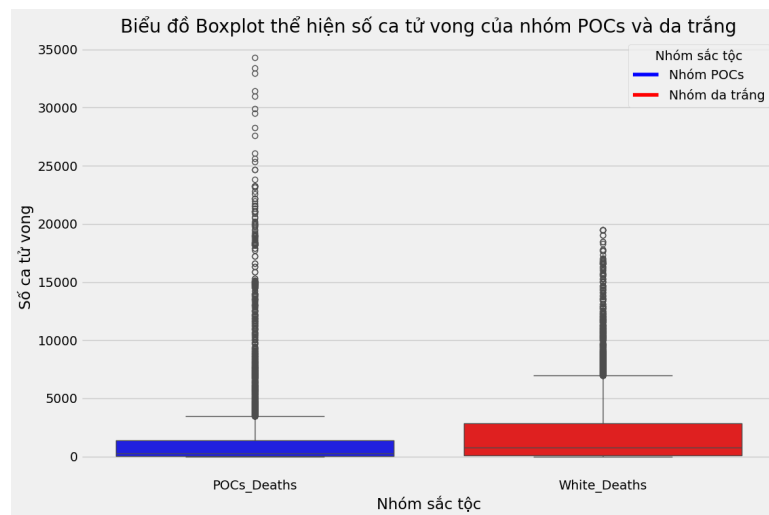
- Ta có thể thấy ở bảng 2 cả 2 nhóm đều có cùng số lượng mẫu là 5320 cho thấy đảm bảo tính cân bằng dữ liệu. Mặt khác, nhóm người da màu (POCs) có số ca tử vong trung bình thấp hơn so với nhóm người da trắng (White) nhưng độ lệch chuẩn của nhóm POCs cao hơn nhưng không đáng kể. Điều đó cho thấy rằng số ca nhiễm trong nhóm POCs này phân tán rộng và có sự khác biệt lớn giữa các giá trị. Đặc biệt là giá trị lớn nhất (max) của nhóm POCs rất cao so với nhóm da trắng, chỉ ra rằng có thể có một số lượng nhỏ khu vực hoặc thời gian có số ca tử vong rất cao trong nhóm POCs. Tuy nhiên, các giá trị ở các phân vị (25%, 50%, 75%) của nhóm da trắng đều cao hơn so với nhóm POCs. Điều này có thể chỉ ra rằng mức độ phân phối của số ca tử vong trong nhóm da trắng có thể dày hơn ở các giá trị trung bình đến cao. Điều này có thể phản ánh tỉ lệ tử vong cao của nhóm POCs thấp hơn người da trắng.

4.4 Trực quan hình ảnh để làm rõ giả thuyết thứ nhất:

4.4.1 Sự phân bố dữ liệu về số ca nhiễm dương tính và số ca tử vong của 2 nhóm sắc tộc POCs và White:



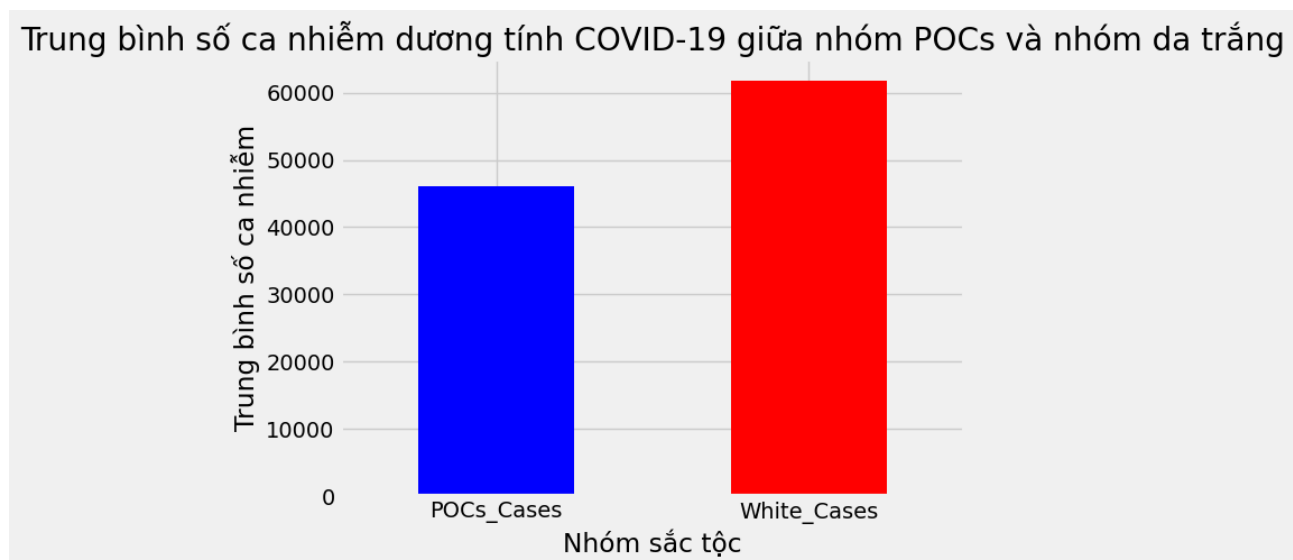
Hình 5. Biểu đồ boxplot thể hiện số ca nhiễm của nhóm POCs và nhóm da trắng



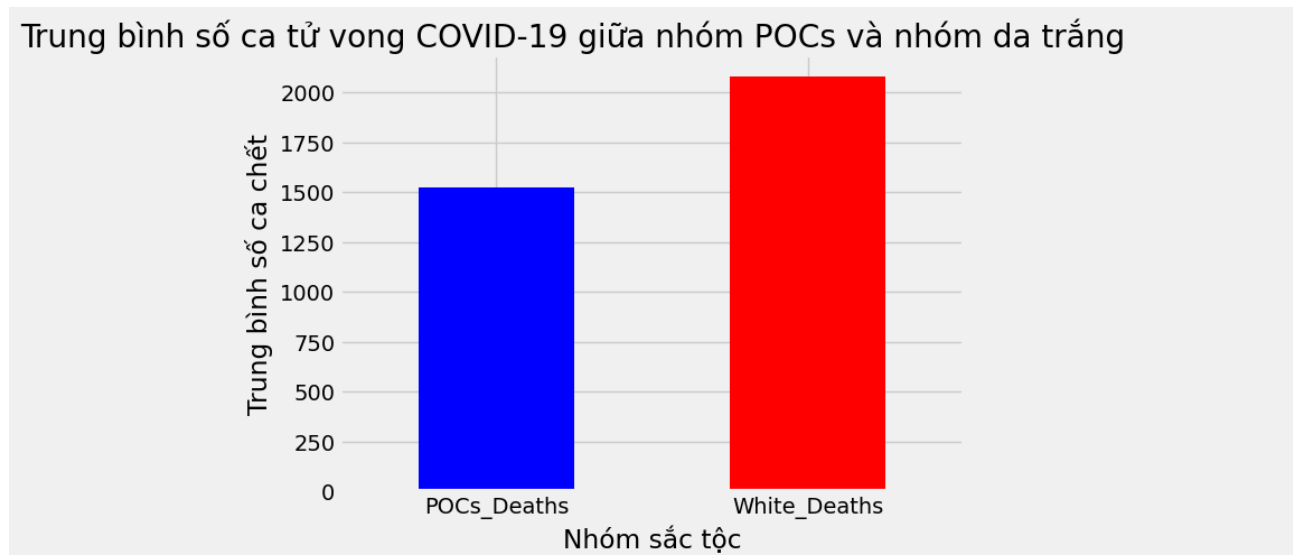
Hình 6. Biểu đồ boxplot thể hiện số ca tử vong của nhóm POCs và nhóm da trắng

- Với biểu đồ boxplot ở hình 5 và 6 ta giúp ta hình dung rõ hơn rằng số ca nhiễm và số ca tử vong ở các phân vị (25%, 50%, 75%) của nhóm da trắng đều cao hơn so với nhóm POCs nhưng các giá trị ngoại lai và giá trị lớn nhất của nhóm POCs rất nhiều và cao hơn so với nhóm da trắng nhưng chiếm rất ít, chỉ ra rằng có thể có một số lượng nhỏ khu vực hoặc thời gian có số ca tử vong rất cao trong nhóm POCs. Điều này có thể phản ánh số ca nhiễm và tỉ lệ tử vong cao của nhóm POCs thấp hơn nhóm da trắng.

4.4.2 Số lượng trung bình về số ca nhiễm dương tính và số ca tử vong của 2 nhóm sắc tộc POCs và White



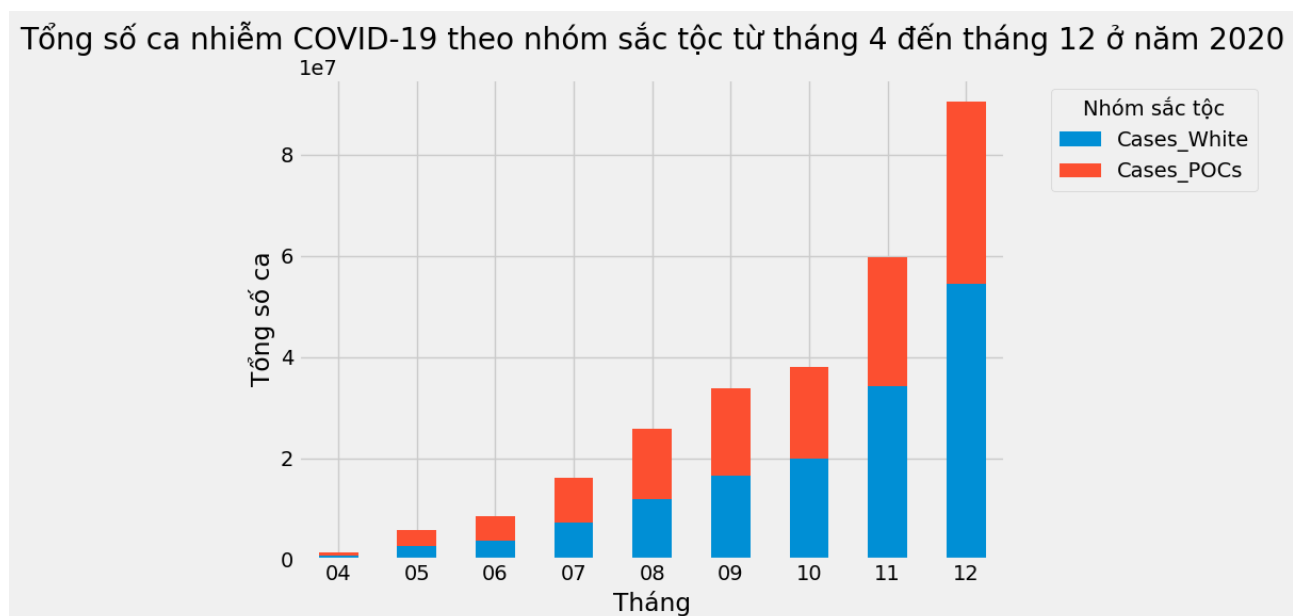
Hình 7. Biểu đồ cột thể hiện số ca nhiễm trung bình của nhóm POCs và nhóm da trắng



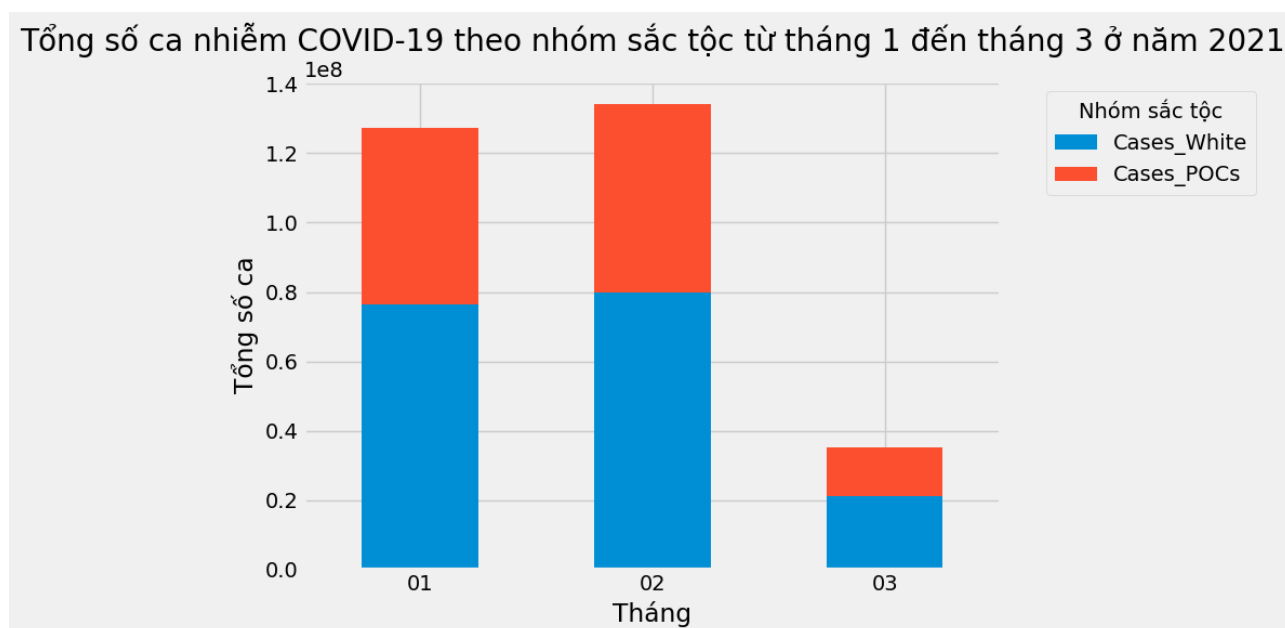
Hình 8. Biểu đồ cột thể hiện số ca tử vong trung bình của nhóm POCs và nhóm da trắng

- Với biểu đồ cột ở hình 7 và 8 giúp ta hình dung rõ hơn rằng số ca nhiễm và số ca tử vong trung bình của nhóm POCs thấp hơn nhóm người da trắng. Trong đó nhóm POCs luôn thấp hơn về mọi mặt. Điều này có thể phản ánh số ca nhiễm và tỉ lệ tử vong của nhóm POCs thấp hơn người da trắng.

4.4.3 Sự phân bố số ca nhiễm dương tính và số ca tử vong qua các tháng trong 2 năm 2020 và 2021

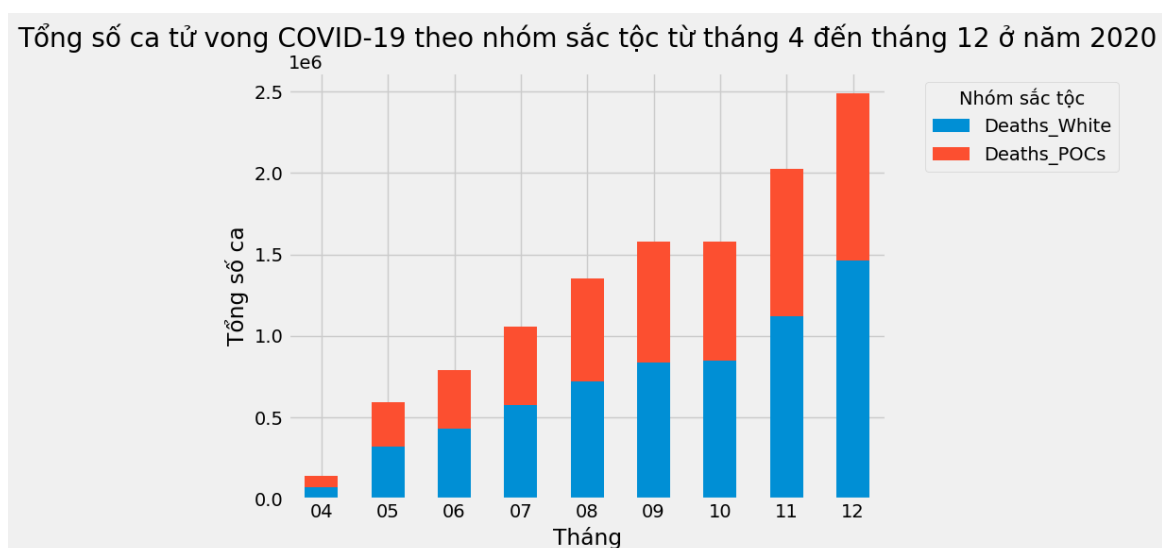


Hình 9. Biểu đồ cột chồng thể hiện sự phân bố số ca nhiễm của nhóm POCs và nhóm da trắng qua các tháng trong năm 2020

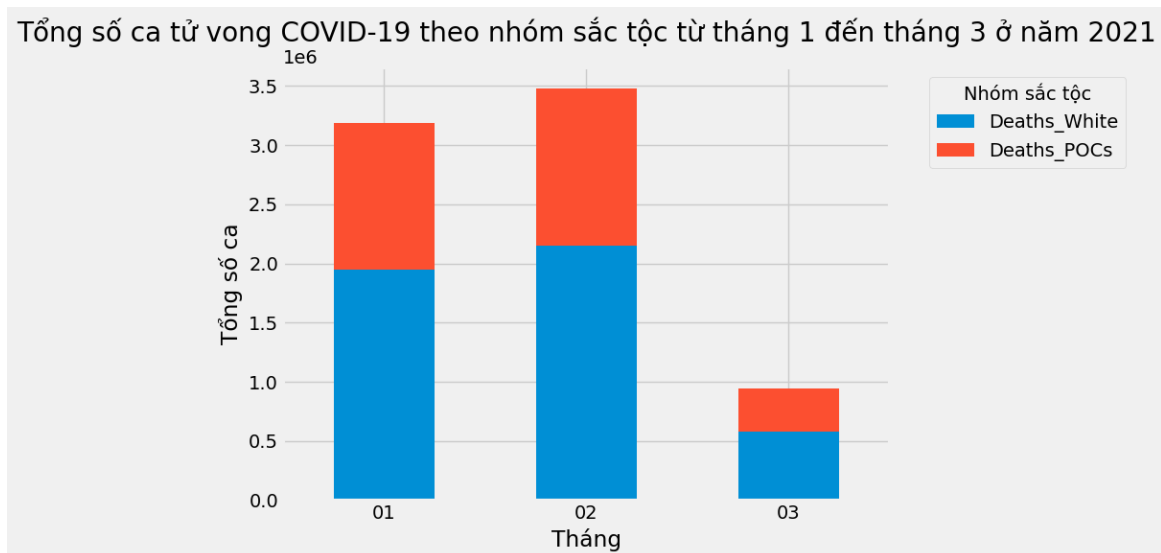


Hình 10. Biểu đồ cột chồng thể hiện sự phân bố số ca nhiễm của nhóm POCs và nhóm da trắng qua các tháng trong năm 2021

- Với 2 biểu đồ cột chồng ở hình 9 và 10 giúp ta hình dung rõ hơn rằng trong năm 2020 đại dịch COVID-19 ghi nhận từ tháng 4 đến tháng 12 và trong năm 2021 đại dịch COVID-19 chỉ ghi nhận trong 3 tháng đầu. Mặc dù từ tháng 4 đến tháng 7 trong năm 2020 các nhóm POCs có số ca nhiễm cao hơn nhóm da trắng nhưng không quá nhiều. Đặc biệt là số ca nhiễm từ tháng 8 năm 2020 trở đi số ca nhiễm của nhóm da trắng luôn cao hơn nhóm POCs. Điều này có thể phản ánh số ca nhiễm của nhóm POCs thấp hơn người da trắng.



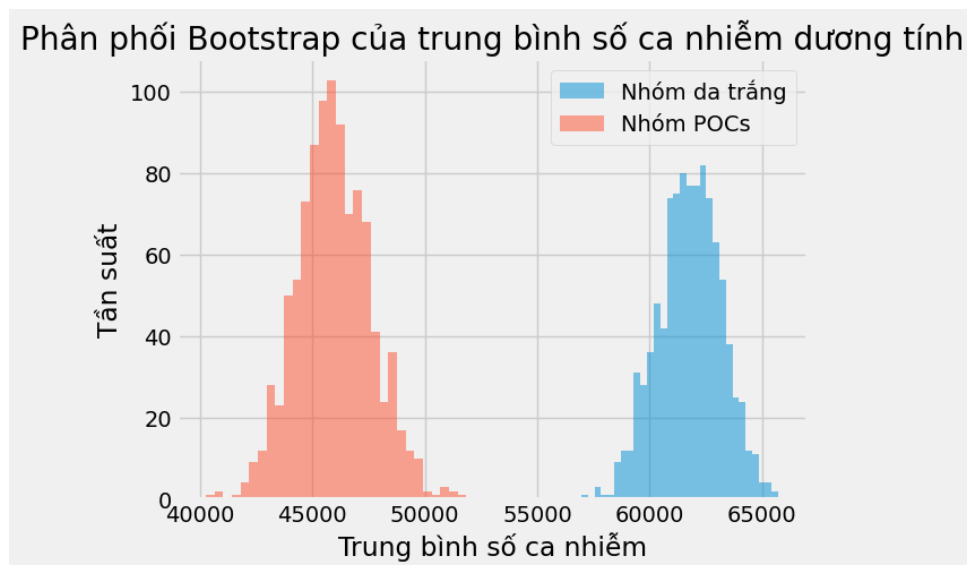
Hình 11. Biểu đồ cột chồng thể hiện sự phân bố số ca tử vong của nhóm POCs và nhóm da trắng qua các tháng trong năm 2020



Hình 12. Biểu đồ cột chồng thể hiện sự phân bố số ca tử vong của nhóm POCs và nhóm da trắng qua các tháng trong năm 2021

- Tương tự với 2 biểu đồ cột chồng ở hình 11 và 12 giúp ta hình dung rõ hơn rằng trong năm 2020 đại dịch COVID-19 cũng được ghi nhận từ tháng 4 đến tháng 12 và trong năm 2021 đại dịch COVID-19 chỉ ghi nhận trong 3 tháng đầu và số ca tử vong của các tháng trong 2 năm đó của nhóm da trắng luôn cao hơn nhóm POCs. Điều này có thể phản ánh tỉ lệ tử vong của nhóm POCs thấp hơn người da trắng.

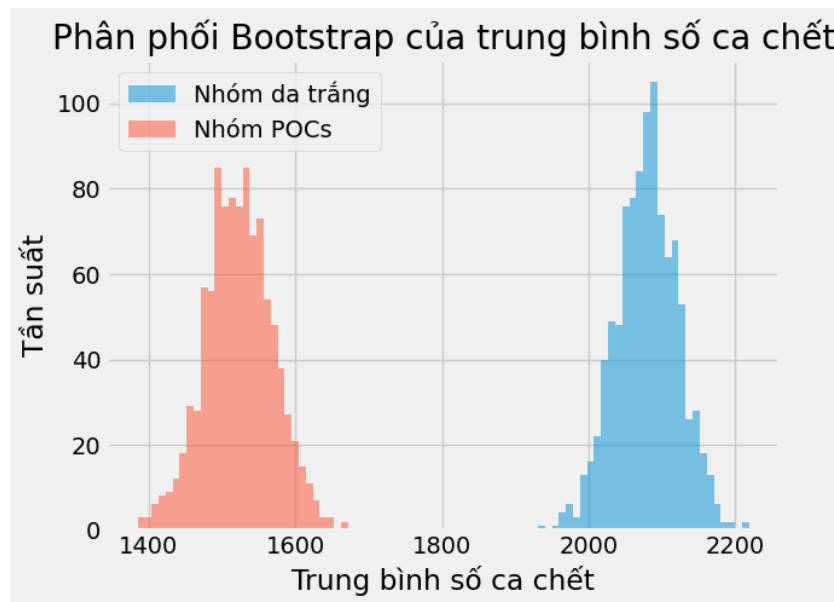
4.5 Mô phỏng tập dữ liệu trên qua phương thức bootstrap để làm rõ giả thuyết thứ nhất



Hình 13. Biểu đồ phân phối bootstrap thể hiện sự phân bố số ca nhiễm trung bình của nhóm POCs và nhóm da trắng

- Hiệu Trung bình: -15829.821037781956
- Xác suất nhóm POCS có số ca nhiễm cao hơn nhóm da trắng: 0.0

- Với biểu đồ phân phối ở hình 13 ta thấy được có sự tách biệt hoàn toàn giữa 2 khối và hiệu trung bình giữa số ca nhiễm trung bình của nhóm POCs và nhóm da trắng là một khoảng rất lớn khoảng 15830 ca. Giá trị p được tính là 0.0, điều này có nghĩa là xác suất số ca nhiễm của nhóm POCs cao hơn nhóm da trắng trong số ca nhiễm COVID-19 là không thể.



Hình 14. Biểu đồ phân phối bootstrap thể hiện sự phân bố số ca tử vong trung bình của nhóm POCs và nhóm da trắng

- Hiệu Trung bình: -556.6058295112778
- Xác suất nhóm POCS có số ca nhiễm cao hơn nhóm da trắng: 0.0

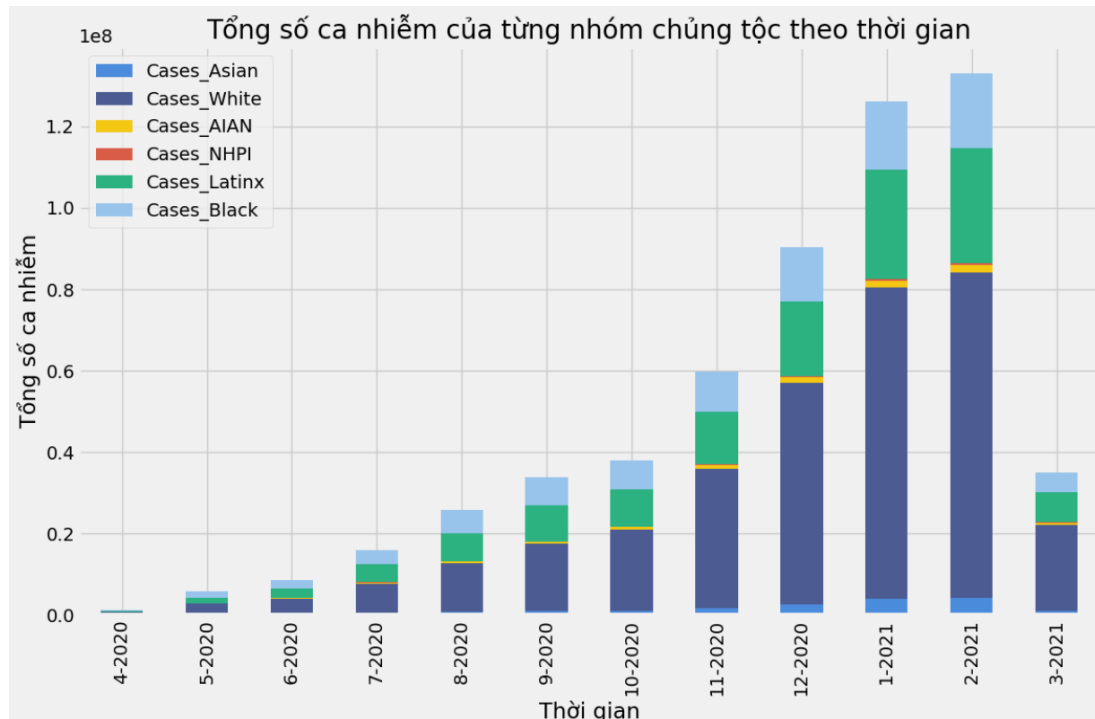
- Với biểu đồ phân phối ở hình 14 ta thấy có sự tách biệt giữa 2 khối và hiệu trung bình giữa số ca tử vong trung bình của nhóm POCs và nhóm da trắng là một khoảng 557 ca. Hơn nữa, giá trị xác suất được tính là 0.0, điều này có nghĩa là xác suất tỉ lệ tử vong của nhóm POCs cao hơn nhóm da trắng trong số ca tử vong COVID-19 là không thể.

4.6 Kết luận giả thuyết thứ nhất

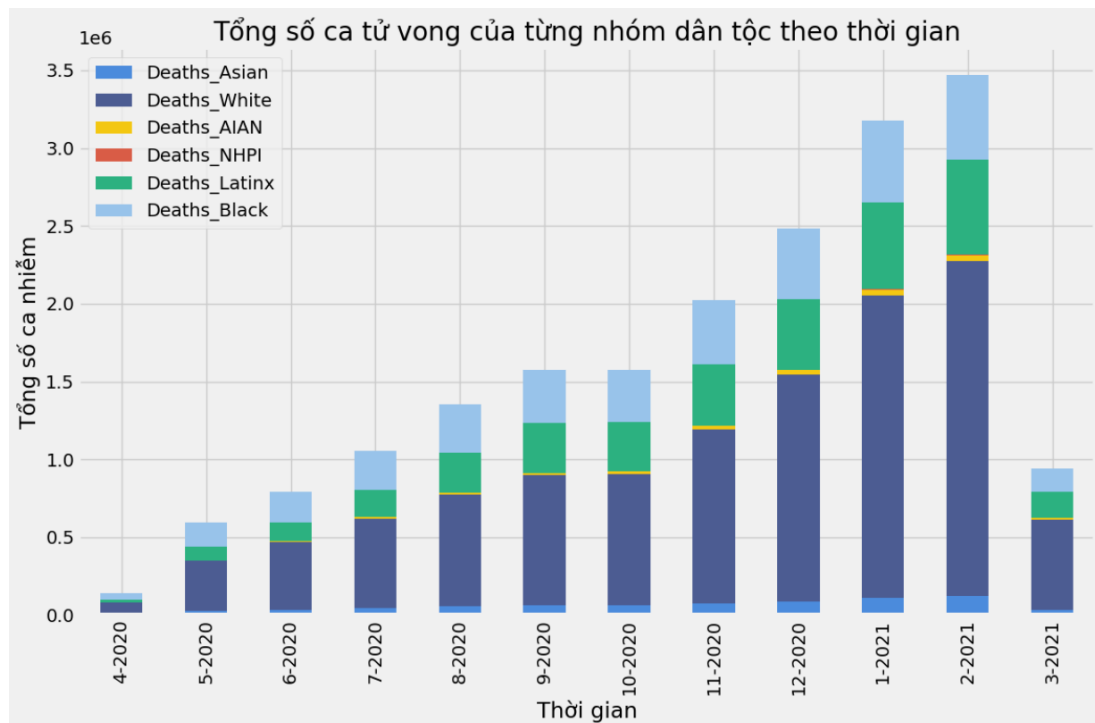
- Dựa theo những phân tích trên ta kết luận được rằng nhóm người Mỹ gốc Phi và những người da màu khác (POCs) có số ca dương tính và tỉ lệ tử vong cao hơn người da trắng ở USA là một giả thuyết sai. Thêm vào đó, để chứng minh và làm rõ hơn phân tích trên là đúng ta có thể thông qua các bài báo uy tín đã được công bố trên internet với kết quả tương tự (bài báo "THE COLOR OF CORONAVIRUS: COVID-19 DEATHS BY RACE AND ETHNICITY IN THE U.S." được đăng trên APM research lab) [5]. Ngoài ra theo bài báo Study: Covid's racial disparities made some white people less vigilant about the virus ở trang NBC news [3] cho thấy những người da trắng được khảo sát ở Hoa Kỳ vào mùa thu năm 2020 ít quan tâm hơn và thậm chí còn có nhiều khả năng trốn tránh các biện pháp phòng ngừa an toàn cho đại dịch sau khi biết về những cách không cân xứng mà nó tác động đến cộng đồng Da đen và các cộng đồng da màu khác vì họ tin rằng đó "không phải là vấn đề của người da trắng" chính vì điều đó cũng có khả năng gây ra việc số ca dương tính và tỉ lệ tử vong của người da trắng cao hơn nhóm POCs ở USA.

4.7 Trực quan hình ảnh để làm rõ giả thuyết thứ hai

4.7.1 Sự phân bố dữ liệu về số ca nhiễm và tử vong của từng nhóm dân tộc theo thời gian



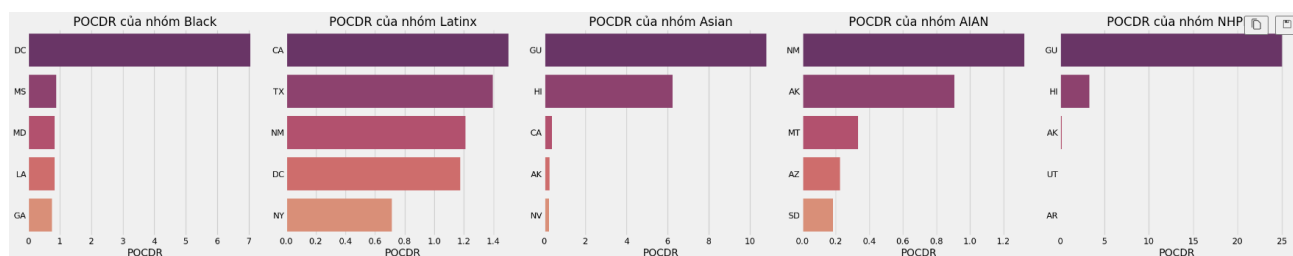
Hình 15. Biểu đồ cột chồng thể hiện tổng số ca nhiễm của từng nhóm chủng tộc theo thời gian



Hình 16. Biểu đồ cột chồng thể hiện tổng số ca tử vong của từng nhóm chủng tộc theo thời gian

- Với biểu đồ ở hình 15 và 16 ta có thể thấy trong xuyên suốt khoảng thời gian 1 năm từ tháng 4/2020 tới tháng 3/2021, nhóm người da trắng luôn luôn chiếm từ 50-60% tổng số ca nhiễm dương tính với COVID-19. Trong khi đó thì nhóm người Latinh và người da đen lần lượt chiếm đứng thứ 2 và 3 trên tổng số ca nhiễm dương tính. Cuối cùng là nhóm người AIAN, người Châu Á và NHPI chiếm trọng số rất ít cùng nhau. Tương tự như số ca nhiễm, người da trắng cũng chiếm tỷ lệ cao nhất về số ca tử vong với COVID-19, chiếm khoảng 50-60% tổng số ca tử vong trong xuyên suốt 1 năm. Và người Latinh và người da đen lần lượt chiếm vị trí thứ 2 và 3 về số ca tử vong, trong khi AIAN, người Châu Á và NHPI chiếm tỷ lệ rất ít cùng nhau.

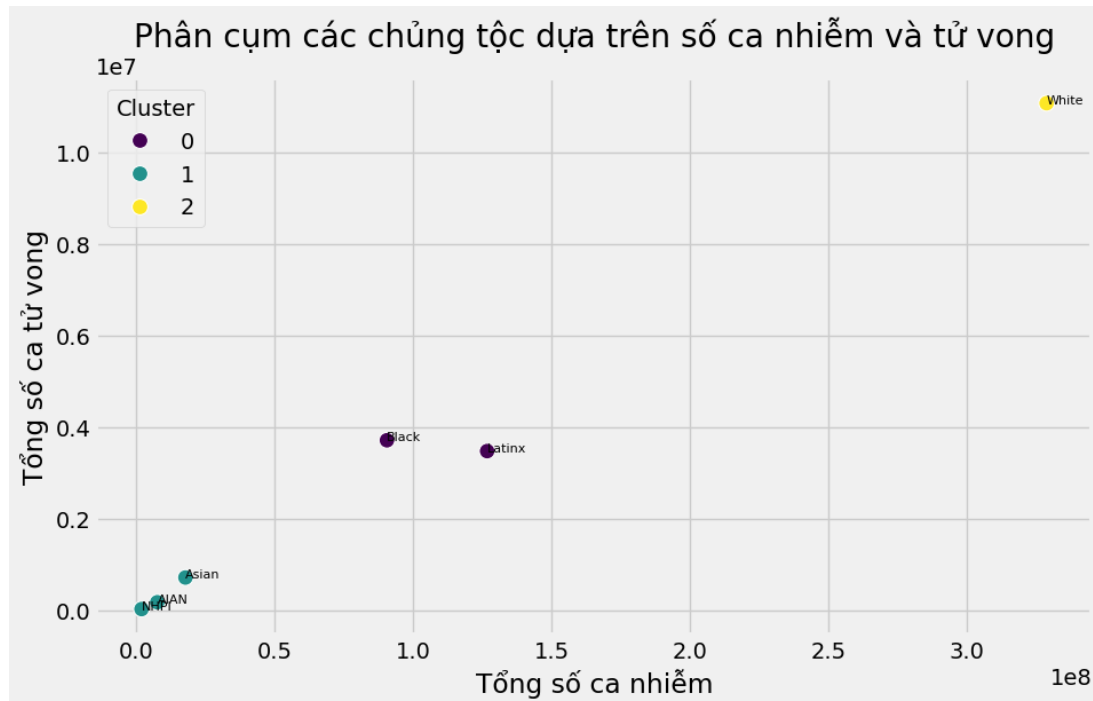
4.7.2 Phân tích chỉ số People of Color Death Rate (POCDR) cho các nhóm sắc tộc ở các bang khác nhau:



Hình 17. Biểu đồ bar thể hiện chỉ số POCDR của các nhóm sắc tộc ở các bang khác nhau

- Dựa theo biểu đồ ở hình 17 cho ta thấy nhóm người da màu (Black) có chỉ số POCDR cao đột biến ở bang DC với giá trị lên tới 7, điều này có nghĩa là tỷ lệ tử vong của người da màu ở bang DC cao hơn 7 lần so với người da trắng ở bang DC. Trong khi đó ở các bang khác của Mỹ thì chỉ số POCDR của nhóm người da màu (Black) luôn ở mức thấp hơn 1, tức là tỷ lệ tử vong của nhóm người da màu (Black) luôn thấp hơn so với người da trắng. Mặt khác, nhóm người (Latinx) có chỉ số POCDR tương đối khiêm tốn khi chỉ số cao nhất của nhóm này là ở bang CA với giá trị khoảng 1.5. Chỉ có duy nhất 4 bang là CA, TX, NM, DC và NY là có chỉ số POCDR của nhóm người Latinx cao hơn 1, còn lại đều thấp hơn 1. Tương tự như nhóm người Latinx, chỉ số POCDR của nhóm AIAN tương đối thấp, chỉ có duy nhất bang NM là có chỉ số POCDR của nhóm người AIAN lớn hơn 1. Cuối cùng là có một sự thú vị trùng hợp ở chỉ số POCDR của nhóm người Asian và NHPI khi cả 2 nhóm này đều có chỉ số POCDR lớn nhất ở 2 bang là GU và HI, và các chỉ số này đều cực kỳ cao, cụ thể chỉ số POCDR của nhóm 'Asian' ở bang GU là hơn 10 còn của nhóm NHPI là đến tận 25. Điều này cho thấy rằng tỷ lệ tử vong của nhóm người Asian và NHPI lớn hơn rất nhiều so với người da trắng ở 2 bang này. Ngoài 2 bang này ra thì chỉ số POCDR của nhóm Asian và NHPI đều thấp hơn 1 ở các bang còn lại.

4.8 Áp dụng phương pháp gom nhóm các chủng tộc theo tổng số ca nhiễm và tử vong làm rõ giả thuyết thứ hai



Hình 18. Biểu đồ phân cụm các chủng tộc dựa theo số ca nhiễm và tử vong

- Với biểu đồ gom nhóm ở hình 18 bằng việc sử dụng chỉ số POCDR và phương pháp phân cụm bằng K-mean dựa trên số ca nhiễm và tử vong, ta có thể thấy được sự tương đồng lớn giữa Latinx và Black với nhau, nhóm còn lại là bao gồm 3 chủng tộc: Asian, AIAN và NHPI. Với điều này, tính đúng đắn của giả thuyết: ‘Sự khác biệt về số ca nhiễm xác nhận và tỷ lệ tử vong giữa các nhóm chủng tộc trên khắp Hoa Kỳ là đáng kể’ càng được củng cố chắc chắn hơn.

4.9 Kết luận giả thuyết thứ hai

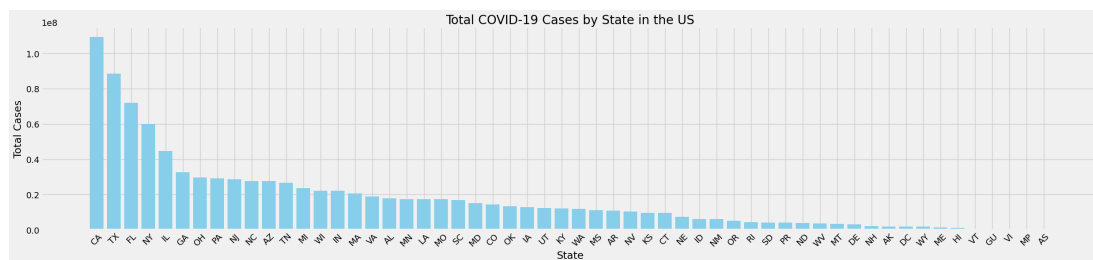
Dựa theo những phân tích trên ta kết luận được rằng có sự khác biệt đáng kể về số ca nhiễm xác nhận và tỷ lệ tử vong giữa các nhóm chủng tộc trên khắp Hoa Kỳ và lãnh thổ và điều đó cũng khẳng định giả thuyết trên là đúng. Ngoài ra từ những phân tích trên ta có thể thấy rằng có sự chênh lệch rất lớn về số ca nhiễm và tử vong giữa các nhóm dân tộc ở Mỹ. Nhóm người da trắng White luôn chiếm tỷ lệ cao nhất và vượt trội hoàn toàn về số ca nhiễm và tử vong so với các chủng tộc POCs.

5 - Phân tích sự ảnh hưởng của các chính sách xã hội của các bang đến tình hình COVID-19 tại Mỹ

- Trong đại dịch COVID-19 được ghi nhận trong hai năm 2020 và 2021 tại Mỹ rất nhiều các chính sách được chính phủ Mỹ ban hành. Và tác động của các yếu tố chính trị và chính

sách đối với số ca bệnh và tử vong của đất nước được thảo luận tiếp theo bằng cách đánh giá sự phân phối của số ca nhiễm và tử vong qua các chính sách được ban hành tại thời điểm đó. Từ đó chúng ta có giả thuyết rằng **liệu sự kiểm soát của đảng chính trị và các chính sách chính phủ có ảnh hưởng đến tỷ lệ ca nhiễm và tử vong giữa các nhóm chủng tộc ở Mỹ hay không?**

5.1 Mô tả dữ liệu và lấy ra bang có số lượng ca nhiễm cao nhất để kiểm định giả thuyết



Hình 19. Tổng số ca nhiễm theo từng bang

Nhận xét tổng thể:

- Dựa vào biểu đồ trên, ta thấy rằng số ca nhiễm của bang California (CA) là vượt trội rất lớn so với những bang còn lại.
- Dựa theo bảng dữ liệu chính sách ở trên ta thấy rằng mỗi bang ở đại dịch COVID-19 ở Mỹ sẽ có thời gian ban hành các chính sách khác nhau.

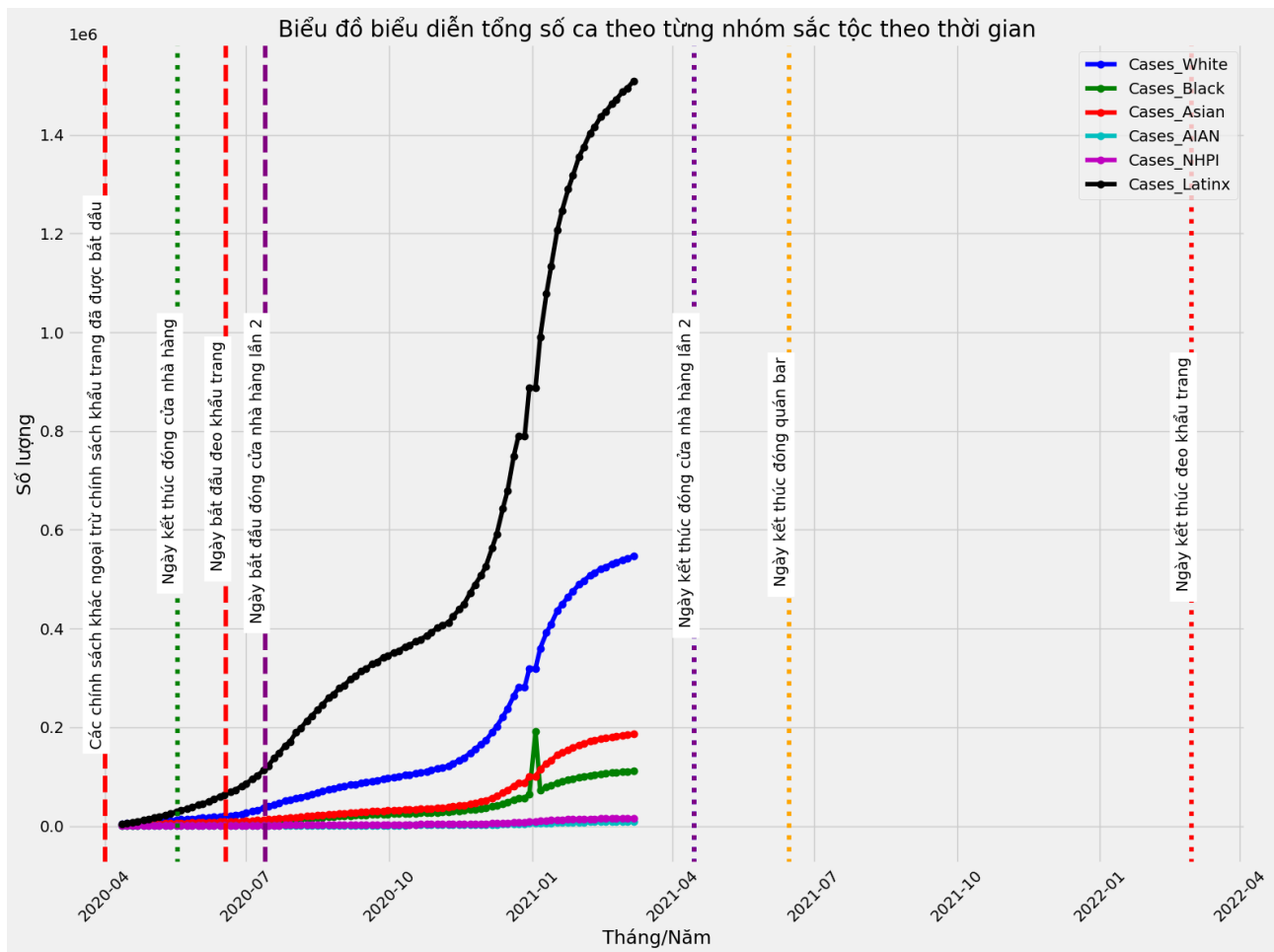
⇒ Từ đó ta có thể dựa vào các chính sách từ bang CA này để chứng minh giả thuyết trên.

5.2 Tiền xử lý dữ liệu

Nhận xét tổng thể:

- Qua quá trình xử lý dữ liệu ta thu được 2 biến dữ liệu chính là `ca_policy_data` và `ca_covid_data`.
- Biến `ca_policy_data` sẽ bao gồm thời gian của 4 chính sách mà chúng tôi rút trích ra là chính sách đeo khẩu trang, chính sách đóng cửa nhà hàng, quán bar và rạp chiếu phim của bang CA.
- Biến `ca_covid_data` sẽ lưu trữ số ca nhiễm và tử vong của các nhóm sắc tộc như là White, Black, Latinx, Asian AIAN, NHPI của bang CA theo từng thời gian. Đặc biệt, có 5 cột mới mang giá trị nhị phân là 0 và 1 được tạo ra là 'mask', 'restaurant', 'restaurant2', 'bar', 'movie' đại diện cho các chính sách nào được ban hành tại date tương ứng.
- Chính sách đeo khẩu trang sẽ bắt đầu từ ngày 18/06/2020 đến ngày 01/03/2022
- Chính sách đóng cửa nhà hàng sẽ bắt đầu từ ngày 16/03/2020 đến ngày 18/05/202
- Chính sách đóng cửa nhà hàng lần 2 sẽ bắt đầu từ ngày 13/07/2020 đến ngày 15/04/202
- Chính sách đóng cửa quán bar sẽ bắt đầu từ ngày 16/03/2020 đến ngày 15/06/202
- Chính sách đóng cửa rạp chiếu phim sẽ bắt đầu từ ngày 19/03/2020 đến ngày 15/04/2021

5.3 Trực quan hóa dữ liệu

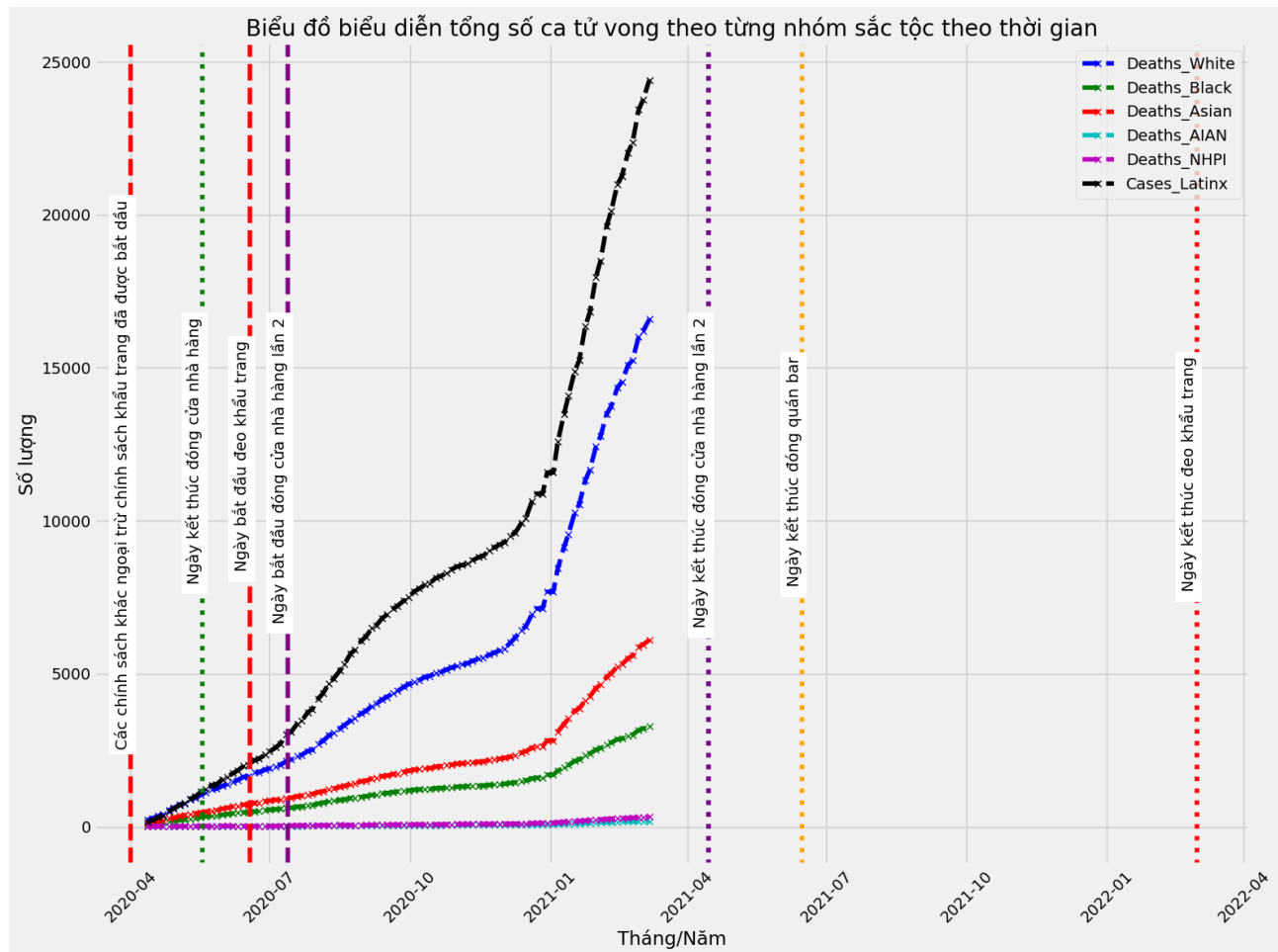


Hình 20. Biểu đồ biểu diễn tổng số ca nhiễm theo từng nhóm sắc tộc theo thời gian

- Nhận xét:

- Dựa theo biểu đồ trên, ta thấy rằng các chính sách được ban hành rất nhiều và có 3 chính sách được ban hành trước thời điểm dữ liệu COVID-19 mà chúng em đang xét là chính sách đóng cửa nhà hàng, quán bar và rạp chiếu phim. Nhưng gần như vẫn không thể kiểm soát nổi được dịch bệnh bùng phát mạnh trong năm 2020 và 2021 tại bang CA của Mỹ.
- Đặc biệt, ta có thể dễ dàng thấy rằng ngay thời điểm chính sách đóng cửa nhà hàng kết thúc cũng chính là thời điểm mà dịch bệnh bắt đầu bùng phát rất mạnh và nhanh. Điều đó cũng cho thấy chính sách đóng cửa nhà hàng cần được ban hành tiếp tục tại thời điểm đó. Vì vậy vào khoảng tháng 7 năm 2020 chính sách đóng cửa nhà hàng lần 2 gấp rút được công bố và áp dụng mãi đến năm 2021.
- Sau lệnh mở cửa nhà hàng thì dịch bùng phát mạnh mẽ và lệnh đeo khẩu trang được ban hành rất nhanh và chính sách này rất hữu dụng và cần thiết mãi được thi hành đến tháng 3 năm 2022.
- Hai chính sách còn lại là đóng cửa rạp chiếu phim và quán bar được thực hiện rất sớm và thời gian kết thúc chính sách này nằm sau khoảng thời gian dữ liệu COVID-19 được ghi nhận ở đây. Điều đó, cũng cho thấy rằng 2 chính sách này cũng rất cần thiết và cần được thi hành lâu dài.

- Trong biểu đồ ta thấy rằng nhóm sắc tộc bùng phát mạnh nhất là Latinx, tiếp theo là White, tiếp đến là Asian, Black, NHPI và cuối cùng là AIAN.



Hình 21. Biểu đồ biểu diễn tổng số ca tử vong theo từng nhóm sắc tộc theo thời gian

- Nhận xét:

- Ta thấy rằng biểu đồ tử vong ở đây có hình dạng như biểu đồ số ca nhiễm ở trên. Điều đó có thể khẳng định dịch bùng phát rất nhanh và các chính sách cũng không thể kiểm soát nổi tại thời điểm đó. Những người bị nhiễm bệnh tại thời điểm đó có nguy cơ chết rất cao.

5.4 Áp dụng phương pháp hồi quy tuyến tính

OLS Regression Results

Dep. Variable:

Cases_Total

R-squared:

0.338

Model:

OLS

Adj. R-squared:

0.316

Method:

Least Squares

F-statistic:

15.45

Date:

Fri, 14 Jun 2024

Prob (F-statistic):

3.33e-08

Time:

23:12:25

Log-Likelihood:

-1438.3

No. Observations:

95

AIC:

2885.

Df Residuals:

91

BIC:

2895.

Df Model:

3

Covariance Type:

nonrobust

coef

std err

t

P>|t|

[0.025

0.975]

mask

1.203e+05

4.69e+05

0.257

0.798

-8.11e+05

1.05e+06

restaurant

-7.14e+04

4.18e+05

-0.171

0.865

-9.02e+05

7.59e+05

restaurant2

1.319e+06

3.69e+05

3.574

0.001

5.86e+05

2.05e+06

bar

5.972e+04

1.55e+05

0.385

0.701

-2.48e+05

3.68e+05

movie

5.972e+04

1.55e+05

0.385

0.701

-2.48e+05

3.68e+05

Omnibus:

11.363

Durbin-Watson:

0.025

Prob(Omnibus):

0.003

Jarque-Bera (JB):

12.888

Skew:

0.890

Prob(JB):

0.00159

Kurtosis:

2.704

Cond. No.

1.19e+16

...

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 2.18e-30. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Mask + Restaurant + Restaurant2 + Bar + Movie -> Cases: MSE = 828449541081.99, R2 = 0.34

Hình 22. Bảng thống kê của mô hình hồi quy tuyến tính cho biến số ca nhiễm

NHẬN XÉT:

- Giá trị R^2 rất thấp cho thấy mô hình này không thực sự tốt để dự đoán số ca nhiễm.
- Ta thấy rằng các biến khác đều có hệ số coef dương ngoại trừ biến restaurant. Lý do là trong lúc tình hình dịch bệnh bùng phát mạnh mẽ và tăng dữ dội thì chính sách này lại không được ban hành. Dẫn tới sự trái ngược nhau nên mang hệ số âm. Điều này cũng chứng tỏ việc ban hành chính sách nhà hàng là rất cần thiết để số ca nhiễm giảm.
- Trong các biến trên thì biến có thể tin tưởng nhất và có ý nghĩa nhất là restaurant2 vì có P-value thấp. Còn các biến còn lại không thật sự có ý nghĩa. Điều đó cũng chứng tỏ rằng việc mở lại chính sách đóng cửa nhà hàng thật sự rất quan trọng. Còn các chính sách khác vẫn đang được thực thi trong lúc dịch tăng mạnh nên không thấy rõ sự ảnh hưởng của nó.

OLS Regression Results

Dep. Variable:

Deaths_Total

R-squared:

0.384

Model:

OLS

Adj. R-squared:

0.364

Method:

Least Squares

F-statistic:

18.93

Date:

Sat, 15 Jun 2024

Prob (F-statistic):

1.27e-09

Time:

09:40:30

Log-Likelihood:

-1017.9

No. Observations:

95

AIC:

2044.

Df Residuals:

91

BIC:

2054.

Df Model:

3

Covariance Type:

nonrobust

coef

std err

t

P>|t|

[0.025

0.975]

mask

1795.6508

5609.641

0.320

0.750

-9347.211

1.29e+04

restaurant

-2486.3232

5003.141

-0.497

0.620

-1.24e+04

7451.801

restaurant2

1.662e+04

4418.476

3.763

0.000

7848.062

2.54e+04

bar

2175.8889

1855.214

1.173

0.244

-1509.266

5861.044

movie

2175.8889

1855.214

1.173

0.244

-1509.266

5861.044

Omnibus:

23.155

Durbin-Watson:

0.030

Prob(Omnibus):

0.000

Jarque-Bera (JB):

30.171

Skew:

1.260

Prob(JB):

2.81e-07

Kurtosis:

4.127

Cond. No.

1.19e+16

...

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 2.18e-30. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Mask + Restaurant + Restaurant2 + Bar + Movie -> Deaths: MSE = 118688444.58, R2 = 0.38

Hình 23. Bảng thống kê của mô hình hồi quy tuyến tính cho biến số ca tử vong

NHẬN XÉT:

- Giá trị R2 rất thấp cho thấy mô hình này không thực sự tốt để dự đoán số ca tử vong.
- Ta thấy rằng các biến khác đều có hệ số coef dương ngoại trừ biến restaurant. Lý do là trong lúc tình hình dịch bệnh bùng phát mạnh mẽ và tăng dữ dội thì chính sách này lại không được ban hành. Dẫn tới sự trái ngược nhau nên mang hệ số âm. Điều này cũng chứng tỏ việc ban hành chính sách nhà hàng là rất cần thiết để số ca tử vong giảm.
- Trong các biến trên thì biến có thể tin tưởng nhất và có ý nghĩa nhất là restaurant2 vì có P-value thấp. Còn các biến còn lại không thật sự có ý nghĩa. Điều đó cũng chứng tỏ rằng việc mở lại chính sách đóng cửa nhà hàng thật sự rất quan trọng. Còn các chính sách khác vẫn đang được thực thi trong lúc dịch tăng mạnh nên không thấy rõ sự ảnh hưởng của nó.

5.5 Kết luận giả thuyết

- Từ những phân tích trên ta thấy rằng sự bùng nổ dịch bệnh của hai năm 2021 và 2021 là rất nhanh và lớn mạnh nên sự kiểm soát của đảng chính trị và các chính sách chính phủ có ảnh hưởng khá thấp đến tỷ lệ ca nhiễm và tử vong giữa các nhóm dân tộc tại thời điểm đó do không kịp thời ngăn chặn. Tuy nhiên, ta cũng có thể thấy sự tác động không nhỏ của việc ban hành các chính sách kịp thời của các chính phủ để phòng ngừa nguy cơ lây nhiễm và tử

vong. Ta có thể thấy rõ hơn các chính sách ảnh hưởng rõ rệt hơn qua các bài báo đã được công bố trên các trang uy tín như là Impact of COVID-19 on the United States Economy and the Policy Response) được đăng trên UN-iLibrary [4]. Tóm lại thì trong thời kì căng thẳng nhất của đại dịch COVID-19, các bang khắp nước Mỹ đã nỗ lực ban hành rất nhiều chính sách để kiểm soát dịch bệnh nhưng hiệu quả thì vẫn rất thấp, số ca nhiễm và tử vong vẫn tăng mạnh mặc dù đôi lúc có giảm nhưng không đáng kể, điều này cho thấy rằng rất có thể việc quản lý và yêu cầu người dân tuân thủ các chính sách diễn ra chưa được hiệu quả cũng như người dân ở đây vẫn chưa thực sự chú trọng đến việc tuân thủ các chính sách phòng chống dịch bệnh.

6 - Phân tích sự ảnh hưởng của yếu tố con người đến tình hình COVID-19 tại Mỹ

- Trong đại dịch vừa qua, giãn cách xã hội luôn là yếu tố quan trọng để có thể nhanh chóng đẩy lùi đại dịch. Cùng bắt đầu như vậy nhiều ý kiến cho rằng những nơi càng đông dân cư, mật độ dân số càng cao thì khả năng lây nhiễm và tử vong càng lớn. Bên cạnh đó ở một quốc gia mà có tỉ lệ người vô gia cư cao như ở Mỹ thì nhiều ý kiến cũng quan ngại rằng số lượng người vô gia cư cao cũng là yếu tố khiến cho dịch bệnh tại Mỹ diễn biến phức tạp và căng thẳng. Một số khác lại cho rằng sức khỏe tinh thần cũng rất quan trọng trong đại dịch và các chuyên gia tâm lý cũng đóng một vai trò hết sức quan trọng trong thời kì khủng hoảng. Vậy chúng ta đặt ra và đi xác thực giả thuyết rằng: **Liệu mật độ dân cư và các yếu tố con người khác của mỗi bang có phải là yếu tố ảnh hưởng lớn đến số ca nhiễm và tỷ lệ tử vong không?**

6.1 Tiền xử lý dữ liệu

- Ta sẽ gộp 2 dataframe có từ 2 bộ dữ liệu **CRDT_Data.csv** và **Policies_Social_COVID_19.xlsx** dựa vào cột 'State' để từ đó tạo thành được 1 dataframe mới chứa các thông tin liên quan đến 1 bang cụ thể ở Mỹ.

- Sau quá trình trên ta tạo thành 1 dataset mới từ 2 dataset cũ chứa các cột giá trị:

Cột	Ý nghĩa
State	Postcode của bang
Cases	Số ca dương tính của bang
Deaths	Số ca tử vong của bang
POPDEN21	Mật độ dân cư năm 2021 của bang
HMLS21	Số người vô gia cư năm 2021 của bang
MH21	Số lượng chuyên gia về sức khỏe tâm thần trên mỗi 100,000 dân vào năm 2021

6.2 Mô tả dữ liệu

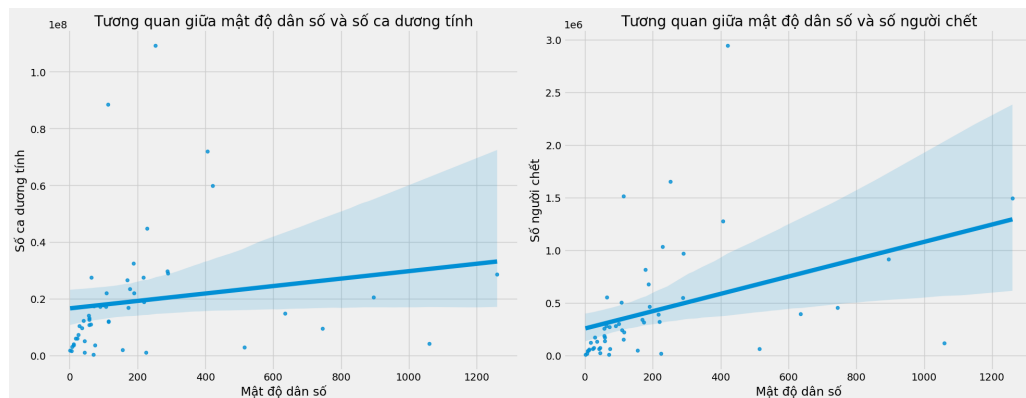
	Cases	Deaths	POPDEN21	HMLS21	MH21
count	5.000000e+01	5.000000e+01	50.000000	50.000000	50.000000
mean	1.934116e+07	4.253429e+05	206.518000	12574.820000	265.862000
std	2.187403e+07	5.551392e+05	274.261603	28801.527673	114.265196
min	3.883060e+05	7.938000e+03	1.280000	532.000000	100.700000
25%	5.323241e+06	6.623500e+04	47.387500	2483.000000	181.350000
50%	1.296777e+07	2.473375e+05	107.955000	4652.000000	244.350000
75%	2.320520e+07	4.942698e+05	222.902500	10087.250000	341.275000
max	1.092985e+08	2.946893e+06	1260.150000	181399.000000	626.600000

Bảng 3. Bảng thống kê các chỉ số của từng bang

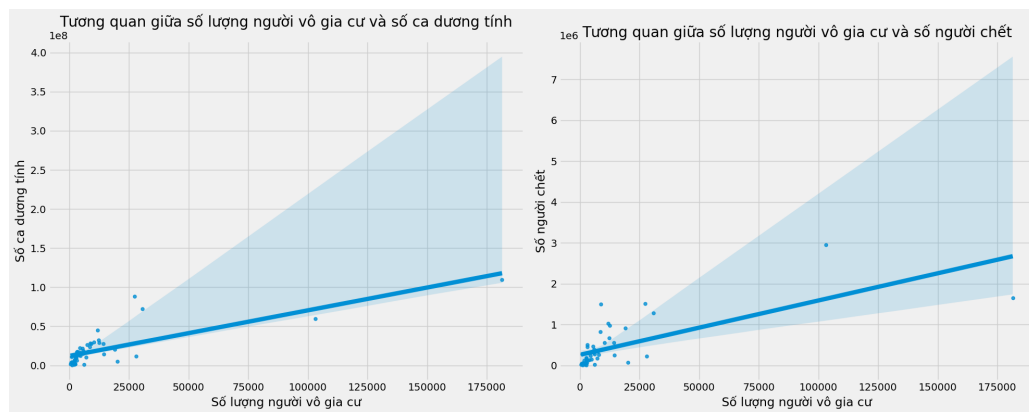
Nhận xét tổng thể:

- Tổng cộng có 50 quan sát, tương ứng với 50 bang của Mỹ.
- ‘Cases’ và ‘Deaths’ có độ lệch chuẩn cao, cho thấy sự phân phối rộng và biến động lớn trong số ca nhiễm và số ca tử vong giữa các quan sát.
- ‘POPDEN21’ và ‘HMLS21’ cũng có độ lệch chuẩn cao, đặc biệt là ‘HMLS21’, cho thấy sự khác biệt đáng kể về mật độ dân cư và số người vô gia cư giữa các khu vực.
- ‘MH21’ có độ lệch chuẩn thấp, cho thấy sự ổn định về số lượng chuyên gia về sức khỏe tâm thần giữa các bang. Giá trị trung bình của ‘MH21’ là xấp xỉ ‘114’, cho thấy trung bình mỗi bang có khoảng 114 chuyên gia về sức khỏe tâm thần trên mỗi 100.000 dân. Đây có thể là một yếu tố quan trọng giúp giảm căng thẳng và tâm lý cho người dân trong thời gian dịch bệnh.
- Các giá trị tại các phân vị 25%, 50% và 75% cho thấy sự phân phối của các biến. Ví dụ, 50% số liệu về số ca nhiễm dưới ‘12,967,770’, nhưng giá trị 75% lên tới ‘23,205,200’, chỉ ra rằng có nhiều quan sát có số ca nhiễm cao hơn mức trung vị đáng kể.
- Tương tự, số ca tử vong cũng có sự phân phối rộng, với giá trị 75% là ‘494,269.8’ nhưng giá trị cao nhất lên tới ‘2,946,893’.
- Các giá trị cực trị (min và max) cho thấy có sự khác biệt lớn giữa các quan sát. Ví dụ, số ca nhiễm tối thiểu là ‘388,306’ trong khi số cao nhất là ‘109,298,500’, cho thấy sự chênh lệch khá cao giữa các bang.

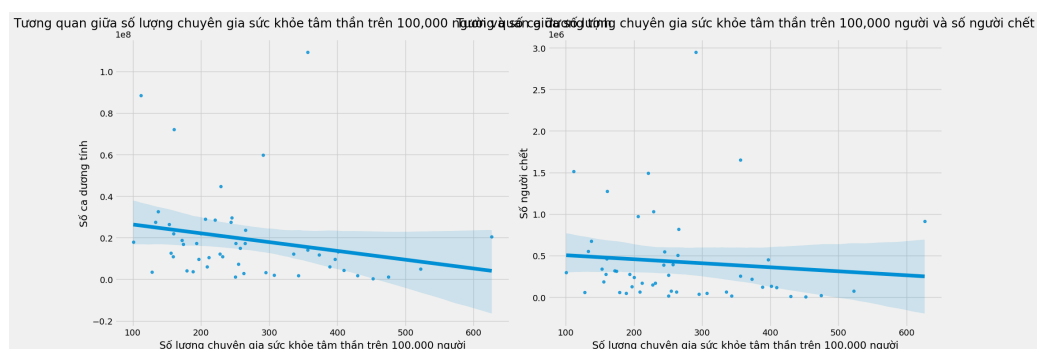
6.3 Trực quan hóa dữ liệu



Hình 24. Sự ảnh hưởng của mật độ dân số tới số ca nhiễm/tử vong



Hình 25. Sự ảnh hưởng của số người vô gia cư tới số ca nhiễm/tử vong



Hình 26. Sự ảnh hưởng của số lượng chuyên gia tâm lý tới số ca nhiễm/tử vong

Nhận xét:

- Dựa vào đường xu thế cũng như sự phân bố của các điểm dữ liệu ta có thể thấy rằng số lượng người vô gia cư càng nhiều thì số ca nhiễm cũng có xu hướng tăng. Điều này càng được thể hiện rõ hơn ở số ca tử vong khi hệ số góc của đường xu hướng lớn hơn. Điều này là dễ hiểu vì người vô gia cư thường không có điều kiện để bảo vệ bản thân khỏi dịch bệnh, họ dễ tiếp xúc với nhiều người khác và không có điều kiện để chăm sóc sức khỏe của mình.
- Trong khi đó ở mật độ dân cư thì cũng có sự tương quan dương với số ca nhiễm và tử vong nhưng lại không được tốt bằng, cụ thể hệ số góc của đường xu hướng nhỏ hơn so với đường xu hướng của số người vô gia cư. Điều này có thể là do mật độ dân cư không phản ánh chính xác số lượng người tiếp xúc với nhau, đặc biệt là ở những nơi có mật độ dân cư cao nhưng có nhiều biện pháp phòng chống dịch bệnh tốt.
- Cuối cùng là số lượng chuyên gia sức khỏe tâm thần, ta thấy rằng số lượng chuyên gia sức khỏe tâm thần càng nhiều thì số ca nhiễm và tử vong càng giảm. Điều này có thể là do sự hỗ trợ tinh thần từ các chuyên gia giúp giảm căng thẳng và lo lắng của người dân, từ đó giúp họ tuân thủ tốt hơn các biện pháp phòng chống dịch bệnh. Đây có thể là một trong những yếu tố quan trọng giúp giảm đi sự ảnh hưởng nặng nề do dịch COVID-19 gây ra tại Mỹ.

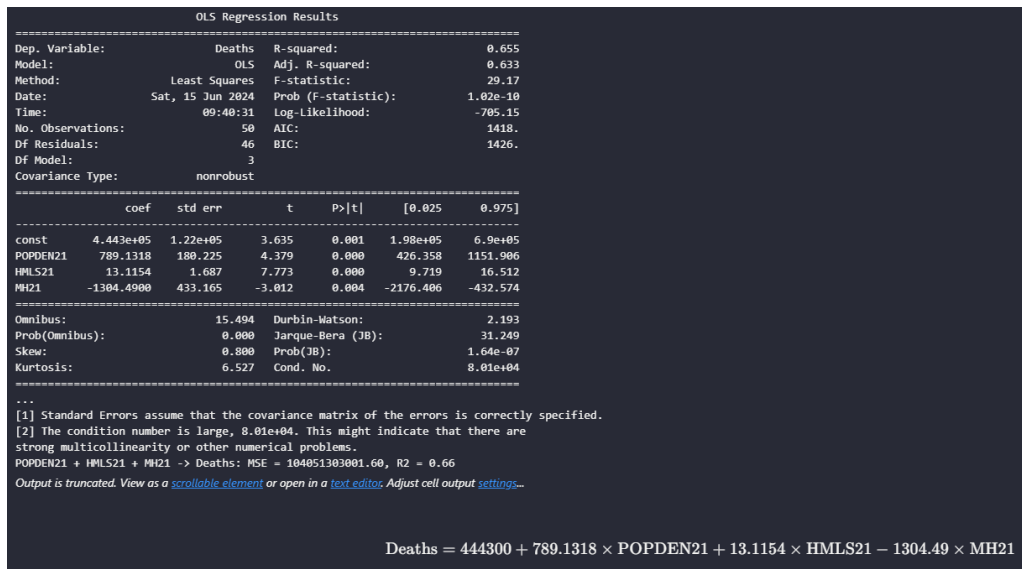
6.4 Áp dụng phương pháp hồi quy tuyến tính

OLS Regression Results						
Dep. Variable:	Cases	R-squared:	0.713			
Model:	OLS	Adj. R-squared:	0.695			
Method:	Least Squares	F-statistic:	38.18			
Date:	Sat, 15 Jun 2024	Prob (F-statistic):	1.53e-12			
Time:	09:40:31	Log-likelihood:	-884.23			
No. Observations:	50	AIC:	1776.			
Df Residuals:	46	BIC:	1784.			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2.717e+07	4.39e+06	6.186	0.000	1.83e+07	3.6e+07
POPDEN21	1.22e+04	6476.100	1.884	0.066	-836.523	2.52e+04
HMLS21	603.4035	60.629	9.952	0.000	481.364	725.443
MH21	-6.746e+04	1.56e+04	-4.334	0.000	-9.88e+04	-3.61e+04
Omnibus:	38.327	Durbin-Watson:	2.001			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	119.579			
Skew:	2.067	Prob(JB):	1.08e-26			
Kurtosis:	9.348	Cond. No.	8.01e+04			
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition number is large, 8.01e+04. This might indicate that there are strong multicollinearity or other numerical problems. POPDEN21 + HMLS21 + MH21 -> Cases: MSE = 134351990030319.55, R2 = 0.71 Output is truncated. View as a scrollable element or open in a text editor . Adjust cell output settings ...						
Cases = 27170000 + 12200 × POPDEN21 + 603.4035 × HMLS21 – 67460 × MH21						

Hình 25. Bảng thống kê cho mô hình hồi quy tuyến tính của biến số ca nhiễm

NHẬN XÉT:

- **R-squared:** 0.713 có nghĩa là 71.3% biến động trong số ca nhiễm Cases được giải thích bởi các biến POPDEN21, HMLS21 và MH21. Đây là một giá trị khá cao, cho thấy mô hình có độ phù hợp tốt.
- **Adjusted R-squared:** 0.695 điều chỉnh cho số lượng biến trong mô hình và thực tế rằng thêm biến vào mô hình sẽ luôn tăng R-squared, điều này cho thấy rằng mô hình vẫn phù hợp tốt với dữ liệu sau khi điều chỉnh.
- **F-statistic:** 38.18 với Prob (F-statistic) = 1.53e-10 chỉ ra rằng mô hình tổng thể là có ý nghĩa thống kê. Điều này có nghĩa là ít nhất một trong các biến độc lập (hoặc cả ba) có tác động đáng kể đến biến phụ thuộc Cases.
- **POPDEN21:** ‘P-value’ là 0.464 cho thấy rằng biến này không có ý nghĩa thống kê (lớn hơn 0.05), do đó không có bằng chứng mạnh mẽ rằng mật độ dân cư ảnh hưởng đến số ca nhiễm.
- **HMLS21:** ‘P-value’ là 0.000 là cực kỳ nhỏ, cho thấy biến này có ý nghĩa thống kê cao.
- **MH21:** ‘P-value’ là 0.000 là cực kỳ nhỏ, cho thấy biến này có ý nghĩa thống kê cao.

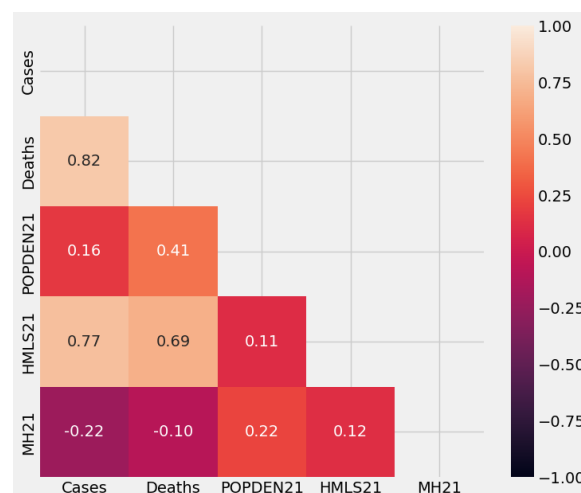


Hình 26. Bảng thống kê cho mô hình hồi quy tuyến tính của biến số ca tử vong

NHẬN XÉT:

- **R-squared:** 0.655 có nghĩa là 65.5% biến động trong số ca tử vong Deaths được giải thích bởi các biến POPDEN21, HMLS21 và MH21.
- **Adjusted R-squared:** 0.633 điều chỉnh cho số lượng biến trong mô hình và vẫn cho thấy mô hình có độ phù hợp tốt.
- **F-statistic:** 29.17 với Prob (F-statistic) = 1.02e-10 chỉ ra rằng mô hình tổng thể là có ý nghĩa thống kê.
- **POPDEN21:** P-value là 0.000 có giá trị nhỏ, cho thấy rằng biến này có ý nghĩa thống kê, có thể thấy rằng đã có sự tác động mạnh mẽ của mật độ dân cư đối với số ca tử vong.
- **HMLS21:** P-value là 0.000 là cực kỳ nhỏ, cho thấy biến này có ý nghĩa thống kê cao
- **MH21:** P-value là 0.004, nhỏ hơn 0.005, cho thấy biến này có ý nghĩa thống kê.

6.5 Sử dụng ma trận hệ số tương quan *Pearson*



Hình 27. Ma trận hệ số tương quan *Pearson*

Nhận xét:

- Quan sát ma trận hệ số tương quan giữa các biến, ta thấy rằng ‘Cases’ và ‘Deaths’ có hệ số tương quan dương cao với ‘HMLS21’ (0.77 và 0.69). Điều này cho thấy rằng số ca nhiễm và tử vong có mối tương quan dương cao với số người vô gia cư. Điều này là dễ hiểu vì người vô gia cư thường không có điều kiện để bảo vệ bản thân khỏi dịch bệnh, họ dễ tiếp xúc với nhiều người khác và không có điều kiện để chăm sóc sức khỏe của mình.
- Trong khi đó mật độ dân cư cũng có hệ số tương quan dương với cả ‘Cases’ và ‘Deaths’, tuy nhiên thấp hơn so với đặc trưng số lượng người vô gia cư. Điều này chứng tỏ mật độ dân cư càng cao thì số ca tử vong càng nhiều, số ca dương tính cũng tăng nhưng thực sự số ca dương tính không bị ảnh hưởng quá nhiều bởi mật độ dân cư.
- Cuối cùng là số lượng chuyên gia sức khỏe tâm thần, ta thấy rằng đặc trưng này có mối tương quan âm với cả ‘Deaths’ và ‘Cases’. Điều này là dễ hiểu vì số lượng chuyên gia sức khỏe tâm thần càng nhiều thì sẽ giúp giảm căng thẳng và lo lắng của người dân, từ đó giúp họ tuân thủ tốt hơn các biện pháp phòng chống dịch bệnh. Đây có thể là một trong những yếu tố quan trọng giúp giảm đi sự ảnh hưởng nặng nề do dịch COVID-19 gây ra tại Mỹ.

6.6 Kết luận giả thuyết

- Đối với mật độ dân cư **POPDEN21**:

- Dựa theo những phân tích trên ta có thể thấy rằng mật độ dân cư cao không phải là một yếu tố chính tác động đến sự lây nhiễm của Covid-19. Mặc dù biểu đồ trực quan cho chúng ta một sự cảm nhận định tính cũng như các kiến thức về xã hội của chúng ta đều khiến chúng ta cảm nhận rằng nơi càng đông dân thì khả năng lây nhiễm càng cao. Tuy nhiên khi phân tích kỹ mô hình hồi quy thì ta có thể thấy điều đó là hoàn toàn sai. Điều này cũng hoàn toàn trùng khớp với kết luận của bài phân tích Are high-density districts more vulnerable to the COVID-19 pandemic? [8]. Bài báo lí giải rằng mặc dù mật độ cao hơn có thể góp phần làm tăng các mối liên hệ xã hội chặt chẽ trong hoàn cảnh bình thường, nhưng điều này không phải lúc nào cũng đúng trong thời kỳ khủng hoảng. Do đó, người dân ở các bang mặc dù có mật độ dân số cao, nhưng đồng thời họ cũng tiếp cận tốt hơn và hiệu quả hơn với cơ sở hạ tầng và dịch vụ đô thị, do đó có thể sống an toàn khi dịch bệnh xảy ra, miễn là họ thực hiện tốt các chính sách giãn cách xã hội. Do đó, các khu vực đô thị và khu đông dân cư cũng có thể an toàn trong đại dịch, cho nên chỉ riêng mật độ không thể được coi là yếu tố rủi ro lớn đối với sự lây lan COVID-19 mà phải xét đến nhiều yếu tố khác.
- Tuy nhiên với tình hình tử vong thì lại khác, mật độ dân cư lớn góp đóng góp lớn trong việc làm tăng số ca tử vong. Lí giải cho việc này ta có thể hiểu rằng, đối với các ca nhiễm mà có khả năng tử vong thì lúc mắc bệnh đa phần là chuyển biến nặng hơn các ca khác và cần điều trị ở cơ quan y tế chứ không thể tự cách ly tại nhà. Mật độ dân số đông tạo nên áp lực cho hệ thống y tế tại địa phương khiến cho khó khăn trong việc điều trị chu đáo cho từng người khiến cho người dân bị thiếu sự nguồn tài nguyên dẫn đến bệnh tình trở nặng và dẫn đến tử vong.

- Đối với số lượng người vô gia cư **HMLS21**:

- Yếu tố này đều hưởng lớn đến cả số ca nhiễm lẫn tử vong. Điều này cũng là hiển nhiên bởi như chúng ta đã đề cập ở trên thì việc giãn cách xã hội là một trong các yếu tố tiên quyết để đẩy lùi đại dịch. Nhưng đối với những người vô gia cư thì lại khó có thể kiểm

soát được điều đó, họ không có nhà, thiếu điều kiện về y tế như khẩu trang, không được tuyên truyền đúng cách dẫn đến việc khiến dịch bệnh lây lan rất nhiều trong cộng đồng những người vô gia cư tại khu vực cũng như lây sang cho các đối tượng khác. Và một khi đã mắc bệnh thì họ cũng khó tiếp cận được với các điều kiện y tế tiên tiến cho nên khả năng sống sót là rất thấp

- Đối với số lượng chuyên gia tâm lý **MH21**:

- Yếu tố này đều hưởng lớn đến cả số ca nhiễm lẫn tử vong. Điều này khá đúng so với những gì chúng ta nhận xét được từ biểu đồ trực quan. Những nơi có nhiều chuyên gia tâm lý thì sức khỏe tinh thần của người dân được chăm sóc tốt hơn do sự hỗ trợ tinh thần từ các chuyên gia giúp giảm căng thẳng và lo lắng của người dân, từ đó giúp họ tuân thủ tốt hơn các biện pháp phòng chống dịch bệnh. Điều này cũng đúng với thông tin mà Viện Y học Quốc gia Hoa Kỳ (National Institutes of Health - NIH) [1] kết luận về sức khỏe tinh thần ảnh hưởng thế nào đối với COVID-19. Cụ thể họ cho rằng những người mắc bệnh hoặc rối loạn tâm thần và sau đó nhiễm COVID-19 có nhiều khả năng tử vong hơn những người không mắc bệnh hoặc rối loạn tâm thần. Sức khỏe tâm thần là trọng tâm nghiên cứu của NIH trong đại dịch COVID-19.

7 - Kết luận

- Qua quá trình phân tích trên, ta có thể rút ra được các kết luận cho mỗi yếu tố ảnh hưởng đến sự lây nhiễm và tử vong của COVID-19 tại Mỹ như sau:

- Tình hình COVID-19 tại Mỹ trong giai đoạn 2020-2021 diễn biến phức tạp và căng thẳng, với số ca nhiễm và tử vong tăng mạnh. Chỉ đến tháng 3/2021 thì mọi thứ mới bắt đầu giảm xuống nhưng vẫn còn cao so với những ngày đầu bùng phát. Điều này cho thấy rằng việc kiểm soát dịch bệnh tại Mỹ chưa thực sự hiệu quả và chậm trễ.
- Về chủng tộc, Mỹ là một quốc gia đa dạng về chủng tộc, chính vì vậy sự ảnh hưởng của COVID-19 tới các nhóm chủng tộc cũng khác nhau. Nhóm người da trắng bị ảnh hưởng nặng nề nhất về số ca nhiễm và tử vong so với các chủng tộc da màu khác (hay còn gọi là 'POCs'). Còn trong nhóm chủng tộc POCs thì người da Mỹ gốc Phi 'Black' và người 'Latinx' bị ảnh hưởng nặng nề nhất so với các chủng tộc da màu còn lại. Tuy nhiên ở đây ta chỉ biết được tổng số ca nhiễm và tử vong chứ không nắm rõ con số ấy chiếm bao nhiêu phần trăm trong dân số nên không thể kết luận rằng chủng tộc nào bị ảnh hưởng nặng nề nhất. Nhưng dựa vào những thông tin trên ta có thể thấy rằng chính sách phòng chống dịch bệnh của Mỹ chưa thực sự hiệu quả đối với cả nhóm chủng tộc da trắng lẫn các nhóm chủng tộc da màu khác.
- Đối với các chính sách xã hội thì ta cũng thấy được chính quyền ở Mỹ cũng rất khẩn trương ban hành nhiều chính sách cách ly nhằm kiểm hãm sự lan tỏa dịch bệnh, tuy nhiên sự hiệu quả đem lại thì không quá đáng kể. Trong các chính sách thì chỉ có chính sách đóng cửa nhà hàng và chỉ cho phép mua về là có ảnh hưởng lớn nhất so với các chính sách còn lại. Tóm lại thì trong thời kỳ căng thẳng nhất của đại dịch COVID-19, các bang khắp nước Mỹ đã rất nỗ lực để ban hành nhiều chính sách kiểm soát dịch bệnh nhưng hiệu quả thì vẫn rất thấp, số ca nhiễm và tử vong vẫn tăng mạnh mặc dù đôi lúc có giảm nhưng không đáng kể, điều này cho thấy rằng rất có thể việc quản lý và yêu cầu người dân tuân thủ các chính sách diễn ra chưa được hiệu quả cũng như người dân ở đây vẫn chưa thực sự chú trọng đến việc tuân thủ các chính sách phòng chống dịch bệnh.

- Còn về yếu tố con người, số lượng người vô gia cư và số chuyên gia tâm lý có ở mỗi bang tác động rất lớn đến số ca nhiễm và tử vong bởi COVID-19, tuy nhiên cách chúng ảnh hưởng lại trái nghịch nhau. Trong khi số lượng người vô gia cư càng lớn thì càng ảnh hưởng tiêu cực thì số chuyên gia tâm lý lại làm điều ngược lại. Điều này là hoàn toàn hợp lý vì như đã phân tích ở trên, người vô gia cư thường không có điều kiện để bảo vệ bản thân khỏi dịch bệnh, họ dễ tiếp xúc với nhiều người khác và không có điều kiện để chăm sóc sức khỏe của mình. Trong khi đó số lượng chuyên gia sức khỏe tâm thần càng nhiều thì sẽ giúp giảm căng thẳng và lo lắng của người dân, từ đó giúp họ tuân thủ tốt hơn các biện pháp phòng chống dịch bệnh. Cuối cùng là mật độ dân số cũng ảnh hưởng ít nhiều đến số ca nhiễm và tử vong. Cụ thể, mật độ dân cư càng cao thì số ca tử vong càng nhiều, số ca dương tính cũng tăng nhưng thực sự số ca dương tính không bị ảnh hưởng quá nhiều bởi mật độ dân cư.

Tài liệu tham khảo

- [1] Mental health during the covid-19 pandemic. *National Institutes of Health*.
- [2] Fda approves first covid-19 vaccine. *FDA*, August 2021.
- [3] Char Adams. Study: Covid's racial disparities made some white people less vigilant about the virus. *NBC News*, April 2022.
- [4] Economic Commission for Latin America and the Caribbean. *Impact of COVID-19 on the United States Economy and the Policy Response*. ECLAC COVID-19 Reports, November 2020.
- [5] ELISABETH GAWTHROP. The color of coronavirus: Covid-19 deaths by race and ethnicity in the u.s. *APM Research Lab*, October 2023.
- [6] Michael Hamilton, Danielle Hamilton, Oluwatamilore Soneye, Olorunshola Ayeyemi, and Raed Jaradat. An analysis of the impact of policies and political affiliation on racial disparities in covid-19 infections and deaths in the usa. *International Journal of Data Science and Analytics*, 13:1–14, 01 2022.
- [7] Tegan Hanlon. Why does alaska have the country's lowest covid-19 death rate? doctors explain. *Alaska Public Media*, September 2020.
- [8] Amir Reza Khavarian-Garmsir, Ayyoob Sharifi, and Nabi Moradpour. Are high-density districts more vulnerable to the covid-19 pandemic? *Sustainable Cities and Society*, 70:102911, 2021.

KẾT THÚC