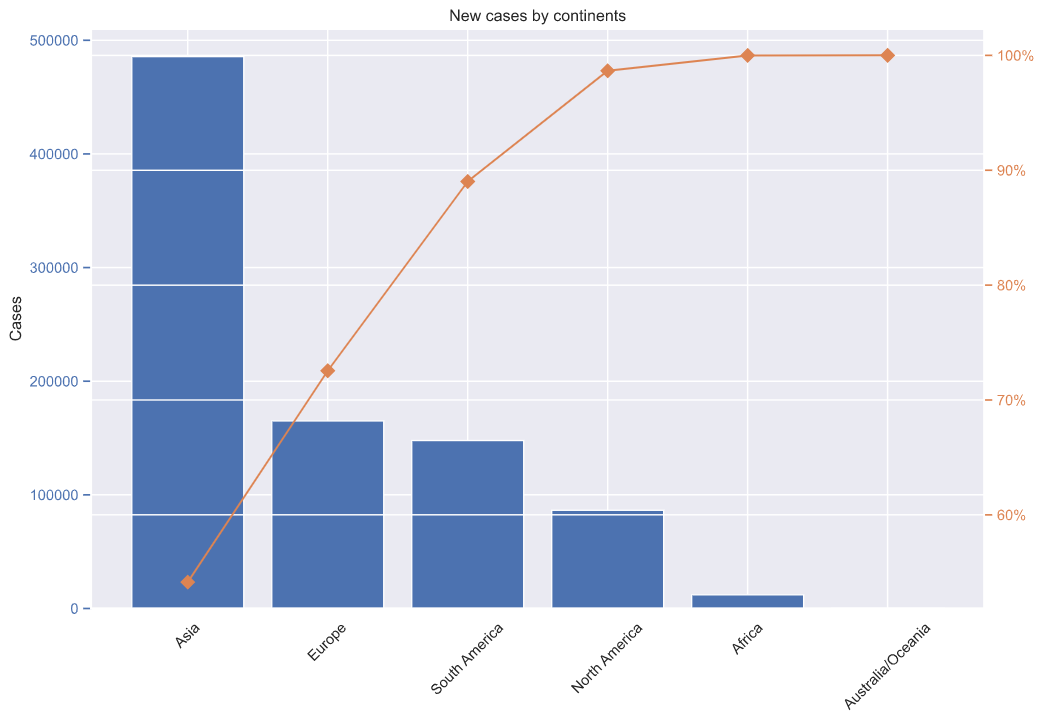
**Pareto Chart.**

Đây là chart được vẽ dựa trên nguyên tắc Pareto: đại đa số mọi thứ trong cuộc sống không được phân phối đều nhau. *Khoảng 80% kết quả là do 20% nguyên nhân gây ra.* Biểu đồ này bao gồm *bar chart* và *line chart*. Các cột trong chart được sắp xếp theo thứ tự từ cao đến thấp theo tần số, còn các giá trị tần suất tích lũy được biểu diễn bằng đường thẳng.

**Quan hệ biểu diễn:** số ca mắc COVID-19 mới trong 1 ngày (23/04/2021) theo lục địa & quốc gia.

**Lý do sử dụng**: Biểu đồ xác định đâu là nơi có nhiều ca mắc mới bộc phát nhất. Mục đích là để tìm ra trong một nhóm nguyên nhân đâu là nguyên nhân quan trọng nhất (lục địa/quốc gia).

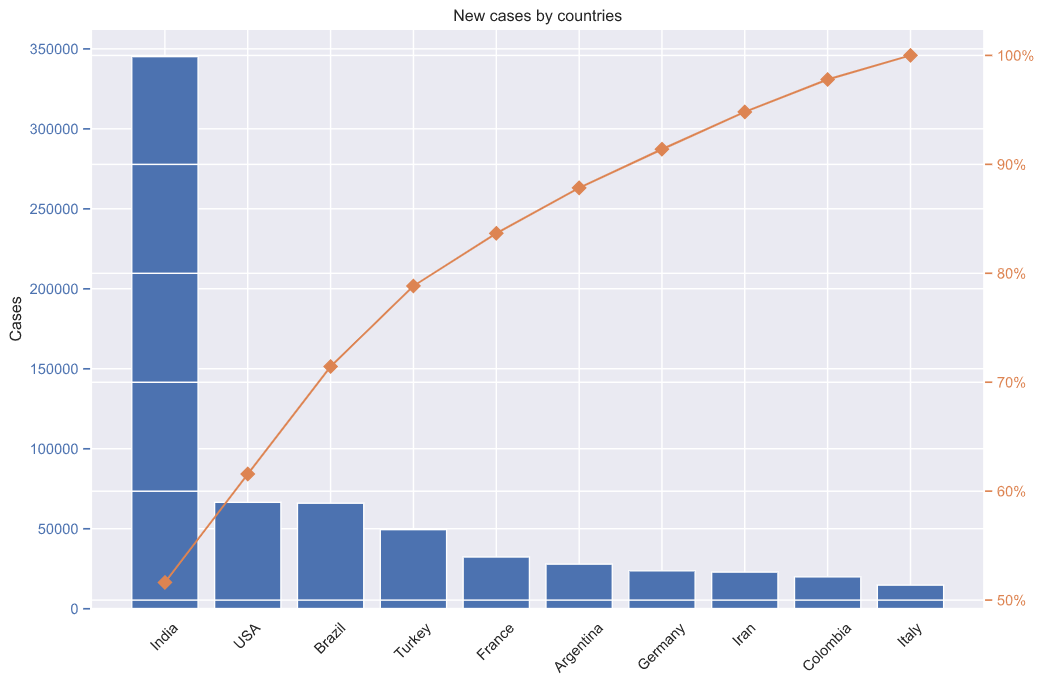
**Nhận xét:** Quan sát biểu đồ, cột trái có đơn vị là tần số, cột phải có đơn vị là tần số tích lũy, đường màu cam thể hiện giá trị tần số tích lũy:

* Chiều cao của cột thể hiện *tần số (frequency)* – đó là số lượng ca mắc mới trong mỗi nhóm châu lục. Ví dụ, có tổng cộng gần 200,000 ca mắc mới ở Châu Âu.
* Đường màu cam thể hiện *tần số tích lũy (cumulative frequency)* – số lượng ca mắc ở các quốc gia có số ca mắc nhỏ hơn và bao gồm số mốc đó. Kết thúc ở 100%.

Dựa vào đường tần số tích lũy này, chiếu xuống trục hoành, ta thấy 80% ca mắc mới thuộc về hầu hết ở 3 châu lục chính: Châu Á, Châu Âu và Nam Mỹ. Điều này có thể chứng tỏ rằng dịch bệnh đang lan rộng và phát triển ở các châu lục này.

Đặc biệt là ở Châu Á, khi số ca mắc mới đạt gần 500,000 trong 1 ngày, cao gấp 2,5 lần so với Châu Âu và cách biệt hoàn toàn so với các châu lục khác.

Ngoài ra, lục địa Châu Úc/Châu Đại Dương không ghi nhận ca mắc mới nào. Đây là mặt tích cực của vấn đề.

🡺 Ta sẽ xem kĩ vào quan hệ với top 10 nước mắc ca nhiễm mới để biết nguyên nhân gây ra số ca nhiễm vượt trội tại Châu Á.

**Nhận xét:** Tập trung vào 80% ca mắc mới thì còn có 4 nước là Ấn Độ, USA, Brazil và Thổ Nhĩ Kỳ.

Đáng chú ý là Ấn Độ, chiếm hơn 50% ca mắc mới trong tổng cộng 10 nước có số ca mắc mới nhiều nhất. Tình hình lây nhiễm tại đây đã bỏ xa cả thế giới. Ngày 23-4 là ngày đầu tiên Ấn Độ ghi nhận số ca nhiễm mới vượt 300.000 ca/ngày, sau chuỗi *8 ngày liên tiếp* nước này có số ca nhiễm mới trên *200.000 ca/ngày.*

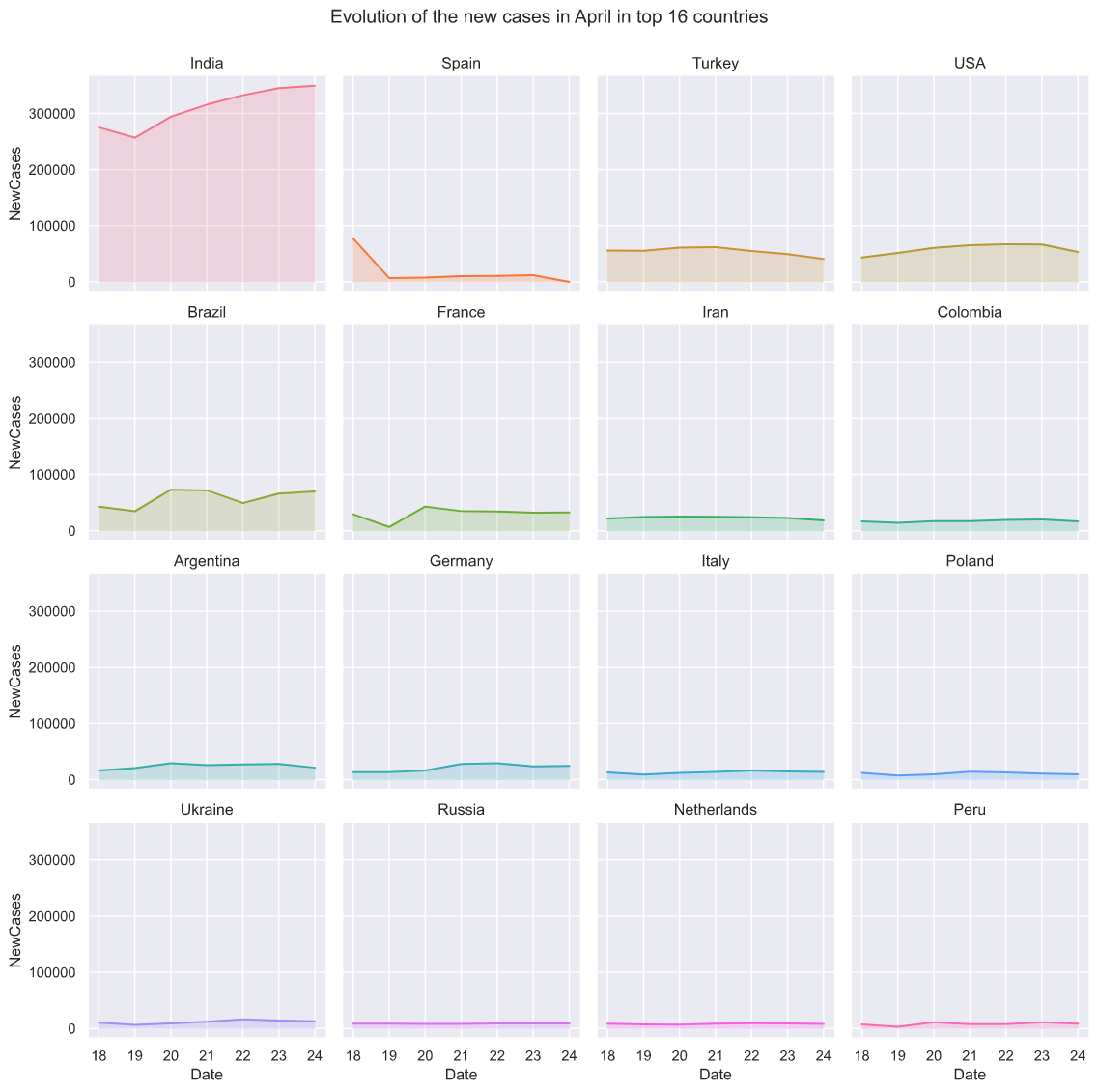
Đây là hậu quả của sự lơ là của người dân trong thực hiện các biện pháp phòng dịch sau khi làn sóng dịch thứ nhất lắng xuống. Người dân lại tụ tập đông trong các sự kiện lễ hội tôn giáo, vận động tranh cử.

🡺 Để xem dữ liệu dưới nhiều góc nhìn, ta tiếp tục sử dụng ***Area chart*** để xem sự thay đổi của số lượng ca mắc mới của các quốc gia trên.

**Area chart.**

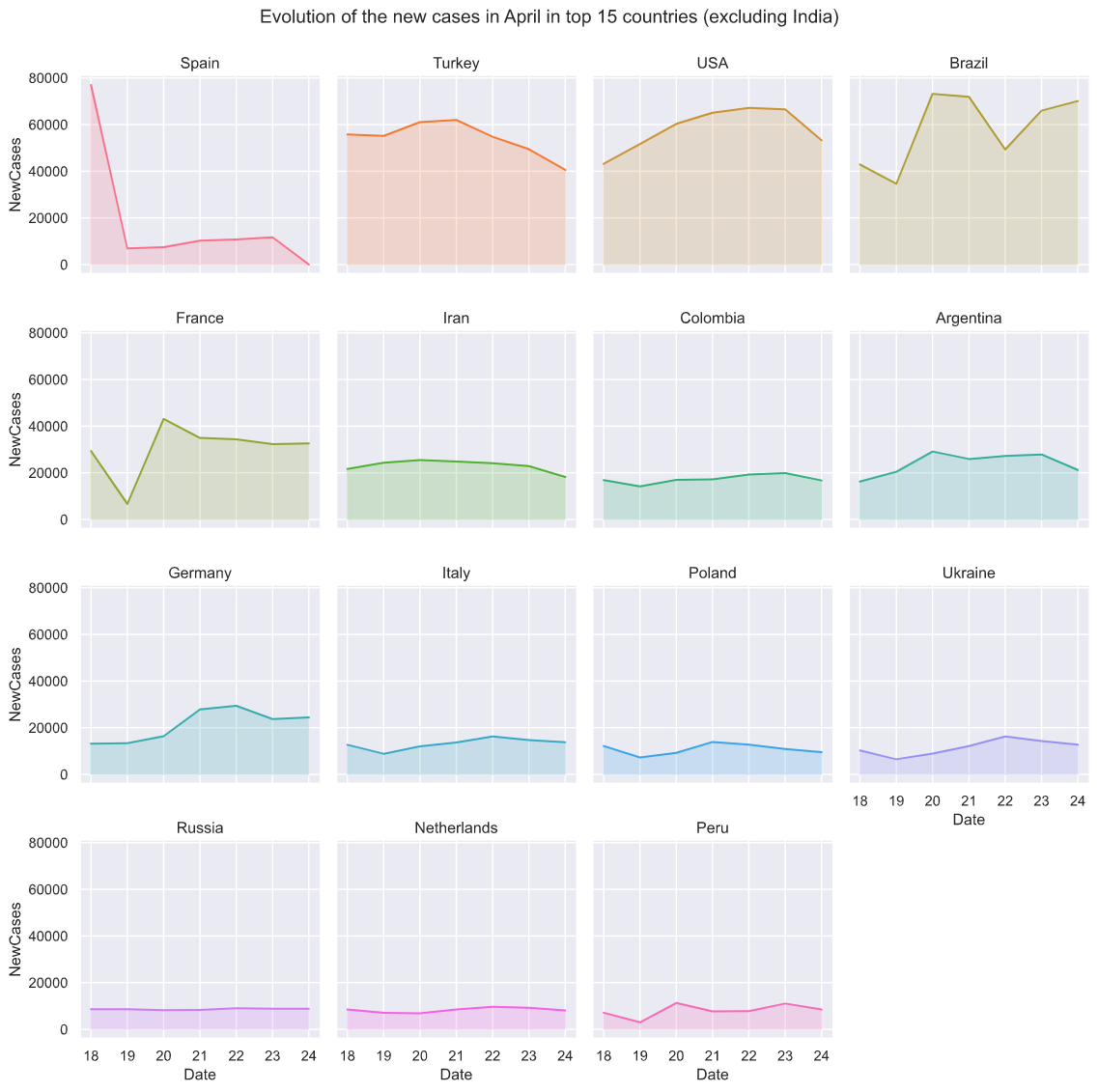
Đây là chart bao gồm một line chart và khu vực giữa line và trục được tô màu để đánh dấu nên có tên gọi là area chart. Thông thường, area chart sẽ được vẽ để xếp chồng lên nhau. Tuy nhiên, vì số lượng chart lớn (16 chart) nên sẽ sử dụng kĩ thuật *faceting*: chia nhỏ các biến dữ liệu trên nhiều ô con và kết hợp các ô con đó thành một hình duy nhất. Vì vậy, thay vì sử dụng thêm một bar chart, chúng ta có thể có quan sát được nhiều biểu đồ hơn, được sắp xếp cùng nhau trong một lưới.

**Quan hệ biểu diễn:** sự thay đổi số ca mắc COVID-19 mới trong 3 ngày (từ 22/4 đến 22/4) của top 16 quốc gia có số lượng ca mắc mới lớn nhất trong ngày.

**Lý do sử dụng:** quan sát được các *tổng số (totals)* theo *thời gian* và tiện lợi để so sánh các quốc gia. Dễ để kiểm chứng với thông tin của ***Pareto Chart*** đã vẽ.

**Nhận xét:** Ấn Độ độc chiếm tier 1, dẫn đầu về số ca mắc mới, gấp nhiều lần so với các nước trong tier 2 như Mỹ, Brazil, Thổ Nhĩ Kỳ, Tây Ban Nha. Đặc biệt từ ngày 19, số ca mắc có chiều hướng tăng mạnh theo từng ngày. Trong các ngày tới, Ấn Độ có thể sẽ cán mốc 400,000 ca/ngày.

Các nước tier 3 như Ukraine, Ba Lan, Peru, Nga, ... thì có số ca mắc mới chỉ bằng 1/2 các nước tier 2.

* Tạm loại bỏ Ấn Độ để xem sự thay đổi của các nước này tốt hơn.

**Nhận xét:** Số ca nhiễm mới ở Mỹ, Brazil và Thổ Nhĩ Kỳ có giá trị cao gấp 2 đến 3 lần so với các nước khác trong top. Song các nước đều có xu hướng giảm đều, chỉ có số ca mắc mới trong ngày ở Brazil đang có chiều hướng tăng trở lại.

Pháp cũng ghi nhận số ca mắc mới tăng đột biến trong ngày 19 nhưng cũng giảm dần ổn định ngay sau đó.

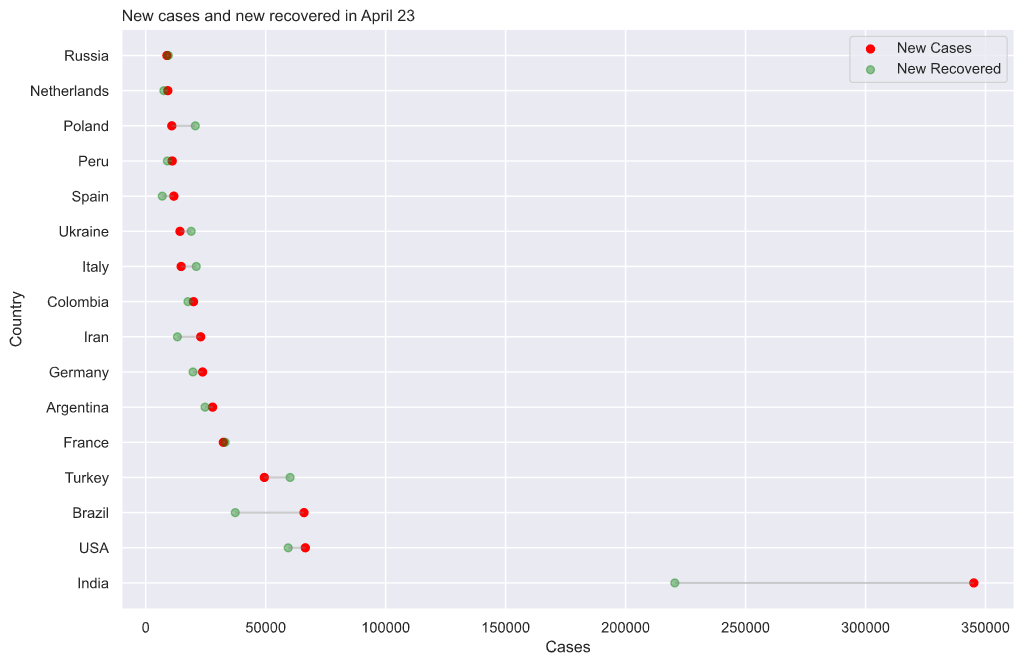
Đặc biệt, ở Tây Ban Nha có một sự giảm ca mới đáng kể trong ngày 19, từ gần 80,000 xuống chỉ còn dưới 20,000 ca. Các nước còn lại cũng ghi nhận số ca mắc mới có biên độ chênh lệch giữa các ngày không biến động nhiều.

**Lolipop chart.**

Đây là một biến thể tiện dụng của *bar chart*; trong đó các bar được thay thế bằng *một line và một dấu chấm ở cuối* (trông như cây lolipop). Cũng giống như bar chart, lollipop chart được sử dụng để so sánh giữa các mục hoặc danh mục khác nhau. Chúng cũng được sử dụng để xếp hạng hoặc hiển thị xu hướng theo thời gian.

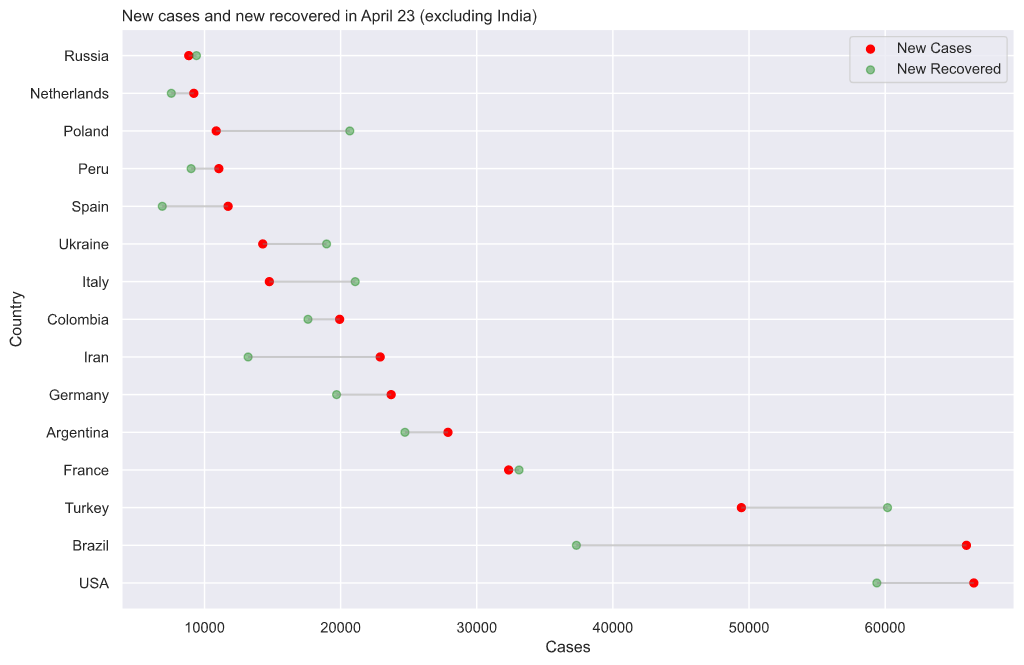
**Quan hệ biểu diễn:** Số ca nhiễm mới xuất hiện và số ca đã hồi phục trong ngày (23/04/2021) của từng quốc gia.

**Lý do sử dụng:** giúp quan sát giá trị quan sát cạnh nhau trên cùng một dòng tốt hơn, bằng cách chỉ hiển thị sự khác biệt của chúng thay vì hiển thị các giá trị trên các dòng khác nhau.



**Nhận xét:** có thể thấy, không chỉ đứng đầu về số ca mới bùng phát trong 1 ngày, Ấn Độ cũng dẫn đầu các nước khác về số ca hồi phục, điều này là hợp lí với tỉ lệ dân số. Tuy nhiên, nếu tập trung về khoảng cách, thì sự chênh lệch giữa ca nhiễm mới và ca hồi phục vẫn còn khá lớn, cho thấy tốc độ hồi phục vẫn còn rất chậm so với tốc độ lây nhiễm. Nếu không có biện pháp khắc phục, Ấn Độ có thể vỡ trận phòng chống COVID-19.

* Do chênh lệch khá lớn. Ta tạm loại bỏ Ấn Độ để xem sự thay đổi của các nước còn lại tốt hơn.



**Nhận xét:** Như vậy, tuy đứng thứ 2 về lượt nhiễm mới, nhưng số ca hồi phục của Mỹ cũng có chiều hướng gần hơn với số ca lây nhiễm.

Đặc biệt là Thổ Nhĩ Kỳ, tuy đứng thứ ba về số ca mắc mới, nhưng số ca hồi lại phục vượt đến hơn 10,000 ca so với ca mắc mới, đây là tín hiệu tốt. Tương tự với quốc gia khác như Ba Lan, Ukraine, Ý.

Các quốc gia còn lại theo xu hướng chung là số ca mới sẽ nhiều hơn một phần so với số ca hồi phục. Đáng lưu ý là Brazil, khi chênh lệch giữa số ca mới và số ca hồi phục là khá lớn. Nếu không cẩn thận phòng chống, rất có thể sẽ kết cục như Ấn Độ.

**Tham khảo:**

<https://stackoverflow.com/questions/53577630/how-to-make-pareto-chart-in-python>

<https://www.datafied.world/web-scraping-live-covid-19-data-and-its-analysis-190>

<https://www.python-graph-gallery.com/242-area-chart-and-faceting>