

KHOA CÔNG NGHỆ THÔNG TIN

ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM

# TRỰC QUAN HÓA DỮ LIỆU



Nhóm thực hiện: Nhóm 11

19127476 - Trần Thị Huế Minh

19127486 - Nguyễn Lê Nguyên

19127125 - Lê Duy Dũng



# MỤC LỤC

<b>I. Phân chia công việc</b>	<b>3</b>
<b>II. Tìm hiểu công cụ Tableau</b>	<b>4</b>
A. Giới thiệu về Tableau	4
1. Tableau là gì?	4
2. Sự ra đời của Tableau?	4
3. Đặc trưng của Tableau	5
4. Ưu nhược điểm của Tableau	6
5. Vai trò của Tableau trong phân tích dữ liệu	7
6. Một số dịch vụ Tableau cung cấp	8
7. Hoạt động của Tableau	8
8. So sánh Tableau và một số công cụ khác	9
B. Tính năng của Tableau	10
1. Tableau Dashboard	10
2. Collaboration and Sharing	14
3. Real-Time Analytics	15
4. Data Blending	16
5. Advanced Visualization	17
6. Map	18
7. Security	19
8. Mobile View	20
9. Ask Data	21
10. Trend Lines and Predictive Analysis	22
<b>III. Trực quan hóa bằng Tableau</b>	<b>23</b>
A. Kết nối dữ liệu nhiều ngày	23
B. Sử dụng biểu đồ để trực quan dữ liệu COVID-19	26
1. Xem xét tình hình số ca nhiễm đang điều trị COVID chưa khỏi bệnh / số ca tử vong / số ca hồi phục trên thế giới	26
Pie Chart	26
Side by side Bar Chart	30
2. Quan sát tình hình dịch bệnh hiện tại của các Quốc gia trên thế giới	36
Map	36
TreeMaps	41
Bar chart	44
Stacked Bar Chart	47
Line Chart	50



3. Một số thuật toán học máy	54
Linear Regression	54
<b>IV. Link tham khảo</b>	<b>58</b>

## I. Phân chia công việc

Họ và tên	MSSV	Công việc			Đánh giá công việc
		Tìm hiểu công cụ Tableau	Trực quan hóa dữ liệu	Báo cáo	
<b>Trần Thị Huệ Minh</b>	19127476	Giới thiệu về Tableau	Trực quan dữ liệu bằng biểu đồ	Viết Báo Cáo	100%
<b>Lê Duy Dũng</b>	19127125	Tính năng của Tableau		Viết Báo Cáo	80%
<b>Nguyễn Lê Nguyên</b>	19127486	Tính năng của Tableau	Áp dụng thuật toán học máy, Map	Viết Báo Cáo	100%

## II. Tìm hiểu công cụ Tableau



### A. Giới thiệu về Tableau

#### 1. Tableau là gì?

Tableau là một công cụ mạnh mẽ và đang được phát triển nhanh nhất, có chức năng phân tích và trực quan hóa dữ liệu trong ngành công nghiệp Business Intelligence. Phần mềm này tổng hợp dữ liệu giống Excel nhưng ở một cấp độ cao hơn và biểu diễn kết quả trực quan nhanh chóng, hiệu quả hơn. Hiểu một cách đơn giản, Tableau chuyển hóa dữ liệu thô dưới dạng bảng thành các định dạng dễ theo dõi và tiếp thu như hình ảnh và biểu đồ mà không cần người dùng biết nhiều kỹ thuật hay kiến thức về lập trình.

#### 2. Sự ra đời của Tableau?

Tableau là kết quả của một dự án khoa học máy tính tại Stanford được ra mắt vào năm 2003 bởi các nhà đồng sáng lập Christian Chabot, Pat Hanrahan và Chris Stolte. Nền tảng này có vai trò cải thiện quá trình phân tích và khiến cho dữ liệu tiếp cận với mọi người dễ dàng hơn.

Qua quá trình phát triển, Tableau đã trở thành giải pháp hữu hiệu với bất kỳ ai làm việc với dữ liệu. Nhờ đó, người dùng nhanh chóng dự đoán được các thông tin chi tiết một cách chính xác nhất.



Năm 2019, Tableau được mua lại bởi Salesforce. Tuy nhiên, sứ mệnh của giải pháp này vẫn là giúp người dùng hiểu về dữ liệu của họ. Đến nay, mọi doanh nghiệp đang áp dụng Tableau vào quy trình hoạt động để tạo ra những lợi ích vượt trội.

### 3. Đặc trưng của Tableau

- **Trực quan hóa dữ liệu:** Tableau cung cấp một loạt các tính năng để tạo ra biểu đồ, bản đồ, bảng tính và các báo cáo động, giúp cho người dùng có thể tương tác với dữ liệu của họ theo nhiều cách khác nhau. Quá trình phân tích dữ liệu diễn ra nhanh chóng, không bị gián đoạn và cập nhật thường xuyên ngay khi có sự thay đổi dữ liệu gốc.
- **Kết nối đến các nguồn dữ liệu khác nhau:** Tableau cho phép người dùng kết nối đến các nguồn dữ liệu khác nhau, bao gồm các cơ sở dữ liệu quan hệ, tệp dữ liệu, các nguồn dữ liệu web và dịch vụ đám mây. Công cụ này cũng hỗ trợ việc kết hợp dữ liệu từ nhiều nguồn khác nhau để tạo ra các báo cáo phức tạp.
- **Giao diện người dùng đơn giản:** Tableau có giao diện đơn giản và dễ sử dụng, vì vậy người dùng không cần có kinh nghiệm lập trình để có thể sử dụng nó. Ngoài ra, Tableau còn cung cấp các công cụ tìm kiếm, lọc dữ liệu và xem trước, giúp người dùng nhanh chóng tìm thấy thông tin mà họ cần.
- **Tính tương tác:** Tableau cho phép người dùng tương tác với dữ liệu của họ bằng cách chọn, lọc và kéo thả các phần tử trên báo cáo. Điều này giúp cho người dùng có thể nhanh chóng khám phá dữ liệu và phát hiện các mẫu và xu hướng.
- **Phân tích dữ liệu và hỗ trợ đưa ra quyết định:** Tableau cung cấp nhiều tùy chọn trực quan hóa, bao gồm biểu đồ, đồ thị và bản đồ. Nó cũng bao gồm các tính năng như ghép dữ liệu, dự báo và hợp tác. Vì vậy, Tableau là một công cụ hữu ích cho các nhà quản lý, nhà phân tích dữ liệu, và các nhân viên làm việc với dữ liệu trong nhiều lĩnh vực khác nhau, bao gồm tài chính, tiếp thị, y tế, và nhiều ngành công nghiệp khác. Các doanh nghiệp, tổ chức và cá nhân sử dụng Tableau để phân tích dữ liệu và đưa ra quyết định.
- **Tính đa nền tảng:** Tableau có thể sử dụng trên nhiều hệ điều hành khác nhau, bao gồm Windows và macOS, và cũng hỗ trợ trên các thiết bị di động như điện thoại thông minh và máy tính bảng.
- **Khả năng chia sẻ và xuất báo cáo:** Tableau cho phép người dùng chia sẻ báo cáo và biểu đồ của họ với các thành viên khác trong tổ chức, cũng như xuất báo cáo dưới dạng tệp PDF, Excel hoặc ảnh.
- Khi sử dụng Tableau, người dùng cần quan tâm đến các chức năng chính như sau:



- Dịch queries thành visualization.
- Quản lý siêu dữ liệu.
- Nhập dữ liệu của tất cả các kích cỡ và phạm vi.
- Đối với người dùng không có kỹ thuật, Tableau cung cấp phương tiện để tạo ra các queries không có code.
- Hỗ trợ phân tích dữ liệu với Big Data.
- Dễ dàng xây dựng các Business Dashboard
- Khám phá dữ liệu hoạt động của doanh nghiệp
- Xây dựng các tài liệu về dữ liệu hoạt động (Data Stories)
- Tính năng mô phỏng dữ liệu (Data visualization)
- Phân tích dữ liệu của mạng xã hội
- Xây dựng bản đồ
- Hỗ trợ các thiết bị di động
- Tích hợp với ngôn ngữ R
- Phân tích khảo sát
- Phân tích theo thời gian

## 4. Ưu nhược điểm của Tableau

### **Ưu điểm:**

- **Trực quan hóa dữ liệu:** Tableau là một công cụ trực quan hóa dữ liệu và cung cấp tính toán phức tạp, trộn dữ liệu và xây dựng dashboard để tạo trực quan hóa dữ liệu đẹp mắt.
- **Tạo nhanh hình ảnh có tính tương tác cao:** Người dùng có thể tạo ra các hình ảnh mà dữ liệu được hiển thị có tính tương tác cao bằng cách sử dụng các chức năng kéo và thả của Tableau.
- **Dễ dàng khi triển khai:** Nhiều loại tùy chọn trực quan hóa có sẵn trong Tableau, giúp nâng cao trải nghiệm người dùng. Tableau rất dễ học so với Python. Những người không có bất kỳ ý tưởng nào về lập trình cũng có thể nhanh chóng tìm hiểu Tableau.
- **Tableau có thể xử lý một lượng lớn dữ liệu:** Tableau có thể dễ dàng xử lý hàng triệu hàng dữ liệu. Một lượng lớn dữ liệu có thể tạo ra các loại trực quan hóa khác nhau mà không ảnh hưởng đến hiệu suất của dashboard. Ngoài ra, có một tùy chọn trong Tableau nơi người dùng có thể tạo kết nối trực tiếp đến các nguồn dữ liệu khác nhau như SQL,...



- **Sử dụng ngôn ngữ Scripting khác trong Tableau:** Ngoài việc kéo thả thì ta cũng có thể thực hiện việc trực quan hóa trên tableau bằng các ngôn ngữ lập trình như Python. Tuy nhiên đây không phải cách tối ưu.
- **Hỗ trợ trên di động và dashboard responsive:** Việc hiển thị dữ liệu trên các thiết bị di động của tableau cũng rất linh hoạt.

### Nhược điểm:

- **Phải tự lên lịch báo cáo:** Tableau không cung cấp lịch báo cáo tự động nên khi muốn cập nhật dữ liệu từ backend thì người dùng phải làm thủ công.
- **Không import được các custom visual.**
- **Định dạng tùy chỉnh trong Tableau:** Định dạng có điều kiện của Tableau và giới hạn 16 cột bảng rất bất tiện cho người dùng. Ngoài ra, để triển khai cùng một định dạng trong nhiều trường, không có cách nào để người dùng có thể thực hiện trực tiếp cho tất cả các trường. Người dùng phải làm điều đó theo cách thủ công cho từng trường, vì vậy rất tốn thời gian.
- **Tham số static và single value:** Các tham số của Tableau là tĩnh và nó luôn chọn một giá trị duy nhất làm tham số. Bất cứ khi nào dữ liệu được thay đổi, các tham số này cũng phải được cập nhật thủ công mỗi lần. Không có tùy chọn nào khác cho người dùng có thể tự động cập nhật các tham số.
- **Độ phân giải màn hình trên Dashboard Tableau:** Bố cục của dashboard được phân phối nếu độ phân giải màn hình của nhà phát triển Tableau khác với độ phân giải màn hình của người dùng.

## 5. Vai trò của Tableau trong phân tích dữ liệu

Phân tích và trực quan hóa dữ liệu là những bước quan trọng trong ngành khoa học dữ liệu và đặc biệt là các ứng dụng của nó trong kinh doanh. Không chỉ là đưa ra các phép so sánh, tổng hợp thông tin quan trọng mà còn giúp dự báo các xu hướng thị trường, hỗ trợ đưa ra các quyết định tối ưu nhất.

Báo cáo truyền thống tốn nhiều thời gian và chưa đem lại hiệu quả thật sự bởi phần lớn dữ liệu thô đều được lưu trữ dưới dạng bảng (excel sheets, google sheets,...) với vô số trường và bản ghi dữ liệu. Các công cụ phân tích và trực quan như Tableau giúp tự động hóa việc phân tích dữ liệu với tốc độ rất nhanh chóng và trực quan, đáp ứng nhu cầu của các doanh nghiệp cần đưa ra đánh giá và quyết định kinh doanh với tốc độ cao.



## 6. Một số dịch vụ Tableau cung cấp

Tableau cung cấp một số dịch vụ sau:

- **Tableau Desktop:** Là phần mềm trên máy tính để bàn cho phép người dùng kết nối, tạo và phân tích dữ liệu để tạo ra các trực quan hóa dữ liệu.
- **Tableau Server:** Là một nền tảng trực tuyến cho phép người dùng chia sẻ và quản lý các bảng điều khiển và trực quan hóa dữ liệu với các thành viên khác trong tổ chức.
- **Tableau Prep:** Là một công cụ để làm sạch và chuẩn bị dữ liệu trước khi phân tích bằng Tableau.
- **Tableau Online:** Là phiên bản trực tuyến của Tableau Server, cho phép người dùng truy cập và chia sẻ các bảng điều khiển và trực quan hóa dữ liệu từ bất kỳ đâu trên thế giới.
- **Tableau Mobile:** Là một ứng dụng di động cho phép người dùng truy cập và xem các bảng điều khiển và trực quan hóa dữ liệu trên điện thoại di động hoặc máy tính bảng.
- **Tableau Public:** Là một dịch vụ miễn phí cho phép người dùng chia sẻ các bảng điều khiển và trực quan hóa dữ liệu của mình trực tuyến với người dùng khác trên toàn cầu.
- **Tableau Reader:** Là một phần mềm miễn phí cho phép người dùng xem các bảng điều khiển và trực quan hóa dữ liệu được tạo bởi Tableau Desktop và được chia sẻ bởi người dùng khác.

## 7. Hoạt động của Tableau

Tableau hoạt động bằng cách kết nối đến các nguồn dữ liệu khác nhau và biến chúng thành các trực quan dữ liệu mạnh mẽ và dễ hiểu. Hoạt động cụ thể của Tableau bao gồm:

- Kết nối đến các nguồn dữ liệu: Tableau có thể kết nối đến nhiều nguồn dữ liệu khác nhau như Excel, CSV, SQL Server, Oracle, Hadoop,...
- Tạo các dashboard và trực quan dữ liệu: Người dùng có thể tạo các bảng điều khiển và trực quan hóa dữ liệu bằng cách sử dụng các tính năng kéo và thả của Tableau. Tableau cung cấp nhiều loại trực quan hóa dữ liệu như biểu đồ cột, biểu đồ đường, bản đồ, v.v.
- Thực hiện phân tích trên dữ liệu: Tableau cho phép người dùng thực hiện các phân tích trên dữ liệu bằng cách sử dụng các tính năng tính toán và lọc dữ liệu.
- Chia sẻ dashboard và các trực quan dữ liệu: Tableau cho phép người dùng chia sẻ các bảng điều khiển và trực quan hóa dữ liệu với người dùng khác trong tổ chức thông qua Tableau Server hoặc Tableau Online.



- Điều chỉnh dashboard và trực quan hóa dữ liệu cho thiết bị di động: Tableau cho phép người dùng điều chỉnh các dashboard và trực quan hóa dữ liệu để phù hợp với các thiết bị di động như điện thoại di động và máy tính bảng.
- Sử dụng các tính năng tùy chỉnh: Tableau cung cấp nhiều tính năng tùy chỉnh để người dùng có thể tạo ra các trực quan hóa dữ liệu độc đáo và phù hợp với nhu cầu của họ.

## 8. So sánh Tableau và một số công cụ khác

### a. Tableau và Excel

Tableau và Excel đều là các công cụ phân tích dữ liệu phổ biến, nhưng có những khác biệt cơ bản giữa chúng:

- Mục đích sử dụng: Excel thường được sử dụng để thực hiện các tính toán và phân tích dữ liệu cơ bản, trong khi Tableau được sử dụng để tạo các trực quan hóa dữ liệu mạnh mẽ và dễ hiểu.
- Quy mô dữ liệu: Excel có thể xử lý và phân tích các bảng tính đơn giản, trong khi Tableau có thể xử lý và phân tích các tập dữ liệu lớn và phức tạp hơn.
- Tính tương tác: Tableau có tính tương tác cao hơn so với Excel. Tableau cho phép người dùng tương tác với các trực quan hóa dữ liệu, như thay đổi các thông số và lọc dữ liệu trực tiếp trên bảng điều khiển, trong khi Excel yêu cầu người dùng thực hiện các thao tác thủ công để thay đổi và lọc dữ liệu.
- Tính linh hoạt: Excel có tính linh hoạt cao hơn trong việc xử lý và sắp xếp dữ liệu, trong khi Tableau có tính linh hoạt cao hơn trong việc tạo các trực quan hóa dữ liệu và chia sẻ chúng với người dùng khác.
- Thời gian đáp ứng: Tableau có thể xử lý và hiển thị dữ liệu nhanh hơn so với Excel, đặc biệt là khi xử lý các tập dữ liệu lớn và phức tạp.

Tóm lại, Excel thích hợp cho các phân tích dữ liệu cơ bản và tính toán, trong khi Tableau là lựa chọn tốt hơn cho việc tạo các trực quan hóa dữ liệu mạnh mẽ và dễ hiểu. Tuy nhiên, sự lựa chọn giữa hai công cụ này phụ thuộc vào mục đích sử dụng và độ phức tạp của dữ liệu cần phân tích.

### b. Tableau và PowerBI

Tableau và Power BI đều là các công cụ phân tích dữ liệu phổ biến, nhưng có những khác biệt cơ bản giữa chúng:

- Tính năng và tính linh hoạt: Tableau có nhiều tính năng hơn so với Power BI và có tính linh hoạt cao hơn trong việc tạo các trực quan hóa dữ liệu, trong khi



Power BI có tính năng tích hợp tốt hơn với các sản phẩm Microsoft và có thể dễ dàng tích hợp với các nguồn dữ liệu từ Microsoft.

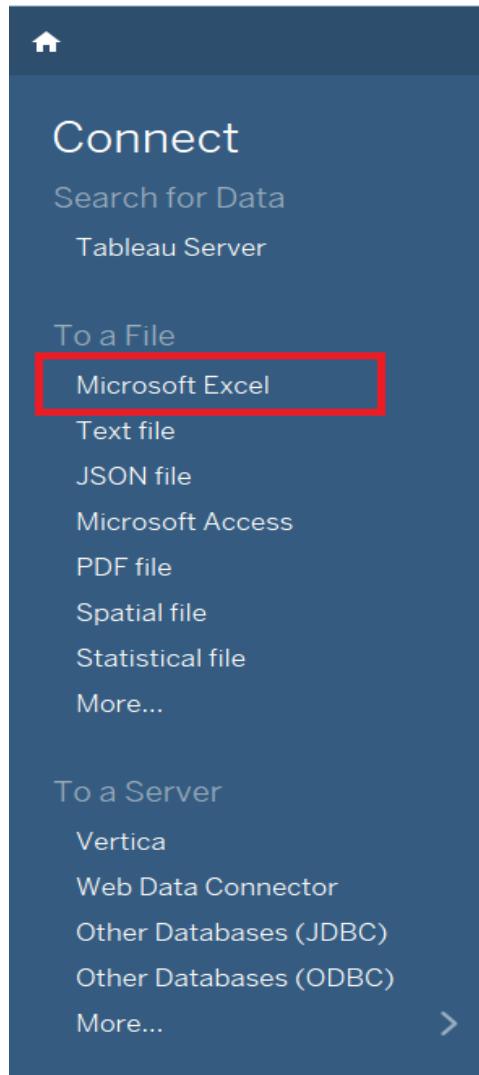
- **Tính tương tác:** Tableau có tính tương tác cao hơn so với Power BI. Tableau cho phép người dùng tương tác với các trực quan hóa dữ liệu, như thay đổi các thông số và lọc dữ liệu trực tiếp trên bảng điều khiển, trong khi Power BI yêu cầu người dùng thực hiện các thao tác thủ công để thay đổi và lọc dữ liệu.
- **Quy mô dữ liệu:** Tableau có thể xử lý và phân tích các tập dữ liệu lớn và phức tạp hơn so với Power BI, nhưng Power BI có tính năng tích hợp dữ liệu tốt hơn và có thể dễ dàng kết nối với các nguồn dữ liệu từ Microsoft.
- **Tính di động:** Power BI có tính di động cao hơn so với Tableau, cho phép người dùng truy cập và xem các báo cáo trực tuyến trên các thiết bị di động.
- **Thời gian đáp ứng:** Power BI có thời gian đáp ứng nhanh hơn so với Tableau, đặc biệt là khi xử lý các tập dữ liệu lớn và phức tạp.

Tóm lại, Tableau thích hợp cho việc tạo các trực quan hóa dữ liệu mạnh mẽ và dễ hiểu, trong khi Power BI có tính năng tích hợp tốt hơn và có tính di động cao hơn. Sự lựa chọn giữa hai công cụ này phụ thuộc vào mục đích sử dụng và độ phức tạp của dữ liệu cần phân tích, cũng như các yêu cầu tích hợp và di động của tổ chức.

## B. Tính năng của Tableau

### 1. Tableau Dashboard

- **Tableau Dashboard:** Tableau Dashboard có Dashboard trực quan với các trình hướng dẫn tự giải thích, cho phép người dùng không có kỹ thuật dễ dàng tạo các trực quan dữ liệu. Chỉ với các thao tác kéo thả đơn giản, ta cũng có thể làm ra được một Dashboard hoàn chỉnh với Tableau. Dưới đây là ví dụ về việc sử dụng Tableau để xây dựng một Dashboard cơ bản.



Đầu tiên khi vừa vào Tableau chọn mục import File Excel. Sau đó ta lựa chọn File data đã chuẩn bị.

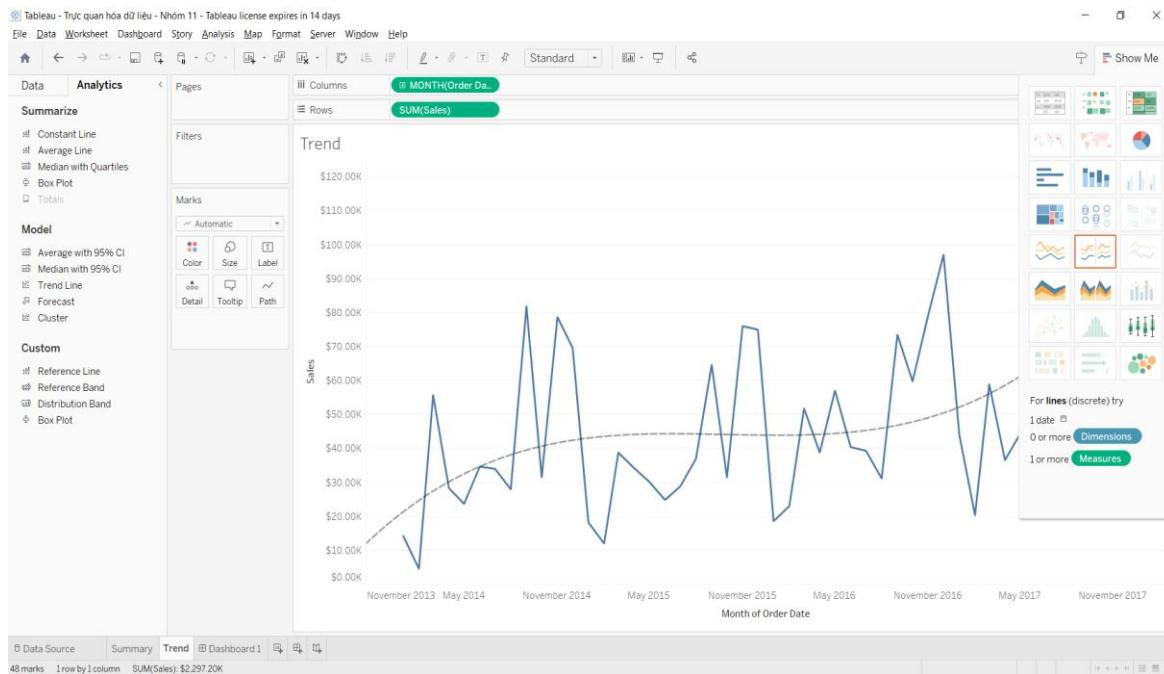


The screenshot shows the Tableau Data Source interface. On the left, there's a sidebar with 'Connections' (Sample - Superstore, Microsoft Excel), 'Sheets' (Orders, People, Returns), and 'New Union', 'New Table Extension'. The main area displays the 'Orders' sheet with a preview of 21 fields and 9994 rows. A table view shows columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, and Customer Name. Below the table, a preview of six rows of data is shown. At the bottom, there are tabs for 'Data Source' and 'Sheet 1'.

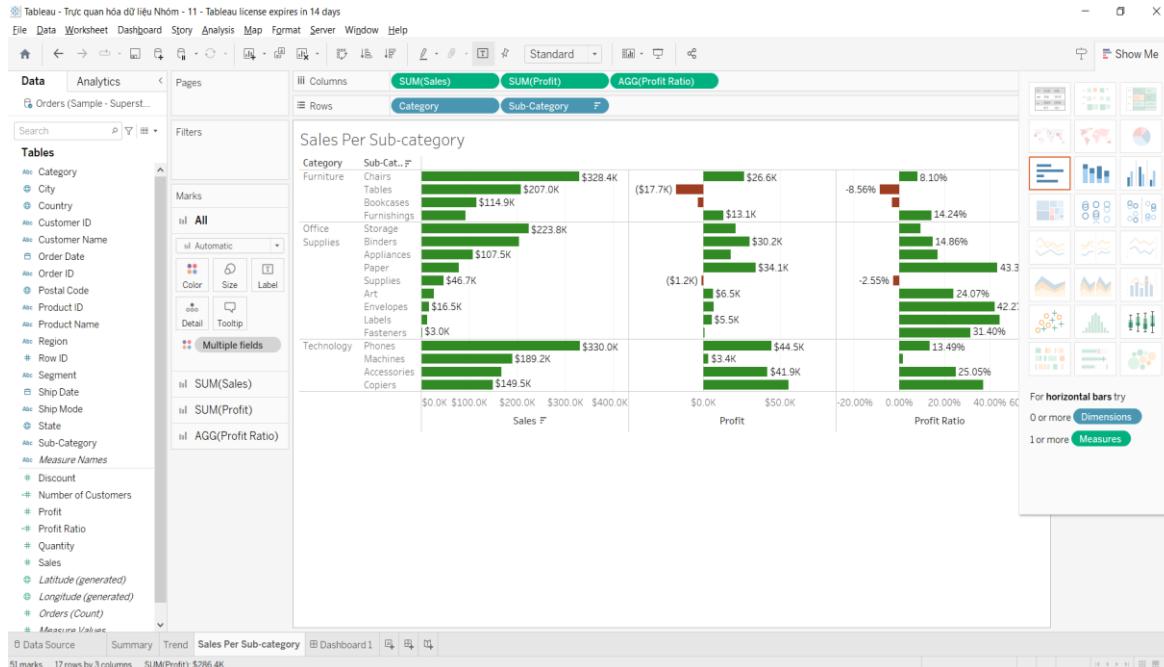
Tiếp theo trong phần Data Source ta kéo thả Sheet mà mình muốn sử dụng để lấy dữ liệu. Sau đó nhấn vào Worksheet để bắt đầu xây dựng Dashboard.

The screenshot shows the Tableau Worksheet interface. On the left, the 'Data' pane lists various dimensions and measures like City, Country, Customer ID, etc. The 'Marks' pane shows 'Automatic' selected with options for Color, Size, and Text. The 'Measures' pane lists AGG(Number of Cus...), SUM(Profit), AGG(Profit Ratio), and SUM(Sales). The main workspace shows a single table with four columns: Number of Customers, Profit, Profit Ratio, and Sales, with values 793, \$286.4K, 12.47%, and \$2,297.20K respectively. To the right, a 'Show Me' panel displays various chart and visualization options.

Sau khi tạo thêm, kéo thả các Measures vào Sheet, chuyển nó về dạng table và định dạng lại thì ta đã có một phần của Dashboard.



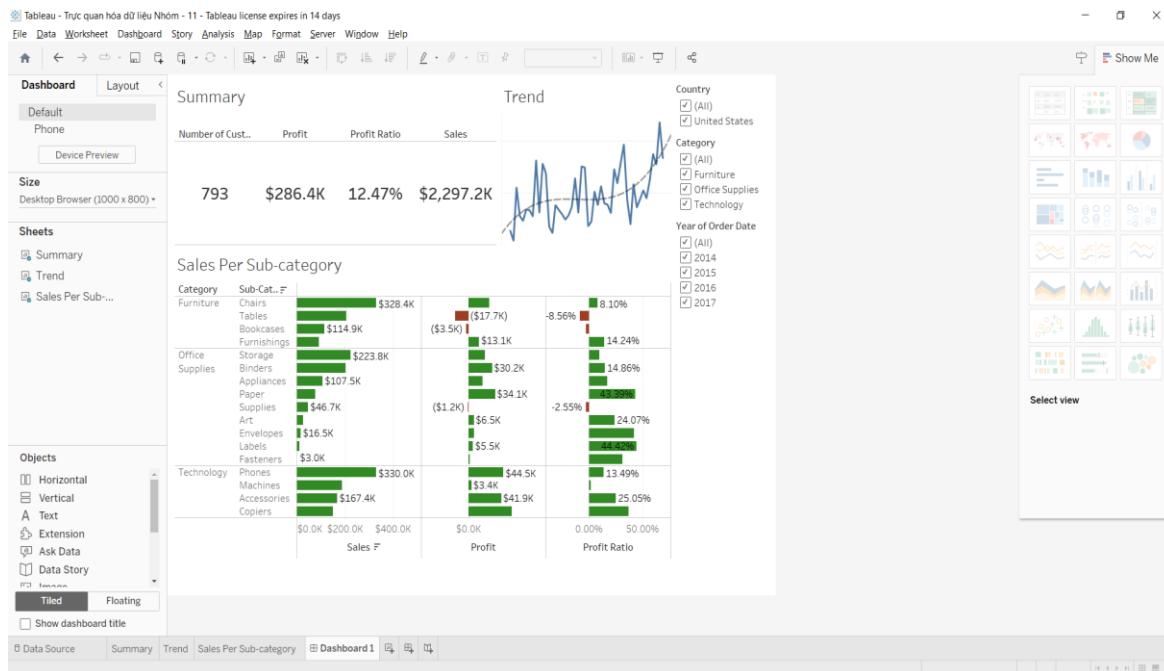
Sau đó ta thêm một Sheet mới và đặt tên nó là Trend. Tiến hành kéo thả Measure Sales và Table Order Date vào sheet sau đó thực hiện chỉnh sửa cũng như thêm Trendline vào sheet thì ta đã có được một Trendline để sử dụng cho Dashboard.



Tiếp theo ta tạo sheet Sales Per Sub-category để thể hiện các thông số bán hàng của mỗi category. Thêm các measures Sales, Profit, Profit Ratio và các table Category, Sub-category vào sheet ta sẽ có được các biểu đồ cột. Tiến hành chỉnh



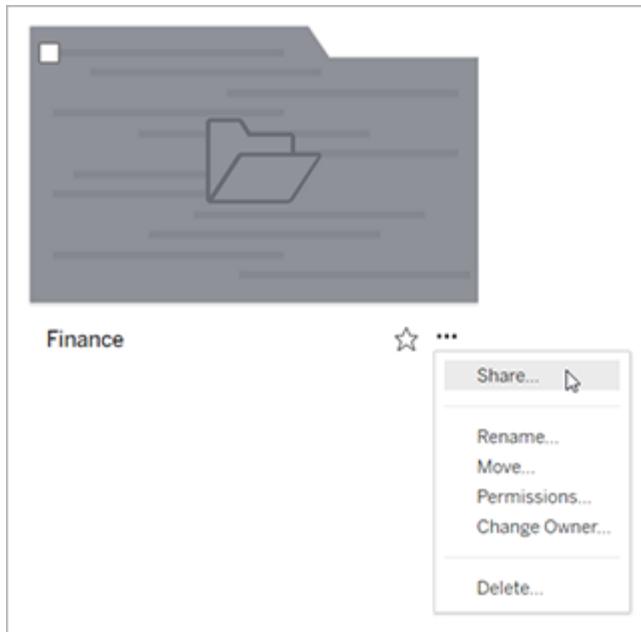
sửa cho khi giá trị của Profit và Profit Ratio là số dương thì sẽ hiển thị màu xanh và màu đỏ khi là số âm.



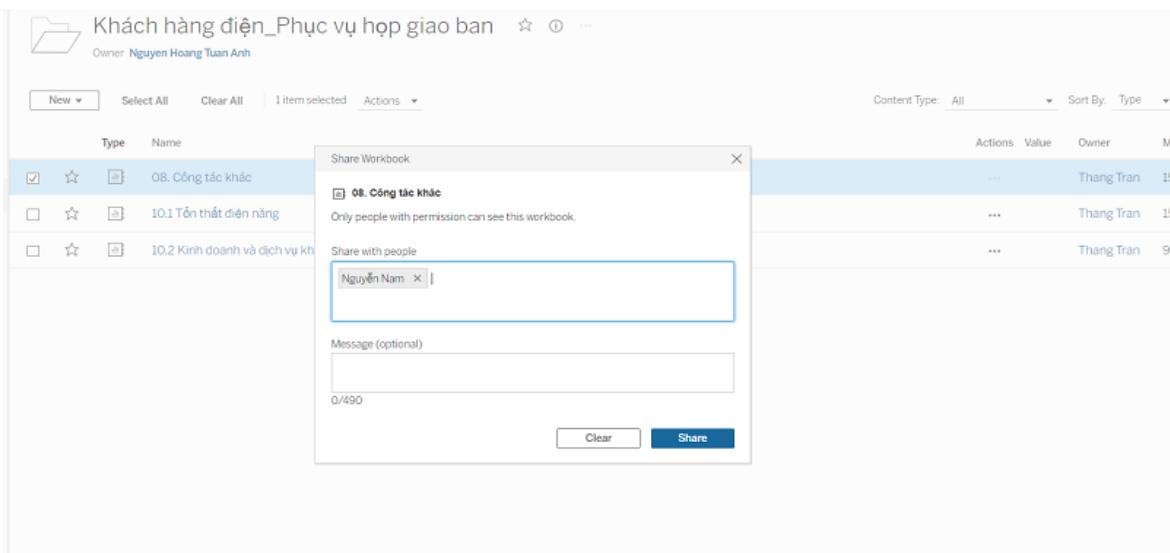
Sau đó ta tạo một Dashboard mới và kéo thả các sheet vừa cấu hình vào và chỉnh sửa chúng. Vậy là chỉ với một số thao tác đơn giản, ta đã có thể tạo ra một Dashboard bằng Tableau.

## 2. Collaboration and Sharing

- Tableau cho phép người dùng cộng tác với các đồng nghiệp của họ để cùng làm việc hoặc đánh giá về một vấn đề hay công việc nào đó. Người dùng cũng có thể chia sẻ lên Cloud, giúp Dashboard có thể truy cập được từ mọi nơi.
- Tuy nhiên để sử dụng được chức năng này ta phải kích hoạt Tableau và có một server để kết nối mọi người thông qua đó.
- Khi sử dụng tính năng chia sẻ với các người dùng khác trong cùng một trang (site), họ sẽ nhận được mail thông báo và có đường dẫn (link) đính kèm đến nội dung được chia sẻ. Ngoài ra, các nội dung được chia sẻ sẽ được lưu trong mục Shared with Me, để họ có thể dễ dàng tìm kiếm lại nội dung.
- Chọn Menu Actions (...) tại nội dung bạn muốn chia sẻ.
- Chọn Share.



- Tại thẻ Share with people, chọn những người bạn muốn chia sẻ.



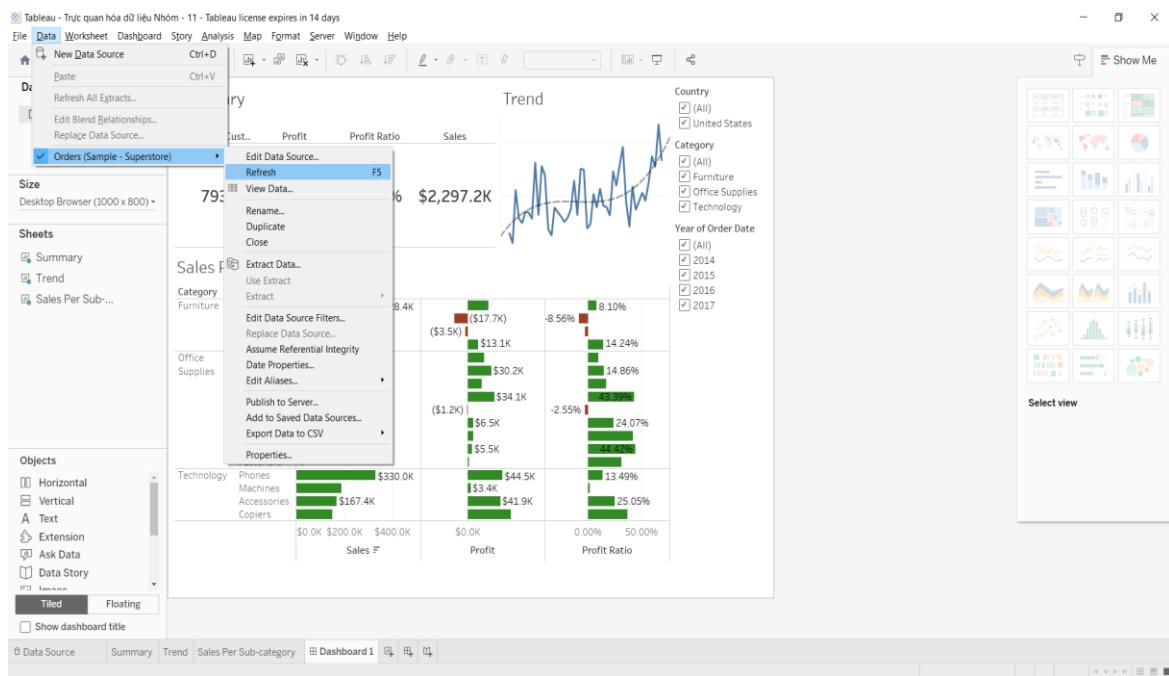
The screenshot shows a SharePoint interface for sharing a workbook. On the left, there's a navigation pane with categories like '08. Công tác khác', '10.1 Tổn thất điện năng', and '10.2 Kinh doanh và dịch vụ KH'. In the center, a 'Share Workbook' dialog box is open, showing the user 'Nguyễn Nam' entered in the 'Share with people' field. At the bottom of the dialog, there's a message input field and a 'Share' button. The background shows a list of users with their names and counts: 'Thang Tran' (15), 'Thang Tran' (15), and 'Thang Tran' (9).

- Chọn Share.

### 3. Real-Time Analytics

Tableau Real Time Data Streaming là một cách để đảm bảo rằng các công ty trích xuất thông tin chi tiết từ dữ liệu cập nhật để đưa ra quyết định kinh doanh đúng đắn.

Đầu tiên ta có thể cập nhật lại dữ liệu bằng tay

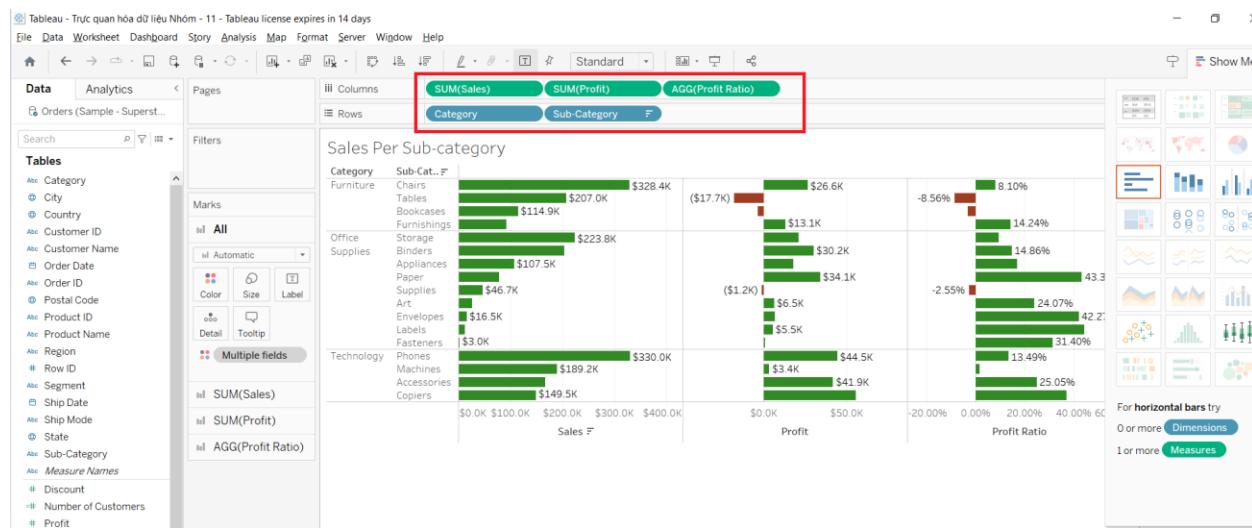


Hoặc có thể được thiết lập bằng 2 phương pháp sau:

- Phương pháp 1: Sử dụng các đoạn mã tùy chỉnh để thiết lập đường truyền dữ liệu theo thời gian thực của Tableau:
  - Tableau Desktop Auto-Refresh
  - Tableau Server Auto-Refresh
- Phương pháp 2: Sử dụng dữ liệu Hevo để thiết lập truyền dữ liệu thời gian thực của Tableau

## 4. Data Blending

Data Blending là việc trộn các dữ liệu với nhau. Với các công cụ khác, công việc này tốn khá nhiều thời gian và công sức. Tuy nhiên với Tableau, chúng ta chỉ cần kéo thả các Measures và Table vào Worksheet và Tableau sẽ đảm nhiệm công việc Data Blending cho chúng ta. Trong ví dụ về Dashboard của nhóm cũng đã có sử dụng tính năng này.



Dữ liệu của các Measures và các table đã được trộn với nhau chỉ thông qua việc kéo thả, cho ra kết quả là một trực quan dữ liệu như hình.

## 5. Advanced Visualization

- Thông qua thư viện hình ảnh, biểu đồ phong phú, đa dạng. Tableau cho phép người dùng trực quan hóa những con số phức tạp, chinh phục được những dữ liệu “khó tính” nhất với những đồ thị sinh động, đầy màu sắc.
- Tableau sở hữu nhiều loại biểu đồ khác nhau như: biểu đồ tròn, biểu đồ cột, biểu đồ cây, gantt chart,... Bên cạnh đó, bạn có thể chuyển đổi các dạng biểu đồ bằng những thao tác cực kỳ đơn giản. Đây được xem là một trong những tính năng nổi bật nhất của Tableau.



The screenshot shows the Tableau desktop application. On the left, there's a sidebar titled 'Tables' listing various data sources like Category, City, Country, etc. The main workspace displays a summary card with metrics: Number of Cust... (793), Profit (\$286.4K), Profit Ratio (12.47%), and Sales (\$2,297.2K). To the right of the workspace is a 'Show Me' panel containing a grid of icons representing different chart types such as bar charts, line graphs, and scatter plots. A tooltip for horizontal bars suggests using dimensions and measures.

Tính năng này nằm bên phải trên giao diện của Tableau và ở trong tab Show Me. Bạn có thể chọn biểu đồ, bảng text, scatter plots, ... Tableau đều có hết.

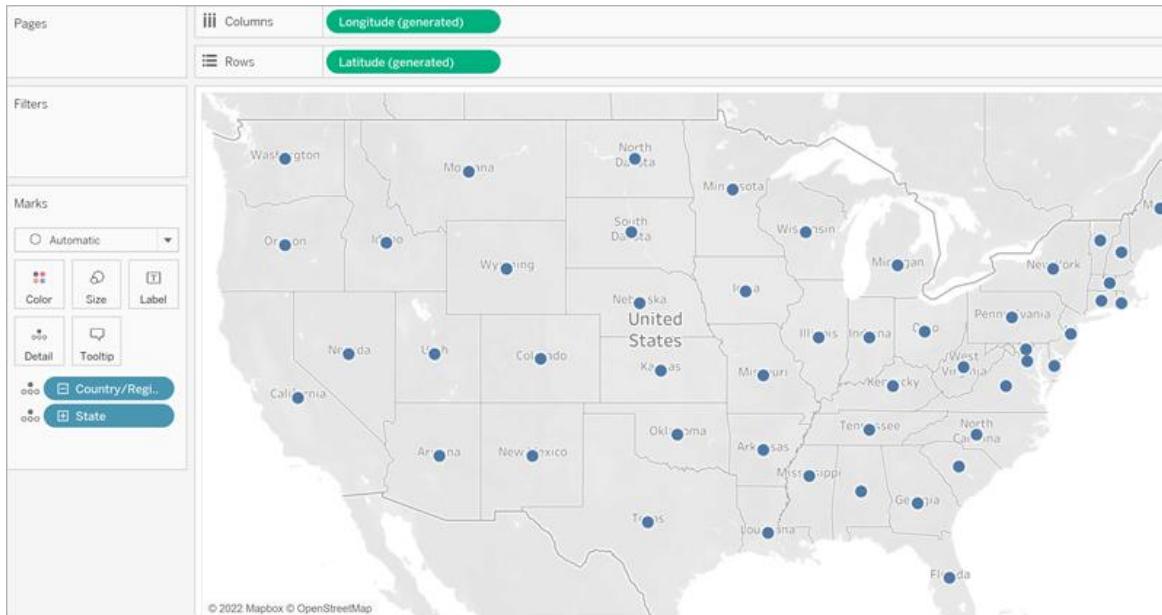
## 6. Map

- Bên cạnh những biểu đồ và đồ thị, Tableau còn có thể giúp nhân sự doanh nghiệp đánh giá xem thời tiết, ảnh hưởng như thế nào đến doanh số bán hàng trong khu vực hoặc thống kê các luật và quy định địa phương, dựa trên **tính năng Lập bản đồ** vô cùng thú vị và mạnh mẽ.
  - Xây dựng một bản đồ điểm đơn giản

B1: Navigate to a worksheet.

B2: Trong ngăn Data, mở mục Vị trí và nhấn vào State

- Ta có kết quả sau:



## 7. Security

- Tableau tuân theo những phương pháp bảo mật dữ liệu tốt nhất, bao gồm mã hóa, bảo mật HTTPS, xác thực tài khoản, mật khẩu được bảo vệ bằng mật mã, xác thực đa yếu tố và khóa sau khi đăng nhập không thành công. Điều này giúp cho doanh nghiệp loại bỏ được lo lắng bị xâm phạm hay Hack dữ liệu.
  - Tableau áp dụng hệ thống phân quyền và xác thực có sẵn. Điều này giúp giảm tình trạng mất dữ liệu. Bên cạnh đó, phần mềm này còn tận dụng “sức mạnh” của giao thức bảo mật khác từ desktop như: Kerberos, Active Directory,...
  - Giả sử bạn có dữ liệu sau đây cho biết tỷ lệ phần trăm số người được tiêm phòng đầy đủ ở các quốc gia khác nhau.

## Sheet 1

Country	
Australia	75.31
Belarus	30.35
Brazil	30.35
Cameroon	2.25
Canada	76.82
Egypt	16.85
India	38.03
Kenya	6.08
UK	68.68
Ukraine	29.35
US	60.63
Venezuela	34.34

- Bạn có thể tạo user filter và mapping user theo cách thủ công với các quốc gia mà họ nên xem. Để tạo user filter, hãy làm theo các bước sau:

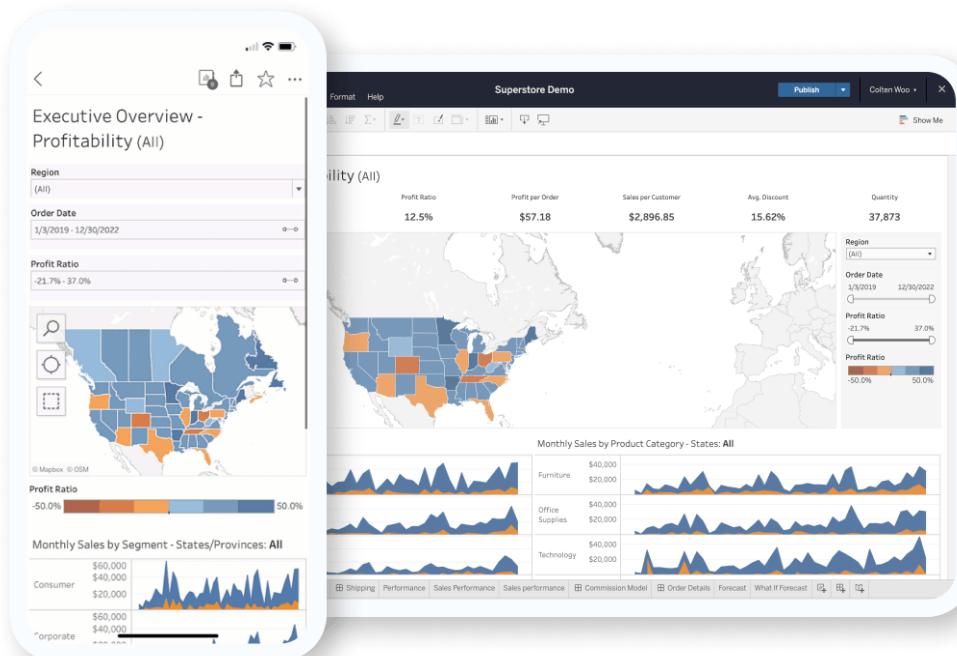
**B1:**Nhập vào Thanh Menu và chọn “Server”

**B2:**Chọn **tạo user filter** nhằm cung cấp danh sách các trường có sẵn nơi bạn có thể tạo biểu mẫu bộ lọc người dùng.

**B3:**Chỉ định các user vào khu vực thích hợp.

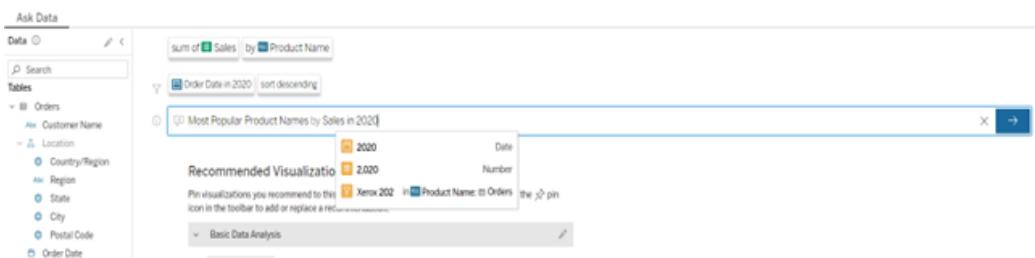
## 8. Mobile View

- Nhằm giúp cho nhân viên văn phòng không cần phải “dán mắt” vào màn hình máy tính quá lâu. Các nhà phát triển của Tableau đã xây dựng nền tính năng Mobile View, cho phép người dùng truy cập và xem Dashboard dữ liệu ngay trên chiếc điện thoại thông minh của mình.

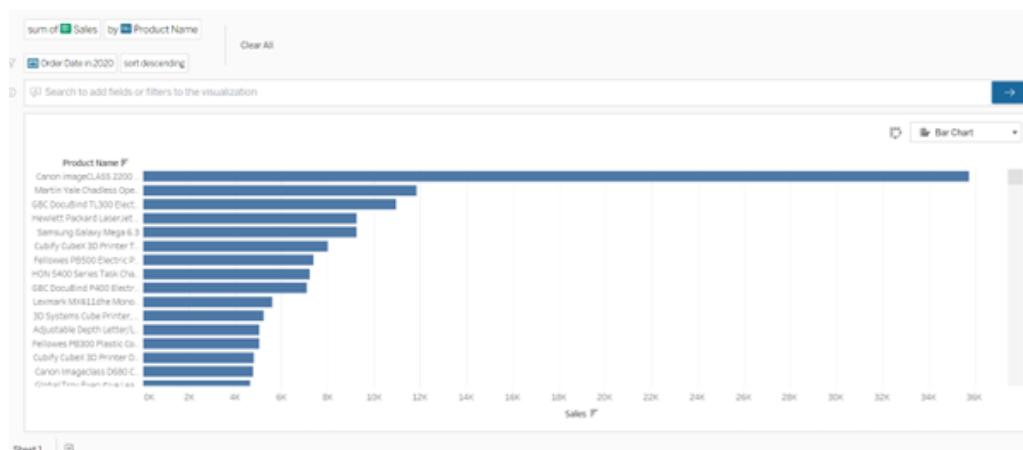


## 9. Ask Data

- Một tính năng mạnh mẽ, siêu tiện lợi của nền tảng Kinh doanh thông minh này là Ask Data – tìm kiếm dữ liệu. Người dùng chỉ cần nhập một truy vấn tìm kiếm, Tableau sẽ tự động trả lời bằng hình ảnh, nó có thể hiểu cả những câu hỏi phức tạp và hỗ trợ các khái niệm mô tả, dự đoán, phân tích. Ask Data thậm chí hiểu được các câu hỏi về thời gian, chẳng hạn như “sớm nhất”, “quý trước”. Ngoài ra, còn cả những câu hỏi về định lượng, “bán chạy nhất”, “phổ biến nhất”, hay những câu hỏi phân tích không gian, “ở đâu”.
- Tìm kiếm dữ liệu “**Most Popular Product Names by Sales in 2020.**”



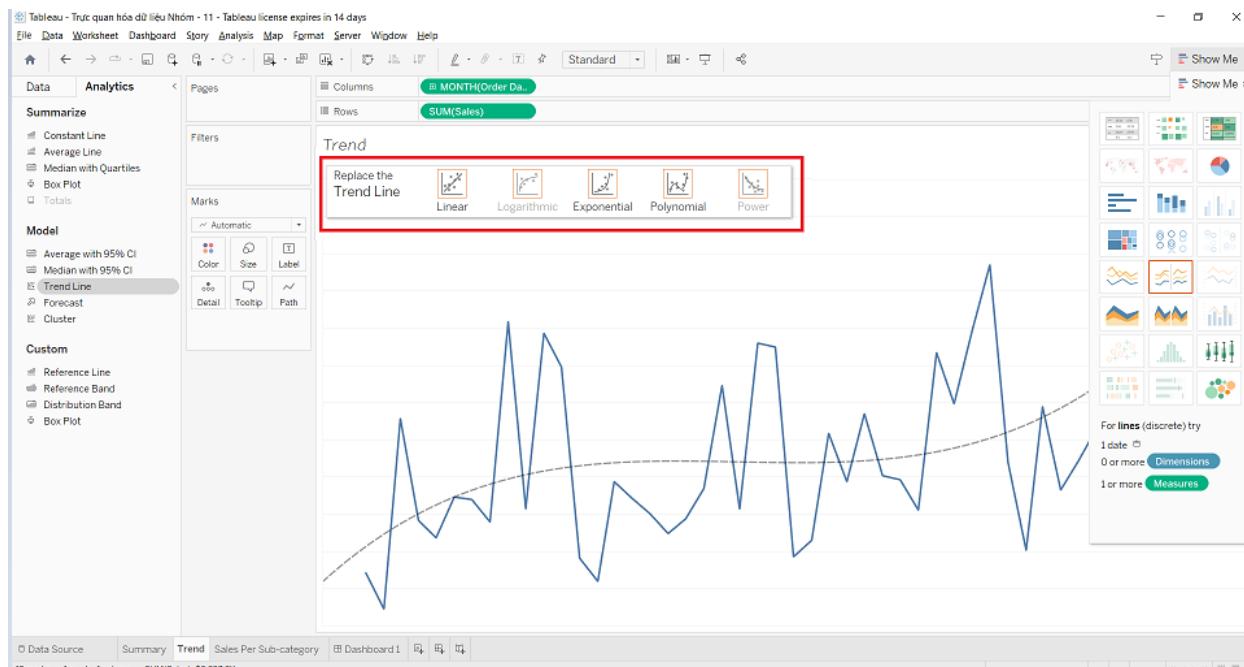
- Nhấn enter, cho ta kết quả bên dưới



Lưu ý: Ask Data feature chỉ hoạt động cho các nguồn dữ liệu đã xuất bản. Bất kỳ ai có giấy phép Tableau – Người sáng tạo, Người khám phá và người xem đều có thể sử dụng tính năng trên.

## 10. Trend Lines and Predictive Analysis

- Tableau có chức năng tạo đường Trendline phục vụ cho việc dự đoán xu hướng.
- Trong Tableau, chúng ta có thể có các đường xu hướng thẳng hoặc cong tùy thuộc vào mô hình bạn chọn từ các tùy chọn khác nhau của các mô hình đường xu hướng có sẵn trong Tableau. Tableau có tổng cộng năm loại đường xu hướng: **Linear, Exponential, Logarithmic, Polynomial, and Power trend line**. Mô tả chi tiết về tất cả các đường xu hướng này được đưa ra trong các phần tiếp theo.

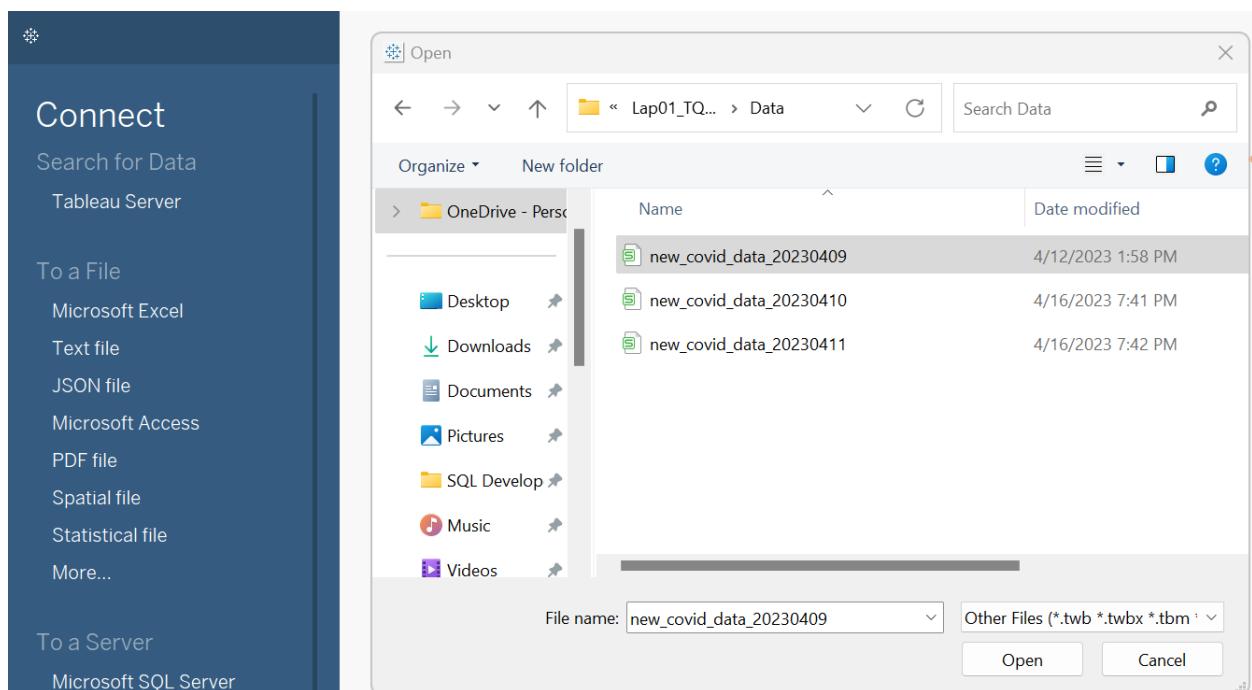


Ở ví dụ về Dashboard nhóm cũng đã sử dụng tính năng này để vẽ đường Trendline phục vụ cho Dashboard. Vào Tab Analytics sau đó kéo và thả Trendline vào Worksheet rồi chọn loại Trendline.

### III. Trực quan hóa bằng Tableau

#### A. Kết nối dữ liệu nhiều ngày

- Để kết nối Dữ liệu thống kê từng ngày ca nhiễm virus Covid-19 từ tổ chức Worldometer, nhóm sử dụng Union của Tableau
- Bước 1: Tại mục Connect → To a file → Chọn More để connect file dữ liệu đầu tiên



- Bước 2: Tại giao diện Data Source, nhấp chuột phải vào File dữ liệu để chọn chế độ Convert to Union

The screenshot shows the Tableau Data Source interface. On the left, there's a sidebar with 'Connections' and 'Files'. Under 'Connections', 'new\_covid\_data\_20230409' is selected. A context menu is open over this connection, with the 'Convert to Union...' option highlighted in blue. The main pane displays the connection details: 'new\_covid\_data\_20230409' with 16 fields and 223 rows. To the right, a preview table shows data for four countries: USA, India, France, and Germany, with columns for Country, Total Cases, and New.

- Bước 3: Kéo thả các file còn lại vào danh sách Union để nối dữ liệu

The screenshot shows the Tableau Data Source interface. A context menu is open over the 'new\_covid\_data\_20230409' connection, with the 'Convert to Union...' option highlighted. A modal dialog titled 'new\_covid\_data\_20230409.csv+' is displayed, showing three files selected for union: 'new\_covid\_data\_20230409.csv', 'new\_covid\_data\_20230410.csv', and 'new\_covid\_data\_20230411.csv'. At the bottom of the dialog, it says 'Tables in union: 3'. There are 'Apply' and 'OK' buttons at the bottom right.

- Bước 4: Sử dụng Calculated Field để tạo một cột dữ liệu Date thay cho cột Table Name để phân biệt dữ liệu thuộc ngày nào



Sử dụng hàm tính toán sau để ánh xạ ngày vào dòng dữ liệu:

```
DATE(DATEPARSE('MM/dd/yyyy', IF [Table Name] = 'new_covid_d THEN '09/04/2023' ELSEIF [Table Name] = 'new_covid_data_20230410.csv' THEN '10/04/2023' ELSEIF [Table Name] = 'new_covid_data_20230411.csv' THEN '11/04/2023' END))
```

- Bước 5: Chuyển kiểu dữ liệu của cột Date vừa thêm được sang dạng date để sử dụng cho công việc trực quan sau này

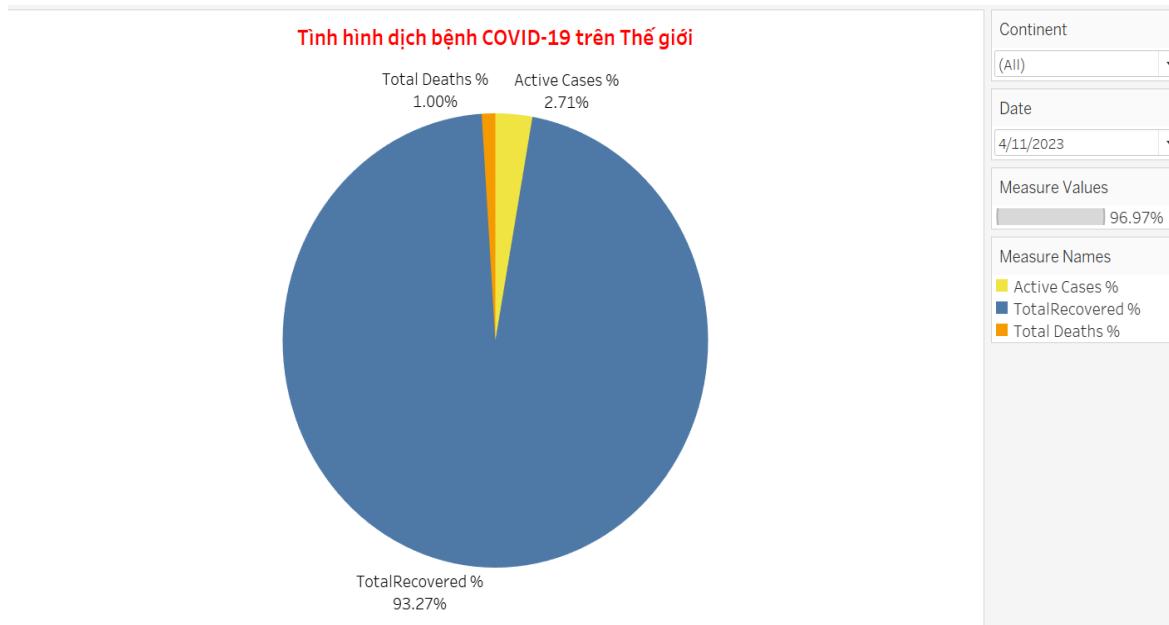
Nhập vào biểu tượng Abc để thay đổi kiểu dữ liệu

## B. Sử dụng biểu đồ để trực quan dữ liệu COVID-19

### 1. Xem xét tình hình số ca nhiễm đang điều trị COVID chưa khỏi bệnh / số ca tử vong / số ca hồi phục trên thế giới

#### Pie Chart

Biểu đồ biểu diễn số ca nhiễm đang điều trị, số ca tử vong và số ca phục hồi do COVID-19 gây ra trên thế giới tính đến ngày 11/4/2023





*Ý nghĩa trực quan:* Như đã phân tích ở Lap 1, Total Death, Active Cases, Total Recovered là các số liệu biểu diễn tình trạng của các ca nhiễm trên thế giới. Vì vậy nhóm sử dụng biểu đồ tròn để thể hiện phần trăm của từng tình trạng ca nhiễm, mỗi phần được phân biệt với nhau qua màu sắc hiển thị.

### *Ý nghĩa sử dụng màu sắc:*

- Đối với dữ liệu Total Recovered cho biết số ca nhiễm đã hồi phục, biểu diễn sự tích cực nên nhóm sử dụng màu xanh dẽ nhìn.
- Đối với Dữ liệu Active Cases cho biết số ca nhiễm đang còn điều trị, là dữ liệu đáng quan tâm khi quan sát tình hình dịch bệnh, nhóm sử dụng màu vàng để làm nổi bật dữ liệu, và đặc biệt là tương phản với màu của dữ liệu Total Recovered
- Đối với Total Deaths, dữ liệu báo động về tình trạng tử vong do COVID-19 được biểu diễn bằng màu nóng và đậm hơn để dễ dàng phân biệt với 2 loại dữ liệu trên.

**Nhận xét:** Nhìn vào biểu đồ ta đã có thể nhận xét rằng tình hình dịch Covid19 hiện nay gần như đã được kiểm soát hoàn toàn khi tỉ lệ hồi phục chiếm tới 93.27%, tỉ lệ đang điều trị là 2.71% và tỉ lệ chết chỉ có 1.00%. Tỉ lệ các ca bị null cũng chỉ có 3.02% nên dù số ca này thuộc nhóm nào thì cũng không phủ nhận được việc dịch Covid 19 đã hoàn toàn được kiểm soát.

### *Hướng dẫn thực hiện trên Tableau:*

B1: tạo các calculate field cho ActiveCase, Total Deaths, Total Recovered

Total Deaths %

`SUM([Total Deaths]) / SUM([Total Cases])`

Active Cases %

`SUM([Active Cases]) / SUM([Total Cases])`

TotalRecovered %

`SUM([Total Recovered]) / SUM([Total Cases])`

B2: Filter các measure name ở trên, chuyển đổi biểu đồ thành pie, cho các màu cho các measure trên

**Filters**

**Measure Names**

**Marks**

Pie

Color      Size      Label

Detail      Tooltip      Angle

**Measure Names**

**Measure Values**

**Measure Values**

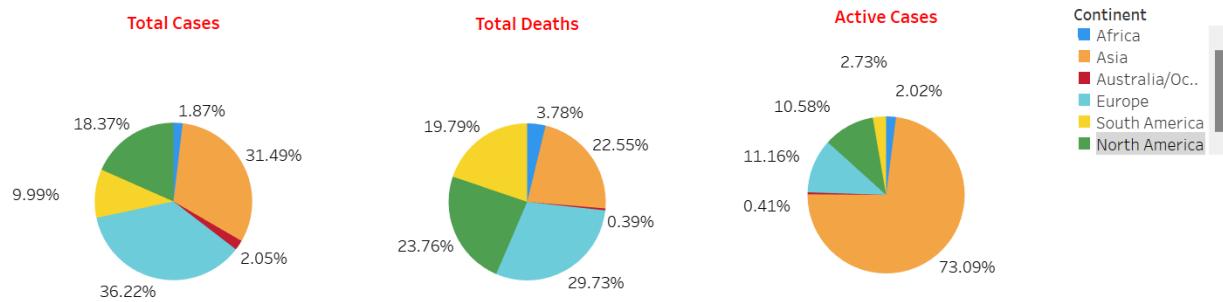
**Measure Values**

AGG(Active Cases %)

AGG(TotalRecovered %)

AGG(Total Deaths %)

### Biểu đồ phần trăm số ca nhiễm COVID-19, số ca tử vong, số ca đang điều trị của từng Châu Lục trên thế giới tính đến ngày 11/4/2023



**Ý nghĩa trực quan:** Dữ liệu ca nhiễm của thế giới có thể được chia theo từng Châu Lục, để quan sát tình hình dịch bệnh cụ thể của từng Châu Lục, nhóm sử dụng 3 biểu đồ tròn biểu diễn tỉ lệ chiếm phần trăm Số ca nhiễm, số ca tử vong, số ca điều trị của từng Châu Lục so với thế giới. Trong đó, số ca tử vong và số ca điều trị là 2 dữ liệu đáng quan tâm nên được đặt chung với biểu đồ biểu diễn số ca nhiễm như trên.

**Ý nghĩa màu sắc:** Mỗi Châu Lục sẽ có màu sắc khác nhau và tương phản nhau, không có sự giống nhau quá nhiều để dễ dàng phân biệt các Châu Lục khác nhau

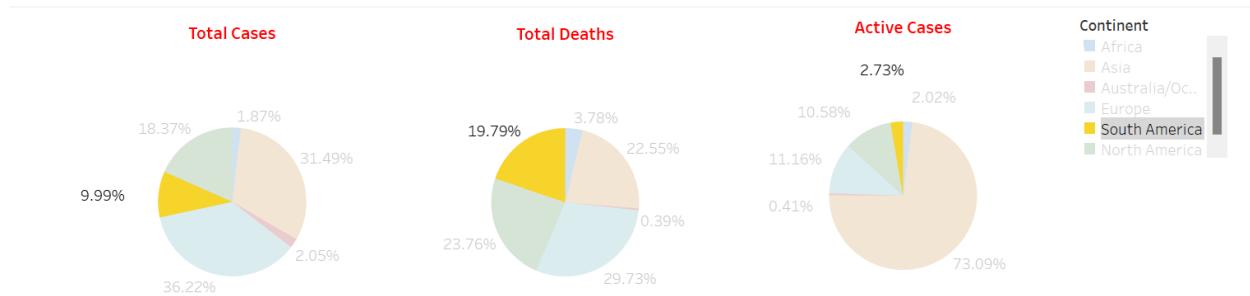
**Nhận xét:** Với sự lựa chọn màu sắc dễ nhìn cũng như việc bố trí các biểu đồ tròn nằm kề nhau, ta có thể dễ dàng nhận thấy một số điểm mấu chốt như:

- North America có số lượng ca đang nhiễm cao nhất, chiếm đến 73.09% trên toàn thế giới. Con số này cho ta biết rằng khu vực này đang lơ là trong việc kiểm soát bệnh dịch dẫn đến số ca nhiễm tăng cao. Tuy nhiên số ca tử vong và tổng số ca nhiễm của khu vực lại không quá cao, chứng tỏ trong quá khứ họ đã kiểm soát dịch bệnh khá tốt.
- Đối với các khu vực khác, số ca đang nhiễm là khá ít do hầu hết mọi khu vực đều đã kiểm soát được dịch bệnh.
- Tỷ lệ tử vong / tổng số ca bệnh của các khu vực là không đồng đều do nhiều yếu tố như y tế hay sự gắt gao trong công tác phòng chống dịch.
- Dữ liệu bệnh dịch giữa các khu vực đều có sự tương quan nhất định mà khi nhìn vào biểu đồ tròn ta có thể dễ dàng nhận thấy.

### Các kỹ thuật sử dụng:

- Select element:

Khi quan sát biểu đồ, nếu chỉ quan tâm số liệu của một Châu Lục thay vì xem tất cả Châu Lục trên thế giới, người dùng có thể chọn duy nhất 1 Châu Lục tại Highlight Selected Element



### Hướng dẫn thực hiện trên Tableau:

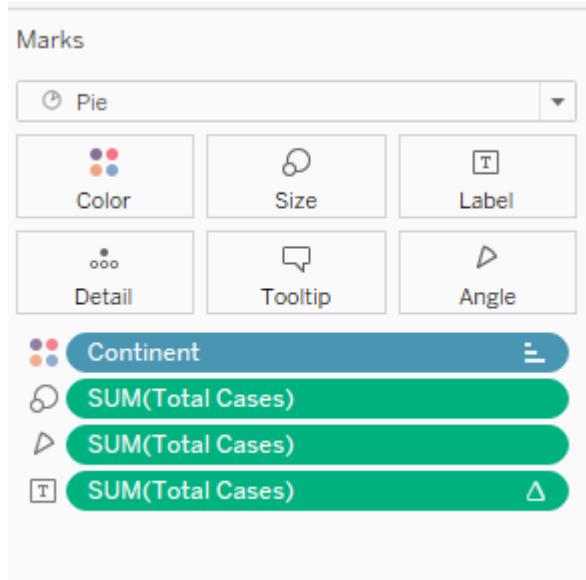
B1:Như biểu đồ trên, ta sẽ tạo biểu đồ Pie, ở Mark các châu lục sẽ được đánh dấu bằng Color

B2: Các size , Angle, Label ta sẽ đánh dấu một trong những measure name sau:

- Total cases

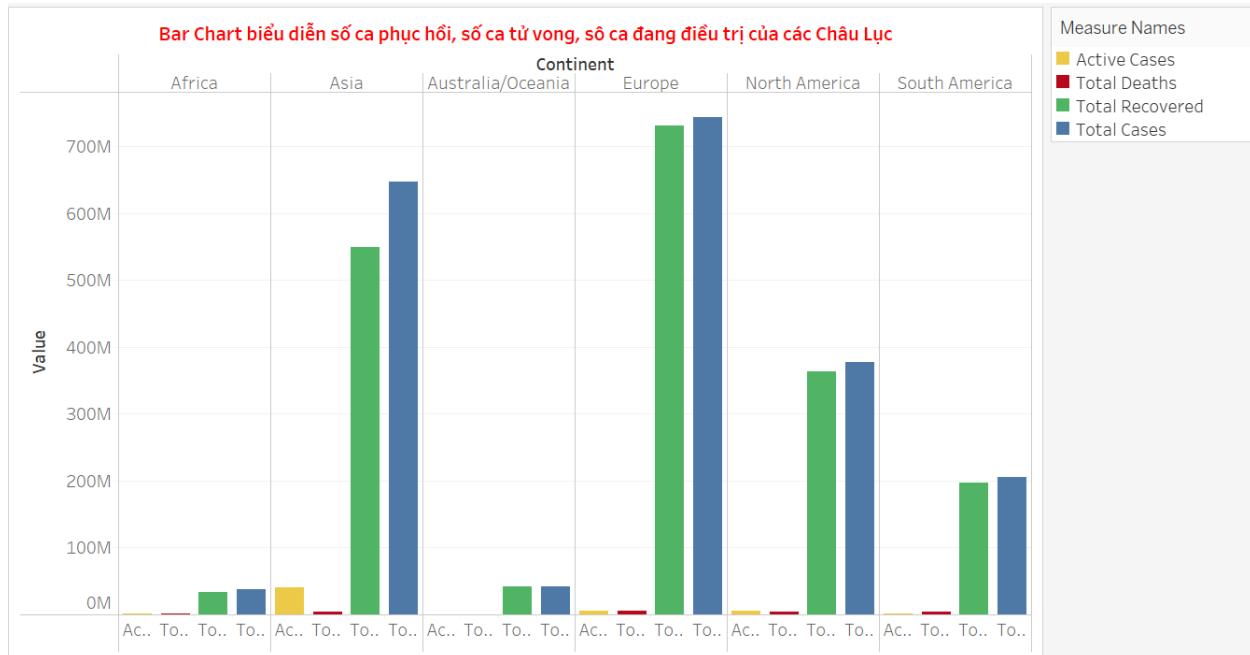
- Total deaths
- Total recovered

Ví dụ:



### Side by side Bar Chart

**Biểu đồ biểu diễn số ca phục hồi, số ca tử vong, số ca đang điều trị của các Châu Lục**



**Ý nghĩa trực quan:** Sau khi quan sát tỉ lệ ca nhiễm, ca tử vong, ca điều trị của các Châu Lục đối với thế giới, nhà phân tích có thể quan tâm đến số liệu cụ thể ca nhiễm đó của từng Châu Lục, vì vậy Bar Chart phù hợp biểu diễn nhiều cột dữ liệu trên cùng một biểu đồ giúp dễ dàng so sánh về số ca nhiễm, số ca phục hồi, số ca tử vong và số ca hiện đang nhiễm giữa các khu vực. Bên cạnh đó việc so sánh giữa các chỉ số trong một khu vực cũng dễ dàng.

### **Ý nghĩa màu sắc:**

- Tương tự như giải thích về Dữ liệu ca đang điều trị, ca tử vong ở trên nên nhóm sử dụng màu vàng và đỏ tương ứng cho 2 loại dữ liệu đó.
- Nhóm sử dụng màu xanh dương - màu thông dụng để biểu diễn số liệu chính là Số ca nhiễm.
- Màu xanh lá để biểu diễn Dữ liệu ca hồi phục mang hướng tích cực.

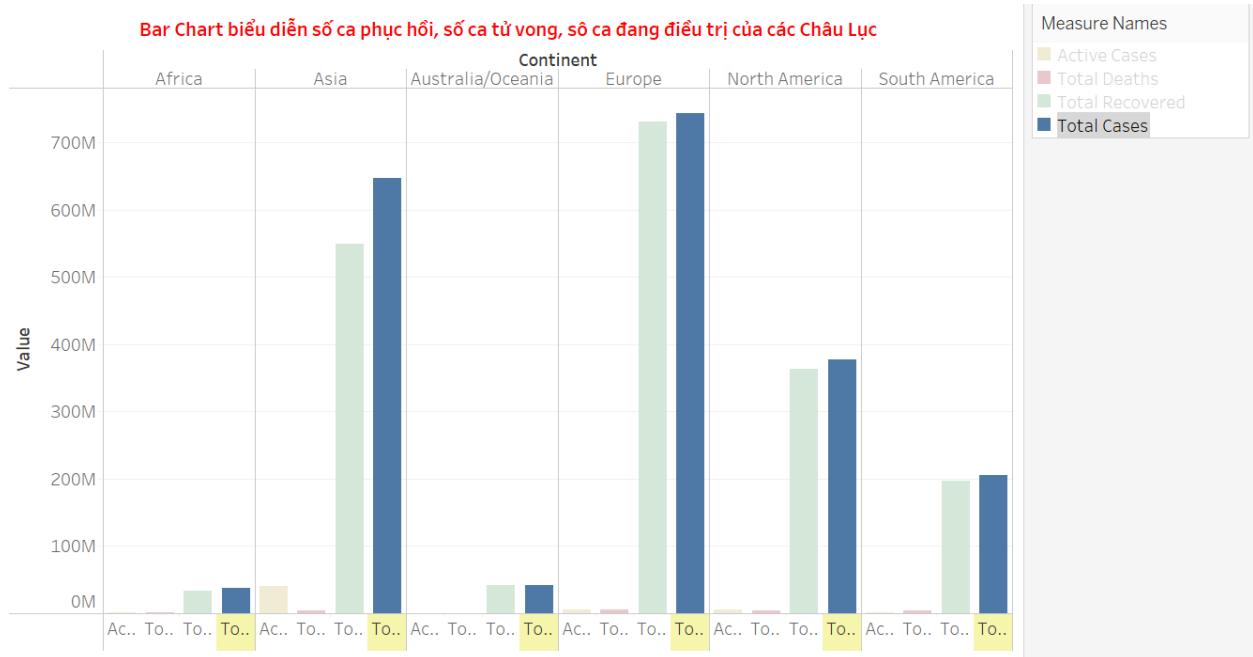
### **Nhận xét:**

- Dựa vào biểu đồ trên, ta có thể những điểm chú ý sau:
  - Europe: có số lượng ca nhiễm lớn nhất hơn 700M ca nhiễm. Với con số này, cho thấy Châu Âu ảnh hưởng nặng về dịch bệnh Covid19. Nhưng nhìn về mặt tích cực, ta có thể thấy số lượng ca phục hồi cũng cao nhất so với các châu lục khác. Điều này chứng tỏ, Các nước Châu Âu đã đề cao cảnh giác về dịch bệnh Covid hiện nay và huy động phòng chống dịch bệnh rất mạnh mẽ

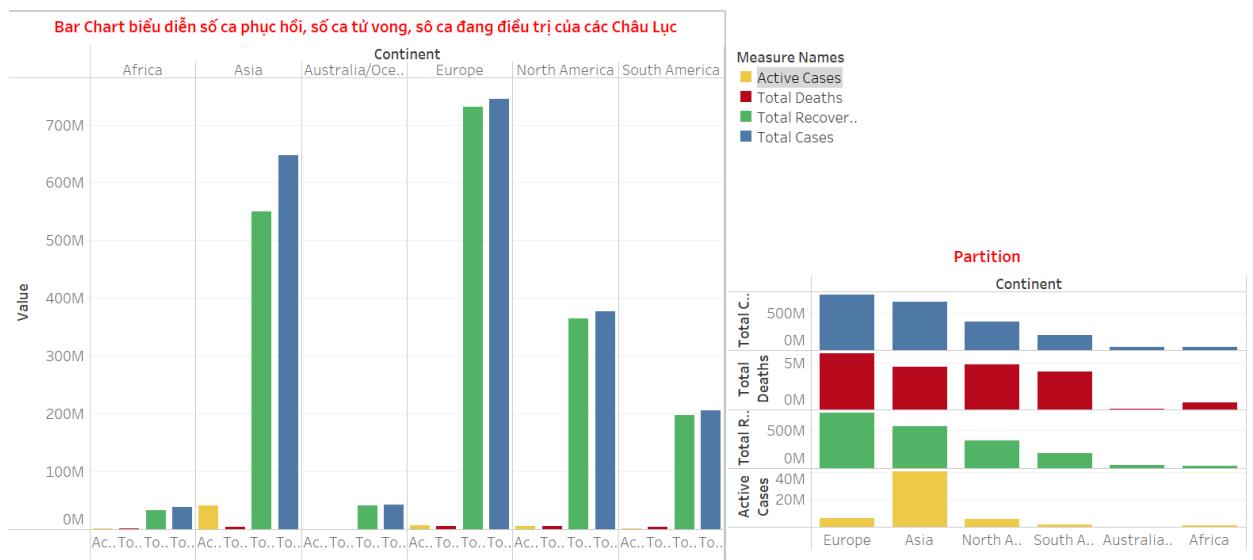
### **Các kỹ thuật sử dụng:**

- Select element:

Khi quan sát biểu đồ, nếu chỉ quan tâm số liệu của ca nhiễm thay vì xem tất cả dữ liệu có trong biểu đồ, người dùng có thể chọn duy nhất 1 trường dữ liệu duy nhất tại Highlight Selected Element



### - Partition into view



## Biểu đồ so sánh các trường dữ liệu Total Cases, Total Deaths và Total Recovered của các Continent.



Bằng cách sắp xếp thứ tự các Continent giảm dần theo Total Cases, người xem có thể dễ dàng so sánh sự liên hệ giữa các trường dữ liệu.

### Nhận xét:

- Dựa vào biểu đồ trên ,đã cho ta có cái nhìn tổng quan về tình hình dịch bệnh Covid của từng châu lục:
  - Điểm đáng chú ý nhất là Europe có Total Cases, Total Deaths và Total Recovered cao nhất. Với cái nhìn đó, đã cho thấy Châu Âu từng không quan tâm về dịch bệnh nên dẫn đến nhiều ca nhiễm và nhiều ca chết thế giới. Nhưng hầu hết các nước Châu Âu đều có cơ sở y tế chất lượng nên cho ta thấy số ca phục hồi cao nhất
  - Điểm đáng chú ý thứ 2 là Asia, ta có thể thấy Châu Á không cao so với Châu Âu nhưng Châu Á lại có Active Case cao nhất. Điều đó dẫn ta đặt câu hỏi rằng: “Liệu điều đó có đáng lo ngại cho các nước Châu Á không?”

### Hướng dẫn thực hiện trên Tableau:

Để biểu diễn dữ liệu các châu lục về Active Cases, Total Deaths, Total Recovered bằng biểu đồ barchart



B1: Về column: ta sẽ đặt các châu lục và Measure Name (Filter các giá trị sau:Active Cases, Total Deaths, Total Recovered)

B2: Về row, ta sẽ để các giá trị của measure value

B3: Để các measure name có giá trị màu khác nhau, ta đánh dấu measure name vào phần Color

The screenshot shows the Tableau Data Source interface with three main sections:

- Filters:** Contains dropdowns for "Continent" and "Measure Names".
- Marks:** Shows settings for "Automatic" mark type, with options for "Color", "Size", "Label", "Detail", and "Tooltip". A button labeled "Measure Nam.." is also present.
- Measure Values:** Lists measures: "SUM(Active Cases)", "SUM(Total Deaths)", and "SUM(Total Recovere..)".

## Partition



### *Hướng dẫn thực hiện trên Tableau:*

B1: Vẽ column, ta đặt các châu lục vào. Vẽ rows, ta đặt các giá trị về Active Cases, Total Recovered, Total Deaths, Total Cases

B2: Filter và marks các giá trị sau:

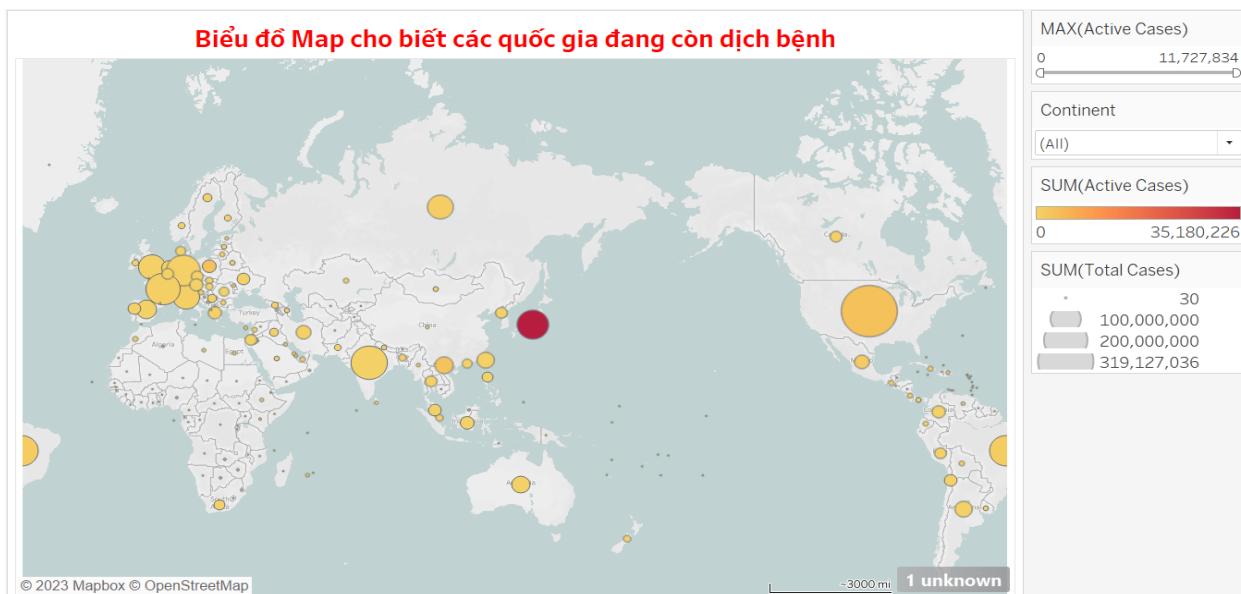


The screenshot shows a software interface for data visualization. On the left, there's a sidebar titled 'Filters' with a dropdown arrow. Below it is a teal button labeled 'Measure Names'. The main area is titled 'Marks' and has a sub-section 'All'. Under 'All', there's a dropdown menu set to 'Automatic'. Below this are four buttons: 'Color' (with a color palette icon), 'Size' (with a magnifying glass icon), 'Label' (with a text icon), 'Detail' (with a grid icon), and 'Tooltip' (with a speech bubble icon). Another teal 'Measure Nam..' button is located below the 'Detail' and 'Tooltip' buttons. The bottom section lists several measures: 'SUM(Total Cas...)', 'SUM(Total Dea...)', 'SUM(Total Rec...)', and 'SUM(Active Ca...)'.

## 2. Quan sát tình hình dịch bệnh hiện tại của các Quốc gia trên thế giới

### Map

Biểu đồ map biểu diễn số ca nhiễm hiện còn đang điều trị của các Quốc gia trên thế giới



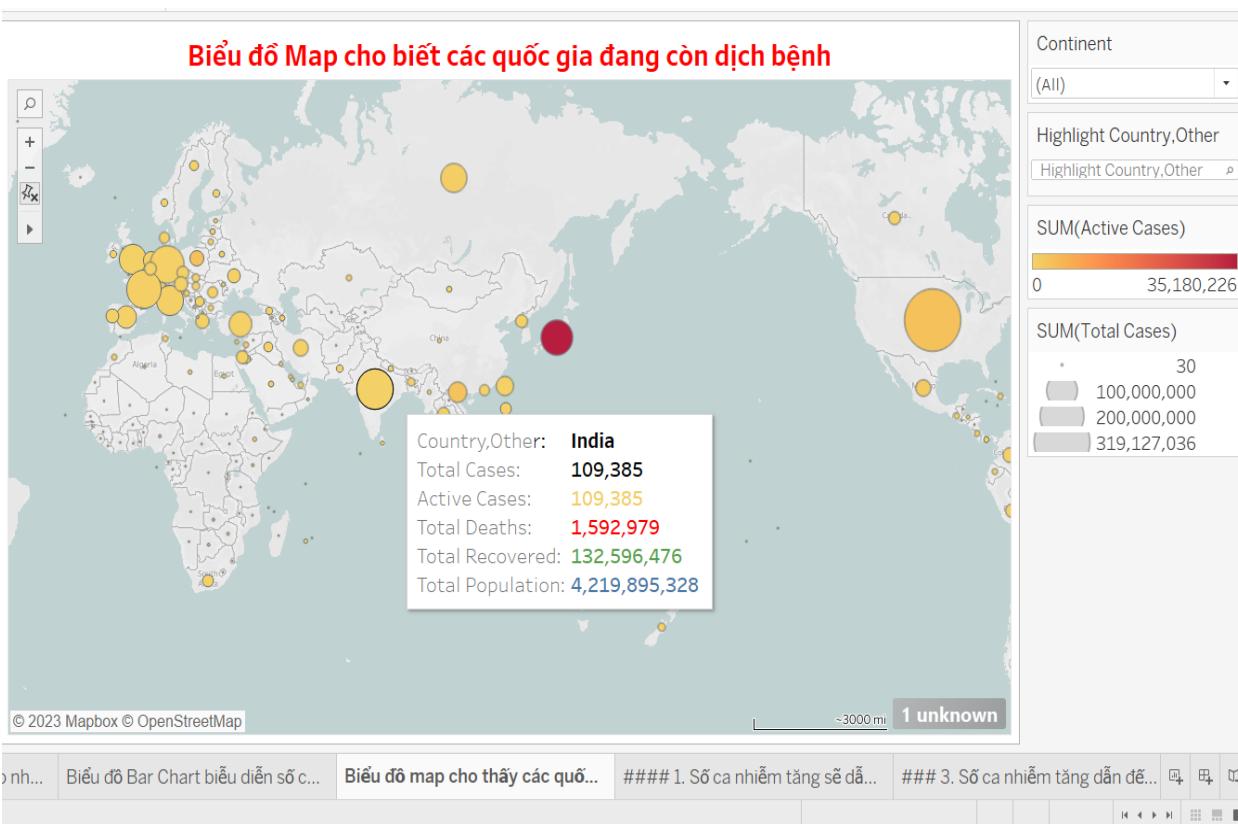
### Ý nghĩa trực quan:

Biểu đồ map mang đến cái nhìn khác về tình hình dịch bệnh, không chỉ tập trung vào số ca nhiễm và tình trạng ca nhiễm, Map cho biết vị trí địa lý của các điểm quốc gia có dịch bệnh và biểu diễn chúng với các dạng khác nhau. Nó sẽ giúp ta hình dung ngay được rằng ở khu vực nào mà tình hình dịch đang chuyển biến xấu hay khu vực nào đã hoàn toàn kiểm soát được dịch mà không cần phải đọc đến các con số hay dữ liệu nào. Đây có lẽ là cách tốt nhất và dễ hiểu nhất để biểu diễn sự phân bố tình hình dịch bệnh hiện nay trên thế giới.

Ở đây nhóm sử dụng điểm dữ liệu để có thể dễ dàng nhìn thấy vị trí quốc gia có dịch bệnh và các đặc điểm khác như số ca đang được điều trị là ít hay nhiều, và số ca tử vong do COVID là thấp hay cao.

- Kích thước điểm biểu diễn phụ thuộc vào số ca điều trị cao hay thấp
- Màu sắc từ vàng sang đỏ biểu diễn số ca tử vong của quốc gia đó càng cao

Ngoài ra tại mỗi điểm dữ liệu còn cung cấp các thông tin khác của quốc gia đó như số ca phục hồi, dân số của quốc gia.



### Ý nghĩa màu sắc:

- Màu sắc của điểm dữ liệu được thay đổi từ vàng đến đỏ để biểu diễn cho sự tăng dần số lượng ca tử vong của quốc gia, số ca tử vong càng cao, càng đáng báo động vì vậy màu đỏ mang cảm giác cần được chú ý mạnh mẽ.

**Nhận xét:** Thông qua một bản đồ thế giới phẳng và dữ liệu dịch bệnh được biểu diễn bằng các hình tròn có màu sắc và độ lớn phụ thuộc vào số lượng các ca nhiễm của khu vực đó thì ta thấy rằng:

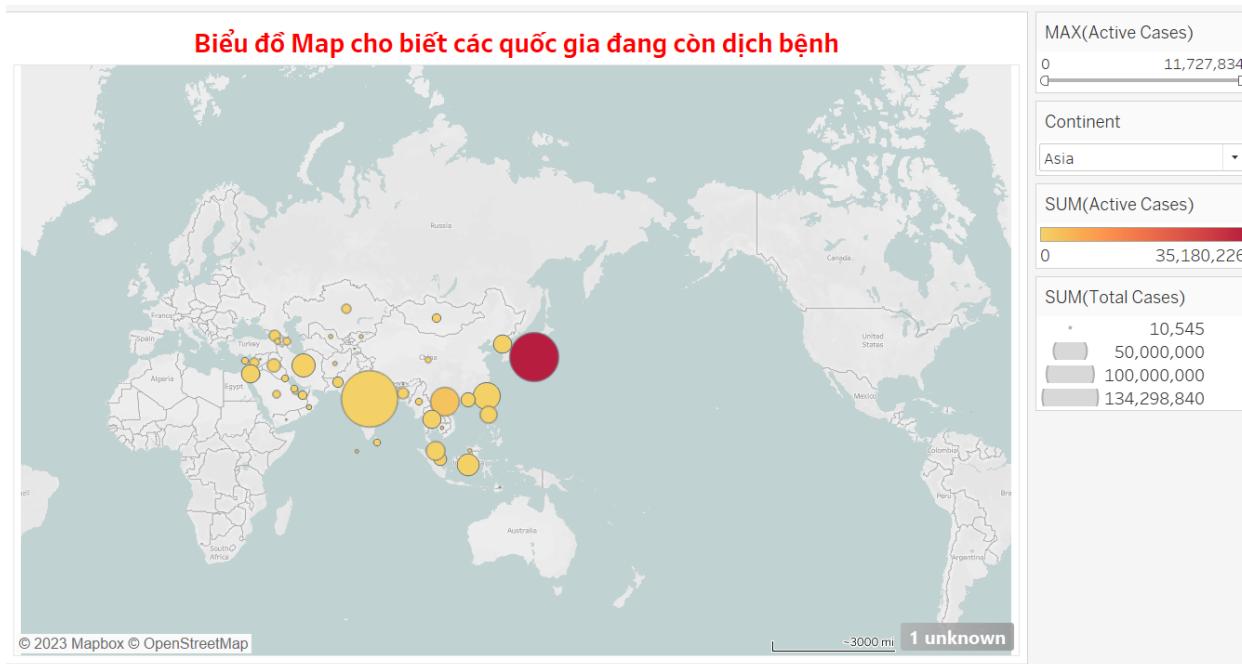
- Khu vực Châu Âu là nơi đang có diễn biến dịch bệnh khá phức tạp khi số ca tử vong rải rác khắp nơi với số lượng từ ít đến nhiều đều có đủ.
- Khu vực Bắc Mỹ cũng tương tự Châu Âu nhưng các điểm dịch ít hơn.
- Nhật Bản trở thành điểm nóng của dịch bệnh khi chỉ duy nhất Nhật Bản được thể hiện bằng màu đỏ tức là số ca tử vong cao và đáng báo động.

### Kỹ thuật sử dụng:

- Reduce items:

Với biểu đồ hiện tại người dùng có thể xem biểu đồ với dạng ít dữ liệu hơn bằng cách sử dụng filter Continent, chức năng này giúp người dùng quan sát dữ liệu của các quốc gia tại một Châu Lục cụ thể, số lượng thuộc tính không thay đổi.

;



- Reduce Attribute:

Để quan sát những quốc gia có số lượng ca nhiễm đang điều trị tại một khoảng nào đó, có thể sử dụng filter MAX(Active Cases), biểu đồ sẽ chỉ hiển thị những điểm dữ liệu thỏa điều kiện

**Biểu đồ Map cho biết các quốc gia đang còn dịch bệnh**



- Geometric Zooming:

Geometric Zooming được trong biểu đồ Map để có thể quan sát rõ từng khu vực trên biểu đồ bằng cách thu phóng biểu đồ



**Hướng dẫn thực hiện trên Tableau:**

B1: Columns và Rows:



The screenshot shows the Tableau interface with two filter panels. The top panel, titled 'Columns', has a dropdown menu and a button labeled 'Longitude (generated)'. The bottom panel, titled 'Rows', also has a dropdown menu and a button labeled 'Latitude (generated)'.

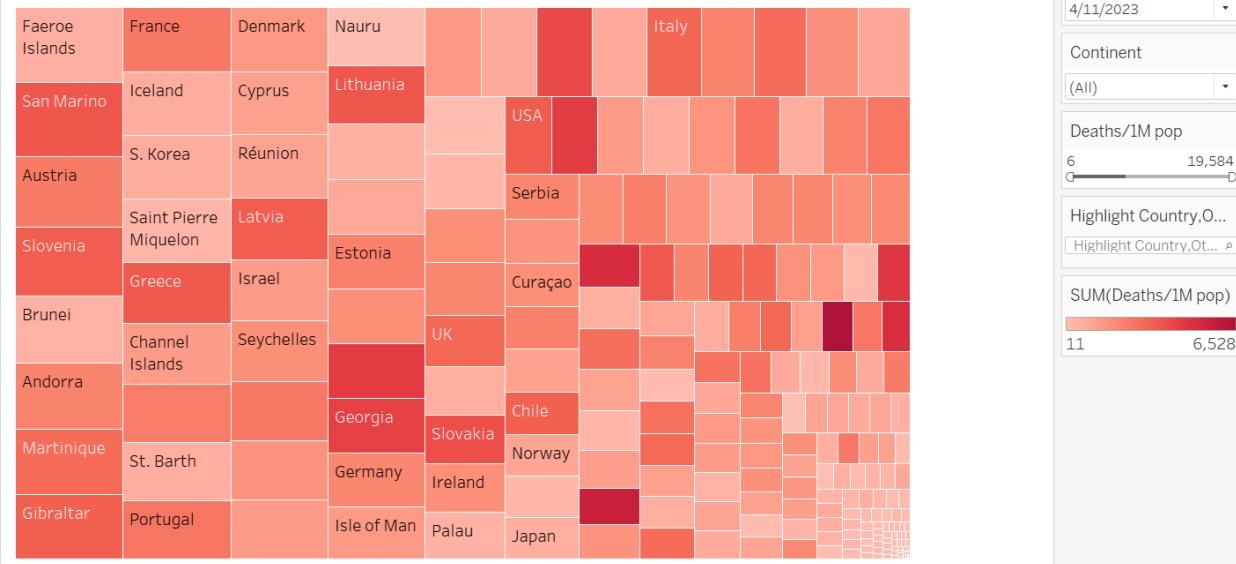
B2: Filter và Marks:

The screenshot shows the Tableau interface with the 'Marks' shelf open. It includes options for 'Automatic' and various mark types: Color, Size, Label, Detail, and Tooltip. Below these, a list of marks is shown, each with a preview icon and a tooltip. The marks listed are: SUM(Active Ca..), SUM(Total Ca..), Country,Other, Continent, SUM(Total De..), SUM(Total Re..), SUM(Active Ca..), and SUM(Total Ca..).

### TreeMaps

Biểu đồ Treemaps cho biết số ca nhiễm trên 1 triệu dân của các quốc gia trên thế giới và biểu diễn số ca tử vong trên 1 triệu dân tại mỗi quốc gia

TreeMaps biểu diễn số ca nhiễm và số ca tử vong / 1M người dân của các quốc gia trên thế giới



### Ý nghĩa trực quan:

Dữ liệu ca nhiễm trên 1 triệu dân của mỗi quốc gia cho biết tình hình diễn biến dịch bệnh của quốc gia đó nghiêm trọng như thế nào, tương tự như vậy mật độ số ca tử vong trên 1 triệu dân thể hiện sự mất kiểm soát do COVID-19 gây ra của quốc gia đó, cũng như thiệt hại về người mà quốc gia đó phải chịu. Tree map thích hợp để biểu diễn cả 2 loại dữ liệu trên trong cùng một biểu đồ mà giúp nhà phân tích dễ dàng so sánh số liệu giữa các quốc gia thông qua màu sắc và kích thước “Ô quốc gia”

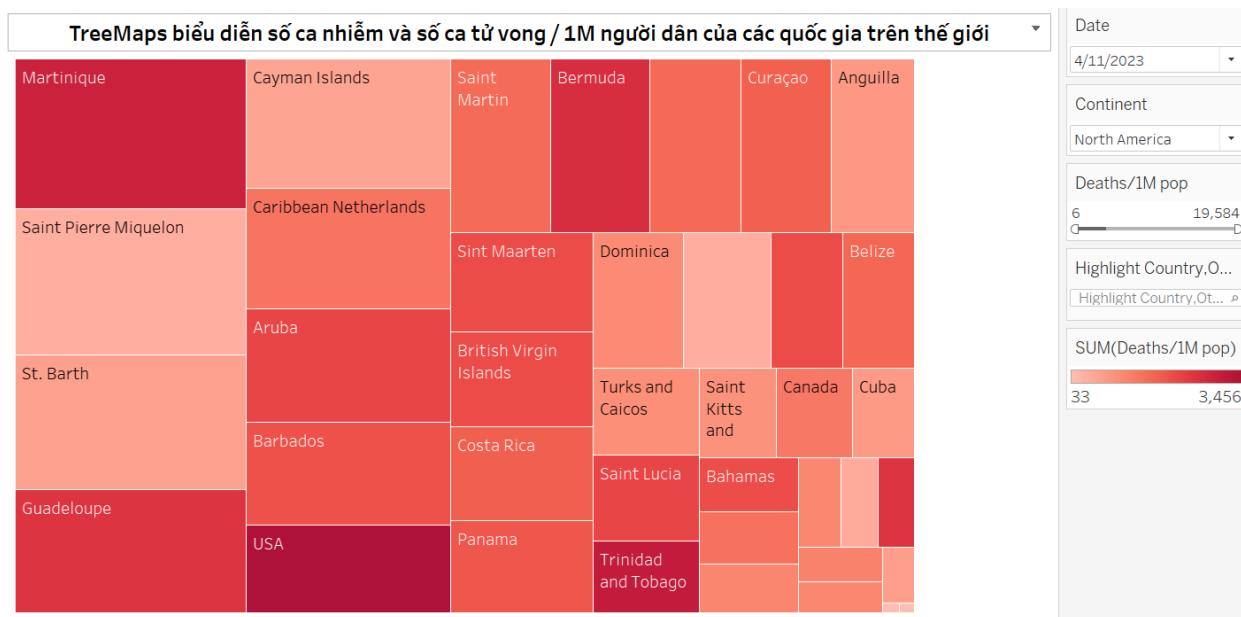
- Kích thước “Ô quốc gia” phụ thuộc vào số ca nhiễm trên 1 triệu dân của quốc gia đó
- Màu sắc “Ô quốc gia” phụ thuộc vào số tử vong trên 1 triệu dân của quốc gia đó

**Ý nghĩa màu sắc:** Sử dụng tone màu nóng để biểu diễn tính quan trọng của các số liệu trên đối với các quốc gia trên thế giới và sự nghiêm trọng do COVID gây ra

### Kỹ thuật sử dụng:

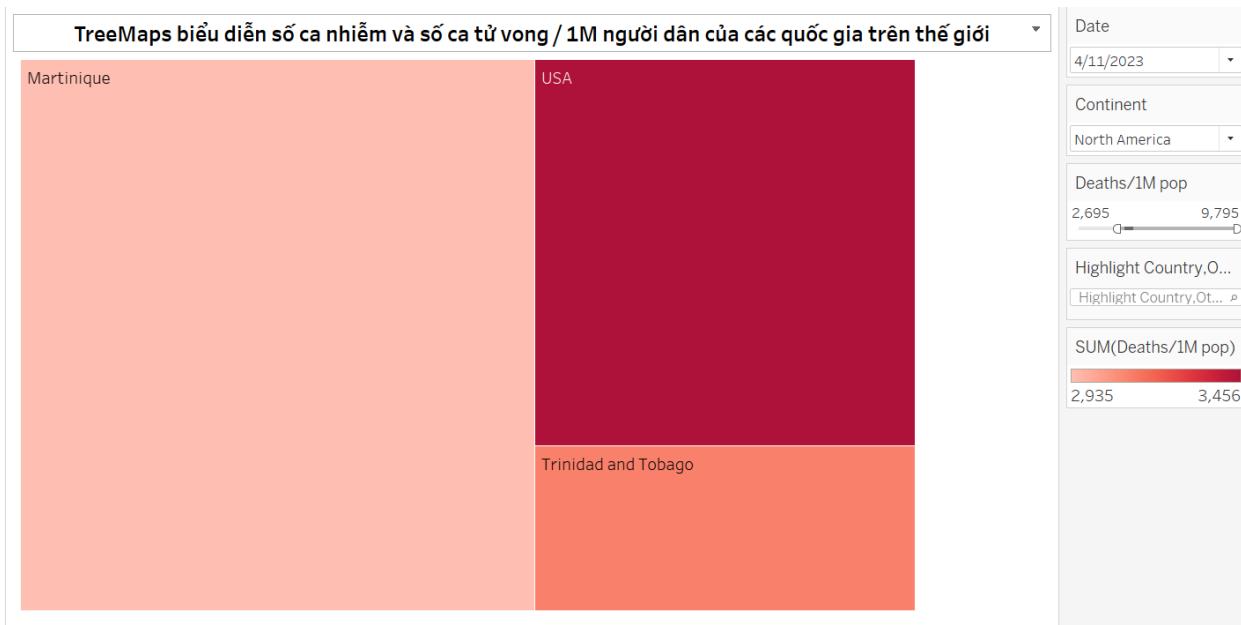
- Reduce items:

Với biểu đồ hiện tại người dùng có thể xem biểu đồ với dạng ít dữ liệu hơn bằng cách sử dụng filter Continent, chức năng này giúp người dùng quan sát dữ liệu của các quốc gia tại một Châu Lục cụ thể, số lượng thuộc tính không thay đổi.



- #### - Reduce Attribute:

Để quan sát những quốc gia có số lượng ca tử vong trên 1 triệu dân tại một khoảng nào đó, có thể sử dụng filter Death/1M pop, biểu đồ sẽ chỉ hiển thị những ô dữ liệu thỏa điều kiện

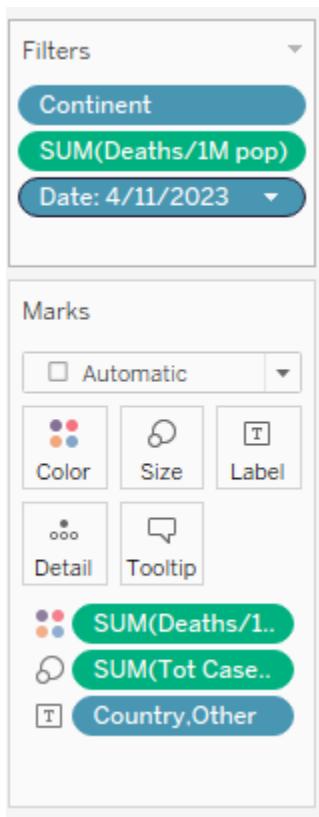


**Nhận xét:** Biểu đồ tree map vừa nhìn sẽ có vẻ hơi khó hiểu nhưng khi ta đã đọc và nắm được các chú thích thì những thông tin mà nó có thể mang đến là rất đáng kể. Ví dụ như với biểu đồ tree map ở trên, ta có thể nắm được trung bình trên 1 triệu người dân thì có bao nhiêu ca nhiễm và bao nhiêu ca tử vong. Số liệu này sẽ biểu diễn chi tiết hơn và hỗ

trợ người xem phân tích tình hình dịch bệnh tốt hơn. Để thấy Mỹ có số ca nhiễm và ca tử vong cao nhất cùng với các nước như Martinique hay Bermuda. Càng thu hẹp số lượng các nước lại ta càng thấy rõ tình hình dịch của các nước khi ở biểu đồ tree map chỉ có USA, Martinique và Trinidad and Tobago thì màu của Martinique và Trinidad and Tobago đã trở nên nhạt đi nhiều so với USA.

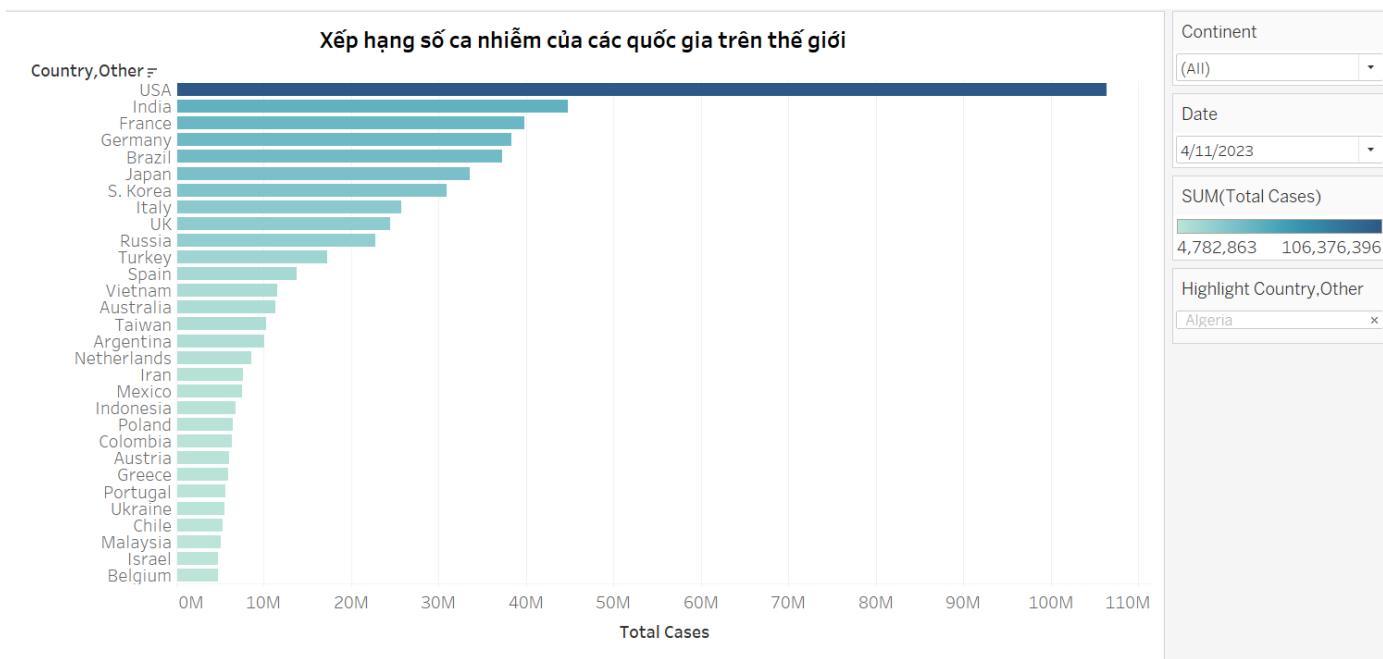
### **Hướng dẫn thực hiện trên Tableau:**

*Filter và Marks các giá trị sau:*



### **Bar chart**

**Biểu đồ biểu diễn top 30 quốc gia có số ca nhiễm cao nhất thế giới**



**Ý nghĩa trực quan:** Với số lượng các miền dữ liệu nhiều, để có thể dễ dàng so sánh, xếp hạng các miền dữ liệu đó thì sử dụng Bar chart ở dạng cột nằm ngang là tối ưu nhất. Ta có thể lựa chọn sắp xếp chúng theo thứ tự thấp đến cao hoặc cao đến thấp nhưng dù là cách nào thì chỉ cần nhìn vào biểu đồ là ta đã có thể biết được quốc gia nào có tổng số ca nhiễm nhiều nhất và quốc gia nào là ít nhất, sự chênh lệch về số lượng các ca giữa các quốc gia cũng được thể hiện tốt hơn.

**Ý nghĩa màu sắc:** Sử dụng tone màu Blue-Teal để cột dữ liệu được biểu diễn theo màu sắc nhạt dần đối với các cột có số lượng ca nhiễm giảm dần, và vì màu thông nhất từ đầu là màu xanh dương là màu thông dụng được dùng để biểu diễn số liệu “Số ca nhiễm” do COVID-19 gây ra.

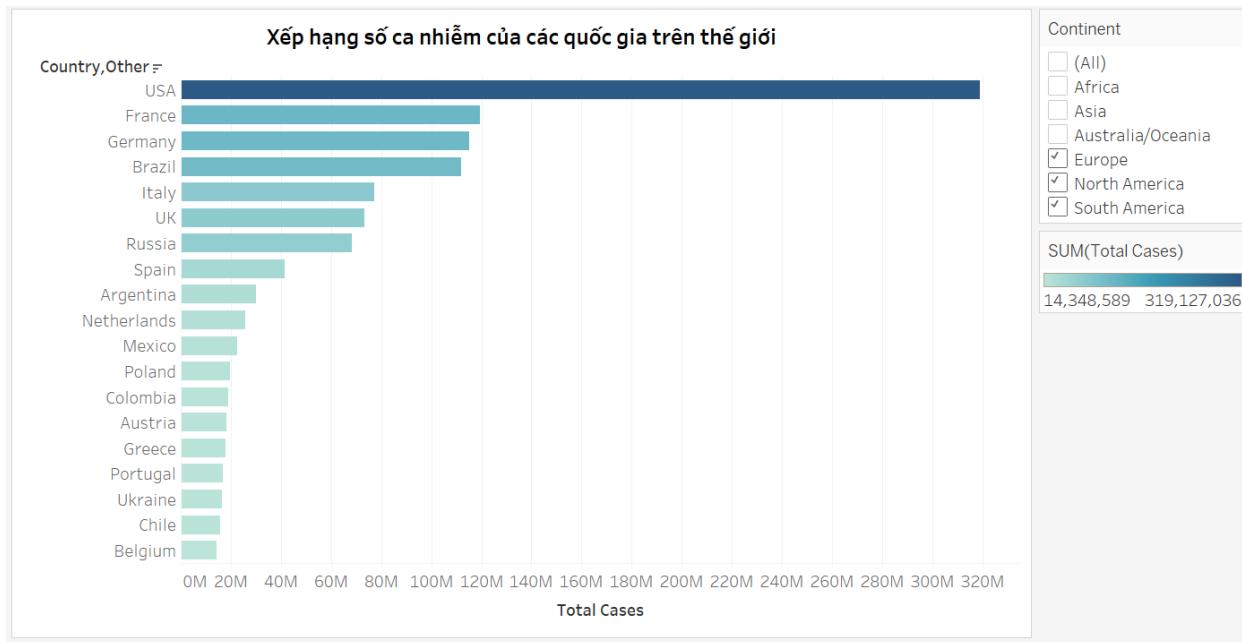
**Nhận xét:** Thông qua biểu đồ trên ta có thể đưa ra một số nhận xét như sau:

- Mỹ có tổng số ca mắc Covid 19 nhiều nhất và vượt xa tất cả các nước còn lại.
- Ấn Độ tuy có dân số rất lớn nhưng số ca nhiễm chỉ bằng một nửa của Mỹ và không hơn các nước Châu Âu với dân số khiêm tốn quá nhiều.
- Biểu đồ này như một bảng xếp hạng cho nên rất hữu ích cho việc so sánh tình hình dịch giữa các nước.

### Kỹ thuật sử dụng:

- Reduce items:

Với biểu đồ hiện tại người dùng có thể xem biểu đồ với dạng ít dữ liệu hơn bằng cách sử dụng filter Continent, chức năng này giúp người dùng quan sát dữ liệu của các quốc gia tại một Châu Lục cụ thể, số lượng thuộc tính không thay đổi.



### Hướng dẫn thực hiện trên Tableau:

- B1: Vẽ columns và rows:

- B2: Filter và Mark



The screenshot shows the Tableau interface with two main shelves:

- Filters:** A blue button labeled "Country, Other".
- Marks:** A section titled "Automatic" with four categories: Color, Size, Label, Detail, and Tooltip. Below these are four measure items:
  - Color: SUM(Tota..)
  - Size: SUM(Tota..)
  - Label: SUM(Acti..)
  - Detail: SUM(Tota..)

- Lưu ý: để màu đậm nhạt theo giá trị từ Cao Nhất → Thấp Nhất, ta sẽ vào Filter → Top chọn measure name mình cần filter và vào Color chọn những thanh màu đậm nhạt

### Stacked Bar Chart

Biểu đồ biểu diễn số ca hồi phục, số ca nhiễm đang điều trị, số ca tử vong của các quốc gia



**Ý nghĩa trực quan:** Từ việc quan sát số ca nhiễm của top 30 quốc gia trên thế giới, nhà phân tích có thể quan tâm đến các số liệu cụ thể của các quốc gia này, và rộng hơn là tất cả các quốc gia trên thế giới. Vì vậy biểu đồ Stacked giúp biểu diễn các số liệu trên trên cùng một cột, mỗi cột là một quốc gia. Bên cạnh đó, biểu đồ này còn có thể quan sát số liệu của các quốc gia theo từng Châu Lục một cách dễ dàng.

**Ý nghĩa màu sắc:** Các màu sắc biểu diễn số ca hồi phục, số ca nhiễm đang điều trị, số ca tử vong đã được thống nhất từ trước vì vậy biểu đồ này cũng áp dụng các màu sắc đó mà không có sự bất tiện nào.

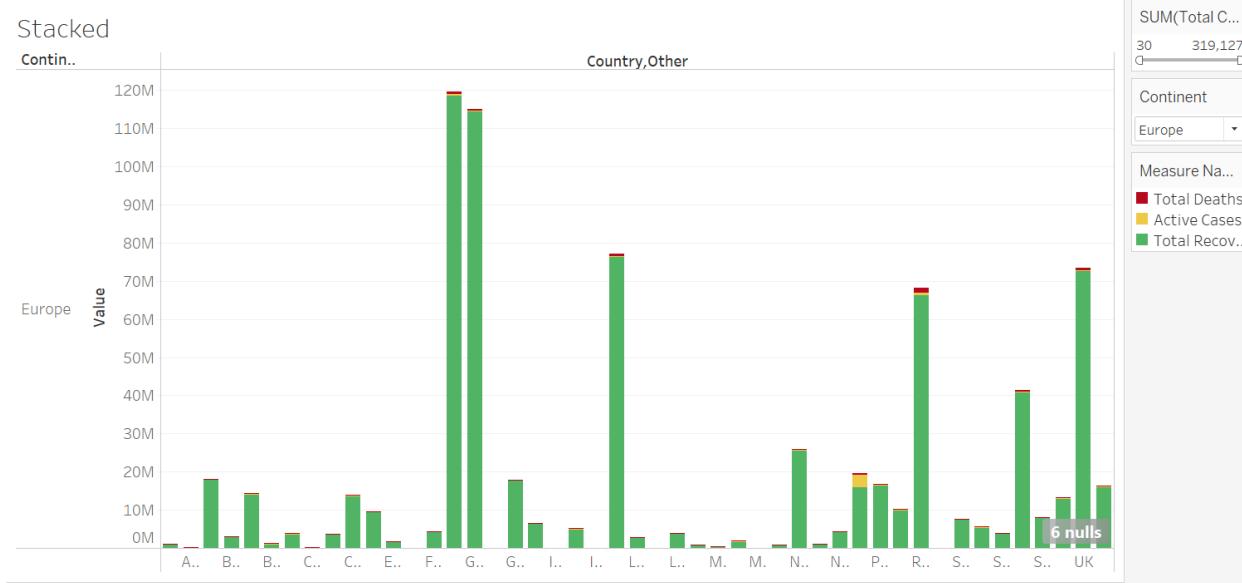
Nhận xét: Thông qua việc sử dụng Stacked Bar Chart ta có thể đưa ra một số nhận xét như sau:

- Nhiều miền dữ liệu được biểu diễn cho ta thấy được sự tương quan giữa số ca hiện đang mắc, số ca hồi phục và số ca tử vong. Số ca tử vong so với số ca hồi phục là rất nhỏ cho thấy công tác chữa trị Covid 19 của các nước là rất đáng khen.
- Số ca hiện đang mắc là rất ít so với số ca hồi phục tức là dịch bệnh đã được kiểm soát.

### Kỹ thuật sử dụng:

- Reduce items:

Với biểu đồ hiện tại người dùng có thể xem biểu đồ với dạng ít dữ liệu hơn bằng cách sử dụng filter Continent, chức năng này giúp người dùng quan sát dữ liệu của các quốc gia tại một Châu Lục cụ thể, số lượng thuộc tính không thay đổi.

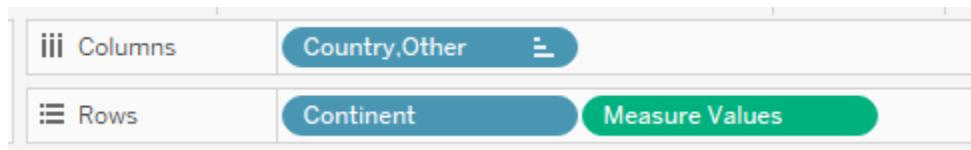


- Changing the alignment in LineUp:

Sắp xếp thứ tự của các thuộc tính (dựa theo giá trị của thuộc tính giảm dần) giúp người xem dễ nhận dạng dễ hơn về các thuộc tính (sự khác biệt về giá trị, độ ưu tiên của giá trị, ...). Ở đây, ta thấy được số ca hồi phục chiếm nhiều nhất so với tổng số ca nhiễm ở tất cả các quốc gia. Từ đó ta có thể đoán rằng diễn biến dịch bệnh đang trong giai đoạn hồi phục, không còn ở đỉnh điểm của dịch bệnh.

### Hướng dẫn thực hiện trên Tableau:

B1: Vẽ Columns và Rows:



B2: Filter và Marks:

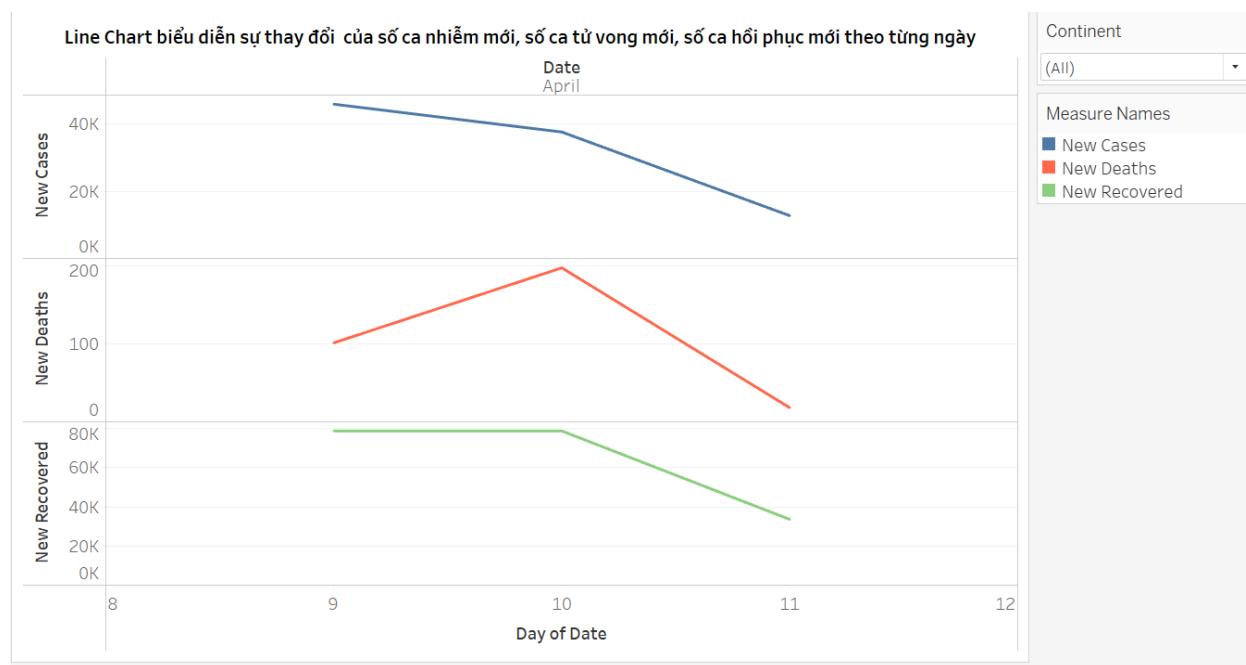


The screenshot shows a data visualization interface with three main sections:

- Filters:** Contains "Measure Names" (highlighted in blue), "Continent", and "SUM(Total Cases)" (highlighted in green).
- Marks:** Shows settings for "Automatic" marks, with options for Color, Size, Label, Detail, and Tooltip. "Color" is highlighted in blue.
- Measure Values:** Contains "SUM(Total Deaths)" (highlighted in green), "SUM(Active Cases)" (highlighted in green), and "SUM(Total Recovered)" (highlighted in green).

### Line Chart

Trực quan dữ liệu biến đổi theo ngày



**Ý nghĩa trực quan:** Để biểu diễn xu hướng, sự thay đổi theo thời gian của số ca tử vong mới, số ca hồi phục mới theo từng ngày thì việc lựa chọn line chart là rất hợp lý. Line chart giống như một đường xu hướng nhưng chi tiết hơn và dễ quan sát hơn. Với các mốc thời gian theo ngày, tuần, tháng, ta có thể biết được xu hướng của ngày hôm nay là tăng hay giảm so với hôm qua hay của tuần này so với tuần trước hay tháng này so với tháng trước.

**Ý nghĩa màu sắc:** Tương tự như Total Cases, Total Deaths, Total Recovered thì dữ liệu New Cases, New Deaths, New Recovered cũng được sử dụng các màu lân lượt là Xanh dương, Đỏ, Xanh lá

**Nhận xét:** Thông qua Line Chart trên ta thấy được xu hướng của dịch bệnh trong 3 ngày được lấy dữ liệu như sau:

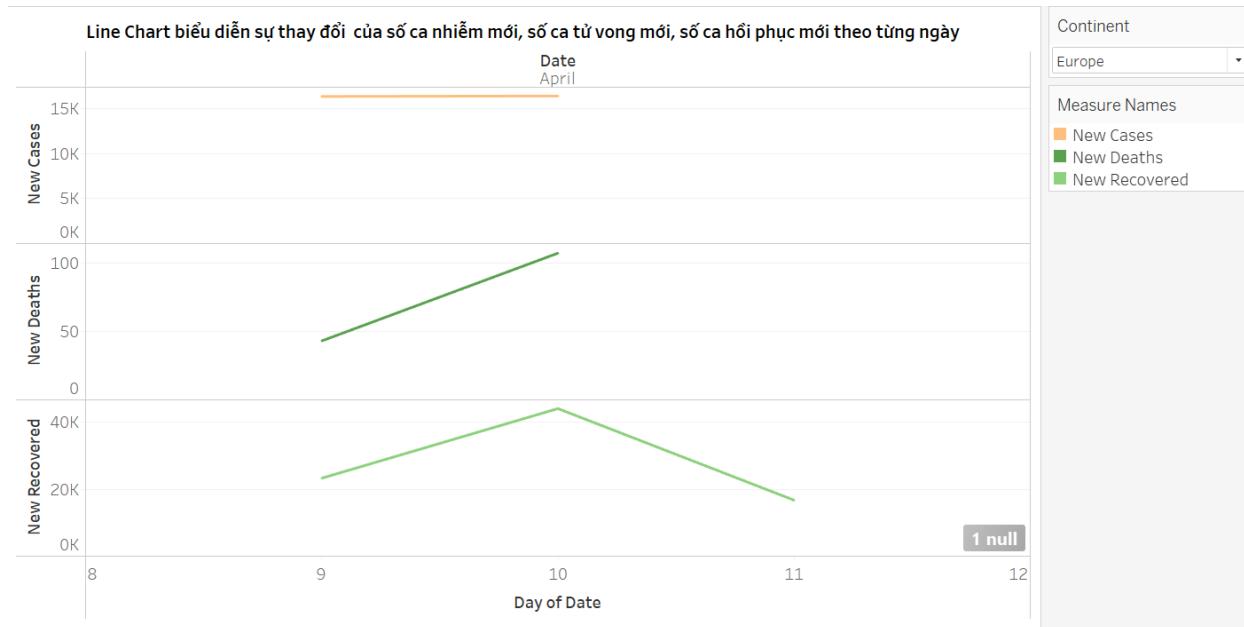
- Số ca nhiễm mới có xu hướng giảm mạnh từ 40 nghìn xuống còn 20 nghìn.
- Số ca tử vong ở ngày 10 tăng từ 100 lên 200 so với ngày 9 nhưng lại giảm mạnh vào ngày 11.
- Số ca hồi phục có xu hướng gần giống với số ca nhiễm mới tức là hầu hết các ca nhiễm đều được chữa khỏi và dần hồi phục.

### Kỹ thuật sử dụng:

- Reduce items:



Với biểu đồ hiện tại người dùng có thể xem biểu đồ với dạng ít dữ liệu hơn bằng cách sử dụng filter Continent, chức năng này giúp người dùng quan sát dữ liệu của các quốc gia tại một Châu Lục cụ thể, số lượng thuộc tính không thay đổi.



*Hướng dẫn thực hiện trên Tableau:*

B1: Vẽ Columns và Rows:

Columns	MONTH(Date)	DAY(Date)	
Rows	SUM(New Cases)	SUM(New Deaths)	SUM(New Recovered)

B2: Filter và Marks:



Filters

Date

Measure Names

Continent: Europe

Marks

~ All

~ Automatic

Color

Size

Label

Detail

Tooltip

Path

Measure Nam..

~ SUM(New Cases)

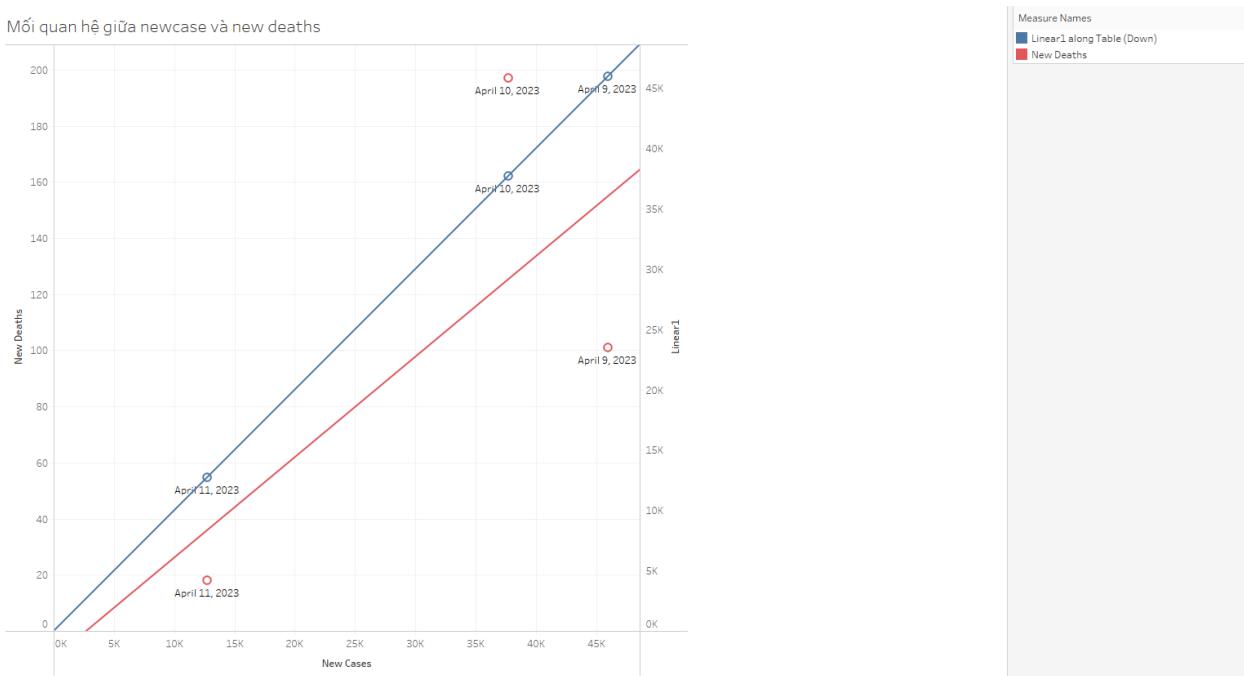
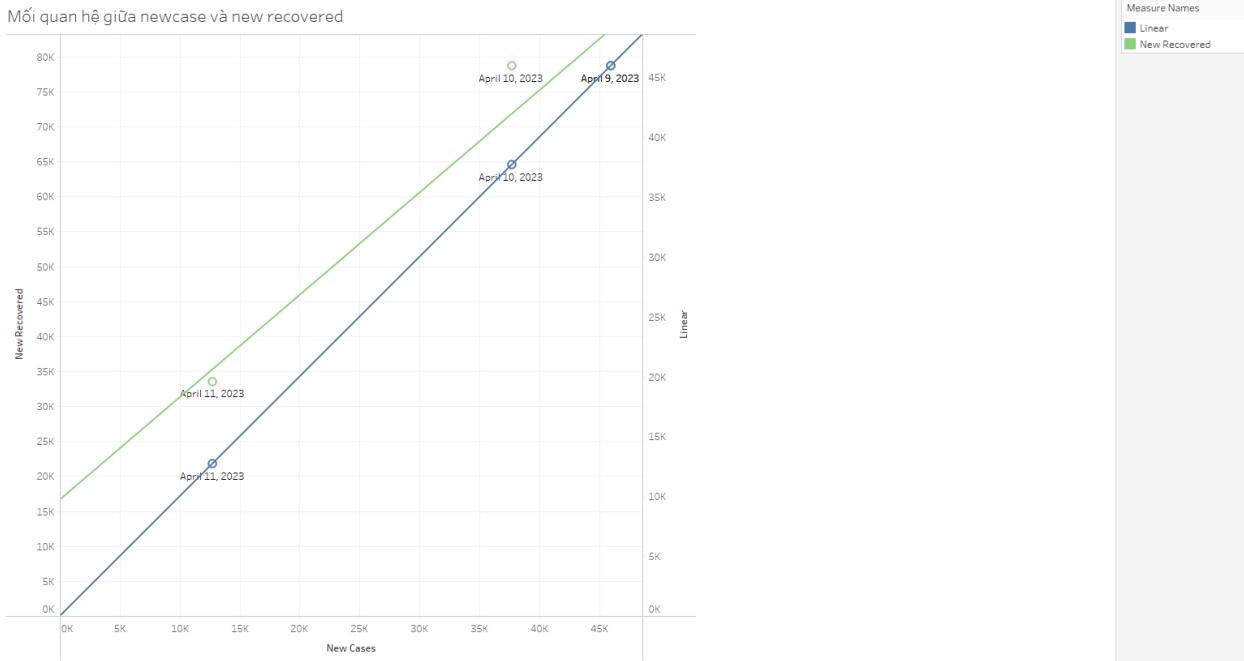
~ SUM(New Deat...)

~ SUM(New Reco...

The screenshot shows the Tableau interface with the 'Filters' shelf at the top containing three items: 'Date', 'Measure Names', and 'Continent: Europe'. Below it is the 'Marks' shelf, which is currently set to 'Automatic'. It includes buttons for Color, Size, Label, Detail, Tooltip, Path, and a dropdown menu for 'Measure Nam..'. Underneath these shelves, there are three summary measures listed: 'SUM(New Cases)', 'SUM(New Deat...)', and 'SUM(New Reco...)'.

### 3. Một số thuật toán học máy

#### Linear Regression



- **Ý nghĩa trực quan:** 2 biểu đồ scatter trên thể hiện mối quan hệ giữa new cases - new recovered và new cases - new deaths. Biểu đồ các mối quan hệ mang cho ta một cái nhìn cụ thể:

- Mật độ new cases và new deaths đang giảm dần theo thời gian

- Mật độ new cases và new recovered đang tăng dần theo thời gian

Đường xu hướng về mối quan hệ new cases - new deaths thấp hơn so với đường hồi quy tuyến tính

Đường xu hướng về mối quan hệ new cases - new recovered cao so với đường hồi quy tuyến tính

- **Ý nghĩa màu sắc:**

- Màu sắc đỏ và xanh ở các điểm dữ liệu thể hiện mối quan hệ giữa new cases - new deaths và new cases - new recovered
- Màu xanh lam thể hiện đường hồi quy tuyến tính giữa tổng new cases - new deaths và new cases - new deaths

- **Nhận xét:**

Biểu đồ trên cho thấy, khi ca nhiễm tăng thì số ca tử vong, số ca hồi phục cũng tăng theo. Điều này có thể đưa ra nhận định rằng tại các quốc gia có số ca nhiễm càng lớn thì hệ thống y tế, chế độ dinh dưỡng và chăm sóc sức khỏe của các quốc gia đó được đẩy mạnh và phát triển, và dẫn đến việc khỏi bệnh sẽ diễn ra nhanh hơn so với các quốc gia bình thường, vì thế dẫn đến số lượng ca phục hồi cũng tăng theo. Và khi các quốc gia tăng cường xét nghiệm thì số ca được ghi nhận cũng tăng theo và dựa vào mối tương quan dương của Số ca nhiễm với số ca tử vong, Số ca nhiễm với số ca hồi phục , thì khi số lần xét nghiệm tăng cũng kéo theo 2 số liệu về ca tử vong và ca phục hồi tăng theo.

Từ biểu đồ trên, ta thấy dịch bệnh covid trên toàn thế giới đang giảm đi theo từng ngày. Ngoài ra cho các đường xu hướng thấy phòng chống dịch bệnh trên toàn thế giới đang diễn ra rất mạnh mẽ và kiểm soát dịch bệnh mà không cần thống kê và phân tích các số liệu cụ thể nào.

### **Hướng dẫn thực hiện trên Tableau:**

#### *a. Setup Tabpy*

- Cài thư viện **tabpy** trong Python nhằm tạo đường hồi quy tuyến tính bằng commandline với câu lệnh **pip install tabpy**

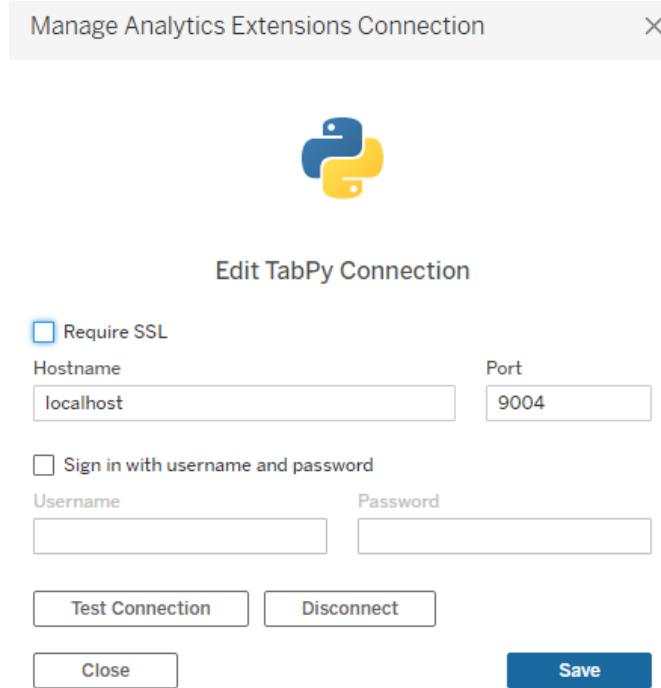
#### *b. Cài đặt Linear Regression*

- Sau khi cài đặt xong, ta sẽ khởi chạy **tabpy** trong commandline



```
C:\Users\DELL>tabpy
2023-04-18,20:07:28 [INFO] (app.py:app:248): Parsing config file E:\Python\lib\site-packages\tabpy\tabpy_server\app\..\common\default.conf
2023-04-18,20:07:28 [INFO] (app.py:app:443): Loading state from state file E:\Python\lib\site-packages\tabpy\tabpy_server\state.ini
2023-04-18,20:07:28 [INFO] (app.py:app:338): Password file is not specified: Authentication is not enabled
2023-04-18,20:07:28 [INFO] (app.py:app:353): Call context logging is disabled
2023-04-18,20:07:28 [INFO] (callbacks.py:callbacks:43): Initializing TabPy...
2023-04-18,20:07:28 [INFO] (callbacks.py:callbacks:64): Initializing TabPy Server...
2023-04-18,20:07:28 [INFO] (app.py:app:133): Done initializing TabPy.
2023-04-18,20:07:28 [INFO] (app.py:app:83): Setting max request size to 104857600 bytes
2023-04-18,20:07:28 [INFO] (app.py:app:110): Web service listening on port 9004
2023-04-18,20:08:25 [INFO] (base_handler.py:base_handler:115): Authentication is not a required feature for API "v1"
2023-04-18,20:08:25 [INFO] (base_handler.py:base_handler:115): Authorization header not found
2023-04-18,20:08:25 [INFO] (web_py:web:2271): 200 GET /info (127.0.0.1) 4.99ms
2023-04-18,20:08:25 [INFO] (base_handler.py:base_handler:115): Authentication is not a required feature for API "v1"
2023-04-18,20:08:25 [INFO] (base_handler.py:base_handler:115): Authorization header not found
2023-04-18,20:08:25 [INFO] (web_py:web:2271): 200 GET /info (127.0.0.1) 3.99ms
2023-04-18,20:08:14 [INFO] (base_handler.py:base_handler:115): Authentication is not a required feature for API "v1"
2023-04-18,20:08:14 [INFO] (base_handler.py:base_handler:115): Authorization header not found
2023-04-18,20:09:33 [INFO] (base_handler.py:base_handler:115): Authentication is not a required feature for API "v1"
2023-04-18,20:09:33 [INFO] (base_handler.py:base_handler:115): Authorization header not found
2023-04-18,20:09:33 [INFO] (web_py:web:2271): 200 GET /info (127.0.0.1) 4.99ms
2023-04-18,20:09:33 [INFO] (base_handler.py:base_handler:115): Authentication is not a required feature for API "v1"
```

- Trở về tableau, ta sẽ vào mục **Help → Setting and Performance → Manage Analytic Extension Connection**



- Ta connect với port 9004 và nhấn Save
- Vào phần worksheet, ta sẽ tạo calculation field như sau:

Linear

new\_covid\_data\_20230409 (2) (2)

Results are computed along Table (across).

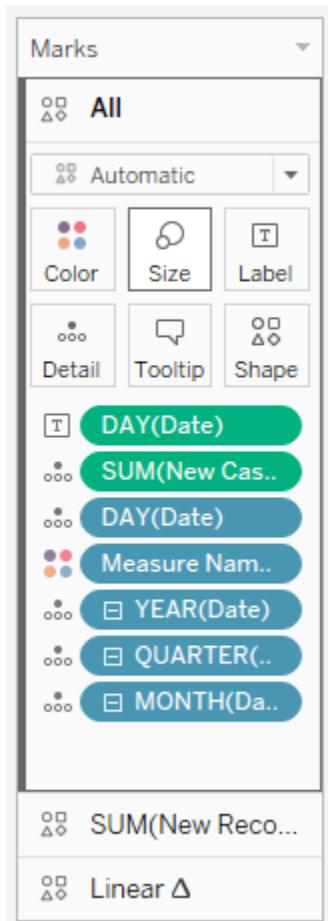
```
SCRIPT_REAL('
import numpy as np
from sklearn import linear_model
clf = linear_model.LinearRegression()
x = np.transpose(np.array([_arg1]))
y = np.array(_arg2)
clf.fit(x,y)
return clf.predict(x).tolist()
',SUM([New Recovered]),SUM([New Cases]))|
```

### c. Trục quan

- B1: Vẽ Columns và Rows



- B2: Filter và Marks





## IV. Link tham khảo

### Link tham khảo:

- [1] [Tableau là gì? Tính năng nổi bật của Tableau](#)
- [2] [Những tính năng hữu ích của Tableau](#)
- [3] [Tableau Real Time Data Streaming: 2 Easy Methods](#)
- [4] [Publish Data Sources and Workbooks](#)
- [5] [Learn to Add Trend Lines in Tableau in Just 3 Steps!](#)
- [6] [Ask Data feature in Tableau](#)
- [7] [Build Charts and Analyze Data](#)
- [8] [What is Tableau Dashboard and How to Build it?](#)