Data Visualization
Lab 2 – Visualizing Data with Tableau

1 Team information

No	Student ID	Student Name	Task	Completion (%)
1	19127242	Đỗ Vương Phúc	- Visualize crawled data to diagrams - Evaluate and give comments on data	100
2	19127142	Trần Thái Đức Hiếu	- Visualize crawled data to diagrams - Evaluate and give comments on data	100
3	19127360	Dương Thị Xuân Diệu	- Write report - Write Introduction of Tableau	100

2 Contents

3   Overview

3.1   About the problem

From about the end of 2019 and the beginning of 2020, a plague spread horribly around the world. Every day, thousands of people are infected and tens to hundreds of people die. Worldometer Organization (www.worldometers.info) has collected statistical data from many sources and many countries reporting daily to compile into a table.

3.2   Self-evaluation

In this lab, we have done all the given task including:
- **Theory: Learning Tableau tool**
  - Introduction to **Tableau**
  - Demonstrating Tableau's features with illustrative examples. The examples are made on existing datasets, not related to the main dataset of the lesson.
- **Practicing: Using Tableau to visualize Woldometer data**
  - Observing data from simple to complex, from single attribute to combined attributes, from independent to dependent relations, etc.
  - Choosing from a wide variety of chart formats, assess conformity, and reuse reasoning in the previous section. Showing these charts in Tableau.
  - Using colors to represent data.
  - Visualizing some day-to-day variable data.
  - Using techniques introduced in Manipulate View, Facet, Reduce, Embed for Tableau demonstration with Worldometer data. Explaining the selection and meaning.
  - Running some simple machine learning algorithms to understand more about data, learning algorithms.
  - The machine is entitled to use the available code, clearly stating the origin.

4 Introduction to Tableau

4.1 What is Tableau?

Tableau is an excellent data visualization and business intelligence tool used for reporting and analyzing vast volumes of data. It is an American company that started in 2003—in June 2019, Salesforce acquired Tableau.

It helps users create different charts, graphs, maps, dashboards, and stories for visualizing and analyzing data, to help in making business decisions.

Tableau has a lot of unique, exciting features that make it one of the most popular tools in business intelligence (BI). Let's learn more about some of the essential Tableau Desktop features. Now that we know what tableau is exactly, let us understand some of its salient features.


4.2 Main Feature of Tableau

4.2.1 Informative Dashboards

Tableau Dashboards combine images, visual objects, text, and other components to present a comprehensive view of your data. Dashboards are extremely useful because they:
- Provide data in the form of stories.
- Allow for the inclusion of various views and objects
- Offer a range of layouts and styles and allow users to apply appropriate filters.

You may even effortlessly duplicate a dashboard or its individual features from one worksheet to another.
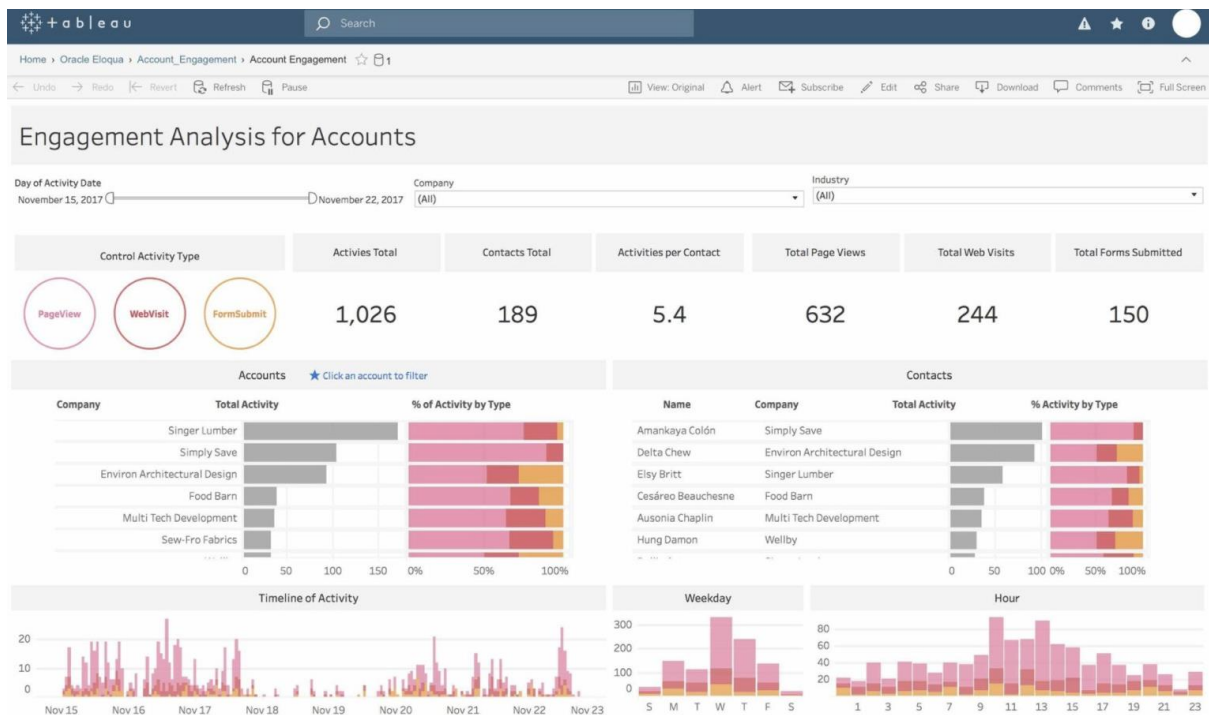
4.2.2   Supports numerous data sources

You may connect to and fetch data from a variety of data sources using Tableau. Tableau supports a wide range of data sources, including local files, spreadsheets, relational and non-relational databases, data warehouses, big data, and on-cloud data.

Any of Tableau's data sources may be readily connected and combined with data from other sources to generate a combinatorial perspective of data in the form of visuals. Tableau also supports a variety of data connections, including Presto, MemSQL, Google Analytics, Google Sheets, Cloudera, Hadoop, Amazon Athena, Salesforce, SQL Server, Dropbox,...

4.2.3   Connectivity with Live and In-Memory Data

Tableau offers in-memory data connection to both live and external data sources. This allows the user to freely combine data from several types of data sources.

By creating live data connections, you may consume data straight from the data source or maintain data in memory by extracting data from a data source as needed. Tableau offers additional data connections capabilities including automated extract refreshes, notification of a live connection failure, and so forth.

4.2.4   Provides Great Security

Tableau takes extra precautions to protect data and users. For data connections and user access, it features a fail-safe security system based on authentication and authorization mechanisms. Tableau also allows you to connect to other security protocols like Active Directory and Kerberos. Tableau employs row-level filtering, which aids in the security of the data.

4.2.5   Easy Collaboration & Sharing

Tableau provides easy ways for users to communicate with one another and exchange data in real-time in the form of visualizations, sheets, dashboards, and so on. It enables you to securely communicate data from a variety of data sources, including on-premise, cloud, hybrid, and so on. Instant and simple cooperation and data sharing aid in obtaining immediate assessments or input on data, resulting in a more comprehensive study.
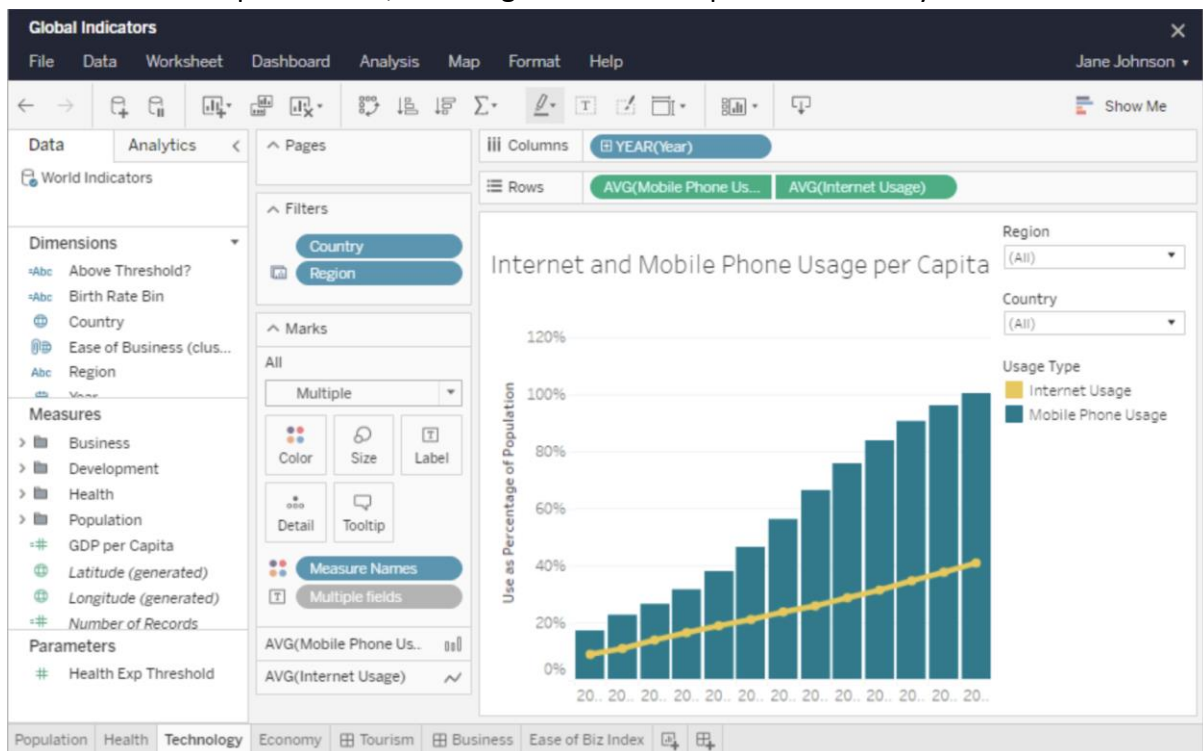

4.2.6   Provides a Mobile Version

Tableau recognizes the importance of mobile phones in today's society and offers a mobile version of the app. Dashboards and reports may be created in a mobile-friendly format. Tableau allows you to adjust mobile layouts for your dashboard based on your mobile device. Customization allows you to add new phone layouts, interactive offline previews, and more. As a result, the mobile view provides Tableau users with a great deal of flexibility and convenience while working with their data on the move.

4.2.7   Advanced Visualization Capabilities

Tableau's wide range of visualizations is one of the primary elements that has contributed to its success. Tableau allows you to create visualizations as simple as a bar chart or a pie chart, as well as more complex ones like a Histogram, Gantt chart, Bullet chart, Motion chart, Treemap, Boxplot, and many others. By selecting the visualization type from the Show Me menu, you can simply pick and create any form of visualization.
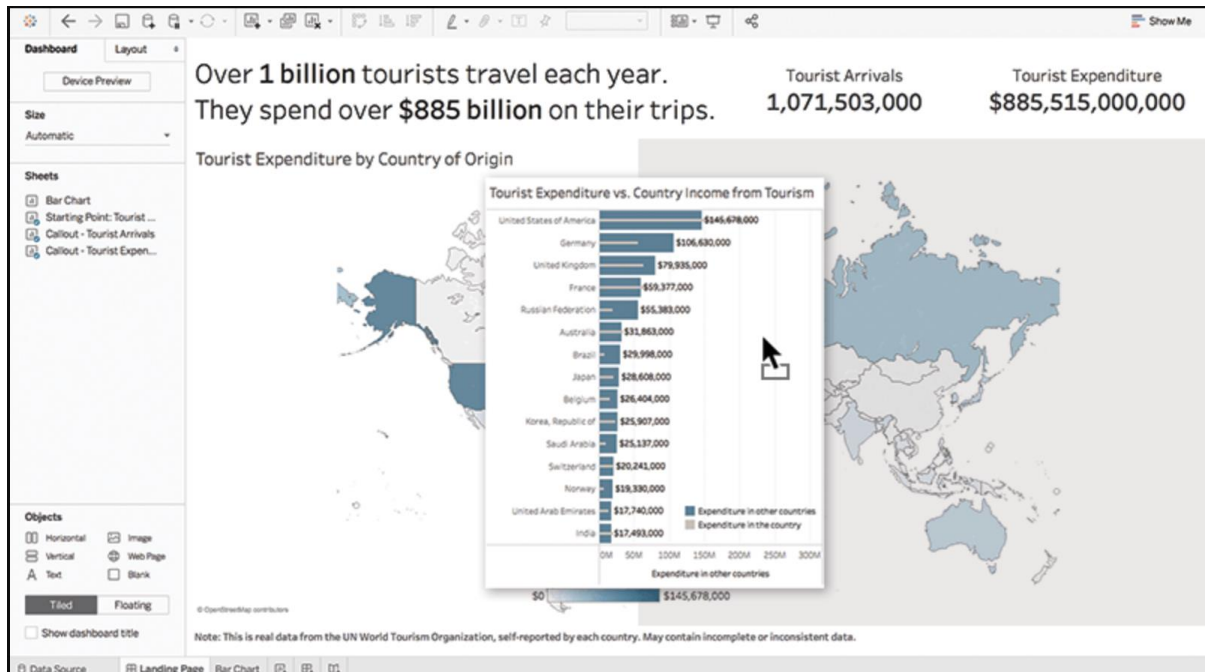
4.2.8   Availability of Maps

The map is yet another key aspect of Tableau. Tableau comes with a lot of pre-installed map data, including cities, postal codes, administrative borders, and so on. As a result, Tableau's maps are extremely comprehensive and insightful.

You may customize the geological layers on the map to meet your needs, and use Tableau to generate meaningful maps with your data. Heat maps, Flow maps, Choropleth maps, Point distribution maps, and more types of maps are accessible in Tableau.

4.2.9   The Ask Data Tool

Tableau's Ask data tool has increased its popularity among users all around the world. This tool simplifies data manipulation by allowing us to conduct basic Google searches. Tableau will give you the most relevant replies if you just enter a question about your data in natural language. The responses are presented not just as text but also as graphics.

For example, if what you're looking for is already in a bar graph, the Ask data option will search for it and open it for you right now. Users may simply go deep into data and uncover new insights and patterns thanks to capabilities like these.

4.2.10 Trend Lines & Predictive Analysis

The use of time series and forecasting by Tableau is another really useful feature. Creating trend lines and forecasts is straightforward with Tableau's powerful backend and dynamic front end. To acquire data predictions such as a forecast or a trend line, just pick particular parameters and drag-and-drop operations employing your concerned fields.
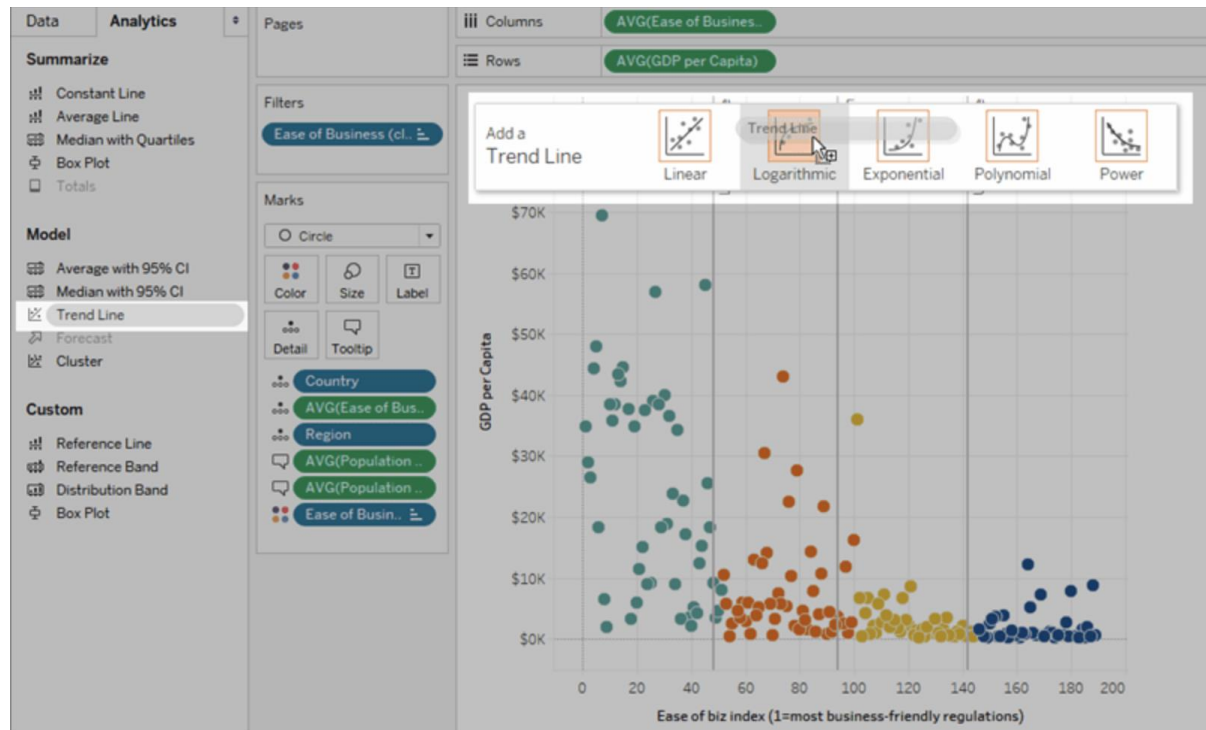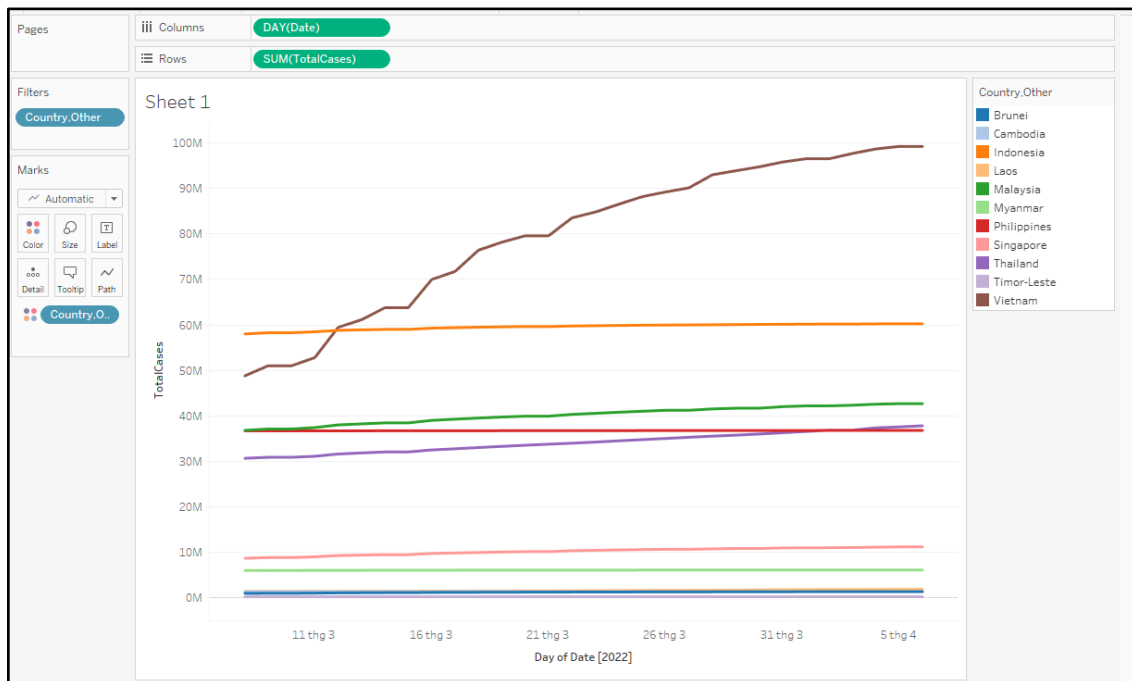
5 Data visualization with Tableau

5.1 About the Data

- The data is collected daily from 09/03/2022 to 13/04/2022 on:
  www.worldometers.info
- Data Preprocessing: We merge all datasets by adding date column to create file **timeseries.csv**

5.2 South-East Asia Cases

First we compare all the Total Cases of 11 countries in SEA with a line graph to have an overview of the situation.

Why line graph:

- Best for timeline
- Easy to compare the values of different countries

However, each country has a different population. Therefore, we should not make a conclusion or evaluate the situation of SEA based on the Total Case
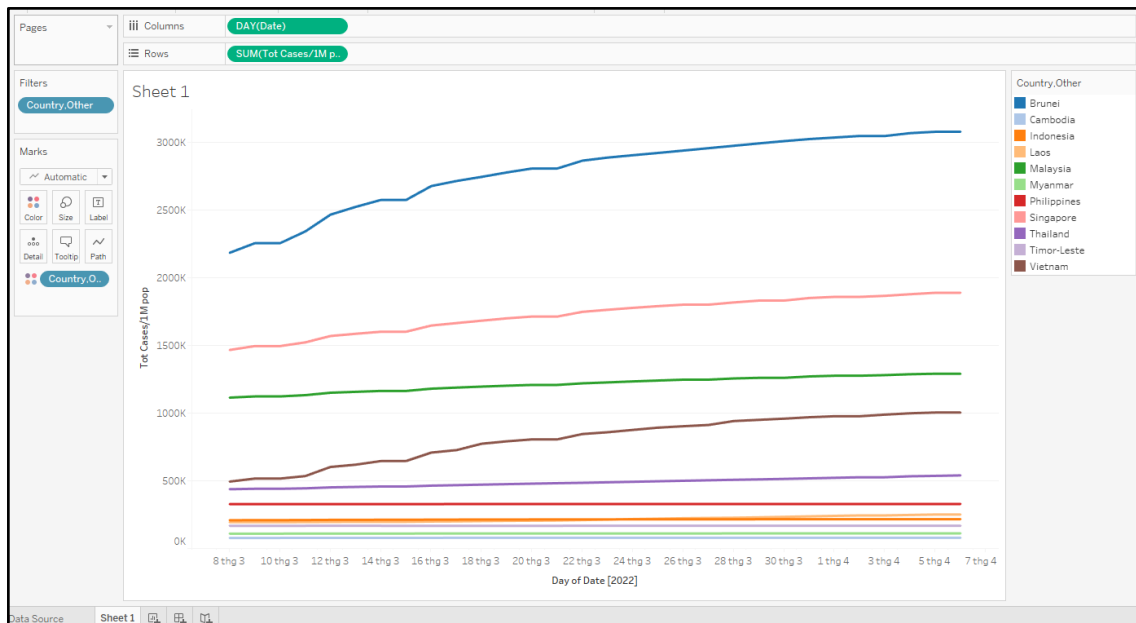
In the next step, we plot the density of Total Cases over the Population (**Total Cases / 1M pop**) to compare the efficiency in preventing Covid-19 in SEA.

As we can see in the line graph, 3 countries that had the highest density of cases were Brunei, Singapore and Malaysia. We will investigate further in these 3 countries.
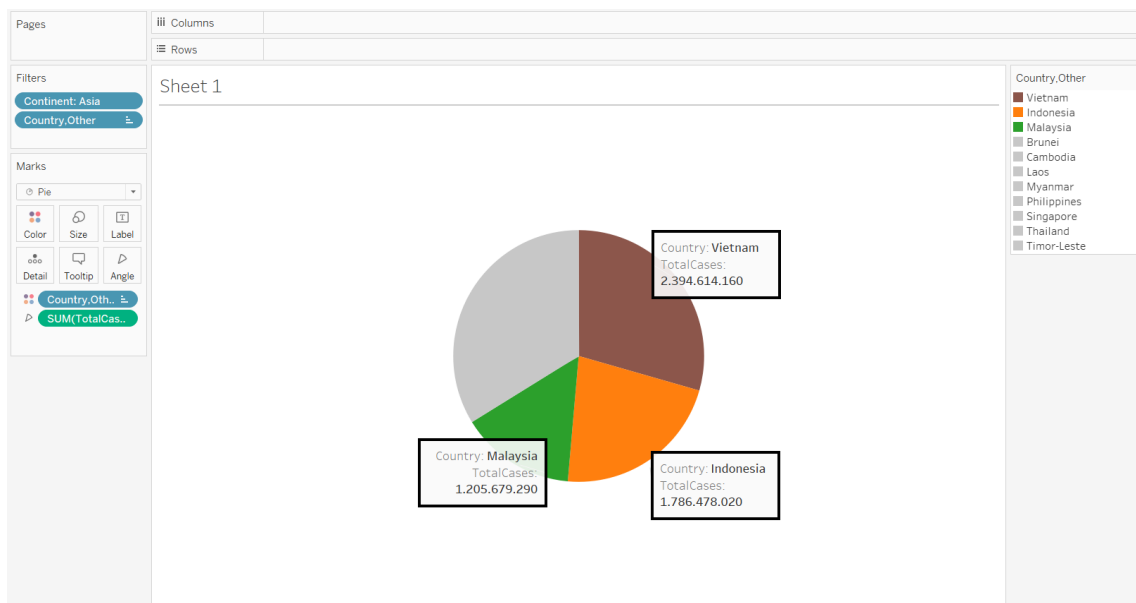

In this pie chart, we can see that Vietnam, Indonesia and Malaysia account for more than 60% of the Total Cases in SEA. However, the next pie chart shows that Brunei, Singapore and Malaysia account for more than ⅔ of the Total Cases Density (Total Cases / 1M pop).
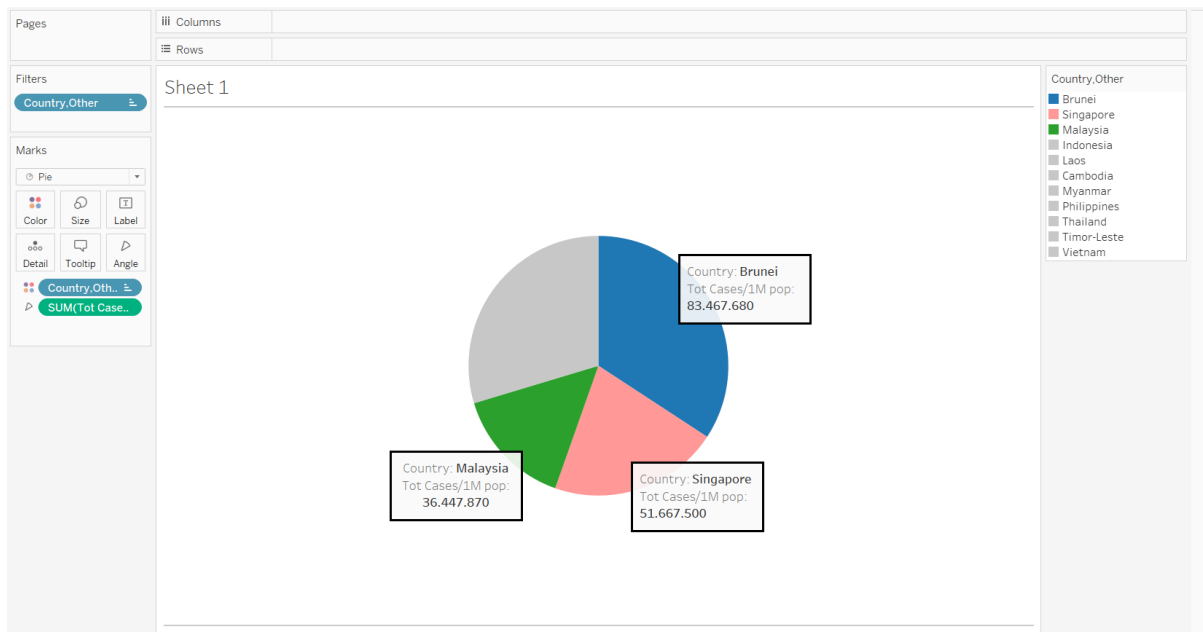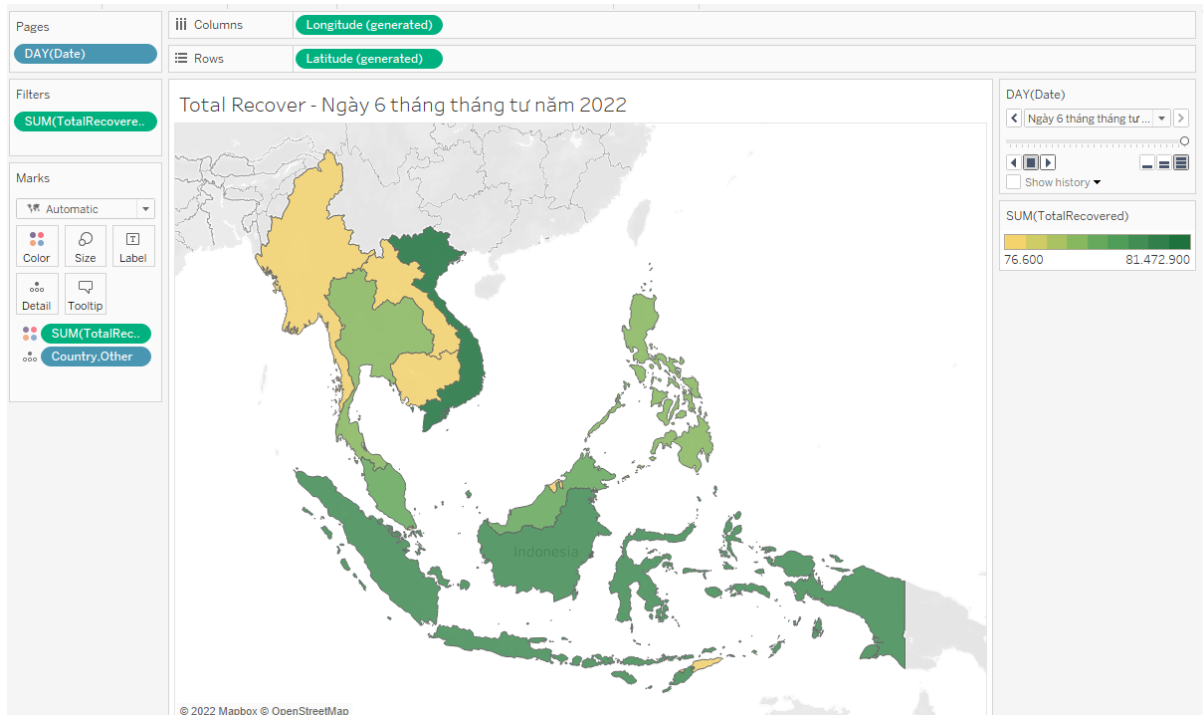
Therefore, we can suppose that the spreading rates of Covid-19 (especially Omiron in this period of time) in Brunei and Singapore are faster than in Vietnam and Indonesia.

5.3   South-East Asia Recovery
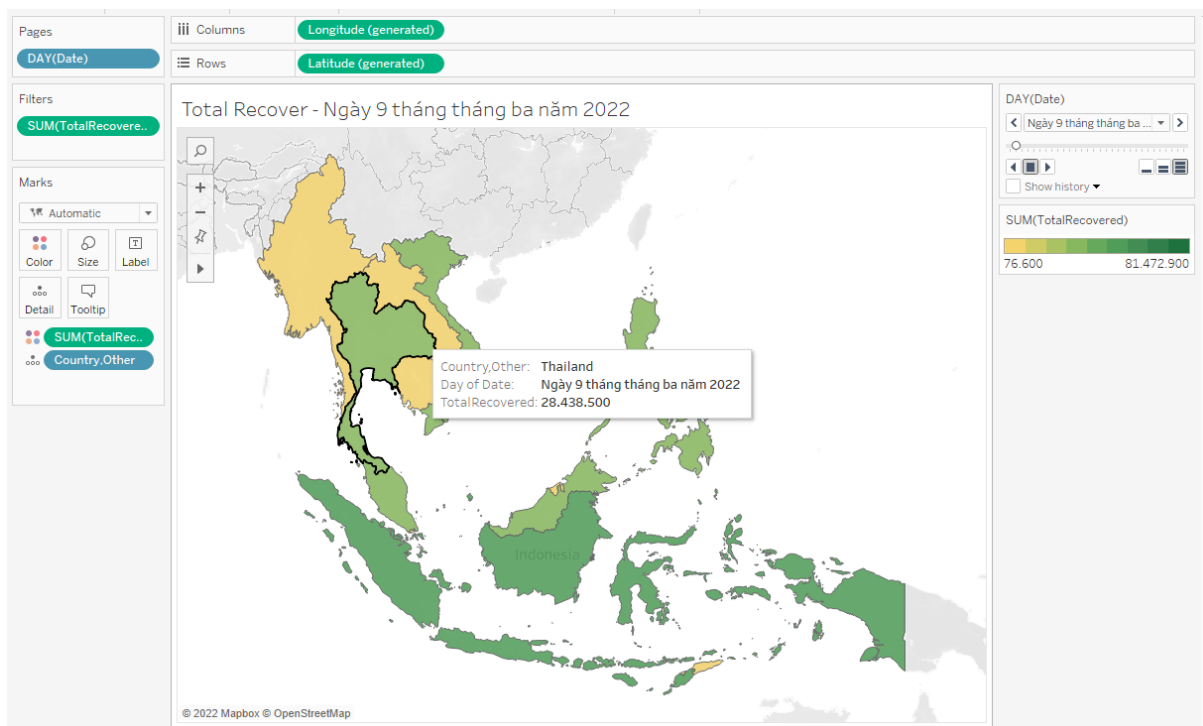
This is the Total Recovery map of SEA on 09/03/2022. Indonesia had the highest Recovery Cases.

However, after nearly a month, on 06/04/2022, Vietnam had the most Total Recovered people. To explain this, we have some ideas:

- In this period, Vietnam had the most Total Cases in SEA. However, Omicron was also popular in time.
- Because Omicron has a short recovery time, most of the cases in this period would recover within a month.

The Total Recovered Cases of Myanmar, Laos and Cambodia did not increase much because they were among the least Total Cases countries.
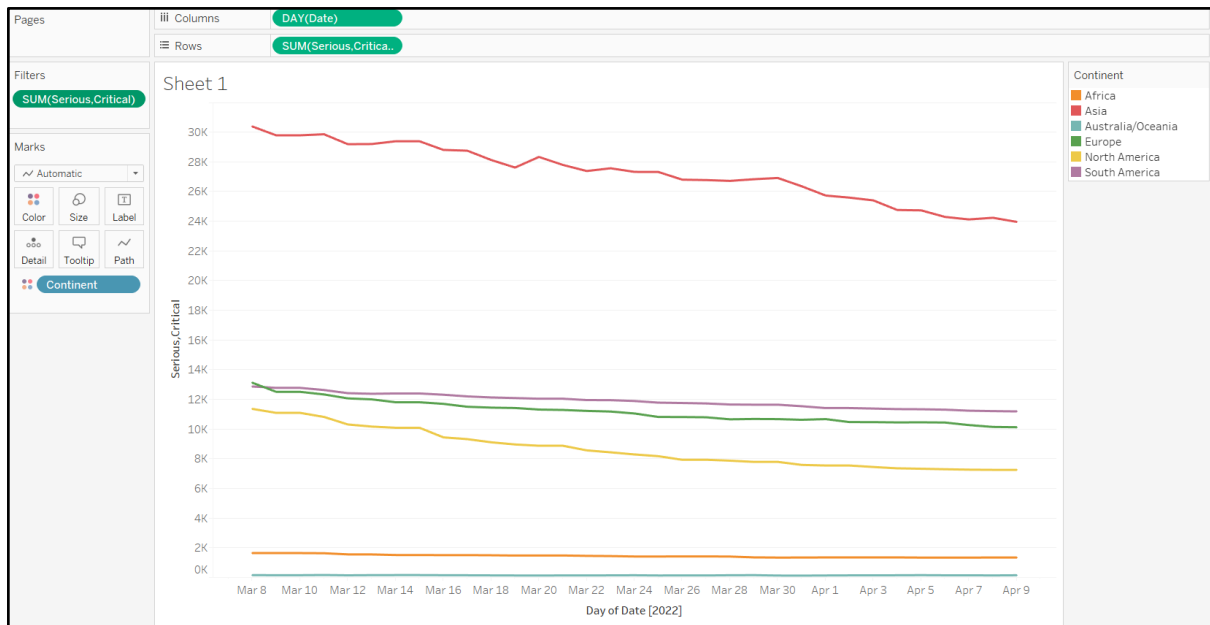
5.4   Total serious cases over continents

Reason for choosing the chart:
Using the line chart can help us to show trends over time. Furthermore, we have already used different colors to specify continents.

Comments:
In general, it can be seen that over time, the number of serious cases are decreasing which can be inferred that the new species of COVID virus - Omicron is less dangerous than the previous one - Delta.

5.5   Recovery of USA

Reason for choosing the chart:

Using the line chart can help us to show trends over time. Furthermore, we have already used different colors to specify continents.
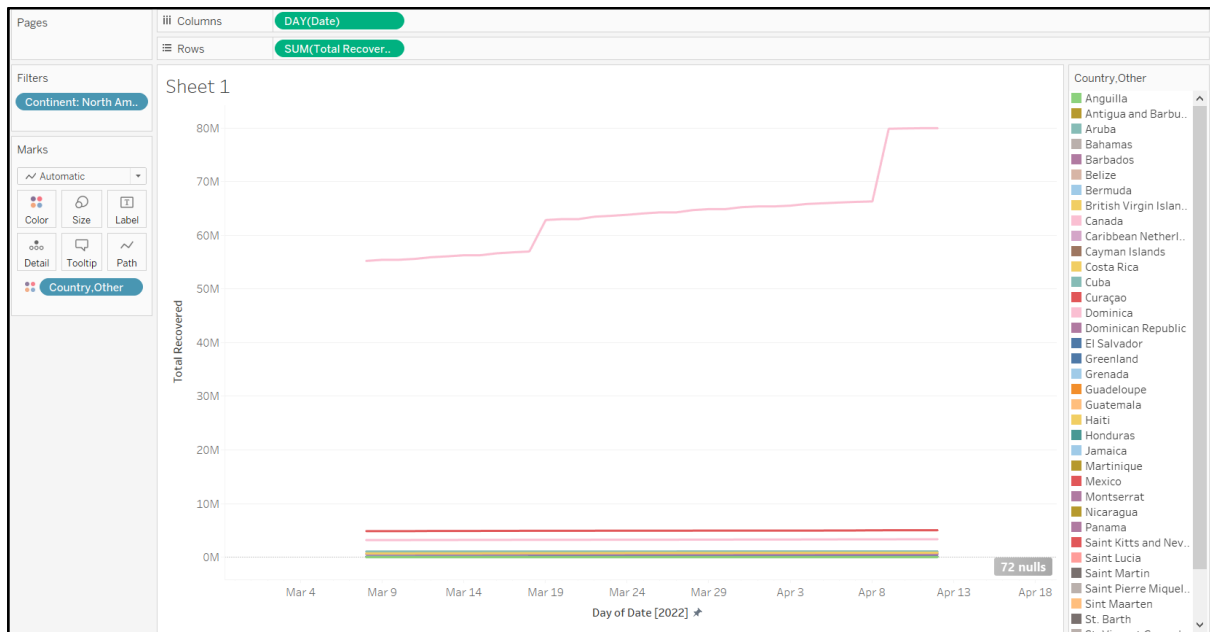
Comments:

With the number of recovery cases, we can see that North America has two sharply rises at 18th March and 8th April. To understand more deeply about this, we filter and visualize the figure of the countries which are located in North America.
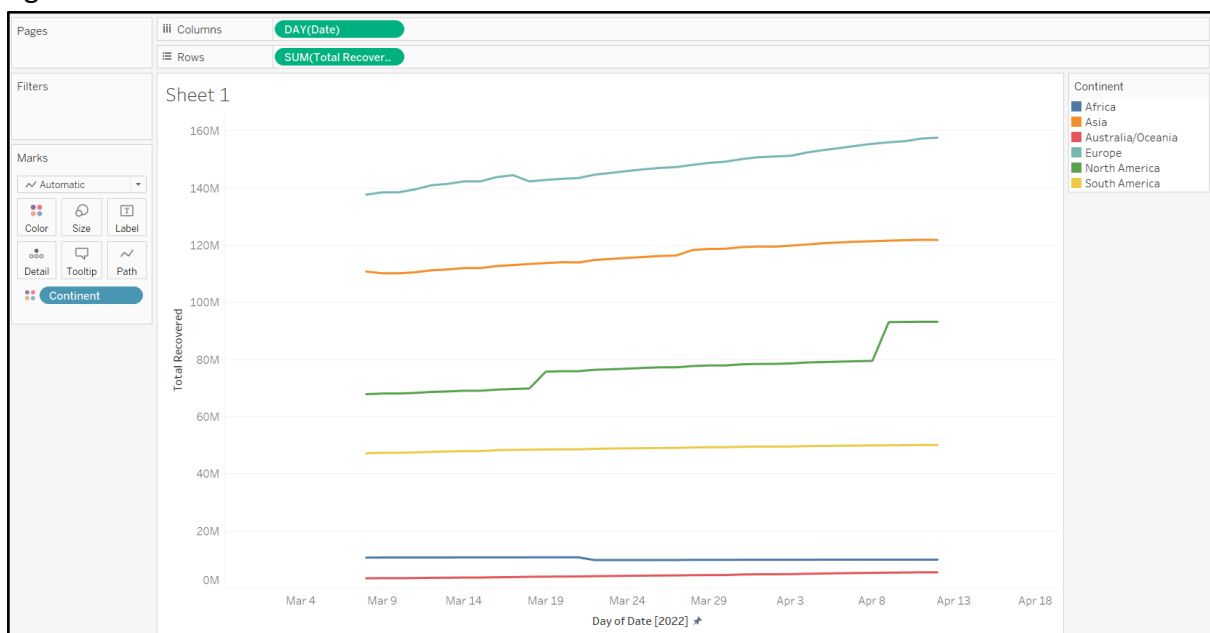
Comments:

As a result, the US recovery cases in those days increased considerably. We can imply this insight that about one-two weeks before there was a vast amount of people being infected by COVID.

Comments:

To verify that insight, we also check for the new cases. The chart depicts clearly that on 11th March the number of new cases in the US reached a peak of nearly 45.000 cases. Also, on the 31st March, and 7th April new cases number are very high. Also, we have found the reason for this situation is that the new variant of Omicron - BA.2 appeared in the US at the beginning of March. "Around the world, infections are largely from the BA.2 version of omicron. In the U.S., BA.2 accounted for about a quarter (23.1%) of the cases for the week ending March 12" - The USA Today network reported from CDC.
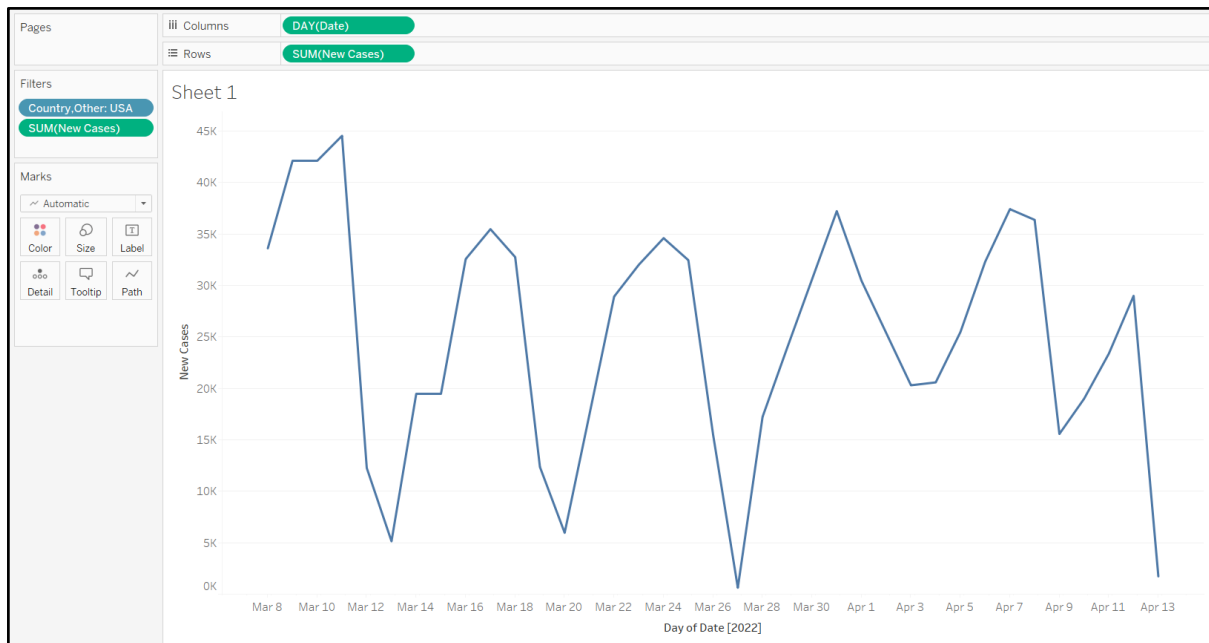
## 5.6   Relationship between rate of death and recovery

Reason for choosing the chart:

With a density chart, it can be used to show the number of samples distributed at each point. This chart can improve the overlapping problem of the scatter plot.

Comments:

In the previous lab, we have already analyzed the rate of death and recovery respective to the total one. However, in that problem, we used the swarm plot to show confidence. Now, with Tableau, we try another kind of plot which is the density plot.

Because the data now is larger due to having many days, the swarm plot cannot adapt to the requirement anymore. So, to show the distribution of data which used to supply our thesis, we choose the density plot. As can be seen, just under 10% of cases were deaded, while the recovered rate is over 70%.
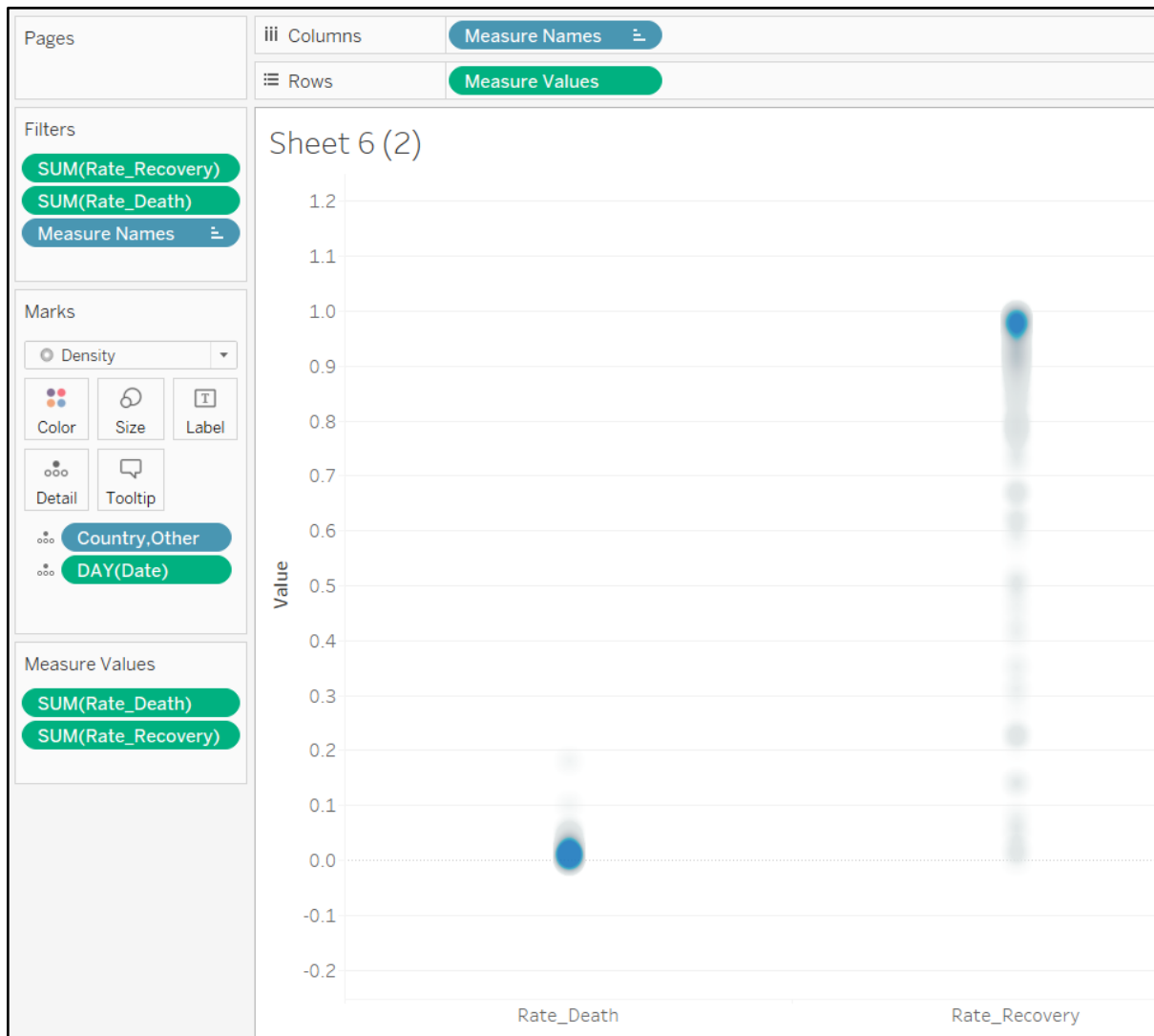
5.7 Chile new cases

Reason for choosing the chart:

Using the line chart can help us to show trends over time.

Comments:

In March, we can see the number of cases of almost every country increased, however, Chile did not. Moreover, their cases were decreasing over time. Obviously, we have done some research to explain this. As a result, there was a newspaper which discussed this problem. According to this paper, Chile has gained community immunity over COVID since they force all adults over 18 year olds to be vaccinated.

This is the according newspaper:

Pháp, Chile ghi nhận các con số trái ngược về ca mắc mới COVID-19

Pháp ghi nhận số ca mắc mới COVID-19 tăng cao nhất kể từ đầu tháng 2 vừa qua, với 180.777 ca trong 24 giờ qua, trong khi số ca mắc mới tại Chile giảm xuống mức thấp nhất trong hơn 2 tháng qua.

Minh Châu (TTXVN/Vietnam+) · 23/03/2022 13:30 GMT+7   https://link.gov.vn/N7H8BLyN   https://link.gov.vn/N7H8BLyN

Trong khi đó, Chile cũng chứng kiến số ca mắc mới trong ngày giảm xuống mức thấp nhất kể từ tháng 4/2020. Với 753 ca nhiễm mới trong ngày 27/7, đây là lần đầu tiên số ca nhiễm mới theo ngày ở Chile xuống dưới 1.000 ca/ngày trong hơn một năm qua.

Hiện quốc gia Nam Mỹ này có tổng cộng 1.611.090 ca mắc COVID-19, trong đó có 35.151 ca tử vong. Kể từ đầu tháng Sáu, Chile đã ghi nhận số ca mắc trong ngày giảm dần và xuống dưới con số 1.500 ca/ngày trung bình cho đến nay.

Một số khu vực tại Chile đã đạt miễn dịch cộng đồng hoàn toàn khi tỷ lệ tiêm vaccine đã phủ rộng hơn 80% dân số cả nước./.

Phương Hoa (TTXVN/Vietnam+)

5.8 Total case forecasting over continents

Reason for choosing the chart:

Using the line chart can help us to show trends over time.

Comments:

We have already seen the trends of the total cases over continents. Now, we try the forecast tool of Tableau to predict the future trend. The result shows that only Europe and Asia may increase gradually, while the others have not much change. Furthermore, we have done some research to figure out what method (algorithm) Tableau uses to forecast. The model that has been used is called DGP (Data generating process).
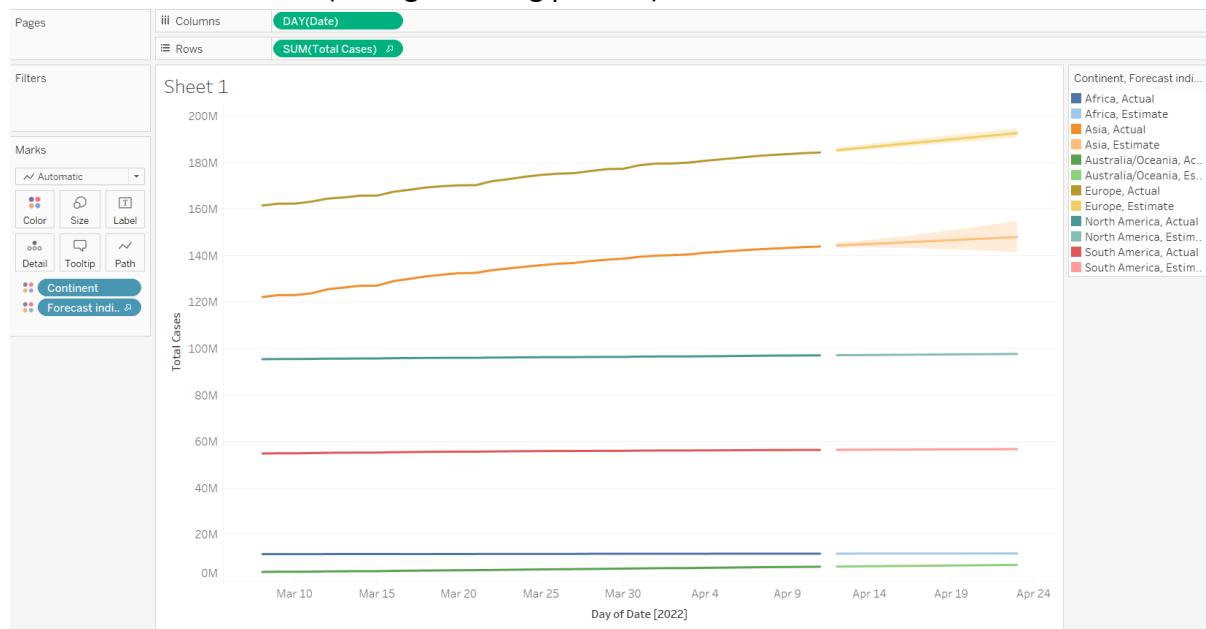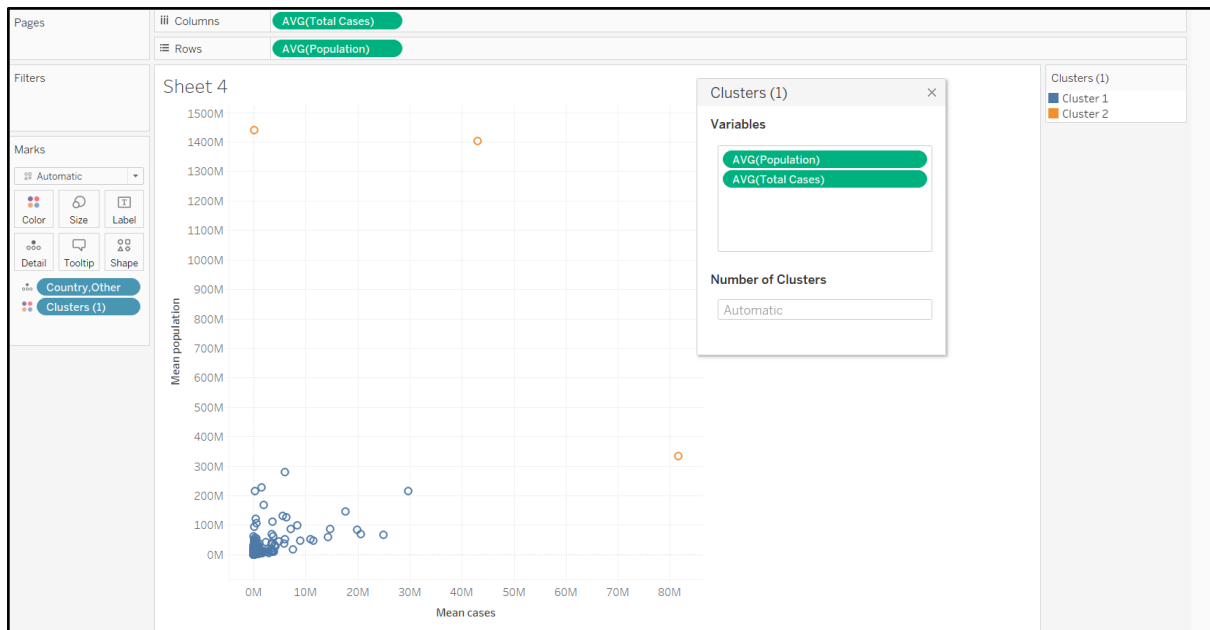

5.9 Segmentation country types in general

Reason for choosing the chart:

Using a scatter plot, we can see clearly the relationship between two variables. In addition, in this observation, we also used a clustering method to segment the data base-on the relationship.

Comments:

We are considering the relationship between the population and the total cases. Indeed, it is easy to see that they are proportional. However, using scatter plot and clustering method, we can see it divided into two groups where the orange cluster may be assumed as the outliers.

Comments:

After that, we excluded the outliers and ran the clustering method again. As a result, an interesting insight has been shown. The data is divided into three groups. We can understand the groups as follows:

- Red cluster: the country which has high population, but the cases is low
- Orange cluster: the country which has small population, but the cases is high
- Blue cluster: the country which is normal

## 5.10 Segmentation country types change over times

Reason for choosing the chart:

Using a scatter plot, we can see clearly the relationship between two variables. In addition, in this observation, we also used a clustering method to segment the data base-on the relationship. Also, the pages feature has been used to show the changes of data points over time.

Comments:

With the above idea, we can explore more information from the data. Instead of using the mean value of cases and population over time, we use the time as a parameter and show the changes by time. From the figure, there are two countries that change the cluster group to the red one. These two countries are VietNam and South Korea.

To explain this we can look back to the scenario of VietNam. Currently, the country has removed the lockdown policy and let people adapt with COVID virus so as to gain immunity. Furthermore, our country has started to open the border gate which can increase the number of cases obviously.
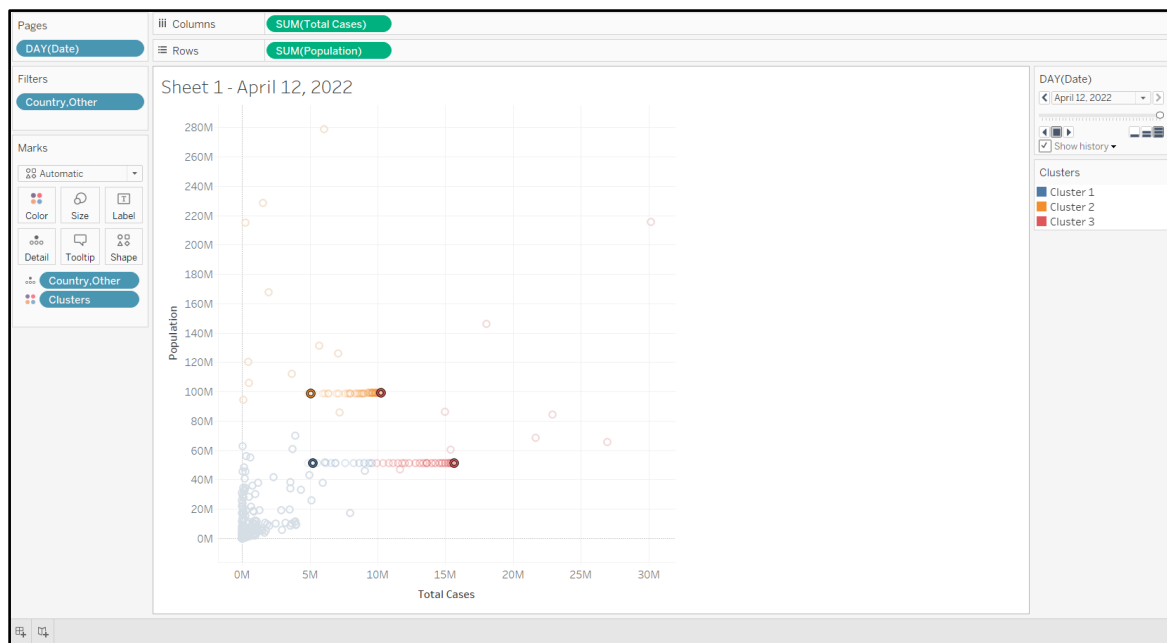
Regarding South Korea, many newspapers said that their government has removed the lockdown policy also; and focus on the serious cases to prevent the death because the Omicron variant is contagious, but not dangerous.

Tình hình COVID-19 ở Hàn Quốc đã diễn biến phức tạp thời gian gần đây khi biến thể Omicron lây lan khắp nước. Các cơ quan y tế Hàn Quốc dự báo đợt dịch hiện tại sẽ đạt đỉnh với khoảng 350.000 ca nhiễm một ngày vào khoảng giữa tháng 3.

Tuy nhiên, Chính phủ Hàn Quốc giờ đây đã từ bỏ chính sách cách ly và điều trị nghiêm ngặt. Họ chuyển sang tập trung vào việc giải quyết các ca nghiêm trọng và ngăn ngừa các trường hợp tử vong, vì biến thể Omicron dễ lây lan nhưng ít gây tử vong.

Source: tuoitre.vn

6   References

- What is Tableau? Working and Key Features, [Online: Analytics steps, 10th Apr]
- New COVID variant is spreading across the US. Here's what you need to know about BA.2 [Online: MSN, 12th Apr]
- Pháp, Chile ghi nhận các con số trái ngược về ca mắc mới COVID-19 [Online: Vietnam Plus, 14th Apr]
- Hàn Quốc ghi nhận trên 240.000 ca COVID-19 trong 3 ngày liền, thêm 3 triệu ca chỉ trong 1 tháng [Online: Tuoitre.vn, 14th Apr]