

Xây dựng báo cáo về tình hình Covid tại Việt Nam

Tổng quan bài tập

Trong ASM này học viên sẽ sử dụng Power BI để xây dựng báo cáo về tình hình COVID19 tại Việt Nam. Trong quá trình làm bài học viên sẽ được thực hành đầy đủ về các giai đoạn trong Workflow khi làm việc với Power BI. Một số nội dung chính như sau:

- Tiền xử lý dữ liệu
- Xây dựng báo cáo

Làm thế nào để hoàn thành bài asm?

- 1.Học viên đọc kĩ phần yêu cầu tổng quan sau đó có thể thực hiện các yêu cầu lần lượt theo hướng dẫn trong phần 2.
- 2.Học viên xem kỹ các phần tài liệu tham khảo.

Tài nguyên

Sẽ có 3 file dữ liệu như sau (trong thư mục data):

- Ca nhiễm theo ngày.csv: Tập dữ liệu này sẽ chứa thông tin dữ liệu theo từng ngày
- Ca nhiễm theo Tỉnh.csv: Tập dữ liệu này sẽ chứa thông tin ca nhiễm theo từng Tỉnh/Thành phố theo từng ngày
- Tổng ca nhiễm trong cả nước.csv: Tập dữ liệu này sẽ chứa thông tin về tổng ca nhiễm tại các tỉnh

Một số DAX function có thể sử dụng trong bài ASM:

- DATESINPERIOD(<Cột chứa ngày>, <Ngày làm mốc>, <Khoảng cách ngày>, <Đơn vị thời gian>):** Giúp trả về những giá trị thời gian nằm trong một khoảng tính từ Ngày bắt đầu tới Ngày cách ngày bắt đầu một khoảng nhất định.
- DATEADD(<Thời điểm làm mốc>, <Khoảng cách thời gian so với mốc>, <Loại đơn vị thời gian>):** Trả về một cột chứa các giá trị về thời gian (ngày, tháng, quý, năm) được chỉ định dựa theo mốc thời gian có sẵn.
- AVERAGEX(<Bảng>, <Biểu thức tính toán>):** Tính trung bình dựa trên <Biểu thức tính toán> trong <Bảng>).

Hướng dẫn hoàn thành dự án

Chức năng và yêu cầu chi tiết

1. Tiền xử lý dữ liệu

- Hãy đảm bảo tất cả các cột ở các file dữ liệu đều đúng định dạng dữ liệu.
- Đối với file **Ca nhiễm theo Tỉnh.csv** thực hiện các công việc sau:
 - Xoá dòng đầu tiên
 - Sử dụng Unpivot để chuyển đổi bảng dữ liệu về định dạng mẫu như sau:

Ngày	Tỉnh Thành	Ca nhiễm trong ngày
4/27/2021	Yên Bái	1
4/29/2021	TP HCM	1
4/29/2021	Hà Nam	5
4/30/2021	Hà Nội	3
4/30/2021	Hưng Yên	2
4/30/2021	Hà Nam	2
5/1/2021	Hà Nam	3
5/2/2021	Hà Nam	2
5/2/2021	Vĩnh Phúc	6
5/3/2021	Hà Nam	2
5/3/2021	Vĩnh Phúc	8

2. Xây dựng data model

2.1. Tạo bảng Thời gian

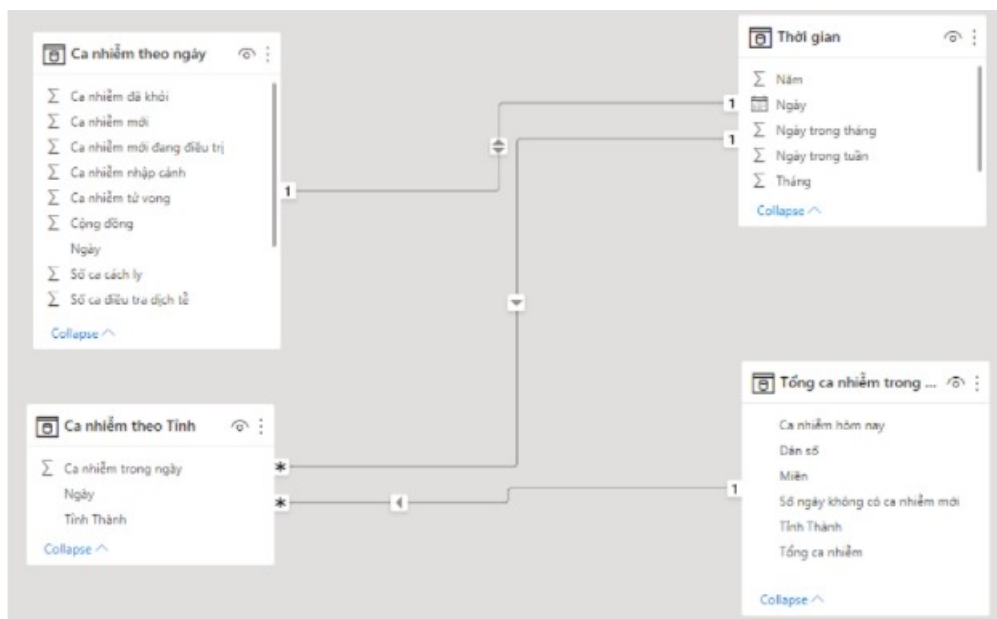
Tạo bảng thời gian với công thức như sau:

```
Thời gian = ADDCOLUMNS(
    CALENDAR(DATE(2021,1,1), MAX('Ca nhiễm theo ngày'[Ngày])),
    "Năm", YEAR([Date]),
    "Tháng", MONTH([Date]),
    "Tuần", FORMAT([Date], "ww"),
    "Ngày trong tuần", WEEKDAY([Date]),
    "Ngày trong tháng", DAY([Date])
)
```

Sau khi tạo bảng xong, hãy đổi tên cột **Date** thành **Ngày**.

2.2. Tạo liên kết giữa các bảng

- Cột **Ngày** trong bảng **Ca nhiễm trong ngày** liên kết với cột **Ngày** trong bảng thời gian.
- Cột **Ngày** trong bảng **Ca nhiễm theo Tỉnh** liên kết với cột **Ngày** trong bảng thời gian.
- Cột **Tỉnh Thành** trong bảng **Ca nhiễm theo Tỉnh** liên kết với cột **Tỉnh Thành** trong bảng **Ca Tổng ca nhiễm trong cả nước**.



2.3. Tạo các Measure

Tạo một bảng mới tên **_Measures** với công thức **_Measures = {""}** để lưu trữ các Measure, sau đó tạo các Measure sau:

- 1.Ca đã khỏi theo ngày
- 2.Ca nhiễm mới theo ngày
- 3.Ca tử vong theo ngày
- 4.Ca đang điều trị theo ngày
- 5.Tổng ca nhiễm đang điều trị
- 6.Ca đã khỏi trung bình trong 7 ngày, 30 ngày
- 7.Ca nhiễm mới trung bình trong 7 ngày, 30 ngày
- 8.Ca tử vong trung bình trong 7 ngày, 30 ngày
- 9.Số ca đã khỏi hôm nay (“ngày hôm nay” được hiểu là ngày lớn nhất trong bộ dữ liệu)
- 10.Số ca đã khỏi hôm qua (bằng “ngày hôm nay” trừ đi 1 ngày)
- 11.Số ca nhiễm mới hôm nay (“ngày hôm nay” được hiểu là ngày lớn nhất trong bộ dữ liệu)
- 12.Số ca nhiễm mới hôm qua (bằng “ngày hôm nay” trừ đi 1 ngày)
- 13.Số ca tử vong hôm nay (“ngày hôm nay” được hiểu là ngày lớn nhất trong bộ dữ liệu)
- 14.Số ca tử vong hôm qua (bằng “ngày hôm nay” trừ đi 1 ngày)
- 15.Số ca đang điều trị hôm nay (“ngày hôm nay” được hiểu là ngày lớn nhất trong bộ dữ liệu)
- 16.Số ca đang điều trị hôm qua (bằng “ngày hôm nay” trừ đi 1 ngày)
- 17.Tổng ca nhiễm/dân số (%)

Lưu ý: Trên đây là những Measure bắt buộc sẽ có, để tính các Measure này, học viên có thể tính trực tiếp thông qua các cột dữ liệu, hoặc thông qua các Measure khác (nếu thấy cần thiết hãy tạo thêm Measure khác để thông qua đó, có thể tính được các Measure theo yêu cầu).

Gợi ý:

- Để thực hiện các Measure 6,7,8: Học viên có thể sử dụng kết hợp hàm DATESINPERIOD, MAX và AVERAGEX
- Để thực hiện các Measure 11,13,15: Học viên có thể sử dụng kết hợp các hàm CALCULATE, FILTER, MAX
- Để thực hiện các Measure 12, 14, 16: Học viên có thể sử dụng kết hợp các hàm CALCULATE, DATEADD.

3. Xây dựng Dashboard

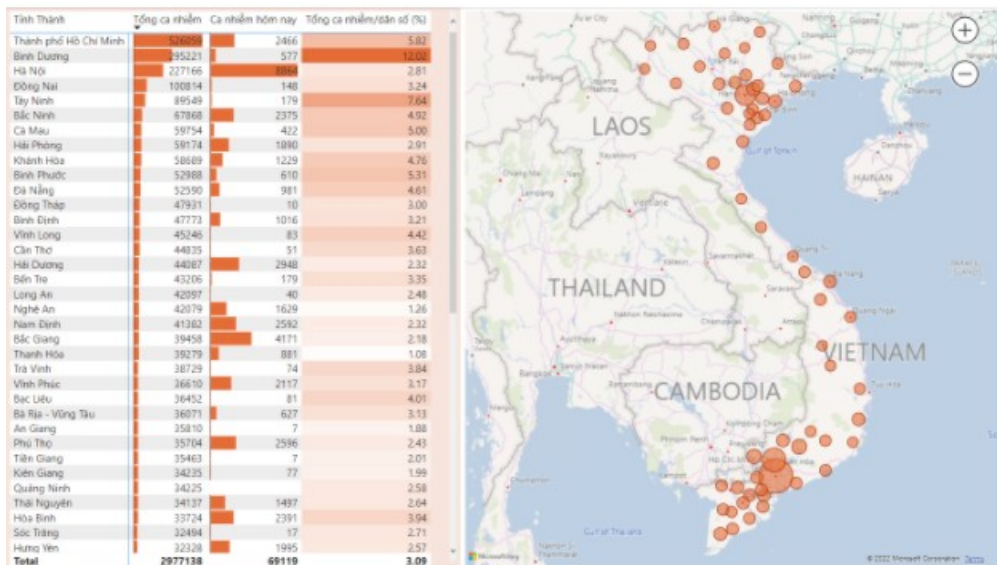
Học viên sẽ thực hiện xây dựng hai Dashboard, một là Dashboard **Tổng quan tình hình Covid19 tại Việt Nam**, một Dashboard là **Tình hình ca nhiễm tại các tỉnh thành**.

Đối với Dashboard **Tổng quan tình hình Covid19 tại Việt Nam**, Học viên sẽ xây dựng các biểu đồ thể hiện các thông tin sau:

- Ca nhiễm mới theo ngày
- Tổng ca nhiễm cộng dồn theo ngày
- Tổng ca nhiễm đang điều trị theo ngày
- Ca nhiễm đã khỏi theo ngày
- Ca nhiễm tử vong theo ngày



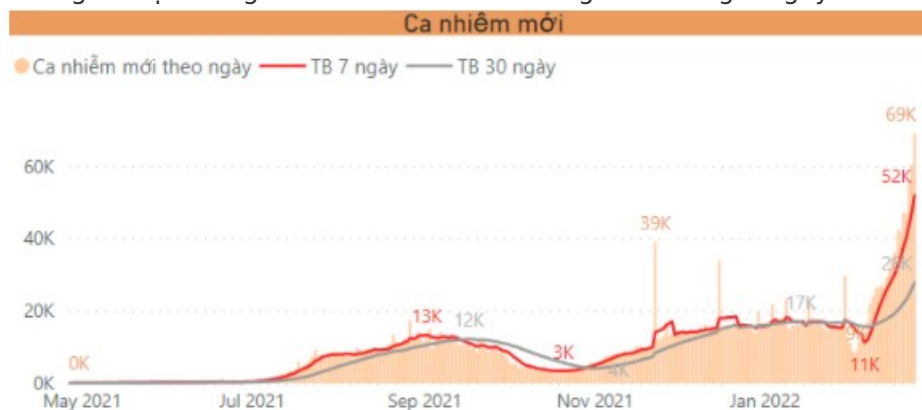
Đối với Dashboard **Tình hình ca nhiễm tại các tỉnh thành** học viên sẽ xây dựng bảng dữ liệu và Map dữ liệu thể hiện cho số ca nhiễm tại các tỉnh thành.



Học viên cần xây dựng giống như biểu đồ mẫu về loại biểu đồ, các thông tin được thể hiện trong từng loại biểu đồ, tên title, căn giữa các title, data label. Đối với bảng dữ liệu mẫu ở Dashboard **Tình hình ca nhiễm tại các tỉnh thành** học viên cần định dạng đúng cho các cột như bảng dữ liệu mẫu (Data bars, Background cho cột được sử dụng). Về màu sắc của các loại biểu đồ trong cả hai Dashboard, học viên có thể tùy chọn màu sao cho hợp lý và dễ quan sát.

3.1 Ca nhiễm mới theo ngày

Biểu đồ thể hiện ca nhiễm mới theo ngày sẽ có các cột dữ liệu là ca nhiễm mới theo ngày, các đường dữ liệu trung bình sẽ là ca nhiễm trung bình trong 7 ngày và 30 ngày.



3.2. Tổng ca nhiễm cộng dồn theo ngày

Tổng ca nhiễm cộng dồn theo ngày sẽ có dạng biểu đồ như sau:



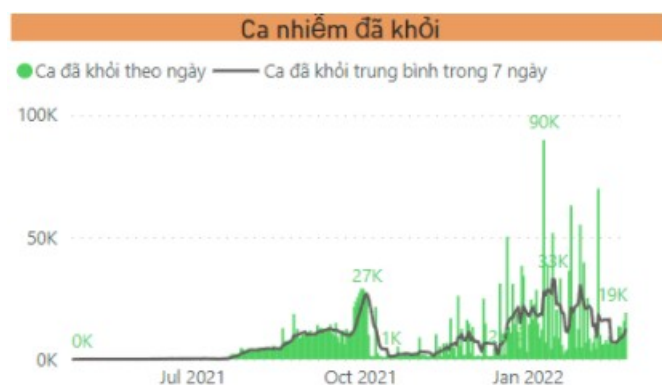
3.3. Tổng ca nhiễm đang điều trị theo ngày

Tổng ca nhiễm đang điều trị theo ngày sẽ có dạng biểu đồ như sau:



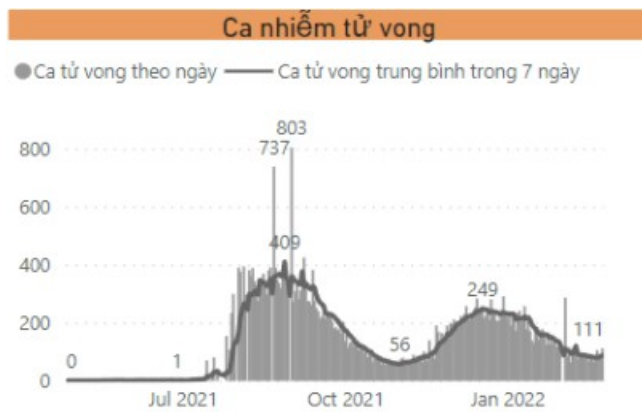
3.4. Ca nhiễm đã khỏi theo ngày

Biểu đồ thể hiện ca nhiễm đã khỏi theo ngày có dạng như sau:

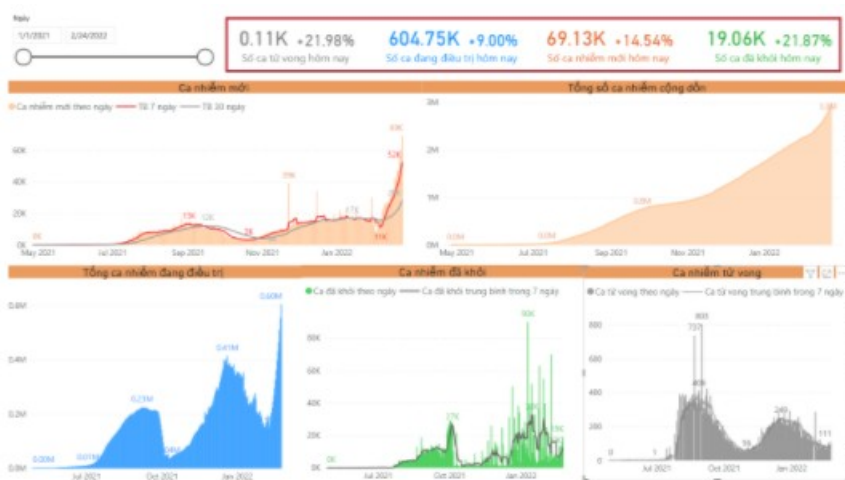


3.5. Ca nhiễm tử vong theo ngày

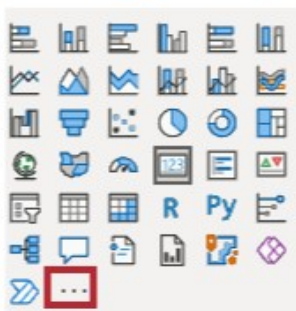
Biểu đồ thể hiện ca nhiễm tử vong theo ngày có dạng như sau:



3.6. Card thể hiện các thông tin



Loại card được sử dụng như hình trên có tên là Card with States by OKViz, Card này sẽ không có sẵn trong Power BI Desktop mà chúng ta phải thêm vào từ Store. Để thêm Card các bạn vào phần khoanh đỏ, sau đó chọn Get more Visuals và chọn mở từ file card_okviz.pbiviz.



Lưu ý:

- Nếu không di chuyển được card, hãy kéo dãn 4 góc của card để đưa card về đúng vị trí.

- Để format card giống như hình: chọn format your visual (tab thứ 2) -> data label -> reference type, đổi thử các loại reference type.

Ví dụ khi thể hiện thông tin về số ca nhiễm mới hôm nay:



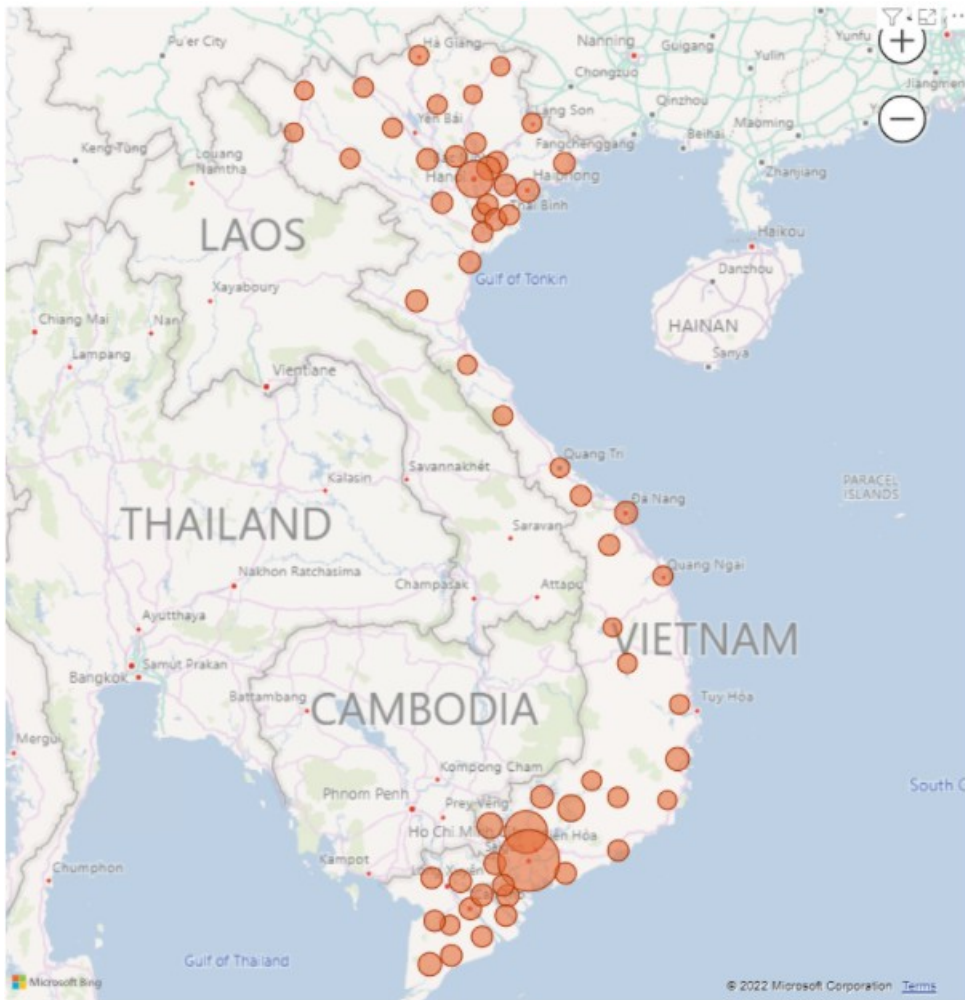
3.7. Bảng dữ liệu thể hiện ca nhiễm tại các Tỉnh/Thành phố

Bảng dữ liệu thể hiện ca nhiễm tại các Tỉnh/Thành phố có định dạng như sau:

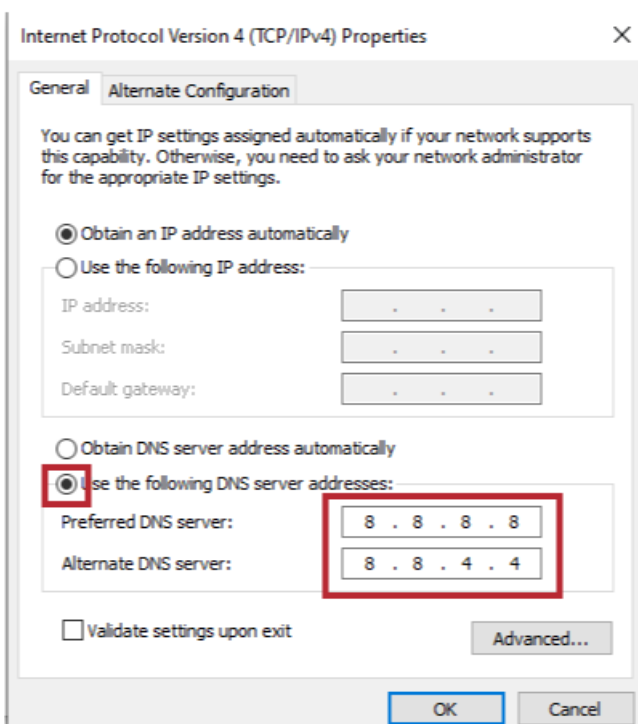
Tỉnh Thành	Tổng ca nhiễm	Ca nhiễm hôm nay	Tổng ca nhiễm/dân số (%)
Thành phố Hồ Chí Minh	526059	2466	5.82
Bình Dương	295221	577	12.02
Hà Nội	227166	8864	2.81
Đồng Nai	100814	148	3.24
Tây Ninh	89549	179	7.64
Bắc Ninh	67868	2375	4.92
Cà Mau	59754	422	5.00
Hải Phòng	59174	1890	2.91
Khánh Hòa	58689	1229	4.76
Bình Phước	52988	610	5.31
Đà Nẵng	52590	981	4.61
Đồng Tháp	47931	10	3.00
Bình Định	47773	1016	3.21
Vĩnh Long	45246	83	4.42
Cần Thơ	44835	51	3.63
Hải Dương	44087	2948	2.32
Bến Tre	43206	179	3.35
Long An	42097	40	2.48
Nghệ An	42079	1629	1.26
Nam Định	41382	2592	2.32
Bắc Giang	39458	4171	2.18

3.8. Map dữ liệu thể hiện tổng ca nhiễm trong cả nước

Map dữ liệu thể hiện tổng ca nhiễm trong cả nước có dạng như sau:



Lưu ý: Trong trường hợp Map không hiển thị, học viên sẽ thực hiện thay đổi DNS server trong máy cá nhân như sau:



Sau khi thay đổi xong, hãy khởi động lại Power BI trên Desktop. Giờ đây Map visual đã có thể được sử dụng.

3.9. Tạo Slicer theo ngày (Không bắt buộc)

Tạo Slicer theo ngày có dạng như sau:



Các tiêu chí đánh giá

STT	Tiêu chí	Đặc tả
Yêu cầu cơ bản		
1	Tiền xử lý dữ liệu	- Bảng dữ liệu từ file Ca nhiễm theo tỉnh.csv được định dạng đúng theo mẫu (Sử dụng Unpivot)
2	Tạo liên kết các bảng dữ liệu	- Tạo bảng Thời gian theo mã được cung cấp. - Các bảng dữ liệu được liên kết đúng theo yêu cầu.
3	Tạo Measure	Tạo được 17 Measure theo yêu cầu.
4	Xây dựng Dashboard	Xây dựng được hai Dashboard theo yêu cầu: - Đối với Dashboard Tổng quan tình hình Covid19 tại Việt Nam, xây dựng các loại biểu đồ sau: + Ca nhiễm mới theo ngày + Tổng ca nhiễm cộng dồn theo ngày + Tổng ca nhiễm đang điều trị theo ngày

		<div>+ Ca nhiễm đã khỏi theo ngày</div> <div>+ Ca nhiễm tử vong theo ngày</div> <div>+ Các Card thể hiện về số ca mới hôm nay, số ca tử vong hôm nay, số ca đang điều trị hôm nay, số ca đã khỏi hôm nay</div> <div>- Đối với Dashboard Tình hình ca nhiễm tại các tỉnh thành, xây dựng được bảng dữ liệu và map dữ liệu về tình hình covid19 tại các tỉnh thành</div> <div>- Đối với Dashboard Tổng quan tình hình Covid19 tại Việt Nam, học viên cần xây dựng giống như biểu đồ mẫu về loại biểu đồ, các thông tin được thể hiện trong từng loại biểu đồ, tên title, căn giữa các title, data label.</div> <div>- Đối với bảng dữ liệu mẫu ở Dashboard Tình hình ca nhiễm tại các tỉnh thành, học viên cần định dạng đúng cho các cột như bảng dữ liệu mẫu (Data bars, Background cho cột được sử dụng). Về màu sắc của cả hai Dashboard, học viên có thể tùy chọn màu sao cho hợp lý và dễ quan sát.</div>
--	--	---

Yêu cầu nâng cao

1	Tạo Slicer	Tạo Slicer theo ngày
---	------------	----------------------

Đáp án các Measure được sử dụng

Ca đã khỏi theo ngày = SUM(Ca nhiễm theo ngày[Ca nhiễm đã khỏi])

Ca đã khỏi trung bình trong 30 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -30, DAY), [Ca đã khỏi theo ngày])

Ca đã khỏi trung bình trong 7 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -7, DAY), [Ca đã khỏi theo ngày])

Ca đang điều trị theo ngày = SUM(Ca nhiễm theo ngày[Ca nhiễm mới đang điều trị])

Ca nhiễm mới theo ngày = SUM(Ca nhiễm theo ngày[Ca nhiễm mới])

Ca nhiễm mới trung bình trong 30 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -30, DAY), [Ca nhiễm mới theo ngày])

Ca nhiễm mới trung bình trong 7 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -7, DAY), [Ca nhiễm mới theo ngày])

Ca tử vong theo ngày = SUM(Ca nhiễm theo ngày[Ca nhiễm tử vong])

Ca tử vong trung bình trong 30 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -30, DAY), [Ca tử vong theo ngày])

Ca tử vong trung bình trong 7 ngày = AVERAGEX(DATESINPERIOD('Thời gian'[Ngày], MAX('Thời gian'[Ngày]), -7, DAY), [Ca tử vong theo ngày])

Số ca đã khỏi hôm nay = CALCULATE([Ca đã khỏi theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]))

Số ca đã khỏi hôm qua = CALCULATE([Ca đã khỏi theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]) - 1)

Số ca đang điều trị hôm nay = CALCULATE([Tổng ca nhiễm đang điều trị], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]))

Số ca đang điều trị hôm qua = CALCULATE([Tổng ca nhiễm đang điều trị], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]) - 1)

Số ca nhiễm mới hôm nay = CALCULATE([Ca nhiễm mới theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]))

Số ca nhiễm mới hôm qua = CALCULATE([Ca nhiễm mới theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]) - 1)

Số ca tử vong hôm nay = CALCULATE([Ca tử vong theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]))

Số ca tử vong hôm qua = CALCULATE([Ca tử vong theo ngày], 'Thời gian'[Ngày] = MAX('Thời gian'[Ngày]) - 1)

Tổng ca nhiễm cộng dồn = SUM(Ca nhiễm theo ngày[Tổng cộng dồn trong cộng dồn])

Tổng ca nhiễm đang điều trị = SUM(Ca nhiễm theo ngày[Tổng ca nhiễm đang điều trị])

Tổng ca nhiễm/dân số = SUM(Tổng ca nhiễm trong cả nước[Tổng ca nhiễm]) * 100 / SUM(Tổng ca nhiễm trong cả nước[Dân số])