**VIET NAM NATIONAL UNIVERSITY HCMC**

UNIVERSITY OF INFORMATION TECHNOLOGY

COMPUTER - SCIENCE

**BUSINESS INTELLIGENCE**

---------------o0o---------------

****

**PROJECT FINAL REPORT**

**DATA STORAGE**

**Assoc. Prof. Phuc Do**

**MSc. Nguyen Thi Kim Phung**

**Student Name: Nguyen Tran Gia Huy**

**Stdent ID: 20520106**

***ACKNOWLEDGEMENT***

Firstly, I would like to express my special appreciation and thanks to Mr. Do Phuc (Lecturer) and Miss. Le Thi Kim Phung (Teacher Assistant as well as our Mento). My project has been inspiring and provided me with various experiences of knowledge and technical skills relating to work and implement my project. Nonetheless, there are still have some mistakes that can not be avoid during this project implementation, fully and delivery it on time. In addition, I would like to appreciate some classmates who also support me in the process of doing this project.

I really looking forward to your comments and suggestions to improve our experiences as well as a valuable knowledge for the following projects.

Sincerely,

Ho Chi Minh City, August 22nd, 2021

**Nguyen Tran Gia Huy**

**MENTOR’S COMMENTS**

Ho Chi Minh City, ……. …………

**Mentor Signature**

**Contents**

1. Data introduction
2. Make SSIS
3. Make SSAS
4. Questions

**Descriptions**

* Microsoft has products for data integration (SSIS), analytics (SSAS), business intelligence (SSRS), and visualization. These tools can be deployed either stand-alone or as part of a Sharepoint system, can be fully integrated with the MS Office desktop products such as Excel.
* Using SSIS to collect the data
* Using SSAS to analyze the data

**What is SSAS and SSIS?**

**SQL Server Analytic Services (SSAS)**

SSAS provides online analytic processing (OLAP) of data from disparate data sources. SSAS allows users to analyze data with a host of tools including SSRS and Excel. In addition, SSAS enables the discovery of data patterns that may not be immediately apparent through the data mining features built into the product.

**SQL Server Integration Services (SSIS)**

Using SSIS, Innovent Solutions can help you implement an appropriate information management foundation that can deliver integrated, accurate, and timely data across your organization. Our data management solutions help ensure you provide trusted data in the areas of business intelligence, data warehousing, data migration, and master data management.

1. **DATA INTRODUCTION:**

**1.1 Get data from this link:**

[Latest COVID-19 Turkey Status | Kaggle](https://www.kaggle.com/xtweyz/latest-covid19-turkey-status)

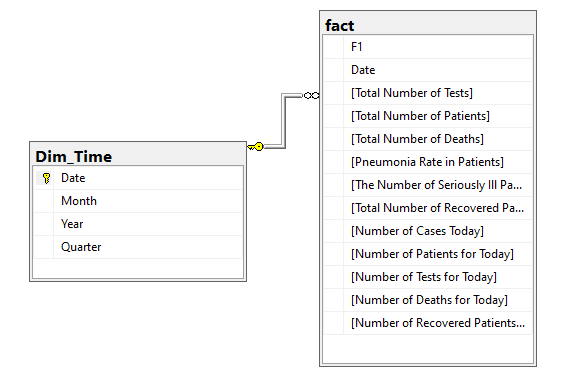
**1.2 Data description:**

This dataset contains latest Covid-19 status on Turkey from Turkey Ministry of Health as on August 17, 2020. This dataset can be used for analyze and visualize covid situtation in Turkey.

**1.3 Information of dataset:**

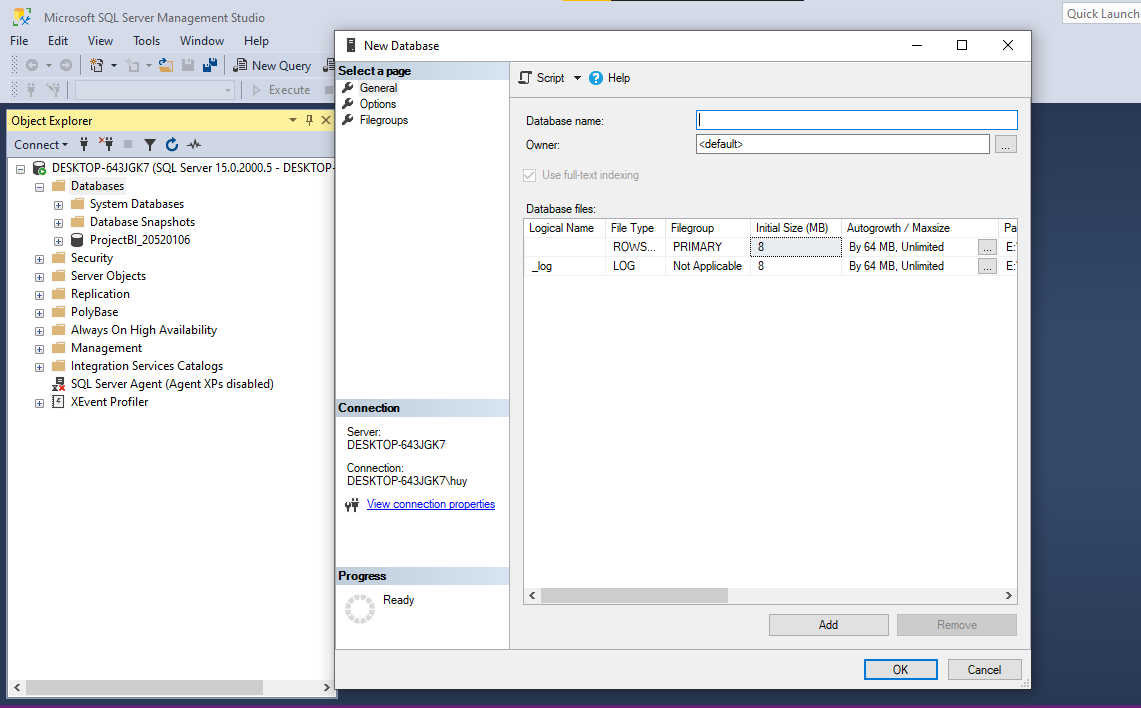
1. Date
2. Total Number of Tests
3. Total Number of Patiens
4. Total Number of Deaths
5. Pneumonia Rate in Patients (%)
6. The Number of Seriously Ill Patients
7. Total Number of Recovered Patients
8. Number of Cases Today
9. Number of Patients for Today
10. Number of Tests for Today
11. Number of Deaths for Today
12. Number of Recovered Patients for Today

**1.4 Diagram (StarSchema):**



1. **MAKING SSIS:**

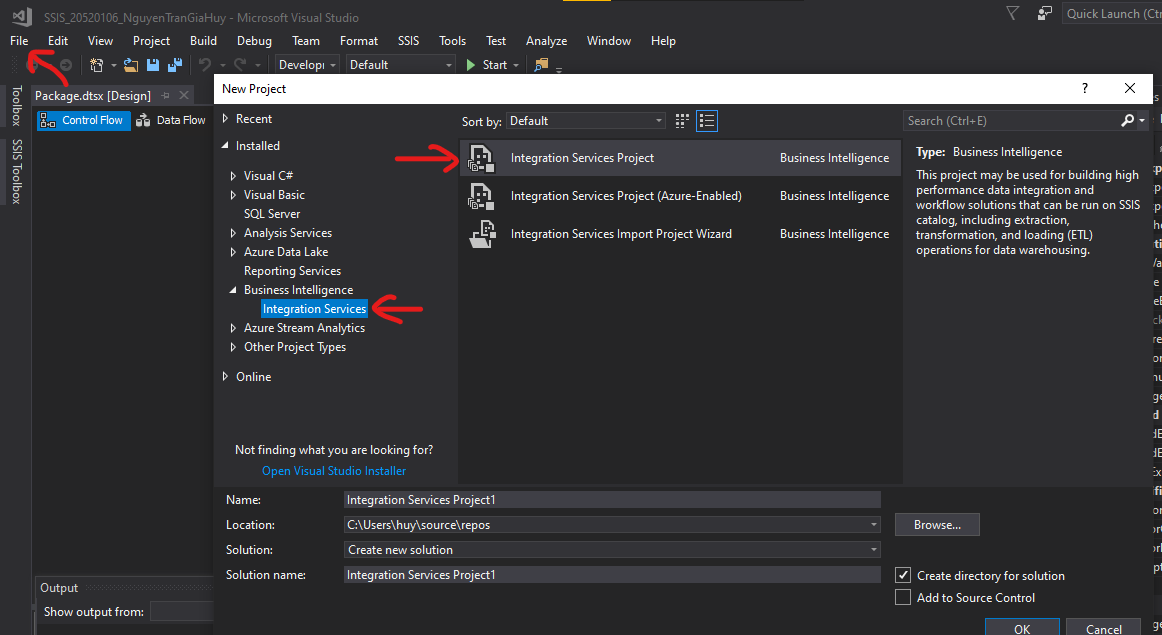
Open SQL Server Management Studio -> right click Database -> new database -> set Name for database.



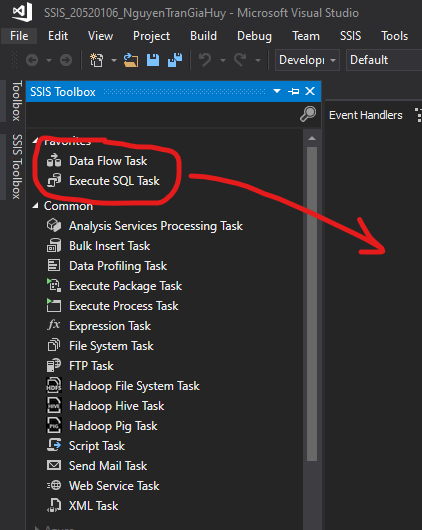
Note: In this project i have already created the database, which name is **ProjectBI\_20520106** in above image.

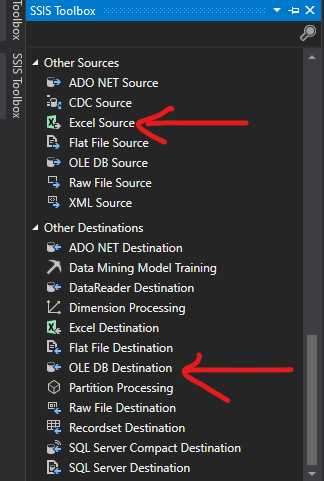
Then open Visual Studio and import data similar with below images.

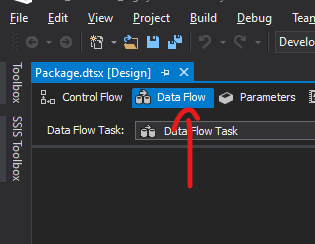
**Step 1:** Choose File -> new -> project -> choose Integration Services Project



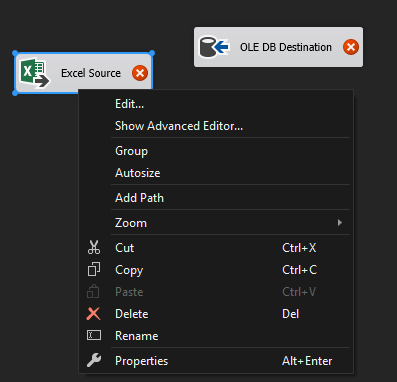
**Step 2:** Choose SSIS Toolbox in the left layout and move **Data Flow Task** and **Execute SQL Task** out the main layout.



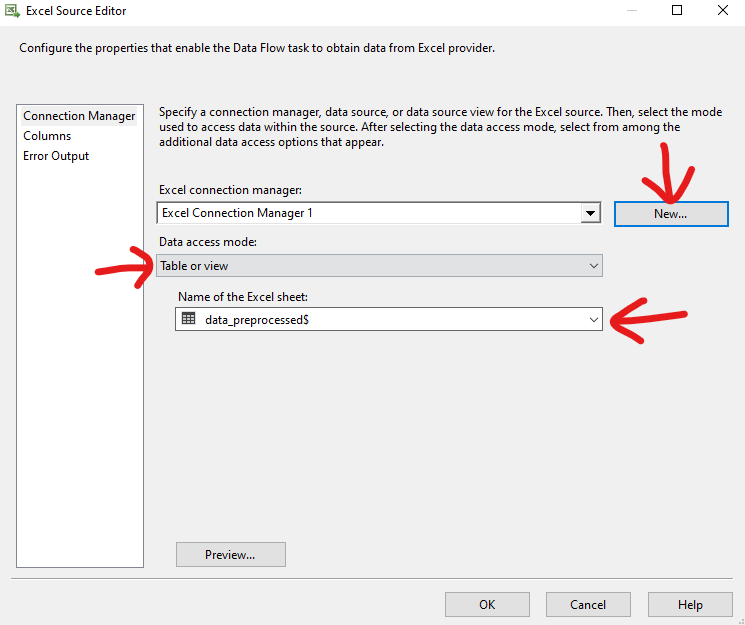
Then, click Data Flow, choose SSIS Toolbox, take **Excel Source** and **OLE DB Destination** to the main layout.



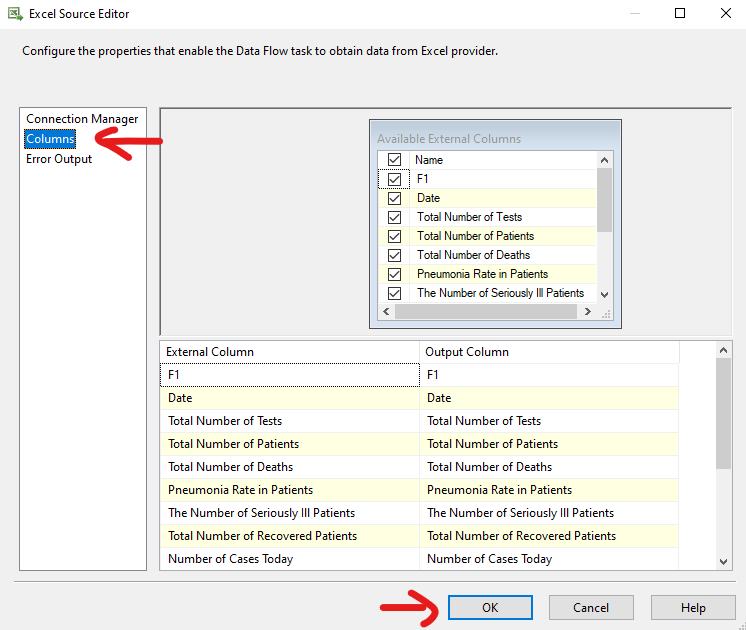
**Step 3:** Then, right click and choose edit the Excel Source.



**Step 4:** Choose **New** -> **Browse** database (data\_preprocessed.xls) you have and choose Table or view in **Data access mode** in below picture. 



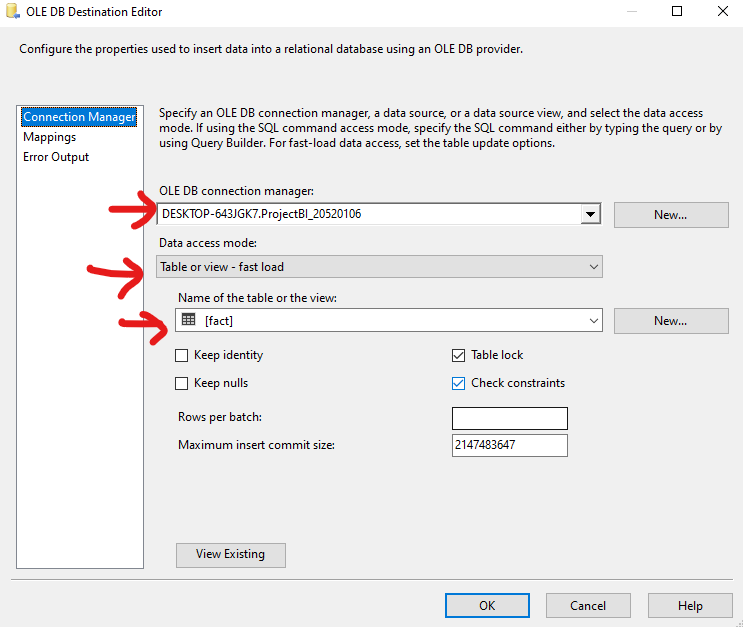
**Step 5:** Go to Columns to edit, because this is the Factable so i choose click all columns, then click Ok.



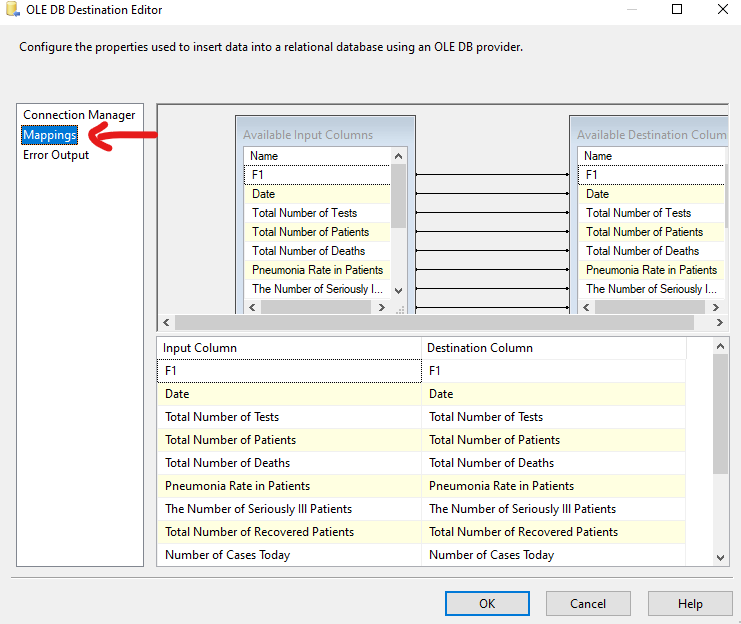
**Step 6:** Move the blue arrow connecting with the destination. However, we should change name its easier to recognize them.



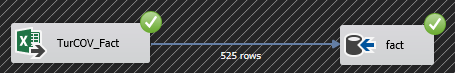
**Step 7:** Next, edit the destination **fact**, before that we go to SQL Sever -> Right click the database -> choose properties -> copy name of both **main server** and **project** and set up like these belowe pictures.



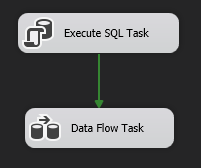
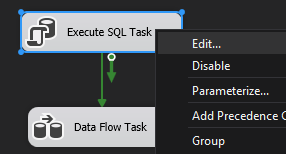
**Step 8:** Go to Mappings to check, if all the left data have already mapped with the right data, it is correct. And otherwise, right click the data which not mapped then it will connect after that.



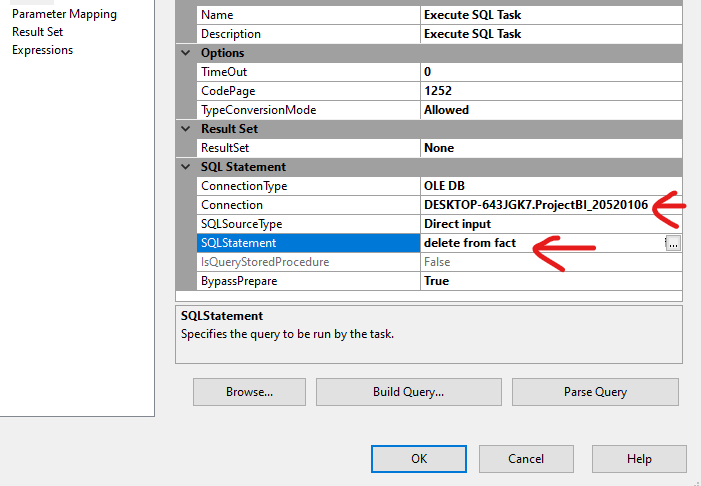
Click **Start** and running to check, then this is your result.



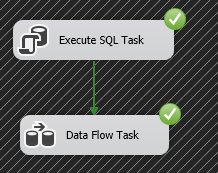
**Step 9:** Go back to the Control Flow, connecting the **Execute SQL Task** with **Data Flow Task**, thn click edit.

Choose the main server of database in **Connection** and fill “delete from fact” in **SQLStatement**.

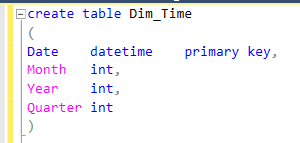


Then click Start again to check. So this is the result.

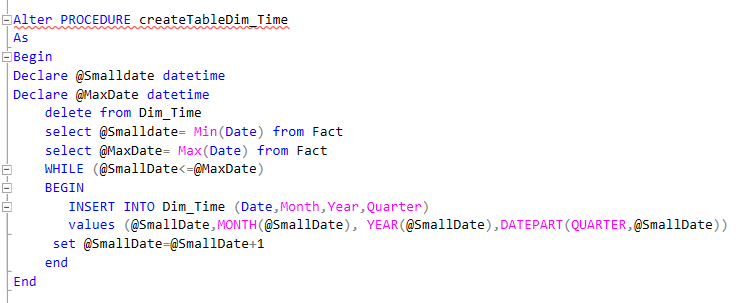


**Step 10:** Create dimension. Following this step, i try to do with the introdution of Miss. Le Thi Kim Phung (Teacher Assistant as well as our Mento) to create a Dim\_Time table with the code in SQL to seperate Date, Month, Year and Quater.

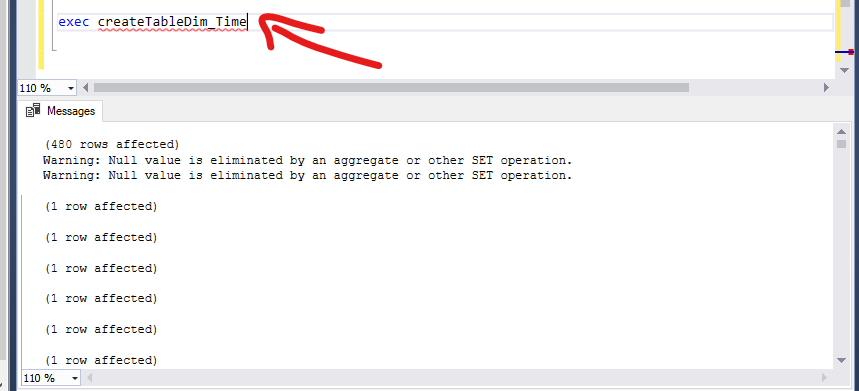
Firstly, i use this code and execute it to create a empty table and check.



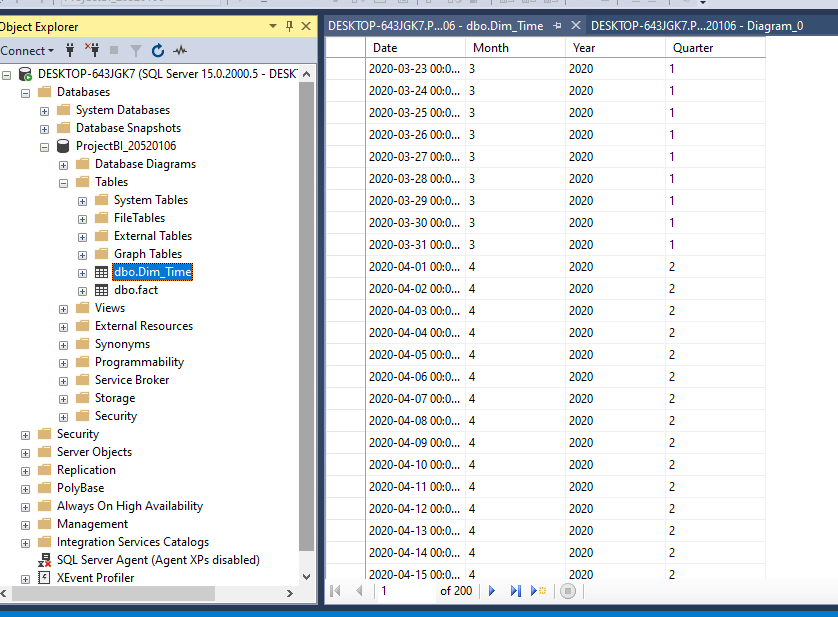
Secondly, deleting the table from the above image and executing this code to seperate the data with Date, Month, Year and Quater respectively.



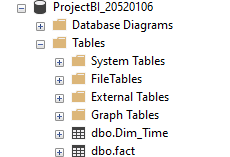
Finally, executing the last code to import all data to the table.



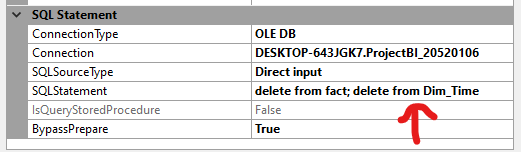
Right click dbo.Dim\_Time -> Edit Top 200 Rows -> Checking

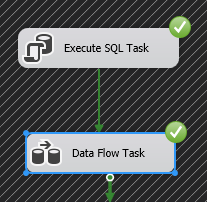
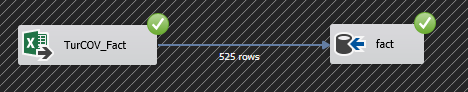


In this project, I create only 1 dimension and 1 fact.



Then, go to Visual Studio and fill “delete from Dim\_Time” one more time and run the program.



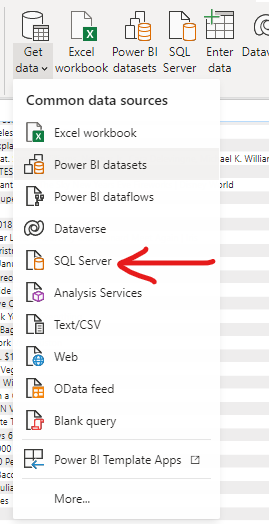
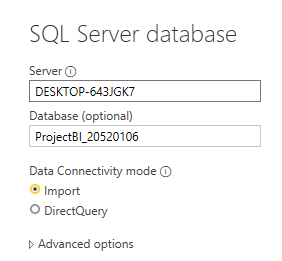
 

1. **MAKING SSAS:**

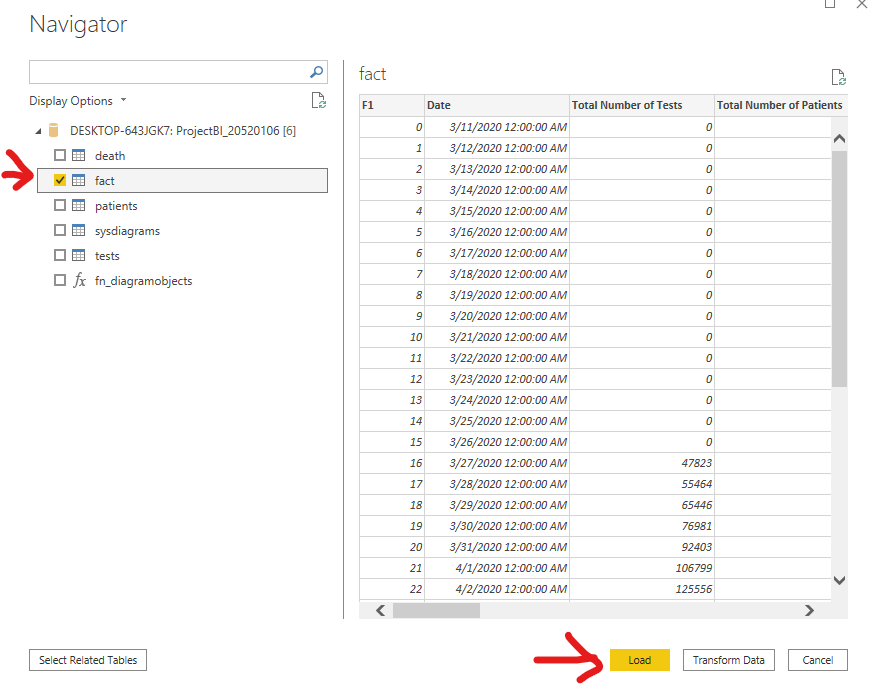
Open Power BI.



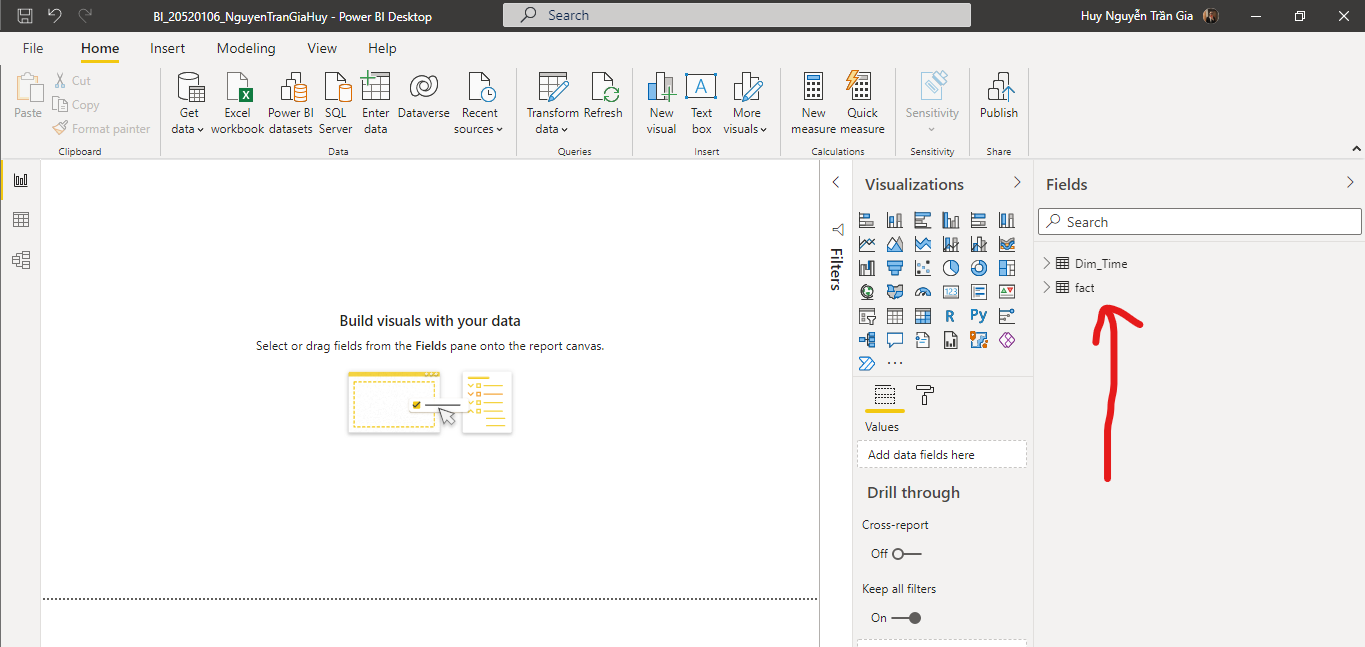
Choose Get data -> SQL Server -> Import name of main Server and Database -> Ok

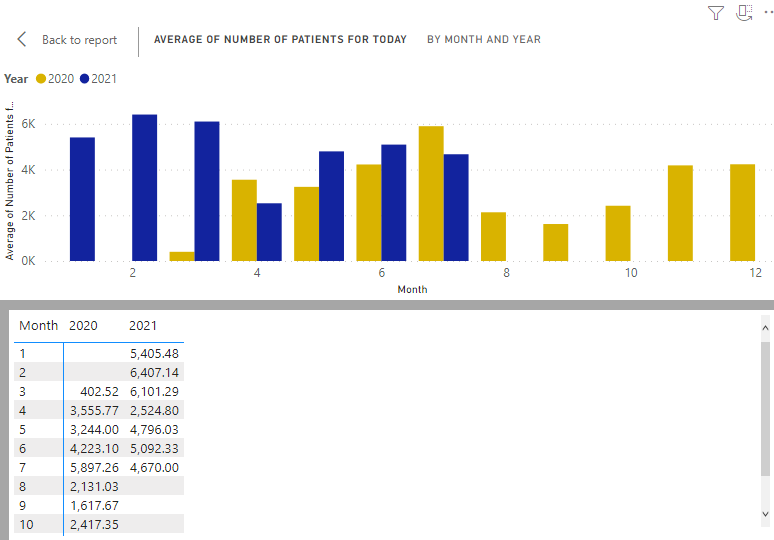
Click choose fact table -> click Load



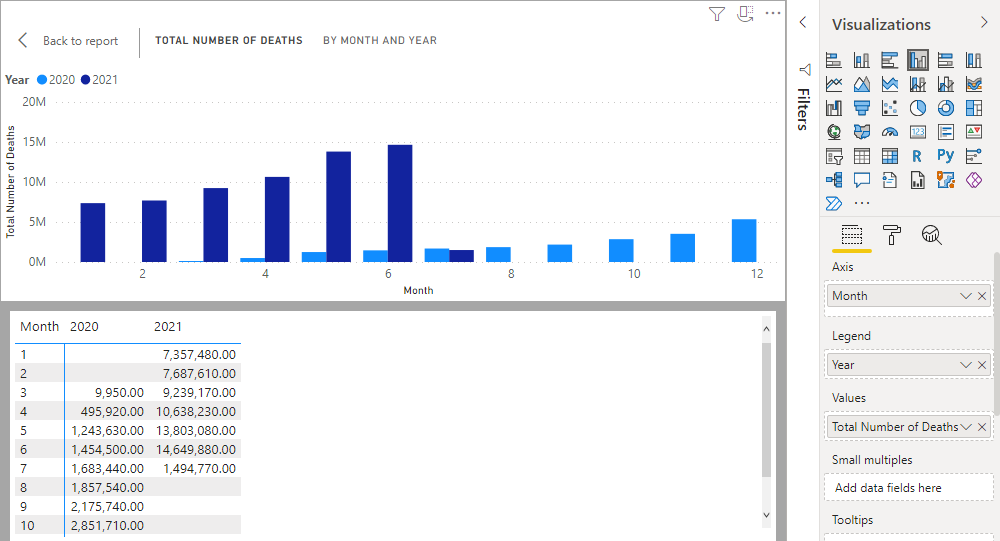
Complete set up Power BI, then this is the result.



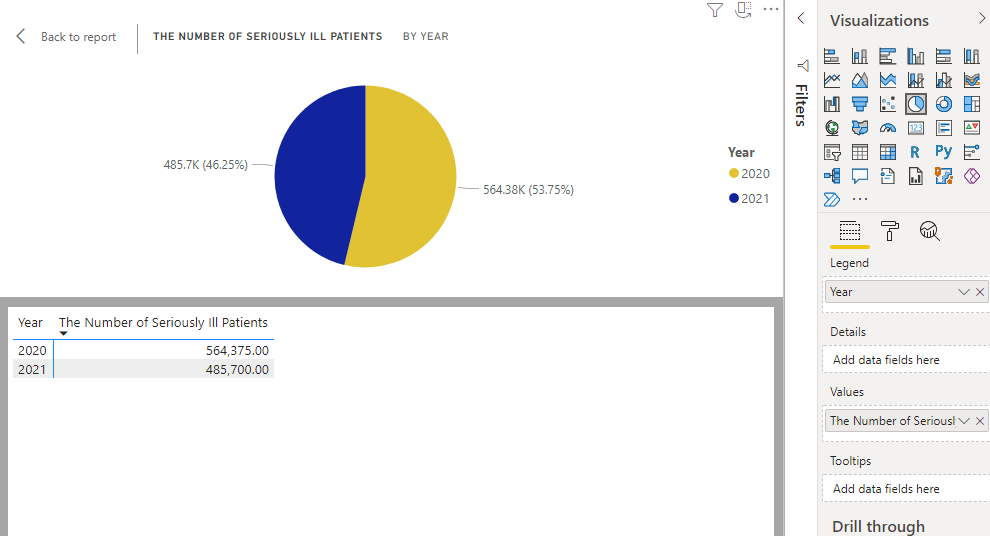
1. **QUESTIONS:** (8 questions)
2. This chart illustrates the Average Number of Patients for Today by Month and Year. The Covid19 outbreak from March in the year 2020 to July in 2021, so the Covid19 just started outbreak from March 2020, except for first 2 months. Additionally, this dataset import at this time (1st August 2021), so we do not have database for following months.

****

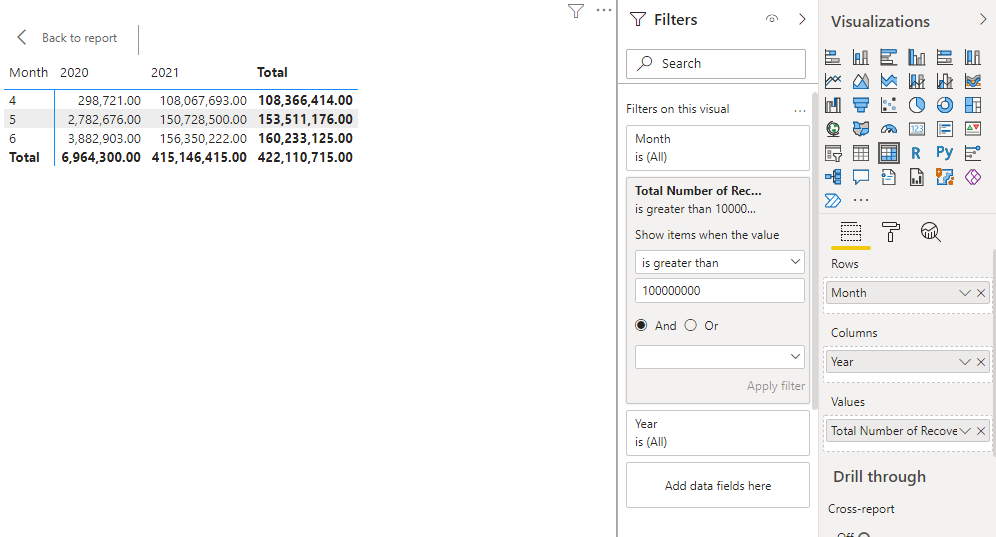
1. Comparison with the question 1, if the Total Number of Patients increased, the Total Number of Deaths also went up gradually from March in 2020 to July in 2021 (from 9,950 to 1,494,770 Number of Deaths).

****

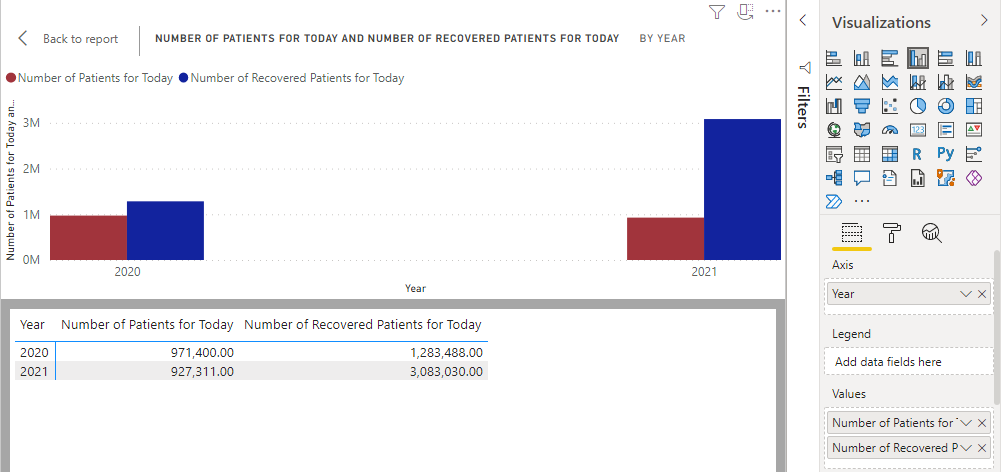
1. This pie graph represents the The Number of Seriously Ill Patients in 2 years 2020 and 2021. The Number of SIP in 2020 occupied above 53% at the figure 564,375 SIP, which was higher than 2021 (over 46% - 485,700 SIP).



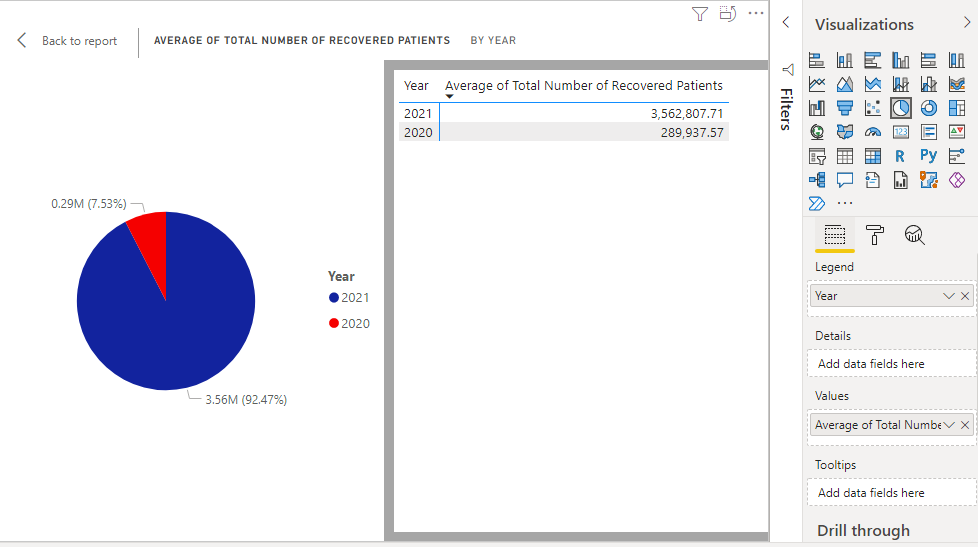
1. The Matric shows the the Total Number of Recovered Patients between 2 years 2020 and 2021, the result that 3 months (April, May and June) have large figures were 108366414, 153511176 and 160233125 respectively after using filters on the visual to find the number greater than 100000000.



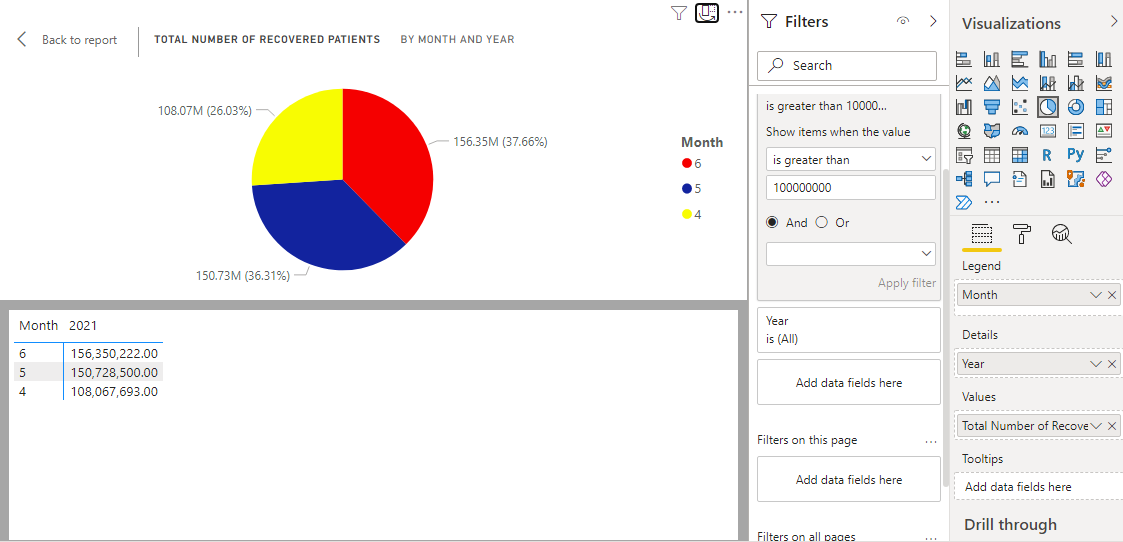
1. Comparison the Number of Patients for Today and Recovered Patients for Today from 971,400.00 and 927,311.00 in 2020 to 1,283.00 and 3,083,030.00 in 2021 namely. The Figure of Patients for Today in 2020 were higher than 2021, except that the Number of Recovered Patients in 2021 were the highest number and larger than 2020.



1. The Average of Total Number of Recovered Patients in the year 2021 accounted for over 92 percent, which was higher than 2020 at approximately 7.6%. Because The Covid19 outbreak from March in the year 2020, so the epidemic situation is getting more and more tense until August 2021.



1. The pie chart illustrates 3 months (April, May and June), which have the Total Number of Recovered Patients over 100000000 Recovered Patients in the year 2021 after using Filters (apply filter when the value is greater than 100000000).



1. The diagram shows Average of Number of Recovered Patients for Today from April 2020 (when the epidemic began to spread) until the present time in July 2021. Specially, the Average of Figure in April 2021 have the highest number at 41,960.40 Recovered Patients, which was comparison with other figures.

