Introduction

Hi, in this project, we will analyze COVID dataset. We will source data from a database that we will develop by joining and by making unions among tables from a uploaded excel file in MS SQL Server. Then we will explore the tables by writing appropriate SQL queries. At the end we will have prepared tables from executed SQL queries for further analysis using dashboards and visualizations.

Project Pipeline

So how are we going to achieve these objectives? By building a data analysis pipeline.

1. **Build a Database**
2. **Explore the data with SQL queries.**
3. **Prepare the custom tables for further analysis (using Power BI, Tableau)**

Let's complete these steps one by one:

Build a Database

Getting the Dataset:

- We will get the dataset from this website: **https://ourworldindata.org/covid-deaths**
- The dataset contains global COVID Data from January 28, 2020 to December 12, 2021. So, it is a pretty comprehensive data.
- The dataset contains 139,383 rows.

Initial Transformation of the Excel file:

- The format of the dataset is CSV, we will change it to excel.
- We will make 2 excel files from the original dataset.
- These datasets will be CovidDeaths and CovidVaccinations.
- We will use only this 2 excel files in the database.

Connecting to the MS SQL Server and Building the Database

- At first, we have to connect to MS SQL Server.
- Then right click on the Databases and select new database.
- Give a name to the database and press add. We will name as "COVID Project" to the database.
- Next, we will import our dataset (hotel_revenue_historical_full.xlsx) using SQL Server Import and Export Wizard.
    - We will choose Microsoft Excel as our data source.

- Then we will browse our Excel file path.
- We will use “Microsoft Excel 2016” version.
- Next, we will use SQL Server Native Client 11.0 as Destination.
- We will enter our database name as Projects.
- Then, we will choose “Copy Data from one or more tables or views” option.
- Now, it will bring us a list of all the tables, and we will select all.
- Finally, we will select Run immediately to finish the process.

- Now we have to refresh the “Projects” database to see the table names showing up in the Projects database.
- Finally, we have created our database.

Data Exploration

This step will be described in the SQL script file.

Prepare a view for further analysis

This step will be described in the SQL script file.