**COVID 19 DATA EXPLORATION**

1. **Introduction:**

In this project, we use SQL server to explore global COVID 19 data following the instruction of [Alex The Analyst](https://www.youtube.com/@AlexTheAnalyst).

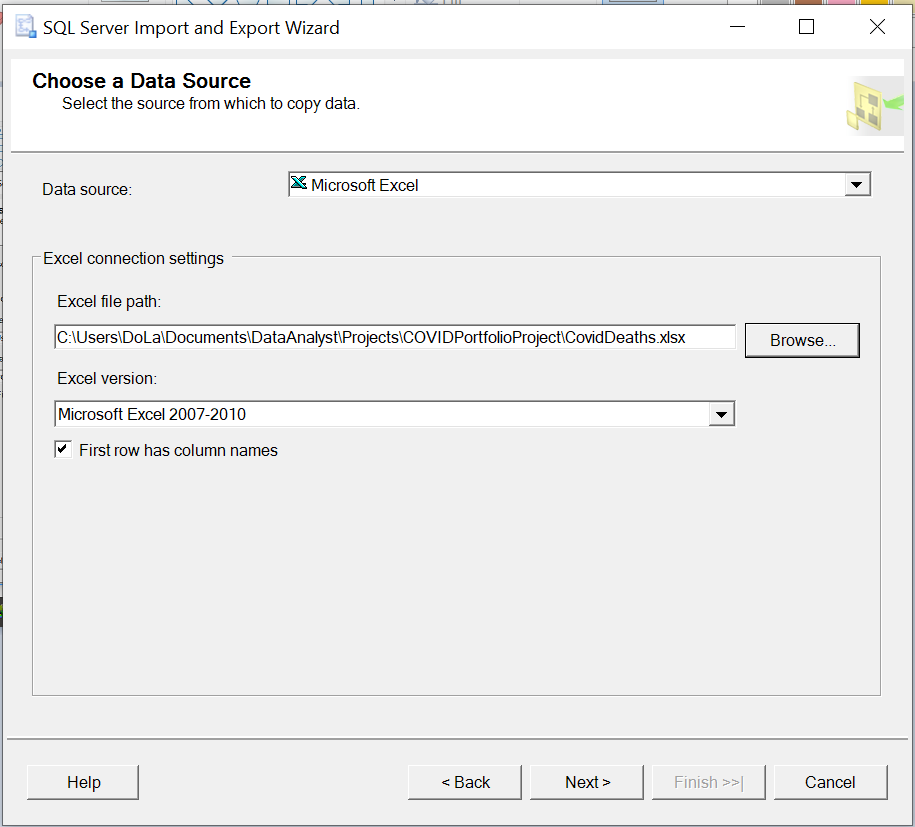
The dataset used in this project is the dataset used in [Data Analyst Portfolio Project | SQL Data Exploration | Project 1/4](https://www.youtube.com/watch?v=qfyynHBFOsM), which included 2 excel files [Covid Deaths](https://github.com/AlexTheAnalyst/PortfolioProjects/blob/main/CovidDeaths.xlsx) and [Covid Vaccinations](https://github.com/AlexTheAnalyst/PortfolioProjects/blob/main/CovidVaccinations.xlsx) contain covid data from January 28 2020 to April 30 2021.

The used dataset originally from [Coronavirus (COVID-19) Deaths - Our World in Data](https://ourworldindata.org/covid-deaths).

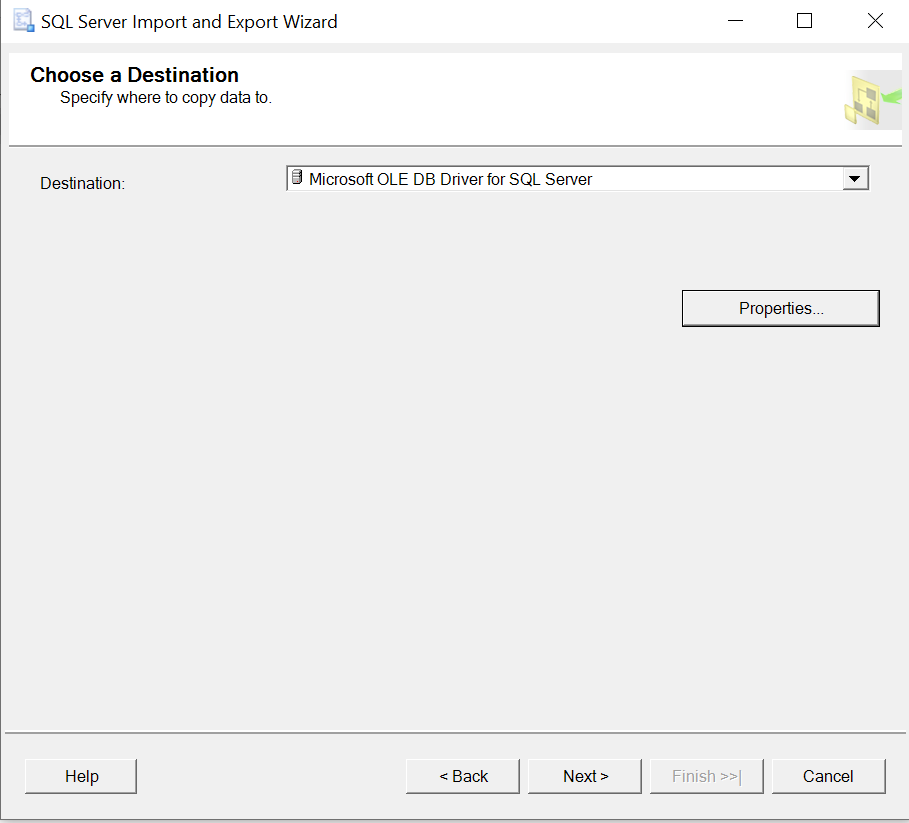
1. **Import Dataset into SQL Server:**

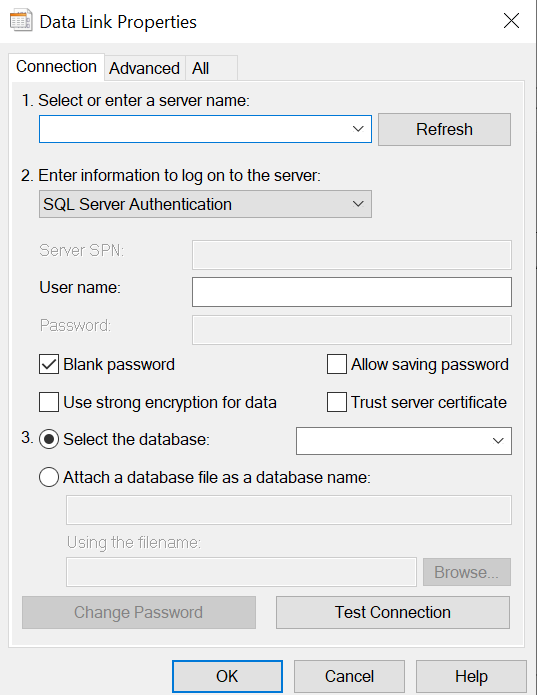
Use Import and Export Wizard in Microsoft SQL Server Management to import 2 excel files into SQL Server database.

* 1. Choose a Data Source. In this case, Microsoft Excel.

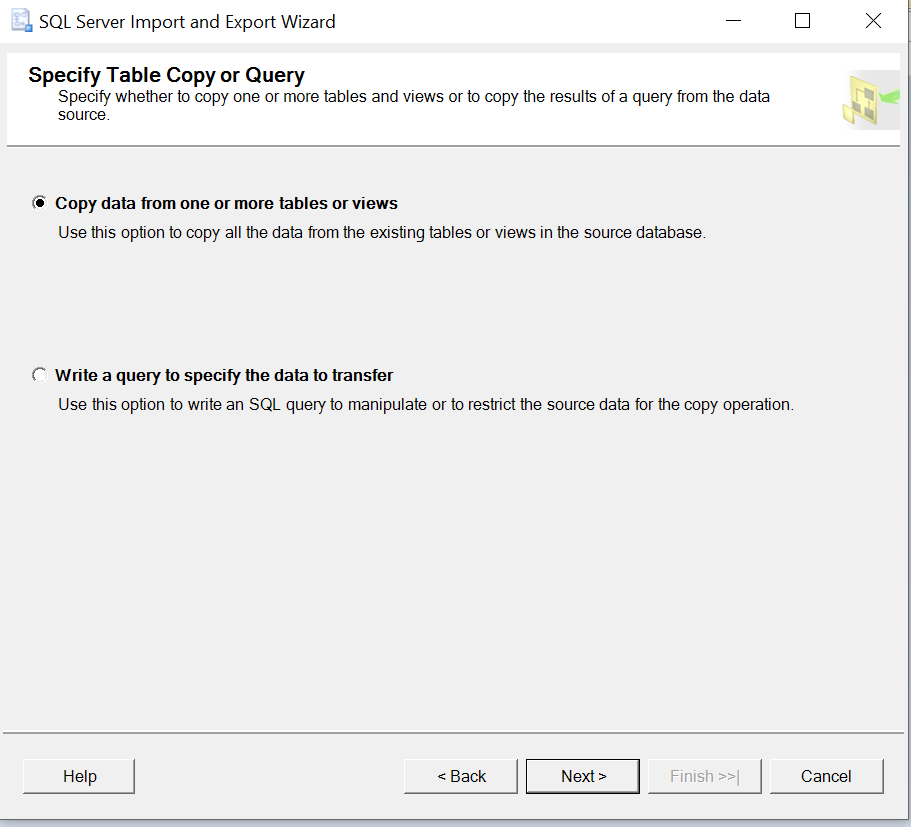


* 1. Choose a Destination, nn this case, Microsoft OLE DB Driver for SQL Server, and define Data Link Properties for the connection.

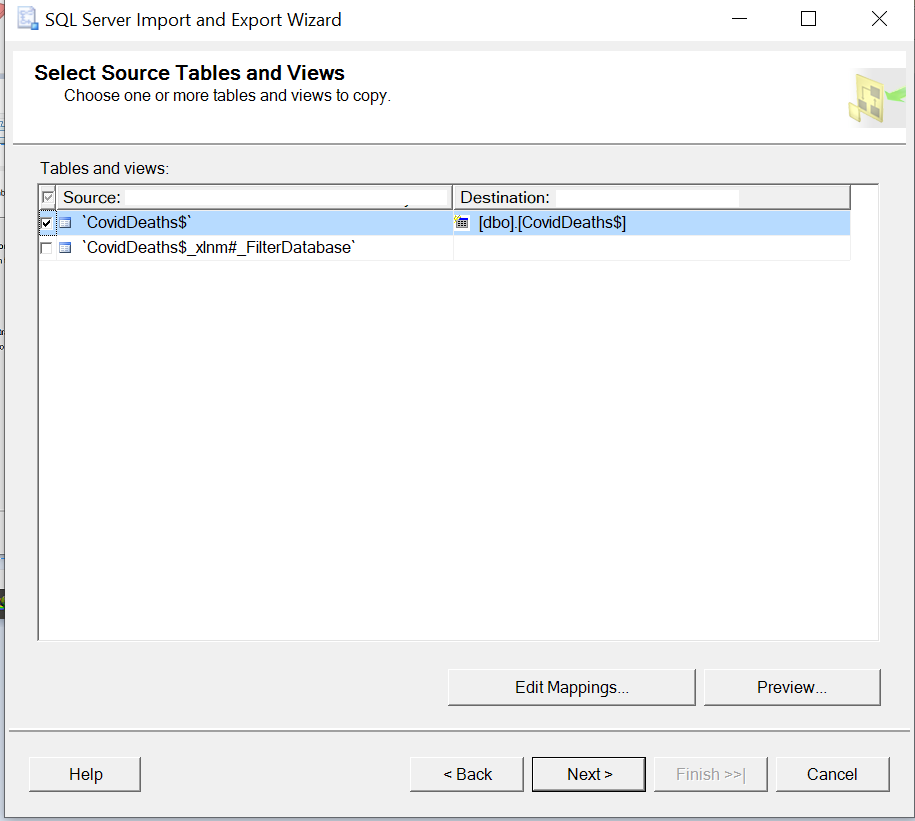




* 1. Choose Copy data from one or more tables or views.



* 1. Select wanted sheet as Source Tables.



Remember to check and edit data type of each column in Edit Mappings.

* 1. Run the import process.

1. **Data Exploration:**
2. **Total Cases vs Total Deaths:**

Find the likelihood of dying if you contract covid in each country.

select location, date, total\_cases, total\_deaths, (total\_deaths/total\_cases)\*100 as death\_percentage

from PortfolioProject..CovidDeaths

where continent is not null

order by location, date;

1. **Total Cases vs Population:**

Find the percentage of population infected with covid in each country.

select location, date, population, total\_cases, (total\_cases/population)\*100 as percent\_population\_infected

from PortfolioProject..CovidDeaths

where continent is not null

order by location, date;

1. **Countries with Highest Infection Rate compared to Population:**

Find countries with highest infection rate compared to population.

select location, population, max(total\_cases) as highest\_infection\_count, max((total\_cases/population))\*100 as highest\_percent\_population\_infected

from PortfolioProject..CovidDeaths

where continent is not null

group by location, population

order by highest\_percent\_population\_infected desc;

1. **Countries with Highest Death Count per Population:**

Find countries with highest death count per population.

select location, max(total\_deaths) as total\_death\_count

from PortfolioProject..CovidDeaths

where continent is not null

group by location

order by total\_death\_count desc;

1. **Contintents with the Highest Death Count per Population:**

Find contintents with the highest death count per population.

select continent, max(total\_deaths) as total\_death\_count

from PortfolioProject..CovidDeaths

where continent is not null

group by continent

order by total\_death\_count desc;

1. **Global Death Numbers:**

Find total infected cases, total deaths, death percentage across the world.

select sum(new\_cases) as total\_cases, sum(new\_deaths) as total\_deaths, (sum(new\_deaths)/sum(new\_cases))\*100 as death\_percentage

from PortfolioProject..CovidDeaths

where continent is not null;

1. **Total Population vs Vaccinations:**

Find population that has recieved at least one covid vaccine.

select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

sum(vac.new\_vaccinations) over (partition by dea.location order by dea.location, dea.date) as rolling\_people\_vaccinated

from PortfolioProject..CovidDeaths dea join PortfolioProject..CovidVaccinations vac

on dea.location = vac.location and dea.date = vac.date

where dea.continent is not null

order by dea.location, dea.date;

1. **Using CTE to perform Calculation on Partition By:**

Find percentage of population that has recieved at least one covid vaccine with CTE.

with pop\_vs\_vac (continent, location, date, population, new\_vaccinations, rolling\_people\_vaccinated)

as

(

select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

sum(vac.new\_vaccinations) over (partition by dea.location order by dea.location, dea.date) as rolling\_people\_vaccinated

from PortfolioProject..CovidDeaths dea join PortfolioProject..CovidVaccinations vac

on dea.location = vac.location and dea.date = vac.date

where dea.continent is not null

)

select \*, (rolling\_people\_vaccinated/population)\*100 as vaccinate\_percentage

from pop\_vs\_vac

order by location, date;

1. **Using Temp Table to perform Calculation on Partition By:**

Find percentage of population that has recieved at least one covid vaccine with temp table.

drop table if exists #percent\_population\_vaccinated

create table #percent\_population\_vaccinated

(

continent nvarchar(255),

location nvarchar(255),

date datetime,

population numeric,

new\_vaccinations numeric,

rolling\_people\_vaccinated numeric

)

insert into #percent\_population\_vaccinated

select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

sum(vac.new\_vaccinations) over (partition by dea.location order by dea.location, dea.date) as rolling\_people\_vaccinated

from PortfolioProject..CovidDeaths dea join PortfolioProject..CovidVaccinations vac

on dea.location = vac.location and dea.date = vac.date

where dea.continent is not null

select \*, (rolling\_people\_vaccinated/population)\*100 as vaccinate\_percentage

from #percent\_population\_vaccinated

order by location, date;