Data Exploration using SQL

Introduction

In this project, I performed data exploration using SQL to analyze COVID-19 data from various angles. I employed Microsoft Excel and SQL Server Management Studio (SSMS) to manipulate and query the dataset obtained from 'Coronavirus (COVID-19) Deaths - Our World in Data'. Below is a summary of my key findings and methodologies.

Initial Data Preparation

1. Downloaded the COVID-19 dataset from 'Our World in Data'.
2. Made data adjustments in Microsoft Excel, including rearranging columns and splitting data into two workbooks: CovidDeaths and CovidVaccinations.

Querying CovidDeaths Table

1. Retrieving Required Columns for First Query:

```
SELECT location, date, total_cases, new_cases, total_deaths, population
FROM dbo.CovidDeaths$
ORDER BY location, date;
```

2. Identifying First COVID-19 Case and Death in Pakistan:

```
SELECT location, date, population, total_cases, new_cases, total_deaths
FROM PortfolioProject..CovidDeaths
WHERE location LIKE '%Pakistan%' ORDER BY date;
```

- According to the dataset, first case reported on 28-02-2020, and the first death on 21-03-2020.

3. Finding Highest Death Count in Pakistan:

```
SELECT location, date, total_cases, total_deaths
FROM PortfolioProject..CovidDeaths
WHERE location LIKE '%Pakistan%'
ORDER BY total_deaths DESC;
```

- Highest deaths recorded on 30-12-2020 with 9992 deaths, as per the dataset.

COVID-19 Impact Analysis

4. Calculating Percentage of Affected Population in Pakistan:

```
SELECT location, date, total_cases, population, ((total_cases/population)*100) AS
AffectedPopulationPercentage

FROM dbo.CovidDeaths$

WHERE location LIKE '%Pakistan%'

ORDER BY date;
```

- Total affected population: 0.37%

These are just some of my observations based on the data queries and their results.