SQL PROJECT ON COVID-19 GLOBAL IMPACT ANALYSIS:

INSIGHTS ON INFECTION RATES, MORTALITY, AND VACCINATION PROGRESS

SQL PROJECT ON COVID-19 GLOBAL IMPACT ANALYSIS: INSIGHTS ON INFECTION RATES, MORTALITY, AND VACCINATION PROGRESS

INTRODUCTION

COVID-19, a highly infectious virus that emerged in late 2019, has left a lasting impact worldwide. The pandemic spread rapidly across countries, leading to unprecedented public health measures, economic shifts, and a global vaccination drive. To understand the scope and effects of COVID-19, I conducted a comprehensive analysis of global COVID-19 data from 2020 to 2021. This project, completed as part of Alex the Analyst's Boot Camp, utilized SQL to uncover insights into infection rates, mortality, and vaccination patterns across various countries.

PROJECT BACKGROUND & APPROACH

Dataset and Tools The dataset for this project was sourced from *Our World in Data*, encompassing global records of COVID-19 cases, deaths, and vaccinations. After some initial formatting in Excel to filter and organize columns, I imported two main tables, COVID Death and COVID Vaccination, into SQL for detailed analysis. This data, covering the years 2020 and 2021, allowed me to examine trends over the course of the pandemic.

Skills Applied: Throughout this project, I used a range of SQL skills, including:

- **Joins** to combine tables for integrated analysis.

- **Common Table Expressions (CTEs)** and **Temporary Tables** for organized and intermediary calculations.
- **Window Functions** for cumulative and rolling metrics, tracking changes over time.
- **Aggregate Functions** and **Views** for summarizing data insights.
- **Data Type Conversion** to ensure consistency in analysis.

ANALYSIS AND KEY INSIGHTS

1. Infection and Mortality Rates

- **Likelihood of Death per COVID Case**: One of the initial analyses focused on understanding the likelihood of death if infected with COVID-19. Calculating the percentage of deaths among total cases by country revealed that the global mortality rate remained constant at around **2%**, with no marked decline despite various containment measures.

```
Select Location, date, total_cases,total_deaths, (total_deaths/total_cases)*100 as DeathPercentage
From PortfolioProject..CovidDeaths
continent is not null
order by 1,2
```

- **Infection Rate by Population**: To gauge the virus's penetration in different countries, I calculated the infection rate relative to each country's population. Small nations such as Andorra, Montenegro, and the Czech Republic emerged as having some of the highest infection rates, underscoring regional vulnerabilities.

```
Select Location, date, Population, total_cases,  (total_cases/population)*100 as PercentPopulationInfected
From PortfolioProject..CovidDeaths'
order by 1,2
```

- **Highest Death Counts Relative to Population**: By analyzing death counts in proportion to population size, I identified countries with particularly high mortality rates, such as the United States, Brazil, Mexico, India, and the United Kingdom. Regionally, Europe, North America, South America, and Asia experienced the highest death counts, with Africa reporting fewer cases and deaths comparatively.

```
Select Location, MAX(cast(Total_deaths as int)) as TotalDeathCount
From PortfolioProject..CovidDeaths
Where continent is not null
Group by Location
order by TotalDeathCount desc
```

2. Global COVID-19 Mortality Statistics

Summing the global data provided a macroscopic view, with approximately **2%** of cases worldwide resulting in death over the analyzed period. This statistic reflects the global impact and severity of the virus, reinforcing the need for continued public health efforts and resources.

```
Select SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths, SUM(cast(new_deaths as int))/SUM(New_Cases)*100 as DeathPercentage
```

```
From PortfolioProject..CovidDeaths
where continent is not null
order by 1,2
```

3. Vaccination Rollout Analysis

- **Vaccination Rates by Population**: Examining the vaccination trends, I calculated the proportion of vaccinated individuals relative to each country's population. Using CTEs and temporary tables, I generated a rolling total to assess the progress of vaccination efforts, which were gradually increasing day by day.

```
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations
, SUM(CONVERT(int,vac.new_vaccinations)) OVER (Partition by dea.Location Order by dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
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 2,3


-- Using CTE to perform Calculation on Partition By in previous query
With PopvsVac (Continent, Location, Date, Population, New_Vaccinations, RollingPeopleVaccinated)
as
(
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations
, SUM(CONVERT(int,vac.new_vaccinations)) OVER (Partition by dea.Location Order by dea.location, dea.Date) as RollingPeopleVaccinated
```

```
--, (RollingPeopleVaccinated/population)*100
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 2,3
)
Select *, (RollingPeopleVaccinated/Population)*100
From PopvsVac

-- Using Temp Table to perform Calculation on Partition By in previous query
DROP Table if exists #PercentPopulationVaccinated
Create Table #PercentPopulationVaccinated
(
Continent nvarchar(255),
Location nvarchar(255),
Date datetime,
Population numeric,
New_vaccinations numeric,
RollingPeopleVaccinated numeric
)
Insert into #PercentPopulationVaccinated
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations
, SUM(CONVERT(int,vac.new_vaccinations)) OVER (Partition by dea.Location Order by dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
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 2,3

Select *, (RollingPeopleVaccinated/Population)*100
From #PercentPopulationVaccinated
```

- **Comparative Analysis of Vaccination Progress**: The data revealed significant differences in vaccination rates across countries, with some nations achieving rapid coverage, while others progressed more slowly, likely due to access disparities.

CONCLUSION

This SQL-based analysis of COVID-19 data provided a comprehensive view of the pandemic's progression and effects on different regions and demographics. Key findings highlighted infection and mortality rates by country, a global death rate of 2%, and steady but varied progress in vaccination rates. Through this project, I developed and applied SQL techniques crucial for turning raw data into meaningful insights, which can guide data-driven decision-making and inform responses to global health crises.

#COVID19 #SQLAnalysis #DataAnalytics #PublicHealth #DataDrivenInsights #OurWorldInData #PortfolioProject