



COVID DATA ANALYSIS PRESENTATION

TEJA ROJNIK MAHER

[LINK TO GITHUB SQL PROJECT](#)

MARCH 2022

TABLE OF CONTENT



Summary



Methodology



Visualization –
Charts



Dashboard

SUMMARY

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. As the virus has spread across the world with record speed

I have evaluated the number of cases, numbers of vaccinations and which countries have been impacted the most.

METHODOLOGY

Collecting data (download dataset from: <https://ourworldindata.org/covid-deaths>)

Re-format the data in Excel (create Covid Deaths, Covid Vaccinations file)

Data Wrangling in SQL (Duplicates and null values were detected and removed, data was standardized and formatted)

Data Analysis (Looking at total cases vs. total deaths, total cases vs. population, Countries with highest infection rate compared to population,...)

Data Visualization (Dashboard was created with different visual charts for easier understanding of the data)

```

--Select Data that we are going to be using
SELECT location,date,total_cases, new_cases, total_deaths, population
FROM PortfolioProject..CovidDeaths
ORDER BY 1,2 --this organizes it by location and date

--Looking at total cases vs. total deaths
SELECT location,date,total_cases, total_deaths, (total_deaths/total_cases)*100 as DeathPercentage
FROM PortfolioProject..CovidDeaths
WHERE location = 'United States'
ORDER BY 1,2

-- Looking at total cases vs population
SELECT location,date, population, total_cases, (total_cases/population)*100 as DeathPercentage
FROM PortfolioProject..CovidDeaths
WHERE location = 'United States'
ORDER BY 1,2

--Looking at Countries with highest infection rate compared to population
SELECT location, population, MAX(total_cases) AS HighestInfectionCount,
MAX((total_cases/population))*100 as PercentPopulationInfected
FROM PortfolioProject..CovidDeaths
--WHERE location = 'United States'
GROUP BY location, population
ORDER BY PercentPopulationInfected desc

--Showing countries with highest death count per population
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

```

EXPLORATION OF DATASET

- Skills used: Joins, CTE's, Temp Tables, Windows Functions, Aggregate Functions, Creating Views, Converting Data Types

EXPLORATION OF DATASET

--GLOBAL NUMBERS / USING AGGREGATE FUNCTIONS

```
SELECT date, SUM(new_cases)
FROM PortfolioProject..CovidDeaths
WHERE continent is not null
GROUP BY date
ORDER BY 1,2  --this will give us numbers on each day for the whole world
```

```
SELECT date, SUM(new_cases), SUM(new_deaths)
FROM PortfolioProject..CovidDeaths
WHERE continent is not null
GROUP BY date
ORDER BY 1,2  --error -not working because new cases column is a float - we have to change it from varchar to integer SEE BELOW
```

```
SELECT date, SUM(new_cases), SUM(cast(new_deaths as int))
FROM PortfolioProject..CovidDeaths
WHERE continent is not null
GROUP BY date
ORDER BY 1,2
```

--global numbers for total cases, total deaths and death percentage, by date

```
SELECT date, 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
GROUP BY date
ORDER BY 1,2
```

--Total global number for total cases, total deaths and death percentage

```
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
```

```
SELECT * from PortfolioProject..CovidVaccinations
```


EXPLORATION OF DATASET

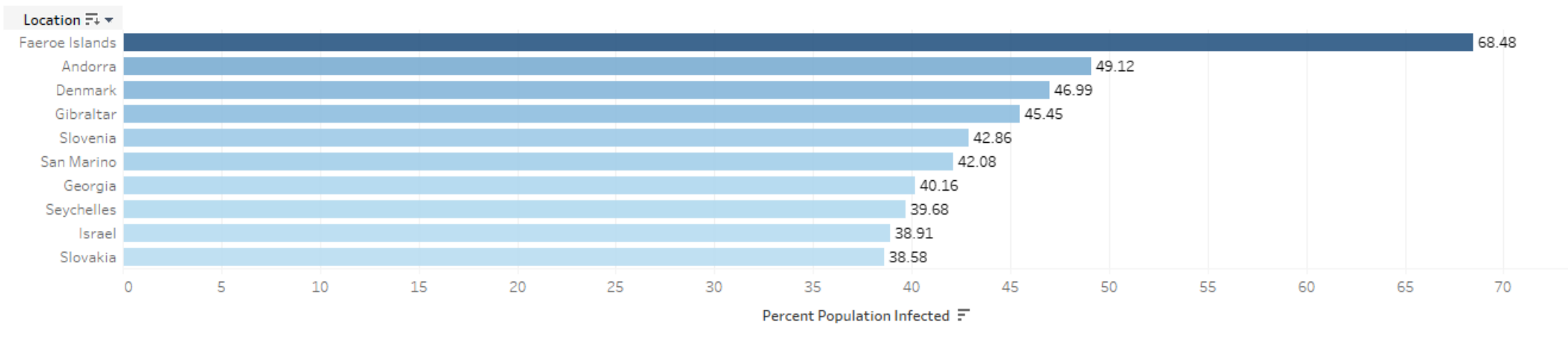
```
--join tables
SELECT *
from PortfolioProject..CovidDeaths dea
join PortfolioProject..CovidVaccinations vac
on dea.location = vac.location
and dea.date = vac.date

--Looking at total population vs vaccinations
SELECT dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations
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

-- Lets JOIN two tables from two data sets(covid vaccinations and covid deaths)
--we will join them on location and on date
Select *
from PortfolioProject..CovidDeaths dea --a little alias so it is shorter
join PortfolioProject..CovidVaccinations vac
on dea.location = vac.location
and dea.date = vac.date --check if it joined them correctly, then continue with analysis

--practice
SELECT dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations --new vaccinations per day
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
```

Top 10 Countries with the Highest Percent of Population Infected

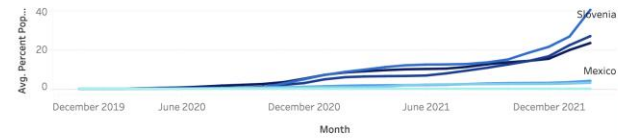


Covid Dashboard Practice by Teja Rojnik Maher

Global Numbers

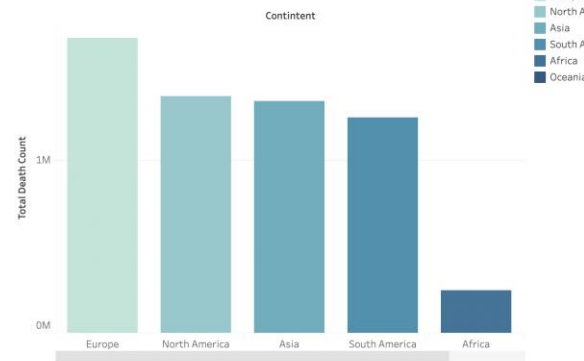
Total Cases	Total Deaths	Death Percentage
432,847,624,000	5,917,680,000	1.37

Percent Population Infected



Location
China India Mexico Slovenia United Kingdom United States

Total Deaths Per Continent



Percent Population Infected Per Country



DASHBOARD