PANDEMIC IN NUMBERS

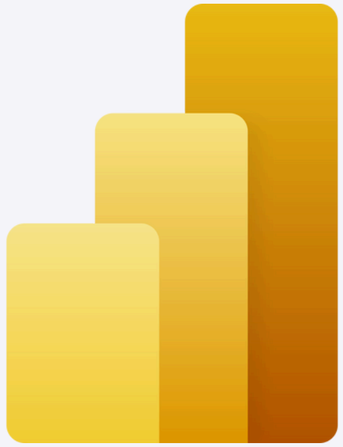
COVID-19 Data Exploration with SQL

Hendryawan Cipta Mulia
hendryawan.cipta@gmail.com
www.linkedin.com/in/hendryawan

ABOUT PROJECT

Explore COVID-19 data using SQL and present visualizations with Power BI.

TOOLS USED IN THIS PROJECT



Power BI
For Data Visualization

SQL Server
For Data Analysis

DATA OVERVIEW

Table 1

covid_deaths

- iso_code
- continent
- location
- date
- population
- total_cases
- new_cases
- new_cases_smoothed
- total_deaths
- new_deaths
- new_deaths_smoothed
- total_cases_per_million
- new_cases_per_million
- new_cases_smoothed_per_million
- total_deaths_per_million
- new_deaths_per_million
- new_deaths_smoothed_per_million
- reproduction_rate
- icu_patients
- icu_patients_per_million
- hosp_patients
- hosp_patients_per_million
- weekly_icu_admissions
- weekly_icu_admissions_per_million
- weekly_hosp_admissions
- weekly_hosp_admissions_per_million

26 Columns

Table 2

covid_vacs

- iso_code
- continent
- location
- date
- new_tests
- total_tests
- total_tests_per_thousand
- new_tests_per_thousand
- new_tests_smoothed
- new_tests_smoothed_per_thousand
- positive_rate
- tests_per_case
- tests_units
- total_vaccinations
- people_vaccinated
- people_fully_vaccinated
- new_vaccinations
- new_vaccinations_smoothed
- total_vaccinations_per_hundred
- people_vaccinated_per_hundred
- people_fully_vaccinated_per_hundred
- new_vaccinations_smoothed_per_million
- stringency_index
- population_density
- median_age
- aged_65_older
- aged_70_older
- gdp_per_capita
- extreme_poverty
- cardiovasc_death_rate
- diabetes_prevalence
- female_smokers
- male_smokers
- handwashing_facilities
- hospital_beds_per_thousand
- life_expectancy
- human_development_index

37 Columns

Total

85172 Rows

KEY SQL QUERIES

SQL Exploration Highlights

```sql
-- Global Cases Numbers in 2020 and 2021
SELECT
    YEAR(date) as year,
    SUM(COALESCE(new_cases, 0)) AS total_cases
FROM
    project_portfolio..covid_deaths
WHERE
    continent IS NOT NULL
GROUP BY
    YEAR(date)
ORDER BY
    1 DESC;
```

	year	total_cases
1	2021	67839879
2	2020	82735098

```sql
-- Global Death Numbers in 2020 and 2021
SELECT
    YEAR(date) as year,
    SUM(COALESCE(new_deaths, 0)) AS total_deaths
FROM
    project_portfolio..covid_deaths
WHERE
    continent IS NOT NULL
GROUP BY
    YEAR(date)
ORDER BY
    1 DESC;
```

	year	total_deaths
1	2021	1355296
2	2020	1824910

```sql
-- Global Case Fatality Rate During 2020 and 2021
SELECT
    YEAR(date) AS year,
    SUM(new_deaths) AS total_deaths,
    SUM(new_cases) AS total_cases,
    ROUND(SUM(new_deaths)/SUM(new_cases)*100, 3) AS death_rate
FROM
    project_portfolio..covid_deaths
WHERE
    continent IS NOT NULL
GROUP BY YEAR(date)
ORDER BY
    1 ASC;
```

	year	total_deaths	total_cases	death_rate
1	2020	1824910	82735098	2,206
2	2021	1355296	67839879	1,998

And many more...

```sql
--Top 5 Country with Highest Total Case in 2020-2021
SELECT TOP 5
    location,
    SUM(COALESCE(new_cases, 0)) AS total_cases
FROM
    project_portfolio..covid_deaths
WHERE
    continent IS NOT NULL
GROUP BY
    location
ORDER BY
    2 DESC;
```

	location	total_cases
1	United States	32346970
2	India	19164969
3	Brazil	14659011
4	France	5677835
5	Russia	4750755

```sql
-- Top 5 Country with Highest Case Fatality Rate in 2020-2021
SELECT TOP 5
    location,
    SUM(COALESCE(new_deaths, 0)) AS total_death,
    SUM(COALESCE(new_cases, 0)) AS total_cases,
    ROUND((SUM(COALESCE(new_deaths, 0)) * 1.0 / NULLIF(SUM(COALESCE(new_cases, 0)), 0)) * 100, 2) AS death_rate
FROM
    project_portfolio..covid_deaths
GROUP BY
    location
ORDER BY
    4 DESC;
```

	location	total_death	total_cases	death_rate
1	Vanuatu	1	4	25
2	Yemen	1226	6317	19,41
3	Mexico	216907	2344755	9,25
4	Syria	1592	22733	7
5	Sudan	2349	33944	6,92

```sql
-- Top 10 Country with Highest Vaccination Rate in 2020-2021
SELECT
    cd.location,
    cd.population,
    SUM(cv.new_vaccinations) as total_vaccinated,
    ROUND((SUM(cv.new_vaccinations)/cd.population*100), 4) as vaccination_rate
FROM
    project_portfolio..covid_deaths cd
    JOIN project_portfolio..covid_vacs cv
    ON cd.location = cv.location AND cd.date = cv.date
WHERE
    cd.continent IS NOT NULL
GROUP BY
    cd.location, cd.population
HAVING
    SUM(cv.new_vaccinations) IS NOT NULL
ORDER BY 4 DESC
OFFSET 2 ROWS
FETCH NEXT 10 ROWS ONLY;
```

	location	population	total_vaccinated	vaccination_rate
1	United Arab Emirates	9890400	9474906	95,799
2	Chile	19116209	14766907	77,2481
3	Malta	441539	317491	71,9055
4	Maldives	540542	376996	69,7441
5	United States	331002647	227590819	68,758
6	United Kingdom	67886004	46070991	67,8652
7	Bhutan	771612	480422	62,2621
8	Hungary	9660350	5419479	56,1002
9	Isle of Man	85032	45192	53,1471
10	Uruguay	3473727	1833763	52,7895

DASHBOARD OVERVIEW

SUMMARY

Between April 2020 and March 2021, the global COVID-19 pandemic resulted in **150.6M total cases**, **3.18M deaths**, and a **4.2% case fatality rate.** Europe reported the highest number of cases among continents (**44.86M**), while the United States led among countries with **32.3M cases.**

Cases peaked globally around the **12th month** but declined before rising again.

Europe, Asia, and North America were most affected, collectively accounting for over

120M

cases

The **US**, **India**, and **Brazil** experienced the highest case counts, emphasizing disparities in pandemic impact.

Let's Connect!

Excited to leverage my skills in SQL and data exploration to uncover valuable insights!

LinkedIn: Hendryawan Cipta Mulia | GitHub: Hendryawan99 | Instagram: @gwe.aan