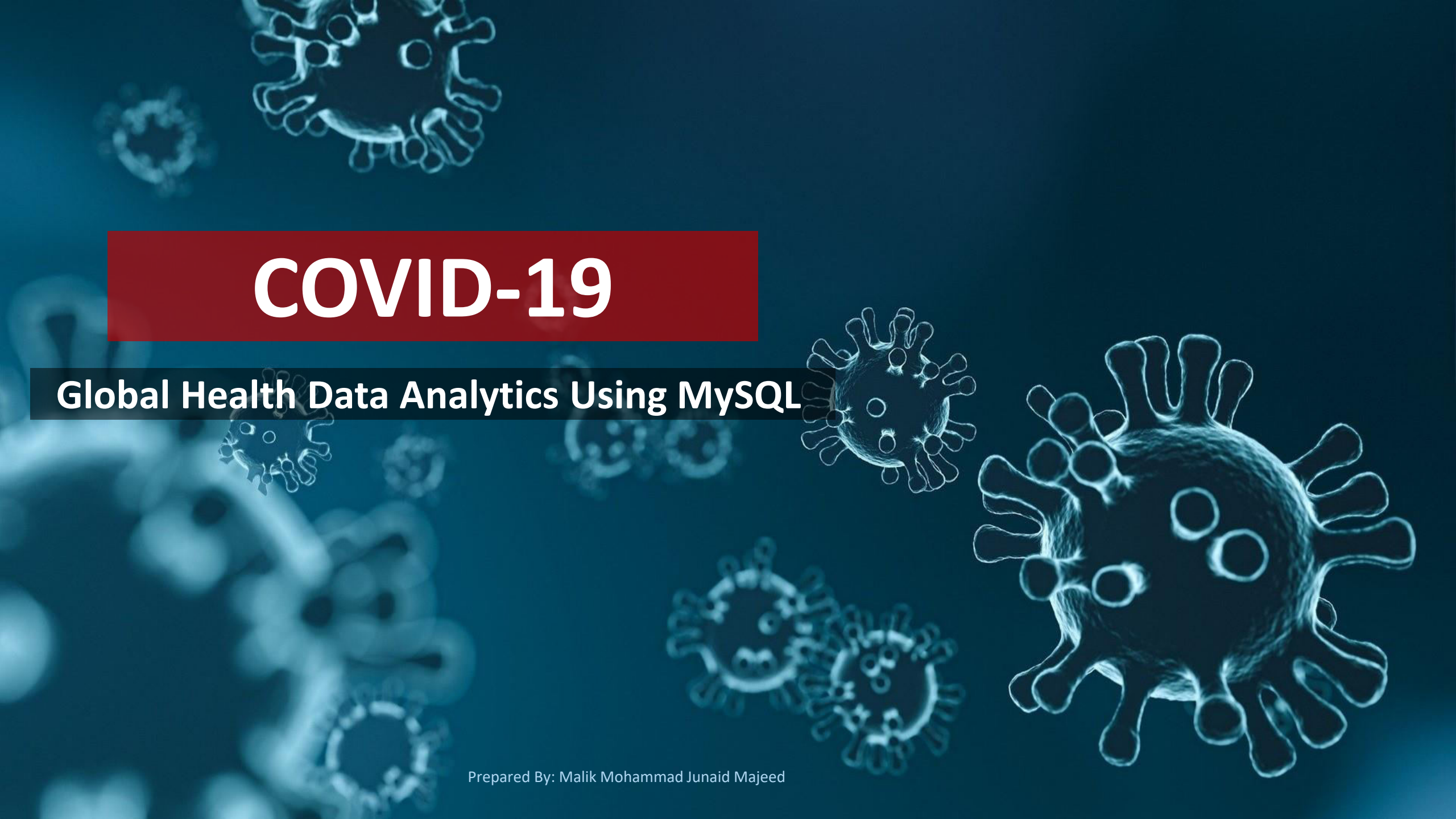


COVID-19

Global Health Data Analytics Using MySQL

Prepared By: Malik Mohammad Junaid Majeed

INTRODUCTION

This presentation showcases a series of advanced SQL queries performed to analyze global health data, with a focus on identifying key trends, high-risk regions, and critical performance metrics. The analysis leverages **MySQL's aggregation, filtering, and analytical capabilities** to transform raw data into actionable insights.

- **Global Total Cases, Deaths, and Recovery Rate** – Summarizing overall impact on a global scale.
- **Top 10 Countries with Highest Active Cases** – Identifying regions with the most ongoing infections.
- **Daily New Cases Globally** – Tracking the daily growth and spread of infections.
- **Case Fatality Rate by Country** – Evaluating the severity of the outbreak across nations.
- **Top 10 Countries by Highest Deaths** – Highlighting the most affected countries by fatalities.
- **Daily Growth Rate per Country** – Analyzing the trend of infections in each country.
- **Peak Cases Date per Country** – Identifying when the outbreak peaked regionally.
- **Mortality vs Recovery Correlation per Country** – Exploring the relationship between deaths and recoveries.

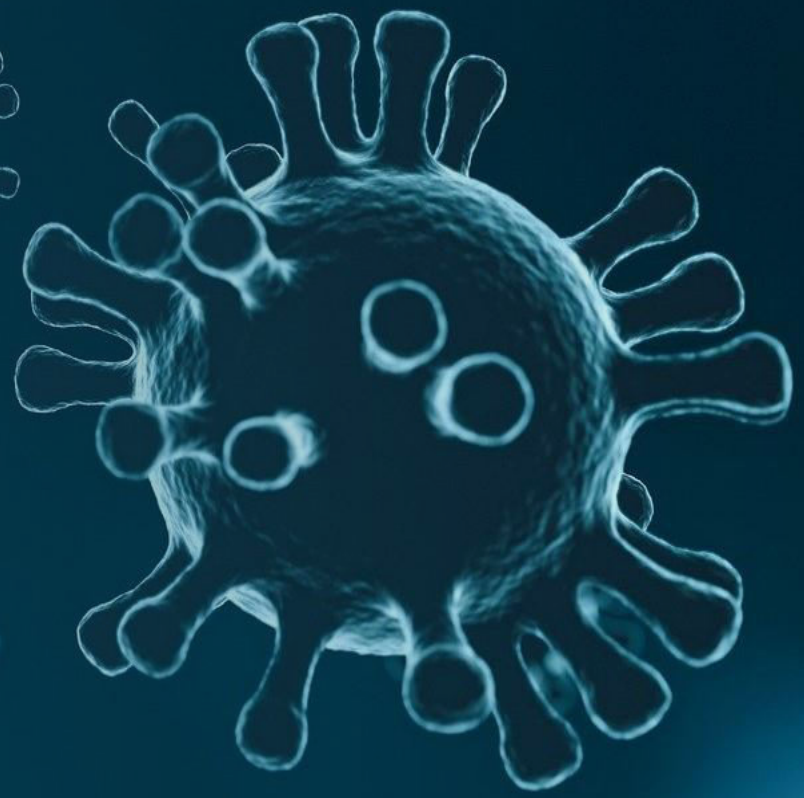
This comprehensive analysis provides a **data-driven understanding of the global situation**, enabling better forecasting, risk assessment, and strategic decision-making.

1. Global Total Cases, Deaths, and Recovery Rate

```
SELECT
    SUM(confirmed) AS total_cases,
    SUM(deaths) AS total_deaths,
    SUM(recovered) AS total_recovered,
    ROUND(SUM(deaths) * 100.0 / NULLIF(SUM(confirmed), 0), 2) AS death_rate,
    ROUND(SUM(recovered) * 100.0 / NULLIF(SUM(confirmed), 0), 2) AS recovery_rate
    FROM covid_19_data;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

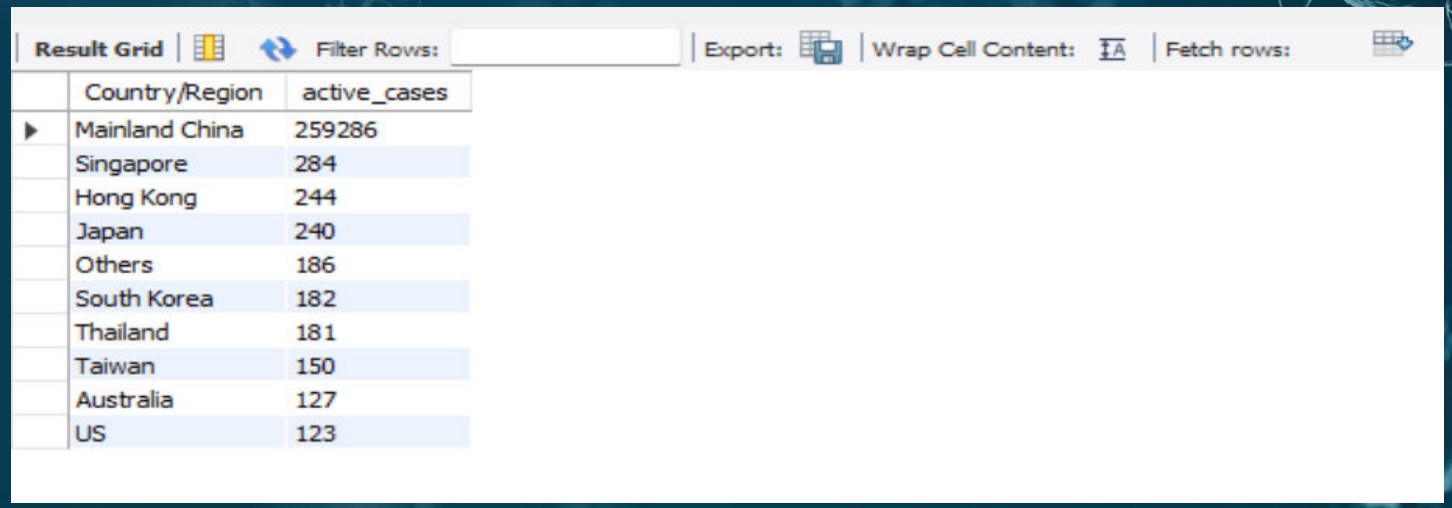
total_cases	total_deaths	total_recovered	death_rate	recovery_rate
281481	6072	13625	2.16	4.84



Shows overall impact and key percentages globally.

2. Top 10 Countries with Highest Active Cases

```
SELECT
    `Country/Region`,
    (SUM(confirmed) - SUM(recovered) - SUM(deaths)) AS active_cases
    FROM covid_19_data
    GROUP BY `Country/Region`
    ORDER BY active_cases DESC
    LIMIT 10;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Country/Region	active_cases
Mainland China	259286
Singapore	284
Hong Kong	244
Japan	240
Others	186
South Korea	182
Thailand	181
Taiwan	150
Australia	127
US	123

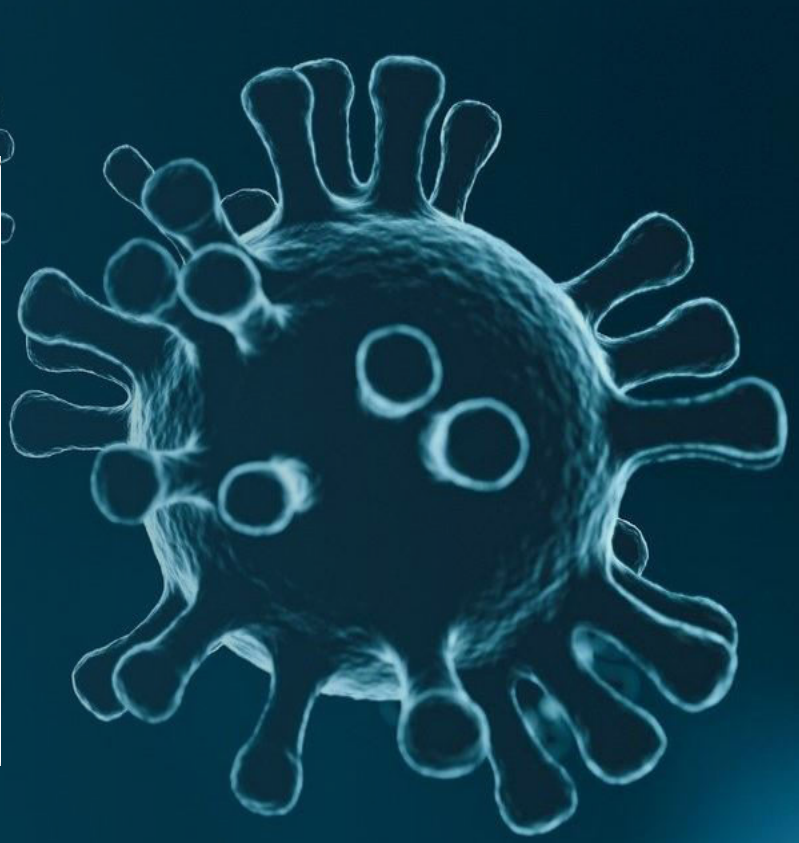
Identifies countries with the most ongoing cases.

3. Daily New Cases Globally

```
SELECT
    ObservationDate,
    SUM(confirmed) - LAG(SUM(confirmed)) OVER (ORDER BY ObservationDate) AS daily_new_cases
    FROM covid_19_data
    GROUP BY ObservationDate
    ORDER BY ObservationDate;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

ObservationDate	daily_new_cases
2/9/2020	2815
2/8/2020	2729
2/7/2020	3574
2/6/2020	3182
2/5/2020	3744
2/4/2020	4011
2/3/2020	3094
2/2/2020	4749
2/1/2020	2113
1/31/2020	1690
1/30/2020	2070
1/29/2020	587
1/28/2020	2651
1/27/2020	809
1/26/2020	681
1/25/2020	496
1/24/2020	156



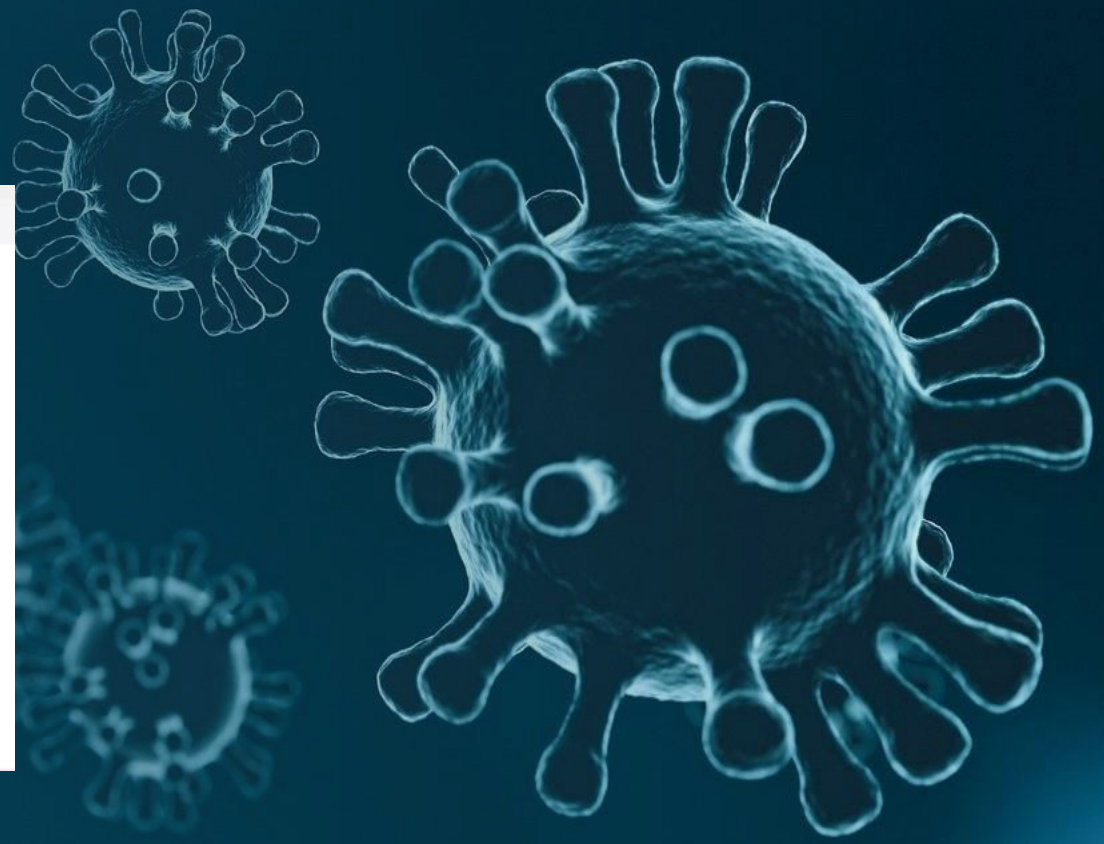
Shows how cases evolved daily.

4. Case Fatality Rate by Country

```
SELECT
    `Country/Region`,
    ROUND(SUM(Deaths) * 100.0 / NULLIF(SUM(Confirmed), 0), 2) AS case_fatality_rate
    FROM covid_19_data
    GROUP BY `Country/Region`
    ORDER BY case_fatality_rate DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

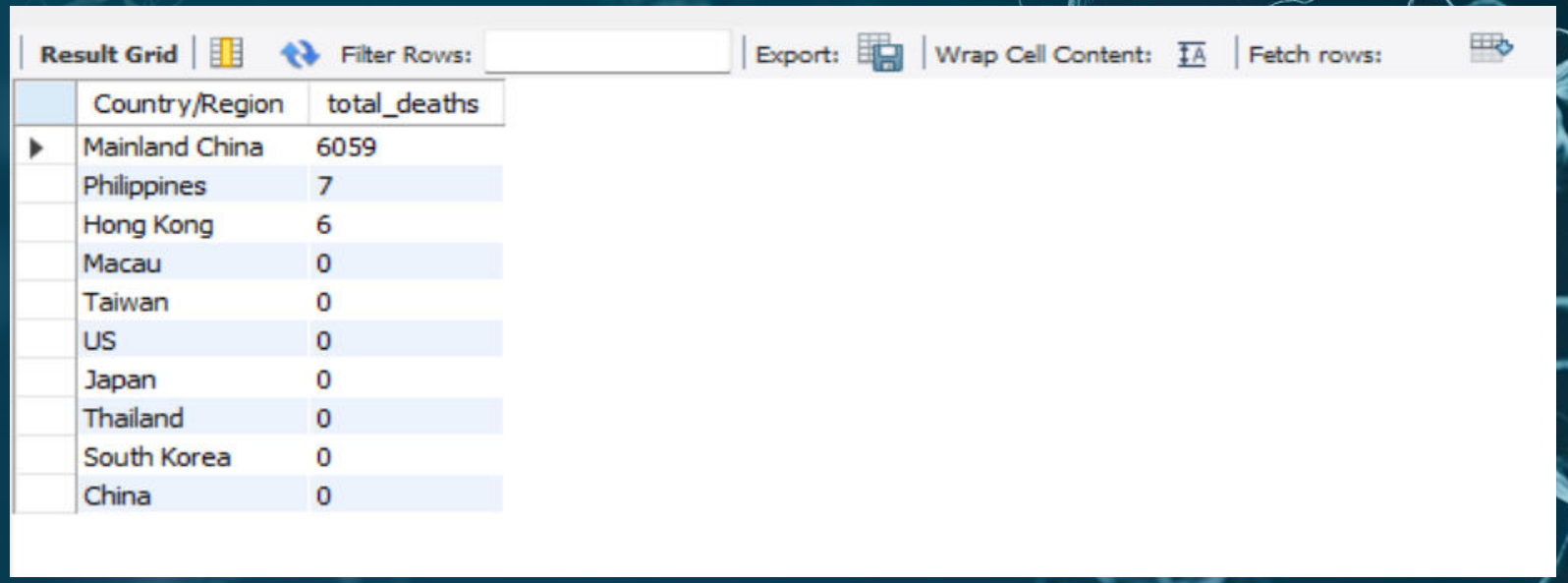
Country/Region	case_fatality_rate
Philippines	36.84
Hong Kong	2.40
Mainland China	2.17
Macau	0.00
Taiwan	0.00
US	0.00
Japan	0.00
Thailand	0.00
South Korea	0.00
Russia	0.00
Sweden	0.00
Singapore	0.00
Malaysia	0.00
Vietnam	0.00
Australia	0.00



Measures severity across countries.

5. Top 10 Countries by Highest Deaths

```
SELECT
    `Country/Region`,
    SUM(Deaths) AS total_deaths
    FROM covid_19_data
    GROUP BY `Country/Region`
    ORDER BY total_deaths DESC
    LIMIT 10;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Country/Region	total_deaths
Mainland China	6059
Philippines	7
Hong Kong	6
Macau	0
Taiwan	0
US	0
Japan	0
Thailand	0
South Korea	0
China	0

Normalizes cases based on population size.

6. Daily Growth Rate per Country

```
SELECT
        `Country/Region`,
        ObservationDate,
        SUM(confirmed) AS total_confirmed,
        ROUND ((SUM(confirmed) - LAG(SUM(Confirmed)) OVER (PARTITION BY `Country/Region`
            ORDER BY ObservationDate))  * 100.0 / NULLIF(LAG(SUM(Confirmed)) OVER (PARTITION BY `Country/Region`
            ORDER BY ObservationDate), 0),  2) AS growth_rate_percentage
        FROM covid_19_data
        GROUP BY `Country/Region`, ObservationDate
        ORDER BY `Country/Region`, ObservationDate DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country/Region	ObservationDate	total_confirmed	growth_rate_percentage
Australia	2/8/2020	15	0.00
Australia	2/7/2020	15	7.14
Australia	2/6/2020	14	7.69
Australia	2/5/2020	13	0.00
Australia	2/4/2020	13	8.33
Australia	2/3/2020	12	0.00
Australia	2/2/2020	12	0.00
Australia	2/1/2020	12	33.33
Australia	1/31/2020	9	0.00
Australia	1/30/2020	9	80.00
Australia	1/29/2020	5	0.00
Australia	1/28/2020	5	0.00
Australia	1/27/2020	5	25.00
Australia	1/26/2020	4	0.00
Australia	1/25/2020	4	NULL
Australia	1/23/2020	0	NULL
Belgium	2/8/2020	1	0.00
Belgium	2/7/2020	1	0.00
Belgium	2/6/2020	1	0.00

Detects acceleration or slowdown of the outbreak.

7. Peak Cases Date per Country

```
WITH daily_data AS
 (SELECT  `Country/Region`, ObservationDate,
SUM(Confirmed) - LAG(SUM(Confirmed)) OVER (PARTITION BY `Country/Region`  ORDER BY ObservationDate ) AS daily_cases
FROM covid_19_data
GROUP BY `Country/Region`, ObservationDate)
SELECT `Country/Region`, ObservationDate,  daily_cases
FROM (
SELECT *, RANK() OVER (PARTITION BY `Country/Region`
ORDER BY daily_cases DESC    ) AS rnk
FROM daily_data) ranked
WHERE rnk = 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country/Region	ObservationDate	daily_cases
Australia	1/25/2020	4
Australia	1/30/2020	4
Belgium	2/8/2020	0
Belgium	2/5/2020	0
Belgium	2/6/2020	0
Belgium	2/7/2020	0
Brazil	1/23/2020	NULL
Cambodia	1/28/2020	0
Cambodia	1/29/2020	0
Cambodia	1/30/2020	0
Cambodia	1/31/2020	0
Cambodia	2/1/2020	0
Cambodia	2/2/2020	0
Cambodia	2/3/2020	0
Cambodia	2/4/2020	0
Cambodia	2/5/2020	0
Cambodia	2/6/2020	0
Cambodia	2/7/2020	0
Cambodia	2/8/2020	0

Finds the date when each country had its highest surge.

8. Mortality vs. Recovery Correlation per Country

```
SELECT
    `Country/Region`,
    ROUND(SUM(Deaths) * 100.0 / NULLIF(SUM(Deaths) + SUM(Recovered), 0), 2) AS Mortality_percentage,
    ROUND(SUM(Recovered) * 100.0 / NULLIF(SUM(Deaths) + SUM(Recovered), 0), 2) AS Recovery_percentage
    FROM covid_19_data
    GROUP BY `Country/Region`
    ORDER BY Mortality_percentage DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country/Region	Mortality_percentage	Recovery_percentage
Hong Kong	100.00	0.00
Philippines	100.00	0.00
Mainland China	31.09	68.91
Macau	0.00	100.00
Taiwan	0.00	100.00
Malaysia	0.00	100.00
Japan	0.00	100.00
Thailand	0.00	100.00
South Korea	0.00	100.00
Vietnam	0.00	100.00
Australia	0.00	100.00
Singapore	0.00	100.00
Sri Lanka	0.00	100.00
China	NULL	NULL

Compares how each country fared in terms of survival vs. mortality.

Conclusion and Future Work

This project demonstrates the effective use of **advanced SQL analytics** to extract, process, and analyze large-scale global health data. By employing structured queries in **MySQL**, it delivers accurate, scalable, and reproducible insights into key metrics such as cases, deaths, recovery rates, and country-wise trends. The work establishes a **robust analytical framework** that supports real-time monitoring and can be adapted for future applications in **infectious disease surveillance, healthcare planning, and risk assessment**.

Building upon this foundation, future enhancements will focus on:

I. **Advanced Analytical Queries:** Incorporating predictive modeling using SQL with machine learning integration.
II. **Dynamic Dashboards:** Connecting the database to visualization tools such as **Power BI or Tableau** for interactive, real-time insights.
III. **Data Enrichment:** Integrating additional datasets (vaccination rates, demographic information, healthcare infrastructure).
IV. **Automated ETL Pipelines:** Enabling seamless data ingestion and periodic updates for continuous analysis.
V. **Scalability and Cloud Deployment:** Migrating the database to cloud-based platforms for global accessibility and higher processing capabilities.

This forward-looking approach ensures the project remains **relevant, extensible, and impactful**, serving as a foundation for **evidence-based research and decision-making** in public health and beyond.