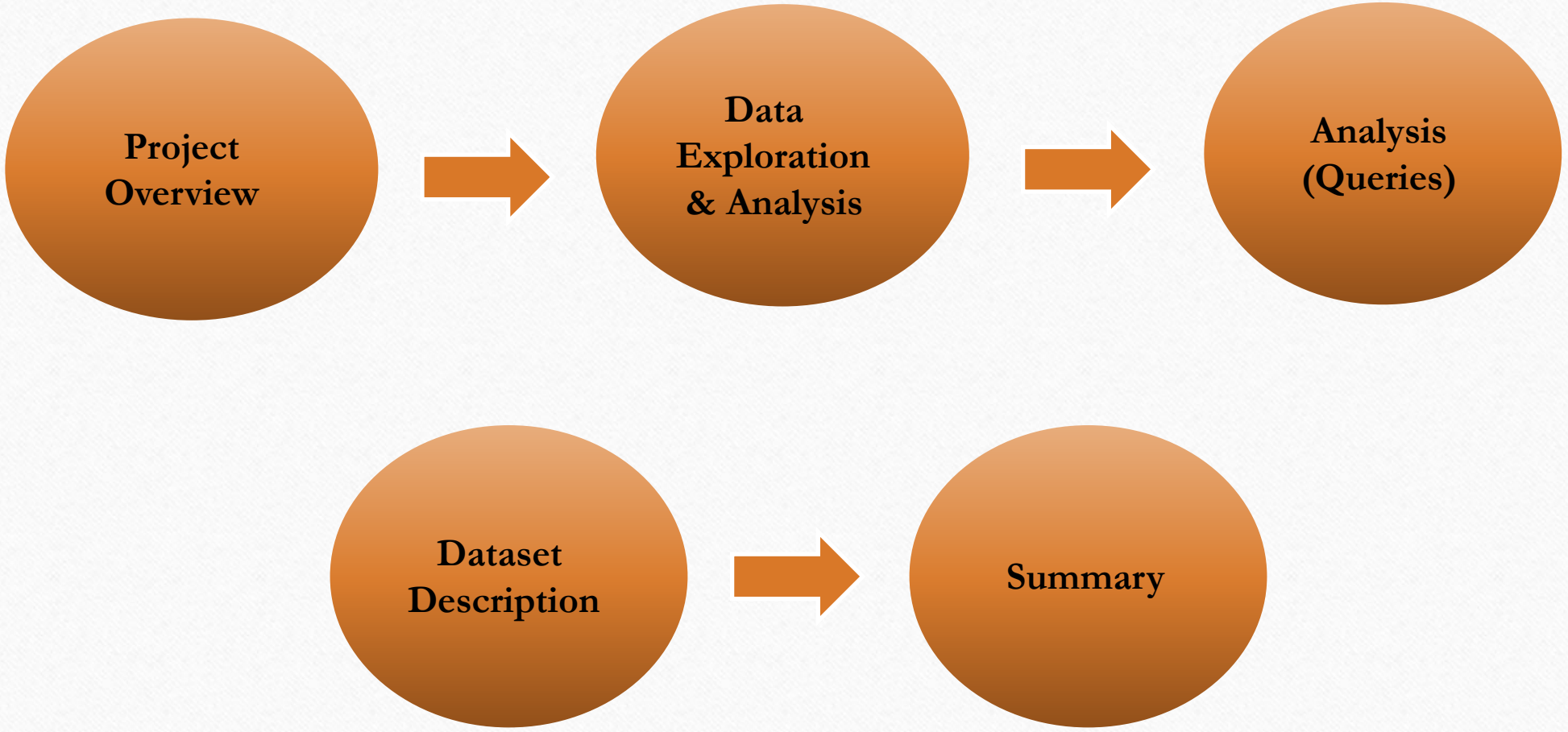
CORONA VIRUS ANALYSIS

SQL (Structure Query Language)

BATCH NUMBER : MIP-DA-11

Presenting By : VISHRUT VERMA

CONTENT


Project Overview
Data Exploration & Analysis
Analysis (Queries)
Dataset Description
Summary

OBJECTIVE

The primary goal of this project is to perform data analysis using SQL. SQL data analysis is to locate the desired data that is necessary.

By using SQL we can efficiently retrieve and update the information.

The analysis reveals significant trends in the spread of COVID-19, highlighting the rapid growth of confirmed cases over time and identifying regions with the highest infection rates through geographic distribution maps.

DESCRIPTION OF DATA

In this project we are going to write some SQL queries to make analysis of data more efficient. Let us see the description of columns in data.

Province: Geographic subdivision within a country/region.
Country/Region: Geographic entity where data is recorded.
Latitude: North-south position on Earth's surface.
Longitude: East-west position on Earth's surface.
Date: Recorded date of CORONA VIRUS data.
Confirmed: Number of diagnosed Corona Virus cases. Deaths: Number of Corona Virus related deaths.
Recovered: Number of recovered Corona Virus cases.

QUERY 1

- **Write a code to check NULL values**

```
SELECT FROM covid_data
        WHERE Confirmed IS NULL
        OR Deaths IS NULL
        OR Recovered IS NULL;
```

QUERY 2

- **If NULL values are present, update them with zeros for all columns.**

```
UPDATE covid_data
SET Confirmed = COALESCE (Confirmed, 0),
Deaths COALESCE (Deaths, 0),
Recovered COALESCE (Recovered, 0)
WHERE Confirmed IS NULL
OR Deaths IS NULL
OR Recovered IS NULL;
```

QUERY 3

- **Check total number of rows**

```
SELECT COUNT(*) AS total_rows FROM covid_data;
```

QUERY 4

- **Check what is start_date and end_date.**

```
SELECT MIN (Date) AS start_date,
MAX (Date) AS end_date
FROM covid_data;
```

QUERY 5

- **Number of month present in dataset.**

```
SELECT COUNT (DISTINCT EXTRACT (YEAR_MONTH
FROM STR_TO_DATE (Date, '%d-%m-%Y')))
AS total_months
FROM covid_data;
```

QUERY 6

- **Find monthly average for confirmed, deaths, recovered**

```
SELECT EXTRACT (YEAR FROM STR TO_DATE (Date, '%d-%m-%Y')) AS year,
EXTRACT (MONTH FROM STR TO DATE (Date, '%d- %m-%Y'))     AS month,
        AVG (Confirmed) AS avg_confirmed,
        AVG (Deaths) AS avg deaths,
        AVG (Recovered) AS avg_recovered
        FROM coviddata
        GROUP BY year, month ORDER BY year, month;
```

QUERY 7

- **Find most frequent value for confirmed, deaths, recovered each month.**

```
SELECT year, month, Confirmed, Deaths, Recovered FROM ( SELECT EXTRACT
(YEAR FROM STR TO DATE (Date, d- %m-%Y')) AS year, EXTRACT (MONTH
FROM STR TO DATE (Date, %d-%m-%Y')) AS month, Confirmed, Deaths,
Recovered,  ROW NUMBER() OVER (PARTITION BY EXTRACT (YEAR FROM
STR TO DATE (Date, %d-%m-%Y')), EXTRACT (MONTH FROM STR TO DATE
(Date, "%d-%m-%Y'))
ORDER BY COUNT (*) DESC) AS row_num FROM covid_data
GROUP BY year, month, Confirmed, Deaths, Recovered) AS subquery
WHERE row_num=1
```

QUERY 8

- **Find minimum values for confirmed, deaths, recovered per year .**

```
SELECT EXTRACT (YEAR FROM STR_TO_DATE (Date, '%d-%m-%Y')) AS year,
MIN (Confirmed) AS min_confirmed,
MIN (Deaths) AS min_deaths,
MIN (Recovered) AS min_recovered
FROM covid_data
GROUP BY year
ORDER BY year ;
```

QUERY 9

- **Find maximum values of confirmed, deaths, recovered per year.**

```
SELECT EXTRACT (YEAR FROM STR_TO_DATE (Date, '%d-%m-%Y'))
AS year, MAX (Confirmed) AS max confirmed,
MAX (Deaths) AS max_deaths,
MAX (Recovered) AS max_recovered
FROM covid_data
GROUP BY year
ORDER BY year;
```

QUERY 10

- **The total number of case of confirmed, deaths, recovered each month.**

```
SELECT
EXTRACT (YEAR FROM STR_TO_DATE (Date, '%d-%m-%Y')) AS year,
EXTRACT (MONTH FROM STR TO_DATE (Date, '%d- %m-%Y')) AS month,
SUM (Confirmed) AS total_confirmed,
SUM (Deaths) AS total_deaths,
SUM (Recovered) AS total_recovered
FROM covid_data
GROUP BY year, month
ORDER BY year, month;
```

QUERY 11

- **Check how corona virus spread out with respect to confirmed case.**

```
SELECT EXTRACT (YEAR FROM STR TO_DATE (Date, '%d-%m-%Y')) AS year,
EXTRACT (MONTH FROM STR_TO_DATE (Date, "%d- %m-%Y')) AS month,
SUM (Confirmed) AS total_confirmed,
AVG (Confirmed) AS avg_confirmed,
VARIANCE (Confirmed) AS variance_confirmed,
STDDEV (Confirmed) AS stddev_confirmed
FROM covid_data GROUP BY year, month
ORDER BY year, month
```

QUERY 12

- **Check how corona virus spread out with respect to death case per month.**

```
SELECT EXTRACT (YEAR FROM STR TO_DATE (Date, '%d-%m-%Y')) AS year,
EXTRACT (MONTH FROM STR_TO_DATE (Date, "%d- %m-%Y')) AS month,
SUM (Deaths) AS total_deaths,
AVG (Deaths) AS avg_deaths,
VARIANCE (Deaths) AS variance_deaths,
STDDEV (Deaths) AS stddev_deaths
FROM covid_data
GROUP BY year, month
ORDER BY year, month;
```

QUERY 13

- **Check how corona virus spread out with respect to recovered case.**

```
SELECT EXTRACT (YEAR FROM STR TO_DATE (Date, '%d-%m-%Y')) AS year,
EXTRACT (MONTH FROM STR_TO_DATE(Date, "%d- %m-%Y')) AS month,
SUM (Recovered) AS total_recovered,
AVG (Recovered) AS avg_recovered,
VARIANCE (Recovered) AS variance_recovered,
STDDEV (Recovered) AS stddev_recovered
FROM covid_data
GROUP BY year, month
ORDER BY year, month;
```

QUERY 14

- **Find Country having highest number of the Confirmed case.**

```
SELECT Country/Region AS country,
SUM (Confirmed) AS total_confirmed
FROM covid_data
GROUP BY country
ORDER BY total_confirmed
ASC LIMIT 1;
```

QUERY 15

- **Find Country having lowest number of the death case.**

```
SELECT Country/Region AS country,
SUM(Deaths) AS total_deaths
FROM covid_data
GROUP BY country
ORDER BY total_deaths
ASC LIMIT 1;
```

QUERY 16

- **Find top 5 countries having highest recovered case.**

```
SELECT `Country/Region' AS country,
SUM (Recovered) AS total_recovered
FROM covid_data
GROUP BY country
ORDER BY total_recovered
DESC LIMIT 5;
```

THANK YOU