SQL PROJECT

The COVID-19 dataset for Nigeria, sourced from Kaggle, spans the years 2020 to 2023 and comprises two tables: "covid-death-year" and "covid-vaccination." This dataset underwent analysis to extract valuable insights specific to Nigeria.

Key performance indicators (KPIs) extracted from the dataset include **Total Death By Years**, **Total Cases by Year**, **Total Cases from Dry and Rainy Seasons and Comparing the total cases and total vaccination**

STEPS TAKEN

Opened my SQL Server (Postgres), connected to the Database by right clicking on the database and giving it my preferred name.

After creating the Database, the next step is to create my tables.

Two tables created are the , **Covid_death_year and covid_vaccination**

CREATING TABLE for Covid_death_year

```
CREATE TABLE covid_death_year(
continent VARCHAR(50),
country VARCHAR(100),
dates VARCHAR(50) ,
total_cases numeric,
new_cases numeric,
new_cases_smoothed numeric,
total_deaths numeric,
new_deaths numeric,
new_deaths_smoothed numeric
)
```

TO IMPORT CSV FILE FROM THE DESKTOP TO POSTGRES SERVER

```
COPY Covid_death_year
FROM 'C\User\HP\Downloads\Covid death years.csv'DELIMITER','CSV
```

While importing this file, I encountered an issue due to a mismatch in the data types. The table was initially created with a VARCHAR data type for the DATE column, while the actual date values in the file were of the DATE data type. To resolve this discrepancy, I imported the data as VARCHAR and later instructed SQL to convert the DATE data type to align with the file's structure.

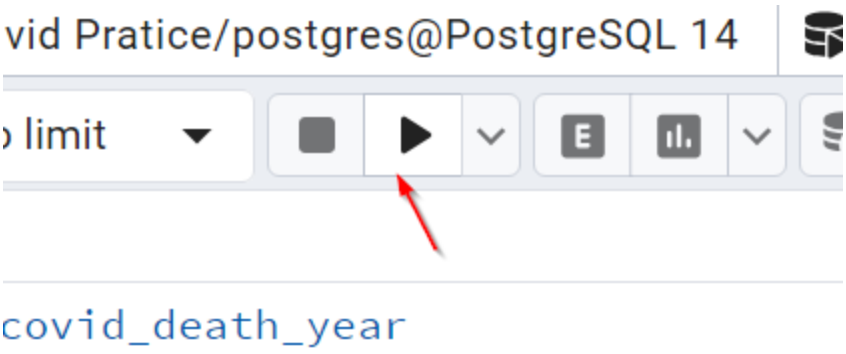
To Do that i used this syntax:

```
ALTER TABLE covid_death_year
ALTER COLUMN "dates" TYPE DATE USING "dates"::DATE;
```

TO VIEW TABLE Covid_death_year

```
SELECT * FROM covid_death_year
```

To RUN the query, On the PostgreSQL Interface



CREATING TABLE for covid_vaccination

```
CREATE TABLE covid_vaccination(
continent VARCHAR(50),
country VARCHAR(100),
total_vaccinations numeric,
people_vaccinated numeric,
people_fully_vaccinated numeric,
new_vaccinations numeric,
new_vaccinations_smoothed numeric
)
```

TO IMPORT CSV FILE FROM THE DESKTOP TO POSTGRES SERVER

```
COPY Covid_Vaccination
FROM 'C\User\HP\Downloads\Covid vaccination.csv'DELIMITER','CSV
```

TO VIEW TABLE Covid_vaccination

```
SELECT * FROM covid_vaccination
```

TECHNICAL ANALYSIS

To Calculate the ***total cases from Dry Season*** that is from ***November to April (11,12,1,2,3,4)***

```sql
SELECT
EXTRACT(MONTH FROM "dates"::DATE) AS month,
COUNT(DISTINCT "dates"::DATE) AS distinct_month_count,
SUM("total_cases") AS total_cases
FROM
covid_death_year
WHERE
EXTRACT(MONTH FROM "dates"::DATE) IN (1, 2, 3, 4, 11, 12)
GROUP BY
month
ORDER BY
month;
```

RUN:

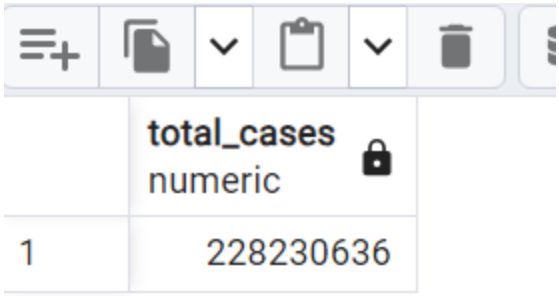
	month numeric	distinct_month_count bigint	total_cases numeric
1	1	122	38671972
2	2	113	37273260
3	3	124	42276806
4	4	120	41213592
5	11	92	33738114
6	12	93	35056892

To Calculate the ***Sum of the total cases of Dry season***

```sql
SELECT
SUM("total_cases") AS total_cases
FROM
```

```
covid_death_year
WHERE
EXTRACT(MONTH FROM "dates"::DATE) IN (11,12,1,2,3,4);
```

RUN:



	total_cases numeric
1	228230636

To Calculate the ***total cases by year***

```
SELECT
EXTRACT(YEAR FROM "dates"::DATE) AS Year,
SUM("total_cases") AS total_cases
FROM
covid_death_year
GROUP BY YEAR
ORDER BY total_cases DESC
;
```

RUN:

	year numeric	total_cases numeric
1	2022	189280678
2	2023	163256796
3	2021	128297722
4	2020	22289314

To Calculate the ***total death by year***

```
SELECT
EXTRACT(YEAR FROM "dates"::DATE) AS Year,
SUM("total_deaths") AS total_deaths
FROM
covid_death_year
GROUP BY YEAR
ORDER BY total_deaths DESC
;
```

	year numeric	total_deaths numeric
1	2022	2294422
2	2023	1930860
3	2021	1651658
4	2020	422924

To compare **the total cases and total vaccination**

```sql
SELECT
EXTRACT(YEAR FROM cd."dates"::DATE) AS Year,
cd."country",
SUM(cd."total_cases") AS total_cases,
SUM(cv."total_vaccinations") AS total_vaccinations
FROM
covid_death_year cd
JOIN
covid_vaccination cv ON cd."country" = cv."country"
GROUP BY
Year, cd."country"
ORDER BY
Year, cd."country";
```

	year numeric	country character varying (100)	total_cases numeric	total_vaccinations numeric
1	2020	Nigeria	31205039600	2050298731208
2	2021	Nigeria	179616810800	2055931420030
3	2022	Nigeria	264992949200	2055931420030
4	2023	Nigeria	228559514400	1723602779532

To Calculate ***the total cases from Rainy Season*** that is ***May-Oct***

```sql
SELECT
EXTRACT(MONTH FROM "dates"::DATE) AS month,
SUM("total_cases") AS total_cases
FROM
covid_death_year
WHERE
```

```
EXTRACT(MONTH FROM "dates"::DATE) IN (5, 6, 7, 8, 9, 10)
GROUP BY
month
ORDER BY
month;
```

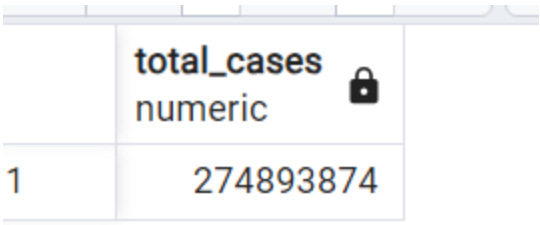
RUN:

	month numeric	distinct_month_count bigint	total_cases numeric
1	5	124	43016662
2	6	120	42426060
3	7	124	45232734
4	8	124	47178624
5	9	120	47275942
6	10	124	49763852

To Calculate the ***Sum of the total cases of rainy season***

```
sum of total cases for rainy season
SELECT
SUM("total_cases") AS total_cases
FROM
covid_death_year
WHERE
EXTRACT(MONTH FROM "dates"::DATE) IN (5, 6, 7, 8, 9, 10);
```

RUN:



	total_cases numeric
1	274893874

FINDINGS

From the above analysis,

- The data indicates that the prevalence of COVID-19 cases was higher during the rainy season (May to October) with a total of 274,893,874 cases. In comparison, the dry season (November to April) had a slightly lower total of 228,230,636 cases. The marginal difference in figures suggests that there is no distinct seasonal pattern, implying that COVID-19 can occur at any time without a strong association with a particular season. This lack of a clear seasonal effect underscores the unpredictability of the virus's occurrence throughout the year.
- In the analysis of annual death rates, it was observed that there was a notable increase in mortality during the year 2022. This surge in deaths might be indicative of a continued impact from individuals who were previously affected by COVID-19, possibly suggesting a gradual progression in mortality rates among those who contracted the virus.
- In the progression of annual COVID-19 cases, 2022 stands out with the highest reported count, contrasting with the lowest figures observed in 2020.
- In the first year, there was a significant rise in both the total number of COVID-19 cases and vaccination rates. This increase can be linked to elevated concerns, as individuals actively sought to prevent contracting the virus.