**SQL for Data Analysis**

**Dataset: 120 years of Olympic history: athletes and result.**

Two csv files “athlete\_events.csv“ and “noc\_regions.csv“.(Kaggle)

**Data from athlete\_events file. We load this data into a table named “OLYMPICS\_HISTORY”.**

**Data from noc\_regions file. We load this data into a table named “OLYMPICS\_HISTORY\_NOC\_REGIONS”.**

**Creating Table.**

DROP TABLE IF EXISTS olympics\_history;

CREATE TABLE IF NOT EXISTS olympics\_history

(

id INT;

name VARCHAR,

sex VARCHAR,

age VARCHAR,

height VARCHAR,

weight VARCHAR,

team VARCHAR,

noc VARCHAR,

games VARCHAR,

year INT,

season VARCHAR,

city Varchar,

sport VARCHAR,

event VARCHAR,

medal VARCHAR

);

DROP TABLE IF EXISTS olympics\_history\_noc\_regions;

CREATE TABLE IF NOT EXISTS olympics\_history\_noc\_regions

(

noc VARCHAR,

region VARCHAR,

notes VARCHAR

)

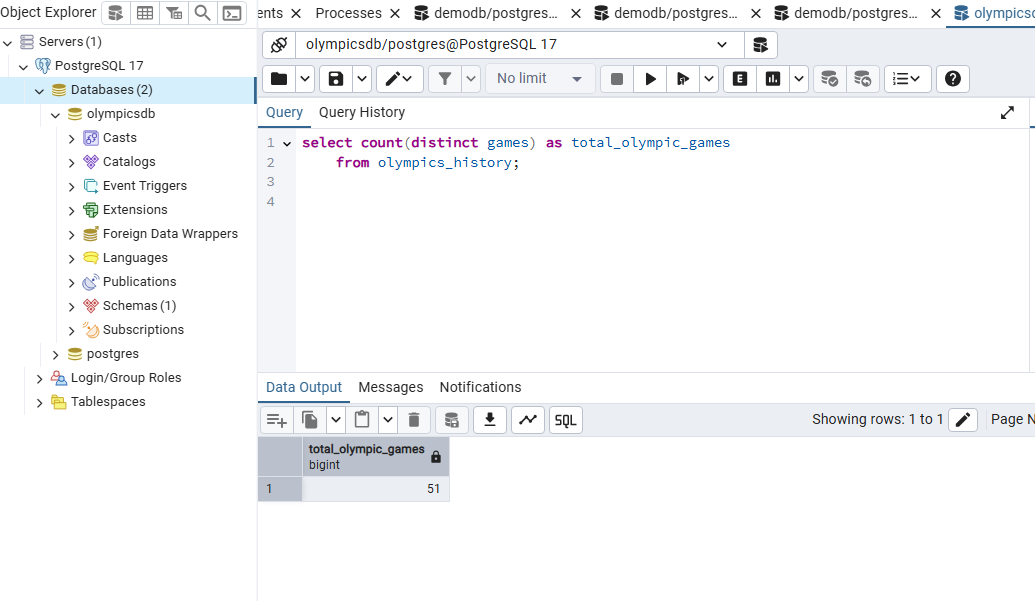
SQL Queries:

**1. How many olympics games have been held?**

**Problem Statement:** Write a SQL query to find the total no of Olympic Games held as per the dataset

select count(distinct games) as total\_olympic\_games

from olympics\_history;



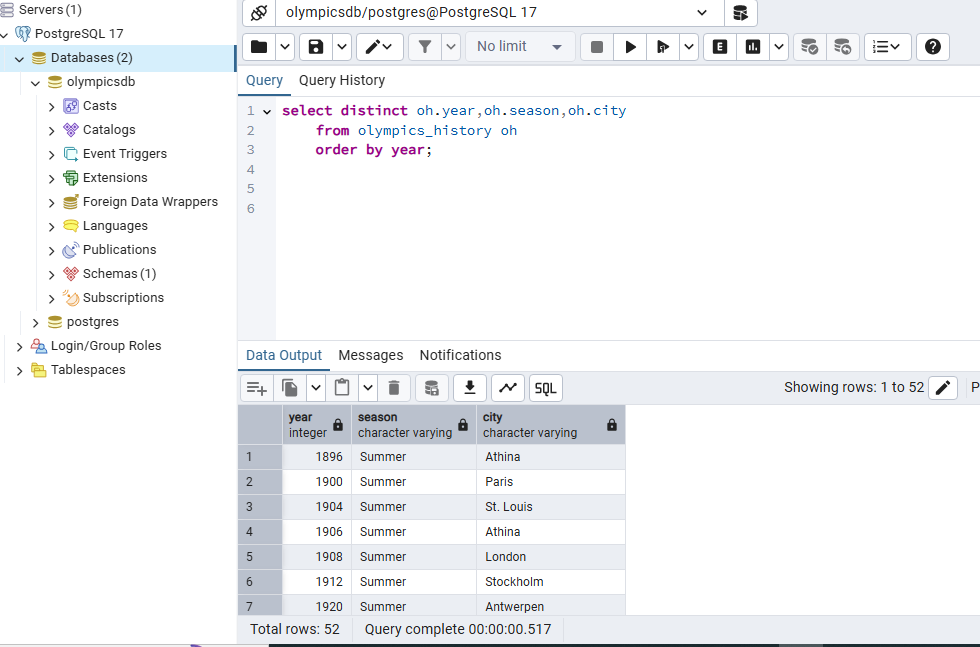
**2. List down all Olympics games held so far.**

**Problem Statement:** Write a SQL query to list down all the Olympic Games held so far.

select distinct oh.year,oh.season,oh.city

from olympics\_history oh

order by year;



**3. Mention the total no of nations who participated in each olympics game?**

**Problem Statement:** SQL query to fetch total no of countries participated in each olympic games.

with all\_countries as

(select games, nr.region

from olympics\_history oh

join olympics\_history\_noc\_regions nr ON nr.noc = oh.noc

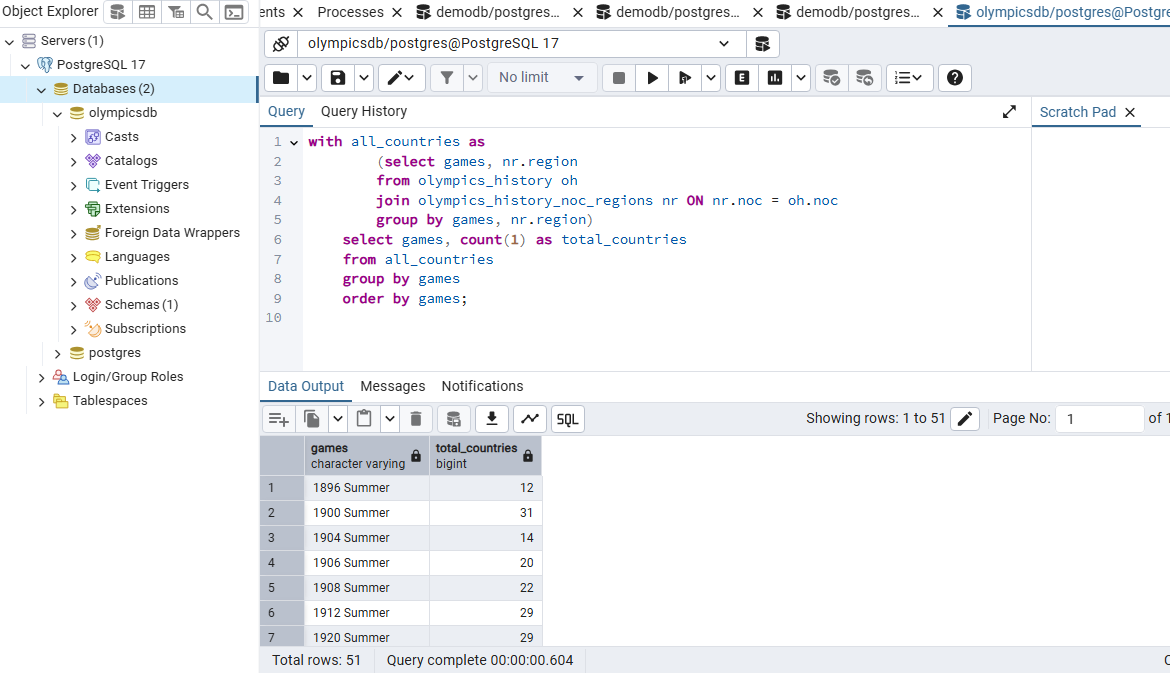
group by games, nr.region)

select games, count(1) as total\_countries

from all\_countries

group by games

order by games;



**4. Which year saw the highest and lowest no of countries participating in olympics**

**Problem Statement**: Write a SQL query to return the Olympic Games which had the highest participating countries and the lowest participating countries.

with all\_countries as

(select games, nr.region

from olympics\_history oh

join olympics\_history\_noc\_regions nr ON nr.noc=oh.noc

group by games, nr.region),

tot\_countries as

(select games, count(1) as total\_countries

from all\_countries

group by games)

select distinct

concat(first\_value(games) over(order by total\_countries)

, ' - '

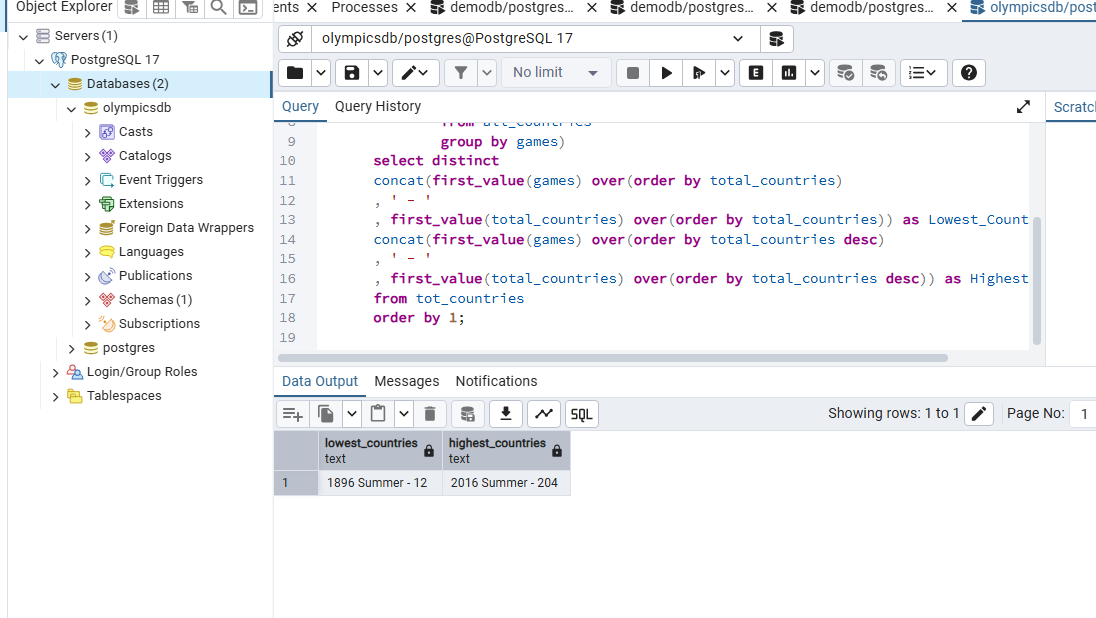
, first\_value(total\_countries) over(order by total\_countries)) as Lowest\_Countries,

concat(first\_value(games) over(order by total\_countries desc)

, ' - ' , first\_value(total\_countries) over(order by total\_countries desc)) as Highest\_Countries

from tot\_countries

order by 1;



**5. Which nation has participated in all of the olympic games**

**Problem Statement**: SQL query to return the list of countries who have been part of every Olympics games.

with tot\_games as

(select count(distinct games) as total\_games

from olympics\_history),

countries as

(select games, nr.region as country

from olympics\_history oh

join olympics\_history\_noc\_regions nr ON nr.noc=oh.noc

group by games, nr.region),

countries\_participated as

(select country, count(1) as total\_participated\_games

from countries

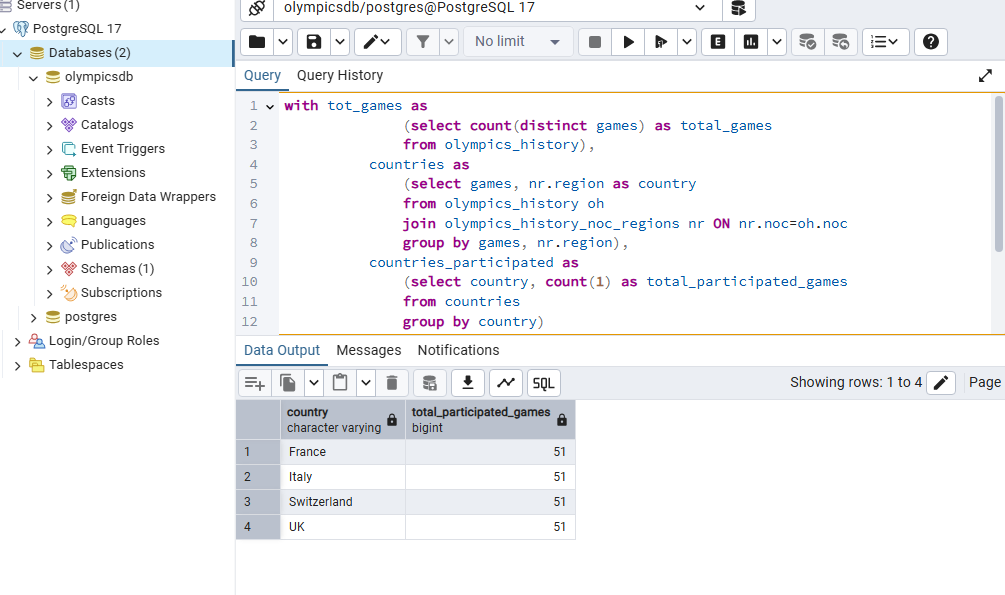
group by country)

select cp.\*

from countries\_participated cp

join tot\_games tg on tg.total\_games = cp.total\_participated\_games

order by 1;



**6. Identify the sport which was played in all summer olympics.**

**Problem Statement**: SQL query to fetch the list of all sports which have been part of every olympics.

with t1 as

(select count(distinct games) as total\_games

from olympics\_history where season = 'Summer'),

t2 as

(select distinct games, sport

from olympics\_history where season = 'Summer'),

t3 as

(select sport, count(1) as no\_of\_games

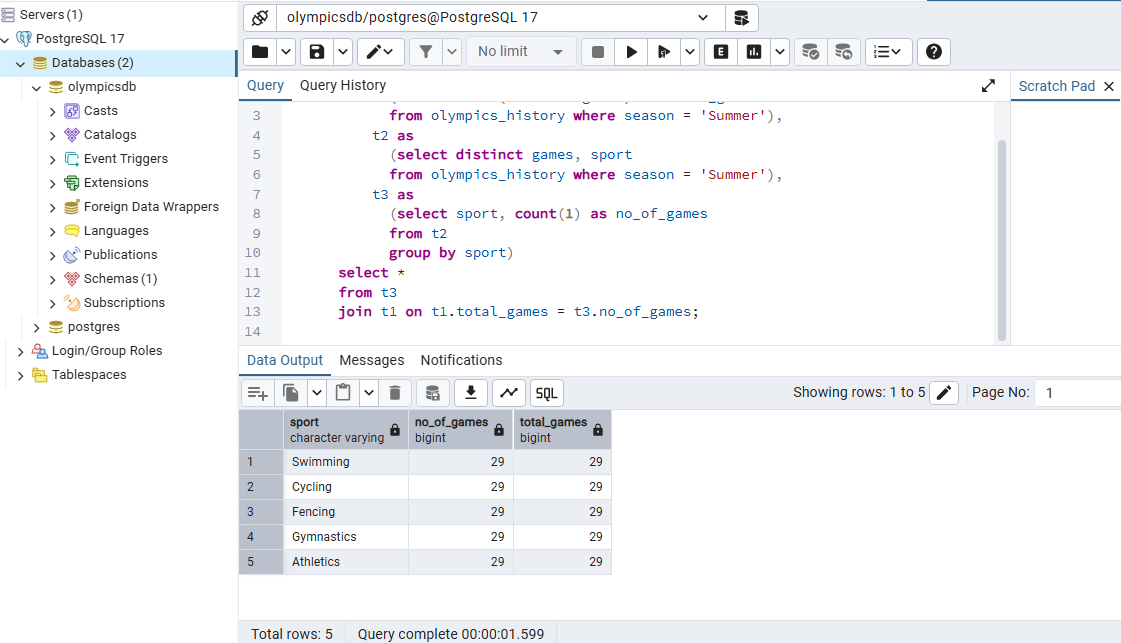
from t2

group by sport)

select \*

from t3

join t1 on t1.total\_games = t3.no\_of\_games;



**7. Which Sports were just played only once in the olympics.**

**Problem Statement:** Using SQL query, Identify the sport which were just played once in all of olympics.

with t1 as

(select distinct games, sport

from olympics\_history),

t2 as

(select sport, count(1) as no\_of\_games

from t1

group by sport)

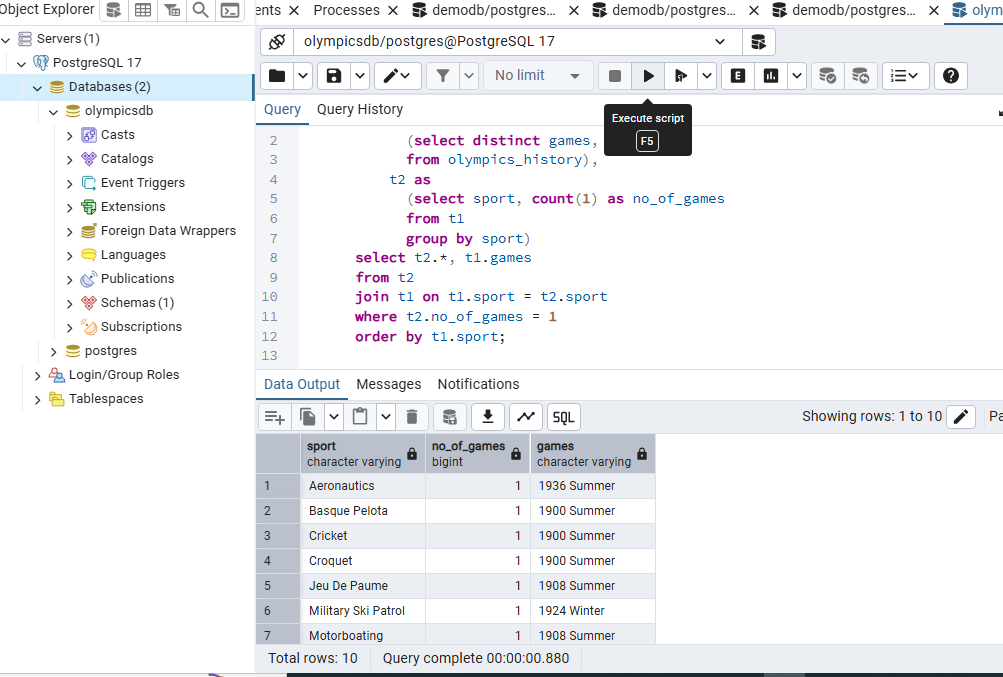
select t2.\*, t1.games

from t2

join t1 on t1.sport = t2.sport

where t2.no\_of\_games = 1

order by t1.sport;



**8. Fetch the total no of sports played in each olympic games.**

**Problem Statement**: Write SQL query to fetch the total no of sports played in each olympics.

with t1 as

(select distinct games, sport

from olympics\_history),

t2 as

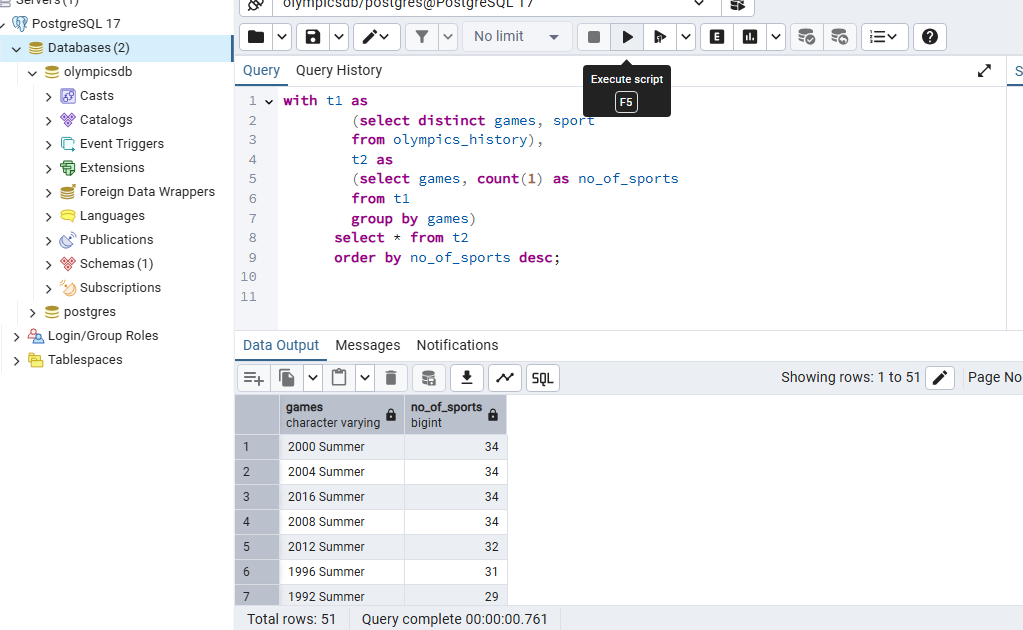
(select games, count(1) as no\_of\_sports

from t1

group by games)

select \* from t2

order by no\_of\_sports desc;



**9. Fetch oldest athletes to win a gold medal**

**Problem Statement:** SQL Query to fetch the details of the oldest athletes to win a gold medal at the olympics.

with temp as

(select name,sex,cast(case when age = 'NA' then '0' else age end as int) as age

,team,games,city,sport, event, medal

from olympics\_history),

ranking as

(select \*, rank() over(order by age desc) as rnk

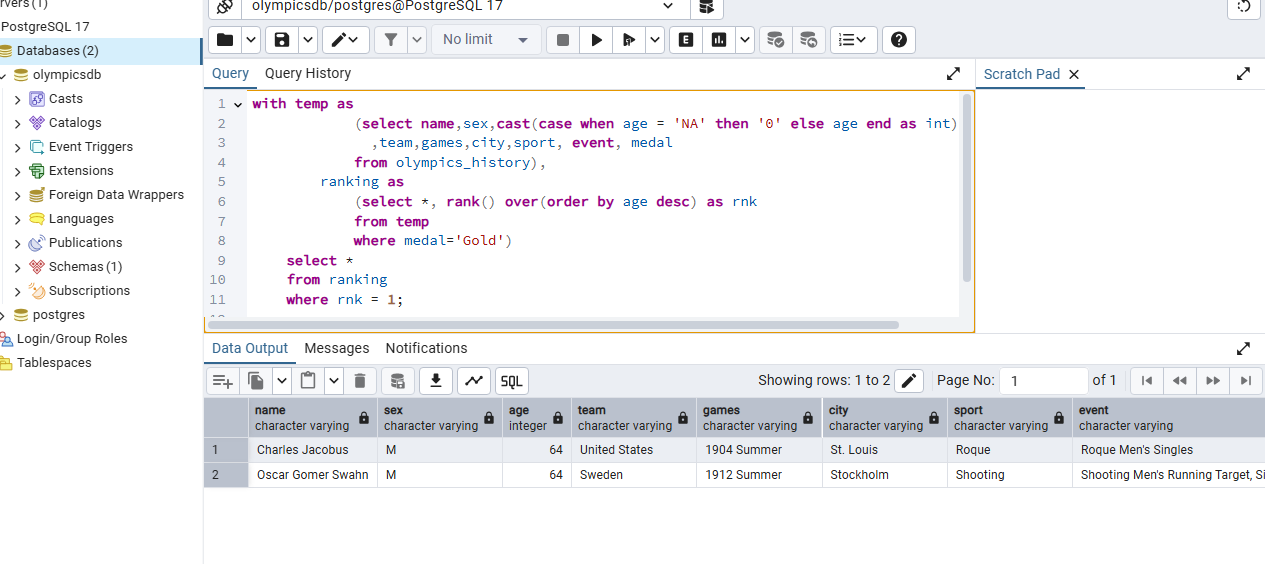
from temp

where medal='Gold')

select \*

from ranking

where rnk = 1;



**10. Find the Ratio of male and female athletes participated in all olympic games.**

**Problem Statement:** Write a SQL query to get the ratio of male and female participants

with t1 as

(select sex, count(1) as cnt

from olympics\_history

group by sex),

t2 as

(select \*, row\_number() over(order by cnt) as rn

from t1),

min\_cnt as

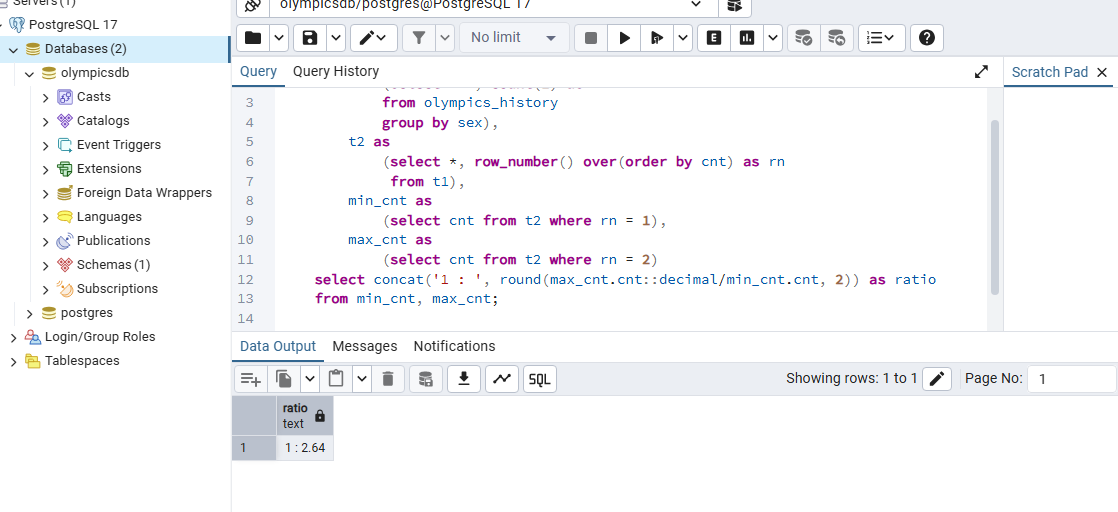
(select cnt from t2 where rn = 1),

max\_cnt as

(select cnt from t2 where rn = 2)

select concat('1 : ', round(max\_cnt.cnt::decimal/min\_cnt.cnt, 2)) as ratio

from min\_cnt, max\_cnt;



**11. Fetch the top 5 athletes who have won the most gold medals.**

**Problem Statement**: SQL query to fetch the top 5 athletes who have won the most gold medals.

with t1 as

(select name, team, count(1) as total\_gold\_medals

from olympics\_history

where medal = 'Gold'

group by name, team

order by total\_gold\_medals desc),

t2 as

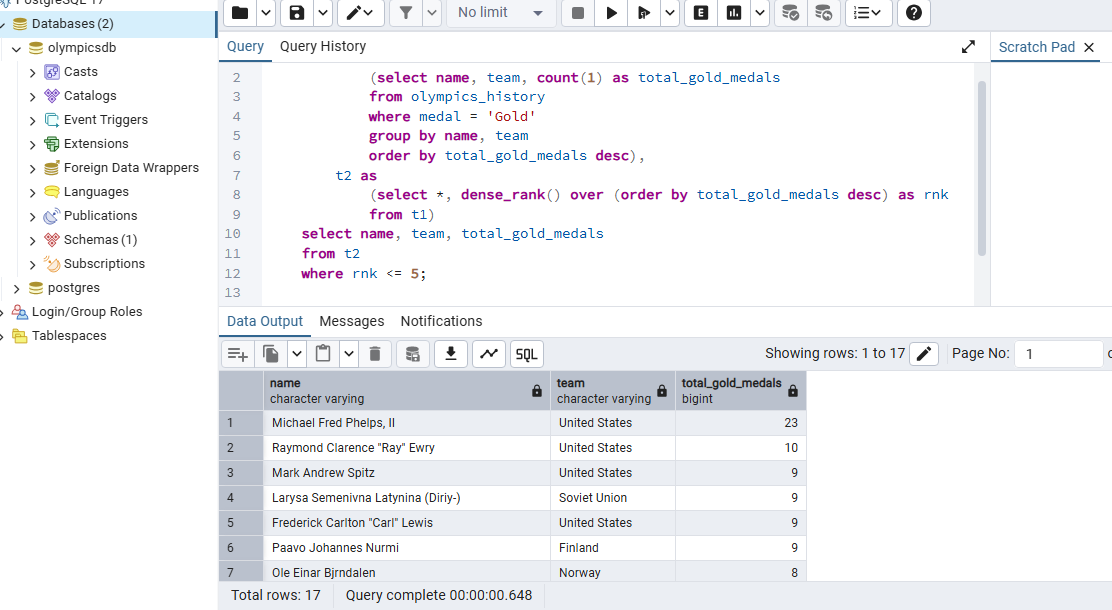
(select \*, dense\_rank() over (order by total\_gold\_medals desc) as rnk

from t1)

select name, team, total\_gold\_medals

from t2

where rnk <= 5;



**12. Fetch the top 5 athletes who have won the most medals (gold/silver/bronze).**

**Problem Statement:** SQL Query to fetch the top 5 athletes who have won the most medals (Medals include gold, silver and bronze).

with t1 as

(select name, team, count(1) as total\_medals

from olympics\_history

where medal in ('Gold', 'Silver', 'Bronze')

group by name, team

order by total\_medals desc),

t2 as

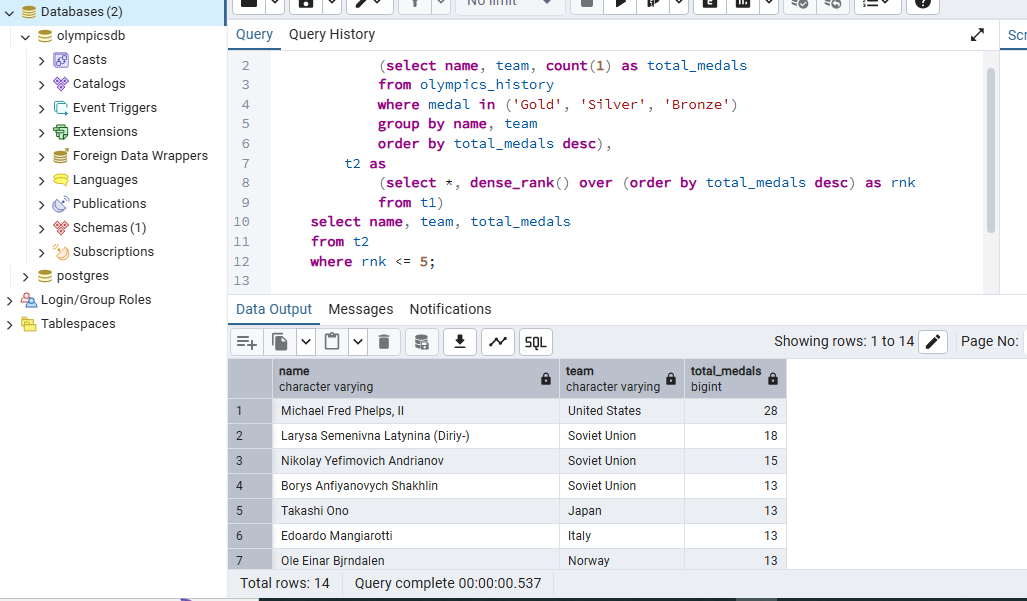
(select \*, dense\_rank() over (order by total\_medals desc) as rnk

from t1)

select name, team, total\_medals

from t2

where rnk <= 5;



**13. In which Sport/event, India has won highest medals.**

**Problem Statement**: Write SQL Query to return the sport which has won India the highest no of medals.

with t1 as

(select sport, count(1) as total\_medals

from olympics\_history

where medal <> 'NA'

and team = 'India'

group by sport

order by total\_medals desc),

t2 as

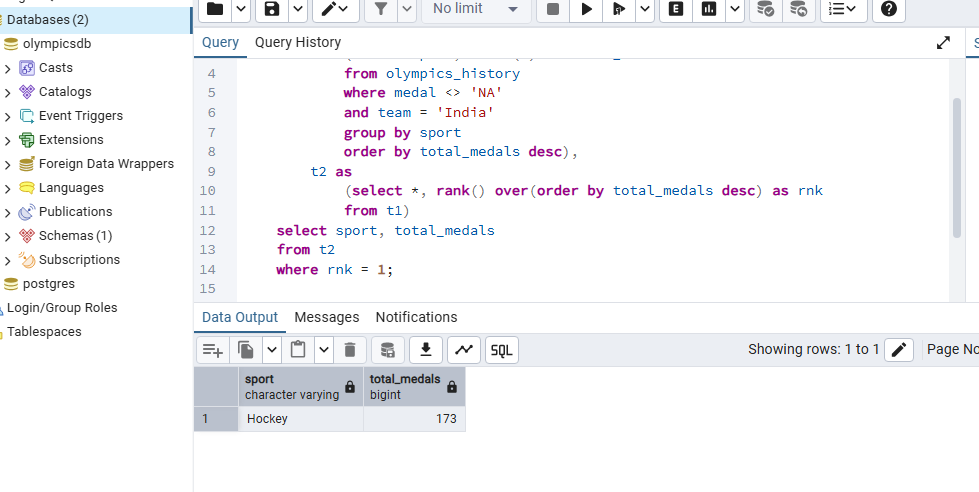
(select \*, rank() over(order by total\_medals desc) as rnk

from t1)

select sport, total\_medals

from t2

where rnk = 1;



**14. Fetch the top 5 most successful countries in olympics. Success is defined by no of medals won.**

**Problem Statement: Write a SQL query to fetch the top 5 most successful countries in olympics. (Success is defined by no of medals won).**

**with t1 as**

**(select nr.region, count(1) as total\_medals**

**from olympics\_history oh**

**join olympics\_history\_noc\_regions nr on nr.noc = oh.noc**

**where medal <> 'NA'**

**group by nr.region**

**order by total\_medals desc),**

**t2 as**

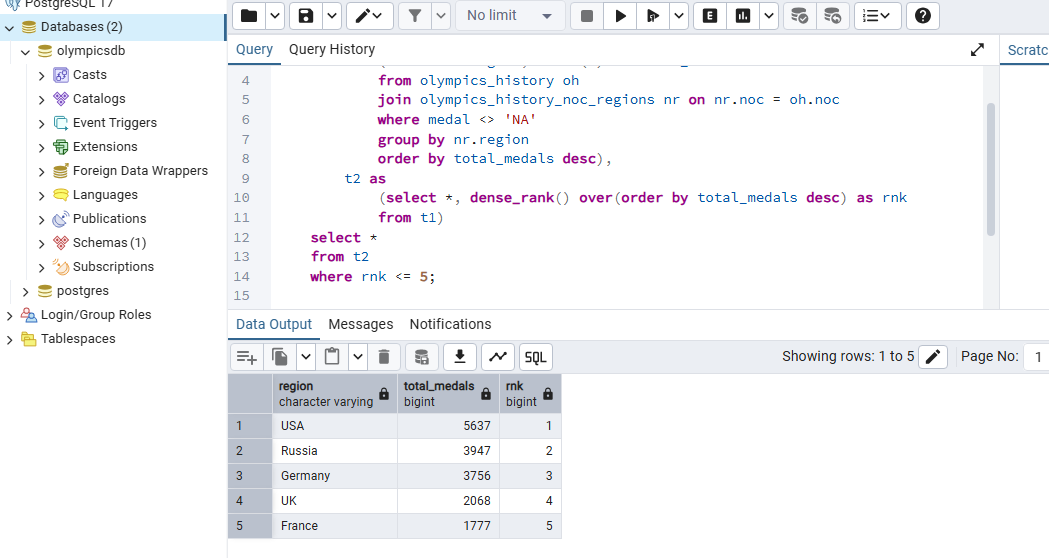
**(select \*, dense\_rank() over(order by total\_medals desc) as rnk**

**from t1)**

**select \***

**from t2**

**where rnk <= 5;**

****