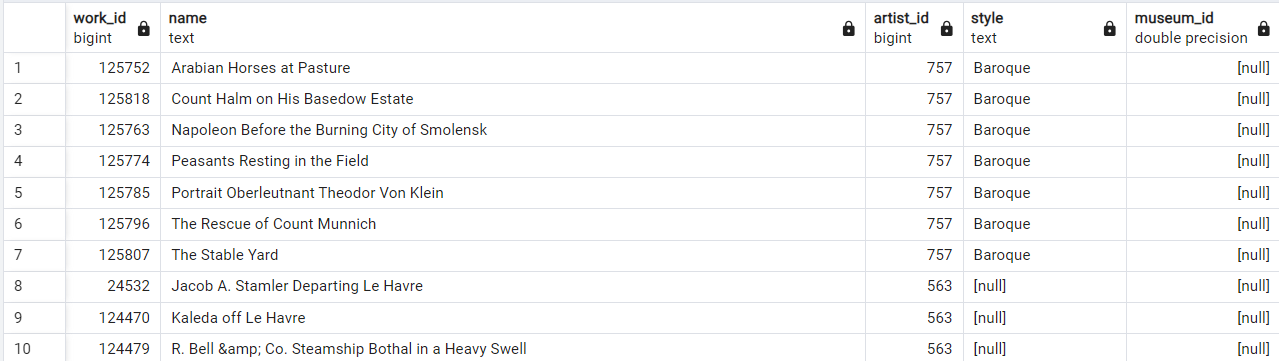
Famous Paintings Project

select \* from artist;  
select \* from canvas\_size;  
select \* from image\_link;  
select \* from museum;  
select \* from museum\_hours;  
select \* from product\_size;  
select \* from subject;  
select \* from work;  
  
Q1) Fetch all the paintings which are not displayed on any museums?  
select \* from work  
where museum\_id is null;  


Q2) Are there museum’s without any paintings?  
select \*  
from museum m  
LEFT JOIN work w  
ON m.museum\_id = w.museum\_id  
Where w.work\_id is NULL;

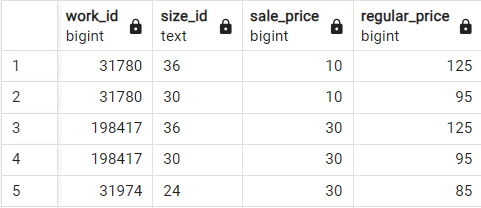
–OR-  
select \* from museum m  
 where not exists (select 1 from work w  
 where w.museum\_id=m.museum\_id)



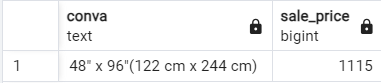
Q3) How many paintings have an asking price of more than their regular price?   
select count(\*) as total  
from product\_size  
WHERE sale\_price > regular\_price;



Q4) Identify the paintings whose asking price is less than 50% of its regular price.  
select \* from product\_size  
where sale\_price < (regular\_price \* 0.5);



Q5) Which canva size costs the most?  
SELECT c.label as canva, p.sale\_price as sale\_price  
FROM product\_size AS p  
JOIN canvas\_size as c ON p.size\_id = c.size\_id::text  
ORDER BY p.sale\_price DESC  
LIMIT 1;  
 --OR --  
select c.label as conva, p.sale\_price as sale\_price  
from ( select \*,  
 rank() over(order by sale\_price desc) as rnk  
 from product\_size) p  
join canvas\_size as c   
on c.size\_id = p.size\_id::bigint  
where p.rnk=1;

  
  
In the product\_size table, the size\_id is in text format, so we typecast it to bigint to ensure that both size\_id columns have the same datatype. This is necessary because the size\_id in the canvas\_size table is of datatype bigint. By typecasting, we maintain consistent data types for proper comparisons and joins.  
  
Q6) Delete duplicate records from work, product\_size, subject and image\_link tables  
delete from work

where ctid not in (select min(ctid)

from work

group by work\_id );

delete from product\_size

where ctid not in (select min(ctid)

from product\_size

group by work\_id, size\_id );

delete from subject

where ctid not in (select min(ctid)

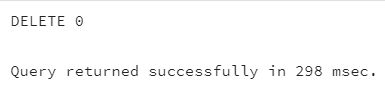
from subject

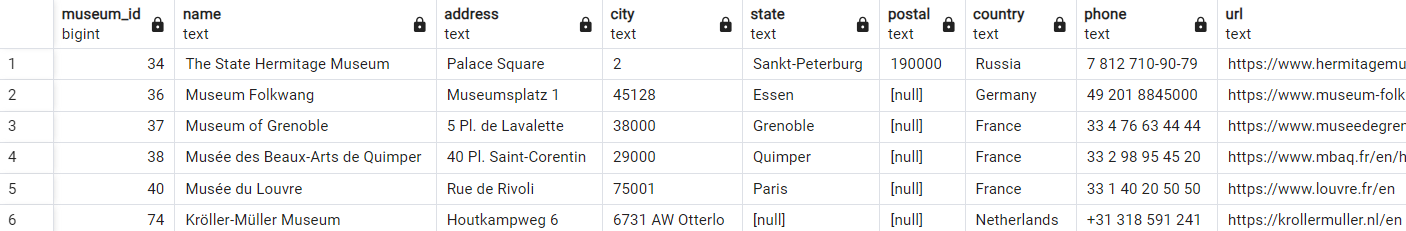
group by work\_id, subject );

delete from image\_link

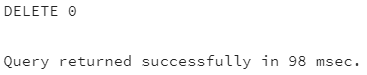
where ctid not in (select min(ctid)

from image\_link

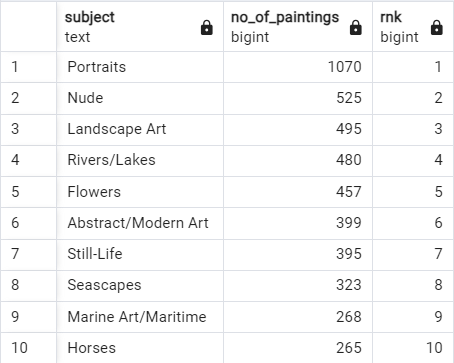
group by work\_id );  


Here,  
ctid is a system column in PostgreSQL that uniquely identifies rows based on their  
physical location in the table. The subquery finds the smallest ctid for each work\_id,  
while the DELETE query removes all other rows, ensuring only one instance per work\_id remains.  
  
Q7) Identify the museums with invalid city information in the given dataset  
select \* from museum   
where city ~ '^[0-9]'  
The ~ operator checks if city starts with a digit, using ^ to mark the beginning   
and [0-9] to match any digit, returning rows where the city field begins with a number,   
typically indicating invalid city names.

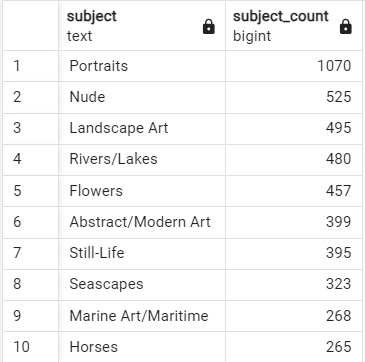
Q8) Museum\_Hours table has 1 invalid entry. Identify it and remove it.  
delete from museum\_hours   
 where ctid not in (select min(ctid)  
 from museum\_hours  
 group by museum\_id, day );



Q9) Fetch the top 10 most famous painting subject.  
1ST With Ranking Window Functions (RANK).  
SELECT \*  
FROM (  
 SELECT s.subject,   
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
   
 FROM work w  
 JOIN subject s ON s.work\_id = w.work\_id  
 GROUP BY s.subject  
) x  
WHERE rnk <= 10;

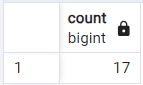


2ND With JOIN and Grouping.  
select distinct subject, count(\*) as subject\_count  
from subject s  
join work w on s.work\_id=w.work\_id  
group by subject  
order by subject\_count desc  
limit 10;  
  
3rd With Simple Aggregation & Grouping  
SELECT subject, COUNT(\*) AS subject\_count  
FROM subject  
GROUP BY subject  
ORDER BY subject\_count DESC  
LIMIT 10;



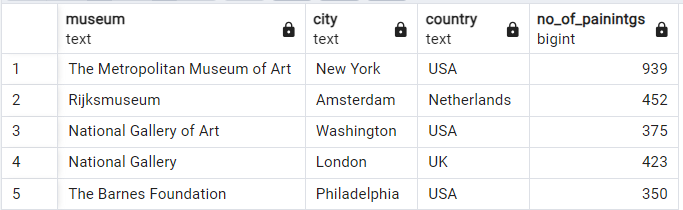
Q10) Identify the museums which are open on both Sunday and Monday. Display museum name, city.  
select m.museum\_id, m.name as museum\_name, m.city   
from museum\_hours mh  
join museum m on m.museum\_id = mh.museum\_id  
where day='Sunday'  
and exists (select 1 from museum\_hours mh2  
 where mh2.museum\_id = mh.museum\_id  
 and mh2.day='Monday');  
 EXISTS is used for subquery filtering, checking if a museum\_id is open on both 'Sunday' and 'Monday', with SELECT 1 simply confirming row existence.

Q11) How many museums are open every single day?  
select count(\*) from (  
 select museum\_id, count(\*)   
 from museum\_hours  
 group by museum\_id  
 having count(\*) =7) x;

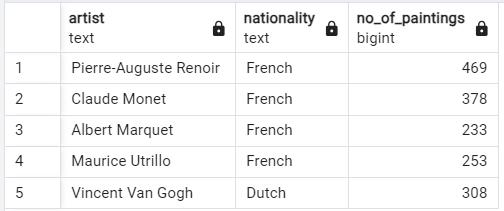


Q12) Which are the top 5 most popular museum? (Popularity is defined based on most no of paintings in a museum)

select m.name as museum, m.city,m.country,x.no\_of\_painintgs  
from ( select m.museum\_id, count(1) as no\_of\_painintgs  
 , rank() over(order by count(1) desc) as rnk  
 from work w  
 join museum m on m.museum\_id=w.museum\_id  
 group by m.museum\_id) x  
join museum m on m.museum\_id=x.museum\_id  
where x.rnk<=5;  
  
 --2nd with CTE--  
WITH ranked\_museums AS (  
 SELECT m.museum\_id,   
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 GROUP BY m.museum\_id  
)  
SELECT m.name AS museum, m.city, m.country, rm.no\_of\_paintings  
FROM ranked\_museums rm  
JOIN museum m ON m.museum\_id = rm.museum\_id  
WHERE rm.rnk <= 5;

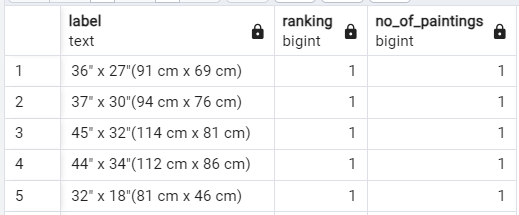


Q13) Who are the top 5 most popular artist? (Popularity is defined based on most no of paintings done by an artist)  
  
SELECT a.full\_name AS artist, a.nationality, x.no\_of\_paintings  
FROM (  
 SELECT a.artist\_id, COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN artist a ON a.artist\_id = w.artist\_id  
 GROUP BY a.artist\_id  
) x  
JOIN artist a ON a.artist\_id = x.artist\_id  
WHERE x.rnk <= 5;  
  
 --2nd with CTE--  
  
WITH ranked\_artists AS (  
 SELECT a.artist\_id,   
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN artist a ON a.artist\_id = w.artist\_id  
 GROUP BY a.artist\_id  
)  
SELECT a.full\_name AS artist, a.nationality, ra.no\_of\_paintings  
FROM ranked\_artists ra  
JOIN artist a ON a.artist\_id = ra.artist\_id  
WHERE ra.rnk <= 5;

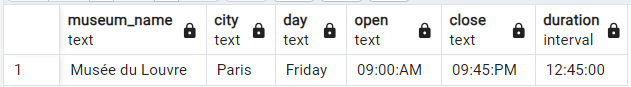


Q14) Display the 3 least popular canva sizes WITH Derived Table.

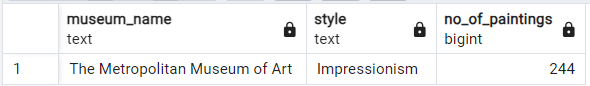
SELECT label, ranking, no\_of\_paintings  
FROM (  
 SELECT cs.size\_id, cs.label, COUNT(1) AS no\_of\_paintings,  
 DENSE\_RANK() OVER (ORDER BY COUNT(1)) AS ranking  
 FROM work w  
 JOIN product\_size ps ON ps.work\_id = w.work\_id  
 JOIN canvas\_size cs ON cs.size\_id::text = ps.size\_id  
 GROUP BY cs.size\_id, cs.label  
) x  
WHERE x.ranking <= 3;  
  
 --2nd with CTE--   
  
WITH ranked\_sizes AS (  
 SELECT cs.size\_id, cs.label, COUNT(1) AS no\_of\_paintings,  
 DENSE\_RANK() OVER (ORDER BY COUNT(1)) AS ranking  
 FROM work w  
 JOIN product\_size ps ON ps.work\_id = w.work\_id  
 JOIN canvas\_size cs ON cs.size\_id::text = ps.size\_id  
 GROUP BY cs.size\_id, cs.label  
)  
SELECT label, ranking, no\_of\_paintings  
FROM ranked\_sizes  
WHERE ranking <= 3;



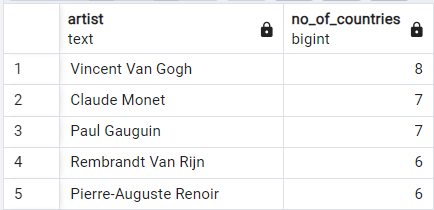
Q15) Which museum is open for the longest during a day. Dispay museum name, state and hours open and which day?  
SELECT museum\_name, state AS city, day, open, close, duration  
FROM (  
 SELECT m.name AS museum\_name, m.state, day, open, close,  
 to\_timestamp(open, 'HH:MI AM') AS open\_time,  
 to\_timestamp(close, 'HH:MI PM') AS close\_time,  
 to\_timestamp(close, 'HH:MI PM') - to\_timestamp(open, 'HH:MI AM') AS duration,  
 RANK() OVER (ORDER BY (to\_timestamp(close, 'HH:MI PM') - to\_timestamp(open, 'HH:MI AM')) DESC) AS rnk  
 FROM museum\_hours mh  
 JOIN museum m ON m.museum\_id = mh.museum\_id  
) x  
WHERE x.rnk = 1;  
  
 --2nd with CTE--   
WITH ranked\_museum\_hours AS (  
 SELECT m.name AS museum\_name, m.state, day, open, close,  
 to\_timestamp(open, 'HH:MI AM') AS open\_time,  
 to\_timestamp(close, 'HH:MI PM') AS close\_time,  
 to\_timestamp(close, 'HH:MI PM') - to\_timestamp(open, 'HH:MI AM') AS duration,  
 RANK() OVER (ORDER BY (to\_timestamp(close, 'HH:MI PM') - to\_timestamp(open, 'HH:MI AM')) DESC) AS rnk  
 FROM museum\_hours mh  
 JOIN museum m ON m.museum\_id = mh.museum\_id  
)  
SELECT museum\_name, state AS city, day, open, close, duration  
FROM ranked\_museum\_hours  
WHERE rnk = 1;



Q16) Which museum has the most no of most popular painting style?  
WITH pop\_style AS   
 (SELECT style,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work  
 GROUP BY style),  
cte AS  
 (SELECT w.museum\_id,  
 m.name AS museum\_name,  
 w.style,  
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 JOIN pop\_style ps ON ps.style = w.style  
 WHERE w.museum\_id IS NOT NULL  
 AND ps.rnk = 1  
 GROUP BY w.museum\_id, m.name, w.style)  
SELECT museum\_name, style, no\_of\_paintings  
FROM cte   
WHERE rnk = 1;

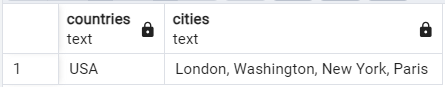


Q17) Identify the artists whose paintings are displayed in multiple countries.  
WITH cte AS (  
 SELECT DISTINCT a.full\_name AS artist,  
 m.country  
 FROM work w  
 JOIN artist a ON a.artist\_id = w.artist\_id  
 JOIN museum m ON m.museum\_id = w.museum\_id  
)  
SELECT artist, COUNT(1) AS no\_of\_countries  
FROM cte  
GROUP BY artist  
HAVING COUNT(1) > 1  
ORDER BY no\_of\_countries DESC;

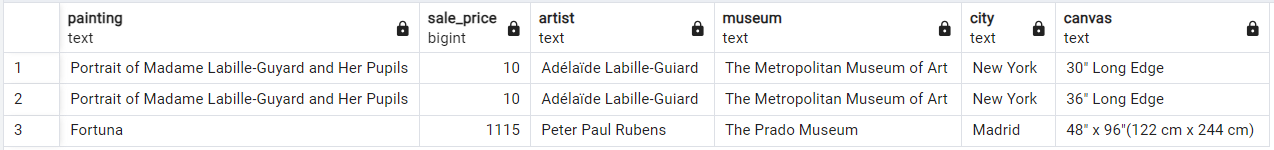


Q18) Display the country and the city with most no of museums. Output 2 seperate columns to mention the city and country. If there are multiple value, seperate them with comma.  
  
WITH cte\_country AS (  
 SELECT country,  
 COUNT(1) AS country\_count,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM museum  
 GROUP BY country  
),  
cte\_city AS (  
 SELECT city,  
 COUNT(1) AS city\_count,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM museum  
 GROUP BY city  
)  
SELECT   
 STRING\_AGG(DISTINCT country.country, ', ') AS countries,  
 STRING\_AGG(city.city, ', ') AS cities  
FROM cte\_country country  
CROSS JOIN cte\_city city  
WHERE country.rnk = 1  
 AND city.rnk = 1;

--WITH SUBQUERY--  
SELECT   
 STRING\_AGG(DISTINCT country, ', ') AS countries,  
 STRING\_AGG(city, ', ') AS cities  
FROM (  
 SELECT country,   
 COUNT(1) AS country\_count,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM museum  
 GROUP BY country  
) AS country\_data  
JOIN (  
 SELECT city,  
 COUNT(1) AS city\_count,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM museum  
 GROUP BY city  
) AS city\_data ON country\_data.rnk = 1 AND city\_data.rnk = 1  
GROUP BY country\_data.rnk, city\_data.rnk;



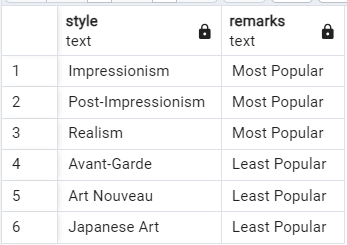
Q19) Identify the artist and the museum where the most expensive and least expensive painting is placed. Display the artist name, sale\_price, painting name, museum name, museum city and canvas label.  
SELECT w.name AS painting,  
 ps.sale\_price,  
 a.full\_name AS artist,  
 m.name AS museum,  
 m.city,  
 cz.label AS canvas  
FROM work w  
JOIN product\_size ps ON ps.work\_id = w.work\_id  
JOIN museum m ON m.museum\_id = w.museum\_id  
JOIN artist a ON a.artist\_id = w.artist\_id  
JOIN canvas\_size cz ON cz.size\_id = ps.size\_id::NUMERIC  
WHERE ps.sale\_price = (SELECT MAX(sale\_price) FROM product\_size)   
 OR ps.sale\_price = (SELECT MIN(sale\_price) FROM product\_size);  
  
 --WITH CTE--  
WITH cte AS (  
 SELECT \*,  
 RANK() OVER (ORDER BY sale\_price DESC) AS rnk,  
 RANK() OVER (ORDER BY sale\_price) AS rnk\_asc  
 FROM product\_size  
)  
SELECT w.name AS painting,  
 cte.sale\_price,  
 a.full\_name AS artist,  
 m.name AS museum,  
 m.city,  
 cz.label AS canvas  
FROM cte  
JOIN work w ON w.work\_id = cte.work\_id  
JOIN museum m ON m.museum\_id = w.museum\_id  
JOIN artist a ON a.artist\_id = w.artist\_id  
JOIN canvas\_size cz ON cz.size\_id = cte.size\_id::NUMERIC  
WHERE rnk = 1 OR rnk\_asc = 1;

  
  
Q20) Which country has the 5th highest no of paintings?  
WITH cte AS (  
 SELECT m.country,  
 COUNT(1) AS no\_of\_Paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 GROUP BY m.country  
)  
SELECT country,   
 no\_of\_Paintings  
FROM cte   
WHERE rnk = 5;

--WITH SUBQUERY--  
SELECT country,  
 no\_of\_Paintings  
FROM (  
 SELECT m.country,  
 COUNT(1) AS no\_of\_Paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 GROUP BY m.country  
) AS ranked\_countries  
WHERE rnk = 5;

  
  
Q21) Which are the 3 most popular and 3 least popular painting styles?  
WITH cte AS (  
 SELECT style,  
 COUNT(1) AS cnt,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk,  
 COUNT(1) OVER () AS no\_of\_records  
 FROM work  
 WHERE style IS NOT NULL  
 GROUP BY style  
)  
SELECT style,  
 CASE   
 WHEN rnk <= 3 THEN 'Most Popular'   
 ELSE 'Least Popular'   
 END AS remarks   
FROM cte  
WHERE rnk <= 3  
 OR rnk > no\_of\_records - 3;

--WITH SUBQUERY--  
SELECT style,  
 CASE   
 WHEN rnk <= 3 THEN 'Most Popular'   
 ELSE 'Least Popular'   
 END AS remarks   
FROM (  
 SELECT style,  
 COUNT(1) AS cnt,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk,  
 COUNT(1) OVER () AS no\_of\_records  
 FROM work  
 WHERE style IS NOT NULL  
 GROUP BY style  
) AS ranked\_styles  
WHERE rnk <= 3  
 OR rnk > no\_of\_records - 3;



Q22) Which artist has the most no of Portraits paintings outside USA?. Display artist name, no of paintings and the artist nationality.  
WITH artist\_paintings AS (  
 SELECT a.full\_name,  
 a.nationality,  
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN artist a ON a.artist\_id = w.artist\_id  
 JOIN subject s ON s.work\_id = w.work\_id  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 WHERE s.subject = 'Portraits'  
 AND m.country != 'USA'  
 GROUP BY a.full\_name, a.nationality  
)  
SELECT full\_name AS artist\_name,  
 nationality,  
 no\_of\_paintings  
FROM artist\_paintings  
WHERE rnk = 1;  
  
 --WITH SUBQUERY--  
SELECT full\_name AS artist\_name,  
 nationality,  
 no\_of\_paintings  
FROM (  
 SELECT a.full\_name,  
 a.nationality,  
 COUNT(1) AS no\_of\_paintings,  
 RANK() OVER (ORDER BY COUNT(1) DESC) AS rnk  
 FROM work w  
 JOIN artist a ON a.artist\_id = w.artist\_id  
 JOIN subject s ON s.work\_id = w.work\_id  
 JOIN museum m ON m.museum\_id = w.museum\_id  
 WHERE s.subject = 'Portraits'  
 AND m.country != 'USA'  
 GROUP BY a.full\_name, a.nationality  
) AS x  
WHERE rnk = 1;

  
  
  
**Conclusion**

The **Famous Paintings SQL Analysis Project** offered a comprehensive exploration of a dataset that encompasses various aspects of artworks, including their pricing and the museums that house them. Through the use of SQL queries, I was able to analyze intricate relationships among paintings, artists, and museums, uncovering significant patterns in pricing strategies and geographical distribution.

This project not only provided me with valuable insights into the art market but also allowed me to deepen my analytical capabilities. I gained a better understanding of how different factors influence artwork valuation, including market trends and artist reputation. Furthermore, I recognized the critical importance of data integrity, as accurate and reliable data is essential for drawing meaningful conclusions.

Overall, this project served as a significant learning experience, enabling me to bridge the gap between theoretical knowledge and practical application in data analysis. The skills and insights I acquired throughout this project will undoubtedly inform my future endeavors in data science and analytics, equipping me to tackle more complex challenges in the field.