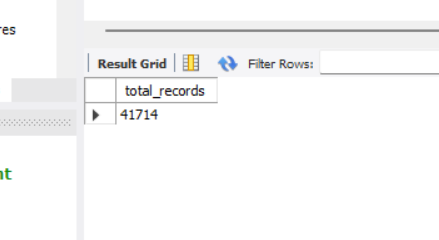
**1. How many records are there in the dataset?**

SELECT COUNT(\*) AS total\_records

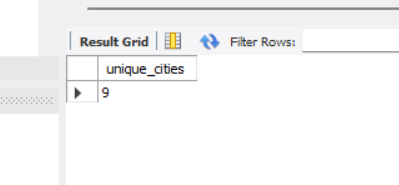
FROM fact\_data;



**2. How many unique cities are in the European dataset?**

SELECT COUNT(DISTINCT CITY) AS unique\_cities

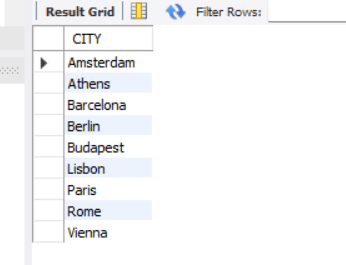
FROM airbnb\_europe\_dataset;



**3. What are the names of the cities in the dataset?**

SELECT DISTINCT CITY

FROM airbnb\_europe\_dataset;

****

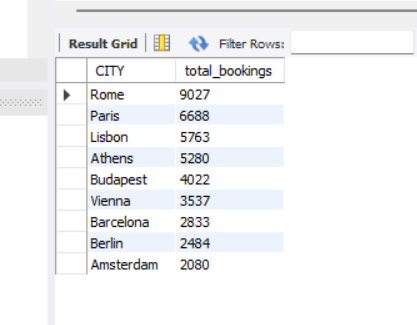
4. **How many bookings are there in each city?**

SELECT CITY, COUNT(\*) AS total\_bookings

FROM airbnb\_europe\_dataset

GROUP BY CITY

ORDER BY total\_bookings DESC;



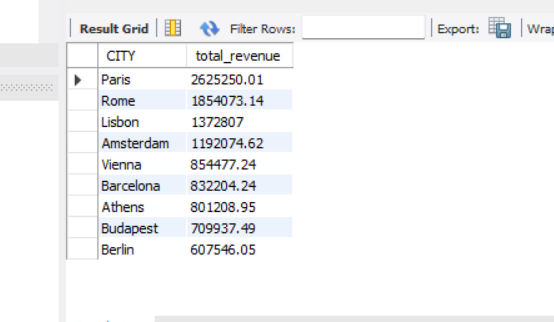
5. **What is the total booking revenue for each city?**

SELECT CITY, ROUND(SUM(PRICE), 2) AS total\_revenue

FROM airbnb\_europe\_dataset

GROUP BY CITY

ORDER BY total\_revenue DESC;



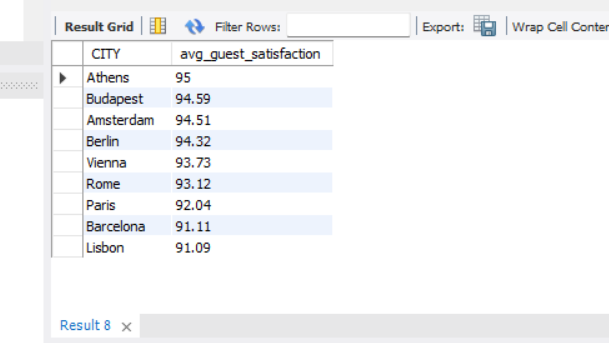
6. **What is the average guest satisfaction score for each city?**

SELECT CITY, ROUND(AVG(Guest\_Satisfaction), 2) AS avg\_guest\_satisfaction

FROM airbnb\_europe\_dataset

GROUP BY CITY

ORDER BY avg\_guest\_satisfaction DESC;



7. **What are the minimum, maximum, average, and median booking prices?**

SELECT

MIN(PRICE) AS min\_price,

MAX(PRICE) AS max\_price,

ROUND(AVG(PRICE), 2) AS avg\_price,

ROUND(

IF(

MOD(COUNT(\*), 2) = 1,

-- For odd counts, get the middle value

SUBSTRING\_INDEX(SUBSTRING\_INDEX(GROUP\_CONCAT(PRICE ORDER BY PRICE), ',', (COUNT(\*) + 1) / 2), ',', -1),

-- For even counts, average the two middle values

(

SUBSTRING\_INDEX(SUBSTRING\_INDEX(GROUP\_CONCAT(PRICE ORDER BY PRICE), ',', COUNT(\*) / 2), ',', -1) +

SUBSTRING\_INDEX(SUBSTRING\_INDEX(GROUP\_CONCAT(PRICE ORDER BY PRICE), ',', COUNT(\*) / 2 + 1), ',', -1)

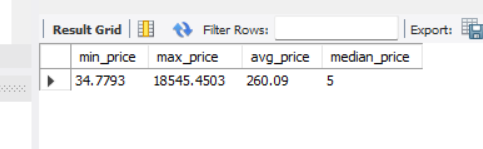
) / 2

),

2

) AS median\_price

FROM airbnb\_europe\_dataset;



8. **How many outliers are there in the price field?**

WITH PriceQuartiles AS (

SELECT

NTILE(4) OVER (ORDER BY PRICE) AS quartile,

PRICE

FROM airbnb\_europe\_dataset

),

IQR\_Calculation AS (

SELECT

MAX(CASE WHEN quartile = 1 THEN PRICE END) AS Q1,

MAX(CASE WHEN quartile = 3 THEN PRICE END) AS Q3

FROM PriceQuartiles

),

OutlierBounds AS (

SELECT

Q1,

Q3,

Q3 - Q1 AS IQR,

Q1 - 1.5 \* (Q3 - Q1) AS lower\_bound,

Q3 + 1.5 \* (Q3 - Q1) AS upper\_bound

FROM IQR\_Calculation

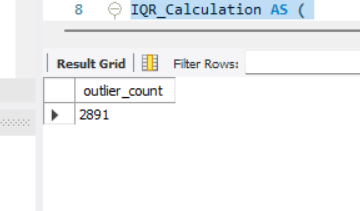
)

SELECT

COUNT(\*) AS outlier\_count

FROM airbnb\_europe\_dataset, OutlierBounds

WHERE PRICE < lower\_bound OR PRICE > upper\_bound;



**9. What are the characteristics of the outliers in terms of room type, number of bookings, and price?**

WITH PriceQuartiles AS (

SELECT

NTILE(4) OVER (ORDER BY PRICE) AS quartile,

PRICE

FROM airbnb\_europe\_dataset

),

IQR\_Calculation AS (

SELECT

MAX(CASE WHEN quartile = 1 THEN PRICE END) AS Q1,

MAX(CASE WHEN quartile = 3 THEN PRICE END) AS Q3

FROM PriceQuartiles

),

OutlierBounds AS (

SELECT

Q1,

Q3,

Q3 - Q1 AS IQR,

Q1 - 1.5 \* (Q3 - Q1) AS lower\_bound,

Q3 + 1.5 \* (Q3 - Q1) AS upper\_bound

FROM IQR\_Calculation

),

Outliers AS (

SELECT \*

FROM airbnb\_europe\_dataset, OutlierBounds

WHERE PRICE < lower\_bound OR PRICE > upper\_bound

)

SELECT

ROOM\_TYPE,

COUNT(\*) AS total\_outliers,

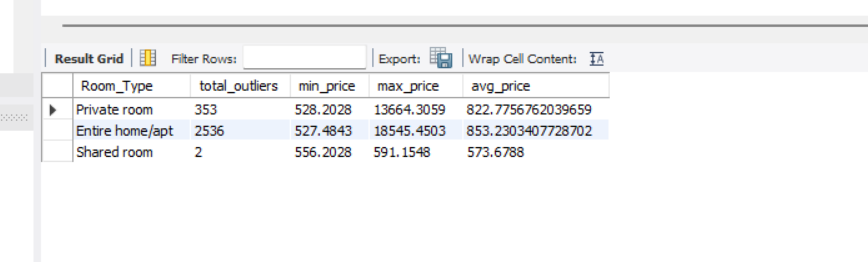
MIN(PRICE) AS min\_price,

MAX(PRICE) AS max\_price,

AVG(PRICE) AS avg\_price

FROM Outliers

GROUP BY ROOM\_TYPE;



10. How does the average price differ between the main dataset and the dataset with outliers removed?

WITH quartiles AS (

SELECT

Price,

NTILE(4) OVER (ORDER BY Price) AS quartile

FROM airbnb\_europe\_dataset

),

iqr\_bounds AS (

SELECT

MIN(CASE WHEN quartile = 1 THEN Price END) AS Q1,

MIN(CASE WHEN quartile = 3 THEN Price END) AS Q3

FROM quartiles

),

outlier\_bounds AS (

SELECT

Q1,

Q3,

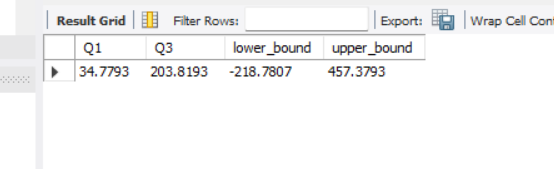
Q1 - 1.5 \* (Q3 - Q1) AS lower\_bound,

Q3 + 1.5 \* (Q3 - Q1) AS upper\_bound

FROM iqr\_bounds

)

SELECT \* FROM outlier\_bounds;



11.

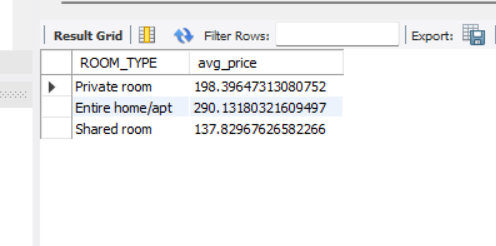
SELECT

ROOM\_TYPE,

AVG(PRICE) AS avg\_price

FROM airbnb\_europe\_dataset

GROUP BY ROOM\_TYPE;



**12. How do weekend and weekday bookings compare in terms of average price and number of bookings?**

SELECT

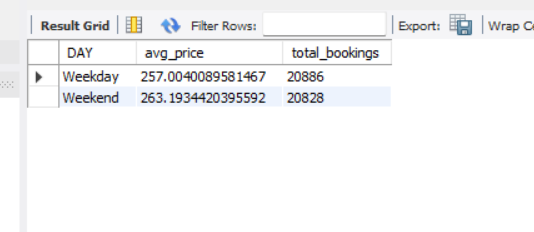
DAY,

AVG(PRICE) AS avg\_price,

COUNT(\*) AS total\_bookings

FROM airbnb\_europe\_dataset

GROUP BY DAY;



**13. Average Distance for Each City**

SELECT

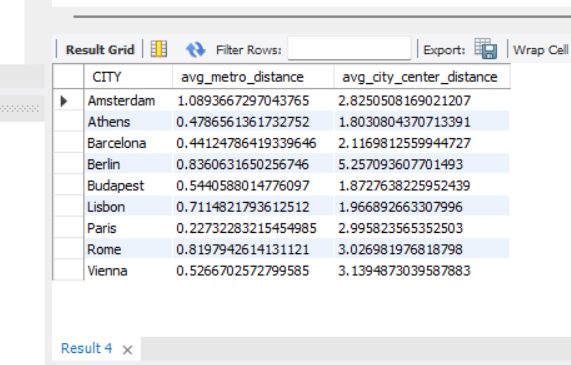
CITY,

AVG(METRO\_DISTANCE\_KM) AS avg\_metro\_distance,

AVG(CITY\_CENTER\_KM) AS avg\_city\_center\_distance

FROM airbnb\_europe\_dataset

GROUP BY CITY;



**14. Bookings by Room Type on Weekdays vs Weekends**

SELECT

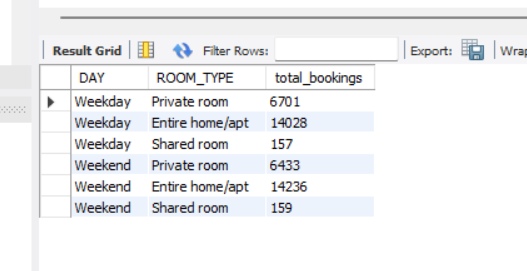
DAY,

ROOM\_TYPE,

COUNT(\*) AS total\_bookings

FROM airbnb\_europe\_dataset

GROUP BY DAY, ROOM\_TYPE;



**15. Booking Revenue by Room Type on Weekdays vs Weekends**

SELECT

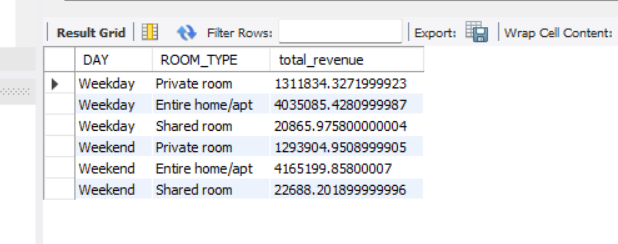
DAY,

ROOM\_TYPE,

SUM(PRICE) AS total\_revenue

FROM airbnb\_europe\_dataset

GROUP BY DAY, ROOM\_TYPE;



**16. Guest Satisfaction Score Summary**

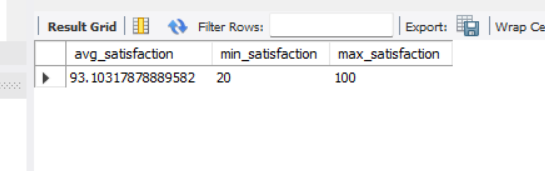
SELECT

AVG(GUEST\_SATISFACTION) AS avg\_satisfaction,

MIN(GUEST\_SATISFACTION) AS min\_satisfaction,

MAX(GUEST\_SATISFACTION) AS max\_satisfaction

FROM airbnb\_europe\_dataset;



**17. Guest Satisfaction Score by City**

SELECT

CITY,

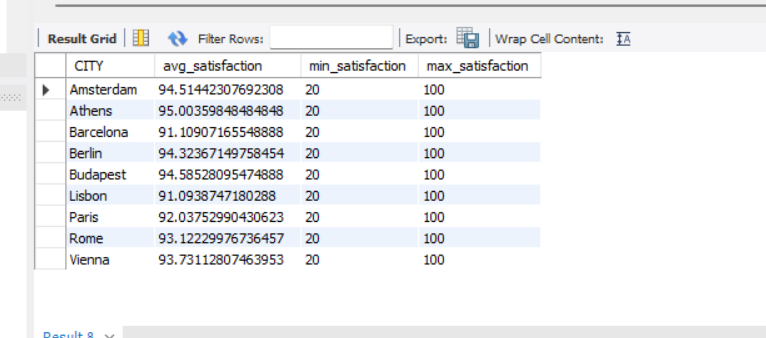
AVG(GUEST\_SATISFACTION) AS avg\_satisfaction,

MIN(GUEST\_SATISFACTION) AS min\_satisfaction,

MAX(GUEST\_SATISFACTION) AS max\_satisfaction

FROM airbnb\_europe\_dataset

GROUP BY CITY;



**19. Average Booking Value Across Cleaned Data**

WITH PriceQuartiles AS (

SELECT

NTILE(4) OVER (ORDER BY PRICE) AS quartile,

PRICE

FROM airbnb\_europe\_dataset

),

IQR\_Calculation AS (

SELECT

MAX(CASE WHEN quartile = 1 THEN PRICE END) AS Q1,

MAX(CASE WHEN quartile = 3 THEN PRICE END) AS Q3

FROM PriceQuartiles

),

OutlierBounds AS (

SELECT

Q1,

Q3,

Q3 - Q1 AS IQR,

Q1 - 1.5 \* (Q3 - Q1) AS lower\_bound,

Q3 + 1.5 \* (Q3 - Q1) AS upper\_bound

FROM IQR\_Calculation

),

CleanedData AS (

SELECT \*

FROM airbnb\_europe\_dataset, OutlierBounds

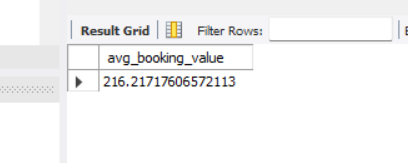
WHERE PRICE BETWEEN lower\_bound AND upper\_bound

)

SELECT

AVG(PRICE) AS avg\_booking\_value

FROM CleanedData;



**20. Average Cleanliness Score Across Cleaned Data**

WITH PriceQuartiles AS (

SELECT

NTILE(4) OVER (ORDER BY PRICE) AS quartile,

PRICE

FROM airbnb\_europe\_dataset

),

IQR\_Calculation AS (

SELECT

MAX(CASE WHEN quartile = 1 THEN PRICE END) AS Q1,

MAX(CASE WHEN quartile = 3 THEN PRICE END) AS Q3

FROM PriceQuartiles

),

OutlierBounds AS (

SELECT

Q1,

Q3,

Q3 - Q1 AS IQR,

Q1 - 1.5 \* (Q3 - Q1) AS lower\_bound,

Q3 + 1.5 \* (Q3 - Q1) AS upper\_bound

FROM IQR\_Calculation

),

CleanedData AS (

SELECT \*

FROM airbnb\_europe\_dataset, OutlierBounds

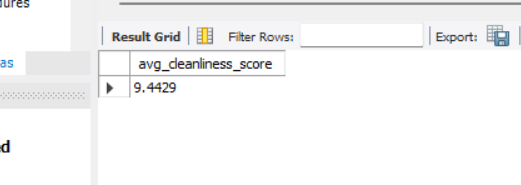
WHERE PRICE BETWEEN lower\_bound AND upper\_bound

)

SELECT

AVG(CLEANLINESS\_RATING) AS avg\_cleanliness\_score

FROM CleanedData;



**21. Cities Ranked by Total Revenue**

SELECT

CITY,

SUM(PRICE) AS total\_revenue,

ROW\_NUMBER() OVER (ORDER BY SUM(PRICE) DESC) AS revenue\_rank

FROM airbnb\_europe\_dataset

GROUP BY CITY

ORDER BY revenue\_rank;

