**Document**

1. How many records are there in the dataset?

Code:

**SELECT count(\*) AS Number\_Of\_Records**

**FROM airbnb\_data;**

Output:



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

Code:

**SELECT COUNT(DISTINCT CITY) AS cities**

**FROM airbnb\_data;**

Output:



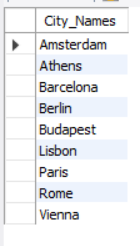
3. What are the names of the cities in the dataset?

Code:

**SELECT DISTINCT city AS City\_Names**

**FROM airbnb\_data;**

Output:



4. How many bookings are there in each city?

Code:

**SELECT city,**

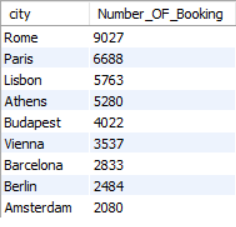
**count(city) AS Number\_OF\_Booking**

**FROM airbnb\_data**

**GROUP BY city**

**ORDER BY 2 DESC ;**

Output:



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

Code:

**SELECT city,**

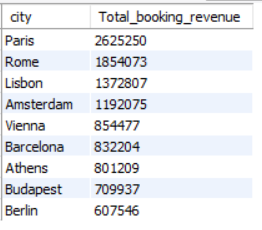
**round (sum(PRICE)) AS Total\_booking\_revenue**

**FROM airbnb\_data**

**GROUP BY city**

**ORDER BY 2 DESC ;**

Output:



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

Code:

**SELECT city,**

**round(avg(GUEST\_SATISFACTION),**

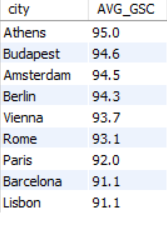
**1) AS AVG\_GSC**

**FROM airbnb\_data**

**GROUP BY city**

**ORDER BY 2 DESC ;**

Output:



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

Code:

**-- Minimum price**

**SELECT ROUND(MIN(PRICE),**

**2) AS minimum**

**FROM airbnb\_data;**

**-- Maximum price**

**SELECT ROUND(MAX(PRICE),**

**2) AS maximum**

**FROM airbnb\_data;**

**-- Average price**

**SELECT ROUND(AVG(PRICE),**

**2) AS average**

**FROM airbnb\_data;**

**-- Median price**

**WITH cte1 AS (**

**SELECT**

**PRICE,**

**NTILE(2)**

**OVER (ORDER BY PRICE) AS price\_group**

**FROM airbnb\_data**

**)**

**SELECT ROUND(AVG(PRICE),**

**2) AS median**

**FROM cte1**

**WHERE price\_group = 1**

**OR price\_group = 2;**

Output:



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

Code:

**WITH cte1 AS (**

**SELECT PRICE,**

**NTILE(4)**

**OVER (ORDER BY PRICE) AS quartile**

**FROM airbnb\_data**

**),**

**cte2 AS (**

**SELECT**

**MAX(**

**CASE**

**WHEN quartile = 1 THEN**

**PRICE END) AS Q1,**

**MAX(**

**CASE**

**WHEN quartile = 3 THEN**

**PRICE END) AS Q3**

**FROM cte1**

**),**

**cte3 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 cte2**

**)**

**SELECT**

**COUNT(\*) AS outliers**

**FROM airbnb\_data,**

**cte3**

**WHERE PRICE < lower\_bound**

**OR PRICE > upper\_bound;**

Output:



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

Code:

**CREATE**

**OR REPLACE VIEW OUTLIER AS**

**(**

**WITH PRICE\_QUARTILES AS (**

**SELECT**

**PRICE,**

**NTILE(4)**

**OVER (ORDER BY PRICE) AS quartile**

**FROM airbnb\_data**

**),**

**FIVE\_NUMBER\_SUMMARY AS (**

**SELECT**

**MIN(PRICE) AS MIN\_ORDER\_VALUE,**

**MAX(PRICE) AS MAX\_ORDER\_VALUE,**

**MAX(**

**CASE**

**WHEN quartile = 1 THEN**

**PRICE END) AS Q1,**

**MAX(**

**CASE**

**WHEN quartile = 2 THEN**

**PRICE END) AS MEDIAN,**

**MAX(**

**CASE**

**WHEN quartile = 3 THEN**

**PRICE END) AS Q3**

**FROM PRICE\_QUARTILES**

**),**

**HINGES AS (**

**SELECT**

**Q1,**

**Q3,**

**(Q3 - Q1) AS IQR,**

**(Q1 - 1.5 \* (Q3 - Q1)) AS LOWER\_HINGE,**

**(Q3 + 1.5 \* (Q3 - Q1)) AS UPPER\_HINGE**

**FROM FIVE\_NUMBER\_SUMMARY**

**)**

**SELECT**

**A.\***

**FROM airbnb\_data A**

**JOIN HINGES H**

**ON A.PRICE < H.LOWER\_HINGE**

**OR A.PRICE > H.UPPER\_HINGE**

**);**

**-- Check outlier data statistics (minimum,**

**average,**

**maximum) grouped by room type**

**SELECT**

**ROOM\_TYPE AS "Room\_Type",**

**COUNT(\*) AS "No\_Of\_Booking",**

**ROUND(MIN(PRICE),**

**1) AS "minimun\_outlier\_price",**

**ROUND(MAX(PRICE),**

**1) AS "maximum\_outlier\_price",**

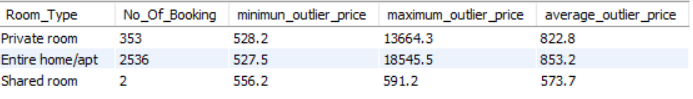
**ROUND(AVG(PRICE),**

**1) AS "average\_outlier\_price"**

**FROM OUTLIER**

**GROUP BY ROOM\_TYPE;**

Output:



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

Code:

**Step 1: Create**

**OR replace a view fOR the cleaned data (without outliers)**

**CREATE**

**OR REPLACE VIEW CLEANED\_AIRBNB\_DATA AS**

**(**

**WITH PRICE\_QUARTILES AS (**

**SELECT**

**PRICE,**

**NTILE(4)**

**OVER (ORDER BY PRICE) AS quartile**

**FROM AIRBNB\_DATA**

**),**

**FIVE\_NUMBER\_SUMMARY AS (**

**SELECT**

**MAX(**

**CASE**

**WHEN quartile = 1 THEN**

**PRICE END) AS Q1,**

**MAX(**

**CASE**

**WHEN quartile = 2 THEN**

**PRICE END) AS MEDIAN,**

**MAX(**

**CASE**

**WHEN quartile = 3 THEN**

**PRICE END) AS Q3**

**FROM PRICE\_QUARTILES**

**),**

**HINGES AS (**

**SELECT**

**Q1,**

**Q3,**

**(Q3 - Q1) AS IQR,**

**(Q1 - 1.5 \* (Q3 - Q1)) AS LOWER\_HINGE,**

**(Q3 + 1.5 \* (Q3 - Q1)) AS UPPER\_HINGE**

**FROM FIVE\_NUMBER\_SUMMARY**

**)**

**--SELECT rows**

**WHERE prices are withIN the bounds to form the cleaned dataset**

**SELECT**

**\***

**FROM AIRBNB\_DATA**

**WHERE PRICE >= (SELECT LOWER\_HINGE**

**FROM HINGES)**

**AND PRICE <= (SELECT UPPER\_HINGE**

**FROM HINGES)**

**);**

**-- Step 2: Calculate the average price fOR both datasets**

**AND compare**

**SELECT**

**'Original Data' AS DATASET,**

**ROUND(AVG(PRICE),**

**2) AS AVG\_PRICE**

**FROM AIRBNB\_DATA**

**UNION ALL**

**SELECT**

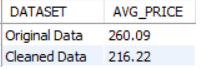
**'Cleaned Data' AS DATASET,**

**ROUND(AVG(PRICE),**

**2) AS AVG\_PRICE**

**FROM CLEANED\_AIRBNB\_DATA;**

Output:



11. What is the average price for each room type?

Code:

**SELECT**

**ROOM\_TYPE,**

**ROUND(AVG(price),**

**1) AS average\_price**

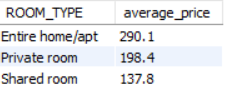
**FROM**

**airbnb\_data**

**GROUP BY ROOM\_TYPE**

**ORDER BY 2 DESC;**

Output:



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

Code:

**SELECT**

**DAY AS Day\_Type,**

**ROUND(AVG(PRICE),**

**0) AS Avg\_Price,**

**COUNT(ROOM\_TYPE) AS Num\_Of\_Booking**

**FROM cleaned\_airbnb\_data**

**GROUP BY DAY;**

Output:



13. What is the average distance from metro and city center for each city?

Code:

**SELECT**

**CITY,**

**ROUND(AVG(METRO\_DISTANCE\_KM),**

**2) AS AVG\_METRO\_DISTANCE\_KM,**

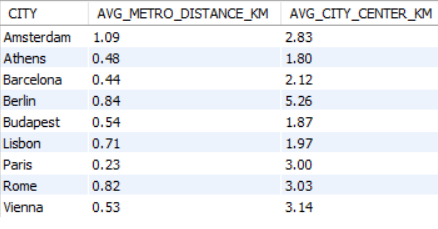
**ROUND(AVG(CITY\_CENTER\_KM),**

**2) AS AVG\_CITY\_CENTER\_KM**

**FROM AIRBNB\_DATA**

**GROUP BY CITY;**

Output:



14. How many bookings are there for each room type on weekdays vs weekends?

Code:

**SELECT**

**CASE**

**WHEN DAY IN ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday') THEN**

**'Weekday'**

**ELSE 'Weekend'**

**END AS DAY\_TYPE,**

**ROOM\_TYPE,**

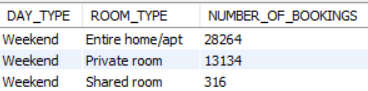
**COUNT(\*) AS NUMBER\_OF\_BOOKINGS**

**FROM AIRBNB\_DATA**

**GROUP BY DAY\_TYPE, ROOM\_TYPE**

**ORDER BY DAY\_TYPE, ROOM\_TYPE;**

Output:



15. What is the booking revenue for each room type on weekdays vs weekends?

Code:

**SELECT**

**CASE**

**WHEN DAY IN ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday') THEN**

**'Weekday'**

**ELSE 'Weekend'**

**END AS DAY\_TYPE,**

**ROOM\_TYPE,**

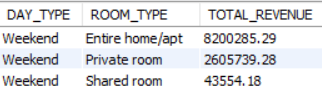
**ROUND(SUM(PRICE), 2) AS TOTAL\_REVENUE**

**FROM AIRBNB\_DATA**

**GROUP BY DAY\_TYPE, ROOM\_TYPE**

**ORDER BY DAY\_TYPE, ROOM\_TYPE;**

Output:



16. What is the overall average, minimum, and maximum guest satisfaction score?

Code:

**SELECT**

**ROUND(AVG(GUEST\_SATISFACTION),**

**2) AS AVG\_GUEST\_SATISFACTION,**

**MIN(GUEST\_SATISFACTION) AS MIN\_GUEST\_SATISFACTION,**

**MAX(GUEST\_SATISFACTION) AS MAX\_GUEST\_SATISFACTION**

**FROM AIRBNB\_DATA;**

Output:



17. How does guest satisfaction score vary by city?

Code:

**SELECT**

**CITY,**

**ROUND(AVG(GUEST\_SATISFACTION),**

**2) AS AVG\_GUEST\_SATISFACTION,**

**MIN(GUEST\_SATISFACTION) AS MIN\_GUEST\_SATISFACTION,**

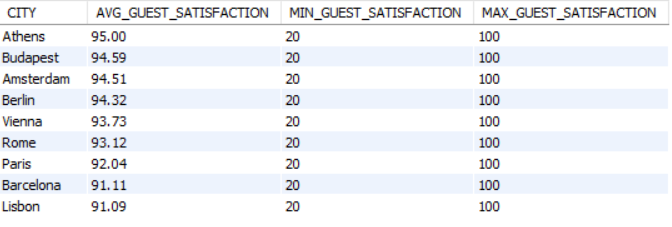
**MAX(GUEST\_SATISFACTION) AS MAX\_GUEST\_SATISFACTION**

**FROM AIRBNB\_DATA**

**GROUP BY CITY**

**ORDER BY AVG\_GUEST\_SATISFACTION DESC;**

Output:



18. What is the average booking value across all cleaned data?

Code:

**SELECT**

**ROUND(AVG(PRICE),**

**2) AS AVG\_BOOKING\_VALUE**

**FROM CLEANED\_AIRBNB\_DATA;**

Output:



19. What is the average cleanliness score across all cleaned data?

Code:

**SELECT**

**ROUND(AVG(CLEANLINESS\_RATING),**

**2) AS AVG\_CLEANLINESS\_SCORE**

**FROM CLEANED\_AIRBNB\_DATA;**

Output:



20. How do cities rank in terms of total revenue?

Code:

**WITH CTE1 AS (**

**SELECT**

**CITY,**

**SUM(PRICE) AS TOTAL\_REVENUE**

**FROM AIRBNB\_DATA**

**GROUP BY CITY**

**)**

**SELECT**

**CITY,**

**TOTAL\_REVENUE,**

**ROW\_NUMBER()**

**OVER (ORDER BY TOTAL\_REVENUE DESC) AS REVENUE\_RANK**

**FROM CTE1**

**ORDER BY REVENUE\_RANK;**

Output:

