

AIRBNB- SQL CODE SNIPPET

```
#AIRBNB SQL PROJECT
# DATA CLEANING AND PREPROCESSING
# Remove duplicates
Select * from `directed-racer-396805.airbnb.airbnb_ny_listing`
limit 5;

SELECT count (*) as row_count_original
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`;

SELECT
    DISTINCT COUNT(*) AS row_count_distinct
FROM
    `directed-racer-396805.airbnb.airbnb_ny_listing`;

-- SHAPE OF DATA
--Row we have already counted
# no of columns
SELECT
    COUNT(*) AS column_count
FROM
    `directed-racer-396805.airbnb.INFORMATION_SCHEMA.COLUMNS`
WHERE
    table_name = 'airbnb_ny_listing';

# Data types
#DATA TYPES
SELECT column_name,
        data_type
FROM
    `directed-racer-396805.airbnb.INFORMATION_SCHEMA.COLUMNS`
WHERE
    table_name = 'airbnb_ny_listing';

-- Check for missing values
SELECT
    SUM(CASE WHEN name IS NULL THEN 1 ELSE 0 END) AS missing_name,
    SUM(CASE WHEN host_id IS NULL THEN 1 ELSE 0 END) AS missing_host_id,
    SUM(CASE WHEN host_name IS NULL THEN 1 ELSE 0 END) AS missing_host_name,
    SUM(CASE WHEN neighbourhood_group IS NULL THEN 1 ELSE 0 END) AS missing_neighbourhood_group,
    SUM(CASE WHEN neighbourhood IS NULL THEN 1 ELSE 0 END) AS missing_neighbourhood,
    SUM(CASE WHEN latitude IS NULL THEN 1 ELSE 0 END) AS missing_latitude,
    SUM(CASE WHEN longitude IS NULL THEN 1 ELSE 0 END) AS missing_longitude,
    SUM(CASE WHEN price IS NULL THEN 1 ELSE 0 END) AS missing_price,
    SUM(CASE WHEN room_type IS NULL THEN 1 ELSE 0 END) AS missing_room_type,
    SUM(CASE WHEN minimum_nights IS NULL THEN 1 ELSE 0 END) AS missing_minimum_nights,
    SUM(CASE WHEN number_of_reviews IS NULL THEN 1 ELSE 0 END) AS missing_number_of_reveiw,
    SUM(CASE WHEN last_review IS NULL THEN 1 ELSE 0 END) AS missing_last_reveiw,
    SUM(CASE WHEN reviews_per_month IS NULL THEN 1 ELSE 0 END) AS missing_reveiw_per_month,
    SUM(CASE WHEN calculated_host_listings_count IS NULL THEN 1 ELSE 0 END) AS
missing_calculated_host_listings_count,
    SUM(CASE WHEN number_of_reviews_ltm IS NULL THEN 1 ELSE 0 END) AS
missing_number_of_reviews_ltm,
```

```

SUM(CASE WHEN license IS NULL THEN 1 ELSE 0 END) AS license,
SUM(CASE WHEN availability_365 IS NULL THEN 1 ELSE 0 END) AS missing_availability_365,
SUM(CASE WHEN rating IS NULL THEN 1 ELSE 0 END) AS rating,
SUM(CASE WHEN bedrooms IS NULL THEN 1 ELSE 0 END) AS bedrooms,
SUM(CASE WHEN beds IS NULL THEN 1 ELSE 0 END) AS beds,
SUM(CASE WHEN baths IS NULL THEN 1 ELSE 0 END) AS baths
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`;

```

--AVERAGE RATING BY NEIGHBOURHOOD

```

SELECT neighbourhood, AVG(cast(rating as decimal)) AS avg_rating
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE rating NOT IN ('No rating', 'New','New ')
GROUP BY neighbourhood
ORDER BY avg_rating DESC
LIMIT 20;
SELECT neighbourhood, AVG(cast(rating as decimal)) AS avg_rating
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE rating NOT IN ('No rating', 'New','New ')
GROUP BY neighbourhood
ORDER BY avg_rating
LIMIT 5;

```

--Distribution of rating

```

SELECT rating, COUNT(*) AS rating_count
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY rating
ORDER BY rating_count desc;

```

-- Average reviews based on room-type

```

SELECT room_type, AVG(number_of_reviews) AS average_reviews
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY room_type
order by average_reviews desc;

```

-- Top 10 listings by number of reviews

```

SELECT id, name, number_of_reviews
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
ORDER BY number_of_reviews DESC
LIMIT 10;

```

-- Hosts with Multiple Listings

```

SELECT host_id, host_name, COUNT(id) AS listings_count
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY host_id, host_name
HAVING COUNT(id) > 1
ORDER BY listings_count DESC;

```

```

SELECT distinct host_id
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`;

```

-- Average Rating by Room Type

```

SELECT room_type, AVG(cast(rating as decimal)) AS average_rating
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
Where rating not in ('No rating', 'New','New ')
GROUP BY room_type

```

```
ORDER BY average_rating DESC;
```

```
-- Review count and average rating by Neighbourhood
```

```
SELECT neighbourhood, COUNT(id) AS review_count, AVG(cast(rating as decimal)) AS  
average_rating  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
Where rating not in ('No rating', 'New','New ')  
GROUP BY neighbourhood  
ORDER BY review_count DESC  
limit 5;
```

```
-- Review count and average rating by Neighbourhood
```

```
SELECT neighbourhood, COUNT(id) AS review_count, AVG(cast(rating as decimal)) AS  
average_rating  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
Where rating not in ('No rating', 'New','New ')  
GROUP BY neighbourhood  
ORDER BY review_count  
limit 5;
```

```
-- Average Price by Neighbourhood
```

```
SELECT neighbourhood, AVG(price) AS average_price  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
GROUP BY neighbourhood  
ORDER BY average_price DESC;
```

```
SELECT neighbourhood, AVG(price) AS average_price  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
GROUP BY neighbourhood  
ORDER BY average_price -- lowest 5 avg price  
limit 5;
```

```
--Price distribution by room_type
```

```
SELECT room_type, MIN(price) AS min_price, MAX(price) AS max_price, AVG(price) AS  
average_price  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
GROUP BY room_type  
order by max_price desc  
limit 5;
```

```
--Seasonal price trends
```

```
SELECT EXTRACT(MONTH FROM last_review) AS month, AVG(price) AS average_price  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
WHERE last_review IS NOT NULL  
GROUP BY month  
ORDER BY average_price desc  
limit 5;
```

```
SELECT EXTRACT(MONTH FROM last_review) AS month, AVG(price) AS average_price  
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`  
WHERE last_review IS NOT NULL  
GROUP BY month  
ORDER BY average_price  
limit 1; -- the season with the lowest average price
```

```

--Pricing and rating correlation
SELECT price, rating
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE rating IS NOT NULL
group by price,rating
order by price desc;

SELECT price, rating
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE rating IS NOT NULL
group by price,rating
order by price
limit 5;-- lowest 5 prices and their rating

-- Competitive Price Analysis
SELECT neighbourhood, room_type, AVG(price) AS average_price
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY neighbourhood, room_type
ORDER BY average_price desc
limit 5;

SELECT neighbourhood, room_type, AVG(price) AS average_price
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY neighbourhood, room_type
ORDER BY average_price
limit 5;-- bottom 5 avg price

-- Average availability by Neighbourhood
SELECT neighbourhood, AVG(availability_365) AS average_availability
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY neighbourhood
ORDER BY average_availability DESC;

SELECT neighbourhood, AVG(availability_365) AS average_availability
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY neighbourhood
ORDER BY average_availability
limit 5;-- top 5 places which are not easily available

--Impact of minimum nights on availability
SELECT minimum_nights, AVG(availability_365) AS average_availability
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY minimum_nights
ORDER BY average_availability;

SELECT minimum_nights, AVG(availability_365) AS average_availability
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY minimum_nights
ORDER BY average_availability desc
limit 5;-- now checking the minimum nights where the listings are mostly available

--Listings with 100% availability
SELECT id, name, neighbourhood, price
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE availability_365 = 365;

```

--Occupancy rate by Room-type

```
SELECT room_type, AVG(availability_365) AS average_availability
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY room_type
ORDER BY average_availability DESC;
```

--Factors Influencing Occupancy Rate

```
SELECT price, rating, number_of_reviews, availability_365
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
order by availability_365;
```

```
SELECT price, rating, number_of_reviews, availability_365
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
```

```
order by availability_365 desc;-- checking for the ones which are generally available
```

-- Percentage of lisings with license information

```
SELECT COUNT(*) AS total_listings, COUNT(license) AS licensed_listings, (COUNT(license) /
COUNT(*)) * 100 AS license_percentage
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`;
```

--listings with expired review

```
SELECT id, name, last_review
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
WHERE last_review < DATE_SUB(CURRENT_DATE(), INTERVAL 1 YEAR);
```

-- Compliance rate by Neighbourhood

```
SELECT neighbourhood, COUNT(id) AS total_listings, COUNT(license) AS licensed_listings,
(COUNT(license) / COUNT(*)) * 100 AS compliance_rate
FROM `directed-racer-396805.airbnb.airbnb_ny_listing`
GROUP BY neighbourhood
ORDER BY compliance_rate DESC;
```