

Airbnb - Exploratory Data Analysis on the City of London Listings

This project analyzes Airbnb host listings data using SQL and visualizations created in Tableau. The goal of this project was to discover insights about the different neighborhood listings from Hosts in the City of London.

Ask Phase

Objective

1. What can we learn about the different hosts and areas in London?
2. Understanding the different types of units listed.
3. Identify key insights on the neighborhoods in London that can be suitable ideas for the company.

Guiding Questions:

1. What can we learn from predictions? (ex. Locations, prices, reviews etc.)
2. Is there any noticeable difference among the areas and what could be the reason for it?
3. Which hosts are the busiest and why?
4. What can the different types of units available tell us?

Prepare Phase:

Data collection: The Datasets provided are from Airbnb data.

Datasets will be focused specifically on the city of London

Sources:

1. Airbnb Data:
<https://www.kaggle.com/datasets/whenamancodes/london-uk-airbnb-open-data>

Process Phase

Key Objectives: Data cleaning and manipulation with the datasets used in this case study.

Tools that will be used in this case study are:

- SQL on BigQuery for data wrangling and manipulation of the data, as well as any cleaning procedures.
- Data Visualization will be used on Tableau.
- Google Slides to present actionable insights to stakeholders.

Data Cleaning and Manipulation Deliverables: Using Sheets

1. Removed unwanted columns from the dataset:
 - a. "Neighborhood group" (empty cells) already have a neighborhood column
 - b. "license" (empty cells)
 - c. Removed "name" column (names of each individual home for rent)
 - d. Removed "ID" column (we already have host_id as a column with distinct values)
 - e. Removed "last review date" column
2. Filtered and removed for any missing values that contain empty cells for "Host" column
3. Trimmed whitespaces for all columns. (only "host" column needed to be adjusted)
4. Remove duplicates from the "Host_id" column (24,122 duplicates rows were found and removed)

SQL using BigQuery cleaning:

1. Removed NULL values under host_id column

Unset

```
DELETE
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
WHERE host_id IS NULL
```

2. Removed any units that had the price of \$0 listed for their Airbnb unit.

Unset

```
DELETE
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
WHERE price = 0
```

Checking new dataset for unique host IDs

Unset

```
SELECT COUNT(DISTINCT host_id) AS Total_ids  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
```

Results:

- Unique Host IDs: 4216


Analysis Phase:

Key Objective: Uncover insights within the dataset.

1. Checking for distinct total hosts and the different properties available

Unset

```
SELECT room_type, COUNT(DISTINCT host_id) AS total_hosts  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`  
GROUP BY room_type
```

See More:  room type per host id

24,681 of the hosts list under Entire Homes and/or apartments followed by Private rooms with 19993 hosts registered. However only 228 hosts have listings under hotels.

2. Checking the avg price of each listing type.

Unset

```
SELECT room_type, AVG(price) AS avg_price  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`  
GROUP BY room_type
```

See More:  avg price per room type

3. Total number of hosts listed by neighborhood

Unset

```
SELECT neighborhood, COUNT(host_id) AS total_hosts
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
ORDER BY total_hosts DESC
```

[See More:](#) [📊 Total number of hosts by neighborhood](#)

As we move closer to the center of London, the number of hosts generally increases, except for the City of London, which has only 215 registered hosts in that area. Westminster has the highest number of listings, followed by Hackney and Tower Hamlets.

4. Avg prices listed by neighborhood

Unset

```
SELECT neighborhood, avg(price) AS avg_price
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
ORDER BY avg_price DESC
```

[See More:](#) [📊 avg price by neighborhood](#)

5. Median price by neighborhood to get a more accurate analysis on what the listing prices look like.

Unset


```
SELECT neighborhood,
APPROX_QUANTILES(price, 2)[OFFSET(1)] AS median_price
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
```

[See More:](#) [📊 median price by neighborhood](#)

6. Max price listed available by the room type per neighborhood in London.

Unset

```
SELECT neighborhood, room_type,  
MAX(price) AS max_price  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`  
GROUP BY neighborhood, room_type
```


See More:  [max price by roomtype in each neighborhood](#)

Key Finding: Some listings were found at a max price \$25K a night whereas 31 of the 34 neighborhoods listed their properties at under \$100 a night. This distortion of the prices easily misconstrued the range in price given by the neighborhood for the average price.

7. Ranking of which neighborhoods and unit type is listed the most

Unset

```
SELECT neighborhood, room_type, COUNT(*) AS listings_count,  
ROW_NUMBER() OVER (PARTITION BY neighborhood ORDER BY COUNT(*)  
DESC) AS unit_type_rank  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`  
GROUP BY neighborhood, room_type
```

See More:  [unit ranked by bookings per neighborhood](#)

8. The top 5 neighborhoods listed, the unit type with median price

Unset

```
WITH popular_neighborhood AS (  
SELECT neighborhood, COUNT(*) AS listing_count,  
ROW_NUMBER() OVER (ORDER BY COUNT(*) DESC) AS popular_rank  
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`  
)
```

```


GROUP BY neighborhood
ORDER BY listing_count DESC
LIMIT 5
),

most_booked_room_type AS (
SELECT neighborhood, room_type, COUNT(*) AS bookings_count,
ROW_NUMBER() OVER (PARTITION BY neighborhood ORDER BY COUNT(*)
DESC) AS unit_type_rank
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood, room_type
),

median_price_per_neighborhood AS (
SELECT neighborhood, APPROX_QUANTILES(price, 2)[OFFSET(1)] AS
median_price
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
)

SELECT p.neighborhood AS popular_neighborhood,
b.room_type AS most_booked_room_type,
m.median_price AS median_price
FROM popular_neighborhood AS p
JOIN most_booked_room_type AS b
ON p.neighborhood = b.neighborhood
AND b.unit_type_rank = 1
JOIN median_price_per_neighborhood AS m
ON p.neighborhood = m.neighborhood;

```

See More:  Top 5 neighborhoods and unit type with median price

9. The top 5 neighborhoods that have the most total reviews, indicating the amount of people visiting these areas.

Unset

```
SELECT neighborhood, SUM(number_of_reviews) AS total_reviews
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
ORDER BY total_reviews DESC
LIMIT 5;
```

Westminster has by far the most reviews out of all the neighborhoods indicating that a majority of visitors visit that area. More analysis is needed to review exactly what the average star rating is in that area.

10. Top 10 neighborhoods reviewed by unit types

Unset

```
SELECT neighborhood, room_type,
SUM(number_of_reviews) AS total_reviews
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood, room_type
ORDER BY total_reviews DESC
LIMIT 10
```

[See More:](#) [+ Top 10 reviewed unit types by neighborhood](#)

This shows that people are leaving more reviews on houses/apts and private rooms in these neighborhoods.

11.

Unset

```
WITH lowest_availability AS (
SELECT neighborhood, MIN(availability_365) AS min_availability
FROM `airbnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
),
```

```

lowest_availability_counts AS (
SELECT l.neighborhood, COUNT(*) AS count_lowest_availability
FROM lowest_availability AS l
JOIN `airnbcasestudy-391303.LondonUK.airbnb_city_listing` AS a
ON l.neighborhood = a.neighborhood
AND l.min_availability = a.availability_365
GROUP BY l.neighborhood
)

SELECT neighborhood, count_lowest_availability
FROM lowest_availability_counts
ORDER BY count_lowest_availability DESC
LIMIT 5;

```

12. The top 5 neighborhoods with the lowest availability throughout the year.

📌 Top 5 neighborhoods with the lowest availability in the year

Hackney Neighborhood has the highest count of being least available throughout the year.

13. Total calculated host listings in each neighborhood

```

WITH top_neighborhoods AS (
SELECT neighborhood, COUNT(*) AS total_listings
FROM `airnbcasestudy-391303.LondonUK.airbnb_city_listing`
GROUP BY neighborhood
ORDER BY total_listings DESC
)

SELECT neighborhood, SUM(calculated_host_listings_count) AS
total_calculated_host_listings_count
FROM `airnbcasestudy-391303.LondonUK.airbnb_city_listing`
WHERE neighborhood IN (

```



```
SELECT neighborhood FROM top_neighborhoods
)
GROUP BY neighborhood
ORDER BY total_calculated_host_listings_count DESC
```

Key finding: Westminster has the highest count of listings at 82350.

The City of London only has 2562.

Westminster also has the median price of 140 a night whereas the city of london has the median price listed at 150 a night.

ACT PHASE:

Key findings:

- Westminster has the highest count of total listings at 82,350. Although a more popular destination, The City of London only has 2,562 listings.
- Westminster has the most hosts registered at 4,276, the most out of all the neighborhoods.
- The majority of hosts (24,681) have entire homes and/or apartments listed in London. Private rooms are made up of 19,993 hosts, with only 228 hosts listing hotels.
- As you move closer to the center of London, the number of hosts generally increases. Only 215 registered hosts have listings in that area. Westminster has the highest number of listings, followed by Hackney and Tower Hamlets.
- After reviewing the max and average prices in all neighborhoods, some listings were found to have a max price of \$25K/night; whereas 31 of the 34 neighborhoods listed their properties at under \$100 a night. Due to various price levels, the average price per listing was inconclusive..
- The median price had a stronger representation of how much the actual prices were.
- Hackney had the lowest availability listed out of all neighborhoods, with only 33 listings.
- **Westminster also has the median price of 140 a night, whereas the city of london has the median price listed at 150 a night.**

In order to draw stronger recommendations, more data is required. Including **revenue and occupancy** rates of each listing,

The “City of London” neighborhood can be a suitable area for someone to list new properties since there is an underwhelming amount of competition than the neighboring Westminster furthermore, the median listing price has a difference of only 10 a night between the two.