

## Table of Contents

1. Research Question .....	3
2. What is Airbnb .....	3
3. Data Acquisition .....	3
3.1. Understanding the dataset .....	4
3.2. Data preparation .....	4
4. Exploratory Data Analysis .....	6
5. Analyzing neighbourhoods .....	7
6. Analyzing property type .....	8
7. Analyzing price .....	10
8. Analysis seasonality.....	12
9. Analyzing reviews .....	13
10. Visualizations and encodings .....	14
10.1. Selection of Colours .....	14
10.2. Selection of the graphs .....	15
10.3. Selection of other visual encodings .....	17
11. Feedback .....	19
12. Conclusion .....	20

## 1. Research Question:

### What are the factors to consider before purchasing a property to rent on Airbnb?

- Where: Preferred area, property types and room types by the guests.
- Average price of the listing broken down by property type, room type and area as the host is expected to set their own price for their own listings.
- The correlation between rating, cleanliness, area, and price
- Seasonality – when is the best time to list a property and take it out for maintenance
- The demand and price of Airbnb rentals changing throughout the year and over time.

## 2. What is Airbnb?

Airbnb as in Air bed and breakfast is a global community marketplace with millions of unique accommodations in over 100,000 cities across 191 countries. In addition to accommodations Airbnb provides unique access to local communities and interests through Experiences, while Places allows users to discover recommendations from locals. Many tourists prefer to stay in a local setting with locals to experience the culture.

It gives a platform for people ('hosts') to rent out ('list') their properties at their convenience without having to go through complicated processes or make large investments. Hosts are expected to determine their own prices. Although Airbnb and other sites offer some broad advice, there isn't any free and accurate service that assist hosts in pricing their properties based on a variety of data elements.

### Airbnb statistics:

Active listings across the world	5.6 million
No. of guests used Airbnb to book accommodation	150 million
Gross revenue 2021 (Revenues for Airbnb + hosts)	\$48.9 billion

## 3. Data Acquisition:

Data source: The data is taken from the Inside Airbnb <http://insideairbnb.com/get-the-data/> which hosts publicly accessible data from the Airbnb website. The Inside Airbnb has gathered information on various cities and countries all over the world.

I chose the London dataset because I live in London and secondly, I want to analyze and visualize one of the world's most popular tourist destinations.

### 3.1 Understanding our dataset

The dataset comprises of three main tables.

**Listings** — Detailed information about each listing, including 74 attributes. However, we only need 14 attributes for our analysis and visualization

Id – listing ID of the property, integer unique for each listing host\_id

– host id, integer unique for each host

host\_since – the date a person first became a host and started listing their properties

neighbourhood\_cleaned – Cleaned neighborhood region of each listing

latitude and longitude – latitude and longitude of the listings

property\_type\_cleaned – Type of the property [apartment, aparthotel, bed and breakfast, Condo, Guesthouse loft residential home, townhouse]

room\_type – Type of the room of the listing [Entire home/apt, Private room, Shared room]

price – Price of the accommodation per night minimum\_nights – Minimum number of nights the guest must book maximum\_nights - Maximum number of nights the guest can book

number\_of\_reviews – Total number of reviews by the guest

review\_scores\_rating – Rating given by the guest who have booked the accommodation in the past.

review\_scores\_cleanliness – Rating for cleanliness by the guest review\_scores\_location – Rating for the location by the guest

**Reviews** – Details of the review given by the guests. Key attributes are:

listing\_id – listing id of each listing date

– date of the review

comment – reviews by the guests

**Neighborhood data** – It has longitudes and latitudes of each neighborhood in London.

Our dataset contains 66,641 records (this includes all the data from London including null values and might require pre-processing).

### 3.2 Data preparation:

Initially, the data was unorganized and unclear. There were many unwanted attributes and non-uniformed data. So, I used excel to clean the data and make it ready for visualization. First step was to delete all the unwanted column and keep only the above mentioned columns.

#### Missing values:

In the price column, there were rows with zero listing price, which is realistically not true. No host will list their property at a cost. I removed those records.

There are a few missing values in reviews, but this can be true, we can have a listing with zero reviews. There are also a few missing values in host\_since column, I have replaced the null with their first review date, as it is difficult to guess the correct date.

### Outliers:

There were also few outliers in the price column, for example there was one listing for 1000 pounds per night for a shared room and 5 pounds per night for the entire apartment, which is practically not correct. It is extremely unlikely that a guest will book a shared room for 1000 pounds per night, and it is also extremely unlikely that a host will rent their apartment for 5 pounds a night. I have used Inter quartile range (IQR) method to remove the outliers in excel for each property type.

Before removing outliers:

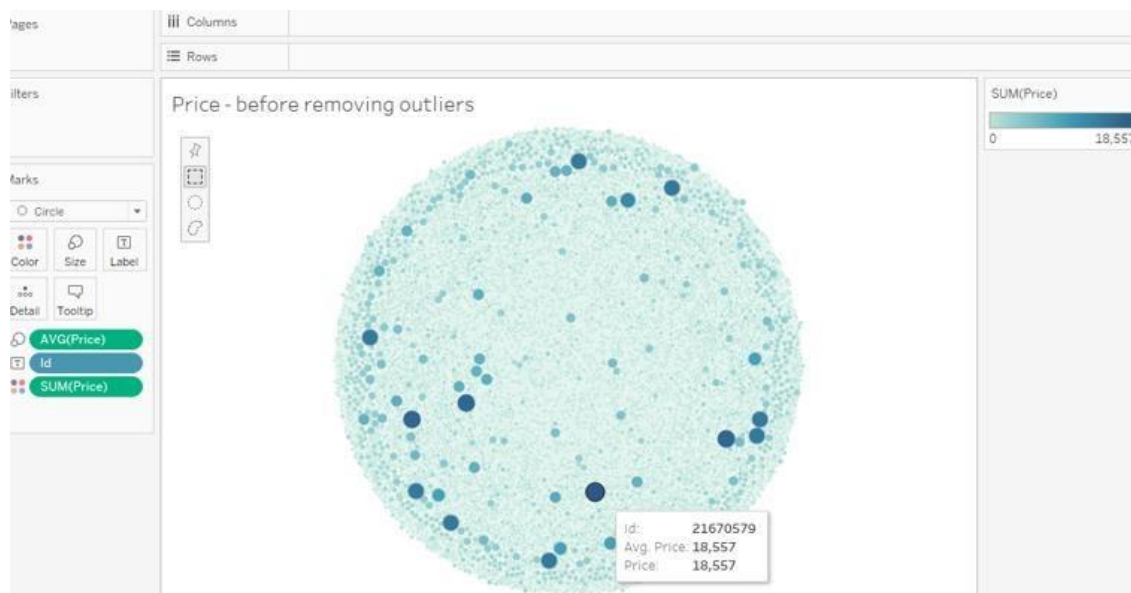


Figure 1 – Price data with outliers After removing outliers:

price column - after removing outliers

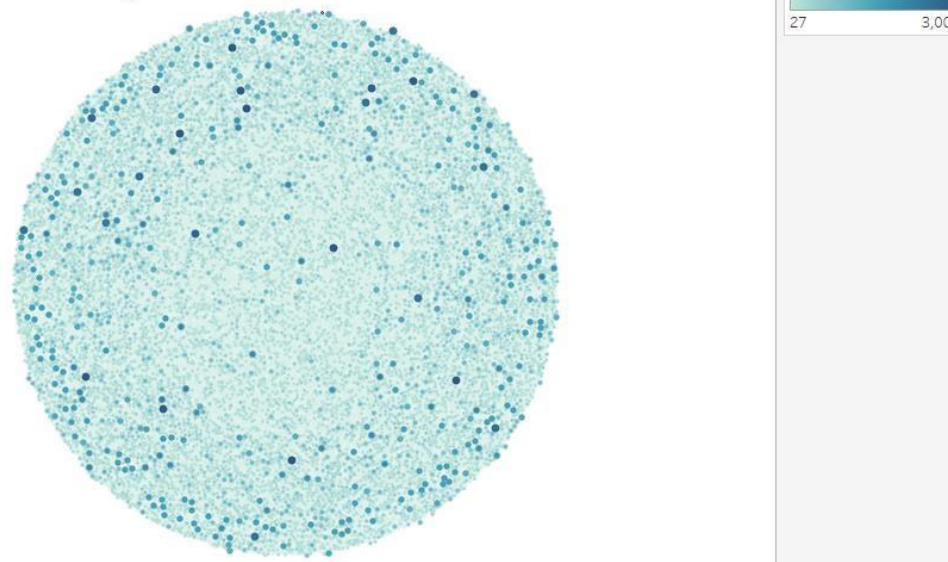


Figure 2 – Price data without outliers

### Cleaning textual column:

There were some names which were same but were written differently, some of them spelt incorrectly or were in all lower case, all upper case etc. Some property types were 'Entire townhouse', 'townhouse' , 'Private room in townhouse' – changed these to 'townhouse' because 'entire' ,private room' or 'shared room' is described in room type.

### 4. Exploratory Data Analysis:

London Airbnb has total 59,680 listings as of September 2021.

Total no. of Listings

59,680

Average Price

£136

Maximun Price

£3,000

Minimun Price

£27

Growth of Airbnb London since 2008:

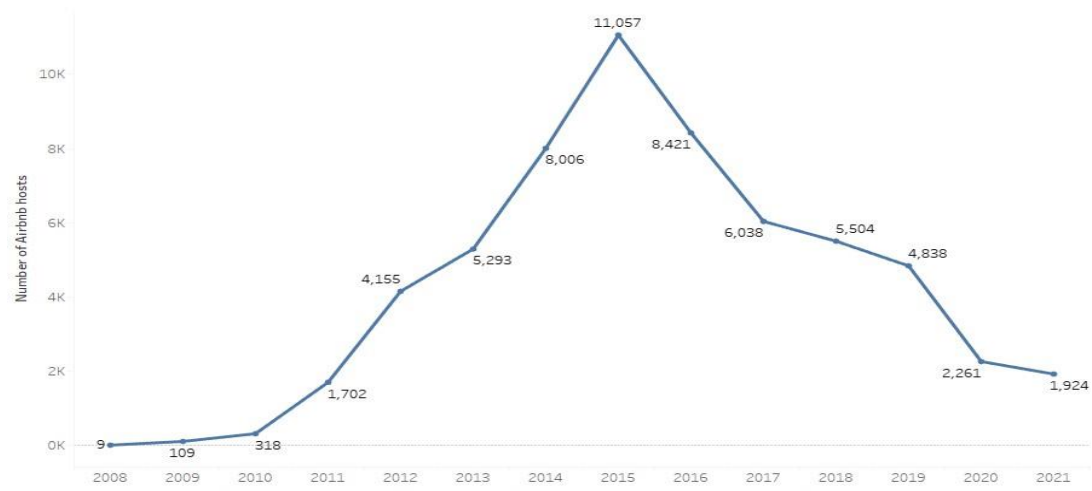


Figure 3 – Airbnb over the years

In the year 2008 there were only a couple of Airbnb hosts, and it became very popular over the years, with highest number of hosts in 2015. However, there was a huge drop in the number of hosts in 2020 due to COVID.

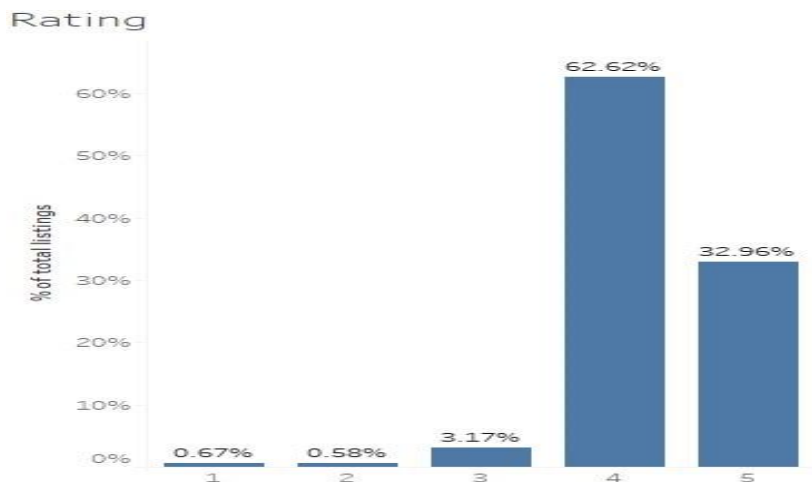


Figure 4 – Rating of Airbnb listings

More than 95% of the postings have a rating of 4 or above, with only 2% having a rating of 2 or below. This indicates that guests are extremely pleased with the type of lodging, pricing, and service provided by Airbnb.

Looking at the popularity of Airbnb, it seems to be a good idea to list properties on Airbnb. Let us now analysis the density of listings by area and popular neighbourhoods.

## 5. Analyzing neighbourhoods:

Number of listings by region

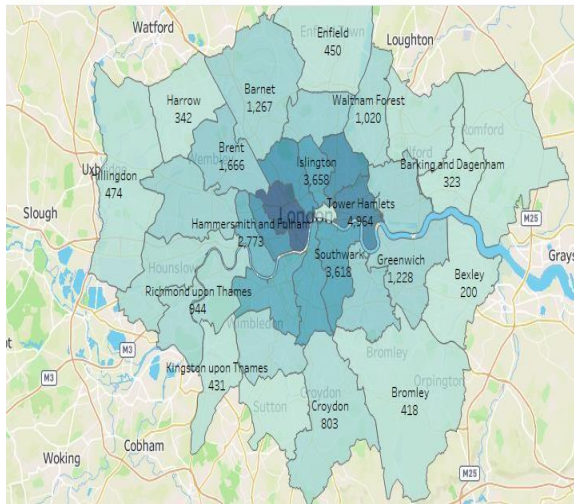


Figure 5- Listing distribution by area

Number of listings by Neighbourhood

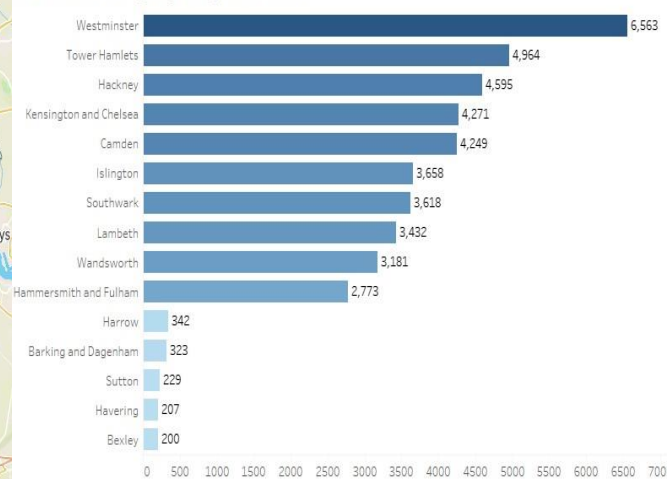


Figure 6 – Popular Neighborhood

Listings are more concentrated in the central part of London and become less dense as we move further out. The reason being their proximity to local attractions and transport links.

Westminster has the highest number of listings(6,563) followed by Tower Hamlets (4,964). Westminster is famous for its historical buildings such as Buckingham Palace, Westminster Abbey and Big Ben. Tower Hamlets is known for its famous tourist attractions such as Tower Bridge, Tower of London and the HMS Belfast.

## 6. Analyzing property types:

After analyzing the neighborhood, let us now check what type of property and rooms are preferred by the guests.

Predominantly, apartment and condo have the highest number of listings among all the property types and the bed and breakfast type has the least number of listings.

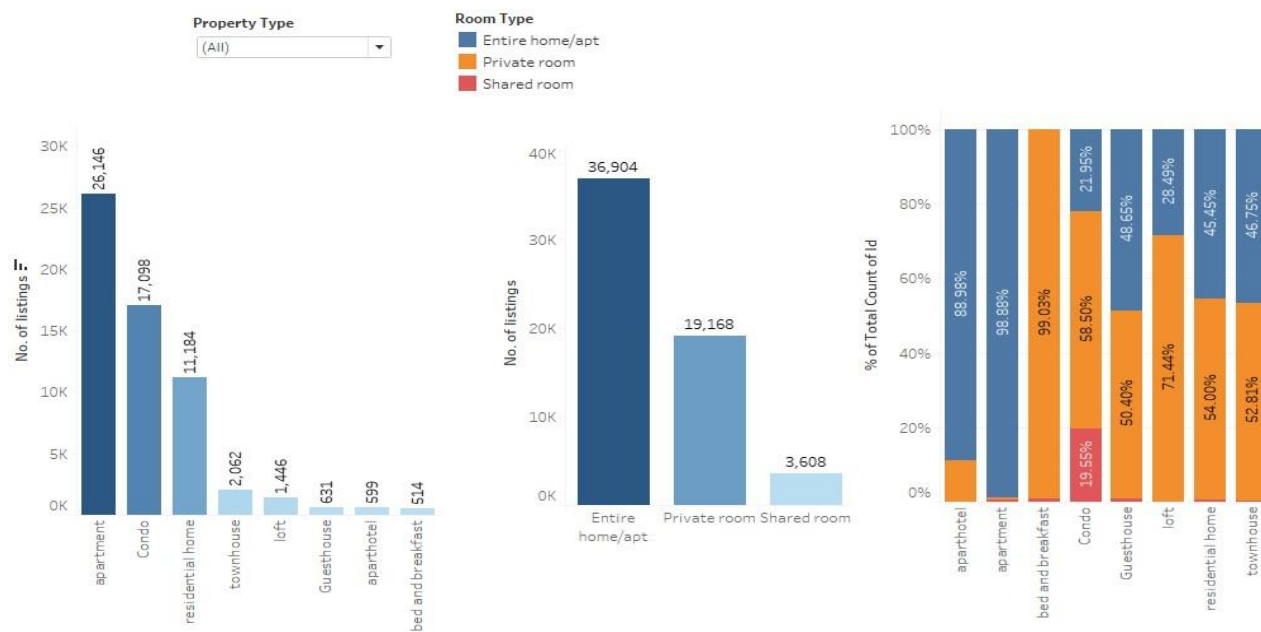


Figure 7- Property analysis  
Demand by room type

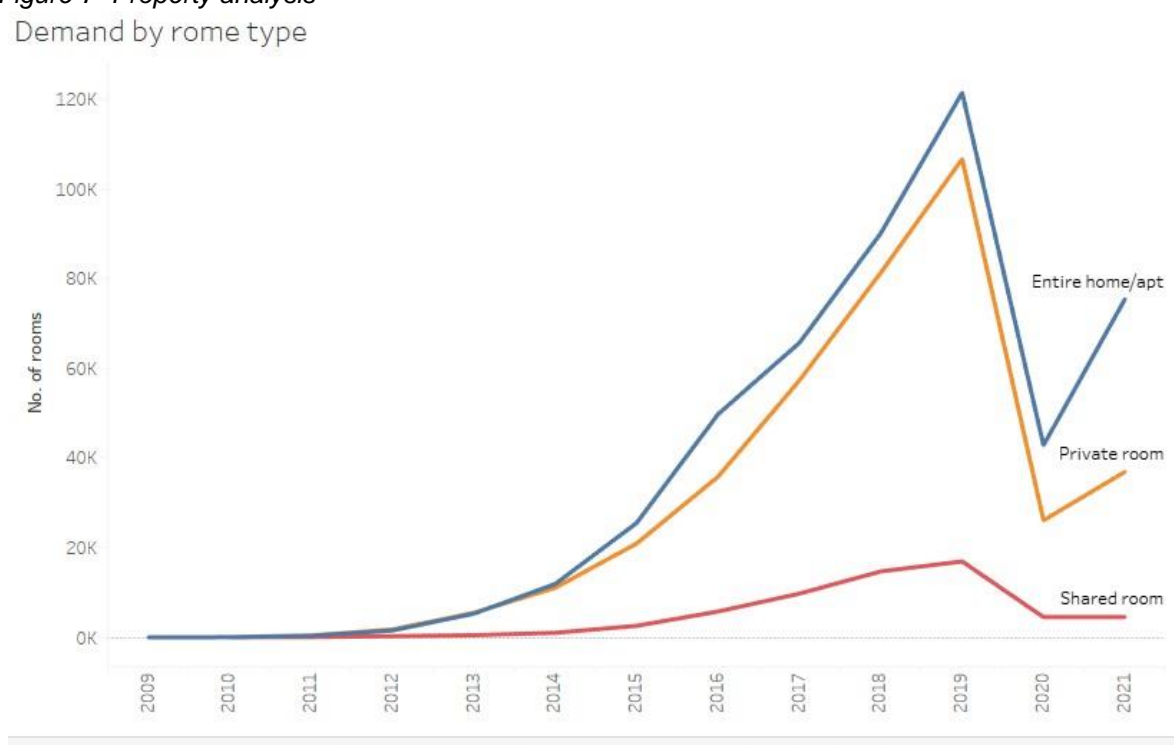


Figure 8- Demand of room types

Entire home/apt has the maximum number of bookings whereas, shared room type has the least number of bookings.

Most of the guests like to book an apartment or condo and very few like to book a shared room, which make sense as a couple or family would book a private room or the entire home, while solo travelers are likely to book a shared room.



About 11,184 hosts list their residential home and, of that 11,184, about 54% of them rent a room in their home. The tourists who like to stay with the locals to have a local experience must like to book this type of accommodation.

For apartments, about 99% of the listings are entire home and for bed and breakfast 99.03 are a private room, looks like host might list a private room where they stay and make breakfast for the guest.

Now we know which areas are suitable for buying a place to rent out on Airbnb, as well as the property types and rooms. Let us now analyze at what price one should list their property based on the location, property type and room type as on Airbnb host must set a price on their listings.

## 7. Analyzing price:

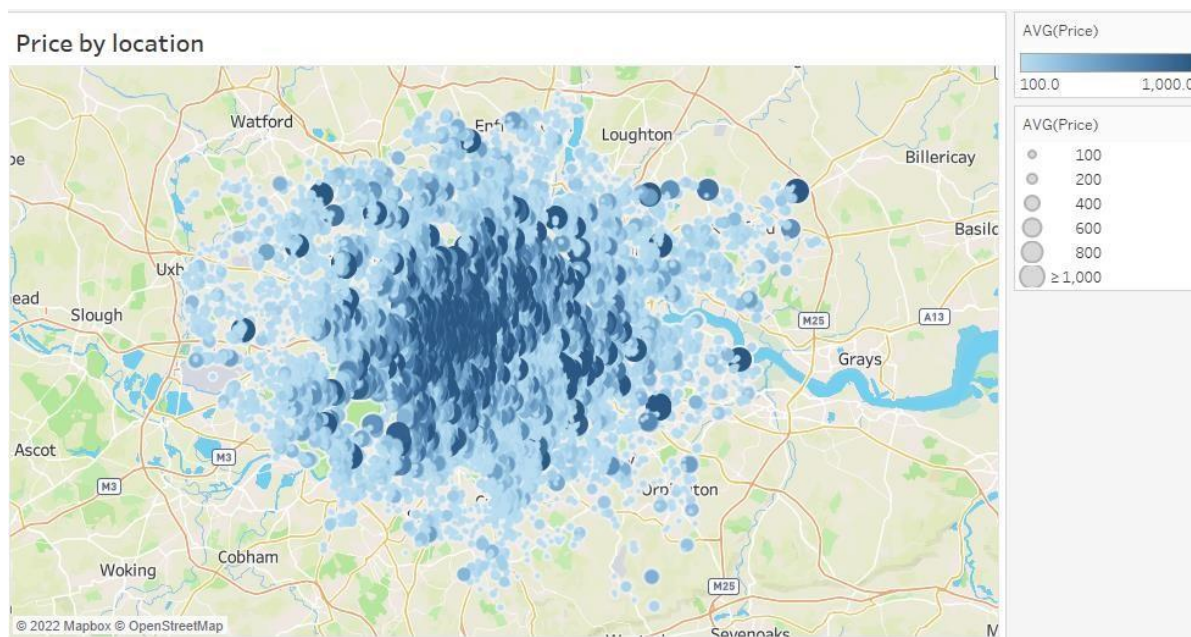


Figure 9 – Variation in price by area

## Price by Neighbourhood

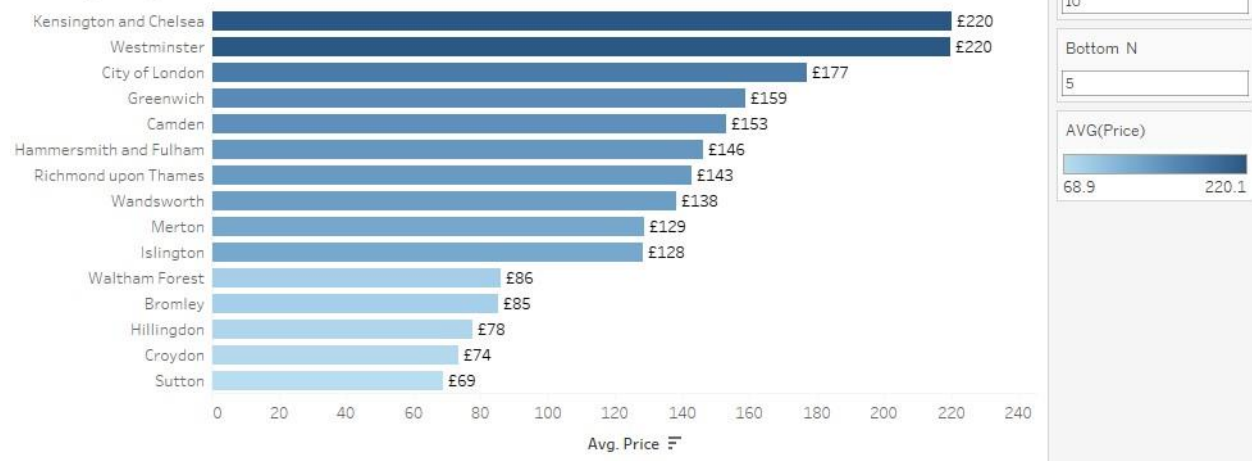


Figure 10 – Variation in price by Neighborhood

Kensington & Chelsea and Westminster have the highest average price of £220, this could be due to the high demand, followed by City of London and Greenwich.

## Price by Property type/Room type

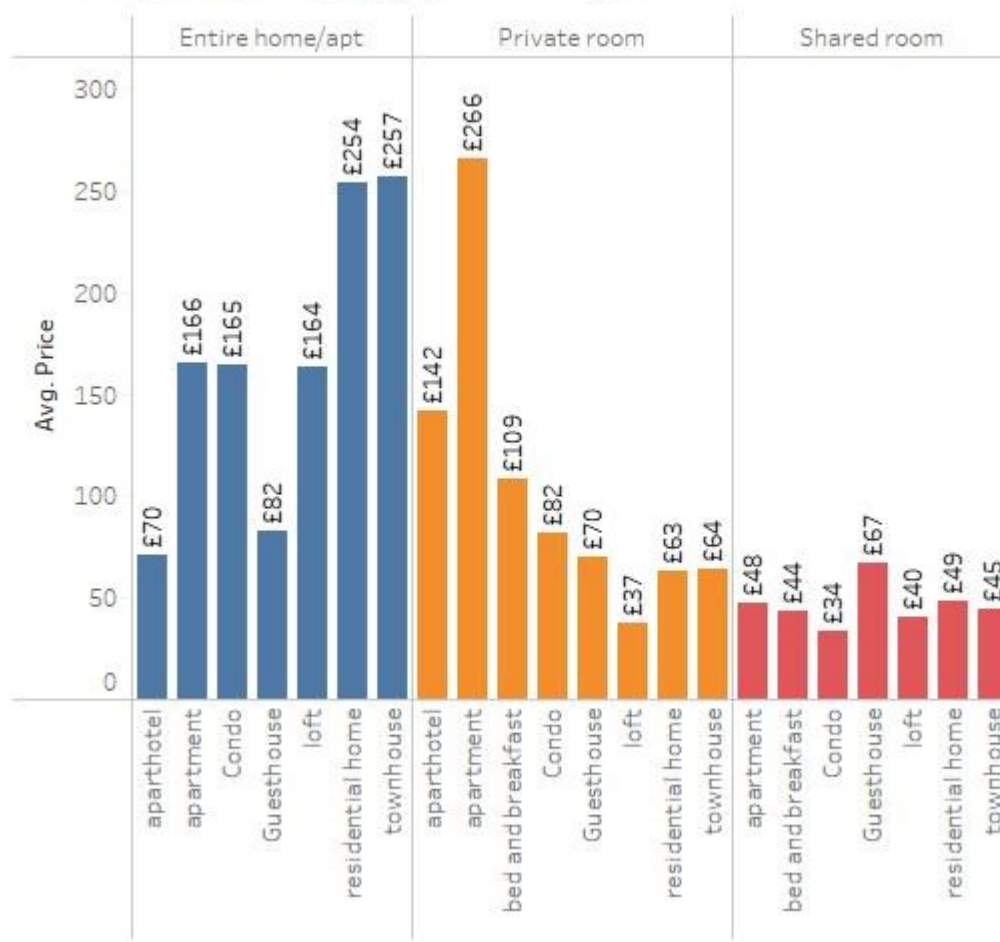


Figure 11- Variation in price by property type/room type

Average price of the entire home is greater than a private room, which is understandable. However, the average price of the entire apartment less than the private room in an apartment.

The host can make more money if they rent the rooms in their apartment as private rooms than rent the entire apartment.

From the graph above we can determine the average price we should list our property based on the type of property and rooms we have and in which neighborhood.

Now we have analyzed the factors hosts should consider if they want to rent a property on Airbnb. Popular neighbourhoods – preferably Westminster, Camden, Kensington and Chelsea or City of London as these areas famous locations and will always be filled with tourists and the type of the properties/room.

We have also determined the average price the host should list their properties based on the property types and room types also considering the location.

## 8. Analysis seasonality:

It is important to know when the demand is high and low so the hosts can rent their properties and take it out for maintenance accordingly.

Airbnb over the time

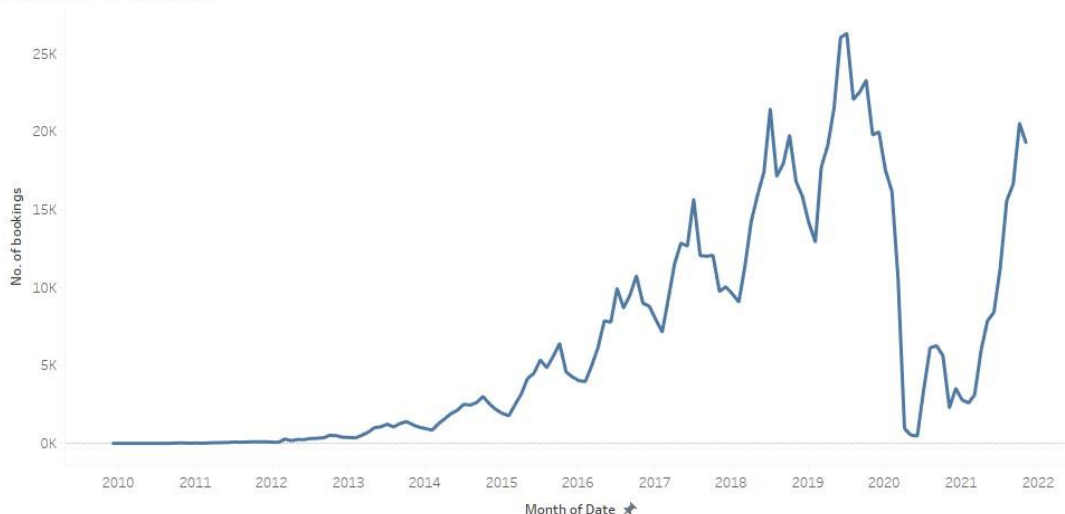


Figure 12 – Seasonality in demand

From the line graph above it is clear that the demand exhibits seasonality and we can also see the drop in the demand in 2020 and 2021 due to covid and has rapidly increased by end of 2021, as soon as the travel restrictions were lifted.

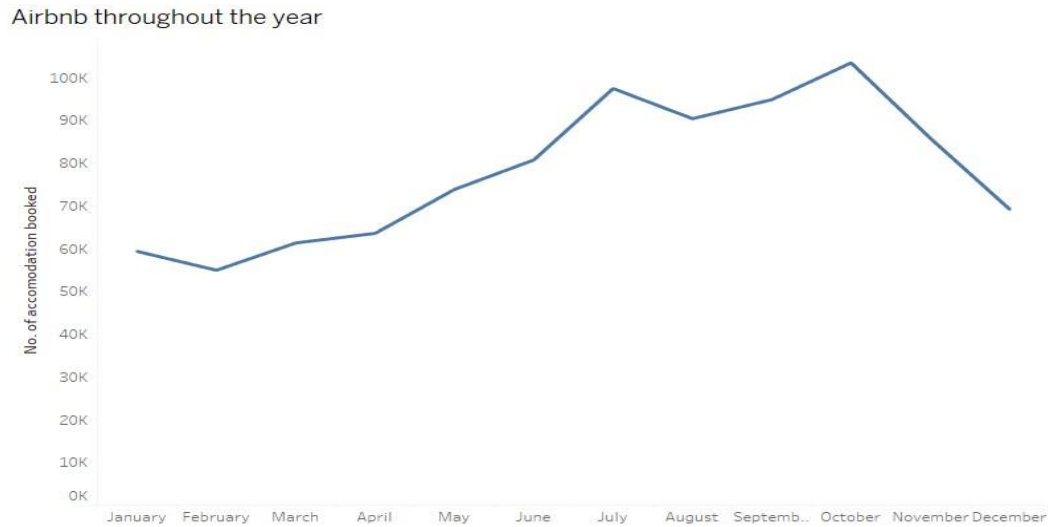


Figure 13 - Demand of Airbnb over the year

Late spring and summer are peak tourist seasons, which is reflected in the graph above. This would be the best time to list the property. The demand is low during winter December through February, this would be the best time to take out the property from maintenance.

## 9. Analyzing reviews:

When we book a service online, reviews play a very important role.



Figure 14- Rating vs Price

There is a positive correlation between the rating and average price, it means higher the rating higher the price.

Number of reviews vs price

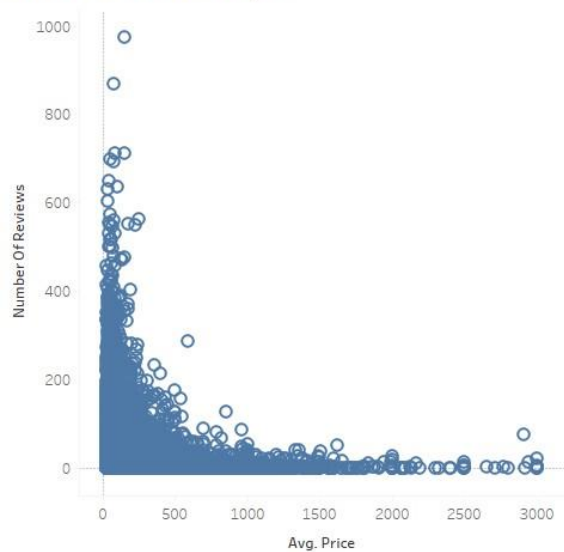


Figure 15- No. of Reviews vs Price

The majority of individuals prefer affordable lodging options, hence they have a higher number of reviews, while a small percentage of people will book expensive accommodations, and consequently have a lower number of ratings.

We can also suggest that most elitists and wealthy people will book expensive accommodations and are less inclined to leave reviews on public forums, which could explain why high-priced listings get fewer reviews.

Cleanliness vs price

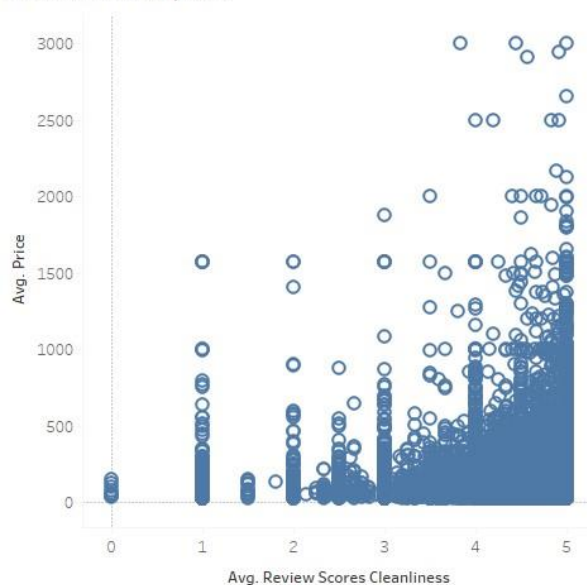


Figure 16- Cleanliness vs Price

Cleanliness affects the price, so to be able to achieve a higher rent, it is important for the host to keep the property clean.

So, if the property has positive reviews, it is likely to attract more guests and have a higher the rating, resulting in a higher rent.

## 10. Visualizations and encodings:

### 10.1 Selection of Colours:

Initially, I used the Airbnb company Colour (red and black) but didn't look visually appealing. Also, red and black are mostly associated with negative/bad/loss.

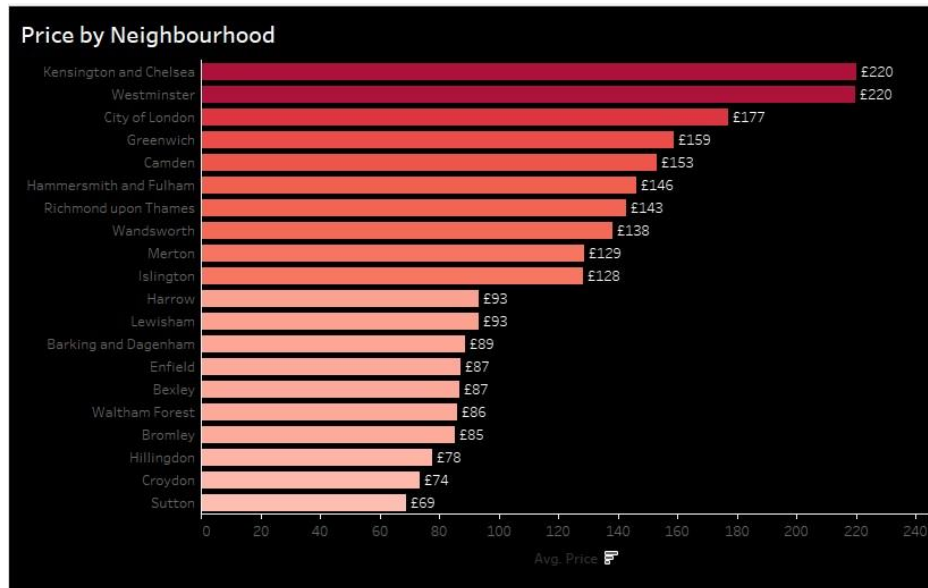


Figure 17 - Airbnb - logo colour

So, I changed the Colour to blue for all my visualization.

Colour is not naturally ordered in our brain. However, brightness and intensity are, it's why I have used different shades to show the variation in price and counts.

Using two divergent Colours is confusing and not very helpful in our case, the reader might mistake it for different measures. Hence, I decided to use different shades of the single Colour; Dark blue being the highest and light blue being the least

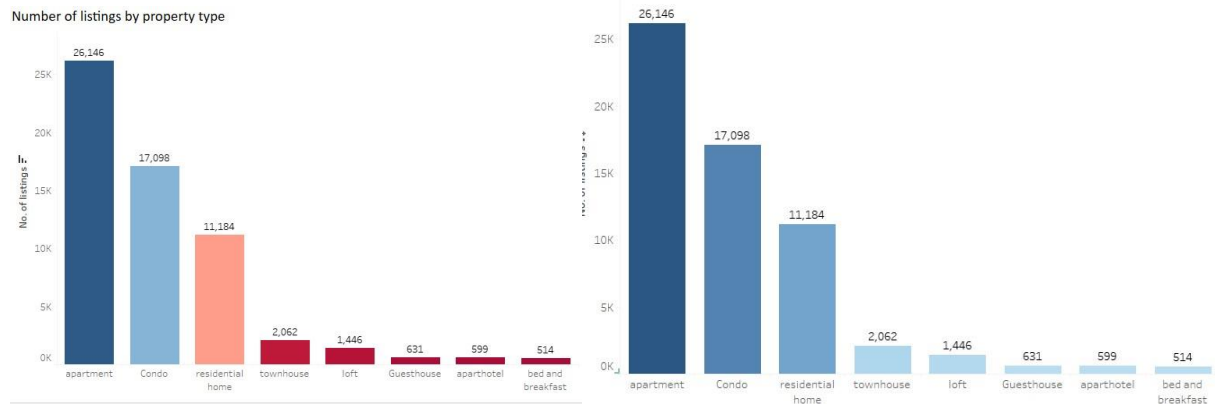


Figure 18 – Divergent vs single Colour

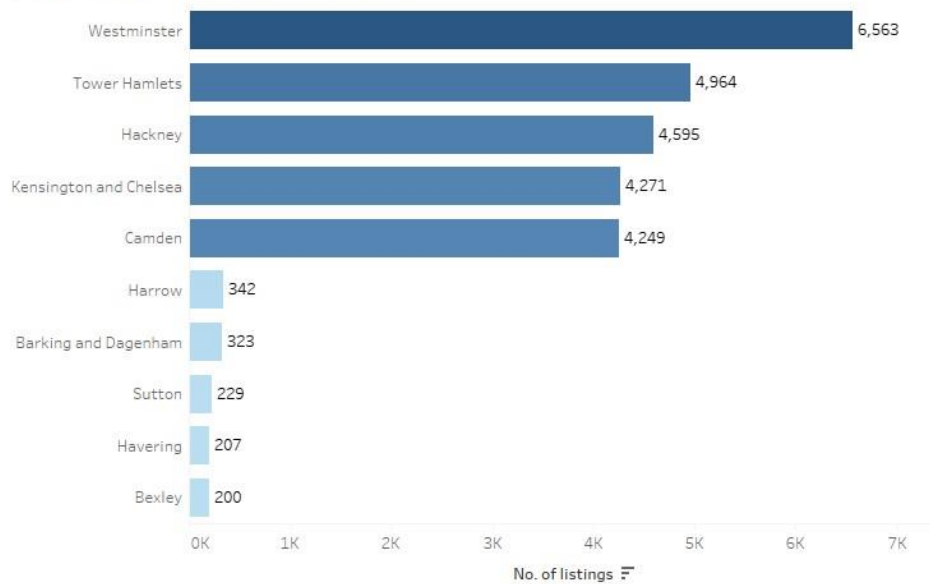
## 10.2 Selection of the graphs:

### Bar Graphs

I have use bar graphs to compare different groups (Neighborhood, property type, room type) as bar graphs are best for representing comparative data, as bar graphs are easy to interpret. These graphs are good for displaying nominal and ordinal data.

For comparing the variables having more than 4 groups like neighbourhoods, I have used horizontal bar graphs as the neighborhood names would be displayed horizontally and our brain is naturally trained to read horizontally from left to right – in the Western World.

Number of listings by Neighbourhood



Number of listings by Neighbourhood

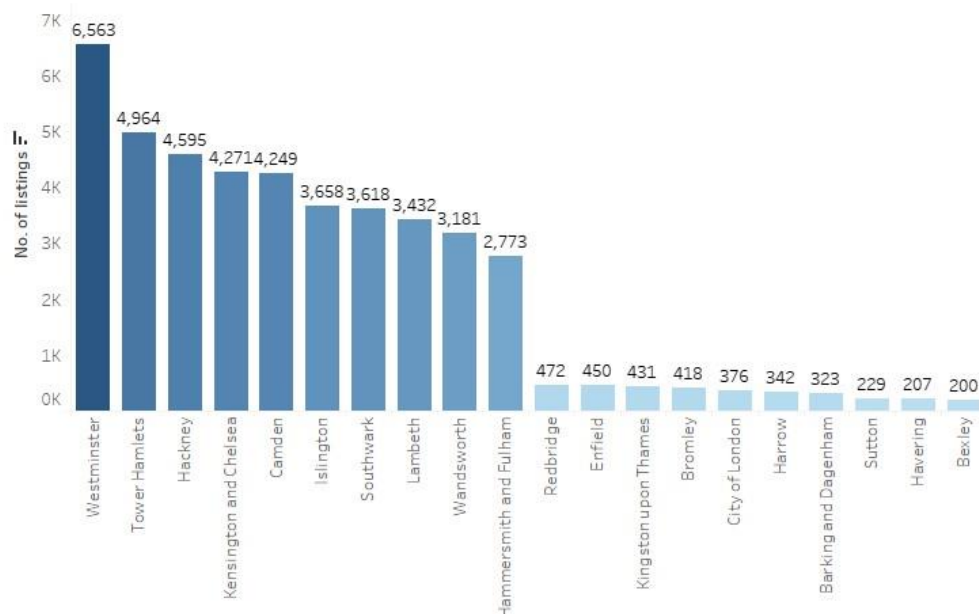


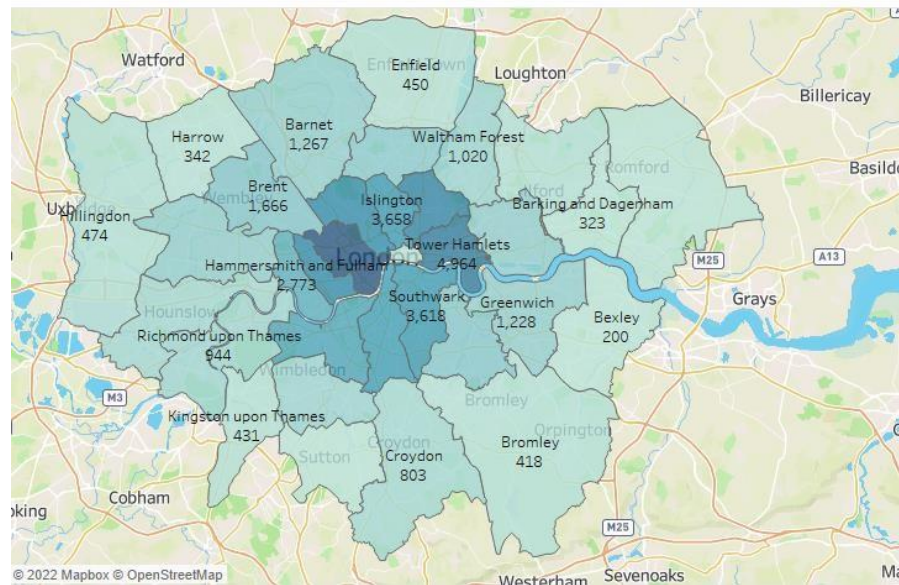
Figure 19 – Horizontal bar graph vs vertical bar graph

## Maps:

I have used the map to visualize the distribution of the number of listings and price by area as it is easy to interpret.



Number of listings by region



Number of listings by region

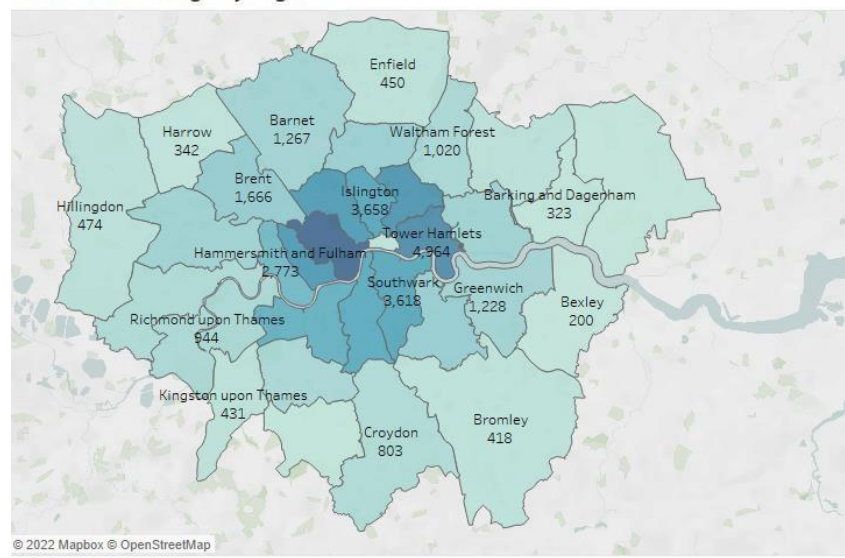


Figure 20 – Map 1 background ‘outdoors’ vs ‘light’

I have used the map background ‘outdoors’ in Map 1. It shows the important features such as water bodies, park, airport, train stations and many more (from Map1 we can easily recognize the river Thames and can visualize the exact area the map is representing). Whereas in light, normal or dark those features are not highlighted

### Pie chart:

It is easier to visualize the distribution of room types from the pie chart when compared to bar graph.

Number of listings by room type

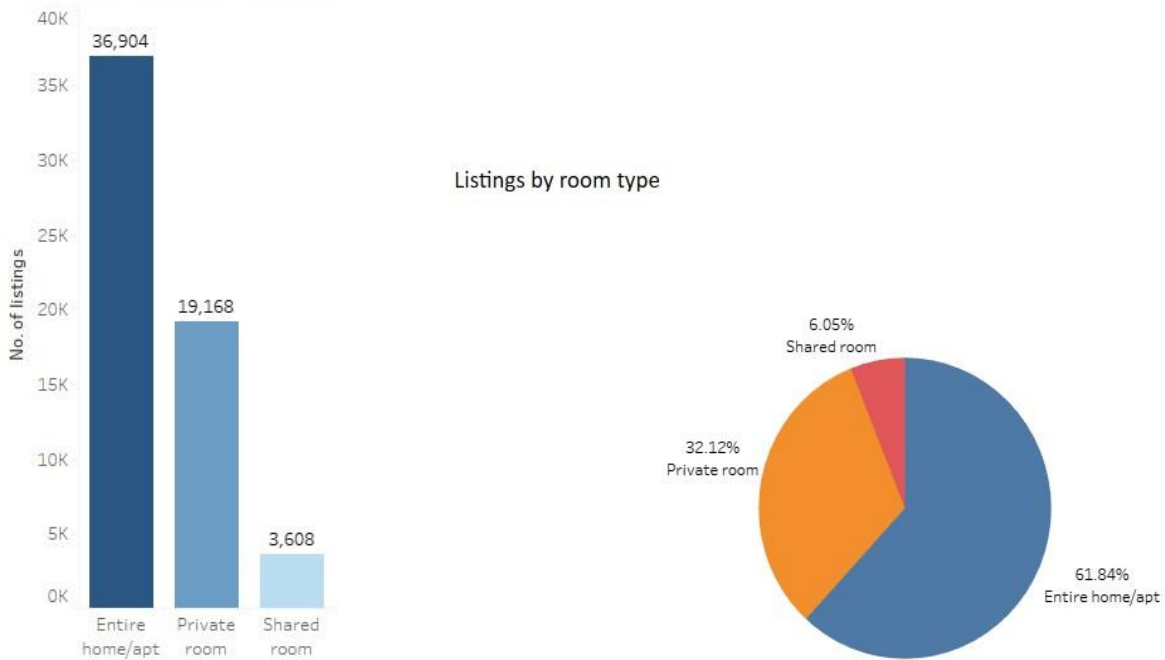


Figure 21 – Bar graph vs pie chart

### Line graphs:

Line graphs are useful when representing changes over time.

Bar graphs are best when changes over time are large. But in our case changes in demand over the months are not huge and it is easier to see seasonality in a line graph than a bar graph

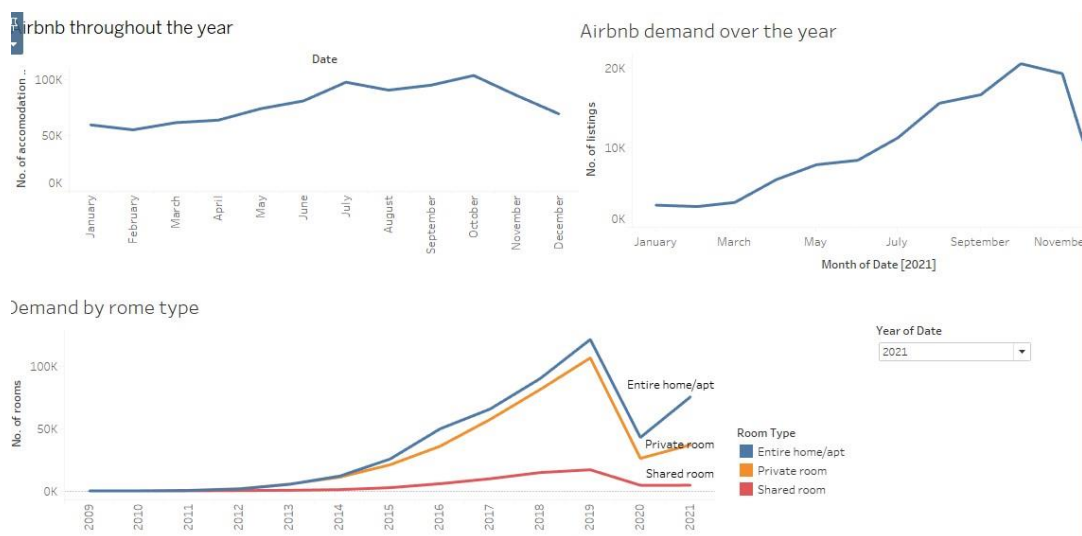


Figure 22 – Line graph – to show changes over time Scatter plot:

We use scatter plots to check the correlation between two measures. Because the reader can easily see if there is any correlation between the variables, if there is then they are positively correlated or negatively correlated.

I have used the scatter plot to check the correlation between the price and the reviews.

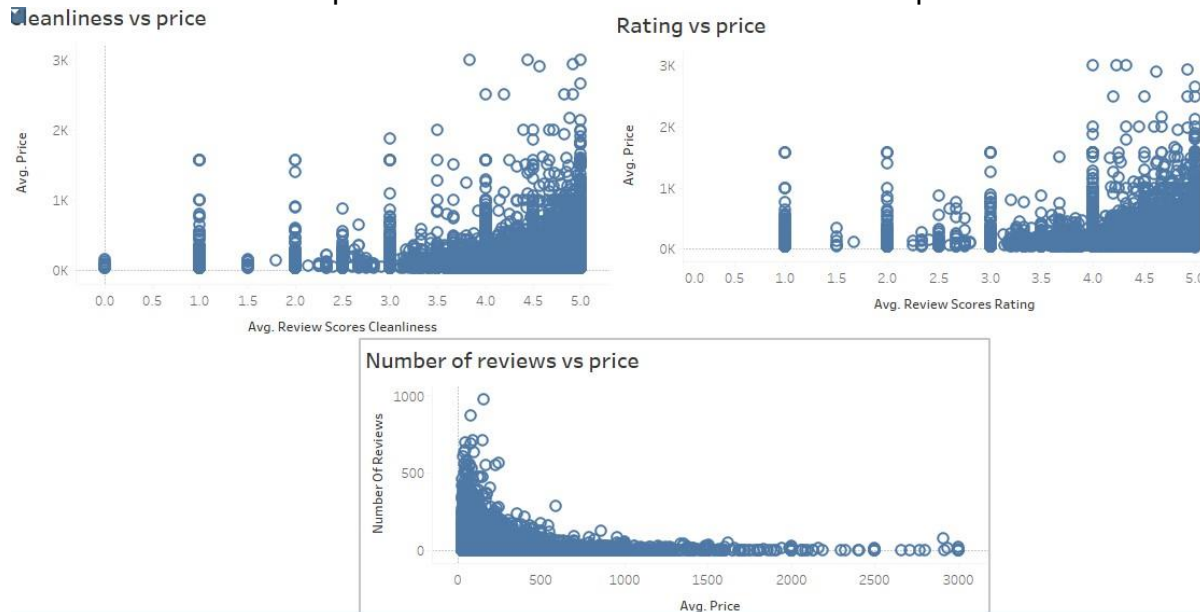


Figure 23 – Scatter plot – correlation analysis

## 11.Feedback:

I showed my work to my partner. He is a Data engineer and is into properties. He understands the important of data analysis and the factors one should consider before investing in properties. The feedback I received for my visualization were:

Visualization was self-explanatory and easy to understand, and the choice of the graphs and Colours are excellent.

For vis X – initially, I visualized all the groups in the neighborhood in a single visualization and for that the feedback was to improve the visualization, it had too many groups and the graph was over-crowded which made it difficult to read and compare.

To address the issue, I added a filter where user can select top N and bottom N groups of their interest.

Another feedback was to improve the visualization of the room type by property type.

Initially, I analyzed the demand of property types and room types separately. Feedback was to create a visualization in such a way that the user should be able to compare and see the demands of each room type across the property types

To address the issue, I created the visualization below, showing different room types by the property type. This would help the host to make decision to list room as private rooms or entire home according to the property type.

Property type by room type

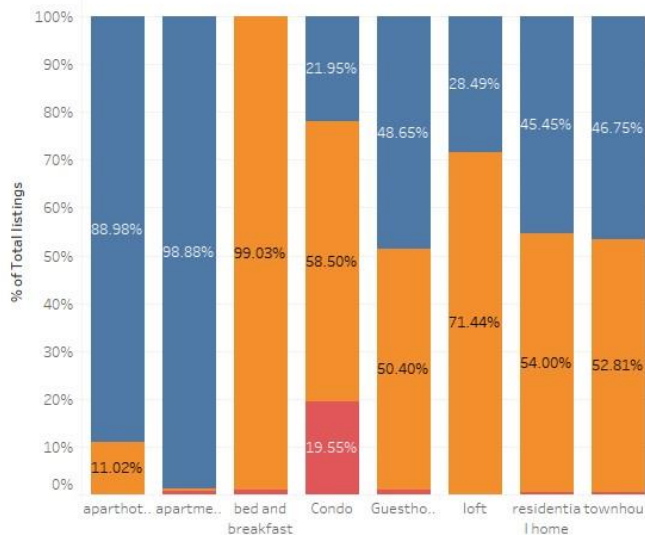


Figure 24 – Demand of the room types by property types

## 12. Conclusion:

The Airbnb has become popular over the years. However, tourism industry was hard hit by the covid in the year 2020 and the recovery began in late 2021 at great speed as soon as the travel restriction were lifted.

About 95% of the listing in London have a rating of 4 and above. This indicates that guests are extremely pleased with the type of lodging, pricing, and service provided by Airbnb. One of the huge advantages of listing property on Airbnb is the host can rent the properties at their convenience without having to go through complicated processes or make large investments. Looking at the popularity of Airbnb, it seems to be a good idea to invest in a property and list it on Airbnb.

However, you should consider a few factors when buying a property to rent on Airbnb and they are:

**WHERE(location)** – Where is the best place to list a property and what are the popular neighbourhoods. From our analysis, Westminster has the highest number of listings followed by Tower Hamlets whereas Bexley and Havering have the lowest number of listings and are least popular. Hence, one should avoid investing in properties in such regions.

**WHAT(what property types)** - What types of properties are in demand and what are the room types preferred by the guests. From the analysis above, the demand and average price for the entire home/apt is more than private and shared room. However, the average price of the entire apartment is less than the private room in an apartment.

**WHEN (Best time)** -When is the best time to list a property and take out for maintenance. From our analysis, demands for the accommodation is high during the late spring and

summer as these are the peak tourist seasons. However, the demand for accommodation starts increasing from March through May when the weather is moderate. increasing from March through May.

And as the demand is low during winter December through February, this would be the best time to take out the property from maintenance

HOW(How to increase booking) – Positive reviews and higher rating will attract more guests. When booking a service online, reviews play an important role. People look through the reviews and rating given by the guest who stayed in the past and if the property has positive reviews, it is likely to attract people to book the accommodation and if they have pleasant experience, then they are likely to give positive review and high rating which will indeed attract more guests, hence more booking. So, it is important to maintain positive reviews and high rating