Analysis of Airbnb Listings in Bristol - UK

Rishabh Mishra  
 *Department of Data Science  
Swinburne University of Technology*Melbourne, Australia  
104101198@student.swin.edu.au

*Abstract*— This research paper examines Airbnb property listings in Bristol city of UK. The paper explains and answers different questions intended for hosts, guests, and organizations using visualizations. I have discussed and validated my questions by elaborating and referencing the work done by other researchers and experts as well. The research paper also reflects the light on limitations of analyzed visualizations and what can be done in the future to overcome those limitations.

# Introduction

# Airbnb is an online marketplace application that provides homestay experiences to travelers around the world. Being a community of more than 4 million hosts and recording more than 1 billion guest arrivals, it is one of the biggest companies in the hospitality and tourism industry. The growth of Airbnb has been remarkable since its foundation in 2007. It has given many people a chance of earning and making money with the property and homes they have. With this much large audience, as a key decision-maker in the company, it is also important to analyze and foresee the past and future trends of how Airbnb has performed. This is why I chose this dataset to show my visualizations. With more than 250 thousand active Airbnb listings, the UK is one of its largest markets. In this research report, I will analyze and visualize Airbnb listings in UK Bristol city.

As I reflected above, Airbnb is such a large community that it becomes very important to take good decisions as a host in order to earn maximum profits. Moreover, as a guest, it is also important to see the price differences in listings in order to get the best deal. Additionally, as a data analyst for the company (Airbnb in this case), it becomes important to find out important factors which affect the business of the company.

So, in this research paper, my aim will be to analyze questions such as:

* Area-wise Listing Frequency in UK’s Bristol city (Listing Frequency) – The intended audience is hosts as they can see and decide, where to list the property and what type of property to list by looking at patterns.
* Price differences in Neighbourhoods(Comparative Analysis) – The intended audience is guests so they can get information about price differences of listings neighborhoods.
* Host Acquisition Rate in Bristol – The intended audience is the company’s data analysts and business-making partners so that they can know the performance of their organization in a particular region.

In this report, I will be discussing and explaining my work in the following order

* How data is created and explanation of data fields
* Visual Analysis performed during initial skimming of the data set
* Work done by other experts to support my research
* Data Processing methods used in the dataset that I have used
* Explanation of my visualizations
* Brief of what my research communicated and future improvements in the work

# Theoretical Foundation

In this section, I will be discussing models and concepts that are put in by some experts that are important to understand this article.

## Domain of the Data

The hospitality and Tourism industry has seen its peak in the last few years, and it is expected to grow even more in the years to come. According to [1], travelers use Airbnb for four main reasons: price, habit, ease of use, and hedonic motivation of experiencing authenticity. One of the key reasons for Airbnb’s success is its focus on relationship building between hosts and guests hence making reviews a critical part of the Airbnb application. Apart from this, Airbnb focuses on making individuals feel that they are ‘travelers’ not ‘tourists’ because it is often seen that word ‘tourists’ make a negative impact on the locals. Back in the day, during the initial days of Airbnb, most of the tourism websites did not have a dedicated image gallery and used to focus mainly on details of amenities but Airbnb changed the perception by showing cleanliness, and housekeeping services and introducing quality image gallery in their application providing a sense of customer satisfaction which leads to extended stays on a particular stay. Introducing licenses, and security policies, and making a host’s details public further grabbed the guest’s attention who was skeptical and paranoid about the safety of this feature introduced the concept of trust dealing on the internet.

Looking into the dataset, it can be clearly seen why columns such as neighborhood, licenses, number of nights, and host details play as an important role in the analysis of Airbnb. With these factors, an individual can analyze and foresee what the guest is looking for and what the hosts can provide in their property to make their listings loved and top-rated.

## Visual Analysis

# For an interactive and effective visualization, data analysis is a very important step. It makes an individual familiar with the type of data and the importance of a particular field in the dataset.

My dataset includes the latitude and longitude of the listings. These values can be used to find the listings by neighborhood. Also, these fields play an important role when visualizing the information on a map. The data collection also includes details of hosts such as the hostname and host id. These fields can be helpful when visualizing the frequency of listings by hosts. There is a price column in the collection which gives us information about the price of the listing and it’s obvious that the price field helps in visualizing many answers as far as guests are concerned. There is also a categorial value in the dataset named room type. This field has 4 values - Entire Home/Apartment, Hotel Room, Private Room, and Shared Room which provide information about the type of listings. The price in the dataset is in ‘pounds and the price is per night for a listing. Further, the data collection also includes the number of reviews columns, this can vaguely give us insight into the number of visits to a listing. Data Source also has a field named ‘last review’ which contains the date of the last review on a listing.

For this work, I have proposed the following techniques to answer my questions

* Comparative Analysis - Use of LOD Expressions
* Acquisition Rate - Line Chart
* Listing Frequency - Symbol Map

# Related Literature

To understand the trends of Airbnb listings, prices, and many other factors that affect the business of hosts and companies, much research has been conducted earlier. In this section, I will be discussing some of the related work done.

In [2], the author did thorough research answering a very important question - what are the locations where listings fetch higher prices? For this analysis, the author used the Airbnb listing data of ‘Seattle’ city. The author answered the question by analyzing information through 2 graphs. Use of a symbol map to analyze the number of listings of each room type in a neighborhood. Through this visualization, the author clearly shows that most of the listings are in central Seattle. To differentiate between the room type in listings, the author has used the color encoding principle to encode the room type column which gives insight into what type of room is listed in which neighborhood. Also, he used a bar chart to show and analyze the number of listings in each neighborhood and its price’. By analyzing the number of listings and prices of each neighborhood, it can be seen which areas/neighborhood in ‘Seattle’ city has expensive listings.

Screenshots of the visualization that the author has performed are attached below (Fig 1 and Fig 2). I will also reference the original article by the author in the references section, you can also refer to the visualizations there if the images are blurry.

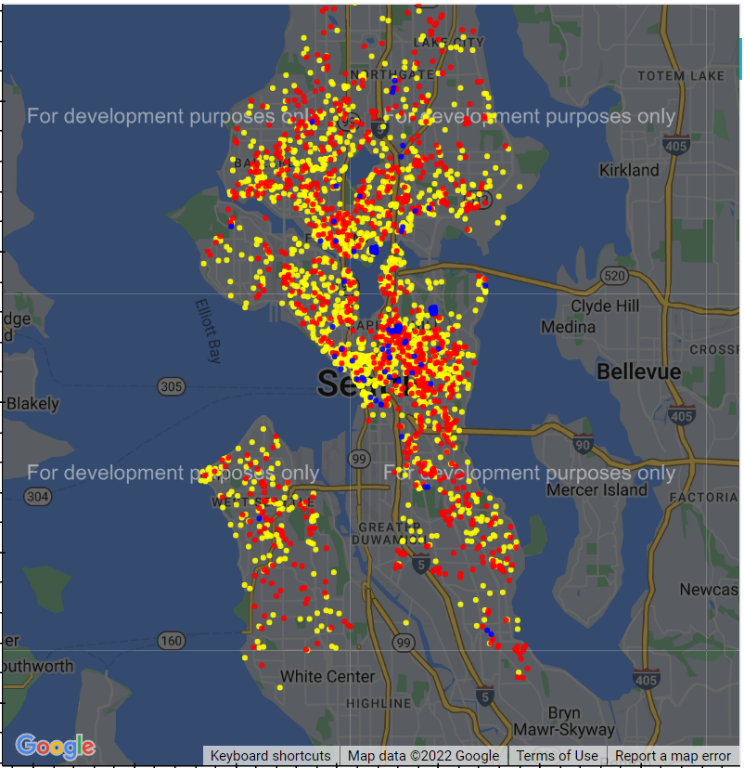


Fig 1

Chart, bar chart

Description automatically generated

Fig 2

Also, in [3], the author answers the question which is very useful for guests’ perspectives. In the analysis, the Airbnb data of ‘Vancouver’ is used. The aim of this analysis was to find inexpensive stays or listings in Vancouver. The author has performed many visualizations to answer different questions. The author analyzed the expense of staying for 1 month in different neighborhoods per room type. For this, the author chose the stacked bar chart to visualize this information.

As stacked bar charts are used when showing comparisons between categories, this method of visualization for the information representation is fine. The author has used neighborhood dimensions on rows and price measurements on columns and it can be easily visualized from the images that the author has shared, which neighborhoods as the least and most expensive listings according to different room categories.

## Chart Description automatically generated

## Fig3

# Methodology

## Data Collection

The data source used in this research report is collected from Inside Airbnb website. The link to the website is - http://insideairbnb.com/get-the-data/. Inside Airbnb, data community is an informal, self-managed and open group where you can ask questions about the data, discuss important community topics, and share resources and news [4]. The data source is not managed by Airbnb directly or indirectly and all the information that is used in the data source is safe to use. The data is picked from Airbnb public details such as property names, property locations, number of reviews on a property, pricing details, neighborhood of the listings

As I mentioned above, the data source is created based on the active listings on Airbnb at that particular time, so if the listing gets deleted, the data in the data source file becomes scrappy data. The location data such as latitude and longitude are exactly up to 150 meters. Details provided are not at all private and are publicly available on Airbnb [5]. The data inside Airbnb data is free and is available to anyone who wants to participate in the community. The data is updated after every quarter to make the most relevant data source possible.

## Data Processing

To build an effective and interactive visualization, the most important step is to clean your data. This is necessary because incorrect values, duplicate values, and missing values in a dataset can result in inconsistent visualization designs, and hence, the aim of doing the visualization gets affected. The data set that I collected from the inside Airbnb data website was already clean, verified, and aggregated. All this saved me a lot of time, therefore choosing the correct and clean dataset becomes a very important step. However, data fields such as ‘neighborhood groups that were present in my dataset were irrelevant to my questions and were of no use, so I decided to remove the values from that field. Moreover, when dealing with large datasets, it is recommended to remove unwanted fields as it increases the workload and processing time of the tableau dashboard. Apart from that, there were some places in the ‘price’ column field where the price was listed with currency notation, which too was removed during the cleaning process. I performed the data cleaning by using the Microsoft Excel program.

After following these 2 steps, the data set was ready to use and perform visualizations that are consistent, interactive, and logical.

# Analysis

During the course of this report, I analyzed and implemented visualizations for 3 questions intended for 3 different audiences - guests, hosts, and the organization itself.

## Listing Frequency

As an Airbnb host, it is important to analyze and decide which areas have the greatest number of listings and which areas have the least number of listings. In order to get more bookings, a host will always want to list the property in the least concentrated areas, that way the listing will be more likely to be discovered due to less competition.

The screenshot of the visualization to answer the above question is attached below in fig 4.

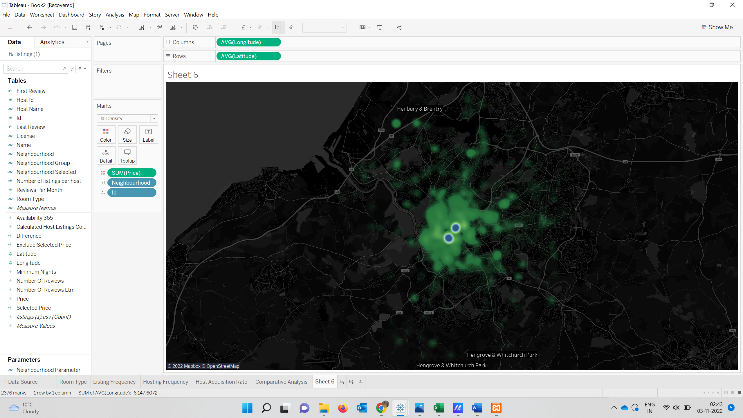


Fig 4

For this visualization, I have used heat maps. This is because when we use heat map or density plot, tableau automatically color codes the region based on the intensity of the clusters in the map i.e., the higher the density - the greater the number of plots within that cluster, and then the more intense the color is going to be. By doing this, we can clearly see which regions in the city have more listings (the dense area and increased color intensity) and which regions in the city have comparatively fewer listings (less dense area with light color intensity).

Through this visualization, we have successfully answered that the listing frequency in the central region of the city is less, and the listing frequency is higher in other regions of the city. As a host, now I have the information and insight about which area would be best if I was to publish my listing.

If you look closely, I have made some changes in the visual encoding of a normal heat map, I changed the background color to black and enabled ‘street, highways, and roads in map background layers. Apart from this, I have changed the color pallet from automatic to ‘green-blue-diverging’. This is done only to make the visualization more interactive as it also shows the road and street connectivity with the neighborhood where the property is listed.

Worksheet Name - Listing Frequency

## Host Acquisition Rate

As a key decision-maker for an organization, it is very important to discuss and analyze the company’s performance, profits, and different strategies. It’s obvious when dealing with statistics of a large organization, the data is too much and sometimes can be complex to understand. This is where the visualization of data helps many individuals in their decision-making. As an Airbnb organizational data analyst, I would like to analyze different factors such as the growth of the company. An organization’s growth (Airbnb in this case) hugely depends on the host retention rate. This shows the trust of a host in the application and the performance of the platform. A low user acquisition rate is generally indicative of a company’s decline and vice versa. Therefore, in this section, I will analyze the host acquisition rate of Airbnb.

The screenshot of the visualization to answer the above question is attached below in fig 5

Graphical user interface, application

Description automatically generated

Fig 5

For this visualization, I have plotted a line graph to indicate the customer acquisition rate. I used line graphs because they are recommended graphs when plotting trends over time. You can generally observe back in your minds, that for this plot - the steeper the line means the higher the acquisition rate.

* Assumption:

To visualize this plot, I needed to create or find an expression from the dataset that returns the date when a host first started renting their property through Airbnb. Looking at the data source, I didn’t have any such field, but I had a date field named ‘last review’. This field contains the date when a listing received its last review. So, for sake of this plot, I assumed that the date of the first review on the property is the date when a host first started renting their property on Airbnb.

* Explanation:

As there was no first review date in the data source, I created a calculated field ‘First Review’. This was accomplished using a fixed expression. For each distinct host id, I determined the minimum review date (last review date). I used the minimum expression to get the earliest value from the last review date field so that we can assume that date as the first review date on a listing of a host.

The expression I used to create the first review date is - {FIXED [Host Id]: MIN([Last Review])}

To plot the desired graph, now we have the desired first review column. I have used count distinct of host id to make groups of hosts that signed up in a particular year. As is clearly visible from the graph, the host acquisition rate has increased slowly and steadily up to the year 2020 and took a huge step after 2020. This means that Airbnb was performing well in Bristol city last year when compared to previous years.

* Worksheet Name - Acquisition Rate

## Comparative Analysis

As guests, when finding properties for vacation, everyone looks for properties that are affordable and have feasible prices. But property prices can vary in neighborhoods based on the time of arrival or area of the neighborhood. So, as a guest, it becomes important to see and analyze which neighborhoods have higher prices and which neighborhoods have low prices. The guest can compare the prices in order to plan a trip within his budget. To answer these problems for a guest, I chose to plot a comparative analysis visualization looking at which, a guest will have an insight into the price differences in the neighborhood.

Chart

Description automatically generated

Fig 6

To make this plot, I created a parameter called ‘Neighborhood Parameter’ and color-coded the selected parameter such that the selected parameter will be highlighted with the darkest red shade and un-selected parameters with a light red shade. To color code the price and relate it with the parameter, I created a calculated field - ‘Neighborhood Selected’. By doing so, the selected neighborhood will get highlighted with the help of color encoding. The aim of doing this was to see the difference between bars when comparing prices with other neighborhoods.

Additionally, to interpret the difference in prices, I created another calculated field called ‘Difference’. When plotting these fields on view, I was able to plot the desired visualization.

* Worksheet Name - Neighborhood Price Differences

# Conclusion

In this research report, I performed visualizations on Airbnb properties that are listed in Bristol city of the UK. The aim of this report was to reflect the visualization knowledge and apply that knowledge and principles to a data source by performing interactive and useful visualizations.

The 3 questions that I answered in this report were intended for each type of audience i.e. the guests, the hosts, and the organization itself.

* 1. In the first visualization - **Listing Frequency**, I used tableau-provided options to make the basic map more interactive by changing the color pallets in order to make the cluster data clearer and more effective visually. This visualization was aimed at hosts who want insight into areas so that they can list their properties accordingly.
  2. In the second visualization - **Acquisition Rate**, I used a line chart to show a trend over time i.e., with which rate the hosts are getting on board with the platform over the years. To get to my aim, I created and used a calculated field. This visualization was aimed at organization people who want to analyze Airbnb performance for a particular region

**Limitations and Future Discussions**

* When performing this visualization needed some field that can give insight into the host’s first onboarding date on the Airbnb platform. As this data was not available in the data source, I created a first review calculated field and presumed the first review date on the listing as the host onboarding date on the platform. So, the information from the visualization is not 100% accurate, it’s just a vague idea. To overcome this and to increase the accuracy of the visualization, we can find the host detail listing from the same data source i.e. inside the Airbnb website that I mentioned in the data source section.
* Tableau provides us the power to join different data source files by using JOINS &UNIONS so that we can gather all the required data in one dataset. Using joins and unions, we can use the original host sign-up date in our dataset, and by doing so, we can make our visualization more accurate and perfect.
  1. In the third visualization - **Comparative Analysis**, I used parameters, calculated fields, LOD Expressions, and color encoding principles to make my visualization. In the end, the price difference between neighborhoods is clearly visible. This visualization was aimed at guests and tourists who want to check different places to plan their trip and expenses.

**Limitations and Future Discussion** -

In this visualization, we can use reference lines to indicate the average price. Apart from this, we can use tooltips with color encoding (red if the average price of a neighborhood is lower than the average price selected neighborhood and green if the average price of a neighborhood is greater than the average price selected neighborhood). Using tooltips shows indications to the new viewers as well.

##### References

This section contains the name and information of all the authors whose work was analyzed, read, and observed.

1. Jia Xin Poon, July 2017, A Theoretical-Framework-in-Determining-Travellers-Intention-of-using-Airbnb-An-Exploratory-Study, Retrieved November 02, 2022, from-

<https://www.researchgate.net/project/A-Theoretical-Framework-in-Determining-Travellers-Intention-of-using-Airbnb-An-Exploratory-Study>

1. Mohammed Irfan, August 2020, Data Science Project, Retrieved - November 02, 2022, from -

<https://mohamedirfansh.github.io/Airbnb-Data-Science-Project>

1. Nikhil Chodimella || Praneet Atwal, Updated April 18, 2019, Final Project - Airbnb Data Visualisation, Retrieved 02 November 2022 from

<https://observablehq.com/@chodimella/final-project-airbnb-data-visualisation>

1. Inside Airbnb Data Community, Retrieved 02 November 2022 from

<http://insideairbnb.com/data-community>

1. Data Assumptions, Retrieved 02 November 2022 from

<http://insideairbnb.com/data-assumptions>