**Airbnb Analysis:**

**Introduction:**

The New York City Airbnb Open Data is a large and detailed dataset that provides information on over 48,000 Airbnb listings in the city. This dataset is important because it offers insights into the Airbnb market in New York City, which is a major tourist destination, and can be used to study the impact of Airbnb on the local housing market and economy.

**Data Dictionary: Source (www.kaggle.com, n.d.)**:

<https://www.kaggle.com/datasets/hasi007/airbnb-newyork-dataset>

| **Variable Name** | **Variable Type** | **Description** |
| --- | --- | --- |
| id | Numeric | A unique identifier for each Airbnb listing |
| name | Text | The name of the Airbnb listing |
| hostß\_id | Numeric | A unique identifier for each host |
| host\_name | Text | The host name |
| Neighbourhood\_group | Text | The borough in which the Airbnb listing is located |
| neighbourhood | Text | The neighbourhood in which the Airbnb listing is located |
| latitude | Numeric | The latitude of the Airbnb listing |
| longitude | Numeric | The longitude of the Airbnb listing |
| room\_type | Text | The type of room being rented |
| price | Numeric | The nightly price for the Airbnb listing |
| minimum\_nights | Numeric | The minimum number of nights a guest can book |
| number\_of\_reviews | Numeric | The number of reviews the Airbnb listing has received |
| Booking\_date | Date | The date of the booking for the Airbnb listing |
| reviews\_per\_month | Numeric | The average number of reviews per month for the Airbnb listing |
| availability\_365 | Numeric | The number of days in a year that the Airbnb listing is available for booking |

Top of Form

**Persona and questions:**

**Persona:** The user is an “Airbnb host” who plays a crucial role in the Airbnb ecosystem by renting out their properties to guests looking for a place to stay. An Airbnb host may be interested in the things could affect their profitability like:

* The demand for Airbnb rentals in various neighbourhoods,
* The average listing price,
* The reviews and ratings of comparable listings.

HCI (Human computer interaction) studies how people use computers and software (Preece, Sharp and Rogers, 2015). UCD (User-Centered Design) software development focuses on end-user demands, wants, and limits. User personas used HCI/UCD theory to this project. The design team met the target audience's needs by creating personas. This method made the solution usable and addressed end-user needs. User testing utilised HCI/UCD principles (Dix, 2003). User testing assessed solution usability and identified opportunities for improvement. This made the solution user-friendly and met end-user needs. The project showed how HCI/UCD could produce effective and user-friendly solutions.

**Research Questions:**

**Q1.** Which neighbourhoods are the most expensive to stay in and have similar rental prices, and what is the availability and range of room types in each of these neighbourhoods?

**R1:** To execute Q1, we must utilise two specific visual charts; a map and box plot. We need to visualise the variables neighbourhoods and average price in a way that conveys the rent of that area.

A good view would be a Map, help visualise geographical data. A map can illustrate New York City neighbourhoods and Airbnb pricing. A map can rapidly identify the most expensive neighbourhoods and compare costs. We also utilise a box plot to investigate the number of days per year an Airbnb listing is open for booking and the availability of different room kinds. The distribution of data for each accommodation type can assist hosts manage reservations and price based on demand.

**Q2.** Which neighbourhood had the highest number of reservations and reviews in the month with the most bookings, and what is the percentage of each neighbourhood group represented in that month?

**R2:** To address Q2, we require a visual representation of a tree map and donut chart are used. We need to visualise the relation between neighbourhood and count of names and try to find the corelation between them.

When displaying hierarchical data, such as the total number of bookings and the monthly average ratings in various neighbourhoods, a tree map chart is effective. On the other hand, a doughnut chart is a fantastic method to rapidly assess the percentage of each neighbourhood group represented in that month.

**Design:**

There are 2 stages of designs, prototype view and final tableau dashboard.

Prototype (Paper Landscape):

Tree map reflects the type of room and its average price of neighbourhood.

Chart

Description automatically generated

The combination of stacked bar with line to display the number of bookings in the month.

Pie chart gives the total share of the neighbourhood group.

Box plot to display the

availability of room types in

Neighbourhood

Figure1: Paper landscape of dashboard

The paper landscape includes four visual charts: a tree map, a box plot, a dual axis of stacked bar with a line, and a pie chart. R1 is met by utilising, the tree map chart is used to identify the most popular and expensive regions based on the average price. The size of each block on the tree map makes it easy to identify which areas are pricier. A box plot map shows the types of rooms available in each community as well as the availability of 365 days.

As a result, hosts may better manage their bookings and modify their pricing strategies to suit demand. R2 is solved by using, based on the booking date, a stacked bar and a line are used to identify which month had the most reservations- a combination chart got an idea from (help.tableau.com, n.d.). This helps to adjust the listings as needed and prepare for periods of peak demand. The percentage of each neighbourhood group and the overall number are also displayed on the pie chart. This offers an accurate picture of the distribution of reservations.

Tableau Design (Final Implementation):

Chart, treemap chart

Description automatically generated

Figure 2: Final dashboard of Tableau

There are total of four charts in the final dashboard that was created in tableau. They are - a map, a box plot, a tree map, and a doughnut chart. The dashboard prototype and final design have only few differences which are listed below:

1. While a tree map may be able to depict the relative sizes of communities based on average price, it may not provide the same level of physical location clarity as a map when displaying the most expensive neighbourhoods in New York City.
2. The prototype showed the highest reservation utilising a dual axis of stacked bars with a line. A tree map may be better than a dual axis of stacked bar and line for showing the highest reservation made for each neighbourhood over all months and their average review. This is done so a tree map can clearly indicate the relative size of each neighbourhood based on the number of bookings and their average rating.
3. A "doughnut chart" is a pie chart with a hole in the middle to maximise space and improve organisation. Doughnut charts may help visualise neighbourhood shares and percentages. Donut charts are more flexible than pie charts (www.beautiful.ai, n.d.) because the hole in the middle allows for fresh information or labels. Doughnut charts are usually more useful and attractive than pie charts.
4. The utilisation of the above charts and clustering gave the tableau implementation a more knowledgeable and superior dashboard than the prototype as the prototype does not make use of new complicated charts or capabilities. In this way, our R1 and R2 can be designed with more clarity.

**Implementation:**

This section goes into detail about how the design is applied to the tableau.

Graphical user interface, map

Description automatically generated

Figure 3: Expensive Neighbourhood based on Average price.

The first chart is a map that shows readers exactly where the neighbourhoods are in relation to one another. The following actions were done to make this map:

1. To generate the map, latitude and longitude were first chosen.
2. Then, the detail part of the marks field was expanded to include Neighbourhood groups and Room types.
3. The average measure of price was then picked, and it was used to determine the average price for each neighbourhood.
4. To depict the most expensive places more accurately, the same average price was used for the text field of the marks.
5. To display the price's intensity, density was set to visible. The graph makes it simple for users to see the top and lowest average rent values in each neighbourhood.

Clustering: This helps discover the costliest neighbourhoods with similar rental prices. Clustering neighbourhoods by average price saves time and improves efficiency. Manually walking over each neighbourhood without clustering is inefficient (Pandey, 2021). Now we have 4 clusters based on average price.

Text

Description automatically generated with medium confidence

Figure 4: Summary of the clusters

Graphical user interface, timeline

Description automatically generated

Figure 5: Models of Clusters

Chart, box and whisker chart

Description automatically generated

Figure 6: Type of room and their Availability

An effective tool for Airbnb hosts to analyse the availability of various room types and spot patterns or trends in the data is the box plot, which is shown in the chart above. The steps below can be used to produce this chart:

1. To determine the number of days a listing is open for reservations, choose the neighbourhood group and availability 365 as the columns and rows, respectively.
2. To get the typical days of availability for each neighbourhood, convert the availability 365 number to an average value.
3. To distinguish between the three sorts of rooms, assign the room type a colour.
4. In the details section, include the average availability.
5. Select the boxplot chart type to examine the demand for various room kinds and, based on the information, modify pricing and availability.

The figure shows that Staten Island has the most room availability, which could be helpful, as we can plan the pricing and marketing efforts for the area.

Chart, treemap chart

Description automatically generated

Figure 7: Monthly reservations and Average review of Neighbourhood

The following diagram is a Tree Map, which is a practical approach to compare data across several communities and identify which ones have the most bookings and ratings.

We took the following actions to make this chart:

1. We first added the "month" field to the column portion of the chart after selecting the "last review" field.
2. Next, we added the "name" and "neighbourhood group" fields to the chart.
3. We counted the number of reservations made in each neighbourhood for the chosen month using the "count" function.
4. The size and text fields of the chart marks were then set to the number of reservations.
5. We coloured the marks according to the number of reservations to give the chart more information. To have the name of the neighbourhood group appear on the chart, the neighbourhood group was assigned to the text field.
6. To display the name of the month, we lastly set the month of the most recent review to the text box.

This graph makes it simple to see which neighbourhood had the most bookings in a particular month, which can assist in tailoring the pricing and marketing strategy.

A picture containing chart

Description automatically generated

Figure 8: Neighbourhood Diversity

This is a doughnut chart, Doughnut charts are like piled bar charts, curled around themselves so that both ends meet and form a circle. This was tough to make because it is a new type of chart which is complex yet gives more clear view, overcame this issue with the help of (Intellipaat Blog, 2021). If any neighbourhood group is selected from filter, doughnut chart will show the percentage of contribution of that area. The above chart is a Doughnut Here are the steps to create a Doughnut Chart in Tableau:

1. Set the base in Tableau by typing sum(0) in the columns field.
2. Drag the neighbourhood group to the colour of the marks field and select the pie chart.
3. Set the neighbourhood to the size and count as the measure.
4. Add the neighbourhood group and its count to the text to display the values.
5. Select the count of neighbourhood and select "percent of total".
6. Duplicate the sum(0) field and rename it to sum(0)(2). Remove all the fields present in it.
7. Select this colour as white and decrease its size to create a hole in the pie chart.
8. Make it a dual-axis chart.

Finally, the chart displays the percentage of total neighbourhood by neighbourhood group.

**Second Implementation: Dashboard in Power BI**

Chart, treemap chart

Description automatically generated

While Power BI Desktop offers a more organised interface that makes creating uniform dashboards across many reports easier, Tableau gives a more flexible approach to constructing dashboard layouts, allowing for precise placement, and scaling of dashboard items. Tableau and Power BI differ significantly in that Tableau is more customizable and enables the creation of custom visualisations like box plots without the need for additional tools, whereas Power BI requires the use of custom visuals that are only accessible with a company or academic email login. Therefore, rather than using box plots, Power BI users may have to make do with simpler visualisations like bar charts. It isn't easy to use Power BI because, unlike Tableau, we cannot display the information we want to say. For example, when displaying a map, we can only show the neighbourhood group; however, we cannot display the average price beside the group name. Tableau offers a better user experience and more colour palette possibilities, while Power BI's dashboard development tools give fewer design options. Beginners may find it more difficult to use Power BI, especially after getting used to Tableau. (Biswal, 2023).

**Walkthrough:**

1. **Which neighbourhoods are the most expensive to stay in and have similar rental prices, and what is the availability and range of room types in each of these neighbourhoods?**

Airbnb hosts may benefit from knowing the costliest neighbourhoods and accommodation types. By knowing the demand for different listings in expensive locations, hosts can adjust their listings to better serve guests. This boosts host income and occupancy. Hosts can improve reservation management and pricing by analysing room types and availability in expensive locations. They can determine which seasons are more popular and adjust their prices or determine which seasons are less popular and schedule maintenance and cleaning. The below map shows the costliest neighbourhood based on average pricing,

Map

Description automatically generated

There are five neighbourhood groups and three types of rooms in NYC.

The second most expensive neighbourhood is the Manhattan with rent of 233 USD per night.

Map

Description automatically generated with low confidence

The most expensive place to stay is Staten Island for Entire room type, where the mean price is about 241 USD per night.

Chart, box and whisker chart

Description automatically generated

On the other hand, despite being the second most expensive group, Manhattan offers mostly shared rooms and entire rooms has least availability in it.

As staten island is expensive, its available for average of 300 days.

But this group doesn’t offer the option of shared rooms.

Queens is cheapest of all neighbourhoods which has shared room at $49 per night and good availability.

Chart, box and whisker chart

Description automatically generated

Cost effective group will be Brooklyn as its average price is also decent and availibilty is less compared to other groups which means it is preffered mostly

1. **Which neighbourhood had the highest number of reservations and reviews in the month with the most bookings, and what is the percentage of each neighbourhood group represented**.

The Airbnb host persona can learn which neighbourhoods’ tourists prefer best by looking at the month with the most reservations and reviews. This data helps them optimise listings and focus marketing. Knowing the amount of each neighbourhood group helps hosts target guest groups and adjust pricing.

We use a tree map to answer- which neighbourhood had the most reservations and reviews during the busiest booking month. This chart compares neighbourhood groups by showing the month with the most bookings, the linked groups, and their ratings. Check the below chart to know which month got highest reservations.

Chart, treemap chart

Description automatically generated

The month with the highest number of bookings was June, with a total of 489 reservations. Most of these bookings were made during the summer season

Chart, treemap chart

Description automatically generated

In June, Manhattan had the highest number of reservations with total of 489 bookings with 1.5 avg review. This suggest that this group is popular and preffered location among guests.

Chart, treemap chart

Description automatically generated

Staten island has very least bookings compared to other neighbourhoods.

A doughnut chart determines the percentage of each neighbourhood in each group for the second half of the question. Hosts can better assess neighbourhood popularity and adjust their strategies with this information.

**Chart

Description automatically generated**

On the other hand, Manhattan and Brooklyn have a high percentage of shared neighbourhoods, indicating a high level of diversity and population density in these areas.

Staten Island has a very low percentage of shared neighbourhoods, so the average prices are high,

which contributes to its low number of bookings.

**Final Analysis:**

The focus is Airbnb Host will be to make informed decisions about how to market and price their rental property to attract more guests.

**1.** Based on the analysis, the costliest neighbourhoods to stay in are Staten Island and Manhattan.

-In Staten Island, the availability of rental rooms is higher compared to other neighbourhoods.

-On the other hand, Brooklyn has the least availability of rental rooms among the five neighbourhoods analysed.

**2.** Analysing the mostly booked neighbourhood and the percent of total neighbourhood count.

- Manhattan and Brooklyn are the most booked neighbourhoods during the month with the highest number of reservations (June).

-These two neighbourhoods’ group has the highest percent of total share, which suggests that there is a high demand for this type of accommodation in these areas.

**Conclusion:**

1. Consider the balance between rental price and availability: Hosts with properties in neighbourhoods like Staten Island and Manhattan may have higher rental prices due to the higher market value, but they should also ensure that the availability of rental rooms is reasonable. If the availability is too high, it may indicate that the rental prices are too high, which may lead to lower demand and lower profits.

2. Adapt to the market trends and preferences: Hosts should analyze the market trends and preferences of potential guests, such as the popularity of shared rooms in highly booked neighbourhoods like Manhattan and Brooklyn. By adapting to these trends and preferences, hosts can cater to the demand and attract more guests, which will ultimately lead to higher profits.

**Comparing Dashboard and story:**

A story is a well-organized narrative that, using the information at hand, offers a clear message or responds to a particular question (www.linkedin.com, n.d.). By examining neighbourhood pricing, availability, and room kinds, it can be utilised to assist Airbnb hosts in maximising their earnings. The performance of a process can be tracked using a dashboard, which is a visual representation of data points (www.linkedin.com, n.d.). The average pricing, availability, reservations, and reviews for each neighbourhood can be shown on an Airbnb dashboard in this case, giving hosts a rapid summary of the important indicators, they need to keep an eye on.

**Reflective Discussion:**

I've learned a lot about using data visualisation tools like Tableau and Power BI because of doing this assignment. I've discovered that carefully considering the data and the intended audience is necessary for producing excellent visualisations. To ensure accurate findings, the data needed to be cleaned and organised before being turned into a visualisation. A more user-centred approach to the design process was made possible using personas and user testing, which produced a more useful and efficient solution.

The ability of these visualisation tools to handle enormous datasets and the feasible degree of customizability are two areas that should be improved for future projects. Overall, working on this project was an amazing opportunity for me to apply the principles I learnt in class and develop my practical data analysis and visualisation abilities. I want to take part in seminars on Tableau's data storytelling, read relevant literature, and exercise creating data visualisations with storytelling elements.

**Conclusion:**

This project analysed the NYC Airbnb dataset to answer two questions on neighbourhood pricing and booking trends. The strategy included determining the most popular neighbourhoods and their preferred room types by analysing the most expensive neighbourhoods' rental costs and room type availability. The findings revealed that while Brooklyn and Manhattan are highly booked with a larger demand for shared accommodations, Staten Island and Manhattan are pricey with a higher market value. To increase profitability, hosts might utilise these insights to modify their pricing and marketing plans.

**Reference list**

Biswal, A. (2023). *Power BI Vs Tableau: Difference and Comparison*. [online] Simplilearn.com. Available at: https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-vs-tableau.

Dix, A. (2003). *Human-computer Interaction*. [online] *Google Books*. Pearson/Prentice-Hall. Available at: https://books.google.co.uk/books?hl=en&lr=&id=IuQxui8GHDcC&oi=fnd&pg=PR14&dq=Dix [Accessed 22 Apr. 2023].

help.tableau.com. (n.d.). *Quick Start: Combination Charts*. [online] Available at: https://help.tableau.com/current/pro/desktop/en-us/qs\_combo\_charts.htm.

Intellipaat Blog. (2021). *Create Donut Chart in Tableau with 10 Easy Steps*. [online] Available at: https://intellipaat.com/blog/tableau-donut-chart/#:~:text=Tableau%20is%20created.- [Accessed 22 Apr. 2023].

Pandey, P. (2021). *Cluster Analysis in Tableau*. [online] Medium. Available at: https://towardsdatascience.com/cluster-analysis-in-tableau-1f19acd0c647 [Accessed 22 Apr. 2023].

Preece, J., Sharp, H. and Rogers, Y. (2015). *Interaction Design: Beyond Human-Computer Interaction*. [online] *Google Books*. John Wiley & Sons. Available at: https://books.google.co.uk/books?hl=en&lr=&id=n0h9CAAAQBAJ&oi=fnd&pg=PR10&dq=Preece [Accessed 30 Mar. 2023].

www.beautiful.ai. (n.d.). *Battle of the Charts: Pie Chart vs. Donut Chart| The Beautiful Blog*. [online] Available at: https://www.beautiful.ai/blog/battle-of-the-charts-pie-chart-vs-donut-chart.

www.kaggle.com. (n.d.). *Airbnb NewYork Dataset*. [online] Available at: https://www.kaggle.com/datasets/hasi007/airbnb-newyork-dataset [Accessed 22 Apr. 2023].

www.linkedin.com. (n.d.). *Explain the difference between Tableau Workbook, Story, Dashboard and Worksheets?* [online] Available at: https://www.linkedin.com/pulse/explain-difference-between-tableau-workbook-story-upendar-sheethala/.