# FINAL EXAM CDA 500
(Shravani Soma - 50477925)

This paper provides an in-depth statistical analysis of the Boston Airbnb dataset, specifically focusing on properties that have received reviews in the past five years to ensure the relevance and accuracy of the data. It explores various factors influencing Airbnb hostings, offering valuable insights to the company(The Board) about potential new properties that could be introduced in strategic Boston locations to boost profitability.

1.Deliverables:

(i) Visual1: Rooms and Availability
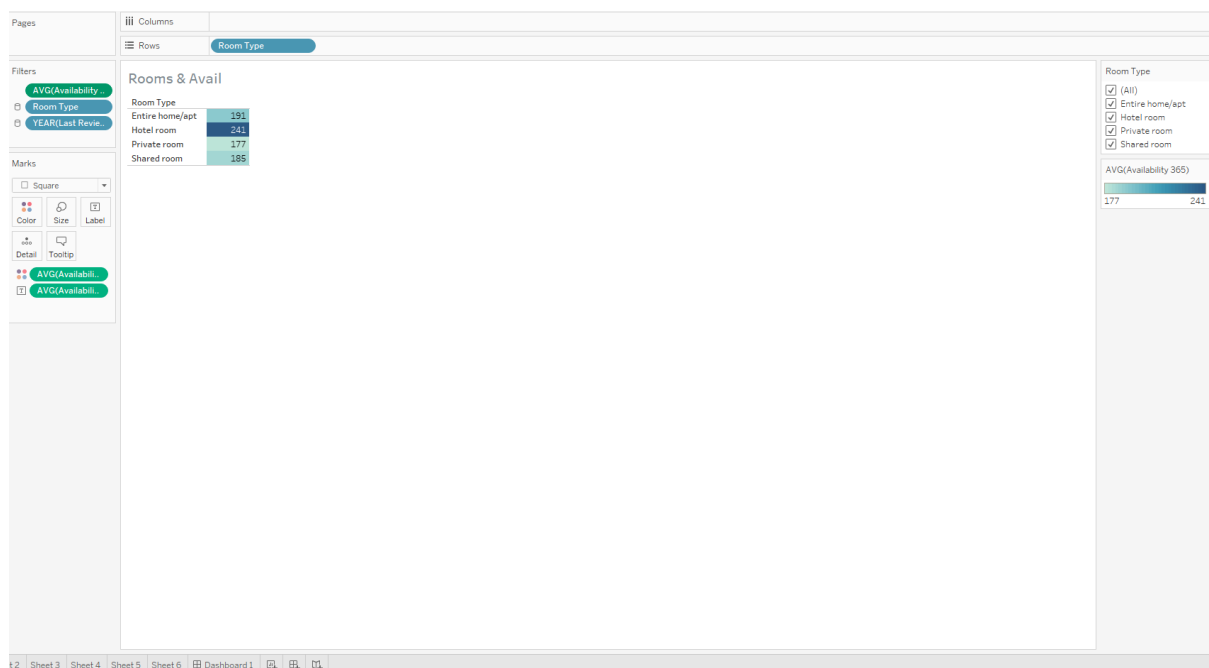
The provided visual represents the average number of days that four different types of rooms are available, specifically filtering for properties that have 365 days of availability annually. Hotel rooms top the list with an average availability of 241 days. And private rooms have the least average availability of 177 days. This visual provides the board with insight into which room categories have the lowest average availability, indicating areas that may require attention or improvement.

I have ensured that this visual rightly follows the Gestalt principles. The absence of lines in the table helps maintain **closure**, allowing the audience to concentrate solely on the numbers. The use of sequential colours further highlights the extremes (the highest and the lowest) to effectively draw attention where it is most needed.

(ii) Visual 2: Property types - Reviews
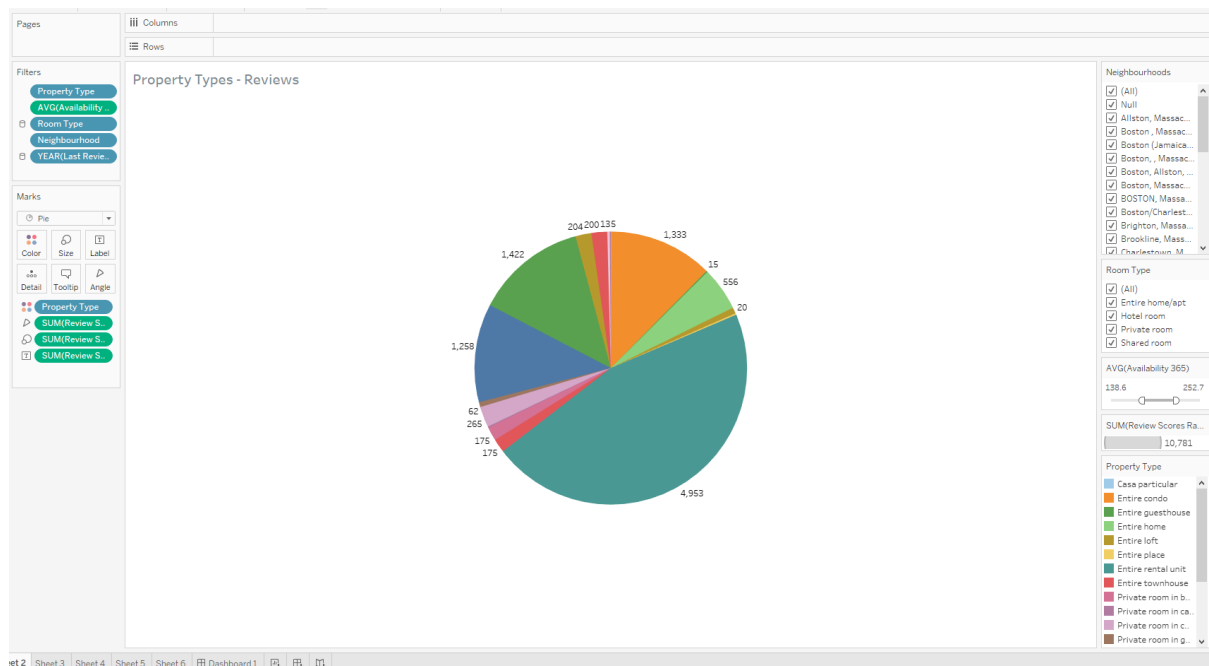
The presented pie chart illustrates the proportion of total reviews for each property type. This analysis has been conducted exclusively on properties reviewed within the past five years. Additionally, there's a neighbourhood checklist filter that allows the selection of specific neighbourhoods for a more focused analysis. The property type "Shared room in boutique hotel" has the lowest number of reviews, while "Entire rental unit" boasts the highest.

The pie chart uses various colours to distinguish between different property types. To prevent any confusion, all numbers are placed outside the chart, given their lesser importance in this context. The primary focus here is on the colours, which indicates each property type's share of the whole chart. To avoid clutter, property types are not written within the pie chart and instead, this information can be accessed as needed by hovering over the relevant chart portion.

(iii) Visual 3: Prices Over Years
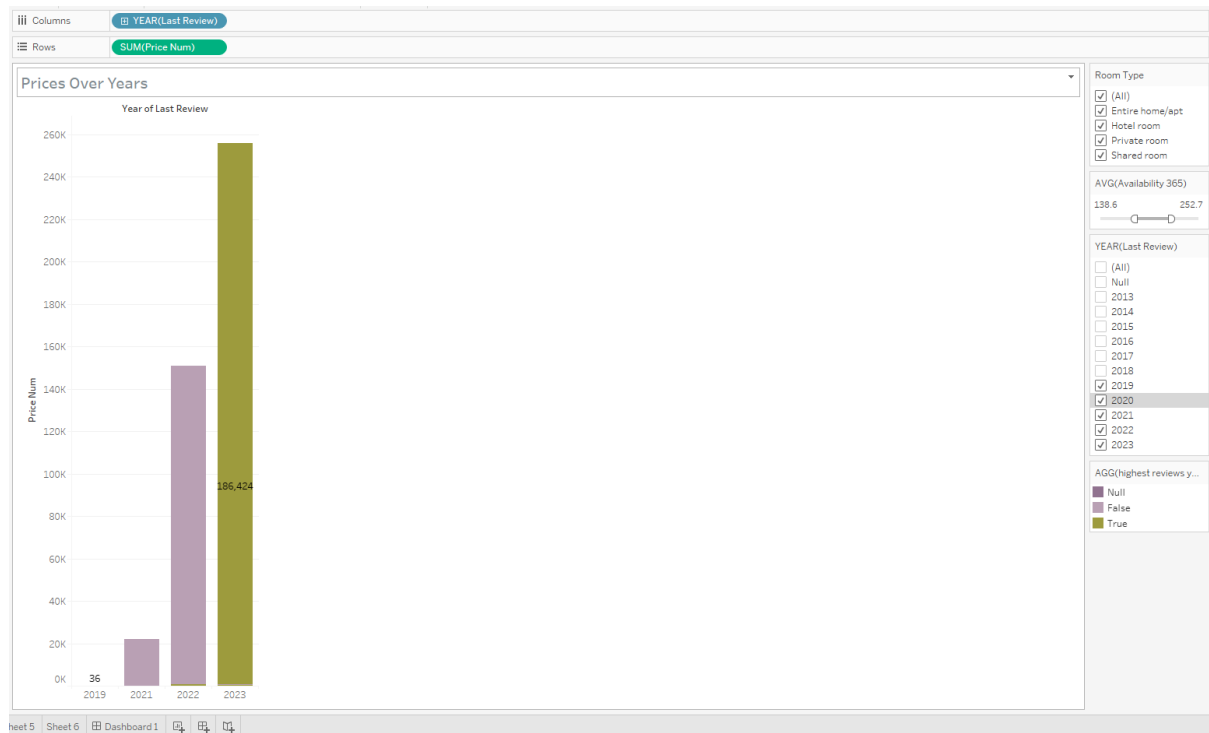
This bar graph presents the price trends of properties over the past five years. It indicates that 2019 had the lowest pricing, at \$36K, while 2023 recorded the highest pricing at \$186,424K.

The application of the Gestalt principle of **similarity** can be observed in this chart. All the years are colour coded similarly, except for the year with the highest pricing, indicating their similar significance. The chart's primary focus is on the years with the highest and lowest prices, hence the unique colour coding for the year with the highest pricing.

(iv) Visual 4: Average number of bedrooms
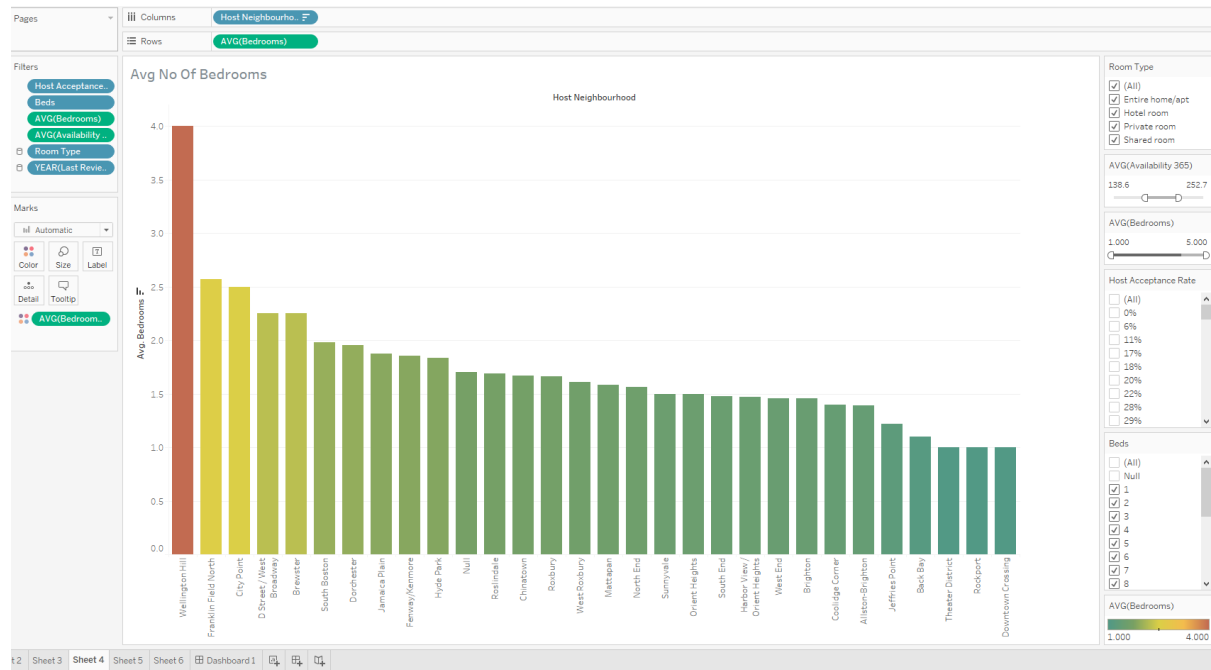
The bar graph presents the average number of bedrooms across different neighbourhoods. It specifically filters out those properties with a host acceptance rate of 50% or above. The chart also offers the ability to further filter for specific room types.

This graph follows the **similarity** principle, all neighbourhoods are indicated using the same colour tone, except for the one with the highest number of bedrooms.The deliberate use of colours plays a crucial role in this graph to portray the trends through a monotone sequential colour palette.

(v) Visual 5: Year Vs Reviews

This above line graph displays two line trends for number of reviews and reviews score rating based on the host year. Red line indicating for the number of reviews trends and brown for the reviews score ratings. Only the maximum and minimum peak points are marked for both the lines.

Colours have been specifically used to distinguish two different lines. **Continuity** among the Gestalt principles is maintained. Without the over usage of colours, this graph has been presented in the most reader-friendly way.

(vi) Visual 6: 15 Expensive property types

The table presented above details the 15 most expensive property types. Each row lists the price and rank for a specific property type. Filters have been applied for 'host identity verified' with a True value (across all worksheets), as well as 'has availability' also with a True value (across all worksheets). The ranking is derived from a new calculated field 'Rank' based on pricing.

In terms of Gestalt principles, **closure** is achieved by omitting unnecessary lines for column separation. And **proximity** is ensured through appropriate spacing, hence enhancing visual clarity and ease of number reading. The table is presented without the use of colours to facilitate a more reader-friendly experience.

2.Deliverables:
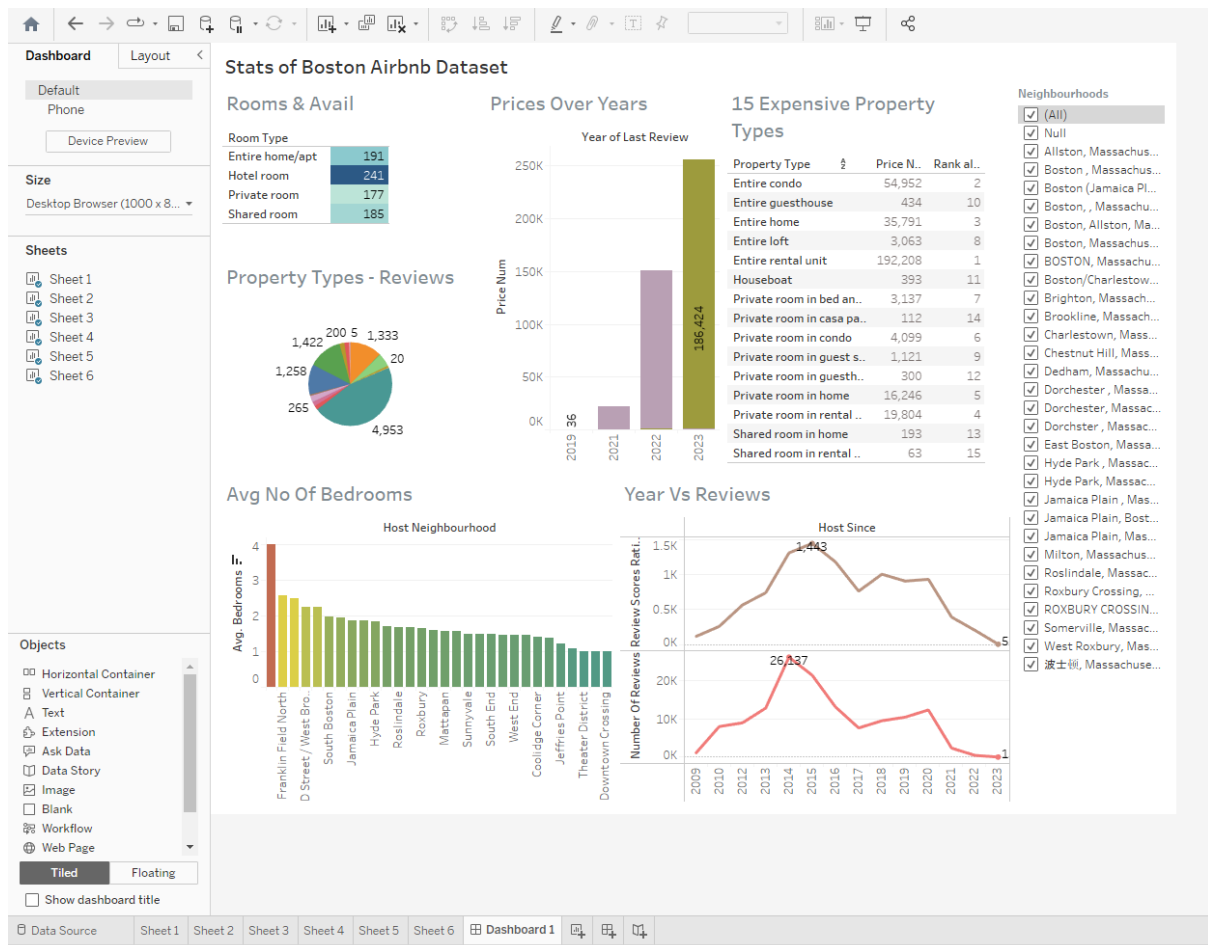
The presented dashboard offers insights into the Boston Airbnb dataset, specifically focusing on properties that have received reviews in the past five years. Comprised of six worksheets, as discussed previously, and an additional neighbourhood checklist filter, it allows for specific area selection or combined neighbourhood analysis.

The layout of the dashboard elements is effectively designed to ensure a better viewer experience. The bar and line graphs mentioned above, that require more expansive space, are placed in the lower horizontal section of the dashboard. The remaining graphs, which do not need broader areas, comfortably fit in the upper section, ensuring a reader-friendly display.

All filters, that do not add a relevance to the dashboard, have been omitted. A differentiation is made between the dashboard title and the worksheet titles via a colour variation, enhancing readability. The careful usage of colours contributes to a smoother and more engaging viewer experience.

3.Deliverables:

The journey from studying the dataset to constructing an interactive dashboard has been enlightening, with its own set of challenges and subsequent solutions. The dataset contained 76 features, including a column titled 'Price' which was of string data type due to the presence of a '$' symbol alongside each price figure. Recognizing its significant value for the analysis, I needed to convert this 'Price' feature into a numeric datatype. A direct manual conversion to a number datatype resulted in a zero value in each row, which led me to create a new calculated field named 'Price Num'. This field was created by removing the '$' symbol, thus enabling a conversion to a numeric datatype. The formula used to calculate this field is provided below.

In order to rank property types in the sixth worksheet, a calculated field titled 'Rank' was created. The formula used to generate this field is provided below.
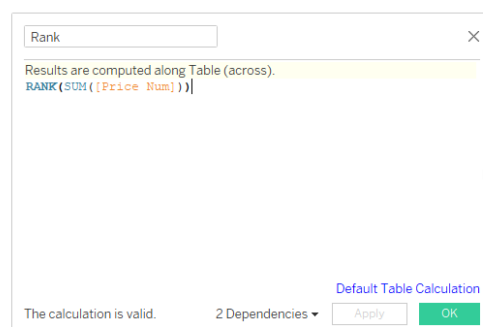
The first visual initially displayed the average days of availability in floating numbers, but this was altered to natural numbers for more relevance. Similarly, the average number of bedrooms was changed from floating to natural numbers to enhance comprehension. Many such subtle modifications were implemented to enhance the analysis and deliver more meaningful insights.

Overall, this comprehensive analysis can assist board members or senior management in understanding the business's performance. This can guide decision-making to improve operations and increase profitability.