Project Documentation

**Project Name**: AIRBNB MARKET ANALYSIS

**Team members:**

* Yepuri Harsha Vardhan
* Amajala Sai Ram
* Sree Nikhil Velicheti
* Vemula Kesavaaditya Gupta
* Myneni Gyanada Chowdary
* Akshitha Gopu
* Isha Narwaria
* Sarah Fathima Barkath

**Skills:**

* Python – numpy, pandas, matplotlib, seaborn, sklearn library
* tableau

**Input Data Size:** 6.59 MB

**Timeline:** 1 Month

**Introduction**

Airbnb is a popular online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in that locale. Instead of booking a hotel, travellers can book a stay directly with the owner of the property, whether it's a single room, apartment, house, or even a unique accommodation like a treehouse. The company acts as a broker and charges a commission from each booking.

Gaining an understanding of the Airbnb industry is the main goal of this data analytics research. Important indicators that will be examined include pricing trends, demand-supply dynamics, customer preferences, and regional distribution.

**Objectives:**

1. Analysing the seasonal demand patterns

2. Giving Airbnb actionable recommendations based on the analysis

3. Type of rooms that are booked the most based on reviews

4. Availability of rooms at a location in a year

5. Calculating the most used room type at nights

6. Customer to be able to book desired room by these visualisations

**Project Components and Skills Covered:**

**Data Collection:** Dataset of Airbnb New York 2019 (obtained from Kaggle, [Source](https://www.kaggle.com/code/chirag9073/airbnb-analysis-visualization-and-prediction/notebook))

**Data Preprocessing:**

Steps performed:

* Dropping unnecessary columns in a DataFrame
* Changing the index of a DataFrame
* Using .str() methods to clean columns
* Using the DataFrame.applymap() function to clean the entire dataset, element-wise
* Renaming columns to a more recognizable set of labels
* Skipping unnecessary rows in a CSV file

The duplicates from the dataset are identified and removed. Finally, the cleaned and preprocessed dataset is saved as ‘cleaned\_airbnb.csv’. The dataset is now prepared for further analysis.

**Data Cleaning:**

Steps performed:

* airbnb.shape reveals the number of rows and columns.
* airbnb.dtypes and airbnb.info() display data types and non-null counts for each column.
* airbnb.duplicated().sum() counts duplicates.
* airbnb.drop\_duplicates(inplace=True) eliminates the duplicates.
* airbnb.isnull().sum() quantifies missing values per column.
* airbnb.drop(['name','id','host\_name','last\_review'], axis=1, inplace=True) eliminates unnecessary columns.
* airbnb.fillna({'reviews\_per\_month':0}, inplace=True) replaces missing values in 'reviews\_per\_month' with 0.
* airbnb.dropna(how='any',inplace=True) discards rows containing any missing values.
* airbnb.describe() provides descriptive statistics for continuous numerical columns.
* airbnb.columns lists all column names.

Imported the cleaned csv file into Microsoft Excel for further analysis. The dataset contains 48,895 entries and 16 columns initially, which is reduced to 12 columns after cleaning.

**Data Exploration:**

Steps performed:

* Used LabelEncoder and OneHotEncoder from the Scikit-learn library to encode categorical variables for the given columns:
* neighbourhood groups in New York City (5 groups: Brooklyn, Manhattan, Queens, Bronx and Staten Island)
* neighbourhoods in the particular groups (ranging from 0 to 220. Ex: Kensington, Harlem, Upper West Side, Chelsea, etc)
* room\_type (3 types: private rooms, entire home/apartment, shared rooms)
* Verified pre-processing results to provide a clean and efficient dataset for analysis and visualization

**Data Visualisations :**

**Price Analysis:**

* Scatterplots were used to compare factors affecting pricing.
* Violin plots were employed to show price distribution across room types.
* Room types -

Entire home/Apartment

Private room

Shared room

* Graphs depicted price distribution across neighborhoods.
* Bar graphs illustrated price frequency distribution.
* Box plots were utilized to identify pricing outliers, supported by scatterplots investigating outliers with minimum nights and price.

**Customer Feedback Analysis:**

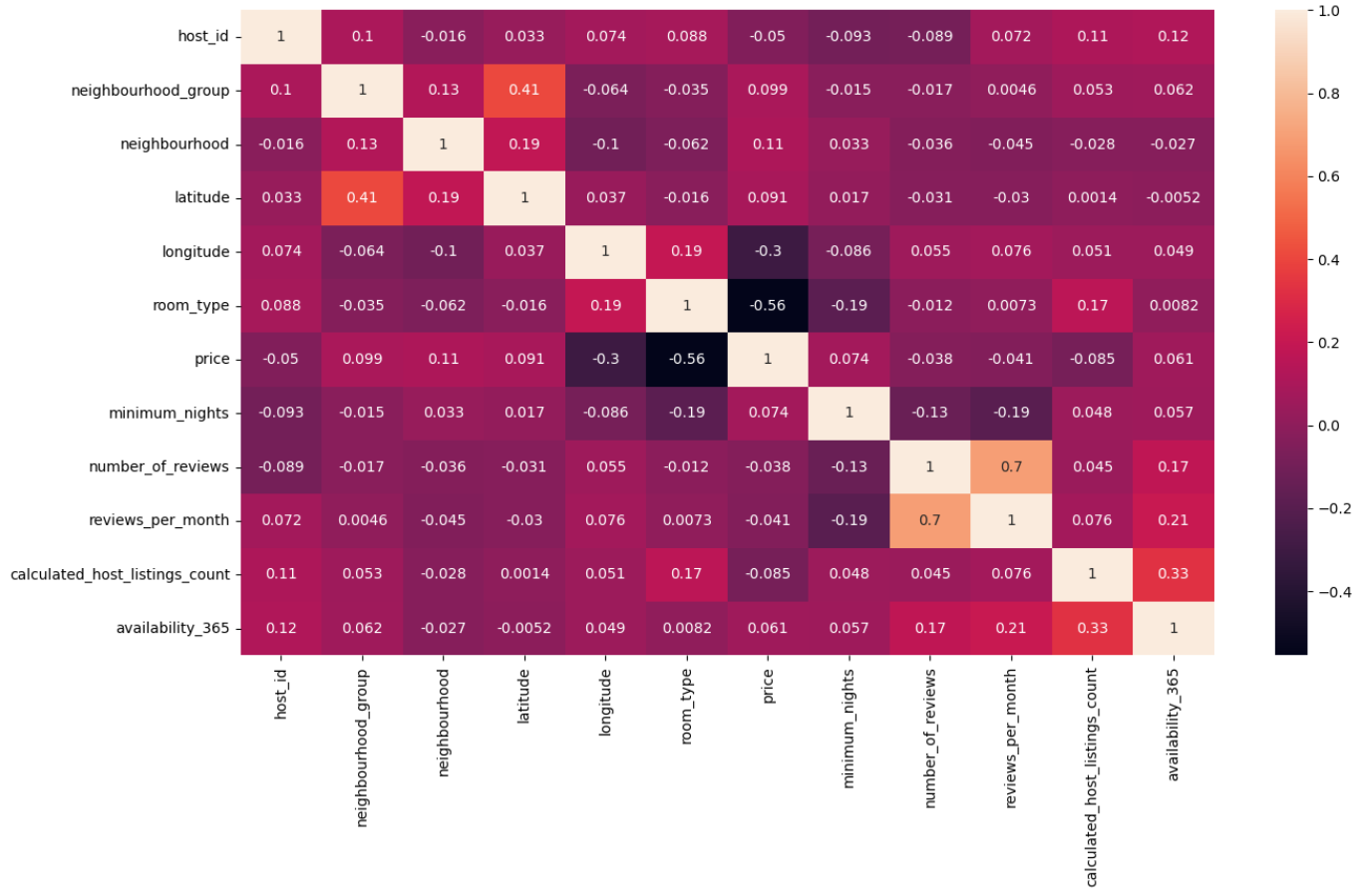
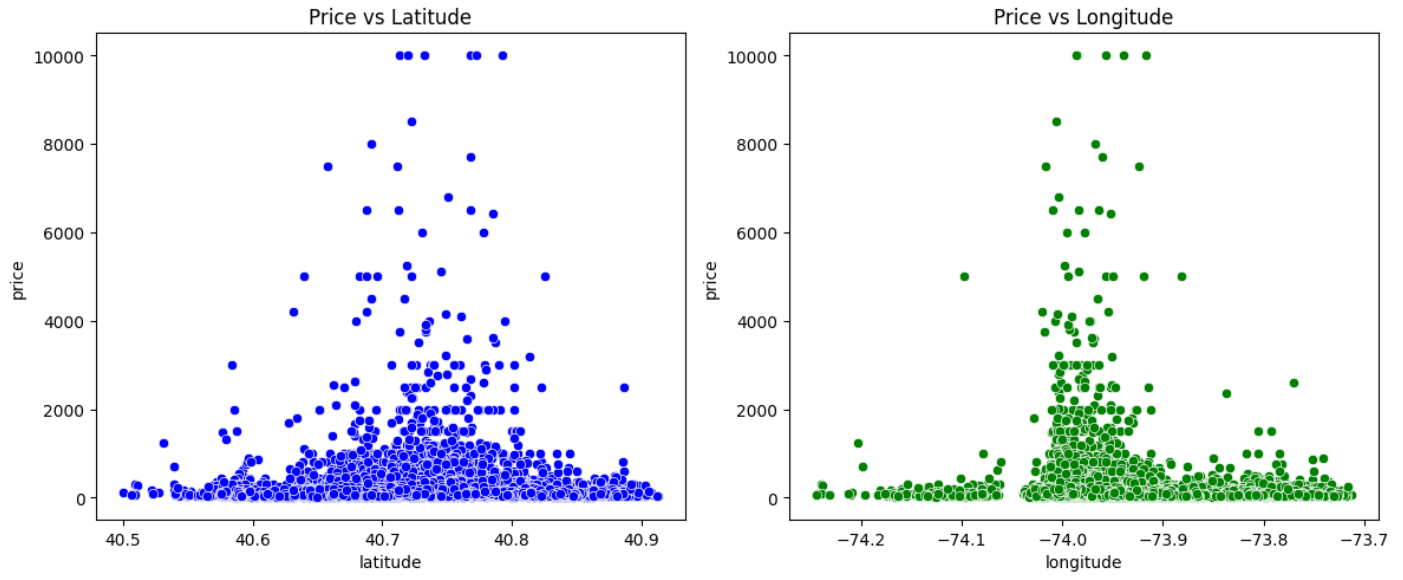
* Bar plots presented average number of reviews and reviews per month.
* Scatter plots compared number of reviews and average reviews per month.
* Pie charts indicated proportion of room types, offering insights into customer preferences.
* Bar plots revealed bookings per room type.
* Stacked bar graphs displayed room type distribution across neighborhoods.
* Pairplot was employed for comprehensive comparisons across different factors.

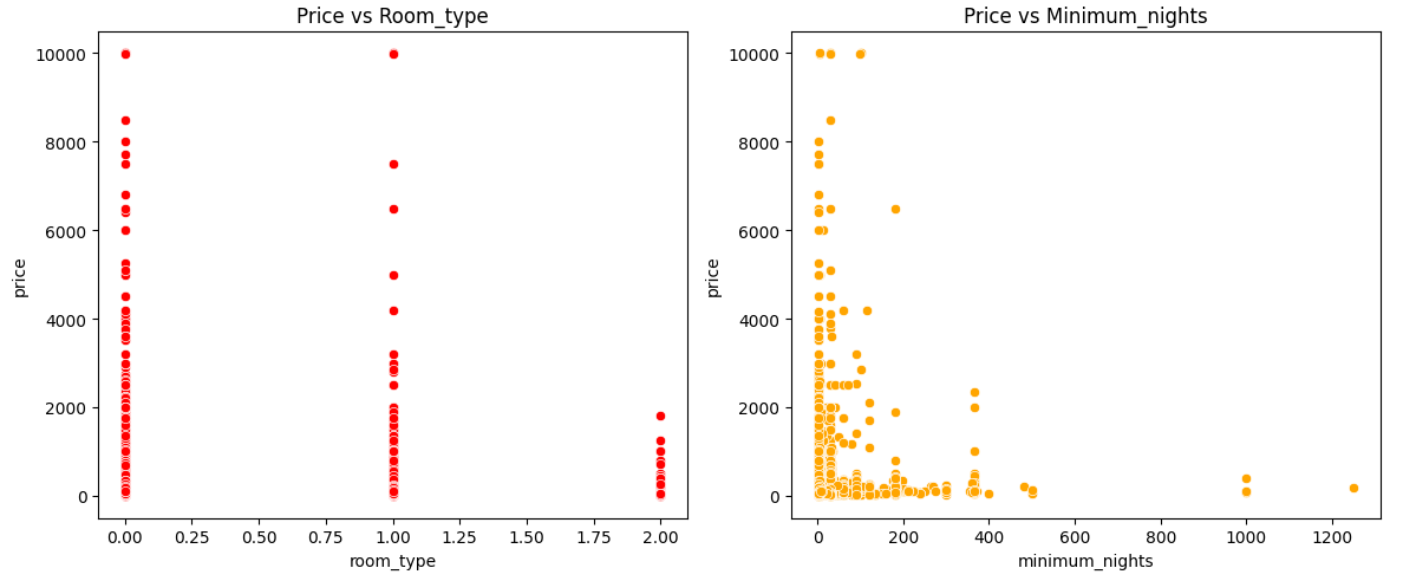
**Host Analysis:**

* Histograms showed listing performance of hosts.
* Scatterplots illustrated the relationship between host activity and listing performance metrics.

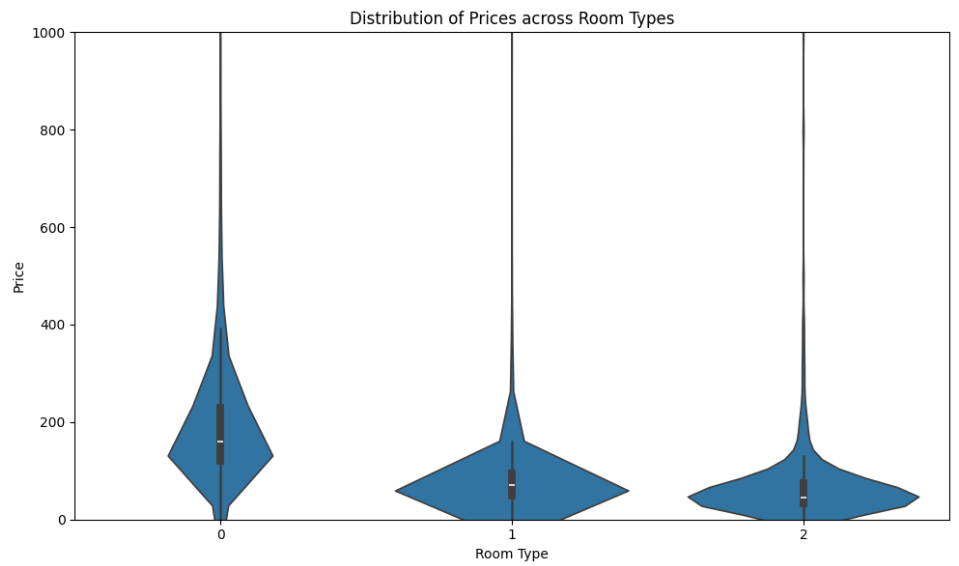
**Feature Engineering:** Label encoding is applied to categorical variables like ‘neighbourhood\_group’, ‘neighbourhood’, and ‘room\_type’. The ranges for encoded features are also provided.

**Exploratory Data Analysis (EDA):**

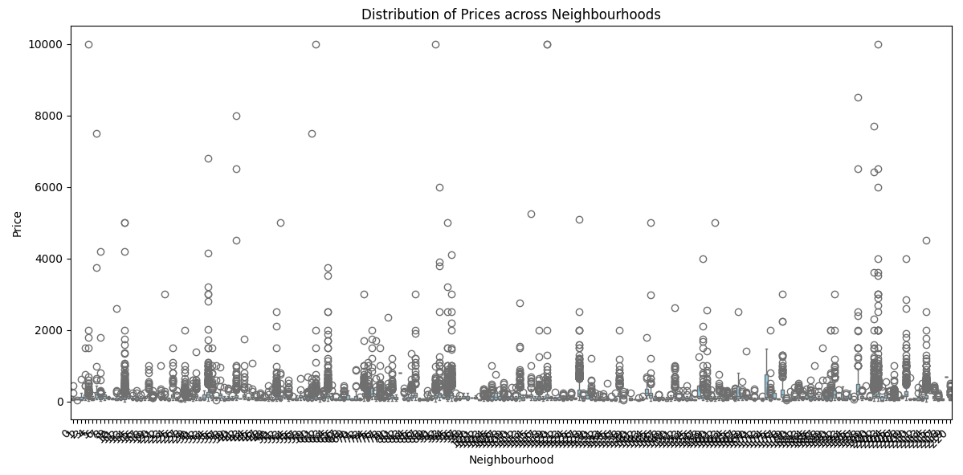
1. Heatmap
2. Scatterplots to visualize factors influencing pricing



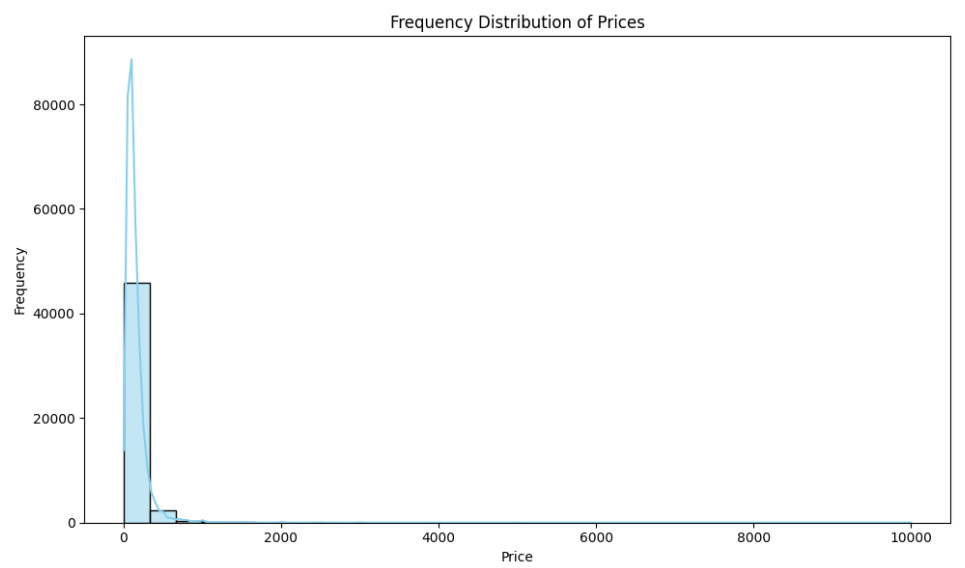
1. Comparing the distribution of prices across different room types using Violin Plot



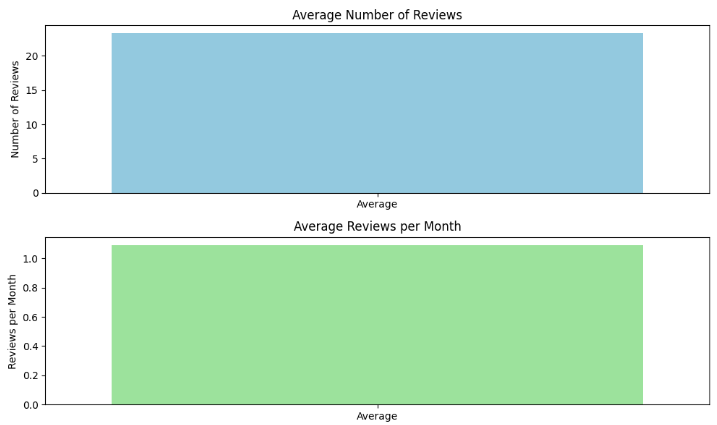
1. Distribution of prices across neighbourhoods using Boxplot



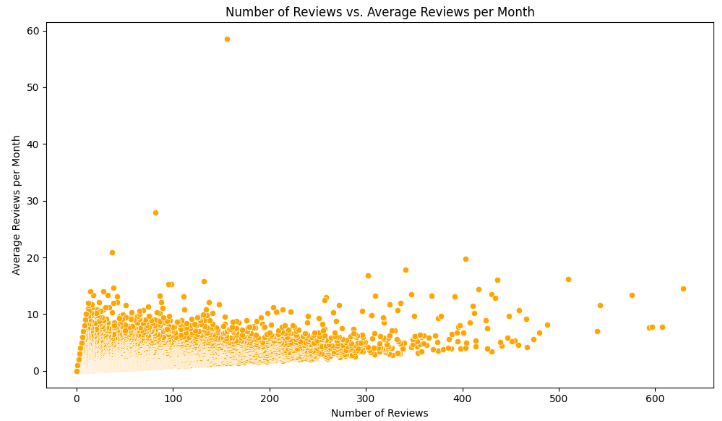
1. Frequency distribution of prices using Histogram



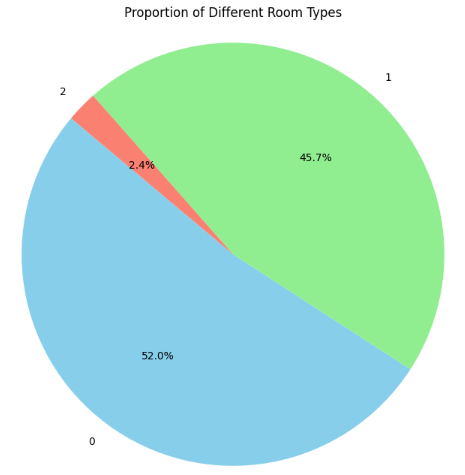
1. Average number of reviews using Barplot



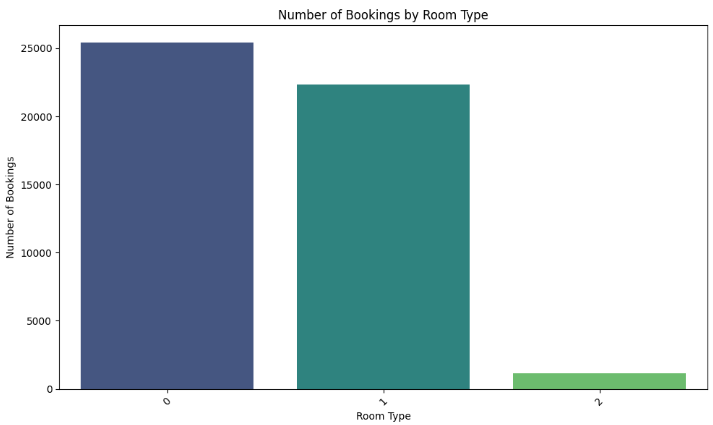
1. Scatterplot used to visualize number of reviews and average reviews per month



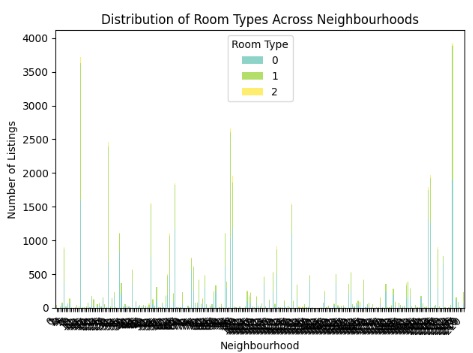
1. Pie chart used to visualize proportion of different room types



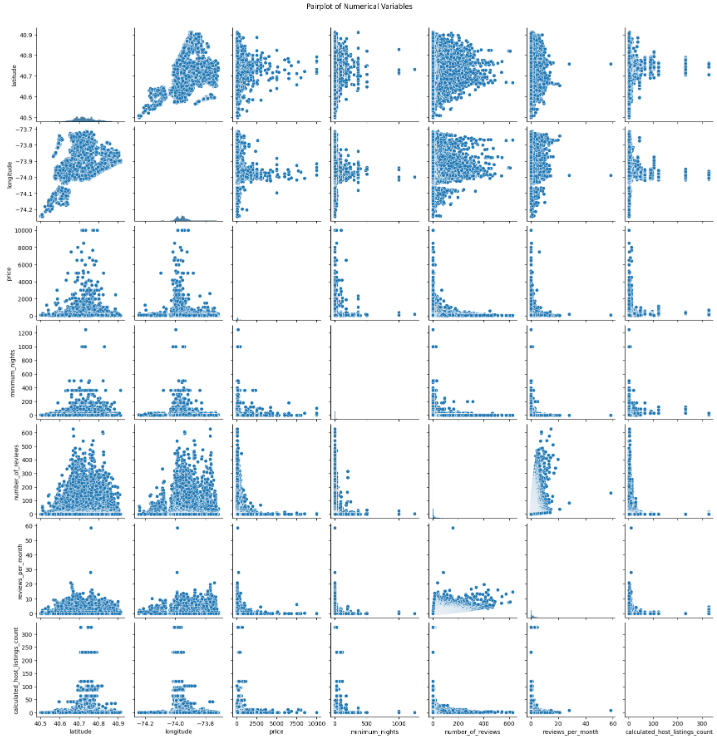
1. Barplot to depict number of bookings by room type



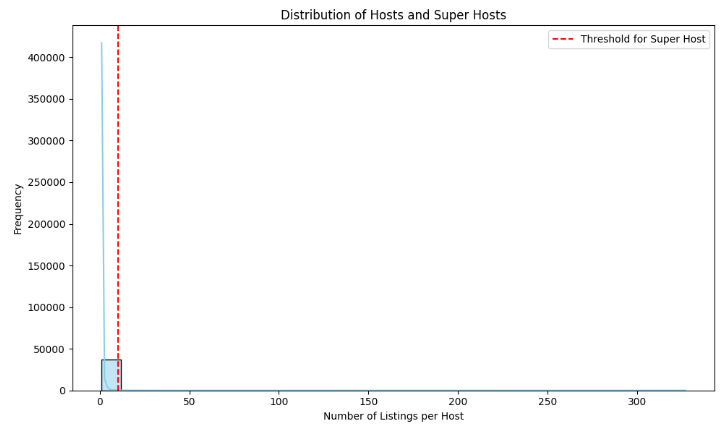
1. Distribution of room types across neighbourhoods



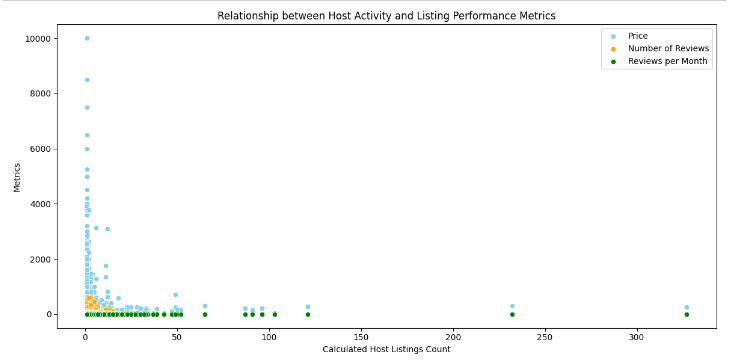
1. Pairplot of numerical variables.



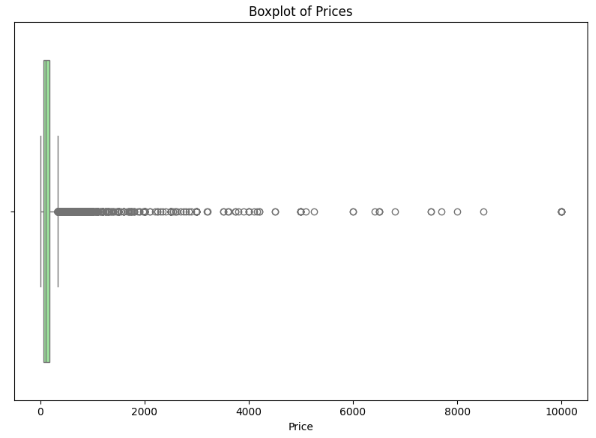
1. Histogram to depict Host distribution.



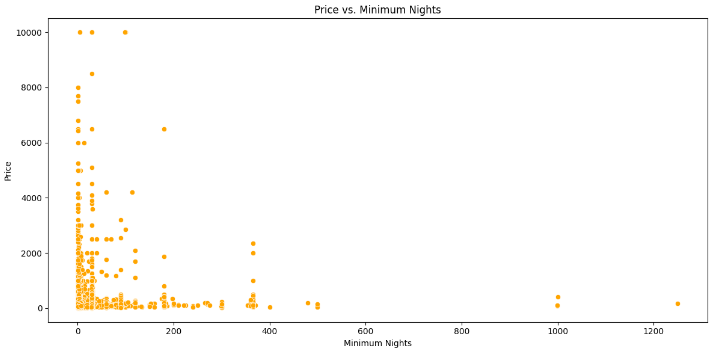
1. Scatterplot to depict relationship between host activity and listing performance metrics.



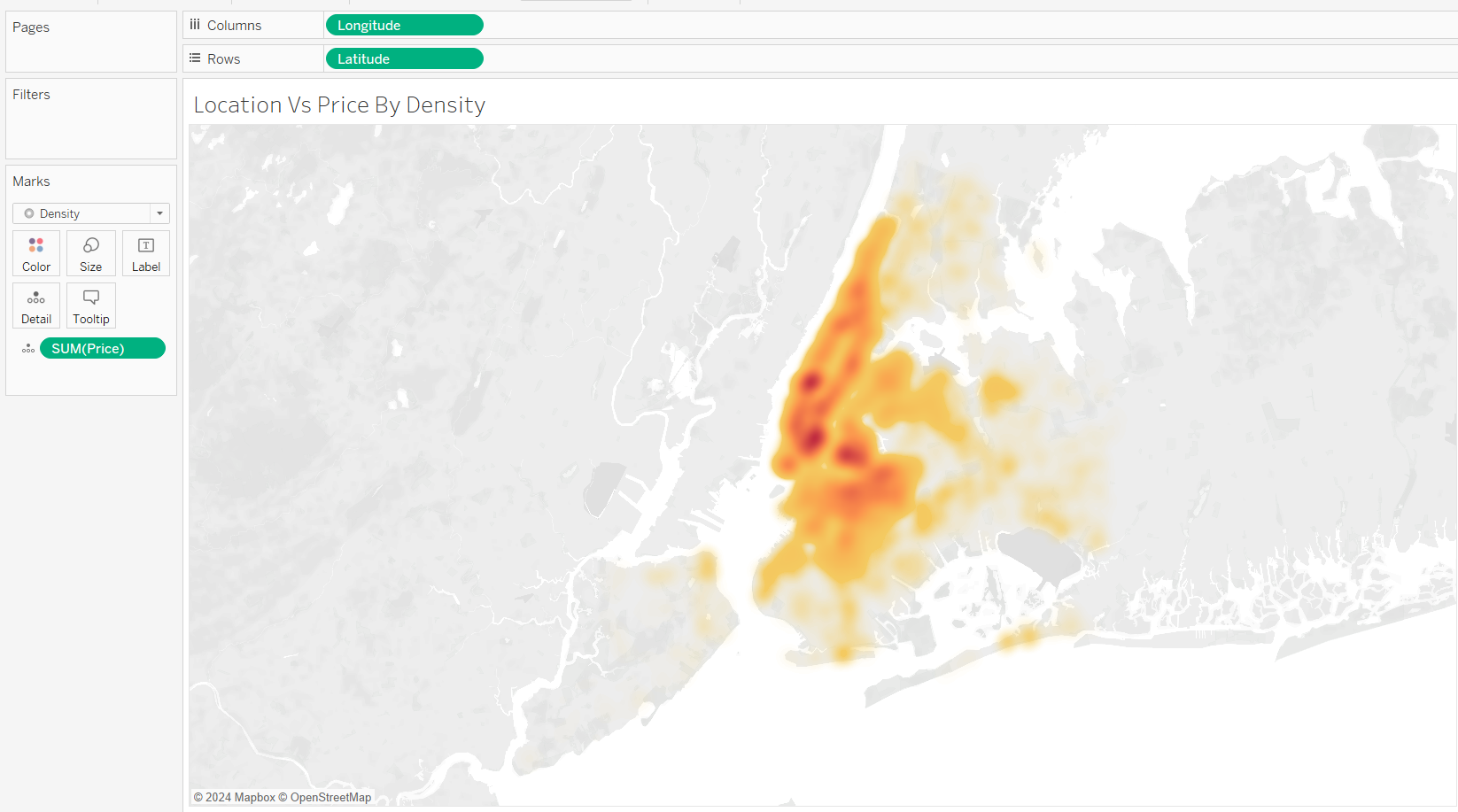
1. Gmplot to identify outliers in prices.



1. Scatterplot used to identify reason behind outliers in pricing.

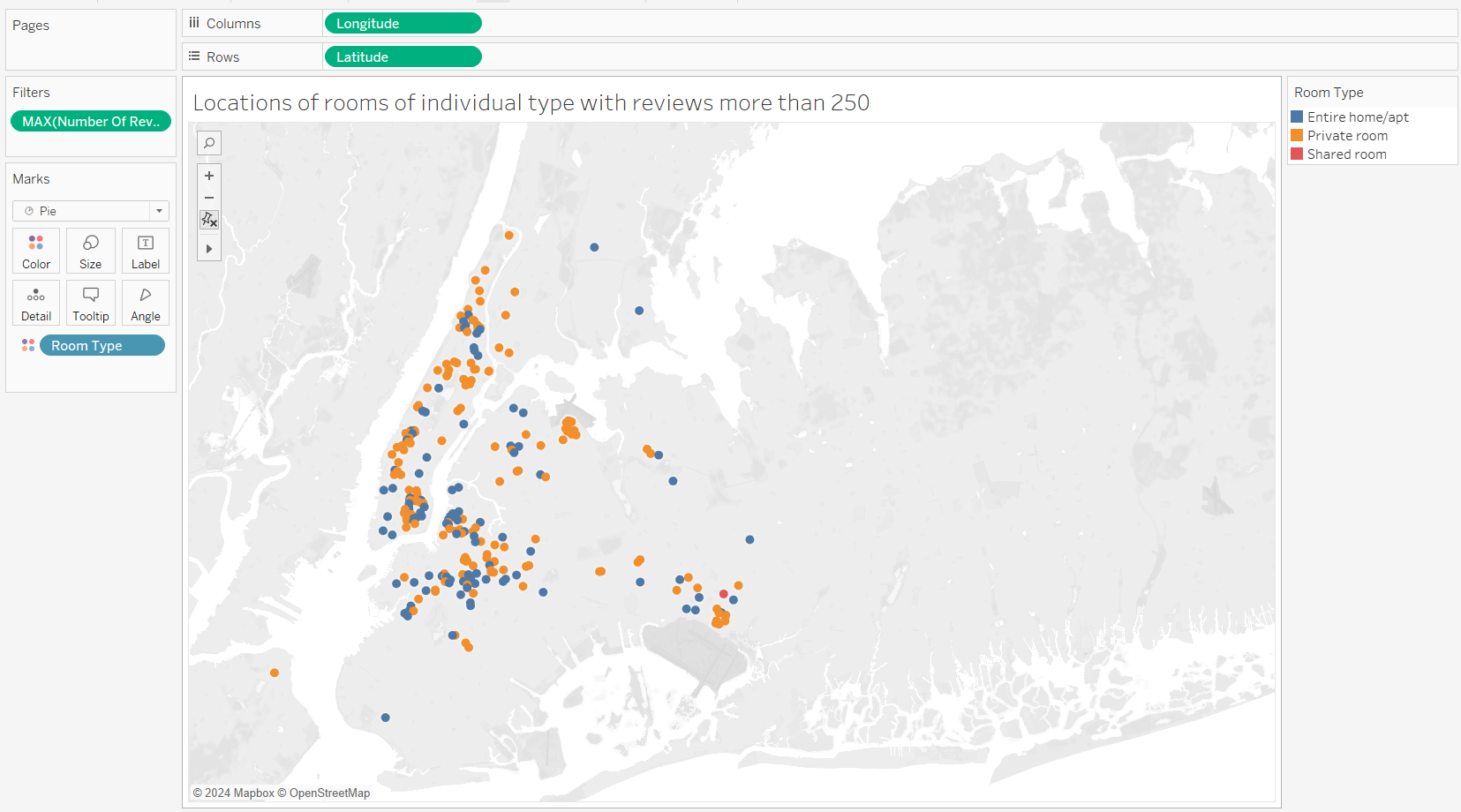


**Tableau Visualisations :**



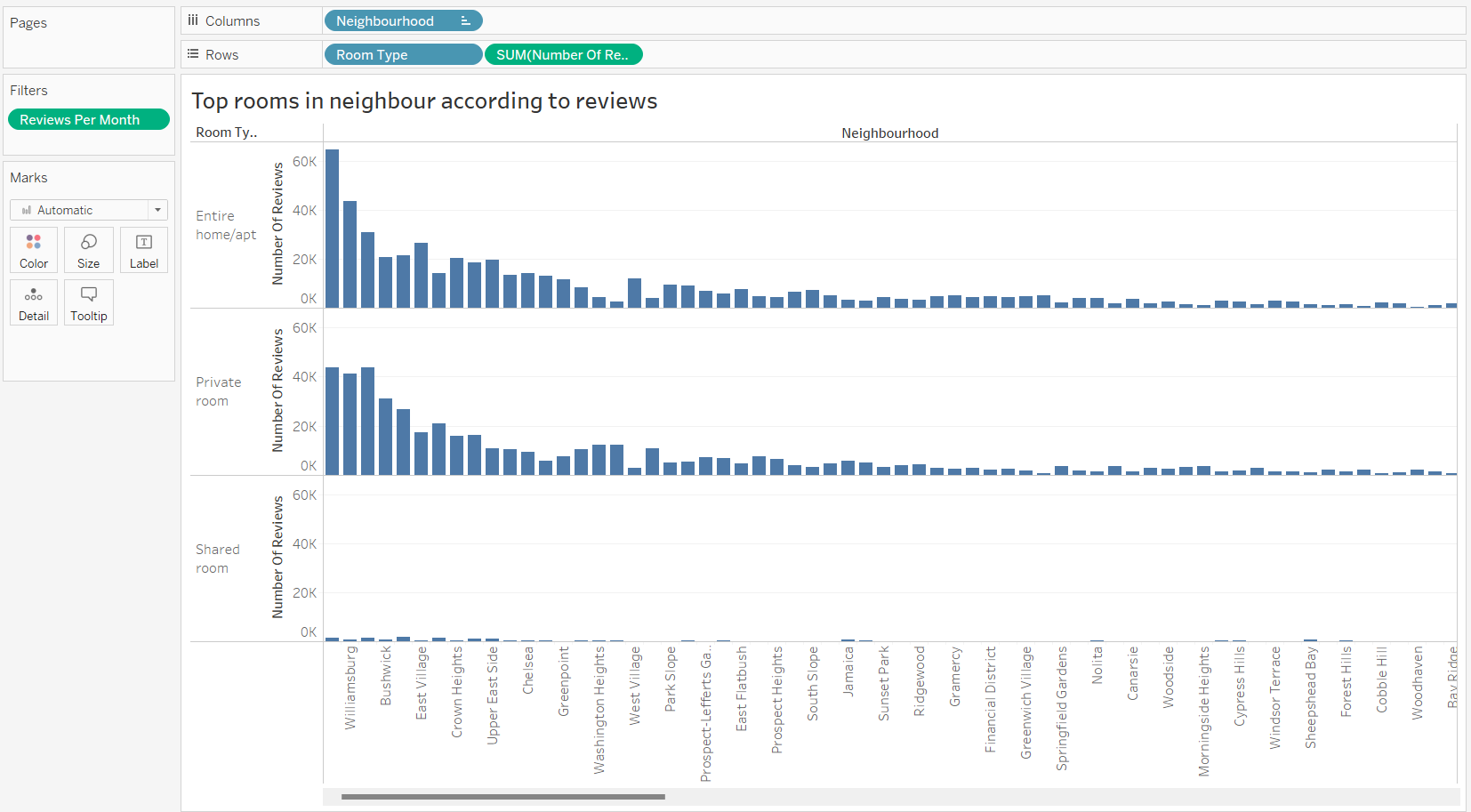
Insights:

Utilizing map visualization to identify location pricing via colour density.



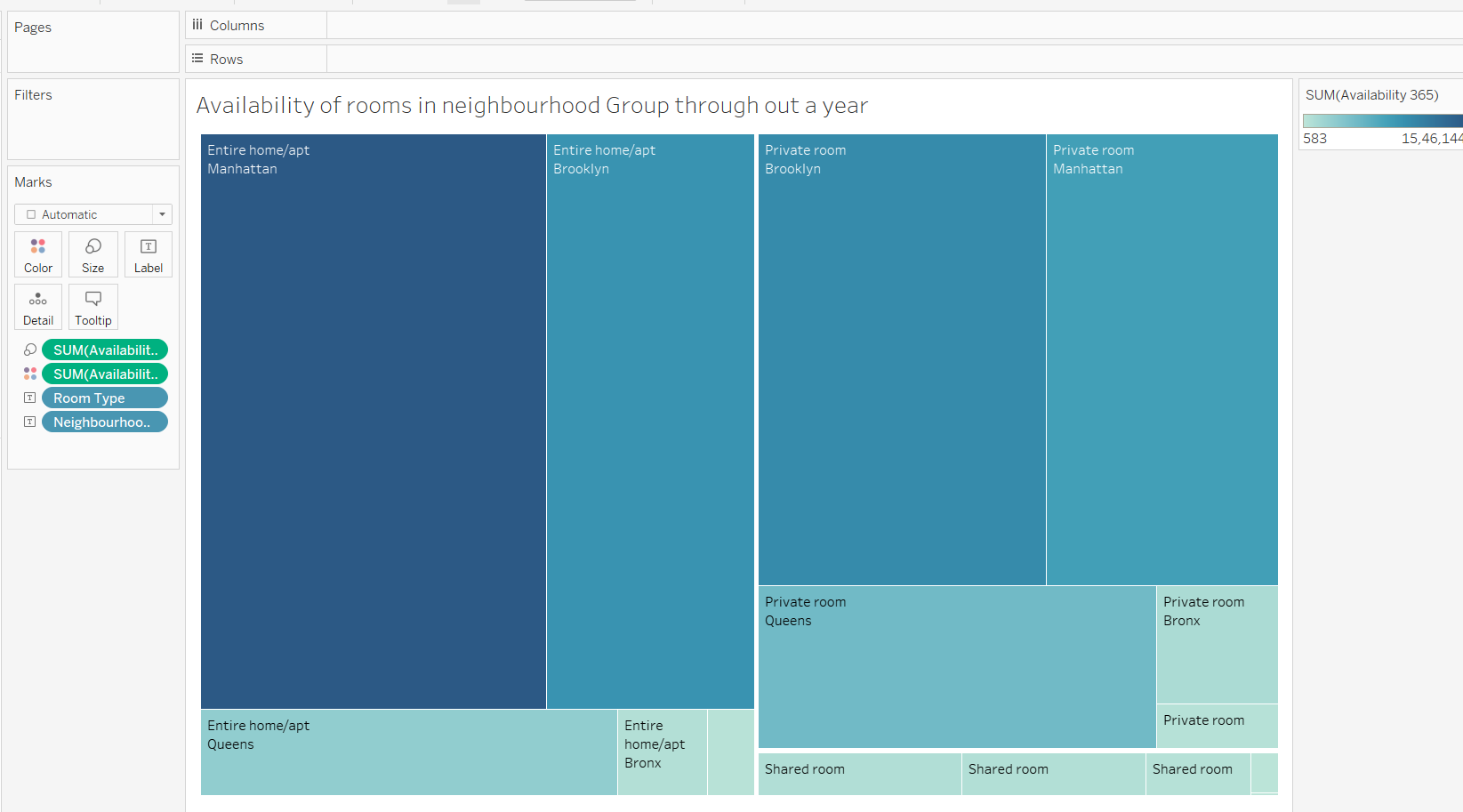
Insights:

Employing map visualization to identify room types with reviews exceeding 250 in the neighbourhood.



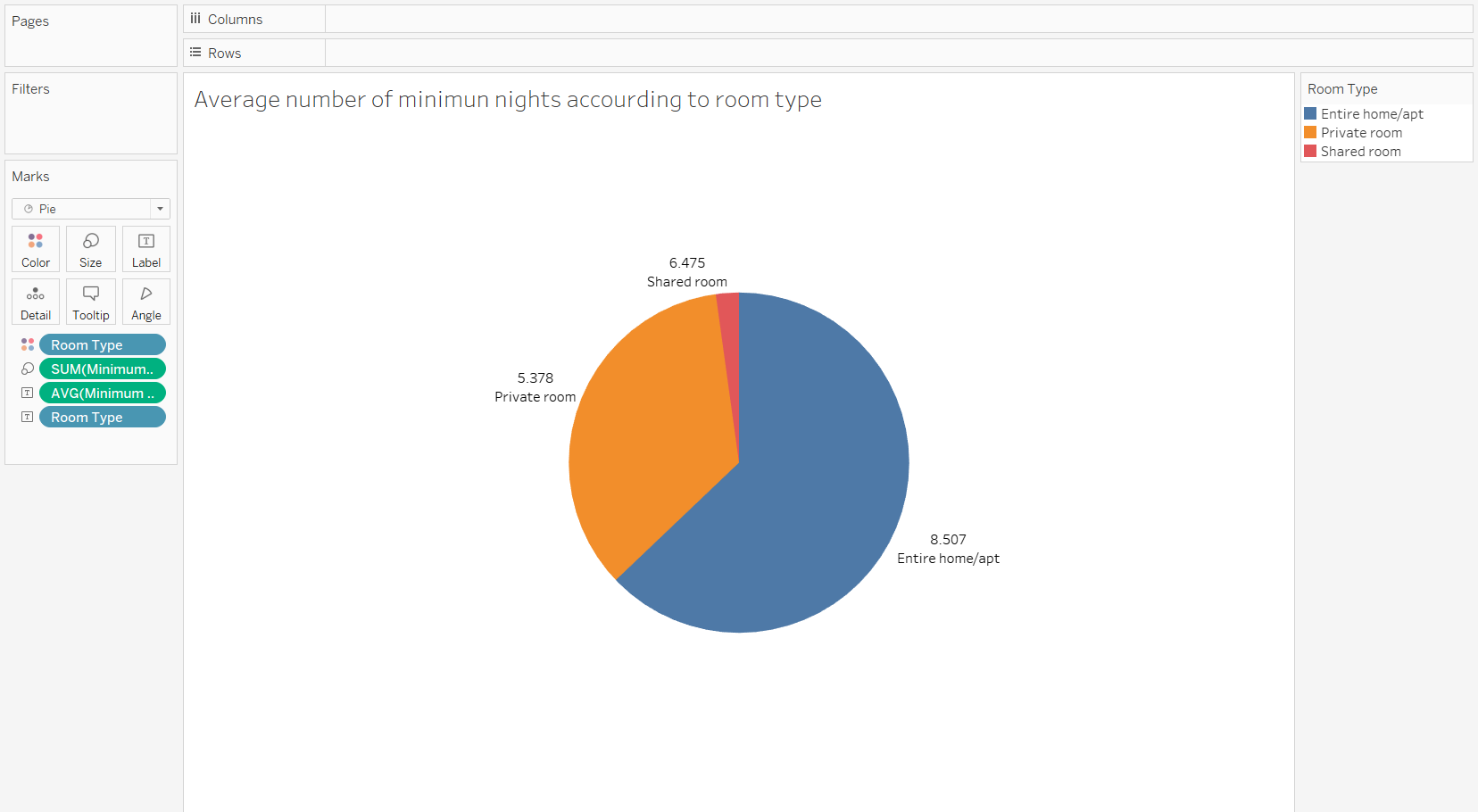
Insights:

Utilizing a bar chart visualization to showcase the top-rated room types in the neighbourhood based on reviews.

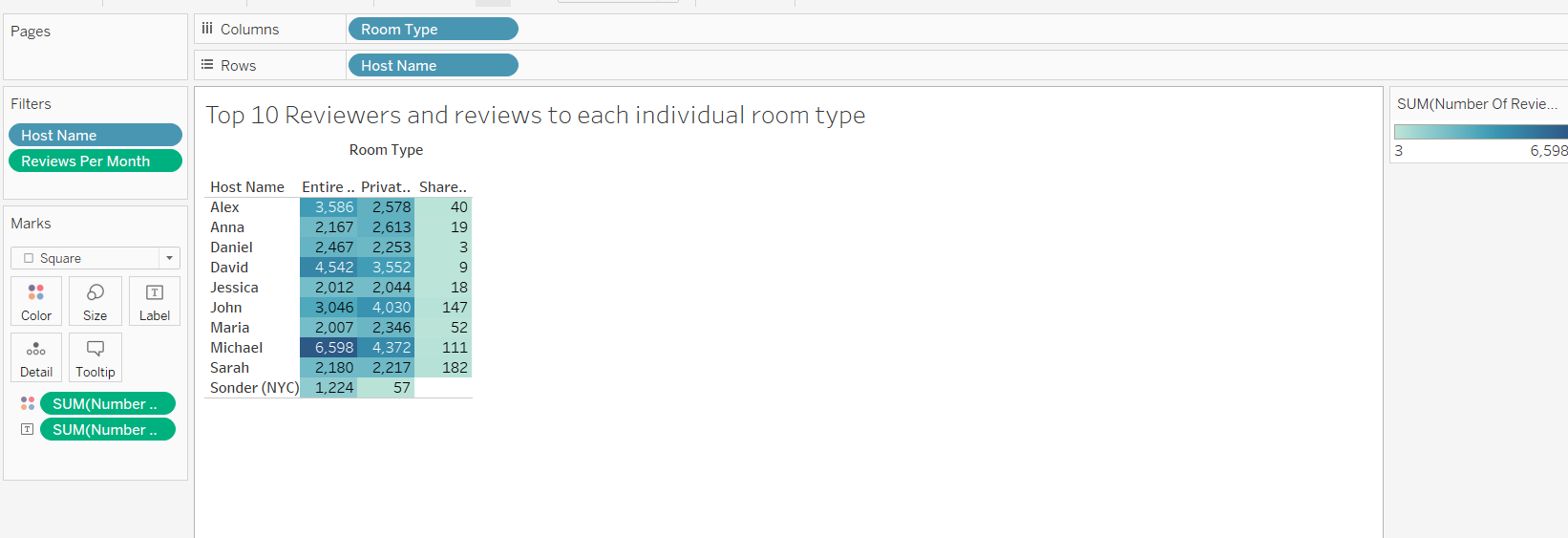


Insights:

Utilizing a heat map visualization to understand the annual availability of room types in the neighbourhood Group.

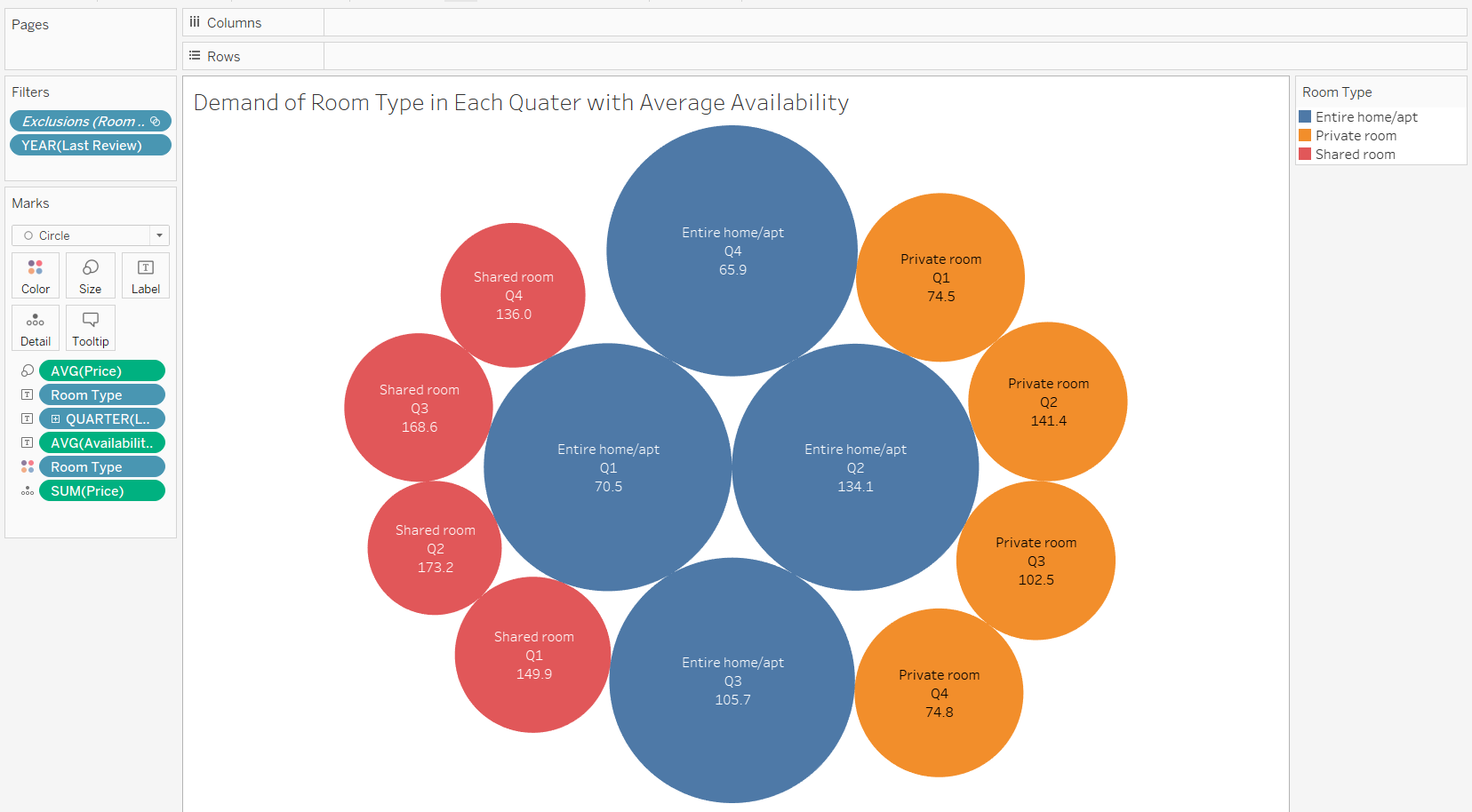


Insights:

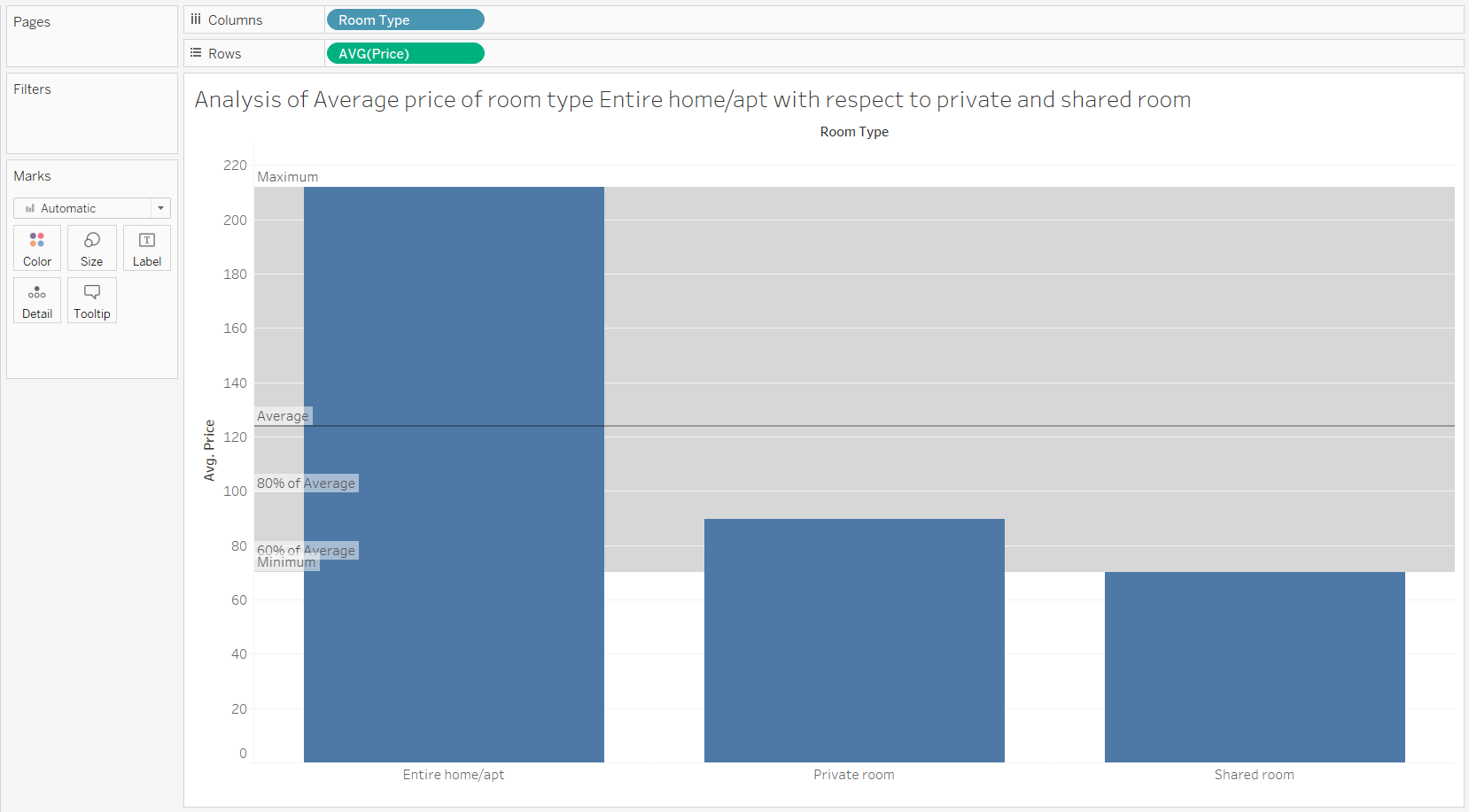
Utilizing a pie chart visualization to identify the average minimum nights per room type.

Insights:

Utilizing highlight tables to identify top reviewers and reviewers for specific room types.

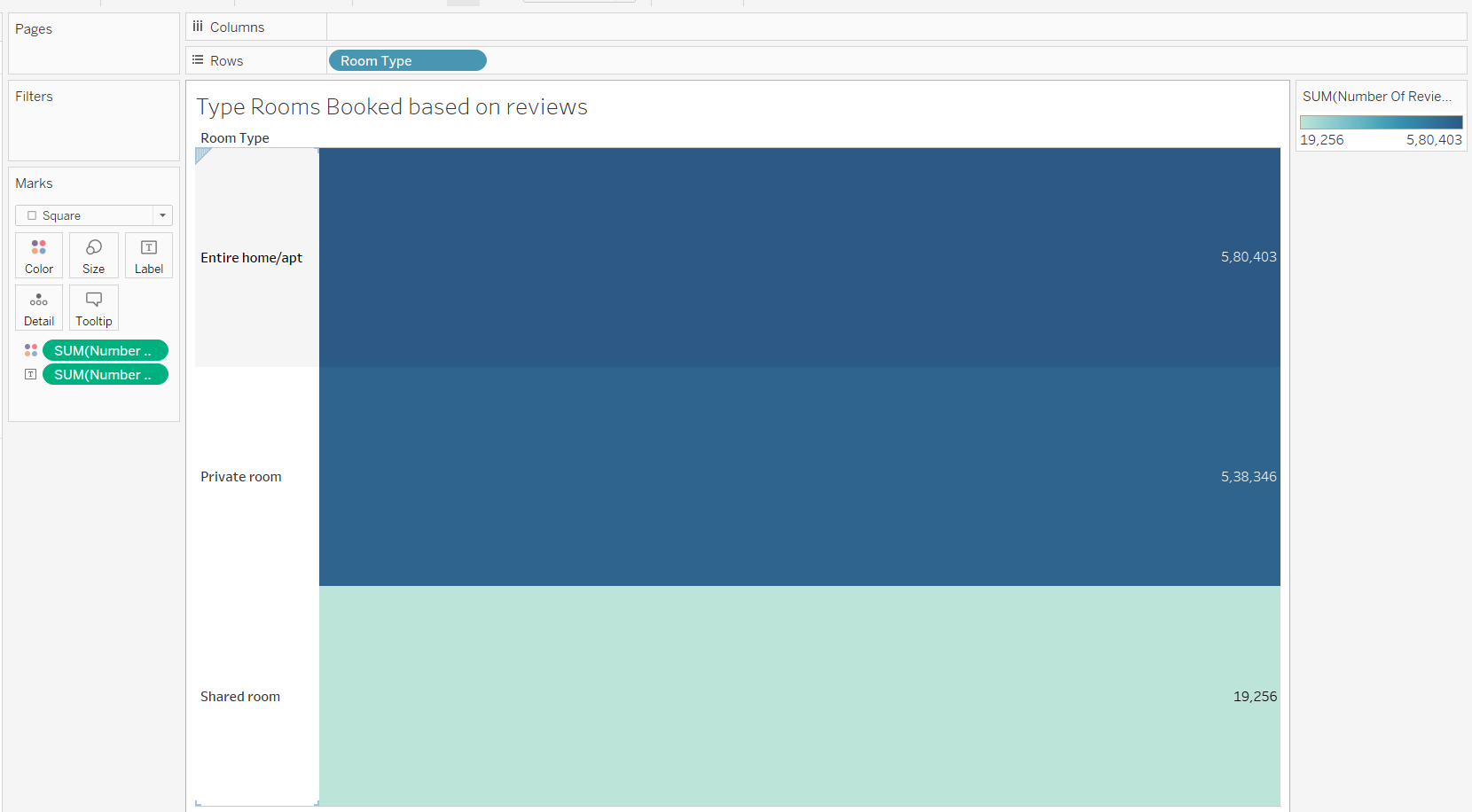


Insights:

Utilizing a bubble chart visualization to identify seasonal demand patterns for each room type based on average availability. 

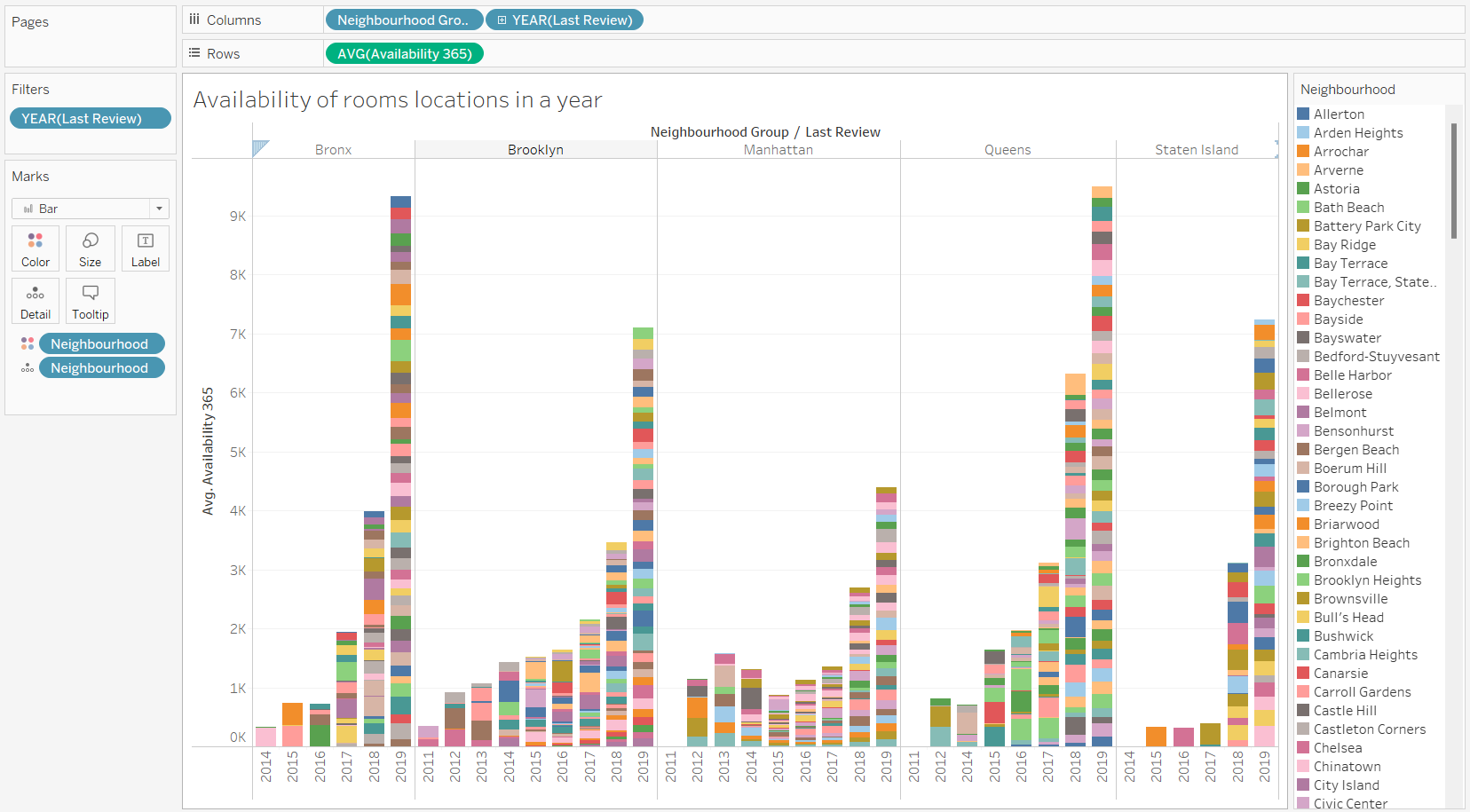
Insights:

Utilizing a bar chart visualization to derive actionable recommendations by comparing average room prices across different room types.



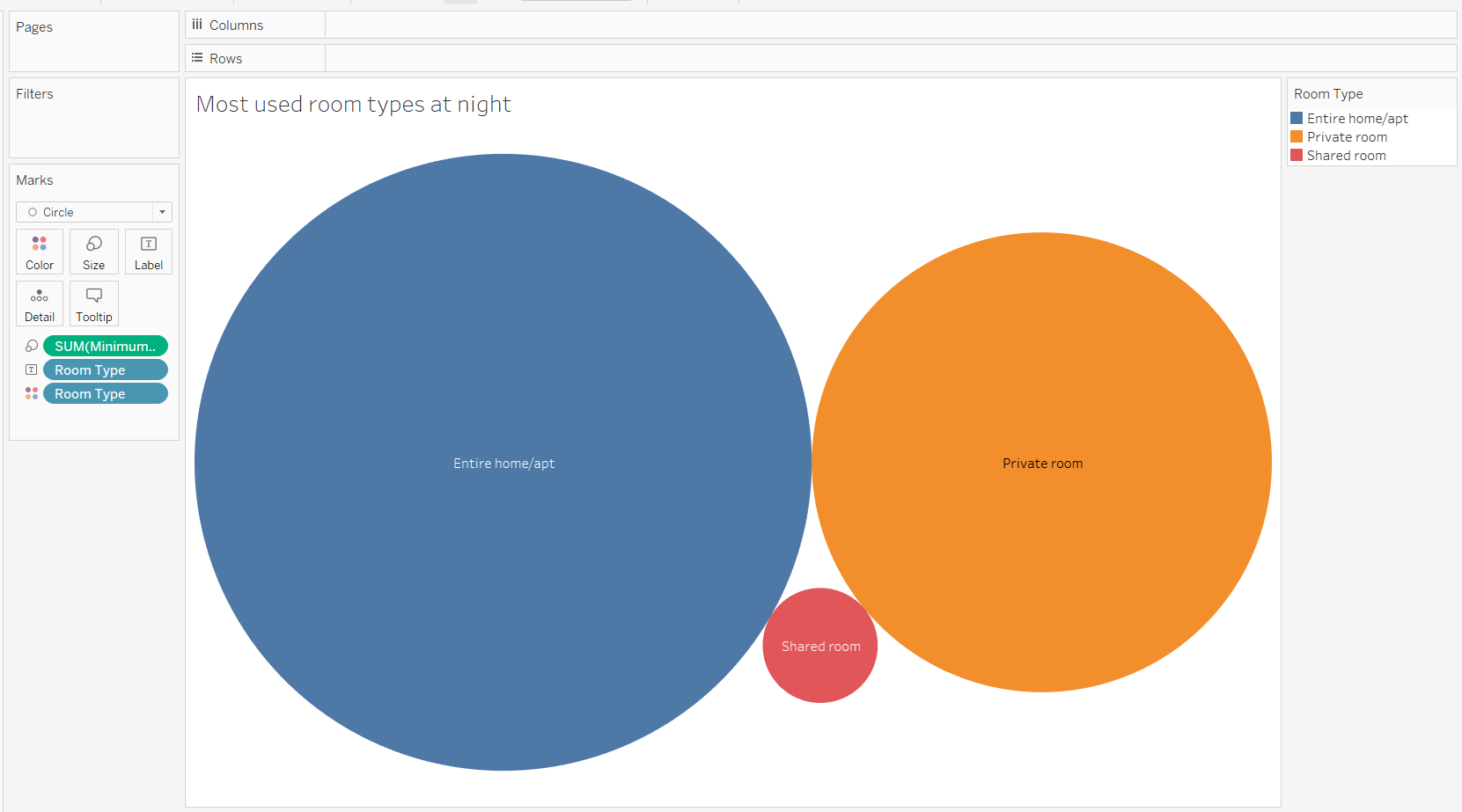
Insights:

Stacked bar chart visualization for the type of rooms that are booked the most based on reviews.



Insights:

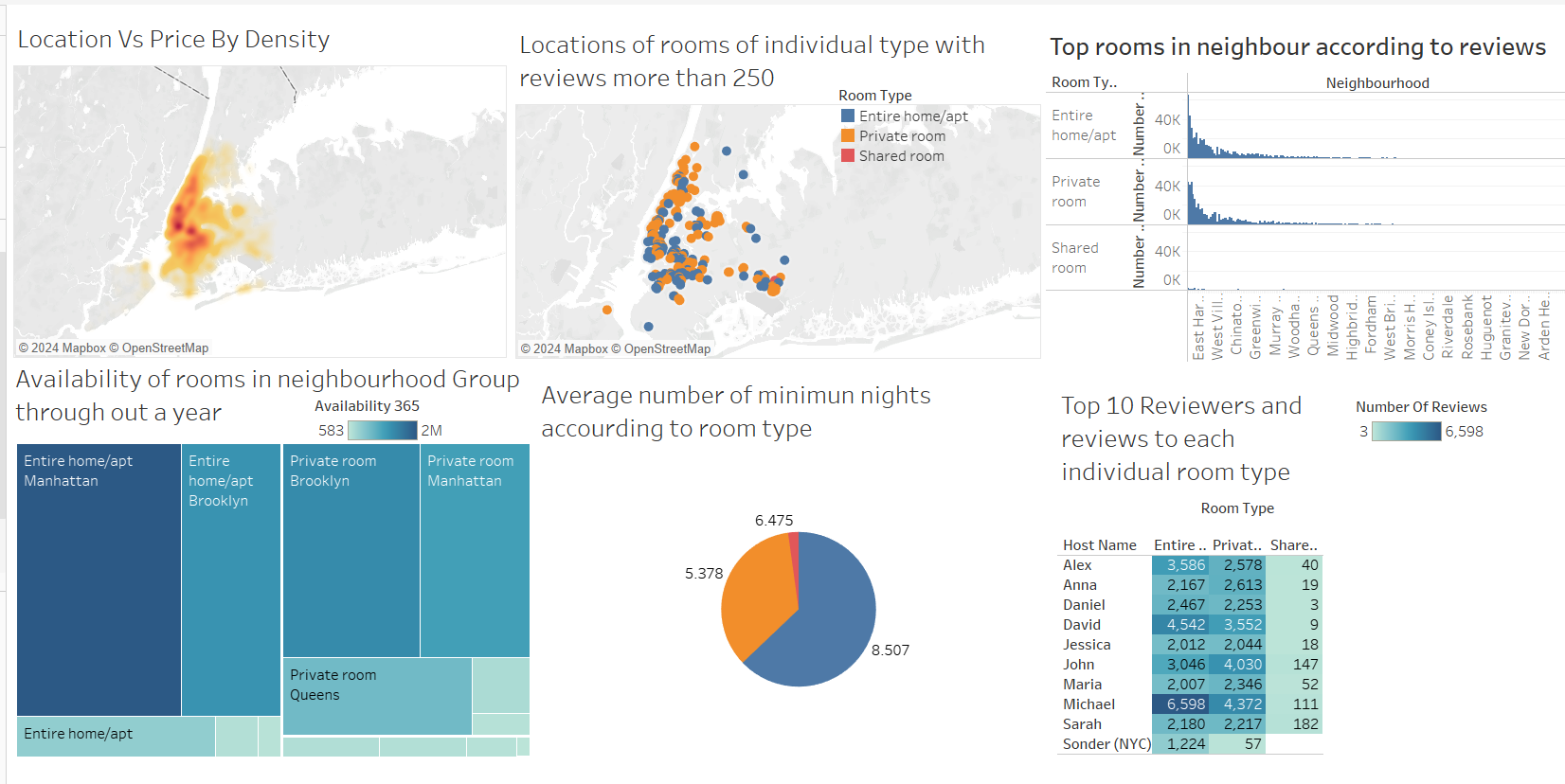
Utilizing stacked bar chart visualizations to identify the most frequently booked room types based on reviews and to assess room availability at a location throughout the year.



Insights:

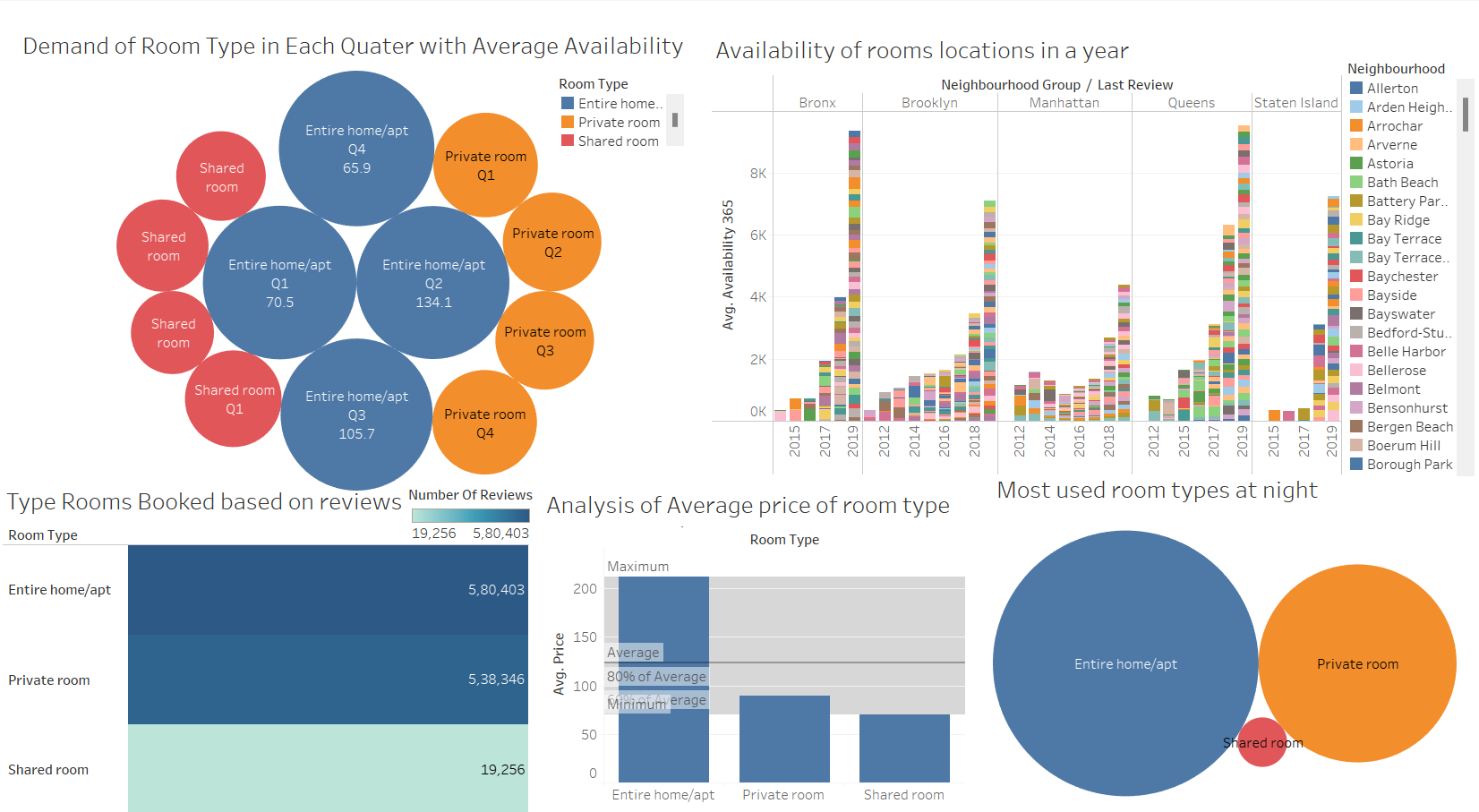
Employing a bubble chart visualization to identify the most utilized room type during nights.

**Tableau Dashboards:**

**Dashboard 1:**

Dashboard 1 Insights:

The dashboard incorporates various visualizations to obtain information insights into Airbnb listings within a neighbourhood. Through map visualizations, users can identify location pricing trends based on colour density, as well as identify room types gathering reviews exceeding 250. A bar chart highlights the top-rated room types in the neighbourhood, offering valuable insights into guest preferences. Additionally, a heat map provides a comprehensive view of room type availability throughout the year, aiding in understanding seasonal demand patterns. The dashboard also features a pie chart illustrating the average minimum nights spend per room type, enabling hosts to optimize their booking strategies. Lastly, highlight tables identify top reviewers for each specific room type, facilitating informed decision-making for hosts aiming to enhance guest satisfaction.

**Dashboard 2:**

Dashboard 2 Insights:

The dashboard presents a comprehensive analysis of Airbnb data, employing various visualization techniques to extract actionable insights. Through a bubble chart visualization, users can identify seasonal demand patterns for each room type, leveraging average availability data. Additionally, a bar chart facilitates the derivation of actionable recommendations by comparing average room prices across different room types, aiding hosts in optimizing pricing strategies. Stacked bar charts further contribute by identifying the most frequently booked room types based on reviews and assessing room availability throughout the year at a specific location. Finally, another bubble chart helps in identifying the most utilized room type during nights, providing valuable information for hosts to tailor their offerings to meet guest preferences effectively.

**Conclusion:**

In conclusion, this project has provided valuable insights into the Airbnb market through user-friendly interfaces and interactive visualizations. It is evident that Entire home/apartment room types are highly preferred by guests, while Shared rooms are less favoured, particularly during nighttime. These findings suggest potential areas for Airbnb to capitalize on and expand its market share, such as focusing on increasing the availability of Shared rooms and implementing dynamic pricing strategies to incentivize bookings during off-peak periods. By leveraging these actionable recommendations, Airbnb can enhance its competitiveness and better meet the diverse needs of its customers in the ever-evolving hospitality industry.

**Results and Insights:**

* This project aims to provide user-friendly interfaces for exploring Airbnb data.
* The most preferred room type is Entire home/apartment followed by a Private room.
* Shared room types are least preferred by customers especially at night.
* Insights and analysis in the Airbnb market are presented through interactive charts and visualizations.

**Actionable Recommendations :**

* Airbnb offering Entire home/apartment room types are attracting guests and getting sufficient reviews. In comparison, Shared room availability is less. This area can be potentially attractive for Airbnb to capitalize on and expand its market share.
* Special promotions and incentives to be provided during off-peak periods to stimulate demand.
* To overcome pricing outliers dynamic pricing strategies can be implemented to incentivize bookings for the less-utilized room types like the shared rooms.