

Data Methodology

Step 1: Storyboarding

- Reviewed the dataset thoroughly to understand its structure and key variables.
- Identified significant fields that would be essential for analysis.
- Created a visual representation to present Insights and Recommendations.
- Developed a basic outline for the slides.

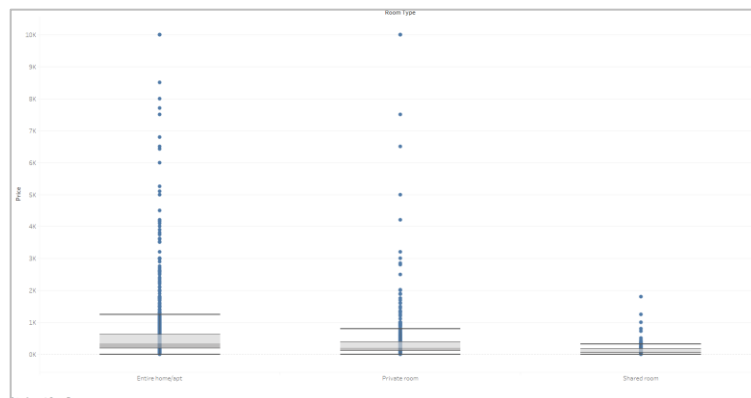
Step 2: Data Wrangling

Univariate Data Analysis:

- Used Tableau to analyse individual variables.
- Examined distributions, unique values, missing data, and potential outliers.

Data Quality Assessment:

- Identified a small percentage of missing values that would not significantly impact the analysis.
- Observed a right-skewed distribution for the price variable.
- Noted the presence of outliers in the price data, as indicated by the boxplot.



- Created a calculated field for grouping the no of nights the customer has booked to understand the distribution



- Created a calculated field of for reviews per host listing using the below formula to understand the spread of reviews

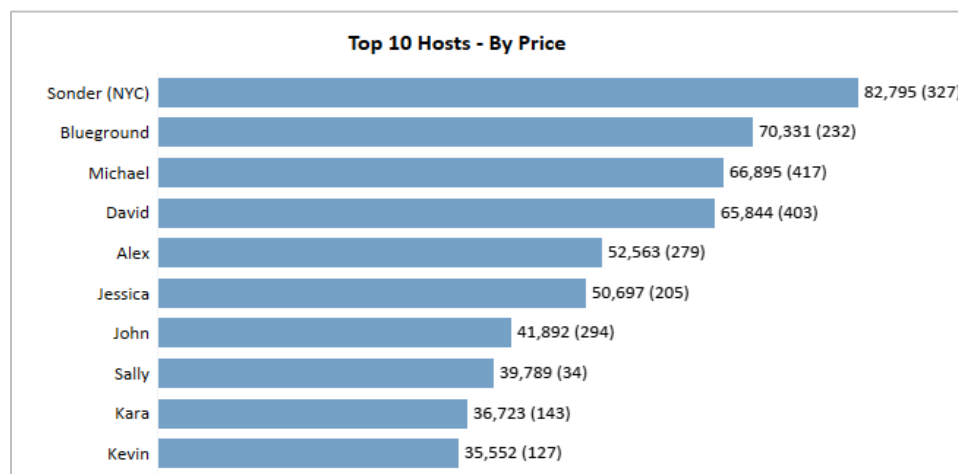
Reviews Per Host Listing

$$\text{Sum}([\text{Number Of Reviews}]) / \text{COUNT}([\text{Calculated Host Listings Count}])$$

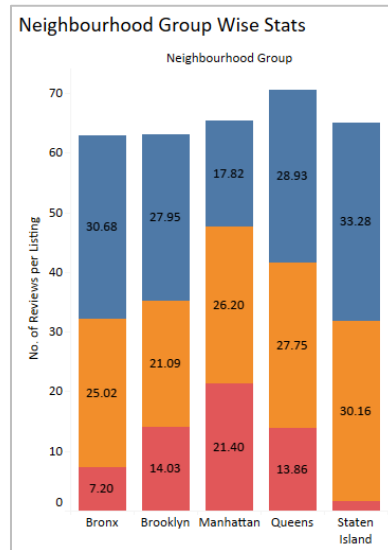
The calculation is valid.
2 Dependencies
Apply
OK

Step 3: Data Analysis

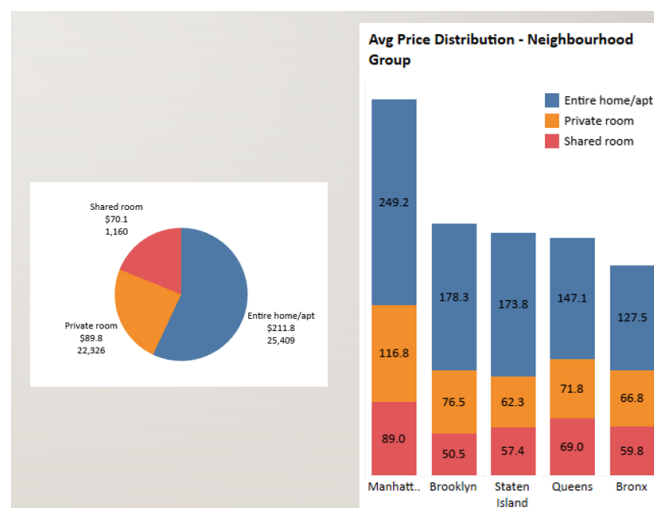
- Create a Bar chart to understand the Top 10 hosts by Sum of Price to understand the impact of the customers and how they are faring against the number of bookings
 - Sonder (NYC)**, despite being the most expensive option, has the highest number of bookings (327). This suggests that customers are willing to pay a premium for Sonder's offerings, possibly due to its location, amenities, or reputation.
 - Michael** and **David**, on the other hand, have a relatively high number of bookings compared to their pricing, indicating they may offer good value for money.
 - While **Sonder (NYC)** commands the highest price, it's worth noting that other hosts with fewer bookings are charging similar rates. This suggests that pricing may not be the sole determinant of popularity.
 - However, there are exceptions, such as **Sally** and **Kara**, who have a relatively low number of bookings despite having a decent number of reviews. This could indicate that these hosts may have recently joined the platform or have a niche market that limits their booking potential.



- Distribution of price and room type by Neighbourhood group



- Created a chart to understand the pricing of each room category against different neighbourhood group
 - Average price of entire Home/Apt is the highest and approx. 100% more than private rooms. For Brooklyn and Staten Island, the average price varies by 150%
 - Private room's price is approx. 10% more than shared rooms except Brooklyn and Manhattan approx. 40%
 - Manhattan is costliest overall and the cheapest are:
 - Entire apt: Bronx
 - Private room: Staten Islands
 - Shared room: Brooklyn



- Insights from above graph:
 - Entire apartments are popular and so are private rooms
 - Except in Manhattan, shared rooms are very unpopular
 - Entire room's reviews per listing for Manhattan is 35% lower than the overall average and number of listings is the highest

Step 4: Presentation

- Made the presentation adhering to best practices and pyramid principle
- Added recommendations for the respective departments