**Airbnb Data Methodology**

**Data Cleaning**

**Dropping the features**

* Dropped “last\_review” feature as it was not needed when we do have “reviews\_per\_month” feature to understand the popularity of listings
* Dropped “host\_name” feature as we are not targeting host names for customer preferences
* Dropped “name” feature as it gives listing names so not relevant in the analysis



**Missing Value Imputation**

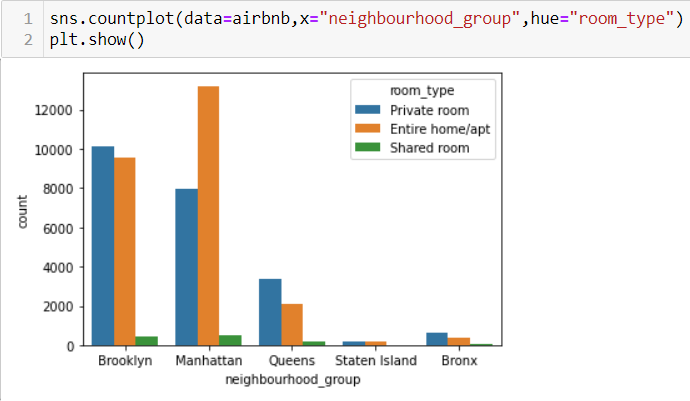
“reviews\_per\_month” feature has 20% null values that is assumed to be 0 review per month so imputed nulls with 0



**Data Analysis**

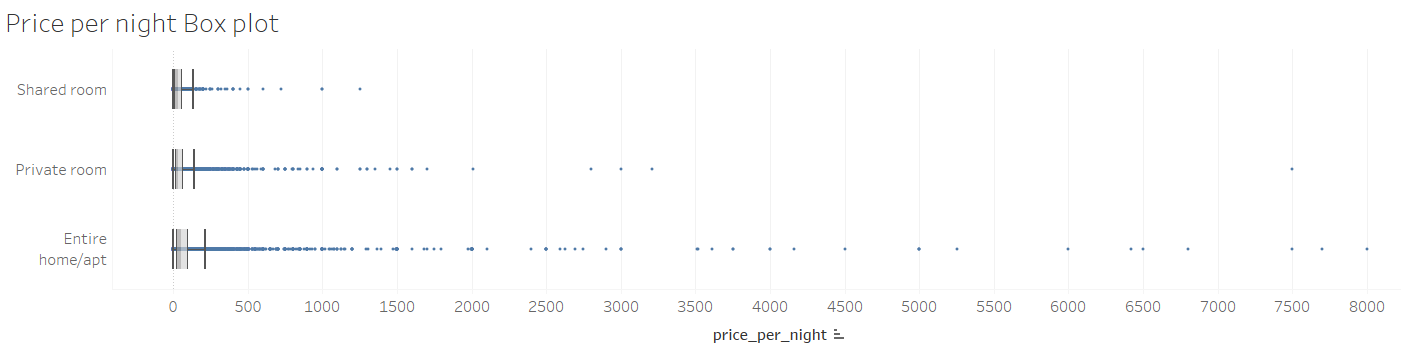
**Univariate Analysis of Continuous variables**

Analyzed Listing count per neighborhood across all types of room

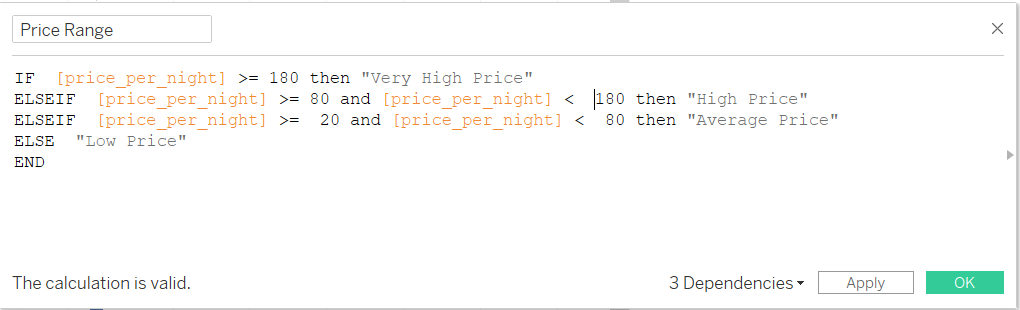


Analyzed price per night:

Created a derived variable to understand price per night using price and minimum nights variables from the data set. It would be fair measure for listing price insights.

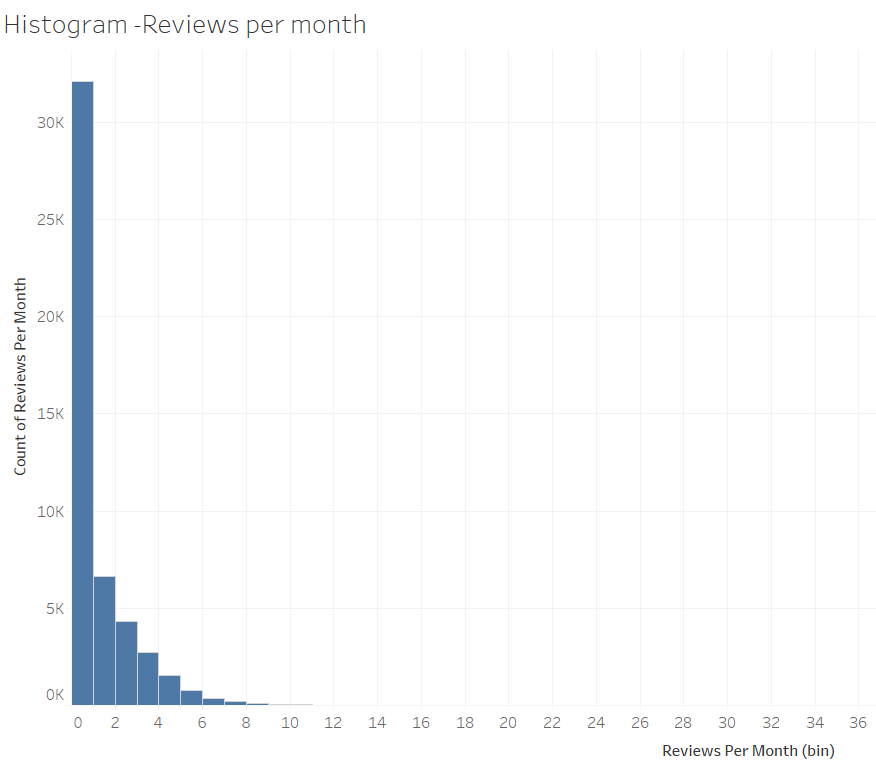


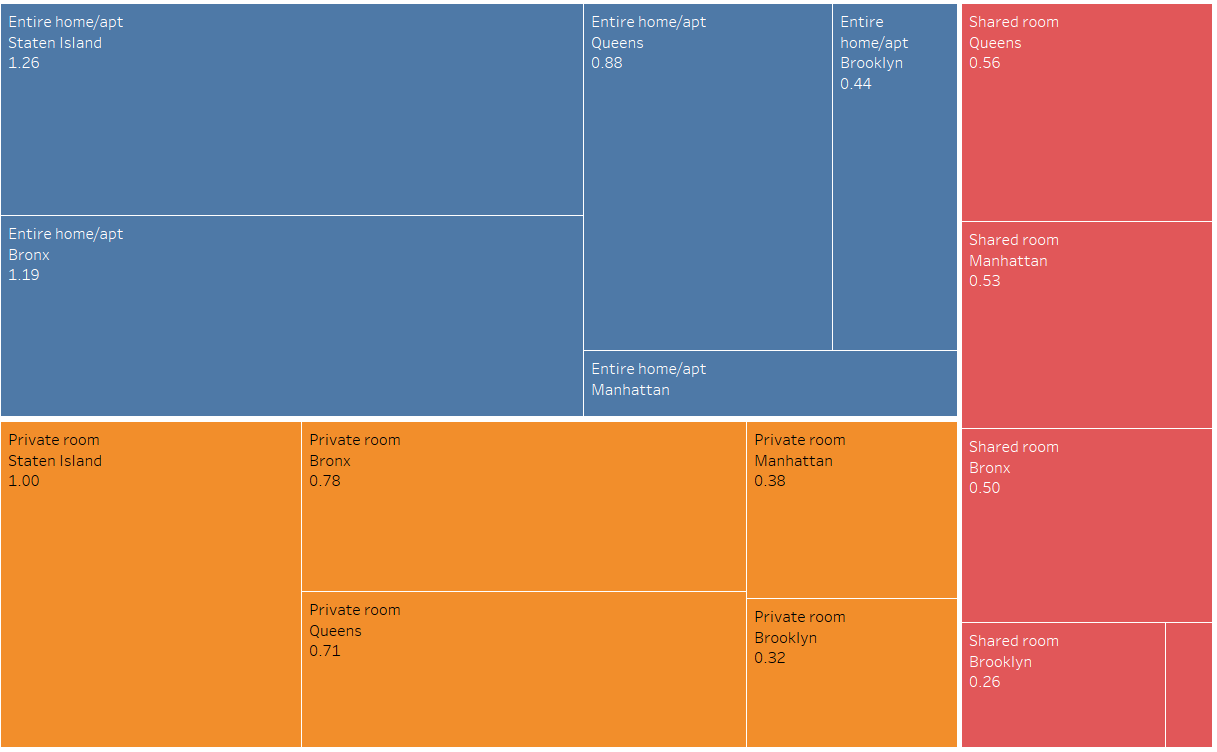
Price variable is highly skewed and observed outliers across all room types, so we have binned the price for further analysis:



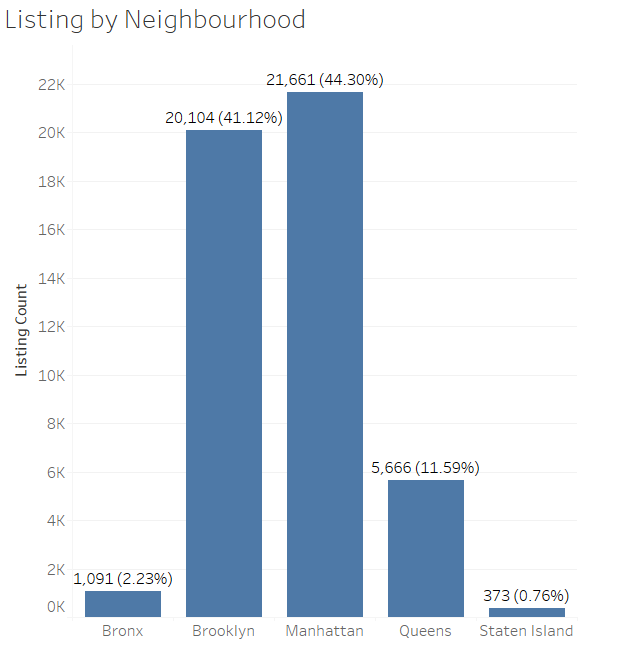
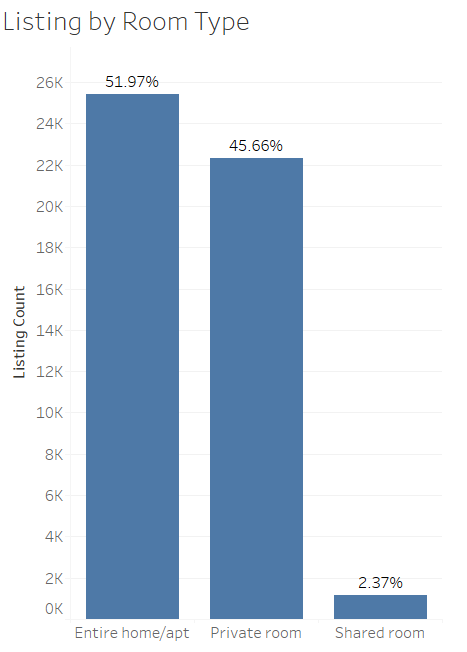
Analyzed reviews per month

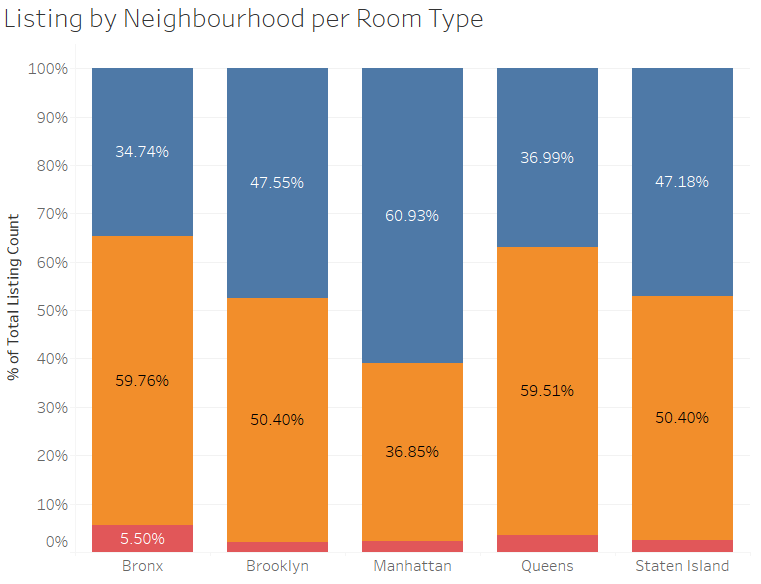
In this case study, it is assumed that the all the reviews are positive reviews. Reviews per month can give good view to measure popularity throughout the year.



This is right skewed variables however median is used to measure popularity of listing. 

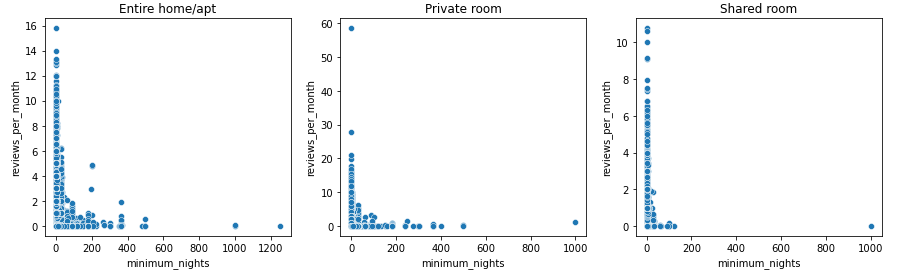
**Univariate Analysis of categorical variables**

Analyzed listing distribution by Neighborhood groups and room types



**Bivariate Analysis**

Scatterplot is used to understand the relationship between reviews per month and minimum night stay across all room types.



Tool used:

**Python** and its visualization libraries like seaborn and matplotlib and **Tableau** are used for data visualization and data preprocessing techniques.