**New York City Airbnb Open Data**

**Visual and Analytical report on the Airbnb listings and metrics in NYC, NY, USA (2019).**

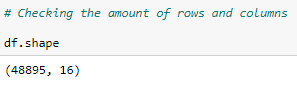
Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019.

The analytics and visualizations are performed in Python programming language using packages like **NumPy**, **Pandas**, **Matplotlib,** and **Seaborn**.

1. **Importing necessary libraries and the dataset.**



1. **Let us take a look at the number of rows and columns of the dataset**

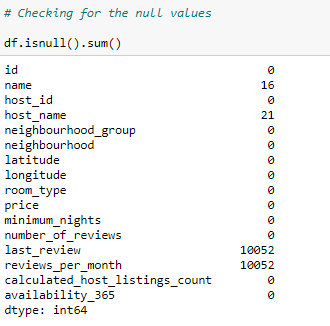


Looking at the dataset we find that the data is contained in **48,895** **rows** and **16 columns**. df.head() function is used to visualise any pandas dataframe’s first 5 rows.

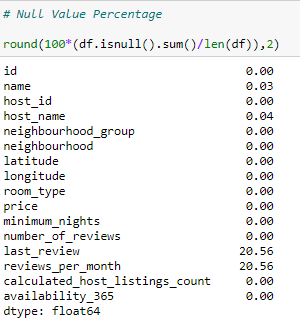
More rows can be visualised by

adding the specific number into the parenthesis i.e., df.head(15) {This will give the first 15 rows of any data frame}

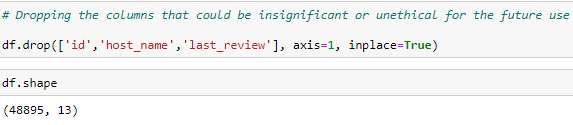
1. **Finding columns with null values in the data frame.**

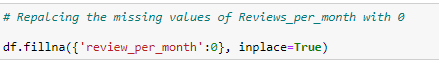


df.isnull().sum() is the function which displays the **sum of total null or NaN entries** in the dataframe. Here, we find that the columns ‘name’, ‘host\_name’, ‘last\_review’, and ‘reviews\_per\_month’ have null or NaN values. These null values need to be removed before further analysis..We have also calculated the percentage of Null Values.



1. **Missing Value treatment**



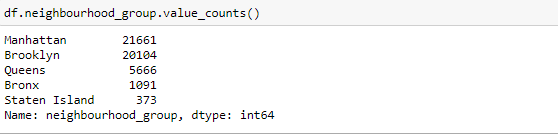


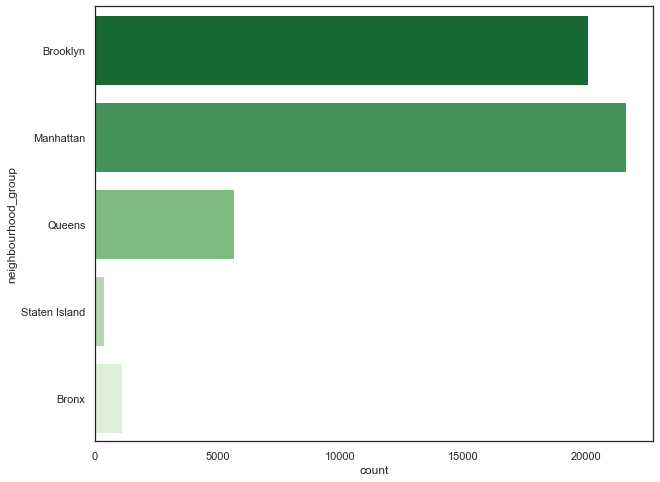
We have **removed** the insignificant columns with respect to our analysis like ‘id’,’host\_name’, ‘last\_review’.

As the reviews are customer-dependent, it's insignificant to **replace** the missing values with the mean or median values that's why replacing them with 0 (zero).

**Analysis of the Neighbourhood\_group column**

1. **The number of neighbourhood groups and their count.**

****

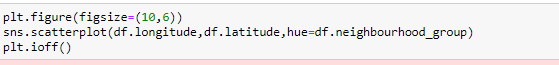


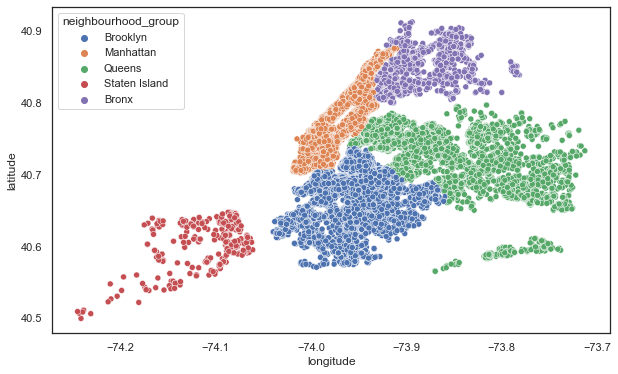
As evident from the count plot, Manhattan has the highest number of Airbnb holdings in **New York (21,661)** followed by **Brooklyn (20,104)**, **Queens (5,666)**, **Bronx (1,091)** and **Staten Island (373)**.

Visualising the neighbourhood groups using libraries like **Seaborn** and **Matplotlib** are very handy and can help in building great dashboards and reports. There are several palette’s available for count plot, one should choose it wisely. Also, we get an option to resize and title the plot according to the requirements.

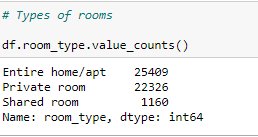
The dataset also provides us with the latitudes and longitudes of the place where each holding is located. This can be used to plot a **Scatter plot** which enables us to visualise the data geospatially.

We infer from the **Scatter Plot** that Manhattan is not the largest parts of New York by area but it has the highest number of Airbnb holdings suggesting that it is a popular tourist place and has huge number of visitors round the year.



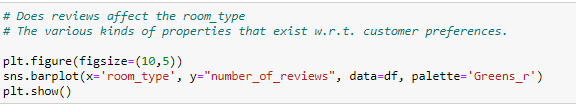


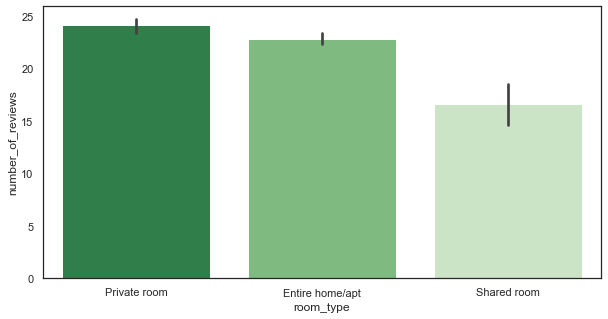
1. **Analysis of Room types in the Airbnb holdings**

****

The rooms are bifurcated into three types: **Shared room (1,106)**, **private room (22,326)**, and **entire home/ apartment (25,409)**.

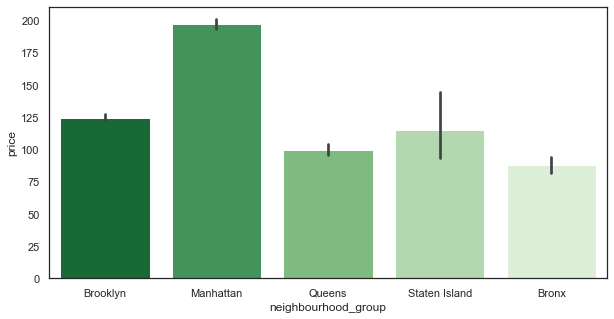
We find that the hosts prefer offering the entire home and private room rather than shared spaces. This can be due to larger and private spaces can yield better incomes for the hosts. Also, managing the costs of utilities for people in shared spaces and other complexities can be a reason for hosts avoiding shared space offerings.



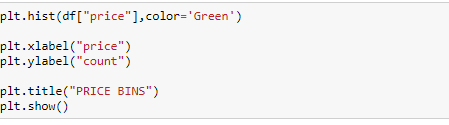


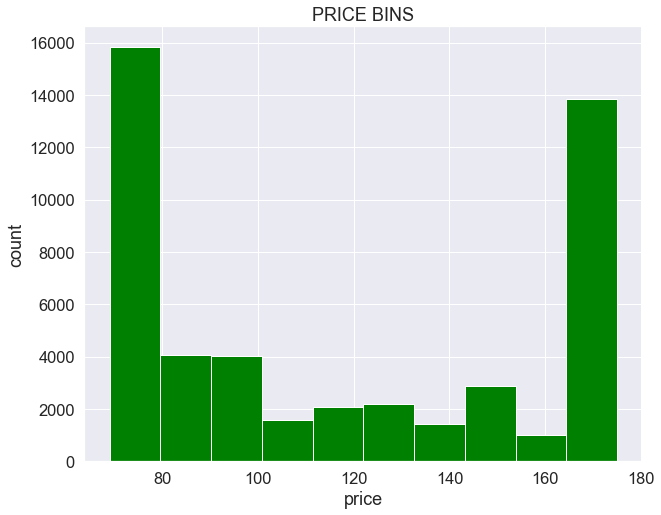
1. **Price Analysis**

****

****

From the above **Seaborn Barplot**, we can observe that Manhattan has gained a significant increase in the price compared to others seeing, Manhattan has been based nearby a river, it has gained popularity when compared to other localities.

****

****

With the help of **Matplotlib Histogram**, we have visualized the Price column by counting the price figures for each property listed in Airbnb.