

Capstone Project-1

Airbnb Bookings Analysis

Team Members:

Gupta Sugandh

Vivek kumar

Soni Rani

Suraj Singh

Akshay Badhe

Outline

1. **Abstract**
2. **Introduction**
3. **Problem Statement**
4. **Play Store Dataset**
5. **Steps Involved**
6. **Data Analysis**
7. **Conclusion**

Abstract :

New York city has been one of the most popular cities for travel and hottest market for Airbnb is an Online-based marketing company that connect people looking for accommodation(Airbnb guest) to people looking to rent their properties(Airbnb hosts) on a short-term or long-term basis. The dataset Contains the real-world data of Airbnb of New York city and describes the listing activity and metrics In NYC,NY for 2019.

Introduction:

- The dataset include listing activity of a metrice in NYC , for 2019. This dataset contains prices for listing
- In different neighborhood groups within different neighborhood cities.
- It also contain different factor Like property type, reviews, and availability of listing; that can affect the price for the listing.
- The data has 48895 observations and 16 attributes. Response variable: Price per night.
- Explanatory variables: price, name host id, host name, neighborhood-group, neighborhood, latitude; longitude, room type, minimum-nights, no. of review, last review, review per month, calculated host listing, and availability 365 days.

Problem Statement:

One of the biggest challenges for Airbnb is to maintain positive customer experience along with having a financially profitable business model for property owners. How factors are affecting the price for the Airbnb listing in NYC? What is the overall location distribution of Airbnb NYC? Which neighbourhood has a better average price for the Airbnb listing? Our goal is to build a statistical model to effectively predict the price for the listings and company can use this model to come up with a price suggestion for the future listings.

AirBnb Dataset

The dataset consists of AirBnb booking system and the data is generated by the airbnb website, which is the best website to book rooms and hotels booking data in this dataset. This dataset has around 49,000 observations in it with 16 columns which specifies about the different types of data in the dataset about the hosts, customer and places and it is a mix between categorical and numeric values. This Data provides Information about the host and room booking done by different customers and details about the hosts and room details booking and reviews and the other details like place, price, reviews, availability and etc. Data about the different hosts and the services given by them at different neighborhood groups and neighbor and the latitude and longitude of the place where they provide the service. Different data about the hosts and the rooms helps to analyze the performance of the hosts and customer about likes and preference of the rooms and hosts.

Steps

Airbnb dataset involves various steps such as:

- ☐ **Loading the dataset**
- ☐ **Performing data preparation**
- ☐ **Cleaning the data**
- ☐ **Extracting statistics from the data**
- ☐ **Performing exploratory analysis**
- ☐ **Performing visualization**
- ☐ **Summarizing inferences**
- ☐ **Writing conclusions**

Data Preparation

Data creation is the process of clearing and replacing raw data before processing and analysis. This is an important step before processing and often involves reforming the data, modifying the data and assembling the data set to enrich the data.

- Collecting data

This step is to get to know the data and understand what needs to be done before the data becomes useful in a particular context. This can be done by reading and initializing the CSV file

- Statistical analysis

Although the dataset appears to have the correct data type for each column, we need to check that. Inconsistent data types will cause problems when handling problems.

- Clear and verify data

This step is important for removing faulty data and filling in the gaps.

Here are the key functions:

- Removal of external data

There is a step in which the important and usual data needs to be filtered

- Fill in the missing values.

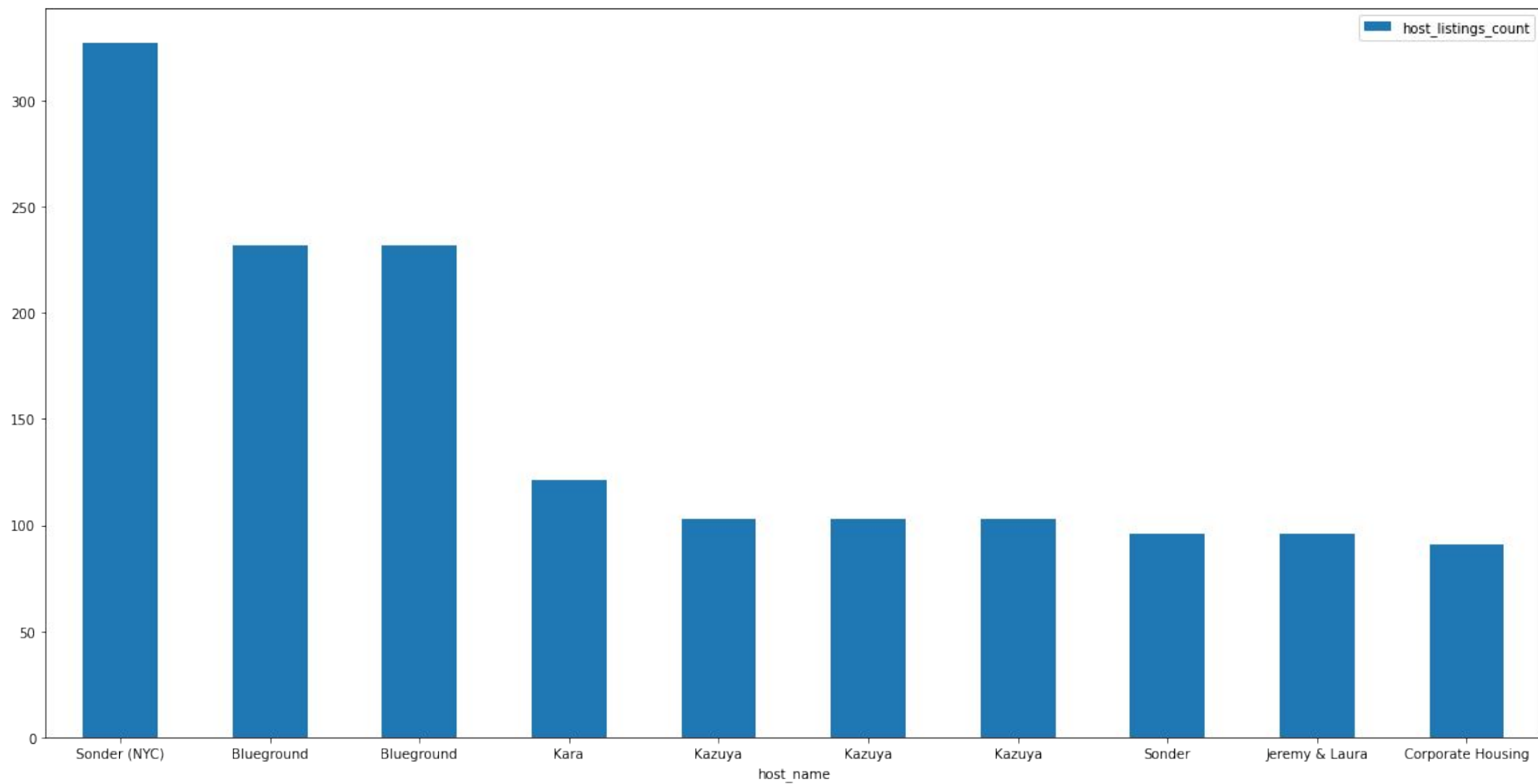
- Adapting data to standardized patterns.

- Transform data

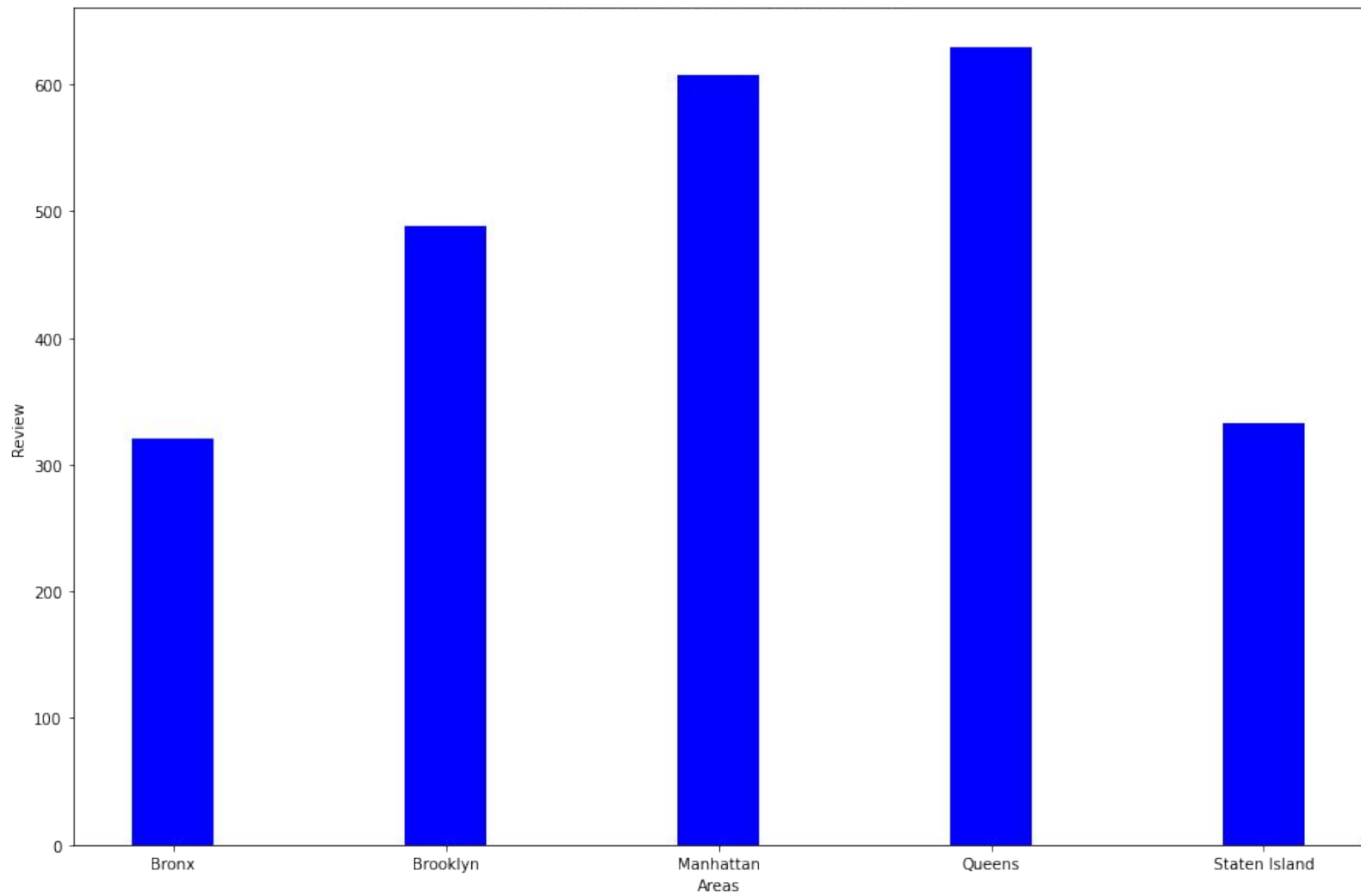
Data transforming is the process of updating format or value entries to reach a well-defined result or to make the data more easily understood by a large audience.

- Exploration analysis and visualization

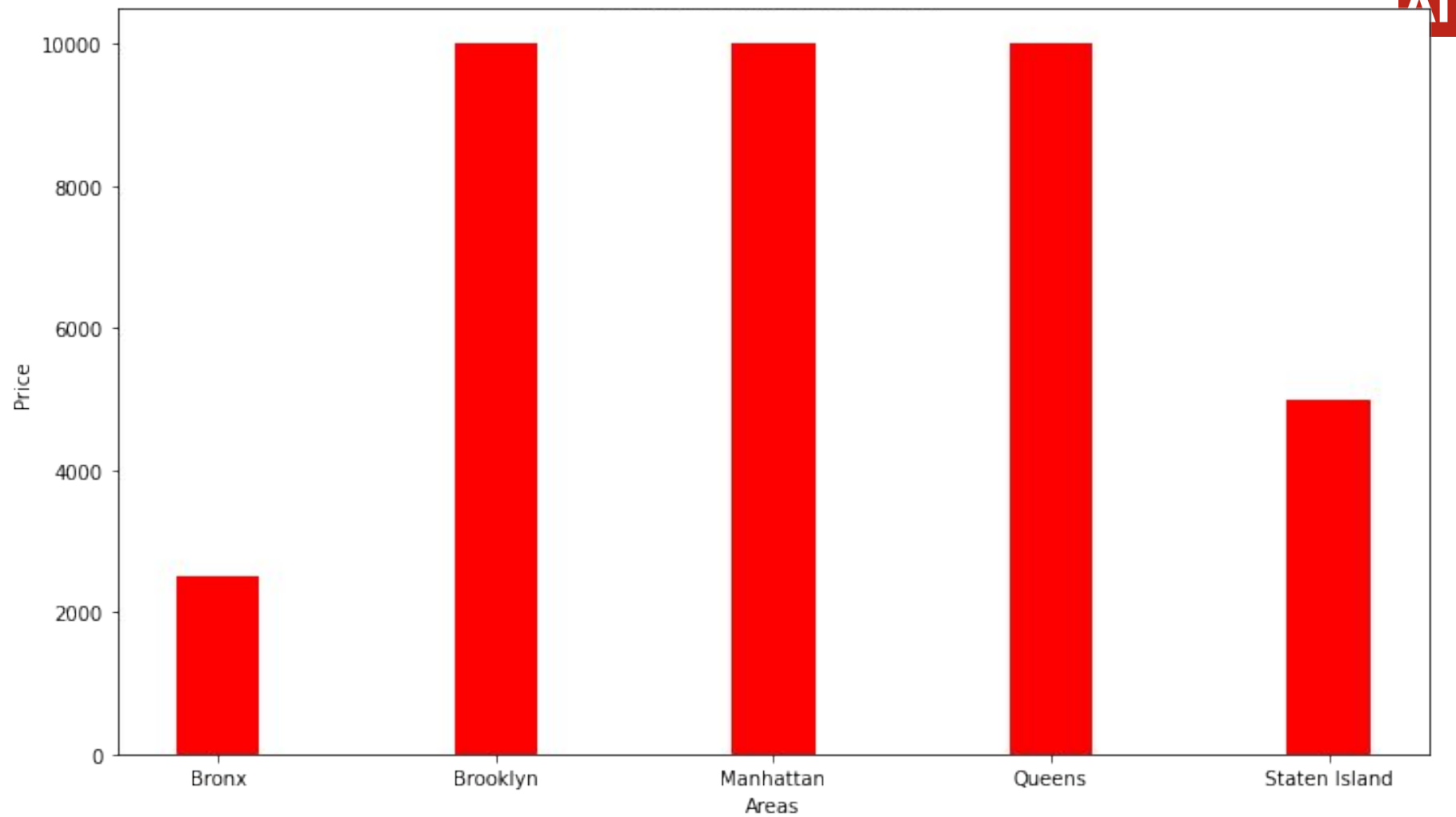
Exploratory Data Visualization (EDVs) are a type of visualization that we combine when we have no clue what information is in our dataset.

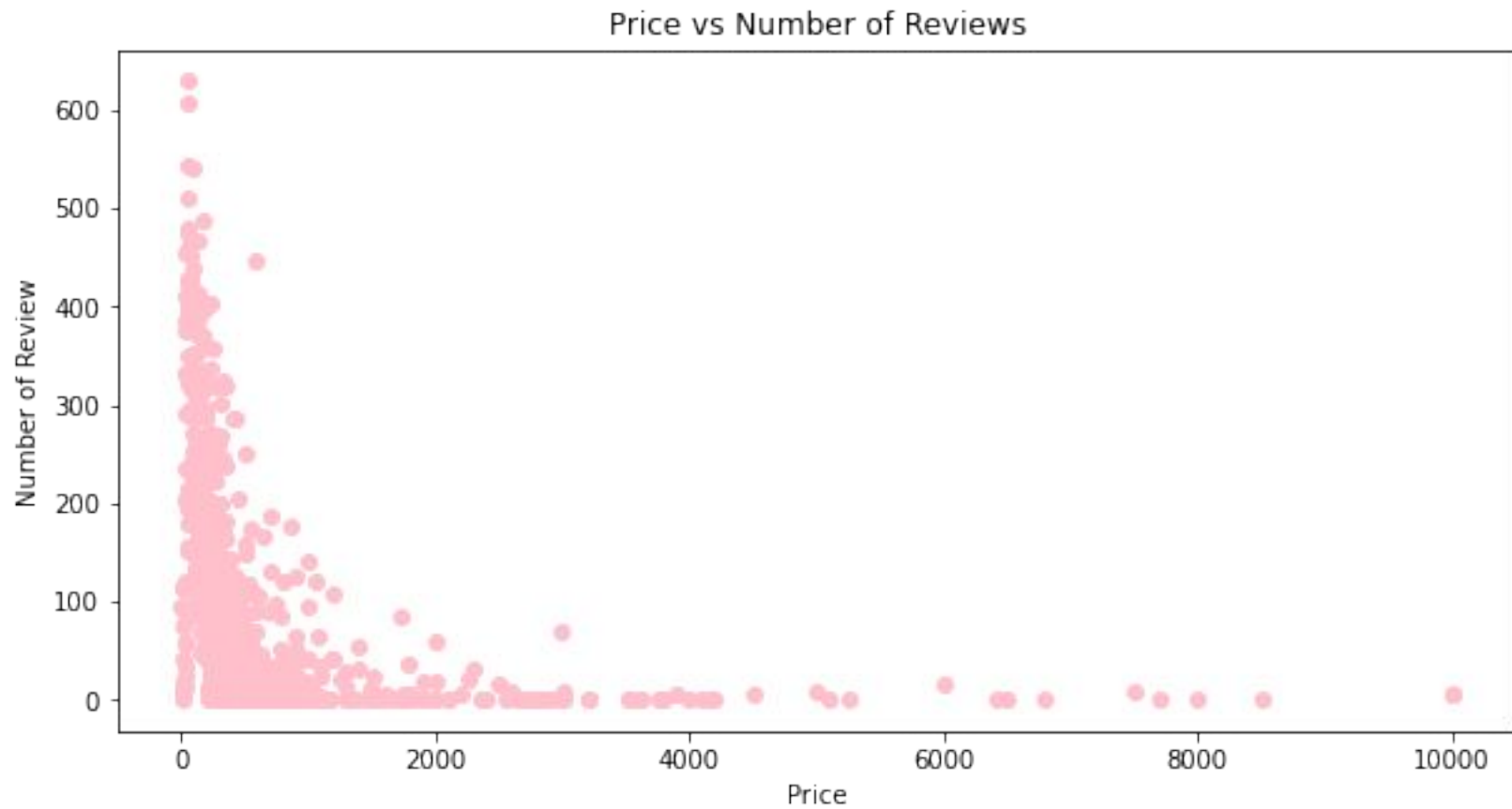


Area and Number of reviews in that area

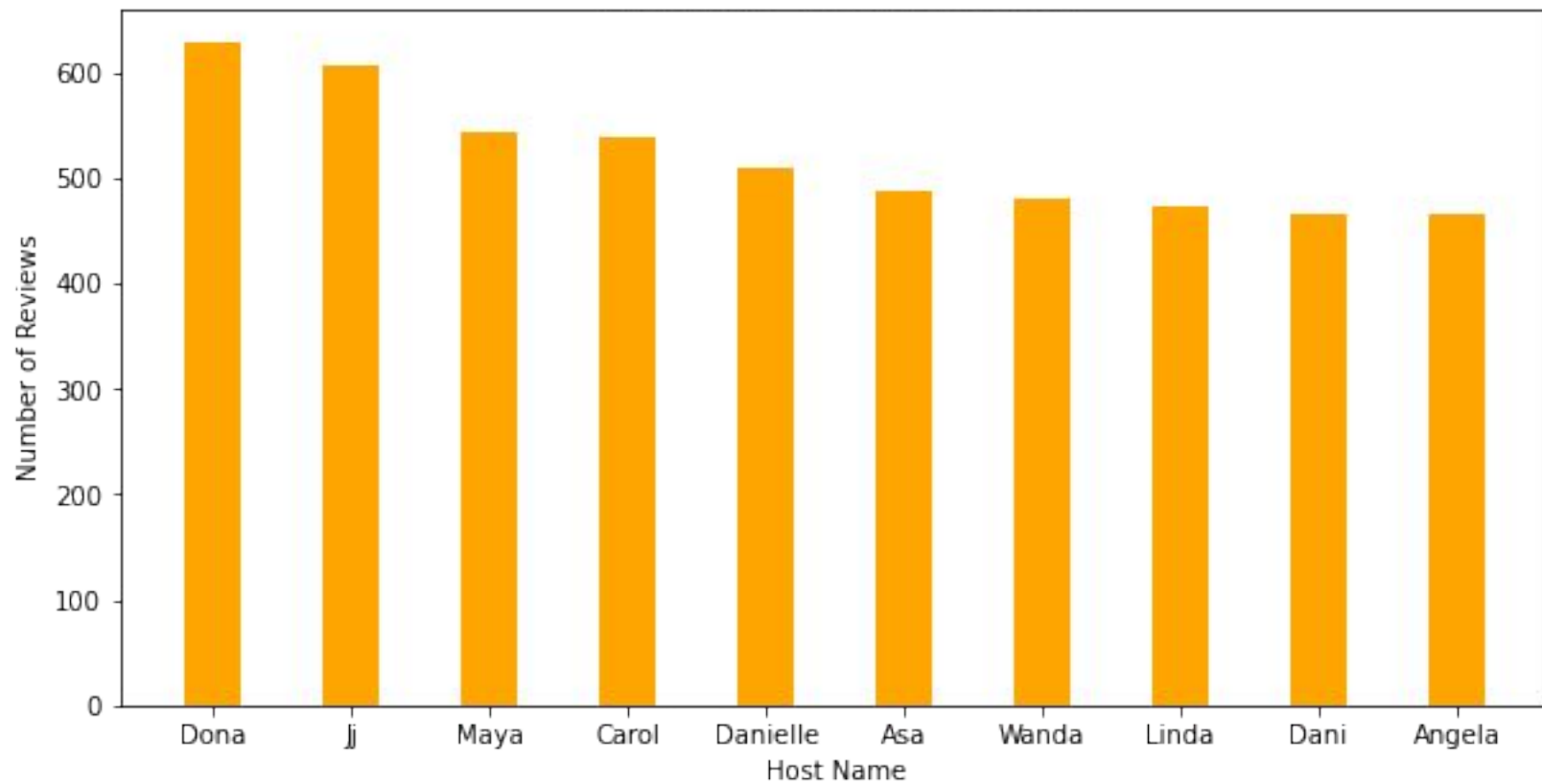


Area and Prices in that area

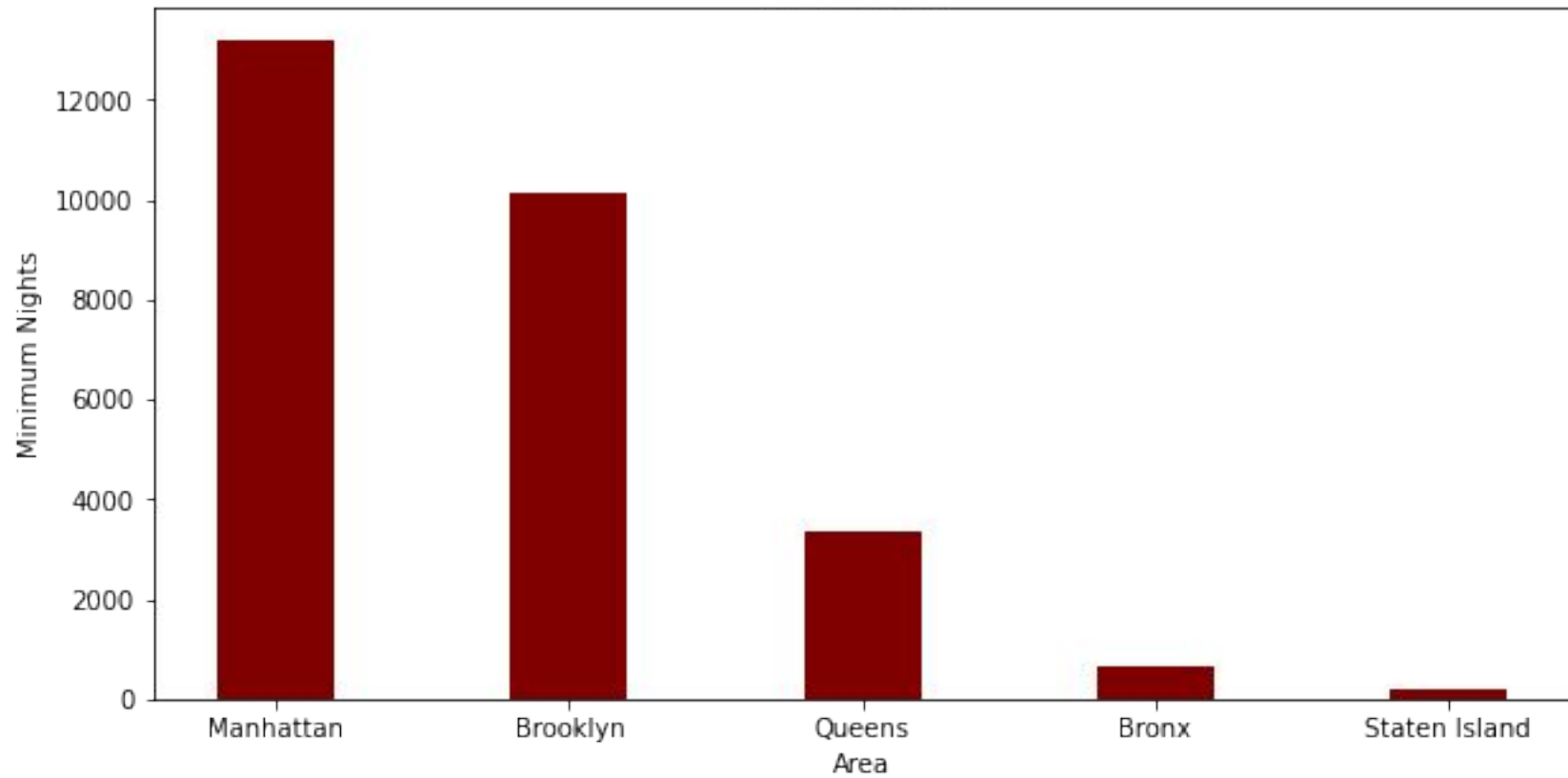




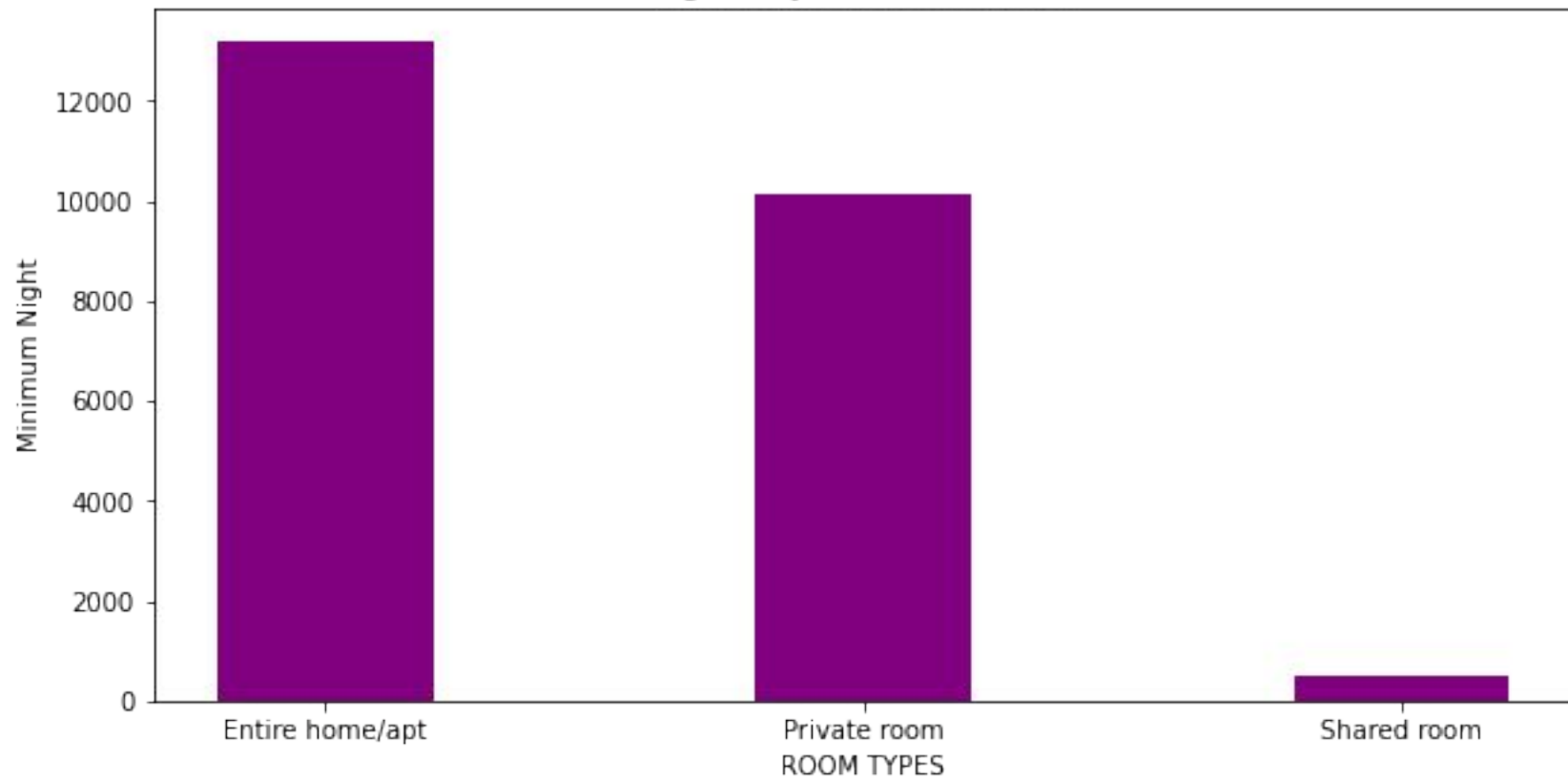
Busiest Hosts based on reviews



Busiest Area



Night stayed in Room Chart



Conclusion:

- ❑ This Airbnb (NYC 2019) Dataset For The Year 2019 Appeared to be a very rich Dataset with a variety of columns that allowed us to do deep data exploration.
- ❑ First, we have found hosts that take good advantage of the Airbnb platform and provide the most listing. Thus, we found that our top host (Sonder) has 327 Listings.
- ❑ On the basis of reviews the neighborhood group, Queens has maximum number of reviews followed by others. So, if there are more number of Reviews for particular Neighborhood group that means that place is a tourist place.
- ❑ Manhattan is the busiest neighborhood place as the minimum night spend is maximum compared to others. So, if people are not staying more than one night means they are travellers.
- ❑ From the above analysis we can conclude that most of the people choose entire home apartment as their first choice.