Capstone Project

Airbnb Bookings Analysis - EDA

Akash Salmuthe

Contents


AI
airbnb

Introduction

What is Airbnb?

- Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.

- Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczykand Joe Gebbia.

Problem Statement

- Airbnb generates a lot of data -data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

- Data analysis on millions of listings provided through Airbnb is a crucial factor for the company.

Key Objectives

- ❑ What can we learn about different hosts and areas?
- ❑ What can we learn from predictions?
- ❑ Which hosts are the busiest and why?
- ❑ Traffic among different areas
- ❑ Finding price difference at different localities

Data Summary

- **Data set**- *Airbnb NYC 2019* -contains observations of bookings done in 5 neighbourhood groups in New York City through Airbnb.
- **Rows**-48895
- **Columns**-16
- **Room types** – Entire homes/apt, private rooms, shared rooms
- **Neighbourhood groups**:-Manhattan, Queens, Staten Island, Bronx, Brooklyn

Data Description

id: unique reference number for each different hotel.
name: name of different hotels of various neighborhoodgroups.
host_id: unique reference id of each individual host.
host_name: name of host hosting different hotels.
neighbourhood_group: aggregate group of neighborhoodcities of some particular regions.
neighbourhood: cities present in NYC.
latitude: latitude is a geographic coordinate that specifies the north–south position of a point on the Earth's surface. Latitude is an angle which ranges from 0°at the Equator to 90°at the poles.

Data Description continued...

longitude: Longitude is a geographic coordinate that specifies the east–west position of a point on the Earth's surface, or the surface of a celestial body.
room_type: Different room types available for booking, which contains Private room, Entire home/apt, Shared room.
price: price per each night stay of different room types at various hotels.
minimum_nights: minimum nights booked in particular hotel.
number_of_reviews: count of reviews got for each hotel.

Data Description continued...

last_review: date of last review got by a customer to a particular hotel.
reviews_per_month: count of reviews getting per month of a particular hotel.
calculated_host_listings_count: It represents total number of listings made by a specific host. In some cases, the properties are same but some of the other features differ like(room_type).
availability_365: number of available days for booking in a year

Data Preparation

1. **Basic data exploration:**

Using *describe()* and *info ()* and size functions of pandas. Gone through a basic exploration of data before entering into EDA.

2. **Dropping unnecessary data:**

Columns like *reviews_per_month* has 10052 missing values but this feature is important for analysis while the other column *last_review* we can drop it.

3. **Verifying Data quality:**

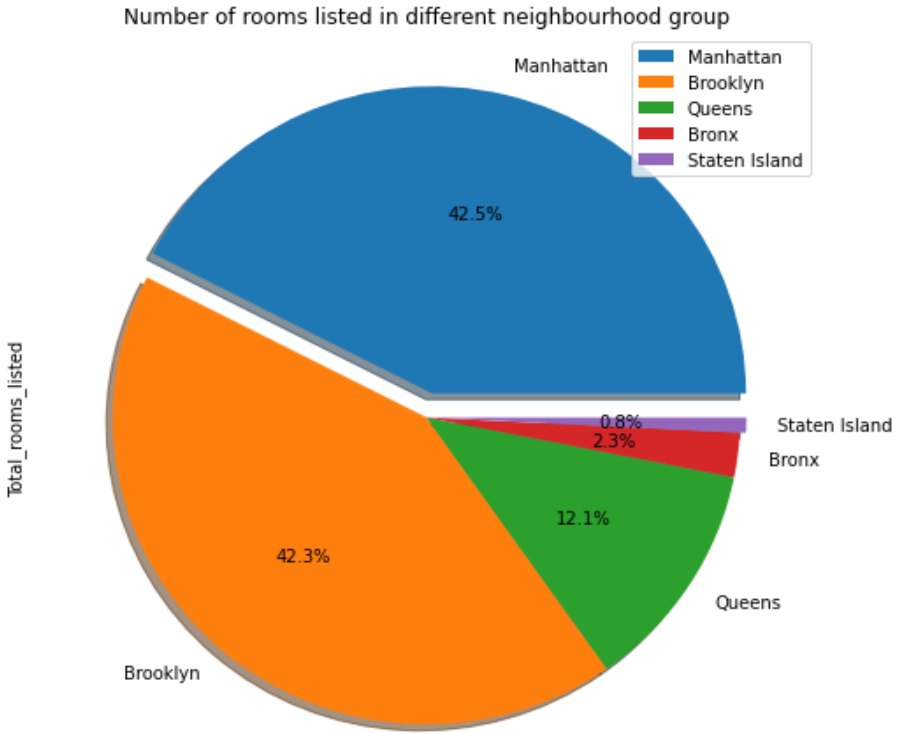
We have gone through whole data and checked null values and reviewed any missing data or wrong data.

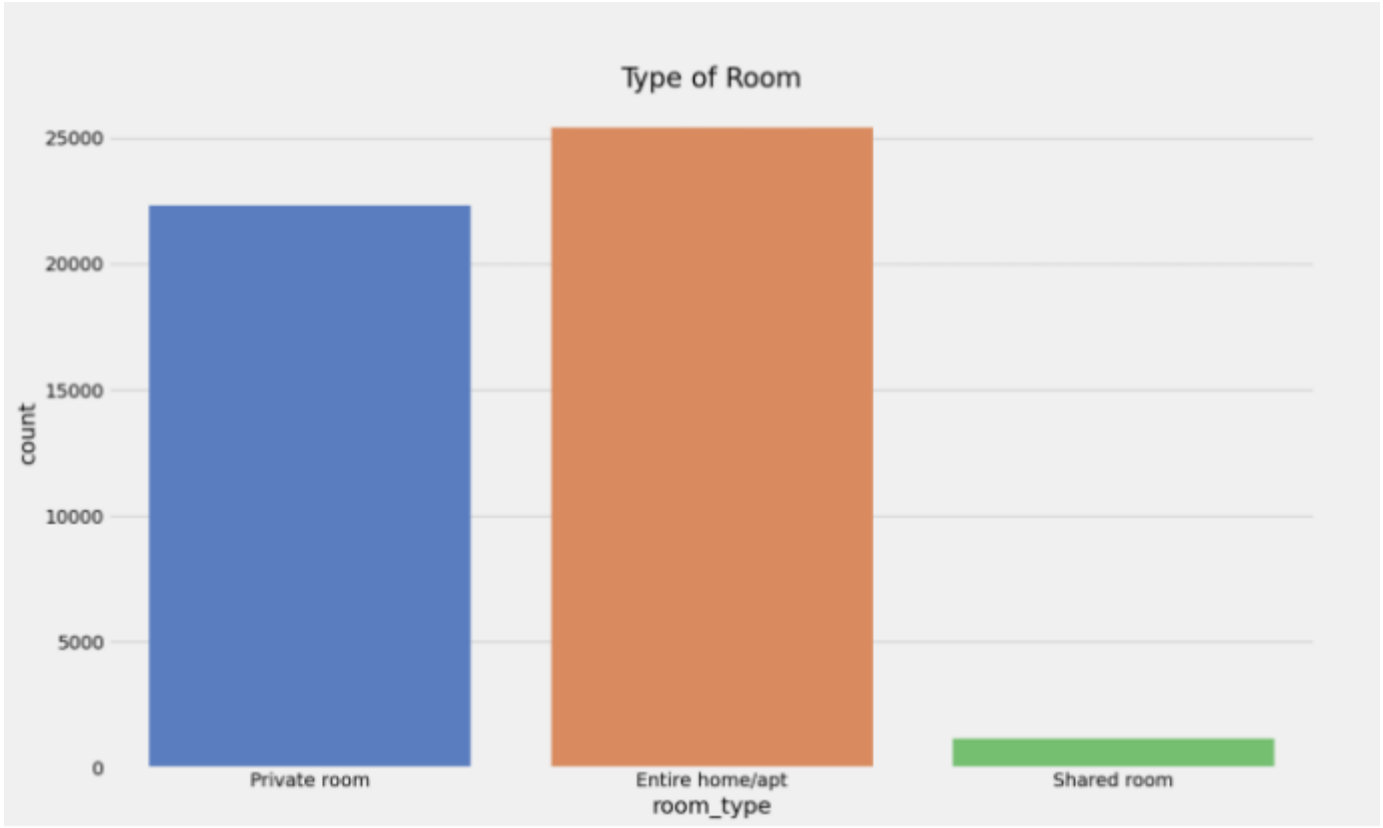
Exploratory Data Analysis

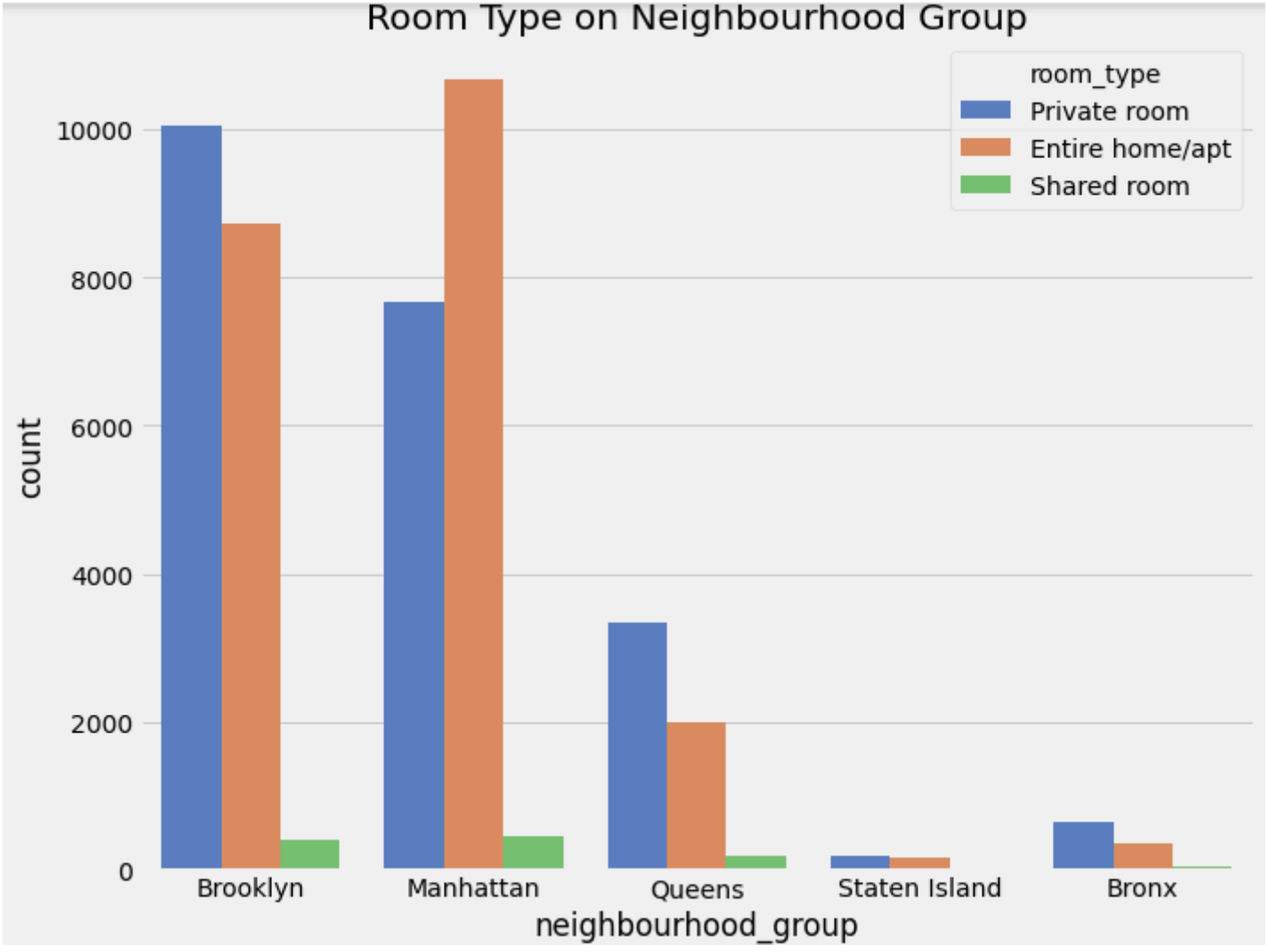
Why EDA ?

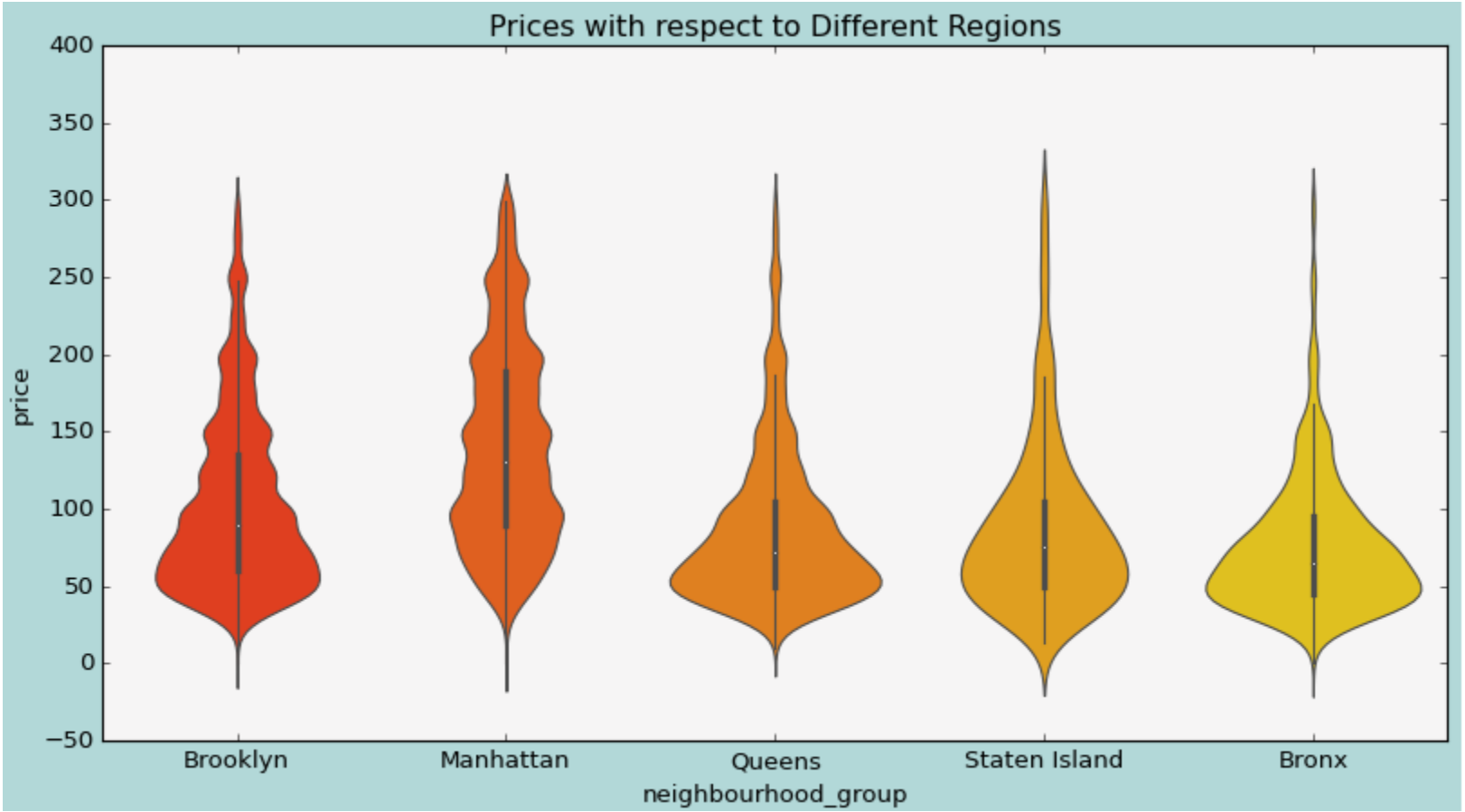
Booking analysis based on price and the most booked room type

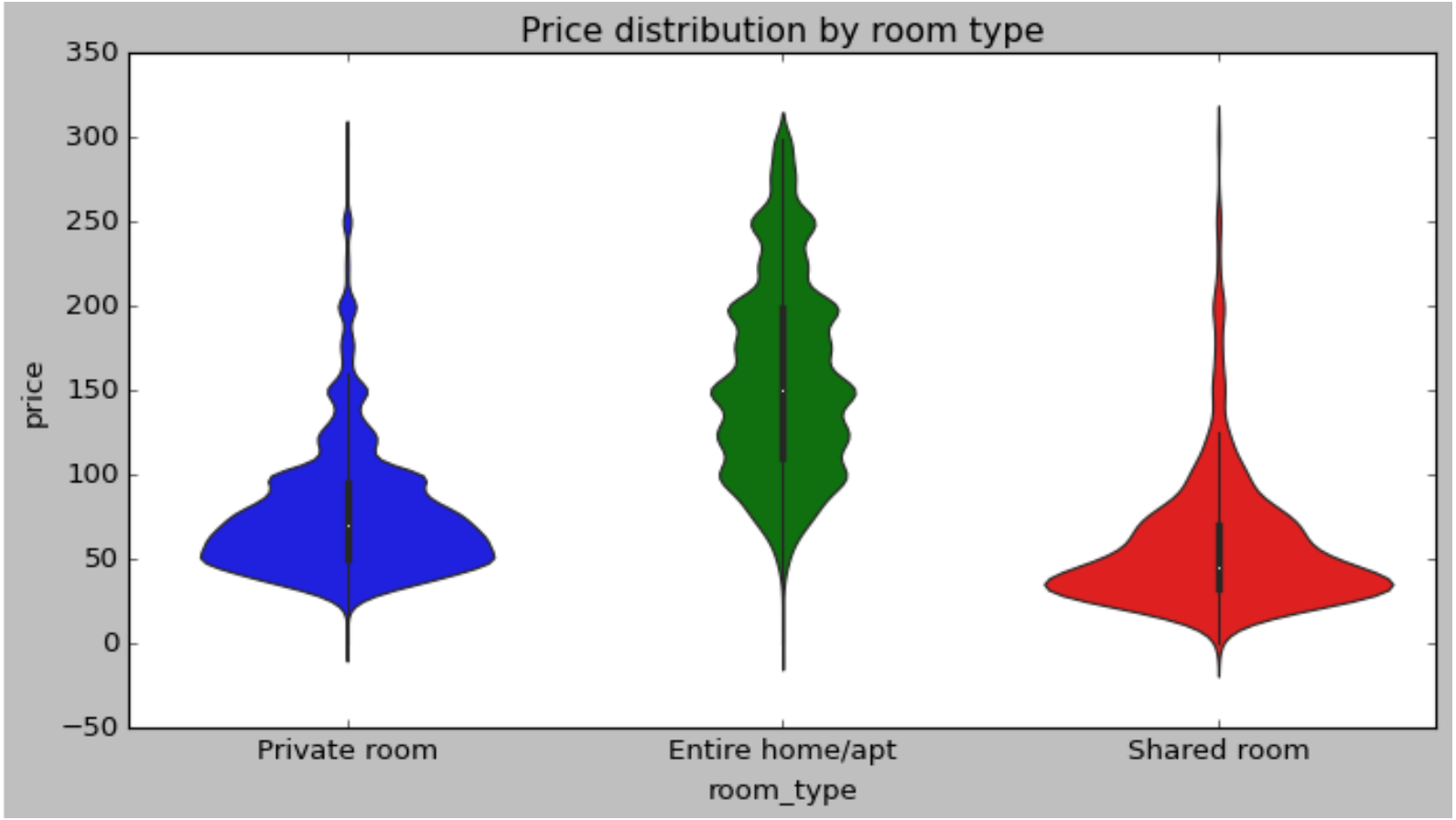
Number of listing in neighbourhood group




Type of Room
25000
20000
15000
10000
5000
0
count
Private room
Entire home/apt
Shared room
room_type


Room Type on Neighbourhood Group
room_type
Private room
Entire home/apt
Shared room
10000
8000
6000
4000
2000
0
count
Brooklyn
Manhattan
Queens
Staten Island
Bronx
neighbourhood_group

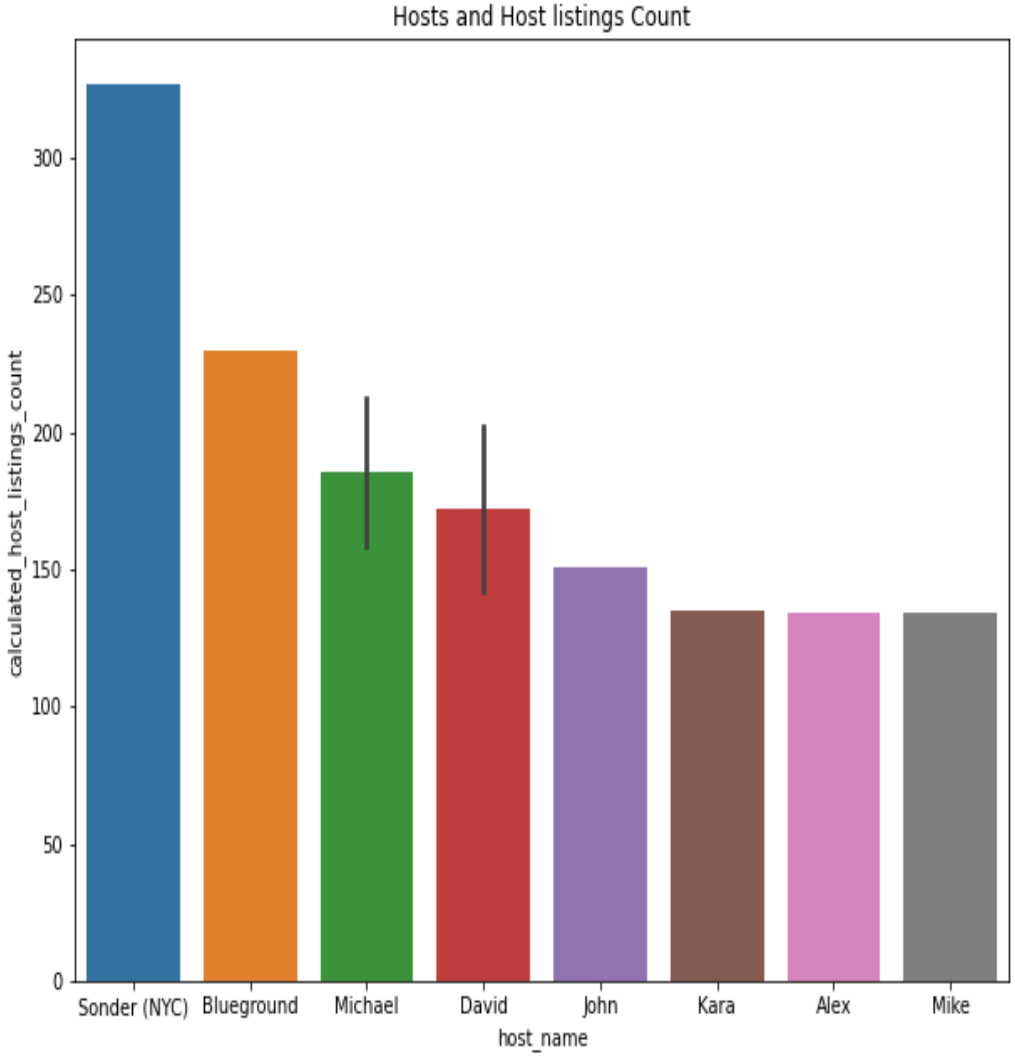

Prices with respect to Different Regions
400
350
300
250
200
150
100
50
0
−50
price
Brooklyn
Manhattan
Queens
Staten Island
Bronx
neighbourhood_group


Price distribution by room type
350
300
250
200
150
100
50
0
−50
price
Private room
Entire home/apt
Shared room
room_type

Top Hosts and their listings count

Most Host:

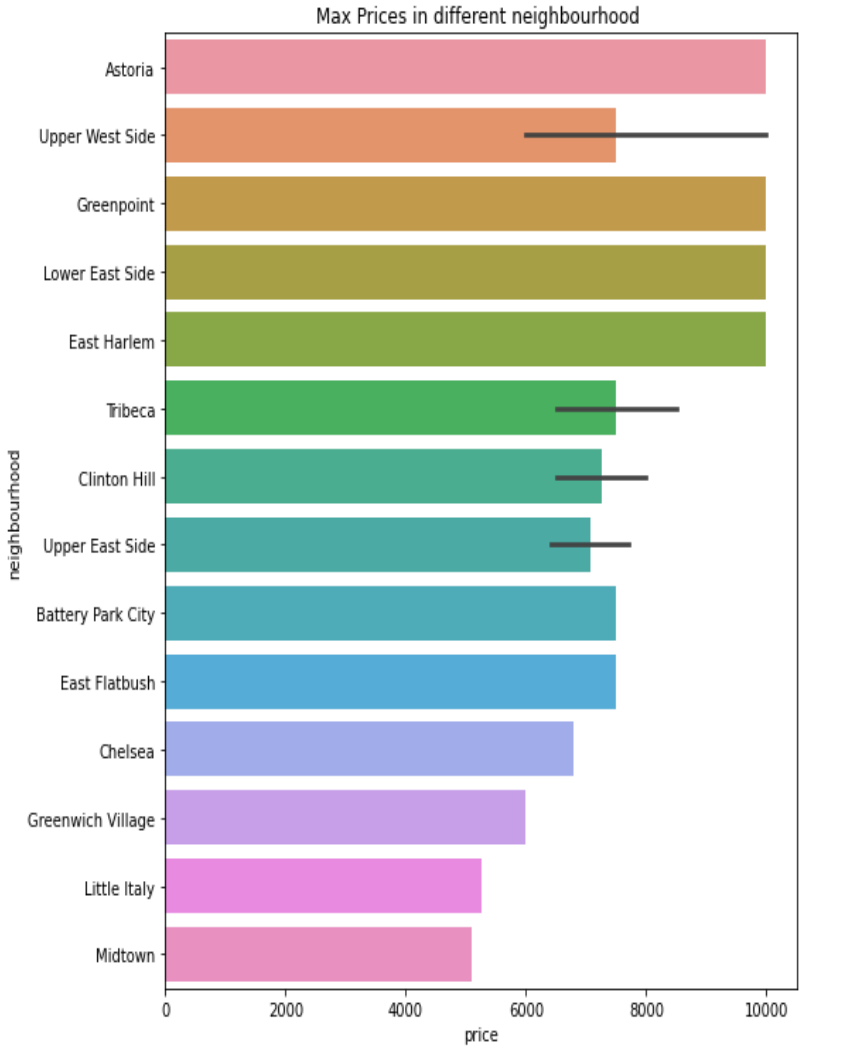
1. Manhattan and Brooklyn



Max prices in different neighborhood

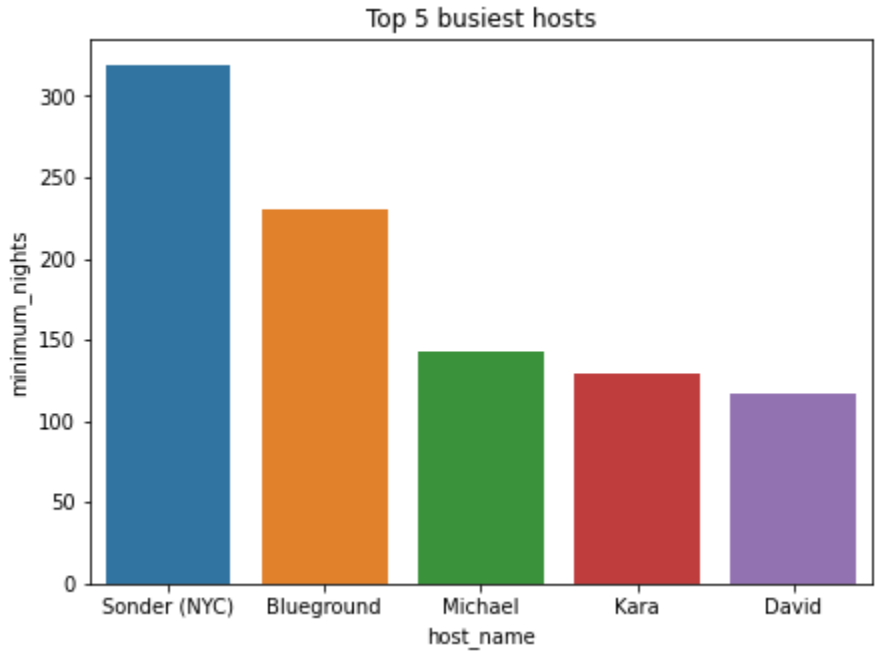
The highest price is 10,000 $

Minimum price is 5000$



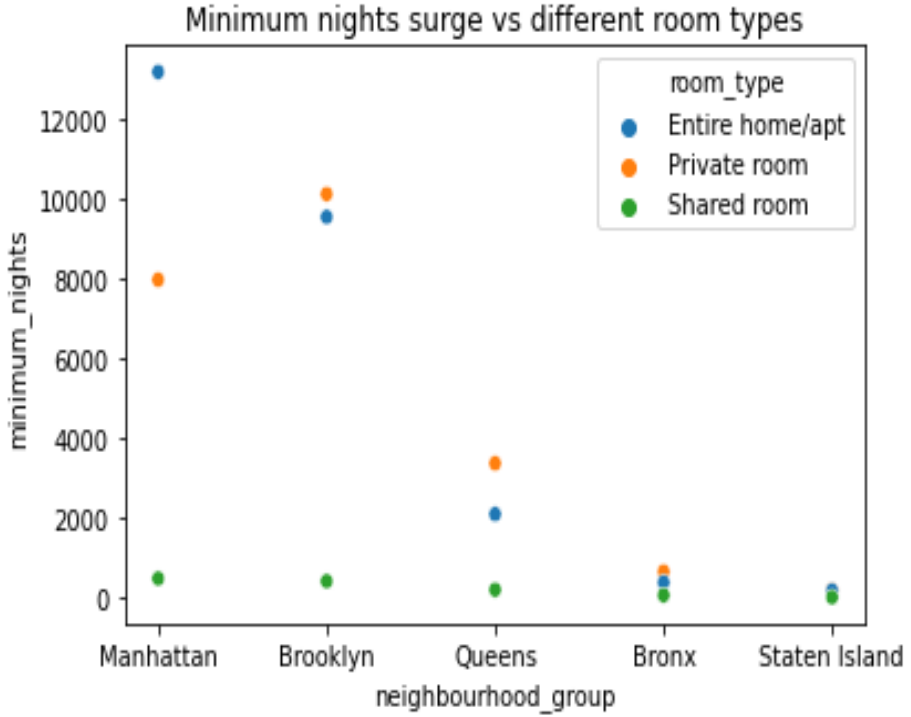
Finding Busiest Host

All 5 busiest host are from manhattan



Traffic among different room types and different neighborhood

- In Manhattan, people are preferring "Entire Home/apt"
- In Brooklyn, Queens and Bronx people are preferring private rooms
- In staten island, people are having equal preference



Challenges faced

- Verifying quality of such huge data and looking for error values.
- Dropping down irrelevant data and making the whole data getting ready for full pledged data analysis.
- Understanding and visualizing complex numerical data, and communicating business solutions.
- Analysing and solving various queries and presenting clear cut outputs.

Conclusion

- Entire home/apt is the most expensive and yet get the most traffic.
- Private room is second highest in terms of booking which suggests people value their privacy.
- Manhattan is the most expensive Neighbourhood group and Brooklyn gets the most traffic.
- Sonder(NYC) is the most successful host, who only caters in Manhattan and only lists apartments and private rooms.