Capstone project

EDA on Airbnb booking By- Paras Chopra

Problem Statement

Airbnb, Inc. is an online marketplace for long & short term rental accomodations founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia.

This Project will include all the process that is required to perform data analysis - Data Discovery & Understanding, Wrangling , visualization & statistical analysis when required.

We analyzed the given dataset which had around 49000 listings. Our main objective is to find key factors that influence the properties listings & relationships between them. We will achieve this in various steps

Data discovery - where we import the data & get some information about the data,

Data cleaning / wrangling - where we will check the data for any inconsistencies & clean the data to make it more useful for later steps,

EDA or exploratory data analysis - where we will analyze the data using charts & statistics [if necessary]. This can help us finding valuable insights that we can use to make decisions that will help our organisation to increase their profits.

Understanding the Data

There are around 49000 listings consisting of variables below :-

ID - Identity number for each listing

name - name of property

HostID - identity number for host/property owner

Host_name - name of host/owner

neighborhood group - a set of neighborhood clubbed together

neighborhood - locality in the city

latitude & longitude - geographical coordinate system

room_type - type & orientation of room

price - rate charged by customer per night

minimum_nights - minimum number of nights stayed by customer

number of reviews - number of reviews given by customers

last review - date when last review was given

reviews per month - number of reviews given by customer per month averagely

calculated_host_listings_count - total count

availability_365 -availability around the year

WorkFlow

Data discovery - where we import the data & get some information about the data,

Data cleaning / wrangling - where we will check the data for any inconsistencies & clean the data to make it more useful for later steps,

EDA or exploratory data analysis - where we will analyze the data using charts & statistics [if necessary]. This can help us finding valuable insights that we can use to make decisions that will help our organisation to increase their profits.

Then we will list out all observations & conclude the project

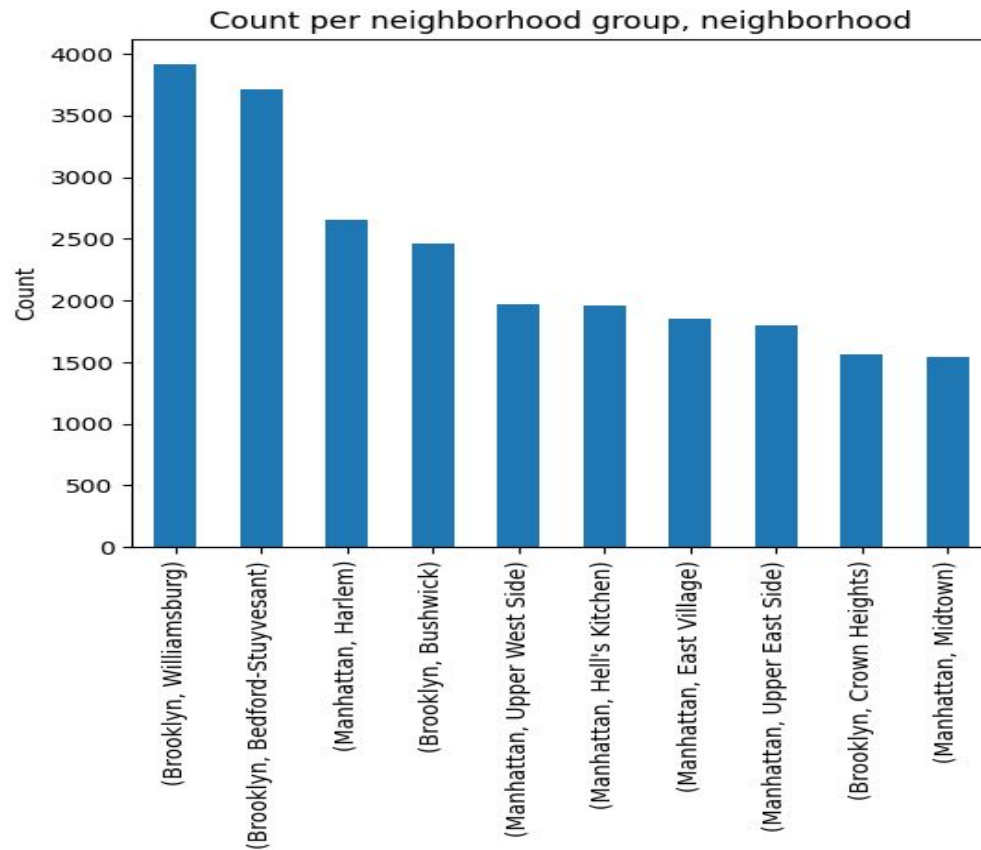# Correlation between all possible Variables
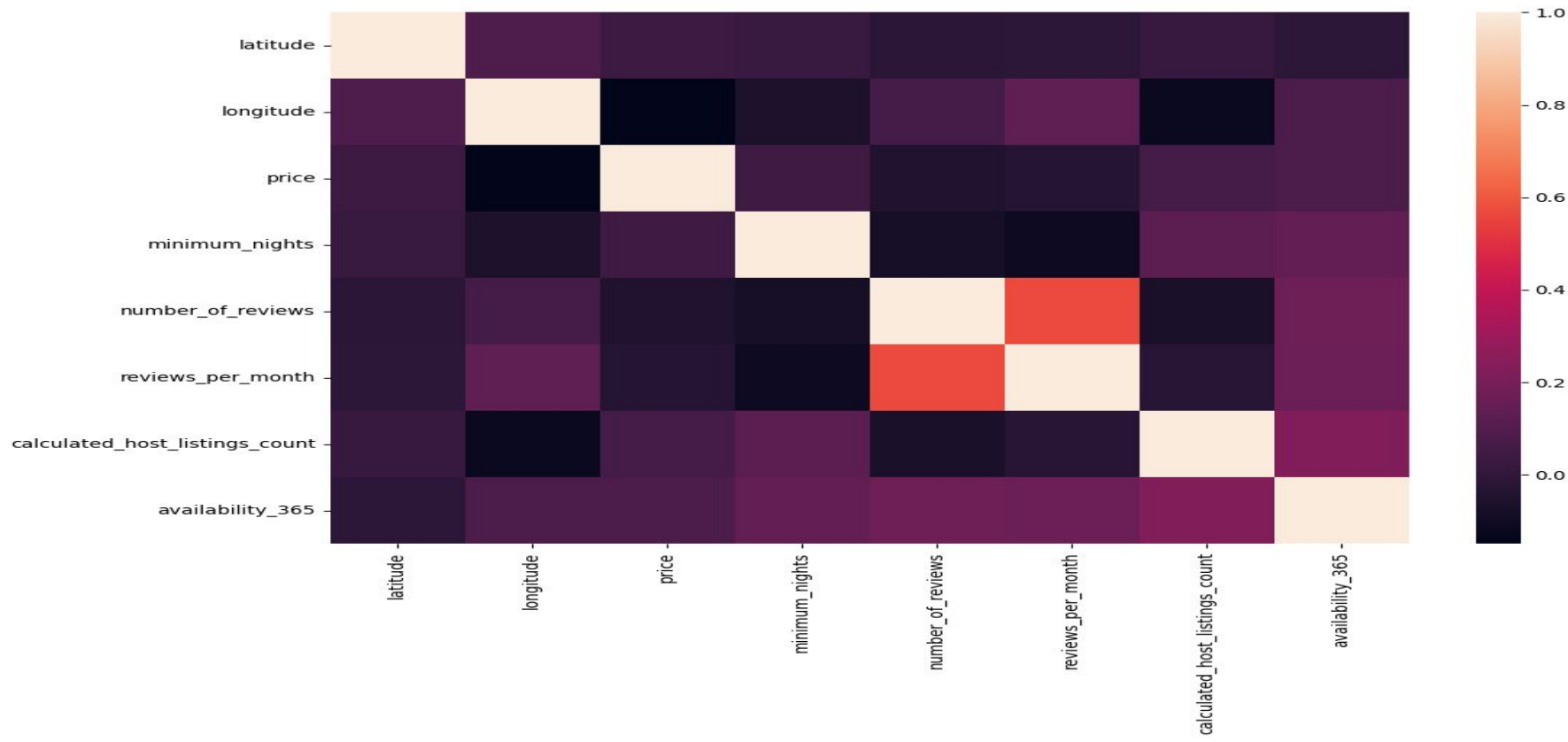
Univariate analysis - count of neighborhood group & neighborhood

Most of listings are in brooklyn or manhattan

If you want to rent out flats ,

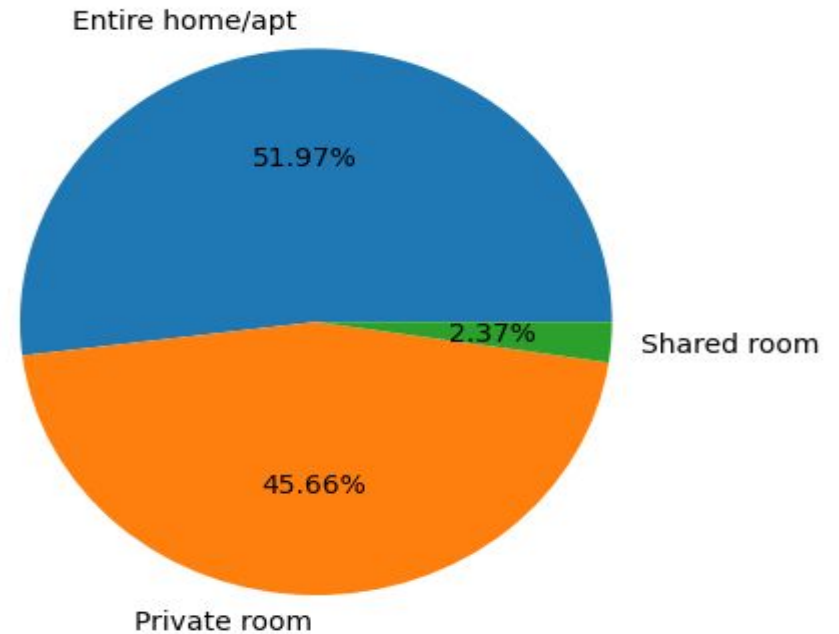then brooklyn & manhattan are the best places

to do it

Count per neighborhood group, neighborhood

Count of room type

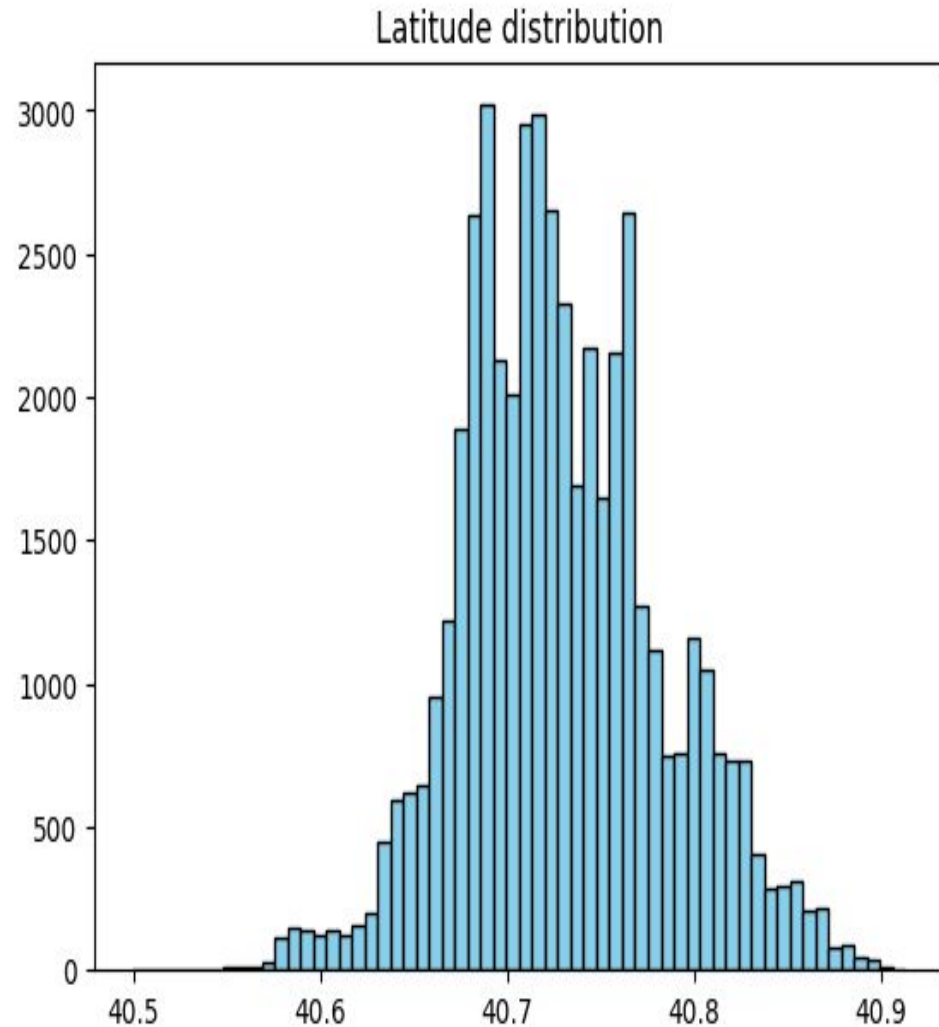Entire homes & private rooms are majority of listings

Seems like private rooms & entire homes are in demand

Latitude distribution

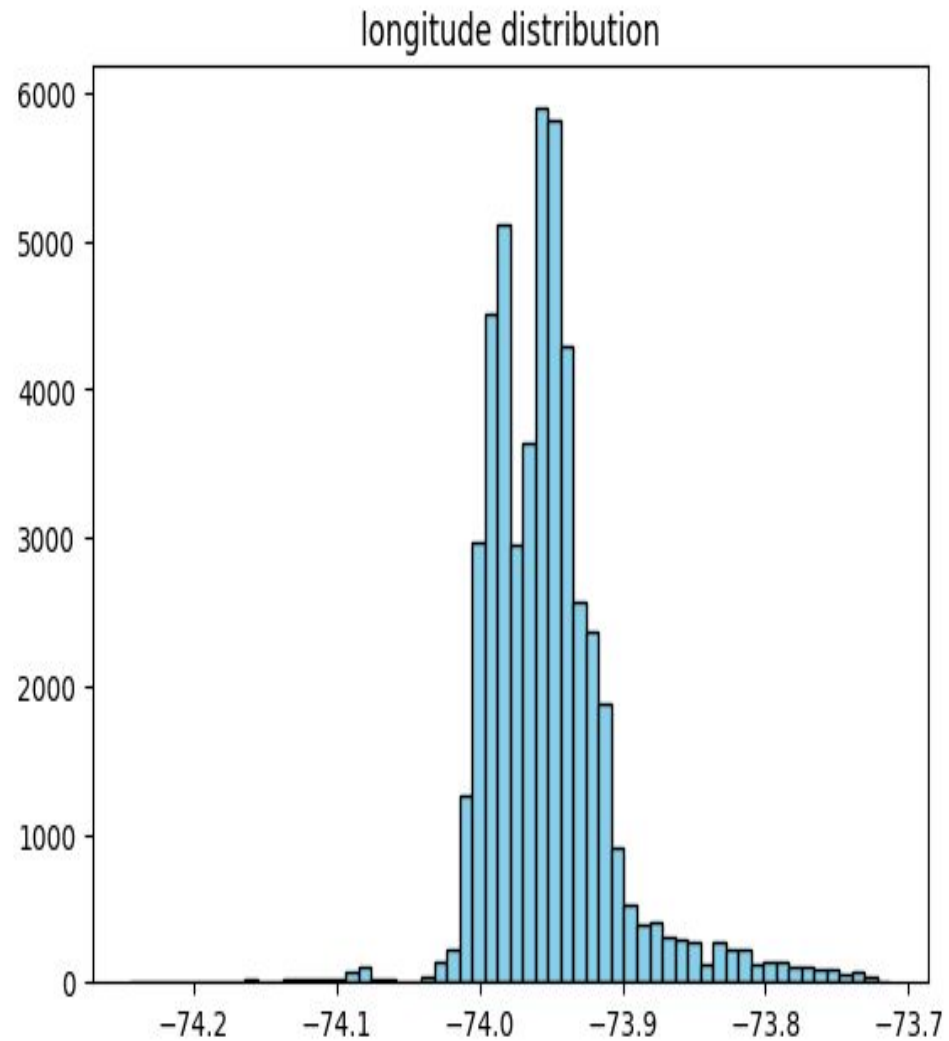Majority of listings lies between latitudes - 40.65 - 40.75

If you want to advertise your flat you can do it in these locations

Latitude distribution

longitude distribution

majority of listings lie in longitude ranges - 74-73.9

this can help us to find where we should advertise our flats to rent

minimum nights distribution - histogram

majority of listings have 0 to 10 min nights, there's also a considerable number of 30 minimum nights

0 to 10 minimum nights & 30 nights package seems to be the industry standards.

minimum_nights distribution
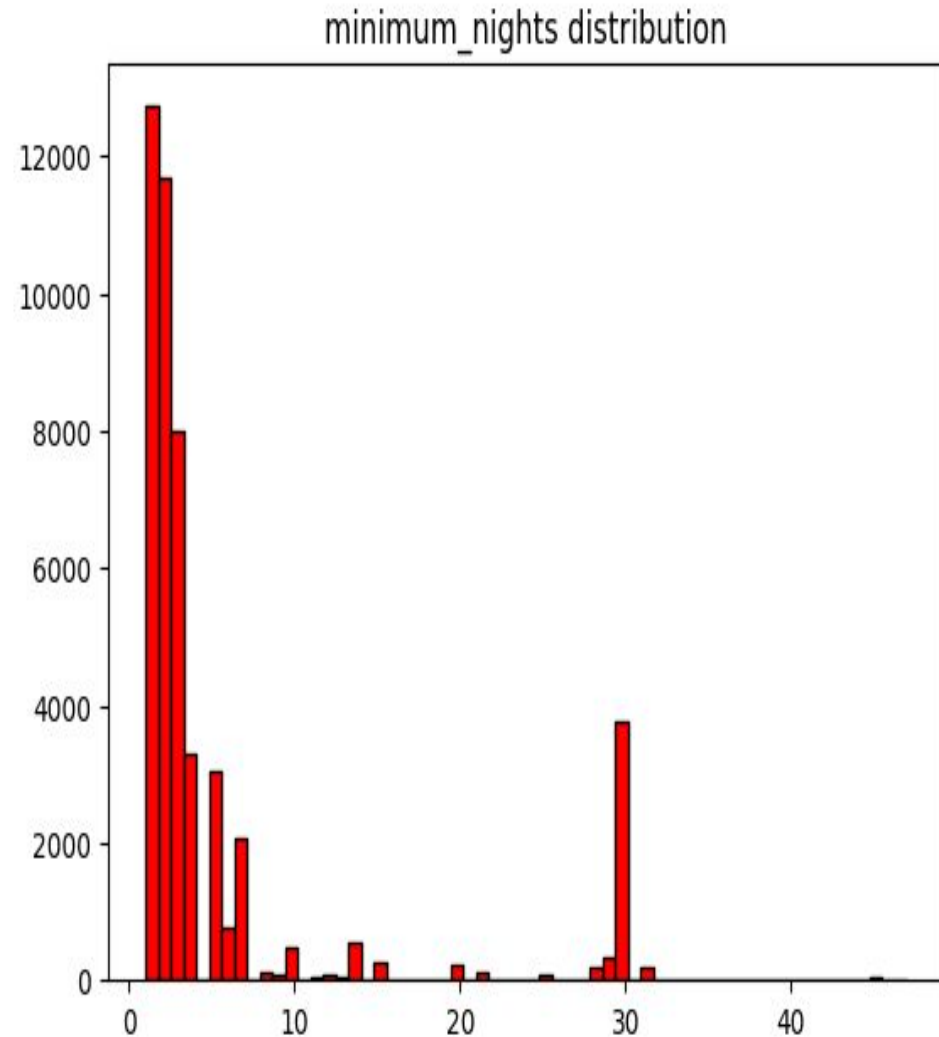
availability 365 distribution

majority of listings are available for 0-100 days per year , most likely weekends [52 weeks per year ie 52 weekends ie around 100 days]. There are also a lot of listings that are available almost entire year.

If you're starting a rental business , then its advisable to make it available for weekends.

availability_365 distribution
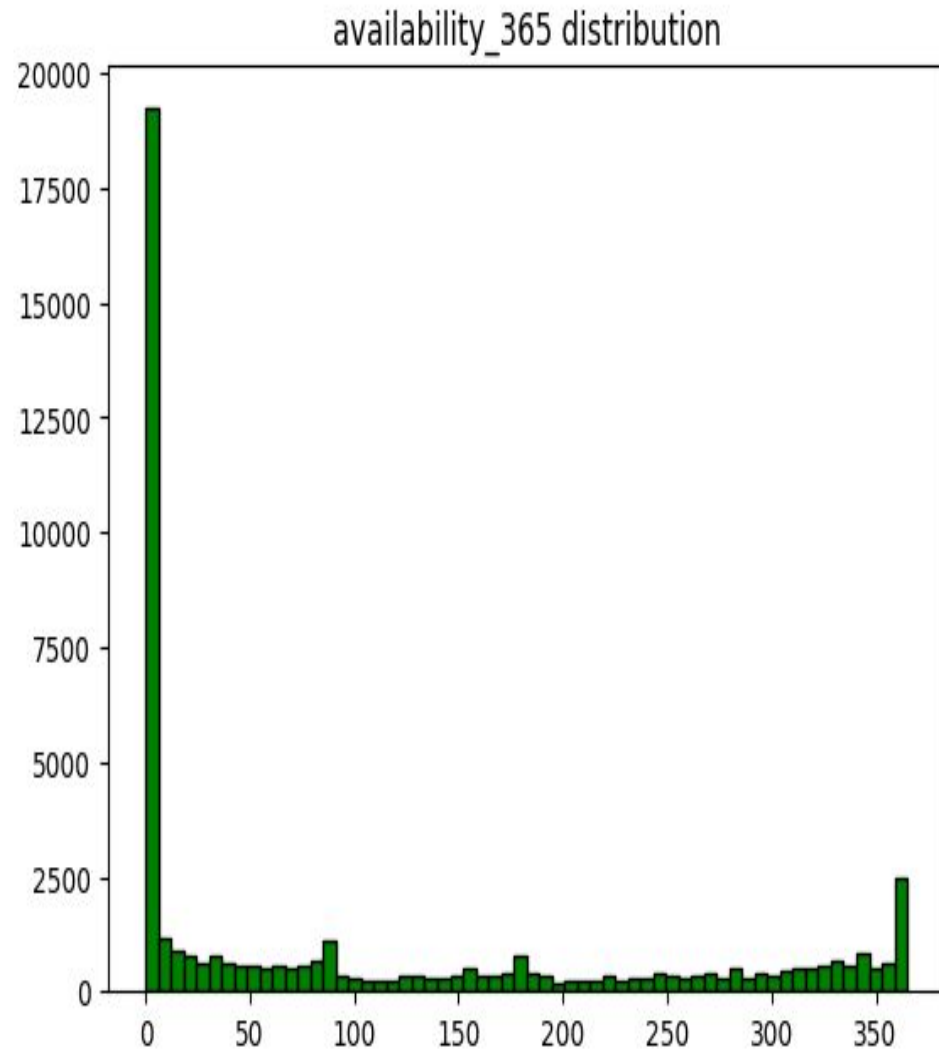
Avg Price per neighborhood group & neighborhood

Staten Island & manhattan are the most expensive , then bronx & brooklyn

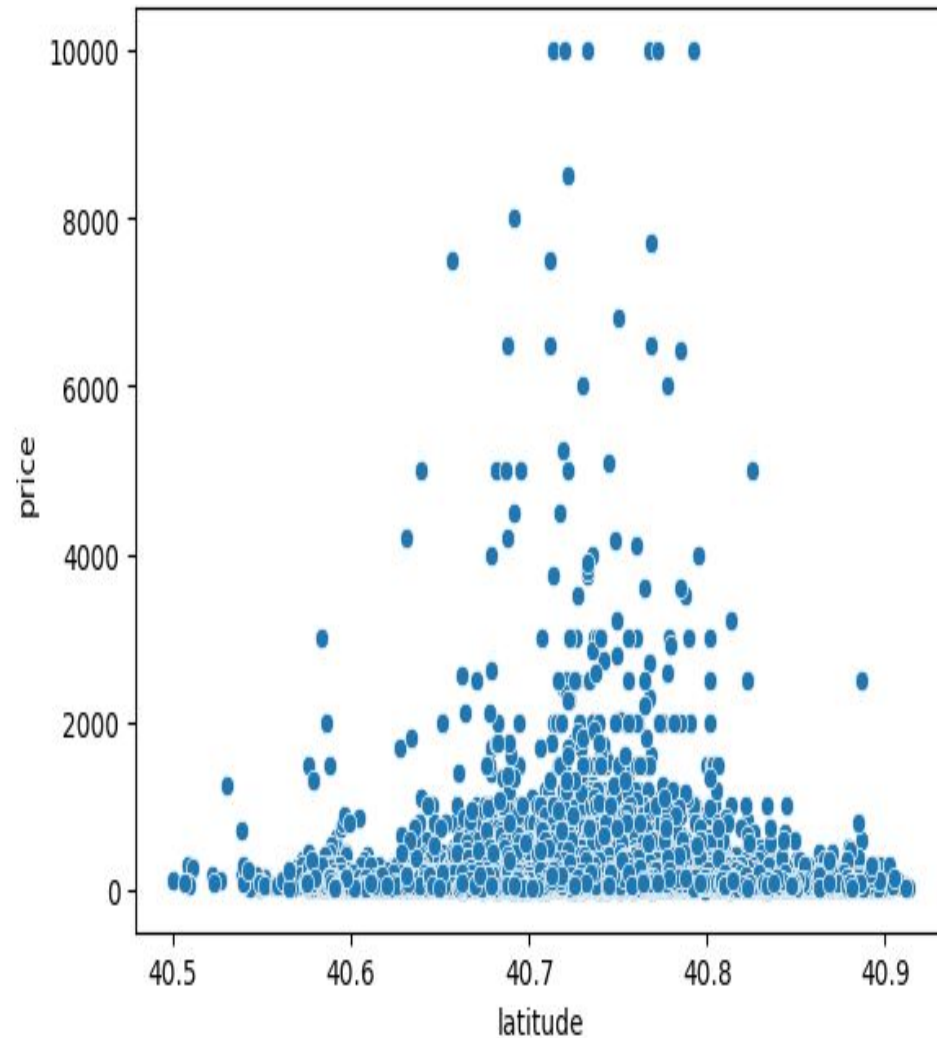If you want to set up a high end apartment for rent theses are the places to do so - Staten islands , manhattan

Count per neighborhood group, neighborhood
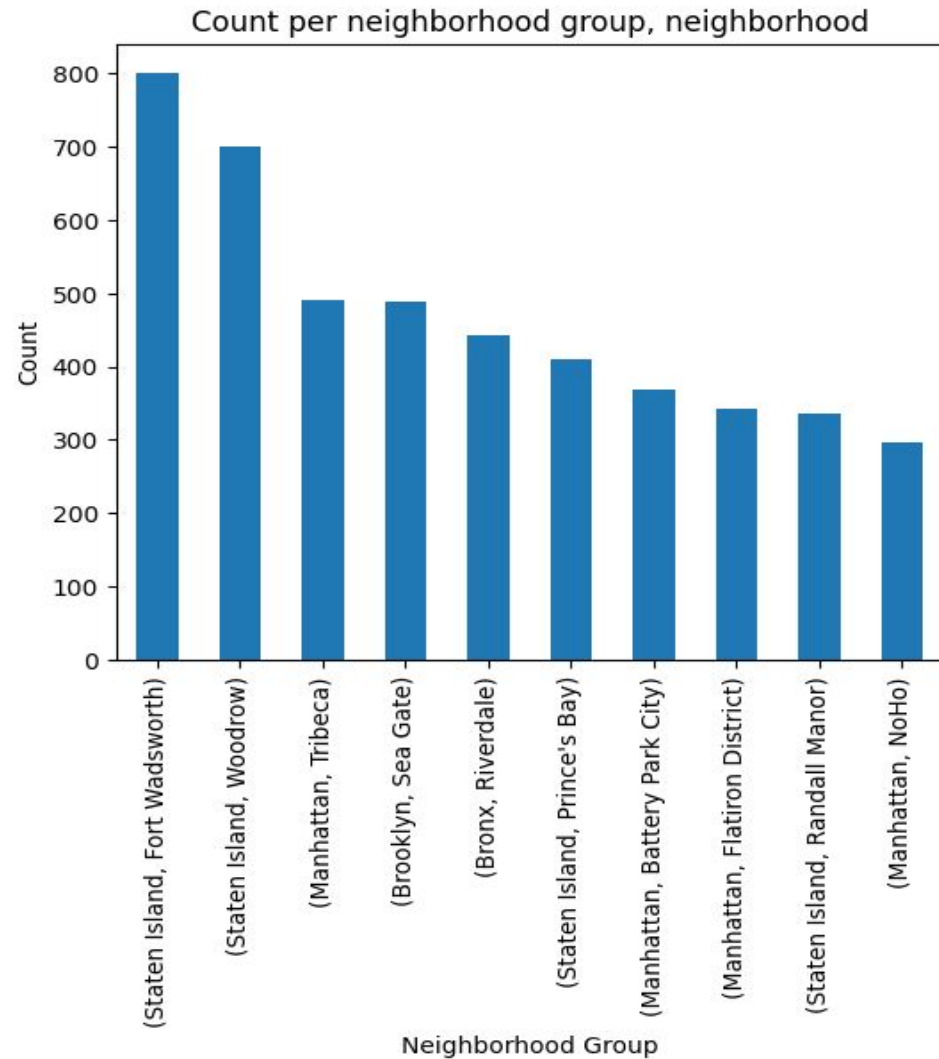
Price vs latitude

the prices & number of listings seems to be higher
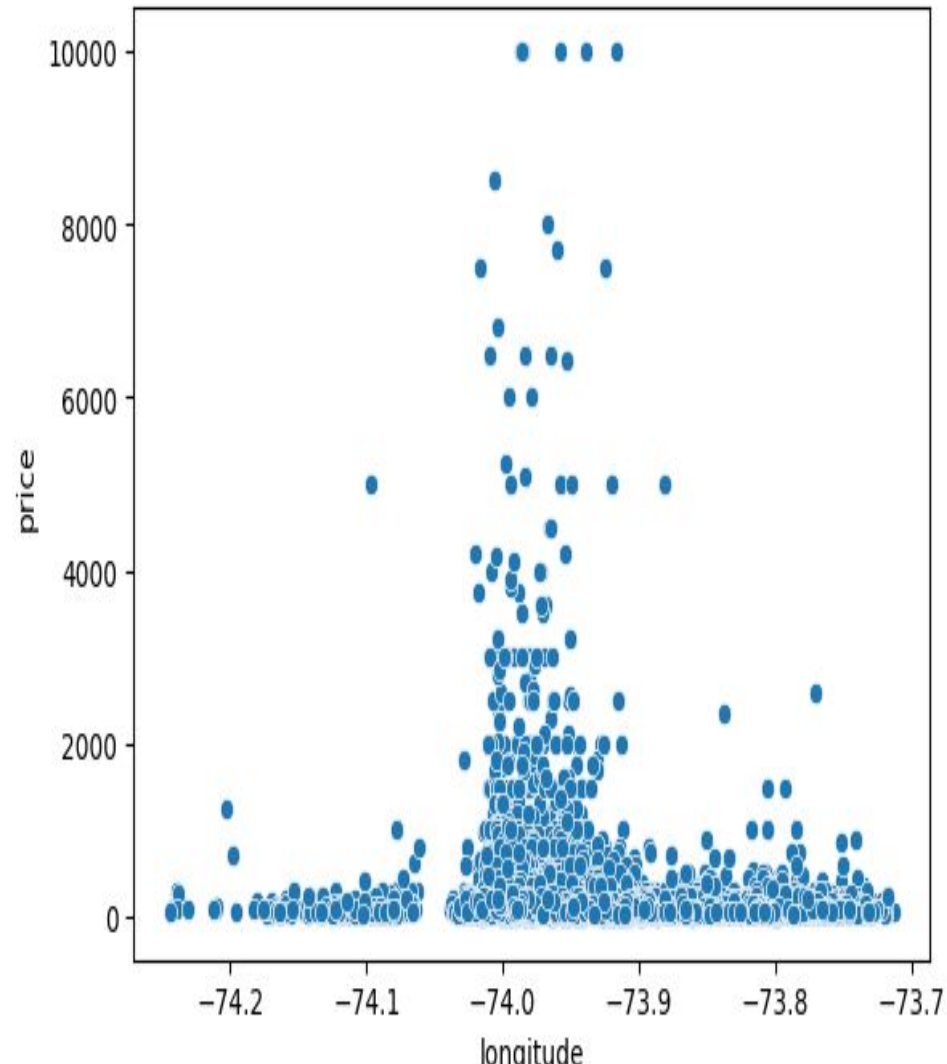between latitudes 40.6 - 40.8

For advertising and setting up your rental business , the
above latitude range is the best

Price vs longitude

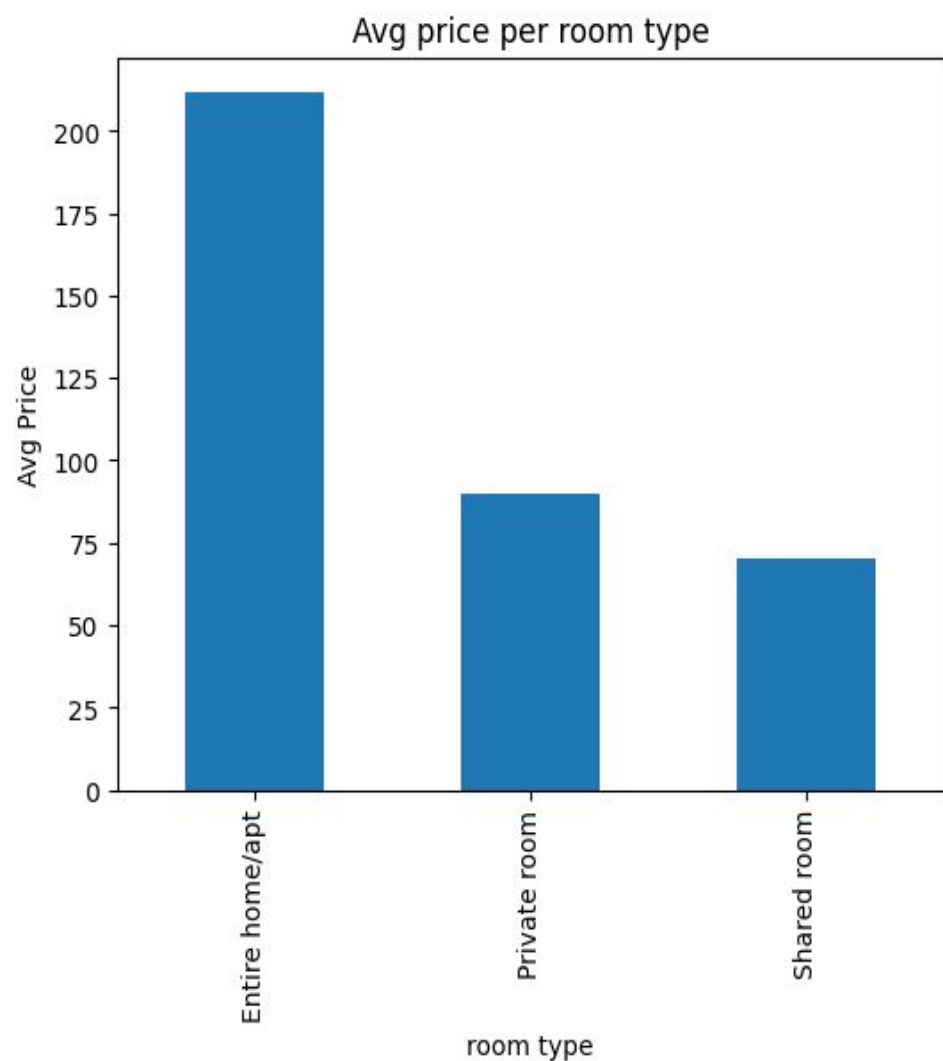Majority of listings lie between longitude -74 to -73.9

The graph shows that the longitude range above has the highest number of renting opportunities & its the best for advertising

Avg price per room_type

Entire homes/apartments are the most expensive, followed by private rooms

For maximum profits , invest in entire apartments/rooms & private rooms

Price vs minimum nights

it seems like listings with lesser minimum nights seems to be more popular & charge more.

seems like 30 nights plan are popular too , followed by 60 & 90 nights plans.
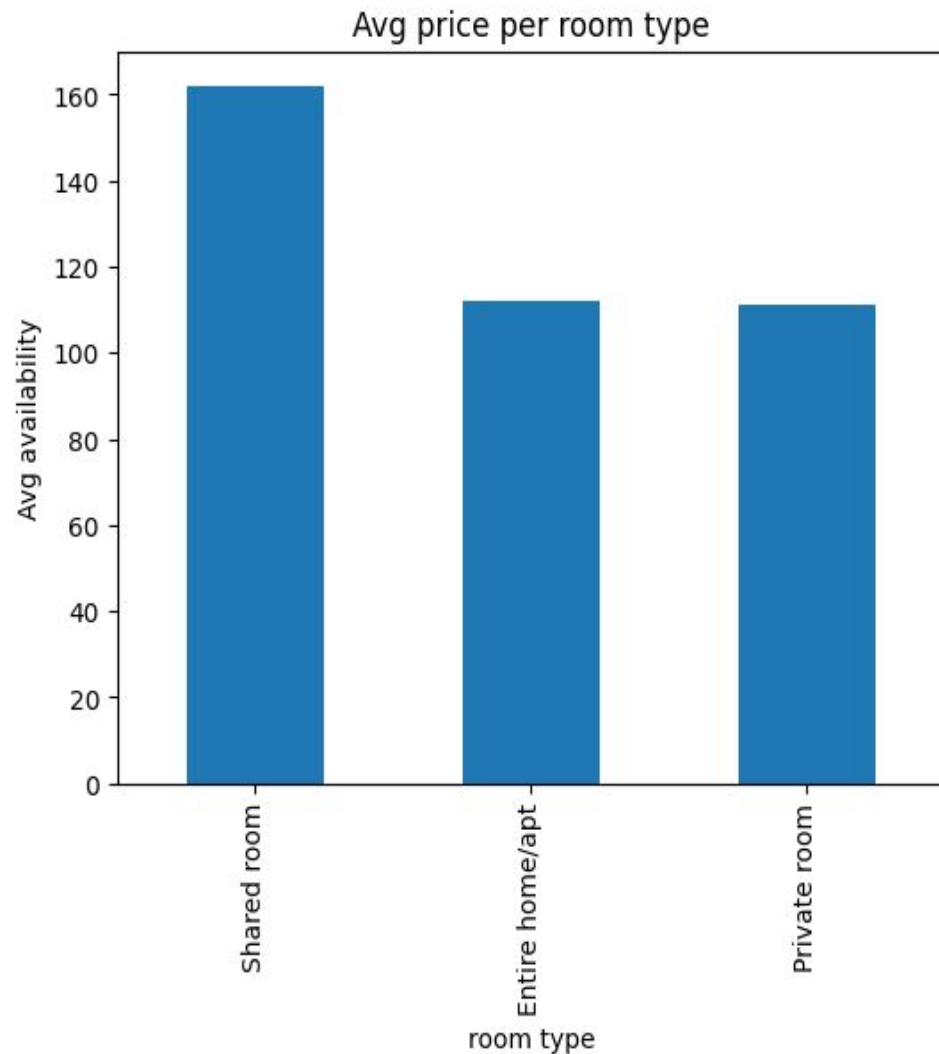
Keep the minimum nights as less as possible while

setting up your rental business. also include 30,60,90

days package
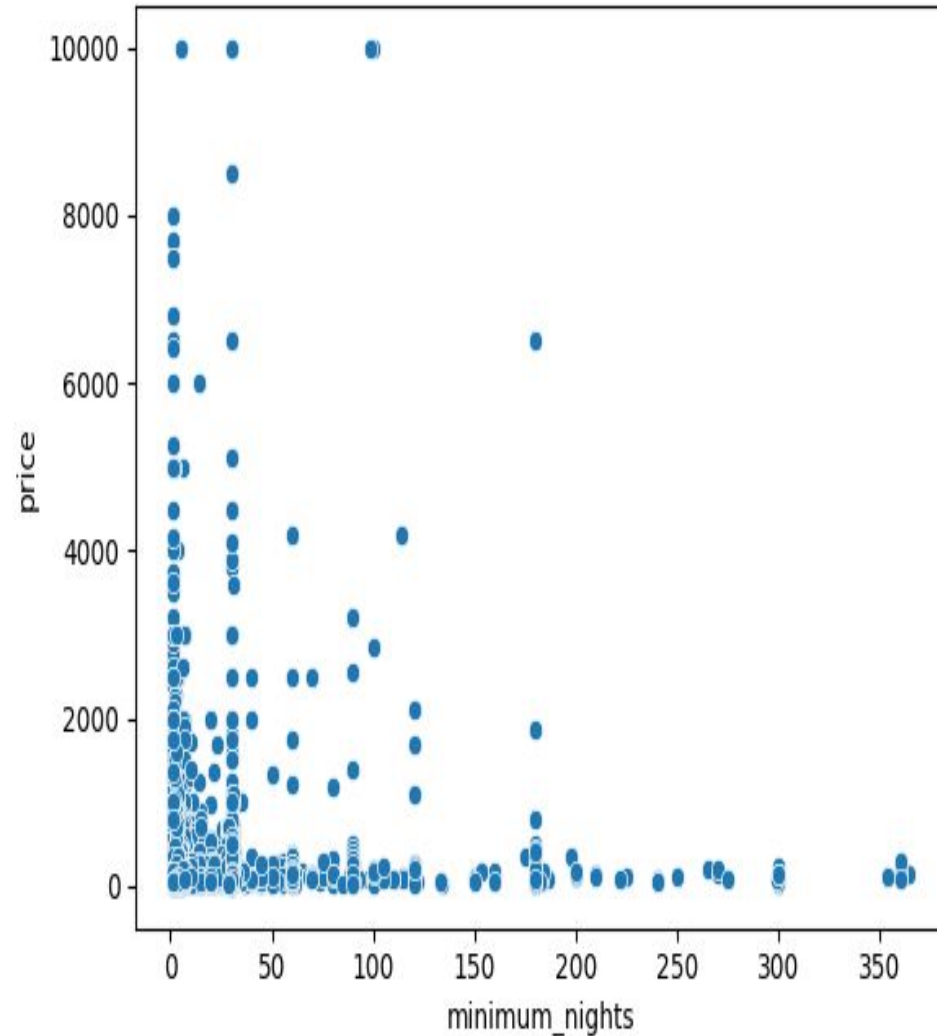
avg Availability per room type

Shared rooms have the most availability , followed by entire home & private rooms

If you want to gain max profits , increase availability in entire rooms since they're more expensive & hence more profitable

Availability per neighborhood group [& neighborhood]

listings in staten islands & bronx seems to have highest availability , followed by brooklyn & queens

avg availability_365 per neighborhood group, neighborhood
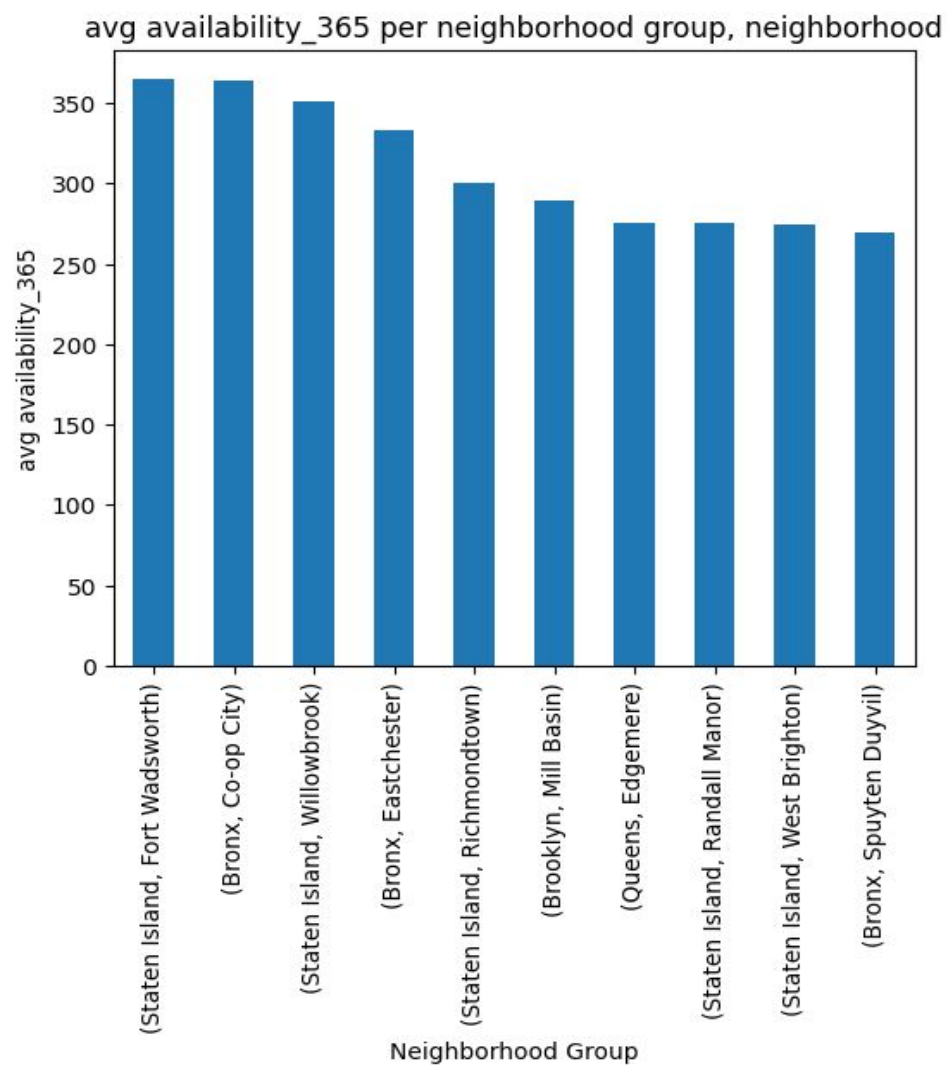
Solutions & Advises for achieving business objectives

Most of listings are in brooklyn or manhattan. If you want to rent out flats , then brooklyn & manhattan are the best places to do it

Entire homes & private rooms are majority of listings. Seems like private rooms & entire homes are in demand

Majority of listings lies between latitudes - 40.65 - 40.75. If you want to advertise your flat you can do it in these locations

majority of listings lie in longitude ranges - 74-73.9. this can help us to find where we should advertise our flats to rent

majority of listings have 0 to 10 min nights, theres also a a considerable number of 30 minimum nights. 0 to 10 minimum nights & 30 nights package seems to be the industry standards

Solution to Business Objective

Staten Island & manhattan are the most expensive , then bronx & brooklyn. If you want to set up a high end apartment for rent theses are the places to do so - Staten islands , manhattan

the prices & number of listings seems to be higher between latitudes 40.6 - 40.8. For advertising and setting up your rental business , the above latitude range is the best

Majority of listings lie between longitude -74 to -73.9. The graph shows that the longitude range above has the highest number of renting opportunities & its the best for advertising.

Entire homes/apartments are the most expensive, followed by private rooms. For maximum profits , invest in entire apartments/rooms & private rooms.

it seems like listings with lesser minimum nights seems to be more popular & charge more. seems like 30 nights plan are popular too , followed by 60 & 90 nights plans. Keep the minimum nights as less as possible while setting up your rental business. also include 30 60 90 days package

Challenges Faced

Understanding & describing the data - variables, data types,

Analysing & dealing with the null values in dataset

Removing Useless columns

Deciding chart types for each case

Deducing inferences from charts

Thank You