Airbnb Booking Analysis

Introduction & Problem Statement

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

This Project will include all the processes that are 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 find valuable insights that we can use to make decisions that will help our organization to increase their profits.

Dataset Discovery

The dataset contains around 49000 rows & 16 columns

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

Steps Involved

1] Data Discovery - this step involves importing the dataset, & gaining some preliminary information. Used functions like describe() , info() to gain insights on the variables. Also found unique values per variable.

2] **Data Cleaning** - this step involves cleaning the data so that it becomes fit for analysis. Includes imputing missing values , removing useless columns. For imputing , we used median [for numerical variables] & 'null' [for categorical variables]

3] **Exploratory data analysis** - This step involves creating graphs & statistics to gain insights from the dataset. Various graphs such as bar graphs, pie chart, scatterplots, histograms have been used in this project. The entire process is divided into 2 phases -

3.1] **Univariate Analysis** - Univariate analysis explores each variable in a data set, separately. It was done on neighborhood group & neighborhood, Count of room type , Latitude , longitude, minimum nights, availability 365.

3.2] **Bivariate Analysis** - Avg Price per neighborhood group & neighborhood, Price vs latitude,Price vs longitude,Avg price per room_type,Price vs minimum nights, avg Availability per room type, Availability per neighborhood group [& neighborhood].

3.3] **Correlation Heatmap -** Correlation heatmap is the best for analyzing correlations between all possible variables.Theres not much correlation between these variables , i.e. they are independent.

3.4] Pair Plot - Pair plot draws all possible graphs for given variables.

Solution to Business Objective

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

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

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

The 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

the majority of listings are available for 0-100 days per year , mosIf you're starting a rental business , then it's advisable to make it available for weekends.t likely weekends [52 weeks per year i.e. 52 weekends i.e. around 100 days]. There are also a lot of listings that are available almost entire year.

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 seem to be more popular & charge more. seems like 30 nights plans 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 a 30,60,90 days package.

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.

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