**AirBnb Booking Analysis**

**AYESHA FIRDOSE W**

**Abstract:**

Online hospitality businesses like Airbnb are a recent phenomenon, their rapid growth has made them a serious competitor in the hospitality industry with significant implications for tourism and tourist destinations. Its potential for further development, and the impact that this development will have on tourism, hotels, and urban destinations. The company has been criticized for enabling bait-and-switch scams, engaging in West Bank settlements, possibly raising house rents, and creating inconvenience for those living near rental properties.

**EDA** can make us understand the role of Airbnb services in different locations of New York City. Also, we will study the different patterns of data affecting the revenue and drawbacks drawn by previous records, how they attack, which weapons they use, etc.

1. **Problem Statement**

* **Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate 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.**
* **This dataset has around 49,000 observations in it with 16 columns and it is a mix of categorical and numeric values.**

**Explore and analyze the data to discover key understandings (not limited to these) such as :**

* What can we learn about different hosts and areas?
* What can we learn from predictions? (ex: locations, prices, reviews, etc.)
* Which hosts are the busiest and why?
* Is there any noticeable difference of traffic among different areas and
* what could be the reason for it?

**From this large amount of data I filtered out some important columns, with which I’m going to analyze the problems. These are the required columns of data mentioned below.**

* **'name'**: It is a column that contains the descriptions/names provided by the respected hosts.
* **'id'**: It is the unique ID given for the statements.
* **'name'**: It is a column that contains the descriptions/names provided by the respective hosts for their customers' reference.
* **'host\_id' & 'host\_name'**: Many hosts offer many properties. These **\*** 'host\_id' & 'host\_name contain records of all the hosts.
* **'neighborhood'** and **'neighbourhood**\_**group'**: These columns contain \* information about the city and areas of properties offered by AirBnb New York.
* **'longitude' and 'latitude'**: As the name suggests, it only contains the \* longitude and latitude of the location of the property.
* **'room\_type'**: Shows the room type of the property (either private room / whole house / shared room ).
* **'price'**: It is an important column that contains the price value of all these properties.
* **'minimum\_nights'**: Gives us information about the minimum number of nights that hosts offer for a specific property.
* **'number\_of\_reviews'** and **'reviews\_month**': Contains information about the number of reviews and reviews per month for these facilities and host hospitality.

**2. Introduction**

Airbnb is an American company that focuses on accommodation, primary homestays for vacation rentals and tourism activities. Based in San Francisco, California, the platform is accessible through a website and mobile app. Airbnb does not own any of the listed properties, instead profiting from the commission income from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

The company is regulated by many jurisdictions, including the European Union and cities such as San Francisco and New Your City.

## **3. Steps Involved in Analysis**

* **Data Wrangling:**

Data Wrangling is the process of gathering, collecting, and transforming raw data into a different format for better understanding, decision making, access, and analysis in less time. It helps us to remove duplicate and unnecessary data to make the further Analysis easy.

* **Exploratory Data Analysis (EDA):**

Once the cleaning and removing is done I performed EDA by comparing different parameters which are involved in the dataset. EDA helps us find various relationships between

parameters. It involves data visualization by comparing different parameters to find out the best of all.

* **Null-Value Treatment:**

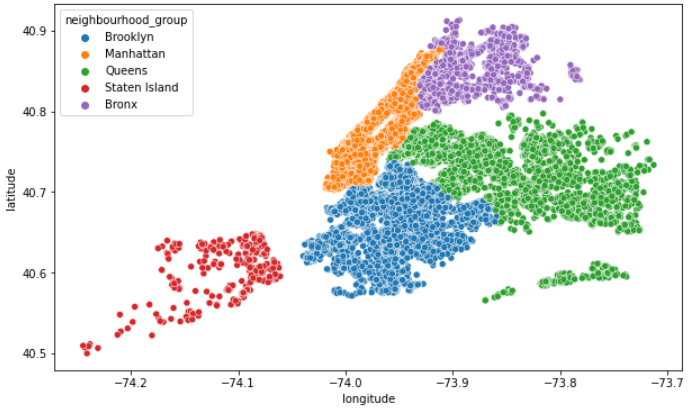
The data set contained many null values which are really not necessary for our analysis but to get the final report we need to treat the null values filling the null values to zero to get the preferred output.

**Outlier Treatment:**

* Outliers are the bad data that may corrupt our analysis, I noticed the price variance to be extreme and so had to treat it with IQR to get the actual result by minimizing the data.

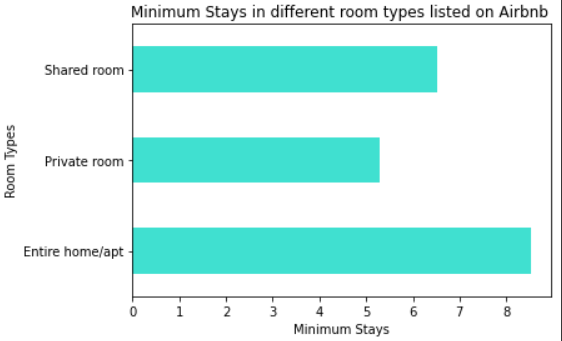
## **4. EDA**

In the process of understanding the data, I first identified the area which was most populated and the reason by taking the longitude and the latitude column.

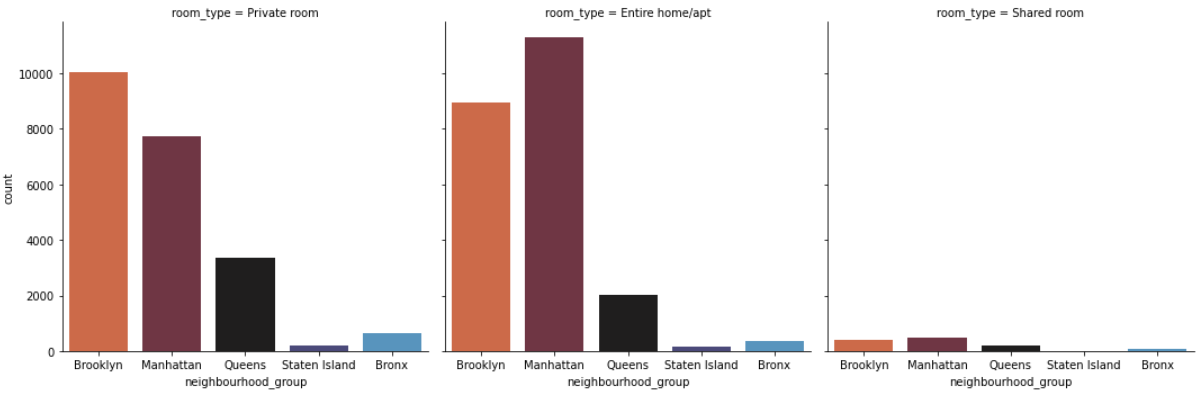


**Manhattan** is the most densely populated neighborhood group by hosts and this may be due to the presence of Manhattan Island bordered by the Hudson, East, and Harlem rivers.

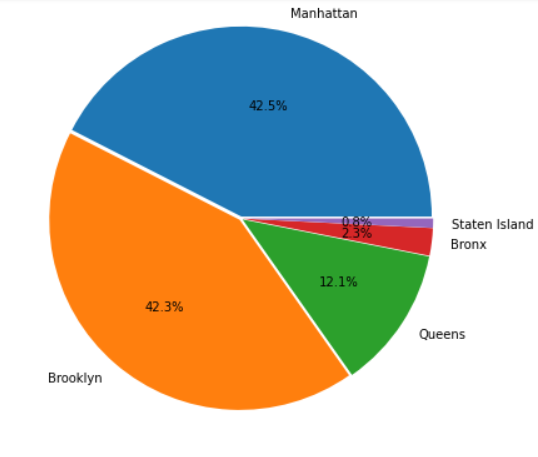
Further Analyzing I found that minimum stay by the tourists was found in **Entire home/apt** room types followed by shared room, as Manhattan is most popular for its tourist attractions and safe for tourists. They have a very welcoming environment and hosts also follow the new trending service to offer it to the tourists and the listed properties on Airbnb.

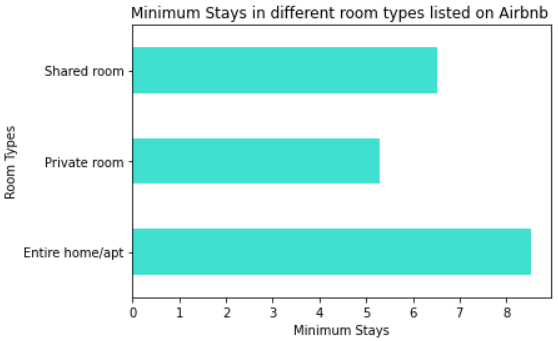


After getting the estimation of the room type stayed with nights, we may also get in which location which room type is preferred more, and here is the data I found.

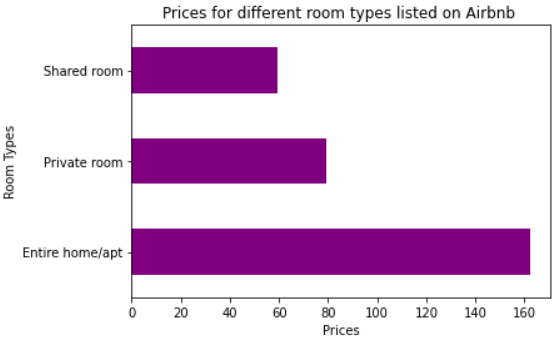


As we can further see the bookings are high in Manhattan followed by Brooklyn, compared to Queens, Bronx, and Staten Island. Staten Island is less preferred by the guests on AirBnb as it is also considered the” forgotten borough” as it may be a neglected neighborhood place compared to all of the other locations by the residents and the government.

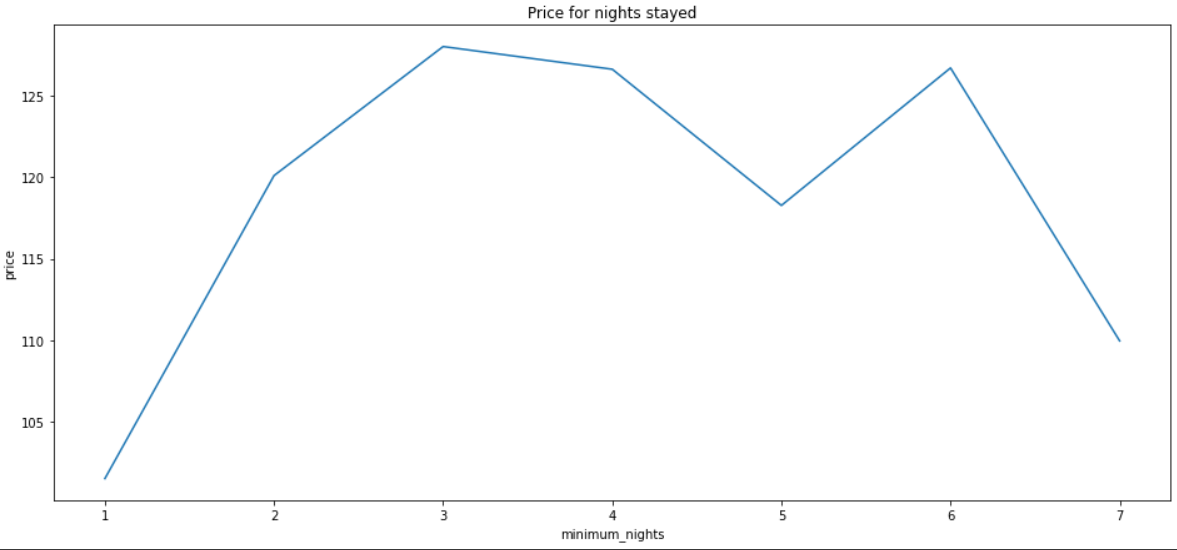




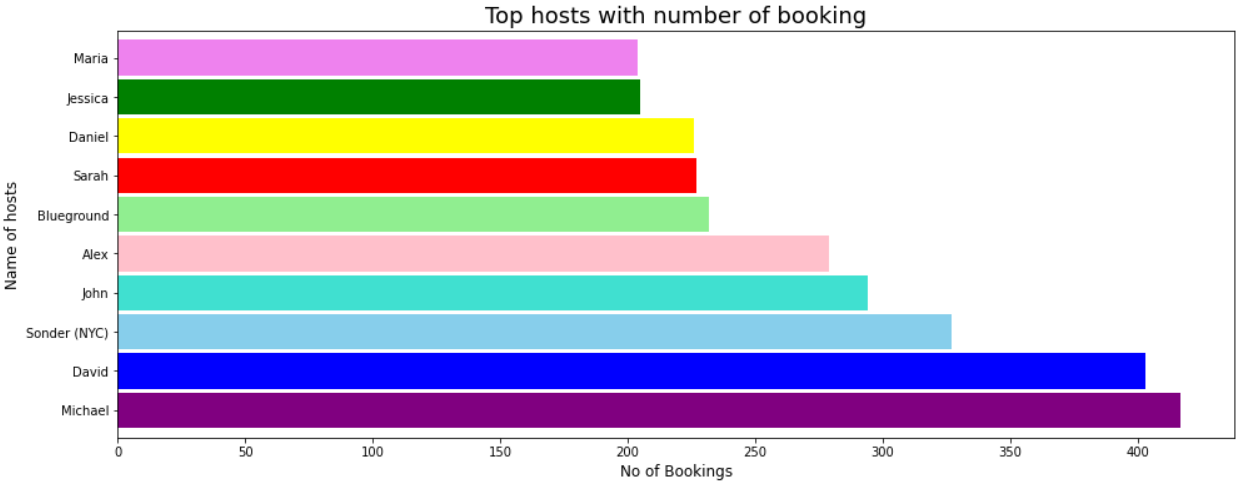
We can see that guests prefer the Entire home/apt to stay for a minimum stay as it. Guests do prefer the Entire home/apt as you get access to the whole house and there will be no one else in it except the people who are with you. So, there will be extra privacy matters that the guests keep in mind when they make bookings via AirBnb.



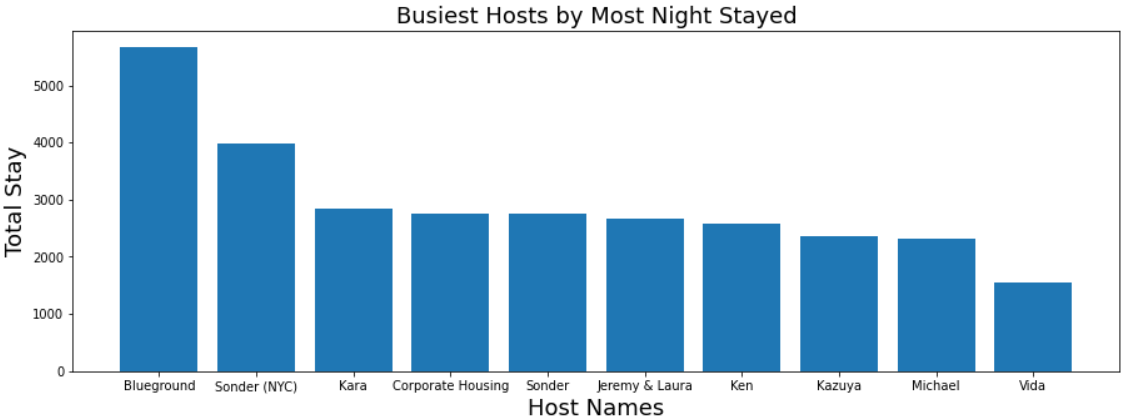
We saw in our previous analysis that most guests prefer Entire home/apt compared to the other two and it is clearly understood that Entire home/apt is more in demand and the prices offered to such rooms on AirBnb will exclusively be higher.



This is a weak trend, we can clearly note the price is slightly getting decreased as the nights are increasing, which means the more they stay, the prices are likely to get decreased as an increase in nights. So, if guests are booking for 1 or maybe 2 nights, there are chances the prices may seem too high on AirBnb platform and likely to be less when the days stayed are more.



With reference to the busiest host in our dataset, I noticed that the hostname **Michael** appears 182 times with respect to the Manhattan neighborhood and 157 rows with respect to Brooklyn, which clearly states that **Michael** is the busiest host among all the neighborhood groups.



When Analyzing the parameter for measuring the busiest host as the total number of nights In the list of hosts according to guests, we see the top 5 hosts as follows:

Blueground

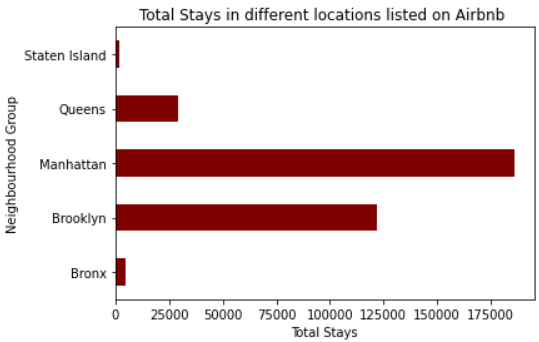
Sonder (NYC)

Kara

Jeremy & Laura

Sonder

We can see here that 'Blueground' is the busiest host and the reason these hosts are the busiest of all is that they belong to the Manhattan neighborhood group and have the Whole House/Apt room type where demand is at its peak. higher price due to its location.



As we can see total nights stayed by the customers are more in Manhattan, Brooklyn, Queens and clearly, we can state that they are the most traffic observed areas as many hosts also belong to the respected neighborhood groups. The more the nights stayed more the traffic observed in these areas as we earlier saw that Manhattan is the busiest group and many hosts belong to that.

**5. Key Findings**

* Hosts from the Manhattan neighbourhood have a higher number of properties listed on Airbnb.
* The Manhattan and Brooklyn neighbourhood groups hold 85% of the total reservation, while Staten Island has the fewest reservations observed.
* The whole house/apt is the most preferred room type, followed by a private room.
* Queens and Staten Island have significantly fewer host listings than Manhattan. So we should take sufficient steps to support Queens and Staten Island host lists.
* Total nights stayed by the customers are higher in the Manhattan group.

**6. Conclusion:**

Finally, here we come to an end of our Analysis, and we can see that Manhattan is the most highly preferred place for tourist attraction as it is highly preferred by the customers and many hosts belong to that place, the reason why Manhattan is most preferable is that it is considered as one of the world's foremost commercial, financial, and cultural centers. So, most of them prefer to stay in Manhattan

We also observed Brooklyn is the second most preferred as this area is definitely worth exploring as it's home to iconic New York attractions like Coney Island, Brooklyn Bridge, and Prospect Park, plus some of New York's best pizza.

Queens, Staten Island, and Bronx are less preferred as the number of nights stayed and the hosts are also very low, it may be because the tourists attract the famous spots which are more advertised, and those are the less famous destinations

Staten Island has sometimes been called the "forgotten borough" by residents who feel neglected by the city government, tourists are attracted to coastal and Island areas because they get a peaceful atmosphere and a clean and green environment there. They get the relaxation to leave their hectic schedule.

Also in the column of minimum\_nights, we observed that in our dataset there were 1 or 2 nights and the prices for those nights were 0, we may tell that AirBnb provides free nights to their Elite or Premium members to attract customers so that they may prefer AirBnb to make their bookings for the next time.

**References-**

1. GeeksforGeeks
2. StackOverflow
3. Pandas.pydata.org
4. https://www.python-graph-gallery.com