Capstone Project-1
Air Booking Analysis

Team Members

1)Pranil Satish Thorat
2)Saransh Srivastava
3)Jai Harish S

AI

Overview of Air booking Analysis:

Introduction:

- Airbnb is an online marketplace for lodging, principally homestays for vacation rentals, and tourism activities based in the United States. Airbnb does not own any of the houses advertised; instead, it makes money from commissions on each booking.

- Airbnb is a database of accommodation bookings. This data is only available for New York City from 2011 to 2019.

- We can grasp all of the features by exploring the data. The objective is to look into the data and then evaluate it with all of the necessary information.

- We have extracted crucial insights from the data through data exploration and data analysis.

- Airbnb is an abbreviated version of AirBedandBreakfast.com, which was the company's

Abstract:

- We have extensive experience with Airbnb Booking Analysis.

- We intend to concentrate solely on New York City.

- The investigation is mostly focused on the boroughs of Brooklyn, Manhattan, Queens, Staten Island, and the Bronx.

- This data includes information about the host, lodging prices, and reviews, among other things.

- With the use of exploratory data analysis, data wrangling, visualization, and other tools, we can extract a lot of information from the analysis.

Problem Statement:

- The customer can book a room through the AirBedandBreakfast.com (airbnb website). To explore more, the customer is moving to the Queens, Bronx, Staten Island, Brooklyn, and Manhattan neighborhoods.

- The customer requires assistance in identifying a number of bookings in each borough country and guiding that customer in selecting the best place to live at an affordable cost, analyzing the room's price so that the customer can afford it, and selecting the appropriate room type based on their needs.

- The quantity of reviews, the number of reviews per month, and the availability of the rooms will all meet these criteria.

- All of these insights will be obtained through EDA and a thorough comprehension of the data.
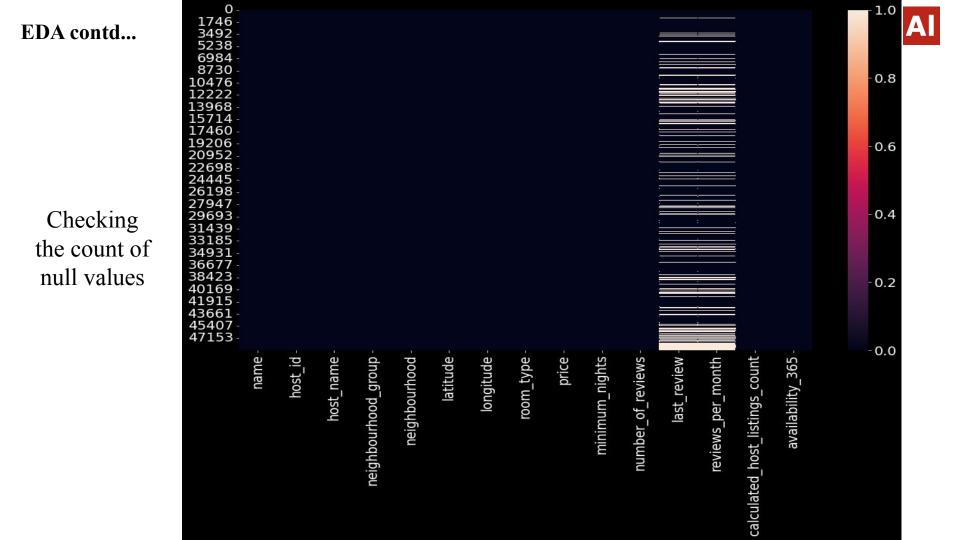
Steps Involved:

1. Data Wrangling

2. Performed Univariate and Bivariate Analysis

3. Checked correlation between each features

4. Treatment of Null values

5. Treatment of Outliers

6. Explored Visualization

7. Transformation
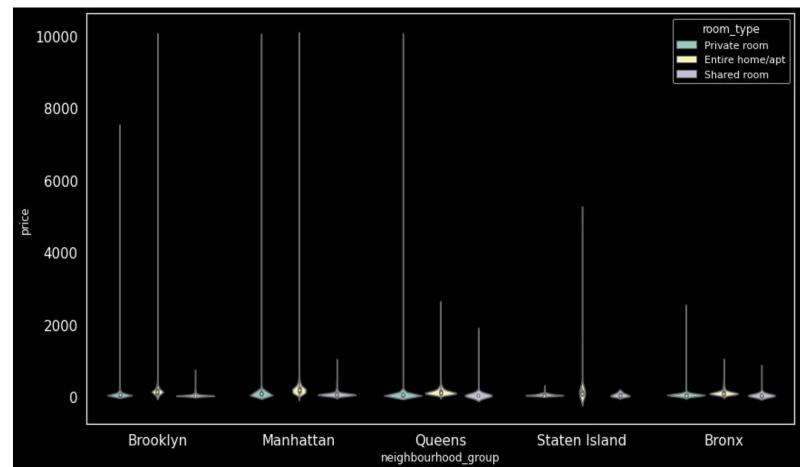
8. Statistical Test

Exploratory Data Analysis:

- Using the pandas package, The data is loaded. Operations performed such as casting the columns into their appropriate data types, to better comprehend the data.

- Statistical data, univariate analysis, and the accuracy of data in each column were all investigated.

- Understanding the relationship of all numerical and categorical columns. In addition, we performed Bivariate Analysis on both numerical and categorical data.

- Performed Univariate analysis for numerical columns to grasp distribution and information and see if it follows a Gaussian distribution.

- Used a variety of plots to gain a better understanding of the data and presented it in an attractive way .

Checking
the count of
null values

Relationship between Neighborhood country vs price

Relationship between all features

EDA contd...
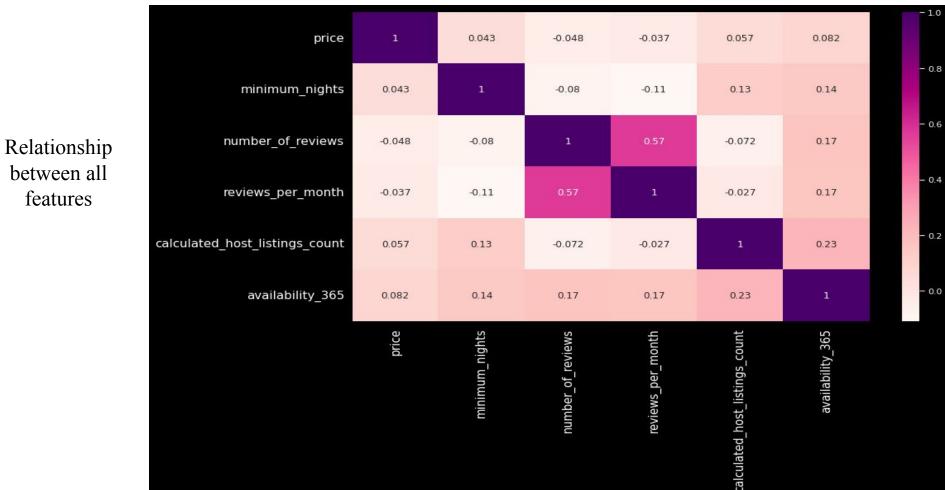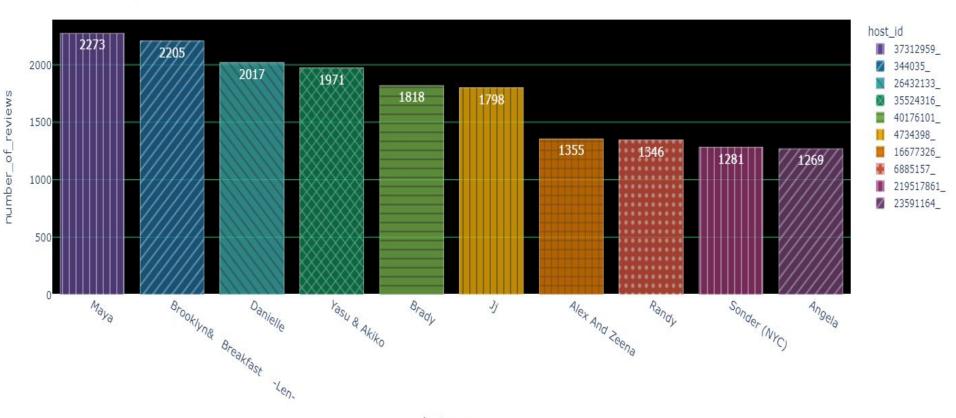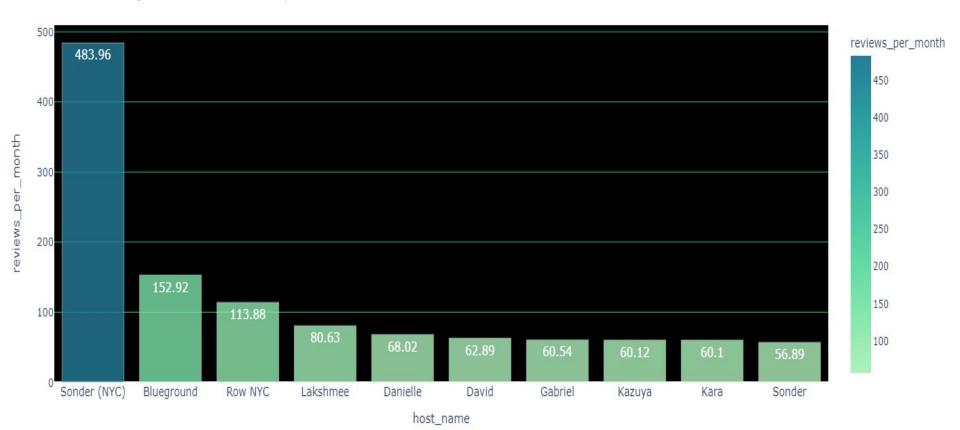

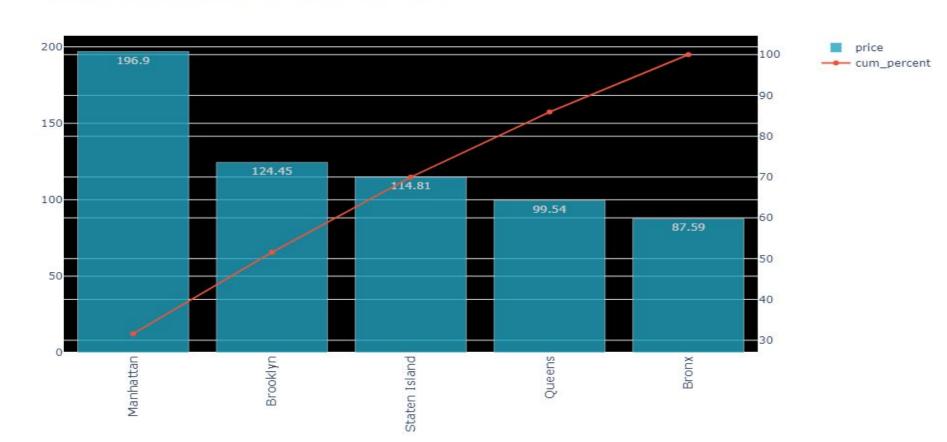Host has the highest number of reviews
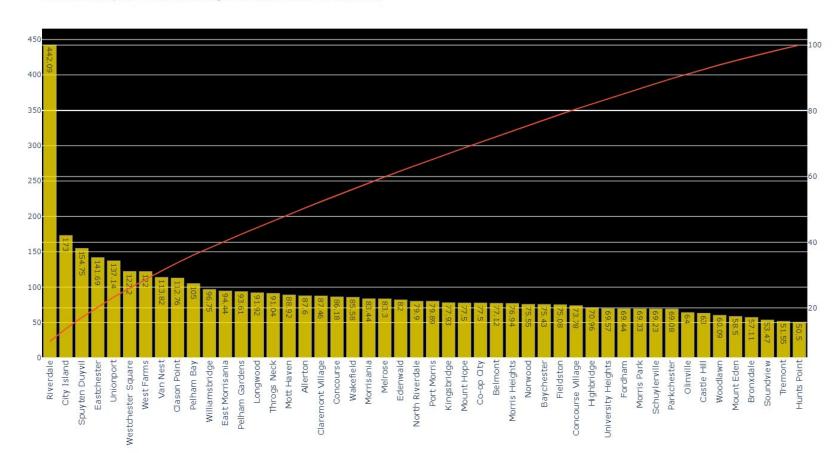
Host has the highest number of reviews per month
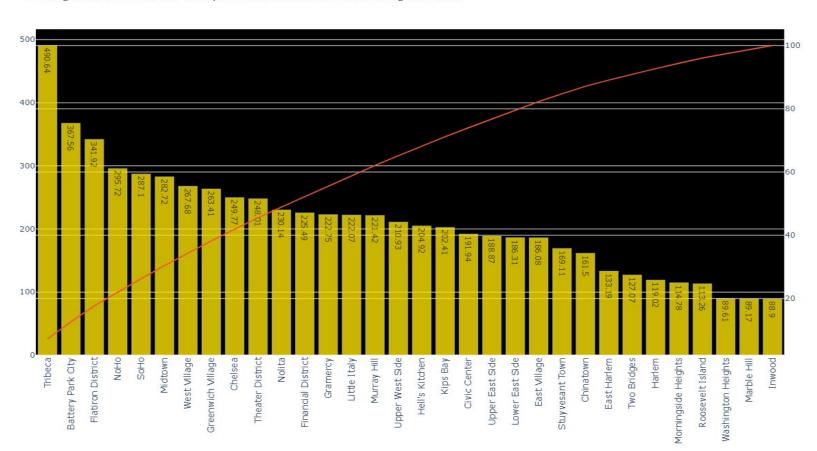
Average price ditribution of Neighbourhood group

EDA contd...

Bronx is Cheapest area and Average Price distribution of that area

EDA contd...

AI

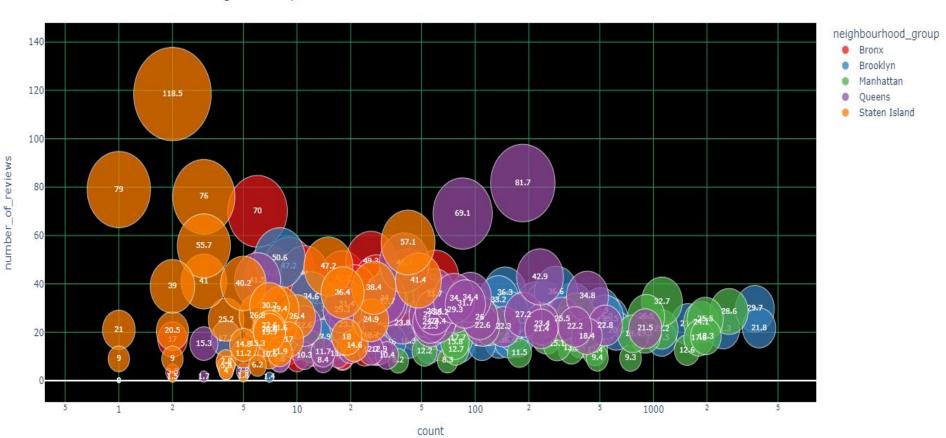Average Price distribution of Expensive areas in Manhattan Neighborhood

Total Bookings of Borough Country vs Average Price

EDA contd...

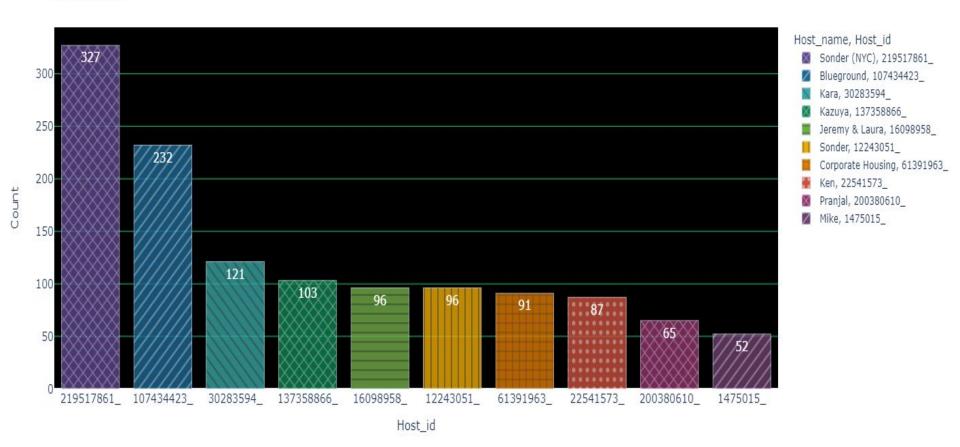Number of reviews for Each Neighborhood place

EDA contd...

Busiest Hosts

EDA contd...
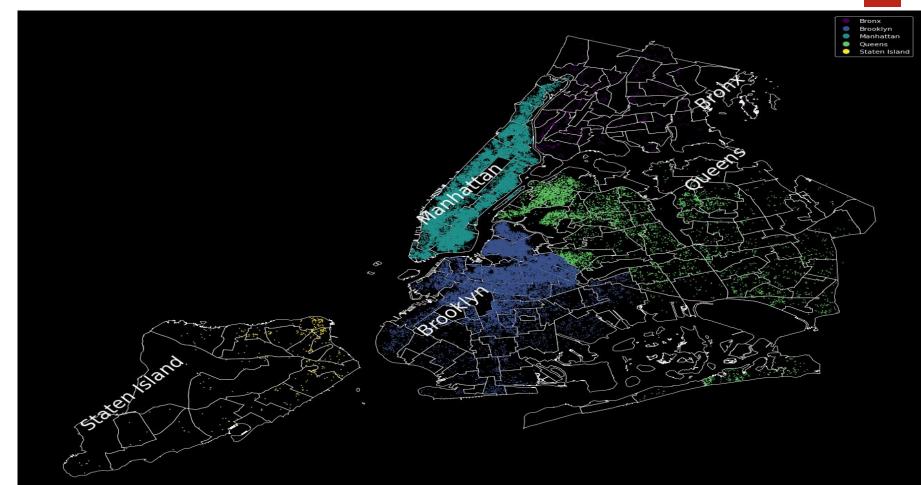
Busiest Hosts Rental place with bookings count and number of reviews
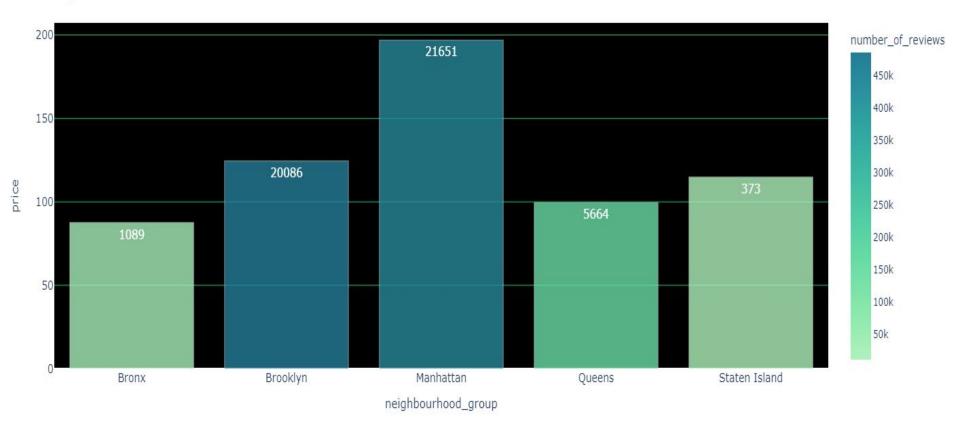
EDA contd...

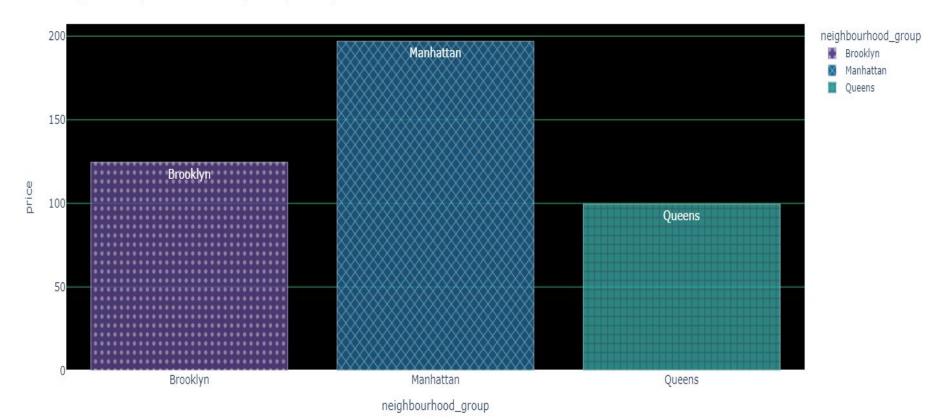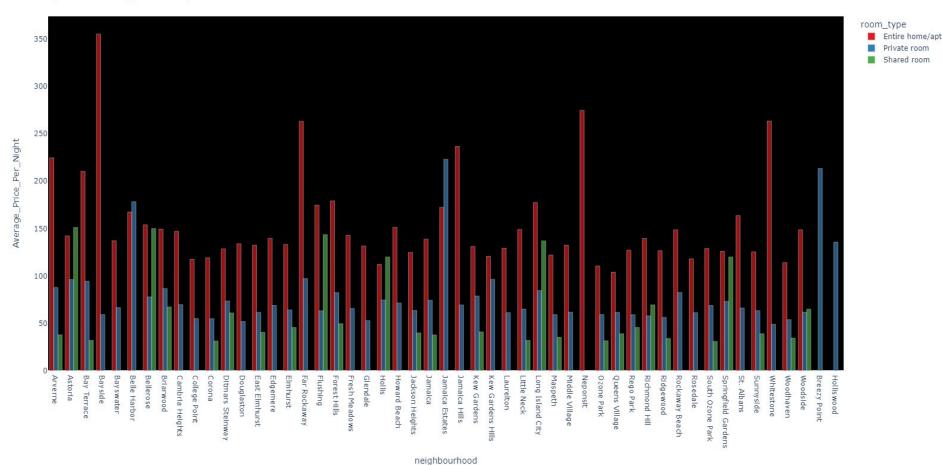Neighborhood Traffic
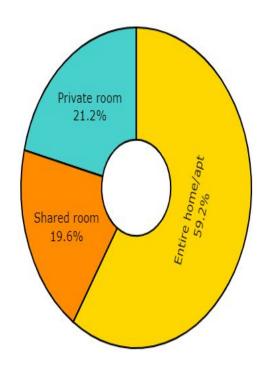
Neighborhood price of Manhattan, Brooklyn and Queens

EDA contd...

Queens Room type and its places to visit at affordable Cost
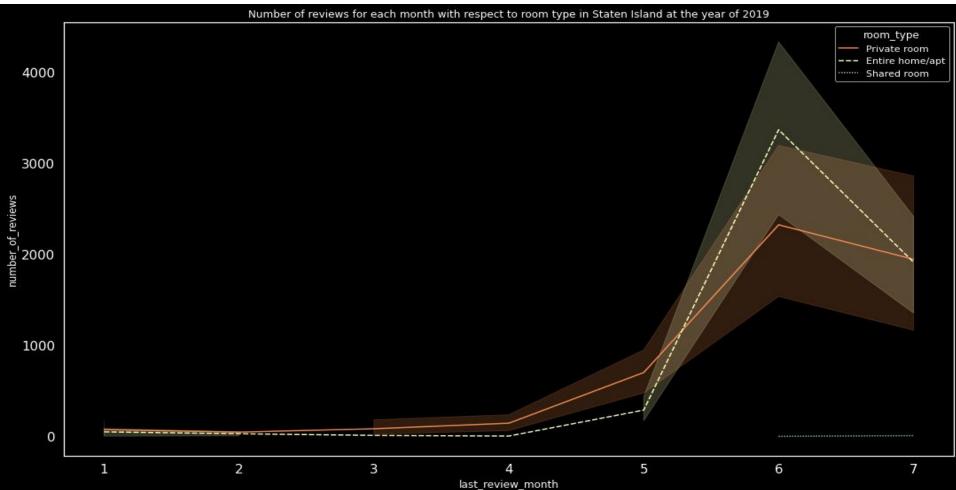
Luxury Rooms in Staten Island and its Room type

Private room
21.2%

Shared room
19.6%

Entire home/apt
59.2%
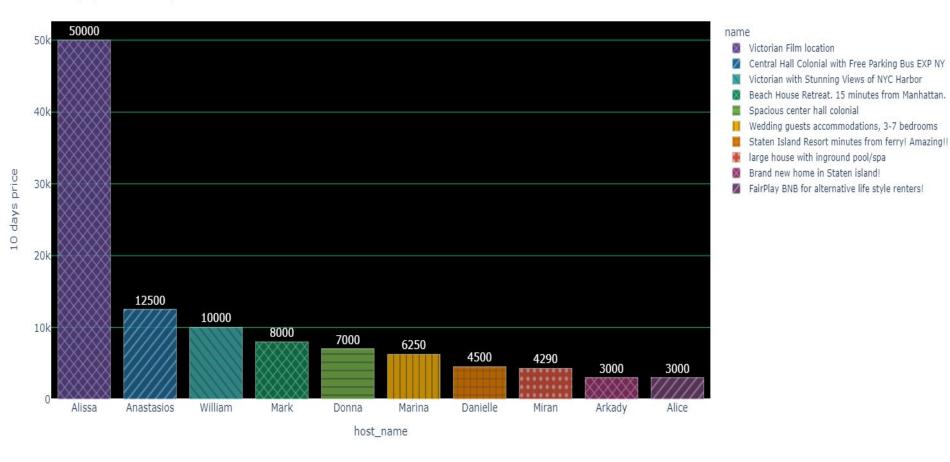
- Entire home/apt
- Private room
- Shared room

Number of reviews for each month with respect to room type in Staten Island at the year of 2019

EDA contd...

AI
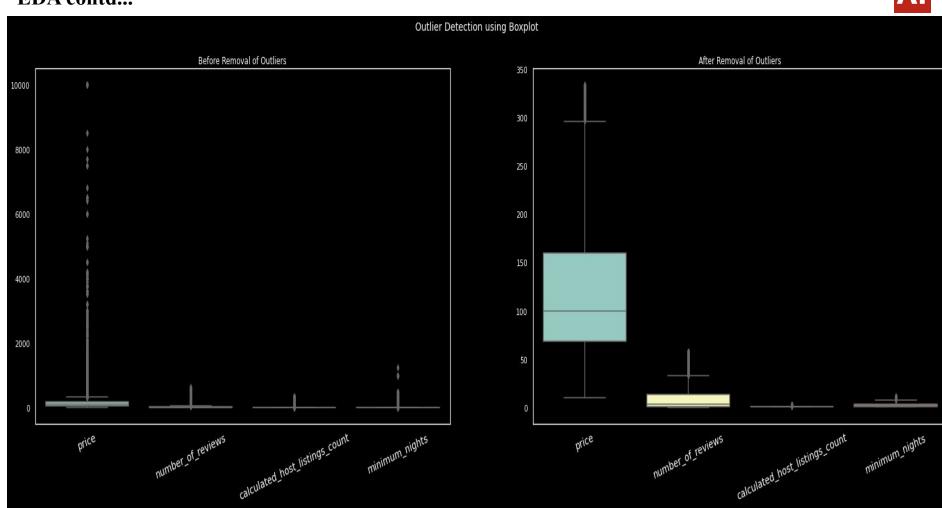
10 days price of Luxury Hotels in Staten Island with their Host Names

name
- Victorian Film location
- Central Hall Colonial with Free Parking Bus EXP NY
- Victorian with Stunning Views of NYC Harbor
- Beach House Retreat. 15 minutes from Manhattan.
- Spacious center hall colonial
- Wedding guests accommodations, 3-7 bedrooms
- Staten Island Resort minutes from ferry! Amazing!!
- large house with inground pool/spa
- Brand new home in Staten island!
- FairPlay BNB for alternative life style renters!

EDA contd...

Outlier Detection using Boxplot

EDA contd...

Transformation into Normal Distribution by applying Central Limit Theorem

Statistical Tests:

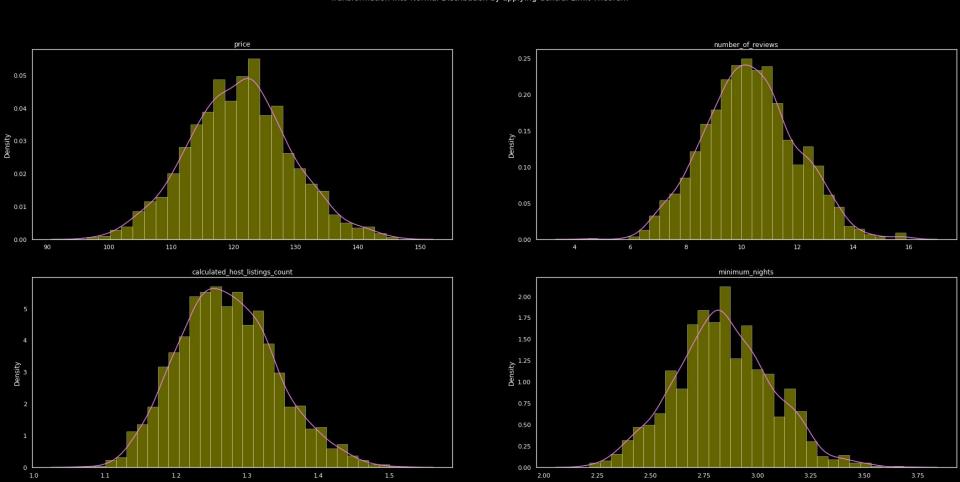- A statistical test is a tool that allows you to make quantitative conclusions about a process or set of activities

- Our goal is to do a statistical test to see if the prices of rooms in different neighbourhoods are similar or not.

- To do so, we built a group of neighbourhoods and took the mean of the prices, then chose two neighbourhood groups with prices that are closer to each other, then we used Hypothesis Testing to get what we wanted. When the prices of rooms in neighbouring groups are equal, the Null Hypothesis is accepted, and when they are not, the Alternate Hypothesis is accepted.

- **Result of the Hypothesis were:**
  - ·   price of rooms in Bronx and Queens are not equal
  - ·   price of rooms in Brooklyn and Staten Island are not equal
  - ·   price of rooms in Queens and Staten island are equal.
  - ·   price of rooms in Manhattan are not equal to any other.
- Another Statistical test we have done to find out whether the room type are similar to neighbourhood groups so to find out we have used Chi Square test.
  - ·   HO: Not Related (no association)
    - ■   o   alpha > 0.05
  - ·   HA: Related (association)
    - ■   o   alpha < 0.05
- *Chi-square = $\sum (fe-fo)2fe$*
- Our goal was to find out the value of alpha and alpha value was less than 0.05 **so we can say neighbourhood groups and room types are related.**

Conclusion:

- Exploratory data analysis was used to aid. Customers can use the data to study and comprehend additional information.

- EDA assigns a host name based on the number of reviews and the amount of reviews every month to identify the best and most customer-friendly hosts.It also shows the host who has availability of 365 days.

- Customers can select the busiest host and learn why they are so busy. EDA provides each host with their own renting space

- It also provided useful information about Staten Island for his upcoming trip by analysing the number of reviews and luxury lodging options.

- With the help of these EDA, a greater amount of valuable information is provided to customers and make better decisions.