

CAPSTONE PROJECT-EDA

Airbnb Booking Dataset

Brief description of Airbnb online booking platform

- Airbnb is a online booking platform
- Here all kinds of basic amenities required by travellers are offered by residents.
- Residents who want to rent their house,car, entire home to make profit will provide their details on the platforms available.

Airbnb Dataset problem statements:

- Which are the neighbourhood groups tourists are preferring to have more visits?
- Dataset airbnb describe host information who are facing less bookings?
- What are the type of rooms having more availability in the neighbourhood?
- What is the relationship between availability of rooms and price ?

Problem Statements continued:

- What are the names of hotels which are face high bookings but having less reviews according to the dataset?
- Check the customers who stayed for short or long term?
- Verify the reviews column with respect to stay of the customers

Price versus Availability analysis:

- Every online business works based on factors like availability, site effectiveness, and price comparisons.
- Here image beside shows top ten records of price column when grouped with availability column.
- As the availability of rooms increases price increases.
- Here the price starts from 136 units.

availability_365

0	136.032111
1	129.509804
2	145.944444
3	136.486928
4	135.759657
5	134.179412
6	123.612245
7	135.123288
8	131.047210
9	146.445596

Name: price, dtype: float64

Analysis on Availability:

- Here the image shows the availability of rooms and their prices.
- We can draw a conclusion that till we reach 365 days availability.
- The price mean has reached to 250 units.

```
availability_365
356    138.000000
357    177.536842
358    212.594444
359    180.925926
360    154.970588
361    157.108108
362    201.096386
363    200.644351
364    260.372709
365    250.769884
Name: price, dtype: float64
```

Analysing Hotels and Customers:

- Here i am trying to retrieve hotel names whose customer bookings were high but average rate of review was less.
- Hillside Hotel although had received high bookings is also lagging to take better reviews.

Hillside Hotel	18
Artsy Private BR in Fort Greene Cumberland	10
IN MINT CONDITION-STUDIOS EAST 44TH/UNITED NATIONS	6
A CLASSIC NYC NEIGHBORHOOD-EAST 86TH/5TH AVENUE	5
Are You Ready for Central Park State of Mind?	4
Modern, Cozy Apt Share for Young Professionals	4
♥ of Manhattan Fantastic 1 Bedroom	4
Sonder The Biltmore Spacious 1BR + Kitchen	4
Sonder The Biltmore Stunning 1BR + Sofa Bed	4
Cozy apartment by Central Park	3
Name: name, dtype: int64	

Analysis room type and mean price:

- Room types according to the dataset are Entire home, Private room and Shared room.
- Image beside gives insights of average price for every room type offered.
- Entire home has 215 , Private room and Shared room have close price mean of about 86 and 78.

```
room_type
Entire home/apt    215.746181
Private room       86.903289
Shared room        78.978102
Name: price, dtype: float64
```


Analysis of minimum nights and number of reviews by customers

- Here i am trying to fetch records of customers who stayed for 1,2 days and comparing their reviews.
- Count of reviews is 12720 of customers who booked for just one day.

minimum_nights

1	12720
2	11696
3	7999
4	3303
5	3034

...

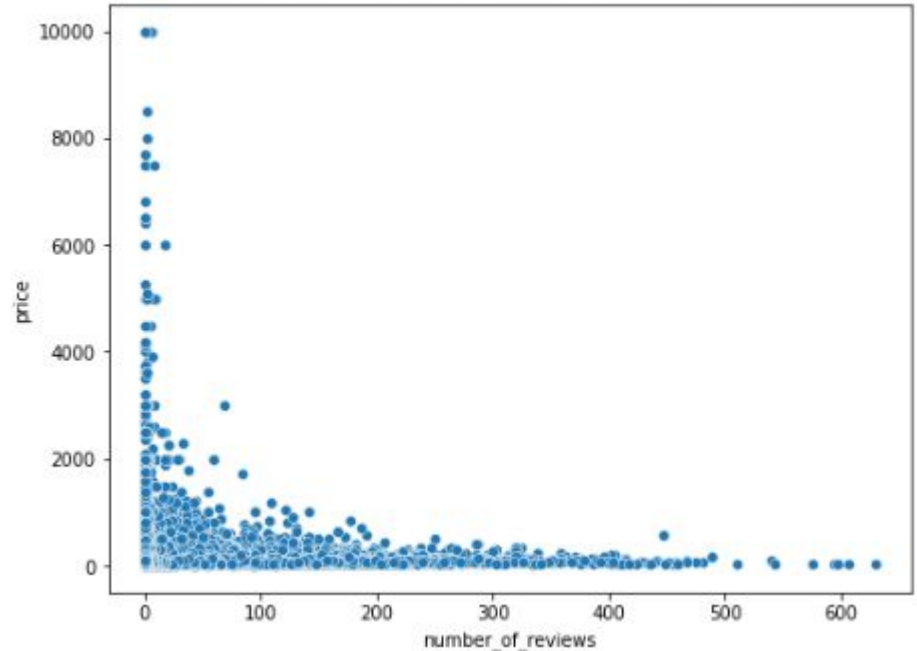
480	1
500	5
999	3
1000	1
1250	1

Name: number_of_reviews, Length: 109, dtype: int64

Reviews as per price range:

- Image shows a scatter plot to visualize the distribution of price and reviews columns.
- Most of the dots are around price below 2000.
- There are only two reviews around price 10,000.
- By this we can assume high price rooms are facing less booking issues.

<matplotlib.axes._subplots.AxesSubplot at 0x7f4ef3b42190>



Relation between per month reviews and price:

- Here the image beside gives more information about the impact of reviews on price.
- Now we can conclude that the high bookings are around lowest price rooms.
- Hence the reviews are also recorded around the low price region.

<matplotlib.axes._subplots.AxesSubplot at 0x7f4ef3cfdc50>

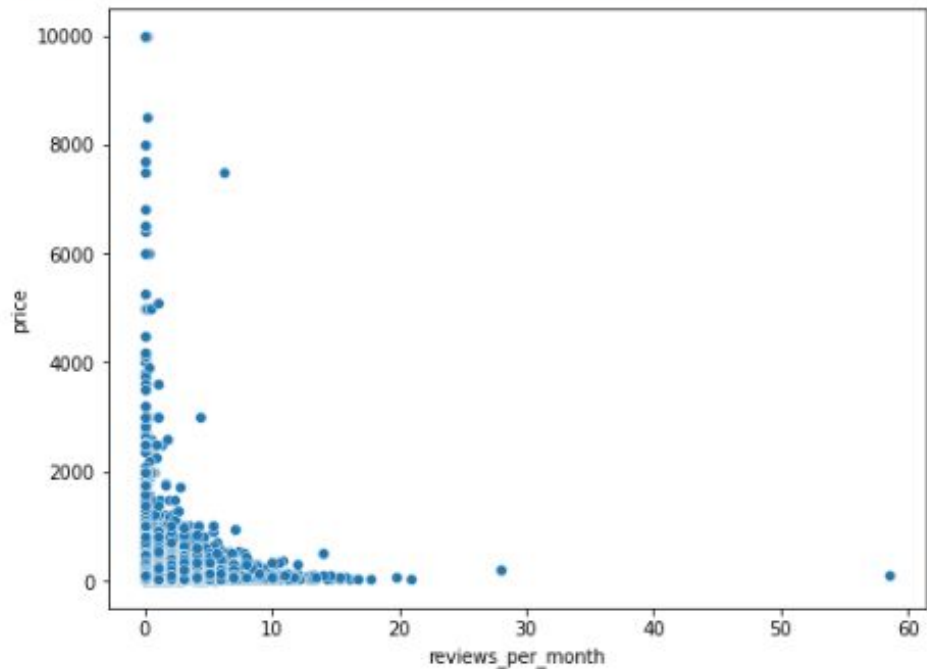
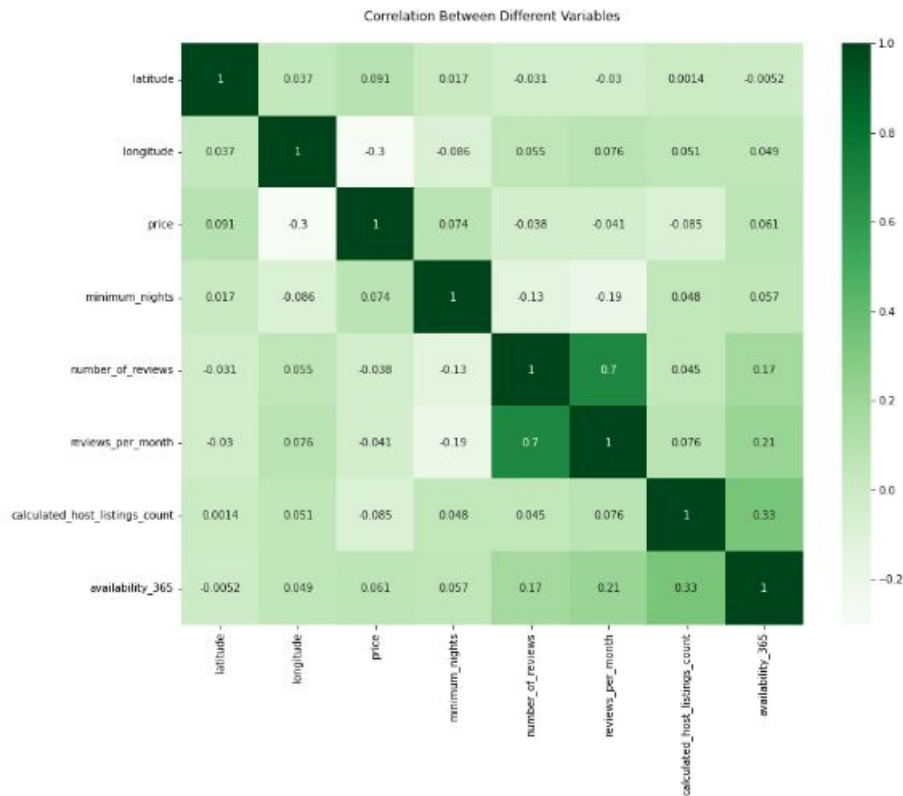


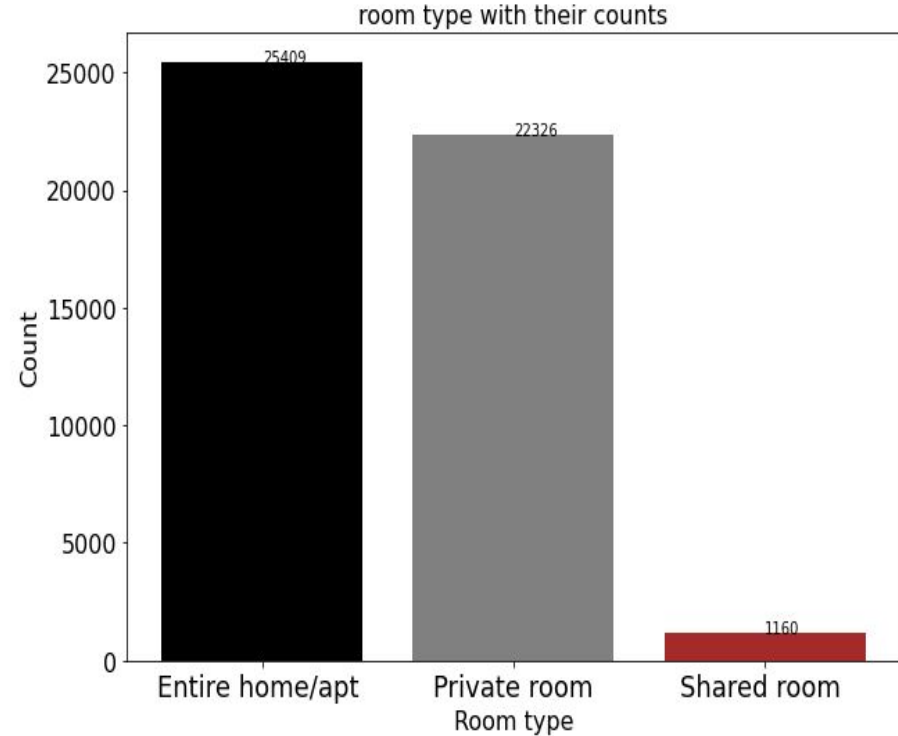
Diagram to show correlation among columns:

- Correlation matrix gives a measure of columns in the form of numeric.
- Here 1 gives high correlation on self columns.
- Longitude and price are showing relatively high correlation compared to other columns from data set.



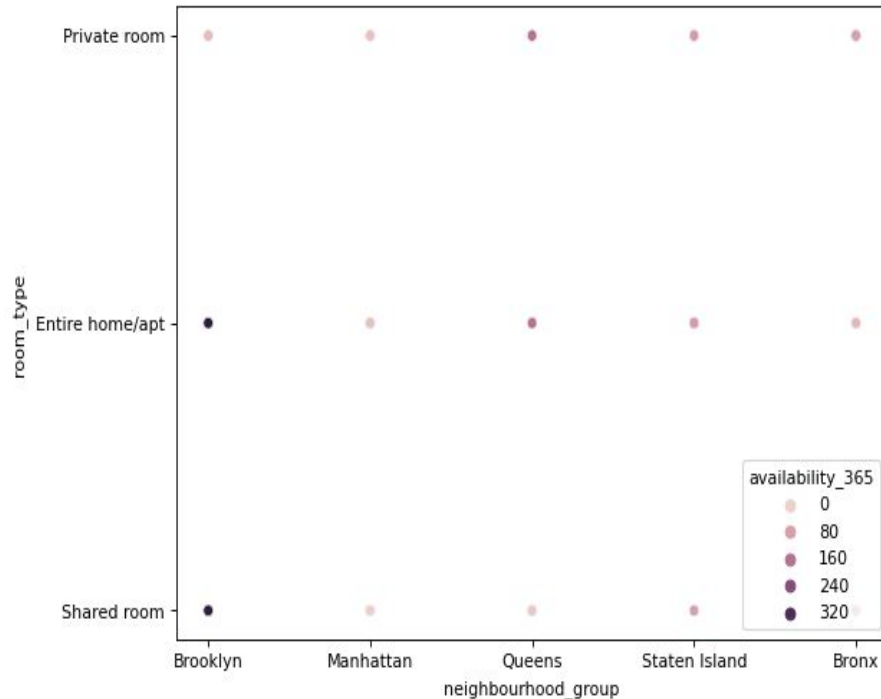
Graph to show count of room types:

- Image beside gives clear view of room type count.
- According to the dataset Entire home has highest count 25400 and Shared room has lowest count of 1180 as per the airbnb data set.



Availability of rooms in the neighbourhood group:

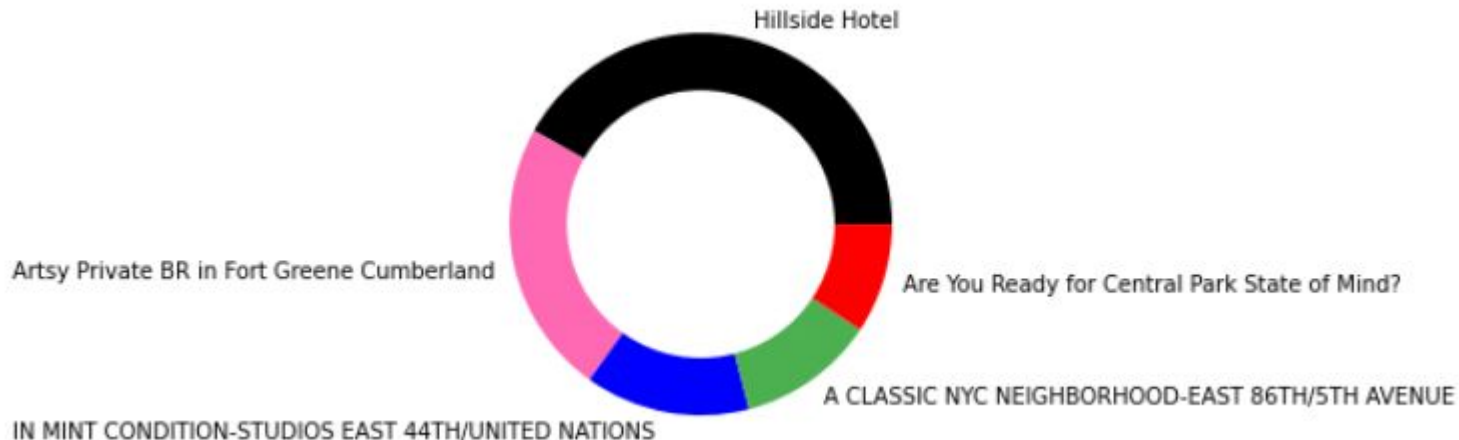
- Image shows a scatterplot of the availability of room types Entire home, Private room and Shared room in the neighbourhood group.
- Brooklyn is been recorded with high availability of Shared and Entire home of about 320 units.



Exploring the hotel names which are busiest:

Hotels which had more bookings

Above average bookings with below average reviews



Analysis over hosts:

According to the records in dataset there are few single reviews.

These could be due to single night stay of customers.

Hence analysis over the hostname column grouped by their stay and review count.

These shows some unhappy customers who never booked again.

```
host_name
(Email hidden by Airbnb)    1
A                             1
A-B                           1
AFI Apartments               4
Aamir                         1
..
思铨                         1
睿                             1
诗月                         1
빈나                         1
소정                         1
Name: minimum_nights, Length: 2461, dtype: int64
```


Conclusion:

- Exploratory analysis of Airbnb dataset has significance as below
- In the dataset we resolved the relationship between the neighbourhood groups which are facing more tourists.
- Price of the rooms and availability of rooms has direct relationship, high availability of rooms has more price comparing to low availability.
- Even though receiving more bookings the reviews offered by customers are very less, Hence care should also be taken for customer satisfaction which has high impact on further bookings.
- Airbnb is a online business so reviews and availability have great impact on business.