**Airbnb Bookings Analysis**

**Technical documentation**

**Biswajit Paul**

**Shreya Pattanayak**

*Data science trainees - AlmaBetter, Bangalore*

**Abstract:**

AirBnb operates on online marketplace for primarily [homestays](https://en.wikipedia.org/wiki/Homestay) for [vacation rentals](https://en.wikipedia.org/wiki/Vacation_rental), and [tourism](https://en.wikipedia.org/wiki/Tourism) activities. It is the most popular rental stay due to its affordability and collaboration with the countries and city all over the world.

After collecting the relevant data of Airbnb hotel listings and performing different statistical analysis and tests, we can have an insight of the factors that impact the most and least on the overall satisfaction level of customers.

***Keywords: Data, Pandas, Matplotlib, Seaborn, Data Visualization, Exploratory data analysis etc.***

**What we have to find out**

* Predict whether a new AirBnb user will effectively book a destination or not.
* Predict which country a new AirBnb user's first booking destination will be.

**Introduction**

The Air platforms adjust their prices using a specific algorithm which is real time and dynamic known as **“Surge Pricing”** or **“Dynamic Pricing”**. It gives output a multiplier which is adjusted along with the base fare, locality and reviews.

Our goal here is to build a predictive model, which could help customers in finding the best option among many.

**Stepwise process :**

1. **Data importing/mounting**

airbnb\_df = pd.read\_csv(“/content/drive/MyDrive/Airbnb EDA Capstone Project 1/Airbnb NYC 2019.csv")

or [Airbnb NYC 2019.csv](file://C:/Users/Zindaram/Downloads/Airbnb%20NYC%202019.csv)

**2.PROBLEM STATEMENT:**

Maximize: The availability of Rooms to the customer.

Minimize: Minimize the bad reviews.

**The main goal of the project is to:**

To create a world where anyone can belong anywhere and we are focused on creating an end-to-end travel platform that will handle every part of your trip. As we work to achieve this goal, we are focused on building for the future, driving strong sustained growth, and creating new businesses that will power long-term success.

**3.DATA DESCRIPTION:**

The data description phase starts with an initial data collection and proceeds with activities in order to get familiar with the data. Identifying data quality problems, discovering first insights into the data and detecting interesting subsets to form hypotheses from hidden information are activities of this step. Data which is collected from a AIRBNB company from Seoul to get analyzed, involves usage details of customers from. The data was taken from Airbnb. It has 49000 rows and 16 columns.

**4.DATASET PREPARATION:**

The Airbnb datasets from room provider company contains 16 features and 49000 observations**.**

**Data-set description**

|  |  |
| --- | --- |
| Column | Dtype |
| id | int64 |
| name | object |
| host\_id | int64 |
| host\_name | object |
| neighbourhood\_group | object |
| neighborhood | object |
| latitude | float6 |
| longitude | float6 |
| room\_type | object |
| price | int64 |
| minimum\_nights | int64 |
| number\_of\_reviews | int64 |
| last\_review | object |
| reviews\_per\_month | float64 |
| calculated\_host\_listings\_count | int64 |
| Availability\_365 | int64 |

**5.FEATURE DESCRIPTIONS:**

**1.Id** – This is the identity number of the property listed by the particular host.

**2.Name** – It stands for the name of the property listed by the host.

**3.Host\_id** – It is the identity number of the host who has registered on Airbnb website.

**4.Host\_name** – These are the names of the hosts who have listed their properties.

**5.Neighbourhood\_group** – These are the names of the neighborhood groups present in NYC.

**6.Neighbourhood** – These are the names of the neighborhoods present in the neighborhood groups in NYC.

**7.Latitude** – These represent the coordinates of latitude of the property listed.

**8.Longitude** - These represent the coordinates of longitude of the property listed.

**9.Room\_type** – This represents the various types of rooms listed by the host.

**10.Price** – This is the rent of the Property listed in USD.

**11.Minimum\_night**s – This represents the minimum number of nights a customer rented the property.

**12.Number\_of \_reviews** – This represents the number of customers who reviewed the property.

**13.Last\_review** – This represents the date when the property was last reviewed.

**14.Review\_per\_month** – It is the count of reviews per month which the property received.

**15.Calculated\_host\_listing\_count** – it is the number of listings done by a particular host.

**16.Availability\_365** – This represents the number of days the property is available.

**● Understanding the data set** Next, we worked on checking the data set.

* How big is the data set?
* How many rows and columns are available?
* What could be the important columns to solve the Problem statement?
* How many null values do we have in the data set?

**● Null value Treatment**

* Dropping the unnecessary columns
* Filling the missing values
* Viewing the correlation of the numerical values

Graphical user interface, chart, treemap chart

Description automatically generated

**7. EXPLORATORY DATA ANALYSIS**

While doing the EDA we used the following analysis to solve the problem statement.

1.What can we learn about different hosts and areas?

2.What can we learn from predictions?

3.Which hosts are the busiest and why?

4.Is there any noticeable difference of traffic among different areas and what could be the reason for it?

**8. OBSERVATION:**

After observing the analysis of data we find the answers to the important questions.

1. The people who prefer to stay in Entire home or Apartment they are going to stay a bit longer in that particular neighborhood only.

2. The people who prefer to stay in Private room won't stay longer as compared to Home or Apartment.

3. Most people prefer to pay less price.

4. If there are more Reviews for a particular Neighborhood group that means that place is a tourist place.

5. If people are not staying more then one night means they are travellers.

**9.CONCLUSION :**

That's it! We reached the end of our exercise.

Starting with loading the data so far, we have done EDA, null values treatment, encoding of categorical columns, and found out major reasons that govern booking and steps on how we can increase it.

**10.REFERENCES**

* Guidance of Alma Better Instructors & Moderators
* Stack Overflow
* Discussion with my team members