AIRBNB PRICE PREDICTION FOR PROPERTY OWNER

RATIONAL STATEMENT

Airbnb is a commercial center, the sum a host can charge on a daily premise is firmly connected to the elements of the marketplace. Valuing an investment property on Airbnb is a difficult undertaking for the proprietor as it decides the quantity of clients for the property. One test that Airbnb has faced is deciding the ideal lease cost. Leaseholders are given a decent determination of postings and can channel by measures like value, number of rooms, room type and other conditions. In addition to this, the expenses to leaseholders and neighborhood probably do not surpass the advantages to travelers and landowners, by which sometimes hosts usually faces lot of problems in renting out their property and faces loss in the business. The hosts do not know about the actual or estimated price to rent out the property which in return results out the extreme loss or extreme benefits to them.

PROBLEM

Hosts wanted to rent out their property on the Airbnb on nightly basis, but being new in the field, they are unaware of the market value of the property. As a result, they would set the price of the property as higher or least. When the price set is high in comparison to market value, he would not get any customers for renting out his property. While, on the contrary, when he will set the least price of the property, more customers would be attracted to rent out his property, but in return he will face the problem in making profits from the property.

In order to resolve this problem, the identified proposed solution is designing a machine learning model to predict the price of the property. This tool will help the host to determine the good price of the property while renting out by which he/she can make profits by attracting more customers.

DATA REQUIREMENTS

All doing some EDA (Exploratory Data Analysis) on the original dataset the main data required for making the model are as follows:

- TYPE OF ROOM: While renting out the property, the type of room is also considered by the tenant. For example, the rent for the private room will be higher as compared to the rent of renting a shared room or renting a full house.

- PRICE OF ROOM: Depending upon the type of room available, the price of the room will also vary.

- LOCATION: The price of the property is also dependent upon the location where it is located. For example: the price of the property in the countryside area with Lake view will charge different amount of money as compared to the amount charged for the property which is in the city or main Downtown area.

- CANCELLATION FEE: Sometimes some of the Airbnb's booking charges some amount of cancellation fee from the customer. This factor is also having impact on the booking.

- TYPE OF PROPERTY: The price varies depending on the type of property. For example: if the property is an Apartment it is going to charge differently as compared to house.

- ACCOMMODATION: Price varies on varying the number the number of accommodation available for a property.

- CLEANING FEE: Some of the Airbnb's booking charges cleaning fee from the customer. This factor is also having impact on the booking.

- CITY: Since the price varies of Airbnb from one city to another. So this factor also has an impact on booking.

DATA SOURCE

Data sources to solve this problem will be obtained from the Kaggle website which was provided for a competition held and the data contain spreadsheet of data.

[Link to data](#)

The data source will consist list of 22 features and after EDA we are left with 14 features.

MISSING VALUES

During the loading of raw data, we could find the list of missing values in multiple features as seen in image below:

```
id                       0
log_price                0
property_type            0
room_type                0
amenities                0
accommodates             0
bathrooms              200
bed_type                 0
cancellation_policy      0
cleaning_fee             0
city                     0
description              0
host_has_profile_pic   188
host_identity_verified 188
host_since             188
instant_bookable         0
latitude                 0
longitude                0
name                     0
neighbourhood         6872
bedrooms                91
beds                   131
```

Features, like bathrooms, host has profile pic, host identity verified, bedrooms and beds, having missing values were filled by using mode values. Neighbourhood feature was dropped as it contains approximately 10% of missing values. Host since feature was also dropped since the price does not depend on this feature.

Hence using the above-mentioned procedure, missing fields were written.

EDA (EXPLORATORY DATA ANALYSIS)

The following were the observation made from the EDA done on the data:

- Using the description from the data we concluded:

    1. Around 40% of the records are from New York city.
    2. Around 70% of the records property type is apartment.
    3. Around 95% of the records contains bed type as real bed.
    4. 99% of host has profile picture.
    5. 60% host identity is verified.

- Using sweetviz library we came with following outputs:

    1. 60% of property available on Airbnb in US is apartment.
    2. 56% of room type on Airbnb in US is entire home/apt.
    3. 40% of properties can make accommodation for 2 peoples.
    4. 80% of properties has 1 washroom.
    5. Almost all the property is furnished and has at least a bed.
    6. 44% of host charges for concellation.
    7. 73% of host charges cleaning fees.
    8. 70% of records are from New York and Los Angeles whereas New York alone contain 40% of all roecords.
    9. 26% of host provide instant booking serivce.
    10. 67% properties contain one bedroom.
    11. 60% properties have just one bed in the property.

- Plotting different features vs Price, we came with following conclusion from the data:

    1. Most of the expensive properties on Airbnb within US is in San Francisco.
    2. Different type of property has different Price. The highest price is of Timeshare property.
    3. Price varies linearly with increase in number of accommodates.
    4. Price varies for different room type and Entire home has the maximum price.
    5. Since there is some dip in graph between no bedroom in house till 1 bedroom and after that there is increasing price as the bedroom count is increasing.
    6. Price is not affected by whether the host identity is verified or not.
    7. Host which doesn't have picture has high price. It could be because of just 226 cases of not having profile picture among 74000 cases.
    8. Price is not affected by whether the instant booking feature is available or not.

- Using heat map of data correlation, we concluded:

    1. It can be seen that accommodates is highly correlated with beds.
    2. It can also be seen that beds have high correlation with bedrooms.
    3. It can also be seen that beds have high correlation with accommodates.

KEY FEATURES

From the list of 22 features after filling missing values, doing EDA and then using correlation heat map we are left with 14 features on which the price of the Airbnb depends.

List of features on which the price depends are:

- log_price
- property_type
- room_type
- amenities
- accommodates
- bathrooms
- bed_type
- cancellation_policy
- cleaning_fee
- city
- description
- latitude
- longitude

TEST PROCESS

I will part the data into three classifications i.e. training dataset, validation dataset and test dataset. The model will be trained using training dataset and then evaluate using validation data. Based on accuracy, I will retrain the model until I get an acceptable accuracy of 70%.

SUBMITTED BY:

SIMRANJEET SINGH DHILLON

100799143