AirBnB Seattle

Alessio
Jun 22 · 2 min read

Introduction

Airbnb, Inc. is an American online marketplace company based in San Francisco, California, United States. Airbnb offers arrangement for lodging, primarily homestays, or tourism experiences. The company does not own any of the real estate listings, nor does it host events; it acts as a broker, receiving commissions from each booking. The task of this project is to pose question and find finding from the open dataset.

Question 1: What is the proportion of room_type per property_type?

As we can see below from the plot, there are 3 categories of room_type:

- Entire home/apt
- Private room

- Shared room

On the other hand, there are 16 different property_type, however for this exercise I have selected the top 5:

- House
- Apartment
- Townhouse
- Condominium
- Loft

Following observation from the plot:

- **House** (property_type) tend to have similar percentages of **Entire home/apt** (room type) and **Private room**(room type)
- **Apartment** (property_type) has greater percentage of **Entire home/apt** (room type) compare to **Private room**(room type)
- **Townhouse** (property_type) has greater percentage of **Private room** (room type) compare to **Entire home/apt**(room type)
- **Condominium** (property_type) has greater percentage of **Entire home/apt** (room type) compare to **Private room**(room type)

Regarding **Shared Room**, there is a low percentage so it is not useful to discuss this

Question 2: Who is the host in the dataset which posses the highest number of properties?

I have started by counting the instances of host(number of properties per host). I have then sorted the the values in descending order and got the first and highest host possessing greater number of properties (8534462)

Filtered data on host_id = 8534462 we can see that the name of the host is is Daniela and owns 46 properties.

Question 3: Which property on the selected database has the worst rating?

I have started to answer this question by plotting BoxPlot showing distribution of values.

We can see from the above that there is a property with rating equal to 20. I have then sorted ascending by rating column and identifed the worst rating and id of the property

Filtering property id on original dataframe we can see that the property_id on dataset with the worst rating is 9183838 and her name is Sam.

Conclusion

This is a summary of the finings from above article:

Question 1: What is the proportion of room_type per property_type?

- **House** (property_type) tend to have similar percentages of **Entire home/apt** (room type) and **Private room**(room type)
- **Apartment** (property_type) has greater percentage of **Entire home/apt** (room type) compare to **Private room**(room type)
- **Townhouse** (property_type) has greater percentage of **Private room** (room type) compare to **Entire home/apt**(room type)
- **Condominium** (property_type) has greater percentage of **Entire home/apt** (room type) compare to **Private room**(room type)

Question 2: Who is the host in the dataset which posses the highest number of properties?

The lucky host is Daniela(id=8534462) with 46 properties. Good for her!!

Question 3: Which property on the selected database has the worst rating?

Sam (id = 9183838) has the worst rating in the dataset. Better work harder :)