Title: Key Features or Amenities Used to Market AirBnB Listings in London

Author Name:

Publication Date:

Section 1: Project Definition

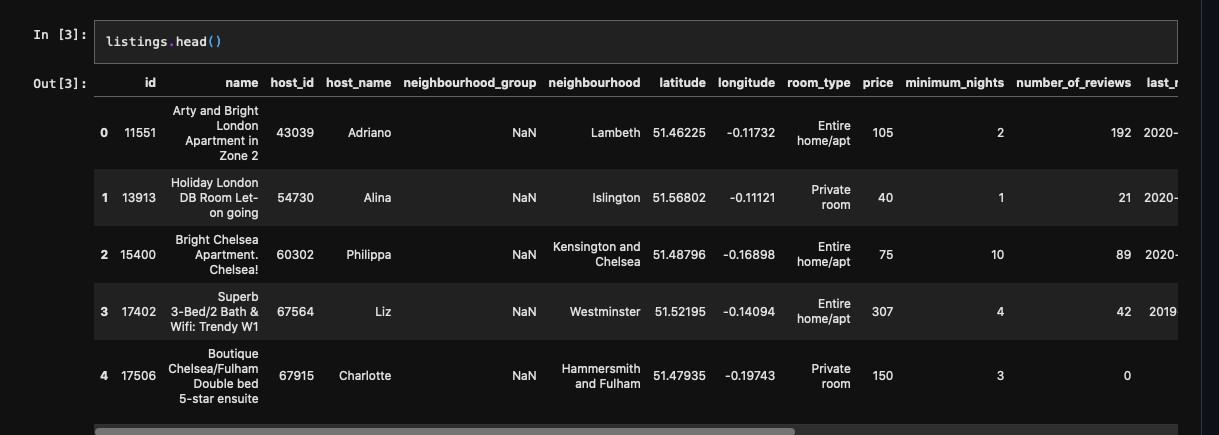
This project aims to identify the key features or amenities in London that are used to market the AirBnB. The analytical research aims to answer some of the following questions:

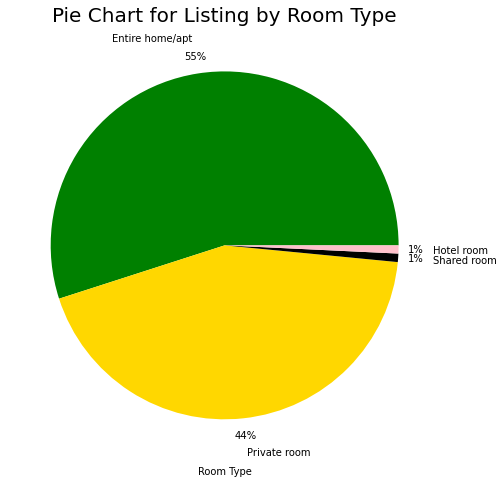
1. What are the most popular listings in London?
2. What are the most influential features or amenities in the popular listings
3. What are the key terms used to identify the most popular listings?

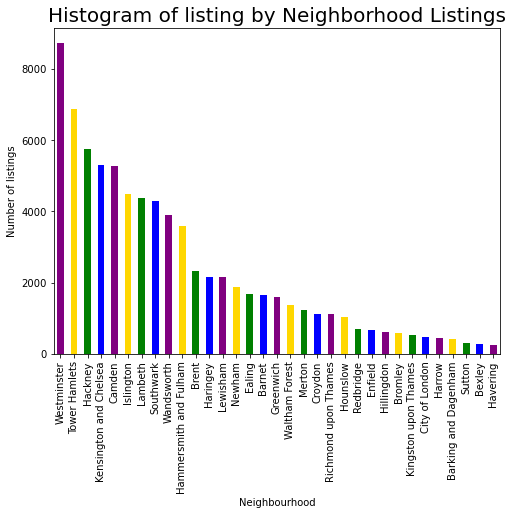
Section 2: Analysis

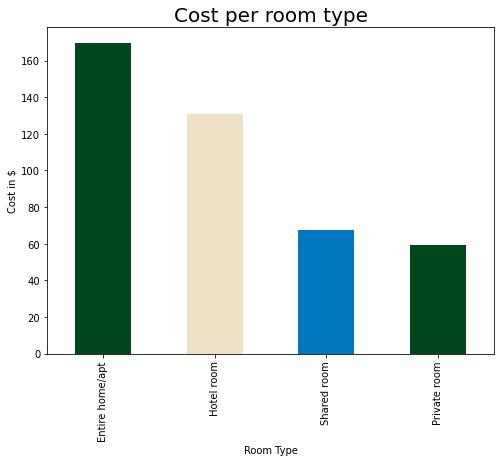
We have three sets of dataset, all related; the listings dataset that has information about the amenities, such as the location, owner and neighborhood group, the reviews dataset has information on reviews for each amenity enlisted and lastly we have the calendar dataset that has information on the booking dates and the charges on the respective dates.

The dataset in question was already cleaned and split into three, the listings, reviews and calendar. See example visualization of listings dataset below.



* 





Section 3: Methodology

The following methodologies and tools were employed:

Python language, the latest, that is version 3 was used. The requisite data processing modules were imported. They include: numpy, pandas, seaborn and matplotlib. Below is a brief discussion of each:

**NumPy** is a Python library used for working with arrays. NumPy comes in handy when we have to convert the dataset in question into an array. Working with arrays makes it so easy to manipulate data.

**Pandas** is built on top of NumPy. Comes in handy when working with arrays. It has tons of methods used to manipulate arrays.

**Seaborn** and **Matplotlib** are used for visualization. Matplotlib is enough but seaborn has some additional, more appealing visualization features that supplement Matplotlib.

Steps Followed:

First the modules were imported as evident on the jupyter notebook.

Second, the three related datasets were loaded.

Third, in order to understand the nature of data, the first five rows of each dataset was visualized in tabular form.

Fourth, visualization was then achieved using matplotlib plots.

Fifth, analysis of the visualizations was done to get deep insight as to what the data represented.

Section 4: Results

In this exercise, there was no need of employing a machine learning algorithm to aid in data analysis. Answers to the research questions did not need employment of AI/ML algorithms, instead, visualizing the plots turned out to conclusively answer the laid out questions.

The following was gathered from the visualization analysis:

1. The most first three populars listings starting from the topmost are Westminster, Tower Hamlets and Hackney.
2. The least last three popular listings, starting with the very last, include Havering, Bexley and Sutton.
3. The most influential amenities are the entire home/apartment with 55 percent rating, followed by the private rooms with 44 percent rating and the least is a tie between hotel and shared rooms.
4. Additionally the cost of the entire home or apartment is the most expensive and the least expensive is the private room.

Section 5: Conclusion

The analytics process was successful since all the questions were satisfactorily answered. However, much more could be done, and that is apartment price comparison across the year, across the neighborhoods. That for sure would make the whole analytics exercise more enriching in terms of the findings.