Estimation of AirBnB rental investments as a profitable business

*Note: Sub-titles are not captured in Xplore and should not be used

1st Natasha Cheri Satish
Department Of CSE
PES University
Bengaluru, India
pes1ug20cs261@pesu.pes.edu

2st Nitish Rathore
Department Of CSE
PES University
Bengaluru, India
pes1ug20cs270@pesu.pes.edu

3st Vishal J Lodha
Department Of CSE
PES University
Bengaluru, India
pes1ug20cs507@pesu.pes.edu

***Abstract*—Just twenty years ago, the sharing economy began to take off. Back then, it was difficult to imagine its significance for many businesses, but today, it is deeply ingrained in our daily lives. This tendency is best exemplified by AirBnB, which introduced a whole new business model to the hotel sector. With a unique pricing model, the AirBnB platform has generated millions of microbusinesses. Professional hosts usually invest in properties at various locations. In this paper we use K-nearest neighbours, a supervised machine learning algorithm to calculate the annual income of a particular location using data of AirBnB in London, we find the best neighborhood to buy a new property to host a AirBnB.**

***Index Terms*—KNN, AirBnB,exploratory data analysis, machine learning**

I. Introduction

The peer-to-peer activity of obtaining, giving, or sharing access to goods and services has seen tremendous rates of growth over the past decade and has now become a crucial component of our daily lives. This sharing, or gig-economy, is defined here as a set of platforms that facilitates this activity. The company AirBnB is at the fore of this expanding economy, looking to change how we use our most valuable asset: housing. AirBnB has become the leading provider of short-term rentals (STRs). The platform, which is already well-known, aims to link short-term tenants with different kinds of landlords in an effort to give travellers a more genuine travel experience while producing financial gains for suppliers. It is probably more profitable for property owners to rent on a daily basis to visitors than on a monthly basis to residents. Studies reveal that STRs are primarily concentrated in tourist destinations and city centres, and that, in these regions, a significant number of AirBnB listings are entire apartments permanently rented to visitors rather than occasional home-sharing. AirBnB is an example of buy-to-let gentrification in which investors purchase properties and obtain returns through the STR market, displacing residents with tourists. Buy-to-let refers to the purchase of property specifically to let it out, that is to rent it out.

Identify applicable funding agency here. If none, delete this.

II. Literature Review

The authors of [1] illustrate how locational strategies (clustering vs. diversification) directly affect the perceived quality and revenue performance of each listing in a professional host's portfolio and how this effect evolves as the portfolio of the professional host expands. The author is focusing on longitudinal AirBnB listings data for New York City from October 2014 to August 2017. Each professional host's number of managed AirBnB listings, the portfolio's geographic breakdown, and host characteristics that may affect a listing's perceived quality and financial performance are all recorded.

Four hypotheses have been put forth by the author. In order to simultaneously quantify the direct impact of locational strategies and the moderating effect of a multi-listing portfolio size with locational strategies on the perceived quality, hypotheses 1 and 3 have been proposed. To investigate the equivalent effects of locational strategies and portfolio size on revenue performance, hypotheses 2 and 4 are proposed. The author's chosen analytical unit is the listing-month. The author discovered that, for each listing in a professional host's portfolio in a given month, its average monthly revenue (AveRevenue), average monthly perceived quality (AveRating), and their interaction (Cluster zip NumMulti), are, respectively, a function of the portfolio size (NumMulti, a continuous variable), locational strategy (Cluster zip, a dummy variable), and their interaction.

According to the author's research, locational clustering of a multi-listing portfolio enhances the perceived quality but has a negative impact on revenue performance in New York City using data from Airbnb. When professional hosts expand their portfolios, these effects are more pronounced for perceived quality and less pronounced for revenue performance.

In [4] the authors are showing how Airbnb which is an online platform through which individuals rent spaces to tourists, has achieved success and become one of the most profitable innovations in the tourism and hospitality industry, has come under examination because of its potential negative impacts, particularly related to social challenges in local communities and impacts on the traditional accommodation sector. An analysis of Airbnb properties listed in Lisbon and other European cities indicated that only 20 per cent of the most booked accommodations appeared to be residents' homes Another analysis identified that tourists in Lisbon are one of the causes for the increase in house prices and rents, which forces homeowners out of the market.

III. Dataset

Dataset has been taken from the inside.airBnB website for UK. The dataset has 16 attributes with 86,358 rows.

```
id                                 0
name                              24
host_id                            0
host_name                         10
neighbourhood_group            86358
neighbourhood                      0
latitude                           0
longitude                          0
room_type                          0
price                              0
minimum_nights                     0
number_of_reviews                  0
last_review                    21516
reviews_per_month              21516
calculated_host_listings_count     0
availability_365                   0
dtype: int64
```

IV. Working of algorithm

We have performed Exploratory Data Analysis to draw inferences from the dataset. We have dropped the columns with null values and used matplotlib and seaborn libraries from Python to plot scatterplots and barplots to make inferences from the dataset. In the EDA we have plotted multiple graphs to understand the data more precisely. The following are the graphs that we have plotted:

Type of room v/s Number of rooms (Bar graph)

Neighbourhood v/s Number of rooms (Bar graph)

Neighbourhood v/s Average price (Bar graph)

Neighbourhood v/s Availability 365 (Bar graph)

Neighbourhood v/s No of days occupied (Bar graph)

Airbnb Properties by Price/Night (Scatter Plot)

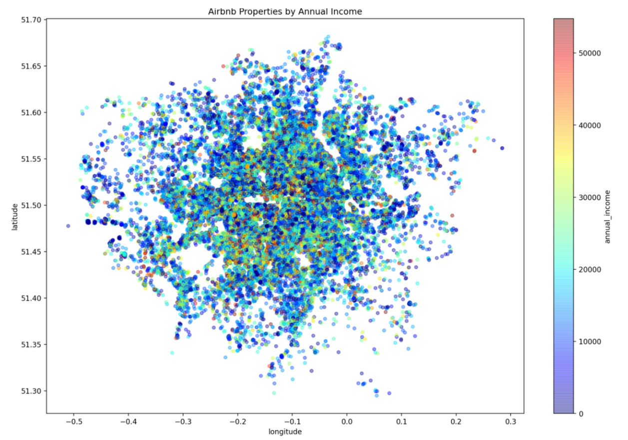
Airbnb Properties availability throughout the year (Scatter Plot)

Airbnb Properties by Annual Income (Scatter Plot)

Airbnb Properties by Total Income per Neighborhood (Scatter Plot)

Annual Income per property of a specific neighborhood (Scatter Plot)

Time to get back investment(in years) (Scatter Plot)



We are calculating the average income of a particular neighborhood and then modeling it to respective latitude and longitude of that area. If the user wishes to calculate the income for a particular latitude/longitude we are implementing weighted KNN algorithm taking the nearest 3 neighbors into consideration.

Let L = (xi , yi) , i = 1, . . . ,n be a training set of observations xi with given class yi and let x be a new observation(query point), whose class label y has to be predicted. Compute d(xi, x) for i = 1, . . . ,n , the distance between the query point and every other point in the training set. Select D' D, the set of k nearest training data points to the query points Predict the class of the query point, using distance-weighted voting. The v represents the class labels. Use the following formula

Distance-Weighted kNN

Might want weight nearer neighbors more heavily...

$$\hat{f}(x_q) \leftarrow \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i}$$

where

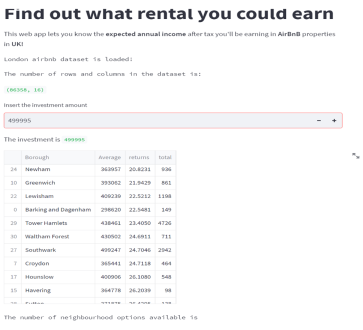
$$w_i \equiv \frac{1}{d(x_q, x_i)^2}$$

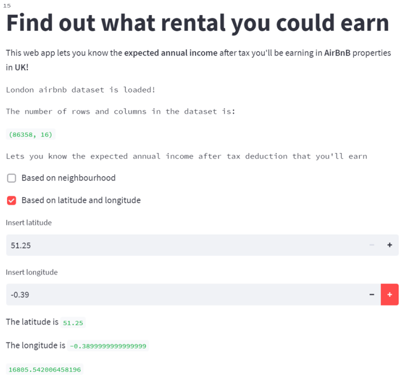
and $d(x_q, x_i)$ is distance between x_q and x_i

Note now it makes sense to use *all* training examples instead of just k

V. Results and Conclusion

From the London AirBnB dataset, we perform exploratory data analysis to estimate the occupancy rate and calculate average income of a neighbourhood based on occupancy rate in a year and the rent per night for different type of rooms. Following which, we have implemented a KNN algorithm to calculate the expected income of a particular property in that neighbourhood. We are using Mean Absolute Error as the performance metric to calculate the effectiveness of the model. We have chosen this metric, as this metric gives us a confidence bound with lower and upper limits fixed, this helps our client to be assured of the worst and the best case scenarios before making an investment related decision.





VI. Future Work

In future we wish to implement a way to calculate the profits for the investors who take the property on lease. Also we plan to expand the business to other cities in other countries as well. We would like to take into consideration the features and facilities provided by the property.

References

[1] Xie, Karen and Mao, Zhenxing. (2019). Locational Strategy of Professional Hosts: Effect on Perceived Quality and Revenue Performance of Airbnb Listings. Journal of Hospitality and Tourism Research. 43. 10.1177/1096348019836277.

[2] A. Zhu, R. Li and Z. Xie, "Machine Learning Prediction of New York Airbnb Prices," 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), 2020, pp. 1-5, doi: 10.1109/AI4I49448.2020.00007

[3] A. Garlapati, K. Garlapati, N. Malisetty, D. R. Krishna and G. Narayana, "Price Listing Predictions and Forthcoming Analysis of Airbnb," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9579773

[4] Petruzzi, Marina and Marques, Gabriela and Carmo, Manuel and Correia, Antónia. (2020). Airbnb and neighbourhoods: an exploratory study. International Journal of Tourism Cities. ahead-of-print. 10.1108/IJTC-08-2019-0119