A Dive into Airbnbs in Edinburgh

Anant K. Gupta and Aryan Sharma

April 21, 2023

1 Overview

This report focuses on analysing Airbnb listings in Edinburgh. Our analysis reveals a significant increase in Airbnb prices during the three weeks of August. The Edinburgh Festival Fringe, which attracts many visitors, is a potential factor contributing to this price hike. During the festival, the availability of Airbnb listings decreases, and owners can charge more for their properties because guests may be willing to pay a higher price to secure a place to stay. This phenomenon can be attributed to the limited options available to potential guests. We also found a relation between an Airbnb's location and its price, with a premium being charged for being in the city center, or interestingly in the outskirts, with a dip in pricing visible between the two regions. Finally, we found that we could predict the pricing of an Airbnb, with an R^2 63.62% by just using the top 10 factors, obtained by our Random Forest regression model. Overall, we gain some key insights into what makes an Airbnb fetch a premium, and discuss some possible reasons behind these price driving factors.

2 Introduction

Context and motivation The study is interesting to investigate because Airbnb has disrupted the traditional hotel industry and has become a popular option for travellers around the world. The study can provide insights into the features that make a property more desirable to potential guests, which can be useful for Airbnb hosts and property owners in Edinburgh. The study can also shed light on how renting prices vary throughout the year in Edinburgh, and thus when it would be most beneficial to advertise properties or put them up for rent. Overall, this study has the potential to provide valuable insights into the intersection of the sharing economy, tourism, and housing markets.

Previous work "Effects of location on Airbnb apartment pricing in Málaga - ScienceDirect"- by Jorge Chica-Olmo, Juan Gabriel González-Morales, José Luis Zafra-Gómez Chica-Olmo et al., 2020.The study investigated how the location of Airbnb apartments in Málaga affects their prices and found that it has a significant impact. Factors such as accessibility to city centre, beach, places of interest, and walkability were found to positively affect the price, with walkability being a new approach. Noise where the apartment is located had a negative impact on the price, while certain ethnic groups in the area had an inverse relationship with the rental price, likely due to lower-income families residing there. This study offers a distinctive view of the spatial distribution of tourist apartment prices.

"The role of specific attributes in determining prices of Airbnb listings in rural and urban locations" - by Martin Falk, Blaise Larpin, Miriam Scaglione Falk et al., 2019.This research analysed the factors that influence Airbnb accommodation prices in rural and urban areas of Switzerland between 2016-2018. The study found that the impact of various attributes on Airbnb prices varied across locations and between low and high-priced listings. Sauna, Jacuzzi & Spa, spaciousness, and uniqueness had a greater impact on rural Airbnb prices than urban ones, while penthouse, chic & designed, and suite had a more significant effect on urban Airbnb prices. The relevance of some attributes was higher for high-priced accommodations, particularly for luxury and suite. Additionally, certain subjective characteristics such as luxury, guest, and chic & designed also increased the willingness to pay in the low-price segment.

Objectives We wish to get some insights into the Edinburgh Airbnb market, and relations of the price of these short-term rentals to other factors. In particular, we are trying to answer the following:

- How does the average price per night of Airbnb in Edinburgh vary over the year? Especially when is there a great demand for rentals?
- Are particular areas or neighbourhoods more expensive than others?
- How well can features of a property listing be used to predict short-term rental prices?

3 Data

Data provenance The data sets for this project have been generated by Airbnb itself and made available at the insideairbnb website "Inside Airbnb: Edinburgh", 2021 and were obtained as csv formatted files. It is available under the Creative Commons Attribution 4.0 International License Commons, 2014, which allows copying and redistribution of the material in any medium or format and also permits one to remix, transform, and build upon the material for any purpose, as long as appropriate credit is given, a link to the license provided, and any changes made indicated.

Data description

- Calendar

  The calendar csv file has 7 columns including listing_id, date, available (True/False), original price, adjusted price (if applicable), minimum number of nights required for booking, and maximum number of nights allowed for booking

- Reviews

  The reviews csv file consists of 4 columns that provide information about a reviewed listing. It includes details such as the listing ID and information about the review, such as the date it was posted, the review text, and the review ID. Additionally, the file also includes information about the reviewer, such as their name and ID.

- Listings

  The reviews csv file consists of 4 columns that provide information about a reviewed listing. It includes details such as the listing ID and information about the review, such as the date it was posted, the review text, and the review ID. Additionally, the file also includes information about the reviewer, such as their name and ID.

- Neighbourhood

  The neighbourhood geojson file includes columns that provide information about different neighbourhoods in a specific region, including neighbourhood name or ID, neighbourhood group, and geometry.

Data processing Significant data cleaning had to be performed. Firstly, there were some columns which had little to no data. These were licence (15 values), bathrooms (0 values), calendar_updated(0 values), and neighbourhood_group_cleansed(0 values). All four except bathrooms were dropped, and bathroom was filled with floats by extracting data from bathroom_text using regular expressions. Next, three categories of columns were removed - ones which were not relevant to questions being asked, ones that could introduce bias in our results, and ones which had extremely high-correlation with another column. Also, some columns which were un-related to our questions or could potentially cause bias were removed. These included, 'host_name', 'listing_url', 'scrape_id', 'last_scraped', 'name', 'description', 'source' etc. Also, we only retained the availability for 90 days (availability_90) column, out of the various

vailibility columns present as the Scottish Government plans to impose regulations that limit rentals to 90 days per year, similar to those in London. These regulations were proposed to address concerns raised by local residents and housing organizations about the rise of short-term rentals in Edinburgh.

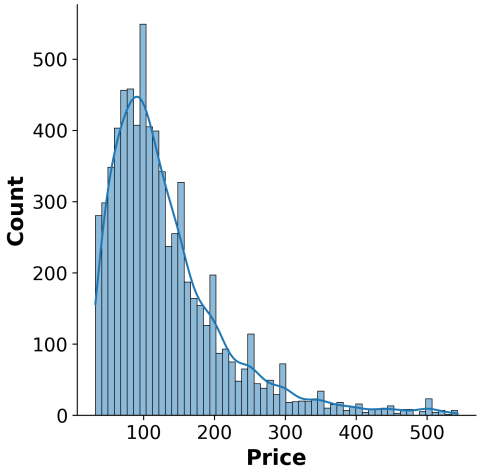

Figure 1: Distribution of Price before applying Log Transform

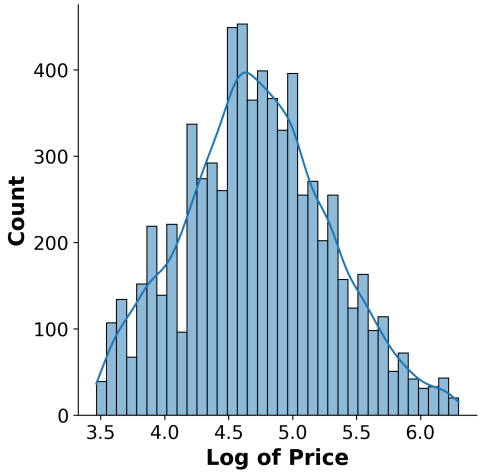

Figure 2: Distribution of Price after applying Log Transform

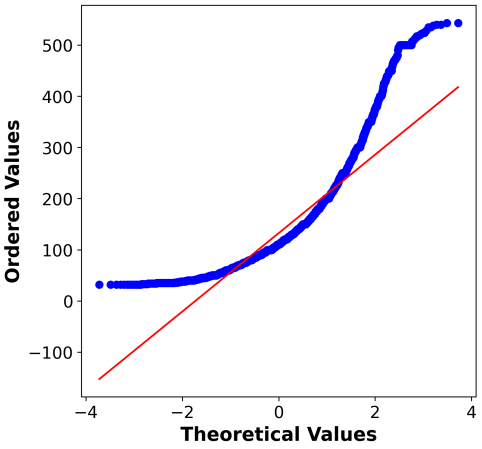

Figure 3: Probability plot of this distribution before transformation.

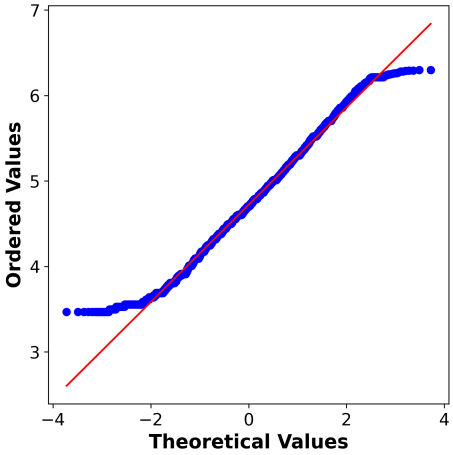

Figure 4: Probability plot of this distribution after transformation.

The initial price distribution for the Airbnb listings was heavily skewed to the right, with a skewness of 1.770251 and a kurtosis of 4.000498 as can be seen in Figure 1. This indicates that most of the prices were concentrated towards the lower end, with a few listings having extremely high prices. To normalize the data and make it suitable for model training, we applied a log transformation to create a new column, log_price as can be seen in 2. This transformation effectively compressed the range of high-priced listings and expanded the range of lower-priced listings, resulting in a more symmetrical and normal distribution of data (Feng et al., 2014) and the same is visualised in the probability plots Figure 3 and Figure 4. The log_price column had significantly lower skewness and kurtosis values (0.152469 and -0.294769, respectively) compared to the original price column. The normalization of the data through log transformation improves model performance by enabling accurate prediction of both high-priced and low-priced listings.

4 Exploration and analysis

The Airbnb dataset offered ample opportunities for exploration and analysis of various research questions. In regard to the variation in demand over the year, we saw considerable variation in the median price of rentals throughout the year, with notable spikes at the end of December/beginning of January and during August, as depicted in 5

As we know demand and price to have a direct relation, the spikes in the pricing indicate a rise in the demand for rentals (Fernando, 2023). The initial spike in December/January, especially around new years can be attributed to the holiday season when people tend to travel to Edinburgh for vacation. However, the spike in August initially appears anomalous. On conducting further investigation into the model, we discovered that this peak was caused due to the Edinburgh Fringe festival, which is known as the biggest art festival, recognized globally. (McLean, 2022).

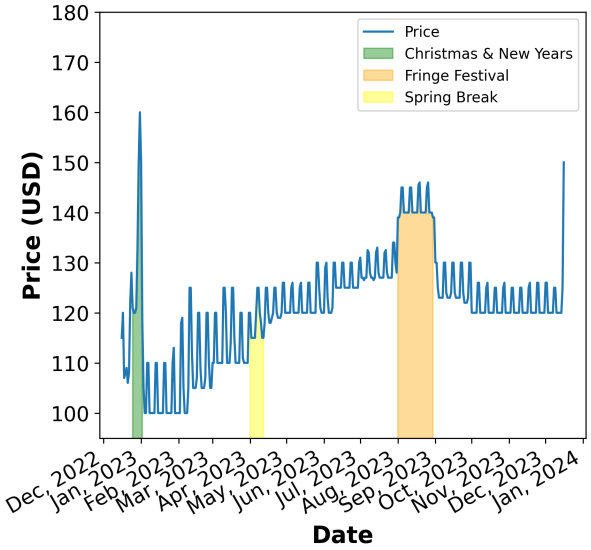

Figure 5: Variation in median price of listings in Edinburgh over the year.

Upon conducting a more in-depth inquiry, we discovered that the reason behind it is the Edinburgh Festival Fringe, which is recognized as the biggest festival for performing arts globally. The festival, which takes place over a period of three weeks in August, contributes to the cause. This festival provides a compelling explanation for the observed spike in median price during that period. The smaller recurrent spikes throughout the year occur on weekends and can thus be attributed to there being greater demand over the weekend. Our null hypothesis is that the fringe festival does not have an effect on the average price on Airbnb bookings as compared to average pricing over the year.

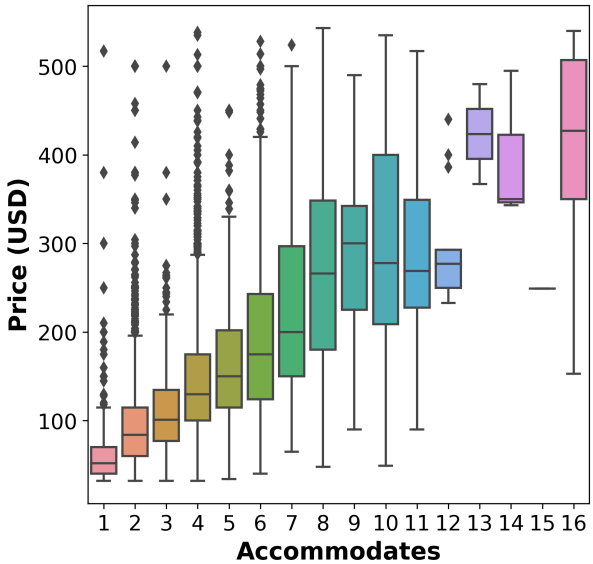

Figure 6: Box-plot price against accomodates

The alternate hypothesis would be that the fringe festival influenced the average pricing during the three weeks of august as compared to the average for the whole year. We conducted a p-test at the 5% significance level with X degrees of freedom for the average price during fringe compared to the average price for the whole year. We got a p-value of 2.55e-19.

This clearly rejects our null hypothesis and proves that fringe festival influences the average price of Airbnb listing in Edinburgh. We can also see in Figure 6 that the price of Airbnb's tends to move up as the number of people it can accommodate increases. This is expected, but we also get to see a lot of outliers in smaller listings, potentially due to them being more luxurious and premium than others.

We then moved on to see the distribution of listings by their location, and its effect on price. On plotting a map of Airbnb listings, Figure 7, we observe that most of the Airbnb listings are centred around Edinburgh's old town. The reason for the high concentration of Airbnb listings in Old Town, Edinburgh seems to be because of its historical and cultural importance, including famous landmarks

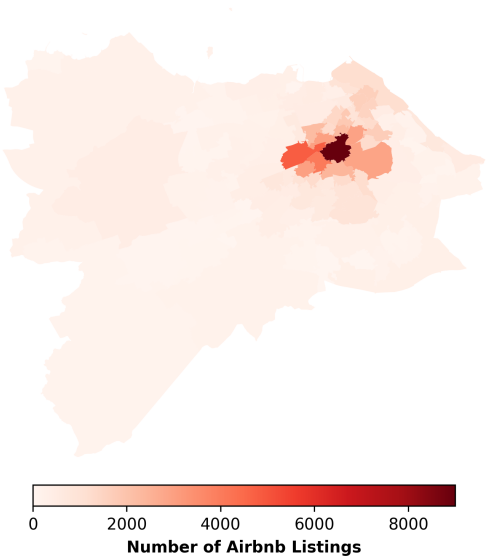

Figure 7: Number of listings in each Edinburgh Neighbourhood

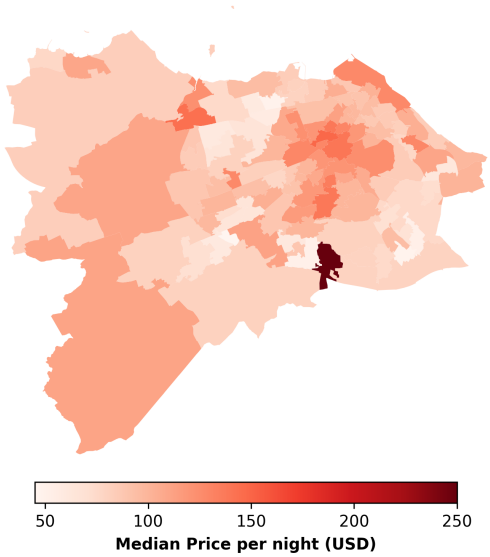

Figure 8: Median price of listings in each Edinburgh Neighbourhood

like Edinburgh Castle and the Scottish Parliament Building, attracting tourists interested in the city's rich history. Additionally, Old Town's lively atmosphere, an abundance of shops and entertainment options, and central location make it a popular choice for visitors who want to explore the city and its surroundings (Galloway, 2020).

There are a total of 7389, of which About 69% of the total listings are "entire"(i.e., you are renting the entire property of your own). Most of the remaining types(about 27.73%) are private rooms(i.e., you are renting a bedroom and possibly also a bathroom, but there might be other people in the property). A very less portion(0.02%) of the properties are shared property rooms(i.e., you are sharing a room with either the property owner or other guests), and the exact counts are present at Table 1 In the map plot, Figure 8, we analyse the median prices of Airbnb's grouped by neighbourhood. From the plot, we can observe that the median pricing of Airbnb's at the centre(around Old town) are relatively more expensive, which is what we expected as we know that it is a popular choice for visitors who want to explore the city and its surroundings, and since there is a high demand for the Airbnb in these areas, the owners are likely to increase the price of these Airbnb's. However, an interesting point can also be observed although the number of listings on the outskirts is less, their pricing is very high. A possible reason for this could be that the houses in these areas are very big and therefore have higher pricing. These prices can be justified by the fact that people usually like to take a break from their day-to-day hectic life and would like to take a break at a silent place, where they could enjoy the scenic beauty. Another interesting observation made from the graph is that area at the bottom southeast of the map(Fairmilehead) has very few listings, but their pricings are absurdly high. The pricing of Airbnb rentals in Fairmilehead, Edinburgh, may be high due to various factors such as high demand among tourists, limited supply, quality of the properties, major events or festivals, and local regulations or taxes. Understanding these factors can help explain why the prices in the area may be elevated. Limited supply could also be a factor, as few Airbnb rentals in the area can result in higher prices due to less competition.

Property Type	Count
Entire	5126
Private	2049
Room	169
Others	30
Shared	15

Table 1: Counts of different room types

The quality of the properties available on Airbnb in the area could also be a factor, as high-quality properties commands higher prices additionally, Fairmilehead and its surrounding areas have historical significance, which can also attract tourists. Places like Roslyn Chapel, Craigmiller Castle, and Gladstone's

Land are some of the historical attractions in the area. The rich history of the area, combined with its scenic beauty, can make it a popular destination for tourists, leading to higher demand for accommodation and potentially higher Airbnb rental prices.

One of the most interesting questions was identifying the factors that have the most significant impact on the price of a particular listing. For this we naturally felt the need to use a learning model, which we could train on the cleaned data, and then inspect to see what importance it gave to the each of the listing's features. We first used a random forest regression model, which gave us an R^2 of 67.78% and an RSME of 0.32. Random Forest Regression is an ensemble learning technique combining multiple decision trees to perform regression and classification tasks, preventing over-fitting. Its ability to handle high-dimensional data with a large number of features and data with complex nonlinear relationships between the features and target variable, was another reason we felt that it was a good learning technique for our model (Team, 2022). Figure 10 plots the actual prices against the predictions, along with a line of best-fit and the distribution of the prices.

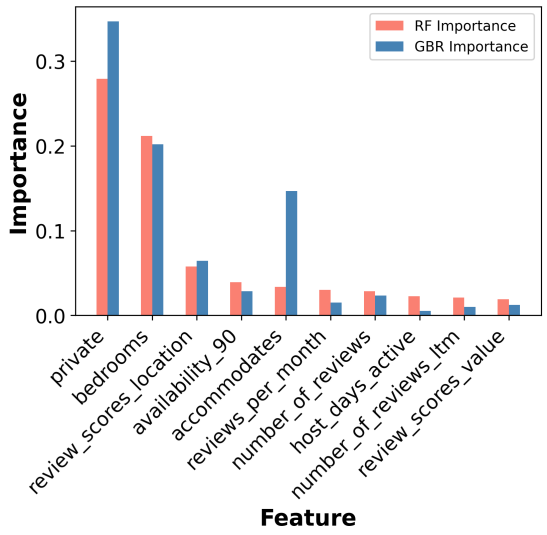

Figure 9: Comparison of importance given to the top 10 features by Random Forest Model and Gradient Boosting Regression Model.

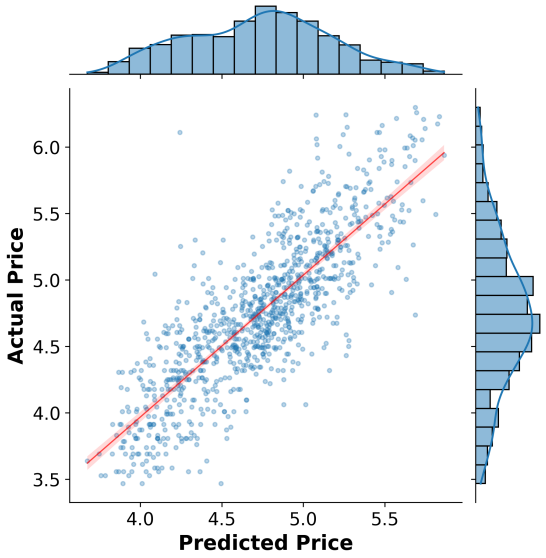

Figure 10: Actual price of the listings against price predicted by Random Forrest model, along with the distribution of price.

To ensure the robustness of our model, we compared it with two additional regression models: Gradient Boosting Regression and Extreme Gradient Boosting. Interestingly, both these models gave us a similar R^2 of 66.72% and 65.55%, indicating that they performed equally well in predicting the price of Airbnb listings in Edinburgh. To gain further insights into the key factors affecting Airbnb pricing in Edinburgh, we compared the feature importance of the 10 most important features as per Random Forest with those of Gradient Boosting Regression. As shown in Figure 9, both models gave similar importance to the factors, which validates our model. However, one anomaly we observed was the high importance given to the private

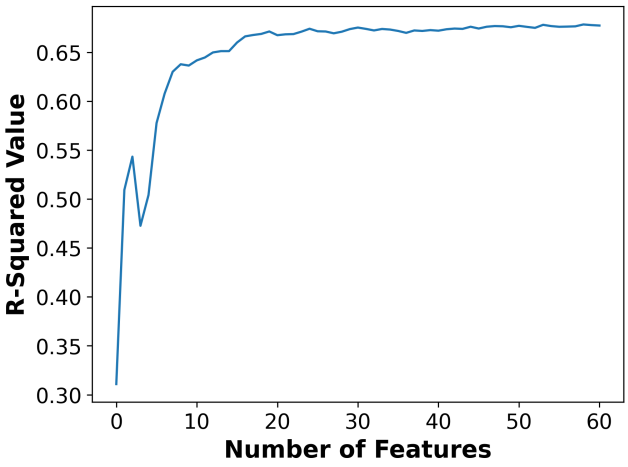

Figure 11: Variation in R-Square of Random Forest Model with increase in Features

property type by Extreme Gradient Boosting. Regression. On running the Random Forest Model, with varying number of features, sorted by their importance in descending order, we saw that in as little as 10 features we can get an R^2 of 63.62%, as can be seen in Figure 11.

5 Discussion and conclusions

Summary of findings In our study, we successfully addressed the three research questions we had formulated. Firstly, we were able to gain insights of the key features that affect an Airbnb's price the most. The top 5 were private, bedrooms, review_scores_location, availibilty_90 and accomodates. We also ran multiple models to ensure robustness. We also ran the chosen Random Forest model, with various number of features, based on their importance, and established that a mere 10 features can enable accurate prediction of the price of a listing with an R-squared value of 63.62%. This finding is significant as it highlights the feasibility of using a limited number of predictors in pricing models for Airbnb listings.

Secondly, we employed a t-test to evaluate the null hypothesis that The Fringe did not have an effect on Airbnb listing prices. Using the p-value from the t-test, we were able to reject the null hypothesis, indicating that Fringe festival does indeed have a statistically significant impact on pricing. This finding is relevant to both Airbnb hosts and customers as it underscores the importance of considering the timing of the booking when pricing a listing or looking for a good deal.

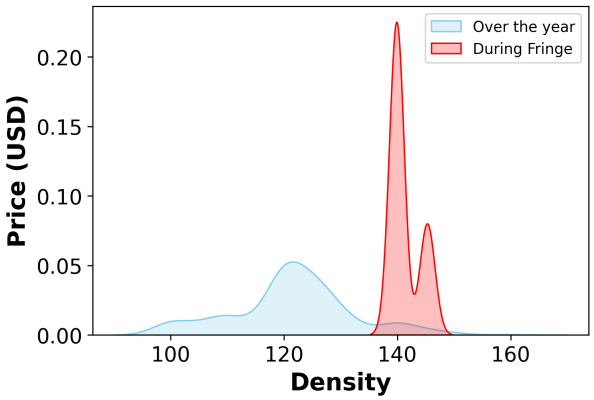

Figure 12: Comparison between distribution of prices during the year and during fringe

Lastly, we also observed that location has a pronounced influence on the pricing of Airbnb listings. Specifically, we found that central areas tend to command higher prices, likely due to their proximity to amenities and attractions. Interestingly, we also noted a rise in pricing in the outskirts of the city, possibly attributable to the growing trend of people seeking to escape the city and immerse themselves in natural surroundings. This finding offers valuable insights for Airbnb hosts looking to price their listings competitively based on location.

Evaluation of own work: strengths and limitations Our team put in a lot of effort and thoroughly analysed the information we're presenting in this report. However, the main issue with our work is that the data we have is only from December 16, 2023, which doesn't accurately represent Airbnb prices. For example, we noticed that prices tend to skyrocket during the New Year, but this could be due to multiple factors, such as high demand and last-minute bookings. Ideally, it would have been better if our dataset included prices from 3 months prior to the booking date. Due to this limitation, our study's conclusions may be affected. Our report also examines the factors that influence Airbnb pricing, and we think that our findings can help property owners improve their listings to increase their market value and generate higher returns. Additionally, our study also provides insights about the number of Airbnb listings at different places in Edinburgh, which could be good for providing suggestions to the owner's as to where setting up Airbnb's could give them a higher return.

Comparison with any other related work The results of our study were similar to the conclusions of a previous study Mwigeka, 2022, except for one key difference: while their study found that the entire home/apartment played the highest role in determining Airbnb prices, our study found that private rooms had the biggest impact. We believe this could be due to variations in traveller preferences across different locations. The size of the market, the type of traveller, and the availability of accommodations could also be factors that influence pricing dynamics. For instance, hosts in a larger market like New York City may

need to price their entire apartments more competitively to attract bookings, while a smaller market like Edinburgh may attract solo travellers or students seeking affordable private rooms.

Our study also supports the observation made in news article "Airbnb guests 'worth £48m' over Edinburgh Fringe", 2018 that Airbnb prices spike during the Fringe period in Edinburgh. This is likely due to the limited supply of available Airbnb properties during this time, with some hosts choosing not to rent out their properties and others renting to festival performers or organizers for higher profits. Properties located in the heart of the city near festival venues are in high demand and can command premium prices. Additionally, hosts may incur additional costs for cleaning and repairs caused by festival guests, further contributing to the increased prices. In conclusion, our study highlights the importance of considering local factors when analysing Airbnb pricing dynamics and shows that the Fringe period is a lucrative time for Airbnb hosts in Edinburgh (VanReenen, 2022).

Improvements and extensions Although the current project involves a variety of statistical methods and addresses several questions, there is potential for improvement and expansion. At present, we have not utilized the reviews dataset provided to us. There are several ways in which this dataset can be used in our analysis, such as utilizing natural language processing to perform sentiment analysis on the reviews and incorporating this data as a factor into our model. As individuals often consult reviews prior to booking, this approach could potentially enhance our model's accuracy. Another possibility is utilizing a map API to determine the distances between Airbnbs and popular destinations within the city. This would offer insights into the quality of the neighbourhoods in which the Airbnbs are situated, as well as the importance individuals attach to proximity to popular destinations. Lastly, the most crucial idea is to attempt to predict factors that impact the price of an Airbnb per person. While larger Airbnbs that can accommodate more people will naturally have a higher price, the price per person is an entirely different story. This aspect concentrates more on the quality of a specific Airbnb and may place greater emphasis on amenities and location.

References

Airbnb guests 'worth £48m' over edinburgh fringe. (2018). https://www.bbc.co.uk/news/uk-scotland-edinburgh-east-fife-44952284

Chica-Olmo, J., González-Morales, J. G., & Zafra-Gómez, J. L. (2020). Effects of location on airbnb apartment pricing in málaga. *Tourism Management*, *77*, 103981. https://doi.org/https://doi.org/10.1016/j.tourman.2019.103981

Commons, C. (2014). Creative commons license deed. https://creativecommons.org/licenses/by/4.0/

Falk, M., Larpin, B., & Scaglione, M. (2019). The role of specific attributes in determining prices of airbnb listings in rural and urban locations. *International Journal of Hospitality Management*, *83*, 132–140. https://doi.org/https://doi.org/10.1016/j.ijhm.2019.04.023

Feng, C., Hongyue, W., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, *26*, 105–9. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Fernando, J. (2023). Law of supply and demand in economics: How it works. https://www.investopedia.com/terms/l/law-of-supply-demand.asp

Galloway, C. (2020). Edinburgh old town found to have 'highest incidence of airbnbs' in uk. https://www.edinburghlive.co.uk/news/edinburgh-news/research-reveals-edinburghs-old-town-17786557

Inside airbnb: Edinburgh. (2021). http://insideairbnb.com/edinburgh

McLean, P. (2022). Edinburgh festival: Biggest arts festival in the world begins. https://www.bbc.co.uk/news/uk-scotland-edinburgh-east-fife-62427442

Mwigeka, S. (2022). What factors drives the airbnb listing's prices? *International Business Economics Studies*, *4*, p26. https://doi.org/10.22158/ibes.v4n1p26

Team, C. (2022). Random forest. https://corporatefinanceinstitute.com/resources/data-science/random-forest/

VanReenen, D. (2022). Airbnb prices more than double in august for edinburgh's fringe. https://www.edinburghlive.co.uk/news/edinburgh-news/edinburgh-fringe-2022-airbnb-prices-24344947.amp