‘Exploratory Analysis of the Social Impact of Airbnb in London Boroughs: Does Airbnb Have a Measurable Influence on Recorded Crime?’

Trupti Kolvekar, Honor McGrigor, Nasra Mohamed, Charlotte Pollard, Samantha Shanthakumar, Hannah Still

INTRODUCTION

Aims and objectives of the project

This study investigates the growing perception in London’s communities that a rise in sharing economy platforms such as Airbnb contributes to a worsening of social conditions, impacting housing, crime and anti-social behaviour (Ke, O’Brien *et al*., 2021). This study aims to use data analysis to measure the impact that the growth of Airbnb has on London’s communities.

The objectives of this study are to:

- Determine the prevalence of entire properties and private rooms listed on Airbnb in London boroughs
- Identify patterns in crime rates, case resolution and anti-social behaviour in the London boroughs over a period of 4 months (11 December 2022 - 14 March 2023)
- Evaluate whether there is a correlation between the top 5 boroughs with the highest and lowest crime rate versus prevalence of Airbnb rentals. Furthermore, whether the crime rate, case resolution and anti-social behaviour has an impact on the rental reviews.
- Assess the comparison of prices in Airbnb dataset and whether there was a trend with crime and anti-social behaviour in certain areas.

Roadmap of the report

This report contains the background to our study and question (*Background*), steps taken to implement it (*Steps Specification*), a detailed description of methodology and process (*Implementation* and Execution), findings (*Result Reporting*), and what we learnt from the project as a whole (*Conclusion*).

BACKGROUND

Airbnb is an online platform that facilitates property sharing for short or long-term letting as alternative accommodation to hotels and hostels. Since its launch in 2009, Airbnb has rapidly gained popularity worldwide due its affordability and ease of use (Hati *et al*., 2021). However, studies have suggested negative impacts on local communities as a result of housing market disruption and increased crime rates (Ke, O’Brien *et al*., 2021, Shabrina *et al*., 2021). The team have all had both connections to London, and experiences of housing pressures. Therefore, this investigation focused on analysing the social impact that Airbnb has across London.

Analysis was conducted using open-source data from the Metropolitan Police, Inside AirBnB, and housing data to determine correlations between Airbnb listings in London boroughs with crime reports in those areas. Visualisations were used to interrogate data and disseminate key findings. These results will help draw inferences on trends, though our conclusion remains circumspect.

The target audience for this study is a non-expert but interested audience comprising members of the public from neighbourhood and community groups, grass-roots organisations, to people looking for housing to buy or rent. A larger study may also eventually be of interest to policymakers. The study seeks to inform interested and invested readers with a basic to good level of understanding of the broad issues at hand, but little technical knowledge.

The results of this study, or of a potential larger study, contribute to a growing body of work investigating the social impacts of large sharing economy platforms. Are these platforms ultimately extractive? Or do they give more back to local communities than they remove?

STEPS SPECIFICATIONS

Framing Questions

Initially, group communication channels were set up. The team conducted a SWOT analysis to assist with assigning roles. A project development board assisted the process of filtering ideas through to decision. Questions and datasets were considered to be of equal importance (i.e. an excellent data set without a thought-provoking or applicable question would be hard to work with, as would a good angle of investigation and an unreliable or too small dataset). The group considered a very diverse range of ideas from open source intelligence, the tech sector, crypto currency, aerospace, healthcare, e-commerce, housing and crime. A further look at housing and crime, while discussing ideas from happiness levels to weather, led to a focus on temporary housing: Airbnb. Once decided, exploratory data testing and analysis was conducted (as can be seen in the 'Initial Stages - Exploratory Analysis' file, in 'Archived_files' on GitHub).

Data Gathering and Sources

Initial research took in Kaggle, government and activist sources, with the focus on open-source, well-provenanced datasets with sufficient depth and breadth, that could feasibly be joined and support detailed analysis. Following discussion, the decision was made to work with data from Inside Airbnb and the Metropolitan police. Inside Airbnb is a data activism project working to advocate for communities affected by Airbnb. The dataset presented us with plenty of information (66,152 rows x 74 columns) and much potential. The crime dataset, (441,293 rows x 12 columns), was sourced from data.police.uk, which holds open source data on crime and policing in the UK, excluding Scotland. The crime dataset contained anomalies such as inclusion of crimes occurring outside of London and results for crime in the City of London had significantly lower crime rates. It was later discovered that separate data files were released for the City of London. The 2022 average property price data was sourced from Rightmove, and summarised in this article. The study compares data and makes findings from a four month period from 11 December 2022 - 14 March 2023. This time period was chosen based on dataset availability and the exigency of a reasonable and manageable period from which to draw conclusions.

Pre-Processing

Following project development, a core team got to work on pre-processing the data. The team worked individually, in pairs, and in tandem. Tasks were assigned based on strengths, weaknesses and priorities for learning. A supportive and invested team worked well on pre-processing; in data cleansing, locations of the crimes were matched to the boroughs in the Airbnb dataset. There were considerable challenges met in matching the borough names, as documented in the 'Implementation Challenges' section below. As the data cleansing was linked to priorities for the in-depth analysis, this was an iterative process; it was important to ensure that the target columns were kept in the dataset, and as the investigation was refined, columns were removed and re-added.

In-Depth Analysis

In-depth analysis of the crime dataset was refined by:

- Overall view of crime type occurring across the 33 London boroughs.
- Following crime trends across the specified time period.
- Highlighting the top 5 boroughs in London for the highest and lowest crime rates.
- Comparing the average price of houses in each area to the rate of crime.
- Investigating case-handling of crimes across crime hotspots and safe havens

The approach to in-depth analysis of the Airbnb dataset was:

- To filter the set to target dates, relevant property types and to remove outliers from the price variable.
- To focus on, in order: the host variable, room types, and the price variable. Review count analysis is also present in the notebook. The variables were examined via top and bottom five boroughs for property count, and top and bottom five boroughs for average price.

IMPLEMENTATION AND EXECUTION

Development approach and team member roles

- Tasks were allocated according to results of the SWOT analysis.
- Collaboration, progress and decision-making was fraught at times, this in itself had an impact on team members and progress. The group effectively managed to work through these challenges.
- As noted below, it was not possible to implement a full agile development process. Work on this project allowed the team to see the value of a project manager or scrum master who is dedicated to supporting the team.
- Team member roles developed organically; team members would take responsibility for a challenge or task they were interested in doing as it came up.

Tools and libraries

- Key data science libraries used: Pandas, Numpy, Matplotlib, Seaborn, Folium, Requests, Scipy
- Online collaboration tools: Miro for project timeline and development, Jam Board - for SWOT analysis
- Code management: Jupyter Notebooks for initial exploratory analysis, Google Colab, Git and Github
- Google Drive for shared documentation tracking, Slack & Zoom for communication channels

Implementation process

Following the initial project development stage and after the steps documented in Implementation Challenges, data collection and preparation for analysis began. The data was cleaned to keep only the information relevant to this project, ensuring a common column for the datasets to be merged and compared (LSOA columns, seen in Data Cleansing files on GitHub). This took time, due to coding issues, see Implementation Challenges. Once cleaned, in-depth analysis could proceed and conclusions could be drawn.

Agile development

Ultimately, the team did not have capacity - either in terms of team members, or time - for a full implementation of agile development principles. However, the elements that were factored in are:

- **An adaptive process**, responding to change - the team faced a number of challenges and was able to adapt expectations and methodologies to support the broader project.
- **Iterative development** - There wasn't capacity to implement a full sprint cycle however, some of the principles of a sprint were followed, for example, tasks were divided up and progressed iteratively. The data cleaning was initially worked on separately with many challenges, so iterative approaches were used to create the final code.
- **Daily Standup** - Some of the principles of the daily standup were evident in team meetings, whilst we also adapted the model to fit individual schedules.
- **Refactoring** - The members of the team working on the data cleaning and data analysis reviewed each other's code and undertook code refactoring methods to bring the code into order. The team members working on analysis also reviewed and shared code, though this was curtailed by time constraints.

Implementation Challenges

Interpersonal/Group challenges
The team faced significant challenges in terms of collaboration and group dynamics. The balance between necessary autonomy and dependence was never resolved and led to negative impacts on the team and process. On reflection, this did lend the project a 'waterfall' rather than 'agile' tendency; in future, it would be wise to avoid this as it does not support the programming process. Importantly, however, the project did progress.

Technical challenges
The team initially had a technical issue with collaboratively using Google Colab to share code, as they were unable to run the shared code and read from the same files. Lottie spent significant time researching and discussing resolutions to this and was able to create a Google Drive for all members to access and retain shared file paths.

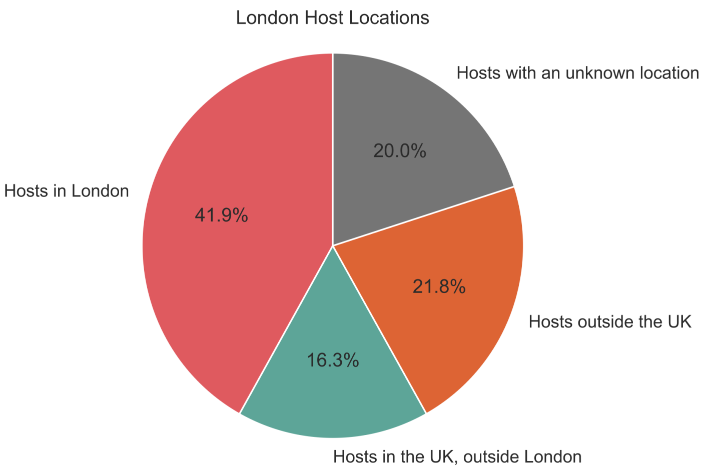
Coding challenges
The team have documented in Table 1, and in 'Archived_files' on GitHub the various issues that were faced with coding. It is important to document the different avenues taken to solve each problem, as they each contribute to an eventual solution. Note that significant coding challenges faced in analysis have not been included due to time pressures, they are documented in the notebooks.

Table 1: Table showing coding challenges the team faced, the location of the attempted code and how this was overcome

What we were trying to do:	Which file the data attempts can be found in	How it went:
Connect to police API to get neighbourhood names from LSOA codes	Archived_files -> neighbourhood_api_attempt.ipynb	We were able to connect to the Police API, and write a working code to convert the LSOA codes in the crime dataset to police neighbourhood names (via the neighbourhood codes). However, the LSOA codes and the neighbourhoods used by the Met police are not the same.

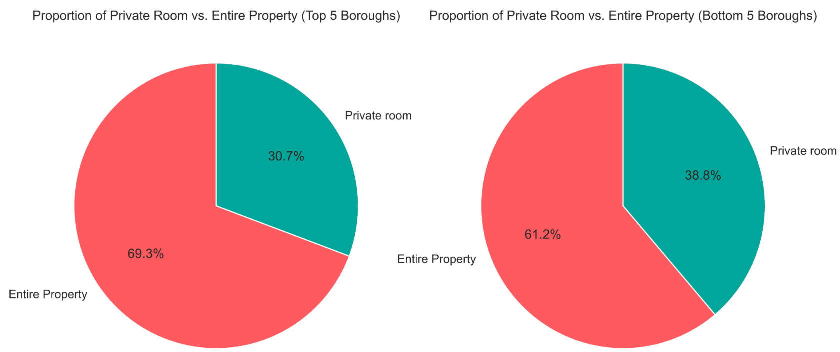
Using 'pip install police-api-client'	Archived_files -> neighbourhood_api_attempt.ipynb	After initially connecting to the Police API, it was discovered that there was a 'pip install' command for the API. This was successfully installed and run, and then used for the next step. However, the same issue was run into with the mismatch between LSOA and neighbourhood codes.
Connect to API to reverse geocode the locations of the crimes using the longitude and latitude of each crime	Archived_files -> neighbourhood_api_attempt_v3_final.ipynb	We were able to connect to the API, and (after working as a team) extract a suburb or district for each entry. However, due to time-space complexity, this process could not be completed because it took too long to run the code based on the amount of data we had.
Locate the crimes within boroughs by using the longitude and latitude, and a database which had the boroughs' boundaries (in easting and northing)	Archived_files -> neighbourhood_api_attempt_v2.ipynb	We connected to an API that could convert longitude and latitudes into easting and northings. After conversion (via function), the latitude and longitude boundaries were established, and a code written to establish the boroughs in which each crime fell. The final function did not fully work, and there were still issues with space-time complexity. At the time of working on improving, another team member developed the data cleaning solution, and so this code was abandoned.
Data cleaning - removing the final numbers at end of LSOA name	Archived_files -> matching_london_boroughs_police_dataset-NOW_INCORPORATED_INTO_DATA_CLEANSING_NOTEBOOK.ipynb	Some of the LSOA areas we saw during our data cleaning were outside of the City of London (despite being a part of the Metropolitan police force data), and boundaries changed. This was resolved by sorting the LOSA area names by stripping five indexes off the end of each value and removing any areas that are not in London using a list.
Data cleaning of Airbnb data	Archived_files -> airbnb_attempt.ipynb	Originally the Airbnb datasets for June 2022 and March 2023 were cleaned separately and the data was explored as separate files. After a code review and discussions, we then decided that it would be better to join the files for each month so that analysis can be done using one file to observe trends.
Use the Stop and Search data from the crime dataset	Archived_files -> 'police_stopandsearch_dataset_cleansing_general.ipynb'	The stop and search dataset only included the longitude and latitude coordinates, and did not have the LSOA names. We could not process this much data with the coordinates alone, as we did not have the capacity to turn this many coordinates through an API to get the London Borough names (time-space complexity, see above).
Trying to get London Borough name from Latitude and Longitude points	Archived_files -> geopy.py	Tried to use the Geopy API to get London Boroughs from Longitude and Latitude. Tests revealed that "city_district" was the closest match but was not consistent. Not enough memory to handle 1000s of data points, would not be able to perform effectively for our data.
Testing plotting points on map of London	Archived_files -> folium_test.py	Code could generate the map but crashed as I had tried to plot too many points at once (wanted to have distribution of crime across London)
Exploratory analysis of crime and airbnb datasets for March	Archived_files -> crime_march_exploratory.ipynb Airbnb_exploratory.ipynb	Explored both datasets, removed null values and attempted to deal with outliers. Made a few visual plots for the group to look at. Was able to change LSOA_name to matched neighbourhoods reliably and accurately. Had a few discussion points for the team meeting on how to move forward with analysis and visualisations

RESULT REPORTING



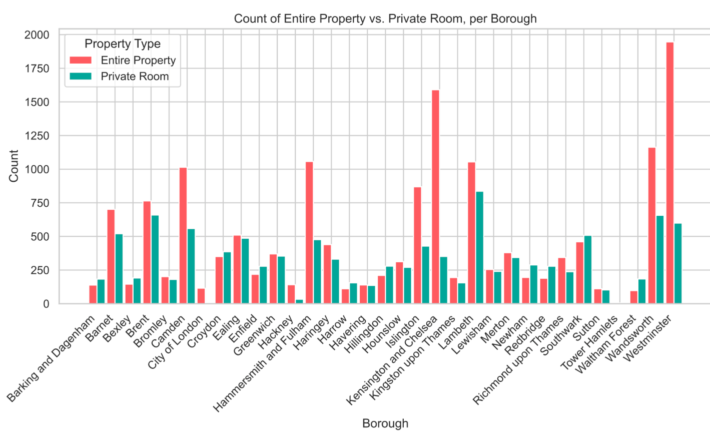
Key insights following thorough analysis of the Airbnb dataset are:

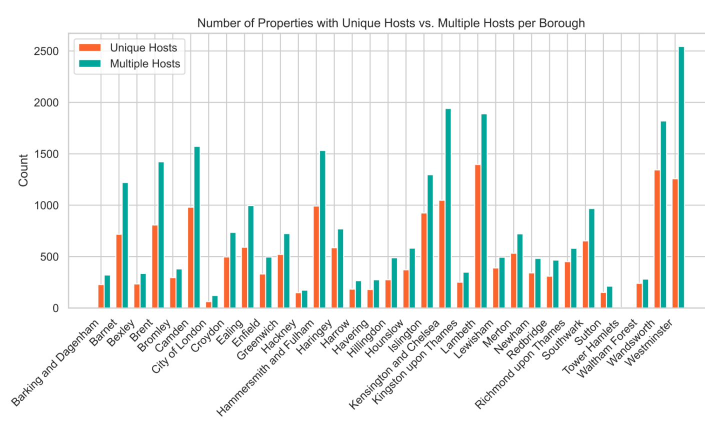
Of the hosts in this set, 42% are located in London and 38% are definitively based outside of London or outside the UK, with 20% having an unknown location. This suggests that the platform facilitates significant trade beyond its sharing economy ethos.



In every London borough, properties listed by hosts with multiple entries outnumber single listings by individuals sharing their homes. Listings for entire properties outnumber listings for private rooms, not only in the top and bottom five boroughs by property count, but across the set. This suggests that Airbnb has a significant overlap with conventional letting agencies. Therefore Airbnb may be amongst a number of factors contributing to the housing shortage.

Kensington & Chelsea and Westminster appear in the top five boroughs for both count of property and average price. City of London appears in the bottom five for property count and the top five for average price. Higher average prices are seen in central areas, attractive to tourists. This suggests significant overlap with the traditional travel industry, while the industry itself may also be drawing benefits from the platform.



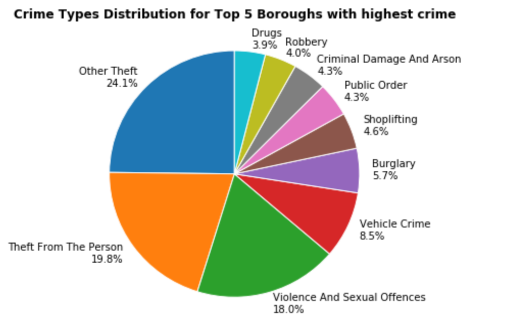


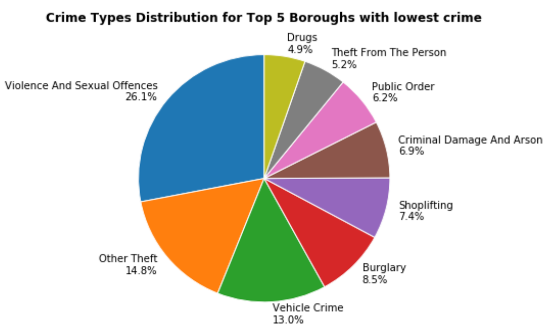
Close examination of pricing reveals unusual values and significant range remains after adjusting high and low values. Further research reveals that this is not unexpected, along with other e-commerce platforms, Airbnb has been probed for facilitating money laundering. In itself, this merits further research.

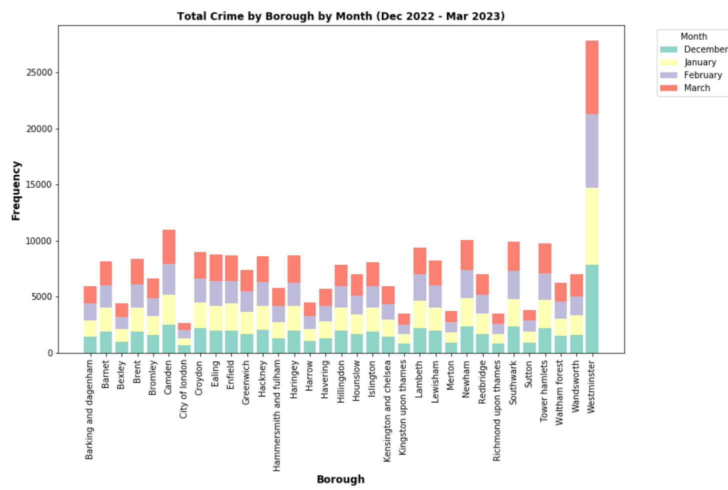
Key insights following thorough analysis of the crime dataset are:

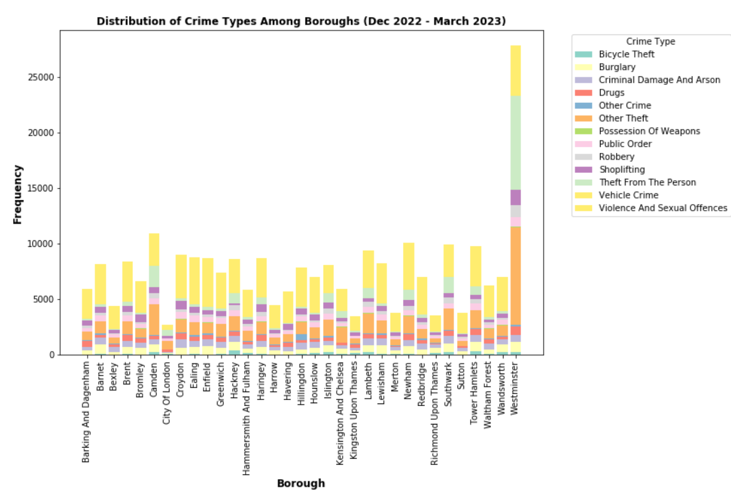
The top five boroughs with highest crime rates are: Westminster, Camden, Newham, Southwark, Tower Hamlets.

The top five boroughs with lowest crime rates are: Sutton, Merton, Richmond upon Thames, Kingston upon Thames, City of London.



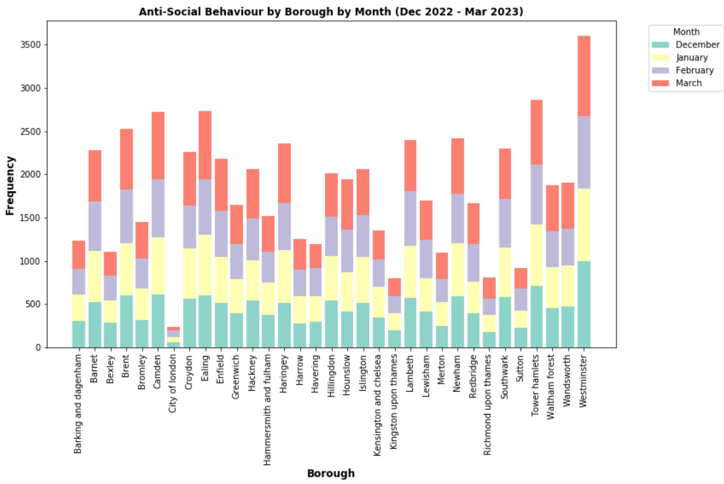






Top three crime types for the boroughs with the highest crime were: Other Theft, Violence and Sexual Offences and Theft from the Person.

Top three crime types for boroughs with the lowest crime were: Violence and Sexual Offences, Other Theft and Vehicle Crime.

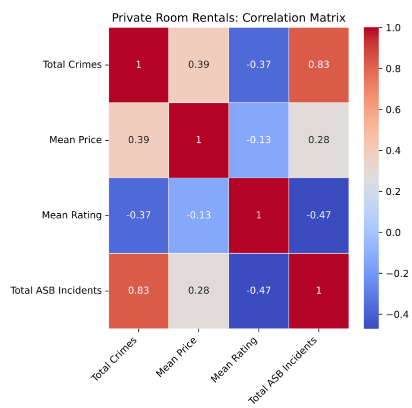


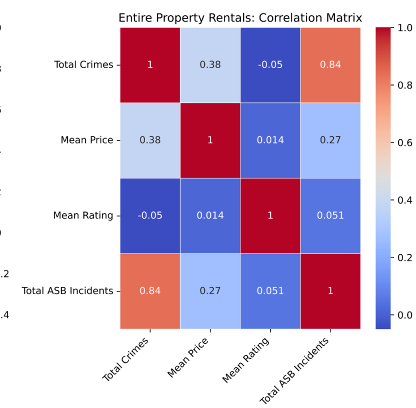
Top five boroughs with highest Anti-Social Behaviour are: Westminster, Tower Hamlets, Ealing, Camden and Brent. Top 5 boroughs with lowest Anti-Social Behaviour are: Merton, Sutton, Richmond upon Thames, Kingston upon Thames and City of london.

Anti-social behaviour (ASB) and crime has increased from December 2022 to March 2023.

CONCLUSION

The project set out to measure and analyse the social impact of Airbnb on London's boroughs. Tentatively, the outcomes find some limited evidence of a measurable impact on social conditions. The correlation matrix reveals positive correlations between total crime and ASB incidents, as expected; it also reveals weaker positive correlations between mean price of private Airbnb rooms and total crime and ASB incidents. The results suggest that further examination of the platform's impact on the city are indicated, and that this should take in a broader sub-set of measures. Larger datasets encompassing greater time intervals might allow for firmer conclusions to be drawn. Further research is indicated into the platform's impact on the housing crisis (defined by rental demand and pricing, and availability and pricing of first-time buyer properties), as well as crime levels, including anti-social behaviour. The outcome stops short of claiming that Airbnb further reduces availability of housing in an already difficult context, but the project contributes to the picture that Airbnb is a player in a London context with some tangible and difficult to measure negative social impacts.





The group would like to acknowledge both the considerable challenges that have been faced during their collaboration, and the fact that the valuable individual skills and insights of all team members, resolve and determination have enabled the project to progress to completion.

References

Ke, L., O'Brien, D.P. and Heydari, B. (2021) "Airbnb and neighbourhood crime: The incursion of tourists or the erosion of local social dynamics?," *PLOS ONE*, 16(7), p. e0253315. Available at: https://doi.org/10.1371/journal.pone.0253315.

Hati, S.R.H. *et al.* (2021) "A decade of systematic literature review on Airbnb: the sharing economy from a multiple stakeholder perspective," *Heliyon*, 7(10), p. e08222. Available at: https://doi.org/10.1016/j.heliyon.2021.e08222.

Shabrina, Z., Arcaute, E. and Batty, M. (2021) "Airbnb and its potential impact on the London housing market," *Urban Studies*, 59(1), pp. 197–221. Available at: https://doi.org/10.1177/0042098020970865.