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

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

## INTRODUCTION

#### Aims and objectives of the project

Hypothesis: Airbnb rentals have a negative social impact on London boroughs.

This study investigates the growing concerns of the general public from London boroughs who are concerned that the rise in home-sharing platforms like Airbnb have had an impact on crime and anti-social behaviour in their neighbourhoods (Ke, O’Brien *et al*., 2021).

The aim of this study is to evaluate the impact that the growth of Airbnb has on London 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

Within this report we have detailed the background to our study and question ([*Background*](#_9jmyvtvidg16)*)*, the steps we took to implement it ([*Steps Specification*](#_o4i4qmlsjiph)), the challenges we faced during this project *(*[*Implementation Challenges*](#_qr7zz5wtxeuv)), what we found ([*Result Reporting*](#_3vqdv0etgjlq)), and what we learnt from the project as a whole ([*Conclusion*](#_8rwob0vgu671)).

## BACKGROUND

Airbnb is an online platform that allows property owners to host their homes (or second properties) for short or long-term lets as alternative accommodation to hotels and hostels. To date, Airbnb has rapidly gained popularity worldwide due its affordability and diverse range of quirky accommodation since its launch in 2009 (Hati *et al*., 2021). However, studies have suggested a negative impact on the local communities as a result of housing market disruption and increased crime rates (Ke, O’Brien *et al*., 2021, Shabrina *et al.*, 2021). With our team all having some connection to London, we passionately feel concerned about the growing Airbnb market. Therefore, we focused our investigation on analysing the social impact that Airbnbs have across London, with a main focus on crime.

We conducted our analysis using open-source data available from the Metropolitan Police, Airbnb and housing datasets to determine correlations between Aribnb listings in London boroughs with crime reports in these areas and housing prices. We will then use visualisations to demonstrate our findings and complete our objectives. These results will help to draw inferences on trends, however it is important to highlight that correlation is disparate from causation.

Our project aims to share our findings to an audience ranging from local politicians and smaller local committees including neighbourhood watch groups, as well as prospective Airbnb owners and people looking to buy or rent properties in London. The results and analysis will be presented clearly in layman's terms for anyone interested to understand, with appealing visual aids to make the results more accessible to everyone, to form their own opinions on the impact of Airbnb in their neighbourhoods.

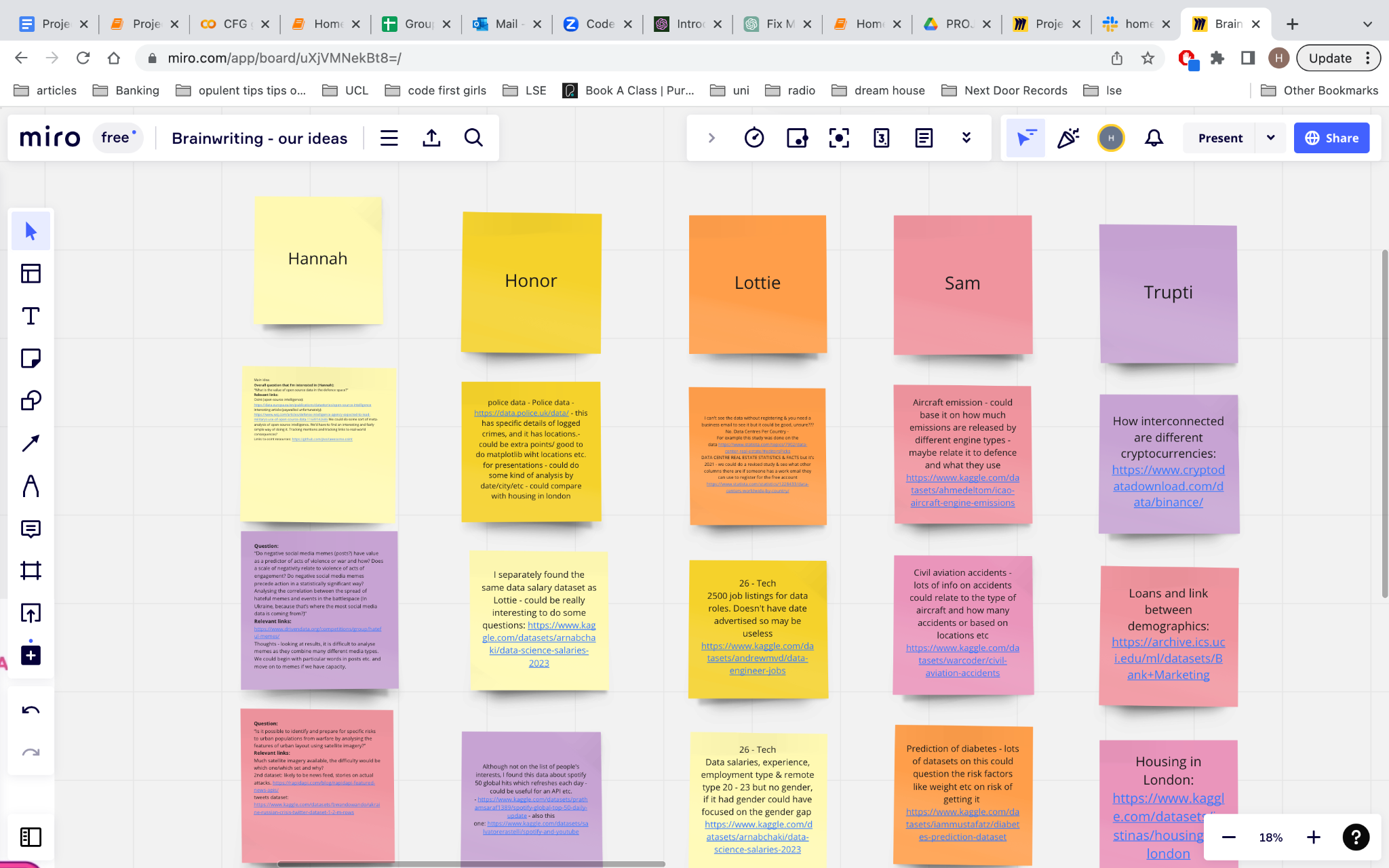
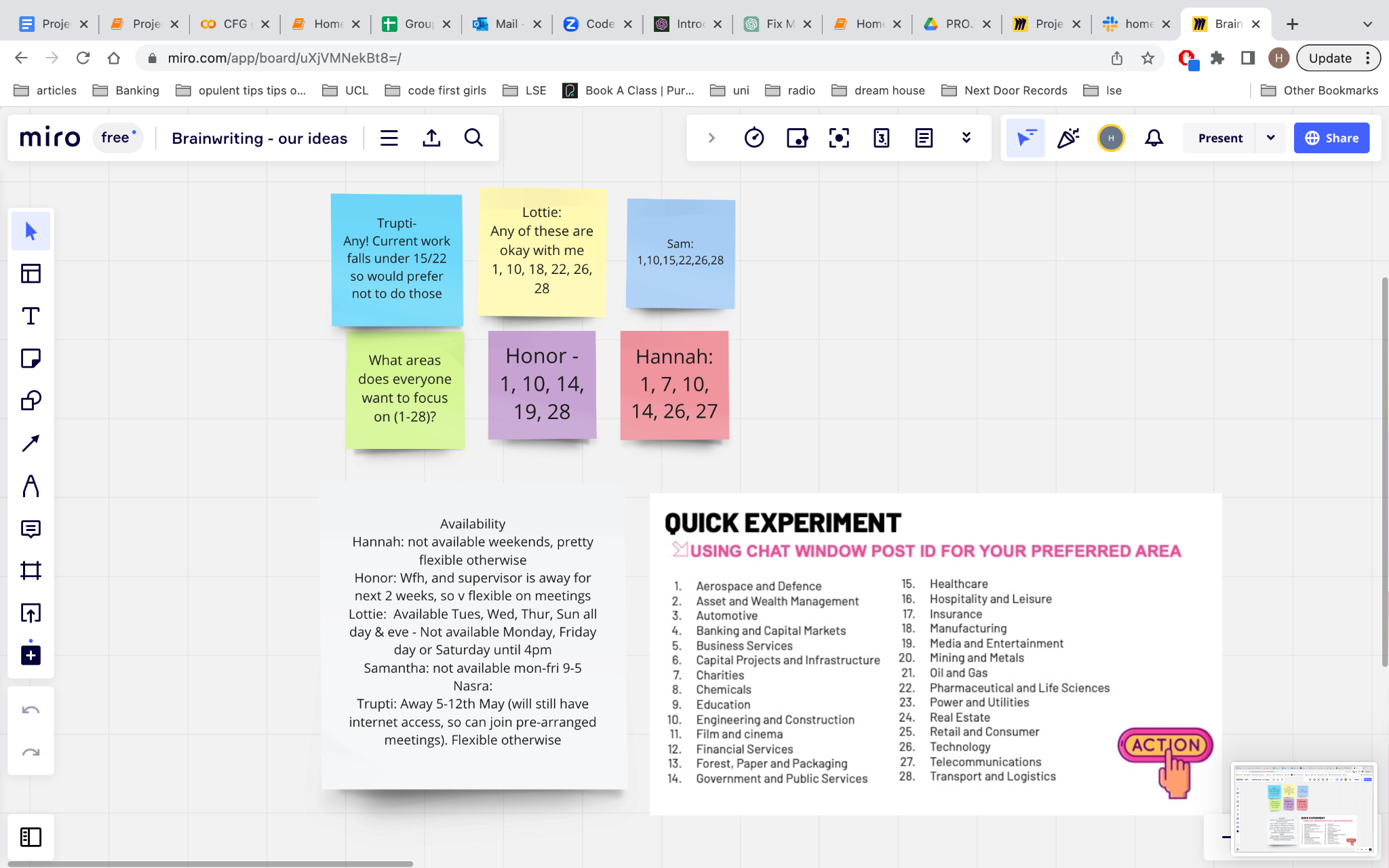
This project aligns with topics covered throughout the Data specialisation pathway of the CFG degree course, with use of many data libraries including NumPy, Pandas and MatPlotLib, as well as coding techniques taught throughout the course, data cleansing/ transformation and visualisation creation to interpret data analysis.

## STEPS SPECIFICATIONS

#### Describe how your team approached each of the key steps of the data analysis: framing questions, data gathering, preprocessing, in-depth analysis etc.

#### Framing Questions

Our team began the group project by getting to know each other and our abilities. We created a Slack channel in which we could communicate, and two Miro Boards. The first was our [self-identified strengths and weaknesses (SWOT)](https://github.com/LottieJane1312/cfg_data2_group7_project/blob/main/homework_week2_specialisation_grouphomework/data/SWOT.png), to assist with assigning roles. On the [second board](https://github.com/LottieJane1312/cfg_data2_group7_project/blob/main/homework_week2_specialisation_grouphomework/data/brainwriting1.png) we each put which areas, from the 1-28 list provided by CFG, we would be interested in working on. We continued to add to the second board with our ideas for the project [as they developed](https://github.com/LottieJane1312/cfg_data2_group7_project/blob/main/homework_week2_specialisation_grouphomework/data/brainwriting3.png) (Fig. 1-2):



*Figure 1 (left): Miro board topics each member was keen to work on. Figure 2 (right) showing project ideas.*

We decided that the questions and data sets were 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). We therefore endeavoured to present each data set with its potential line of questioning.

Ideas that we discussed, which evolved into our final question, were on housing, crime statistics, and open source intelligence. We concluded that, due to the time constraints of the project, any questions relating to open source intelligence would not be feasible at this stage. After looking into housing and crime (while discussing ideas from happiness levels to weather) we focused on temporary housing; Airbnb. Once we decided on our topic of focus, we conducted exploratory data testing and analysis (as can be seen in the ‘Initial Stages - Exploratory Analysis’ file, in ‘Archived\_files’ on GitHub).

#### Data Gathering and Sources

Our research began by looking at Kaggle, government and policy sources, as well as activist sources. We looked for open-source verified datasets with sufficient entries for our project that we could feasibly produce code to join datasets and create a detailed analysis to meet the project brief set by CFG.

Following multiple discussions within the group and with our supervisor Kosy, we settled on datasets, with extensive data on Airbnb, crime and housing. [Inside Airbnb](http://insideairbnb.com) is a data activism project working to advocate for communities affected by Airbnbnb. The data presented us with a lot of information (66,152 rows x 74 columns) and avenues for questioning. The crime dataset was sourced from <https://data.police.uk/data/>, again offering a wide range of anonymous information from the London Metropolitan Police central database (441,293 rows x 12 columns). We noted that the crime dataset contained anomalies such as inclusion of crimes occurring outside of London, and that the 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 <https://www.rightmove.co.uk/house-prices-in-London.html>, and summarised in [this article](https://www.standard.co.uk/homesandproperty/property-news/london-house-prices-every-borough-ranked-rise-fall-b1074007.html). We will be comparing data over a four month period from 11 December 2022 - 14 March 2023 to interpret our findings. This time period was chosen based on the datasets we had found available and what would be a representative time period to draw conclusions from.

We had also looked at using Stop and Search data provided by the Metropolitan Police to include in our analysis, however upon further inspection we noticed that the data did not have LSOA codes to identify locations which we had planned to use to match our different datasets with. Therefore this data was omitted from our project, but could be used for further studies.

#### Pre-Processing

As detailed in the project tracker, after deciding on our question and the datasets we would be using, we assigned tasks and got to work! Team members worked individually as well as in pairs, working in tandem on different tasks. These tasks were assigned based on each person's skills and weaknesses, discussed on the Miro boards, and what each person wanted to do. When anyone got to an obstacle in their task, we would share code and discuss on Slack to help each other.

The main task of pre-processing was data cleansing, where we matched the locations of the crimes to the boroughs in the Airbnb data and decided on relevant data for our analysis. We had many issues with matching the borough names, as documented in the ‘Implementation Challenges’ section below. As the data cleansing was linked to the in-depth analysis, it was an iterative process; we had to ensure that the columns we wanted to question were kept in the dataset, so as we refined our investigation we removed and re-added some columns.

#### In-Depth Analysis

* Overall view of type of crime occurring across the 33 London boroughs
* Followed crime trends across the specified time period
* Highlighted the top 5 boroughs in London for the highest and lowest crime rates
* Compared the average price of houses in each area compared to rate of crime
* Investigated case handling of crimes across crime hotspots and safe havens

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

* To filter the set to target dates, relevant property types and to remove outliers from the price variable.
* Following this, analysis focused on the host variable, room types, and the price variable. The variables were examined via top and bottom five boroughs for count, and top and bottom five boroughs for average price. The top and bottom five boroughs for count the room types; finally, the price variable was examined in detail, including top and bottom five boroughs for price.

## IMPLEMENTATION AND EXECUTION

#### Development approach and team member roles

* The team’s goals were twofold; we wanted to learn as much as possible from the project, but also complete it in the tight time frame, so allocated tasks according to the SWOT analysis.
* As we had never worked together before, we discussed in the first meeting about focusing on collaboration, and everyone’s voices being heard - this eventually led to issues, because we were not always able to make decisions and struggled to schedule meetings with everyone present at once.
* We did not use a scrum system as everyone preferred to work equally without a chosen leader/scrum master; throughout the project, however, we could see the value of having the scrum system after the issues we faced with making decisions and maintaining deadlines.
* Team member roles developed quite organically; team members would take responsibility for a challenge or task they were interested in doing as it came up, with us keeping in mind the strengths and weaknesses we documented to start.
* [A timeline](https://github.com/LottieJane1312/cfg_data2_group7_project/blob/main/homework_week2_specialisation_grouphomework/data/timeline.png) was created using a Miro board with tasks to be completed each week (Fig 3).

#### 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

#### 

*Figure 3: Miro board used as a timeline of key tasks to be completed each week.*

#### Implementation process

We began by framing the problem, in the first week, and presenting ideas. After the delays documented (Implementation Challenges below), we began data collection, and preparation for analysis. The data was cleaned to keep only the information we thought relevant to this project, and made sure there was a common column for the different datasets to be compared against (LSOA columns, seen in Data Cleansing files on GitHub). This took the most time, due to coding issues (Implementation Challenges). Once cleaned, we proceeded with an in-depth analysis 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, as documented below, 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 one final code (explained in Implementation challenges).
* **Daily Standup** - Some of the principles of the daily standup were evident in team meetings, whilst we also adapted the model to fit our schedules.
* **Refactoring** - The members of the team working on the data cleaning, reviewed each other's codes and undertook code refactoring methods to make the code look uniform and neater.

#### 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, the project was delivered.

**Technical challenges**

We initially had a technical issue with collaboratively using Google Colab to share our code, as we were unable to run a shared code and read the same files. Lottie spent a significant amount of time researching and discussing with course supervisors how to solve this issue, and was able to create a Google drive for all members to access so that the file paths would be the same.

**Coding challenges**

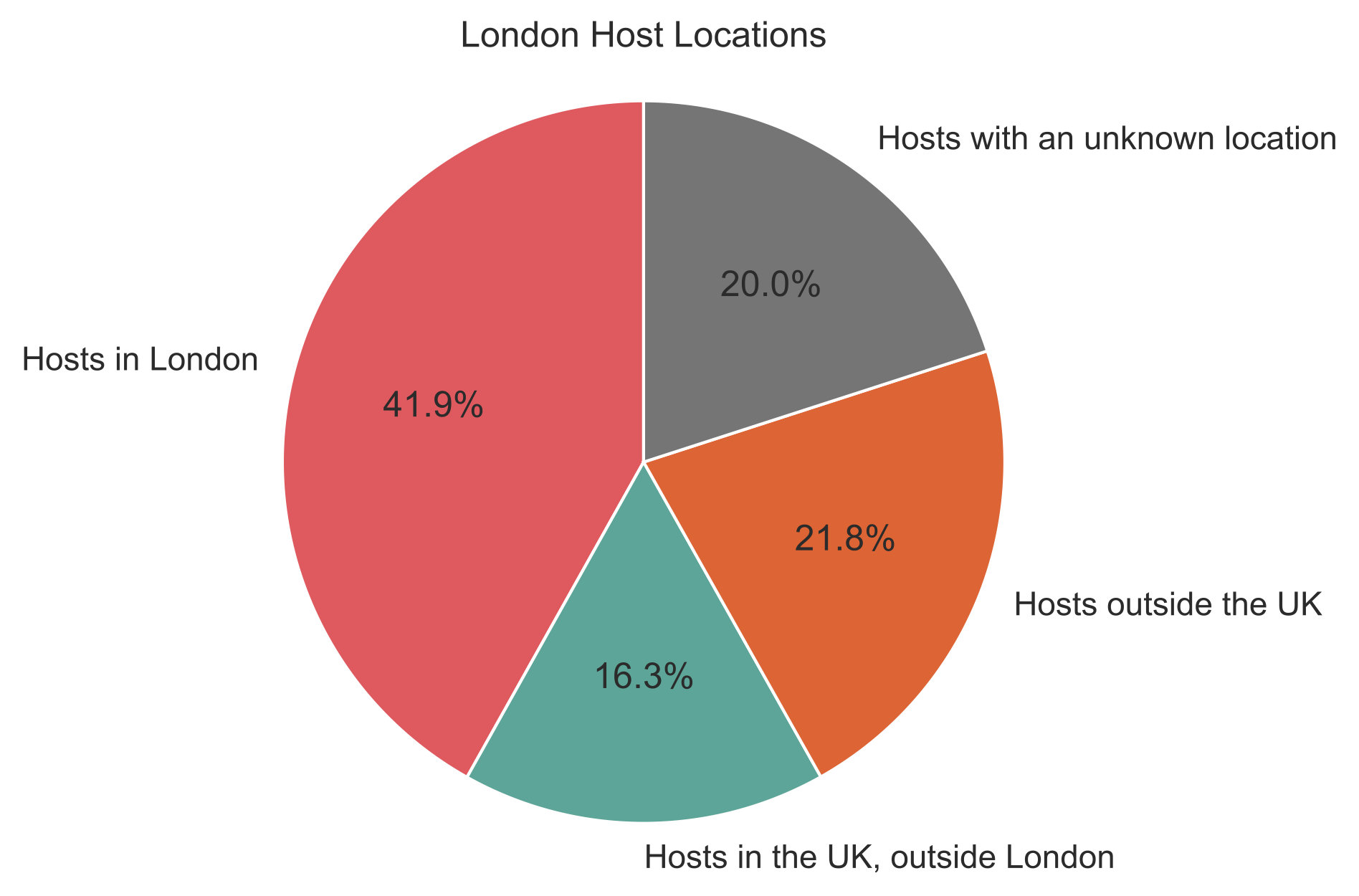
We have documented in Table 1, and in ‘Archived\_files’ on GitHub the various issues that we had with coding. We feel that it is important to document the different avenues we took to solve each problem, as they each contributed to directing us to our eventual solution.

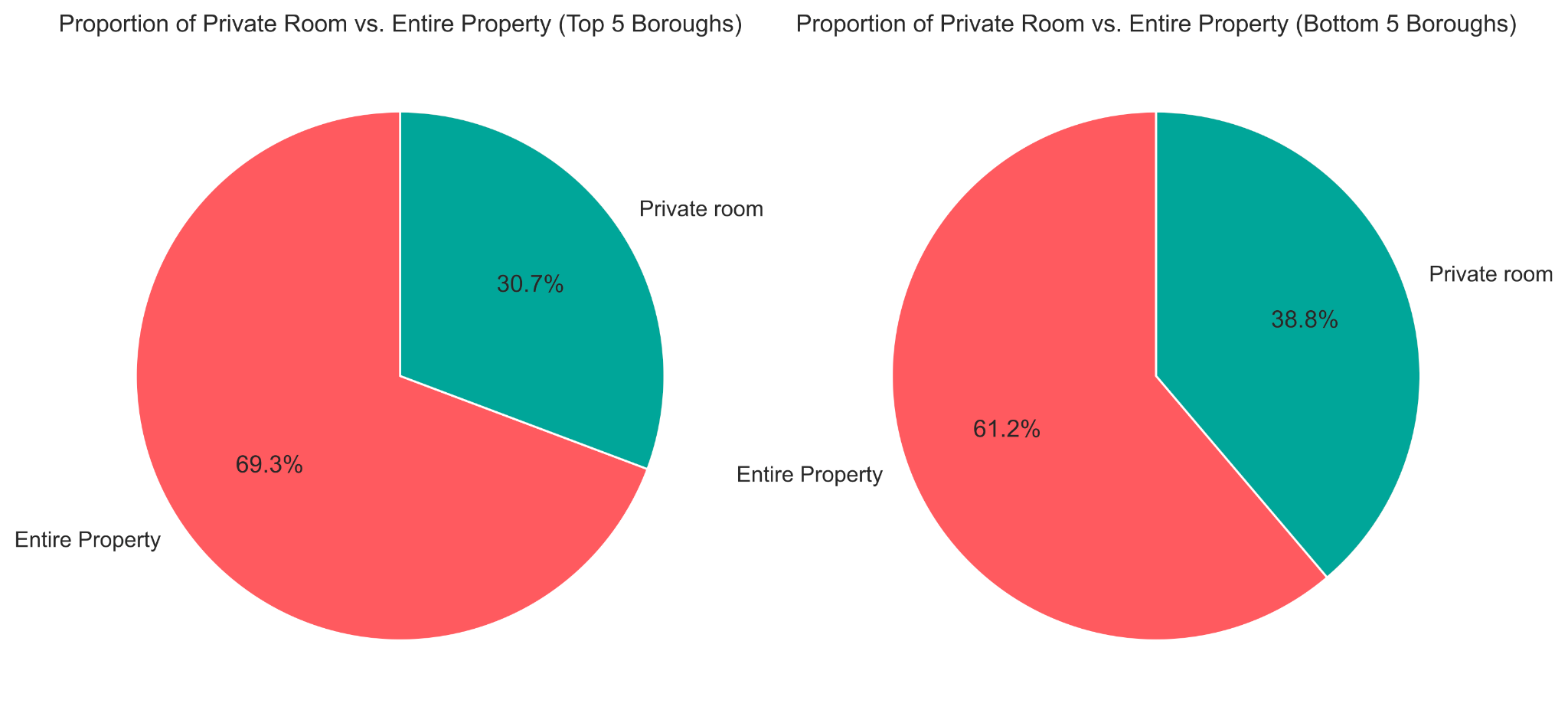
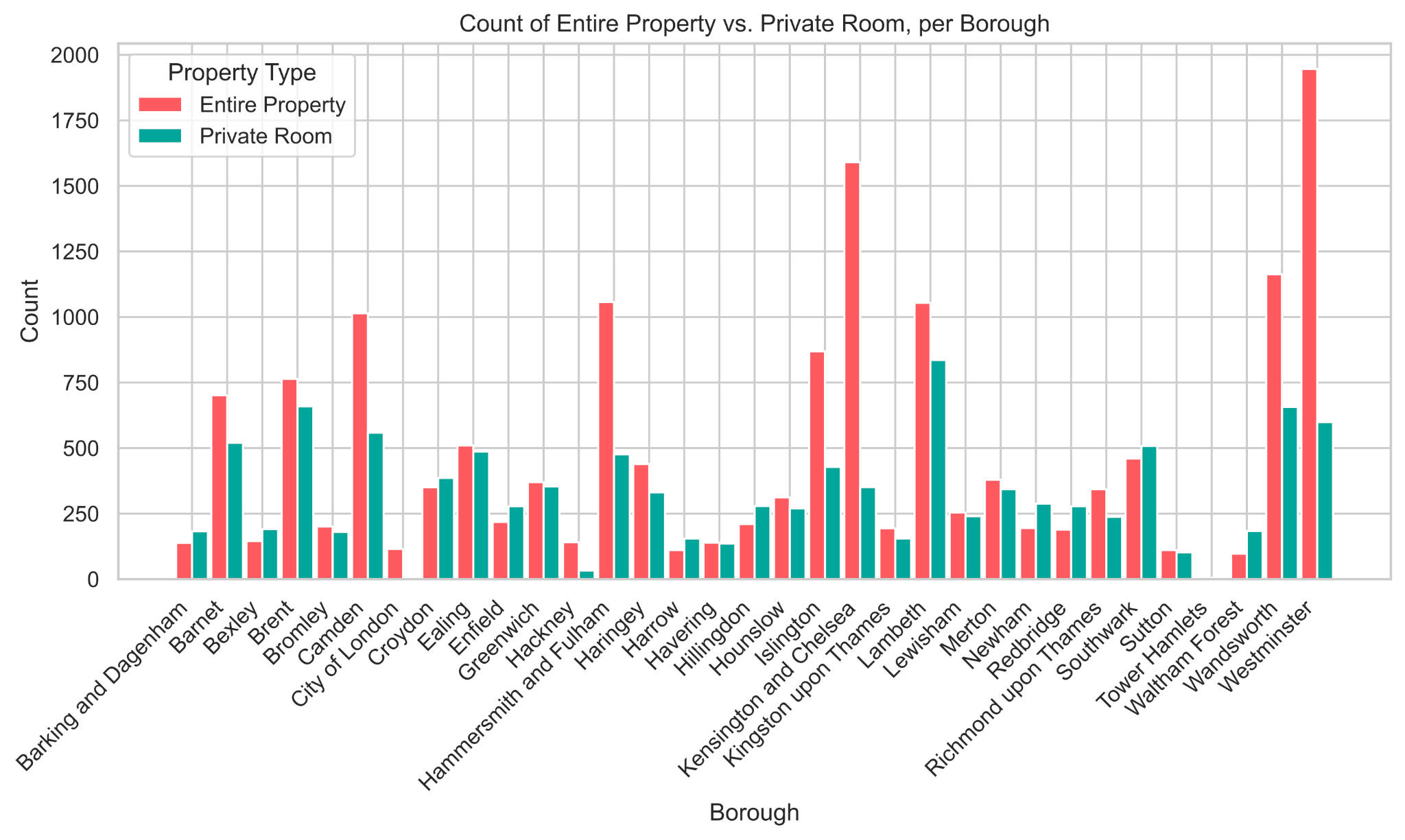
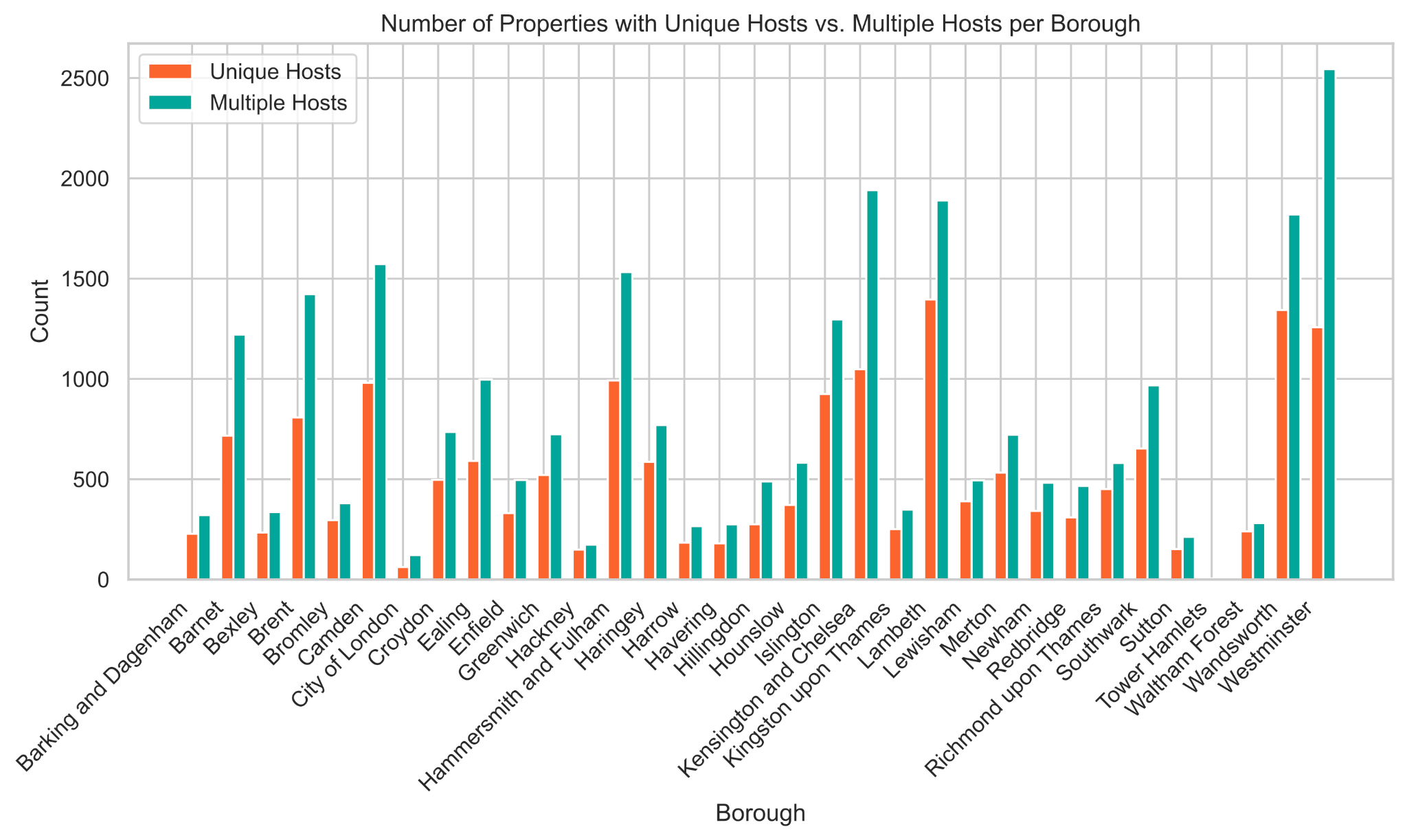
*Table 1: Table showing coding challenges the team faced, the location of the attempted code and how this was overcome.*

| **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 |

## 

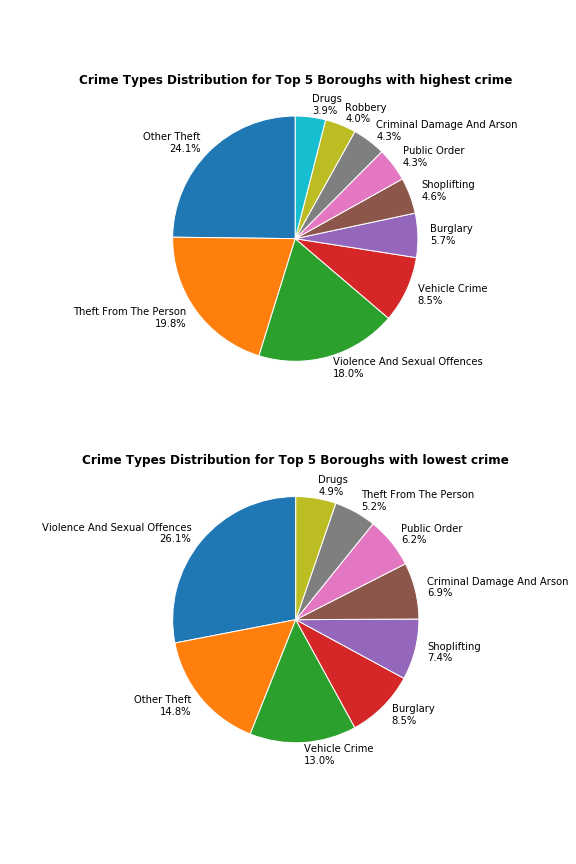
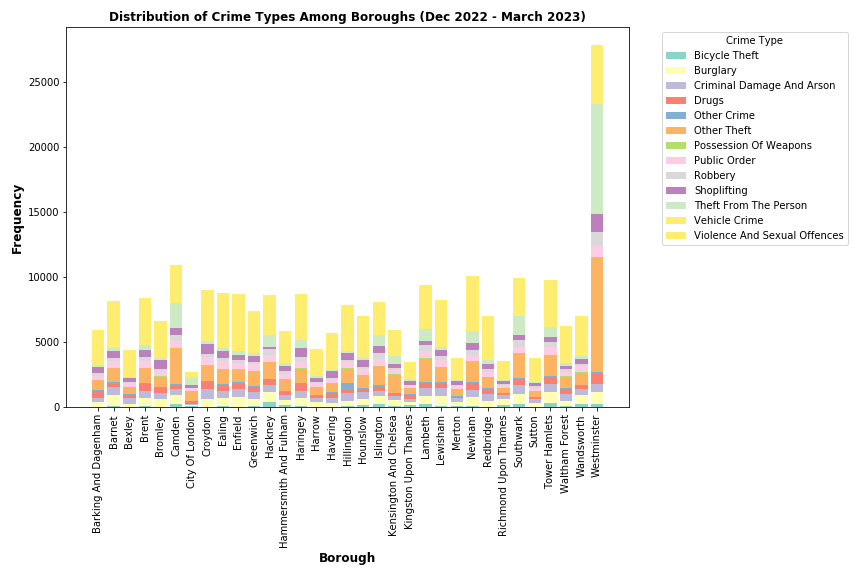
## RESULT REPORTING

**Key insights following thorough analysis of the Airbnb dataset are:**

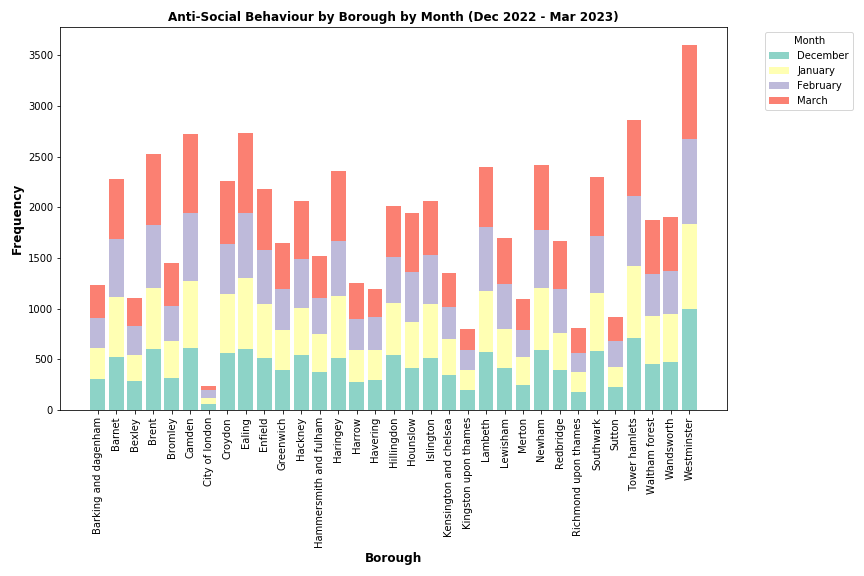
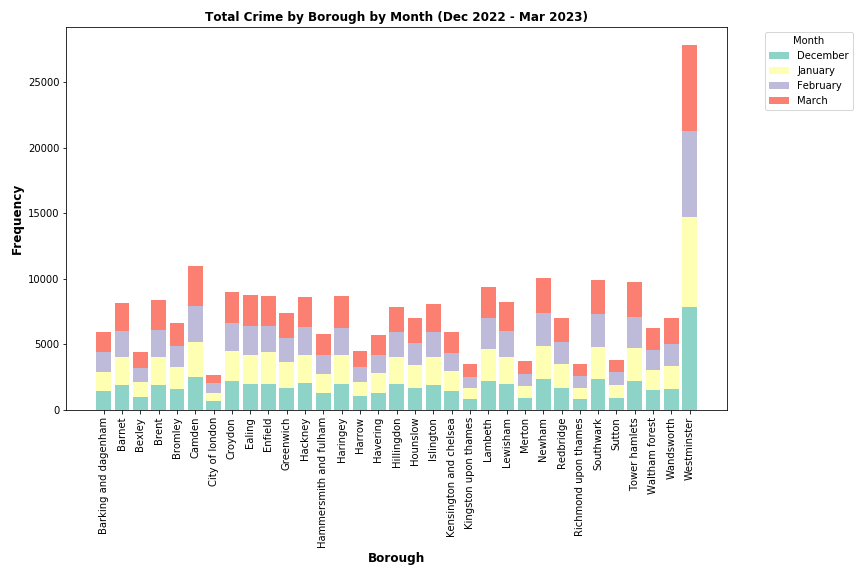
* Of the hosts in this set, 42% are located in London and 38% are definitively based outside of London or outside the UK, with 20% having an unknown location. This suggests that the platform facilitates significant trade beyond its sharing economy ethos. 
* In every London borough, properties listed 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](https://www.thetimes.co.uk/article/financial-conduct-authority-probes-airbnb-links-to-criminal-cash-2k0q6bpjh). In itself, this merits further research. 

**Key insights following thorough analysis of the crime dataset are:**

* Top 5 boroughs with highest crime: westminster, camden, newham, southwark, tower\_hamlets. Top 5 boroughs with lowest crime: sutton, merton, richmond upon thames, kingston upon thames, city of London. Top 3 crime types for the boroughs with the highest crime was Other theft, violence and sexual offences and theft from a person. Top 3 crime types for boroughs with the lowest crime was violence and sexual offences, other theft and vehicle crime, shown in Fig.

****

* Top 5 boroughs with highest ASB: westminster, tower hamlets, ealing, camden and brent. Top 5 boroughs with lowest asb 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

Suggested 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 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.

Due to the time constraints of this project and data freely available, we were only able to use limited data and conducted a month by month analysis. In the future we would like to conduct further research into the social impact of Airbnb, possibly using data spanning across a few years. It would also be interesting to further examine data and draw conclusions about the impact of Airbnb on the housing crisis, including:

* To analyse the impact of the Airbnb platform on the property rental sector
* To analyse house price changes in boroughs with a high or low burden of Airbnb properties.

Overall our team has come together to create a very interesting investigation on the social impact of Airbnb in London. Individually we have all enjoyed our experience working collaboratively and learnt a lot:

* Honor: *“I have really appreciated working in a group, and seeing how different people's perspectives and skill sets help, and take the project in different directions.”*
* Samantha: *“I have really enjoyed working with my group where we could all share different ideas and methods of overcoming challenges we faced to work together to create this project.”*
* Lottie: *“Working in a group, remotely, under such time constraints posed it’s challenges, however it’s proven extremely fulfilling to see our strengths shine through as the project took shape”*
* Trupti: *“I have enjoyed testing my python coding abilities, collaborating with and learning from my other team mates.”*

## 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.