Group Name's Group Project

Declaration of Authorship

We, `505 not found`, pledge our honour that the work presented in this assessment is our own. Where information has been derived from other sources, we confirm that this has been indicated in the work. Where a Large Language Model such as ChatGPT has been used we confirm that we have made its contribution to the final submission clear.

Date: 17/12/2024

Student Numbers: 24082251 24065306 24133780 24057655 24046535

Research propose

Exploring the Impact Profile of on London Communities

Research objectives: To assess the impact of Airbnb across different regions by geodemographic categorisation of different London neighbourhoods and to provide scientific evidence for policy makers to understand which neighbourhoods are affected by Airbnb short-term rentals (STL) and which are likely to face this impact in the future. Differentiated policy recommendations will be proposed to avoid one-size-fits-all regulation.

Response to Question

1. Who collected the InsideAirbnb data?

Inside Airbnb data was collected by Murray Cox, a data activist and the project's founder, John Morris, the website designer and report producer, and Taylor Higgins, a master's student focusing on sustainable tourism at the Università degli Studi di Firenze.

2. Why did they collect the InsideAirbnb data?

The purpose of InsideAirbnb data is to provide data-driven insights into the impact of Airbnb on residential housing markets, thereby contributing to public discourse on the regulation and effects of short-term rental platforms in urban areas.

3. How did they collect it?

The data is collected by utilizing web scraping techniques such as self-made bots, inside Airbnb and AirDNA (Pawlicz and Prentice, 2021) (Prentice and Pawlicz, 2023) to extract publicly available information from Airbnb's website, focusing on various aspects of listings such as location, price, availability, and host details. This approach allows for the assembly of comprehensive datasets, which are then cleansed and organized to facilitate thorough analysis.

4. How does the method of collection (Q3) impact the completeness and/or accuracy of the InsideAirbnb data? How well does it represent the process it seeks to study, and what wider issues does this raise?

The data collection method used by InsideAirbnb raises data quality issues, mainly related to data incompleteness, reliance on website structure, and technical challenges. In terms of accuracy, data is automatically retrieved from the website, which possess the risk of capturing inaccurate or outdated information due to the dynamic nature of web content (Krotov and Johnson, 2023).

In addition, due to privacy measures, the geographic coordinates provided by Airbnb may not reflect the exact location of the listing, which adds a layer of inaccuracy. And as web scraping depends heavily on the structure of the Airbnb website. Changes to the website layout or measures to block scraping activities may disrupt data collection efforts, like Airbnb's anti-scrap measures including CAPTCHA or IP bans, pose additional challenges (Prentice and Pawlicz, 2023). This burdens data analysts by requiring them to constantly develop and maintain scraping scripts.

In the discussion of the structure of InsideAirbnb data, it contains all aspects of the Airbnb market, including the distribution and characteristics of listings, pricing models, and the impact of Airbnb on the local housing market. And it is a relatively complete dataset and can assist with comprehensive analysis study.

Besides the accuracy concerns, the use of InsideAirbnb data raises technical, legal, and ethical issues. Legally, as discussed in Sobel (Sobel, 2021), scraping faces challenges in different jurisdictions, depending on how it intersects with privacy laws and terms of service agreements. This could affect the legality of the Inside Airbnb data collection process, especially if it violates Airbnb's terms of service. Scraping also raises ethical issues, particularly regarding the consent of data subjects (Airbnb hosts and guests) whose information is collected without explicit permission. This raises significant privacy issues, as highlighted in the study by Xie and Karan (Xie and Karan, 2019), where users' awareness and concerns about how their data is used influence their privacy management behaviours.

5. What ethical considerations does the use of the InsideAirbnb data raise?

The use of the InsideAirbnb database does raise several ethical considerations.

Firstly, there are issues of legal compliance. Web scraping can conflict with legal standards and ethical norms, particularly when data is collected without explicit consent, potentially leading to legal actions (Krotov and Johnson, 2023).

Secondly, privacy concerns for individuals must be addressed. Although the data might be publicly accessible, individuals typically do not anticipate their rental information being extensively aggregated and analyzed (Brenning and Henn, 2023).

In many instances, data subjects (hosts and guests) are neither directly informed nor asked for consent when their data is scraped and analyzed. This presents a significant ethical dilemma: using their information without explicit permission, especially when such data might be utilized to draw conclusions or influence policies that could directly impact them.

Moreover, there is the issue of how policymaking might be influenced by the data. Since the scraped data can contain errors, issues with accuracy and potential misrepresentation may lead to misleading conclusions that could negatively affect Airbnb hosts, guests, and policy decisions.

Additionally, the misuse of data poses a significant ethical concern. When analyzing Inside Airbnb data, it is crucial to ensure that the data is not used for purposes unintended by the original data providers, such as market manipulation, unfair competition, or research that adversely impacts hosts and guests.

Lastly, transparency and accountability are crucial. Ethical research involving data scraping should clearly disclose its methodologies, the specific data collected, and how this data is utilized. Such transparency is especially important for accountability, particularly if the research has the potential to influence public opinion or policy (Brenning and Henn, 2023).

6. With reference to the InsideAirbnb data (*i.e.* using numbers, figures, maps, and descriptive statistics), what does an analysis of Hosts and the types of properties that they list suggest about the nature of Airbnb lettings in London?

6.1 Analysis of Hosts

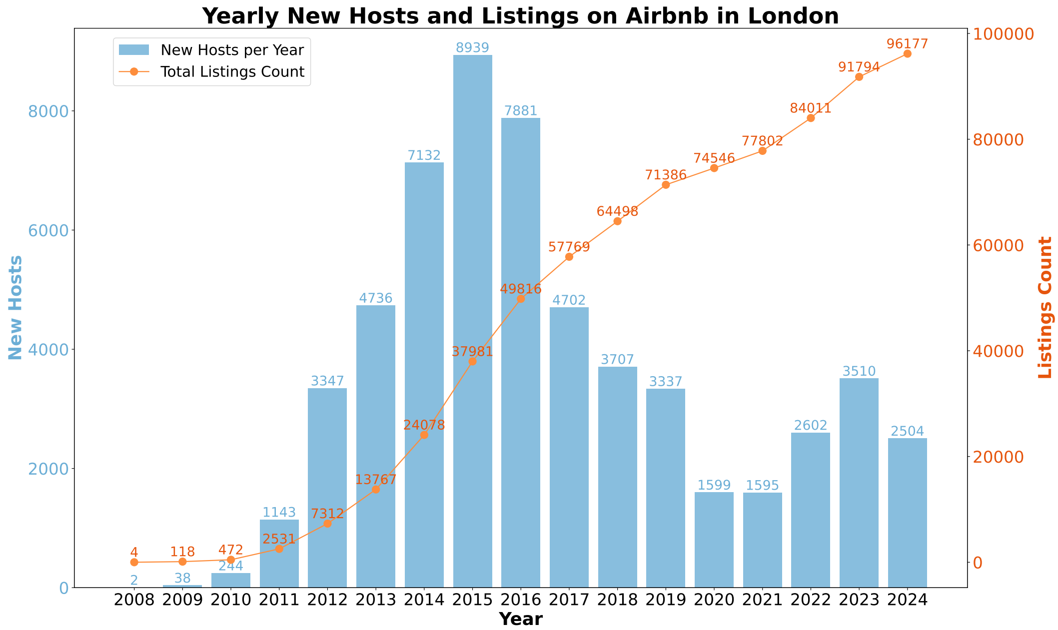
6.1.1 Distribution of the Number of Listings per Host

In London, only 47.8% (45,932 listings) are owned by single-listing hosts, while the remaining 52.2% are held by multi-listing hosts.
Notably, hosts with 10 or more listings account for 20.8% (20,038 listings) of the total.

conclusion:

1. Prevalence of multi-listing hosts: more than half of all listings owned by multi-listing hosts, indicating that multi-listing is common in london.
2. Professional landlords: hosts who owned 10+ listings owned more than one fifth listings, suggesting a significant presence of professional landlords in the market.

6.1.2 Changes in the number of landlords and renters over the years



- Based on Airbnb's dataset for London, the whole story started in 2008 and the growth in hosts and listings was slow between 2008 and 2010, **accelerating sharply from 2011 to 2016**. The peak occurred in 2015, with 8,939 new hosts, while 2014 and 2016 saw increases of over 7,000 hosts each. By 2016, total listings neared 50,000.
- However, **growth slowed in subsequent years**, with 2020 and 2021 adding only around 1,600 hosts and 3,000 listings annually—nearly half the growth seen in 2019—largely due to the pandemic's impact on the rental market. From 2022 to 2024, post-pandemic recovery is evident, but growth remains far below peak levels.

6.2 Analysis of property

6.2.1 Distribution of room types of property

Room type of property is divided into four categories.

- Entire home/apt: 63.8%
- Private room: 35.6%
- Shared room: 0.45%
- Hotel room: 0.2%

Conclusion:

1. The high proportion of entire homes/apt indicates that many guests prefer independent accommodations for greater privacy and autonomy.
   This aligns with a broader shift in tourism, where more visitors are opting for alternative lodging options instead of traditional hotels to enjoy a more spacious and private environment (Zervas, Proserpio and Byers, 2017).
2. The 35.6% share of private rooms suggests that some guests are still willing to choose more affordable accommodations, even if it means sharing common spaces. These listings cater to budget-conscious travelers.

3. The low percentages of shared rooms and hotel rooms indicate that Airbnb's core market in London, a well-established market, tends to favor more private lodging options.

6.2.2 Distribution of Minimum Nights for renting property

Based on the dataset, 93550 listings have a minimum night stay of less than the STR threshold (30 days), making up 97.3% of the total. Additionally, listings with a minimum stay of less than 7 days account for 92.3% of the total. **The London rental market on airbnb is dominated by short-term rentals.**

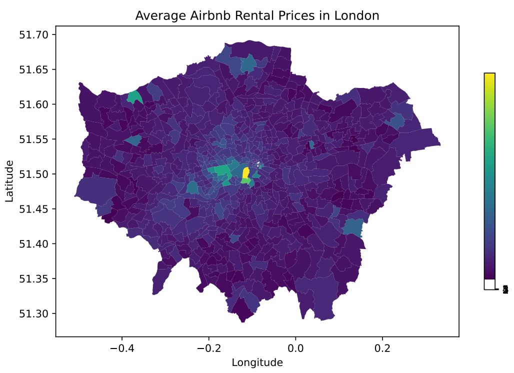
Chaudhary had illustrated some drawbacks of short term renting (Chaudhary, 2021).

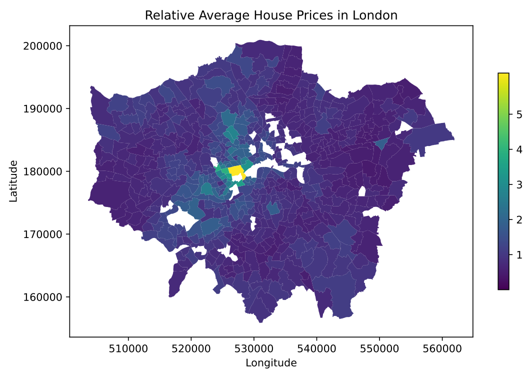
1. **Reduced long-term housing supply**: Due to higher profits from short-term rentals (e.g., Airbnb), many landlords prioritize short-term leases over long-term rentals, exacerbating London's housing crisis and driving up rents, especially for low- and middle-income residents.
2. **Community impacts**: A high volume of short-term rentals can disrupt neighborhoods, increasing noise and tourist traffic, making communities less appealing for long-term residents and undermining stability and safety.

7. Drawing on your previous answers, and supporting your response with evidence (*e.g.* figures, maps, EDA/ESDA, and simple statistical analysis/models drawing on experience from, e.g., CASA0007), how *could* the InsideAirbnb data set be used to inform the regulation of Short-Term Lets (STL) in London?

7.1 The Correlation between Airbnb Rental Prices and Housing Prices

This part aims to explore the **spatial correlation** between **Airbnb rental prices** and **housing prices** across various wards in London. Wards are considered as the smallest unit of analysis for this research. Initially, K-means clustering is employed to categorize properties based on their rental prices. Subsequently, average Airbnb rental prices and average housing prices for each ward are calculated.





By comparing these two metrics visually on a map, it is observed that the area with **the highest Airbnb rental prices is Bishop's**, which paradoxically reflects a **relatively low average housing price**.

Conversely, Knightsbridge and Belgravia, located in proximity to the city center, exhibit the highest average housing prices, with a **noticeable decline** as one moves outward from the central area.

Importantly, the districts that report the highest Airbnb rental prices do not coincide with those that have the highest housing prices.

Nevertheless, **both** metrics are significantly concentrated around the ward of Knightsbridge and Belgravia. Furthermore, some suburban wards demonstrate relatively high Airbnb rental prices; however, the housing prices in these areas remain comparable to those of their neighboring regions, suggesting limited impact.

7.2 Multiple linear regression

In order to more intuitively prove the impact of Airbnb on the local community and explore the extent of the impact, we used the method of constructing a multiple linear regression model, where we calculated the median number of **housing price**, **population density** and **house sales** of each ward, and took them as the **dependent variables**. We calculated the median number of **Airbnb price**, **monthly number of reviews**, **annual availablity**, **review value**, and **aribnb count** as **independent variables**.

	Variable	VIF
0	Intercept	1650.880650
1	Airbnb_price	1.600156
2	Airbnb_availability_365	1.155116
3	Reviews_per_month	1.089707
4	Review_scores_value	1.120966
5	Airbnb_count	1.757443

After calculating the VIF of the independent variables, we find that there are no variables that exceed the threshold, so there may be no obvious multicollinearity, and the model results are as follows:

```
                                 OLS Regression Results                                  
=========================================================================================
Dep. Variable:     Population_per_square_kilometre   R-squared:                       0.324
Model:                                         OLS   Adj. R-squared:                  0.318
Method:                              Least Squares   F-statistic:                     54.13
Date:                             Tue, 17 Dec 2024   Prob (F-statistic):           6.68e-46
Time:                                     12:32:25   Log-Likelihood:                -5573.4
No. Observations:                              570   AIC:                         1.116e+04
Df Residuals:                                  564   BIC:                         1.118e+04
Df Model:                                        5                                         
Covariance Type:                         nonrobust                                         
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                  4.132e+04   7303.028      5.658      0.000     2.7e+04    5.57e+04
Airbnb_price             -4.3365      5.857     -0.740      0.459     -15.841       7.168
Airbnb_availability_365 -24.2943      3.169     -7.667      0.000     -30.518     -18.071
Reviews_per_month      1560.8176    584.249      2.671      0.008     413.248    2708.388
Review_scores_value   -6301.3219   1475.036     -4.272      0.000   -9198.556   -3404.088
Airbnb_count             20.4364      2.575      7.937      0.000      15.379      25.494
==============================================================================
Omnibus:                       35.280   Durbin-Watson:                   1.495
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               85.630
Skew:                           0.307   Prob(JB):                     2.55e-19
Kurtosis:                       4.796   Cond. No.                     9.92e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.92e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:      Houseprice_median   R-squared:                       0.532
Model:                            OLS   Adj. R-squared:                  0.528
Method:                 Least Squares   F-statistic:                     128.3
Date:                Tue, 17 Dec 2024   Prob (F-statistic):           1.34e-90
Time:                        12:32:26   Log-Likelihood:                -7616.0
No. Observations:                 570   AIC:                         1.524e+04
Df Residuals:                     564   BIC:                         1.527e+04
Df Model:                           5                                         
Covariance Type:            nonrobust                                         
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                 -1.108e+05   2.63e+05     -0.422      0.674   -6.27e+05    4.06e+05
Airbnb_price           3198.5010    210.820     15.172      0.000    2784.413    3612.589
Airbnb_availability_365 -229.8990   114.050     -2.016      0.044    -453.913      -5.885
Reviews_per_month     -1.968e+04    2.1e+04     -0.936      0.350    -6.1e+04    2.16e+04
Review_scores_value    7.096e+04   5.31e+04      1.337      0.182   -3.33e+04    1.75e+05
Airbnb_count            574.4705     92.678      6.199      0.000     392.435     756.506
==============================================================================
Omnibus:                      349.748   Durbin-Watson:                   1.688
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8767.965
Skew:                           2.228   Prob(JB):                         0.00
Kurtosis:                      21.690   Cond. No.                     9.92e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.92e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Housesales_median   R-squared:                       0.078
Model:                            OLS   Adj. R-squared:                  0.070
Method:                 Least Squares   F-statistic:                     9.530
Date:                Tue, 17 Dec 2024   Prob (F-statistic):           9.68e-09
Time:                        12:32:26   Log-Likelihood:                -3344.1
No. Observations:                 570   AIC:                             6700.
Df Residuals:                     564   BIC:                             6726.
Df Model:                           5
Covariance Type:            nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                    -286.7489    146.180     -1.962      0.050    -573.872       0.374
Airbnb_price                0.1353      0.117      1.154      0.249      -0.095       0.366
Airbnb_availability_365     0.0137      0.063      0.216      0.829      -0.111       0.138
Reviews_per_month         -31.8929     11.695     -2.727      0.007     -54.863      -8.923
Review_scores_value        89.6425     29.525      3.036      0.003      31.651     147.634
Airbnb_count                0.2284      0.052      4.432      0.000       0.127       0.330
==============================================================================
Omnibus:                      575.162   Durbin-Watson:                   1.609
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            34626.521
Skew:                           4.415   Prob(JB):                         0.00
Kurtosis:                      40.148   Cond. No.                     9.92e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 9.92e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The model fits well, including: - The Houseprice_median model performed best, explaining 53.2% of the fluctuations - Population_per_square_kilometre model was second, explaining 32.4% of the fluctuations. - The Housesales_median model performs the worst, explaining only 7.8%.

- Airbnb_count is significant in all three models and the effect is positive.
- Reviews_per_month and Review_scores_value are significant in some models, but in different directions.
- Airbnb_price is only significant in the Houseprice_median model.

7.3 The Impact of Airbnb on London Neighborhoods: A K-Means Geodemographic Classification

7.3.1 Data processing

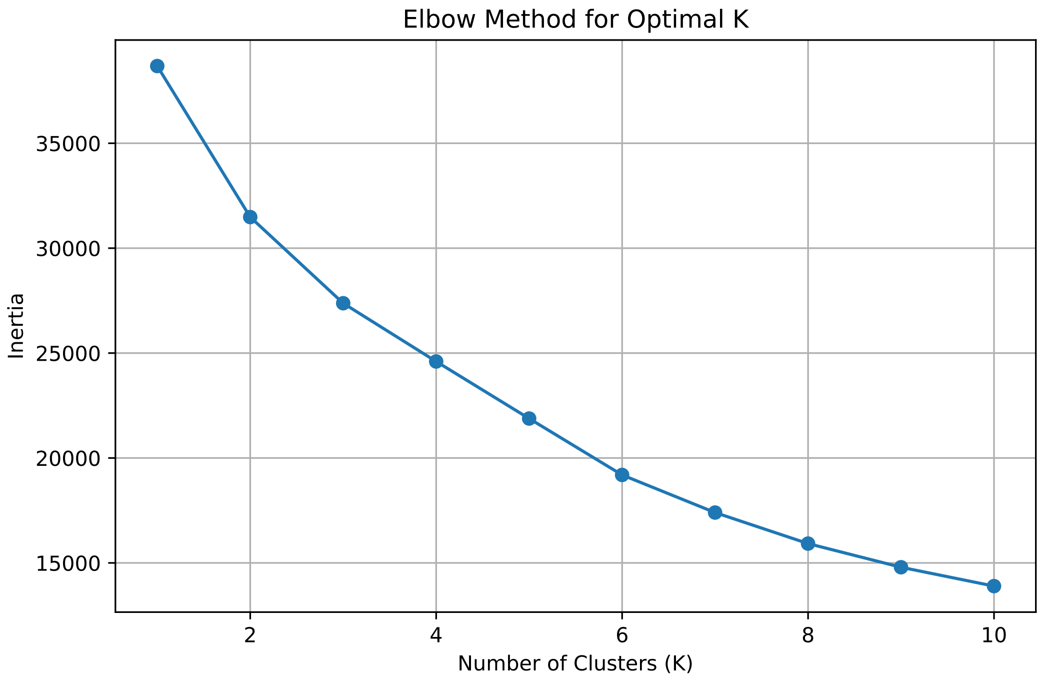
key variables - Airbnb Average Price per Night: Reflects the economics of the short-term rental market. - Airbnb Density: shows how well Airbnb's are distributed in a given neighborhood. - Changes in Social Listing Prices Over a Five-Year Period: Indirectly reflects Airbnb's potential impact on the local real estate market. - Population Density: Indicates the size and density of a community's population. - Income Deprivation Index: Higher values indicate greater income deprivation, which may be - correlate with the economic status and social vulnerability of the community - Hotel Density: measures the competitive environment for traditional lodging establishments. - Attraction Density: Reflects the community's tourist appeal and potential demand for Airbnb's. - Public Transportation Accessibility: Demonstrates the community's accessibility to tourists.

Variables that need to be processed :

1. Calculate the density of tourist attractions in each LSOA in London
2. Calculate hotel density for each LSOA unit
3. Calculate five-year house price changes for each LSOA unit
4. Read in and process Airbnb data
5. Calculate Airbnb Density and Average Price per Night for LSOA Units
6. Merge all data
7. Check and clean the data

7.3.2 Standardization (Z-score scaling)

7.3.3 Elbow Method to caculate the k value



- Based on the information provided in the graphs, the optimal number of clusters (K) appears to be 3.

7.3.4 K-means Clustering

Results of K-means clustering

1. Cluster 0: “Low Impact — Stable Residential Areas”

- Lowest Airbnb Density (-0.42) and Hotel Density (-0.13) indicate minimal tourism and Airbnb presence.
- Highest Price Change (19.39) reflects significant housing market growth, possibly driven by local development or gentrification.
- Low Population Density (6,688 people/km^2) and Poor Public Transport Access (7.98) make these areas less attractive to short-term rentals.
- Moderate Income Deprivation (0.11) suggests economic vulnerability.

2. Cluster 1: “Moderate Impact — Balanced Zones”

- Moderate Airbnb Density (0.44) and Hotel Density (-0.06) indicate balanced tourism and residential dynamics.
- Moderate Price Change (10.75) and Population Density (16,761 people/km^2) suggest a stable housing market with mixed use.
- Moderate Public Transport Access (19.40) supports tourism without overwhelming local infrastructure.
- Higher Income Deprivation (0.19) reflects economic challenges despite balanced Airbnb activity.

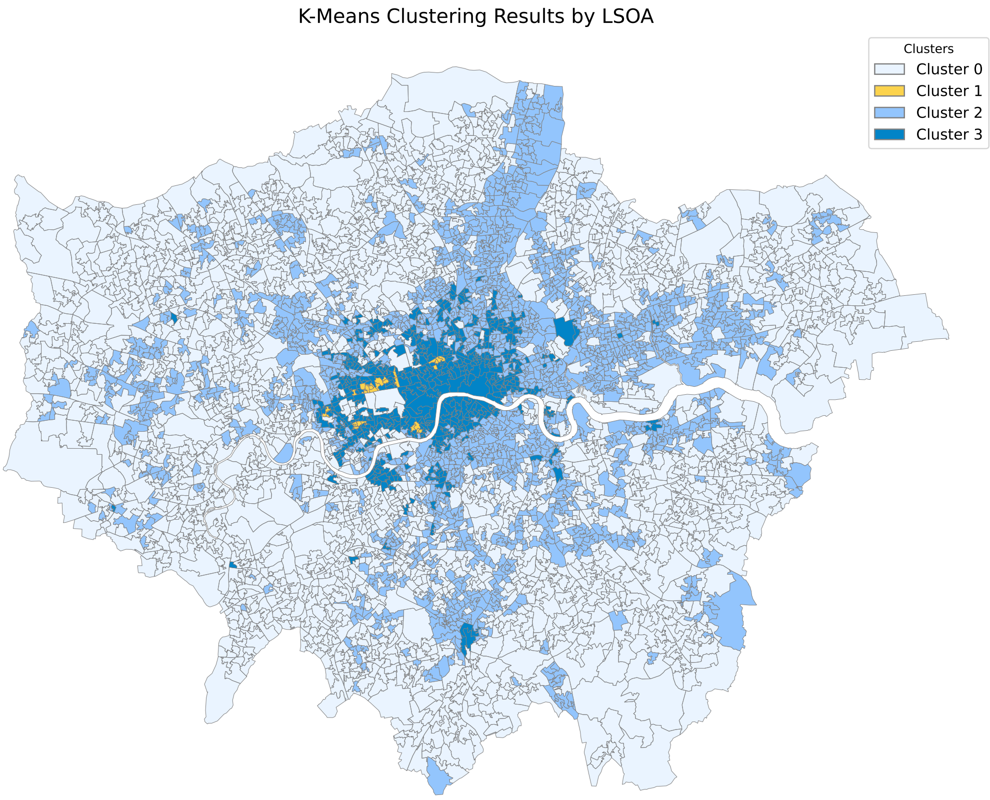
3. Cluster 2: “Emerging Impact — High Price Pressure Areas”

- Unusually High Average Price (18.90) indicates significant economic pressures on the housing market.
- Moderate Airbnb Density (0.48) and Tourist Attraction Density (0.19) reflect emerging tourism activity.
- High Population Density (17,579 people/km^2) combined with Good Public Transport Access (30.45) supports growth in short-term rentals.
- Moderate Income Deprivation (0.16) suggests economic tension within these areas.

4. Cluster 3:“High Impact — Tourist Hotspots”

- Highest Airbnb Density (3.51) and Hotel Density (3.11) mark these areas as major tourism accommodation hubs.
- Highest Average Price (0.54) suggests strong profitability for short-term rentals.
- High Tourist Attraction Density (1.63) and Population Density (20,976 people/km^2) highlight their appeal to visitors.
- Convenient Public Transport Access (53.91) facilitates heavy visitor traffic.

7.3.5 Plot K-Means Clustering Result



- *From the map:*
- Cluster 0 dominates the outer boroughs of London and peripheral areas.
- Cluster 1, which is Small, isolated neighborhoods near central areas showing signs of Airbnb growth.
- Cluster 2 is widely distributed across inner London and intermediate zones surrounding central areas.It forms a ring around the core city center, particularly dense in the inner boroughs like Hackney, Islington, and Camden.
- Cluster 3, is concentrated in central London, including areas like Westminster, the City of London, and parts of Southwark.

7.3.6 Conclusion on Airbnb Impact and Policy Recommendations:

1. Low Impact — Stable Residential Areas:

- Airbnb Impact:These neighborhoods are largely residential with minimal tourism influence but are at risk of future Airbnb expansion due to rising housing prices.
- Policy Recommendations: Proactively monitor Airbnb growth. Implement measures to prevent housing market disruption, such as rental density caps.

2. Emerging Airbnb Areas — "Expansion Zones":

- Airbnb Impact: These neighborhoods have a moderate Airbnb presence that coexists with local housing needs.
- Policy Recommendations: Encourage sustainable tourism while safeguarding affordable housing. Adopt policies that prevent excessive short-term rental concentration.

3. High Impact — Tourist Hotspots:

- Airbnb Impact: These areas are already heavily impacted by Airbnb activity, creating pressure on housing availability and affordability.
- Policy Recommendations: Introduce stricter regulations, such as rental caps, zoning restrictions, and taxation on short-term rentals to protect long-term residents.

4. Emerging Impact — High Price Pressure Areas:

- Airbnb Impact: These neighborhoods are emerging hotspots, where Airbnb activity and housing pressures are growing.
- Policy Recommendations: Implement early intervention strategies to balance tourism growth and housing stability. Monitor housing price trends closely.

Sustainable Authorship Tools

Using the Terminal in Docker, you compile the Quarto report using `quarto render <group_submission_file>.qmd`.

Your QMD file should automatically download your BibTeX and CLS files and any other required files. If this is done right after library loading then the entire report should output successfully.

Written in Markdown and generated from Quarto. Fonts used: Spectral (mainfont), Roboto (sansfont) and JetBrains Mono (`monofont`).

References

Brenning, A. and Henn, S. (2023) 'Web scraping: A promising tool for geographic data acquisition', *arXiv preprint arXiv:2305.19893*.

Chaudhary, A. (2021) 'Effects of airbnb on the housing market: Evidence from london.', *Available at SSRN 3945571*.

Krotov, V. and Johnson, L. (2023) 'Big data: Challenges related to data, technology, legality, and ethics', *Business Horizons*, 66(4), pp. 481–491.

Pawlicz, A. and Prentice, C. (2021) 'UNDERSTANDING SHORT-TERM RENTAL DATA SOURCES – a VARIETY OF SECOND-BEST SOLUTIONS', *Tourism in Southern and Eastern Europe*. Available at: https://api.semanticscholar.org/CorpusID:246571127.

Prentice, C. and Pawlicz, A. (2023) 'Addressing data quality in airbnb research', *International Journal of Contemporary Hospitality Management*. Available at: https://api.semanticscholar.org/CorpusID:258644931.

Sobel, B. L. (2021) 'The new common law of web scraping', *Lewis & Clark L. Rev.*, 25, p. 147.

Xie, W. and Karan, K. (2019) 'Consumers' privacy concerns and privacy protection on social networking sites in the era of big data: Empirical evidence from college students', *Journal of Interactive Advertising*, 19(3), pp. 187–201.

Zervas, G., Proserpio, D. and Byers, J. W. (2017) 'The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry', *Journal of marketing research*, 54(5), pp. 687–705.