505 not found'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

Initial Research scope

1. Research Topic:

Exploring the Impact Profile of London's Neighbourhoods

2. Methodology:

1. Correlation between Airbnb Rental Prices and Housing Prices
2. Multiple Linear Regression Analysis
3. K-Means Geodemographic Classification

3. Research objectives:

1. Investigate the Impact of Airbnb on Housing Prices and Availability:
2. Identify Affected Areas and Community Profiles
3. Summarize Policy Recommendations

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?

InsideAirbnb's data collection method faces quality issues due to incomplete data and dependency on Airbnb's website structure, risking outdated or inaccurate information from dynamic web content (Krotov and Johnson, 2023). Privacy measures may also skew geographic accuracy, and changes in Airbnb's website structure or anti-scraping measures like CAPTCHA and IP bans further complicate collection, necessitating ongoing script maintenance (Prentice and Pawlicz, 2023).

The InsideAirbnb dataset comprehensively captures the Airbnb market, including listing distribution, characteristics, pricing models, and their impact on local housing, supporting detailed analyses.

However, the use of InsideAirbnb data introduces technical, legal, and ethical challenges. Legally, the variability in scraping laws across jurisdictions could render some data collection illegal, especially if it violates Airbnb's terms of service (Sobel, 2021). Ethically, scraping without data subjects' consent (Airbnb hosts and guests) raises privacy concerns, influenced by their awareness and data management practices (Xie and Karan, 2019).

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

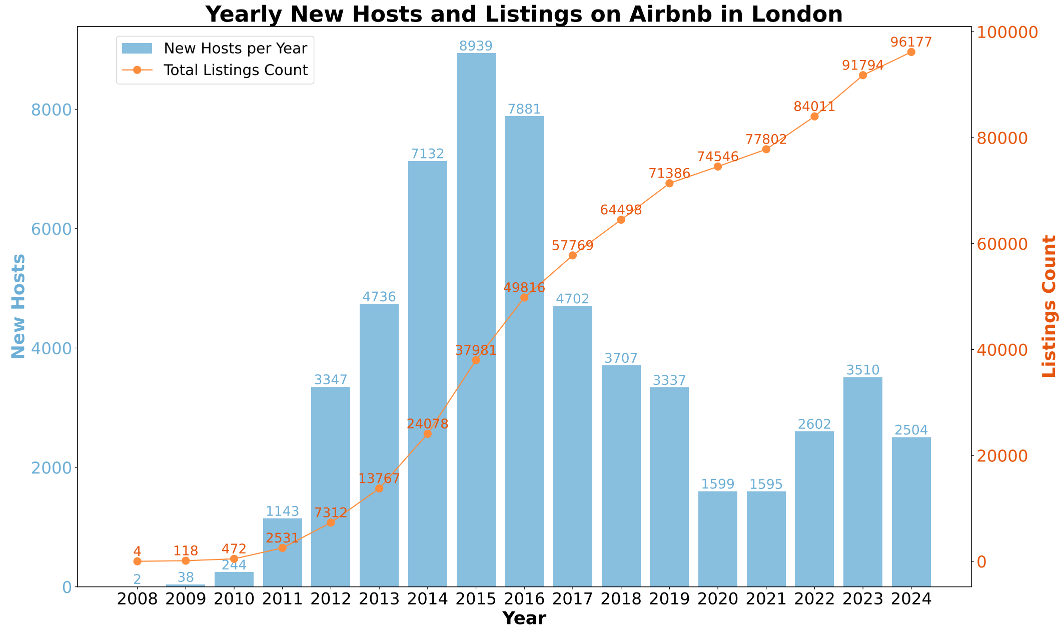
6.1.1 Distribution of the Number of Listings per Host

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



- Airbnb's growth in London started slowly between 2008 and 2010, then **accelerated sharply from 2011 to 2016**. The peak was in 2015 with 8,939 new hosts, and both 2014 and 2016 saw over 7,000 new hosts. By 2016, total listings reached nearly 50,000.
- However, **growth slowed in the following years**, with 2020 and 2021 adding only around 1,600 hosts and 3,000 listings annually, partly due to the pandemic. From 2022 to 2024, post-pandemic recovery occurred, but growth remained well 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.

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.**

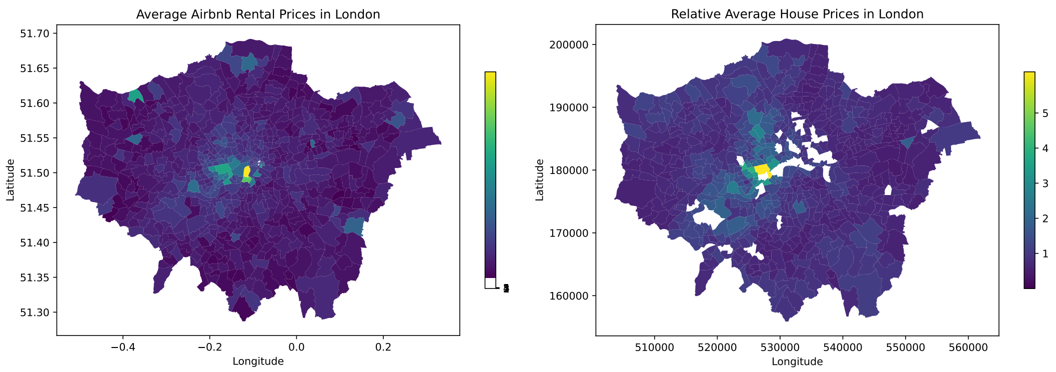
Chaudhary had illustrated some drawbacks of short term renting (Chaudhary, 2021).

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:

```
Regression results for Population_per_square_kilometre:
```

	Coefficient	Standard Error	t-value	P-value	R-squared
const	41322.899	7303.028	5.658	0.000	0.324
Airbnb_price	-4.337	5.857	-0.740	0.459	0.324
Airbnb_availability_365	-24.294	3.169	-7.667	0.000	0.324
Reviews_per_month	1560.818	584.249	2.671	0.008	0.324
Review_scores_value	-6301.322	1475.036	-4.272	0.000	0.324
Airbnb_count	20.436	2.575	7.937	0.000	0.324

```
Regression results for Houseprice_median:
```

	Coefficient	Standard Error	t-value	P-value	R-squared
const	-110809.008	262870.114	-0.422	0.674	0.532
Airbnb_price	3198.501	210.820	15.172	0.000	0.532
Airbnb_availability_365	-229.899	114.050	-2.016	0.044	0.532
Reviews_per_month	-19683.477	21029.858	-0.936	0.350	0.532
Review_scores_value	70961.269	53093.433	1.337	0.182	0.532
Airbnb_count	574.471	92.678	6.199	0.000	0.532

```
Regression results for Housesales_median:
```

	Coefficient	Standard Error	t-value	P-value	R-squared
const	-286.749	146.180	-1.962	0.050	0.078
Airbnb_price	0.135	0.117	1.154	0.249	0.078
Airbnb_availability_365	0.014	0.063	0.216	0.829	0.078
Reviews_per_month	-31.893	11.695	-2.727	0.007	0.078
Review_scores_value	89.643	29.525	3.036	0.003	0.078
Airbnb_count	0.228	0.052	4.432	0.000	0.078

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: K-Means Geodemographic Classification

7.3.1 Data processing

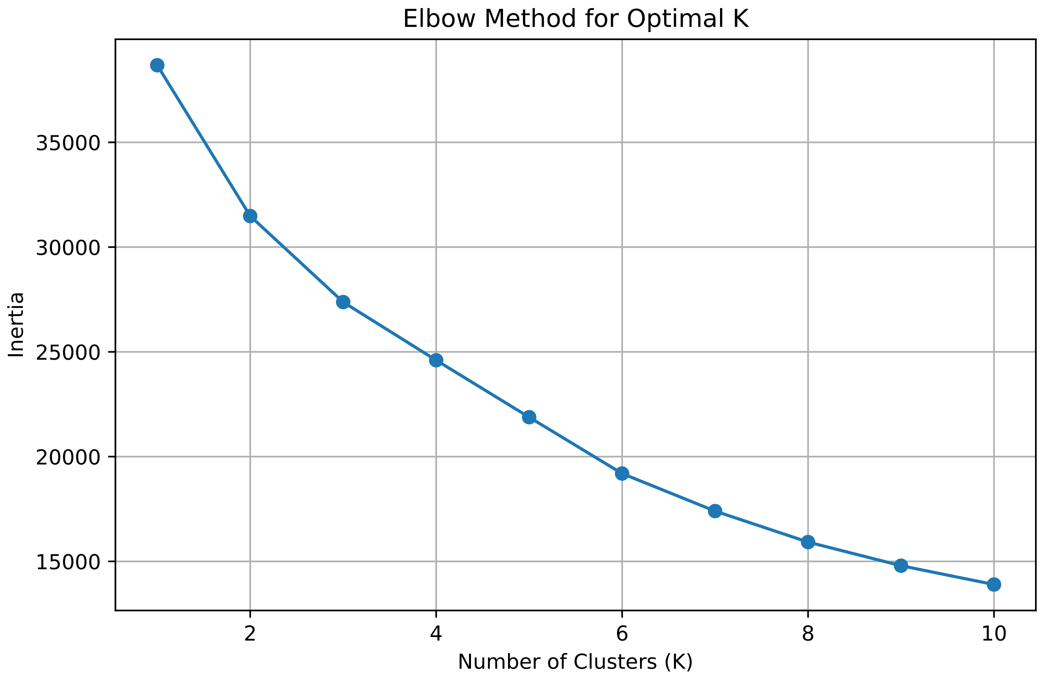
key variables:
- average_price: Average nightly price of Airbnb
- airbnb_density
- Price_Change: Five-year housing price change in the community
- People per Sq Km: Population density
- Income Score (rate): Income deprivation score
- hotel_density
- tourist_attraction_density
- AvPTAI2015: Public Transport Accessibility Index

Variables that need to be processed :

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

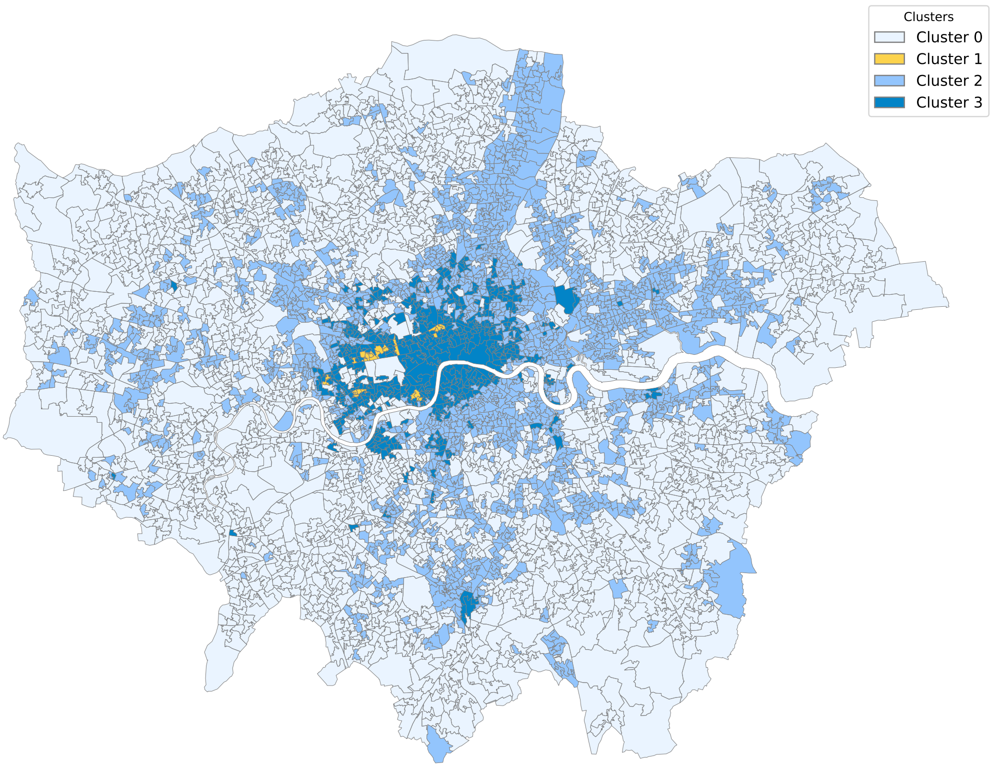
7.3.4 K-means Clustering

Results of K-means clustering

1. Cluster 0: "Low-Impact Peripheral Areas" Low Airbnb activity, poor transport access, rising housing prices, minimal tourism.
2. Cluster 1: "Heavily Impacted Tourist Hubs": Extremely high Airbnb density, central locations, high hotel and tourist attraction density, excellent transport access.
3. Cluster 2: "At-Risk Transition Zones": Moderate Airbnb activity, rising housing prices, income deprivation, low tourism and hotel presence.
4. Cluster 3: "Emerging Impact Zones": Growing Airbnb density, rising housing prices, near-central areas, moderate tourism and transport access.

7.3.5 Plot K-Means Clustering Result

K-Means Clustering Results by LSOA



- *From the map:*
- Cluster 1 - Concentrated in the central core of London, these are the high-impact tourist hubs. Such as, Southwark, Tower Hamlets and Camden.
- Cluster 2 - Located between Cluster 1 and Cluster 0, these are the transitional at-risk zones.
- Cluster 3 - Situated near the center, like Westminster, Kensington. These are the emerging impact zones on the periphery of the core.
- Cluster 0 - Found in the outer peripheral areas, these are the low-impact neighborhoods.

7.3.6 Conclusion on Airbnb Impact and Policy Recommendations:

1. Heavily Impacted Central Tourist Hubs:

- Airbnb Impact: Severe housing pressures, tourism-driven displacement, and rising rents. Airbnb dominates short-term rentals, reducing long-term housing availability.
- Policy Recommendations: Immediate regulation to control Airbnb density. Implement rental caps and enforce zoning restrictions. Protect affordable housing for local residents.

2. At-Risk Transition Zones:

- Airbnb Impact:Emerging Airbnb activity increases risks of gentrification, especially in deprived areas. Communities face housing affordability issues as prices rise.
- Policy Recommendations: Introduce early interventions to stabilize housing costs. Encourage long-term rentals over short-term stays.

3. Emerging Impact Zones:

- Airbnb Impact: Rising Airbnb density puts growing pressure on housing availability and affordability. Areas are on the path to becoming heavily impacted.
- Policy Recommendations: Balance tourism opportunities with protecting housing for locals.

4. Emerging Impact — High Price Pressure Areas:

- Airbnb Impact: Airbnb activity is very low, with little effect on housing or rents. Poor public transport makes these areas less attractive to visitors.
- Policy Recommendations: Focus on improving transport and infrastructure to support balanced development.

8. Conclusion and Reflection

This study explores Airbnb's impact on London's neighbourhoods.The analysis indicates that high Airbnb rental prices do not always correspond with high housing prices, with areas like Bishop's showing higher Airbnb rents despite lower housing costs. Multiple linear regression models reveal that Airbnb listings, reviews, and availability significantly influence housing prices and population density. K-means clustering categorizes areas into tourist hubs, at-risk zones, emerging areas, and low-impact zones. The study recommends regulating Airbnb density in high-impact areas, protecting affordable housing, and balancing tourism with long-term housing needs. Moreover, challenges related to accuracy, legality, and ethics must be acknowledged. Future research should combine more comprehensive data to enhance the result.

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.