CIS5450 Term Project Proposal

**Date:** 2 November 2023

**Team:** Claudia Peinado, Chukwudire Uba, and Jeff Robson

**Draft Title:** Evaluation of the relationship between housing prices and short-term rental sector activity in London, UK

# 1. Introduction

Short-term rental (STR) services are a major driver of urban transformation. Since its founding in 2008, Airbnb – as a model STR – has expanded to over 4.5 million listings across the world, and has yielded over 300 million unique stays and USD $41 billion in revenue to hosts.[[1]](#footnote-0)

In the United Kingdom, Greater London has experienced significant growth pressures exacerbated by the growth of the STR sector. A 2020 report on the influence of Airbnb on London’s housing market revealed that over 80,000 properties were listed on the service, with 23 percent thought to be in violation of the local authority’s 90-day rental limit for STRs.[[2]](#footnote-1) London’s traditional predominance within the United Kingdom’s labor market, coupled with the growth of STRs and the ongoing financialization of the region’s housing market have strained the ability of Londoners to afford safe and stable housing within their means.[[3]](#footnote-2)

Using regional housing purchase price data, STR listing data, and supplemental indicator data for measures of deprivation and taxation, we will apply data cleaning, exploratory data analysis (EDA), linear and logistic regression modeling techniques to build a predictive model for median housing prices by borough given local STR sector activity. Given the body of research on the relationship between housing affordability and STRs, we reason that an active STR sector within a borough is positively correlated with increased housing costs.[[4]](#footnote-3)

**Note:** A borough in London is one of 32 local authority districts governed by an individual council that, with the City of London, form the Greater London area.

# 2. Methodology

## Team responsibilities

The study will be conducted by three MCIT students: Claudia Peinado, Chukwudire Uba, and Jeff Robson. If possible, we would like to be assigned to Zijian Zhang as the project’s teaching assistant. Team roles and responsibilities are as follows:

* Claudia Peinado: Data cleaning, EDA, and logistic model development.
* Chukwudire Uba: Data cleaning, EDA, and model development.
* Jeff Robson: Data acquisition, data cleaning, model development review, and reporting.

### Description of responsibilities

**Data acquisition:** Data will be collected from identified sources in the following section.

**Data cleaning:** Data from disparate sources will be cleaned, standardized, and synthesized for use within the EDA and model development stages. Data will then be uploaded to a database.

**Exploratory data analysis:** A series of data exploration tasks will be performed on the cleaned cumulative data set. Where appropriate, EDA tasks will be visualized according to data visualization best practices highlighted in lecture. Data exploration and feature development may include:

* From individual purchase prices: mean, median purchase price by borough
* Individual STR listings: mean, median STR cost per night by borough, number of STR by borough
* Measures of deprivation by sub-borough level: Mean index of deprivation, deprivation indicators for crime, barriers to housing and services, living environment (subset of indices)
* Individual Council tax bands by borough: Median council tax band by borough

**Model development:** Per the following section “Model development approach”, two standard linear regression models and one logistic regression model will be developed and compared to determine an optimal modeling approach for our given research problem.

**Reporting:** Each team member will conduct regular notebook development as part of the commission of the above tasks; additional contextual reporting and writing will be provided in an online blog format.

## 

## Data sources and description

**Assumptions:** Project scope is narrowed from the Greater London area (32 boroughs) to the inner London boroughs (13) of Camden, Greenwich, Hackney, Hammersmith and Fulham, Islington, Kensington and Chelsea, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Wandsworth, Westminster, and the City of London. This will focus the project with an appropriately scaled scope and minimize STR sparsity.

Project scope is narrowed to 2019 data to mitigate distorting effects of the COVID-19 pandemic on both the short-term rental sector and the housing market, and to align with the most recent available UK government data on indices of deprivation.

The proposed primary data sources are as follows:

**United Kingdom property price data, 2019 (England and Wales)**. [Data](https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads) is collected and maintained by HM Land Registry, UK Government and available through a Creative Commons license. The dataset exceeds 50,000 rows for the year 2019 and contains features for sale price, transfer date, postcode, property type, sale duration, street, locality (borough), town/city as primary points of interest.

**Airbnb listings data, March 2020.** Data is collected and maintained by data-driven advocacy group Inside Airbnb and is available under Creative Commons A4: Attribution. The dataset exceeds 50,000 rows for the period 2019-2020 and contains features for borough and geolocation, property type, room type, number beds and baths, amenities, price per night, minimum and maximum nights, long-term availabilities (30, 60, 90, 365 days), unique host identifier, account initialization date, licencing, and review instances.

**Supplemental data sources, 2019.** Additional small-scale data sources for English Measures of Deprivation (UK Government, 2019) and local borough taxation bands (Greater London Authority, 2019) will be consulted to provide additional context for the study.

## 

## Data modeling approach

To satisfy our hypothesis, we will conduct linear and logistic regression modeling to determine the direct relationship between STR frequency and housing purchase costs by borough. We will explore three regression model types (vanilla, lasso, ridge), with an assessment of each to determine the model with best predictive fit for the relationship between median housing purchase prices and STR sector activity.

We denote the independent variable as the number, or frequency, of STR listings and the dependent variable as purchase price for housing (with the assumption that purchase price shapes future rental takings). We will also explore the correlation between purchase price and STR frequency with borough-level measures of deprivation (a subset of measures will include indicators for crime, barriers to housing and services, and quality of living environment) and borough council taxation rates to explain variance in the overall relationship.

We intend to use either k-Fold Cross-Validation or Bootstrap techniques to evaluate our models and the mean squared error (MSE) as the ultimate performance metric.

# 3. Conclusion

## Challenges

Several prospective challenges may be encountered during the completion of the term project. These include, but are not limited to:

* **Scale ambition to time.** Given the inherent complexity of using data to model a housing market, adjustments may need to be made in the approach and ambition to keep the project in scope and scale to its submission deadline.
* **Aligning disparate datasets.** Two primary and two supplemental datasets are noted as the basis for the project. Work will need to be undertaken to read, parse, and prepare a cleaned dataset that joins together disparate data by borough name as the join key.
* **Scale of the work.** To this point in the course, each team member has worked only on individually developed autograded and generally linear Google Colab notebooks. This term project will require close collaboration, self-directed research using open source data without the “sanity checks” offered by the autograder. This will require a different perspective on work and project management to ensure the project proceeds to schedule.
* **Balancing flexibility and rigidity.** While this proposal offers a potential way forward for the term project and is sufficiently detailed to begin preliminary project tasks, we cannot know everything without doing the work and so will need to maintain a balance of flexibility in the face of emergent challenges while adhering to the general terms established in this document.

## Anticipated product, work-flow and deadlines

The team will provide as an anticipated product the required and complete notebook(s) as **Deliverable 1** and will complete as **Deliverable 2** a blog-based narrative description of the term project, including but not limited to: research background and rationale, methodology, outcomes, and potential future directions for study.

The following deadlines are required in satisfaction of the term project:

1. **Proposal Check-in:** Nov 2, 2023, 11:00 am ET (US)
   1. *Team members will have read the proposal thoroughly, added their thoughts and come to a consensus. Individual Model selection must also take place at this time.*
   2. Proposal Due: Monday, November 6, 2023, 11:59 pm Eastern (US) - *Jeff submits*
2. **Intermediate Check-in:**  Nov 16, 2023, 11:00 am ET (US)
   1. *All Data Cleaning is complete at this time so members can move on to complete their model section.*
3. **Model-Progress Check-in:** Nov 30, 2023, 11:00 am ET (US)
   1. *All individual Models are due at this time. As a team we will decide which model is most effective based on MSE during this meeting.*
4. **Final Check-in:** Dec 7, 2023, 11:00 am ET (US)
   1. *Team deadline for finishing the complete notebook and blog sections. Team will use this time to make final adjustments.*
   2. NOTE: This is the weekend of the Penn final period; the assignment is functionally due the preceding week for this reason
5. **Final Deliverable** (Options): Dec 9, 2023

As the final deliverable submission dates are scheduled for **Saturday, December 9, 2023**, concurrent with the MCIT final exam period, the team will endeavor to complete the project ahead of the fixed submission date.

4. References

Deboosere, R., Kerrigan, D.J., Wachsmuth, D., and El-Geneidy, A. (2019). “Location, location and professionalization: A multilevel hedonic analysis of Airbnb listing prices and revenue”. *Regional Studies, Regional Science* 6.1: 143-156.

Greater London Authority Housing and Land (2020). “Housing Research Note 2020/04: Short-term and holiday letting in London,” Policy report, authored by Cosh, G. Available at: <https://www.london.gov.uk/sites/default/files/housing_research_note_4-_short-term_and_holiday_letting_in_london.pdf> (accessed October 30, 2023).

Mayor of London. (May 2018). “London Housing Strategy”. Available at: <https://www.london.gov.uk/sites/default/files/2018_lhs_london_housing_strategy.pdf>   
(accessed October 30, 2023).

Shabrina, Z., Arcaute, E., and Batty, M. (2021). “Airbnb and its potential impact on the London housing market.” *Urban Studies* 1-25. Available at: [https://discovery.ucl.ac.uk/id/eprint/  
10120512/1/0042098020970865.pdf](https://discovery.ucl.ac.uk/id/eprint/10120512/1/0042098020970865.pdf) (accessed October 30, 2023).

5. Data Repositories

Inside Airbnb Data Repository: <http://insideairbnb.com/get-the-data/>

Ministry of Housing, Communities, and Local Government (UK), “Council Tax Charges - Bands, Borough”, <https://data.london.gov.uk/dataset/council-tax-charges-bands-borough>.

Ministry of Housing, Communities and Local Government (UK), “Indices of deprivation”, <https://data.london.gov.uk/dataset/indices-of-deprivation>

UK Government, “Price Paid Data”, <https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads>

1. Deboosere, R., Kerrigan, D.J., Wachsmuth, D., and El-Geneidy, A., “Location, location and professionalization: A multilevel hedonic analysis of Airbnb listing prices and revenue”. *Regional Studies, Regional Science* 6.1 (2019): 143-156. [↑](#footnote-ref-0)
2. Greater London Authority – Housing and Land, “Housing Research Note 2020/04: Short-term and holiday letting in London,” Policy report, authored by G. Cosh (2020). Available at: <https://www.london.gov.uk/sites/default/files/housing_research_note_4-_short-term_and_holiday_letting_in_london.pdf> (accessed October 30, 2023). [↑](#footnote-ref-1)
3. Mayor of London, “London Housing Strategy” (May 2018), available at: <https://www.london.gov.uk/sites/default/files/2018_lhs_london_housing_strategy.pdf> (accessed October 30, 2023). [↑](#footnote-ref-2)
4. See for example: Shabrina, Z., Arcaute, E., and Batty, M.. “Airbnb and its potential impact on the London housing market.” *Urban Studies*, 1-25 (2021). Available at: <https://discovery.ucl.ac.uk/id/eprint/10120512/1/0042098020970865.pdf> (accessed October 30, 2023). [↑](#footnote-ref-3)