Relationship between housing affordability and short-term rentals

**Data links:**

<https://www.kaggle.com/datasets/willianoliveiragibin/uk-property-price-data-1995-2023-04/data>

<https://landregistry.data.gov.uk/app/ppd/>

<http://insideairbnb.com/explore>

<http://insideairbnb.com/london>

<https://data.london.gov.uk/dataset/local-authority-average-rents>

<https://data.london.gov.uk/dataset/housing-london>

<https://data.london.gov.uk/dataset/housing-tenure-borough>

<https://www.gov.uk/government/statistics/english-housing-survey-2021-to-2022-private-rented-sector/english-housing-survey-2021-to-2022-private-rented-sector>

<https://www.kaggle.com/datasets/justinas/housing-in-london>

<https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads#yearly-file>

<https://data.london.gov.uk/dataset/indices-of-deprivation>

<https://assets.publishing.service.gov.uk/media/5d8b399a40f0b609946034a4/IoD2019_Infographic.pdf>

<https://data.london.gov.uk/dataset/average-private-rents-borough>

**Subject information:**

<https://www.london.gov.uk/sites/default/files/2018_lhs_london_housing_strategy.pdf>

<https://link.springer.com/chapter/10.1007/978-3-030-11674-3_2>

<https://www.mdpi.com/2413-8851/7/2/45>

<https://citymonitor.ai/environment/housing/wrap-it-up-and-start-again-the-future-of-londons-housing>

<https://www.wired.co.uk/article/airbnb-growth-london-housing-data-insideairbnb>

<https://en.wikipedia.org/wiki/Affordability_of_housing_in_the_United_Kingdom>

<https://upgo.lab.mcgill.ca/publication/>

<https://en.wikipedia.org/wiki/London_boroughs>

<https://www.london.gov.uk/sites/default/files/housing_research_note_4-_short-term_and_holiday_letting_in_london.pdf>

<https://www.wired.co.uk/article/airbnb-london-short-term-rentals>

<https://journals.sagepub.com/doi/full/10.1177/23998083211001836>

<https://en.wikipedia.org/wiki/Inner_London>

**Coding:**

<https://practicaldatascience.co.uk/data-science/how-to-geocode-and-map-addresses-in-geopy>

<https://www.kaggle.com/code/korfanakis/housing-in-london-eda-with-pandas-and-gif>

<https://medium.com/airbnb-engineering/using-machine-learning-to-predict-value-of-homes-on-airbnb-9272d3d4739d>

<https://www.analyticsvidhya.com/blog/2021/10/end-to-end-predictive-analysis-on-airbnb-listings-data/>

<https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression>

**Examples:**

<https://andrew-siu12.github.io/2019-06-15-Airbnb-Data-Analysis/>

<https://github.com/yalinyener/EDA-Airbnb-London>

<https://medium.com/analytics-vidhya/exploratory-data-analysis-on-airbnb-properties-in-london-39eb80da6d15>

**Project Structure:**

1. Introduction
2. EDA (Exploratory Data Analysis)
3. Modeling
4. Description of challenges / obstacles faced
5. Potential next steps / future directions

Deliverable schedule:

1. Proposal: Monday, November 6, 2023 - 11:59PM ET
2. Intermediate Check-in: **week** of November 13 - 19, 2023 - 11:59PM ET
3. Code Notebook: Saturday, December 9, 2023
   1. NOTE: This is the weekend of the Penn final period; this is functionally due the preceding week for this reason
4. Final Deliverable (Options): Saturday, December 9, 2023

**Proposal Structure** (~300 words in total)

1. Team composition and responsibilities
2. Data source
   1. >= 50,000 rows after cleaning and dropping NULL values
   2. Set of features properties useful in predicting outcomes
3. Project plan
   1. Explain what we intend to study
   2. What is the objective
   3. What types of models are being considered
4. Why is this project interesting / important?
5. What challenges and obstacles might we anticipate in this project
6. TA assignment -> Zijian Zhang

**Deliverable package:**

1. Code notebook, organized
2. Blog post (Medium)

## 

| **Milestone** | **Status** | **Date Due** | **Member Responsible** |
| --- | --- | --- | --- |
| Proposal Draft → Final | Not started | Nov 2, 2023, 11:00 am ET (US) | All |
| Data Cleaning Completed | Not started | Nov 16, 2023, 11:00 am ET (US) | See team roles & responsibilities |
| Individual Model section completed in notebook | Not started | Nov 30, 2023, 11:00 am ET (US) | See team roles & responsibilities |
| Notebook Completed | Not started | Dec 7, 2023, 11:00 am ET (US) | See team roles & responsibilities |
| Blog Completed | Not started | Dec 7, 2023, 11:00 am ET (US) | See team roles & responsibilities |

Inner London boroughs (13 + City of London)

Camden, City of London, Greenwich, Hackney, Hammersmith and Fulham, Islington, Kensington and Chelsea, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Wandsworth, Westminster

Narrow to 2019 only data to mitigate distorting effects of COVID-19 pandemic on short-term rental sector and property speculation, alignment with most recent UK government data on indices of deprivation.

Conduct linear regression modeling to determine direct relationship between STR frequency within borough and housing purchase costs.

Exploration of three regression model types (vanilla, lasso, ridge) to research question, with r-squared assessment to determine model with best fit to relationship.

Independent variable: number of STR listings (frequency)

Dependent variable: purchase / housing costs (acknowledgment that purchase cost shapes future rents).

Exploration of correlation between IMD score and borough tax rate as explanation for variance in relationship between borough, purchase price.

Data cleaning and management:

* Chunking and parsing housing purchase prices by borough
* Chunking and parsing STR data by borough
* Aggregation of UK IMD indices (generating borough median from sub-borough level data) and integration with borough data
* Integration of borough Council tax data

Features:

* 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

Additional data available for:

* Floor area for all housing sold in a given year (2019) [requires parsing and join on address]
* Median rent by borough for given year (2019)