Using Large Language Models to Predict Neighbourhood Change

Nick Malleson*[1,2], Thanh Bui Quang[3], Hang Nguyen Thi Thuy[4], Phe Hoang Huu[5], Minh Kieu[6], Molly Asher[1], Yiyu Wang[1] and Alexis Comber[1]

[1]School of Geography, University of Leeds, UK
[2]Alan Turing Institute, London, UK
[3]Faculty of Geography, VNU University of Science, Hanoi, Vietnam
[4]VNU Vietnam Japan University, Hanoi, Vietnam
[5]R&D Consultants, Hanoi, Vietnam
[6]Department of Civil and Environmental Engineering, University of Auckland, New Zealand

GISRUK 2025

Summary

This paper uses a large-language model (LLM) to read user-contributed text data – in this case property listings on Airbnb – and assign a score to indicate the degree to which the property's neighbourhood is potentially undergoing gentrification. Preliminary results for a case study in Bristol are evaluated. These scores will ultimately be used as an input in a model of dynamic spatio-temporal neighbourhood change with the aims of (i) quantifying the impact that neighbourhood reputation has on historical house prices and (ii) predicting the emergence of new house price bubbles.

KEYWORDS: Large-language model; neighbourhood change; gentrification; hedonic house price model;

*Corresponding author: n.s.malleson@leeds.ac.uk

1 Introduction

This paper forms part of a larger project entitled Integrated Analysis of Social Media and Hedonic House Prices for Neighbourhood Change (INTEGRATE). The aim of INTEGRATE is to analyse change in neighbourhoods over time and space to identify emerging gentrification. It will do this by combining user-generated text content (e.g. from social media, discussion forums, property reviews, etc.) with traditional hedonic house price models (HHPMs). Historically, HHPMs have focussed on estimating house prices based on characteristics of the individual property – such as such as age, number of bedrooms, floor area (Follain and Jimenez, 1985) – and characteristics of the location – such as amenities (schools, parks, shops, etc.) and distances to transportation hubs, workplaces, central business districts, etc. (Osland, 2010; Poudyal et al., 2009) or features like crime rates, ethnic composition, or pollution levels (Lynch and Rasmussen, 2001; Hui et al., 2007). A drawback with traditional HHPMs is that they neglect the impact of "intangible" (Huu Phe and Wakely, 2000) features related to "place" (Agnew, 2011), such as a neighbourhood's reputation. Unlike the traditional features used in HHPMs these 'intangibles' can be extremely difficult to estimate quantitatively.

This paper will present one aspect of the project: namely the process of using a large-language model to read the descriptions of properties listed on Airbnb and to estimate 'scores' that indicate the degree to which the property's neighbourhood has undergone (or is undergoing) gentrification. Ultimately we aim for these scores to capture some of the 'intangible' aspects that help to drive house price dynamics. Preliminary results suggest that the apparent 'reasoning' that the LLM undertakes to estimate a gentrification score is logical, but further validation is needed before the method can be used in earnest. The work also raises questions about whether the concept of 'gentrification' is still relevant in many areas, and to what extent there is sufficient gentrification-related 'signal' in noisy text data.

2 Methods and Data

We use the `Meta Llama 3.3 70B Instruct Turbo` model; a large, modern (at the time of writing!), open-source LLM. The model was accessed using an API provided by the service together.ai. The script to run the analysis is publicly available on the project's GitHub repository. We use a paid API because it provides access to models that are far too large and computationally expensive to be run locally.

There are numerous user-generated text-based data sources that could potentially reveal insight into gentrification patterns, including social media (Twitter, Facebook, Foursquare, etc.), discussion boards (Reddit), and many others. Here we experiment with property listings posted to Airbnb and made available through the service Inside Airbnb. Our hypothesis is that the language and terms used to advertise a property in an area that has been gentrified, or is currently undergoing gentrification, will be distinguishable from those used to advertise properties in other types of neighbourhoods.

The following illustrates the system prompt that is passed to the LLM (developed, somewhat ironically, with help from ChatGPT). A number of listings are then appended to the end of the prompt.

```
You are an expert in urban studies with a deep understanding of gentrification
and how it is discussed in public discourse.  I will provide you with some
Airbnb listings.  Your task is to analyse their text and determine the extent
to which they suggest that the neighbourhood or area referenced is experiencing
gentrification.  Specifically:

Read the listings closely and identify any words, phrases, or implications
that might indicate signs of gentrification, such as mentions of new luxury
developments, rising rents, displacement of long-time residents, upscale amenities
(e.g., artisanal coffee shops, craft breweries), changing demographics, or
neighbourhood 'revitalisation'.

Consider both explicit and implicit cues.  Explicit cues directly mention
new businesses or rising prices, while implicit cues might reflect subtle
neighbourhood changes.

Assign a score from 1 to 5, where 1 means not suggestive of gentrification
and 5 means highly suggestive.

Explain your reasoning in 1-2 sentences, referencing the specific words or
phrases in the tweet that led you to your conclusion.

Provide your answer strictly in the format '1.  Score.  Reasoning', '2.  Score.
Reasoning', '3.  Score.  Reasoning', etc., without any additional explanation
or commentary.
```

We obtain 2644 property listings accessible on 23 September 2024 from Inside Airbnb for the region of Bristol, UK (chosen, arbitrarily, because it is the location of the 2025 GISRUK conference). Figure 1 illustrates the locations of these listings along with their gentrification 'score' (discussed in Section 3). Each listing contains columns that capture its spatial location as well as a `description` of the property and a `neighbourhood_overview`. We combine those two text columns into a single one and append that to the prompt. Only 1,495 of the listings have both the `description` and `neighbourhood_overview`, so we drop the remainder from the remaining analysis.

The LLM reads the concatenated text column and returns both a score and an explanation for each listing. The code has been designed in such a way that multiple listings can be passed simultaneously to reduce the number of API calls and the total number of tokens passed to the LLM (the API used charges per token). As the LLM returns text there are some additional steps required to distinguish between the responses to the different listings and to extract the scores and the justifications from the raw text. Initially the LLM behaved erratically, returning scores in various formats that were often difficult to parse, but after some trial and error with the prompt the output stabilised such that all listings were assigned whole numbers and output was returned in a consistent format.

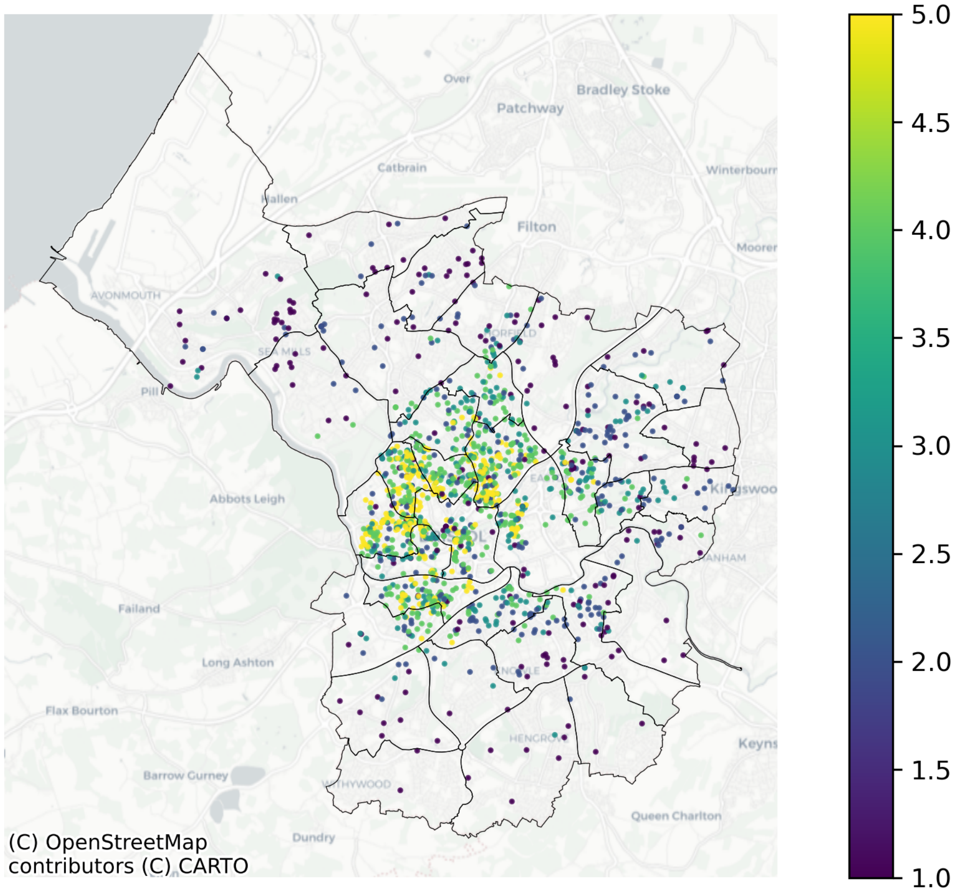

Figure 1: The Airbnb listing locations in Bristol and their gentrification scores

3 Results

3.1 Spatial Distribution of the Gentrification Scores

Figure 1 illustrated the spatial locations of the scores returned by the LLM. Whilst robust validation is necessary before drawing any firm conclusions, it is interesting that listings in the central areas of Bristol appear to receive higher scores than those on the outskirts. This is to be expected and, given that all the LLM knows about each listing is its text description, the possibility of a definable spatial pattern that aligns with theoretical expectations is encouraging.

3.2 Example LLM Outputs

We ask the LLM to return an explanation for the choice of its score, and these seem qualitatively sensible. The LLM is able to refer to features of the listing and neighbourhood that are (or are not) indicative of gentrification. Examples of LLM outputs are not included here due to space limitations, but will be presented during the conference for discussion. Again this is encouraging, but requires much more considered validation before drawing any firm conclusions.

3.3 Output Variability

Although the LLM justifications for its scores appear reasonable, it is worth noting that a reasonable justification does not mean that other, different, scores could not also have been justified. To explore this, and to explore output variability more generally, we repeatedly ask the LLM to estimate the gentrification score for each listing and then compare the scores across runs. The LLM is probabilistic, so even with the same inputs and parameter settings the output will vary slightly. We run the LLM over all listings three times and calculate the Fleiss Kappa (Fleiss, 1971) which is a Kappa statistic that is well suited to problems where there are more than two 'annotators'. The score is 0.74 which indicates "substantial" (Landis and Koch, 1977) agreement across the three experiments.

4 Discussion and Future Work

This paper explores the use of a large language model (LLM) as a means of assessing neighbourhood change through textual analysis of Airbnb property descriptions, with a focus on identifying signs of gentrification. While the initial findings suggest potential in this approach, several limitations and challenges warrant further scrutiny.

With respect to data, the work is limited to the study of cities that are available through the Inside Airbnb service. In addition, the very presence of AirBnB listings may be indicative of gentrification, regardless of the neighbourhood descriptions. Currently we do not attempt to control for this effect, although it will be considered during future validation stages (discussed below).

The concept of gentrification itself may require reexamination. Its applicability and relevance can vary significantly across contexts, both geographically and temporally. For example, in cities where housing markets are already highly unaffordable, the traditional markers of gentrification may no longer be meaningful or may require alternative interpretations. Similarly, the understanding of gentrification in non-Western settings could differ, raising questions about the universality of the approach. For example, the INTEGRATE project has partnered with the Vietnam National University, Hanoi, and it is unclear whether gentrification is a meaningful concept in an environment where all land is technically owned by the state and rented by citizens.

Detecting gentrification through text introduces inherent difficulties. While property descriptions may hint at changing neighbourhood characteristics, they often lack the depth or consistency needed for robust analysis. This is compounded through the use of a Likert scale (1–5) where distinguishing between the different levels can be difficult. Future work is experimenting with more concrete definitions of different types/stages of gentrification, rather than relying on an abstract scale.

Finally, validation remains a key concern. Although some alignment between LLM-generated scores and theoretical expectations may have been observed, we have not yet attempted to implement any proper testing against independent benchmarks. Immediate future research should address these limitations to better understand the feasibility and limitations of this approach.

5 Acknowledgements

This research is supported by UKRI (ESRC) funding ES/Y006259/1 under the Digital Footprints scheme.

References

Agnew, J. (2011). Space and place. In J. Agnew & D. Livingstone (Eds.), *The SAGE Handbook of Geographical Knowledge* (pp. 316–331). London: SAGE Publications.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382, https://doi.org/10.1037/h0031619.

Follain, J. R. & Jimenez, E. (1985). Estimating the demand for housing characteristics: A survey and critique. *Regional Science and Urban Economics*, *15*(1), 77–107, https://doi.org/10.1016/0166-0462(85)90033-X.

Hui, E. C., Chau, C., Pun, L., & Law, M. (2007). Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Building and Environment*, *42*(6), 2333–2343, https://doi.org/10.1016/j.buildenv.2006.05.004.

Huu Phe, H. & Wakely, P. (2000). Status, Quality and the Other Trade-off: Towards a New Theory of Urban Residential Location. *Urban Studies*, *37*(1), 7–35, https://doi.org/10.1080/0042098002276.

Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159, https://doi.org/10.2307/2529310.

Lynch, A. K. & Rasmussen, D. W. (2001). Measuring the impact of crime on house prices. *Applied Economics*, *33*(15), 1981–1989, https://doi.org/10.1080/00036840110021735.

Osland, L. (2010). An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling. *Journal of Real Estate Research*, *32*(3), 289–320, https://doi.org/10.1080/10835547.2010.12091282.

Poudyal, N. C., Hodges, D. G., & Merrett, C. D. (2009). A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy*, *26*(4), 975–983, https://doi.org/10.1016/j.landusepol.2008.11.008.