DISSERTATION

Unpacking the Nuances of London's Neighbourhood Change & Gentrification Trajectories

By Joshua Yee Pin Yi

August 2019

This dissertation is submitted in part requirement for the MSc in Smart Cities and Urban Analytics at the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, UCL.

Submitted: 30 August 2019

Supervisor: Dr Adam Dennett

Module ID: BENVGSC6

Student ID: 18121142

Word Count: 11,999 (from pages 13 – 117)

Abstract

Amid the simmering global economic uncertainties and geopolitical instabilities of our present day, deepening social inequalities have nonetheless remained a constant in our contemporary urban landscapes. Particularly, within global cities, like London for instance, the disparities have become increasingly apparent. The gravity of the situation is underscored by recently published reports suggesting London's concurrent embodiment as home to not only the majority of the UK's richest 1%, or indeed the 0.1%, but also its urban poor. In spite of such polarisations, a sense of indifference towards the needs of the average Londoner, not to mention the underclasses, seemingly prevails throughout London's hyper-commodified housing market, which is typified by sky-rocketing property prices and a chronic undersupply of truly affordable housing options.

Gentrification, emanating from the nexus of such trends, have thus been endemic in London; propelling socio-spatial transformations in localised neighbourhoods and displacing incumbent residents. Despite gentrification manifesting in more accentuated and diversified ways than ever before, recent research has only been tepid in honing the critical edge necessary for effectively distinguishing gentrification from other forms of neighbourhood change and to rigorously dis-aggregate gentrification's nuances. This has consequently limited the depth of contributions that academia can make towards robust policy-making.

Addressing such issues, this paper employs a novel empirical approach, that synergises recent advancements in Machine Learning, data availability and spatial analysis, to systematically examine the variegated trajectories of neighbourhood change

across London, before filtering down to critically identify gentrifying neighbourhoods with precision and clarity. The nature of gentrification's mutations and their spatial patterning in London are further extracted using a combination of Principal Component Analysis, K-Means clustering and in-depth spatial analysis. Machine Learning is subsequently adopted to model gentrification's observed trends and predict its future frontiers; thereby offering academia and policy-makers unprecedented and highly-contextualised insights into gentrification's dynamics and geographies.

Declaration of Authorship

I, Joshua Yee Pin Yi, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 11,999 words in length from the introduction to concluding chapter (inclusive).

Signed: _____

Date: 30/08/2019

Table of Contents

Chapter 1: Introduction 13

Chapter 2: Research Overview 15

2.1 Purpose 15

2.2 Scope 16

2.3 Objective 18

2.4 Outline 18

Chapter 3: Literature Review 20

3.1 Gentrification as an Evolving Process of Neighbourhood Change 20

3.2 Gentrification as an Urban Phenomenon Entailing Stark Socio-Spatial Consequences 23

3.3 Gentrification as a Single Typology of Neighbourhood Ascent 28

3.4 Gentrification as a Subject of Empirical Analysis 29

Chapter 4: Methodology 34

4.1 Data Collection & Wrangling 34

4.1.1 ONS' Census and Housing Transactions Data 34

4.1.2 GLA's Planning Permissions Data 36

4.1.3 CDRC's Population Churn Data 40

4.2 Workflow & Methods 42

4.2.1 Stage 1: Scoring LSOAs to Identify their Broad Neighbourhood States ... 44

4.2.2 Stage 2: Clustering Ascending LSOAs to Differentiate Typologies of Neighbourhood Ascent 48

4.2.3 Stage 3: Clustering Gentrifying LSOAs to Differentiate Typologies of Gentrification 52

4.2.4 Stage 4: Building a ML model to Classify and Predict Gentrifying LSOAs and their Typologies 54

Chapter 5: Analysis and Results 58

5.1 Exploring the States of London's LSOAs 58

5.2 Defining Contextualised Neighbourhood Typologies 62

5.2.1 Typologies of Neighbourhood Ascent 62

5.2.2 Typologies of Gentrification 72

5.3 Uncovering Spatial Patterns 81

5.3.1 Typologies of Neighbourhood Ascent 81

5.3.2 Typologies of Gentrification 85

5.4 Predicting Future Neighbourhood Trajectories 90

5.4.1 Model to Predict Future Gentrifying LSOAs 90

5.4.2 Model to Predict Typologies of Future Gentrifying LSOAs 101

Chapter 6: Discussion 108

6.1 Contextualising the Results 108

6.1.1 'Icebergs' of Fulham's Super-Gentrifying Neighbourhoods 112

6.1.2 The Battlegrounds of Newham's North-east 114

Chapter 7: Conclusion 116

Bibliography 118

List of Figures

Figure 1. Reference Map of Boroughs & LSOAs in the Greater London Area 16
Figure 2. Tract Typologies Generated & Used By UDP 27
Figure 3. Example of Regularising Data Aggregated Using 2001 LSOAs to the Updated 2011 LSOAs 35
Figure 4. Frequencies of Planning Permissions in LDD Dataset 37
Figure 5. Sample of Raw Datapoints in LDD Dataset 38
Figure 6. Planning Permissions for Redevelopment / Conversion of Existing Residential Properties 39
Figure 7. Planning Permissions for New-Built Residential Developments 39
Figure 8. Overview of Entire Workflow 43
Figure 9. Dot & Whisker Plots for 4 Proxy Variables (inclusive of insets) 46
Figure 10. Equation of Robust Scaler 46
Figure 11. Equations of Scoring Metrics 55
Figure 12(a)-(b). Statistics of Principal Components (PC) Derived From PCA 58
Figure 13. Composite Index Scores of all LSOAs in 2001 59
Figure 14. Composite Index Scores of all LSOAs in 2011 59
Figure 15. Neighbourhood States of all LSOAs 61
Figure 16. Correlation Heatmap for Input Variables in Stage 2 63
Figure 17. Underlying Distributions of the Data Before and After Transformation 64
Figure 18. Results of Silhouette Analysis to Determine Optimal Number of Clusters (K) in Stage 2 65
Figure 19. Calinski-Harabasz Index Scores for Clusters Derived Through Different Dataset Permutations (Stage 2) 65
Figure 20. Characteristics of Gentrification Cluster 66
Figure 21. Characteristics of Incumbent Upgrading Cluster 68
Figure 22. Characteristics of Re-urbanisation Cluster 70
Figure 23. Correlation Heatmap for Input Variables in Stage 3 72
Figure 24. Underlying Distributions of the Data Before and After Transformation 73

Figure 25. Results of Silhouette Analysis to Determine Optimal Number of Clusters (K) in Stage 3 74
Figure 26. Calinski-Harabasz Index Scores for Clusters Derived Through Different Dataset Permutations (Stage 3) 74
Figure 27. Characteristics of Super-Gentrification Cluster 75
Figure 28. Characteristics of Marginal Gentrification Cluster 77
Figure 29. Characteristics of Mainstream Gentrification Cluster 79
Figure 30. LSOAs in Typologies of Neighbourhood Ascent 82
Figure 31. ArcMap Outputs for Spatial Autocorrelation Test (Global Moran's I) 83
Figure 32. Gentrification's Hotspots 84
Figure 33. LSOAs in Typologies of Gentrification (inclusive of insets for Kensington and Chelsea and Islington) 86
Figure 34. Spatial Autocorrelation Test Results and Mapping of Super-Gentrification's Hotspots 87
Figure 35. Spatial Autocorrelation Test Results and Mapping Marginal Gentrification's Hotspots 88
Figure 36. Spatial Autocorrelation Test Results and Mapping Mainstream Gentrification's Hotspots 89
Figure 37. Base Models' Prediction Accuracy on Test Set 91
Figure 38. Variable Importance for Predicting Gentrifying LSOAs 93
Figure 39. Model's Results at Every Iteration of Variable Pruning (Stage 2) 93
Figure 40. Pruned Models' Prediction Accuracy on Test Set and Entire 2001-2010 Dataset 95
Figure 41. Model's Errors in Predicting Gentrifying LSOAs (Before Pruning) 96
Figure 42. Comparison of (a) PC1 Scores and (b) Locations of Gentrifying LSOAs 97
Figure 43. Illustration of the Score Assignments to Replicate the Clustering Effect 98
Figure 44. Accuracy of Pruned Model with Geog. Covariates Predicting on Entire 2001-2010 Dataset 99
Figure 45. Model's Predictions on Future Gentrifying LSOAs 99
Figure 46. Base Models' Optimal Hyperparameters and Prediction Accuracy on Test Set 102
Figure 47. Variable Importance for Predicting Typologies of Gentrifying LSOAs 103

Figure 48. Model's Results at Every Iteration of Variable Pruning (Stage 3)..... 103
Figure 49. Model's Errors in Predicting LSOAs' Gentrification Typologies 105
Figure 50. Accuracy of Model at Every Stage of the Refinement Process..... 106
Figure 51. Model's Predictions on the Typologies of Future Gentrifying LSOAs..... 107
Figure 52. Map Showing Outline of London Docklands and the Typologies of Neighbourhood Ascent..... 109
Figure 53. Conceptual Positionings of Neighbourhood Ascent and Gentrification Typologies under the schema of Neighbourhood States and Change 111
Figure 54(a)-(c). Gentrification Typologies and 'Iceberg Basements' within the Borough of Hammersmith and Fulham Borough 113
Figure 55. Gentrification Typologies in the Borough of Newham..... 115

List of Tables

Table 1. Summary of Socio-spatial Impacts Produced by Gentrification..... 24
Table 2. Variables Incorporated in Research..... 41
Table 3. Labels of variables used for scoring LSOAs 45
Table 4. Input Variables for Clustering Ascending LSOAs 50
Table 5. Breakdown of Data Transformation and Re-scaling Methods 51
Table 6. Input Variables for Clustering Gentrifying LSOAs 53
Table 7. Hyperparameters of Chosen ML Models 56
Table 8. Top 5 Boroughs with Highest Share of LSOAs in Ascent and Decline 62
Table 9. Best Hyperparameter Combinations Extracted from GridSearchCV 90
Table 10. Least Important Variables Pruned from Model (Stage 2) 94
Table 11. Least Important Variables Pruned from Model (Stage 3) 104

List of Acronyms and Abbreviations

CDRC	Consumer Data Research Centre
CI	Composite Index
FY	Financial Year
GLA	Greater London Authority
K	Number of Clusters
LDD	London Development Database
LSOA	Lower Super Output Area
MAUP	Modifiable Areal Unit Problem
ML	Machine Learning
MSOA	Middle Super Output Area
NS-SEC	National Statistics – Socio-economic Classification
OA	Output Areas
ONS	Office of National Statistics
PC	Principal Component
PCA	Principal Component Analysis
RF	Random Forest
SD	Standard Deviation
SES	Socio-economic Status
UDP	Urban Displacement Project
UK	United Kingdom
US	United States of America

Acknowledgements

This dissertation would not have been possible without the wealth of thoughts, contributions and encouragements of many others. Above all, my special thanks and deepest gratitude is owed to my supervisor, Dr Adam Dennett, for the astute guidance and insights he had provided throughout the course of the dissertation. In particular, Adam's wise and timely feedback was a key and irreplaceable factor that led to the success of this research. I am also grateful to Dr Cao Mengqiu, Mr Alexander Vassilev and Ms Leong Sze Yuan for the initial groundwork they undertook in the Summer of 2018 to clean and prepare a collection of Census datasets which were eventually incorporated as part of the research.

Part of the data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 334, ES/L011840/1; ES/L011891/1. Additional data sources include the Office of National Statistics and Greater London Authority who, like the Consumer Data Research Centre, I owe my thanks and appreciation for the use of their data to enable the research implemented in this dissertation.

I would also like to express my thankfulness to the lecturers and staff at UCL's Centre for Advanced Spatial Analysis (CASA) for their enthusiasm in teaching and imparting the knowledge that was critical in driving my academic progression.

Finally, I am hugely indebted to the Housing and Development Board (HDB) of Singapore for granting me this opportunity to further my skills and expertise in CASA, as well as to my wife and family for their continuous encouragement and support that sustained me at every step of this incredible journey.

1 Introduction

Arguably amongst the most contentious of issues, gentrification commonly evokes strong debates in academia and policy circles across the world. More than simply a topic of debate, gentrification is critically a process of neighbourhood change which has often served to fulfil the residential and lifestyle preferences of specific societal groups, but disparagingly, at the expense of others (Slater, 2011). This systemic (re)production of winners and losers across society and space, which is typically entrenched in the stratifications caused by socio-economic status (SES), race and gender disparities, thus lies precisely at the heart of gentrification's controversial and highly-politicised nature (Lees et al., 2008).

Although studies on gentrification are nothing new, several past works have tended to overlook the nuanced typologies of neighbourhood change that co-exist with gentrification (Hochstenbach and van Gent, 2015). Consequently, gentrification has typically been conflated with other forms of neighbourhood change, which inhibit the robust generation of knowledge and insights to support policy-making. Atkinson (2008, p.2634) aptly outlined the fallacies of this situation as he criticised that research has 'tended to label too many kinds of neighbourhood change as gentrification and this elasticity has reduced the bite of critical studies of its localised appearance and has diminished policy-maker interest'.

Besides needing to better distinguish gentrification from other typologies of neighbourhood change, tremendous utility can be yielded to both theory-building and real-world applications if research is able to efficaciously diagnose gentrification's varying manifestations (Lees, 2009). Particularly in global cities where staggering social

inequalities exist (Massey, 2017), there is growing evidence to suggest their emerging role as epicentres for gentrification's latest variants (Rérat et al., 2010). Hence, to ensure the discipline's continued relevance amidst gentrification's increasingly fluid expressions, scholars have since been urged to 'readjust their lens of enquiry and analyses' and thoroughly consider gentrification's present-day trends in research (Smith and Butler, 2007, p.2).

This paper is therefore an attempt to re-instil the 'bite' of gentrification research and deliver contemporarily relevant insights which will enable robust policy-making and deepen scholarly understandings concerning neighbourhood change, gentrification and their heterogenous typologies. Leveraging on a repertoire of novel Machine Learning (ML), spatial analytical techniques and multi-dimensional data, this study will comprehensively examine London, chosen as this paper's study area, to unpack the continuum of active neighbourhood change typologies, before drilling down to identify neighbourhoods which are genuinely gentrifying. Deep-dives will be conducted to further segment the diverse gentrifying processes unfolding in London, model their patterns and eventually predict their potential trajectories.

2 Research Overview

2.1 Purpose

The underlying motivations for this study are 3-fold. Firstly, as explained above, a well-defined categorisation of neighbourhood change and gentrification typologies is urgently required to meaningfully enrich existing discourses surrounding gentrification, and serve as a catalyst for robust decision-making by urban planners and city administrators. To address this gap, our analysis is driven to critically deconstruct the diversified patterns of neighbourhood change and gentrification, and systematically identify their spatial underpinnings.

Secondly, in the vein of applied research, this paper intends to galvanise forward-looking predictions of gentrification's future frontiers, as a basis for formulating pre-emptive and highly-effectual policies targeting the phenomenon's ill-effects. On ground, these insights can also help stimulate much-required discussions amongst local community members regarding their neighbourhood's future and mobilise support for defending against gentrification's negative impacts in areas where its 'waves' are expected to buffet (Hackworth and Smith, 2001, p.465).

Thirdly, recognising that gentrification's patterns have become increasingly challenging to pinpoint, largely due to their proliferating variations, this study aspires to elevate current methodological approaches in ways that can negotiate the mounting complexities. Specifically, by employing a unique workflow that integrates ML with spatial analysis, and undergirded by the incorporation of novel datasets, this paper

hopes to push the envelope of traditional empirical methods and showcase refreshing ways in which gentrification can be effectively analysed at scale.

2.2 Scope

This study is scoped to quantitatively analyse all neighbourhoods located within the Greater London area *(see Figure 1)*, with a focus on identifying areas experiencing neighbourhood change and gentrification between 2001-2011, before predicting future scenarios. Although a definitive timeline regarding the realisation of these predictions cannot be foretold, it is expected that they could occur within the short to medium-term following 2016, since the most recent data utilised for predictions were from that year.

Figure 1. Reference Map of Boroughs & LSOAs in the Greater London Area

The decision to adopt London as the contextual anchor for analysis was a thoroughly considered one. Given London's status as a leading cosmopolitan city that attracts global capital and people from all facets of the socio-economic and cultural

spectrums (Massey, 2017), it is hardly surprising that hyper-commodification has set into London's residential property market, and has reared the chronic lack of affordable housing as its dark underbelly (Watt and Minton, 2016). This is further substantiated by recent reports listing London as simultaneously home to not only the largest proportion of the UK's super-rich (Roantree et al., 2019) but also its urban poor (Social Metrics Commission, 2019); thereby adding to the hypothesis that such gaping societal gulfs have fuelled the rise of gentrification's latest mutations in the city (Butler, 2011). London hence provides opportunities for our analysis to engage directly with gentrification's latest evolutions and develop situated understandings for a context which requires them most pressingly.

However, as elucidated by van Ham & Manley (2012), a key challenge for any research concerning neighbourhoods has been to suitably define the 'neighbourhood' unit for analysis, since 'neighbourhoods' are more to do with fluid, social constructs than they are as objective entities which can be easily categorised. Nonetheless, aligned to Galster's (2001, p.2112) argument that neighbourhoods could be conceived as a 'bundle of spatially based attributes associated with clusters of residences', the 2011 Lower Layer Super Output Areas (LSOAs) was eventually rationalised as the optimal spatial unit to facilitate our analysis.

Purposefully designed by the Office for National Statistics (ONS) as a set of relatively fine-grained, statistical areas that offered consistency in its population sizes (averaging around 1,600 residents) and some degree of coherence in their social composition (ONS, 2004), LSOAs were ideal for analytical purposes. While other geographical units had been developed, such as the finer output areas (OA) or coarser

Middle Super Output Areas (MSOA), many of the datasets which this study sought to use were unavailable at the OA-scale while MSOAs could not offer the same level of spatial detail.

With respect to the study's temporal timeframe, 2001 and 2011 were chosen to tie in with the UK's 2 most recent census. The advantage of this selection was the readiness of rich, high-resolution census data capturing the state of London's socio-physical landscape, which are essential for examining gentrification. Access to the same quality of information outside the census periods are notably scarce or otherwise entirely unattainable.

2.3 Objective

As the objective, this study aims to address the following research questions:

1) How can the broad processes of neighbourhood change (i.e. neighbourhood ascent and decline) be understood in the context of London's neighbourhoods?

2) How can ML and novel non-census-based data help identify gentrification?

3) Which typologies of neighbourhood ascent and gentrification are found in London and what are their spatial patterns?

4) What predictions can be made about future gentrification trends in London?

2.4 Outline

The remaining chapters are outlined below:

- *Chapter 3* provides a literature review covering the key concepts, issues and academic works surrounding neighbourhood change and gentrification
- *Chapter 4* discusses in-depth the empirical methodology used in the analysis

- *Chapter 5* presents the data analysis and its findings, particularly expounding on the derived typologies of neighbourhood change and gentrification, their spatial patterns and predictions
- *Chapter 6* involves a discussion that puts the study's findings into context with case examples
- *Chapter 7* summarises the implication and limitations of this research

3 Literature Review

3.1 Gentrification as an Evolving Process of Neighbourhood Change

Gentrification, a term traced historically to sociologist Ruth Glass (1964), was initially coined to describe a complex, but distinctive, pattern of socio-spatial transformations observed around parts of inner London during the 1960s. In its purest form, gentrification referred to the process by which pre-existing working-class neighbourhoods were increasingly being taken over by the gentry of middle-class status, that concurrently spurred the rehabilitation of dated residential properties, upward shifts in housing prices and the ensuing displacement of residents with lower SES from such areas (Hamnett, 2003; Lees et al., 2008).

Concomitant with the identification of gentrification as a unique process of neighbourhood change, explanations concerning its causes were theorised and intensely debated from an early stage (Slater, 2011). Traditionally, scholars were divided between 2 ideological camps – those who advocated for 'production' reasons as the impetus for gentrification versus others who viewed it as an outcome of 'consumption' factors (Chapple et al., 2017, p.14). Briefly, production-driven explanations posited gentrification as deeply entwined with capital markets, cycles of investment and disinvestment, and facilitated by a wide network of actors, including 'developers, builders, mortgage lenders, government agencies, real estate agents' (2017, p.15). Augmenting this argument, Smith (1979, p.545), in his seminal work, mooted the 'rent gap' as gentrification's essential basis, wherein substantial differences between the cost

of purchasing properties in disinvested neighbourhoods and the potential return on investment that gentrifiers could yield (through sale or rent) were the main mechanisms in attracting bouts of capital inflows and eventually middle-class residents. Conversely, proponents of consumption-driven reasonings, such as Ley (1996) and Butler (1997), tended toward the socio-cultural orientations of the burgeoning middle-class and emphasised the role of the individual, particularly their preferences for certain lifestyles, neighbourhood aesthetics and character, as the forces that mediated residential location choices and perpetuated gentrification's manifestations.

Notwithstanding the coinage of the term 'gentrification' more than 5 decades ago, it remains highly-relevant in conceptualising the distinct trajectory of neighbourhood transformations still experienced across our urban landscapes today, albeit having evolved in scope to encompass several developments that have since crystallised in the discourse (Lees, 2007; Smith and Butler, 2007). The key assertions put forth by academics explaining the structural transformations that have (re)moulded gentrification's scope are summarised in 3 points.

Firstly, underpinned by waves of globalisation in recent history, the breadth and depth of gentrification have grown immensely (Lees et al., 2015). In facilitating the seamless movement of people and capital between places, and catalysing cycles of global economic restructuring, globalisation is cited as a key accelerator driving the spread of gentrification to places conventionally untouched by this phenomenon (Lees et al., 2016); permeating into cities within the Global South and secondary urban areas of the North, like Bristol (Bridge, 2003) and Leeds (Dutton, 2003). Unsurprisingly, such trends have evoked proclamations that 'gentrification is now global' (Atkinson and

Bridge, 2005, p.1) and cemented gentrification's status as a research area of paramount importance.

Secondly, in parallel with its broadening reach, scholars have highlighted gentrification's increasingly diverse variants over the years (Van Criekingen and Decroly, 2003; Butler and Lees, 2006). Although a pluralism of examples exist, 2 contemporary derivatives of gentrification, that juxtapose well against classical models and have proven pertinent in global city contexts, are discussed in this review. 'Super-gentrification' describes the further gentrifying of previously gentrified neighbourhoods by the globally-connected elites at the pinnacle of the socio-economic ladder (Lees et al., 2008, p.130). Under such circumstances, the victims of displacement are consequently middle-class residents, and not the working-class that one expects from normative gentrifying processes. Even where disinvested, working-class neighbourhoods have been gentrified, works have shown that incoming gentrifiers have not necessarily conformed to the wealthy, middle-class stereotype (Rose, 1984). Instead, relatively marginal segments of society who lack deep financial resources, but are nonetheless attracted to the low rents, locational and aesthetic appeals of working-class neighbourhoods, have sought to gentrify such areas en masse – a differentiated gentrifying process labelled 'marginal gentrification' (Owens, 2012, p.347).

Thirdly, in contrast to the production-consumption dichotomy that had earlier split the discipline, there is now a greater appreciation amongst scholars that both paradigms are mutually valid in explaining the causes of gentrification (Rigolon and Németh, 2019). Partly due to the deepening nuances of gentrification's manifestations and its new frontiers, academics now agree that this urban phenomenon cannot be

characterised through siloed explanations but integrated ones; which perhaps reflect the realisation of Hamnett's (1984, p.284) earlier thesis asserting gentrification's embodiment as 'simultaneously a physical, economic, social and cultural phenomenon'. Such developments thus set the stage for all research, including this paper, to recognise gentrification as a multi-faceted issue that needs to be spatially and temporally situated.

3.2 Gentrification as an Urban Phenomenon Entailing Stark Socio-Spatial Consequences

While gentrification's manifestations are in themselves intriguing lines of inquiry, it has undeniably been gentrification's implications that have propelled wide interest from academia and policy circles in this topic. Atkinson & Bridge (2005) provided an excellent summary of gentrification's extensive impacts, which is reproduced in *Table 1*.

Table 1. Summary of Socio-spatial Impacts Produced by Gentrification

Positive	*Negative*
	Displacement through rent/price increases
	Secondary psychological costs of displacement
Stabilisation of declining areas	Community resentment and conflict
Increased property values	Loss of affordable housing
	Unsustainable speculative property price increases
	Homelessness
Increased local fiscal revenues	Greater take of local spending through lobbying/articulacy
Encouragement and increased viability of further development	Commercial/industrial displacement
Reduction of suburban sprawl	Increased cost and changes to local services
	Displacement and housing demand pressures on surrounding poor areas
Increased social mix	Loss of social diversity (from socially disparate to rich ghettos)
Rehabilitation of property both with and without state sponsorship	Under-occupancy and population loss to gentrified areas

Source: (Atkinson & Bridge, 2005, p.5)

Evidently, gentrification creates existential repercussions, both positive and negative, on our social and physical landscapes. Insofar as previous attempts were made to paint a rosy picture of gentrification by obfuscating its downsides, particularly by private developers and amidst state-sanctioned programmes (Lees, 2008), critical scholars have, through evidence-based research, exposed gentrification as a 'largely

negative process' where its costs severely outweigh any potential benefit (Atkinson, 2004, p.126).

Amongst the impacts, displacement arguably ranks most consequential of all, given the profound devastations it heaps on gentrification's losers (Slater, 2011). While urban displacement can basically be understood as the removal and replacement of particular communities from an area, it is worthwhile noting that gentrification-induced displacement can be disaggregated into 3 broad threads (Marcuse, 1985), of which 2 are covered here due to their relevance. The first thread, 'direct displacement', was introduced to elucidate the displacement of incumbent residents via economic or physical mechanisms, such as landlords increasing rents or depriving tenants of utilities (1985, p.205). 'Exclusionary displacement', on the other hand, signified the inability of new households to access housing in previously affordable neighbourhoods due to gentrification-related spikes in housing prices and rents (1985, p.206).

Regardless of form, the negativities resulting from urban displacement are onerous and grave, to which Marcuse (1984, p.931) puts concisely – 'at worst it leads to homelessness, at best it impairs a sense of community'. Given this situation, gentrification and the displacement it entails cannot be left unignored, thereby framing the imperatives for critical research to be focused on this area.

To combat these prevailing trends, several studies have acknowledged the need for tools that effectively detect gentrification (Chapple and Zuk, 2016). Early-warning systems such as those pioneered under the Urban Displacement Project (UDP) are positive exemplars, whereby various US cities were analysed for their vulnerability to gentrification and displacement (Zuk and Chapple, 2015a). Employing primarily census

and built environment data to discern urban processes, UDP's early-warning systems assigned census tracts to distinct typologies depending on whether they were exposed to risks of gentrification and displacement or were already in their advanced stages *(see Figure 2)*.

A survey of recent works however revealed that the majority of such efforts have focused on US cities, which poses 2 theoretical and methodological barriers hampering the direct transposition of applications to other contexts. Firstly, as datasets comparable to the ones used for identifying urban trends in US-based studies may not be readily available in other countries, there are, at best, limited opportunities for studies outside the US to replicate their methodologies. Secondly, as gentrification is a contextually-specific phenomenon (Freeman et al., 2016), insights drawn from US-based studies cannot be generalised to encompass other places. Such reflections are strikingly apparent in the UK's case since the nature of gentrification has proven dissimilar to the US (Lees, 1994), while certain datasets adopted in UDP's methodology, such as those explicitly concerning low-income households, do not possess equivalents in the UK (White and McLaren, 2009). These are hence existing research gaps in the field that can only be plugged by undertaking rigorous, in-depth studies that are tailored toward localised settings.

Figure 2. Tract Typologies Generated & Used By UDP

Displacement and Gentrification Census Tract Typologies	
Typology (tract income level)	**Typology Criteria**
Not Losing Low-Income Households (Low Income)	• Low Income Tract in 2015 • Not classified as At Risk of, Ongoing, or Advanced Exclusion

At Risk of Gentrification or Displacement (Low Income)	• Low Income Tract in 2015
	• Vulnerable to gentrification in 2000 (see appendix)
	• Has 2 out of 4 risk factors:
	  • Fixed rail station
	  • Historic housing stock
	  • Employment center
	  • "Hot housing market" (see appendix)
	• Not currently undergoing displacement or ongoing gentrification

Ongoing Gentrification / Displacement of Low-Income Households (Low Income)	• Low Income Tract in 2015
	• Vulnerable to gentrification in 2000 (see appendix)
	• Loss of low income households 2000-2015
	• Has 1 out of 2:
	  • A "Hot housing market" (see appendix)
	  • Decline in low Income migration rate
	• Or Gentrified in 1990-2000 or 2000-2015 (See appendix)

Advanced Gentrification (Moderate to High Income)	• Moderate to High Income Tract in 2015 • Gentrified in in 1990-2000 or 2000-2015
Not Losing Low-Income Households	• Moderate to High Income Tract in 2015 • Not classified as At Risk of, Ongoing, or Advanced Exclusion

At Risk of Exclusion (Moderate to High Income)	• Moderate to High Income Tract in 2015
	• Has 2 out of 4 risk factors:
	  • Fixed rail station
	  • Historic housing stock
	  • Employment center
	  • "Hot housing market" (see appendix)
	• Not currently undergoing exclusion

Ongoing Exclusion / Displacement of Low-Income Households (Moderate to High Income)	• Moderate to High Income Tract in 2015
	• Loss of Low Income households 2000-2015
	• Has 1 out of 2:
	  • A "Hot housing market"
	  • Decline in low income migration rate

Advanced Exclusion (Moderate to High Income)	• Moderate to High Income Tract in 2015 • Low and declining proportion of low-income households since 2000 • Low low-income migration in 2015

Source: (Zuk & Chapple, 2015a)

3.3 Gentrification as a Single Typology of Neighbourhood Ascent

Despite the keen interest in gentrification, this urban phenomenon should not be mis-conceived as being the only form of neighbourhood ascent, or indeed neighbourhood change. In an enlightening piece, Owens (2012, p.347) clarified this by delineating 'incumbent upgrading', 'neighbourhood upgrading' and gentrification as some of the typologies constituting neighbourhood ascent. Moving up the hierarchy, 'neighbourhood ascent' and 'decline' are then categories nested underneath the broad banner of neighbourhood change (2012, pp.347–348).

Although neighbourhood enhancements are inherent in all typologies of neighbourhood ascent, subtle differences distinguish these typologies. 'Incumbent upgrading' referred to in-situ improvements experienced by existing residents, potentially in terms of their SES or housing conditions, over time (2012, p.347). Therefore, contrary to gentrification, 'incumbent upgrading' does not entail displacement (Van Criekingen and Decroly, 2003, p.2456). Separately, 're-urbanisation' is another form of urban change that became prevalent in the UK since the 1990s when the 'urban renaissance' movement was vigorously promoted (Boddy, 2007, p.90). Characterised by the injection of new-built residences to regenerate city-centres and brownfield sites, re-urbanised areas typically attracted specific demographic groups – such as 'younger single people or childless couples' (2007, p.95) in Bristol's case – but did not entail direct displacement since existing residences were unaffected. Exclusionary displacement was nonetheless a potentiality if the introduction of new

developments and populations made, otherwise financially accessible, houses in the vicinity unaffordable for low-income groups (2007, p.99).

Given the nuances inherent in neighbourhood ascent trajectories, Owens (2012, p.364) convincingly argued for the merits of being able to discern them, wherein she asserted that a holistic comprehension of the typologies 'sharpens the concept of gentrification and provides a fuller and more accurate depiction of neighbourhood ascent'. In practical terms, such insight allows neighbourhood issues to be properly framed, and facilitates the strategic deployment of resources and infrastructure required to treat the issues. In light of these advantages, it is thus surprising that the typologies of neighbourhood ascent have not been duly considered or interrogated in the majority of gentrification studies. A case in point is Reades et al.'s (2019) recent study of gentrification trends in London. Notwithstanding the novel and insightful use of ML to model urban trends, an approach discussed in the following sub-section, the study essentially assumed relative 'uplifts' observed in neighbourhoods as affirmative signs that these areas were gentrifying (2019, p.3). However, as explained earlier, these changes may well have been due to 'incumbent upgrading' or 're-urbanisation' instead, thereby lending to the possibility that the reported incidence of gentrification had been overstated in their study. This deficiency is thus another pressing research gap in the domain that begs attention.

3.4 Gentrification as a Subject of Empirical Analysis

Across the literature, numerous empirical methods have been employed to examine gentrification, including techniques from both sides of the quantitative-qualitative divide. In exploring the significance of research methodologies in the context

of gentrification studies, Barton's (2016) commentary outlined the trade-offs associated with the varying strategies and the conundrums typically faced by researchers in selecting optimal approaches.

In particular, Barton (2016, p.93) highlighted that although qualitative methods were useful in eliciting people's rich, lived experiences to understand gentrification's intangible aspects, such as the changing 'social character' of gentrifying neighbourhoods, they were labour-intensive and time-consuming. Hence, the application of qualitative approaches was mostly restricted to small areas or samples. Moreover, as a greater degree of subjectivity was inherent in qualitative methods, there was a greater risk of inconsistent outcomes being produced.

Contrastingly, quantitative methods offered a more comprehensive means of analysing neighbourhoods, and with the availability of appropriate data, can be implemented to cover extensive areas (Barton, 2016). Additionally, as quantifiable data undergird the workings of quantitative techniques, more consistent results are typically derived. Despite the superiority of quantitative approaches in its ability to objectively analyse gentrification trends and at scale, 2 criticisms were levelled against them by Barton (2016). Firstly, quantitative approaches tended to depend overwhelmingly on census data that were heavily orientated towards socio-economic aspects. Resultantly, quantitative studies were constrained in their perspectives, as they generally steered toward mono-dimensional assessments of gentrification (2016, p.93). Secondly, since census data is typically produced only once per decade and usually published without any personal identifiers, quantitative analyses were inevitably restricted in their capacity to actively monitor gentrification and had no means of differentiating neighbourhoods

which ‘underwent naturally occurring improvements (incumbent upgrading)’ from those that were gentrifying (2016, p.96).

Insofar as the criticisms of quantitative approaches are acknowledged, the recent advent of ‘Big Data’ is a critical pathway that can enable research to overcome its longstanding, methodological weaknesses (Brunsdon and Singleton, 2015, p.322). Specifically, with the seemingly ever-enlarging diversity and availability of data in our current age, opportunities for researchers to venture beyond traditional data sources, like the census, have correspondingly proliferated in magnitude. Researchers have recently begun to recognise such opportunities and now incorporate previously unavailable datasets into their analysis, including administrative, retail, and even crowd-sourced, social data (Hristova et al., 2018; Longley et al., 2018). Gentrification analyses can thus be expected to, and indeed should, persist in leveraging ‘Big Data’ for the betterment of scholarship.

Rising with the dawn of ‘Big Data’, ML is a particular subset of artificial intelligence that has gained traction throughout academia. Unlike classic statistical approaches, ML algorithms possess the capacity to learn from data inputs and adapt its processes to optimise its performance at given tasks without the need for much human manipulation (Witten et al., 2011). Furthermore, due to their set-up, many ML algorithms can handle extensive quantities of high-dimensional data effectively, even those exhibiting issues of multi-collinearity, and are capable of modelling non-linear relationships between components of the dataset which rudimentary linear regression analysis cannot achieve (Witten et al., 2011). In the realm of urban analytics, several

examples showcase ML's impressive ability to extract patterns from 'noisy' spatial data, model and project complex spatial phenomena, such as Wei and Knox (2014).

However, ML's adoption in gentrification studies is still in its infancy and can be improved. Reades et al.'s (2019) study, as introduced earlier, is one of the pioneering examples that exploited ML for the purpose of understanding gentrification. In their paper, a specific ML algorithm called 'Random Forest' (RF) was first applied to model patterns of neighbourhood 'uplift' across London using past census data, and subsequently used to predict potential neighbourhood states by 2021. In terms of performance, the paper revealed that RF outperformed traditional multi-linear regression models by at least '10%' across all evaluation metrics (2019, p.12). Other gentrification studies which have successfully incorporated ML into their workflows include Ilic et al.'s (2019, p.3) use of 'machine mapping' on Google Street View images to identify visible improvements to housing facades as signs of gentrification, and Chermesh et al.'s (2018) utilisation of cluster analysis to tease out gentrifying neighbourhoods across New York.

Considering gentrification's innately spatial nature, spatial methods are well-positioned to complement ML, which are conventionally aspatial in their workings (Georganos et al., 2019), and boost the overall methodological robustness of research studies (Lauren, 2017). Importantly, spatial methods such as Global Moran's I, which will be further discussed under the methodology section, glean from the underlying spatial structures and layer crucial spatial perspectives onto the analysis. Kiely & Bastian's (2019) recent publication paves the way forward in this respect, as they innovatively

combined Geographically-Weighted Regression with ML to predict gentrification in New York.

However, the synergising of ML with spatial analysis is again a vastly untapped opportunity. It is therefore the vision of this paper to capitalise on the advancements in data availability and methodological approaches to bridge the perennial gaps in gentrification research and engender applications that will truly benefit society.

4 Methodology

4.1 Data Collection & Wrangling

The datasets utilised in this study were derived from 3 main sources, namely the ONS, the Greater London Authority (GLA) and the Consumer Data Research Centre (CDRC). The obtained datasets and the steps taken to pre-process them are outlined in the proceeding paragraphs. Data wrangling and downstream analysis were carried out using Python and ArcMap.

4.1.1 ONS' Census and Housing Transactions Data

ONS' data outputs from the 2001 and 2011 censuses were used as the cornerstone for analysing the changing states of London's neighbourhoods. This dataset alone consisted 248 features and covered a potpourri of themes, ranging from those describing the socio-economic make-up of neighbourhoods to its housing characteristics.

A major caveat in concurrently engaging with data from the 2001 and 2011 censuses, however, was the non-identical LSOA boundaries used to aggregate information in each year that rendered direct comparisons of LSOA-level data between both years unachievable. The dissimilarities stemmed from the ONS splitting and merging certain 2001 LSOAs to produce the 2011 version. To overcome this issue, a crucial pre-processing step was undertaken to regularise all data aggregated using the 2001 LSOAs to match the updated 2011 boundaries. A re-aggregation weighting scheme generated by the UK Data Service (2015) proved useful as it gave guidance on how the

2001 LSOAs should be apportioned so as to sync with the 2011 version. An example of the regularisation is shown in *Figure 3*.

Figure 3. Example of Regularising Data Aggregated Using 2001 LSOAs to the Updated 2011 LSOAs

Besides having to regularise census data across years due to inconsistencies in LSOA boundaries, data features within the same year, particularly those represented as absolute counts, required standardisation since LSOAs varied in population size and land area. All absolute counts in the datasets were therefore converted into percentages for inter-LSOA comparability.

Additionally, with the economic value of residential properties and housing turnovers being an integral factor and product of gentrification's cycles respectively, ONS datasets comprising median house prices and counts of residential sales transactions that transpired yearly at the LSOA-level (between 2001-2016) were also

collated for study. The dataset was divided into 2 groups – (1) for data dated in-between the censuses (i.e. 2001-2010), and (2) for those dated 2011-2016.

4.1.2 GLA's Planning Permissions Data

The GLA's London Development Database (LDD) is a data repository containing digital records of planning permissions granted to development projects. Although LDD's data have not appeared to feature much in past studies, they are an excellent resource for urban research given their richness in documenting development projects, including residential (re)development works, around the city. Since the redevelopment/conversion of existing houses and introduction of new-built residential properties in concentrated numbers were tendencies of different neighbourhood change typologies, this LDD dataset was ideal for distinguishing differing trajectories and was hence co-opted for further analysis.

To be analysed suitably, the LDD dataset had to be pre-processed in 3 ways. Firstly, the dataset was scrubbed to remove records which either did not pertain to residential properties or possessed erroneous entries. While the intent was to retain all records from its initial year (i.e. Financial Year (FY) 2004) to FY2011, a visual inspection of the data revealed that planning permissions prior to FY2006 were comparatively sparse and therefore unlikely to be reliable for research. Hence, only records dated between FY2006-FY2011 were kept. By the end of the data cleaning, approximately 25,000 records out of around 70,600 original entries remained. Subsequently, the data was segmented into 2 categories – (1) for planning permissions granted to redevelopment/conversion works on pre-existing properties, and (2) for new-built

residential projects. *Figure 4* shows the statistical frequencies of these categories in their respective years.

Figure 4. Frequencies of Planning Permissions in LDD Dataset

Thirdly, as the original dataset came as point data, they had to be aggregated according to the 2011 LSOA boundaries for alignment with the other datasets. A sample of the raw datapoints is provided in *Figure 5.*

Figure 5. Sample of Raw Datapoints in LDD Dataset

Using the 'Display X-Y Data' and 'Spatial Join' tools in ArcMap, the datapoints were plotted according to their Eastings-Northings and aggregated according to the LSOA they were situated in. The aggregated data was then standardised using the 2006 dwelling unit figures and presented as the rate of planning permissions granted per 1,000 dwelling units. The choropleth maps in *Figures 6-7* depict the geographical variations in the rate of planning permissions granted for redevelopment/conversion and new-built projects respectively.

Figure 6. Planning Permissions for Redevelopment / Conversion of Existing Residential Properties

Figure 7. Planning Permissions for New-Built Residential Developments

4.1.3 CDRC's Population Churn Data

Harnessing fine-grained data from public electoral registers, consumer databases and put together using 'bespoke data linkage techniques' (Lansley et al., 2018), CDRC's population churn dataset contained informative year-on-year estimates of population turnover at the LSOA-scale, which was highly-novel since no other similar dataset existed. With population churn being a direct indicator signalling changes to the incumbent residents within a neighbourhood, this dataset was valuable for analysing gentrification and readily incorporated into the study. The methodology adopted in producing this dataset are elaborated by Lansley et al. (2018), while the digitised population churn maps are viewable at https://maps.cdrc.ac.uk/#/indicators/churn/default/BTTTFFT/10/-0.1500/51.5200/.

As CDRC's dataset was already in the form required for analysis, little pre-processing had to be done, except to segregate data into 2 sets – (1) for data dated between 2001-2010 and (2) for data between 2011-2016. Average churn rates for both time periods were calculated and used for analysis.

The complete suite of variables incorporated into this research is tabulated in *Table 2*.

Table 2. Variables Incorporated in Research

Census Data for both 2001 & 2011

Cat	Variable	Label
Population & Demographics	Population Size	Pop_size
	% of residents aged 0 – 4	age_0_4
	% of residents aged 5 – 7	age_5_7
	% of residents aged 8 – 9	age_8_9
	% of residents aged 10 – 14	age_10_14
	% of residents aged 15	age_15
	% of residents aged 16 – 17	age_16_17
	% of residents aged 18 – 19	age_18_19
	% of residents aged 20 – 24	age_20_24
	% of residents aged 25 – 29	age_25_29
	% of residents aged 30 – 44	age_30_44
	% of residents aged 45 – 59	age_45_59
	% of residents aged 60 – 64	age_60_64
	% of residents aged 64 – 74	age_65_74
	% of residents aged 75 – 84	age_75_84
	% of residents aged 85 – 89	age_85_89
	% of residents aged 90 & above	age_90_over
Household	% of 1-person household with all elderly occupants	1-personHse_aged
	% of 1-person household with occupants of other age groups	1-personHse_other
	% of families with all elderly occupants	1-family_aged
	% of married couple households with no children	1-family_married_noKids
	% of married couple households with dependent children	1-family_married_depKids
	% of married couple households with no dependent children	1-family married_noDepKids
	% of cohabitating couple households with no children	1-family_cohab_noKids
	% of cohabitating couple households with dependent children	1-family_cohab_depKids
	% of cohabitating couple households with no dependent children	1-family _cohab_noDepKids
	% of lone-parent households with dependent children	1-family_lone_depKids
	% of lone-parent households with no dependent children	1-family_lone_noDepKids
	% of other households with dependent children	otherHse_depKids
	% of other households with all students	otherHse_students
	% of other households with all elderly occupants	otherHse_aged
	% of other households with occupants of other aged groups	otherHse_other
	Average household size	avg_hse_size
Place of Birth	% of residents born in the UK	born_UK
	% of residents born in Ireland	born_IRE
	% of residents born in the EU	born_EU
	% of residents born in Europe but outside the EU	born_restEUR
	% of residents born in Africa	born_Africa
	% of residents born in Asia	born_Asia
	% of residents born in North America and the Caribbean	born_NthAm
	% of residents born in South America	born_SthAm
	% of residents born in Oceania	born_Oceania
Ethnicity	% of residents with 'White' ethnicity	eth_white
	% of residents with 'Mixed' ethnicity	eth_mixed
	% of residents with 'Asian' ethnicity	eth_asian
	% of residents with 'Black' ethnicity	eth_black
Religion	% of residents who are Christian	Christian
	% of residents who are Buddhist	Buddhist
	% of residents who are Christian	Hindu
	% of residents who are Jewish	Jewish
	% of residents who are Muslim	Muslim
	% of residents who are Sikh	Sikh
	% of residents who are of other religions	Oth_religion
	% of residents who have no religion	No_religion
	% of residents who did not state their religion	Religion_not_stated
Car Ownership	% of residents with no cars or vans	no_carsVans
	% of residents with 1 car or van	1_carsVans
	% of residents with 2 cars or vans	2_carsVans
	% of residents with 3 cars or vans	3_carsVans
	% of residents with 4 and more cars or vans	4_more_carsVans
Density & Dwelling	Population density	Pop_density
	Number of dwelling units	dwellings_no
	% of dwellings with occupancy ratings of -1 or less	occup_rating
	% of dwellings with central heating	central_heating
	Average number of rooms	avg_rooms
	% of dwellings with usual residents	hsehold_usual_residents

Cat	Variable	Label
Tenure	% of tenures which are owned outright	tenure_owned_outright
	% of tenures which are owned with a mortgage or loan	tenure_owned_ mortLoan
	% of tenures on a shared ownership	tenure_shared
	% of tenures which are rented through councils	tenure_rent_council
	% of tenures which are rented through housing associations	tenure_rent_hsgAssc
	% of tenures which are privately rented	tenure_rent_private
	% of tenures which are rented from other sources	tenure_rent_other
Dwelling Type	% of dwellings which are detached houses	hse_detached
	% of dwellings which are semi-detached houses	hse_semiDetached
	% of dwellings which are terraced houses	hse_terraced
	% of dwellings which are flats	hse_flats
	% of dwellings which are in a converted / shared house	hse_converted_shared
	% of dwellings which in commercial buildings	hse_commercial
Income & Economic Activeness	Median household income	median_income
	% of residents who are economically active	economically_active
Occupation	% of residents in occupational class 1 (Managers, directors and senior officials)	Managers_directors_ senior_occup
	% of residents in occupational class 2 (Professional occupations)	Professional_occup
	% of residents in occupational class 3 (Associate professional and technical occupations)	Associate_professional_ technical_occup
	% of residents in occupational class 4 (Administrative and secretarial occupations)	Administrative_ secretarial_occup
	% of residents in occupational class 5 (Skilled trades occupations)	Skilled_trades_occup
	% of residents in occupational class 6 (Caring, leisure and other service occupations)	Caring_leisure_other_ service_occup
	% of residents in occupational class 7 (Sales and customer service occupations)	Sales_customer_ service_occup
	% of residents in occupational class 8 (Process, plant and machine operatives)	Process_plant_machine_ operatives_occup
	% of residents in occupational class 9 (Elementary occupations)	Elementary_occup
NS-SEC	% of residents in NS-SEC class 1	NS-SEC_1
	% of residents in NS-SEC class 2	NS-SEC_2
	% of residents in NS-SEC class 3	NS-SEC_3
	% of residents in NS-SEC class 4	NS-SEC_4
	% of residents in NS-SEC class 5	NS-SEC_5
	% of residents in NS-SEC class 6	NS-SEC_6
	% of residents in NS-SEC class 7	NS-SEC_7
	% of residents in NS-SEC class 8	NS-SEC_8
	% of residents not classified under NS-SEC	NS-SEC_noClass
Work Conditions	% of male residents in part-time jobs working <15hrs	Males_ptTime_less15hrs
	% of male residents in part-time jobs working 16 – 30hrs	Males_ptTime_16_30hrs
	% of male residents in full-time jobs working 31 – 48hrs	Males_FTime_31_48hrs
	% of male residents in full-time jobs working >49hrs	Males_FTime_more49hr s
	% of female residents in part-time jobs working <15hrs	Females_ptTime_less15 hrs
	% of female residents in part-time jobs working 16 – 30hrs	Females_ptTime_16_30h rs
	% of female residents in full-time jobs working 31 – 48hrs	Females_FTime_31_48h rs
	% of female residents in full-time jobs working >49hrs	Females_FTime_more49 hrs
Qualifications	% of residents with no qualifications	No_qual
	% of residents with level 1 qualifications	Lv1_qual
	% of residents with level 2 qualifications	Lv2_qual
	% of residents with level 3 qualifications	Lv3_qual
	% of residents with level 4 qualifications	Lv4_qual
	% of residents with other qualifications	Other_qual
Travel to Work	% of residents who work from home	toWork_home
	% of residents who travel to work via the underground or light rail	toWork_underground_ light_rail
	% of residents who travel to work via the train	toWork_train
	% of residents who travel to work via the bus	toWork_bus
	% of residents who travel to work via motorcycle	toWork_motorcycle
	% of residents who travel to work via car or van	toWork_carVan
	% of residents who travel to work as a passenger in a car or van	toWork_passenger_ carVan
	% of residents who travel to work via the taxi	toWork_taxi
	% of residents who travel to work via bicycle	toWork_bicycle
	% of residents who travel to work via walking	toWork_walk
	% of residents who travel to work via other means of transport	toWork_other

Data Sources: (ONS, 2016) & (ONS, 2011)

Housing Price & Transactions Data

Cat	Variable	Label
House Prices & Sales	Median house price (in base year)	Median_HsePrice_ BaseYr
	Median house price (in final year)	Median_HsePrice_EndYr
	Average housing sales transactions	avg_hse_sales

Data Source: (ONS, 2018)

London Development Database Data (GLA)

Cat	Variable	Label
Planning Permissions	No. of planning permissions granted for redevelopment / conversion of pre-existing residential properties (per 1,000 dwelling units)	conv_rates
	No. of planning permissions granted for new-built residences (per 1,000 dwelling units)	newblt_rates

Data Source: (GLA, 2019)

Population Churn

Cat	Variable	Label
Population Churn	Average population churn	avg_pop_churn

Data Source: (CDRC, 2018)

4.2 Workflow & Methods

Diverse methods, straddling across statistics, ML and spatial analysis, were chained into a seamless, multi-level workflow with every phase building upon results of the former. *Figure 8* illustrates the workflow diagrammatically while explanations of individual methods are detailed in the following paragraphs.

Figure 8. Overview of Entire Workflow

4.2.1 Stage 1: Scoring LSOAs to Identify their Broad Neighbourhood States

Following the pioneering examples of Reades et al. (2019) and Owens (2012), a handful of neighbourhood variables was stringently selected as proxies for quantifying neighbourhood states, particularly to determine if neighbourhoods had been ascending, declining or stable over a period of time. Specific to the UK's context, Reades et al. (2019) operationalised 4 pairs of variables, including 'household income', 'property sales value', the proportion of residents possessing the highest educational 'qualifications' and those within the leading 'occupational classes', for assessing LSOAs.

While this paper agreed with the first 3 variables and similarly adopted them, it is argued that the last-mentioned variable relating to the top-tier of occupational classes should be supplanted with that of the NS-SEC (i.e. NS-SEC Class 1) instead. Recognising that the NS-SEC went beyond pure employment categories to consider 'socio-economic differences' in the layering of its classifications (ONS, n.d.), the NS-SEC was likely to offer a more holistic reflection of a LSOA's socio-economic conditions and thus better suited for analysing neighbourhood states. *Table 3* provides a tabulation of these 4 variable pairs.

Table 3. Labels of variables used for scoring LSOAs

Pair	*Variables*	*Labels*
1	Median Housing Price (2001)	hseP_01
	Median Housing Price (2011)	hseP_11
2	Median Income (2001)	inc_01
	Median Income (2011)	inc_11
3	% of residents with Level 4 Qualifications & Above (2001)	LV4_01
	% of residents with Level 4 Qualifications & Above (2011)	LV4_11
4	% of residents in NS-SEC Class 1 (2001)	SEC1_01
	% of residents in NS-SEC Class 1 (2011)	SEC1_11

Having established the proxy variables, the next step was to create a Composite Index (CI) that integrated these attributes to score the LSOAs. As it was ideal that the CI was a single value for ease of scoring, Principal Component Analysis (PCA) was carried out, using Python's Scikit-Learn package, to reduce the multiple dimensions embedded within the variables. The component that accounted for the 'largest possible variance in the data' (Jaadi, 2019), also known as the 1st Principal Component (PC1), was adopted as the scores.

Nonetheless, as the variables were on differing scales and contained outliers *(see Figure 9)*, they had to be first re-scaled before being processed by PCA. Robust Scaler was chosen for this task given its superiority in handling outliers during the re-scaling process *(see equation in Figure 10)*.

Figure 9. Dot & Whisker Plots for 4 Proxy Variables (inclusive of insets)

Figure 10. Equation of Robust Scaler

$$\text{Robust Scaler} = \frac{X_i - median\ (x)}{Q_3(X) - Q_1(X)}$$

, where:
X = Variable Set
i = Specific Datapoint i in set
Q = Quantile

Source: (Lin, 2017)

Once CI scores were generated for all LSOAs based on their 2001 and 2011 proxy attributes, the 2 following approaches helped distil whether LSOAs had been in relative ascent, decline or stable between 2001-2011:

a) Changes in Rank

LSOAs were ranked according to their scores for each year and had their rank difference calculated as the change between their 2001 and 2011 rankings. LSOAs with positive rank differences (rank upgrading) that varied >1 standard deviation (SD) from the mean were designated as ascending, while the opposite was true for LSOAs with negative rank differences (rank downgrading) which had a SD of <-1. Similar to Reades et al. (2019, p.13), rank differences within ±1 SD were treated as stable LSOAs exhibiting 'random fluctuation' .

b) Changes in Score

However, a limitation in singularly depending on rank changes to identify neighbourhood states was the bounded nature of ranking systems. LSOAs ranked sufficiently close to the extremities of the rank systems in 2001 were inadvertently restricted in their ability to move further upwards (for originally high-ranked LSOAs) or downwards (for originally low-ranked LSOAs) in rankings even though considerable improvement or downgrading may have occurred by 2011. Consequently, such LSOAs evaded detection and was not flagged as ascending or descending, as they should otherwise have been. To circumvent this issue, a supplementary measure was taken to ascribe LSOAs with score changes within the top 5th percentile as ascending, and those within the bottom 5th percentile as declining. The 5th percentile thresholds were intentionally chosen for prudency in ensuring that only LSOAs with the most significant of score changes were sieved through this process.

4.2.2 Stage 2: Clustering Ascending LSOAs to Differentiate Typologies of Neighbourhood Ascent

Stages 2 and 3 made up a 2-step clustering process which focused on grouping together LSOAs undergoing similar forms of socio-spatial transformations, albeit targeting different aspects of the neighbourhood change schema. In tinkering with the methods, the field of Geodemographics was a useful reference since their analyses, which typically sought to cluster people and places encompassing similar traits, shared broad principles with this study.

In particular, the K-Means clustering technique, which is regularly implemented in Geodemographics, was adopted for discovering clusters within the data. K-Means functions through an iterative approach that partitions datapoints into natural groupings based on similarities in their features (Royall and Wortmann, 2015). Depending on the number of clusters (K) selected, K-number of centroids are planted and iteratively adjusted to fit the datapoints. Based on their nearest centroid at the end of the iteration process, datapoints are grouped into clusters. Silhouette scores were computed for 1-10 clusters to establish the optimal K, while the algorithm was initiated 5,000 times to mitigate against any effects arising from the random initialisation of centroids.

Equally important in cluster analysis was the choice of variables given as inputs for K-Means to define the clusters. Since the aim was to cluster similar neighbourhood change trajectories, change variables representing variations in a feature's values between 2001-2011 were computed to facilitate the clustering. However, it was crucial that only change variables pertinent to the phenomena under study were included, as

non-essential variables might become 'noise' distorting the ideal formation of clusters (Alexiou and Singleton, 2015, p.140). For ease of interpretation, it was also ideal that similarly-themed change variables were merged under broader categories.

For Stage 2, since neighbourhood ascent typologies were the focus, only ascending LSOAs identified in Stage 1 and change variables key for distinguishing between the various forms of neighbourhood ascent, such as population turnover, development works and alterations to a LSOA's socio-demographic composition, were adopted in the clustering.

Notwithstanding the utility of change variables in indicating the magnitude and directionality of neighbourhood transformations, they inevitably provided no contextual background concerning the change which hindered one from discerning the qualitative differences between varying transformation trajectories. Put simply, while change variables could show a significant influx of well-paid, highly-qualified professionals into a neighbourhood, they are unable to convey if the receiving neighbourhood is originally an averagely-priced, working-class neighbourhood or a wealthy, up-scale one. Depending on the receiving neighbourhood's context, the situations could be qualitatively dissimilar and belong to divergent gentrification typologies (e.g. super-gentrification as opposed to mainstream gentrification). Hence, status variables connoting a LSOA's prior conditions (i.e. in 2001 for analysing changes between 2001-2011) from which the changes originated were added to complement the analysis. The variables curated for Stage 2's clustering are tabulated in *Table 4*.

Table 4. Input Variables for Clustering Ascending LSOAs

	Variables	*Labels*	*Composition (where applicable)*
Population Domain (7)			
Change Variables (Difference between 2001 and 2011 values)	Change in % of residents aged 65+	agePop_chg	This variable was aggregated based on the following data: - % of residents aged 65 to 74 - % of residents aged 75 to 84 - % of residents aged 85 to 89 - % of residents aged 90 and over
	Change in % of households with no children	hse_noKids_chg	This variable was aggregated based on the following data: - % of married couple households with no children - % of cohabiting couple households with no children
	Change in % of households with dependent children	hse_depKids_chg	This variable was aggregated based on the following data: - % of married couple households with dependent children - % of cohabiting couple households with dependent children - % of lone-parent households with dependent children - % of other households with dependent children
	Change in % of households with no dependent children	hse_noDepKids_chg	This variable was aggregated based on the following data: - % of married couple households with no dependent children - % of cohabiting couple households with no dependent children - % of lone-parent households with no dependent children
	Change in % of households made up of residents aged 65+	hse_age_chg	This variable was aggregated based on the following data: - % of 1-person household with all elderly occupants - % of families with all elderly occupants - % of other households with all elderly occupants
	Population churn between 2001 & 2011	avg_churn_01_10	-----
Status Variable	% of residents aged 65+ in 2001	agePop_01	-----
Urban Development Domain (2)			
Change Variables (Difference between 2001 and 2011 values)	Rate of planning permissions granted for conversion of existing residential properties (per 1,000 dwellings in LSOA)	conv_rates	-----
	Rate of planning permissions granted for new-built residential properties (per 1,000 dwellings in LSOA)	newblt_rates	-----

To optimise the performance of K-Means, highly-correlated variables had to be first removed, while features with non-normal distributions and on distinctly different scales needed to be transformed and re-scaled suitably (Alexiou and Singleton, 2015).

Similar to the framework utilised by the ONS (2015) in producing the classifications for OAs, various transformation and re-scaling techniques *(see Table 5 for a breakdown)* were employed to permutate the data preparation in 12 ways in order to find the optimal solution. To boost the methodology's rigour, only transformed and re-scaled permutations that derived no highly-skewed data would be picked for actual clustering. Once the clusters from the different permutations were obtained, the Calinski-Harabasz Index, which measures how dense and well-separated clusters were (Caliński and Harabasz, 1974), was used to evaluate the quality of the clusters and the best scoring set was designated as the optimal outcome.

Table 5. Breakdown of Data Transformation and Re-scaling Methods

Purpose	Method	Equation	Description	Source
Data Transformation	Yeo-Johnson Power Transformation	$y_i^{(lambda)} = \begin{cases} [(y_i+1)^{(lmbda)} - 1] \,/\, (lambda), & \text{if lambda} \neq 0, \text{y} \geq 0 \\ \log(y_i+1), & \text{if lambda} \neq 0, \text{y} \geq 0 \\ -[(-y_i+1)^{(2-lmbda)} - 1] \,/\, (2-lambda), & \text{if lambda} \neq 2, \text{y} < 0 \\ -\log(y_i+1), & \text{if lambda} \neq 2, \text{y} < 0 \end{cases}$	▪ Variant of the Box-Cox Transformation that allows for 0 and negative figures in the data	(Yeo and Johnson,2000)
	Cube Root Transformation	$y = \sqrt[3]{x}$	▪ Works well with non-normal distributions and right-skewed data	(Cox, 2001)
	Inverse Hyperbolic Sine Transformation	$y = \log(\sqrt{x^2+1} + x)$	▪ Works well with skewed data and datasets with 0 and negative figures	(Friedline et al., 2015)
	No Transformation	---	▪ No transformation as an option	---
Data Re-scaling	z-score	$\frac{x - mean(x)}{standard\ deviation\ (x)}$	▪ Scales data centred around 0 and using standard deviation intervals	(Lin, 2017)
	Robust Scaler (interquartile range)	$\frac{X_i - median\ (x)}{Q_3(X) - Q_1(X)}$	▪ Scales data by removing median and dividing by inter-quartile range ▪ Robust handling of outliers	
	Robust Scaler (inter-decile range)	$\frac{X_i - median\ (x)}{Q_{90^{th}\ percentile}(X) - Q_{10^{th}\ percentile}(X)}$	▪ Scales data by removing median and dividing by inter-decile range ▪ Robust handling of outliers	

With the optimal set of clusters generated, spatial analysis was done to explore their underlying spatial patterns. In particular, Global Moran's I, which is a statistic that

‘measures spatial autocorrelation based on both feature locations and feature values simultaneously’ (ESRI, n.d.), was used to examine if spatial clustering, dispersion or randomness was evident in each cluster. For computing the Global Moran’s I in ArcMap, the ‘Conceptualisation of Spatial Relationships’ parameter had to be defined – to which, this study adopted a strict definition of a LSOA’s ‘neighbourhood’ and its spatial relationships with other LSOAs, whereby only abutting LSOAs (queen’s contiguity) could interact with it.

Hotspot analysis based on the Getis-Ord Gi* statistic was also implemented to extract the hotspots of gentrifying LSOAs. Essentially, hotspot analysis works by comparing the value of a particular LSOA and that of its neighbours to an expected threshold value. Hotspots are identified when high values are observed across the feature and its neighbours, and at levels surpassing the expected threshold (Anselin, 2019).

4.2.3 Stage 3: Clustering Gentrifying LSOAs to Differentiate Typologies of Gentrification

The empirical procedures implemented in Stage 2 were analogous to that of Stage 3, except for the pool of LSOAs and input variables used to define the clusters. With Stage 3’s goal to cluster the typologies of gentrification, only LSOAs which identified as gentrifying in Stage 2 were extracted for this phase. Additionally, variables which could critically distinguish between the various forms of gentrification, such as those reflecting the nature and extent of socio-economic changes in LSOAs, were used as inputs for the analysis. The variables curated for clustering in Stage 3 are tabulated in *Table 6*.

Table 6. Input Variables for Clustering Gentrifying LSOAs

	Variables	*Labels*	*Composition (where applicable)*
Socio-economic Domain (9)			
Change Variables (Difference between 2001 and 2011 values)	Change in % of residents in the Higher Managerial, Administrative & Professional Occupation category (NS-SEC Class 1)	SEC1_chg	-----
	Change in % of residents in the Lower Managerial, Administrative & Professional Occupation category (NS-SEC Class 2)	SEC2_chg	-----
	Change in % of residents in the Intermediate Occupation category (NS-SEC Class 3)	SEC3_chg	-----
	Change in % of residents in the Small Employers & Own Account Workers category (NS-SEC Class 4)	SEC4_chg	-----
	Change in % of residents in the Lower Supervisory & Technical Occupation category (NS-SEC Class 5)	SEC5_chg	-----
	Change in % of residents in the Semi-Routine Occupation category (NS-SEC Class 6)	SEC6_chg	-----
	Change in % of residents in the Routine Occupation category (NS-SEC Class 7)	SEC7_chg	-----
	% change in median income	inc_chg	-----
Status Variable	Median income in 2001	inc_01	-----
Housing Domain (5)			
Change Variables (Difference between 2001 and 2011 values)	Change in % of residential properties owned by occupants	owned_chg	This variable was aggregated based on the following data: - % of tenures which are owned outright - % of tenures which are owned with a mortgage or loan_pct
	Change in % of residential properties privately rented by occupants	pteRent_chg	-----
	Change in % of residential properties socially rented by occupants	socRent_chg	This variable was aggregated based on the following data: - % of tenures which are rented through councils - % of tenures which are rented through housing associations
	% change in median house prices	hseP_chg	-----
Status Variable	Median house price in 2001	hseP_01	-----

4.2.4 Stage 4: Building a ML model to Classify and Predict Gentrifying LSOAs and their Typologies

With gentrifying LSOAs between 2001-2011 and their associated variants identified through Stages 2 and 3 respectively, Stage 4 focused on building a multi-class classification model to predict future gentrifying LSOAs and their typologies. The model development process comprised 2 parts. The 1st part was designed to train, test and validate a ML model to predict LSOAs in gentrification by 2011 and their corresponding forms. To produce the predictions, the model would be trained to crunch the full range of 2001-2010 data, such as the 2001 Census, house prices and population churn data up till 2010, and 'learn' from the underlying patterns embedded within the outputs of Stages 2 and 3. Once trained and calibrated, the 2nd part of the process would use 2011-2016 data as sole inputs for the model to predict future gentrification scenarios.

A pair of ML algorithms that utilised different approaches in 'learning' from data – RF and AdaBoost – was experimented in this study (Witten et al., 2011). In brief, RF utilises a collection of decision trees that randomly samples the dataset and puts splits of data features through a series of true/false conditions to systematically narrow down to a decision – which in this analysis, were predictions of whether a LSOA was gentrifying, and if so, its typology. The robustness of RF is forged through the 'ensembles' of decision trees created and their ability to vote collectively for the optimal decision (2011, p.357). In contrast to growing an extensive forest of decision trees, AdaBoost, otherwise known as 'Adaptive Boosting', iteratively constructs 'weak classifiers' to model the data and 'learn' from the mistakes of its last iteration to cumulatively produce a 'strong classifier' (2011, p.358). Similar to RF, AdaBoost is highly-robust as it corrects the errors made by 'weak classifiers' at every iteration and takes a

weighted average of all internal classifiers to output a prediction. Both models were developed through Python's imbalance-learn package which specially adapted models to handle data with imbalanced classes.

Although both models were initialised, the idea was to pick only the better-performing base model for advanced calibration and eventually adopt its predictions as this stage's results. The metrics selected for evaluating the models included the weighted F1-score and the geometric mean, which measured the model's ability to make accurate predictions *(see the underlying equations of these metrics in Figure 11)*.

Figure 11. Equations of Scoring Metrics

Metric	Equation		Source
Weighted F1-score	$\text{Score} = 2 \text{ x} \frac{(recall\ x\ precision)}{(recall + precision)}$	where, Recall = Accurate predictions of Gentrifying LSOAs out of total accurate predictions of both Gentrifying and Non-Gentrifying LSOAs Precision = Accurate predictions of Gentrifying LSOAs out of total predictions of Gentrifying LSOAs	(Swalin, 2018)
Geometric Mean	$\text{Score} = \sqrt{recall\ x\ specificity}$	where, Recall = Accurate predictions of Gentrifying LSOAs out of total accurate predictions of both Gentrifying and Non-Gentrifying LSOAs Specificity = Accurate predictions of Non-Gentrifying LSOAs out of total predictions of Non-Gentrifying LSOAs	(Lemaitre et al., 2016)

To build the ideal model, 3 key considerations necessitated attention. Firstly, to ensure robustness in the model-building process, the input dataset given to the model for 'learning' had to be divided into a training and test set. This split helped made sure that a subset of the data which the model had not trained on (i.e. test set) could be used to objectively gauge the model's performance.

Secondly, both RF and AdaBoost contained hyperparameters which needed tuning to optimise their performance. Hyperparameters dictated the functioning of ML algorithms and hence required careful calibration to produce optimal results. *Table 7*

tabulates the key hyperparameters of both models. To comprehensively search for the best combination of hyperparameters, Scikit-Learn's 'GridSearchCV' library was deployed to systematically check all hyperparameter permutations within a pre-defined space.

Table 7. Hyperparameters of Chosen ML Models

ML Model	Hyperparameter	Description	Source
RF	Number of trees in forest	The number of trees controls the learning of the data, whereby having more trees allow for more learning to be done. However, too many trees can lead to overfitting of the model.	(Ben Fraj, 2017)
	Maximum features for splitting of each leaf	This hyperparameter limits the number of features that can used for getting the best split on a leaf node.	
	Criterion	The criteria used for measuring the quality of a split at each leaf node.	
AdaBoost	Number of estimators	This hyperparameters control the number of weak estimators to be trained by the model, whereby more estimators allow for more learning. However, similar to trees in RF, too many estimators can lead to overfitting.	(Navlani, 2018)
	Learning Rate	It controls the weights which are assigned to each weak learner which are iteratively trained on the data.	

Thirdly, besides hyperparameter tuning, feature engineering can also further optimise ML models. While feature engineering entails a broad categorisation of techniques (Witten et al., 2011), an effective method to raise a model's performance was to prune away variables deemed to be insignificant for predicting purposes by the algorithm, after an initial round of modelling. The maximum number of variables that could be pruned was capped at 50% of the input dataset, in line with normative standards. With pruning, only significant variables were crunched by the model for making predictions, which could therefore lead to potentially more accurate outcomes.

Separately, capitalising on the inherently spatial nature of gentrification, another feature engineering technique innovatively implemented by this study was the incorporation of 'geographic covariates' into the models (Bergen and Lindstrom, 2019,

p.2). Acknowledging that ML models are generally aspatial in nature, the integration of relevant geographic covariates, which can be perceived as spatial attributes relating to the spatial patterning of gentrification, enable ML models to recognise and include the phenomenon's spatial structures in their workings. Specifically, results from Stages 2 and 3's spatial analysis will be crafted into the model's geographic covariates.

5 Analysis and Results

5.1 Exploring the States of London's LSOAs

Applying PCA on the 4 proxy variables selected for quantifying neighbourhood conditions, PC1 was extracted and adopted as the CI to score LSOAs based on their states in 2001 and 2011. As evinced in *Figures 12(a)-(b)*, PC1 accounted for a reasonably high proportion of variance within the data (78.4%) while its composition was well-spread across all 4 proxy attributes, albeit house prices contributing a bigger share.

Figure 12(a)-(b). Statistics of Principal Components (PC) Derived From PCA

LSOAs mapped according to their CI scores in 2001 and 2011 are provided in *Figures 13 and 14* respectively. Observing each year individually, the highest-scoring LSOAs in 2001 were concentrated around Central London and along spines that extended out towards the city's north and south-western edges. Conversely, the lowest-scoring LSOAs were typically located within outer boroughs towards the east, including Barking and Dagenham and Havering.

Figure 13. Composite Index Scores of all LSOAs in 2001

Figure 14. Composite Index Scores of all LSOAs in 2011

By 2011, a general uplift in scores was noticeable across large swathes of the city, particularly in previously mediocre-scoring LSOAs situated around the north-west, east and south of London. However, this paper cautions against making the flawed

assumption that all LSOAs exhibiting these broad, absolute score increments could definitively be branded as being in genuine ascent, since most LSOAs throughout London are likely to have undergone some degree of improvements and social upgrading with time, in tandem with the upward-trending house prices and socio-economic conditions that have been generally evident across the city over the decades (Reades et al., 2019). It is nonetheless telling from the map that the highest-scoring LSOAs in 2011 were established around Holland Park, Mayfair, Hampstead Heath and Wimbledon.

Employing the pre-defined criteria to methodically sieve through every LSOAs in London, neighbourhoods in ascent, decline and steady states were identified and mapped in *Figure 15*. Insofar as labels such as 'ascent' and 'decline' were used, it is nonetheless clarified that these labels should be interpreted in relative, and not absolute, terms since neighbourhood states were determined as function of comparative differences in LSOAs' CI scores and ranks. Particularly with knowledge that absolute uplifts have been the general tendency across most of London, descending LSOAs, for instance, need not necessarily be regressing in actual terms but were simply improving at a pace below the norm.

Figure 15. Neighbourhood States of all LSOAs

Between 2001-2011, 732 LSOAs were highlighted as ascending, 619 LSOAs in decline, whereas the remaining 3,484 LSOAs were considered stable. While LSOAs in ascent and decline were dotted throughout the city, ascending LSOAs tended toward locations in Central and East London while a denser congregation of declining LSOAs was situated toward the western peripheries. The top 5 boroughs consisting the largest proportion of ascending and declining LSOAs are featured in *Table 8.*

Table 8. Top 5 Boroughs with Highest Share of LSOAs in Ascent and Decline

Top 5 boroughs with highest share of LSOAs in ascent				Top 5 boroughs with highest share of LSOAs in decline			
Borough	*Total LSOAs*	*LSOAs in ascent*	*Share*	*Borough*	*Total LSOAs*	*LSOAs in decline*	*Share*
Kensington and Chelsea	103	55	53.40%	Ealing	196	69	35.20%
Westminster	128	46	35.94%	Hounslow	142	37	26.06%
Newham	164	55	33.54%	Brent	173	45	26.01%
Hackney	144	46	31.94%	Croydon	220	48	21.82%
Redbridge	161	49	30.43%	Kingston upon Thames	98	21	21.43%

5.2 Defining Contextualised Neighbourhood Typologies

In line with the vision of operationalising sharper and more targeted neighbourhood change categorisations, dedicated rounds of cluster analysis were undertaken to unearth the varying typologies of neighbourhood ascent and gentrification specific to London's urban landscape.

5.2.1 Typologies of Neighbourhood Ascent

To recap, the selection of input variables for the cluster analysis had to be first checked for multi-collinearity then suitably transformed and re-scaled in order to optimise the process. Benchmarking highly-correlated variables as those with correlation coefficients beyond ±0.8, the correlation heatmap in *Figure 16* surfaced 3

variables that fell into this category – *'youngPop_chg'*, *'agePop_chg'* and *'hse_age_chg'*. Since 2 of the 3 variables – *'agePop_chg'* and *'hse_age_chg'* – were related to the ageing population changes, only one of them – *'agePop_chg'* – was retained to rectify the issue of multi-collinearity. *'youngPop_chg'* and *'hse_age_chg'* were consequently excluded from further analysis.

Figure 16. Correlation Heatmap for Input Variables in Stage 2

With respect to the data transformation, the Cube Root, Yeo-Johnson Power and Inverse-Hyperbolic Sine transformations successfully produced datasets that contained no heavily-skewed variables bearing a skewness score beyond ±1. *Figure 17* provides a sample of the underlying data distributions before and after transformation.

Figure 17. Underlying Distributions of the Data Before and After Transformation

Each of the 3 transformed datasets was subsequently re-scaled using 3 different scalers and their permutations were individually used as inputs for clustering. Silhouette analysis, which was conducted to determine the optimal K for conducting K-Means, showed that 3 clusters were ideal *(see Figure 18)*. K-Means was hence implemented on all 9 dataset permutations using the parameter *k*=3, and the quality of their derived clusters was measured with the Calinski-Harabasz Index *(see scores in Figure 19)*.

Figure 18. Results of Silhouette Analysis to Determine Optimal Number of Clusters (K) in Stage 2

No. of Clusters	Silhouette Scores
2	0.30957
3	0.32272
4	0.17312
5	0.18406
6	0.16823
7	0.17071
8	0.16203
9	0.15251
10	0.15271

Figure 19. Calinski-Harabasz Index Scores for Clusters Derived Through Different Dataset Permutations (Stage 2)

Evidently, the permutation transformed using Yeo-Johnson Power and scaled with Robust Scaler (inter-decile range) edged ahead in performance. The 3 clusters catalysed through this dataset permutation and their corresponding profiles are interpreted in the upcoming pages.

Cluster: Gentrification

Figure 20. Characteristics of Gentrification Cluster

Variables	conv_ rates	newblt_ rates	avg_churn _01_10	agePop_ chg	hse_ noKids_ chg	hse_ depKids_ chg	hse_ noDepKids _chg	agePop_01
Values in actual terms	+5.36 / 1,000 dus	+3.08 / 1,000 dus	+2.47%	+0.18%	+0.86%	+0.82%	+0.74%	10.49%

As shown in *Figure 20* which plots the average values of this cluster's centroid against the means of the entire dataset, LSOAs comprised in this cluster had distinctively above-average population churn rates and planning permissions granted for redevelopment/conversion projects. These values also far exceeded the figures seen in the remaining clusters. The about-average growth of households at all stages of the lifecycle signalled that diverse types of households were coming into these LSOAs, though inclined towards older demographics given the higher-than-average increases of the ageing population.

The high intensities of population turnover and works involving the re-adaptation of existing residences, potentially for new inhabitants, were hence indicative of gentrification where incumbent residents had been displaced and existing houses revamped to suit the needs of gentrifiers.

Cluster: Incumbent Upgrading

Figure 21. Characteristics of Incumbent Upgrading Cluster

Variables	conv_ rates	newblt_ rates	avg_churn _01_10	agePop_ chg	hse_ noKids_ chg	hse_ depKids_ chg	hse_ noDepKids _chg	agePop_01
Values in actual terms	+0.77 / 1,000 dus	+1.39 / 1,000 dus	+2.27%	-1.03%	-1.07%	+3.15%	+0.87%	14.39%

With its centroid averages presented in *Figure 21*, this cluster contained LSOAs which exhibited below-average population churn rates as well as planning permissions granted for redevelopment/conversion works and new-built housing. Instead, conspicuous increases were seen in the percentage of households with children (both dependent and non-dependent) which potentially implied a growth of households in the middle to latter stages of their lifecycles. LSOAs in this cluster were relatively older in 2001 and experienced average changes to its ageing population by 2011.

Considering these trends, this cluster resonated with the incumbent upgrading typology, wherein existing residents experienced in-situ uplifts in their SES and were

progressing along the family lifecycle as they stayed in-place over time. Substantial levels of population turnover or direct displacement were therefore untypical of incumbent upgrading, which was reflected through this cluster's nominal population churn and planning permissions granted for redeveloping/converting existing houses.

Cluster: Re-urbanisation

Figure 22. Characteristics of Re-urbanisation Cluster

Variables	conv_ rates	newblt_ rates	avg_churn _01_10	agePop_ chg	hse_ noKids_ chg	hse_ depKids_ chg	hse_ noDepKids _chg	agePop_01
Values in actual terms	+1.58 / 1,000 dus	+5.89 / 1,000 dus	+2.04%	-5.00%	+2.65%	-5.90%	-2.52%	10.57%

Juxtaposed against the earlier clusters, this cluster's centroid averages, as displayed in *Figure 22*, demonstrated that its LSOAs had distinctly higher rates of planning permissions granted for new-built residential developments and growths in the proportion of households with no children. Coupled with losses in the percentage of ageing residents (-5% in actual terms), these incoming households were likely of younger demographics and at relatively early stages of the family lifecycle. The lower-than-average population churn and planning permissions for redevelopment/conversion of existing houses signified that population turnover and the direct displacement of incumbent residents were not likely prevalent.

The influx of new-built developments and young families, combined with the attendant lack of population turnover, were consistent with the re-urbanisation typology where regeneration efforts were commonly made to introduce new developments and attract new populations. Despite not entailing direct displacement, LSOAs in this cluster should be monitored by city officials given their susceptibility to exclusionary displacement, as explained in the literature review.

5.2.2 Typologies of Gentrification

The data pre-processing for the 2nd phase of clustering revealed 2 highly correlated variables, *'inc_01'* and *'hseP_01'*, which marginally crossed the threshold of 0.8 *(see Figure 23)*. However, as both variables were status indicators only to provide context to the observed changes, and were essential for pinpointing the different gentrification typologies downstream, a considered decision was made to retain them.

Figure 23. Correlation Heatmap for Input Variables in Stage 3

As for the data transformation, only the Yeo-Johnson Power transformation yielded a dataset containing no highly-skewed variables this round *(see Figure 24 for the underlying distributions of the before and after transformation)* and was put through for re-scaling and clustering.

Figure 24. Underlying Distributions of the Data Before and After Transformation

Underlying Data Distributions Prior to Transformation

Underlying Data Distributions after Yeo-Johnson Power Transformation

Based on the silhouette analysis *(see Figure 25)*, 2 or 3 clusters appeared to be optimal for this dataset, albeit *k=2* having a very slight edge. Nonetheless, as having 3 clusters was bound to be more informative than simply 2 clusters, and since the trade-off in terms of silhouette scores was reasonably minute, *k=3* was eventually adopted for catalysing the cluster analysis.

Figure 25. Results of Silhouette Analysis to Determine Optimal Number of Clusters (K) in Stage 3

No. of Clusters	Silhouette Scores
2	0.53615
3	0.51797
4	0.45786
5	0.49240
6	0.48585
7	0.43735
8	0.43662
9	0.39954
10	0.39831

Clusters derived from the dataset permutations were again measured for their quality using the Calinski-Harabasz Index. Judging from the results *(see Figure 26)*, 2 permutations were tied at top place – those re-scaled using (1) z-scores and (2) robust scaler (inter-decile range). It was confirmed that the clusters obtained through both dataset permutations were identical, and clusters from the *'yeojohnson_zscore'* permutation are outlined accordingly in the proceeding pages.

Figure 26. Calinski-Harabasz Index Scores for Clusters Derived Through Different Dataset Permutations (Stage 3)

Cluster: Super-Gentrification

Figure 27. Characteristics of Super-Gentrification Cluster

Variables	SEC1_ chg	SEC2_ chg	SEC3_ chg	SEC4_ chg	SEC5_ chg	SEC6_ chg	SEC7_ chg
Values in actual terms	+2.73%	+4.09%	+1.48%	+2.36%	-0.14%	+0.27%	+0.27%

Variables	hseP_chg	inc_chg	owned_ chg	socRent_ chg	pteRent_ chg	inc_01	hseP_01
Values in actual terms	+163.9%	+46.65%	-2.78%	-2.01%	+2.05%	£37.24K	£352.5K

Viewing *Figure 27* that depicts the cluster's centroid averages, the LSOAs in this cluster could be inferred as being originally wealthy and expensive neighbourhoods, given their higher-than-average income levels and house prices in 2001. By 2011, the proportion of residents ascribed with the NS-SEC 1 and 2 categories increased notably, though only moderate percentage changes in income and house prices were registered. It is nonetheless highlighted that these percentage changes in income and house prices were based on fairly large denominator values (i.e. 2001 figures), and if computed in absolute terms, their growths by 2011 would in fact be relatively substantial. Above-

average changes were also seen in the proportion of owned and socially-rented houses, whereas changes in private rental tenures dropped below average.

These trends, particularly the influx of residents from the top NS-SEC tiers into already affluent LSOAs with expensively-priced residences, were reminiscent of the super-gentrification typology wherein reasonably well-off, middle-class neighbourhoods were increasingly taken over by incomers possessing even higher SES and deeper capital resources.

Cluster: Marginal Gentrification

Figure 28. Characteristics of Marginal Gentrification Cluster

Variables	SEC1_ chg	SEC2_ chg	SEC3_ chg	SEC4_ chg	SEC5_ chg	SEC6_ chg	SEC7_ chg
Values in actual terms	+1.30%	+0.65%	+1.33%	+3.05%	+0.13%	+1.47%	+1.45%

Variables	hseP_chg	inc_chg	owned_ chg	socRent_ chg	pteRent_ chg	inc_01	hseP_01
Values in actual terms	+181.5%	+45.23%	-7.34%	-2.05%	+8.48%	£27.14K	£124.3K

Contrastingly, this cluster's centroid averages *(see Figure 28)* denoted that the LSOAs here had below-average incomes and house prices in 2001. Moreover, by 2011, it was residents from the NS-SEC 4-7 tiers, instead of the top categories, that were expanding at above-average rates. Nonetheless, house prices seemed to have also experienced above-average growth by 2011. In terms of tenures, houses on private and social rents both witnessed above-average changes, as ownership rates receded.

From this vantage point, this cluster resembled the marginal gentrification typology whereby incomers who did not conform to the typical profile of an affluent,

middle-class gentrifier increasingly took over affordably-priced neighbourhoods that appealed to them.

Cluster: Mainstream Gentrification

Figure 29. Characteristics of Mainstream Gentrification Cluster

Variables	SEC1_ chg	SEC2_ chg	SEC3_ chg	SEC4_ chg	SEC5_ chg	SEC6_ chg	SEC7_ chg
Values in actual terms	+3.81%	+4.95%	+1.26%	+1.87%	+0.26%	-0.85%	+0.27%

Variables	hseP_chg	inc_chg	owned_ chg	socRent_ chg	pteRent_ chg	inc_01	hseP_01
Values in actual terms	+143.6%	+54.27%	-2.00%	-11.36%	+11.36%	£25.13K	£139.1K

Consisting 45 LSOAs, this cluster, akin to marginal gentrification, starts off with relatively lower income levels and house prices in 2001 *(see centroid averages in Figure 29)*. However, unlike the preceding cluster, this cluster registered well-above average gains in income and percentage of high SES residents – from NS-SEC 1 and 2 – by 2011. With reference to tenurial shifts, this cluster experienced higher-than-average changes in housing ownership and those on private rentals, whereas socially-rented housing rapidly declined (-11.36% in actual terms). This could be suggestive of acute council housing losses in these LSOAs.

Consequently, the steep inflows of residents from the top NS-SEC tiers into originally inexpensive and non-wealthy neighbourhoods, coupled with the rising income levels and diminishing stock of socially rented housing, were hallmarks of mainstream gentrification whereby traditionally working-class communities in affordable neighbourhoods were displaced by people of higher SES and which in turn propelled rising income levels in these areas.

5.3 Uncovering Spatial Patterns

Following the cluster analyses, the respective typologies were mapped and analysed for patterns that underpinned their spatial manifestations.

5.3.1 Typologies of Neighbourhood Ascent

The geographical distributions of the 3 neighbourhood ascent typologies are mapped in *Figure 30*.

Observing their geographies, some degree of spatial autocorrelation was suspected to exist amongst gentrifying LSOAs as they were typically located in London's inner boroughs, albeit some exceptions along the city's fringes. Results from the Global Moran's I test *(see Figure 31)* confirmed that gentrifying LSOAs were spatially clustered at statistically-significant levels (p-value <0.01) under a queen's contiguity configuration. This result, by extension, pointed towards the high possibility that gentrification's 'diffusion' effects were at play (Redfern, 1997, p.1335), wherein the spread of gentrifying LSOAs was catalysed through directly-abutting neighbourhoods.

Figure 30. LSOAs in Typologies of Neighbourhood Ascent

Figure 31. ArcMap Outputs for Spatial Autocorrelation Test (Global Moran's I)

Given the z-score of 29.4214361451, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Although LSOAs experiencing re-urbanisation were similarly situated around the inner boroughs, they were centred around central-eastern boroughs, such as Tower Hamlets and Hackney. Conversely, LSOAs identified as undergoing incumbent upgrading were generally sited within the ring of London's outer boroughs.

The hotspots of gentrifying LSOAs were elicited through the Getis-Ord Gi* tool, and results *(see Figure 32)* substantiated the claim that denser clusters were located around the central-east, in boroughs like Westminster, Chelsea and Kensington and Camden. Sizeable hotspot islands were also observed in specific parts of Newham, Lewisham, Southwark and Merton.

Figure 32. Gentrification's Hotspots

5.3.2 Typologies of Gentrification

Figure 33 maps out the geographies of the 3 gentrification typologies. Evidently, super-gentrifying LSOAs appeared to manifest strongly in a collection of large LSOAs along the city's peripheries and Central London, including places where super-gentrification has been documented such as Richmond Avenue of Barnsbury (Butler and Lees, 2006) and Portland Road (Moore, 2012). Contrastingly, LSOAs in marginal gentrification gravitated towards boroughs in East London, such as a constellation of LOSAs bunched together east of Stratford. Spatial patterns for LSOAs experiencing mainstream gentrification were however less apparent, though they were mostly contained within London's inner boroughs.

Global Moran's I confirmed that all 3 gentrification typologies exhibited spatial clustering at statistically-significant levels, while Getis-Ord Gi* extracted the hotspots for each typology (see *Figures 34-36)*.

Figure 33. LSOAs in Typologies of Gentrification
(inclusive of insets for Kensington and Chelsea and Islington)

Figure 34. Spatial Autocorrelation Test Results and Mapping of Super-Gentrification's Hotspots

Spatial Autocorrelation Test for **Super-Gentrification**

Moran's Index	z-score	p-value	Conclusion
0.326009	38.594145	0.0000000	Given the z-score of 38.594145, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

Hotspot Analysis
for **Super-Gentrification**

Figure 35. Spatial Autocorrelation Test Results and Mapping Marginal Gentrification's Hotspots

Spatial Autocorrelation Test for **Marginal Gentrification**

Moran's Index	z-score	p-value	Conclusion
0.123745	14.642906	0.0000000	Given the z-score of 14.642906, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

Hotspot Analysis
for **Marginal Gentrification**

Figure 36. Spatial Autocorrelation Test Results and Mapping Mainstream Gentrification's Hotspots

Spatial Autocorrelation Test for **Mainstream Gentrification**

Moran's Index	z-score	p-value	Conclusion
0.076428	9.119292	0.0000000	Given the z-score of 9.119292, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

Hotspot Analysis
for **Mainstream Gentrification**

Hotspots for super-gentrifying LSOAs were mainly found in Westminster, Kensington and Chelsea boroughs, Hampstead Heath and Wimbledon as well as Cockfosters and Totteridge in the north. Despite observing fewer hotspots for marginally gentrifying LSOAs, there were sizeable concentrations in Newham's north-east, Peckham and Telegraph Hill. Mainstream gentrification's hotspots were less contiguous and existed as small islands, which were generally situated in the inner boroughs north of the Thames as well as places abutting the river's southern banks in Southwark and Woolwich.

5.4 Predicting Future Neighbourhood Trajectories

ML algorithms were trained on the observed trends and spatial patterns of neighbourhood ascent and gentrification that unfolded between 2001-2011, with the aim of predicting which LSOAs will gentrify in the near future and their corresponding typologies. Results from the model-building process are outlined below.

5.4.1 Model to Predict Future Gentrifying LSOAs

With the full range of 2001-2010 neighbourhood variables serving as inputs, 2 separate base models developed using RF and AdaBoost algorithms were trained, calibrated and tested on the outputs of the earlier cluster analysis that discriminated between the various forms of neighbourhood ascent. The best hyperparameter combination for both models extracted through GridSearchCV are summarised in *Table 9*.

Table 9. Best Hyperparameter Combinations Extracted from GridSearchCV

Random Forest (RF)		*AdaBoost*	
Hyperparameter	*Best Candidate*	*Hyperparameter*	*Best Candidate*
No. of Trees	*600*	*No. of Estimators*	*2800*
Maximum Features	*Auto*	*Learning Rate*	*0.1*
Criterion	*Entropy*		

The tuned models were then applied on the test set to predict gentrifying LSOAs in 2011. Predictions were compared against the 'true' labels from the earlier cluster analysis and results are presented in the form of a confusion matrix *(see Figure 37)*. Comparatively, RF was the better-performing model as it predicted both gentrifying and non-gentrifying LSOAs with balance and good accuracy (≥ 80% for both classes).

Although AdaBoost predicted more non-gentrifying LSOAs correctly (96%), it came with significantly larger false-negative errors wherein gentrifying LSOAs were mis-classified as non-gentrifying (53%).

Figure 37. Base Models' Prediction Accuracy on Test Set

Given its superior performance, insights pertaining to the most important variables for predicting gentrifying LSOAs were extracted from the RF model. As over 120 variables were used in the modelling, for brevity, variables were grouped according to their overarching categories, and only the top-10 most important ones are labelled in *Figure 38*. Median house prices in the last known year before the prediction timeline

featured as most important in predicting gentrifying LSOAs, while socio-economic-related variables, such as NS-SEC, Occupation and Work Conditions, collectively constituted almost half of the top-10 list.

Pruning was subsequently done to remove variables which the model deemed insignificant for prediction. Based on tests conducted to determine the optimal quantum of insignificant variables for pruning, it was found that removing the 57 bottom-ranked variables was most ideal for enhancing the model's prediction capabilities *(see score peaks in Figure 39)*.

Figure 38. Variable Importance for Predicting Gentrifying LSOAs

Figure 39. Model's Results at Every Iteration of Variable Pruning (Stage 2)

The 57 variables that were pruned are listed in *Table 10*:

Table 10. Least Important Variables Pruned from Model (Stage 2)

Rank	Variable	Rank	Variable
72	Oth_religion	101	toWork_taxi
73	avg_hse_sales	102	NS-SEC_2
74	hsehold_usual_ residents	103	toWork_bus
75	Skilled_trades_occup	104	Religion_not_stated
76	Males_ptTime_less15hrs	105	NS-SEC_4
77	tenure_owned_outright	106	age_15
78	Professional_occup	107	toWork_motorcycle
79	Hindu	108	4_more_carsVans
80	1-family_cohab_depKids	109	eth_mixed
81	1-family_married_noKids	110	toWork_walk
82	toWork_bicycle	111	Associate_professional_technical_occup
83	NS-SEC_7	112	Sales_customer_service_occup
84	Jewish	113	age_60_64
85	born_restEUR	114	hse_detached
86	age_0_4	115	age_16_17
87	NS-SEC_1	116	age_30_44
88	1-family_lone_depKids	117	born_Asia
89	eth_asian	118	tenure_rent_other
90	NS-SEC_6	119	hse_commercial
91	age_5_7	120	tenure_shared
92	age_8_9	121	occup_rating
93	Muslim	122	otherHse_students
94	born_SthAm	123	toWork_other
95	Buddhist	124	otherHse_aged
96	No_religion	125	Sikh
97	Caring_leisure_other_service_occup	126	1-family_cohab_noDepKids
98	age_10_14	127	avg_rooms
99	1-personHse_other	128	avg_hse_size
100	2_carsVans		

The pruned model was applied not only to the test set, but the entire dataset as well, to obtain a sense of how the model had improved and its overall performance respectively. Based on the results in *Figure 40*, the pruned model was more effective than its unpruned version in predicting gentrifying LSOAs in the test set, and achieved very encouraging results when predicting on the entire dataset (i.e. 95% and 83% accuracy in predicting gentrifying and non-gentrifying LSOAs respectively).

Figure 40. Pruned Models' Prediction Accuracy on Test Set and Entire 2001-2010 Dataset

To further optimise the model, geographic covariates which layer on spatial perspectives and can enable the model to recognise the spatiality of gentrification's manifestations were incorporated in 2 ways. Firstly, mapping the pruned RF model's prediction errors geographically, it was observed that errors mostly laid within the boundaries of London's inner boroughs and Richmond upon Thames *(see Figure 41)*. Distances between each LSOA and these boundaries were calculated and co-opted as the 1st geographic covariate for the model to 'learn' the spatial structures of its errors and correct for them.

Figure 41. Model's Errors in Predicting Gentrifying LSOAs (Before Pruning)

Secondly, since spatial clustering patterns amongst gentrifying LSOAs were evinced through Global Moran's I, the 2nd geographic covariate was designed to be an approximate for these clusters. To produce this covariate, PCA was applied to the top-10 predictors of gentrifying LSOAs, which was previously extracted from the base model. PC1 containing the largest share of variances in the data was extracted and adopted as a crude indicator of a LSOA's likelihood to gentrify.

A sanity check was done by comparing PC1's scores against the locations of gentrified LOSAs. Based on maps in *Figure 42*, it was verified that high PC1 scoring areas generally resonated well with gentrification's geographies. Hence, to replicate the observed clustering patterns and produce the 2nd geographic covariate, each LSOA was attributed with the sum of PC1 scores across all its neighbours defined under the queen's contiguity matrix (i.e. abutting LSOAs). *Figure 43* provides an illustration showing how the score assignment worked.

Figure 42. Comparison of (a) PC1 Scores and (b) Locations of Gentrifying LSOAs

(a) PC1 Scores

(b) Locations of Gentrifying LSOAs

Figure 43. Illustration of the Score Assignments to Replicate the Clustering Effect

After including both geographic covariates, the RF model was tested again on the entire dataset. The results *(see Figure 44)* were testament to the prowess of geographic covariates as the model now predicted gentrifying LSOAs in 2011 perfectly and were also highly-accurate in predicting non-gentrifying LSOAs (84%). Notwithstanding the excellent results, the analysis acknowledged that predictions for non-gentrifying LSOAs incurred a 16% error margin which equated to 720 LSOAs. This error margin was therefore a factor that needed to be considered for predictions churned by this model.

Figure 44. Accuracy of Pruned Model with Geog. Covariates Predicting on Entire 2001-2010 Dataset

Finally, by applying the model to the 2011-2016 dataset, predictions on LSOAs that could potentially gentrify in future were produced and mapped in *Figure 45.*

Figure 45. Model's Predictions on Future Gentrifying LSOAs

Gentrifying LSOAs in future can potentially be expected to cluster within the confines of London's inner boroughs, Richmond upon Thames and Haringey. Juxtaposed against the gentrifying LSOAs of 2011, gentrification processes along northern parts of Barnet and Enfield are likely to continue, whereas ongoing episodes of gentrification along London's eastern, western and southern fringes are forecasted to halt in the coming years. Nevertheless, as mentioned earlier, readers will need to be mindful that these projections are inadvertently inflated by errors incurred by the model when predicting non-gentrifying LSOAs.

5.4.2 Model to Predict Typologies of Future Gentrifying LSOAs

The process for building a model to predict the typologies of future gentrifying LSOAs closely mirrored that of the earlier phase, and hence the results are succinctly elaborated in this sub-section.

By way of tuning and testing both RF and AdaBoost base models to predict the varying typologies of gentrification for gentrifying LSOAs in 2011, the RF model outperformed AdaBoost and was thus adopted for further analysis *(see results in Figure 46)*.

Figure 46. Base Models' Optimal Hyperparameters and Prediction Accuracy on Test Set

RF Hyperparameter	No. of Trees	Max. Features	Criterion
RF Best Candidate	230	Log2	Entropy

AdaBoost Hyperparameter	No. of Estimators	Learning Rate
AdaBoost Best Candidate	90	0.1

The most important variables for predicting the diversified typologies of gentrification were extracted from the RF model, which revealed that NS-SEC variables as dominant within the top-10 *(see Figure 47)*.

Figure 47. Variable Importance for Predicting Typologies of Gentrifying LSOAs

To prune variables, tests established that removing the 53 least important variables optimised the model's predictions. Test results and the list of pruned variables are indicated in *Figure 48* and *Table 11* respectively.

Figure 48. Model's Results at Every Iteration of Variable Pruning (Stage 3)

Table 11. Least Important Variables Pruned from Model (Stage 3)

Rank	Variable	Rank	Variable
76	toWork_walk	103	Other_qual
77	born_Asia	104	toWork_bicycle
78	Pop_size	105	Males_ptTime_less15hrs
79	Sikh	106	age_65_74
80	toWork_train	107	hse_commercial
81	Females_ptTime_16_30hrs	108	eth_asian
82	1-personHse_aged	109	age_20_24
83	born_NthAm	110	1-family_married_depKids
84	central_heating	111	Caring_leisure_other_service_occup
85	Christian	112	born_Oceania
86	toWork_motorcycle	113	age_85_89
87	age_10_14	114	otherHse_aged
88	age_75_84	115	Professional_occup
89	age_45_59	116	1-family_cohab_noDepKids
90	otherHse_depKids	117	age_25_29
91	Associate_professional_technical_occup	118	age_60_64
92	1-family_cohab_noKids	119	otherHse_other
93	born_Africa	120	age_30_44
94	tenure_shared	121	Females_ptTime_less15hrs
95	tenure_rent_private	122	otherHse_students
96	1-family_lone_noDepKids	123	toWork_other
97	avg_rooms	124	age_5_7
98	tenure_rent_other	125	avg_pop_churn
99	No_religion	126	born_UK
100	Pop_density	127	born_IRE
101	Buddhist	128	avg_hse_size
102	dwellings_no		

Geographic covariates were subsequently formulated and added into the model, in a manner analogous to the previous modelling process. However, as the spatial configurations of this model's erroneous predictions were noticed to reside mainly within the inner boroughs *(see Figure 49)*, the 1st geographic covariate for this phase was tweaked to be based on the distance between each LSOA and the inner borough boundaries.

Figure 49. Model's Errors in Predicting LSOAs' Gentrification Typologies

The performances of the model at every step of its evolution are provided in *Figure 50*. The pruning of insignificant variables and inclusion of geographic covariates clearly elevated the model's prediction capacities from its base state. Resultantly, the final version of this multi-class classification model perfectly predicted LSOAs undergoing mainstream gentrification in 2011, while simultaneously attaining highly-accurate predictions for super-gentrification (90%) and marginal gentrification (79%) as well.

Figure 50. Accuracy of Model at Every Stage of the Refinement Process

The plethora of 2011-2016 variables was eventually fed into the model to predict the specific typologies of future gentrifying LSOAs. According to the predictions mapped

in *Figure 51*, super-gentrifying trends will potentially retain a stronghold over LSOAs in central-western London, around Hampstead Heath, Richmond Park and the northern edges of Barnet and Enfield. Previously isolated islands of super-gentrification near Chiswick, Clapham South and Dulwich are nonetheless anticipated to expand. LSOAs experiencing marginal gentrification in future are likely to stay in East London, although potentially becoming more extensive around Brockley and Kensal Green. Separately, future mainstream forms of gentrification are predicted to be domineering within London's inner boroughs and north of the Thames, in boroughs such as Camden, Islington and Hackney.

Figure 51. Model's Predictions on the Typologies of Future Gentrifying LSOAs

6 Discussion

6.1 Contextualising the Results

As a dynamic region at the forefront of our rapidly evolving world, neighbourhood change has been rife across London's urban landscape between 2001-2011. As put forward by our analysis, the 2 broad processes of neighbourhood change – neighbourhood ascent and decline – collectively accounted for 1,351 LSOA or almost 30% of all LSOAs in the city, which concurred with Reades et al. (2019) whose figures were in the same region. No borough, except for the City of London, was spared from these processes, thereby underscoring the extensiveness of socio-spatial transformations happening in London.

Insightful as this quantification of neighbourhood change may be, it is but a starting point for delving into gentrification, where robust research is most urgently needed. As neighbourhood change, is a broader category encompassing gentrification as one of its sub-typologies, it requires filtering down for the purpose of unambiguously identifying gentrifying neighbourhoods. Population churn and planning permissions data, as proxy indicators for population turnover and direct displacement, were shown our analysis to be crucial in the filtration process.

The 2 non-gentrifying forms of neighbourhood ascent found active in London – re-urbanisation and incumbent upgrading – though not the prime focus of this paper – in themselves offer interesting stories of neighbourhood uplift and urban rejuvenation occurring in the city, ostensibly without causing direct displacement. For instance, the London Docklands, mined through our cluster analysis as an area predominantly

undergoing re-urbanisation *(see Figure 52)*, is described by Butler (2007) as a poignant example that pulls together narratives of de-industrialisation and urban renewal, as its disused ports were thoroughly re-created into the new residential areas existing today.

Figure 52. Map Showing Outline of London Docklands and the Typologies of Neighbourhood Ascent

Pivoting quickly to gentrifying LSOAs as our main focus, this typology comprised around 15% of all LSOAs in London; affecting over half-a-million residents. Percolating through traditionally richer, upscale West London boroughs like Kensington and Chelsea and Westminster, as well as boroughs in East London that are conventionally viewed as working-class and less desirable, gentrification in London can literally be described as knowing no boundaries. Such observations are substantiated by other researchers, ranging from Butler and Lees' (2006) study that exposed gentrification's workings within affluent neighbourhoods in Barnsbury to Butler et al.'s (2013) empirical examination of gentrification in the deprived parts of East London. These contrasting, and indeed

almost conflicting, urban backdrops against which gentrification has materialised clearly hint toward its existence in variegated forms – to which our study has shown that super-gentrification, marginal gentrification and mainstream gentrification were simultaneously taking place in London and are unlikely to diminish in the near future according to our model's predictions.

Taken together, the contextualised typologies of neighbourhood ascent and gentrification uncovered in London could be positioned under the overarching schema of neighbourhood change in the manner shown in *Figure 53*.

Figure 53. Conceptual Positionings of Neighbourhood Ascent and Gentrification Typologies under the schema of Neighbourhood States and Change

Similar to the Docklands, each locale exhibiting any of gentrification's typologies will almost certainly contain rich accounts detailing the entanglements between place particularities and history which can help contextualise and enable situated insights into the emergence of gentrification across space. However, given the limits of this paper, only 2 accounts can be briefly discussed in conjunction with our study's findings below. Nevertheless, to ensure that insights can be continually gleaned far beyond this paper's conclusion, a supplementary data visualisation platform displaying our research outputs

(*https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/*) and codes are uploaded online (*https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes*) for anyone wishing to dive deeper into the details.

6.1.1 'Icebergs' of Fulham's Super-Gentrifying Neighbourhoods

Parsons Green, centred around Fulham's south, has traditionally been an affluent neighbourhood that enjoys proximity to Central London, reputable international schools like the Ecole Marie d'Orliac, and count historic 'Victorian and Edwardian' and the highly sought-after 'lion houses' amongst its residential stock (Casey, 2015). Notwithstanding the steep house prices, super-gentrification has been creeping into Parsons Green, driven by international investors and capital in particular (The Resident, 2014). Capturing such trends in our analysis, *Figures 54(a)-(b)* show that while super-gentrification appeared in only a slither of LSOAs to the east of Parsons Green between 2001-2011, the phenomenon is slated to expand markedly according to our model's predictions.

More interestingly, though, these trends have seemingly been accompanied by a proliferation of hyper-luxurious basements in 'iceberg houses' (Baldwin et al., 2018, p.5) dug to accommodate the opulent amenities, from swimming pools to cinemas, of London's super-rich in recent years. Juxtaposing the geographies of 'iceberg houses' in the borough *(see Figure 54(c))* against our model's predictions of where super-gentrification will likely spread into, there is a clear mirroring between the two; which is hence symbolic of the 'global excesses of wealth' and new 'spatial expression' that have

and will continue to define the exclusive trajectories of super-gentrification that are unique to London (Baldwin et al., 2018, p.17).

Figure 54(a)-(c). Gentrification Typologies and 'Iceberg Basements' within the Borough of Hammersmith and Fulham Borough

Source: (Baldwin et al., 2018, p.31)

6.1.2 The Battlegrounds of Newham's North-east

Quintessentially East London, Newham's north-east have housed London's working-class over the decades. Despite its proximity to Stratford which was revitalised for the 2012 Olympics and the planned Elizabeth Line stations at Manor Park and Forest Gate, house prices in this part of Newham have remained resolutely affordable through the years (EastBlam!, 2015), to the extent that realtors have recently touted it as 'a place where real Londoners can still afford to buy' (Bloomfield, 2018).

However, with this confluence of events, socio-spatial transformations have started to occur. Based on our findings, marginal gentrifiers, who are attracted by the appeal of affordably-priced neighbourhoods in accessible locations (Mendes, 2013), had started to enter the area between 2001-2011. Projecting forward, marginal gentrification, likely catalysed by the opening of the Elizabeth Line, is postulated to inundate the area, while the minority of mainstream gentrifying LSOAs will gravitate closer towards the new stations *(see Figure 55)*. Notwithstanding these forecasts, it is recognised that the future opening of the Elizabeth Line, and indeed other oncoming major infrastructural projects elsewhere (e.g. High-Speed 2, etc.), could bring about unprecedented impacts on house prices and advance gentrification in ways that our model can only partially speculate at the present stage. Further in-depth modelling and scenario-casting will necessarily be required to better gauge such impacts and refine the predictions; which are beyond the ambit of this paper but are certainly viable avenues for future research.

Figure 55. Gentrification Typologies in the Borough of Newham

Though a boon for gentrifiers, such developments have spelt disaster for incumbent residents, especially low-income council tenants, as many have been reportedly evicted to make way for incomers (Hancox, 2014). Consequently, an ongoing struggle exists between segments of Newham's community and the local council; which most tangibly surfaced through the 2014 protests led by the 'Focus E15' movement (Parkinson and Domokos, 2014).

7 Conclusion

Addressing urgent calls for academia to cultivate deeper scholarly insights that are sensitive to the spatialities and intricacies of urban change, this paper has meticulously distilled the nuanced geographies and dyanmics of neighbourhood change and gentrification, using a combination of ML and spatial analysis under a data-driven approach. In identifying, modelling and predicting the various neighbourhood states and gentrification typologies operating across London, the findings this analysis will likely contribute substantially towards equipping policy-makers with tools that allow them to comprehend and strategically tackle gentrification and displacement, both present and future. It is also hoped that the innovative methodology employed here will inspire future works to tinker with new empirical approaches going forward.

Notwithstanding its successes, this paper acknowledges certain limitations inherent in the analysis which could be improved on, potentially through future research. Firstly, as with most spatial analyses, the Modifiable Areal Unit Problem (MAUP) is an issue (Lloyd, 2016, p.1189). Though our analysis operationalised LSOA units which are already relatively fine-grained, it is recognised that not all dwellings and households in the LSOA necessarily conformed to identical trends. More precise data visualisations and granular analysis, assuming if such high-resolution data was available, are hence imperatives to overcome MAUP. Secondly, as noted earlier, an error margin was inevitable in the ML model's predictions, which inadvertently inflated the forecasts of future gentrifying LSOAs. Potentially, the linkage to more relevant datasets and even other models could be done to further enhance the current analysis and our model's predictions. Finally, acknowledging that gentrification is a process which potentially

spans over decades (Zuk and Chapple, 2015b), extending our analysis' timeframe retrospectively to the 1990s, or even the 1960s when gentrification was first observed by Ruth (1964), would have enabled even greater insights. However, given the dearth of quality data, comparable to the ones used in this paper, for those time-periods culled such intentions. Nonetheless, with more historical data gradually being curated through projects like the Layers of London (Institute of Historical Research, 2018), detailed analysis on past gentrification cycles may not be too far a stretch in time to come.

Bibliography

Alexiou, A., Singleton, A., 2015. Geodemographic Analysis, in: Brunsdon, C., Singleton, A.D. (Eds.), Geocomputation: A Practical Primer. SAGE Publications, Inc., London, UK, pp. 136–151. https://doi.org/10.4135/9781473916432

Anselin, L., 2019. Local Spatial Autocorrelation - Univariate Local Statistics [WWW Document]. URL https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html (accessed 02.08.19).

Atkinson, R., 2008. Commentary: Gentrification, Segregation and the Vocabulary of Affluent Residential Choice. Urban Studies 45, 2626–2636. https://doi.org/10.1177/0042098008097110

Atkinson, R., 2004. The Evidence on the Impact of Gentrification: New Lessons for the Urban Renaissance? European Journal of Housing Policy 4, 107–131. https://doi.org/10.1080/1461671042000215479

Atkinson, R., Bridge, G., 2005. Gentrification in a Global Context, 1st ed. Routledge, New York, NY. https://doi.org/10.4324/9780203392089

Baldwin, S., Holroyd, E., Burrows, R., 2018. Mapping the Subterranean Geographies of Plutocratic London: Luxified Troglodytism? (Preprint).

Barton, M., 2016. An Exploration of the Importance of the Strategy used to Identify Gentrification. Urban Studies 53, 92–111. https://doi.org/10.1177/0042098014561723

Ben Fraj, M., 2017. In Depth: Parameter Tuning for Random Forest [WWW Document]. In Depth: Parameter Tuning for Random Forest - All Things AI - Medium. URL https://medium.com/all-things-ai/in-depth-parameter-tuning-for-random-forest-d67bb7e920d (accessed 18.07.19).

Bergen, S., Lindstrom, J., 2019. Comprehensive Tutorial for the Spatio-Temporal R-package [WWW Document]. URL https://cran.r-project.org/web/packages/SpatioTemporal/vignettes/ST_tutorial.pdf (accessed 07.08.19).

Bloomfield, R., 2018. Watch this space: Zone 3 First-time Buyer Homes in “Real” East End of London Outpost [WWW Document]. Homes and Property. URL https://www.homesandproperty.co.uk/property-news/buying/first-time-buyers/zone-3-firsttime-buyer-homes-in-real-east-end-of-london-outpost-a117671.html (accessed 17.08.19).

Boddy, M., 2007. Designer Neighbourhoods: New-Build Residential Development in Nonmetropolitan UK Cities—The Case of Bristol. Environment and Planning A: Economy and Space 39, 86–105. https://doi.org/10.1068/a39144

Bridge, G., 2003. Time-Space Trajectories in Provincial Gentrification. Urban Studies 40, 2545–2556. https://doi.org/10.1080/0042098032000136200

Brunsdon, C., Singleton, A., 2015. Conclusion: The Future of Applied Geocomputation, in: Geocomputation: A Practical Primer. SAGE Publications, Inc., 1 Oliver's Yard, 55 City Road London EC1Y 1SP, pp. 320–326. https://doi.org/10.4135/9781473916432

Butler, T., 2011. Gentrification in London — Modes of Middle-class Establishment in a Global City, in: Herrmann, H., Keller, C., Neef, R., Ruhne, R. (Eds.), Die Besonderheit Des Städtischen. VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 265–284. https://doi.org/10.1007/978-3-531-93338-2_13

Butler, T., 2007. Re-urbanizing London Docklands: Gentrification, Suburbanization or New Urbanism? International Journal of Urban and Regional Research 31, 759–781. https://doi.org/10.1111/j.1468-2427.2007.00758.x

Butler, T., 1997. Gentrification and the Middle Classes. Ashgate, Aldershot, UK.

Butler, T., Hamnett, C., Ramsden, M.J., 2013. Gentrification, Education and Exclusionary Displacement in East London: Gentrification, education and exclusionary displacement in East London. International Journal of Urban and Regional Research 37, 556–575. https://doi.org/10.1111/1468-2427.12001

Butler, T., Lees, L., 2006. Super-gentrification in Barnsbury, London: Globalization and Gentrifying Global Elites at the Neighbourhood Level. Transactions of the Institute of British Geographers 31, 467–487. https://doi.org/10.1111/j.1475-5661.2006.00220.x

Caliński, T., Harabasz, J., 1974. A Dendrite Method for Cluster Analysis. Communications in Statistics 3, 1–27. https://doi.org/10.1080/03610927408827101

Casey, C., 2015. Parsons Green Rides Out Slump in Fulham Property Prices [WWW Document]. Financial Times. URL https://www.ft.com/content/1a897206-0f8e-11e5-b968-00144feabdc0 (accessed 16.08.19).

CDRC, 2018. Population Churn Index. URL https://data.cdrc.ac.uk/dataset/cdrc-population-churn-index (accessed 01.06.19).

Chapple, K., Loukaitou-Sideris, A., Waddell, P., Chatman, D., Ong, P., 2017. Developing a New Methodology for Analyzing Potential Displacement.

Chapple, K., Zuk, M., 2016. Systems for Gentrification. Cityscape: A Journal of Policy Development and Research 18, 109–130.

Chermesh, D., Hao, X., Gerardo Rodríguez, V., Hambardzumyan, R., Bianco, F., Chapple, K., Kleiman, N., Sobolevsky, S., 2018. Map of Gentrification and Displacement for The Greater New York (Capstone Project). New York University.

Cox, N.J., 2011. Stata Tip 96: Cube Roots. The Stata Journal 11, 149–154.

Dutton, P., 2003. Leeds Calling: The Influence of London on the Gentrification of Regional Cities. Urban Studies 40, 2557–2572. https://doi.org/10.1080/0042098032000136219

EastBlam!, 2015. Let's all move to East Ham, E6 - Find out more about life in Newham. EastBLAM! URL http://www.eastblam.co.uk/2015/02/east-ham-newham-e6-houses/ (accessed 18.08.19).

ESRI, n.d. How Spatial Autocorrelation (Global Moran's I) works—Help | ArcGIS Desktop [WWW Document]. URL http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm (accessed 20.07.19).

Freeman, L., Cassola, A., Cai, T., 2016. Displacement and gentrification in England and Wales: A quasi-experimental approach. Urban Studies 53, 2797–2814. https://doi.org/10.1177/0042098015598120

Friedline, T., Masa, R.D., Chowa, G.A.N., 2015. Transforming wealth: Using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement. Social Science Research 49, 264–287. https://doi.org/10.1016/j.ssresearch.2014.08.018

Galster, G., 2001. On the Nature of Neighbourhood. Urban Studies 38, 2111–2124. https://doi.org/10.1080/00420980120087072

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International 1–16. https://doi.org/10.1080/10106049.2019.1595177

GLA, 2019. London Development Database SQL Extract. URL https://data.london.gov.uk/dataset/london-development-database-sql-extract (accessed 01.06.19).

Glass, R., 1964. London: Aspects of Change. MacGibbon & Kee, London.

Hackworth, J., Smith, N., 2001. The Changing State of Gentrification. Tijdschrift voor economische en sociale geografie 92, 464–477. https://doi.org/10.1111/1467-9663.00172

Hamnett, C., 2003. Gentrification and the Middle-class Remaking of Inner London, 1961-2001. Urban Studies 40, 2401–2426. https://doi.org/10.1080/0042098032000136138

Hamnett, C., 1984. Gentrification and Residential Location Theory: A Review and Assessment, in: Herbert, D., Johnston, R.J. (Eds.), Geography and the Urban

Environment: Progress in Research and Applications. Wiley & Sons, London, UK, pp. 283–319.

Hancox, D., 2014. East London "Regeneration" vs. Young Mothers. Vice. URL https://www.vice.com/en_uk/article/7baz79/Newham-single-mothers-protest-gentrification-olympic-inspire-a-generation (accessed 12.08.19).

Hochstenbach, C., van Gent, W.P., 2015. An Anatomy of Gentrification Processes: Variegating Causes of Neighbourhood Change. Environment and Planning A: Economy and Space 47, 1480–1501. https://doi.org/10.1177/0308518X15595771

Hristova, D., Aiello, L.M., Quercia, D., 2018. The New Urban Success: How Culture Pays. Frontiers in Physics 6. https://doi.org/10.3389/fphy.2018.00027

Ilic, L., Sawada, M., Zarzelli, A., 2019. Deep Mapping Gentrification in a Large Canadian City Using Deep Learning and Google Street View. PLOS ONE 14, 1–21. https://doi.org/10.1371/journal.pone.0212814

Institute of Historical Research, 2018. Layers of London | Recording the Layers of London's Rich Heritage [WWW Document]. URL https://www.layersoflondon.org/ (accessed 11.08.19).

Jaadi, Z., 2019. A Step by Step Explanation of Principal Component Analysis [WWW Document]. Medium. URL https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2 (accessed 05.07.19).

Kiely, T.J., Bastian, N.D., 2019. The Spatially-Conscious Machine Learning Model. arXiv:1902.00562 [cs, stat].

Lansley, G., Li, W., Longley, P., 2018. Modelling Small Area Level Population Change from Administrative and Consumer Data. Presented at the GIS Research UK (GISRUK) 2018, The University of Leicester.

Lauren, B., 2017. Machine Learning in ArcGIS [WWW Document]. ESRI. URL https://www.esri.com/arcgis-blog/products/analytics/analytics/machine-learning-in-arcgis/ (accessed 17.07.19).

Lees, L., 2009. Urban Renaissance in an Urban Recession: The End of Gentrification? Environment and Planning A: Economy and Space 41, 1529–1533. https://doi.org/10.1068/a4275

Lees, L., 2008. Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance? Urban Studies 45, 2449–2470. https://doi.org/10.1177/0042098008097099

Lees, L., 2007. Progress in Gentrification Research? Environment and Planning A: Economy and Space 39, 228–234. https://doi.org/10.1068/a39329

Lees, L., 1994. Gentrification in London and New York: An Atlantic gap? Housing Studies 9, 199–217. https://doi.org/10.1080/02673039408720783

Lees, L., Morales, E.L., Shin, H.B., 2016. Planetary Gentrification. Polity Press, Malden, MA.

Lees, L., Shin, H.B., López-Morales, E. (Eds.), 2015. Global Gentrifications: Uneven Development and Displacement, 1st ed. Bristol University Press, Bristol, UK. https://doi.org/10.2307/j.ctt1t894bt

Lees, L., Slater, T., Wyly, E.K., 2008. Gentrification. Routledge, New York, NY.

Lemaitre, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. Journal of Machine Learning Research 18, 1–5.

Ley, D., 1996. The New Middle Class and the Remaking of the Central City. Oxford University Press, Oxford, UK.

Lin, J., 2017. Feature Scaling [WWW Document]. URL https://jovianlin.io/feature-scaling/ (accessed 1.7.19).

Lloyd, C.D., 2016. Spatial Scale and Small Area Population Statistics for England and Wales. International Journal of Geographical Information Science 30, 1187–1206. https://doi.org/10.1080/13658816.2015.1111377

Longley, P., Cheshire, J., Singleton, A.D., 2018. Consumer Data Research. UCL Press.

Marcuse, P., 1985. Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City. Journal of Urban and Contemporary Law 28, 195–240.

Marcuse, P., 1984. To Control Gentrification: Anti-displacement Zoning and Planning for Stable Residential Districts. Review of Law & Social Change 931–952.

Massey, D., 2017. Prologue: “World City,” in: Ren, X., Keil, R. (Eds.), The Globalizing Cities Reader. Routledge, London, UK, pp. 1–512.

Mendes, L., 2013. Marginal Gentrification as Emancipatory Practice: An Alternative to the Hegemonic Discourse of the Creative City? RCCS Annual Review. https://doi.org/10.4000/rccsar.513

Moore, K., 2012. Is this the UK’s Most Gentrified Street? BBC News. URL https://www.bbc.com/news/magazine-18394017 (accessed 12.08.19).

Navlani, A., 2018. AdaBoost Classifier in Python [WWW Document]. AdaBoost Classifier in Python (article) - DataCamp. URL

https://www.datacamp.com/community/tutorials/adaboost-classifier-python (accessed 18.07.19).

ONS, 2018. Median House Prices for Administrative Geographies: HPSSA Dataset 9. URL https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianhousepricefornationalandsubnationalgeographiesquarterlyrollingyearhpssadataset09/current (accessed 01.06.19).

ONS, 2016. 2011 Census Aggregate Data. UK Data Service (Edition: June 2016). DOI http://dx.doi.org/10.5257/census/aggregate-2011-1 (accessed 01.06.19).

ONS, 2015. Methodology Note for the 2011 Area Classification for Output Areas. URL https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables (accessed 01.06.19).

ONS, 2011. 2001 Census Aggregate Data. UK Data Service (Edition: 2011). DOI http://dx.doi.org/10.5257/census/aggregate-2001-2 (accessed 01.06.19).

ONS, 2004. Census 2001: Definitions. TSO, London.

ONS, n.d. The National Statistics Socio-economic classification (NS-SEC) [WWW Document]. URL https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 (accessed 12.7.19).

Owens, A., 2012. Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent: CITY & COMMUNITY. City & Community 11, 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Parkinson, J.N., Domokos, J., 2014. London's Housing Crisis: The Story of the Focus E15 mums. The Guardian. URL https://www.theguardian.com/society/video/2014/oct/08/london-housing-newham-mums-focus-e15-video (accessed 13.08.19).

Reades, J., De Souza, J., Hubbard, P., 2019. Understanding Urban Gentrification Through Machine Learning. Urban Studies 56, 922–942. https://doi.org/10.1177/0042098018789054

Redfern, P.A., 1997. A New Look at Gentrification: 2. A Model of Gentrification. Environment and Planning A: Economy and Space 29, 1335–1354. https://doi.org/10.1068/a291335

Rérat, P., Söderström, O., Piguet, E., 2010. New Forms of Gentrification: Issues and Debates. Population, Space and Place 16, 335–343. https://doi.org/10.1002/psp.585

Rigolon, A., Németh, J., 2019. Toward a Socioecological Model of Gentrification: How People, Place, and Policy Shape Neighborhood Change. Journal of Urban Affairs 1–23. https://doi.org/10.1080/07352166.2018.1562846

Roantree, B., Pope, T., Joyce, R., 2019. The Characteristics and Incomes of the Top 1% (IFS Briefing Note No. BN 254). Institute of Fiscal Studies. https://doi.org/10.1920/BN.IFS.2019.BN0254

Rose, D., 1984. Rethinking Gentrification: Beyond the Uneven Development of Marxist Urban Theory. Environment and Planning D: Society and Space 2, 47–74. https://doi.org/10.1068/d020047

Royall, E., Wortmann, T., 2015. Finding the State Space of Urban Regeneration: Modeling Gentrification as a Probabilistic Process using k-Means Clustering and Markov Models. Presented at the 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM), Cambridge, MA, USA, pp. 1–24.

Slater, T., 2011. Gentrification of the City, in: Bridge, G., Watson, S. (Eds.), The New Blackwell Companion to the City, Wiley-Blackwell Companions to Geography. Wiley-Blackwell, Malden, MA, pp. 571–585.

Smith, D.P., Butler, T., 2007. Conceptualising the Sociospatial Diversity of Gentrification: 'To Boldly Go' into Contemporary Gentrified Spaces, the 'Final Frontier'? Environment and Planning A: Economy and Space 39, 2–9. https://doi.org/10.1068/a39322

Smith, N., 1979. Toward a Theory of Gentrification: A Back to the City Movement by Capital, not People. Journal of the American Planning Association 45, 538–548. https://doi.org/10.1080/01944367908977002

Social Metrics Commission, 2019. Measuring Poverty 2019 (Report). Social Metrics Commission, London, UK.

Swalin, A., 2018. Choosing the Right Metric for Evaluating Machine Learning Models - Part 2 [WWW Document]. URL https://medium.com/usf-msds/choosing-the-right-metric-for-evaluating-machine-learning-models-part-2-86d5649a5428 (accessed 17.07.19).

The Resident, 2014. Area Focus: Fulham [WWW Document]. URL https://www.theresident.co.uk/homes-interiors/area-focus-fulham/ (accessed 15.08.19).

UK Data Service, 2015. GeoConvert [WWW Document]. URL http://geoconvert.mimas.ac.uk/ (accessed 11.07.19).

Van Criekingen, M., Decroly, J.-M., 2003. Revisiting the Diversity of Gentrification: Neighbourhood Renewal Processes in Brussels and Montreal. Urban Studies 40, 2451–2468. https://doi.org/10.1080/0042098032000136156

van Ham, M., Manley, D., 2012. Neighbourhood Effects Research at a Crossroads. Ten Challenges for Future Research Introduction. Environment and Planning A: Economy and Space 44, 2787–2793. https://doi.org/10.1068/a45439

Watt, P., Minton, A., 2016. London's Housing Crisis and its Activisms: Introduction. City 20, 204–221. https://doi.org/10.1080/13604813.2016.1151707

Wei, F., Knox, P.L., 2014. Neighborhood Change in Metropolitan America, 1990 to 2010. Urban Affairs Review 50, 459–489. https://doi.org/10.1177/1078087413501640

White, I., McLaren, E., 2009. The 2011 Census Taking Shape: The Selection of Topics and Questions. Population Trends 135, 8–19. https://doi.org/10.1057/pt.2009.4

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. ed, Morgan Kaufmann series in data management systems. Morgan Kaufmann, Burlington, MA.

Yeo, I.-K., Johnson, R., 2000. A New Family of Power Transformations to Improve Normality or Symmetry. Biometrika 87, 954–959. https://doi.org/10.1093/biomet/87.4.954

Zuk, M., Chapple, K., 2015a. Urban Displacement Project. https://www.urbandisplacement.org/

Zuk, M., Chapple, K., 2015b. Case Studies on Gentrification and Displacement in the San Francisco Bay area.