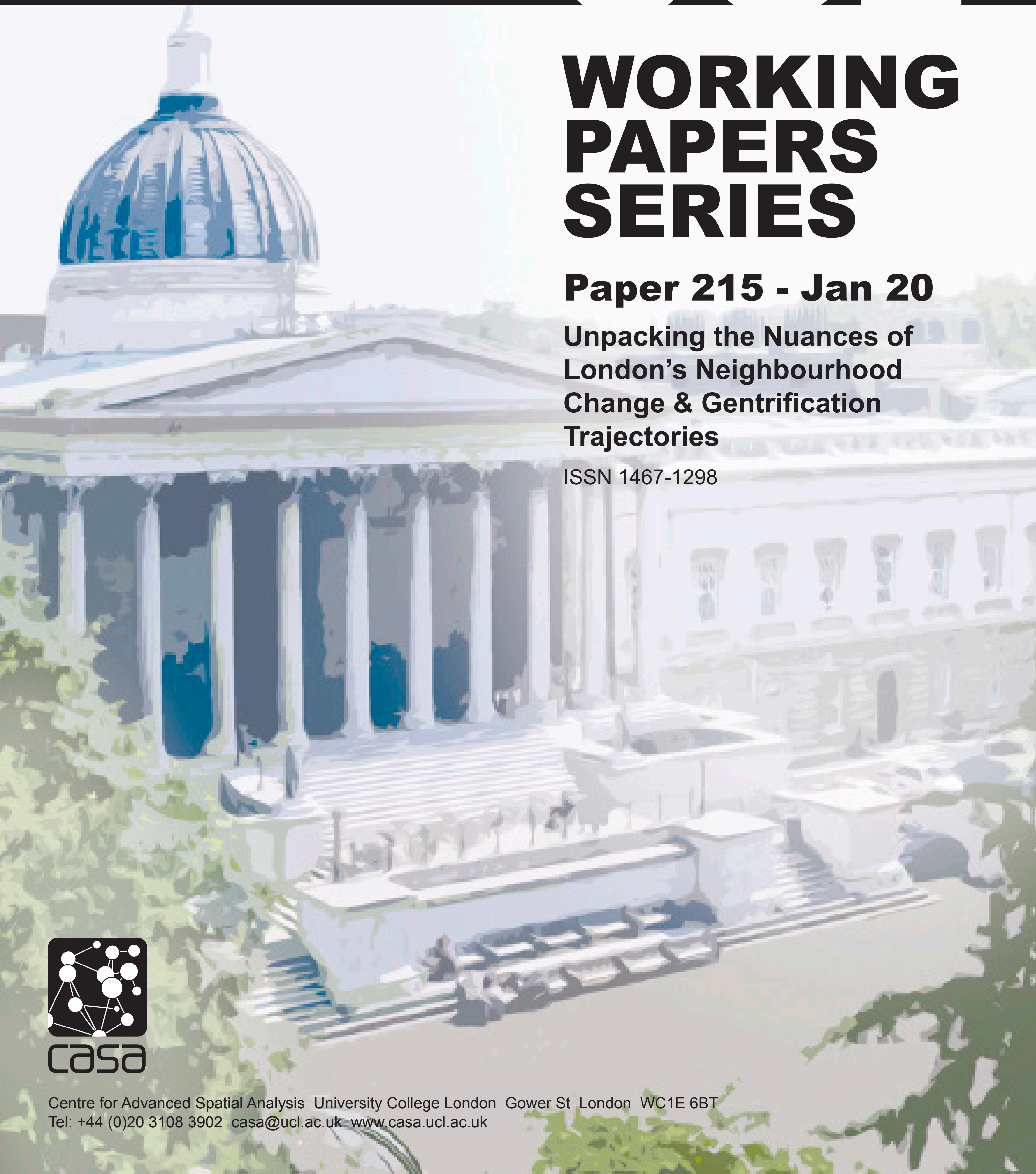
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL



WORKING PAPERS SERIES

Paper 215 - Jan 20

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

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London Gower St London WC1E 6BT
Tel: +44 (0)20 3108 3902 casa@ucl.ac.uk www.casa.ucl.ac.uk

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

Joshua Yee

University College London

ORCID ID: https://orcid.org/0000-0003-4650-0855

Adam Dennett

University College London

ORCID ID: https://orcid.org/0000-0001-8036-0185

Abstract

Amid the simmering global economic uncertainties and geopolitical instabilities of our present day, deepening social inequalities have remained a constant in our contemporary urban landscapes. Particularly within global cities like London, the disparities have become increasingly apparent. The gravity of the situation is underscored by recently published reports suggesting London's position as home to not only the majority of the UK's richest 1%, or indeed the 0.1%, but also a large proportion of its urban poor. In spite of such polarisations, a sense of indifference towards the needs of the average Londoner 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, has 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, new data sets and spatial analysis techniques, to systematically examine the variegated past and future trajectories of neighbourhood change across London. The nature of gentrification's mutations and its 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 policy-makers unprecedented and highly-contextualised insights into gentrification's projected dynamics and geographies.

1 Introduction

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, which concurrently spurred the rehabilitation of dated residential properties, precipitated upward shifts in housing prices and the resulted in the displacement of residents with lower SES from such areas (Hamnett, 2003; Lees et al., 2008). Gentrification commonly evokes strong debates in academia and policy circles across the world as it is a process of neighbourhood change which frequently serves to fulfil the residential and lifestyle preferences of specific societal groups (those towards the top of the socio-economic pile) at the expense of others (those towards the bottom) (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 studies have tended to overlook the nuanced typologies of neighbourhood change, such as 'incumbent upgrading' (Van Criekingen & Decroly, 2003, p.2452) and 're-urbanisation' (Buzar et al., 2007, p.64), that co-exist with gentrification (Hochstenbach & van Gent, 2015). Consequently, gentrification has typically been conflated with these non-gentrifying forms of neighbourhood change, which inhibit the robust generation of knowledge and insights to support policy-making. Atkinson (2008, p.2634) outlined these issues, noting 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 stark 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. Our aims are threefold: firstly, to identify, characterise and locate neighbourhoods which have undergone recent gentrification, specifically disaggregating the different types of change revealed by the data; secondly, to explore which neighbourhoods are likely be next in line; and thirdly, in the process present and make available data, code and novel interactive visualisations as a comprehensive tool for supporting policy and decision making in the city – accessible here[1,2]. Undergirding the originality of this study is a repertoire of novel Machine Learning (ML), spatial analytical techniques and new sources of multi-

[1] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes

[2] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/

dimensional data, leveraged to comprehensively examine the diversified patterns of neighbourhood change and gentrification trajectories across London. By operationalising a highly-integrative workflow tailored to critically deconstruct the nature and spatial underpinnings of neighbourhood change and gentrification's varying forms, the analysis first systematically identifies the continuum of neighbourhood change typologies active at the Lower Super Output Area (LSOA) scale, before drilling down to identify neighbourhoods which are genuinely gentrifying. Targeted in-depth investigations are then conducted to further segment the diverse gentrifying processes unfolding in London, model their patterns and eventually predict their potential trajectories.

2 Literature Review

2.1 Gentrification as an Evolving Process of Neighbourhood Change

Notwithstanding the coinage of the term 'gentrification' more than five 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 & Butler, 2007). The key assertions put forth by academics explaining the structural transformations that have (re)moulded gentrification's scope are summarised below.

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 UK, 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 (Van Criekingen & Decroly, 2003; Butler & Lees, 2006). Although a pluralism of examples exist, two 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, research has 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). Consequently, such developments set the stage for all research, including this paper, to recognise gentrification as a highly-dynamic and nuanced urban phenomenon that needs to be spatially situated.

2.2 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) provide an excellent summary of gentrification's extensive impacts, including the positive such as new investment and building, rising prosperity and changing social mix, but offset against the negative such as rising rents, community displacement, homelessness, psychological damage and rising costs.

Insofar as some attempts have been made to paint a rosy picture of gentrification by obfuscating its downsides, particularly by private developers and amidst state-sanctioned urban reneweal programmes (Lees, 2008), critical scholars have, through evidence-based research, characterised gentrification as a 'largely negative process' where its costs severely outweigh potential benefits (Atkinson, 2004, p.126).

Amongst the impacts, displacement arguably ranks most consequential of all, given the profound impacts is has on individuals and communities who are forced to leave places they once thought of as home (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 three broad threads (Marcuse, 1985), of which two are covered here due to their relevance. The first, '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 profound. Marcuse (1984, p.931) notes – '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 & 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 & 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 – viewable here[3].

A survey of recent research, however, reveals that the majority of such efforts have focused on US cities, which poses two 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, Cassola, & Cai, 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 & 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.

2.3 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. Owens (2012, p.347) adds helpful clarity 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

[3] https://www.urbandisplacement.org/map/sf

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 following 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 attract specific demographic groups – such as ‘younger single people or childless couples’ (2007, p.95) in Bristol’s case – but do not entail direct displacement since existing residences were unaffected. Exclusionary displacement is nonetheless possible if the introduction of new developments and populations precipitate a local uplift making previously 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) argues for the merits of being able to discern them, wherein she asserts 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 assumes 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.

2.4 Gentrification as a Subject of Empirical Analysis

Insofar as various quantitative methods have traditionally been used to study gentrification, they entail certain drawbacks in their implementation. In particular, Barton (2016) criticised that quantitative approaches tended to depend overwhelmingly on census data that were heavily orientated towards socio-economic aspects. Resultantly, quantitative studies are constrained in their perspectives, as they generally steer toward mono-dimensional assessments of gentrification (2016, p.93). Furthermore, since census data is typically produced only once per decade and usually published without any personal identifiers, quantitative analyses are inevitably restricted in their capacity to actively monitor gentrification and have no means of differentiating neighbourhoods which 'underwent naturally occurring improvements (incumbent upgrading)' from those that were gentrifying (2016, p.96).

The recent advent of 'Big Data', however, is a critical pathway that can enable quantitative research to overcome some of these longstanding, methodological weaknesses (Brunsdon and Singleton, 2015, p.322). Specifically, with the seemingly ever-enlarging diversity and availability of data, particularly in cities like London, opportunities for researchers to venture beyond traditional data sources, like the census, have correspondingly proliferated. Administrative, retail, and even crowd-sourced, social data (Hristova et al., 2018; Longley, Cheshire, & Singleton, 2018) are now offering new opportunities for understanding human patterns and processes. Gentrification analyses can thus be expected to, and indeed should, persist in leveraging 'Big Data' for the betterment of scholarship.

In parallel with the dawn of 'Big Data', advances in analytical techniques that are able to process and detect signals in these ever more vast and messy sources of information have been rapid in recent years. Machine Learning (ML) is a particular subset of artificial intelligence that has gained traction. Unlike classic statistical approaches, ML algorithms can learn from data inputs and adapt their processes to optimise performance for 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 more rudimentary linear regression analyses cannot achieve (Witten et al., 2011). In the realm of urban analytics, studies such as that carried out by Wei and Knox (2014) showcase ML's ability to extract patterns from 'noisy' spatial data, model and project complex spatial phenomena..

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 a pioneering example that has exploited ML for the purpose of understanding gentrification. In their paper, a specific ML algorithm called a '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) 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 statistics such as 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.

3 Data Collection

The datasets utilised in this study were derived from three main sources, namely the Office for National Statistics (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 following paragraphs. Data ‘wrangling’ and downstream analysis were carried out using Python and ArcMap.

3.1 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’s GeoConvert tool (2015) proved useful as it gave guidance on how data recorded for 2001 LSOAs could be re-weighted for 2011 boundaries.

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.

3.2 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 2 ways. Firstly, the dataset was scrubbed to remove records which either did not pertain to residential properties or possessed erroneous entries. Secondly, 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.

3.3 CDRC's Population Churn Data

Harnessing fine-grained data from public electoral registers, consumer databases and put together using 'bespoke data linkage techniques' (Lansley, Li, & Longley, 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 displayed here[4].

The complete suite of variables incorporated into this research is tabulated in *Table 1*.

[4] https://maps.cdrc.ac.uk/#/indicators/churn/default/BTTTFFT/10/-0.1500/51.5200/

Table 1. 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_more49hrs
	% of female residents in part-time jobs working <15hrs	Females_ptTime_less15hrs
	% of female residents in part-time jobs working 16 – 30hrs	Females_ptTime_16_30hrs
	% of female residents in full-time jobs working 31 – 48hrs	Females_FTime_31_48hrs
	% of female residents in full-time jobs working >49hrs	Females_FTime_more49hrs
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

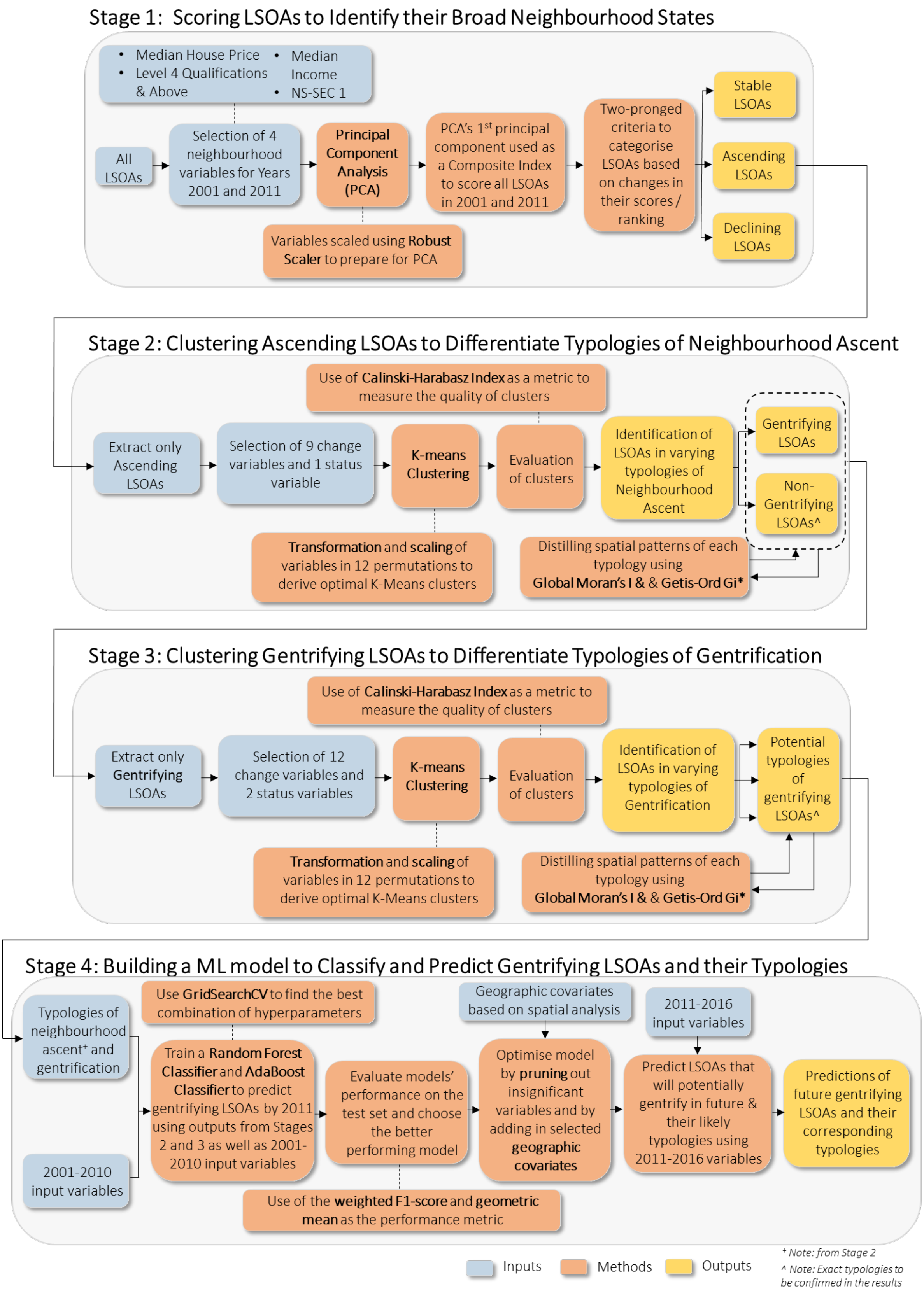
Data Source: (CDRC, 2018)

4 Methodology

Diverse methods, straddling across statistics, ML and spatial analysis, were chained into a multi-level workflow with every phase building upon results of the former in order to unpack the diversity and trajectories of gentrification across neighbourhoods in London. *Figure 1* illustrates the workflow diagrammatically while explanations of individual methods are detailed in the supplementary material. The full datasets and analysis code are available for those wishing to reproduce the analysis, here[5].

[5] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes

Figure 1. Overview of Entire Workflow



5 Analysis and Results

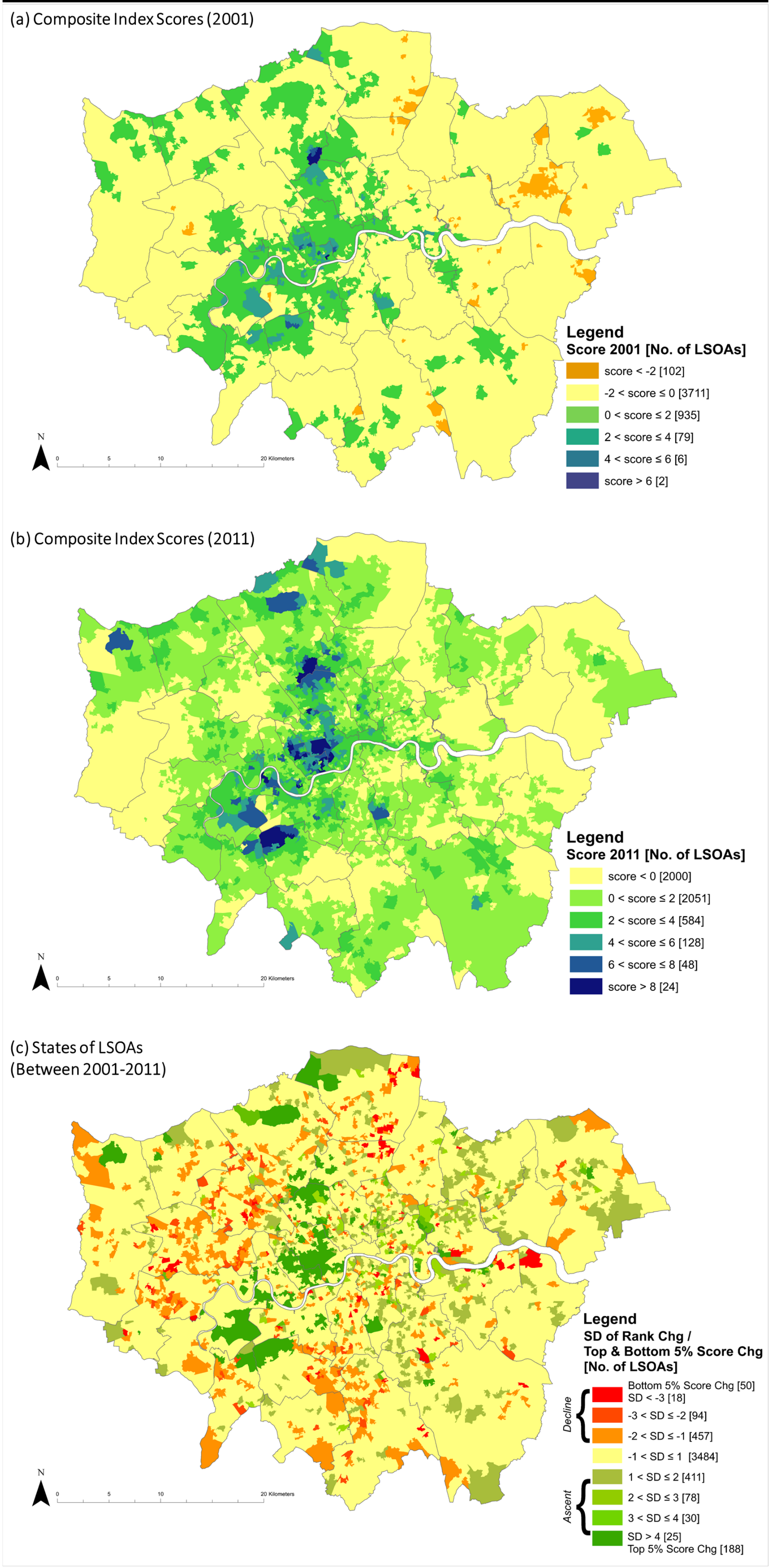
5.1 Exploring the States of London's LSOAs

The first stage in the analysis is to identify areas of recent neighbourhood change in the city. Following the examples of Reades et al. (2019) and Owens (2012), Median house price, income, degree-level (level 4) qualifications and those in the highest socio-economic class selected as proxies for quantifying neighbourhood states, particularly to determine if neighbourhoods had been ascending, declining or stable over a period of time and could be studied using data from the 2001 and 2011 Censuses and the ONS. Principal Components analysis was employed (as described in the supplementary material) to combine and create a single composite index variable for each year, allowing for changes to be visualised.

LSOAs mapped according to their Composite Index scores in 2001 and 2011 are provided in *Figures 2a-b respectively* – and can be explored interactively here[6]. 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.

[6] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page4

Figure 2. LSOAs' Composite Index Scores and Neighbourhood States between 2001-2011



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, we caution against the 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 2c (and interactively[7])*. 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.

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

---

[7] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page4

London while a denser congregation of declining LSOAs was situated toward the western peripheries.

5.2 Defining and Contextualising Typologies of Neighbourhood Ascent

A *k*-means algorithm was used to isolate clusters of neighbourhoods within the ascending areas, with similar ascent characteristics (see supplementary material and GitHub[8,9] for details). Three clusters were identified: Gentrification, Incumbent-upgrading and re-urbanisation - their profiles shown in *Figure 3* (and interactively[10]).

[8] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/4%20Data%20Preparation%20for%20Clustering.ipynb

[9] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/5a%20Clustering%20Ascending%20LSOAs.ipynb

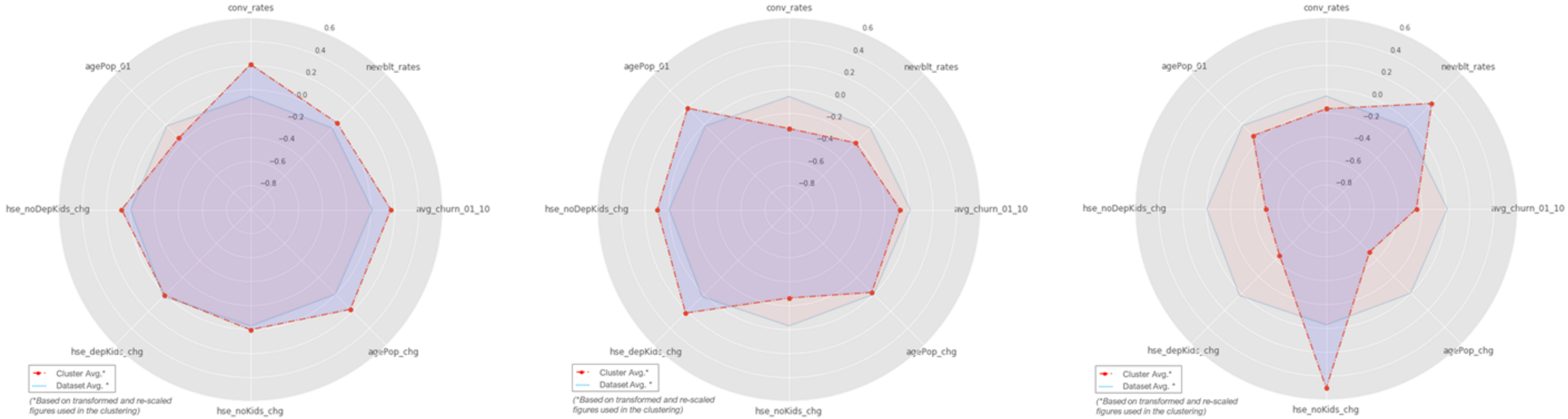
[10] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page3

Figure 3. Clusters of Ascending Neighbourhoods

(a) Gentrification Cluster
Total LSOAs: 334

(b) Incumbent Upgrading Cluster
Total LSOAs: 290

(c) Re-Urbanisation Cluster
Total LSOAs: 108



Variables	Labels	Gentrification	Incumbent Upgrading	Re-Urbanisation
		Values in actual terms		
Rate of planning permissions granted for conversion of existing residential properties (per 1,000 dwellings)	conv_rates	+5.36	+0.77	+1.58
Rate of planning permissions granted for new-built residences (per 1,000 dwellings)	newblt_rates	+3.08	+1.39	+5.89
Average Population churn	avg_churn_01_10	+2.47%	+2.27%	+2.04%
Change in % of residents aged 65+	agePop_chg	+0.18%	-1.03%	-5.00%
Change in % of households with no children	hse_noKids_chg	+0.86%	+-1.07%	+2.65%
Change in % of households with dependent children	hse_depKids_chg	+0.82%	+3.15%	-5.90%
Change in % of households with no dependent children	hse_noDepKids_chg	+0.74%	+0.87%	-2.52%
% of residents aged 65+ in 2001	agePop_01	10.49%	14.39%	10.57%

Figure 3a plots the average values of the Gentrification cluster's centroid against the means of the entire dataset. LSOAs 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.

The incumbent upgrading cluster is presented in *Figure 3b*. It contains LSOAs which exhibit 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.

Juxtaposed against the earlier clusters, the re-urbanisation cluster's centroid averages, as displayed in *Figure 3c*, demonstrate 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.

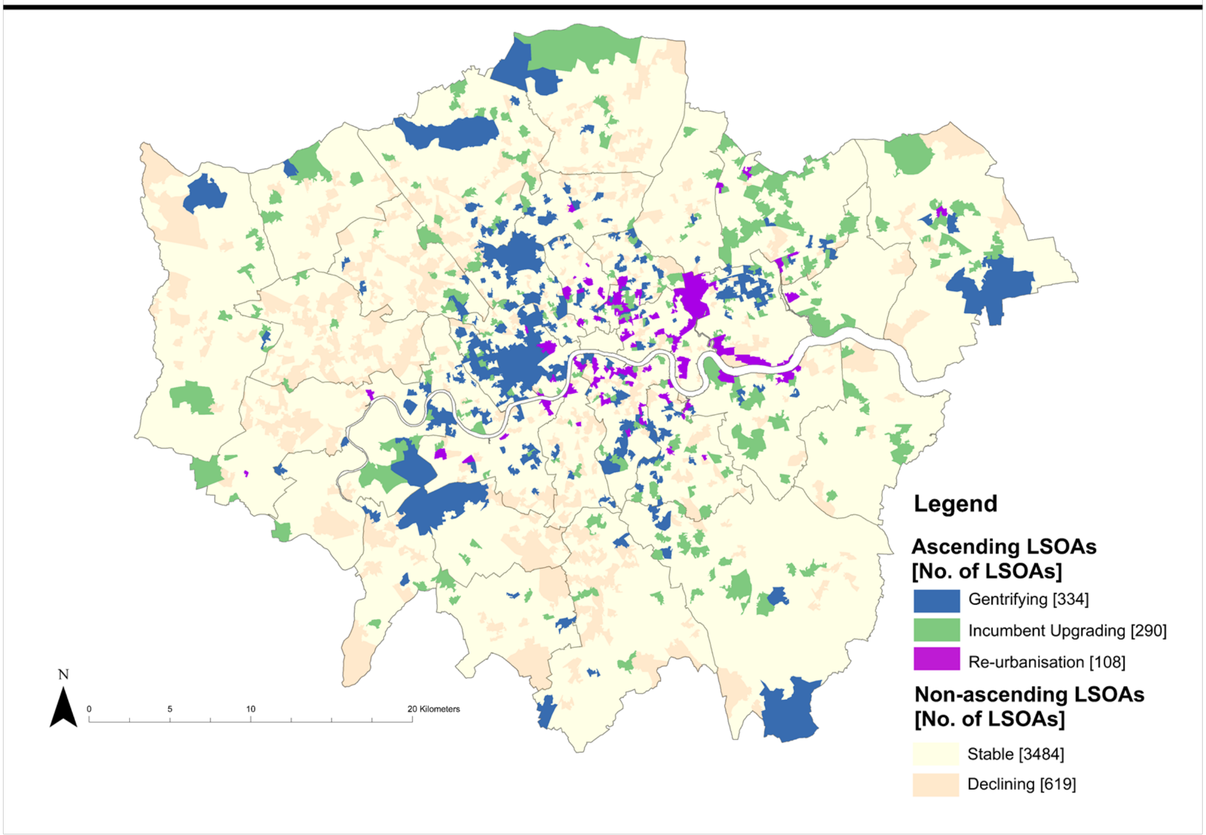
The influx of new-built developments and young families, combined with the attendant lack of population turnover, were consistent with the re-urbanisation typology where regeneration efforts were commonly made to introduce new developments and attract new populations. Despite not entailing direct displacement,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

The geographical distributions of the 3 neighbourhood ascent typologies are mapped in *Figure 4 (and interactively[11])*. There appears to be some degree of spatial autocorrelation amongst re-urbanising LSOAs in East London (particularly around the Olympic Park and the central-eastern boroughs of Tower Hamlets and Hackney), LSOAs undergoing incumbent upgrading especially in Redbridge and gentrifying LSOAs to the West of the city centre - albeit some exceptions along the city's fringes.

---

[11] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page3

Figure 4. LSOAs in Typologies of Neighbourhood Ascent & their Spatial Autocorrelation Statistics



Typology	Moran's Index	z-score	p-value	Conclusion
Gentrifying	0.249456	29.421436	0.00000	Given the z-score of 29.421436, there is a less than 1% chance that this clustered pattern could be a result of random chance.
Incumbent Upgrading	0.073596	8.699261	0.00000	Given the z-score of 8.699261, there is a less than 1% chance that this clustered pattern could be a result of random chance.
Re-Urbanisation	0.209344	24.772292	0.00000	Given the z-score of 24.772292, there is a less than 1% chance that this clustered pattern could be a result of random chance.

These visual observations are confirmed by a Global Moran's I test confirming that gentrifying, incumbent upgrading and re-urbanising 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.

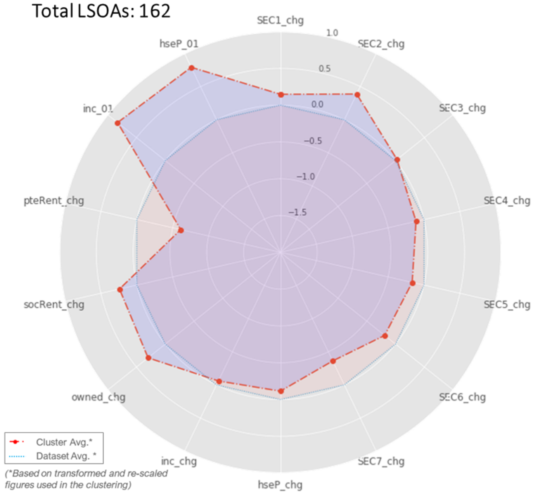
5.3 Defining Contextualised Typologies of Gentrification

In the final stage of the analysis, in order to ascertain the different types of gentrification occurring within the gentrification cluster, LSOAs in this cluster were re-classified using a similar methodology to the second stage, incorporating new variables (see the supplementary material and GitHub[12] for details). Three sub-categories of gentrification are identified within this cluster and described below.

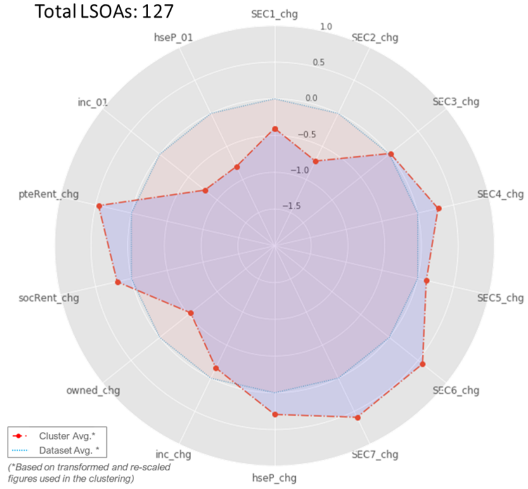
12 https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/5b%20Clustering%20Gentrifying%20LSOAs.ipynb

Figure 5. Clusters of Gentrifying Neighbourhoods

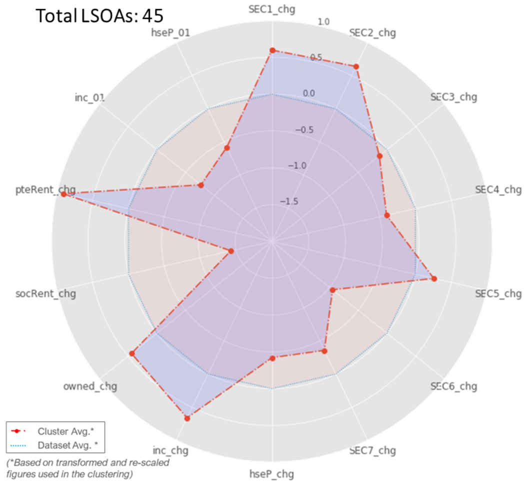
(a) Super-Gentrification Cluster
Total LSOAs: 162



(b) Marginal Gentrification Cluster
Total LSOAs: 127



(c) Mainstream Gentrification Cluster
Total LSOAs: 45



Variables	**Labels**	**Super-Gentrification**	**Marginal Gentrification**	**Mainstream Gentrification**
		Values in Actual Values		
Change in % of residents in the Higher Managerial, Administrative & Professional Occupation (NS-SEC Class 1)	SEC1_chg	+2.73%	+1.30%	+3.81%
Change in % of residents in the Lower Managerial, Administrative & Professional Occupation (NS-SEC Class 2)	SEC2_chg	+4.09%	+0.65%	+4.95%
Change in % of residents in the Intermediate Occupation (NS-SEC Class 3)	SEC3_chg	+1.48%	+1.33%	+1.26%
Change in % of residents in the Small Employers & Own Account Workers (NS-SEC Class 4)	SEC4_chg	+2.36%	+3.05%	+1.87%
Change in % of residents in the Lower Supervisory & Technical Occupation (NS-SEC Class 5)	SEC5_chg	-0.14%	+0.13%	+0.26%
Change in % of residents in the Semi-Routine Occupation (NS-SEC Class 6)	SEC6_chg	+0.27%	+1.47%	-0.85%
Change in % of residents in the Routine Occupation (NS-SEC Class 7)	SEC7_chg	+0.27%	+1.45%	+0.27%
% change in median house prices	hseP_chg	+163.9%	+181.5%	+143.6%
% change in median income	inc_chg	+46.65%	+45.23%	+54.27%
Change in % of residential properties owned by occupants	owned_chg	-2.78%	-7.34%	-2.00%
Change in % of residential properties socially rented by occupants	socRent_chg	-2.01%	-2.05%	-11.36%
Change in % of residential properties privately rented by occupants	pteRent_chg	+2.05%	+8.48%	+11.36%
Median income in 2001	inc_01	£37.24K	£27.14K	£25.13K
Median house price in 2001	hseP_01	£352.5K	£124.3K	£139.1K

The Super-Gentrification cluster (*Figure 5a)* reveals that 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 growth 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.

Contrastingly, the marginal gentrification cluster's centroid averages *(see Figure 5b)* denote 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.

Consisting 45 LSOAs, the Mainstream Gentrification cluster, akin to Marginal Gentrification, starts off with relatively lower income levels and house prices in 2001 *(see centroid averages in Figure 5c)*. However, unlike Marginal Gentrification, this cluster contains areas with well-above average gains in higher income and 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.

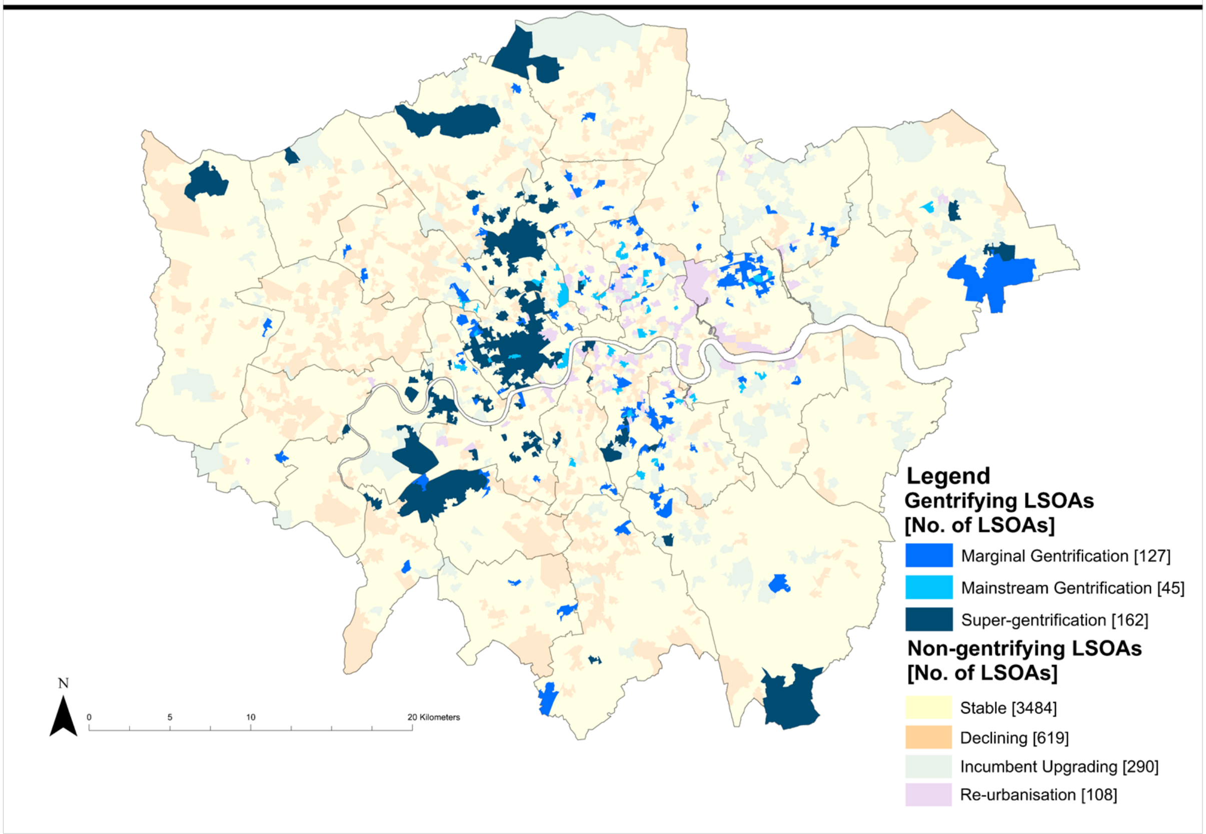
5.4 Uncovering Spatial Patterns

Figure 6 (and interactively[13]) maps out the geographies of the 3 gentrification sub-typologies. Evidently, super-gentrifying LSOAs appear 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 & Lees, 2006) and Portland Road (Moore, 2012). Contrastingly, LSOAs in marginal gentrification are found towards boroughs in East London, such as a constellation of LOSAs bunched together

[13] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page6

east of Stratford. Spatial patterns for LSOAs experiencing mainstream gentrification were however less apparent, though they were mostly contained within London's inner boroughs.

Figure 6. LSOAs in Gentrifying Typologies & their Spatial Autocorrelation Statistics



Typology	Moran's Index	z-score	p-value	Conclusion
Super-Gentrifying	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.
Marginal Gentrifying	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.
Mainstream Gentrifying	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.

5.5 Predicting Future Gentrification

Having identified clear areas of recent gentrification, our attention is turned to exploring where the next phases of gentrification may occur in the city. The first stage of any prediction is to try and understand the factors influencing past trends. Recent advances in multidimensional analysis have proven that machine learning algorithms can be particularly adept at this sort of challenge and so a suite of ML models 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, the full description in the supplementary material and code for the analysis is available here[14,15].

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 two ways. Firstly, though the addition on an inner London dummy and secondly through a gentrification neighbour dummy (see supplementary material for details). The addition of geographic covariates improved the prediction of gentrifying and non-gentrifying areas to 100% and 85% respectively.

A similar modelling methodology (see supplementary material and here[16]) incorporating additional variables can be used to predict the specific typologies of future gentrifying LSOAs. According to the predictions mapped in *Figure 7 (and for an interactive version, here[17])*, 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

---

[14] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/6%20Data%20Prep%20for%20Modelling.ipynb

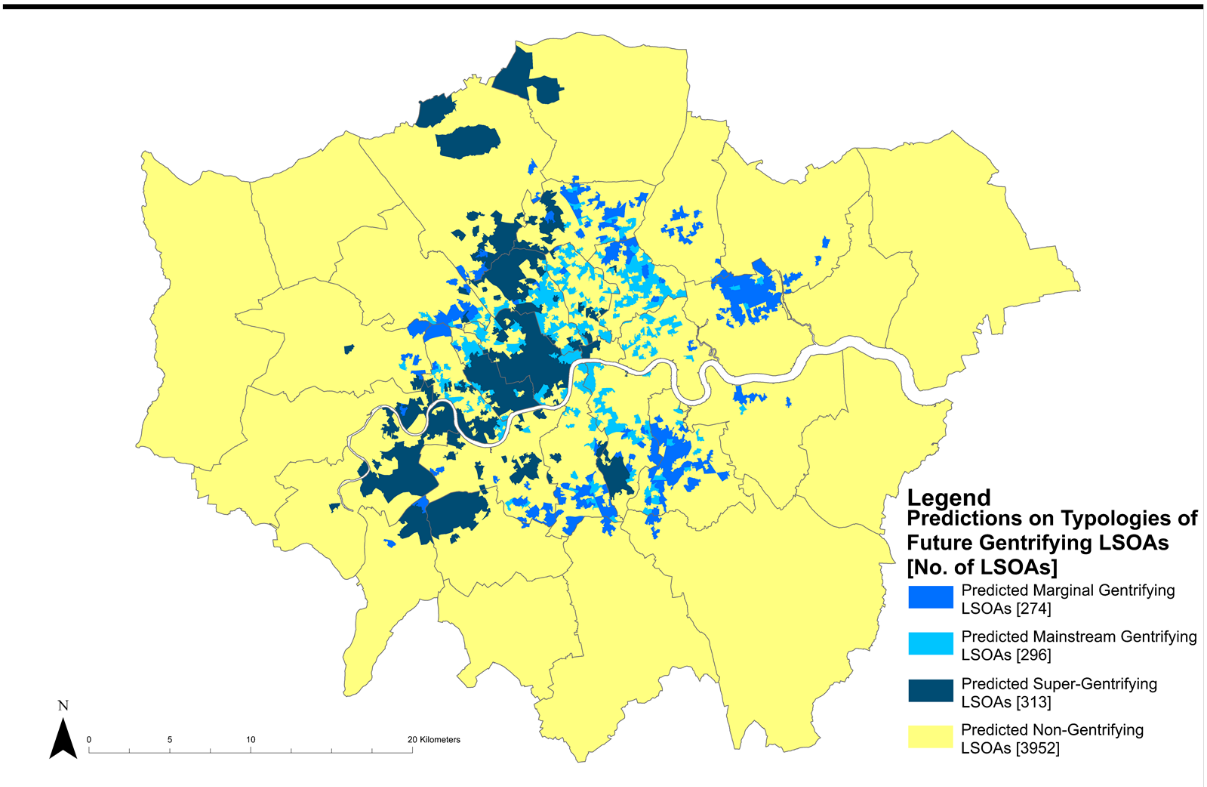
[15] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/7a%20Model%20%26%20Predict%20Ascending%20LSOAs.ipynb

[16] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes/blob/master/7b%20Model%20%26%20Predict%20Gentrifying%20LSOAs.ipynb

[17] https://jytg17.github.io/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories/#page6

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



6 Discussion

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 presented by our analysis, the two 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 , as this paper has demonstrated, the neighbourhood change process can be viewed more broadly, encompassing gentrification as one of its sub-typologies. Only through the analysis of new and previously unstudied population churn and planning permissions data were we able to tease out the nuance in this process.

Gentrifying LSOAs were found to comprise 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. 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.

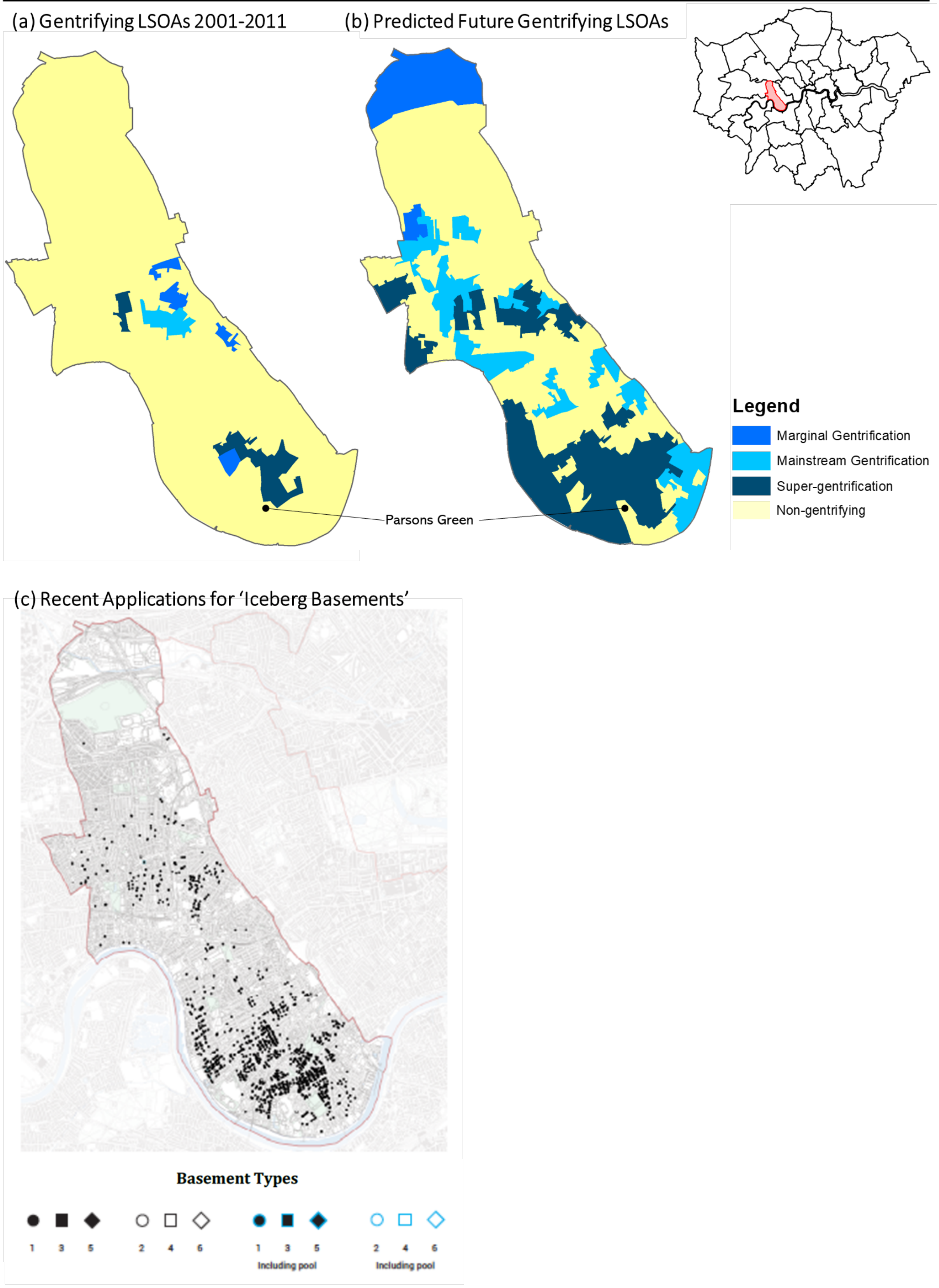
Each locale exhibiting any of gentrification's typologies will almost certainly contain rich accounts detailing the entanglements between place particularities and history which can help contextualise and enable situated insights into the emergence of gentrification across space. However, given the limits of this paper, only two 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 data visualisation platform displaying our research outputs and codes are uploaded online for anyone wishing to dive deeper into the details.

6.1 '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, fee-paying international schools like the Ecole Marie d'Orliac, and counts historic 'Victorian and Edwardian' and the highly sought-after 'lion houses' amongst its residential stock (Casey, 2015). Notwithstanding the high residential property values, 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 8a-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 8c)* 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 8. Gentrification Typologies and 'Iceberg Basements' within the Borough of Hammersmith and Fulham Borough



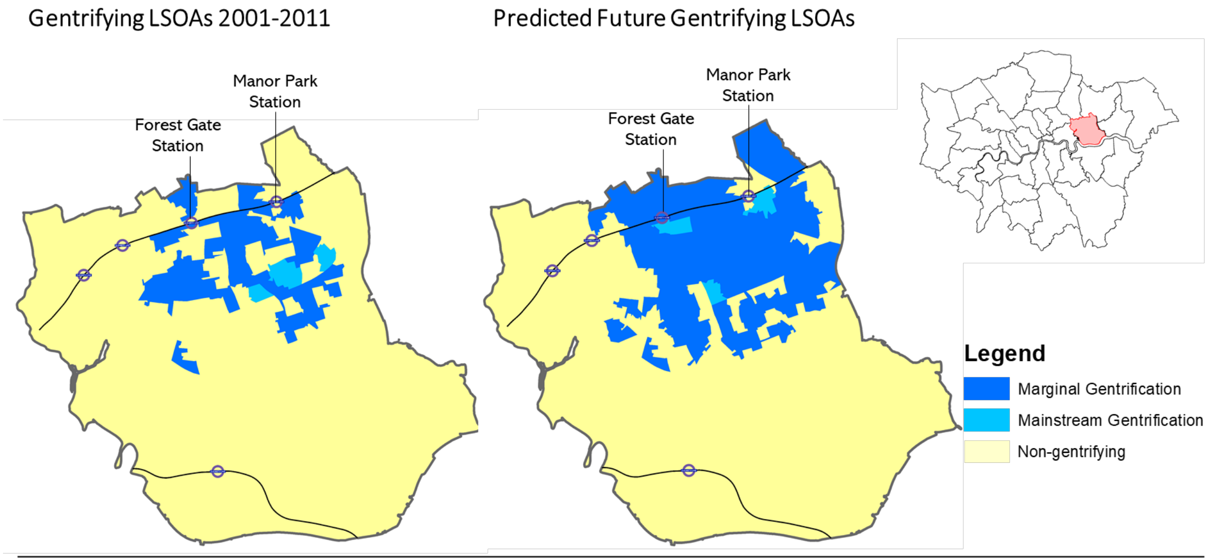
Source: (Baldwin et al., 2018, p.31)

6.2 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 estate agents 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 9)*. 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 9. 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 & Domokos, 2014).

7 Conclusion

In this paper, using a data-driven approach, the nuanced geographies and recent dynamics of neighbourhood change and gentrification in London have been explored, using a combination of ML and spatial analysis methods. In identifying, modelling and predicting the various neighbourhood states and gentrification typologies operating across London, the findings of this analysis should contribute substantially towards equipping policy-makers with tools that allow them to comprehend and strategically tackle gentrification and displacement, both present and future. We have shown very clearly that in places such as Fulham and Newham, while experiencing very different types of neighbourhood change, the challenges faced by the borough councils in dealing with significant change are not going to abate and will require concerted efforts to mitigate the challenges that such change will bring.

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 & 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.

References

Atkinson, R. (2004). The Evidence on the Impact of Gentrification: New Lessons for the Urban Renaissance? *European Journal of Housing Policy*, *4*(1), 107–131. https://doi.org/10.1080/1461671042000215479

Atkinson, R. (2008). Commentary: Gentrification, Segregation and the Vocabulary of Affluent Residential Choice. *Urban Studies*, *45*(12), 2626–2636. https://doi.org/10.1177/0042098008097110

Atkinson, R., & Bridge, G. (2005). *Gentrification in a Global Context* (1st ed.). https://doi.org/10.4324/9780203392089

Baldwin, S., Holroyd, E., & Burrows, R. (2018). *Mapping the Subterranean Geographies of Plutocratic London: Luxified Troglodytism?* [Preprint]. Retrieved from https://www.researchgate.net/publication/325046741_Mapping_the_Subterranean_Geographies_of_Plutocratic_London_Luxified_Troglodytism

Barton, M. (2016). An exploration of the importance of the strategy used to identify gentrification. *Urban Studies*, *53*(1), 92–111. https://doi.org/10.1177/0042098014561723

Bloomfield, R. (2018). Watch this space: Zone 3 First-time Buyer Homes in 'Real' East End of London Outpost. Retrieved 17 August 2019, from Homes and Property website: 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

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*(1), 86–105. https://doi.org/10.1068/a39144

Bridge, G. (2003). Time-Space Trajectories in Provincial Gentrification. *Urban Studies*, *40*(12), 2545–2556. https://doi.org/10.1080/0042098032000136200

Brunsdon, C., & Singleton, A. (2015). Conclusion: The Future of Applied Geocomputation. In *Geocomputation: A Practical Primer* (pp. 320–326). https://doi.org/10.4135/9781473916432

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*(2), 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*(4), 467–487. https://doi.org/10.1111/j.1475-5661.2006.00220.x

Buzar, S., Hall, R., & Ogden, P. E. (2007). Beyond Gentrification: The Demographic Reurbanisation of Bologna. *Environment and Planning A: Economy and Space*, *39*(1), 64–85. https://doi.org/10.1068/a39109

Casey, C. (2015, June 19). Parsons Green rides out slump in Fulham property prices. Retrieved 16 August 2019, from Financial Times website: https://www.ft.com/content/1a897206-0f8e-11e5-b968-00144feabdc0

CDRC. (2018). *Population Churn Index*. Retrieved from https://data.cdrc.ac.uk/dataset/cdrc-population-churn-index

Chapple, K., & Zuk, M. (2016). Systems for Gentrification. *Cityscape: A Journal of Policy Development and Research*, *18*(3), 109–130.

Chermesh, D., Hao, X., Gerardo Rodríguez, V., Hambardzumyan, R., Bianco, F., Chapple, K., … Sobolevsky, S. (2018). *Map of Gentrification and Displacement for The Greater New York* (Capstone Project, New York University). Retrieved from http://www.udpny.org/static/media/report.8f3f1564.pdf

Dutton, P. (2003). Leeds Calling: The Influence of London on the Gentrification of Regional Cities. *Urban Studies*, *40*(12), 2557–2572. https://doi.org/10.1080/0042098032000136219

EastBlam! (2015, February 27). Let’s all move to East Ham, E6—Find out more about life in Newham. Retrieved 18 August 2019, from EastBLAM! website: http://www.eastblam.co.uk/2015/02/east-ham-newham-e6-houses/

Freeman, L., Cassola, A., & Cai, T. (2016). Displacement and gentrification in England and Wales: A quasi-experimental approach. *Urban Studies*, *53*(13), 2797–2814. https://doi.org/10.1177/0042098015598120

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., … 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*. Retrieved from https://data.london.gov.uk/dataset/london-development-database-sql-extract

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

Hamnett, C. (2003). Gentrification and the Middle-class Remaking of Inner London, 1961-2001. *Urban Studies*, *40*(12), 2401–2426. https://doi.org/10.1080/0042098032000136138

Hancox, D. (2014, January 21). East London ‘Regeneration’ vs. Young Mothers. Retrieved 12 August 2019, from Vice website: https://www.vice.com/en_uk/article/7baz79/Newham-single-mothers-protest-gentrification-olympic-inspire-a-generation

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*(7), 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*(3), 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. Retrieved 11 August 2019, from https://www.layersoflondon.org/

Kiely, T. J., & Bastian, N. D. (2019). The Spatially-Conscious Machine Learning Model. *ArXiv:1902.00562 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1902.00562

Lansley, G., Li, W., & Longley, P. (2018, January). *Modelling Small Area Level Population Change from Administrative and Consumer Data*. Presented at the GIS Research UK (GISRUK) 2018, The University of Leicester. Retrieved from http://leicester.gisruk.org/programme/

Lauren, B. (2017). Machine Learning in ArcGIS. Retrieved 17 July 2019, from ESRI website: https://www.esri.com/arcgis-blog/products/analytics/analytics/machine-learning-in-arcgis/

Lees, L. (1994). Gentrification in London and New York: An Atlantic gap? *Housing Studies*, *9*(2), 199–217. https://doi.org/10.1080/02673039408720783

Lees, L. (2007). Progress in Gentrification Research? *Environment and Planning A: Economy and Space*, *39*(1), 228–234. https://doi.org/10.1068/a39329

Lees, L. (2008). Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance? *Urban Studies*, *45*(12), 2449–2470. https://doi.org/10.1177/0042098008097099

Lees, L. (2009). Urban Renaissance in an Urban Recession: The End of Gentrification? *Environment and Planning A: Economy and Space*, *41*(7), 1529–1533. https://doi.org/10.1068/a4275

Lees, L., Morales, E. L., & Shin, H. B. (2016). *Planetary Gentrification*. Malden, MA: Polity Press.
Lees, L., Shin, H. B., & López-Morales, E. (Eds.). (2015). *Global Gentrifications: Uneven Development and Displacement* (1st ed.). https://doi.org/10.2307/j.ctt1t894bt

Lees, L., Slater, T., & Wyly, E. K. (2008). *Gentrification*. New York, NY: Routledge.

Lloyd, C. D. (2016). Spatial Scale and Small Area Population Statistics for England and Wales. *International Journal of Geographical Information Science*, *30*(6), 1187–1206. https://doi.org/10.1080/13658816.2015.1111377

Longley, P., Cheshire, J., & Singleton, A. D. (2018). *Consumer Data Research*. Retrieved from www.ucl.ac.uk/ucl-press

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

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.

Massey, D. (2017). Prologue: "World city". In X. Ren & R. Keil (Eds.), *The Globalizing Cities Reader* (pp. 1–512). Retrieved from https://doi.org/10.4324/9781315684871

Mendes, L. (2013). Marginal Gentrification as Emancipatory Practice: An Alternative to the Hegemonic Discourse of the Creative City? *RCCS Annual Review*, (5). https://doi.org/10.4000/rccsar.513

Moore, K. (2012, June 27). Is this the UK's Most Gentrified Street? *BBC News*. Retrieved from https://www.bbc.com/news/magazine-18394017

ONS. (2011). *2001 Census aggregate data*. Retrieved from http://dx.doi.org/10.5257/census/aggregate-2001-2

ONS. (2016). *2011 Census aggregate data*. Retrieved from http://dx.doi.org/10.5257/census/aggregate-2011-1

ONS. (2018). *Median house prices for administrative geographies: HPSSA dataset 9*. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianhousepricefornationalandsubnationalgeographiesquarterlyrollingyearhpssadataset09/current

Owens, A. (2012). Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent: CITY & COMMUNITY. *City & Community*, *11*(4), 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Parkinson, J. N., & Domokos, J. (2014, October 8). London's Housing Crisis: The Story of the Focus E15 mums. *The Guardian*. Retrieved from https://www.theguardian.com/society/video/2014/oct/08/london-housing-newham-mums-focus-e15-video

Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies*, *56*(5), 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*(8), 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*(5), 335–343. https://doi.org/10.1002/psp.585

Rose, D. (1984). Rethinking Gentrification: Beyond the Uneven Development of Marxist Urban Theory. *Environment and Planning D: Society and Space*, *2*(1), 47–74. https://doi.org/10.1068/d020047

Slater, T. (2011). Gentrification of the City. In G. Bridge & S. Watson (Eds.), *The new Blackwell companion to the city* (pp. 571–585). Malden, MA: Wiley-Blackwell.

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*(1), 2–9. https://doi.org/10.1068/a39322

The Resident. (2014, July 29). Area Focus: Fulham. Retrieved 15 August 2019, from https://www.theresident.co.uk/homes-interiors/area-focus-fulham/

UK Data Service. (2015, July 29). GeoConvert. Retrieved 11 July 2019, from http://geoconvert.mimas.ac.uk/

Van Criekingen, M., & Decroly, J.-M. (2003). Revisiting the Diversity of Gentrification: Neighbourhood Renewal Processes in Brussels and Montreal. *Urban Studies*, *40*(12), 2451–2468. https://doi.org/10.1080/0042098032000136156

White, I., & McLaren, E. (2009). The 2011 Census Taking Shape: The Selection of Topics and Questions. *Population Trends*, *135*(1), 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). Burlington, MA: Morgan Kaufmann.

Zuk, M., & Chapple, K. (2015a). *Urban Displacement Project*. Retrieved from https://www.urbandisplacement.org/

Zuk, M., & Chapple, K. (2015b). *Case Studies on Gentrification and Displacement in the San Francisco Bay area*. Retrieved from https://www.urbandisplacement.org/sites/default/files/images/case_studies_on_gentrification_and_displacement-_full_report.pdf

Supplementary Material

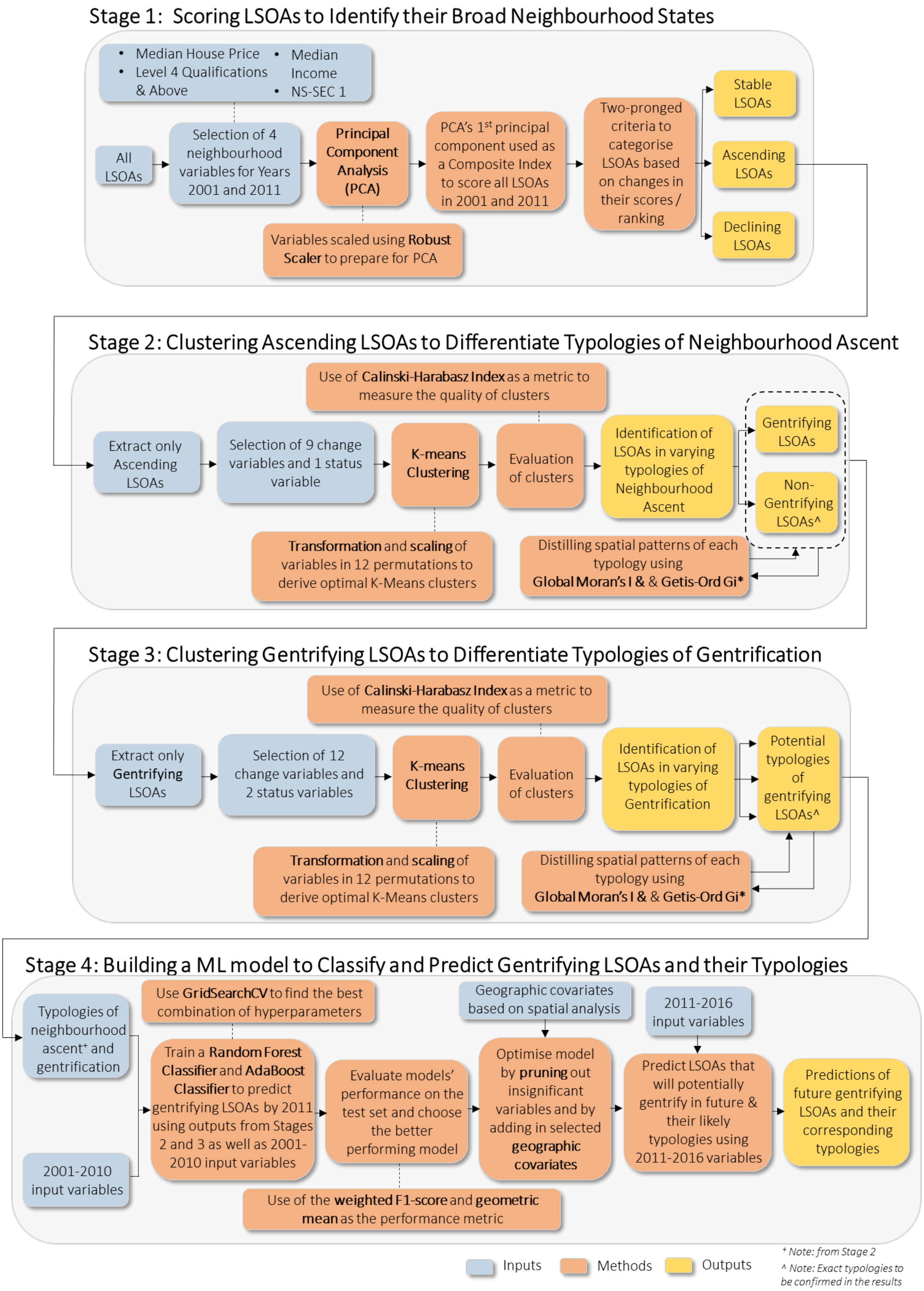
Identifying Neighbourhood Change and areas of Gentrification in London

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

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 in order to unpack the diversity and trajectories of gentrification in London. *Figure 1* illustrates the workflow diagrammatically while explanations of individual methods are detailed in the following paragraphs. The full datasets and analysis code are available for those wishing to reproduce the analysis, here[18].

[18] https://github.com/jytg17/Unpacking-the-Nuances-of-Londons-Neighbourhood-Change-Gentrification-Trajectories-codes

Figure 1. Overview of Entire Workflow



Following the examples of Reades et al. (2019) and Owens (2012), a handful of neighbourhood variables were 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.

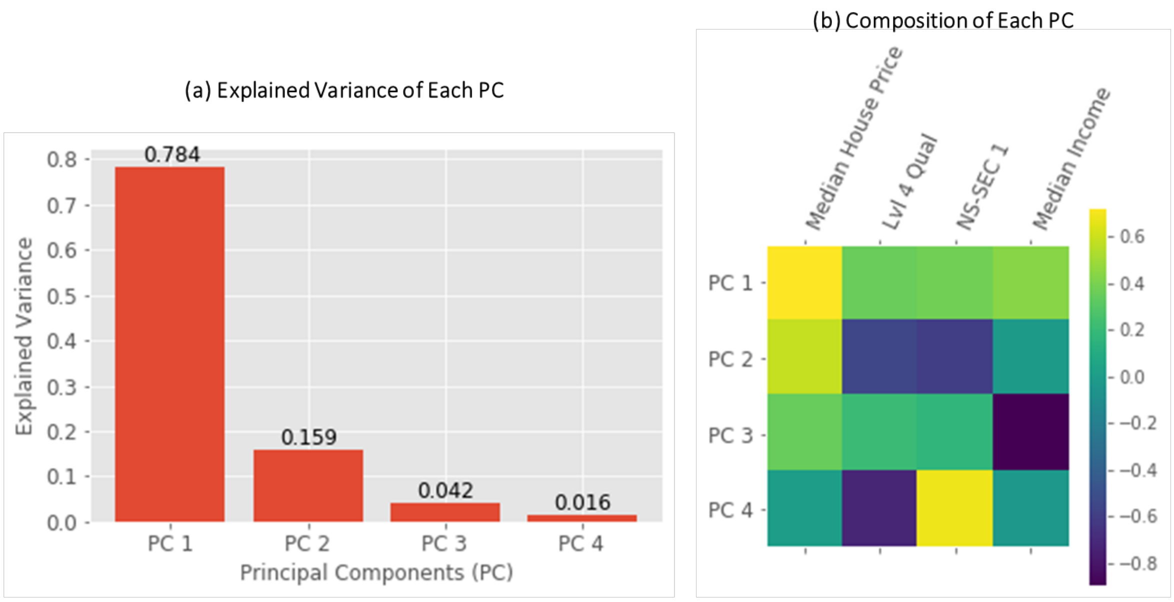
While this paper agreed with the first 3 variables and similarly adopted them, it is argued that the last-mentioned variable relating to the top-tier of occupational classes should be supplanted with that of the NS-SEC (i.e. NS-SEC Class 1) instead. Recognising that the NS-SEC went beyond pure employment 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 1* provides a tabulation of these 4 variable pairs.

Table 1. 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. The variables were first re-scaled then subsequently processed using Principal Component Analysis (PCA) 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), accounted for a reasonably high proportion of variance within the data for both years (78.4%) and its composition was well-spread across all 4 proxy attributes (*Figures 2a-b)*, so was adopted as the composite change index (CI) variable.

Figure 2. Statistics of Principal Components (PC) Derived From PCA



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) <u>Changes in Rank</u>

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

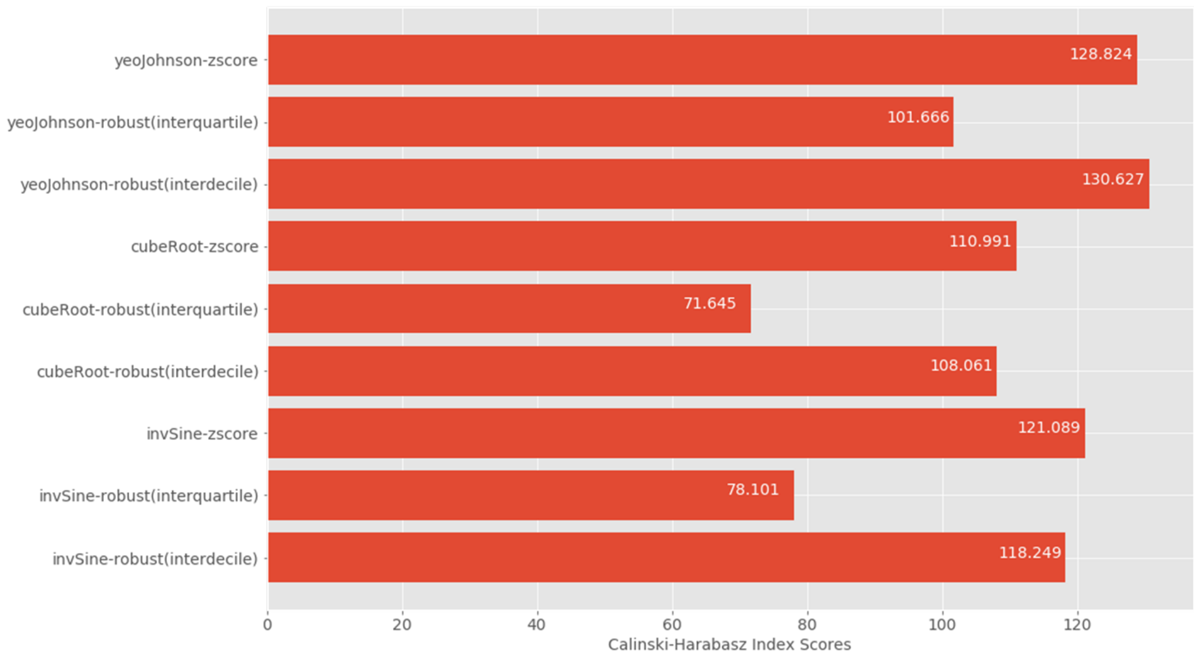
However, a limitation in singularly depending on rank changes to identify neighbourhood states was the bounded nature of ranking systems. LSOAs ranked sufficiently close to the extremities of the rank systems in 2001 were inadvertently restricted in their ability to move further upwards (for originally 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 Stage 2: Clustering Ascending LSOAs to Differentiate Typologies of Neighbourhood Ascent

With regard to the first round of cluster analysis to mine the situated typologies of neighbourhood ascent, a check for multi-collinearity amongst the variables surfaced 3 data features that fell into this category – *'youngPop_chg' (Change in % of residents aged below 65)*, *'agePop_chg' (Change in % of residents aged 65+)* and *'hse_age_chg' (Change in % of households made up of residents aged 65+)*. 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.

The Cube Root, Yeo-Johnson Power and Inverse-Hyperbolic Sine transformations applied on the variables successfully produced datasets that contained no heavily-skewed variables bearing a skewness score beyond ±1. These datasets were subsequently re-scaled using 3 different scalers and 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. 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 3)*.

Figure 3. 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.

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. Specifically, the K-Means clustering technique was adopted for discovering clusters within the data. 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. 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 2*.

Table 2. 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	-----

Similar to the framework utilised by the ONS (2015) in producing the classifications for OAs, various transformation and re-scaling techniques *(see Table 3 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 & Harabasz, 1974), was used to evaluate the quality of the clusters and the best scoring set was designated as the optimal outcome.

Table 3. 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	
	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	(Lin, 2017)
	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).

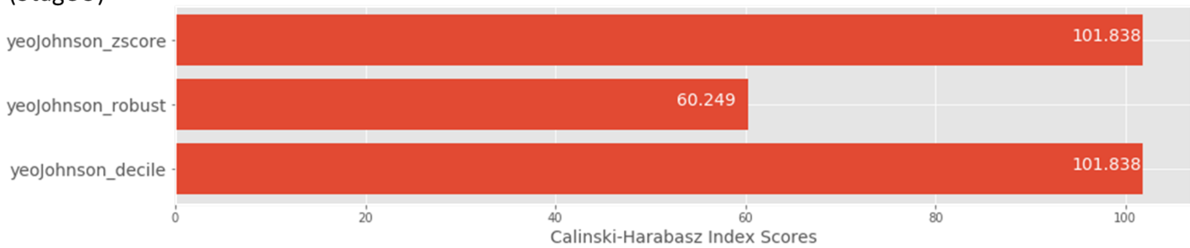
4.3 Stage 3: Clustering Gentrifying LSOAs to Differentiate Typologies of Gentrification

The data pre-processing for the 2nd phase of clustering revealed 2 highly correlated variables, *'inc_01' (Median Income in 2001)* and *'hseP_01' (Median House Prices in 2001)*. 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.

As for the data transformation, only the Yeo-Johnson Power transformation yielded a dataset containing no highly-skewed variables this round and was put through for re-scaling and clustering. Based on the silhouette analysis, 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.

Clusters derived from the dataset permutations were again measured for their quality using the Calinski-Harabasz Index. Judging from the results *(see Figure 4)*, 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 4. Calinski-Harabasz Index Scores for Clusters Derived Through Different Dataset Permutations (Stage 3)



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 4*.

Table 4. 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	-----

4.4 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.

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 5* 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 5. 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.

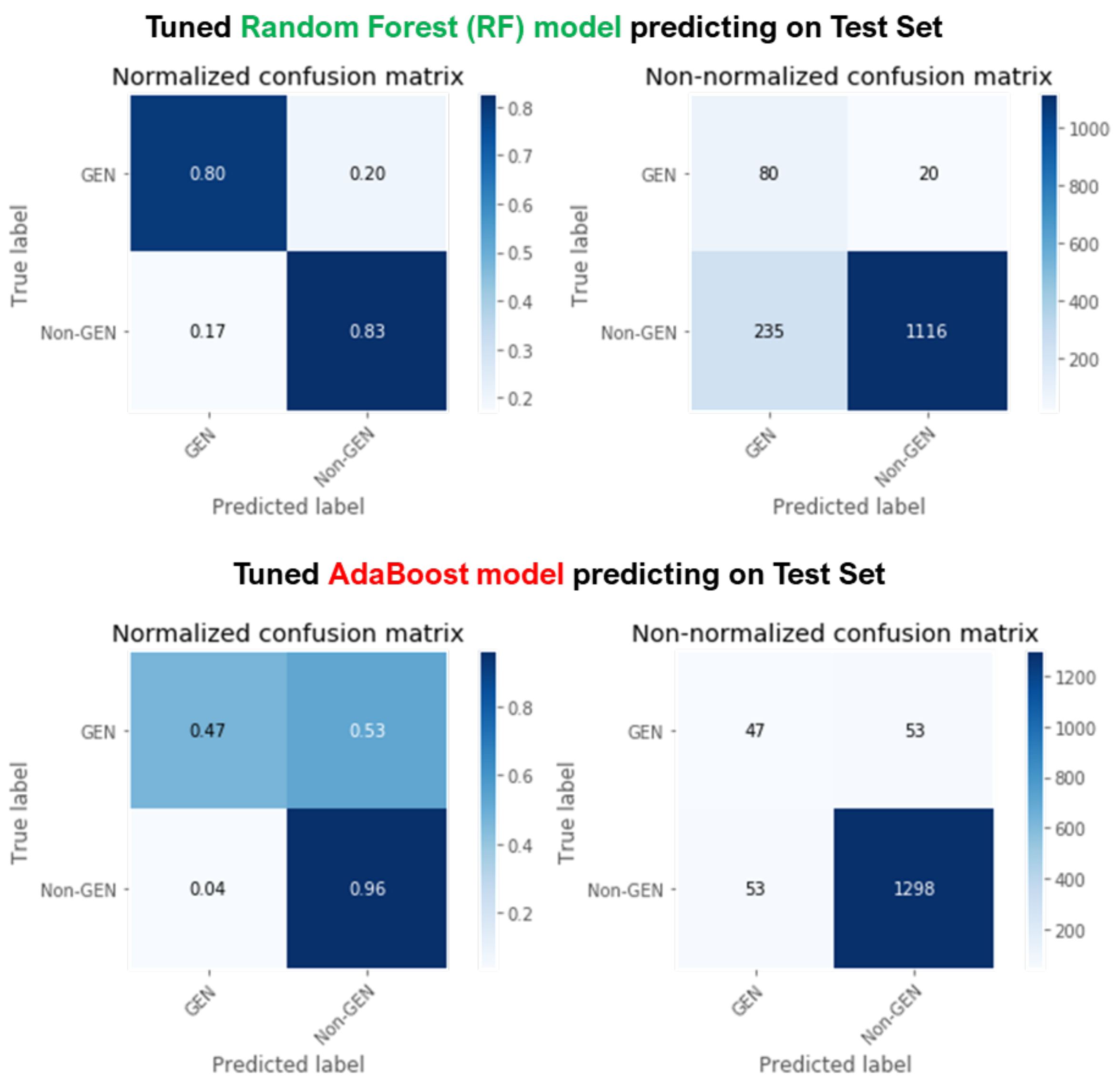
Separately, capitalising on the inherently spatial nature of gentrification, another feature engineering technique innovatively implemented by this study was the incorporation of 'geographic covariates' into the models (Bergen and Lindstrom, 2019, p.2). Acknowledging that ML models are generally aspatial 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.

Predicting Future LSOAs in Gentrification

With the full range of 2001-2011 neighbourhood variables[19] serving as inputs, two separate base models developed using Random Forest (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 aim being to predict gentrifying LSOAs in 2011. Selecting a test sub-set of the original data to enable model predictions to be compared to observed changes, 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 5)*. 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%).

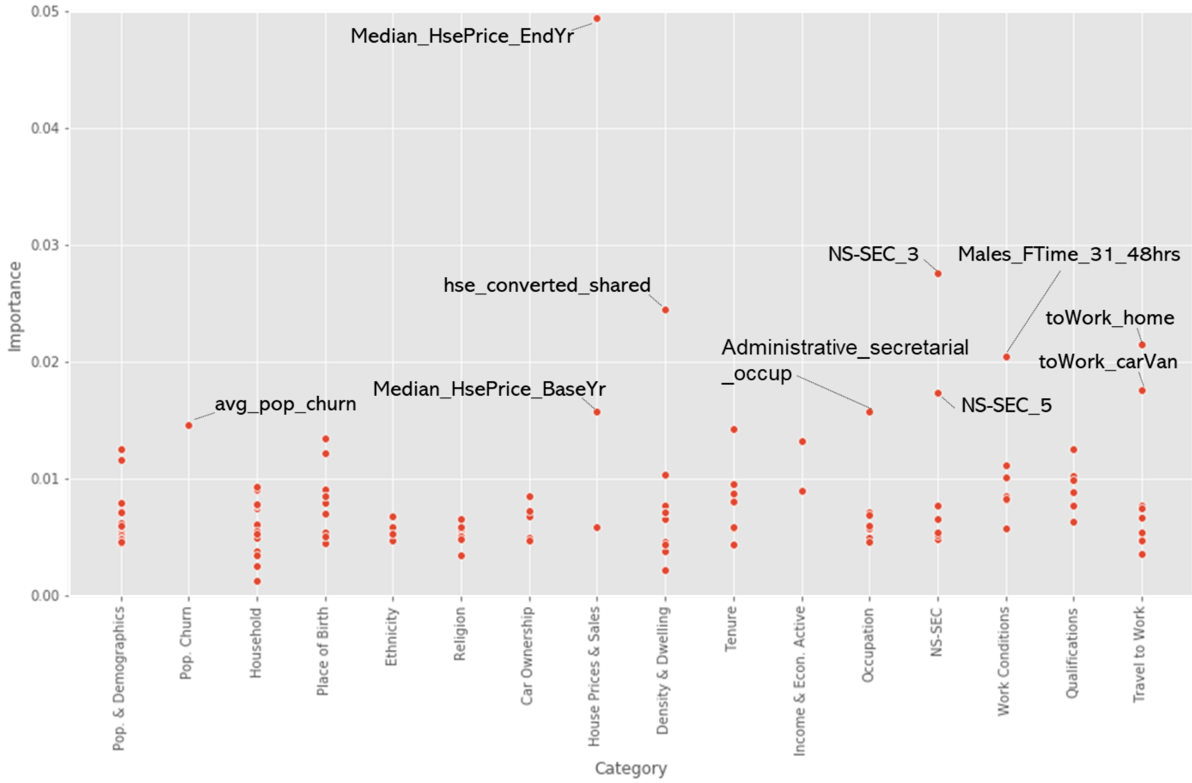
[19] Refer to Table 1 of the main text

Figure 5. 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. The top-10 most important variables for predicting gentrification are shown in *Figure 6*. 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.

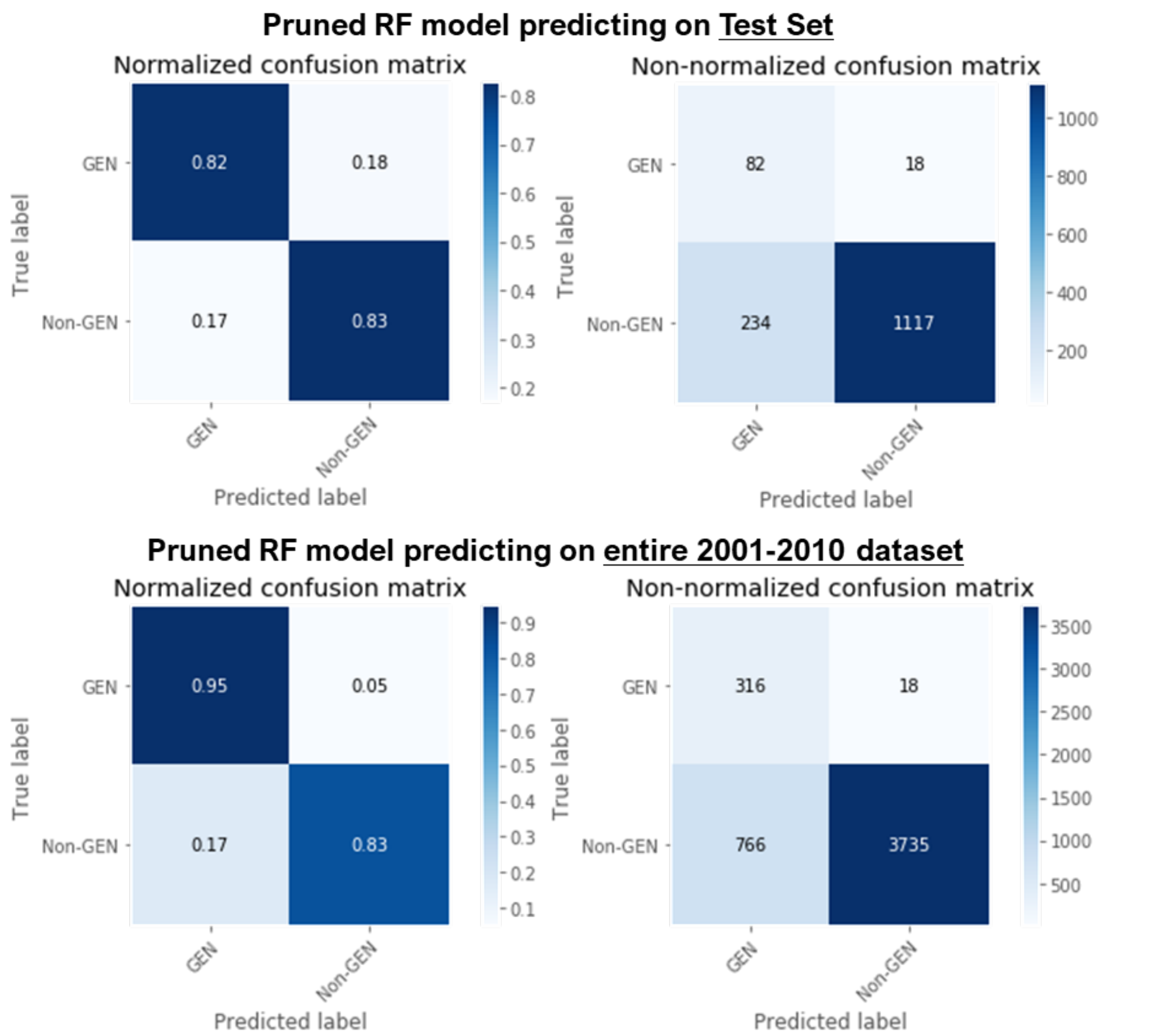
Figure 6. Variable Importance for Predicting Gentrifying LSOAs (Stage 2)



Pruning was subsequently carried out 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.

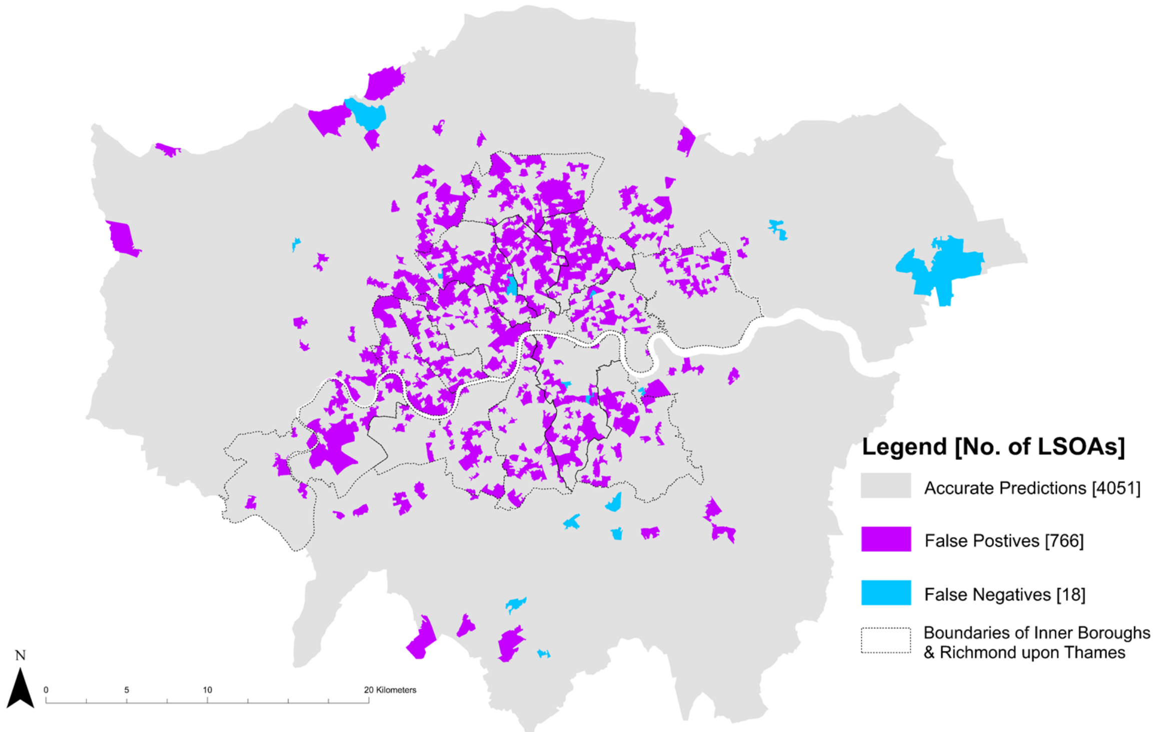
The pruned model was applied not only to the test set, but the entire dataset as well, to obtain a sense of how the model had improved and its overall performance respectively. Based on the results in *Figure 7*, the pruned model was more effective than its unpruned version in predicting gentrifying 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 7. 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 two 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 8)*. 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.

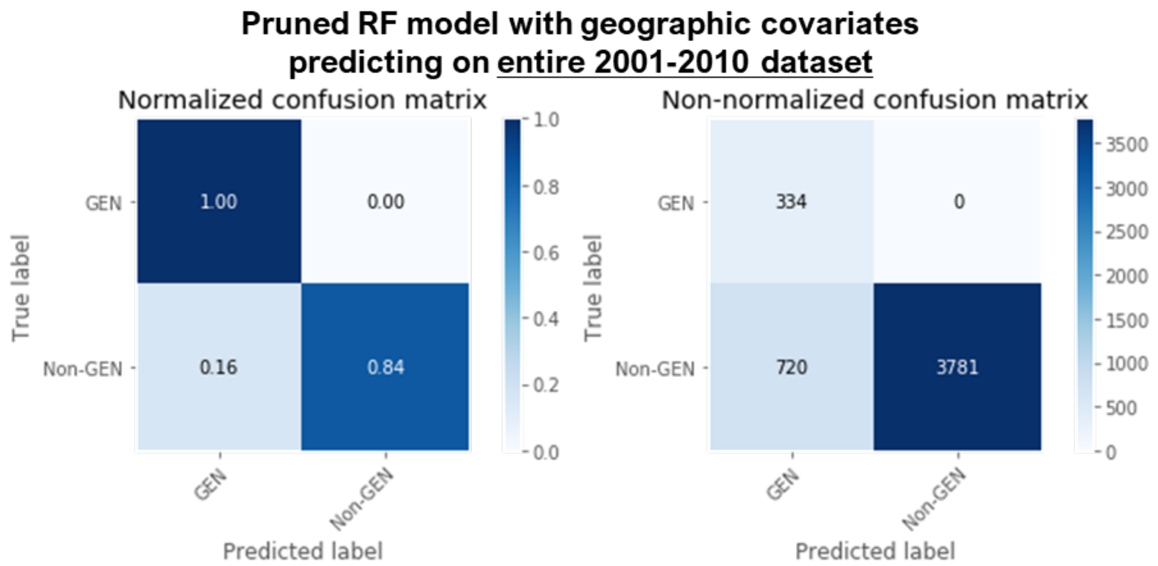
Figure 8. Model's Errors in Predicting Gentrifying LSOAs (Before Pruning)



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. 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).

After including both geographic covariates, the RF model was tested again on the entire dataset. The results *(see Figure 9)* 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.

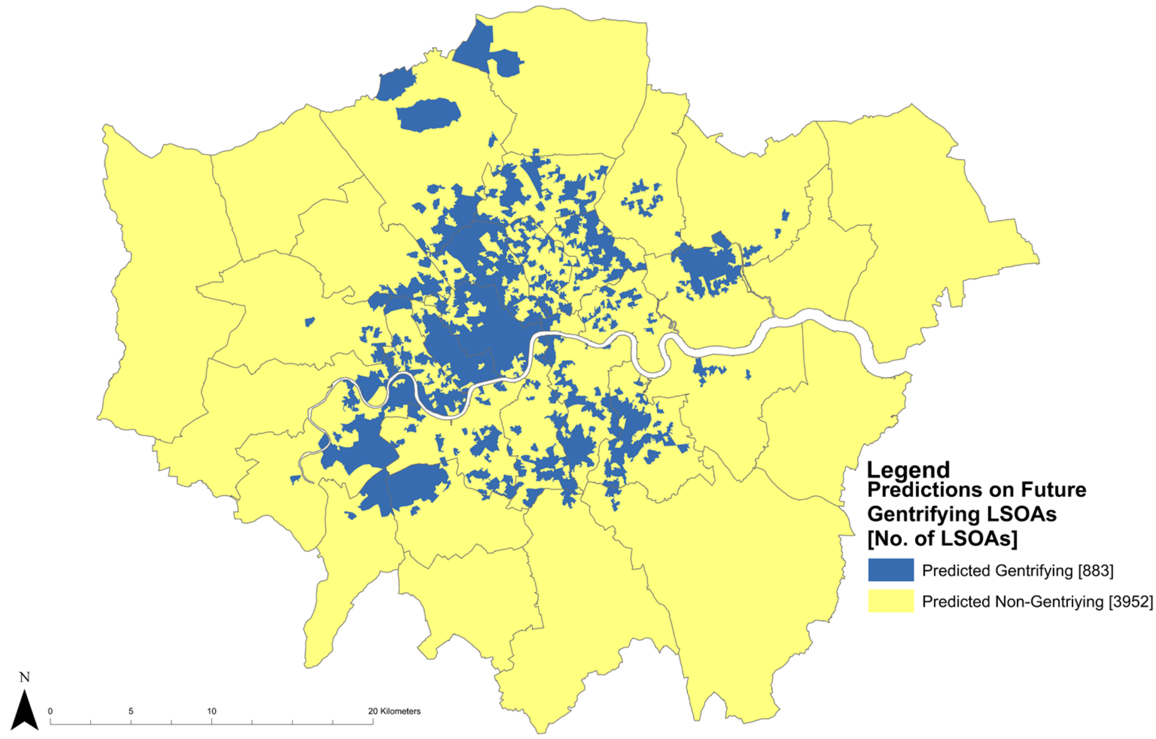
Figure 9. Accuracy of Pruned Model with Geog. Covariates Predicting on Entire 2001-2010 Dataset



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 10.*

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.

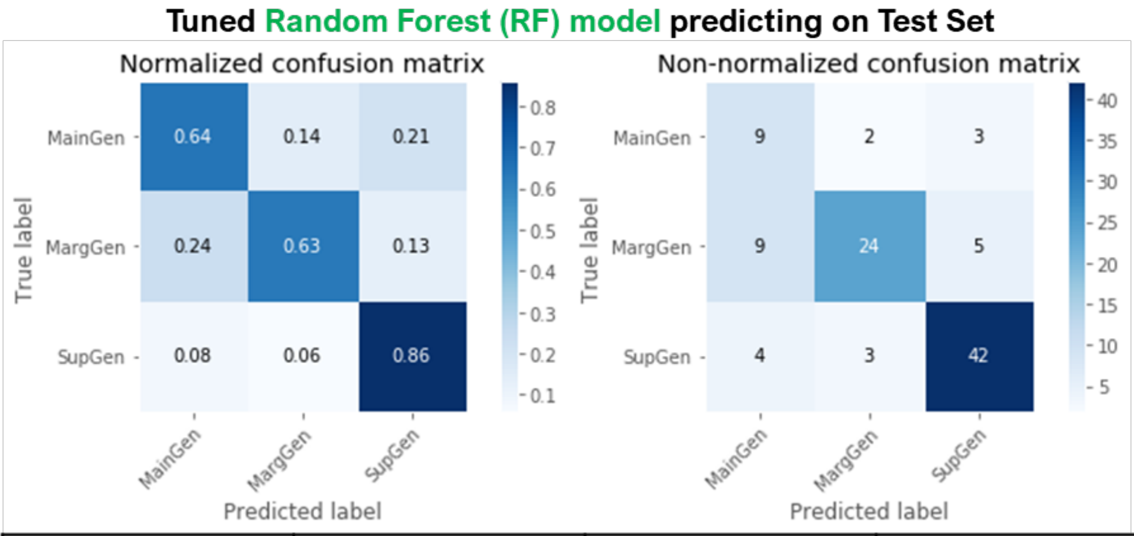
Figure 10. Model's Predictions on Future Gentrifying LSOAs



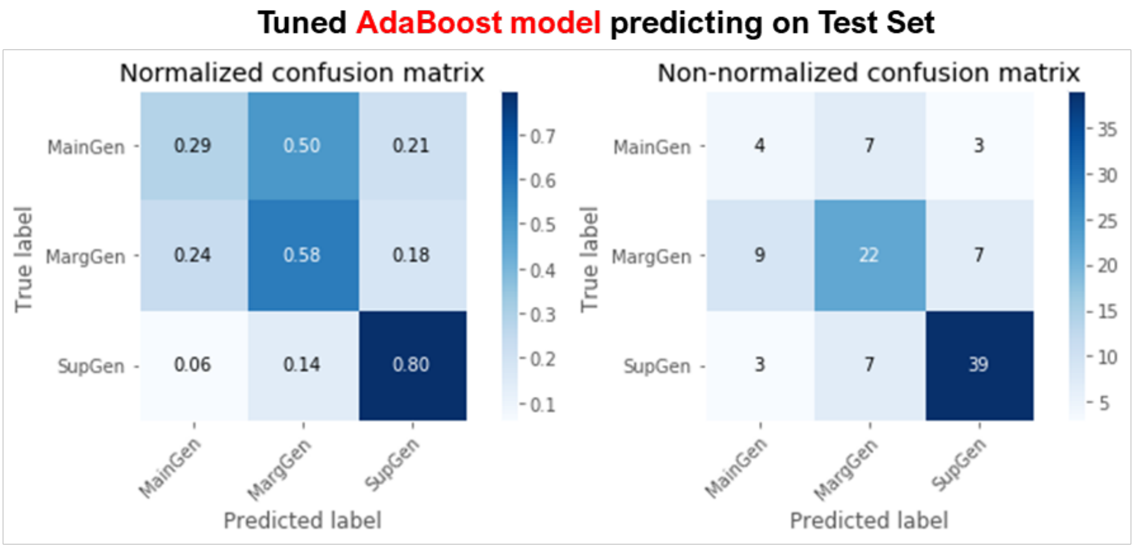
Predicting Typologies of Future Gentrifying LSOAs

By way of tuning and testing another pair of 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 11)*.

Figure 11. Base Models' Optimal Hyperparameters and Prediction Accuracy on Test Set



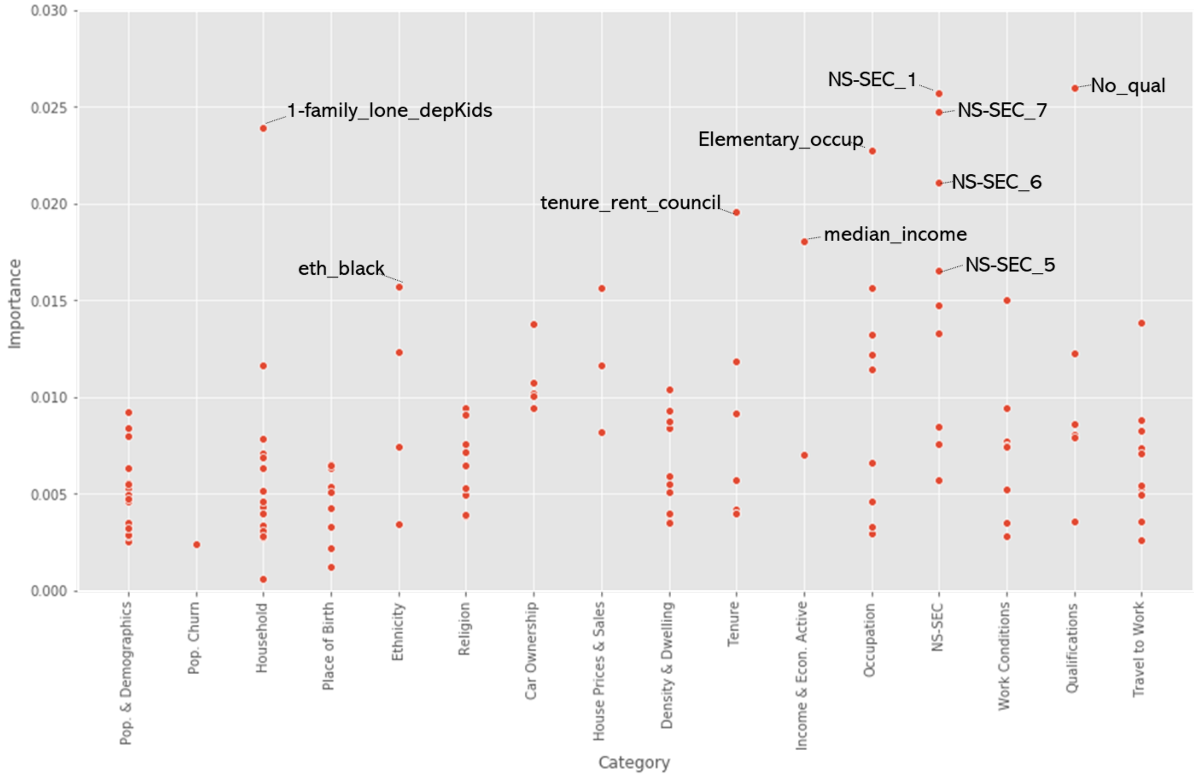
RF Hyperparameter	No. of Trees	Max. Features	Criterion
RF Best Candidate	230	Log2	Entropy



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

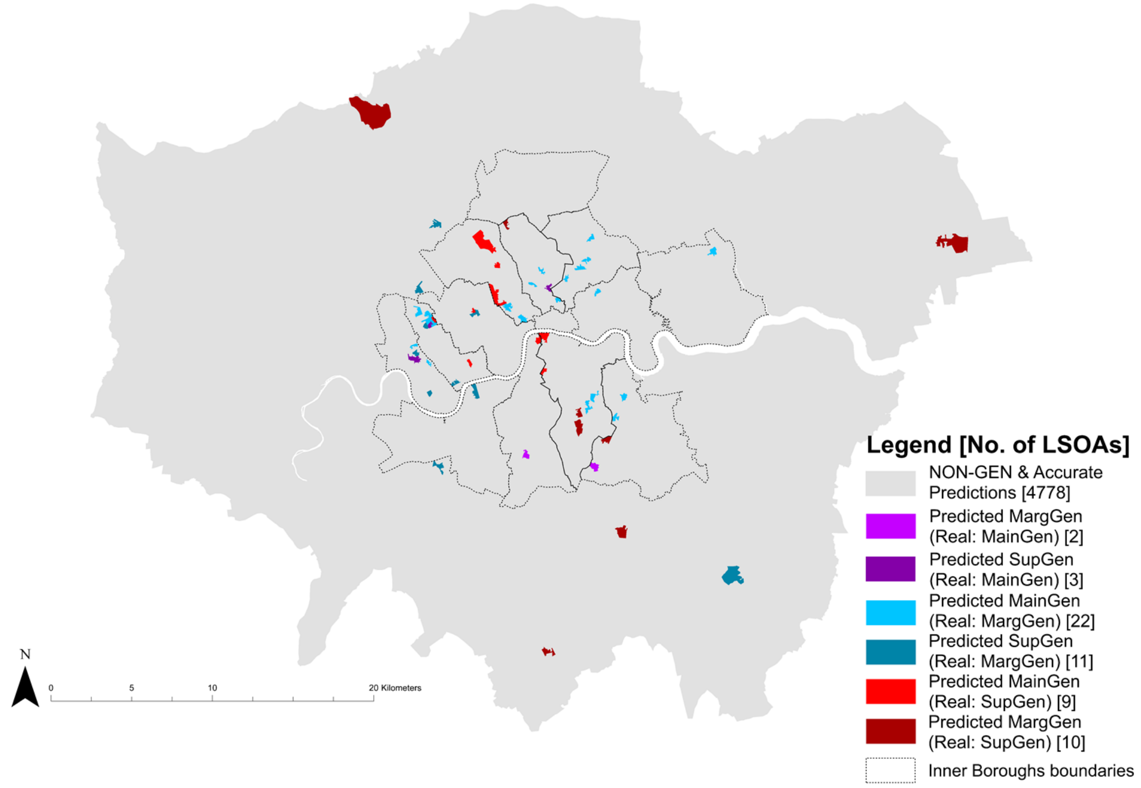
The most important variables for predicting the diversified typologies of gentrification were extracted from the RF model, which revealed that NS-SEC variables as dominant within the top-10 *(see Figure 12)*. To prune variables, tests established that removing the 53 least important variables optimised the model's predictions.

Figure 12. Variable Importance for Predicting Typologies of Gentrifying LSOAs



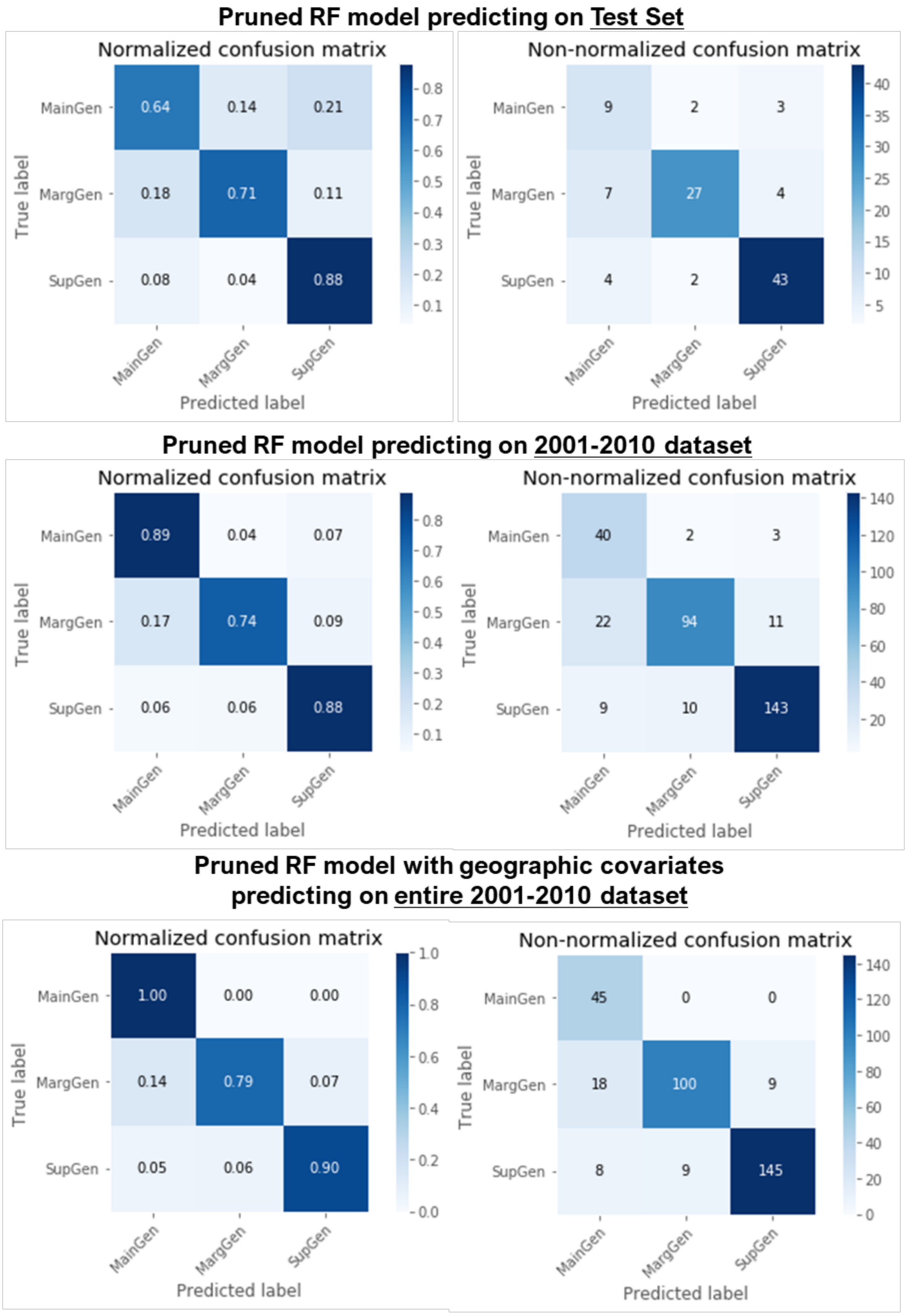
Geographic covariates were subsequently formulated and added into the model, in a manner analogous to the previous modelling process. However, as the spatial configurations of this model's erroneous predictions were noticed to reside mainly within the inner boroughs *(see Figure 13)*, the 1st geographic covariate for this phase was tweaked to be based on the distance between each LSOA and the inner borough boundaries.

Figure 13. Model's Errors in Predicting LSOAs' Gentrification Typologies



The performances of the model at every step of its evolution are provided in *Figure 14*. 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 14. Accuracy of Model at Every Stage of the Refinement Process



References

Alexiou, A., & Singleton, A. (2015). Geodemographic Analysis. In C. Brunsdon & A. D. Singleton (Eds.), *Geocomputation: A Practical Primer* (pp. 136–151). https://doi.org/10.4135/9781473916432

Anselin, L. (2019). Local Spatial Autocorrelation—Univariate Local Statistics [Workbook]. Retrieved 2 August 2019, from https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html

Ben Fraj, M. (2017). In Depth: Paramter Tuning for Random Forest. Retrieved 18 July 2019, from In Depth: Parameter Tuning for Random Forest—All Things AI - Medium website: https://medium.com/all-things-ai/in-depth-parameter-tuning-for-random-forest-d67bb7e920d

Bergen, S., & Lindstrom, J. (2019). Comprehensive Tutorial for the Spatio-Temporal R-package [E-Book]. Retrieved 7 August 2019, from https://cran.r-project.org/web/packages/SpatioTemporal/vignettes/ST_tutorial.pdf

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Cox, N. J. (2011). Stata Tip 96: Cube Roots. *The Stata Journal*, *11*(1), 149–154.

ESRI. (n.d.). How Spatial Autocorrelation (Global Moran's I) works—Help | ArcGIS Desktop. Retrieved 20 July 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

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

Jaadi, Z. (2019, May 24). A Step by Step Explanation of Principal Component Analysis. Retrieved 5 July 2019, from Medium website: https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2

Navlani, A. (2018). AdaBoost Classifier in Python. Retrieved 18 July 2019, from AdaBoost Classifier in Python (article)—DataCamp website: https://www.datacamp.com/community/tutorials/adaboost-classifier-python

ONS. (n.d.). The National Statistics Socio-economic classification (NS-SEC). Retrieved 12 July 2019, from https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010

ONS. (2015). *Methodology Note for the 2011 Area Classification for Output Areas*. Retrieved from https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables

Lin, J. (2017, July 23). Feature Scaling. Retrieved 1 July 2019, from https://jovianlin.io/feature-scaling/

Owens, A. (2012). Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent: CITY & COMMUNITY. *City & Community*, *11*(4), 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. *Urban Studies*, *56*(5), 922–942. https://doi.org/10.1177/0042098018789054

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed). Burlington, MA: Morgan Kaufmann.

Yeo, I.-K., & Johnson, R. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954