Beyond The Tracks: Evaluating Transport-led Regeneration Impact on Businesses' Dynamics, Sectoral Shifts, and Agglomeration, The Case of Northern Line Extension.

By

Lama Alswliman

Supervised by Prof. Adam Dennett

CASA0010 - Dissertation

Submission Date: 22-8-2025

Word Count: 11,885

This dissertation is submitted in part requirement for the MSc in Urban Spatial Science in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

Abstract

In recent decades, strategic transport investment has been increasingly employed to support regeneration initiatives and stimulate economic activity. Yet, their impact on businesses' dynamics, sectoral shifts, and agglomeration remains unexplored, particularly in London context. As these projects are usually justified based on their economic impact, assessing their outcome is essential to better inform policymakers. This research investigates the impact of the Northern Line Extension on business dynamics, sectoral composition, and agglomeration in Vauxhall, Nine Elms, and Battersea Opportunity Area. Drawing on unique firm-level datasets, including the Business Census and Business Rates, and applying methods like Adjusted Interrupted Time Series and Spatial Clustering (HDBSCAN), this analysis evaluates businesses' dynamics, agglomeration patterns, and sectoral shifts, focusing on two points in time (construction starting year and opening year) while considering trends following those points. The findings suggest a redistribution of economic activity rather than net growth, with retail sectors benefiting the most. The anticipation effect mainly drove this pattern. At the same time, the findings indicate increased agglomeration of businesses around the stations, specifically in the form of urbanisation economies, reflecting a diverse mix of industries. The study contributes to understanding the impact of transport-led regeneration by providing one of the first evaluations of this impact in London and offers insights into future urban policy.

Keywords: transport investment, firms, agglomeration, new firm creation, economic activity redistribution, anticipation effect, location decision, stations, startup, cluster.

Declaration

I hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 11,885 words in length. I also acknowledge the use of ChatGPT for research note summarisation and Grammarly for draft proofreading. Finally, accounting for research reproducibility, all code used in this analysis is available on GitHub at: (Dissertation_NLE_Impact_Analysis_TFL) Lastly, for reproducibility purposes, and as some datasets used in this analysis are safeguarded or private, limiting the ability to be shared, a mock dataset is provided instead. This means, results produced on the back of such mock data will not to any extent match the behaviour of the real dataset.

Lama Alswliman

Contents

1 Introduction: 13

1.1 Research Focus and Aim: 13

1.2 Research Structure: 14

2 Literature Review: 15

2.1 Transport Investment as a Catalyst for Regeneration: 15

2.2 London Policy: 15

2.3 Methods Transport Investment Affects Economic Activity: 16

2.4 Economic Activity Measures: 18

2.5 Debates Around Transportation Impact: 20

2.6 Conclusion: 21

3 Data and Area: 22

3.1 Area of Study: 22

3.2 Data: 23

3.3 Spatial Unit: 27

4 Methodology 28

4.1 Data Preprocessing 28

4.2 Features Engineering 33

4.3 Model Selection and Evaluation 35

4.3.1 Control Area Selection: 35

4.3.2 Modelling Techniques: 36

4.4 Ethical Considerations: 43

5 Results: 44

5.1 Overall Businesses AITS Models Results 44

5.2 Industry Specific Businesses AITS Models Results 49

5.3 Models Evaluation 52

5.4 Businesses Clustering 54

6 Discussion: 61

7 Conclusion: 65

Appendix A: 71

Appendix B: 80
Appendix C: 82

List of Figures:

Figure 1: Northern Line Extension Scheme with VNEB Opportunity Area Border. **Source**: (TFL, 2013) 23
Figure 2: PTAL Score for VNEB Opportunity Area, on the Left for year 2015, while on the right is a projection for year 2031. **Source:** (TFL, 2013a, 2015) 26
Figure 3: Methodology flow chart 28
Figure 4: population over time VNEB vs OKR 30
Figure 5: Ethnicity changes over time, VNEB vs OKR 30
Figure 6: Residential Churn VNEB vs OKR 31
Figure 7: In and Out deprivation Index VNEB vs OKR 31
Figure 8: Job count over time VNEB vs OKR 32
Figure 9: Property prices/rents over time VNEB vs OKR 32
Figure 10: Business counts/rates over time VNEB vs OKR 34
Figure 11: Average business distance (dispersion) over time VNEB vs OKR 34
Figure 12: Old Kent Road opportunity area border and proposed stations for the Bakerloo Line Extension **Source:** (Southwark Council, 2017) 35
Figure 13: Moved business origin 37
Figure 14: VNEB business count by accessibility over time 38
Figure 15: Moved Business vs Independent Variables Joint Distribution Plots 40
Figure 16: New Business vs Independent Variables Joint Distribution Plots 41
Figure 17: VNEB vs OKR OAs Asian Ethnicity, Business Counts, and Average Business Distance of 2015-2021 42
Figure 18: Correlation heatmap 44
Figure 19: Overall business AITS coefficient 48
Figure 20: industry-specific business AITS coefficient for new business 50
Figure 21: industry-specific business AITS coefficient for moved businesses 51
Figure 22: Model residuals plots 53
Figure 23: Business density over time 54
Figure 24: Business density over time by NLE stations 54
Figure 25: business density by station and sector 55
Figure 26: Clusters Silhouette score over time and search space 56
Figure 27: Business cluster count over time 56
Figure 28: Cluster and business counts within stations buffers 57
Figure 29: Cluster and business counts within stations buffers by sector 57
Figure 30: Cluster sector/accessibility diversity 58
Figure 31: agglomeration types of clusters over time 58
Figure 32: Distance of agglomeration-type clusters from stations 59
Figure 33: Map of business clusters over time (years 2012, 2015 and 2018) 59
Figure 34: Map of business clusters over time (years 2021 and 2024) 60

Figure 35: Models ranking **Source:** (*MTEB Leaderboard - a Hugging Face Space by mteb*, no date)..... 80
Figure 36: Embedding and Cosine similarity method **Source:** (Sivaprakasam, 2025) 81

List of Tables:

Table 1: Overview of used datasets 24
Table 2: Key variables of Business Census data 25
Table 3: Key variables of WHYTHAWK data 25
Table 4: Business features list 33
Table 5: Explanation of AITS model formula symbols 36
Table 6: Studied dependent Variables 37
Table 7: AITS independent variables list 39
Table 8: VIF removed features 44
Table 9: Overall new and moved business AITS model coefficients 45
Table 11: Model-fit evaluation metrics 52
Table 12: Overall new and moved business AITS model coefficients 72
Table 13: Retail business AITS model coefficients 74
Table 14: Office business AITS model coefficients 75
Table 15: Industrial business AITS model coefficient 76
Table 16: Leisure business AITS model coefficient 78
Table 17: Industry-specific model-fit evaluation metrics 79

List of Acronyms and Abbreviations:

TFL	Transport For London
GLA	Greater London Authority
VNEB	Vauxhall, Nine Elms, Battersea Opportunity Area
VNEB-OAPF	Vauxhall, Nine Elms, Battersea Opportunity Area Planning Framework
OKR	Old Kent Road
NLE	Northern Line Extension
JLE	Jubilee Line Extension
PTAL	Public Transport Access Level
CAZ	Central Activity Zone
OA	Output Area
LSOA	Lower Super Output Area
VOA	Valuation Office Agency
ONS	Office National Statistics
CDRC	Consumer Data Research Centre
BPS	Battersea Power Station
AITS	Adjusted Interrupted Time Series
VIF	Variance Inflation Factor
UK	United Kingdom
USA	United States of America

List of Definition:

Term	Description
New Business	Businesses that are newly formed and registered
Moved Business	Businesses relocated into the VNEB Opportunity Area from the rest of London boroughs, outside of London, Wandsworth and Lambeth boroughs (excluding the opportunity area), and from low accessibility areas to high accessibility areas inside the border of the VNEB Opportunity Area.
PTAL	Accessibility score used to quantify how well a location is connected to the public transport network.
High Accessibility	6b, 6a, 5 PTAL scores are classified as high accessibility
Low Accessibility	4, 3, 2, 1b, 1a PTAL scores are classified as low accessibility.
Urbanisation	Businesses from different sectors cluster together. In this analysis it is also referred to as diverse clusters.
Localisation	Businesses from the same sectors cluster together. In this analysis it is also referred to as homogeneous clusters.
Average Business Distance	Extra measure of agglomeration calculated by the average companies' distance from the business centre for each OA
Average business distance per industry	Extra measure of agglomeration calculated by the average companies' distance from the business centre per industry for each OA
Treatment	Dummy variable that captures differences between treatment and control areas.
Intervention	Dummy variable that captures differences before and after the intervention (construction or opening year intervention)
Treatment post-intervention	Interaction variable of treatment and intervention capturing differences before and after the intervention between treatment and control areas.
Trend	A variable that captures the dependent variable trends over time
Trend post-intervention	A variable that captures the dependent variable trends over time after the intervention occurred
Trend- treatment post-intervention	Interaction variable of trend, treatment and intervention capturing differences in trends before and after the intervention between treatment and control areas.

Acknowledgements:

My heartfelt thanks go to my supervisor, Dr. Adam Dennett, for his insightful guidance, encouragement, and invaluable support in shaping every stage of this dissertation. Most of all, I am grateful for his help in enabling me to explore a topic that I am passionate about. I am also thankful for the Transport for London (TFL) team for their support and for giving me the opportunity to work on such a unique project. I would like to thank Gavin Chait for his support and for sharing his data (Whythawk), which was a big part of my analysis.

Most importantly, I am deeply grateful for my partner, Ayham. Without his unwavering support, patience, and encouragement to pursue my passion and dreams, I could not accomplish this. I owe it all to him. Lastly, I would like to thank my mom and dad for their unwavering belief in me and for their continuous emotional support, as well as my sisters and brother for their support and encouragement.

1 Introduction:

1.1 Research Focus and Aim:

Rapid urbanisation and cities' expansion have posed various challenges. Governments face increasing pressure to accommodate the population's growing demands, including economic vitality, inclusive development, and effective land use. To address these challenges, policymakers have employed regeneration plans supported by strategic transport investment as a tool for urban development, catalysing economic growth. Therefore, substantial funding is allocated for these investments to aid regeneration initiatives. For example, the UK government recently announced a long-term capital investment of £2.167 billion to support TFL's capital renewal from 2026 to 2030 (DFT, 2025). These investments are justified by their potential to foster economic growth, productivity and agglomeration. Therefore, assessing their impact is essential for evaluating whether they met their objectives in delivering economic, social and spatial benefits. These insights enable policymakers to develop more effective investment strategies and support evidence-based monitoring and evaluation of regeneration outcomes.

Researchers and economists have extensively studied the transportation investments' impact at the regional level, consistently finding a positive link with productivity, employment, and output growth. However, their impact on the micro local economy remains insufficiently researched. Scholars have examined these investments' effect on local economic activities, focusing mainly on changes in property prices and employment, due to their measurability and relevance to policy objectives. Instead, limited attention has been given to changes in the businesses' demographics and dynamics, despite their significant role in gauging economic vitality. Furthermore, the transport investments' impact remains a subject of debate among researchers, who argue the overestimation of its effects on business dynamics, and suggest that it results in a reorganisation of economic activities rather than a genuine overall economic activity increase (Chatman and Noland, 2011).

Recently, the UK has seen significant growth in transport infrastructure investments, primarily supporting regeneration initiatives within London. The London Plan underscores transport investment as a crucial catalyst for regeneration and economic development, especially in designated growth zones. Despite this surge, a limited body of research has examined the impact of transport investments on London context.

Ultimately, this study aims to contribute to the transport economics by using the Northern Line Extension in Vauxhall, Nine Elms, Battersea Opportunity Area as a case study. It leverages unique datasets, mainly Business Census and Business Rate, to study businesses' dynamics, sectoral shifts, and business concentration and agglomeration. This study employs statistical methods, including Adjusted Interrupted Time Series and spatial clustering, to address the research aim and objectives. This dissertation has the following research aim:

Examine the transport investments' impact on businesses' demographics and dynamics, evaluate their effect on local economic activity, and assess which economic sectors experience the greatest benefits.

The above aim results in the following research objectives:

- Examine the transport investments' impact on shaping the businesses' dynamics and demographics in the different accessibility levels.
- Evaluate whether these investments generate a net business growth or redistribute existing economic activities and assess the role of the anticipation effect in driving the observed behaviour.
- Examine if transport investment drives sectoral shifts in targeted areas and around stations.
- Evaluate if the impact is more profound at the investment announcement or the actual operational date.
- Assess the impact of transport investment on changes in business density and clustering patterns around stations, identifying any agglomeration effect.
- Assess the level of transport-led regeneration alignment with policy goals.

1.2 Research Structure:

Chapter 2 reviews relevant existing literature on how transport investment boosts local economic activities and explores debates around transport investment.

Chapter 3 explores the datasets used and the reasoning behind study area and spatial unit decisions.

Chapter 4 describes the data preprocessing steps used in this research, as well as the methods employed in the analysis.

Chapter 5 presents the results.

Chapter 6 discusses the results, the analysis faced limitations, potential areas for future research work and policy recommendations.

Chapter 7 presents the research conclusion.

2 Literature Review:

2.1 Transport Investment as a Catalyst for Regeneration:

Over the years, the United Nations has identified a global trend of increasing urbanisation, intensifying demand for infrastructure, housing, services, and economic opportunities (United Nations Department of Economic and Social Affairs, 2018). In response, transport-led regeneration has emerged as an urban policy instrument to accommodate and direct urban growth. As a strategic approach, it employs targeted transport infrastructure investments, such as new railways or stations, as a catalyst for urban regeneration. This strategy aims to optimise land use, stimulate economic growth, and improve environmental sustainability in the targeted area (Department for Transport, 2021).

Transport investments can play a pivotal role in restructuring the urban fabric, particularly in areas marked by deprivation, underutilisation, or socio-spatial inequalities, conditions often exacerbated by the absence of transport infrastructure. By increasing accessibility, such schemes can boost private investment, attract businesses, unlock development, and enhance labour market access (D. Knowles and Ferbrache, 2016). However, such strategies' impact remains a subject of scholarly debate, owing to their complexity. While the benefits can be significant, a vast body of research has highlighted potential adverse outcomes. Such investment can alter neighbourhood dynamics and profiles, leading to gentrification, displacement and intensification of social and economic inequalities (Zuk *et al.*, 2018; Lin and Yang, 2019).

Recently, a growing number of governments, including the UK, Australia, and the United States, have been incorporating strategic transportation investments into their urban policies and frameworks. They promote these investments as ways to stimulate economic activity, attract investment, create jobs, and support regeneration, while aligning with the principles of Smart Growth and New Urbanism (Dittmar and Ohland, 2003; HM Treasury, 2020; FTA, 2025). This policy shift emphasises the importance of evaluating how these transportation investments impact economic activity changes within urban areas. The following section discusses transport investments' role within London policy.

2.2 London Policy:

The Docklands area regeneration is one of the earliest examples of transport-led regeneration in London. The Docklands Light Railway (DLR) significantly improved connectivity between that area and Central London, thereby stimulating economic development and attracting private investment. However, (Church, 1990) critiqued the absence of a holistic transport plan integrated during the early regeneration phase, arguing that investments were reactively implemented under investors' pressure following the increased demand.

In contrast, transport investment is now a core component of The London Plan. With London's population projected to exceed ten million by the mid-2030s, the Mayor of London

identified Opportunity Areas[1] as key zones for accommodating this growth. However, some Opportunity Areas lack adequate transportation infrastructure, leading to stalled development and limited investments (GLA, 2015). To address these challenges, the Mayor of London proposed targeted transport investments as a regeneration and economic growth catalyst.

Within the framework of the 'Good Growth', The London Plan (Greater London Authority, 2021) positions strategic transport investment as a mechanism for driving economic development, business growth, and agglomeration across the city, while promoting equitable prosperity. This is achieved through focused investments in the Opportunity Areas, thereby extending London's Central Activity Zone (CAZ)[2] economic agglomeration benefits. Through targeted enhancement in accessibility, such investments unlock developments and support the strategic plan of transforming Opportunity Areas into hubs for employment and commercial development.

2.3 Methods Transport Investment Affects Economic Activity:

Transport investments impact local economies through four mechanisms: expanding the labour market and productivity, increasing accessibility, promoting agglomeration, and broadening market demand(Zhang and Cheng, 2023). For this research, the focus is on discussing accessibility and agglomeration.

Agglomeration:

Transport investment can shape the local economic activities by influencing spatial relationships, increasing accessibility, and reducing effective distance. Consequently, these changes affect the economic geography by creating new opportunities and increasing proximities, leading firms to cluster (Laird and Venables, 2017). Agglomeration refers to the spatial concentration of economic activities, where businesses benefit from being near others through knowledge spillover, shared resources and infrastructure, and labour market pooling, resulting in increased productivity. Spatial clusters' benefits are known as agglomeration externalities. Evidence suggests that transport investment induces economic agglomeration by increasing accessibility, elevating economic productivity, especially in larger cities, which tend to see bigger gains (Venables, 2007; Chatman and Noland, 2014). However, Venables (2007) argues that such gains generate economic surplus, which may be capitalised into land and property values. Primarily, two agglomeration mechanisms have been identified based on within-industry concentration vs cross-industry diversity. Localisation economies occur when productivity gains arise from firms co-locating within the same industry cluster. In contrast, Urbanisation economies are characterised by productivity gains arising from the clustering of cross-sectoral industries (Chatman and Noland, 2011). These mechanisms help illustrate why certain areas thrive economically and how they shape businesses' dynamics. Moreover,

[1] Opportunity Areas, according to The London Plan,' are areas with development capacity to accommodate new housing, commercial development and infrastructure linked to existing or potential improvement in public transport connectivity.'

[2] The London Plan identified CAZ as London's commercial core, encompassing the central business district and employment hubs. Thus, it is a key driver for London's economic growth.

agglomeration can attract new firms, increase competitiveness, and influence the sectoral shifts of local economies. Chatman, Noland and Klein (2016) investigated the influence of new light rail on firm formation, while considering agglomeration effects. His findings indicate that urbanisation economies positively correlate with new firm formation, while localisation economies have the opposite effect.

Scholars argue that transportation investments may influence business density and shape agglomeration typology, specifically around stations. In their analysis of Metrosur, (Mejia-Dorantes, Paez and Vassallo, 2012) found that areas surrounding new stations witnessed higher business densities overall, specifically retail activities, exhibiting the most significant clustering gains around stations. Similarly, (Song *et al.*, 2012) investigated the relationship between sectoral agglomeration and accessibility in Seoul, concluding that enhanced transportation connectivity not only promoted business concentration but also facilitated the transformation and redistribution of sectoral clusters. Particularly, stations with high subway accessibility encouraged greater sectoral-diversity within clusters.

Accessibility:

An immediate impact of transportation investment is the accessibility enhancement within and across targeted areas. This enhances market potential, increasing these sites' attractiveness and stimulating economic activity and land value (Knowles, Ferbrache and Nikitas, 2020). Numerous factors influence firms' location decisions (accessibility, transport costs, rents…), which have been examined within the academic literature. Accessibility and rent received the most focus, as they are the primary determinants in firms' location decisions, influencing the pursuit of spatial advantages.

The relationship between accessibility and a firm's location is conceptualised through location theory. It indicates that firms strategically select locations to maximise benefits, such as profits, by enhancing suppliers, labour market, and consumers' accessibility, while simultaneously minimising transportation costs (Weber and Friedrich, 1929). Transport investments increase land value by capitalising on accessibility, demonstrating a clear correlation between land value and proximity to stations (Ko and Cao, 2013).

Enhancing accessibility through transport investment increases targeted areas' attractiveness, triggering inward capital investment and attracting developers and businesses. Empirical evidence established the transport accessibility role in attracting new businesses and promoting more balanced economic growth, thereby supporting regional economic development (Ozmen-Ertekin, Ozbay and Holguin-Veras, 2007). However, different industries (retail, manufacturing, etc.) exhibit varying willingness to trade economic capital for accessibility (Alonso, 1964). This variation influences how sectors are spatially reorganised in relation to transportation stations. Iseki and Jones (2018) found that specific sectors, such as finance, insurance, and professional services, have higher concentrations around metro stations and exhibit relocation patterns towards higher accessible areas. Nevertheless, such patterns do not always result in net business gains for those areas. Some scholars argue that boosting local economic activity requires further enhancements beyond accessibility improvements. For example, De Bok and Van Oort (2011) argue that

agglomeration economies, supported by accessibility enhancements, have a significant impact on firms' behaviour. All of this demonstrates the crucial accessibility effect on business demographics and growth.

2.4 Economic Activity Measures:

Property Prices:

Property prices are extensively examined as transport investments' impact on economic activity, particularly at a microscale. The transportation accessibility enhancements increase the location's attractiveness, especially when supported by policy or regeneration plans. This attracts investors and developers, increasing property values. The literature revealed a positive correlation between property prices and proximity to stations. However, the proximity effects vary depending on the property type, with commercial property prices increasing more in value than residential properties (Debrezion, Pels and Rietveld, 2007). Nevertheless, some studies have shown a negative or no significant effect on properties, which is attributed to several factors, including location, accessibility dynamics, proximity, and rail system maturity (Mohammad *et al.*, 2013). These complexities underscore the need for direct indicators to capture transport investments' impact on economic activity accurately.

Employment:

Employment is another standard indicator in academic literature to examine transport investments' impact on economic activity. Through increasing accessibility, reducing commuting costs, and encouraging economic activities' agglomeration around stations, transport investments can increase employment density, expand access to employment opportunities, and improve firm labour matching. However, this relationship remains a subject of ongoing academic debate. Some scholars argue that the impact on employment highly relies on various factors, including neighbourhood characteristics and complementary policies. Schuetz (2015) observed that new stations had no significant effect on retail employment in some areas, while they correlated negatively in others, underscoring the importance of local context. Others contend that observed employment growth near new stations might result from spatial employment relocation rather than net gains. Pogonyi, Graham and Carbo (2021) investigated how the Jubilee Line Extension affected local employment. The findings reveal a 6.6% increase in employment within 750 meters from the station. In contrast, areas 750 to 2000 meters away experienced a 1.6% decrease, suggesting that JLE led to economic activity redistribution rather than employment growth. This highlights that transport investments influence employment in surrounding areas. Still, the outcomes depend on factors like initial employment density, local characteristics, and policies, meaning they may not always directly reflect economic activity.

New Firm Creation:

Scholars have increasingly focused on examining transportation investments' influence on firm creation. Areas with better access to transportation act as incubators for new firms, as proximity to transit stations reduces transportation costs, improves market access, and fosters firms' agglomeration, thereby collectively contributing to firms' growth. Several scholars investigated how a new station influences firms' creation, accounting for sectoral differences in how firms benefit from proximity. Credit (2017) evaluated the new light rail system's impact on new business creation. The results indicate that new businesses experience a notable rise approximately 0.25 miles from the station; however, this effect diminishes with increasing distance from the station (beyond 1 mile). This increase was mainly evident in three sectors: retail, services, and knowledge, whereby knowledge increased more than the increases in retail and services. Similarly, Yao and Hu (2020) assessed the impact of urban transit on startups, finding that startup formation is concentrated around stations. Moreover, retail saw the highest startup creation increase, while technology firms experienced the lowest increase. Nevertheless, both studies emphasised the transport investments' role as a catalyst for economic activity and regeneration plans when they are synergistically incorporated. This shift in focus towards investigating the transportation stations' impact on new businesses has enriched the literature, paving the way for further research. However, accounting solely for business creation captures a limited dimension of transport investment's broader influence on economic activity.

The Importance of Measuring Business Demographics and Dynamics:

Extensive research highlights transport investments' role in boosting economic growth by meeting demands and acting as catalysts for economic development. Investments are justified by conventional cost-benefit analyses (CBA), which focus on direct benefits like travel costs and time savings. While these traditional methods are helpful for project appraisal, scholars have criticised them for failing to account for wider economic impact (WEI), such as the agglomeration effect, economic revitalisation, and business activity (Lakshmanan, 2011).

Measuring WEI is essential for both policymakers and academics, as it provides valuable insights into indirect effects, such as changes in spatial economic patterns, land use, and the labour market (Wang, Zhong and Hunt, 2019). Since (WEI) often manifests through shifts in employment, productivity, and spatial economic patterns, it is essential to investigate transport investments' impact on changes in businesses' dynamics, as an underlying mechanism underpinning economic activity. These investments influence businesses' location decisions, leading to changes in the sectoral composition and spatial clusters of businesses, which ultimately boost productivity and innovation through agglomeration (Lee, 2021). Additionally, businesses play a critical role in employment dynamics through firm birth, death, expansion, and relocation, which underpin job creation and economic renewal (Reynolds, 1994; Neumark, Zhang and Wall, 2005).

Although extensive research has examined the macroeconomic effects of transportation investments, studies assessing their impact on local economies remain limited. Furthermore, there is limited attention to changes in the business demographics and dynamics. Examining

businesses' changes offers critical insights into how transport investment reshapes the spatial distribution and dynamics of economic activity, offering a clearer indicator of transport investment's impact. Evidence suggests that transport investment, specifically when embedded within a broader regeneration framework, can have several effects on businesses' dynamics. These include attracting new businesses, influencing firms' survival, reshaping sectoral composition, and diversifying the local economy through distinct economic activities clustering (Tornabene and Nilsson, 2021; Champagne and Dubé, 2023).

Finally, most empirical studies have investigated the transport investments' impact in the USA, Spain, and China, with few analyses focusing on London, thereby resulting in a substantial gap in the existing research body. In recent years, the UK has witnessed a surge in transport investments, particularly in London. These investments are highlighted in The London Plan policy and the Mayor's Transport Strategy as vital tools to promote economic growth, support regeneration, and manage London's ongoing growth. Therefore, it is crucial to analyse how such investments shape businesses' composition and activities. This reveals how improved accessibility influences businesses' location decisions, firm creation, sectoral shifts, and business density and agglomeration around stations, thereby providing a meaningful assessment of policy effectiveness and addressing the gap in the London-specific literature.

2.5 Debates Around Transportation Impact:

Growth or Redistribution:

Transport investments are typically justified by their positive influence and capacity to drive economic growth. But what if this claim is not valid, and transport investment redistributes economic activity rather than producing net growth?.

Therefore, it is essential to differentiate between the actual transport infrastructure impact on economic activity growth, defined as the net increase in economic output, or firm numbers, and the spatial redistribution of economic activity without a genuine increase in it (Redding and Turner, 2015). This distinction is crucial for the accurate assessment of transport infrastructure and policies' success. Although this distinction is significant for evaluating transport investments' impact, relatively little research is conducted on this idea. In this context, Pogonyi, Graham and Carbo (2021) investigated the JLE impact on economic development, explicitly distinguishing between net growth and spatial firms' redistribution. They concluded that the extension has only spatially shifted the economic activity closer to the station from adjacent neighbourhoods without generating local-level growth, consistent with displacement rather than net growth. Similarly, Lindgren, Pettersson-Lidbom and Tyrefors (2021) examined the railway impact on economic activity in Sweden, measured by income, employment and productivity, population size, and land value. Their findings indicate that railway generated significant net growth rather than redistribution, and this was supported by the absence of spillover effects to neighbouring untreated areas. Overall, these contradicting results highlight the importance of further analysis to clarify whether transport investment truly stimulates net economic growth or shifts economic activity elsewhere.

Anticipation Effects:

A key consideration in assessing the economic impact of transport investment is the trajectory of economic activity, which can be shaped by the anticipation effect. Anticipation is defined as the changes in behaviour and investment patterns that might occur in response to the expected improvement in accessibility. This effect encompasses both the pre-opening and post-opening phases of a transport investment, reflecting the excitement, novelty, and perceived accessibility gains surrounding new stations. Such dynamics may manifest in changes in relocation behaviours and investment decisions before and after the initiation of the transport infrastructure. However, this effect tends to decline once the transport infrastructure is fully integrated and stabilised within the system (Golub, Guhathakurta and Sollapuram, 2012).

Evidence from the housing market has illustrated this effect. Comber and Arribas-Bel (2017) in their investigation of Crossrail's impact found that the scheme announcement led to a significant increase in house prices, reflecting forward-looking behaviour in the housing market. Similarly, Mohammad *et al.* (2013) examined the transport system maturity in relation to property value, represented by three distinct time intervals: construction, immediately after operation, and after stabilisation. Their findings suggest that increases in property value are often more pronounced around the time of the announcement than after the system stabilises.

In the business context, Credit (2017) presents one of the earliest empirical investigations into the 'novelty factor' within the context of new business formation. The results revealed that at the opening time, the area with proximity to the station experienced a surge in new businesses; however, over time, this increase diminishes, suggesting the existence of a 'novelty factor'.

These studies highlight the importance of accounting for temporal heterogeneity when assessing transport investment outcomes. Incorporating anticipation effects into transport investment assessment can provide valuable insights into the economic sustainability of transport-led regeneration impact.

2.6 Conclusion:

The literature review has examined the mechanism by which transport investment impacts economic activity, and the indicators used to measure that impact. It also delves into the debates around transport investments. The literature largely overlooks using changes in businesses' dynamics as indicators of economic activity to assess the impact of transport investments. To address this, a framework has been established to assess how transport investment influences sectoral changes, firm creation and relocation, and changes businesses' spatial concentration and agglomeration mechanism, all while investigating the ongoing debates about its impact.

3 Data and Area:

3.1 Area of Study:

VNEB Opportunity Area:

This research focuses on the Vauxhall, Nine Elms, Battersea Opportunity Area (VNEB) for the following reasons. First, VNEB represents one of London's largest regeneration zones, with significant growth potential following Canary Wharf extension (GLA, 2012a). The area's regeneration plan, supported by transport investments, makes it ideal for studying how such investments boost economic growth in regeneration areas. Second, historically, VNEB has been characterised by strong industrial uses, including a low-value industrial and designated strategic industrial locations. Despite its central location in London, the area was relatively isolated and lacked transport connectivity with surrounding neighbourhoods. It was marked by social and spatial fragmentation and surrounded by pockets of deprivation (Merz, 2009; GLA, 2012b). Thus, studying this area offers the opportunity to comprehend transport investments' influence on economic activities and business diversity, altering the businesses' dynamics, and making the area more accessible and attractive for inward investments.

Acknowledging accessibility barriers, GLA, TFL, and key stakeholders positioned transport investments as crucial to unlocking the area's development. TFL undertook a major transport study, considering different scenarios to support VNEB development with and without the extension. The Northern Line Extension (NLE) was constructed to create two new stations (Nine Elms and Battersea Power Station) branching from Kennington station. The NLE was considered essential for catalysing economic growth across the VNEB and London more broadly, through enabling business formation, residential areas, and leisure districts. This would contribute to achieving CAZ-level density of activity within VNEB and effectively extending the CAZ into its last undeveloped segment (TFL, 2013). Finally, NLE is the first significant transport investment since the JLE in 1999 and is expected to replicate the extension's stimulative impact on economic growth. Therefore, examining its impact on economic activity is crucial, particularly since no empirical research has yet been conducted to assess its impact on economic activity, especially on businesses.

To study the NLE impact on VNEB, a control area for comparative analysis is used: the Old Kent Road (OKR) Opportunity Area. Like VNEB, OKR has a regeneration plan supported by strategic transport investments; however, this transport investment has not been established yet (detailed reasoning for selecting this area discussed later). This approach assesses and isolates the strategic transport investments' impact, like NLE, on economic activities within London regeneration zones.

Figure 1: *Northern Line Extension Scheme with VNEB Opportunity Area Border.* ***Source****: (TFL, 2013)*

3.2 Data:

To understand how NLE influenced businesses' demographics and dynamics in the VNEB area, several datasets from different sources are used. The two main businesses' datasets are the Business Census and WHYTHAWK datasets. **Table-1** summarises all datasets, combining both public and private sources.

Data	Source	Description
Business Census	CDRC (2012-2024) for all UK registered businesses.	Offers a longitudinal register of all UK-registered corporate entities. This data provides an annual snapshot at the address level.
Business Rate	Whythawk (2010-2024) for Wandsworth, Lambeth, and Southwark boroughs. Provided and produced by Gavin Chait at Whythawk.	Provides data on the address level for commercial properties, including business rates, ratepayer details, and business sector.
PTAL Accessibility indicator	TFL	This Data is an indicator of how well connected an area is and is calculated by TFL.
Residential Mobility and Deprivation Index	CDRC (1997-2023)	At the LSOA level, it contains an estimation of residential mobility and deprivation levels.
Property Characteristics, Prices and Rents	CDRC (2007-2023)	At the LSOA level, it contains an estimation of

		property sale prices and rent values.
Modelled Ethnicity Proportions	CDRC (1997-2023)	At the LSOA level, it contains an estimation of ethnic distribution percentages.
Business Register and Employment Survey	Nomis (2010-2023)	At the LSOA level, it contains the number of employees in a geography and sector.
Residential Mobility Index	CDRC (1997-2023)	At the LSOA level, it includes data on estimates of residential churn as a percentage of 2023 levels.
Output Area Population	Nomis (2011-2022)	Estimates of the resident population at the OA level.
Output Area Boundary	ONS	Contains the OAs' boundaries.
Opportunity Area Boundary	TFL	Contains the boundaries of London's Opportunity Areas.

Table 1: *Overview of used datasets*

Business Census:

This dataset covers all UK-registered businesses from 2012 to 2024, provided by Companies House via Fusion Data Science. The dataset is at the firm's address level, providing the finest spatial level granularity needed for detailed, small-area business dynamics analysis over time. It captures businesses' life-cycle, including formation, dissolution, and redistribution. The dataset contains essential company information, like SIC[3] Code that is used to identify companies' sectors. It allows for developing additional variables useful for the analysis, such as tracking firms' locations. **Table-2** summarises the key variables.

Variable	Description
ID	a unique identifier for each company linking the company's record across all entity files.
Company Number	Company number as specified by Companies House
Incorporation Date	This refers to the date that a company is formed.
Company Name	Name of the company
Dissolution Date	Company dissolution date
Postcode	Postcode for the registered address of the company
AddressLine1	First line of the registered address of the company
SIC Code	Company Standard Industrial Classification Code
Postcode Longitude	Longitude for the company postcode
Postcode Northing	Latitude for the company postcode

[3] SIC code is the standard industrial classification code.

Company Category	Refers to the registered legal status of the company (Private Limited Company, ..)
Country of origin	The country where the business originates from
Company status	Includes the company's status if it's still active or is in another stage.

Table 2: *Key variables of Business Census data*

WHYTHAWK:

This is a quarterly dataset from 2010 to 2024 for commercial properties in England and Wales. It is produced using data from VOA, ONS, and Local Authorities in England and Wales. While it is not a firm-level specific, it provides unique variables, like rates, that can be used to analyse the changes in businesses' dynamics and decisions. However, to do this, the data should be approached and wrangled with care to create meaningful variables. As this dataset is on the branch-level per business, duplicated businesses are dropped for this analysis to focus on companies with one location to avoid any over-representation and adhere to Census granularity. Dataset is on the address-level, which provides the opportunity to analyse a small spatial unit as well. **Table-3** summarises the dataset key variables.

Variable	Description
Rates Code	Local authority hereditament billing code, only unique in combination with 'rates authority'
Rates Authority	Local authority/ billing authority code
Location Code	UARN as defined by the VOA, should be unique
Floor Area	Commercial property floor area in square meters
Pc Pcs	Postcode
Name	Name of commercial property occupier
Rental Valuation	Property rental valuation based on VOA
category	SCat highest-level category
SCat Code	VOA Special Category code (numeric)
Sub Category	Higher-level SCat category
Rates Expected	Rates expected to be paid
Address no	First line of the address
Status Date	The start date of the current occupation

Table 3: *Key variables of WHYTHAWK data*

PTAL Indicator:

PTAL is a numeric score (0 – 6b) developed by TFL, measuring how well a location is connected and served by public transport, where zero represents the lowest score and 6b is the highest score. It is widely used to support local and strategic planning decisions. According to TFL, the PTAL calculation involves both walking time to the nearest station[4], and the average waiting times based on service frequency and reliability, which are combined to calculate total access time. This is then converted to equivalent doorstep frequency (EDF), which is used to derive an access index and finally mapped to the PTAL (TFL, 2015). Therefore, using the PTAL to evaluate how a new station affects businesses' dynamics provides a more realistic measure of accessibility, a critical driver for businesses' location decisions, than relying on Euclidean proximity. A business may be close to a station yet poorly served due to infrequent train services or encountered walking network barriers, issues that PTAL explicitly account for.

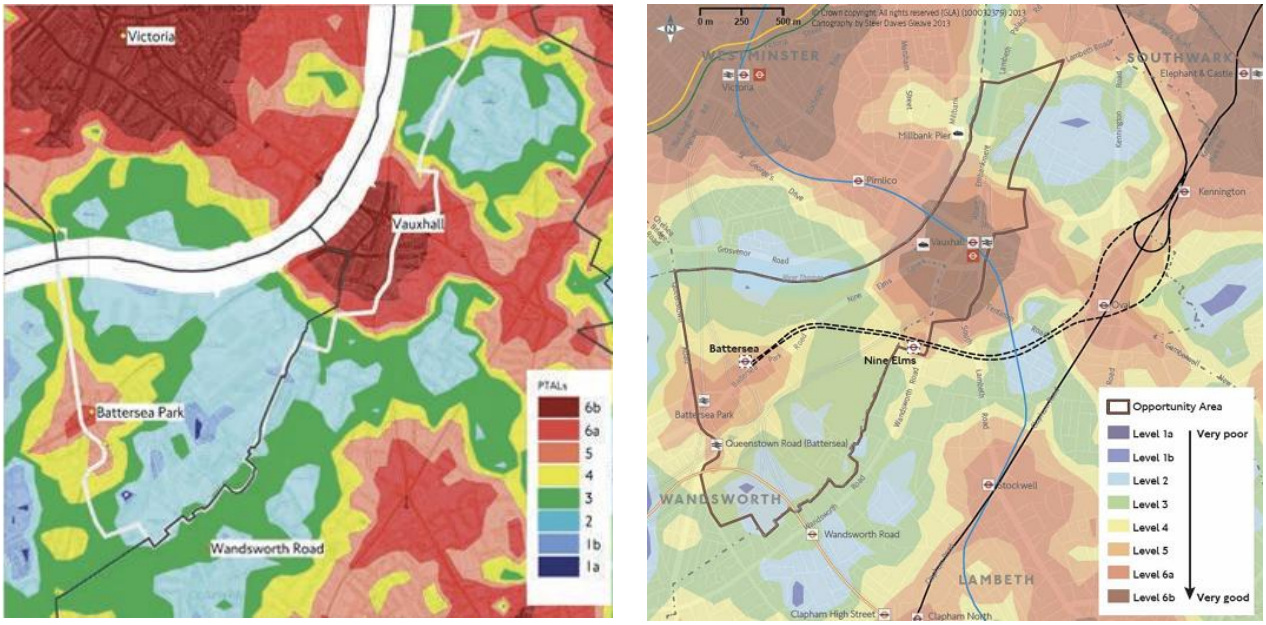


Figure 2: PTAL Score for VNEB Opportunity Area, on the Left for year 2015, while on the right is a projection for year 2031.
Source: (TFL, 2013a, 2015)

Other Datasets:

Other datasets have been used in the analysis to account for spatial heterogeneity and potential confounders like population and demographics. Those assist in understanding how socioeconomic factors affect transport investment planning and regeneration initiatives. For example, property characteristics, prices, and rents data can serve as a proxy for people's affluence. Modelled Ethnicity Proportions data contains ethnic composition (White British, White Irish, Asian, African, etc.) on LSOA-level from 1997 to 2023. Finally, the Residential Mobility and Deprivation Index data offer yearly estimates of the differences in deprivation percentile scores between incoming and outgoing residents. This helps determine whether

[4] This can also be referred to as service access points (SAPs)

areas attract wealthier or poorer residents, thereby indicating economic shifts and changes in demand and businesses' growth.

3.3 Spatial Unit:

As this analysis examines how transport-led regeneration influences businesses' dynamics, selecting the appropriate spatial unit is crucial. The study uses the Output Area (OA), the smallest administrative spatial unit in London. This choice is driven by the following reasons:

- The two primary datasets, the Business Census and WHYTHAWK, contain data at the address level. This enables precise mapping of each firm's postcode to its respective OA. This facilitates a high-resolution analysis of spatial change and thus captures the localised impact of NLE on the VNEB opportunity area.
- OAs are stable, standardised statistical units that provide a good granularity level of spatial data for analysing business dynamics compared to postcodes, which are not consistent, frequently changing and too small, causing data to be very sparse and highly sensitive.
- The VNEB area covers several LSOAs across the boundary of Wandsworth and Lambeth boroughs, meaning some LSOAs may be partially inside and partially outside the opportunity area. Using OA-level as a spatial unit ensures more precise analysis within the VNEB opportunity area, thereby minimising aggregation bias risk in estimating NLE impact.

One thing to highlight is that some used datasets are on the LSOA-level and are disaggregated to the OA-level employing different assumptions based on data types (discussed below). Those assumptions might over- or underrepresent some quantities (for example, ethnic groups) on the OA-level. Hence, estimated coefficients of such variables might not 100% reflect the true relationship of those independent variables with the studied dependent variables. However, due to the limited availability of time series data at the OA-level, these assumptions were made.

4 Methodology

This chapter covers methodologies used to preprocess various datasets and create independent variables of interest. Additionally, it includes the two techniques used in modelling business dynamics to understand the NLE impact on businesses, such as business creation, redistribution, sectoral shifts and business spatial concentration around new stations. **Figure-3** outlines the complete methodology workflow for clarity.

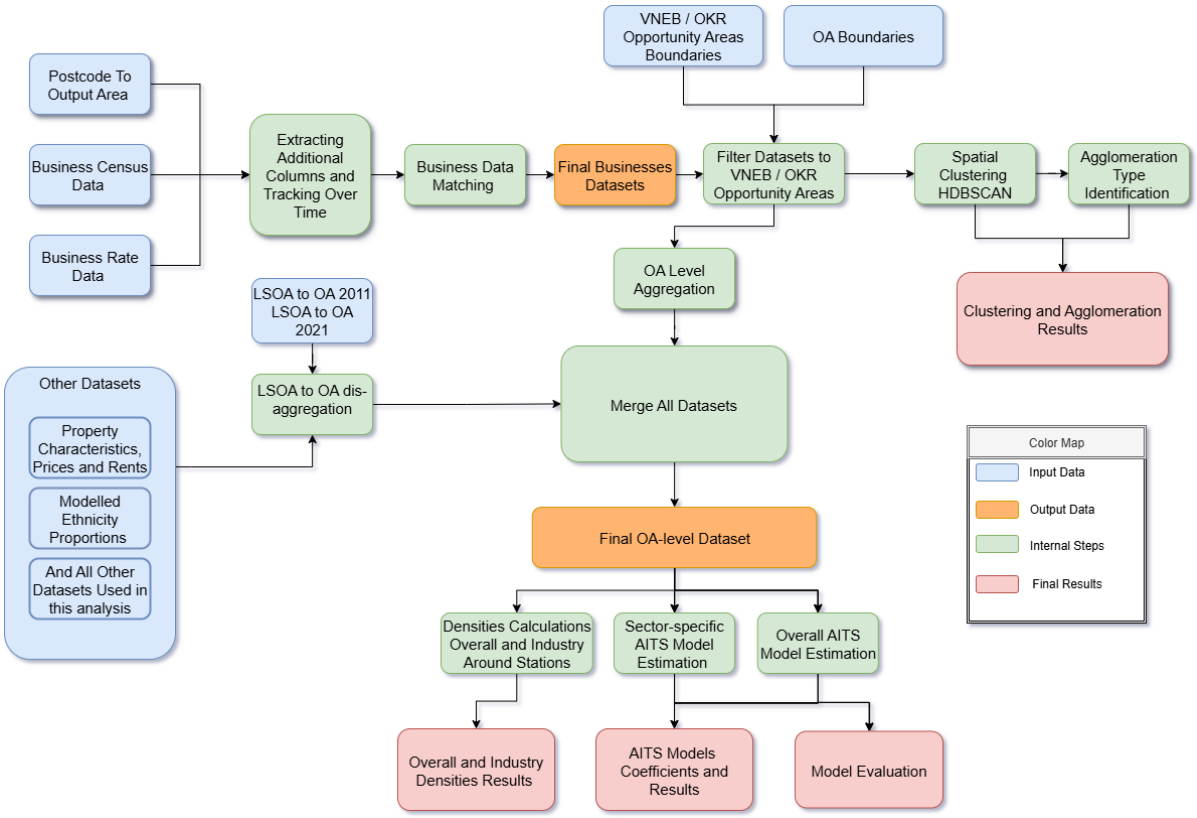


Figure 3: *Methodology flow chart*

4.1 Data Preprocessing

Business Data:

As mentioned earlier, the **Business Census** dataset has fields used in constructing the final set of model features. Each business has a detailed address, including the first line of address and postcode. By joining businesses' postcodes to the "2021 LSOA-OA-Postcode" mapping, each business is mapped to its respective OA, latitude and longitude. Furthermore, this dataset is annual, and has a business unique identifier, "CompanyID". This field helps in tracking business changes over time, including locations' changes. By grouping businesses based on ID and comparing their postcodes across consecutive years, relocated businesses are identified, like those moving to high accessibility areas. This is used to calculate moved business density at the OA-level later.

Additionally, this dataset includes fields such as dissolution and incorporation dates, which are used to identify yearly new and dissolved companies, and eventually calculate new

business density, dissolution and survival rates when aggregating data at the OA-level. Finally, business industries based on SIC classification are mapped to SCAT classification using a static mapping between the two classifications.

Similarly, **Business Rate** data has the first line of address and postcode, which are used to link businesses to their respective OA, latitude and longitude. As this dataset is quarterly, each year's fourth quarter data is used as that year's existing businesses. Moreover, as no unique identifier exists, "rate code", "location code" and "name" fields are used to create businesses' IDs, which are used to track business changes over time.

Although business rate data includes "status date" as the registration date, it lacks a separate dissolution date field. Hence, by grouping businesses over time, the latest year the business appears in is the dissolution year (except 2024 data). Ultimately, those fields are used to calculate new, moved and dissolved business densities and rates. Additionally, the business rates dataset includes additional fields such as expected rates, business rent valuation, and floor area, which are used to calculate OAs' average business rates and rent valuation.

Matching Business Rate and Business Census Datasets:

Since there is no common identifier across datasets to merge on, businesses' names and full addresses are used instead. Since these elements are not standard in both datasets, a semantic similarity matching approach is employed to map the business datasets accurately. Therefore, a deep neural-net sentence-embedding model pre-trained for bilingual language, leveraging the robust capabilities of Multilingual-MiniLM-L12-H384 (Wang *et al.*, 2020) and fine-tuned using Siamese BERT-Networks with 'sentence-transformers' (Yi, 2023) and Augmented SBERT with Pair Sampling Strategies (Thakur *et al.*, 2021) is used. Business names and addresses are embedded into 384 dimensions vectors, which are used to calculate cosine similarity, identifying the most similar names or addresses. To ensure minimal levels of false positives, 85% similarity score threshold is applied to match addresses and business names. This threshold is selected by randomly sampling data and testing if the false positive rate stays below 1%. Finally, both datasets, whether matched or non-matched businesses, are combined in the final business dataset. Full details of the matching process are explained in **Appendix-B**.

Demographic Data:

Population data is collected on the OA-level for each age category and overall, with two caveats worth mentioning. Firstly, population data is based on the 2011-OA classification. Hence, a mapping from 2011-OA to 2021-OA classifications is applied to get 2021-OA populations. This introduces three scenarios that need to be outlined. The first involves a one-to-one mapping between 2011-OA and 2021-OA. The second occurs when multiple 2011-OAs are mapped to only one 2021-OA, with the 2021-OA population being the total population of all involved 2011-OAs. The third involves splitting a 2011-OA into multiple 2021-OAs. In this scenario, the population is assumed to be split equally between the different new 2021-OAs. The second caveat is that population data is available until 2022;

therefore, an assumption is made that the population remained the same after 2022. The final dataset includes the total and working-age population (19-64 years) at the OA-level.

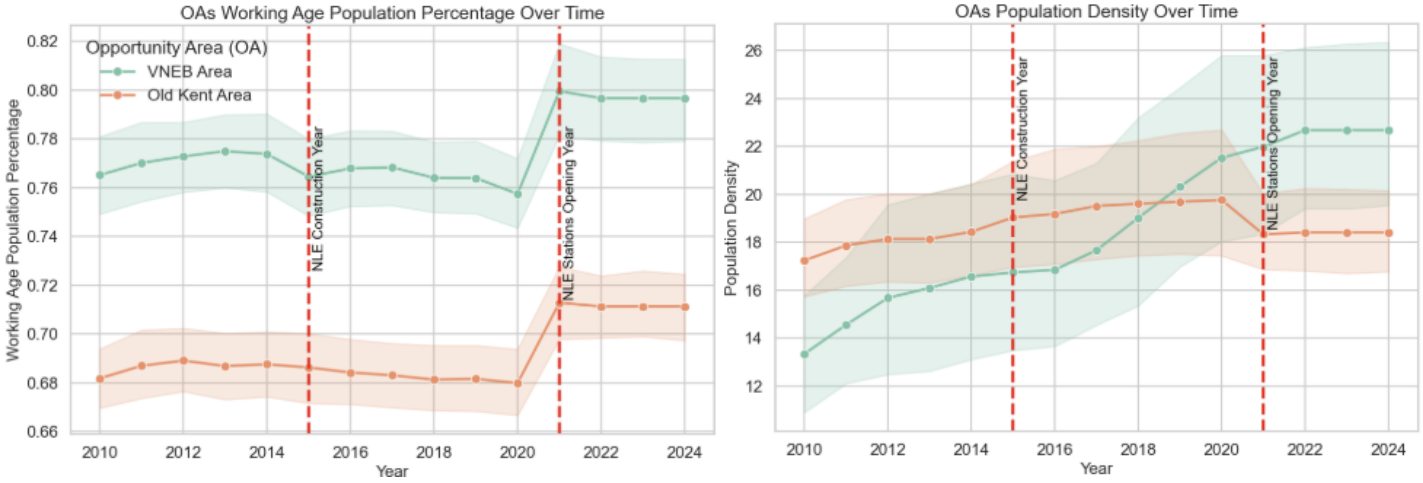


Figure 4: *population over time VNEB vs OKR*

Modelled Ethnicity Proportions data includes percentages of different ethnicities, including White, Asian and Black, at the LSOA-level. To calculate ethnicities on the OA-level, and due to the lack of more granular ethnicity data, an assumption is made that percentages hold the same on the OA-level for each LSOA.

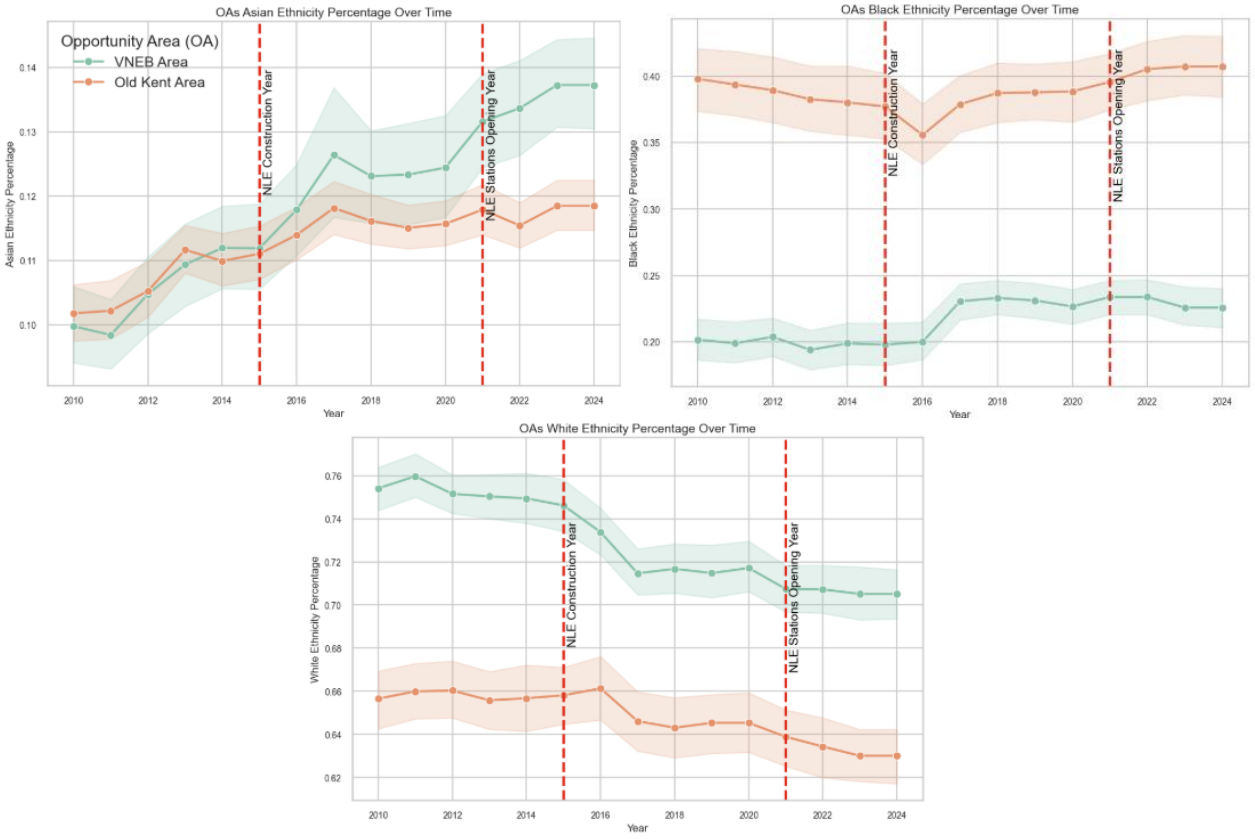


Figure 5: *Ethnicity changes over time, VNEB vs OKR*

The Residential Mobility Index (RMI) is calculated based on the household difference ratio compared to 2023 at the LSOA-level. **Figure-6** shows that RMI is higher earlier in the timeline compared to 2023, due to using 2023 levels as the reference point. To calculate RMI on the OA-level, an assumption is made that percentages hold the same on the OA-level for each LSOA. As 2024 is missing, 2023 levels are used, assuming there is no change in 2024.

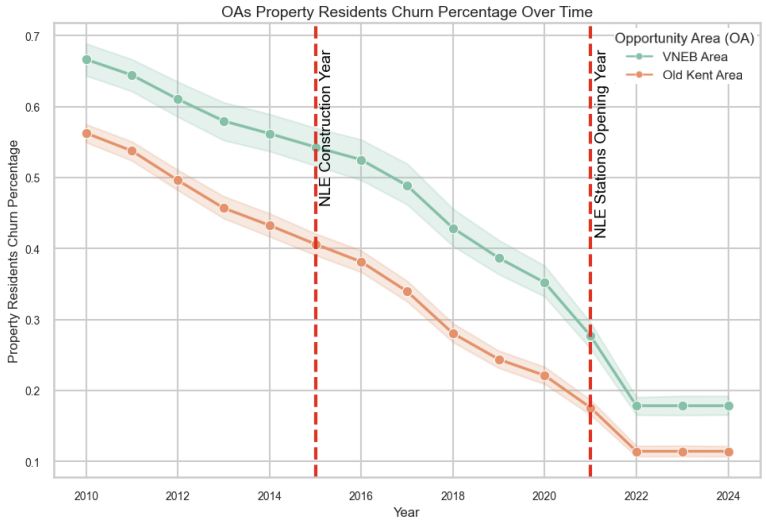


Figure 6: *Residential Churn VNEB vs OKR*

Resident Movement and Deprivation Index data contain deprivation index values of people moving in and out of each LSOA. Like the demographics data, OA-level data is calculated by assuming the same level of deprivation for every LSOA holds for all its OAs.

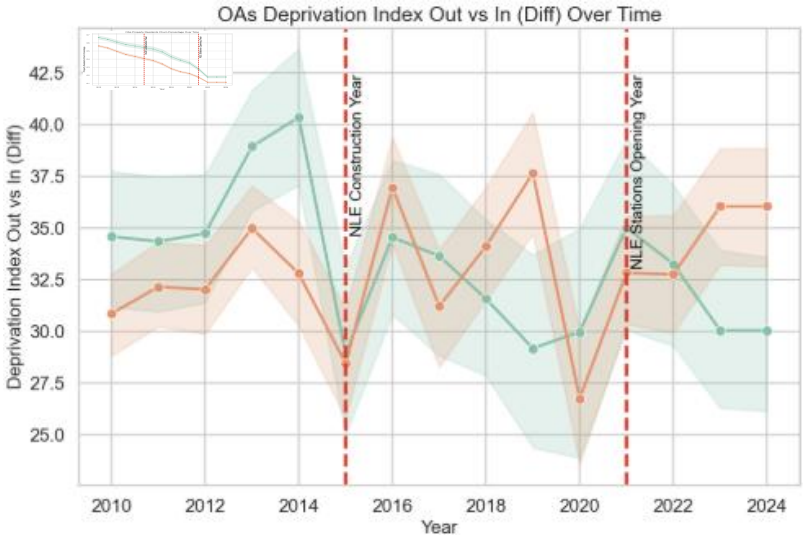


Figure 7: *In and Out deprivation Index VNEB vs OKR*

Other Datasets:

The Business Register and Employment Survey dataset contains job numbers on the LSOA-level for each SIC industry. To calculate OA-level jobs count per industry and overall, the population data is used to estimate the job numbers for each OA proportionally by OAs population percentage. Then the SIC to SCAT mapping is used to calculate jobs per SCAT industry.

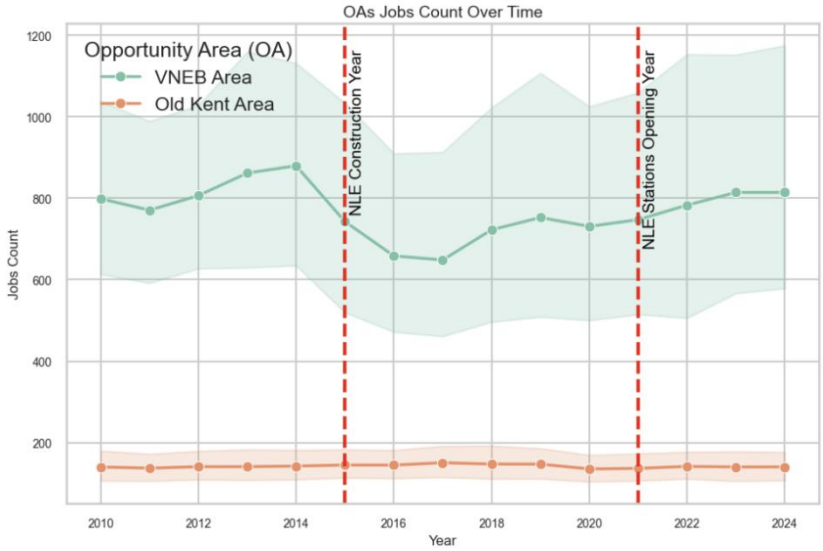


Figure 8: *Job count over time VNEB vs OKR*

Finally, the **Property Characteristics, Prices and Rents** dataset contains properties' average prices and rents on the LSOA-level. Like the demographics dataset, property prices and rent values are assumed to hold on the OA-level for the same LSOA, as there is no information on the split within LSOAs.

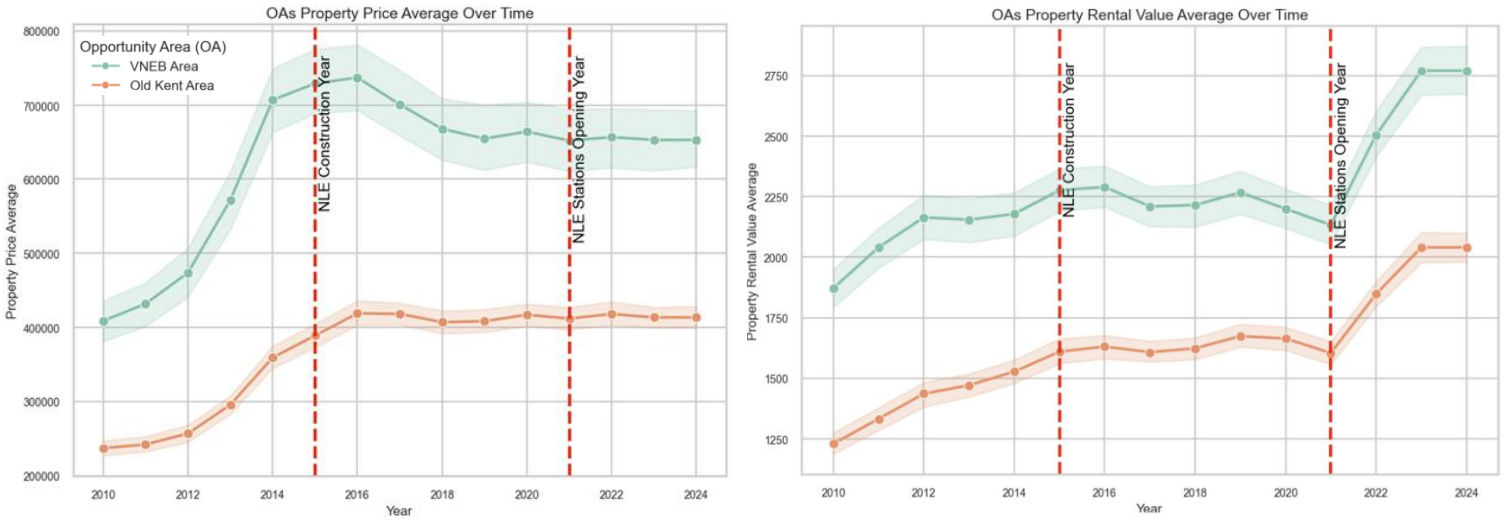


Figure 9*: Property prices/rents over time VNEB vs OKR*

4.2 Features Engineering

Before discussing the different estimated models, the feature engineering section is dedicated to explaining the aggregation performed on the businesses dataset, whereby several features are extracted from each business to represent businesses' characteristics on OA-level.

Businesses Aggregation at OA Level:

To study the NLE impact on VNEB (the treatment area) compared to Old Kent Road (the control area, explained below) regarding the overall and industry-specific businesses' dynamics, the independent variables listed in **table-4** are calculated on OA-level to be used in the model estimation.

Feature Name	Feature Description	Granularity
OA21CD	Identifier of OA based on 2021 classification	OA level
Year	Year of calculation	OA level
OA PTAL	PTAL 2023 classification of each OA	OA level
Area (m^2)	OA area in square meters	OA level
Businesses Count	Count of businesses for a given OA and a given year	Overall / Per Industry
New Businesses Count	Count of new businesses for a given OA and a given year	Overall / Per Industry
Moved Businesses Count	Count of businesses moved from their locations for a given OA and a given year	Overall / Per Industry
Closer or Same PTAL Moved Businesses Count	Count of businesses moved to a higher or same level of PTAL accessibility for a given OA and a given year	Overall / Per Industry
Dissolved Businesses Count	Count of dissolved businesses for a given OA and a given year	Overall / Per Industry
Dissolved Business Rate	Percentage of dissolved businesses for a given OA and a given year	Overall / Per Industry
Survival Rate	Percentage of businesses that survived over the full study years for a given OA	Overall / Per Industry
Business Rates Per m^2	Average of business rates per m^2 floor area for a given OA and a given year	Overall / Per Industry
Rent Valuation Per m^2	Average of business rent valuations per m^2 floor area for a given OA and a given year	Overall / Per Industry

Table 4: *Business features list*

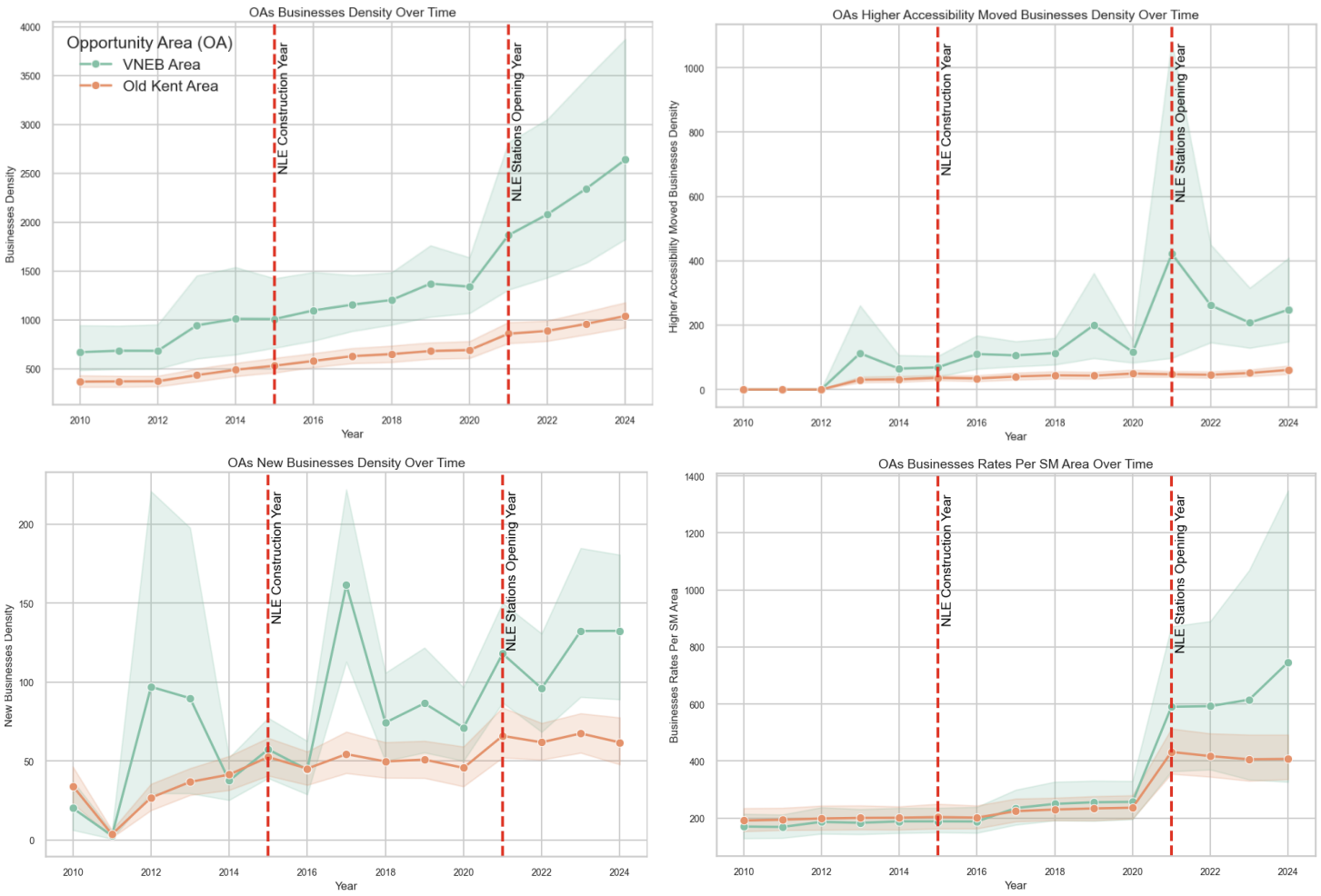


Figure 10: *Business counts/rates over time VNEB vs OKR*

Agglomeration Measures:

Following (Chatman, Noland and Klein, 2016), overall and per-industry business counts are used to estimate the agglomeration effect. Additionally, an extra measure is added by calculating the companies' average distance from the business centre for each OA (business dispersion). This is calculated for overall firms and per industry.

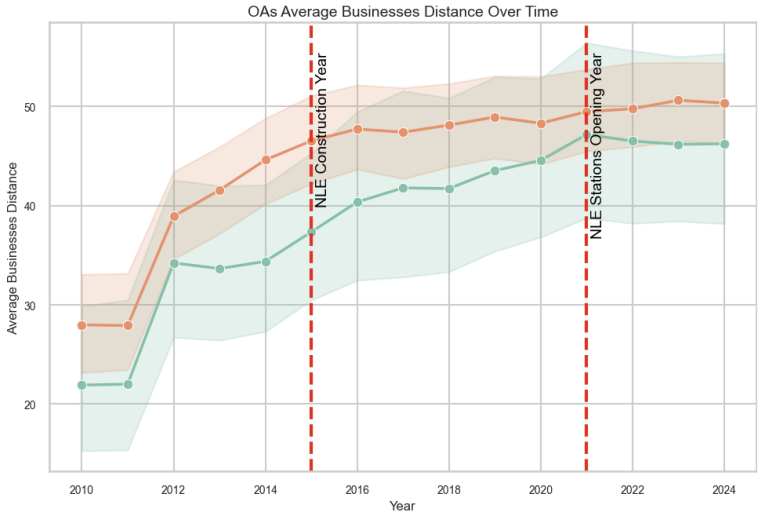


Figure 11: *Average business distance (dispersion) over time VNEB vs OKR*

4.3 Model Selection and Evaluation

4.3.1 Control Area Selection:

When using models such as difference-in-difference and adjusted interrupted time series (AITS) to infer a specific event's causal impact on an outcome within a particular area, it is crucial to select an appropriate control area. These models rely on the assumption that, in the treatment's absence, the treatment and control areas would follow a similar trend over time. Hence, the control area should closely resemble the treatment area in the pre-intervention period to limit selection bias. In this analysis, the Old Kent Road Opportunity Area is selected as the control area because its characteristics before the intervention are like those of the VNEB (treatment area). Like VNEB, high deprivation levels and industrial use characterise OKR. It is designated as a strategic industrial location (SIL), has limited connectivity, and is allocated for a regeneration plan supported by transport investments (Bakerloo Line Extension) (Southwark Council, 2017; Greater London Authority, 2021). However, the area has not yet experienced its planned catalytic intervention, making it a suitable control area candidate to assess the NLE impact on VNEB.

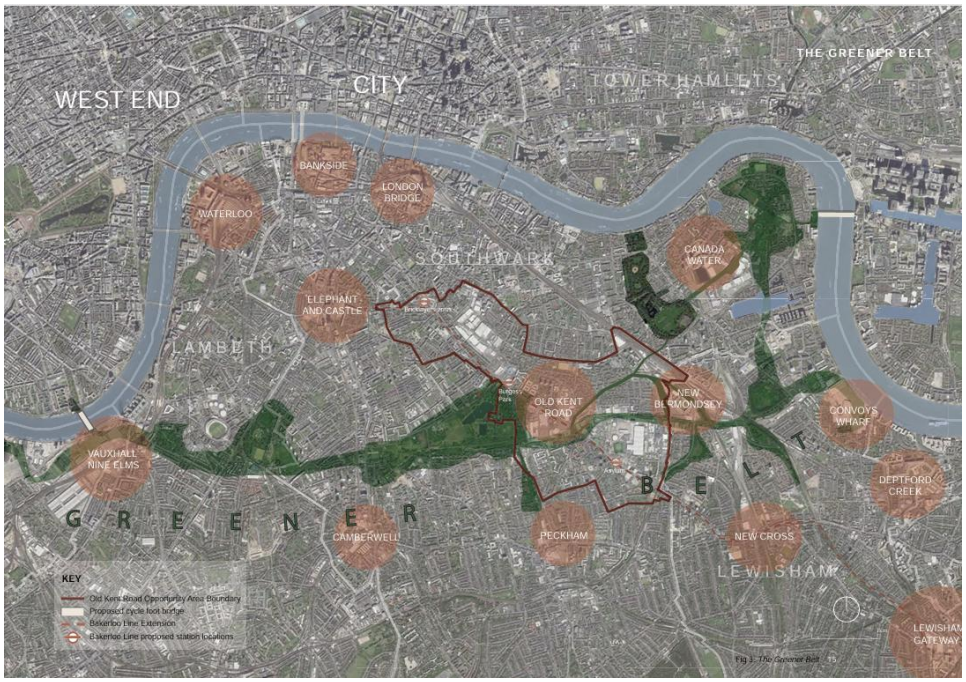


Figure 12: Old Kent Road opportunity area border and proposed stations for the Bakerloo Line Extension **Source:** (Southwark Council, 2017)

4.3.2 Modelling Techniques:

4.3.2.1 Adjusted Interrupted Time Series (AITS):

To capture NLE impact over time on businesses' dynamics, this analysis uses the AITS methodology, which is used by (Credit, 2017). AITS is a statistical method that leverages the regression framework, whether by using the standard linear regression or more advanced types of regressions such as Generalised Linear Model (GLM) with different distributions like Negative Binomial or Poisson, and uses time series data to study the intervention impact on treatment versus control areas over time.

The advantage of using AITS over the regular DiD model comes from using time series data in studying the causal relationship between transport interventions and the different studied business demographics dependent variables, outlined later in this section. AITS accounts for both the changes in absolute values and trends of dependent variables over long periods before and after the intervention, enabling tracking and identifying the anticipation effect. This is crucial for studying transport interventions as they take time to accumulate their impact; hence, AITS enables the study of pre- and post-treatment trends without relying on subjectively selecting two random snapshots in time to study the intervention effect in DiD model case.

AITS Models Design:

In this analysis, several AITS models are estimated to study two dependent variables outlined in **table-6** for different years of intervention and accessibility levels (further explained later). GLM framework is used for model estimation with either Poisson or Negative Binomial distributions (dependent variables are business counts based). Experiments show that the Poisson distribution suffers from overdispersion across all estimated models; hence, the Negative Binomial distribution is used in the final models, showing very little to no dispersion. For each AITS model, a few dummy and trend variables are used to represent treatment versus control areas, the intervention point, trend over time, and intervention interactions with treatment dummy and trend. AITS formula is:

$$Y_{it} = \beta_0 + \beta_1 Trend_t + \beta_2 Treat_i + \beta_3 Post_t + \beta_4(Trend_t * Post_t) + \beta_5(Treat_i * Post_t) + \beta_6(Treat_i * Post_t * Trend_t) + \gamma X_{it} + \epsilon$$

Symbol	Description
$\boldsymbol{\beta_0}$:	Intercept
$\boldsymbol{\beta_1}$:	Baseline trend for control group
$\boldsymbol{\beta_2}$:	Baseline level difference between treated and control
$\boldsymbol{\beta_3}$:	Level change post-intervention for control
$\boldsymbol{\beta_4}$:	Trend change post-intervention for control
$\boldsymbol{\beta_5}$:	Level change for treated group post-intervention (treatment effect)
$\boldsymbol{\beta_6}$:	Trend change in the treated group post-intervention
$\boldsymbol{\gamma X_{it}}$:	Time-varying covariates

Table 5: *Explanation of AITS model formula symbols*

Dependent Variable	Distribution	Level	Importance
New Business Count per km^2 (new business density)	Negative Binomial	Overall and per industry	Analysing these variables enhances understanding of how transport investments affect new business creation and relocation decisions, and highlights which sectors benefit most. This allows for assessing the localised impact of these investments on economic activity. Examining both variables provides a broader insight into the overall impact.
Moved Business Count per km^2 (moved business to higher/same PTAL density)	Negative Binomial	Overall and per industry	

Table 6: *Studied dependent Variables*

A careful consideration is needed for intervention year selection in AITS, as this will impact the model's coefficient estimation. Hence, to measure the impact of NLE announcement versus completion, two intervention years are selected, 2015 (construction kicked off) and 2021 (stations completion). The 2011 announcement year is not used because there are insufficient data points for the studied variables before 2011. To measure each intervention year's impact, a separate model is estimated for each year and dependent variable, which allows a comparison of effects between the two intervention years.

To study the NLE impact on OAs based on accessibility levels, OAs are classified into high and low accessibility with PTAL levels five or higher considered high accessibility areas, while OAs with levels four or lower considered low accessibility areas. This splits the VNEB Opportunity Area equally (53 OAs high and 58 OAs low). To measure the difference in NLE impact on accessibility levels, a separate model is estimated for each accessibility and dependent variable, allowing for an easy comparison of NLE impact on different accessibility classes.

Figure-13 illustrates business relocation origins to VNEB, highlighting that the most significant pattern, correlated with NLE major interventions, especially in 2021, is the movement from other London boroughs.

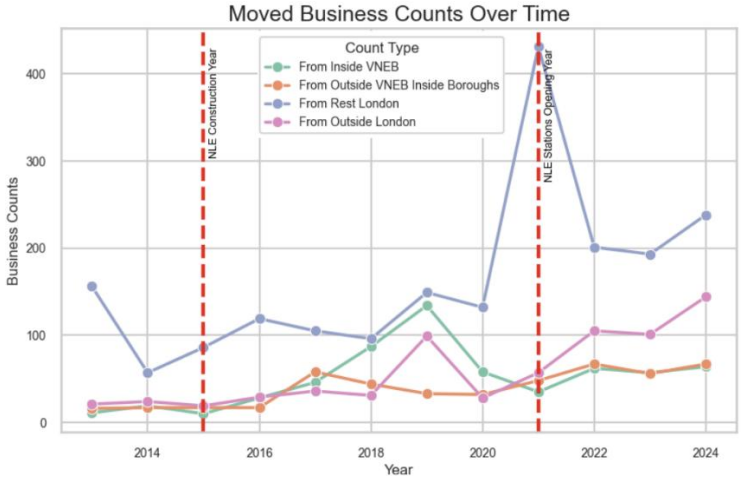


Figure 13: *Moved business origin*

Figure-14 illustrates how different accessibility classes exhibit distinct patterns in moved, new, and overall business counts. New businesses experienced a significant surge in high-accessibility areas following the start of construction in 2015, with a lesser impact when stations opened in 2021. However, this trend gradually diminished in subsequent years. In contrast, moved businesses experienced a significant increase in 2021 in terms of high-accessibility compared to low-accessibility areas, a trend that persisted until 2024.

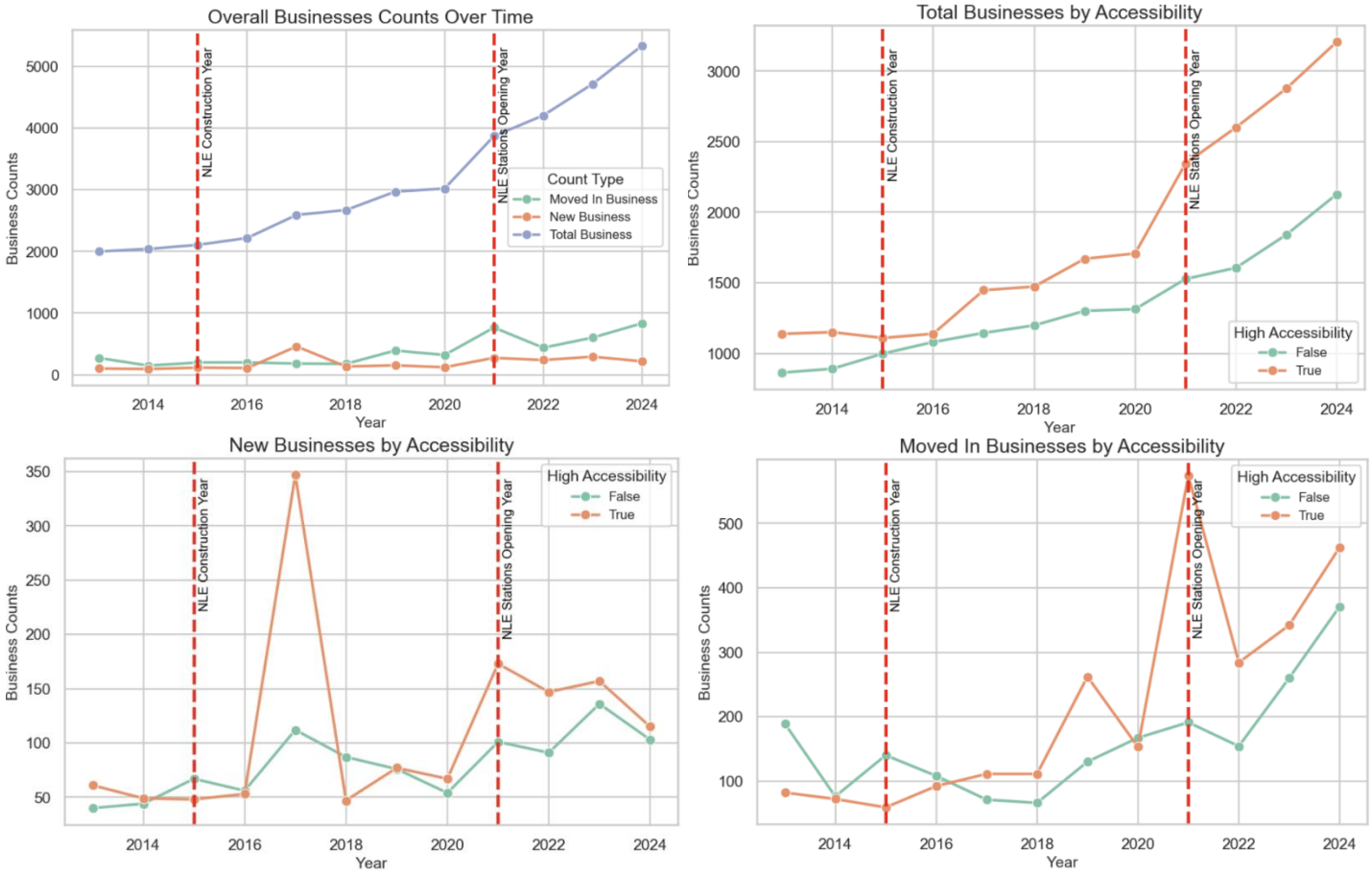


Figure 14: *VNEB business count by accessibility over time*

Table-7 shows a list of OA-level independent variables used to estimate each dependent variable. New, moved, and overall business densities variables are one year lagged to incorporate them as features. Additionally, a COVID-19 dummy variable is added for 2020 and 2021 to measure its effect on dependent variables.

Feature Category	Feature	Description
Demographics Features	Population (km^2)	Ensures that the analysis is not affected by underlying population characteristics that might influence economic activity.
	Working-age population	
	Asian percentage	
	Black percentage	
	White percentage	
	Changes in the deprivation index (people moving in and out)	

Local Agglomeration Features	Business density 1lag	It is well established that agglomeration economies (urbanised and localised) impact firms' creation and relocation. The included features capture multiple dimensions of local agglomeration and serve as proxies for agglomeration economies at the micro-level. Including these variables enables the analysis to account for the impact of agglomeration effects on businesses' demographics and dynamics.
	New business density 1lag	
	Moved business density 1lag	
	Average business cluster distance	
	Jobs count	
	Business density 1lag per industry	
	Average business cluster distance per industry	
	Jobs percentage per industry	
Business Dynamics Features	Business dissolution rate	To account for the local business environment, economic health, and vitality conditions that would influence business creation and relocation.
	Business survival rate	
	Average business rate per floor area (m^2)	
	Average rental valuation per floor area (m^2)	
Properties Features	Average property prices	To account for socio-economic dynamics, residential affluence, and cost of living, which may affect the attractiveness of an area and footfall.
	Average property rental value	
	Property residents churn percentage	
OA Features	Area in km^2	To account for the OAs area impact.
AITS Features	Treatment area dummy	Those dummy and trend features are important to study the impact of intervention on the absolute count and trend of the treatment area compared to the control area.
	Post-intervention year dummy (2015 and 2021)	
	Treatment area post-intervention year dummy (2015 and 2021)	
	Trend over the years	
	Trend over the years post-intervention (2015 and 2021)	
	Trend over the years for treatment area post-intervention (2015 and 2021)	
Other Features	COVID-19 dummy	To consider Covid's impact on firms' creation and relocation, and to prevent Covid-related distortions from confounding the estimates.

Table 7: *AITS independent variables list*

Figures 15 and **16** demonstrate joint distribution plots of moved and new businesses against the most important independent variables.

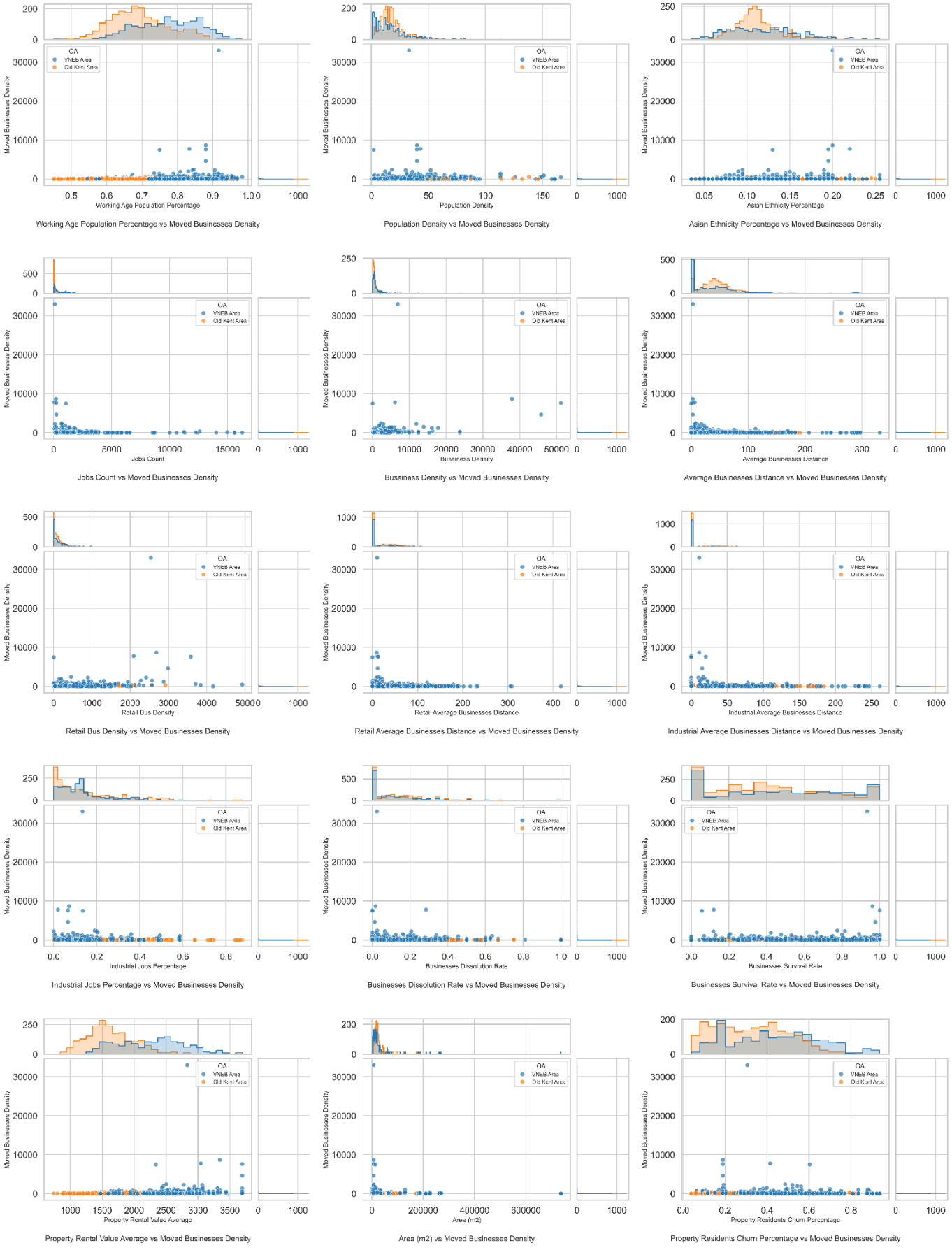


Figure 15: *Moved Business vs Independent Variables Joint Distribution Plots*

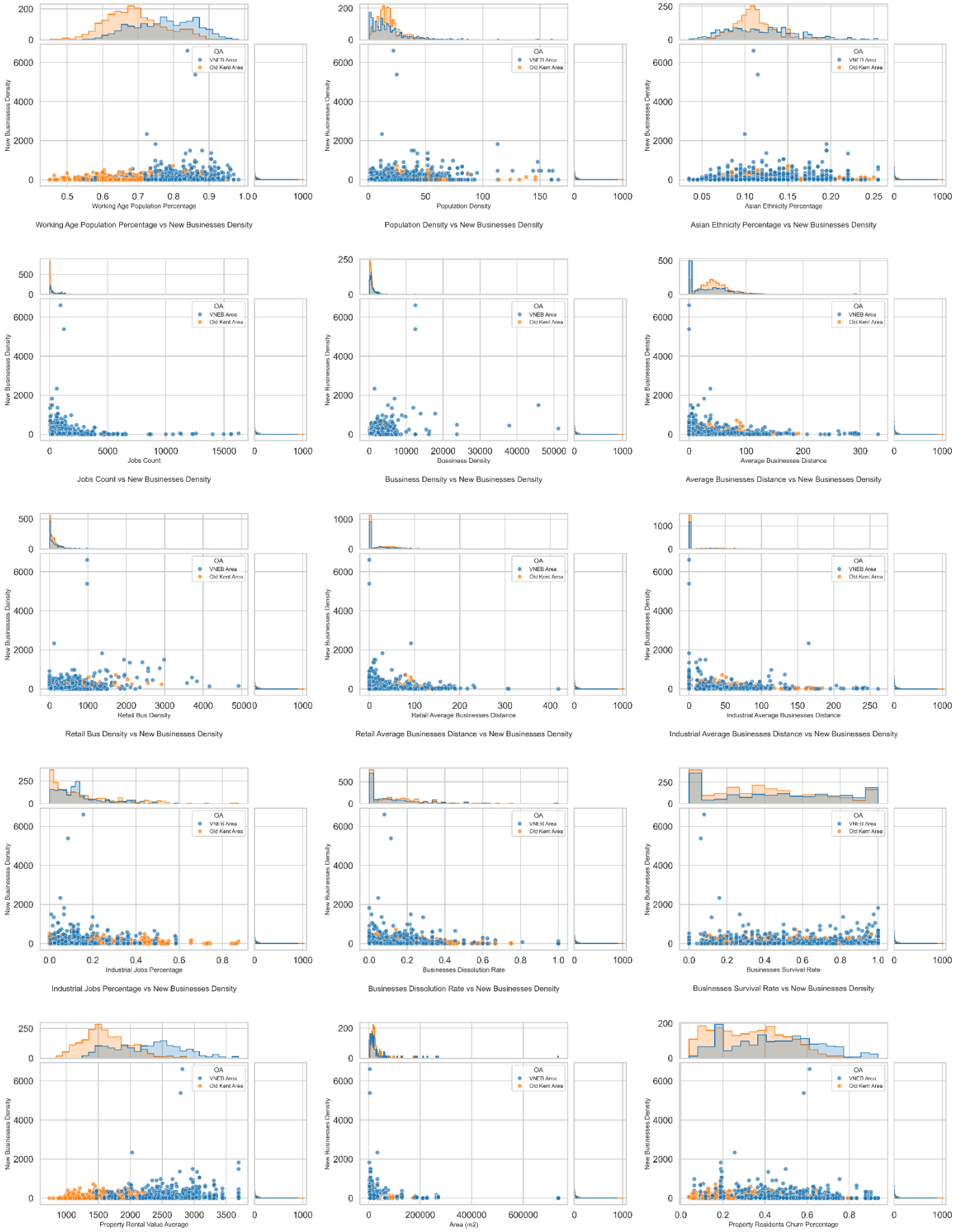


Figure 16: *New Business vs Independent Variables Joint Distribution Plots*

Figure-17 exhibits a few independent variables maps between 2015 and 2021 of VNEB vs OKR, including Asian ethnicity percentage, business counts and average business distance.

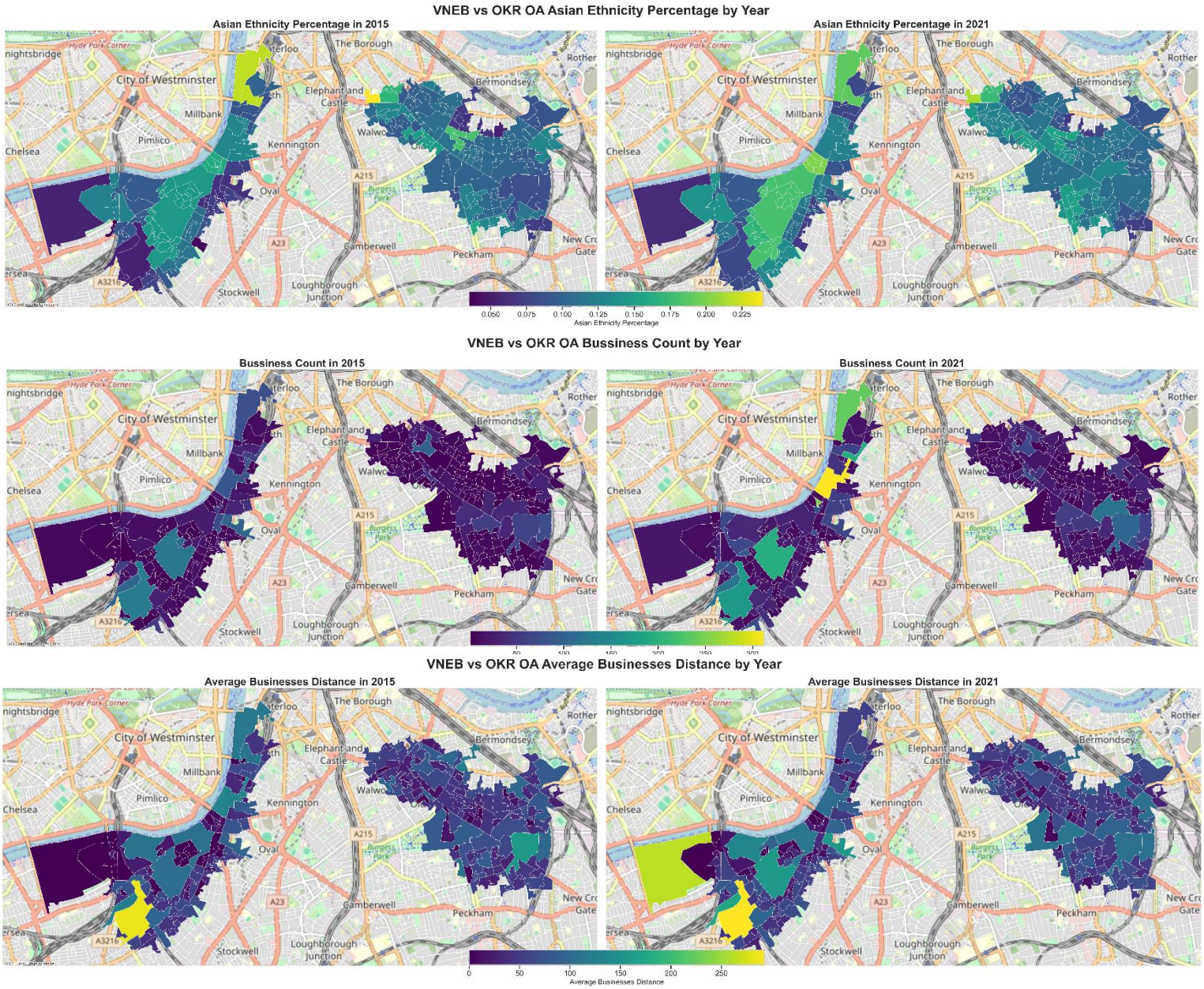


Figure 17: *VNEB vs OKR OAs Asian Ethnicity, Business Counts, and Average Business Distance of 2015-2021*

Multicollinearity and Model Evaluation:

To address multicollinearity between independent variables, the Variance Inflation Factor (VIF) is used to drop features using a threshold of 8. Then, following each AITS model estimation, feature coefficients with a p-value of 0.05 or less are considered. Additionally, both McFadden and Cox-Snell R^2 are used to check Negative Binomial models' fitness. To compare NB models against Poisson models, the Likelihood Ratio Test (LR) and p-value are used to measure how well NB fits compared to Poisson and if the improvement is statistically significant. Finally, model dispersion and residuals distributions (QQ plot and residuals-vs-fitted) are used to check model fitness.

4.3.2.2 Spatial Clustering:

To capture business concentration changes and agglomeration composition around the new stations over time, business spatial clustering is employed based on their locations. Several clustering methods are good potentials; however, HDBSCAN can identify clusters with varying densities and is robust to noise and outliers, which positions it well as an option for real-world cases like studying business agglomeration. To this end, business cluster formation is studied overall and around stations (Nine Elms and Battersea Power stations) with a radius distance of 500 and 750 meters to measure the NLE impact on business clusters.

To uncover changes in sectoral diversity within clusters around stations, entropy is used to evaluate clusters' purity based on sectoral composition. This enables distinguishing between two types of business clusters, Localisation versus Urbanisation, measuring the NLE impact on business agglomeration types.

Entropy formula:

$$H = \sum_{i=1}^{k} p_i \, log_2 \, p_i$$

K: sectors number per cluster, $\boldsymbol{p_i}$**:** the proportion of points in sector i for a given cluster.

To produce stable and good-quality clusters, hyperparameter tuning is applied for each year based on the most important parameters (minimum cluster size and epsilon distance). The evaluation metric used in tuning those parameters is the silhouette score, defined as:

$$S = \frac{(b - a)}{\max(a, b)}$$

a: the mean intra-cluster distance, **b:** the mean nearest-cluster distance

4.4 Ethical Considerations:

This research uses a combination of safeguarded datasets obtained via CDRC, secured through a formal data access agreement, along with private and publicly accessible datasets. All data are anonymised and aggregated, containing no personal identifiers, and outputs have been processed to reduce the risk of re-identification. The findings will be shared with the partner organisation, TFL, to support evidence-based policy making and urban development strategies. The UCL Ethics Committee reviewed the data and methodology as presenting minimal risk and did not require additional review.

5 Results:

In this section, the results of all models, whether overall or industry-specific business behaviours, are presented to identify any emerging patterns.
The first step in estimating any regression-type model is to address any multicollinearity that exists in the data. **Figure-18** shows the features' correlation heatmap, which exhibits some level of multicollinearity that needs to be handled for every model:

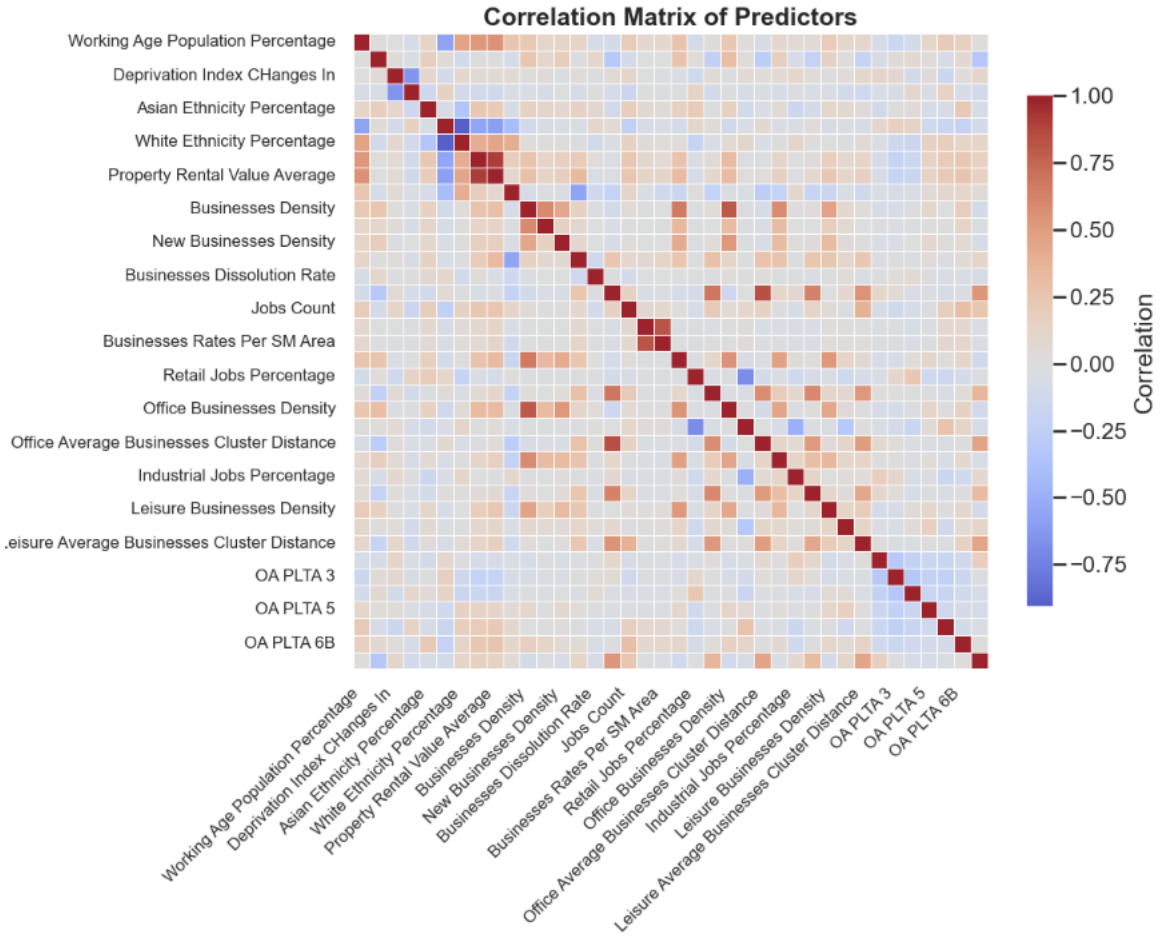


Figure 18: *Correlation heatmap*

5.1 Overall Businesses AITS Models Results

To address multicollinearity of studied dependent variables at the overall business level, VIF with a threshold of 8 is used to drop highly correlated features as shown in **table-8**:

Feature	Max VIF Value
White Ethnicity Percentage	60
Office Sector Jobs Percentage	30
Average Property Prices	14
Final list of features	**7**

Table 8: *VIF removed features*

Then, a model is estimated for both new and moved businesses' dependent variables of each intervention year and each accessibility class, resulting in the full coefficient values table in

Appendix A. **Table-9** [5]highlights the most important coefficient values, from which the following key findings can be drawn. Percentages in the key findings section are calculated based on the incident rate ratios (IRR), which are computed as e raised to the power of every given coefficient value. Finally, the **treatment and trend-treatment post-intervention** features are the key variables capturing the NLE impact on treatment versus control areas.

Table 9: *Overall new and moved business AITS model coefficients*

Dependent Variable	**New Business Density**				**Moved Business Density**			
Accessibility	**high**		**low**		**high**		**low**	
year	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	35	35	34	34	35	35	34	34
Pearson Chi2	3780	3110	7140	6860	3370	5030	19600	48700
Pseudo R-Square	0.689	0.692	0.457	0.443	0.956	0.931	0.834	0.68
Demographics Features								
Population	0.192**	0.205**	0.252**	0.245**	0.098*	0.110**	0.138**	0.113**
Working-age population	0.837	0.915	1.783**	1.571**	1.315**	1.271**	2.314**	2.339**
Asian %	1.037	0.718	-2.161*	-2.314*	9.348**	8.895**	-5.635**	-4.495**
Local Agglomeration Features								
Average distance	-0.143	-0.048	-0.072	-0.058	-0.290**	0.019	-0.411**	-0.421**
Jobs count	0.162**	0.202**	-0.052*	-0.065**	0.096	0.319**	0.178**	0.274**
Retail Business density 1lag	0.253**	0.262**	0.304**	0.306**	0.445**	0.451**	0.078**	0.073**
Retail average distance	0.341**	0.341**	0.238**	0.226**	0.478**	0.364**	0.055	0.059
Industrial average distance	0.244**	0.234**	0.154**	0.145**	0.028	0.046	0.095**	0.083**
Industrial Jobs percentage	0.948**	0.721*	0.636**	0.492**	1.354**	1.012**	0.141	-0.228
Business Dynamics Features								
Dissolution rate	-0.690*	-0.444	0.095	-0.044	-0.225	1.341**	0.408**	1.372**
Survival rate	0.17	0.345*	0.570**	0.344**	0.246	1.143**	-0.141	0.079
Properties Features								
Residents churn %	-0.732	-1.079*	0.042	0.525	-0.426	-1.164**	1.490**	1.500**
AITS Features								
Treatment post-intervention	0.775**	0.309	0.478**	-0.14	0.785**	0.923**	-1.126**	-0.826**
Trend- treatment post-intervention	-0.066*	-0.233*	0.011	0.131	-0.003	-0.231*	0.049**	0.247**
Other Features								
Area	-0.524**	-0.584**	-0.392**	-0.406**	-0.539**	-0.870**	-0.452**	-0.540**

[5] One star (*): Coefficient values are less or equal 0.05 and greater than 0.01. Two stars (**): coefficient values are les or equal 0.01

COVID-19 dummy	-0.006	-0.502**	0.027	-0.393**	0.067	-0.898**	0.092	-0.624**

AITS Models Key Findings:

New Businesses:

- 2015 intervention year shows a significant positive impact on treatment vs control areas for both high and low accessibility (with 117% for high accessibility compared to only 61% for low accessibility). The intervention has a negative impact on trend for high-accessibility, indicating a yearly decrease of 6% after the intervention.
- 2021 intervention year shows no significant impact on both high and low accessibility, demonstrating no clear trend pattern for new businesses after stations opened.

Moved Businesses:

- 2015 intervention year shows a significant positive impact on treatment vs control areas for high-accessibility (with 119% increase), while a negative impact for low-accessibility (with 68% decrease). The intervention has a positive impact on trend for low-accessibility, indicating a 5% yearly increase after intervention, whereas no significant impact for high-accessibility.
- 2021 intervention year witnesses even higher positive impact on treatment vs control for high-accessibility (with 152% increase), while a negative impact on low-accessibility (with 56% decrease). Intervention impact on trend is positive for low-accessibility with 28% yearly increase after intervention, while it is negative for high-accessibility with a 21% yearly decrease after intervention.

Demographics Variables:

- Working-age population variable has the most consistent positive coefficient for both new and moved businesses across the two years of intervention; the total population also had a consistent positive coefficient, but less in magnitude compared to the working-age population.
- Asian ethnicity has a very high positive coefficient for high accessibility areas (mainly for moved businesses), while a negative coefficient for low accessibility areas for both years. This may be associated with the increase in the Asian ethnicity percentage observed earlier in **Figure-4**. Black ethnicity and changes in the deprivation index have no significant coefficients (with only small coefficients for moved businesses in 2021).

Agglomeration Variables:

- Job counts have a consistent positive coefficient for both new and moved business for high and low accessibility areas for both years (except a negative coefficient for low accessibility for new companies in 2015). Businesses' distance has a negative

coefficient for moved businesses, mainly in 2021, for both low and high accessibility areas.

- Retail variables of business density and distribution have the most consistent positive coefficient for new and moved businesses for high and low accessibility areas for both years. Industrial distance and job count variables mainly have a positive coefficient for new businesses for both years, suggesting the increase in new businesses is associated with more scattered industrial companies. Offices and leisure variables have no consistent coefficient for both years and types of businesses (except for a positive office businesses coefficient for high accessibility and new companies in 2015).

Business Dynamics Variables:

- Business dissolution rate has a positive coefficient for moved businesses in 2021 for both high and low accessibility areas, suggesting that moved businesses benefited from dissolved firms in moving to higher PTAL areas. The survival rate has a positive coefficient for new businesses, particularly in 2021, indicating that it encourages new businesses creation.

OA and Properties Variables:

- OA area shows a consistent negative coefficient for both new and moved businesses in high and low accessibility areas for both years, suggesting that bigger OAs witnessed a lower increase in new and moved businesses.
- Residents' churn has a negative coefficient for both new and moved businesses of high accessibility areas in 2021, while a positive coefficient for low accessibility of the same year, suggesting that business creation and redistribution in low areas benefited from residents' movement, contrary to businesses in high areas.

COVID-19:

COVID-19 showed a negative coefficient in 2021 only on both new and moved businesses for both high and low accessibility areas, with decreasing percentages as (40%, 33%, 59%, 47%) respectively, affecting moved businesses in high accessibility areas the most.

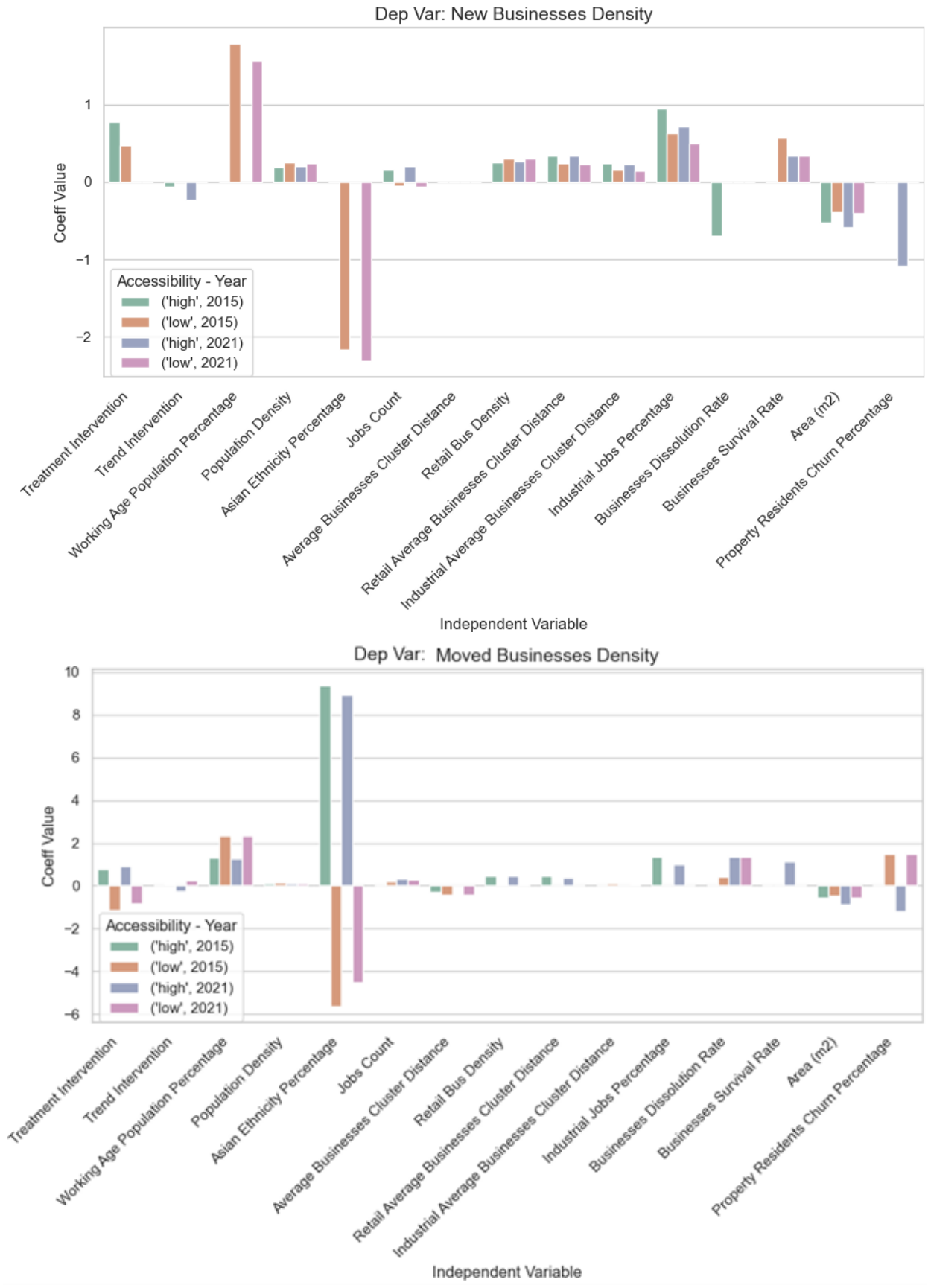


***Figure 19**: Overall business AITS coefficient*

Those results show the intervention's impact on the overall businesses with distinctive patterns; however, zooming in on each sector can help in understanding which sectors benefited the most from both interventions.

5.2 Industry Specific Businesses AITS Models Results

To address multicollinearity of the studied dependent variables at the industry-level, VIF with a threshold of 8 is used to drop highly correlated features. This resulted in dropping mainly two features: an ethnicity feature (White Ethnicity Percentage for the high accessibility model and Black Ethnicity Percentage for the low model), and the other is Average Property Prices.

A model is then estimated for each dependent variable, intervention year, accessibility classification and industry, resulting in the coefficient tables of industry-specific models in **Appendix A**, from which the following key findings can be drawn.

New Businesses:

- 2015 intervention year shows a positive impact on treatment vs control areas for low-accessibility for retail, industrial and leisure sectors with 78%, 135% and 79% increase, respectively, and a significant positive impact for retail and industrial high-accessibility areas with 500% and 111% increase. The intervention has a negative impact on trend for low-accessibility of industrial and leisure, and high-accessibility of retail, resulting in yearly decreases of 5%, 5% and 12%, after the intervention. In contrast, retail low-accessibility shows a positive impact, indicating 6% annual increase after intervention.
- 2021 intervention year shows a positive impact on treatment vs control areas for low-accessibility for retail, industrial and leisure sectors, with 63%, 184% and 63% increase respectively, and a negative impact for leisure high-accessibility and office low-accessibility with 82% and 35% decrease. The intervention has a negative impact on trend for industrial and leisure low-accessibility areas, indicating a yearly reduction of 30% and 22% after the intervention, while a positive impact for office low-accessibility with a 24% annual increase.

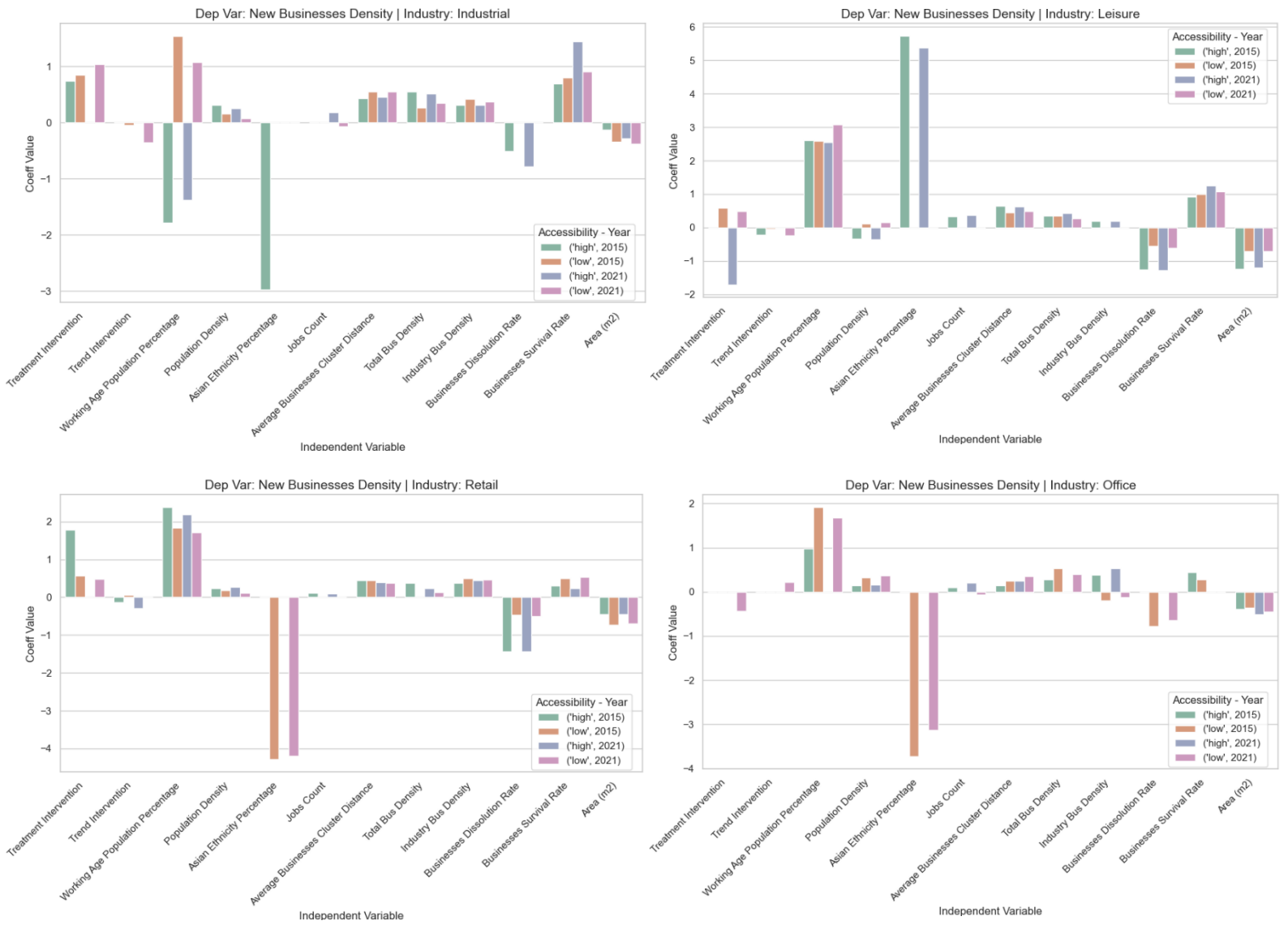


***Figure 20**: industry-specific business AITS coefficient for new business*

Moved Businesses:

- 2015 intervention year shows a significant positive impact on high-accessibility for retail, industrial, and leisure and low-accessibility for retail only, with an increase of 1230%, 497%, 2080% and 42% respectively, while a negative impact on low-accessibility for office, industrial and leisure with a decrease of 81%, 80% and 85% respectively. Intervention impact on trend is positive for low-accessibility of office, industrial, and leisure, with 14%, 8%, and 19% increase, respectively. In contrast, it has a negative impact on trend for industrial and leisure high-accessibility and retail low-accessibility, indicating a yearly decrease of 11%, 10% and 7%, respectively.
- 2021 intervention year witnesses a significant positive impact for high-accessibility for retail, office, industrial and leisure sectors with an increase of 519%, 289%, 458% and 268% respectively, while a negative impact for low-accessibility of office, industrial and leisure industries with a decrease of 40%, 83% and 73% respectively.

Intervention impact on trend is negative for high-accessibility areas of retail, office and industrial, with a decline of 31%, 33% and 50%. In comparison, it is positive for low-accessibility areas of office, industrial and leisure, with an increase of 24%, 73% and 91%.

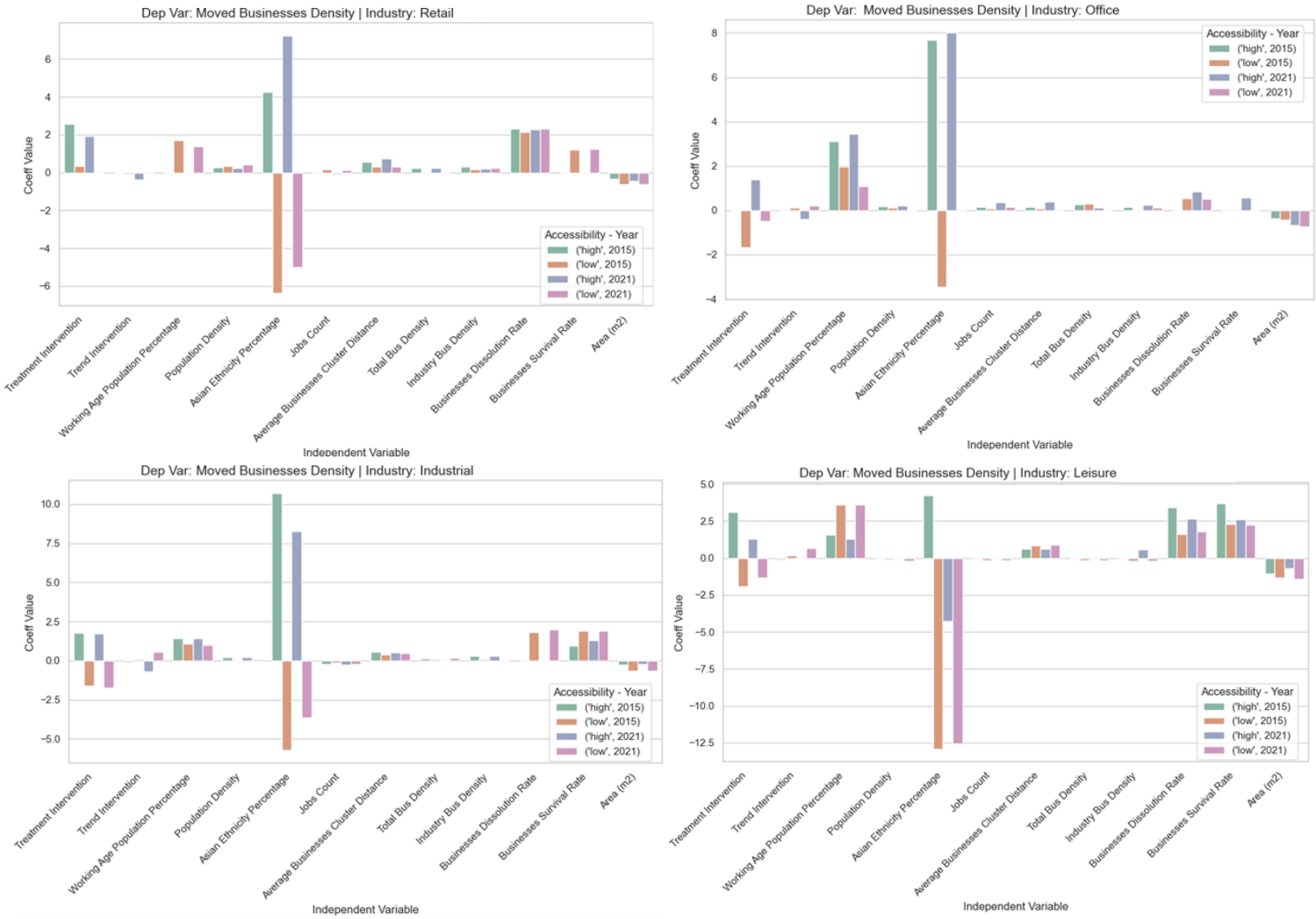


Figure 21: *industry-specific business AITS coefficient for moved businesses*

Conclusion:

Retail is the highest winning sector of both intervention years (2015 shows a profound impact for both new and moved), benefiting the most from businesses migrating to higher accessibility areas. On the contrary, the office sector is the least benefited one from new businesses and mainly benefited from moved businesses in the 2021 intervention year. Industrial low-accessibility areas benefited equally from new business for both years, whereas high-accessibility areas benefited more from business redistribution for both years. Leisure had no or a negative impact on higher accessibility for new companies (benefiting low accessibility areas specifically in 2015), while a massive uptake occurred in the business redistribution towards highly accessible areas.

Agglomeration Factors:

Overall and industry-specific business densities and average business distance have positive coefficients for both new and moved businesses of each industry, whether for high or low accessibility areas. This indicates that higher business density encourages business creation and redistribution, as they spatially expand to accommodate more incoming businesses, providing evidence for agglomeration. Additionally, the survival rate has a positive relationship for both new and relocated businesses, indicating that surviving businesses encourage more businesses to move in or get created. On the contrary, the business dissolution rate has a negative relationship with new businesses, suggesting that dissolved businesses discourage the creation of new businesses. However, it has a positive impact on the moved businesses, implying that dissolved businesses created a space for more businesses to move in.

Studied agglomeration factors in AITS models provide insight into the impact of industry-specific business densities and spatial distribution on business formation and redistribution. However, augmenting these results with other agglomeration angles, such as business clustering, diversity, and sectoral mix, using business clustering analysis is crucial for obtaining a comprehensive understanding.

5.3 Models Evaluation

Several statistical tests are employed, including the p-value of Poisson vs NB, Cox-Snell R2, McFadden's R2, and dispersion. Additionally, residuals versus fitted and Q-Q plots are used to assess model fit. **Table-10** shows the model fit statistics for each of the overall businesses' estimated models. A similar table is added in **Appendix A** to show similar statistics for industry-specific models.

Table-10 demonstrates that the Negative Binomial models show a moderate to solid fit of overall business density with acceptable levels of dispersion and huge improvements over Poisson models. This indicates that Negative Binomial models fit the data much better than the Poisson models, with high Cox–Snell values, and LR test results reinforcing that predictors are explaining a lot more variation than the null model.

Model	Cox-Snell R2	McFadden's R2	dispersion	(Poisson vs NB) p-value / LR Test
Overall New Business High-15	0.69	0.1	3.47	0 / 165305
Overall New Business Low-15	0.46	0.06	2.74	0 / 192632
Overall New Business High-21	0.69	0.1	2.86	0 / 165052
Overall New Business Low-21	0.42	0.06	2.63	0 / 203999
Overall Moved Business High-15	0.96	0.26	3.1	0 / 199956
Overall Moved Business Low-15	0.83	0.18	7.5	0 / 234373
Overall Moved Business High-21	0.93	0.22	4.62	0 / 199453
Overall Moved Business Low-21	0.66	0.11	7.5	0 / 255770

***Table 10**: Model-fit evaluation metrics*

Figure-22 illustrates examples of Q-Q plots and residual distributions for the new and moved business models in two different years. Those graphs demonstrate that no clear bias or pattern exists in residuals, and they are, to a reasonable extent, following the theoretical line in the Q-Q plot. However, there are abnormal residuals, especially at the extremes of the residuals distribution. After investigation, it appears that these patterns are driven by outliers in the dependent variables (high values of new and moved business density, especially in the treatment area, which drives such a pattern). Those outliers can be addressed by removing them or applying some capping; however, for this study, the decision was to keep all data points to study the NLE impact on all OAs over time without any exclusion.

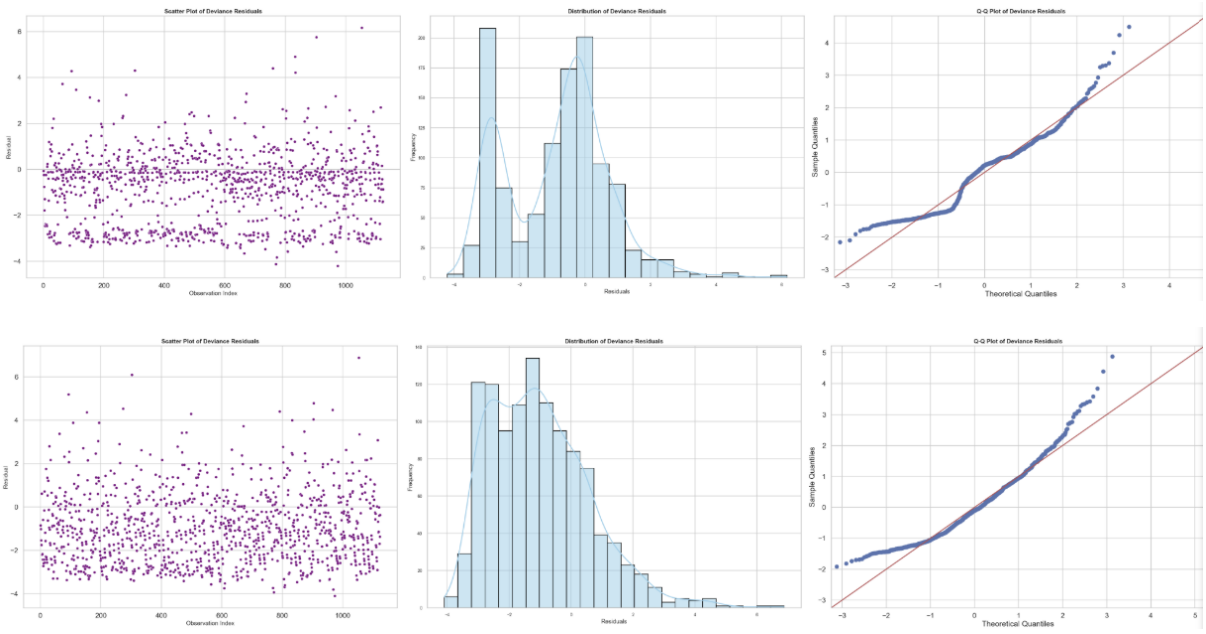


Figure 22: *Model residuals plots*

5.4 Businesses Clustering

AITS exhibited that NLE has a positive impact on increasing business counts, whether from new or moved firms, overall and per industry across the accessibility classes. The impact dynamic differs between the two intervention years; however, to better understand the spatial dynamics of such an increase, a deep dive into businesses' density around the new stations and how firms cluster is crucial. To this end, calculating business density in the two buffer areas around each station shows that density increased from 444 and 332 in 2010 for 500 and 750-meter buffers, to 1316 and 1125 in 2024.

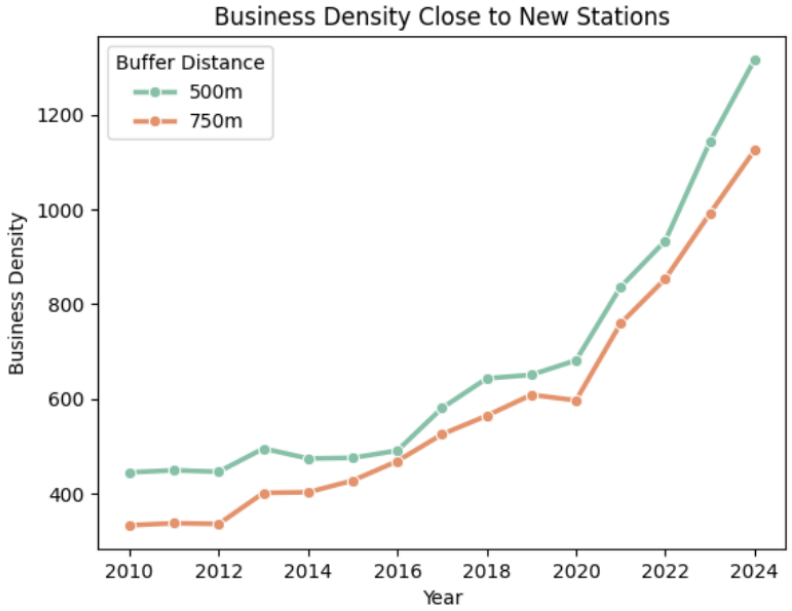


Figure 23: *Business density over time*

But is this increase in density equal for the two new stations? By looking at each station separately in **Figure-24**, Nine Elms seems to have a higher density over time for both buffers compared to Battersea Power Station. However, Battersea Power Station witnessed a sharper increase in density after the two stations were in operation, mainly for a 500m buffer.

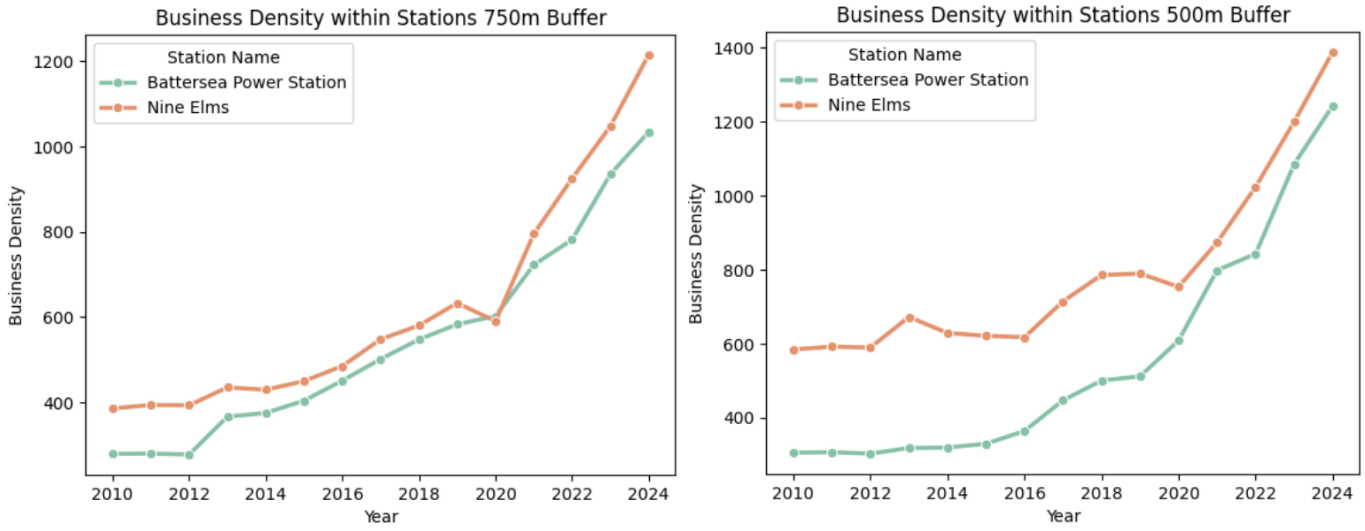


Figure 24: *Business density over time by NLE stations*

Examining the sectors' split of density for both stations reveals that office and retail are the two sectors with the highest increase in density, with the gap between them narrowing over time. However, looking at the two stations separately shows that the Retail density picked up more around Battersea Power Station, surpassing the retail density around Nine Elms, while the gap between the two stations' Offices narrowed over time. Lastly, **figure-25** shows that the Leisure sector started to pick up after the stations opened at a similar rate for both stations.

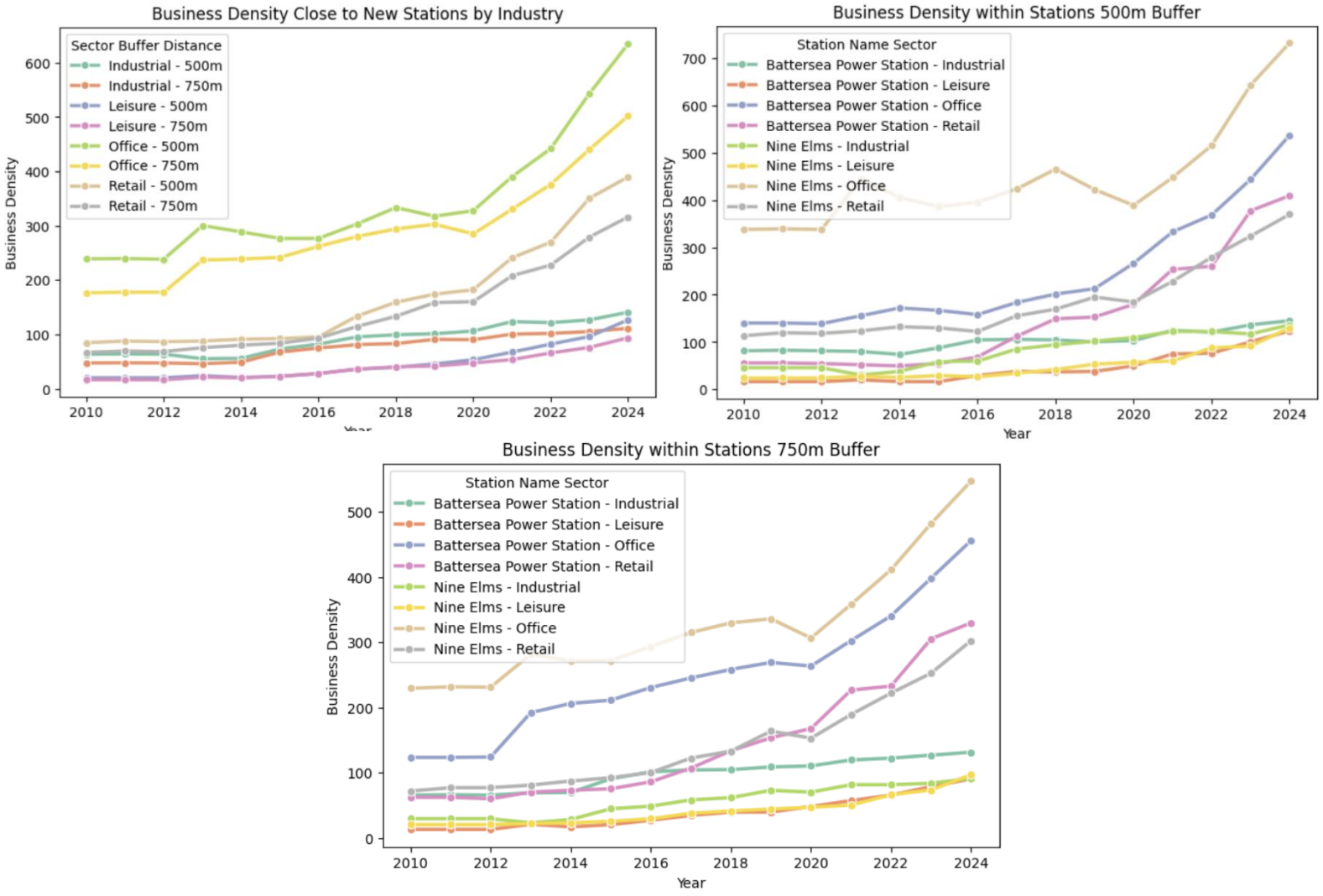


Figure 25: *business density by station and sector*

To understand how NLE stations support the agglomeration and formation of clusters, spatial clustering of businesses is employed. Hyperparameter tuning of the minimum cluster size and cluster selection epsilon is performed, to generate the most consistent and stable clusters, by grid searching the two-parameter space (min cluster size ranges from 15 to 100, and epsilon ranges from 0 to 200 meters), resulting in the best minimum cluster size of 15 and best epsilon distance of 0 meters across the years. The silhouette score is used to evaluate clusters' quality, whereby **Figure-26** shows the score over the years and for a given year across the search space.

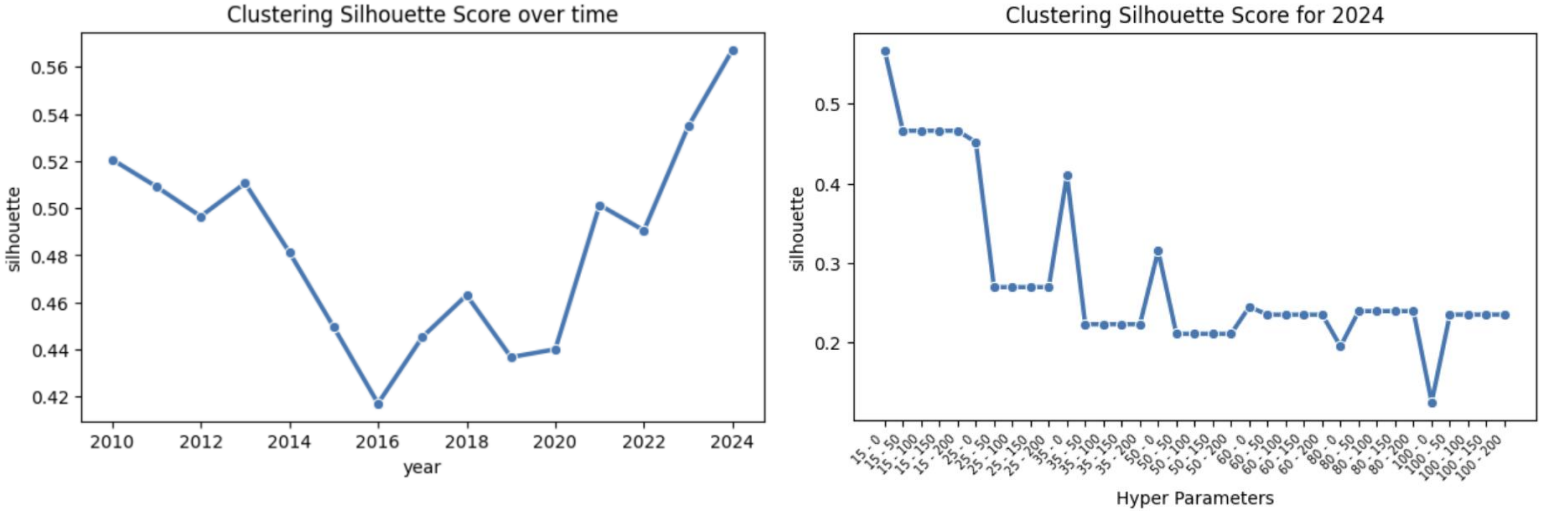


Figure 26: *Clusters Silhouette score over time and search space*

Overall Business Clusters:

Figure-27 shows that the number of overall business clusters started to pick up after the construction work began in 2015, and a sharp increase can be identified after the two stations are in operation (with a slight slowdown due to COVID-19 in 2020).

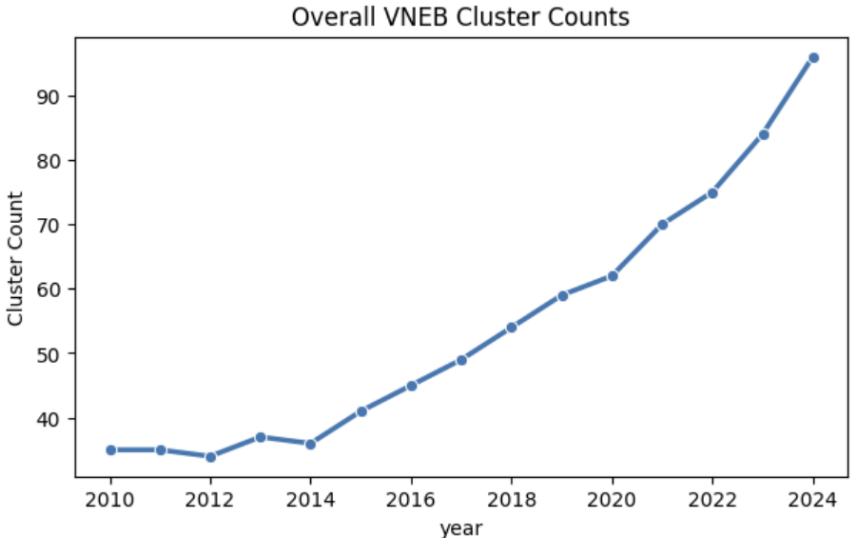


Figure 27*: Business cluster count over time*

To better understand the impact of new stations on cluster formation, clusters' number within a 500- and 750-meter buffers around stations is calculated over time, which shows an increase from 13 and 21 in 2010 to 40 and 63 in 2024 for 500 and 750-meter buffers. Total businesses in clusters around stations increased significantly, from 590 and 859 in 2010 for 500 and 750-meter buffers, to 1739 and 2891 in 2024.

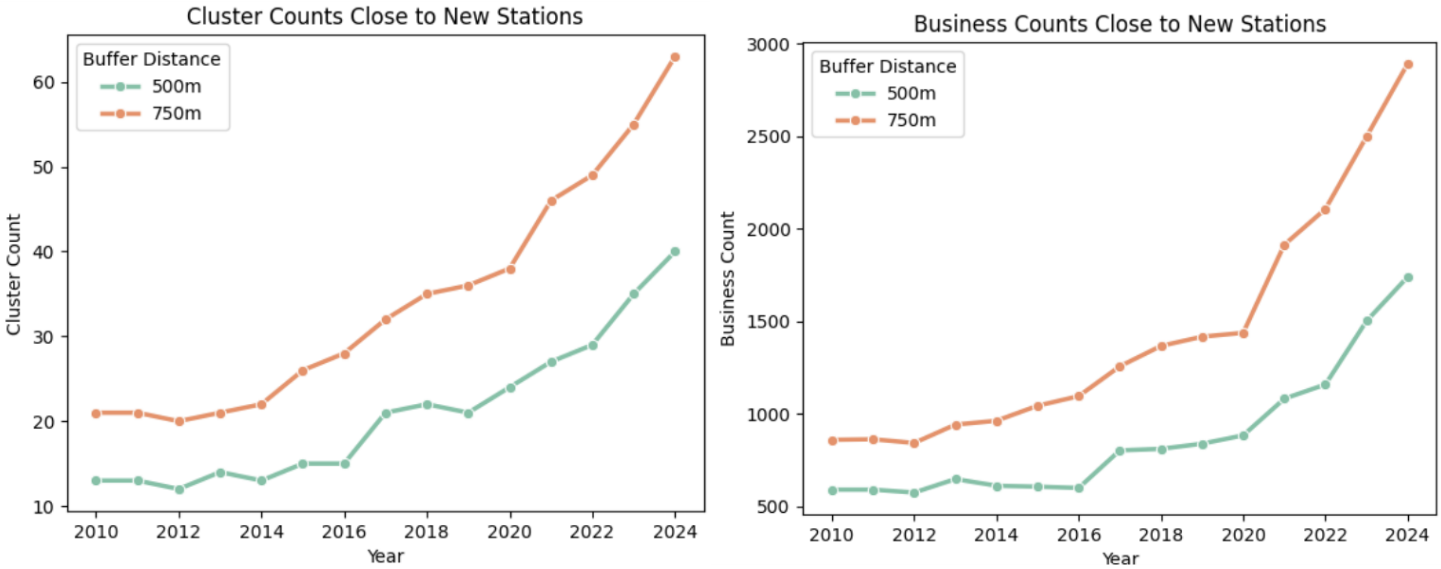


Figure 28: *Cluster and business counts within stations buffers*

Zooming in on the highest share sector for each cluster reveals that office-dominant clusters have increased since 2015 for both buffers. Retail-dominant clusters mainly picked up after the stations were in operation in 2021, while industrial-dominant clusters decreased around the opening after they had witnessed an increase since the construction work started. Finally, leisure dominates very few clusters, mainly in the last few years (only one cluster for 500 meters in 2024), indicating that leisure businesses are scattered across other industries' dominant clusters.

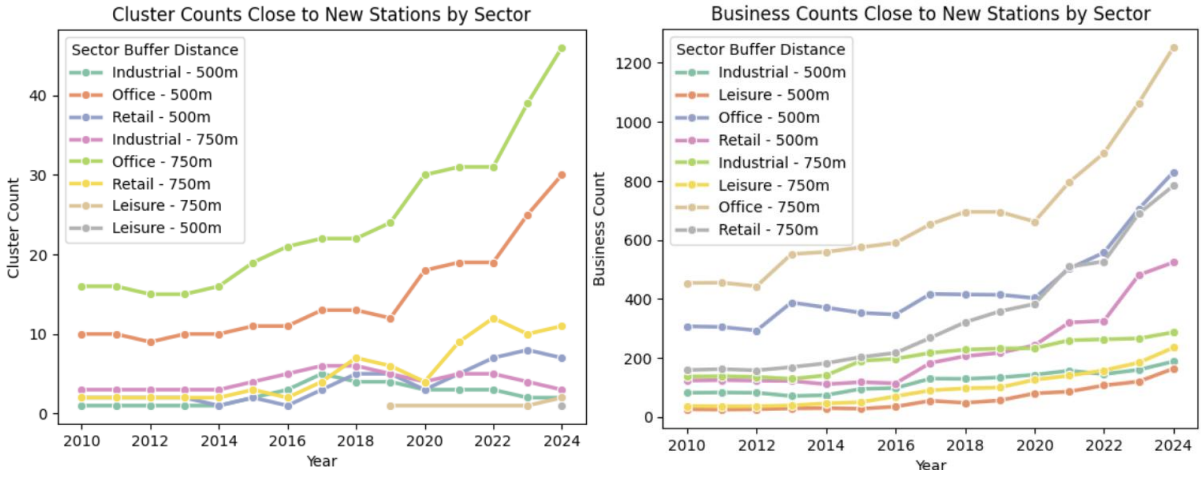


Figure 29*: Cluster and business counts within stations buffers by sector*

Cluster entropy based on industries is calculated to identify whether clusters are, on average, more diverse or homogeneous. Similarly, entropy for clusters based on accessibility is calculated to determine whether clusters are spread across different PTAL classes or not. **Figure-30** illustrates the increase in average sector diversity over time, indicating the formation of more multi-industry clusters. In contrast, the average accessibility diversity remained essentially unchanged, with a notable decline after 2021, remaining predominantly on the lower end, and suggesting a specific PTAL class dominating those clusters.

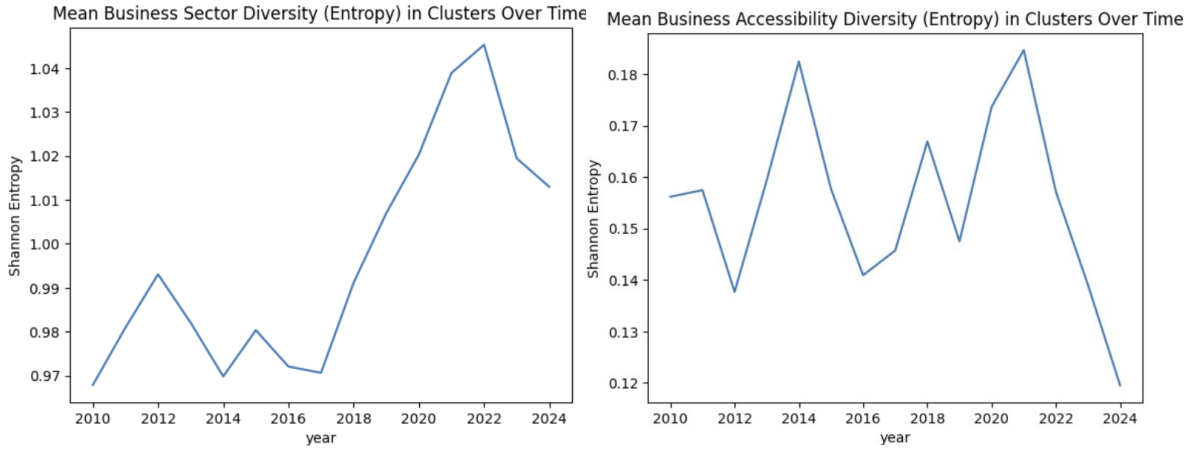


***Figure 30**: Cluster sector/accessibility diversity*

To identify the agglomeration type of formed clusters based on diversity and homogeneity (urbanisation versus localisation), entropy thresholds of the lower 30% quantiles (homogeneous) and the higher 70% quantiles (high-diversity) are used. Based on that, **figure-31** shows that the formation of highly diverse clusters picks up after 2015 and especially after the opening of new stations in 2021. Similarly, moderately diverse clusters, which represent a transformation from homogeneous to more diverse clusters, increase over time with sharp spikes around 2015 and 2021. On the contrary, homogeneous cluster formation fluctuated over time, with a drop in 2015 representing a transformation to more diverse clusters, and a slight increase after 2021. However, the gap between diverse and homogeneous clusters widens over time.

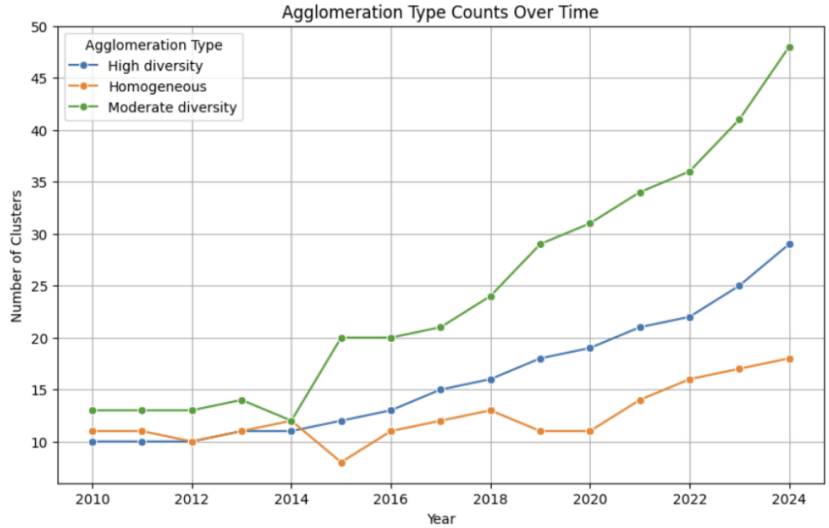


***Figure 31:** agglomeration types of clusters over time*

Figure-32 shows that diverse (high and moderate) clusters are formed closer to stations, with an average distance of around 500 meters from NLE stations. In contrast, homogeneous clusters are formed further away, on average approximately 900 meters from NLE stations.

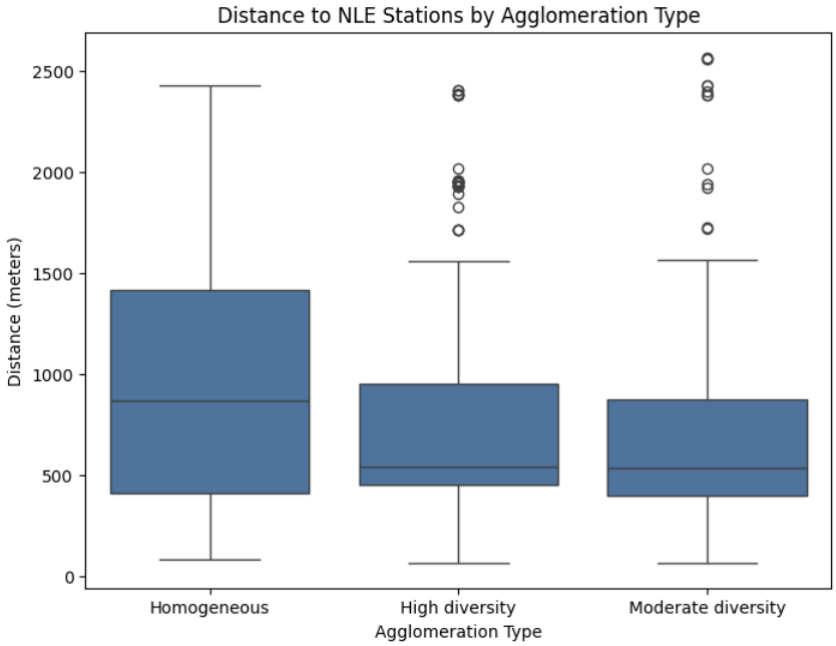


Figure 32: *Distance of agglomeration-type clusters from stations*

Figure-33 and **34** shows maps of highly diverse, moderately diverse, and sector-based homogeneous clusters distribution over the years around NLE stations. It demonstrates the increase in diverse cluster formation around Nine Elms and Battersea Power Station, whilst homogeneous clusters formed around stations are mainly of retail and office sectors.

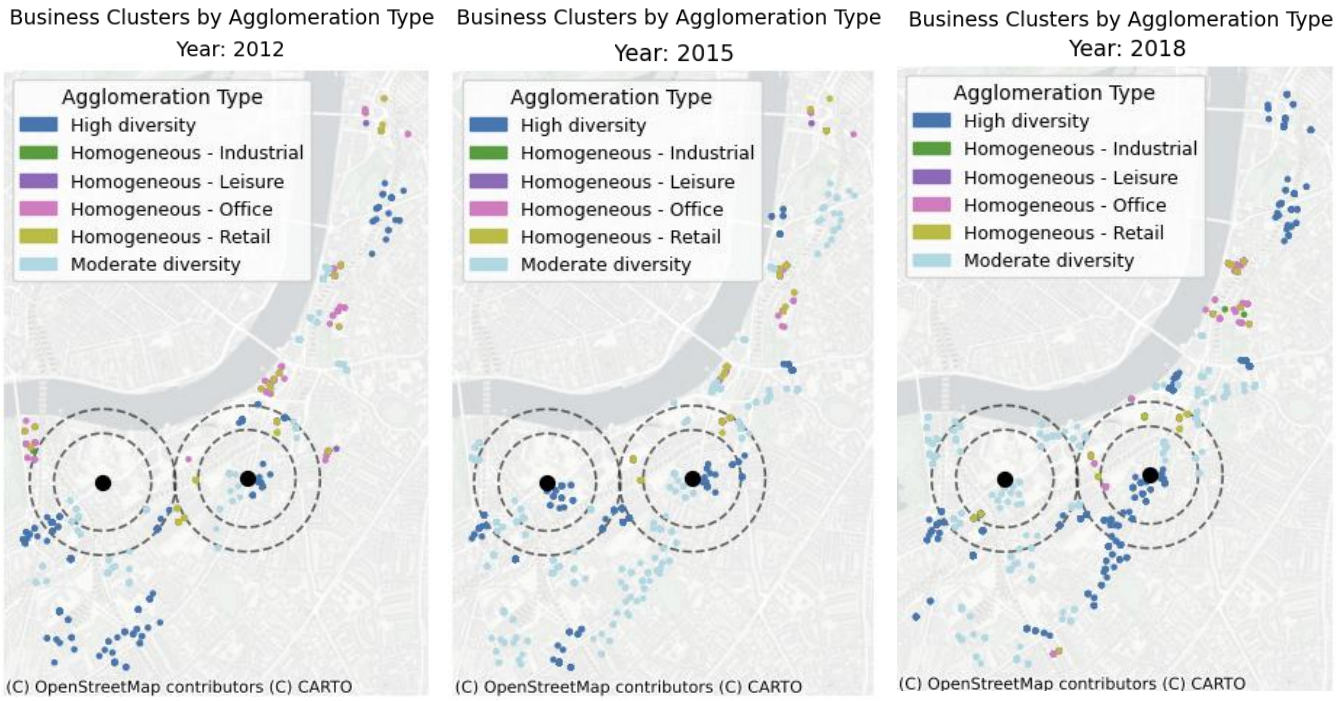


Figure 33: *Map of business clusters over time (years 2012, 2015 and 2018)*

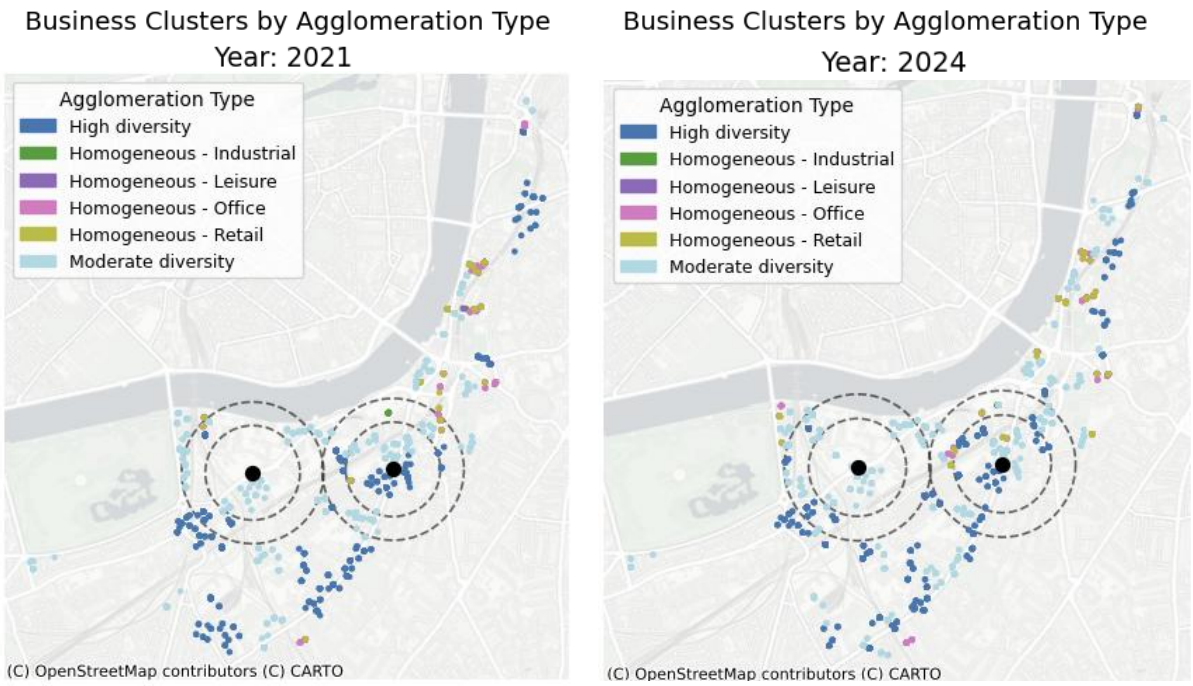


Figure 34: *Map of business clusters over time (years 2021 and 2024)*

6 Discussion:

This analysis's importance stems from being one of the first empirical studies focusing on how transport-led regeneration reshapes business dynamics in London, and the first to evaluate the Northern Line Extension (NLE). This analysis has examined the NLE impact on businesses' dynamics, focusing on different angles like business growth and redistribution, sectoral shifts, and business spatial distribution and agglomeration. The results showed that NLE had a positive impact on business formation mainly in 2015, suggesting that expectations drove new businesses. Given the time businesses require to get established, the increase in new businesses was notably pronounced during the NLE development initial phase. This aligns with (Credit, 2017) findings that a surge in new businesses occurred before the system stabilised; however, for the NLE case, the timing of this surge differed, occurring more around the construction kick-off, which can be attributed to the anticipation effect. In 2021, when NLE stations opened, London was still recovering from COVID-19. Furthermore, commercial rents and business rates had already experienced significant increases, which might be attributed to the NLE regeneration plan, thereby diminishing any potential influence NLE might have on business formation.

Conversely, when examining the NLE impact on businesses' relocation, a positive effect on higher accessibility areas in 2015 and 2021 was found. This effect was stronger when stations opened, indicating that businesses aimed to boost their visibility and capitalise on agglomeration by moving nearer to NLE stations. Most businesses relocated from other London boroughs to the highly accessible areas within the VNEB Opportunity Area to benefit from increased opportunities and exposure.

These results align with (Pogonyi, Graham and Carbo, 2021) findings, as the NLE opening didn't cause a net increase in business counts; instead, it led to economic activities redistribution. NLE opening had a highly significant impact on the number of relocating businesses. By capturing business movements between treated and untreated areas over time, this analysis goes beyond the approach employed by (Pogonyi, Graham and Carbo, 2021) which inferred displacement indirectly based on the spatial pattern of net changes in firm counts across distance bands. A key strength of this analysis lies in focusing on tracking individual businesses' relocation and movements, whether from low to high accessibility areas, or from outside to inside VNEB. This enabled the identification of the NLE impact on the different accessibility areas, whereby high areas benefited more than low areas, proving that accessibility improvements brought by transport investment have a key role in business relocation and formation decisions, as suggested by (Iseki and Jones, 2018)

This methodology's advantage, mainly the AITS model, stems from its ability to track the NLE impact trends after the intervention year. Although NLE announcements (whether construction phase or opening) led to a surge in new and relocated businesses, this increase didn't sustain itself over time and began to decline. This indicates an anticipation effect, possibly due to excitement and businesses' overestimation of NLE benefits, which started to normalise and decrease over time, revealing a mismatch between expected and actual

benefits. Another factor to consider is that businesses' data has only a few years of data after the stations opened, which might not be enough to draw a clear conclusion about trends.

Regarding years of intervention, the results demonstrate that each year showed a very distinct behaviour for NLE impact on businesses. While 2015 had a significant and profound effect on new business formation, 2021 was the year that drove a massive business relocation to high-accessibility areas introduced by NLE.

Zooming in on how NLE impacted sectoral shifts, the results show that the retail sector emerged as the winner, both in terms of new business in 2015 and relocated businesses in both 2015 and 2021. This indicates that some elements of sector shifts towards retail occurred, benefiting from increased footfall due to accessibility enhancements, and mainly concentrated around stations in high accessibility areas, aligning with (Yao and Hu, 2020). However, the office sector witnessed an increase in business relocation in 2021, the year when stations opened, highlighting the sensitivity of the office sector to having a reliable commuting solution for its staff and the completion of major regeneration projects, depending less on footfall compared to the retail sector. Finally, an interesting pattern for new businesses in the industrial sector was observed for both 2015 and 2021, whereby low accessibility areas attracted new businesses from this sector. This is aligned with the VNEB Opportunity Area Planning Framework (VNEB-OAPF) in promoting retail, office and leisure businesses mainly around NLE stations.

Spatially, the NLE impact on business density around stations showed an overall increase in business density over time, with higher density concentrated closer to stations (within 500 meters compared to 750 meters). Initially, there was a notable density gap between Battersea Power Station (BPS) and Nine Elms before the NLE project. Over time, this gap narrowed, highlighting the significant advantages NLE has delivered to both stations, especially BPS, which is designated as a ‘growth pole’. This aligns with (Banister and Thurstain-Goodwin, 2011) findings of transport contribution to concentration and density levels through firms' relocation to higher productivity areas. This increase in business density can foster productivity growth, as suggested by (Venables, 2007); however, it is hard to measure due to the lack of additional productivity data.

Results of sector-based density around stations show that both the retail and office sectors witnessed a spike over time (after the construction started and throughout stations’ opening). This can be evidence that NLE was successful in transforming VNEB to be an extension of the London CAZ area as planned by the TFL business case and the VNEB Opportunity Area framework. The VNEB-OAPF focuses on delivering two ‘growth poles’ at BPS and Vauxhall, with a focus towards retail and office businesses at BPS. This is evident in the density and formation results of retail and office business clusters within a 500-meter buffer around the station. However, for Nine Elms, while the planning framework focuses on residential and mixed-use development, results show that office business is the biggest winner within a 500-meter area of the station. This suggests that the VNEB-OAPF has achieved its goals with respect to commercial development.

Finally, to understand if there is any agglomeration effect, industry-specific AITS results show that overall business density has a positive impact on new and moved businesses. These findings align with the agglomeration economies theory that businesses located in dense areas likely benefit from labour pooling, customer access, and knowledge spillovers. These results align with (Chatman, Noland and Klein, 2016) findings on the overall business density impact on business formation. Additionally, new and relocated businesses have a positive relationship with business spatial dispersion within OAs, which means that increases in moved and new businesses caused the business landscape within OAs to expand and create more business' hubs and clusters. AITS findings align with clustering results, which show how NLE boosted the number of clusters around the two stations massively, creating more diverse clusters over time. By using entropy to identify whether those clusters are urbanisation or localisation, results show that more urbanisation clusters are formed over time and positioned around stations. However, localisation cluster counts witnessed a slow increase, remaining low in numbers and positioned further away from stations, aligning with (Song *et al.*, 2012) . This suggests that NLE supported the mixed-use economic development around stations with an increase in diverse clusters fostering innovation and resilience for economic shocks and benefiting from cross-sector interactions.

Reflecting on the London policy, these findings suggest that transport-led regeneration did indeed catalyse economic activity. However, this boost was predominantly due to the relocation of businesses rather than new investment, especially after the completion of NLE. This highlights the need for more targeted policies supporting new business formation, such as reducing business rates, rather than solely relying on the benefits of transport investment. Nevertheless, NLE enabled GLA to achieve several goals set by the VNEB-OAPF, mainly in unlocking development potential, nurturing commercial activities, and extending the London CAZ area.

Throughout this analysis, a few limitations were faced, which are worth acknowledging as follows:

- **Data availability on fine spatial scale**: to better understand changes in business patterns, this analysis needed fine granularity. Data at the OA-level, especially longitudinal data over years, was limited. During preprocessing, assumptions were made to disaggregate LSOA datasets to the OA-level. These datasets, though not business-related, were important as confounding factors. The disaggregation preserved proportional distributions and aggregate values at the LSOA level, but this method may influence the results.
- **Business datasets limitation:** the availability of business datasets is limited, especially those that are publicly available. No single dataset provides ratable value, rents, and other firm-level attributes simultaneously. To overcome this, Business Census data were merged with Business Rates data. Additionally, business datasets, both used and publicly available, lacked key variables such as turnover and employee counts that could further enrich the analysis.

- **Old Kent Road as control area:** A critical decision in the AITS model is choosing a suitable control area. Old Kent Road Opportunity Area was chosen because its regeneration vision, policy framework, demographics and other features closely mirror those of VNEB Opportunity Area. Yet, it has not benefited from its planned transport investment. However, TFL's recent planning policy restricts high-density development in Old Kent Road unless the required transportation investments are provided, thereby introducing a caveat. This planning restriction may affect businesses' patterns and growth in OKR, potentially impacting the AITS model results.
- Another limitation that might have an impact on the conclusions drawn from the clustering analysis is the absence of business productivity data. If such data exists, then clustering analysis can be computed more accurately by weighting data points using their productivity to produce clusters based on productivity concentration rather than solely depending on their locations, which might not necessarily truly reflect a high productivity concentration.

Although the dissertation examines multiple dimensions of transport-led regeneration's effects on businesses' dynamics and demographics, future research could expand the scope by investigating its impact on other aspects, such as productivity and businesses' survival rates. A key area for further study is identifying which businesses benefit most from such investment, through analysis of market-share shifts between local businesses and newly attracted enterprises, while distinguishing between chains, independent businesses, and business sizes. Such analysis could employ metrics like entry and exit rates, revenues, proportions of various business categories, and changes in footfall compared to competitors. This helps to assess whether transport-led regeneration fosters inclusive economic growth or leads to displacement and increased competitive pressure on existing firms. Future work might also explore whether such investments attract foreign-owned businesses, and how this may impact the wider London economy and the productivity of the CAZ zone. Finally, constructing a longitudinal dataset that incorporates detailed spatial and economic firm-level attributes would facilitate further empirical analysis and support future work.

7 Conclusion:

Uncovering the impact of transport-led regeneration on economic activities is a complex problem to crack, especially with a minimal body of research targeting investments in London. This analysis aimed to address this topic by studying the impact of NLE on business dynamics and demographics, focusing on business formation and redistribution, sectoral shifts and spatial concentration and agglomeration. To this end, it employed two main techniques, statistical modelling using AITS and spatial clustering using HDBSCAN. The analysis provides a nuanced understanding of how businesses and individual sectors benefited from such investment. Additionally, it demonstrates how new stations influenced agglomeration topology and business concentration. The analysis concluded that although NLE brought business formation at the start of the construction phase, it contributed mainly to attracting businesses from other parts of London after its completion, emphasising more on the redistribution of economic activities with a decaying trend over time. It demonstrated how sectors benefited from NLE at different capacities, whereby the retail sector benefited the most over time. Moreover, businesses in both retail and office sectors were the main dominant ones around the two new stations. Furthermore, NLE shaped business spatial concentration in the VNEB Opportunity Area, extending the London CAZ, and encouraging diverse business agglomeration around the two new stations while maintaining homogeneous business agglomeration further away from stations.

Finally, these findings reinforce debates around the role of transport-led regeneration in shifting economic activities rather than boosting business net growth and maintaining business growth after the system stabilises (anticipation effect). It highlights the need for complementary policies to existing transport-led regeneration policies that support new business formation and sustainable economic growth and ensure benefits are widely distributed.

References:

Alonso, William. (1964) *Location and land use : toward a general theory of land rent / William Alonso*. Cambridge, Mass: Harvard University Press

Banister, D. and Thurstain-Goodwin, M. (2011) 'Quantification of the non-transport benefits resulting from rail investment', *Journal of Transport Geography*, 19(2), pp. 212–223. Available at: https://doi.org/10.1016/j.jtrangeo.2010.05.001

Champagne, M.-P. and Dubé, J. (2023) 'The impact of transport infrastructure on firms' location decision: A meta-analysis based on a systematic literature review', *Transport Policy*, 131, pp. 139–155. Available at: https://doi.org/10.1016/j.tranpol.2022.11.015

Chatman, D.G. and Noland, R.B. (2011) 'Do Public Transport Improvements Increase Agglomeration Economies? A Review of Literature and an Agenda for Research', *Transport Reviews*, 31(6), pp. 725–742. Available at: https://doi.org/10.1080/01441647.2011.587908

Chatman, D.G. and Noland, R.B. (2014) 'Transit Service, Physical Agglomeration and Productivity in US Metropolitan Areas', *Urban Studies*, 51(5), pp. 917–937. Available at: https://doi.org/10.1177/0042098013494426

Chatman, D.G., Noland, R.B. and Klein, N.J. (2016) 'Firm Births, Access to Transit, and Agglomeration in Portland, Oregon, and Dallas, Texas', *Transportation Research Record: Journal of the Transportation Research Board*, 2598(1), pp. 1–10. Available at: https://doi.org/10.3141/2598-01

Church, A. (1990) 'Transport and urban regeneration in London Docklands', *Cities*, 7(4), pp. 289–303. Available at: https://doi.org/10.1016/0264-2751(90)90027-5

Comber, S. and Arribas-Bel, D. (2017) '"Waiting on the train": The anticipatory (causal) effects of Crossrail in Ealing', *Journal of Transport Geography*, 64, pp. 13–22. Available at: https://doi.org/10.1016/j.jtrangeo.2017.08.004

Credit, K. (2017) 'Transit-oriented economic development: The impact of light rail on new business starts in the Phoenix, AZ Region, USA', *Urban Studies*, 55(13), pp. 2838–2862. Available at: https://doi.org/10.1177/0042098017724119

D. Knowles, R. and Ferbrache, F. (2016) 'Evaluation of wider economic impacts of light rail investment on cities', *Journal of Transport Geography*, 54, pp. 430–439. Available at: https://doi.org/10.1016/j.jtrangeo.2015.09.002

De Bok, M. and Van Oort, F. (2011) 'Agglomeration economies, accessibility and the spatial choice behavior of relocating firms', *Journal of Transport and Land Use*, 4(1), p. 5. Available at: https://doi.org/10.5198/jtlu.v4i1.144

Debrezion, G., Pels, E. and Rietveld, P. (2007) 'The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis', *The Journal of Real Estate Finance and Economics*, 35(2), pp. 161–180. Available at: https://doi.org/10.1007/s11146-007-9032-z

Department for Transport (2021) ‘Understanding and quantifying transformational impacts from transport interventions: literature review’.

DFT, D. (2025) ‘Transport for London Spending Review Phase 2 Outcome’.

Dittmar, Hank. and Ohland, Gloria. (2003) *The new transit town : best practices in transit-oriented development / edited by Hank Dittmar, Gloria Ohland.* Washington, D.C. ; Island Press.

FTA (2025) *Transit-Oriented Development | FTA*. Available at: https://www.transit.dot.gov/TOD (Accessed: 4 July 2025)

Galster, G. *et al.* (2004) ‘Measuring the Impacts of Community Development Initiatives: A New Application of the Adjusted Interrupted Time-Series Method’, *Evaluation Review*, 28(6), pp. 502–538. Available at: https://doi.org/10.1177/0193841X04267090

GLA, G. (2012a) *Mayor hails finance agreement for Northern Line extension | London City Hall*. Available at: https://www.london.gov.uk/press-releases-4780 (Accessed: 28 July 2025)

GLA, G. (2012b) *Vauxhall Nine Elms Battersea Opportunity Area Planning Framework*. Available at: https://www.wandsworth.gov.uk/media/11751/vneb_oapf_2012_0.pdf

GLA, G. (2015) ‘Transport-led regeneration Has TfL’s Growth Fund risen to the challenge?’ Available at: https://www.london.gov.uk/sites/default/files/growth_fund_report_-_final_without_logo.pdf

Golub, A., Guhathakurta, S. and Sollapuram, B. (2012) ‘Spatial and Temporal Capitalization Effects of Light Rail in Phoenix: From Conception, Planning, and Construction to Operation’, *Journal of Planning Education and Research*, 32(4), pp. 415–429. Available at: https://doi.org/10.1177/0739456x12455523

Greater London Authority (2021) ‘The London Plan’. Available at: https://www.london.gov.uk/sites/default/files/the_london_plan_2021.pdf

HM Treasury, H.T. (2020) ‘National Infrastructure Strategy’. Available at: https://assets.publishing.service.gov.uk/media/5fbd810dd3bf7f5736c1a18f/NIS_final_web_single_page.pdf

Iseki, H. and Jones, R.P. (2018) ‘Analysis of firm location and relocation in relation to Maryland and Washington, DC metro rail stations’, *Research in Transportation Economics*, 67, pp. 29–43. Available at: https://doi.org/10.1016/j.retrec.2016.11.003

Knowles, R.D., Ferbrache, F. and Nikitas, A. (2020) ‘Transport’s historical, contemporary and future role in shaping urban development: Re-evaluating transit oriented development’, *Cities*, 99, p. 102607. Available at: https://doi.org/10.1016/j.cities.2020.102607

Ko, K. and Cao, X. (2013) ‘The Impact of Hiawatha Light Rail on Commercial and Industrial Property Values in Minneapolis’, *Journal of Public Transportation*, 16(1), pp. 47–66. Available at: https://doi.org/10.5038/2375-0901.16.1.3

Laird, J.J. and Venables, A.J. (2017) ‘Transport investment and economic performance: A framework for project appraisal’, *Transport Policy*, 56, pp. 1–11. Available at: https://doi.org/10.1016/j.tranpol.2017.02.006

Lakshmanan, T.R. (2011) ‘The broader economic consequences of transport infrastructure investments’, *Journal of Transport Geography*, 19(1), pp. 1–12. Available at: https://doi.org/10.1016/j.jtrangeo.2010.01.001

Lee, J.K. (2021) ‘Transport infrastructure investment, accessibility change and firm productivity: Evidence from the Seoul region’, *Journal of Transport Geography*, 96, p. 103182. Available at: https://doi.org/10.1016/j.jtrangeo.2021.103182

Lin, J.-J. and Yang, S.-H. (2019) ‘Proximity to metro stations and commercial gentrification’, *Transport Policy*, 77, pp. 79–89. Available at: https://doi.org/10.1016/j.tranpol.2019.03.003

Lindgren, E., Pettersson-Lidbom, P. and Tyrefors, B. (2021) ‘The Causal Effect of Transport Infrastructure: Evidence from a New Historical Database□’

Mejia-Dorantes, L., Paez, A. and Vassallo, J.M. (2012) ‘Transportation infrastructure impacts on firm location: the effect of a new metro line in the suburbs of Madrid’, *Journal of Transport Geography*, 22, pp. 236–250. Available at: https://doi.org/10.1016/j.jtrangeo.2011.09.006

Merz, S.K. (2009) ‘Vauxhall Nine Elms Battersea Opportunity Area Transport Study’.

Mohammad, S.I. *et al.* (2013) ‘A meta-analysis of the impact of rail projects on land and property values’, *Transportation Research Part A: Policy and Practice*, 50, pp. 158–170. Available at: https://doi.org/10.1016/j.tra.2013.01.013

Neumark, D., Zhang, J. and Wall, B. (2005) ‘Employment Dynamics and Business Relocation: New Evidence from the National Establishment Time Series’.

Ozmen-Ertekin, D., Ozbay, K. and Holguin-Veras, J. (2007) ‘Role of Transportation Accessibility in Attracting New Businesses to New Jersey’, *Journal of Urban Planning and Development*, 133(2), pp. 138–149. Available at: https://doi.org/10.1061/(asce)0733-9488(2007)133:2(138)

Pogonyi, C.G., Graham, D.J. and Carbo, J.M. (2021) ‘Metros, agglomeration and displacement. Evidence from London’, *Regional Science and Urban Economics*, 90, p. 103681. Available at: https://doi.org/10.1016/j.regsciurbeco.2021.103681

Redding, S.J. and Turner, M.A. (2015) ‘Transportation Costs and the Spatial Organization of Economic Activity’, in *Handbook of Regional and Urban Economics*. Elsevier, pp. 1339–1398. Available at: https://doi.org/10.1016/b978-0-444-59531-7.00020-x

Reynolds, P. (1994) ‘Autonomous Firm Dynamics and Economic Growth in the United States, 1986–1990’, *Regional Studies*, 28(4), pp. 429–442. Available at: https://doi.org/10.1080/00343409412331348376

Schuetz, J. (2015) ‘Do rail transit stations encourage neighbourhood retail activity?’, *Urban Studies*, 52(14), pp. 2699–2723. Available at: https://doi.org/10.1177/0042098014549128

Song, Y. *et al.* (2012) 'Industrial agglomeration and transport accessibility in metropolitan Seoul', *Journal of Geographical Systems*, 14(3), pp. 299–318. Available at: https://doi.org/10.1007/s10109-011-0150-z

Southwark Council, S.C. (2017) 'Old Kent Road Area Action Plan Opportunity Area Planning Framework'. Available at: https://www.southwark.gov.uk/sites/default/files/2024-11/Old%20Kent%20Road%20Area%20Action%20Plan%20Opportunity%20Area%20Planning%20Framework.pdf

TFL, T. (2013) 'Northern Line Extension Economic & Business Case'.

TFL, T. (2015) 'Assessing transport connectivity in London'. Available at: https://content.tfl.gov.uk/connectivity-assessment-guide.pdf

Thakur, N. *et al.* (2021) 'Augmented SBERT: Data Augmentation Method for Improving Bi-Encoders for Pairwise Sentence Scoring Tasks'. arXiv. Available at: https://doi.org/10.48550/arXiv.2010.08240

Tornabene, S. and Nilsson, I. (2021) 'Rail transit investments and economic development: Challenges for small businesses', *Journal of Transport Geography*, 94, p. 103087. Available at: https://doi.org/10.1016/j.jtrangeo.2021.103087

United Nations Department of Economic and Social Affairs (2018) 'World Urbanization Prospects The 2018 Revision'.

Venables, A.J. (2007) 'Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation', *Journal of Transport Economics and Policy*, 41.

Wang, W. *et al.* (2020) 'MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers'. arXiv. Available at: https://doi.org/10.48550/arXiv.2002.10957

Wang, W., Zhong, M. and Hunt, J.D. (2019) 'Analysis of the Wider Economic Impact of a Transport Infrastructure Project Using an Integrated Land Use Transport Model', *Sustainability*, 11(2), p. 364. Available at: https://doi.org/10.3390/su11020364

Weber, Alfred. and Friedrich, C.J. (1929) *Theory of the location of industries / Alfred Weber ; English edition with introduction and notes by Carl J. Friedrich.* Chicago ; University of Chicago Press.

Yao, L. and Hu, Y. (2020) 'The impact of urban transit on nearby startup firms: Evidence from Hangzhou, China', *Habitat International*, 99, p. 102155. Available at: https://doi.org/10.1016/j.habitatint.2020.102155

Yi, Z. (2023) 'Use Siamese BERT-Networks to fine-tune minBERT with downstream tasks'.

Zhang, Y. and Cheng, L. (2023) 'The role of transport infrastructure in economic growth: Empirical evidence in the UK', *Transport Policy*, 133, pp. 223–233. Available at: https://doi.org/10.1016/j.tranpol.2023.01.017

Zuk, M. *et al.* (2018) ‘Gentrification, Displacement, and the Role of Public Investment’, *Journal of Planning Literature*, 33(1), pp. 31–44. Available at: https://doi.org/10.1177/0885412217716439

Appendix A:

This appendix will include full coefficient tables for all the dependent variables investigated in this analysis and an evaluation table for industry-specific models. Coefficient tables include the full coefficient table for overall new and moved businesses, as the table included in the result section includes only the most significant variables. As well as coefficient tables for industry-specific models.

Overall Business Coefficients:

Dependent Variable	New Business Density				Moved Business Density			
Accessibility	high		low		high		low	
year	2015	2021	2015	2021	2015	2021	2015	2021
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	35	35	34	34	35	35	34	34
Pearson Chi2	3780	3110	7140	6860	3370	5030	19600	48700
Pseudo R-Square	0.689	0.692	0.457	0.443	0.956	0.931	0.834	0.68
Demographics Features								
Population	0.192**	0.205**	0.252**	0.245**	0.098*	0.110**	0.138**	0.113**
Working-age population	0.837	0.915	1.783**	1.571**	1.315**	1.271**	2.314**	2.339**
Asian %	1.037	0.718	-2.161*	-2.314*	9.348**	8.895**	-5.635**	-4.495**
Black %	0.67	0.42	0	0	-0.459	-1.228*	0	0
In Deprivation	-0.054	-0.06	0.023	0.033	-0.015	-0.103*	0.05	0.069**
Out Deprivation	0.05	0.062	0.075**	0.082**	-0.05	-0.01	0.036	0.095**
Local Agglomeration Features								
Business density 1lag	0.204**	0.105	0	0	0.004	-0.02	0	0
New business density 1lag	0	0	0	0	0.07	0.088*	0.009	-0.015
Moved business density 1lag	-0.093*	-0.069	0.01	0.014	0	0	0	0
Average distance	-0.143	-0.048	-0.072	-0.058	-0.290**	0.019	-0.411**	-0.421**
Jobs count	0.162**	0.202**	-0.052*	-0.065**	0.096	0.319**	0.178**	0.274**
Office Business density 1lag	0.245**	0.286**	0.038	0.028	0.099	0.1	0.074**	0.115**
Office average distance	0.039	0.006	-0.012	0.014	0.184*	0.256**	0.263**	0.349**
Retail Business density 1lag	0.253**	0.262**	0.304**	0.306**	0.445**	0.451**	0.078**	0.073**
Retail average distance	0.341**	0.341**	0.238**	0.226**	0.478**	0.364**	0.055	0.059
Retail Jobs percentage	0.549	0.908**	0.21	0.163	-0.919**	-0.762**	-0.023	-0.077
Leisure Business density 1lag	0.026	0.026	-0.051*	-0.032	-0.150**	-0.188**	-0.005	0.001

Leisure average distance	0.025	0.015	0.001	0.001	0.043	0.039	0.201**	0.215**
Leisure Jobs percentage	-0.044	0.015	0.383	0.14	-0.907*	-0.399	-0.376	-0.931**
Industrial Business density 1lag	-0.077	-0.068	0.160**	0.153**	0.056	0.062	0.090**	0.090**
Industrial average distance	0.244**	0.234**	0.154**	0.145**	0.028	0.046	0.095**	0.083**
Industrial Jobs percentage	0.948**	0.721*	0.636**	0.492**	1.354**	1.012**	0.141	-0.228
Business Dynamics Features								
Dissolution rate	-0.690*	-0.444	0.095	-0.044	-0.225	1.341**	0.408**	1.372**
Survival rate	0.17	0.345*	0.570**	0.344**	0.246	1.143**	-0.141	0.079
Business rate	-0.046	-0.051	-0.015	-0.033	-0.038	0	-0.009	0.01
Rental valuation	0.054	0.065	-0.043	-0.025	0.021	-0.019	-0.063	-0.133**
Properties Features								
Property rent	0.083	0.184**	0.062	0.074*	-0.044	0.332**	0.128**	0.202**
Residents churn %	-0.732	-1.079*	0.042	0.525	-0.426	-1.164**	1.490**	1.500**
AITS Features								
Treatment	-0.311*	-0.054	-0.493**	-0.148*	-0.123	-0.21	1.179**	0.446**
Intervention	-0.195	0.151	0.288**	0.434**	-1.444**	-0.949**	-1.306**	-0.031
Treatment post-intervention	0.775**	0.309	0.478**	-0.14	0.785**	0.923**	-1.126**	-0.826**
Trend	0.263**	0.072**	0.131**	0.088**	2.367**	0.346**	2.498**	0.354**
Trend post-intervention	-0.290**	-0.269*	-0.206**	-0.380**	-2.390**	-0.676**	-2.432**	-0.567**
Trend-treatment post-intervention	-0.066*	-0.233*	0.011	0.131	-0.003	-0.231*	0.049**	0.247**
Other Features								
Area	-0.524**	-0.584**	-0.392**	-0.406**	-0.539**	-0.870**	-0.452**	-0.540**
COVID-19 dummy	-0.006	-0.502**	0.027	-0.393**	0.067	-0.898**	0.092	-0.624**

Table 11: *Overall new and moved business AITS model coefficients*

Retail Business Coefficients:

Industry	**Retail**							
Dependent Variable	**Moved Business Density**				**New Business Density**			
Accessibility	**High**		**Low**		**High**		**Low**	
Year	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	24	24	25	25	24	24	25	25

Pearson Chi2	6370	9210	24100	35300	7660	7860	19600	20300
Pseudo R-Square	0.938	0.919	0.82	0.767	0.7	0.681	0.587	0.576
Demographics Features								
Population	0.266**	0.251**	0.336**	0.414**	0.231**	0.265**	0.176**	0.112**
Working-age population	0.81	-0.229	1.709**	1.392**	2.375**	2.192**	1.834**	1.712**
Asian %	4.247**	7.227**	-6.375**	-4.996**	-1.469	-0.803	-4.288**	-4.197**
Black %	-0.283	-1.106*	0	0	-0.242	-0.336	0	0
In Depriv	-0.158**	-0.142**	0.103**	0.118**	0.018	0.044	0.016	0.033
Out Depriv	-0.006	-0.011	0.065*	0.053	0.204**	0.178**	0.094**	0.090**
Local Agglomeration Features								
Business density 1lag	0.239**	0.238**	-0.034	-0.05	0.368**	0.237**	0.051	0.133**
Industry Business density 1lag	0.323**	0.199**	0.180**	0.237**	0.372**	0.441**	0.491**	0.464**
New business density 1lag	0.120**	0.197**	0.02	-0.055*	0	0	0	0
Moved business density 1lag	0	0	0	0	-0.067	-0.075	-0.069**	-0.067**
Average distance	0.568**	0.741**	0.309**	0.321**	0.455**	0.399**	0.445**	0.383**
Jobs count	-0.006	-0.096*	0.160**	0.149**	0.118**	0.095*	0.023	0.018
Business Dynamics Features								
Dissolution rate	2.304**	2.292**	2.160**	2.322**	-1.426**	-1.440**	-0.467**	-0.499**
Survival rate	-0.116	-0.027	1.225**	1.229**	0.310**	0.232*	0.491**	0.527**
Business rate	0	0	-0.347**	-0.294**	0	0	-0.008	-0.034
Rental valuation	-0.095*	-0.109**	-0.015	-0.095*	-0.038	-0.087*	-0.170**	-0.146**
Properties Features								
Property rent	0.011	0.157**	0.356**	0.334**	0.049	0.062	-0.002	-0.024
Residents churn %	2.354**	2.058**	-0.694*	-0.325	0.515	0.683	1.390**	1.653**
AITS Features								
Treatment	-1.113**	0.622**	0.678**	0.622**	-1.107**	-0.350**	-0.683**	-0.240**
Intervention	-2.176**	-1.484**	-0.753**	-0.069	-0.082	0.124	0.421**	0.303*
Treatment post-intervention	2.588**	1.934**	0.349*	-0.238	1.791**	0.639	0.577**	0.489*
Trend	1.910**	0.442**	1.628**	0.228**	0.059	0.214**	-0.01	0.134**
Trend post-intervention	-1.749**	-0.258*	-1.652**	-0.487**	0.025	-0.398**	0.046	-0.304**
Trend- treatment post-intervention	0.016	-0.372**	-0.074**	0.036	-0.129**	-0.296*	0.055**	0.033
Other Features								
Area	-0.325**	-0.456**	-0.618**	-0.612**	-0.442**	-0.445**	-0.722**	-0.696**

COVID-19 dummy	0.013	-0.642**	0.248**	-0.425**	-0.157	-0.837**	0.186**	-0.143

Table 12: *Retail business AITS model coefficients*

Office Business Coefficients:

Industry	**Office**							
Dependent Variable	**Moved Business Density**				**New Business Density**			
Accessibility	**High**		**Low**		**High**		**Low**	
Year	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	25	25	25	25	25	25	25	25
Pearson Chi2	4490	6070	52200	25100	6120	6100	13100	12300
Pseudo R-Square	0.893	0.856	0.816	0.637	0.74	0.725	0.529	0.477
Demographics Features								
Population	0.191**	0.198**	0.132**	0.023	0.148**	0.165**	0.319**	0.373**
Working-age population	3.122**	3.455**	1.970**	1.100**	0.976*	0.761	1.912**	1.682**
Asian %	7.684**	8.006**	-3.465**	-1.65	0.642	0.199	-3.730**	-3.133**
Black %	-1.288*	-1.982**	0	0	-0.086	-0.207	0	0
In Depriv	-0.102*	-0.136**	-0.05	0.018	-0.247**	-0.250**	0.075**	0.063*
Out Depriv	-0.003	0.071	0.004	0.156**	-0.103*	-0.086	0.034	0.047
Local Agglomeration Features								
Business density 1lag	0.257**	0.121*	0.309**	-0.001	0.286**	0.06	0.536**	0.402**
Industry Business density 1lag	0.138*	0.251**	-0.038	0.124**	0.387**	0.537**	-0.202**	-0.116*
New business density 1lag	0.045	0.068*	0.008	-0.004	0	0	0	0
Moved business density 1lag	0	0	0	0	0.043	0.065	-0.036	-0.017
Average distance	0.160**	0.394**	0.092**	0.043	0.151**	0.254**	0.252**	0.352**
Jobs count	0.134**	0.361**	0.082**	0.150**	0.102*	0.200**	-0.043	-0.061*
Business Dynamics Features								
Dissolution rate	-0.298	0.831**	0.527**	0.498**	-0.389	-0.141	-0.784**	-0.641**
Survival rate	0.138	0.559**	0.087	-0.024	0.439**	0.28	0.277**	0.172
Business rate	-0.423**	-0.089	0.154**	0.579**	-0.409**	-0.322**	0.09	0.103*
Rental valuation	0.263**	-0.088	-0.221**	-0.948**	0.323**	0.267**	-0.175**	-0.194**
Properties Features								
Property rent	-0.001	0.283**	0.201**	0.584**	0.284**	0.380**	0.223**	0.247**

Residents churn %	-0.815*	-1.619**	2.309**	1.827**	-2.736**	-2.561**	-0.607*	-0.28
AITS Features								
Treatment	-0.323	-0.401**	1.217**	0.985**	-0.035	-0.052	0.118	0.152*
Intervention	-0.906**	-1.078**	-1.004**	-0.447**	0.128	-0.103	0.029	0.324*
Treatment post-intervention	0.337	1.381**	-1.662**	-0.503*	0.107	0.442	0.017	-0.431*
Trend	2.185**	0.383**	2.264**	0.324**	0.334**	0	0.496**	0.111**
Trend post-intervention	-2.222**	-0.683**	-2.224**	-0.450**	-0.505**	-0.117	-0.611**	-0.485**
Trend- treatment post-intervention	0.035	-0.397**	0.131**	0.214**	-0.013	-0.208	0.019	0.217**
Other Features								
Area	-0.371**	-0.683**	-0.414**	-0.719**	-0.385**	-0.512**	-0.359**	-0.450**
COVID-19 dummy	0.014	-1.022**	0.052	-0.527**	0.472**	-0.109	-0.079	-0.654**

Table 13: *Office business AITS model coefficients*

Industrial Business Coefficients:

Industry	**Industrial**							
Dependent Variable	**Moved Business Density**				**New Business Density**			
Accessibility	**High**		**Low**		**High**		**Low**	
Year	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**	**2015**	**2021**
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	25	24	25	25	25	24	25	25
Pearson Chi2	11300	16500	30700	41600	15800	13600	30900	30900
Pseudo R-Square	0.883	0.861	0.702	0.659	0.571	0.608	0.534	0.531
Demographics Features								
Population	0.217**	0.219**	-0.013	0.061*	0.315**	0.254**	0.165**	0.083**
Working-age population	1.408*	1.419*	1.051**	0.985**	-1.785**	-1.370**	1.538**	1.079**
Asian %	10.673**	8.241**	-5.743**	-3.663**	-2.976*	-0.829	1.033	1.138
Black %	-0.342	0.305	0	0	2.582**	1.689**	0	0
In Depriv	0.009	-0.003	0.068*	0.106**	0.192**	0.147**	0.103**	0.167**
Out Depriv	-0.169**	-0.180**	0.107**	0.187**	0.367**	0.215**	0.075**	0.107**
Local Agglomeration Features								
Business density 1lag	0.118*	0.027	0.093**	0.144**	0.555**	0.515**	0.263**	0.352**
Industry Business density 1lag	0.311**	0.300**	0.075*	0.06	0.321**	0.313**	0.422**	0.377**

New business density 1lag	-0.038	-0.029	-0.003	-0.02	0	0	0	0
Moved business density 1lag	0	0	0	0	0.065	0.089*	-0.068**	-0.106**
Average distance	0.562**	0.514**	0.385**	0.450**	0.436**	0.454**	0.548**	0.555**
Jobs count	-0.232**	-0.264**	-0.159**	-0.225**	0.006	0.180**	-0.029	-0.069*
Business Dynamics Features								
Dissolution rate	-0.264	-0.074	1.828**	2.002**	-0.508**	-0.783**	-0.086	-0.136
Survival rate	0.925**	1.300**	1.881**	1.912**	0.698**	1.451**	0.808**	0.910**
Business rate	-0.765**	-0.217**	0.058	0.074*	0.279**	0.021	-0.076	-0.099*
Rental valuation	0.517**	0	0	-0.052	-0.433**	0	-0.097*	-0.093*
Properties Features								
Property rent	0.012	0.283**	0.096*	0.159**	0.017	-0.088	-0.202**	-0.133**
Residents churn %	1.762**	1.220*	2.942**	3.042**	1.470**	1.244**	0.382	0.962**
AITS Features								
Treatment	-0.650*	0.424**	1.373**	0.339**	-0.006	0.137	-0.680**	-0.360**
Intervention	-1.184**	-1.256**	0.674**	0.217	-0.275	0.568	0.402**	0.517**
Treatment post-intervention	1.787**	1.719**	-1.600**	-1.758**	0.746**	0.207	0.854**	1.045**
Trend	3.017**	0.442**	1.437**	0.347**	0.256**	0.160**	-0.026	0.049**
Trend post-intervention	-2.909**	-0.479**	-1.464**	-0.756**	-0.250**	-0.729**	-0.024	-0.339**
Trend- treatment post-intervention	-0.116**	-0.697**	0.076**	0.549**	0.005	0.114	-0.048*	-0.358**
Other Features								
Area	-0.289**	-0.215**	-0.657**	-0.668**	-0.122*	-0.286**	-0.341**	-0.371**
COVID-19 dummy	0.389**	-0.734**	0.1	-0.718**	0.206*	-0.606**	-0.002	-0.448**

Table 14: *Industrial business AITS model coefficient*

Leisure Business Coefficients Table:

Industry	Leisure							
Dependent Variable	Moved Business Density				New Business Density			
Accessibility	High		Low		High		Low	
Year	2015	2021	2015	2021	2015	2021	2015	2021
Model Details								
No. Observations	1125	1125	2640	2640	1125	1125	2640	2640
Model DoF	25	25	25	25	25	25	25	25
Pearson Chi2	25400	17400	55000	83300	15600	16600	34000	36100
Pseudo R-Square	0.859	0.839	0.765	0.716	0.777	0.771	0.599	0.543
Demographics Features								
Population	-0.017	-0.078	-0.125**	-0.219**	-0.338**	-0.350**	0.107**	0.155**
Working-age population	1.555*	1.300*	3.590**	3.598**	2.610**	2.550**	2.582**	3.071**
Asian %	4.218**	-4.295**	-12.935**	-12.543**	5.727**	5.371**	-1.113	-1.522
Black %	1.064	4.905**	0	0	6.820**	5.898**	0	0
In Depriv	0.02	0.139*	-0.126**	-0.073*	0.094	0.139*	-0.251**	-0.229**
Out Depriv	0.052	0.017	-0.121**	-0.099**	-0.199**	-0.051	-0.239**	-0.180**
Local Agglomeration Features								
Business density 1lag	0.015	-0.064	-0.161**	-0.165**	0.356**	0.428**	0.351**	0.264**
Industry Business density 1lag	0.068	0.570**	-0.191**	-0.200**	0.190**	0.193**	0.04	0.044
New business density 1lag	0.160**	-0.016	0.225**	0.238**	0	0	0	0
Moved business density 1lag	0	0	0	0	0.212**	0.166**	-0.006	-0.005
Average distance	0.631**	0.621**	0.855**	0.874**	0.639**	0.619**	0.447**	0.495**
Jobs count	0.114	0.063	-0.156**	-0.157**	0.336**	0.362**	0.04	-0.003
Business Dynamics Features								
Dissolution rate	3.414**	2.638**	1.622**	1.791**	-1.260**	-1.271**	-0.552**	-0.611**
Survival rate	3.701**	2.615**	2.294**	2.261**	0.926**	1.248**	1.003**	1.086**
Business rate	0.329**	0.139	-0.136**	-0.196**	0.092	0.154	0.004	0.006
Rental valuation	-0.356**	-0.044	0.014	0.148**	0.042	0.051	0.03	0.054
Properties Features								
Property rent	0.653**	0.561**	-0.047	-0.097*	0.214**	0.265**	-0.384**	-0.300**
Residents churn %	0.695	3.969**	1.902**	3.399**	0.541	0.571	-0.156	0.304

AITS Features								
Treatment	-2.509**	-0.828**	1.437**	0.392**	1.798**	0.855**	-0.359**	-0.205*
Intervention	-3.235**	-1.195**	0.042	0.901**	-0.3	1.470**	-0.575**	0.380*
Treatment post-intervention	3.082**	1.304**	-1.913**	-1.327**	-0.378	-1.719**	0.583**	0.486*
Trend	1.263**	0.256**	1.420**	0.293**	0.679**	0.154**	0.906**	0.179**
Trend post-intervention	-1.224**	-0.102	-1.500**	-0.865**	-0.534**	-0.608**	-0.941**	-0.536**
Trend- treatment post-intervention	-0.104**	-0.185	0.171**	0.647**	-0.217**	0.076	-0.049*	-0.245**
Other Features								
Area	-1.065**	-0.686**	-1.353**	-1.432**	-1.233**	-1.210**	-0.701**	-0.716**
COVID-19 dummy	-0.663**	-1.196**	0.203*	-0.685**	0.134	-0.608**	0.068	-0.649**

Table 15: *Leisure business AITS model coefficient*

Industry-Specific Models Evaluation:

Model	Cox-Snell R2	McFaddens R2	Dispersion	(Poisson vs NB) p-value / LR Test
Retail New Business high-15	0.6955	0.1389	6.97	0.0000 / 68476.46
Office New Business high-15	0.7403	0.1385	5.57	0.0000 / 135178.98
Industrial New Business high-15	0.5713	0.1310	14.40	0.0000 / 35515.64
Leisure New Business high-15	0.7767	0.2589	14.16	0.0000 / 22242.69
Retail New Business low-15	0.5873	0.1163	7.51	0.0000 / 113610.50
Office New Business low-15	0.5290	0.0921	5.03	0.0000 / 129197.38
Industrial New Business low-15	0.5339	0.1216	11.82	0.0000 / 69433.18
Leisure New Business low-15	0.5987	0.1794	13.00	0.0000 / 40347.18
Retail Moved Business high-15	0.9383	0.3227	5.79	0.0000 / 68120.70
Office Moved Business high-15	0.8927	0.2141	4.09	0.0000 / 137311.36
Industrial Moved Business high-15	0.8825	0.3303	10.30	0.0000 / 31472.95
Leisure Moved Business high-15	0.8590	0.3214	23.14	0.0000 / 24499.04
Retail Moved Business low-15	0.8196	0.2480	9.22	0.0000 / 80530.08
Office Moved Business low-15	0.8160	0.1935	19.98	0.0000 / 196687.18
Industrial Moved Business low-15	0.7020	0.2243	11.75	0.0000 / 44367.93
Leisure Moved Business low-15	0.7653	0.2869	21.03	0.0000 / 39549.45
Retail New Business high-21	0.6806	0.1333	7.15	0.0000 / 68268.72
Office New Business high-21	0.7251	0.1326	5.55	0.0000 / 140466.95
Industrial New Business high-21	0.6081	0.1449	12.33	0.0000 / 35353.30
Leisure New Business high-21	0.7713	0.2548	15.06	0.0000 / 22282.19
Retail New Business low-21	0.5761	0.1128	7.76	0.0000 / 113847.97
Office New Business low-21	0.4771	0.0793	4.71	0.0000 / 132914.54
Industrial New Business low-21	0.5313	0.1207	11.81	0.0000 / 69762.32
Leisure New Business low-21	0.5426	0.1537	13.82	0.0000 / 40777.76
Retail Moved Business high-21	0.9191	0.2913	8.37	0.0000 / 69741.31

Office Moved Business high-21	0.8564	0.1861	5.53	0.0000 / 134199.14
Industrial Moved Business high-21	0.8605	0.3040	15.02	0.0000 / 31497.88
Leisure Moved Business high-21	0.8393	0.3000	15.82	0.0000 / 24728.02
Retail Moved Business low-21	0.7670	0.2109	13.49	0.0000 / 81878.77
Office Moved Business low-21	0.6374	0.1159	9.61	0.0000 / 204872.72
Industrial Moved Business low-21	0.6585	0.1991	15.92	0.0000 / 44980.92
Leisure Moved Business low-21	0.7155	0.2488	31.88	0.0000 / 40112.57

Table 16: *Industry-specific model-fit evaluation metrics*

Appendix B:

This appendix covers the details of the matching algorithm used to map businesses between the Business Census and Business Rates datasets. As explained earlier, no overlapping identifier exists between the two datasets. Hence, a combination of the business's name and the business's full address is used to merge the two datasets. Names and addresses might be represented differently in each dataset, even though they are the same business in reality.

For example, a business exists in business census data with the name of "Johns walker bro limited"[6] and the address of "40 arches, New Malden, London, KT3 3PB", while business rates data has the same business with the name of "John's Walk. Brothers ltd" and the address of "Arches of 40, New Malden, London KT33PB". Using old approaches of fuzzy matching, such as traditional NLP techniques for cleaning text of stop words and stemming, might help increase the matching rate over an exact match; however, using an LLM-based deep neural network model is more accurate and bullet-proof in matching texts.

To this end, an advanced deep neural network model named "Lajavaness/bilingual-embedding-small" is used. This model is pre-trained leveraging the robust capabilities of Multilingual-MiniLM-L12-H384 (Wang *et al.*, 2020) and fine-tuned using Siamese BERT-Networks with 'sentence-transformers' (Yi, 2023) and Augmented SBERT with Pair Sampling Strategies (Thakur *et al.*, 2021). This model is pre-trained to capture semantics in full sentences and represent their meanings as a high-dimensional vector space of 384 latent variables. This model is chosen by ranking all available open-source models in Hugging Face leaderboard based on Semantic Textual Similarity (STS) task and then selecting the model that can fit in a development machine (which is the 384 dimensions model that is of 117 million parameter size). The figure below shows the top models by STS and model size.

Rank (Bo…	Model	Zero-shot	U…	Number of P…	Embedding D…	Mean (TaskT…	Bitext …	Reranking	Retrieval	STS
4	Qwen3-Embedding-0.6B	99%		595M	1024	56.01	72.23	61.41	64.65	76.17
9	text-multilingual-embedding-002	99%		Unknown	768	54.25	70.73	61.22	59.68	76.11
5	Linq-Embed-Mistral	99%		7B	4096	54.14	70.34	64.37	58.69	74.86
27	bilingual-embedding-base	98%		278M	768	50.61	69.98	59.41	52.96	74.84
13	Cohere-embed-multilingual-v3.0	⚠ NA		Unknown	1024	53.23	70.50	64.07	59.16	74.80
8	SFR-Embedding-Mistral	96%		7B	4096	53.92	70.00	64.19	59.44	74.79
31	bge-m3-custom-fr	98%		567M	1024	50.57	72.16	60.37	53.26	74.69
33	bilingual-embedding-small	98%		117M	384	49.69	69.48	59.31	49.55	74.14
22	bge-m3	98%		568M	1024	52.18	79.11	62.79	54.60	74.12

Figure 35: Models ranking **Source:** (*MTEB Leaderboard - a Hugging Face Space by mteb*, no date)

The model is used first to embed each business address and name. Then, using the cosine similarity metric, a similarity score is calculated between businesses based on addresses and names. Finally, a threshold of 0.85 is used to filter out all noisy matches. This threshold is

[6] Examples of business names and addresses are hypothetical here for explanation purposes only, this is because the business datasets used are safeguarded and data leakage is not permitted

determined by analysing sample datasets to find the optimal value that keeps the false positive rate below 1%, thereby ensuring high accuracy in business matching. The figure below demonstrates how embedding and cosine similarity work.

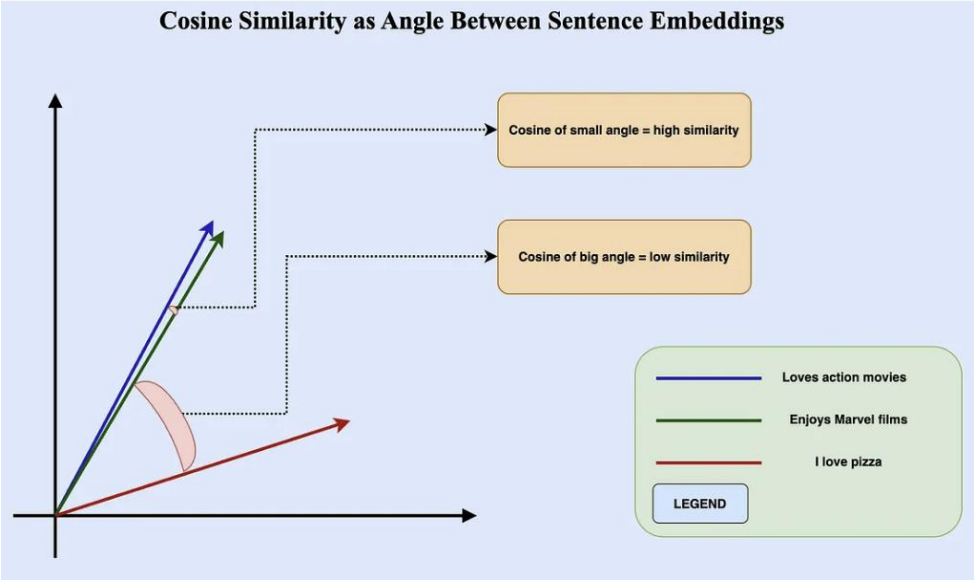
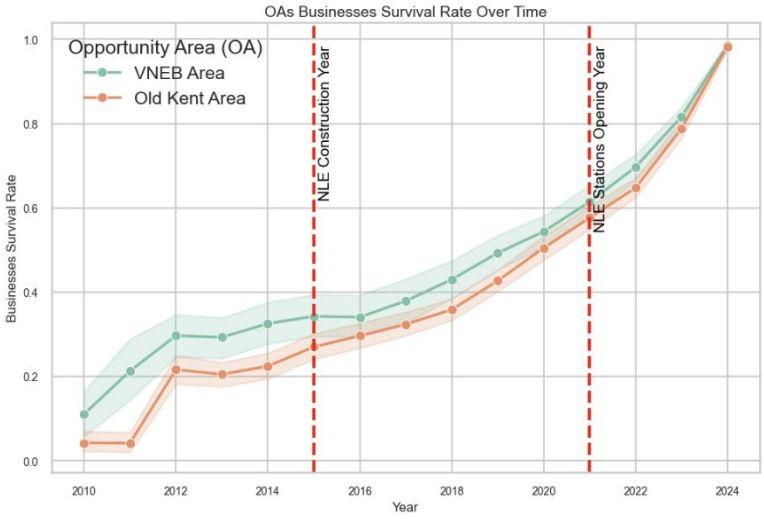


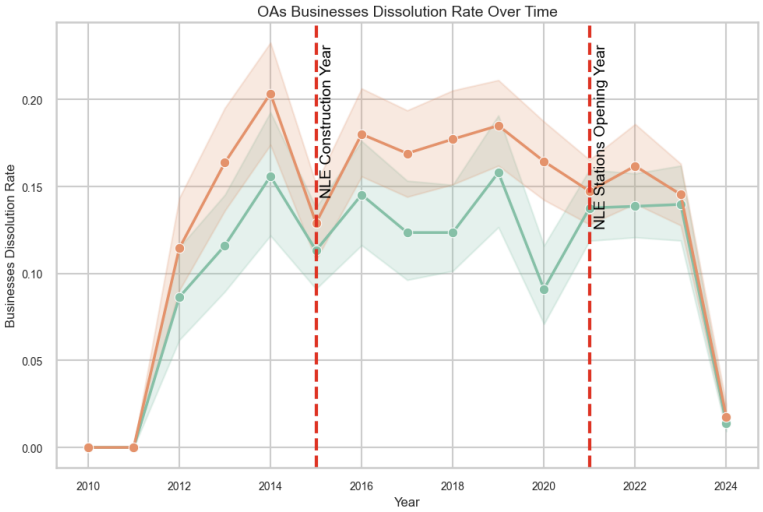
Figure 36: *Embedding and Cosine similarity method* ***Source:*** *(Sivaprakasam, 2025)*

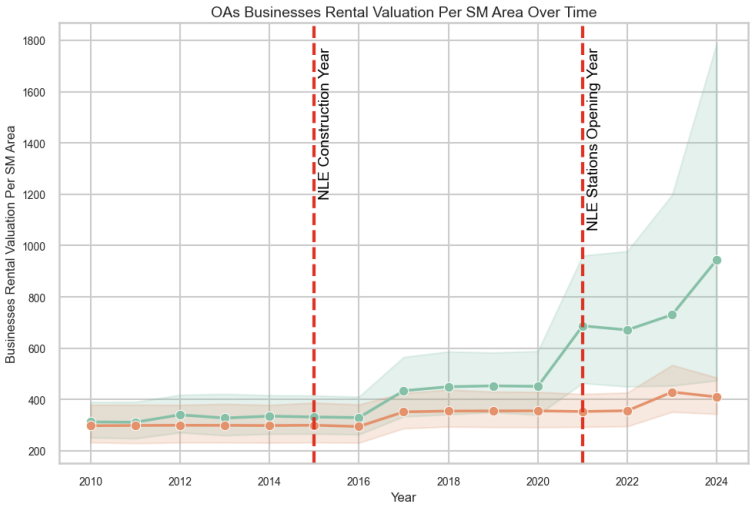
Appendix C:

This appendix contains additional explanatory analysis images, which were not added due to space constraints. It includes variable plots for VNEB vs OKR, as well as joint distribution plots of moved and new businesses against all variables.

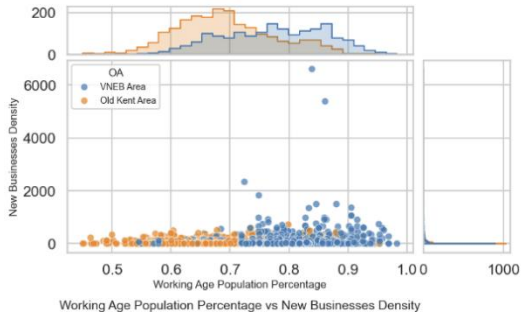
VNEB vs OKR variables:



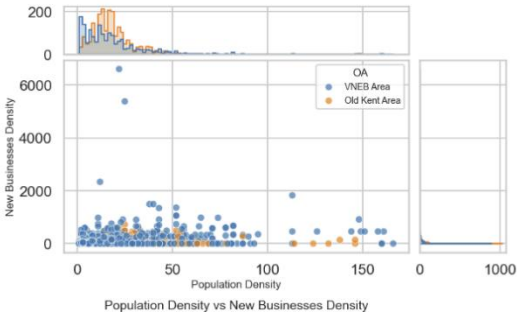




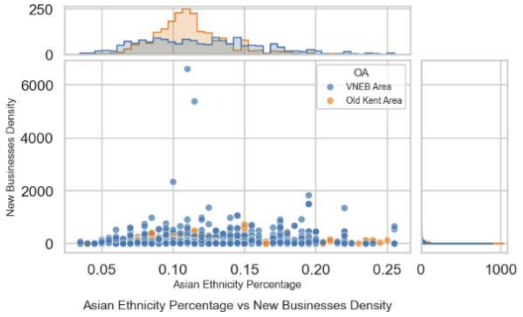
New Business against all variables joint distribution plots:



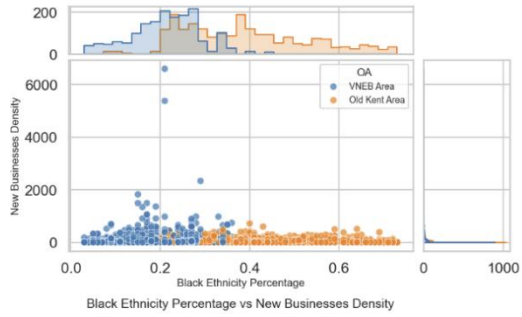
Working Age Population Percentage vs New Businesses Density



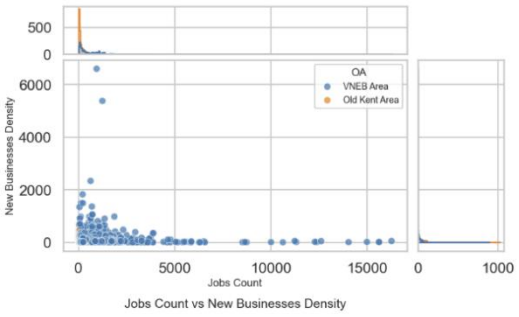
Population Density vs New Businesses Density



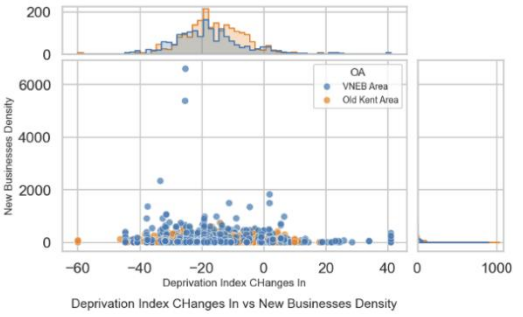
Asian Ethnicity Percentage vs New Businesses Density



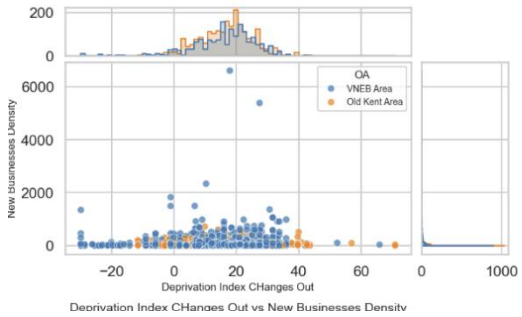
Black Ethnicity Percentage vs New Businesses Density



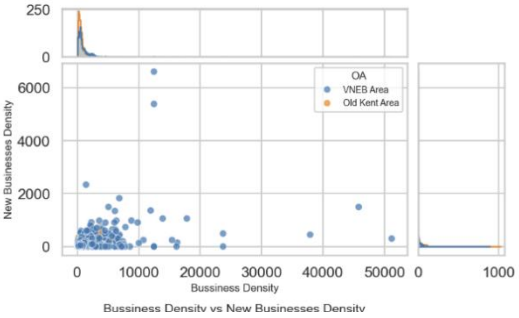
Jobs Count vs New Businesses Density



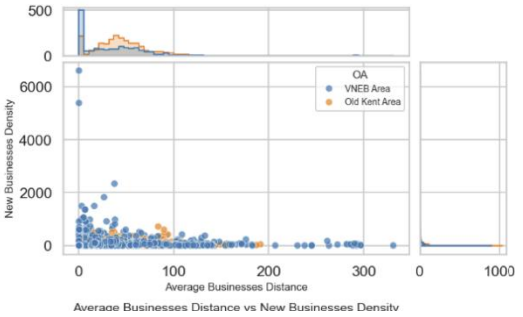
Deprivation Index CHanges In vs New Businesses Density



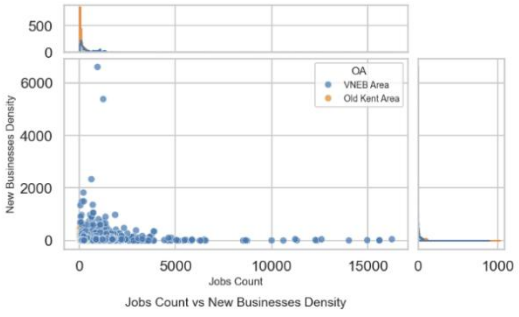
Deprivation Index CHanges Out vs New Businesses Density



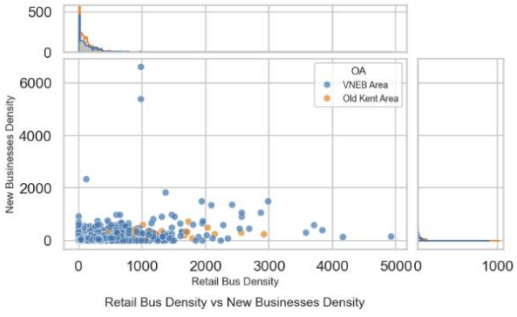
Bussiness Density vs New Businesses Density



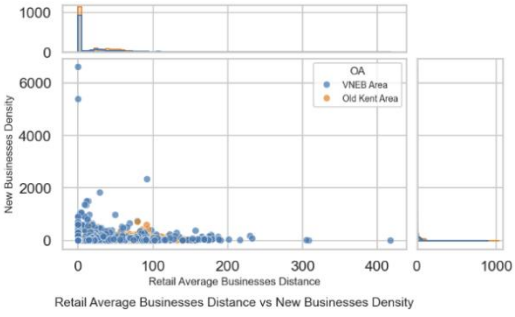
Average Businesses Distance vs New Businesses Density



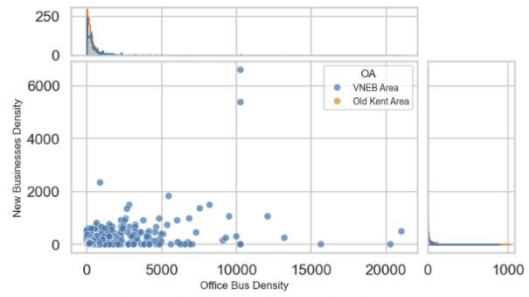
Jobs Count vs New Businesses Density



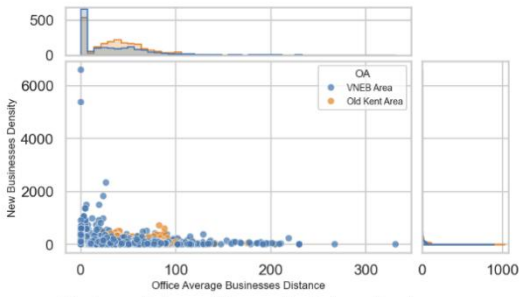
Retail Bus Density vs New Businesses Density



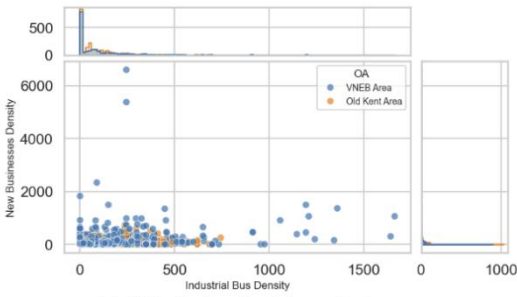
Retail Average Businesses Distance vs New Businesses Density



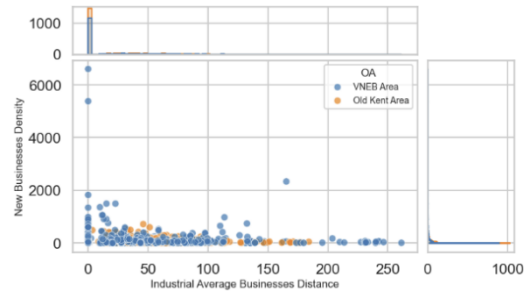
Office Bus Density vs New Businesses Density



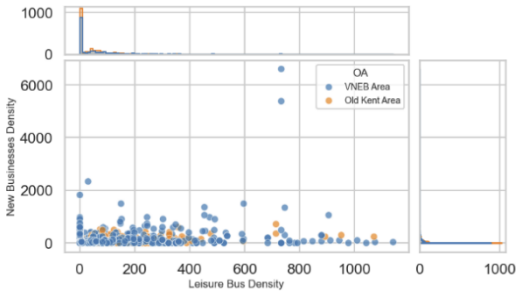
Office Average Businesses Distance vs New Businesses Density



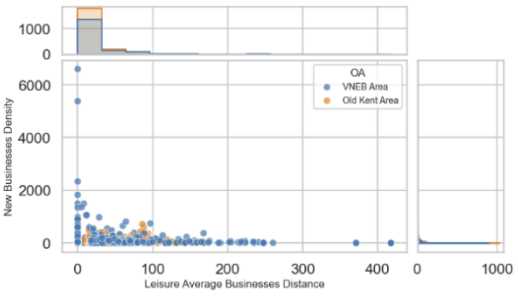
Industrial Bus Density vs New Businesses Density



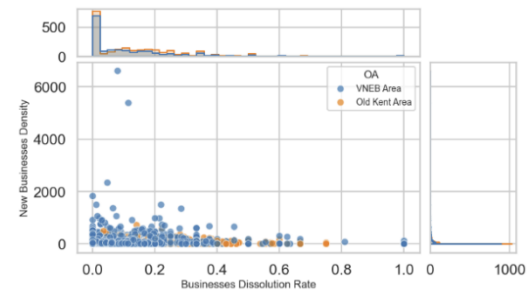
Industrial Average Businesses Distance vs New Businesses Density



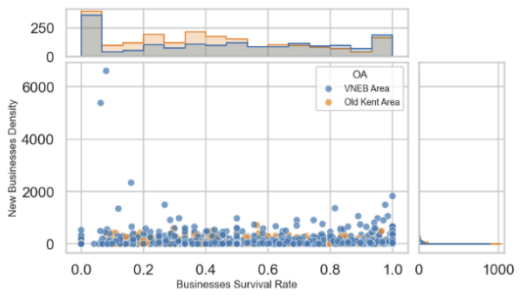
Leisure Bus Density vs New Businesses Density



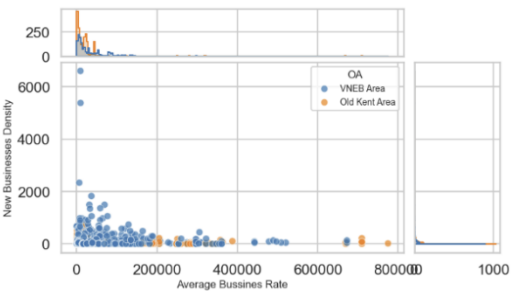
Leisure Average Businesses Distance vs New Businesses Density



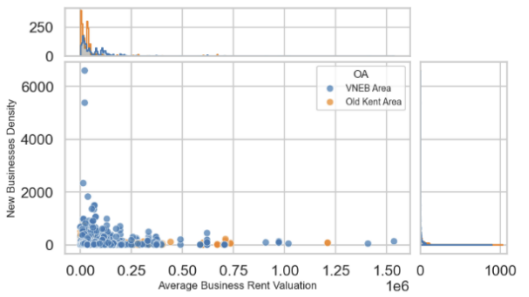
Businesses Dissolution Rate vs New Businesses Density



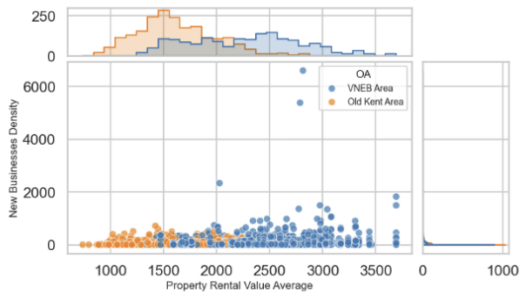
Businesses Survival Rate vs New Businesses Density



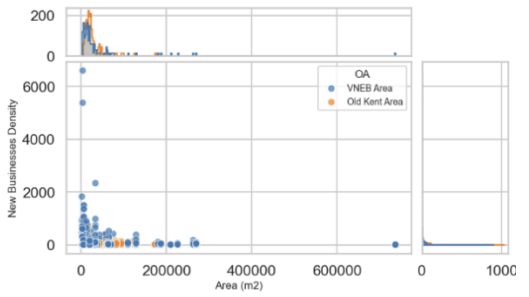
Average Bussines Rate vs New Businesses Density



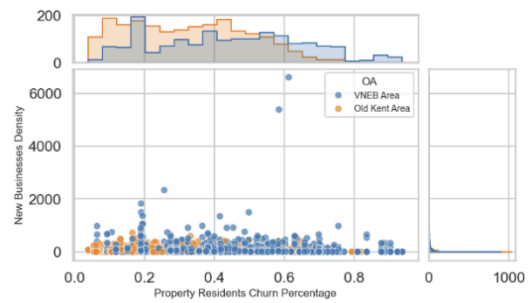
Average Business Rent Valuation vs New Businesses Density



Property Rental Value Average vs New Businesses Density

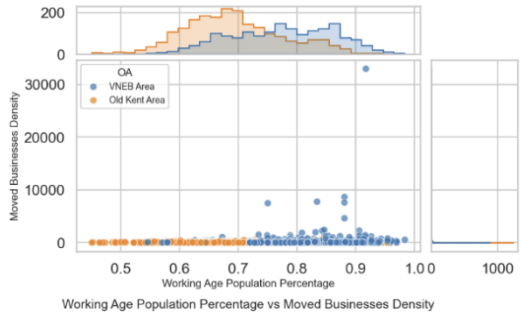


Area (m2) vs New Businesses Density



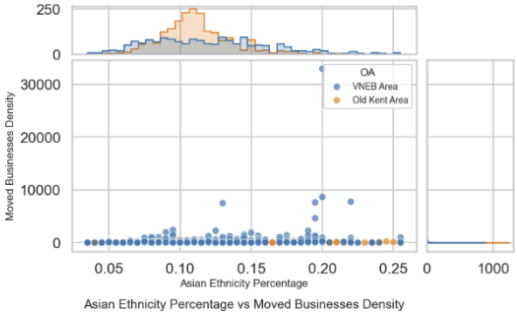
Property Residents Churn Percentage vs New Businesses Density

Moved Business against all variables joint distribution plots:

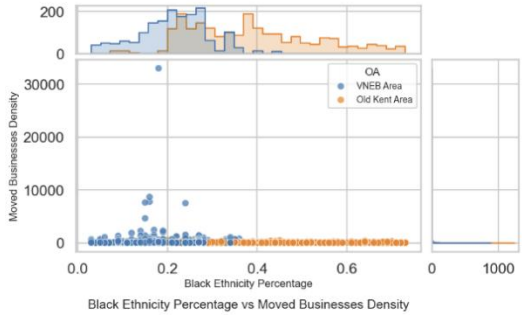


Working Age Population Percentage vs Moved Businesses Density

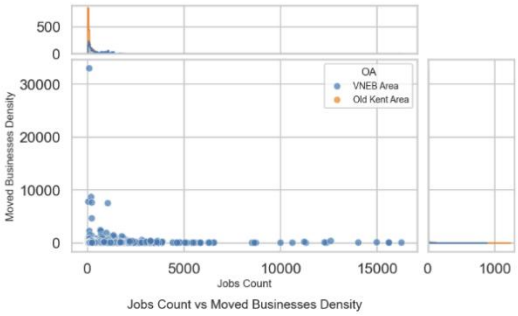
Population Density vs Moved Businesses Density



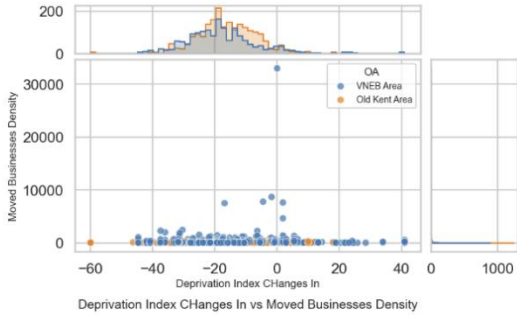
Asian Ethnicity Percentage vs Moved Businesses Density



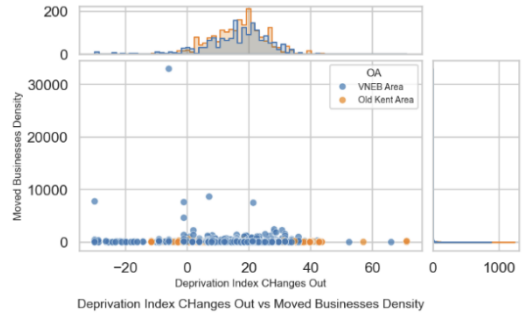
Black Ethnicity Percentage vs Moved Businesses Density



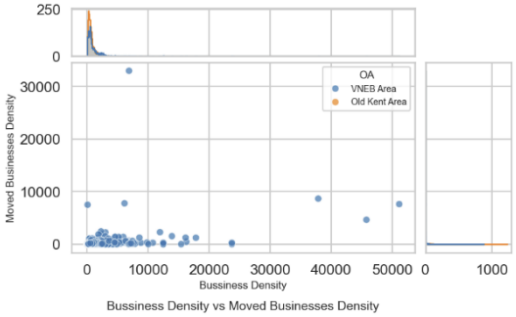
Jobs Count vs Moved Businesses Density



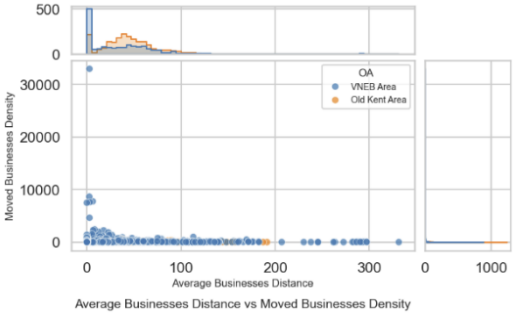
Deprivation Index CHanges In vs Moved Businesses Density



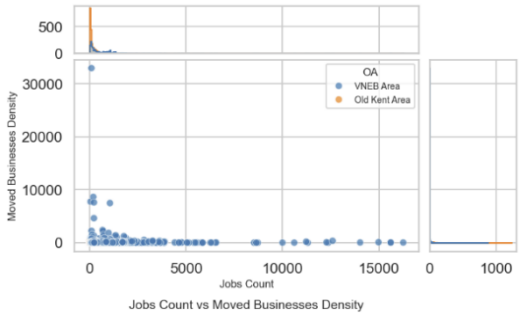
Deprivation Index CHanges Out vs Moved Businesses Density



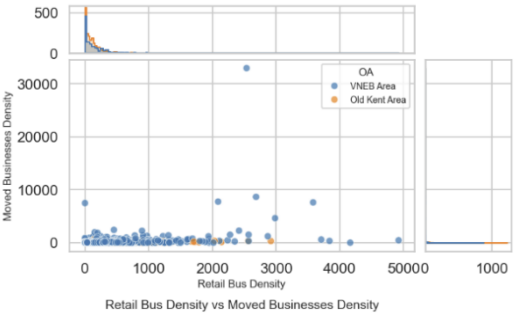
Bussiness Density vs Moved Businesses Density



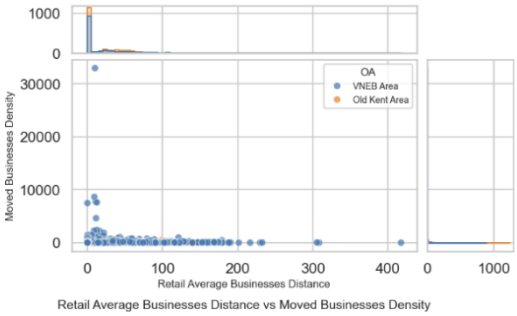
Average Businesses Distance vs Moved Businesses Density



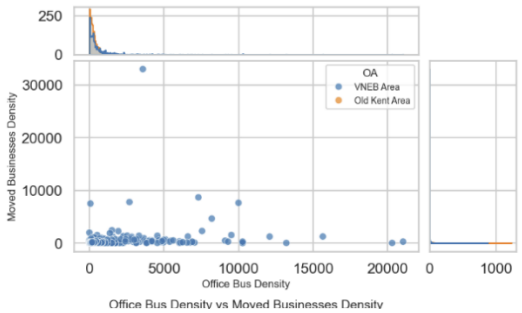
Jobs Count vs Moved Businesses Density



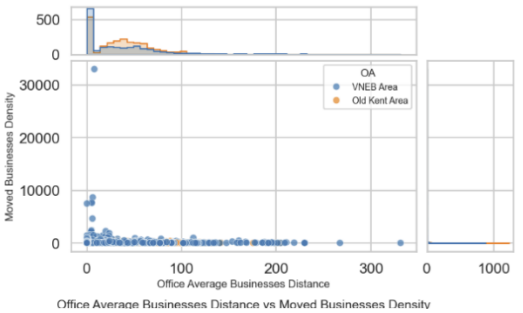
Retail Bus Density vs Moved Businesses Density



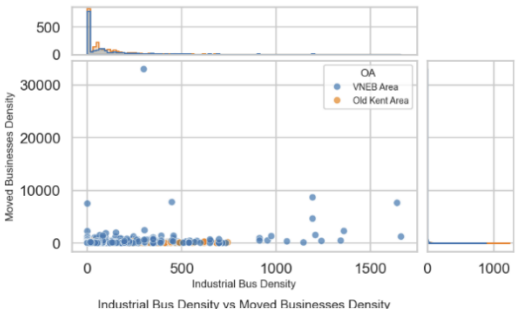
Retail Average Businesses Distance vs Moved Businesses Density



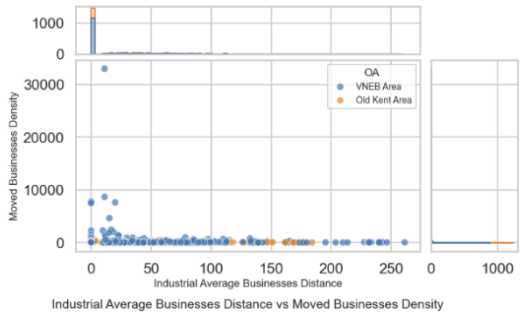
Office Bus Density vs Moved Businesses Density



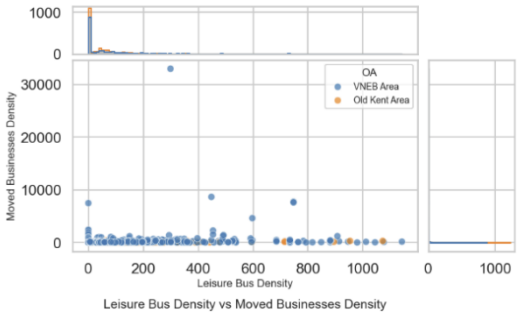
Office Average Businesses Distance vs Moved Businesses Density



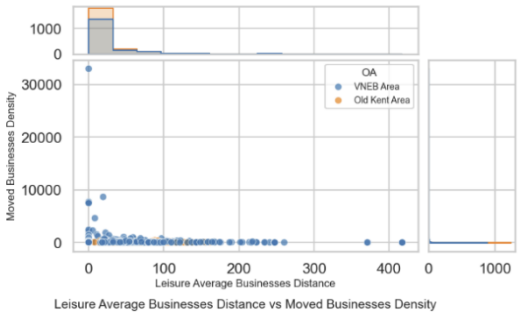
Industrial Bus Density vs Moved Businesses Density



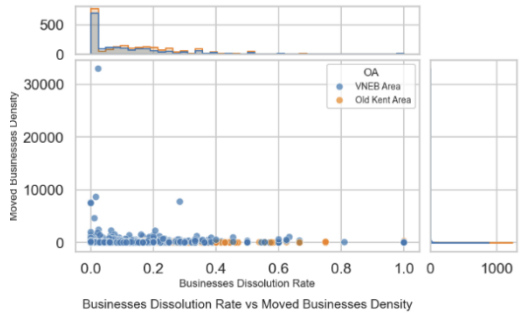
Industrial Average Businesses Distance vs Moved Businesses Density



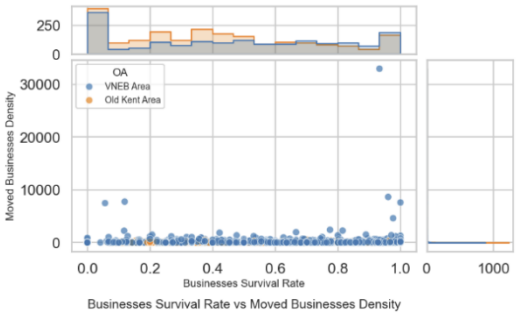
Leisure Bus Density vs Moved Businesses Density



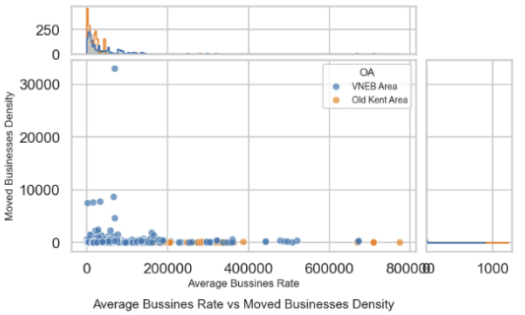
Leisure Average Businesses Distance vs Moved Businesses Density



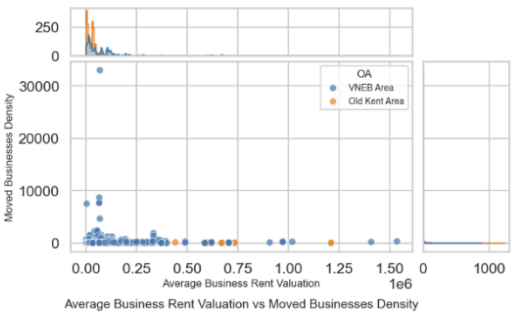
Businesses Dissolution Rate vs Moved Businesses Density



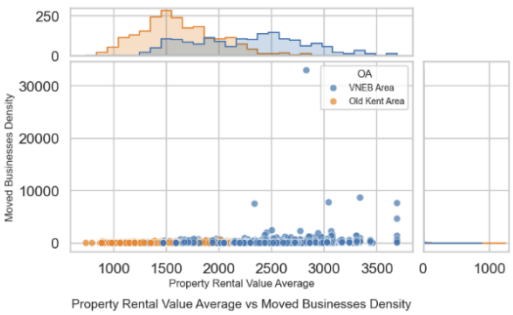
Businesses Survival Rate vs Moved Businesses Density



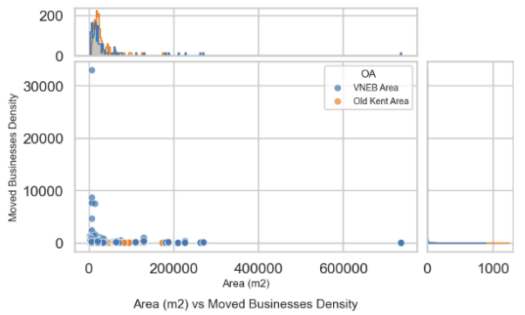
Average Bussines Rate vs Moved Businesses Density



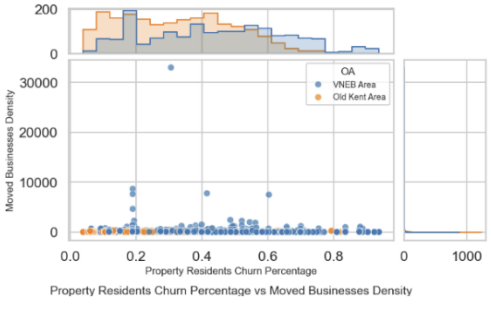
Average Business Rent Valuation vs Moved Businesses Density



Property Rental Value Average vs Moved Businesses Density



Area (m2) vs Moved Businesses Density



Property Residents Churn Percentage vs Moved Businesses Density