Tong Li's Spatial Data Analysis Projects

· **Understanding the Impact Restrictions on Mobility and Working from Home has had on Greater London Train Station Hubs as Desirable Retail Environment Locations**

A Master's thesis in collaboration with Local Data Company (LDC) and Consumer Data Research Centre (CDRC)
London
08/2022

· **Lost Children, Stolen Life**
— Child Trafficking in China

A spatial data visualisation and interactive webpage porject
London
05/2022

· **Potential Correlation Between the Number of Airbnb Listings and Growth of House Prices in London**

A spatial data analysis and visualisation porject
London
01/2022

Understanding the Impact Restrictions on Mobility and Working from Home has had on Greater London Train Station Hubs as Desirable Retail Environment Locations

A Master's thesis in collaboration with Local Data Company (LDC) and Consumer Data Research Centre (CDRC), London, 08/2022

The cdata dashboard made by Tableau can be viewed here

INTRODUCTION

Since the outbreak of COVID-19 and the accompanying restrictions on mobility have caused sustained retail losses and long-lasting changes to consumer travel and shopping habits. By addressing the new trend of using spatial big data and clustering methods for retail site-selection, this study sought to assess the performance of the retail environment around 11 major London station hubs during the various phases of the outbreak of COVID-19 using key indicators, tries to identify those areas that shown robustness, vulnerabilities and resilience in facing of external impacts. The results show that Liverpool Street station was more vulnerable during the lockdown phases and did not show sufficient resilience to face the post-pandemic era. In contrast, the Westfield-based retail environment around Stratford station hub, which had the most passengers' volume across Greater Britain during the lockdown, demonstrated robustness and resilience in the face of the impact. This study allows for the interpreting, extending and validating the results using more data, and has exhibited the potential to apply the indicators detecting retail health and the methods identifying retail environments with particular features to other retail areas.

RESEARCH OBJECTS

This study aims to investigate the impact of the decline in transportation usage on the retail environments around stations, considering operating status in different retail sectors and property vacancy rates. It seeks to use appropriate methods and criteria to develop metrics for quantifying this impact and classifying retail environments. The goal is to identify the locations of more resilient, robust, or vulnerable retail environments during and after the COVID-19 pandemic, and to understand the impact of the pandemic on business operations, retail location choices, and government development policies around station hubs.

STUDY AREAS

This study selects the top 10 station hubs in London by usage (number of exits and entries) during the outbreak of COVID-19 (from April 2020 to April 2021, ORR) and other station hubs which play essential roles in national or international transportation (e.g., King's Cross and St Pancras), and stations located in retail or commuter centre areas of London and indeed the UK (e.g., Oxford Circus).

Name	Reasons for choice	Services	Location
Kings Cross	Has national railways, connects the whole UK	Train; tube	central of London
St Pancras	International railways Eurostar	Train; tube	central of London
Liverpool street	In the heart of the City financial district, 3rd busiest before pandemic	Train; tube	central of London
Stratford	Became the busiest station during the pandemic	Overground; train; tube; DLR	eastern London
Waterloo	Biggest and busiest (before pandemic), connects southwestern England, commuter services around west and southwest London	Train; tube	southwestern London
Victoria	The 2nd busiest, commuter station, links peripheral London	Train; tube; coach	southwestern London
Paddington	Connects western London (Notting Hill and Little Venice), Bristol, Bath, western England. final stop on the Heathrow Express.	Train; tube	western London
Highbury & Islington	Became the 6th busiest station during the pandemic	Overground; train; tube	northern London
Clapham Junction	Busiest UK station for interchanges between services, became the 7th busiest station during the pandemic	Train; overground; tube (Elizabeth line, being constructed)	southwestern London
Barking	Interchange station, became the 9th busiest station during the pandemic	Overground; train; tube	eastern London
Oxford Circus	Retail, tourists and some workers in the area	Tube	central of London

Table 1: List of station hubs as research objectives

This study used an 800m radius around the station as the scope for data collection and analysis. This criterion is based on the central place theory of Applebaum and Cohen, as well as Moreno's (2016) concept of a 15-minute walking city based on pedestrian walking speed. Within this radius, the retail environments are able to reach a diverse range of consumers based on the purpose of using public transportation, including transfer passengers, commuters, tourists, and residents

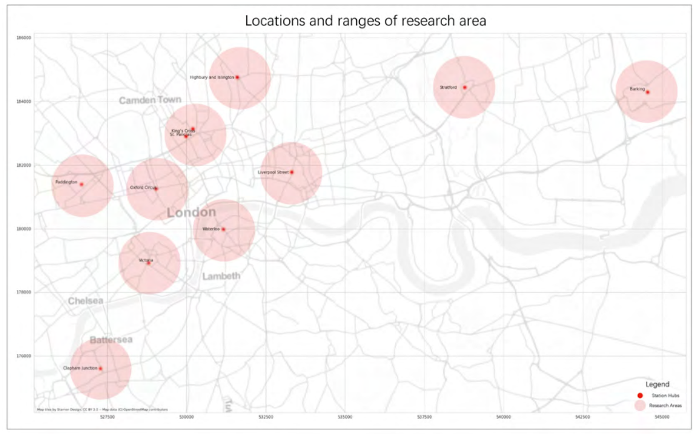


Fig 1: Locations and ranges of research area

STUDY PEROIDS

The period of the pandemic in London is divided into seven phases according to UK government coronavirus lockdown and measures:

Period	Status	Date
phase I	before the Breakout of COVID-19	October 2019 to February 2020
phase II	the First Lockdown	March 2020 to June 2020
phase III	the First Easing	July 2020 to October 2020
phase IV	the second Lockdown	November 2020 to March 2021
phase V	the Second Easing	April 2021 to July 2021
phase VI	the Recovering	August 2021-December 2021
phase VII	the Post COVID-19	January 2022 to May 2022

Table 2: List of study peroids

DATA

POI data for records of retailing facilities within an 800m radius centred on each of these stations since 1999. POI data includes fields which can extract helpful indicators for evaluating retail environments.Table 3.3 shows raw information of POI data and gives descriptions for each field.

Field Name	Description	Example	Non-Null Count	Unique Value Count
Source Premise Id	Station ID	52887830	14754	11
Station	Station name	Kings Cross	14754	11
Target Premise Id	Property ID	50988543	14754	8886
Target Occupier ID	Retail ID	12453184	14754	13644
Name	Retail name	Runners Need	14754	8794
Multiple Name	Name of retailing in chains	Runners Need	4795	1497
Street No	Location and address	200	11684	1321
Street		Pentonville Road	14754	679
Post Code		N1 9JP	14754	3547
Town		London	14754	3
Region		Greater London	14754	1
Longitude		-0.11806	14754	4993
Latitude		51.53116	14754	3916
Business Type	Retailing in chains (= Multiple) or not (=Independency) or vacant properties(=Dummy)	Multiple	14754	3
Sub Category	Third level classification	Sports Goods Shops	14754	326
Category	Second level classification	Sports, Toys, Cycle Shops & Hobbies	14754	42
Classification	Main level classification	Comparison	14754	6
Care Of	Retailer of this retail	Cycle Surgery	462	244
Created Date	date the retail began to operate	2010-08-26	14754	2315
Closed Date	date the retail closed	2020-03-05	5446	790
Is New	Is this property being let for the first time as a retail business	TRUE	14754	2
Retail Mix	the quality (based on price point, customer profile and location) of offer for comparison (general retailing) retailers	Mass	1671	3

Table 3: Description of POI data

DATA DESCRIPTION

■ Impacts on different retail sectors

Generally speaking, the number of stores in all retail sectors was growing before the pandemic, but during the first lockdown and easing period, the overall number of stores that opened greatly decreased, although the decrease trend slowed down afterwards, but overall, still more stores were closed than newly opened. This means that as of May 2022, the impact of the pandemic and lockdown measures on the retail industry around the station continues.

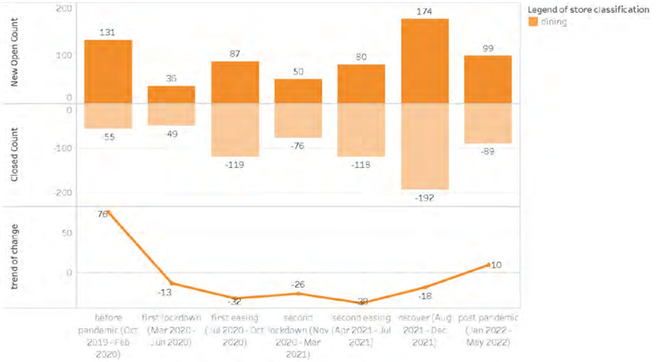


Fig 2 (a) : Change of store number in dining sector

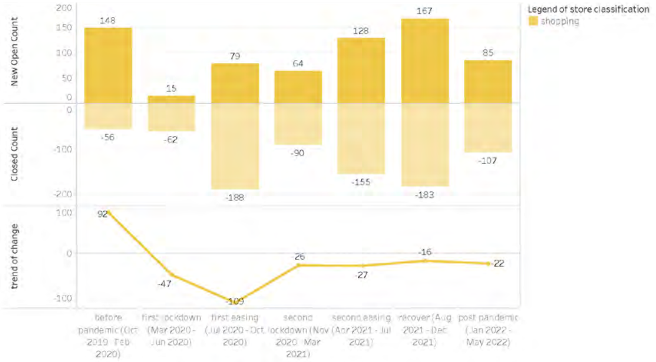


Fig 2 (b) : Change of store number in shopping sector

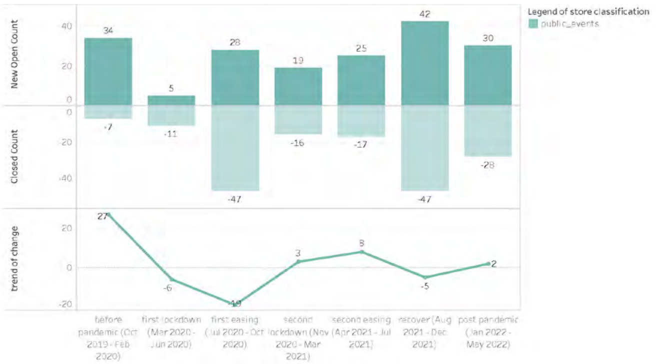


Fig 2 (c) : Change of store number in public events sector

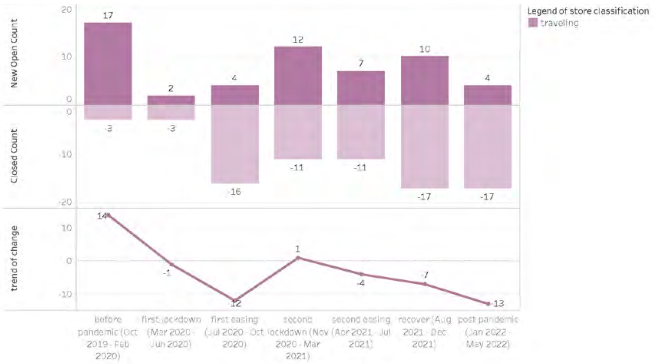


Fig 2 (d) : Change of store number in travelling sector

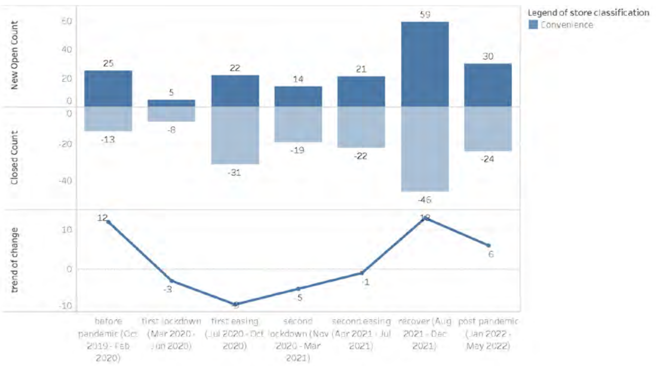


Fig 2 (e) : Change of store number in convenience sector

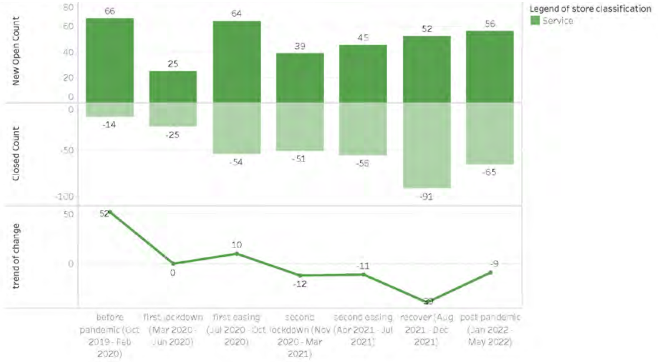


Fig 2 (f) : Change of store number in service sector

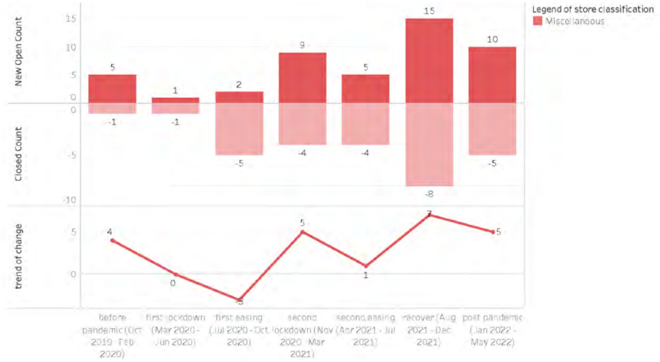


Fig 2 (g) : Change of store number in miscellaneous sector

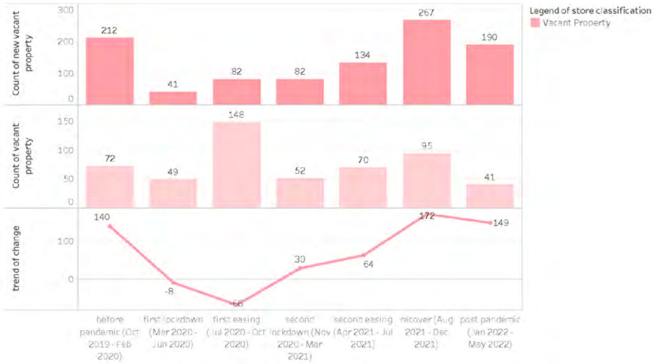


Fig 2 (h) : Change of vacant properties' number

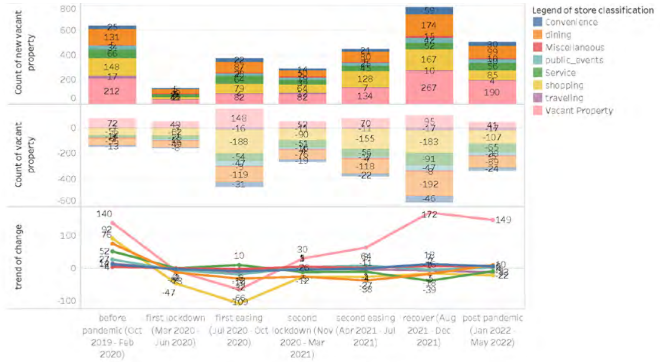


Fig 2 (i) : Change of store number in all sectors

Fig 2 (a - i) : Change of store number by different retail sectors

The two sectors with the largest proportions (shopping and dining) showed a general trend of decreasing total number of stores throughout the five phases of the study, with shopping being the sector most affected by the pandemic, showing a significant decrease in the total number of stores during the first easing period and then stabilizing but still decreasing overall. The dining sector, on the other hand, showed less significant changes, even showing a growth trend in the total number of stores during the post pandemic phase.

According to Fig. 3 and fig. 4, it can be seen that the change in the number of stores in shopping sectors happened mainly around oxford circus, followed by Stratford; the change in the number of dining facilities is most evident around Liverpool street station. Although the convenience sector saw a decrease in total number during the lockdown and recovery phases, the number of stores began to rise in subsequent stages. The overall number of service stores increased during the first easing phase and continued to grow in subsequent stages, but still showed a trend of decreasing overall.

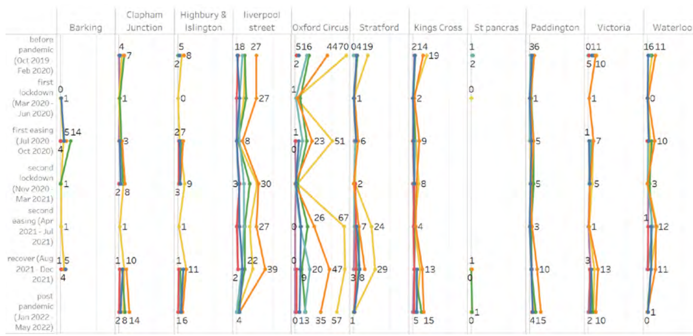


Fig 3: Trends of number of new-opened stores by station hubs

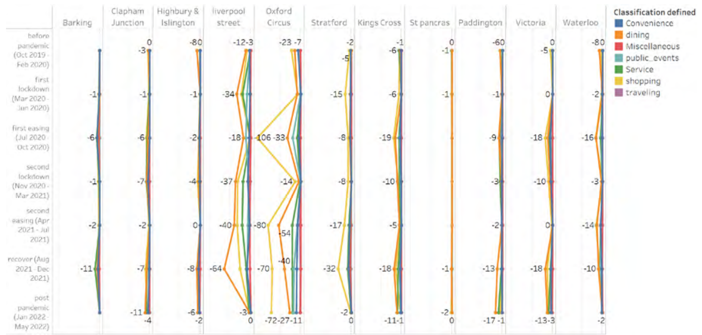


Fig 4: Trends of number of closed stores by station hubs

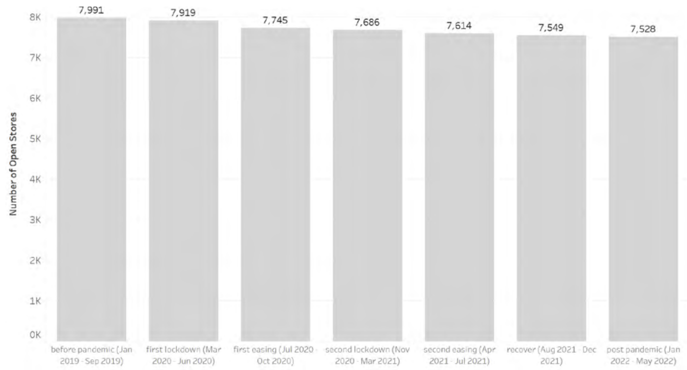


Fig 5: Trends of number of opened stores in each phase

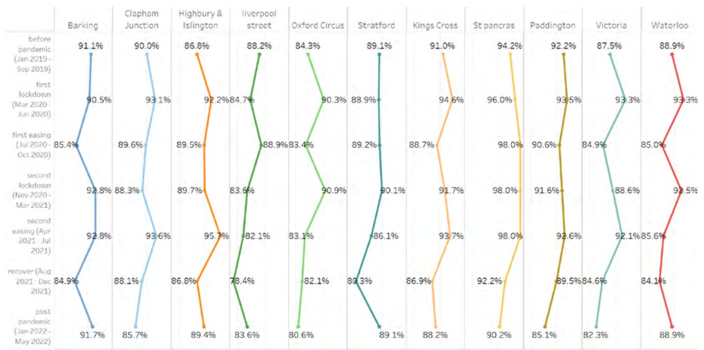


Fig 6: Trends of open rate in each station hubs

■ Growth potential in each station hub

□ Change of total number of stores

According to Fig. 5, since the start of the pandemic, the total number of opened retail facilities has been on a downward trend.

□ Opened stores and open rate

With the exception of Liverpool Street Station, whose open rate has clearly decreased, the open rate of other stations remained stable or increased during the first lockdown phase, but the situation was the opposite during the first easing phase. Lowest Open rate in most stations happened during recovery and post pandemic phases. Aside from Highbury Islington, Barking, and Stratford, all stations have failed to recover their open rate to pre-pandemic levels.

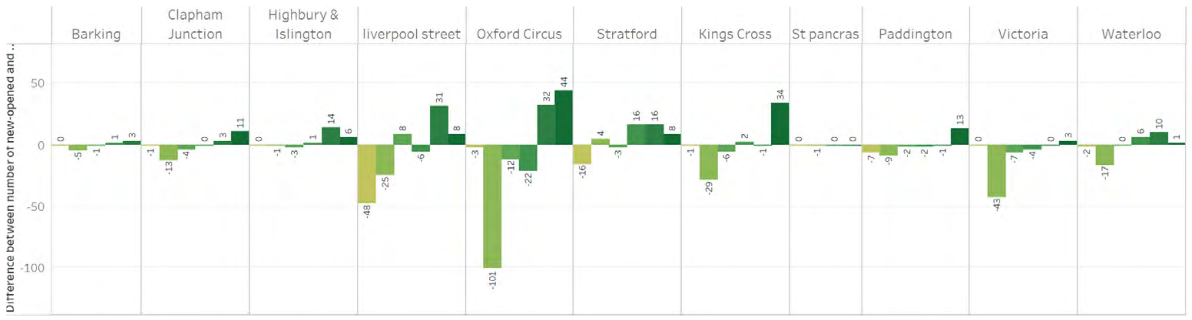


Fig 7: Actual number of added stores during each study phase by stations

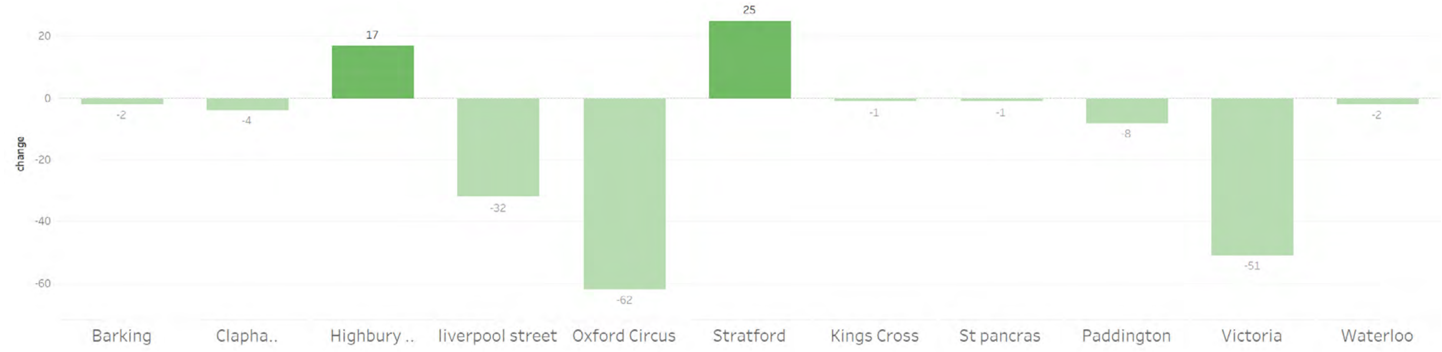


Fig 8: Actual number of stores added during the total study phases

□ Actual growth of retail facilities

The change trend of the number of retail facilities operating within a region can reflect its retail industry's growth potential. Fig. 7 reflects the actual growth trend of the number of stores operating around each station at different phases (excluding vacant stores, the difference between new-opened and closed stores). It can be seen that most stations experienced a significant decrease in the number of opened stores during the pandemic, especially during the first easing phase (except for Liverpool Street Station). According to Fig. 8, the first station to show a turnaround trend was Stratford, which became the station with the largest increase in the number of opened stores, while other stations decreased after the pandemic.

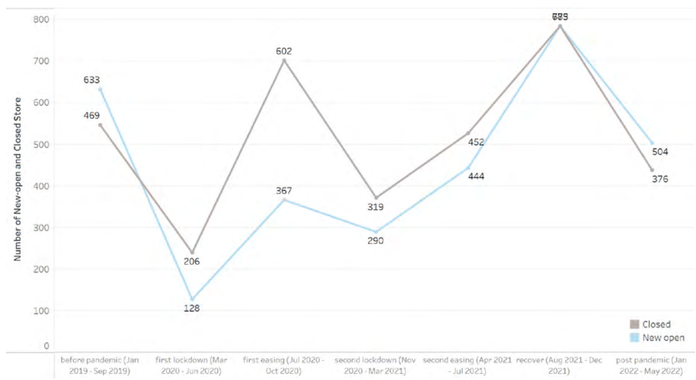


Fig 11: Trends of number of closed and newly opened retail facilities in total

□ Closed and newly opened retail facilities

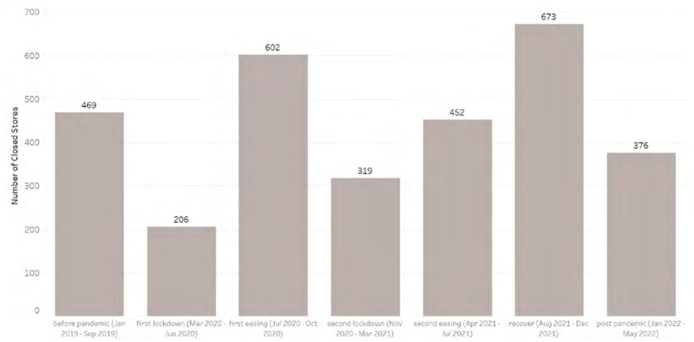


Fig 9: Trends of number of closed retail facilities in total

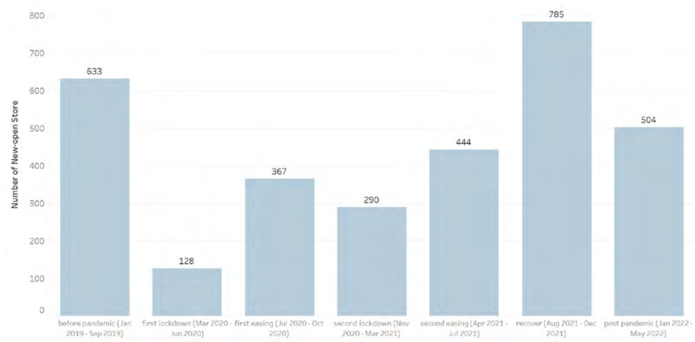


Fig 10: Trends of number of newly opened retail facilities in total

The number of closed and newly opened facilities show the same trend, which is obviously due to the fact that new stores often take over after the previous ones close, without any new property marketing. In a good retail market, the number of new stores should be equal to or greater than the number of closed stores. Fig. 10 shows the significant gap between the number of newly opened retail facilities and the number of closed facilities during the two lockdowns and easing phases, reflecting the impact of the pandemic and lockdown measures on the retail environment around the stations.

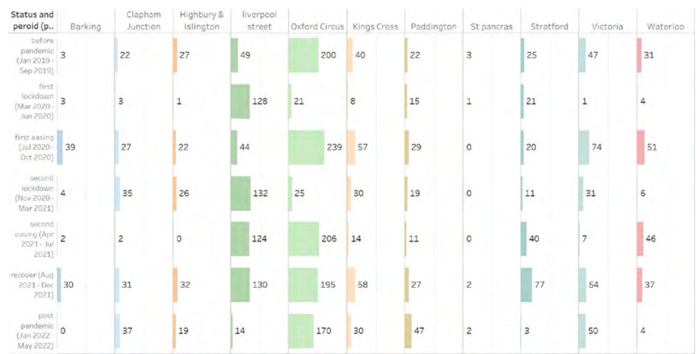


Fig 12: Trends of number of closed retail facilities in each station hubs

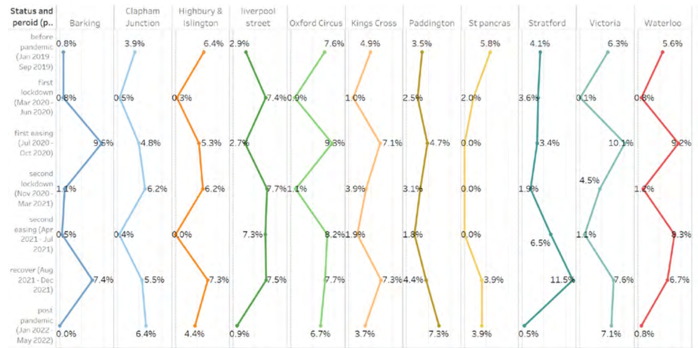


Fig 13: Trends of closure rate in each station hubs

During the first lockdown phase, the only station with a rising closure rate was Liverpool Street Station, meaning that the stores nearby had the most direct reaction to the pandemic and the lockdown measures. Other stations saw an increase in closure rates during the first easing phase. Throughout the pandemic, the number of closed stores around all stations was greater than the number of new stores until the recovery phase. The largest wave of store closures occurred during the first easing phase.

Fig 14: Trends of newly opened retail facilities in each station hubs

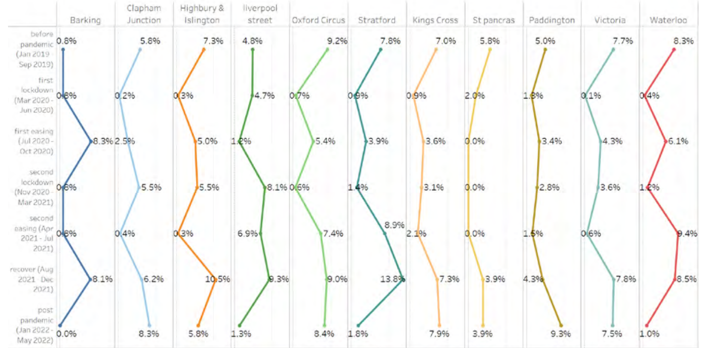


Fig 15: Trends of newly opened rate in each station hubs

Throughout the entire study period, new stores were mainly concentrated at Liverpool Street Station and Oxford Circus. During the two lockdown phases, the number of new stores at stations other than Liverpool Street Station increased but showed signs of stagnation. During the recovery phase, the number of new stores at all stations returned to or exceeded pre-pandemic levels. During the first lockdown phase, the new open rate at all stations decreased. In the first easing phase, the new open rate at stations except Liverpool Street Station and St Pancras Station decreased, while the rest of the stations showed an increase. Another peak of new open rate growth was in the recovery phase, and the rate during this phase was generally higher than before the pandemic. Stations with significant growth of retail facilities were Stratford, Liverpool Street and Barking.

■ Vacancy rate

The vacancy rate is an important indicator of the economic performance of retail facilities in an area. This mainly includes the long-term vacancy rate (number), short-term vacancy rate (number), and total vacancy rate (number) of the research stations in each phase.

□ Long-term vacancy rate (number)

This indicator refers to the number and proportion of retail facilities that have not yet been occupied by merchants as of the current research phase, within the scope of the stations under study. Stations with high levels of long-term vacancy are facing decline and are more vulnerable to external factors such as pandemics and lockdowns, leading to persistent and even increasing levels of long-term vacancy.

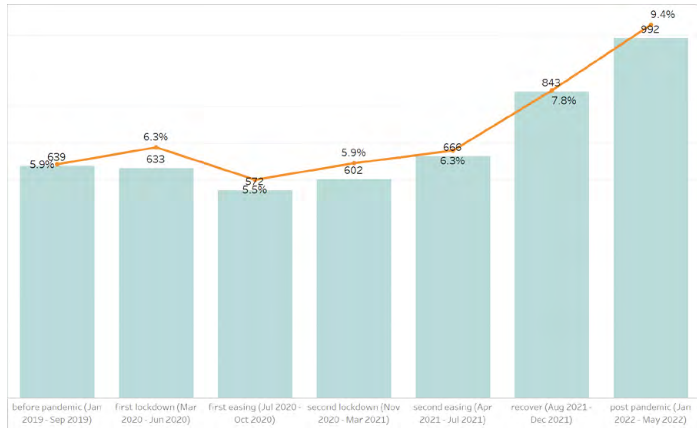


Fig 16: Trends of average long-term vacancy rate and number of vacant properties in total

According to Fig. 16, overall, the long-term vacancy rate shows an upward trend during most phases, except for a slight decline during the first easing phase. The number of properties with long-term vacancy decreased from pre-pandemic to first easing phase, and then increased, showing significant growth in the last two research phases.

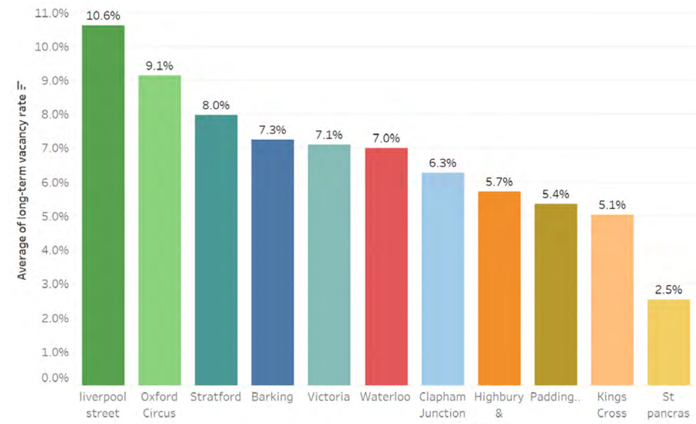


Fig 17: Average of long-term vacancy rate in each station

As shown in Fig. 17, the average long-term vacancy rate is highest at Liverpool Street Station during the research period, while the actual average rate is lowest at King's Cross Station due to the limited number of retail facilities at St Pancras Station.

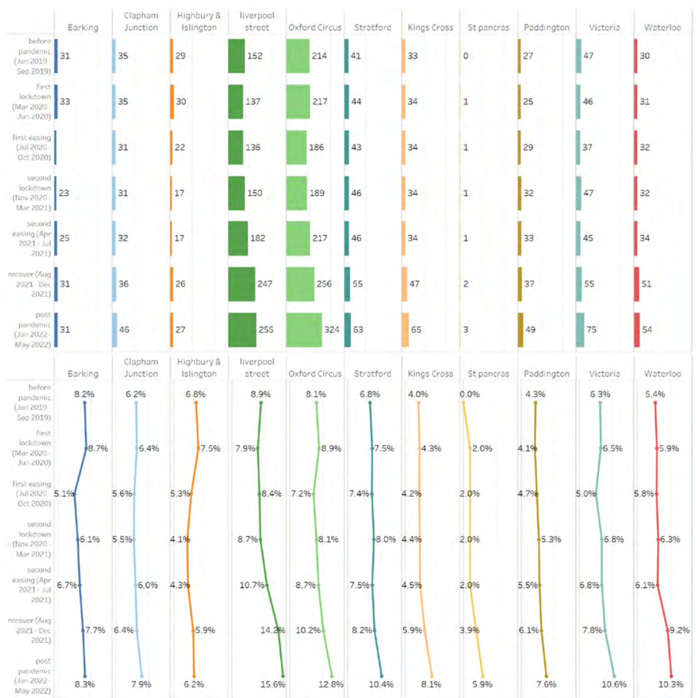


Fig 18: Trends of long-term vacant properties in each station hubs

According to Fig. 18, the long-term vacancy rate for most stations during the post-pandemic phase is significantly higher than pre-pandemic levels, by 1.5 to 2 times, with an increasing trend throughout the whole research period. The exception is at Barking and Highbury & Islington, where the long-term vacancy rate shows little change from pre-pandemic to post-pandemic, while Clapham Junctionhas the most stable change in long-term vacancy rate.

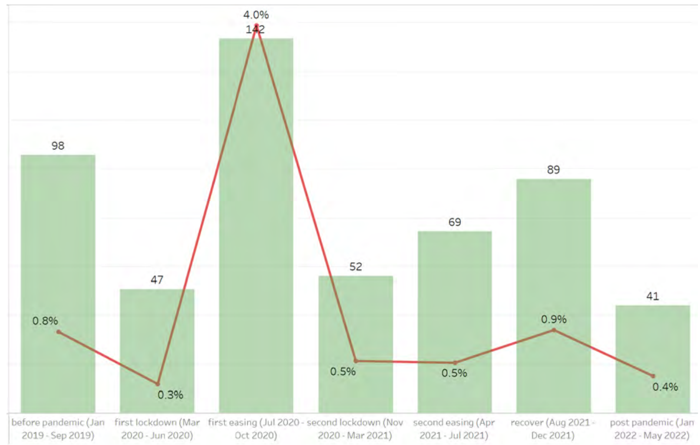


Fig 19: Trends of average short-term vacancy rate and number of vacant properties in total

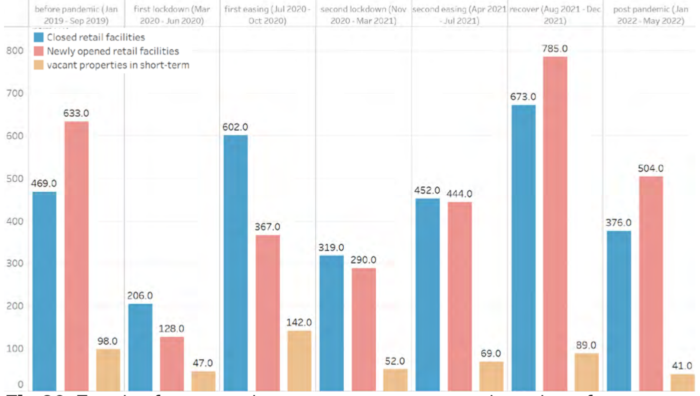


Fig 20: Trends of average short-term vacancy rate and number of vacant properties in total

□ Short-term vacancy rate (number)

This indicator means recently closed and no new retailers have taken over or as a new property, there are no businesses yet to become vacant stores. Short-term vacancies mean a high turnover of retailers, and higher levels of short-term vacancy are typically found in areas with more active retail businesses. Fig 19 shows that the short-term vacancy rate rises significantly during the first easing phase, then falls back quickly to a lower level, and then rises again slightly during the recovery phase. Combining Fig. 11 and Fig. 20, it can be seen that noticeably more shops became vacant after the first easing phase than during the recovery phase, suggesting that the frequency of retail facilities turnover declined during the first easing phase, but regained during the recovery and post pandemic phases.

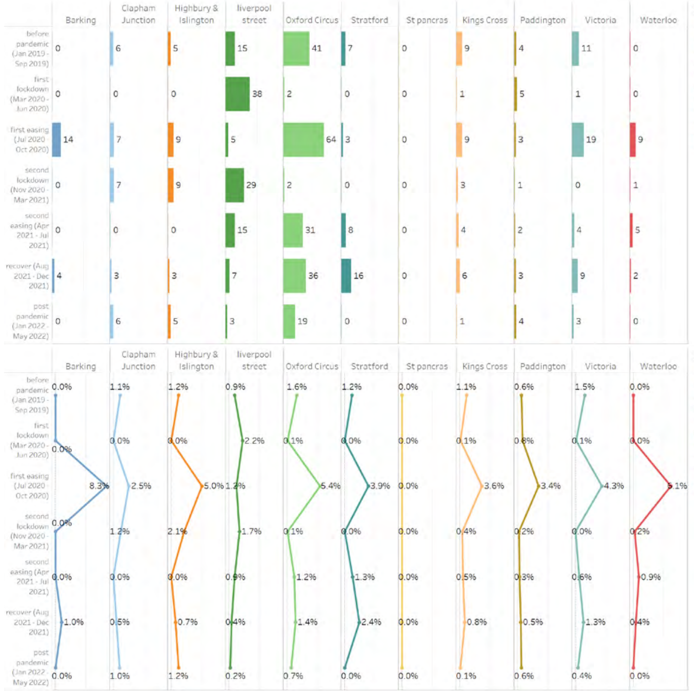


Fig 21: Trends of short-term vacant properties in each station hubs

Fig.21 shows that the change in the number of short-term vacant retail facilities occurs mainly at Liverpool Street station and Oxford Circus station. Most stations have a peak in short-term vacant retail facilities during the first easing phase, but similar to the closure rate (number), Liverpool Street station has a peak in short-term vacant retail facilities during the first easing phase. Short-term vacancy rates at all stations were either flat or lower than pre-pandemic levels in the post-pandemic phase.

□ Overall vacancy rate (number)

The overall vacancy rate (number) refers to the proportion (number) of all vacant retail facilities (including short- and long-term vacancies) at the current phase. According to Fig. 22, the overall vacancy rate (number) shows an overall upward trend, with two peaks of increase appearing during the first easing and recovery phases.

Fig 22: Trends of average overall vacancy rate and number of vacant properties in total

Fig. 23 shows the trend of the number and proportion of vacant properties at each station. The number and overall vacancy rate of vacant properties at Liverpool Street station and Oxford Circus station are the highest among all stations. Both indicators for all stations generally show a growing trend throughout the whole research period.

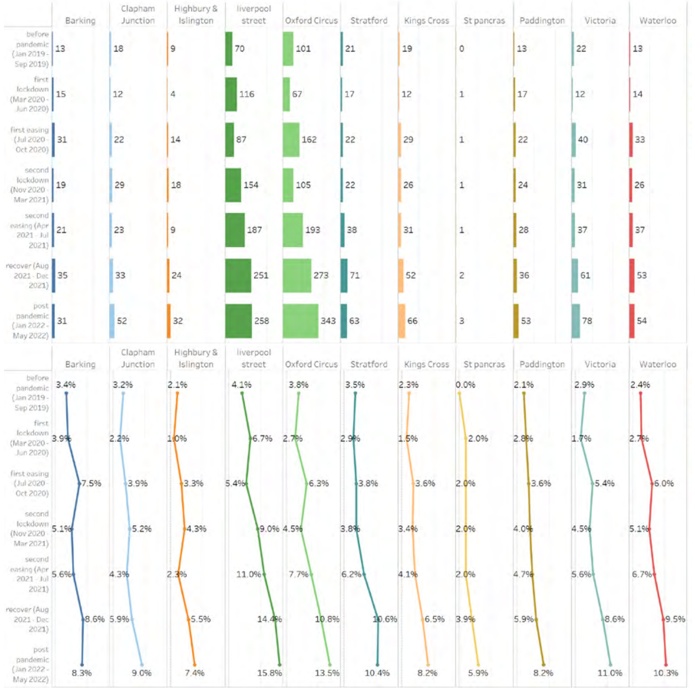


Fig 23: Trends of overall vacant properties in each station hub

Fig. 24 shows the average overall vacancy rate of the study subjects, as well as their respective overall vacancy rates and the difference with the Greater London vacancy rate at the same phase. The average overall vacancy rate is always lower than Greater London's benchmark, but the gap is decreasing, indicating that the retail environment around the stations is becoming less ideal than in the past. Specifically for each station, Liverpool Street station's overall vacancy rate surpassed Greater London's level during the second easing phase and continued to rise. Then during the recovery phase, Oxford Circus station's overall vacancy rate also surpassed Greater London's level. In the post-pandemic phase, the overall vacancy rates of Liverpool Street, Oxford Circus and Victoria stations all surpassed Greater London's level.

■ Summary

1. Retail sectors

The number of retail facilities in all sectors during the research period is declining, with the shopping sector being the most affected by the pandemic, most notably during the first easing phase. The closed

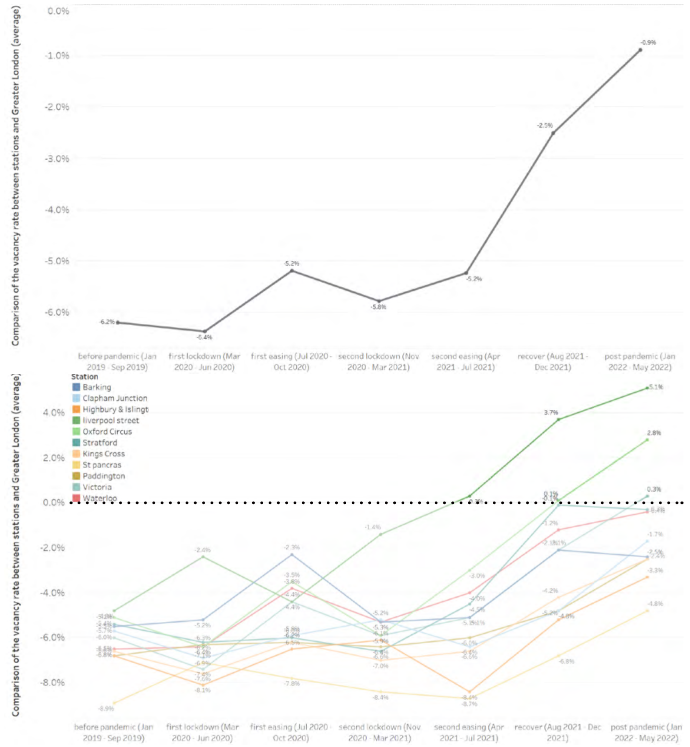


Fig 24: Trends of average overall vacancy rate and number of vacant properties in total

stores are mainly located around Oxford Circus and Stratford stations.

2. Growth potential

Most stations experienced a significant decline in the number of operating stores during the pandemic; Liverpool Street Station was the first to react to the lockdown measures, with its closure rate starting to increase during the first lockdown phase. Except for Highbury Islington, Barking, and Stratford, rest of stations have failed to recover to their pre-pandemic open rate. Since the recovery phase began, the number of newly opened retail facilities at all stations has begun to recover or exceed pre-pandemic levels. The stations with the most noticeable increase in newly opened rate are Stratford, Liverpool, and Barking.

3. Vacancy

Aside from the first easing phase, the long-term vacancy rate of all stations showed an upward trend. The station with the highest average long-term vacancy rate was Liverpool Street Station, and the lowest was Kings Cross Station. the trend of short-term vacancy rate showed a decrease in store turnover rate at each station during the first easing phase, but there was some recovery during the recovery and post-pandemic phases, and the short-term vacancy rate remained flat or lower than pre-pandemic levels after the pandemic.

RESEARCH METHODOLOGY

■ Preperation

Using appropriate spatial division methods to establish a connection between data and concepts of the retail environment. After geo-projection, the cleaned POI records are presented as dense points on the map. Considering that the subject of this study is the desirable retail environment, it is necessary to use a suitable method to link point records to the geospatial space. The grid was finally divided into units of 50m x 50m. the spatial gathering of retail stores can be clearly observed at this scale of units.

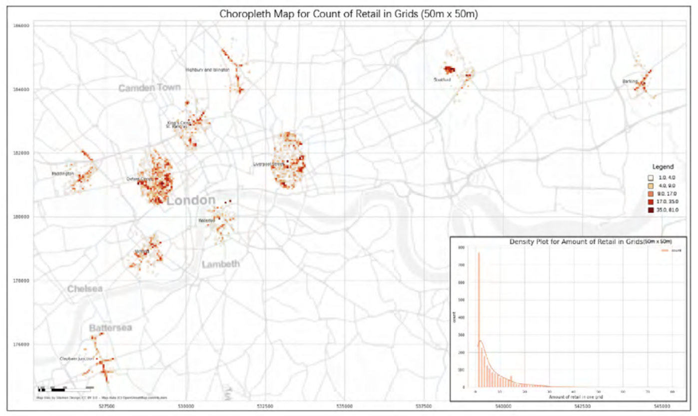


Fig 25: Choropleth Map for Count of Retail in Grids (50m x 50m)

■ Build feature variables

1. Retail health

· The number of new open or operating, or closed retail facilities in a research unit and their respective proportions of total facilities;

· reflecting the level of activity and resilience of retail environment;

· The number of closed, open and new retail facilities in each phase is calculated as a proportion of the total (including closed, open and long-term vacant) in that phase for each unit.

Items in POI data	Create date	Closed date	Classification	Description
Closed retail facilities	The time in the field is within the current interval of the phase		Exclude Vacant properties	Been closed during the current phase
Operat-ing retail facilities	The time in the field is before or within the current interval of the phase	The time in the field is before the current interval of the phase or has NaN value	Exclude Va-cant proper-ties	Been operating during the current phase
Recent opened retail facilities	The time in the field is within the current interval of the phase		Exclude Va-cant proper-ties	Began to operate during the current phase

Table 4: Description of indicators for retail health

2. Retail occupancy

· Can be measured by both the long-term and the recent vacancy rates (also known as short-term vacancy rates) during a specific phase;

· Short-term vacancies, which reflect a high turnover of retailers, can indicate a high level of activity in the retail environment;

· long-term vacancies can signal structural problems and indicate that retail businesses in the area are facing a decline and may be more vulnerable to external influences

Items in POI data	Closed date	Classification	Description
Recent vacant retail facilities	The time in the field is within the current interval of the phase	Vacant prop-erties	Recently closed and no new retailers have taken over or as a new prop-erty, there are no businesses yet to become vacant stores.
Long-term vacant retail facilities	The time in the field is before the current interval of the phase	Vacant prop-erties	became vacant before the current phase.

Table 5: Description of indicators for retail occupancy

3. Retail mix

· Use Shannon information entropy to measure the variety of retail formats in each unit;

· The higher the Shannon entropy, the more diverse the retail formats. Therefore, units with higher Shannon entropy have a stronger ability to cope with external impacts;

· The entropy of retail mix is calculated based on the categories in the POI data. Assuming one unit has A rows of POI data with N (N=42) types of retail, A1+ A2 + ... +An = Σi Ai (i = 1, 2 ... N, An represents the number of retail that type is n), then the probability can be defined as:

$$Pi = \frac{Ai}{A} = \frac{Ai}{\sum_{i}^{N} Ai}$$

Since ΣiPi = 1, the entropy of retail mix can be defined as:

$$H = -\sum_{i=1}^{N} Pi \times LnPi$$

H (H ≥ 0) is Shannon's entropy diversity index. The higher the H, the more various the retail mix of the unit is.

■ Labelling retail units

· The retail units in the different phases of the pandemic can be broadly classified as **robust, resilient, vulnerable** to external impacts. the clustering algorithm in unsupervised machine learning, which can extract patterns and find outliers from the dataset based on similarities and differences between features, can be used to group each retail unit;

· Long-term vacancies can signal structural problems and indicate that retail businesses in the area are facing a decline and may be more vulnerable to external influences.

■ Implementation

A total of nine variables were first used to reduce the dimensionality of the three indicators of retail health, retail occupancy and retail mix for the six phases since First Lockdown, and the components that explained the greatest variance in the data were clustered using K-means. The clustering results were interpreted by correlating the results with the variables. The clustering results of each phase were then used as a new feature variable to repeat the above operation for 2062 retail units. The new clustering results obtained describe the categories to which each retail unit belongs throughout the pandemic.

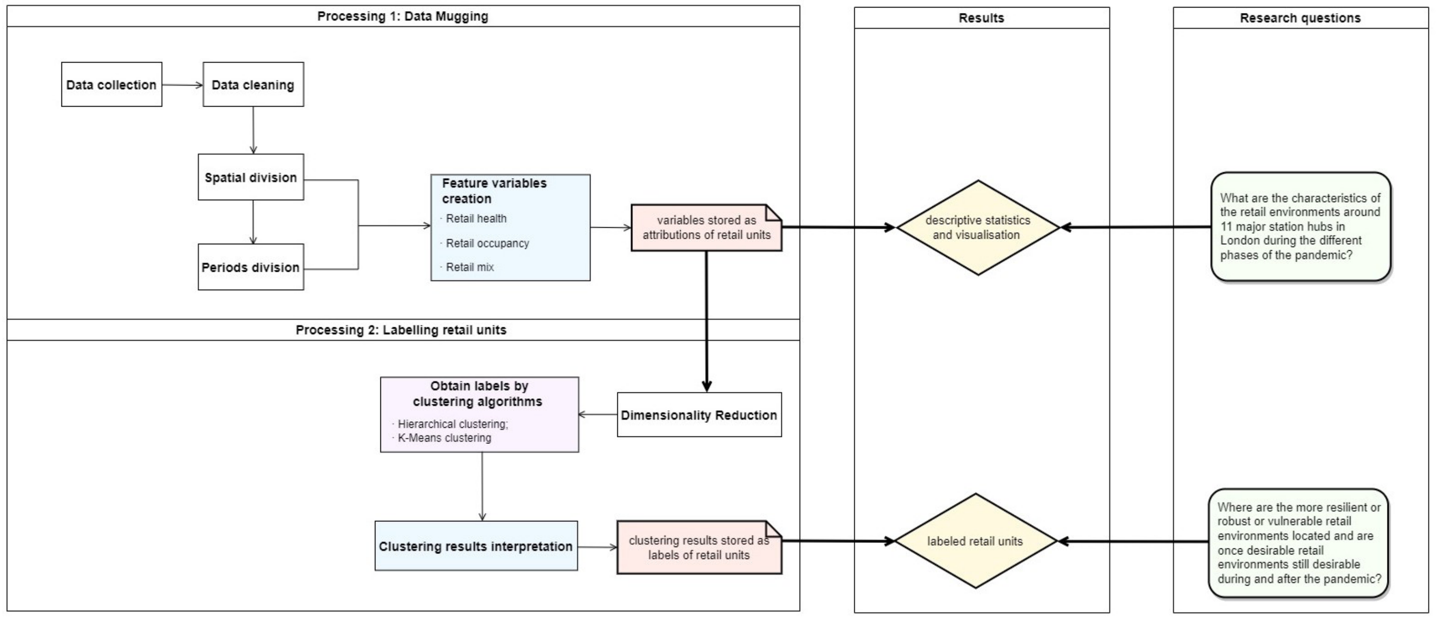


Fig 26: Methodology Flowchart

RESULTS

■ Clustering results

Cluster labels No.	Cluster labels	Changes of rates Compare to the phase-I: before the outbreaks of COVID-19 (2019.10-2020.02)						
		retail openings rate	recent retail openings rate	retail closures rate	long-term vacancy rate	current phase vacancy rate	SHDI	average distance to station hub
C_0	Robust							
C_1	Resilient	+	+	-	-	-	+	+
C_2	Active	+	+	+	-	+	+	+
C_3	Stagnant		-	-		-		
C_4	Vulnerable	-	+	+	+		-	-
C_5	Abnormal	+	+					+

Table 6: Feature variables clustering results descriptions

□ Robust: Values around 0 for each characteristic variable; implying that the retail units within this cluster had not changed significantly from the pre-epidemic period in terms of various features.

□ Resilient: Clusters with an increase in openings rate, recent openings rate and SHDI, and a decrease in closures rate and vacancy rate compared to the pre-epidemic period; resilient to impacts.

□ Vulnerable: With the opposite features to resilient ones; more vulnerable to impacts.

Clusters that are outside of the predictions:
□ Active: the openings rate, recent openings rate, SHDI and current phase vacancy rate have all increased, but the long-term vacancy rate has decreased; the turnover of retail businesses within this cluster is relatively frequent.

□ Stagnant: have similar features to the robust cluster, with most features remaining unchanged from the pre-pandemic period. However, the closures rate, recent openings rate and vacancy rate at current phase have decreased; fewer new retail facilities are being added to these units, but the existing businesses are also remaining in operation during the current phase, thus maintaining a low turnover rate.

□ Abnormal: The change in openings rate, recent openings rate are all close to or equal to 1. Most of these units have only one or two new openings within them, which makes their data distribution heavily skewed. This cluster label is not very meaningful for measuring retail environments.

■ Spatial distribution of results

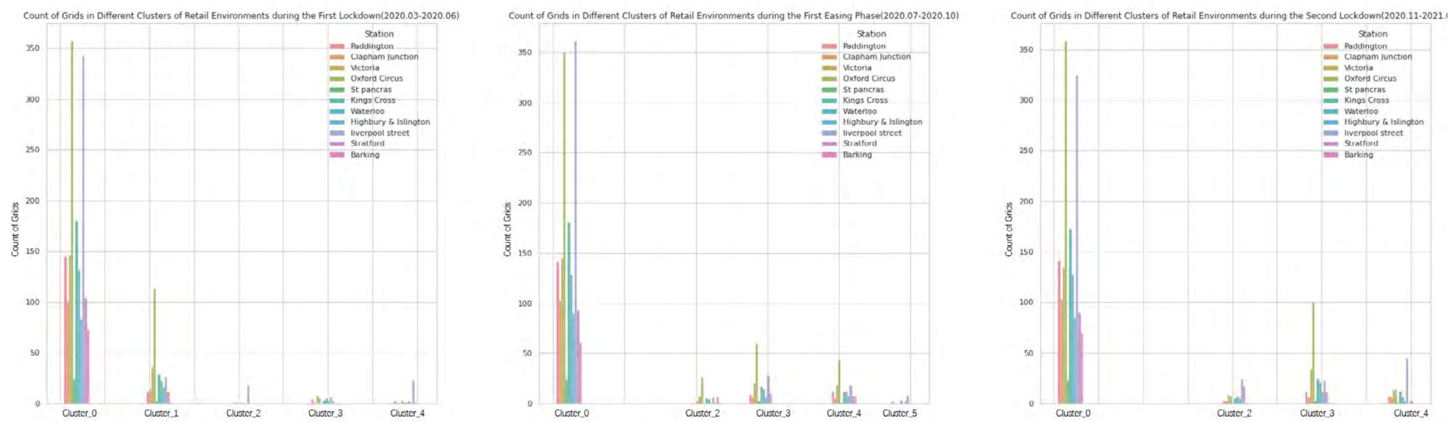
Fig 4.1 shows the spatial distribution of different clusters of retail units in each research phase. Fig 4.2 illustrates the distribution of the number of units from different station hubs within each cluster. These visualisations show how the capacity of each unit to resist impacts varies over time, and represent the spatial distribution characteristics of the different clusters.

Fig 27: Distribution maps of retail units for feature variables clustering results

□ Phase II - First lockdown
· Most of the retail units around the station hubs were relatively robust;
· But several retail units around Liverpool Street station were clearly vulnerable or active.
· Most vulnerable retail units are concentrated near to the centre of stations (station hubs which are offering national rail services, including Paddington, Victoria and Waterloo);

□ Phase III-First easing
· Vulnerable retail units are appearing around every station hub;
· There are significantly more retail units in stagnant status;
· Oxford Circus were more vulnerable than Liverpool Street;

□ Phase IV-Second lockdown
· More retail units became stagnant;
· Liverpool Street were more vulnerable than Oxford Circus, some of features of the retail facilities around Liverpool Street contributed to its vulnerability



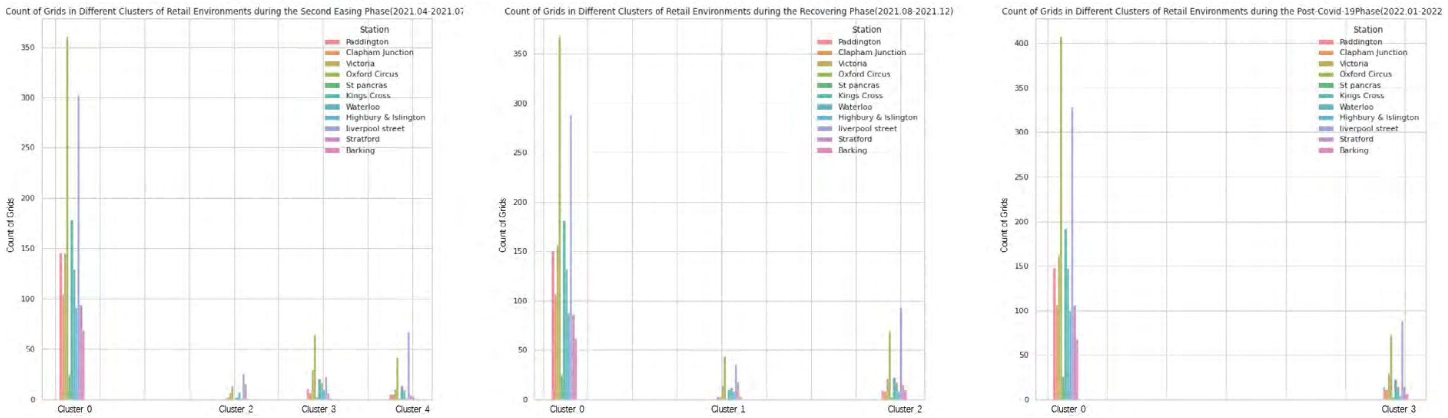


Fig 28: Count of retail units in different clusters during each phase

□ Phase V-Second easing
· More areas around Liverpool Street and Oxford circus went from stagnant to vulnerable.

□ Phase VI-Recovering
· More retail units became more active than before the pandemic, with all rates increasing except for the long-term vacancy rate;
· Around 7% of retail units showed resilience after the impacts, mainly around Liverpool Street, Oxford Circus and Stratford, with improved retail health and retail mix indicators.

□ Post COVID-19 phase
· 86.6% of retail units showed no significant change compared to the pre-pandemic period, so it can be assumed that these areas have gradually returned to their original levels;
· Some retail units are stagnant, because retailers usually focus on new store openings or review their estates after the Christmas in the first quarter.

■ Whether once desirable retail environments still desirable during and after the pandemic?

Cluster labels No.	Number of grids	Clusters the grids belong to in each phase					
		Phase II	Phase III	Phase IV	Phase V	Phase VI	Phase VII
a	1427	C_0	C_0	C_0	C_0	C_0	C_0
b	324	C_1	C_3	C_3	C_3	C_0	C_0
c	137	C_0	C_4	C_4	C_4	C_2	C_3
d	174	C_0	C_0	C_0	C_0	C_3	C_3

*C_0: Robust; C_1: Resilient; C_2: Active; C_3: Stagnant; C_4: Vulnerable;

Table 7: Description of clusters for retail environments from phase II to phase VII

Investigate the changes in the belonging clusters of each grid throughout the research phase. Using the labels of the clusters of all the grids at each stage as categorical variables and clustering them again using K-Means method.

□ Cluster_a
· grids were robust across all phases;
· consistently robust across all phases of the impacts.

□ Cluster_b
· grids that were resilient.

□ Cluster_c
· grids that were vulnerable during the lockdown and easing phases;
· are more vulnerable to external shocks and less able to adapt to the changed external environment.

□ Cluster_d
· grids that remained robust during the pandemic but stagnated during the recovering and post-COVID19 phases;
· grids were able to resist the impacts from the pandemic but lacked the ability to adapt to the new normal.

RESULTS FINDINGS

■ The characteristics of the retail environments during the different phases of the pandemic

This study finds that retail environments around different station hubs do not always respond immediately to external shocks such as lockdowns in the immediate phase, and most have lags. The exception is Liverpool Street station, where retail environments around this station are more sensitive to lockdowns, with more vulnerable units present during this phase compared to other station hubs. This difference may be related to the fact that the main types of consumers (tourists and commuters) for the retail facilities around these two stations are different, but more research is needed to further investigate this. For those regional station hubs connected to suburbs and other cities (Paddington, Victoria, Waterloo, King's Cross and St Pancras), their resilience to impacts has gradually diminished since the beginning of the First lockdown but is gradually returning as the various restrictions are lifted during the recovery phase. However, there are still some retail units that continue to show vulnerability to change.

■ Are once desirable retail environments still desirable during and after the pandemic?

Stratford, the station hub whose most passengers' volume across GB from 2020 to 2021, exhibits similar characteristics to Oxford Circus station. It was also found that the retail environments at Westfield, a large commercial complex located near the station, are more desirable than the retail units located along the street to the southeast of the station. the research area in southwest Liverpool Street station is the City of London, densely distributing high-rise commercial office buildings. Retail units here were less resilient to impacts during the lockdown phases and became stagnant after the pandemic. A possible explanation for this phenomenon is that the main consumers in the area, commuters working in the area, were forced to work from home during the lockdown, so the loss of this clientele was an obvious blow to nearby retail. And as studies and reports have shown that many employees chose to continue to work remotely after the pandemic, it can be inferred that the traffic usages in these areas will not return to pre- pandemic levels, at least soon. The same phenomenon has

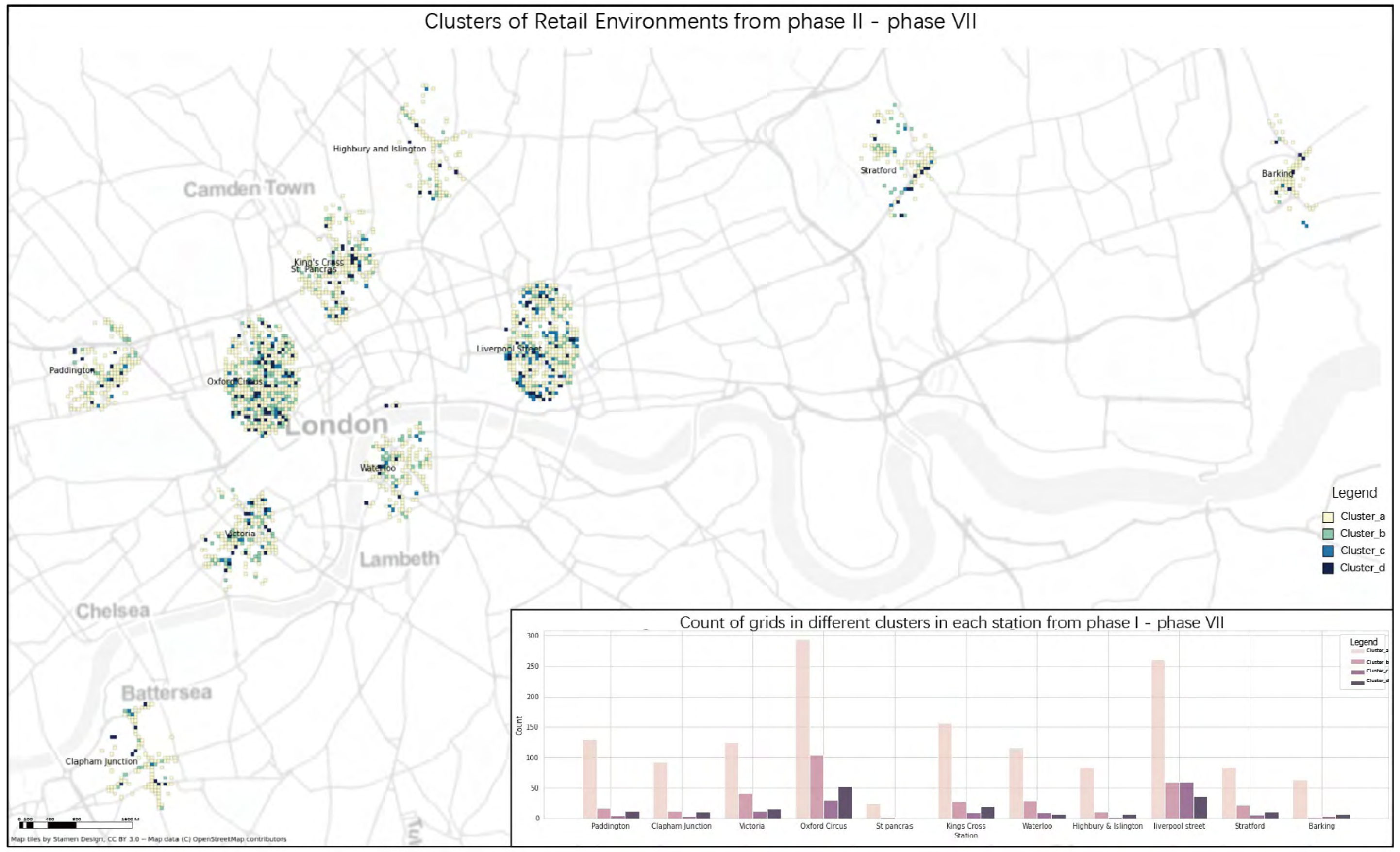


Fig 29: Distribution maps of clusters for retail environments from phase II to phase VII

been observed around Victoria and Waterloo. Around Victoria station hub, the retail units along the railways with higher density of accommodation sector are vulnerable or stagnant, too. But the more densely populated units around Waterloo, along the Thames, have withstood the impact and shown resilience. The retail facilities in the units around Paddington are predominantly hotels, but their response to the impact has been disparate. Some units are more vulnerable, others more robust or even resilient. But in general, those with more catering sectors are more vulnerable to the impacts.

The comparison between King's Cross and St Pancras shows that the retail units within St Pancras are more resilient in terms of comparison but more vulnerable in terms of catering, showing similar characteristics to Paddington. The retail units within King's Cross station have withstood the impact of the pandemic at all stages and are more desirable than in St Pancras. However, the more catering based retail units on the east side of King's Cross station hub are vulnerable.

LIMITATIONS OF THE RESEARCH

The limitations of this study come from two aspects. Firstly, due to data and time constraints, this study only used a single POI data and limited analytical tools to describe and generalise the phenomena presented in the dataset. Further validation of the above research findings would still require additional external data for correlation or regression analysis. The shortcomings also stem from the fact that the outliers have not been adequately dealt with; the units classified as less resilient to shocks in Transit and commuter hubs often have only a few shops within them, and the closure or opening of individual shops can significantly interfere with the features of these units. This means that these results are bias and therefore not sufficiently informative. Also, the failure to exclude these units from the clustering analysis may cause inaccuracy in the clustering results.

APPENDIX

The tools used in this analysis include QGIS, Tableau and Excel.
The code used to undertake this analysis is accessible here: https://github.com/tong19940114/CASAdissertation.

RECOMMENDATIONS FOR FUTURE WORK

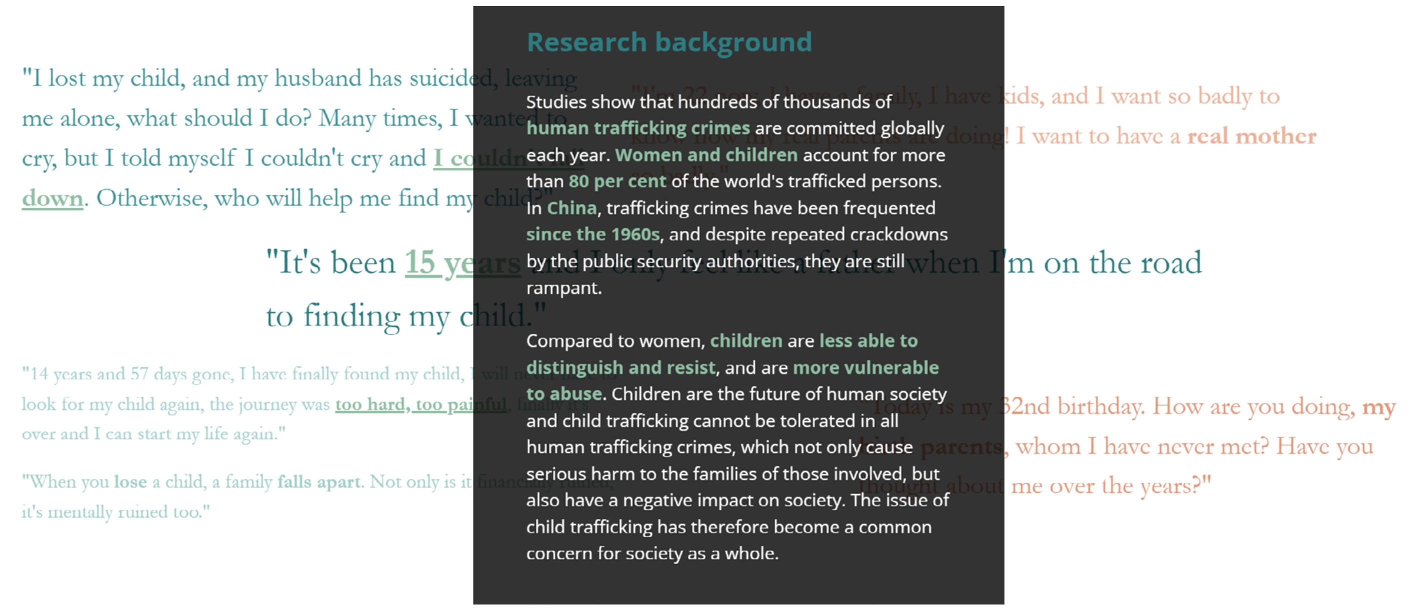
The above research findings and limitations also provide some insights for further work; for instance, the reasons why the retail environments of different station hubs fall into the same cluster may vary, depending on the function of the surrounding area and the main types of passengers. More fine-grained correlations can be built upon by collecting retail sales, social media check-in data, GPS footfall data and demographic data within a smaller area. Besides, it is also possible to compare the resilience of retail environments in different zones around the same station hub. The correlation of elements such as functional urban zoning, pedestrian accessibility, urban form and the resilience of the retail environment in different areas around the same station hub can also be further explored.

CONCLUSION

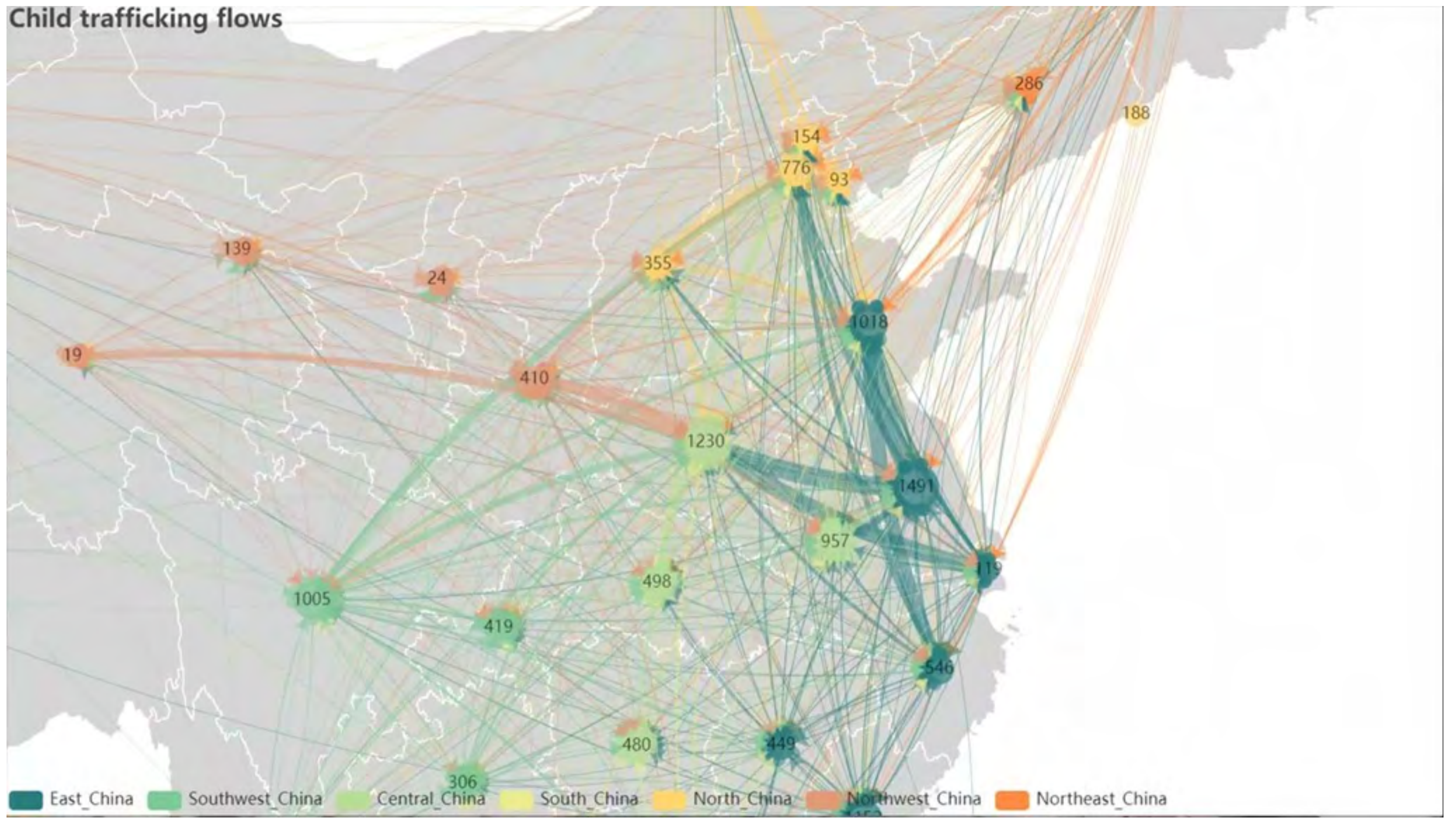
Since the outbreak of COVID-19 and the accompanying restrictions on mobility have caused sustained retail losses and long-lasting changes to consumer travel and shopping habits. This study sought to assess the performance of the retail environment around 11 major London station hubs during the various phases of the lockdown using key indicators, and successfully identified those areas that were more vulnerable and resilient, building the basis for a more microscopic assessment of the impact of the lockdown policy and the pandemic on the retail environment, as well as providing a basis for subsequent commercial location, area and transport planning. This research addresses the new trend of using machine learning methods for retail site-selection, driven by spatial big data. The clustering methods show the potential to identify the retail environments with particular features. Furthermore, this study allows for the interpreting, extending and validating the results using more data.

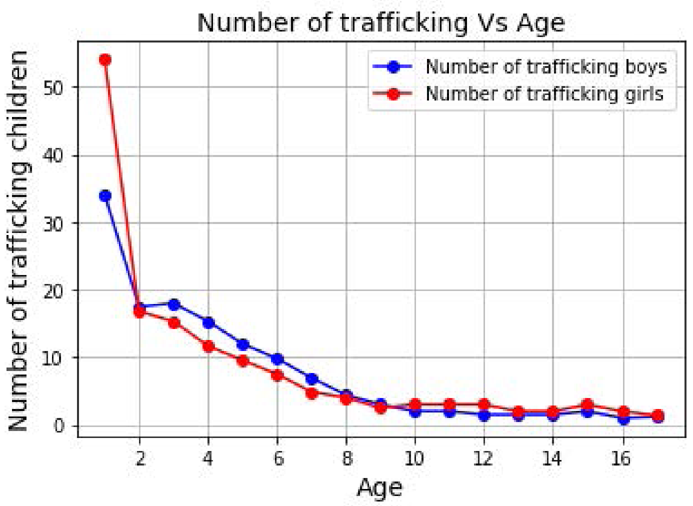
Lost Children, Stolen Life — Child Trafficking in China

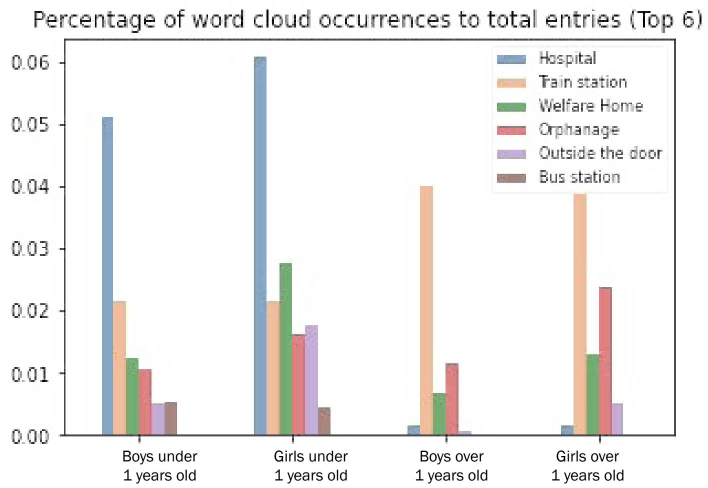
A spatial data visualisation and interactive webpage porject, London, 05/2022
View the webpage in here: https://rollingcasaer.github.io

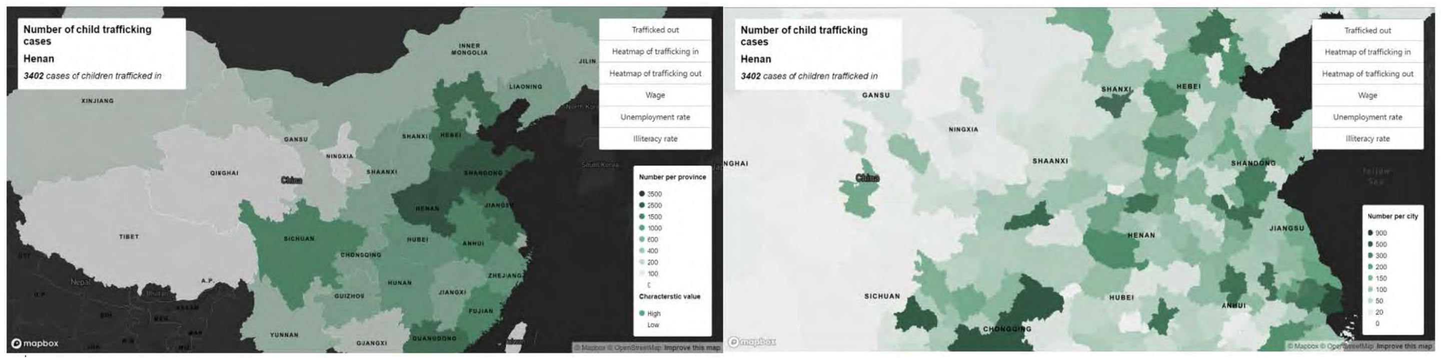


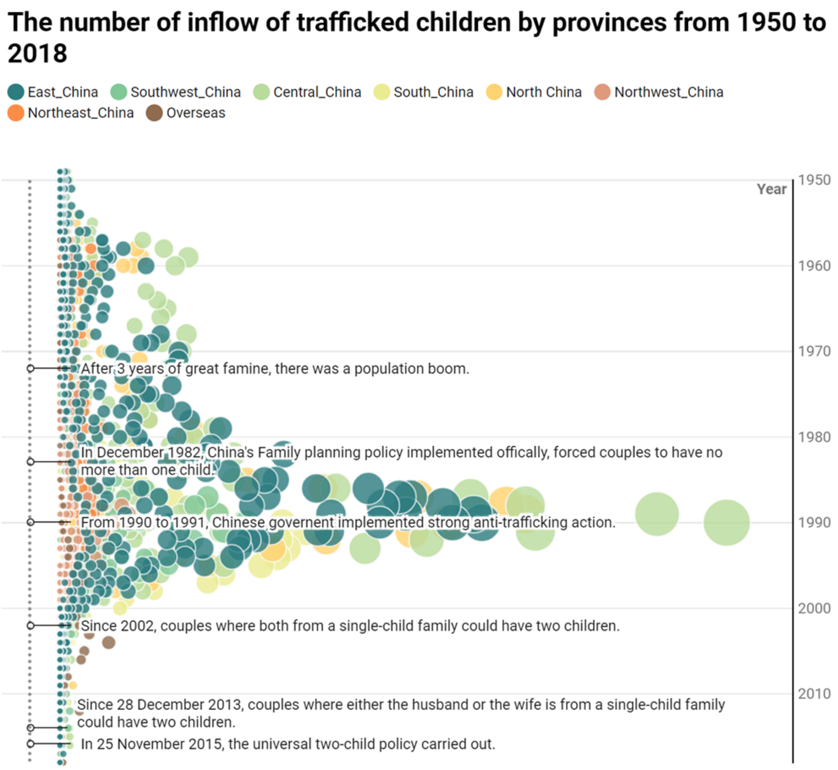
Children trafficking is considered one of the most serious crimes of the 20th century (Wang, 2015), Looking back at the trend of child trafficking around the world, most countries link the purpose of child trafficking with sexual slavery and labour exploitation (Martinho, Gonçalves and Matos, 2020), but considering the differences in time, space and culture, the main purpose of child trafficking does not seem to be the same in China. The phenomenon of child trafficking in China has a long history and has for a long time been a neglected crime, even considered a 'tradition' (Tan, 2018). Child abduction and trafficking is a serious and irreversible danger to the abducted child, the family of origin, the family that bought the child, and society at large (Wu, Liu and Duan, 2017). Of particular note is that, China experienced a large number of child trafficking incidents from the 1980s to the early 2000s, where the ultimate purpose of these cases is more linked to family inheritance and getting rid of the moral humiliation of being childless (Liu, 2018), so the perpetrator is both a participant in trafficking and possibly an illegal “guardian” who protects the child’s growth (Ren, 2004) . This makes child trafficking in China complex with specific characteristics and factors. The purpose of this project is to analyse the characteristics of child trafficking in China through a visual and interactive approach. In section two, we introduce the background information and conduct literature review. Then we describe the data processing in section three. Detailed data analysis, visualisation methods and interactive integration are explained in sections four, five and six respectively, based on which we form our conclusions in section seven.

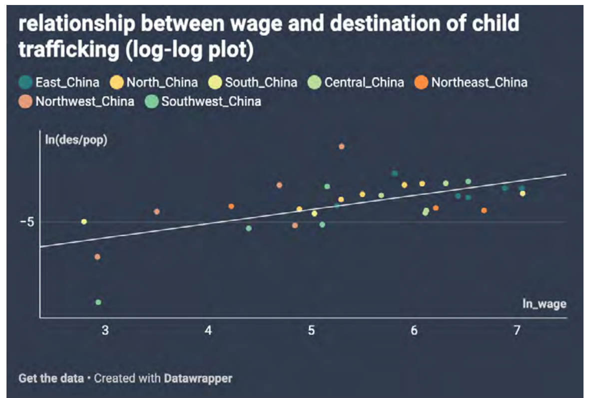


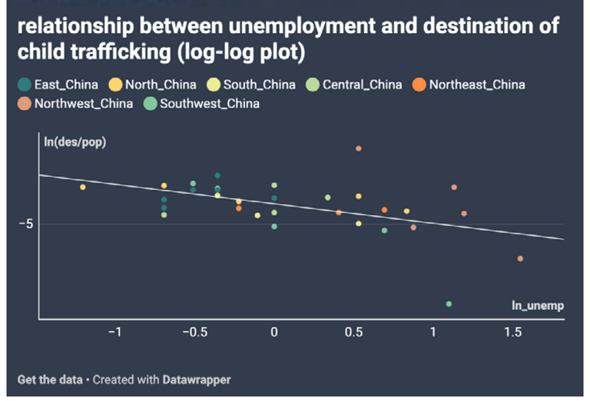








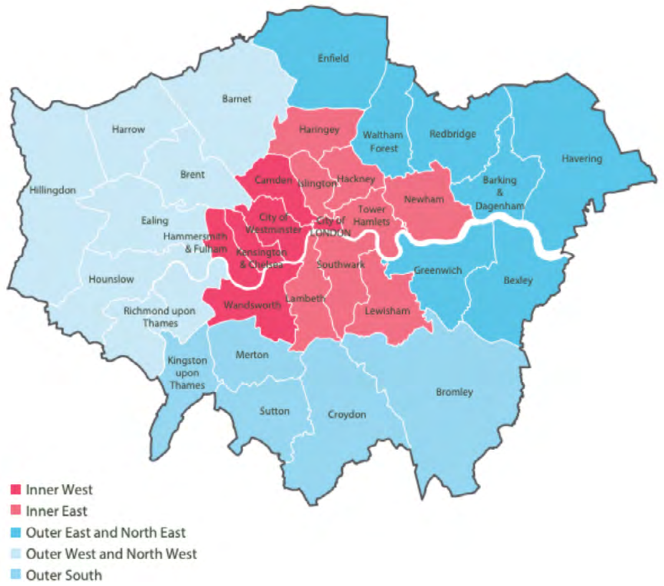




Potential Correlation Between the Number of Airbnb Listings and Growth of House Prices in London

A spatial data analysis and visualisation porject, London, 01/2022

Map of London Boroughs

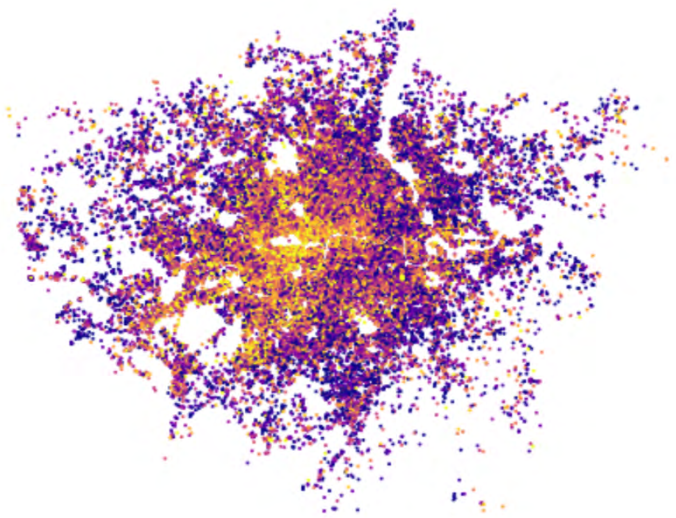


BACKGROUND

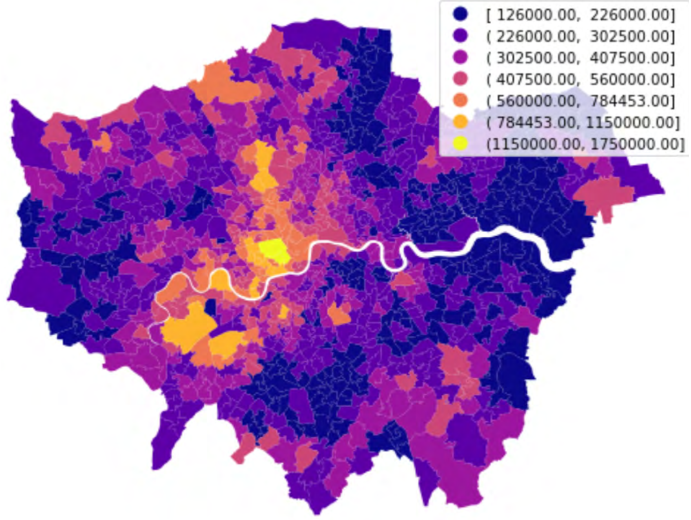
As the study of Sheppard and Udell (2016) states, in New York, Airbnb listings have a statistically significant correlation with local house price. And as long as Airbnb properties are not the negative source primarily, this correlation would be positive. So, investors are very likely to make speculative purchase of residential properties for potential capital benefits once the expansion of Airbnb has more positive impact on the area where these properties located. And as a kind of speculative purchase, the investors tend to rent properties in short-term(in other words, becoming listings in Airbnb) for higher profits [Ryan and Ma, 2020]. It will further enhance the growth of Airbnb listings in these areas [Todd, James, et al. 2021].

For the purpose of determining which areas in London have a potential correlation between the growth of Airbnb listings and speculative purchases attracted by rising house prices over the same period, the median house price was used as the independent variable and the growth of airbnb listings as the dependent variable for the LSOA resolution in London(LSOA is a suitable geographic level for narrowing the targeted areas in this analysis because it's level is smaller than other resolutions relatively, with sufficient available data, and has generally equal population in one of areas[ocsi, 2019]). Then a local regression analysis was conducted for each study area using the geographically weighted regression(GWR) to finally filter out the areas with the highest correlation between the two variables.

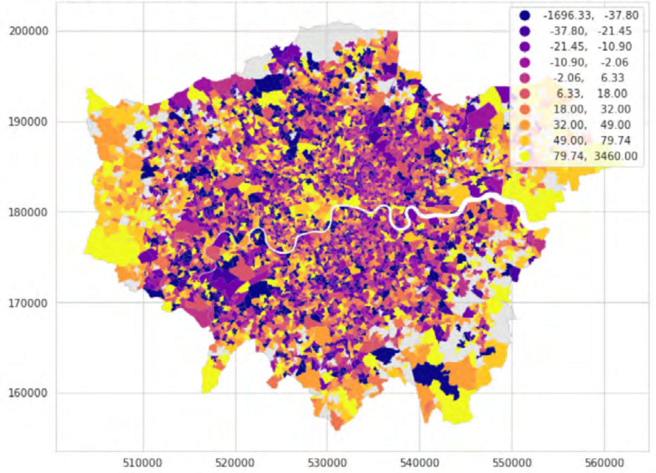
Distribution Map of Airbnb Price Per Night in London



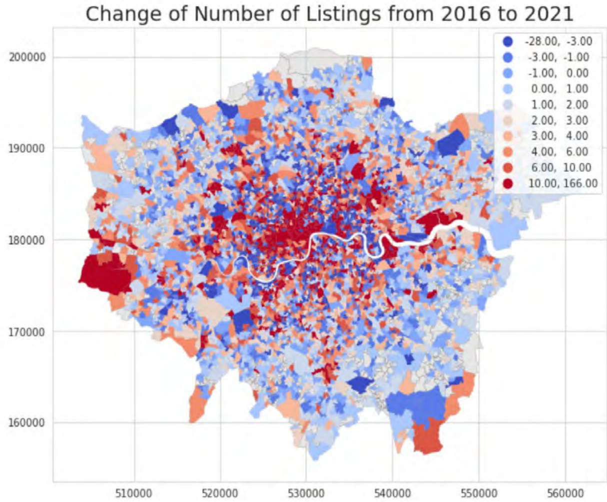
Distribution Map of Airbnb Price Per Night in London by Boroughs



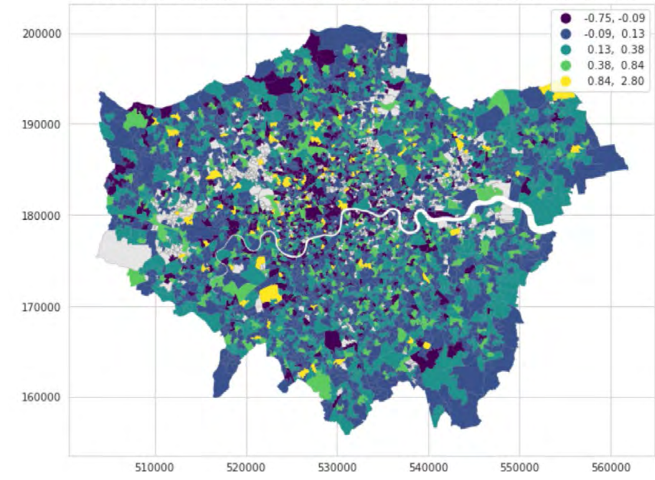
Change of Mean Price of airbnb from 2016 to 2021 in London



Change of Number of Listings from 2016 to 2021



Median House Price Growth Rate from 2016 to 2021

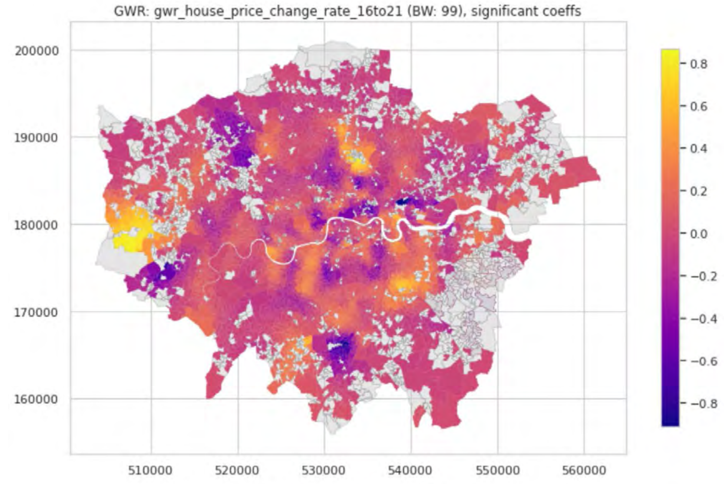


APPROACH

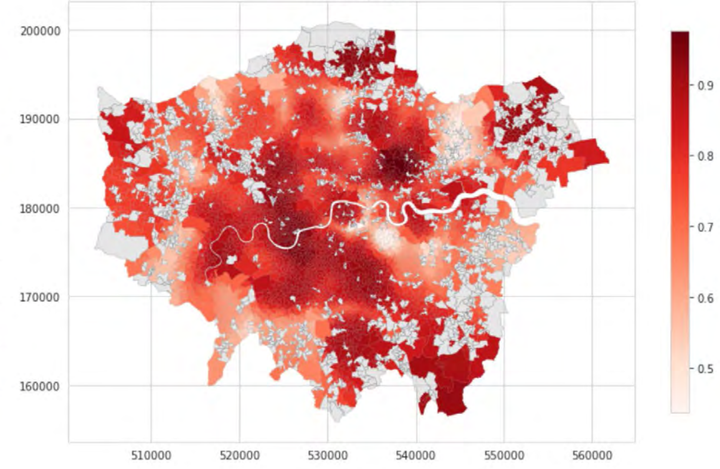
Geographically weighted regression (GWR) as a spatial statistical technique, is the primary approach to study the problem I raised in previous section. Unlike the traditional global regression model, GWR is a local regression model which captures process's spatial heterogeneity [Fotheringham, Brunsdon and Charlton, 2002], and considers that the data which close to the research targets (in this analysis, the targets are the centre of gravity in each LSOA area) have heavier weight in the model than data further away [Fotheringham, Brunsdon and Charlton, 2002]. In each targeted area, GWR would create a dataset by using nearby data then run a regression model on that area. By calculating or specifying the parameter that controls locality (called bandwidth), it is determined how close to the target area the observations are to be considered relevant.

As a local regression model, GWR increases model fit and reduces residual spatial auto-correlation [Oshan, T. et al., 2019]. Besides, because that every area has its own local regression model, we can evaluate the model accuracy in each area by checking their own the local R square.

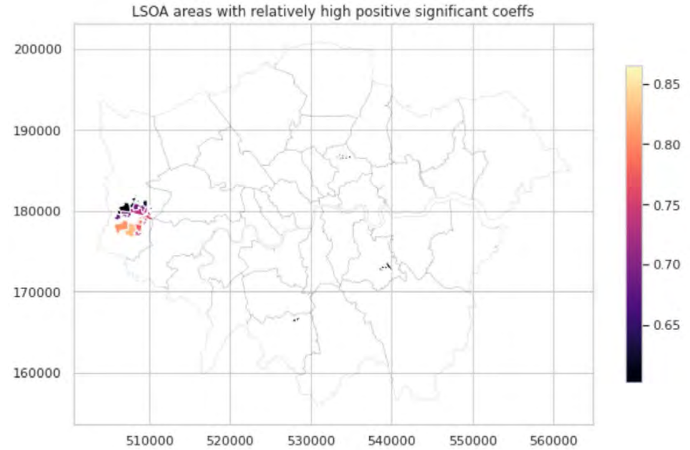
GWR: gwr_house_price_change_rate_16to21 (BW: 99), significant coeffs



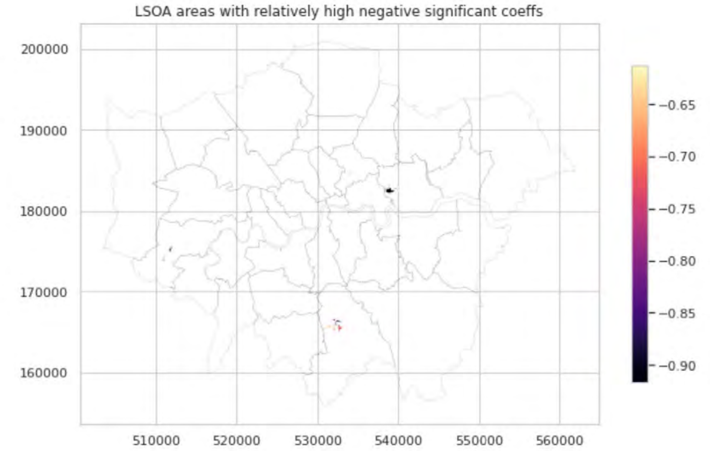
GWR: local R2



LSOA areas with relatively high positive significant coeffs



LSOA areas with relatively high negative significant coeffs



RESULTS

To filter the estimates for statistically significant parameters based on a significant threshold α, the mgwr package allows to build a matrix of the significant local coefficient estimates and sets all non-significant results at each site to zero, which is help to plot maps only colouring observations with significant local coefficient.

According to plots above, the areas with relative higher positive significant local coefficients(>0.6) could be filtered and mapped plotted(Figure 5.). The filtered areas are concentrated in Hillingdon, Hackney, Lewisham and Sutton respectively. That is to say, the growth rate of median house price in these areas has more positive impact on their change of number of Airbnb listings than other areas. In contrast, areas with higher negative significant local coefficients(< -0.6) are located in Croydon, Hackney, Hounslow, Newham and Tower Hamlets.