Chen *et al. Computational Urban Science* (2025) 5:30
https://doi.org/10.1007/s43762-025-00191-z

Computational Urban Science

ORIGINAL PAPER

Open Access

The impacts of specific place visitations on theft patterns: a case study in Greater London, UK

Tongxin Chen[1,3*], Kate Bowers[2] and Tao Cheng[3]

Abstract

Exploring the impacts of population place visitation on crime patterns is crucial for understanding crime mechanisms and optimising resource allocation in crime prevention. While recent studies have broadly examined dynamic population activities at specific places from geo big data, limited crime-related studies have utilised this measurement to disentangle the impact of specific place visitation on urban crime patterns. This study aims to investigate the impact of population activities at different urban functional places on theft levels across different urban areas and distinctive social changing contexts. We utilised geo big data (mobile phone GPS trajectory records) collected from millions of anonymous users to measure footfalls (counts of visitations) attached to place types on weekdays and weekends. An explainable machine learning approach was applied to analyse the impacts of place visitations on theft levels: the 'XGBoost' algorithm trained a high-performance regression model and 'SHapley Additive exPlanations' (SHAP) values were measured to identify the contributions of different visitation variables to theft levels at specific spatial and temporal scales. Using the police records and geo big data in Greater London from 2020 to 2021, the optimised model revealed that visitation to 'Accommodation, eating and drinking' services during weekdays had the most significant impact compared to 17 other types of place visitations. Further, the influence of place visitations on theft varied across different local urban areas corresponding with changes in social restrictions during the pandemic. Specifically, the urban areas where theft was most impacted by visitation at specific types of places (e.g., accommodation, eating and drinking services) shifted to outer London during the first national lockdown compared to normal times. The findings provide further evidence from direct micro-level analysis and contribute to tailoring policing strategies in places with different contexts and urban visitation patterns.

Keywords Mobile phone GPS data, Human mobility, Ambient population, Explainable machine learning, Urban vitality, Geo big data

*Correspondence:
Tongxin Chen
tongxin.chen@hull.ac.uk
[1] Centre of Excellence for Data Science, Artificial Intelligence and Modelling, University of Hull, Cottingham Road, Hull HU6 7RX, UK
[2] Department of Security and Crime Science, University College London, Tavistock Square, London WC1H 9EZ, UK
[3] SpaceTimeLab for Big Data Analytics, University College London, Gower Street, London WC1E 6BT, UK

1 Introduction

The routine activities of citizens in urban places are closely intertwined with available opportunities for crime occurrence, hence they significantly influence urban crime patterns in space and time (Brantingham and Brantingham 2016). Traditionally criminological research has used a simple count of the specific types of places to represent high-volume population activity in urban settings (e.g., how many shopping centres, bars or public transit stations are present). These proxy methods, which use static land uses to represent activities at place have revealed strong relationships with crime patterns (Bernasco and Block 2011; Brantingham and Brantingham 1995; Barnum et al. 2017; Curman et al. 2015). With the growth of location-based services and applications, the accumulation of different types of geo big data enables the measurement of population activities through a more direct footprint of citizen activity. Recent studies have revealed significant connections between population dynamics sensed from geo big data (e.g., mobile phone data and social media data) and crime patterns in urban areas (Andresen 2011; He et al. 2020; Song et al. 2019; Chen et al. 2023a). Despite this growing body of research, gaps remain in our understanding of the intricate interplay between crime and population activity behaviours in specific urban contexts, particularly across different types of places, local areas and under varying socio-economic conditions such as during the pandemic. Further investigation is required to examine variations in the place-specific of population activity (which we refer to here as place visitation) on crime patterns in local urban areas. Simply put, to further understand crime problems, we need combined models that measure both the size and variation of dynamic populations in time and space, but also distinguish between the activities of those populations by place function.

To address these research gaps, this study aims to answer the following specific research questions: (1) How do dynamic place visitations influence specific crime patterns in different types of places? (2) Does the impact of specific place visitation on crime level vary under different social contexts, such as during the COVID-19 pandemic? (3) How does the influence of place visitation on crime differ across local urban areas?s?

We investigate these questions through a comprehensive analysis of theft incidents in London neighbourhoods, represented by Lower layer Super Output Areas (LSOAs) covering he period from 2020 to 2021. First, place visitation variables were quantified into monthly footfalls (sensed from large-scale anonymous mobile phone GPS data) at various places in different urban areas. Subsequently, the impacts of place visitation variables on theft levels were investigated by employing an explainable machine learning approach consisting of training tree-based XGBoost models and implementing a 'SHapley Additive exPlanations' (SHAP) strategy. Last, both global and local impacts based on SHAP values were employed to investigate population activity-based factors in explaining crime levels at different types of places across urban areas during several defined pandemic periods.

The structure of this paper is outlined as follows: The Sect. 2 reviews the theoretical foundations and empirical research in crime pattern analytics using emerging geo big data sets and discusses related studies during the pandemic. The Sect. 3 provides a description of the study area and related data used. The Sect. 4 illustrates the generation of place visitation variables and the explainable machine learning strategy used in the study. The Sect. 5 presents the experiments conducted in the study area and the obtained results. The Sect. 6 discusses implications and limitations based on the empirical findings. Finally, the Sect. 7 summarises the contributions and highlights further work.

2 Background

Crime opportunity theories reveal that citizens' daily routine activities intertwined with the available crime opportunities (e.g., targets) in urban settings play a key role in distributing crime patterns in space and time (Cohen 1981). In particular, the places attracting large volumes of population visitations are highly associated with crime occurrences. Overall, opportunity theories, such as crime pattern theory (CPT) and routine activity theory (RAT), provide insights into how crime opportunities can be affected by dynamic population changes in particular urban areas. RAT implies that the spatial patterns of crimes in urban areas are shaped by varied crime opportunities generated by the complex interplay of different population's activities in urban settings (Cohen and Felson 1979; Felson 2016). To detail, crime opportunities arise from the interactions between motivated offenders, suitable targets/victims (e.g., population) and the absence of capable guardians in a specific time and space.

CPT also emphasises the role of population activity in creating crime opportunities, noting that these typically arise in busy areas familiar to offenders and where there are suitable targets and victims (Brantingham and Brantingham 1984; Brantingham et al. 1981). Further, CPT describes the places with higher levels of crime as following two distinct mechanisms: 1) crime generators which are the places that generate crime by attracting large volume populations to a particular location (e.g., shopping malls and transportation stations) – these places tend to have a high volume of foot traffic and a concentration of businesses and services, which can create opportunities for criminal activity; and 2) crime attractors which are places that may not attract a large population but are still associated with higher levels of criminal activity due to their characteristics (e.g., abandoned buildings) which are likely to attract offenders to engage in criminal activity (Brantingham et al. 1981; Brantingham and Brantingham 2016).

Before the widespread use of mobility datasets, earlier studies often utilised the number of crime generator locations (e.g., shopping centres or retail places) as a proxy for measuring the levels of visitor/population-based crime opportunities across urban areas (Sherman et al. 1989; Eck and Weisburd 1995). For example, Groff and Lockwood (2014) investigated the influence of criminogenic facilities (e.g., bars, liquor stores and public transportation) on crime rates across street segments in Philadelphia, USA. They used place data (point of interest) to identify the locations of these facilities and analysed their impacts on crime patterns in the surrounding area. Within the same city, Haberman and Ratcliffe (2015) examined the impact of potentially criminogenic places on street robbery counts in census blocks. They examined the connections between aggregate crime levels and census block locational characteristics during both daytime and nighttime. Their findings showed that certain places have a greater influence during the day, while others are more significant at night. Furthermore, Barnum et al. (2017) investigated both similarities and differences in relationships between place features and crime across urban settings in the US cities: Newark, Los Angeles and Cincinnati – demonstrating variation and the context-specific nature of the crime-place connection. Schnell et al. (2019) demonstrated that facilities generating crime opportunities are significantly associated with the spatial distribution of violent crime occurrences on street segments. Specifically, the presence of particular place types, such as retail and commercial properties, was found to be related to higher levels of violent crime. A clear limitation of these approaches is that relying on static representations from place counts cannot accurately reflect real population activities, which are inherently linked to the ever-changing dynamics of crime opportunities across time and space.

By using geo big data, researchers can more effectively track and analyse the dynamic trends of the population's routine activity, providing deeper insights into population shifts and their implications on various societal issues. In this context, utilising geo big data offers a valuable advantage to capture population dynamics and link the sensed patterns to crime incidents (Shaw et al. 2016; Wu et al. 2018; Chen et al. 2018). For instance, Malleson and Andresen (2015) demonstrated that the inclusion of social media data (geo-tagged Twitter data) provides a more accurate representation of the ambient population, leading to a better understanding of the spatial patterns of crime. Kontokosta and Johnson (2017) also used Twitter data to simulate ambient population distributions and revealed the spatial relationship between ambient population and crime. In addition, Hanaoka (2018) estimated the ambient population using mobile phone data and revealed the strong relationship between population dynamics and certain types of street crimes, such as theft and assault. Further studies have also examined a strong correlation between property crime (e.g., theft) incidents and dynamic ambient populations sensed from mobile phone data collected from cellular towers (He et al. 2020; Song et al. 2018; Johnson et al. 2021). Focusing on transportation settings, Zahnow and Corcoran (2021) used transit smart card data to measure the population's activity at bus stops and revealed a significant relationship with crime occurrences around bus stops. Specifically, higher usage levels at bus stops and the usage of certain amenities (e.g., seating) were found to be related to increased crime risk.

The COVID-19 pandemic had profound influences on the population's routine activities in cities due to the implementation of various restrictive policies aimed at curbing the spread of the virus. Measures such as lockdowns, social distancing, travel restrictions, and closures of non-essential businesses drastically altered the way people interact, work, and socialise, leading to significant changes in urban dynamics (Chen et al. 2023b; Cheng et al. 2022). It logically followed that the distribution of opportunities for crime in urban areas was also influenced as a consequence of the shifts in the population's activities (Halford et al. 2020; Stickle and Felson 2020). Empirically, the COVID and crime studies have revealed that property crime experienced a decrease while some violent crimes saw an increase (Ashby 2020; Mohler et al. 2020). However, there are only a limited number of studies that have explored the relationships between crime and population activity – especially at specific types of places or venues – during the COVID-19 pandemic. Halford et al. (2020) analysed police data and Google community mobility data to determine how changes in population activities influenced crime patterns at a region level. Then, Chen et al. (2022) employed a spatio-temporal stratified model to assess the relationships between the urban population's activities and crime patterns across urban areas (census block groups) of San Francisco during the COVID-19 stay-at-home mandate. Similarly, Cheung and Gunby (2022) found that variations in mobility were associated with shifts in property crime rates in New Zealand cities during the pandemic.

In summary, existing studies on the impact of population mobility on crime remain limited, as they tend to focus on large regions or city-level analyses, often overlooking localised interactions within specific urban settings across different areas. To address this gap, this study contributes to mobility-related crime research by examining how directly measured place visitations influence crime across different contexts – including different place types, urban areas and time periods.

3 Data

3.1 Study area and time period

In this research, Greater London as the study area is the largest city in the UK, housing over 8.8 million people in mid-2021[1]. The UK's Office for National Statistics (ONS) has structured a hierarchical system of census units, aiding in the compilation and examination of area-level demographic details[2]. London comprises 33 Local Authority areas and 4,835 designated local census zones named 'Lower super open areas' (LSOAs), which were used as the unit of analysis in this study.

Amid the global pandemic starting in 2020, London became one of the cities most affected by COVID-19 with a high number of infections in the UK. To counter this, the government rolled out a nationwide lockdown on March 23, 2020, including measures like a stay-at-home policy, pausing public transit, and the closure of non-vital businesses. As the situation evolved, restrictions were adjusted based on the infection rates which led to a second lockdown from November to December 2020, and a subsequent third phase from January to March 2021. Given the significance of these restrictions on population mobility, these months are used as temporal markers in the analysis that follows.

3.2 Mobile phone GPS trajectory data

Anonymous mobile phone GPS trajectory data for the London areas collected from broadly mobility-related application apps (e.g., navigation, route planning, outdoor sports) were provided by Location Sciences AI[3]. This data collection takes place following user agreements established under the General Data Protection Regulation (GDPR) framework, ensuring the privacy and protection of individual user information. The mobile phone GPS data used in this analysis includes 1,979,081 users (about 22% of the total London resident population) in Greater London during the observed two-year period (2020 and 2021). Abundant in terms of the diversity of trajectory collection apps and the sample rate of user numbers, our GPS dataset demonstrated a good representation for measuring the mobility activity for the London population (Chen et al. 2023b).

3.3 POI data

The Point of Interest (POI) data of London, which represents place locations and place types, are provided by Ordnance Survey[4]. The classification scheme of the POI dataset offers three hierarchical levels of information for POI types, consisting of nine groups, 52 categories, and 600 classes[5]. In this research, nine distinct POI types were selected according to the first-level classification (nine groups) from the dataset, including: 'Accommodation, eating, and drinking (AED)', 'Transport (TRA)', 'Commercial services (CS)', 'Attractions (ATT)', 'Sport and entertainment (SE)', 'Education and health (EH)', 'Public infrastructure (PI)', 'Manufacturing and production (MP)' and 'Retail (RET)'. To clarify, each POI is assigned to only one category from the nine categories with no overlap – each POI belongs exclusively to a single category at this classification level. For example, a POI classified as 'Commercial services (CS)' will not also be labelled under another category such as 'Retail (RET)'.

3.4 Police record data

Theft from the person data sourced from London police records was obtained from the Metropolitan Police Service in the UK's online police data portal[6]. The records detail each theft incident, specifying its location (latitude, longitude, LSOA index) and time (month and year). Due to the 'geomasking' process for keeping location details anonymous in the police records prior to public sharing, the LSOA is the most specific geospatial unit with reliable spatial precision for cumulative counts (Tompson et al. 2015).

4 Methods

This section presents the analysis framework including two components: (1) The generation and measurement of the place visitation variables. This describes the detection of an individual's stays/stops from the raw mobile phone GPS trajectory. Next, we linked detected stays to place locations for each geospatial area (LSOA) and aggregated to place visitations in LSOA at monthly level. (2) Explainable machine learning which was used to model the relationships between theft and place visitations.

4.1 Generation of place visitation variables

In this study, a place was delineated by a polygon-based representation named area of interest (AOI), which was generated based on the point of interest (POI) topology for each geospatial unit of analysis (LSOA). Figure 1 shows the process of generating areas of interest (AOI) from points of interest (POI) within each LSOA, which is a widely used geospatial approach for representing

[1] London datastore: https://data.london.gov.uk/dataset/londons-population

[2] Office for National Statistics: https://www.ons.gov.uk/

[3] Location Sciences AI (now known as Sorted): https://sorted.com/

[4] Ordnance Survey: https://www.ordnancesurvey.co.uk/

[5] The official Ordnance Survey POI classification scheme document explains the hierarchical levels in more detail: https://www.ordnancesurvey.co.uk/documents/product-support/user-guide/points-of-interest-classification-schemes-v3.4.pdf

[6] Data.police.uk: https://data.police.uk/

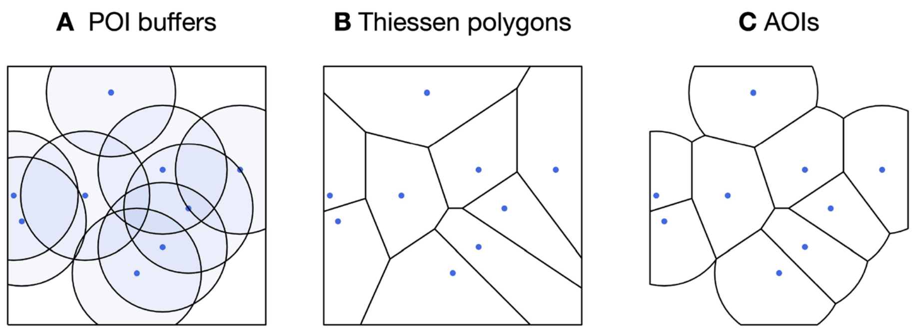

Fig. 1 Generating areas of interest (AOIs) from points of interest (POIs) in one geospatial area (represented by the square boundary). To clarify, there is no geographical boundary overlap among the created AOIs

different types of places or functional areas (Liu et al. 2019; Chen et al. 2020; Li et al. 2022). For POIs in one defined geospatial area, the intersecting areas as the represented AOIs (Fig. 1C) are generated from their Thiessen (Voronoi) polygons (Fig. 1B) then intersected with POI's buffer zones – the circle areas with a radius of 50 meters from each POI (Fig. 1A). This is a common strategy used to create the catchment areas to represent places (Jiang et al. 2015; Kucukpehlivan et al. 2023).

In this research, nine distinct AOI categories were produced from the first-level classification (the 9 groups introduced in Sect. 3.3) in the POI dataset (i.e., each AOI that was generated represented a single place that was labelled with a place type). The AOIs here include the physical area of the visitation places themselves but are also assumed as the 'catchment' area for the population visiting that place. In other words, AOIs in the LSOAs represent the on-the-street populations assumed to be associated with places as well as the populations at the places themselves.

Figure 2 shows the steps involved in detecting place visitations based on GPS trajectory data and AOIs in one geospatial area (LSOA) as an example. For each individual's GPS trajectory (Fig. 2A), a stay detection algorithm proposed by Hariharan and Toyama (2004); Pappalardo et al. (2019) was implemented to retrieve the stays (represented by the centre of red points in Fig. 2B) where a single user spends some time at a location (Zheng 2015; Zhao et al. 2016). In this work, each stay/stop was defined when a user spent at least 5 minutes within a 50-meter radius according to GPS point records. This parameters set is based on the assumption that a stop detected from the GPS points represents the natural range of human stops when visiting a place location in urban areas and is commonly found in related urban analysis studies (Zhao et al. 2015).

By spatially linking with the AOIs, all the stays attached with place information were then aggregated to footfalls representing population activity in places (i.e., place visitations) in space and time (Fig. 2C). It is important to note that three distinct types of stays, which are not strongly related to place visitation at locations (such as staying at home or working at a workplace) were excluded from consideration: (1) Stays in the early morning from 0 AM to 6 AM. The early morning hours typically do not showcase social activity at places due to significantly reduced activity and closed businesses, making these stays appear less representative of human place-visiting activity. Such human activity patterns have also been identified in various urban studies through the use of geo

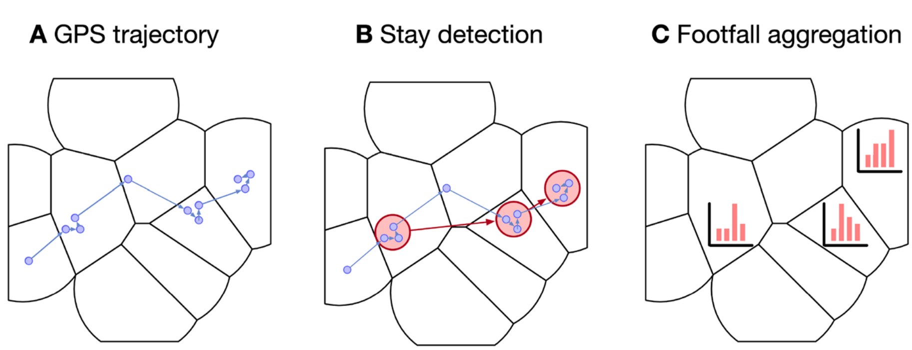

Fig. 2 Detecting place visitations based on GPS trajectory data

big data (Traunmueller et al. 2018; Sulis et al. 2018); (2) A user's home location, i.e., the stay location that a user visitations most frequently during the night-time period (from 11 PM to 6 AM) (Pappalardo et al. 2016; Verma et al. 2024); (3) A user's workplace which refers to the location of the stay's duration time above 6 hours from 7 AM to midnight as working behaviours are not the same as visiting behaviours and would confound the analysis. This distinction was also used in related studies to infer the work location from mobile phone data (Kung et al. 2014; Yan et al. 2019).

4.2 Measuring place visitation variables

For the measurement of the place visitation variables, the stays at each place (i.e., AOI) were separated into weekdays and weekends as there are evidenced differences in population activity across urban areas during these distinct period types (e.g., Niu and Silva (2023)). The stays were then separately aggregated as footfalls (stay counts) for the predefined geospatial unit (LSOAs in this study) and temporal unit (daily in this study) for weekdays and weekends. Next, the monthly daily average footfall (MDAF) for each type of place on weekdays (WD) and weekends (WE) was constructed as the place visitation variables for each month during the two years. The rationale for this monthly aggregation is to align this data with the police record data (described in Sect. 3.4), which is only provided at a monthly temporal resolution.

To elaborate, it is helpful to take the measurement of MDAF at a place type on weekdays in a single LSOA as an example. We first calculated the sum of footfalls at one place type such as retail (there might be multiple AOIs in one LSOA) on weekdays across a month, then calculated the daily average by dividing by the total number of weekday days in that month. This aggregation process resulted in 18 categories (nine types of AOIs for WD and WE) of place visitation variables (MDAF) for each LSOA during the 24 months in this study. Note that if a specific type of AOI is not present within the LSOA, the corresponding MDAF value will be assigned to zero.

4.3 Explainable machine learning

Machine learning techniques have been widely used in crime studies due to their strong predictive capabilities and versatility (Rummens et al. 2017; Alves et al. 2018). However, the black-box nature prevents a deeper understanding of the underlying factors driving crime mechanisms (Berk and Bleich 2013; Guidotti et al. 2018). While some machine learning methods (e.g., decision trees) can provide feature importance scores, these measures cannot reflect the true underlying relationships between explanatory variables and responders, particularly in examining their interactive effects (Lundberg and Lee 2017). Thus, there is a demand for explainable machine learning – an advanced approach focusing on developing and offering comprehensible justifications for model predictions – to help better identify factors associated with crime opportunities in space and time (Zhang et al. 2022; Campedelli 2022).

This study employed the ensemble learning method known as the XGBoost model, followed by the explainable machine learning strategy called 'SHapley Additive exPlanations' (SHAP). This section first describes the XGBoost for modelling the relationship between place visitation variables (MDAF) and theft levels, then introduces the explainable strategy SHAP, especially in terms of how this approach reveals the specific impacts of different place visitation variables on theft levels in the XGBoost model.

4.3.1 XGBoost regression model

XGBoost. This study selected the XGBoost (short for 'Extreme Gradient Boosting') regressor (regression model) to fit the theft levels and place visitation variables due to its efficiency and scalability in handling large data sets. As an ensemble learning method, XGBoost uses tree-based models as base learners and implements gradient boosting machines (GBMs) to iteratively combine the predictions of multiple weak learners (i.e., decision trees) with improving accuracy and generalisation capabilities (Freund et al. 1999; Chen and Guestrin 2016). The XGBoost algorithm works by adding new decision trees to the ensemble models with each decision tree attempting to correct the errors of the previous trees. The XGBoost model can be denoted as:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \tag{1}$$

where f is a function in a set of functions (i.e., function space) $\mathcal{F}$, i.e., the set of all possible decision trees. Then, $\mathcal{F}$ can be denoted as:

$$\mathcal{F} = \{f(x) = w_{q(x)}\} \quad (q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T) \tag{2}$$

In this context, q represents a function that navigates through the structure of a decision tree and outputs the corresponding leaf index (i.e., the unique identifier assigned to each leaf node in a tree), T is the leaf numbers in the decision tree, w is the leaf weight and K is the total number of the trees in the model.

Overall, XGBoost adds decision trees to minimise a specific objective function. The objective function combines a loss function (i.e., mean absolute error) that measures the difference between the predicted and actual values of the target variable and a regularisation term that penalises complex models to avoid overfitting. Thus, the objective function can be denoted as:

$$\mathcal{L}^{(t)} = \sum_{i}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega\left(f_t\right) \tag{3}$$

where $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ is the loss function that measures the discrepancy between the true label y_i and the predicted label $\hat{y}_i^{(t-1)} + f_t(x_i)$. $f_t(x_i)$ is a new function (e.g., a decision tree) to be added, where t is the current iteration. The sum $\sum_{i=1}^{n}$ adds up these losses for all n samples. $l^{(t)}$ is the objective function at the t-th iteration that the algorithm aims to minimize. $\Omega(f_t)$ is a regularisation term that controls the complexity of the model and prevents over-fitting, which can be denoted as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \tag{4}$$

γT is a regularisation component related to the complexity of the model. Here, T typically denotes the number of terminal nodes (leaves) in the tree-based model, and γ is the regularisation parameter controlling the influence of the tree complexity. $\frac{1}{2}\lambda\|w\|^2$ is another component called L2 regularisation related to the weights w of the model. Here, λ controls the influence of the weight regularisation, and $\|w\|^2$ represents the squared L2 norm (i.e., the sum of the squares of the elements of weight vector w).

Fundamentally, XGBoost is attempting to identify a new function $f_t(x)$ at each step and incorporate it into the existing model. This process minimises the cumulative loss of decision trees while ensuring the ensemble model's complexity remains controlled. During training, XGBoost determines the optimal split for each decision tree based on the feature that significantly reduces the loss function. Once the decision trees are added to the ensemble, XGBoost uses boosting to update the training samples' weights to lead to better overall performance. In summary, XGBoost determines the optimal split for each decision tree based on the feature that provides the most significant reduction in the objective function (indicating better prediction) and uses boosting to update the weights of the training samples.

Model training. The XGBoost regression model was prepared by linking the explanatory matrix in the shape of (4,835 LSOAs × 24 months) rows and nine columns (place visitation variables for each place type) with the counts of theft at LSOA and monthly level. In the model training setup, the first 19 months of data from January 2020 to July 2021 (80 % of the total dataset) was selected as the training set. The remaining dataset encompassing five months from August 2021 to December 2021 was selected as the testing set. In terms of data preparation, the training set and testing set (as part of the explanatory matrix X (i.e., the place visitation variables) and response variable y (i.e., theft counts)) were separately implemented using z-score standardisation. The model performance metrics for the XGBoost regressor are root mean square error (RMSE) and the coefficient of determination (R^2) for both the training set and testing set. Briefly, the R^2 indicates the proportion of the target variable's variance that the model can explain, while the RMSE quantifies the difference between the model's predictions and the actual observations. Therefore, a higher R^2 value and a lower RMSE indicate better model performance. For the hyperparameter tuning of the XGBoost regressor, grid search and cross-validation methods were utilised to optimise the parameter settings (10-fold cross-validation was set in GridSearchCV[7] at this step).

4.3.2 Model explanation using SHAP

Though traditional feature importance in tree-based models can provide useful insights, it is still limited in providing full interpretability. This is mainly due to the fact that feature importance calculation is based on heuristic methods (e.g., Gini importance, mean decrease impurity), which cannot measure the complex interactions across features. Another concern is that the feature importance can also be biased towards the preferential treatment of features with a large number of categories. Further, feature importance does not provide information about the directionality of the impacts (i.e., whether an increase in a feature value leads to an increase or decrease in the predicted value).

Shapley additive explanation (SHAP) as an advanced machine learning interpreter leverages the concept of Shapley values from cooperative Game Theory (Chalkiadakis et al. 2011) to fairly distribute the contribution of each feature towards the prediction for each individual instance, thereby providing detailed global and local interpretability in machine learning models (Lundberg and Lee 2017; Lundberg et al. 2018). In general, considering a machine learning model as a 'game' where the features used in the model are 'players', the SHAP strategy aims to calculate the values of the contributions of target features on prediction in the model. Then, the SHAP value $\varnothing_i(v)$ for each feature i can be denoted as:

$$\varnothing_i(v) = \sum_{S \subseteq N\backslash\{i\}} \frac{|S|!(n-|S|-1)!}{n!}(v(S\cup\{i\}) - v(s)) \tag{5}$$

Where N is the set of all features and n is the total number of features, S is a subset of features not

[7] GridSearchCV: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

including feature i, and v is the model function that gives the prediction for each subset of features. So, the $v(S \cup \{i\}) - v(s)$ represents the prediction changes after we include the new feature i in the model and $\frac{|S|!(n-|S|-1)!}{n!}$ represented the associated weight (i.e., marginal contribution). Then, $\sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!}$ is the weight by summing up the weights from all possible subsets S.

To summarise, the SHAP value for feature i represents the average contribution of feature i in the model's prediction when it is added to different subsets of features, weighted by the probability of each subset forming before feature i is added. Hence, an absolute SHAP value represents the magnitude or strength of the impact that a feature has on the model's prediction compared to the baseline prediction. For example, a larger feature SHAP value (which can be positive or negative) indicates that the feature is more important for contributing to the model (i.e., increases the model performance or makes the model more predictable) compared to other input features. Positive SHAP values for a feature i mean that the higher value of this feature contributes to increasing the dependent variable value in prediction (it implies that this type of feature obtains a positive impact on the model's predicted values). Conversely, negative SHAP values for a feature i indicate that the higher feature value in a given instance contributes to decreasing the model's dependent variable value in prediction (it implies that the feature obtains a negative impact on the model's predicted values).

5 Results

The nine types of AOIs were constructed from POIs for each LSOA of London through the methods described in Sect. 4.1. In total, 252,685 AOIs including the nine types representing places were generated across 4,835 LSOAs in London. Figure 10 in Appendix gives an example by depicting the map of generated AOIs for the six LSOAs within the City of London (as one local authority in Greater London). By aggregating and attaching stays detected from the mobile phone GPS trajectory data, the daily footfalls for AOIs were generated from 2020 to 2021 in London. Then place visitation variables were obtained by calculating the MDAF (monthly daily average footfalls) for each type of place (AOI) for each LSOA. This process results in MDAF as place visitation variables (18 categories in total) with nine place types separated for weekdays and weekends over 24 months and 4,835 LSOAs in London. In training, the XGBoost regressor, modelling demonstrated a maximum tree depth of 13, a learning rate of 0.04, and a column sub-sample ratio of 0.3 during tree construction. The best model performance metric achieved offered a RMSE value of 0.19 and an R^2 of 0.96 within the 19-month dataset from January 2020 to July 2021 (i.e., the training set). Next, during model testing, the performance of the trained XGBoost regressor was evaluated with a RMSE value of 0.54 and an R^2 of 0.73. These metrics were obtained by comparing the trained model's predicted theft levels with the actual data over the last five-months of the dataset from August 2021 to December 2021 (i.e., the testing set). Last, the SHAP strategy was employed to give more in-depth model explanation for the trained XGBoost regressor with application to both the training and testing set. The SHAP values were utilised to quantify the impacts of place visitation levels on theft levels both on a global and local scale. In particular, the local SHAP values allowed for examining the impacts of visitation at places on theft levels for one specific LSOA and month.

In the remainder of this section, Sect. 5.1 describes the global impacts of varying levels of population visitations in different categories of places on theft in the context of the changing restrictions or relaxation policies during the pandemic from 2020 to 2021. Subsequently, Sect. 5.2 delineates the local interpretability of the model by focusing on the impacts of the population's visitation at various places on theft levels at local LSOAs, especially focusing on the variations of these impacts in London LSOAs for different observation months relevant to pandemic policies.

5.1 Global impacts

In the trained 'Theft' XGBoost regressor, the coefficient of determination (R^2) from the training set (0.96) and the testing set (0.73) both show promising performance, as well as indicating the strong relationships between the explanatory variables (the MDAF of place visitations) and the response variables (theft levels) across London LSOAs. Further, the SHAP values quantify the contribution/impact of a particular variable (i.e., each type of place visitation variable) towards the prediction outputs (theft levels) across all months and LSOAs. Figure 3 shows the mean absolute SHAP values of different types of place visitations on weekdays/weekends, representing their impact on theft levels. The figure shows that the population's visitation levels at the venues categorised as 'Accommodation, eating and drinking' (AED) on weekdays and weekends have the top two impacts on theft levels with mean SHAP values of 0.038 and 0.030, respectively. Conversely, weekday visitations to the 'Education and health' (EH) venues appear to have the weakest impact on the theft levels. It is also of note that place visitations categorised under 'Retail' (RET) and 'Attractions' (ATT) on weekends and weekdays have generated a consistently high level of impact on the predictability of theft levels compared to other visitation categories with the SHAP values of 0.015 and 0.026.

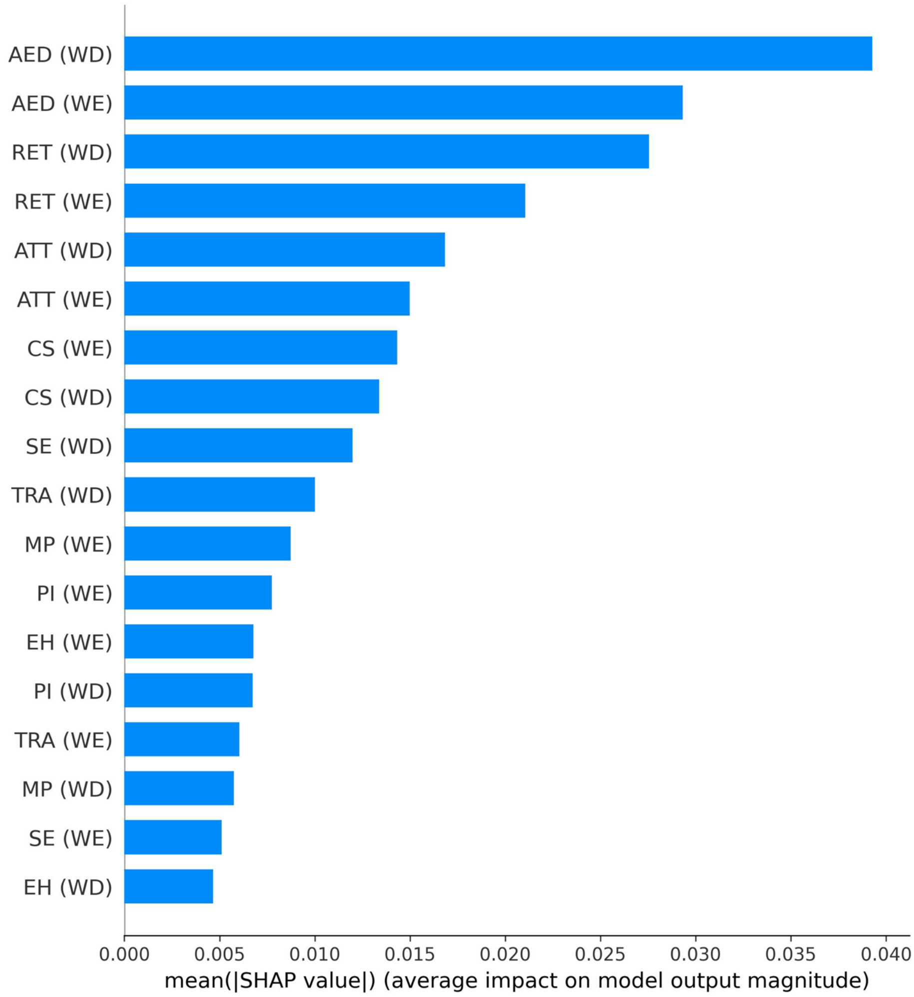

Fig. 3 The global impacts (measured by the mean absolute SHAP values) of different types of place visitations (on weekdays/weekends) on theft incident levels in London from 2020 to 2021. The full AOI names are 'Accommodation, eating, and drinking' (AED), 'Retail' (RET), 'Attractions' (ATT), 'Commercial services' (CS), 'Sport and entertainment' (SE), 'Transport' (TRA), 'Manufacturing and production' (MP), 'Education and health' (EH) and 'Public infrastructure' (PI)

To understand the impact of place visitation by type of place on theft, Fig. 4 plots the SHAP values of each place visitation variable (measured by MDAF) across all spatial and temporal units/grids, i.e., 4835 LSOAs × 24 months (represented by individual dots). The x-axis in the figure represents the influence (SHAP values) of place visitation variables on the theft counts. The positive and negative SHAP values (above or under 0) indicate whether the population visitations at a particular place type contribute to increasing or decreasing predicted theft levels. The legend on the y-axis represents the types of place visitation levels (i.e., the MDAF value levels) and is visually distinguished by the colour, with red signifying higher values and blue denoting lower values. Overall, across the units tested, increasing visitor population levels increase predicted theft levels – as most of the dots with high place visitation levels are to the right of the zero line – although the picture is more balanced for 'Transport' (TRA), 'Education and Health' (EH) and 'Manufacturing and Production' (MP). However, in several situations, compared to all observation samples (i.e., all LSOAs over 24 months), those units with higher levels of population visitations (indicated by red dots) at certain places also decreased theft levels (i.e., had negative impacts), such as 'Retail' on weekends (RET(WE)) and 'Attractions' on weekdays and weekends (ATT(WD) and ATT(WE)) (showing red dots to the left-hand side of the zero line). It can also be observed that the low-level population visitations (indicated by blue dots) are clustered around SHAP values

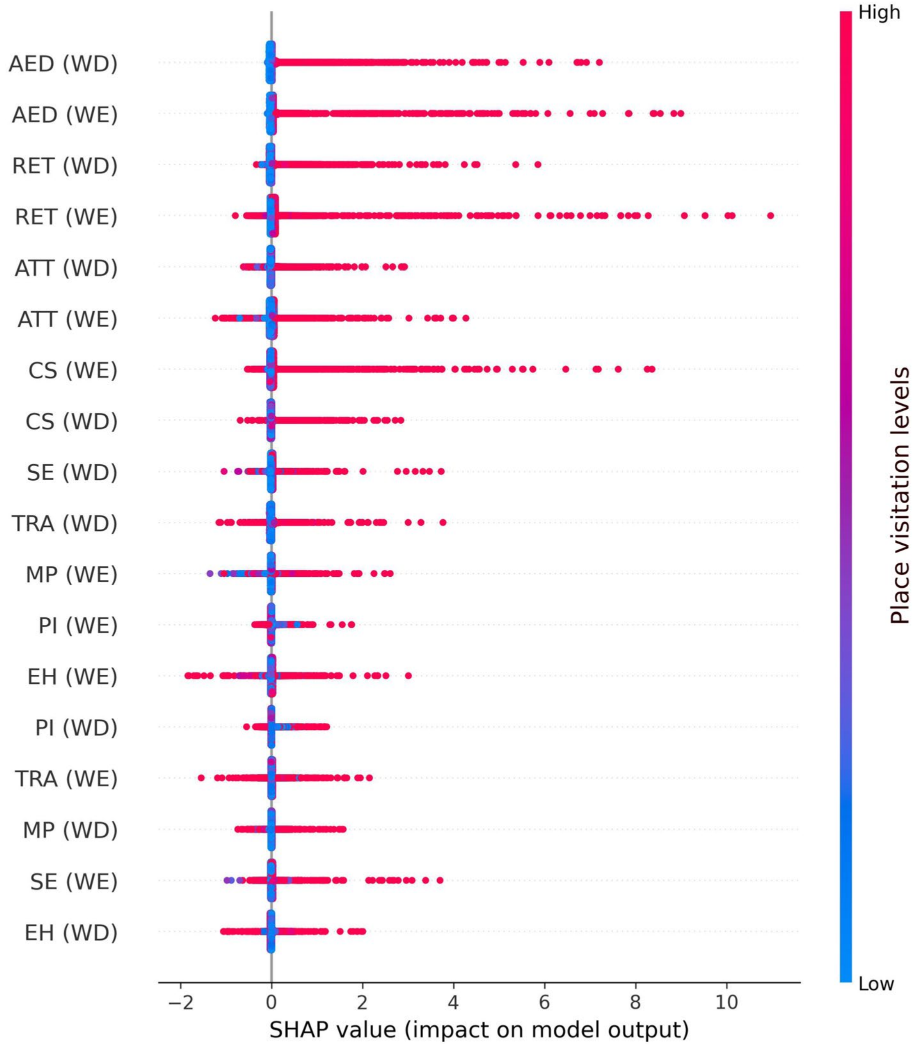

Fig. 4 The SHAP values of place visitation variables across all spatial and temporal units/grids (4835 LSOAs × 24 months in total). The legend represents the levels of place visitation variable values and is visually distinguished by the colour, with red signifying high values and blue denoting low values

of 0, which means limited place visitation levels have negligible impacts on theft levels.

To more fully understand the impacts of place visitation on theft incidents amid the pandemic restriction and relaxation policies, Figs. 5, 6 and 7 show the global impact of the population's visitation at three types of places ('Accommodation, eating and drinking', 'Retail' and 'Attractions') on theft levels in London during the two years by month. Here, the global impact is represented by the mean of absolute SHAP values of all London LSOAs by month within the two-year period. Three vertical grey lines denote the specific national lockdown months in the UK, including the 'First national lockdown' (from March 23, 2020 to June 23, 2020), 'Second national lockdown' (from November 5, 2020 to December 2, 2020) and 'Third national lockdown' (from January 6, 2021 to February 22, 2021). In general, it shows that the global impacts of population visitation at various places on theft significantly changed due to pandemic-related restrictions. The impact of place visitation on theft level started to rise in correspondence with the relaxation of pandemic constraints and declined when the restrictions were reintroduced. The figures illustrate that the global impact of the population's visitation at these three specific places (accommodation eating and drinking, retail and attractions) on theft reached the highest levels in February 2020 (before the first national lockdown) while declining to the lowest level in April 2020 during the first national lockdown. Furthermore,

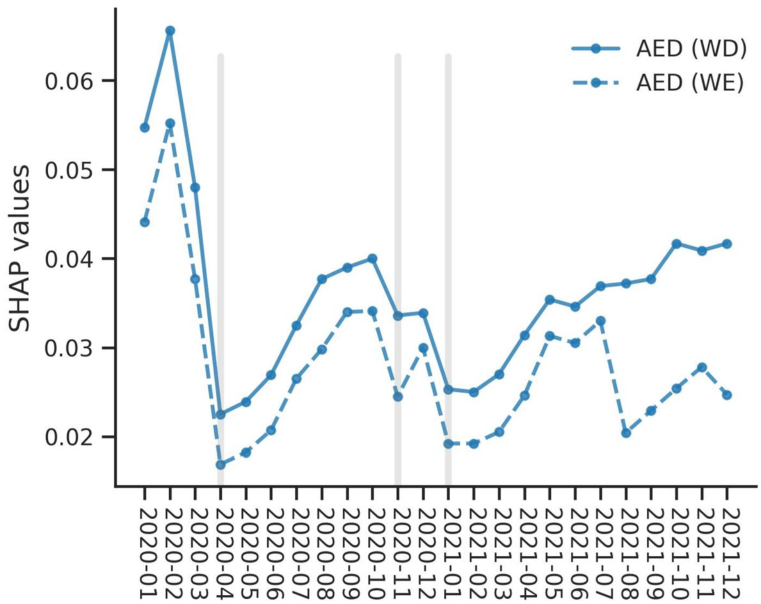

Fig. 5 The global impact (measured by the mean of absolute SHAP values of all LSOAs) of the population's visitation at 'Accommodation, eating and drinking' on theft outcomes of London from 2020 to 2021 (24 months). The visitation on weekdays (WD) and weekends (WE) are plotted as a solid line and a dashed line, respectively

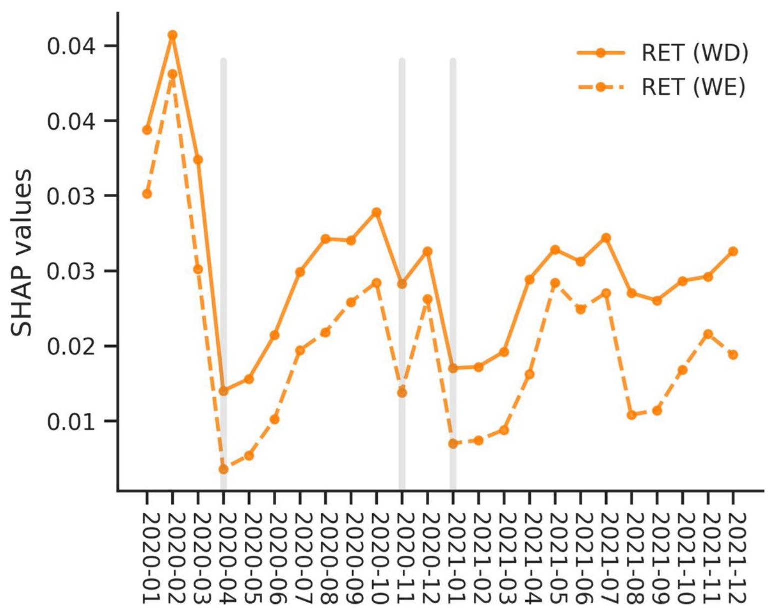

Fig. 6 The global impact (measured by the mean of absolute SHAP values of all LSOAs) of the population's visitation at 'Retail' on theft outcomes of London from 2020 to 2021 (24 months). The visitation on weekdays (WD) and weekends (WE) are plotted as a solid line and a dashed line, respectively

it indicates the reinforcements in the impacts of place visitation on crime levels during eased restriction periods following the lockdowns, specifically between May 2020 and September 2020 after the first national lockdown, and between February 2021 and May 2021 after the third national lockdown. Further, the visitation to these locations on weekdays led to a higher impact on theft compared to visits during weekends over the two years across all the months studied.

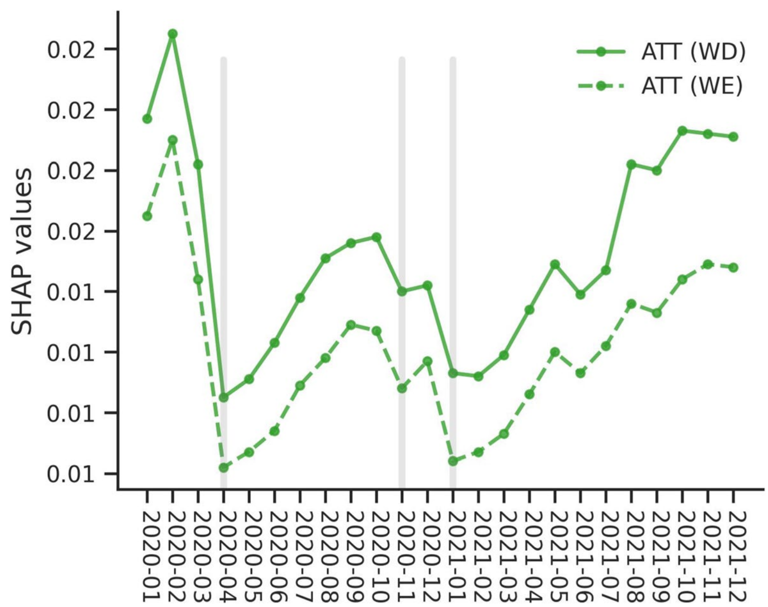

Fig. 7 The global impact (measured by the mean of absolute SHAP values of all LSOAs) of the population's visitation at 'Attractions' on theft outcomes of London from 2020 to 2021 (24 months). The visitation on weekdays (WD) and weekends (WE) are plotted as a solid line and a dashed line, respectively

5.2 Local impacts

The local impacts explore the contribution of each type of place visitation measured by SHAP values in the XGBoost model's output at each spatio-temporal unit/grid (i.e., per LSOA and month in this study). In this manner, the heterogeneous impacts of place visitations on theft incident levels can be captured across London LSOAs for different pandemic contexts.

To distinguish the varied impacts of place visitation on theft across London LSOAs, Figs. 8, 11 and 12 (see Figs. 11 and 12 in Appendix) highlight the LSOAs with the top 30 positive SHAP values for three types of place visitation ('Accommodation, eating and drinking', 'Retail' and 'Attractions') on weekdays and weekends in different pandemic contexts, respectively. The maps also indicate the boundaries between the LSOAs in inner London (1,901 LSOAs) and outer London (2,934 LSOAs). To clarify, these maps only depict the LSOAs which have seen positive and high-level impacts of selected place visitations on thefts (as indicated by extremely high SHAP values). For illustration, the top positive SHAP values in each map are broken into three distinct tiers (see the legends for reference). Furthermore, the four months represented in each map relate to unique months in pandemic restriction/relaxation policies (including February 2020 with the before lockdown, April 2020 with the first national lockdown, August 2020 with the first lockdown restriction eased, and November 2020 with the second national lockdown) to illustrate the evolving spatial patterns of the place visitations' influences on theft.

Figure 8 displays the spatial distributions of LSOAs in which theft levels are most affected by the place visitations at 'Accommodation, eating and drinking' venues during four unique pandemic circumstances. Notably, in normal times (i.e., before lockdown), the top SHAP values for visitation at 'Accommodation, eating and drinking' venues on weekdays are concentrated in the city centre of inner London. Following the government's implementation of social policy during the first national lockdown, it illustrates that the top SHAP values for 'Accommodation, eating and drinking' visitation on weekdays shifted and were observed in the northeastern outer London LSOAs. However, it is important to note that the SHAP values in the first national lockdown were lower than those before the lockdown. A similar pattern of displacement in the first national lockdown can be seen in the map for visitation on weekends (see the map of 'AED (WE) 2020-04') where the high SHAP values appear prominently in the northeastern LSOAs, and some new LSOAs with high SHAP values are observed in the western outer London. As the restrictions of the first national lockdowns started to ease in the summer of 2020, the place visitation levels increased theft (shown as the increased SHAP values) and re-clustered in the city centre.

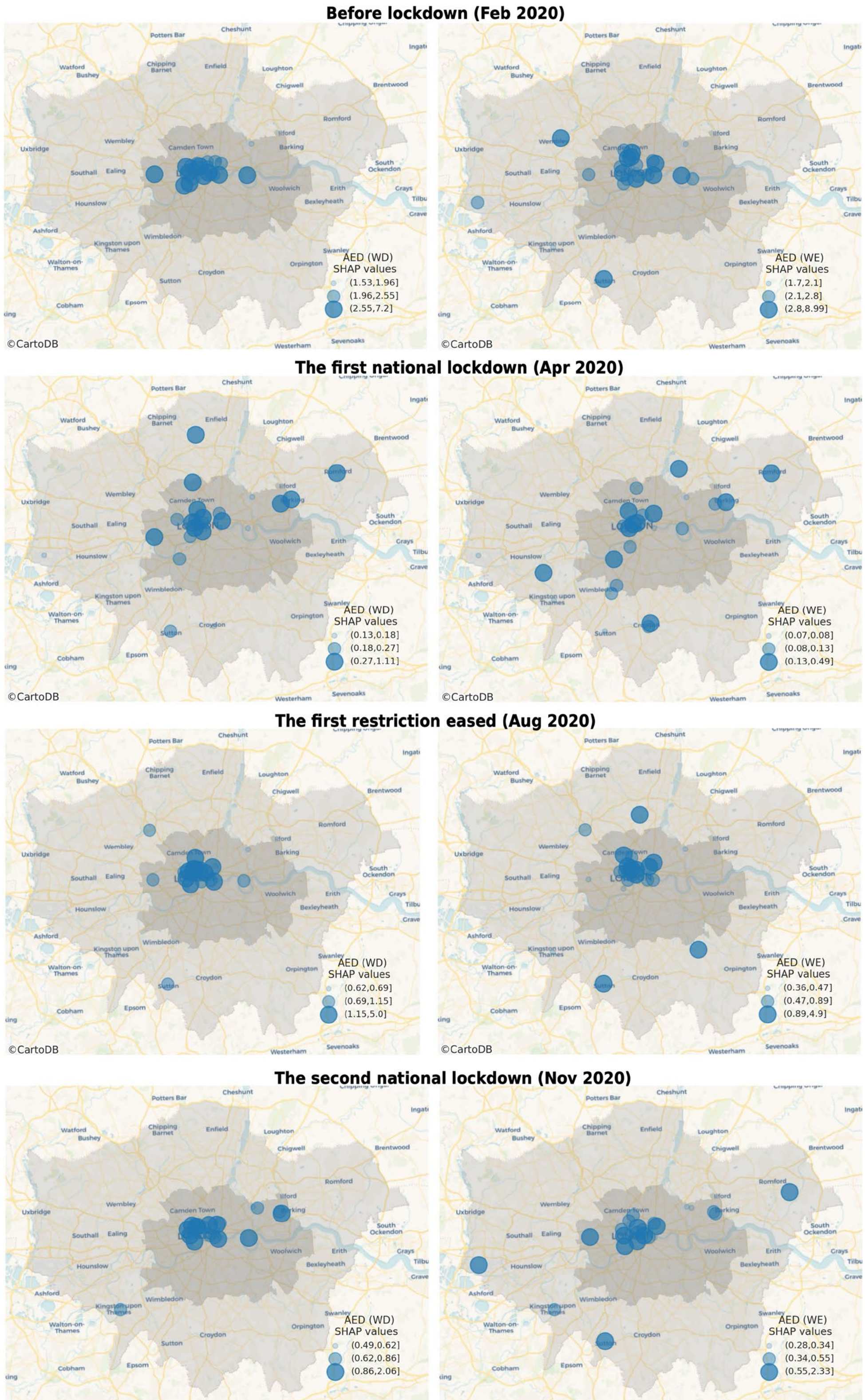

Fig. 8 The spatial distributions of top 30 positive SHAP values of the place visitation at 'Accommodation, eating and drinking' on weekdays (WD) and weekends (WE) in inner (the darker central area) and outer London (the lighter surrounding area). The markers represent the centroid of LSOAs and their size and transparency correspond to the magnitude of the top positive SHAP values, which have been grouped into three distinct levels for visualization purposes (as indicated in the legends)

Figures 11 and 12 in Appendix show the similar patterns as Fig. 8 for 'Retail' and 'Attraction' places, but denote nuances by type of place. For example, highly influential population visitations in driving theft levels for 'Retail' appear to be more widely distributed across all LSOAs during the periods examined, but visitations at 'Attraction' tend to have a greater impact on theft levels in inner London compared to outer London areas. This shows that there are differences both in terms of the spatial distribution of highly influential populations, but also in terms of how these changed over the pandemic period.

To further explore the varied impacts of place visitations in specific urban areas during the different pandemic contexts, Fig. 9 plots the place visitation's SHAP values for three selected LSOAs made up of distinct urban functional settings both before lockdown (Feb 2020) and during the first national lockdown (Apr 2020). The top left map of London shows the physical locations of the three LSOAs (A, B and C) selected from inner London and outer London. The top table indicates each LSOA's AOI numbers representing different urban functional contexts. LSOA A (coded as E01004765) is located

The AOI counts in selected LSOAs

	AED	RET	CS	ATT	TRA	PI	SE	EH	MP
A (E01004765)	72	76	238	46	14	39	17	49	18
B (E01001640)	17	12	12	10	12	11	5	6	0
C (E01001043)	79	55	165	10	26	44	15	31	11

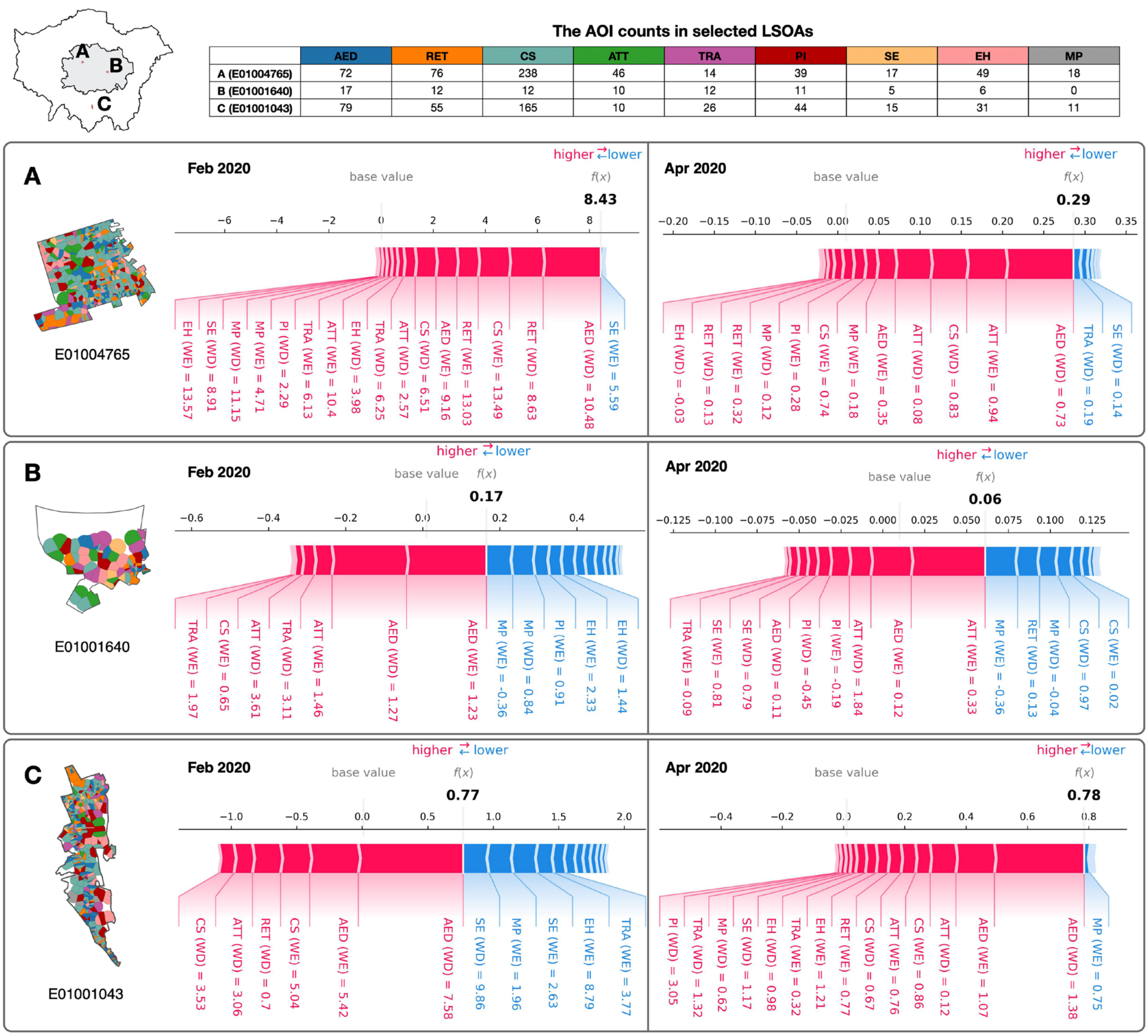

Fig. 9 The SHAP values of place visitation variables at three selected LSOAs before and during the first national lockdown months are illustrated as force plots. The length of the bars in the force figure denotes the impact extent (the magnitude of the SHAP values) of place visitation variables on theft levels with red specifying positive impacts (more visitations meaning more thefts) and blue specifying negative impacts (more visitations meaning less theft)

in a highly-developed commercial and entertainment district (traditionally known as the 'Oxford circus area' next to Soho) in the centre of London, and mainly contains 238 'Commercial services', 76 'Retail' and 72 'Accommodation, eating and drinking' sites. Unlike the high-density AOIs in LSOA A, LSOA B (coded as E01001640) is a typical residential and entertainment area in the Greenwich main district of inner London. This area includes 17 'Accommodation, eating and drinking', 12 'Retail', 12 'Commercial service' and 10 'Attractions' AOIs. Finally, LSOA C (coded as E01001043) is a part of the high street area of Croydon (in outer London) and includes 165 'Commercial services', 79 'Accommodation, eating and drinking' and 55 'Retail' sites.

Figure 9 contains large amounts of information, all of which demonstrates the nuances in data trends and hence we devote some time to interpretation here and provide some examples. In the subfigure of LSOA A, the two force plots illustrate the impacts of place visitation variables on theft levels in February 2020 (left) and April 2020 (right), respectively. The visitation at 'Sport and Entertainment' on weekends (labelled as SE (WE)) was the only type of place visitation that had a negative impact on theft in Feb 2020, but this was a weak influence. In addition, the values of place visitation variables (standardised MDAF) are plotted underneath the corresponding bars, so the MDAF for SE (WE) in LSOA A in Feb 2020 was 5.59. On the x-axis, the dark bold number represents the theft level value (standardised) predicted from the trained XGBoost regressor for selected LSOA and month (e.g., LSOA A and February 2020). The base value (0 as labelled in the x-axis) is the expected predicted theft level (i.e., standardised mean value) of all samples (24 months and 4,853 LSOAs). Hence, predicted theft levels in LSOA A were higher in Feb 2020 than in April (8.43 compared to 0.29) but also predicted levels in LSOA A were higher than those in other units like LSOA B.

Figure 9 also shows that the majority of the place visitation variables imposed positive impacts on theft before lockdown (Feb 2020), especially the visitation at 'Accommodation, eating and drinking' on weekdays obtaining the highest SHAP values than other places. Despite experiencing a significant drop in all place visitation levels (MDAF) in the first national lockdown (Apr 2020), the population visitations at 'Accommodation, eating and drinking' in LSOA A continued to influence thefts positively on weekdays (AED (WD) visitation level decreased from 10.48 to 0.73). However, the weekday visitation at 'Transportation' sites changed to having a negative impact on theft in LSOA A during the lockdown, from a positive impact during normal times (TRA (WD) visitation levels decreased from 6.25 to 0.19).

Although the decline in visitation to LSOA B was relatively smaller compared to LSOA A, the impact of 'Attractions' on weekends in LSOA B reached the highest level in the first national lockdown (April 2020). On the other hand, the influence of visitation to 'Accommodation, eating, and drinking' on weekdays (AED (WE)) dropped to the second position in April 2020, after holding the most influence in February 2020. LSOA C also experienced the shifts in place visitations' impacts on theft, with the impact of 'Attractions' on weekdays moving up to third place during the lockdown. Interestingly, the impact of 'Transportation' visitation on weekends shifted from a negative to a positive impact on the thefts between normal time and the lockdown periods (TRA (WE) visitation level reduced from 3.77 to 0.32). This appeared to be a more general trend in this LSOA, with a number of the visitation variables having a positive impact on theft numbers during the lockdown period.

6 Discussion

This study has examined the interplay between theft incidents and the population's place visitations sensed from geo big data across London's urban areas during various social changes with a particular emphasis on periods of the pandemic. The analytical methodologies used here have highlighted substantial disparities in the influence of dynamic place visitations on theft levels across different circumstances. Using SHAP values as a measure of the impacts/contributions of place visitation on theft levels, the results underscore the strong association between place visitations and theft levels. The results demonstrate the strong relationship across London LSOAs between 'Accommodation, eating, and drinking' places and levels of theft on weekdays in particular, with higher visitation typically leading to increased theft levels. However, the findings from local analyses indicate variations in the impact of visitation across London LSOAs – with areas where visits had a significant influence on theft shifting from inner London to outer London during the pandemic restriction periods. When closely examining specific urban areas (LSOAs), noticeable changes can be observed in the influences of visitation on theft at different place types under various pandemic contexts and demonstrate that there is no general local trend. This detailed research enables a more nuanced understanding of the dynamic influence of population activity related to different urban places on theft levels, particularly in a 'natural experiment' condition such as a pandemic.

The benefit of sensing population activities from big data is that crime opportunities can be dynamically assessed which allows for the analysis of the relationship between specific crime types, the type of places in which they occur, and their busyness across any defined urban areas and temporal periods. For example, theft is strongly linked to specific types of place visitation, particularly the 'Accommodation, eating, and drinking' sites in this study, where the measured high-volume population activity (footfalls) creates opportunities during normal (i.e., non-pandemic) times. The large-scale opportunities created by high foot traffic in accommodation, eating, and drinking venues attract thieves searching for suitable targets, such as the properties of customers in these places. Such target selection preference and the population creating opportunities turns these place locations into crime generators in certain urban areas based on crime pattern theory (Brantingham and Brantingham 2016).

The place visitation variables also capture the temporal variations that directly influence the availability of crime opportunities at specific types of location. This approach delineates the opportunities in micro-places characterized by variations in population activity across different temporal periods, such as distinct phases during the pandemic, or between weekend and weekday contexts. For example, in the case study, direct evidence shows a global decline in population activities at accommodation, eating and drinking, retail, and attraction venues, resulting in a noticeable reduction in the impact of these visitations on thefts during the implementation of pandemic restrictions. Conversely, the criminogenic influences arising from place visitations exhibit a rebound during periods of relaxation as social activities recover leading to increased crime opportunities. We can conclude that using dynamic variables representing variations in activity levels at place rather than purely using place categorisations offer advantage in understanding the interactions between places, populations and levels of crime."

Importantly, the pattern that the opportunities are not uniformly distributed across London LSOAs, which reflects the variations in population routine activity (e.g., mobility/movement patterns) across urban areas. These mobility behaviour changes reflect that some locations experienced a drastic reduction in mobility and/or impacts on theft during the pandemic while others did not. For instance, during the pandemic restrictions, the places where visitations have a relatively high impact on crime appear to have spatially shifted from inner to outer London. This shift might be due to the fact that commercial and entertainment areas in inner London no longer attracted large volumes of the population for dining, working or shopping compared to pre-pandemic levels. During this time, outer London residents seemed to exhibit activity more around their residential area due to limited mobility, such as reduced commuting and outdoor activities in compliance with pandemic restriction policies.

Another explanation is that theft levels influenced by place visitations are further modulated by local guardianship across various urban neighbourhoods. Under the reduction of crime opportunities alongside social activity restrictions (due to low levels of visitor population), some urban neighbourhoods with different guardianship levels showed varied capacities in their ability to protect their areas against crime (Andresen and Hodgkinson 2022; Campedelli et al. 2020; Chen et al. 2023a).

Additionally, the local interpretability findings obtained from the XGBoost regressor in this study also demonstrate that the variations in the population guardianship could be related to the local places within distinct contexts. The negative SHAP values may represent protective impacts of population visitation against theft crime with different levels as distinguished by urban functionality. For example, there are differences in the type of places delivering the negative (protective) influences on theft between the commercial and entertainment district (LSOA A) and the local residential areas (LSOA B) in normal (non-pandemic) times (see Fig. 9). Also, under certain contexts, the impacts of population visitation at the same type of place on theft would be different. For example, the impact of the population's visitation at 'Transportation' sites (on weekends) on thefts turned from negative (protective) to positive (promoting) during the first national lockdown.

Disentangling the intricacies of the dynamic population visitation's impact on theft incidents at diverse place settings within disparate social contexts can be instrumental in shaping the implications for crime prevention strategies. Predominantly, based upon the explainable machine learning strategy, these methods can not only pinpoint potential areas of elevated theft activity through sensing high-volume population visitation derived from geo big data but also dynamically identify the type of local place that has either a positive or negative influence on theft levels. Then, identifying the locations of theft occurrences significantly influenced by place visitation could serve as a tailored tool for resource allocation. This strategy would help plan the deployment of specific management resources at specific places, thereby offering a more targeted approach to intervention. Within the context of dynamic societal shifts (particularly health crises or natural hazards), understanding the influences of local visitation on criminal activities can help to calibrate routine preventative strategies (such as police patrolling) in alignment with changing requirements. The strategy adjustments should account not only for the spatial differences in how population activity influences crime but also for the specific context and surrounding environment of each urban area, as high-impact locations may shift between different types, as demonstrated by this case study. Moreover, it is crucial to align strategy modifications to consideration of the characteristics of each urban region, as the dominant impact tends to shift between different types of place visitations during times.

While this study provides valuable insights, there are limitations when interpreting the findings and designing future research. First, the geo big data employed (mobile phone GPS data) in this study contains a general bias, which can be related to the socio-economic conditions of the population groups sharing the data (Pappalardo et al. 2023). Second, the detection of place visitation behaviour from geo-big data also generates bias in this work. One bias arises from the parameters used in stay detection, as well as the heuristic parameters employed to infer home and work locations based on mobile phone GPS data. Another issue is the sensed place visitation levels are obtained on the basis of POIs which are likely to have different densities across urban areas, meaning that there will be subsequent variation in the size of the AOIs. Beyond visit frequency, other important factors such as the duration of stay are also critical components of visitation behaviour, but have not been examined in this study. The emphasis on visitation behaviours is also limited by the exclusion of the working population and residents from the analysis, though one could argue that it is important to avoid conflating these population groups. Third, this analysis does not include dynamic guardianship factors, such as police patrol records. The absence of this data limits our ability to fully track the interaction between guardianship and targets as theorised by the opportunity theories of crime. Fourth, while the study delineates spatial units into urban neighbourhood areas, the temporal units are aggregated into monthly levels due to police record data limitation. The monthly-level temporal resolution obstructs the development of crime prediction models that operate on a daily or hourly basis that could be facilitated by high-resolution geo big data. One of the barriers here is encountered in accessing corresponding crime incident data in high resolution in space and time from open UK police data sources. Another is the time and computing power cost of undertaking all possible forms of analysis on distinct forms of aggregated mobility data. Lastly, the study primarily analyses the interoperability from the trained machine learning models, i.e., named relationships or impacts, but it does not provide causal inferences. In the face of dynamic social contexts, such an approach needs careful verification in interpretation for policy development through incorporating an understanding of the local neighbourhoods to ensure significance.

7 Conclusions

In conclusion, this study has evaluated the impacts of the population's visitation to various places on theft by applying explainable machine learning techniques in the London LSOAs over the 2020-2021 period. The population visitations at places sensed from geo big data show a strong association with theft in the study area, particularly the population's visitation at 'Accommodation, eating, and drinking' on weekdays which incurs the highest (positive) impact on theft levels. Moreover, the impact of place visitations on crime can fluctuate depending on the circumstances during different pandemic phases with changes in influence across various types of place visitations at local urban areas. This study suggests the importance of considering the dynamic characteristics of population activity at places over time and in different contexts when developing targeted crime prevention strategies. Future research could focus on characterising place visitation behaviour and integrating more dynamic guardianship factors, which could help elucidate crime opportunities and their interplay with target populations at various places. The data and analysis required in this research make it computationally challenging, but an obvious further research pursuit would be to replicate this method in other countries and regions.

Appendix

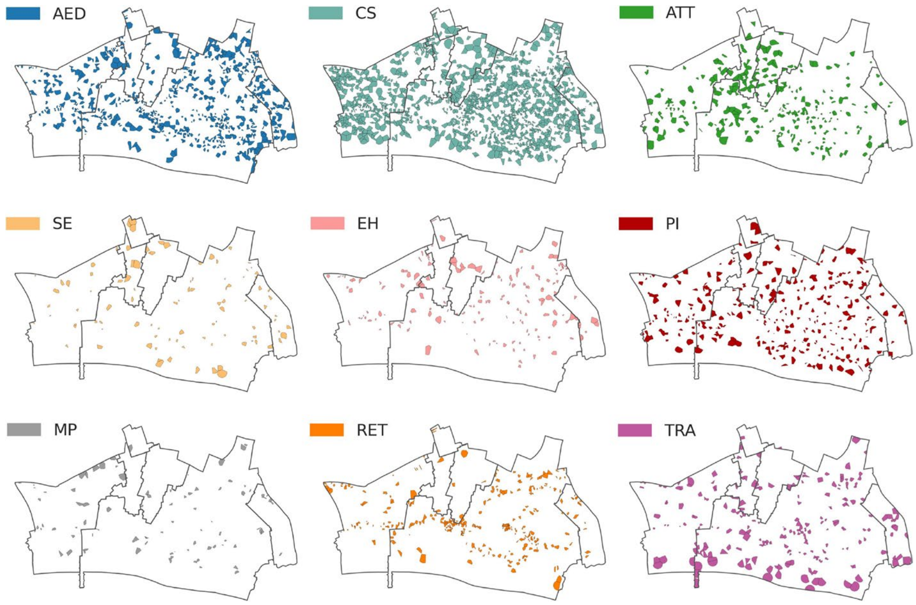

Fig. 10 The generated nine types of AOIs in City of London (including six LSOAs) are displayed. Each LSOA boundary is plotted by grey lines. Each AOI type is represented using a different colour in the plot. The full AOI names are 'Accommodation, eating, and drinking' (AED), 'Commercial services' (CS), 'Attractions' (ATT), 'Sport and entertainment' (SE), 'Education and health' (EH), 'Public infrastructure' (PI), 'Manufacturing and production' (MP), 'Retail' (RET) and 'Transport' (TRA). The distributions of commercial services and accommodation, eating and drinking exhibit denser patterns in the central areas of the City of London compared to other places. Additionally, some transportation sites are located near the southern boundary of the city, while other place types have not been observed as present in this region

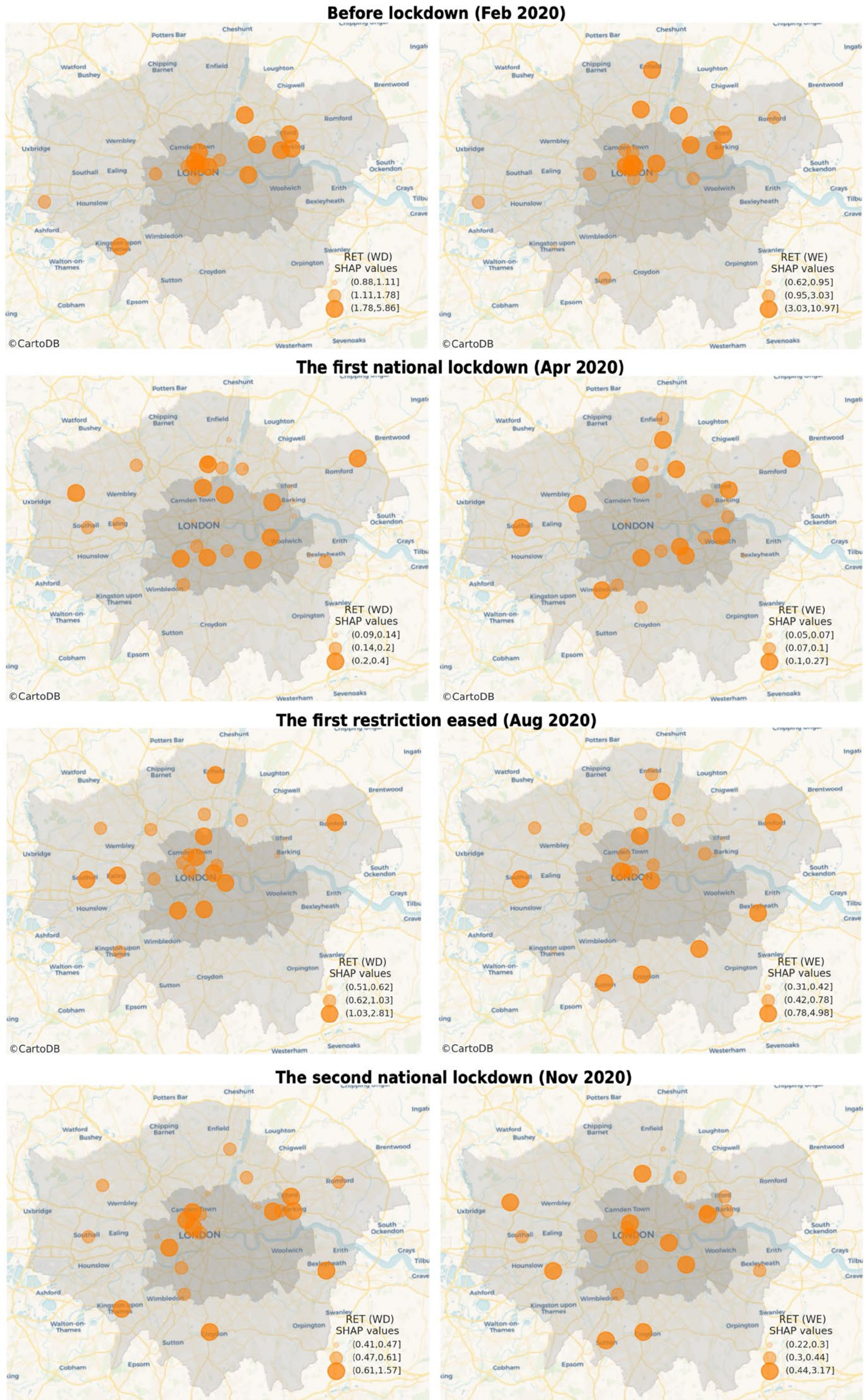

Fig. 11 The spatial distributions of top 30 positive SHAP values of the place visitation at 'Retail' on weekdays (WD) and weekends (WE) in inner (the darker central area) and outer London (the lighter surrounding area). The markers represent the centroid of LSOAs and their size and transparency correspond to the magnitude of the top positive SHAP values, which have been grouped into three distinct levels for visualization purposes (as indicated in the legends)

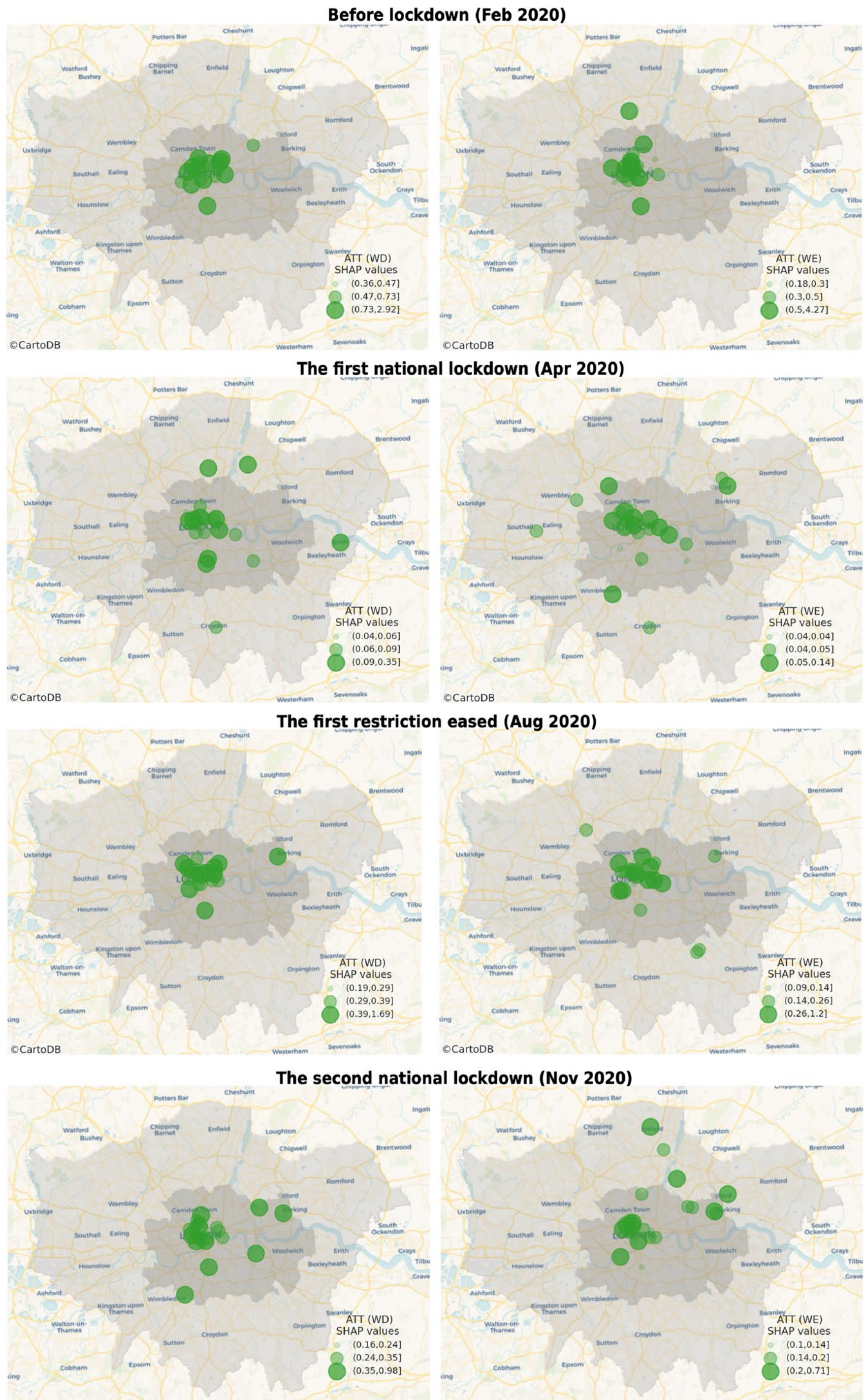

Fig. 12 The spatial distributions of top 30 positive SHAP values of the place visitation at 'Attractions' on weekdays (WD) and weekends (WE) in inner (the darker central area) and outer London (the lighter surrounding area). The markers represent the centroid of LSOAs and their size and transparency correspond to the magnitude of the top positive SHAP values, which have been grouped into three distinct levels for visualization purposes (as indicated in the legends)

Acknowledgements
The authors thank the editor and the anonymous reviewers for their comments to improve this work.

Authors' contributions
Conceptualization: Tongxin Chen, Kate Bowers and Tao Cheng; Methodology: Tongxin Chen; Formal analysis and investigation: Tongxin Chen; Writing - original draft preparation: Tongxin Chen; Writing - review and editing: Tongxin Chen, Kate Bowers and Tao Cheng; Funding acquisition: Kate Bowers and Tao Cheng; Resources: Kate Bowers and Tao Cheng; Supervision: Kate Bowers and Tao Cheng. All authors read and approved the final manuscript.

Funding
This research was partially supported by the U.K. Economic and Social Research Council Consumer Data Research Centre (CDRC) under Grant ES/L011840/1.
The second author is funded by the Economic and Social Research Council under the U.K. Research and Innovation open call on COVID-19 under Grant ES/V00445X/1.

Data availability
The data and materials in this work are available upon request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 4 December 2024 Revised: 14 May 2025 Accepted: 23 May 2025
Published online: 03 June 2025

References

Alves, L. G., Ribeiro, H. V., & Rodrigues, F. A. (2018). Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Applications, 505*, 435–443.
Andresen, M. A. (2011). The ambient population and crime analysis. *The Professional Geographer, 63*(2), 193–212.
Andresen, M. A., & Hodgkinson, T. (2022). In a world called catastrophe: the impact of COVID-19 on neighbourhood level crime in Vancouver, Canada. *Journal of experimental criminology*, *19*, 487–511.
Ashby, M. P. (2020). Changes in police calls for service during the early months of the 2020 coronavirus pandemic. *Policing: A Journal of Policy and Practice, 14*(4), 1054–1072.
Barnum, J. D., Caplan, J. M., Kennedy, L. W., & Piza, E. L. (2017). The crime kaleidoscope: A cross-jurisdictional analysis of place features and crime in three urban environments. *Applied Geography, 79*, 203–211.
Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology & Pub. Pol'y, 12*, 513.
Bernasco, W., & Block, R. (2011). Robberies in chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of research in crime and delinquency, 48*(1), 33–57.
Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan New York.
Brantingham, P. J., & Brantingham, P. L. (1995). Criminality of place: Crime generators and crime attractors. *European Journal of Criminal Policy and Research, 3*, 5–26.
Brantingham, P. J., & Brantingham, P. L. (2016). The geometry of crime and crime pattern theory. *Environmental criminology and crime analysis* (pp. 117–135). Routledge.
Brantingham, P. J., Brantingham, P. L., amp, et al. (1981). *Environmental criminology*. Sage Publications.
Campedelli, G. M. (2022). Explainable machine learning for predicting homicide clearance in the united states. *Journal of criminal justice, 79*, 101898.
Campedelli, G. M., Favarin, S., Aziani, A., & Piquero, A. R. (2020). Disentangling community-level changes in crime trends during the covid-19 pandemic in chicago. *Crime Science, 9*(1), 1–18.
Chalkiadakis, G., Elkind, E., & Wooldridge, M. (2011). *Computational aspects of cooperative game theory*. Morgan & Claypool Publishers.
Chen, J., Pei, T., Shaw, S.-L., Lu, F., Li, M., Cheng, S., Liu, X., & Zhang, H. (2018). Fine-grained prediction of urban population using mobile phone location data. *International Journal of Geographical Information Science, 32*(9), 1770–1786.
Chen, T., Bowers, K., & Cheng, T. (2023). Applying dynamic human activity to disentangle property crime patterns in london during the pandemic: An empirical analysis using geo-tagged big data. *ISPRS International Journal of Geo-Information, 12*(12), 488.
Chen, T., Bowers, K., Zhu, D., Gao, X., & Cheng, T. (2022). Spatio-temporal stratified associations between urban human activities and crime patterns: A case study in san francisco around the covid-19 stay-at-home mandate. *Computational Urban Science, 2*(1), 1–12.
Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). San Francisco. Association for Computing Machinery.
Chen, T., Zhu, D., Cheng, T., Gao, X., & Chen, H. (2023). Sensing dynamic human activity zones using geo-tagged big data in greater london, uk during the covid-19 pandemic. *PloS one, 18*(1), e0277913.
Chen, W., Zhou, Y., Wu, Q., Chen, G., Huang, X., & Yu, B. (2020). Urban building type mapping using geospatial data: A case study of beijing, china. *Remote Sensing, 12*(17), 2805.
Cheng, T., Chen, T., Liu, Y., Aldridge, R. W., Nguyen, V., Hayward, A. C., & Michie, S. (2022). Human mobility variations in response to restriction policies during the covid-19 pandemic: An analysis from the virus watch community cohort in england, uk. *Frontiers in Public Health*, *10*, 3617.
Cheung, L., & Gunby, P. (2022). Crime and mobility during the covid-19 lockdown: A preliminary empirical exploration. *New Zealand Economic Papers, 56*(1), 106–113.
Cohen, L. E. (1981). Modeling crime trends: A criminal opportunity perspective. *Journal of Research in Crime and Delinquency, 18*(1), 138–164.
Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American sociological review, 44*(4), 588–608.
Curman, A. S., Andresen, M. A., & Brantingham, P. J. (2015). Crime and place: A longitudinal examination of street segment patterns in vancouver, bc. *Journal of Quantitative Criminology, 31*, 127–147.
Eck, J. E., & Weisburd, D. (1995). *Crime and place* (vol. 4). Criminal Justice Press.
Felson, M. (2016). The routine activity approach. *Environmental criminology and crime analysis* (pp. 106–116). Routledge.
Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence, 14*(771–780), 1612.
Groff, E. R., & Lockwood, B. (2014). Criminogenic facilities and crime across street segments in philadelphia: Uncovering evidence about the spatial extent of facility influence. *Journal of Research in Crime and Delinquency, 51*(3), 277–314.
Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR), 51*(5), 1–42.
Haberman, C. P., & Ratcliffe, J. H. (2015). Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology, 53*(3), 457–483.
Halford, E., Dixon, A., Farrell, G., Malleson, N., & Tilley, N. (2020). Crime and coronavirus: Social distancing, lockdown, and the mobility elasticity of crime. *Crime science, 9*(1), 1–12.
Hanaoka, K. (2018). New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users' locations. *Environment and Planning B: Urban Analytics and City Science, 45*(2), 295–311.
Hariharan, R., spsampsps Toyama, K. (2004). Project lachesis: parsing and modeling location histories. In *International Conference on Geographic Information Science* (pp. 106–124). Springer.
He, L., Páez, A., Jiao, J., An, P., Lu, C., Mao, W., & Long, D. (2020). Ambient population and larceny-theft: A spatial analysis using mobile phone data. *ISPRS International Journal of Geo-Information, 9*(6), 342.
Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., Jr., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems, 53*, 36–46.
Johnson, P., Andresen, M. A., & Malleson, N. (2021). Cell towers and the ambient population: A spatial analysis of disaggregated property crime. *European Journal on Criminal Policy and Research, 27*(3), 313–333.

Kontokosta, C. E., & Johnson, N. (2017). Urban phenology: Toward a real-time census of the city using wi-fi data. *Computers, Environment and Urban Systems, 64*, 144–153.
Kucukpehlivan, T., Cetin, M., Aksoy, T., Kurkcuoglu, M. A. S., Cabuk, S. N., Pekkan, O. I., Dabanli, A., & Cabuk, A. (2023). Determination of the impacts of urban-planning of the urban land area using gis hotspot analysis. *Computers and Electronics in Agriculture, 210*, 107935.
Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PloS one, 9*(6), e96180.
Li, X. Y., Sinniah, G. K., & Li, R. (2022). Identify impacting factor for urban rail ridership from built environment spatial heterogeneity. *Case Studies on Transport Policy, 10*(2), 1159–1171.
Liu, L., Xia, B., Wu, H., Zhao, J., Peng, Z., & Yu, Y. (2019). Delimitating the natural city with points of interests based on service area and maximum entropy method. *Entropy, 21*(5), 458.
Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*, 4765-4774.
Malleson, N., & Andresen, M. A. (2015). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science, 42*(2), 112–121.
Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E., Uchida, C. D., & Brantingham, P. J. (2020). Impact of social distancing during covid-19 pandemic on crime in los angeles and indianapolis. *Journal of criminal justice, 68*, 101692.
Niu, H., & Silva, E. A. (2023). Understanding temporal and spatial patterns of urban activities across demographic groups through geotagged social media data. *Computers, Environment and Urban Systems, 100*, 101934.
Pappalardo, L., Manley, E., Sekara, V., & Alessandretti, L. (2023). Future directions in human mobility science. *Nature computational science, 3*(7), 588–600.
Pappalardo, L., Rinzivillo, S., & Simini, F. (2016). Human mobility modelling: Exploration and preferential return meet the gravity model. *Procedia Computer Science, 83*, 934–939.
Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R. (2019). scikit-mobility: A python library for the analysis, generation and risk assessment of mobility data. arXiv preprint arXiv:1907.07062.
Rummens, A., Hardyns, W., & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied geography, 86*, 255–261.
Schnell, C., Grossman, L., & Braga, A. A. (2019). The routine activities of violent crime places: A retrospective case-control study of crime opportunities on street segments. *Journal of criminal justice, 60*, 140–153.
Shaw, S.-L., Tsou, M.-H., & Ye, X. (2016). Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science, 30*(9), 1687–1693.
Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27–56.
Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. *Journal of Quantitative Criminology, 35*(4), 831–854.
Song, G., Liu, L., Bernasco, W., Xiao, L., Zhou, S., & Liao, W. (2018). Testing indicators of risk populations for theft from the person across space and time: The significance of mobility and outdoor activity. *Annals of the American Association of Geographers, 108*(5), 1370–1388.
Stickle, B., & Felson, M. (2020). Crime rates in a pandemic: The largest criminological experiment in history. *American Journal of Criminal Justice, 45*(4), 525–536.
Sulis, P., Manley, E., Zhong, C., & Batty, M. (2018). Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science, 2018*(16), 137–162.
Tompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2015). Uk open source crime data: Accuracy and possibilities for research. *Cartography and geographic information science, 42*(2), 97–111.
Traunmueller, M. W., Johnson, N., Malik, A., & Kontokosta, C. E. (2018). Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems, 72*, 4–12.
Verma, R., Mittal, S., Lei, Z., Chen, X., & Ukkusuri, S. V. (2024). Comparison of home detection algorithms using smartphone gps data. *EPJ Data Science, 13*(1), 6.
Wu, C., Ye, X., Ren, F., & Du, Q. (2018). Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in shenzhen, china. *Cities, 77*, 104–116.
Yan, L., Wang, D., Zhang, S., & Xie, D. (2019). Evaluating the multi-scale patterns of jobs-residence balance and commuting time-cost using cellular signaling data: A case study in shanghai. *Transportation, 46*, 777–792.
Zahnow, R., & Corcoran, J. (2021). Crime and bus stops: An examination using transit smart card and crime data. *Environment and Planning B: Urban Analytics and City Science, 48*(4), 706–723.
Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. *Computers, Environment and Urban Systems, 94*, 101789.
Zhao, F., Ghorpade, A., Pereira, F. C., Zegras, C., & Ben-Akiva, M. (2015). Stop detection in smartphone-based travel surveys. *Transportation research procedia, 11*, 218–226.
Zhao, K., Tarkoma, S., Liu, S., and Vo, H. (2016). Urban human mobility data mining: an overview. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1911–1920). IEEE.
Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology (TIST), 6*(3), 1–41.

Publisher's Note