Allocating London Police Forces Through Predictive Policing To Reduce Burglary Rates

Tobias Hijdra[1774492]**, David Mandado Loureiro**[1978276]**, Alicja Napieralska**[1896024]**,**
Rik Olde Bijvank[1963287], **Melissa Selamet**[1921495], **Tenzing Sherpa-Gregory**[1834274],
Group 38

Abstract

While burglary being one of the most critical crime-related issues in London, the goal of our project is to distribute police resources effectively, enabling them to improve community safety, reduce burglary rates, and rebuild public trust in the police. To achieve this goal, we have developed a predictive model using XGBoost to forecast burglaries for the next month. These predictions are used to create an agenda for police officers, allowing them to know which region to patrol. To determine if our methods are rebuilding public trust, we have created a feedback form which will allow the police force to get insights on several topics. Our predictions are roughly consistent with actual values, giving promising results that our model will be able to help reduce burglaries in London.

1 Introduction

All the technical content covered in this project can be viewed at our GitHub repository, as well as the initial GitHub repository we migrated from because of blocked history for several members in case of wanting to see more activity.

Predictive policing has been a topic of frequent discussion, as decision-makers need to find a balance between public safety and ethics. One side argues that using data to, for example, find hotspots in crime is beneficial for society and should be incorporated into the police force. Others think data about the citizens should not be used in predictive models since it might lead to biases, such as over-policing certain areas. While both sides have valid arguments, burglary rates in London have been a considerable concern for the police force and the public. Since large open datasets are available, predictive models could be trained to help reduce these rates. Although using this data to create accurate and interpretable predictions is challenging, the recent development of machine learning models, this task has become increasingly achievable. To not neglect the ethical side of this discussion, it is also important to gain details about public opinion. This could be in the form of satisfaction levels, perceived safety, or at a baseline incorporation of societal factors. It will give the police force insights into if this form of policing is making a difference and has a positive social impact. This project seeks to answer a central research question: **How can we give recommendations for allocating the London police force such that burglary rates decline?** To support this, we explore four different sub-questions that will be addressed through our work:
1. How can we predict the distribution of burglaries across London?
2. Which social factors are highly associated with burglaries in London?
3. How can we involve the community in crime prediction by integrating their safety perceptions?
4. How does distribution of burglaries influence the number of police officers sent?

This report gives the methodology and results of our research. Several datasets have been used and details will be provided for what is included in the predictive model. We have tested several machine learning models, and XGBoost has shown the best results, as seen later in the paper. Our predictions are based on an LSOA level for the next month, allowing for the allocation of police forces on the same granularity. We will explain how the allocation of police officers is calculated in later sections.

Every feature was ranked using the MoSCoW method, which meant that the effort put into each item would match how important it was. *Must* (Core Features) are the essential components that couldn't be compromised on included LSOA-level burglary prediction, a ward-level police-allocation tool, and a resident survey with feedback visualization. These parts form the minimum viable system that connects actual data to practical deployment decisions and community feedback. Should (High-

Priority) are components such as dataset uploads, time-sliders, feature-importance for transparency, sentiment scoring, combined survey analytics, and special-operation logic. These additions would make the insights deeper and more transparent, transforming the basic tools into something that could be used practically on a daily basis. *Could* (Nice-to-Haves) features such as a shift calendar visualization, mobile-friendly interface, charts that could be exported, extra models, and feeding survey data back into predictions. Each of these would improve how easy the system was to use or make it more analytically powerful, but they were not essential for the initial launch. *Won't* (Out of Scope) are aspects that were deliberately left out, like predictions for a specific day, duties beyond burglary, including payed or private datasets and risk scores for individuals. Leaving these out helped avoid certain ethical issues, and data access problems, keeping the project manageable and on track.

2 Background

Burglary remains a persistent issue in London, both in terms of its societal impact, as well as the demand it places on limited police resources. According to the Office for National Statistics (Office for National Statistics, 2025a), only around 4.9% of residential burglary lead to charges in England and Wales and over 70% of investigations close due to no suspect. London continuously has some of the highest burglary rates among of the cities in the United Kingdom, adding to the pressure on local police forces. In response to this growing demand, predictive policing systems have started being used as a method to predict where and when this crimes might happen (Loya et al., 2024). People who support this technology claim that predictive policing can actually stop crimes before they take place, with studies showing crime reductions of up to 7.4 percent when compared to regular patrol methods, with results that are statistically more efficient (Mohler et al., 2015). However, these technologies are not without their issues, as predictive-policing tool raise important concerns about invasive surveillance, biased algorithms, and to which extend law enforcement agencies can be held responsible for their actions (Ferguson, 2017).

The foundation for our predictive methods comes from extensive research that demonstrates how burglary and similar crimes do not happen randomly across time and location, but instead show patterns that are stable enough to allow for short-term predictions (Andresen and Malleson, 2011), (Weisburd and Foundation, 2008). Moreover, tree-based models that use gradient boosting, such as XGBoost and LightGBM, along with ensemble approaches like Random Forests, have been consistently recognized as some of the best-performing for prediction problems as they can capture complicated, non-linear patterns and work effectively with diverse kinds of input features (Caruana and Niculescu-Mizil, 2016). Additionally, recent studies have indicated that publicly available, large-scale datasets offer a clear and expandable basis for developing these prediction models (Povala et al., 2019). In particular, incorporating demographic features such as income deprivation and housing density has been shown to significantly improve burglary forecasts at the LSOA level (Povala et al., 2020).

Community surveys consistently reveal that residents' perceptions of safety often diverge from what is captured by official crime records, influenced more by emotional and cultural factors, or by high-profile local incidents, rather than by actual crime rates (Blumstein and Nakamura, 2009). This gap between perception and reality indicates that we need to combine statistical crime predictions with insights from the community itself if police departments want to keep their credibility and maintain public trust. Finally, over-policing can reduce public cooperation and long-term compliance, ultimately undermining crime control (Berg and Huebner, 2011), which eventually weakens crime prevention efforts (Berg and Huebner, 2011). On the other hand, community-based interventions that focus on building collective effectiveness and local participation have consistently proven more successful at reducing property crimes than strategies that only react to crimes after they happen through enforcement (Nubani et al., 2023).

3 Methodology

3.1 Dataset Construction & Preprocessing

Auxiliary Datasets

To explore and ultimately predict burglary trends across London to answer **RQ 1**, we began by gathering publicly available datasets. Our primary data source was from Police.uk which had all accounts for past crime incidents in the United Kingdom (UK Police Data). This dataset provided monthly-level information on reported crimes, including the crime type, location, and date of report.

We then searched for other reliable datasets to enhance our predictions with different factors that might be affecting burglary rates. Once again, we obtained the stop-and-search dataset from the Police.uk website. This allowed us to track monthly counts of stop-and-search incidents per LSOA. On top of this, we added the latest available Index of Multiple Deprivation (IMD) dataset (Greater London Authority, 2019), which was from 2019, which provided decile rankings for each LSOA across various deprivation domains. Furthermore, we included population estimates from the Office for National Statistics (ONS) to account for population-driven crime exposure (Office for National Statistics, 2025b). Finally, to support geospatial analysis and visualization, we used GeoJSON boundary files for both LSOAs and London wards (Greater London Authority, 2025).

While collecting supporting data, we noted that many of these other datasets had 2019 reports as the most recent versions. Therefore, we used January 2019 as the starting point for our dataset to ensure consistency. Ultimately, the dataset spans a period of 75 months (2019 January - 2025 March), enabling us to model seasonal variation, long-term trends, and potential lagged effects of social and economic conditions on burglary rates.

Pre-processing and Feature Engineering

We applied different pre-processing and feature engineering steps for all the respective datasets we chose.

Crime Dataset Preprocessing Our primary dataset was monthly crime data at the incident level. To begin, we filtered the dataset to include only records with valid timestamps, LSOA codes, and non-negative burglary counts. Then, we changed the date column into a datetime format and aggregated monthly by LSOA. As our predictive task is burglary forecast, we computed the total burglary count per LSOA-month, which became the target variable for our model.

Stop-and-Search Mapping The stop-and-search dataset included longitude and latitude coordinates. These were spatially joined to their corresponding LSOAs using official GeoJSON boundary files to align stop-and-search activity with the LSOA-level granularity of our target. Counts were aggregated monthly per LSOA, and missing months were forward-filled where necessary to preserve continuity for time-series modelling.

Socioeconomic Features We incorporated the 2019 Index of Multiple Deprivation (IMD), which includes decile-level scores across areas such as income, employment, and crime. These deciles were added as static attributes for each LSOA. Since IMD is released infrequently, we treated these as time-invariant over our 2019–2025 time period. To validate this approach, we compared 2015 and 2019 IMD deciles and found that over half of LSOAs remained in the same decile, with very few of them changing by more than one. This confirmed that IMD rankings are generally stable over time and still provide meaningful contextual information when combined with dynamic features.

Population Estimates Population estimates were collected from the Office for National Statistics. These were also assigned as static features per LSOA to normalize crime counts and derive per capita crime rates. Since population data was not available monthly, we used the most recent data available and held them constant over time.

Geo-spatial Consistencies For geospatial considerations, spatial boundaries for LSOAs were consistent throughout the study period. Ward boundaries were not consistent. We evaluated different years of ward boundaries, by calculating the percentage of area of each LSOA in a ward, to find the year where LSOAs fit within wards best. Next, aligning LSOAs to ward boundaries required some preprocessing due to a few mismatches. These edge cases were manually solved, ensuring accurate aggregation and alignment across geographic levels.

Feature Engineering To capture short- and long-term temporal patterns, we created lag features for burglary counts (1, 2, 3, 6 and 12 months) and also added rolling statistics (mean, standard deviation, sum) over the same time frames. Seasonal features, such as the month since, cosine transforms, and quarter indicators were also added to create the cyclical variation. We also introduced derived indicators such as the time since the last burglary event for each LSOA and entropy scores based on the distribution of other crime types.

Interaction Terms To reflect social dynamics, we created interaction features combining temporal, demographic, and spatial data. For example, we multiplied IMD deciles with the engineered feature with time-since-burglary to expose non-linear relationships that might be missed by additive features alone.

3.2 Modeling Approach

To predict monthly burglary counts across London's LSOAs, we approached the task as a supervised regression problem. Our main goal was to estimate the number of burglary counts per LSOA in any given month. Our target value, therefore, was "burglary_count", a value that we had calculated from our primary crime dataset.

Given the structured nature of our dataset and taking into account the importance of capturing non-linear relationships and temporal dependencies, we selected XGBoost as our predictive model. XGBoost is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It works efficiently well with tabular, large datasets, like the one that we have created.

We specifically configured the XGBoost algorithm for regression since we wanted to be able to predict the actual burglary rates per LSOA and hence allocate our police according to this predicted value. In the process of choosing the most suitable model, we also looked into working with LightGBM and XGBoost with different objectives, like Tweedie regression and Poisson regression.

Ultimately, we found that using Optuna, which is a hyperparameter optimization framework, significantly improved the performance of our XGBoost regression model. Optuna allowed us to automatically search for the best combination of model parameters (such as learning rate, maximum depth, and regularization terms) through efficient sampling strategies. We worked with The final model, trained with 1000 estimators and 100 optimization trials, achieved the strongest results across the training and test evaluation sets, as can be seen in Table 1.

Model	MAE	RMSE	R^2
XGBoost Regression	0.131	0.431	0.764
XGBoost Tweedie	0.112	0.453	0.740
XGBoost Poisson	0.124	0.472	0.717
LightGBM Tweedie	0.117	0.447	0.757
LightGBM Poisson	0.303	0.704	0.097

Table 1: Performance of different models on the test set using MAE, RMSE, and R^2 as evaluation metrics.

Workings of XGBoost and its parameters XGBoost is a gradient boosting machine learning model. It creates many decision trees, where each tree is trained to predict the negative gradient of the loss function with respect to the current model's predictions. For some loss functions, such as MSE, this negative gradient can also be seen as the residual error. Each new tree is added to the model, and the learning rate hyperparameter controls how much it contributes to the final prediction. The updated prediction equals the previous prediction plus the multiplication of the learning rate and the new tree's output. It does this as many times as the n_estimator hyperparameter specifies. We found that the most important hyperparameters to tune included the before mentioned, but also max_depth, subsample, and colsample_bytree. Max depth controls the depth of the trees and should not be set too high, as the model might overfit, but also not too low, as it will not be accurate. Subsample and colsample_bytree set the fraction of training data and features, respectively, randomly sampled for each tree. The former helps prevent overfitting, and lower values provide more regularization, while the latter prevents the trees from relying too heavily on the same features. Since our dataset has many features, we do not want the model to only rely on a few to make its predictions. Both of these metrics help XGBoost generalize its predictions without sacrificing too much accuracy.

3.2.1 Performance Metrics and Validation Strategy

To evaluate the model's performance fairly, we split our dataset temporally into three different splits:

- **Training data:** All LSOA-level monthly data from **2019 to 2022**.
- **Validation data:** January–December **2023**. This validation set was also used for hyperparameter tuning with Optuna.
- **Test set:** Starting from **January 2024** and onwards. This split was also used to report final performance metrics shown in Table **??**.

As can be seen in Table 1, we used three different performance metrics to see clarity, accuracy, and interpretability.

Mean Absolute Error (MAE): calculates the average magnitude of the errors between our model's predictions and the actual observed values. The absolute values are taken over this value to ensure that the direction does not impact the overall measure of accuracy (ScienceDirect Topics, 2024a).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Root Mean Squared Error (RMSE): measures the square root of the average of the squared differences between the predicted values from the model output and the actual observed values. We have also used our RMSE as our optimization criterion when using Optuna (Optuna Team, 2024). We did this because RMSE penalizes larger prediction errors more strongly compared to MAE. RMSE can be considered as smooth and continuous, ensuring a stable convergence during Optuna hyperparameter tuning. Finally, it is also widely used in regression and is interpretable in the same units as burglary counts (ScienceDirect Topics, 2024b).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Explained Variance (R^2): measures the proportion of variability in the dependent variable that our model can explain. It is especially essential for demonstrating the model's effectiveness:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

3.2.2 Feature Importance

This section directly answers **RQ 3**.

We visualized the importance of features using XGBoost's accuracy metrics to improve model transparency and ensure trust in model outputs. This helped us identify which variables most influenced burglary risk predictions. By looking at Figure 1, we can see that the most influential feature is the months since the last burglary. This shows that recent incidents are a strong signal of future risk. Next is IMD x months-since-burglary, an interaction capturing how deprivation levels amplify crime persistence. This also helps us validate the value of our engineered features. Finally, we also see several rolling averages and lagged counts, which reflect temporal trends—burglary patterns that evolve over months. Interestingly, 'other theft' is also among the top features, suggesting broader crime patterns may share underlying drivers with burglary.

3.3 Community Dataset

To examine how community perceptions could be incorporated into predictive crime modeling, a synthetic dataset of survey responses was created at the LSOA level. Edge-case areas were manually chosen to show deliberately differences between actual crime rates and how safe people felt, so that LSOAs with low crime but where people felt unsafe, and vice versa. For each LSOA, between two and five synthetic responses were generated, including answers to nine Likert-scale questions (on a 1-5 scale) and one open-ended question. In the edge-case areas, responses were fixed to extremes (all high or all low), while in other areas responses were randomly generated to show natural variation.

The open-text responses were sourced from Reddit from subreddits about UK and London: "london," "UKCrime," "AskUK," and "BritishProblems". Posts that mentioned London and relevant topics like burglary, policing, or housing issues were filtered, and these were used as responses for the open-ended survey question. This synthetic survey data was then processed and analyzed using the Community Tool that will be described in the following section.

4 Solution Design & Implementation

4.1 Predictive Model

After finalizing our XGBoost regression model and explaining the technical information regarding it in the Methodology phase, we will now discuss how we have integrated our model into a live pipeline that works directly with the police tool.

The model is directly embedded in our website (https://policeallocation.live) in the dashboard, which is accessible to the police officers. As new monthly data are released for the Metropolitan Police Force on the Police.uk website, officers can upload new monthly crime data in a CSV format through an admin interface. Once uploaded, new data appended to the master dataset and goes through the preprocessing. In doing so, new variables and interaction terms are created automatically. Finally, updated predictions for each LSOA are generated for the upcoming month, and these predictions are reflected in the police allocation scheduling tool (See Appendix 2).

4.2 Community Feedback Tool

The Community Feedback Tool was creaed to adress **RQ 3** and is composed of two components, two pages: Safety Overview and Resident Survey (See Appendix 4, 5, 6 and 7).

4.2.1 Safety Overview

The Safety Overview dashboard includes two interactive visualizations (See Appendix 4 and 5): a line chart to display temporal patterns and a bar chart to illustrate the spatial distribution of burglaries. From a dropdown menu at the top, a user can choose a desired ward, and the data on the visuals is updated accordingly. The line chart shows burglary trends for the selected ward over a past year, excluding three most recent months, with the ward's count in a solid orange line and the London mean (across all wards) in a dashed blue line. Using both distinct colors and line styles improve readibility when the two series overlap, allowing residents to clearly see how their neighbourhood's crime trend compares to the city average. Below, a bar chart displays burglary counts for each Lower Layer Super Output Area (LSOA) within the selected ward. By presenting every LSOA side by side, users can see the distribution of burglaries and identify local hotspots, gaining a more granular overview of where burglaries concentrate. As properties remain vulnerable for up to two months post-burglary (Bowers and Johnson, 2004), data from the most recent three months is excluded from both visualizations to protect active investigations and resident safety.

The aim of the Safety Overview is to provide actual, data-driven insights into burglary risk. Based on this information, residents can provide informed feedback in the survey described in the following subsection, so that community input is grounded in real crime data.

4.2.2 Resident Survey

The Resident Survey is a core component for gathering feedback on the police allocation strategy (See Appendix 6 and 7), combining quantitative and qualitative feedback to assess how well current deployments align with community concerns.

Seven quantitative questions on a 1–5 Likert scale to calculate, what is referred in this paper as a "perceived safety" score, focused on how citizens feel about burglary safety in their LSOA. As "safety" can be interpreted differently by each respondent, the targeted questions were created to narrow the scope and generate a comparable perceived-safety value for every LSOA-month tuple, which then would be contrast with XGBoost burglary-risk predictions, described in details in the following section.

An eighth open-ended question allows residents to share broader concerns more freely. While all survey responses would be delivered to police officers, for a more comprehensive overview, the open answers are also grouped into predetermined topics, that were identified through literature research as correlated with increased burglary, such as: empty homes (holidays), lighting/CCTV, vacant properties, drug hotspots, police visibility/response times, reporting (ease/barriers), and community watch.

Then, BERT model was used to identify and cluser exact words that would fall within each of the topics. BERT was chosen (Devlin et al., 2019), as it is trained on social media data, which makes it well-suited for interpreting informal, nuanced open surevy resposnses, more effectively than frequency models like Word2Vec or TF-IDF. Subsequently, sentiment model was run on each topic to assign a mean sentiment score. NLTK sentiment model was chosen for this task, as not only it is an industry standard, but also it is a light and quick model making it suitable to run an analysis after collecting survey responses at the end of each month. Topics with more negative sentiment are displayed as higher priority for officers, as detailed later.

4.3 Police Tool

The police tool was created to answer **RQ 4** and combines the outputs from both the predictive XGBoost model and the community survey to provide a resource allocation and interface for the law enforcement.

4.3.1 Choropleth Map

Aimed to provide police officers with spatial understanding by depicting burglary data in its actual geographical context, the main visualization of the police tool is a choropleth map of London divided by ward boundaries. This map displays *Past Data* (historical burglary incidents), *Predicted Data* (outputs of the XG Boost model for the next month) and *Predicted-vs-perceived Data* (See Appendix 9, 10 and 3). In the Filter options, users can additionally select a time range via a *Slider* to view patterns over years or specific periods.

The tool also supports continuous updates: Users click the *Upload Data* button to add new crime records (e.g., monthly police reports), which are automatically appended to the existing CSV files. After uploading, the model can be refreshed via the *Rerun Model* button, updating XGBoost predictions so allocation recommendations remain current. In the default view, users may click the *Down-*

load Schedule button to obtain officer deployment plans across all wards.

When a specific ward is selected—by choosing it in the *Ward Dropdown*—the display splits into two panels. The left panel retains the full London ward map, while the right panel zooms into that ward's LSOAs, showing the distribution of burglaries within it (See Appendix 11). This side-by-side presentation removes reliance on short-term memory, allowing officers to see overall ward counts in the London context alongside detailed LSOA breakdowns. Deeper red shading indicates higher burglary incidence. Since LSOAs are small administrative units (approximately 1,500 residents) and local officers are familiar with their areas, this visualization pinpoints the exact streets or locations that require focused patrols.

4.3.2 Incorporating Community Feedback

On the same map, users can click the *Predicted Data* button to display the gap between perceived safety and predicted burglary. This is calculated by comparing the safety scores from the quantitative survey questions with the model's predicted burglary counts, both of which are normalized for comparison on one scale. Each area is indexed was $i = 1, 2, \ldots, n$, where n is the total number of LSOAs. For each area i:

$P_i \in \mathbb{Z}_{\geq 0}$	Raw predicted burglary count
$S_i \in [1, 5]$	Mean perceived safety score

Since these two measures are on different scales, we use min-max normalization, to translate both values to the range $[0, 1]$ for comparison. The predicted burglary count is normalized:

$$\overline{P}_i = \frac{P_i - P_{\min}}{P_{\max} - P_{\min}},$$

where $P_{\min}$ and $P_{\max}$ are the minimum and maximum predicted values across all areas. Analogically, the perceived safety score is normalized:

$$\overline{S}_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}},$$

where S_i is a survey score, and $S_{\min}$ and $S_{\max}$ are the bounds of the Likert scale, in our case 1 and 5. The perceived vs predicted safety gap G_i is then calculated as the difference between the normalized predicted risk and the normalized perceived safety:

$$G_i = \overline{P}_i - \overline{S}_i.$$

Thus, LSOAs where the predicted burglary count vastly exceeds the perceived score could point to underreported crime or policing blind spots, whereas LSOAs where the perceived score is much larger than the predicted count, suggesting people feel less safe than the model predicts, might show heightened community anxiety skewing perceptions.

Alongside the choropleth map, users can click the *Perception Analysis* tab (See Appendix 13), which reveals a diverging bar chart of predefined topics. The topic displayed farthest to the left—bearing the most negative sentiment score—signals the community's greatest concern or disappointment. Police officers can then integrate this insight with other data sources, such as incident reports on that topic, and take targeted action. The view also presents mean sentiment scores for each topic across all wards or time periods, offering a high-level overview of how community feelings evolve and where attention is most needed.

4.3.3 Scheduling

Below the choropleth map, the tool displays a police allocation calendar based on the constraint of 100 officers available per ward (See Appendix 11 and 12). Officer deployment occurs in two-hour slots between 06:00 and 22:00, with each officer limited to two hours per day for up to four days per week. The allocation methodology operates on a proportional basis, by distributing officers according to the relative burglary risk across LSOAs within each ward. The system calculates the total predicted burglaries for all LSOAs within a ward, then allocates officer resources proportionally based on each LSOA's percentage contribution to the total risk. As a result, higher-risk areas get more police presence while still covering all areas, as the system also has a baseline allocation of one shift per week for areas with zero predicted burglaries, to account for potential model errors and allow for basic coverage and police presence everywhere. Additionally, the system takes into account that more burglaries happen during afternoon and evening hours, so it randomly skews allocation toward evening time slots to increase officer presence during higher-risk timeslots.

To make sure we utilise all tools available for our project, our project comes with a research-based guideline for the police to follow regarding picking special operation dates. As publicly available data does not allow for an exact time and location of any burgalries, instead of givng a specific number o police officers for allocation, we instead decied to crete aguildenes based on the analysis of the data avaale to use through publcily avaibale datasets and literature resarch. It also contains a method to adhere to for the actual police allocation, given a chosen date within the provided guideline. There is a 7 percent increase in property crime in a borough hosting a home Premier League football game for every 10,000 fans (Marie, 2010). There are seven London-based Premier League teams, three of them are outliers in terms of average attendance (Arsenal, Tottenham Hotspur, and West Ham United). Our guideline informs the Police to pick a home game of one of the teams above. Moreover, they should pick one date in December (peak burglary season), prioritise weekend evening kick-off (higher attendance), and focus on London derbies (more intense and loaded match). All of these factors increase the Police displacement, thus encouraging burglars in that borough. When a special operation date has been picked according to these guidelines, the Police should randomly choose 5 out of the 10 wards furthest away from the stadium (within the same borough) to patrol extensively during the football match. This ensures burglars will not discover any pattern. We have created lists of target wards, which we will provide to the police.

5 Analysis of Results

5.1 Predictive Accuracy (RQ 1)

Table 1 depicts that although LightGBM-Tweedie delivers the lowest MAE (0.117), XGBoost Regression posts the best overall balance asit has the highest R² (0.764) and the lowest RMSE (0.431). An MAE of 0.131 means our predictions deviate, on average, by roughly one burglary per eight LSOA-months, which is an acceptable resolution for monthly patrol planning. The RMSE of 0.431 confirms that large errors are rare; this matters because we tuned the model precisely on that metric. Capturing 76% of variance (R²) shows the model explains most of the month-to-month fluctuation in burglaries despite the crime's inherent noise. Figure Figure 2 shows the model's predictions track real burglary counts closely (correlation 0.88) and stay accurate in the most frequent 0–5 burglaries-per-month range, meaning it can be used for planning monthly patrols.

5.2 Most Impactful Features (RQ 2)

Importance scores show that most impactful on a burlarly prediction features are: months_since_burglary which flags the repeat-victimisation window, indicating that patrols should remain in areas where a burglary just occured; imd2019_x_msb which links deprivation and stop-and-search to higher risk, which signalls that already deprived and policed areas are at higher risk of burgalries; rolling_mean_3 reflects the model's reliance on a 3-month average over monthly counts, possibly due to repeated victimization and the need to smooth random spikes for steadier patrol planning. This findings build a basis for suggestions for improvements in police allocations in future work.

5.3 Community Perception Alignment (RQ 3)

Our current Community Tool insight implementation Figure Figure 3in the appendix relies on synthetic citizen responses, so it serves purely as a proof-of-concept and cannot be yet used to make any claims. However, the predicted-perceived safety map based on the survey answers proved to be a good way to display data as it allows to discover geographic patterns that raw numbers miss, allowing the police to not only rely in a model with 0.76 R^2, but also have information on which they can act upon at their own judgement. Additionally it allowed to create a pipeline for topic and sentiment models, for now, validating that this is technically possible.

5.4 Patrol Allocation Efficacy (RQ 4)

We simulated one month of patrols using 2 principles: *Random Baseline* (minimum uniform random shifts for each LSOA) and *Model-guided* (our proportional allocation). Following previous research and this project's interest in community satisfaction, a minimum patrol time per LSOA was determined, the remaining officers (the majority) where assigned proportionally inside each ward. Although it is not possible to test effectiveness, research is evidence enough that a balance between minimum allocation and proportional focus on hotspots can leverage the police resources to protect those LSOAs with highest burglaries, while not neglecting the ones with a lower prediction.

5.5 Limitations & Future Work

1. **Static socioeconomic features.** IMD deciles and population are fixed at 2019/2021 values,

so fast-changing neighbourhoods can be misclassified. Next step: refresh those fields annually and expose a simple "update-CSV" button in the admin panel for this specific data, which can be processed into the dataset as our other upload data function.

2. **Synthetic perception data.** Because we lacked time and resources for real survey feedback, the sentiment layer is based on Reddit text and manually created edge-cases. Next step: Conduct the survey and share through several means to obtain real citizen answers. If that would prove not possible, then at least conducting an interview study that would validate our idea and approach.

3. **Evaluation limited to one metric set.** We report MAE/RMSE/R^2 city-wide, which can hide ward-level bias, which is an issue as the social aspect of our analysis and potential bias toward more deprived areas would compromise this goal. Next step: compute ward-specific data to identify wards whose error derives the most from the median.

4. **Patrol schedule is purely proportional.** Patrols are currently assigned strictly in proportion to predicted burglary share, and one-off events such as football matches are covered separately by Special Operations units; the next release should let planners tag an event date, venue and scale, then automatically blend that footprint with baseline risk, recalculate patrol weights, and display the outcome on richer visuals—e.g., layered heat-maps, event buffers and officer-load charts—for a more nuanced, data-driven allocation.

These enhancements would not only strengthen predictive accuracy but also improve fairness, interpretability, and real-life effectiveness, aligning with continuous- improvement principles in the ALGO-CARE framework. Although we can not test real results from our tools, our progress, research and resources all indicate burglary could be reduced.

6 Ethical analysis

To ensure our tool meets UK ethical, legal, and operational standards, we assessed it using the ALGO-CARE framework, a UK-developed guide for responsible algorithm use in policing (Marion Oswald, 2017). This section maps our design choices to its core principles.

Our model is designed solely to support, not replace, human decision making. The community tool is the only aspect of this project that is open to the public. The police tool, including the predictive system, is restricted to internal police use only where its output informs patrol planning, but does not make binding decisions. This tool is deployed solely as a decision support system. Human officers retain complete discretion over final patrol plans and can adjust allocations, incorporate community feedback, and contextual knowledge, such as local events or intelligence, to supplement model output. As part of the tool, a community tool serves as a mechanism for feedback through a survey that captures resident perceptions of safety and police presence. Ensuring that the model serves as one input within a broader community-informed framework. These mechanisms safeguard against over-reliance and support the advisory nature of the system.

All data sources used in this project are publicly available and anonymized at most to the LSOA level, ensuring that no personal identifiers are involved. This is the same standard that the London police use when publishing their data (UK Police Data). We followed data minimisation and necessity principles by focusing on features with relevance to burglary prediction, such as stop-and-search counts, IMD scores, and population data. This is in line with the police code of practice (College of policing code of practice). The decision to exclude the most recent three months of data from the public-facing dashboard further reflects our commitment to operational security and the protection of active investigations.

Model outputs are generated at the LSOA level, which is a statistical unit designed to preserve anonymity while offering operational specificity (Park, 2021). This level of granularity is allows for precise patrol planning without infringing on individual privacy. The model uses multiple years worth of historical data, enhanced with temporal lags, seasonal features, and derived indicators to ensure contextual accuracy. Model performance is monitored over time, and data is refreshed as often as the police release data, with exclusions applied to recent months for the public tool to protect investigations and ensure ethical compliance (Kate Bowers, 2005).

As for ownership: the model, preprocessing pipeline, and dashboard interface is open-source and publicly hosted on GitHub. This gives any stakeholders, such as police departments, researchers, or overseeing entities the ability to audit or modify our tool. We encourage them to inspect its operation and verify its compliance with ethical and technical standards. By making use of only publicly available as data, source code, and preprocessing logic, we attempt to address concerns around black-box proprietary policing tools (Rudin, 2019).

We invite our stakeholders to challenge any decision made while developing this tool. For the public, we have given them direct access to both the code and the ability to give feedback to the police via open questions. For further transparency on how to access our code, a read me file has been included. The police force can inspect feature importance for each predictive variable via SHAP visualizations. Combining feature importance and community feedback can help identify discrepancies that can be acted on. This aligns with recommendations from the UK Centre for Data Ethics and Innovation, which emphasizes public involvement and transparency in AI oversight (Review into bias in algorithmic decision-making).

As for the accuracy of our tool, XGBoost was selected for its well-documented prediction abilities (Tianqi Chen, 2016). Model performance was validated using RMSE, MAE and R^2, ensuring alignment with the purpose of the tool: to forecast burglary counts in London. These are reflective of prediction ability and are standard in academic evaluation of forecasting models (Rob J Hyndman, 2021). We mitigated risk of false negatives is through baseline patrol assignments, while false positives are managed through human oversight.

Our tool represents a responsible balance between predictive models and human policing. By ensuring a minimum patrol presence for all areas, integrating community feedback, and excluding recent data from public visualizations, we minimize risks of over or under policing. The model is deployed transparently, can be updated regularly, and is made openly available. These measures align with the Information Commissioner's Office's AI auditing framework (Guidance on AI and data protection) and findings on community-based legitimacy in predictive policing (Kristian Lum, 2016). Ultimately, the tool's design supports both crime prevention and trust-building, addressing dual mandates of operational effectiveness and ethical policing.

To minimize the black-box effect and maximize explainability, we show the importance of each predictive variable. In addition, the public-facing components translate historical data into intuitive visuals, allowing residents to understand and respond to the broader trends shaping burglaries in their areas.

7 Reflection

Ethical safeguards guided every design decision; technical convenience came second. We used only fully anonymised, open data, which meant that we had to opt out from employing street-level heat-maps and individual-level risk scores. LSOA granularity was the smallest unit that protected privacy while remaining operationally useful for officers, who we assumed know their neighbourhoods.

We also removed sensitive predictors—such as ethnicity and detailed stop-and-search outcomes—to avoid reinforcing historical bias, which in turn decreased our accuracy. Similarly, we decided to forecast monthly burglary counts rather than daily events because the public dataset lacks precise time-stamps, and interpolating them would give a false sense of precision, intruding additional level of potential uncertainty.

Deployment followed the same logic. A minimum-patrol rule allowed for a baseline coverage in every LSOA, offsetting the model's tendency to ignore "cold" areas, but also meant that we had to allocate more officers overall, allocating almost all 100 officers available each month. For the Community Tool burglary data is displayed with a 3 month lag as well and shown only as bar and line charts, not interactive map, to avoid tipping off potential offenders or exposing possible repeat-victimisation hotspots. Moreover, although we aimed at framing each topic as well as the words that fall within it in a neutral manner, it should be noted that certain topics can have more negative sentiment be default (e.g., "Drug Hotspots" vs "Empty Houses/Away from the city") which should be kept in mind during evaluation. However, topics still were an effective manner to summarize survey response for more actionable feedback, so this limitation did not ultimately eliminated this feature entirely. Additionally, the entire codebase is open-source, inviting external feedback, transparency and preventing "black-box" policing.

Several tensions remain: city-wide error metrics can hide ward-level bias, confidence intervals are absent, and our community-perception layer is synthetic. Future work will prioritise fairness diagnostics, probabilistic outputs, and real community surveys—keeping ethics the primary driver rather than an afterthought.

8 Group Dynamics and Efficiency

Beginning of the project

When we started this project, our team managed to work well together even though we were initially a bit confused about where to start. Right from the beginning, we all knew we wanted to bring social elements into our work, and pretty quickly two main ideas came up: working with socioeconomic datasets and getting community feedback. These became pur collective main goals that influenced our entire project and that everyone on the team worked on throughout the whole project.

Early on, we noticed at what each team member was good at and what they struggled with, which really helped us later when we had to divide up tasks. Some people on our team were stronger at programming, while others felt more comfortable dealing with machine learning models or handling big datasets. Since we had four data science students and two computer science students, we ended up with a nice balance of skills in data analysis, coding, and organizing workflows. This early recognition of our different strengths helped with splitting up the work, even when we were still quite unsure about who should be doing what exactly. As initially we were having trouble with coordination, we decided to rotate who would chair meetings and take notes each week, so that each person at some point would be responsible for managing and overseeing the whole work being done.

Interim Presentation

Getting ready for the intermediate presentation became a major turning point in how we functioned as a team. Since that was the first moment hen we really had to bring all our individual work together into one presentation, it made us think more about the overall project and how our separate pieces connected with each other. By creating initial demos and rough versions of what our final solution would look like, we started to develop a shared understanding of where the project overall was heading.

This phase helped us move away from just working on our own tasks separately to actually collaborating as a real team. The intermediate presentation also helped us understand how to communicate and work together more effectively moving forward, like being more realistic about the time we can devote to the project and informing the others when the work was done or halted because one of us encountered an issue.

We were strategic about using people's strengths and addressing weaknesses. Since some team members were not comfortable with presenting, we decided they can handle the intermediate presentation, as at that point the ideas were still developing and did not need detailed technical explanations yet. It gave them an opportunity to get more experience and feel more confident about public speaking in a situation that was not as high-pressure.

There was also really good collaboration between the computer science and data science students. The CS students usually brought organized approaches to figuring out what tasks to prioritize and how to handle programming, while the DS students walked us through how machine learning models, statistics, and visualization tools worked. By helping each other out like this, we could fill in the gaps in our knowledge and create a more unified way of tackling problems.

Final Presentation

Once the intermediate presentation was over, we had a much better sense of how to work together. People naturally fell into roles as we learned more about each other's strengths, what interested us, and how we liked to work.

We also started working together, in parallel on the tasks that required more attention, like when our model was not performing to the standard we all worked on it individually but at the same time, trying out different solutions, so we could bring fresh perspectives on improvements. By the time we reached the final stages of the project, our group had become much more efficient. We had developed a better workflow clearer task divisions and most importantly the exact scope of what each task meant in practice. At the end, we are satisfied with what we achieved and how we developed as a group throughout the project. We are all very different people with different backgrounds, courses taken, and objectives, but we managed to use this diversity to our advantage as everyone contributed their unique skills to deliver the final result.

References

Martin A. Andresen and Noel Malleson. 2011. Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency*, 48(1):58–82.

Mark T. Berg and Beth M. Huebner. 2011. Reentry and the ties that bind: An examination of social ties, employment, and recidivism. *Justice Quarterly*, 28(2):382–410.

Alfred Blumstein and Kiminori Nakamura. 2009. Redemption in the presence of widespread criminal background checks. *Criminology*, 47(2):327–359.

Kate J. Bowers and Shane D. Johnson. 2004. Domestic burglary repeats and space-time clusters. *European Journal of Criminology*, 2(1):67–92.

Rich Caruana and Alexandru Niculescu-Mizil. 2016. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pages 161–168.

College of policing code of practice. 2023. [link].

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Andrew Guthrie Ferguson. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press, New York, NY.

Greater London Authority. 2019. Indices of deprivation (2019).

Greater London Authority. 2025. Lsoa atlas.

Guidance on AI and data protection. 2023. Guidance on ai and data protection.

Shane D Johnson Kate Bowers. 2005. Domestic burglary repeats and space-time clusters: The dimensions of risk. *European Journal of Criminology 2(1)*, 1(2).

William Isaac Kristian Lum. 2016. To predict and serve? *Significance*, 13:14–19.

Alberto Loya, Anastasios Noulas, and John Shawe-Taylor. 2024. Xai for trees: Interpretable forecasting for burglary crimes with explainable tree-based models. `https://papers.ssrn.com/abstract=4917220`. SSRN working paper.

Olivier Marie. 2010. Police and thieves in the stadium: Measuring the (multiple) effects of football matches on crime. CEP Discussion Paper 1012, Centre for Economic Performance, London School of Economics and Political Science.

Sheena Urwin Marion Oswald. 2017. Written evidence submitted by marion oswald, senior fellow in law and director of the centre for information rights, university of winchester, and sheena urwin, head of criminal justice, durham constabulary (alg0030).

George O. Mohler, Matthew B. Short, Steven Malinowski, Michael E. Johnson, George E. Tita, Andrea L. Bertozzi, and Paul J. Brantingham. 2015. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399–1411.

Linda Nubani, Heather Fierke-Gmazel, Holly Madill, and Anthony De Biasi. 2023. Community engagement in crime reduction strategies: A tale of three cities. *Journal of Participatory Research Methods*, 4(1).

Office for National Statistics. 2025a. Crime in england and wales: Appendix tables. Release date: 24 Apr 2025.

Office for National Statistics. 2025b. Lsoa mid-year population estimates - national statistics.

Optuna Team. 2024. Optuna: A hyperparameter optimization framework.

Neil Park. 2021. Population estimates by output areas, electoral, health and other geographies, england and wales: mid-2020.

Joona Povala, Sami Virtanen, and Mark Girolami. 2019. Burglary in london: Insights from statistical heterogeneous spatial point processes. arXiv preprint arXiv:1910.05212.

Joona Povala, Sami Virtanen, and Mark Girolami. 2020. Burglary in london: Insights from statistical heterogeneous spatial point processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5):1067–1090.

Review into bias in algorithmic decision-making. 2020. Review into bias in algorithmic decision-making.

George Athanasopoulos Rob J Hyndman. 2021. *Forecasting: Principles and Practice (3rd ed)*. Monash University.

Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*.

ScienceDirect Topics. 2024a. Mean absolute error.

ScienceDirect Topics. 2024b. Root mean square error.

Carlos Guestrin Tianqi Chen. 2016. Xgboost: A scalable tree boosting system. *KDD'16*, pages 785–794.

UK Police Data. 2025. Data downloads.

David Weisburd and Police Foundation. 2008. Ideas in american policing. in place-based policing.

Appendix

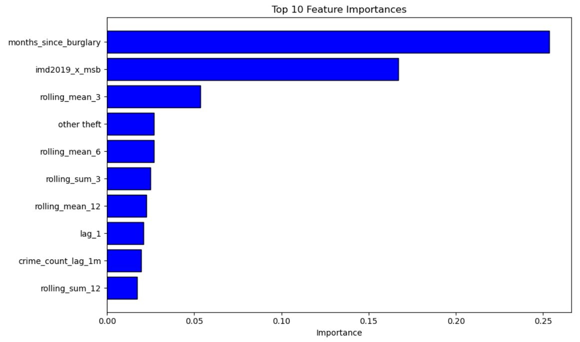

Figure 1: Feature Importance Bar Chart from XGBoost

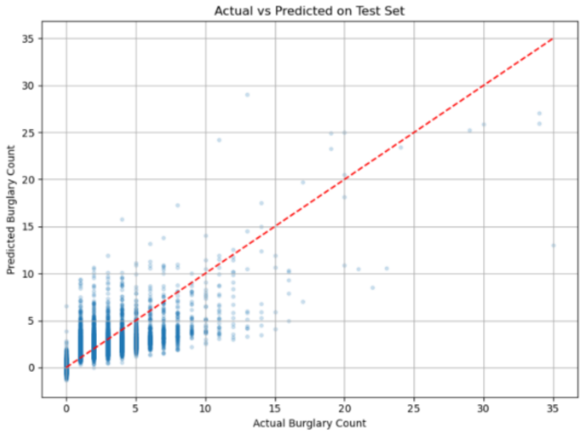

Figure 2: Actual vs. Predicted Burglaries

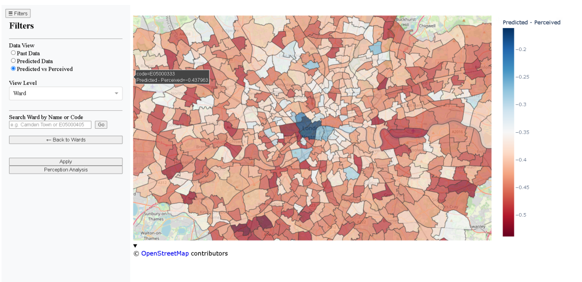

Figure 3: Perceived vs. Predicted Safety (Choropleth)

Figure 4: Community Tool Dashboard pt 1

Figure 5: Community Tool Dashboard pt 2

Figure 6: Community Tool Survey pt 1

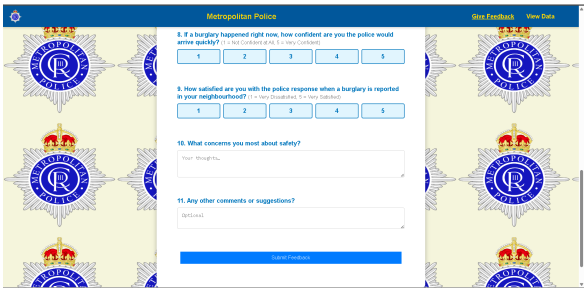

Figure 7: Community Tool Survey pt 2

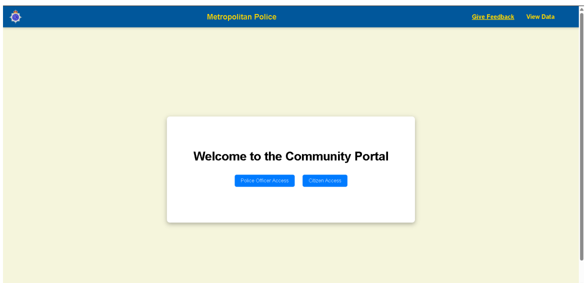

Figure 8: Community Tool Main Page

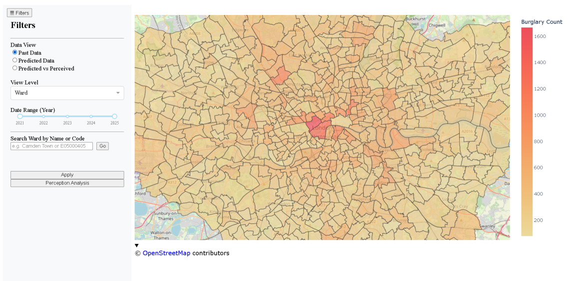

Figure 9: Police Dashboard pt 1

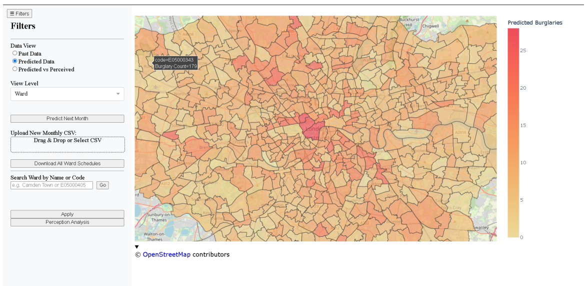

Figure 10: Police Dashboard pt 2

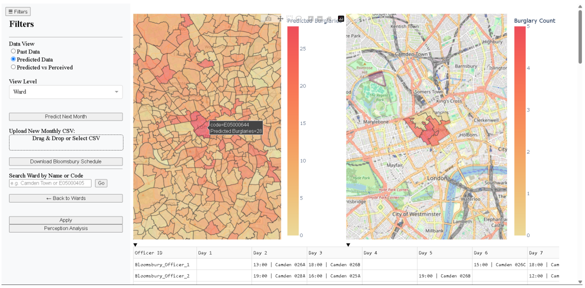

Figure 11: Police Dashboard pt 3

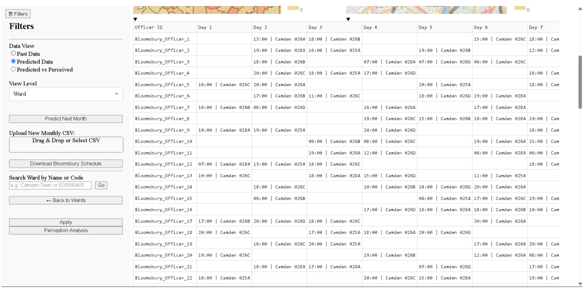

Figure 12: Police Dashboard pt 4

Figure 13: Police Dashboard pt 6