PROJECT REPORT

UNIVERSITY OF GUELPH

London Crime Map

An R-Shiny app to alert users about potential crime hotspots

BY

Atharva Vichare, 1320832

Prathamesh Walawalkar,

DATA*6500 Project Report

Under the guidance of

Dr. Justin Slater

Table of Contents

Abstract

Urban crime risk in London varies markedly across neighbourhoods and seasons, complicating personal safety decisions and local resource allocation. This study developed an interactive mapping approach that integrates police-recorded crime with socio-economic indicators to summarize neighbourhood-level risk at the Lower-layer Super Output Area (LSOA) scale. We combined LSOA and street-level crime records with the English Indices of Multiple Deprivation (2019), Census 2021 labour-market indicators, and mid-year population estimates to analyze spatial temporal patterns and produce an interpretable risk score. Exploratory analyses (choropleths, hotspot detection, and time-series decomposition) revealed persistent clusters of high crime frequency and seasonal surges across multiple boroughs. We then defined a two-component score that separates likelihood from vulnerability: likelihood **L = 0.70*CF + 0.30*C** (crimes per day CF, total crimes C), vulnerability **V = 0.60*IMD + 0.40*U** (deprivation IMD, unemployment U), and overall risk **R = 0.60*L + 0.40*V**. Higher deprivation and unemployment generally coincided with higher crime frequency. The resulting R Shiny "London Crime Map" enables non-technical users to explore risk over space and time, compare boroughs, and communicate patterns transparently. The framework is extensible to additional covariates (e.g., transport accessibility, workplace density) and can inform targeted prevention and local planning.

1. Introduction

London's crime risk varies across neighbourhoods and seasons, reflecting differences in activity patterns, deprivation, and the built environment. These differences are not random: nightlife districts, transport hubs, and retail corridors exhibit distinct temporal rhythms compared with primarily residential streets, and local socio-economic conditions shape both opportunity and guardianship in ways that evolve over time. At the same time, London offers unusually rich open data police-recorded crime at street and LSOA levels, socio-economic indicators, official boundaries, and population estimates that make fine-scale spatial–temporal analysis feasible.

Yet the abundance of data does not automatically yield insight. Crime records are event-based and time-stamped but deliberately anonymized at street level to protect privacy; socio-economic indicators arrive at different cadences (e.g., decennial Census, periodic deprivation updates) and spatial supports (LSOA, MSOA, borough); and boundary updates necessitate careful harmonization. Simple choropleths of raw counts can be misleading because they conflate population exposure with incidence and can mask within-area heterogeneity. Rate maps improve comparability but raise questions about the correct denominator (resident vs. ambient population), the stability of rates for small counts, and the treatment of seasonal circulation patterns. Temporal aggregation choices (weekly vs. monthly) determine sensitivity to short-lived spikes versus persistent trends, while the modifiable areal unit problem (MAUP) means interpretations can shift with the choice of spatial unit. Ignoring spatial autocorrelation or seasonal components risks biased inference and blurred change detection.

In practice, stakeholders from residents planning safer routes to local authorities prioritizing prevention need outputs that explain where and when risk is elevated and why, not just heatmaps of recent incidents. The methodological challenge is to convert heterogeneous, multi-resolution inputs into interpretable, decision-relevant summaries without sacrificing essential spatial–temporal structure. This motivates an approach that (i) distinguishes persistent hotspots from seasonal surges through baseline comparisons and clustering diagnostics, (ii) aligns event data with area-level indicators via transparent resampling and boundary-aware joins, (iii) normalizes indicators to improve cross-neighbourhood comparability when explicit exposure measures are lacking, and (iv) prioritizes interpretability by separating **likelihood** (what is happening now and how often) from **vulnerability** (structural conditions that heighten consequences and recurrence). London's open data ecosystem enables such a framework, while also requiring care around privacy jitter, small-number variability, and regime changes. In short, the goal is to turn rich but heterogeneous open data into a neighbourhood level summary of risk that is timely, comparable, and understandable supporting everyday decision-making and local prioritization while being explicit about limitations and uncertainty.

1.1 Problem statement

Stakeholders need neighbourhood-level summaries that explain **where** and **when** risk is elevated and **why** not just raw counts. However, several constraints complicate this task: (i) crime varies across both space and time, creating persistent and seasonal hotspots; (ii) indicators live at different geographic and temporal resolutions and require careful harmonization; (iii) standard choropleths can overemphasize high-population areas if rates and exposure are not handled properly; and (iv) highly complex models may be accurate but opaque for public communication.

A practical, interpretable framework is needed to aggregate recent crime frequency and socio-economic vulnerability into a transparent neighbourhood risk summary.

1.2 Motivation

Prior research links crime trends to economic incentives, area deprivation and social cohesion, macroeconomic shocks, and welfare changes, suggesting that vulnerability indicators help explain spatial disparities in crime. In a city with substantial open data infrastructure, there is an opportunity to integrate recent crime frequency with deprivation and unemployment to produce a **meaningful, communicable measure of local exposure**. An interactive map that exposes space–time patterns, supports comparison across LSOAs and boroughs, and remains interpretable for non-technical users can aid personal safety planning and local prioritization.

1.3 Objectives and Research Question

The primary aim of this project was to produce an interpretable, map-based summary of neighbourhood crime risk for London by integrating recent crime frequency with socio-economic vulnerability and operationalizing the result in an interactive R Shiny application. The objectives were formulated to address limitations of raw count maps (which conflate exposure with incidence) and highly complex models (which can be accurate yet opaque), and to promote transparent, data-driven communication that non-technical users can trust and act upon.

The specific objectives of the project were to:

1) Assemble and harmonize police-recorded crime, deprivation, labour-market, boundary, and population datasets at the LSOA level
2) Quantify recent crime **likelihood** using stable, time-windowed indicators (e.g., crimes per day and total incidents)

3) Represent **vulnerability** using area deprivation and unemployment
4) Combine these components into a concise risk score with clear weights and units
5) Implement an interactive map that reveals spatial temporal patterns, supports comparisons across boroughs and LSOAs, and communicates assumptions and limitations clearly.

Guided by this aim, the study addressed two research questions designed to connect recent patterns to structural context. First, **how are crime frequencies distributed across London's LSOAs over time, and where do persistent or seasonal hotspots emerge?** Second, **to what extent do socio economic indicators specifically deprivation and unemployment align with higher crime frequency at the LSOA level?** Together, these questions test whether a transparent combination of likelihood and vulnerability can yield a neighbourhood-level summary that is both empirically grounded and practically useful.

2. Background & Related Work

Urban crime risk is uneven across space and time, and a large empirical literature connects these patterns to local economic incentives, area deprivation, social cohesion, and macroeconomic shocks. When legitimate opportunities contract or are unevenly distributed, property crime especially tends to rise where expected gains exceed perceived risks; conversely, stronger social cohesion and community resources correlate with lower crime. Recessions and labour-market scarring have been linked to elevated crime that can persist beyond the initial downturn as job loss and reduced mobility become entrenched. Policy shifts that narrow welfare support can further exacerbate crime in vulnerable areas by tightening household budgets and weakening stabilizing institutions. In a London context, these results motivate summarizing recent crime frequency alongside structural vulnerability proxied by deprivation and unemployment at a neighbourhood scale that is meaningful for residents and local practitioners. A map that only colors LSOAs by counts risks reproducing where people live, work, or travel rather than illuminating the underlying risk; a more useful summary is both time-sensitive (capturing current patterns and seasonal surges) and structure-aware (reflecting socio-economic context).

London's open data ecosystem makes such analysis feasible by providing complementary sources at multiple resolutions. Police-recorded crime is available as street-level events (with privacy-preserving spatial displacement) and as LSOA aggregates over regular time steps . Structural context comes from the English Indices of Multiple Deprivation (2019) and Census 2021 labour market indicators, while ONS boundaries and mid-year population estimates supply authoritative geographic frames and denominators for rates . Additional layers such as transport accessibility or job density can help interpret ambient exposure and routine activity patterns . These assets, however, bring practical considerations: positional uncertainty in street-level points cautions against over-interpreting micro-hotspots; heterogeneous update cadences (monthly events vs. periodic structural indicators) complicate temporal alignment; versioned geographies require careful handling of splits/merges and boundary vintages; and small-number volatility in lightly populated LSOAs can distort rate maps. Any pipeline must therefore address temporal alignment (common windows), spatial harmonization (consistent CRS and boundary versions), exposure normalization (rates or standardized indicators), and uncertainty communication (documenting jitter, small counts, and assumptions) .

Analytically, urban crime studies commonly begin with exploratory spatial data analysis (ESDA). Global Moran's I gauges overall spatial autocorrelation; local indicators (LISA) identify clusters and outliers; and Getis-Ord Gi* statistic provides hot and cold spots. Results depend on the spatial weights matrix (e.g., contiguity or k-nearest neighbors), so robustness checks and explicit reporting are good practice. Time structure is examined via decomposition of trend and seasonality, seasonal indices, and changepoint detection to separate persistent shifts from short lived spikes (events or targeted operations). For modeling, panel regressions (fixed/random effects) and spatial econometric models (e.g., spatial lag SAR, spatial error SEM) incorporate space time dependence and correct inference when residuals are spatially correlated. Each approach has limits in public-facing settings: choropleths of raw counts overstate risk where exposure is higher; rate maps depend on denominators (resident vs. ambient population) that may not reflect the population at risk and can be unstable for small counts; and complex predictive models, while accurate, can be opaque to non-specialists and difficult to validate with community stakeholders. Dashboards often present multiple unconnected layers (counts, rates, socio-economic indices) without a principled way to summarize them into a single, communicable signal.

This project responds to that gap by constructing a transparent, two-component measure that separates **likelihood** (how often incidents occur over a recent, stabilized window) from **vulnerability** (structural conditions associated with higher impact and persistence), implemented with London's open data and guided by socio-economic evidence . Likelihood is proxied by crimes per day and total incidents over a defined horizon; vulnerability is represented by deprivation and unemployment. The resulting score is not a causal claim, but an interpretable summary aligned with how non-technical users' reason about risk, while adhering to spatial temporal good practice (aggregation to stable supports, consistent windows, and explicit clustering/seasonality diagnostics).

Several design choices flow from the literature and data properties and are carried through the report. We prioritize stable pooling windows to balance recency with variance, especially for small areas; aggregate to the LSOA level to mitigate positional jitter while acknowledging the modifiable areal unit problem (MAUP); quantify and report spatial autocorrelation in underlying indicators and model residuals; document exposure assumptions (using resident population where ambient measures are unavailable) and avoid over-interpreting small denominators; and emphasize interpretability over black box prediction to support public communication and decision-making .

3. Data

This project integrates complementary open datasets for London at two spatial supports—street-level events and LSOA aggregates—to summarize neighbourhood risk while preserving key space–time structure. Core sources include Metropolitan Police (MPS) crime counts at LSOA level (recent and historical), UK Police street-level records, structural indicators from the English Indices of Multiple Deprivation (IMD) and Census 2021, authoritative small-area boundaries from ONS, and small-area population denominators. Additional context layers (transport accessibility and job density) support interpretation where relevant.

1) **MPS LSOA Level Crime — most recent 24 months (counts, LSOA)** : Monthly recorded crime counts per LSOA for the latest two years. This is the primary "likelihood" signal (e.g., crimes/day, total incidents) for current risk because it avoids the positional jitter in street-level points and aligns cleanly with LSOA-based structural indicators.
2) **MPS LSOA Level Crime — Historical (counts, LSOA – historical)** : An extended time series of the same small area counts beyond 24 months. Used to establish baselines, check seasonality, and identify hotspots that recur across years rather than appearing as one-off spikes.

3) **Indices of Deprivation 2019 — Scores (LSOA)** : Overall IMD score ranking small areas by relative deprivation across domains (income, employment, education, health, housing, crime, living environment). Serves as a **vulnerability** proxy that changes slowly over time and provides structural context.

4) **Indices of Deprivation 2019 — Domains of Deprivation** : Domain-level components of IMD (e.g., income, employment, health, education, crime). Used diagnostically to interpret which underlying dimensions co-vary with crime in flagged LSOAs; not all domains feed directly into the summary score.

5) **Census 2021 — Labour Market: Economic Activity (LSOA)** : Small-area labour-market indicators, including unemployment. Provides a second **vulnerability** proxy complementing IMD by capturing recent labour-market stress at the same spatial support as crime counts.

6) **LSOA boundaries (Dec 2021) — ONS Open Geography** : Authoritative polygon geometries for LSOAs. A single boundary vintage is applied across all joins to ensure geographic alignment and to avoid artifacts from splits/merges or boundary drift.

7) **Jobs and Job Density** : Workplace counts and jobs-to-resident ratios by area. Offers additional context for daytime inflow and opportunity structure beyond resident population, helping explain high-incidence LSOAs with low residential counts.

Data limitations and challenges

1. **Positional uncertainty (street-level data).** Street-level points are intentionally displaced to protect privacy. Fine-grained hotspot surfaces can over-interpret this displacement. We therefore aggregate to LSOA or use points qualitatively, not for micro-block inference .
2. **Different update cadences.** Crime events are monthly and current; IMD updates are infrequent; Census 2021 is a structural snapshot; population estimates are annual. Analyses use **common time windows** and treat slow-moving indicators as contextual, not as high-frequency drivers .
3. **Boundary versions.** LSOA boundaries evolve. We adopt a single ONS vintage (Dec 2021) for all joins to avoid misalignment; historical series are treated carefully to minimize artifacts from splits/merges .
4. **Small-number volatility.** In lightly populated LSOAs, rates can be unstable. We prefer **pooled time windows** (e.g., multi-month) and report counts alongside rates to avoid over-interpreting rare events .
5. **Recording practices and category mix.** Changes in police recording, campaign-driven reporting, or category redefinitions can shift counts independent of true incidence. We note such possibilities when interpreting short-run changes and rely on **multiple indicators** (counts, rates, persistence) rather than a single metric.

Data Preprocessing -

Crime CSV files were read and standardised to ensure consistent column naming. Socioeconomic datasets were cleaned to retain only relevant attributes — IMD scores, unemployment percentages — and borough names were harmonised across datasets. Monthly crime counts, stored in wide format (one column per month), were converted to long format using `tidyr::pivot_longer`.

This allowed easy filtering and time-series analysis. Column names such as `"202308"` were converted into proper `Date` objects using `lubridate` functions for temporal plotting.

Borough-level monthly totals were computed using `dplyr::group_by` and `summarise`, ensuring each borough had complete time-series data. Crime types were grouped into simplified categories (e.g., Theft, Violence, Vehicle Crime, Drugs, Other) for more interpretable visualisation.

A **risk score** was calculated by combining crime frequency and socioeconomic vulnerability. This was implemented as:

likelihood L = 0.70*CF + 0.30*C (crimes per day CF, total crimes C),

vulnerability V = 0.60*IMD + 0.40*U (deprivation IMD, unemployment U),

overall risk R = 0.60*L + 0.40*V.

4. Tools and Technologies

The implementation of the London Crime Map relied on a compact R-based stack selected for small-area spatial analysis, reproducible ingestion from open APIs, and responsive, map-centric interactivity.

- **R Shiny** — Framework for the interactive application (modular structure: app.R, ui.R, server.R). Handles reactive filtering (time window, category, borough/LSOA), computes the likelihood–vulnerability risk score on the server, and renders maps/charts without page reloads.

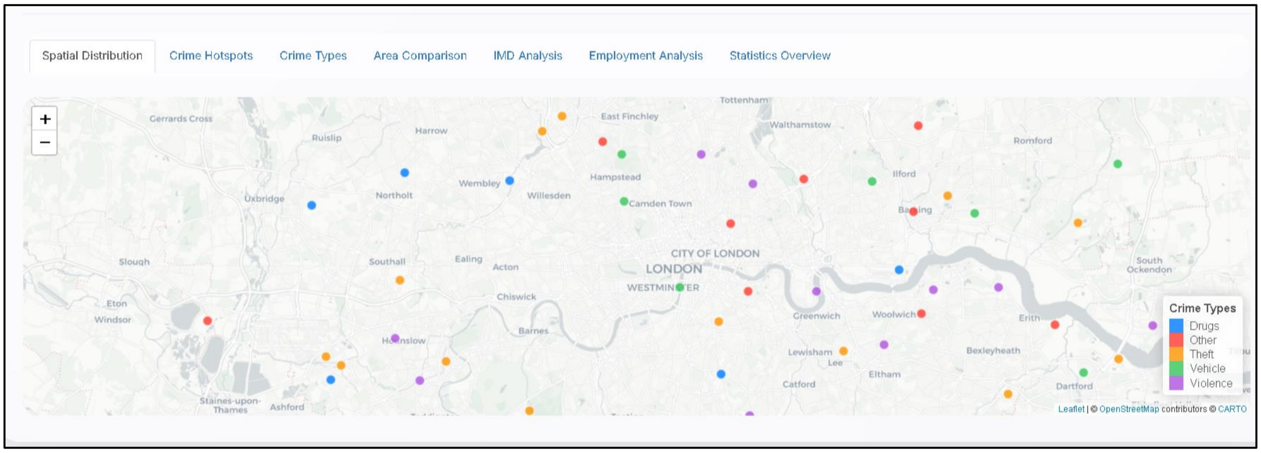
- **tidyverse (dplyr, tidyr, readr, stringr, purrr)** — Data wrangling pipeline for cleaning, joining, reshaping, and summarizing monthly crime counts and socio-economic indicators into an analysis-ready LSOA panel.
- **data.table** — High-performance aggregation for large street-level event files and fast windowed summaries (e.g., crimes/day over rolling horizons).
- **sf (+ s2, rmapshaper)** — Spatial data model and geometry engine used to read/validate ONS LSOA polygons, harmonize CRS, perform spatial joins (street points → LSOA), and simplify geometries for web mapping while preserving topology.
- **spdep (and classInt)** — Spatial diagnostics and weights construction (contiguity/k-NN) for Global Moran's I, Local Indicators of Spatial Association (LISA), and cluster classification; classInt for consistent choropleth breaks.
- **leaflet (+ leaflet.extras)** — Web map rendering with tiled basemaps, choropleths at LSOA level, popups/tooltips, and hotspot overlays; supports smooth interaction on modest hardware.
- **tmap (static) + ggplot2 / plotly (interactive)** — Descriptive maps for the report (tmap), and interactive time-series/distribution charts in the app (ggplot2 for grammar, converted to plotly for tooltips/hover).
- **httr / curl + jsonlite** — Robust ingestion from the UK Police street-level crime API and other HTTP endpoints; JSON parsing to tibbles, with error handling and rate-limit backoff.
- **lubridate** — Date parsing and calendar arithmetic for month indexing, rolling windows, and seasonality summaries.
- **arrow (or qs)** — Fast, compressed on-disk caches of intermediate tables to speed up iterative development and app startup without re-pulling raw files.

- **renv** — Project-local package management to capture exact versions and ensure the app reproduces across machines.
- **Git + GitHub** — Version control, issues, and pull requests; supports transparent change tracking and collaborative review.
- **R version 4.4.x** — Primary computational environment for data processing, diagnostics, and application runtime.

5. Results

Landing page

The **borough selection control**, **geospatial crime distribution map**, and **risk evaluation summary** are brought together to provide an immediate overview of safety conditions for the chosen location. The borough filter dynamically updates both visual and numerical outputs, ensuring all displayed information is relevant to the selected area. This design was motivated by the need for a user-centric interface that prioritizes localized decision-making, as crime patterns can differ significantly even within the same city. By integrating risk scoring alongside the map, the user can quickly interpret not only the location of incidents but also their broader safety implications.

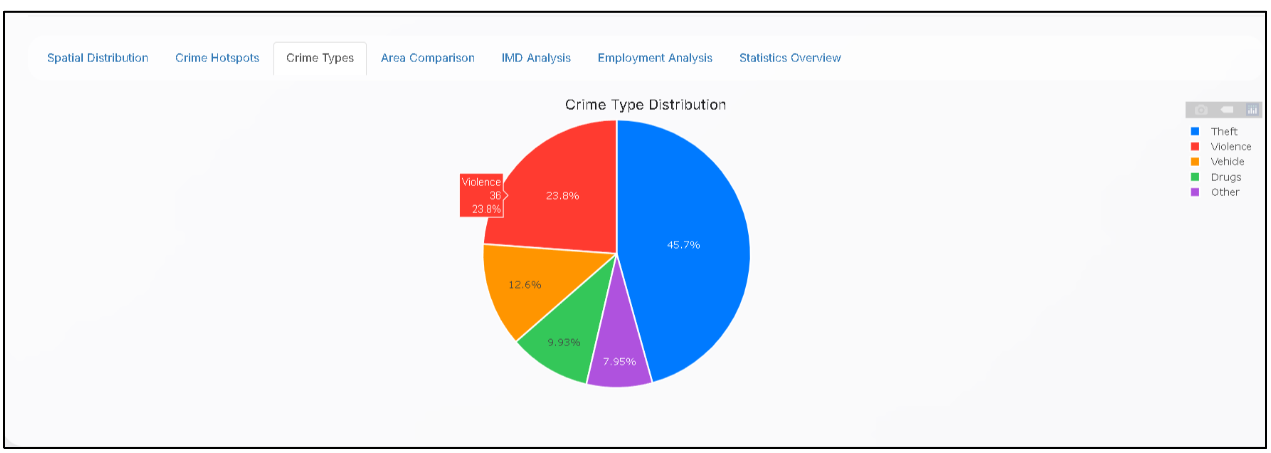


Spatial distribution of crimes in London

The **spatial distribution map** presents individual crime incidents as georeferenced points, each color-coded according to offence type. This fine-grained view provides precise locational data that can be cross-checked with other urban datasets, such as street lighting coverage or public transport routes, to identify possible contributing factors. Unlike aggregated heatmaps, this method supports case-level analysis, enabling investigative teams to see exactly where incidents are occurring and detect micro-level trends. The inclusion of crime type colors also ensures that users can distinguish between different forms of criminal activity at a glance.

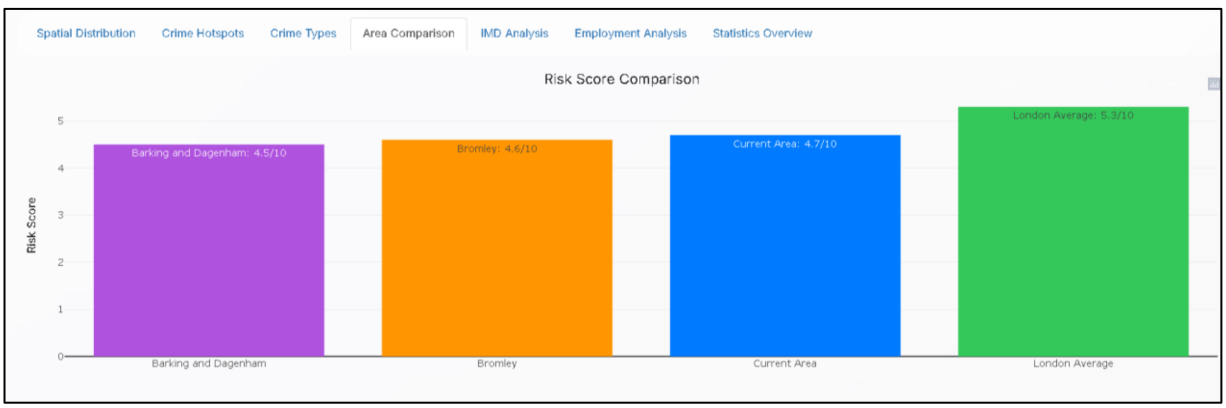
KDE hotspot prediction

For this heatmap, spatial density analysis is used to identify clusters of high criminal activity within the geographic area. The intensity-based color gradient enables immediate recognition of the most problematic zones, supporting the prioritization of patrol routes, CCTV placement, or environmental design changes. The heatmap's value lies in its ability to reveal patterns that are less obvious in point-based maps—helping decision-makers focus resources where they can achieve the greatest impact. This layer also supports proactive policing by highlighting areas at risk before crime levels escalate further.



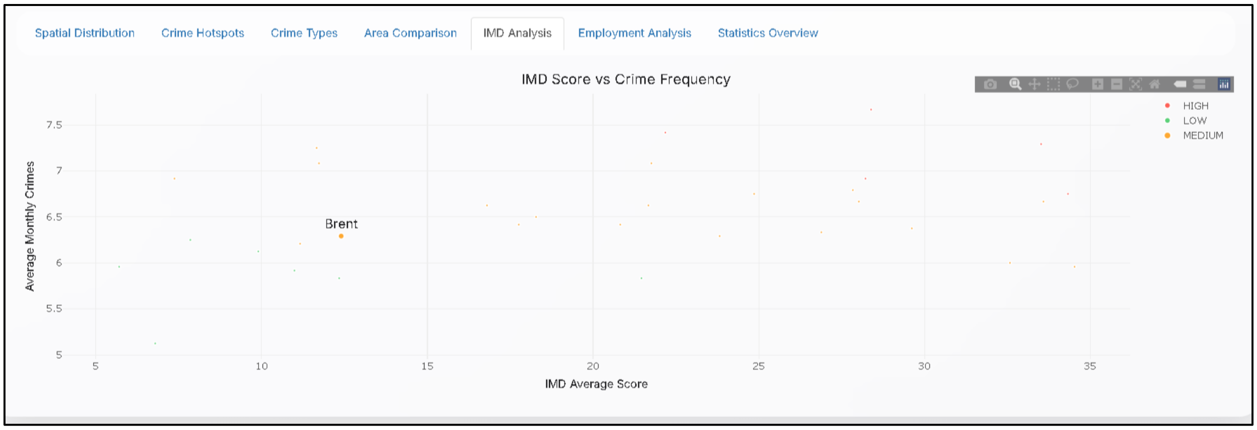
Crime type distribution

The **crime type distribution** pie chart breaks down all reported incidents into major categories such as theft, violence, vehicle crime, drug-related offences, and others, showing their proportional contribution to the total crime volume. The inclusion of this breakdown is important because different crime categories require different intervention strategies, policing approaches, and community initiatives. For example, a borough with a high proportion of theft might focus on security awareness campaigns, while one with elevated violence rates might invest in conflict prevention programs. This visualization provides a straightforward foundation for such strategic planning.



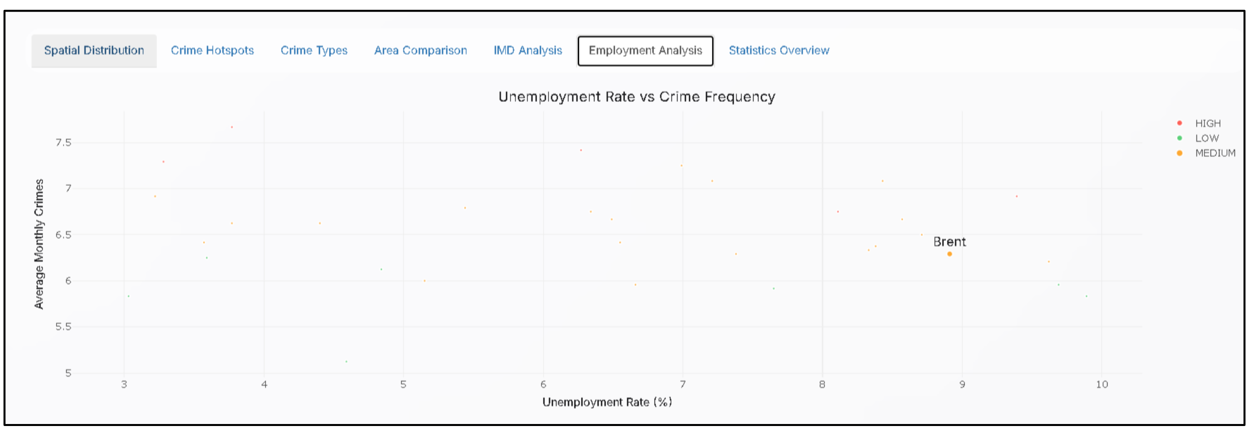
Risk comparison with neighbouring areas

Here we compare the selected borough against neighboring areas and the overall London average, translating numerical assessments into a clear visual hierarchy. This comparative method was selected to provide instant clarity—rather than interpreting an isolated score, users can immediately understand whether their area is performing above or below the metropolitan norm. The side-by-side layout encourages quick benchmarking while also revealing performance gaps that might not be obvious from raw figures alone.



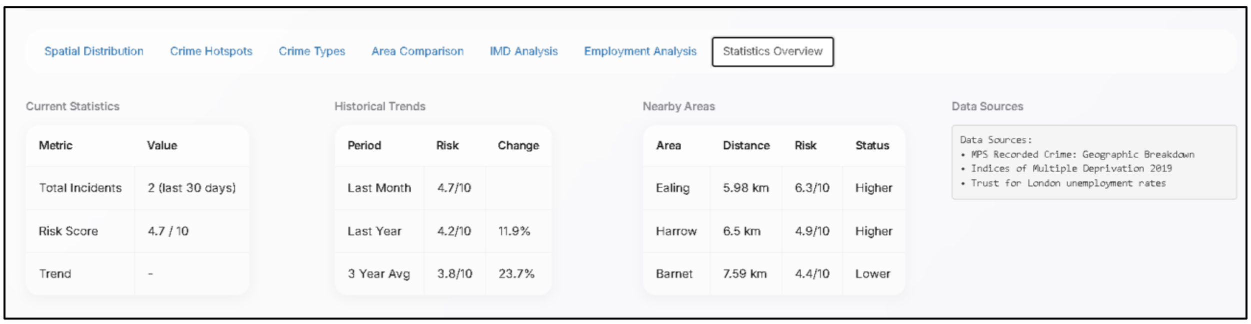
Relation between IMD score and crime frequency

In this, the **IMD score vs. crime frequency** relationship is analyzed, comparing socio-economic deprivation—measured by the UK's Index of Multiple Deprivation—with recorded crime levels. IMD is a composite measure encompassing factors such as income, employment, education, and housing quality, making it an ideal benchmark for assessing community vulnerability. By mapping this against crime rates, the dashboard can identify boroughs where high deprivation is strongly linked to crime, which can inform more targeted and efficient allocation of public safety resources.



Relation between unemployment rate and crime frequency in that area

The scatter plot plots boroughs based on their average monthly crime counts and unemployment percentages, while assigning color codes for risk categories. This visualisation was created to explore potential socio-economic correlations, given the relationship between economic stressors and certain crime types. Presenting both variables in a unified space allows for quick identification of boroughs where high unemployment coincides with elevated crime. This insight is essential for policymakers who wish to address the root causes of safety concerns through both law enforcement and economic development strategies.



Statistical summary

Here we have the **statistics overview**, which consolidates real-time crime metrics, multi-year historical trend data, comparative analysis with nearby boroughs, and the original data sources. This single-pane structure minimizes navigation effort, allowing the user to cross-reference current statistics with long-term patterns without switching views. Including distance-based comparisons with surrounding boroughs was intentional, as residents and stakeholders often assess safety relative to neighboring areas.

6. Impact and Broader Implications

The London Crime Map translates multiple open datasets into an interpretable, neighbourhood-level summary of risk, enabling non-specialists to reason about **where** and **when** exposure is elevated and **why**. Because it is built on official, regularly updated sources police-recorded crime (street and LSOA), deprivation, labour-market indicators, authoritative boundaries, and population denominators the approach is transparent and reproducible, supporting routine monitoring as new data arrive . Framing results through a simple likelihood vulnerability construct aligns with established evidence that area deprivation and labour-market stress co-move with crime, making the outputs easier to communicate and to connect to actionable levers beyond enforcement .

For residents and visitors, the application provides a practical safety aid: users can explore recent patterns by LSOA, observe seasonal rhythms, and compare adjacent neighbourhoods before choosing routes, meeting points, or travel times. Presenting both **recent frequency** (likelihood) and **structural context** (vulnerability) helps avoid common misinterpretations of raw counts or unstable rates, while popups and tooltips make assumptions explicit (e.g., resident population as a denominator where ambient exposure is unknown) . The emphasis on explainability clear definitions, consistent windows, and documented limits improves trust relative to opaque predictive scores.

For local authorities and practitioners, the map offers a quick triage view that complements formal analysis. Persistent hot areas and seasonal surges can be prioritized for targeted prevention, environmental design changes, or outreach, while stable small-area panels enable before–after tracking of interventions at the LSOA scale . Because indicators are harmonized to a common boundary vintage, cross-borough comparisons are more defensible, and recurring patterns can be distinguished from short-lived anomalies. Importantly, the tool is not a substitute for formal risk assessment or deployment models; instead, it provides a **decision support layer** that highlights where deeper analysis, and engagement would be most valuable.

Policy implications extend beyond policing. The consistent alignment between deprivation, unemployment, and elevated crime frequency underscores the potential value of

place-based socio-economic measures from youth employment and training to housing quality and local services in reducing risk over time . By situating recent patterns within structural context, the application supports conversations that include public health, education, transport, and community organizations alongside law enforcement.

Transparency and accountability are strengthened by relying on open, auditable inputs and by documenting processing steps (temporal alignment, boundary harmonization, and exposure assumptions) . Because all sources are public and versioned, external stakeholders can replicate the pipeline, scrutinize choices (e.g., spatial weights, smoothing), and propose improvements. This openness encourages community engagement residents can validate face validity, highlight blind spots, or suggest contextual layers that improve interpretation.

Equity and ethics considerations are integral. Area-level risk maps can unintentionally stigmatize communities or create feedback loops if used to justify disproportionately aggressive enforcement. The design therefore emphasizes **interpretation guidance** (e.g., "risk ≠ blame"), reports count alongside stabilized indicators, and flags uncertainty in small-number areas. Where feasible, complementary context layers (e.g., transport access, job density) can help explain high incidence in low-residence zones without attributing risk to residents themselves . Continuous monitoring for distributional impacts (e.g., whether communication or resource allocation patterns systematically disadvantage certain areas) is recommended.

Generalisability is high. The method can be ported to other UK cities or regions by swapping in the relevant street-level feeds, small-area boundaries, and socio-economic indicators; the reliance on standard geographies and public releases simplifies reuse . For London itself, the framework is extensible to additional covariates (e.g., footfall proxies, business opening hours) and to modest modelling enhancements, provided interpretability remains central. Future iterations can incorporate explicit uncertainty bands, ambient population estimates where available, and user-facing documentation that links local findings to the broader socio-economic literature .

Finally, while the observed associations between deprivation, unemployment, and crime frequency are consistent with prior research, the project's aim is **communication and decision support**, not causal attribution. Recording practices, category definitions, and population dynamics can shift measured incidence independently of true underlying risk . Keeping these limits visible helps ensure that the tool is used to **coordinate prevention and community investment**, rather than to over-interpret short run fluctuations or to label neighbourhoods in ways that could be counterproductive.

7. Conclusion and Future Work

This study set out to turn London's rich but heterogeneous open data into an interpretable, neighbourhood level picture of crime risk. Using official small-area counts (LSOA), street level events, deprivation indices, Census labour-market indicators, authoritative boundaries, and population denominators, we assembled a reproducible pipeline that surfaces spatial–temporal patterns across the city . Exploratory analysis highlighted persistent clusters and seasonal surges, while comparisons across boroughs and adjacent LSOAs showed that some areas experience consistently higher incident frequency than their neighbours. Against this backdrop, we found that areas with greater deprivation and higher unemployment tended to exhibit higher recent crime frequency an association consistent with long-standing empirical evidence linking concentrated disadvantage, labour-market stress, and crime outcomes .

To communicate these patterns in a way that non-specialists can use, we built a transparent, two-component **risk score** that separates **likelihood** (crimes per day and total incidents over a stable window) from **vulnerability** (deprivation and unemployment). The mechanics are intentionally simple and documented so that assumptions can be inspected and debated. Delivered through an R Shiny interface, the map visualizes hotspots and risk scores at the LSOA level, enabling residents, visitors, and practitioners to explore where risk is elevated and how it changes over time without needing to interpret raw counts or opaque model outputs. While the score is not causal, it provides a compact, decision-support summary that aligns with how people reason about exposure and context, and it can be replicated or extended as datasets update.

At the same time, we remain cautious about limitations: positional displacement in street-level data, cadence mismatches across sources, small-number volatility in lightly populated areas, and exposure proxies based on resident population rather than true footfall. These constraints argue for careful interpretation, pooled windows for stability, and transparent documentation of boundary vintages and normalization choices. With those caveats, the framework contributes a clear, reproducible way to summarize neighbourhood crime risk that is both empirically grounded in open data and aligned with socio-economic insights from the literature .

Future Work

Extend beyond IMD and unemployment to examine domain-level IMD components (income, education, housing, health), youth demographics, housing tenure, and indicators of the night-time economy (e.g., licensed premises density), alongside existing context layers such as PTAL and job density. Where feasible, incorporate proxies for ambient population (footfall, mobility) to refine exposure.

Refine the risk score.

- Perform **sensitivity analysis** on weights (e.g., grid search, cross-validated regression to learn weights from held-out periods) and report uncertainty bands.
- Calibrate by borough or cluster of similar LSOAs to account for structural heterogeneity.
- Explore category-specific likelihood (e.g., theft vs. violence) with a composite display that preserves interpretability.
- Add **stability controls** (minimum counts, Empirical Bayes smoothing) to mitigate small-number volatility.

Complete and harden the R Shiny app.

- Implement caching (e.g., arrow/qs) and pre-computed tiles for fast startup; add accessibility features (keyboard navigation, high-contrast themes).
- Provide in-app methodology notes and a downloadable "data dictionary" and changelog to support transparency and reproducibility.

Deeper street-level integration.

- Use street-point data more fully while respecting positional displacement: aggregate to LSOA for metrics and optionally provide *illustrative* kernel density overlays with disclaimers.
- Where categories and timestamps allow, add time-of-day or day-of-week views and small multiples to show routine activity rhythms .

Evaluation and ethics.

- Conduct user testing with residents and practitioners to assess clarity, usefulness, and potential misinterpretations; iterate on wording and visual encodings accordingly.
- Monitor for distributional impacts and avoid stigmatization: emphasize "risk ≠ blame," show counts alongside stabilized indicators, and document uncertainty.

- Track data-quality issues (recording changes, category redefinitions, boundary updates) and flag breaking changes in the UI.

Generalisation.

- Port the framework to other UK cities by swapping in local boundaries and feeds; maintain the same transparent likelihood–vulnerability structure to preserve comparability.
- Where richer exposure data are available (e.g., footfall), test whether incorporating them improves calibration without compromising interpretability.

References

1. MPS LSOA Level Crime — most recent 24 months. London Datastore. https://data.london.gov.uk/dataset/recorded_crime_summary
2. Indices of Deprivation 2019 — IMD Scores (LSOA). London Datastore. https://data.london.gov.uk/dataset/indices-of-deprivation-2019
3. Census 2021 (GLA/ONS) — Labour Market: Economic Activity (LSOA). https://data.london.gov.uk/dataset/2021-census-lsoa-labour-market
4. ONS Open Geography — LSOA boundaries (Dec 2021). https://geoportal.statistics.gov.uk/
5. GLA — Jobs and Job Density. https://data.london.gov.uk/dataset/jobs-and-job-density
6. M. Draca and S. Machin, Economic incentives and crime, 2015.
7. I. Kawachi, B. P. Kennedy, and R. G. Wilkinson, Deprivation/inequality, social cohesion, and crime, 1999.
8. B. Bell, A. Bindler, and S. Machin, "Crime Scars: Recessions and crime," 2018.
9. O. Melander and M. Miotto, Welfare cuts and crime, 2022.