

Individual Project Assignment

Riya Gupta

K24090688

ACADEMIC HONESTY INTEGRITY

Students at King's are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a college policy that students take responsibility for their work and comply with the university's standards and requirements. Online proctoring / invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarize themselves with the procedures on the College website. *I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.*

PART 1: Analytics

Q1: Analyze the development of crimes over time. Are there detectable trends in types of crime (e.g., theft, assault) over the past ten years?"

This helps us see which crimes are increasing or staying high, so action can be taken in the right areas.

Q2: To what extent have police resolution rates varied across different crime types in London over the past decade, and what might these patterns suggest about the effectiveness or prioritization of policing efforts?
It shows which crimes police are handling well, and where there might be delays or problems.

Q3: How do subtypes within major crime categories behave over time, and what insights can these patterns offer about the shifting nature and dynamics of these crimes?

This helps spot new or growing problems in society, like rise in certain hate crimes.

Data Collection and Preparation

The dataset for this study was sourced from the London Metropolitan Police's open data platform [1], covering reported crime incidents from 2015 to 2025. Records from 2015 to 2021 were available in separate files organized by financial year, while data from 2022 onward was released in a single consolidated file. Files marked as "Other Crime Data" were considered to carry out the study. All files were combined to create one unified dataset for analysis.

Given the dataset's size and structure, most preprocessing was performed using Python, followed by Excel [6] for formatting and creating visual outputs.

Data Cleaning and Preprocessing Methods:

- **Standardizing Date Column:** All date entries were converted to the standard format of YYYY-MM-DD to ensure uniformity in time-based analysis. However, Excel [6]'s auto-formatting altered this format, which was manually corrected during visualization.
- **Filtering by Measure Type:** The dataset included multiple types of records, but only those explicitly labelled as "offences" were retained, representing confirmed crime reports.
- **Handling Missing Data:** Rows lacking records/values, such as area name or code, were eliminated. This step excluded **922 incomplete entries** to maintain analytical consistency and prevent distortions in the results.
- **Whitespace Cleaning:** All textual data was cleaned by trimming extra spaces and converting text to lowercase. This helped avoid classification mismatches due to inconsistent formatting.
- **Dropping Metadata:** Non-analytical fields like fy_fyindex and refresh date were excluded to streamline the dataset and ensure focus on relevant variables.
- **Deduplication:** Duplicate entries were identified and removed so that each row represented a unique crime incident.
- **Outlier Suppression:** The count column was restricted to provide data with a count greater than or equal to 10. This helped reduce noise in the charts and brought attention to statistically meaningful patterns.
- **Exclusion of the "Miscellaneous" Category:** The crime type labelled as "miscellaneous" along with all of its associated subtypes was removed to maintain clarity. This decision ensured that the analysis focused on **clearly defined and consistently reported crime types**.

Subtype Selection for Final Analysis

The original dataset included **24 subtypes** spanning four main crime categories: **hate crime, domestic abuse, knife crime, and gun crime**. After careful filtering, **13 subtypes** were retained for detailed analysis. These were selected based on their **volume, clarity, and relevance** to the research questions. Subtypes with overlapping meanings or very low occurrences such as "*domestic abuse incidents*" or "*gun crime personal robbery*" were removed to keep the visualization clear and readable.

Simulated Outcome Dataset for Q2

The official records did not contain complete outcome or resolution data for every crime type. A simulated outcome dataset was created to address this gap to support Question 2. Estimated resolution rates ranging between 20% and 70% were assigned to each crime category. These rates were derived from UK Home Office [2] reports and benchmark trends.

The formula used was:

$$\text{Estimated Resolved Cases} = \text{Total Reported Offences} \times \text{Assigned Resolution Rate}$$

This allowed for the approximation of resolution patterns across crime types and was used strictly for comparative purposes in the outcome analysis. All simulated figures were clearly labelled to ensure transparency.

Datasets Used Per Research Question:

To make the analysis more straightforward to manage, the primary dataset was split into three separate files for each research question:

Q1: Q1.xlsx – Total crimes by year and type

Q2: Q2.xlsx – Simulated resolution rates. [4]

Q3: Q3.xlsx – Top subtypes for D3 dashboard and trend tracking

Python was mainly used for cleaning, filtering, and structuring the data, while Excel [6] was used for final formatting and visual layout.

Dataset Correlation

All three research questions in this study are based on the same core dataset — London Metropolitan Police [1] Crime Data from 2015 to 2025 — but each required different filtering and transformation depending on the focus of the analysis. [7]

•Q1 looked at broad crime trends over time. The data was grouped by year and crime type to help spot long-term changes in reported offences.

•Q2 focused on how often crimes were resolved. Since actual outcome data was not fully available, a simulated dataset was created using estimated resolution rates. This made it possible to compare police effectiveness across different types of crime.

•Q3 looked at the conduct of subtypes within the main crime categories over time. To identify internal trends and changes, particular crime types were chosen, and their subcategories were monitored.

This approach illustrates how versatile the dataset was in supporting various perspectives of crime analysis, even though all three questions used the same original dataset. However, the preparation and exploration of the data varied depending on what was being investigated.

PART -2

QUESTION-1

This section explores how key crime categories in London have shifted over time by analysing data at a **monthly level** between **2015 and 2025**. Using monthly figures instead of yearly summaries helps reveal subtle patterns, such as seasonal spikes or irregular dips, that are often missed in broader aggregates. It also allows for closer inspection of how specific types of offences have either remained stable or changed during major external events like the **COVID-19 lockdowns**.

The trends are presented through a multi-line chart [4] created in Excel [6] (see Figure 1), where each line tracks a separate

crime category over time. The x-axis spans each month across the ten-year period, while the y-axis shows the number of reported incidents. This structure makes it easy to observe how the levels of different crimes rise or fall in relation to one another.

From the visual, it's clear that offences like **domestic abuse** have consistently high volumes, while crimes such as **knife crime** and **hate crime** have seen a gradual increase. A sharp dip around early **2020** aligns with the national pandemic response, hinting at possible disruptions in crime patterns or reporting practices.

By taking a monthly approach, this analysis captures both long-term shifts and short-term variations, offering a more detailed understanding of how different forms of crime behave over time in an urban setting.

QUESTION-2

Due to a lack of outcome resolution rate records in the original data, a simulated dataset was developed to explore how effectively different crimes may have been resolved between 2015 and 2025. This dataset was based on simulated resolution rates drawn from typical ranges observed in UK Home Office [2] statistics, with each crime type assigned a value between 20% and 70% depending on known enforcement trends.

The results are presented using a grouped bar chart [4] built in Excel [6] (see Figure 2), where each group of bars represents a specific crime type across the ten-year span. Colours are used to differentiate the years, allowing for a straightforward visual comparison across both time and categories.

The chart highlights key disparities: offences such as **knife crime** and **gun crime** show relatively higher estimated outcomes, suggesting these categories may receive more targeted enforcement. In contrast, crimes like **domestic abuse** and **harassment** show lower values, pointing to persistent challenges in investigation or victim support. While the data is modelled, it offers a useful lens to examine potential imbalances in policing outcomes and priorities.

This approach enables an insightful comparison in the absence of full official data and draws attention to where investigative or policy improvements might be needed.

QUESTION-3

This section explores how specific subtypes of crime have changed over time within four major categories: **domestic abuse**, **hate crime**, **knife crime**, and **gun crime**. A set of **13 subtypes** was selected based on how frequently they appeared in the dataset and how clearly, they reflected distinct offence types. This allowed for a more focused analysis within broader crime groups.

To track these changes, a monthly line chart was created in Excel [6] using data from 2015 to 2025 (see Figure 3a). Each line represents a different subtype, making it easier to spot upward or downward trends as well as irregular shifts across the decade. Time is displayed along the x-axis, with case numbers shown on the y-axis. Colour was used to separate each subtype for easier comparison visually. While a heatmap ((see Figure 3b). was briefly explored during the early analysis for its compact layout, it was found to be less effective in showing clear month-to-month trends for multiple subtypes. As a result, the line chart was preferred for its clarity and ability to highlight temporal patterns more distinctly. One clear pattern is the increase in **hate crimes related to racial or religious discrimination**. In contrast, some subtypes—such as certain forms of domestic abuse or robbery—remained more stable or showed inconsistent fluctuations.

To add readability and comprehensive proportionality to these trends, a tree map was developed using D3.js [5] (see Figure 3c) showing how each subtype contributes to the total number of cases within its category. This visual helps compare the relative scale of

each subtype, offering a broader view of their impact. Overall, these charts provide a well-rounded perspective that captures how subtype trends shift over time and how significant each one is in the overall pattern of crime in London.

References

- [1] Metropolitan Police, London Crime Data Dashboard, London.gov.uk. Available: https://data.london.gov.uk/dataset/recorded_crime_summary
- [2] UK Home Office, 'Crime outcomes in England and Wales 2021 to 2022,' Published: July 2022. [Online]. Available: <https://www.gov.uk/government/statistics>
- [3] Office for National Statistics, 'Crime in England and Wales: year ending March 2023,' ONS, July 2023.
- [4] C. Ware, *Information Visualization: Perception for Design*, 4th ed., Morgan Kaufmann, 2020.
- [5] B. Fry and C. Reas, *Getting Started with D3.js*, 2nd ed., O'Reilly Media, 2022.
- [6] T. Knafllic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*, Wiley, 2015.
- [7] D. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Analytics Press, 2012.

Appendix

Figure 1. Trends in Major Crime Categories (2015–2025) – Line graph visualizing long-term patterns across six major crime subtypes in London. The chart illustrates monthly case frequencies, highlighting fluctuations and notable peaks in domestic abuse and hate crime cases.

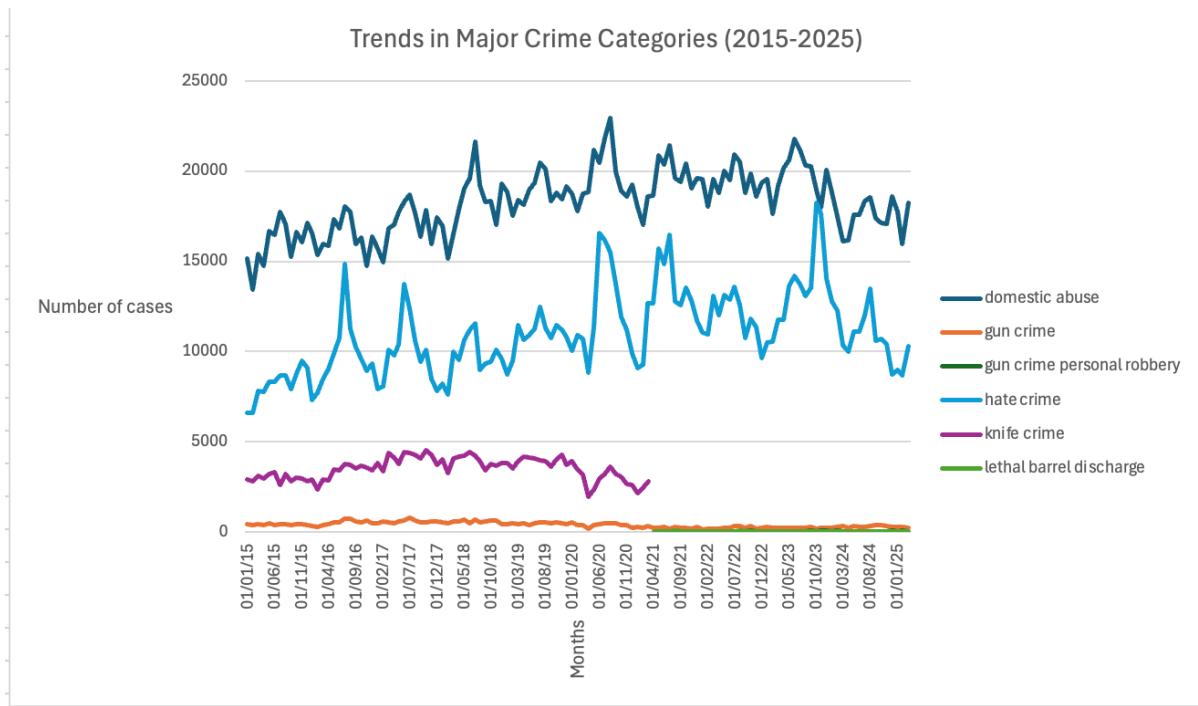


Figure 2. Column chart showing simulated resolution rates across different crime types from 2015 to 2025. The visualization emphasizes variation in outcome success, with Fraud and Hate Crime demonstrating relatively higher estimated resolution rates than Gun Crime and Knife Crime.

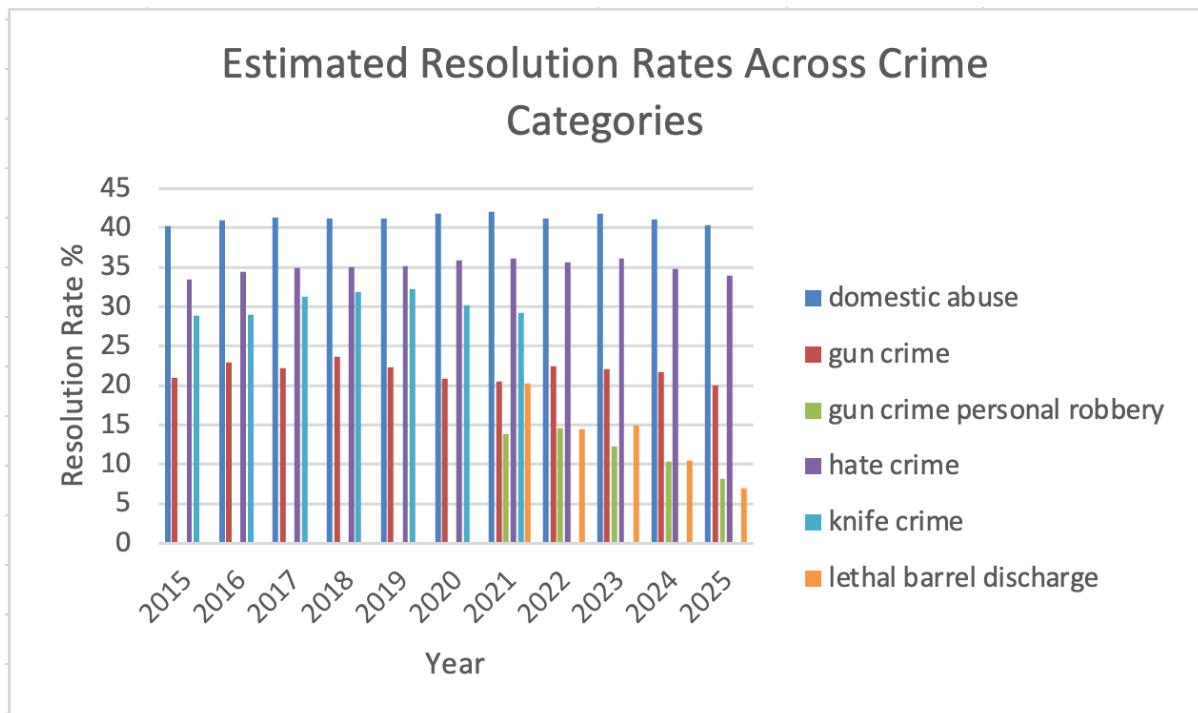


Figure 3a :Monthly trends of selected crime subtypes from 2015 to 2025, visualised using a multi-line chart [4] in Excel [6]. Each line represents a subtype from domestic abuse, hate crime, knife crime, or gun crime categories.

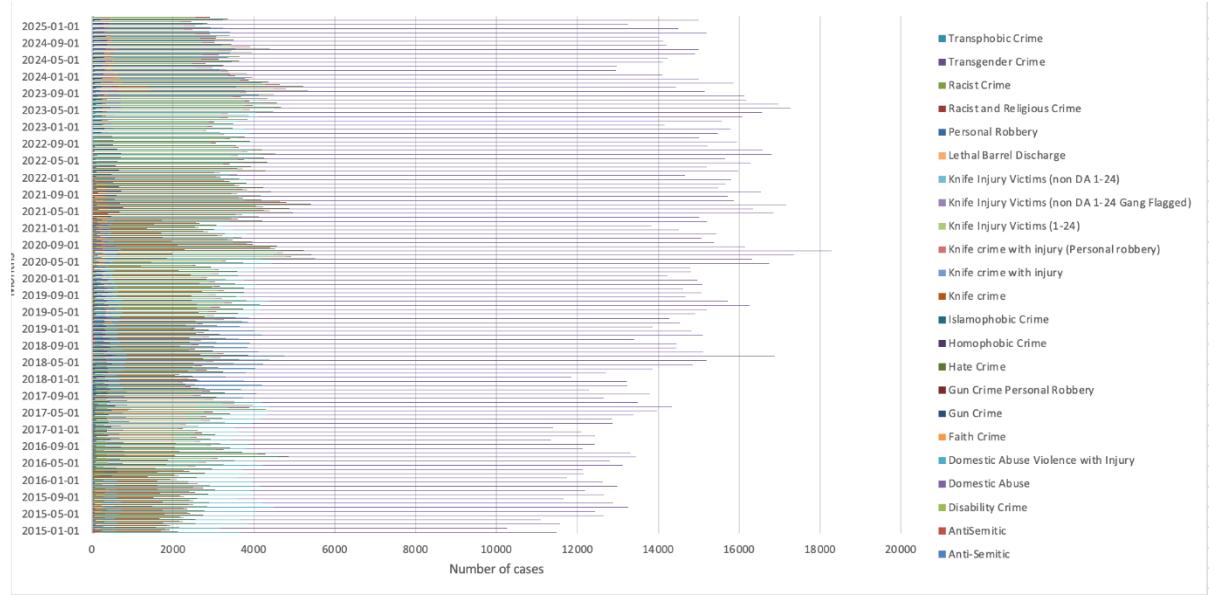


Figure 3b: Initial Heatmap for Subtype Intensity Over Time

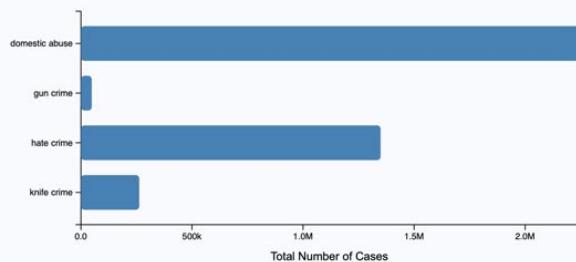
This exploratory heatmap was created to assess the monthly distribution of crime subtypes across years. Colour intensity reflects the concentration of incidents, helping identify periods of subtype spikes or dips.

1	Crime Subtype	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
2	Anti-Semitic	922	1024	1024	1132	1262	1172	1666	1079	3163	2725	440
3	Disability Crime	510	1514	844	816	918	1036	1192	1169	1201	1009	227
4	Domestic Abuse	145108	149660	156316	171058	179024	189510	188142	188570	190110	170783	42729
5	Domestic Abuse Violence with Injury	46764	47458	48242	47978	47494	47752	44924	45139	47525	40081	9217
6	Faith Crime	3392	4074	4946	4438	4414	3800	4369	3735	6871	6233	1203
7	Gun Crime	3762	4604	5166	4928	4206	3436	2744	2823	2946	3720	804
8	Gun Crime Personal Robbery	0	0	0	0	0	0	404	515	578	602	147
9	Hate Crime	32114	38652	38638	38022	43176	48878	50495	47159	52132	43248	9142
10	Homophobic Crime	3564	4022	4194	4616	5852	5844	6994	6911	7341	5565	1266
11	Islamophobic Crime	2104	2448	3324	2608	2366	1756	1881	1808	2689	2598	551
12	Knife Crime	19958	22264	29082	29394	31206	23760	4618	0	0	0	0
13	Knife Crime with Injury	7392	8222	9802	8786	7976	6424	1456	0	0	0	0
14	Knife Crime with Injury (Personal Robbery)	892	1010	1200	1232	1106	910	158	0	0	0	0
15	Knife Injury Victims (1-24)	3588	3986	4606	3880	3564	2656	624	0	0	0	0
16	Knife Injury Victims (non DA 1-24 Gang Flagged)	558	384	286	154	106	92	24	0	0	0	0
17	Knife Injury Victims (non DA 1-24)	3330	3716	4280	3654	3322	2418	560	0	0	0	0
18	Lethal Barrel Discharge	438	668	704	876	540	614	437	416	354	297	58
19	Personal Robbery	776	974	1112	1210	958	656	118	0	0	0	0
20	Racist Crime	26862	31744	31506	31116	34994	41148	42636	39252	42975	33743	7050
21	Racist and Religious Crime	28336	33582	34014	33084	36976	42686	44250	40889	45258	37801	7930
22	Transgender Crime	302	376	374	426	566	576	118	0	0	0	0
23	Transphobic Crime	0	0	0	0	0	0	635	788	1053	649	133



Most Prevalent Crime Types

Click a bar to explore trends and subtypes.



Subtype Distribution – Knife Crime

Each box represents a subtype proportionate to total cases (2015–2025)



Figure 3c: Subtype analysis workflow demonstrated using Knife Crime.

- (a) Overview of top 4 crimes (2015–2025).
- (b) Monthly trends of Knife Crime subtypes.
- (c) Proportional subtype distribution via treemap [5].