

A PYTHON EDA PROJECT

---

# Crime Patterns Around the M25: An Exploratory Data Analysis

---

*Prepared By: Zharys Consolacion*

Client: Nadine Green

[Project Repository on GitHub](#)

Date: November 10, 2025

# 1 Executive Summary

This report examines monthly crime-rate (per 10,000 people) across the Police Forces that work around the M25. Its purpose is to provide supporting evidence on locations to live, a short train distance from central London. Data on Street-level crimes and Stop & Searches provided by data.police.uk were used in this analysis. Summary statistics were produced and displayed through boxplots, and temporal trends of the data were analysed for the period September 2023 - October 2025. Hotspots of crime in different regions were visualised.

## Key findings include:

- Thames Valley Police and Surrey Police presented with the lowest crime rates (66.28 and 59.93 per 10,000 people respectively).
- Crime rates across all forces showed simultaneous rises and falls, suggesting region-wide influences (e.g seasonal effects or socio-economic conditions).
- Violence and sexual offences were the most prevalent category of crime, with Kent Police and Essex Police having the highest crime rate in this category.
- Thames Valley Police had the highest number Stop & Searches being performed despite mid-low level crime rates. Surrey Police performed the least number of Stop & Searches.

Thames Valley Police and Surrey Police are put forward as recommendations of further analysis to better understand the underlying drivers of their relatively low crime rates and contrasting policing operations. Deeper breakdowns by neighbourhood, crime type and socio-economic indicators can also be explored to further refine housing-location advice.

## Contents

1	Executive Summary	2
2	Introduction	4
3	Methodology	5
3.1	Data Description . . . . .	5
3.2	Data Pre-processing . . . . .	5
3.2.1	Loading and Joining of CSVs . . . . .	6
3.2.2	Street-level Data . . . . .	6
3.2.3	Stop and Search Data . . . . .	7
3.3	Anomalies . . . . .	7
3.4	Workflow . . . . .	7

<b>4</b>	<b>Exploratory Data Analysis</b>	<b>8</b>
4.1	Univariate Analysis . . . . .	8
4.2	Multivariate Analysis . . . . .	9
4.2.1	Monthly Crime Rate by Police Force . . . . .	9
4.2.2	Crime Type . . . . .	9
4.2.3	Stop & Search Analysis . . . . .	10
4.2.4	Hotspot Analysis . . . . .	11
<b>5</b>	<b>Key Insights and Findings</b>	<b>12</b>
<b>6</b>	<b>Recommendations</b>	<b>13</b>
<b>7</b>	<b>References</b>	<b>14</b>
<b>A</b>	<b>Trello Board</b>	<b>14</b>

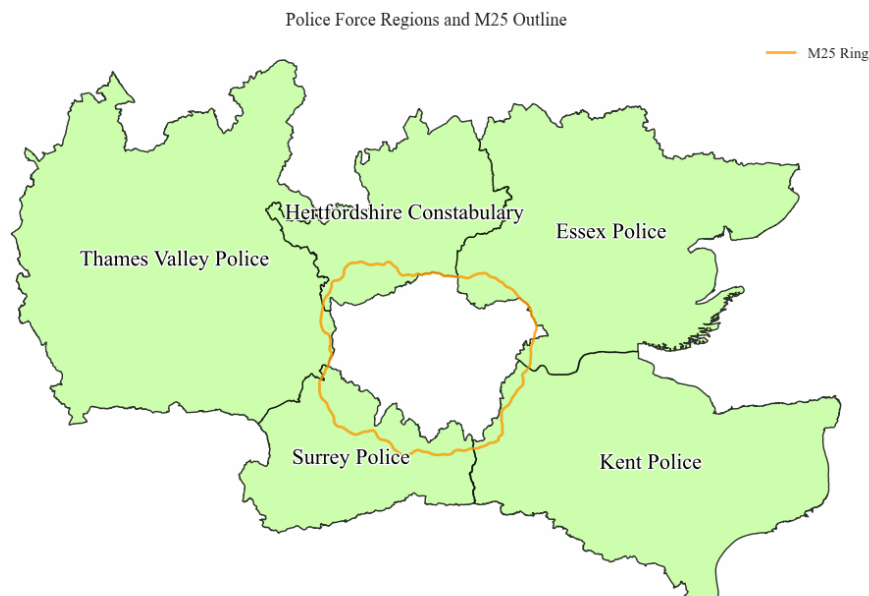
## 2 Introduction

This project is an analysis on the crime in the UK. The stakeholder, Nadine Green has requested this to give insight on which locations are most and least desirable to live in. The purpose is to give insight which will lead to more informed decisions around sales within the real estate company.

This analysis focuses on the crime patterns of the Police forces that work around the M25. This includes the following forces,

- Essex Police
- Kent Police
- Surrey Police
- Hertfordshire Constabulary
- Thames Valley Police

A visual on the regions we will be looking at are found within figure 1.



**Figure 1:** Police force region visual

Further, this analysis focuses on the time period October 2023 - September 2025. This is to give insight on what crime has been like in the short term period.

This project aims to address the following 3 objectives,

- What crimes are the most prevalent in each region?
- Which regions have the most crimes reported? Over time?
- Are there any crime hotspots in these regions?

## 3 Methodology

### 3.1 Data Description

The main source of data is data.police.uk [1], and the following data files are taken from this site,

- Street-level crime data - A log of crimes happening in the jurisdiction of each force.
- Stop and search data - A log of all stop and searches carried out by each force.
- Force boundaries - Kml files defining the boundaries of the UK in which each of the forces serve

Other data sources outside data.police.uk are used. These include the following,

- Population statistics - the population that each police force serves. This data was found on HMICFRS [2]
- M25 boundary - this data was found on OpenData [3]

### 3.2 Data Pre-processing

Data cleaning and pre-processing contains the following main steps

1. Loading and joining of CSVs to form one for Street-level data and one for Stop & Search Data
2. Creating copies of raw-data after joining
3. For each CSV locate NaN values, and where appropriate
  - Drop rows or columns
  - Replace values
4. Locate anomalies

This report will summarise briefly the steps taken for pre-processing. A more in depth description pre-processing steps can be found within Crime-Data-Preprocessing&Cleaning Jupyter Notebook.

### 3.2.1 Loading and Joining of CSVs

Data downloaded has the following structure,

```
raw-data/
├── 2023-10/
│   ├── 2023-10-essex-stop-and-search.csv
│   ├── 2023-10-essex-street.csv
│   └── (other regions)...
└── (other year-month combinations)...
```

The aim is to join all data grouping them only based on whether they are Street level data or Stop and search data. The end result will look like,

```
combined-data/
├── combined-allregions-stop-and-search.csv
└── combined-allregions-street.csv
```

From this process, 240 csv files will become 2. Resulting files will be too big to open up in Excel, therefore cleaning and pre-processing is further continued on Python.

### 3.2.2 Street-level Data

The following table 1 shows the distribution of NaNs across different columns. **Note:** Columns that had not NaNs within them are not included.

**Table 1:** Proportion of Missing Values by Police Force and Variable

Police Force	Crime ID (%)	Long/Lat & LSOA(%)	Last outcome category (%)	Context (%)
Essex Police	9.08	1.63	9.08	100
Hertfordshire Constabulary	21.53	2.16	21.53	100
Kent Police	15.26	1.15	15.26	100
Surrey Police	15.78	1.38	15.78	100
Thames Valley Police	10.15	2.14	10.15	100

To address the NaNs in the data, the following steps were made,

1. Rows missing geospatial data (i.e **Longitude**, **Latitude** and **LSOA name**), were dropped, as this type of data was key in our later analysis.
2. Column **Context** were dropped as we did not need this data for our analysis.
3. For **Last Outcome Category** NaNs were replaced appropriately with 'Status update unavailable'.
4. **Crime ID** was kept as signified unique identifiers for each entry, rows missing this were not removed as took up a significant proportion of data

### 3.2.3 Stop and Search Data

The following tables 2 and 3 shows the distribution of NaNs across different columns.

**Note:** Columns that had not NaNs within them are not included.

**Table 2:** Missing Values for Stop & Search by Region — Part A

Region	Longitude/Latitude (%)	Gender (%)	Age range (%)	Self-defined ethnicity (%)
Essex Police	0.79	8.16	23.53	8.16
Hertfordshire Constabulary	18.11	4.07	8.11	4.81
Kent Police	0.88	4.40	10.68	6.29
Surrey Police	0.53	2.21	10.41	14.70
Thames-Valley Police	6.61	2.21	7.80	1.45

**Table 3:** Missing Values for Stop & Search by Region — Part B

Region	Officer-defined ethnicity (%)	Legislation (%)	Object of search (%)	Removal of more than outer clothing (%)
Essex Police	8.39	0.65	100.00	100.00
Hertfordshire Constabulary	8.48	0.00	0.00	0.03
Kent Police	4.34	0.09	0.00	0.00
Surrey Police	1.70	0.00	22.46	100.00
Thames-Valley Police	1.45	0.01	0.03	1.51

To address the NaNs in the data, the following steps were made,

1. For geospatial data (i.e **Longitude and Latitude**) - NaNs were replaced. Hertfordshire Constabulary had a disproportionate amount of data missing in comparison to other regions. Replacements of NaNs made using the following steps,
  - (a) Finding appropriate donor values from other rows
  - (b) Applying random noise to these values to prevent overplotting
2. The following columns were dropped - **Gender, Age range, Self-defined ethnicity, Officer-defined ethnicity, Legislation, Object of search, Removal of more than outer clothing**

## 3.3 Anomalies

Within the Street-level data, there were instances where crimes happening in one police region would be followed up or transferred to a different police force. In these cases, transfers would have geospatial data that would lie outside of the region. These points would result in These points were filtered out using the following steps,

- Loading in Kml files for each region
- Filtering out data that did not lie within the bounds of each region

## 3.4 Workflow

Depicted in figure 2 is the workflow of data.

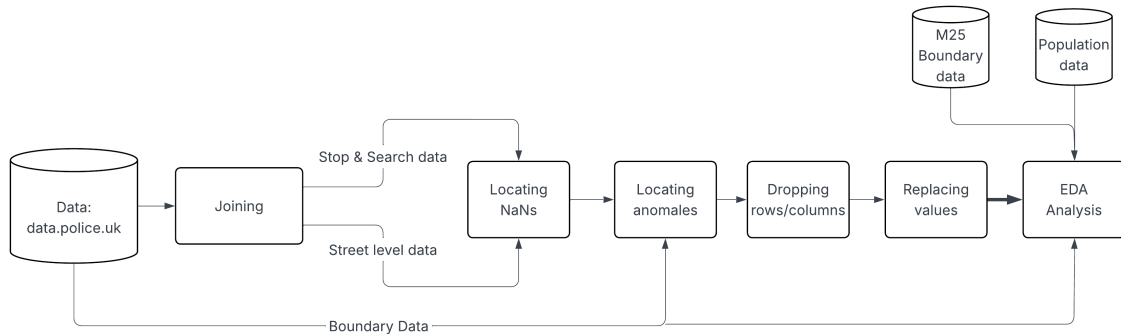


Figure 2: Workflow of Data

## 4 Exploratory Data Analysis

To begin the exploratory analysis, the number of recorded crimes was normalised by the population within each Police force region. Crime rate was expressed as Crimes per 10,000 residents. This adjustment ensured that crime rate was comparable across Police Force regions with different population sizes.

### 4.1 Univariate Analysis

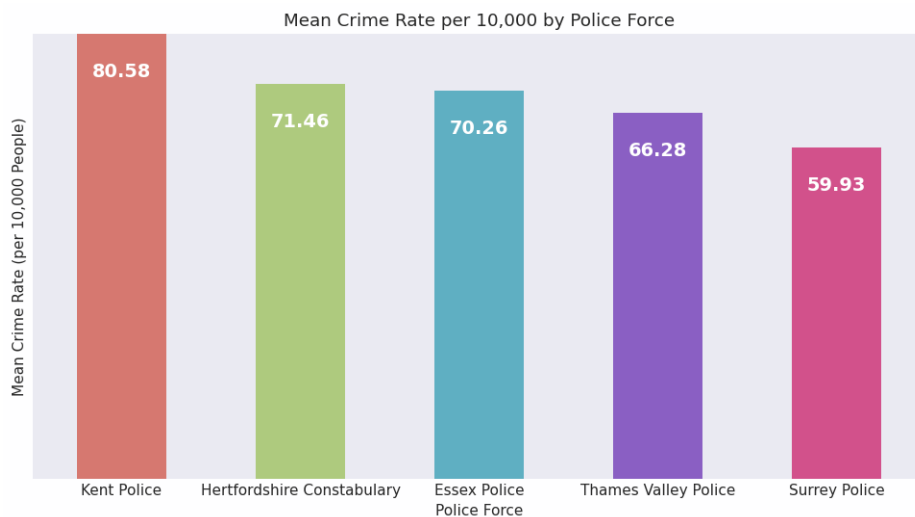


Figure 3: Mean Crime Rate (per 10,000 persons) by Region

Figure 3 highlights the mean crime rate for each police force over the given period. Crime rate is calculated individually for each month and averaged. Figure 4 displays a boxplot for each police force. This displays the summary statistics of Crime rate, indicating the mean, IQR, minimum and maximum.

Overall, these display Kent Police to have the highest average crime rate and Surrey Police to have the lowest among the group.



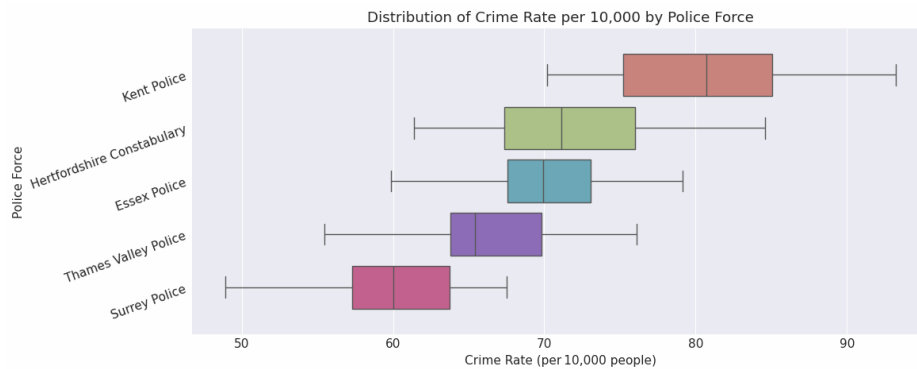


Figure 4: Boxplot of Crime rate by Region

## 4.2 Multivariate Analysis

### 4.2.1 Monthly Crime Rate by Police Force

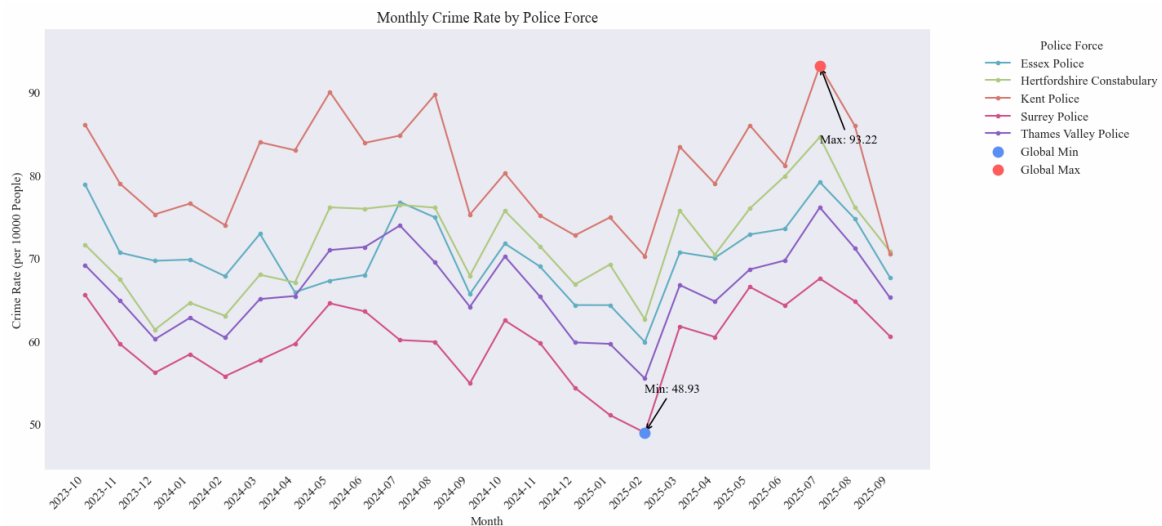


Figure 5: Crime Rate Over Time by Region

Significantly, within figure 5, the Monthly Crime Rate by Police Force. Further, the global minimum and maximum overall are highlighted. This depicts crime rate to be quite volatile around an equilibrium for each Police Force. This also displays simultaneous rises and falls in rates across all Police Forces, which suggest that a shared external factor, rather than force specific policing actions alone may be driving underlying dynamics.

### 4.2.2 Crime Type

As seen within figure 6, which displays the total counts of all crimes by Crime types, Violence and sexual offences are clearly the most prevalent category of crime. This is followed by Anti-social behaviour, which carries fewer than half the number of incidents compared to violence and sexual offences. The next most frequent categories

are Shoplifting, Criminal damage and arson, and Other theft.

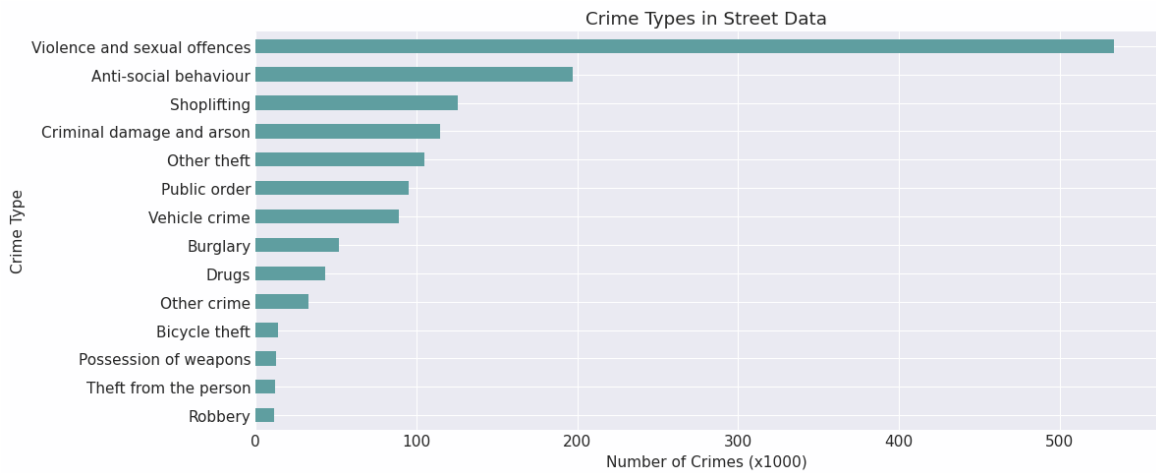


Figure 6: Crime Types in Street Data

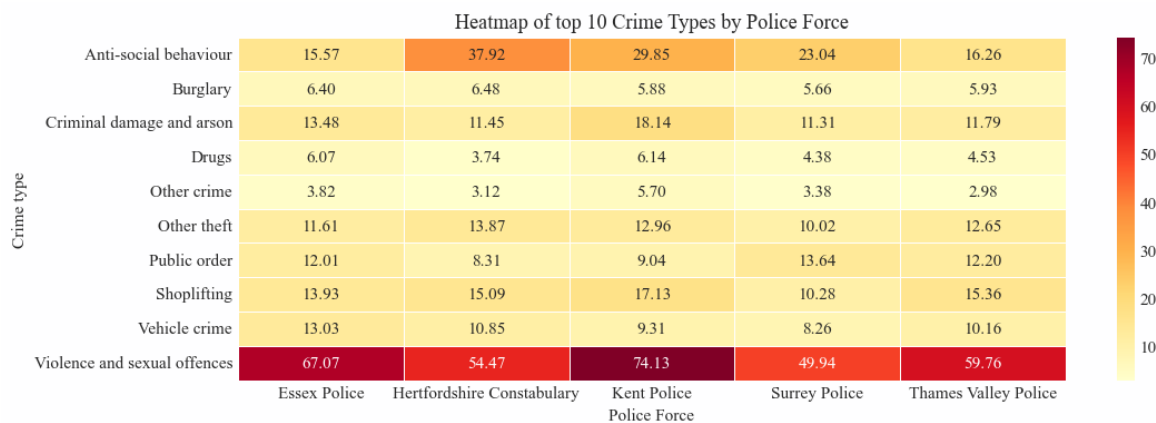
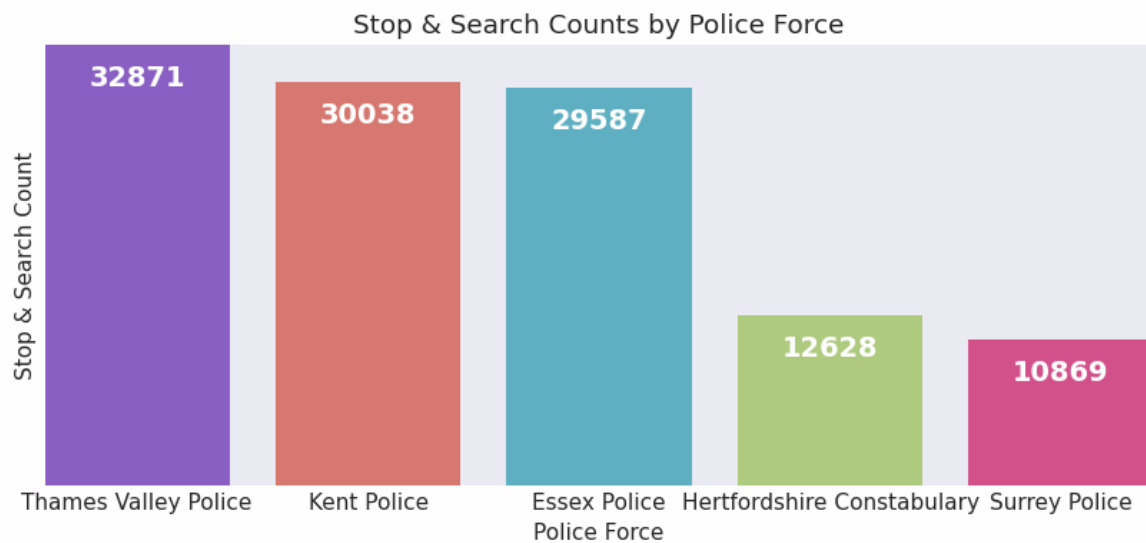


Figure 7: Top 10 Crime Types by Police Force heatmap

As shown in figure 7 (Top 10 Crime Types by Police Force heatmap), Kent Police recorded the highest rate of Violent and sexual offences at 74.13 per 10,000 with Essex Police coming in second with 67.07 per 10,000. By contrast, Surrey Police had the lowest rate in this category, standing at 49.94 per 10,000. In the Anti-social behavior category, Herfordshire Constabulary had the highest rate, with Kent Police coming in a close second. Lowest rates in this category come from Essex Police and Thames Valley Police.

### 4.2.3 Stop & Search Analysis

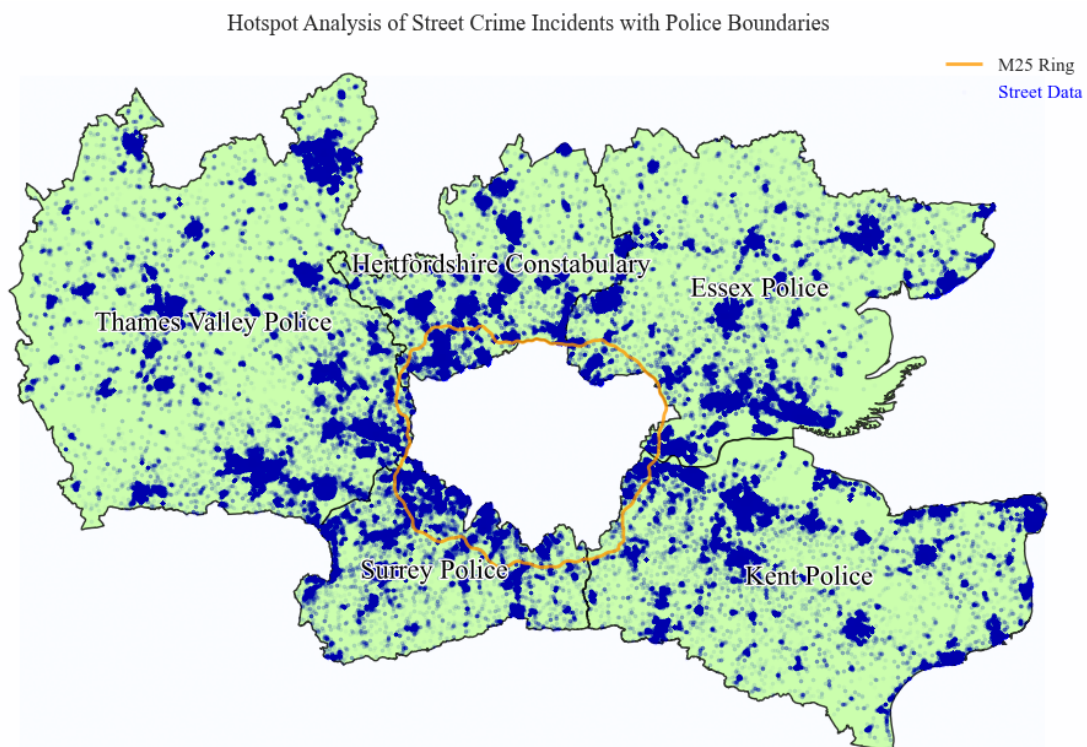
Within figure 8, displays the total number of Stop & Searches occurring over the period, grouped by Police Force. It displays Thames Valley Police to perform the highest number of stop and searches at roughly 32,000 across the two years, whereas



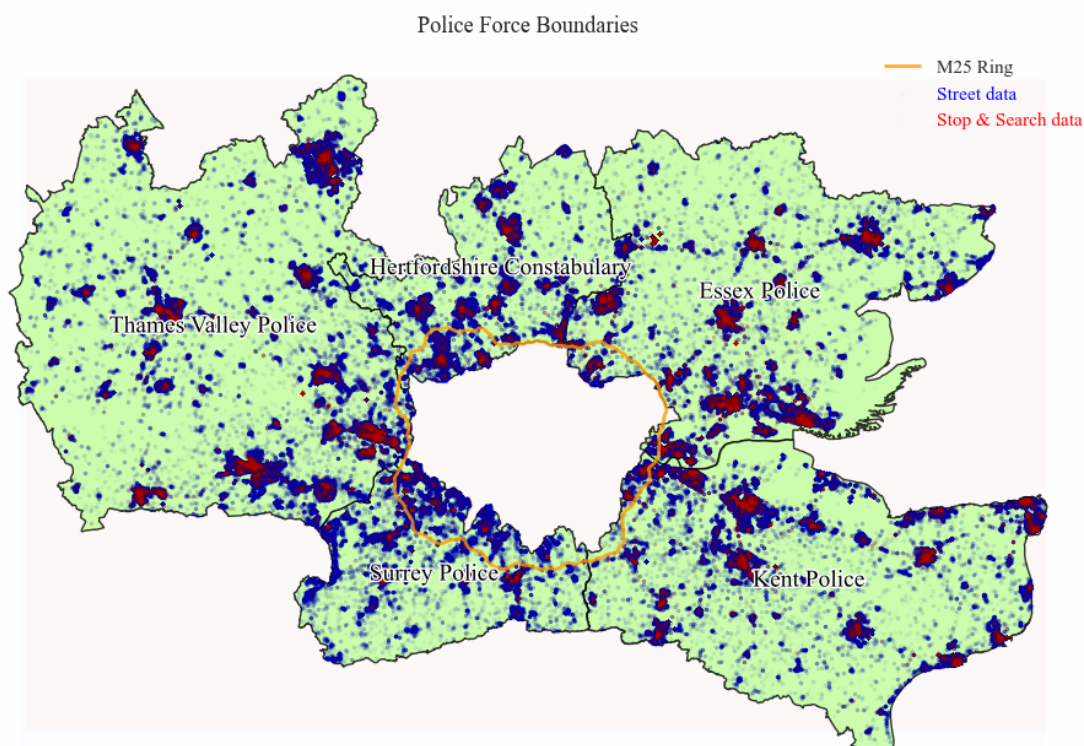
**Figure 8:** Stop & Search counts by Police Force

Surrey Police only performed roughly 11,000, just over a third of Thames Valley Police.

#### 4.2.4 Hotspot Analysis



**Figure 9:** Hotspot Map with Street-level data



**Figure 10:** Hotspot Map with Street-level and Stop & Search data

Figure 9 maps the geospatial coordinates provided by each entry of the crime data, and plots it onto our map of regions, giving a visual representation of where crimes have happened and further highlights any potential hotspots of repeated activity. By providing spatial context, this helps agents make more informed decisions for their clients, identifying areas with lower incident volume and advising accordingly on safe places to live.

Further, figure 10 maps the geospatial coordinates of Stop & Search data, indicating police presence within the area. This overlay gives a clearer view of whether resources are being deployed effectively to address identified crime hotspots.

## 5 Key Insights and Findings

Key insights from this analysis can be summarised as,

- Thames Valley Police and Surrey Police presented with the lowest mean crime rates, 66.28 and 59.93 per 10,000 respectively.
- Crime rates across all forces showed simultaneous rises and falls, suggesting a shared external factor.
- Violence and sexual offences were the most prevalent category of crime, followed by Anti-social behaviour.

- Kent Police and Essex Police had the highest rate of Violence and sexual offences crime happening across the period. Surrey Police had the lowest.
- Thames Valley Police had the highest number of Stop & Searches being performed. Surrey Police had the lowest, with its count being just over a third of Thames Valley Police's count.

## 6 Recommendations

Given the key findings and analysis, Thames Valley Police and Surrey Police are put forward as recommendations for further analysis. Surrey Police overall, produced the lowest mean Crime rate, at 59.93 crimes per 10,000 persons despite displaying a low police presence through their Stop & Search data. Thames Valley Police came close second at 66.28 crimes per 100,000 persons. Thames Valley Police also displayed to be most active in terms of Stop & Searches, providing confidence through their strong police presence.

Further analysis, into both Police Forces can hopefully lead to a more informed decision in regards to which of the two are better. This can happen through possible analysis of outcome data from the two forces and maybe a further look into the hotspots of crime within the two regions.

## 7 References

- 1.Data.Police.UK. Data downloads — data.police.uk [Internet]. Police.uk. 2024. Available from: <https://data.police.uk/data/>
- 2.HMICFRS. Police forces [Internet]. His Majesty's Inspectorate of Constabulary and Fire & Rescue Services. 2025. Available from: <https://hmicfrs.justiceinspectorates.gov.uk/police-forces/>
- 3.Motorway Boundary Files – Fusion Data Science Open Data [Internet]. Fusion-datascience.com. 2019 [cited 2025 Nov 9]. Available from: <https://opendata.fusiondatascience.com/2019/02/16/motorway-boundary-files/>

## A Trello Board

This section displays figures showcasing the state of the Trello board over time. Further, the Trello board can be accessed through the link.

<https://trello.com/invite/b/690a57bc5451bf2609711f8f/ATTI2d6c5b1e59bdcd54e88d95d604b91/crime-data-exploratory-data-analysist-management>

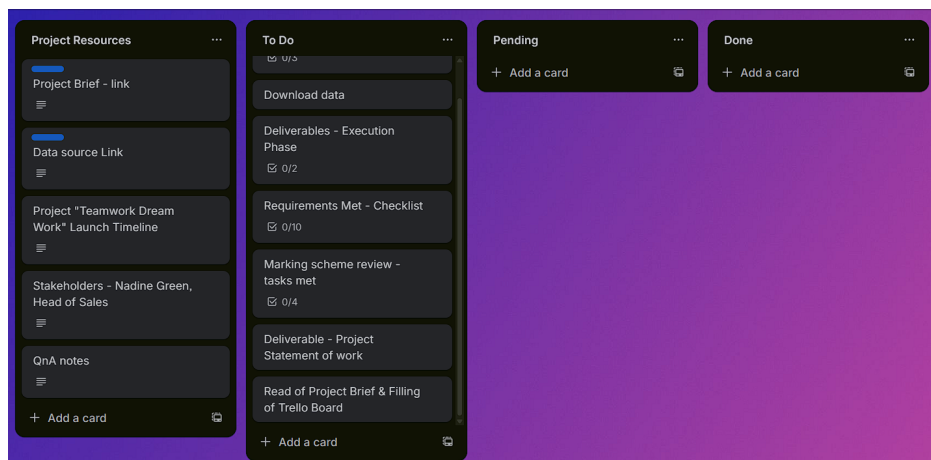


Figure 11: Beginning of Trello Board

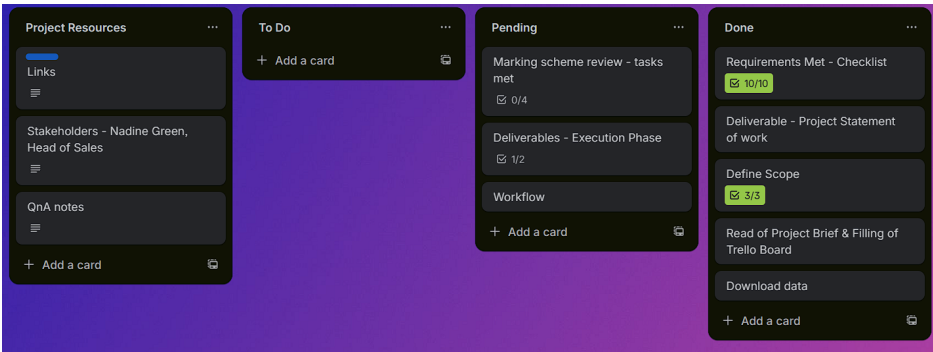


Figure 12: Midway of Trello Board pt 1

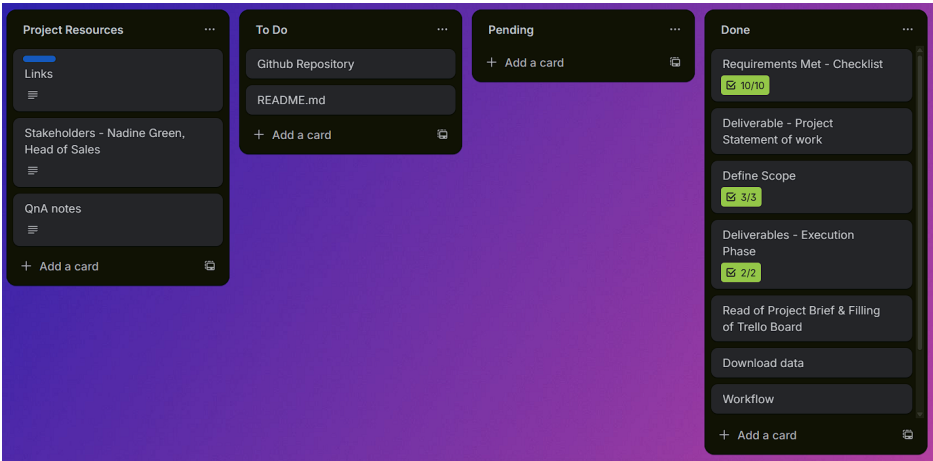


Figure 13: Midway of Trello Board pt 2

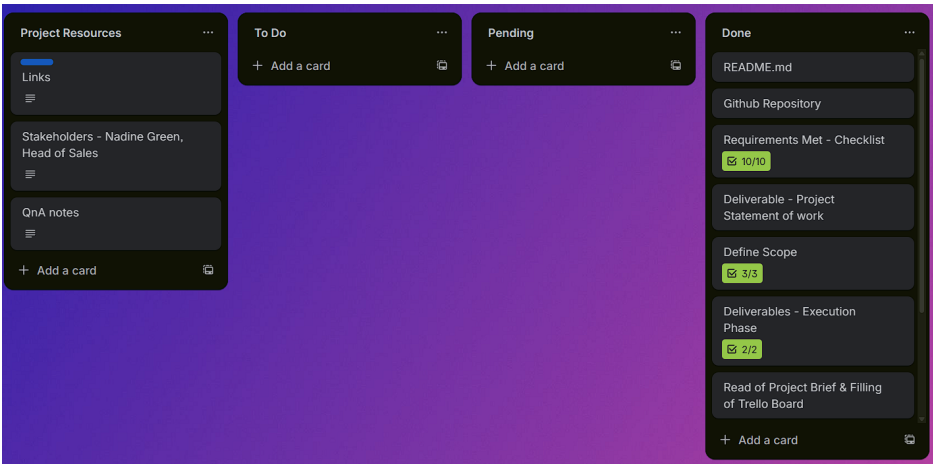


Figure 14: End of Trello Board