**Executive Summary**

This report investigates the spatial, socioeconomic, and predictive patterns of crime across four UK police forces- Metropolitan Police, City of London Police, West Midlands Police, and West Yorkshire Police- to support strategic real estate decision-making. Using street-level crime data (Feb 2023–Feb 2025) merged with the English Indices of Deprivation (2019), the analysis offers a granular, LSOA-level perspective on where crime is most concentrated, how it correlates with deprivation indicators, and how these trends are forecasted to evolve.

**Key Findings:**

* Crime is spatially concentrated in urban cores, especially the City of London, Westminster, Birmingham, and Manchester. Conversely, rural and suburban areas consistently report lower crime volumes.
* Income and education deprivation were the strongest correlates of crime; areas with multiple overlapping disadvantages experienced the highest crime frequencies.
* Surprisingly, poor service access correlated with lower crime, likely due to underreporting in rural or disconnected zones.
* The City of London recorded the highest crime per LSOA, yet also had the highest resolution rate, pointing to dense commercial activity coupled with effective policing.
* Predictive modelling revealed stable upward trends in crime, especially in West Midlands and West Yorkshire, suggesting emerging investment risk if left unaddressed.

**Strategic Real Estate Recommendations:**

* **High-crime areas** (e.g. inner London, Manchester): Target investors and corporate tenants with secure, high-density developments like build-to-rent flats with concierge and surveillance.
* **Moderate-crime transitional zones** (e.g. West Midlands): Promote regeneration schemes and value-add properties, particularly in areas with forecastable crime growth.
* **Low-crime suburbs and rural areas**: Market family homes and long-term ownership opportunities to risk-averse buyers, highlighting school quality and green amenities.
* **Deprivation-aware targeting**: In high-deprivation zones, focus on institutional buyers or social landlords. Avoid marketing to families in violent crime hotspots.
* **Crime resolution matters**: High-crime/high-resolution areas (e.g. City of London) can offer stable investment environments if paired with infrastructure and enforcement. Conversely, low-resolution zones (e.g. Metropolitan) require stronger buyer risk disclosure.

**Limitations:**

* Socioeconomic data used for modelling (IoD 2019) may not reflect post-pandemic realities.
* Missing values in deprivation indicators were imputed using force-level averages, which may obscure hyperlocal variation.
* Predictive models were trained on only one year of data, limiting seasonality detection and long-term forecasting.
* Some crime types and resolution details were not included due to dataset limitations.

**Next Steps:**  
The report recommends further modelling on two contrasting jurisdictions:

* **City of London**: To understand how concentrated crime interacts with high resolution and dense commercial land use.
* **West Midlands**: To explore scalable predictive strategies in a post-industrial, socioeconomically diverse region facing steady crime growth.

Together, these focal points offer a foundation for data-informed, risk-adjusted property listing, pricing, and development strategies tailored to both buyers and institutional stakeholders.

**Introduction**

**Background and Literature Context**

Crime is an influential factor in urban development and real estate dynamics. A substantial body of literature has demonstrated that increased crime rates are consistently associated with declining property values, reduced buyer interest, and lower long-term community investment (Gibbons, 2004; Ihlanfeldt & Mayock, 2010; Tita et al., 2006). Crime functions as a major deterrent in housing markets, influencing both perceptions of neighbourhood safety, accelerating residential turnover (Linden & Rockoff, 2008; Gibbons, 2004). However, this relationship does not occur in isolation; it is often mediated by broader socioeconomic indicators such as income deprivation, educational opportunity, and access to public services.

Income deprivation has long been identified as a structural determinant of both crime and housing devaluing. Areas with high levels of economic disadvantage often experience elevated crime rates, which in turn deter investment and depress housing demand (Cheshire & Sheppard, 2004). These trends are further exacerbated by limited access to quality public services. For example, inadequate transport connectivity, poor access to healthcare etc. contributes to higher crime rates and disinvestment (Kelaher et al., 2010; Wikström & Loeber, 2000).

Educational access is another critical factor in the crime-real estate relationship. Research has repeatedly shown that low educational attainment and the presence of underperforming schools are strongly correlated with increased youth crime and long-term delinquency (Lochner & Moretti, 2004; Deming, 2011).

Together, these factors- crime, income deprivation, service access, and educational opportunity, interact to shape both the lived experience of urban residents and broader real estate market trends. Crucially, Nadine Green, Head of Sales at our real estate firm, is interested in identifying high-risk and low-risk property zones to optimise listing decisions. Therefore, this analysis will cover only descriptive and predictive exploratory analysis.

**Research Questions I will cover in my EDA:**

* Which Local Super Output Areas (LSOAs) consistently report higher/lower crime volumes?
* How does socioeconomic deprivation (income, education, environmental, and homelessness rates) correlate with local crime indexes and crime types?
* How does accessibility to services (GPs, schools, supermarkets) affect crime incidence?
* What are the most common crime types per police force?
* Which jurisdictions have the best crime resolutions/ outcomes?
* How is future crime forecasted to increase per police jurisdiction?

**Methodology**

**Data Sources**

This project integrates two complementary data sources: street-level crime data (Feb 2023 – Feb 2025) from [data.police.uk](https://data.police.uk/data/), and the English Indices of Deprivation (2019), focusing on domain-level ranks and deciles.

**1. Street-Level Crime Data (data.police.uk)**

This dataset provides time-stamped records of individual crimes reported to UK police forces. By spanning four distinct police forces and a two-year window:

The four police forces were strategically selected:

* **Metropolitan Police** (Greater London): High-density urban area with complex socioeconomic stratification.
* **City of London Police**: Financial hub, small geography but uniquely low-volume high-impact crimes.
* **West Midlands Police**: Post-industrial city with diverse income and infrastructure disparities.
* **West Yorkshire Police**: Mix of rural and urban profiles with known property crime hotspots.

These regions together offer a cross-sectional representation of UK urban crime diversity.

**2. English Indices of Deprivation (IoD, 2019)**

The IoD dataset offers a multi-dimensional profile of each LSOA, including: Income Deprivation, Education Deprivation, Crime Deprivation and more. Allowing analysis to go beyond crime incidents by incorporating structural socioeconomic conditions that both cause and are affected by crime.

**3. LSOA-Based Integration**

Merging datasets via LSOA ensures spatial accuracy in cross-referencing crime frequency with underlying deprivation. This supports:

* Context-aware recommendations for property development or withdrawal
* More meaningful insights for Nadine’s sales strategy, such as targeting undervalued areas with improving trends or avoiding high-crime, structurally deprived zones

**Preprocessing Overview**

The pre-processing pipeline was designed to ensure the integrity, usability, and spatial alignment of UK street-level crime data enriched with deprivation indicators. The following steps were executed in Python within a modular Jupyter notebook:

**1. Data Ingestion and Consolidation**

* Crime data was structured in monthly folders, with CSVs for each police force.
* A nested loop was implemented using pathlib.Path to iterate through months and forces.
* For each file:
  + Metadata (month and force name) was extracted from filenames.
  + Each CSV was loaded into a Data Frame with appended Month and Force columns.
* All monthly Data Frames were concatenated using pd.concat() to form a master dataset.
* Deprivation and civic infrastructure datasets were loaded separately, and LSOA identifiers were renamed for consistency.

**2. Merging External Metadata**

* Using outer-joining, the deprivation and service access datasets were joined on LSOA code.
* This metadata was left-joined to the crime street Data Frames, ensuring no crime records were lost while adding contextual indicators.

**3. Initial Cleaning**

* Columns with high nullity, redundancy (e.g., location coordinates, "context"), or irrelevance to modelling were dropped.
* Object-type columns representing numerical values (e.g., deprivation ranks/deciles) were converted using pd.to\_numeric(..., errors='coerce').
* Crime type and Force were cast to the category data type.
* Month strings were parsed into datetime objects via pd.to\_datetime() to facilitate time series analysis.

**4. Handling Missing and Duplicate Data**

* Rows with missing LSOA code (<1%) were dropped, as mean imputation depended on geographic anchors.
* Missing Crime ID entries were filled using uuid.uuid4() to preserve row uniqueness.
* Full-row duplicates were removed using df.drop\_duplicates().
* Remaining duplicate Crime IDs, resulting from police data reuse were handled by generating unique row-level UUIDs.

**5. Imputation of Socioeconomic Variables**

* Rank and decile fields with nulls were first imputed using group-level means by LSOA code.
* If LSOA group-level mean data was unavailable or incomplete, a fallback imputation was conducted using Force-level means via groupby().transform().
* Variables still containing missing values post-imputation were removed from the final dataset.

**6. Feature Engineering**

* **crime\_count**: Calculated via value\_counts() on LSOA code to derive crime volume per area.
* **lagged\_crime\_count**: Introduced temporal dynamics by shifting crime\_count within each force group using .shift(1), creating monthly lagged crime frequency variables (useful for machine learning prediction).
* **service\_access**: Computed as the average road distance from each LSOA to four services, GP, school, supermarket, and post office.
* **crime\_category**: A new variable that aggregated detailed crime types into broader conceptual groups (e.g., “Violent”, “Property”) for interpretability.
* **Decile/Rank Rescaling**: All ranks and deciles were reversed e.g., crime deprivation/ income deprivation (e.g., 11 - decile) as the original data meant low values = high deprivation. This was done to ensure higher values consistently denoted worse conditions, improving interpretability for visual and statistical analysis.

**7. Export**

* The fully cleaned and engineered Data Frame was exported as final\_data.csv using df.to\_csv(..., index=False) for downstream modelling and visualization.

**Exploratory Data Analysis Overview:**

The structured workflow includes five analytical tiers: Univariate Analysis, Bivariate Analysis, Grouped Analysis, Multivariate Analysis, Domain specific analysis, and Predictive Modelling.

1. **Univariate analysis**

For my univariate analysis I computed descriptive statistics on crime\_count (analysing overall crime frequency). This analysis examined the skew, kurtosis, mean, median and so forth using describe(). This analysis examined mean crime frequency across each 4 police forces using groupby(), allowing for descriptive statistics on a force level aggregation. A visualisation was created using a faceted histogram plot using seaborn and sns.FacetGrid() to visualise the crime frequency distribution per police jurisdiction.

Moving on, analysis examined the highest/ lowest crime LSOA’s to pinpoint what specific areas had the most/least crime using a combination of groupby() and sort\_values(ascending =True/False). These variables were visualised using this with sorted bar charts using sns.barplot(). I finished by analysing/ visualising the most frequent types of crime overall using sns.countplot() and value\_counts(), and what crimes were most common within each police force using sns.catplot().

1. **Bivariate analysis**

The process began by conducting a correlation matrix between crime deprivation deciles and income, education, and environmental deprivation. Alongside service access and homelessness rates using the corr() method. These variables were visualised using a line plot using sns.lineplot().

1. **Grouped analysis**

Following univariate and bivariate analysis, I conducted some more advanced grouped analysis. This analysis examined crime resolution rate per police force, by creating a temporary variable of resolved/unresolved crime outcome categories, transforming them into a binary value (1= resolved crime, 0= unresolved crime) using astype(int). This step was visualised as a table using packages including IPython.display, then as a bar chart using sns.barplot() to analyse the crime resolution rate by force.

Lastly, this analysis examined the top crime categories by income deprivation tiers. A visualisation was created using 3 tiers of income deprivation categorised low/ medium/ high by creating custom functions implementing if/ elif/ else statements and applying this to my dataframe using apply(). This allowed me to group by income tiers and crime categories, visualising the top 3 crime categories (e.g., violent crime, property crime, anti-social crime) by low, medium, or high-income deprived areas using sns.barplot().

1. **Multivariate analysis**

Multivariate analysis I conducted was an extension of the prior bivariate analysis I conducted- mainly analysing interactions between strong correlators of crime simultaneously. Focused on analysing the interaction between income/ education/ and crime deprivation deciles These variables were visualised using a heatmap, using scatter() and a line plot. To visualise my line plot I first created a temporary variable using a custom function involving an if statement, to categorise education into 3 bins- low, medium and high, and applied this to my data frame using apply(). I then used groupby(), and sns.lineplot() to group and visualise education bands against income and crime deprivation.

1. **Domain specific analysis**

The domain-specific analysis involved a crime map and time series visualisations. I conducted a time series analysis on normalised crime count (averaged on LSOA populations) and it’s monthly fluctuations between the 4 police forces. Using groupby() based off force and month, and the mean() of ‘crime\_count’ which is computed based off LSOA crime sums, These variables were visualised using all 4 forces using sns.lineplot().

A visualisation was created using a crime map by installing geopandas/mapclassify using pip, and using LSOA boundary shapefiles provided by: <https://geoportal.statistics.gov.uk/datasets/ons::lower-layer-super-output-areas-december-2021-boundaries-ew-bsc-v4-2/about>. I filtered my data for duplicates and made sure each row contained the LSOA code and sum of crime\_count, merged the shapefile with my crime data, and plotted a choropleth map using the mapclassify package.

1. **Predictive modelling**

To conclude the analysis phase, I conducted a machine learning model using ‘crime\_count’ and it’s monthly lagged counterpart ‘lagged\_crime\_count’. The process began by creating a training set between February 2023- February 2024, and a testing set between February 2024- February 2025. I then used the package sklearn.linear\_model and sklearn.metrics to conduct a regression and compute MAE coefficients for model evaluation. I used a for loop to create a linear model for each police force, fit the model on our training data and computed model summaries, followed by appending our linear model to our testing set. A visualisation was created using a faceted residual plot of the actual data against predicted values using sns.scatterplot() and facetgrid().

**Results and real estate applications:**

1. **Univariate analysis**

*Table 1: Descriptive stats on crime frequency:*

|  |  |  |
| --- | --- | --- |
| Mean | Standard deviation | Median |
| 1702 | 4161 | 561 |

The average crime count per record is approximately 1,702, but the median is much lower at 561, indicating a right-skewed distribution. This indicates that while most areas experience moderate crime levels, a small number of areas have very high crime rates that raise the overall average.

*Table 2: Mean according to police jurisdiction*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | West Midlands | West Yorkshire | Metropolitan | City of London |
| Mean | 962 | 1301 | 1967 | 10571 |

The **City of London** has the highest average, but this may be skewed due to:

1. A very small number of LSOAs (but high crime activity in a commercial hub)
2. Data aggregation issues (e.g., centralised reporting)
3. The Metropolitan Police (covering Greater London) also shows high values, which is expected for an urban force.

West Midlands and West Yorkshire show lower average crime counts, suggesting either less crime or more distributed patterns.

*Figure 1: Faceted plot of crime frequency distribution per police jurisdiction*

A graph of different numbers

AI-generated content may be incorrect.

**City of London**

* Much smaller LSOA count overall (note low y-axis).
* The concentration around 500–2,000 crimes suggests crime is localized but intense in some central zones.
* Reflects the compact geography and commercial density of the City.

**Metropolitan (Greater London)**

* Extremely high number of LSOAs with low-to-mid crime counts (especially <1,500).
* The long tail (up to 10,000) shows that some areas, like inner boroughs, see disproportionately high crime.
* This spread reflects the socio-economic and spatial diversity across London, from low-crime outer boroughs to high-crime urban cores.

**West Midlands**

* Majority of LSOAs fall within the 0-1,500 range, with a few extending beyond 2,000.
* Distribution shows a consistently moderate spread, with no sharp peaks.

**West Yorkshire**

* Similar pattern to West Midlands - tight clustering below 1,500, steep drop-off after that.
* This indicates that while crime is present across most areas, it's less concentrated in extreme outliers.

*A bar graph with numbers and a number of numbers

AI-generated content may be incorrect.Figure 2: Top 10 LSOA’s by total crime count*

The most crime frequented LSOA's are 6 neighbourhoods in Westminster. One postcode in the city of London area, Hillingdon, Birmingham, and Leeds. Alternately, 4 Neighbourhoods in Ashford have the least amount of crime, one in Wychavon, Wyre, Sefton, Brentwood, and Wrexham.

**Solution:**

High-crime areas: Recommend secure apartments, gated developments, or BTL flats with concierge/security.

Lower-crime areas: Market detached/semi-detached homes or family-friendly townhouses.

*Figure 3: Frequency of each crime category*

A graph with blue and white bars

AI-generated content may be incorrect.

The top 5 most frequent crimes are sexual/violent offences, anti-social behaviour, theft, vehicle crime, and shoplifting. Whereas the least frequent are- possession of weapons, bicycle theft, other, robbery, and drugs.

*A group of colorful bars

AI-generated content may be incorrect.Figure 4: Crime categories and their frequency, per police force*

**Violent crime (e.g. West Midlands, West Yorkshire)**

* Potential safety concerns, especially for families.
* **Solution:** Focus on selling to investors, developers, or rental landlords seeking value in undervalued zones.

**Property crime (e.g. City of London, Metropolitan)**

* Busy commercial/residential mix, high footfall.
* **Solution:** Sell to young professionals, executives, or students looking for centrality and convenience.

**Anti-social behaviour (e.g. Metropolitan)**

* Possible community disruption or youth disengagement.
* **Solution:** Consider buyers interested in regeneration zones, council partnerships, or medium-term capital gain.
* Anti-social-heavy zones: Prioritize new builds or refurb projects aimed at elevating community feel.

1. **Bivariate analysis:**

*Table 3: correlation matrix between crime deprivation decile, service access, homelessness rate, and income/ education/ environment/ deprivation*

|  |  |
| --- | --- |
| Variable | Correlation to Crime Deprivation Decile |
| Income Deprivation Decile | 0.48 |
| Education Deprivation Decile | 0.38 |
| Environment Deprivation Decile | 0.23 |
| Homelessness Rate | -0.04 |
| Service Access | -0.23 |

Correlation matrix summary

* Income Decile (0.48): Lower-income (more income deprived) areas show higher crime index.
* Education Decile (0.38): Worser educated areas show higher crime, typically due to high youth disengagement/ weak social infrastructure or capital.
* Homelessness (-0.04): Practically no relationship.
* Service access (-0.23): The worser the service access, the lower crime deprivation. This is an unexpected finding. One reason for this could be an issue of under reporting. With the lack of services it's harder to reach out for help/ report crime. Alternately, the 'rural effect' means that service deserts have stronger local ties and thus less crime.
* Environment Decile (0.24): The more polluted an area, with less recreation and green space, faces elevated crime also.

**Solution:**

* Position low-income area properties to social landlords, developers, and patient investors with a regeneration mindset- help mitigate crime in the long run.
* Recommend buyer due diligence on neighbourhood-level crime alongside school ratings when selling to families.
* Highlight nearby parks, tree cover, or planned greening as selling points, especially in flats or dense housing areas.
* Avoid making assumptions or using homelessness rates as a risk signal in buyer guidance.

Subsequently, this report visualises the top 2 strongest correlators of crime.

*Figure 5: Correlation matrix line plot between Crime deprivation and Income deprivation*

A graph with blue lines and numbers

AI-generated content may be incorrect.

*Figure 6: Correlation matrix line plot between Crime index and Education deprivation*

A graph with blue lines and numbers

AI-generated content may be incorrect.

1. **Grouped analysis**

*Table 4: Resolution rate per police jurisdiction*

|  |  |
| --- | --- |
| Force | Resolution rate |
| City of London | 3.81% |
| Metropolitan | 1.86% |
| West Midlands | 2.46% |
| West Yorkshire | 2.41% |

*Figure 7: Bar chart signifying resolution rate by police force*

A graph showing a number of different colored rectangular objects

AI-generated content may be incorrect.

City of London- 3.81%- Highest - though still low, suggests more closure per crime.

Metropolitan- 1.86%- Very low - may affect buyer perception of safety/trust.

West Midlands- 2.46%- Low but consistent with national trends.

West Yorkshire- 2.41%- Similar story - low resolution, high volume areas.

**Solutions:**

* Flag resolution rates in your area summary, especially for risk-averse or long-term buyers
* A high-crime, high-resolution area may be more stable than one with unresolved volume
* Metropolitan has high crime and low resolution- if leasing/selling here, make sure: prioritise buildings with more security, buyers with higher risk tolerance such as investors as opposed to vulnerable populations like families/ single person households.
* City of London- highest resolution indicates strong policing, surveillance infrastructure etc. prioritise buyers who are more at risk of crime e.g., elderly, families with children etc.

*Figure 8: Top crime categories by income deprivation tiers*

A graph of different colored bars

AI-generated content may be incorrect.

Violent crime is concentrated in poorer/ more deprived areas, with nearly 5.4x more crime in higher deprived areas than low. Property crime is prevalent across all tiers, but is highest in higher income/ less deprived areas because of perceived luxury/ valuable goods in these areas. Anti-social behaviour also follows deprivation gradient.

**Solution:** For families, emphasise gated communities and avoid high income deprived areas. Investors who consider medium/low deprived areas should prioritise security to protect their assets due to the prevalence of property crime.

1. **Multivariate analysis**

*Figure 9: Correlation between education deprivation, income deprivation, and crime deprivation*

A graph of a crime

AI-generated content may be incorrect.

This plot confirms that areas facing multiple disadvantages, low income and poor educational attainment, are far more likely to experience elevated levels of crime. It visually reinforces the cumulative impact of deprivation across domains.

**Solution:** Nadine should analyse multiple indices for real estate decision, by factoring them in together she can make informed decisions on avoiding areas with high crime. If you’re selling from a low income area, by informing sales based on better education services could help lessen the risk of crime and vice versa.

1. **Domain specific visuals**

*Figure 10: Time series analysis grouped by police forces (and their respective LSOA monthly crime frequency)*

A graph of a number of people

AI-generated content may be incorrect.

City of London has extremely high crime per LSOA, with average monthly counts around 10,000+ crimes per LSOA, it's an extreme outlier. This is due to the very small number of LSOAs in the City of London, each covering a dense commercial hub with disproportionately high footfall and crime. While total crime volume is low, per-LSOA intensity is exceptionally high, a classic example of data distortion from small area counts.

Metropolitan Police Has Significantly Higher LSOA-Level Crime Than Others, averaging around 1,800- 2,400 crimes per LSOA per month, with minor seasonal variation. This reflects not just population density, but urban complexity, socioeconomic diversity, and transport-driven footfall in Greater London.

West Midlands and West Yorkshire Remain Close in Crime Density. Both fluctuate in the 1,000-1,500 range, with West Yorkshire slightly higher overall. These forces show more stability and less pronounced peaks, indicating a more even spread of crime across LSOAs.

**Solution:** This can guide buyers concerned with local experience or the density of incidents as opposed to force-wide trends. For instance, a family looking to move to London may be mistaken by their low crime count, unaware of the crime density instead. Per LSOA/area

*Figure 11: Crime map*

A map of england with red spots

AI-generated content may be incorrect.

Crime is Highly Concentrated in Urban Areas. These dark red clusters (highest crime counts) are clearly visible around London, Manchester, Birmingham, Leeds or Bradford. These areas have both high population density and intense urban activity, which typically correlate with higher crime volumes.

Moderate Crime Spread in Secondary Urban Zones. Orange-tinted regions represent mid-to-high quantile crime counts, Parts of Bristol, Nottingham, Sheffield, Liverpool, Newcastle, and Portsmouth/Southampton.

Low Crime Dominates Rural and Suburban Areas, these regions correspond to rural areas, small towns, and affluent suburbs, where lower crime rates are typical.

**Solution:** this spatial crime distribution suggests a tiered investment strategy. In high-crime urban areas like London, Manchester, and Birmingham, prioritising build-to-rent developments with integrated security and targeting investors or tenants with higher risk tolerance or short-term leases (e.g., young professionals, corporates) may be more viable. Meanwhile, low-crime rural and suburban zones are ideal for family housing, retirement communities, or long-term ownership sales, appealing to risk-averse buyers seeking safety and stability.

1. **Predictive modelling**

*Table 6: Model coefficients for predicted increases in crime, according to police force*

|  |  |  |  |
| --- | --- | --- | --- |
| Force | Intercept | Slope | MAE |
| West Midlands | 172 | 0.58 | 170 |
| West Yorkshire | 167 | 0.63 | 182 |
| Metropolitan | 166 | 0.63 | 205 |
| City of London | 1757 | 0.39 | 2341 |

To assess the predictability of crime trends across different regions, linear regression models were developed for each police force using lagged\_crime\_count as the independent variable and crime\_count as the dependent variable. The slope, intercept, and Mean Absolute Error (MAE) provide insight into the crime dynamics and model fit per region.

**West Midlands**

* **Slope**: 0.59
* **Intercept**: 172
* **MAE**: 171

The model indicates modest but consistent crime growth. With a relatively low MAE and strong linear fit, crime trends in the West Midlands appear predictable and stable. This makes the region suitable for confident real estate planning, particularly for long-term residential or mixed-use developments where risk mitigation is a priority.

**West Yorkshire**

* **Slope**: 0.64
* **Intercept**: 167
* **MAE**: 182

West Yorkshire exhibits slightly faster crime growth than the West Midlands, but still follows a clear lagged pattern. The model performs well, suggesting crime changes are responsive to recent history, offering reliable signals for value-add investments and regeneration planning.

**Metropolitan Police (Greater London)**

* **Slope**: 0.63
* **Intercept**: 166
* **MAE**: 205

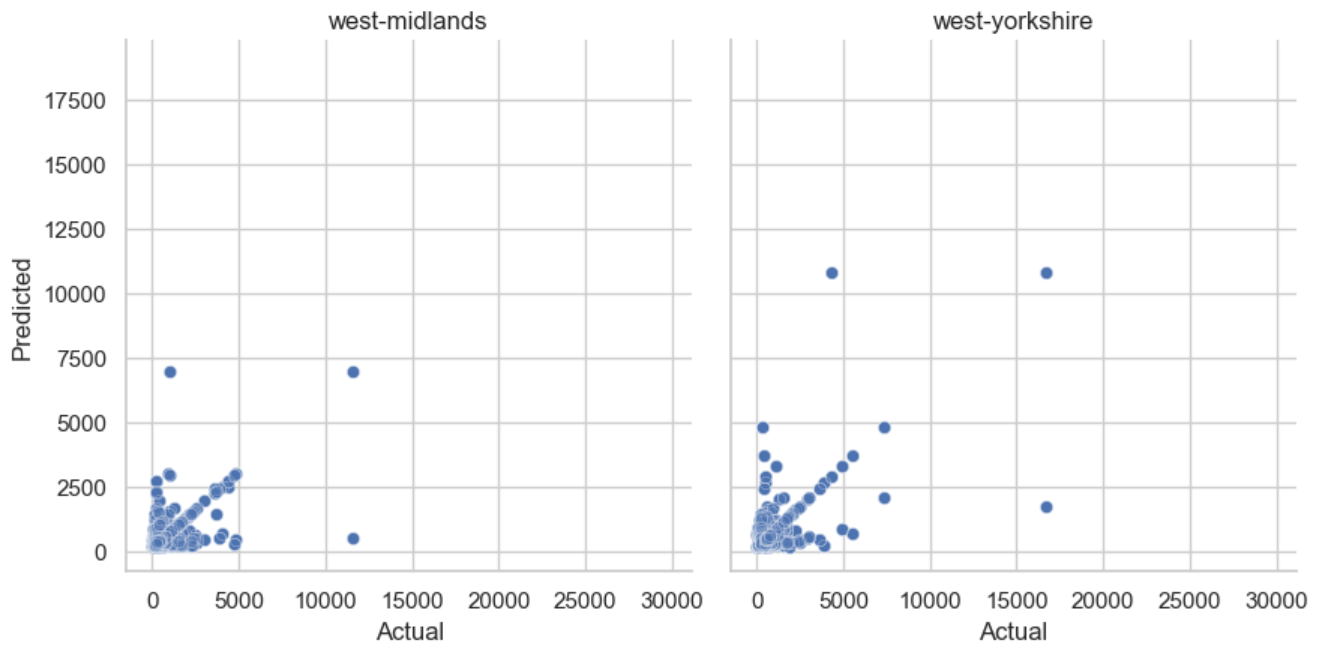
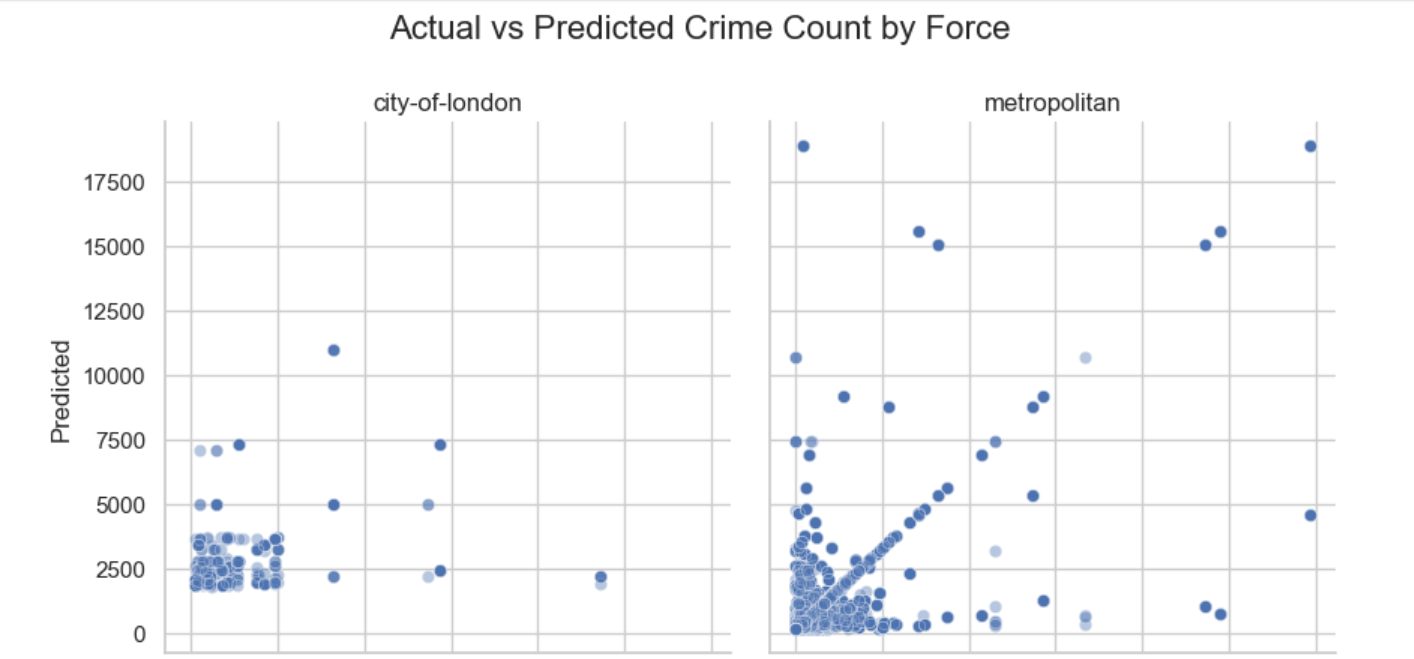
Greater London demonstrates high and rising crime volumes, yet the model achieves a reasonable fit. The positive slope indicates ongoing upward pressure, making this area suitable for build-to-rent developments, smart surveillance infrastructure, or high-security housing. Predictability enables planners to anticipate short-term fluctuations, though long-term forecasting remains complex due to the city's heterogeneity.

**City of London**

* **Slope**: 0.39
* **Intercept**: 1,757
* **MAE**: 2,342

The model fit is weakest in the City of London. With an unusually high intercept and large residual error, crime patterns here appear unpredictable and highly concentrated, likely driven by commercial activity, nightlife, and event-based footfall. Real estate decisions in this area should be approached with caution; investors should prioritize adaptive, high-yield commercial or short-term rental strategies over long-term residential placements.

Figure 12: F*aceted residual plot by police force*



City of London- Model is overestimating crime counts for many cases. Suggests poor fit for this force - possibly due to small sample size or outlier sensitivity.

Metropolitan- Dense cluster around the origin, but a significant number of predictions are consistently lower or higher than actuals (i.e., some underestimation and overestimation).

West Midlands- Model is less reliable at high crime levels, possibly underfitting.

West Yorkshire- Appears to slightly underpredict high actual counts (points below the y=x line). Nonetheless, performs better than City of London and West Midlands in terms of spread.

**Solution:**

All forces exhibit positive crime growth, indicating that higher crime in one period tends to persist into the next.

West Midlands and West Yorkshire show modest and predictable crime increases with strong model fit, ideal for conservative or mid-risk investment strategies.

Metropolitan (Greater London) presents high crime levels but reasonable forecast accuracy, making it suitable for security-oriented or institutional investment.

City of London stands out as volatile and structurally unique, with poor model performance, warranting specialised strategies tailored to short-term or commercial leasing markets.

**Discussion**

**Summary of Key Findings and Strategic Recommendations**

This analysis identified clear spatial and socioeconomic patterns in crime distribution across England. High crime volumes were concentrated in dense urban areas such as Greater London, Birmingham, and Manchester, while lower crime frequencies were found in rural and suburban regions. Socioeconomic deprivation, particularly income and education, showed the strongest positive correlations with crime rates, whereas service access displayed an inconsistent relationship, potentially influenced by underreporting in remote areas.

Violent crime were disproportionately present in areas of high-income deprivation, while property crime and anti-social behaviour were more evenly distributed, particularly in commercial and high-footfall zones. Predictive modelling confirmed positive crime growth trends across all police forces, with particularly stable and forecastable increases in the West Midlands and West Yorkshire. The City of London, despite its limited geographic scope, exhibited the highest crime intensity per LSOA and the highest resolution rate, indicating a uniquely concentrated and well-policed environment.

The spatial and socioeconomic patterns uncovered in this analysis present clear implications for sales strategy and investor positioning:

1. Investment by Crime Intensity

* High-crime urban areas (e.g. London, Manchester, Birmingham): Prioritise secure developments such as gated communities, build-to-rent flats, and properties with concierge or surveillance features. Target institutional investors, young professionals, or tenants with higher risk tolerance.
* Mid-crime secondary cities (e.g. Nottingham, Sheffield): Focus on regeneration-led investments, refurbishments, and value-add properties in transitioning neighbourhoods.
* Low-crime rural and suburban zones: Position detached or semi-detached homes for families, retirees, and long-term buyers seeking stability and safety.

2. Deprivation-Specific Strategies

* In highly income-deprived areas, prioritise buyers with a regeneration focus (e.g. housing associations, social landlords) and long-term investors. Emphasise community uplift potential while mitigating risk with added security.
* Avoid marketing to risk-averse buyers (e.g. families) in zones with extreme income/ education deprivation and violent crime prevalence.
* In medium- and low-deprivation areas, emphasise the need for secure storage and home protection due to elevated rates of property crime.

3. Crime Type Differentiation

* Violent crime zones: Market primarily to investors or developers, not families, given safety concerns.
* Property crime hotspots: Emphasise centrality and access for target groups like students and professionals but advise on added security.

4. Use of Crime Resolution Rates

* Include local crime resolution rates in sales material. A high-crime area with a high-resolution rate (e.g. City of London) may still represent a secure, well-policed investment.
* In areas with low resolution (e.g. Metropolitan), guide risk-averse buyers toward more secure buildings and focus listings toward investors or short-term tenants.

5. Neighbourhood-Level Contextualisation

* Use a multi-factor approach- education, income, service access, and environment, to assess risk and guide pricing.
* Emphasise proximity to parks, greenery, and well-rated schools when targeting family buyers.

**Limitations:**

* The underlying deprivation data is from 2019 and may not reflect post-pandemic socioeconomic changes.
* Missing values in deprivation metrics were imputed using force-level means, potentially obscuring finer LSOA-level differences and weakening variable precision.
* The predictive models were trained on only one year of data (Feb 2023–Feb 2024), which may not capture seasonal cycles, long-term effects, or structural shifts.
* Additionally, not all crime types or outcome fields were analysed due to dataset scope constraints.

These limitations suggest that while the findings are meaningful, caution is advised in operationalising them at a micro level. Future iterations should incorporate updated deprivation indices, seasonal adjustments, and extended historical data to strengthen predictive power and local relevance.

**Proposed Areas for Further Analysis**

In alignment with stakeholder objectives, further analysis is proposed for two contrasting police jurisdictions:

1. **City of London Police**: Despite a small number of LSOAs, this area reports extremely high crime counts per unit and the highest crime resolution rate (3.81%). These characteristics make it a valuable case for understanding the dynamics of concentrated commercial crime and the impact of policing infrastructure on real estate risk.
2. **West Midlands Police**: As a large, socioeconomically diverse region with forecastable crime increases and strong machine learning performence/ crime predictivity (low MAE), the West Midlands presents a strategic opportunity to explore data-driven planning for regeneration, investment, and risk-adjusted pricing strategies.

Focusing on these two jurisdictions will provide deeper insight into both concentrated and distributed crime environments, supporting targeted real estate planning and investment guidance.

References:

Cheshire, P., & Sheppard, S. (2004). *Capitalising the value of free schools: The impact of supply constraints on house prices*. Economic Journal, 114(499), F397–F424.

Deming, D. (2011). *Better schools, less crime?* Quarterly Journal of Economics, 126(4), 2063–2115.

Gibbons, S. (2004). *The costs of urban property crime*. Economic Journal, 114(499), F441–F463.

Linden, L., & Rockoff, J. E. (2008). *Estimates of the impact of crime risk on property values from Megan’s laws*. American Economic Review, 98(3), 1103–1127.

Lochner, L., & Moretti, E. (2004). *The effect of education on crime: Evidence from prison inmates, arrests, and self-reports*. American Economic Review, 94(1), 155–189.