# **Statement of Work**

**Project Title:** Exploratory Analysis of UK Crime Data for Real Estate Decision-Making  
**Stakeholder:** Nadine Green, Head of Sales, Real Estate Company  
**Prepared by:** EO   
**Date:** 10/05/2025

## **1. Project Objectives**

* Identify the most and least desirable UK regions based on crime trends (over two years)
* Provide actionable insights to guide real estate sales strategies.
* Recommend two police forces for deeper future analysis.

**Key Questions Answered:**

* Which regions have the lowest violent crime rates?
* Are there seasonal spikes in property crimes?
* Where are the safest LSOAs for development?

## **2. Scope of Work**

### **Included**

* **Data Sources:** [Street-level crime data](https://data.police.uk/data/) (Mar 2022–Mar 2024) from:
  + Hertfordshire Constabulary
  + Kent Police
  + Lincolnshire Police
  + Northamptonshire Police
* **Analysis Focus:**
  + Crime frequency and type distribution.
  + Temporal trends (monthly/seasonal).
  + Crime hotspots
  + Property crime focus (burglary, vehicle crime, arson).
* **Deliverables:**
  + Jupyter Notebook with preprocessing and analysis code.
  + EDA report with visualisations and executive summary.
  + [Trello board](https://trello.com/invite/b/6819cb865c15b50ae9a7595a/ATTIb21fcd8e7b451d537f5b37681c9428e25441E457/crime-project)

### **Excluded**

* Stop-and-search or outcome data
* Correlation with non-crime factors (e.g., property prices, demographics).

## **3. Methodology**

### **Data Collection**

* Download data from [data.police.uk](https://data.police.uk/data/) for:
  + **Forces:** Hertfordshire Constabulary, Kent Police, Lincolnshire Police, and Northamptonshire Police.
  + **Timeframe:** March 2022 – March 2024.

### **Analysis Approach**

1. Data Cleaning: Handle missing values, standardise formats.
2. Descriptive Statistics: Crime totals, top crime types per force.
3. Time-Series Analysis: Monthly trends and seasonality.
4. Geospatial Visualisation: Crime density maps (using folium).

### **Tools**

* Pandas (data wrangling).
* Seaborn/Matplotlib (statistical visualisations).
* Folium (interactive hotspot maps).
* NumPy (numerical computations, array operations)
* Jupyter Notebook for reproducibility.

## **4. Deliverables**

* Jupyter Notebook: Code for data extraction, cleaning, and analysis.
* EDA Report: PDF or Markdown file with: Executive Summary, Key Findings, Visualisations.
* Crime Maps: Maps highlighting hotspots.
* Completed [Trello Board](https://trello.com/invite/b/6819cb865c15b50ae9a7595a/ATTIb21fcd8e7b451d537f5b37681c9428e25441E457/crime-project)

## **5. Timeline**

* Day 1: Data collection, pipeline setup, Trello board.
* Day 2: Data cleaning, basic analysis (totals, crime types).
* Day 3: Data cleaning, basic analysis
* Day 4: Advanced analysis (temporal trends, geospatial maps).
* Day 5: Report writing, final visualisations, and quality checks.

## **6. Success Criteria**

* Stakeholder can identify top 2 forces for further study.
* Visualisations clearly highlight safest/riskiest regions.
* Report is actionable.

## **7. Assumptions & Risks**

* **Assumptions:**
  + Data is complete and accurate for the selected timeframe.
* **Risks/Mitigations:**
  + **Risk:** Large dataset size slows processing.  
    **Mitigation:** Test on a subset first, then scale.