

1 Introduction & Aim

Crime in cities does not spread evenly. It concentrates in a small number of places known as hotspots. UK public crime tools such as Police.uk show where crime has happened, but they do not identify hotspots or forecast future crime. This project addresses that gap for Swansea.

AIM

To analyse and forecast short term crime patterns in Swansea at LSOA level by combining crime, weather and deprivation data, presented through an interactive Power BI dashboard.

OBJECTIVES

- Build a SQL pipeline integrating four open data sources
- Identify crime zones using **K Means clustering**
- Forecast next month crime using three models
- Present results in an interactive dashboard

WHY THIS MATTERS

Gives Swansea Police and local planners a clearer view of where crime concentrates and where to focus resources.

2 Methodology

A quantitative case study of **24,921 crime records** across **150 Swansea LSOAs** from January 2024 to January 2025.

DATA SOURCES

Police.uk ONS WIMD 2025 Met Office MIDAS

PIPELINE



EVALUATION

K Means clusters validated with the Elbow Method and Silhouette Score. Forecast models tested using a split based on time and measured with **MAE** and **RMSE**.

3 Hotspot Detection

K Means clustering grouped Swansea's 150 LSOAs into **four distinct crime zones**. The Elbow Method and Silhouette Score both confirmed K equals 4 as optimal, with crime concentrated around the city centre.

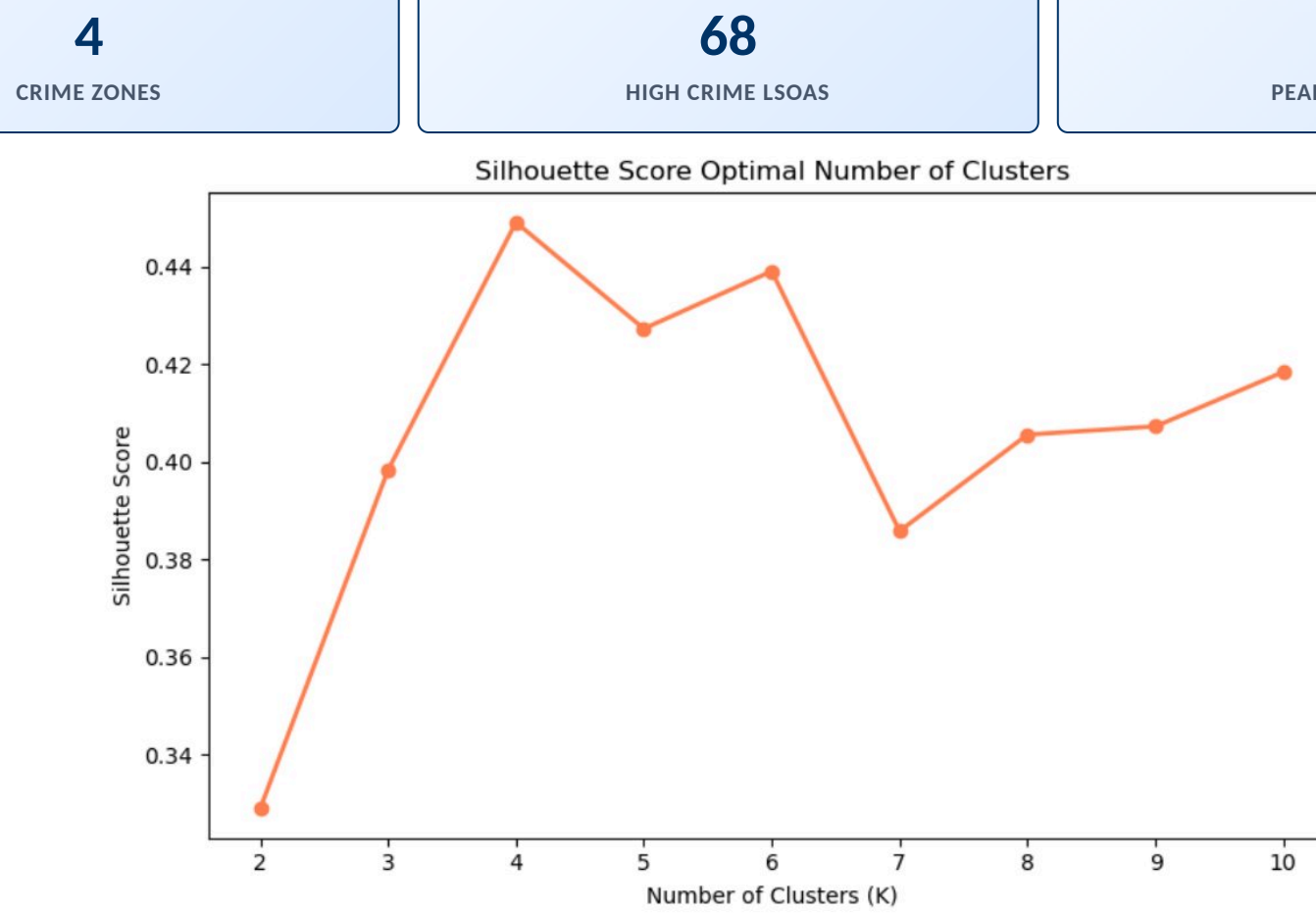


Figure 1. Silhouette Score peaks at K equals 4, confirming the choice of four clusters.

4 Forecasting Results

Three models compared on a time based split. The simple baseline performed surprisingly well.

Model	MAE	RMSE
Baseline (Rolling Avg)	3.411	4.875
Random Forest	3.439	4.789
XGBoost	3.774	5.821

Baseline achieved the lowest MAE. Random Forest achieved the lowest RMSE.

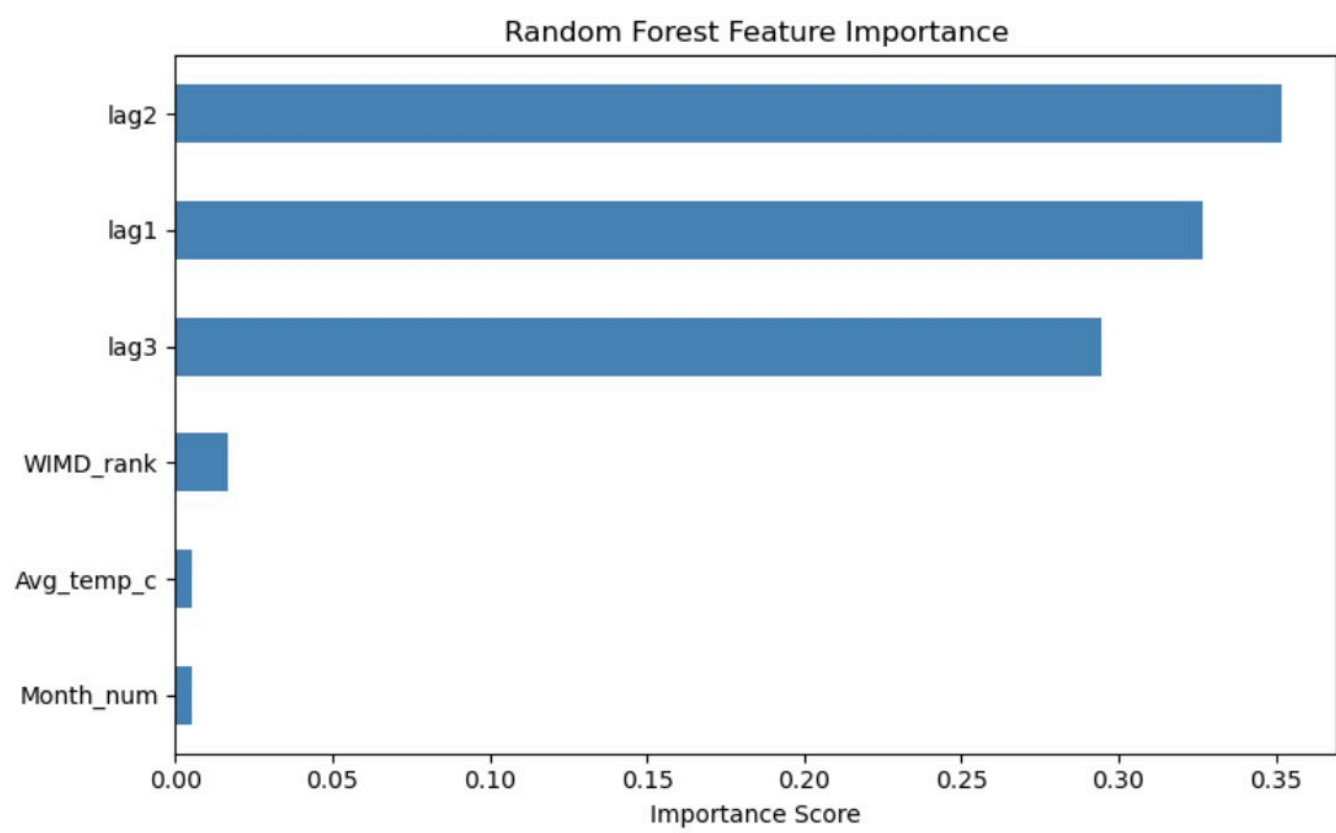


Figure 2. Recent past crime (lag1 to lag3) dominates feature importance.

5 Interactive Power BI Dashboard

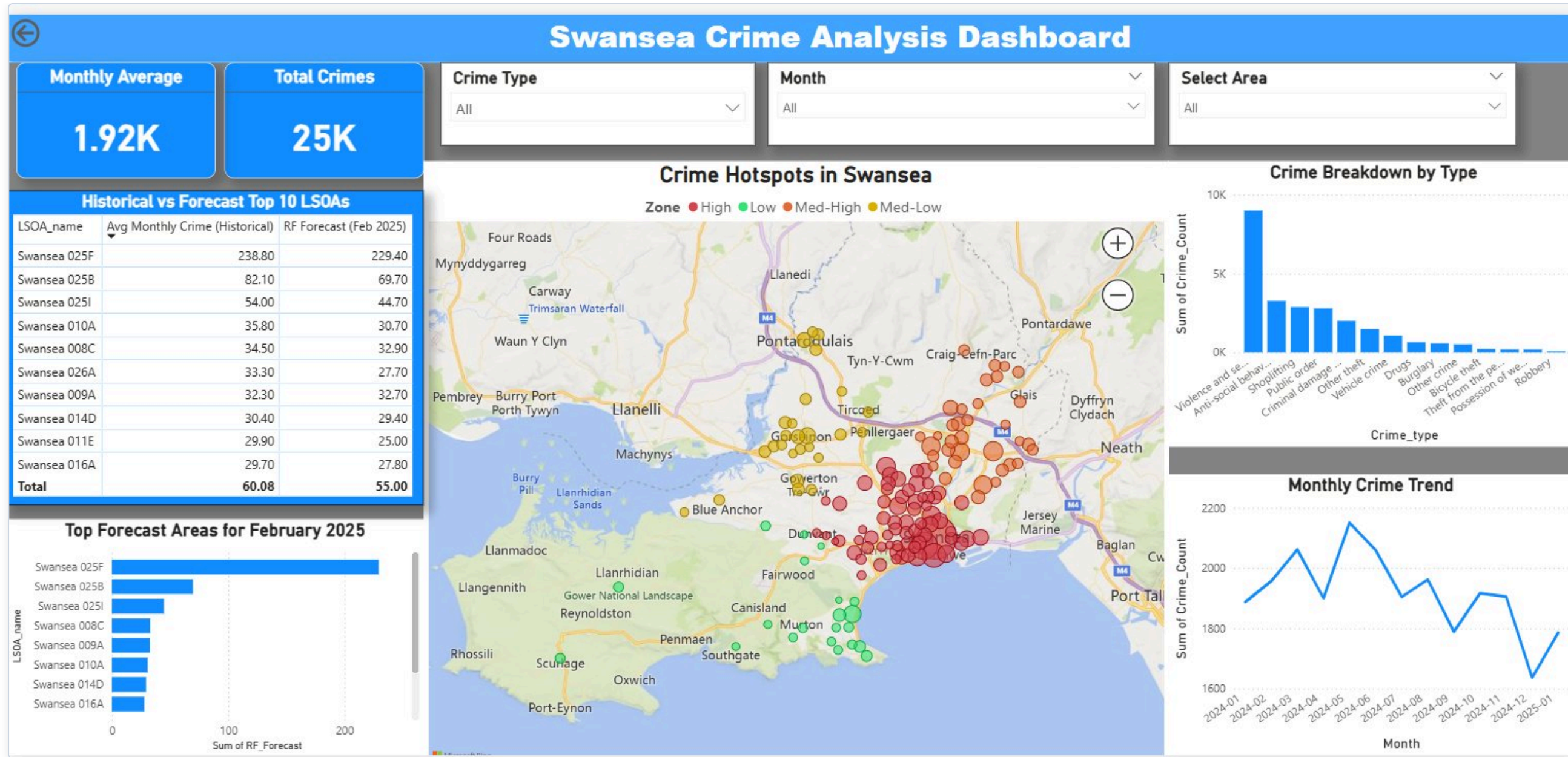


Figure 3. Interactive Power BI dashboard for the Swansea crime analysis system.

WHAT THIS SHOWS & WHY IT MATTERS

The dashboard brings hotspots, monthly trends, crime breakdown and the February 2025 forecast into one interactive view. Filters by area, month and crime type let analysts and planners quickly identify priority zones and emerging risks.

6 February 2025 Forecast

Crime risk remained concentrated in city centre LSOAs. Top predicted hotspots from Random Forest:

Swansea 025F	229.4
Swansea 025B	69.7
Swansea 025I	44.7
Swansea 008C	32.9
Swansea 009A	32.7
Swansea 010A	30.7

TOP RISK AREA
Swansea 025F is the highest risk LSOA, far exceeding all others in the forecast.

7 Limitations & Future Work

LIMITATIONS

- Only 13 months of data, limiting seasonal patterns
- Forecasts cover total crime, not by individual type
- Weather data is a 2023 seasonal proxy
- LSOA level only due to Police.uk anonymisation

FUTURE WORK

- Extend the dataset to three or more years
- Build crime type specific forecasting models
- Apply hyperparameter tuning and add features
- Automate dashboard refresh with new data

8 Key Findings & Conclusion

- Crime in Swansea is highly concentrated in a small number of city centre LSOAs, supporting Weisburd's law of crime concentration.
- K Means identified 4 distinct crime zones at LSOA level (high, medium high, medium low, low).
- Baseline had the lowest MAE and Random Forest had the lowest RMSE. Added complexity did not clearly improve forecasts.
- Recent past crime levels were the strongest predictors, while temperature and deprivation had limited influence.
- Swansea 025F remained the highest risk LSOA in the February 2025 forecast.