# Theft of Motor Vehicles

Data Analytics Lab

February 2023

This project aims to use crime data held by WMP to predict hotspots of offending for Theft of Motor Vehicle (TOMV) offences.

# Table of Contents


1 Introduction ..... 3
2 Overview of Theft of Motor Vehicle Offences ..... 4
2.1 Summary of Data Used ..... 4
2.2 Escalation of offending ..... 5
2.3 Locations of offending ..... 6
2.4 Vehicle Data ..... 7
2.4.1 Most vulnerable vehicles in the region ..... 7
2.4.2 Age of vehicles stolen ..... 9
2.4.3 Difference in locations of most vulnerable vehicles ..... 9
2.5 Clustering and Hot Spot Analysis ..... 11
3 Methodology ..... 14
3.1 Spatio-Temporal Analysis ..... 14
3.2 GAM Workflow ..... 15
3.3 Model Validation and Selection ..... 15
3.4 Forecasting ..... 17
3.5 Model Results ..... 18
4 Project Output ..... 19
5 Appendix ..... 22
5.1 Other Methods Considered ..... 22
6 References ..... 23

# 1 Introduction

Theft of Motor Vehicles (TOMV) has become a 2022 force tactical priority due to the escalation of offending over the last three years. A vehicle crime task force has been created to focus efforts on this type of criminality.

The increased level of offending is largely being driven by the international supply chain shortage affecting vehicle and vehicle part production and availability within the United Kingdom (UK). A shortage of computer chips in car production, as well as other essential materials such as copper, aluminum and cobalt, has led to fewer new vehicles being produced and consumers waiting a long period of time for new vehicles. The current economic climate, logistical impact of Covid-19, Britain's exit from the European Union (EU) and the war in Ukraine are potentially contributing factors which have exacerbated supply issues. Consequently, the global demand for second hand vehicles and parts has increased significantly making vehicle theft a high reward and low risk crime for offenders, with just over 1% of recorded offences achieving a charge. Nationally in 2021, the most stolen vehicle make was Ford followed by Jaguar Land Rover (JLR). Ford is the most commonly owned vehicle in the UK; having easily interchangeable parts between models makes them an attractive target for their parts.

The current work from the Vehicle Crime Taskforce has revealed intelligence around a large number of chop shops across the force area. A problem profile has been created in respect of the vehicle types and patterns of offending, however, this is driven by reporting data alone.

This project uses crime data held by WMP in order to predict hotspot offending within the WMP region. This is in order to target our Neighbourhood Policing Unit (NPU) activity effectively together with appropriate crime prevention operations, thereby providing us with a more informed picture to tackle the TOMV issue. The analysis will focus on all vehicles, and highlight the most vulnerable vehicle types.

The intended output of this project will be a predictive model visualised in a Business Insight (Qlik) dashboard. This will be available to the vehicle crime task force and to intelligence analysts to assist with making recommendations about the deployment of resources to reduce vehicle theft. The information from the prediction will be used in conjunction with other information such as the locations of car dealerships and where more vulnerable vehicle types are registered according to the Police National Computer (PNC). In addition, local knowledge of offending patterns and intelligence about known vehicle offender's current activity will form part of the decision-making process.

The vehicle crime task force already utilises a range of operational and preventative tactics to reduce offending and a number of departments contribute to this work. Access to this predictive tool will not change the tactics used, but will ensure that resources are targeted effectively towards the most vulnerable locations and vehicle types.

The format of this report is as follows: section 2 gives an overview of the data used in the project, section 3 describes the methods used for creating the predictions and section 4 explains the project output.

# 2 Overview of Theft of Motor Vehicle Offences

## 2.1 Summary of Data Used

Data used in this project is based on crime investigations classified as 'Theft of Motor Vehicle' (TOMV), where there is a vehicle linked to the crime, and location information available. Only vehicles linked to an investigation where they are the 'STOLEN' vehicle are included in this dataset. Vehicles that were damaged, used in a crime or found at the crime are excluded from the dataset. The locations used in this project are the original crime locations, as eastings and northings where this was not available, locations were inferred from other address information available such as the postcode. Where this was not available either, records were excluded from the modelling as accurate location was a necessary part of the data.

All offences considered in this project are classified under the crime sub group of 'theft or unauthorised taking of a motor vehicle', this includes three distinct offence titles:

- Unauthorised taking of a motor vehicle (does not include 'driving or being carried knowing motor vehicles has been taken')
- Theft of a motor vehicle
- Attempted – Theft of a motor vehicle

This project does not include crimes classified as 'Theft from a motor vehicle', aggravated vehicle taking or residential burglaries where car keys were taken.

Vehicle make and model information was cleaned to ensure there was no duplicated values (e.g. FORD, FORD (UK) and FORD (EUROPE) were combined into one), and any misspellings were corrected where possible. Vehicles were categorized into 8 categories: Supermini, Van, Small Family Vehicle, Large Family Vehicle, SUV, Motorcycle, Executive/Sports and Other.

The premises type of incident location was not recorded well enough to be useful in this project as approximately 96% of values were 'UNKNOWN'.

When this dataset is aggregated to a weekly level, taking values from the last 6 months there was an average of 209 vehicles stolen per week, with a maximum of 261 and a minimum of 167 per week.

## 2.2 Escalation of offending

Since the beginning of 2021 there has been a continuous increase in the number of TOMV offences recorded in the Force area. Minimal monthly seasonal trends have been seen in recent years as the number of offences each month is higher than the previous month. Due to this, vehicle offences are now a force priority. Looking at the recorded data in the dataset for this project, there was a sudden level shift of the data in April 2021, most likely this was caused by the implementation of Connect. Prior to this there are two troughs in the data, most likely caused by the two national Covid-19 lockdowns we had in March 2020 and November 2020.

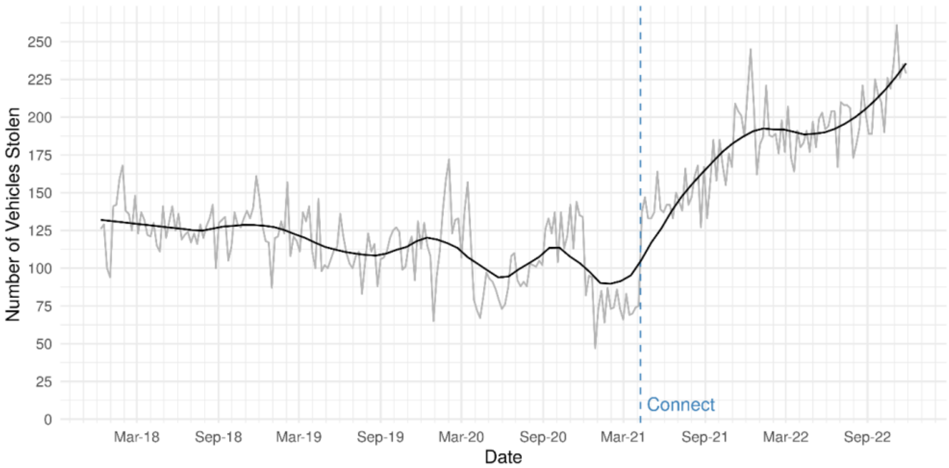



*Figure 1: Line chart showing smoothed line for weekly counts of number of vehicles stolen that are linked to investigations in the West Midlands Police Force area. Blue dashed line shows the date that Connect went live.*

Across the Neighbourhood Policing Units (NPUs) similar trends to the force total are seen, with increasing levels of TOMV offences over the last two years. The biggest increase has been seen in Birmingham West (BW), which is also the NPU that has the highest levels of this type of offending.

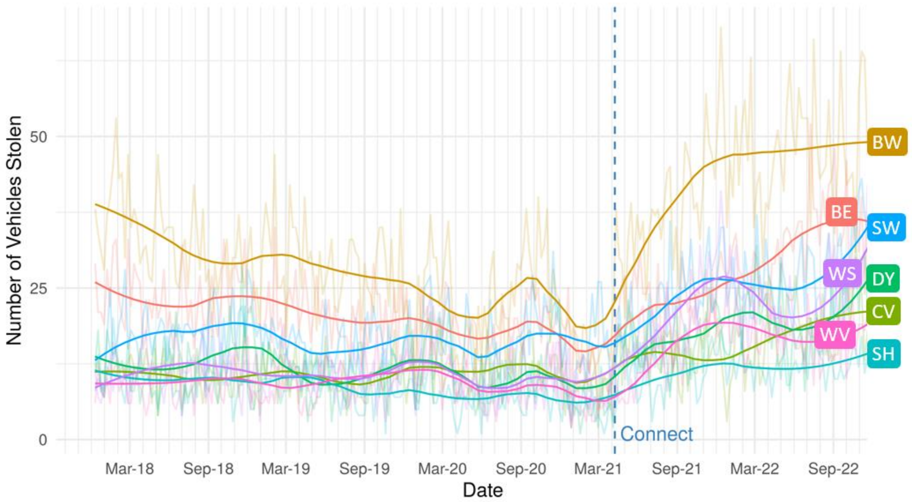



*Figure 2: Line chart showing smoothed lines for the weekly counts of number of vehicles stolen that are linked to investigations in WMP Force area, split up by NPU. Blue dashed line shows the date that Connect went live.*

## 2.3 Locations of offending

The aim of this project is to predict the locations of theft of motor vehicles, and not the locations where they are found, TOMV offences with vehicle locations linked where they have been found are excluded as we are only interested in the locations of the crime where a vehicle was stolen, not for example where a vehicle was used in a pursuit of another crime type. There is spatial clustering of offending over time in certain neighbourhoods and areas, with areas in Birmingham West showing higher levels of offending than other NPUs over the last 12 months when counts are totaled for each neighbourhood.

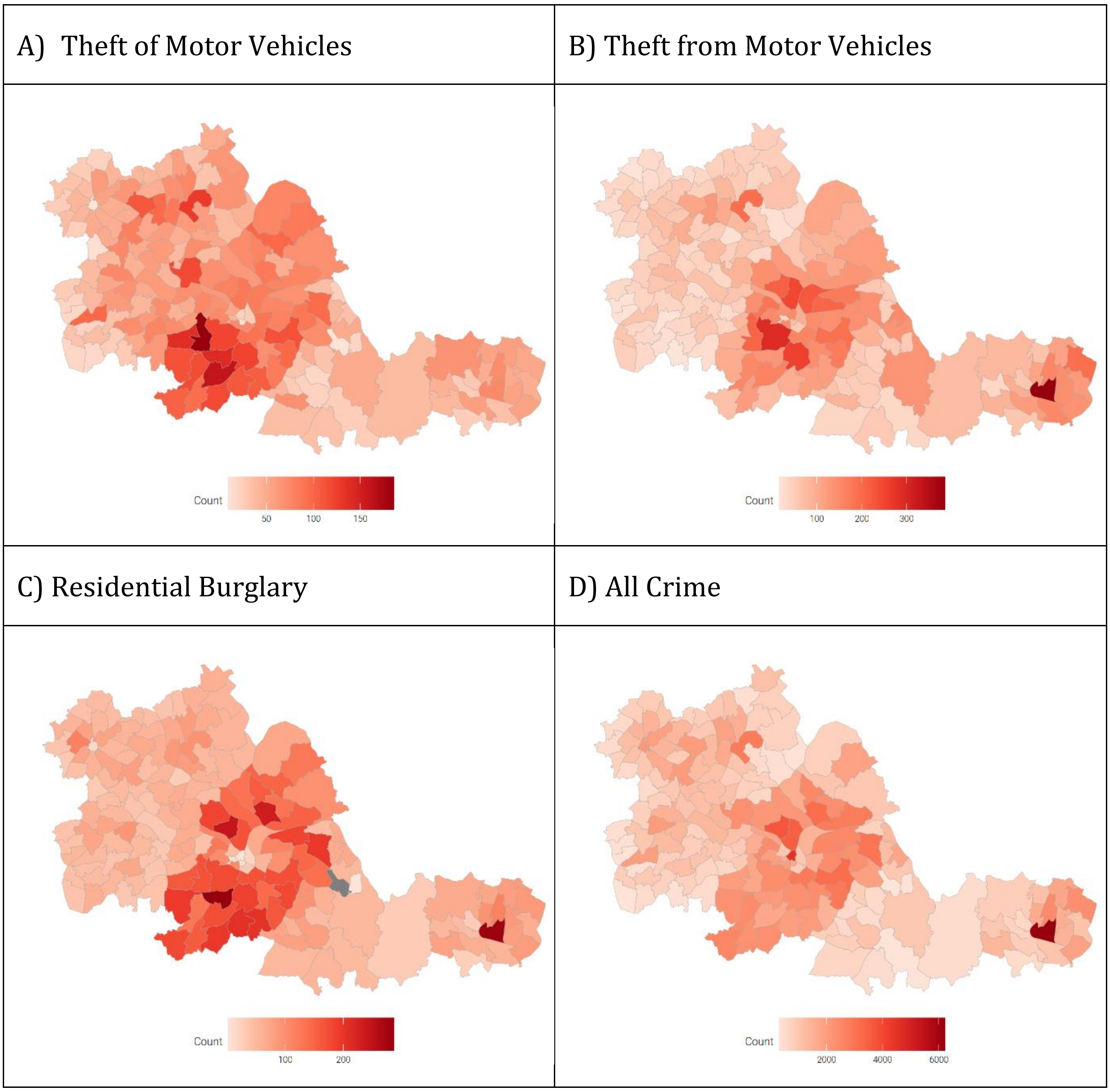



*Figure 3: Total count of a) number of cars stolen that are linked to a TOMV offence, b) count of theft from motor vehicle offences, c) count of residential burglaries and d) count of all crime, over the last 12 months per neighbourhood of WMP area. Darker colours show higher counts.*

## 2.4 Vehicle Data

### 2.4.1 Most vulnerable vehicles in the region

Within the WMP region, over the last 5 years, the manufacturer of vehicles that were most commonly stolen is Ford, with 31% of all stolen vehicles being this make. Mercedes and Land Rover are the next most taken. There are 110 unique vehicle manufacturers, but only 28 of these have had more than 100 cars stolen in this time period. The manufacturer of the vehicle is 'unknown', 'other' or blank in 4% of records. The most registered vehicle manufacturer in the UK for 2021 and 2022 to date is Ford, accounting for around 8% of vehicle and van registrations (SMMT, 2022).

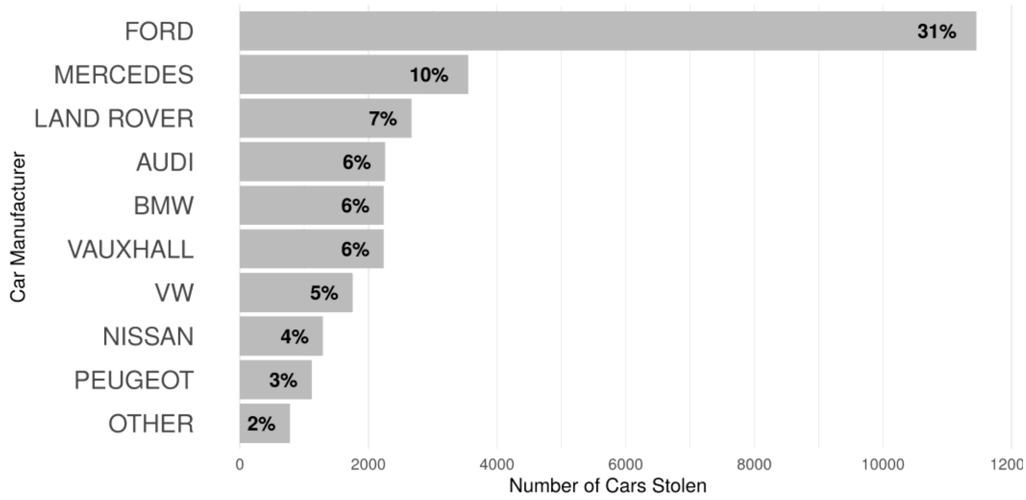



*Figure 4: Ranking of top 10 vehicle manufacturers based on total number of cars stolen over the last 5 years. Percentages displayed on the figure are the percentage out of all vehicles stolen, but only the top 10 manufacturers are shown here.*

When the data is broken down further into the vehicle model as well as manufacturer, three of the top 10 are Ford manufactured models. This is based on 5 years of data in total.

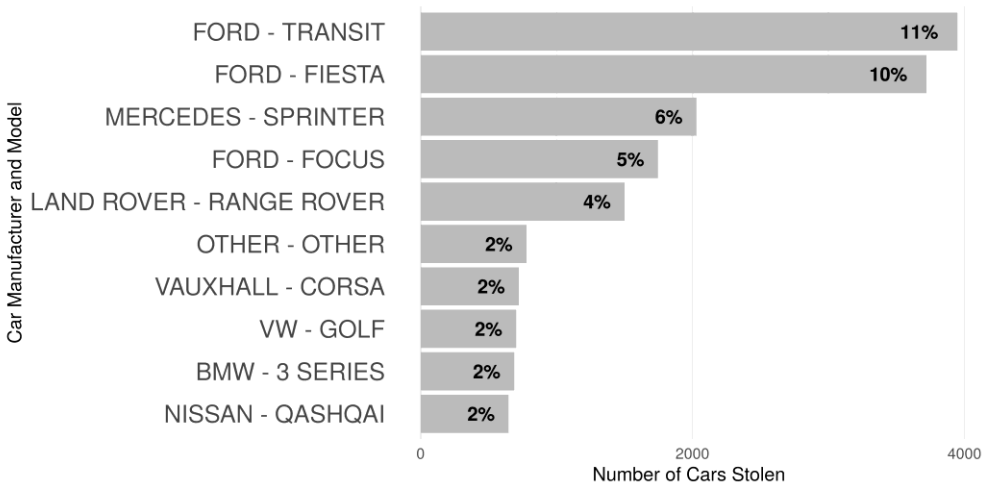



*Figure 5: Bar chart showing the ranking of the top 10 vehicle manufacturers and models that have been stolen in WMP region over the last 5 years. Percentages displayed on the figure are percentage of total cars stolen, but only the top 10 are shown here.*

Over the last 5 years, Ford Fiesta, Ford Transit Van and Mercedes Sprinter Van have always been in the top 10 list of makes and models stolen, with either Ford Fiesta or Ford Transit consistently in the top position. The Ford Fiesta was the best-selling car between 2009 and 2020 and so it is not surprising that it is also one of the most commonly stolen, as there are a larger number of them registered than other cars (Vehicle Licensing Statistics, 2021). This also matches with the new vehicle registration and used car sales data for the UK, with Ford Fiesta, Ford Transit and Ford Focus also featuring in the top 10 for 2021 and 2022 (SMMT, 2022).

Vehicles can also be categorized into larger categories, not based on the manufacturer but on the size and type of the vehicle, these follow the market segments. The category 'Supermini' is the most taken vehicle, this aligns with this segment being the largest in the used vehicle market, with Superminis accounting for one third of all bought vehicles (SMMT, 2022).

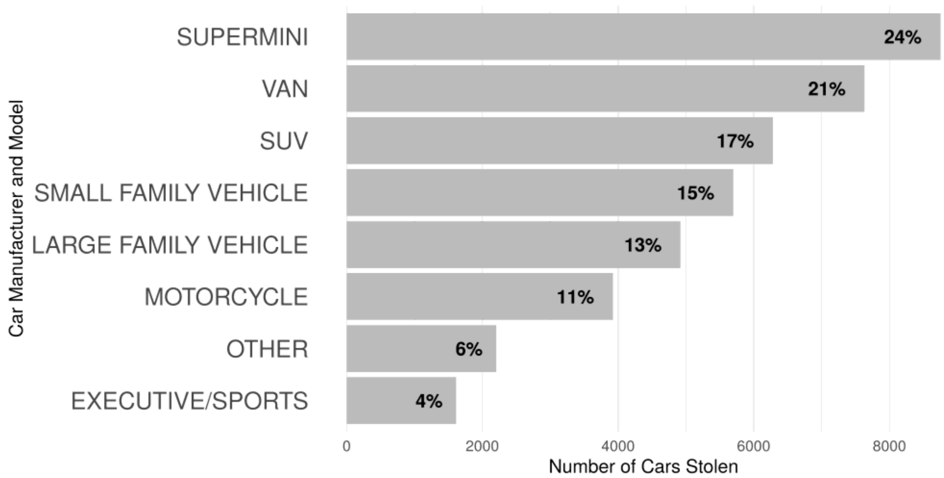



*Figure 6: Bar chart showing the breakdown of all vehicles into six categories and the proportion of vehicles that are stolen from each. Based on 5 years of data.*

### 2.4.2 Age of vehicles stolen

The age of vehicle data field is not mandatory and therefore the majority of vehicles do not have this information in the system. However, the age of the vehicle can be inferred from the registration plate, assuming that most vehicles will have their original registration plate on them. Where registration plates were 7 characters long (in the standard form XX00 XXX, where X=character and 0=number), the third and fourth characters were extracted as these give the approximate year of registration.

Aggregating all vehicles stolen over the last five years, there is a peak in the number of vehicles stolen that were registered between 2015 and 2017. If the data is split into yearly groups, this overall trend does not change, with these years being the most commonly stolen.

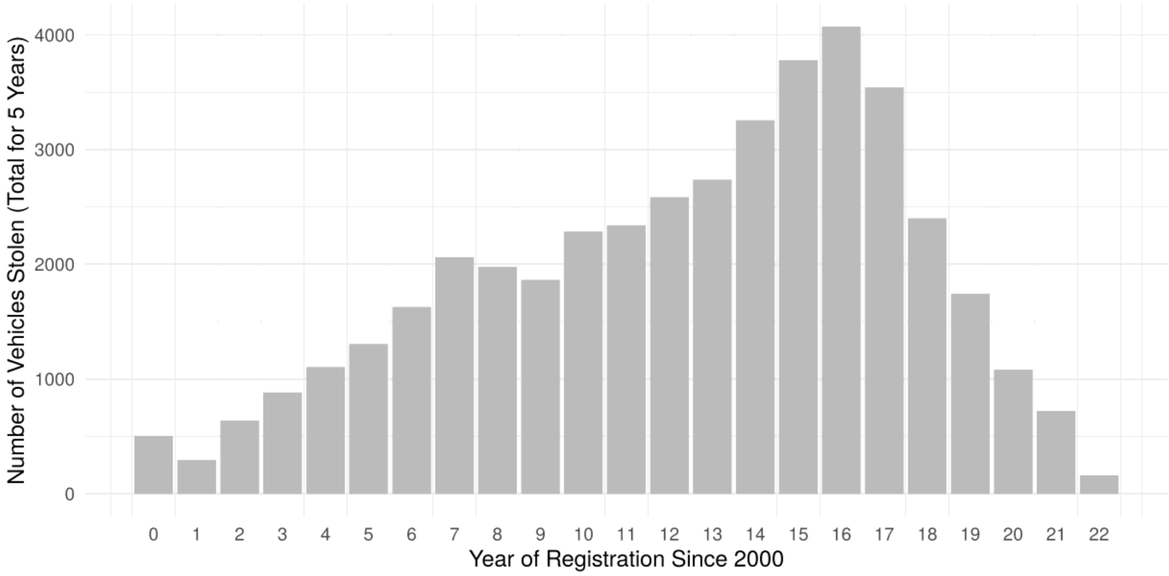


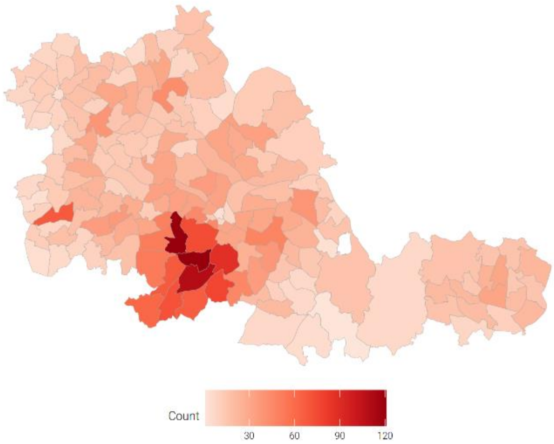

*Figure 7: Total number of vehicles over the last 5 years that were registered in each approximate year. Excludes registration plates that were not in the standard format or where the registration plate was unknown.*

As this data was aggregated and sparse, it was not useful or possible to include it in the spatiotemporal predictive model.

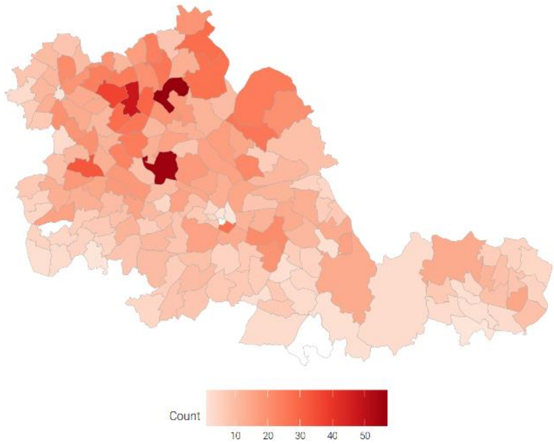
### 2.4.3 Difference in locations of most vulnerable vehicles

Potential differences in hot spot locations between different categories and types of vehicles are seen, looking at the last 12 months of data. Dark red areas show higher levels of offending, and white areas have lower levels. It is clear that depending on the vehicle type there are different parts of the region where more of those vehicles are stolen. The scales in the below plots are unique to each category, but as shown in Figure 6 the number of vehicles stolen in each category is different, with 'Superminis' making up the largest proportion of stolen vehicles.

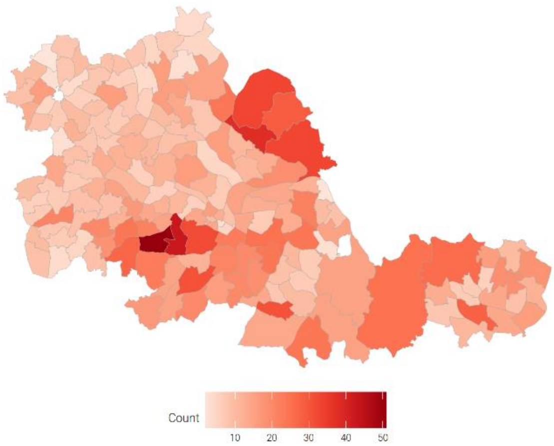
Supermini (24%)



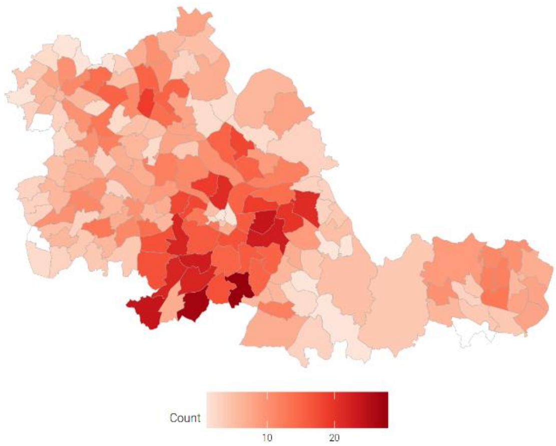

Van (21%)



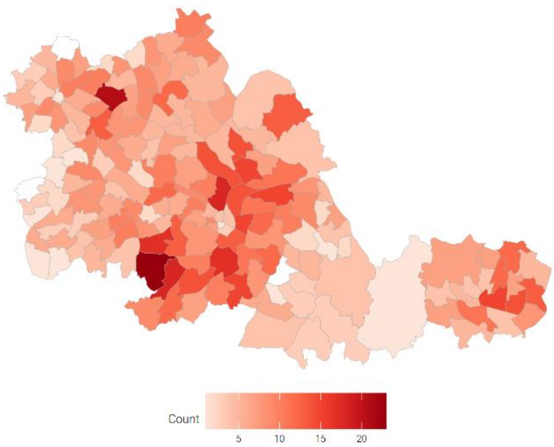

SUV (17%)



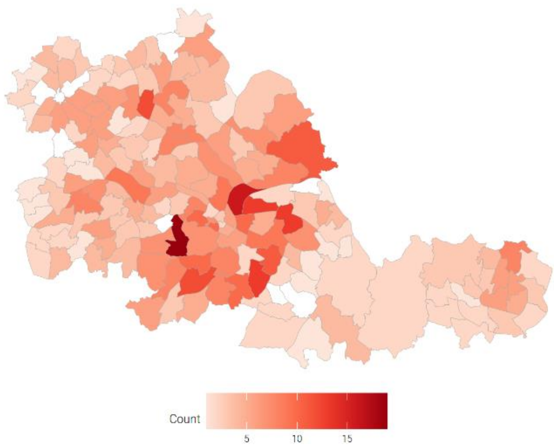

Small Family Vehicle (15%)




Other (including unknown, large luxury and executive, sports and MPV) (14%)




Large Family Vehicle (10%)



In order to maintain maximum accuracy, all data is used together for forecasting purposes, as well as being split into the six categories shown above. When breaking the data up further into makes and models, it becomes too sparse to use for forecasting purposes.

## 2.5 Clustering and Hot Spot Analysis

Using the WMP defined neighbourhoods as polygons, Moran's I analysis (Cressie et al., 2019) was carried out to assess the degree of spatial association between neighbourhoods. A coefficient of 0.49 was returned when using data from the last 6 months. This positive value equating to an upwards slope on a scatterplot (i.e. positive spatial autocorrelation) suggests that as the number of TOMV offences in a said neighbourhood polygon increases, so does the number of offences in neighbouring neighbourhood polygons. This coefficient suggests that the number of TOMV offences in neighbourhoods are clustered and not randomly distributed and dispersed. When using 12 or 60 months of data, the coefficient returned was 0.51 and 0.54, suggesting similar findings over different time periods.

Local Indicators of Spatial Association (LISA) is another method that can be used to assess the degree of spatial association, it allows for the decomposition of Moran's I into the contribution of each individual observation (Anselin et al., 1995). The LISA for each observation gives an indication of the extent of significant spatial clustering of similar values around that observation. Local spatial clusters (hot spots) may be identified as those locations for which the LISA is significant. Locations with significant positive local spatial autocorrelation are the core of a cluster, the actual cluster includes neighbours as well as the core. Figure 8 shows the output of a LISA, where red neighbourhoods are those with high values of TOMV offences, surrounded by other neighbourhoods with high values. Blue areas are those with low levels of TOMV offences. This output confirms that there is spatial clustering of observations in the WMP region.

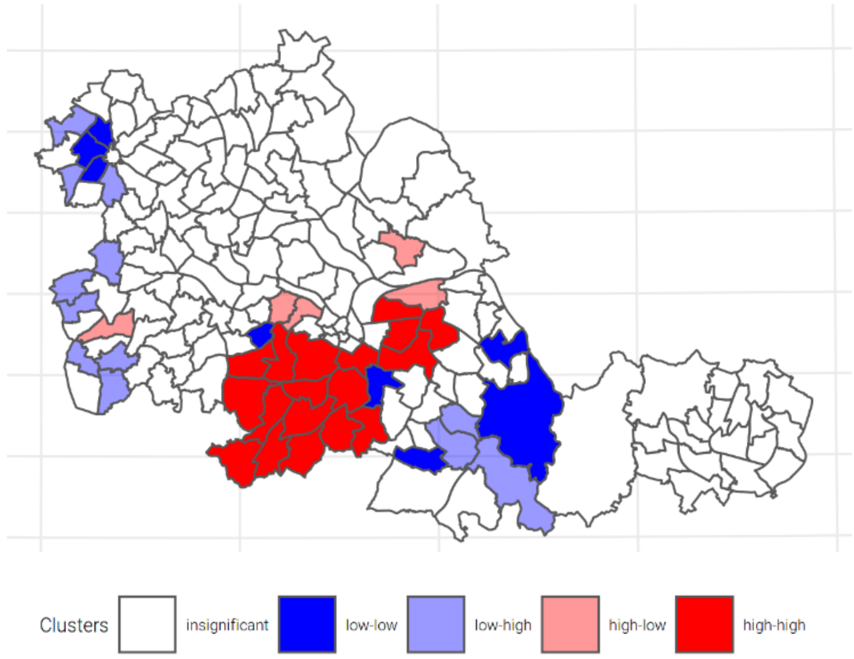



*Figure 8: Local Indicators of Spatial Association (LISA) plot for neighbourhood regions of WMP area.*

When data is aggregated into polygons, the modifiable areal unit problem may occur. This is when data outputs potentially vary depending on where the polygon boundaries are drawn. To escape this problem, kernel density estimation (KDE) can be used to create maps showing hotspots of TOMV offences based on the original point data.

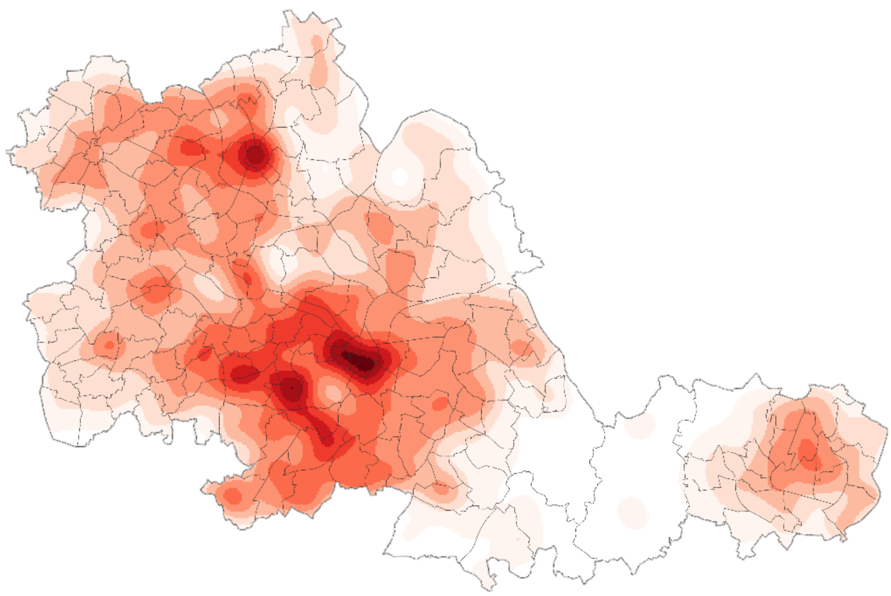


*Figure 9: KDE heat map of all stolen vehicles over last 12 months. Highlights the city centre and Birmingham West as areas where there are a large number of incidents.*

Hot spot analysis can be used to inform locations that are likely to have offences, however this approach assumes that current clusters of crimes (hot spots) persist, with future incidents predicted to happen in the same places. To test the accuracy of this approach, during testing of the model outputs, incidents' locations from the previous week are compared to the predicted locations outputted from the model.

# 3 Methodology

## 3.1 Spatio-Temporal Analysis

Spatio-temporal analysis is used to model the relationship using both spatial and temporal information, and produce forecasts of a response variable. In this project the spatial element is the location of offences within the WMP region, and the temporal element is the date that the offence took place. The response variable to be modelled is the number of TOMV offences (per week), in the form of count data.

Originally the data in the dataset was formatted as point referenced data, with locations given as eastings and northings. This was then split into 1,367 equally sized grid cells of approximately 0.75km$^2$. Within each of these grid cells, the number of TOMV offences to occur each week was counted. This created a low count, discrete dataset, heavily skewed with zero values where no offences occurred.

Undertaking purely spatial analysis does not allow for forecasting into the future, but only considers what has already happened across space. Methods such as universal kriging can be used to interpolate the response variable across space with unknown values (Graler et al., 2016). This requires a number of explanatory variables in the modelling process and is most useful in a Geostatistical setting (Moraga 2019). In the current problem, the response variable is made up of count data, rather than a continuous numeric variable. Predictive spatial methods alone are not suited to this as we are not trying to interpolate things that have already happened, but instead we are trying to forecast things that are yet to happen. Additionally, there is no requirement to predict the occurrence of offences outside of the force area, and there is a lack of available explanatory variables.

In order to spatio-temporally forecast crime occurrences in the future, there must be dependence over both space and time. Temporal dependence can exist in the form of repeated patterns such as seasonality, or trends. Due to the large increases seen in the number of TOMV offences in the region, minimal seasonality has been seen over the last two years. Instead the number of cars stolen each month is continuing to increase. Spatial dependence exists in the form of spatial autocorrelation which was confirmed to be present in the data (Section 2.5, Clustering and Hot Spot Analysis2.5).

The selected method for this project was Generalised Additive Modelling (GAM) which was compared against other methods and selected for its speed and accuracy (Wang and Brown, 2012)(for explanations of other methods considered please see section 5, Appendix).

## 3.2 GAM Workflow

Data was formatted for modelling by counting the total number of incidents in each 0.75m$^2$ grid cell, each week, with any grid cells having no incidents being assigned a zero count. Any points lying outside of the force area were removed from the dataset. Five years of data equated to 261 weekly groupings. For the final GAM modelling only 12 weeks of data were used at a time. Due to the increasing level of this crime type over the last two years, minimal seasonality has been seen, therefore more accurate results were seen when a shorter, more recent, number of weeks were used to create the GAM model.

This modelling method assumes additivity between predictor variables (but takes non-linearity into account). The formula used for modelling was

$$y \sim s(grid) + s(week) + linked\ crimes + max\ weeks$$

Where y is the response variable to predict (count of TOMV incidents) and 'linked crimes' and 'max weeks' are the explanatory variables provided to the models. Two smooth terms (s) are included, 'grid', which gives an index number of the grid cell for the count and 'week' which gives the week number of the count. A negative binomial family distribution was used for model fitting.

Linked data: for each offence included in the TOMV dataset, any crime committed by the same suspect or offender in the same time period, within the West Midlands region was grouped into a linked crime dataset. These offences were only those that were not TOMV offences to avoid duplicated counts. This dataset was counted across the same grid as the TOMV offences, to give a weekly count, per grid. A lag of one week as added to this variable (e.g. week 14's linked crime count was used with week 15's TOMV count). This explanatory variable was added to better inform the model using available data on offenders of vehicle offences.

Max weeks: the second explanatory variable included in the GAM model was the maximum number of weeks across the whole time period that one grid cell saw no TOMV offences take place within it. To calculate this the number of consecutive zeros for each grid cell were counted, and the maximum number used for this variable.

A neighbourhood matrix was created using the grid object, giving a list of the neighbouring cells of each cell. This is used within the modelling to account for spatial autocorrelation between grid cells. It is included within the smooth for the 'grid', along with a Markov random fields smoothing basis spline which is appropriate for spatial data.

## 3.3 Model Validation and Selection

The response variable was modelled using 12 weeks of data (training set), and predictions outputted for the following week. The results of this were validated against the real values for the following week (test set). The central prediction from this model was converted to a binary variable (0 or 1), however as all values outputted from the prediction were greater than 1, the 'top values' were selected and converted to have a value of 1, with the rest assigned a value of 0. Multiple sets of 'top values' were selected based on the top percentiles between 90 and 70. Each set of values were compared against the test set, and a confusion matrix calculated. Two metrics were calculated from the confusion matrices for each set of values, and the set that gave the most accurate results was selected.

$$\boldsymbol{Metric\ 1}: ((Precision * 32000) - ((1 - Specificity) * 32000)) * (Sensitivity/0.5)$$

$$\boldsymbol{Metric\ 2}: Precision * 32000$$

To select the top percentile of values that gave the best model output, metric 1 was divided by metric 2 and the percentile of values that gave the highest value was selected.

This metric was defined specifically for this project to balance model predictions against resource cost for WMP, considering both specificity and sensitivity in relation to precision. The value of 32000 is the approximate cost of a PC per year and was used to assign a cost value to the model results[1].

The model was then rerun using the 12 most recent weeks of data and used to forecast forward one week (future week). The selected percentile of values was taken forward and adjusted using a time series prediction of the forecasted week. Each time the model is run, the most accurate split of top percentile of values is chosen, this is a dynamic selection based on the results of the training – testing model run.

The time series used to forecast a weekly value for the forecasted week was a Bayesian structural time series using a semi local linear trend and a seasonal component with 52 seasons (to allow for the weekly seasonality). The data was modelled using a training data set of 5 years minus one week, and validated using the held-out week. The model was then rerun using all available data to forecast the next 4 weeks. The first value from this prediction was used to adjust the spatial predictions to ensure they summed to the same value. The spatial forecasts were adjusted using the time series so that the sum of the predictions equals the sum of the time series forecast value for the same week.

[1] This can be thought of as return – cost / return (with an adjustment for sensitivity), or essentially the "profit" as a proportion of the total return. If the false positives (1 – specificity) are used in the denominator, the same ranking of models is derived.

## 3.4 Forecasting

One week ahead forecasts for TOMV offences, displayed as a hot spot grid map over the region of the West Midlands. The values were split into low (cream), medium (pink), high (red) and very high (dark red) categories.

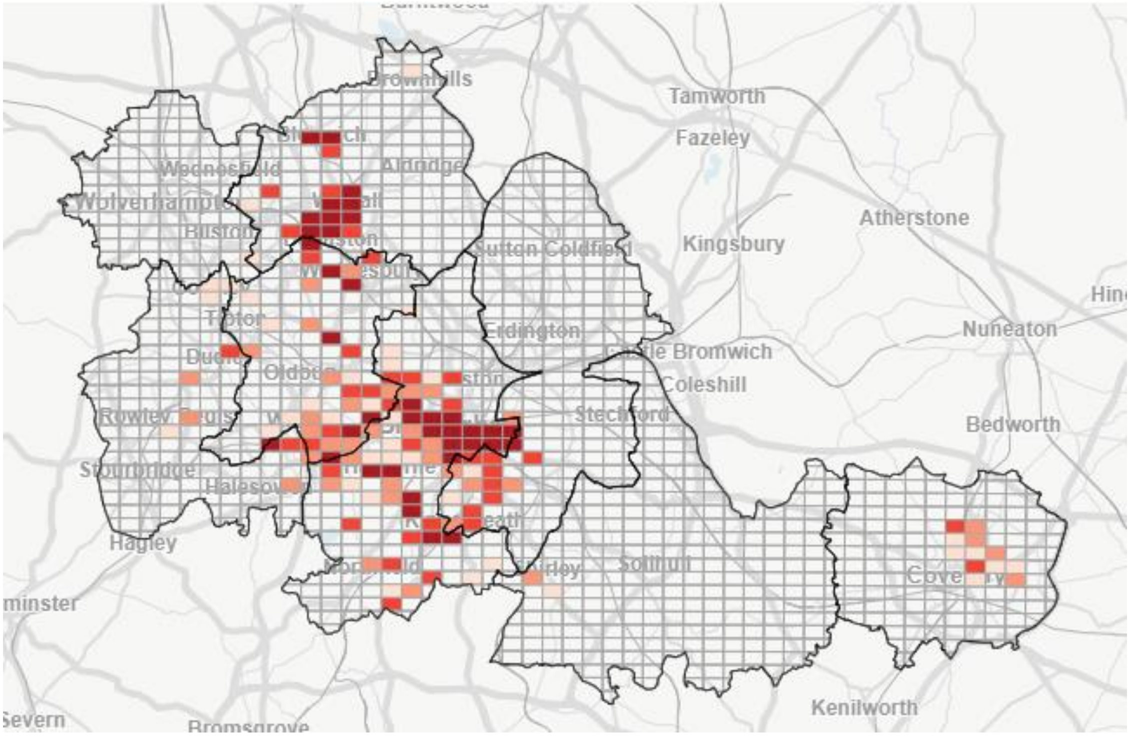



*Figure 10: Example output of one-week ahead predictions of most vulnerable areas for stolen vehicles in the West Midlands region. Dark red = very high, red = high, pink = medium, cream = low.*

## 3.5 Model Results

During model testing, predicted values from the model were compared with known values for the predicted week. These results can be compared against just taking a hot spot approach by using incident locations from the previous week to be the values for the following week. Using predicted values outputted by the GAM model yielded more precise results and a higher final overall metric, that considers sensitivity, specificity and precision. This shows that using the modelling method to output one-week ahead forecasts is more accurate than using the previous weeks' locations. When comparing model results using 12 weeks of data to using 52 weeks (1 year), a higher overall metric and model accuracy was seen. This also allowed the predictions to be more dynamic and update reflecting more recent activity.

| | **Sensitivity** | **Specificity** | **Precision** | **Metric 1** | **Metric 2** | **Overall Metric** | **Accuracy** |
|---|---|---|---|---|---|---|---|
| **Predicted Values from Model (using 12 weeks data)** | 0.32 | 0.89 | 0.35 | 5000 | 11225 | **0.44** | **80%** |
| **Predicted values using previous Week's Values** | 0.22 | 0.85 | 0.22 | 985.60 | 7040 | 0.14 | 76% |
| **Predicted Values using 1 year of data** | 0.36 | 0.86 | 0.33 | 4646 | 10560 | 0.43 | 79% |

# 4 Project Output

The results of this project were output into a Business Insights (Qlik) dashboard. The aim of this dashboard was to display the one-week ahead forecasts highlighting the areas of the West Midlands TOMV offences are more likely, as well as displaying available information on past offences in an easily accessible way. The dashboard has been made available for alpha testing to end users to ensure its features support operational activity. Their feedback will be used for further development of the dashboard.

### Theft of Motor Vehicles (TOMV) Dashboard

**Sheet 1:** One-week ahead forecasts highlighting vulnerable areas for TOMV incidents. Forecasts available for 'All vehicles' or eight vehicle categories. For details of which vehicles are in each category, please view available past data. Accuracy of these spatial forecasts is approximately 80%.

Past incident locations can be viewed alongside incident details, with a summary chart of the most vulnerable vehicle makes and models and total number of vehicles stolen for the selected time period.

**Sheet 2:** Weekly count of TOMV incidents with a 4-week-ahead time series forecast. This is for 'all vehicles' total or can be split into the same eight categories as the spatial forecasts. Approximate accuracy is 95% (Mean Absolute Percentage Error = 5%)

Details of the data included in this dashboard:

| | Included in Dataset | Excluded from Dataset |
|---|---|---|
| **Crime Type** | **Theft or unauthorised taking of a motor vehicle:**<br>• Unauthorised taking of a motor vehicle<br>• Theft of a motor vehicle<br>• Attempted – Theft of a motor vehicle | • Theft from motor vehicle<br>• Aggravated Vehicle Taking<br>• Residential Burglary where car keys were stolen |
| **Location** | Crimes with location information available | Crimes with unknown, NA or inaccurate location |
| | | Crimes committed outside force area |
| **Role of Vehicle in Crime** | **STOLEN** vehicles | Damaged / Used / Recovered / Found Vehicles |

| Vehicle Categories: |
|---|
| Supermini |
| Van |
| SUV |
| Small Family Vehicle |
| Large Family Vehicle |
| Motorcycles |
| Executive/Sports |
| Other Vehicles |

See next page for examples of vehicles in each category.

*Figure 11: Introduction Page for Qlik Dashboard giving overview of each sheet and information about which data is included in the dashboard.*

| Vehicle Categories: | Examples of vehicles included in category |
|---|---|
| Supermini | Ford Fiesta, Vauxhall Corsa, Vauxhall Astra, Fiat 500, Toyota Yaris |
| Van | Ford Transit, Mercedes Sprinter, Peugeot Boxer, Citroen Relay, Renault Traffic |
| SUV | Land Rover Range Rover, Land Rover Discovery, Ford EcoSport, Nissan QashQai |
| Small Family Vehicle | Ford Focus, VW Golf, Audi A3, BMW 1 Series, Honda Civic, Toyota Corolla |
| Large Family Vehicle | BMW 3 Series, Audi A4, Mercedes C Class, Vauxhall Vivaro, Ford Mondeo |
| Motorcycles | Motorcycles and Mopeds |
| Executive/Sports | Audi A6, Jaguar XJ, BMW 5 Series |
| Other Vehicles* | Where vehicle is unknown, or only manufacturer is known but not the vehicle model |

*Figure 12: Examples of types of vehicle included in each category for the dashboard visualizations. This is not an exhaustive list but the most common vehicles in each category.*

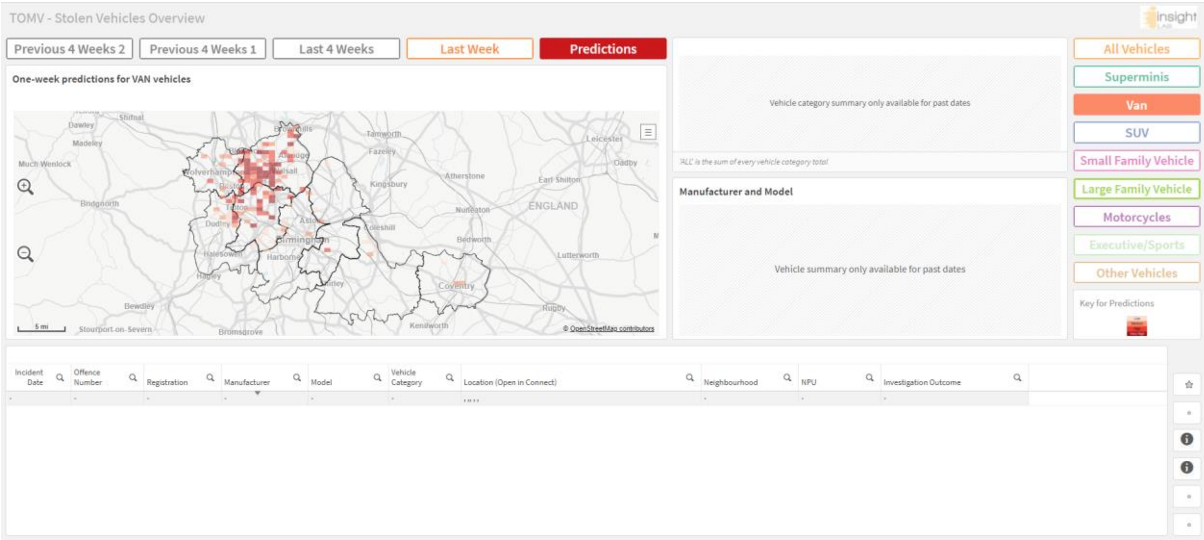



*Figure 13: Example of Qlik Dashboard, filtered to vehicle category 'Van', showing the predicted one-week-ahead values for most vulnerable areas for this type of stolen vehicle.*

TOMV offences having occurred in the past are provided; where available the offence number and location were hyperlinked to the Connect records, which when clicked means they automatically open in the Connect App. It also shows summary charts grouped by vehicle manufacturer and model, and by vehicle category, to highlight the vehicles more likely to be taken in the selected time period. There are four options for available time periods to select: last one week, last four weeks, and then the two four-week periods before that one. When selected these will show locations of all offences that took place in these time periods.

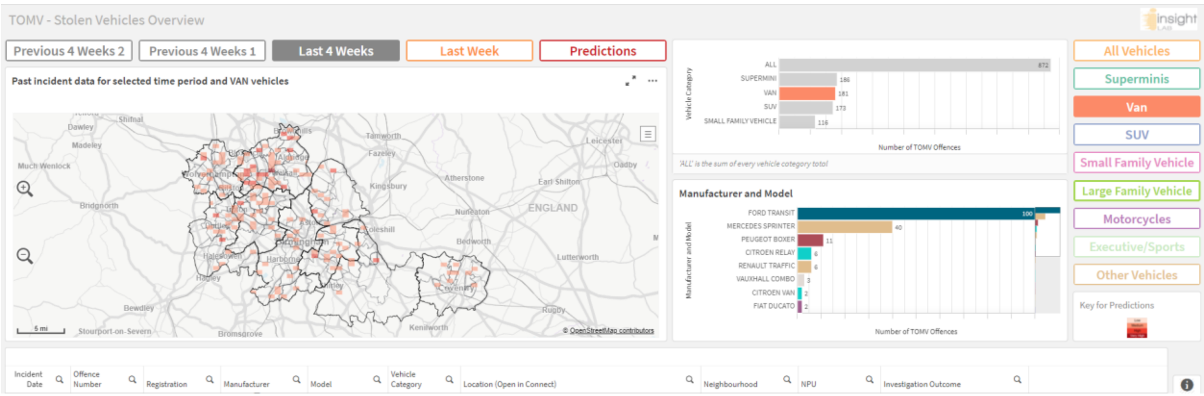



*Figure 14: Example of Qlik Dashboard, filtered to vehicle category 'Van', showing the previous four weeks offences, summary of vehicle manufacturers and models, and offence information below.*

Data displayed on the dashboard can be filtered down by any of the selectable items, such as NPU, neighbourhood, vehicle manufacturer and model. When an NPU or neighbourhood is selected, the map will automatically zoom into the selection and show the exact point locations of the TOMV offences.

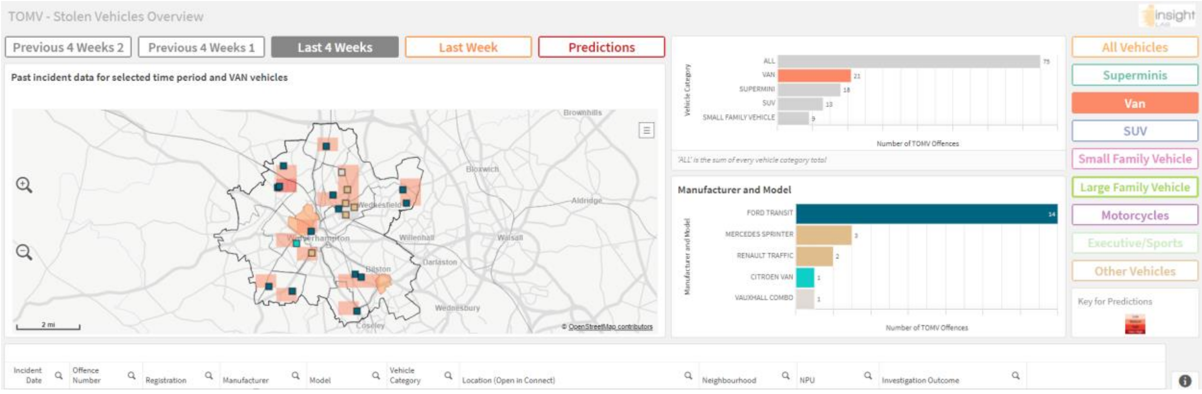



*Figure 15: Example of Qlik Dashboard showing all Vans stolen in Wolverhampton NPU over the last four weeks, highlighting point locations coloured by vehicle manufacturer.*

The second sheet of the dashboard shows four-week ahead forecasts for total weekly count of all TOMV offences (including those in the spatial forecasts and those that we have no location information for). Above the time series chart there is a breakdown of the number of offences for two time periods, last week and last four weeks. This shows the total offences count, the number of those offences that have location information available and the number of vehicles recovered. This sheet can also be filtered down into the eight vehicle categories.

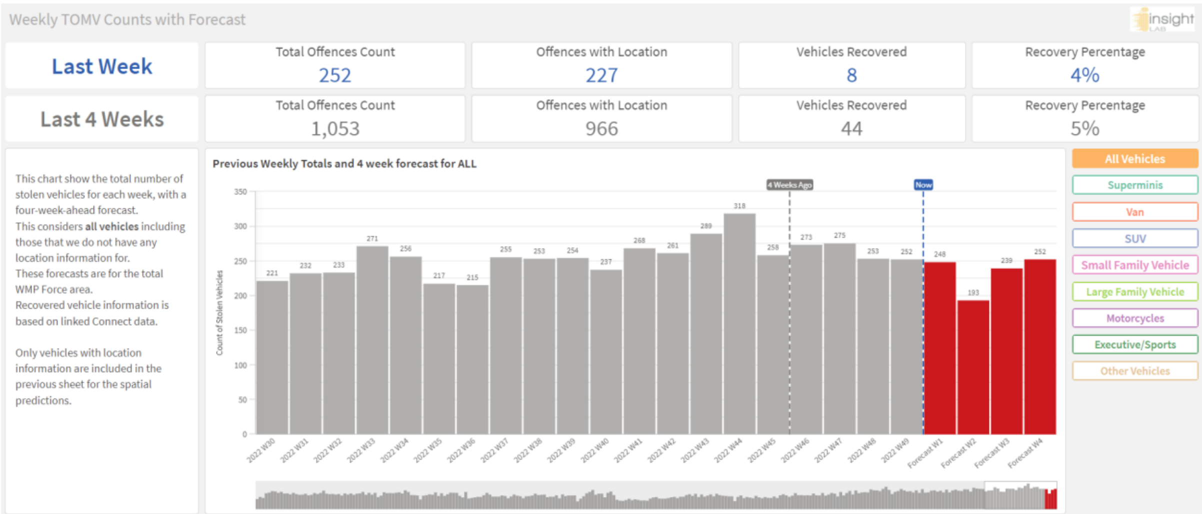



*Figure 16: Example of Sheet 2 of the Qlik Dashboard showing a weekly time series count for all stolen vehicles and a four-week ahead time series forecast.*

# 5 Appendix

## 5.1 Other Methods Considered

Integrated Nested Laplace Approach (INLA) approximates the posterior distribution of Bayesian analyses via inferences. Structural spatial correlation was modelled using a Besag, York, Mollie model (this includes spatially structured and spatially unstructured elements). This model accounts for spatial autocorrelation in that the grid cells that are adjacent in space show more similar numbers of TOMV offences than areas that are not neighbours. Spatial models consider spatial autocorrelation in order to separate the general trend from the random spatial variable. Spatial autocorrelation suggests that observations that are close in space are likely to show similar values. INLA modelling can be approached in two ways, the SPDE with a mesh approach, and the standard approach. In both cases, the response values to be predicted are included in the original model as NAs (missing values), and they are then treated as any other parameter in the model. The predictive distribution of the fitted values is automatically computed and the marginal fitted values of the missing observations can be returned.

The SPDE approach considers the exact locations of observations as it does not bin them into cells as INLA does. A triangulated mesh is created and it can more easily handle non-rectangular regions without wasting computational effort. On top of the mesh a Gaussian Markov Random Field representation is built. The correlation between the INLA model and GAM model predictions was 98%, however the GAM modelling approach was significantly faster and therefore GAM was taken forward as the chosen model.

# 6 References


Anselin, L., (1995) Local Indicators of Spatial Association, Geographical Analysis, Vol 27(2), pp93-115

Cressie, N., Zammit-Mangion, A. and Wikle, C.K. (2019) *Spatio-temporal statistics with R*. Chapman and Hall.

Graler, B., Pebesma, E. and Heuvelink, G. (2016) Spatio-Temporal Interpolation using gstat. R Journal, 8(1), p.204.

Moraga, P. (2019) Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. Chapman and Hall / CRC Biostatistics Series.

Society of Motor Manufacturers and Traders (SMMT). (2022) SMMT Car Dales Data for Q4, FY 2021 and Q3 2022.

Vehicle Licensing Statistics (2021) April to June 2021 (publishing.service.gov.uk)

Wang, X. and Brown, D. (2012) The spatio-temporal modeling for criminal incidents. Security Informatics 1(1), p.1-17.