WMP Briefing Paper

Predicting Theft of Motor Vehicle Offences

This project is at the proposal stage and is presented to the committee 'in principle' so that any immediate concerns can be raised.

The finer details of the methodology, exact data to be used and mode of communicating the results will not be determined until after the exploratory data analysis (EDA) phase has been undertaken.

Once the analyses have been completed the project will be presented to the Committee again so that the data used, methodology, findings, intention for deployment and communication plans can be examined in more detail.

Legal opinion has been sought and the Data Protection Impact Assessment (DPIA) is being reviewed by the Force Data Protection Officer (DPO).

Tasking

This project was requested by Superintendent Jim Munro on 07/06/2022 in response to a recommendation in the recent vehicle crime problem profile completed by the Intelligence Department and endorsed by the Force Performance Panel. Superintendent Munro leads the Vehicle Crime Task Force.

Purpose

The aim of the project is to predict the locations of Theft of Motor Vehicle (TOMV) offences for the top five vulnerable vehicle types.

The purpose is to provide the Force with a more informed picture of TOMV offending, which has been escalating and is a tactical priority. This will enable the Force to tackle offending by targeting Neighbourhood Policing Unit (NPU) activity effectively, together with appropriate crime prevention operations.

Context

The volume of Theft of Motor Vehicle[1] offending has escalated over the last three years (*Fig 1*) and tackling this crime type has been a Force tactical priority throughout 2022. A Vehicle Crime Task Force has been created to focus efforts on this type of criminality.
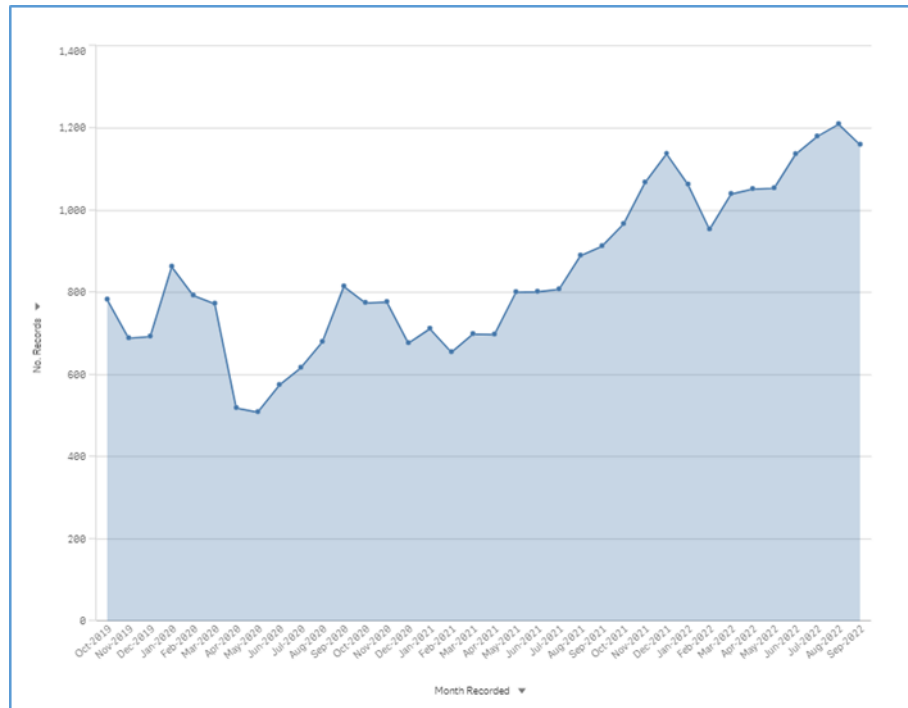

Figure 1: Theft of Motor Vehicle Offences in the West Midlands Oct 2019 - Sept 2022

At present, there is an international supply chain shortage affecting vehicle production and availability within the United Kingdom (UK). This has resulted in a shortage of computer chips in car production, as well as other essential materials such as copper, aluminium and cobalt, leading to fewer new vehicles being produced and consumers waiting a long period of time for new vehicles.[2] 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.[3] Nationally in 2021, the most stolen vehicle make was Ford followed by Jaguar Land Rover (JLR)[4]. Ford is the most commonly owned vehicle in the UK[5] and has easily interchangeable parts between models and other makes, making them an attractive target for their parts.

---

[1] The Home Office Counting Rules (HOCR) for Recorded Crime state that Theft of Motor Vehicle offences include Theft of a Motor Vehicle (48/1) and Unauthorised taking of a motor vehicle (130/1) under the Theft Act 1968. It does not include offences where a house is burgled in order to steal the vehicle, or where the offence is a robbery, for example where the victim is forced out of their car at knife point
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1099880/hocr-complete-july-22.pdf

[2] Second Hand Car Prices Surge, BBC News, 2021.

[3] Positive outcome rate for TOMV 2022/23 to date

[4] This is Money, Britain's Most Nicked Motors of 2021, 2022 & The Most Stolen UK vehicles in 2021, Express, 2022 and REVEALED: Thieves' favourite cars in West Midlands - West Midlands Police & Crime Commissioner (westmidlands-pcc.gov.uk)

[5] Vehicle Licensing Statistics: April to June 2021 (publishing.service.gov.uk) 2021.

The Force has been undertaking wider intelligence and investigative work to understand why vehicles are being stolen, the prevalence of 'chop shops'[6], whether vehicles remain intact, whether vehicles or parts remain in the UK, whether they are used to commit further crime, how they are stolen and whether they are stolen to order as part of wider organised criminality, which may involve the exploitation of vulnerable individuals.

To complement this wider work, the Data Analytics Lab (DAL) has been tasked to build a model which predicts where offending hotspots are more likely to be located. The analysis will be focused on the top five most vulnerable vehicle types for theft as identified by the motor industry, based on their knowledge of current supply issues around specific vehicle types and parts, as well as West Midlands Police (WMP) crime data.

Intended activity resulting from the project

The intention is that the predictive model will be 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 offenders' 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.

Ethical considerations

The DAL asks the Committee to advise whether there are any ethical concerns if WMP were to make resourcing decisions to tackle vehicle theft using information based on a statistical predictive model, in conjunction with other, more traditional approaches. The likely outcome is that some areas of the Force will receive more and others less of the resources available to tackle vehicle crime.

---

[6] Chop shops are workshops where car thieves can sell stolen vehicles where they are stripped for parts and can be used to repair a car that has been written off by an insurance company. WMP: What is a chop shop?

Data

Data to be used:

Level of analysis:

☐ Individual
   Individuals aggregated?
   ☐ Yes
   ☐ No
☐ Specific Area:
   ☐ Output Areas
   ☐ Super Output Areas - Lower
   ☐ Super Output Areas - Mid
   ☐ Wards
   ☐ Districts
☑ West Midlands
☐ Other (type of crime)

The unit of analysis will be crimes recorded as Theft of Motor Vehicle. Data relating to individuals who are the victims, offenders or suspects will not be included in the analyses.

Reliability of data:

An extensive exploratory data analysis (EDA) phase will be undertaken to examine the extent of any data quality issues. The data comes from the WMP systems CONNECT and ControlWorks, used for recording crimes and incidents respectively. In addition, information supplied by the industry relating to the theft of car makes and types will be considered.

Discussions with subject matter experts (SMEs) will be undertaken to capture additional requirements and to sense check the analyses.

Sample or entirety: Entirety

Type of analysis:

☐ Exploratory
☐ Explanatory
☑ Predictive
☐ Optimisation

Proposed methodology:

1. What are the top five types / makes of cars that are stolen?
2. Undertake relevant EDA, including mapping, spatial cluster analysis, temporal analysis, proportion of which manufacturers' cars get stolen, etc.
3. Spatio-temporal prediction model build; consider use of spatio-temporal predictive kriging.
4. Presentation of results on a Qlik app.
5. Qlik app design and build in conjunction with Visualisation and Business Insight Developer

Will the project eventually be automated:

☑ Yes
☐ No
It is envisaged that a dashboard will be built which updates on a regular basis.

Means of evaluation: The performance of the prediction will be monitored on an on-going basis.

ALGO-CARE considerations

As this project is at the proposal stage and is presented to the committee 'in principle' in order that any immediate concerns can be raised, the finer details of the methodology will not be determined until after the EDA. Once the analyses have been completed the projects will be presented to the Committee again so that findings and methodology can be examined in more detail.

Advisory	
If applicable, are the outputs from the algorithm to be used in an advisory capacity?	The output of the prediction would be visualised in a Business Insight (Qlik) dashboard. This would be provided to Intelligence Analysts to feed into their threat and risk assessments of vehicle crime. In conjunction with other data and intelligence, the analysts will provide recommendations to decision makers about where best to focus resources to tackle this crime type. This forms part of force tasking processes at force-wide and local level, where resourcing decisions are taken.
Does a human officer retain decision-making discretion?	Yes, the predictions are for information and will contribute to wider analysis. Resourcing decisions will be made using the normal tasking forums taking this and other information, such as local intelligence, into account.
Lawful	
What is the policing purpose justifying the use of the algorithm (means and ends)?	This project supports the Force Strategy and the *Precision Policing Doctrine* by ensuring that resourcing decisions are based on data and evidence.
	Reducing Theft of Motor Vehicle has been a Force priority throughout 2022. The continuing escalation of offending means that WMP must make effective use of its resources and ensure they are targeted at the right locations. In addition to financial losses for victims of vehicle crime, the prevalence of 'chop shops' means that there is potential for unsafe cars to be sold and driven by unwitting purchasers; and the involvement of organised crime groups means vulnerable individuals may at risk of exploitation.
Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes?	There will be no interference with the privacy of individuals since the analyses will examine events rather than nominals (victims, suspects or offenders).

In what way will the tool improve the current system and is this demonstrable?	There are currently no means of predicting the locations of TOMV within WMP more widely, the expertise and tools to build a predictive model sits within the Data Analytics Lab.
	Once the model is developed an assessment can be made as to its performance, compared to current methods of reviewing the vehicle theft hotspots.
Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim?	The data are from WMP systems and are collected as part of normal operational activity.
Is the operation of the tool compliant with national guidance?	The analyses proposed would accord with the Government Digital Service Data Ethics Framework 2020[7]
Granularity	
Does the algorithm make suggestions at a sufficient level of detail given its purpose and the nature of the data processed?	Given the overall aim of predicting the number and locations of TOMV offences, using crime data aggregated over time and to bespoke spatial units would enable better decision making within WMP.
Are data categorised to avoid broad-brush grouping and results and therefore issues of potential bias?	Using aggregated counts would best suit the aims of the project, but would not involve any other form of categorisation (given that crimes would be the unit of analysis).
Do the potential benefits outweigh any data quality uncertainties or gaps?	The project will include an extensive EDA element and this should highlight areas of heightened uncertainty in the data or where particular gaps exist. Should any such issues be identified, these would be addressed as a part of the project. Given the benefits of reducing TOMV offending it is not expected that any data quality issues would be of such a magnitude as to warrant not undertaking the project.
Is the provenance and quality of the data sufficiently sound?	The data have been gathered during the day-to-day work investigative of WMP and will enable analyses of the type envisioned for this project.
If applicable, how often are the data to be refreshed?	To be agreed with end users – likely to be monthly as a minimum, probably more frequently.

[7] https://www.gov.uk/government/publications/data-ethics-framework

If the tool takes a precautionary approach in setting trade-offs, what are the justifications for the approach taken?	Ultimately any model developed would aim to maximise specificity whilst trying to gain as high a sensitivity as possible. This approach would mean that we could best allocate WMP resources whilst ensuring a minimisation of false positives.
Ownership	
Who owns the algorithm and the data analysed?	WMP would own the analyses and the data.
Does WMP need rights to access, use and amend the source code and data?	No
Are there any contractual or other restrictions which might limit accountability or evaluation?	No
How is the operation of the algorithm kept secure?	The data and the analyses are contained wholly within the WMP system and the security measures employed therein.
Challenge	
What are the post-implementation oversight and audit mechanisms, e.g. to identify any bias?	For any model that is developed and productionised, checks will be made as to its accuracy on an on-going basis (overall accuracy, sensitivity, specificity, etc.) as well as any consistent patterns that may represent biases.
If the algorithm is to inform criminal justice disposals, how are individuals notified of its use?	Not applicable
Accuracy	
Does the specification of the algorithm match the policing aim and decision policy?	The model would aim to produce information to aid decision making within WMP which would be in line with its aims and policies.
Can the accuracy of the algorithm be validated periodically?	The productionisation of any model resulting from the project would include checking its accuracy on an on-going basis.

Can the percentage of false positives / negatives be justified?	Not yet known, however any model developed would aim to maximise specificity whilst trying to gain as high a sensitivity as possible.
How was the method chosen as opposed to other available methods?	Currently the broad approach has been identified due to the nature of the business question and the data available.
What are the (potential) consequences of inaccurate forecasts?	The main issues arising from inaccurate forecasts would be (a) potential for actions for WMP that may not be necessary (including in particular locations) and (b) WMP resources being allocated ineffectively.
Does this represent an acceptable risk?	Any model arising from this project would seek to balance the advantages against the risks arising from inaccurate predictions partly via balancing the model's sensitivity and specificity and partly through assessing the types of decisions for which any such model would be effective for and any actions that would arise from these decisions. This would be subject to periodic review.
How are the results checked for accuracy and how is historic accuracy fed back into the algorithm for the future?	For any model that was productionised, it's accuracy would be assessed on an on-going basis via measuring its accuracy (sensitivity, specificity, etc.) as well as producing histograms of counts so that any degradation of the model could be tracked and the model rebuilt if necessary.
How would inaccurate or out-of-date data affect the result?	This is partly dependent on the nature of any model should one be capable of being built. Generally inaccurate or out-of-date data could detrimentally impact on the model's performance (in terms of accuracy) and lead to inefficient decision making and resource deployment.
Responsible	
Would the operation of the algorithm be considered fair?	During the development of any model, the presence of any biases in the underlying data or for predictions to produce biases would be fully examined and mitigated if the potential was present.
Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review?	The details of any model arising from this project would be provided and, as mentioned above, when productionised there would be on-going checks as to model performance.
Would it be considered to be used in the public interest and to be ethical?	The aim of the project is to enable the reduction of vehicle theft. Given the financial cost and potential for harm to local communities caused by this offence type it is likely to be viewed positively and as an ethical use of police data.
Explainable	

Is information available about the algorithm / decision-making rules and the impact of each feature?	A technical report will be produced which will include information about the methods used and assumptions made.

Appendix 1: Glossary of Terms

WMP / Law Enforcement Terminology	
DAL	Data Analytics Lab
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
FET	Force Executive Team
HOCR	Home Office Counting Rules (for recorded crime)
NPU	Neighbourhood Policing Unit
OPCC	Office of the Police and Crime Commissioner
PCC	Police and Crime Commissioner
SME	Subject Matter Expert
TOMV	Theft of Motor Vehicle
WMP	West Midlands Police

Data Science Terminology	
ALGO-CARE	All projects have used the ALGO-CARE to consider ethical implications: Advisory, Lawful, Granularity, Ownership, Challenge, Accuracy, Responsible, Explainable
AUC	AUC stands for 'area under the curve' of a ROC (Receiver Operating Characteristics) curve. Essentially, ROC is a probability curve and the AUC tells us how good the model is at distinguishing between different groups within the data. This is a statistical test for the accuracy of the model that has been built.

EDA	Exploratory Data Analysis
Productionise	To 'productionise' means that once we are satisfied that the model works well, we would automate the process of providing predictions on a regular basis.
Spatio-temporal model	Spatio-temporal models use patterns evident over space as well as patterns evident through time – "everything is related to everything else, but near things are more related than distant things" (Tobler's first law of geography). This is often true in time as well as over space.
Sensitivity	Refers to the ability of the model to identify the 'true positives' as a rate. It measures the proportion of actual positives that are correctly identified as such. The greater the sensitivity of a model, the less the specificity will be.
Specificity	Refers to the ability of the model to identify the 'true negatives' as a rate. It measures the proportion of actual negatives that are correctly identified as such. The greater the specificity of a model, the less the sensitivity will be.