# **Crime Seasonality Planner**

Data Analytics Lab

July 2022

This project aims to redesign the current seasonality planner using time series methods and output an interactive Qlik dashboard for use within the force.

# **Table of Contents**

1	Int	troduction3				
2	Ove	erview of Current Seasonality Planner	4			
	2.1	What is it used for	4			
	2.2	Methods	4			
	2.3	Why does it need updating?	6			
3	Dat	ta	7			
	3.1	Crime data	7			
	3.2	Covid-19 related data	8			
4	Me	thodology	11			
	4.1	Time Series Methods	11			
	4.2	Model Validation and Selection	11			
	4.3	Forecasting	13			
5	Pro	eject Output	15			
	5.1	Welcome Page	15			
	5.2	Seasonality Profile Overview	15			
	5.3	Seasonality Profile Comparison	16			
	5.4	Time Series	16			
6	Fut	ure Updates	18			
7	Apj	pendix	19			
	7.1	Crime Types	19			
	7.1	.1 Serious Violence (VWI, Knife Used, Injury Caused)	19			
	7.1	.2 Rape	19			
	7.1	.3 Stalking and Harassment	20			
	7.1	.4 Other Crimes	20			
	7.1	.5 Anti-Social Behaviour (incidents)	21			
	7.2	Details of Time Series Methods	22			
	7.2	.1 Model Validation and Selection	22			
	7.2	.2 Classical Decomposition	22			
	7.2	.3 Simple Exponential Smoothing	22			
	7.2	.4 Croston's Method	23			
	7.2	.5 ARIMA	23			
	7.2	.6 Bayesian Structural Time Series	23			
8	Ref	erences	30			

### 1 Introduction

The seasonality planner is used to show monthly variations in reported crime demand for individual crime types across the Force and at an NPU (Neighbourhood Policing Units) level. Seasonality can be described as the expected variation from the trend over time, and is calculated using past data. Some crimes are highly seasonal, and others only present weak seasonality. In a policing context, calculating seasonality of crime can assist in planning resource allocations on a force-wide strategic basis.

The existing seasonality planner is calculated using basic time series decomposition methods in a spreadsheet. It involves time-consuming manual data inputs and as a result has historically only been updated once per year. The output is a static seasonality grid listing the seasonality of all selected crime types at a force level, and broken down by individual NPUs. Due to the impacts of the Covid-19 pandemic on crime levels, the seasonality planner has not been updated since the beginning of 2020. Using more sophisticated methods, the aim of this project is to redesign the seasonality planner to create a more user-friendly output that considers the effects the pandemic has had on crime and continues to show seasonality of a wider range of crime types. The new output will be in the form of a Business Insights (Qlik) dashboard and it will have the possibility of more frequent updates if required.

The format of this report is as follows; section 2 gives an overview of the current seasonality planner, and the need for its redesign. Section 3 includes information about the data used in the new seasonality planner. Section 4 explains the methodology, workflow and results of the new seasonality planner. Section 5 introduces the project output, an interactive Business Insights (Qlik) dashboard. Section 6 is the Appendix and contains more detailed information about the methods used and any supplementary information.

# 2 Overview of Current Seasonality Planner

#### 2.1 What is it used for

The current seasonality planner is used to calculate demand indexes for each month of the year, for selected crime types or offences at Force (WMP) and Neighbourhood Policing Unit (NPU) level. It provides a narrative for crime trends within the West Midlands Police (WMP) region. It is widely used to assist the monthly Force Tasking and Delivery Board (FTDB), and quarterly Strategic Tasking and Co-ordination Group (STCG) as well as other governance boards and for planning relevant resource allocations on a force-wide strategic basis. It is used to see if expected crime levels are met, or if they are deviating from previously seen trends.

Some crimes are already known to show a strong seasonal pattern:

- Burglaries tend to increase in the Autumn with the advent of 'darker nights'.
- Violence in public spaces tends to increase during the summer months when there are more opportunities for socializing mixed with alcohol consumption.
- Robberies are seen to increase at the start of Autumn and Spring terms when teenagers return to school with new CRAVED (Concealable, Removable, Available, Valuable, Enjoyable and Disposable) items, such as mobile phones.

Examples of crime types which may not be suitable for inclusion in the seasonality planner would be crimes that are reported retrospectively (e.g. child abuse), or crimes which are relatively rare occurrences (e.g. homicide).

#### 2.2 Methods

The existing methods for the seasonality planner are shown in Figure 1. To summarise, they involve extracting monthly count data from the WMP systems and manually inputting it into a spreadsheet, where the seasonality index is calculated. These values are exported into the 'seasonality grid' which is the current output. This process must be repeated for each crime type, at a force and NPU level. The index values are split into three categories: high, medium and low. The grid is colour coded into these three categories as well as showing the percentage value. The cut off for the categories are: low is less than -5%, medium between -5% and 5% and high is any values greater than 5%.

Five years of data are used to calculate the seasonality indices. Classical time series decomposition is used in which a centered moving average (order 12) of the monthly crime counts is calculated as the trend-cycle component. This component is subtracted from the total values to leave the seasonal irregular component. The seasonal irregular component is then averaged over each month of the year to give a monthly seasonal index, which is normalized and converted to a percentage.

The seasonality planner has not been updated or recalculated since January 2020, due to the effects that the Covid-19 pandemic has had on crime numbers. Analysing seasonality at areas smaller than NPU level, such as neighbourhood level, is not currently practiced due to insufficient data availability to reliably see trends.

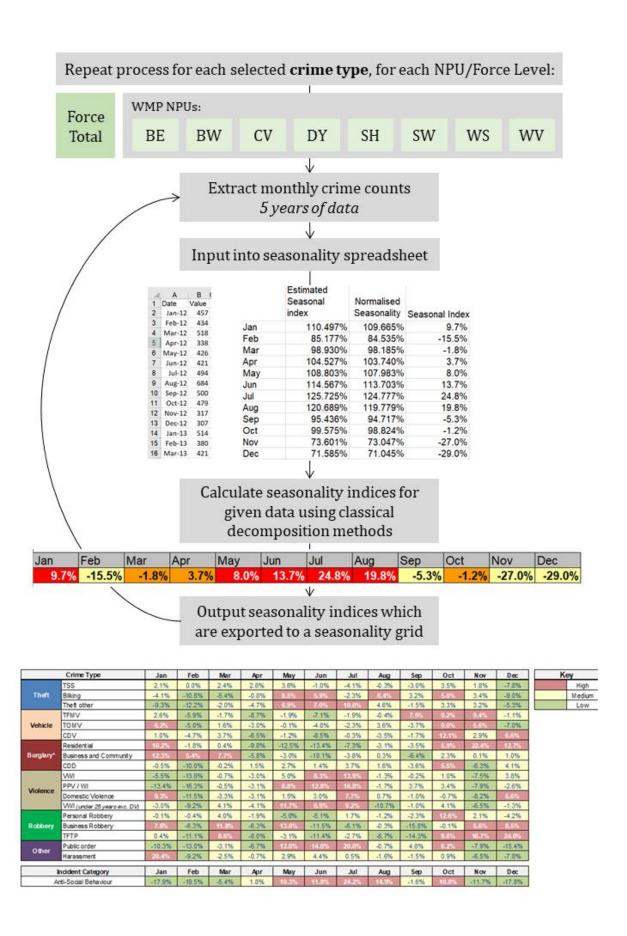


Figure 1 Workflow and methods for the current seasonality planner output.

# 2.3 Why does it need updating?

The seasonality planner output is important for many aspects of Force business, in particular for resource planning, annual leave embargos, timing of crime prevention public campaigns, planning partnership activities with other agencies, and minimizing the impact of internal changes during periods of expected high demand. The existing seasonality planner was developed some years ago using the tools and expertise available at the time, before the existence of the Data Analytics Lab (DAL). In addition, it is apparent that some traditional patterns and volume of offending have been disrupted by restrictions imposed during the Covid-19 pandemic. This project uses more sophisticated prediction of seasonal offending patterns, considering any potential impact of recent changes to these patterns.

### 3 Data

### 3.1 Crime data

Where possible, data for this project was sourced from the WMP Connect system, which has details about each crime that has been recorded. This system was introduced to the force on 2021-04-08. Data from the previous 'Crimes' system was combined with Connect data in situations where recording practices altered post-Connect changeover and the legacy data inputted into Connect did not fulfill the requirements. Anti-social behaviour incident data was sourced from the Controlworks (and the previous Oasis) system (the command and control system for WMP).

Unless otherwise stated, the date that the incident occurred was used, rather than the date reported or the date the record was created on the system as this is more likely to accurately reflect the seasonal patterns of crime occurrences.

Data between 2017-04-01 and 2022-03-31 is used for all crimes to be presented in the seasonality planner, this is 5 years of data. This data is aggregated on a monthly scale in the form of counts, of unique offence numbers. Data was calculated as a force-wide total for each crime type, as well as splitting it into the 8 NPUs (BE, BW, CV, DY, SH, SW, WS, WV). Seasonality on individual NPU level is not added together to create the force wide total, this is calculated separately. The force wide total may include more offences than the individual NPUs combined as some offences do not include exact location data. When incident location information was missing, but the incident postcode was present, the postcode was used to infer the NPU and fill in the missing data to ensure the NPU level data was as complete as possible.

The crimes included within the planner are those that were decided as part of the key performance indicators for 2022/23. Unless otherwise specified, crime types were defined using the home office main, sub and sub-sub groupings that are allocated on the Connect system for each crime. More details of the crime definitions used for these crime types can be found in the appendix (Section 7). For crime types such as domestic abuse and child abuse, 'non-crime' records were excluded. Cancelled crimes ('no crime' records) were excluded from all counts.

Selected crime types were also reported using two victim sub-categories: 'female victim' and 'victim under 25 years old' (Table 1). The sub-categories are included within the total for that crime type. The crime type 'Violence with Injury' (VWI) is further split into two sub categories of VWI DA and VWI Non-DA, as well as VWI All (which includes both DA and Non-DA offences). These sub-categories were included to assist with the local and national priorities to tackle violence against women and girls (VAWG) and serious youth violence.

Table 1 Crimes included in the seasonality planner. Selected crimes also reported with two victim sub-categories: female victim and victim under 25 years old.

Category	Category Crime Name			
Theft	Theft Shops and Stalls			
	Bilking			
	Theft from Motor Vehicle			
	Theft of Motor Vehicle			
Neighbourhood	Residential Burglary			
Crimes	Business and Community Burglary			
Crimes	Personal Robbery	$\checkmark$	$\checkmark$	
	Business Robbery			
	Theft from the Person			
	Violence with Injury	$\checkmark$	$\checkmark$	
	Violence with Injury (DA)			
	Violence with Injury (non-DA)			
	Violence without Injury	$\checkmark$	$\checkmark$	
	Serious Violence (*VWI, knife used,	$\checkmark$	✓	
Violence	injury caused)			
	Rape	$\checkmark$	$\checkmark$	
	Stalking (DA)	$\checkmark$	$\checkmark$	
	Stalking (non-DA)	$\checkmark$	$\checkmark$	
	Harassment (DA)	$\checkmark$	$\checkmark$	
	Harassment (non-DA)	✓	✓	
	Public Fear, Alarm or Distress			
	Hate Crime	$\checkmark$	$\checkmark$	
Other	Child Abuse			
Other	Domestic Abuse			
	Online Crime			
	Anti-Social Behaviour			

#### 3.2 Covid-19 related data

Restrictions imposed due to the Covid-19 pandemic had an impact on the level and composition of crime in England, in particular during periods of 'lockdown' in which people were told to stay at home (Kirchmaier and Villa-Llera, 2020). Langton et al. (2021) suggest that opportunities for residential burglary were decreased due to more 'daytime guardians' being present when stay at home measures were in place (March to May 2020 and again January to March 2021). Declines in the level of shoplifting, theft and assault have also been seen when restrictions were imposed, most likely associated with changes in mobility and disruption of interactions between offenders, targets and guardians (Halford et al. 2020). Overall the balance of the benefits and costs from committing a crime were altered due to pandemic restrictions, in most cases reducing the benefits and increasing the costs (Neanidis and Rana 2021).

Previous studies have shown that not all crime types showed the same response, with drug related offences and anti-social behaviour reported to have increased in England as a result of the pandemic (Langton et al. 2021). Anti-social behaviour comes under the

category of crimes against society, which may have included offences associated with individual's refusal to comply with lockdown measures (Neanidis and Rana 2021). This may explain some of the increased level of anti-social behaviour incidents reported. The increased level of drug offences could be explained by an increase in police presence, who are the main source of detection for such crimes (Buil-Gil et al. 2021).

As well as the introduction of restrictions, the removal of those restrictions has further impacted the level of crime away from the normal seasonal patterns. A rise in unemployment and other financial stresses resulting from the pandemic increased the incentives for the involvement in crime during the last two years (Wang et al. 2021).

When using 5 years of data for the seasonality planner, the time period from March 2020 to the present day has been disrupted by the pandemic and therefore may not have shown the expected crime levels, or seasonal patterns. Without accounting for these restrictions using explanatory variables, the seasonality indexes produced may be skewed and misleading. In order to account for the varying impacts to crime, a selection of explanatory variables related to the pandemic were selected to be included in the time series modelling. Accuracy of models containing covid explanatory variables can be compared against models without any explanatory variables. Some crime types may require different explanatory variables, depending on the effect that the pandemic had on the crime level.

A selection of covid response indicators were taken from the 'Oxford Covid-19 Government Response Tracker' dataset for the United Kingdom (Hale et al. 2021). These were a mixture of single indicators and aggregated indices and were used as explanatory variables in the time series models. They are based on the most stringent government policy in place in the country. These values were originally in the form of daily data, from 2020-01-01 to 2022-03-31, and were averaged using a mean calculation over each month to create a monthly variable to match the format of the crime time series data (Table 2).

Table 2 Covid explanatory variables used in the time series modelling (Hale et al. 2021). Variables are referred to from here on using their number, e.g. Covid1 = Stay Home.

No.	Covid variable	Description
1	STAY HOME	Single indicator. Records orders to stay and confine to the home. Values are 0 = no measures, 1 = recommend not leaving the house and 2 = requirement to not leave the house except for essential trips or activities.
2	GOVERNMENT RESPONSE	Aggregated indicator of the overall response by the government, includes all containment, closure, economic response and health system indicators.
3	STRINGENCY INDEX	Records the strictness of lockdown style closures and containment policies that primarily restrict people's behavior. Aggregated indicator.
4	HEALTH INDEX	Aggregated indicator to show how many and how forceful the measures to contain the virus and protect people's health are (combining lockdown restrictions and closures with health measures such as testing policy and contact tracing)
5	ECONOMIC SUPPORT	Aggregated indicator showing how much economic support has been made available (such as income support and debt relief).

# 4 Methodology

#### 4.1 Time Series Methods

Time series data is composed of regularly spaced observations ordered by time (or date). When data is formatted as a time series, certain methods can be used to analyse and forecast the observations. Time series forecasts are built with the assumption that the previously seen trends will continue into the forecast period (but changes in trends can be accommodated quickly). As it is already known that crime data displays seasonal patterns, *seasonal* time series methods were used where possible.

The modelling workflow is summarised in Figure 2 and explained in more detail in the following sections and the appendix. In brief: for every crime type, at a force and individual NPU level, multiple time series forecasting methods were used, and evaluated for their accuracy. The most accurate method selected during model validation was used to calculate the seasonal indices and forecast future values. Methods with and without covid explanatory variables were compared, as different crime types may have been impacted differently by covid restrictions. A more in-depth explanation of the individual methods used, and a table of the selected models is provided in Section 7.2 (Details of Time Series Methods, Appendix).

### 4.2 Model Validation and Selection

Accuracy of time series models was assessed using a training and testing data set approach. Models were initially run using a training data set (a subset of the original data) and forecast over the test data set time period (also known as the unseen, or withheld data). The trained models have no knowledge of the data in the testing period and so the forecasts can be directly compared with the 'real' data from that time period. The accuracy is measured by comparing the forecast data with the real observations, the accuracy measure used was Mean Absolute Percentage Error (MAPE).

The accuracy measure MAPE was chosen due to its easy interpretability, and ability to compare across different time series methods and data sets. The model with the smallest error on test data is likely to have the smallest errors for future forecasts, as long as there are no significant changes in the data generation processes that occur (large shocks of some kind).

Equation 1: Calculation of Mean Absolute Percentage Error (MAPE). Where n=number of observations in the time series)

$$MAPE = \frac{\sum (Observed - Forecast) / Observed}{n}$$

Table 3 Results from model validation for each crime type at a force level, MAPE: Mean Absolute Percentage Error. For full results please see appendix.

Crime Name	MAPE	Model Type
Theft Shops and Stalls	9.60	Croston
Bilking	11.12	BSTS + Covid5
Theft from Motor Vehicle	8.61	ARIMA + Covid2
Theft of Motor Vehicle	8.67	BSTS + Covid1
Residential Burglary	6.30	ARIMA + Covid3
Business and Community Burglary	7.97	ARIMA + Covid2
Personal Robbery	4.88	ARIMA
Personal Robbery (Victim U25)	9.46	Croston
Personal Robbery (Victim Female)	10.35	ARIMA + Covid1
Business Robbery	13.02	ARIMA + Covid5
Theft from the Person	19.88	BSTS + Covid1
Violence with Injury	8.27	ARIMA + Covid1
Violence with Injury (DA)	5.54	ARIMA + Covid1
Violence with Injury (Non-DA)	14.47	ARIMA + Covid3
Violence with Injury (Victim U25)	8.10	BSTS + Covid1
Violence with Injury (Victim Female)	8.74	ARIMA + Covid1
Violence without Injury	6.02	BSTS + Covid4
Violence without Injury (Victim U25)	9.04	BSTS
Violence without Injury (Victim Female)	19.61	ARIMA + Covid1
Serious Violence	22.89	BSTS + Covid1
Rape	12.31	BSTS + Covid3
Rape (Victim U25)	13.24	ARIMA + Covid1
Rape (Victim Female)	35.76	ARIMA + Covid2
Stalking (DA)	9.05	ARIMA + Covid2
Stalking (DA, Victim U25)	12.34	ETS
Stalking (DA, Victim Female)	10.82	BSTS + Covid1
Stalking (Non-DA)	18.81	ARIMA + Covid3
Stalking (Non-DA, Victim U25)	24.42	BSTS + Covid1
Stalking (Non-DA, Victim Female)	25.09	ARIMA + Covid1
Harassment (DA)	10.88	Croston
Harassment (DA, Victim U25)	11.68	ETS
Harassment (DA, Victim Female)	16.47	BSTS + Covid3
Harassment (Non-DA)	20.55	BSTS + Covid1
Harassment (Non-DA, Victim U25)	22.82	BSTS
Harassment (Non-DA, Victim Female)	21.06	ETS
Public Fear, Alarm or Distress	15.68	ARIMA
Hate Crime	12.04	ETS
Hate Crime (Victim U25)	18.49	ARIMA + Covid1
Hate Crime (Victim Female)	21.94	ARIMA + Covid1
Child Abuse	8.25	ETS
Domestic Abuse	7.93	BSTS
Online Crime	18.86	ARIMA + Covid1
Anti-Social Behaviour (Incidents)	10.05	ARIMA

# 4.3 Forecasting

The most accurate model for each crime type was selected and re-run using all available data (5 years: April 2017 to March 2022) to produce a forecast for a future period of 12 months (April 2022 to March 2023). The forecast contains a point forecast of the predicted values, along with prediction intervals for the future time period.

When calculating seasonality values from the model outputs, the values are given in the form of crime counts, above or below the trend. In order to make these seasonality values comparable across different NPUs or crime types, they are converted into a percentage by dividing them by the mean of past volumes<sup>1</sup>. A seasonal index shows the degree of variation, comparing increases and decreases to the average of the season.

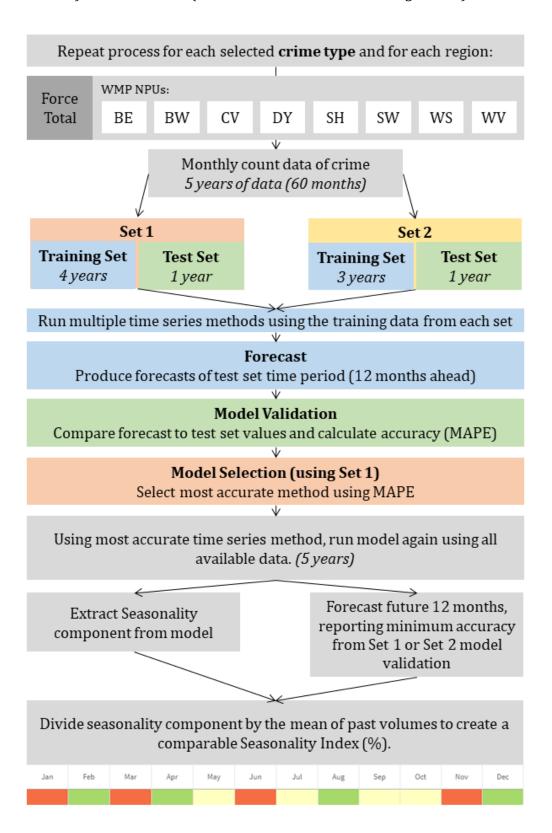
The raw seasonality values are also output; these are the actual number of crimes above or below (if negative) the trend for the given crime type. It should be made clear that the raw values are not the total predicted value for a crime type in a given month, but the value above or below the trend that has occurred.

The forecasted values in this project are based upon the previous data and the covid explanatory variables (where they are selected) and do not consider any other explanatory variables. They should only be used in combination with expert knowledge when used for decision making. The main output of this project is the seasonality component rather than the forecasted values. The forecast error will be displayed alongside the forecast values in order to give an indication of the confidence of the forecasts. In cases where the forecast error is high, there may be less data available to produce a reliable forecast, or the selected crime does not show regular or seasonal patterns. Forecast error may also be higher where there has been a level shift in the number of crimes due to system changes (e.g. implementation of Connect or ControlWorks) or changes in crime recording practices or counting rules.

\_

 $<sup>^{1}</sup>$  Here the mean is adjusted by the proportionate contribution of a year's contribution to the whole time series in order to mitigate non-stationarity.

Figure 2: Modelling Workflow of the process used, details of individual model types can be found in Table 5. (MAPE: Mean Absolute Percentage Error)



# 5 Project Output

The end output was a Business Insight (Qlik) dashboard displaying the crime seasonality planner. Dashboards are already used within the force and their interactive capabilities and easy accessibility make them useful. The dashboard has the advantage of being more regularly updated if required. The layout and content of the dashboard was developed with consultation from a number of experienced analysts, who will be end users. Observing their interaction with the dashboard and their questions about the underlying methodology has helped to refine the information on the 'Welcome Page', and the notation provided on 'pop-ups', to assist with interpretation. A briefing at Strategic Tasking also highlighted the queries senior leaders will have when using the dashboard, and changes implemented to cater for these.

# 5.1 Welcome Page

The first sheet of the Crime Seasonality Planner dashboard is a welcome page containing an overview of the dashboard and important information about its contents and usage.

# 5.2 Seasonality Profile Overview

The next sheet of the dashboard was designed to match the existing seasonality planner output. It is formatted to show the seasonality (by month) for each crime type, with the option to select the required NPU. The default selection is to display 'Force' level data. All crimes are initially shown, but the user is able to select a subset of the crime types for comparison purposes. This sheet is restricted so that only one NPU can be selected at a time.

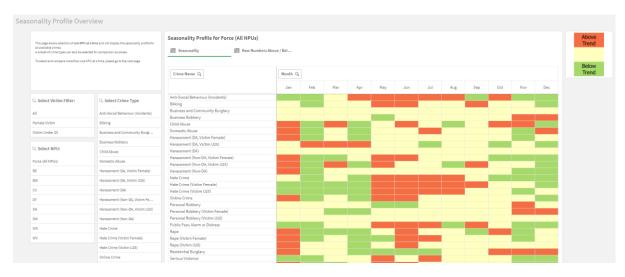


Figure 3 Seasonality Profile Overview Sheet: screenshot from Qlik App.

The seasonality of each crime per month is shown with a coloured scale. The scale has been designed to match the existing seasonality planner for continuity. With High where values are greater than 5%, medium where values are between 5% and - 5% and low being where vales are less than -5%. There is a second tab in the graph container showing the seasonality values as raw values instead of as coloured blocks. The raw crime values are the number of crimes on average above or below the trend for that crime type, compared to other months. The raw values are colour coded using the same measure of

seasonality as the plain coloured blocks, for continuity. It should be noted that each crime type has raw values of different magnitudes depending on what the crime type is.

# **5.3 Seasonality Profile Comparison**

The second sheet of the dashboard has the same layout as the previous sheet, but allows the selection of multiple NPUs at a time for comparison purposes. When no selection is made, data for all crimes and all NPU options are shown. Using this sheet, the seasonality of multiple crimes across multiple NPUs can be compared.

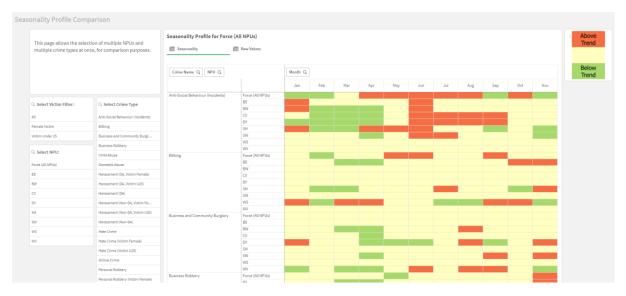


Figure 4 Seasonality Profile Comparison Sheet: screenshot from Qlik app. This sheet allows selection of multiple NPUs and Crime Types at one time.

#### 5.4 Time Series

The third page of the dashboard requires the selection of only one crime and NPU at a time. It displays the seasonality profile, again either as plain coloured blocks or raw crime values. Below this the time series data for the selected crime type is shown. With 5 years of 'real' data shown plus a 12-month forecast of this crime type. This time series is shown to supplement the seasonality values, to show the trends over time. As the raw seasonality values are the values above or below the trend, the time series gives a more complete pictures of the number of crimes occurring for each selection.

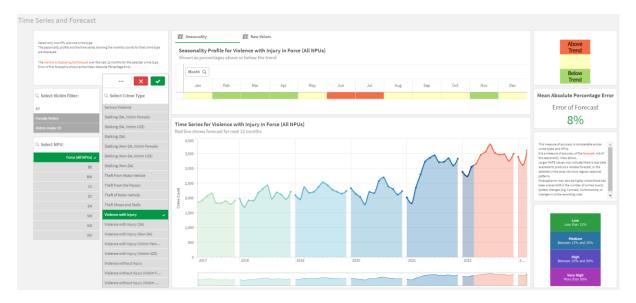


Figure 5 Seasonality Time Series and Forecast Sheet: screenshot from Qlik app. This sheet allows selection of one NPU and one crime type at a time.

The error of the forecasted value is also displayed. This is shown as the Mean Absolute Percentage Error and is the same measure of error that the modelling methods were evaluated on. The higher the value of MAPE, the less confidence there is in the forecasts. High values (above 25%) may be caused by a lack of data for the selected crime, if it is a less frequent occurrence. Time series methods are more accurate when they have more data points available. Higher error may also be caused by crimes that do not show a strong and regular seasonal pattern. As seasonal time series methods were used, there is a focus on seasonality. For crimes that are retrospectively reported such as Rape, Stalking, Harassment, etc the 'incident from' date may not be truly accurate. By investigating the data for certain crime types in more depth it was clear that the 1st January each year, as well as the 1st of each month shows peaks in the number of these crimes. This is due to retrospective reporting when exact dates are not known. Due to this the data may be skewed and not reflect the true seasonality of the crime occurring. An additional box providing information to support interpretation of the seasonality profile was provided where required.

# **6 Future Updates**

The seasonality planner forecasts are planned to be updated every month going forward. However, for continuity the seasonality grid element will be updated less frequently, suggested every 6-12 months. By updating the forecasts monthly, the original forecasts can be compared with the subsequent 'real' data and the accuracy compared. This will allow further evaluation of the methods selected, and if necessary, adjustment of the workflow. As time moves on, still only 5 years of data will be used, so the data going into the models will move in a sliding window.

# 7 Appendix

# 7.1 Crime Types

### 7.1.1 Serious Violence (VWI, Knife Used, Injury Caused)

The definition of serious violence in this project is the same as the Project Guardian Violence with Injury (VWI) Knife Used Application. This is different from other definitions of 'serious violence' or 'most serious violence' that have been previously used for other projects.

Due to the changeover to the new WMP Connect System in April 2021, the way offences involving weapons are recorded on the system has changed. Despite legacy data being transferred to the new connect system, this transferred data did not contain the same level of detail as data in the old crimes system. Therefore, in order to have a complete time series containing 5 years of data for this crime type, data from the old Crimes system was combined with data from the new Connect system. For incidents that occurred between 2017-04-01 and 2021-04-07, the old crimes system was used. For any incidents between 2021-04-08 and 2022-03-31, the new system was used. This was considered the best way to ensure the most accurate counts were made, without the risk of missing any relevant incidents.

Filters for this crime type were matched as closely as possible between the old 'crimes' system data and the new 'Connect' system data to ensure a seamless time series dataset was created. These two sets of data were combined to create a new table with the crimes data and connect data, this combined data set was used for all further analysis. The following are the main filtering rules applied:

 The specific offences were matched using home office main codes (homc) and home office offence codes (hooc). These codes were extracted from the Connect system for offences of Violence Against the Person > Violence With Injury. These codes are contained in

- Table 7 (Section 6. Appendix).
- Any offences flagged as 'Domestic Abuse' or 'Domestic Violence' were excluded.
- Only incidents where a weapon was 'used causing injury' were included Weapon type filtered to those contained in
- Table 6 (Section 6. Appendix).

### **7.1.2** Rape

This category includes all crimes where the main group is SEXUAL OFFENCES and the sub group is RAPE.

Due to the nature of this crime type, crimes are often reported retrospectively. This means that the 'incident from' date is not always accurate, with a large number of crimes having this date listed as the 1st of January or the 1st day of any month. This is assumed to be when the crime is historical and reported at a later date, with the exact dates not being recorded, but instead a wider time period given, or when the crime has occurred over a longer period of time. Using this date information, the seasonality planner will be skewed to show high levels of seasonality with larger numbers of these offences occurring in January. This is not a true reflection of the crime occurrences. This supports the idea that Rape and Sexual Offences are more difficult to calculate seasonality with, due to the longer time gap between occurrence and reporting than other offences. As previously mentioned, additional information boxes were presented on the Qlik app to support interpretation where required.

### 7.1.3 Stalking and Harassment

Both Stalking and Harassment have been split into two categories – crimes related to domestic abuse, and crimes unrelated to domestic abuse. The split was made using keywords list and crime flags. To filter out only domestic abuse related crimes these two columns were searched through using the string 'DOMESTIC ABUSE' and any crimes that contained this as a keyword or crime flag were included.

From the 1st April 2018 the Home Office changed the recording practices for Harassment and Stalking offences in order to better understand these cases, in particular where harassment escalates to stalking. The new rules mean that stalking / harassment should be recorded in addition to the most serious additional victim-based offence involving the same victim-offender relationship. The old rules meant that the substantive offence was recorded instead of harassment.

From the 1st April 2020 the crime recording rules were updated again. This update meant that every domestic abuse (DA) related offence of harassment should instead be recorded as stalking (Sec. 2), unless the Force has a clear rationale for the lesser offence. Looking at the counts over time for Harassment and Stalking offences it is clear that these changes in recording practices have had an impact on the recording of these crimes. Filtered only to domestic abuse cases, there is a drop in the number of harassment cases recorded from Jan 2020, and a concurrent increase in stalking cases, reflecting the updated rules.

These changes in recording practices are likely to have had an effect on the number of offences recorded. Stalking and Harassment also encounter a January skew due to retrospective reporting and these offences taking place over a longer period of time.

#### 7.1.4 Other Crimes

Some crime types do not follow the home office classification codes, and instead required filtered by 'keywords' or 'crime flags' available on the crime records. For Child Abuse and Domestic Abuse, the 'offence type' category was also included as a filter. This category only contains values for records created pre-Connect. The crime flag and keywords categories were introduced with Connect and so inclusion of 'offence type' ensures all possible records for the selected crime types are included in the crime counts.

Table 4 Filtering rules for Hate Crime, Child Abuse, Domestic Abuse and Online Crime.

CRIME TYPE	KEYWORD LIST	KEYWORDS	CRIME FLAG	OFFENCE TYPE
HATE CRIME			'НАТЕ'	
CHILD ABUSE	'CHILD SEXUAL ABUSE'	CHILDABUSE		CA
DOMESTIC ABUSE		DOMESTIC ABUSE	DOMESTIC ABUSE	DV
ONLINE CRIME	ONLINE CRIME		ONLINE CRIME	

Child Abuse and Domestic Abuse crime types excluded any crimes where the home office main group code was 'NON-CRIME'. A non-crime is not a substantive offence and merely a record, therefore they are not included in the crime counts for this seasonality planner.

### 7.1.5 Anti-Social Behaviour (incidents)

This is the only 'crime type' that does not use data from the Connect system. Instead data from the ControlWorks system is used, this contains reports of incidents, rather than crimes. Anti-social behaviour is recorded in the Connect system (NON-CRIME>ASB) however incidents are only given an ASB non-crime number if case management is required. Therefore, to gain a clearer picture of the demand of all ASB incidents, data is sourced from the Controlworks (and Oasis – for pre-March 2020) database.

The ControlWorks system was introduced on 3<sup>rd</sup> March 2020, the previous system was called Oasis. To be able to have 5 years of data for this crime type data from the two systems was joined together. Oasis was used for incidents between 2017-04-01 and 2020-03-03, and ControlWorks data was used for incidents between 2020-03-04 and 2022-03-31.

Anti-social behaviour incidents were classified slightly differently in the two systems. In Oasis anti-social behavior incidents classified as one of the following were included in the counts for this project:

- ANTI-SOCIAL BEHAVIOUR NUISANCE
- ANTI-SOCIAL BEHAVIOUR ENVIRONMENTAL
- ANTI-SOCIAL BEHAVIOUR PERSONAL
- DISORDER

The below categories were included from data in the new Controlworks system:

- ASB NUISANCES
- ASB ENVIRONMENT
- ASB PERSONAL
- ASB DISORDER

The Anti-Social Behaviour, Crime and Policing Act 2014 was revised in January 2021, this change should also be considered when viewing the time series of the data for this crime type.

### 7.2 Details of Time Series Methods

#### 7.2.1 Model Validation and Selection

Models were selected using model accuracy from a training and testing data set. Two different training and testing set splits were used for model validation. The first training data set was made up of 4 years of data (April 2017 to March 2021) and a testing data set was made up of 1 year of data (April 2021 to March 2022). This is an 80:20 split between training and testing data. This data set was used to select the most accurate model for forecasting.

A second training and testing set was used, composed of a training set with 3 years of data (April 2017 to March 2020) and a testing set with 1 year of data (April 2020 to March 2021). This second training/testing set was also run on the data and the MAPE compared to the first training and testing set outputs. Due to the introduction of new systems such as Connect and Controlworks, the accuracy of the first training/testing split may have been skewed. For this reason, in some cases the accuracy of the second training/testing split was displayed instead of the first training/testing set. This reduces the impacts of the new systems on the accuracy measures where there were severe level differences between crime levels in 2020 and 2021.

### 7.2.2 Classical Decomposition

The simplest method for analysing time series data is classical decomposition. This splits the observations out into three components: trend, irregular and error. This method can be in additive or multiplicative format:

(A) 
$$y_t = \mu_t + \gamma_t + \varepsilon_t$$
,  $t = 1, ..., n$   
(B)  $y_t = \mu_t * \gamma_t * \varepsilon_t$ 

Equation 2 Additive (A) and multiplicative (B) time series decomposition

This method is similar to the method that is currently used in the existing Seasonality Planner. The trend component is calculated by a moving average, usually centered with 12 months of data (when using monthly data). The irregular component is the original

data minus the trend component. The irregular component of classical decomposition can be used to calculate a seasonality index for each month (or time period) of time series data. When this method is used on its own it cannot consider any additional explanatory variables. Classical decomposition can be used in conjunction with other methods to calculate the seasonality components of the fitted values from the more complex modelling methods.

### 7.2.3 Simple Exponential Smoothing

Simple Exponential Smoothing (SES) uses the weighted averages of past observations in order to provide a forecast, with more weight given to the most recent observations. No explanatory variables can be incorporated into this method, meaning that the effects of the covid-19 pandemic on crime counts cannot be accounted for. This method has been included to compare with the more complex methods that include explanatory variables, to see what effect the pandemic has had on the data and whether it is always necessary to include the extra variables. For some crime types that have a stable time series with regular seasonal patterns, a simpler method may be just as, or more, accurate than more complex methods.

#### 7.2.4 Croston's Method

Croston's method is included as in some cases the crime data includes gaps where there are zero occurrences, and other time series methods do not forecast well with these intermittent gaps. Croston's method handles this by splitting the time series into two, forecasting the demand level where it exists, alongside forecasting the average time in between the demand occurrences. This method is carried out in the same way as SES and ARIMA, with two training and testing sets. The fitted values of this model are used to create the seasonality index where this is selected as the most accurate model.

### **7.2.5 ARIMA**

ARIMA (Auto Regressive Integrated Moving Average) is a more complex method of time series forecasting. It uses autoregressive (AR) and moving average (MA) terms to predict the future values, based on past information. An ARIMA contains three values that can be specified (p, d and q). Where an ARIMA model (p, d, q) will contain the AR term, p, MA term, q, and number of differences, d. For seasonal arima models, three additional terms (uppercase P, D and Q) are included for the seasonal part of the model, it becomes an ARIMA (p, d, q) (P, D, Q) $_{\rm m}$ . Where m is the number of observations per year (e.g. 12 when using monthly data).

An automated step-wise method identifies the best model fit by minimizing the Akaike Information Criterion (AIC). This considers the differencing requirements of the ARIMA model, where stationary data is required and foregoes the need to alter data before inputting it into the model. Data is still differenced when required, but this does not need to be done manually. Differencing the data is done to make the data stationary, where the mean, standard deviation and autocorrelation are constant over time.

ARIMA modelling also allows the inclusion of explanatory variables. The covid variables described in Section 3.2 are each included in an ARIMA model, and their accuracy compared to each other, an ARIMA model with no explanatory variables and the other modelling methods.

To calculate the seasonality from the ARIMA models, the fitted values from the model were decomposed and averaged over each month. Using this method ensured that seasonality captured the covid explanatory variables as well as the time series data. As the fitted values are one-step-ahead forecasts, the seasonality values were appropriately adjusted to ensure the correct months were specified in the seasonality index.

### 7.2.6 Bayesian Structural Time Series

The final time series method used in this project is Bayesian Structural Time Series. Structural time series methods examine the separate components of the data – trend, seasonal, and the random component. The trend component shows any slowly changing parts of the series over time, in contrast to the seasonal component which captures periodic elements of the time series. They are considered a more flexible method of prediction than traditional or ARIMA methods. Structural time series methods are based on the construction of a state space model, which assumes the changes in the time series observations  $(y_1, ..., y_n)$  are determined by an unobserved series of vectors  $(\alpha_1, ..., \alpha_n)$  and specifies this relationship.

This uses the Markov Chain Monte Carlo (MCMC) method to sample from the posterior distribution of a Bayesian structural time series model. It is composed of a regression formula and a state specification containing the required components of state (trend and seasonal components). The number of iterations relates to the number of MCMC draws from the posterior distribution. This method also utilises the Kalman filter to adjust forecasts.

When this method is used without any explanatory variables it is an ordinary state space time series model. When explanatory variables are added, a spike-and-slab prior is used for the static regression component of the model. The spike-and-slab component and the Kalman filter require observations and state variables to be Gaussian. An advantage of bsts models compared with ARIMA models is that they do not have a requirement for stationary data, reducing the risk of losing valuable information during the differencing process.

There are different types of trend component available that can be used within a bsts model. The simplest is the local level model, which just captures the level of the series. In this project a semi-local linear trend component was chosen which is a more complex way of describing the trend. It includes the level as well as a slope, which gives more information to the model about how fast or slow the level changes. Semi local linear trend model is reported as the most useful for long-term forecasting (Scott 2021). It assumes the level component moves according to a random walk model, but the slop component moves according to an auto-regressive (1) process centered on a potentially non-zero value. There are four independent components for the prior distribution of a semi-local linear trend, and the prior distributions of each of these parameters can be specified if required. It was however found that specifying the prior distributions did not improve the accuracy any further than using the default priors selected by the modelling package.

The seasonal component of the model can be thought of as a regression on 12 dummy variables with coefficients constrained to sum to 1. Seasonal state component for monthly data represents the contribution of each month to the annual cycle.

$$Y_t = u_t + \theta_t + \varepsilon_t$$
  $\varepsilon_t \sim N(0, \sigma_{\epsilon}^2)$ 

$$\theta_{t+1} = -\sum_{s=0}^{s-2} \theta_{t-s} + \gamma_{\theta,t}$$

Where S is the number of seasons and  $\gamma_t$  indicates the joint contribution to the observed response,  $Y_t$ .

There is only one prior distribution required for the seasonal component of the state space model, the seasonal sigma prior. The sigma value for seasonal data should be lower to capture the behaviour of the seasons. To compare the effects of estimating the priors, a default model with no prior estimation was included in the modelling run. Priors can be estimated using a training and testing data set, using a cross-validation process.

The covid variables were used in this BSTS modelling in the same way they were used in the ARIMA modelling. One model was created for each covid variable, with their accuracies compared to each other, a model with no explanatory variables and the other modelling methods.

The seasonality of the bsts models was calculated by averaging the state contributions of the seasonal component of each model over each month. This outputs a value for each month in the form of raw values on the same scale as the crime counts. The seasonal values were divided by the mean of the past volumes (adjusted to mitigate non-stationarity) to give a comparable seasonal index in percentage form.

*Table 5: List of model types and inclusion of covid explanatory variables.* 

Model Type	Covid Variable	Description
ETS	-	No covid variables as unable to add explanatory variables to this method
ARIMA	-	No covid variables – for comparison
ARIMA	Stringency	
ARIMA	Government Response	
ARIMA	Health Index	
ARIMA	Economic Support	
ARIMA	Stay at Home	
BSTS	-	No covid variables – for comparison
BSTS	Stringency	
BSTS	Government Response	
BSTS	Health Index	
BSTS	Economic Support	
BSTS	Stay at Home	
CROSTON	-	Method used for count data, can improve accuracy where there are many 0 values.

Table 6 Weapon Types for 'Serious Violence' (Crime 12) that were included in the query, comparison of crime system and connect system definitions.

Weapon Name	-	Vame	Weapon Name	Weapon Name
Crimes System	Code Crimes Syste	***	Connect System	Code
SHARP INSTRUMENTS	02	111		Connect System
KNIFE – DAGGER	09		-	-
KNIFE - FLICK	10		-	-
KNIFE - KITCHEN	10		-	-
KNIFE - KITCHEN KNIFE - MACHETE	12		-	-
			-	-
KNIFE - PEN	13		-	-
KNIFE – CRAFT	14		-	-
RAZOR	19		- -	-
CROSS BOW	30		Cross Bow	WEAPON TYPE 30
KNIFE (N/K)	33		-	-
SCREWDRIVER	41		Screwdriver	WEAPON TYPE 41
SHARP INSTRUMENT	69		Sharp Instrument	WEAPON TYPE 69
DAGGER	70		Dagger	WEAPON TYPE 70
FLICK KNIFE	71		Flick Knife	WEAPON TYPE 71
KITCHEN KNIFE	72		Kitchen Knife	WEAPON TYPE 72
MACHETE	73		Machete	WEAPON TYPE 73
PEN KNIFE	74		Pen Knife	WEAPON TYPE 74
CRAFT KNIFE	75		Craft Knife	WEAPON TYPE 75
LOCK KNIFE	76		Lock Knife	WEAPON TYPE 76
KNIFE - UNKNOWN	77		Knife - Unknown	WEAPON TYPE 77
RAZOR/BLADE	78		Razor Blade	WEAPON TYPE 78
SWORD	83		Sword	<b>WEAPON TYPE 83</b>
AXE	84		Axe	WEAPON TYPE 84
SCISSORS	85		Scissors	<b>WEAPON TYPE 85</b>
MEAT CLEAVER	87		Meat Cleaver	WEAPON TYPE 87
BROKEN BOTTLE OR	90		-	-
BROKEN GLASS				
ZOMBIE KNIFE	95		Zombie Knife	WEAPON TYPE 95
CYCLONE KNIFE	96		Cyclone Knife	WEAPON TYPE 96
COMBAT KNIFE	97		Combat Knife	WEAPON TYPE 97

Table 7 Home office classification codes for 'serious violence' offences

Crimes Classification Code	Home Office Main Code	Home Office Classification Code	Full Offence Title
2	002	00	Attempted murder
5D	005	06	Assault with Intent to cause Serious Harm - Causing bodily injury by explosion
5D	005	01	Attempted - s.18 - Assault with Intent to cause Serious Harm - Wounding with intent to do grievous bodily harm
5D	005	01	s.18 - Assault with Intent to cause Serious Harm - Wounding with intent to do grievous bodily harm
8N	800	02	Assault with Injury - Administering poison with intent to injure or annoy
8N	800	01	Assault with Injury - s.20 - Malicious wounding: wounding or inflicting grievous bodily harm
8N	800	06	Assault with Injury - s.47 - Assault occasioning actual bodily harm
8N	008	21	Owner or person in charge allowing dog to be dangerously out of control in any place in England or Wales (whether or not a public place) injuring any person or assistance dog
8P	8P 008 60	60	Racially or religiously aggravated assault or assault occasioning actual bodily harm
8P	800	59	Attempted - Racially or religiously aggravated wounding or grievous bodily harm
8P	800	59	Racially or religiously aggravated wounding or grievous bodily harm
8S	005	01	Assault Police - s.18 - Wounding with intent to resist/prevent arrest
8S	800	01	Assault Police - Assault with Injury - s.20 - Malicious wounding: wounding or inflicting grievous bodily harm
8T	800	06	Assault Emergency Worker - Assault with Injury - s.47 - Assault occasioning actual bodily harm
8T	005	01	Assault Emergency Worker - s.18 - Assault with Intent to cause Serious Harm - Wounding with intent to do grievous bodily harm

Table 8 Home Office Main Group / Sub Group and Sub-Sub Groupings for Crime Types included in the seasonality planner.

CATEGORY	CRIME NAME	MAIN GROUP	SUB GROUP	SUB SUB GROUP
THEFT	Theft Shops and Stalls	THEFT	SHOPLIFTING	SHOPLIFTING
THEFT	Bilking	THEFT	OTHER THEFT	MAKING OFF WITHOUT PAYMENT
NEIGHBOURHOOD CRIMES	Theft from Motor Vehicle	VEHICLE OFFENCES	THEFT FROM A VEHICLE	THEFT FROM A MOTOR VEHICLE
NEIGHBOURHOOD CRIMES	Theft of Motor Vehicle	VEHICLE OFFENCES	THEFT OF UNAUTH TAKING OF A MOTOR VEH	THEFT OF UNAUTH TAKING OF A MOTOR VEH
NEIGHBOURHOOD CRIMES	Residential Burglary	BURGLARY	BURGLARY - RESIDENTIAL	BURGLARY - RESIDENTIAL
NEIGHBOURHOOD CRIMES	Business and Community Burglary	BURGLARY	BURGLARY - BUSINESS AND COMMUNITY	BURGLARY - BUSINESS AND COMMUNITY
NEIGHBOURHOOD CRIMES	Personal Robbery	ROBBERY	ROBBERY OF PERSONAL PROPERTY	ROBBERY OF PERSONAL PROPERTY
NEIGHBOURHOOD CRIMES	Business Robbery	ROBBERY	ROBBERY OF BUSINESS PROPERTY	ROBBERY OF BUSINESS PROPERTY
NEIGHBOURHOOD CRIMES	Theft from the Person	THEFT	THEFT FROM THE PERSON	THEFT FROM THE PERSON
VIOLENCE	Violence with Injury	VIOLENCE AGAINST THE PERSON	VIOLENCE WITH INJURY	
VIOLENCE	Violence without Injury	VIOLENCE AGAINST THE PERSON	VIOLENCE WITHOUT INJURY	
VIOLENCE	Serious Violence	VIOLENCE AGAINST THE PERSON	VIOLENCE WITH INJURY	
VIOLENCE	Rape	SEXUAL OFFENCES	RAPE	
VIOLENCE	Stalking (DA)	VIOLENCE AGAINST THE PERSON	STALKING AND HARASSMENT	STALKING
VIOLENCE	Stalking (Non-DA)	VIOLENCE AGAINST THE PERSON	STALKING AND HARASSMENT	STALKING
VIOLENCE	Harassment (DA)	VIOLENCE AGAINST THE PERSON	STALKING AND HARASSMENT	HARASSMENT
VIOLENCE	Harassment (Non-DA)	VIOLENCE AGAINST THE PERSON	STALKING AND HARASSMENT	HARASSMENT
OTHER	Public Fear, Alarm or Distress	PUBLIC ORDER OFFENCES	PUBLIC FEAR, ALARM OR DISTRESS	PUBLIC FEAR ALARM OR DISTRESS

Table 9 List of acronyms used in this report.

ARIMA	Auto-Regressive Integrated Moving Average (time series method)
ASB	Anti-Social Behaviour
BSTS	Bayesian Structural Time Series (time series method)
DA	Domestic Abuse
DAL	Data Analytics Lab
ETS	Exponential Smoothing
FTDB	Force Tasking Delivery Board
HOMC	Home Office Main Code
HOOC	Home Office Offence Code
MCMC	Markov Chain Monte Carlo
NPU	Neighbourhood Policing Unit
STCG	Strategic Tasking and Coordination Group
VWI	Violence with Injury
WMP	West Midlands Police

#### 8 References

- Buil-Gil, D., Zeng, Y. and Kemp, S. 2021. Offline crime bounces back to pre-COVID levels, cyber stays high: interrupted time-series analysis in Northern Ireland. *Crime science*. https://doi.org/10.1186/s40163-021-00162-9
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S. and Tatlow, H. 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour. https://doi.org/10.1038/s41562-021-01079-8
- Halford, E., Dixon, A., Farrell, G., Malleson, N. and Tilley, N. 2020. Crime and coronavirus: social distancing, lockdown, and the mobility elasticity of crime. *Crime Science* **9**, 11. https://doi.org/10.1186/s40163-020-00121-w
- Kirchmaier, T. and Villa-Llera, C. 2020. Covid-19 and changing crime trends in England and Wales. Centre for Economic Performance, (013).
- Langton, S., Dixon, A. and Farrell, G. 2021. Six months in: pandemic crime trends in England and Wales. *Crime science*, 10(1), pp.1-16.
- Neanidis, K.C. and Rana, M.P., 2021. Crime in the era of COVID-19: Evidence from England.
- Scott, S. 2021. Bayesian Structural Time Series. CRAN. https://cran.r-project.org/web/packages/bsts/bsts.pdf
- Wang, J., Fung, T. and Weatherburn, D. 2021. The impact of the COVID-19, social distancing, and movement restrictions on crime in NSW, Australia. *Crime science*, 10:24.