

Data-Driven Method for Identification of Top Offenders

Criminal justice research indicates that a small percentage of offenders commit a disproportionate amount of crime, often presented as X% of offenders responsible for Y% of crime. As a result, various policing strategies—predominately intelligence-led policing—recommend a focus on the key percentage of top, repeat, or high-profile offenders. In operationalizing this concept, law enforcement organizations have developed various methods of identifying top offenders. The use of simple metrics, such as number of convictions or arrests, is one of the most basic methods. However, through crime and intelligence analysis, identification of top or repeat offenders can be more robust. The following describes one sophisticated, data-driven mechanism for determining offenders.

Through the use of crime analysts with powerful data systems, law enforcement organizations have developed various methodologies for identification of top offenders. A method grown in popularity through the crime analysis profession was initially introduced by Sonoma County, California, analyst Matt Harris in 2010. Mr. Harris’s method weighs each person’s involvement in each incident with three variables: the seriousness of the offense, the role the individual played, and the recency. Each individual’s values for all incidents are then tallied and a final list produced. Sonoma County operationalized the list for a variety of enforcement purposes, including warrant and parole sweeps.

Top Offender Search Wildcard - *

Data Updated: **9/14/2010**

LAST	FIRST	MIDDLE

CITY	Res - ESZ	PROBATION	PAROLE
ALAMEDA	1010100000	Felony Formal	ACTIVE
ANDERSON	1010200000	Felony Informal	DISCHARGED
AVONDALE	1010300000	Misd Formal	PRE-PAROLE
BEL	1010400000	Misd Informal	REVOCATION

AGENCY	OFFENSE TYPE	Off - Beat
COP	ARSON	
JCP	ASSAULT - AGGRAVATED	
PEP	ASSAULT - INTIMIDATION	10101
RPP	ASSAULT - SIMPLE	10102
SCSD	BAD CHECKS	10103

HEIGHT	WEIGHT	AGE	RACE	SEX	HAIR	EYE

Other Reports Summary Report Data Report Data Report w/INV Clear Form

Min Points	150
Gang	<input type="checkbox"/>
Sex	<input type="checkbox"/>
Drug	<input type="checkbox"/>
Violent	<input type="checkbox"/>

Figure 1: Sonoma County, California, crime analyst Matt Harris is a pioneer in this area. In the mid-2000s, he created a database to allow quick ranking and searching of top offenders using a formula of his own design.

The Cambridge, Massachusetts, Police Department adopted Harris's base model for a ranking and prioritization scheme associated with a Smart Policing Initiative (SPI) grant in 2012. Harris's three variables remained at the core of the model, although additional weights were afforded for any gang, gun, or drug involvement in the past crimes. The resulting list of offenders was used not for enforcement but for focused deterrence, consisting of "intervention" meetings with high-ranked offenders that both threatened them with future police attention and prosecution and introduced them to social service providers who could assist with their core issues (Smart Policing, 2014).

Foundation Technology to Support a Data-Driven Method for Identification of Top Offenders

Identification of top offenders using a data-driven approach can be accomplished using various software platforms. Microsoft Access is typically best suited for smaller to mid-sized agencies based on data volume in computer aided dispatch (CAD) and records management systems (RMS) and limited expertise in programming. Larger-sized law enforcement agencies with more data and multinetwork user capability will likely find Microsoft SQL or Oracle better suited for performance.

For this project, Microsoft Access was chosen for its ability to (a) link to remote data, (b) store data internally, (c) answer complex law enforcement questions based on multiple datasets, and (d) offer tools to build forms and reports to manage and report on the database top offender records. Microsoft Access also has the advantage of being easily portable to other agencies because of widespread access to and use of Microsoft Office licensing.

Identification of top offenders using a data-driven approach typically begins with use and integration of RMS data. A best practice to begin this approach is to copy the RMS data into a new Microsoft Access database and implement cleaning and reformatting routines as described elsewhere in the Data section of the Crime Analysis Toolkit. *Note: Many of the queries needed for this process are too complex to utilize production or "live" data from the RMS system directly.* Upon finalizing the dataset, it is important to ensure that the database is secured for use only by appropriate personnel.

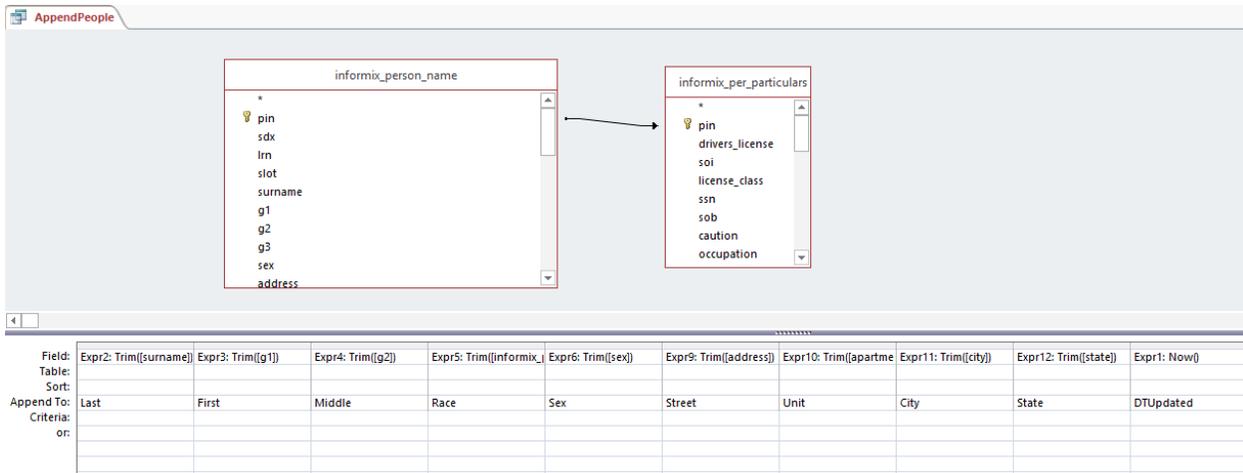


Figure 2: A series of “append” queries in Microsoft Access copy data from the police RMS into the secure offender database.

Development of a Data-Driven Algorithm for Identification of Top Offenders

The Harris formula of identifying top offenders considers past known criminal involvements and scores them based on the seriousness of the crime, the role the individual played, and the recency. The basic calculation for each incident is:

$$\text{MAX (Offense Weight * Role Weight – Date Decay Total,0)}$$

The table below shows several examples of the formula in action, assuming a current date of September 15, 2016:

Offense	Offense Weight	Role	Role Weight	Date	Days Since	Decay Constant	Decay Total	Total
Robbery	8	Arrest	1	08/31/2016	15	0.002	0.03	7.97
Auto theft	6	Suspect	0.9	09/15/2015	365	0.002	0.73	4.67
Shoplifting	4	Arrest	1	09/15/2010	2,190	0.002	4.38	0

Although there are default values for the offense weight, role weight, and date decay constant, these values can be edited by the priority of the offenses for the department’s goals and objectives in identification. Further, values for the recency of crimes can be edited to identify more recently active offenders.

The *offense weight* is assigned on a scale of 1 to 10 where 10 represents the most serious offense (murder) and 1 represents the least serious (miscellaneous traffic offenses, generally). The tables below indicate the default scores for various crimes in the database. Offense weights can be expanded to a scale of 1 to 100 or 1 to 1,000 if you are using a statute table containing all statutes.

Offense	Default Weight
Murder	10
Sexual assault	9
Armed robbery	8
Burglary	7
Auto theft	6
Counterfeiting	5
Vehicle burglary	4
Drug possession	3
Disorderly conduct	2
Traffic offenses	1

The weights should generally fall in line with UCR/IBR hierarchies but, more important, be based on the department’s priorities. Although it is not possible to eliminate all subjectivity in the weighting process, Ratcliffe (2014) has recommended tying such offense weights to sentencing guidelines, a sentiment echoed by Larry Sherman (2013, p. 423) as “justified on good democratic grounds as reflecting the will of the people.” It is important to remember that the level of subjectivity in the weighting system improves on systems that determine top offenders by weighting all offenses equally, by total crimes or arrests and/or anecdotally.

In the development of weights, it is important to consider differences in the quality of two offenses in the same category. For example, an armed robbery in which a victim is shot and the offender steals \$2 million is more severe than one in which the offender simply brandishes a knife and steals a mobile phone. In the example project, weighting additional factors is deemed too complex in the identification of top offenders.

It is common for law enforcement incidents to contain more than one offense. For example, a suspect may break into a home, steal keys, and subsequently steal a car from the garage. Offenses of breaking and entering as well as theft of an automobile would be documented on the same report. In the identification of top offenders, the weighting counts only the most serious offense in each incident. In a system that complies with IBR standards and that records multiple offenses only when they are independent of each other (i.e., an assault and a theft are assumed to be part of a robbery), all involved offenses should be included in the weighting. However, it is important to maintain consistency in weighting the most serious offense or weighting all offenses.

Step 1: Set Weights for Offenses and Roles

Offense Weights

Default Exclude Domestic

Offense	Weight	ucr	
MURDER	10	0901	C
DRIVE BY SHOOTING	9	1314	1
HOMICIDE-NEGLIGENT	9	0910	C
MANSLAUGHTER	9	0910	1
SEXUAL ASSAULT	9	1199	C
ARSON-STRUCTURE	8	2099	1
KIDNAPPING	8	1099	C
ROBBERY-ARMED	8	1299	C
ROBBERY-STRONG ARM	8	1206	C
TERRORISM	8	1602	C
AGGRAVATED ASSAULT	7	1315	C
AGGRAVATED ASSAULT DV	7	1315	1
AGGRAVATED ASSAULT-POLICE C	7	1311	C
ARSON-MOBILE	7	2099	C
ARSON-OTHER PROPERTY	7	2099	2
BURGLARY FORCE	7	2202	C
BURGLARY NO FORCE	7	2204	C
HUMAN TRAFFICKING SERVITUD	7	1099	2
HUMAN TRAFFICKING SEX ACTS	7	1199	1
UNLAWFUL IMPRISONMENT	7	1099	1

Role Weights

Off. Dflt. Vic. Dflt.

Role	Weight	Include
ARRESTEE	1	<input checked="" type="checkbox"/>
JUV-ARRESTE	1	<input checked="" type="checkbox"/>
SUSPECT	0.9	<input checked="" type="checkbox"/>
JUV-SUSPECT	0.9	<input checked="" type="checkbox"/>
SUMMONS	0.9	<input checked="" type="checkbox"/>
JUV-SUMMONS	0.9	<input checked="" type="checkbox"/>
VICTIM/ARRE	0.8	<input checked="" type="checkbox"/>
JUV-VIC/ARR	0.8	<input checked="" type="checkbox"/>
JUV-VIC/SUS	0.7	<input checked="" type="checkbox"/>
JUV-INVOLVE	0.7	<input checked="" type="checkbox"/>
INVOLVED/IL	0.7	<input checked="" type="checkbox"/>
VICTIM/SUSP	0.7	<input checked="" type="checkbox"/>
JUV-SUBJECT	0.5	<input checked="" type="checkbox"/>
SUBJECT	0.5	<input checked="" type="checkbox"/>
RUNAWAY	0.4	<input checked="" type="checkbox"/>

Record: 1 of 28 No Filter

Include Categorization Weights

Pull Latest RMS Data

Compute and Store Values

Figure 3: Offense and involvement weights appear on the main database form and are editable by the user. A button restores the defaults.

The *involvement weights* serve as a multiplier against the offense weight. The table below lists the default values. Those arrested for a crime receive the full weight of the offense (x 1) for involvement, while those suspected or summonsed to court receive a lesser weight.

While arrested, suspected, or summonsed weights provide value in identification of top offenders, it is important to consider including other involvements such as witness, victim, or complainant. The role of witness, victim, or complainant should not be included to count “against” an individual in the identification process but should be considered based on research suggesting that offenders also may be victims of specific crimes. For example, gang members involved in inter-gang violence routinely alternate among victim, witness, and suspect involvement within a variety of law enforcement reports.

Involvement	Default Multiplier
Arrested	1.0
Suspect	0.9
Summonsed	0.9
Involved other	0.5
Witness	0.3
Victim	0.1
Reporting	0.1

The final weighting factor, the *date decay* variable, subtracts a small amount of weight (default = 0.002) for every day that passes since the incident occurred. This factor identifies more recently active top offenders and ensures that each individual, no matter how serious his or her offenses, will eventually stop being identified if he or she stops committing crimes.

Note: The resulting identification of top offenders contains one weakness. It identifies only known offenders. Individuals who escape detection for crimes—who avoid showing up in the records system as arrestees or suspects—will not be identified.

PersonID	Name	DOB	Address	Zone	OffWeights	CatWeights	CustomWeig	Total	Sta
1549					83	0	0	83	
3309				2	64	0	0	64	
12003					55	0	0	55	
7345				1	54	0	0	54	
12505					53	0	0	53	
261				2	51	0	0	51	
12381				DT	49	0	0	49	
6759					48	0	0	48	
2442				2	47	0	0	47	
3629					47	0	0	47	
17398				1	47	0	0	47	

IncNum	RecordType	Date	Agency	Location	Offense	Role	OffWeight	RoleWeight
TE201635444	Incident	03/25/2016	TE		CRIMINAL DAMAGE	JUV-SUSPECT	3	0.9
TE20164035	Incident	01/10/2016	TE		ASSAULT	JUV-SUMMONS	6	0.9
TE20163684	Incident	01/09/2016	TE		ASSAULT	JUV-SUMMONS	6	0.9
STE201525495	Street Check	12/11/2015	TE		STREET CHECK	SUBJECT	1	0.5
STE201525243	Street Check	11/25/2015	TE		STREET CHECK	SUBJECT	1	0.5
STE201525239	Street Check	11/24/2015	TE		STREET CHECK	SUBJECT	1	0.5
STE201521908	Street Check	10/25/2015	TE		STREET CHECK	SUBJECT	1	0.5
TE2015123952	Incident	10/07/2015	TE		DISORDERLY CONDUCT	JUV-ARRESTE	2	1
STE201517002	Street Check	08/29/2015	TE		STREET CHECK	SUBJECT	1	0.5

Figure 4: The database generates a ranked list of top offenders and offers the ability to "expand" each offender's history to see what specific involvements make up his or her score.