Our goal: predict whether your case is likely to be solved.

Prediction benefits:

* It can tell law enforcement when to allocate additional resources to aid investigation or when to stop investigation; whether the resources are well-spent or if the case is likely to go unsolved, unnecessarily draining resources.
* Recent unsolved cases that are predicted to be solved can be reanalyzed to be solved, making it a highly valuable investment of law enforcement time to reinvestigate cases. If the available toolset changed significantly, however, the prediction may not apply due to difference in evidence quality. Also, the witnesses and other information may not be available anymore.

What are we working with?

* Our dataset: the complete Uniform Crime Report data summarizing all homicides and homicide clearances reported from 1976 to 2020
* Attention: ID=one incident (789,664 in total)

row=one victim (827,219 in total)

Describing each of the 31 variables in MAP's version of the Supplementary Homicide Report:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| variable | reason to exclude | 827,219 rows:  no. of groups | definition  and  comments | include/  exclude | group by |
| ID | - | 789,664 IDs | unique record identifier, based on the YEAR, MONTH, INCIDENT and ORI | IN | first entry |
| CNTYFIPS | purely administrative | 3,065 counties | state and county of the reporting law enforcement agency | EX | - |
| ORI | Too disaggregated; see state | 13,037 agencies | the originating agency making the report (state + county + number within the state) | EX | - |
| INCIDENT | purely administrative | 1,000 incidents | used to assist in building a unique record number for each case and to differentiate each case reported within the same month | EX | - |
| ACTIONTYPE | purely administrative |  | Whether it was a normal update or an adjustment to a previous report  adjustment overwrites normal update | EX | - |
| MSA | Since the definition of the metropolitan statistical areas, represented by the feature MSA, changed over time, the feature should be excluded as well. |  |  | EX | - |
| AGENCY | agencies may have the same name, too disaggregated | 9,687 agencies | name of the law enforcement agency | EX | - |
| 1 agency = unknown;  names not uniform? |
| STATE |  | 52 states | state of the originating agency making the report  USA has 50 states, PAPSP8 is not one of them; Washington D.C. is neither a state itself nor does it belong to any state | IN | first entry |
| STATENAME | has 804,751 instances of missing values recorded | 52 states | the state of the reporting agency  ? for missing value (counts as one of the 52 state) | EX | - |
| FILEDATE | describes the date a record was reported, not the date of the occurrence of the crime. Also, in some cases, date was not reported and has been estimated. In some other cases, the date indicates when local or state police provided the record to MAP. | 5,569 dates | describes when the homicide was reported  2,510 missing values | EX | - |
| AGENTYPE | - | 8 types | the type of law enforcement agency  “4” is not part of the FBI numbering scheme -> 4 wrong entries | IN | first entry |
| SOURCE | The Supplementary Homicide Reports (abbreviated SHR) is a homicide report database run by the FBI. Florida is the only state not participating in the program but instead reporting their cases in a separate tally. The feature SOURCE will therefore record FBI for all states except Florida, which will be recorded as MAP. As such, the feature doesn’t provide any additional value for our analysis. | 2 sources | if record provided by FBI (SOURCE=1) or was obtained by MAP under the Freedom of Information Act | EX | - |
| SOLVED | - | 2 outcomes | whether the offender was identified at time report was made or not identified | PREDICT |  |
| YEAR | - | 45 years | year of homicide occurrence or when the victim’s body was recovered | IN | first entry |
| MONTH | - | 12 Months | month of homicide occurrence or when the victim’s body was recovered | IN | first entry |
| HOMICIDE | arbitrary distinction | 2 classes | variable defining whether the report was murder/non-negligent manslaughter or manslaughter by negligence  default: murder/non-negligent manslaughter | EX | - |
| SITUATION | the value can be false due to the true number of offenders (and in few cases victims) being unknown | 6 combinations | as the body is the clue to the crime, there are only: single victim/multiple victims but single offender, multiple offenders, unknown offender(s)  [default: single victim/unknown offender(s)] | EX | - |
| VICAGE | - | 101 age groups | 1-98 is the actual age, 99 is anyone 99+, 0 is anyone below one year of age, 999 is any unknown age  12,000 unknown age variables | IN | mean/first |
| VICSEX |  | 3 sexes | unknown possible | IN  statistical analysis needed | ? |
| VICRACE |  | 6 races | Races comprise White, Black/African American, Asian, American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, or “other.”  unknown possible | IN  statistical analysis needed | ? |
| VICETHNIC | many agencies decline reporting the ethnicity of victims and offenders due to it being only Hispanic/Latino and Not Hispanic/Latino. |  |  | EX | - |
| RELATIONSHIP | unknown if case is not solved |  | the relationship between the victim and the offender; if any  only the first offender included in the SHR report is being cited in the data; relationship not determined possible | EX | - |
| OFFAGE | 1. not all offenders have to be known 2. Unlike victims, the MAP does not create multiple case reports for additional offenders. Only the first offender listed in the original SHR report is included. For an incident with more than 11 victims, there are multiple SHR reports.- |  |  | EX | - |
| OFFSEX | 1. not all offenders have to be known 2. Unlike victims, the MAP does not create multiple case reports for additional offenders. Only the first offender listed in the original SHR report is included. For an incident with more than 11 victims, there are multiple SHR reports. |  |  | EX | - |
| OFFRACE | 1. not all offenders have to be known 2. Unlike victims, the MAP does not create multiple case reports for additional offenders. Only the first offender listed in the original SHR report is included. For an incident with more than 11 victims, there are multiple SHR reports. |  |  | EX | - |
| OFFETHNIC | many agencies decline reporting the ethnicity of victims and offenders due to it being only Hispanic/Latino and Not Hispanic/Latino. |  |  | EX | - |
| CIRCUMSTANCES | theory to the crime – speculative; not needed | 32 circumstances | the circumstances (or theory) for the crime  Circumstances undetermined (for unknown) other (catch-all term) | EX | - |
| SUB-  CIRCUMSTANCE | see circumstance | 8 subcircumstances | several conditions in which the victim is reported to have been a criminal offender  ? (for non-felons)  Not enough information to determine  (for felons) | EX | - |
| VICCOUNT | number will be included in the diversity of data | 0-11; 21 | number of additional victims  The MAP’s database captures all reported homicide victims as separate cases. Associated victims will have identical ID numbers. | EX | - |
| OFFCOUNT | unknown if not yet solved | 0-12; 17; 21; 40 | number of additional offenders | EX | - |
| WEAPON |  | 17 weapons | default: Other or type unknown  grouping possible:  Firearm (Shot wound)   * Firearm, type not stated * Handgun – pistol, revolver, etc. * Rifle * Shotgun * Other gun   Blunt Force (BF trauma)   * Blunt object – hammer, club, etc. * Personal weapons, including beating * Pushed or thrown out of window   Toxins (contact with toxins)   * Narcotics or drugs, sleeping pills * Poison, does not include gas   Burns   * Explosives * Fire   Oxygen deficiency   * Drowning * Strangulation or hanging * Asphyxiation – includes death by gas   Other or type unknown weapon | IN |  |