How Police uses Data to Predict Crime

Preamble

Los Angeles Police Department is one of dozens of cities across the country that's trying to predict where crime will happen—and who those future criminals will be—based on past crime and arrest data. One effort, known as Operation LASER, which began in 2011, crunches information about past offenders over a two-year period, using technology developed by the shadowy data analysis firm Palantir, and scores individuals based on their rap sheets. If you've ever been in a gang, that's five points. If you're on parole or probation? Another five. Every time you're stopped by police, every time they come knocking on your door, that could land you more points. The higher the points, the more likely you are to end up on something called the Chronic Offender Bulletin, a list of people the data says are most at risk of reoffending and ought to be kept on close watch.

The city says this so-called "predictive policing" approach can help the department efficiently target resources and help reduce crime. But civil rights advocates worry that all this fancy technology is just a glossy veneer on old-school racial profiling.

On the other hand, San Francisco City Council had develop an initiative called DataSF with the mission to empower the use of data. DataSF seeks to transform the way the San Francisco City works through the use of data. They believe that

use of data can improve the operations and the services of San Francisco City Council. This ultimately leads to increased quality of life and work for San Francisco residents, employers, employees and visitors.

The Challenge

The Police Department of San Francisco have been collecting incident reports from 2018 to 2020. This dataset includes police incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases. Reports included are those for incidents that occurred starting January 1, 2018 onward and have been approved by a supervising officer.

The fields included in the data set are described below.

Field Name	Description	Type
Incident Datetime	The date and time when the incident occurred	Date & Time
Incident Date	The date the incident occurred	Date & Time
Incident Time	The time the incident occurred	Plain Text
Incident Year	The year the incident occurred, provided as a convenience for filtering	Plain Text
Incident Day of Week	The day of week the incident occurred	Plain Text
Report Datetime	Distinct from Incident Datetime, Report Datetime is when the report was filed.	Date & Time
Row ID	A unique identifier for each row of data in the dataset	Plain Text

Incident ID	This is the system generated identifier for incident reports. Incident IDs and Incident Numbers both uniquely identify reports, but Incident Numbers are used when referencing cases and report documents.	Plain Text
Incident Number	The number issued on the report, sometimes interchangeably referred to as the Case Number. This number is used to reference cases and report documents.	Plain Text
CAD Number	The Computer Aided Dispatch (CAD) is the system used by the Department of Emergency Management (DEM) to dispatch officers and other public safety personnel. CAD Numbers are assigned by the DEM system and linked to relevant incident reports (Incident Number). Not all Incidents will have a CAD Number. Those filed online via Coplogic (refer to “Filed Online” field) and others not filed through the DEM system will not have CAD Numbers.	Plain Text

Report Type Code	A system code for report types, these have corresponding descriptions within the dataset.	Plain Text
Report Type Description	The description of the report type, can be one of: Initial; Initial Supplement; Vehicle Initial; Vehicle Supplement; Coplogic Initial; Coplogic Supplement	Plain Text
Filed Online	Non- emergency police reports can be filed online by members of the public using SFPD's self-service reporting system called Coplogic Values in this field will be "TRUE" if Coplogic was used to file the report. Please reference the link below for additional info: (http:// sanfranciscopolice.org/ reports).	Checkbox

Incident Code	Incident Codes are the system codes to describe a type of incident. A single incident report can have one or more incident types associated. In those cases you will see multiple rows representing a unique combination of the Incident ID and Incident Code.	Plain Text
Incident Category	A category mapped on to the Incident Code used in statistics and reporting. Mappings provided by the Crime Analysis Unit of the Police Department.	Plain Text
Incident Subcategory	A subcategory mapped to the Incident Code that is used for statistics and reporting. Mappings are provided by the Crime Analysis Unit of the Police Department.	Plain Text
Incident Description	The description of the incident that corresponds with the Incident Code. These are generally self-explanatory.	Plain Text

Resolution	The resolution of the incident at the time of the report. Can be one of: • Cite or Arrest Adult • Cite or Arrest Juvenile* • Exceptional Adult • Exceptional Juvenile* • Open or Active • Unfounded Note: once a report is filed, the Resolution will not change. Status changes and/or updates must be provided using a Supplemental Report *Incidents identifying juvenile information are not included in this dataset. Please see the Juvenile Data section for more information.	Plain Text
Intersection	The 2 or more street names that intersect closest to the original incident separated by a backward slash (\). Note, the possible intersections will only include those that satisfy the privacy controls.	Plain Text

CNN	The unique identifier of the intersection for reference back to other related basemap datasets. For more on the Centerline Node Network see https://datasf.gitbook.io/draft-publishing-standards/standard-reference-data/basemap/street-centerlines-nodes	Plain Text
Police District	The Police District where the incident occurred. District boundaries can be reviewed in the link below. Please note this field is entered by officers and not based on the point. Reference here: https://data.sfgov.org/d/wkhw-cjsf	Plain Text

Analysis Neighborhood	This field is used to identify the neighborhood where each incident occurs. Neighborhoods and boundaries are defined by the Department of Public Health and the Mayor's Office of Housing and Community Development. Please reference the link below for additional info: https://data.sfgov.org/d/p5b7-5n3h Please note this boundary is assigned based on the intersection, it may differ from the boundary the incident actually occurred within.	Plain Text

Supervisor District	There are 11 members elected to the Board of Supervisors in San Francisco, each representing a geographic district. The Board of Supervisors is the legislative body for San Francisco. The districts are numbered 1 through 11. Please reference the link below for additional info: https://data.sfgov.org/d/8nkz-x4ny Please note this boundary is assigned based on the intersection, it may differ from the boundary the incident actually occurred within.	Plain Text
Latitude	The latitude coordinate in WGS84, spatial reference is EPSG:4326	Number
Longitude	The longitude coordinate in WGS84, spatial reference is EPSG:4326	Number
point	The point geometry used for mapping features in the open data portal platform. Latitude and Longitude are provided separately as well as a convenience.	Location

The San Francisco City Council in conjunction with the Police Department want to develop an interactive map where the citizens can detect risk areas or obtain

information about the crimes that occur so they can take precautions. Additionally, they want citizens to help collect more information about the crimes that are committed in order to alert the Police Department and even as they do in Los Angeles to predict a crime.

You can download the date from the following link (click here)

Delivery method

This activity is individually. Send the links of your web app using Plotly & Streamlit.