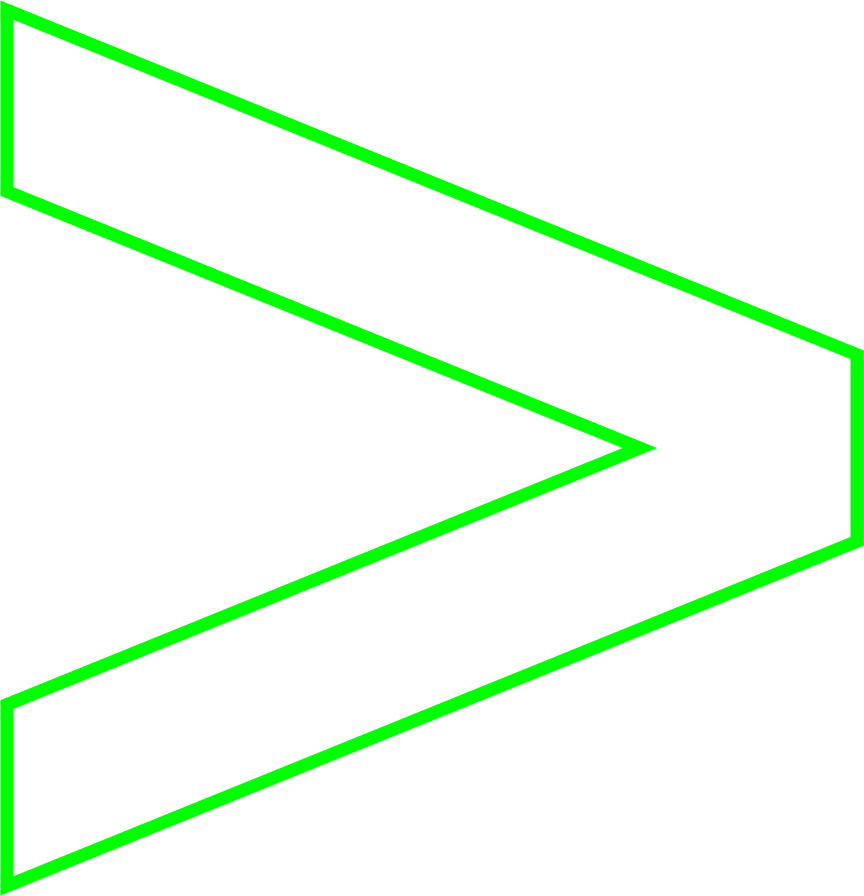
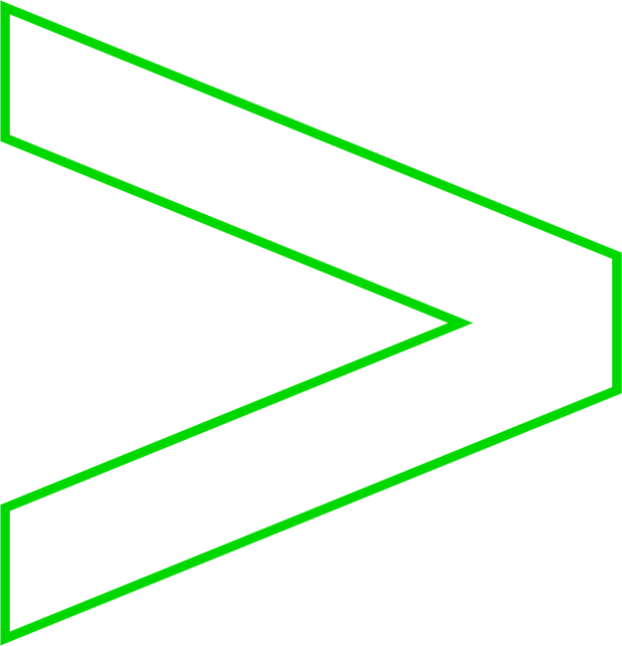


**AWS DATA ANALYTICS SPECIALTY**



**CASE STUDY**



**CHICAGO CRIMES DATASET**

**Table of Contents**

[1 Background 5](#_Toc51676679)

[2 Challenges 6](#_Toc51676680)

[3 Business Need 7](#_Toc51676681)

[4 Proposed Solution 8](#_Toc51676682)

[5 Data Model, Metadata and Sample Data 9](#_Toc51676683)

# 1 Background

* For the past few years, Chicago has been considered one of the most dangerous cities in the US. The motivation behind this project is to help the Chicago Police Department make better decisions with the use of data. It is important for them to receive key insights from this dataset. This dataset contains information from the Chicago Police Department(CPD) from 2012 to 2016.
* From the dataset the CPD wants to obtain significantly important information like: Most common type of crimes - crime rate across the years - Crimes that resulted in Arrests - Top IUCR Codes - Most common areas where crimes occur - Top FBI Codes and so forth.
* CPD wants to draw reliable and definite conclusions regarding % of the criminals are arrested ,which is the most common crime, which locality needs more patrols going around the area and accordingly set a specific perimeter for the areas where crime scene is extremely high.
* CPD has decided to allocate this “crime data analytics” project to **Accenture** to help it gain detailed insights into the crime trends to prevent crimes and to protect the masses.
* **Accenture** is responsible for storing, developing and maintaining the data and analytical systems for reduction and prevention of crimes in Chicago.
* Accenture wishes to leverage the market leader in the Cloud space - the **Amazon Web Services** to fulfill the requirements
* Based on the analysis of the Chicago crimes dataset, we should offer several recommendations to the Chicago Police Department on the several aspects of crimes in order to prevent them.

# 2 Challenges

* Currently, the CPD has to put in a lot of time and effort in manually performing the analysis of the huge number of crimes which get reported every day.
* There is no system in place to record and track the crime rates on a real time basis and hence the CPD is not able to contain the crime rates within the city in a near-real time fashion which is causing a huge damage to the public and society in general.
* The senior most officers of the CPD are not getting a big picture understanding of different types of crime happening in the city and hence they are not able to come up with preventive measures like having more patrols going around the most vulnerable crime areas to mitigate the crimes.

# 3 Business Need

* The requirement is to stage, process and store the crimes data and transform it to the analytical needs. Finally build reports/visualizations which provide actionable insights.
* Another requirement is to ingest the crime data to get near real-time recommendations to prevent crimes in the city.
* Additionally, none of the data should be discarded. In fact, it should be preserved for enabling use-cases related to regulatory compliance, public safety, insurance, prevention and deduction of crimes in the city.
* CPD wants Accenture to build a Cloud-based common platform to store all data related to crimes in order to analyze the data for better visibility on crime trends, to minimize and prevent crimes and to ensure public safety in whole of Chicago.

# 4 Proposed Solution

* Considering the Volume, Velocity and Veracity aspects of the data, Accenture has decided to build a Cloud-based storage and analytics system using **Amazon Web Services**.
* Accenture has analyzed various available options and delivered some Proof of Concepts based on the data received from the CPD.
* In a one-year contract with CPD, Accenture will build a cloud-based solution on AWS using services like S3, Kinesis, Dynamodb, EMR, Redshift, Athena, QuickSight etc.

# 5 Data Model, Metadata and Sample Data

**Data Model:**

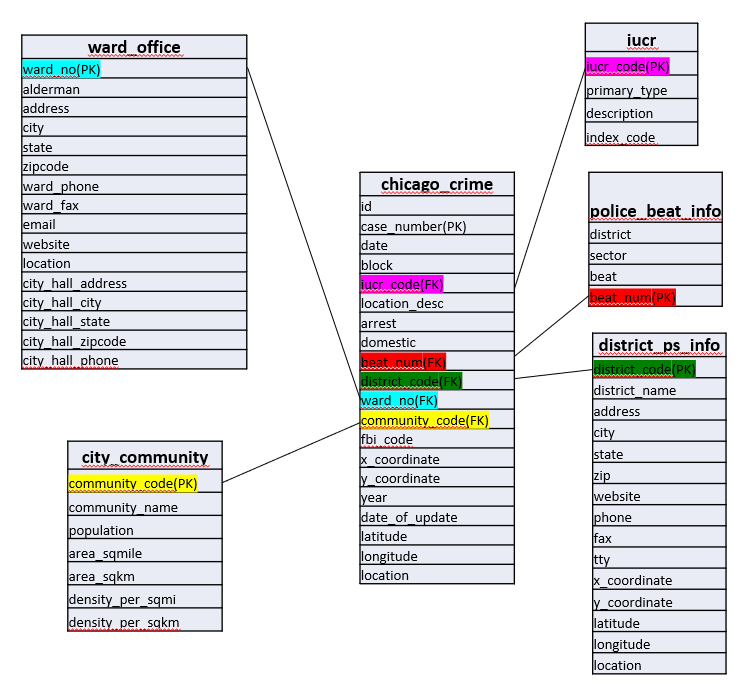
The below data model represents the master and dependent files relationship for the Chicago crimes dataset.

The complete dataset is stored in multiple CSV files.

The master data is contained in file “chicago\_crime\_dataset.csv”

The datasets listed below are referenced in the master dataset.

* chicago\_district\_ps\_info.csv
* chicago\_ward\_offices.csv
* iucr\_codes.csv
* chicago\_city\_community.csv
* chicago\_police\_beat\_info.csv



**Metadata:**

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| id | Unique identifier for the record. |
| case\_number | The Chicago Police Department RD Number (Records Division Number)  which is unique to the incident. |
| date | Date when the incident occurred. |
| block | The partially redacted address where the incident occurred,  placing it on the same block as the actual address. |
| iucr\_code | The Illinois Uniform Crime Reporting code.  This is directly linked to the Primary Type and Description. |
| primary\_type | The primary description of the IUCR code. |
| description | The secondary description of the IUCR code, a subcategory  of the primary description. |
| location\_desc | Description of the location where the incident occurred. |
| arrest | Indicates whether an arrest was made. |
| domestic | Indicates whether the incident was domestic related as defined  by the Illinois Domestic Violence Act. |
| beat\_num | Indicates the beat where the incident occurred. A beat is the smallest police geographic area  – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. |
| district\_name | Indicates the police district where the incident occurred. |
| ward\_no | The ward (City Council district) where the incident occurred. |
| community\_code | Indicates the community area where the incident occurred. |
| fbi\_code | Indicates the crime classification as outlined in  the FBI's National Incident-Based Reporting System (NIBRS). |
| x\_ coordinate | The x coordinate of the location where the incident occurred in State Plane  Illinois East NAD 1983 projection. This location is shifted from the actual location  for partial redaction but falls on the same block. |
| y\_coordinate | The y coordinate of the location where the incident occurred in State Plane  Illinois East NAD 1983 projection. This location is shifted from the actual location  for partial redaction but falls on the same block. |
| year | Year the incident occurred. |
| date\_of\_update | Date and time the record was last updated. |
| latitude | The latitude of the location where the incident occurred. This location is  shifted from the actual location for partial redaction but falls on the same block. |
| longitude | The longitude of the location where the incident occurred. This location is shifted  from the actual location for partial redaction but falls on the same block. |
| location | The location where the incident occurred in a format that allows for creation of maps and other  geographic operations on this data portal. This location is shifted from the actual location for  partial redaction but falls on the same block. |

**Data Samples:**

***Sample Data: chicago\_district\_ps\_info.csv***

|  |  |
| --- | --- |
| DISTRICT\_CODE | 1 |
| DISTRICT\_NAME | Central |
| ADDRESS | 1718 S State St |
| CITY | Chicago |
| STATE | IL |
| ZIP | 60616 |
| WEBSITE | http://home.chicagopolice.org/community/districts/1st-district-central/ |
| PHONE | 312-745-4290 |
| FAX | 312-745-3694 |
| TTY | 312-745-3693 |
| X\_COORDINATE | 1176569.052 |
| Y\_COORDINATE | 1891771.704 |
| LATITUDE | 41.85837259 |
| LONGITUDE | -87.62735617 |
| LOCATION | (41.8583725929, -87.627356171) |

***Sample Data: chicago\_ward\_offices.csv***

|  |  |
| --- | --- |
| WARD\_NO | 1 |
| ALDERMAN | La Spata, Daniel |
| ADDRESS | 1958 North Milwaukee Avenue |
| CITY | Chicago |
| STATE | IL |
| ZIPCODE | 60647 |
| WARD\_PHONE | (872) 206-2685 |
| WARD\_FAX | (312) 448-8829 |
| EMAIL | info@the1stward.com |
| WEBSITE | https://www.the1stward.com/ |
| LOCATION | (41.917215, -87.68799) |
| CITY\_HALL\_ADDRESS | 121 North LaSalle Street, Room 200, Office 13 |
| CITY\_HALL\_CITY | Chicago |
| CITY\_HALL\_STATE | IL |
| CITY\_HALL\_ZIPCODE | 60602 |
| CITY\_HALL\_PHONE | (312) 744-3063 |

***Sample Data: iucr\_codes.csv***

|  |  |
| --- | --- |
| IUCR\_CODE | 110 |
| PRIMARY\_TYPE | HOMICIDE |
| DESCRIPTION | FIRST DEGREE MURDER |
| INDEX\_CODE | I |

***Sample Data: chicago\_crime\_dataset.csv***

|  |  |
| --- | --- |
| id | 10508693 |
| case\_number | HZ250496 |
| date | 5/3/2016 23:40 |
| block | 013XX S SAWYER AVE |
| iucr\_code | 486 |
| location\_desc | APARTMENT |
| arrest | TRUE |
| domestic | TRUE |
| beat\_num | 1022 |
| district\_code | 10 |
| ward\_no | 24 |
| community\_code | 29 |
| fbi\_code | 08B |
| x\_coordinate | 1154907 |
| y\_coordinate | 1893681 |
| year | 2016 |
| date\_of\_update | 5/10/2016 15:56 |
| latitude | 41.86407316 |
| longitude | -87.70681861 |
| location | (41.864073157, -87.706818608) |

***Sample Data: chicago\_city\_community.csv***

|  |  |
| --- | --- |
| community\_code | 1 |
| community\_name | Rogers Park |
| population | 55,062 |
| area\_sqmile | 1.84 |
| area\_sqkm | 4.77 |
| density\_per\_sqmi | 29,925.00 |
| density\_per\_sqkm | 11,554.11 |

***Sample Data: chicago\_police\_beat\_info.csv***

|  |  |
| --- | --- |
| DISTRICT | 1 |
| SECTOR | 1 |
| BEAT | 1 |
| BEAT\_NUM | 111 |

Copyright © 2019 Accenture

All rights reserved.

Accenture and its logo are   
trademarks of Accenture.