

# **Databases and SQL for Data Science**

Assignment:

Chicago Crime Data

## Dataset Used in this Assignment

A real-world dataset provided by the Chicago Data Portal. This assignment involves 3 datasets for the city of Chicago obtained from the Chicago Data Portal:

### 1. Chicago Socioeconomic Indicators

This [dataset](#) contains a selection of six socioeconomic indicators of public health significance and a *hardship index*, by Chicago community area, for the years 2008 – 2012.

### 2. Chicago Public Schools

This [dataset](#) shows all school level performance data used to create CPS School Report Cards for the 2011-2012 school year.

### 3. Chicago Crime Data

This [dataset](#) reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days.

## Scenario

You have been hired by a non-profit organization that strives to improve educational outcomes for children and youth in the City of Chicago. Your job is to analyze the census, crime, and school data for a given neighborhood or district. You will identify causes that impact the enrollment, safety, health, environment ratings of schools.

# Tasks

## 1. Review the datasets

Before you begin, you will need to become familiar with the datasets. Snapshots for the three datasets in .CSV format can be downloaded from the following links:

- Chicago Socioeconomic Indicators: [Click here](#)
- Chicago Public Schools: [Click here](#)
- Chicago Crime Data: [Click here](#)

## 2. Load the datasets into a database

You will need to create 3 tables in the database, one for each dataset, named as follows, and then load the respective .CSV file into the table:

1. CENSUS\_DATA
2. CHICAGO\_PUBLIC\_SCHOOLS
3. CHICAGO\_CRIME\_DATA

## 3. Write and execute queries

- Problem 1: Find the total number of crimes recorded in the CRIME table.
- Problem 2: List community areas with per capita income less than 11000.
- Problem 3: List all case numbers for crimes involving minors?
- Problem 4: List all kidnapping crimes involving a child?(children are not considered minors for the purposes of crime analysis)
- Problem 5: What kind of crimes were recorded at schools?
- Problem 6: List the average safety score for all types of schools.

- Problem 7: List 5 community areas with highest % of households below poverty line.
- Problem 8: Which community area(number) is most crime prone?
- Problem 9: Use a sub-query to find the name of the community area with highest hardship index.
- Problem 10: Use a sub-query to determine the Community Area Name with most number of crimes?