Scenario

Imagine you have been hired as a Data Science Consultant by a non-profit organization to help analyze socioeconomic conditions, and identify insights to improve quality of education and reduce crime in the city of Chicago. Your first task is to review several Chicago datasets and perform some exploratory analysis.

Tasks

For the exploratory analysis you will perform the following tasks:

1. Download the datasets and become familiar with what they contain.

2. LOAD the datasets into a database

3. Utilize Python based Jupyter notebooks to compose and run SQL queries that retrieve specific results.

Once you have performed these tasks you will be tested in demonstrating working knowledge of SQL and correctness of queries and their result sets.

About the Datasets

To complete the assignment problems in this notebook you will be using three datasets that are available on the city of Chicago's Data Portal:

1. Socioeconomic Indicators in Chicago

2. Chicago Public Schools

3. Chicago Crime Data

1. Socioeconomic Indicators in Chicago

This dataset contains a selection of six socioeconomic indicators of public health significance and a "hardship index," for each Chicago community area, for the years 2008 – 2012.

A detailed description of this dataset and the original dataset can be obtained from the Chicago Data Portal at: https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

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. This dataset is provided by the city of Chicago's Data Portal.

For this assignment you will use a snapshot of this dataset which can be downloaded from: https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhplzt01qtz0s7ew7.csv

A detailed description of this dataset and the original dataset can be obtained from the Chicago Data Portal at: https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t

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.

This dataset is quite large - over 1.5GB in size with over 6.5 million rows. For the purposes of this assignment we will use a much smaller subset of this dataset which can be downloaded from: https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv

A detailed description of this dataset and the original dataset can be obtained from the Chicago Data Portal at: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

Download and Review the datasets

NOTE: For the purposes of this assignment ensure you download the datasets using the links below instead of directly from the Chicago Data Portal. The versions linked here are subsets of the original datasets. For example the original Crime dataset is over 1.5GB in size with over 6.5 million rows, however to make it easier to complete this assignment we will utilize a much smaller sample of this dataset. In some cases the versions of datasets below have had some of the column names modified to be more database friendly which will make it easier to complete this assignment.

Click on the links below to download and save the datasets (.CSV files):

1. **CENSUS_DATA:** https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv

2. **CHICAGO_PUBLIC_SCHOOLS** https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhpl zt01qtz0s7ew7.csv

3. **CHICAGO_CRIME_DATA:** https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgsw qemfpm1x7f.csv

Once you have downloaded the above datasets, preview them using a text editor (e.g. Notepad / TextEdit) or a spreadhseet (Excel/Numbers/Google Sheets) - but do NOT modify or save them. If you make any changes, the data may not load properly in the database.