Introduction

Using this Python notebook you will:

1. Understand three Chicago datasets
2. Load the three datasets into three tables in a SQLIte database
3. Execute SQL queries to answer assignment questions

Understand 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.

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.

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 the datasets

This assignment requires you to have these three tables populated with a subset of the whole datasets.

In many cases the dataset to be analyzed is available as a .CSV (comma separated values) file, perhaps on the internet.

Use the links below to read the data files using the Pandas library.

- Chicago Census Data

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCensusData.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01

- Chicago Public Schools

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoPublicSchools.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01

- Chicago Crime Data

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCrimeData.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01

NOTE: Ensure you use the datasets available on the links above instead of directly from the Chicago Data Portal. The versions linked here are subsets of the original datasets and have some of the column names modified to be more database friendly which will make it easier to complete this assignment.

Execute the below code cell to avoid prettytable default error.

In [1]:

```
!pip install ipython-sql prettytable

import prettytable

prettytable.DEFAULT = 'DEFAULT'
```

```
Requirement already satisfied: ipython-sql in /opt/conda/lib/python3.11/site-packages (0.5.0)
Requirement already satisfied: prettytable in /opt/conda/lib/python3.11/site-packages (3.12.0)
Requirement already satisfied: ipython in /opt/conda/lib/python3.11/site-packages (from ipython-sql)
(8.22.2)
Requirement already satisfied: sqlalchemy>=2.0 in /opt/conda/lib/python3.11/site-packages (from
ipython-sql) (2.0.30)
Requirement already satisfied: sqlparse in /opt/conda/lib/python3.11/site-packages (from ipython-sql)
(0.5.3)
Requirement already satisfied: six in /opt/conda/lib/python3.11/site-packages (from ipython-sql) (1.16.0)
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.11/site-packages (from
ipython-sql) (0.2.0)
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.11/site-packages (from prettytable)
(0.2.13)
Requirement already satisfied: typing-extensions>=4.6.0 in /opt/conda/lib/python3.11/site-packages (from
sqlalchemy>=2.0->ipython-sql) (4.12.2)
Requirement already satisfied: greenlet!=0.4.17 in /opt/conda/lib/python3.11/site-packages (from
sqlalchemy>=2.0->ipython-sql) (3.0.3)
Requirement already satisfied: decorator in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (5.1.1)
Requirement already satisfied: jedi>=0.16 in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (0.19.1)
Requirement already satisfied: matplotlib-inline in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (0.1.7)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in /opt/conda/lib/python3.11/site-packages
(from ipython->ipython-sql) (3.0.42)
Requirement already satisfied: pygments>=2.4.0 in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (2.18.0)
Requirement already satisfied: stack-data in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (0.6.2)
Requirement already satisfied: traitlets>=5.13.0 in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (5.14.3)
Requirement already satisfied: pexpect>4.3 in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (4.9.0)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /opt/conda/lib/python3.11/site-packages (from
jedi>=0.16->ipython->ipython-sql) (0.8.4)
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.11/site-packages (from
pexpect>4.3->ipython->ipython-sql) (0.7.0)
Requirement already satisfied: executing>=1.2.0 in /opt/conda/lib/python3.11/site-packages (from
stack-data->ipython->ipython-sql) (2.0.1)
Requirement already satisfied: asttokens>=2.1.0 in /opt/conda/lib/python3.11/site-packages (from
stack-data->ipython->ipython-sql) (2.4.1)
Requirement already satisfied: pure-eval in /opt/conda/lib/python3.11/site-packages (from
stack-data->ipython->ipython-sql) (0.2.2)
```

Store the datasets in database tables

To analyze the data using SQL, it first needs to be loaded into SQLite DB. We will create three tables in as under:

1. **CENSUS_DATA**
2. **CHICAGO_PUBLIC_SCHOOLS**
3. **CHICAGO_CRIME_DATA**

Load the `pandas` and `sqlite3` libraries and establish a connection to `FinalDB.db`

In [2]:

```
%load_ext sql
```

In [3]:

```
import csv, sqlite3

conn = sqlite3.connect('FinalDB.db')
cur = conn.cursor()
!pip install pandas
```

```
Requirement already satisfied: pandas in /opt/conda/lib/python3.11/site-packages (2.2.3)
Requirement already satisfied: numpy>=1.23.2 in /opt/conda/lib/python3.11/site-packages (from pandas)
(2.2.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.11/site-packages (from
pandas) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-packages (from pandas)
(2024.1)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.11/site-packages (from pandas)
(2024.2)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
```

Load the SQL magic module

In [4]:

```
!pip install ipython-sql
```

```
Requirement already satisfied: ipython-sql in /opt/conda/lib/python3.11/site-packages (0.5.0)
Requirement already satisfied: prettytable in /opt/conda/lib/python3.11/site-packages (from ipython-sql)
(3.12.0)
Requirement already satisfied: ipython in /opt/conda/lib/python3.11/site-packages (from ipython-sql)
(8.22.2)
Requirement already satisfied: sqlalchemy>=2.0 in /opt/conda/lib/python3.11/site-packages (from
ipython-sql) (2.0.30)
Requirement already satisfied: sqlparse in /opt/conda/lib/python3.11/site-packages (from ipython-sql)
(0.5.3)
Requirement already satisfied: six in /opt/conda/lib/python3.11/site-packages (from ipython-sql) (1.16.0)
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.11/site-packages (from
ipython-sql) (0.2.0)
Requirement already satisfied: typing-extensions>=4.6.0 in /opt/conda/lib/python3.11/site-packages (from
sqlalchemy>=2.0->ipython-sql) (4.12.2)
Requirement already satisfied: greenlet!=0.4.17 in /opt/conda/lib/python3.11/site-packages (from
sqlalchemy>=2.0->ipython-sql) (3.0.3)
Requirement already satisfied: decorator in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (5.1.1)
Requirement already satisfied: jedi>=0.16 in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (0.19.1)
Requirement already satisfied: matplotlib-inline in /opt/conda/lib/python3.11/site-packages (from
ipython->ipython-sql) (0.1.7)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in /opt/conda/lib/python3.11/site-packages
(from ipython->ipython-sql) (3.0.42)
```

```
Requirement already satisfied: pygments>=2.4.0 in /opt/conda/lib/python3.11/site-packages (from ipython->ipython-sql) (2.18.0)
Requirement already satisfied: stack-data in /opt/conda/lib/python3.11/site-packages (from ipython->ipython-sql) (0.6.2)
Requirement already satisfied: traitlets>=5.13.0 in /opt/conda/lib/python3.11/site-packages (from ipython->ipython-sql) (5.14.3)
Requirement already satisfied: pexpect>4.3 in /opt/conda/lib/python3.11/site-packages (from ipython->ipython-sql) (4.9.0)
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.11/site-packages (from prettytable->ipython-sql) (0.2.13)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /opt/conda/lib/python3.11/site-packages (from jedi>=0.16->ipython->ipython-sql) (0.8.4)
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.11/site-packages (from pexpect>4.3->ipython->ipython-sql) (0.7.0)
Requirement already satisfied: executing>=1.2.0 in /opt/conda/lib/python3.11/site-packages (from stack-data->ipython->ipython-sql) (2.0.1)
Requirement already satisfied: asttokens>=2.1.0 in /opt/conda/lib/python3.11/site-packages (from stack-data->ipython->ipython-sql) (2.4.1)
Requirement already satisfied: pure-eval in /opt/conda/lib/python3.11/site-packages (from stack-data->ipython->ipython-sql) (0.2.2)
```

Use `Pandas` to load the data available in the links above to dataframes. Use these dataframes to load data on to the database `FinalDB.db` as required tables.

Establish a connection between SQL magic module and the database `FinalDB.db`

In [5]:

```
import pandas as pd
df1 = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCensusData.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01.csv")
df2 = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoPublicSchools.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01.csv")
df3 = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoPublicSchools.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01.csv")

df1.to_sql("CHICAGO_CENSUS_DATA",conn,if_exists='replace',index=False,method="multi")
df2.to_sql("CHICAGO_CRIME_DATA",conn,if_exists='replace',index=False,method="multi")
```

```
df3.to_sql("CHICAGO_PUBLIC_SCHOOLS",conn, if_exists='replace', index=False,method="multi")
```

Out[5]:

```
566
```

In [6]:

```
%sql sqlite:///FinalDB.db
```

In [7]:

```
%sql SELECT name FROM sqlite_master WHERE type='table'
```

```
* sqlite:///FinalDB.db
Done.
```

Out[7]:

name
CHICAGO_CENSUS_DATA
CHICAGO_CRIME_DATA
CHICAGO_PUBLIC_SCHOOLS

You can now proceed to the the following questions. Please note that a graded assignment will follow this lab and there will be a question on each of the problems stated below. It can be from the answer you received or the code you write for this problem. Therefore, please keep a note of both your codes as well as the response you generate.

Problems

Now write and execute SQL queries to solve assignment problems

Problem 1

Find the total number of crimes recorded in the CRIME table.

In [12]:

```
%sql select count(*) from CHICAGO_CRIME_DATA
```

```
* sqlite:///FinalDB.db
Done.
```

Out[12]:

count(*)
566

Problem 2

List community area names and numbers with per capita income less than 11000.

In [9]:

```
%sql SELECT COMMUNITY_AREA_NAME, PER_CAPITA_INCOME FROM
CHICAGO_CENSUS_DATA WHERE PER_CAPITA_INCOME <11000;
```

```
* sqlite:///FinalDB.db
Done.
```

Out[9]:

COMMUNITY_AREA_NAME	PER_CAPITA_INCOME
West Garfield Park	10934
South Lawndale	10402
Fuller Park	10432
Riverdale	8201

Problem 3

List all case numbers for crimes involving minors?(children are not considered minors for the purposes of crime analysis)

In [10]:

```
%sql SELECT CASE_NUMBER FROM CHICAGO_CRIME_DATA WHERE DESCRIPTION LIKE
MINORS;
```

```
* sqlite:///FinalDB.db
(sqlite3.OperationalError) no such column: CASE_NUMBER
[SQL: SELECT CASE_NUMBER FROM CHICAGO_CRIME_DATA WHERE DESCRIPTION LIKE
MINORS;]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

Problem 4

List all kidnapping crimes involving a child?

In [11]:

```
%sql SELECT CASE_NUMBER FROM CHICAGO_CRIME_DATA WHERE PRIMARY_TYPE =
"KIDNAPPING" AND DESCRIPTION LIKE "%CHILD%";
```

```
* sqlite:///FinalDB.db
(sqlite3.OperationalError) no such column: CASE_NUMBER
[SQL: SELECT CASE_NUMBER FROM CHICAGO_CRIME_DATA WHERE PRIMARY_TYPE =
"KIDNAPPING" AND DESCRIPTION LIKE "%CHILD%" ;]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

Problem 5

List the kind of crimes that were recorded at schools. (No repetitions)

In [ ]:

Problem 6

List the type of schools along with the average safety score for each type.

In [ ]:

Problem 7

List 5 community areas with highest % of households below poverty line

In [ ]:

Problem 8

Which community area is most crime prone? Display the coumminty area number only.

In [ ]:

Double-click **here** for a hint

Problem 9

Use a sub-query to find the name of the community area with highest hardship index

In [ ]:

Problem 10

Use a sub-query to determine the Community Area Name with most number of crimes?

In [ ]:

Author(s)

Hima Vasudevan
Rav Ahuja
Ramesh Sannreddy

Contribtuor(s)

Malika Singla
Abhishek Gagneja

<!-- ## Change log <table> Date Version Changed by Change Description 2023-10-18 2.6 Abhishek Gagneja Modified instruction set 2022-03-04 2.5 Lakshmi Holla Changed markdown. 2021-05-19 2.4 Lakshmi Holla Updated the question 2021-04-30 2.3 Malika Singla Updated the libraries 2021-01-15 2.2 Rav Ahuja Removed problem 11 and fixed changelog 2020-11-25 2.1 Ramesh Sannareddy Updated the problem statements, and datasets 2020-09-05 2.0 Malika Singla Moved lab to course repo in GitLab 2018-07-18 1.0 Rav Ahuja Several updates including loading instructions 2018-05-04 0.1 Hima Vasudevan Created initial version
-->