



Assignment: Notebook for Graded Assessment

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-Chicn9c-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?tzp-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?](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)

[utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01](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?](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)

[utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork20127838-2021-01-01](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 [18]: !pip install ipython-sql prettytable

import prettytable

prettytable.DEFAULT = 'DEFAULT'
!pip install pandas

Requirement already satisfied: ipython-sql in /opt/conda/lib/python3.12/site-packages (0.5.0)
Requirement already satisfied: prettytable in /opt/conda/lib/python3.12/site-packages (3.15.1)
Requirement already satisfied: ipython in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (8.31.0)
Requirement already satisfied: sqlalchemy>=2.0 in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (2.0.37)
Requirement already satisfied: sqlparse in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (0.5.3)
Requirement already satisfied: six in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (1.17.0)
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (0.2.0)
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.12/site-packages (from prettytable) (0.2.13)
Requirement already satisfied: greenlet==0.4.17 in /opt/conda/lib/python3.12/site-packages (from sqlalchemy==2.0->python-sql) (3.1.1)
Requirement already satisfied: typing-extensions==4.6.0 in /opt/conda/lib/python3.12/site-packages (from sqlalchemy==2.0->python-sql) (4.12.2)
Requirement already satisfied: decorator in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (5.1.1)
Requirement already satisfied: jedi==0.16 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (0.19.2)
Requirement already satisfied: matplotlib-inline in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (0.1.7)
Requirement already satisfied: pexpect<4.3 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (4.9.0)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (3.0.50)
Requirement already satisfied: pygments>=2.4.0 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (2.19.1)
Requirement already satisfied: pexpect==4.3 in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (0.1.0)
Requirement already satisfied: traitlets>=5.13.0 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (5.14.3)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in /opt/conda/lib/python3.12/site-packages (from jedi==0.16->python->python-sql) (0.8.4)
Requirement already satisfied: ptyprocess==0.5 in /opt/conda/lib/python3.12/site-packages (from pexpect<4.3->python->python-sql) (0.7.0)
Requirement already satisfied: executing==1.2.0 in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (2.1.0)
Requirement already satisfied: asttokens>=2.1.0 in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (3.0.0)
Requirement already satisfied: pure_eval in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (0.2.3)
Collecting pandas
  Downloading pandas-2.2.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (89 kB)
Collecting numpy==1.26.0 (from pandas)
  Downloading numpy-2.2.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (62 kB)
Requirement already satisfied: python-dateutil==2.8.2 in /opt/conda/lib/python3.12/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz==2020.1 in /opt/conda/lib/python3.12/site-packages (from pandas) (2024.2)
Collecting tzdata==2022.7 (from pandas)
  Downloading tzdata-2025.1-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.12/site-packages (from python-dateutil==2.8.2->pandas) (1.17.0)
Downloading pandas-2.2.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.7 MB)
12.7/12.7 MB 165.5 MB/s eta 0:00:00
Downloading numpy-2.2.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (16.1 MB)
16.1/16.1 MB 176.7 MB/s eta 0:00:00
Downloading tzdata-2025.1-py2.py3-none-any.whl (346 kB)
Installing collected packages: tzdata, numpy, pandas
Successfully installed numpy-2.2.3 pandas-2.2.3 tzdata-2025.1
```

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 [19]: import sqlite3
import pandas

con = sqlite3.connect("FinalDB.db")
cur = con.cursor()
```

Load the SQL magic module

```
In [20]: !pip install ipython-sql
!load_ext sql

Requirement already satisfied: ipython-sql in /opt/conda/lib/python3.12/site-packages (0.5.0)
Requirement already satisfied: prettytable in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (3.15.1)
Requirement already satisfied: ipython in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (8.31.0)
Requirement already satisfied: sqlalchemy>=2.0 in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (2.0.37)
Requirement already satisfied: sqlparse in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (0.5.3)
Requirement already satisfied: six in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (1.17.0)
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.12/site-packages (from ipython-sql) (0.2.0)
Requirement already satisfied: greenlet==0.4.17 in /opt/conda/lib/python3.12/site-packages (from sqlalchemy==2.0->python-sql) (3.1.1)
Requirement already satisfied: typing-extensions==4.6.0 in /opt/conda/lib/python3.12/site-packages (from sqlalchemy==2.0->python-sql) (4.12.2)
Requirement already satisfied: decorator in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (5.1.1)
Requirement already satisfied: jedi==0.16 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (0.19.2)
Requirement already satisfied: matplotlib-inline in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (0.1.7)
Requirement already satisfied: pexpect<4.3 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (4.9.0)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (3.0.50)
Requirement already satisfied: pygments>=2.4.0 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (2.19.1)
Requirement already satisfied: stack-data in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (0.6.3)
Requirement already satisfied: traitlets>=5.13.0 in /opt/conda/lib/python3.12/site-packages (from ipython->python-sql) (5.14.3)
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.12/site-packages (from prettytable->python-sql) (0.2.13)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in /opt/conda/lib/python3.12/site-packages (from jedi==0.16->python->python-sql) (0.8.4)
Requirement already satisfied: ptyprocess==0.5 in /opt/conda/lib/python3.12/site-packages (from pexpect<4.3->python->python-sql) (0.7.0)
Requirement already satisfied: executing==1.2.0 in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (2.1.0)
Requirement already satisfied: asttokens>=2.1.0 in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (3.0.0)
Requirement already satisfied: pure_eval in /opt/conda/lib/python3.12/site-packages (from stack-data->python->python-sql) (0.2.3)
The sql extension is already loaded. To reload it, use:
%reload_ext sql

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.
```

```
In [21]: df = pandas.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")
df2 = pandas.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")
df3 = pandas.read_csv("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")

df.to_sql("CENSUS", con, if_exists='replace', index=False, method='multi')
df2.to_sql("SCHOOLS", con, if_exists='replace', index=False, method='multi')
df3.to_sql("CRIME", con, if_exists='replace', index=False, method='multi')
```

Out [21]: 533

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

```
In [22]: %sql sqlite:///FinalDB.db
```

You can now proceed to 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 [23]: %sql SELECT COUNT(ID) FROM CRIME

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

Out [23]:

```
COUNT(ID)
533
```

Problem 2

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

```
In [24]: %sql SELECT COMMUNITY_AREA_NAME, COMMUNITY_AREA_NUMBER FROM CENSUS WHERE PER_CAPITA_INCOME < 11000

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

Out [24]:

```
COMMUNITY_AREA_NAME COMMUNITY_AREA_NUMBER
West Garfield Park 26.0
South Lawndale 30.0
Fuller Park 37.0
Riverdale 54.0
```

Problem 3

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

```
In [25]: %sql select * from crime where PRIMARY_TYPE = 'OFFENSE INVOLVING CHILDREN'

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

Out [25]:

```
ID CASE_NUMBER DATE BLOCK IUCR PRIMARY_TYPE DESCRIPTION LOCATION_DESCRIPTION ARREST DOMESTIC BEAT DISTRICT WARD COMMUNITY_AREA_NUMBER FBICODE X_COORDINATE Y_COORDINATE YEAR LATITUDE LONG
```

```
5768654 HNS67387 2007-09-02 026XW BLMONT AVE 1754 OFFENSE INVOLVING CHILDREN AGG SEX ASLT OF CHILD FAM MBR CHURCH/SYNAGOGUE/PLACE OF WORSHIP 0 0 1411 14 1.0 21.0 2 1158166.0 1921161.0 2007 41.93941471 -87.69
```

```
6986273 HR391350 2009-06-23 015X S KOHLIN AVE 1753 OFFENSE INVOLVING CHILDREN SEX ASLT OF CHILD BY FAM MBR APARTMENT 0 1 1012 10 24.0 29.0 2 1147638.0 1892092.0 2009 41.85985521 -87.73
```

```
5176248 HM768251 2006-12-11 076X S CAMPBELL AVE 1751 OFFENSE INVOLVING CHILDREN CRIM SEX ABUSE BY FAM MEMBER RESIDENCE 1 0 835 8 18.0 70.0 20 1161039.0 1853860.0 2006 41.75467441 -87.6
```

```
8159639 HT394616 2011-07-01 077X S SOUTH SHORE DR 1752 OFFENSE INVOLVING CHILDREN AGG CRIM SEX ABUSE FAM MEMBER RESIDENCE 0 0 421 4 7.0 43.0 20 1197175.0 1854844.0 2011 41.76655146 -87.55
```

Problem 4

List all kidnapping crimes involving a child?

```
In [26]: %sql select * from crime where PRIMARY_TYPE = 'KIDNAPPING' and DESCRIPTION LIKE '%CHILD%'

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

Out [26]:

```
ID CASE_NUMBER DATE BLOCK IUCR PRIMARY_TYPE DESCRIPTION LOCATION_DESCRIPTION ARREST DOMESTIC BEAT DISTRICT WARD COMMUNITY_AREA_NUMBER FBICODE X_COORDINATE Y_COORDINATE YEAR LATITUDE LONG
```

```
5276766 HNT44162 2007-01-26 050XW WIAN BUREN ST 1792 KIDNAPPING ABDUCTION/STRANGER CHILD STREET 0 0 1533 15 29.0 25.0 20 1143050.0 1897546.0 2007 41.87490841 -87.3
```

Problem 5

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

```
In [27]: %sql select distinct(PRIMARY_TYPE) as type_crimes_at_schools from crime where LOCATION_DESCRIPTION LIKE '%SCHOOL%'

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

Out [27]:

```
type_crimes_at_schools
BATTERY
CRIMINAL DAMAGE
NARCOTICS
ASSAULT
CRIMINAL TRESPASS
PUBLIC PEACE VIOLATION
```

Problem 6

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

```
In [28]: %sql select DISTINCT("Elementary, Middle, or High School"), avg(SAFETY_SCORE) from schools group by "Elementary, Middle, or High School"

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

Out [28]:

```
Elementary, Middle, or High School avg(SAFETY_SCORE)
ES 49.52038369304557
HS 49.6235294176471
MS 48.0
```

Problem 7

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

```
In [29]: %sql select COMMUNITY_AREA_NAME, PERCENT_HOUSEHOLDS_BELOW_POVERTY from census order by PERCENT_HOUSEHOLDS_BELOW_POVERTY desc limit 5

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

Out [29]:

```
COMMUNITY_AREA_NAME PERCENT_HOUSEHOLDS_BELOW_POVERTY
Riverdale 56.5
Fuller Park 51.2
Englewood 46.6
North Lawndale 43.1
East Garfield Park 42.4
```

Problem 8

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

```
In [46]: %sql select COMMUNITY_AREA_NUMBER, count(PRIMARY_TYPE) from crime group by COMMUNITY_AREA_NUMBER order by count(PRIMARY_TYPE) desc limit 1

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

Out [46]:

```
COMMUNITY_AREA_NUMBER count(PRIMARY_TYPE)
25.0 43
```

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 [36]: %sql select COMMUNITY_AREA_NAME, Hardship_index from census where Hardship_index = (select Hardship_index from census order by Hardship_index desc limit 1)

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

Out [36]:

```
COMMUNITY_AREA_NAME HARDSHIP_INDEX
Riverdale 98.0
```

Problem 10

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

```
In [45]: %sql select community_area_name, COMMUNITY_AREA_NUMBER from census where COMMUNITY_AREA_NUMBER = (select COMMUNITY_AREA_NUMBER from crime group by COMMUNITY_AREA_NUMBER order by count(PRIMARY_TYPE) desc limit 1)

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

Out [45]:

```
COMMUNITY_AREA_NAME COMMUNITY_AREA_NUMBER
Austin 25.0
```

Author(s)

Hima Vasudevan

Rav Ahuja

Ramesh Sanreddy

Contributor(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 Modified 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 Sanreddy 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 -->