Introduction

Using this Python notebook I 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 I 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=100
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=100
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=100
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.

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 [1]:
```
# Import necessary libraries
import pandas as pd
import sqlite3 as sq

# Create a connection to the SQLite database
conn = sq.connect('FinalDB.db')

# Create a cursor object
cursor = conn.cursor()
```

Load the SQL magic module

In [2]:
```
%load_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 [3]:
```
# URLs for the datasets
census_data_url = 'https://cf-courses-data.s3.us.cloud-object-storage.appdom
schools_data_url = 'https://cf-courses-data.s3.us.cloud-object-storage.appdc
crime_data_url = 'https://cf-courses-data.s3.us.cloud-object-storage.appdoma

# Load the datasets into pandas DataFrames
census_df = pd.read_csv(census_data_url)
schools_df = pd.read_csv(schools_data_url)
crime_df = pd.read_csv(crime_data_url)

# Display the first few rows of each dataframe to confirm successful load
print(census_df.head())
```

```
print(schools_df.head())
print(crime_df.head())
```

```
   COMMUNITY_AREA_NUMBER COMMUNITY_AREA_NAME  PERCENT_OF_HOUSING_CROWDED  \
0                    1.0         Rogers Park                         7.7
1                    2.0          West Ridge                         7.8
2                    3.0              Uptown                         3.8
3                    4.0      Lincoln Square                         3.4
4                    5.0        North Center                         0.3

   PERCENT_HOUSEHOLDS_BELOW_POVERTY  PERCENT_AGED_16__UNEMPLOYED  \
0                              23.6                         8.7
1                              17.2                         8.8
2                              24.0                         8.9
3                              10.9                         8.2
4                               7.5                         5.2

   PERCENT_AGED_25__WITHOUT_HIGH_SCHOOL_DIPLOMA  \
0                                          18.2
1                                          20.8
2                                          11.8
3                                          13.4
4                                           4.5

   PERCENT_AGED_UNDER_18_OR_OVER_64  PER_CAPITA_INCOME  HARDSHIP_INDEX
0                              27.5              23939            39.0
1                              38.5              23040            46.0
2                              22.2              35787            20.0
3                              25.5              37524            17.0
4                              26.2              57123             6.0
   School_ID                                    NAME_OF_SCHOOL  \
0    610038                  Abraham Lincoln Elementary School
1    610281  Adam Clayton Powell Paideia Community Academy ...
2    610185                Adlai E Stevenson Elementary School
3    609993                   Agustin Lara Elementary Academy
4    610513                     Air Force Academy High School

  Elementary, Middle, or High School         Street_Address     City State  \
0                                 ES        615 W Kemper Pl  Chicago    IL
1                                 ES  7511 S South Shore Dr  Chicago    IL
2                                 ES     8010 S Kostner Ave  Chicago    IL
3                                 ES     4619 S Wolcott Ave  Chicago    IL
4                                 HS        3630 S Wells St  Chicago    IL

   ZIP_Code    Phone_Number  \
0     60614  (773) 534-5720
1     60649  (773) 535-6650
2     60652  (773) 535-2280
3     60609  (773) 535-4389
4     60609  (773) 535-1590

                                              Link  \
0  http://schoolreports.cps.edu/SchoolProgressRep...
1  http://schoolreports.cps.edu/SchoolProgressRep...
2  http://schoolreports.cps.edu/SchoolProgressRep...
3  http://schoolreports.cps.edu/SchoolProgressRep...
4  http://schoolreports.cps.edu/SchoolProgressRep...

                Network_Manager  ... Freshman_on_Track_Rate__  \
```

```
0        Fullerton Elementary Network  ...                    NDA
1           Skyway Elementary Network  ...                    NDA
2           Midway Elementary Network  ...                    NDA
3         Pershing Elementary Network  ...                    NDA
4  Southwest Side High School Network  ...                   91.8

  X_COORDINATE Y_COORDINATE   Latitude  Longitude COMMUNITY_AREA_NUMBER  \
0  1171699.458  1915829.428  41.924497 -87.644522                    7
1  1196129.985  1856209.466  41.760324 -87.556736                   43
2  1148427.165  1851012.215  41.747111 -87.731702                   70
3  1164504.290  1873959.199  41.809757 -87.672145                   61
4  1175177.622  1880745.126  41.828146 -87.632794                   34

  COMMUNITY_AREA_NAME  Ward Police_District                    Location
0        LINCOLN PARK    43              18  (41.92449696, -87.64452163)
1         SOUTH SHORE     7               4  (41.76032435, -87.55673627)
2             ASHBURN    13               8  (41.74711093, -87.73170248)
3            NEW CITY    20               9    (41.8097569, -87.6721446)
4       ARMOUR SQUARE    11               9  (41.82814609, -87.63279369)

[5 rows x 78 columns]
         ID CASE_NUMBER        DATE                    BLOCK IUCR  \
0   3512276    HK587712  2004-08-28        047XX S KEDZIE AVE  890
1   3406613    HK456306  2004-06-26  009XX N CENTRAL PARK AVE  820
2   8002131    HT233595  2011-04-04        043XX S WABASH AVE  820
3   7903289    HT133522  2010-12-30      083XX S KINGSTON AVE  840
4  10402076    HZ138551  2016-02-02           033XX W 66TH ST  820

  PRIMARY_TYPE                    DESCRIPTION         LOCATION_DESCRIPTION
\
0        THEFT                  FROM BUILDING           SMALL RETAIL STORE
1        THEFT                 $500 AND UNDER                       OTHER
2        THEFT                 $500 AND UNDER  NURSING HOME/RETIREMENT HOME
3        THEFT  FINANCIAL ID THEFT: OVER $300                   RESIDENCE
4        THEFT                 $500 AND UNDER                      ALLEY

   ARREST  DOMESTIC  ...  DISTRICT  WARD  COMMUNITY_AREA_NUMBER  FBICODE  \
0   False     False  ...         9  14.0                  58.0        6
1   False     False  ...        11  27.0                  23.0        6
2   False     False  ...         2   3.0                  38.0        6
3   False     False  ...         4   7.0                  46.0        6
4   False     False  ...         8  15.0                  66.0        6

  X_COORDINATE  Y_COORDINATE  YEAR   LATITUDE  LONGITUDE  \
0    1155838.0     1873050.0  2004  41.807440 -87.703956
1    1152206.0     1906127.0  2004  41.898280 -87.716406
2    1177436.0     1876313.0  2011  41.815933 -87.624642
3    1194622.0     1850125.0  2010  41.743665 -87.562463
4    1155240.0     1860661.0  2016  41.773455 -87.706480

                      LOCATION
0    (41.8074405, -87.703955849)
1  (41.898279962, -87.716405505)
2  (41.815933131, -87.624642127)
3  (41.743665322, -87.562462756)
4  (41.773455295, -87.706480471)
```

```
[5 rows x 21 columns]
```

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

In [12]:
```
# Connect to the SQLite database 'FinalDB.db'
conn = sq.connect('FinalDB.db')

# Create a cursor object to interact with the database
cursor = conn.cursor()

# Load the dataframes into the database as tables
census_df.to_sql('CENSUS_DATA', conn, if_exists='replace', index=False)
schools_df.to_sql('CHICAGO_PUBLIC_SCHOOLS', conn, if_exists='replace', index
crime_df.to_sql('CHICAGO_CRIME_DATA', conn, if_exists='replace', index=False

# Commit the transaction and close the connection
conn.commit()

print("Data successfully loaded into FinalDB.db")
```

```
Data successfully loaded into FinalDB.db
```

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 [13]:
```
%%sql
SELECT COUNT(*) AS total_crimes
FROM CHICAGO_CRIME_DATA;
```

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

Out[13]:

total_crimes
533

Problem 2

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

In [14]:
```
%%sql
ITY_AREA_NUMBER, COMMUNITY_AREA_NAME
```

```
FROM CENSUS_DATA
WHERE PER_CAPITA_INCOME < 11000;
```

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

Out[14]:

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

Problem 3

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

In [15]:

```
%%sql
SELECT CASE_NUMBER
FROM CHICAGO_CRIME_DATA
WHERE DESCRIPTION LIKE '%MINOR%';
```

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

Out[15]:

CASE_NUMBER
HL266884
HK238408

Problem 4

List all kidnapping crimes involving a child?

In [16]:

```
%%sql
SELECT *
FROM CHICAGO_CRIME_DATA
WHERE PRIMARY_TYPE = 'KIDNAPPING'
AND DESCRIPTION LIKE '%CHILD%';
```

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

Out[16]:

ID	CASE_NUMBER	DATE	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION
5276766	HN144152	2007-01-26	050XX W VAN BUREN ST	1792	KIDNAPPING	CHILD ABDUCTION/STRANGER

Problem 5

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

In [17]:
```
%%sql
SELECT DISTINCT PRIMARY_TYPE
FROM CHICAGO_CRIME_DATA
WHERE LOCATION_DESCRIPTION = 'SCHOOL';
```

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

Out[17]:

PRIMARY_TYPE

Problem 6

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

In [18]:
```
%%sql
SELECT "Elementary, Middle, or High School", AVG(SAFETY_SCORE) AS avg_safety
FROM CHICAGO_PUBLIC_SCHOOLS
GROUP BY "Elementary, Middle, or High School";
```

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

Out[18]:

Elementary, Middle, or High School	avg_safety_score
ES	49.52038369304557
HS	49.62352941176471
MS	48.0

Problem 7

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

In [19]:
```
%%sql
SELECT COMMUNITY_AREA_NAME
FROM CENSUS_DATA
ORDER BY PERCENT_HOUSEHOLDS_BELOW_POVERTY DESC
LIMIT 5;
```

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

Out[19]:

COMMUNITY_AREA_NAME
Riverdale
Fuller Park
Englewood
North Lawndale
East Garfield Park

Problem 8

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

In [20]:
```
%%sql
SELECT COMMUNITY_AREA_NUMBER
FROM CHICAGO_CRIME_DATA
GROUP BY COMMUNITY_AREA_NUMBER
ORDER BY COUNT(*) DESC
LIMIT 1;
```

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

Out[20]:

COMMUNITY_AREA_NUMBER
25.0

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 [21]:
```
%%sql
SELECT COMMUNITY_AREA_NAME
FROM CENSUS_DATA
WHERE HARDSHIP_INDEX = (SELECT MAX(HARDSHIP_INDEX) FROM CENSUS_DATA);
```

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

Out[21]:

COMMUNITY_AREA_NAME
Riverdale

Problem 10

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

In [22]:
```
%%sql
SELECT COMMUNITY_AREA_NAME
FROM CENSUS_DATA
WHERE COMMUNITY_AREA_NUMBER = (
    SELECT COMMUNITY_AREA_NUMBER
    FROM CHICAGO_CRIME_DATA
    GROUP BY COMMUNITY_AREA_NUMBER
    ORDER BY COUNT(*) DESC
    LIMIT 1
);
```

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

Out[22]:

COMMUNITY_AREA_NAME
Austin