Assignment 3

(Using DB SQL Various Statements to Retrieve Crime Data)

CSCE 4357/5933: Database Systems Security

(Spring 2023)

**Due Date: 04/26 @ 11:59pm | Points:** **100**  **| Late Policy: 10%**

**NOTE: Please write everything by your own and do not copy/paste, avoid plagiarism.**

### **Introduction**

Using this Python notebook you will:

1. Understand three Chicago datasets
2. Load the three datasets into three tables in a Db2 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](https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
2. [Chicago Public Schools](https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
3. [Chicago Crime Data](https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)

### **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 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](https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01&cm_mmc=Email_Newsletter-_-Developer_Ed%2BTech-_-WW_WW-_-SkillsNetwork-Courses-IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork-20127838&cm_mmca1=000026UJ&cm_mmca2=10006555&cm_mmca3=M12345678&cvosrc=email.Newsletter.M12345678&cvo_campaign=000026UJ)

### **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 can be found from the Chicago Data Portal at: [https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t](https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01&cm_mmc=Email_Newsletter-_-Developer_Ed%2BTech-_-WW_WW-_-SkillsNetwork-Courses-IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork-20127838&cm_mmca1=000026UJ&cm_mmca2=10006555&cm_mmca3=M12345678&cvosrc=email.Newsletter.M12345678&cvo_campaign=000026UJ)

### **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](https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01&cm_mmc=Email_Newsletter-_-Developer_Ed%2BTech-_-WW_WW-_-SkillsNetwork-Courses-IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork-20127838&cm_mmca1=000026UJ&cm_mmca2=10006555&cm_mmca3=M12345678&cvosrc=email.Newsletter.M12345678&cvo_campaign=000026UJ)

### **Download the Datasets**

**The assignment requires you to have 3 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. Click on the links below to download and save the datasets (.CSV files):

* [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-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
* [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-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
* [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-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)

**NOTE**: For the learners who are encountering issues with loading from **.csv** in **DB2** on **Firefox**, you can download the .**txt** files and load the data with those:

* [Chicago Census Data](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCensusData.txt?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
* [Chicago Public Schools](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoPublicSchools.txt?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)
* [Chicago Crime Data](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCrimeData.txt?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkDB0201ENSkillsNetwork22-2022-01-01)

**NOTE:** Ensure you have downloaded the datasets using the links above instead of directly from the Chicago Data Portal. 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 for this assignment.

### **To Store the CSV Datasets in Database Tables**

This task you have already completed in the class.

### **Connect to the Database**

This task you have already completed in the class connecting from Jupiter to DB directly.

**Graphical user interface, text, application

Description automatically generated**

## **Problems**: Now Write and Execute SQL Queries to Solve Assignment Problems

### **[10pts] Problem 1**

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

**SELECT count(\*) FROM assignment\_3.chicagocrimedata;**

Python Query:

|  |
| --- |
| totalnumberofcrimes = panda.read\_sql\_query("SELECT count(\*) FROM chicagocrimedata",conn)  print(totalnumberofcrimes) |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 2**

* List community areas with per capita income less than **11000**.

**SELECT COMMUNITY\_AREA\_NAME,PER\_CAPITA\_INCOME FROM assignment\_3.chicagocensusdata where PER\_CAPITA\_INCOME <11000;**

Python Query:

|  |
| --- |
| areas = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NAME,PER\_CAPITA\_INCOME FROM chicagocensusdata where PER\_CAPITA\_INCOME <11000 ",conn)  areas.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 3**

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

**SELECT DISTINCT CASE\_NUMBER,DESCRIPTION FROM assignment\_3.chicagocrimedata WHERE DESCRIPTION LIKE '%MINOR%';**

Python Query:

|  |
| --- |
| totalnumberofcrimes = panda.read\_sql\_query("SELECT DISTINCT CASE\_NUMBER, DESCRIPTION FROM chicagocrimedata WHERE DESCRIPTION LIKE '%MINOR%'",conn)  totalnumberofcrimes.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 4**

##### List all kidnapping crimes involving a child?

**SELECT DISTINCT CASE\_NUMBER, PRIMARY\_TYPE, DATE, DESCRIPTION FROM assignment\_3.chicagocrimedata WHERE PRIMARY\_TYPE='KIDNAPPING' and DESCRIPTION LIKE '%CHILD%';**

Python Query:

|  |
| --- |
| kidnapping = panda.read\_sql\_query("SELECT DISTINCT CASE\_NUMBER, PRIMARY\_TYPE,DESCRIPTION FROM chicagocrimedata WHERE PRIMARY\_TYPE='KIDNAPPING' and DESCRIPTION LIKE '%CHILD%' ",conn)  kidnapping.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 5**

##### What kinds of crimes were recorded at schools?

**SELECT DISTINCT PRIMARY\_TYPE FROM assignment\_3.chicagocrimedata WHERE LOCATION\_DESCRIPTION LIKE '%SCHOOL%';**

Python Query:

|  |
| --- |
| school = panda.read\_sql\_query("SELECT DISTINCT PRIMARY\_TYPE FROM chicagocrimedata WHERE LOCATION\_DESCRIPTION LIKE '%SCHOOL%'",conn)  school.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 6**

For this problem, the original column name in the dataset was not as per the standards, so I have changed the column name from **Elementary or Middle or High School** to **Elementary\_Middle\_High\_School**

##### List the average safety score for each type of school.

**SELECT Elementary\_Middle\_High\_School, AVG(SAFETY\_SCORE) AVERAGE\_SAFETY\_SCORE FROM assignment\_3.chicagopublicschools group by Elementary\_Middle\_High\_School;**

Python Query:

|  |
| --- |
| school = panda.read\_sql\_query("SELECT Elementary\_Middle\_High\_School, AVG(SAFETY\_SCORE) AVERAGE\_SAFETY\_SCORE FROM chicagopublicschools group by Elementary\_Middle\_High\_School",conn)  school.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 7**

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

**SELECT COMMUNITY\_AREA\_NAME, PERCENT\_HOUSEHOLDS\_BELOW\_POVERTY FROM assignment\_3.chicagocensusdata ORDER BY PERCENT\_HOUSEHOLDS\_BELOW\_POVERTY DESC LIMIT 5;**

Python Query:

|  |
| --- |
| poverty = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NAME, PERCENT\_HOUSEHOLDS\_BELOW\_POVERTY FROM chicagocensusdata ORDER BY PERCENT\_HOUSEHOLDS\_BELOW\_POVERTY DESC LIMIT 5 ;",conn)  poverty.head() |

**Output**: Please attach the screenshot of the output below.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 8**

##### Which community area is most crime prone?

**SELECT COMMUNITY\_AREA\_NUMBER, COUNT(COMMUNITY\_AREA\_NUMBER) FROM assignment\_3.chicagocrimedata GROUP BY COMMUNITY\_AREA\_NUMBER ORDER BY COUNT(COMMUNITY\_AREA\_NUMBER) DESC LIMIT 1;**

Python Query:

|  |
| --- |
| crimeprone = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NUMBER, COUNT(COMMUNITY\_AREA\_NUMBER) FROM chicagocrimedata GROUP BY COMMUNITY\_AREA\_NUMBER ORDER BY COUNT(COMMUNITY\_AREA\_NUMBER) DESC LIMIT 1",conn)  crimeprone.head()hardshipindex = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NAME, HARDSHIP\_INDEX FROM chicagocensusdata WHERE HARDSHIP\_INDEX = (SELECT MAX(HARDSHIP\_INDEX) FROM chicagocensusdata)",conn)  hardshipindex.head() |

**Output**: Please attach the screenshot of the output below.



Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 9**

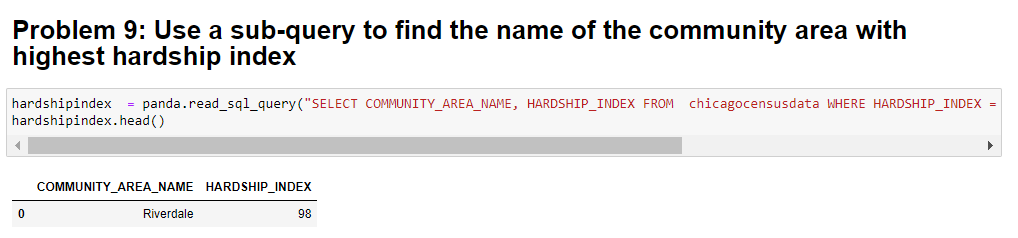
##### Use a sub-query to find the name of the community area with highest hardship index

**SELECT COMMUNITY\_AREA\_NAME, HARDSHIP\_INDEX FROM assignment\_3.chicagocensusdata WHERE HARDSHIP\_INDEX = (SELECT MAX(HARDSHIP\_INDEX) FROM assignment\_3.chicagocensusdata);**

Python Query:

|  |
| --- |
| hardshipindex = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NAME, HARDSHIP\_INDEX FROM chicagocensusdata WHERE HARDSHIP\_INDEX = (SELECT MAX(HARDSHIP\_INDEX) FROM chicagocensusdata)",conn)  hardshipindex.head() |

**Output**: Please attach the screenshot of the output below.



Graphical user interface, text, application, email

Description automatically generated

### **[10pts] Problem 10**

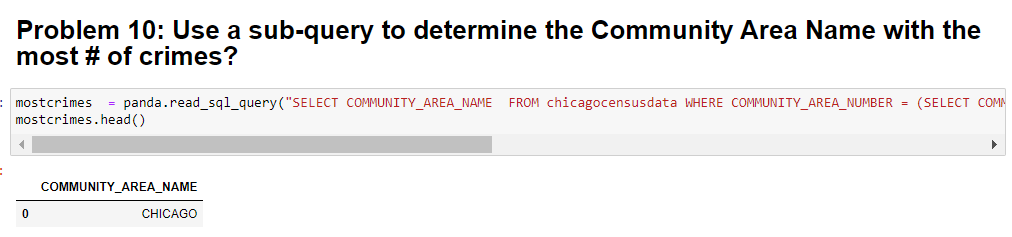
##### Use a sub-query to determine the Community Area Name with the most # of crimes?

**SELECT COMMUNITY\_AREA\_NAME FROM assignment\_3.chicagocensusdata WHERE COMMUNITY\_AREA\_NUMBER = (SELECT COMMUNITY\_AREA\_NUMBER FROM assignment\_3.chicagocrimedata GROUP BY COMMUNITY\_AREA\_NUMBER ORDER BY COUNT(COMMUNITY\_AREA\_NUMBER) DESC LIMIT 1) LIMIT 1;**

Python Query:

|  |
| --- |
| mostcrimes = panda.read\_sql\_query("SELECT COMMUNITY\_AREA\_NAME FROM chicagocensusdata WHERE COMMUNITY\_AREA\_NUMBER = (SELECT COMMUNITY\_AREA\_NUMBER FROM chicagocrimedata GROUP BY COMMUNITY\_AREA\_NUMBER ORDER BY COUNT(COMMUNITY\_AREA\_NUMBER) DESC LIMIT 1) LIMIT 1",conn)  mostcrimes.head() |

**Output**: Please attach the screenshot of the output below.



Graphical user interface, text, application, email

Description automatically generated