

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.

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`

```
import csv, sqlite3

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

```
!pip install -q pandas==1.1.5
```

```
%load_ext sql
```

The sql extension is already loaded. To reload it, use:
`%reload_ext sql`

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

```
import pandas
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")
df.to_sql("CENSUS_DATA", con, if_exists='replace',
index=False,method="multi")

df = pandas.read_csv("https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-
SkillsNetwork/labs/FinalModule_Coursera_V5/data/ChicagoCrimeData.csv")
df.to_sql("CHICAGO_CRIME_DATA", con, if_exists='replace', index=False,
method="multi")

df = pandas.read_csv("https://cf-courses-data.s3.us.cloud-object-
storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB0201EN-
SkillsNetwork/labs/FinalModule_Coursera_V5/data/
ChicagoPublicSchools.csv")
df.to_sql("CHICAGO_PUBLIC_SCHOOLS_DATA", con, if_exists='replace',
index=False, method="multi")

/home/jupyterlab/conda/envs/python/lib/python3.7/site-packages/
pandas/core/generic.py:2882: UserWarning: The spaces in these column
names will not be changed. In pandas versions < 0.14, spaces were
converted to underscores.
    method=method,
```

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

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

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

[('CENSUS_DATA',), ('CHICAGO_CRIME_DATA',),
 ('CHICAGO_PUBLIC_SCHOOLS_DATA',)]
```

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.

```
%sql SELECT COUNT(*) AS TOTAL_CRIMES \
      FROM CHICAGO_CRIME_DATA

* sqlite:///RealWorldData.db
Done.

[(533,)]
```

Problem 2

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

```
%sql select community_area_name from CENSUS_DATA Where
per_capita_income < 11000

* sqlite:///RealWorldData.db
Done.

[('West Garfield Park',),
 ('South Lawndale',),
 ('Fuller Park',),
 ('Riverdale',)]
```

Problem 3

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

```
%%sql
select
  id,
  case_number,
```

```

    primary_type,
    description
from
    CHICAGO_CRIME_DATA
where
    description like '%minor%'

* sqlite:///RealWorldData.db
Done.

[(3987219, 'HL266884', 'LIQUOR LAW VIOLATION', 'SELL/GIVE/DEL LIQUOR
TO MINOR'),
 (3266814, 'HK238408', 'LIQUOR LAW VIOLATION', 'ILLEGAL CONSUMPTION BY
MINOR')]

```

Problem 4

List all kidnapping crimes involving a child?

```

%sql select * FROM CHICAGO_CRIME_DATA WHERE primary_type like '%KIDNAP
%' AND description like '%CHILD%'

* sqlite:///RealWorldData.db
Done.

[(5276766, 'HN144152', '2007-01-26', '050XX W VAN BUREN ST', '1792',
'KIDNAPPING', 'CHILD ABDUCTION/STRANGER', 'STREET', 0, 0, 1533, 15,
29.0, 25.0, '20', 1143050.0, 1897546.0, 2007, 41.87490841, -
87.75024931, '(41.874908413, -87.750249307)')]

```

Problem 5

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

```

%sql SELECT distinct(primary_type) from CHICAGO_CRIME_DATA where
Location_description like 'SCHOOL%'

* sqlite:///RealWorldData.db
Done.

[('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.

```
%%sql select "Elementary, Middle or High School" ,AVG(safety_score)
from CHICAGO_PUBLIC_SCHOOLS_DATA
GROUP BY "Elementary, Middle or High School"

* sqlite:///RealWorldData.db
Done.

[('Elementary, Middle or High School', 49.50487329434698)]
```

Problem 7

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

```
%%sql SELECT community_area_name ,percent_households_below_poverty
FROM census_data
ORDER BY percent_households_below_poverty DESC Limit 5

* sqlite:///RealWorldData.db
Done.

[('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.

```
%%sql select community_area_number, count(*) as frequency from
CHICAGO_CRIME_DATA
GROUP BY community_area_number order by frequency DESC LIMIT 1

* sqlite:///RealWorldData.db
Done.

[(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

```
%%sql
select community_area_name from CENSUS_DATA
where hardship_index = (select max(hardship_index) from CENSUS_DATA)
```

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

[('Riverdale',)]
```

Problem 10

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

```
%%sql
select community_area_name from CENSUS_DATA where commun
```

Author(s)

Contributor(s)

© IBM Corporation 2023. All rights reserved.