Skills Network

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-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=
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=
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=
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 pandas as pd
        import sqlite3

        conn = sqlite3.connect('FinalDB.db')
```

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 [6]: # Load datasets into pandas DataFrames
        census_data = pd.read_csv('ChicagoCensusData.csv')
        public_schools = pd.read_csv('ChicagoPublicSchools.csv')
        crime_data = pd.read_csv('ChicagoCrimeData.csv')

        # Load DataFrames into the database
        census_data.to_sql('CENSUS_DATA', conn, if_exists='replace', index=False)
        public_schools.to_sql('CHICAGO_PUBLIC_SCHOOLS', conn, if_exists='replace', index=False)
        crime_data.to_sql('CHICAGO_CRIME_DATA', conn, if_exists='replace', index=False)
```

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

```
In [7]: %sql sqlite:///FinalDB.db
```

```
Out[7]: 'Connected: @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 [9]: %sql SELECT COUNT(*) AS TOTAL_CRIME FROM CHICAGO_CRIME_DATA
```

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

TOTAL_CRIME

533

Problem 2

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

In [10]: `%sql SELECT community_area_name, community_area_number FROM CENSUS_DATA WHERE per_capita_income`

* sqlite:///FinalDB.db
Done.

Out[10]:

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 [11]: `%sql SELECT case_number FROM CHICAGO_CRIME_DATA WHERE description LIKE '%CHILD%';`

* sqlite:///FinalDB.db
Done.

Out[11]:

CASE_NUMBER
HN567387
HR391350
HN144152

Problem 4

List all kidnapping crimes involving a child?

In [12]: `%sql SELECT case_number, primary_type, description FROM CHICAGO_CRIME_DATA WHERE primary_type =`

* sqlite:///FinalDB.db
Done.

Out[12]:

CASE_NUMBER	PRIMARY_TYPE	DESCRIPTION
HN144152	KIDNAPPING	CHILD ABDUCTION/STRANGER

Problem 5

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

In [13]: `%sql SELECT DISTINCT(primary_type) FROM CHICAGO_CRIME_DATA WHERE location_description LIKE '%SCH`

Out[13]:

PRIMARY_TYPE
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 [14]:
```
%sql SELECT "Elementary, Middle, or High School", AVG(safety_score) as average_safety_score FROM
```

* sqlite:///FinalDB.db
Done.

Out[14]:

Elementary, Middle, or High School	average_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 [15]:
```
%sql SELECT community_area_name FROM CENSUS_DATA ORDER BY percent_households_below_poverty DESC
```

* sqlite:///FinalDB.db
Done.

Out[15]:

COMMUNITY_AREA_NAME
Riverdale
Fuller Park
Englewood
North Lawndale
East Garfield Park

Problem 8

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

In [17]:
```
%sql SELECT community_area_number, COUNT(*) AS number_of_crimes FROM CHICAGO_CRIME_DATA GROUP BY
```

* sqlite:///FinalDB.db
Done.

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 [18]: `%sql` `SELECT community_area_name FROM CENSUS_DATA WHERE hardship_index = (SELECT MAX(hardship_inde`

* sqlite:///FinalDB.db
Done.

Out[18]:

COMMUNITY_AREA_NAME
Riverdale

Problem 10

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

In [19]: `%sql` `SELECT CD.community_area_name FROM CENSUS_DATA CD JOIN CHICAGO_CRIME_DATA CCD ON CD.communit`

* sqlite:///FinalDB.db
Done.

Out[19]:

COMMUNITY_AREA_NAME
Austin

In [20]:
```
# Close connection for the database
conn.close()
```

Author(s)

Hima Vasudevan

Rav Ahuja

Ramesh Sannreddy

Contribtuor(s)

Malika Singla

Abhishek Gagneja

Change log

Date	Version	Changed by	Change Description
2023-10-18	2.6	Abhishek Gagneja	Modified instruction set

Date	Version	Changed by	Change Description
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