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. Click on the links below to download and save the datasets (.CSV files):

- Chicago Census Data
- Chicago Public Schools
- Chicago Crime Data

NOTE: Ensure you have downloaded the datasets using 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**

Let us now load the ipython-sql extension and establish a connection with the database

- Here you will be loading the csv files into the pandas Dataframe and then loading the data into the above mentioned sqlite tables.
- Next you will be connecting to the sqlite database **FinalDB**.

Refer to the previous lab for hints .

Hands-on Lab: Analyzing a real World Data Set

```
%load_ext sql
```

```
import csv, sqlite3

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

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

```
     |████████████████████████████████| 9.5 MB 17.4 MB/s
```

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



```

```
'Connected: @RealWorldData.db'
```

```
import pandas
df = pandas.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DB020
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-DB020
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-DB020
df.to_sql("CHICAGO_PUBLIC_SCHOOLS_DATA", con, if_exists='replace', index=False, method="multi")
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/generic.py:2882: UserWarning: The spaces in these column names wi
  both result in 0.1234 being formatted as 0.12.
```

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(*)
from
  CHICAGO_CRIME_DATA
```

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

count(*)
533

Problem 2

List community areas with per capita income less than 11000.

```
%%sql
select
 community_area_number,
 community_area_name,
 per_capita_income
from
  CENSUS_DATA
where
  per_capita_income < 11000
```

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

COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME	PER_CAPITA_INCOME
26.0	West Garfield Park	10934
30.0	South Lawndale	10402
37.0	Fuller Park	10432
54.0	Riverdale	8201

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.
```

ID	CASE_NUMBER	PRIMARY_TYPE	DESCRIPTION
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
  id,
  case_number,
  primary_type,
  description
from
  CHICAGO_CRIME_DATA
where
  primary_type like '%kidnapping%'
  and
  description like '%child%'
```

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

ID	CASE_NUMBER	PRIMARY_TYPE	DESCRIPTION
5276766	HN144152	KIDNAPPING	CHILD ABDUCTION/STRANGER

Problem 5

What kinds of crimes were recorded at schools?

```
%%sql
select
  distinct primary_type,
  /* description ,*/
  location_description
from
  CHICAGO_CRIME_DATA
where
  location_description like '%school%'
order by
  primary_type
```

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

PRIMARY_TYPE	LOCATION_DESCRIPTION
ASSAULT	SCHOOL, PUBLIC, GROUNDS
BATTERY	SCHOOL, PUBLIC, GROUNDS
BATTERY	SCHOOL, PUBLIC, BUILDING
CRIMINAL DAMAGE	SCHOOL, PUBLIC, GROUNDS
CRIMINAL TRESPASS	SCHOOL, PUBLIC, GROUNDS
NARCOTICS	SCHOOL, PUBLIC, GROUNDS
NARCOTICS	SCHOOL, PUBLIC, BUILDING
PUBLIC PEACE VIOLATION	SCHOOL, PRIVATE, BUILDING
PUBLIC PEACE VIOLATION	SCHOOL, PUBLIC, BUILDING

Problem 6

List the average safety score for each type of school.

```
%%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	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

```
%%sql
select
  community_area_name,
  max(percent_households_below_poverty) as highest_percent_households_below_poverty
from
  census_data
where
  percent_households_below_poverty is not null
group by
  community_area_name
order by
  highest_percent_households_below_poverty desc
limit 5
```

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

COMMUNITY_AREA_NAME	highest_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?

```
%%sql
select
  community_area_number,
  count(community_area_number) as CrimeNumber
from
  CHICAGO_CRIME_DATA
where
  community_area_number is not null
group by
  community_area_number
order by
  crimenumber desc
limit 1
```

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

COMMUNITY_AREA_NUMBER	CrimeNumber
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_number,
  community_area_name,
  hardship_index
from
  census_data
where
  hardship_index = (
    select max(hardship_index)
    from census_data
  )
```

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

COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME	HARDSHIP_INDEX
54.0	Riverdale	98.0

Problem 10

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

```
%%sql
select
  distinct ccd.community_area_number, cd.community_area_name
from
  CHICAGO_CRIME_DATA ccd
join
  census_data cd
on
  ccd.community_area_number = cd.community_area_number
where
  ccd.community_area_number = (
    select community_area_number
    from CHICAGO_CRIME_DATA
    where community_area_number is not null
    group by community_area_number
    order by count(community_area_number) desc
    limit 1
  )
```

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

COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME
25.0	Austin

Author(s)

Hima Vasudevan