DB0201EN-Week4-2-2-PeerAssign-v5-py

February 15, 2019

Assignment: Notebook for Peer Assignment

1 Introduction

Using this Python notebook you will: 1. Understand 3 Chicago datasets
1. Load the 3 datasets into 3 tables in a Db2 database 1. Execute SQL queries to answer assignment questions

1.1 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 1. Chicago Public Schools 1. Chicago Crime Data

1.1.1 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.

For this assignment you will use a snapshot of this dataset which can be downloaded from: https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv

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

1.1.2 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.

For this assignment you will use a snapshot of this dataset which can be downloaded from: https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhplzt01qtz0s7ew7.csv

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

1.1.3 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.

This dataset is quite large - over 1.5GB in size with over 6.5 million rows. For the purposes of this assignment we will use a much smaller sample of this dataset which can be downloaded from: https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv

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

1.1.4 Download the 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): 1. **CENSUS_DATA:** https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv 1. **CHICAGO_PUBLIC_SCHOOLS** https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhplzt01qtz0s7ew7.csv 1. **CHICAGO_CRIME_DATA:** https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv

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.

1.1.5 Store the datasets in database tables

To analyze the data using SQL, it first needs to be stored in the database.

While it is easier to read the dataset into a Pandas dataframe and then PERSIST it into the database as we saw in Week 3 Lab 3, it results in mapping to default datatypes which may not be optimal for SQL querying. For example a long textual field may map to a CLOB instead of a VARCHAR.

Therefore, **it is highly recommended to manually load the table using the database console LOAD tool, as indicated in Week 2 Lab 1 Part II**. The only difference with that lab is that in Step 5 of the instructions you will need to click on create "(+) New Table" and specify the name of the table you want to create and then click "Next".

Now open the Db2 console, open the LOAD tool, Select / Drag the .CSV file for the first dataset, Next create a New Table, and then follow the steps on-screen instructions to load the data. Name the new tables as folows:

1. **CENSUS_DATA**
2. **CHICAGO_PUBLIC_SCHOOLS**
3. **CHICAGO_CRIME_DATA**

1.1.6 Connect to the database

Let us first load the SQL extension and establish a connection with the database

```
In [1]: %load_ext sql
```

In the next cell enter your db2 connection string. Recall you created Service Credentials for your Db2 instance in first lab in Week 3. From the **uri** field of your Db2 service credentials copy everything after db2:// (except the double quote at the end) and paste it in the cell below after ibm_db_sa://

```
In [2]: # Remember the connection string is of the format:
        # %sql ibm_db_sa://my-username:my-password@my-hostname:my-port/my-db-name
        # Enter the connection string for your Db2 on Cloud database instance below
        %sql ibm_db_sa://ttk07945:kk41nf3cg7lr9s-7@dashdb-txn-sbox-yp-dal09-04.services.dal.blue
```

```
Out[2]: 'Connected: ttk07945@BLUDB'
```

CHICAGO_CRIME_DATA

```
In [3]: # type in your query to retrieve all column names in the CHICAGO_CRIME_DATA table along
        %sql select distinct(name), coltype, length from sysibm.syscolumns where tbname = 'CHICA
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[3]: [('ARREST', 'VARCHAR ', 5),
         ('BEAT', 'SMALLINT', 2),
         ('BLOCK', 'VARCHAR ', 35),
         ('CASE_NUMBER', 'VARCHAR ', 8),
         ('COMMUNITY_AREA_NUMBER', 'SMALLINT', 2),
         ('DATE', 'VARCHAR ', 22),
         ('DESCRIPTION', 'VARCHAR ', 46),
         ('DISTRICT', 'SMALLINT', 2),
         ('DOMESTIC', 'VARCHAR ', 5),
         ('FBICODE', 'VARCHAR ', 3),
         ('ID', 'INTEGER ', 4),
         ('IUCR', 'VARCHAR ', 4),
         ('LATITUDE', 'DECIMAL ', 18),
         ('LOCATION', 'VARCHAR ', 29),
         ('LOCATION_DESCRIPTION', 'VARCHAR ', 33),
         ('LONGITUDE', 'DECIMAL ', 18),
         ('PRIMARY_TYPE', 'VARCHAR ', 33),
         ('UPDATEDON', 'VARCHAR ', 22),
         ('WARD', 'SMALLINT', 2),
         ('X_COORDINATE', 'INTEGER ', 4),
         ('YEAR', 'SMALLINT', 2),
         ('Y_COORDINATE', 'INTEGER ', 4)]
```

```
In [4]: # type in your query to retrieve the number of columns in the CHICAGO_CRIME_DATA table
        %sql select * from syscat.columns where tabname = 'CHICAGO_CRIME_DATA';
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

Out[4]:
```
[('TTK07945', 'CHICAGO_CRIME_DATA', 'ID', 0, 'SYSIBM  ', 'INTEGER', 4, 0, None, None, No
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'CASE_NUMBER', 1, 'SYSIBM  ', 'VARCHAR', 8, 0, 'OCTE
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'FBICODE', 14, 'SYSIBM  ', 'VARCHAR', 3, 0, 'OCTETS'
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'X_COORDINATE', 15, 'SYSIBM  ', 'INTEGER', 4, 0, Non
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'Y_COORDINATE', 16, 'SYSIBM  ', 'INTEGER', 4, 0, Non
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'YEAR', 17, 'SYSIBM  ', 'SMALLINT', 2, 0, None, None
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'UPDATEDON', 18, 'SYSIBM  ', 'VARCHAR', 22, 0, 'OCTE
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'LATITUDE', 19, 'SYSIBM  ', 'DECIMAL', 18, 8, None,
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'LONGITUDE', 20, 'SYSIBM  ', 'DECIMAL', 18, 8, None,
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'LOCATION', 21, 'SYSIBM  ', 'VARCHAR', 29, 0, 'OCTET
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'DATE', 2, 'SYSIBM  ', 'VARCHAR', 22, 0, 'OCTETS', 2
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'BLOCK', 3, 'SYSIBM  ', 'VARCHAR', 35, 0, 'OCTETS',
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'IUCR', 4, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTETS', 4,
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'PRIMARY_TYPE', 5, 'SYSIBM  ', 'VARCHAR', 33, 0, 'OC
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'DESCRIPTION', 6, 'SYSIBM  ', 'VARCHAR', 46, 0, 'OCT
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'LOCATION_DESCRIPTION', 7, 'SYSIBM  ', 'VARCHAR', 33
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'ARREST', 8, 'SYSIBM  ', 'VARCHAR', 5, 0, 'OCTETS',
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'DOMESTIC', 9, 'SYSIBM  ', 'VARCHAR', 5, 0, 'OCTETS'
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'BEAT', 10, 'SYSIBM  ', 'SMALLINT', 2, 0, None, None
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'DISTRICT', 11, 'SYSIBM  ', 'SMALLINT', 2, 0, None,
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'WARD', 12, 'SYSIBM  ', 'SMALLINT', 2, 0, None, None
 ('TTK07945', 'CHICAGO_CRIME_DATA', 'COMMUNITY_AREA_NUMBER', 13, 'SYSIBM  ', 'SMALLINT',
```

In [5]:
```
import pandas

get = pandas.read_csv('https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7
get.head()
```

Out[5]:

	ID	CASE_NUMBER	DATE	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC
0	3512276	HK587712	08/28/2004 05:50:56 PM	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING	SMALL RETAIL STORE	False	False
1	3406613	HK456306	06/26/2004 12:40:00 PM	009XX N CENTRAL PARK AVE	820	THEFT	$500 AND UNDER	OTHER	False	False
2	8002131	HT233595	04/04/2011 05:45:00 AM	043XX S WABASH AVE	820	THEFT	$500 AND UNDER	NURSING HOME/RETIREMENT HOME	False	False
3	7903289	HT133522	12/30/2010 04:30:00 PM	083XX S KINGSTON AVE	840	THEFT	FINANCIAL ID THEFT: OVER $300	RESIDENCE	False	False
4	10402076	HZ138551	02/02/2016 07:30:00 PM	033XX W 66TH ST	820	THEFT	$500 AND UNDER	ALLEY	False	False

```
                ...                WARD  COMMUNITY_AREA_NUMBER  FBICODE  \
0               ...                14.0                   58.0        6
1               ...                27.0                   23.0        6
2               ...                 3.0                   38.0        6
3               ...                 7.0                   46.0        6
4               ...                15.0                   66.0        6

   X_COORDINATE Y_COORDINATE  YEAR              UPDATEDON   LATITUDE  \
0     1155838.0    1873050.0  2004  02/10/2018 03:50:01 PM  41.807441
1     1152206.0    1906127.0  2004  02/28/2018 03:56:25 PM  41.898280
2     1177436.0    1876313.0  2011  02/10/2018 03:50:01 PM  41.815933
3     1194622.0    1850125.0  2010  02/10/2018 03:50:01 PM  41.743665
4     1155240.0    1860661.0  2016  02/10/2018 03:50:01 PM  41.773455

   LONGITUDE                      LOCATION
0 -87.703956    (41.8074405, -87.703955849)
1 -87.716406  (41.898279962, -87.716405505)
2 -87.624642  (41.815933131, -87.624642127)
3 -87.562463  (41.743665322, -87.562462756)
4 -87.706480  (41.773455295, -87.706480471)

[5 rows x 22 columns]
```

CENSUS_DATA

```
In [6]: import pandas
        census = pandas.read_csv('https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd3
        print('census data imported successfully!')
```

```
census data imported successfully!
```

```
In [ ]: # %sql PERSIST census
```

```
In [7]: census.head()
```

```
Out[7]:    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
```

CHICAGO_PUBLIC_SCHOOLS

```
In [108]: ch_schools = pandas.read_csv('https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhplzt0
          print('ch_schools data imported successfully!')

ch_schools data imported successfully!


In [75]: %sql PERSIST ch_schools

 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB


Out[75]: 'Persisted ch_schools'

In [109]: ch_schools.head()

Out[109]:    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             ...               \
0        Fullerton Elementary Network              ...
1           Skyway Elementary Network              ...
2           Midway Elementary Network              ...
3         Pershing Elementary Network              ...
4  Southwest Side High School Network              ...

  Freshman on Track Rate % X_COORDINATE Y_COORDINATE   Latitude  Longitude  \
0                     NDA  1171699.458  1915829.428  41.924497 -87.644522
1                     NDA  1196129.985  1856209.466  41.760324 -87.556736
2                     NDA  1148427.165  1851012.215  41.747111 -87.731702
3                     NDA  1164504.290  1873959.199  41.809757 -87.672145
4                    91.8  1175177.622  1880745.126  41.828146 -87.632794

  COMMUNITY_AREA_NUMBER COMMUNITY_AREA_NAME  Ward Police District  \
0                     7        LINCOLN PARK    43              18
1                    43         SOUTH SHORE     7               4
2                    70             ASHBURN    13               8
3                    61            NEW CITY    20               9
4                    34       ARMOUR SQUARE    11               9

                     Location
0  (41.92449696, -87.64452163)
1  (41.76032435, -87.55673627)
2  (41.74711093, -87.73170248)
3    (41.8097569, -87.6721446)
4  (41.82814609, -87.63279369)

[5 rows x 78 columns]
```

In [132]:
```
%sql select * from syscat.columns where tabname = 'SCHOOLS';
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[132]: [('TTK07945', 'SCHOOLS', 'School_ID', 0, 'SYSIBM  ', 'INTEGER', 4, 0, None, None, None
           ('TTK07945', 'SCHOOLS', 'NAME_OF_SCHOOL', 1, 'SYSIBM  ', 'VARCHAR', 65, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'Elementary, Middle, or High School', 2, 'SYSIBM  ', 'VARCHAR
           ('TTK07945', 'SCHOOLS', 'Street_Address', 3, 'SYSIBM  ', 'VARCHAR', 30, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'City', 4, 'SYSIBM  ', 'VARCHAR', 7, 0, 'OCTETS', 7, None, 'Y
           ('TTK07945', 'SCHOOLS', 'State', 5, 'SYSIBM  ', 'VARCHAR', 2, 0, 'OCTETS', 2, None, '
           ('TTK07945', 'SCHOOLS', 'ZIP_Code', 6, 'SYSIBM  ', 'INTEGER', 4, 0, None, None, None,
           ('TTK07945', 'SCHOOLS', 'Phone_Number', 7, 'SYSIBM  ', 'VARCHAR', 14, 0, 'OCTETS', 14
           ('TTK07945', 'SCHOOLS', 'Link', 8, 'SYSIBM  ', 'VARCHAR', 78, 0, 'OCTETS', 78, None,
           ('TTK07945', 'SCHOOLS', 'Network_Manager', 9, 'SYSIBM  ', 'VARCHAR', 40, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'Collaborative_Name', 10, 'SYSIBM  ', 'VARCHAR', 34, 0, 'OCTE
           ('TTK07945', 'SCHOOLS', 'Adequate_Yearly_Progress_Made_', 11, 'SYSIBM  ', 'VARCHAR',
           ('TTK07945', 'SCHOOLS', 'Track_Schedule', 12, 'SYSIBM  ', 'VARCHAR', 12, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'CPS_Performance_Policy_Status', 13, 'SYSIBM  ', 'VARCHAR', 1
           ('TTK07945', 'SCHOOLS', 'CPS_Performance_Policy_Level', 14, 'SYSIBM  ', 'VARCHAR', 15
           ('TTK07945', 'SCHOOLS', 'HEALTHY_SCHOOL_CERTIFIED', 15, 'SYSIBM  ', 'VARCHAR', 3, 0,
           ('TTK07945', 'SCHOOLS', 'Safety_Icon', 16, 'SYSIBM  ', 'VARCHAR', 11, 0, 'OCTETS', 11
           ('TTK07945', 'SCHOOLS', 'SAFETY_SCORE', 17, 'SYSIBM  ', 'SMALLINT', 2, 0, None, None,
           ('TTK07945', 'SCHOOLS', 'Family_Involvement_Icon', 18, 'SYSIBM  ', 'VARCHAR', 11, 0,
           ('TTK07945', 'SCHOOLS', 'Family_Involvement_Score', 19, 'SYSIBM  ', 'VARCHAR', 3, 0,
           ('TTK07945', 'SCHOOLS', 'Environment_Icon', 20, 'SYSIBM  ', 'VARCHAR', 11, 0, 'OCTETS
           ('TTK07945', 'SCHOOLS', 'Environment_Score', 21, 'SYSIBM  ', 'SMALLINT', 2, 0, None,
           ('TTK07945', 'SCHOOLS', 'Instruction_Icon', 22, 'SYSIBM  ', 'VARCHAR', 11, 0, 'OCTETS
           ('TTK07945', 'SCHOOLS', 'Instruction_Score', 23, 'SYSIBM  ', 'SMALLINT', 2, 0, None,
           ('TTK07945', 'SCHOOLS', 'Leaders_Icon', 24, 'SYSIBM  ', 'VARCHAR', 11, 0, 'OCTETS', 1
           ('TTK07945', 'SCHOOLS', 'Leaders_Score', 25, 'SYSIBM  ', 'VARCHAR', 3, 0, 'OCTETS', 3
           ('TTK07945', 'SCHOOLS', 'Teachers_Icon', 26, 'SYSIBM  ', 'VARCHAR', 11, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'Teachers_Score', 27, 'SYSIBM  ', 'VARCHAR', 3, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'Parent_Engagement_Icon', 28, 'SYSIBM  ', 'VARCHAR', 7, 0, 'O
           ('TTK07945', 'SCHOOLS', 'Parent_Engagement_Score', 29, 'SYSIBM  ', 'VARCHAR', 3, 0, '
           ('TTK07945', 'SCHOOLS', 'Parent_Environment_Icon', 30, 'SYSIBM  ', 'VARCHAR', 7, 0, '
           ('TTK07945', 'SCHOOLS', 'Parent_Environment_Score', 31, 'SYSIBM  ', 'VARCHAR', 3, 0,
           ('TTK07945', 'SCHOOLS', 'AVERAGE_STUDENT_ATTENDANCE', 32, 'SYSIBM  ', 'VARCHAR', 6, 0
           ('TTK07945', 'SCHOOLS', 'Rate_of_Misconducts__per_100_students_', 33, 'SYSIBM  ', 'DE
           ('TTK07945', 'SCHOOLS', 'Average_Teacher_Attendance', 34, 'SYSIBM  ', 'VARCHAR', 6, 0
           ('TTK07945', 'SCHOOLS', 'Individualized_Education_Program_Compliance_Rate', 35, 'SYSI
           ('TTK07945', 'SCHOOLS', 'Pk_2_Literacy__', 36, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTETS',
           ('TTK07945', 'SCHOOLS', 'Pk_2_Math__', 37, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTETS', 4,
           ('TTK07945', 'SCHOOLS', 'Gr3_5_Grade_Level_Math__', 38, 'SYSIBM  ', 'VARCHAR', 4, 0,
           ('TTK07945', 'SCHOOLS', 'Gr3_5_Grade_Level_Read__', 39, 'SYSIBM  ', 'VARCHAR', 4, 0,
           ('TTK07945', 'SCHOOLS', 'Gr3_5_Keep_Pace_Read__', 40, 'SYSIBM  ', 'VARCHAR', 4, 0, 'O
           ('TTK07945', 'SCHOOLS', 'Gr3_5_Keep_Pace_Math__', 41, 'SYSIBM  ', 'VARCHAR', 4, 0, 'O
           ('TTK07945', 'SCHOOLS', 'Gr6_8_Grade_Level_Math__', 42, 'SYSIBM  ', 'VARCHAR', 4, 0,
           ('TTK07945', 'SCHOOLS', 'Gr6_8_Grade_Level_Read__', 43, 'SYSIBM  ', 'VARCHAR', 4, 0,
           ('TTK07945', 'SCHOOLS', 'Gr6_8_Keep_Pace_Math_', 44, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OC
           ('TTK07945', 'SCHOOLS', 'Gr6_8_Keep_Pace_Read__', 45, 'SYSIBM  ', 'VARCHAR', 4, 0, 'O
           ('TTK07945', 'SCHOOLS', 'Gr_8_Explore_Math__', 46, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTE
           ('TTK07945', 'SCHOOLS', 'Gr_8_Explore_Read__', 47, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTE
```

```
('TTK07945', 'SCHOOLS', 'ISAT_Exceeding_Math__', 48, 'SYSIBM  ', 'DECIMAL', 4, 1, Non
('TTK07945', 'SCHOOLS', 'ISAT_Exceeding_Reading__', 49, 'SYSIBM  ', 'DECIMAL', 4, 1,
('TTK07945', 'SCHOOLS', 'ISAT_Value_Add_Math', 50, 'SYSIBM  ', 'DECIMAL', 3, 1, None,
('TTK07945', 'SCHOOLS', 'ISAT_Value_Add_Read', 51, 'SYSIBM  ', 'DECIMAL', 3, 1, None,
('TTK07945', 'SCHOOLS', 'ISAT_Value_Add_Color_Math', 52, 'SYSIBM  ', 'VARCHAR', 6, 0,
('TTK07945', 'SCHOOLS', 'ISAT_Value_Add_Color_Read', 53, 'SYSIBM  ', 'VARCHAR', 6, 0,
('TTK07945', 'SCHOOLS', 'Students_Taking__Algebra__', 54, 'SYSIBM  ', 'VARCHAR', 4, 0
('TTK07945', 'SCHOOLS', 'Students_Passing__Algebra__', 55, 'SYSIBM  ', 'VARCHAR', 4,
('TTK07945', 'SCHOOLS', '9th Grade EXPLORE (2009)', 56, 'SYSIBM  ', 'VARCHAR', 4, 0,
('TTK07945', 'SCHOOLS', '9th Grade EXPLORE (2010)', 57, 'SYSIBM  ', 'VARCHAR', 4, 0,
('TTK07945', 'SCHOOLS', '10th Grade PLAN (2009)', 58, 'SYSIBM  ', 'VARCHAR', 4, 0, 'O
('TTK07945', 'SCHOOLS', '10th Grade PLAN (2010)', 59, 'SYSIBM  ', 'VARCHAR', 4, 0, 'O
('TTK07945', 'SCHOOLS', 'Net_Change_EXPLORE_and_PLAN', 60, 'SYSIBM  ', 'VARCHAR', 3,
('TTK07945', 'SCHOOLS', '11th Grade Average ACT (2011)', 61, 'SYSIBM  ', 'VARCHAR', 4
('TTK07945', 'SCHOOLS', 'Net_Change_PLAN_and_ACT', 62, 'SYSIBM  ', 'VARCHAR', 3, 0, '
('TTK07945', 'SCHOOLS', 'College_Eligibility__', 63, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OC
('TTK07945', 'SCHOOLS', 'Graduation_Rate__', 64, 'SYSIBM  ', 'VARCHAR', 4, 0, 'OCTETS
('TTK07945', 'SCHOOLS', 'College_Enrollment_Rate__', 65, 'SYSIBM  ', 'VARCHAR', 4, 0,
('TTK07945', 'SCHOOLS', 'COLLEGE_ENROLLMENT', 66, 'SYSIBM  ', 'SMALLINT', 2, 0, None,
('TTK07945', 'SCHOOLS', 'General_Services_Route', 67, 'SYSIBM  ', 'SMALLINT', 2, 0, N
('TTK07945', 'SCHOOLS', 'Freshman_on_Track_Rate__', 68, 'SYSIBM  ', 'VARCHAR', 4, 0,
('TTK07945', 'SCHOOLS', 'X_COORDINATE', 69, 'SYSIBM  ', 'DECIMAL', 13, 3, None, None,
('TTK07945', 'SCHOOLS', 'Y_COORDINATE', 70, 'SYSIBM  ', 'DECIMAL', 13, 3, None, None,
('TTK07945', 'SCHOOLS', 'Latitude', 71, 'SYSIBM  ', 'DECIMAL', 18, 8, None, None, Non
('TTK07945', 'SCHOOLS', 'Longitude', 72, 'SYSIBM  ', 'DECIMAL', 18, 8, None, None, No
('TTK07945', 'SCHOOLS', 'COMMUNITY_AREA_NUMBER', 73, 'SYSIBM  ', 'SMALLINT', 2, 0, No
('TTK07945', 'SCHOOLS', 'COMMUNITY_AREA_NAME', 74, 'SYSIBM  ', 'VARCHAR', 22, 0, 'OCT
('TTK07945', 'SCHOOLS', 'Ward', 75, 'SYSIBM  ', 'SMALLINT', 2, 0, None, None, None, '
('TTK07945', 'SCHOOLS', 'Police_District', 76, 'SYSIBM  ', 'SMALLINT', 2, 0, None, No
('TTK07945', 'SCHOOLS', 'Location', 77, 'SYSIBM  ', 'VARCHAR', 27, 0, 'OCTETS', 27, N
```

1.2 Problems

Now write and execute SQL queries to solve assignment problems

1.2.1 Problem 1

Find the total number of crimes recorded in the CRIME table

In [159]:
```
# Rows in Crime table

%sql select count(ID) as total_number_of_crimes_recorded from CHICAGO_CRIME_DATA;
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

Out[159]:
```
[(Decimal('533'),)]
```

1.2.2 Problem 2

Retrieve first 10 rows from the CRIME table

```
In [12]: %%sql

         select ID from CHICAGO_CRIME_DATA LIMIT 10;
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[12]: [(3512276,),
          (3406613,),
          (8002131,),
          (7903289,),
          (10402076,),
          (7732712,),
          (10769475,),
          (4494340,),
          (3778925,),
          (3324217,)]
```

1.2.3 Problem 3

How many crimes involve an arrest?

```
In [13]: %%sql

         select count(ARREST) from CHICAGO_CRIME_DATA
             where (ARREST = 'TRUE');
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[13]: [(Decimal('163'),)]
```

1.2.4 Problem 4

Which unique types of crimes have been recorded at GAS STATION locations?

```
In [14]: %%sql

         select PRIMARY_TYPE, LOCATION_DESCRIPTION from CHICAGO_CRIME_DATA
             where (LOCATION_DESCRIPTION = 'GAS STATION');
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[14]: [('THEFT', 'GAS STATION'),
          ('THEFT', 'GAS STATION'),
          ('NARCOTICS', 'GAS STATION'),
          ('ROBBERY', 'GAS STATION'),
          ('ROBBERY', 'GAS STATION'),
          ('CRIMINAL TRESPASS', 'GAS STATION')]
```

Hint: Which column lists types of crimes e.g. THEFT?

1.2.5 Problem 5

In the CENUS_DATA table list all Community Areas whose names start with the letter 'B'.

```
In [15]: %%sql

         select COMMUNITY_AREA_NAME from CENSUS_DATA
             where (COMMUNITY_AREA_NAME like 'B%');
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[15]: [('Belmont Cragin',),
          ('Burnside',),
          ('Brighton Park',),
          ('Bridgeport',),
          ('Beverly',)]
```

1.2.6 Problem 6

Which schools in Community Areas 10 to 15 are healthy school certified?

```
In [16]: %%sql

         select COMMUNITY_AREA_NUMBER, NAME_OF_SCHOOL, healthy_school_certified from SCHOOLS
             where COMMUNITY_AREA_NUMBER between 10 and 15
                 and (healthy_school_certified = 'Yes');
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[16]: [(10, 'Rufus M Hitch Elementary School', 'Yes')]
```

1.2.7 Problem 7

What is the average school Safety Score?

```
In [112]: %%sql

          select avg(SAFETY_SCORE) as AVG_SCHOOL_SAFTEY_SCORE from SCHOOLS;
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[112]: [(Decimal('49.504873'),)]
```

1.2.8 Problem 8

List the top 5 Community Areas by average College Enrollment [number of students]

```
In [122]: %%sql
          select COMMUNITY_AREA_NUMBER, avg(COLLEGE_ENROLLMENT) as Average_College_Enrollment fr
              group by COMMUNITY_AREA_NUMBER
               order by Average_College_Enrollment
                  fetch first 5 rows only;
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[122]: [(36, Decimal('140.000000')),
           (48, Decimal('261.333333')),
           (37, Decimal('265.500000')),
           (29, Decimal('321.625000')),
           (50, Decimal('324.000000'))]
```

1.2.9 Problem 9

Use a sub-query to determine which Community Area has the least value for school Safety Score?

```
In [158]: %%sql

          select COMMUNITY_AREA_NUMBER from SCHOOLS
              where (select min(SAFETY_SCORE) from SCHOOLS)
                  group by COMMUNITY_AREA_NUMBER
                      order by min(SAFETY_SCORE)
                          fetch first 1 rows only;
```

```
 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.
```

```
Out[158]: [(40,)]
```

1.2.10 Problem 10

[Without using an explicit JOIN operator] Find the Per Capita Income of the Community Area which has a school Safety Score of 1.

```
In [11]: %%sql
         select c.per_capita_income, s.community_area_number, s.community_area_name from census_
                                  where c.community_area_number = s.community_area_number
                                          and safety_score = 1;

 * ibm_db_sa://ttk07945:***@dashdb-txn-sbox-yp-dal09-04.services.dal.bluemix.net:50000/BLUDB
Done.


Out[11]: [(13785, 40, 'WASHINGTON PARK')]
```