

# DB0201EN-Week4-1-1-RealDataPractice-v4-py

February 13, 2019

Lab: Working with a real world data-set using SQL and Python

## 1 Introduction

This notebook shows how to work with a real world dataset using SQL and Python. In this lab you will: 1. Understand the dataset for Chicago Public School level performance 1. Store the dataset in an Db2 database on IBM Cloud instance 1. Retrieve metadata about tables and columns and query data from mixed case columns 1. Solve example problems to practice your SQL skills including using built-in database functions

### 1.1 Chicago Public Schools - Progress Report Cards (2011-2012)

The city of Chicago released a dataset showing all school level performance data used to create School Report Cards for the 2011-2012 school year. The dataset is available from the Chicago Data Portal: <https://data.cityofchicago.org/Education/Chicago-Public-Schools-Progress-Report-Cards-2011-/9xs2-f89t>

This dataset includes a large number of metrics. Start by familiarizing yourself with the types of metrics in the database: <https://data.cityofchicago.org/api/assets/AAD41A13-BE8A-4E67-B1F5-86E711E09D5F?download=true>

**NOTE:** Do not download the dataset directly from City of Chicago portal. Instead download a more database friendly version from the link below. Now download a static copy of this database and review some of its contents: <https://ibm.box.com/shared/static/f9gvj1gjmxxzycdhplzt01qtz0s7ew7.csv>

#### 1.1.1 Store the dataset in a Table

In many cases the dataset to be analyzed is available as a .CSV (comma separated values) file, perhaps on the internet. 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 the previous lab, 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 CHICAGO PUBLIC SCHOOLS dataset and load the dataset into a new table called SCHOOLS.

### 1.1.2 Connect to the database

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

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

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

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

### 1.1.3 Query the database system catalog to retrieve table metadata

You can verify that the table creation was successful by retrieving the list of all tables in your schema and checking whether the SCHOOLS table was created

```
In [4]: # type in your query to retrieve list of all tables in the database for your db2 schema
```

```
%sql select TABSCHEMA, TABNAME, CREATE_TIME from syscat.tables where tabschema='TTK07945'
```

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

```
Out[4]: [('TTK07945', 'INSTRUCTOR', datetime.datetime(2019, 2, 7, 0, 23, 43, 252860)),
('TTK07945', 'EMPLOYEES', datetime.datetime(2019, 2, 3, 21, 4, 20, 481536)),
('TTK07945', 'JOB_HISTORY', datetime.datetime(2019, 2, 3, 21, 4, 20, 879875)),
('TTK07945', 'JOBS', datetime.datetime(2019, 2, 3, 21, 4, 21, 324450)),
('TTK07945', 'DEPARTMENTS', datetime.datetime(2019, 2, 3, 21, 4, 21, 772202)),
('TTK07945', 'LOCATIONS', datetime.datetime(2019, 2, 3, 21, 4, 22, 227941)),
('TTK07945', 'PETSALe', datetime.datetime(2019, 2, 5, 0, 35, 15, 48406)),
('TTK07945', 'INTERNATIONAL_STUDENT_TEST_SCORES', datetime.datetime(2019, 2, 11, 16, 54, 9)),
('TTK07945', 'CHICAGO_SOCIOECONOMIC_DATA', datetime.datetime(2019, 2, 11, 18, 28, 54, 9)),
('TTK07945', 'SCHOOLS', datetime.datetime(2019, 2, 13, 19, 21, 43, 161457))]
```

Double-click [here](#) for a hint

Double-click [here](#) for the solution.

### 1.1.4 Query the database system catalog to retrieve column metadata

The SCHOOLS table contains a large number of columns. How many columns does this table have?

```
In [9]: # type in your query to retrieve the number of columns in the SCHOOLS table
%sql select * from syscat.columns where tablename = 'SCHOOLS';
```

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

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

```
( 'TTK07945', 'SCHOOLS', 'Gr6_8_Keep_Pace_Math_', 44, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Gr6_8_Keep_Pace_Read__', 45, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Gr_8_Explore_Math__', 46, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Gr_8_Explore_Read__', 47, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'ISAT_Exceeding_Math__', 48, 'SYSIBM ', 'DECIMAL', 4, 1, None,
( 'TTK07945', 'SCHOOLS', 'ISAT_Exceeding_Reading__', 49, 'SYSIBM ', 'DECIMAL', 4, 1, None,
( '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, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'ISAT_Value_Add_Color_Read', 53, 'SYSIBM ', 'VARCHAR', 6, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Students_Taking__Algebra__', 54, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Students_Passing__Algebra__', 55, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', '9th Grade EXPLORE (2009)', 56, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', '9th Grade EXPLORE (2010)', 57, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', '10th Grade PLAN (2009)', 58, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', '10th Grade PLAN (2010)', 59, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Net_Change_EXPLORE_and_PLAN', 60, 'SYSIBM ', 'VARCHAR', 3, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', '11th Grade Average ACT (2011)', 61, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Net_Change_PLAN_and_ACT', 62, 'SYSIBM ', 'VARCHAR', 3, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'College_Eligibility__', 63, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Graduation_Rate__', 64, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'College_Enrollment_Rate__', 65, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'COLLEGE_ENROLLMENT', 66, 'SYSIBM ', 'SMALLINT', 2, 0, None,
( 'TTK07945', 'SCHOOLS', 'General_Services_Route', 67, 'SYSIBM ', 'SMALLINT', 2, 0, None,
( 'TTK07945', 'SCHOOLS', 'Freshman_on_Track_Rate__', 68, 'SYSIBM ', 'VARCHAR', 4, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'X_COORDINATE', 69, 'SYSIBM ', 'DECIMAL', 13, 3, None, None, None,
( 'TTK07945', 'SCHOOLS', 'Y_COORDINATE', 70, 'SYSIBM ', 'DECIMAL', 13, 3, None, None, None,
( 'TTK07945', 'SCHOOLS', 'Latitude', 71, 'SYSIBM ', 'DECIMAL', 18, 8, None, None, None,
( 'TTK07945', 'SCHOOLS', 'Longitude', 72, 'SYSIBM ', 'DECIMAL', 18, 8, None, None, None,
( 'TTK07945', 'SCHOOLS', 'COMMUNITY_AREA_NUMBER', 73, 'SYSIBM ', 'SMALLINT', 2, 0, None,
( 'TTK07945', 'SCHOOLS', 'COMMUNITY_AREA_NAME', 74, 'SYSIBM ', 'VARCHAR', 22, 0, 'OCTETS',
( 'TTK07945', 'SCHOOLS', 'Ward', 75, 'SYSIBM ', 'SMALLINT', 2, 0, None, None, None, 'Y',
( 'TTK07945', 'SCHOOLS', 'Police_District', 76, 'SYSIBM ', 'SMALLINT', 2, 0, None, None, None,
( 'TTK07945', 'SCHOOLS', 'Location', 77, 'SYSIBM ', 'VARCHAR', 27, 0, 'OCTETS', 27, None,
```

Double-click [here](#) for a hint

Double-click [here](#) for the solution.

Now retrieve the the list of columns in SCHOOLS table and their column type (datatype) and length.

```
In [8]: # type in your query to retrieve all column names in the SCHOOLS table along with their
        %sql select distinct(name), coltype, length from sysibm.syscolumns where tbname = 'SCHOOLS'

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

```
Out[8]: [('10th Grade PLAN (2009)', 'VARCHAR ', 4),
         ('10th Grade PLAN (2010)', 'VARCHAR ', 4),
```

('11th Grade Average ACT (2011)', 'VARCHAR ', 4),  
 ('9th Grade EXPLORE (2009)', 'VARCHAR ', 4),  
 ('9th Grade EXPLORE (2010)', 'VARCHAR ', 4),  
 ('AVERAGE\_STUDENT\_ATTENDANCE', 'VARCHAR ', 6),  
 ('Adequate\_Yearly\_Progress\_Made\_', 'VARCHAR ', 3),  
 ('Average\_Teacher\_Attendance', 'VARCHAR ', 6),  
 ('COLLEGE\_ENROLLMENT', 'SMALLINT', 2),  
 ('COMMUNITY\_AREA\_NAME', 'VARCHAR ', 22),  
 ('COMMUNITY\_AREA\_NUMBER', 'SMALLINT', 2),  
 ('CPS\_Performance\_Policy\_Level', 'VARCHAR ', 15),  
 ('CPS\_Performance\_Policy\_Status', 'VARCHAR ', 16),  
 ('City', 'VARCHAR ', 7),  
 ('Collaborative\_Name', 'VARCHAR ', 34),  
 ('College\_Eligibility\_\_', 'VARCHAR ', 4),  
 ('College\_Enrollment\_Rate\_\_', 'VARCHAR ', 4),  
 ('Elementary, Middle, or High School', 'VARCHAR ', 2),  
 ('Environment\_Icon', 'VARCHAR ', 11),  
 ('Environment\_Score', 'SMALLINT', 2),  
 ('Family\_Involvement\_Icon', 'VARCHAR ', 11),  
 ('Family\_Involvement\_Score', 'VARCHAR ', 3),  
 ('Freshman\_on\_Track\_Rate\_\_', 'VARCHAR ', 4),  
 ('General\_Services\_Route', 'SMALLINT', 2),  
 ('Gr3\_5\_Grade\_Level\_Math\_\_', 'VARCHAR ', 4),  
 ('Gr3\_5\_Grade\_Level\_Read\_\_', 'VARCHAR ', 4),  
 ('Gr3\_5\_Keep\_Pace\_Math\_\_', 'VARCHAR ', 4),  
 ('Gr3\_5\_Keep\_Pace\_Read\_\_', 'VARCHAR ', 4),  
 ('Gr6\_8\_Grade\_Level\_Math\_\_', 'VARCHAR ', 4),  
 ('Gr6\_8\_Grade\_Level\_Read\_\_', 'VARCHAR ', 4),  
 ('Gr6\_8\_Keep\_Pace\_Math\_', 'VARCHAR ', 4),  
 ('Gr6\_8\_Keep\_Pace\_Read\_\_', 'VARCHAR ', 4),  
 ('Gr\_8\_Explore\_Math\_\_', 'VARCHAR ', 4),  
 ('Gr\_8\_Explore\_Read\_\_', 'VARCHAR ', 4),  
 ('Graduation\_Rate\_\_', 'VARCHAR ', 4),  
 ('HEALTHY\_SCHOOL\_CERTIFIED', 'VARCHAR ', 3),  
 ('ISAT\_Exceeding\_Math\_\_', 'DECIMAL ', 4),  
 ('ISAT\_Exceeding\_Reading\_\_', 'DECIMAL ', 4),  
 ('ISAT\_Value\_Add\_Color\_Math', 'VARCHAR ', 6),  
 ('ISAT\_Value\_Add\_Color\_Read', 'VARCHAR ', 6),  
 ('ISAT\_Value\_Add\_Math', 'DECIMAL ', 3),  
 ('ISAT\_Value\_Add\_Read', 'DECIMAL ', 3),  
 ('Individualized\_Education\_Program\_Compliance\_Rate', 'VARCHAR ', 7),  
 ('Instruction\_Icon', 'VARCHAR ', 11),  
 ('Instruction\_Score', 'SMALLINT', 2),  
 ('Latitude', 'DECIMAL ', 18),  
 ('Leaders\_Icon', 'VARCHAR ', 11),  
 ('Leaders\_Score', 'VARCHAR ', 3),  
 ('Link', 'VARCHAR ', 78),  
 ('Location', 'VARCHAR ', 27),

```
( 'Longitude', 'DECIMAL ', 18),
( 'NAME_OF_SCHOOL', 'VARCHAR ', 65),
( 'Net_Change_EXPLORE_and_PLAN', 'VARCHAR ', 3),
( 'Net_Change_PLAN_and_ACT', 'VARCHAR ', 3),
( 'Network_Manager', 'VARCHAR ', 40),
( 'Parent_Engagement_Icon', 'VARCHAR ', 7),
( 'Parent_Engagement_Score', 'VARCHAR ', 3),
( 'Parent_Environment_Icon', 'VARCHAR ', 7),
( 'Parent_Environment_Score', 'VARCHAR ', 3),
( 'Phone_Number', 'VARCHAR ', 14),
( 'Pk_2_Literacy__', 'VARCHAR ', 4),
( 'Pk_2_Math__', 'VARCHAR ', 4),
( 'Police_District', 'SMALLINT', 2),
( 'Rate_of_Misconducts__per_100_students_', 'DECIMAL ', 5),
( 'SAFETY_SCORE', 'SMALLINT', 2),
( 'Safety_Icon', 'VARCHAR ', 11),
( 'School_ID', 'INTEGER ', 4),
( 'State', 'VARCHAR ', 2),
( 'Street_Address', 'VARCHAR ', 30),
( 'Students_Passing__Algebra__', 'VARCHAR ', 4),
( 'Students_Taking__Algebra__', 'VARCHAR ', 4),
( 'Teachers_Icon', 'VARCHAR ', 11),
( 'Teachers_Score', 'VARCHAR ', 3),
( 'Track_Schedule', 'VARCHAR ', 12),
( 'Ward', 'SMALLINT', 2),
( 'X_COORDINATE', 'DECIMAL ', 13),
( 'Y_COORDINATE', 'DECIMAL ', 13),
( 'ZIP_Code', 'INTEGER ', 4)]
```

Double-click **here** for the solution.

### 1.1.5 Questions

1. Is the column name for the "SCHOOL ID" attribute in upper or mixed case?
2. What is the name of "Community Area Name" column in your table? Does it have spaces?
3. Are there any columns in whose names the spaces and paranthesis (round brackets) have been replaced by the underscore character "\_"?

## 1.2 Problems

### 1.2.1 Problem 1

**How many Elementary Schools are in the dataset?**

```
In [16]: %%sql
select count(*) from SCHOOLS
        where "Elementary, Middle, or High School" = 'ES';
```

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

```
Out[16]: [(Decimal('462'),)]
```

Double-click [here](#) for a hint  
Double-click [here](#) for another hint  
Double-click [here](#) for the solution.

### 1.2.2 Problem 2

**What is the highest Safety Score?**

```
In [18]: %%sql
        select max(SAFETY_SCORE) from SCHOOLS as MAX_SAFETY_SCORE;
```

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

```
Out[18]: [(99,)]
```

Double-click [here](#) for a hint  
Double-click [here](#) for the solution.

### 1.2.3 Problem 3

**Which schools have highest Safety Score?**

```
In [22]: %%sql

        select NAME_OF_SCHOOL, Safety_Score from SCHOOLS
        where Safety_Score = 99;
```

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

```
Out[22]: [('Abraham Lincoln Elementary School', 99),
          ('Alexander Graham Bell Elementary School', 99),
          ('Annie Keller Elementary Gifted Magnet School', 99),
          ('Augustus H Burley Elementary School', 99),
          ('Edgar Allan Poe Elementary Classical School', 99),
          ('Edgebrook Elementary School', 99),
          ('Ellen Mitchell Elementary School', 99),
          ('James E McDade Elementary Classical School', 99),
          ('James G Blaine Elementary School', 99),
          ('LaSalle Elementary Language Academy', 99),
          ('Mary E Courtenay Elementary Language Arts Center', 99),
```

```
( 'Northside College Preparatory High School', 99),
( 'Northside Learning Center High School', 99),
( 'Norwood Park Elementary School', 99),
( 'Oriole Park Elementary School', 99),
( 'Sauganash Elementary School', 99),
( 'Stephen Decatur Classical Elementary School', 99),
( 'Talman Elementary School', 99),
( 'Wildwood Elementary School', 99)]
```

Double-click [here](#) for the solution.

#### 1.2.4 Problem 4

**What are the top 10 schools with the highest "Average Student Attendance"?**

```
In [25]: %%sql
        select Name_of_School, AVERAGE_STUDENT_ATTENDANCE from SCHOOLS
           order by Average_Student_Attendance desc nulls last limit 10;

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

```
Out[25]: [('John Charles Haines Elementary School', '98.40%'),
          ('James Ward Elementary School', '97.80%'),
          ('Edgar Allan Poe Elementary Classical School', '97.60%'),
          ('Orozco Fine Arts & Sciences Elementary School', '97.60%'),
          ('Rachel Carson Elementary School', '97.60%'),
          ('Annie Keller Elementary Gifted Magnet School', '97.50%'),
          ('Andrew Jackson Elementary Language Academy', '97.40%'),
          ('Lenart Elementary Regional Gifted Center', '97.40%'),
          ('Disney II Magnet School', '97.30%'),
          ('John H Vanderpoel Elementary Magnet School', '97.20%')]
```

Double-click [here](#) for the solution.

#### 1.2.5 Problem 5

**Retrieve the list of 5 Schools with the lowest Average Student Attendance sorted in ascending order based on attendance**

```
In [28]: %%sql
        select Name_of_School, AVERAGE_STUDENT_ATTENDANCE from SCHOOLS
           order by Average_Student_Attendance fetch first 5 rows only;

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



```
Out[28]: [('Richard T Crane Technical Preparatory High School', '57.90%'),
          ('Barbara Vick Early Childhood & Family Center', '60.90%'),
          ('Dyett High School', '62.50%'),
          ('Wendell Phillips Academy High School', '63.00%'),
          ('Orr Academy High School', '66.30%')]
```

Double-click [here](#) for the solution.

### 1.2.6 Problem 6

Now remove the '%' sign from the above result set for Average Student Attendance column

```
In [29]: %%sql
        select Name_of_School, replace(AVERAGE_STUDENT_ATTENDANCE, '%', '') from SCHOOLS
        order by Average_Student_Attendance fetch first 5 rows only;

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

```
Out[29]: [('Richard T Crane Technical Preparatory High School', '57.90'),
          ('Barbara Vick Early Childhood & Family Center', '60.90'),
          ('Dyett High School', '62.50'),
          ('Wendell Phillips Academy High School', '63.00'),
          ('Orr Academy High School', '66.30')]
```

Double-click [here](#) for a hint

Double-click [here](#) for the solution.

### 1.2.7 Problem 7

Which Schools have Average Student Attendance lower than 70%?

```
In [36]: %%sql
        select Name_of_School, AVERAGE_STUDENT_ATTENDANCE from SCHOOLS
        where Average_Student_Attendance < '70%';

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

```
Out[36]: [('Barbara Vick Early Childhood & Family Center', '60.90%'),
          ('Chicago Vocational Career Academy High School', '68.80%'),
          ('Dyett High School', '62.50%'),
          ('Manley Career Academy High School', '66.80%'),
          ('Orr Academy High School', '66.30%'),
          ('Richard T Crane Technical Preparatory High School', '57.90%'),
          ('Roberto Clemente Community Academy High School', '69.60%'),
          ('Wendell Phillips Academy High School', '63.00%')]
```

Double-click [here](#) for a hint

Double-click [here](#) for another hint

Double-click [here](#) for the solution.

## 1.2.8 Problem 8

### Get the total College Enrollment for each Community Area

```
In [41]: %%sql
        select COMMUNITY_AREA_NAME, sum(COLLEGE_ENROLLMENT) as Total_Enrollment from SCHOOLS
        group by Community_Area_Name;

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

```
Out[41]: [('ALBANY PARK', 6864),
          ('ARCHER HEIGHTS', 4823),
          ('ARMOUR SQUARE', 1458),
          ('ASHBURN', 6483),
          ('AUBURN GRESHAM', 4175),
          ('AUSTIN', 10933),
          ('AVALON PARK', 1522),
          ('AVONDALE', 3640),
          ('BELMONT CRAGIN', 14386),
          ('BEVERLY', 1636),
          ('BRIDGEPORT', 3167),
          ('BRIGHTON PARK', 9647),
          ('BURNSIDE', 549),
          ('CALUMET HEIGHTS', 1568),
          ('CHATHAM', 5042),
          ('CHICAGO LAWN', 7086),
          ('CLEARING', 2085),
          ('DOUGLAS', 4670),
          ('DUNNING', 4568),
          ('EAST GARFIELD PARK', 5337),
          ('EAST SIDE', 5305),
          ('EDGEWATER', 4600),
          ('EDISON PARK', 910),
          ('ENGLEWOOD', 6832),
          ('FOREST GLEN', 1431),
          ('FULLER PARK', 531),
          ('GAGE PARK', 9915),
          ('GARFIELD RIDGE', 4552),
          ('GRAND BOULEVARD', 2809),
          ('GREATER GRAND CROSSING', 4051),
          ('HEGEWISCH', 963),
          ('HERMOSA', 3975),
          ('HUMBOLDT PARK', 8620),
          ('HYDE PARK', 1930),
          ('IRVING PARK', 7764),
          ('JEFFERSON PARK', 1755),
          ('KENWOOD', 4287),
```

```

('LAKE VIEW', 7055),
('LINCOLN PARK', 5615),
('LINCOLN SQUARE', 4132),
('LOGAN SQUARE', 7351),
('LOOP', 871),
('LOWER WEST SIDE', 7257),
('MCKINLEY PARK', 1552),
('MONTCLARE', 1317),
('MORGAN PARK', 3271),
('MOUNT GREENWOOD', 2091),
('NEAR NORTH SIDE', 3362),
('NEAR SOUTH SIDE', 1378),
('NEAR WEST SIDE', 7975),
('NEW CITY', 7922),
('NORTH CENTER', 7541),
('NORTH LAWNSDALE', 5146),
('NORTH PARK', 4210),
('NORWOOD PARK', 6469),
('OAKLAND', 140),
('OHARE', 786),
('PORTAGE PARK', 6954),
('PULLMAN', 1620),
('RIVERDALE', 1547),
('ROGERS PARK', 4068),
('ROSELAND', 7020),
('SOUTH CHICAGO', 4043),
('SOUTH DEERING', 1859),
('SOUTH LAWNSDALE', 14793),
('SOUTH SHORE', 4543),
('UPTOWN', 4388),
('WASHINGTON HEIGHTS', 4006),
('WASHINGTON PARK', 2648),
('WEST ELSDON', 3700),
('WEST ENGLEWOOD', 5946),
('WEST GARFIELD PARK', 2622),
('WEST LAWN', 4207),
('WEST PULLMAN', 3240),
('WEST RIDGE', 8197),
('WEST TOWN', 9429),
('WOODLAWN', 4206)]

```

Double-click **here** for a hint

Double-click **here** for another hint

Double-click **here** for the solution.

### 1.2.9 Problem 9

**Get the 5 Community Areas with the least total College Enrollment sorted in ascending order**

In [44]: %%sql

```
select COMMUNITY_AREA_NAME, sum(College_Enrollment) as TOTAL_ENROLLMENT from SCHOOLS
group by Community_Area_Name
order by TOTAL_ENROLLMENT asc
fetch first 5 rows only;
```

\* ibm\_db\_sa://ttk07945:\*\*\*@dashdb-txn-sbox-yp-dal09-04.services.dal.ibm.com:50000/BLUDB  
Done.

Out[44]: [('OAKLAND', 140),  
 ('FULLER PARK', 531),  
 ('BURNSIDE', 549),  
 ('OHARE', 786),  
 ('LOOP', 871)]

Double-click [here](#) for a hint

Double-click [here](#) for the solution.

### 1.2.10 Problem 10

**Get the hardship index for the community area which has College Enrollment of 4638**

In [59]: %%sql

```
select hardship_index
from chicago_socioeconomic_data as CD, schools as CPS
where CD.ca = CPS.community_area_number
and college_enrollment = 4368;
```

\* ibm\_db\_sa://ttk07945:\*\*\*@dashdb-txn-sbox-yp-dal09-04.services.dal.ibm.com:50000/BLUDB  
Done.

Out[59]: [(6.0,)]

Double-click [here](#) for the solution.

### 1.2.11 Problem 11

**Get the hardship index for the community area which has the highest value for College Enrollment**

In [64]: %%sql

```
select ca, community_area_name, hardship_index from chicago_socioeconomic_data
where ca in (select community_area_number from schools order by college_enrollment
```

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

```
Out[64]: [(5.0, 'North Center', 6.0)]
```

Double-click [here](#) for the solution.

### 1.3 Summary

In this lab you learned how to work with a real word dataset using SQL and Python. You learned how to query columns with spaces or special characters in their names and with mixed case names. You also used built in database functions and practiced how to sort, limit, and order result sets, as well as used sub-queries and worked with multiple tables. Copyright © 2018 [cognitiveclass.ai](#). This notebook and its source code are released under the terms of the [MIT License](#).