CS340 – Assignment 1
(Dr. Naveed Arshad)

Read the handout very carefully.

(Deadline: 26th September 2019 11:55pm)

Important Guidelines:

1- All your code should be in "**CS340_your-rollno_A1.sql**". You are not supposed to submit more than one file. Just upload it.
2- You are supposed to do this assignment **alone.** Any kind of collaboration, except for discussing what was taught in the class is strictly prohibited.
3- Any case of cheating (if caught) will be reported to the Disciplinary Committee without any delay. **(This includes trying to trick the auto-grader by any means)**
4- There are some useful links at the bottom of this document. Read them after you completely read the handout.
5- Any assignment which is even one second late, or has any other file except the above mentioned will not be accepted.
6- If your code does not compile, do not argue for the grade.

Environment Setup:

We will be loading our datasets and writing SQL queries on the Db2 provision on IBM Cloud, but first we need to setup a Db2 Object on the Cloud. Follow the steps below and you'll be able to setup your Db2 Object in about in a few minutes.

1- Go to the following link https://www.ibm.com/cloud and signup for your IBM Cloud account.
2- Initialize a free Cloud-storage-Object on the Lite Plan.
    - Navigate to the Dashboard and Click on the "Create Resource" button on the top right corner.
    - Scroll down and click on "Object Storage". Then, the process is self-explanatory.
3- Initialize a Db2 Object
    - Navigate to the Dashboard and type "Db2" in the Search bar and initialize a Db2 Object like the initialization of the Cloud-Storage-Object.
4- Go to "Resource List" from the side bar and click on the Db2 Object you created.
5- Click on "Open Console" and you are good to go.

NOTE: A separate guide has been provided in the assignment folder for loading datasets and running SQL queries.

Assignment Instructions:

The following datasets from the Chicago Data Portal are provided to you for the assignment.

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.

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.

3. Chicago Crime Data

This dataset reflects reported incidents of crime (except for murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days.

NOTE: The original datasets are very large. Hence, a subset of the datasets has been provided to you in the assignment folder.

Load the datasets and initialize the table names as shown below: (Strictly follow these table names else marks will be deducted by the auto-grader)

1- **SCHOOLS**
2- **CRIME**
3- **CENSUS_DATA**

You are required to write SQL queries to answer the following questions in the "Run SQL" tab of the Db2 setup. Once you've executed the queries copy all your code in a file named "CS340_your-rollno_A1.sql" and upload your submission on LMS.

[Total Marks: 100]

a) Find the total number of crimes recorded in the CRIME table. **(5)**
b) Retrieve first 10 rows from the CRIME table. **(5)**
c) How many crimes involve an arrest? **(5)**
d) What is the average school Safety Score? **(5)**
e) In the CENUS_DATA table list all Community Areas whose names start with the letter 'B'. **(5)**
f) Which unique types of crimes have been recorded at GAS STATION locations? **(5)**
g) Which schools in Community Areas 10 to 15 are healthy school certified? **(5)**
h) List the top 5 Community Areas by average College Enrollment [number of students] **(10)**
i) Use a sub-query to determine which Community Area has the least value for school Safety Score? **(10)**
j) [Without using an explicit JOIN operator] Find the name of the community area and the Per Capita Income of the Community Area which has a school Safety Score of 1. **(15)**
k) Find the School ID, Name of School and City of all the schools with a lower than average safety score. **(10)**
l) Find the name of the community area and the Per Capita Income of the Community Area which has School Safety Score of 1, **using a JOIN operator. (20)**