DATA MANAGEMENT
Team Exercise #5a – Topic Review

In this exercise your team will create a data model, prepare a database, and develop some SQL queries.

Getting the benefit from this exercise requires that you invest some individual time. Once everyone on your team has done that, you should compare notes and develop an agreed-upon model. At this point you may decide to temporarily split the work up for questions 2 and 3. **However**, by the end of the exercise each person on your team should have a very clear idea about how everything was accomplished.

Please take time to help explain concepts to one another. If someone seems uncertain, spend a bit of extra time explaining. These are really important and valuable skills, and you'll need them for your project work (not to mention career opportunities).

1. Create a high-fidelity data model for Charity Intelligence (Ci) based on the attached case, using Barker notation.
2. Using the ATOM text editor, write SQL statements to implement your database. Your file should include all CREATE statements to generate the db, tables, columns, foreign keys, relationships, as well as all INSERT statements necessary to seed each table with five rows of test data.
   a) Name your database as follows: "CI_TEAM##" where ## is your team number (01, 02, etc).
   b) Do not merely use the phpMyAdmin export function to dump your final SQL code. You need to actually type this out, and confirm it works by pasting and running the entire file in SQL.
   c) Add a semi-colon (;) after each and every SQL command, as this makes it much easier to copy-and-test your code.
3. Once again using ATOM, write out SQL queries to answer the following:
   a) Provide an alphabetical list showing common and legal names of all Ontario charities.
   b) Generate a list with employee first name, last name, and number of articles published.
   c) How many times was each article viewed, from most to least popular?
   d) List the common name of each charity, along with its sector and sum of all donations made during 2020, starting with the highest donation total.

Collect together your team's data model, database (saved in an ATOM sql file), and SQL queries (saved in an ATOM sql file), combine them into a single archive / zip file, and **submit this single zip file as a Learn assignment** on behalf of your team.

Notice: your team must work independently. You are allowed to refer to course materials and online sources to learn more about SQL syntax. But you must *not* confer about the case, your model, your queries, or share/use any other information from any person outside of your team.

Due Thursday at 11:59 pm.

Charity Intelligence (Ci)*

Kate Bahen, retired Equity Research Associate and founder of the online Canadian watchdog Charity Intelligence (Ci), was on the phone and staring at a three-year-old sticky note on her desk labelled "FUNDRAISE!!" The note served a constant reminder of her goal of having Ci funded entirely by donations. Ironically, the purpose of her current call was to instruct her financial advisor to withdraw $80,000 from her retirement savings to finance Ci's deficit, for the 10th year in a row. Donations to Ci had plateaued, and Kate didn't know what to do.

Canadians collectively donated $16 billion annually to over 86,000 registered Canadian charities.† Ci's mission was "to provide Canadian donors with information to help them make informed and intelligent giving decisions to have the greatest impact." This was accomplished through publication of easy-to-read analytical reports on their website featuring each of the 700 largest charities in Canada, as well as articles on charitable giving entitled "Views on Charity News".

Ci's message was spreading. Website visits had grown 83 percent annually, with over 260,000 hits in the current year. Despite this growth, Ci was chronically underfunded. Bahen had a sense that certain types of charity advertisements, articles, messages, and efforts must be more effective than others – and Ci seemed to have a mountain of website visitor data. How was it that they were always flying in the dark?

Operationally, all charity data was currently stored on Excel spreadsheets, one spreadsheet per charity profile. The hundreds of resulting Excel files were inconsistent, and many contained buggy VBA macros. When an intern completed a charity profile, they logged in to Ci's website account and manually input the final calculated information from the Excel spreadsheet into that page's source code. Thus, all the comments, financial information, and scores for each charity were essentially hardcoded into the HTML code of each charity profile's webpage.

Bahen had learned from her intern, Ethan Smart, that the non-profit was sitting on a "data lake" of unstructured user session information that was not very useful in its current form. If Ci were to organize its data and user session information into a structured database, Smart argued, it would be possible to analyze donor web behaviours to learn what kinds of content was associated with higher donation amounts.

Bahen barely understood technology, or what Smart was talking about – she much preferred to keep everything on nice clean spreadsheets! But if she kept spending her retirement savings to keep CI afloat, she realized she wouldn't be able to stay retired for long. And she did see value in knowing more about what kinds of content donors were responding to. Maybe a centralized database really was the solution to her biggest problem. Bahen asked Smart to dig into the problem and make a recommendation.

---

* Based on a case study developed by Matthew Scierko and Kevin Silberberg under the supervision of Professor Derrick Neufeld.
† http://sectorsource.ca/research-and-impact/giving-research

Data Requirements

After conversations with Bahen and the full-time Ci employees, Smart assembled the following facts:

Each charity that Ci tracks has at minimum a registration number and legal name, and possibly additional elements (i.e., common name, mailing address, and website URL). Each charity belongs to at least one, and possibly multiple sectors (e.g., Animal Welfare, Arts & Culture, Education, etc.), and each sector contains many different charities.

Ci updates every charity's profile every year, and so a given charity includes multiple annual profiles (i.e., one each for 2020, 2019, 2018, etc.), while each annual profile relates to just one particular charity. Characteristics of each profile include year (which along with registration number helps to uniquely identify that profile), 'about' text, financial review text, a financial transparency score (out of 100), and optionally, preparer notes.

Ci has several full-time employees, and occasional part-time interns who are treated as employees. Employee data includes name, SIN, mailing address, phone number, and email address. Every employee is responsible for authoring multiple annual profiles, and profiles are often authored by multiple employees. Each employee is also responsible for authoring multiple articles during the year, and articles are frequently authored by more than one employee. Article data includes publication date, title, and article text.

Whenever someone visits the Ci website, a new 'session' is generated. Each session includes a cookie id, the date and time of cookie creation, and the date and time of cookie expiry. During a session, one or more charities may be opened, and one or more articles may be viewed. The specific date and time of these activities is recorded.

Finally, during their session a user may decide to make one or more donations, for which the following data is captured: transaction id, date, time, amount. Each donation is associated with just one session. (Payments and tax receipt invoicing are handled through a third-party processor, so Ci does not collect identity or credit card information.)