CEE 412 / CET 522 Transportation Data Management and Visualization

Assignment 3 | Winter 2020

CEE 412 students will have a total possible score of 50 points and CET 522 students will have a total possible score of 60.

Questions in this assignment are based on tables used in SQL Exercise Part 3 & 4. You are not required to finish all steps in the exercise to answer the following questions, and all you need to do is to copy the tables into your own database (the first step in both exercises). However, lessons learned in the exercises might be helpful in answering some questions.

1. Based on the tables we used in SQL Exercise Part 3 (i.e., CEOs, Companies, Countries), write SQL queries to answer the following questions. Please **include both your SQL query and result** (*screenshots are fine*) in the answer [5 points each; 10 points in total for CEE 412 students; 20 points in total for CET 522 students].

   Note: test your answer by execute the query in your own database.

1) Find all countries with higher population than 200,000,000. Your result should be a table of country names by population. Try convert your population values to some more readable format (e.g., use "0.32 Billion" to represent 320,314,000).

2) Find the market value of the company whose CEO receives the highest one year payment. Also include the company name, the CEO name, and the corresponding payment in your result.

3) [For CET 522 Students ONLY] The level of development for each country is determined by the following criteria:

   *A country is categorized as a **Developed** country if its GDP Per Capita is higher than $12,000, otherwise it is categorized as a **Developing** country.*

   Add a new column into the Countries table to describe the level of development for each country. Values in that column should be either **Developed** or **Developing** (no need to show query result for this question, just paste your query statement here).

4) [For CET 522 Students ONLY] In each level of development, count the number of countries and calculate the average GDPPC. Make use of the new column you just created (don't worry about the result, neither the criteria nor the GDPPC data are completely correct).

2. Based on the tables we used in SQL Exercise Part 4 (i.e., Cyclists, Locations, Weather), write SQL queries to answer the following questions [5 points each; 15 points in total].

1) Find a primary key in each table. Compose SQL queries to define the primary keys (no need to show query result for this question, just paste your query statements here).
   Note that you should already established the primary key for the Locations table if you have finished the SQL Exercise Part 4.

2) Establish a foreign key constraint between Weather and Locations table (no need to show query result for this question, just paste your query statements here).
   Note that this should be a constraint on the Locations table, meaning that all dates you have in your Locations table must exist in the Weather table.

3) Try to insert the following row into the Locations table. What happens when you insert this row? Why?

```
INSERT INTO Locations(LocationID, Road, Direction, Date)
VALUES (50, 'Stevens Way & Benton Ln', 'NB', '2017-02-08')
```

3. Using the same tables (Cyclists, Locations, Weather) in the last question, write queries to answer the following questions [25 points]:

1) What is the total cyclist count for each hour of the day on date 2010-09-23 (across all locations where data have been collected on that day)?
   Hint: use DATEPART(HOUR, TimeInt) to return the hour part of the TimeInt column.

2) Building on the previous query, write a query to return only the rows with hourly count greater than 300.

3) What is the hourly cyclist count for days with minimum temperature below 10 degrees?

4) Over all days, what is the average cyclist count during 8:00 – 8:59 AM?
   Note: There are multiple ways to answer this, one way would be to write a subquery to return the 8AM count grouping by date. Then, write an outer query to find the average count over all days.

5) What is the fraction of male cyclists?