ERG3010 Project: Highway Bridges in U.S.

Purpose

-Review database management concepts and practice SQL programming skills with real data.

Introduction

-Based on the understanding of the raw data table, find out an interesting aspect or some easily overlooked hidden information, for instance, cost, safety, relationship among structure type and maintenance, relationship among owner and vendor etc., set a goal to derive certain valuable information by designing a small database and specific queries.

-You can use a part of attributes and records of the big raw data table to reconstruct a smaller database which might consist of several tables. Sometimes, you may need to create a new table that contains data from other sources to provide additional information for your new database.

Assessment scheme **: 15% weight

-This project is the assignments of the last four weeks.

Schedule

Start from Oct 12

Week 1st, Preparation

-To form a team with 3-4 students. (each team preferably has both CS and DS students)

-MySQL environment installation

-Raw data importing and cleaning

- Analyze supporting materials, attributes reading and understanding

Week 2nd-3rd, Conceptual design

-Objectives/goals setting

-Attributes selection and logical designing with ERD (hand drawing or using MySQL workbench to illustrate the diagram)

-The new database establishing and structure optimization

-Using extra raw data download from Internet for designing a specific DB is OK.

Week 4th -5th Coding and database performance evaluation

-Queries coding and optimization

-Queries' performance evaluation

Week 6th, **Presentation**, <**15min**

-Due day***: Submit the PPT of your team before **8 pm, Nov 25**.

-Pre day***: **Nov 26-27**

-PPT should contains Background, BD System Introduction, Objective, ERD, Queries Code, Results, Performance evaluation, and further work, and Work assignment and contribution of team members

-Run your SQL queries to get the result after you presentation

-It will be better to chose one of team member to give the presentation

-You can use your own PC.

-If your project has finished ahead of schedule, you can give your presentation in any class

before the due day. Pls inform the instructor in advance. **

Basic Objectives/goals of the project***

-To get the year for every state in US when the largest number of bridges were built in that state.

-To find out the main factors about bridges' maintenance/improvement cost, e.g. structure type, condition of main parts, length, location, traffic, etc., and prove it with information derived from DB quantitatively.

Evaluation

-Accomplish the above two basic goals with correct ERD and Queries in detail base on concepts and skills you have learned from this course, 50%. **

-Define a new interesting problem/request, and fulfill it with feasible queries to get a valuable information, 25%

-Design elaborated queries to handle tough query requests/objectives with impressive efficiency or effectiveness, 15%

-Apply extra knowledge or skills beyond this course, 10%

Raw data

It is the annual inspection data of all highway bridges in U.S. provided by Federal Highway Administration (FHWA), Department of Transportation, U.S.. The related docs and the database are downloaded from the following website:

https://www.fhwa.dot.gov/bridge/nbi.cfm

https://www.fhwa.dot.gov/bridge/nbis.cfm

You can use the updated data of 2016, or other historical data of the last two decades.

Guide book

This document introduce all attributes involved, you should read it carefully..

https://www.fhwa.dot.gov/bridge/mtguide.pdf

And there are lots of useful information or reading materials you can find from the website.

Hints

-The most difficult part of DB designing is understanding the system you are facing and organizing the untreated/semistrutured raw data.

-For Data Science students

-SQL can help you derive information from many tables based on relationships among them

-You can use R, MATLAB or other statistic tools to show the result graphically, but do not use them to do the data analysis part.

-For Computer Science students

-Try to enhance the query efficiency as much as possible

-You can use Python, Java, C# or other language to design an UI of your database, but do not spend too much time on it.