QMB 6358: Software Tools for Business Analytics
Department of Economics
College of Business
University of Central Florida
Fall 2023

Assignment 7

Due Monday, November 20, 2023 at 11:59 PM
in your GitHub repo.

Instructions:

Complete this assignment within the space on your GitHub repo in a folder called `assignment_07`. When you are finished, submit your code by pushing your changes to your GitHub repo, following the instructions in Question 3. You are free to discuss your approach to each question with your classmates but you must `git push` your own work.

Question 1:

Create a database `TractorDataBase.db` to store tables of information related to the sale of used tractors. The data are are collected into three tables. The first table, called `tractor_sales.csv` contains the characteristics and conditions of tractors and includes the following variables.

$saleprice_i$ = the price paid for tractor i in dollars
age_i = the number of years since tractor i was manufactured
$enghours_i$ = the number of hours of use recorded for tractor i
$johndeere_i$ = an indicator of whether tractor i is manufactured by John Deere
$spring_i$ = an indicator of whether tractor i was sold in April or May
$summer_i$ = an indicator of whether tractor i was sold between June and September
$winter_i$ = an indicator of whether tractor i was sold between December and March

You also have access to another dataset `tractor_specs.csv`, which an intern has painstakingly compiled from research about the models of tractors that were sold.

$horsepower_i$ = the horsepower of tractor i
$diesel_i$ = an indicator of whether tractor i runs on diesel fuel
fwd_i = an indicator of whether tractor i has four-wheel drive
$manual_i$ = an indicator of whether tractor i has a manual transmission

The above datasets did not indicate whether or not the tractors had enclosed cabs, which would be convenient for the operator. Now suppose that a coworker obtained the data by inspecting photographs from the advertisements and saved a list of tractors with enclosed cabs in the file `tractors_with_cabs.csv`.

a) Write an SQL script that defines the schema for each of the three tables TractorSales, TractorSpecs, and TractorCabs, and reads in the tables. It may help to create temporary files without headers.

b) Write an SQL query that calculates the minimum, average, and maximum price of tractors and separately tabulates these figures for tractors made by John Deere and the others.

c) Write an SQL query that joins the data in TractorSales to those in TractorSpecs by SaleID.

d) Write an SQL query that joins the data in TractorSales to those in TractorSpecs, and then adds a binary variable to the joined table to indicate whether each tractor has an enlosed cab or not.

e) Write an SQL query that calculates the minimum, average, and maximum horsepower, the minimum, average, and maximum price of tractors, and separately tabulates these figures for tractors with different levels of the following variables: whether the tractors run on diesel fuel, whether the tractors have four-wheel drive, whether the tractors have a manual transmission, whether the tractors were made by John Deere, and whether the tractors have an enclosed cab.

Question 2:

Now that you have created the SQL commands for Question 1, use these to perform the analysis on other platforms.

a) Perform the SQL queries in R.

   i) Read the csv files into R as you would for any other data frames.

   ii) Use the sqldf package to execute the queries.

b) Perform the SQL queries in python.

   i) Read the csv files into python and insert the records into tables following the procedure in the example in demo_17_SQL_w_Python_Advanced in the course repository QMB6358F23.

   ii) Use the sqlite3 module to execute the queries.

Question 3:

Push your completed files to your GitHub repository following these steps. See the `README.md` and the `GitHub_Quick_Reference.md` in the folder `demo_02_version_control` in the QMB6358F23 course repository for more instructions.

1. Open GitBash and navigate to the folder inside your local copy of your git repo containing your assignments. Any easy way to do this is to right-click and open GitBash within the folder in Explorer. A better way is to navigate with UNIX commands.
2. Enter `git add .` to stage all of your files to commit to your repo. You can enter `git add my_filename.ext` to add files one at a time, such as `my_filename.ext`. in this example.
3. Enter `git commit -m "Describe your changes here"`, with an appropriate description, to commit the changes. This packages all the `add`ed changes into a single unit and stages them to `push` to your online repo.
4. Enter `git push origin main` to push the changes to the online repository. After this step, the changes should be visible on a browser, after refreshing the page.