QMB 6358: Software Tools for Business Analytics

Department of Economics
College of Business
University of Central Florida
Fall 2023

Assignment 4

Due Monday, October 9, 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_04`. In this folder, save a copy of the sample file called `A4Q1_data.sh` that will contain your shell script for Question 1, and a copy of the sample file called `A4Q2_data.sh` that will contain your shell script for Question 2. Samples are available in the `assignment_04` folder within the code repository QMB6358F23.

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:

Recall the repository QMB6358_car_auction_data that contains 96 `.csv` files, which contain the records of monthly used car auctions over the years 2010 to 2017. You will revisit the task of collecting these files to form one large dataset and print some summary statistics, as before, except this time you will organize the data files into the dataset using UNIX commands. .

Complete the shell script `A4Q1_data.sh` to assemble the dataset in two different ways.

a) Use the `cat` command within a for loop to output the full dataset in the file `A4Q1a_full.csv`.

b) Use the `cat` command in a single line to output the full dataset in the file `A4Q1b_full.csv`.

c) Add some commands to the script `A4Q1_tests.R` to test the two datasets. Use commands similar to those at the bottom of `A3Q1_data.R` for testing the dataset in Question 1 of Assignment 3. Add a line at the bottom of `A4Q1_data.sh` to run this script and output the results to `A4Q1_results.out`.

Finally, running the script `A4Q1_data.sh` will assemble the datasets and calculate `summary` statistics and a `table` of observations for each method in both versions of Question 1 from Assignment 3 and 4. Use the output to verify that the datasets from this Assignment 4, Question 2 are the same as those from Assignment 3, Question 2.

Note: Do not save a large dataset in a folder within your GitHub repository for your assignments. The GitHub repository is designed to store text files with code but not data; there is a limit on the files size, as well. Please save any large files to a location on your computer outside of the repository.

Question 2:

Continue the project of building a model of the value of used airplanes in an online marketplace, using the same three `.csv` files in the folder `assignment_03`, relating to the sales and characteristics of airplanes. As before, your job is to join these files to form one full dataset and print some output from a regression model. This time, however, you will organize the data files into the dataset using UNIX commands. Complete the shell script `A4Q2_data.sh` to assemble the dataset in two different ways.

a) Use the `paste` command to join the datasets into the file `A4Q2a_full.csv`.

b) Use the `join` command to join the datasets into the file `A4Q2b_full.csv`.

c) Add some commands to a script you will call `A4Q2_tests.R` to test the two datasets. Use commands similar to those at the bottom of `A3Q2_data.R` for testing the dataset in Assignment 3, Question 2, which will read in the full dataset and estimate the full model. Add a line at the bottom of `A4Q2_data.sh` to run this script and output the results to `A4Q2_results.out`.

Finally, running the script `A4Q2_data.sh` will join the datasets and verify that the datasets are formed correctly by running the code from the bottom of `A3Q2_data.R` (again, after copying that code into a new script called `A4Q2_tests.R`) and checking that the commands `lm` and `summary(lm_model_1)` print output without errors. Use the output to verify that the datasets from this Assignment 4, Question 2 are the same as those from Assignment 3, Question 2.

Note: These datasets are small enough that it is fine to save them within a code repository.

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.