QMB 6358: Software Tools for Business Analytics
Executive Development Center
College of Business
University of Central Florida
Fall 2020

Assignment 4

Due Wednesday, October 7, 2020 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 your answer to Question 1 in a file called `A4Q1_data.R`. In the same folder, save 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 QMB6358F20.

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:

The folder `assignment_04` contains two `.csv` files. Your job is to join these files to form one full dataset and print some output from a regression model. Use the file `A4Q1_data.R` as a starting point. Complete it in stages by following these steps:

a) Read in the `tractor_sales.csv` dataset and store it in a data frame called `tractor_sales` in memory in the R workspace.

b) Read in the `tractor_specs.csv` dataset and store it in a data frame called `tractor_specs` in memory in the R workspace.

c) Use the `merge` command to join the datasets to form the data frame `tractor_full` in memory in the R workspace.

d) Verify that the dataset is formed correctly by running the script at the bottom of `A4Q1_data.R` and checking that the commands `lm` and `summary(lm_model_1)` print output without errors.

Question 2:

Now 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 the script `A4Q2_tests.R` to test the two datasets. Use commands similar to those at the bottom of `A4Q1_data.R` for testing the dataset in Question 1. 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 `A4Q1_data.R` and checking that the commands `lm` and `summary(lm_model_1)` print output without errors. Use the output to verify that the datasets are the same in both Questions 1 and 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_04_version_control` in the QMB6358F20 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 master` to push the changes to the online repository. After this step, the changes should be visible on a browser, after refreshing the page.