

## LARSON—INFO 790—CLASSROOM WORKSHEET 08

### A Modified Example

We will modify the Titanic example on Steve Essi's data set.

#### 1. Log in to VCU's Athena cluster.

The following directions assume you have an Athena account, that you have set up Sage, and that you have set up (using `make`) the CONJECTURING program.

- (a) Start the Chrome browser.
- (b) If you are off-campus, you'll need to connect to the VPN first.
- (c) Then go to `https://athena3.hprc.vcu.edu`
- (d) Login using your VCU EID as your username, and your corresponding VCU password.
- (e) Click the Apps button and a Sage session. The default options are fine. This will take a couple of minutes.
- (f) Click the Apps button and start an "athena shell access" session (this will give you a terminal window, where we can issue commands).
- (g) Your Sage session will first say "Queued", then "Starting". When it is ready you will see a button that says, "Connect to Sage". Click that.
- (h) You should then get an "untitled" interactive-Python notebook (ipy nb), or the last file you had open the previous time you used Athena.
- (i) When your notebook opens look on the upper-right to make sure the SageMath kernel is running (if it isn't you can change the *kernel*).

2. **Reminders for setting up Expressions and Conjecturing.** In each case, for each experiment, we will make a folder in your root directory; we will need a copy of the "expressions" compiled executable in that folder; and we will use an `.ipy nb` located in that notebook. When we call the CONJECTURING program we will use the version in the `~/conjecturing` folder downloaded from github (what you did with the github command; if there are ever new files on github, using the command `git pull` will update your files).

#### 3. Setting up the c08 example on Athena

- (a) In your Athena shell tab, make a `C08_experiment` directory;  
run: `mkdir C08_experiment`
- (b) Check that you have this directory by running the directory command: `ls`
- (c) Change into your `C08_experiment` directory: `cd C08_experiment`
- (d) Copy the expressions file to your `C08_experiment` directory:  
`cp ~/conjecturing/c/build/expressions ./`

#### 4. Getting the c08 data and worksheet.

- (a) Go to <https://math1um.github.io/Teaching/>
- (b) Scroll down and find the INFO 790 files. Download the c08\_data.xlsx (excel) and .ipynb (script) files.
- (c) These two files need to be in your C08\_experiment directory on Athena. On your Athena Jupyter notebook, there is a button for *uploading* files. Upload the c08 .xlsx and .ipynb files. Check that they are in the C08\_experiment directory.

#### 5. Notes from Steve Essi

Here's what we're told about the data:

- (a) The following variables are all continuous (so they will be used in invariant-relation conjectures): "Tx\_R\_O1", "N\_Tx\_Ctr1", "N\_R\_Ctr1", "N\_Center", "M\_R\_O1", "M\_R\_ST1", "N\_PT\_TRANS\_S", "PT\_TRANS\_T1", "N\_PTSURV\_SUM", "N\_PT\_HOS\_S", "VAVF\_F", "N\_DP\_HGBD12", "N\_DP\_HGBD", "HOMEHD1", "PD1", "HD1", "TOTSTAS".
- (b) The following variables are categorical: (so they will be used in property-relation conjectures): "SMR", "SHR", "HDKTVPM12\_F", "PTSURV\_T", "PT\_HOS\_T".
- (c) The *targets* are "PTSURV\_T" and "PT\_HOS\_T".

#### 6. Loading the c08 script

- (a) In your open Sage session tab, you will see a list of files on the left. Double-click on your C08\_experiment folder, then double click on c08\_experiment.ipynb.
- (b) Check on the upper-right that Sage is the kernel. If not, change it to Sage.
- (c) Run the commands one cell at a time (go to the first cell, and click RUN, or SHIFT-ENTER).

#### 7. Test Data

- (a) Note something new at the end of this worksheet: we name and evaluate a specific conjecture for all of the testing data.
- (b) What is the name of the testing examples?
- (c) Where are the invariant conjectures stored?
- (d) Where are the property conjectures stored?
- (e) Only the 0th conjecture is investigated. Now investigate the 1th conjecture.

#### Final Note

Dr Brooks wrote this file so that it could be easily imitated for a wide-variety of tabular data-files. You should read each cell carefully and **ask questions** about what the commands do. **You will be doing this with your own data files.**