ANLT5030 – Unit 8 Assignment 1 Tutorial

SAS Studio

Instructions

- Prepare the data set for analysis and identify variables that could be used in analysis.

- Run a logistic regression on the data, as appropriate, in order to respond to the question.
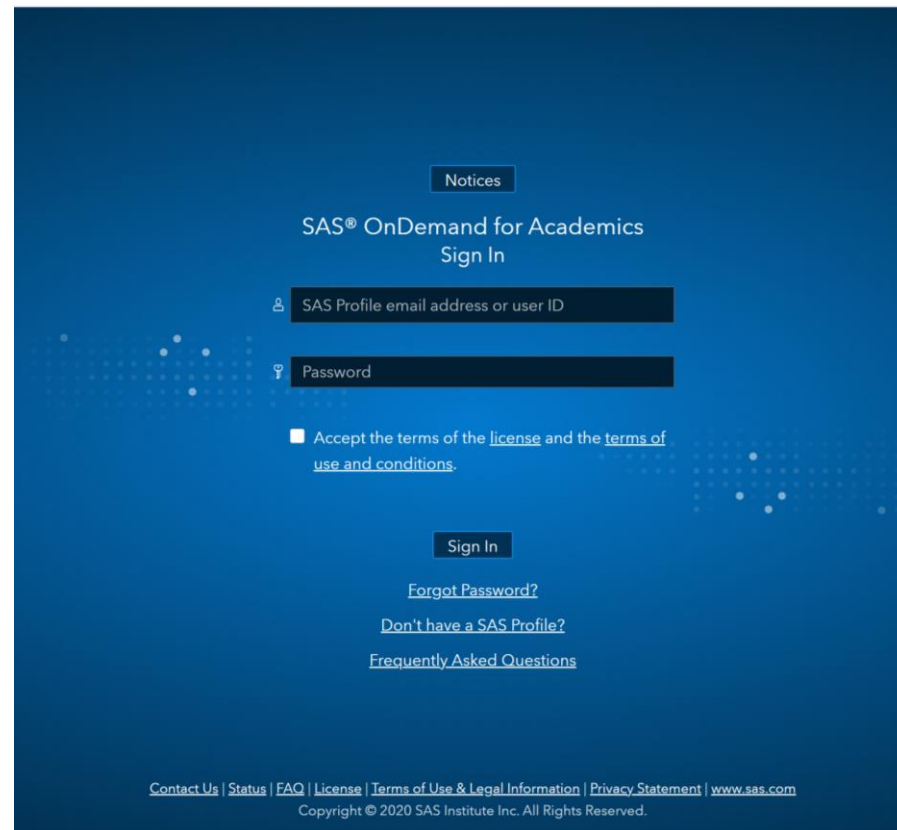
Dataset

- Download the data file provided in the Unit 8 Welcome announcement in the course announcements. (https://archive.ics.uci.edu/ml/datasets/Bank+Marketing)

- Reference for dataset:

    S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

- Instructions specific to the dataset (from the Unit 8 announcement):

    - Please fit a logistic regression model on this dataset to predict if the client will subscribe (yes/no) to a term deposit (variable y). Note that you may need to translate this into a binary (0/1) variable in order to do so.

Access the SAS OnDemand for Academics Control Center

https://odamid.oda.sas.com/SASODAControlCenter

SAS OnDemand for Academics (SODA) Control Center

Click on Files(Home)

© Stefanie G. Reay, MS, PhD, Capella University, 2021

The Upload button will display in dark blue
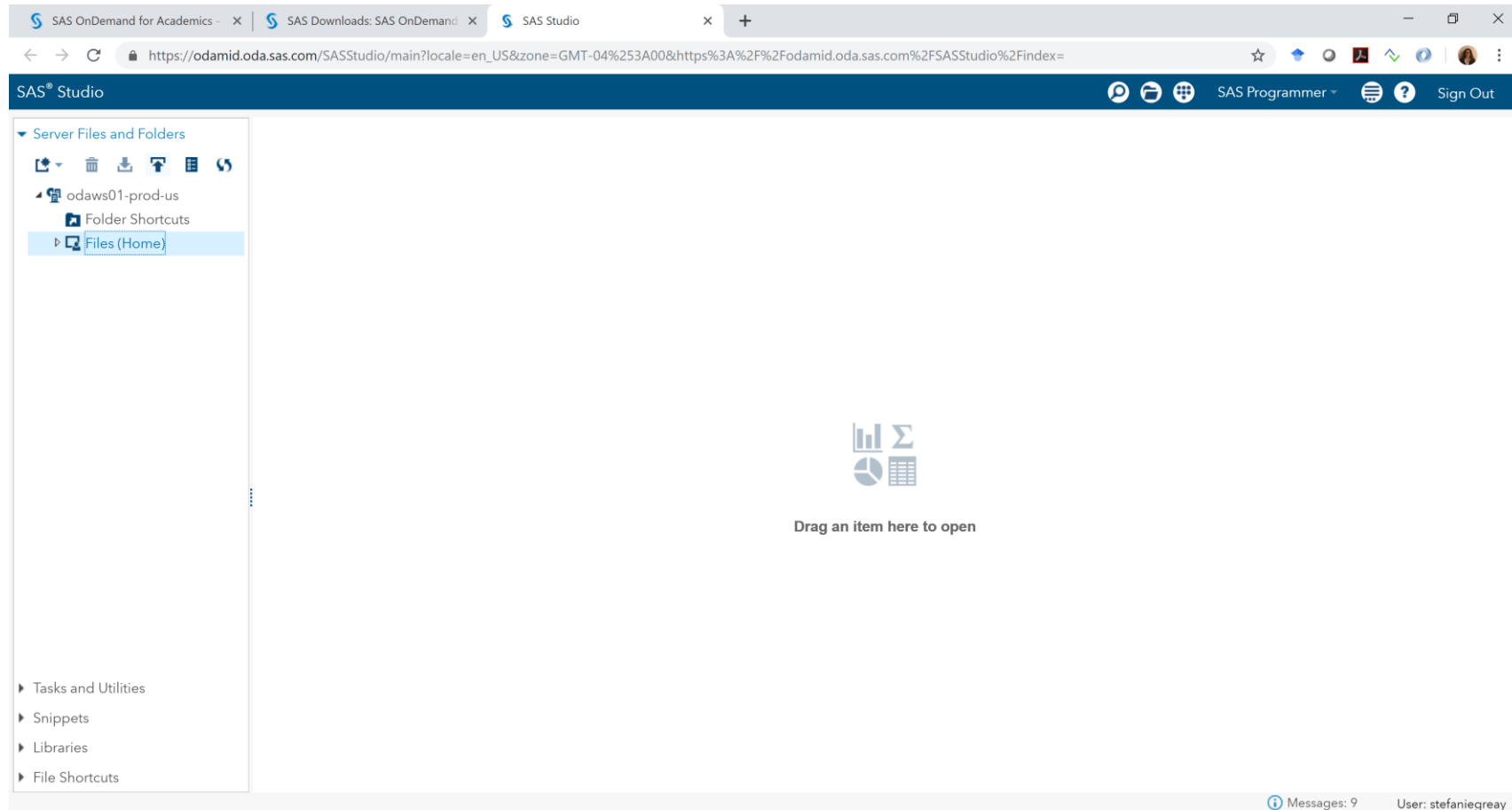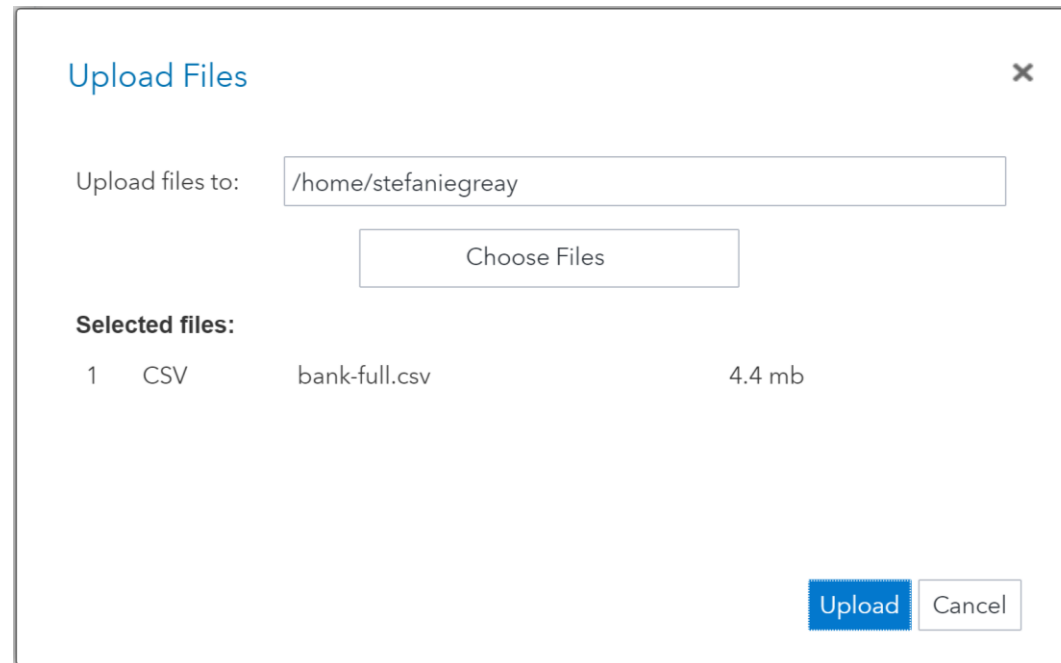
You can create a folder at this point, if you wish, or simply upload to your home directory.
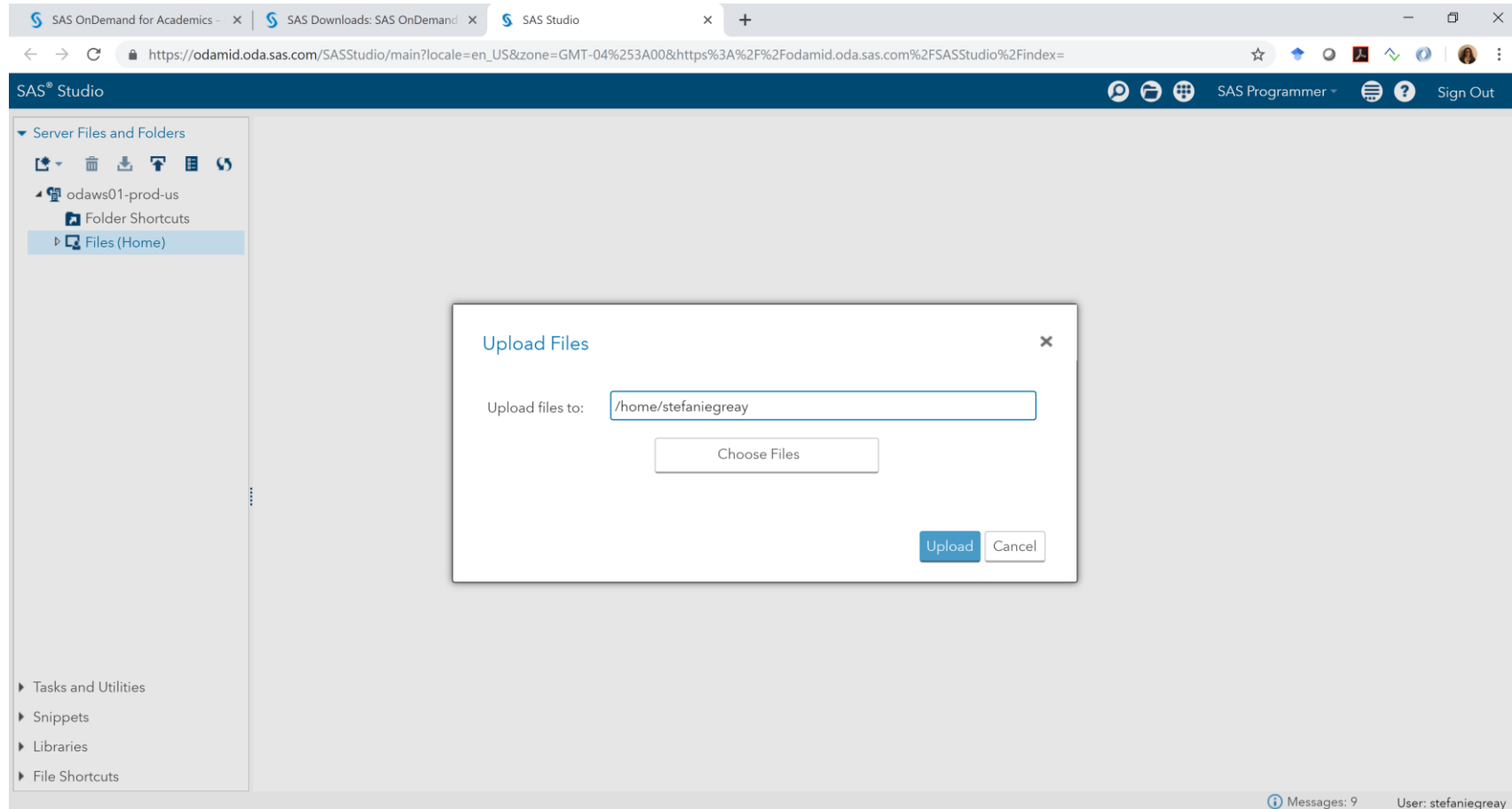
Select "Choose Files" to browse your computer for the dataset you want to upload. Once the dataset has been selected, click "Upload."

You will be able to view your files by clicking on "Files(Home)" to verify that your file successfully uploaded.
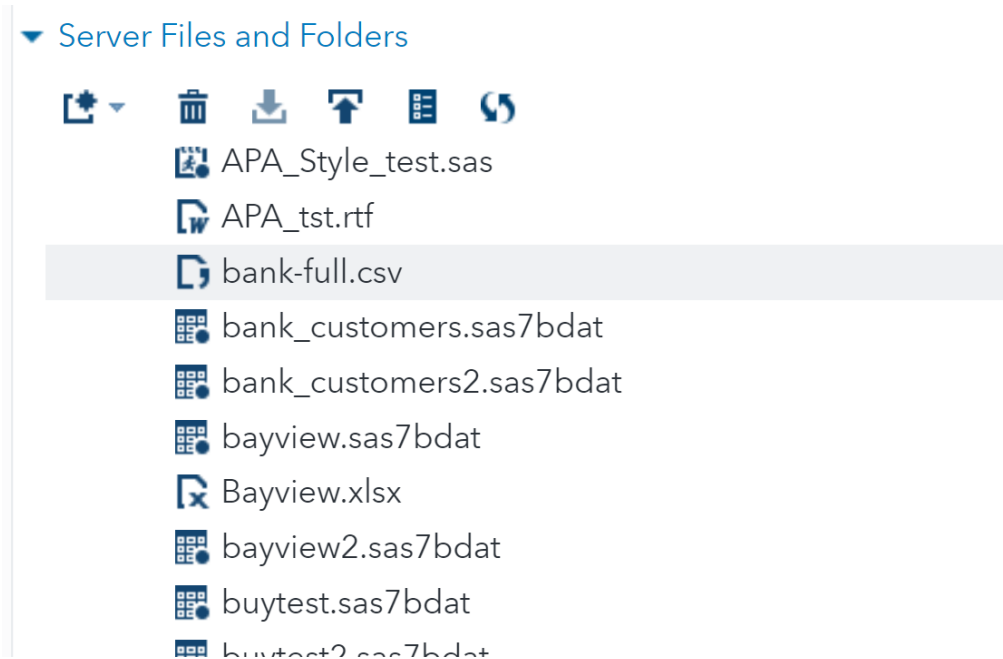
To get started with the SAS portion of the assignment, start a new SAS program.
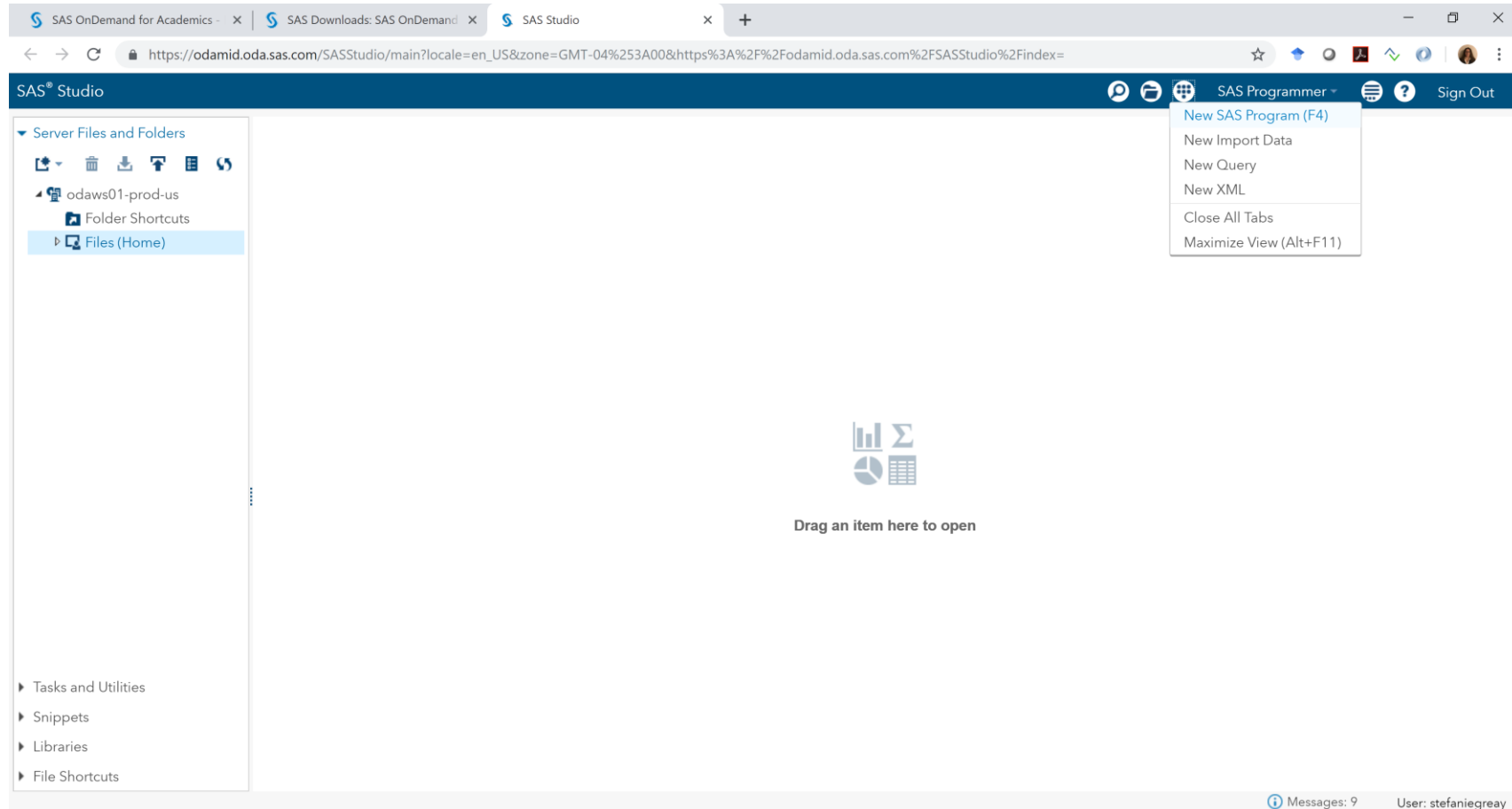
Import the dataset into a SAS dataset format (from the current csv format)
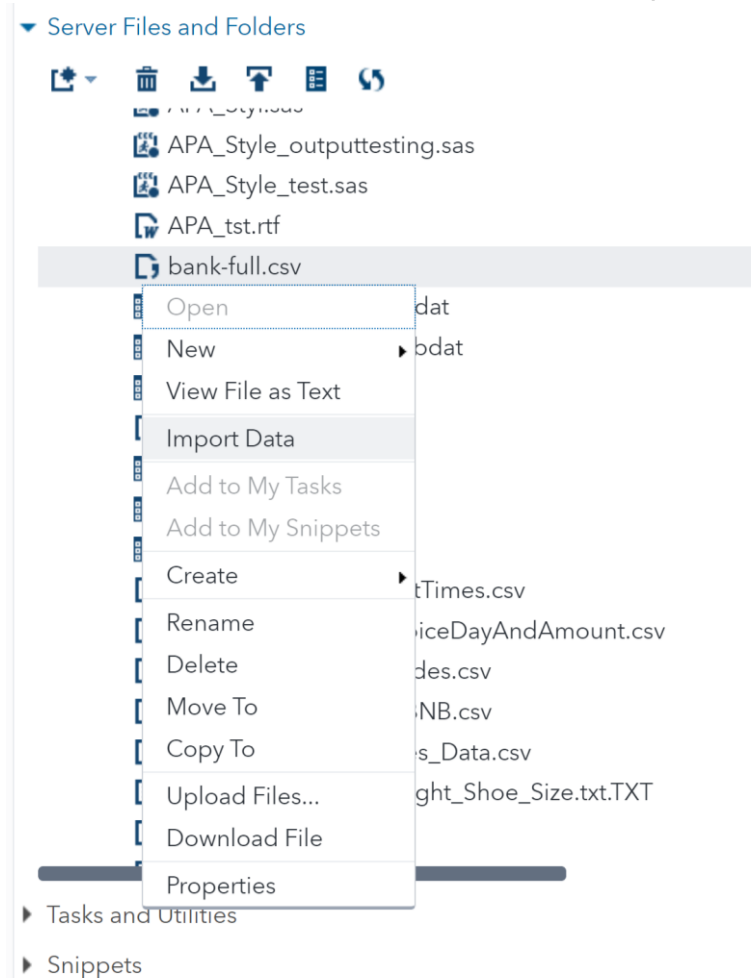
The Proc Import code will be written for you (save this as a template to use for future imports!)

NOTE: you have to run this code for the data to actually import.

You will need to adjust some of the settings to tell SAS that the file is separated by semicolons and that there are quotes around them.

Changes:
1) The file type needs to be "DLM (Delimited file)."
2) The delimiter should be a semicolon (i.e. ;).
3) The "Quote delimiter value" option should be checked.

© Stefanie G. Reay, MS, PhD, Capella University, 2021

To run the code, click the icon that looks like a guy running.
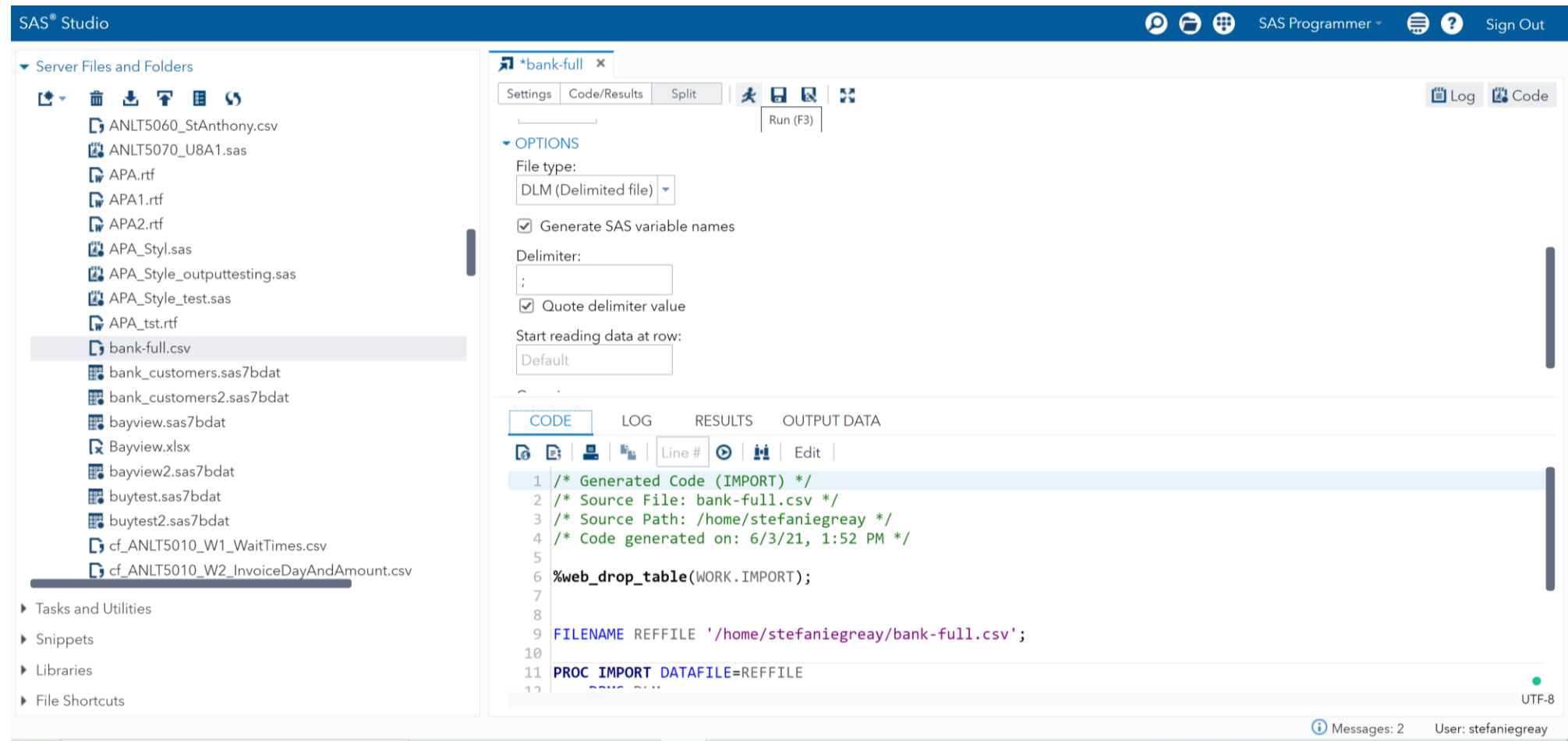
To create a SAS Library for your Files(Home) folder, you need to use a libname statement

You can determine what the path is for your Files(Home) Folder by right clicking on Files(Home) and selecting Properties. (Yours will not be the same as mine – it will include your userID, not mine.)

© Stefanie G. Reay, MS, PhD, Capella University, 2021
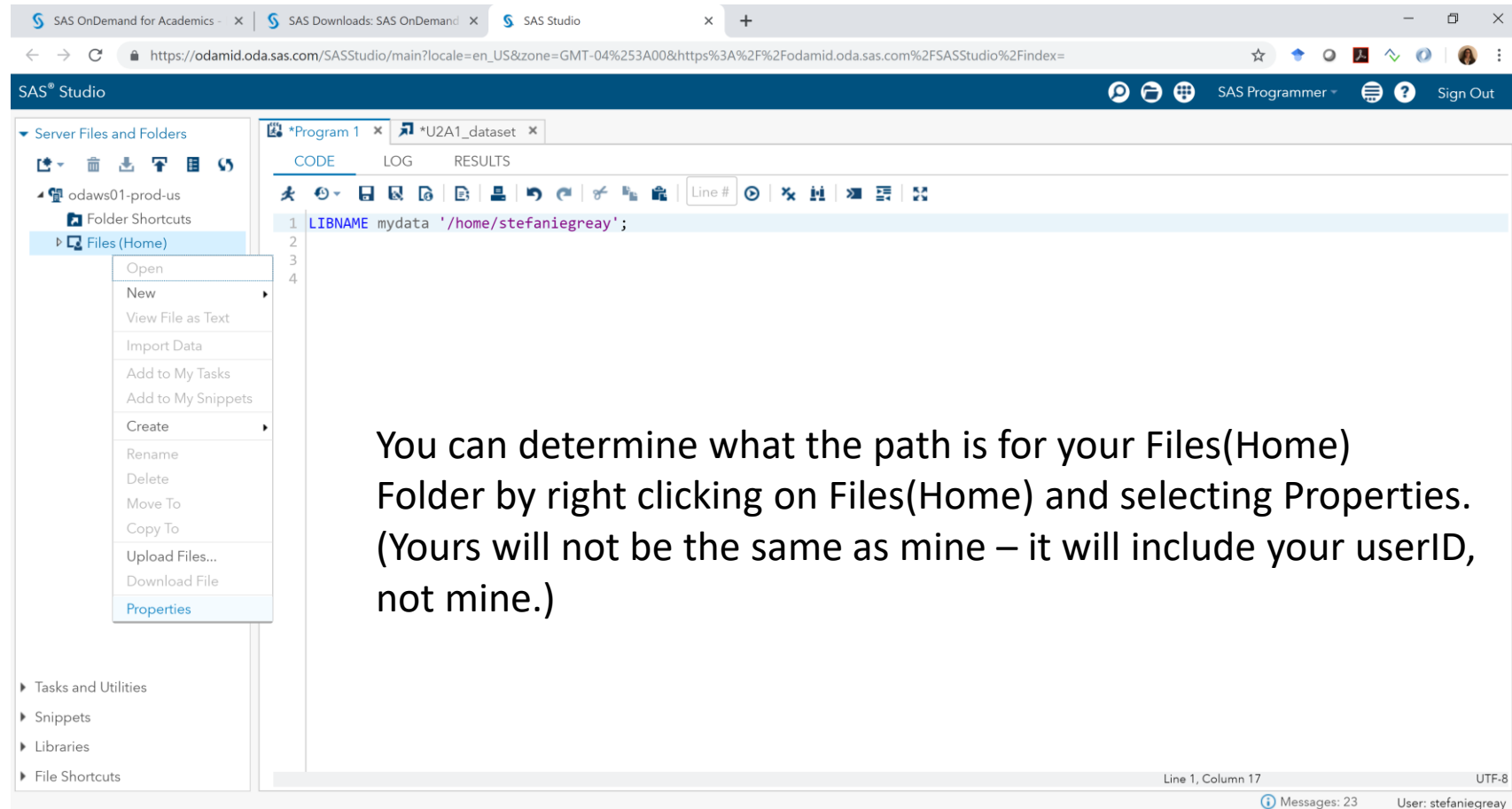
Save the temporary SAS datasets created by the import to your library using the following sample code.

You will need to make sure your libname represents
Your USERID, not mine. The rest should be the same.
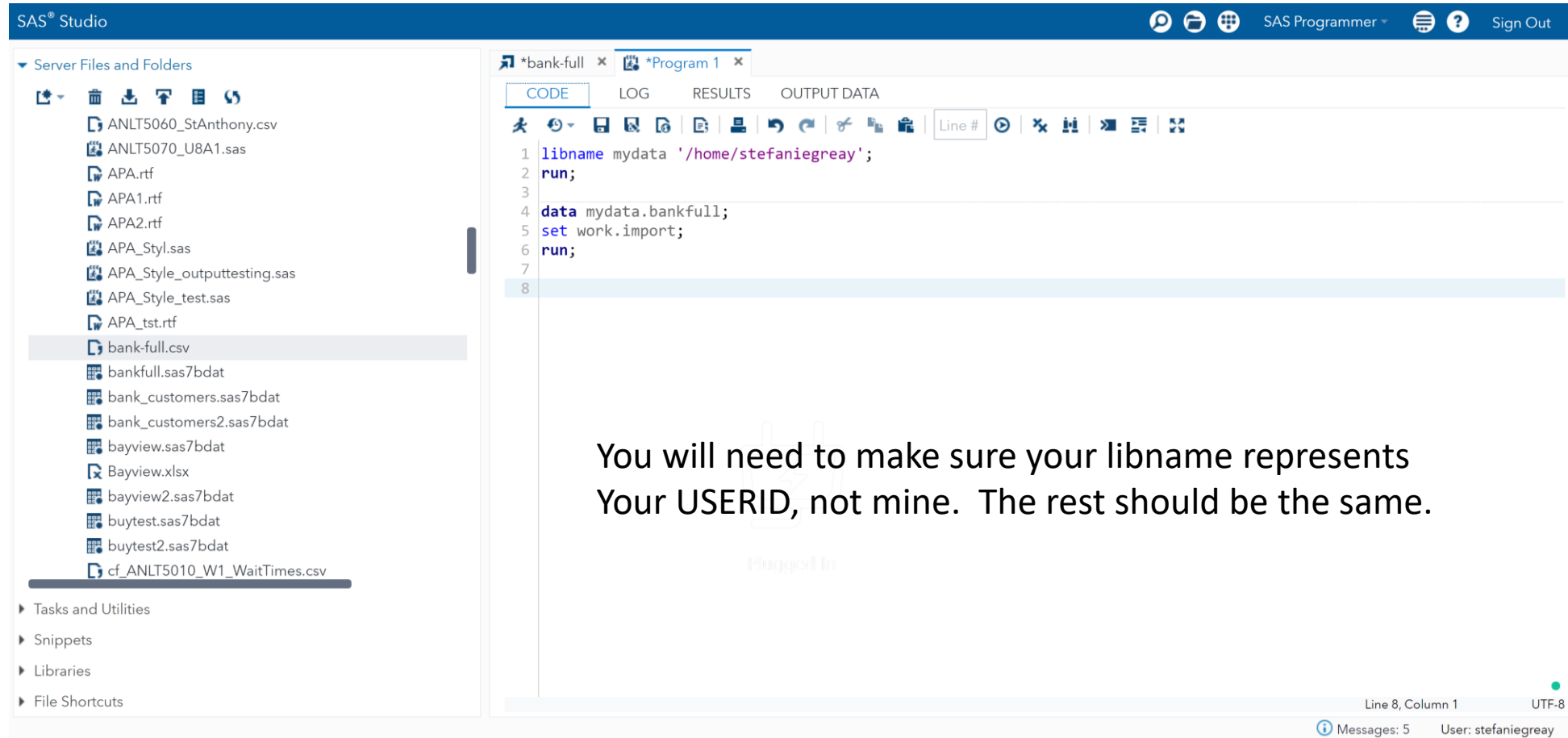
When you run the code, you will see the dataset in the ouput data window and can verify its success.

You can now run any procedures against that dataset via the code window.


```
    *bank-full  ×      *ANLT5030_U8A1_sample_code.sas  ×

    CODE        LOG        RESULTS        OUTPUT DATA

 1  libname mydata '/home/stefaniegreay';
 2  run;
 3
 4  data mydata.bankfull;
 5  set work.import;
 6  run;
 7
 8  proc freq data=mydata.bankfull;
 9  tables campaign contact default education housing job loan marital month outcome y;
10  run;
11
12  proc univariate data=mydata.bankfull;
13  var age balance day duration pdays previous;
14  run;
15
16  proc corr data=mydata.bankfull;
17  var age balance day duration pdays previous;
18  run;
19
20  proc logistic data=mydata.bankfull descending;
21  class campaign contact default education housing job loan marital month poutcome;
22  model y= age balance day duration pdays previous
23  campaign contact default education housing job loan marital month poutcome;
24  run;
```

© Stefanie G. Reay, MS, PhD, Capella University, 2021

Sample Code for the assignment

```
libname mydata '/home/stefaniegreay';
run;

data mydata.bankfull;
set work.import;
run;

proc freq data=mydata.bankfull;
tables campaign contact default education housing job loan marital month poutcome y;
run;

proc univariate data=mydata.bankfull;
var age balance day duration pdays previous;
run;

proc corr data=mydata.bankfull;
var age balance day duration pdays previous;
run;

proc logistic data=mydata.bankfull descending;
class campaign contact default education housing job loan marital month poutcome;
model y= age balance day duration pdays previous
campaign contact default education housing job loan marital month poutcome;
run;
```

Note that this proc logistic is just a sample and includes all variables...you will need to check the assumptions and significance of contributions, and adjust the variables in the model accordingly.

Additional Resources for Logistic Regression and SAS's Proc Logistic

- SAS's Proc Logistic documentation:

  https://support.sas.com/documentation/onlinedoc/stat/141/logistic.pdf

- Institute for Digital Research and Education Statistical Consulting Proc Logistic annotated output:

  https://stats.idre.ucla.edu/sas/output/proc-logistic/

Assumptions for Logistic Regression

Assumptions
1) The response variable must be binary (for simple, binary logistic regression).
2) None or very little multicollinearity exists between the variables.
3) There is a linear relationship between the independent/explanatory variables and the log odds.
4) Large sample size.
5) The observations are independent.