

# 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.



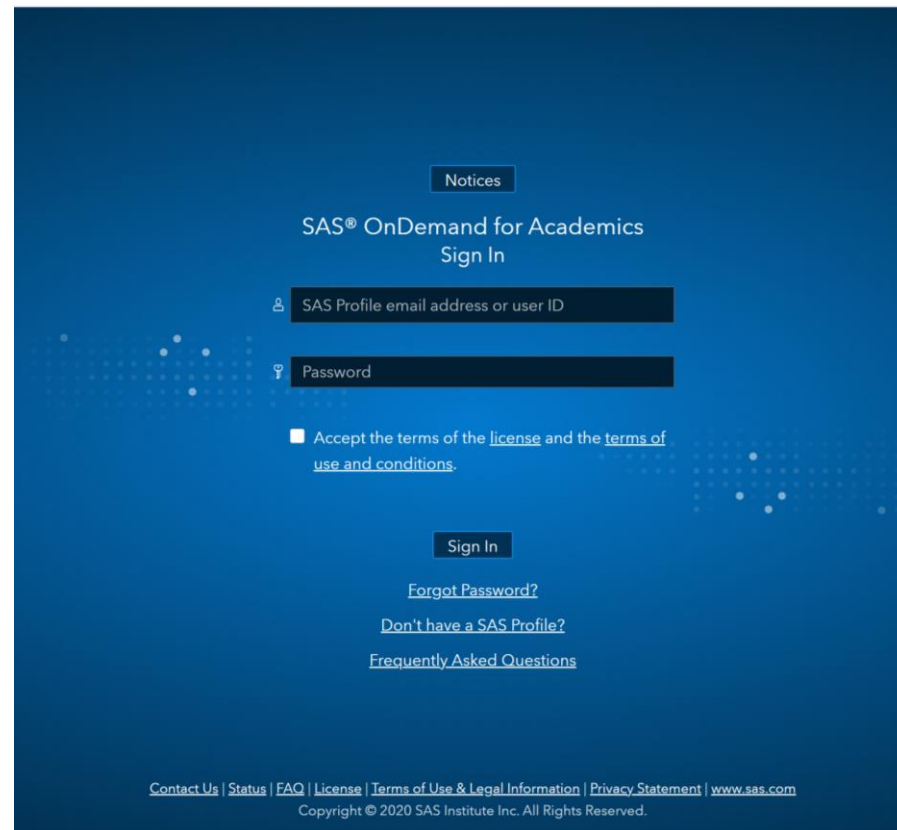
# Dataset

- 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>



The screenshot shows the SAS OnDemand for Academics Sign In page. The background is a dark blue gradient with a subtle pattern of white dots. At the top, there is a 'Notices' button. Below it, the text 'SAS® OnDemand for Academics' and 'Sign In' are displayed. The sign-in form includes two input fields: 'SAS Profile email address or user ID' and 'Password'. Below the password field is a checkbox labeled 'Accept the terms of the [license](#) and the [terms of use and conditions](#)'. A 'Sign In' button is positioned below the checkbox. At the bottom of the form, there are three links: 'Forgot Password?', 'Don't have a SAS Profile?', and 'Frequently Asked Questions'. The footer contains a row of links: 'Contact Us', 'Status', 'FAQ', 'License', 'Terms of Use & Legal Information', 'Privacy Statement', and 'www.sas.com', followed by the copyright notice 'Copyright © 2020 SAS Institute Inc. All Rights Reserved.'



# SAS OnDemand for Academics (SODA) Control Center

The screenshot displays the SAS OnDemand for Academics (SODA) Control Center dashboard. At the top, the SAS logo is on the left, and the user's location (United States) and name (Stefanie Reay) are on the right. The main heading is "SAS® OnDemand for Academics Dashboard". Below this, there are tabs for "Planned Events" and "Notices". A navigation bar includes "Applications", "Enrollments", and "Courses". The "Applications" tab is active, showing a list of SAS products with their descriptions and actions. On the right, there is a "Reference" section with links to the Support Site, Step-by-Step Reference Guides, and Frequently Asked Questions. Below that, the "Quotas" section shows progress bars for the Home Directory (1% of 46.5MB/5120MB) and Course Directory (7% of 207.0MB/3072MB). At the bottom, there is a link to "Other Ways to Access SAS® OnDemand for Academics Resources".

**SAS® OnDemand for Academics Dashboard**

United States | Stefanie Reay

Planned Events | Notices

Applications | Enrollments | Courses

**SAS® Studio**  
Write and run SAS code with a Web-based SAS development environment.  
*Actions:* [Clear my saved tabs.](#)

**SAS® Enterprise Guide®**  
Deliver the power of SAS from an easy-to-use, point-and-click interface. ([Download Required](#))

**SAS® Enterprise Miner™**  
Reveal valuable insights with powerful data mining software. ([Configuration Steps Required](#))  
*Actions:* [Clear my project locks.](#)

**SAS® Forecast Studio**  
Generate large numbers of high-quality forecasts automatically. ([Configuration Steps Required](#))  
*Actions:* [Manage your personal environment.](#)

**JMP® Software access to SAS® hosted servers**  
Statistical discovery software. Users must have a copy of JMP® software. ([Configuration Steps Required](#))

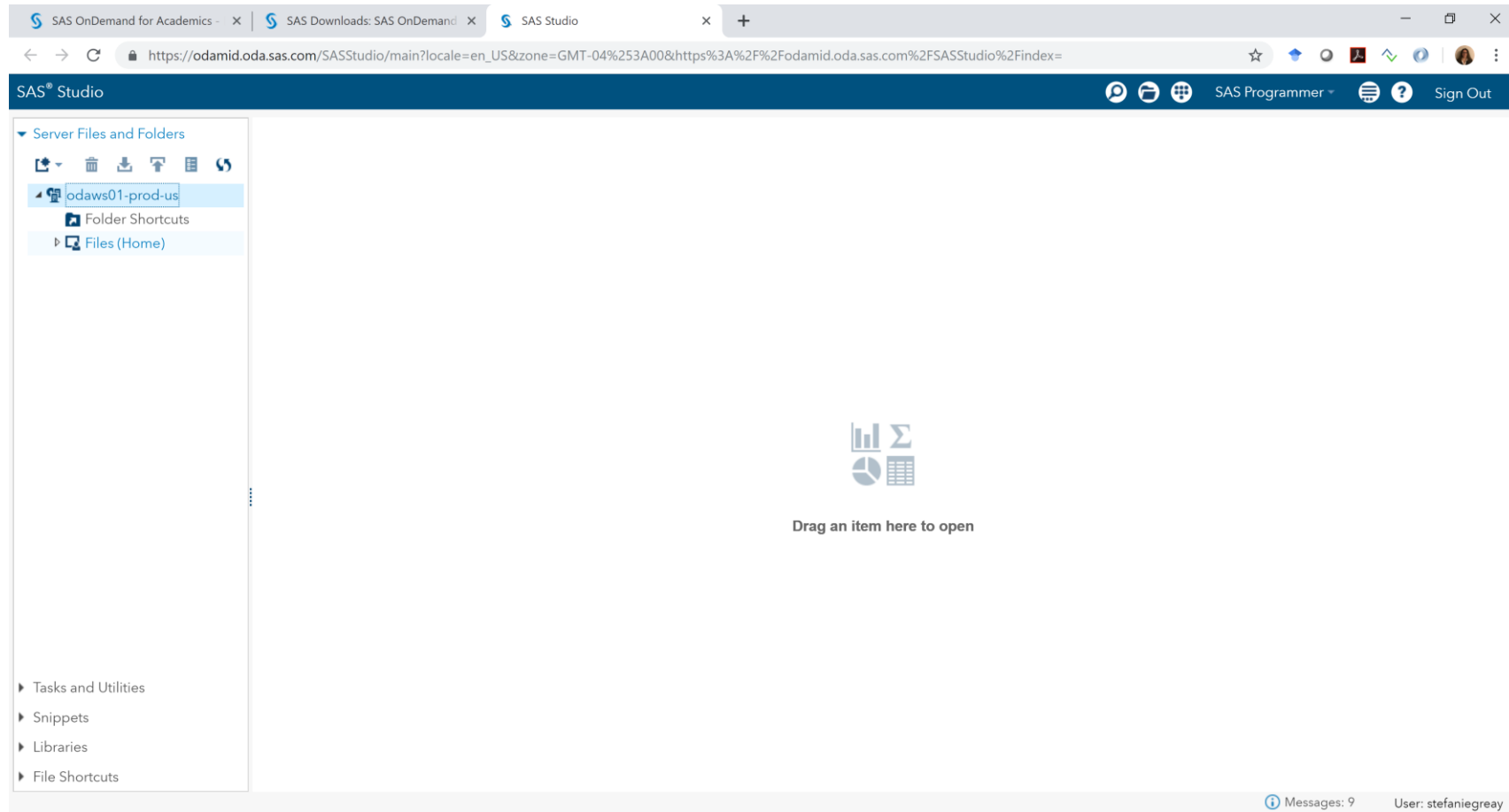
**Reference**  
[Support Site](#)  
[Step-by-Step Reference Guides](#)  
[Frequently Asked Questions](#)

**Quotas ([learn more](#))**  
Home Directory (46.5MB/5120MB)  
1%  
Course Directory (207.0MB/3072MB)  
7%

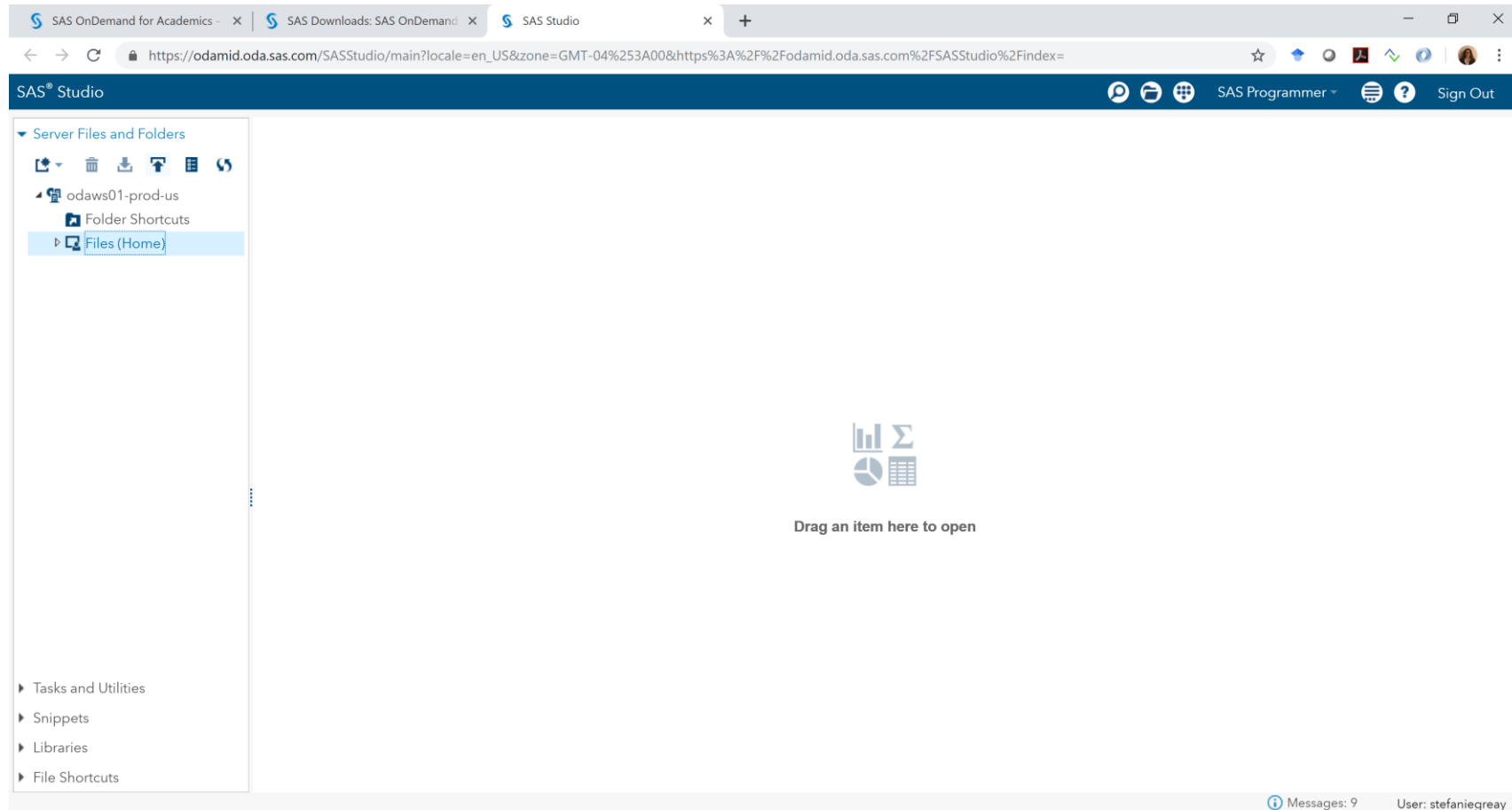
[Other Ways to Access SAS® OnDemand for Academics Resources](#)



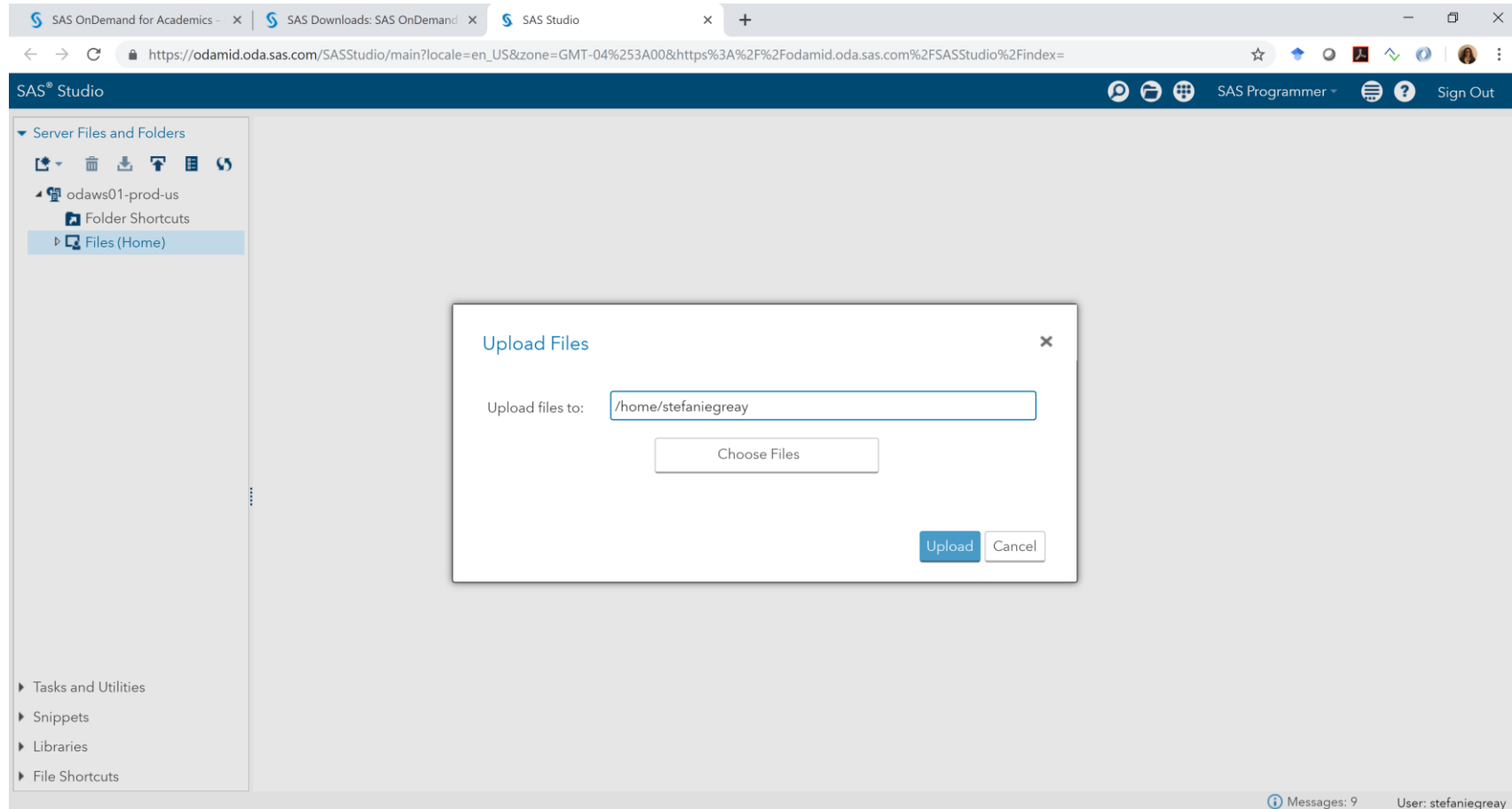
# Click on Files(Home)



# The Upload button will display in dark blue

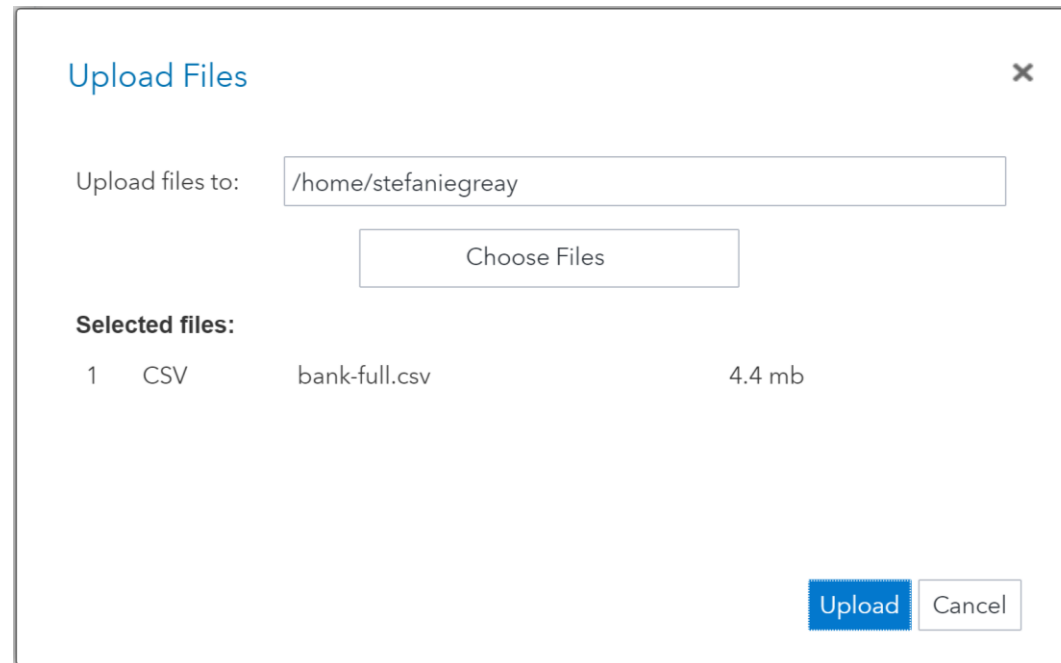


You can create a folder at this point, if you wish, or simply upload to your home directory.





Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”

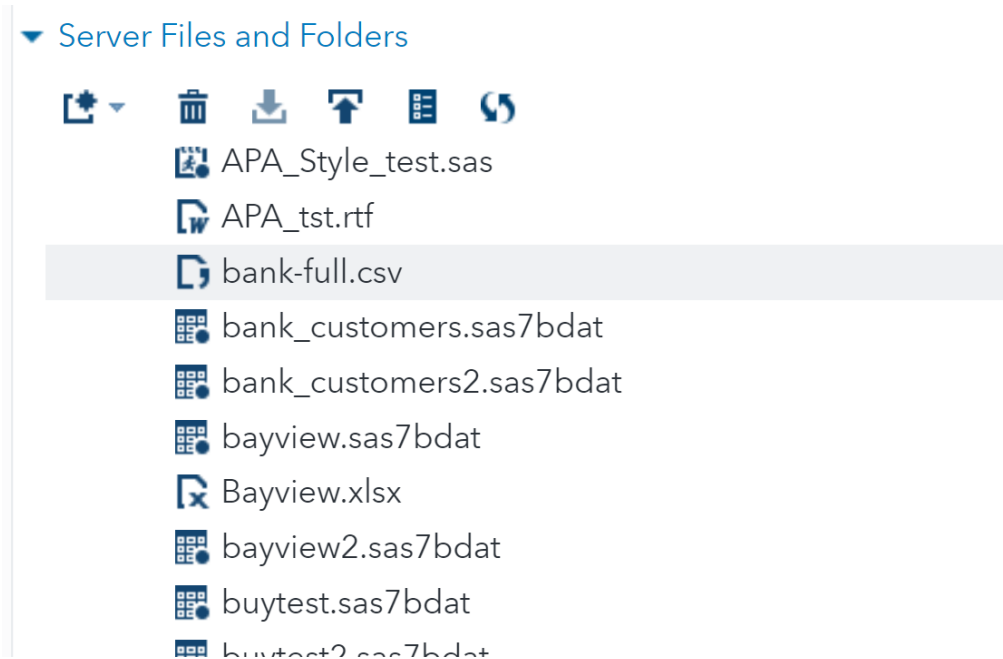


The screenshot shows a web-based 'Upload Files' dialog box. At the top left is the title 'Upload Files' in blue, and at the top right is a close button 'x'. Below the title, there is a label 'Upload files to:' followed by a text input field containing the path '/home/stefaniegreay'. Underneath the input field is a button labeled 'Choose Files'. Below this, the section 'Selected files:' is followed by a table listing one file. The table has four columns: an index '1', a file type 'CSV', a filename 'bank-full.csv', and a size '4.4 mb'. At the bottom right of the dialog are two buttons: 'Upload' (highlighted in blue) and 'Cancel'.

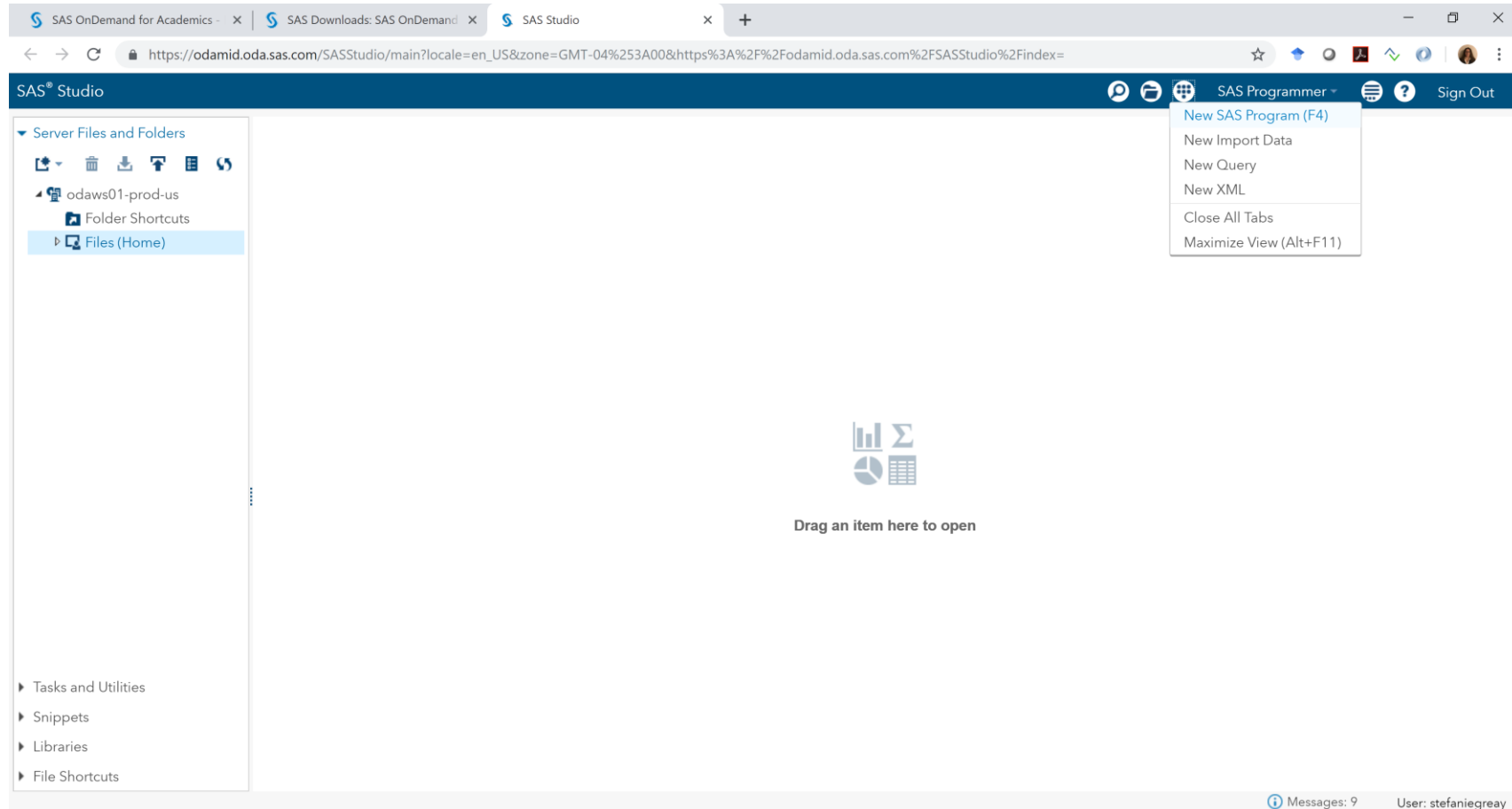
Selected files:			
1	CSV	bank-full.csv	4.4 mb



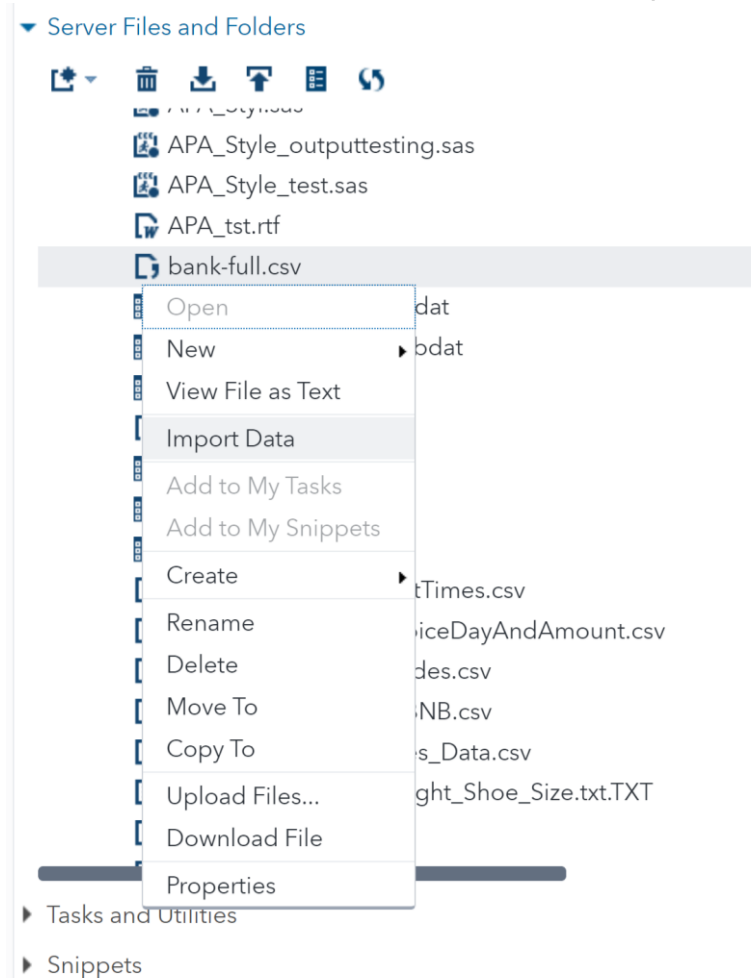
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



# 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!)

The screenshot shows the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'bank-full.csv' selected. The main window is titled '\*bank-full' and shows the 'FILE INFORMATION' tab. It displays the source file 'bank-full.csv' located at '/home/stefaniegreay'. The 'OUTPUT DATA' section shows the SAS server 'SASApp', data set name 'IMPORT', and library 'WORK'. The 'CODE' tab is active, showing the Proc Import code. The 'RESULTS' tab is also visible, showing the 'OUTPUT DATA' table 'WORK.IMPORT' with 45211 rows and 17 columns. The columns are: age, job, marital, education, default, and ba. The first five rows of data are displayed.

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

	age	job	marital	education	default	ba
1	58	management	married	tertiary	no	
2	44	technician	single	secondary	no	
3	33	entrepreneur	married	secondary	no	
4	47	blue-collar	married	unknown	no	
5	33	unknown	single	unknown	no	



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.

The screenshot shows the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'bank-full.csv' selected. The main window is titled '\*bank-full' and has tabs for 'Settings', 'Code/Results', and 'Split'. The 'Settings' tab is active, showing the 'OPTIONS' section. The 'File type' is set to 'DLM (Delimited file)'. The 'Generate SAS variable names' checkbox is checked. The 'Delimiter' is set to ';'. The 'Quote delimiter value' checkbox is checked. The 'Start reading data at row:' is set to 'Default'. Below the settings, the 'OUTPUT DATA' tab is active, showing a table named 'WORK.IMPORT'. The table has 17 columns and 45211 rows. The first five rows of data are displayed.

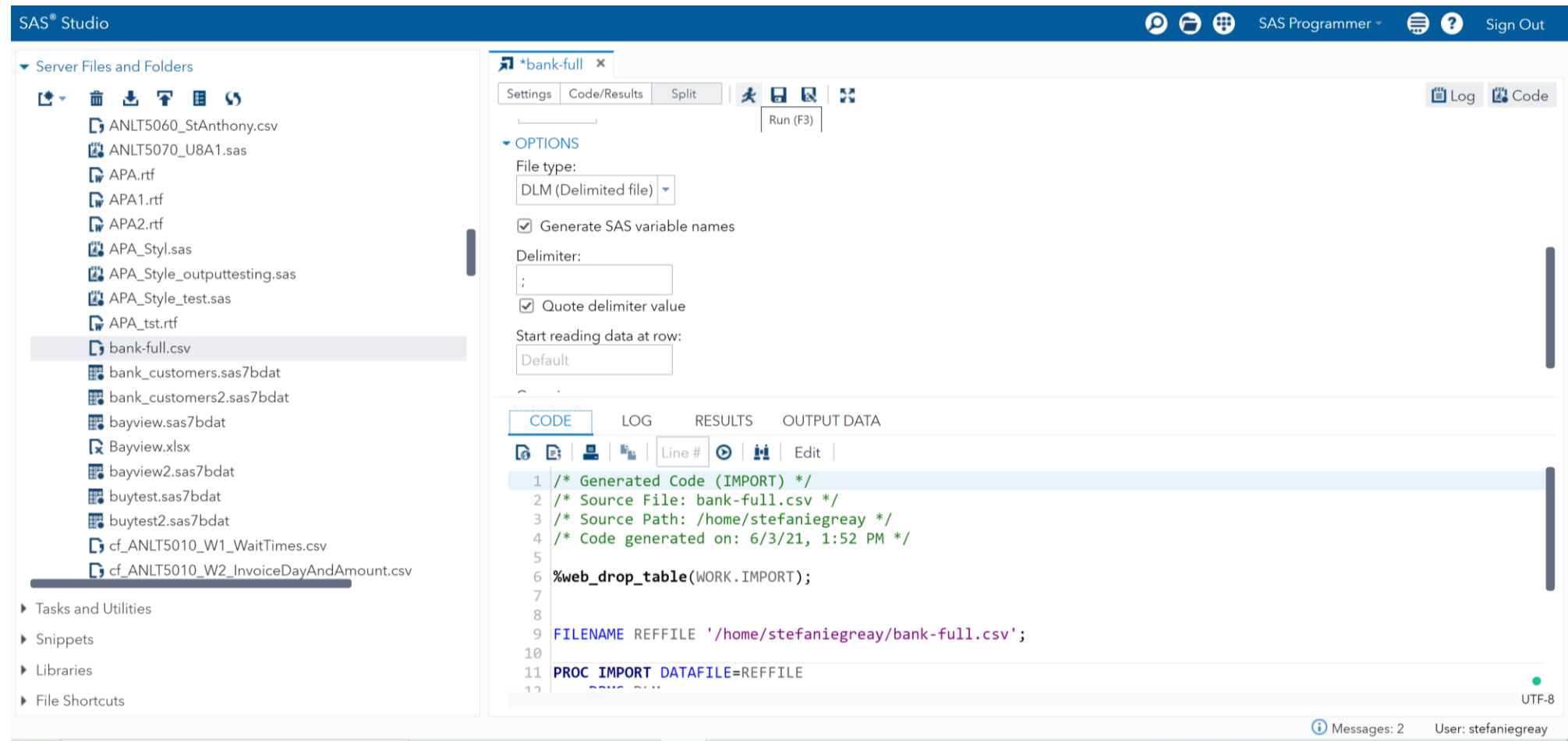
Changes:

- 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.

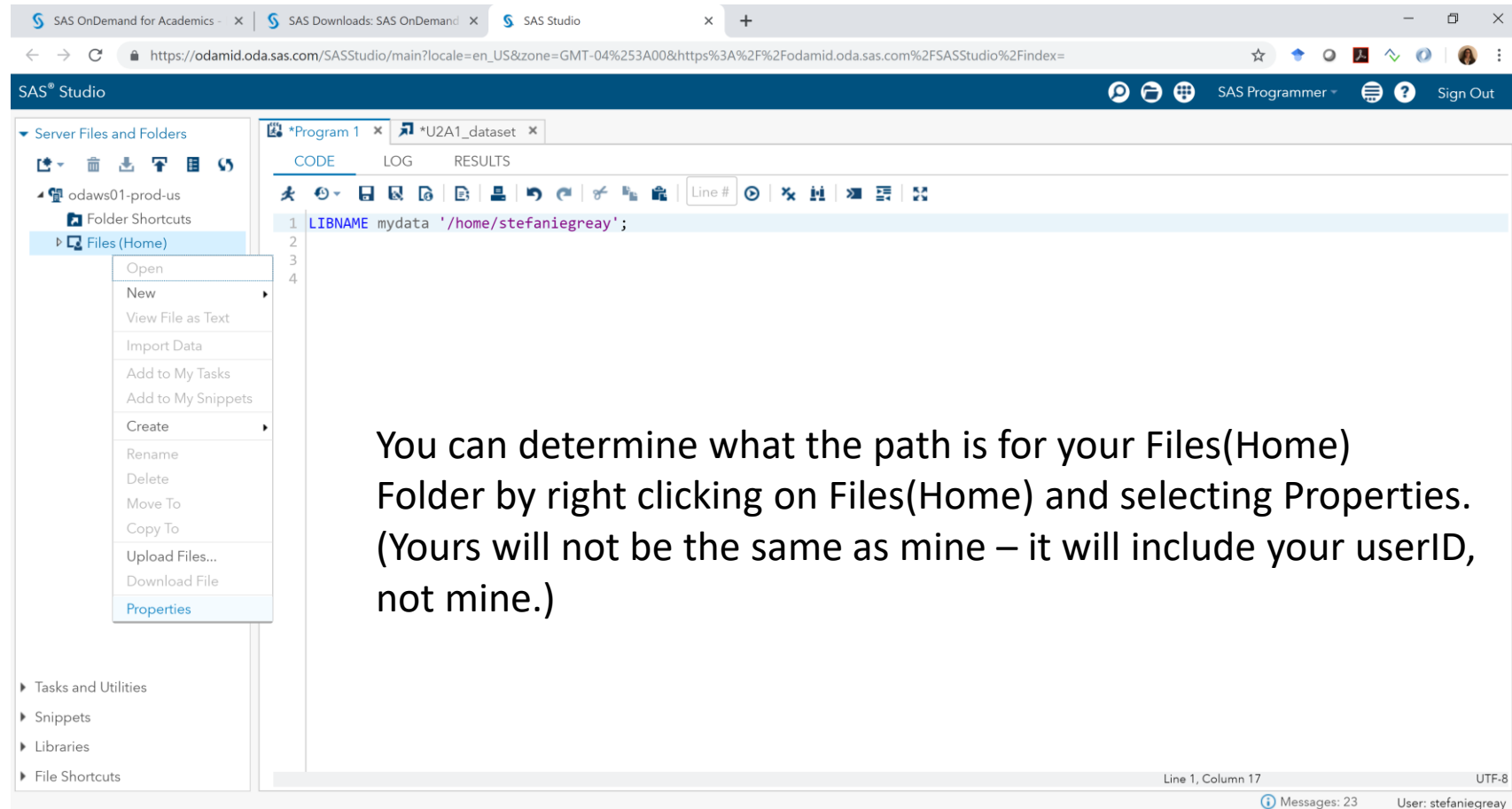
	age	job	marital	education	default	ba
1	58	management	married	tertiary	no	
2	44	technician	single	secondary	no	
3	33	entrepreneur	married	secondary	no	
4	47	blue-collar	married	unknown	no	
5	33	unknown	single	unknown	no	



To run the code, click the icon that looks like a guy running.

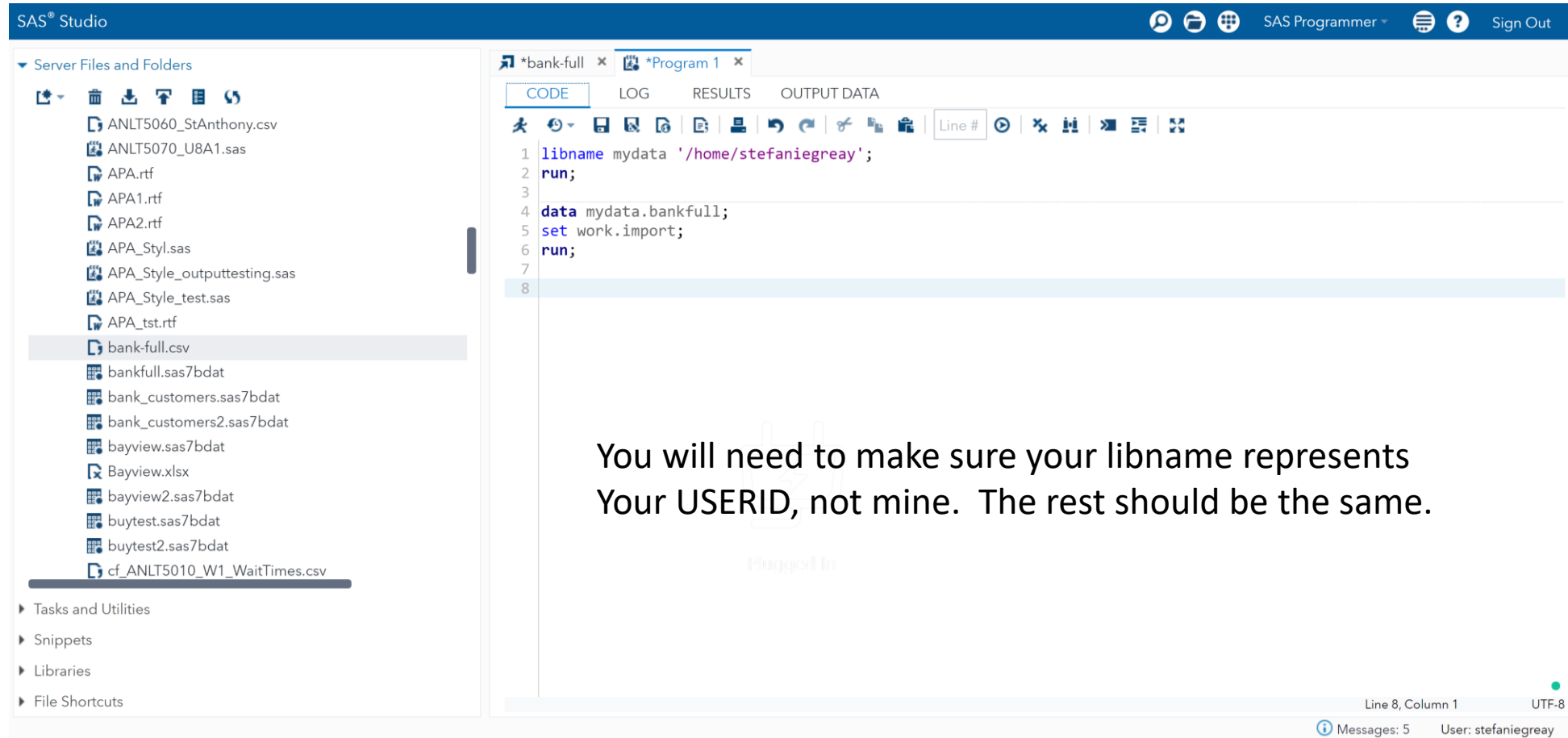


# To create a SAS Library for your Files(Home) folder, you need to use a libname statement





# Save the temporary SAS datasets created by the import to your library using the following sample code.



When you run the code, you will see the dataset in the output data window and can verify its success.

SAS® Studio

Server Files and Folders

- ANLT5060\_StAnthony.csv
- ANLT5070\_U8A1.sas
- APA.rtf
- APA1.rtf
- APA2.rtf
- APA\_Styl.sas
- APA\_Style\_outputtesting.sas
- APA\_Style\_test.sas
- APA\_tst.rtf
- bank-full.csv
- bankfull.sas7bdat
- bank\_customers.sas7bdat
- bank\_customers2.sas7bdat
- bayview.sas7bdat
- Bayview.xlsx
- bayview2.sas7bdat
- buytest.sas7bdat
- buytest2.sas7bdat
- cf\_ANLT5010\_W1\_WaitTimes.csv

Tasks and Utilities

Snippets

Libraries

File Shortcuts

\*bank-full \*Program 1

CODE LOG RESULTS OUTPUT DATA

Table: MYDATA.BANKFULL View: Column names Filter: (none)

Columns

- Select all
- age
- job
- marital
- education
- default
- balance
- housing
- loan
- contact

Property Value

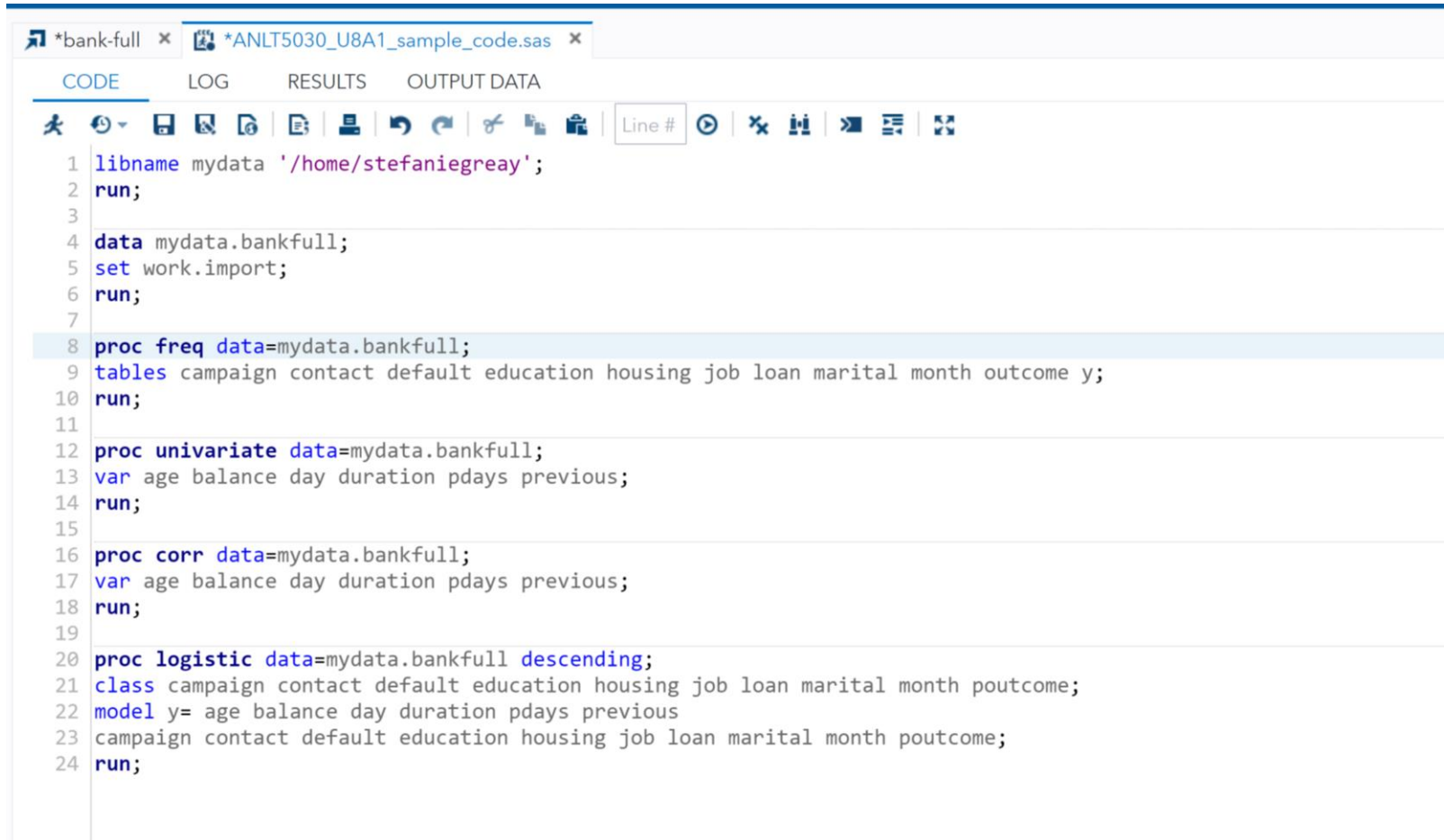
	age	job	marital	education	default	balance
1	58	management	married	tertiary	no	
2	44	technician	single	secondary	no	
3	33	entrepreneur	married	secondary	no	
4	47	blue-collar	married	unknown	no	
5	33	unknown	single	unknown	no	
6	35	management	married	tertiary	no	
7	28	management	single	tertiary	no	
8	42	entrepreneur	divorced	tertiary	yes	
9	58	retired	married	primary	no	
10	43	technician	single	secondary	no	
11	41	admin.	divorced	secondary	no	
12	29	admin.	single	secondary	no	
13	53	technician	married	secondary	no	
14	58	technician	married	unknown	no	
15	57	services	married	secondary	no	
16	51	retired	married	primary	no	
17	45	admin.	single	unknown	no	

Total rows: 45211 Total columns: 17 Rows 1-100

Messages: 5 User: stefaniegreay



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

A screenshot of the SAS Studio Code Window. The window has a title bar with two tabs: '\*bank-full' and '\*ANLT5030\_U8A1\_sample\_code.sas'. Below the title bar are four tabs: 'CODE', 'LOG', 'RESULTS', and 'OUTPUT DATA'. The 'CODE' tab is active. Below the tabs is a toolbar with various icons for file operations, editing, and execution. The main area of the window contains SAS code. Line 8, 'proc freq data=mydata.bankfull;', is highlighted in blue. The code includes: 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;

# 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.

