ANLT5020 – Unit 3 Assignment 1 Tutorial

SAS Studio

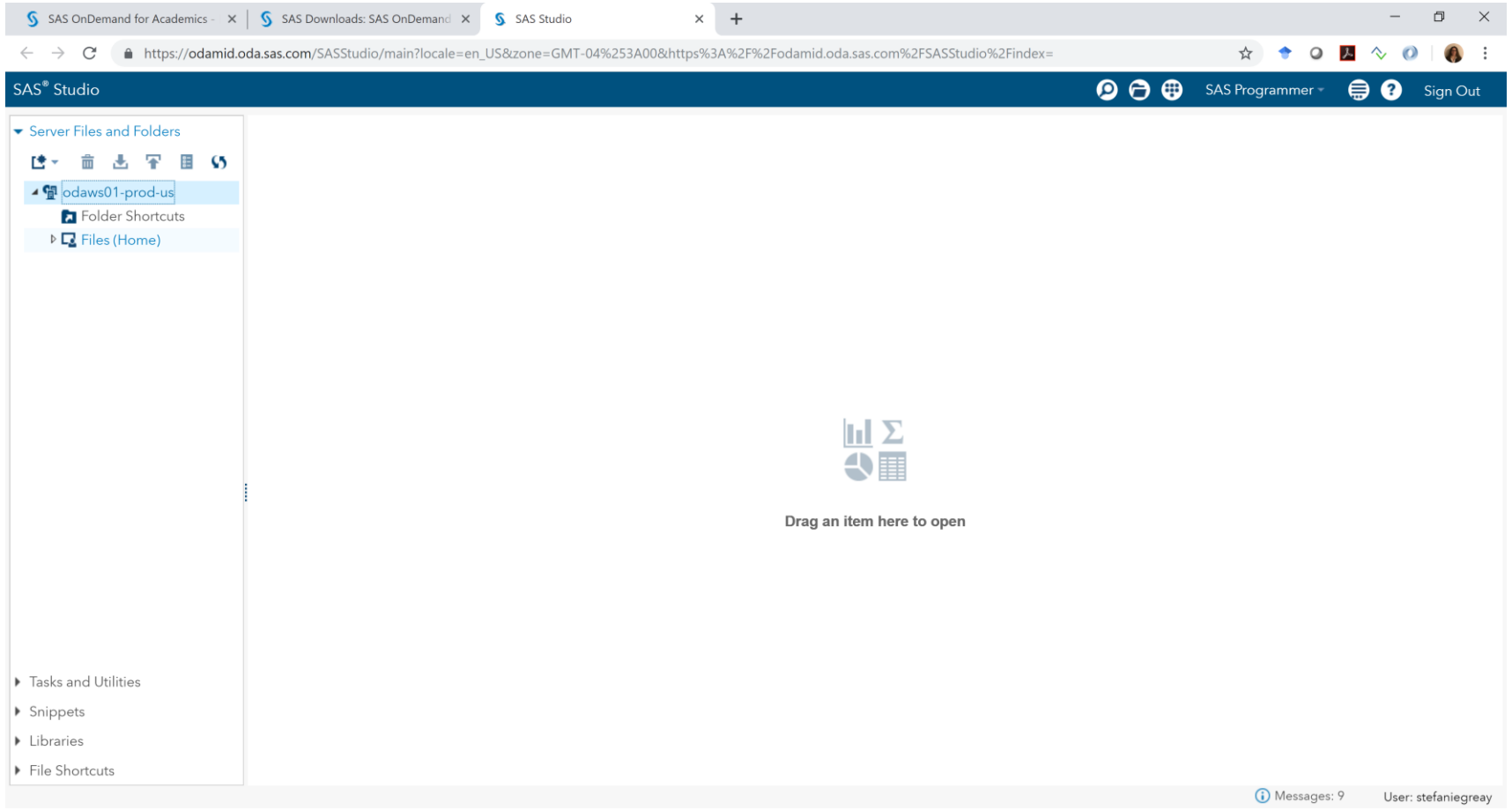
Instructions

- Create a new SAS dataset using data from Teachers.xlsx.
- Complete the following using this dataset:
- For the character variables FirstName, LastName, Gender, and Subject, run an SAS procedure that will detect the invalid values within these variables. List the invalid values for each variable, if any.
- For the variable Salary, run an SAS procedure to find any salaries that seem to be unusual compared to the rest of the dataset. List the name of the employee with an invalid Salary value.
- For the variable LastName, note that there are extra blanks in some of the last names. Use a DATA step with a SET statement and a COMPRESS command to create a new dataset without these blanks. (Hint: You will need to create a new variable name for the compressed LastName.)
- Write another DATA step to fix other errors in the new dataset you created. (Hint: Make use of the SET statement.) There are three errors in the new dataset:
- Some teacher names are spelled incorrectly.
- All Genders should be capital letters.
- The actual salary of Employee_ID=E030 is $73,000.
- Use a PROC PRINT statement to print out the new dataset with the errors correcte.

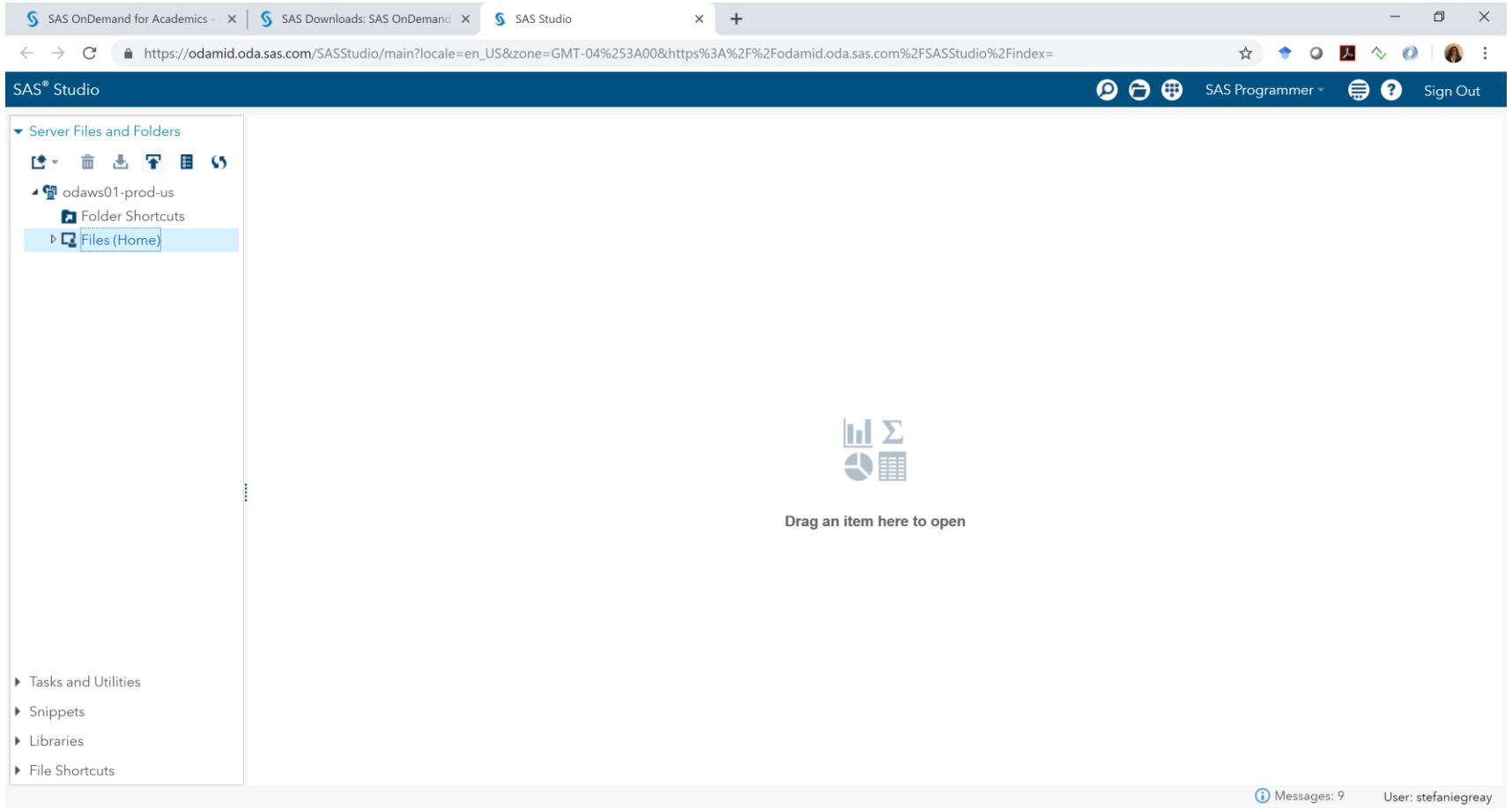
Dataset

- Download the Teachers.xlsx file from the course datasets zip file or from the Unit 3 Welcome announcement in the course announcements.

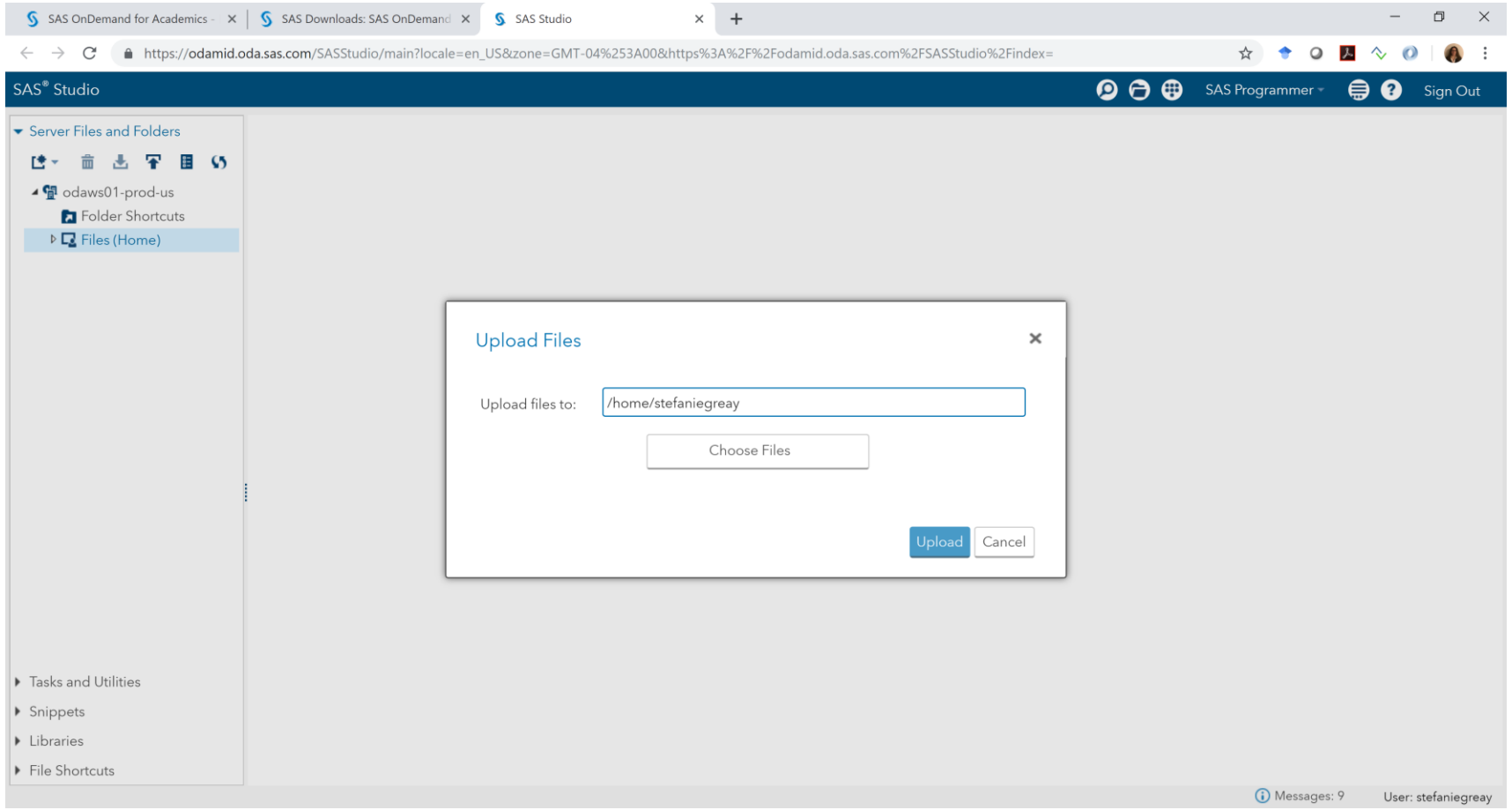
Click on Files(Home)



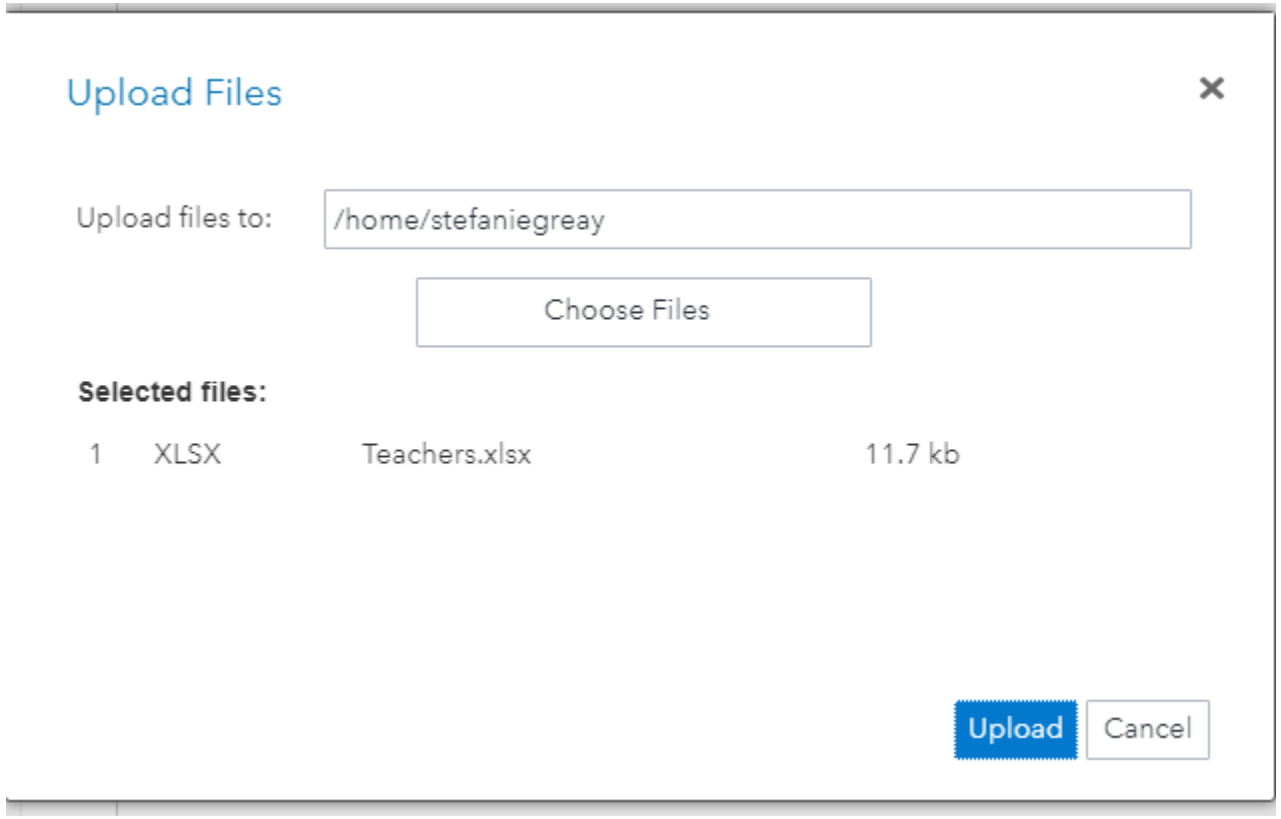
The Upload button will display in dark blue



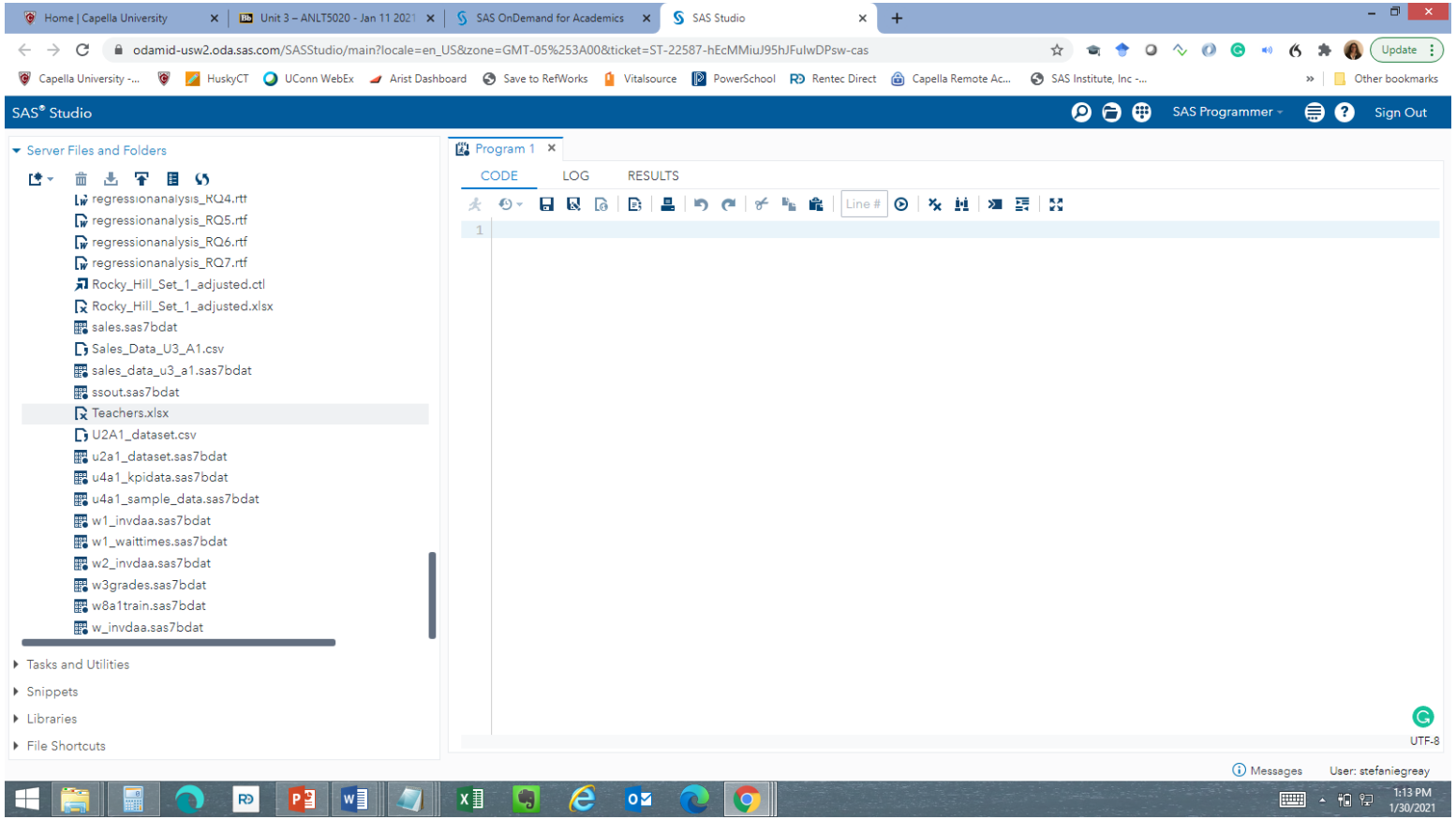
You can create a folder at this point, if you wish, or simply upload to your home directory.



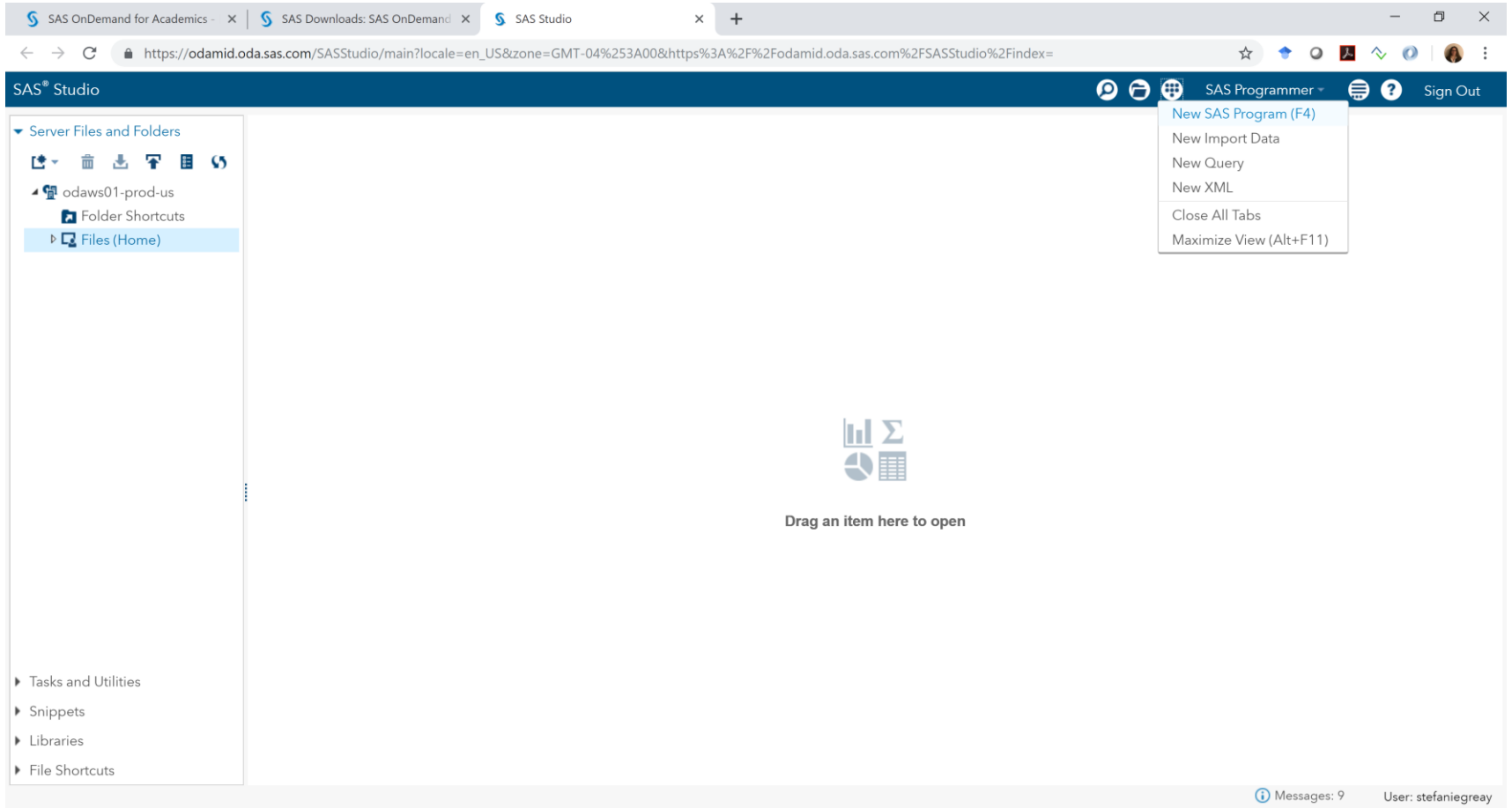
Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”



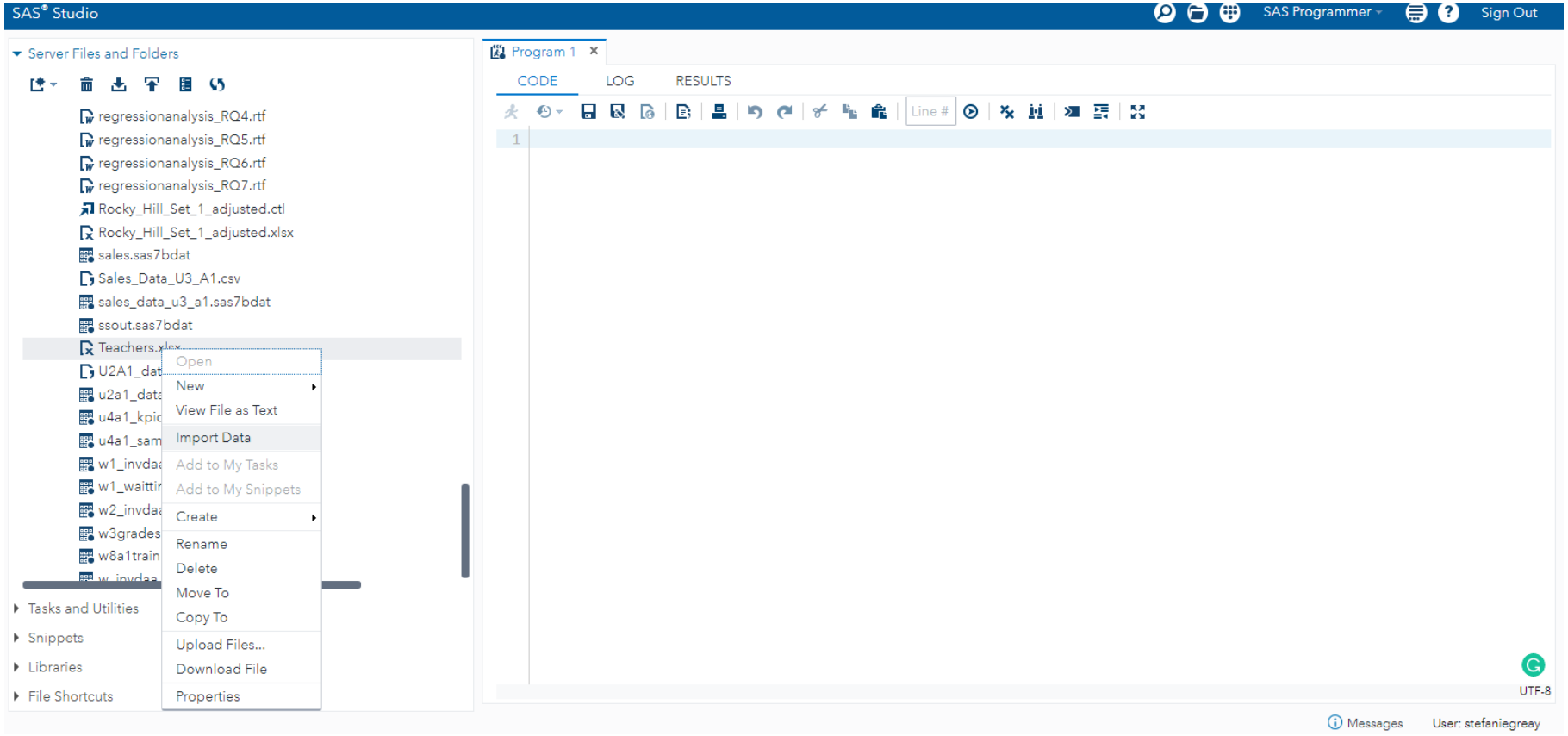
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



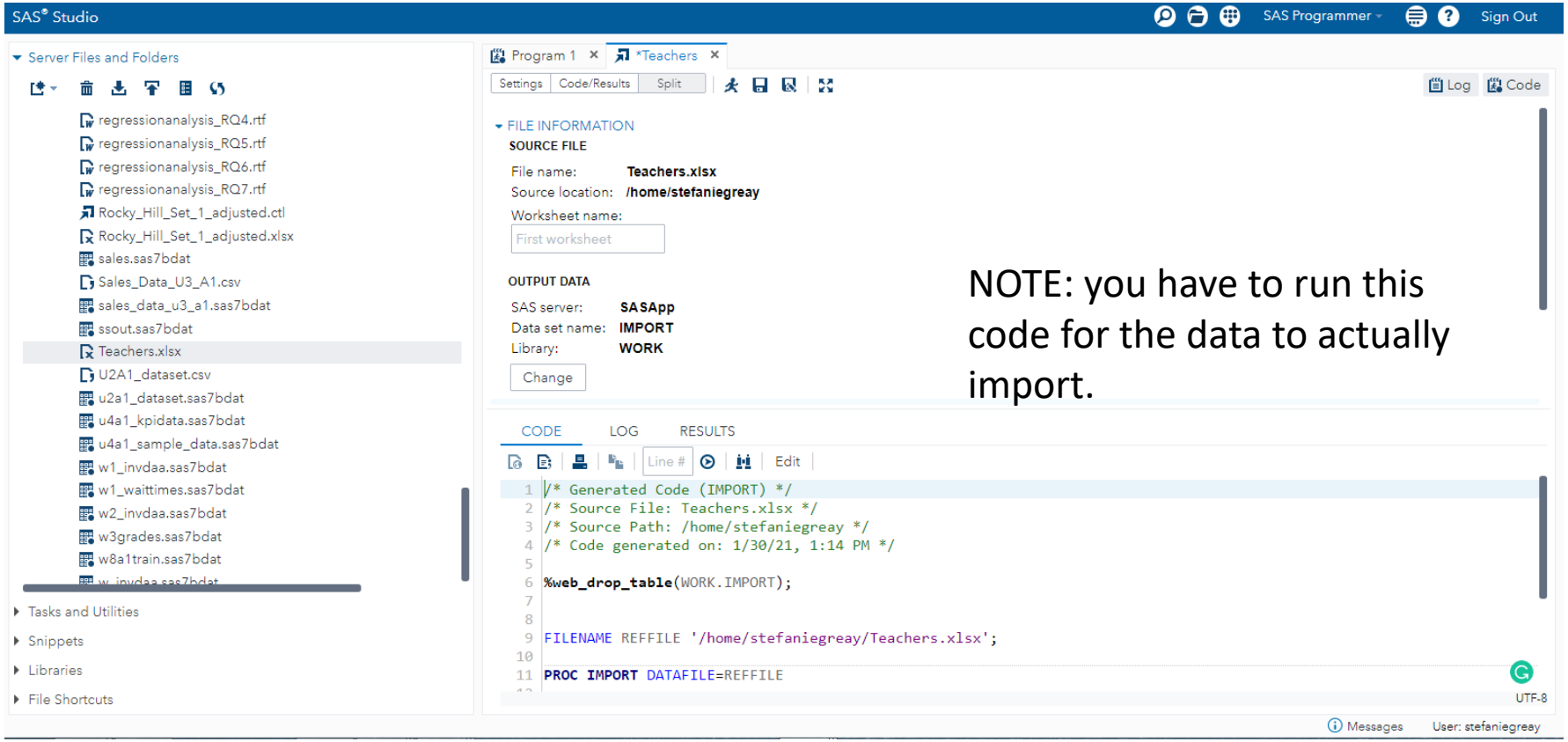
To get started with the SAS portion of the Unit 3 Assignment 1 assignment, start a new SAS program.



Import the dataset into a SAS dataset format (from the current xlsx format)

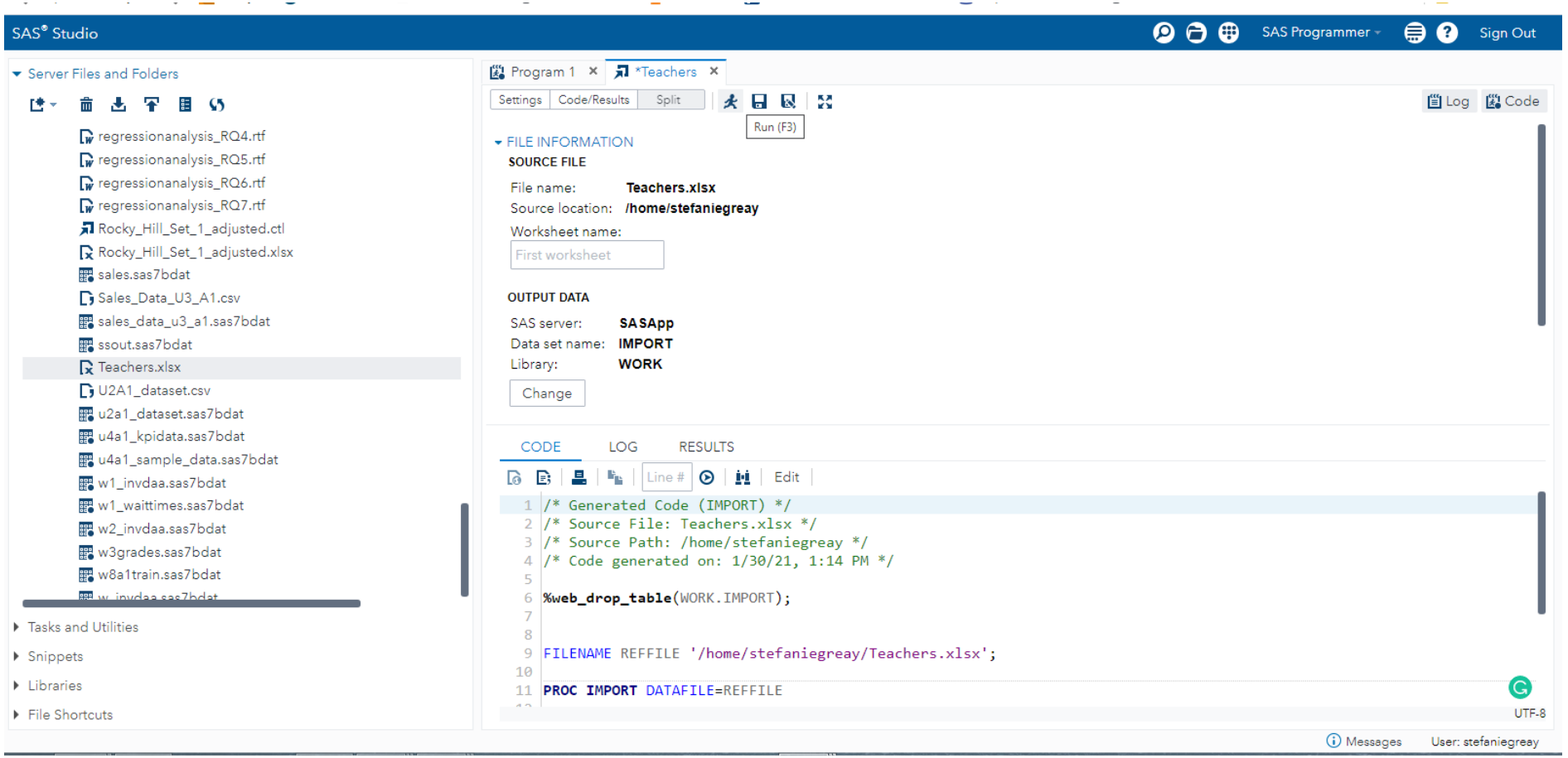


The Proc Import code will be written for you (save this as a template to use for future imports!)

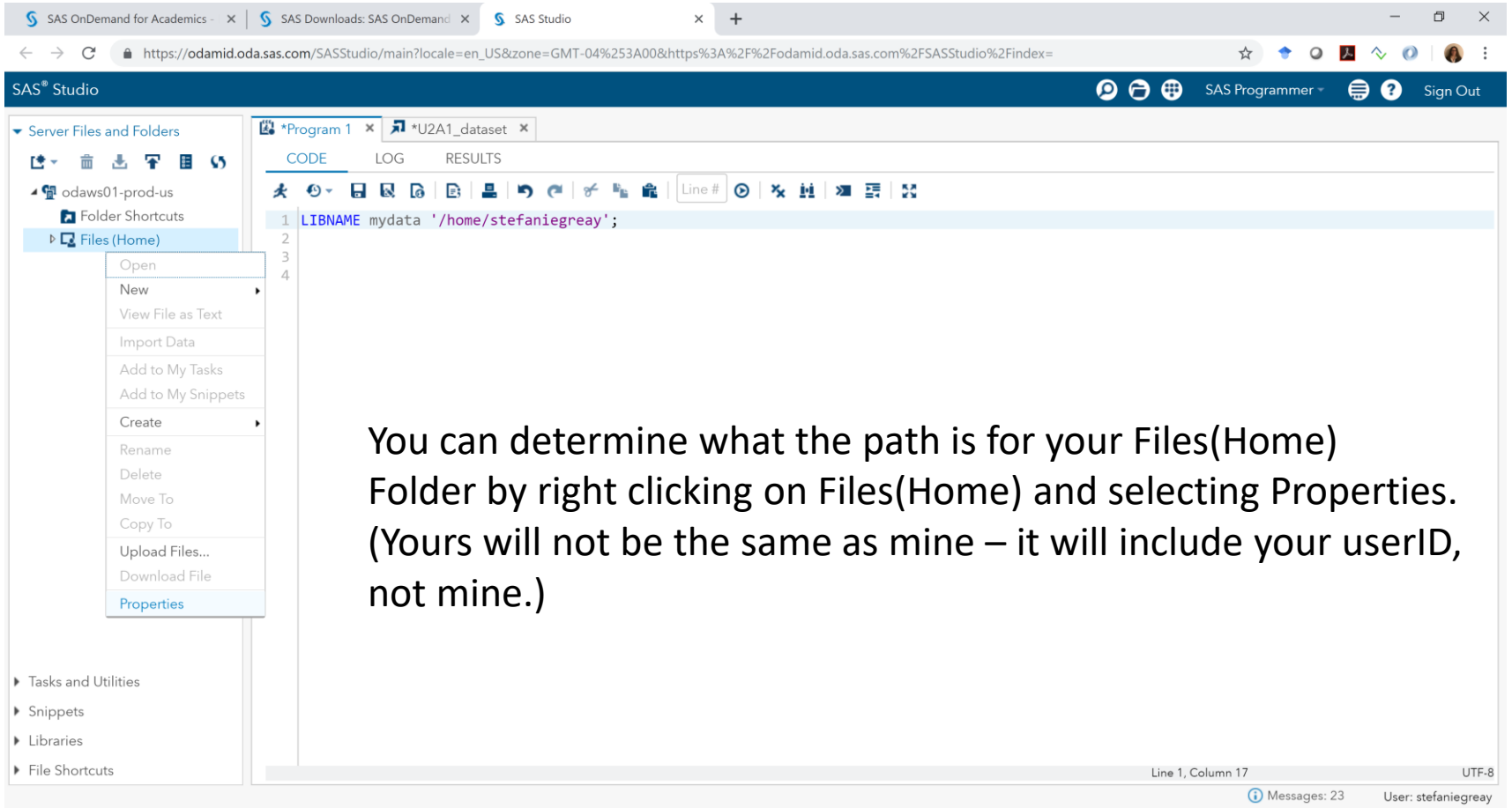


NOTE: you have to run this code for the data to actually import.

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

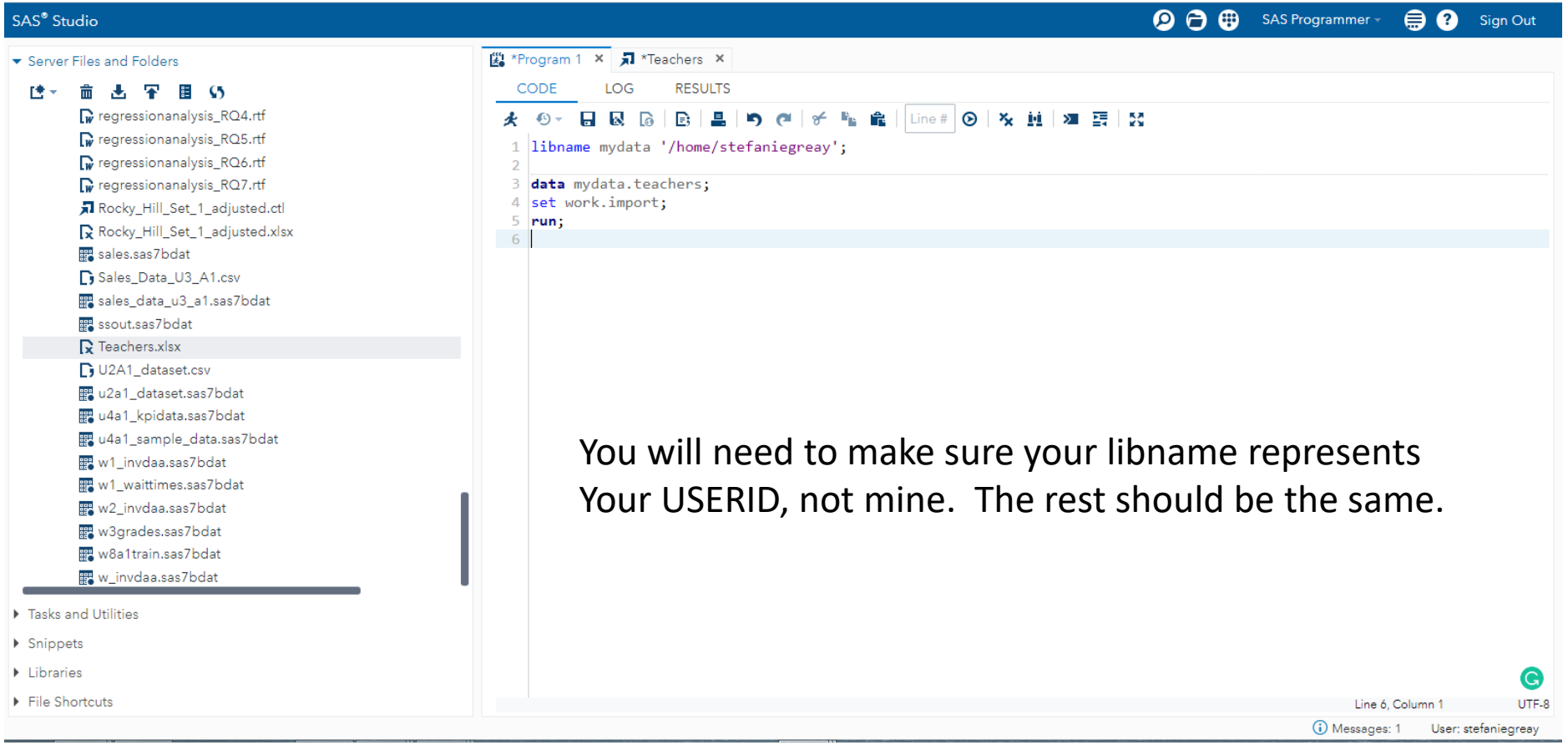


To create a SAS Library for your Files(Home) folder, you need to use a libname statement

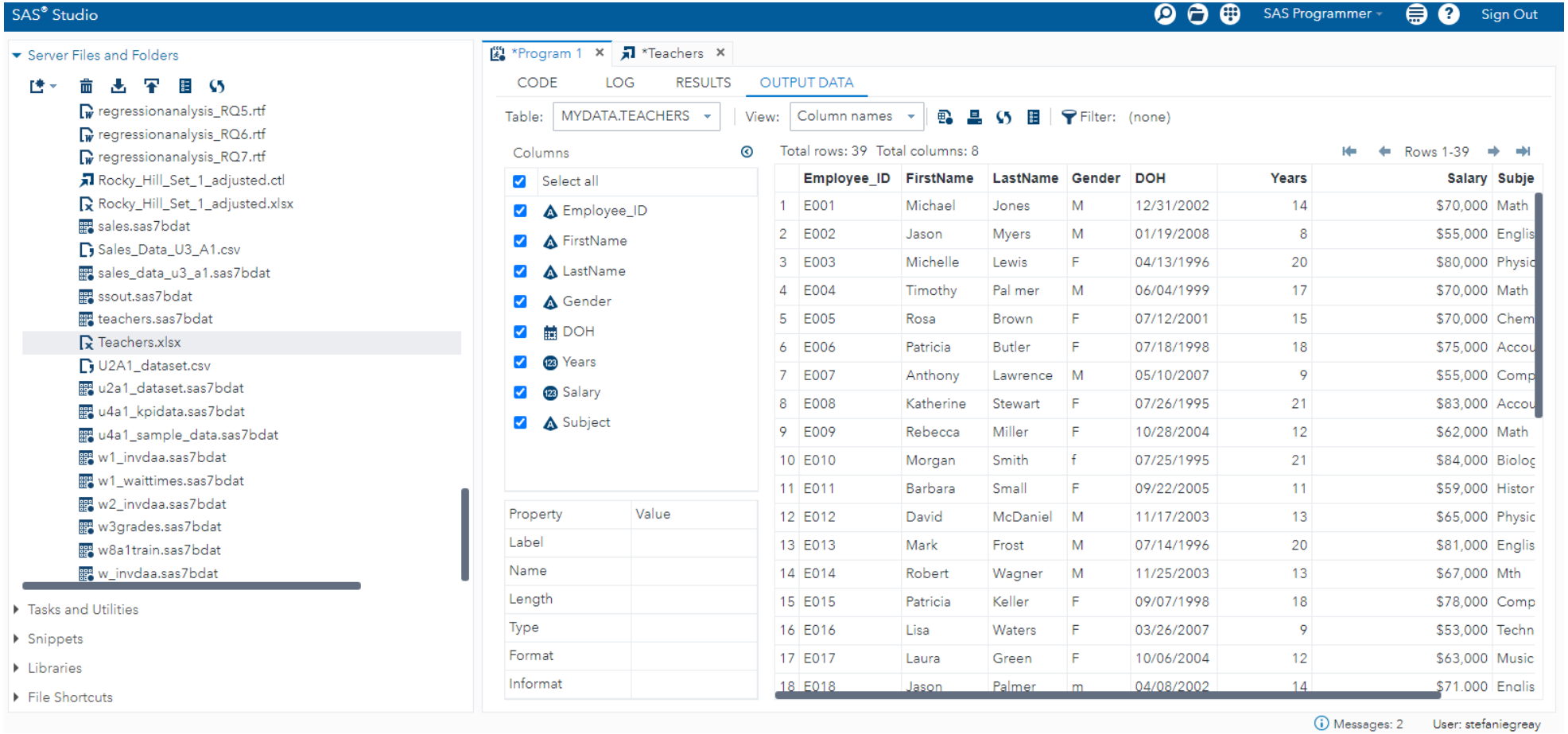


You can determine what the path is for your Files(Home) Folder by right clicking on Files(Home) and selecting Properties. (Yours will not be the same as mine – it will include your userID, not mine.)

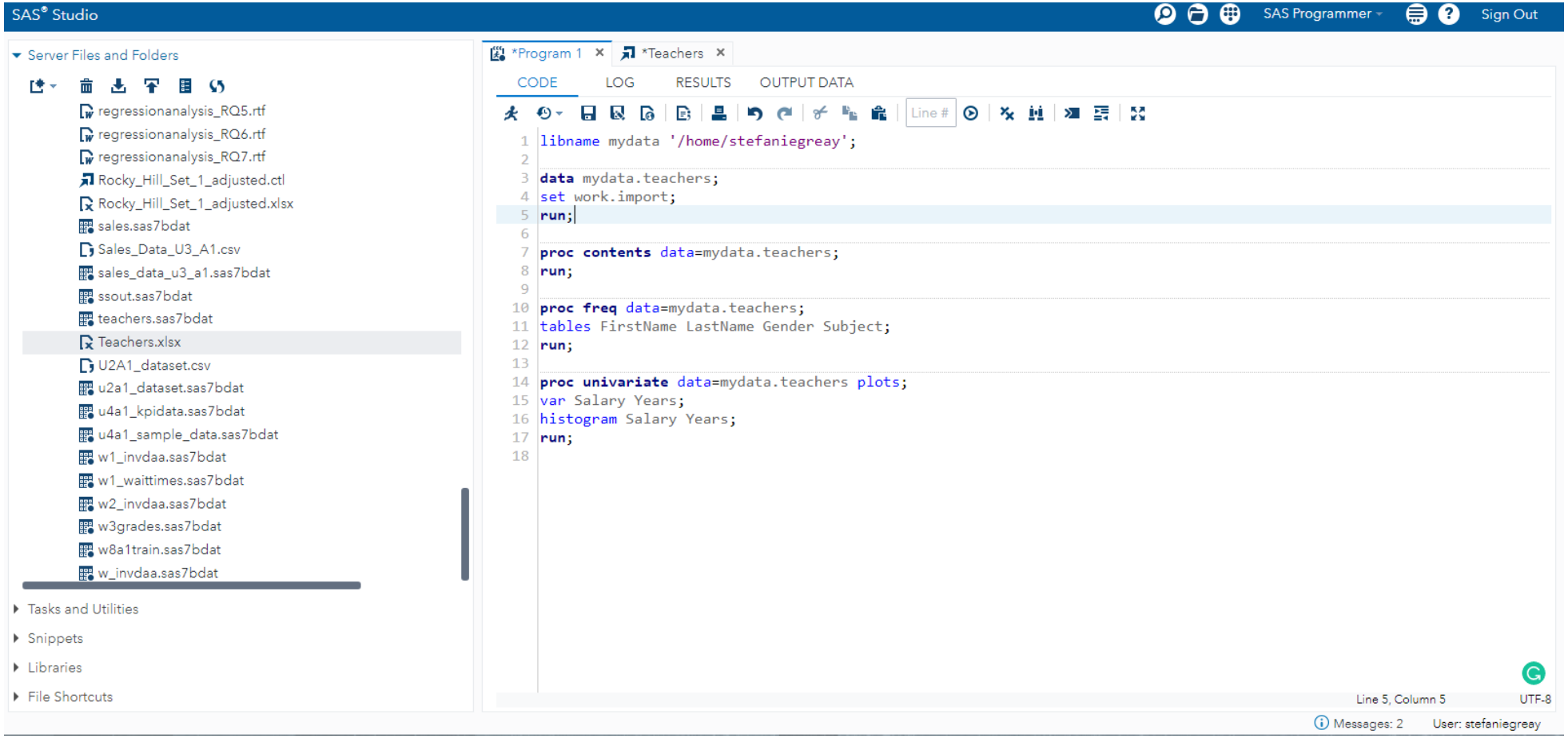
Save the temporary SAS dataset created by the import to your library using the following sample code.



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



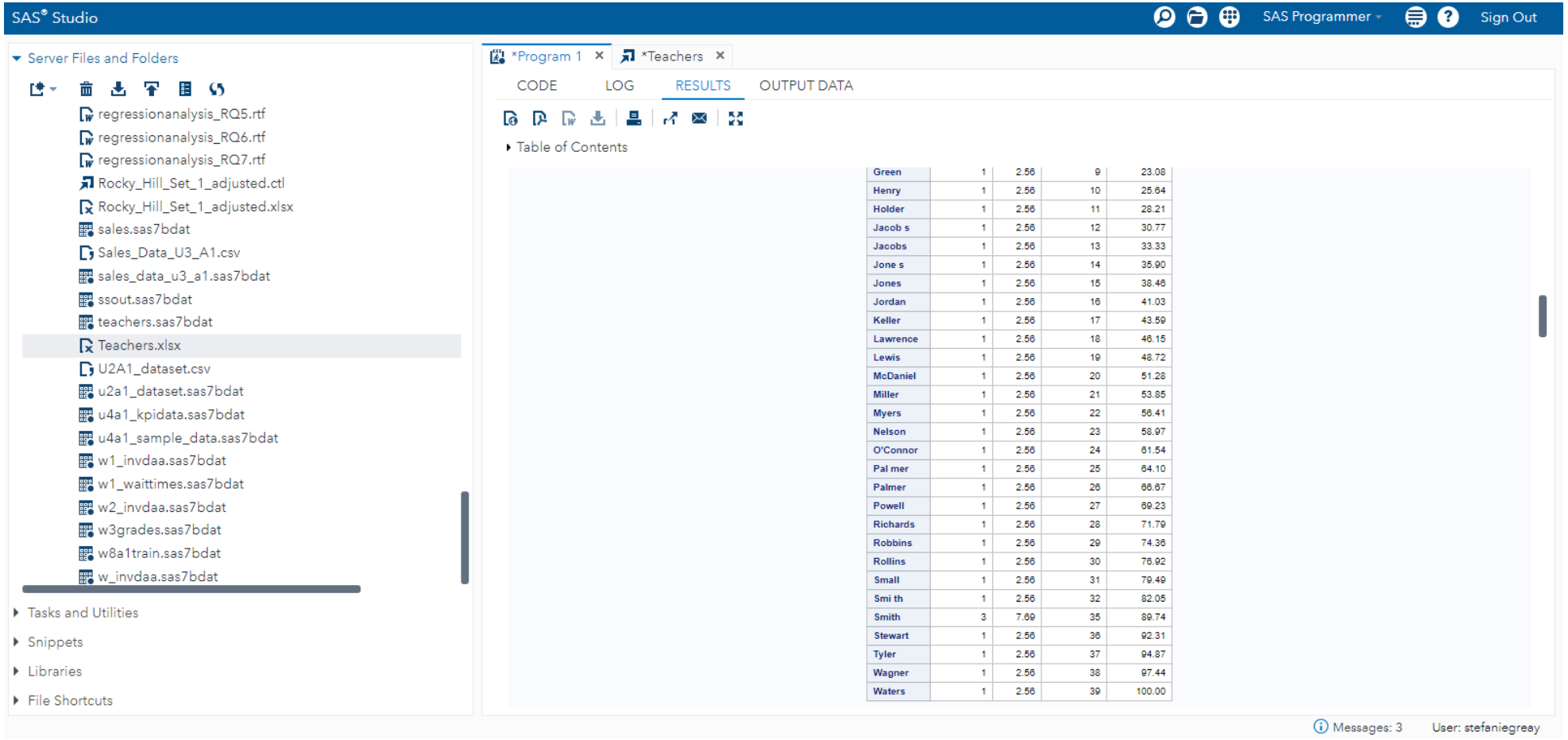
You can now run any procedures against that dataset via the code window.



Sample Code for identifying issues

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

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

proc contents data=mydata.teachers;
run;

proc freq data=mydata.teachers;
tables FirstName LastName Gender Subject;
run;

proc univariate data=mydata.teachers plots;
var Salary Years;
histogram Salary Years;
run;
```

Once you run the code, you can review the output to identify data issues, like those in the LastName variable.



Sample Code

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

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

proc contents data=mydata.teachers;
run;

proc freq data=mydata.teachers;
tables FirstName LastName Gender Subject;
run;

proc univariate data=mydata.teachers plots;
var Salary Years;
histogram Salary Years;
run;
```

```
data mydata.teachers2;
set mydata.teachers;
LastName_c = Compress(LastName);
run;

data mydata.teachersc;
set mydata.teachers2;
Gender_c = UPCASE(Gender);
/* The teachers' names errors resolve by the
compression in the last name (i.e. removing
spaces)*/
/* There are issues in the spelling of the subjects,
though. */
if Subject='Acconting' then Subject='Accounting';
if Subject='Eglish' then Subject='English';
if Subject='Mth' then Subject='Math';
if Employee_ID='E030' then Salary = 73000;
run;
```