

ANLT5020 – Unit 6

Assignment 1 Tutorial

SAS Studio

Instructions

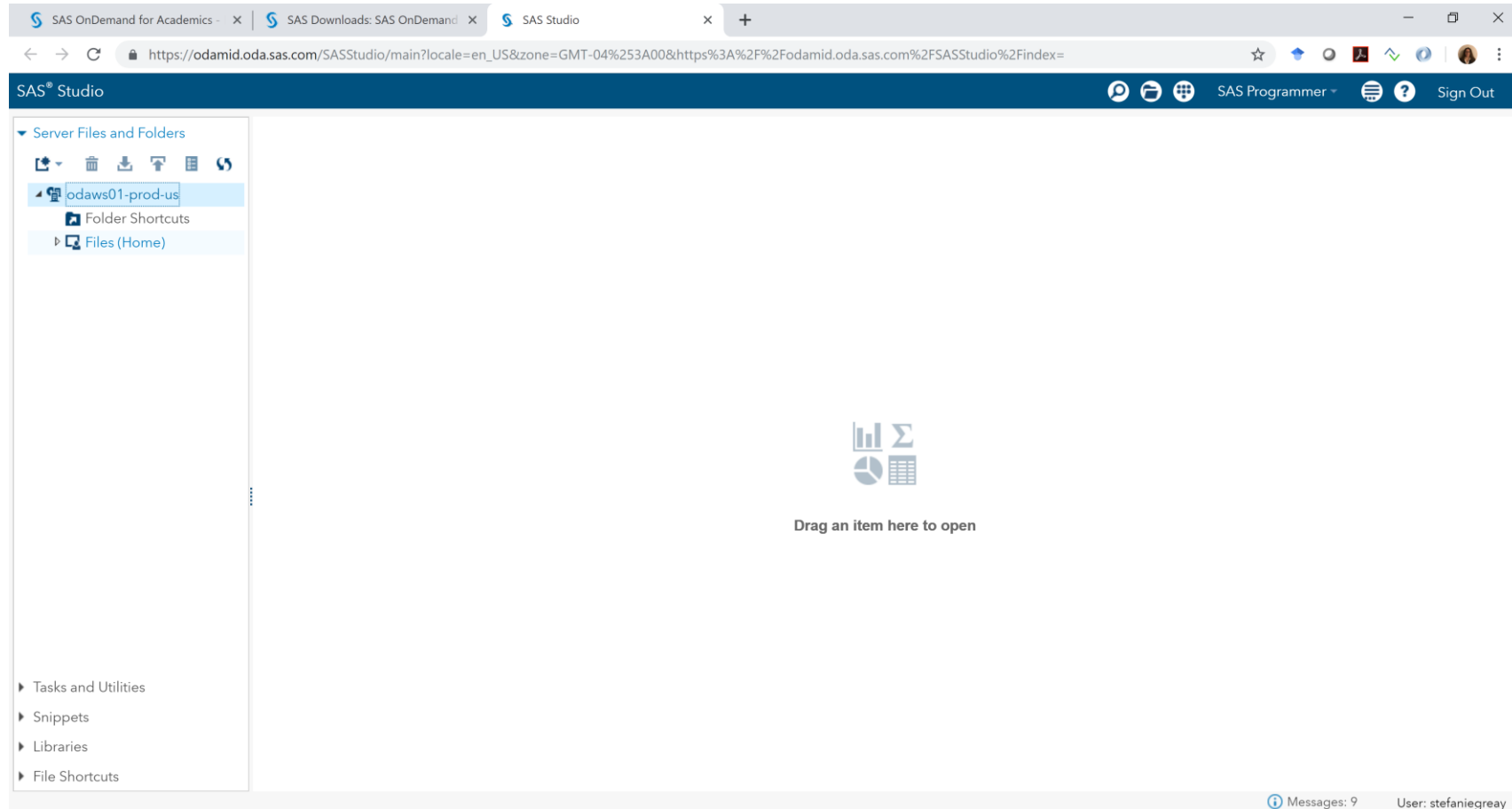
Complete the following tasks:

- Create a SAS dataset containing this data using a SET statement.
- Clarify the labels on the MOS, DOS, YOS, and DRG by using a LABEL statement to give the columns more meaningful names than the three-letter variable names.
- Use a SAS procedure to reformat the Charges variable to show dollars and cents, including a dollar sign.
- Change the length of the LastName variable to 36 characters using a LENGTH statement to allow for longer last names that may flow through this field/variable in future data.
- Combine the MOS, DOS, and YOS fields into a single date field, manipulate that field so that SAS understands it as a date, and use a FORMAT statement to format the new combined date field like this: 02/14/2016.
- Create a list of at least five best practices for managing big data based on your evaluation of the assigned readings and your work in SAS. Provide a description for each best practice. Be sure to address privacy, security, and ethical concerns in your recommendations.

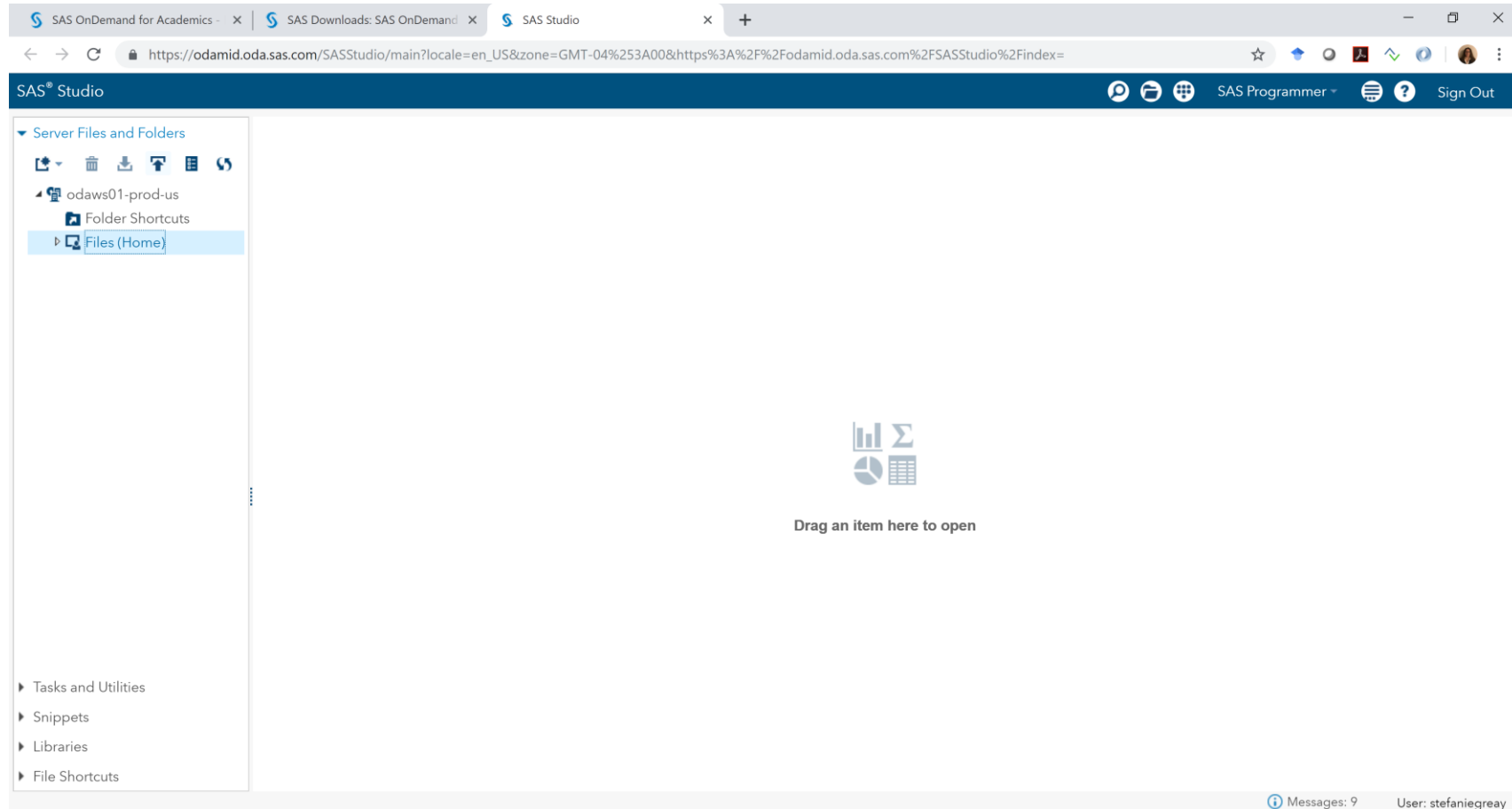
Dataset

- Download the Claims.txt file from the course datasets zip file or from the Unit 6 Welcome announcement in the course announcements.

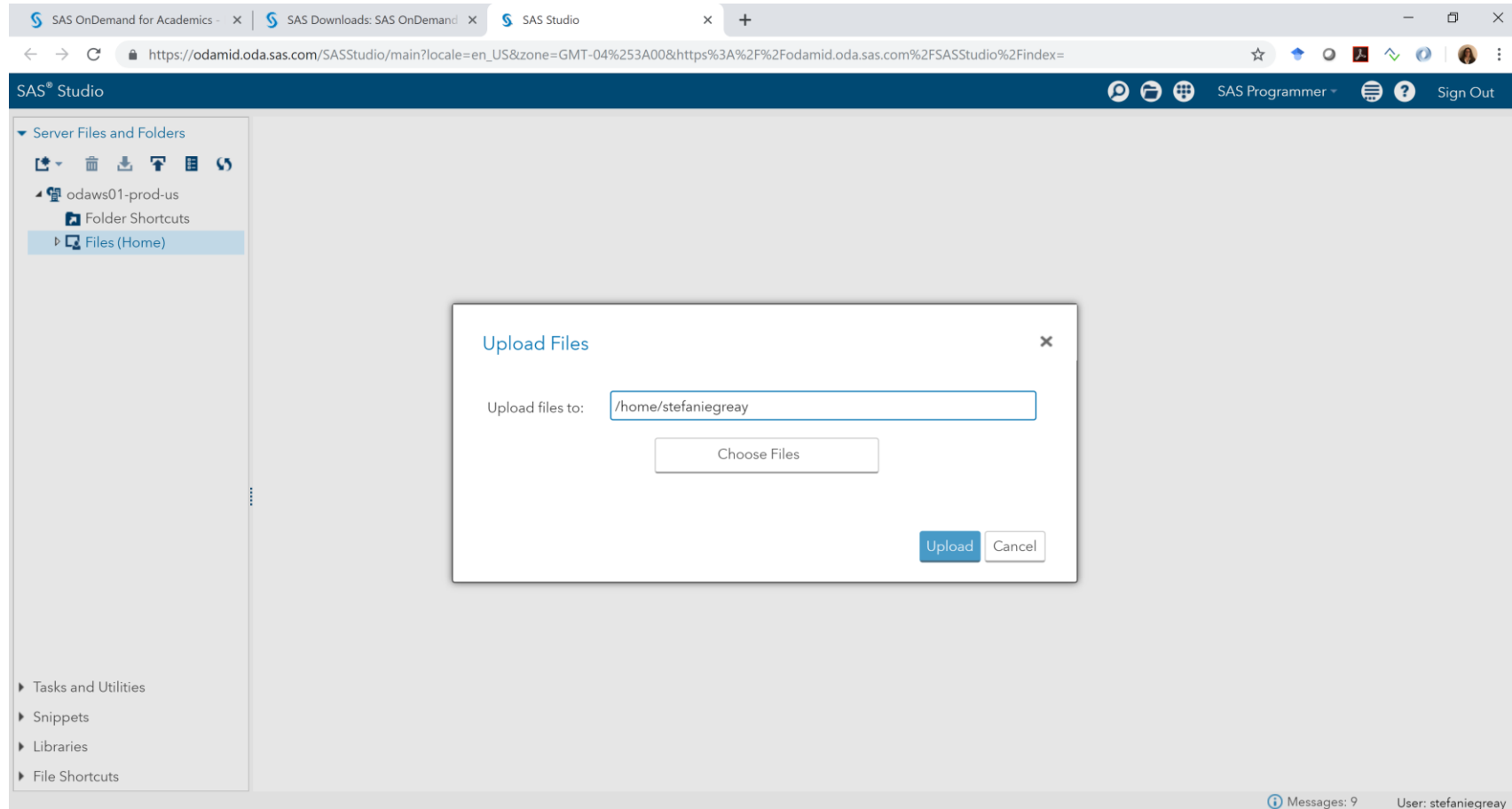
Click on Files(Home)



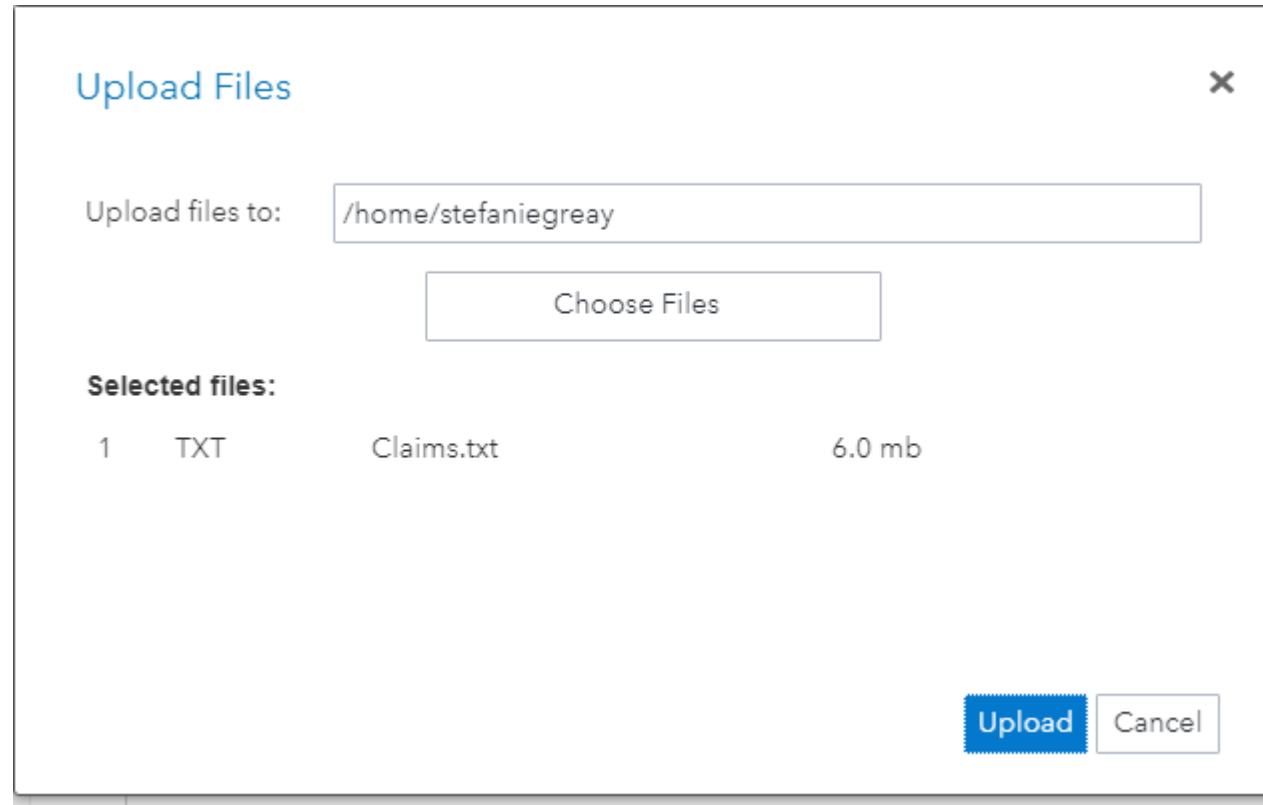
The Upload button will display in dark blue



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



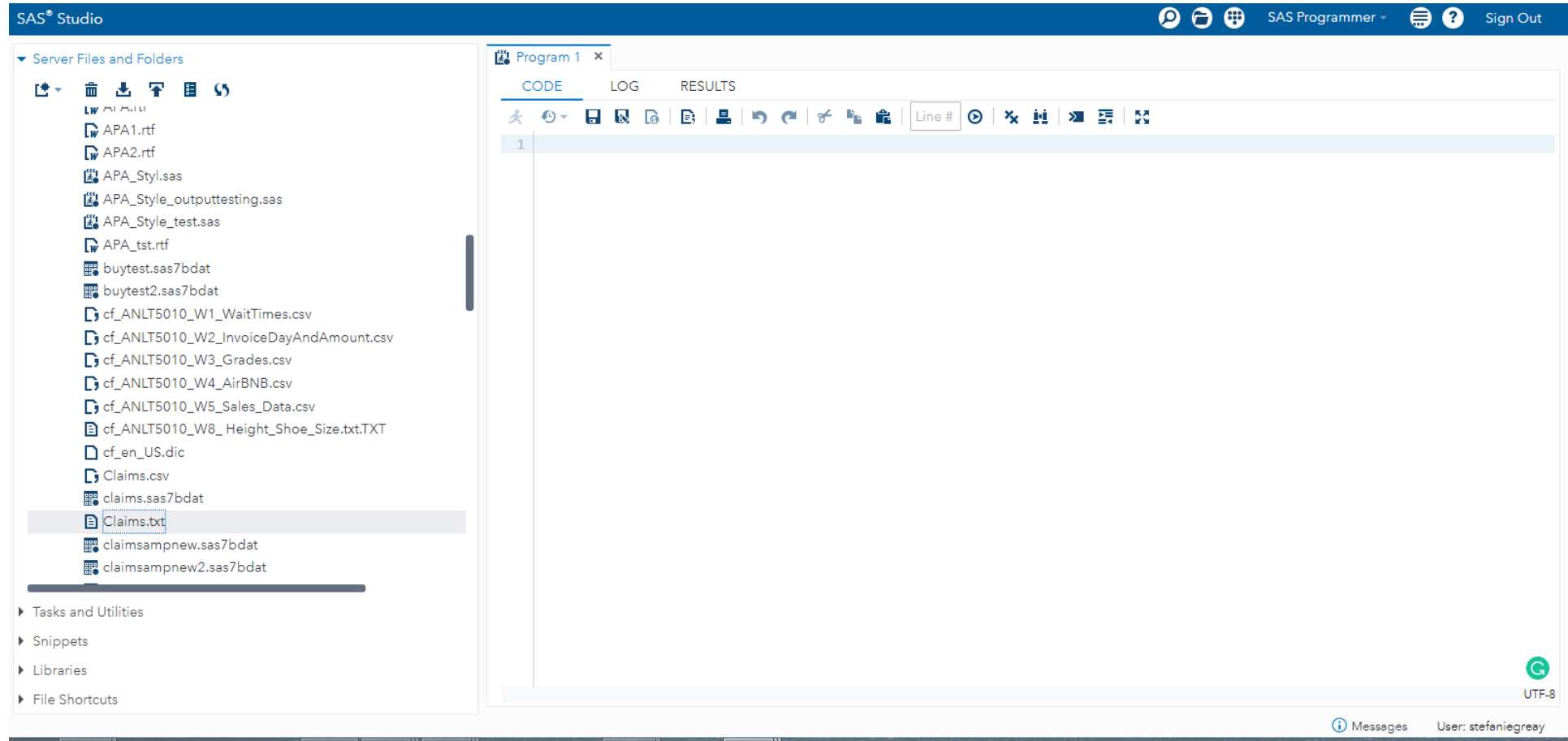
Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”



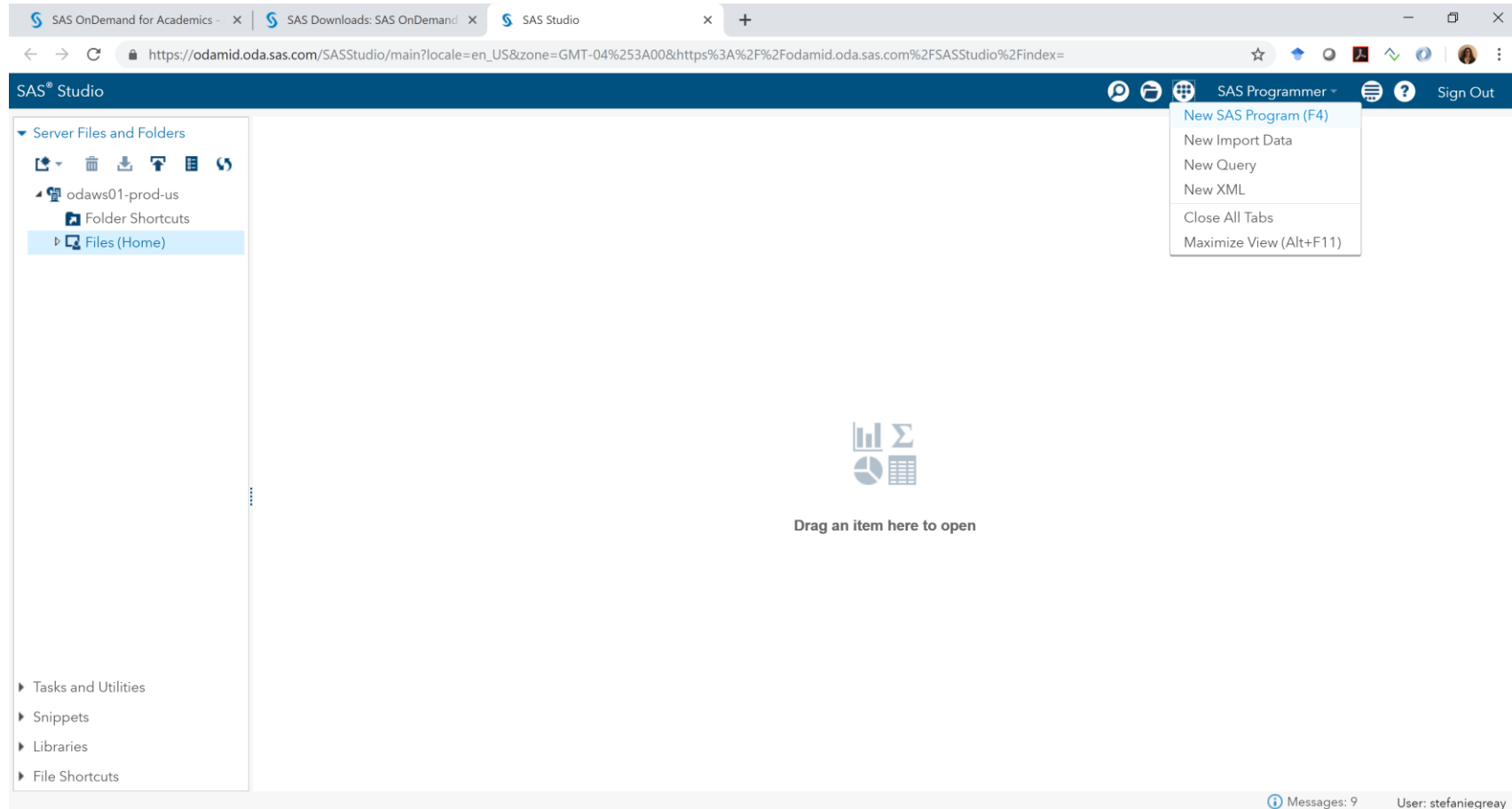
The screenshot shows a web-based 'Upload Files' dialog box. At the top left is the title 'Upload Files' in blue, and at the top right is a close button 'x'. Below the title, there is a label 'Upload files to:' followed by a text input field containing the path '/home/stefaniegreay'. Underneath the input field is a button labeled 'Choose Files'. Below this, there is a section titled 'Selected files:'. Under this title, there is a table with one row of data: a number '1', the file type 'TXT', the filename 'Claims.txt', and the file size '6.0 mb'. At the bottom right of the dialog box, there are two buttons: a blue 'Upload' button and a white 'Cancel' button with a grey border.

Selected files:			
1	TXT	Claims.txt	6.0 mb

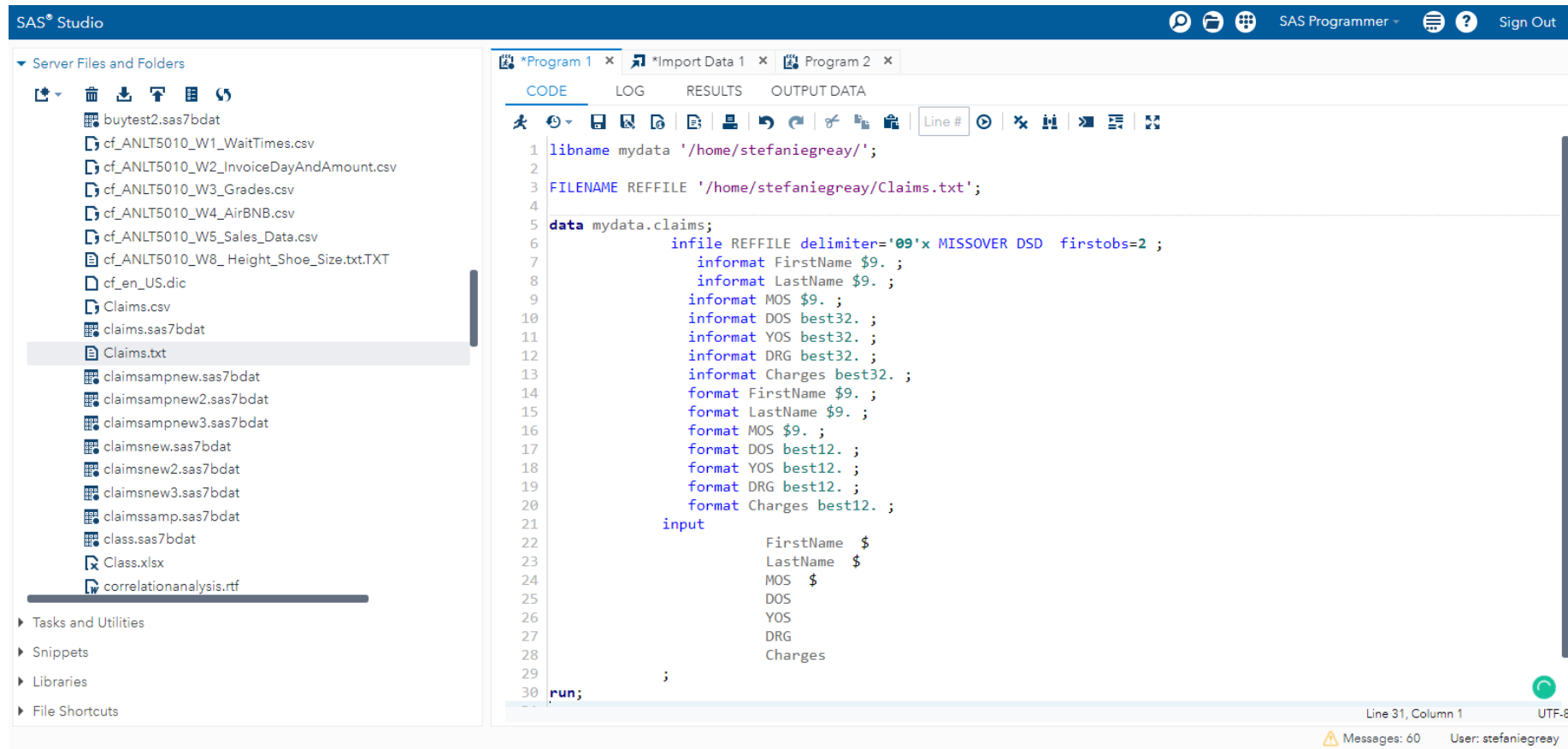
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



To get started with the SAS portion of the assignment, start a new SAS program.



Import the dataset into a SAS dataset format using a SAS data step, as requested in the assignment. (You will need to change the location of your folder and the file you uploaded.)



The screenshot shows the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'Claims.txt' selected. The main editor displays a SAS program with the following code:

```
1 libname mydata '/home/stefaniegreay/';
2
3 FILENAME REFFILE '/home/stefaniegreay/Claims.txt';
4
5 data mydata.claims;
6     infile REFFILE delimiter='09'x MISSOVER DSD firstobs=2 ;
7     informat FirstName $9. ;
8     informat LastName $9. ;
9     informat MOS $9. ;
10    informat DOS best32. ;
11    informat YOS best32. ;
12    informat DRG best32. ;
13    informat Charges best32. ;
14    format FirstName $9. ;
15    format LastName $9. ;
16    format MOS $9. ;
17    format DOS best12. ;
18    format YOS best12. ;
19    format DRG best12. ;
20    format Charges best12. ;
21    input
22        FirstName $
23        LastName $
24        MOS $
25        DOS
26        YOS
27        DRG
28        Charges
29    ;
30 run;
```

The status bar at the bottom indicates 'Line 31, Column 1' and 'UTF-8'. The user is identified as 'stefaniegreay'.

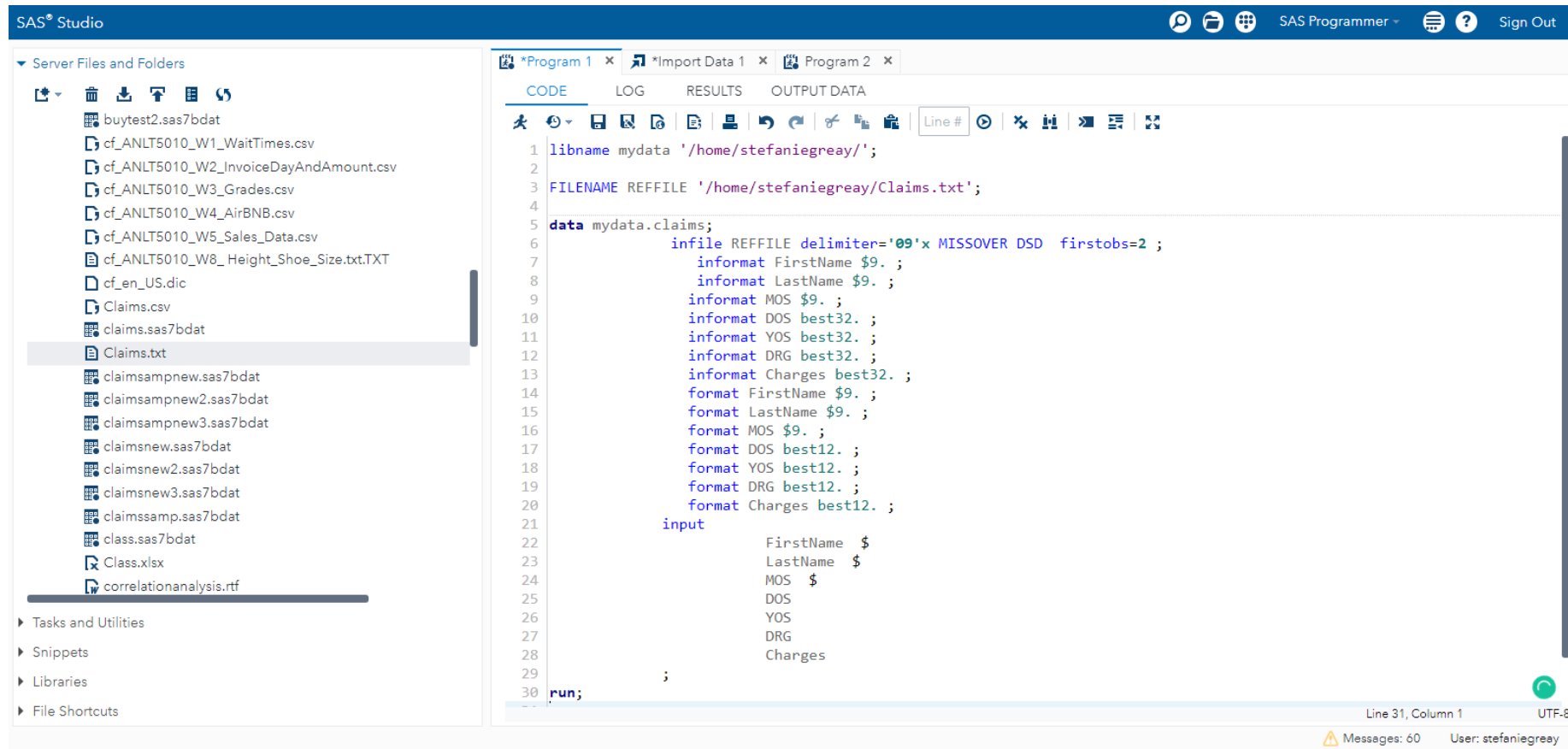
Code to import data in tab-delimited text file using data step.

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

FILENAME REFFILE '/home/stefaniegreay/Claims.txt';

data mydata.claims;
    infile REFFILE delimiter='09'x MISSOVER DSD  firstobs=2 ;
        informat FirstName $9. ;
        informat LastName $9. ;
        informat MOS $9. ;
        informat DOS best32. ;
        informat YOS best32. ;
        informat DRG best32. ;
        informat Charges best32. ;
        format FirstName $9. ;
        format LastName $9. ;
        format MOS $9. ;
        format DOS best12. ;
        format YOS best12. ;
        format DRG best12. ;
        format Charges best12. ;
    input
        FirstName $
        LastName $
        MOS $
        DOS
        YOS
        DRG
        Charges
    ;
run;
```

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

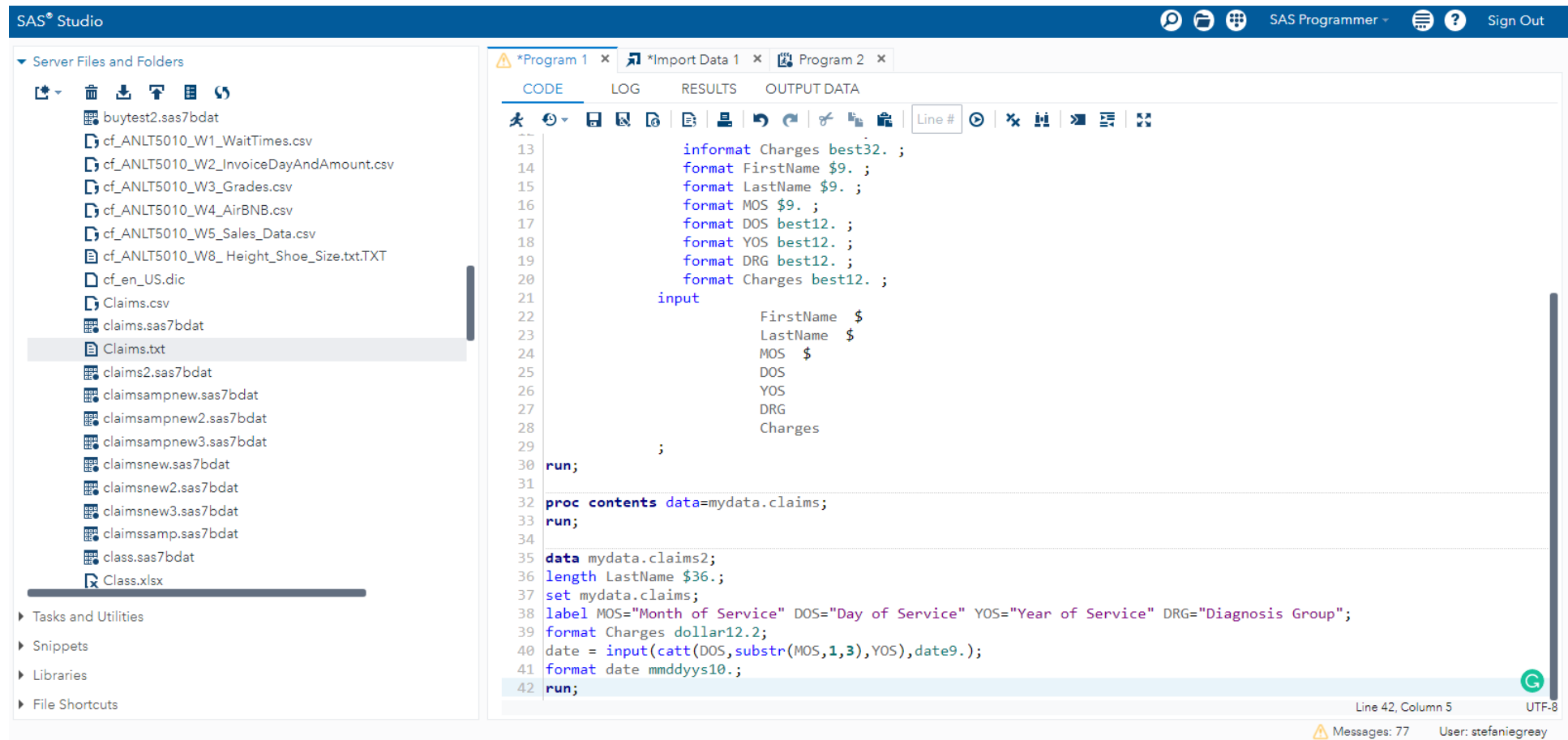


When you run the code, you will see the dataset in the output data window and can verify its success.

The screenshot displays the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'Claims.txt' highlighted. The main window is divided into tabs: CODE, LOG, RESULTS, and OUTPUT DATA. The OUTPUT DATA tab is active, showing a table named 'MYDATA.CLAIMS'. The table has 7 columns: FirstName, LastName, MOS, DOS, YOS, and two unnamed columns. The first 18 rows of data are visible, showing names like Nancy Garcia, Gail Davis, Joan Jones, etc. The bottom status bar indicates 'Messages: 61' and 'User: stefaniegreay'.

	FirstName	LastName	MOS	DOS	YOS		
1	Nancy	Garcia	September	8	2015		
2	Gail	Davis	October	21	2015		
3	Joan	Jones	January	22	2015		
4	Jim	Brown	June	21	2015		
5	Bob	Williams	December	18	2015		
6	Gail	Brown	July	19	2015		
7	Tom	Garcia	January	28	2015		
8	Thomas	Hernandez	June	22	2015		
9	Saly	Brown	January	11	2015		
10	Tom	Garcia	February	28	2015		
11	Jack	Hernadnez	February	9	2015		
12	Bob	Brown	August	26	2015		
13	Thomas	Brown	May	9	2015		
14	Joan	Hernandez	April	12	2015		
15	Jim	Miller	September	14	2015		
16	Gail	Garcia	December	23	2015		
17	Johnathon	Hernadnez	April	18	2015		
18	Sally	Jones	May	18	2015		

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



Sample Code for completing requested data manipulation tasks

```
data mydata.claims2;  
length LastName $36.;  
set mydata.claims;  
label MOS="Month of Service" DOS="Day of Service" YOS="Year of Service" DRG="Diagnosis Group";  
format Charges dollar12.2;  
date = input(catt(DOS,substr(MOS,1,3),YOS),date9.);  
format date mmddyys10.;  
run;
```