**Steps in Assignment 1**

1. Create your account on PythonAnywhere.com with the userID and password provided by PythonAnywhere. I have posted this file on Canvas under Python related materials. Be sure to change your password after account creation.
2. Upload the scripts (with extension .py) posted on Canvas to PythonAnywhere.com

You will need the following scripts for Assignment 1:

**pa\_crawler.py** – this script fetches the posts from Edmunds.com forums. Specify the forum URL and choose the number of pages. I would choose 200 pages. The output of this script is edmunds\_extraction.csv, which will be saved on PythonAnywhere.com itself.

**pa\_wordfrequency.py** – creates a list of all words (minus stopwords) in the input file (currently shown as edmunds\_new.csv, change it to the name of the output file created by pa\_crawler.py). There are two output files from pa\_wordfrequency.py – word\_freq.csv and word\_pair\_freq.csv. Ignore the second output file for now, and work with word\_freq.csv.

**pa\_find\_and\_replace.py** – this script can replace a set of words with another set of words. I have provided a list of car models and brands in the file models.csv. You can add to the list as needed. Note that the script does not know or care about what models or brands are. It simply takes a word from the 2nd column (B) and replaces it with the word mentioned in the 1st column (A).

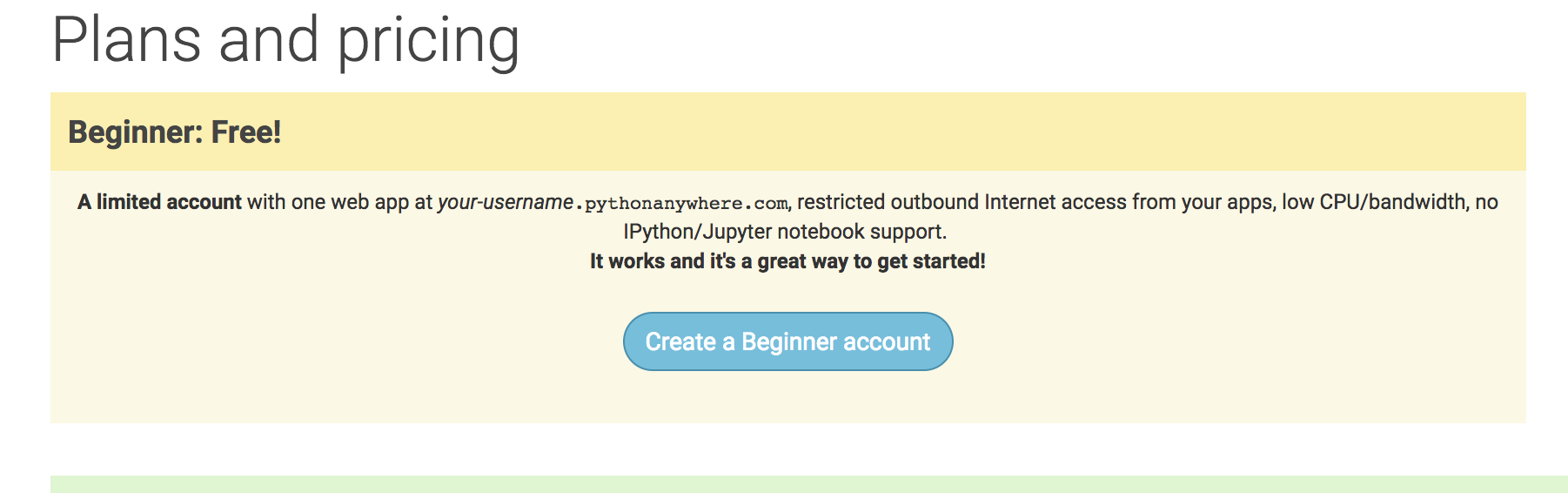
**pa\_lift.py** – this script calculates lift values. It uses two input files – edmunds\_new.csv and a text file edmunds\_pair\_keys.txt where you specify all the words for which you want pairwise lift calculation. E.g., bmw,lexus,luxury,performance (use all lowercase, no space between the words). To make sure that the script finds at least one instance of all pairs of words, introduce a line of fake data in the edmunds\_new.csv file with the following data: bmw lexus performance bmw (for this example, yours will be different depending on lift values you are calculating). The lift values are saved in an output file lift\_values.csv

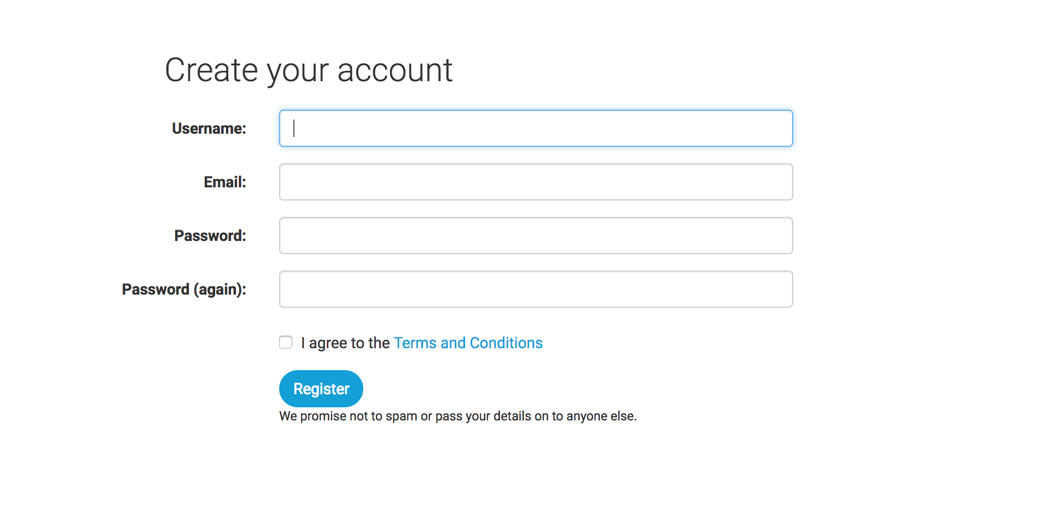
**pa\_mds.py** – this script plots the MDS graph. It uses sample\_dissimilarity\_matrix.xlsx as input. The input file can be constructed from a lift table (you have to create it manually from the output of the pa\_lift.py script. See similarity\_to\_dissimilarity.csv to understand the process; the similarity table has to be constructed from lift\_values.csv. On the same spreadsheet that you created the similarity table, create the dissimilarity table. Now delete the similarity table and retain the dissimilarity table, which will be the input to the mds script. The output of the pa\_mds.py script is mds.png (image file).

**Creating your account on PythonAnywhere.com and running Python scripts**

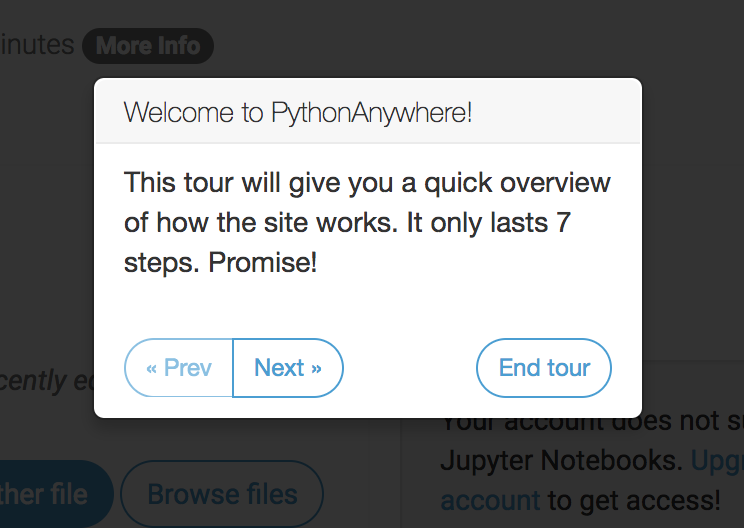
1. Open URL <https://www.pythonanywhere.com/>
2. Click on “Start running Python Online in less than a minute”



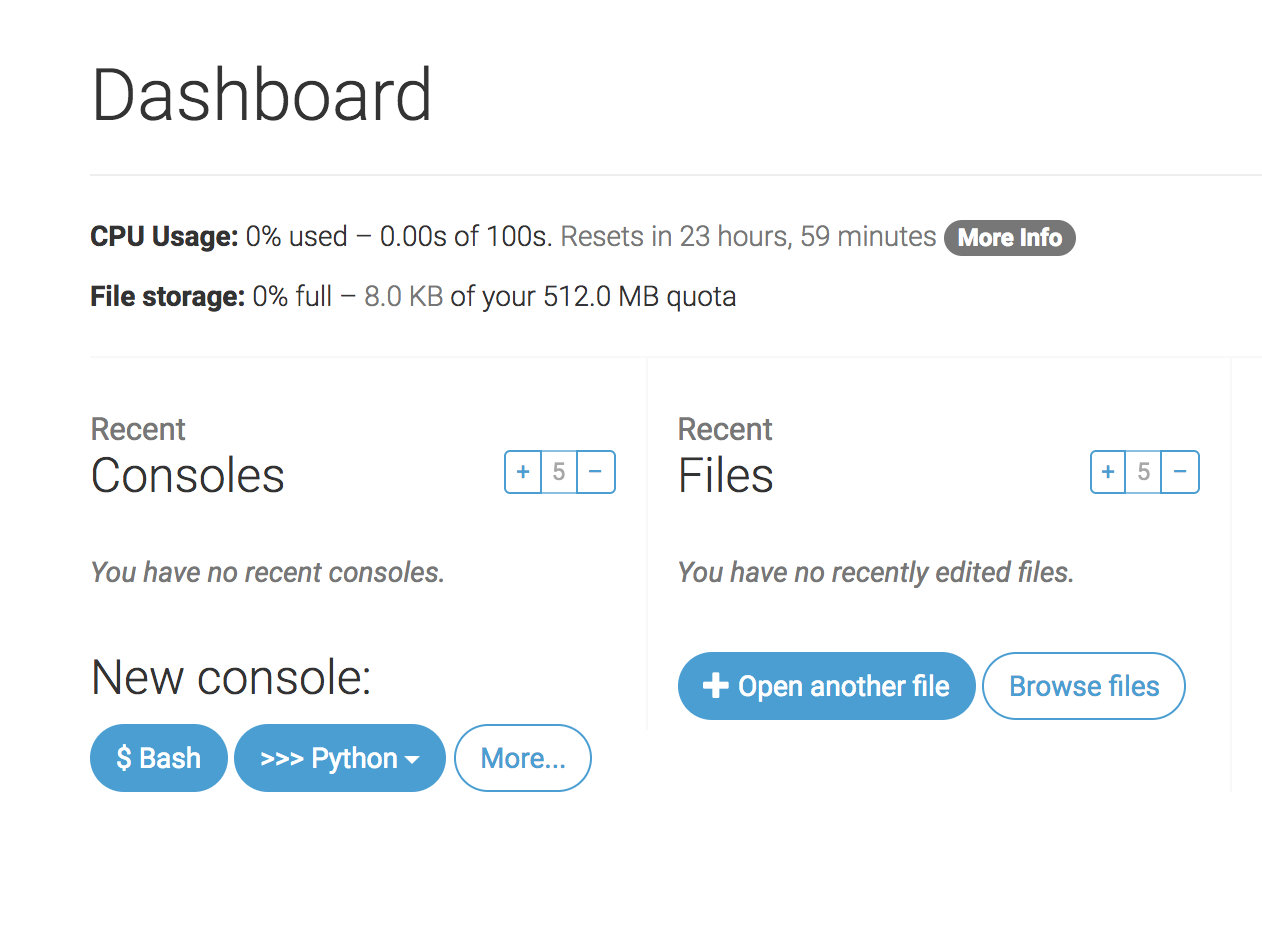
1. Click on Create a Beginner account
2. Register yourself



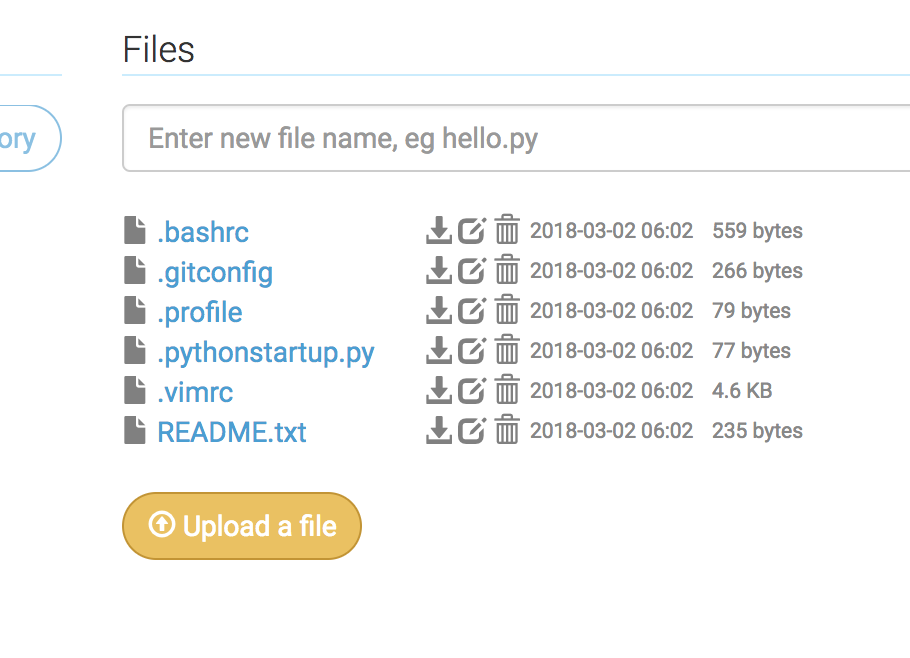
1. Click on End Tour



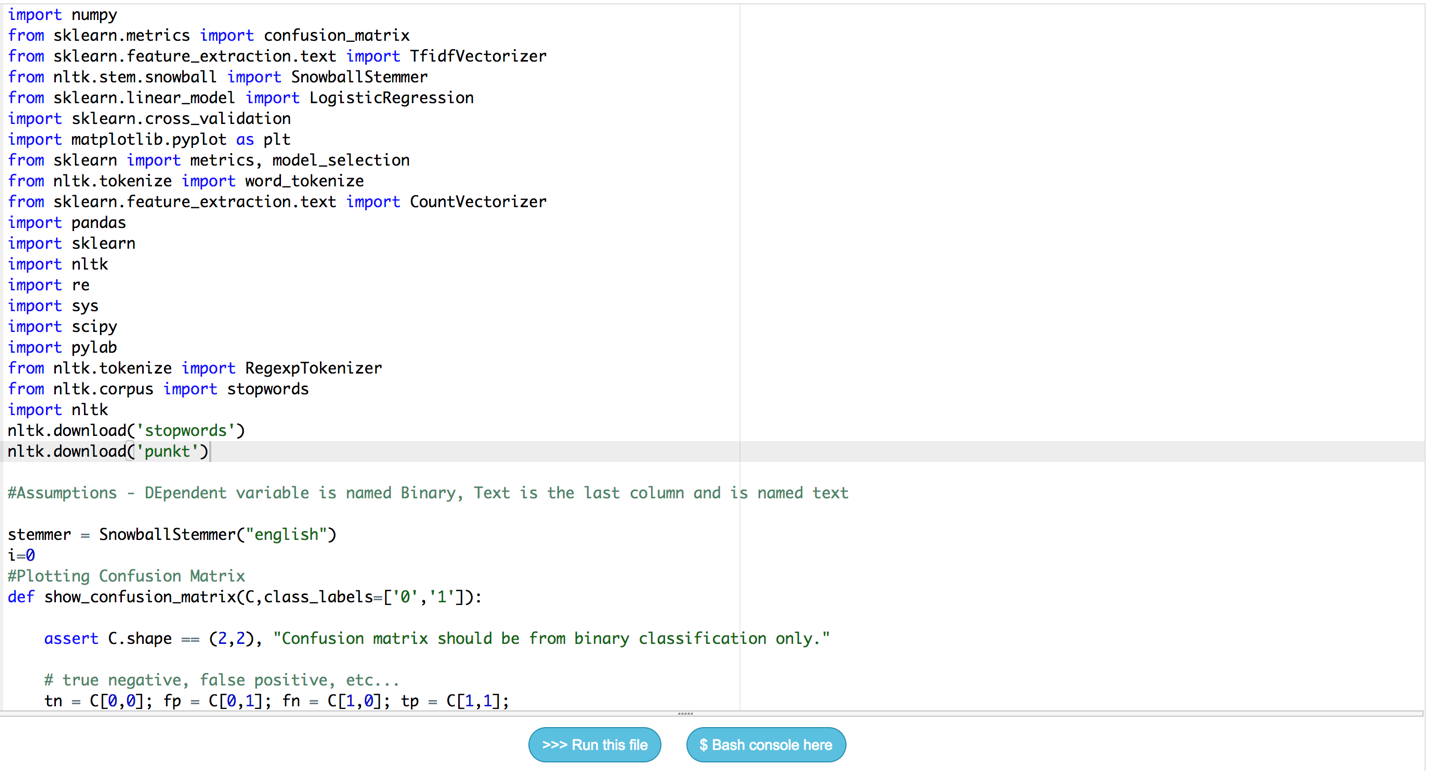
1. Click on Browse files



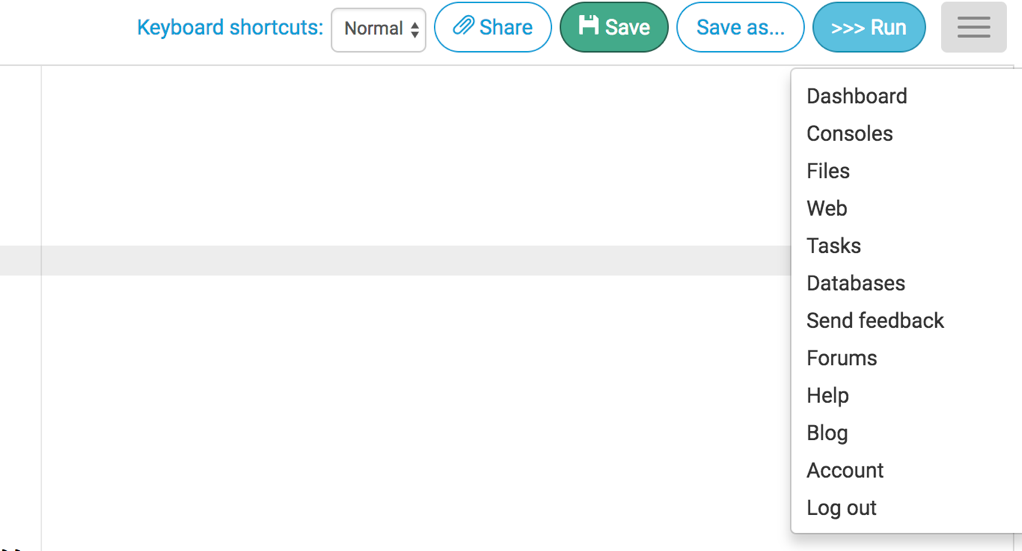
1. Click on Upload a file and upload your Python script and Data Files (if needed) one at a time. E.g., the script pa\_crawler.py does not need an input file. Similarly pa\_wordfrequency.py uses the output file from crawler.py. You may need to change the name of the input and/or output files inside the scripts.

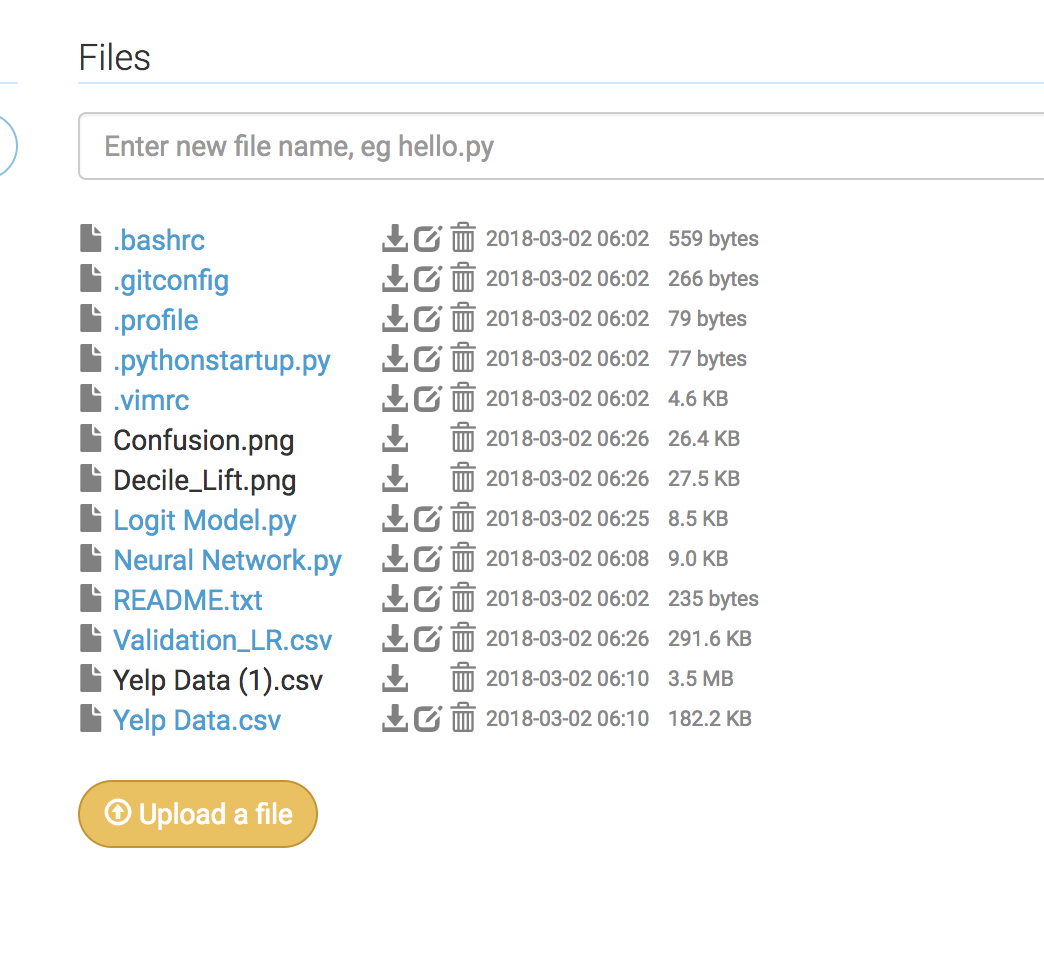


1. To run your code, click on the file name after you upload it (e.g., pa\_crawler.py) and then click on ‘Run this file’ (when you try to rerun again the Run button will appear on the top right)

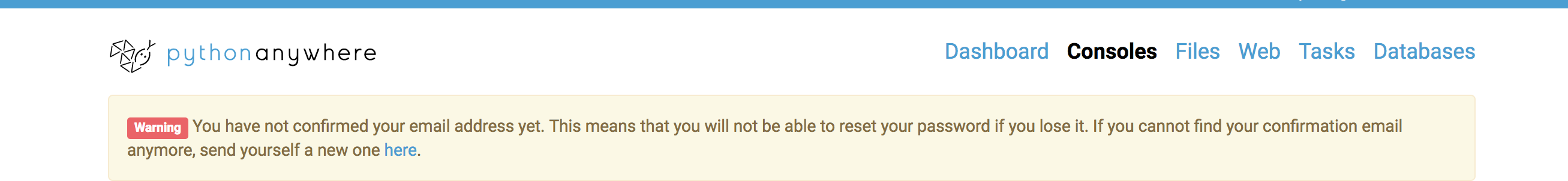


1. On the top right corner click on the three lines and select Files

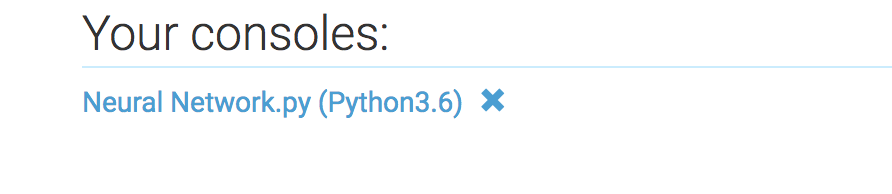


1. Your output files, if any, are saved here (Files Section)
2. To save Memory:

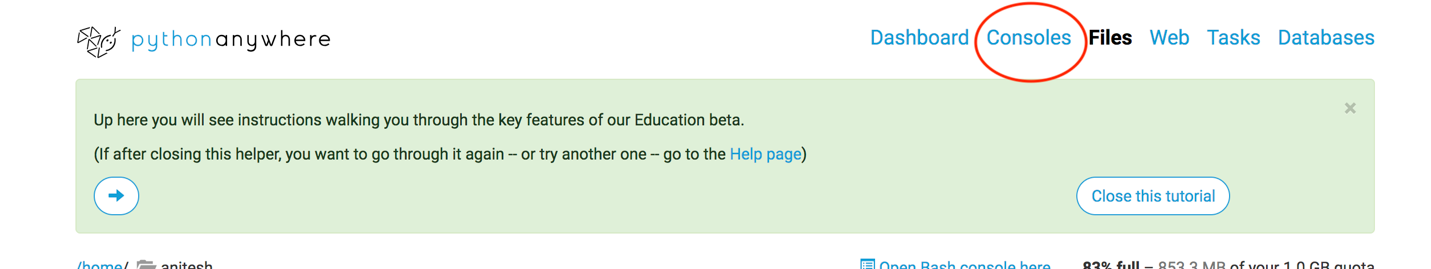
The memory and space is limited. So, make sure you kill your processes once you are done using it. Click on Console



And the click on the cross mark to kill the process



1. In case you get missing library error, click on Consoles



Under Start a new console section click on Bash, and type **pip install –user** followed by the name of the library.

E.g., **pip install –user selenium** to install selenium

