



High-level Instructions

*** See page 2 for step-by-step instructions. ***

- **STEP 1** (Week 10: 16 November approximately):
 - Download a dataset of your choice from online as a **CSV** or **ODS** file.
 - Load the dataset into the spreadsheet program *LibreOffice Calc* and calculate some summary statistics.
- **STEP 2** (Week 11: 23 November approximately):
 - Take the provided template Python program that can load the CSV file and run it.
 - Change the program to do some simple calculations on the data, and print out some values.
- **STEP 3** (Week 12: 30 November approximately):
 - Change your Python program so that it can load the CSV file and output an HTML file that contains some of the data and the results of some calculations.
 - Upload the Python program to *PythonAnywhere*, run it there, and set up *PythonAnywhere* to share the resulting HTML file.
- **STEP 4** (Week 13: **7 December 9:30am**):
 - Change the Python program so that that can load the CSV file, output the entire file as HTML, and generate some summary statistics on the file.
 - Upload this all to the correct folder in *PythonAnywhere* **by the due date and time** so that it can be viewed from the web.
- Ensure the files are in the correct place so that your submission can be marked.

Submission Notes

- All steps of this assignment will be evaluated only after the final submission time (7 December 9:30am) has passed. However, it is recommended that you get the earlier steps done by their respective due dates.
- **By the final due date and time, you need to upload your files to the place specified in the assignment document so that they can be downloaded by our markers.** The correct places for the files are (substituting YOURUSERNAME for your own PythonAnywhere username):
 - `/home/YOURUSERNAME/assignment2/a2_input.csv`
 - `/home/YOURUSERNAME/assignment2/a2_output.html`
 - `/home/YOURUSERNAME/assignment2/a2_process_data.py`
 - `/home/YOURUSERNAME/assignment2/a2_processing.ods`
- **To have your assignment counted against your grades, attend your demonstration session**, which will be announced in a separate schedule and will be after the due date of the assignment.
- Check the separate evaluation form to see on what basis your markers will be grading you.
 - Use all of the data processing techniques shown in the evaluation form.
 - Use all of the Python techniques shown in the evaluation form.
- Have fun.

*** Keep your eye on the separate evaluation form, for the marks. ***

Step 1

Finish by approximately 16 November.

Download a dataset

Search the internet for a spreadsheet dataset that catches your interest. For example, if you are interested in the subject of climate change you might check the data at <http://datahub.io/organization/climatedata>, or if you are interested in socio-economic gender equality, take a look at <http://datacatalog.worldbank.org/?Topics=Gender>. And so on. Most data in this kind of repository can be exported in at least one of CSV, ODS, XLS or XLSX file formats.

Some important points:

- If the data is provided in CSV format you will need to keep the CSV file for processing with Python (see steps 2-4) and also use LibreOffice Calc to convert it to ODS for processing with LibreOffice Calc.
- If the data is provided in ODS format you will need to keep the ODS file for processing with LibreOffice Calc and also use LibreOffice Calc to convert it to CSV for processing with Python (see steps 2-4).
- If the data is provided in a format like XLS or XLSX, you need to use LibreOffice Calc to convert it to both CSV and ODS.

When it comes time to submit your assignment, the resulting CSV file should be named `a2_input.csv` and the ODS file (which should also include the data processing activities from this assignment) should be named `a2_processing.ods`.



Accompanying the assignment PDF, the zip file `a2_template.zip` contains CSV and ODS files with the desired names (you will need to replace these with the files that you download).

Calculate summary statistics

Taking particular attention to the evaluation criteria in the separate evaluation criteria sheet, load your ODS file into *LibreOffice Calc* and do some exploratory data analysis.

For example, try calculating some summary statistics using functions such as COUNT, SUM, AVERAGE, or similar (the easy set), as well as COUNTIF, SUMIF, SUMPRODUCT, or similar (the difficult set). Also, try sorting, filtering, making charts and formatting the spreadsheet for readability. The resulting file will be called `a2_processing.ods`.



You can create multiple “worksheets” in the spreadsheet if you wish to do different operations on different worksheets.

Think about what these statistics and visualisations might tell us about the topic of the data.

Step 2

Finish by approximately 23 November.

Use Python to run the program that loads the file

In the file accompanying the PDF (`a2_template.zip`), there is a Python program file called `a2_process_data.py`. Ensure that that file, together with the CSV data file from the previous step (`a2_input.csv`) are in the current directory. Now if you run the Python program file, you should obtain the following output:

```
This assignment (assignment 2) hasn't been finished.  
All it can do is print out the contents of a couple of cells of the file a2_input.csv:  
Cell at index 0,0:  
Ada  
Cell at index 0,1:
```

Lovelace
Cell at index 1,0:
Grace

Note: To run the Python program from the command line you would use the `cd` command to reach the correct directory and then run the Python command – for example, if your files are in the directory `/home/djduffstudent2/assignment2`, you would write:
`cd /home/djduffstudent2/assignment2`
`python3 a2_process_data.py`
Note that we use the `python3` command as we are using Python version 3.4.



**FOR THIS ASSIGNMENT WE ARE USING PYTHON VERSION 3.4.
Do not develop your code for any version less than 3.3 because it probably
won't work for the markers.**

Make some changes to the Python program

Now you can edit the Python program file to change its output. In the program, the code to read the CSV file is already written and you do not need to change it. Thereafter, the variable `contents` will contain the content of the CSV file. `contents` will be a “list of lists”.

*** Make sure that you read the comments in the Python file so that you know which parts you should change. ***

Try making the following changes:

- Print the variable `contents`.
- Print the variable `contents[0]`.
- Print the variable `contents[0][0]`.
- Print the type of the variable `contents` using the `type()` function.
- Print the type of the variable `contents[0]`.
- Print the type of the variable `contents[0][0]`.
- Print some other members of the `contents` list of lists, if they exist. E.g. `contents[5][6]`.
- Print the type of some other members of the `contents` list of lists, if they exist. E.g. `contents[5][6]`.
- Multiply, subtract, exponentiate, and add some of the numerical values; save the result in variables with descriptive names (like `agetotal`, for example).

*** Saving a value to a variable saves it in memory; it is not the same as saving to a file. ***

- Try to calculate a mean average of some variables and save the result in a variable.
- Print the variables that you have saved.
- See what happens if you try to print variables that don't exist, such as `contents["chickenchicken"]`. Keep a note of the error message.
- Concatenate some of the strings and print the result.
- See what happens when you try to concatenate a number and a string. Keep a note of the error message.
- See what happens when you try to multiply a number with a string.
- Print the types of the results of these calculations.
- Run the `help()` function on some of the variables you have created.

Note: If you would like to work in the `ipython3` development environment, make sure you run `ipython3` from the directory containing your program and data files. For example, if your files are in the directory `/home/djduffstudent2/assignment2`, you would write:
`cd /home/djduffstudent2/assignment2`
`ipython3`

Then run the following *ipython3* command:

```
%run a2_process_data.py
```

Thereafter, the `contents` variable will be available to you to manipulate as you please. For example, try typing:

```
contents
```

```
contents[0]
```

```
contents[0][0]
```

Note: Python programs are not normally run via *ipython3*; only for purposes of “debugging” them.

Once you have finished with *ipython3*, you can put any commands you need to run again into your program file.

Step 3

Finish by approximately 30 November.

Output HTML from your Python program

Edit your program `a2_process_data.py` so that it produces valid HTML 5 as output instead of plain text. If the output of the program is saved to file (called `a2_output.html`), that file can then be loaded into a browser and displayed as a web-page or uploaded to an HTML 5 validator like <http://validator.w3c.org/>.

Using the command-line the necessary command would be:

```
python a2_process_data.py > a2_output.html
```

*** The “>” character is called a “redirection” and redirects program output to the given file-name. ***

As an example, if your program previously outputted the text on the left, it should now output the text on the right:

Non-HTML	HTML
Mean temperature in India for 2012 was 24.23. Mean temperature in India for 2012 was 24.69.	<!DOCTYPE html> <html> <head> <meta charset="utf-8"> <title>India Temperature Analysis</title> </head> <body> <p>Mean temperature in India for 2012 was 24.23.</p> <p>Mean temperature in India for 2012 was 24.69.</p> </body> </html>

Upload everything to PythonAnywhere and publish it

Using the *PythonAnywhere* file manager, create a folder called `assignment2` in your home directory. You learnt how to do this for assignment 1.

Using the “Web” tab in the *PythonAnywhere* dashboard, create the static files mapping, mapping from the URL `YOURUSERNAME.pythonanywhere.com/assignment2/` to the path `/home/YOURUSERNAME/assignment2`. You learnt also how to do this for assignment 1.

Now upload the files that you have created so far so that typing the following addresses into a web browser should download or display the files that you uploaded:

- http://YOURUSERNAME.pythonanywhere.com/assignment2/a2_input.csv
- http://YOURUSERNAME.pythonanywhere.com/assignment2/a2_output.html
- http://YOURUSERNAME.pythonanywhere.com/assignment2/a2_process_data.py
- http://YOURUSERNAME.pythonanywhere.com/assignment2/a2_processing.ods

As previously, in all of the above, `YOURUSERNAME` should be changed to your *PythonAnywhere* username.

Step 4

Finish by 7 December 9:30am.

Output the whole input CSV as HTML

In the previous steps you changed the provided Python program that loaded the input CSV so that it did some calculations and outputted some HTML. Now you will add to that program another feature.

Now change the the program so that it loads the input CSV and prints the *whole file* as appropriately formatted HTML. Since the CSV represents a grid of data it would be advisable to print HTML that presents the data in a grid.

To do this properly you will need to use flow-control. In particular, you will need to write either `while` loops or `for` loops.

Calculate and output summary statistics

In previous steps you have not been asked to calculate statistics on the data because doing so properly requires you to use flow control – in particular, `while` loops or `for` loops.

In the first step you will calculate a summary statistic (such as a spreadsheet may calculate in a `SUM` or `COUNT` or `AVERAGE` function). For example, you might average a column of numbers.

Next, you will calculate a more sophisticated summary statistic (such as a spreadsheet may calculate in one of the more advanced functions like `SUMIF`, `COUNTIF` or `SUMPRODUCT`). For example, you might average numbers in a column matching a condition..

The output of this will be formatted as HTML as you have been doing since step 3.

Make the output into a readable document

You now have a program that outputs HTML containing the whole dataset presented nicely, as well as some statistics calculated on that dataset. Now change the program so that the output HTML is in the form of a report containing the data as a table, the summary statistics and a brief (1-2 paragraphs) report about this investigation. In particular, write about what questions about the data itself your brief investigation may have answered. For example, if your dataset was about India temperatures for the last 100 years, has India warmed a lot since the 1960s? Do summer and winter temperature trends follow the same pattern?

Upload to PythonAnywhere

Upload the four files you have created (the CSV, ODS, PY and HTML files) to the assignment2 directory on *PythonAnywhere* that you created in step 3. Ensure that they are in the correct location to be collected for marking. Pay special attention to getting the file-names correct and uploading well in advance of the deadline. The uploaded files will be marked and will be the subject of your demonstration session (which you should ensure to attend to get your marks).



If you get the code and output uploaded to PythonAnywhere in the correct location by the due date and time, it will be marked.