CM50266 Applied Data Science

Lab 1: Weather Visualisation

Deadline

Lab Due	8pm 30th October 2020. (End of week 5)
Peer Assessment Due	8pm 13th November 2020. (End of week 7)
Marks	/10 (10% of overall unit mark.)

Data

You are provided with four data files that contain daily values of outdoor temperature, indoor temperature + humidity, barometric pressure and rainfall covering a year. They are in CSV format. They have been uploaded to the unit Moodle page and can be found in lab1data.zip.

Task1 (5 marks)

Write a python program to parse the CSV files. Read the data from the CSV files into an appropriate python data structure. For the file handling part of this task you should **NOT** use an existing CSV reading and parsing library such as PANDAS. Instead you should code the capability from scratch. You are recommended to structure your code to allow re-use of the core parser as you may find it useful in the future. You may use libraries including PANDAS once you have read the data, including to convert datatypes from their raw input type.

Your code should correctly parse the CSV file provided. If a file is read that does not conform to the variant of the CSV file provided your program should not generate any unhandled exceptions.

Compute the minimum, maximum, mean and standard deviation for each component of the weather data and report this. In task 2 you may wish to account for gaps or other issues in the data provided, but in task 1 you should compute these values for all the data provided.

Using either a suitable python library to plot pairs of the data or Microsoft Excel, visually check for correlations between the weather components.

Your program will be assessed according to the following criteria:

Assessment Criteria for Task 1	Marks Awarded
Does the program correctly parse the csv file into a python data structure?	1
Does the program correctly deal with mal-formed input, that is avoid throwing unhandled exceptions?	1
Is the code that reads the csv file wrapped in a function that takes a filename as input and returns the data? And does it not make use of any existing csv reading and parsing library such as Pandas for the file handling?	1
Does the program correctly output the minimum, maximum, mean and standard deviation for each component?	1
Has a complete set of plots been provided to allow identification of correlations between variables.	1

Task 2 (5 marks)

Extend your code to summarise the data into monthly and seasonable values (min, max, mean and SD, or mean and confidence intervals) for the weather components Humidity, Indoor Temperature, Outdoor Temperature, Rainfall and Barometer. From this summary data construct two infographics (one monthly and one seasonal). These

infographics should visualise the summarised data concisely and clearly. You should incorporate any additional data into your infographic that helps you to convey a clear and interesting message for the reader.

Your infographics will be assessed according to the following criteria:

Assessment Criteria for Task 2	Marks Awarded
Do both of the infographics illustrate the required summary data (either min, max, mean, SD, or mean and confidence intervals) for the weather components Humidity, Indoor Temperature, Outdoor Temperature, Rainfall and Barometer without misrepresenting the data?	1
Do the infographics employ effective visual mappings (encodings of data as visual properties) according to Mackinlay's ranking of visual properties, and are these visual mappings justified in a short explanation to accompany the infographic?	1
Do the infographics enable effective (fast and accurate) comparisons between monthly and seasonal values? (i.e. The data should be visualised such that it makes comparisons easy to perform – tables are not effective)	1
Do the infographics use colour effectively to draw attention and encode data, and is the use of colour justified in a short explanation to accompany the infographic?	1
Do the infographics integrate text/labels effectively without adding unnecessary 'chart junk'?	1

Submission

Lab Submission:

- Via Moodle ZIP containing your code, your infographics in PDF format, and a few sentences of accompanying text to provide explanations to your peer marker about your decisions regarding visual mappings and use of colour in Task 2 (in a clearly labelled text file or on separate page within the PDF document).
- The set of graphs and result requested in task 1 (included as clearly labelled files in your submission or inserted within a PDF document)

Peer Assessment Submission:

- Via online form.
- Complete one entry for each submission you are reviewing.
- You may also submit an entry for your own work

Peer Assessment

This unit will make use of peer assessment. This means that after the initial deadline for a piece of coursework you will be allocated the work of three other students to examine and assign a mark. This will allow you to see how others have tackled the same problem. The purpose of this is to expose you to issues you may not have identified for yourself and improve your understanding of the problem being tackled.

You will be provided details of how to download the three submissions. You are expected to examine these and compare them to the assessment specification given in this document. Each of the criteria is designed to be a simple pass/fail assessment where the submission either meets the requirement or it does not. Where any criteria are not met, you must indicate why you have reached this conclusion.

You will be given a link to an online form where you can submit an entry for each submission you examine. You must also submit an entry for your own work. You are strongly recommended to assess your own work after you have reviewed the work of the other students. You must submit all the forms by the peer assessment deadline.

There are no additional marks for completing the peer assessment. However, a penalty of up to 50% will be applied to your lab mark should you fail to complete the peer assessment satisfactorily.

A satisfactory assessment entry means you will have completed a form for each submission allocated to you and provided a valid justification for each of the criteria you have labelled as not met.

The work you submit should be anonymous and not include your name or userid. You should remove any reference to your username in any pathname in your code. Replace it with 'username'. You must not engage in discussion of your mark or the marks you will allocate to your peers with your peers. You should report any attempt by others to influence the marking process.

Mark Calculation

Your mark will be calculated in the following way:

1. The two closest peer marks given will be used. If the three marks are equally spaced, the pair closest to your own estimate will be used.

2. If your own mark estimate lies above the peer marks you will receive the mean of the peer marks.

3. Otherwise, you will receive the mean of the two highest marks. (The two peer marks and you own estimate.)

Your mark will be returned to you once this processing has been done. You will also receive the details of the marks allocated by your peers. This will include their reasoning. This is a provisional mark. If you do not consider the mark to be fair, you can contact the lecturer and ask for it to be reviewed. Your work will be re-marked and where the lecturer determines a different mark, the peer marking will be checked and any unsatisfactory marking will have the penalty applied. Should your request for a review not be justified, a penalty may be applied to your mark as you will have further demonstrated that you have not properly understood the material or the feedback you have received.

After the review period the coursework mark will be finalised. To maximise your marks, you should attempt to be as accurate in the marking of both the peer work and your own.

Extensions

If any student is granted an extension, they will still have to undertake peer marking of others work after their updated deadline, with appropriate extensions. Their own work may be peer marked or assessed by the lecturers/tutors depending on the availability of peer markers at that time.

Plagiarism

Plagiarism is a serious offence - please check: http://www.bath.ac.uk/library/help/infoguides/plagiarism.html and http://www.bath.ac.uk/students/support/academic/academic-integrity/

KMC/2020