MATH36031 Project 3 - deadline 16th December 2022, time 1100hrs.

For this project you need will need to download the *bananas.csv* file on Blackboard, and it will located in the Projects folder in the Project 3 section. This is a very large datafile with almost 12000 entries, so do not try print the file! The first few lines of the file are as shown in figure 1. The *Origin* header denotes the country of origin of the bananas. The *Date* header

	Origin	Date	Price	Units
1	'costa_rica'	2022-11-11	1	'£/kg'
2	'guatemala'	2022-11-11	0.9100	'£/kg'
3	'dollar_bananas'	2022-11-11	0.9600	'£/kg'
4	'all_bananas'	2022-11-11	0.9600	'£/kg'
5	'colombia'	2022-11-04	0.9700	'£/kg'

Figure 1: The first few lines of the *bananas.csv* file.

shows the date for the data listed, the *Price* header column shows the cost in pounds sterling for a unit of the product recorded on that date. The *Units* header indicates the unit (such as pounds per kilograms in the above sample).

You need to process the file using MATLAB to answer the following questions:

1. From the data produce a list of all the distinct entries under the *Origin* header and check how many distinct *Units* there are.

2. The price of the bananas fluctuate a lot during the year. For each name listed under the *Origin* header find the mean of that variety.

3. Produce a grouped box plot comparing the variation of the prices of the different varieties of bananas with *Origin:*

   'colombia', 'costa_rica', 'dominican_republic', 'honduras', 'jamaica', 'windward_isles', 'mexico'

   and comment on your results.

4. Taking the time series for the variety with *Origin* 'colombia', analyze the time series and comment on any seasonal trends.

5. Use the **corrcoef** function to calculate the correlation coefficients between the fluctuation of prices for the variety with *Origin* 'colombia' and 'costa_rica'.

Outputs required You are required to submit a report (maximum 8 pages including any appendices) in pdf form via the submission box on Blackboard. Additionally you need to submit your m-files used for the MATLAB codes in answering the above questions.

Additional information and guidelines

1. All coding must be done in MATLAB.
2. Keep your page length **not exceeding eight A4 pages,** with a font size no smaller than 11, and page margins no smaller than 2cm. There is no need for a title page for a relative short report like this. **If more than 8 pages are submitted only the first 8 pages will be marked and the rest of the submission will be ignored.**
3. List the complete code at the end of each question, or in an appendix. Make your source code more readable, by keeping the indentation and stylistic features, and can be copied from your submitted. Your published results should be reproducable from the code attached.
4. Have a look at the generic rubric about how your report will be marked, and also the intended learning outcomes about what you are expected to achieve in the end.
5. Avoid copying (too many) sentences directly from the project description, and try to restate the problem with your own words or examples if possible.
6. You may use your report in the future as evidences of written work, so take it seriously.
7. Your target audience is a fellow student on your course: explain the questions so that the report can be understood without this project description and your approach can be implemented in another computer. The report should indicate to the reader how well you understand the problem and the approach you have taken, the validation and other checks that you made to ensure your results are credible. Reports submitted containing codes only and with no explanations of how the problem was solved, will result in a failing mark, even though the codes may work perfectly well and give the correct answers.
8. Balance the explanation of the approach and the comments in the code. Avoid under-commenting and over-commenting.
9. Aim for precision and clarity of writing.
10. Since there is no final exam, you are advised to spend at least 15 hours on each project, with additional self-study if you are less experience with computer programming. Remember for a 10 credit module like this one, you are expect to spend $100 = 10 \times 10$ hours in total (including lectures, labs, self-study and coursework).
11. Please do not put any personal information on the report, only your student ID number.
12. The submission box (via Blackboard and Turnitin) for each project will be open two weeks before the deadline, and you are encouraged to submit an early draft to see how Turnitin works on Blackboard. Only your last submission will be marked, and DO NOT submit anything after the deadline. Any late penalty will be applied by the Teaching and Learning Support Office according to the Undergraduate Student

Handbook, and any extension has to be approved from the Office too (not from the lecturer).

13. This coursework is meant to be an individual piece of work and the reports must be written by yourself. Your attention is drawn to the section 4.10 on *Academic Malpractice* on page 36 of the undergraduate handbook. Cases of plagairism will be treated seriously and individuals will be reported to the University Disciplinary Committee.