Programming for Data Analytics
Big Project

Worth 50% of overall marks (This is marked out of 100)
Due date: Monday 13rd January 2025
This should give me enough time to correct it and get the marks into the Program and Awards Board.
Please create and copy a link to your repository as soon as possible, you can continue to modify the project up to the due date.

Description:

Write a Notebook that demonstrates that you, can perform data analysis on some data. I will allow a lot of flexibility in this project, so that you can use it as an opportunity to do something that is useful for your work.

If you cannot think of a project to do you may Analysis the wind speed around the country with a view to windfarm:

- You may look for your own source of historic weather information, and/or use the Met Eireann one (Historical Data - Met Éireann - The Irish Meteorological Service). Click on the download button to get a zip file that contains a CSV file.
- You may need to clean and normalize the data before doing analysis
- Questions you can ask:
    - How much wind power is there at a particular location?
        - This is quite open ended, is this just the mean wind speed for an hour/day/month/year, or should you take into account that there are wind ranges that the windfarms can operate in. (min max speeds)
        - Some analysis of what power when would be useful (time of day/year)
    - Are the wind speeds likely to be the same in 10 years in the future? ie is there a trend in recorded wind speeds over the last few decades.
    - Is there any other weather metric worth analyzing (eg rain, temp)
    - What will the power output of the windfarms in Ireland be like next week, according to the weather forecasts? (ok that is a tricky one, because you would need to get, or make up, information about the size and locations of the wind farms in Ireland, one find/makeup the windspeed to power output equation.
    - Anything else you can think of?

Assessment strategy

You have flexibility as to what you do for this project. I understand that this can cause confusion as to what you should do, so here schedule of some of the features you may want to implement.

	Description
Minimum project	1. Read in/acquire/ make up, a number of good data sets 2. Analyze the data using various techniques, we have covered in this module (I do want you to explore with this) 3. Some nice-looking plots to illustrate your findings.
	Features you can add for more marks
	• Increase the range of data you are analyzing
	• Use more features of SciKitLearn and other packages that you find useful
	• Provide meaningful insights
	▪ Increase the complexity of your plots without reducing readability
	• Use a database

To get over 70% your application should be very well laid out, look good and work efficiently and well.
The project should be well laid out and the notebook(s) should be **already run**.
Marks may be deducted for:

- Poorly formatted code, that I find hard to read (Do not over comment your code),
- If I find it hard to run,
- I have to read too much fluffy writing. Keep your writing short and to the point
- I find it hard to understand your GitHub layout. (a README file is handy).

Percentage	For	Description
40%	Code	Well laid out, efficient code. A good range or techniques used.
30%	Documentation	The readme/notebook layout and analysis, descripting succinctly what you have done and any conclusions you have come to. I do not want to read lots of waffle, keep it succinct, but do demonstrate you know what your code is doing.
20%	Research	The range of resources you use. Do tell me where you used the resources. Just putting a load of links at the end of the notebook/readme will not cut it 😊
10%	Consistency	Regular, well commented GitHub commits

Hand up:

A link to the GitHub repository that contains the project (and only this project). The repository should contain:

1. Your notebook(s), and code (if any).
2. A "ReadMe" file:
    i. Explain what repository contains
    ii. Any imports that are required
    iii. How to use the code in the repository (if it is not straightforward.
3. Any other documentation you feel is appropriate.

Deadline:

The deadline is **13th January 2025**.

Best of Luck.

Email me if you have any questions.