Credit Task -I (Unsupervised learning)

Step-1

This task is designed to assess the Credit level expectations.

Step-2

Your tutor will then review your submission and will give you feedback. If your submission is incomplete the tutor will ask you to include missing parts. Tutor can also ask follow-up questions, either to clarify something that you have submitted or to assess your understanding of certain topics.

Feedback and submission deadlines

Feedback deadline: Friday 30 August (No submission before this date means no feedback!)

Submission deadline: Before creating and submitting portfolio.

Required documents

1. Submit a report (pdf format) in Ontrack (https://ontrack.deakin.edu.au)
2. Complete the problem credit task and submit your code file (.ipynb) separately in the OnTrack (https://ontrack.deakin.edu.au).

Background

Urban planning is crucial for creating a comfortable and sustainable environment. Weather plays a key role in developing such an environment. With a microclimate dataset, you can create location-specific plans to enhance the quality of life for the population.

Datasets Description

This microclimate sensor data is publicly available for download from the City of Melbourne's Open Data Portal: https://data.melbourne.vic.gov.au/explore/?sort=modified.

Search for "**Microclimate sensors data**" on the portal to find the dataset. The data includes various environmental readings, such as: Ambient air temperature, Relative humidity, Atmospheric pressure, Wind speed and direction, Gust wind speed, Particulate matter 2.5, Particulate matter 10, Noise.

Evidence of Learning- SIT307

1. We are interested in finding optimal number of groups for this dataset, where the ground truth grouping information is represented by `**sensor location'** feature.
    a. What method shall we use for solving this problem and why?
    b. Is it possible to reduce the number of features used in a clustering model while still maintaining informative clusters? If so, what are some techniques we can use to achieve this without losing significant information?
    c. Find out optimal number of groups, report the outcome and justify your findings.
2. Implement two alternative solutions of Q1 (c), one using DBSCAN. Compare and report the findings.
3. Evaluate quality of the groupings that you have reported as a solution of Q1 (c) and Q2. Based on the evaluation outcomes, report the best solution and explain the results.
4. Quantify and print the relationship among independent variables of this dataset. Calculate two collective variables, using two different methods, that represent the same dataset. Create two two-dimensional plots to display the relationship between these new variables and explain the plot.
5. Is there any loss of information due to the transformation performed in Q4? Explain your answer with evidence.