TAYLOR' S UNIVERSITY
Wisdom • Integrity • Excellence

Module Code: *ITS61504 (Feb 2025)*

Module Name: *Data Mining*

Assignment No./Title	**Assignment (Individual)** **30% (1 – CASE STUDIES ANALYSIS –MLO1)**
Course Tutor/Lecturer	Ts. Dr Anbuselvan Sangodiah
Section No	*[Provide your Section No]*
Student Name, ID and Signature	

Declaration *(need to be signed by the student, otherwise, the assessment will not be evaluated)*

Certify that this assignment is entirely my own work, except where I have given fully documented references to the work of others, and that the material contained in this assignment has not previously been submitted for assessment in any other formal course of study. I have also read and accept the content of the ***documentation and submission instructions section*** *of the assignment 1 question.*

Marks/Grade:	**Evaluated by:**
Evaluator's Comments:	

* Please include this cover page for your project submission

CASE STUDIES ANALYSIS

PURPOSE (MLO1)

The purpose of this assignment is to:

1. **MLO1:** *Demonstrate the ability to preprocess and explore datasets to uncover patterns and insights for a given scenario.*

Objectives

- **Understanding Data Set for Data Mining Applications:**
  - Begin by describing the given data set by your instructor. [the data set name]. *(Example: Provide description of the data set → data set size, the nature of data set, number of columns and etc.*
  - Provide a sample of records of the data set *(Example: screenshot of the first 20 records)*
- **Problem Formulation and Objective Definition:**
  - Define some specific problems that can be formulated from the given data set. The identified problems should be solved using data mining techniques.
  - Clearly state the objectives that are aligned with the identified problems. *(The identified objectives must be addressing the identified problems)*
- **Application of Data Mining Techniques:**
  - Provide detail justifications as to how and why data mining techniques *(classification, clustering, association rule mining, outlier detection)* can be applied to achieve the stated objectives.
- **Implementation of Data Preprocessing:**
  - Apply data preprocessing techniques such as ***data cleaning, data transformation, data integration and data reduction*** to the given data set. *(Provide detail explanation with justifications the techniques/methods that are used in each data preprocessing technique)*
  - Use pre-processing algorithms using *Python* to analyze the dataset. Document the steps taken in implementing these techniques and explain why they were necessary. *(screenshot Python codes when document the steps)*
- **Data Visualization and Interpretation:**
  - Explore the given data set – by using some visualization techniques *(Example: histogram, bar chart, scatter chart, box plot and etc).*
  - Interpret relevant patterns and insights discovered from data visualization. *(Students are expected to provide detail explanation)*
- **Ethical Considerations in Data Mining:**
  - Reflect on any ethical concerns related to the data mining process, including privacy, bias, and fairness.
  - Explain how you ensured the responsible use of data and safeguarded sensitive information.
- **Report Writing:**
  - Prepare a clear and concise report documenting the entire data mining process, from data exploration to the final interpretation of results.
  - Include any challenges faced during the process, the data preprocessing methods applied, and your final recommendations.
  - Present your findings in a structured and professional manner, ensuring clarity for both technical and non-technical audiences. Also, adherence to proper file naming. This includes proper submission of Python codes.
  - **Deadline:** Ensure that the assignment is submitted by the due date via the *MyTimes*.

You are required to carry out this Assignment 1 individually. Generally, in this assignment, you will be analyzing the given data set by your instructor ***(you are neither not allowed or cannot use any other data set other than provided by your instructor)*** by understanding the data set, identifying how data mining techniques can be applied to the given data set to address some specific problems, implementing data pre-processing techniques to ensure the data set are suitable to be applied by data mining techniques.

CASE STUDIES ANALYSIS (30%)

Report Outline

1. Data Set Description
2. Problem Statement and Objectives
3. Application of Data Mining Techniques
4. Implementation of Data Preprocessing
5. Data Visualization and Interpretation
6. Ethical Considerations in Data Mining

Deliverables

1. A well-structured and properly formatted academic document that contains the detailed specifications for the Assignment 1.
2. Python codes (PyCharm Project File/Python Files) & Cleaned data set (in excel format) after pre-processing

***Deliverable 1** must be submitted in two formats *doc* and *pdf*
File format: **AssignmentPart1_Report_SectionNo_StudentIDNo**

***Deliverable 2** must be submitted in zip format
File format**: AsignmentPart1_Codes_SectionNo_StudentIDNo**

Assignment 1 Due Date: 27/2/2025 (Thursday by 5.00pm - MYT) submit via **mytimes.taylors.edu.my** submission link.

Important Milestones for your assignment work (Assignment 1 and Assignment 2)

% of completion work of assignment (CW)

Milestone 1 (Week 1) – CW - 20%: Understanding data set

- Completion of description of the data set.
- Identification and completion of problem statements and objectives.
- Identification and completion of application of data mining techniques.
- Identification and completion of data preprocessing techniques for the given data set.

Milestone 2 (Week 2) – CW - 40%: Implementing data preprocessing

- Implementation of data preprocessing algorithms and documentation of data preprocessing steps.

Milestone 3 (Week 3) – CW - 70%: Data Exploration

- Completion of data visualization and interpretation.
- Identification of ethical considerations.
- Commencement of Assignment 2

Milestone 4 (Week 4) – CW - 100%: Data Exploration

- Fine tuning the work on data visualization and interpretation. *(con't)*
- Fine tuning the work on identification of ethical considerations. *(con't)*
- Completion of final document – Assignment 1
- Identification of algorithms for association analysis (Assignment 2)

Documentation & Submission Instructions

As you prepare your assignment for submission, please adhere to the following guidelines to ensure your work is clear, well-documented, and demonstrates your coding process effectively:

1. Use the provided front page template for your assignment. It should include the full details of yourself. Make sure all information is accurate and clearly presented.
2. As for documenting Python codes, accompany every segment of code with detailed comments. These should explain the function and mechanism of the code, the expected outputs, and any important considerations or potential issues that might arise.
3. The academic Honors Code applies to this assignment. Plagiarized work would not be marked.
4. All assignments are due on the dates given by the instructor on *MyTimes* or a certain day of the week as specified in the lesson plan.
5. All work must be original and if, taken from any works other than yours must be properly referenced using APA Referencing System.
6. Before proceeding to the final submission, you must first submit your PDF document to Turnitin for plagiarism check. Ensure you review the similarity report and make any necessary adjustments to your work to adhere to academic integrity guidelines.
7. Take note: Submission of assignment **DOES NOT** mean that you pass assignment by default.
8. Final Review: Before submission, review your PDF/Word document to confirm that all parts of the assignment are included and that the document is not missing any content.
9. The following circumstances can result in your assignment being rejected or no marks are awarded for assignment:
    a. **Fail to comply with assignment submission due date**
    b. **Incomplete submission of assignment (example: files are missing)**
    c. **Submission of corrupted assignment files**
    d. **Fail to adhere to proper file naming format**
    e. **Submission of plagiarized assignment work**
    f. **Fail to attach/incomplete Turnitin report**

MARKING SCHEME – ASSIGNMENT 1

Items	Marks
Data Set Description	2
Problem Statement and Objectives	4
Application of Data Mining Techniques	5
Implementation of Data Preprocessing	7
Data Visualization and Interpretation	7
Ethical Considerations in Data Mining	2
Overall documentation	3
TOTAL	**30**

MARKING RUBRICS – ASSIGNMENT 1

Criteria	**8-10 marks**	**6-7 marks**	**5 marks**	**1-4 marks**	**0 marks**
Data Set Description	Comprehensive and insightful description of the dataset. Clear identification of relevant attributes and their data types. Thorough exploration of data characteristics (e.g., distribution, missing values, outliers).	Adequate description of the dataset. Identification of key attributes. Some exploration of data characteristics.	Basic description of the dataset. Limited identification of attributes. Superficial exploration of data characteristics.	Insufficient description of the dataset. Incomplete or inaccurate identification of attributes. No exploration of data characteristics.	No description of the dataset provided.
Problem Statement & Objectives	Clearly defined and well-articulated research problem. Specific, measurable, achievable, relevant, and time-bound (SMART) objectives. Justification for the chosen problem and its significance.	Adequate problem statement and objectives. Objectives are generally clear and relevant. Some justification for the problem.	Basic problem statement and objectives. Objectives may lack clarity or specificity. Limited justification for the problem.	Vague or unclear problem statement and objectives. Objectives are not well-defined or relevant. No justification for the problem.	No problem statement or objectives defined.
Application of Data Mining Techniques	Appropriate selection and justification of data mining techniques. Demonstration of a deep understanding of the chosen techniques. Effective application of techniques to address the research problem.	Suitable selection of data mining techniques. Good understanding of the chosen techniques. Effective application of techniques in most cases.	Basic selection of data mining techniques. Limited understanding of the chosen techniques. Some application of techniques to address the problem.	Inappropriate selection of data mining techniques. Poor understanding of the chosen techniques. Limited or ineffective application of techniques.	No data mining techniques applied.
Implementation of Data Preprocessing	Comprehensive and effective data preprocessing steps (e.g., handling missing values, outlier detection, feature scaling,	Adequate data preprocessing steps. Justification for most preprocessing steps. Good implementation of	Basic data preprocessing steps. Limited justification for preprocessing steps. Some implementation of preprocessing	Insufficient or incorrect data preprocessing steps. No justification for preprocessing steps. Poor implementation of	No data preprocessing performed.

	transformation). Justification for each preprocessing step. Proper implementation of preprocessing techniques.	preprocessing techniques.	techniques.	preprocessing techniques.	
Data Visualization & Interpretation	Effective use of various visualization techniques (e.g., charts, graphs, plots). Clear and insightful interpretation of visualizations. Visualizations effectively communicate findings and support conclusions.	Good use of visualization techniques. Clear interpretation of most visualizations. Visualizations effectively communicate some findings.	Basic use of visualization techniques. Limited interpretation of visualizations. Visualizations may lack clarity or effectiveness.	Poor use of visualization techniques. Insufficient or inaccurate interpretation of visualizations. Visualizations are ineffective or misleading.	No visualizations provided.
Ethical Considerations	Clear identification and discussion of potential ethical implications of the project. Demonstration of awareness of privacy, bias, fairness, and other ethical concerns. Steps taken to mitigate potential ethical risks.	Identification of some ethical implications of the project. Awareness of key ethical concerns. Some discussion of steps to mitigate risks.	Limited identification of ethical implications. Basic awareness of ethical concerns. Limited discussion of risk mitigation.	No identification of ethical implications. Lack of awareness of ethical concerns. No discussion of risk mitigation.	No consideration of ethical implications.
Overall documentation	Well-organized and professional presentation of the project. Clear and concise writing style. Properly formatted and cited sources. Effective use of tables, figures, and diagrams.	Good organization and presentation of the project. Clear and generally concise writing. Properly formatted sources. Effective use of some tables, figures, and diagrams.	Adequate organization and presentation of the project. Some clarity and conciseness in writing. Some formatting and citation issues. Limited use of tables, figures, and diagrams.	Poor organization and presentation of the project. Unclear and difficult to read. Significant formatting and citation issues. Ineffective use of tables, figures, and diagrams.	Poorly organized and presented project. Unclear and difficult to understand. No proper formatting or citations. No use of tables, figures, or diagrams.

----------END OF ASSIGNMENT 1----------