DESCRIPTION OF COURSEWORK

Course Code	DSC 101
Course Name	Introduction to Data Science
Lecturer	Toa Chean Khim
Academic Session	2025/04
Assessment Title	Assignment

A. Introduction/ Situation/ Background Information

This course provides a thorough introduction to Data Science. The assessment will cover the data analytical techniques, which used to solve the data related problem. The assessment is aim to impart a wide range of techniques to the students

B. Course Learning Outcomes (CLO) covered

At the end of this assessment, students are able to:

CLO 1 Apply a wide range of techniques in data science to solve data-related problems

C. University Policy on Academic Misconduct

1. Academic misconduct is a serious offense in Xiamen University Malaysia. It can be defined as any of the following:
   i. **Plagiarism** is submitting or presenting someone else's work, words, ideas, data or information as your own intentionally or unintentionally. This includes incorporating published and unpublished material, whether in manuscript, printed or electronic form into your work without acknowledging the source (the person and the work).
   ii. **Collusion** is two or more people collaborating on a piece of work (in part or whole) which is intended to be wholly individual and passed it off as own individual work.
   iii. **Cheating** is an act of dishonesty or fraud in order to gain an unfair advantage in an assessment. This includes using or attempting to use, or assisting another to use materials

that are prohibited or inappropriate, commissioning work from a third party, falsifying data, or breaching any examination rules.

2. All assessments submitted must be the student's own work, without any materials generated by AI tools, including direct copying and pasting of text or paraphrasing. Any form of academic misconduct, including using prohibited materials or inappropriate assistance, is a serious offense and will result in a zero mark for the entire assessment or part of it. If there is more than one guilty party, such as in case of collusion, all parties involved will receive the same penalty.

D. Instruction to Students

This assignment is **a Group assignment (5 persons per group).** Students must work together to create 1 report per group, and each individual will be evaluated separately. The students should submit the following to the Moodle:

1. **Document A:** Includes Assignment Cover Page and Report of the Task. Saved in GroupNo.docx and Group No.pdf format. For example: Group 1.docx and Group 1.pdf.
2. **Document B**: Export your source code into the format Group_NO.html. For example: Group 1.html.
3. **Power Point C**: Create Power point slide for your report. Saved as Group No.pptx. For example: Group 1.pptx. If the file large, upload to online storage (XMUMDrive, OneDrive, etc…) and submit the link.
4. The submission deadline of Assignment is on 23:55, 6 Jun 2025. Overdue penalty (20 marks deduction per day) will be given to project that submitted after deadline. Submission type: Softcopy

1. Submission type: **Softcopy**.

E. Evaluation Breakdown

No.	Component Title	Percentage (%)
1.	**Task 1**	100
	TOTAL	**100**

F. Task 1

The field of data science has gained a lot of attraction from various fields in recent years. The main reason is because of the increase in the amount of data being generated by individuals and businesses. The ability to perform data preprocessing and extract the insight from the data is essential in solving data-related problem. This can help business gain competitive edge in their respective field. In this assignment, each group will be exploring various knowledge and techniques learn in Chapter 1 to Chapter 7 to solve data-related problems. You may refer to the following outline as a guide for structuring your group assignment.

Problem Identification and Planning:
Define the problem your group wanted to solve and the goals needed to achieved. Provide description on the research question, the context, and the expected outcomes.

Data Collection:
Gather the relevant data to solve the problem. The data could be collected from website, database, APIs, surveys, or external sources. Describe the data that you collected.

Data Cleaning and Pre-processing:
Prepare the collected data for analysis. This involves handling missing values, correcting errors, normalizing data, and transforming variables. Briefly describe the preprocessing techniques you applied. Present your clean data in CSV file or another format.

Exploratory Data Analysis (EDA):
Using analysis and analytic (for example descriptive, diagnostic, etc.) to uncover patterns, correlations, and insights of the data. Using visualization (bar chat, pie chart, line chart, etc.) to explore the relationships between the variables. Describe your finding from the EDA.

Model Building and Evaluation:
Provide predictive model (eg. Linear regression) to predict data performance. Evaluate the performance of the model using appropriate metrics (eg: accuracy, r-squared, mean squared error, etc). Describe on the model's performance.

Communicating the Data and Model Results:
Present the finding of your data to the stakeholder. Each group will be required to deliver using slides created with graphic presentation software, like Microsoft PowerPoint. Each team member is responsible for presenting the section of the report that they have contributed to. Include your name in the footer of your presentation slides. The overall presentation will be 20 minutes (15 minutes for slide presentation and 5 minutes for Q&A session).

Documentation:
Ensure that your code, analysis, and decision-making processes are thoroughly documented. Create a visually engaging report by incorporating **images and tables** to enhance the effective communication of content. The report content should be organized with titles and subtitles.

- Titles and subtitles – 14 points font size, Times New Roman.
- The content body - 12 points font size, Times New Roman, 1.5 spacing and **justify** aligned.
- Spell and grammar checked.

APPENDIX 1

MARKING RUBRICS

Component Title		Assignment				Percentage (%)	100	
Criteria		Score and Descriptors					Weight (%)	Marks
		Excellent	Good	Average	Need Improvement	Poor		
		5	4	3	2	1		
Group	**Clarity of Problem Statement and Planning**	The problem statement and planning are exceptionally clear, well-defined, and aligns perfectly with the collected dataset. The goals provided achievable.	The problem statement and planning are clear and well-defined. The goals provided are mostly achievable.	The problem statement and planning are somewhat clear but may be vague. The goals provided are generally achievable.	The problem statement and planning are unclear and poorly defined. The goals provided are minimally achievable.	The problem statement and planning are missing and not defined. The is no clear goals provided	10	
	Effectiveness of Data Cleaning and Pre-processing	Data cleaning and pre-processing are performed thoroughly and meticulous. All the dirty data are handled appropriately.	Data cleaning and pre-processing are well-executed. Most of the dirty data are handled appropriately.	Data cleaning and pre-processing are adequate, had minor issues and lack thoroughness. Adequate dirty data are handled appropriately.	Data cleaning and pre-processing are incomplete, had major issues, and lack thoroughness. Some dirty data are handled appropriately.	Little or no effort had been made for the data cleaning and pre-processing. Dirty data poorly or does not handle.	15	
	Insightfulness of Exploratory Data Analysis (EDA)	EDA is comprehensive and insightful, uncovering key patterns, correlations, and trends. Interesting and clear visualization provided	EDA is thorough and insightful, revealing sufficient patterns, correlations, and trend. Clear visualization provided.	EDA covers the basics and provides some insights, but may miss some patterns, correlations or trend. Adequate visualization provided	EDA is incomplete and missing key insight. Miss to provide most of the patterns, correlation, or trend. Unclear visualization provided.	Little or no EDA conducted. Poor or no visualization provided.	20	
	Appropriateness of Model Selection, Building, and Evaluation	The model selection is highly appropriate for the problem, with a well-justified choice.	The model selection is appropriate and generally justified.	The model selection is somewhat appropriate and the justification may be weak or unclear.	The model selection is inappropriate or poorly justified.	No model selected and Poorly justified.	10	
	Content and Format of Report	Report is exceptionally well-structured, covering all required analyses in depth, with clear explanations, strong logic, and excellent formatting. Tables	Report is well-organized, covering most required analyses with clear explanations. Minor issues	Report covers basic analyses but lacks depth in explanations or contains some structural issues. Tables and figures are present but	Report lacks structure, several analyses missing or incomplete. Formatting is inconsistent, and tables/figures are	Report is poorly structured, missing key content, and poorly formatted. Tables and	15	

		and figures are perfectly integrated and well-labeled.	with structure or formatting. Tables and figures are mostly clear and correctly labeled.	may be poorly integrated or labeled.	unclear or mislabeled.	figures no present or irrelevant.						
Individual								**Marks:**	**Marks**	**Marks**	**Marks**	**Marks**
	Content delivery	The presentation is engaging and professional, with a clear and confident delivery. The presenter demonstrates a deep understanding of the subject matter.	The presentation is clear and well-organized, with a good delivery. The presenter demonstrates a good understanding of the subject matter.	The presentation is generally clear, but lack of organized. The presenter demonstrated a general understanding of the subject matter.	The presentation is unclear and disorganized. The presenter demonstrated a weak understanding of the subject matter.	Poor or no presentation was provided. The presenter provides little or no understanding of the subject matter.	**20**					
	Teamwork and Collaboration	Present active engagement with all team members within group assignment.	Present general engagement with most of the team members within group assignment.	Present engagement up to limited extend and had minor issues with most of the team members within group assignment.	Present poor engagement and had major issues with most of the team members within group assignment.	Present no engagement with most of team members within group assignment.	**5**					
	Questions answering	Present excellent knowledge of the subject. Able to answer all the questions. Interesting explanation of the model deployment in real-world scenario.	Present good knowledge of the subject. Able to answer most of the questions. Good understanding of the model deployment in the real-world scenario.	Present adequate knowledge of the subject. Able to answer some of questions Adequate understanding of the model deployment in the real-world scenario.	Present limited knowledge of the topic. Able to answer some of questions. Poor understanding of the model deployment in the real-world scenario.	Present less or no knowledge of the subject. Unable to answer the question.	**5**					
						TOTAL	**100**					