CDS6214 Data Science Fundamentals

Project (40%)

The aim of this project is to propose a problem that can be solved by undergoing the data science process. You are to be involved in the entire process from figuring out the problem and question you intend to ask, understand, collect data, perform cleaning/pre-processing, explore the data, build data-driven models and visualize the data meaningfully. The data science process also involves the ability to pitch your problem and to convey the message convincingly to the target audience.

Deadline: 11.59pm, 22 June 2025, Sunday of Week 13.

General Instructions

1. This is a group project to be completed in a group of a max of 4 students within the SAME tutorial section. Please refer to your tutor on instructions to register as a group for Project.

2. Based on your tutorial section, your group can only work on dataset within the domain that has been assigned.

Tutorial Section	Domain
TT1L	Health and Medicine
TT2L	Education
TT3L	Sales and Marketing
TT4L	Biology
TT5L	Economics
TT6L	Climate and Environment
TT7L	Engineering
TT8L	Physics and Chemistry
TT9L	Social Sciences
TT10L	Transportation

3. With the assigned domain, propose a *problem* to solve. You should try to uncover insights from your data and do some literature review to gain necessary domain knowledge of the proposed problem. Under that main problem, ask FOUR questions that you would like to answer (note: go for a variety of different types of questions).

4. To make your project engaging, consider creating a scenario within an organizational context where data analysis is utilized to address a problem or to guide future decision-making.

5. The selected dataset should be structured and have at least 10 attributes (columns) and 1,000 instances (rows). The dataset should be less than 50MB in total file size.

Deliverables

Submission and evaluation components will be in the form of:

1. Written Report (PDF)
   - Note: Your report should not be more than 15 pages.
   - The Cover Page, Table of Content and References are not counted as part of the 15 pages. Please use the template cover page.
   - For References, please use APA style. Excluding figures and tables, the content of your report to be written within Normal margin with Times New Roman, font size 12.
   - **Submit only PDF file format.**

2. Python Notebook and dataset *(IPYNB and CSV/XLSX/TXT)*
   - You are REQUIRED to use **Python** for code development and data visualization. Make sure that all the codes are clearly documented and consolidated in a Python Notebook.
   - Include the dataset in the appropriate format as part of your submission. The **dataset link** should be included on the Cover Page of the report.

3. Presentation *(PPTX)*
   - Record your group's presentation, upload it to YouTube and kept as unlisted. The maximum duration is 10 minutes per group. All members in the group must do presentation and introduce themselves during the presentation with camera on. The **YouTube link** MUST be included on the Cover Page of the report.
   - It is your responsibility to ensure the submitted link is accessible and working. Your examiner is not responsible for requesting a new video link if the one you submitted is non-functional. If your link is non-functional, you will receive 0 mark for the presentation.
   - Submit your **presentation slides** in pptx format.
   - Your examiner might require your group to conduct a physical presentation in addition to the recorded presentation if deemed necessary.
   - In your presentation, discuss the topics below:
     (1) Introduction
     (2) Your question/challenge/problem
     (3) Your data analysis and answer to the question/challenge/problem, include visual as much as possible.
     (4) Repeat (2) and (3) for all question/challenge/problem
     (5) Discussion on findings/insight/business decision/challenge/etc
     (6) Conclusion

Softcopy submissions are to be done via **eBwise** before the deadline. Late submissions are allowed but will incur a penalty of 10 marks per hour. Submissions made after 4 hours will receive zero marks.

Do NOT submit any zip file. Do not resubmit after the deadline. Resubmission after the deadline will be treated as late submission and the penalty applies.

Penalties

- 10 marks will be deducted for each hour late after the deadline.
- 0 mark will be awarded for this Project if the content of this Project is plagiarised from any sources
- 0 mark will be awarded for this Project if the group submit the Project 5 hours late.
- 3 marks will be deducted for the video that exceeds 10 minutes
- 3 marks will be deducted for exceeding 15 pages of the report excluding cover page and references
- 3 marks will be deducted for submitting a slide deck that exceeds 10 slides.
- 3 marks deducted for not having a Cover Page.
- 20 marks will be deducted for working with data in different domain other than the one assigned to your tutorial section.

Be aware that plagiarism is a serious offence. Cite all your references! This includes, but not limited to:
- Materials taken from websites, articles,
- Research papers, books,
- Images, videos (YouTube etc.) and other media.

Evaluation Mark Breakdown

Deliverable		Marks (/40)	Totals
Report	- Introduction and objectives: Problem statement, objectives and scope, background information - Dataset - The FOUR questions with answers - Discussions and Insights: Discussion about the implication of the findings/insights gain from the FOUR questions, limitation and assumption, practical recommendation. - Conclusion and References: Technical references with citation, conclusion of key findings - Professionalism: Well-structured report, grammatical and typo graphical error free, labelled and integrated figures	15	
Code	- Data pre-processing/cleaning - Data Analysis/statistical study - Modelling - Evaluation - Code structure and readability - Innovation: creative approach to problem - Professionalism	10	
Visualization	- Data - Results - Models' performance	5	
Presentation	- Content: Clarity and structure, depth of knowledge, accuracy and relevance, originality - Delivery: Clarity and articulation, confident and poise - Visual Aids: Design and visuals used on the slides, easy to read and appealing visuals - Time Management: Adherence to time limit and time allocation to each section	10	
	TOTAL		40

Note: Not all group members will get the same marks. The **contribution** (in %) you filled in the cover page will be used to weight the marks among the group members. Please ensure all members agree to the submitted percentage.