DATA7001
Group Project

Multiple submissions and milestone due dates
30 Marks Group | 10 Marks Individual | 40 Marks Total

Description
The group project can be undertaken in groups of 4-5. You are free to make the groups as you choose. Each group is expected to identify and agree on a specific data science problem, which can be sourced from your domain of choice or discussed with your lecturers. The project is expected to follow the data science process:

Problem solving with data: Use design thinking to formulate an authentic data science problem.

Getting the data I need: Identify what dataset(s) are needed and design a feasible acquisition and ingestion strategy keeping in mind issues relating to data privacy and sampling.

Is my data fit for use: Explore, transform and enrich the data as needed to demonstrate the fitness of the dataset(s) for the intended problem.

Making the data confess: Design and implement analytical methods to interrogate the data and produce insights and/or foresights relating to the data science problem.

Storytelling with data: Create visual representations and narratives from analytical results that facilitate effective decision-making.

Assessment Criteria
Below are the learning objectives for the course as provided in the course profile.
1. Apply design thinking methodology to data science problems
2. Design effective data science processes from question formulation to persuasive story telling with data
3. Develop data-centric approaches to complex business and scientific problems
4. Reason with the fitness of basic computational and analytical models in data science scenarios
5. Work in teams with diverse backgrounds towards authentic data science solutions

There are four milestones and respective deliverables for the project. The following explains the assessment criteria for each of the deliverables:

Project Pitch: The pitch discussions in class are informal (no slides), should be brief (2 minutes or less), and focus on communicating the main idea of your project. You then submit a 300 words summary on blackboard based on the feedback received on the pitch. This carries no marks, but will be used as the agreed scope. Deviations from the pitch may occur as the project progresses. A section in the final report will have to be added, explaining what/why deviations occurred.

Group Presentation: Prepare a 7-minute presentation, which outlines the key elements of your project. The presentation should be structured along the data science process. Where possible include live demos of any computational or statistical methods that you may have designed and implemented. You can also prepare video of the demo ahead of time. The presentation will be assessed on clarity of presentation, and completeness and creativity of the project.

Peer Review: Prepare 3-5 recommendations for improvement of three other projects presented during the in-class group presentations, that is at least one recommendation for each project. Thus attendance of other group presentations will be required. Each recommendation should be no more than one paragraph. Your recommendations should be actionable, specific, and objective. The peer review will be assessed on the above three factors, i.e. are they actionable, specific and objective; as well as the creativity of the recommendation.

Project Report: The project report is the most significant deliverable in the project. The report should incorporate the feedback and recommendations received after the presentations and peer review. There is no minimum or maximum page limit for the report. The report will be assessed on the innovation of the data science solution, technical correctness of the methods designed or implemented, completeness of the overall project, and quality of the report presentation.

There are four criteria that will be used in the assessment of the report:
1. Innovation of the data science solution
   The problem domain you have chosen, the approach you have taken to generate the data science solution, and the significance of your results, are all indicators of the creativity and innovation of your project. Innovation can be demonstrated in the initial description of the problem domain as well as the story telling from the results of your analytical methods.

2. Technical correctness of the methods designed or implemented
   Technical correctness can be demonstrated by providing exploratory results that profile the data in various ways and help the reader understand the fitness of the data for the task at hand. The way the data is captured/retrieved and ingested (getting the data I need) also needs to be explained so that the generative process can be understood. Last but not least, the analytical methods (statistical or computational) applied in making the data confess should be justified against the questions you are proposing to ask from the data.

3. Completeness of the overall project
   The report should include a section to explain all five parts of the data science process. Additionally there should be a section introducing the problem domain and a section outlining related works (to the best of your knowledge) where the problem domain may have been studied by other data scientists. The implementation of your project in jupyter (or a justified programming environment of your choosing) is essential to demonstrate the completeness (and technical feasibility and correctness) of your project. You may include various code snippets, screen shots, visualizations etc. within your report from the implementation as needed. A link to the implementation can be provided and/or a demo (live or recorded) can be made during the group presentation.

4. Quality of the report presentation
   The quality of the report will be assessed on overall presentation including clear and succinct writing, good flow or arguments and report structure.

While there is some freedom in how you structure the final report, it is expected that there will be at least four parts:

I. The essential story or executive summary, including the main aims, the key insights, which may include key graphs/tables, and take away message.
II. The main part of the report, which includes all necessary technical details, following the data science process.
III. A summary section detailing how you responded to feedback from the teaching team and peers, as well as outlining any deviations from the project pitch.
IV. An appendix listing the names and sources of the data sets as well as code libraries, or tools you used in your project

In addition, on a blank page immediately after the title page please include the following statement after discussion with the group.
   We give consent for this to be used as a teaching resource. / We DO NOT give consent for this to be used as a teaching resource.

Reflective Essay: The reflective essay is the final deliverable of the project and should outline in 1 A4 page your most significant learning from the project. Reflect on the choices you made within the project and highlight what went right and what you would do differently and why. The essay is individual and hence can contain a candid account of group dynamics as well as any practical issues faced during the project. The essay will be assessed on clarity and insight shown in a set of well-supported reflections which evidence your learning from experience.

Milestones and Submission Schedule
The following schedule outlines the due dates for the formal milestones for the project. Carefully plan the intermediate or informal milestones (inline with the data science process) with your group members to ensure timely completion.

Project pitch presentation	Pass/Fail	In-class	Wed 16 Sep	Group
Project pitch summary	Pass/Fail	On blackboard	Fri 18 Sep 11.59pm	Group
Group presentation	15%	In-class	Wed 28 Oct	Group
Peer review	5%	On blackboard	Mon 2 Nov 11.59pm	Individual
Project report*	15%	On blackboard	Fri 6 Nov 11:59pm	Group
Reflective essay	5%	On blackboard	Mon 9 Nov 11:59pm	Individual

*For the final project submission, please nominate one group member to upload to Blackboard by the due date. Your submission will consist of a single .zip file containing:

1. your report,
2. any code, jupyter notebooks, tableau files, in a \code\ directory, and
3. your datasets (if small enough) or a *sample* of your datasets (if they are too large --- say 1,000 records) in a \data\ directory.