MAST90106 Data Science Project Pt1: Assessment

Task B: Written report	5% of assessment in MAST90106 and MAST90107 combined

Your group will produce a written report (approximately 2500 words), which describes the problem domain, discusses the related literature and presents a concrete plan for semester 2.

The deadline for submission is **Friday May 27, 11:59pm**

We recommend the use of LaTeX to create the written report. LaTex will allow you to create a professional standard document very easily. It also has a particularly good automatic bibliography processing tools (bibtex). Please use an 11 pt font and 1.5 line spacing. We also recommend that you use Overleaf, a collaborative tool for writing in LaTeX.

The report should contain the following sections (suggested lengths in parentheses):

1. **Introduction (1 page)**: What is your problem? Why is it important?

   How is the problem difficult from a Data Science perspective? For example, difficulty might stem from sheer size, sparsity, degree of noise, structure, etc.

2. **Related work (2 pages)**: What research from the literature tackles the same or similar problems?

   We recommend that you cite at least 10 different sources. It is important that you identify the key resources/research papers describing models/methods that will be most useful when developing a solution for your problem. You may need to consider if there are any commercial systems that are relevant to your problem, and include these in your review.

   In your description of the related work, you should paint a coherent picture of the background rather than simply saying *. . . 'paper X said blah, web site Y said blah'*.

3. **Data analysis / preliminary model development (8-10 pages**): What data do you have? What models have you explored?

   Describe your exploratory data analysis – include data cleaning and processing steps, basic analysis of the data (from the size and scale, to ranges of values for response variables and input features, degrees of correlations etc). If you have preliminary results (from your statistical/computational models), you should include the results in this section.

   Note: all tables and figures must be documented appropriately (e.g. with captions, figure axes labelled, acronyms explained etc.)

4. **Proposal for semester 2 (2 pages)**: How do you propose to answer the client's question?

   How do you plan to use the data (in terms of models, algorithms and other technique) for analysis and prediction?

   What techniques from the **Related work** section will you use? How might you adapt them? How will you evaluate success? What hurdles are you likely to face?

5. **Timeline (0.5 page)**: Include a clear project plan (a Gantt chart perhaps).

   Key tasks and milestones should be listed. The role of each group member should be clearly documented.

The report should be written such that it is accessible to others in your student cohort, as well as to your client.

The group written report is a **hurdle** task. That is, your group must achieve a satisfactory mark on the group written report if you are to continue with the project in semester 2.