QBUS6600
Data Analytics for Business Capstone
Semester 2, 2023

Assignment 2 (Group assignment)

1. Key information

Required submissions:

- Team responsibilities outline (due: September 14; required but not marked).
- Peer review 1: September 15
- Progress report (due: October 4; required but not marked).
- Written report (due: October 16).
- Presentation video (due: October 16).
- Python Code (due: October 16; required but not marked).
- Peer review 2: October 20

Weight: 40% of your final grade.

Length: Your written report should have a maximum of **15 pages** (single spaced, 11pt). Cover page, references, table of contents and appendix (if any) will not count towards the page limit. Please keep in mind that making good use of your audience's time is an essential business skill: every sentence, table or figure should serve a purpose.

2. Problem description

Please start by carefully reading through the Project Outline pdf document posted on the page containing the full dataset for your industry project. **Focus on the Problem Description section of the Project Outline**, especially the corresponding second and third bullet points (Modelling and Strategy). Both your analysis and your recommendations should be in line with the requirements/suggestions provided in the Project Outline.

Your task as a group is to build on the insights discovered in the first assignment and to use any appropriate statistical/machine learning modelling tools to address the questions relevant to your industry project. Ensure that you justify the selection of your final model(s) and interpret the model(s) in terms of the key attributes or features. If your best-performing models are too complex for this interpretation, we suggest that you also consider well-performing interpretable models.

Use the results from your analysis to outline a strategy and provide recommendations to the management team corresponding to your industry project. Please note that your strategy and recommendations should **significantly** expand on anything you proposed in your preliminary recommendations of the first assignment.

3. Written report

The purpose of the report is to describe, explain, and justify your findings to the management team corresponding to your selected industry project. You may assume that team members have training in business analytics, however, they are not experts in statistics or machine learning. The team's time is important: please be concise and objective.

Executive briefing.

You are asked to summarise your findings so that they can be shared with the wider business and, in particular, senior management. This one-page briefing should concisely describe your findings to a non-technical audience and primarily address the business problem. In the briefing you should also outline your recommendations for acting on your findings.

You are limited to **a maximum of 1 page** (which is included in the overall 15 pages).

Suggested potential outline for the main parts of the report (further details below):

1. Business context and problem formulation.
2. Data processing, EDA, and feature engineering.
3. Model building.
4. Conclusions and recommendations.
5. Executive briefing.

Please note that the 'Model building' section should be the **most substantial** part of your written report. You should consider breaking down the longer parts into smaller sections.

4. Presentation video

Create a video recording of an oral presentation to summarise your work and your findings for an executive or managerial audience. For example, the video can be a Zoom recording, in which the team members present while the presentation slides are shared. Please aim for **about 10 minutes** in length. All group members must speak during the presentation, ideally for equal amounts of time. Remember, your business audience will not be experts in machine learning or statistics. Your presentation should translate the analysis, modelling, and findings to business impact.

5. Marking Scheme

Business context and problem formulation	5%
Data processing, EDA, and feature engineering	20%
Model building	40%
Conclusions and recommendations	15%
Executive briefing	10%
Presentation video	10%
Total	**100%**

6. Rubric (basic requirements)

Business context and problem formulation. Your report gives a detailed description of the problem that is being investigated, providing the context and background for the analysis.

Data processing, EDA, and feature engineering. You describe the data processing steps clearly and in sufficient detail, justifying and explaining your choices and decisions. You describe and explain your feature engineering process. Your choices and decisions are justified by data analysis, domain knowledge, logic, and trial and error (if necessary). You describe your EDA process, presenting selected results. Your analysis is sufficiently rich, and your visualizations are insightful. You explain the relevance of the EDA results to the underlying business problem. You interpret the statistical outputs that you provide.

Model building. You clearly describe and justify the models, methods, and algorithms in your analysis. The choice of methods is logically related to the substantive problem, underlying theoretical knowledge, and data analysis. You interpret the estimated models. You note crucial assumptions and whether they are potentially violated. Your overall analysis is rich, comprehensive, thorough, and logical. You implement a sound model selection process. You obtain a high standard of predictive accuracy in line with what is achievable with the methods at the level of experience expected from students taking this unit. You are not misled by overfitting and explicitly acknowledge the limitations of the data and/or methods.

Conclusions and recommendations. The reasoning from the analysis and results to your conclusions and recommendations is logical and convincing. Your conclusions and recommendations are written in plain language appropriate for non-technical audience. Your recommendations are well thought out, carefully developed, and well supported.

Executive briefing. You appropriately summarize your key findings to a non-technical audience, addressing the underlying business problem. Your clearly outline your suggestions for acting on your findings. Your executive briefing is well presented and logically structured.

Writing and presentation of the report. Your writing is concise, clear, precise, and free of grammatical and spelling errors. Your paragraphs and sentences follow a clear logic and are well connected. If you use an abbreviation or label, you define it first. Your report is well organised and professionally presented, as if it had been prepared for a client later in your career. There are clear divisions between sections and paragraphs. Your tables are appropriately formatted and clearly presented. They do not contain irrelevant information. The tables are placed near the relevant discussion in your report. Your figures are easy to understand and have informative titles, captions, labels, and legends. The figures are well formatted and laid out. The figures are placed near the relevant discussion in your report. Your figures have appropriate definition and quality. All numerical results are reported to suitable precision (typically no more than three decimal places, in some cases fewer). The text of your report is **entirely free of Python code**. You follow the University of Sydney referencing rules and guidelines.

Presentation video. Your presentation is well-structured, clear, and insightful. Your presentation is well-performed in terms of logic, confidence, passion, and timing.

6. Python code (required but not marked).

Your group is required to submit the Python code used for the analysis, as a Jupyter notebook or Python script. The code is submitted separately from the report. The code may be examined to verify that your group has done the work. Your code should have comments that clearly indicate which parts correspond to which sections of your report. You should explicitly acknowledge when you borrow pieces of code from external sources.

8. Team responsibilities outline. Progress report (required but not marked).

Your group is required to submit a brief 'Team responsibilities outline' (due September 14) and a brief 'Progress report' (due October 4) – we encourage everyone to start thinking as early as possible about the project tasks and the division of labour. There are **no templates** for these documents – you are free to use **any format** or structure within the guidelines given below.

Team responsibilities outline. Have a group meeting as soon as possible to agree on the division of labour. After the meeting, submit one pdf document for your group, providing your group number and listing the names of the group members. For each name on the list, report whether or not they attended the group meeting, indicate their group assignment task (e.g. EDA), and provide the corresponding expected date of completion. Please try to keep the document under ¾ of a page. This submission will not be marked; however, there is a 5-mark group assignment penalty for not submitting the document by the due date.

Peer reviews. While we work with others, we bring our strengths but also our weaknesses into the space. There are two rounds of peer reviews that are designed for you to benefit from peer feedback. The second review will also be useful when you work on your Assignment 3. Peer review submission will not be marked; however, there is a 5-mark assignment penalty if the review is not submitted by the due date. The penalty is only applied to the individual who is late with their review.

Progress report. Have another group meeting to discuss the progress. Submit your update on the team responsibilities outline: for each task, indicate the percentage completed and the updated expected completion date. Feel free to include any additional comments, such as changes to the tasks assigned and/or division of labour. In addition, provide attendance for the group meeting (yes/no for each student), and the date of the meeting. Please keep the document under one page. This submission will not be marked; however, there is a 5-mark assignment penalty for not submitting the document by the due date.

7. Late Submission of the report/code/presentation video.

Late submissions are subject to a deduction of 5% of the maximum mark for each calendar day after the due date. After ten calendar days late, a mark of zero will be awarded.

8. Disputes

If there is a dispute within a group, please notify the teaching team as soon as possible and provide evidence, so that the dispute may be resolved quickly and equitably. If the teaching team is notified sufficiently early, uncooperative and/or uncontactable group members will be removed from groups and will complete their assignment individually. All group members are expected to make a meaningful and reasonable contribution to the team.