

Marking Rubric Summary

Section	Recommended page length	Marks
Introduction	0.5	5
Data processing	1	5
Exploratory data analysis	4-5	10
Feature engineering	2	10
Methodology	4-5	28
Model validation and comparisons	0.5-1	7
Conclusion	0.5	3
References statement of contribution, meeting minutes	-	7
Writing and presentation	-	10
Jupyter notebook	-	5
Kaggle competition	-	10
	Maximum: 15 pages	Total: 100

Detailed Marking Rubric

	Description	Marks	Total marks
Introduction (0.5 pages)	A detailed summary of the business context is provided alongside a complete description of the business problem.	2	5
	A thorough description of the contents and aim of the report is provided and describes how the report will address the business problem.	2	
	A concise description of the results contained in the report is provided.	1	
Data processing (1 page)	A concise and complete description of the data that is provided.	1	5
	A concise and complete description of how the data is processed is provided.	2	
	A thorough description of how missing values are handled is provided, and the way in which missing values are handled is appropriate.	2	
	The units of relevant variables are not explained e.g. it is not mentioned that 'price' is in AUD.	-1	
Exploratory data analysis (EDA) (4-5 pages)	A thorough description of the EDA process is included and has been carefully curated to present a selection of key results. The EDA studies key variables individually and pairwise using appropriate figures and descriptive statistics.	5	10
	The EDA clearly identifies features in the data that are relevant to model building.	1	
	The EDA clearly identifies outliers or anomalies that can affect analysis and provides a detailed description of how these are handled.	2	
	The report details the relevance of the EDA results to subsequent feature engineering and model building.	2	
	A figure is provided without accompanying analysis.	-1 per figure	
	Analysis techniques have been misapplied.	-1 per technique	
	Visualisation is not appropriate for the data, e.g., a bar chart for time series data.	-1 per figure	
	The business context has not clearly informed the analysis.	-2	
Feature engineering (2 pages)	A complete and thorough description of the feature engineering process is provided. It is clear that substantial effort has gone into designing new feature, which have been appropriately engineered.	5	10
	Choices made during the feature engineering process are well justified.	3	
	A new location-related variable has been created using existing variables and is well justified.	2	
	A feature engineering technique has been misapplied.	-1 per technique	
Methodology (4-5 pages)	The choice of models is clear and well justified.	2	28
	A correct and detailed description of the fitting process for model 1 is provided.	5	
	A correct and detailed interpretation of model 1 with reference to the business context is provided.	3	
	A correct and detailed description of the fitting process model 2 is provided.	5	
	A correct and detailed interpretation of model 2 with reference to the business context is provided.	3	
	A correct and detailed description of the fitting process model 3 is provided.	5	
	A correct and detailed interpretation of model 3 with reference to the business context is provided.	3	
	Crucial assumptions are correctly discussed with reference to the business context.	2	
	Incorrect model interpretation.	-1 per interpretation	
	Does not include an interpretable linear model.	-5	
	Does not include a regression tree.	-5	
	Does not include an advanced model (e.g. bagging, random forest, boosting or a stacked model).	-5	

Model validation and comparisons (0.5 – 1 pages)	The report includes training and validation scores (which have been correctly interpreted from Kaggle).	1	7
	A detailed and insightful comparison of the models is provided with reference to the business context.	3	
	A detailed and insightful analysis on the limitations of the models is provided with reference to the business context.	3	
Conclusion	A brief and concise summary of the work contained in the report is provided and includes all key findings	2	3
	The conclusion clearly relates to the business context.	1	
Writing and presentation	The writing is clear and concise and generally free of grammatical and spelling errors. The paragraphs and sentences follow a clear logical flow and are well connected. The language uses technical terminology correctly.	5	10
	The report is well organised and professionally presented as if it were prepared for a client later in your career.		
	Excessive grammatical and spelling errors.	-1	
	Abbreviations used without being defined.	-1	
	Divisions between sections and paragraphs are unclear.	-1	
	The report contains Python code.	-2	
	Numerical results are not reported to suitable precision.	-1	
	Report is longer than 15 pages	-5 per page	
	Figures and tables are appropriately formatted and contain relevant information. Figures and tables are described in the text (with referencing) and are appropriately captioned.	5	
	Figures and/or tables are of a low quality e.g., are poorly formatted screenshots.	-1 per figure/table	
	Figures and/or tables are of a low resolution and labels are difficult to read.	-1 per figure/table	
	Figures are missing labels or legends.	-1 per figure/table	
	Figures appear in the report which are not referenced in the text.	-1 per figure/table	
	Figures are convoluted and do not convey meaning.	-1 per figure/table	
References	Sources are appropriately referenced, both in text and in the references list.	1	1
Statement of contribution	A statement of contribution has been provided and the contribution each team member made to the project is clear.	1	2
	The statement of contribution clearly describes how generative AI (e.g. chatGPT) was used in the project.	1	
Meeting minutes	Group expectations form was completed and submitted by the due date.	1	4
	Appendix contains complete and detailed meeting minutes of at least 3 meetings.	3	
Jupyter notebook	Notebook is runnable without error.	1	5
	Notebook is well formatted.	1	
	Code is clear and consistent with best practices. For example, includes descriptions and appropriate code comments.	2	
	Results of the notebook are consistent with the written report, i.e., all figures and results are reproducible in the notebook.	1	
	Notebook contains cells with excessively long outputs.	-1 per cell	
	Notebook uses packages that are not allowed in the scope of the assignment. The list of allowed packages is provided on Canvas.	-2	
Kaggle competition	As described in section 5.1 of the assignment description.	10	10
Late submissions	Late submissions are subject to a penalty of 5% for each calendar day after the due date. After ten calendar days late, a mark of zero will be awarded.	-5 per day	
Failure to contribute to group work	This will be judged on a case-by-case basis, taking into account meeting minutes, the statement of contribution, and other evidence.	Up to -100 for no contribution.	