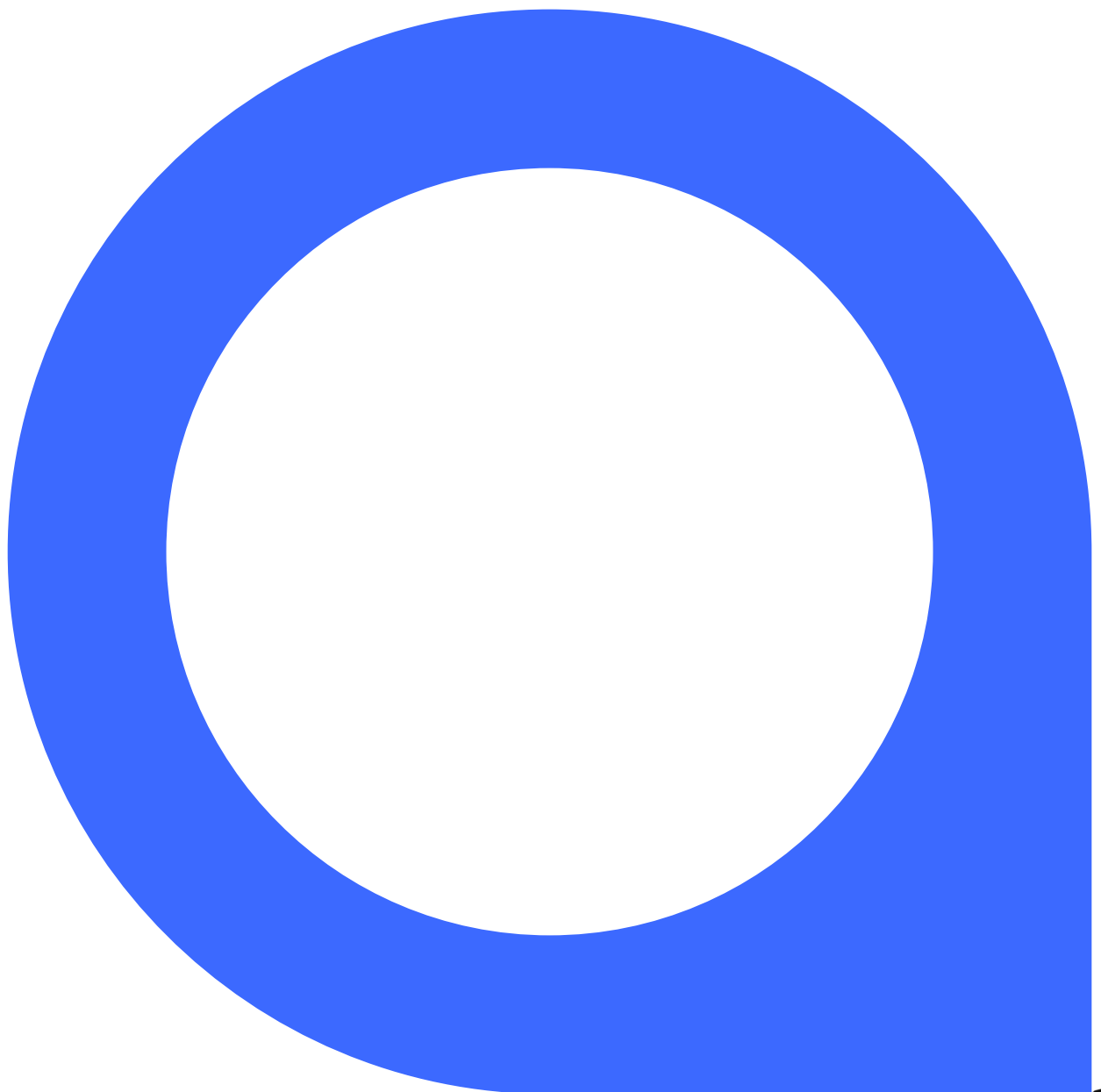


Data Science Applications

Assignment Semester 2 2024 – Marking Rubric





Marking Rubric

The following rubric outlines the specific standards required for each question in the assignment.

Criteria in bold at the top of each rubric box provide a holistic view on what differentiates answers under each rubric grade. Other criteria in each rubric box provide more specific guidance about components of an answer that are usually required to meet each holistic criterion.

Criteria designated as ‘hurdle’ in the left most column are those that must be met to gain a certain grade (usually either a grade of 2 or 3).

Criteria designated as ‘differentiator’ in the left most column are those that help to distinguish between assignment responses at each rubric grade.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1a	5%	Demonstrates a very strong understanding of the dataset.	Demonstrates a good understanding of the dataset.	Demonstrates some understanding of the dataset.	Demonstrates a limited understanding of the dataset.	Does not demonstrate an understanding of the dataset.
Hurdle (for a 4)		Splits the data into training, validation, and test sets at an appropriate stage in the data exploration to avoid leakage in the model to be built in Question 3.		Does not split the data into training, validation, and test sets, or does not split the data at an appropriate stage in the data exploration to avoid leakage in the model to be built in Question 3.		
Hurdle (for a 3)		Accurately applies and checks at least five different types of appropriate cleaning steps. Cleaning the 'Description' column does not count towards the target of five cleaning steps.			Applies some cleaning steps, although it is not clear whether these steps are appropriate for the problem context or the output of the cleaning is not checked or there are errors in the steps undertaken.	
Differentiator		Examines the dataset by applying appropriate data exploration techniques and visualisations that are tailored to each variable.	Examines the dataset by applying appropriate data exploration techniques and visualisations.	Summarises the dataset by applying appropriate data exploration techniques.	Summarises the dataset but some of the data exploration techniques are not appropriate.	Summarises the dataset but most or all the data exploration techniques applied are not appropriate.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1b	5%	Demonstrates a very strong ability to apply text vectorisation.	Demonstrates a good ability to apply text vectorisation.	Demonstrates some ability to apply text vectorisation.	Demonstrates a limited ability to apply text vectorisation.	Does not demonstrate an ability to apply text vectorisation.
Hurdle (for a 2)		Accurately cleans and vectorises the 'Description' feature in the training, validation, and test datasets.				Vectorises the 'Description' feature with some errors in the application of the vectorisation method.
Differentiator		Justifies the chosen vectorisation method in relation to preparing the data for the analysis that will follow.	Describes the vectorisation method used with strong reference to preparing the data for the analysis that will follow.	Describes the vectorisation method used with some reference to preparing the data for the analysis that will follow.	Describes the vectorisation method used but does not link this description to the analysis that will follow.	Outlines the vectorisation method used but does not link this outline to the analysis that will follow.
Differentiator		Correctly interprets the outcomes of at least two types of appropriate checks on the vectorisation completed, in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the vectorisation completed, in a thorough and insightful way.	Correctly interprets the outcomes of at least one appropriate check on the vectorisation completed.	Applies at least one appropriate check on the vectorisation completed but does not interpret its outcome or does not interpret its outcome correctly.	Applies at least one check on the vectorisation completed but the check is not appropriate, and the outcome of the check is not interpreted or not interpreted correctly.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1c	5%	Demonstrates a very strong ability to communicate key findings to relevant stakeholders.	Demonstrates a good ability to communicate key findings to relevant stakeholders.	Demonstrates some ability to communicate key findings to relevant stakeholders.	Demonstrates a limited ability to communicate key findings to relevant stakeholders.	Does not demonstrate an ability to communicate key findings to relevant stakeholders.
Differentiator		Summarises key characteristics of the dataset, with a strong link to the problem context.	Summarises key characteristics of the dataset, with some link to the problem context.	Outlines key characteristics of the dataset, with some link to the problem context.	Outlines key characteristics of the dataset, but there is no link to the problem context.	Does not outline key characteristics of the dataset.
Differentiator		Communicates the summary in a way that is highly suitable for the Membership team at the Institute.	Communicates the summary in a way that is suitable for the Membership team at the Institute.	Communicates the outline in a way that is mostly suitable for the Membership team at the Institute.	Communicates the outline in a way that is mostly not suitable for the Membership team at the Institute.	Communicates the outline in a way that is not suitable for the Membership team at the Institute.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2a	5%	Demonstrates a very strong ability to exercise critical thinking to plan a model before building it.	Demonstrates a good ability to exercise critical thinking to plan a model before building it.	Demonstrates some ability to exercise critical thinking to plan a model before building it.	Demonstrates a limited ability to exercise critical thinking to plan a model before building it.	Does not demonstrate an ability to exercise critical thinking to plan a model before building it.
Differentiator		Suggests a sensible set of features in relation to the business problem to be solved with clustering.	Describes a sensible set of features with strong reference to the business problem to be solved with clustering.	Describes a sensible set of features with some reference to the business problem to be solved with clustering.	Describes a sensible set of features but does not link this description to the business problem to be solved with clustering.	Outlines a set of features but these do not seem sensible given the business problem to be solved with clustering.



Question	Weight	Significantly above pass level 5	Above pass level 4	Pass level 3	Below pass level 2	Significantly below pass level 1
2b	5%	Demonstrates a very strong ability to apply a clustering algorithm.	Demonstrates a good ability to apply a clustering algorithm.	Demonstrates some ability to apply a clustering algorithm.	Demonstrates a limited ability to apply a clustering algorithm.	Does not demonstrate an ability to apply a clustering algorithm.
Hurdle (for a 2)		Accurately applies a clustering algorithm to the dataset.				Applies a clustering algorithm to the dataset with errors in the way the algorithm has been applied.
Hurdle (for a 3)		Examines the clustering outputs using internal validation, including a correct interpretation of the validation output.			Does not examine the clustering outputs or does not interpret the validation output or does not interpret the validation output correctly.	
Differentiator		Justifies the type of clustering algorithm used and the selection of hyperparameters (as relevant) such as the distance method, linkage method, and/or number of clusters.	Discusses the type of clustering algorithm used and the selection of hyperparameters (as relevant) such as the distance method, linkage method, and/or number of clusters, and the selections seem reasonable.	States the type of clustering algorithm used and the selection of hyperparameters (as relevant) such as the distance method, linkage method, and/or number of clusters, and the selections seem reasonable.	States the type of clustering algorithm used and/or the selection of hyperparameters (as relevant) such as the distance method, linkage method, and/or number of clusters, but the selections do not seem reasonable.	Does not select hyperparameters such as the distance method, linkage method, or number of clusters.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2c	10%	Demonstrates a very strong ability to examine a clustering algorithm's outputs.	Demonstrates a good ability to examine a clustering algorithm's outputs.	Demonstrates some ability to examine a clustering algorithm's outputs.	Demonstrates a limited ability to examine a clustering algorithm's outputs.	Does not demonstrate an ability to examine a clustering algorithm's outputs.
Hurdle (for a 2)		Examines the clustering output using manual validation, including a correct interpretation of the output of the manual validation.				Does not examine the clustering output, does interpret the validation output, or does not interpret the validation output correctly.
Differentiator		Describes the key characteristics of each cluster with a strong link to the problem context.	Describes the key characteristics of each cluster with some link to the problem context.	Outlines the key characteristics of each cluster with some link to the problem context.	Outlines some characteristics of the clusters but these are not key characteristics or there is no link to the problem context.	Outlines some characteristics of the Institute's CPD assets but these are not based on the output of the clustering algorithm.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2d	5%	Demonstrates a very strong ability to apply clustering outcomes to a business problem.	Demonstrates a good ability to apply clustering outcomes to a business problem.	Demonstrates some ability to apply clustering outcomes to a business problem.	Demonstrates a limited ability to apply clustering outcomes to a business problem.	Does not demonstrate an ability to apply clustering outcomes to a business problem.
Differentiator		Suggests a sensible, simpler way to group the CPD Hub's assets, based on the examination of the clustering outputs.	Describes a sensible, simpler way to group the CPD Hub's assets, based on the examination of the clustering outputs.	Outlines a sensible, simpler way to group the CPD Hub's assets, based on the examination of the clustering outputs.	Outlines a sensible, simpler way to group the CPD Hub's assets, but these are hypothetical and not related to the examination of the clustering outputs.	Outlines a way to group the CPD Hub's assets, but this is not a sensible and simpler way to group the assets.
Differentiator		Communicates in a way that is highly suitable for the Membership team of the Institute.	Communicates in a way that is suitable for the Membership team of the Institute.	Communicates in a way that is mostly suitable for the Membership team of the Institute.	Communicates in a way that is mostly not suitable for the Membership team of the Institute.	Communicates in a way that is not suitable for the Membership team of the Institute.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a	10%	Demonstrates a very strong ability to construct a response variable that is suitable for the given problem.	Demonstrates a good ability to construct a response variable that is suitable for the given problem.	Demonstrates some ability to construct a response variable that is suitable for the given problem.	Demonstrates a limited ability to construct a response variable that is suitable for the given problem.	Does not demonstrate an ability to construct a response variable that is suitable for the given problem.
Hurdle (for a 3)		Constructs a suitable response variable to use in the model.			Constructs a response variable but it is not suitable for use in the model.	
Differentiator		Justifies the response variable constructed with strong reference to the problem context.	Justifies the response variable constructed with some reference to the problem context and previous data exploration conducted.	Describes the response variable constructed with reference to the problem context and previous data exploration conducted.	Describes the response variable constructed but does not reference the problem context or previous data exploration conducted.	Does not describe the response variable constructed.
Differentiator		Correctly interprets the outcomes of at least two types of appropriate checks on the response variable constructed, in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the response variable constructed, in a thorough and insightful way.	Correctly interprets the outcomes of at least one appropriate check on the response variable constructed.	Applies at least one appropriate check on the response variable constructed but does not interpret its outcome or does not interpret its outcome correctly.	Applies at least one check on the response variable constructed but the check is not appropriate and the outcome of the check is not interpreted or is not interpreted correctly.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3b	5%	Demonstrates a very strong ability to exercise critical thinking to plan how a model will be evaluated.	Demonstrates a good ability to exercise critical thinking to plan how a model will be evaluated.	Demonstrates some ability to exercise critical thinking to plan how a model will be evaluated.	Demonstrates a limited ability to exercise critical thinking to plan how a model will be evaluated.	Does not demonstrate an ability to exercise critical thinking to plan how a model will be evaluated.
Differentiator		Suggests four evaluation metrics to use that are highly suitable for the business problem to be solved with classification.	Describes four suitable evaluation metrics, with strong reference to the business problem to be solved with classification.	Describes four suitable evaluation metrics, with some reference to the business problem to be solved with classification.	Describes four suitable evaluation metrics, but does not link this description to the business problem to be solved with classification.	Outlines four evaluation metrics but these do not seem sensible given the business problem to be solved with classification.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3c	20%	Demonstrates a very strong ability to construct a classifier.	Demonstrates a good ability to construct a classifier.	Demonstrates some ability to construct a classifier.	Demonstrates a limited ability to construct a classifier.	Does not demonstrate an ability to construct a classifier.
Hurdle (for a 2)		Correctly constructs a neural network to classify assets as having 'low', 'medium', or 'high' popularity.				Constructs a neural network but does not classify assets as having 'low', 'medium', or 'high' popularity.
Hurdle (for a 2)		Experiments with different model architectures, features, and hyperparameters.				Does not experiment with different model architectures, features, or hyperparameters.
Hurdle (for a 4)		Correctly interprets appropriate check(s) on the final classifier's outcomes and demonstrates an understanding of the model's behaviours and drivers		Applies at least one check on the final classifier's outcomes but does not demonstrate an understanding of the model's behaviours or drivers.		
Differentiator		Improves predictions between the initial and final iterations, using a range of appropriate metrics.	Attempts to improve predictions between the initial and final iterations, using a range of appropriate metrics.	Takes an iterative approach, but more appropriate metrics could have been used.	Does not take an iterative approach but takes steps to prevent the classifier from over or under fitting to the training data.	Does not take steps to prevent the classifier from over or under fitting to the training data.
Differentiator		The code is structured and very easy to read.	The code is structured and easy to read.	The code is structured and mostly easy to read.	The code lacks structure and is mostly difficult to read.	The code lacks structure and is difficult to read.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3d	5%	Demonstrates a very strong ability to interpret relevant measures of success for a classifier.	Demonstrates a good ability to interpret relevant measures of success for a classifier.	Demonstrates some ability to interpret relevant measures of success for a classifier.	Demonstrates a limited ability to interpret relevant measures of success for a classifier.	Does not demonstrate an ability to interpret relevant measures of success for a classifier.
Hurdle (for a 1)		Calculates how good the final classification model's predictions are using the evaluation metrics suggested in Question 3b.				
Hurdle (for a 3)		Compares the final classification model's outcomes to those under a suitable (simple) benchmark model.			Does not compare the final classification model's outcomes to those under a suitable (simple) benchmark model.	
Differentiator		Correctly interprets the evaluation metrics calculated, with strong reference to the business problem to be solved and with language that is highly suitable for the Membership team at the Institute.	Correctly interprets the evaluation metrics calculated, with strong reference to the business problem to be solved and with language that is mostly suitable for the Membership team at the Institute.	Correctly interprets the evaluation metrics calculated, with some reference to the business problem to be solved and with language that is mostly suitable for the Membership team at the Institute.	Correctly interprets the evaluation metrics calculated, but with no reference to the business problem to be solved or using language that is mostly not suitable for the Membership team at the Institute.	Does not correctly interpret the evaluation metrics calculated. Does not reference the business problem to be solved. Does not use language suitable for the Membership team at the Institute.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4a	5%	Demonstrates a very strong ability to critique the output of a GenAI tool.	Demonstrates a good ability to critique the output of a GenAI tool.	Demonstrates some ability to critique the output of a GenAI tool.	Demonstrates a limited ability to critique the output of a GenAI tool.	Does not demonstrate an ability to critique the output of a GenAI tool.
Differentiator		Discusses at least four material issues with the document summary produced by GenAI and weighs these against useful aspects of the summary. The discussion is directly relevant to the problem context.	Discusses three material issues with the document summary produced by GenAI and weighs these against the useful aspects of the summary. The discussion is directly relevant to the problem context.	Outlines three material issues with the document summary produced by GenAI and weighs these against the useful aspects of the summary. The outline is directly relevant to the problem context.	Outlines three material issues with the document summary produced by GenAI but does not weigh these against the useful aspects of the summary or the outline is not directly relevant to the problem context.	Outlines at least three issues with the document summary produced by GenAI but does not weigh these against the useful aspects of the summary and the outline is not relevant to the problem context.
Differentiator		Communicates clearly and concisely in a way that is highly suitable for the Membership team at the Institute.	Communicates clearly and concisely in a way that is suitable for the Membership team at the Institute.	Communicates clearly and concisely in a way that is mostly suitable for the Membership team at the Institute.	Communicates in a way that is mostly not suitable for the Membership team at the Institute.	Communicates in a way that is not suitable for the Membership team at the Institute.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4b	5%	Demonstrates a very strong ability to use GenAI to create a suitable document summary.	Demonstrates a good ability to use GenAI to create a suitable document summary.	Demonstrates some ability to use GenAI to create a suitable document summary.	Demonstrates a limited ability to use GenAI to create a suitable document summary.	Does not demonstrate an ability to use GenAI to create a suitable document summary.
Differentiator		Explains a suitable process followed to produce a more engaging summary of the chosen asset, using GenAI, which addresses the issues uncovered in Question 4a.	Describes a suitable process followed to produce a more engaging summary/ of the chosen asset, using GenAI, which addresses the issues uncovered in Question 4a.	Describes a suitable process followed to produce a more engaging summary of the chosen asset, using GenAI, which addresses most of the issues uncovered in Question 4a.	Outlines a suitable process followed to produce a more engaging summary of the chosen asset, using GenAI, which addresses some of the issues uncovered in Question 4a.	Outlines a process followed to produce a summary of the chosen asset, using GenAI, but the process followed does not seem suitable.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5	10%	Demonstrates a very strong ability to communicate model outcomes to a non-technical audience.	Demonstrates a good ability to communicate model outcomes to a non-technical audience.	Demonstrates some ability to communicate model outcomes to a non-technical audience.	Demonstrates a limited ability to communicate model outcomes to a non-technical audience.	Does not demonstrate an ability to communicate model outcomes to a non-technical audience.
Hurdle (for a 1)		Prepares a presentation to summarise findings in a video format.				
Hurdle (for a 3)		Demonstrates an understanding of the findings from Questions 2, 3, and 4 in the assignment.			Does not demonstrate an understanding of the findings from Questions 2, 3, and 4 in the assignment.	
Differentiator		The video has a clear start, middle, and end. There are clear transitions between all sections.	The video has a clear start, middle, and end. There are clear transitions between most sections.	The video has some structure. There are transitions between some sections.	The video has some structure. There are no transitions between sections.	The video lacks structure.
Differentiator		Communicates clearly and concisely in a way that is highly suitable for the Membership team at the Institute.	Communicates clearly and concisely in a way that is suitable for the Membership team at the Institute.	Communicates clearly and concisely in a way that is mostly suitable for the Membership team at the Institute.	Communicates in a way that is mostly not suitable for the Membership team at the Institute.	Communicates in a way that is not suitable for the Membership team at the Institute.
Total	100%					

Students are advised that a mark of zero (0) will be allocated for any question (or sub-question) where there is no attempt made or the marker finds the attempt is completely unsatisfactory.