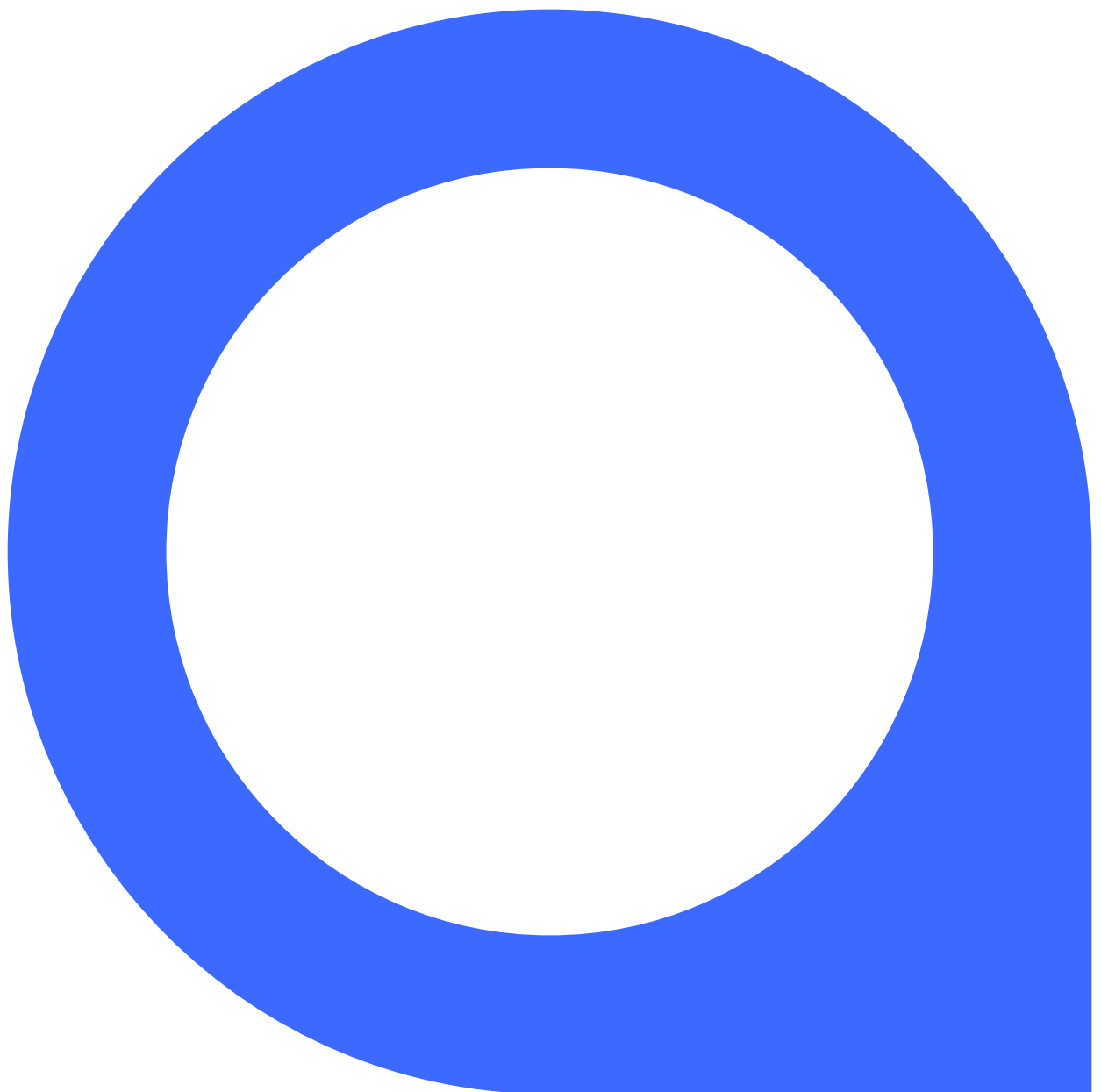


Data Science Applications

Assignment Semester 1 2025 – 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 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 'Transcript' column, using word embeddings and TF-IDF.				Vectorises the 'Transcript' column with some errors in the application of the vectorisation methods or only uses one of the two methods.
Differentiator		Accurately applies and checks at least five different types of appropriate text cleaning steps.	Accurately applies and checks at least four different types of appropriate text cleaning steps.	Accurately applies and checks at least three different types of appropriate text cleaning steps.	Accurately applies and checks at least two different types of appropriate text cleaning steps.	Accurately applies and checks no more than 1 different type of text cleaning steps, no accurate text cleaning applied.
Differentiator		Correctly interprets the outcomes of at least two types of appropriate checks on the vectorisations completed, in a thorough and insightful way.	Correctly interprets the outcomes of one appropriate check on the vectorisations completed, in a thorough and insightful way.	Correctly interprets the outcomes of at least one appropriate check on the vectorisations completed.	Applies at least one appropriate check on the vectorisations completed but does not interpret its outcome or does not interpret its outcome correctly.	Applies at least one check on the vectorisations 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
1b	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 internal 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 all required hyperparameters (distance method, linkage method, and number of clusters) with comparison to alternative choices, and with reference to the business context.	Discusses the type of clustering algorithm used and the selection of all required hyperparameters (distance method, linkage method, and number of clusters), and the selections seem reasonable.	States the type of clustering algorithm used and the selection of some required hyperparameters (distance method, linkage method, and number of clusters), and the selections seem reasonable.	States the type of clustering algorithm used and the selection of some required hyperparameters (distance method, linkage method, and 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
1c	5%	Demonstrates a very strong ability to build and interpret an adversarial model.	Demonstrates a good ability to build and interpret an adversarial model.	Demonstrates some ability to build and interpret an adversarial model.	Demonstrates a limited ability to build and interpret an adversarial model.	Does not demonstrate an ability to build and interpret an adversarial model.
Hurdle (for a 2)		Accurately applies a random forest algorithm to the cluster analysis and calculates variable importance.				Applies random forest algorithm to the cluster analysis and calculates variable importance, with errors how the algorithm has been applied.
Hurdle (for a 3)		Accurately calculates variable importance.			Calculates the variable importance, with errors in how the algorithm has been applied.	
Differentiator		Validates the top 10 keywords and their relationship to different clusters.	Discusses the top 10 keywords and their relationship to different clusters.	States the top 10 keywords.	States the calculated variable importances.	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
1d	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 transcripts, 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
1e	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 actionable insights from how the clusters relate to the three management issues, based on the examination of the clustering outputs.	Describes how the clusters relate to the three management issues, based on the examination of the clustering outputs.	Outlines how the clusters relate to the three management issues, based on the examination of the clustering outputs.	Outlines the cluster attributes, but these are not related to all three management issues.	Outlines the cluster attributes, but these are not related to any management issues.
Differentiator		Communicates in a way that is highly suitable for the management team of Bigtel.	Communicates in a way that is suitable for the management team of Bigtel.	Communicates in a way that is mostly suitable for the management team of Bigtel.	Communicates in a way that is mostly not suitable for management team of Bigtel.	Communicates in a way that is not suitable for the management team of Bigtel.



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 construct and apply an LLM prompt.	Demonstrates a good ability to construct and apply an LLM prompt.	Demonstrates some ability to construct and apply an LLM prompt.	Demonstrates a limited ability to construct and apply an LLM prompt.	Does not demonstrate an ability to construct and apply an LLM prompt.
Hurdle (for a 2)		Correctly constructs an LLM prompt for all three outputs, and the output matches specifications.				Does not examine the clustering output, does interpret the validation output, or does not interpret the validation output correctly.
Differentiator		Constructs and applies an LLM prompt, the output matches specifications, and the prompt contains all 5 components of a strong prompt.	Constructs and applies an LLM prompt, and the output matches specifications, but the prompt contains only 4 components of a strong prompt.	Constructs and applies an LLM prompt, and the output matches specifications, but the prompt contains only 3 components of a strong prompt.	Constructs and applies an LLM prompt, and the output matches specifications, but the prompt contains only 1 or 2 components of a strong prompt.	Constructs an LLM prompt, but does not apply it, or the output does not match the specifications.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2b	5%	Demonstrates a very strong ability to apply generative AI to a business problem.	Demonstrates a good ability to apply generative AI to a business problem.	Demonstrates some ability to apply generative AI to a business problem.	Demonstrates a limited ability to apply generative AI to a business problem.	Does not demonstrate an ability to apply generative AI to a business problem.
Differentiator		Suggests actionable insights from how the LLM outputs relate to the three management issues.	Describes how the LLM outputs relate to the three management issues.	Outlines how the LLM outputs relate to the three management issues.	Outlines the LLM outputs, but these are not related to all three management issues.	Outlines the LLM outputs, but these are not related to any management issues.
Differentiator		Communicates in a way that is highly suitable for the management team of Bigtel.	Communicates in a way that is suitable for the management team of Bigtel.	Communicates in a way that is mostly suitable for the management team of Bigtel.	Communicates in a way that is mostly not suitable for management team of Bigtel.	Communicates in a way that is not suitable for the management team of Bigtel.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a	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 'Transcript' 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
3b	5%	Demonstrates a very strong ability to propose a unit of analysis that is suitable for the given problem.	Demonstrates a good ability to construct a unit of analysis that is suitable for the given problem.	Demonstrates some ability to construct a unit of analysis that is suitable for the given problem.	Demonstrates a limited ability to construct a unit of analysis that is suitable for the given problem.	Does not demonstrate an ability to construct a unit of analysis that is suitable for the given problem.
Hurdle (for a 3)		Proposes a suitable unit of analysis to use in this business problem.			Proposes a unit of analysis but it is not suitable for use in this business problem.	
Hurdle (for a 3)		States the entity and the basis of the timestamps.			Does not state the entity and/or does not state the basis of the timestamps.	
Differentiator		Proposes a suitable unit of analysis with strong reference to the problem context.	Proposes a suitable unit of analysis with some reference to the problem context.	Describes a suitable unit of analysis with reference to the problem context.	Describes a suitable unit of analysis but does not reference the problem context.	Does not describe a suitable unit of analysis.
Differentiator		Justifies a suitable selection of timestamps (regular/event/both) with strong reference to the problem context.	Justifies a suitable selection of timestamps (regular/event/both) with some reference to the problem context.	Describes a suitable selection of timestamps (regular/event/both) with some reference to the problem context.	Describes a suitable selection of timestamps (regular/event/both) but does not reference the problem context.	Does not describe the selection of timestamps.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3c	5%	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
3d	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
3e	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 customer churn as a binary outcome.				Constructs a neural network but does not classify customer churn as a binary outcome.
Hurdle (for a 2)		Experiments with different model architectures, features, regularisation, and hyperparameters.				Does not experiment with different model architectures, features, regularisation 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
3f	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 3d.				
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 management team at Bigtel.	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 management team at Bigtel.	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 management team at Bigtel.	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 management team at Bigtel.	Does not correctly interpret the evaluation metrics calculated. Does not reference the business problem to be solved. Does not use language suitable for the management team at Bigtel.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3g	5%	Demonstrates a very strong ability to interpret the behaviours of a classifier.	Demonstrates a good ability to interpret the behaviours of a classifier.	Demonstrates some ability to interpret the behaviours of a classifier.	Demonstrates a limited ability to interpret the behaviours of a classifier.	Does not demonstrate an ability to interpret the behaviours of a classifier.
Hurdle (for a 1)		Calculates feature importance.				
Hurdle (for a 3)		Calculates partial dependence and SHAP feature explanations.			Does not calculate partial dependence and SHAP feature explanations.	
Differentiator		Correctly interprets the model explanation outputs, prioritising which details to calculate and share, with strong reference to the business problem to be solved, and with language highly suitable for the management team at Bigtel.	Correctly interprets the model explanation outputs, prioritising which details to calculate and share, with strong reference to the business problem to be solved, and with language suitable for the management team at Bigtel.	Correctly interprets the model explanation outputs, prioritising which details to calculate and share, with some reference to the business problem to be solved, and with language mostly suitable for the management team at Bigtel.	Correctly interprets the model explanation outputs, prioritising which details to calculate and share, with no reference to the business problem to be solved, or with language not suitable for the management team at Bigtel.	Does not correctly model explanation outputs. Does not reference the business problem to be solved. Does not use language suitable for the management team at Bigtel.



Question	Weight	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4	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 Management team at Bigtel.	Communicates clearly and concisely in a way that is suitable for the Management team at Bigtel.	Communicates clearly and concisely in a way that is mostly suitable for the Management team at Bigtel.	Communicates in a way that is mostly not suitable for the Management team at Bigtel.	Communicates in a way that is not suitable for the Management team at Bigtel.
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.