



DATA ANALYTICS APPLICATIONS

ASSIGNMENT MARKING RUBRIC SEMESTER 1 2022



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 information about components of an answer that are usually required to meet each holistic criterion.

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
1 Domain knowledge	15 marks	Demonstrates a very strong ability to acquire domain knowledge about an industry. Explains at least five key characteristics of the book publishing industry. Explains how these characteristics are relevant to this analysis. Lists the source(s) used for each of the characteristics explained.	Demonstrates a good ability to acquire domain knowledge about an industry. Explains three or four key characteristics of the book publishing industry. Explains how these characteristics are relevant to this analysis. Lists the source(s) used for each of the characteristics explained.	Demonstrates some ability to acquire domain knowledge about an industry. Outlines at least three key characteristics of the book publishing industry. Outlines how these characteristics are relevant to this analysis. Lists the source(s) used for most of the characteristics identified.	Demonstrates a limited ability to acquire domain knowledge about an industry. Outlines at least three characteristics of the book publishing industry. Does not outline how these characteristics are relevant to this analysis or it is not clear that these characteristics are relevant. Lists some source(s) of information.	Does not demonstrate an ability to acquire domain knowledge about an industry. Does not outline at least three relevant characteristics of the book publishing industry. Does not list any of the source(s) of the information acquired.



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
2 Exploratory data analysis	10 marks	<p>Demonstrates a very strong understanding of the features in the dataset.</p> <p>Examines at least six features in the dataset that are relevant in answering Questions 3 and 5.</p> <p>Describes each of these features in relation to their usefulness for the analysis to be undertaken.</p> <p>Identifies a range of issues with the data and explains how these will be managed in the analysis.</p>	<p>Demonstrates a good understanding of the features in the dataset.</p> <p>Examines at least six features in the dataset that are relevant in answering Questions 3 and 5.</p> <p>Describes each of these features, with some link to the problem context.</p> <p>Identifies one or two issues with the data and explains how these will be managed in the analysis.</p>	<p>Demonstrates some understanding of the features in the dataset.</p> <p>Summarises some of the features in the dataset that are relevant in answering Questions 3 and 5.</p> <p>Outlines most of these features, with some link to the problem context.</p> <p>Identifies one or two issues with the data and states how these will be managed in the analysis.</p>	<p>Demonstrates a limited understanding of the features in the dataset.</p> <p>Summarises some of the features in the dataset that are relevant in answering Questions 3 and 5.</p> <p>Outlines these features but with no link to the problem context.</p> <p>Identifies some issues with the data but does not state how these will be managed in the analysis.</p>	<p>Does not demonstrate an understanding of the features in the dataset.</p> <p>Summarises some of the features in the dataset that are relevant in answering Questions 3 and 5.</p> <p>Does not comment on these features.</p> <p>Does not identify any issues with the data.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3 Vectorising features	13 marks	<p>Demonstrates a very strong ability to apply natural language processing.</p> <p>Successfully tokenises the title and description features.</p> <p>Accurately applies at least four data cleaning techniques and explains why these are suitable for analysis in Question 4.</p> <p>Accurately applies stemming or lemmatisation.</p> <p>Accurately calculates vectorised features that represent the title and description features.</p> <p>Explains each of these steps and why they have been undertaken.</p> <p>The code is well structured and easy to read.</p> <p>Demonstrates that a range of data checks have been performed on each step undertaken.</p>	<p>Demonstrates a good ability to apply natural language processing.</p> <p>Successfully tokenises the title and description features.</p> <p>Accurately applies three reasonable data cleaning techniques and explains why these are suitable for analysis in Question 4.</p> <p>Accurately applies stemming or lemmatisation.</p> <p>Accurately calculates vectorised features that represent the title and description features.</p> <p>Explains each of these steps and why they have been undertaken.</p> <p>The code is well structured and easy to read.</p> <p>Demonstrates that data checks have been performed on each step undertaken.</p>	<p>Demonstrates some ability to apply natural language processing.</p> <p>Successfully tokenises the title and description features.</p> <p>Accurately applies one reasonable data cleaning technique and explains why this is suitable for analysis in Question 4.</p> <p>Accurately applies stemming or lemmatisation.</p> <p>Accurately calculates vectorised features that represent the title and description features.</p> <p>Outlines each of these steps.</p> <p>The code has structure and is mostly easy to read.</p> <p>Demonstrates that data checks have been performed on most steps undertaken.</p>	<p>Demonstrates a limited ability to apply natural language processing.</p> <p>Tokenises the title and description features.</p> <p>Applies at least one data cleaning technique.</p> <p>Applies stemming or lemmatisation.</p> <p>Calculates vectorised features.</p> <p>There are some errors in the way the above steps are undertaken.</p> <p>Does not outline each of these steps.</p> <p>The code lacks structure or is difficult to read.</p> <p>Does not demonstrate that any data checks have been performed.</p>	<p>Does not demonstrate an ability to apply natural language processing.</p> <p>Does not complete each step of tokenising, cleaning, stemming/lemmatisation, and vectorisation.</p> <p>or</p> <p>Demonstrates no or limited understanding of these NLP steps.</p> <p>The code lacks structure and is difficult to read.</p> <p>Does not demonstrate that any data checks have been performed.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
4 Clustering	7 marks	<p>Demonstrates a very strong ability to apply and evaluate clustering algorithms.</p> <p>Accurately applies several clustering algorithms to group the books into a small number of distinct clusters based on the title of each book.</p> <p>Explains this step and why it has been undertaken.</p> <p>Justifies the selection for the number of clusters.</p> <p>Examines the patterns detected in book titles using internal, external, and manual validation.</p> <p>Considers the problem context when performing these evaluations.</p> <p>The code is well structured and easy to read.</p>	<p>Demonstrates a good ability to apply and evaluate clustering algorithms.</p> <p>Accurately applies several clustering algorithms to group the books into a small number of distinct clusters based on the title of each book.</p> <p>Discusses this step.</p> <p>Discusses the selection for the number of clusters and the selection seems reasonable.</p> <p>Examines the patterns detected in book titles using two of internal, external, and manual validation.</p> <p>Considers the problem context when performing these evaluations.</p> <p>The code is well structured and easy to read.</p>	<p>Demonstrates some ability to apply and evaluate clustering algorithms.</p> <p>Accurately applies a clustering algorithm to group the books into a small number of distinct clusters based on the title of each book.</p> <p>States the selection for the number of clusters.</p> <p>Applies two of internal, external, and manual validation to evaluate the clusters.</p> <p>Outlines the patterns detected in book titles.</p> <p>Considers the problem context when performing these evaluations.</p> <p>The code has structure and is mostly easy to read.</p>	<p>Demonstrates a limited ability to apply and evaluate clustering algorithms.</p> <p>Applies a clustering algorithm with a few errors in the way the algorithm has been applied.</p> <p>Does not make a selection for the number of clusters.</p> <p>Does not state how the number of clusters was selected.</p> <p>Applies one of internal, external, or manual validation to evaluate the clusters.</p> <p>No consideration of the problem context was incorporated into the evaluation.</p> <p>The code lacks structure or is difficult to read.</p>	<p>Does not demonstrate an ability to apply and evaluate clustering algorithms.</p> <p>Does not apply a clustering algorithm to group the books into a small number of distinct clusters.</p> <p>Demonstrates no or limited understanding of clustering.</p> <p>Does not apply internal, external, or manual validation to evaluate the clusters.</p> <p>The code lacks structure and is difficult to read.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5a Response variable	10 marks	<p>Demonstrates a very strong ability to choose and create a response variable that is suitable for the given problem.</p> <p>Calculates a suitable response variable to use in the model.</p> <p>Justifies the response variable selected.</p> <p>Demonstrates that a range of checks have been performed on the response variable calculated.</p>	<p>Demonstrates a good ability to choose and create a response variable that is suitable for the given problem.</p> <p>Calculates a suitable response variable to use in the model.</p> <p>Explains the response variable selected.</p> <p>Demonstrates that checks have been performed on the response variable calculated.</p>	<p>Demonstrates some ability to choose and create a response variable that is suitable for the given problem.</p> <p>Calculates a suitable response variable to use in the model.</p> <p>Describes the response variable selected.</p> <p>Demonstrates that some checks have been performed on the response variable calculated.</p>	<p>Demonstrates a limited ability to choose and create a response variable that is suitable for the given problem.</p> <p>Calculates a response variable to use in the model.</p> <p>Does not discuss what response variable has been used or it is not clear whether the response variable is suitable.</p> <p>Does not demonstrate that any checks have been performed on the response variable calculated.</p>	<p>Does not demonstrate an ability to choose and create a response variable that is suitable for the given problem.</p> <p>Does not calculate a suitable response variable to use in the model.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5b Classification model	15 marks	<p>Demonstrates a very strong ability to construct classification models.</p> <p>Correctly constructs several types of classifiers to predict book sales.</p> <p>Experiments with different types of classifiers, model architectures, and hyperparameter selections.</p> <p>Takes an iterative approach to build each classifier, where evaluation outcomes are used to improve predictions after each iteration.</p> <p>Takes steps to prevent the classifiers from overfitting to the training data.</p> <p>Justifies the final model, architecture, hyperparameters, and features selected.</p> <p>Demonstrates that a range of checks have been performed on the final classifier constructed.</p> <p>The code is well structured and easy to read.</p>	<p>Demonstrates a good ability to construct classification models.</p> <p>Correctly constructs several types of classifiers to predict book sales.</p> <p>Experiments with different types of classifiers, model architectures, or hyperparameter selections.</p> <p>Takes steps to prevent the classifiers from overfitting to the training data.</p> <p>Describes the final model architecture, hyperparameters, and features selected.</p> <p>Demonstrates that checks have been performed on the final classifier constructed.</p> <p>The code is well structured and easy to read.</p>	<p>Demonstrates some ability to construct classification models.</p> <p>Correctly constructs a classifier to predict book sales.</p> <p>States the type of classifier, model architecture, hyperparameter, and features selected.</p> <p>Demonstrates that some checks have been performed on the classifier constructed.</p> <p>The code has structure and is mostly easy to read.</p>	<p>Demonstrates a limited ability to construct classification models.</p> <p>Constructs a classifier with some errors in the code that negatively impacts the model's predictions.</p> <p>Does not state the type of classifier, model architecture, hyperparameter, or features selected.</p> <p>Does not demonstrate that any checks have been performed on the classifier constructed.</p> <p>The code lacks structure or is difficult to read.</p>	<p>Does not demonstrate an ability to construct classification models.</p> <p>Does not construct a classifier that predicts book sales.</p> <p>The code lacks structure and is difficult to read.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5c Model evaluation	10 marks	<p>Demonstrates a very strong ability to evaluate classification models.</p> <p>Evaluates how good the final classification model's predictions are using at least three appropriate measures of success.</p> <p>Justifies the measures used given the problem context.</p> <p>Compares outcomes under the classifier's built to those under a benchmark model.</p> <p>Demonstrates, via comments in the notebook, a very strong understanding of how predictive the model is.</p>	<p>Demonstrates a good ability to evaluate classification models.</p> <p>Evaluates how good the final classification model's predictions are using two appropriate measures of success.</p> <p>Explains why these measures have been used.</p> <p>Demonstrates, via comments in the notebook, a good understanding of how predictive the model is.</p>	<p>Demonstrates some ability to evaluate classification models.</p> <p>Evaluates how good the final classification model's predictions are using one appropriate measure of success.</p> <p>Demonstrates, via comments in the notebook, some understanding of how predictive the model is.</p>	<p>Demonstrates a limited ability to evaluate classification models.</p> <p>Calculates how good the final classification model's predictions are using one appropriate measure of success.</p> <p>Demonstrates a limited understanding of how predictive the model is.</p>	<p>Does not demonstrate an ability to evaluate classification models.</p> <p>Does not select an appropriate measure to evaluate how good the final classification model's predictions are or does not appear to understand how to interpret any of the calculated measures.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
6 Limitations	10 marks	Demonstrates a very strong understanding of the limitations of the analysis. Explains at least five key limitations of this analysis that are directly relevant to the problem context. Explains steps that could be taken to overcome each of these limitations.	Demonstrates a good understanding of the limitations of the analysis. Explains three or four key limitations of this analysis that are directly relevant to the problem context. Explains steps that could be taken to overcome each of these limitations.	Demonstrates some understanding of the limitations of the analysis. Outlines at least three key limitations of this analysis that are relevant to the problem context. Outlines steps that could be taken to overcome each of these limitations.	Demonstrates a limited understanding of the limitations of the analysis. Outlines at least three limitations of this analysis but either the limitations are not material or there is no discussion of the steps that could be taken to overcome each of these limitations.	Does not demonstrate an understanding of the limitations of the analysis. Outlines fewer than three limitations of the analysis.



Data Analytics Applications

Assignment Marking Rubric Semester 1 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
7 Executive summary	10 marks	<p>Demonstrates a very strong grasp of the concepts being assessed in this assignment.</p> <p>Demonstrates a very strong ability to communicate model findings to a non-technical audience.</p> <p>The executive summary is presented as a video summary.</p> <p>The summary has good structure and presents all points clearly and concisely.</p> <p>Conclusions drawn are supported by the summarised findings.</p>	<p>Demonstrates a good grasp of the concepts being assessed in this assignment.</p> <p>Demonstrates a good ability to communicate model findings to a non-technical audience.</p> <p>The executive summary is presented as a video summary.</p> <p>The summary has good structure and presents most points clearly and concisely.</p> <p>Conclusions drawn are supported by the summarised findings.</p>	<p>Demonstrates some grasp of the concepts being assessed in this assignment, but the language used could convey the information more clearly in some cases.</p> <p>The executive summary is presented as a video summary.</p> <p>The summary has structure and presents some points clearly and concisely.</p> <p>Conclusions drawn have some link to the summarised findings.</p>	<p>Demonstrates a limited grasp of the concepts being assessed in this assignment.</p> <p>The executive summary is not presented as a video summary or the summary lacks structure or does not present points clearly or concisely.</p> <p>Conclusions drawn are not supported by the summarised findings.</p>	<p>Does not demonstrate a grasp of the concepts being assessed in this assignment.</p> <p>The answer to this question is not an executive summary.</p>
Total	100 marks					