



DATA ANALYTICS APPLICATIONS

ASSIGNMENT MARKING RUBRIC

SEMESTER 2 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 guidance 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. Applies at least two strategies to acquire knowledge about the real estate industry. Explains at least five key drivers of property sales prices. Lists the source(s) used for each of the drivers explained. Explains how well the property dataset captures these key drivers. Uses language suitable for communication with your friend in real-estate.	Demonstrates a good ability to acquire domain knowledge about an industry. Applies at least one strategy to acquire knowledge about the real estate industry. Explains three or four key drivers of property sales. Lists the source(s) used for each of the drivers explained. Explains how well the property dataset captures these key drivers. Uses language suitable for communication with your friend in real-estate.	Demonstrates some ability to acquire domain knowledge about an industry. Applies at least one strategy to acquire domain knowledge about the real estate industry. Describes at least three key drivers of property sales. Lists the source(s) used for most of the drivers outlined. Outlines how well the property dataset captures these key drivers. Mostly uses language suitable for communication with your friend in real-estate.	Demonstrates a limited ability to acquire domain knowledge about an industry. Applies at least one strategy to acquire domain knowledge about the real estate industry. Describes at least three drivers of property sales. Lists at least one source of information. Does not outline how well the property dataset captures these key drivers. Does not use language suitable for communication with your friend in real-estate.	Does not demonstrate an ability to acquire domain knowledge about an industry. Applies at least one strategy to acquire domain knowledge about the real estate industry. Outlines some drivers of property sales. Does not list any of the source(s) of the information acquired. Does not outline how well the property dataset captures these key drivers. Does not use language suitable for communication with your friend in real-estate.



Data Analytics Applications

Assignment Marking Rubric Semester 2 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 Deployment decisions	10 marks	Demonstrates a very strong understanding of the deployment decisions that are relevant to this problem. Discusses at least five key deployment decisions that are directly relevant to the planning of the app roll-out. Uses language suitable for communication with your friend in real-estate.	Demonstrates a good understanding of the deployment decisions that are relevant to this problem. Discusses three or four key deployment decisions that are directly relevant to the planning of the app roll-out. Uses language suitable for communication with your friend in real-estate.	Demonstrates some understanding of the deployment decisions that are relevant to this problem. Outlines at least three key deployment decisions that are relevant to the planning of the app roll-out. Mostly uses language suitable for communication with your friend in real-estate.	Demonstrates a limited understanding of the deployment decisions that are relevant to this problem. Outlines at least three deployment decisions but either the decisions are not material or it is not clear that these are relevant to the planning of the app roll-out. Does not use language suitable for communication with your friend in real-estate.	Does not demonstrate an understanding of the deployment decisions that are relevant to this problem. Outlines one or two deployment decisions but these are not relevant to the planning of the app roll-out. Does not use language suitable for communication with your friend in real-estate.



Data Analytics Applications

Assignment Marking Rubric Semester 2 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
3a Exploratory data analysis	5 marks	Demonstrates a very strong understanding of the features in the dataset. Examines more than three features in the dataset. Describes key characteristics of these features in relation to their usefulness for the analysis to be undertaken, using language suitable for sharing with your friend. Identifies at least three issues with the data and explains how these will be managed in the analysis. The code is well structured and easy to read.	Demonstrates a good understanding of the features in the dataset. Examines three features in the dataset. Describes key characteristics of these features, with some link to the problem context, using language suitable for sharing with your friend. Identifies one or two issues with the data and explains how these will be managed in the analysis. The code is mostly well structured and easy to read.	Demonstrates some understanding of the features in the dataset. Summarises at least three features in the dataset. Outlines key characteristics of these features, with some link to the problem context, using language mostly suitable for sharing with your friend. Identifies one or two issues with the data and states how these will be managed in the analysis. The code has structure and is mostly easy to read.	Demonstrates a limited understanding of the features in the dataset. Summarises one or two features in the dataset. Outlines characteristics of these features but with no link to the problem context or outlines characteristics using language that is overly technical and not suitable for sharing with your friend. Identifies one or two issues with the data but does not state how these will be managed in the analysis. The code lacks structure or is difficult to read.	Does not demonstrate an understanding of the features in the dataset. Summarises one or two features in the dataset. Does not comment on these features. Does not identify any issues with the data. The code lacks structure and is difficult to read.
3b Data cleaning and pre-processing	5 marks	Demonstrates a very strong ability to clean data prior to using it in a model. Accurately applies at least four reasonable data cleaning or pre-processing techniques and explains why these have been performed. The code is well structured and easy to read.	Demonstrates a good ability to clean data prior to using it in a model. Accurately applies three reasonable data cleaning or pre-processing techniques and explains why these have been performed. The code is mostly well structured and easy to read.	Demonstrates some ability to clean data prior to using it in a model. Accurately applies at least three reasonable data cleaning or pre-processing techniques and outlines why these have been performed. The code has structure and is mostly easy to read.	Demonstrates a limited ability to clean data prior to using it in a model. Accurately applies at least three reasonable data cleaning or pre-processing techniques but does not outline why these have been performed. The code lacks structure or is difficult to read.	Does not demonstrate an ability to clean data prior to using it in a model. Applies some data cleaning techniques with errors in the application of the techniques. Does not outline why these data cleaning steps have been performed. The code lacks structure and is difficult to read.



Data Analytics Applications

Assignment Marking Rubric Semester 2 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 Vectorising features	10 marks	<p>Demonstrates a very strong ability to apply natural language processing.</p> <p>Accurately applies steps in the NLP process, as appropriate, such as tokenisation, data cleaning, stemming or lemmatisation, and vectorisation.</p> <p>Explains each NLP step and why it has (or has not) been undertaken in relation to preparing the data for use in modelling.</p> <p>The code is well structured and easy to read.</p> <p>Demonstrates that each step has been performed correctly by interpreting the outcomes of at least two types of checks.</p>	<p>Demonstrates a good ability to apply natural language processing.</p> <p>Accurately applies steps in the NLP process, as appropriate, such as tokenisation, data cleaning, stemming or lemmatisation, and vectorisation.</p> <p>Explains each NLP step and why it has (or has not) been undertaken in relation to preparing the data for use in modelling.</p> <p>The code is mostly well structured and easy to read.</p> <p>Demonstrates that reasonable checks have been performed on each step undertaken.</p>	<p>Demonstrates some ability to apply natural language processing.</p> <p>Accurately applies steps in the NLP process, as appropriate, such as tokenisation, data cleaning, stemming or lemmatisation, and vectorisation.</p> <p>Outlines each NLP step that has been undertaken in relation to preparing the data for use in modelling.</p> <p>The code has structure and is mostly easy to read.</p> <p>Demonstrates that reasonable checks have been performed on the output of most steps undertaken.</p>	<p>Demonstrates a limited ability to apply natural language processing.</p> <p>Accurately applies steps in the NLP process, as appropriate, such as tokenisation, data cleaning, stemming or lemmatisation, and vectorisation.</p> <p>Does not outline each of these steps.</p> <p>The code lacks structure or is difficult to read.</p> <p>Demonstrates that reasonable checks have been performed on some of the steps undertaken.</p>	<p>Does not demonstrate an ability to apply natural language processing.</p> <p>Applies steps in the NLP process such as tokenisation, data cleaning, stemming or lemmatisation, and vectorisation.</p> <p>There are some errors in the ways the above steps are undertaken.</p> <p>Does not outline each of these steps.</p> <p>Demonstrates no or limited understanding of these NLP steps.</p> <p>The code lacks structure and is difficult to read.</p> <p>Demonstrates only one or two cursory checks having been performed.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 2 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	5 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 the response variable has been calculated correctly by interpreting the outcomes of at least two types of checks.</p> <p>The code is well structured and easy to read.</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 the response variable has been calculated correctly by interpreting the outcome of one type of check.</p> <p>The code is mostly well structured and easy to read.</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 reasonable checks have been performed on the response variable calculated.</p> <p>The code has structure and is mostly easy to read.</p>	<p>Demonstrates a limited 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>Does not describe what response variable has been selected or it is not clear whether the response variable is suitable.</p> <p>Does not demonstrate that any reasonable checks have been performed on the response variable calculated.</p> <p>The code lacks structure or is difficult to read.</p>	<p>Does not demonstrate an ability to choose and create a response variable that is suitable for the given problem.</p> <p>Calculates a response variable but it is not suitable for use in the model.</p> <p>Does not state what response variable has been selected.</p> <p>Does not demonstrate that any reasonable checks have been performed on the response variable calculated.</p> <p>The code lacks structure and is difficult to read.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 2 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 Additional feature	10 marks	<p>Demonstrates a very strong ability to use an external data source to gather additional data to improve a model.</p> <p>Constructs a suitable additional feature that is not already available in the property dataset.</p> <p>Justifies the additional feature constructed based on the domain research conducted in Question 1.</p> <p>Demonstrates that the additional feature has been constructed correctly by interpreting the outcomes of at least two types of checks.</p> <p>The code is well structured and easy to read.</p>	<p>Demonstrates a good ability to use an external data source to gather additional data to improve a model.</p> <p>Constructs a suitable additional feature that is not already available in the property dataset.</p> <p>Explains the additional feature constructed with reference to the domain research conducted in Question 1.</p> <p>Demonstrates that the additional feature has been constructed correctly by interpreting the outcome of one type of check.</p> <p>The code is mostly well structured and easy to read.</p>	<p>Demonstrates some ability to use an external data source to gather additional data to improve a model.</p> <p>Constructs a suitable additional feature that is not already available in the property dataset.</p> <p>Describes the additional feature constructed.</p> <p>Demonstrates that some reasonable checks have been performed on the additional feature constructed.</p> <p>The code has structure and is mostly easy to read.</p>	<p>Demonstrates a limited ability to use an external data source to gather additional data to improve a model.</p> <p>Constructs an additional feature that is not already available in the property dataset.</p> <p>Does not describe what additional feature has been constructed or it is not clear whether the additional feature is suitable.</p> <p>Does not demonstrate that any reasonable checks have been performed on the additional feature calculated.</p> <p>The code lacks structure or is difficult to read.</p>	<p>Does not demonstrate an ability to use an external data source to gather additional data to improve a model.</p> <p>Constructs an additional feature but it is not suitable for use in the model or it does not rely on an external data source.</p> <p>Does not demonstrate that any reasonable checks have been performed on the additional feature calculated.</p> <p>The code lacks structure and is difficult to read.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 2 2022

5c Classification model	15 marks	<p>Demonstrates a very strong ability to construct classification models.</p> <p>Correctly constructs at least two different types of classifiers to predict the sales price band that a property will fall within.</p> <p>Experiments with different model architectures, features, 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 over or under fitting to the training data.</p> <p>Justifies the final model, architecture, hyperparameters, and features selected.</p> <p>Demonstrates that the final classifier has been constructed correctly by interpreting the outcomes of at least two types of checks.</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 at least two different types of classifiers to predict the sales price band that a property will fall within.</p> <p>Experiments with different model architectures, features, or hyperparameter selections.</p> <p>Takes steps to prevent the classifiers from over or under fitting to the training data.</p> <p>Describes the final model architecture, hyperparameters, and features selected.</p> <p>Demonstrates that the final classifier has been constructed correctly by interpreting the outcome of one type of check.</p> <p>The code is mostly well structured and easy to read.</p>	<p>Demonstrates some ability to construct classification models.</p> <p>Correctly constructs one type of classifier to predict the sales price band that a property will fall within.</p> <p>States the type of classifier, model architecture, hyperparameter, and features selected.</p> <p>Demonstrates that some reasonable 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 one classifier to predict the sales price band that a property will fall within.</p> <p>There are some errors in the code that negatively impact 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 reasonable 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>Constructs one classifier but it does not predict the sales price band that a property will fall within.</p> <p>The code lacks structure and is difficult to read.</p>
-------------------------------	-------------	--	---	--	---	---



Data Analytics Applications

Assignment Marking Rubric Semester 2 2022

Question	Marks	Significantly above pass level	Above pass level	Pass level	Below pass level	Significantly below pass level
		5	4	3	2	1
5d 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 classifiers built to those under a suitable benchmark model.</p> <p>Demonstrates, via comments in the notebook, a very strong understanding of how predictive the model is.</p> <p>The code is well structured and easy to read.</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 the measures that have been in relation to the problem context.</p> <p>Demonstrates, via comments in the notebook, a good understanding of how predictive the model is.</p> <p>The code is mostly well structured and easy to read.</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>There is some link to the problem context in the evaluation of the model.</p> <p>Demonstrates, via comments in the notebook, some understanding of how predictive the model is.</p> <p>The code has structure and is mostly easy to read.</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>The code lacks structure or is difficult to read.</p>	<p>Does not demonstrate an ability to evaluate classification models.</p> <p>Calculates how good the final classification model's predictions are but the measure of success chosen is not appropriate for this problem.</p> <p>or</p> <p>Does not appear to understand how to interpret any of the calculated measures.</p> <p>The code lacks structure and is difficult to read.</p>



Data Analytics Applications

Assignment Marking Rubric Semester 2 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 Video communication	5 marks	Demonstrates a very strong ability to communicate technical concepts to a non-technical audience. The video has good structure, with a clear beginning, middle, and end. All points are presented clearly and concisely using language suitable for communication with your friend in real-estate.	Demonstrates a good ability to communicate technical concepts to a non-technical audience. The video has good structure, with a beginning, middle, and end. Most points are presented clearly and concisely using language suitable for communication with your friend in real-estate.	Demonstrates some ability to communicate technical concepts to a non-technical audience. The video has some structure. Some points are presented clearly and concisely using language suitable for communication with your friend in real-estate.	Demonstrates a limited ability to communicate technical concepts to a non-technical audience. The video lacks structure. Does not present points clearly or concisely using language suitable for communication with your friend in real-estate.	Does not demonstrate an ability to communicate technical concepts to a non-technical audience. The answer is not presented as a video.
6 Video content	5 marks	Demonstrates a very strong understanding of the model and its limitations. Explains the business idea, performance of the model, and recommendation. The performance of the model and recommendation reference the problem context. Explains at least three material limitations of the analysis and steps to overcome them. These are directly relevant to the problem context.	Demonstrates a good understanding of the model and its limitations. Describes the business idea, performance of the model, and recommendation. The performance of the model and recommendation reference the problem context. Explains two material limitations of the analysis and steps to overcome them. These are directly relevant to the problem context.	Demonstrates some understanding of the model and its limitations. Outlines the business idea, performance of the model, and recommendation. The performance of the model and recommendation reference the problem context. Outlines at least two material limitations of the analysis and steps to overcome them. These are relevant to the problem context.	Demonstrates a limited understanding of the model and its limitations. Outlines the business idea, performance of the model, and recommendations but at least one of these does not make sense or is difficult to understand. Outlines at least two limitations of the analysis and steps to overcome them but these are not material limitations.	Does not demonstrate an understanding of the model and its limitations. The video does not cover the four required components. or Outlines the business idea, performance of the model, and recommendation but these do not make sense or are difficult to understand. Outlines one or more limitations of the analysis and steps to overcome them but these are not material and lack relevance to the problem context.



Data Analytics Applications

Assignment Marking Rubric Semester 2 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 Proposed refinement	5 marks	Demonstrates a very strong understanding of the impact of the proposed refinement. Explains at least two relevant advantages and two relevant disadvantages. These are directly linked to the problem context. Uses language suitable for communication with your friend in real-estate.	Demonstrates a good understanding of the impact of the proposed refinement. Explains one relevant advantage and one relevant disadvantage. These are relevant to the problem context. Uses language suitable for communication with your friend in real-estate.	Demonstrates some understanding of the impact of the proposed refinement. Describes one relevant advantage and one relevant disadvantage. These are relevant to the problem context. Mostly uses language suitable for communication with your friend in real-estate.	Demonstrates a limited understanding of the impact of the proposed refinement. Describes one advantage and one disadvantage but it is not clear that these are relevant to the problem context. Does not use language suitable for communication with your friend in real-estate.	Does not demonstrate an understanding of the impact of the proposed refinement. Outlines some advantages and disadvantages but these are not relevant to the problem context. Does not use language suitable for communication with your friend in real-estate.
Total	100 marks					