COM00148M

Department of Computer Science

Big Data Analytics

SUMMATIVE ASSESSMENT BRIEF



|  |  |
| --- | --- |
| **Author** | Dr Phoebe Barraclough |
| **Assessment type** | Summative assignment |
| **Weighting** | 100% |
| **Release** | Week 3 |
| **Deadline** | Monday following Week 8, 13:00 (UK time) \* |

\* If this date falls on a UK public holiday or a University of York closure day, the submission date will change. Please check the submission point in the ‘Assignments’ area of the module in Canvas for the exact submission deadline.

# Module Learning Outcomes

The module learning outcomes (MLO’s) for this module are as follows:

1. Create a data set using modern database models and technology
2. Manipulate a data set to extract statistics and features.
3. Critically evaluate and apply data mining techniques/tools to build a classifier or regression model and predict values for new examples.
4. Analyse and communicate issues with scaling up to large data sets and use appropriate techniques to scale up the computation.
5. Critically discuss the need for privacy, identify privacy risks in releasing information, and design techniques to mediate these risks.

This assessment addresses **all** the module learning outcomes listed above.

# Assessment Background/Scenario

Rental prices may increase by up to 2% by the end of 2025. However, in some cases, if landlords apply cumulative rent update coefficients from the past three years, rent could rise by as much as 11% [1].  In light of this upward trend, you have been assigned the task of identifying, analysing, and proposing solutions to three key problem areas that were identified following a Rent “price” investigation. You are required to use data mining techniques (regression/classification) and tools (WEKA version 3.8.5 or Python 3) covered in the Big Data Analytics module.  Only the provided dataset, "apartment\_for\_rent", can be used; it should be cleaned and prepared to generate meaningful, targeted results.

## Data Set (.CSV)

The apartment rental dataset is provided under the CC0: Public Domain license. It is a combined dataset consisting of two CSV files: "apartment\_for\_rent\_train" and "apartment\_for\_rent\_test". One is intended for training and the other for testing. You may use one or both, depending on your chosen methodology; however, it is essential to explicitly state which dataset(s) were utilised and in what context within your submission.

## Reference:

[1] Supercasa, (2025), *Real Estate Market News* [Online]. Available: https://supercasa.pt/en-gb/noticias/rents-could-rise-by-up-to-11-in-225-in-specific-cases/n6385 [Accessed: Mar. 1, 2025]

|  |
| --- |
| At this level of study, there may be multiple “right answers” to any given problem. You may have to make assumptions about ambiguous or incomplete information, interpret the information provided, or make justifiable decisions about approaches and techniques to apply. These should flow from the information provided, your understanding of the course materials and your further research around the subject area under assessment. To support your work and meet the requirements of a passing grade and potentially achieve higher grades, you must state assumptions, interpretations and decisions, with justifications, clearly in your answer in order to assist the marker in understanding your work and giving you appropriate credit for it.  Equally, you must incorporate or refer to other authors' research to support your assumptions, interpretations and decisions, with justifications. Academic writing relies on evidence from various sources to build a strong argument. Instead of summarising individual sources, integrate key evidence into a cohesive discussion and be sure to use citations and references appropriately. This approach shows that your work is based on more than just personal opinion. |

# Assessment Tasks

## Rent “price investigation: (MLO2 & MLO3) (40%)

The apartment manager has posed the following question: Which characteristics influence apartment prices? To answer this, the manager suggests investigating the following:

1. Among the discrete variables (e.g., Balcony), identify three that are most likely to contribute to high-priced properties. Compare your selections and justify why you have selected them for the activity.
2. Assess whether there is a correlation—positive or negative—between the number of rooms, price, and rental duration of an apartment.
3. Determine whether an apartment’s address is a significant predictor of price increases, and identify three additional attributes that are associated with high rental rates. Justify your selections by comparing them to other variables and provide supporting evidence.
4. You are required to use WEKA or Python 3 to build a classification or regression model to support this analysis.

## Storing data and scalable solutions: (MLO1 & MLO4) (40%)

### Part 1: Design a relational database

The apartment manager is considering replacing the current flat file (CSV) system used to store most of their data. You have been assigned the task of designing a relational database to store the provided ‘apartment\_for\_rent’ dataset currently in CSV format. You must decide which features to include or adapt in order to store all the provided information, and justify your choices. To help the apartment manager assess the feasibility of this transition, you are required to provide the following:

1. A database design in the form of a UML-standard ER diagram, normalised to Third Normal Form (3NF).
2. Sample SQL based on your ER diagram, including:
3. A statement to insert a new row of data, covering all relevant attributes.
4. A query to extract all apartments priced at $1000 or less, that include an elevator, and use ‘USD’ as the currency.
5. A query to extract the average price for each currency, for comparison purposes.

### Part 2: Consider scaling

The apartment manager is also considering long-term solutions to support their plans for establishing international offices worldwide. This expansion is expected to generate significantly more data—potentially tens of megabytes. To manage and utilise this data effectively, the business will require a rapid-response system capable of operating efficiently in a global rental market. Assume that certain automated analyses must trigger immediate alerts—for example, when the count of a specific type of item exceeds a defined threshold. Using relevant examples from the dataset, propose a solution that distributes the processing load across multiple computers. Justify your choice of technologies and explain why this approach would be effective in this context.

## Considering customer-facing applications: (MLO5) (20%)

The apartment manager is considering developing a customer-facing application to provide instant customer service and make it easier for potential customer to browse and select available listings.  As part of this initiative, the manager also plans to collect personal information from potential customers through an online form, in order to offer personalised recommendations to returning users. Before proceeding, the manager must address three key privacy concerns. These concerns should be evaluated in the context of the new application, the business’s data analysis goals (Task 1), and the planned move toward permanent data storage (Task 2).

You are required to identify the three most critical privacy issues and, for each one, propose appropriate strategies for managing the risks. Your discussion should relate to the given scenario and the dataset provided.

# Deliverables

Your assignment must be laid out following the formatting guidelines that are specified in the ‘Submission Formatting’ page in Canvas. This includes restrictions on the length of the appendices, expectations on how your work must be presented and any penalties when these expectations are not met.

* Your submission must consist of a **report and supplementary files, either the final data files in ARFF** **format if you are using WEKA, or the Python scripts and CSV files if you opted to use Python 3**. ARFF, Python code and CSV files must be submitted as part of a .zip archive, which is separate from the report.

Given the tasks above, you must produce a report detailing solutions and justifying your decisions. You will need to provide supporting evidence for each solution in the form of images/screenshots of the practical work you must have undertaken to complete this assessment, including anything that is specifically requested. Your decisions and justifications must be supported by the current literature.

Your report must not exceed **3,000 words** in total and consist of three clear sections – one for each task. Further formatting details and essential points are given below.

## Task 1: Rent “price” investigation (MLO2 & MLO3 40%)

(Suggested word count for this section: 1500, i.e. 500 words per solution)

You can import the given data files into a spreadsheet to initially scrutinise and review the data, as well as perform any cleaning before you translate it into the .ARRF format for use in Weka, or CSV file for use in Python. A notepad application can also be used to do this.

You are required to submit the .ARRF (or CSV and Python files) that you have used to perform the required analysis, so we are able to verify your results where necessary.

Your report must discuss and present the following for this task:

* Any assumptions you made about the scenario or areas of investigation.
* Any pre-processing you have undertaken to make the data fit for purpose.
* Clearly state the specific analysis techniques you have employed in your solutions.
* Justification for the selection of techniques/approach, given the nature of the data and the requirements of the investigation, which is effectively supported by literature.
* Summarise the overall findings from your analysis and include the specific results.
* Consider how your results, individually and as a whole, answer the question posed by the apartment manager.
* Critically evaluate the approach you have taken, and the techniques selected in the context of the given data set and scenario. You must reflect on what you have learned from the process and identify what was effective/ineffective. An honest appraisal of an ineffective approach will gain credit. This discussion must be supported by literature.
* For each solution, you must provide images/screenshots that demonstrate the tool you have selected in Weka, any relevant settings, and the output produced by that tool.
* For each solution, you must provide (an) additional file(s) containing the final data structure you used in Weka (in Weka’s .ARFF format), or CSV files and Python scripts if you used Python.

## Task 2: Storing data and possible solutions (MLO1 & MLO4) (40%)

(Suggested word count for this section: 1000, i.e. 400 words for Part 1, 600 for Part 2)

### Part 1:

You can either type out SQL statements OR build and screen capture (image) the SQL from a live database. Note, however, that typed SQL statements will contribute to your word count.

You may choose to demonstrate the normalisation process using specific examples of your approach. Any further visual aids (tables/models/diagrams etc.) may also be presented in the appendices, not in the main body of the report.

You must discuss the approach you have taken to creating the relational database structure, referring to the key aspects of your design, such as why an attribute/variable was selected as a primary key, or why you have elected to contain a specific set of variables/attributes with the same table.

### Part 2:

You must discuss and present the following for this task:

* Discuss and present any additional assumptions you have made about the scenario or analysis requirements. You may also restate any assumptions from Task 1 that are particularly relevant here.
* Justification for the technology/technique selected, and the approach to your solution(s), given the nature of the data set and the context of the scenario.
* Clear comparison of benefits and limitations against other potential technology/techniques/solutions.
* It is expected that this section will be supported by literature, with effective use of citations (and attached reference list) to support your claims.

To obtain a **PASS** in this task, you **MUST** present the requested ERD design and sample SQL and use correctly formatted citations and a supporting reference list.

## Task 3: Considering customer-facing application (ML05) (20%)

(Suggested word count for this section: 500)

* Clearly state the three privacy issues you intend to discuss.
* Give clear reasoning as to why these are potential issues, and what evidence you have drawn on to identify them.
* Clearly present potential mitigation strategies for each identified issue. Where relevant, compare these strategies to those used in similar scenarios or issues to support their effectiveness
* Effective use of citations (and attached reference list) to support your claims.

To attain a **Pass** in this task you **MUST** support your discussion with relevant literature using citations and a reference list, using the IEEE format.

## Note about Figures and Appendices

All diagrams, images or screenshots for any task must be presented within the body of the report and close to the page(s) where that task is answered. They must be numbered (e.g. Fig 1, Table 1 etc.) and referred to in the associated text.

For Task 1, 3 to 6 figures are sufficient for illustration. You must decide on which figures are most relevant and place other supplementary ones in the unmarked appendix. For Task 2, part 1, the ERD and SQL are required to obtain a Pass. Figures or tables of the normalisation process may be shown in the unmarked appendix.

Note that all figures, whether in the body of the report or in the appendix, **MUST** be referenced and discussed within the report. However, only those figures in the body of the report are required to obtain a Pass.

## Referencing

You are required to use the [IEEE referencing style](https://subjectguides.york.ac.uk/referencing-style-guides/ieee) for citing books, articles, and all other sources (such as websites) used in your assignment.

Good referencing is essential in order to meet the standards of academic integrity set by the University. All your sources must be acknowledged, regardless of whether you’ve included direct quotes or not. Visit your **Academic Integrity Tutorial module** in Canvas for additional guidance on effective referencing.

# Marking Criteria

|  |  |  |  |
| --- | --- | --- | --- |
| **Learning Outcome** | **Section/Task** | **Criteria** | **Available marks** |
|  | **Task 1: Rent “price” investigation** | | |
| 2/3 | 1 | **Approach and results:**  The evidence and discussion present a clear investigation of the data as requested.  Appropriate techniques and pre-processing have been utilised to undertake this.  Results are clear and credible (i.e. not obviously invalid). | **20** |
| 3 | 1 | **Justification:**  There is a clear and appropriate justification of the approach/techniques used.  This is supported by the literature. | **10** |
| 3 | 1 | **Critical evaluation:**  The approach has been critically evaluated, which is in alignment with the results generated.  Any identified issues are valid, accurately described, and of genuine concern.  This is supported by the literature. | **10** |
|  | **Task 2: Storing data and scalable solutions** | | |
| 1 | 2 | **Database design:**  Is correctly presented and appropriately normalised to 3NF.  The discussion supports and clarifies the approach taken and the decisions made. | **10** |
| 1 | 2 | **Sample SQL statements**:  Are correct in relation to the presented design. | **10** |
| 4 | 2 | **Scalable solution:**  A viable approach is described that uses multiple technologies and would be likely to achieve a worthwhile improvement in performance (given coordination overhead etc). | **20** |
|  | **Task 3: Considering customer-facing application** | | |
| 5 | 3 | **Privacy issues:**  Are of genuine concern and are appropriate given the context of the scenario and tasks. | **10** |
| 5 | 3 | **Mitigation strategies:**  The strategies discussed to deal with the privacy issues offer realistic solutions and are supported by the literature and current standards. | **10** |
|  |  | **TOTAL:** | 100 |

# 

# Marking Criteria: Grade breakdown

The grade breakdown is written for assessment markers. All statements in each grade band need to be met to be awarded that band.

## Task 1: Rent “price” investigation – Approach and results – 20%

*MLO2: Manipulate a data set to extract statistics and features.*

*MLO3: Critically evaluate and apply data mining techniques/tools to build a classifier or regression model and predict values for new examples.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Either no evidence has been presented OR little to none of the presented evidence supports the required areas (discussion, approach and results) effectively. * Either no discussion has been presented OR little to none of the discussion presented for the areas of investigation demonstrates appropriate techniques/approaches, given the nature of the data set and scenario. * Either no results have been presented OR little to none of the results presented for the required areas of investigation are clear or credible, given the nature of the technique/approach used. |
| 40-49% | Marginal fail | * Evidence has been presented which supports some of the required areas (discussion, approach or results) to some degree. * The discussion presents some techniques/approaches for the required areas of investigation, some of which are appropriate, but no single area is wholly complete, given the nature of the data set and scenario. * Some results presented for the required areas of investigation are appropriate, but not wholly complete or credible for any single area, given the nature of the technique/approach used. |
| 50%-59% | Pass | * Evidence has been presented which supports most of the required areas (discussion, approach or results), with some areas of weakness. * The discussion clearly indicates that at least **one** of the required areas of investigation has utilised an appropriate technique/approach, given the nature of the data set and scenario. * The results for at least **one** of the required areas of investigation are clear and credible, given the nature of the technique/approach used. |
| 60%-69% | Merit | * Evidence has been presented which supports all the required areas (discussion, approach or results), with few weak areas. * The discussion clearly indicates that at least **two** of the required areas of investigation have utilised appropriate techniques /approaches, given the nature of the data set and scenario. * The results for at least **two** of the required areas of investigation are clear and credible, given the nature of the technique/approach used. |
| 70%-100% | Distinction | * Evidence has been presented which supports all the required areas (discussion, approach or results), with little to no areas of weakness. * All **three** of the required areas of investigation have utilised appropriate techniques/approaches, given the nature of the dataset and scenario. * The results for all **three** of the required areas of investigation are clear and credible, given the nature of the technique/approach used. |

## Task 1: Rent “price” investigation – Justification – 10%

*MLO2: Manipulate a data set to extract statistics and features.*

*MLO3: Critically evaluate and apply data mining techniques/tools to build a classifier or regression model and predict values for new examples.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Few or no citations or reference list are presented, none of which is in the correct format. * The justification for the techniques/approach used is either not present, incomplete or wholly ineffective/inappropriate for all areas of investigation, given the nature of the data set and scenario. * The justification is not supported by the literature, or the justification is supported in a limited way (many areas of ineffectiveness/inappropriateness). |
| 40-49% | Marginal fail | * Some citations and a reference list are presented, but are not necessarily in the correct format. * The justification for the techniques/approach used has some areas of effectiveness/appropriateness, but no single area of investigation is wholly complete, given the nature of the data set and scenario. * The justification is supported by the literature, with a few areas of effectiveness/appropriateness. |
| 50%-59% | Pass | * Most citations and the reference list are correctly formatted using IEEE (with minor errors). * There is an effective justification for the techniques/approach used for at least **one** of the investigation areas. * The justification is supported by the literature, with some areas of weakness (in terms of effectiveness/appropriateness). |
| 60%-69% | Merit | * Citations and the reference list are correctly formatted using IEEE. * There is a clear and appropriate justification for the techniques/approach used for at least **two** of the investigation areas. * The justification is supported by the literature, with few areas of weakness (effectiveness/appropriateness). |
| 70%-100% | Distinction | * Citations and the reference list are correctly formatted using IEEE. * There is a clear and appropriate justification for the techniques/approach used for all **three** of the investigation areas. * The justification is supported by the literature, with few or no areas of weakness (in terms of effectiveness/appropriateness). |

## Task 1: Rent “price” investigation – Critical Evaluation – 10%

*MLO2: Manipulate a data set to extract statistics and features.*

*MLO3: Critically evaluate and apply data mining techniques/tools to build a classifier or regression model and predict values for new examples.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Few or no citations or reference list are presented, none of which is in the correct format. * The critical evaluation for the techniques/approach used is either not present, incomplete, misaligned with the presented results, or wholly ineffective/inappropriate for all areas of investigation, given the nature of the data set and scenario. * The critical evaluation is not supported by the literature, or critical evaluation is supported in a limited way (many areas of ineffectiveness/inappropriateness). |
| 40-49% | Marginal fail | * Some citations and a reference list are presented but are not necessarily in the correct format. * The critical evaluation for the techniques/approach used has some areas of effectiveness/appropriateness, and some areas may be aligned with the presented results, but no single area of investigation is wholly complete, given the nature of the data set and scenario. * The critical evaluation is supported by the literature, with a few areas of effectiveness/appropriateness. |
| 50%-59% | Pass | * Most citations and the reference list are correctly formatted using IEEE (with minor errors). * The critical evaluation provides an accurate and rational appraisal of the effectiveness of the techniques/approach used for at least **one** of the areas of investigation, which is aligned with the presented results. * The critical evaluation is supported by the literature, with some areas of weakness (effectiveness/appropriateness). |
| 60%-69% | Merit | * Citations and the reference list are correctly formatted using IEEE. * The critical evaluation provides an accurate and rational appraisal of the effectiveness of the techniques/approach used for at least **two** of the areas of investigation, which is aligned with the presented results. * The critical evaluation is supported by the literature, with few areas of weakness (effectiveness/appropriateness). |
| 70%-100% | Distinction | * Citations and the reference list are correctly formatted using IEEE. * The critical evaluation provides an accurate and rational appraisal of the effectiveness of the techniques/approach used for all **three** of the areas of investigation, which is aligned with the presented results. * The critical evaluation is supported by the literature, with few or no areas of weakness (effectiveness/appropriateness). |

## Task 2: Storing data and scalable solutions – Database design – 10%

*MLO1: Create a data set using modern database models and technology.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * The requested ERD design has either not been presented or is not readable. * The ERD diagram demonstrates little or no effective relational database structure (tables and naming are not rational), which demonstrates little or no effective normalisation, and little or no correct notation. * Either no discussion is presented or what is presented describes a limited approach to the decisions made during the construction of the relational database. |
| 40-49% | Marginal fail | * The requested ERD design has been presented but may not be wholly readable. * The ERD diagram demonstrates a limited relational database structure (tables and naming are not always rational) which has not achieved 3NF, but demonstrates some effective normalisation, and mostly uses correct notation. * The discussion presents the approach to the decisions made during the construction of the relational database, which is rational to some degree. |
| 50%-59% | Pass | * The requested ERD design has been presented and is readable. * The ERD diagram demonstrates a mostly clear and accurate relational database structure (tables and naming are rational) which has been normalised to 3NF, and mostly uses correct notation. * The discussion presents a mostly rational approach to the decisions made during the construction of the relational database. |
| 60%-69% | Merit | * The requested ERD design has been presented and is readable. * The ERD diagram demonstrates a clear and accurate relational database structure (tables and naming are rational) which has been normalised to 3NF, and uses correct notation. * The discussion presents a rational approach to the decisions made during the construction of the relational database. |
| 70%-100% | Distinction | * The requested ERD design has been presented and is readable. * The ERD diagram demonstrates clear and accurate relational database structure (tables and naming are rational) which has been normalised to 3NF, and uses correct notation. * The discussion presents a clear and rational approach to the decisions made during the construction of the relational database. |

## Task 2: Storing data and scalable solutions – Sample SQL statements – 10%

*MLO1: Create a data set using modern database models and technology.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Either no SQL has been presented, or all presented SQL statements are incorrect and poorly formed to a greater degree (minor and/or major errors across all three statements). |
| 40-49% | Marginal fail | * The presented SQL demonstrates mostly correct (minor errors) for at least one statement, input or output. |
| 50%-59% | Pass | * The presented SQL demonstrates mostly correct (minor errors) for both input and at least one output statement. |
| 60%-69% | Merit | * The presented SQL demonstrates correct construction for both input and at least one output statement. |
| 70%-100% | Distinction | * The presented SQL demonstrates correct construction for both input and all output statements. |

## Task 2: Storing data and scalable solutions – Proposing a scalable solution – 20%

*MLO4: Analyse and communicate issues with scaling up to large data sets and use appropriate techniques to scale up the computation.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Few or no citations or reference list are presented, none of which is in the correct format. * Either no approach for a scalable solution has been presented or the approach that is presented is unfeasible, with no potential gains identified, and a number of unidentified limitations. * The discussion either presents no justification for the approach, or what is presented is limited, with few or no areas of effectiveness/appropriateness. * The discussion either presents no comparison with alternative technologies/solutions, or what is presented is limited, with few or no areas of effectiveness/appropriateness. * The discussion is not supported by the literature, or the discussion is only supported in a limited way (many areas of ineffectiveness/inappropriateness). |
| 40-49% | Marginal fail | * Some citations and a reference list are presented but are not necessarily in the correct format. * A somewhat viable approach has been presented for a scalable solution, which does not clearly identify any potential gains, or has unidentified limitations. * The discussion presents a somewhat rational justification for the approach, with a few areas of effectiveness/appropriateness. * There is a somewhat clear comparison with some alternative technologies/solutions, which supports the solution to some degree, demonstrating a few areas of effectiveness/appropriateness. * The discussion is supported by the literature, with a few areas of effectiveness/appropriateness. |
| 50%-59% | Pass | * Most citations and the reference list are correctly formatted using IEEE (with minor errors). * A mostly viable approach has been presented for a scalable solution, identifying at least one potential gain. * The discussion presents a mostly clear rationale and justification for the approach, with some areas of weakness (effectiveness/appropriateness). * There is a mostly clear and effective comparison with some alternative technologies/solutions, that mostly supports the solution, with some areas of weakness (effectiveness/appropriateness). * The discussion is supported by the literature, with some areas of weakness (effectiveness/appropriateness). |
| 60%-69% | Merit | * Citations and the reference list are correctly formatted using IEEE. * A viable approach has been presented for a scalable solution, with potential gains and few areas of weakness. * The discussion presents a clear rationale and justification for the approach, with few areas of weakness (effectiveness/appropriateness). * There is a clear and effective comparison with alternative technologies/solutions, that clearly supports the presented solution, with few areas of weakness (effectiveness/appropriateness). * The discussion is supported by the literature, with few areas of weakness (effectiveness/appropriateness). |
| 70%-100% | Distinction | * Citations and the reference list are correctly formatted using IEEE. * A viable and effective approach has been presented for a scalable solution, with clear gains and few or no areas of weakness. * The discussion presents a clear rationale and justification for the approach, with few or no areas of weakness (effectiveness/appropriateness). * There is a comprehensive comparison with alternative technologies/solutions, which clearly supports the presented solution, with few or no areas of weakness (effectiveness/appropriateness). * The discussion is supported by the literature, with few or no areas of weakness (effectiveness/appropriateness). |

## Task 3: Considering customer-facing application – Privacy issues – 10%

*MLO5: Critically discuss the need for privacy, identify privacy risks in releasing information, and design techniques to mediate these risks.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Few or no citations or reference list are presented, none of which is in the correct format. * Either the privacy issue(s) identified are not valid/relevant to the task and scenario’s context, OR they are only discussed in a limited way. * The identified privacy issues are not supported by the literature, or they are only supported in a limited way (many areas of ineffectiveness/inappropriateness). |
| 40-49% | Marginal fail | * Some citations and a reference list are presented but are not necessarily in the correct format. * Although valid privacy issues have been identified that are potentially relevant to the task and scenario, none are discussed in a complete and comprehensive way. * The identified privacy issues are supported by the literature, but with few areas of effectiveness/appropriateness |
| 50%-59% | Pass | * Most citations and the reference list are correctly formatted using IEEE (with minor errors) * At least **one** valid privacy issue has been identified that is relevant to the task and scenario’s context and is discussed in a complete and comprehensive way. * The identified privacy issues are supported by the literature, with some areas of weakness (effectiveness/appropriateness). |
| 60%-69% | Merit | * Citations and the reference list are correctly formatted using IEEE. * At least **two** valid privacy issues have been identified that are relevant to the task and scenario’s context and are discussed in a complete and comprehensive way. * The identified privacy issues are supported by the literature, with few areas of weakness (effectiveness/appropriateness). |
| 70%-100% | Distinction | * Citations and the reference list are correctly formatted using IEEE. * At least **three** valid privacy issues have been identified that are relevant to the task and scenario’s context and are discussed in a complete and comprehensive way. * The identified privacy issues are supported by the literature, with few or no areas of weakness (effectiveness/appropriateness). |

## Task 3: Considering customer-facing application – Mitigation strategies – 10%

*MLO5: Critically discuss the need for privacy, identify privacy risks in releasing information, and design techniques to mediate these risks.*

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | * Few or no citations or reference list are presented, none of which is in the correct format. * Either no mitigation strategies at all are presented, OR if they are, they are not feasible, OR do not clearly align with any given privacy issues, OR are not relevant given the context of the task and scenario. * The mitigation strategies are not supported by the literature, or they are only supported in a limited way (many areas of ineffectiveness/inappropriateness). |
| 40-49% | Marginal fail | * Some citations and a reference list are presented but are not necessarily in the correct format. * Mitigation strategies are presented, some of which are feasible, but they either do not clearly align with given privacy issues, or the context of the task and scenario. * The mitigation strategies are supported by the literature, but with few areas of effectiveness/appropriateness |
| 50%-59% | Pass | * Most citations and the reference list are correctly formatted using IEEE (with minor errors) * At least **one** of the mitigation strategies presented is feasible, and aligns clearly with relevant given privacy issues, and the context of the task and scenario. * The mitigation strategies are supported by the literature, with some areas of weakness (effectiveness/appropriateness). |
| 60%-69% | Merit | * Citations and the reference list are correctly formatted using IEEE. * At least **two** of the mitigation strategies presented are feasible, and align clearly with relevant given privacy issues, and the context of the task and scenario. * The mitigation strategies are supported by the literature, with few areas of weakness (effectiveness/appropriateness). |
| 70%-100% | Distinction | * Citations and the reference list are correctly formatted using IEEE. * At least **three** of the mitigation strategies presented are feasible, and clearly align with relevant given privacy issues, and the context of the task and scenario. * The mitigation strategies are supported by the literature, with few or no areas of weakness, (effectiveness/appropriateness). |

# Assessment Submission

You will submit your assessment in the ‘Assignments’ area of the module in Canvas. Please check your Canvas module for the specific submission date for this assignment.

This assessment requires you to anonymously upload your submission to Canvas. If you are submitting multiple files, you must upload all files simultaneously to ensure that they are marked as a single submission. If you want to resubmit one component of your work, you need to re-upload all other files at the same time: every submission must include **all** files required by the assessment brief.

We recommend that you allow at least 30 minutes before the deadline to upload your submission, as failure to upload your assessment file within the allotted time is not admissible as an exceptional circumstance.

The webpage [How do I submit an online assignment?](https://community.canvaslms.com/t5/Student-Guide/How-do-I-submit-an-online-assignment/ta-p/503) provides further technical information on how to upload an assessment. The advice given here comes directly from Canvas. We do not recommend uploading assignments by mobile. We recommend you view the submission immediately after uploading your work to ensure the correct file has been submitted and no technical errors have occurred.

If you face any technical difficulties whilst trying to submit this assessment, then contact Canvas support on [support@instructure.com](mailto:support@instructure.com) or +44 80 0060 8442 (available 24 hours) in advance of the deadline. You should also email [york-online-assessment@york.ac.uk](mailto:york-online-assessment@york.ac.uk) as a matter of urgency to report the issue and receive further instruction.

# Assessment Policies

This assessment is subject to the policies stated on the ‘Summative Assessment Policies’ page in Canvas. These policies include (but are not limited to):

* Academic Integrity and submission of student work to Turnitin
* Advice on anonymising your assessment
* Penalties for late submission
* Marking policy for multiple submissions
* The Fit to Sit / Submit policy
* Passing mark and module reassessment

Please ensure that you have read and understood these policies before starting the assessment.