



Assignment 2 Specification

SWE5204

Advanced Databases and Big Data

Course/Program	BEng Software Engineering & BSc (Hons) Computing		
Module Name	SWE5204: Advanced Database and Big Data		
Assessment Number	2 of 2		
Assessment Type (and weighting)	Portfolio Component (50% of overall mark)		
Assessment Name	Data Science and Big Data		
Issue Date	w/c 20/11/23		
Assessment Submission Date	Assessment item	Due Date	Weight
	Assignment 2 of 2	12/01/24 by 23:59	50%

Learning Outcomes Assessed

LO3: Apply appropriate database concepts and techniques to solve given problems.

LO4: Demonstrate the application of appropriate Big Data tools for advanced analytics

Scenario:

Movies R Us is an on-demand movie company which operate all over the world via the internet. They have generated a huge amount of data on movie ratings over the years.

You have recently been recruited by Movies R Us as a Junior Data Scientist. The company are new to data analytics and have heard how beneficial this can be therefore, they want you to explore the dataset so they can write an article analysing movie ratings by the critics and audiences. The report should also justify the budget of the movies plus additional analysis such as profits by specific categories. They are hoping to obtain a more update data set in the next 3 months where similar tasks can be performed to help boost their on-demand sales by offering the most popular on demand films, past and present.

The data set for this assessment is in the assignment section of Moodle.

Portfolio Component 3: Solving Data Science Problem (25% of overall Mark)

Part 1 - Exploring the Data:

You must analyse the data very carefully by running some simple tests in R and you have set up the following tasks, which are agreed upon by your team leader. For the following tasks, you must write code and generate graphs (at least one graph for tasks 2-6).

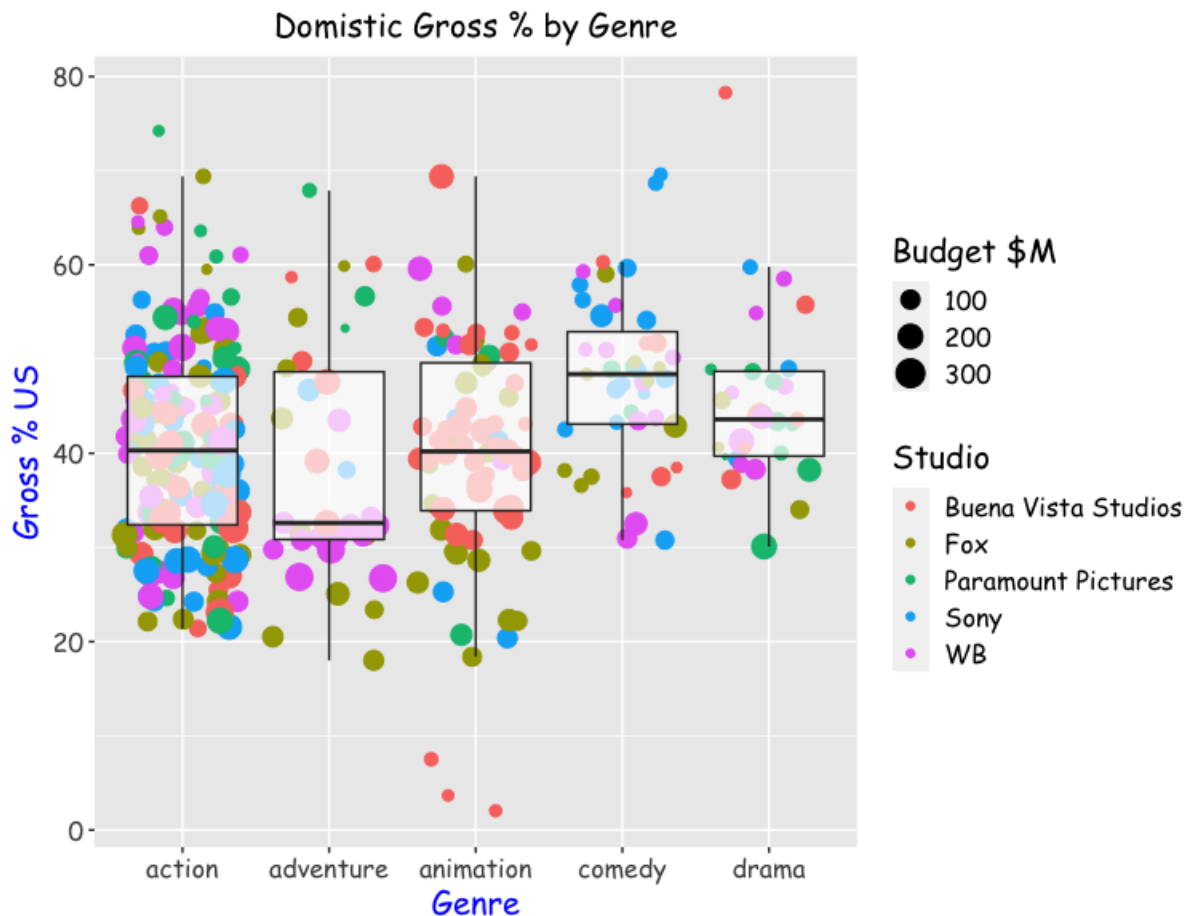
1. Explore your dataset (using str, nrow etc.) and explain your understanding.
2. How Genre impacts the budget of the movie?
3. Is there any relation between the critic rating and the budget?
4. Is there any relationship between the audience ratings and the budget?
5. Show the correlation between audience and critic ratings has evolved throughout the years by movie genre. (Request from the CEO)
6. Create a graph to show the number of films from the dataset categorised by Genre.

Part 2 - Advanced Analytics:

Once you have completed the above tasks, your manager gives you an extended Movie data set. The dataset contains more columns than the previous one. Using this new data set, you should complete the following tasks.

For the following tasks, you must write code and generate graphs (at least one graph for tasks 1-5).

1. They give you the following graph image as the R code is not found and they would like to remove the spelling mistake in the Graph Title and rename the Graph Title "Gross Percentage By Genre". You need to recreate the graph by writing R code. You must use the Grammar of Graphics to recreate the following graph. You must also explain your code and display the output at each step.



2. Write R code to find the trend of the Day of the week that most/least movies were released compared to other days.
3. Identify if the profit of a movie depends on any of the features in this data set i.e. genre, director, profit etc
4. Use ggplot and boxplot to identify if there is an anomaly / any anomalies in the data?
5. Find if there is any further insight you can find from this data set, marks will be awarded by the number of further useful graphs as well as their complexity.

Note: You must provide code, graphs and appropriate analysis of the findings for each of the sub-tasks as well as explaining your code. Copy and paste your R code in your report for each task (main body), screenshots will not be accepted. Each graph & code should be clearly labelled.

Demonstration: You must demonstrate your data science solution via Zoom. You will show evidence of your code being executed, explain the code and analyse your findings. The date and time of the demonstration will be published via Moodle and will take place during your timetable session w/c 15/01/24. **(All tasks must be finished and submitted in your assignment before the demonstration, the demonstration does not mean an extension to carry out further work – this will be checked after submission)**

Important: No demonstration means Zero marks for task 1.

The explanation of your code and the demonstration will provide evidence that this is your own work as well as:

Part 3 – 500 Word Reflection:

You also should write a reflective piece of at least 500 words, where you reflect on the process of completing the R analysis tasks. Explain honestly what challenges you faced while carrying out the work, what you feel you managed to do well and what you feel you could have done better and how you might achieve that in future. Provide actual evidence and specific examples here, do not merely say, for example “I found the R language difficult because I’m not good at it”, this needs to be personal and specific to you!

Portfolio Component 4: Big Data Tools and Techniques (25% of overall Mark)

Evaluate alternative appropriate Big Data technologies for Movies R Us to develop a database solution. Since the company wants to analyse both structured and unstructured data in real-time to check the performance and popularity of movies on demand, they need a system that can deploy effective data analytics. Analyse big data analysis and visualisation techniques that influence the organisation’s decision-making in a cost-effective database solution.

Write a report of 2500 words to inform the company management about the technologies available and how they will fit for the company’s new database solutions. The report should identify and compare various Big Data visual tools and techniques suggest three suitable visual tools and/ or techniques to meet the company’s future need.

Word Count (Component 4): The report should have a word count of **2500 words**.

Expected Number of Sources: The Report should have at least 10 references of which 3 should be relevant peer-reviewed journal/ conference papers.

Secondary Research Requirements:

Secondary research support is expected should be correctly cited using Harvard Referencing for both in-text citations and Reference Structure. For further details please see <https://leaponline.bolton.ac.uk/My-Academic-Development/My-Writing-Techniques/Referencing/Level-2/Harvard-Referencing.aspx>

Submission: You must submit **Part 1 and 2 in a single (MS Word) document** through the appropriate Moodle Turnitin link by **23:59 on 12 January 2024**.

Grading

A percentage mark will be provided based on *General Assessment Guidelines for Written Assessments*. Grading is as follows:

A:	70 - 100%
B:	60 - 69%
C:	50 - 59%
D:	40 - 49%

Marks below 40% will be classed as fail.

Specific Assessment Criteria:

70% and above

- Have performed an excellent analysis, understanding, and implemented database systems for a specific problem,
- Critically analysed and evaluated domain-specific solutions and a justified conclusion,
- Have provided an excellent review of current and state-of-the-art big data tools and techniques and demonstrate understanding,
- Extensive research demonstrating the use of a range of current and quality secondary research resources.

60%- 69

- Have performed a comprehensive analysis, understanding, and implemented database systems for a specific problem,
- Clearly analysed and evaluated domain-specific solutions and a justified conclusion,
- Have provided a comprehensive review of current and state-of-the-art big data tools and techniques and demonstrate understanding,
- Research demonstrating use of a wide range of current and quality secondary research resources.

50%-59

- Demonstrates sound analysis, understanding, and implemented database systems for a specific problem,
- Presents a logical analysis and evaluated domain-specific solutions and a justified conclusion,
- Demonstrates a sound and breadth review of current and state-of-the-art big data tools and techniques and demonstrate understanding,
- Research demonstrating use of a range of relevant current and quality secondary research resources.

40%-49

- Demonstrates sufficient analysis, understanding, and implemented database systems for a specific problem,
- Presents a valid analysis and evaluated domain-specific solutions and a sufficient conclusion,
- Demonstrates a sufficient review of current and state-of-the-art big data tools and techniques and demonstrate understanding,
- Research demonstrating use of limited relevant current and quality secondary research resources.

Note: This assignment will also be assessed by using the **General Assessment Guidelines for Assessments Level HE5**.

Guidelines for the Preparation and Submission of Written Assessments

1. Written assessments should be word-processed in Arial or Calibri Light font size 12. There should be double-spacing and each page should be numbered.
2. There should be a title page identifying the programme name, module title, assessment title, your student number, your marking tutor and the date of submission.
3. You should include a word-count at the end of the assessment (excluding references, figures, tables and appendices).
Where a word limit is specified, the following penalty systems applies:
 - Up to 10% over the specified word length = no penalty
 - 10 – 20% over the specified indicative word length = 5 marks subtracted (but if the assessment would normally gain a pass mark, then the final mark to be no lower than the pass mark for the assessment).
 - More than 20% over the indicative word length = if the assessment would normally gain a pass mark or more, then the final mark will capped at the pass mark for the assessment.
4. All written work should be referenced using the standard University of Bolton referencing style : **Harvard** – see: <https://libguides.bolton.ac.uk/resources/referencing/>
5. Unless otherwise notified by your Module Tutor, electronic copies of assignments should be saved as word documents and uploaded into Turnitin via the Moodle class area. If you experience problems in uploading your work, then you must send an electronic copy of your assessment to your Module Tutor via email BEFORE the due date/time.
6. Please note that when you submit your work to Moodle, it will automatically be checked for matches against other electronic information, as well as for hidden text characters and AI generated text. You will be able to see similarity matches but not currently flags

for hidden characters and AI-generated text. The outcomes of Turnitin reports may be used as evidence in an academic misconduct investigation (see Section 14).

7. Late work

Late work will be subject to the following penalties:

- Up to 7 calendar days late = 10 marks subtracted but if the assignment would normally gain a pass mark, then the final mark to be no lower than the pass mark for the assignment.
- More than 7 calendar days late = This will be counted as non-submission and no marks will be recorded.

Late submission of assessments on refer and those which are graded Pass/Fail only, is not permitted unless an extension is approved. See below.

8. Extensions

In the case of exceptional and unforeseen circumstances, an extension of up to 14 days after the assessment deadline may be requested using the standard University Extension Request Form. For approval there would need to be an explanation and evidence of relevant circumstances.

Longer extensions for individual assessments, projects and artefacts may be granted, at the discretion of the Programme Leader.

Requests for extensions which take a submission date past the end of the module (normally week 15) must be made using the Mitigating Circumstances procedure.

Some students with registered disabilities will be eligible for revised submission deadlines. Revised submission deadlines for disability adjustments do not require the completion extension request paperwork. However, students should request these in writing in advance.

Please note that the failure of data storage systems is not considered to be a valid reason for an extension. It is therefore important that you keep multiple copies of your work on different storage devices before submitting it.

Academic Misconduct

Academic misconduct may be defined as any attempt by a student to gain an unfair advantage in any assessment. This includes plagiarism, collusion, commissioning (contract cheating) amongst other offences. In order to avoid these types of academic misconduct, you should ensure that all your work is your own and that sources are attributed using the correct referencing techniques. You can also check originality through *Turnitin*.

Please note that penalties apply if academic misconduct is proven. See the following link for further details:

<https://www.bolton.ac.uk/student-policy-zone/student-policies-2023-24/academic-misconduct-regulations-and-procedures-23-24>

GENERAL ASSESSMENT GUIDELINES – LEVEL HE5

	Relevance Learning outcomes must be met for an overall pass	Knowledge and Understanding	Analysis, Creativity and Problem-Solving	Self-awareness and Reflection	Research/Referencing	Written English	Presentation and Structure
Class I (Exceptional Quality) 85% - 100%	Work is directly relevant and expertly addresses the requirements of the brief. Learning outcomes are met.	Demonstrates an exceptional breadth and depth of knowledge and understanding of theory and practice beyond the threshold expectation for the level. Demonstrates sophisticated conceptual understanding of specialised areas.	Presents a perceptive synthesis and critical evaluation of findings' drawing on a broad range of relevant sources in order to draw clear, systematic, justified conclusions. Provides a critical insight and interprets complex matters and ideas comprehensively. Demonstrates creative flair and a high level of originality. Demonstrates exceptional problem-solving skills and initiative.	Provides insightful reflection and critical self-awareness in relation to the outcomes of own work and personal responsibility.	A wide range of contemporary and relevant reference sources selected and drawn upon. Sources cited accurately in both the body of text and in the Reference List/ Bibliography.	Writing style is clear and appropriate to the requirements of the assessment. An exceptionally well written answer with competent spelling, grammar and punctuation. For example, paragraphs are well structured and include linking and signposting. Sentences are complete and different types are used. A wide range of appropriate vocabulary is used.	The presentational style and layout are correct for the type of assignment. Evidence of planning and logically structured. Where relevant, effective inclusion of, and reference to, figures, tables and images.
Class I (Excellent Quality) 70% - 84%	Work is relevant and comprehensively addresses the requirements of the brief. Learning outcomes are met.	Demonstrates an excellent breadth and depth of knowledge and understanding of theory and practice for this level. Demonstrates in-depth understanding of key areas in different contexts.	Presents an excellent appraisal of findings through the critical analysis of information sources. Draws clear, systematic, justified and perceptive conclusions. Demonstrates creative flair, a high level of originality and initiative. Demonstrates a critical understanding of problem-solving approaches and applies excellent problem-solving skills.	Provides excellent reflection and self-awareness in relation to the outcomes of own work and personal responsibility.	A wide range of relevant reference sources selected and drawn upon. Sources cited accurately in both the body of text and in the Reference List/ Bibliography.	Writing style is clear and appropriate to the requirements of the assessment. An excellently well written answer with competent, spelling, grammar and punctuation. For example, paragraphs are well structured and include linking and signposting. Sentences are complete and different types are used. A wide range of appropriate vocabulary is used.	The presentational style and layout are correct for the type of assignment. Evidence of planning and logically structured. Where relevant, effective inclusion of, and reference to, figures, tables and images.
Class II/i (Very Good Quality) 60% - 69%	Work is relevant and addresses most of the requirements of the brief. Learning outcomes are met.	Demonstrates a thorough breadth and depth of knowledge and understanding of theory and practice for this level. Demonstrates very good understanding of key areas in different contexts.	Presents a perceptive appraisal of findings through the critical analysis of information. Draws clear, justified and thoughtful conclusions. Demonstrates a very good level of creativity and originality. Demonstrates a critical understanding of problem-solving approaches and applies strong problem-solving skills.	Provides very good reflection and self-awareness in relation to the outcomes of own work and personal responsibility, as required by the assessment.	A range of appropriate reference sources selected and drawn upon. Sources cited accurately in the main both the body of text and in the Reference List/Bibliography.	Writing style is clear and appropriate to the requirements of the assessment. A very well written answer with competent spelling, grammar and punctuation. For example, paragraphs are well structured and include linking and signposting. Sentences are complete and different types are used. A range of appropriate vocabulary is used.	The presentational style and layout are correct for the type of assignment. Evidence of planning and logically structured in the main. Where relevant, effective inclusion of, and reference to, figures, tables and images.

Class II/i (Good Quality) 50% - 59%	<p>Work addresses key requirements of the brief. Some irrelevant content.</p> <p>Learning outcomes are met.</p>	<p>Demonstrates a sound breadth and depth of knowledge and understanding of theory and practice for this level. Demonstrates sound understanding of key areas in different contexts.</p>	<p>Presents a logical appraisal of findings through the critical analysis of information. Draws clear and justified conclusions. Demonstrates creativity and originality. Demonstrates a critical understanding of problem-solving approaches and applies effective problem-solving skills.</p>	<p>Provides valid reflection and self-awareness in relation to the outcomes of own work and personal responsibility, as required by the assessment.</p>	<p>Relevant reference sources selected and drawn upon.</p> <p>Most sources accurately cited both the body of text and in the Reference List/Bibliography.</p>	<p>Writing style is mostly appropriate to the requirements of the assessment. Grammar, spelling and punctuation are generally competent and minor lapses do not pose difficulty for the reader. Paragraphs are structured and include some linking and signposting. Sentences are complete. A range of appropriate vocabulary is used.</p>	<p>The presentational style and layout are correct for the type of assignment. Logically structured in the most part. Inclusion of figures, tables and images but not all relevant or referred to.</p>
Class III (Satisfactory Quality) 40% - 49%	<p>Work addresses the requirements of the brief, although superficially in places. Some irrelevant content.</p> <p>Learning outcomes are met.</p>	<p>Demonstrates a sufficient breadth and depth of knowledge and understanding of theory and practice for this level.</p> <p>Demonstrates understanding of a range of key areas in different contexts.</p>	<p>Presents valid appraisal of findings in parts following a critical appraisal of information. Interprets a number of matters and ideas but with some descriptive passages evident which lack clear purpose. Demonstrates creativity and originality in places. Demonstrates a critical understanding of problem-solving approaches and applies sufficient problem-solving skills.</p>	<p>Provides some reflection and self-awareness in relation to the outcomes of own work and personal responsibility, as required by the assessment.</p>	<p>Some relevant reference sources selected and drawn upon.</p> <p>Some weaknesses in referencing technique.</p>	<p>Writing style is occasionally not appropriate for the assessment. Grammar, spelling and punctuation are generally competent, but may pose minor difficulties for the reader. Some paragraphs may lack structure, and there is limited linking and signposting. Some appropriate vocabulary is used.</p>	<p>The presentational style and layout are largely correct for the type of assignment. Adequately structured. Inclusion of some figures, tables and images but not all clear, relevant and/or referred to.</p>
Borderline Fail 35% - 39%	<p>Work addresses some of the requirements of the brief. Irrelevant and superficial content.</p> <p>One or more learning outcomes have not been met.</p>	<p>Demonstrates a lack of knowledge and understanding of theory and practice for this level. Demonstrates insufficient understanding of a range of areas in different contexts.</p>	<p>Presents a descriptive overview of findings following some analysis of information. Some passages lack clear purpose. Demonstrates scant creativity or originality. Demonstrates a lack of critical understanding of problem-solving approaches and insufficient problem-solving skills.</p>	<p>Provides limited reflection and self-awareness in relation to the outcomes of own work and personal responsibility, when required.</p>	<p>Sources selected are limited and/or lack relevance.</p> <p>Poor referencing technique employed.</p>	<p>Writing style is unclear and does not match the requirements of the assessment in question. Deficiencies in spelling, grammar and punctuation makes reading difficult and arguments unclear in places. Paragraphs are poorly structured.</p>	<p>For the type of <u>assignment</u> the presentational style, layout and/or structure are lacking. Inclusion of figures, tables and images but not clear, relevant and/or referred to.</p>
Fail <34%	<p>Work does not address the requirements of the brief. Irrelevant and superficial content.</p> <p>One or more learning outcomes have not been met.</p>	<p>Demonstrates inadequate knowledge and understanding of theory and practice for this level.</p> <p>Demonstrates insufficient understanding of relevant areas.</p>	<p>Presents a descriptive overview of findings with little consideration of the quality of information drawn upon. Demonstrates little or no creativity. Demonstrates a lack of understanding of problem-solving approaches and limited problem-solving skills.</p>	<p>Provides inadequate reflection and self-awareness in relation to the outcomes of own work and personal responsibility, when required.</p>	<p>An absence of academic and relevant sources selected and drawn upon.</p> <p>Poor referencing technique employed.</p>	<p>Writing style is unclear and does not match the requirements of the assessment in question. Deficiencies in spelling, grammar and punctuation makes reading difficult and arguments unclear. Unstructured paragraphs.</p>	<p>For the type of <u>assignment</u> the presentational style, layout and/or structure are lacking. Inclusion of figures, tables and images but not clear, relevant and/or referred to.</p>