

COMP3008

Big Data Analytics

20 CREDIT MODULE

ASSESSMENT: 100% Coursework **W1: 30% Report Outline**
 W2: 70% Report

MODULE LEADER: Dr Lauren Ansell

MODULE AIMS

- To introduce students to the fundamentals of non-relational (NoSQL) databases.
- To critically evaluate the differences between relational and NoSQL databases.
- To gain experience in pre-processing data files adequately for the use of NoSQL databases.
- To gain experience with NoSQL databases through hands-on projects.

ASSESSED LEARNING OUTCOMES (ALO):

1. Critically compare and contrast the differences between relational and non-relational databases.
2. Critically appraise NoSQL database strengths and weaknesses.
3. Demonstrate the ability to perform all the CRUD operations (namely, create, retrieve, update and delete) on NoSQL databases.

Overview

This document contains all the necessary information pertaining to the assessment of COMP3008 *Big Data Analytics*. The module is assessed via **100% coursework**, across two elements: **30% Report outline** and **70% Report**.

The sections that follow will detail the assessment tasks that are to be undertaken. The submission and expected feedback dates are presented in Table 1. All assessments are to be submitted electronically via the respective DLE module pages before the stated deadlines.

	Submission Deadline	Feedback
Report Outline (30%)	3rd March 2026 (15:00)	31st March 2026
Report (70%)	5th May 2026 (15:00)	3rd June 2026

Table 1: Assessment Deadlines

All assessments will be introduced in class to provide further clarity over what is expected and how you can access support and formative feedback prior to submission. Whilst the assessment information is provided at the start of the module, it is not necessarily expected you will start this immediately – as you will often not have sufficient understanding of the topic. The module leader will provide guidance in this respect.

Moderation

This assignment brief has been moderated in line with the university policy.

Moderator Name	Moderation Date
Craig McNeile	09/01/2026

Useful Assignment Information

Please see below some useful information regarding submissions for your modules.

Good Coding Practices

Like all things, no code is perfect. Just because your code compiles and runs does not mean it is perfect, there is always room for improvement. No code will achieve 100% of the available marks. Where code submission is required by a module, it will be assessed against the following criteria:

1) Functionality

- a. Does your code meet all the specified requirements of the assignment?
- b. Does your code behave correctly across typical and edge-case scenarios?
- c. Does your code have appropriate error handling that does not lead to it crashing or undefined behaviour?

2) Efficiency

- a. Is your code optimised in terms of performance and resource use?
- b. Does your code handle functions, variables, data calls efficiently?
- c. Is there any unnecessary repetition, complexity or processing reducing efficiency within your code?

3) Readability

- a. Is your code logically structured and easy to read?
- b. Have you maintained standard formatting conventions like indentations, spacing and naming, within your code?
- c. Are variables, functions, and classes clearly named and purposeful?

4) Documentation

- a. Are there helpful comments explaining non-obvious parts of your code?
- b. Do your comments document your process, development or rationale clearly?
- c. Could someone unfamiliar with the code understand the approach you have taken?

Use of Generative AI for Creating Code

Each assignment element for each module will have a clear indication of the permitted level of use of AI (solo, assisted or partnered) including the generation of code. Please ensure that you **read and understand the permitted use**. If you are unsure of a particular use, please reach out to the Module Leader and ask.

We strongly encourage you to read and explore beyond the core content delivered within the modules. However, this must be done correctly following academic process. Wherever code is drawn from an outside source (e.g. you have not written it yourself), regardless of whether this is AI generated or from an online repository (GitHub, Stack Overflow etc) **you must reference the original source**. This can be done in your documentation as comments or part of your write ups. Students found to be utilising code without claiming could face an academic offence.**If in doubt speak to your Module Leader.**

Versioning

A critical part of a development (software or academic) is versioning. Keeping a number of iterations over the course of development ensures you always have backups to fall back on. It also provides clear demonstration of the development process you implemented.

Using online/cloud-based storage solutions and repositories provide additional peace of mind, alongside being industry practice. The university provide OneDrive to all students which offers inbuilt version history for Microsoft products. GitHub is the university recommended repository for versioning code, which includes great integration with a number of IDEs.

Submitting Code

It is vital that you confirm that all code that you submit is in the **correct format** and **compiles correctly or runs without error**. If you clean up your code before submission, please ensure that all dependencies (libraries, functions and variables) are included so that the marker can compile your code.

We can only mark what has been submitted. If the code does not run correctly, it is not our responsibility to spend time error handling to award you marks.

It is also important to **show your working**. Tidying up your code is a critical part of coding practices, but if you remove the workings that provides you with the required output, we cannot see how you got there.

In a worst-case scenario, if you remove a key part of your working, and on compiling your code it gives an error or different result, we have no way of confirming what went wrong, or indeed if you fabricated those outputs.

Please note, just because a piece of code has given the wrong output, it does not mean it should be deleted. Seeing these attempts with comments documenting your attempt allows us to understand where you may have made an error and provide feedback that addresses this. In some cases, you could also be awarded marks for the attempt.

Assessment 1: Set Exercises

Task:

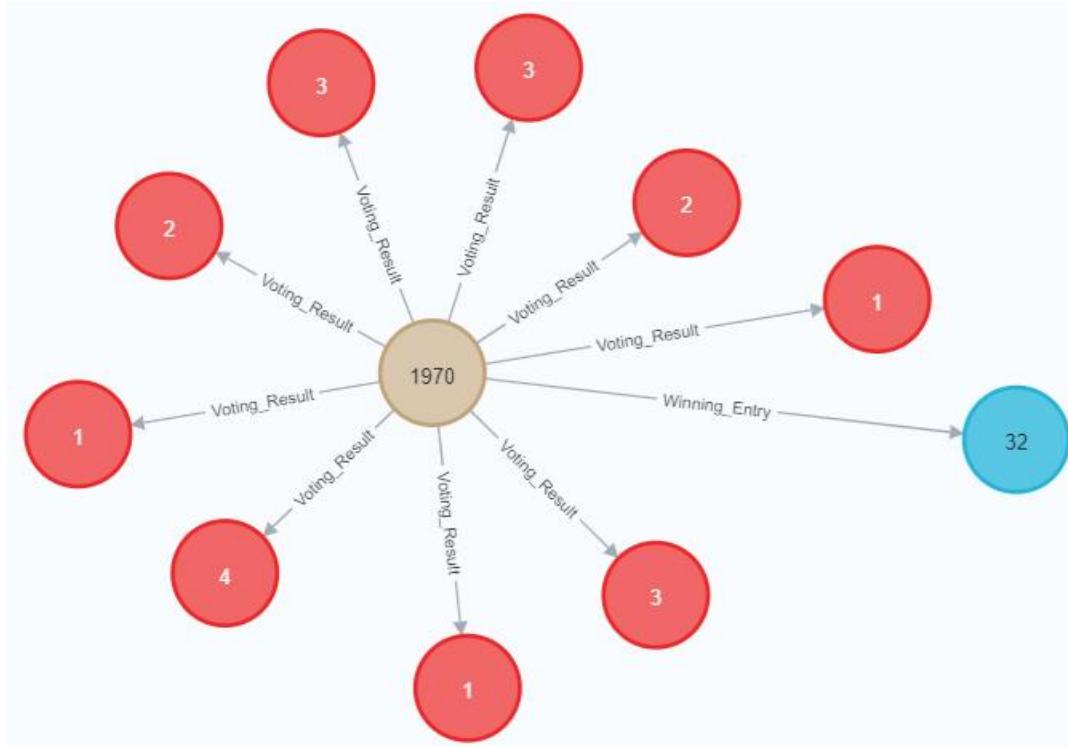
The Eurovision Song Contest is an annual international music competition where countries primarily from Europe (and some beyond) compete by performing original songs live.

You have been provided with three CSV files, one containing the information about location and year of the contest, the winning act from 1956 to 2025 and a file containing the voting results from 1956 to 2025, on the DLE (look for the Coursework section).

Using the data provided, you should carry out the following exercises.

Exercise 1 (15 marks)

Create a Neo4j database to store the data comprised in the CSV files. The database should respect the data model displayed in the example in the figure below, please note that the image below has had some nodes hidden for clarity. You have to provide all the commands needed to create the database and populate it with the data in the CSV files, and you must provide them in the exact order you propose to execute them. If you create indexes, you must also include the commands for index creation. Your database will be recreated, and the only way to do so is by following the commands that you will provide, in the order in which you provide them.



Exercise 2 (5 marks)

Produce a Neo4j query to list all the countries that have won the competition more than twice. The query should include the country and be listed in descending order, that is the country which has won the greatest number of times should be listed first, followed by the country which has the second greatest, and so on. Your answer must show the query followed by the result with the columns appropriately named.

Exercise 3 (5 marks)

Produce a Neo4j query to find all the host countries which then also went on to win the contest that year. The query should list the winning nations, the song they won with and the year in which they won returned in chronological order. You should submit the query and the result with the columns appropriately named.

Exercise 4 (5 marks)

Produce a Neo4j query to identify all the persistent friendships between countries. The query should list both countries and the number of points given. The query result should be listed in alphabetical order by the country giving the points. You should submit the query and the result with the columns appropriately named.

Submission Requirements:

- Query and results submitted at as single TXT document to DLE
- Appendix with completed Generative AI Declaration

Acceptable level of generative AI tool use in this assessment element

The acceptable level of GenAI use for this element is detailed with the allowed uses listed below. This is split into three categories (**Solo Work** – work must be your own with no AI support, **Assisted Work** – some uses of AI tools allowed, **Partnered Work** – AI tools integral part of the work). If you have any questions, please contact the module leader.

Solo Work	You must not use generative AI tools.	<input checked="" type="checkbox"/>
Assisted Work	You are permitted to use generative AI tools in an assistive role.	<input type="checkbox"/>
Partnered Work	Generative AI tool use is required as an integral part of the assessment, but transparency is required.	<input type="checkbox"/>

Table 2: Acceptable Level of generative AI use

Inline with the indicated acceptable level of AI use above, the following uses are acceptable:

- **S1 - Generative AI tools have not been used for this assessment.**

Any use of AI in your work must be declared within your documentation. You **must also include a signed Generative AI Declaration as an appendix to your submission**. The declaration form can be found on DLE (or Here on the Programme Page). This form will not be included in any word count associated with this assignment.

Assessment Criteria:

The following marking scheme will be used for this part of the assessment:

- Creation of database (15 Marks)
- Execution of queries (5 Marks each)

Assessment 2: Report

Task:

You have been provided with data from the UK Annual Population Survey from January 2019 to December 2024 in two separate csv files. The UK government wants to understand how the COVID-19 pandemic has impacted employment patterns and regional economies.

To do this you need to use a range of big data analysis techniques. The report should contain commentary on the results of each of the analysis techniques you apply to the data and your final insights and recommendations, with justification.

Your report should include how the data has been pre-processed, **at least** three different methods for exploratory analysis and **at least** two methods for predictive modelling, where you need to comment on the suitability and accuracy of each of the methods.

REPORT REQUIREMENTS

The report should fulfil the following requirements:

- You should write your report so that it is understandable and accessible, you can assume the reader is familiar with the COMP3008 material and do not wish to see it reiterated in your report.
- Your report should provide details of the methods applied for the predictive modelling.
- Your work on this assignment should involve considerable reading and research from multiple appropriate sources—the Internet is a good place to start, but do not stop there; consider also books and preferably academic articles. Include a reference list, and make sure that your report contains citations to the sources in your reference list and using Harvard style referencing.
- When citing sources, ensure that you provide clear and explicit information about the contents of the cited source, so that the relationship between the contents of the source and the points being made are clear.

Consider the following points before starting your write-up:

The length of the report should be *3500 words (+/-10%)*. Include a word count immediately after your report title—a penalty of 5 marks will be applied for omitting the word count. Reports exceeding the maximum word length will be penalised. Your reference list is not included in the word count. If you produce appendices, these are not included in the word count.

Please note that a significant amount of substantive content is expected: Do not waste words on “conversational”, verbose, or rambling English. Be concise!

You will be writing a report, *not* a journal or conference paper—hence, there is *no* need for an Abstract. You are not writing a dissertation or a book either. Thus, there is *no* need for Terms of Reference; Aims and Objectives; Methodology; Foreword.

Please avoid overly long sentences: 30 words is a long sentence and needs strong use of proper punctuation to be intelligible. Reconsider sentences once they go over 20 words.

Ensure that your writing is clear, and concise. Keep it simple, readable, and clear. Conciseness helps the reader to understand the intended message and will also help you to meet the word limit.

Ensure that your report has a clear structure in terms of numbered sections, sub-sections (if necessary) and paragraphs.

Ensure that any plots included are legible at 100% zoom.

Submission Requirements:

- 3500-word limit
- Report submitted as a single PDF document to DLE
- Appendix with completed Generative AI Declaration

Acceptable level of generative AI tool use in this assessment element

The acceptable level of GenAI use for this element is detailed with the allowed uses listed below. This is split into three categories (**Solo Work** – work must be your own with no AI support, **Assisted Work** – some uses of AI tools allowed, **Partnered Work** – AI tools integral part of the work). If you have any questions, please contact the module leader.

Solo Work	You must not use generative AI tools.	<input type="checkbox"/>
Assisted Work	You are permitted to use generative AI tools in an assistive role.	<input checked="" type="checkbox"/>
Partnered Work	Generative AI tool use is required as an integral part of the assessment, but transparency is required.	<input type="checkbox"/>

Table 2: Acceptable Level of generative AI use

Inline with the indicated acceptable level of AI use above, the following uses are acceptable:

- A1 - Idea Generation and Problem Exploration
- A5 - Language Refinement
- A7 - Code Generation for Learning Purposes

Any use of AI in your work must be declared within your documentation. You **must also include a signed Generative AI Declaration as an appendix to your submission**. The declaration form can be found on DLE (or Here on the Programme Page). This form will not be included in any word count associated with this assignment.

Assessment Criteria:

The following marking scheme will be used for this part of the assessment:

- Literature review (10 Marks)
- Overview of the methods used in report (10 Marks)
- Pre-processing of the data (10 Marks)
- Application of the methods (20 Marks)
- Results (10 Marks)
- Conclusion (5 Marks)
- Reference list (5 Marks)

An illustration of the feedback template is presented in Table 3.

ASSESSMENT CRITERIA

	Criteria	Weighting
1	Is there a properly constructed reference list and are the elements of such a list cited appropriately in the report?	5 marks
2	Is the research carried out appropriate in terms of breadth and depth? The relation between the cited sources and the report are clearly explained.	10 marks
3	Does the report provide an overview of the specific methods (at least 3 exploratory methods and at least 2 predictive methods) used to analyse the data, and is this clearly and explicitly based upon evidence and citations from the literature?	10 marks
4	Does the report explain how the data has been pre-processed and the effects of this?	10 marks
5	Does the report provide a detailed analysis that adequately describes the results of applying the different methods? Is the solution supported by appropriate evidence?	20 marks
6	Does the report propose insights and recommendations? Does the report identify reasons for these? Is this clearly and explicitly supported by evidence? Are such examples presented and discussed in sufficient depth?	10 marks
7	Is there a set of conclusions, and do they provide a reasonable summary – at an appropriate level of abstraction – based upon the contents of the report?	5 marks

Table 3: Report Assessment Marking Criteria

GRADE CRITERIA

When awarding marks, the following guidelines will be employed.

Mark	Grade Criteria
Unprofessional (0-39%)	The quality of the work has not met the learning outcomes. Understanding and application of fundamental concepts and techniques is questionable. Work of this quality would not be acceptable in professional employment.
Poor (40-49%)	The quality of work has only met the threshold level but still requires further work to get it to a better standard. Your submission contains logical and analytical errors related to analysis and design techniques. Also, it only demonstrates a basic understanding of the subject competence. Further improvement is required to demonstrate personal thoroughness, effort and independent learning.
Fair (50-59%)	The quality of work submitted suggests that you have demonstrated a fair understanding of the analysis and design techniques. Still, the work you have submitted contains some errors and incomplete analysis and design. Also, it demonstrates you are able to apply your knowledge but need to improve understanding of the subject competence and personal thoroughness, effort and independent learning.
Good (60-69%)	The quality of the work submitted suggests that you are able to apply the analysis and design techniques well. The work you have submitted is substantially correct and complete. Also, it demonstrates a good understanding of subject competence and personal thoroughness, effort and independent learning.
Excellent (70-100%)	The quality of work is outstanding with no significant flaws. It demonstrates a high level of subject knowledge and competence; personal thoroughness, effort and independent learning; and possibly significant additional analytical/critical thought. Well done!

Acceptable levels of AI use:

The table below provides the acceptable use categories for GenAI. Each assessment element may allow different uses. Please check the brief for each element carefully to see what uses are allowed.

Solo Work	S1 - Generative AI tools have not been used for this assessment.
	A1 – Idea Generation and Problem Exploration Used to generate project ideas, explore different approaches to solving a problem, or suggest features for software or systems. Students must critically assess AI-generated suggestions and ensure their own intellectual contributions are central.
	A2 - Planning & Structuring Projects AI may help outline the structure of reports, documentation and projects. The final structure and implementation must be the student's own work.
	A3 – Code Architecture AI tools maybe used to help outline code architecture (e.g. suggesting class hierarchies or module breakdowns). The final code structure must be the student's own work.
	A4 – Research Assistance Used to locate and summarise relevant articles, academic papers, technical documentation, or online resources (e.g. Stack Overflow, GitHub discussions). The interpretation and integration of research into the assignment remain the student's responsibility.
Assisted Work	A5 - Language Refinement Used to check grammar, refine language, improve sentence structure in documentation not code. AI should be used only to provide suggestions for improvement. Students must ensure that the documentation accurately reflects the code and is technically correct.
	A6 – Code Review AI tools can be used to check comments within the code and to suggest improvements to code readability, structure or syntax. AI should be used only to provide suggestions for improvement. Students must ensure that the code accurately reflects their knowledge and is technically correct.
	A7 - Code Generation for Learning Purposes Used to generate example code snippets to understand syntax, explore alternative implementations, or learn new programming paradigms. Students must not submit AI-generated code as their own and must be able to explain how it works.
	A8 - Technical Guidance & Debugging Support AI tools can be used to explain algorithms, programming concepts, or debugging strategies. Students may also help interpret error messages or suggest possible fixes. However, students must write, test, and debug their own code independently and understand all solutions submitted.

	<p>A9 - Testing and Validation Support AI may assist in generating test cases, validating outputs, or suggesting edge cases for software testing. Students are responsible for designing comprehensive test plans and interpreting test results.</p>
	<p>A10 - Data Analysis and Visualization Guidance AI tools can help suggest ways to analyse datasets or visualize results (e.g. recommending chart types or statistical methods). Students must perform the analysis themselves and understand the implications of the results.</p>
	<p>A11 - Other uses not listed above Please specify:</p>
Partnered Work	<p>P1 - Generative AI tool usage has been used integrally for this assessment Students can adopt approaches that are compliant with instructions in the assessment brief. Please Specify:</p>

General Guidance

Extenuating Circumstances

There may be a time during this module where you experience a serious situation which has a significant impact on your ability to complete the assessments. The definition of these can be found in the University Policy on Extenuating Circumstances here:

<https://www.plymouth.ac.uk/student-life/your-studies/essential-information/exams/exam-rules-and-regulations/extenuating-circumstances>

Plagiarism

All of your work must be of your own words. You must use references for your sources, however you acquire them. Where you wish to use quotations, these must be a very minor part of your overall work.

To copy another person's work is viewed as plagiarism and is not allowed. Any issues of plagiarism and any form of academic dishonesty are treated very seriously. All your work must be your own and other sources must be identified as being theirs, not yours. The copying of another persons' work could result in a penalty being invoked.

Further information on plagiarism policy can be found here:

Plagiarism: <https://www.plymouth.ac.uk/student-life/your-studies/essential-information/regulations/plagiarism>

Examination Offences: <https://www.plymouth.ac.uk/student-life/your-studies/essential-information/exams/exam-rules-and-regulations/examination-offences>

Turnitin (<http://www.turnitinuk.com/>) is an Internet-based 'originality checking tool' which allows documents to be compared with content on the Internet, in journals and in an archive of previously submitted works. It can help to detect unintentional or deliberate plagiarism.

It is a formative tool that makes it easy for students to review their citations and referencing as an aid to learning good academic practice. Turnitin produces an 'originality report' to help guide you. To learn more about Turnitin go to:

<https://help.turnitin.com/new-links.htm?Highlight=guide>

Referencing

The University of Plymouth Library has produced an online support referencing guide which is available here: <http://plymouth.libguides.com/referencing>.

Another recommended referencing resource is [Cite Them Right Online](#); this is an online resource which provides you with specific guidance about how to reference lots of different types of materials.