

<b>Module code &amp; title:</b>	COM7064: Programming for Data Science	<b>Module leader:</b>	Dr Nicholas Day <a href="mailto:Nicholas.Day@bnu.ac.uk">Nicholas.Day@bnu.ac.uk</a>
<b>Assignment No. and type:</b>	CW1 Programming Assignment	Assessment weighting:	100%
<b>Submission time and date:</b>	Week 10: Friday 12th December 2025 2pm	Target feedback time and date:	Friday 16th January 2026 2pm (Inclusive of Christmas Closure)

Assignment task

This assignment will simulate the initial stages of a data investigation – from data collection to preparatory steps, to exploration of the data (descriptive statistics, further statistics, and visualisation of data analysis) and even modelling and forecasting. You are required to select a research question to ask of a dataset or scenario and utilise either Python or R to produce the pipeline which takes the raw data to analysis. If adopting a statistical R-studio route – form hypothesis as part of your questioning of the data.

Follow a structured process, which could be CRISP-DM (or BizML), and document the development journey. You are required to write a 1000-word report that should outline the scenario, data collection, exploratory data analysis (EDA), rationale for your choice of descriptive statistical tests, and their results, and visualisation. You should utilise established metrics for statistical tests and also cite relevant published research where relevant to background context or the results you have produced.

You are advised to document these stages of the approach in one Jupyter Notebook file (.ipynb) using both markdown and code cells. You could write your 1000-word report as a markdown cell of this Notebook. You can upload either the Jupyter Notebook file (.ipynb) or the converted pdf (see guidance on Blackboard – prefer options 'print' or 'print preview') and then upload this to Turnitin.

Alternatively, you could write your 1000-word report in Microsoft Word and submit this to Turnitin. You would need to include your Python code/R Studio code in this Word document.

Marking Criteria:

- Introduction – data collection and outline the problem (20 marks)
- Preparation and Exploratory Data Analysis (EDA) on the dataset (40 marks)
- Evaluation of the Data Story – the results in context (40 marks)
- verify that your work is your own, you will be asked to provide an in-class demonstration and explanation of your project.

If you are registered with the Disability and Inclusion Services or have a disability, please contact your module leader to discuss how your reasonable adjustments will be applied to this assignment.

Getting Support for Your Assignment

Your first point of advice should be the person setting the assignment or the module leader.

If you need more general support with your studies (e.g. academic writing, referencing, critical thinking) then the [Student Learning and Achievement Team](#) can support development of your academic skills.

This assignment has been designed to provide you with an opportunity to demonstrate your achievement of the following module learning outcomes:

LO1: Determine the appropriate statistical method for analysing a data set, based on an objective assessment of options, context and need.

LO2: Apply programming techniques to prepare, analyse and present data.

LO3: Evaluate the results achieved in terms of accuracy, limitations and usefulness of the data story.

Practicalities: Referencing, presenting and submitting your work

You must submit the source code to the Turnitin submission point on Blackboard (either as a Notebook file or a converted PDF). You are advised to document these stages of the approach in one Jupyter Notebook file (.ipynb) using both markdown and code cells. You can upload either the Jupyter Notebook file (.ipynb) or the converted pdf (see the guide on Blackboard – prefer options to 'print' or 'print preview') and then upload this to Turnitin. Alternatively, you could write your 1000-word report in Microsoft Word and submit this to Turnitin. You would need to include your Python code in this Word document.

Please ensure that your work is referenced in Harvard Referencing style. Guidance on using this style correctly is available at <https://www.citethemrightonline.com>

Confidentiality

You should maintain and respect confidentiality in relation to the protection of personal, technical and/or commercial information of a sensitive nature in your assessed work.

For further information and guidance, please see the University [Academic Confidentiality guidance](#).

Academic integrity

Academic integrity means taking responsibility for your own work and not acting in a way to give you an unfair advantage over another learner. This includes writing in your own words and acknowledging sources you have used. It also covers the appropriate use of Artificial Intelligence.

When you submit an assignment, you are making a declaration that it is your own work and that you have acknowledged the contribution of others and their ideas in its development (for example, by referencing them appropriately).

For further information, please see the following student facing guidance on the University website:

[BNU Student Academic Integrity Guidance](#)

[BNU Artificial Intelligence Guidance for Students](#)

There are also links to resources to support your understanding of Academic Integrity within your Blackboard module sites in the Assessment Information area.

Categorical Mark and Grade	0,1,10,20,32,35,38% (F)	42,45,48% (D)	52,55,58% (C)	62,65,68% (B)	72, 75,78% (A)	82, 85, 88, 92, 95, 98, 100% (A+)
Pass status	Fail Learning Outcomes have not been met	Marginal Fail Learning Outcomes have not been met	Pass - Satisfactory Learning Outcomes have been met	Pass – Good/Very Good Learning Outcomes have been exceeded	Pass – Excellent Learning Outcomes have been exceeded	Pass – Outstanding Learning Outcomes have been exceeded
Criterion 1 (LO1) Determine the appropriate statistical method for analysing a data set, based on an objective assessment of options, context and need.	Either a non-submission or a lack of understanding of data science concepts. Please arrange an appointment with your tutor.	An attempt is made but there is a lack of understanding of data science concepts. Please arrange an appointment with your tutor.	The fundamental concepts have been understood and satisfactory evidence presented.	A good level of comprehension is demonstrated, but there may be one or two concepts that are not fully understood.	A very strong comprehension of most data science concepts is demonstrated. There may be some minor inconsistencies or misconceptions with a concept.	An excellent comprehension of data science concepts is demonstrated, which will facilitate the learner's ability to write solutions to problems.
Criterion 2 (LO2) Apply programming techniques to prepare, analyse and present data.	The solution (if one submitted at all) does not fulfil the requirements and the code does not compile into a functioning program. There is no evidence that students have investigated why the program does not compile/run.	There is evidence that an attempt has been made, but the source code does not compile and/or does not fulfil the basic requirements of the application. There is minimum to no evidence that the solution has been tested.	The solution does compile and create a functioning program that fulfils the basic requirements, and there is evidence that these basic features have been tested.	The solution compiles fulfils most (not all) of the stated requirements. The implementation of these features is supported by appropriate test evidence.	The solution compiles, works without exception and fulfils all the requirements and evidence is presented of these features being fully tested.	The solution compiles, works without exception and fulfils all the requirements and evidence is presented of these features being fully tested. Furthermore, this work goes beyond the requirements expected and truly 'stands out'.
Criterion 3 (LO3) Evaluate the results achieved in terms of accuracy, limitations and usefulness of the data story.	Either a non-submission or a lack of understanding of data science concepts. Please arrange an appointment with your tutor.	An attempt is made but there is a lack of understanding of data science concepts. Please arrange an appointment with your tutor.	The fundamental concepts have been understood and satisfactory evidence presented.	A good level of comprehension is demonstrated, but there may be one or two concepts that are not fully understood.	A very strong comprehension of most data science concepts is demonstrated. There may be some minor inconsistencies or misconceptions with a concept.	An excellent comprehension of data science concepts is demonstrated and applied to the development of the final solution.