BCS LEVEL 4 DIGITAL MODULAR PROGRAMME IN DATA ANALYSIS

LEARNER GUIDE TO THE PROJECT

bcs
The Chartered Institute for IT

April 2022 V1.5
Unit Code: H/616/8041

Are you ready to showcase your skills?

Well done for all of your hard work so far on the Digital Modular Programme (DMP). It's now time to begin planning and preparing your synoptic project. This is your opportunity to apply everything you have learned so far, and we want to make sure that you are as prepared as possible.

Within this document, you will find information about the requirements and format for this part of the assessment for the Level 4 Digital Modular Programme in Data Analysis.

NB: This synoptic project is part of the mandatory assessment strategy for all qualifications in the DMP suite of qualifications. This project **will not be required** by learners accessing any element of the DMP as part of an Apprenticeship.

Overview of the project

Working in a project-based environment, data analysts collate, analyse and present data in order to help inform business decisions.

This project will enable you to demonstrate your analytical skills, your use of varied methods of presenting data analysis outputs, and your ability to distil key data analysis findings into a presentation. It will require you to draw upon experience gained during your day-to-day work and design a project to address a data analysis task that is relevant to your role, although an alternative approach is available for learners who are not currently working in a data analyst role.

What should I focus on within my project?

The project should address a specific business problem, recurring issue or idea/opportunity. There are two options available for how you may undertake this project:

1. **A real-world business solution.** Your choice of project will focus on a real work situation relevant to your current job role and where you have identified an opportunity to undertake data analysis to provide specific information and insight that can support a business decision to be made.

   Whilst the basis for the project will be data available through your role, you are encouraged to also utilise other open source data freely available for use online, such as government published data, market data or the datasets provided by BCS where these could be relevant.

   As part of the project you will be expected to manipulate, analyse and model data. You should ensure that the data is of substantial size and include time-series based elements for the purposes of prediction and forecasting. The data sets you use should **not** be pre-cleaned as it is important to demonstrate how you allow for any errors, duplicates or missing values during the process you undertake.

   You can also source datasets outside of your workplace to further your analysis and detail any assumptions you make.

2. **A solution to a given scenario**. You can undertake your project using a given scenario, data and set of requirements provided by BCS. You will be required to execute a process of refining the requirements through discussions with other stakeholders and in this situation you should simulate this process using people available to you, such as your tutor.

   To ensure that you cover all the required criteria you will have to manipulate multiple data sets including, but not limited to, the provided sets. You are encouraged to research and justify use of further data sets to link to and enhance your analysis.

   To undertake this project using the scenario provided by BCS, you will be given access to the relevant files in advance. These files will include the scenario, high-level requirements and source data provided in a CSV format.

Whichever option you chose, it should:

- Clearly state the business situation and modelling requirements.
- Include a summary view of the initial data (redaction of specific confidential items is acceptable).
- Forecast future data points over a specified time-scale appropriate to your business.
- Enable the data to be visualised easily by non-technical audiences and senior decision makers.
- The model should be reusable so that it can be use with other datasets.

You will need to produce:

- A clear requirements statement based on stakeholder discussions.
- Details of the processes used for migration and collation of data.
- A design and documentation for an appropriate statistical model using a tool of your choice.
- A summary and visualisation of the model output.
- A stakeholder presentation of the results and learnings.

The project must map, in an appendix:

- How it evidences the relevant Knowledge, Skills and Behaviours for this assessment method (as described on pages 12-13).
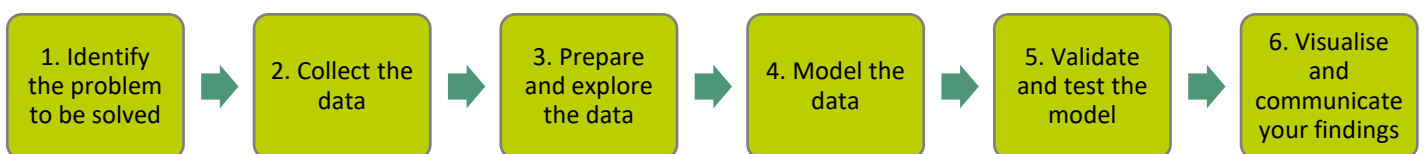- Where you have provided justifications for any decisions made during the project.

You should produce a presentation that includes the outputs of your analysis and recommendations that could then be delivered to the relevant decision makers. Please also use this as an opportunity to evaluate your own skills and identify areas for personal improvement.

We highly recommend that a mentor/tutor supports you throughout your project. When you meet with them you will complete a log of your meetings to track the progress and feedback of your project where each party signs and dates to allow the log to be submitted alongside your project. This will enable you to capture any feedback given to you by your mentor/tutor on your individual performance, how you have worked with others, and provide opportunity for you to record how you have overcome any challenges during each stage of the project.

It is important that you aim to draw conclusions that can form the basis of a presentation report. This report should be designed to enable a group of stakeholders to make a particular business decision – so it's important that you consider your target audience.

How should I structure my work?

The project will follow the same structure whether you choose to focus on a real-world solution or a given scenario. To ensure you meet the requirements of this assessment, the files you submit for your project should reference the Data Analysis Lifecycle throughout. You should set out your project as follows:

1. Identify the problem to be solved	→	2. Collect the data	→	3. Prepare and explore the data	→	4. Model the data	→	5. Validate and test the model	→	6. Visualise and communicate your findings

Top tip: These elements are not equally weighted – this means that you would not be expected to spend the same amount of time on or produce as much for each of the six sections.

The guidance provided on page 6 onwards clearly identifies the scope of criteria you need to include within your work for each of the section of your project.

How long should my project be?

The project must represent a substantial piece of work; as such, the suggested time requirement for this part of the assessment is 30 hours. These hours are included in the Total Qualification Time (TQT) of 407 hours. You do not have to be supervised when you complete all the project work but you may wish to have your tutor available for any questions.

As a minimum all project outputs must include:
- an introduction
- the scope of the project (including key performance indicators)
- project outcomes and how the outcomes were achieved
- a project plan
- consideration of legislation, regulation, industry and organisational policies, procedures and requirements
- analysis
- research and findings
- recommendations and conclusions

A word limit of **3500 words** applies to your project, with a 10% tolerance either way. Any additional reports, surveys or supporting documentation can be included as an annex, which would not contribute to the overall word count. If the project exceeds the tolerance allowed, then the project will be returned unmarked and a restructure before the second submission would be requested.

How will I submit my work?

Your project must be submitted electronically and contain notes and guidance for the assessor so that they can easily identify where you have met the learning objectives and assessment criteria in your work. The final submission must be in a format that is straightforward to access, so that your work can be assessed against the stated criteria – these might include PowerPoint, Word, PDF or Excel.

What will I need to demonstrate to pass this assessment?

The following table outlines the tasks you will be required to undertake for each stage of your project, the pass and distinction criteria used for this assessment, with reference to the knowledge, skills and behaviours (KSBs) you will be able to demonstrate by undertaking this project. These KSBs are listed in full in the Occupational Standard section of this document. The syllabus criteria has been provided to show you how the learning undertaken through completion of the Digital Core module and your occupationally focussed module will support you to meet these criteria.

---

1. Identify the problem to be solved

You will describe the problem, providing a clear explanation for how data analysis will enable the problem to be solved. You should outline the approach you will take to undertake data analysis, detailing each of the steps you will undertake. You will be required to:

    a. Provide an overview of the project you will undertake;
        i. identifying the audience and the outputs required (K7, K9, B2, B4)
        ii. identifying the scope of the work to be completed (K3, S5)
        iii. identifying the stakeholders with whom you will be working (S12, B3) *If you are not within a work context this may be a teacher*

    b. Develop a plan for how you will undertake each stage of the data analysis lifecycle (S2,S7,B5)

Pass criteria	Distinction criteria	Reference to learning
You can demonstrate that you are able to:	You can demonstrate that you are able to:	The learning towards this criteria can be supported within the following syllabus areas:
Identify and describe the problem to be solved and explain how data analysis can present a solution. Describe the intended audience for this project, listing their specific requirements for the data analysis project.	Evaluate the key characteristics of the audience and their requirements, highlighting the most important characteristics that will influence the choice of solution to the problem. Justify the key priorities and dependencies within each stage, evaluating any risks to the overall plan.	BCS Level 4 Module in Digital Core: 1.5, 2.2 BCS Level 4 Module in Data Analysis:

Produce a plan for how the problem will be solved, identifying key priorities and making clear and appropriate reference to the data analysis lifecycle. Describe the stakeholders required to undertake the project including their role and responsibility.		3.1, 3.2, 3.5, 3.6, 6.1-6.6

2. Collect the data

You should provide evidence towards how you have collected the data. This includes detail of the specific extra data sources you have used and the means by which you have collected the data. You should include reference to any sources you have used if using open data sources, with consideration towards any terms of use. You will be required to:

a. Identify the sources of the data to be used. You should detail any requirements for security and licensing for each set of data (K1, K2, K4, K15).
b. Identify potential risks and issues with collecting and using the data and how you will manage these (K2, K8, S8, B5). The provided datasets have limitations in their quality that should be identified *(Note – you do not need to do this for all of the given data, only the elements you intend to use in your analysis.)* You may include a copy of a risk and issues log in your annex.
c. Collect and migrate the data as identified, following relevant standards, policies and procedures. (K1, S1)

Pass criteria	Distinction criteria	Reference to learning
Identify the most suitable sources of data required to solve the problem, describing the means by which you will collect the data. Explain how the data will be sourced securely, describing which legal requirements and organisational policies and procedures you will need to adhere to. Identify potential risks and issues with collecting and using the data, explaining how you will manage these.	Compare multiple data sources, justifying the choice of data used for the project. Analyse the legal and ethical impact of collecting and using the data and the potential implications for stakeholders and the business.	BCS Level 4 Module in Data Analysis: 5.1, 5.2, 6.1

Demonstrate the process by which the data has been successfully collected and migrated for use in the analysis process.		

3. Prepare and explore the data

You should outline the process you have taken to prepare the data for analysis, discussing how you have addressed any errors, omissions or duplication in the data. You should provide evidence of migration of the data from the sources to a relational database that you have designed. This should include examples of specific design documentation that you have used in preparing the data. You will be required to:

a. Classify the data from all the sources (S3).
b. Identify the data structures and formats of the data (S4, K6).
c. Prepare and analyse the data in logical stages and inline with the project definition/scope (S10, S14, B4, S5, S15).
d. Migrate the data from the sources using a relevant tool (S3, S4, S8, K6, K10, K11).
e. Log any issues and obstacles or problems with the analysis of the data and how they were resolved (B2, B5, B6, B7, S6). You should detail the steps taken to cleanse the data. You may also include a copy of a risk and issues log in your annex.

Pass criteria	Distinction criteria	Reference to learning
Explain how you have correctly classified the data structures of the data you are working with. Demonstrate the process by which you have successfully cleansed and migrated the data using a relevant tool. Explain any issues with the data or problems encountered during preparation and exploratory analysis, and the steps you have taken to overcome these challenges.	Evaluate and justify your chosen data cleansing processes inline with the project scope and the needs of the intended audience. Evaluate the reproducibility of the data cleansing process used and it can be used in future data analysis activities.	BCS Level 4 Module in Data Analysis: 2.1, 2.3, 5.3, 4.3, 4.7, 5.4, 6.1, 6.3

4. Model the data

You should research applicable models that are relevant to the context of the organisation or scenario to design, develop and execute a modelling exercise to identify and predict trends within the data. The output of your modelling should include relevant visualisations. You will be required to:

    a.   Research applicable models in relation to the domain context and requirements of the audience. (S7, S12)
    b.   Using appropriate models, you should create visualisations of the trends and patterns in the data. (B4, S7, S10, S11, S13, S14, S15, K13, K14)
    c.   Building upon the trends you should use appropriate techniques to forecast/predict future trends. (B4, S13)

Pass criteria	Distinction criteria	Reference to learning
Explain the choice of modelling techniques used. Apply appropriate statistical techniques to forecast/predict trends from the data. Demonstrate your ability to model the data and produce a visualisation of the results.	Justify the choice of data modelling techniques including comparison to alternative methods. Evaluate the outputs of the statistical techniques used listing the strengths and weaknesses of the model.	BCS Level 4 Module in Data Analysis: 4.4, 4.6, 5.1, 5.2, 6.1 6.3, 6.4

5. Validate and test the model

You should provide evidence of how you have tested the model, the process you have followed and your observations when testing. You should explain your approach to fixing or addressing any issues encountered whilst testing. You will be required to:

a. Identify validation checks that you have or will carry out. (B4),
b. Carry out the testing of your model in an appropriate manner, documenting and resolving any issues encountered. (B5, S15) You may wish to include a copy of a testing log or any other additional information when testing in your annex.

Pass criteria	Distinction criteria	Reference to learning
Demonstrate and explain the process through which you have successfully validated and tested the model. Explain how you have resolved any issues encountered during testing of your model.	Evaluate how well the model fits the data and its suitability for modelling from future/alternative sources.	BCS Level 4 Module in Data Analysis: 5.1-5.4, 6.1, 6.5

6. Visualise and communicate your findings

You should provide a summary of your findings, drawing conclusions that will inform on the business decision making process. You should include copies of reports that have been created to share with key stakeholders. These should include visualisations of the data included in the report such as graphs, charts or screenshots of dashboards created.

       a. Prepare the outputs for your identified audience in line with the original plan. (S5, S12, S14, K7, B3)
       b. Review the outputs of the work and your own performance, reflecting on the experience of carrying out the analysis and any lessons learnt. (B3, B6)

Pass criteria	Distinction criteria	Reference to learning
Present the outputs of your analysis, explaining in a clear and comprehensible manner how it solves the problem. Identify and list potential areas for improvement through an evaluation and review of the outputs of your work and your performance, identifying any lessons learnt.	Justify the approach taken to deliver the recommended outcome with rationale for the conclusions drawn and any omissions.	BCS Level 4 Module in Digital Core: 1.3 2.1, 5.1, 7.4 BCS Level 4 Module in Data Analysis: 2.5, 3.6, 6.6

Occupational standard

The following table lists each of the Knowledge, Skills and Behaviours as defined within the Data Analyst occupational standard as referenced in the sections above.

Knowledge
K1: current relevant legislation and its application to the safe use of data
K2: organisational data and information security standards, policies and procedures relevant to data management activities
K3: principles of the data life cycle and the steps involved in carrying out routine data analysis tasks
K4: principles of data, including open and public data, administrative data, and research data
K5: the differences between structured and unstructured data
K6: the fundamentals of data structures, database system design, implementation and maintenance
K7: principles of user experience and domain context for data analytics
K8: quality risks inherent in data and how to mitigate or resolve these
K9: principal approaches to defining customer requirements for data analysis
K10: approaches to combining data from different sources
K11: approaches to organisational tools and methods for data analysis
K12: organisational data architecture
K13: principles of statistics for analysing datasets
K14: the principles of descriptive, predictive and prescriptive analytics
K15: the ethical aspects associated with the use and collation of data
Skills
S1: Use data systems securely to meet requirements and in line with organisational procedures and legislation including principles of Privacy by Design
S2: implement the stages of the data analysis lifecycle
S3: apply principles of data classification within data analysis activity
S4: analyse data sets taking account of different data structures and database designs
S5: assess the impact on user experience and domain context on data analysis activity
S6: identify and escalate quality risks in data analysis with suggested mitigation or resolutions as appropriate
S7: undertake customer requirements analysis and implement findings in data analytics planning and outputs
S8: identify data sources and the risks and challenges to combination within data analysis activity
S9: apply organizational architecture requirements to data analysis activities
S10: apply statistical methodologies to data analysis tasks
S11: apply predictive analytics in the collation and use of data
S12: collaborate and communicate with a range of internal and external stakeholders using appropriate styles and behaviours to suit the audience
S13: use a range of analytical techniques such as data mining, time series forecasting and modelling techniques to identify and predict trends and patterns in data
S14: collate and interpret qualitative and quantitative data and convert into infographics, reports, tables, dashboards and graphs
S15: select and apply the most appropriate data tools to achieve the optimum outcome
Behaviours
B1: maintain a productive, professional and secure working environment
B2: show initiative, being resourceful when faced with a problem and taking responsibility for solving problems within their own remit

B3: work independently and collaboratively
B4: logical and analytical
B5: identify issues quickly, investigating and solving complex problems and applying appropriate solutions. Ensures the true root cause of any problem is found and a solution is identified which prevents recurrence.
B6: resilient - viewing obstacles as challenges and learning from failure.
B7: adaptable to changing contexts within the scope of a project, direction of the organisation or Data Analyst role.

Top tips you help you prepare

Finally, here are our key pieces of advice for preparing your project:

1. Plan your time wisely – e.g. you should approach this piece of work as if it were a real-life work project.
2. Document what you do – e.g. things will fail, but this does not mean that they are not worth writing about. For example, you might like to discuss why they failed? What did you learn? How did this have an impact on your next steps?
3. Use tools that you are familiar with – e.g. do not be tempted to use new or unfamiliar tools for the project, as this could waste time.
4. Don't be afraid to make assumptions – e.g. data will not be perfect. Do not be tempted to cleanse it beforehand, as this would leave a large chunk of work undocumented and you may not be able to evidence many of the key behaviours above.
5. Show your thought process – e.g. think about how you have used logic or problem-solving techniques to approach and break down a problem. Make sure you document this!

You can find more information in the Data Analyst syllabus or the DMP Qualification Guide.

Good luck!