

BPP Business School

Coursework Cover Sheet

Please use this document as the cover sheet of for the 1st page of your assessment.

Please complete the below table – the grey columns

Module Name	Programming for Data Analysts (PDA)
Programme Name	MSc Management with Data Analytics
Student Reference Number (SRN)	<Your SRN>
Assessment Title	<Your assessment title> CW10[S]

Please complete the yellow sections in the below declaration :

<p><u>Declaration of Original Work:</u></p> <p>I hereby declare that I have read and understood BPP's regulations on plagiarism and that this is my original work, researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Business and Technology.</p> <p>The word count, excluding contents table, bibliography and appendices, is words.</p> <p>Student Reference Number: Date: </p>	
---	--

By submitting this coursework you agree to all rules and regulations of BPP regarding assessments and awards for programmes.

Please note that by submitting this assessment you are declaring that you are fit to sit this assessment.

BPP University reserves the right to use all submitted work for educational purposes and may request that work be published for a wider audience.

MSc Management with Data Analytics

Programming for Data Analysts

Coursework Assessment Brief SUMMATIVE CW10[S]

1. General Assessment Guidance

- Your summative assessment for this module is made up of this coursework submission which accounts for 100% of the marks
- **Please note late submissions will not be marked.**
- You are required to submit all elements of your assessment via **Turnitin online access**. Only submissions made via the specified mode will be accepted and hard copies or any other digital form of submissions (like via email or pen drive etc.) **will not be accepted**.
- For coursework, the submission word limit is **2,500** words. You must comply with the word count guidelines. You may submit LESS than 2,500 words but not more. **Word Count guidelines can be found on your programme home page and the coursework submission page.**
- **Do not put your name or contact details anywhere on your submission**. You should only put your student registration number (SRN) which will ensure your submission is recognised in the marking process.
- A total of 100 marks are available for this module assessment, and you are required to achieve minimum **50%** to **pass** this module.
- You are required to use only Harvard Referencing System in your submission. Any content which is already published by other author(s) and is not referenced will be considered as a case of plagiarism.
You can find further information on Harvard Referencing in the online library on the VLE. You can use the following link to access this information: <http://bpp.libguides.com/Home/StudySupport>
- BPP University has a strict policy regarding authenticity of assessments. In proven instances of plagiarism or collusion, severe punishment will be imposed on offenders. You are advised to read the rules and regulations regarding plagiarism and collusion in the GARs and MOPP which are available on VLE in the Academic registry section.
- You should include a completed copy of the **Assignment Cover sheet**. Any submission without this completed Assignment Cover sheet may be considered invalid and not marked.

2. SUMMATIVE Assessment Brief

2.1. Assessment learning outcomes

This assessment is designed to gauge your understanding, skills and application of common data analysis techniques used in business and other organisations today. As such you need to demonstrate your attainment in these areas according to the THREE Module Learning Outcomes (LOs):

LO 1: Critically evaluate the principles of programming and apply them in a business context.

LO 2: Critically evaluate the use of code libraries in programming for a business context.

LO 3: Construct a programming solution to solve a defined business problem.

2.2. Assessment tasks

This assessment is made up of TWO Parts

- Part 1 - A business report.
- Part 2 - A coding exercise in data analysis using a Python notebook.

You will have worked on both these Parts for your Formative Assessment. Now update both Parts for your Summative Assessment as set out below. You should act on any feedback you received on your Formative submission, together with your own further learning and development across the module.

2.3. Scenario

HIRAD Financial Services (HFS) is a local company that provides small business loans. Last year, loan applications increased by over 200%, largely because of a concerted online campaign to establish a strong digital presence. Almost all loan applications and business leads are generated from search engines and digital advertisements, reflecting the decision to increase advertising spend on [SEO \(Search Engine Optimisation\)](#) channels such as Google, Facebook, LinkedIn and similar platforms.

Despite a strong digital marketing approach, the current loan application process remains manual. It requires the online completion of information, including gender, marital status, number of dependents, education, income etc. Loan decisions are manually categorised as either “approved” or “rejected.” To date, several of these factors have been considered in the approval decision. All applications are reviewed and approved by the loan team which, given the recent increase in volumes, has resulted in skills shortages, longer loan approval times and increased potential operational and control risk. The current operating model constrains further growth.

You are employed by HFS as a Data Team Manager, with programming and data analytics knowledge, as well as a deep appreciation for the need to balance business growth with a robust operational control environment. You will be leading this project with a team of programmers, with various levels of experience, and have been tasked with providing a scalable solution – that addresses key resourcing and control risks.

Specifically, the Board has instructed you to develop and maintain several partial automation processes that will help the existing loans team, freeing up their time for greater one-on-one customer contact. You need to provide a data-driven solution while working with a variety of key stakeholders each with varying objectives such as marketing, internal audit and compliance. You need to assure the board that you have consulted widely in drawing up requirements and have considered how any code produced will be further developed and maintained. You also need to consider that any code developed will need to be accessible to (or used by) other staff such as sales, customer service and support personnel.

An in-house database administrator (DBA) was able to compile a PDF of past applications which the loans team are hoping to map to previous loan approval outcomes.

The two files provided by the DBA are:

- A file in PDF format called 'HIRAD_Loans_Database_Table.pdf'
- An Excel file, called 'HIRAD Loans Data.xlsx'

(Note: These same two files are used in the Formative and the Summative).

The first file has been extracted from business loan records from the previous year, and it includes a loan approval status field (Yes/No) for each application, allowing the business to map inputs to outcomes for a possible supervised machine learning exercise.

The Excel file is maintained by the Sales team, and it is currently being saved in a shared folder. This increases the chance of duplication and missing values.

You will need to reflect the learnings throughout this module and consider the learning outcomes particularly LO 3: Construct a programming solution to solve a defined business problem as you create your answer.

Report - Business Proposal (LO1, LO2)

Using the scenario given, develop a report that proposes and justifies the continued development and maintenance of the code that has been written by the programming team.

Task 1 (LO1 and LO2) Business Report:

- Identify the problem issues in the given scenario.
- Discuss requirements, challenges, and data quality of the data sets.
- Explain what steps should be taken to implement the proposed solution.
- Your **Approach** should include **cleaning** and **preparation, design, test, and maintenance**.
- State the potential benefits of your approach and the implications of not addressing the problem using programming principles (e.g. OOP, performance optimisation, version control, code maintenance, etc)
- Provide clear recommendations based on your insights to enable automation of the loan approval.

Programming Solution (LO3)

Task 2 (LO3): Programming Solution (this is not included in the word count)

- Load both data files and combine them.
- Explain the steps taken to load the data, how you treated the PDF data, how you cleaned the data, the libraries called and the overall procedure.
- Show outputs from the **EDA** (exploratory data analysis).

Report Structure and References

Follow the guidelines given in Section 3 Report Structure and Referencing

Business Report Guidance notes:

The format of the report should include:

TASK 1 – Business Report (LO1 and LO2):

(2500 words)

First set out the purpose of the report and its structure. It must then cover the current business environment of companies like HFS, the problems that the programming solution has addressed, and what impact and benefits your proposed development and maintenance of the code might have on the business. You should also mention the implications of not proceeding and the kind of human resources needed. **There is no need to discuss project finances**, but your introduction must comment on the following:

- what **user requirements** have been considered (what tools will be used);
- the source and nature of the **supplied dataset** and its suitability for **data analytics**;
- the **challenges** the team faces in further developing and maintaining the code, the benefits of using reusable code and any regulatory and ethical implications.

Analyse the **approach** you took to implement your solution. This should be based on a suitably named **data analysis lifecycle or framework** (e.g. PPDAC, CRISP-DM, SDLC or similar frameworks). Discuss, especially, the **key steps taken** in loading and preparing the data for analysis. Use the **outputs** (calculated values and any charts and visuals) from your **EDA in Part 2** to explain the key findings. Key findings can relate to any business insights that might be gained from the analysis of the historical loan records, as well as any general lessons about the approach being taken for this project (may include comparisons/contrasts to similar real-life examples). The HFS board wants to know if your approach works, is sustainable and worth building on.

Your Approach discussions should include:

- The code libraries used (what these are and why they are needed);
- The reasons for choice of language and platform;
- The design (e.g., using pseudo code) to aid understanding and allow the design to be implemented in other languages if needed (code agnostic);
- Testing the code (e.g., by using Excel) to confirm that the programme works; and
- Maintenance of any code to allow controlled and efficient development (e.g., by version control using Git and good commentary)

Recommendations for future work: Briefly explain how your solution could be further developed to build a predictive model using the historical data provided. (i.e. A model that can be trained to predict the outcome of any future loan application.)

This should include a short **explanation** of the techniques, libraries, tools, and objective functions used to evaluate the precision of your recommended predictive model.

Note: There is no need to implement your recommendation ideas in code or build a predictive model

Your recommendation must also discuss:

- how existing data collection and data quality may be improved;
- if other libraries might be used to increase model accuracy;
- additional training or support the staff will need in order to use these new tools.
- safeguards needed to ensure compliance to regulatory and ethical standards.

Conclusions: A brief conclusion summarising the main points in the report.

In writing your report, use the insight and knowledge provided in this module but also leverage sound independent academic research to support your report. As you develop your work, you should self-evaluate your developing draft against the criteria set out in the Marking Guide below (See Section 5).

Programming Solution Guidance notes:

TASK 2: Construct a Programming Solution (LO3)

You must deliver a pre-run **Jupyter** notebook also known as an **Interactive Python Notebook** (a `.ipynb` file) using Google Colab, with the code used, with **comments**, to explain the scripts, the libraries used, and the logic. All such commentary should be **written using the built-in markup language** (Markdown text) or embedded using the standard Python syntax (preceded by a `#`).

The notebook which you create should highlight some of the key findings which you found from the given data and their insights which you can provide to the business.

The Python Notebook should include the following:

1: Loan Data Solution Development

Python Code: Copy and paste the entire contents of your `.ipynb` file – screenshots of the notebook with the **executed code** and markdown text.

Test Results: Copy and paste any output from Excel showing if it has confirmed the correct working of your code.

Pseudo code: to communicate the design in a logical manner.

The Python Code should be presented in a new `.ipynb` Jupyter notebook within Google Colab . Programmatically extract the two datasets from these two files which contains information about past loan records. The integers shown in each column of the loan datasets are to be interpreted as follows and you may read them as strings or numbers in your code as you deem appropriate.:

- Gender: 1-Male, 2-Female
- Married: 0-Single, 1-Married
- Dependents: 0, 1, 2, 3+
- Graduate: 0-No, 1-Yes
- Self_Employed: 0-No, 1-Yes
- Credit_History: 0-No, 1-Yes
- Property_Area: 1-Urban, 2-Semiurban, 3-Rural

You should use Python to load the information of these datasets in memory. You should also add comments to your notebook, explaining the steps taken to load the data, how you treated the PDF data, the libraries called and the overall procedure, bearing in mind this will be used for training colleagues in future.

2: Descriptive analysis

First, check the datasets and make sure the data that comes from these two files is valid. Ensure your loan data is correctly indexed on the LoanID column.

Then, clean the loan data. Provide an explanation of the steps taken to ensure data preparation for analysis such as the correction of duplicates, missing values, outliers etc.

Next, carry out Descriptive analysis on the current loan data. Your notebook file should contain the code and output of some basic Exploratory Data Analysis (EDA) which *must* include the following:

- The total amount loaned by HFS.
- The average amount loaned.

- The average loan term.
- The total number of applicants broken down into Approved and Rejected, the number of males and female in each case, and shown in an appropriate chart.
- The maximum and minimum of amounts loaned and shown in an appropriate chart.
- The number of self-employed who had their loan approved, expressed as a percentage of all who had their loan approved, and shown in an appropriate chart.

Optional

- The income distribution of all main applicants, showing average and standard deviation.
- The top ten applicants by loan amount.
- The distribution of properties (rural, urban etc) of all loan applicants, shown in an appropriate chart.

Note: Ensure currency is appropriately presented (use £ for UK currency)

The code must be commented using the built-in markup facility in Colab or embedded commentary so that it can be maintained in future. The output of calculations and visualisations should be shown in the notebook and in the business report. This code and outputs should then be copied and pasted as Appendix 1 in your Part 1, business report. Also, this pre-run code file (Python Jupyter notebook) should ideally be embedded or a link to it presented in your report. Details of how you can do that is provided in the next section.

You must include an accessible link to your code to pass this module.

3. Report Structure and Referencing

In addition, ten marks are awarded for the overall professionalism of your report and the adoption of academic standards.

Guidelines:

- Your report should follow the section naming structure and order set out in the Brief. You should also add your own sub-headings as you see fit to demonstrate your ability to develop structure and content.
- Your report should include an auto-generated contents page including section headings and sub-headings. The contents page should also include a page-referenced list all tables, charts and figures provided in your report. Remember to number all pages in your report, for example 'Page 8 of 12'.
- Ensure you develop your discussion in a logical progression: Findings, inferences, conclusions, recommendations.
- Do not make general assertions without supporting evidence.
- Zero spelling errors and grammatical mistakes
- Cite all your sources in the body of the text and in the Referencing using the Harvard Referencing style <http://bpp.libguides.com/Home/StudySupport>.
- Include a blend of industry research, case studies and academic references.

You should set out your Business Report in one **word** document, according to the following heading structure. The sections in **bold** count towards your word count.

- University Cover Page
- Table of Contents
- **Introduction**
- **Approach**
- **Recommendations**
- **Conclusions**
- Programming Solution
- Appendices (Optional)

You should add sub-headings under this overall structure as you feel fit to demonstrate your ability to develop the section themes and to provide meaningful sub-structure. But you must use this overall structure to provide a consistent framework against which your marker will allocate marks.

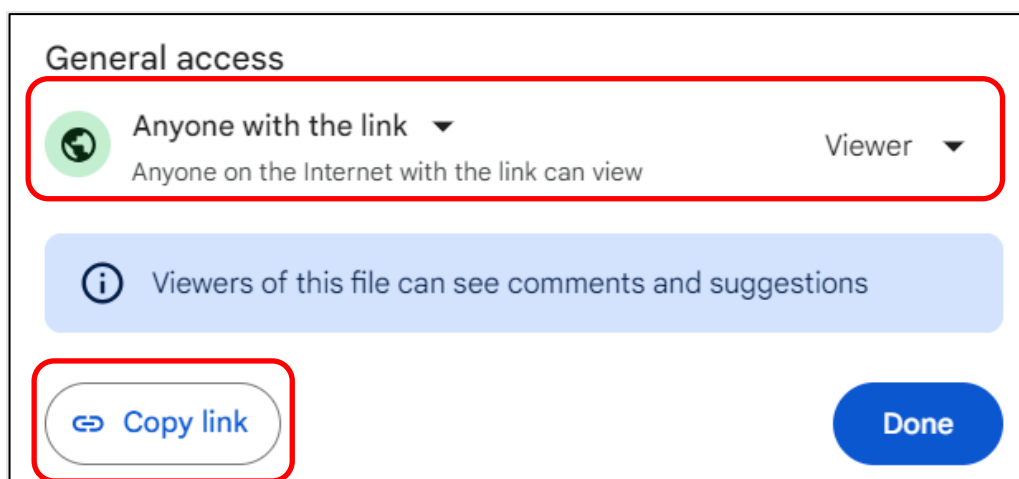
You will be deducted marks if you do not follow this structure. Also, note that there is NO requirement for producing an Executive Summary.

Total word count: **2,500**. The Cover Page, Table of Contents, References, Appendices (Optional), Tables, Charts and Figures do *not* count towards word count.

The code output, test validation, pseudocode and content of the Python Notebook is *not* included in the word count.

IMPORTANT: Adding your pre-run code to your report prior to uploading to TurnItIn

Locate the report file and **embed your *Pre-run* Python notebook** into your **Word document**. If you are unable to embed your python notebook for any reason, you **must** provide a *shared* link to the file. This is easily done within Google Colab by selecting the 'Share button' in the top right-hand corner of the screen:



NOTE: If you do not embed your notebook or provide a link you will lose marks.

4. Mapping Learning Outcomes to Assessment Tasks

The table below sets out the mapping between the three Learning Outcomes and the key tasks in your Summative Assessment which are designed to test your achievement against these Learning Objectives.

Learning Outcome	Mapping to Summative Assessment Tasks
LO 1: Critically evaluate the principles of programming and <i>apply</i> them in a business context	TASK 1 – Business Report: <ul style="list-style-type: none"> Introduction (including a background to the business problem) Approach, Conclusion and Recommendations
LO 2: Critically evaluate the use of code libraries in programming for a business context	
LO 3: Construct a programming <i>solution</i> to solve a defined business problem	TASK 2 – Programming Solution

5. Marking Guide

The assignment is marked out of 100 and counts towards 100% of your module mark. The following table shows the tasks, marks and marking rubric. *You should iteratively self-assess your performance against the Marking Guide as you develop your draft submission to evaluate your performance against your target grade.*

LEARNING OUTCOMES	Fail (0-39%)	Marginal Fail (40-49%)	Pass (50-59%)	Merit (60-69%)	Distinction (70-79%)	High Distinction (80-100%)
(LO1)– Critically evaluate the principles of programming and apply them in a business context.	<p>Weak report. The work does not identify the business case. It shows no justification of the proposed solution. No mention of user needs, challenges, and data quality. Has no mention of programming principles to help ensure a robust software solution.</p>	<p>The work has limited identification of the business case. Has some justification of the proposed solution. There is limited mention of user needs, challenges, and data quality.</p> <p>There is limited mention of programming principles to help ensure a robust software solution.</p>	<p>The work shows Satisfactory identification of the business case that can be used to justify the proposed solution. It mentions user needs, challenges and data quality.</p> <p>There is some basic discussion on programming principles.</p>	<p>The work shows Good identification of the business case that can be used to justify the proposed solution. It includes some discussion of user needs, challenges and data quality.</p> <p>There is good discussion on programming principles. Includes some justification of principles used.</p>	<p>Excellent presentation of a business case that can be used to justify the proposed solution. Includes good discussion of user needs, challenges and data quality.</p> <p>There is detailed discussion on programming principles, elaborating their benefits. Includes justification of use of the principles applied.</p>	<p>Outstanding presentation of a business case that can be used to justify the proposed solution. Excellent discussion of user needs, challenges and data quality.</p> <p>There is extensive discussion on programming principles, elaborating their benefits. Includes justification of use of the principles applied.</p> <p>Has some references to real life cases/studies.</p>

<p>(LO2) – Critically evaluate the use of code libraries in programming for a business context</p>	<p>Work indicates poor knowledge base with many omissions and/or lack of theory relating to the use of programming for data analysts (e.g. relevant code libraries) within a software.</p> <p>Has no discussion on framework used.</p> <p>Has no justification of software tools used.</p> <p>No discussion on code reusability and maintenance.</p>	<p>Work is inadequate and has implicit knowledge base. Has some omissions and/or lack of theory relating to the use of programming for data analysts such as use of code libraries within a software development business context.</p> <p>Has limited discussion on framework used.</p> <p>Has limited justification of software tools used.</p> <p>Has limited discussion on code reusability and maintenance.</p>	<p>Satisfactory work. It has basic knowledge base; explores and explicitly analyses the data analytics discipline, its theory relating to the use of programming. Discusses some relevant code libraries with some originality and detail within a software development and business context.</p> <p>Has basic discussion on framework used.</p> <p>Has basic justification of the software tools used.</p> <p>Has basic discussion on code reusability and maintenance.</p>	<p>Has good knowledgebase, explores and analyses data analytics discipline, its theory relating to the use of programming with some originality, detail and autonomy within a software development and business context. Some benefits of predictive modelling are explained.</p> <p>Has good discussion on framework used.</p> <p>Has good justification of the software tools used.</p> <p>Has good discussion on code reusability and maintenance.</p>	<p>Excellent: Has very good knowledgebase, exploring and analysing data analytics discipline, its theory relating to the use of programming with clear originality, detail and autonomy within a software development and business context. The merits of different predictive models are discussed.</p> <p>Has very good discussion on framework used.</p> <p>Has very good justification of software tools used.</p> <p>Has very good discussion on code reusability and maintenance.</p> <p>Demonstrates some extensive reading.</p>	<p>Outstanding: Strong knowledgebase, exploring and analysing data analytics discipline, its theory relating to the use of programming with clear originality, detail and autonomy within a software development and business context. The merits of different predictive models are discussed. Thoroughly examines a wide range of sources, including some real life examples/case studies.</p> <p>Has strong and elaborate discussion on the framework used. Cites some real life examples.</p> <p>Shows extensive reading.</p> <p>Has strong justification of the software tools used.</p> <p>Has strong discussion on code reusability and maintenance.</p>
<p>(LO3) - Construct a programming solution to solve a defined business problem</p>	<p>Weak code: Has vague or no programming solution to solve the business problem.</p> <p>Does not show and loading of data sets.</p> <p>Has no pasted executed/run codes.</p> <p>No merging of data sets. No visualization.</p>	<p>Code shows limited programming solution to solve the business problem.</p> <p>Has code for loading of data sets.</p> <p>Has some pasted executed/run codes. No merging of data sets.</p> <p>Has some visualisations. Lacks insights.</p>	<p>Code shows basic programming solution to solve the business problem. Has code for loading of data sets. Has pasted executed/run codes. Shows merging of data sets. Has some satisfactory visualisations, but proper EDA insights not discussed or needs elaboration. Has</p>	<p>Shows a good programming solution to solve some of the business problem. Has code for loading of data two datasets. Has pasted executed/run codes. Shows merging of data sets. Has some good visualisations with EDA insights discussed. Includes some reasonable explanation</p>	<p>Excellent: A detailed programming solution to solve most of the business problems. Has very good code for loading of data two datasets. Has pasted executed/run codes. Shows merging of data sets. Has some very good, varied and relevant visualisations. Has good EDA insights with detailed steps. Includes good reasonable explanation,</p>	<p>Outstanding: Comprehensive programming solution to effectively solve the business problems. Has very good/advanced code for loading of data two datasets. Has pasted executed/run codes. Shows merging of data sets. Has excellent, varied and relevant visualisations. Has elaborate EDA insights with detailed steps to achieve the results.</p>

			some logical diagram to illustrate design stage.	and some comments. Presents a logical diagram, could be pseudocode.	markup, and some comments. Has good self-written pseudocode or logical representation of approach.	Includes good reasonable explanation, with mix of comments and markup and some comments. Has well-written logical pseudocode.
Report Structure and References (Applies across all LOs tasks)	<p>Weak use of a wide range of personal research which is critically evaluated for key conceptual issues and is largely consistent throughout.</p> <p>Weak references and notes with significant errors or omissions. Unprofessional appearance of report.</p>	<p>Limited, no clear evidence of considerable personal research with inappropriate references. Limited range of range of appropriate resources. Contains problems with inconsistencies with conceptual evaluation. Reference has some minor error or insignificant errors.</p>	<p>Satisfactory and substantial research. Evidence of a range of personal research with conceptual evaluation. There are a limited number of references, but the correct format is used, albeit with some errors. There may be some errors in formatting and presentation, but the report is reasonably professional in appearance. References not in alphabetical order.</p>	<p>Good evidence of an innovative or original use of extensive personal research which has been thoroughly evaluated conceptually. Good referencing has few if any errors. Some sources correctly cited and has complete set of references in the correct format and in alphabetical order. There is evidence of wider reading and research.</p>	<p>Excellent. Has evidence of an innovative or original use of extensive personal research which has been thoroughly critically evaluated both conceptually and methodologically.</p> <p>Excellent with precise, full and appropriate references and notes at near-publishing standard. Sources correctly cited and there is a complete set of references in the correct format and in alphabetical order. Report is professional throughout, as expected in a business or consultancy report.</p>	<p>Outstanding evidence of innovative, and original use of extensive personal research. Report has been thoroughly critically evaluated, conceptually and methodologically with deep reflection. Detailed and consistent approach to headings, tables and graphs. Sources correctly cited and there is a complete set of references in the correct format and in alphabetical order. Consistent formatting and presentation professional throughout, as expected in a business or consultancy report.</p>