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)
Student Reference Number (SRN)	<Your SRN>
Assessment Title	<Your assessment title> CW2[S]

Please complete the yellow sections in the below declaration:

Declaration of Original Work:

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.

The word count, excluding contents table, bibliography and appendices, is _____ words.

Student Reference Number: _____ Date: _____

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

Programming for Data Analysts (PDA)

SUMMATIVE Coursework Assessment Brief [CW2]

Submission mode: Turnitin online access

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 Hub. 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 General Academic Regulations (GAR) and Manual of Academic Procedures (MOPP) which are available on the Hub under Student Services | Help & Support | BPP
- 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 coding exercise in data analysis using a Python notebook.
- Part 2 - writing a business report.

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

Zappy Financial Services (ZFS) 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 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 ZFS as a Consultant Data Analyst, with programming and data analytics knowledge, as well as a deep appreciation for the need to balance business growth with a robust 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 the any code produced will be further developed and maintained. You also need to consider that any code developed will need to be accessible to (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 'Loans_Database_Table.pdf'
- An Excel file, called 'Zappy Loan 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.

2.4. Part 1: Construct a Programming Solution (30 marks) (LO3)

In Part 1, you will deliver 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.

The notebook which you create should highlight some of the key findings which you have in the data and the insights which you can provide to the business. The tasks which need to be completed in the Python Notebook include the following:

Task 1: Loan Data Automation

Create a new .ipynb notebook within Google Colab and load the TWO data files provided by the DBA. 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. Recall this will be used for training colleagues in future.

Task 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 ZFS.
- 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 (additional marks will be given for the following)

- 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.

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 showed in the notebook and in the report in Part 2. This code and outputs should then be copied and pasted as Appendix 1 in your Part 2 report.

2.5. Part 2: Report - Business Case (60 marks) (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.

The format of the report should include:

a) **Introduction**: This should first set out the purpose of the report and its structure. It must then cover the current business environment of companies like ZFS, 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:

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

b) **Approach**: Describe the approach you took to implement your solution. This should be based on a suitable **data analysis lifecycle or framework,** such as one covered in the module, Describe, 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 1 to explain the key findings. Key findings can relate to any business insight 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. The ZFS board want to know if your approach works, is sustainable and worth building on.

You should provide discussion of

- 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)

C) 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 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.

Appendices - (Note: these do not add to your word count.)

Appendix 1 – Code: Copy and paste the entire contents of your *.ipynb* file – screenshots of the notebook with the code and markdown text.

Appendix 2 – Test Results: Copy and paste any output from Excel showing if it has confirmed the correct working of your code.

Appendix 3 – Your pseudo code for the design.

In writing your report, use the insight and knowledge provided in this module but also leverage sound 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).

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 our 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 PDF document, according to the following heading structure.

- University Cover Page
- Table of Contents
- Introduction
- Approach
- Recommendations
- Conclusions
- Appendices

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, Tables, Charts and Figures do not count towards word count.

The content of the Python Notebook is not included in the word count.

4. Mapping Learning Outcomes to Assessment Tasks

The table below sets out the mapping between the three Leaning 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 apply them in a business context	Part 2
LO 2: Critically evaluate the use of code libraries in programming for a business context	Part 2
LO 3: Construct a programming solution to solve a defined business problem	Part 1

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, in order to evaluate your performance against your target grade.*

Assignment task	Distinction (70-100%)	Merit (60-69%)	Pass (50-59%)	Fail (0-49%)
Part 1 - Construct a Programming Solution – (LO3) – 30 marks	Guidelines: • 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).			
	Student correctly displays a programming solution to solve a business problem, the EDA outputs and explains in detail the steps taken to achieve the results	Student correctly displays a programming solution to solve a business problem, and the EDA outputs, with reasonable explanation and comments.	Student correctly displays a programming solution to solve a business problem.	Student fails to display a programming solution to solve the business problem.
Part 2 – Report Business Case (Introduction) (LO2) – 30 marks	Guidelines: • Identify the problem you are hoping to address and the workplace context. • Define the solution (a high-level description) • State what the benefits will be and the implications of not addressing the problem. • Discuss requirements, challenges and data quality,			
	Excellent presentation of a business case that can be used to justify the proposed solution. Good discussion of user needs, challenges and data quality.	Good presentation of a business case that can be used to justify the proposed solution. Some discussion of user needs, challenges and data quality.	Satisfactory presentation of business case that can be used to justify the proposed solution. Mentions user needs, challenges and data quality.	Weak answer. No justification of the proposed solution. No mentions user needs, challenges and data quality.

Assignment task	Distinction (70-100%)	Merit (60-69%)	Pass (50-59%)	Fail (0-49%)
Part 2 – Report Business Case (Approach, Recommendations, Conclusions) (LO1, LO2) – 30 marks	Guidelines: • Explain what steps should be taken to implement the proposed solution. • Steps should include cleaning and preparation, design, test and maintenance. • Clear recommendations as to what should be done to enable automation of loan approval			
	Excellent 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.	Good knowledgebase, exploring and analysing data analytics discipline, its theory relating to the use of programming with some originality, detail and autonomy within a software development and business context. The benefits of predictive modelling are explained	Satisfactory knowledgebase; explores and explicitly analyses the data analytics discipline, its theory relating to the use of programming (and relevant code libraries) with some originality and detail within a software development and business context.	Inadequate and often implicit knowledgebase with some omissions and/or lack of theory relating to the use of programming for data analysts (and relevant code libraries) within a software development business context.
Report Structure and References (Applies across all LOs tasks) - 10 marks	Follow the guidelines given in Section 3 Report Structure and Referencing			
	For a distinction the report will use a consistent approach to headings, tables and graphs. Sources will be correctly cited and there will be a complete set of references in the correct format and in alphabetical order. There is evidence of extensive independent reading and research. Formatting and presentation is professional throughout, as expected in a business or consultancy report.	Referencing has few if any errors. The report is reasonably well presented but could be improved by greater attention to detail. There is evidence of wider reading and research.	Referencing is satisfactory. 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.	Weak research with inappropriate references. No professional appearance of report