BPP Coursework Cover Sheet

Please use the table below as your cover sheet for the 1st page of the submission. The sheet should be before the cover/title page of your submission.

Programme	
Module name	
Student Reference Number (SRN)	
Report/Assignment Title	
Date of Submission *(Please attach the confirmation of any extension received)*	

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, and that I have researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Business.

- I declare that where I have used any AI tools, it was for the following reasons (highlight as appropriate):
  - To review a draft
  - To correct language errors
  - Other (please describe)

  _____

- I have copied & retained for BPP University's reference, all AI prompts used in the creation of AI content and all AI-generated responses in support of my assignment and attached relevant evidence in the appendices.

- I understand that I may be required to participate in a *viva voce,* where I will be questioned on any aspect of my assignment, including key concepts, theories, examples used, & any sources included.

- 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, submission is your declaration you are fit to sit.**

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

BPP School of Business

BPP UNIVERSITY BUSINESS SCHOOL

MSc Management with Data Analytics

Applied Modelling and Visualisation

Coursework Assessment Brief

Submission mode: **Turnitin online access**

1. General Assessment Guidance

- Your summative assessment for this module is made up of this **2,500** words 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 **2500** words. You must comply with the word count guidelines. You may submit LESS than **2500** 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: https://bpp.libguides.com/home/business
- 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 UPPs which are available on the HUB in the Help and Support section under Documents and Forms.
- **Use of AI in assessments is only allowed for the purposes of reviewing a draft, correcting language errors or if specified in the summative assessment brief.** If you have used AI for any of these purposes, you should indicate this on the Assignment Cover sheet. For more information regarding acceptable and unacceptable use of AI, please enrol onto the Generative AI Foundations course on the HUB.
- 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. Assessment Brief

Source: https://stock.adobe.com/uk/images/airplane/150585380

For this assignment you are working as a Data Analytics Consultant for the SilkySky Airways and have been asked to prepare a Consultancy Report based on the airline's passenger 'satisfaction' Data Set. This report and your findings will be used in a 'visually appealing' presentation to the CEO, Senior Flight personnel and Cabin Crew in the Annual Staff Conference and it has been proposed some *interactive* elements will be placed securely on the company intranet.

Summative Submission

You are provided with a set of data **SILKYSKY_DATA_CW2.csv** that summarises the levels of passenger 'satisfaction'. The file contains over 103,000 rows of information from the SilkySky Airways database system for the current calendar year. Your objective is to use machine learning principles to model and visualise key data with a view to helping staff better understand what factors impacted levels of 'satisfaction' for passengers using the airline. Each feature is listed below:

Field	Data Description
Ref	Number
id	Number
Gender	TEXT: Male/Female
Satisfied	Y = Satisfied N = Unsatisfied
Age	Number
Age Band	Under 18 25 to 34 55 to 64 45 to 54 35 to 44 18 to 24

	65 or over
Type of Travel	Business travel Personal Travel
Class	Business Eco Eco Plus
Flight Distance	Number: Distance in Miles
Destination	Text: Destination Country Name
Continent	Asia Europe Africa North America Europe/Asia (Eurasia) South America
Inflight Wi-Fi service	Number rating: 0 to 5 *(where 0 is low/poor)*
Departure/Arrival time convenient	Number rating: 0 to 5 *(where 0 is low/poor)*
Ease of Online booking	Number rating: 0 to 5 *(where 0 is low/poor)*
Gate location	Number rating: 0 to 5 *(where 0 is low/poor)*
Food and drink	Number rating: 0 to 5 *(where 0 is low/poor)*
Online boarding	Number rating: 0 to 5 *(where 0 is low/poor)*
Seat comfort	Number rating: 0 to 5 *(where 0 is low/poor)*
Inflight entertainment	Number rating: 0 to 5 *(where 0 is low/poor)*
On-board service	Number rating: 0 to 5 *(where 0 is low/poor)*
Leg room service	Number rating: 0 to 5 *(where 0 is low/poor)*
Baggage handling	Number rating: 0 to 5 *(where 0 is low/poor)*
Check-in service	Number rating: 0 to 5 *(where 0 is low/poor)*
Inflight service	Number rating: 0 to 5 *(where 0 is low/poor)*
Cleanliness	Number rating: 0 to 5 *(where 0 is low/poor)*
Departure Delay in Minutes	Number
Arrival Delay in Minutes	Number

Your summative submission should be a written report in ***MSWord format*** (**NOT** a PDF file) and should be at most **2,500** words. It should describe how applied modelling and visualisation can be used to present summaries of passenger data. Your report will inform a corporate presentation so should be appropriately tailored to a rich and varied audience consisting of CEO, Senior Flight

personnel and Cabin Crew. You are also required to carry out independent research into the deferent categories of 'satisfaction' and techniques used to analyse and forecast data in your report.

You must complete all the following tasks:

(ILO1 - Formulate innovative data driven solutions to commercial problems)

TASK 1: Develop a data-driven solution to the given scenario (ILO1).

The solution must use two analytical models to predict the scale and accuracy of the airline's data using the Python programming language and relevant Python libraries taking into consideration the following guidance notes.

Task 1 - Data-Driven Solution Guidance notes:

You should provide a data-driven solution that:

- ✓ Follows an established design methodology (e.g. PPDAC or CRISP-DM), including flowcharts and pseudocode
- ✓ Performs an Extract, Transform, and Load (ETL) process (including import, clean and prepare the data for analysis, whilst ensuring that the relevant test, validation and training sets are created).
- ✓ Performs Exploratory Data Analysis (EDA) with appropriate visualisations
- ✓ Trains and tests TWO analytical models
- ✓ Evaluates the models based on your choice of loss function
- ✓ Produces appropriate visualisations of your results
- ✓ Describes the solution development process

You should choose two from the following models:

- Logistic regression
- Naïve Bayes
- Decision Tree
- Bagging
- Random Forest
- AdaBoost
- XGBoost
- Artificial neural network
- Another appropriate state-of-the-art algorithm

(ILO2 – Critically evaluate the use of algorithms and model when developing analytical solutions)

Task 2: Critically analyse the two models chosen for your solution in Task 1 (ILO2)

Critically analyse the two models chosen for your solution in Task 1, and in particular, the strengths and limitations of each model using the guidance notes provided below with references to the relevant literature.

Task 2 Guidance notes:

Your critical analysis must also include:

- ✓ An explanation of your chosen loss function
- ✓ A short discussion of the accuracy metrics
- ✓ A **summary table** of the of the accuracy metrics of the two chosen models to support the selection of the best model

(ILO3 – Critically appraise the concepts, tools and techniques for data visualisation)

Task 3: Communicate your findings supported by several outputs from Task 1 (ILO3)

Communicate your findings supported by several outputs from Task 1, including graphical outputs such as **correlation matrix**, **heat map**, and **confusion matrix** using the guidance notes provided below.

Task 3 Guidance notes:

Your critical appraisal should be based on your findings in Task 1, and must also include:

- ✓ An analysis of how the Exploratory Data Analysis (EDA) output guided your selection of the analytical models
- ✓ An explanation of the justification for performing EDA and the use of appropriate descriptive statistics and visualisations to understand the results of that analysis
- ✓ A recommendation of the use of one model for sustaining or increasing the rate of 'satisfaction'

3. Research and Referencing

Your report should include a list of references used to develop the report and research to support the suggested approach. The list should use only the *Harvard Referencing System* as highlighted in the *General Assessment Guidance* section of this document. All the figures/tables used in the report must have captions and, wherever needed, properly referenced, and explained in your submission.

***Suggested* report format**

Cover page (University cover sheet)

Table of Contents

List of Abbreviations (if appropriate)

Introduction (Scope and Background)

Key Factors that impact on passenger 'satisfaction'

Tasks (with Technical Details and Independent Research)

Recommendations

Next steps

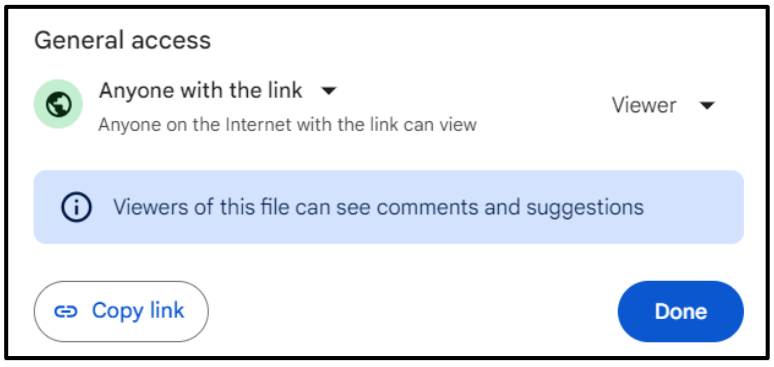
References

Appendix

The sections in **bold** contribute to the word count of **2,500** words

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

Locate the report file **and** embed your ***Pre-run*** Python notebook. If you are unable to embed your python notebook in your MS Word document 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:



IMPORTANT: **If you do not embed your notebook or provide a link you will lose marks**

4. Marking Guide

Modelling and Visualisation		Fail 0-39%	Marginal Fail 40-49%	Pass 50-59%	Merit 60-69%	Distinction 70-79%	High Distinction 80-100%
30%	**Formulate data-driven solutions to commercial problems (ILO1)**	Provided data-driven solution fails to execute, fails to display the options, or halts during execution. The data-driven solution showed inadequate demonstrati on of implicit knowledge base with some omissions and/or lack of theory relating to the use of ETL processes. No discussion of ambiguities, assumptions or anomalies. Data-driven solution fails to produce any outputs which can be used to communicate your findings	Data-driven solution correctly loads the input data file provided for the solution. No comments are given on the method used for their data-driven solution. Data-driven solution can demonstrate the concept of EDA, as well as comparisons of the outputs of the appropriate model outcomes and metrics but with no explanation or comments. The data-driven solution showed weak demonstration of implicit knowledge base with some omissions and/or lack of theory of the use of modelling and visualisation for a data project (and relevant code libraries). Data-driven solution correctly uses a package to produce communication tools but does not contain any explanation or commentary.	Data-driven solution correctly loads the input data file provided for the solution. Comments are given on the approach taken. Data-driven solution correctly handles duplicate values as well as EDA. Comments are given. The solution achieves the prediction for the 'satisfaction' likelihood and also correctly outputs appropriate model outcomes and metrics with reasonable level of commentary and explanation. Solution correctly uses a model to produce communication tools, with reasonable explanations and comments.	Data-driven solution correctly loads the input data file into a Python data structure. Comments and explanations are given with detail on the extract phase of the project. Data-driven solution handles duplicate values, missing values as well as descriptive statistics explaining the steps taken to reach the results. Notebook also achieves prediction for the 'satisfaction' Likelihood with good explanation and comments about the method used. There are model evaluation metrices outputted alongside predictions. Solution correctly uses a model to produce communication tools with good explanation and comments about the method used.	Data-driven solution correctly loads the input data file provided for the solution. The comments provided cover technical details of the extract phase of the project, demonstrating extensive knowledge on generic ETL process. Solution handles duplicate values, missing values and explains in detail the steps taken to reach the results. Correctly uses a model to achieve prediction for the 'satisfaction' likelihood and outputs the appropriate model outcomes and metrics. Explanations are detailed and profound. Solution correctly uses a model to produce communication tools, with very detailed explanation and comments about the model output and your chosen method of communication conveys this.	Data-driven solution correctly loads the input data file provided for the solution in a modular fashion. The comments provided cover exceptional technical details of the extract phase of the project, demonstrating extensive knowledge on ETL processes and their peculiarities. Solution handles duplicate values, handles missing values, correctly uses a model to achieve prediction for the future trends and outputs the appropriate model outcomes, metrics as well as an example of the prediction in action for a new mock entries and scenarios. Comments provided are profound in detail. Explain in detail the steps taken to reach the results with further explanation of methods to expand the steps taken or process followed. Also explains rationale behind the methods used. Solution correctly uses a model to produce communication tools with very detailed explanation and comments about the method used including examples of similar practices and suggestions to further enhance the communication of results.

Modelling and Visualisation		Fail 0-39%	Marginal Fail 40-49%	Pass 50-59%	Merit 60-69%	Distinction 70-79%	High Distinction 80-100%
30%	**Critically evaluate the use of models, analysing the strengths and weaknesses (ILO2)**	**Inadequate** and often implicit knowledge base with some omissions and/or lack of theory *relating to the use of programming for predictive modelling. No explanation of loss function, accuracy metrics, or recommendation of model for sustaining or increasing 'satisfaction' rate.*	**Weak and** often implicit knowledge base with some omissions and/or lack of theory *relating to the use of programming for predictive modelling. Weak explanation of loss function, accuracy metrics, or recommendation of model for sustaining or increasing 'satisfaction' rate.*	**Satisfactory knowledge base** that begins to explore and analyse the theory relating to the use of programming for predictive modelling. Satisfactory explanation of loss functions, accuracy metrices and comparative strengths of models based on ability to sustain or increase 'satisfaction' rate drawing on the academic literature.	**Good knowledge base** that explores and analyses the theory relating to the use of programming for predictive modelling. Good explanation of loss functions, accuracy metrices and comparative strengths of models based on ability to sustain or increase 'satisfaction' rate drawing on the academic literature with originality and autonomy.	**Excellent knowledge base** that explores and analyses the theory relating to the use of programming for predictive modelling. Excellent explanation of loss functions, accuracy metrices and comparative strengths of models based on ability to sustain or increase 'satisfaction' rate drawing on the academic literature with considerable originality and autonomy.	**Outstanding knowledge base** that explores and analyses the theory relating to the use of programming for predictive modelling. Excellent explanation of loss functions, accuracy metrices and comparative strengths of models based on ability to sustain or increase 'satisfaction' rate drawing on the academic literature with outstanding originality and autonomy at the cutting edge of current scholarship.
30%	**Critically using and appraising data visualisation techniques (ILO3).**	**Inadequate** and often implicit knowledge base with some omissions and/or lack of theory *relating to the use of EDA, descriptive statistics and data visualisation. There are no data visualisations, neither in the notebook nor the report.* *The student did not explain the justification for performing EDA, did not present appropriate descriptive statistics and has not explained how EDA guides model selection.*	**Weak and** often implicit knowledge base with some omissions and/or lack of theory *relating to the use of data visualisation. There isn't sufficient evidence of useful data visualisations, neither in the notebook nor the report.* *There is weak explanation for performing EDA, coming up with appropriate descriptive statistics and how EDA guides model selection.*	**Satisfactory knowledge base** that begins to explore and analyse the theory relating to the use of data visualisation. The student has presented several appropriate data visualisations, communicating insights visually both in the report and the notebook. There is satisfactory explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.	**Good knowledge base** that explores and analyses the theory relating to the use of data visualisation. The student has presented several appropriate data visualisations, communicating insights visually both in the report and the notebook. There is good explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.	**Excellent knowledge base** that explores and analyses the theory relating to the use of data visualisation techniques. The student has presented several high-quality data visualisations, excellently communicating insights visually both in the report and the notebook. There is excellent explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.	**Outstanding knowledge base** that explores and analyses the theory relating to the use of data visualisation. The student has presented several outstanding data visualisations, excellently communicating insights visually both in the report and the notebook. There is outstanding explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection. There are examples of data visualisation techniques at the cutting edge of industry using a variety of methods.

Modelling and Visualisation		Fail 0-39%	Marginal Fail 40-49%	Pass 50-59%	Merit 60-69%	Distinction 70-79%	High Distinction 80-100%
5%	Academic Research and Referencing Skills	**Inadequate** critical analysis or evaluation with some difficulties. Largely imitative and descriptive. Some difficulty with structuring the line of logical argument and accuracy in expression of argument.	**Limited** critical analysis and/or evaluation with reflection and broad evidence-based critique. Solid structure or argument including line of logical reasoning and accuracy in expression of argument.	***Satisfactory*** critical analysis and/or evaluation. Good reflection and solid, well-reasoned judgements forming from evidence-based critique. Consistent logical structure of argument including the line of reasoning and accuracy in expression of argument.	***Good*** *critical analysis and/or evaluation skills. Demonstrates intellectual originality and imagination* *Assumptions are clearly stated.*	***Excellent*** *critical analysis and/or evaluation* *skills. Demonstrates intellectual originality, integrity, coherence and imagination.* *Assumptions are clearly stated.*	***Outstanding*** *critical analysis and/or evaluation. Demonstrates intellectual originality, integrity, coherence, creativity and imagination working consistently in the higher cognitive domains to a professional standard.* *Assumptions are clearly stated.*
5%	*Follow the guidelines given in Section 3 Research and Referencing*	**Inadequate** references and notes but may contain inconsistencies, errors or omissions.	**Limited** and full and appropriate references and notes with minor or insignificant errors	***Satisfactory*** with precise, full and appropriate references and notes.	**Good** with precise, full and appropriate references and notes at a high standard.	**Excellent** with precise, full and appropriate references and notes at near-publishing standard.	**Outstanding** with precise, full and appropriate references and notes at publishing standard.
Overall Grade	***Overall Grade***	**Your submission demonstrates a limited understanding of the key concepts and theories covered in the module. Your analysis is superficial and lacks critical thinking or evidence-based support. Your writing is unclear, disorganised, and contains significant errors.**	**Your submission demonstrates a limited understanding of some (you could include a specific number) of the key concepts and theories covered in the module. You have attempted to apply these concepts to the chosen project or case study, but your analysis is limited and lacks depth. Your writing is somewhat unclear and contains some errors.**	***Your submission demonstrates a satisfactory understanding of the key concepts and theories covered in the module. You have applied these concepts to the chosen project or case study with some success. Your analysis is generally clear and logical but may lack depth or critical thinking. Your writing is mostly clear and organised, with few errors.***	**Your submission demonstrates a good understanding of the key concepts and theories covered in the module. You have effectively applied these concepts to the chosen project or case study, providing a well-structured and insightful analysis. Your writing is clear, organized, and free of significant errors. You have demonstrated a strong understanding of the subject matter.**	**Your submission demonstrates an excellent understanding of the key concepts and theories covered in the module. You have applied these concepts to the chosen project or case study with exceptional clarity, depth, and critical thinking. Your analysis is highly insightful and demonstrates a strong grasp of the complexities involved in your field of study. Your writing is exemplary, clear, concise, and free of errors. You have exceeded expectations and demonstrated an outstanding understanding of the subject matter.**	**Your submission demonstrates an exceptional understanding of the key concepts and theories covered in the module. You have applied these concepts to the chosen project or case study with deep insight, originality, and critical thinking. Your analysis demonstrates a deep understanding of the innovative advancements in your field of study. Your writing is outstanding, setting a new standard for clarity, conciseness, and originality. You have exceeded all expectations and demonstrated an unparalleled mastery of the subject matter.**