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	Applied Modelling and Visualisation
Student Reference Number (SRN)	
Assessment Title	MAV - Skywards International Airlines Report – CW3 [F]

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.

BPP UNIVERSITY SCHOOL OF BUSINESS AND TECHNOLOGY

MSc Management with Data Analytics

Applied Modelling and Visualisation

Coursework Assessment Brief - CW3 [F]

Submission mode: **Turnitin online access**

1. General Assessment Guidance

- Your formative assessment for this module is made up of this Coursework submission which does NOT count toward your final marks.
- **Please note late submissions will not be marked.**
- You are required to submit 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 **formative submission word limit is 1,000 words.** You must comply with the word count guidelines. You may submit LESS than 1000 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.
- 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. Formative Assessment Brief

Source: https://ik.imgkit.net/3vlqs5axxjf/NMG/uploadedImages/Articles/News/Industry/airplane-takeoff-Getty%20Images.png?tr=w-920%2Cfo-auto

For this assignment you are working as a Data Analytics Consultant for the Skywards International Airlines Corporation 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.

Formative Submission

Your formative submission will be a written report (at most **1,000** words) that should attempt tasks 1 and 2 and select **one** relevant analytical model to classify whether a customer is satisfied or not and critically analyse the model, as described in task 3. You MUST **include a code appendix** that performs the associated tasks.

You are provided with a set of data **SKYWARDS_DATA_CW3.csv** that summarises the levels of passenger 'satisfaction'. The file contains over 103,000 rows of information from the Skywards 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	18 to 24 25 to 34 35 to 44 45 to 54 55 to 64 65 or over Under 18
Type of Travel	Business travel Personal Travel
Class	Business Eco Eco Plus
Flight Distance	Number: Distance in Miles
Destination	Text: Destination Country Name
Continent	Africa Asia Europe Europe/Asia (Eurasia) North America 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 formative submission should be a written report (at most **1,000** words) that describes 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 also **include a code appendix** that can perform the following tasks:

1. Provide a rationale of the steps taken during each step of the **Extract, Transform and Load (ETL)** phase of the project, discussing any ambiguities, assumptions, and anomalies in the provided data and how you should deal with them (ILO1).

2. Explain the justification for performing **Exploratory Data Analysis (EDA)** and the use of appropriate descriptive statistics and visualisations to understand the results of that analysis, and critically analyse how the EDA process will guide your selection of analytical model (ILO3).

3. Select **ONE** relevant analytical model to predict the scale and accuracy of the airline's data (ILO2). Critically analyse the strengths and limitations of each model with references to the relevant literature. You should choose from the following models:

    i. Linear or Logistic regression

    ii. Decision Tree

    iii. Bagging

    iv. Random Forest

    v. AdaBoost

    vi. XGBoost

    vii. Artificial neural network

    viii. Another appropriate state-of-the-art algorithm

4. **NOT REQUIRED FOR FORMATIVE**
   Provide a **critical evaluation of each model** selected in the previous task by using your test data set (ILO2).

   a. Including an explanation of your chosen loss function.

   b. A short discussion of the accuracy metrics.

   c. Cast the accuracy metric, the number of correct predictions, and the number of incorrect prediction results of all the models to a table to allow for comparison.

   d. Based on your findings, make a critically justified recommendation for the use of one model for reducing the rate of 'satisfaction'.

5. **Communicate your findings** (ILO3). Provide several graphical outputs (with commentary) such as a ***correlation matrix, a heat map or confusion matrix*** of your results, in order to illustrate your analytical outputs in a visually pleasing manner.

6. A code appendix that performs the following tasks (ILO1):

   i. *Import, clean and prepare* the data for analysis, ensuring the relevant **test, validation and training sets are prepared**

   ii. *Perform* **Exploratory Data Analysis** with appropriate visualisations

   iii. *Train* and test the **ONE analytical model**

   iv. *Evaluate* the **ONE model** based on your choice of loss function

   v. *Produce* appropriate **visualisations** of your results

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

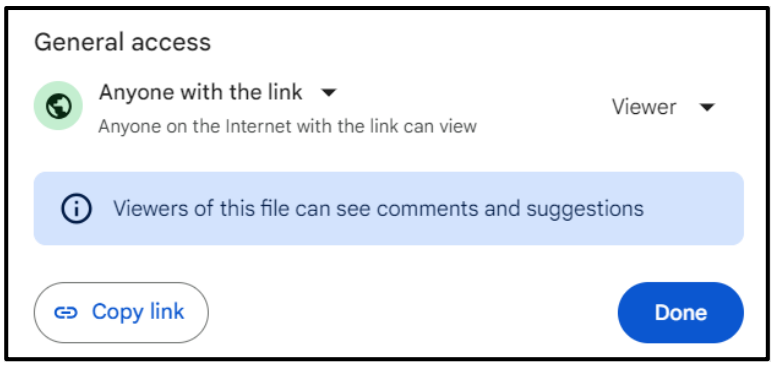
References

Appendix

The sections in **bold** contribute to the word count of **1,000** 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 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.

3. Marking Guide

Modelling and Visualisation		**0-39% Fail**	**40-49% Low Fail**	**50-59% Pass**	**60-69% Merit**	**70-79%80-100% Distinction**	
30%	Formulate data-driven solutions (ILO1) Attach code appendix that performs the outlined tasks and discuss ambiguities, assumptions and anomalies.	Notebook fails to execute, fails to display the options, or halts during execution. **Inadequate** and often 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. Notebook fails to produce any outputs which can be used to communicate your findings	Notebook correctly loads the input data file into a Python data structure. No comments are given on the method used. Notebook uses a package to conduct EDA, as well as comparisons of the outputs of the appropriate model outcomes and metrics but with no explanation or comments. **Weak** and often 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)* *Notebook correctly uses a package to produce communication tools but does not contain any explanation or commentary.*	Notebook correctly loads the input data file into a Python data structure. Comments are given on the approach taken. Notebook correctly handles duplicate values as well as EDA. Comments are given. The script achieves prediction for the 'satisfaction' likelihood and also correctly outputs appropriate model outcomes and metrics with reasonable level of commentary and explanation. Notebook correctly uses a package to produce communication tools, with reasonable explanations and comments.	Notebook 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. Notebook 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. Notebook correctly uses a package to produce communication tools with good explanation and comments about the method used.	Notebook correctly loads the input data file into a Python data structure. The comments provided cover technical details of the extract phase of the project, demonstrating extensive knowledge on dataframe imports. Notebook handles duplicate values, missing values and explains in detail the steps taken to reach the results. Correctly uses a package to achieve prediction for the 'satisfaction' likelihood and outputs the appropriate model outcomes and metrics. Explanations are detailed and profound. Notebook correctly uses a package to produce communication tools, with very detailed explanation and comments about the model output and your chosen method of communication conveys this.	Notebook correctly loads the input data file into a Python data structure in a modular fashion. The comments provided cover exceptional technical details of the extract phase of the project, demonstrating extensive knowledge on dataframe imports and their peculiarities. Notebook handles duplicate values, handles missing values, correctly uses a package 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. Notebook correctly uses a package 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		**0-39% Fail**	**40-49% Low Fail**	**50-59% Pass**	**60-69% Merit**	**70-79%80-100% Distinction**	
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 'satisfaction' reduction.*	**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 'satisfaction' reduction.*	**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 reduce '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 reduce '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 reduce '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 reduce 'satisfaction' rate drawing on the academic literature with outstanding originality and autonomy at the cutting edge of current scholarship.

Modelling and Visualisation		**0-39% Fail**	**40-49% Low Fail**	**50-59% Pass**	**60-69% Merit**	**70-79%**	**80-100% Distinction**
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		0-39% Fail	40-49% Low Fail	50-59% Pass	60-69% Merit	70-79%80-100% Distinction	
10%	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. **Inadequate** references and notes but may contain inconsistencies, errors or omissions.	**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. **Limited** and full and appropriate references and notes with minor or insignificant errors	***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. ***Satisfactory*** with precise, full and appropriate references and notes.	***Good*** *critical analysis and/or evaluation skills. Demonstrates intellectual originality and imagination* *Assumptions are clearly stated.* **Good** with precise, full and appropriate references and notes at a high standard.	***Excellent*** *critical analysis and/or evaluation skills. Demonstrates intellectual originality, integrity, coherence and imagination.* *Assumptions are clearly stated.* **Excellent** with precise, full and appropriate references and notes at near-publishing standard.	***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.* **Outstanding** with precise, full and appropriate references and notes at publishing standard.