

7BUI024W.2 Business Analytics Portfolio (Sem 2 2021/22)

Module leader	Thierry Chaussalet	
Unit	Business Analytics Portfolio	In-Class Test (14/12/2022) (Onsite)
Weighting:	70%	30%
Qualifying mark	40%	
Description	The Business Analytics Portfolio is made of: - Coursework - Final deadline: 05/01/23, 1pm - Weekly Online Multiple-Choice Test (only the best 6 marks count) – see Blackboard Note that only Excel and Palisade Decision Tools (StatTools, @Risk, PrecisionTree) are allowed to conduct analyses.	
Learning Outcomes:	LO1 select critically and apply quantitative modelling concepts for problem solving and decision-making. LO2 use appropriate business analytics techniques for real-world problems and data. LO3 select and use suitable software packages to analyse data and build models. LO4 write comprehensive and critical reports evaluating and interpreting obtained results.	
Handed Out:	Fri 21 October 2022	
Due Date	Thu 05 January 2023, 1pm – Final Deadline	
Expected deliverables	Electronic submission is required via BB by 05/01/23, 1pm . The main document including appendices should be only in ONE file (MS Word or pdf). The document should be self-contained and include answers to all questions/problems labelled accordingly , and it must not exceed 3,000 words and 12 pages (min font Arial/Calibri 10pts) excluding appendices. In addition, models must be submitted separately in ONE single ZIP file. <i>See additional guidelines at the end of document.</i>	
Method of Submission:	online via Blackboard	
Type of Feedback and Due Date:	Written/Oral feedback and marks 15 working days (3 weeks) after the submission. Marks will remain provisional until formally agreed by an Assessment Board.	

Copying and plagiarism

Any external sources utilised should be correctly referenced using a common referencing technique (e.g., the Harvard technique). For more details on referencing please visit <http://www.westminster.ac.uk/study/current-students/support-and-facilities/library-it-services/referencing>.

Copying and plagiarism carry severe penalties. Please note that the University offers an online learning tutorial designed to help students understand and avoid plagiarism. This can be accessed by any student under My Organisation on Blackboard. The tab is labelled 'Avoiding Plagiarism'.

Penalty for Late Submission

If you submit your coursework late but within 24 hours or one working day of the specified deadline, 10 marks will be deducted from the final mark, to minimum of the pass mark (50%), as a penalty for late submission. If you submit your coursework more than 24 hours after the specified deadline you will be given a mark of zero for the work in question unless a claim of Mitigating Circumstances has been submitted and accepted as valid.

It is recognised that on occasion, illness or a personal crisis can mean that you fail to submit a piece of work on time. In such cases you must inform the Faculty Registry Office in writing on a mitigating circumstances form, giving the reason for your late or non-submission. You must provide relevant documentary evidence with the form. This information will be reported to the relevant Assessment Board that will decide whether the mark of zero shall stand. For more detailed information regarding University Assessment Regulations, please refer to the following website: <http://www.westminster.ac.uk/study/current-students/resources/academic-regulations>

Coursework Description

7BUI024W

Given 24/10/22

Note: *All questions must be answered correctly, and the coursework must be returned by the due date above for full credit. All datasets, together with their descriptions, can be found in the Assessment section of the BlackBoard module site. Poor presentation will be penalised. Remember that this is meant to be an individual coursework and that any indication that it is otherwise may result in severe penalties. See additional guidelines at the end of the document.*

Important: Electronic submission is required via BB by **05/01/23, 1pm**. This main report including appendices should be only in ONE file (MS word or pdf: named **SURNAME_7BUI024W_CW_Main**, where SURNAME must be replaced by your own surname). **This document should be self-contained and fully comprehensible without using the additional ZIP file which contains your models** – this is a separate **single ZIP** file (named **SURNAME_7BUI024W_CW_Models**), which must be submitted at the same time through a separate submission link.

Problem 1 (42 marks)

Aware of your expertise in data analysis, the **Best University Guide (BUG)**, a guide publishing useful information for prospective University students across the World, has contacted you to undertake the analysis of data related to the 2023 QS World University Rankings [file **2023 QS World University Rankings V2.1 (For qs.com).xlsx**], using only the Top 500 Universities. The data contain [variables](#) such as Academic reputation, Employer reputation, Faculty/student ratio, Citations per faculty, International student ratio, International faculty ratio, International Research Network, Employment Outcomes, and Overall score. There are also variables about the location and the [classification](#) of Universities (Size, Subject Range (Focus), Research Intensity (Res.), Age, and Status).

In particular, they would like you to answer the questions below. For each question, **indicate what assumptions you have made and justify why you choose a particular approach** (graph, hypothesis test, both, etc.), and **summarise your findings in plain English** (no more than 2 lines).

a) Overall Scaled Score

- i. Use suitable summary statistics (e.g., mean, min, max, mode, median, 1st and 3rd quartile) and chart(s) to **analyse the distribution** of the overall score.
- ii. Use suitable summary statistics (e.g., mean, min, max, mode, median, 1st and 3rd quartile) and chart(s) to **compare** the overall score based on *Size*.
- iii. Repeat the analysis in (ii) to **compare** the overall score based respectively on Subject Range (*Focus*), Research Intensity (*Res.*), and *Age*.
- iv. Determine confidence intervals for the mean overall score respectively by *Size*, Subject Range (*Focus*), Research Intensity (*Res.*), and *Age*.
- v. Using appropriate hypothesis tests, help BUG find out whether there is a significant difference in overall score between:
 1. New (Less than 10 years old) and Historic (100 years old and more) Universities
 2. Small (Fewer than 5,000 students) and Extra Large (More than 30,000 students) Universities
 3. Specialist (2 or fewer faculty areas) and Full comprehensive (All 5 faculty areas + medical school) Universities

b) Overall Scaled Score by continents

- i. Add a variable named *Continent* to indicate which continent each University is located – you may want to use the file **countryContinent.xlsx** to this effect.

- ii. Use suitable charts to compare the overall score by *Continent*.
- c) Relationships between variables.
 - i. Calculate all correlation coefficients where appropriate.
 - ii. Explore the relationship between each of the variable and the overall score.
- d) Model(s) to forecast overall score in terms of other variables
 - i. Develop a model using all variables. Comment on your results and potential issues.
 - ii. Develop a model using only significant variables. Comment on your results and potential issues.
 - iii. Develop a model using only the variables not used to calculate the overall score. Comment on your results.

Problem 2 (18 marks)

Please answer the following questions after carefully reading the paper:

Bedair, H., Ting, N., Bozic, K.J., Della Valle, C.J., Sporer, S.M. (2011) Treatment of Early Postoperative Infections after THA - A Decision Analysis, *Clinical Orthopaedics and Related Research*, 469:3477–3485.

- a) What kind of model was developed? What was the aim(s) of the model?
- b) Briefly describe the possible decision alternatives, outcomes, and criterion for selecting the best decision.
- c) Using a suitable package such as Precision Tree, draw, for this problem, a decision tree such as the one in Figure 1. Include all relevant parameters.
- d) Fold back the tree and determine the best option. Is it the same as in the paper? Justify why or why not.
- e) Undertake the same one-way and two-way sensitivity analyses as in the paper, as well as any additional sensitivity analysis as you see fit (justify). You might use graphics of your choice to display results. Comment on your results, in particular on:
 - how they compare with those in the paper
 - the usefulness of sensitivity analysis in decision analysis.

Problem 3 (40 marks)

- a) A company sells air conditioners to four regions labelled as East, South, North, and West. They have estimated that the monthly demand for air conditioners is as follows: East: 9,500; South: 16,000; North: 11,500; and West: 9,000.

The company is considering building the air conditioners in four different cities (City 1, City 2, City3, and City 4). The variable cost of producing an air conditioner in a city and shipping it to a region of the country is given in the following table.

Unit variable cost (£)				
	East	South	North	West
City 1	206	225	230	290
City 2	225	206	221	270
City 3	230	221	208	262
City 4	290	270	262	215

Any factory can produce up to 15,000 air conditioners monthly.

Formulate algebraically a linear programming (LP) model to assist the company in minimising the annual variable cost of meeting demand for air conditioners. Solve using Excel Solver.

- b) The monthly fixed cost of operating a factory in each city is shown in the following table.

City	Monthly fixed cost (£ thousands)
City 1	600
City 2	620
City 3	580
City 4	560

Revise your formulation to minimise the monthly total (variable + fixed) cost of meeting demand for air conditioners. Solve using Excel Solver.

- c) Additionally, at least 5,000 units of North demand must come from either City 1 or from City 2. Revise your formulation in (b) to incorporate this constraint and solve using Excel Solver.

Show all details of your formulation and include copies of your software models and answer/sensitivity reports for assessment.

- d) The company has gathered historical data of the monthly demand for air conditioners over the past 9 years 9 months (since 2013)- see file **Data_Pb3.xlsx**.

i. Exploring the time series for each region.

1. Plot the data for each region as time series.
2. Determine whether these exhibit trend and/or seasonality?
3. If/Where relevant, calculate seasonal indexes for each region.
4. State with justification for each region what would be appropriate forecasting methods to forecast monthly demand until September 2023.

ii. Apply two appropriate forecasting methods for each region to forecast monthly demand between October 2022 and September 2023.

1. For each region, compare performance of the two methods using MAE, RMSE, and MAPE.
2. Check and comment on forecasting errors for randomness and autocorrelation
3. Using the best method, produce monthly forecast for each region from October 2022 to September 2023.

- e) Describe how you could use your results in d) to determine a better monthly shipping schedule between October 2022 and September 2023, i.e., how you would combine your results in d) with the optimisation models developed in a), b), and c) (use only a) for the sake of illustration).

Indicative Marking Scheme (in general, at least 50% of the mark is given to explanations / justifications). Results shown somewhere else than the report, e.g., on the spreadsheet will not be given any marks.

Problem 1 (42 marks)

a	b	c	d
22 (i: 3; ii: 3; iii: 6; iv: 6; v: 4)	4	4	12 (i: 4; ii: 4; iii: 4)

Problem 2 (18 marks)

a	b	c	d	e
2	2	7	3	4

Problem 3 (40 marks)

a	b	c	d	e
9	6	3	19 (i:10 [2 2 3 3]; ii:9 [4 3 2])	3

General points

- Your main report must be **self-contained and contain all the answers to the questions** including justifications and explanations. For instance, you cannot simply say “see filexyz.xlsx for results”. This will result in a **zero mark**.
- Results should be presented in a clear and concise manner. Poor presentation will be penalised.
- Answer questions as they are asked, i.e., **label your answers a, b, c, ...etc.** Not doing so will result in **substantial loss of marks**.
- **Do not put all the package outputs as your answer.** You should only include the relevant parts and in a decent table or graph format, not as a software snapshot. Plots are usually fine but only select the key ones for the main part of the answers. Put the rest in an appendix (and refer to them in the main text). If the question is “Use suitable charts to ...”, then you must include charts in the answer to that question, not just in the appendix.
- Remember that the whole report (excluding appendices) should not be more than **3,000 words**.

Problem 1

All relevant graphs and tables should be in the main text, not in the appendix. Note that software outputs often need some editing for acceptable presentation before being included in the text.

In some questions, charts are not enough. You need some inferential stats either using confidence intervals or hypothesis testing.

When using hypothesis testing, state clearly null (H_0) and alternative hypotheses (H_a). Note the conclusion is either you reject H_0 (there is sufficient evidence to do so at the chosen significance level) or you can't reject H_0 (there is insufficient evidence to do so). And remember that the null hypothesis has always “=”, “≤”, or “≥” in it.

Problem 2

Answers to a) and b) should be in your own words, not just cut and pasted. Each should be between 50 and 100 words.

Include the whole tree in the appendix if it is too large. However, you can present parts of it in the main text, especially as subtrees tend to be quite replicated.

Any Precision Tree output (tree, sensitivity analysis, etc.) must be justified and described, not just included.

Problem 3

a-c) You need to show and explain your algebraic formulation. The copy of the spreadsheet / Solver outputs should be in the appendix of the report.

d) i.4. State what methods are appropriate and why here. You are not asked to apply them in this question. This is in the next question.

d) ii. This is where you apply and compare the methods suggested in d) i.4.

e) You are asked to describe what you could do, not to perform any calculation.