University of Exeter

Introduction to Business Analytics

BEM2031

Module Handbook 2023-24

Module Convener: Alison Harper

2nd Floor, Building:One, CSAM

a.l.harper@exeter.ac.uk

Book office hours here

Introduction to Business Analytics Module Handbook

BEM2031 2023-24 Term 2

Module Description

This module will explore the role of information and analytics in supporting the development of strategies, and the practical techniques managers can use to design effective information flows.

Information is the lifeblood of business. Companies that manage information effectively can improve efficiency, be more responsive to market opportunities, achieve competitive advantage and operate more sustainably. As businesses drive towards sustainable strategies, they are looking for better information to guide decisions. A critical next step is to build information systems and data analytics capabilities that will turn raw data into actionable insights. This will enable companies to identify which actions more effectively are achieving their goals, detect risk or opportunity early, evaluate possible outcomes, allocate resources to achieve greatest returns, and measure the true impact of products.

Internationalisation: the module will draw on recent scholarship in the areas of data and analytics published by researchers internationally (the UK, Europe, the United States) and case studies based on a variety of national contexts.

Employability: the module will offer an opportunity to acquire knowledge and develop analytical skills for those pursuing careers in planning and analytics.

Module Aims

The module aims to enhance your understanding of the application of data in organisations, and to start the process of building your capability in designing, structuring, and analysing data.

Specifically, we will consider:

- How businesses use data to build, understand and report on their activities
- How to apply current concepts in data and analytics to real examples
- The use of 'Design Thinking' to create information management systems
- The initial tools for analysing numbers and text

ILO: Module-specific skills

- Critically evaluate current approaches used for collection, management, communication and analysis of commercial, operational and sustainability data, and how this data is used to support decision-making.
- Apply Design Thinking techniques to the analysis of a specific business challenge and use these to identify required information flows.
- Use data visualisation techniques to share original content and insight with a general management audience .
- Demonstrate familiarity with analytical tools available for the analysis of numerical and textual data and use these to find, derive and evaluate information.
- Discuss current developments and thinking in the information management industry, specifically around big data management, analytics, cloud, and visualisation techniques.

ILO: Discipline-specific skills

- Describe key terms and concepts in data and information management and be able to apply these to a typical business situation.

ILO: Personal and key skills

- Critical and reflective thinking.
- Demonstrate effective independent study and research skills.

General Support

- General administrative UEBS queries: info.buildingone@exeter.ac.uk
- Student timetable queries: student.timetable.buildingone@exeter.ac.uk
- Other general queries (SID): www.exeter.ac.uk/sid/ (please note SID email address no longer used)
- Business School welfare team: welfare.buildingone@exeter.ac.uk
- Accessibility (e.g. ILPs): www.exeter.ac.uk/wellbeing/accessability/support/
- Exams and ILPs: https://www.exeter.ac.uk/students/wellbeing/resources-and-services/exams-and-ilps/
- Mitigation (extensions and deferrals): https://www.exeter.ac.uk/students/infopoints/yourinfopointservices/mitigation/

Module resources

- Download and install R and RStudio: RStudio Desktop - Posit
- Start learning with Posit Cloud Primers and R cheatsheets
- Module textbook: Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Provost, Foster ; Fawcett, Tom (2013)
  Hard copies available at Forum Library, or available online at: Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking - University of Exeter
- R for Data Science is an excellent free book by Wickham and Grolemund.
- For GGPlot2 refer to the GGPlot2 book by Wickham.
- We will use Tidy Text Mining with R by Silge and Robinson.
- And Interpretable Machine Learning by Chistoph Molnar.
- You can find more information about R Markdown and its options on the website R Markdown (rstudio.com) or the book R Markdown: The Definitive Guide (bookdown.org).

Course overview 2024

Week	Tasks	Overview
T2: Week 1 16 January Workshop 1 (video)	Textbook Ch.1&2 A short talk about an algorithm for human attraction: Christian Rudder: Inside OKCupid: The math of online dating A great (also short) talk about using data to tell stories: Making data mean more through storytelling | Ben Wellington	**Data analytic thinking:** A broad overview of the different topics in business analytics. Business analytics as a leadership problem. The goal of this class is to prepare you to lead in a data-driving organization, or to help create the vision of a data-driven organization. How do you decide which models are most reliable? How do you recruit or manage a data science team? How do you persuade other colleagues and management about the proper course of action using data?
T2: Week 2	A data analytics pipeline: A Beginner's Guide to the Data Science Pipeline	**Managing and cleaning data:** Managing the data pipeline from the creation of new data, to processing

23 January	• An overview of data pre-processing: What Is Data Preprocessing? 4 Crucial Steps to Do It Right	the data, to producing results. What are the different kinds of data? How is data cleaned, stored, and made ready for analysis?
T2: Week 3 30 January	• Video: Dominic Bohan - Turning Bad Charts into Compelling Data Stories • Video: Hans Rosling, The best stats you've ever seen • Read: Storytelling with Data • Listen: Data is Personal (it was hard to pick an episode from this podcast, it's great) • RStudio primer on visualisation	**Data visualisation:** We will cover the basic elements of data visualization. We will focus on using the ggplot package. It's the most popular and most powerful visualization software used across the industry. This is the software both the BBC and the New York Times use to create their graphics.
T2: Week 4 6 February	• Textbook Ch.6 • Watch: StatQuest: K-means clustering Watch: StatQuest: Hierarchical Clustering • Watch StatQuest: PCA main ideas • Watch StatQuest: Principal Component Analysis (PCA), Step by Step • Play: Visualizing K-Means Clustering • Play Visualizing DBSCAN • Play: Principal Component Analysis • Read this great description of Hierarchical Clustering • And this and this useful descriptions of distance metrics	**Clusters and similarity:** A basic task in data exploration considers the similarity and groups in data. We will also examine dimension reduction through PCA

T2: Week 5 13 February Peer-reviewed Homework Due 16 February 2024	• Textbook Ch. 3, 4 • Watch StatQuest: Decision Trees • Watch StatQuest: Random Forests Part 1 • Watch StatQuest: Random Forests Part 2 • Watch Decision Trees and Random Forests lectures from Nando de Freitas for more detailed explanations • Play A Visual Introduction to Machine Learning • Play Random Forest Playground • Play Linear Regression (try clicking and dragging on points)	**Predictive modelling:** We will attempt to predict classes and continuous outcomes
T2: Week 6	READING WEEK	
T2: Week 7 27 February	• Textbook Ch. 5, 7, 8 • Watch StatQuest: Bias and Variance • Watch StatQuest: ROC and AUC Clearly Explained • Watch StatQuest: Cross validation • Watch StatQuest: Sensitivity and Specificity • Read AUC-ROC: a really good article	**Metrics of Evaluation.** What is a good model? How do you know if a predictive model is actually a good model and will perform well in the future?
T2: Week 8 5 March	• Textbook Ch. 10 • For reference, Text Mining for R	**Text Analytics** Digitized text is an incredibly common yet underutilized source of data in organizations. We will cover

	• Listen: Text Mining in R	some fundamentals of text analytics.
T2: Week 9 12 March Assignment Due 15 March 2024 Time: 15:00 hours	• Textbook Ch. 9, 11 • Review: Cohen et al. 2018	**Data Driven Decisions** Now that you have evidence, what option should you take? Cohen MC, Guetta CD, Jiao K, Provost F (2018) Data-Driven Investment Strategies for Peer-to-Peer Lending: A Case Study for Teaching Data Science. Big Data 6(3):191– 213
T2: Week 10 19 March	Textbook Ch. 12, 13 Watch: Introduction to Ethical AI • Listen: Talking Machines - AI for Good and The Real World Watch: Getting Specific about Algorithmic Bias Watch: 7 minutes to understand AI – A set of UNESCO videos For more detail: • Watch: Deep Learning State of the Art (2020) | MIT Deep Learning Series	**AI and Data Ethics** What are the risks and rewards of AI in our organizations? How do we create systems that create a better working environment while also improving productivity? How will the nature of work change as these technologies enter the workplace?
T2: Week 11 26 March Final project due 28 March 2024 Time: 15:00 hours	**Final project due Thursday 28 March. Time: 15:00 hours**	

Assessments

In Week 4-5 I will ask you to complete a **peer-reviewed homework assignment**. You can do this alone or in a group, and I will ask you to peer review each other's work. This is a **formative assessment**, which will help you with your final project. I would not expect it to take more than 2-3 hours, and I will provide a video to walk you through the code required.

There are **two summative assessments** for this module:

(a) The **assignment** is intended to develop and practice analytic skills. It is an **worth 30%** of your final grade. Outline for Critique Length: 300-500 words

Assignment Due: 15 March 2024 Time: 15:00 hours

(b) A summative assessment in the form of a **single final project is worth 70%** of your final grade. Analytics Report Critique Word Count: 3,000 words

Final Project Due: 28 March 2024 Time: 15:00 hours

(a) The assignment will be very similar to what was done in class but will use different datasets. There will be several sections which will be marked using the scale listed below.

Fully correct answers that complete the task in the expected manner will be given a high distinction of 8/10. For a full 10/10 I have left some room for innovation and personal exploration. Students who go above the expected, integrate a new package, attempt a new plot, try a new analysis, can be rewarded here.

Score	Description
0	The problem was not attempted.
2	The problem was attempted but largely incomplete or incorrect.
4	Concepts are understood, but not well explained in the context of the problem. Calculations yield the wrong answer due to minor or major errors. Plots are incorrectly generated.
6	The approach is generally correct. Calculations yield the wrong answer due to minor errors. Plots are roughly correct.
8	The solution is correct, well-documented, and the writing is clear. Reproducible code provides a correct step-by-step solution and is easy to follow. Plots are correct, detailed, and clearly explained.

10	The solutions are exceptional, clear, and creative. The solutions provided innovate and expand on existing knowledge.

(b) For the **final project**, you will be given a report similar to what may be provided in a business setting along with a dataset.

Your task is to critique the report and provide your own report. You will provide additional or corrected visualizations and analyses, and recommendations and conclusions to top management regarding the most prudent course of action based on the data.

The full details are in the separate assignment brief.

Additional Information

Late Submissions:

There are significant penalties for submitting work late.

For coursework:

- Work submitted up to one hour late will receive a 5% reduction in marks, down to a minimum score of the module pass mark
- Work submitted between 1 hour and 24 hours late will be capped at the pass mark
- Work submitted more than 24 hours late will receive a mark of zero

(NOTE: Where an exceptional three-week extension has been granted, work submitted at any point beyond the extended submission deadline will receive a mark of zero. Any students requiring additional time should submit a further application for mitigation within 24 hours of the extended deadline in order to be granted a deferral.)

Please always check you're submitting the right piece of work to the right place. A Late Submission of Coursework FAQs is also available within the TQA Manual: section 2.11.

Further information: FAQ | Student hubs | University of Exeter

Mitigation:

Mitigation works by giving you extra time to complete your assignment.
Two types of mitigation are possible:

(i) For coursework assignments, you can have an evidence-free extension of 72 hours (3 days). This option is available once per assessment. You can use it up to four

times during the academic year; any further extensions required after this must be applied for through the evidence-based process detailed below.
(ii) If you need an assessment extension of more than 72 hours and/or if you've used all four evidence-free extensions, you need to apply for evidence-based Mitigation.

Please see here for further information:
Mitigation | Student hubs | University of Exeter

Academic misconduct:

Academic misconduct is taken extremely seriously in all modules at the University of Exeter. Misconduct is said to have taken place when student has not been academically honest. This can be intentional or unintentional. Offences include plagiarism, collusion, fabrication, misrepresentation.

Further details on the School's plagiarism policy can be found here:
Academic Misconduct Procedure | Academic Study Administration | University of Exeter

Please be aware that you must not cut and paste sentences/passages from sources except if you are directly quoting from those sources, and you have indicated as such in your writing. Similarly, using any sort of 'writing service' that you may find online, colluding on assignments with classmates, or other such tactic is wholly inappropriate, and is viewed very seriously indeed. Your assignments must be written solely by you.

Generative AI policy

Using generative Artificial Intelligence (AI) tools such as ChatGPT in academic work - Referencing - LibGuides at University of Exeter

Referencing

For all work on this module, you must adopt the APA style of referencing.

Referencing Styles - Referencing - LibGuides at University of Exeter

Marks will be deducted from assignments with incorrect or incomplete referencing.

Do please be aware that lecture notes are not an appropriate academic reference. Academic journal articles and serious publications (online or print) are strongly recommended.

Internet sites should only be used if you can be certain about the academic credibility of the source, for example the Office for National Statistics (ONS), the Chartered

Institute of Personnel and Development (CIPD), and similar are credible and relevant sources. Business balls, Wikipedia, tutor2u.net, netmba.com, and similar websites are NOT acceptable academic references.