



UCD Michael Smurfit
Graduate Business School

MKT 42370: Advanced Analytics & Big Data

Summer Trimester 2023/24

Time (Room):

Wednesday	10:00-13:00 (C201-GSB)
Thursday	14:00-17:00 (C201-GSB)
Friday	10:00-13:00 (C201-GSB)

Office Hours (Room):

Wednesday	13:30-14:30 (<u>D010-GSB</u>)
Thursday	17:30-18:30 (<u>D010-GSB</u>)
Friday	13:30-14:30 (<u>D010-GSB</u>)

Instructor: David DeFranza
Email: david.defranza@ucd.ie

Every day, an increasing variety of new information is created at a larger volume and faster velocity than ever before. This information presents incredible opportunities for businesses but contending with its size and speed poses considerable technical and strategic challenges. Since much of this newly generated data captures the thoughts, opinions, and behaviors of individual consumers, marketers play an important role in addressing these challenges. Indeed, businesses capable of extracting knowledge from large data sets can achieve a considerable competitive advantage, and marketers capable of facilitating this process will find they have an advantage in the job market. In this module, we will engage with business problems, specifically those which might be faced by marketers, using data analytic thinking. Along the way, we will discuss the fundamental principles guiding the extraction of knowledge from information, challenges posed by very large data sets (including “Big Data”), and some of the most common techniques and technologies used to manage and mine such data.

Learning Outcomes

This module aims to prepare you for the challenges and opportunities today’s marketers face when dealing with “Big Data” and extracting knowledge from large data sets more generally. At the end of this module, you should be able to:

1. Explain the unique characteristics of and challenges posed by Big Data
2. Apply industry standard best practices for the organization and documentation of large datasets
3. Summarize the data mining process within the context of a business problem
4. Identify an appropriate analysis method based on a description of the business problem and available data

5. Assess the performance of a model or analysis based on common diagnostic metrics
6. Explain foundational data mining methods and machine learning algorithms

Required & Recommended Readings

This course will draw on two textbooks, readings from provide a background for topics discussed in class. The first book is available from the library and new, used, and electronic editions are available from most book sellers. The second book is available as an eBook from the library or as an electronic (PDF) edition. The books are:

Provost, F., & Fawcett, T. (2013) *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly.

- Publisher link: <https://www.oreilly.com/library/view/data-science-for/9781449374273/>
- UCD library link: <https://go.exlibris.link/gVGY4zh>

Zhou, H. (2020). *Learn data mining through Excel: A step-by-step approach for understanding machine learning methods*. Springer.

- Publisher link: <https://link.springer.com/book/10.1007/978-1-4842-5982-5>
- UCD library link: <https://go.exlibris.link/WJRJNnGW>

Chapters will be assigned as background reading. **Required reading** for each week will be posted to Brightspace.

Cases are provided through Harvard Business Publishing. Please use the links provided via Brightspace to download a copy of the cases. Note that many cases include a supplemental data file which you should also download.

Title	Authors
Allianz: Optimizing Customer Acquisition Strategy using Machine Learning	Bjarne Brie, Tineke Distelmans, Kristof Stouthuysen, Tim Verdonck, Christopher Grumiau, Thoppan Sudaman
HomeZilla: Attracting Homebuyers through Better Photos	Gregory S. Zaric, Homgmei Sun
A Technical Note on Data Preparation and Model Building with a Real Estate Dataset	Mehmet A. Begen, Cici Sun, Viola Qiu, Linke Bai
Customer Analytics at Bigbasket - Product Recommendations	Paul Abraham, Manaranjan Pradhan, Lakshminarayanan S, Ganesh Iyer, Dinesh Kumar Unnikrishnan
Cluster Analysis for Segmentation	Rajkumar Venkatesan
Linear Programming Basics	Victor Martinez de Albeniz Margalef
Generative AI Value Chain	Andy Wu, Matt Higgins
Assessing Prediction Accuracy of Machine Learning Models	Michael W. Toffel, Natalie Epstein, Kris Ferreira, Yael Grushka-Cockayne

Required Software

This module assumes access to and competency with spreadsheet software such as Microsoft Excel or Google Sheets. Please note that both tools are available for free to students through [UCD Connect](#). Applications will be discussed specifically in Microsoft Excel, which is recommended.

If you are not comfortable with the basics of Microsoft Excel (particularly cell references and functions) you can review these concepts through DataCamp courses which are provided as part of this module. Please note that completion of these courses is optional, and your progress will not be monitored or evaluated by instructors.

You can access DataCamp via the URL provided on Brightspace.

Assessment

This module has three assessment components, each with a specific weighting and marks totally 100%. There will be no exam. The weighting assigned to and responsibility for (i.e., whether an assessment is to be completed by an individual or group) each component is shown below. Following, each assessment component is discussed in detail.

Assessment Component	Weighting	Individual / Group	Deadline
1. In-Class Quizzes	30%	Individual	12 June & 21 June
2. Class Participation	20%	Individual	23:59 after each lecture
3. Data Analysis Project	50%	Group	15 July

In-Class Quizzes

In-class quizzes will test your knowledge of the course material. For each quiz you will have to answer 15 multiple-choice type questions about the content covered in the preceding weeks. All quizzes will be conducted in-class, during the first 45 minutes of class.

Please note that quizzes will only be available for the time specified and that failure to conduct the quiz will automatically result in zero marks. During that window you a limited amount of time to correctly answer as many questions as possible. Please note that you will answer a small subset taken from a large pool of questions, and that the questions will be randomly allocated to each student. This means that students are unlikely to see the same questions and/or questions in the same order. Quizzes may cover any material discussed in class *or in the required reading*.

Class Participation

Students are, of course, encouraged to participate actively during lectures by asking and/or answering questions. Formal assessment of class participation will take place through the submission of 10 “minute essays.” After each lecture, an assignment will be made available on Brightspace. This assignment will consist of two questions. The first question will ask you to

analyze a concept from the lecture or required readings. The second question will ask: What question(s) do you have about the material covered in today's class?

Responses will be due before 23:59 that day. If you have completed the required reading and attended the lecture, you should be able to answer each question in fewer than 200 words written in only a few minutes. No late submissions will be considered.

Data Analysis Project

Working in randomly assigned teams, you will be asked to conduct a comprehensive data analysis addressing a specific business problem. The problem is detailed in the required case "Allianz: Optimizing Customer Acquisition Strategy using Machine Learning." In this case, you will learn that online sales for the company have fallen sharply and unexpectedly. Working as marketing analysts, your team must utilize the available data to understand consumer segments, predict future sales, and make recommendations to senior management. Your analysis should follow the template provided by Provost and Fawcett (2013; see above), the CRISP-DM framework, and other strategic frames discussed in class.

One final submission for each group should consist of three artifacts. The first artifact (20% of project grade) is a detailed data sheet and/or dictionary documenting the underlying dataset. The second (30% of project grade) is a well-organized and documented Excel file containing your complete analysis of the data. Finally, the third artifact (30% of project grade) is a presentation (either PDF or PowerPoint format) which carefully details the results of applying the CRISP-DM to the case. The presentation file should be created in the style of a corporate presentation which your team might deliver to senior management. In addition, each individual group member will be asked to submit peer-evaluation rating relative contribution to the final product (20% of project grade).

Fundamentally, the analysis and report should answer the following questions:

1. What is the business problem in this case and how is this problem transformed into a machine learning problem?
2. Who are T&B customers? Describe the most important characteristics of the customers and their outstanding insurance policies.
3. Compare T&B customers to the online visitors that did not convert into purchasing customers. Can you identify the main drivers of customer conversion?
4. Can you identify customer segments using unsupervised learning? Can you predict customer conversion?
5. Evaluate the outcome of the predictive models you used. How effective are they in predicting customer conversion?
6. Develop a customer acquisition strategy for Allianz using the analysis results and insights from assignment questions 2 to 5.

Keep in mind that these questions should be answered within the context of the CRISP-DM.

Final reports and peer-evaluations should be submitted via Brightspace no later than 15 July, 2024. The assignment will be graded in accordance with a rubric which will be provided in advance. Additional details will be provided in class and via Brightspace.

Assessment Criteria and Grade Descriptors

This module utilizes criterion referencing and UCD grade descriptors. Before attempting the assessments for this module, you are encouraged to review the grade descriptors. A copy of the UCD grade descriptors can be downloaded from:

<https://www.ucd.ie/registry/t4media/UCD%20Module%20Grade%20Descriptors.pdf>

Please note that this course utilizes the “**Alternative Linear Conversion Grade Scale**” for conversion of numeric grades to letter grades, including calculation of a final grade from the course average. The scale can be found here:

<https://www.ucd.ie/registry/t4media/Standard%20Conversion%20Grade%20Scale.pdf>

Protocol for submitting your assignments: All continuous assessment should be submitted electronically via Brightspace or handed to the instructor in class, as directed, by the deadline specified. Please do not email assignments directly to the teaching team, unless explicitly directed to do so.

Statement of Inclusion

This module strives to be a model of inclusion. We respect and value student diversity in all of the modules we offer. We aim to provide and promote equitable access and opportunity to all students regardless of disability, race, age, gender, sexuality or socio-economic status. Students are encouraged to approach staff to discuss their learning needs. Any information disclosed will be treated confidentially.

University Policies

You should ensure you are familiar with the following UCD protocols:

- **Plagiarism and Academic Integrity:** UCD and the College of Business take academic integrity extremely seriously. All work must be your own, be completed specifically for this module and not have been submitted elsewhere. It should also be accompanied by a signed own work statement, such as the following:

I declare that all materials included in this essay/report/project/dissertation is the end result of my own work and that due acknowledgement have been given in the bibliography and references to ALL sources be they printed, electronic or personal.

The university's plagiarism and academic integrity policy is available from: <https://www.ucd.ie/secca/studentconduct/>

- **Harvard Referencing Style:** UCD College of Business uses the Harvard style of referencing. The UCD library has developed some resources on avoiding plagiarism and on how to reference correctly using the Harvard style. These resources are available from: <https://libguides.ucd.ie/academicintegrity>
- **Assessment Submission Form:** When submitting a piece of assessment, you are asked to attach an assessment submission form. This form is available from: <https://www.ucd.ie/t4cms/assessment%20submission%20form.pdf>
- **Late Submission of Coursework:** This policy outlines the steps you should take where you know in advance that you will not be in a position to meet a submission deadline and the penalties imposed in such circumstances. See https://hub.ucd.ie/usis!/W_HU_MENU.P_PUBLISH?p_tag=GD-DOCLAND&ID=137
- **UCD Extenuating Circumstances policy:** If, during the course of this module, you encounter any serious unforeseen circumstances that are beyond your control and which prevent you from meeting the requirements of the module, you should consult this policy. A student guide to this policy is available from: <https://www.ucd.ie/students/studentdesk/extenuatingcircumstances/>
- **UCD Student Code:** The UCD Student Code establishes the University's regulations and expectations in respect of student behaviour and conduct. The Student Code is available from: <https://www.ucd.ie/secca/studentconduct/>

Module Topics

The schedule below outlines the planned themes, by lecture. Updates and additions will be notified in class and on Brightspace. Required reading and additional information will be posted to Brightspace.

Lecture	Date	Topic
1	29 May	Course Overview; Intro to Big Data & Machine Learning
2	30 May	Finding and Managing Data; Intro to Feature Engineering
3	31 May	Feature Engineering & Applications
4	5 June	Association Rules & Market Basket Analysis
5	6 June	Clustering Methods & Applications
6	7 June	Introduction to Tree-Based Models
7	12 June	Quiz 1 ; Tree-Based Model Applications
8	13 June	Linear Regression Review
9	14 June	Logistic Regression
10	19 June	Support Vector Machines
11	20 June	Neural Networks
12	21 June	Quiz 2 ; Evaluating Model Performance
	15 July	Data Analysis Project Due

How to Find D010-GSB



UCD Blackrock Campus

Building Index	No.	Grid
Blackrock Examination Centre	1.	H5
Copi-Print / ILTG	2.	D3
D Building	3.	D5
E Building	4.	C4
East Hall	5.	C3
Graduate School of Business	6.	D3
Library	7.	E3
Liguori House (Grey House)	8.	E1
Management House Executive Education	9.	D2
Marketing Development Programme (Granite Room)	10.	D3
Oratory	11.	B3
Proby House	12.	D1
Restaurant	13.	B4
Services Desk	14.	C4
West Hall	15.	B3
Laundry Room	16.	D4

LEGEND

- Building
- Walkways
- Roadways
- Parking

UCD Unicare:
 Our campus, Our care...
 Emergency Line: (01 716) 7999

