Course Unit File

NOVA IMS Information Management School

Data Science for Marketing

Level:	**Second cycle**	**Semester:**	**Fall**
Curricular Unit Code:	**200201**	**Duration:**	**S1**
Year:	**1**	**Academic year:**	**2022-23**

Total of hours

Total of Hours Spent	Contact Hours	Nr. of hours per week	ECTS
210	; 30 TP	2	7.5

T - Theoretical; TP - Theoretical and practical ; PL - Practical and Laboratorial; TC - Field Work; S - Seminar; E - Internship; OT - Tutorial Orietation; O - Other

Responsible teaching staff

Nuno Miguel da Conceição António

Other teaching staff

Ana Isabel Neves Edmundo, Diogo de Bernardes Henriques e Almeida Diogo, Nuno Miguel da Conceição António

Office hours

Monday, Wednesday, and Thursdays: 19h00 to 20h30 (requires previous scheduling by email)

General objectives

Data science uses interdisciplinary techniques, such as statistics, data visualization, database systems, and machine learning to identify original, useful, and understandable patterns in data.
This course will familiarize students with Data science applications and analytical projects' lifecycle. Students will learn techniques for understanding and preparing data before building analytical models, such as data characterization/description, RFM, or association rules (e.g., market basket analysis).

Enrolment Requirements

Familiarity with the main theme of the course is not required. But it is highly recommended that the students have knowledge of Inferential Statistics and good computer user skills.
Students without previous training or experience with Python should complete the three following Datacamp online courses before the third week of this course (first practical class): *Introduction to Python, Intermediate Python* , and *Data manipulation with pandas* . The instructor will provide information on how to have free access to the Datacamp platform.

Learning outcomes of the curricular unit (LO)

Upon completion of this course, students should be able to:
LO1. Understand the fundamental concepts associated with Data Science
LO2. Recognize and identify Data Science applications in Marketing
LO3. Understand a Data Science project lifecycle
LO4 . Identify the most common types of data and how to collect data from SQL databases
LO5 . Use Excel PowerPivot to do basic and intermediate data description/characterization tasks
LO6 . Use Python as an analytical tool
LO7 . Be able to execute basic and intermediate data preparation and preprocessing tasks
LO8 . Be able to use data visualization as both a data exploration tool and a communication tool
LO9. Calculate and explain the most relevant performance measures used in association rules and describe the Apriori algorithm
LO10 . Calculate and explain data similarity and dissimilarity measures
LO11 . Calculate and interpret the RFM model
LO12 . Be able to design a simple experiment and identify a statistical test to measure it

Syllabus

LU1. Introduction to Data Science
LU2. CRISP-DM process model
LU3 . Common data types and introduction to SQL
LU4 . Data characterization and description
LU5. Data understanding
LU6 . Communication and Data visualization
LU7 . Data preparation
LU8 . Association rules and the Apriori algorithm
LU9 . Data similarity and dissimilarity measures
LU10 . RFM model
LU11 . Introduction to hypothesis testing
LU12 . Introduction to Excel Power Pivot
LU13 . Introduction to the Python programming language

Curricular unit Planning

Week	Type	Activities
Before week 7 (18/07)	-	Datacamp online courses: • *Introduction to Python* • *Intermediate Python* • *Data manipulation with pandas*
1 (06/09)	T	• Course overview • LU1 - Introduction to Data Science • LU2 - CRISP-DM process model
2 (13/09)	T	• LU3 - Common data types and introduction to SQL • LU4 - Data characterization and description

3 (20/09)	P	• SQL exercises • Data characterization/description exercise
4 (27/09)	P	• Data characterization/description exercises (cont.)
5 (04/10)	T	• LU5 - Data understanding • LU6 - Communication and Data visualization
6 (11/10)	T	• LU6 - Communication and Data visualization (cont.) • LU7 - Data preparation
7 (18/10)	P	• Python quiz • Data understanding and data visualization
8 (05/11)	T	• LU8 - Association rules and the Apriori algorithm
9 (08/11)	P	• Association rules (market basket analysis) exercise (including data understanding and data preparation)
10 (15/11)	T	• LU9 - Data similarity and dissimilarity measures
11 (22/11)	T	• LU10 - RFM Model • LU11 - Introduction to hypothesis testing
12 (29/11)	P	• Data similarity and dissimilarity measures exercise
13 (06/12)	P	• RFM model exercise
14 (13/12)	P	• Group project Q&A
Exam season	P	• Group project presentation and discussion

T: Theoretical P: Practical

Demonstration of the syllabus coherence with the curricular unit's learning objectives

The learning units (LU) cover the learning outcomes (LO) as presented in the following table:

	LO1	LO2	LO3	LO4	LO5	LO6	LO7	LO8	LO9	LO10	LO11	LO12
LU1	X	X										
LU2			X									
LU3				X								
LU4					X							
LU5							X					
LU6								X				
LU7								X				
LU8									X			
LU9										X		
LU10											X	
LU11												X
LU12					X							
LU13						X	X	X				

Teaching methodologies

The course is based on theoretical and practical classes. Several teaching strategies are applied, including slides presentation, step-by-step instructions on approaching practical examples, and questions and answers. The practical component is oriented towards exploring the tools introduced to students (Microsoft Excel and Python) and the project's development.
Applications used: Microsoft Excel, Python, Jupyter notebook, Microsoft visual code.

Evaluation

Due to the application-based design of the course, evaluation is continuous and applies to both the theory and practical components. There is no ?one only exam? with a single weight of 100%.
All evaluation grades are on a scale of 0-20.

- Python Quiz:
  - Individual - with materials consultation
  - 10% weight
- Group project:
  - The minimum grade is 8.0
  - 50% weight
- Exam:
  - Individual - with materials consultation
  - The minimum grade is 8.0
  - 1 st season or 2 nd season: 40% weight

All submissions should be made via Moodle. Submissions after the deadline will be rejected.

Demonstration of the coherence between the teaching methodologies and the learning outcomes

The theoretical presentation of concepts and methodologies, followed by the application in a project, provides students with the knowledge, skills, and competencies listed as learning objectives at the beginning of classes. Concepts and methods will be demonstrated with business cases so students can better understand their application in real-world business problems.
Together with the group project, the final exam requires students to integrate the entire content of the course and prove their mastery of it.

Bibliography

[A] Miller, T. W. (2015). Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python, Pearson
[B] Keller, G. and Gaciu, N. (2020). Statistics for Management and Economics (2nd edition), Cengage Learning
[C] Han, J., Kamber, M., Pei, J. (2012). Data Mining - Concepts and Techniques (Third edition), Morgan Kaufmann. Available online using the NOVA IMS VPN
http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=377411
[D] Linoff, G. S., and Berry, M.J.A (2011). Data Mining Techniques for marketing, sales, and customer support (Third edition). Wiley Publishing, Inc. Available online using the NOVA IMS VPN
http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=520245
[E] Provost, F., and Fawcett, T. (2013). Data Science for Business, O?Reilly
[F] Materials and references URLs provided in class by the instructors