Barcelona School of Management

MSc in Management
(2024/2025)

Practical Data Science

Professor: Daniel Lozano
Office hours: by appointment

Course Description

In today's data-driven world, businesses rely heavily on data science and analytics to inform strategic decisions. However, understanding how to effectively utilize data science tools and techniques for management purposes can be challenging. That's where our Practical Data Science for Management course comes in.

This course will provide you with the practical knowledge and skills needed to leverage data effectively in your management role.

Objectives

Through a combination of lectures and hands-on exercises, participants will learn how to:

1) Develop practical proficiency in Python programming for data science applications.
2) Understand the fundamentals of data science and analytics, including data collection, cleaning, and analysis techniques.
3) Interpret and communicate key insights derived from data to support decision-making processes.
4) Learn how to use techniques such as regression analysis, decision trees, random forests, and logistic regression to forecast trends, identify patterns, and make informed predictions about future outcomes.

Methodology

The course requires the use of Python as a programming language. Apart from the use of Python in the lectures, there will be 6 additional hours devoted to an introduction to Python. These hours will be held during the second quarter (see schedule below) and are of mandatory attendance for students who want to take this class.

There will be 20 one and half-hour lectures where we will dive deep into various topics such as data acquisition, exploratory data analysis, predictive modeling, machine learning algorithms, and model evaluation. Each lecture will blend theoretical concepts with practical examples and interactive discussions to ensure a comprehensive understanding of how data science can be effectively applied in management scenarios.

Class slides will be available on the eCampus after the corresponding class.

Evaluation criteria

Regular Assessment

To pass the course, you must earn at least 50 points out of 100, according to the following distribution:

Data project I: 15 points. This will be a group project based on the application of regression models.

Data project II: 15 points. This will be a group project based on the application of classification models.

Data project III: 20 points. This will be a group project based on the application of neural networks.

Final exam (TBA): 50 points.

Re-take

If the student has not passed the course, the final exam can be retaken during the retake period. The final grade will be computed using the grade in the retake exam.

Students are required to attend 80% of classes. Failing to do so without justified reason will imply a Zero grade in the participation/attendance evaluation item and may lead to suspension from the program.

Students who fail the course during the regular evaluation are allowed ONE re-take of the evaluation, in the conditions specified above. If the course is again failed after the retake, the student will have to register again for the course the following year.

In case of a justified no-show to an exam, the student must inform the corresponding faculty member and the director(s) of the program so that they study the possibility of rescheduling the exam (one possibility being during the "Retake" period).

Plagiarism is to use another's work and to present it as one's own without acknowledging the sources in the correct way. All essays, reports or projects handed in by a student must be original work completed by the student. By enrolling at any UPF BSM Master of Science and signing the "Honor Code," students acknowledge that they understand the schools' policy on plagiarism and certify that all course assignments will be their own work, except where indicated by correct referencing. Failing to do so may result in automatic expulsion from the program."

Calendar and (approximate) Contents

Session	*Content*	*Week*
	Python crash-course	
March 14 (17:30)	Intro to Python I (3-hour session)	
March 21 (17:30)	Intro to Python II (3-hour session)	
	Topic I Introduction and First Steps	
April 3 (19:00)	Introduction to the subject	1
April 4 (17:30)	Data acquisition and cleaning	1
April 8 (19:00)	Data preprocessing techniques	2
	Topic II Regression Models	
April 10 (19:00)	Linear Regression	2
April 15 (19:00)	Lasso, Ridge and ElasticNet	3

April 17 (19:00)	SVR and Decision Trees	3
April 22 (19:00)	Evaluation techniques (Loss, Precision, Confusion Matrix, F1-score)	4
	Topic III Classification Models	
April 24 (19:00)	Naïve Bayes	4
April 26 (23:55)	Last Day to submit **Project I**	4
April 29 (19:00)	Logistic Regression	5
May 6 (19:00)	Decision trees and random forests	6
May 8 (19:00)	Other classification techniques	6
May 13 (19:00)	Recommendation Models I	7
May 15 (19:00)	Recommendation Models II	7
May 17 (23:55)	Last day to submit **Project II**	7
May 22 (19:00)	Neural Networks I	8
May 27 (19:00)	Neural Networks II	9
May 29 (19:00)	Neural Networks III	9
	Topic IV Practical issues for a data scientist	
June 3 (19:00)	Model Evaluation and Validation	10
June 5 (19:00)	Handling imbalanced data and bias	10
June 9 (23:55)	Last Day to Submit Project III	
June 10 (19:00)	Ethical considerations in Data Science	11
June 12 (19:00)	Summary, questions, etc.	11
TBA	**FINAL EXAM**	12

Reading Materials/ Bibliography/Resources

Textbook

Data Science for Business (ISBN 978-1449361327)

Class slides provided by professor

Code samples provided by professor

Bio of Professor

Prof. Daniel Lozano holds a master's degree in Quantum Physics from the University of Copenhagen. After finalizing his master's studies, he started to apply all the skills

acquired to the fields of Machine Learning and AI to become a Data Scientist/ ML Engineer. Currently, his main areas of expertise are NLP (Natural Language Processing) and computer vision using Python.