

FOUNDATIONS OF DATA SCIENCE IN MARKETING

Course code	GRAV035
Course title	<i>Foundation of data science in marketing</i>
Type of course	<i>Compulsory</i>
Level of course	<i>Graduate</i>
Year of study	<i>1st</i>
Semester	<i>1st</i>
Number of credits	6 ECTS
Lecturer	<i>Dr. Quentin Gallea, quentin.gallea@protonmail.com</i>
Course prerequisites	<i>Statistical Data Analysis, Multivariate Statistics</i>
Form of studies	<i>Consecutive (evening)</i>
Teaching language	<i>English</i>

Objective of the course

Where are we, and where are we going? These are potentially the most fundamental questions we can ask ourselves, we can ask regarding our company or professional projects, or even what politicians would wonder at the country level. In order to answer those questions, an obvious solution is to measure and track progress, in other words translate the reality numerically. Even though numbers help to bring a sense of objectivity and facilitate comparison they can be equally misleading.

This class will guide you through the fascinating world of data science and will provide you with applied knowledge and tools to make better decisions, prevent misinformation and optimize data communication. This class is composed of four parts. First, we will explore how to define a Key Performance Indicator (KPI) robust to manipulation and bias. Second, we will dive into the analysis of the KPIs by learning and applying a standard approach for Exploratory Data Analysis in order to reduce the risk of statistical bias. Third, we will discover how to assess causality, an essential aspect to track progress and make decisions. Finally, we will cover data storytelling and communication.

Students will learn to code in Python. Python is currently the most widely used and versatile programming language. Hence discovering Python will open many doors to the participants (from Machine Learning to data analytics, including valuable research skills). In addition, we are going to use extensively Generative AI to support our journey.

Learning outcomes

At the end of the course the students will be able to:

Course learning outcomes (CLO)	Study methods	Assessment methods
CLO1. Demonstrate knowledge and understanding of the different statistical method for exploratory data analysis, causal inference, statistical and behavioural bias.	Lectures, practical assignments	Group project, practical assignments
CLO2. Apply causal inference method, exploratory data analysis method to marketing related questions	Lectures, practical assignments	Group project, practical assignments
CLO3. Determine correct statistical methods for decision-making.	Lectures, practical assignments	Group project, practical assignments
CLO4. Interpret the results of statistical analysis and present them in a rigorous, understandable, clear and concise manner.	Lectures, practical assignments	Group project, practical assignments
CLO5. Argue the choice of a key performance indicator: strength and weaknesses	Lectures, practical assignments	Group project, practical assignments

Learning methods

The course is taught entirely in English and is designed to achieve its aims through a combination of lectures and practical exercises, and a group project. Exercises and lectures are designed to encourage active participation, co-operative and creative work, interactive communication, as well as critical and statistical thinking.

Quality issues

The lecturer assures a variety of teaching and learning methods, interim knowledge assessment, and supply of learning material to students, as well as discussions of practical and individual work in class during the course.

Cheating issues

Individual testing, supervised practical assignments and individual project ensure studying quality and are forms to prevent cheating. The ISM regulations on academic ethics, including cheating (see, *ISM University regulations*) are fully applied in the course.

Course content

Day	Topic	Readings	In-class hours
1.	Introduction, introduction to GenAI and introduction to Python	[3]	4
2.	Part I: Robust and Meaningful KPI	[1]; [2]	4
3.	Part II: Exploratory Data Analysis:	[3]; [4]; [5]	4
4.	Part II: Marketing Mix Models		4
5.	Part III: Causal inference: The fundamental problem of causal inference and Directed Acyclic Graphs	[6] 1, 3 and 4; [7] 1 and 4;	4
6.	Part III: Causal inference: A/B testing and power analysis	[7] 2;	4
7.	Part III: Causal inference: Diff-in-Diff	[6] 9; [7] 13;	4
8.	Part IV: Data Storytelling: A good data story	[8] 4	4
9.	Part IV: Data Storytelling: A good data story	[8] 7, 8	4
			Total: 36

Self-study and assessment

Method	Total hours	Percentage of the final grade %
Group project presentation	62	50
Group project report	62	50
Total:	124	100

Assignments and evaluation

Assessment for this course will have two components:

The course final grade and overall assessment involves a group project. The group project will cover all the aspect of the class. The presentation will take place one week after the class on ZOOM on the 29th of January.

Students must score for all tasks of the course at the specified time. Precision of composite evaluations is left intact (up to 2 decimal places) until the end of semester and only the final evaluation will be subject to rounding.

For the project, you will receive a dataset. You will have to produce a report and a presentation analysing the dataset and the situation based on all the tools and sections covered in the class (KPI definition, EDA, Analysis, Causal analysis, Data storytelling and reporting). The presentation should summarize your report in a clear and concise way.

Re-take of the exam

In case of failing final evaluation, student has a possibility to correct the written report and resubmit (**100%** of the final grade).

Literature (course materials and readings)

1. Bergstrom, Carl T., and Jevin D. West. *Calling bullshit: The art of skepticism in a data-driven world*. Random House Trade Paperbacks, 2021.
2. Wheelan, Charles. *Naked statistics: Stripping the dread from the data*. WW Norton & Company, 2013.

3. McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.
4. Downey, Allen. *Think stats: Exploratory data analysis*. " O'Reilly Media, Inc.", 2014.
5. Bruce, Peter, Andrew Bruce, and Peter Gedeck. *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media, 2020.
6. Cunningham, Scott. *Causal inference: The mixtape*. Yale university press, 2021.
7. Facure, Matheus. *Causal Inference in Python*. " O'Reilly Media, Inc.", 2023.
8. Dykes, Brent. *Effective data storytelling: how to drive change with data, narrative and visuals*. John Wiley & Sons, 2019.