

Overview, Housekeeping

6G7V0026 Principles of Data Science

Luciano Gerber

Senior Lecturer
Department of Computing and Mathematics
Faculty of Science and Engineering
Manchester Metropolitan University

Block 1, Week 1 (20 to 24-Sep-21)

Welcome from the Unit Team

- ▶ Firstly, a very warm welcome! We hope that you find it all fulfilling.
- ▶ Me: [Luciano Gerber](#), L.Gerber@mmu.ac.uk, E127.
 - ▶ Bookable one-to-ones: Tue 2-3pm, Thu 11am-12pm.

Calendar, Timetable

- ▶ See our [Programme's Moodle Area](#)
- ▶ In brief:
 - ▶ Six teaching weeks: 20-Sep to 29-Oct
 - ▶ One assessment week (no teaching): 1 to 5-Nov

Teaching and Learning Strategy

- ▶ **Timetabled sessions** (the *lectorials*; typically, practical, problem-solving-focused, interactive sessions):
 - ▶ **Fri 9am-10:40am at JDC0.17**
 - ▶ **Fri 12pm-1:40pm at JDC2.04**
- ▶ **Bookable one-to-ones:**
 - ▶ Support with directed tasks (via MS Teams or in-person at JDE1.27)
 - ▶ Thu 11am-12pm; Fri 3-4pm

- ▶ Asynchronous activities
 - ▶ **Directed tasks** (e.g., work on exercises from webinar, engage with tutorial videos and study materials)
 - ▶ **Independent study time** (e.g., work on additional problem-solving tasks)
 - ▶ I would like to **encourage everyone** to **use the Discussion Forum** or our **MS Teams Channel** for asking questions, sharing answers and advice, and the like.

Attendance

- ▶ Please touch card at reader just before the start of each timetabled session.
- ▶ If something fails, please speak to me and I'll check the online system.

Support Software

- ▶ Microsoft Teams
- ▶ Padlet
- ▶ Mentimeter
- ▶ :

Health and Safety, COVID-19

- ▶ Please make sure to observe the **Covid-19 safety guidance** in the [COVID-19 Safety page](#).
 - ▶ “Continue to wear face coverings in enclosed and crowded indoor spaces. . .”
 - ▶ “take two COVID-19 rapid lateral flow tests every week. . .”
- ▶ Please see [Health and Safety section of the Programme's Moodle page](#).

Assessment

- ▶ Details will be published on Moodle within the next couple of weeks. You can make a start as soon as it is handed-in and, then, be able to obtain feedback before assessment week.
- ▶ The unit's assessment is 100% coursework.
- ▶ A data science project: you are asked to work on a case study, employing problem-solving skills, creativity, and initiative.
- ▶ Students are expected to obtain, process, visualise, and analyse datasets, and present results and insights in the form of a report.
- ▶ Submission's deadline is 5-Nov-21 (it should appear on Moodle already)

Diversity in Cohort

- ▶ Traditionally, we have students with different degrees of specialisation in some core skills (e.g., computational, statistical).
- ▶ Also, coming from different areas: some, Computer Science, Mathematics, Statistics; others, Biology, Zoology, Chemistry; also, Engineering, Business, and Social Sciences.
- ▶ A great opportunity for inter-disciplinary exchanges; for benefiting from complementary skills.
- ▶ There will be additional material (independent study time) for the more experienced students, and for those who might need additional support with some of the foundational skills.

Unit Content, In Broad Terms

- ▶ Types of data science problems addressed:
 - ▶ Understand data distributions
 - ▶ Extract, clean data
 - ▶ Summarise and aggregate data
 - ▶ Identify associations
 - ▶ Understand differences in groups of data
 - ▶ :

- ▶ Tasks (and supporting packages) in the data science process
 - ▶ Data Manipulation (e.g., loading, saving, subsetting) (pandas)
 - ▶ Data Exploration, Preparation, Cleaning (pandas, seaborn)
 - ▶ Data Grouping and Aggregation (pandas)
 - ▶ Data Visualisation (matplotlib, seaborn)
 - ▶ :