



L-Università  
ta' Malta

1<sup>st</sup> University of Malta

# Data Science Summer School

18-29  
July 2022



## WHY DATA SCIENCE?

Data science has been described as the 'sexiest job of the 21<sup>st</sup> Century'. It is expected to be one of the most sought-after skills in the next three years. Data science combines computer science, statistics, information science, AI, and statistics, together with their related methods, processes, algorithms, and systems for the purpose of extracting knowledge, patterns, trends, and insights from data. Over the past few years, data science has become increasingly popular and is now widely applied in business, industry, and academia.

The Data Science Platform (DSP) is organising the 1<sup>st</sup> UM Data Science Summer School which will be held in the afternoons in the two-week period Monday **18 July to Friday 29 July**. The school will be held at the University of Malta at the Msida campus. Sessions will be from **13:00 to 17:30** and consist of a two-hour lecture will be followed by a two-hour practical tutorial. There will be a 30-minute refreshment & networking break in between.

The school is targeted at undergrad and postgrad students in ICT and related areas as well as past graduates, practitioners, and professionals who are interested in learning more about this exciting new field. The DSP is committed to helping build a more diverse, and inclusive, data science community and also welcomes postgraduate students and professionals/practitioners in non-ICT areas.

Our lecturing team consists of academics and researchers in computer science, AI, computer and communications engineering, statistics, and information systems.

## LEARNING OUTCOMES

Upon successful completion of the school attendees should be able to:

- ⊗ Outline the challenge of working with big data using statistical methods;
- ⊗ Integrate the insights from data analytics into knowledge generation and decision making;
- ⊗ Analyse how to acquire data, both structured and unstructured, process it, store it, and convert it into a format suitable for analysis;
- ⊗ Apply the basics of statistical inference including probability and probability distributions;
- ⊗ Explain classification methods and related methods for assessing model fit and cross validating predictive models;
- ⊗ Identify, and illustrate, the main concepts of machine learning algorithms and their application;
- ⊗ Outline the difference between supervised and unsupervised learning approaches;
- ⊗ Summarize the quantitative methods of text analysis, including mining social media and other online resources;
- ⊗ Compare, and contrast, the various data interpretation and visualization tools;
- ⊗ Use Python as the main programming language to perform data science; and
- ⊗ Summarize the various privacy and ethics issues in Data Science.

## Language of Instruction

English

## Admission Requirements

UM Data Science Summer School applicants should have a good background, and proficiency, in any of the following or related areas computer science, software, IT, or other related areas A first degree in one of these areas is recommended but professional experience may also be sufficient to make up for the lack of a first degree.

A good knowledge of, and proficiency in, the Python programming language are recommended.

## Tuition Fee & Accommodation

Attendees pay a fixed tuition fee of **€260**. The tuition fee covers the cost of all the academic activities (classes, workshops, and tutorials etc ), as well as the attendance certificate and refreshments. Applicants from outside Malta must make their own arrangements for travel and accommodation.

## Assessment & Certification

Attendees who attend all the sessions, and complete all the coursework in the tutorial sessions, will be awarded a certificate of attendance Attendees will not be evaluated or graded.

## Schedule

Weekdays 13:00 to 17:30 at the Msida campus. A 2 hour lecture is followed by 2 hour practical with a 30 minute networking and tea & coffee break in between.

Attendees are expected to bring their own laptop for the practical sessions.

Session slides will be made available to the attendees.

## Registration

Attendees can register for the summer school at the link below.

👉 [um.edu.mt/events/datascience2022](https://um.edu.mt/events/datascience2022)

Payment is required at the time of registration.

Seats are allocated on a first come first served basis.

For group registrations please contact us as listed below.

## Contact Us

For more information and questions regarding the Summer School please do not hesitate to contact us on [ict@um.edu.mt](mailto:ict@um.edu.mt)

SCAN ME



# THE PROGRAMME

## Week 1

### Day 1 – Introduction to Data Science

Overview.

Introduction to Data Science.

Applications of Data Science.

Types of Data.

The Five Steps of Data Science.

Descriptive, Predictive, and Prescriptive Analytics.

Big Data.

Installing Anaconda.

---

### Day 2 – Python Programming

Python for Data Science.

List comprehensions, Lambda expressions, etc.

Pandas, Numpy, Matplotlib, Seaborn, etc.

Data Handling.

*Attendees will be assumed to have a background in programming.*

---

### Day 3 – Statistics and Probability

Concepts of Statistics.

Basis of Experimentation, normalization, and random sampling.

Hypothesis testing, confidence intervals, interpretation of p-values.

Introduction to Probability Theory.

---

### Day 4 – Data Science Concepts

Correlation vs Causation.

Classification vs Regression.

Evaluation of Classifiers and Regressors.

---

### Day 5 – Time Series, NLP, and Visualisation

Time Series and Properties.

Natural language Processing (NLP).

Visualisation using Matplotlib, Seaborn, and Plotly.

## Week 2

### Day 6 – Machine Learning

What is Machine Learning?

Supervised vs Unsupervised Learning.

Classification vs Regression.

Bias and Variance.

Linear Regression.

Logistic Regression.

Neural Networks.

---

### Day 7 – Clustering and Decision Trees

Agglomerative Clustering.

Divisive Clustering.

K-Means and DBSCAN.

EM Clustering.

Decision Trees and ID3.

---

### Day 8 – Ensembles, Bagging, and Boosting

Ensemble Learning.

Bagging and Boosting.

Random Forests.

AdaBoost, XGBoost, LightGBM, and CATBoost.

---

### Day 9 – Deep Learning

What is Deep Learning.

Building deep learning models with Keras.

Recurrent Neural Networks (including LSTMs).

Convolutional Neural Networks.

---

### Day 10 – Miscellaneous Topics

Naïve Bayes and SVMs.

Data Privacy and Ethics.

Data Science on the Cloud.