

MSc Data Science Orientation



Sept. 23, 2025

Agenda

- Introduction.
- Student support, learning community.
- Programme, course selection.
- Handy tools/skills to pick up; FAQ.

Introduction

Introduction [notation: [link](#)]

- Sarah McManus:

- MSc Programmes Manager. ⇒ See Sarah's slides.
- Office: COL 6.11 ([campus map](#)); e-mail: stats-msc@lse.ac.uk



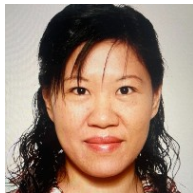
Sarah

Introduction [notation: [link](#)]

- Sarah McManus:
 - MSc Programmes Manager. ⇒ See Sarah's slides.
 - Office: COL 6.11 ([campus map](#)); e-mail: stats-msc@lse.ac.uk
- Lai Ching To (Sharon):
 - Programmes Administrator. ⇒ General programme enquiries, coursework submissions, newsletters and exams administration.
 - Office: COL 6.14; e-mail: l.to2@lse.ac.uk



Sarah



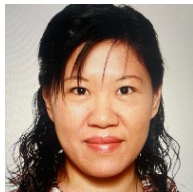
Sharon

Introduction [notation: [link](#)]

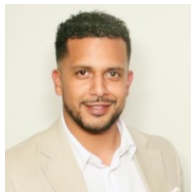
- **Sarah McManus:**
 - MSc Programmes Manager. ⇒ See Sarah's slides.
 - Office: COL 6.11 ([campus map](#)); e-mail: stats-msc@lse.ac.uk
- **Lai Ching To (Sharon):**
 - Programmes Administrator. ⇒ General programme enquiries, coursework submissions, newsletters and exams administration.
 - Office: COL 6.14; e-mail: l.to2@lse.ac.uk
- **Franks Reyes-Sarria:**
 - Department Senior Student Adviser ⇒ See Franks' slides.
 - Office: COL 5.09; e-mail: f.r.reyes-sarria@lse.ac.uk



Sarah



Sharon



Franks

Introduction – continued

- Ieva Kazlauskaitė:
 - Deputy Programme Director for MSc Data Science,
 - Office: COL 5.12; e-mail: i.kazlauskaite@lse.ac.uk



Ieva

Introduction – continued

- **Ieva Kazlauskaitė:**
 - Deputy Programme Director for MSc Data Science,
 - Office: COL 5.12; e-mail: i.kazlauskaite@lse.ac.uk
- **Zoltan Szabo:**
 - Programme Director and Academic Mentor of all MSc data science students $\Rightarrow$ next slide,
 - Office: COL 5.14; e-mail: z.szabo@lse.ac.uk



Ieva



Zoltan

Introduction – continued

- **Ieva Kazlauskaitė:**
 - Deputy Programme Director for MSc Data Science,
 - Office: COL 5.12; e-mail: i.kazlauskaitė@lse.ac.uk
- **Zoltan Szabo:**
 - Programme Director and Academic Mentor of all MSc data science students ⇒ next slide,
 - Office: COL 5.14; e-mail: z.szabo@lse.ac.uk



Ieva



Zoltan

Email sending recipe

- To = **Zoltan** (primary contact), Cc = **Ieva**.
- Do not send independent duplicate emails to Zoltan & Sarah!

Role of academic mentor

Responsibilities:

- 1 To provide advice on and give final approval of course choices at the beginning of Autumn and Winter Term.

Role of academic mentor

Responsibilities:

- ① To provide advice on and give final approval of course choices at the beginning of Autumn and Winter Term.
- ② Discuss any academic challenges you may experience.

Role of academic mentor

Responsibilities:

- ① To provide **advice** on **and** give **final approval of course choices** at the beginning of Autumn and Winter Term.
- ② Discuss any **academic challenges** you may experience.
- ③ Maintain **regular contact** with you through direct one-to-one meetings and through regular email contact as needed.

Role of academic mentor

Responsibilities:

- ① To provide **advice** on **and** give **final approval of course choices** at the beginning of Autumn and Winter Term.
- ② Discuss any **academic challenges** you may experience.
- ③ Maintain **regular contact** with you through direct one-to-one meetings and through regular email contact as needed.
- ④ Follow up on any concerns regarding your **attendance and progress**.

Role of academic mentor

Responsibilities:

- ① To provide **advice** on **and** give **final approval of course choices** at the beginning of Autumn and Winter Term.
- ② Discuss any **academic challenges** you may experience.
- ③ Maintain **regular contact** with you through direct one-to-one meetings and through regular email contact as needed.
- ④ Follow up on any concerns regarding your **attendance and progress**.

Office hours

- by appointment, email Zoltan to arrange,
- please **give context** and **be to-the-point**.

Student support+

Academic

- lecturers, class teachers,
- Moodle,
- academic mentor (Zoltan) – as covered.

Administrative

- Department of Statistics graduate handbook,
- programmes manager (Sarah).

- MSc study room: COL 7.02a,
- seminars:
 - Statistics and Data Science seminar,
 - Joint Econometrics and Statistics seminar.
 - Joint Risk and Stochastics and Financial Mathematics seminar.
 - Industrial seminar (coming soon).

- MSc study room: COL 7.02a,
- seminars:
 - Statistics and Data Science seminar,
 - Joint Econometrics and Statistics seminar.
 - Joint Risk and Stochastics and Financial Mathematics seminar.
 - Industrial seminar (coming soon).
- social media:
 - Instagram: @lsestatistics,
 - X: @LSEStatistics,
 - LinkedIn: Department of Statistics,
 - Student Hub.

(School-wise) student support, learning
community

School-wide (academic, pastoral, wellbeing) support and services

① LSE Student Support Map:

- "A student guide to [navigating help](#), resources, and support @ LSE".
- [download the Student Hub \(app\)](#).

School-wide (academic, pastoral, wellbeing) support and services

① LSE Student Support Map:

- "A student guide to [navigating help](#), resources, and support @ LSE".
- [download the Student Hub \(app\)](#).

② LSE Student Services Centre:

- various services / advice $\ni$ enrolment, immigration and visas, deferrals, course selection, exams, appeals, degree certificates, graduation.
- [1st contact point on general enquiries](#).

School-wide (academic, pastoral, wellbeing) support and services+

3 LSE LIFE:

- "to discover and **develop skills** you will use as a student and beyond",
[Ex: effective reading, argumentative writing, critical thinking, time management, revising for exams, avoiding plagiarism, research ethics, work-life balance]

School-wide (academic, pastoral, wellbeing) support and services+

3 LSE LIFE:

- "to discover and **develop skills** you will use as a student and beyond",
[Ex: effective reading, argumentative writing, critical thinking, time management, revising for exams, avoiding plagiarism, research ethics, work-life balance]
- availability: online & ground floor of **LSE library**
[Ex: books, study spaces, PCs, laptop loan service, eduroam, photocopying, printing].

School-wide (academic, pastoral, wellbeing) support and services+

3 LSE LIFE:

- "to discover and **develop skills** you will use as a student and beyond",
[Ex: effective reading, argumentative writing, critical thinking, time management, revising for exams, avoiding plagiarism, research ethics, work-life balance]
- availability: online & ground floor of **LSE library**
[Ex: books, study spaces, PCs, laptop loan service, eduroam, photocopying, printing].

4 LSE Digital Skills Lab:

- to develop key **digital skills**,
- availability: online (Teams & Zoom).

School-wide (academic, pastoral, wellbeing) support and services+

3 LSE LIFE:

- "to discover and **develop skills** you will use as a student and beyond",
[Ex: effective reading, argumentative writing, critical thinking, time management, revising for exams, avoiding plagiarism, research ethics, work-life balance]
- availability: online & ground floor of **LSE library**
[Ex: books, study spaces, PCs, laptop loan service, eduroam, photocopying, printing].

4 LSE Digital Skills Lab:

- to develop key **digital skills**,
- availability: online (Teams & Zoom).

5 LSE Careers:

- **career** advisory services and **support**.

School-wide (academic, pastoral, wellbeing) support and services+

3 LSE LIFE:

- "to discover and **develop skills** you will use as a student and beyond",
[Ex: effective reading, argumentative writing, critical thinking, time management, revising for exams, avoiding plagiarism, research ethics, work-life balance]
- availability: online & ground floor of **LSE library**
[Ex: books, study spaces, PCs, laptop loan service, eduroam, photocopying, printing].

4 LSE Digital Skills Lab:

- to develop key **digital skills**,
- availability: online (Teams & Zoom).

5 LSE Careers:

- **career** advisory services and **support**.

6 LSE Generate:

- developing **entrepreneurial skills**.

School-wide (academic, pastoral, wellbeing) support and services++

⑥ LSE Student Wellbeing Service:

- disability and mental health service,
- student counselling service: stress, personal & emotional difficulties.

School-wide (academic, pastoral, wellbeing) support and services++

⑥ LSE Student Wellbeing Service:

- disability and mental health service,
- student counselling service: stress, personal & emotional difficulties.
- $\ni$ Peer Support Scheme: "trained student volunteers who support their peers by offering a non-judgmental, confidential listening space to any LSE student".

School-wide (academic, pastoral, wellbeing) support and services++

6 LSE Student Wellbeing Service:

- disability and mental health service,
- student counselling service: stress, personal & emotional difficulties.
- $\ni$ Peer Support Scheme: "trained student volunteers who support their peers by offering a non-judgmental, confidential listening space to any LSE student".

7 LSE Faith Centre:

- religious activities, wellbeing (yoga, Tai Chi).

School-wide (academic, pastoral, wellbeing) support and services++

6 LSE Student Wellbeing Service:

- disability and mental health service,
- student counselling service: stress, personal & emotional difficulties.
- $\ni$ Peer Support Scheme: "trained student volunteers who support their peers by offering a non-judgmental, confidential listening space to any LSE student".

7 LSE Faith Centre:

- religious activities, wellbeing (yoga, Tai Chi).

$\Rightarrow$ your learning community: help others / speak with others (LSE students, alumni, staff, ...).

Programme, course selection

Programme structure: 4 units; regulations

- Terms [AT/WT/ST = Autumn/Winter/Spring Term; dates]:
 - 2×11 weeks (AT, WT; teaching; W6 = reading week),
 - 6-week term: ST (main exam period).

Programme structure: 4 units; regulations

- Mandatory courses:

- ST443 Machine Learning and Data Mining (H),
- ST445 Managing and Visualising Data (H),
- ST447 Data Analysis and Statistical Methods (H),
- ST498 Capstone Project (details later; F).

$$\Rightarrow 2.5 = 3 \times 0.5 + 1.$$

Programme structure: 4 units; regulations

- Mandatory courses:

- ST443 Machine Learning and Data Mining (H),
- ST445 Managing and Visualising Data (H),
- ST447 Data Analysis and Statistical Methods (H),
- ST498 Capstone Project (details later; F).

$$\Rightarrow 2.5 = 3 \times 0.5 + 1.$$

- Optional courses: $1.5 = 3 \times 0.5$ (≥ 0.5 unit of ST***).

- Data science options [ST = statistics]:
 - ST444 Computational Data Science ([H](#); [suspended 2025/26](#)),
 - ST446 Distributed Computing for Big Data ([H](#)),
 - ST449 Artificial Intelligence ([H](#)),
 - ST451 Bayesian Machine Learning ([H](#)),
 - ST455 Reinforcement Learning ([H](#)),
 - ST456 Deep Learning ([H](#)),
 - ST457 Graph Data Analytics and Representation Learning ([H](#)).

- Other statistics options [ST = statistics]:
 - ST405 Unsupervised Machine Learning and Multivariate Data Analysis (H),
 - ST411 Generalised Linear Modelling and Survival Analysis (H),
 - ST418 Advanced Time Series Analysis (H),
 - ST429 Statistical Methods for Risk Management (H),
 - ST436 Financial Statistics (H),
 - ST454 Bayesian Data Analysis (H; suspended 2025/26),
 - ST458 Financial Statistics II (H),
 - ST459 Quantum Computation and Information (H).

- Outside options [MA = maths, MY = methodology]:
 - MA407 Algorithms and Computation ([H](#)),
 - MY459 Computational Text Analysis and Large Language Models ([H](#)),
 - MY461 Social Network Analysis ([H](#)),
 - MY470 Computer Programming ([H](#)).

- Most courses in the Dept. of Statistics: [controlled access](#)
 - ⇒ limited numbers and you must meet their pre-requisites.
 - You can sign up to seminar groups *after* accepting your controlled access course offer.
- Many other departments operate similarly: [complete list](#).
- Some courses allow [auditing](#):
 - you can attend lectures,
 - seminar, assessment: no.

Essential

Select your courses ASAP to maximize your chances of being accepted!

Assessments

- Classical exams, projects, homework, continuous assessment.
- Submission: **GitHub** or **Moodle**.
- GitHub: make sure to **follow the GitHub anonymity naming convention**; see pages 36 of the **handbook**.
- **Continuous assessments**:
 - a solution for **each assessment** needs to be **submitted**!
 - otherwise, it will be zero absent/incomplete $\Rightarrow$ overall course



Plagiarism, academic misconduct

- pretty serious offense on any submitted work,
- submission must be yours,



- you must acknowledge and refer to all sources used: websites, ChatGPT, generative AI, ...
- do not copy word-by-word (unless quoted), transform it to your world; inspiration $\in$ learning,
- further info: **LSE regulations**.

Capstone projects

real-world data science project with a company (Nov - Aug)



Capstone project – continued

- Process: bidding, followed by joint happiness maximization.



- Important dates:
 - Nov. 2024: assignment of project topics to groups,
 - Aug. 2025: project report submission deadline.
- Project coordinator: Marcos Barreto.

Handy tools/skills to pick up, FAQ

- Q: Could you help in my visa-related question X?
- A: Sarah and Me/leva are legally *not* allowed to give you visa advice ⇒ **log a query** with the Student Advice and Engagement Team.

- Q: Could you help in my visa-related question X?
 - A: Sarah and Me/leva are legally *not* allowed to give you visa advice ⇒ **log a query** with the Student Advice and Engagement Team.
-
- Q: Can I go on family visit/vacation after Capstone submission?
 - A: Yes; let us (= dept.) know, then enjoy the trip! Note: stay in London within reasonable distance during term-time.

- Q: Could you help in my visa-related question X?
- A: Sarah and Me/leva are legally *not* allowed to give you visa advice ⇒ **log a query** with the Student Advice and Engagement Team.

- Q: Can I go on family visit/vacation after Capstone submission?
- A: Yes; let us (= dept.) know, then enjoy the trip! Note: stay in London within reasonable distance during term-time.

- Q: Are mandatory courses mandatory?
- A: Yes.

- Q: Softwares, tools and techniques to absorb?
- A: Selection = strong quantitative background: ✓, congrats!

Let us dive into this question!

Typesetting: $\text{\LaTeX}$

- high-quality typesetting system ($\text{\TeX}$) with macro packages ($\text{\LaTeX}$),
- creator:
 - The Art of Computer Programming (book, 7 volumes): publisher with poor typesetting $\Rightarrow$ Donald Knuth ($\text{\TeX}$, '78),
 - version number: 3.141592653 (in 2021) $\rightarrow \pi$,
 - coding errors in $\text{\TeX}$: $\$2.56 = 256$ cents, $\times 2$.



- Leslie Lamport ($\text{\LaTeX}$, '84).

$\text{\TeX}$ = high-quality article/book/TR/poster/... with minimal effort

- gold standard in ML,
- cross-platform: Windows, MacOS, Linux,
- CTAN = packages, ...
- collaborative editing: Overleaf ($\leftrightarrow$ Git),
- style: WYSI $\underbrace{\text{WYW}}$ instead of (all)WYSI $\underbrace{\text{WYG}}$,
what you want what you get
- 1/2h intro to L^AT_EX.

Top ML venues $\Rightarrow$ pick up the writing 'etiquette'

- Terminology:

AI $\approx$ ML $\approx$ data science $\approx$ applied maths $\approx$ data-centric engineering

Top ML venues $\Rightarrow$ pick up the writing 'etiquette'

- Terminology:

AI $\approx$ ML $\approx$ data science $\approx$ applied maths $\approx$ data-centric engineering

- Mathwise:

$\underbrace{\text{statistics}}_{\text{guarantees}}, \underbrace{\text{optimization}}_{\text{computation}}, \underbrace{\text{functional analysis}}_{\text{function spaces}}$

Top ML venues $\Rightarrow$ pick up the writing 'etiquette'

- Terminology:

AI $\approx$ ML $\approx$ data science $\approx$ applied maths $\approx$ data-centric engineering

- Mathwise:

$\underbrace{\text{statistics}}_{\text{guarantees}}, \underbrace{\text{optimization}}_{\text{computation}}, \underbrace{\text{functional analysis}}_{\text{function spaces}}$

- Top-tier **conferences** [acceptance $\sim 20 - 25\%$]:

- Conference on Neural Information Processing Systems (**NeurIPS**),
- International Conference on Machine Learning (**ICML**),
- International Conference on Artificial Intelligence and Statistics (**AISTATS**),
- Conference on Uncertainty in Artificial Intelligence (**UAI**),
- Conference on Learning Theory (**COLT**).

Top ML venues $\Rightarrow$ pick up the writing 'etiquette'

- Terminology:

AI $\approx$ ML $\approx$ data science $\approx$ applied maths $\approx$ data-centric engineering

- Mathwise:

$\underbrace{\text{statistics}}_{\text{guarantees}}, \underbrace{\text{optimization}}_{\text{computation}}, \underbrace{\text{functional analysis}}_{\text{function spaces}}$

- Top-tier **conferences** [acceptance $\sim 20 - 25\%$]:

- Conference on Neural Information Processing Systems (**NeurIPS**),
- International Conference on Machine Learning (**ICML**),
- International Conference on Artificial Intelligence and Statistics (**AISTATS**),
- Conference on Uncertainty in Artificial Intelligence (**UAI**),
- Conference on Learning Theory (**COLT**).

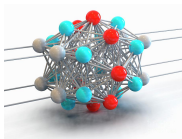
- Top **journal**:

- Journal of Machine Learning Research (**JMLR**).

Top conferences+

Great 'app-specific' conferences :

- NLP:
 - Empirical Methods in Natural Language Processing (EMNLP).
- deep & representation learning:
 - International Conference on Learning Representations (ICLR).
- computer vision:
 - Conference on Computer Vision and Pattern Recognition (CVPR),
 - International Conference on Computer Vision (ICCV).



Publication morale, mentality

- conference-oriented! (vs stats),
- extension to journal (+theory/app),
- open (typically),
- reproducibility: more & more emphasis.

History:

- open-access repo of preprints,
- hosted by Cornell University,
- launched in 1991 (Paul Ginsparg; T_EX made it possible),

History:

- covered areas: **mathematics**, **computer science**, physics, quantitative biology, quantitative finance, **statistics**, electrical engineering and systems science, economics.

History:

- covered areas: [mathematics](#), [computer science](#), physics, quantitative biology, quantitative finance, [statistics](#), electrical engineering and systems science, economics.
- [submission rate](#) (2024): 244,041 articles/year; [further stats](#).

- Categories [1st approximation]:
 - stat.ML: statistical machine learning ([stat](#)),
 - cs.LG: learning ([CS](#)),
 - math.OC: optimization ([maths](#)).
- Description of sub-categories in statistics and CS, classification:
 - [statistics](#), [CS](#),
 - [ML classification guide](#).

- Categories [1st approximation]:
 - stat.ML: statistical machine learning (**stat**),
 - cs.LG: learning (**CS**),
 - math.OC: optimization (**maths**).
- Description of sub-categories in statistics and CS, classification:
 - **statistics**, **CS**,
 - **ML classification guide**.
- Each paper has 1 primary and typically 1-2 secondary categories.
- In addition: **MSC** (=maths) and **ACM** (=CS) classes.

Mailing lists

- ML:
 - [ML-News](#),
 - [UAI](#).
- Statistics: [allstat](#).
- ML + neuroscience: [Connectionists](#).
- Academic social networking: [LinkedIn](#).
- Computing job announcements: [CRA](#).
- Jobs in the UK: [jobs.ac.uk](#).

Programming languages

- **Python:**
 - gold standard in data science.
 - general-purpose language $\xrightarrow{\text{specialized}}$ ML.
 - Good book: Mark Lutz. **Learning Python**, 6th Edition. O'Reilly Media, Inc., 2025.
- **R:**
 - widely-used in statistics.

@ LSE

Digital Skills Lab: pre-sessional courses (**Python**, **R**, $\text{\LaTeX}$).

OpSystem

- : Windows, MacOS, Linux, else.



OpSystem

- : Windows, MacOS, Linux, else.



- Linux:
 - idea: user freedom $\Rightarrow$ you can own your computer.

OpSystem

- **?**: Windows, MacOS, Linux, else.



- Linux:
 - idea: user freedom $\Rightarrow$ you can own your computer.
 - free $\Rightarrow$ community-driven, versatile (e.g. Python-customizable WM), secure (no viruses) & private (no spying / dummy ads), resource-efficient (old laptop $\rightarrow$ supercomputers), ...

- **?**: Windows, MacOS, Linux, else.



- Linux:
 - idea: user freedom $\Rightarrow$ you can own your computer.
 - free $\Rightarrow$ community-driven, versatile (e.g. Python-customizable WM), secure (no viruses) & private (no spying / dummy ads), resource-efficient (old laptop $\rightarrow$ supercomputers), ...

Getting Started with Linux mini-course

- CKK 2.04, 14:00-16:30, Friday (Sept. 26),
- You are welcome to sign up and join!

Enjoy your time at the department and LSE!

