Course proposal printable view

DS105L Data for Data Science (2022/23)

1 General

1.1) Title of the Course: Consideration should be given to the course title in the context of existing provision of taught courses by all academic departments in the School. Proposers and editors should cross-check proposed titles with the list of available courses in the Calendar and consult with colleagues as necessary, especially if a course with a similar or closely related title already exists. This is to ensure that no significant overlap or duplication occurs.

Data for Data Science

1.2) Departments/institutes/groups

Department *	Is this the lead department? *
Data Science Institute	yes

1.3) Unit value (1 = full-unit, 0.5 = half-unit, 0 = non-assessed / non-degree language course)

0.5

1.4) What level is the course?

Undergraduate 1

1.5) Proposed academic session in which course is first to be taught

2021

1.6) Will this course be taught over 1 or 2 academic sessions?

1

1.7) Please specify for how many years this course will run (not specified means indefinitely)?

1.8) If this course shares teaching with another course, please provide the course title or code below by searching in the box below.

1.9) Pre-requisites: Please indicate any specific requirements students will need in order to take this course. Courses listed in this table will be displayed in the course guide as being required pre-requisites. If multiple courses are listed, they will be displayed as "course 1 and course 2" in the course guide.

1.10) Pre-requisites - additional content for the Course Guide. If general pre-requisites are required, or if a combination of courses is required that should be displayed in the course guide as "course 1 or course 2" or "course1 or equivalent" for example, please enter the text here. Please write references to specific courses using the following format: '*Course Title* (course code)'.

1.11) General notes for colleagues when considering this proposal (for example, you may wish to propose a code for this course)

The impetus for introducing this module is the creation of the BSc in Politics and Data Science. But we also anticipate that it would be a suitable foundational course for any BSc in X + Data Science, or for students in quantitative or data science-related fields who want to know more about data manipation from a data science perspective.
We propose the course code DS105.

2 Demand

2.1) Why is the course being proposed? How does it complement and/or extend existing provision within the programme(s) it will be taught on?

Understanding data and its structure is a core element of any foundational education in data science. This module will cover these fundamentals in a comprehensive and systematic way, including the import, manipulation, conversion, linkage, storage, and retrieval of structured and semistructured data. It also covers key skills in publishing and collaborating online, which are not just hallmarks of the modern information age, but also the principal means for data scientists to work together on development and analytic projects and to disseminate results. Knowing how the core systems of online platforms work is a core part of a solid practical education in data science. It also provides tools and understanding of technical aspects of data that will be key elements for use in more advanced data science courses.

2.2) Please identify any existing course(s) to be discontinued by the introduction of this course

2.3) Please give expected approximate number of full-time students of the course over the next five years (n.b. the maximum number class/seminar size is 15 students per group)

Term *	Number of students *
Year 1	40
Year 2	50
Year 3	55
Year 4	60
Year 5	60

2.4) How many seminar/class groups will be offered (n.b. the maximum class/seminar class is 15 students per group)? Additional information about the School's policy on seminar and class size limits can be accessed here.

Term *	Number of groups *
Year 1	3
Year 2	
Year 3	4
Year 4	4
Year 5	4

2.5) Please explain the basis of above approximated figures

These are based on the intake in Year 1 of 29 students in the BSc in Politics and Data Science, plus a few other students who have been admitted by request. In later years we expect this course to expand as demand increases.

3 Content

3.1 Course content

This course will cover the fundamentals of data, with an aim to understanding how data is generated, how it is collected, how it must be transformed for use and storage, how it is stored, and the ways it can be retrieved and communicated. The course will also cover workflow management for typical data transformation and cleaning projects, frequently the starting point and most time consuming part of any data science project. This course uses a project-based learning approach towards the study of online publishing and group-based collaboration, essential ingredients of modern data science projects.

It introduces the principles and applications of the electronic storage, structuring, manipulation, transformation, extraction, and dissemination of data. This includes data types, how data is stored and recorded electronically, the concept and fundamentals of databases. It also covers how data is formatted and communicated. It presents basic methods for obtaining data from the Internet, including simple methods for web scraping and the use of APIs to submit queries that return structured data. Finally, it covers methods for formatting and publishing data.

Sharing and publishing data will also form a key part of this module and will include key skills in on-line publishing, including the elements of web design, the technical elements of web technologies and web programming, as well as the use of revision-control and group collaboration tools such as GitHub. Each student will build an interactive website based on content relevant to their domain-related interests, and will use GitHub for accessing and submitting course materials and assignments. The final project will involve group work to create a data-based website published on GitHub.

This module is not designed to be a hands-on introduction to the use of databases, but does introduce the concepts of databases. For more detailed learning on databases, we will encourage students to take *ST207 Databases*.

3.2 Availability

3.2.1) Please identify programmes on which course is to be listed

Programme	Programme stream	Compulsory	Paper number	Year
BSc in Politics and Data Science (UBPDS)		no	4	1

3.2.2) Will this course be available to General Course Student? (for UG courses only)

yes, this course is freely available

3.2.3) Will this course be available as an outside option or to students in other departments?

yes, this course is freely available

3.2.4) Availability - additional content for the Course Guide

Rationale

09/11/2021 TQARO: Added to BSc in Politics and Data Science (UBPDS).

3.2.5) Some graduate courses are designated as 'controlled access' due to limited places and/or prerequisites that are required in order to study the course. Students will need to apply to the teaching department for permission to take the course in LSE for You before being allowed to register for it. Will you control access to this course?

yes

3.2.6) On the LSE for You course choice system for undergraduate students some courses are labelled as 'capped'. As soon as the number of students registered reaches the capped number, the status of these courses will change to "full" and no one else will be able to select them. They are managed on a first come first served basis. If you intend to cap your course, what is the maximum number of students you will be able to teach?

45

3.3 Teaching

3.3.1) Teaching: Please state the number of hours of teaching each term by lecture and seminar/class teaching hours. Please click here for additional guidance about teaching. Please note that it is the responsibility of the department to ensure that the information in the timetable and the course guide matches.

Type	MT sessions	LT sessions	ST sessions
Lecture	0 x 0 min	20 x 50 min	0 x 0 min
Class	0 x 0 min	9 x 90 min	0 x 0 min

3.3.2) Information included here will appear in the course guide. Please add further information about core teaching detailed in 3.3.1 if necessary. If you will provide structured learning activities as part of this course (i.e. activities the department makes available to students during a Week 6 reading week that do not form part of the core teaching but which would benefit students' learning experiences), please provide details here. Please click here for additional guidance about teaching.

A combination of classes and lectures totalling a minimum of 33.5 hours (counting 50 mins as an hour) across Lent Term, with a reading week in Week 6.

Rationale

3.3.3) Learning outcomes: please provide details of the learning outcomes for the course. Please click here for additional guidance about learning outcomes. Students will be able to ...

By the end of this course students will be able to:

- Understand the basic structure of data types and common data formats
- Show familiarity with international standards for common data types
- Manage a typical data acquisition, cleaning, structuring, and analysis workflow using practical examples
- Clean data, and diagnose common problems involved in data corruption and how to fix them
- Understand the concept and fundamentals of databases
- Link data is linked from different sources
- Manage a typical data acquisition, cleaning, structuring, and analysis workflow using practical examples and real database applications.
- Use the collaboration and version control system GitHub, based on the git version control system.
- Understand the fundamentals of “markup” languages, including Hypertext Markup Language (HTML), Extensible Markup Language (XML), and the Markdown format for formatting documents and web pages.
- Create and maintain simple websites using HTML and CSS
- Use APIs to send and retrieve data from Internet sources

3.3.4) Teaching and learning methods: please describe how the teaching and learning methods help students attain the learning outcomes. Please also explain how the teaching on this course will use a variety of methods to enable all students to engage with the teaching and learning process. For example, through the use of interactive learning methods such as workshops and online activities or diverse approaches to giving feedback such as face-to-face or online communication.

The class will meet weekly for 10 x 2 hour lectures, plus 10 corresponding classes consisting of structured lab exercises lasting 1.5 hours each. Lab sessions will be allocated to structured exercises that students will complete and submit following the lab sessions, for assessment.

Student exercises and labs will use cloud computing platforms in the form of Jupyter Notebooks and/or RStudio Server platforms.

GitHub Classrooms will provide the means of distributing and submitting coursework.

3.3.5) Please provide a rationale for the course assessment (both formative and summative assessment), including its timing. Please explain how the methods of assessment have been chosen to test whether students are working towards/have met the course learning outcomes and take into account inclusive approaches to assessment. For example, is there a range of different assessments including group work, presentations, etc., and sufficient training and development for the assessments that are being used, for example, presenting skills? Are there clear guidelines for different methods of assessment processes? You should consult the LSE Assessment Toolkit for guidance on broadening student assessment.

This course is designed to provide practical learning alongside a introduction to the basics of the technologies covered. Interactive and handson learning through practical exercises will supplement the readings and lectures. Just as with learning a language, we see the practical component as essential, and expect that students will have to participate in this part very actively for effective learning to take place.

Summative assessment will take place through a final assignment that consists of publishing a web page involving data and the dissemination of data using on-line tools. Publication will take place using a student repository on GitHub to

publish a website that is linked to a database, using GitHub pages. This will be prepared through working in assigned groups. Each page will have mandatory components but be tailored to the student's particular interests and expertise. Students will work on their data and the web page throughout the second half of the course, but be expected to have it completed by the start of the exam period.

3.4 Summative assessment

3.4.1) Summative Assessment

Assessment 1

Coursework

Weight (%) *	Type *	Number of words	Timing of submission
60.0	coursework	1000	ST

Others

Weight (%) *	Type *	Further details	Timing of submission
40.0	group project	Website based on data	ST

3.4.2) Summative assessment - additional content for the Course Guide

Rationale

3.5 Formative coursework

3.5.1) Formative coursework is an essential part of the teaching and learning experience at the School. It should be introduced at an early stage of a course and normally before the submission of assessed coursework. Students will normally be given the opportunity to produce essays, problem sets or other forms of written work in preparation for summative assessment. Feedback on formative work should help students to understand how they can improve their performance in readiness for summative assessment.

Please enter details about formative coursework in the table below ('type', 'number' and 'term' will appear in the course guide, along with any additional information entered in the 'additional content' free text box below).

When proposing or making modifications to formative coursework, departments should consider how issues of inclusivity and accessibility have been considered in all areas of teaching, learning and assessment. Have students had sufficient development and practice to prepare them for completing the summative assessment/s? Please see resources at Teaching and Learning Centre or contact tlc@lse.ac.uk or lti@lse.ac.uk for more information.

Type *	Title	Number *	Term *	Notes

Type *	Title	Number *	Term *	Notes
Other	Problem sets	5	LT	Could be groupwork-based

3.5.2) Formative coursework - additional content for the Course Guide

Students will work on weekly, structured problem sets in the staff-led class sessions. Example solutions will be provided at the end of each week.

3.6 Reading list

3.6.1) Reading list: this should list 5-10 essential readings only.

- Duckett, Jon. *HTML and CSS: Design and Build Websites.* New York: Wiley, 2011.
- Lake, Peter. *Concise Guide to Databases: A Practical Introduction.* Springer, 2013.
- Sklar, David Learning PHP 5 O'Reilly, 2004. GitHub Guides at https://guides.github.com, including: "Understanding the GitHub Flow", "Hello World", and "Getting Started with GitHub Pages".
- Jacobson, Daniel *APIs: A Strategy Guide*. O'Reilly: 2012.
- Zafarani, R., Abbasi, M. A. and Liu, H. (2014) *Social Media Mining: An introduction*. Cambridge University Press.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

3.6.2) Reading list - additional content for the Course Guide

4 Management

4.1) Teacher(s) responsible

Person *	Room
Benoit, Kenneth (KBENOIT)	PEL.4.01C

4.2) Teacher(s) responsible - additional content for the Course Guide

5 Resources

5.1 Staffing

5.1.1) Will the introduction of the course require any net addition to department teaching resources? In other words will your department need to put a bid to APRC for additional resource in order for this course to be delivered?

no

5.1.2) Academic

Number *	Level/grade *	Source of funding *
1	SBA1	MSL

5.1.3) Part-time

5.1.4) Technical support

Number *	Level/grade *	Source of funding *
1	SB06	MSL

5.2 Teaching and accommodation

5.2.1) Are the lectures shared with other courses?

no

5.2.2) The maximum size for a class/seminar is 15 students. Only under exceptional circumstances will a course receive exemption from this rule. Does this new course require a three year exemption from the rule? Additional information about the School's policy on seminar and class size limits can be accessed here.

no -

5.2.3) Will any additional teaching space be required?

no

5.3 Library

5.3.1) Is the new course in a subject that falls outside the main social sciences already taught/researched at the LSE?

no

5.3.2) Does the new course require material in languages outside those already collected by the Library?

no -

5.3.3) Does the new course require significant recurrent expenditure on print/electronic journals/datasets?

no -

5.4 IT Services

5.4.1) Does the proposed course require use of computer software other than word processing and spreadsheet applications for front of house (teacher's computer in a classroom)?

yes - Some interactive data science exercises using cloud computing methods, but these will be configured and funded by DSI staff.

5.4.2) Does the proposed course require use of computer software other than word processing and spreadsheet applications for student classroom computers?

no -

5.4.3) Is this a joint course with other institutions where access to electronic resource at the LSE is required for students physically based at another institution?

no -

5.4.4) Does the software and hardware which is currently installed on all standard computers in classrooms and computer rooms meet your course requirements?

no

5.4.5) Will students need to be taught in a PC classroom?

no

5.4.6) Will lecture capture be required?

yes

5.4.7) Will any teaching occur outside normal working hours (0900-1800) or at weekends?

no

6 Skills

6.1) Personal Development Skills

Quality *	Includes *	Justification
Leadership	no	
Self-management	yes	Through completing coursework
Team working	yes	Final project is team-based.
Problem solving	yes	Problem sets involve problem-solving.
Application of information skills	yes	Data manipulation is an information skill.
Communication	yes	Presenting results involves communication.
Application of numeracy skills	yes	Data aggregation involves basic numeracy skills.
Commercial awareness	no	
Specialist skills	yes	Using APIs is a specialist skill

6.2) HECoS codes

Code *
Data management (I260)
programming (100956)
internet technologies (100373)

7 Consultation

7.1) Has this proposal been discussed and endorsed at a departmental/institute meeting?

yes - 04/12/2020

The general approval for introducing basic modules in data science was approved as part of the DSI's strategic education plan. This plan was approved at the 1 April 2020 Education Committee meeting at the 29 April 2020 Research Committee meeting, and on the 13 May 2020 Academic Board meeting. That detailed strategic plan included indicative content for two DS modules, and this content closely matches that provided in this proposal. It arose from a

year-long Data Science Steering Group chaired by Professor Pauline Barrieu of the Department of Statistics. It received written endorsement by its members, which included Professor Milan Vojnovic (ST), Professor Kenneth Benoit (MY), Professor Lazslo Vegh (MA), and a half dozen other senior professorial members of the DSSG. It also included an explicit letter of support by the head of the departments of ST, MA, and MY. After the formation of the Data Science Institute on 1 September 2020, the Data Science Education Forum (DSEF) was formed and given the task of overseeing the development of this proposal, in part to accompany the new BSc in Politics and Data Science programme proposal. This group met on: 11 November, 24 November; 4 December 2020, and 4 January 2021. At the last two of these meetings, the content of this course was discussed and approved.

7.2) Has this proposal been discussed at a staff/student meeting?

no

Not applicable because MY has no undergraduate students who might have attended such a meeting.

7.3) Has this proposal been discussed with the Teaching and Learning Centre?

yes - 04/12/2020

7.4) Colleagues with related interests in other departments/institutes will need to have been consulted. You might find it useful to refer to the LSE Experts information. The Sub-Committee Secretaries are happy to provide individual advice on who to consult with.

Person *	Department/Institute	Consultation date *	Notes	Were any objections raised? *
Jackson, Jonathan (JACKSOJP)	Methodology	18/12/2020		no
Steele, Fiona (STEELEF)	Statistics	04/12/2020	We revised the content based on comments from Statistics (greatly reduced the database emphasis).	no
Schonhardt-Bailey, Cheryl (SCHONHAR)	Government	04/12/2020		no

Consultation occurred through the Data Science Education Forum, meeting of 4 December 2020, which included representatives from the following departments: - Department of Statistics - Department of Mathematics - Department of Government - Department of Geography and Environment The DSEF is also chaired by the Director of the Data Science Institute, who has an overview of data science education at the School. Clare Gordon of the Eden Centre and Jeni Brown of the Digital Skills Lab also attended.

version: 1.3.13