**UCL SUMMER SCHOOL**

**DATA SCIENCE AND BIG DATA ANALYTICS**

**Key Information**

<table>
<thead>
<tr>
<th>Module code</th>
<th>ISSU0053</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taught during</td>
<td>Session Two: Monday 22 July - Friday 9 August 2019</td>
</tr>
<tr>
<td>Module workload</td>
<td>45 teaching hours plus approximately 100 study hours</td>
</tr>
<tr>
<td>Module leader</td>
<td>Lisa Chalaguine</td>
</tr>
<tr>
<td>Department</td>
<td>Department of Computer Science</td>
</tr>
<tr>
<td>Credit</td>
<td>15 UCL credits, 7.5 ECTS, 4 US</td>
</tr>
<tr>
<td>Level</td>
<td>Level 2, second year Undergraduate</td>
</tr>
<tr>
<td>Pre-requisites</td>
<td>Successful completion of a first year undergraduate level module in statistics and experience of using statistical computer software</td>
</tr>
<tr>
<td>Assessment</td>
<td>Practical assessments (50%)</td>
</tr>
<tr>
<td></td>
<td>Final exam (50%)</td>
</tr>
</tbody>
</table>

**Module Overview**

Data Science is an exciting new area that combines scientific inquiry, statistical knowledge, substantive expertise, and computer programming. One of the main challenges for businesses and policy makers when using big data is to find people with the appropriate skills. Students taking this module will be introduced to the most fundamental data analytic tools and techniques, and learn how to use specialised software to analyse real-world data and answer policy-relevant questions.

**Week One**

- Principles of research design;
- Probability framework and statistical inference;
- Linear regression models.

**Week Two**

- Classification models;
- Resampling methods;
- Model selection.

**Week Three**

- Non-linear models;
- Tree-based models;
- Unsupervised learning;
- Unstructured data analysis.

Please note that this module description is indicative and may be subject to change.
Module Aims
This course aims to provide an introduction to the data science approach to the quantitative analysis of data using the methods of statistical learning, an approach blending classical statistical methods with recent advances in computational and machine learning. The course will cover the main analytical methods from this field with hands-on applications using example datasets, so that students gain experience with and confidence in using the methods we cover. It also covers data preparation and processing, including working with structured databases, key-value formatted data (JSON), and unstructured textual data.

Teaching Methods
Lectures and seminars (computer labs), student presentations, classroom debates, private reading and outlines/assignments. Reading lists will be available online via the UCL library site and the readings will be regularly revised and updated. Student support will be provided via seminars/tutorials (computer lab sessions) and office hours. Relevant materials and forums will also be housed on the Moodle system.

Learning Outcomes
Upon successful completion of this module, students will:

- Have a sound understanding of the field of data science and develop the ability to analyse real-world data using some of its main methods;
- Become comfortable applying regression models for continuous and limited outcome variables;
- Explore more complex models, such as the widely-used panel data models;
- Develop familiarity with descriptive and predictive analytics, and their application to big data problems;
- Explore methods of text analytics and automated data acquisition;
- Have received a solid foundation for more advanced or more specialised study.

Assessment Methods
- Practical assessments (50%)
- Final exam (50%)

Key Texts
The primary texts are:

The following are supplemental texts which you may also find useful: