Subject Description Form

Subject Code	EIE1003 (for 42477)
Subject Title	Foundations of Data Science
Credit Value	3
Level	1
Pre-requisite/ Co-requisite/ Exclusion	Nil
Objectives	Being able to discover useful knowledge and information from a large amount of data is very critical to industry, business and government. This subject aims to provide students the fundamental concepts of data science and the basic technologies for data analytics. It provides hands-on experiences in data analytics and case studies in applications of data science in engineering, social science, healthcare, business and government. It also prepares students with the right mentality towards data and the ability to leverage data for decision-making.
Intended Subject Learning Outcomes	**Upon completion of the subject, students will be able to:** Category A: Professional/academic knowledge and skills 1. Understand the basic concepts and technologies of data science. 2. Acquire the basic technical know-how on data analytics. Category B: Attributes for all-roundedness 3. Understand the data-driven process for problem solving. 4. Demonstrate how to harness and process data for decision-making.
Subject Synopsis/ Indicative Syllabus	1. Introduction to Data Science • Data science vs. big data vs. data analytics • Benefits of data science • Skill sets required • Privacy, security and ethics • Example applications and case studies 2. Technologies for Data Science • Basic concepts in summary statistics • Graphs and plots for data analytics, e.g., box plots, scatter plots, histograms, run charts, etc. • Example case studies of exploratory data analytics for data science • Fundamental of machine learning for data science • Cloud technologies 3. Tools for Data Science • Data cleaning, e.g., OpenRefine • Machine learning tools, e.g., Microsoft ML Studio, Weka • Data visualization tools, e.g., Google Chart, Tableau 4. Applications with Case Studies • Recommendation systems • Spam filtering • Stock prediction • Social networks • Sentiment analysis

<table>
<tr><td>Teaching/Learning Methodology</td><td>
Lectures: The subject matters will be delivered through lectures (both in-person and online ones). Students will be engaged in the lectures through Q&A, discussions and specially designed classroom activities. Practitioners and software vendors will be invited to give guest lectures.

Tutorials and Workshops: Students will work on data analytics projects using software tools. Students will start from small and easy projects in the first half of the subject. In the second half, students will work on a more realistic project that solves real-world problems, using the knowledge and know-hows that they have learnt from the small projects.

Assignment: Students will need to do a group-based mini-project on data science.
</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Subject Learning Outcomes</td><td>
<table>
<tr><th rowspan="2">Specific Assessment Methods/Tasks</th><th rowspan="2">% Weighting</th><th colspan="4">Intended Subject Learning Outcomes to be Assessed (Please tick as appropriate)</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>1. Continuous Assessment (total: 100%)</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>• Mini-project (proposal, report and presentation)</td><td>30%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>• Exercises</td><td>16%</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>• Tests</td><td>30%</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>• Laboratory activities</td><td>24%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Total</td><td>100%</td><td colspan="4"></td></tr>
</table>
Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

Laboratory exercises and mini-project will require students to apply what they have learnt to solve problems. There will be open-ended questions that allow students to exercise their creativity in solution design.

Tests and Exercises assess students' achievement of the learning outcomes in a more formal manner.

Mini-project is group-based and weights 30% of the whole assessment. Among the 30% weight, 7% is for proposal, 13% is for final report, and 10% is for presentation (in the form of a 10-minutes video). Proposal and report (20% in total) are evaluated based on group, while presentation (10%) is evaluated individually. Each group member will present the part he/she is responsible for in the mini-project. The mini-project will make use of publicly available tools such as Microsoft Azure Machine Learning Studio so that requirements on programming knowledge is kept to a minimum, i.e., no programming background is assumed. Students will perform drag and drop of data sources, machine learning models, analytic methods, and evaluation methods from the tool to solve data science problems. Enthusiastic students could use the cloud-based API to perform more complex tasks.

Tests and Exercises weight 46% and they are individual assessments.

Laboratory activities weight 24% and they are individual assessments.

Overall, 80% of the assessment is individual assessment and 20% is group-based assessment.
</td></tr>
</table>

Student Study Effort Expected	**Class contact (time-tabled):**	
	• Lectures (In-person and online)	22 Hours
	• Tutorial/Laboratory/Practice Classes	17 Hours
	Other student study effort:	
	• Lecture: preview/review of notes; homework/assignment; preparation for test/quizzes	30 Hours
	• Tutorial/Laboratory/Practice Classes: preview of materials, revision and/or reports writing	36 Hours
	Total student study effort:	**105 Hours**
Reading List and References	**Reference Materials:** 1. L. Cao, *Data Science Thinking: The Next Scientific, Technological and Economic Revolution*. Cham: Springer International Publishing, 2018. 2. L. Igual and S. Sequi, *Introduction to Data Science: a Python Approach to Concepts, Techniques and Applications*. Cham, Switzerland: Springer, 2017. 3. S. Alan and V. Torra, *Data Science in Practice*. Cham, Switzerland: Springer 2019. 4. G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*. Cham, Switzerland: Springer 2019. 5. P. Kromer and R. Jurney, *Big Data for Chimps*. O'Reilly, 2016. 6. T. Ojeda et al., *Practical Data Science Cookbook*. Packt Publishing Ltd, 2014. 7. A. Adhikari and J. DeNero, *Computational and Inferential Thinking: The Foundations of Data Science*. https://www.inferentialthinking.com/	
Last Updated	Aug 2021	
Prepared by	F. Leung, Man-Wai Mak	