

## DATA 300 – Introduction to Data science

We are living in an era in which the kind of problems that could be solved using data are driving a huge wave of innovations in various industries – from healthcare to education, and from finance to policymaking. Data Literacy is becoming a fundamental skill for everyone. With that motivation, this course introduces the field of Data Science in a practical manner, using a hands-on approach that assumes no prior knowledge on the subject. The focus is on solving problems using Python or R or any other language of choice. Students will be introduced to several real-life problems that involve collecting and analysing both small and big data. They will also study common methods for quantitative and computational analytics through which they can have an overview of key concepts, skills, and technologies used by data scientists. It presents material that will continue to be relevant even when tools and technologies change. Hands-on exercises include analyses of social media data from services such as Twitter (now X) and YouTube, as well as working with large Yelp datasets with more than a million records.

Data Science already has seen its impact across a spectrum of societal activities – in medicine, in transportation, in policy making, in business development, in education, and so on. In most cases, Data Science tends to advocate automation. In doing so, it creates new markets and new job opportunities where insights developed by machines supplement human intelligence and decision making. Data Science has a big say in upcoming changes in social norms influenced by the Fourth Industrial Revolution (4IR) and Cognification. Subsequently, way of life such as contemporary business processes, data privacy, and governmental regulations will be forced to undergo disruptive changes. A deeper understanding of Data Science will assist the society in general adapt to such changes.

### **Learning Outcomes:**

CLO1	Assess and articulate the relevance of data for a particular business, healthcare, educational or other societal challenges.
CLO2	Collect, store, retrieve, and preprocess data.
CLO3	Undertake different kinds of data analysis and data analytics.
CLO4	Exhibit familiarity with data science methods by learning and experiencing essential algorithms and approaches.
CLO5	Identify data-driven analytics problems, and design solutions, and develop code to solve them.

Every study activity of this course, whether it be a reading activity or an exercise or an assignment, is aimed at one or more of these learning outcomes.

## Text:

- A hands-on introduction to Data Science, by Chirag Shah, 2020, Cambridge Press. <https://www.cambridge.org/highereducation/books/a-hands-on-introduction-to-data-science/9D55C29C653872F13289EA7909953842#overview>

## Programming Languages:

Students are encouraged to use Python to complete the exercises and assessment problems. Examples in the textbook employ R. With the instructor's permission, students are welcome to use Python, R, or any other language of their choice.

## Course Outline

Students are encouraged to follow the timeline and content flow outlined in the table below. Do connect with the instructor in case of significant lapse in study timeline. As always, connect either with the Academic Expert/Tutor or the Course Coordinator regarding any content/concept clarification.

Unit		Title/Topic(s)	Sections/chapters in the book
1	Week 1	Part I <ul style="list-style-type: none"><li>• Course orientation</li><li>• Contextualising Data Science</li><li>• Quiz 0 due</li></ul>	<ul style="list-style-type: none"><li>• Sections 1.1-1.5, 1.7-1.8</li></ul>
2	Week 2	<ul style="list-style-type: none"><li>• Data types</li><li>• Data collections and storage</li><li>• Data pre-processing</li></ul>	<ul style="list-style-type: none"><li>• Sections 2.1-2.4</li></ul>
3	Week 3	<ul style="list-style-type: none"><li>• Data analysis</li><li>• Descriptive analysis</li><li>• Exploratory analysis</li><li>• Mechanistic analysis</li></ul>	<ul style="list-style-type: none"><li>• Sections 3.1-3.2</li><li>• Section 3.3</li><li>• Section 3.7</li><li>• Section 3.8</li></ul>
4	Week 4	<ul style="list-style-type: none"><li>• Data Analytics</li><li>• Diagnostic analytics</li><li>• Predictive analytics</li><li>• Prescriptive analytics</li></ul>	<ul style="list-style-type: none"><li>• Section 3.4</li><li>• Section 3.5</li><li>• Section 3.6</li></ul>
	Week 5	Quiz 1 is due	
5	Week 6	Part III <ul style="list-style-type: none"><li>• Machine learning introduction and regression</li></ul>	<ul style="list-style-type: none"><li>• Section 8</li></ul>

6	Week 7	<ul style="list-style-type: none"> <li>Supervised learning</li> </ul>	<ul style="list-style-type: none"> <li>Section 9</li> </ul>
7	Week 8	<ul style="list-style-type: none"> <li>Supervised learning</li> </ul>	<ul style="list-style-type: none"> <li>Section 9</li> </ul>
8	Week 9	<ul style="list-style-type: none"> <li>Unsupervised learning</li> </ul>	<ul style="list-style-type: none"> <li>Section 10</li> </ul>
9	Week 10	Quiz 2 is due Time to start the project	
10	Weeks 11 & 12	Part IV <ul style="list-style-type: none"> <li>Solving problems with big data</li> </ul>	<ul style="list-style-type: none"> <li>Sections 11.1-11.4</li> </ul>
11	Week 13	<ul style="list-style-type: none"> <li>Evaluating the goodness of a model</li> </ul>	<ul style="list-style-type: none"> <li>Section 12.4</li> </ul>
12	Weeks 14, 15 and 16	<ul style="list-style-type: none"> <li>Project due</li> <li>Preparations for [take-home + oral assessment] final exam</li> </ul>	

## Assessments

- Quiz 0 – 5%
- Quiz 1 – 15%
- Quiz 2 – 20%
- Project – 40%
- Take-home final exam with oral assessment – 20%