Week	Tasks	Overview
T2: Week 1 13 January Workshop 1 (video)	Textbook Ch.1&2 A short talk about an algorithm for human attraction: Christian Rudder: Inside OKCupid: The math of online dating A great (also short) talk about using data to tell stories: Making data mean more through storytelling | Ben Wellington	**Data analytic thinking:** A broad overview of the different topics in business analytics. Business analytics as a leadership problem. The goal of this class is to prepare you to lead in a data-driving organization, or to help create the vision of a data-driven organization. How do you decide which models are most reliable? How do you recruit or manage a data science team? How do you persuade other colleagues and management about the proper course of action using data?
T2: Week 2 20 January	CRISP_DM – Read and review – you will be using this for your final project. A data analytics pipeline: A Beginner's Guide to the Data Science Pipeline • An overview of data pre-processing: What Is Data Preprocessing? 4 Crucial Steps to Do It Right	**Managing and cleaning data:** Managing the data pipeline from the creation of new data, to processing the data, to producing results. What are the different kinds of data? How is data cleaned, stored, and made ready for analysis?
T2: Week 3 27 January	• Video: Dominic Bohan - Turning Bad Charts into Compelling Data Stories • Video: Hans Rosling, The best stats you've ever seen • Read: Storytelling with Data • Listen: Data is Personal (it was hard to pick an episode from this podcast, it's great) • RStudio primer on visualisation	**Data visualisation:** We will cover the basic elements of data visualization. We will focus on using the ggplot package. It's the most popular and most powerful visualization software used across the industry. This is the software both the BBC and the New York Times use to create their graphics.
T2: Week 4	• Textbook Ch.6	**Clusters and similarity:** A basic task in data exploration considers the similarity and groups in

3 February	• Watch: StatQuest: K-means clustering Watch: StatQuest: Hierarchical Clustering • Watch StatQuest: PCA main ideas • Watch StatQuest: Principal Component Analysis (PCA), Step by Step • Play: Visualizing K-Means Clustering • Play Visualizing DBSCAN • Play: Principal Component Analysis • Read this great description of Hierarchical Clustering • And this and this useful descriptions of distance metrics	data. We will also examine dimension reduction through PCA
T2: Week 5 10 February Peer-reviewed Homework Due 14 February 2024	• Textbook Ch. 3, 4 • Watch StatQuest: Decision Trees • Watch StatQuest: Random Forests Part 1 • Watch StatQuest: Random Forests Part 2 • Watch Decision Trees and Random Forests lectures from Nando de Freitas for more detailed explanations • Play A Visual Introduction to Machine Learning • Play Random Forest Playground	**Predictive modelling:** We will attempt to predict classes and continuous outcomes

	• Play Linear Regression (try clicking and dragging on points)	
T2: Week 6	READING WEEK	
T2: Week 7 24 February	• Textbook Ch. 5, 7, 8 • Watch StatQuest: Bias and Variance • Watch StatQuest: ROC and AUC Clearly Explained • Watch StatQuest: Cross validation • Watch StatQuest: Sensitivity and Specificity • Read AUC-ROC: a really good article	**Metrics of Evaluation.** What is a good model? How do you know if a predictive model is actually a good model and will perform well in the future?
T2: Week 8 3 March	• Textbook Ch. 10 • For reference, Text Mining for R • Listen: Text Mining in R	**Text Analytics** Digitized text is an incredibly common yet underutilized source of data in organizations. We will cover some fundamentals of text analytics.
T2: Week 9 10 March Assignment Due 14 March 2024 Time: 15:00 hours	• Textbook Ch. 9, 11 • Review: Cohen et al. 2018	**Data Driven Decisions** Now that you have evidence, what option should you take? Cohen MC, Guetta CD, Jiao K, Provost F (2018) Data-Driven Investment Strategies for Peer-to-Peer Lending: A Case Study for Teaching Data Science. Big Data 6(3):191– 213
T2: Week 10 17 March	Textbook Ch. 12, 13 Watch: Introduction to Ethical AI • Listen: Talking Machines - AI for Good and The Real World	**AI and Data Ethics** What are the risks and rewards of AI in our organizations? How do we create systems that create a better working environment while also improving productivity? How will the

	Watch: Getting Specific about Algorithmic Bias Watch: 7 minutes to understand AI – A set of UNESCO videos For more detail: • Watch: Deep Learning State of the Art (2020) | MIT Deep Learning Series	nature of work change as these technologies enter the workplace?
T2: Week 11 28 March Final project due 28 March 2024 Time: 15:00 hours	**Final project due Friday 28 March. Time: 15:00 hours**	