MONASH University

Faculty of Information Technology | 25 YEARS

MONASH BUSINESS SCHOOL

DATA SCIENCE FOR MANAGERS

A course for professionals seeking to harness the potential of Data Science.

11-13 APRIL 2016

DATA SCIENCE FOR MANAGERS

A complete tour of Data Science for the professional seeking to understand how Big Data is revolutionising the business landscape and aiming to build out Data Science capabilities within their organisations or to manage Data Science and data-analytic functions.

COURSE OVERVIEW

The impact of Data Science on modern business is second only to the introduction of computers. And yet, for many businesses the barrier of entry remains too high due to lack of knowhow, organisational inertia, difficulties in hiring the right manpower, an apparent need for upfront commitment, and more.

This course is designed to address these barriers, giving the necessary knowledge and skills to flesh out and manage Data Science functions within your organisation, taking the anxiety-factor out of the Big Data revolution and demonstrating how data-driven decision-making can be integrated into one's organisation to harness existing advantages and to create new opportunities.

Assuming minimal prior knowledge, this course provides complete coverage of the key aspects, including data wrangling, modelling and analysis, predictive-, descriptive- and prescriptive-analytics, data management and curation, standards for data storage and analysis, the use of structured, semi-structured and unstructured data as well as of open public data, and the data-analytic value chain, all covered at a fundamental level.

COURSE DESCRIPTION

The past several years have been marked by a paradigm shift in the role of data in organisations. Whereas, historically, data was retained for compliance reasons or where needed for day-to-day business operations, the advent of cheap, readily-available storage options has made organisations more inclined not to erase stored data, and the boom of equally-cheap, equally-available processing power has opened the door to advanced analytics on this stored data, unlocking the business value hidden in the bits.

Today, this trend has been taken to its extremes: data is collected by any available means, far beyond what is necessary for standard business operations or, indeed, beyond information that has clearly-defined future uses; deep analysis is done both retrospectively and on-the-fly, often driving split-second business decisions; insights are gained by combining otherwise unrelated – and historically siloed – data sets, including ones that are publicly available or that are purchasable, such as social media archives or demographic data; and long-held rules-of-thumb are being systematically replaced by quantitatively-superior data-driven decision-making.

Use of data analysis has become ubiquitous, from traditional uses such as risk analysis by banks and insurance companies to new domains such as consumer-behaviour analysis, churn prediction and efficacy measurement and optimisation for all types of customer incentives. Also emerging are intra-organisational applications and uses, such as in The Internet of Things: Big Data analytics over telemetry data from industrial appliances and networked devices (e.g., smart meters) are now used in every vertical, from manufacturing to mining, from transportation to health, from energy to cyber-security. Wherever a digital footprint can be created, data is gathered and analysed in order to model behaviours, understand causes and effects, predict the future and allow decision optimisation for profit maximisation and cost minimisation, which is why even small and medium businesses today are accumulating Big Data and experimenting with cloud-based data analytics, and why this data is proving vital for creating and maintaining their competitive advantages.

The Data Science for Managers course, encompassing an intensive 24 hours over the course of 3 days, is an offering uniquely focused on the needs of professionals in managerial positions in the data-driven world. Drawing on Monash's world-leading expertise, the experience of industry-leading practitioners and real-world case studies, the course is designed to give professionals an understanding regarding where data relevant to decision-making can be found, how it can be harnessed, what wrangling is required to make it usable and how predictive or prescriptive models can be generated from it. The course will furthermore address how to approach data privacy, how to manage issues of data storage and accessibility, and the use of data in real-time decision-making.

Upon course completion, attendees will have

- gained confidence in the management of data-analytic projects,
- learned the skills necessary to allow their organisations a pain-free migration into the "data-driven enterprise" world and to increase their organisation's foothold in data analysis, and
- acquired an understanding of the key trends in Data Science and how these are influencing the future of business.

11-13 APRIL 2016

Monash University Law Chambers
555 Lonsdale St,
Melbourne CBD

DAY 1 / MON 11 APRIL

Module 1
Introduction to Data Science

Module 2
Data as an Asset

Module 3
Data Life Cycles

Module 4
Privacy and Ethics in Big Data

DAY 2 / TUE 12 APRIL

Module 5
Data Engineering for Analysis

Module 6
Data Wrangling and Exploratory Analysis

Module 7
Statistical Modelling I Fundamentals of Statistics

Module 8
Statistical Modelling II Model Creation and Validation

DAY 3 / TUE 13 APRIL

Module 9
Visualisation

Module 10
Operationalisation and the Model Life Cycle

Module 11
Building a Data-Driven Enterprise

Module 12
Case Study

> "[The course] was absolutely invaluable to me and came at the exact right point for where we are heading in our organisation. There was a lot of stuff I had picked up through research but without actually practising anything it is always hard to have visibility of what does and doesn't work. For that alone this course has probably saved JLL many times the course fee."
>
> **– Dayton Brown**
> Project Manager
> Energy and Sustainability Services
> JLL

COURSE STRUCTURE

DAY 1 / MON 11 APRIL

Module 1 Introduction to Data Science

- Origins of Data Science and a Brief history of the Big Data revolution.
- The Big Data landscape.
- How much data is there really, and does it matter?
- Un-siloing data: use paradigms for organisational data and public data.
- Descriptive, predictive and prescriptive analysis.
- From recommendations to insights: black-box and white-box analytics.

Module 2 Data as an Asset

- The V's of Big Data: Volume, Velocity, Variability, Veracity.
- Data business strategies.
- Data sources, synergies and differentiators.

Module 3 Data Life Cycles

- The analytics value chain.
- Overview of the data analysis cycle: connecting data science to the business problems.
- Work cycle of a data scientist: wrangling, modeling and validation.
- Managing research.

Module 4 Privacy and Ethics in Big Data

- The societal impacts of Big Data.
- Privacy in Australia and global perspectives.
- Big Data ethics: history and current thinking.
- Opportunities and risks for organisations and individuals.

DAY 2 / TUE 12 APRIL

Module 5
Data engineering for Analysis

- Data Science engineering and its drivers for change.
- Data volumes, data structures, and how they vary.
- Data Science architectures: the common stages.
- The usual suspects: Distributed File Systems, Map Reduce, Spark

Module 6
Data Wrangling and Exploratory Analysis

- Determining data quality. Data cleansing.
- Entity matching.
- Imputation.
- Background modelling.
- Exploratory analysis.

Module 7
Fundamentals of Statistics

- Types of data: numerical, categorical, ordinal.
- Statistical summaries: mean, standard deviation, quantiles, correlation.
- Simple data visualisation: histograms, boxplots, time plots and scatterplots.
- Cross-tabulations.
- Causality vs. association, independence.
- Randomisation and random sampling.
- Statistical inference using bootstrapping.

Module 8
Model Creation and Validation

- Prediction: linear regression, nonparametric regression, *k*-NN.
- Forecasting: auto.arima and Error-Trend-Seasonal exponential smoothing algorithms.
- Hold-out sets, cross-validation, AIC.
- Classification: logistic regression, classification trees, SVM.
- Clustering: *k*-means, hierarchical clustering.
- Supervised vs. unsupervised vs. semi-supervised learning.
- Dimension reduction: principal components.
- Languages and environments (e.g. R, Python, MATLAB or even Excel) and standards (PMML).

DAY 3 / WED 13 APRIL

Module 9
Visualisation

- Practical and effective visualisation: beyond bar charts.
- Finding the unexpected: the role of visualisation in exploratory analysis.
- Communicating findings: the role of visualisation in communicating Data Science outputs.
- Standard tools: R, Tableau, D3.

Module 10
Operationalisation and the Model Life Cycle

- Determining the needs: on how much data must decisions be taken, how often and how quickly must they be made, how often must models be refreshed?
- Plugging into existing data paths and choosing appropriate technologies.
- Stale models and model refreshing.
- Operationalisation from a business perspective: determining value and making Data Science outputs part of standard business and decision-making processes.

Module 11
Panel: Building a Data-Driven Enterprise

- Data Science as a process, rather than as a point event.
- The role of high-level management in enabling data-driven decisions.
- The role of direct management: on the un-Gantt-ability of research.

Module 12
Case Study

SPEAKERS

Michael Brand is an Associate Professor of Data Science at Monash University.

Prior to this, he served as Senior Principal Data Scientist and the Data Science Lead in charge of Infrastructure for Data Science at Pivotal, as the CTO Group Algorithm Leader at PrimeSense Ltd., and as Chief Scientist of the Advanced Technologies Group and Head of the Speech Research Team at Verint Systems, Inc.. His publications range from number and graph theory to computation theory and bioinformatics. He holds eight patents and 10+ pending patents on topics ranging from Big Data analysis to information retrieval and from machine vision to man-machine interfaces.

Course Coordinator

Dennis Claridge is a co-founder of doubleIQ and its Business Director since 2000, concentrating on delivering sophisticated marketing and analytics capability to customers.

In the last six years his focus has been on the delivery of data engineering/analytics capability as a service. This has culminated in the development of doubleIQ's Trusted Data Marketplace which is now being used by some of Australia's largest companies. Claridge's interest lies in the use of data engineering/ analytics to create new business opportunities, on the edge or between businesses. Increasingly these ideas simply arise from the digitisation of everything and the external delivery of many of the core business functions.

Di Cook is Professor of Business Analytics at Monash University.

She is a Fellow of the American Statistical Association and Editor of the Journal of Computational and Graphical Statistics. She is well-known for her work in data visualisation, exploratory data analysis and data mining, and for developing open source software. Prof Cook has authored more than 100 papers, chapters or books, and delivered more than 100 invited talks and workshops (details can be found at **http://dicook.github.io**). Her current research includes developing new types of random forests for classification, visualisation, clustering and significance testing of massive biological data, tennis and soccer statistics and enabling statistical inference with visual exploration of data.

James Horton is a data technology advisor, strategist and connector.

He is Managing Director at Datanomics, an advisory and collaborative innovation practice focused on the development of new business and industry development models based on trust-permission based shared data exchanges. Horton has over 25 years of experience in developing and leading data-centric technology initiatives for both multinational and start-up enterprises across Australia and the Asia-Pacific region. He assists private and public sector organisations in developing strategies to maximise the value of their data and data-technology-related assets.
Horton has a Bachelor of Business from the University of South Australia and an MBA from ANU.

Rob J Hyndman is Professor of Statistics at Monash University, Director of the Monash University Business and Economic Forecasting Unit, and Editor-in-Chief of the International Journal of Forecasting.

He has received several awards for his research including the 2007 Moran Medal from the Australian Academy of Science. Prof Hyndman has authored over 140 papers, chapters or books on statistical topics, and more than 200 consulting reports (see complete publication list on **robjhyndman.com**). His current research projects include time series visualisation, forecasting large hierarchies of time series, electricity demand forecasting, and machine learning strategies for improving multi-step forecasting.

Kim Marriott is Professor of Computer Science at Monash University.

Marriott is an internationally recognised researcher in information visualisation and optimisation. He completed his PhD in 1989 at the University of Melbourne and then spent four years as a Research Scientist at IBM TJ Watson Research Center in New York. He returned to Australia in 1993 to take up a position at Monash University. He is a co-director of the Monash FIT Research Flagship in Modelling, Optimisation and Visualisation and one of the leaders of the Monash Immersive Analytics initiative which is exploring how new technologies like the Oculus Rift and Monash's $1.9M CAVE2™ immersive visualisation facility can support data analytics and decision making.

™ CAVE2 is a trademark of the University of Illinois Board of Trustees.

Mark Stammers has over 25 years of experience in data management, data warehousing, client server and web-based interface development projects.

As a co-founder of doubleIQ and its Technical Director since 2000, his key role has been as data modeller/architect. In this role Stammers has had to keep abreast of the latest data engineering and analytical techniques and how best to apply these to specific business problems. Stammers's interests lie in using the semantics of the data to automate and inform underlying data orchestration and exploitation processes, to deliver economic value to end users and downstream business processes.

WHO SHOULD ATTEND

This 3-day course is a unique "starter kit" for Data Science adoption, covering all aspects from data collection to operationalisation. It was designed specifically for participants with little or no prior knowledge in the field who wish to gain an understanding of the real-world organisational role and the life-cycle of quantitative analysis, but will be equally suitable to managers of BI functions who wish to extend their expertise into advanced analytics and Big Data.

It is relevant to professionals from all verticals who are:

- seeking to understand how Big Data is revolutionising the business landscape,
- aiming to build out Data Science capabilities within their organisations,

OR

- wishing to manage Data Science and data-analytic functions.

LEARN

why Data Science has become an intrinsic feature of modern business.

BECOME

your organisation's expert on the utilisation of data and analytic resources.

MAKE

effective business decisions throughout the data and model life cycles.

UNDERSTAND

the value of data and its organisational availability.

COURSE COST

Early Bird	**Standard Registration**
Registration until Friday 18 March, 2016	Registration until Thursday 7 April, 2016
$3,500 + GST	**$4,000 + GST**
3 or more attendees from organisation **$3,300 + GST** *(Save $700)*	3 or more attendees from organisation **$3,800 + GST** *(Save $200)*

Maximum number of attendees: 30

WHEN

11-13 April 2016

WHERE

Monash University Law Chambers
555 Lonsdale St, Melbourne CBD

For further information and to enrol

VISIT

it.monash.edu/data-science

CONTACT

Martine Holberton
Faculty of Information Technology
Monash University, Vic 3800, Australia
P +61 3 990 52346
E martine.holberton@monash.edu

In association with

GCSAGILE
catapulting you into the future

doubleIQ

MONASH University

GROUP OF EIGHT AUSTRALIA

Group of 8 is an alliance of Australia's best universities, recognised for their excellence in teaching and research.