## CONTENTS

### ABOUT
- Become a data ninja 3
- Our training clients 4
- Our faculty 5
- Meet our Founder: Isaac Reyes 6
- What is data science? 7
- The skills of the data scientist 8

### COURSES

#### DATA STORYTELLING FOR BUSINESS
- Course Overview 9
- Course Outline – Day 1 10
- Course Outline – Day 2 11

#### EXCEL ANALYTICS NINJA
- Course Overview 12
- Course Outline – Day 1 13
- Course Outline – Day 2 15

#### ADVANCED VISUALIZATION & DASHBOARD DESIGN
- Course Overview 16
- Course Outline – Day 1 17
- Course Outline – Day 2 18

#### INTRO TO R PROGRAMMING FOR B.I.
- Course Overview 19
- Course Outline – Day 1 20
- Course Outline – Day 2 21

#### INTRO TO DATA SCIENCE & MACHINE LEARNING
- Course Overview 22
- Course Outline – Day 1 23
- Course Outline – Day 2 24
Data is useless without the skill to analyse it.

Data alone is merely a commodity. It’s data scientists and analysts who breathe life into this data and create value, advantage and impact. And the business world agrees—McKinsey predicts that the United States alone faces a shortage of 140,000-190,000 people with deep analytical skills.

We train analytics professionals so that they are prepared to face the challenges and opportunities posed by the new data environment.

Our difference - real business datasets.

Computer science and statistics courses from the university sector do not create professionals who are prepared for the rigors of commercial data. Real business data is often large (millions of rows), high dimensional (hundreds of variables), unstructured and high velocity. It is also rarely clean, awash with missing values, data breaks and outliers.

All of our courses utilize real commercial datasets that will prepare you for the information you will encounter in your next role as a data scientist or analyst.

“You cannot give me too much data. I see big data as storytelling — whether it is through information graphics or other visual aids that explain it in a way that allows others to understand across sectors. I always push for the full scope of the data over averages and aggregations — and I like to go to the raw data because of the possibilities of things you can do with it.”

Mike Covaretta
Data Scientist and Manager, Ford Motor Company
Data science and analytics is revolutionizing business across all industry verticals.

Since 2015, we’ve trained over 300 companies, government departments and NGOs in fundamental data science skills. From banking to telcos and retail to real estate: we’ve trained people in your field.
OUR FACULTY

Learn from the thought leaders in the field

DataSeer is an analytics and data science training provider that has been offering innovative public and private training courses since 2015.

ISAAC REYES, Data Scientist

Prior to founding DataSeer, Isaac held the position of Head of Data Science at Altis, Australia’s largest information management consultancy. At Altis, Isaac led a team of data scientists who design analytics and machine learning solutions for enterprise clients throughout AU/NZ.

A former university lecturer in statistics at the Australian National University, Isaac is also a TEDx speaker and a regular keynote at big data conferences.

Isaac holds a Master’s Degree in Statistics from the Australian National University and a Bachelors Degree in Actuarial Science from Macquarie University.

JAY MANAHAN, Data Storytelling Expert

A data storytelling expert, Jay is concurrently a trainer at DataSeer and Head of Operations at Magpie.IM, an online payments startup.

In his prior role, Jay was the Head of the Manila Shared Services Center for Kforce (Nasdaq: KFRC) and a Business Development Director at analytics company, Sencor.

Jay holds an MBA and B.S. in Mathematics from Ateneo de Manila University. He was a winner of the 2017 Grab Data Visualization challenge.

CARL CALUB, Data Scientist

One of the top data scientists in the APAC region, Carl has employed an extensive arsenal of data science methods (e.g. statistics, graph analysis, spatial and NLP) for different applications (such as impact evaluation, campaign tactical support and intelligent policy design) in a diverse set of industries (banking, finance, telecommunications, education, medicine, disaster and sports).

He has pioneered models for conglomerate risk mapping, exotic derivatives valuation, operational risk measurement and customer relationship management while in Ernst & Young. He has also worked with Big Data for mobility patterns and geographic distribution models while in the PLDT-Smart Group.

Carl is also an active R thought leader in the Philippines, evangelizing it not just in local communities but also thru pioneering work he has done in previous roles.

DOMINIC BOHAN, Analytics Lead

Dom brings a wealth of data storytelling experience from his career at QBE, one of Australia’s largest insurance companies, where he was a leader in analytics, procurement and business improvement.

He has been responsible for negotiating multi-million dollar contracts with suppliers, presenting data driven strategy recommendations to the company’s senior executives, and producing reporting for the Group Board of Directors.
Meet our Founder:

ISAAC REYES

“I live and breathe it.”

This is how Isaac Reyes describes his decade long relationship with data. And with over 3,000 hours of data science training experience at the world’s leading institutions, the numbers certainly add up.

“Teaching is definitely a passion”, says Isaac. “I’ve always kept one foot in the education sector and one foot in the commercial sector. A trainer who is unfamiliar with the commercial application of his methods risks becoming too esoteric in their teaching. On the other hand, a practitioner who doesn’t teach misses out on the peer review process that occurs when presenting to a smart audience.”

Yet another method is used to validate DataSeer training outcomes: global data science competitions such as Kaggle. “I was excited when the first batch of DataSeer graduates entered their first Kaggle competition”, Isaac says. “Our first batch of DataSeer graduates ended up ranked in the top 2.6% of Data Scientists worldwide! I couldn’t have been more proud of our graduates.” “Since then, DataSeer graduates have continued to post top 5% finishes in the world’s most competitive machine learning competitions, including those issued by industrial equipment giant, Caterpillar, and the large European retailer, Rossman.

More recently, Isaac shared his vision for Data Science with perhaps the biggest teaching stage of them all – TEDx. “Speaking about the intersection of Data Science and world issues at a TED event was something that I’ve always wanted to do. My TED talk focused on how we can use Data Science to measure how much we really care about the issues that matter.” Since his TEDx talk, Isaac has also delivered keynote talks on data visualization at Strata, the world’s largest data conference.

So what does Isaac have in store for DataSeer training in 2019?

“2019 is the year we implement all of the feedback we collected from our course attendees in 2018.

We plan on creating more realistic workshop problems around commercial datasets that reflect the digital, high throughput and unstructured data environment of 2019. We are also set to deliver domain specific custom trainings that provide our corporate partners with the exact outcomes they need for their specific industry vertical or departmental needs.

Finally, the vision is for our clients and course attendees to keep winning. Our training attendees will continue to win because they build analytical skills that increase their value in the labor market. Our clients will continue to win because they end up with staff capable of playing at the highest levels of the analytics value chain.”

Isaac holds a Bachelor’s Degree in Actuarial Science from Macquarie University and a Master’s Degree in Statistics from the Australian National University. He is currently the Head of Data Science at Altis Consulting.

“Our clients will continue to win because they end up with staff capable of playing at the highest levels of the analytics value chain.”
WHAT IS DATA SCIENCE?

Data Science is one of the fastest growing disciplines in the business sector today. New findings from MIT research show that companies with data-driven decision making environments had 4% higher productivity and 6% higher profits than other businesses.

In 2008, Dr DJ Patil and Jeff Hammerbacher, heads of analytics and data at LinkedIn and Facebook respectively, coined the term ‘data science’ to describe the emerging field of study that focused on teasing out the hidden value in the data that was being collected from touchpoints all over the retail and business sectors.

Data Science is now the umbrella term used for a discipline that spans programming, statistics, data mining, artificial intelligence, networking, analytics, business intelligence, visualization and a host of other subject areas. The science is constantly changing and evolving, as it moves to keep abreast of technology and business practices alike. Data Science has applications not only in business decisions, but also across a wide range of verticals including biostatistics, astronomy and molecular biology. Wherever you find large amounts of information, you’ll find an application for data science.

You can’t hide from data

The combination of distributed processing power in the cloud, ultra-fast internet and cheap storage has made one thing clear: data is here to stay. Unprecedented amounts of data are now being collected, saved, and stored safely in the cloud. As exabyte upon exabyte is stored, a new discipline grows to tunnel through the mountain of data to find the nuggets of gold: data science.

“Without big data, companies are blind and deaf, wandering out onto the web like deer on a freeway.”
- Geoffrey Moore
Never before have these three been so closely aligned: Coding to query and manipulate large datasets. Statistics to run robust analyses. Business expertise to know how to ask the right questions and create usable insights. But data science isn’t just a static flowchart - it’s a conglomeration of skills in individuals who can use data to let companies know how to move forward and along which vertices.

Every good data scientist knows that the quality of insights are dependent on the quality of data input. The first task of any analytics project is to extract data, whether that data is stored within an on premise data warehouse or housed alongside terabytes in the cloud. Coding skills in languages like SQL, R, Spark and Python are required to extract, clean and prepare data for analysis.

Data Scientists are a rare class among their technical brethren: they need to have excellent client facing and human interfacing skills to complement their technical skills. Often the point of contact between the C-suite and analytics teams, data scientists must have a firm grasp on core business processes, costs, project management methodologies, production systems and corporate culture. The creation of actionable, positive ROI recommendations, backed by solid analysis and good data is the end game, and is the primary reason the profession has grown to be one of the most desirable skillsets in corporate circles today.

While statistics is hardly a new field, today’s data scientists have experienced a paradigm shift in statistical application. Where once the field of statistics concentrated on achieving valid results with small samples, today, with a torrent of information, modern data scientists face the challenge of separating the signal from the noise. Judicious application of statistical methods, coupled with rigorous mathematical theory allow data scientists to create models that power actionable insights.
Data Storytelling is predicted to be the top business skill of the next 5 years.

Well told data stories are change drivers within the modern organisation. But how do we find the most important insights in our business data and communicate them in a compelling way? How do we connect the data that we have to the key underlying business issue?

This course takes students from the fundamentals (what should we be measuring and why?) through to the elements of good visualisation design (what does a good chart look like?) through to proficiency in data storytelling. By the end of the course, participants will know how to produce engaging, cohesive and memorable data stories using Excel and PowerPoint. The course also teaches attendees the importance of producing statistically robust visualisations and insights.

Suitable for

This is our most popular course. It’s suited towards any professional who works with data and charts. If you need to tell better stories with your data, then this course is for you.
COURSE OUTLINE - DAY 1

I. Introductions, Ice Breaker (9:00am - 9:20am)

II. Overview of the Four Keys to Data Storytelling (9:20am - 9:30am)
   - Audience
   - Data
   - Visuals
   - Narrative

III. The Audience (9:30am -10:00am)
   - Exploring versus explaining: PowerPoint decks versus interactive dashboard versus infographics
   - The audience is the 'hero' of your data story
   - The Gestalt Principle of Proximity
   - The Gestalt Principle of Similarity
   - The Cleveland-McGill Hierarchy of Perceptual Tasks
   - Unlocking your inner data journalist: The power of storytelling titles

IV. Preparing Your Data for Data Storytelling (10:00am - 10:15am)
   - Step 1 - Know the context behind your data
   - Step 2 - Variable classification
   - Step 3 - Handle missing values and outlier values
   - Step 4 - Sanity checks
   - Step 5 - Univariate and bivariate analysis
   - Step 6 - Data Storytelling

V. Q&A / Break (10:15am - 10:30am)

VI. The Data-Ink Ratio and the Other Gestalt Principles of Visual Perception (10:30am - 11:00am)
   - Tuft’s Data-Ink Ratio
   - Hybrid Titles
   - Tuft’s Lie-Factor
   - Perceptual issues with 3D charts
   - The Gestalt Principle of Connectedness
   - The Gestalt Principle of Closure
   - The Gestalt Principle of Enclosure

VII. The Visualisation Arsenal (including industry best practices) (11:00am - 12:00pm)
   - Tables Versus Charts Versus Single Metrics
   - Situations when a table outperforms a chart
   - Situations when a single headline metric outperforms a chart
   - Chart selection: choosing the right visual for the job
   - The Bar Chart and Stacked Bar Chart
   - The Pie Chart

VIII. Lunch (12:00pm - 1:00pm)

IX. The Visualisation Arsenal cont. (including industry best practices) (1:00pm - 1:30pm)
   - The Line Chart and Dual Axis Line Chart
   - The Combo Chart and Pareto Chart
   - The Histogram
   - The Scatter Plot

X. Workshop: Team Activity (1:30pm - 4:00pm)

XI. Presentations, Feedback and Day 1 Wrap Up (4:00pm - 5:00pm)
COURSE OUTLINE - DAY 2

I. The How of Data Storytelling  
   • A step-by-step data storytelling chart makeover in Excel  

II. Turning Facts Into Stories: The Power of Narrative  
   • How should a multiple slide data story flow?  
   • Staging your data stories - smartphone sales example  
   • Focusing attention with color  

III. Common Data Storytelling Pitfalls  
   • The Giant Outlier Problem  
   • The Pasta Line Chart Problem  
   • The Occlusion Problem  
   • Effort versus reward: Macros for speeding up the data storytelling process  

IV. Q&A / Break  

V. Applied Walkthrough: How One of Our Lead Data Storytellers or Lead Data Scientists Created a Data Story for a Major Global Client  

VI. Lunch  

VII. The Statistics Behind Good Data Storytelling  
   • Sample size and inference - Why it is important  
   • Correlation and causation - Applied examples  

VIII. Workshop: Team Activity and Presentation  

IX. Presentations Feedback and Wrap Up
EXCEL ANALYTICS NINJA

Course duration: 2 Days
Laptop Specs: intel i3 processor, 2GB RAM. Either Mac or Windows OS.
Required Software: Excel and Powerpoint 2013 or later.

Learn the fundamentals of business analytics in this two day intensive program.

Some executives mistakenly believe that the majority of value in business datasets is only unlocked by applying advanced statistical and machine learning techniques. In practice, most of the value in business data is derived by asking relatively simple questions that can be answered using basic data manipulation and common metrics (e.g. averages, totals, counts and percentages).

That said, the ability to ask the right business questions and answer them with the right metrics is a fundamental analytics skill that is sorely lacking in the skillset of most data analysts and managers. Why? University statistics and math programs don’t prepare graduates for the challenges and pace of the business setting. In this Excel based course, participants will learn how to progress through the full data driven decision making process, from identifying the business question through to hypothesis development, data manipulation and presenting of results.

Suitable for

This is our second most popular course. It’s suited to any professional who needs to make decisions using business data.
COURSE OUTLINE - DAY 1

I. Introductions, Ice Breaker (9:00am - 9:15am)
   - What is the end goal of this course?

II. Keys to Effective Analytics: Exploratory Data Analysis (EDA) (9:15am - 9:30am)
   - What is EDA?
   - Context: understanding the data and its source
   - Variables: knowing and classifying data into various data types
   - Wrangling: performing basic data munging to address missing values, outliers, input errors
   - Analysis: discovering univariate and bivariate relationships in the data
   - Visualization: using charts and graphs to present your analytics

III. Context and Variables: Understanding the Data (9:30am - 9:45am)
   - Questions to ask of your dataset
   - What are the different types of data?
   - Classifying the variables of the course dataset
   - Numeric variables: continuous and discrete
   - Categorical variables
   - Date variables
   - Categorical variables that appear numeric and vice-versa
   - Continuous variables that appear discreet and vice-versa
   - Dummified data: what they look like and why they exist
   - Formatting data according to their variable types

IV. Wrangling: Using Formulae, Filtering, and Sorting to Manipulate Data (9:45am - 10:15am)
   - Querying your data
   - Sorting data according to various dimensions and multiple levels
   - Performing a sanity check on the data
   - Addressing missing values, outliers, and input errors
   - List-wise deletion and case-wise deletion
   - Identifying and extracting metrics needed to generate or prove certain insights
   - Manipulating text or string data
   - Working with dates
   - Wrangling data through arrays

V. Q&A / Break (10:15am - 10:30am)

VI. Univariate Analysis: Leveraging Excel Features for Analyzing Data (10:30am - 10:45am)
   - Measures of central tendency: Where is the middle of my data?
   - Measures of dispersion: How varied are the values of my data?
   - The Histogram and appropriate binning
   - Statistical summaries: sum, count, min, max, range, mean, median, mode, quartiles
   - Performing analytics on your data by grouping, sorting, filtering, and applying statistical transformations
COURSE OUTLINE - DAY 1

VII. Bivariate and Multivariate Data Analysis: Pivot Table Jiu Jitsu Purple Belt

- Creating new variables and metrics that illustrate inter-variable relationships
- Performing advanced analytics by applying statistical transformations on multiple variables
- Making complex queries on your data to make business-relevant analysis (e.g. top x%, above/below average, between x and y, contains x, etc.)
- When are proportions and averages better than counts or totals for comparisons?
- Working on a large telco dataset

(10:45am - 12:00pm)

VIII. Lunch

(12:00pm - 1:00pm)

IX. Bivariate and Multivariate Data Analysis (cont’d)

- Summarizing data in various ways, according to the insight to be supported
- The difference between count, sum, and average
- Various ways to calculate percentages and their meanings
- Showing increases, decreases, growth rates, decline rates
- Showing running totals and ranks

(1:00pm - 2:00pm)

X. Dynamic Data Visualization for Rapid-Fire EDA

(2:00pm - 2:15pm)

XI. Workshop

(2:15pm - 4:15pm)

XII. Group Presentations and Wrap Up

(4:15pm - 5:00pm)
COURSE OUTLINE - DAY 2

I. Day 1 Recap (9:00am - 9:15am)
   • Lessons from Day 1 workshop: context, metric selection, averages, percentages

II. Dynamic Data Visualization (cont’d) (9:15am - 9:45am)
   • Modifying data visualizations on the fly
   • When to use Stacked Bar Charts, Combination Charts, Dual-Axis Charts
   • Slicing data and charts according to categorical, numeric, and date variables
   • Building a dynamic dashboard

III. Polishing Your Data Visualization (9:45am - 10:15am)
   • Critiquing charts and learning best practices (e.g. formatting, sorting, etc.)
   • How do I select the right chart for a given objective?
   • Selecting the data to include in your chart
   • Sorting your chart
   • Best practices in stating key insights in your charts

IV. QA / Break (10:15am - 10:30am)

V. Elegant Data Visualization (10:30am - 12:00pm)
   • Formatting your data table effectively: best practices, aesthetics, conditional formatting
   • Sparklines and other inline graphics
   • When to use and how to create non-standard data visualizations
   • Tornado Chart / Divergent Bar Chart / Bi-Directional Bar Chart
   • Dual Axis Charts
   • Funnel Chart
   • Reference lines
   • Bullet graph
   • Area Chart
   • Waterfall
   • Burn Down Chart

VI. Lunch (12:00pm - 1:00pm)

VII. Presenting Your Analysis to an Executive Audience (1:00pm - 2:00pm)

VIII. Workshop (2:00pm - 4:00pm)

IX. Group Presentations and Wrap Up (4:00pm - 5:00pm)
ADVANCED VISUALIZATION & DASHBOARD DESIGN

Course duration: 2 Days
Laptop Specs:
Windows 7/8/10 or MacOS El Capitan/Sierra.
At least 2GB RAM.
Minimum 2GB free disk space.

Required Software:
Microsoft Excel 2013 or later.
Tableau Public or Power BI

Take your visualization and dashboard skills to the next level

Advanced Visualization and Dashboard Design is aimed at the professional who already possesses fundamental data visualization and data storytelling skills. A natural continuation point from our Data Storytelling for Business and Excel Analytics Ninja courses, this course provides participants with the skills needed to produce stunning, understandable business dashboards and graphs. Taught using a variety of visualization tools including PowerBI and Tableau, the course covers the keys to designing for interactivity and drill down effects. The course also covers less commonly used but valuable visualization methods, including methods for visualizing networks and flows. Dashboard design is covered in detail, with participants creating a dashboard ‘makeover’ during the class practical workshop.

Suitable for

This course is suitable for any professional who wants to analyze and extract value from business data using sophisticated data visualization and interactive dashboards that convey insights with clarity.
COURSE OUTLINE - DAY 1

I. Introductions (9:00am - 9:15am)
II. Grammar of Graphics (9:15am - 10:00am)
   • Think like a graph
   • Look at bars, lines, and scatter plots from new perspective
   • Learn the intuition of visualization software
III. Q&A/Break (10:00am - 10:15am)
IV. Visualizing Comparisons (10:15am - 11:15am)
   • Expand your visual vocabulary beyond bar charts
   • Compare more than one metric simultaneously
   • Compare across not just one but two classifications
V. Visualizing Parts (11:15am - 12:00pm)
   • Say goodbye to pie charts
   • Show multiple and deeper levels of comparisons
VI. Lunch (12:00pm - 1:00pm)
VII. Visualizing Trends (1:00pm - 1:45pm)
   • Beyond simple line charts for visualizing time series
   • Recognise the inadequacy of time plots for time series data
   • Emphasize either annual trends or seasonal patterns
VIII. Visualizing Maps (1:45pm - 2:30pm)
   • Create maps using data
   • Demystify how data points are superimposed on maps
   • Recognise how to enhance addresses using coordinates
   • Identify data considerations for drawing heatmaps
IX. Q&A/Break (2:30pm - 2:45pm)
X. Workshop (2:45pm - 4:30pm)
XI. Day 1 wrap-up (4:30pm - 5:00pm)
COURSE OUTLINE - DAY 2

I. Intro: What is a Dashboard? (9:00am - 9:30am)
   - Examples of Good vs. Bad Dashboards
   - Keys to Effective Dashboards: UX & Interactivity, Metrics, Visuals & Design
   - Class-specific behaviour
   - User-defined/custom functions

II. On-the-fly Critiquing / Makeover of Participants’ Own Dashboards (9:30am - 10:00am)

III. QA / Break (10:00am - 10:15am)

IV. User Experience & Interactivity (10:15am - 11:30am)
   - Overview of Interactivity: tooltips, sorting, filtering, highlights, zoom, force
   - Demo: building interactivity in Tableau and Power BI

V. Metrics (11:30am - 12:00pm)
   - What is a good metric?
   - Action & Accountability
   - Context: via goals, via time period comparisons, via relatable units
   - Comprehensibility

VI. Lunch (12:00pm - 1:00pm)

VII. Visuals & Design (1:00pm - 2:00pm)
   - Review of visualization concepts: data-ink ratio, chart-junk, Gestalt principles
   - Layout and positioning
   - Focusing attention in dashboards
     - Size
     - Color
     - Big numbers
     - Human forms
   - The Bullet Graph

VIII. Workshop (2:00pm - 4:15pm)

IX. Group Presentations and Wrap Up (4:15pm - 5:00pm)
INTRO TO R FOR BUSINESS INTELLIGENCE

Course duration:
2 Days

Laptop Specs:
Windows 7/8/10 or MacOS El Capitan/Sierra or Ubuntu 14.04/16.04
At least 2GB RAM.
Minimum 2GB free disk space.

Required Software:
Latest Microsoft R Open for 64-bit systems or latest R for 32-bit systems.
Latest version of RStudio.

R is the world's leading data science and statistics programming language.

In this introduction to R for business intelligence: you will master the fundamentals of this beautiful open-source language, including factors, lists and data frames. After completing the course, you will be ready to undertake your very own end-to-end data analysis projects using the world’s most sophisticated data analysis tool. R itself is completely free and can be used to extend the capabilities of data warehousing software such as SQL Server 2016 and Microsoft Azure ML Studio. Working on business datasets in class, you will leverage the power of R to inform business decision making. Join millions of R-users worldwide in a user community that is growing by 40% every year.

Suitable for

This course is suitable for business professionals who want a crash course in an end-to-end data management workflow. It is also suitable for professionals who seek to understand the ecosystem and community behind R and make it a powerful and cost-effective application for their enterprise.
COURSE OUTLINE - DAY 1

I. Introductions (9:00am - 9:15am)

II. R Overview (9:15am - 9:45am)
   • Who develops R
   • Harnessing the power of the R community
   • Acknowledging the limitations of R
   • Discovering more potent distros of R
   • Enhancing the R experience

III. Q&A / Break (9:45am - 10:00am)

IV. Interactive Programming (10:00am - 10:45am)
   • Interacting with R
   • Saving values to variables
   • Naming R objects properly
   • Annotating R code
   • Scripting multiple R commands

V. Workshop (10:45am - 11:15am)
   • Working with object assignment, operators, and functions

VI. Data Types (11:15am - 12:00pm)
   • Distinguishing between how objects are stored and parsed
   • Numeric, logical, and character data
   • Understanding the quirks of R

VII. Lunch (12:00pm - 1:00pm)

VIII. Data Types (cont.) (1:00pm - 1:30pm)
   • Creating dates and factors in R
   • Learning how to index in R

IX. Functions (1:30pm - 2:00pm)
   • Specifying inputs to functions
   • Getting help on functions
   • Writing custom functions

X. Workshop (2:00pm - 2:45pm)

XI. Q&A / Break (2:45pm - 3:00pm)

XII. Workshop (cont.) (3:00pm - 3:30pm)

XIII. Data Structures (3:30pm - 4:00pm)
   • Knowing what structures can be used to organize data
   • Indexing different data structures

XIV. Vectorised Loops (4:00pm - 4:30pm)
   • Looping vs vectorisation
   • Avoiding loops in R

XV. Workshop (4:30pm - 4:55pm)

XVI. Day 1 Wrap up (4:55pm - 5:00pm)

• Who develops R
• Harnessing the power of the R community
• Acknowledging the limitations of R
• Discovering more potent distros of R
• Enhancing the R experience

• Interacting with R
• Saving values to variables
• Naming R objects properly
• Annotating R code
• Scripting multiple R commands

• Working with object assignment, operators, and functions

• Distinguishing between how objects are stored and parsed
• Numeric, logical, and character data
• Understanding the quirks of R

• Creating dates and factors in R
• Learning how to index in R

• Specifying inputs to functions
• Getting help on functions
• Writing custom functions

• Knowing what structures can be used to organize data
• Indexing different data structures

• Looping vs vectorisation
• Avoiding loops in R

• Workshopping
### COURSE OUTLINE - DAY 2

**I. Working with csv-files**  
- What are csv-files
- Common variations of csv-files
- Importing csv-files into R
- Ensuring integrity of metadata on import

**II. Wrangling Data**  
- Think like a table
- Nested functions vs. chained commands
- Best practices in script writing

**III. Q&A / Break**

**IV. Wrangling Data (cont.)**  
- Typical code chunks for data wrangling
- Exporting data.frames to a csv-file

**V. Working with xlsx-files**  
- Importing xlsx-files
- Saving interim data.frames as rds-files

**VI. Workshop**

**VII. Lunch**

**VIII. Working with xlsx-files (cont.)**  
- Saving the workspace as RData-files
- Importing several xlsx-files simultaneously
- Exporting multiple data.frames as sheets in an xlsx-file

**IX. Handling Multiple Tables**  
- Data wrangling for analysis
- Crafting a master dataset from separate tables
- Knowing what type of join to use
- Understanding how joins treat duplicates

**X. Q&A / Break**

**XI. Workshop**

**XII. Day 2 Wrap up**
Learn how to think like a data scientist to solve business problems: translate business problems to and derive value from data science models.

Introduction to Data Science and Machine Learning is aimed at the professional who wants to learn how to think like a data scientist with a strong focus on business applications. By the end of the course, participants will gain an understanding of what data science is and how it can be used to create value for an organisation. In addition, they will also be exposed to various machine learning problems and how they can be relevant to a business.

Taught using a variety of open source and cloud technologies, the course provides a sneak peek at an end-to-end data science workflow as well as the process and technologies involved.

Suitable for

This course is suitable for managers who want a better understanding of how data science can be used to aid business decision making. The course is also aimed at any person who wants a high- to mid-level appreciation view of data science and its applications.
COURSE OUTLINE - DAY 1

I. Introduction
   (9:00am - 9:15am)

II. Data Science Overview
    (9:15am - 10:00am)
    - Data science for the individual
    - Data science for the organization
    - Data scientist vs data science teams
    - Generating value from data science
    - Solving problems the data science way

III. Q&A/Break
     (10:00am - 10:15am)

IV. Introduction to Machine Learning & Azure ML Studio
    (10:15am - 10:30am)
    - Machine learning vs traditional programming
    - Drivers to the rise of machine learning
    - Types of machine learning problems
    - Working in Azure ML Studio

V. Anomaly Detection
    (10:30am - 11:15am)
    - Use cases for anomaly detection
    - Anomaly detection in Azure ML Studio
    - Adjusting sensitivity to anomalies
    - Sense checking anomalies detected

VI. Recommender
    (11:15am - 12:00pm)
    - Schema of typical data source for recommenders
    - User-item ratings and implied ratings
    - Matchbox recommender in Azure ML Studio
    - Recommendation engines in practice

VII. Lunch
     (12:00pm - 1:00pm)

VIII. Segmentation
      (1:00pm - 1:45pm)
      - Numerical manifestation of segments
      - Helping the algorithm detect more distinct segments
      - Cluster analysis in Azure ML Studio
      - Identifying the number of segments to detect
      - Profiling identified segments

IX. Best Practices in Segmentation
    (1:45pm - 2:15pm)
    - Ensuring relevance of segments to be discovered
    - Practicing judicious segment profiling
    - Rendering actionable findings from segmentation

X. Q&A/Break
    (2:15pm - 2:30pm)

XI. Workshop
    (2:30pm - 4:30pm)

XII. Day 1 wrap-up
     (4:30pm - 5:00pm)
COURSE OUTLINE - DAY 2

I. Supervised Learning Overview
   • What is supervised learning
   • What kinds of predictions can you make
   • How do predictions look like for different types
   • How are predictions assessed for different types

II. Q&A / Break

III. Supervised Regression Models
   OLS for Supervised Regression
   • Intuition of linear regression
   • Optimising RMSE vs optimising MAE
   • Understanding the feature weights of linear regression
   • Crafting a more flexible linear regression model
   • Presenting the feature weights as outputs
   Decision Trees & Random Forests for Regression
   • Intuition of decision tree predictions
   • Underlying principles in calibrating decision trees
   • How decision trees build up into a random forest
   • Understanding the hyperparameters of random forests
   Neural Networks for Regression
   • How neural networks generate predictions
   • Underlying principles in calibrating neural networks
   • Understanding the hyperparameters of neural networks

IV. Model and Feature Selection
   • The dangers of overfitting
   • Knowing when your model is too good to be true
   • Measuring model performance adjusted for overfitting
   • Identifying nuisance and irrelevant features
   • Model and feature selection

V. Lunch

VI. Using Categorical Features
   • Operationalising categorical or string features
   • Common errors when working with dummies

VII. Supervised Classification
   • Intuition behind calibrating classification models
   • Deriving classification error measures
   • Considerations in selecting the error measure to track

VIII. Q&A/Break

IX. Workshop

X. Day 2 Wrap-up
Need help?
Contact us today for enrollment inquiries.

dataseer.com
info@dataseer.com
1 Scotts Road #24-10,
Shaw Centre SG 228208

PH: +632 821 8471 / +632 959 6761 / +63 917 677 3825
SG: +65 6817 8652