IfATE

Shaping skills training

Data scientist (integrated degree)

DATA SCIENTIST (INTEGRATED DEGREE)

Reference Number: ST0585

Details of standard

Occupational Profile

Data Science is a broad and fast-moving field spanning maths and statistics, software engineering and communications. Data Scientists blend experience and knowledge from a wide range of fields and organisations, and continuously seek to expand their range of technical skills.

Data Scientists find information in diverse datasets to address complex problems and improve organisational processes. They are inquisitive, they explore and visualise data of all kinds, find and present 'stories' within the data in a meaningful way to a range of technical and non-technical audiences. They make recommendations to inform strategic and operational decision making through sourcing, accessing and manipulating data, and engineering data processes. They identify and address data biases, and handle private data ethically and appropriately, complying with (inter)national privacy regulations. They use the insights gathered about the data they have analysed to inform and achieve organisational goals. They achieve desired outcomes by planning, organising and managing resources effectively.

Data Scientists are dynamic and adaptable, addressing varied problems with varied techniques. They actively explore innovative ways to use existing and new statistical, algorithmic, predictive, machine learning and artificial intelligence tools and techniques, to find significant and valuable patterns in data and transform this into information for their organisation. They gather new sources of data, and combine datasets to increase their value. Using a scientific approach, they perform statistical analysis, build and validate models from the data, use programming practices, and maintain data, tools and processes to implement robust and valuable data solutions.

Data Scientists have an impact at a strategic and operational level by building and maintaining strong collaborative relationships with key stakeholders, subject matter experts and colleagues at all levels. They engage with the wider Data Science community to share ideas, techniques and experiences. They can work in any sector, public or private, and will often work in a multi-disciplinary team with domain experts, Data Architects, Data Engineers, Analysts, and Technology Professionals.

Typical Roles

Data Scientist, Informatics, Data Engineer

Entry Requirements

Individual employers will set the selection criteria, but this is likely to include 3 A Levels, including 1 STEM subject, a Level 3 Apprenticeship or other relevant qualifications. Other relevant or prior experience may also be considered by the employer.

Knowledge

A Data Scientist must understand:

1. The context of Data Science and the Data Science community in relation to computer science, statistics and software engineering. How differing schools of thought in these disciplines have driven new approaches to data systems.

2. How Data Science operates within the context of data governance, data security, and communications. How Data Science can be applied to improve an organisation's processes, operations and outputs. How data and analysis may exhibit biases and prejudice. How ethics and compliance affect Data Science work, and the impact of international regulations (including the General Data Protection Regulation.)

3. How data can be used systematically, through an awareness of key platforms for data and analysis in an organisation, including:

   1. Data processing and storage, including on-premise and cloud technologies.

   2. Database systems including relational, data warehousing & online analytical processing, "NoSQL" and real-time approaches; the pros and cons of each approach.

   3. Data-driven decision making and the good use of evidence and analytics in making choices and decisions.

4. How to design, implement and optimise analytical algorithms – as prototypes and at production scale – using:

   1. Statistical and mathematical models and methods.

   2. Advanced and predictive analytics, machine learning and artificial intelligence techniques, simulations, optimisation, and automation.

   3. Applications such as computer vision and Natural Language Processing.

   4. An awareness of the computing and organisational resource constraints and trade-offs involved in selecting models, algorithms and tools.

   5. Development standards, including programming practice, testing, source control.

5. The data landscape: how to critically analyse, interpret and evaluate complex information from diverse datasets:

   1. Sources of data including but not exclusive to files, operational systems, databases, web services, open data, government data, news and social media.

   2. Data formats, structures and data delivery methods including "unstructured" data.

   3. Common patterns in real-world data.

Skills. A Data Scientist is able to:

1. Identify and clarify problems an organisation faces, and reformulate them into Data Science problems. Devise solutions and make decisions in context by seeking feedback from stakeholders. Apply scientific methods through experiment design, measurement, hypothesis testing and delivery of results. Collaborate with colleagues to gather requirements.

2. Perform data engineering: create and handle datasets for analysis. Use tools and techniques to source, access, explore, profile, pipeline, combine, transform and store data, and apply governance (quality control, security, privacy) to data.

3. Identify and use an appropriate range of programming languages and tools for data manipulation, analysis, visualisation, and system integration. Select appropriate data structures and algorithms for the problem. Develop reproducible analysis and robust code, working in accordance with software development standards, including security, accessibility, code quality and version control.

4. Use analysis and models to inform and improve organisational outcomes, building models and validating results with statistical testing: perform statistical analysis, correlation vs causation, feature selection and engineering, machine learning, optimisation, and simulations, using the appropriate techniques for the problem.

5. Implement data solutions, using relevant software engineering architectures and design patterns. Evaluate Cloud vs. on-premise deployment. Determine the implicit and explicit value of data. Assess value for money and Return on Investment. Scale a system up/out. Evaluate emerging trends and new approaches. Compare the pros and cons of software applications and techniques.

6. Find, present, communicate and disseminate outputs effectively and with high impact through creative storytelling, tailoring the message for the audience. Use the best medium for each audience, such as technical writing, reporting and dashboards. Visualise data to tell compelling and actionable narratives. Make recommendations to decision makers to contribute towards the achievement of organisation goals.

7. Develop and maintain collaborative relationships at strategic and operational levels, using methods of organisational empathy (human, organisation and technical) and build relationships through active listening and trust development.

8. Use project delivery techniques and tools appropriate to their Data Science project and organisation. Plan, organise and manage resources to successfully run a small Data Science project, achieve organisational goals and enable effective change.

Behaviours. A Data Scientist demonstrates:

1. An inquisitive approach: the curiosity to explore new questions, opportunities, data, and techniques; tenacity to improve methods and maximise insights; and relentless creativity in their approach to solutions.

2. Empathy and positive engagement to enable working and collaborating in multi-disciplinary teams, championing and highlighting ethics and diversity in data work.

3. Adaptability and dynamism when responding to varied tasks and organisational timescales, and pragmatism in the face of real-world scenarios.

4. Consideration of problems in the context of organisation goals.

5. An impartial, scientific, hypothesis-driven approach to work, rigorous data analysis methods, and integrity in presenting data and conclusions in a truthful and appropriate manner.

6. A commitment to keeping up to date with current thinking and maintaining personal development. Including collaborating with the data science community.

Duration:

Typically 36 months

Qualifications:

BSc in Data Science

Upon successful completion of the apprenticeship and upon receipt of the apprenticeship certificate, individuals are eligible to apply for The Science Council for Registered Scientist (RSci) through a shortened application route. Individuals also need to be a member of a professional body that is licensed by the Science Council to be awarded this status. Further information is on the Science Council's website.

English and Maths:

Apprentices without Level 2 English and Maths will need to achieve this level prior to taking the end-point assessment. For those with an education, health and care plan or a legacy statement the apprenticeships English and maths minimum requirement is Entry Level 3 and British Sign Language qualification are an alternative to English qualifications for whom this is their primary language.

Level:

This is a Level 6 apprenticeship

Review Date:

After two years

Version log

VERSION	CHANGE DETAIL	EARLIEST START DATE	LATEST START DATE	LATEST END DATE
1.1	Occupational standard revised	24/04/2024	Not set	Not set
1.0	Approved for delivery	17/08/2018	23/04/2024	Not set