

 www.spectraanalytics.com

 info@spectraanalytics.com



From Complexity Science to Data Science

By

Dr Marcus Ong & Dr Dan Sprague



Contents

- I. Data Science / AI / ML Market**
- II. Delivering an end-to-end solution**
- III. Why Complexity Science is great for Data Science**
- IV. Should I go into industry?**
- V. Should I work in-house?**
- VI. What skills are required?**

Data Science / AI / ML Market

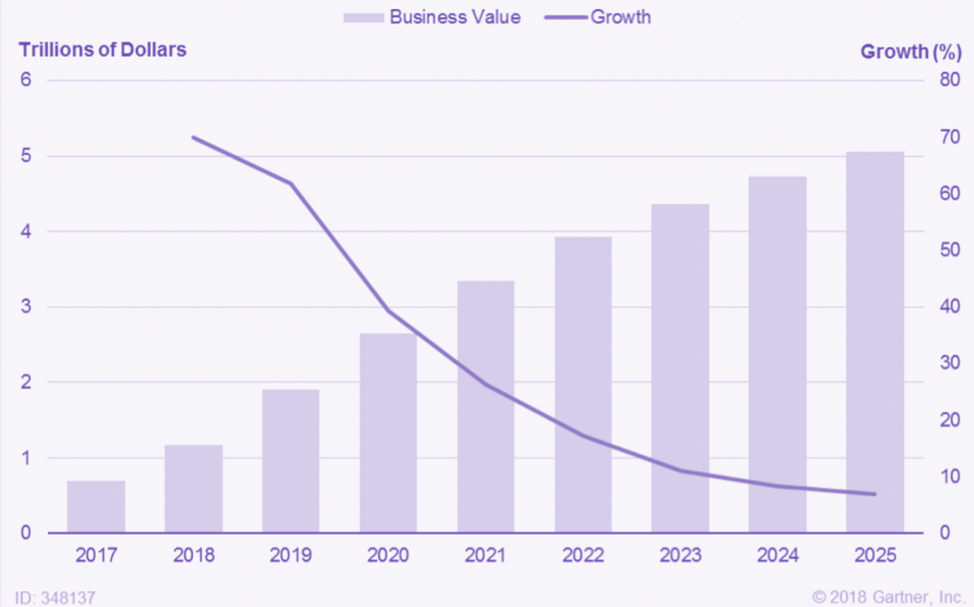
I. AI / ML Adoption

Business value derived from AI will be over \$5trn by 2025

Growth in AI will quickly normalize from 70% in 2017 to just 7% by 2025

- Gartner

Business Value Forecast for AI



By 2019, startups will overtake Amazon, Google, IBM and Microsoft in driving the AI economy with disruptive businesses

- Gartner

I. Data Science Hiring

Estimated 22,000 PhDs capable of building AI systems globally

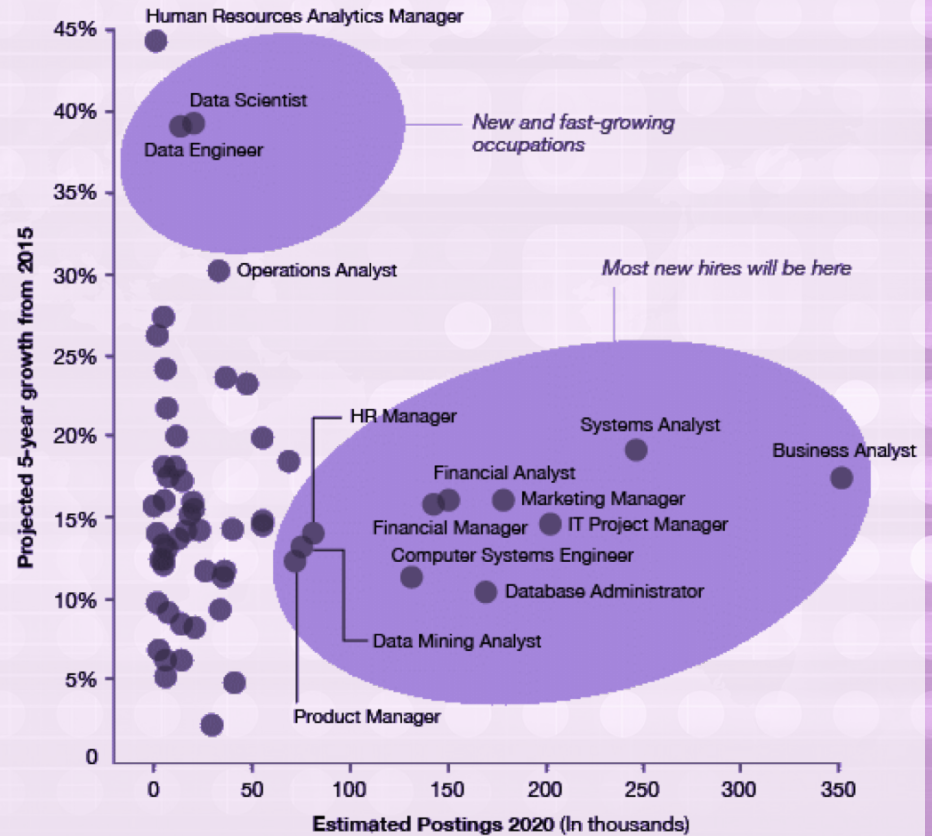
Only 3,000 looking for a job

10,000+ job vacancies in US

- Element AI

Situation likely to deteriorate as roles increase by 40% over the next 5 years

- PWC



Note: Each dot represents an occupation in the US jobs market where data science and analytics skills are required. Source: PwC analysis based on Burning Glass Technologies data, January 2017.

I. Data Science Hiring

Data Science requires advanced skills in a range of domains which are in scarce supply and time consuming to develop

- PWC

■ Fundamental (conceptual literacy) ■ Intermediate (practical application) ■ Advanced (applied)

Skills	Analytics-enabled jobs		Data science jobs		
	Data-driven decision makers	Functional analysts	Data analysts	Data engineers (hardware and software)	Data scientists and advanced analysts
Domain knowledge Research or business	Advanced	Advanced	Intermediate	Intermediate	Advanced
Visualization The story in the data	Intermediate	Advanced	Advanced	Intermediate	Advanced
Data governance Including ethics and security	Intermediate	Intermediate	Intermediate	Advanced	Advanced
Engineering Hardware, software, storage		Intermediate	Intermediate	Advanced	Advanced
Management/Curation Sourcing, cleaning, manipulating		Intermediate	Intermediate	Advanced	Advanced
Analytical approaches Level of precision	Intermediate	Intermediate	Advanced	Intermediate	Advanced
Machine learning Teach computers to recognize patterns			Intermediate	Advanced	Advanced

Source: PwC analysis based on Burning Glass Technologies data, January 2017.

Delivering an end-to-end solution

II. End-to-End Solution



Collect the data that you are using for the analysis.

- Data Files
- Web Scraping
- Database
- Forms

Consider ethics

Store the data that you've collected

- SQL (PostgreSQL)
- NoSQL (MongoDB)
- Timeseries (Casandra)
- Distributed (Hadoop)
- Consider security and data governance

Analyse the data

- Python, R, Matlab, Excel
- Data pipelines (Scikit-Learn)
- Requires domain/business input

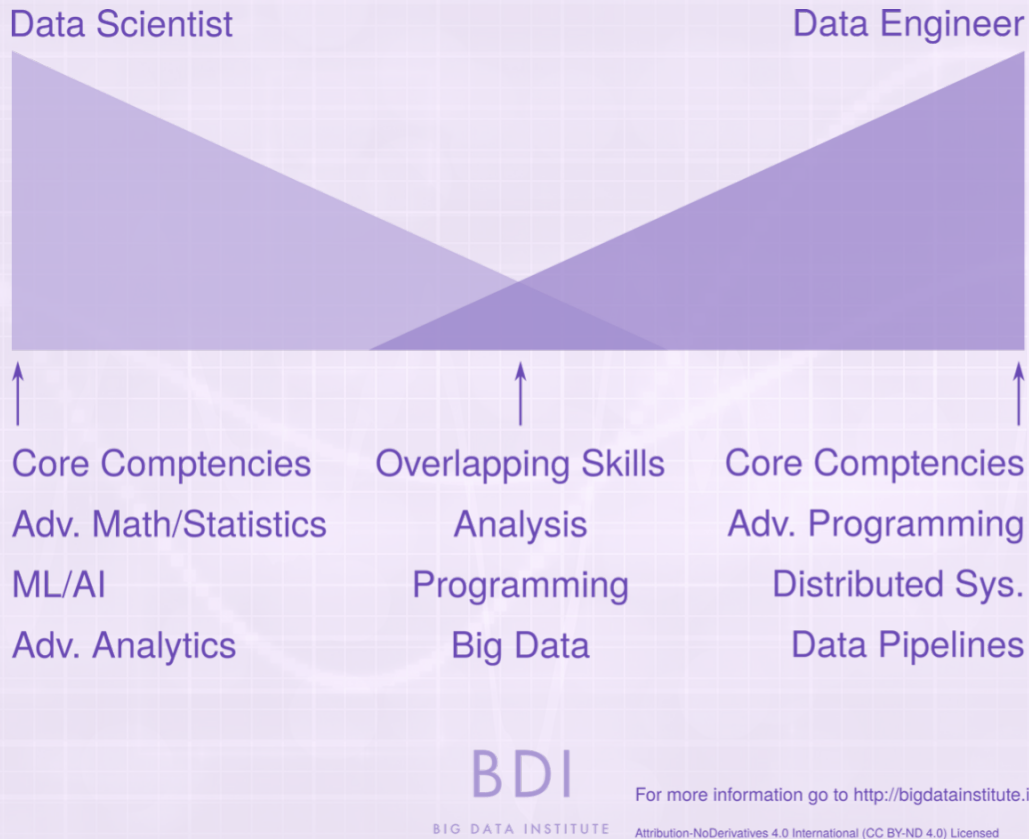
Deliver product to the business

- Webserver (AWS)
- Executable Files
- Scripts
- Need to consider maintenance / support, and task scheduling, security

Displaying static & interactive graphs via a GUI

- GUI Design
- Webapps (HTML, Javascript, d3)

II. Roles

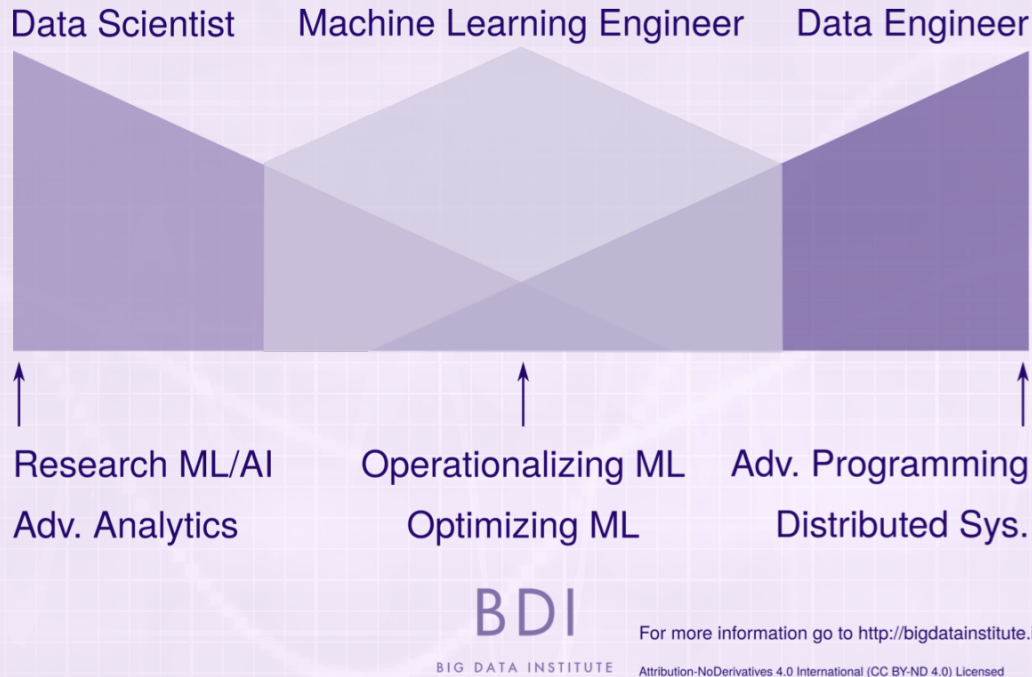


Large companies will employ both Data Engineers and Data Scientists

At smaller companies you may be required to fulfil both roles

Expectation that Data Scientists should provide end-to-end solution

II. Roles



Machine Learning Engineer a new role that straddles the Data Scientist and Data Engineer roles

Another important role is Business Analyst / Business Intelligence

Why Complexity Science is great for Data Science

III. General Skills

- Open-minded and curious - stay at forefront rapidly developing research
- Teaching and team working - share knowledge and work collaboratively.
- Strong mathematical background - understand the algorithms and statistical properties of the problem
- Independent research - often no existing solution, requires independent investigation
- Experience of high level rapid prototyping languages such as Python, R and Matlab

III. Complexity Skills

- Recognise analogies between systems in different disciplines
- Interdisciplinary: can talk to domain experts to understand the problems they care about
- Comfortable dealing with a variety of real-world systems
- Comfortable with nonlinear and non-stationary systems

Should I go in to industry?

IV. Industry V Academia

Criteria	Industry	Academia
Focus	Solving real world problems with immediate impact. Great to see things put into practice.	Teaching, publishing papers and submitting grant applications. Great to help people learn and contribute to the 'body of knowledge'.
Culture	Fast, dynamic, results oriented culture.	Slower, driven by intellectual curiosity and funding opportunities. Lots of freedom, independence and more collaborative.
Job Security	If you're in a well-established company and perform well job security will be good	Good once you achieve lecturer level and tenure. Fixed-term contracts for post docs can be stressful

IV. Industry V Academia

Criteria	Industry	Academia
Career Progression	Well-structured giving stability.	Dependent upon funding and academic requirements.
Work life balance	Long hours	Flexible hours
Cultural diversity	Work with people from a range of backgrounds; varies between industries.	Work with mainly very highly educated people largely from your field of focus

Should I work in-house?

V. In-house V Consulting

Criteria	In-house	Consulting
Projects	Range of projects is narrow as you are only focussing on the needs of one company	Range of projects is wide as you'll be working on a variety of projects with different clients
Knowledge	Highly specific for your individual industry	More general knowledge across a range of industries
Enjoyment	If you choose your industry correctly you will generally be working on projects you're interested in	If broad range of interests this suits a curious mind but you may have to work on less interesting projects from time to time
Senior Contact	Generally your main contact will be your line manager	Access to higher level management because they do not have the skills in-house and have hired you

What skills are required?

based on work by

Dr Ben Dias, Head of Data Science at Royal Mail

VI. Education

Criteria	Requirement	Evidence	Why
Education	Strong educational background in a heavily mathematical subject	Hold a Masters or PhD in a mathematical based subject.	Being a data scientist you need to understand the algorithms that you are using and the statistical properties of the problem you are analysing
Research	Experience of conducting independent research	Have completed a Masters or PhD research project.	Much of the work you will be required to do does not have an existing solution. This means you must be able to work independently to investigate the problem and where possible generate a solution

VI. Programming Skills

Criteria	Requirement	Evidence	Why
Programming Skills	<p>Experience of different types of languages:</p> <ol style="list-style-type: none">1. A high-level rapid prototyping language such as Python or R2. A low-level, deployment language such as Java, C/C++3. A scalable/Big Data language such as Scala/Spark	<p>Example of projects worked on in these languages with the number of years and proficiency.</p> <p>Open-source code bases, such as Github, beneficial. Links to online work.</p>	<p>For junior data scientists it's important to programme confidently in a high level language to conduct your analysis. For senior data scientists it's important to implement and productionise solutions i.e. increasing computational efficiency - a low-level language - and/or deploying at scale - a big data language.</p>

VI. Programming Skills

Criteria	Requirement	Evidence	Why
Programming Skills	5. Frontend development using HTML, CSS, Javascript, VUE and D3	Example of projects worked on in these languages with the number of years and proficiency.	Displaying and visualising your findings is beneficial which is where frontend development is important
	6. Databases: SQL (PostgreSQL), NoSQL (MongoDB), Timeseries (Kx), Graph (Neo4j)	Open-source code bases, such as Github, beneficial. Links to online work.	Important to interact with databases to access data for the analysis
	7. Data Pipelines		Data pipelines ensure consistency and reproducibility

VI. Technical Skills

Criteria	Requirement	Evidence	Why
Technical Skills	Knowledge of a range of methods for forecasting, optimising and simulating	Examples of using different methods and why certain methods were chosen. External Accreditations such as CMath, CSci, CStat.	It is important to have a wide knowledge of the different methods, where to apply them and their strengths and weaknesses. This will help you choose the most appropriate methods when solving a problem.

VI. Technical Skills

Criteria	Requirement	Evidence	Why
Managing a server	Can keep a server (e.g. a webserver or database server) up and running	Personal website, internships.	Data science is more than just running a script - you may need to deploy your products to a server for others to interact with
DevOps Tools	Experience or knowledge of production technologies such as Docker (for deploying servers), Airflow (for scheduling tasks), and unit testing	Some of these can be used during your PhD. Can also gain experience from internships	Docker is becoming increasingly popular across the data science world. Experience with other DevOps tools demonstrates that you can integrate your work with other systems

VI. Business Skills

Criteria	Requirement	Evidence	Why
Commercial Tools & Processes	<p>Knowledge of Agile software development (e.g. Scrum, Kanban etc) and associated tools (e.g. JIRA, Assembla etc), PRINCE2 project management.</p> <p>Experience of using different environments (Linux, Windows, Hadoop, Cloud etc)</p> <p>Experience of using version control (e.g. Git) and documenting (e.g. Wiki)</p>	Examples of projects you managed that used these tools / methodologies and why they were suited	These are generally required for a Senior data scientist in order to manage and deliver projects.

VI. Business Skills

Criteria	Requirement	Evidence	Why
Commercial Acumen	Experience of delivering real impact in a project	Give examples of the impact of your work on specific projects	It is important to demonstrate the benefits of your work and prove that it has a real impact.
Leadership Skills	Experience in coaching, mentoring and line management	Examples of managing people, mentoring or teaching	This is for senior data scientists who would be expected to manage junior employees

VI. Soft Skills

Criteria	Requirement	Evidence	Why
Soft Skills	Excellent written and communication skills	Evidence of stakeholder management, influencing senior managers, presentations to business/non-technical audiences. Official training courses and mentoring / coaching received	Data science is more than just doing the analysis. You need to get your work actioned in the real world for it to have any consequence. This requires being able to communicate effectively and influencing decision makers.

VI. Personality Traits

Criteria	Requirement	Evidence	Why
Personality Traits	An open mindset and curiosity	Regular MOOC's or other training courses completed, conferences and workshops attended	To stay at the forefront of data science it is important to always be keen to learn new things as the area is changing very rapidly
	Teaching others and working in a team	Examples of team projects and teaching / mentoring	You can't know everything, so it's important to share knowledge and work collaboratively.
	Hard work and determination		Other team members may be dependent upon your analysis so it's important to keep on schedule

CONTACT DETAILS



Dr Marcus Alexander Ong
Chief Executive Officer
marcus.ong@spectraanalytics.com



www.spectraanalytics.com



info@spectraanalytics.com



0207 866 0604



Dr Dan Sprague
Chief Science Officer
dan.sprague@spectraanalytics.com



Steve Williams
Head of Business Development
steve.williams@spectraanalytics.com