# Data Science Skills in the Energy Sector:

## Survey Results

Dr Stephen Haben & Elisa Evans
Digital & Data Consultants

**March 2023**

# CONTENTS

# Key Insights from the survey

**40%**

found it very difficult to hire data scientists with the skills required

**95%** of data scientists believe the

Most common modelling technique implemented is

**forecasting**

**39%** had teams of 4 or fewer data scientists

In the last **5 years**

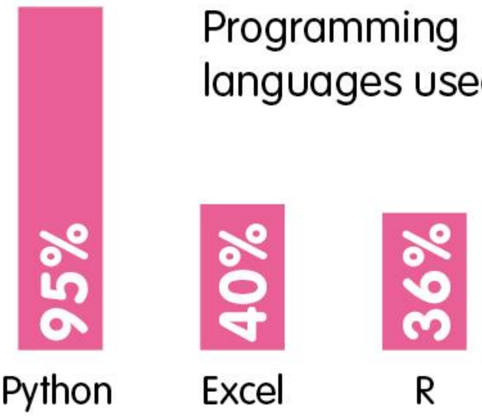
**68%**

of data science teams were created

**50%**

of data scientists had a masters level degree or above

Programming languages used



# 1. EXECUTIVE SUMMARY

Digitalisation is driving vast opportunities for new services and innovations in the energy sector. It will enable new business models, better consumer products, and support the UK government and industries in meeting their Net Zero ambitions. Applications could include:

- Sophisticated demand side response to help support load balancing.
- Smart storage control algorithms to improve utilisation of renewable energy.
- Smart meters to reduce costs through dynamic prices or improve energy efficiency.
- Electric vehicles enabled with two-way communications to provide distributed control and reduce network strain.

Proper and optimal implementation of these products and services will require personnel with the necessary data and digital skills. In particular, it will require new roles in data science. This is the subject of this survey report.

A data scientist's role is to analyse data so as to extract actionable knowledge, and requires expertise from a whole host of different disciplines, including statistics, machine learning, data analysis and computing. The tasks vary but include: developing visualisation tools to better illustrate and understand relationships and patterns in data; creating and training advanced machine learning algorithms for modelling and prediction; selecting and engineering new features to improve the accuracy and the understanding of a problem; and developing error metrics and measures for meaningful assessments and evaluation.

Compared to many other areas and applications such as fintech and social media, data science within the energy sector is a comparatively immature discipline. Data sources have been fairly limited, and challenges have been more suited to traditional engineering approaches. This is expected to rapidly change as society moves towards a low carbon economy. The roll out of advanced metering infrastructure and smart meters have increased visibility and data concerning the networks and consumers, inviting more data-driven solutions. Further to this, the roll out of controllable low carbon technologies (such as electric vehicles and heat pumps) with two-way communications capabilities have enabled a whole host of new and exciting challenges which can only be optimised through advanced algorithms and methods from artificial intelligence and machine learning.

The recency of data science within the energy sector means there is very little information and insight into the skills landscape, including the respective approaches and challenges. This report is an initial attempt to bridge this gap and better understand a number of questions including:

- How is data science being implemented with the energy sector?
- Who are the leaders of the emerging data science teams?
- What are the challenges in recruitment, and continuing development and training?

The answers to these questions have been gathered through a short survey which have targeted data scientist practitioners in the energy sector or those actively managing them. Although the survey has focused mainly on the UK there are also some responses from international organisations. Although this is a relatively small survey sample, the difficulty in recruiting hundreds of participants is perhaps indicative of the energy sector immaturity in this area. We expect that as the opportunities from the energy sector become evident, there will be a rapid uptick in organisations trying to build their data capabilities. This will of course only put further stress on the data skills gap.

The responses gathered in this report have gathered some insight into the data science and analytics landscape in the energy sector. The following are some of the key findings:

**Respondents:** Most of the respondents were directly involved with the data science team or were experienced with data science/analytical work. This strengthens the evidence contained within this report since it has come from those with direct knowledge of data science, data analytics or the process of using the outputs.

**Implementation:** The value of data science across the energy sector is apparent. The methods and techniques are being applied to a wide range of applications and data sources, some of which are expected (demand forecasting), but also in very new areas (text analysis of social media data). To implement the impressive array of complex and advanced machine learning and statistical techniques highlights the need for a wide range of skills and knowledge in both data science and the technology stack.

**Data Science Teams:** Data scientists teams within organisations tend to be relatively small and all concentrated within a single team or a couple of teams. These teams also typically developed in the last few years, although some emerged from technically adjacent disciplines.

Data scientists/analysts typically require technically skilled individuals with most having at least an undergraduate degree in a technical subject such as engineering or mathematical sciences. Many others have postgraduate degrees. For those who did not join straight from academia, they came from other energy industries or other data-driven organisations. This perhaps highlights the desire for energy domain knowledge in addition to data skills.

**Leadership:** Most leads and managers of the data science/analyst teams are data scientists themselves with postgraduate degrees, suggesting technical skills are important to leading some of these teams. This is supported by other responses where two-thirds suggest that the data science lead/manager should have high technical skills. However, this is not the only requirement. Many respondents suggest that although technical skills are important so are managerial, communication and people skills.

**Recruitment and Training:** There is a large need for data science skills within the energy sector. However, the responses to our questionnaire have shown that organisations are struggling to train, recruit and upskill the talent to meet the demand. Although advanced data science expertise is one prominent area of difficulty, it is other skills that organisations are finding it most challenging to recruit. Domain knowledge and coding skills are two of the most prominent weaknesses and are hindering efforts to produce operational implementation of the algorithms data scientists develop.

The survey was not explicitly designed to develop recommendations, rather it aimed to understand the current landscape so that deeper questions and analysis could be supported. However, based on this survey there are clearly some key areas which will need focus going forward. The following are therefore some of the high-level recommendations which are apparent from the responses to this survey:

1. **Enabling training for future data scientists**. Prospective data scientists must have all the resources and training available to them to ensure they can be immediately valuable to the energy sector. As seen in the recruitment section this means also ensuring that domain knowledge and communications skills are also included, not just the highly technical data science skills.
2. **Upskilling**: many organisations don’t have the time to keep the team up to date with the latest methods and technologies. There needs to be easily accessible resources so that the

cutting-edge research and development in the areas of data science can be identified and used across the sector. This includes making the outputs from academic research openly accessible but also ensuring that the research is shared in easily digestible ways to save time and resources.

3. **Reskilling**: to reach Net Zero, decarbonisation will be a gradual transition by integrating existing technology with increasing deployment of low carbon and renewable technology using a whole-systems approach. As well as upskilling the workforce, there will be a need to repurpose existing skills for emerging technologies such as hydrogen and carbon capture and storage. This will be challenging as we go into uncharted territory full of uncertainty. Data science plays a crucial role in leveraging what we know from existing vectors and can learn from early adoption of new technology. We will need a workforce that adapts, continually develops data science approaches, and embraces digital tools (such as digital twins) to accelerate us to Net Zero.
4. **Support building data science/analytics teams**: Recruiting, upskilling and reskilling are irrelevant if the frameworks and infrastructure are not in place for the data scientists to flourish and optimise their outputs. This includes understanding the skills needed within the organisation, that the leadership is chosen appropriately, and that the team has the right mix of personnel.

This paper is just the first step in understanding the data skills gap and have hopefully built a solid foundation for further investigations. The next steps will be to go both deeper and wider with the investigation, analysing the personal challenges within different organisations, and incorporating more views from around the world. However, it will also require looking much more closely at the interactions between the data science teams and the wider businesses. What are the skills that are required to understand and make maximum use of the data-driven insights? What are the other digital skills which are needed to facilitate dissemination and wider learning? Perhaps more importantly, what are the ways that businesses can create an environment where data scientists can thrive and extract the maximum value from the data they analyse?

# 2. INTRODUCTION

The digitalisation of the energy system will lead to many opportunities in the energy sector. From new products such as dynamic pricing, new ways to control and manage the network, and completely different business models (for example, local energy markets, and heating as a service).

These different products and business models will also require different skills. Data scientists have an increasingly desirable role in this context. The Global Energy Talent Index [1] surveyed 10,000 energy professionals of 144 nationalities across 161 countries. In the 2022 edition, the importance of technical skills gap was highlighted: "Technical skills[1] are now the most transferrable skills for all energy professionals and the most important for future-proofing all energy workforces." Artificial Intelligence (AI) and automation is becoming increasingly important throughout the energy sector to support the transition to Net Zero.

Despite this, there has been a consistent shortage in the required personnel. For example, in 2020 QuantHub found[2] there were three times as many job postings for data scientists than job searches. Skills are a major component of the UK's National Data Strategy [2], whose findings included:

- The need for the formal and vocational education system to **better prepare those leaving school, further education and university for increasingly data-rich lives and careers** [3]
- **Industries needing to develop their understanding of their own data skills needs**, including how to define and source these requirements, and how to develop or source employees with the right mix of sector and specialist knowledge

The Covid pandemic has accelerated the digitalisation process in the wider economy [4], but, without the digital and data skills within the energy sector it will be impossible to tackle the challenges Net Zero presents. There has been increasing discussion of the general skills issue for Net Zero, however, there is very little information on the specific area of data science skills in the energy sector.

The aim of this report is to begin to improve our understanding of the data science skills landscape in the energy sector through the use of a survey targeting those on the cutting edge of implementing advanced data analytics in industry. The aim of this is to understand:

- What is the skills landscape in the energy sector?
- What are some of the methods and techniques deployed within high quality data science teams?
- What makes a good data science lead, and what qualities are less useful?
- What are the difficulties in recruiting members of data science teams in the energy sector?

This survey will hopefully help inform further investigation and recommendations for tackling the skills gap. The digital skills shortage costs the UK economy £12.8 billion a year according to Cebr economic modelling for Virgin Media O2 [5].

Before proceeding to the results of the survey we provide a brief overview of some of the related surveys in data science and the energy sector.

[1] For this report "technical skills" will refer to specialised expertise and knowledge for practical tasks. For our purposes these tasks will include data modelling, analysis and evaluation.
[2] https://quanthub.com/data-scientist-shortage-2020/

### 2.1. DEFINITIONS

For this survey we are considering data science and data analytics applications within energy focused organisations. These include consultancies, distribution network operators, research organisations and commercial businesses.

We purposefully stated data science and data analytics, as the main focus is on creating value from data. Hence, we didn't want to exclude the important data analysis, visualisations and statistical modelling that would be part of business but not fit under the stricter data science umbrella.

### 2.2. DATA SKILLS GAP: SHORT REVIEW OF RELATED LITERATURE

In an increasingly digitalised society, there are concerns with the gap between skill needs and supply. To better understand the skills gap, several surveys have been conducted and the majority of them are sector agnostic. This section will discuss a few of the findings from some of these surveys relevant to this report.

#### 2.2.1. SURVEY REVIEWS

An early report from 2015 by Nesta showed that "Two–thirds of datavores[3]who tried to recruit analysts in the previous 12 months struggled to fill at least one vacancy." [6]. As mentioned above, a similar report was found by QuantHub in 2020 which found three-times as many job listings as job searches [7]. In addition, other findings from the same from QuantHub included:

- Companies are trying to capitalise on the current workforce by upskilling and reskilling employees, so they become more self-sufficient and less reliant on a small selection of expensive experts.
- Universities and higher education institutions have been developing data science and data analytics courses to prepare school-leavers for the market.

In 2020 the UK's Department for Digital, Culture, Media and Sport commissioned Opinium to perform an extensive survey to understand the UK data skills gap [8][4]. They contacted 1,045 businesses, 5,000 workers and 1,000 students in higher education or training across the UK. Some of the insights highlighted:

- Small businesses were most significantly affected as they generally had lower performance across all data skills.
- Small businesses were less likely to have existing data roles or be recruiting for new data roles.
- Cost is a major barrier to establishing data teams within businesses, due to a shortage of suitably skilled people, budgets are stretched to attract talent or work must be outsourced where it is cheaper.
- Businesses with a Chief Data Officer (CDO) are twice as likely to have a clear digital strategy than those without.
- Students don't receive adequate guidance on how to become a data scientist.
- The hard skills data gap is larger than the soft skills gap[5].

---

[3] Datavore – Nesta's name for "those businesses which rely on data for commercial decision-making"
[4] It should be noted that this report took a broader definition of data skills than this survey and included Basic IT skills, Database management, programming, as well as machine learning and advanced statistics.
[5] The report defines soft skills as those such as soft skills such as Collaboration, Project management and Storytelling etc.

- A gender divide emerges where male workers feel more confident than female workers regarding their hard skill capabilities, their most highly rated soft skills are industry/sector expertise and leadership whereas females feel their strengths are professionalism and curiosity
- The gap will only be closed when education and training is provided covering both hard and soft skills. A variety of training options will be needed which can also cater for those constrained by budget.

Another useful survey on data science skills more generally is the annual Kaggle survey. The most recent 2022 results (using responses from 23,997 Kaggle members from 173 countries) revealed [9]

- Respondents identified as predominantly male.
- Python and SQL were the most common programming skills.
- Sci-kit learn was the most popular machine learning tool.
- AWS, Google Cloud Platform and Microsoft Azure again saw year-on-year growth.

A caveat to this survey is that it may be focused primarily on those entering Kaggle competitions rather than those which are implementing data science within their profession.

There are also some surveys which look at the energy sector skills more generally. The Energy Barometer from the Energy Institute released a survey looking at the skills in the energy sector [10]. In the 2021 edition (which received responses from over 400 professionals across the energy system) many concerns about skills were raised, including:

- Skills and human capacity issues may eclipse technology as the biggest barrier to meeting Net Zero.
- Availability of appropriate training courses is lacking, either due to time or financial constraints.
- Oil and gas workers are most likely to be unsure of their training requirements as they transition to a greener future.
- The UK energy sector will need to fill 400,000 jobs between now and 2050 – most will be new roles and will have a long lead time to bring these skilled workers 'online'.
- Limited numbers of young people are choosing STEM qualifications, there needs to be a cultural shift in favour of vocational qualifications as university degrees might not be necessary and gender norms challenged.

There are also investigations into skills in specific subsectors of the energy system. In 2021, Robert Gordon University together with OGUK released a report on data and digital skills for the oil and gas industry [11]. The report showed rapid changes where "increasing data fluency and data dependency within organisations is challenging established organisation and leadership models". There was also great uncertainty as to what data and digital skills they have today will be necessary in the near future, and concern from data professionals if these organisations will help support their skills development. In the fast-moving world of digitalisation, keeping up to date with the latest technology and techniques is becoming a common problem across emerging and established data-driven organisations.

In November 2022 The Energy Institute was commissioned to complete a Hydrogen Skills Gap Study Landscape Review [12] assessing the missing skills required for a future hydrogen economy. They identified an ever-increasing skills gap through the loss of skilled and experienced workers due to retirement which isn't being replenished with new entrants. Part of the solution would be to increase the variety and potential routes for trainee data scientists, not neglecting the current workforce who can be reskilled or upskilled to meet demand. Another suggested way of increasing the number of data scientists would be to provide fast-track pathways that are funded and provide contextualised knowledge training via a variety of methods, not necessarily requiring lengthy stints in academia.

Hydrogen as an emerging energy vector also presents some unique challenges as it can be formed via different routes and storage/transport requires careful safety monitoring and management. The amount of data required to manage the supply chain and forecast will be non-trivial and hydrogen will complement other energy vectors. The need for multi-discipline collaboration will be required, leveraging data science knowledge gained from other low carbon technologies. In times of uncertainty, we will be looking to data scientists to reach our Net Zero goals.

The rapid digitalisation of the energy sector will certainly mean that data skills are one component of the skills gap in the energy sector.

### 2.2.2. DATA SCIENCE SKILLS IN THE ENERGY SECTOR

Unfortunately, information outlining the data science skills required specifically by the energy sector is scarce. There is some coverage in the paper 'Data Science and Energy: Some Lessons From Europe on higher Education Course Design and Delivery [13]. The authors concluded that the energy sector struggles to attract and train data scientists mainly due to a lack of emphasis on data science in graduate programs and high barriers of entry for data scientists from other sectors. The reasons why are likely to be multi-dimensional and includes factors such as financial incentives, challenge maturity, data availability and different regulatory environments. For example, unlike the FinTech sector where there are free employee-embedded data science training options delivered through partners [14], financial support is less readily available in the energy sector.

The culture of utilising data and machine learning is not as embedded as other areas such as bioinformatics which began over 50 years ago [15]. Advances in bioinformatic tools, and the arrival of biological Big Data have allowed computer science and biology to become intertwined and fully integrated into many university curriculums. However, there appears to be some progress in the areas of energy analytics, with many more specialised courses and degrees in these areas.

Another challenge is the significant paucity of data compared to other sectors. As a regulated industry, energy data has often been siloed or stored, for privacy, commercial reasons, and no requirement has been made it more open. Further, there simply wasn't as much data collected. However, having said that, energy is now at the forefront of everyone's mind as we search for ways to tackle the climate crisis. The Energy Data Taskforce [16], and similar initiatives such as Italy's Digital Transformation [17] and around the rest of the world are addressing the need for data to be made available to researchers, innovators and others in the energy sector. For these reasons the energy sector is now moving into a modern digitalised era.

This is an exciting period for energy where more data is being made available than ever, and new technologies are emerging. However, this produces numerous issues:

- Organisations and employees in energy sectors, especially management positions, are less familiar with data-driven methods, and even less with machine learning and the latest data science techniques.
- Further, such individuals may be very mistrusting of the uncertain nature of some of the outputs and models. Energy has traditionally been a sector which has built-in contingencies to ensure nothing fails.
- The supporting data infrastructure, processes and organisation required to deliver production data science are often immature or non-existent.
- Organisations may be unaware of the potential of data science and the opportunities available.

- Those in charge of hiring data science teams may be unaware of how to find, recruit or train the best talent.
- In a rapidly changing area of data science, new technologies may be emerging all the time. The technology stack to use can be a daunting situation.

The following survey will show parallels with many of the above literature cited, but will also show some of the issues specific to the energy sector.

# 3. METHODOLOGY

The aim of our survey is to gain an insight of data science/analytics practices with energy focused organisations. For these reasons we were not centred on academia, government or general organisations. Primarily we were interested in energy organisations which had some data science/analytics staff, whether permanent or contractors. We typically targeted data science/analytics staff, and, if possible, the leads or managers of such teams. Hence the methodology was primarily to target individuals in our immediate networks. However, to add some breadth to the search we also utilised our newsletter and other data focused email groups such as Power Swarm [18] where lots of energy organisations are signed up.

The survey has some limitations, organisations will primarily be UK based as a natural consequence of the density of such contacts within our networks. However, where possible we tried to expand the search to organisations which are further afield. In addition to this, to encourage more open answers to our survey questions we ensured that no individual could be identified, and the answers could be submitted anonymously. The drawback to the approach is that we could not clarify any remarks if they did not provide their contact details, and of course we couldn't infer any other information about their answers. It should also be noted that this survey was first shared on July 2021 and made available until we started processing the results at the end of 2022. This means care should be taken with interpreting some of the results, especially those which discuss the age of the teams etc.

The limitations to the results are deemed acceptable for this initial survey. As there is very little information in the specific area of data science skills in the energy industry the drawbacks to the answers are small in comparison to the evidence that will be provided by this survey. The hope is that this is a preliminary report from which more extensive and rigorous methodologies will be possible.

# 4. ABOUT THE PARTICIPANTS

The first set of questions were to understand who was taking part in the survey. This included their position in the company, what type of organisation it was, and whether this individual was part of the data science/analytics team. This ensures the answers given throughout the survey are by those who are directly or indirectly involved in the data science/analyst implementations and can give informed answers.

The first level of investigation is the organisation itself. The categories are based on the standard ones used by ESC to classify organisations and are relatively broad to maintain anonymity of the respondents. The organisation types are shown in the Table 1 and the proportion is shown in the pie chart Figure 1.

*Table 1. Count of the different interview respondents by Organisation Type.*

| Organisation Type | Number of Respondents |
|---|---|
| Utilities or network operator | 18 |
| Small to medium sized enterprise (SME) | 11 |
| Large Business | 6 |
| Research organisation or business school | 2 |
| International organisation | 4 |
| Catapult Network | 1 |
| **Total** | **42** |

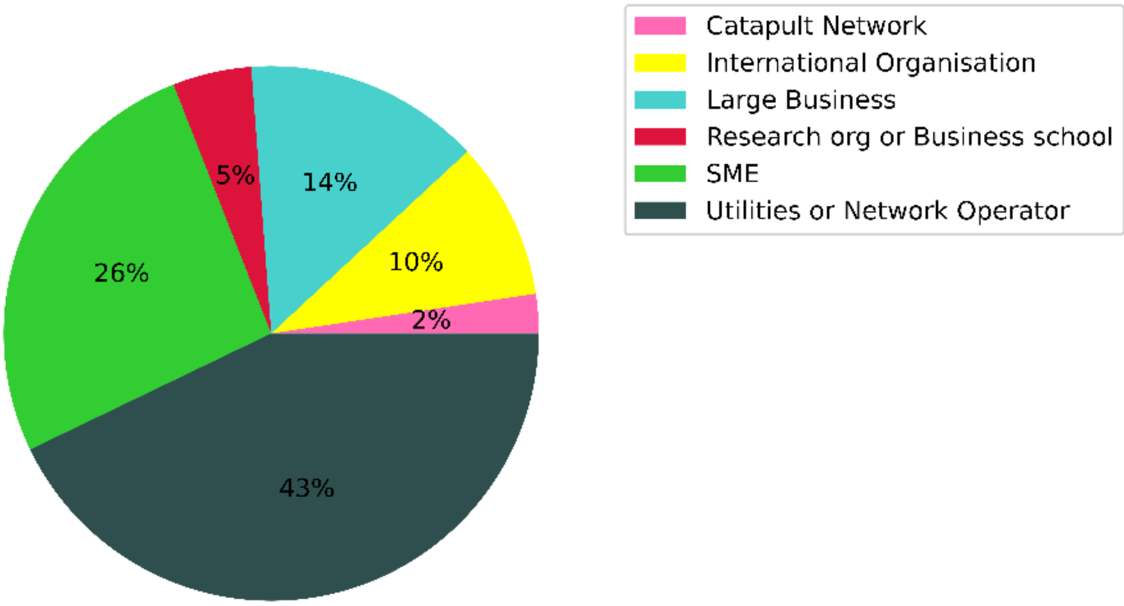



*Figure 1. Composition of the survey respondents by organisation type.*

Figure/Table 1 shows that primarily respondents were individuals from the utilities or network operator sector (18 in total), with the next largest responses from SMEs with 11 individuals. This is encouraging since these organisations are ones where data-driven innovation is expected or needed to drive the energy sector towards a modern digitalised system.

We next asked the individuals "**What is your job title?".** Reassuringly the most common answer was Data Scientist although this was broken down into other roles within the team, e.g. Senior Data Scientist, Head of Data Science, etc. In addition, it was found that not all of those who led the data scientist team had data science in their title. Hence Table 2 only includes those who gave their role as "Data Scientist" without "Head", "Lead" or "Manager" in the title. For these reasons Heads, Leads and Manager roles were not included in the initial count of data scientists as shown in Table 2. Similarly for the Analysts.

*Table 2. Breakdown of the role of the respondents to the skill survey.*

| Main Role in title (grouped) | Number of participants |
|---|---|
| Data Scientist (Non lead/Managerial/Head) | 13 |
| Data Analyst | 2 |
| Researcher | 1 |
| Consultant | 2 |
| Vice President/Chief Technical Officer | 2 |
| Engineer | 6 |
| **Total** | **26** |

Since managers and data science leads were not included in the data science category above, it underestimates the number of respondents directly involved with the data science/analysts. In fact, 17 of the respondents led the data science team (more information on leadership is in Section 7). Of these, 2 gave their roles as Data Scientist, and 1 as Analyst. Accounting for these, it turns out there were at least two-thirds (28) out of the total 42 participants who were directly involved with the data science team either as analysts, data scientists or the lead.

To be more explicit about the team structure we also asked **Are you part of the data science/analytics team in your organisation?** Of the responses, 32 of the 42 responses (i.e., over three quarters) were from the data science/analytics team in their organisation. However, note this does not mean that the other 10 did not implement data science themselves since several of the responses noted that there is no data science team in the organisation (See Section 6). The 10 not in the data science team had a mix of other professions:

1. Business Development Manager
2. Product Manager
3. Vice President of Product
4. Chief Technical Officer
5. Head of Energy Assessment
6. Head of Smart Networks
7. Head of Architecture
8. Research Engineer

Note that 1 of the above who stated they were *not* in the data science team led the data science team. Of the 10 not part of a data science team, 8 did highly technical data work, or explained their experience working in their role in the previous data science team. One person didn't give a response and the final participant worked directly with the data science team.

**Summary:** Most respondents were directly involved with the data science team or were experienced with data science/analytical work. This strengthens the evidence contained within this report since it has come from those with direct knowledge with data science, data analytics or the process of using the outputs.

# 5. IMPLEMENTING DATA SCIENCE

The next questions were aimed at understanding how and why data science is applied within the organisation. This includes looking at the applications, the type of data analysed, the type of algorithms implemented, and the skills required to perform the modelling and analysis.

The first question asked the respondent to **Please provide a short summary of your own experience with implementing data analytics/science within your organisation.** This was a free text option, so we manually compiled the main responses to highlight both common actions but also some other less typical responses.

The primary activity mentioned was forecasting or predictive modelling. This is applied to many different aspects of the energy system including load, rooftop solar generation, solar farm generation, wind farm generation, and prices. In addition, climate forecasting was mentioned by one respondent for more medium-term forecasts (up to six months). In fact, dealing with weather variables is referenced by many of the respondents.

In addition to the forecasting tasks there were many applications mentioned, some of which would require forecasts to achieve. The tasks include expected demand side response estimates, network management, network fault prediction, network planning (in particular, to handle the future uptake of low carbon technologies), anomaly detection, Electric Vehicle (EV) route planning algorithms, layout optimisation for offshore wind farms, and energy management systems (including battery energy storage systems).

With the increased installation and availability of smart meter data, there were also increased applications at the individual consumer level. These again include demand side response and energy management at the consumer level, but also Internet of Things (IoT) platforms, consumer churn analysis, house energy system modelling, and electricity and gas customer segmentation and characterisation.

There were different levels of data science capabilities and maturity in organisations. There are some firms with large consultancy teams supporting other organisations such as the National Grid Electricity System Operator (NGESO) and Distribution Network Operators (DNOs), and there are other firms which are at relatively early stages and are focusing on the low hanging fruit.

Machine Learning Operations (MLOps) is a common feature in the responses. It was stated that it is important to have a "robust and well-planned data architecture in place" before the data scientist gets on board. However, many organisations have pulled data scientists into doing data engineering and architecture jobs. Similarly, data scientists are being employed on tasks which could be completed by less specialised people. These jobs take up a lot of time and it makes it difficult to implement their speciality. Information management (data quality, and systems integration) is another related task mentioned and performed by our respondents.

Difficulties were also mentioned: some data sources e.g. Balancing Mechanism Reporting Service (BMRS) were stated as difficult to interpret, and weather data can be hard to find and expensive. There was also a surprising lack of big data from sensors at the transmission level. Poor documentation was another difficulty.

Finally, there was mention of working across the organisations. Many senior leaders want to use AI to "transform the business" but lack the understanding to translate this into effective action. However, others reported a good collaboration with the business team and software developers but scaling this collaboration to the entire company has proved difficult. There was a need to manage the translation between teams, especially business units and IT support.

The next question **Is there a particular department/business activity that typically relies on the insights generated by the data science team?** This was designed to dive deeper into the activities of the organisation, beyond the person answering the survey and to find out who derived the most value out of the data science/analytics insights. Of the 42 responses, 2 said no, and 4 said yes without further clarification. However, the other 35 provided some further details. They can be split into internally and externally used insights.

| Internal |
| --- |
| Trading and Operations |
| All Teams/Departments |
| Finance |
| Sales and Marketing |
| Product |
| Consumer Insights |
| Board Level Requests |
| Business Team |
| Decision makers from each business unit |
| Commercial |
| Innovation |

| External |
| --- |
| Clients e.g. planning teams |
| Utility companies |
| Energy Markets |
| Financial Institutions |
| Electric grids |
| Energy Systems |
| |
| |
| |
| |
| |

Trading & Operations was by far the most common department mentioned by 7 out of 42 participants. Three respondents stated that all teams relied on the data science teams insights to some extent.

Most respondents used their insights for internal use supporting other areas within their organisation to make data-driven decisions that would affect their bottom line. This may not be surprising as >90% of the organisations were commercially focused rather than being engaged in research or education.

One respondent said: "Not really. Data science isn't really doing 'insights'. This is an interesting answer as data is processed in different ways according to a need. The definition or interpretation of what constitutes data science or 'insights' may vary across the sector. This respondent processed time series (smart meter and pricing) customer and customer performance data, and they identified as a utilities or network operator. Perhaps the survey could have asked if the *outputs* of the data science team are relied upon by other teams around the business.

We next asked for some **High-level examples of some typical insights.** These could be grouped into roughly five areas, which are listed below.

**<u>Modelling & Forecasting</u>**

- Energy consumption, peak demand and usage forecasting.
- EV charging and electricity prices.
- Fault prediction.
- Market trends, pricing and trading recommendations.
- Predicted demand of low carbon technology over time.
- Scenario modelling.

- Consensus modelling.
- Life expectancy of assets.

**Performance evaluation**

- Forecast validation and benchmarking.
- Substation rankings.
- Understand the impact of adverse weather event and lessons learned.
- Cost to benefit analysis of optimisations to justify fees.
- Photovoltaic (PV) power plant efficiency analysis to identify shortfall (if any).
- Plant behaviour changes due to predictive models.

**Optimisation**

- Asset optimisation.
- Underutilisation of energy assets during high PV generation within a locality.
- Improve load forecasting performance.
- Reduce uncertainties and generate value.
- Calibrating energy system models.

**Research & Development**

- Development of applications that manage microgrids and hybrid parks.
- Private wire opportunities.
- Machine Learning (ML) research to develop new solar PV forecasting techniques.

**Consumer Affairs**

- Understanding user behaviour in relation to EV charging.
- Creation of a linguistic rule taxonomy using Natural Language Processing (NLP) to understand and remedy customer complaints.
- Identifying high-risk customers.
- Identifying which customers and type of customers might be a top performer for a particular energy program.

A couple of responses also expanded on how text analysis was used:

“Typical questions we might help with: Why is customer contact so high this week? What has been the trend in retail pricing this week? Can you help us target Warm Home Discount communities?”

“Use Natural Language Processing for sentiment analysis in emails. Human characteristics and behavioural analysis for vulnerability. Many projects looking to make predictions from data of poor quality, i.e. high uncertainty.”

The follow up question asked **How would you describe the core machine learning models the data scientists/analysts frequently apply?** This would enable a deeper understanding of how the applications, services and insights are achieved with data science. Relatively high-level categories in machine learning were given for respondents to choose from including supervised learning (Sup Learn), unsupervised learning (Unsup Learn), optimisation, control, forecasting, reinforcement learning (Reinfor Learn) and Other. The results are summarised in Figure 2. Given the above discussion, it is perhaps unsurprising that large number of respondents identified forecasting as the most common machine learning technique applied within their organisation. Further, since regression is a major component of forecasting, it should not be a surprise that Supervised learning is also very common, although it can obviously be applied to many other applications. The next most common is optimisation. Optimisation is a fundamental component to network modelling, asset management and demand management so this is to be expected. Optimisation has traditionally been applied by mathematical sciences rather than data scientists, however as the energy system becomes more complex, it would be interesting to learn if more meta-heuristic techniques such as evolutionary algorithms are increasingly being applied.

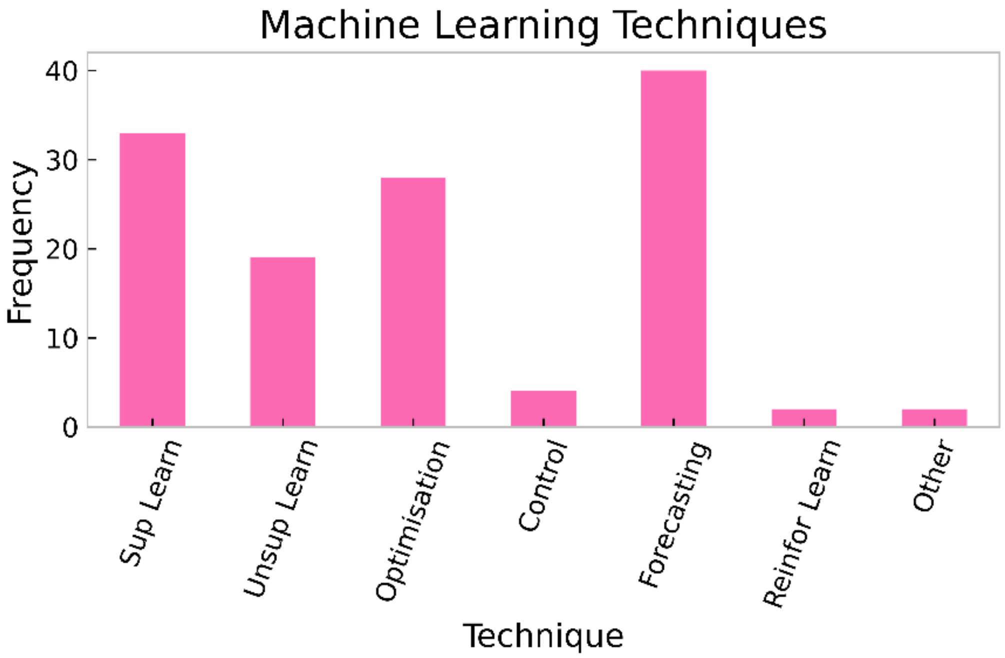



*Figure 2. Count of the core machine learning models which the data scientists/analysts apply, tagged by the survey participants.*

Two areas of machine learning, reinforcement learning and control are very sparsely applied in the organisations who responded to the survey. Since reinforcement learning is relatively new compared to forecasting and optimisation, perhaps this is natural. However, control is a big component of engineering solutions so could be more common than was exhibited in the responses. It certainly would be a component of building energy management systems, and network control. The reason it did not feature prominently in this survey could be because it has traditionally been applied within engineering roles rather than data science, although this may change as the control systems utilise more data and become increasingly complex.

The next question asks, **what are the most common types of data the team analyses?** There was a wide range of data types that were included in the survey response: Time series, Geographic Information System (GIS), asset data, etc. Time series data is the most common data type analysed. This included demand, generation, smart meters, pricing, EV charging, and meteorological data. However, also quite common was GIS data, and data associated to consumers such as survey information, tariff data, and demographic data. This focus on consumers also extended to email clicks and social media information.

Asset data was also very common for both condition monitoring application and to understand the degradation of batteries. Further, staff data was also mentioned to improve resource allocation. More novel data sets included image and text data. Satellite data is very useful for PV forecasting, and text analysis was useful for sentiment analysis and better customer service. Large volumes of text would lend itself to techniques such as Natural Language Processing. Considering fuel poverty and vulnerable customers is of heightened importance during these turbulent times it is perhaps unsurprising that 'demographic', 'property, 'health', 'social' and 'media' also make an appearance (albeit minor). This links with consumer insight activities using AI to understand customer behaviour.

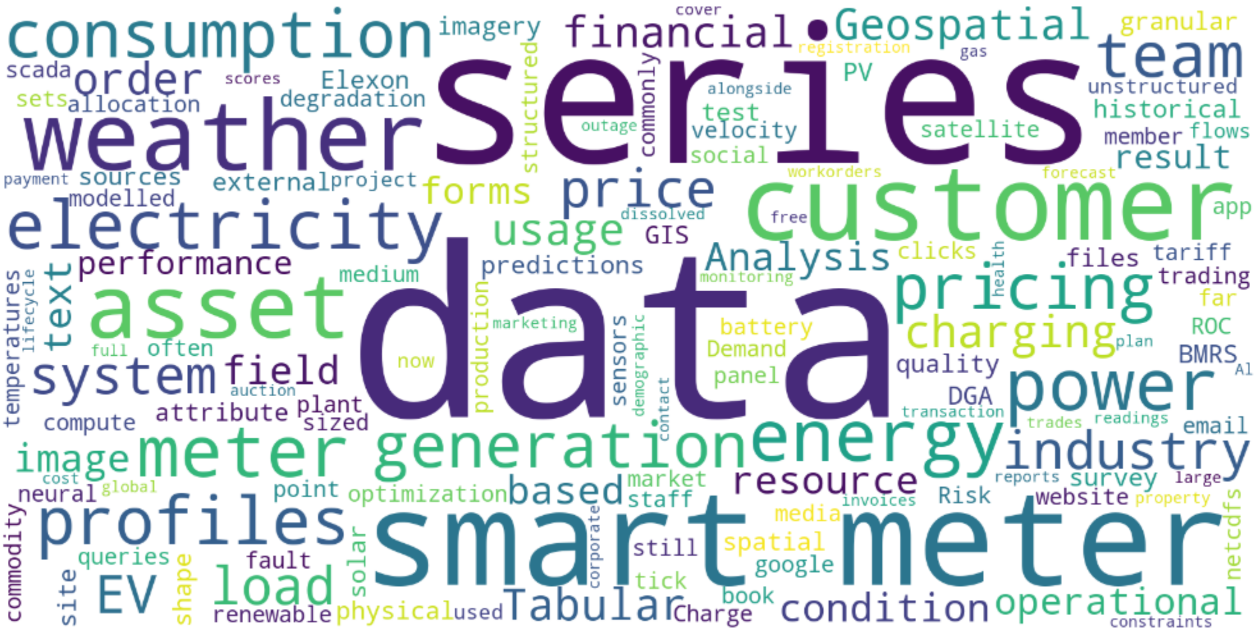



*Figure 3. Word cloud showing the most common words describing the data the teams analyse.*

To summarise the most prominent text in the responses, the word cloud in Figure 3 was created. Note that this shows that 'series' data was the most common datatype. Due to the variety of ways 'time series' can be written including typing errors, the word 'time' was filtered out for clarity - however 'series' does infer 'time series'. Related to that, the most prominent words describe the data e.g., electricity data from customers/assets.

**What is the typical size of the data processed?**

The increased amount of monitoring and development of tools to analyse data means that many organisations are become more data-driven. This brings new challenges in terms of being able to manage, pre-process and develop models. The current amount of data in the energy sector is unlikely to be as large as other data-driven organisations in social media and finance.  This question tried to generate a rough idea of magnitude of the data that these organisations are now having to deal with.

The answer to this question was given in a free text box so the answers are diverse and often given as a range. Most responses gave an approximate answer in terms of the byte size. Typical data sizes handled were therefore categorised by the maximum size using the normal naming convention in steps of $10^3$ bytes, e.g., kilobyte (kB), megabyte (MB), gigabyte (GB), terabyte (TB) etc as per Figure 4. Pie chart showing the typical orders of magnitude of the data analysed by the data scientists in the organisations who responded to the survey.

The size of data processed varied widely from small CSV files (kBs) to TBs as it depended on the purpose for processing the data and the type of application/project. For time series it depended on the resolution of the data (seconds/minutes/hours/days) and the duration e.g. (months/years). For energy trading across multiple markets and commodities, large time series datasets accumulated. Data wasn't limited to numerical types, there was also large volumes of text such as one respondent mentioned for analysing all comments on websites such as Reddit.

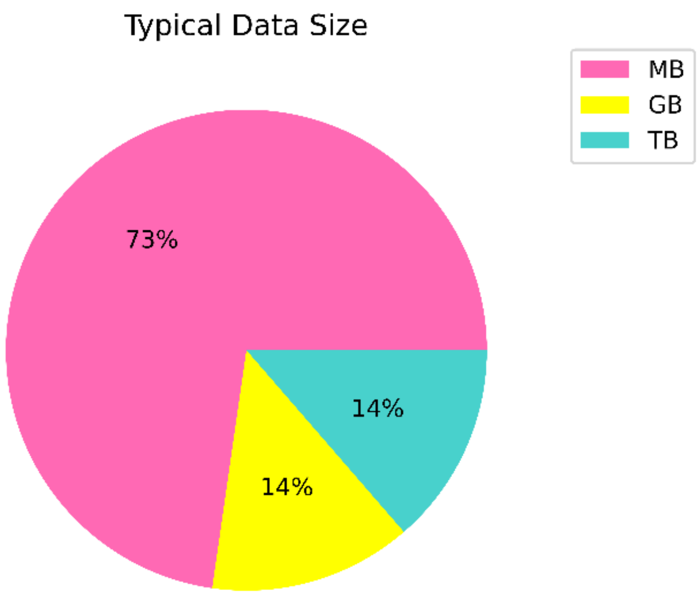



*Figure 4. Pie chart showing the typical orders of magnitude of the data analysed by the data scientists in the organisations who responded to the survey.*

There were no respondents who only processed kilobytes-worth of data, the majority processed at least MBs of data with the rest split evenly between GBs and TBs of data. None of the respondents processed data large enough to warrant measuring in petabytes (PBs) or beyond. This shows that the size of data was relatively small within the respondent's organisations, perhaps suggesting the lack of data available, collected or processed.

The variety of applications, machine learning techniques, and data types and sizes suggest a variety of data science and analytical skills are required. For this reason, we asked **How important are each of the following technical skills in your team?** From which we provided the following list of skills which the respondents needed to score from 1 to 5, with 5 being the most important.

- Databases
- Coding
- Analysis and statistics
- Algorithms and models
- Domain knowledge
- Visualisations
- Feature engineering
- Software development
- Data architecture
- Data engineering

The proportion of responses for each score for each skill is shown in Figure 5, and they are ordered in terms of the skills that participants thought were most important, in this case analysis and statistics.

It is interesting that the two highest rated skills, analysis and statistics, and coding, are relatively dissimilar. The first is related to advanced techniques for understanding the data and is essential for understanding and developing machine learning models. In contrast, the coding is essential for implementing models, but may not require any knowledge of data analysis or statistical modelling. In fact, no participant rated coding any lower than 3 which suggests that, although some skills are not important to some organisations, *all* organisations rated coding highly. The top three answers also align with the results from the recruitment section (Section 8) where advanced data science skills, proper coding skills and domain knowledge were all listed as the most desirable traits that organisations struggled to recruit in.

The low score of software development and data architecture is perhaps not surprising since these are not primary tasks for a data science professional, although it does appear that data scientists are required to do them rather than someone more qualified as seen when we asked about the respondents' experience in the job as outlined above.

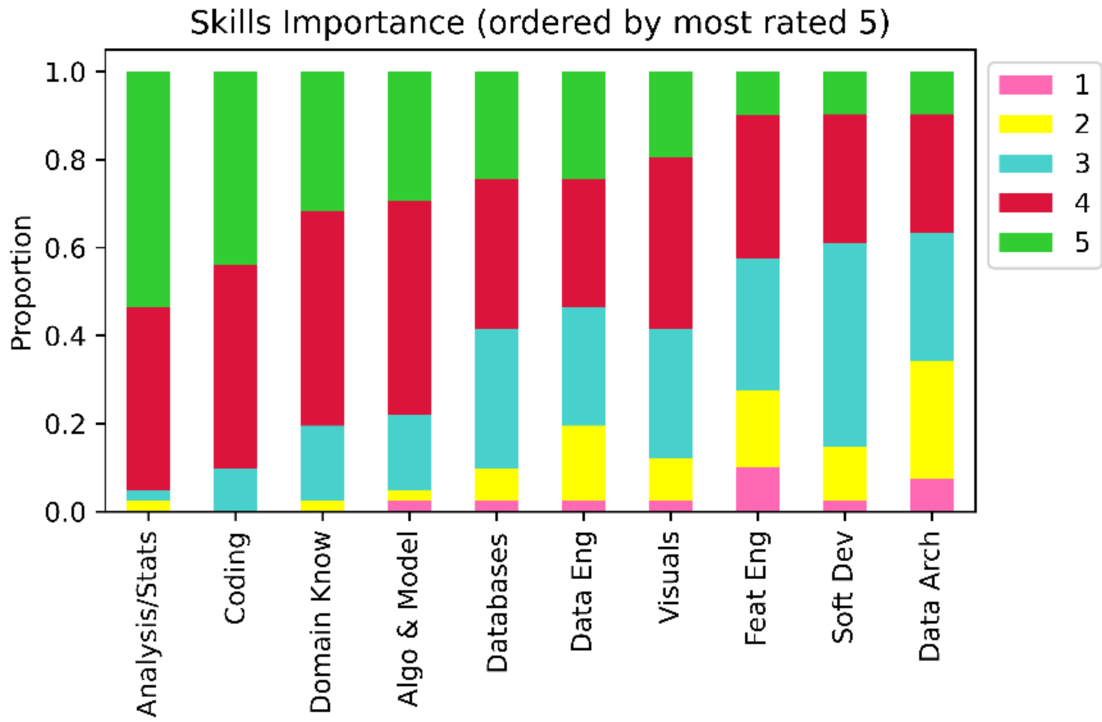



*Figure 5. Proportion of ratings for each technical skill for data scientists/analysts.*

There are also some surprising rankings in the list. Visualisation is a fundamental component of proper data analysis and hence the development of models. However, it is listed below data engineering and databases in the ranking. However, this could be an artifact of the way the skills have been ordered. In fact, if only ratings of 4 and 5 are considered then visualisation is rated as more important. Feature engineering is another surprisingly low ranked skill given that extracting the correct features for input to a machine learning model is as important, if not more so, than choosing the right algorithm. It could be that the participants see the visualisation and analysis & statistics skills as core elements for doing the job and hence have ranked feature engineering as relatively low. Alternatively, many machine techniques include feature selection/extraction automatically (for example Least Absolute Shrinkage and Selection Operator) thus reducing its importance as a skill in its own right.

Although the above techniques may be applied, they could be very basic. For this reason, we also asked **How advanced do you view the algorithmic/modelling techniques being applied by**

**your data science team in the following areas?** (Scale of 1 – 5 with 5 being most advanced). This question focused on the four core areas of data science algorithms/modelling techniques:

- Visualisation
- Data pre-processing and cleaning
- Statistical and data analysis
- Core machine learning models development

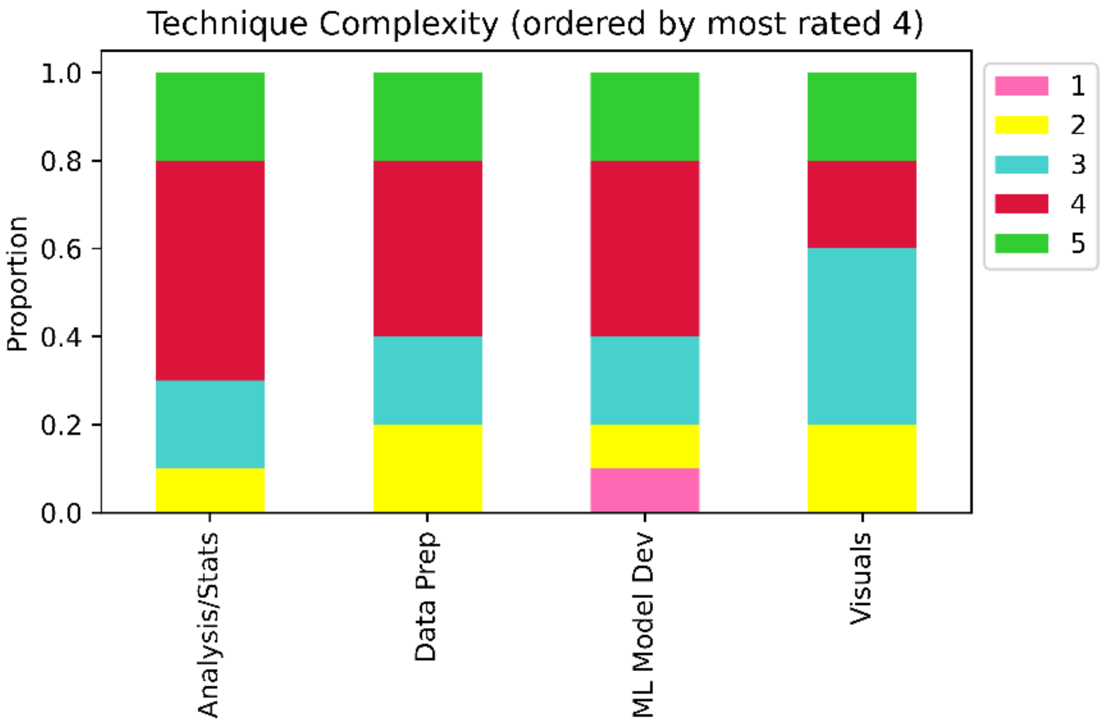



*Figure 6. Proportion of technique complexity for four main data science areas. Ranked according to most ratings of four.*

Figure 6 shows the proportion of ranking of complexity for each technique area. These are ordered by the most complexity ratings of four since each technique had equal proportions of 20% with a rating of 5. The complexity of the visualisation techniques is relatively low. This could be because visualisation techniques are often most useful when they convey simple ideas and show the main features within the data. With time series, this will simply mean time series plots, autocorrelations and pair plots. In contrast the most complex techniques perceived by the respondents is for data and statistical analysis. In contrast to machine learning model developments, advanced statistical analysis requires deeper understanding of core mathematical models and theory to be applied. Perhaps surprising, data pre-processing techniques are seen as more complex. Unfortunately, there is little information in the survey to understand what sort of complex techniques are being applied. It's likely due to the noisy and unclean nature of real data, it requires careful approaches and understanding of the patterns in the data to be able to make it useful for the models developed.

The final question on the implementation of data science, focused on the programming tools used to develop the models and analysis and asked **What programming technologies/tool do the data scientists/analysts most frequently use within your organisation?** This is important for several reasons. As shown above, the size and complexity of the data is increasing, and it is vital that the right tools are utilised to ensure computationally feasible analysis and advanced algorithms can be developed. The count of the various programming languages used by data scientists/analysts by the organisations represented by the survey are shown in Figure 7.

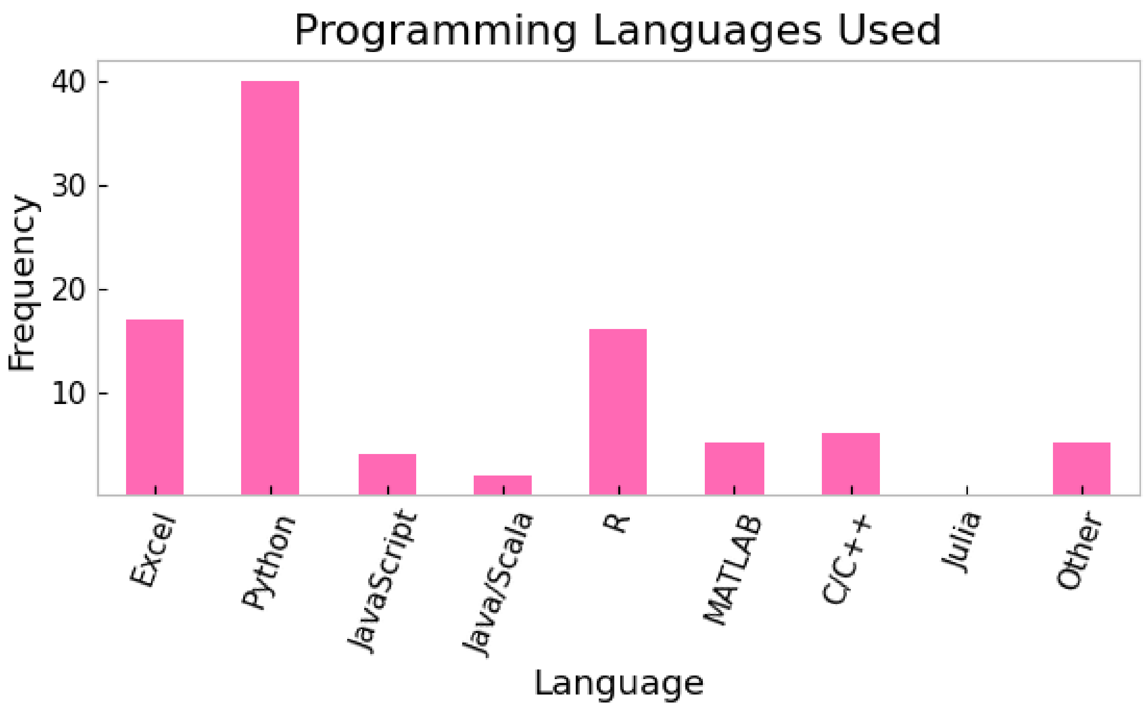



*Figure 7. Count of the types of languages used by the data scientists/analysts in the difference organisations.*

Python and R are, as expected, two of the most used languages, although Python is by far the most popular. In fact, all but two of the respondents stated they used Python (the other two used R) and has become a very versatile tool for analysis, statistical modelling, and machine learning, implemented using a simple and intuitive language (in contrast to C which has a much steeper learning curve). R is traditionally more popular in academia and has more statistical analysis functionality but in fact is less popular than Excel. Excel has historically been one of the most popular tools for data analysis, although perhaps may be ill-equipped for more complicated and computationally expensive modelling. JavaScript and Java/Scala are traditionally more popular for software developers and may suggest the type of tasks data scientists/analysts are being asked to implement within their organisation. MATLAB also had a few users, although the proprietary nature of this software may mean it is often not a common choice.

In the 'Other' languages label this included:

- Golang
- Fortran
- C#
- SAS
- SPSS

In short, although there is a variety of programming languages being implemented, Python is by far the front runner.

**Summary:** The value of data science across the energy sector is apparent. The methods and techniques are being applied to a wide range of applications and data sources, some of which are expected (demand forecasting), but also in very new areas (text analysis of social media data). To implement the impressive array of complex and advanced machine learning and statistical techniques highlights the need for a wide range of skills and knowledge in both data science and the technology stack.

# 6. ABOUT THE DATA SCIENCE TEAMS

The previous section showed how data science is implemented within the organisations of the respondents. The aim of the next questions was to understand a little bit more about the composition of the teams themselves, and how they are structured.

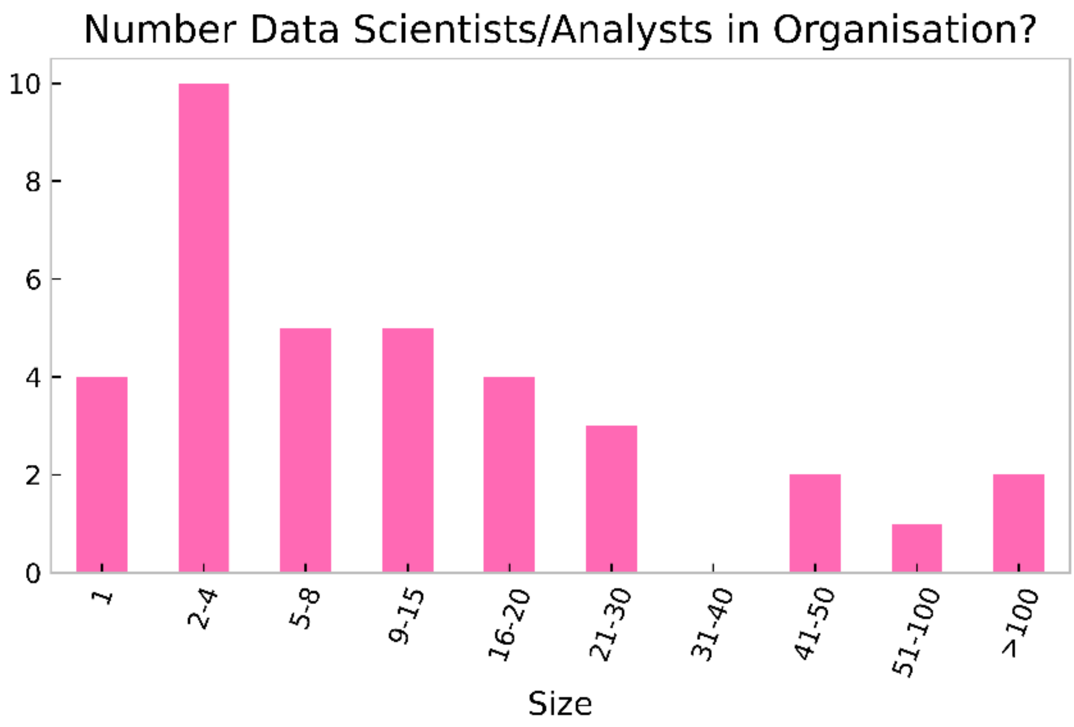



*Figure 8. Count of how many data scientists/analysts there were throughout the organisation.*

The first question was **how many data scientists/analysts are in the organisation**? The results are plotted in Figure 8. Most of the sizes are quite small, with 14 out of the 36 responses had 4 or fewer data scientists/analysts. Two thirds of the organisations had 15 or fewer. However, there were also 3 organisations with over 50 data scientists/analysts with one of about 100, another of 170, and one with over 1000. These are large organisations, and the one with >1000 data scientists is a large international corporation. Although in some responses the number of data scientists/analysts is low, in at least one case, 4 out of 5 members of their entire organisation were data scientists. This was not explored further as there was no dedicated question to probe further. We only know that data scientists/analysts can form a large proportion of the organisation. A future question on the size of the organisation would help to understand this further.

The next question tried to understand **Is there a dedicated data science/analytics team within your organisation?** Within larger organisations it may be expected that there are analytical individuals spread throughout the company but there may also be a specific dedicated team. If there are teams then they may focus on a specific task, or they may support the rest of the organisation with data related analysis and modelling. Individual data scientists/analysts are likely to be supporting a particular team for their data needs rather than the whole organisation. This question was a multiple-choice and the summarised answers are shown in Table 3.

*Table 3. How many data science were part of a team within the organisations*

| Response | Total Responses | Percentage of Responses |
|---|---|---|
| Yes – all data scientists/analysts are within the same team | 13 | 30.1 |
| Yes – but there are a few teams with their own data analyst/scientists | 20 | 47.6 |
| No – but there are individuals within non-dedicated teams | 6 | 14.3 |
| No – we don't have any dedicated data scientists | 2 | 4.8 |
| Other - Lots of dedicated teams from different countries and across various parts of the organization | 1 | 2.4 |

As mentioned in Section 4, three quarters of those individuals who answered the survey came from organisations with dedicated data science teams. By combining with answers to the previous question, we note that the size of teams (whether dedicated teams or one of a few teams) varied from only 2 all the way up to 100. One (which answered as "Other" in the table) had a large numbers of data scientists (40+) and is one of only three international organisations who took part in the survey.

Perhaps of more interest are the 8 organisations which did not have dedicated teams. The number of data scientists was left blank in the questionnaire for 2 of the 8 organisations, and of the other six, four had only 1 data scientist, one had 2, and the last had 3 data scientists and 15 analysts. It could be data science teams would be formed if they grew further, but this is purely speculation as we have no questions regarding specific team development.

Given some of the shortages in skills (see the Recruitment Section 8) an important question is how many of the data science teams are fully employed and how many are contractors. We asked **what percentage of the data scientists/analysts are contractors?** The is shown in Table 4. We found that most organisations had zero contractors. There were only 2 organisations that consisted entirely of contractors, 1 of which only had 2 data scientists/analysts and was an SME. Surprisingly, the largest organisation with over 1000 data scientists/analysts only had 10% contractors which meant that they employed most of the data science talent they needed.

*Table 4. Count and percentage of data scientists/analysts which are contactors.*

| How many of the data scientists/analysts are contractors? | Total Responses | Percentage of Responses |
|---|---|---|
| None | 25 | 59.5 |
| Less than 10% | 6 | 14.3 |
| Between 10 – 50% | 8 | 19.0 |
| Between 50-90% | 1 | 2.4 |
| All of them | 2 | 4.8 |
| **Total** | **42** | **100** |

Data science is a relatively recent discipline, to find out when data science were first established at the organisation, we asked **when was your data science team created? If there is no dedicated team, when were data scientists first recruited on a continual basis?** Our survey showed that some of the data science teams evolved from previous departments and/or developed very recently, e.g., since 2016 as shown in Table 5.

Of those that developed from previous teams they typically originated from teams which have significant overlap with data science skills (optimisation, statistics, data analysis). For example, one respondent mentioned that their analytics team existed for 15 years previously and became the Data Science team in 2018. Another mentioned a Mathematical Modelling team created in the early 2000s and then gained a specialist Data Science team from 2020. Other data science teams existed from the outset as the organisation only came into being recently e.g., 2019.

*Table 5. Count of year of founding of the data science/analytics team within the organisations.*

| **Year data science team created (to nearest year)** | **Total Responses** | **Percentage of Responses** |
|---|---|---|
| 2010 - 2011 | 2 | 6.4 |
| 2012 - 2013 | 1 | 3.2 |
| 2014 - 2015 | 1 | 3.2 |
| 2016 - 2017 | 6 | 19.4 |
| 2018 - 2019 | 13 | 42.0 |
| 2020 - 2021 | 8 | 25.8 |
| **Total** | **31** | **100** |

These results overlap with those found in the Kaggle 2021 report concerning the length of experience of the data scientists [19]. At the time of our survey, most teams had only existed for a relatively short time (less than five years), and extending this to leadership (see Section 7), we find that most data science leads/managers have also less than five years' experience in the role. This is perhaps unsurprising for a relatively recent discipline but shows that even if the sector has a shortage more generally, the situation is likely more difficult when seeking highly experienced data scientists.

The next set of questions aimed to investigate the qualifications, background and skills of the data science team. The previous Section 5 showed that the type of data the data science teams needed to handle, as well as the importance of the various skills. In other words, this showed the desired qualifications needed to implement the data science operations within the organisations. The following questions are more about what the skills/experience of the actual data scientists/analysts are. There is obviously going to be a significant overlap and it is likely that the skills needed are also driven by what skills are available, and vice versa. Gaps in the skills is explored in the recruitment part of Section 8.

To understand the academic background to the data scientists the next question asked, **Among the data scientists/analysts, what is the rough percentage of individuals with the following qualifications (highest attainment):**

- **Postdoctoral researcher**
- **PhD**
- **Masters**
- **Undergraduate**
- **Other**

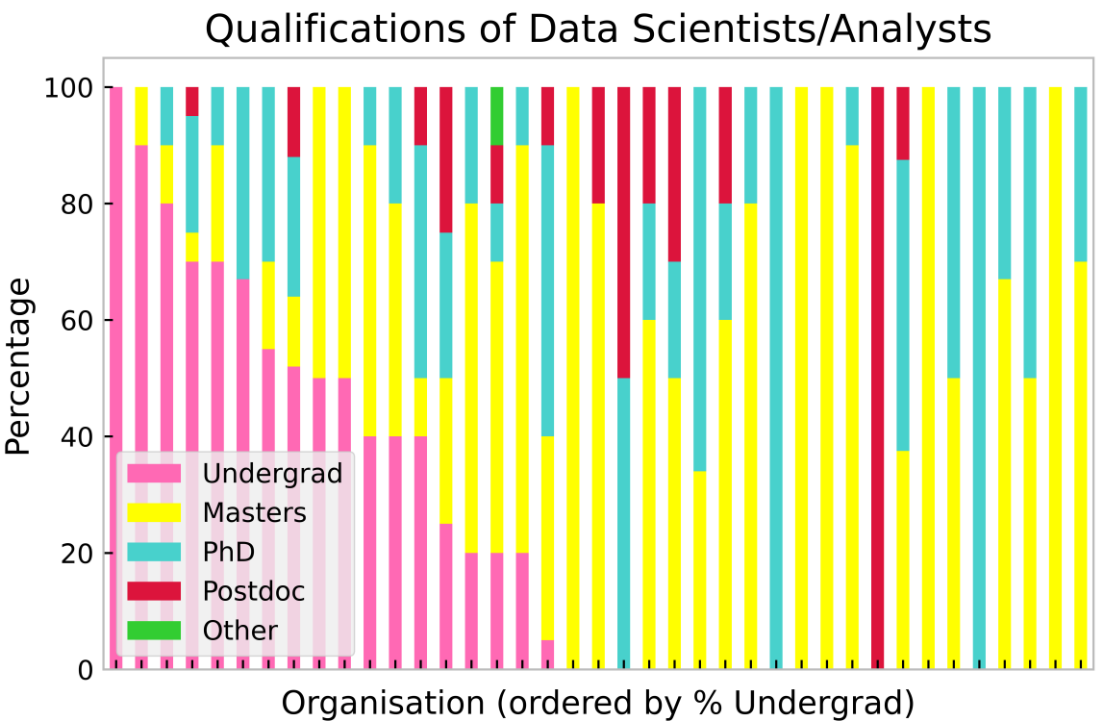



*Figure 9. Percentage of data scientists/analysts within 39 organisations with different qualifications.*

Figure 9. Percentage of data scientists/analysts within 39 organisations with different qualifications.Figure 9 shows that often there is often a mix of data scientists/analysts with undergraduates, masters and PhD qualifications. Less commonly there are also postdoctoral experience/qualifications. Over 50% (21) of organisations seem to consist purely of at least a postgraduate degree (Masters, or above) without any sole undergraduates. Four organisations only consisted of postdocs and PhDs; these tend to have small numbers of data scientists/analysts (<4). Only one organisation stated that 10% of their data scientists' highest qualification attained was 'Other'. Unfortunately, further clarification was not provided, however this could include any qualification other than a degree such as vocational, diploma etc.

**If typical data scientists/analysts have at least an undergraduate degree, what is their discipline?** Figure 10 shows the main discipline the data scientists/analysts in the organisation studied (according to the person filling out the survey) if they studied at least an undergraduate degree. They mainly fit within the more technical disciplines. Surprisingly few were involved in computer science compared to mathematics, physics and engineering. These technical disciplines will likely provide a solid foundation for the practitioners with the scientific method, meaning they have skills in analysing relationships, building models, and testing hypotheses.

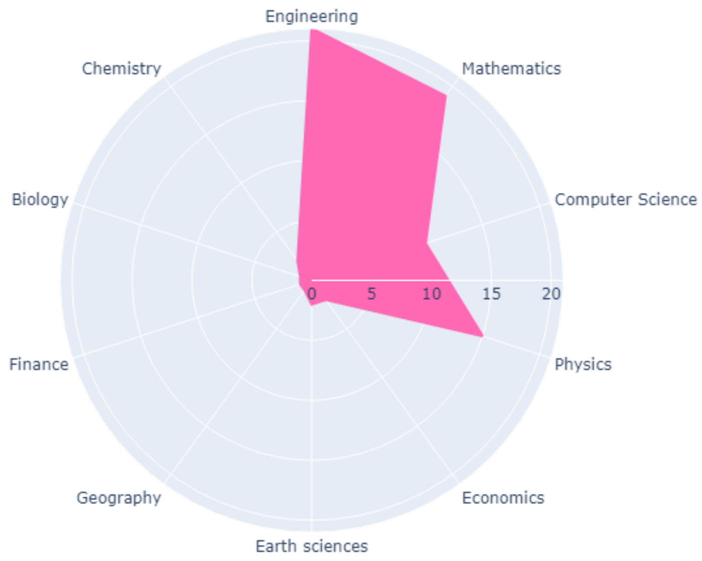



*Figure 10. Plot showing the main discipline data scientists studied if they have at least an undergraduate degree.*

**If team members do not come straight from academia, what is their experience/background?**
Figure 11 shows the background of data scientists who didn't come directly from academia. Most come from other data organisations and/or energy industry. See the Appendix in Section 10 for further detail regarding these areas. This is significant as it aligns with the outcomes of some of the recruitment-based questions in Section 8. Domain knowledge was seen as one of the most important skills for a data scientists/analyst that organisations struggle to recruit. However here it is shown that a large amount of those who do not come from academia are from the energy industry which means they likely have significant domain knowledge. This suggests that those who come from academia and "Other data organisations" may not have much energy domain knowledge.

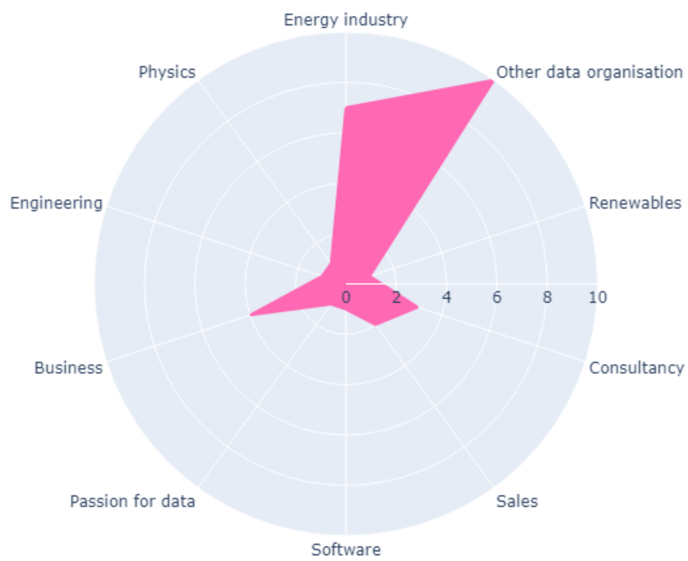



*Figure 11. Background of data scientists for those who were not employed directly from academia.*

There are very few data scientists from business, sales, consultancy or even software backgrounds. However, given the potential skills gap in data science it may be worth understanding this in more detail in future work. Enabling those from less technical backgrounds to transition to a new career by reskilling in data analysis or science could open avenues for building capacity with teams and across the sector.

Data science is a complex field which requires many technical skills from model development, data analysis, statistics, feature engineering etc. but also skills for reporting and presenting a clear narrative to non-technical colleagues and employers. The next question asked **Would you describe the data scientists/analysis in the team as 'all-rounders' having skills across the data science stack (ML and analytics, visualisation, databases, Extract Transform Load (ETL) etc.) or are the skills split across the team?**

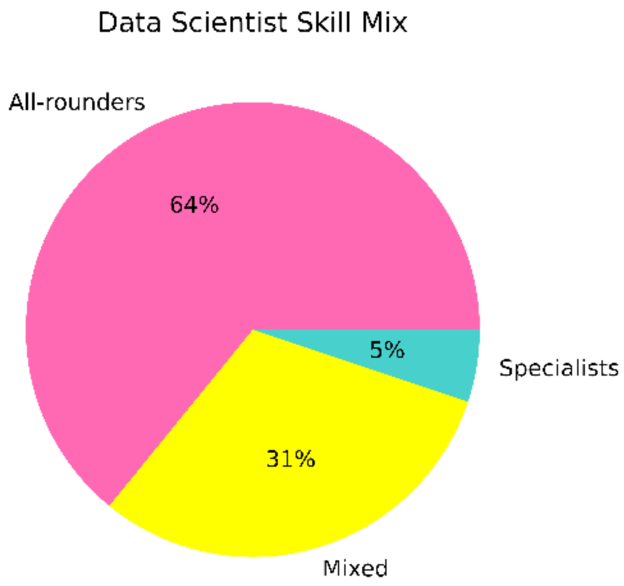



*Figure 12. Pie chart showing proportion of respondents who believe whether the majority of their data scientists/analysts are all-rounders, specialists or a mix.*

All-rounders are likely to have a decent high-level knowledge of many areas of data science (and adjacent areas in data engineering and software development) but will not perhaps understand a particular topic very deeply. In contrast, specialists will have expert knowledge on a particular area of data science, e.g., this could be the general topic of forecasting, or it could be of a particular model type such as deep neural networks or generalised additive models. The results are presented in Figure 12 and show that almost two thirds of organisations employ all-rounders. Only one organisation (which accounts for about 5% of the responses) considered the team as only specialists but there was no further elaboration. However, about a third of the organisations have a mix of all-rounders and specialists. Of those that elaborated further, it appears that the specialism often focused on either modelling or data engineering-type skills. The former focused on science, analysis, visualisations and data models. The latter focused ETL, databases, and the deployment stack. This perhaps suggests that these organisations are interested in tackling many diverse problems rather than producing an over-engineered or absolutely optimal solution.

**Summary:** Data scientists within organisations tend to be relatively small and all concentrated within a team or a couple of teams. These teams also typically developed in the last few years, although some emerge from technically adjacent disciplines.

Data scientists/analysts typical require technically skilled individuals with most having at least an undergraduate degree in a technical subject such as engineering or mathematical sciences. Many others have postgraduate degrees. For those who did not join straight from academia, they came from other energy industries or other data-driven organisations. This perhaps highlights the desire for energy domain knowledge in addition to data skills.

# 7. LEADERSHIP

The aim of this section is to understand the leadership structure of data science teams. For this reason, the focus is on those organisations which stated that they had dedicated data science/analytics teams. However, the analysis will also include the opinions of all respondents on what makes good leaders and leadership. Recall from Section 4 that 17 of the respondents led the data science team and hence many the respondents had direct experience in this topic.

The first question focused on whether, of those with dedicated teams, **Who leads the data science team within the organisation?** The aim was to see if they are data scientists themselves or come from other roles or disciplines.

*Table 6. Count and percentage of responses answering: "Who leads the data science team within the organisation?".*

| Role | Number of Responses | Percentage of Responses |
|---|---|---|
| The lead data scientist/analyst | 23 | 54.8 |
| Engineer | 1 | 2.38 |
| Director/Head | 4 | 9.5 |
| Manager | 5 | 11.9 |
| Other | 1 | 2.38 |
| No dedicated Team | 8 | 19.0 |
| **Total** | **42** | **100** |

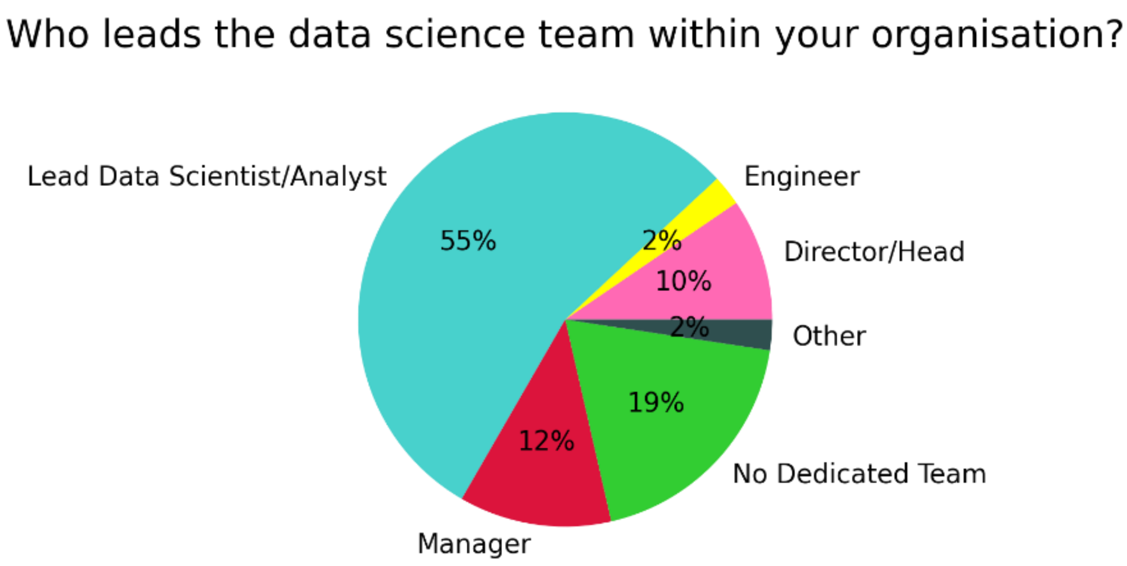



*Figure 13. Pie chart showing the proportion of those leading the data science team within each role.*

The answers are shown in Table 6 and the pie chart in Figure 13. The majority of those leading are lead data scientists/analysts themselves. Suggesting a high technical skill is required. However, if we group the Director, Heads and Managers these make up around 22%, almost a quarter, of the organisations. If the organisations with no dedicated teams are ignored, this means two-thirds of the teams are headed by a technical lead, and a quarter by managerial roles (NB: that assumption is based on the title, which of course may not be entirely accurate).

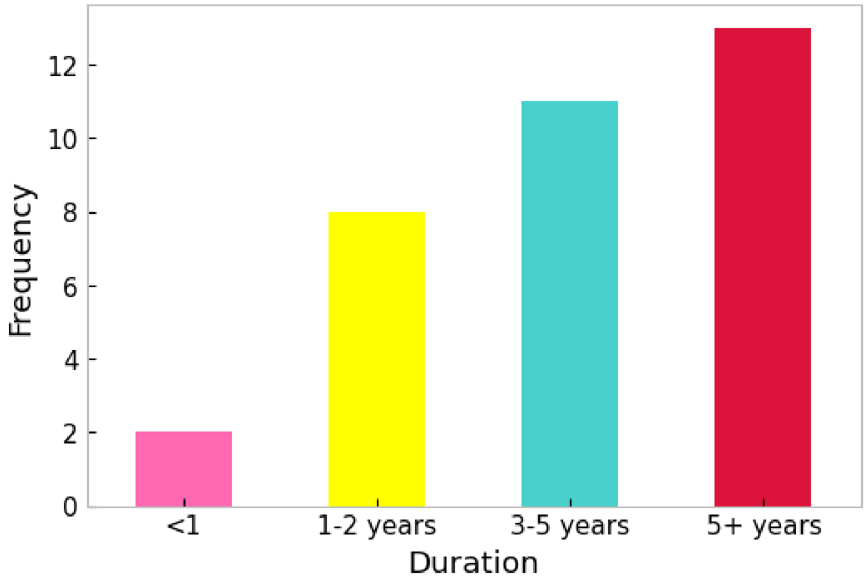



*Figure 14. Histogram of how long data science leads/managers have been in this type of role in total.*

The next question asked **if there is a lead/manager, how long have they/you been in this type of role (not just current role)?** The results are shown in Figure 14. Of the 34 respondents, most leads/managers (21) had been in this type of role less than 5 years. This ties in with the increase in data science teams forming from 2016 onwards. This also suggests that there may be a shortage of data scientists with long term managerial experience. This is unsurprising given the relative recency of data science as a discipline in general, and in the energy sector. It may also suggest it may be desirable to help develop the current data scientists to help fulfil the needs for talented data science leads and managers.

The next question asks **If there is a lead/manager of the data science team, what level/type of background/experience do they/you have?** The answers depended on how the question was interpreted to mean qualifications/experience and whether technical or soft skills were highlighted.

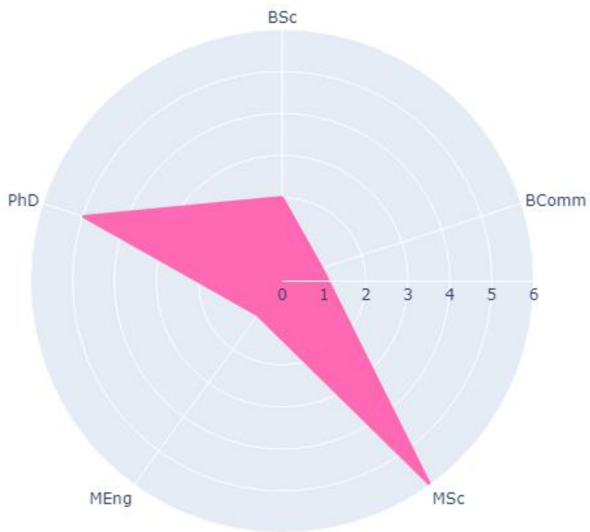



*Figure 15. Spider plot showing the academic qualifications for the lead/manager of the data science team.*

The responses were roughly split into two main categories, those relating to formal qualifications or previous experience in other roles. Of the 42 respondents, 15 included details about the lead/manager's academic background. The minimum qualification cited was an undergraduate degree, the majority held a postgraduate degree or doctorate as per Figure 15. This aligned with the results about the data science/analyst teams in Section 6 which showed that many of them had at least an undergraduate degree. The skew for leads was towards postgraduate degrees such as PhDs and Masters degrees.

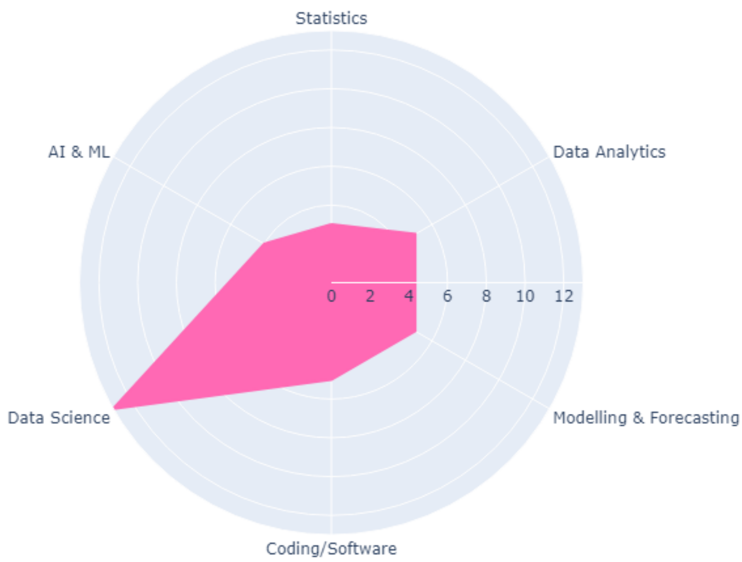



*Figure 16. Spider plot of the various technical subject areas associated to the data science lead/manager.*

Of the 35 respondents that described the lead/manager's experience in terms of technical skills, general data science was most popular as opposed to specific skill areas shown in Figure 16 which could suggest all-round skills, much like was preferred for the data science teams themselves (Section 6).

Of those that described the lead's background through previous roles, the answers are shown in grouped categories in Figure 17. Interestingly 35 respondents did not describe the journey of leads/managers stemming from the energy systems/utilities sectors. Experience was concentrated in the engineering/business sectors with a spike of general vocational 'on-the-job' experience (as opposed to a dedicated role in another discipline) suggesting that academic qualifications may not be strict pre-requisites and transferable skills might be more important in a leader.

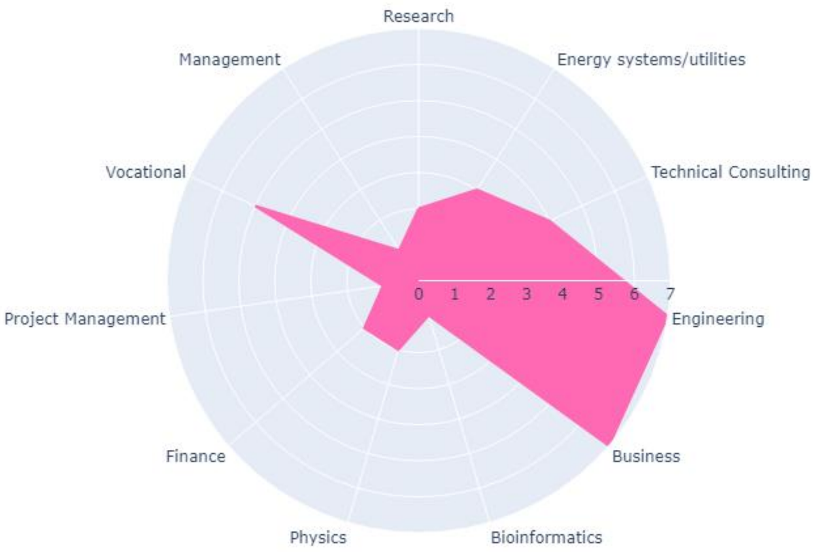



*Figure 17. Spider plot of the various areas experience has been gained associated by the lead data scientist.*

The above answers suggest that there is a mix of preferences between technical and managerial skills. To be more explicit about the required skills the respondents were asked **Generally speaking, is it important for the lead to have a high-level of technical skill?** As shown in Figure 18, the majority of the 42 respondents said it *was* important for the lead to have a high-level of technical skill, although nearly a fifth said it wasn't. This may tally with the 14% of respondents who were managers and may not be technically skilled. Those that replied 'Other' are shown below as the answer was partially yes and partially no.

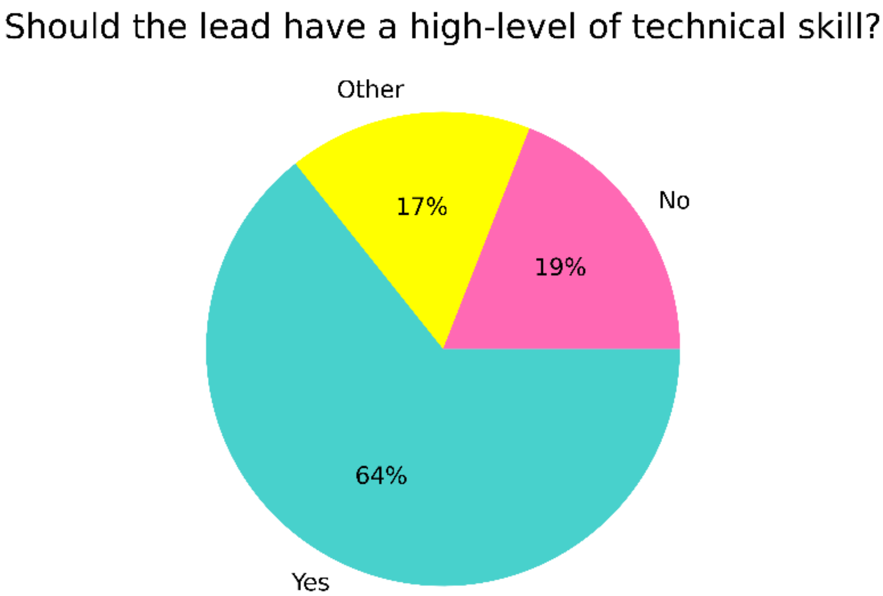



*Figure 18. Pie chart showing the proportion of each answer to whether the lead of the data science team should have a high level of skill.*

Response given as 'Other'

- Team leader does not need to have the best technical skill in that department.
- In our case, the manager is not the data science lead and I like it that way. It can be two persons – one who is empathetic and one who knows the technical stuff.
- Depends, but they do at least need to have a basic understanding of the data science process and accept it can be useful.
- Yes, for a Data Science team, less so for a Risk team where subject matter expertise can compensate a lesser tech knowledge.
- Not a necessity but helps in coaching and mentoring those new to business.
- Yes, to a degree but not at the expense of the 'bigger picture'.
- Need to have enough skill to be able to understand the data scientists and be respected by them.

*Table 7. List of advantageous and disadvantageous characteristics for a data science lead according to survey respondents.*

| **Advantageous characteristics** | **Disadvantageous characteristics** |
|---|---|
| Technical/analytical skills | Lack of technical skills |
| Domain knowledge | Lack of domain knowledge |
| Communication skills | Poor communication skills |
| Collaborative | Not collaborative |
| Empathy | |
| Responsible | |
| Nurturing | |
| Time management | |
| Stakeholder management | |
| Business acumen | Lack of business knowledge |
| Leadership | Micromanagement |
| Project management | Over-promise |
| Organised | |
| People management | Unable to delegate/overcomplicate |
| Intelligent | |
| Curious | |
| Patient | |
| Decisive | Weak to execute |
| Self-learner | |
| Open-minded | Narrow-minded |
| Motivational | Lack of initiative |
| Strategist | |
| Generalist | |
| Visionary | Short-term focus |
| Influencer | |
| Passionate | No ambition |
| Trouble-shooter | |
| Confidence | Arrogant/headstrong/defensive |
| Experience | Lack of experience |
| | Easily distracted |

These seem to be more balanced responses as we'll see in the next question regarding advantageous/disadvantageous characteristics. A good leader will need a variety of skills which include 'soft' skills as technical skills alone would seem insufficient. From the responses, technical skill isn't mandatory but does help. What is more important, is how the team functions under the lead as technical expertise may be covered by other team members.

The final question asked**, in your opinion, what are some of the most advantageous/disadvantageous characteristics for a lead to have?** To simplify the analysis to this question the answers were grouped into advantageous and disadvantageous characteristics in Table 7. Where possible, dichotomous characteristics have been paired to show which are considered advantageous/disadvantages. None of the characteristics were ever cited as being both advantages and disadvantages.

Interestingly, the only unmatched disadvantageous characteristic was 'easily distracted' as the expected opposite would be 'focused' or similar. However, there were numerous advantageous characteristics that stood alone, 'generalist' could be perceived as a disadvantage but wasn't specifically mentioned as such.

Data scientist specific attributes are relatively sparse. There is a mention of technical skills as being important, but by far management, people skills and communication skills are much more prominent. This highlights almost the opposite to the previous question where almost two thirds thought that technical skill was important. However, this would be misleading because in fact many of the respondents mentioned technical analytical skills. Hence technical skill is important to most respondents but there just aren't many diverse ways of representing this information in the answers.

**Summary:** Most leads and managers of the data science/analyst teams are data scientists themselves with postgraduate degrees, suggesting technical skills are important to leading some of these teams. This is supported by other responses where two-thirds suggest that the data science lead/manager should have high technical skills. However, this is not all that is important, many suggest that although technical skills are important so are managerial, communication and people skills.

# 8. RECRUITMENT AND DEVELOPMENT

The previous sections have shown that a range of skills are highly desirable for data scientists/analysts working with the energy sector. However, very little is known about how difficult it is to recruit talent with these skills. In addition, data science is a rapidly progressing area and new technologies and techniques are constantly being developed. This section considers responses related to recruiting data scientist/analysts, but also focus on upskilling and developing within the organisations themselves.

## 8.1. TEAM DEVELOPMENT

Data science can move quickly, for this reason we asked **How do you ensure the continuous development of the skills within the data science team?** There were 24 individual responses to this.

Everyone who responded is interested in this but there is a wide variety of success. Three respondents are trying to but are struggling to solve it, another stated it is an active area of discussion. Hence it is clearly seen as quite important, but it doesn't appear to be an easy thing to incorporate into an operational business. Some upskilling/training is only implemented when it is required for new projects and incorporated during task sprints. Trying to incorporate training into new projects isn't always successful though, and high workloads limit the opportunities for skills development especially in areas where the individuals are "not very knowledgeable".

A whole host of different ways to train the team were put forward:

- Hackathons
- Sharing articles
- Reading papers on recent research
- Webinars
- Regulatory white papers
- Internal Knowledge sharing
- E-learning/online courses
- Attending workshops
- Work on toy problems
- Peer review of Code
- Analysis of current market tools

Most of the responses suggest training is relatively informal but supported through the organisations. Individuals are encouraged to share articles and reading around their topic, and in one case the participant mentioned they have a "budget for courses, books and workshops". Many individuals in the team are encouraged to pursue their passions and, in some cases, have total freedom to explore their own ideas and projects. This correlates with some of the desired characteristics in the recruitment Section 8.2, which highlight motivated and passionate individuals.

More formal upskilling is through hackathons, courses and workshops, and some institutions have specific courses for new starters. Typically, there is some regular meet ups where ideas are shared or through reading clubs, but many find it difficult to find the resources. One response highlighted a very focused implementation on upskilling via a "data skills squad" which involved the data science team and other departments. They also had a skills audit, upskilling prioritisation, and a mixed upskilling program leveraging existing internal knowledge, external providers, partnerships with suppliers and an Apprenticeship Levy. Two participants also mentioned working with universities to facilitate upskilling. One used the relationship to maintain up-to-date state-of-the-art knowledge of the latest development, another used universities for MSc/PhD part-time learning. Academic-Industry collaborations to bridge some of the skills gap was also highlighted in our recent report [20].

Since conferences are an important way to aid development, we asked **Do the data analysts/scientists regularly participate or attend external conferences and events? If yes, what type of events?** The responses from 29 participants are given in Table 8 and shows a relatively even split between them, although slightly more seem to attend conferences.

*Table 8. Count of organisations whose data analysts/scientists attend external conferences and events.*

| Response | Total Responses | Percentage of Responses |
|---|---|---|
| Yes | 17 | 58.6 |
| No | 12 | 41.4 |

Of the type of events attended the following were mentioned:

- ESC webinars
- Industry conferences
- Technical conferences
- Energy conferences
- Data science/analytics conferences
- Energy system/storage/flex focused rather than general data science
- Data science and Python conferences
- Regular international conferences
- Academic conferences
- Technology solution themed conferences
- AWS events
- Big Data week
- Data science workshops e.g., Royal Statistical Society
- Specific Major conferences such as: CIRED, Enlit, Energy Networks Innovation Conference, ODSC, Gartner, SAS Global Forum, Google Cloud Next
- Informal e.g., local ad hoc meetups

Of those that said no, some said this was due to Covid restrictions rather than any other reason and the desire to attend events was otherwise there.

Finally, we asked **Do the team regularly publish results?** Although this is obviously not a priority for most businesses, publishing results has several potential development opportunities. Firstly, it increases the documentation of methods and techniques employed within the organisation. This makes it easier to newcomers, team members and recruits to get quickly up to speed with the implementations and potentially learn new techniques for something they themselves are not developing directly. Secondly, it helps develop communication skills for the data scientists. As will be shown in the following section, team working, and communication skills are viewed as important in addition to the technical ability.

From the results shown in Figure 19, it appears that most of the teams do in fact publish. Most only publish internally as expected however, several teams also publish externally with a small number publishing in peer-reviewed (PR) journals.

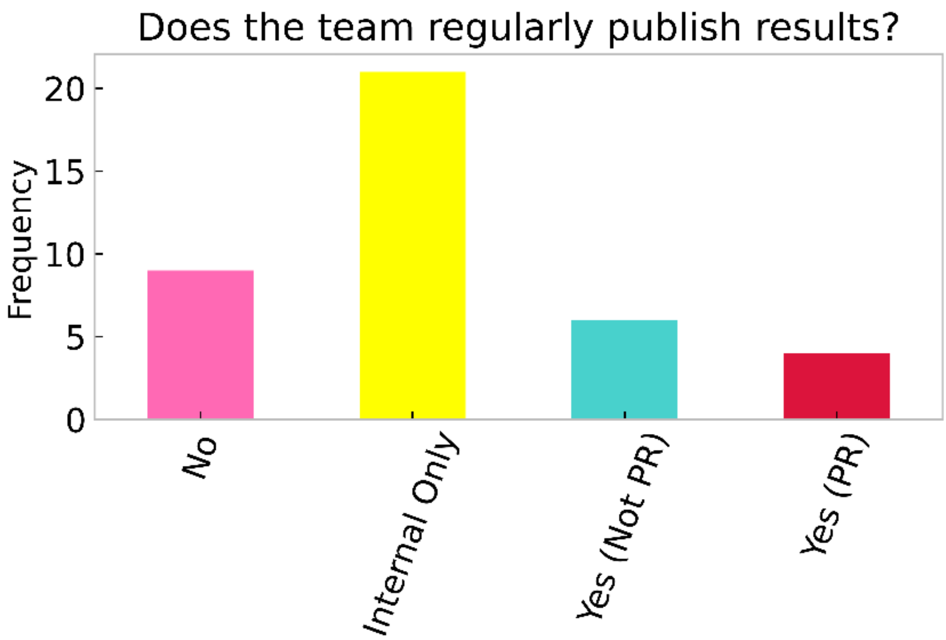



*Figure 19. Responses for whether the data science/analysts team publish their results. The choices are no, internal only, yes but not peer reviewed (PR), yes peer reviewed.*

## 8.2. RECRUITMENT

This section focuses on recruitment and upskilling. We first asked whether the participant answering the survey was **part of the recruitment process for the data science/analytics team?**

*Table 9. Answers to whether the participant in the survey was involved in the recruitment process.*

| Answer | Number of Responses | Percentage |
|---|---|---|
| Yes | 28 | 66.7 |
| No | 14 | 33.3 |
| **Total** | **42** | **100%** |

The answers in Table 9 show that two thirds were involved. This ensures that the answers in the rest of this section are well-informed to the recruitment process.

Several questions related to the difficulty in recruiting the necessary skills for data science. This has become an important topic of late beyond energy, as many sectors are struggling to recruit the necessary talent. We first asked participants **What are the most important characteristics you look for in the individuals you recruit?** The answers can be broken down roughly into four main areas:

- Technical skills
- Team working skills
- Specific knowledge and,
- Personality traits.

Being a quick learner and having critical thinking skills was one of the most prominent characteristics answered as part of this question. This did not necessarily mean an advanced degree was needed. In fact, one participant stated explicitly "No need for PhD or Master, just need to see their ability to tackle all type of problems in a smart way." Unsurprisingly good coding skills were a common desired technical characteristic as was an ability to communicate. Other more technical skill answers included modelling experience, data wrangling skills, being "python savvy", and an "ability to implement stat models".

In addition to technical knowledge the respondents wished for their data science and analytics teams to have quite broad capabilities (aligning with the desired abilities highlighted in Section 6). A major desire was for recruits to have a wider understanding of the domain, including business comprehension. As noted by one participant, "having a commercial view of the problem is helpful" this may help remove time consuming rabbit holes. An interdisciplinary background was also stated as a desired characteristic.

Common across the board is that the person should be "easy to work with" and being a team player. Many data scientists do not work in isolation and hence a cooperative nature and ability to communicate both within the data science team and externally to clients is essential.

Other personality traits commonly stated was about being a motivated and passionate individual who is always keen to learn and is very curious about the area and the problems they are solving. Motivation was a common term used for this question. Other less common responses included ensuring they have an interest in the energy domain and enjoying coding. Although intellect and innovative thinking was mentioned, it wasn't a prominent answer, and one participant explicitly mentions that "intelligence is overrated". Suggesting that it is not about being hyper-intelligent but more about having a positive and enthusiastic personality.


*Figure 20. Word cloud showing most common words in the response to what the most important characteristics are in recruits.*

The main words collated from the responses are shown in the word cloud in Figure 20. These highlight the same topics identified by manual investigation of the responses including communication, domain knowledge, technical skills etc.

Given that the above skills and characteristics were desired the next question we asked was **"How easy has it been to recruit the required skills in the area of data science/analytics for your company? (1 being very easy, 5 being very difficult)".** The results are shown in Figure 21. All but two of the 39 participants answered this question. As the results show, the experience varies widely but 15 of the 39 (almost 40%) found it difficult or very difficult to hire the skills they desired. It does appear that one person incorrectly answered a Very Easy rather than Very difficult as the later question indicated they had difficulty recruiting in all areas. We left in the above as we cannot confirm.


*Figure 21. Response counts for the question about difficulty in recruiting required data science skills within their organisation.*

The areas that organisations struggled with were investigated in more detail via follow up questions. We asked **What specific areas/skills have you had the most difficulty in recruiting suitable individuals?** And **can you give possible reasons you think why it is difficult to recruit in these areas?** We tie the responses to these questions together as they are obviously linked.

Starting with those who do not have any difficulties recruiting (e.g. answered OK, good, or very good for the previous question). Since these organisations don't have problems with recruiting data scientists, their responses focus on three main areas of difficulty:

- Getting candidates with sufficient domain knowledge
- Software development skills
- Significant seniority and advanced skills in data science

Although domain knowledge was seen to be one of the most important characteristics in a data scientist, it seems it is a sparse property within the pool of candidates that these organisations identify. One option is to train subject matter experts in AI but one participant noted that this is a steep learning curve and many do not have the time or inclination to learn. Further, domain experts are likely valuable to the department/team they are already a member of which is a further potential blocker to reskilling such candidates.

Software development skills, including data and software engineering were one of the most common skills that the participants organisations struggled to recruit. Operations is a main focus of the organisations and hence it is important to turn the data science algorithms into a useful and reliable business product. One participant mentioned they have higher demand for these skills since they were relatively young. High demand is one reason why organisations struggle to hire talent in software development, another was that the focus of data scientists was often analytical skills, and it was difficult to find all-rounders who had domain knowledge, analytical skills and coded with best software practices.

It seems that these organisations had little problem recruiting data scientists in general, but it was difficult to find good quality candidates with "both significant modelling and programming skills", as one participant mentioned. There are not many candidates with clear significant experience and capabilities. Another participant suggested this was because data analytics/science are not taught extensively at university. 'Senior Data Scientist' was mentioned as one position that was difficult to fill because of the competitive market.

A few other areas were also mentioned by the participants:

- Difficult to attract graduates who want to work on complex physical problems, and instead want to work in web related topics.
- The energy sector must compete with digital native companies who can offer improved salary and location.

In other words, competition with other sectors could also increase the skills gap within the energy sector.

Now turning to those who noted they found it difficult to recruit (scored 4 or 5 on the previous question). Domain knowledge and software/data engineering skills were again heavily mentioned topics which one participant identifying difficulties finding individuals who can productionise their code. Some organisations can find data scientists but with limited domain knowledge and vice-versa, but they can't find one with both. The reason given is that they believe energy industry lags behind many of the other digital cutting-edge sectors. The stronger candidates who possess multiple skills seem to focus on these other finance or tech-based organisation which often offer better pay. This was echoed by another participant who mentioned that the energy sector was at an earlier place than banks/telecoms etc. and therefore doesn't' appeal to everyone. The energy sector isn't currently dealing with massive data sets or necessarily need very sophisticated AI/machine learning.

Despite many candidates having good skills on paper, this group, who struggled to recruit, pointed out many more technical weaknesses in the candidates they recruited including:

- Poor coding skills
- Lack of deep understanding of methods,
- Being unable to think probabilistically
- Not having basic data cleaning and visualisation skills

Further, one participant mentioned it was difficult to find someone who can perform the entire data science pipeline from data analysis to model building, productionising, testing, and evaluation.

Although recent graduates could be a pool of technical individuals which could be used to build capacity, one participant mentioned bad experiences with graduates from data science courses and that academics often lack confidence.

The prominent desirable softer skills mentioned in the previous question didn't appear in the list of skills organisations struggled to recruit, these included team players, motivated individuals, and being a quick learner. One participant did mention team working as a specific area of difficulty. For them it was difficult finding candidates who were able to extract information or requirements from colleagues.

Due to some of the difficulties in recruiting skilled data scientists, an attractive alternative is to upskill from other teams internally. Hence, we also asked **Have you upskilled within the organisation to recruit data scientists as well as externally advertised? How do you compare these two approaches?**

- Yes, internally upskilled only - 0

- Both - 13
- No, external only – 8

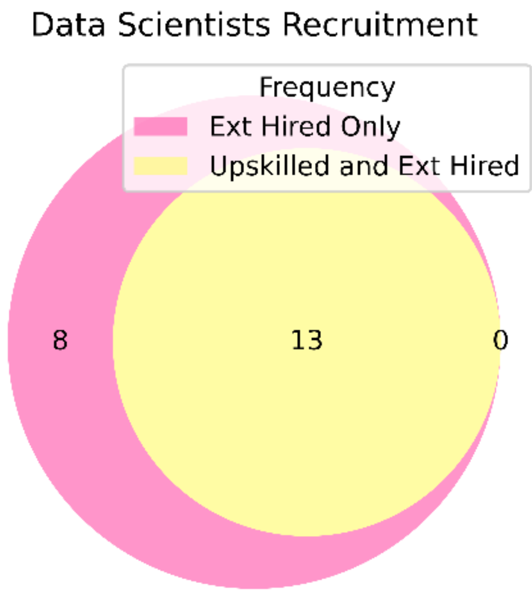



*Figure 22. Upskilling and external recruitment counts within organisation participating in the survey.*

The breakdown of the results is shown in Figure 22. None of the respondents recruited exclusively from within, but if the whole organisation is already data-focused, new staff must be externally recruited.

Internal recruitment is only possible if an organisation is sufficiently large, otherwise it just shifts the recruitment need. Some organisations try both approaches as upskilling is more successful due to in-house expertise and building upon existing domain knowledge. However, external candidates bring new ideas and fresh perspectives.

Some skillsets may be stronger depending on the background of external candidates such as formal technical training. Upskilling internal recruits may mean teaching fundamentals usually taken for granted. It can take longer to upskill from within but may be more rewarding (teaching/receiving training) and provide more job satisfaction.

One approach used is to recruit externally for entry level roles but upskill an existing employee for a senior role as this is more efficient. A combination of internal and external recruitment depending on the role requirement seems to be the general strategy. Recruiting from within is also constrained because current employees already occupy a niche and may struggle to upskill alongside their current commitments, or their rate of learning is below expectations which favours hiring external candidates. One respondent mentioned they were largely self-taught and learnt python programming on their own using several courses in Udemy and Coursera to gain ML skills.

We finished the survey with any other thoughts from the participants via the question: **Please add any other thoughts or comments with regard to data analytics/science skills applied within your organisation.**

The responses highlight the idea that the amount and experience of data science skills in the energy sector is still insufficient to meet the large requirements. Many data-science dependent companies are still without a focused data science team. The pace of development is also difficult to match and facilitate, and retaining experienced people is difficult since turnover is one of the only ways for individuals to progress.

The benefits of AI/machine-learning are still yet to be fully realised and hence those data scientist skills are not being used to their fullest yet. There is a suggestion that energy network companies need to start making more progress with incorporating more advanced analytics. This may in turn encourage more of the data science talent to engage with the energy sector. However, there is a caution that these problems may always be relatively small and not "big data" challenges.

University collaborations are a potential resource for skilled individuals but there are roadblocks for some organisations. One participant mentioned trying to collaborate with academics but have been unable due to IP concerns and limitations of what can be shared.

**Summary:** There is a large need for data science skills within the energy sector. However, the responses to our questionnaire have shown that organisations are struggling to train, recruit and upskill the talent to meet the demand. Although some advanced data science expertise is one area of difficulty it is some of the wider skills which organisation are finding difficult to recruit. Domain knowledge and coding skills are two of the most prominent weaknesses and are hindering efforts to produce operational implementation of the algorithms data scientists develop.

# 9. DISCUSSION

This survey is one of the first looking at the intersection of data science and the energy sector. The increased digitalisation of the energy sector means there is a requirement to understand the growing need for data science skills and personnel. This report is an attempt to start to understand the gaps and requirements as well as the current landscape.

Some of the results in this survey overlap with those included in our literature review. Notably that many of the participants mentioned they struggled to recruit the talent that they needed and both hard and soft skills are in demand.

Significantly given our sectoral focus, the survey participants highlighted that the energy sector may struggle to attract the best talent due to financial competitiveness from other sectors such as FinTech.

Our survey results above also show that students leaving academia to join data science teams are missing some vital skills suggesting there is a lack of guidance on the tools and techniques used in industry. In our survey there was frustration that certain skills, such as proper coding skills, are not adequately taught or trained, even within data science courses. However, in terms of hard versus soft skills, the responses were more evenly distributed in our surveys compared to the surveys in our literature review. Although technical skills were certainly valued, domain knowledge and other soft skills such as communication and team working were also highlighted.

Utilising university leavers to help fill the skills gap was something highlighted by those interviewed as part of our "From Academia to Industry" reports [21]. However, further investigation into specific data science master courses in that report highlighted, at least in the UK, that there was very little focus on coding skills.

There is, however, so much more which needs to be understood and tackled. Data science in the energy sector is relatively new, especially compared to social media organisations, FinTech and many other sectors. The energy sector is also at a relative disadvantage to many of these areas, with traditionally less data available, and less financial reward. However, the energy sector is now one of the most exciting and important places to work. Due to the need for countries around the world to move to a net zero energy system utilising novel services and innovations.

There is an unwarranted perception that there are not interesting data science problems to be tackled in the energy sector. This is clearly not true as evidenced by the type of work presented in this report. Many of the organisations are looking at a variety of complex data sets and the interplay between energy vectors, services and dynamic requirements provide plenty of challenges in forecasting, customer insights, control, optimisation and image processing, amongst others. Organisations themselves are partly responsible for some of the malaise as they have been slow to promote the interesting problems they are tackling and have not traditionally shared the extremely valuable data that they possess.

There are many next steps that could be taken based on this research. However, the following are some of the most pertinent areas:

1. **Enabling training to future data scientists**. Prospective data scientists need to have the resources and training available to them to ensure they can be immediately valuable to the energy sector. As seen in the recruitment section this also means ensuring that domain knowledge and communications skills are also a part of this, not just the highly technical data science skills.
2. **Upskilling**: many organisations don't have the time to keep the team up to date with the latest methods and technologies. Resources must be more accessible so that the cutting-edge research

and development in the areas of data science can be identified and used across the sector. This includes making the outputs from academic research openly accessible but also ensuring that the research is shared in easily digestible ways to save time and resources.

3. **Reskilling:** as we go through the transition of decarbonisation towards Net Zero this will include adapting the current workforce to suit the demands of low carbon and emerging technologies such as hydrogen and carbon capture. Some skills will be directly transferable, others will require more support especially if digital literacy itself an issue or staff find themselves in new environments working alongside disciplines that are new to them, such as data science.
4. **Building teams**: Recruiting and upskilling are irrelevant if the frameworks and infrastructure are not in place for the data scientists to flourish and optimise their outputs. This includes understanding the skills needed within the organisation, that the leadership is chosen appropriately, and that the team has the right mix of personnel.

Companies that are leaders in their industry are always looking ahead and don't have difficulties recruiting data scientists. Organisations should prioritise attracting and retaining the best talent to remain competitive in a seller's market. This means dedicating time and resource developing their recruitment strategy. Attractions include exciting data problems to be explored along with the necessary data to facilitate innovation. The power of data science is through storytelling. If an organisation has a clear vision and can communicate compelling stories from previous implementation of data science, this will encourage staff to join them on their voyage of discovery.

# 10. APPENDIX

## 10.1. ADDITIONAL INFORMATION

Figure 11. Background of data scientists for those who did not come straight from academia, respondents acquired their experience from the following areas:

- **Energy industry** – such as utility companies, electricity grid, domain expertise
- **Other data organisation** – non-energy related organisation such as finance, private sector data management
- **Renewables** – exclusively and no other industries
- **Consultancy** – general analytics in a consultancy environment
- **Sales** - unspecified
- **Software** – such as Google, coding
- **Passion for data** – led individuals to upskill
- **Business** – optimisation, stats/analytics
- **Engineering** – energy, electrical or not-specified
- **Physics** – astrophysics, NASA

## 10.2. ACRONYMS

| | | |
|---|---|---|
| **AI** | - | Artificial Intelligence |
| **BMRS** | - | Balancing Mechanism Reporting Service |
| **CIRED** | - | Centre International de Recherche sur l'Environnement et le Développement |
| **CSV** | - | Comma-separated values |
| **DNOs** | - | Distribution Network Operators |
| **ENIC** | - | European Network of Information Centres |
| **ESC** | - | Energy Systems Catapult |
| **ETL** | - | Extract Transform Load |
| **EV** | - | Electric Vehicle |
| **GB** | - | Gigabyte |
| **GIS** | - | Geographic Information System |
| **IoT** | - | Internet of Things |
| **kB** | - | Kilobyte |
| **MB** | - | Megabyte |
| **ML** | - | Machine Learning |
| **MLOps** | - | Machine Learning Operations |
| **NGESO** | - | National Grid Electricity System Operator |
| **NLP** | - | Natural Language Processing |
| **ODSC** | - | Open Data Science Conference |
| **PB** | - | Petabyte |
| **PV** | - | Photovoltaic |
| **SME** | - | Small or Medium sized Enterprise |
| **TB** | - | Terabyte |
| **WHD** | - | Warm Home Discount |

# 11. BIBLIOGRAPHY

[1] Energy Jobline, "Global Energy Talent Index," 2022. [Online]. Available: https://www.getireport.com/.

[2] Dept for Digital, Culture, Media & Sport, "National Data Strategy," 2020. [Online]. Available: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy#data-1-2.

[3] Royal Society, May 2019. [Online]. Available: https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills-report.pdf?la=en-GB&hash=212DAE7D599B0A48687B372C90DC3FEA.

[4] Open Access Government, "Maintaining momentum on digital investment is key to dealing with the skills shortage," [Online]. Available: https://www.openaccessgovernment.org/digital-investment-is-key-to-dealing-with-the-skills-shortage/146247/.

[5] Virgin Media O2, 18 October 2022. [Online]. Available: https://news.virginmediao2.co.uk/get-online-week-digitals-skills-shortage-costs-the-uk-economy-and-workers-12-8-billion-as-brits-grapple-with-higher-bills-during-the-cost-of-living-crisis/.

[6] Nesta, "Analytic Britain: Securing the right skills for the data-driven economy," 2015.

[7] Quant Hub, "The Data Scientist Shortage in 2020," [Online]. Available: https://quanthub.com/data-scientist-shortage-2020.

[8] Dept for Digital, Culture, Media & Sport, "Quantifying the UK Data Skills Gap," 18 May 2021. [Online]. Available: https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report.

[9] Kaggle, "2022 Kaggle Machine Learning & Data Science Survey," [Online]. Available: https://www.kaggle.com/competitions/kaggle-survey-2022.

[10] The Energy Institute, "The net zero skills issue: Bridging the gap to a low carbon workforce," 2021. [Online]. Available: https://www.energyinst.org/barometer/2021.

[11] Robert Gordon University, "The Impact of Digitalisation on Data Professionals," 2021.

[12] Energy Institute, "https://idric.org/wp-content/uploads/9.4_Hydrogen-Skills-Gap-Study-1.pdf," [Online].

[13] H. Kazmi, Í. Munné-Collado, K. Tidriri, L. Nordström, F. Gielen and J. Driesen, "Data Science and Energy: Some Lessons From Europe on Higher Education Course Design and Delivery," *Harvard Data Science Review,* vol. 4, no. 1, 2022.

[14] Hewlett Packard Enterprise, "Closing the data science talent gap," [Online]. Available: https://www.hpe.com/us/en/insights/articles/closing-the-data-science-talent-gap-2203.html.

[15] J. Gauthier, A. Vincent, S. Charette and N. Derome, "A brief history of bioinformatics," *Brief Bioinform,* vol. 20, no. 6, 2019.

[16] Government, “Energy Data Taskforce,” 2019. [Online]. Available: https://www.gov.uk/government/groups/energy-data-taskforce.

[17] Business 20 (B20), “Digital Transformation: Taskforce,” [Online]. Available: https://www.b20italy2021.org/taskforce/digital-transformation/.

[18] Power Swarm, [Online]. Available: https://powerswarm.co.uk/.

[19] Kaggle, “State of Data Science and Machine Learning 2021,” [Online]. Available: https://www.kaggle.com/kaggle-survey-2021.

[20] S. Haben and S. Hinton, “Data Science: From Academia to Industry, Making Impact in the Energy Sector,” Energy Systems Catapult, 2022.

[21] Energy Systems Catapult and Arenko, “Data Science: From Academia to Industry,” https://es.catapult.org.uk/report/data-science-from-academia-to-industry/, 2022.

[22] Airswift, [Online]. Available: https://www.airswift.com/blog/skills-green-workforce-of-the-future.

[23] Energy Institute, “https://idric.org/wp-content/uploads/9.4_Hydrogen-Skills-Gap-Study-1.pdf,” [Online].

**Energy Systems Catapult**
7th Floor, Cannon House
18 Priory Queensway
Birmingham
B4 6BS

es.catapult.org.uk
livinglab@es.catapult.org.uk