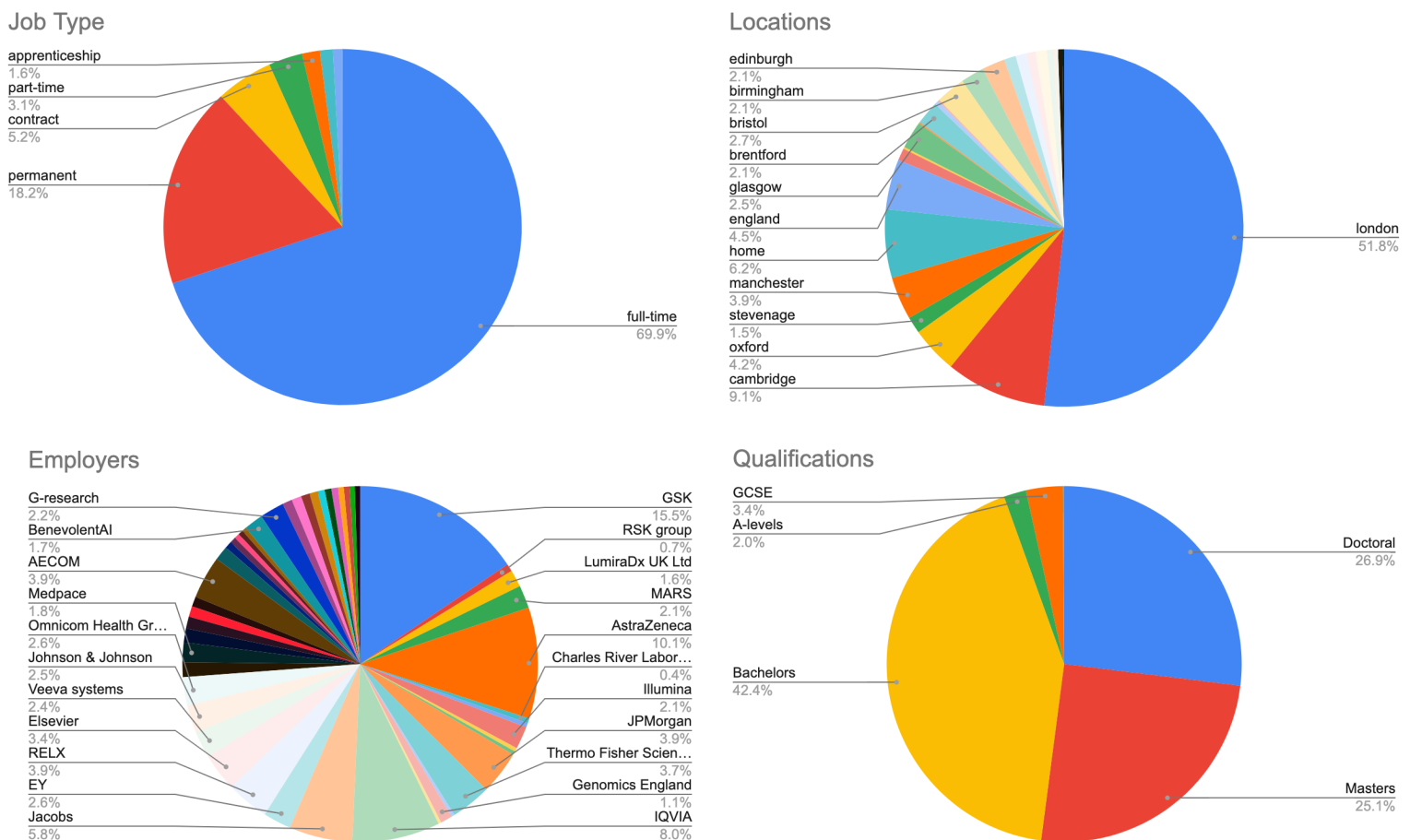


Health data science careers event: sharpening objectives

Section 1 - Identifying suitable career pathways

The health data science careers pathways landscape

To first gain an overview of the jobs available in the health data science (HDS) space, we searched on the 'indeed' careers website, as this is one of the most popular job listings websites in the UK. We understood that relevant roles may be titled as something other than 'health data scientist', and so in order to ensure a thorough overview, we performed searches using the following: health data scientist; health data science; medical data science; medical data scientist; machine learning medicine; machine learning biology; and machine learning healthcare. We hoped that this range of searches would encompass a variety of roles. In total, this returned **16054** jobs. It is likely that not all of the results are relevant, and many are overlapping, but this was not so important at this stage, as we wanted a general overview to guide our decisions about the pathways on which to focus. The results are presented below.



Even if these do not accurately represent typical HDS career paths as there are potential biases (are the search terms accurate? Do some employers not advertise on these sites? Are start-ups represented?), we can see some general trends: the majority of jobs are full time; most are based in London; there are a large range of employers, with pharmaceutical giants dominating the landscape; the majority of jobs require university-level qualifications but there is a large representation for both undergraduate and postgraduate training required. We used this as a guide for our further investigations into HDS pathways, to see if these core elements and trends are visible in: previous research; specific career pathway example; and as advertised by employers.

Health data science careers literature analysis

To further understand HDS career pathways, we searched the literature for publications containing the terms "health data scientist" or "health data science" and "careers" or "jobs". There was a lack of relevant research published, likely due to the recent increase in HDS careers. The majority of publications focused on academic HDS pathways (Munevar, 2017) or HDS academic course structure (Aldridge, 2019) with a few statements about HDS careers in industry. The most relevant publication characterised the US HDS job market (Meyer,

2019). The report focused on analysing recent job postings which we believed could be useful for our research. Of particular interest were the author’s results to the question “What are the types of data scientist positions for which U.S. healthcare organisations are hiring and what is the focus of the work?” Figure 1 summarises the key skills required for HDS jobs in the US. These skills are broken down into junior and senior data scientist levels. Statistics, R, machine learning, storytelling, and Python were the most common keywords used in job postings.

Data Scientist Skills	Overall (All Job Postings) n = 3218	Overall Percentage Distribution	Data Scientist Level n = 2006	Senior Data Scientist Level n = 1094
Statistics (eg, general linear model, analysis of variance)	138	4%	94	40
R	136	4%	87	44
Applying machine learning techniques	133	4%	85	44
Storytelling; delivering actionable results	132	4%	91	38
Python	125	4%	79	41
Communicating findings	117	4%	79	36
Developing products	117	4%	75	39
Data-driven problem solving	112	3%	71	39
Data manipulation	108	3%	70	34
Developing algorithms	106	3%	66	34
Setting up/maintaining data platforms	97	3%	65	28
SQL	95	3%	63	27
Implementing models into production	89	3%	54	33
SAS	84	3%	54	27
Work in multidisciplinary teams	84	3%	55	28
Creating visualizations	73	2%	45	25
Identifying business problems to address	64	2%	38	23
Big and Distributed Data	61	2%	34	24
Hadoop	61	2%	25	34
Unstructured Data (eg, noSQL, text mining)	56	2%	34	20
Other	1230	38%	742	436
TOTAL	3218	100%	2006	1094

Figure 1: The range of data science skills mentioned in US HDS job postings (Meyer, 2019).

Additionally, the author defined sub-categories of roles within HDS job postings: performance improvers; product developers; modellers; and innovators (Figure 2). Performance improvers worked on areas such as quality measures, financial performance, and patient outcomes. Product developers focused on areas including population health, digital health, decision support, speech/language solutions, behavioural health, and claims analytics. Modeller job postings were nonspecific and required core data science skills, often focusing on machine learning. Innovators addressed areas such as health standards, personalised or precision medicine, genomics, and biology. These could offer us a useful way to further break down the HDS careers within the UK when defining our career pathways.

Sub-Category Details:

Category	Sub-Category	
Innovation	Data Scientist Level	11 5.6%
	Senior Data Scientist Level	3 1.5%
Non-Specific	Data Science Associate Level	3 1.5%
	Data Scientist Level	37 18.7%
	Senior Data Scientist Level	22 11.1%
Performance Improvement	Data Science Associate Level	7 3.5%
	Data Scientist Level	50 25.3%
	Manager/Director	2 1.0%
	Senior Data Scientist Level	15 7.6%
Product Development	Data Science Associate Level	1 0.5%
	Data Scientist Level	27 13.6%
	Senior Data Scientist Level	20 10.1%

Degree Required		
Advanced Degree	8	4.0%
Bachelor’s and Above	82	41.4%
Master’s and Above	76	38.4%
PhD and MD or Equivalent	18	9.1%
Not Listed	14	7.1%
Total	198	100.0%

Figure 2: Sub-categories and degree required found within HDS job postings. Each sub-category has different degree requirements which are described in more detail in the report. (Meyer, 2019).

Different skills were required within each category. ‘Product developer’ and ‘modeller’ careers require more data science domain expertise and can apply those skills to a wider range of areas. On the other end of the spectrum were the innovators, who had more training in research and science, particularly precision medicine. 5–7 years of experience were required in more than half of these positions, however in every category there were roles open to those with 0 years of experience. The top hiring organisations included biotech, vendors, and recruiters. Due to the lack of a similar analysis for the UK, we believe a similar publication based on an analysis of UK HDS job postings would be useful.

Linkedin individual career pathways

To have a better understanding of pathways for health data scientists, we performed case studies using LinkedIn user profiles. We searched for keywords including “data scientist”, “healthcare”, “statistician” and “bioinformatics” in job titles, and searched for healthcare institutions and companies that are well-known in the UK. After searching for more than 50 profiles, we picked two examples as listed below. They are not comprehensive due to the diversity of HDS pathways, but they represent two typical pathways. Other pathways include staying in academia throughout one’s career, or creating a start-up, but these are less common.

Example 1: Academic route (PhD + Postdoc + Research Associate, diverting to industry).

Job title	Education	Previous career	Listed Skills
Senior Data Scientist in Bioinformatics at a pharmaceutical company	<ul style="list-style-type: none"> - Master’s degree in Chemistry - PhD in Bioinformatics and Molecular Biology 	<ul style="list-style-type: none"> - Postdoctoral Fellow at a bioinformatics institute (1 year) - Research Associate at a cancer research institute (4 years) 	<ul style="list-style-type: none"> - Healthcare related: Bioinformatics, Genomics, Computational Biology, Molecular Biology, Biochemistry, etc - Programming/Data Science related: R, Python, Unix, Data Management, etc

Example 2: Industrial route, with applied HDS jobs. A PhD is not necessarily required.

Job title	Education	Previous career	Listed Skills
Data Scientist at a pharmaceutical company	<ul style="list-style-type: none"> - Bachelor’s degree in Chemistry - Master’s (research) in Drug Discovery and Development - Master’s (taught) in Health Data Science 	<ul style="list-style-type: none"> - Data Science Graduate Scientist at current company (2 years 1 month) - Internship experience: 4-month internship in HDS company 	<ul style="list-style-type: none"> - Healthcare related: pharmaceutical industry, UV/Vis spectroscopy, HPLC, NMR, medicinal, chemistry, etc - Programming/data science related: machine learning, R, Python, SQL, STATA, SAS, MongoDB, PyTorch, etc

From these 50+ profiles, the following can be summarised:

- Most of them have acquired at least a Master’s degree. A PhD degree is common. These degrees often have an interdisciplinary nature, where healthcare and data science are both important components.
- HDS professionals are usually proficient in multiple programming languages and data science skills.
- Knowledge in at least one healthcare domain is usually present in their education history.
- Internship experience at relevant healthcare companies or institutions is common when a more industrial or applied pathway is taken.

Employers’ perspectives

We then searched for companies that facilitated jobs within HDS or offered transitional routes from academia, as well as recruitment agencies specific to HDS.

In terms of **career transitions from academia into industry**, there were a few schemes set up to allow for this.

- *Faculty AI* <[link](#)> provides a fellowship programme for new graduates/post-doctoral researchers within STEM based fields to fast track careers within data science. They have a specific health & life sciences branch that has connections with companies such as: IQVIA, NHS England, BHF, Breast cancer now. These allow for 8 week placements within their customer companies with the main goal being employment post placement.
- *AstraZeneca* <[link](#)> provides a data sciences & AI graduate programme that allows for three 8 month placement periods giving new graduates a chance to be involved in a multitude of projects.
- *NHS* <[link](#)> provides a NHS Graduate Digital, Data & Technology scheme that offers 3 distinct pathways. The most relevant to our event would be the Data arm in which they employ data analysts

and information analysts. The scheme lasts for 2 years with the end goal being employment within the NHS in your desired position

- *Storm3* <[link](#)> is a recruitment agency specifically for the HealthTec arena. They have 2 specific consultants for data & analytics as well as other areas within HealthTec i.e. engineering & devops, sales & marketing, product management. They recruit for companies all over the world.
- *WeEngage* <[link](#)> is another recruitment agency specifically for HealthTec with a specific specialism within data & analytics.

In terms of pursuing **academia fellowships & research** as a career role, there are a few national institutes that try and support this.

- *NIHR Fellowships* <[link](#)> offer a range of fellowship opportunities for people in different stages of their career. Most relevant to our talks are HDS opportunities for post-doctoral students. This falls under the advanced fellowship options.
- *HDRUK Fellowships* <[link](#)> offer specific post-doctoral fellowship opportunities exclusively to HDS.
- *BHF Fellowships* <[link](#)> offer post-doctoral fellowship opportunities for an array of audiences e.g. clinical/non-clinical. For our event we would be focusing on data science related awards such as: BHF-Turing Cardiovascular Data Science Awards, Cardiovascular Catalyst Awards.
- *Wellcome Fellowships* <[link](#)> offer a vast array of awards and fellowships for different career stages. The one at most interest to us would be the Early-Career Awards that fund people for health related disciplines up to 3 years after their PhD.
- Various other charitable foundations offer fellowships e.g. Cancer Research UK etc - Plenty of fellowship opportunities for people who want to pursue research as a career.

Section 2 - Which pathways are we addressing?

From our research findings, we will aim to address the industry-focused pathways (biotech, pharma, HealthTech, healthcare organisations), rather than academia pathways. We will focus on the pathways that are open to those with 0 years of industry experience and that require a wide range of academic qualifications (bachelor, master and PhD level), since we are aiming our event at current students without experience.

Section 3 - Critical skills and experiences

As we are keen to focus on students transitioning from bachelor/masters/PhD studies into industry, we believe it would be valuable to have representatives from each of these levels, either as company representatives or as speakers at the event. Industry-specific skill building, such as CV coaching and headshots, would also be a useful addition. We are keen to have keynote speakers that are industry career-focused, with a deep understanding of the skills required for our defined pathways. The Association of the British Pharmaceutical Industry (ABPI) represents over 120 pharmaceutical companies and has a team dedicated to career development. Therefore a member of this team, such as Andrew Croydon (Skills & Education Policy and Examinations Director at ABPI), would likely be appropriate as a keynote speaker. Additionally, to gain a deeper understanding of the HealthTech industry, we would consider a speaker from Storm3, a HealthTech recruitment company.

References

- Aldridge, R.W. (2019) 'Research and training recommendations for public health data science', *The Lancet Public Health*, 4(8), p. e373. doi:10.1016/S2468-2667(19)30112-4.
- Meyer, M.A. (2019) 'Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings', *Journal of the American Medical Informatics Association : JAMIA*, 26(5), pp. 383–391. doi:10.1093/jamia/ocy181.
- Munevar, S. (2017) 'Unlocking Big Data for better health', *Nature Biotechnology*, 35(7), pp. 684–686. doi:10.1038/nbt.3918.