A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

THE DATA SCIENTIST: WHAT COMPANIES EXPECT FROM RECENT GRADUATES AND THE ROLE OF BUSINESS SCHOOLS

CAROLINA LOURENÇO MONTEIRO, 29216

Work project carried out under the supervision of:

Elizabete Cardoso

03-01-2021

Abstract

This research aims to understand if universities are preparing students well enough to cater to companies' current data science needs, particularly in terms of skills. To do so, the study reviews data science programs in Portuguese business schools and follows a mixed-method approach, by interviewing data science managers, and surveying recent data scientists online. It was concluded that there is room to improve the skillsets of graduates and recommendations are provided to business schools, who can have a strategic role in bridging the skill gap for future graduates, in what is consensually perceived as strategic, fast growing professional occupation.

Keywords: Data Science, Recent Graduates, Skillsets, Business Schools

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Data Science is a rapidly developing field, emerging in a context of unprecedented volume, variety and velocity of data, "every day, 2.5 quintillion bytes of data are created" (Zikopoulos et al, 2012), exceeding the capacity of the conventional analysis and management tools. Companies' uses of data have expanded yielding greater potential for efficiency and value creation. Thus, skills of those who can extract valuable insights from data are in high demand, with IBM predicting that "by 2020 the number of job openings for data scientists will increase by 364000 to 2720000 in the US". However, supply of graduates with the required knowledge seems not be meeting the demand. In Portugal, Fernando Matos leader of the Data Science Portuguese Association (DSPA) claimed in August 2020 that today "there were a few thousands missing". This mismatch can be driven by the lack of a common skillset framework (Markow et al, 2017). In fact, no concrete definition about the function has been reached, being commonly referred to as "a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming and systems engineering expertise" (Chang and Grady, 2019).

Further analysis is necessary to understand the challenges employers face, but the basis of training the future workforce remains in universities. Business schools in particular can have a strategic role as they are able to provide a multidisciplinary understanding. While there is a number of ways schools can play a role in increasing the employability of their graduates, this research will mainly focus on recommendations on the curricular aspects of data science programs within business schools. Portuguese business schools, are no exception, having introduced in the last couple of years bachelors, masters and post-graduation programs in data science and related area (Appendix 1). However, it is unclear if they are doing it in a useful way for companies, by addressing their needs. Therefore, this study starts by analyzing existing literature about the data science gaps;

reviews the available data science master's programs offered by Portuguese business schools (Appendix 2); purposes the assessment of the perception of companies in this regard, by means of interviewing data science managers and to confirm the findings, surveys the perspective of recent data scientists. Finally, recommendations, grounded on the insights from the analysis, are elaborated for business schools to help bridge the skill gap felt in the area.

2. Literature Review

Section 2.1: Data science - The field

Data Science is "an interdisciplinary field aiming to turn data into real value" (Van der Aalst, 2016) that has become popular in the past 20 years in a world increasingly dominated by data, "just like computer science emerged as a discipline when computers became widely available, data science is emerging as organizations are struggling to make sense of torrents of data." (Van der Aalst, 2014). It essentially covers two types of skills – technical, including data extraction, preparation, exploration, transformation, storage, retrieval and mining; and business related, including the understanding of business issues, the ability to provide useful insights, to frame the appropriate analytical solutions and to communicate them (Chiang et al, 2012). In fact, with the volume and variety of today's data, in addition to more powerful computers and emerging algorithms enabling deeper analysis, companies in almost every industry are exploiting data for competitive advantage. This results in both companies realizing they need to hire data scientists and universities establishing the programs to train them. (Provost and Fawcett, 2013).

Having first shown up in literature over 50 years ago, in John W. Turkey's "The Future of Data Analysis", the concept is presented as an "as-yet unrecognized science" whose subject of interest was learning from data (Donoho, 2017). In the late 1990's, the term started to gain more relevance in debates that called for the “need for statisticians to join with computer scientists to bring mathematical rigor to the computational analysis of large data sets” (Kelleher and Tiernery, 2018). They argued that a merger between the two “would produce a powerful force for innovation” (Cleveland, 2001). Some urged for the reformation of academic statistics, Professor J.C Jeff Wu even proposed to rename it “Data Science” as a way to expand it beyond its classical domain. Since then, data science has broadened beyond statistics, focusing more on data prediction and presentation (Chambers, 1993) rather than inference (Breiman, 2001). Nowadays, it is an area that has led Hal Varian, Google’s Chief Economist in 2009, to state that it allows its experts “…of being able to access, understand, and communicate the insights you get from data analysis”, in accordance with the article “Data Scientist: The sexiest Job of the 21st century”, which highlights the increasing demand for professionals with competencies in the field (Davenport and Patil, 2012).

Moreover, the fact that the term Data Science has had many attempting to define it, its use and its related terms throughout the years, alongside with it being closely intertwined with other prominent fields such as Big Data and Business Analytics (Provost and Fawcett, 2013), led to a lack of consensus regarding its definition and laid ground to the existing ambiguity in establishing the required skills for the role, as stated by Warden (2011) “there is no widely accepted boundary for what’s inside and outside of data science’s scope”.

Section 2.2: How Business Schools have catered to Business needs

Business schools have historically been associated with the goal of preparing their students for employment after graduation (Bennis and O’Toole, 2004). Thus, implying that the curriculum must be developed by taking into account the needs of businesses. In fact, the first curriculum of the world’s first business school - ESCP Europe, established in 1819 in Paris (Blanchard, 2009), reflected a combination of both theoretical and practical approaches, that included pedagogical games; promoted an international scope, as well as a social and demand-oriented approach to management to deliver value for society (Kaplan, 2014). Most of these values are still present in the foundations of today's European business schools, in which Management education according to Kaplan (2014) "Entails cross-cultural, societal management based on an interdisciplinary approach", and has evolved to reflect and adapt to the needs of the competitive reality of global businesses. Nowadays, their contribution to society keeps growing and can be translated in providing the right set of skills and the competencies sought by organizations, as referred by the preamble of the Association to Advance Collegiate Schools of Business's Eligibility Procedures and Accreditation Standards for Business Accreditation, "...management education must prepare students to contribute to their organizations…" (AACSB, 2006).

Nonetheless, what most literature on the subject has reported is that business schools have not always achieved success in catering to these business needs, or in other words, are not preparing their students well enough. For instance, Sadri (2002) surveyed both recent graduates and their employers on whether 7 identified core business school competencies were being taught correctly or if there was a need for more training, having concluded that the schools were not doing an "overly effective job teaching the competencies deemed important". Similarly, Dierdoff and Rubin (2006), developed a list of 6 managerial competencies and found significant mismatch between the importance allocated by managers to these same competencies and the extent to which they were covered in MBA courses.

When it comes to Data Science in the context of universities, it mainly existed in relation to Computer Science and Engineering programs. However, the demand for highly trained professionals and the growing interest from students led to the fast creation of programs specifically in the area (Vicario, Coleman, 2019). Moreover, Demchenko et al (2016) argue that, at present time, most of the existing curriculum still only offers a limited set of competencies compared with the breadth of knowledge future graduates are expected to possess in the data science professional life. Hence, suggesting that there is a significant gap between the knowledge of the available talent and the required one for the positions to be filled, particularly in what concerns business skills (Mikaelf et al, 2018). Building on that, Adam et al's 2019 study identified three competency groups (data science analytics, data science engineering and domain knowledge) whose balance throughout a data science program would be determinant to classify it as successful, the results showed that 59% of European programs were significantly imbalanced. Thus, despite there being a lack of understanding in the effective ways of implementing the ideal educational curriculum to fulfill these skill requirements (Mikaelf et al, 2018), the role of academia is apparent in ensuring this gap is bridged, as Asamoah, (2017) reinforces "There is a need for a curriculum that inculcates data-science related principles in business-focused majors". In addition, beyond producing talent, universities can also play the role of aiding in establishing a scientific definition of the data scientist profile (Volpe, Esposito, 2019). Which, in turn, highlights the need for collaboration between universities and companies to attain the best match between the curriculum and the integration in the labor market.

However, even with the interest in filling this gap, current literature has only scarcely focused on the development of academic curricula for data science programs within business schools. From the available body of knowledge, research on how to improve future data science academic curriculum and pedagogy development show that, firstly the understanding of the exact skills that are necessary and of the gap that currently exists is of outmost importance (Mikalef et al, 2018). Secondly, (Lu, J. 2019) adopted The Skills Framework for the Information Age (SFIA) for the novel interpretation of data science employability skills to create the appropriate corresponding academic modules and lastly, (Stanton, Stanton 2020) and (Volpe, Esposito, 2019), propose the use of job postings and position qualifications to establish the Data Scientist profile and to provide universities with recommendations on how to adjust.

Section 2.3: How companies' needs have changed

The emergence of big data analysis increasingly represented a competitive advantage, in almost every industry. In fact, data, provided that it is turned into actionable insight, is the main driver in leading companies to the digital, some cases include - optimizing enterprise operational processes, supporting innovation in industry, market demand predictions, decision making (Yin, Kaynak, 2015) and providing advanced personalized customer-centric service delivery (Demchenko et al, 2016). Moreover, McAfee and Brynjolfsson's (2012) large-scale study supports the claim of data being essential when noting "the more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results". Thus, the increasing adoption of data and the growth of its dependency to face competitive business' needs, implied a shift not only in the job functions related to the area but also in the existing tools. Indeed, this meant that the conventional methods and applications to manage and analyze data sets were no longer sufficient to handle the new amounts of data, of such volume and variety (Fawcett, Waller, 2013), being it critical to modernize infrastructure and leverage new disruptive technology. In the past two decades data science and its related fields of big data and business analytics have developed and started to mature (Stanton, Stanton, 2020). Today, according to the world's most popular data science platform Anaconda, the steady journey to maturity of Data Science in organizations is ongoing, creating the need for interdisciplinary teams and crossing corporate silo boundaries with the expected goal of in the next couple of years becoming a strategic business function of its own.

Companies consider it critical to seek professionals capable of unlocking the value in data (Marrow, 2018). Several evidence points towards increasing demand for highly trained analytics talent, with the skill set of cross-functional skills, business acumen (Stanton, Stanton, 2020) and deep knowledge of data intensive technologies (Demchenko et al, 2016), starting with the rise in demand for data scientists in LinkedIn's "Analytics and Data Science Job Growth" (Appendix 3) (Patil, 2011) to being ranked number 3 in LinkedIn's 2018 "Top Emerging job analysis" based on members with public profiles that held a full time position for 5 years in the US. In addition, in a recent study conducted by the Graduate Management Admission Council, 76% of Fortune Global 100, 80% of Fortune Global 500, and 72% of for-profit public companies plan to hire Master's qualified analytics talent in 2019 (GMAC, 2019).

As for the required skills sought by employers, different studies have concluded that, in addition to requiring deep domain knowledge (Waller, Fawcett, 2013) – which means business understanding specific to a particular field - analysis, machine learning, statistics and data mining were found to be the top hard skills required, along with the programming languages Python, SQL, R, Hadoop and Java (Hale, 2018) (Lee, 2016). As for soft skills, interpersonal skills, communication, teamwork and leadership were commonly noted as key for the position (Stanton, Stanton, 2020). In spite of these results, there is still no consolidated denomination for the role and the skills, as most of them overlap with the ones of other data related job positions (Volpe, Esposito, 2019).

Finally, regardless of the apparent opportunity big data presents to organizations, it can also pose challenges directly related to the data scientist profession. Namely, there is a shortage of talent (Lee, 2017), as companies struggle with finding "available and qualified data scientists who can make sense of big data with a proper understanding of the domain and who are comfortable using analytical tools" (Storey, Song, 2017). This gap is noted for both entry-level positions (Stanton, Stanton, 2020) and for senior positions. Related to this, Volpe and Esposito (2019) highlight the fact that while analytics are becoming more sophisticated, data scientists are still in their infancy.

Section 2.4: Research Questions

As aforementioned, the prominence of data in the future is almost certain and expected to keep becoming more critical, especially in the ever-changing business environment. Therefore, considering the relevance of the data science function for the future, the reported existing gap in finding skilled talent, for the current demand, becomes urgent to be fixed. Thus, building on the existing literature, which is limited, especially in what refers to business school graduates going for entry-level data science jobs, this research aims to contribute to helping both companies and universities bridge this gap. More precisely, the following research questions have been developed:

RQ1: What do data science managers think are the main data science skills gaps?

RQ2: What do graduates, recently employed, think are the main data science skills training gaps?

RQ3: How can Business Schools help in bridging those gaps?

3. Methodology

The present study follows a **critical realism** research philosophy. This research paradigm is rooted in a realist or positivist ontology, in that it recognizes the existence of a measurable reality that is external and independent to us, but it also feeds from the subjective or interpretivism epistemology, in as much as it acknowledges that our knowledge of this reality is filtered and limited by our own interpretations and perceptions. The starting point is the ontological, departing from an initial theory – the data science skill gaps found in literature - which is then subject to being supported, elaborated on, modified or denied, in order to help build a more accurate understanding of reality

(Fletcher, 2016) and to identify the causal mechanisms driving the social events (Archer et al, 1998). In this case, this is accomplished by analyzing the perceptions of both managers and recent graduates, on what is lacking on hired talent from the former, and what is missing from university training in the field, from the latter. After understanding the reality through both those lenses, this study will be able to suggest ways business schools can contribute to fixing the gaps.

Given the research questions and studies reviewed in the literature, it was decided to use a sequential mixed-method structure, first targeting the data science managers population with semi-structured interviews (online, for covid-19 compliance), and then recent data scientists with an online survey.

The qualitative data was collected through semi-structured interviews during November and December of 2020, including a set of open-ended and predetermined questions targeted at a preferred specific group of respondents. The use of open-ended questions gives the participants complete freedom when answering, without being limited to fixed options, which in turn, allows for richer and more explanatory responses and even ones not anticipated by the researcher. (Mack et al, 2019). The full interview script can be consulted in Appendix 4. Starting with general questions about the data science area in the company of the interviewee, the first section focuses on the current importance of the role within the organization and on identifying the needs still being felt. Then, the second section aims to pinpoint whether recruitment issues are felt. Followed by a third section, more comprehensive and focused on the necessary competencies and skill sets for the job. The fourth section is dedicated to retrieving opinions about the possible role business schools can have in helping companies bridge the gap. Finally, to wrap the interview up, the respondents were asked to provide any additional comments they considered relevant. In total 10 interviews were conducted, held in English with the exception of one, which was conducted in

Portuguese per respondent's preference, lasting for approximately 30 to 40 minutes, via Microsoft Teams. All were recorded and later transcribed, and in one case translated, to perform content analysis of the results (Excerpt of the content grid in Appendix 5).

After gathering the experts' insights, 4 hypotheses were formulated to be confirmed in the next phase, by measuring the perception of the recent graduates working in data science through a self-completed online questionnaire. The preferred target were recent graduates, newly working in data science – up to 3 years of professional experience - as they provide a direct link between the company's outlook on the area and whether the preparation received in university was enough. As such, this target was understood to have a unique positioning in the present analysis – not only can they provide another perspective on the data science skill needs experienced in organizations, as data scientists, which may match, complement or disagree with the experts' (Lisá et al, 2019) but at the same time, as recent graduates, they also have the advantage of still having present their first-hand university experience, provided that there might be some bias, making them strategic in the process of designing recommendations on bridging the gap.

The questionnaire starts off with an eliminatory question to ensure the participants have worked or are currently working in the field of data science, if the participant confirms so, then it surveys through multiple-choice questions about their position and background, asks to rate the level of agreement with statements about data science and the importance of different types of skills, finishing with a final section pertaining to the demographic information about the participants. The full questionnaire can be consulted in Appendix 6. It was distributed on social media platforms Facebook and LinkedIn, through private messaging using snowball sampling, asking participants to share with colleagues who shared the intended profile, and through public posts in groups dedicated to data science, analytics and similar areas, which were understood to have a considerable size of people fitting into the target profile. The quantitative data collected during this second stage was then analyzed in IBM SPSS to perform descriptive and inferential statistics.

4. Analysis and Discussion

4.1. Qualitative Findings

This section will first characterize the interviews sample, and then present the results in the different themes, before summarizing the answer to RQ1 -What do data science managers think are the main data science skills gaps? – and raising the hypotheses for the subsequent survey.

The conducted interviews, included 5 lead data scientists, 2 directors, one of data science and the other of advance analytics, 1 business intelligence and analytics manager, 1 product manager and 1 advisor on digital technology and innovation policies in the European Commission. From consulting agencies on analytics (BiLD Analytics), business and IT (Everies Portugal), digital (Wyoming Interactive, Impossible), marketing (Merkle) and cloud solutions (Amazon Web Services), to online marketplaces (JustEats), to software provider (JustPark) and to insurance (Fidelidade). The first 2 and the last one are mainly based in Portugal. (Appendix 7 summarizes this information).

On whether the function will increasingly become more relevant, there is a definitive affirmative consensus among all interviewees. With data becoming more sophisticated and efficient every day, data science is able to showcase better results, more companies are starting to adopt it, making it a clear part of the future. As society keeps moving into a digital era, more and more acquiring the right digital skills is of outmost importance. However, at the current exponential growth of data creation, the rate of our capacity to analyze it must strive to keep up with that pace, presenting an opportunity for more professionals to come into the area and a need to keep learning more about it and how to reap its associated benefits. *"The gap between the data that we generate and store and* *our capacity as humans to extract value from it is getting bigger each year, so, I think that the opportunity for professionals will grow more and more because we need to keep the pace of the data creation"* (I1: Lead Data Scientist, Analytics consulting, PT). It is worth mentioning, that this growth might not be as accelerated as it was a few years ago, when the field of data science was first generating hype, in conformity with Stanton and Stanton, (2020) who claim business analytics have started to mature, one respondent remarks that *"I do think that although it's not going to be, in my opinion, an adoption and a growth as steep and as exponential as it was before, it's still going to continue to increase."* (I4: Director of DS, Digital consulting, not PT)

Concerning the data science needs companies experience, (4/10) interviewees mention the need for more senior level data scientists, with deeper experience in both projects and commercial aspects, to mentor the more junior members. "*One of our current major needs are senior people, people with experience in machine learning projects, actual experience in with the processes and that have really deep knowledge of the things and can solve problems."* (I7: Lead Data Scientist, Business and IT Consulting, PT)*; "(...) there needs to be someone to talk to stakeholders and identify potential use cases and communicate those and pitch those project ideas to the stakeholders in the company."* (I6: Lead Data Scientist, Online Market place, not PT). A better mix of social and engineering skills, improving the best practices of data scientists in data engineering and in dealing with big data are desired, as well as need for new sources of data. Another ability companies would benefit from is being able to spot opportunities for data science applications, I5 reports that, despite recognizing its added value, some companies are not even aware of their own needs in this area.

Regarding the recruitment of data scientists, from the majority of those who have hired before (8/10), (6/8) mention that one of the reasons for it to be challenging is because with the broadening of the term “data scientist” there seems to be a lot of people in the market calling themselves after it, when they are only proficient in coding and lacking the necessary background to understand the basis of what they do and how to communicate it. *“People hear about data science and they know it is a hot topic, so they watch a couple videos online and they call themselves data scientists, when they have no idea of even the most basic questions about math, anything about the algorithms.”* (I7: Lead Data Scientist, Consulting, PT).

In regard to entry level data scientists, recruiters felt that it can be easier to hire them than more experienced staff. For example, by running graduate programs or going straight to universities, they manage to attract a considerable number of candidates. This can depend on the attractiveness of the company, traditional industries seem to struggle more on this field, or on the geography, as I8 from Portugal puts it *“the best talent is taken by foreign companies”.* In contrast, there is special reference to the shortage of senior data scientists, which despite improving every year, it is still agreed that for the demand for specialization, there might be a problematic lack of experienced and skilled talent, supporting the findings of Storey and Song (2017). This seems to be affecting the industry, especially the SMEs – as shown by the higher costs associated with hiring these experts which I3, I4 and I5 make reference to. *“The easiest trade-off analysis that we can do is how much costs an IT expert in a company and how much a marketing, finance or commercial expert in the company and we see the balance of the IT guys to be much more well paid than the other category, so obviously we’re lacking experts on that areas.”* (I3: Digital technology and innovation policies advisor).

On the topic of required skills for recent graduates, they can be grouped into four main categories. To start, all interviewees made reference to the importance of having a strong quantitative and analytical background in math and statistics for data science. Coding skills are also a requirement, with the most mentioned languages being Python and SQL, in accordance with Hale (2018) and Lee (2016). Business knowledge is essential so to ask the right questions, define the commercial problems and find the respective best solutions, as well as understand how to face the business challenges for data science and being able to translate the technical terms into business ones. Finally, soft skills are comprehensively highlighted, resilience and problems solving appear to be especially relevant for this function. *"A lot of times our first attempts don't work, and it is important to keep going and not lose the enthusiasm."* (I8: Director of Advanced Analytics, Insurance, PT). Being able to learn autonomously and willing to learn and adapt continuously to new tools are also key. Plus, knowing how to communicate and present findings in clear and understandable ways is critical. Other soft skills that were pointed out as important include: being passionate about data, ability to innovate, skills in media literacy to tackle misinformation, and bringing societal and ethical concerns into coding and data. Although overall all should be mandatory for a data scientist, depending on the complexity of the role, for a junior there is more emphasis on dominating the technical skills, given that the remaining can be acquired with experience and maturity, while for a more senior data scientist or a manager role the business and soft skills matter more.

As far as the skills recent graduates seem to be lacking in the most, generally, business understanding and awareness about how decisions work is lacking, also experience in programming beyond basic knowledge of a language is noted. In addition, half of the interviewees alert graduates to be aware of the differences between academia (theoretical problems) and companies (open-ended problems). The latter implies a certain process flow along with teamwork with different people from various areas and roles. Plus, "real-world" data is messy and often times unavailable. Usually, graduates underestimate the amount of time they will be spending trying to gather data, cleaning and preparing it. *"80% of your time you're just cleaning data, chasing for* *data and that's a big part of the job that sometimes we feel like on interviews when you have like case study for new hires, they struggle a little bit to understand the challenges of data."* (I10: Lead Data Scientist, Marketing Consulting, not PT).

When it comes to how satisfied employers currently are with the skills of graduates, by acknowledging that some sort of additional training is always necessary, respondents answer this question. While ideally this would preferably not be required, it is still naturally expected. *"They definitely need training because real life is much more complex than theoretical life and data is super messy in real life as opposed to what people usually get from school."* (I2: BI manager, software provider, not PT). The type of training provided is tailored depending on the background and experience of the graduates. For instance, one might be proficient in one programming language but might need to develop further some other useful languages or tools used in the market. I6 adds the need for improved data engineering best practices, to make code more efficient. *"But usually they're not that good in for instance, cloud technologies, like Azure or AWS. So, we need to give them training on these tools. Because these are the tools that the market is using. So, you need to know them otherwise you won't be able to work on the market"* (I1: Lead Data Scientist, Analytics Consulting, PT). Besides the technical and coding skills, coaching in soft skills is also considered critical, namely when it comes to stakeholder management, communication, translation of data terms and how to work in teams. In the same line, business knowledge, commercial and subject-matter expertise also might be taught. *"When we had the graduate program, the 1st month was just dedicated to training, most of it just in the commercial side of it. Industry specific jargon."* (I4: Director of Data Science, digital consulting, not PT).

Regarding the academic background of the talent, the reported ones as most common are engineering, electronics, biomedical, computer science, mathematics, physics, statistics, information systems and occasionally business, economics and psychology. However, most mention that there are multiple paths to data science, which could indicate that the first degree of education might not be relevant if later some specialization (Master's degree, PhD, bootcamps) is pursued. In fact, as I2 puts it *"Data science is not about being a pure tech developer, you don't need to know how a computer works to do good data science, you just need to be really keen to learn and always learning all the time, because it is always evolving."*

While nowadays there is much more awareness about the data science area, which led to the creation of new courses, post-graduations and master's degrees, there appears to be a general agreement that schools are lagging behind. Not only do schools take too long to adapt to the current market needs but they should be focusing on skills the market will need in the future. These results go in line with those of Mikaelf et al (2018), that the understanding of the ways of implementing the ideal educational curriculum to fulfill these skill requirements is missing. On whether business schools could have a role in bridging this gap, I2 states *"I think they should have done it yesterday."* I1 supports this opinion as it is his belief that in the future any manager will need to have some level of knowledge of data science, as most decisions will be taken based on data. To succeed, all state that there must be a focus on both business acumen and technical knowledge, as well as, experience with practical applications - An example would be debugging exercises to spot errors. *"In this data science position any base of academics on both development and business can provide its own synergies."* (I5: Product Manager, digital consulting, PT).

Still, at the moment, (3/10) show preference for traditional engineering/math/statistics/physics schools over business schools, mainly because commercial and communication skills are easier to teach on the job. To increase business school graduates' employability some suggest: integrating collaborations with enterprises in recruitment fairs; having experience with projects, via internships or bootcamps for example, along with creating GitHub accounts to share them, since it allows recruiters to know tangibly what a graduate can do by asking questions such as "Why did you do it that way?"; "What would you have done differently?" to figure out if one can actually understand code. *"Even if it is not your best project ever, there are lessons that you learned doing it that are very important, even over quality.*" (I7: Director of Advanced Analytics, Insurance, PT).

With this analysis completed, **RQ1** is addressed and in sum, managers report the main skills gaps of data science graduates to be related to the lack of efficient programming and to the limited business understanding, stakeholder management and communication competencies. Awareness of the challenges of data and the reality of working with it, is also missing. Furthermore, the following hypothesis to be elaborated and in the next data collection phase further developed, as well as validated:

H1: For the success of a data scientist, the 4 types of skills are key

H2: Companies need more experienced staff and better engineering practices, in data science

H3: Graduates are lacking mostly in business understanding, coding expertise and knowledge of "real-world" data challenges

H4: Graduates need in-job training in multiple types of skills, depending on background

4.2. Quantitative Findings

This section starts with insight into the obtained sample, before describing general findings and summarizing the answer to RQ2 - What do graduates, recently employed, think are the main data science skills training gaps?

In spite of various efforts involving the identification and contact of dozens of recent data scientists, the fact remains that there are few data science programs in Portugal, taking in limited cohorts. If we further refine the population to those entering the market in the last 3 years, we are dealing with a group with few hundred individuals at most. This unfortunately led to the online questionnaire obtaining only 49 responses, with no more than 34 considered valid for the purpose of this analysis. Consequently, the sample size limits the robustness and statistical significance of the following findings, and results should be taken as exploratory.

In terms of the sample characterization, 31.3% identify as female and 6 nationalities are represented - American, Dutch, Indian, Italian, Pakistani, Portuguese – the latter including the vast majority of respondents (83.9%). Given the target, expectedly most of the sample was aged between 18 and 24 (45.9%) and between 25 and 34 (34.4%). Most participants held a master's degree (65.6%), with the remaining 31.3% reporting their highest level of education to be a bachelor's degree and 3.1% a PhD.

Regarding how long they have been working in the area, 94.1% have been in it up to 3 years - the amount of experience considered for them to be recent graduates in this research. Among those, 58.8% have up to 1 year of experience and 26.5% have 2 years of tenure. Related to this, 55.9% of the sample's level of seniority is entry-level, followed by 26.5% having junior positions. Along the same line, the vast majority (76.5%) has no people who report back to them. Some sort of correlation could be expected between being in the job for less time and having less people reporting to them, and indeed running a Fisher-exact test (because the requirements for a chi-square test were not met as most cells had values below 5) rejects the independence of both variables (p-value: 0.008). Furthermore, this sample represents different industries, with IT being the most popular (32.4%), followed by banking (23.5%) and consulting (17.6%).

The first academic degree of participants is fairly distributed, confirming the findings from the qualitative interviews that there are many paths to data science, with the most common being

Mathematics, Statistics or Physics (29.4%); Engineering (26.5%) and Business, Economics or Finance (26.5 %). In terms of additional training to work in the field (excluding in-job training), 55.9% completed a postgraduate or master's program, while 44.1% instead took other subject-specific courses. By cross-tabulating both variables, it is observable that only respondents who studied Engineering or Information Systems, participated in no training at all, however no statistically significant correlation was found between the two. (Fisher exact test p-value: 0.865).

For the next section, Table 4 summarizes the following results. On the extent respondents agree with data science being an important function in their current organization, on a scale from 0 (Completely disagree) to 10 (Completely agree), results show an average rating of 7.69, alike the opinion of managers. When testing this statement against other variables, gender produced significant results (ANOVA p-value 0.018), as it seems that females tend to agree to a lesser extent about the importance of data science for their organizations. Moreover, graduates agree to a stronger extent that data science will increasingly gain relevance within companies in the years to come, once more following what was reported by the experts. As such, these two statements relate to one another (r=0.535**). On the topic of skills, participants agreed that data-related skills (e.g.: data engineering, data analysis including data visualization, modelling) are the most important for a data scientist, but neither agreed or disagreed that those skills are enough for the success of a data scientist. This indicates the relevance of the technical competencies but shows that they might not be sufficient for success in the field. Indeed, when introducing a new type of skill, business-related skills (e.g.: marketing concepts, communication, customer experience) and questioning if both are equally important, the average rating was 7.18, meaning that there is a tendency to think a combination of both is key for a data science professional.

Looking deeper, at a set of 13 skills, which included 3 out of the 4 groups of skills identified during interviews - ones related to business, math and statistics, and coding, participants rated them, on a scale from 0 (Not important at all) to 10 (Extremely important). Table 5 summarizes the following results, with the top 3 most important being “data preparation and management skills”, “data visualization skills” and “experience with real-world data science applications”. While the 2 less important were “database engineering” and “coding in R”. These ratings go in line with the qualitative insights, particularly the relevance given to experience and the fact that data engineering, while evidently helpful, might not be a priority for success in the area. In addition, no skill group seems to considerably outweigh the other, as for instance, “business and strategy knowledge”, “research thinking, hypothesis formulation and statistical analysis” and “coding in SQL” have very similar mean values, supporting once more the combination of all. Testing all the competencies against academic background resulted in significant values for “experience with real-world data applications” (ANOVA p-value: 0.022), it looks like the Business/Economics/Finance and Information Management/Information Systems graduates consider this to be less important for data science than the graduates from Computer Science, Engineering and Math/Statistics/Physics. Comparing the industry differences in means of the importance of the competencies, “data visualization skills” shows statistical significance (ANOVA p-value: 0.034) - all represented industries, with the exception of retail, rated it important (ranging between 7.00 and 8.72) but the highest mean value came from IT.

In terms of soft skills, a list of 9 was analyzed, (Table 6 summarizes the next results) “problem solving” and “analytical thinking” were reported as the most important, followed by “critical thinking”. For these 3 skills, no ratings below 7 were given. “Resilience” comes in fourth place with a mean of 8.40. Once more, these results confirm the statements from the managers.

Interestingly, while only being mentioned in one of the 10 interviews, “ethical behavior” has a mean value of 8.16, scoring slightly above “teamwork” and “communication”, which were brought up by all interviewees. This suggests that ethics in regard to data is an issue that should be further discussed. Finally, “leadership” is considered the least important, which was somewhat expected from graduates in mostly junior roles. As per interviews, if a recruiter wants to hire a data science manager, leadership skills will be critical but not necessarily for a pure data scientist. Analyzing the importance of the soft skills against gender, produced a significant effect for “communication” (ANOVA p-value: 0.022): females consider it less important than males. The type of training incurred in also seems to slightly affect “Critical thinking”, (ANOVA p-value: 0.049) with those who selected “I did not need additional training”, rating this skill as less important than those who completed master’s programs or other courses.

Overall what graduates wish they had learned more in depth, in order of most voted skills, lays in “experience with real-world data science applications”, “experience in building and using algorithms”, “coding in python”, “working with big data” and “business and strategy knowledge” (full list in Appendix 11). These opinions correspond to what managers reported as the most lacking skills in recent graduates – experience in real-life projects and in coding. Moreover, when asked which of the skills they think companies are struggling the most with, the participants’ response shifts a little from the previous question (full list in Appendix 12). While “experience with real-world data science applications”, “experience in building and using algorithms” and “working with big data” are still among the most voted, “data preparation and management” is voted up, and understandably “business and strategy knowledge” gets less votes. “Coding in python”, is also voted down, against interviewees reporting problems with efficiency in coding. One respondent used the open comments feature on the online survey to elaborate that businesses lack the modern data infrastructures for what they require data scientists to practice and that business leaders still do not know how to make data-driven decisions. While this is not statistically significant, it does resonate with anecdotal reports in the field.

Lastly, from the 55.9% graduates who completed a postgraduate or master's program in the data science area, on a scale from 0 (Completely disagree) to 10 (Completely agree), (Table 9 summarizes the following results) respondents believe that their university's core curriculum addressed well enough the needs of companies. Similarly, in their perspective, university prepared them well enough for the data science job. Nonetheless, on average they feel that they were better prepared in terms of data-related skills than in business skills. Furthermore, when seeking additional training 57.9% reported they looked at it for both types of skills and 31.6% focused mainly on data-related skills. A relationship between feeling prepared in regard to any of the types of skills and academic background was suspected, as reported in the qualitative analysis, and statistically confirmed through an ANOVA (p-value=0.018 for data-related skills, p-value= 0.022 for business skills). Respondents whose first degree is in business, report they were the best prepared in terms of business skills (mean=6.8, SD=0.98) than the rest and conversely, business graduates rate their preparedness in terms of data-related skills an average 5.8 compared to the others. The ones who feel better prepared in terms of data-related skills are information management/systems graduates (mean=7.50, SD=2.12).

In sum, this analysis answers **RQ2**, with data suggesting that key issues with training reside in better preparation in business understanding and more experience with real-world applications and handling big data. Moreover, these findings **do not reject H1**, as graduates do not rank one particular type of skill above the other, although there is some fluctuation on the ranking of the soft skills compared to the stated in interviews; **do not reject H2**, since graduates' report of the skills they wished they had learned more in depth matches the ones from managers; **partially reject H3**, in the sense that while there is convergence on need for more experience, graduates do not report engineering skills issues; **do not reject H4**, given that there is statistical significance between both variables.

5. Conclusions and Recommendations

Overall, the findings from the present study confirm the ones in the existing body of knowledge, contributing to it by adding more insights about data science needs within companies. During the discussion, some divergence emerged when the survey alerted to the importance of ethics in regard to data that had emerged less in the interviews and to the effects that gender may have in certain aspects of the data science area. Grounded on these, **RQ3 -** How can Business Schools help in bridging those gaps? **-** is answered, in the form of recommendations: ultimately, there is room for business schools to play a role in bridging the skills gap expressed, which managers made clear and survey respondents reflected in their need for additional training and not agreeing above a mean value of 7 that universities prepared them well enough. Additionally, from the revision of Portuguese business schools' data science programs, it appears that the curriculums in place are in the right track to bridge the gap, with the majority of them incorporating and addressing most of the needs and concerns reported in the interviews with managers and from the survey.

Nevertheless, schools must ensure they are agile and responsive to the needs of a continually evolving field. The ideal curriculum must have: technical quantitative background thoroughly covered; programming taught beyond the basics, promoting efficiency in coding and focusing on practical problems (examples – open-ended problems, exercises on debugging and spotting errors in code), currently Python and SQL should be the requirement but this could rapidly change, so the foundations of coding must be a priority to allow them to be easily transferable; must provide business understanding and know-how, particularly in terms of translating data terms into non-technical terms; must develop student's soft skills, mainly resilience, problem solving, critical and analytical thinking, and must bring awareness to ethics in regards to data. The curriculum should also reinforce the realities of dealing with data, such as the time consumed in chasing for it, cleaning it and understanding it. Further, it should promote an environment that encourages students to learn independently; experimentation, by requiring internships, offering upskilling training modules or boosting participation in data competitions. As a basis, it must keep track of what are the in-demand tools to upskill their students. In addition, beyond curriculum schools can help structure and define the function of data science, in accordance with Volpe and Esposito (2019), for instance, by collaborating with companies in recruitment fairs to inform students about the wide range of opportunities when working with data science.

As for limitations, besides the small sample size not being representative enough of the research population, the lack of diversity in nationalities and industries, in both research phases, may also have influenced results. The fact that the qualitative data is not directly comparable with the quantitative, also remarks as a limitation when rejecting or not the elaborated hypothesis. In addition, the questionnaire included rating and ranking questions which are dependent on a subjective nature of interpretation.

Future research should therefore focus on the perspectives of a more diverse sample of industries, namely include some traditional companies and SMEs, less fluent in data, as they start incorporating the field, to confirm if they struggle more than well-known tech companies and startups. Another interesting issue raised worth exploring, is the problem of lack of qualified seniors in data science and specialization of staff in a world where data keeps getting more complex, at a very fast pace.

6. References

"Anaconda Releases 2020 State of Data Science Survey Results." Anaconda. Accessed October 17, 2020. https://www.anaconda.com/press/anaconda-releases-2020-state-of-data-science-survey-results.

"Business Analytics." Curso em Detalhe | Católica Porto Business School. Accessed January 2, 2021. https://www.catolicabs.porto.ucp.pt/pt/curso-detalhes/business-analytics/9300.

"Business Education." FT.com. Accessed January 2, 2021. http://rankings.ft.com/businessschoolrankings.

"Master (MSc) in Data Science." Iscte. Accessed December 13, 2020. https://www.iscte-iul.pt/course/297/master-msc-in-data-science/presentation.

"Master of Science in Business Analytics." CATÓLICA-LISBON. Accessed December 13, 2020. https://clsbe.lisboa.ucp.pt/masters-science/master-science-business-analytics?_ga=2.209627348.798887544.1607817912-1960832129.1607817912.

"Master's in Business Analytics Overview". NOVA SBE. Accessed December 13, 2020. https://www2.novasbe.unl.pt/en/programs/masters/business-analytics/overview.

"Masters in Data Analytics for Business" ISEG Lisbon. Accessed December 13, 2020. https://www.iseg.ulisboa.pt/aquila/cursos/mdab.

Abraham, Steven Eric, and Lanny A. Karns. 2009. "Do Business Schools Value the Competencies That Businesses Value?" *Journal of Education for Business* 84(6): 350–56. https://doi.org/10.3200/joeb.84.6.350-356.

Archer, Margaret., Bhaskar, Roy, Collier, Andrew, Lawson, Tony and Norrie, Alan. (Eds.). 1998. *Critical Realism: Essential Readings* (1st ed.). Routledge. https://doi.org/10.4324/9781315008592

Asamoah, Daniel Adomako, Ramesh Sharda, Amir Hassan Zadeh, and Pankush Kalgotra. 2017. "Preparing a Data Scientist: A Pedagogic Experience in Designing a Big Data Analytics Course." *Decision Sciences Journal of Innovative Education* 15(2): 161–90. https://doi.org/10.1111/dsji.12125.

Belloum, Adam S.z., Spiros Koulouzis, Tomasz Wiktorski, and Andrea Manieri. 2019. "Bridging the Demand and the Offer in Data Science." *Concurrency and Computation: Practice and Experience*. https://doi.org/10.1002/cpe.5200.

Bennis, Waren., O'Toole, James. "How Business Schools Lost Their Way." *Harvard Business Review*, August 1, 2014. Accessed October 7, 2020 https://hbr.org/2005/05/how-business-schools-lost-their-way.

Bichler, Martin, Armin Heinzl, and Wil M. P. Van Der Aalst. 2016. "Business Analytics and Data Science: Once Again?" *Business & Information Systems Engineering*, 59(2): 77–79. https://doi.org/10.1007/s12599-016-0461-1.

Bogner, Alexander, Littig, Beate and Menz Wolfgang. 2009 "Introduction: Expert Interviews — An Introduction to a New Methodological Debate". In: Bogner A., Littig B., Menz W. (eds) *Interviewing Experts. Research Methods Series.* Palgrave Macmillan, London. https://doi.org/10.1057/9780230244276_1

Chambers, J. M. (1993), "Greater or Lesser Statistics: A Choice for Future Research," *Statistics and Computing*, 3: 182–184.

Chang, Wo L., and Nancy Grady. 2019. "NIST Big Data Interoperability Framework" Big Data Taxonomies 2 https://doi.org/10.6028/nist.sp.1500-2r2.

Chiang, Roger H. L., Paulo Goes, and Edward A. Stohr. 2012. "Business Intelligence and Analytics Education, and Program Development." *ACM Transactions on Management Information Systems*, 3(3): 1–13. https://doi.org/10.1145/2361256.2361257.

Danermark, Berth, and Ekström Mats. 2019. "Explaining Society: Critical Realism in the Social Sciences". Abingdon, Oxon: Routledge.

Dignan, Larry. "Data Science Dominates LinkedIn's Emerging Jobs Ranking." ZDNet. ZDNet, December 10, 2019. https://www.zdnet.com/article/data-science-dominates-linkedins-emerging-jobs-ranking/.

Donoho, David. 2017. "50 Years of Data Science" *Journal of Computational and Graphical Statistics*, 26(4): 745-766. https://doi.org/10.1080/10618600.2017.1384734

Fletcher, Amber J. 2016 "Applying Critical Realism in Qualitative Research: Methodology Meets Method." *International Journal of Social Research Methodology* 20(2): 181–94. https://doi.org/10.1080/13645579.2016.1144401.

Jifa, Gu and Zhang Lingling. 2014. "Data, DIKW, big data and data science" *Procedia Computer Science*, 31: 814-821. doi: 10.1016/j.procs.2014.05.332

Kaplan, Andreas. 2014. "European Management and European Business Schools: Insights from the History of Business Schools." *European Management Journal* 32(4): 529–34. https://doi.org/10.1016/j.emj.2014.03.006.

Kelleher, John D., and Brendan Tierney. 2018."What Is Data Science?". In *Data Science*. Cambridge, MA: The MIT Press. 19-22.

Lee, In. 2017. "Big Data: Dimensions, Evolution, Impacts, and Challenges." *Business Horizons* 60(3): 293–303. https://doi.org/10.1016/j.bushor.2017.01.004.

Lisá, Elena, Katarína Hennelová, Denisa Newman. 2019. “Comparison between employers’ and students’ expectations in respect of employability skills of university graduates.” *International Journal of Work-Integrated Learning*, 2019, 20(1): 71-82.

Lu, Jing. 2019. “Data Science in the Business Environment: Skills Analytics for Curriculum Development.” *Machine Learning, Optimization, and Data Science Lecture Notes in Computer Science*. 116–28. https://doi.org/10.1007/978-3-030-13709-0_10.

Mack, Natasha, and Cynthia Woodsong. 2005. *Qualitative Research Methods: A Data Collector's Field Guide*. North Carolina: FLI.

Markow, Will, Braganza, Soumya, Bledi Taska, Miller, Steven M. and Hughes, Debbie, 2017. “The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market”. Technical Report. Burning Glass Technologies.

Mikalef, Patrick, Michail N. Giannakos, Ilias O. Pappas, and John Krogstie. 2018. “The Human Side of Big Data: Understanding the Skills of the Data Scientist in Education and Industry.” *2018 IEEE Global Engineering Education Conference (EDUCON)*. https://doi.org/10.1109/educon.2018.8363273.

Patil, D. J. “1. Building Data Science Teams.” Chapter. In *Building Data Science Teams: the Skills, Tools and Perspectives behind Great Data Science Groups*. Sebastopol, CA: O'Reilly, 2011.

Press, Gil. “A Very Short History of Data Science.” Forbes. Forbes Magazine, October 15, 2014. Accessed October 4. https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/.

Provost, Foster, and Tom Fawcett. 2013. “Data Science and Its Relationship to Big Data and Data-Driven Decision Making.” *Big Data*, 1(1): 51–59. https://doi.org/10.1089/big.2013.1508.

Stanton, Wilbur W., and Angela D'auria Stanton. 2020. “Helping Business Students Acquire the Skills Needed for a Career in Analytics: A Comprehensive Industry Assessment of Entry-Level Requirements.” *Decision Sciences Journal of Innovative Education* 18(1): 138–65. https://doi.org/10.1111/dsji.12199.

Storey, Veda C., and Il-Yeol Song. 2017. “Big Data Technologies and Management: What Conceptual Modeling Can Do.” *Data & Knowledge Engineering* 108: 50–67. https://doi.org/10.1016/j.datak.2017.01.001.

Thomas H. Davenport and D.J. Patil, and Andrew McAfee and Erik Brynjolfsson. “Data Scientist: The Sexiest Job of the 21st Century.” Harvard Business Review, May 26, 2017. Accessed October 7. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

Tomé, João. “‘Faltam a Portugal Alguns Milhares De Cientistas De Dados.’” Dinheiro Vivo, August 21, 2020. https://www.dinheirovivo.pt/empresas/tecnologia/faltam-a-portugal-alguns-milhares-de-cientistas-de-dados-12894459.html. Consulted on December 13, 2020

Van der Aalst, Will.M.P. 2014. "Data Scientist: The Engineer of the Future". In: *Enterprise Interoperability VI*, Mertins K., Bénaben F., Poler R., Bourrières JP. (eds). Proceedings of the I-ESA Conferences, 7: 13-26. Springer, Cham. https://doi.org/10.1007/978-3-319-04948-9_2

Vicario G, Coleman S. 2020. "A review of data science in business and industry and a future view". *Appl Stochastic Models Bus Ind.* 36: 6–18. https://doi.org/10.1002/asmb.2488

Volpe, Maddalena Della, and Francesca Esposito.2019. "The Data Scientist Job in Italy: What Companies Require." *Advances on P2P, Parallel, Grid, Cloud and Internet Computing Lecture Notes in Networks and Systems*. 894–903. https://doi.org/10.1007/978-3-030-33509-0_84.

Warden, Peter. "Why the Term 'Data Science' Is Flawed but Useful." OReilly Radar, May 9, 2011. Accessed October 7, 2020. http://radar.oreilly.com/2011/05/data-science-terminology.html.

Y. Demchenko et al., 2016 "EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry," 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Luxembourg City: 620-626. https://doi: 10.1109/CloudCom.2016.0107.

Y. Gil. 2014. "Teaching Parallelism without Programming: A Data Science Curriculum for Non-CS Students," *Workshop on Education for High Performance Computing,* New Orleans, LA. 42-48. doi: 10.1109/EduHPC.2014.12.

Yin, Shen, and Okyay Kaynak. 2015. "Big Data for Modern Industry: Challenges and Trends [Point of View]." *Proceedings of the IEEE* 103(2): 143–46. https://doi.org/10.1109/jproc.2015.2388958.

Zikopoulos, Paul; deRoos, Dirk; Parasuraman, Krishnan; Deutsch, Thomas; Giles, James; and Corrigan, David. 2012. *Harness the Power of Big Data the IBM Big Data Platform*. McGraw-Hill Osborne Media

7. Appendices

Appendix 1 (1. Introduction)

Table 1: *Data Science programs in Portuguese Business Schools, including vacancies and year of launch when available on the schools' website*

University	Financial Times ranking	Programs	Launch year
NOVA SBE	26 (2020)	• MSc in Business Analytics • Executive Education Data Science for managers	• 2020 • -
ISTCE	66 (2019)	• BSc in Data Science • MSc in Data Science (70 vacancies)	• 2019
CATÓLICA LISBON	31 (2020)	• MSc in Business Analytics • Executive Education: Business Analytics: Data Science and Big Data	• 2021 • 2016
CATÓLICA PORTO	79 (2020)	• Executive Education: Business Analytics	• -
ISEG	-	• MSc in Data Analytics for Business (35 vacancies) • Post-grad in Data Science Business Analytics	• 2017

Appendix 2 - Revision of Data Science Master's programs in Portuguese Business Schools (1. Introduction)

Reviewing four of the offered master's programs in data science and related areas from business schools in Portugal, revealed that overall the programs were built to respond to the rapidly growing demand for specialists in this field, with the ability to extract value from the huge volume of data that exists, leveraging it to identify (managerial) problems and formulate solutions. Preferably, candidates are required to hold a first degree in Sciences, Technology, Engineering, Mathematics, Economics, Finance or Management.

Looking at the four curriculums (below, Table 2 details each program) in general, they seem to be designed following an interdisciplinary and a multidisciplinary approach:

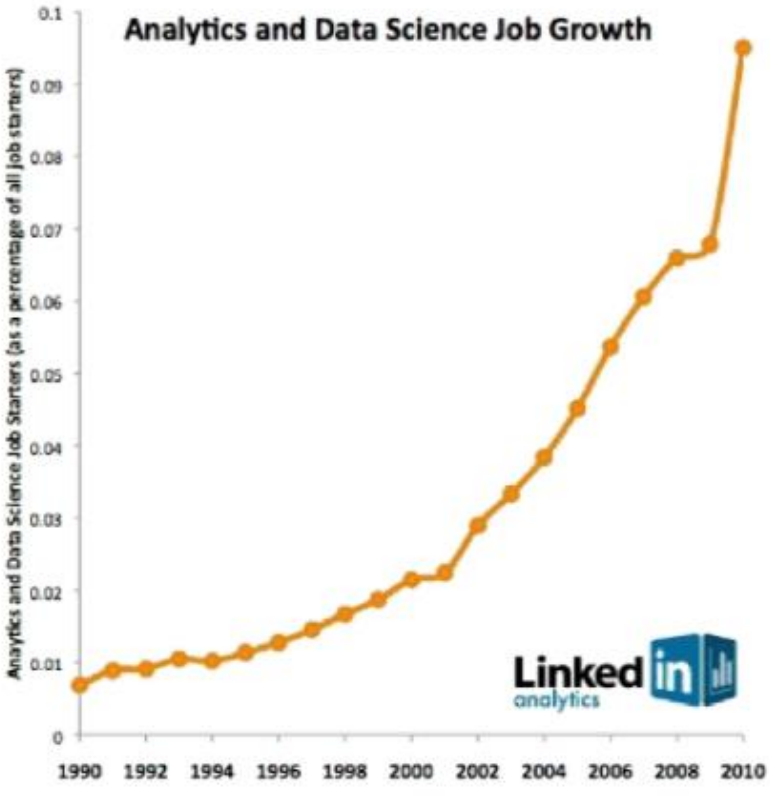
- The development of solid and current technical skills, as well as advanced quantitative knowledge. Examples of mandatory courses on this topic are – Advanced Data Analytics, Big Data Management, Machine Learning, Advanced Statistical Methods, Forecasting Methods, Programming for Data Science, Data Visualization, Predictive Analytics.
- Training in Business-related skills and soft skills such as leadership, communication and critical thinking competencies. Some examples include, Modelling Business Decisions, Managing People, Decision Making and Optimization.
- Courses dedicated to the ethics related to working with data, such as Data Ecosystems and Governance in Organizations, Cyber Law, Data Privacy and Ethics
- Focus on the development of research skills (Research and Development Seminars at ISEG) and in the real-life application (Nova SBE's program including project-based learning, which addresses real-world data challenges in the development of a project in partnership with organizations).

In spite of this, it is worth noting that the courses covering hard skills represent the majority of the curriculums in all programs.

Table 2: *Summary of the Master's programs in Portuguese Business Schools, including program's focus and curriculum*

UNIVERSITY	PROGRAM	AIM / FOCUS	CURRICULUM
NOVA SBE	MSc in Business Analytics	"Aims at the development of solid technical, organizational, leadership, critical thinking and communication competencies for future managers and leaders of organizations (public or private, for-profit or not-for-profit) in a context increasingly dependent on algorithms and hybrid (human-machine) systems for decision making. Focused on the education of translators, i.e. people who understand organizations and the managerial problems they face and know how to use technology and leverage data to solve them."	**Mandatory courses** Advanced Data Analysis Digital Markets Data Curation Modelling Business Decisions Data Visualization Data Ecosystems and Governance in Organizations Machine Learning **Electives** Advanced Programming for Data Big Data Analysis Digital Marketing Digital Strategy and Transformation E-commerce Empirical Methods for Finance Financial Econometrics Fintech Ventures Innovation Management Marketing Analytics Modeling Business Decisions Open Innovation Project Scoping (mandatory to apply for Project-Based Learning) Product Design and Development Technology Strategy WEB and Cloud Computing
ISCTE	MSc in Data Science	"Aims to fill a gap that has been growing increasingly in the Portuguese business fabric (eg, finance, public policy, insurance, fisheries and agriculture, energy, telecommunications, tourism, health). with the challenges inherent in extracting knowledge and value from the huge wealth of data that exists, both in business and on the Internet."	Big Data Management Data Science Methodologies and Technologies Pattern Recognition Prediction Models Big Data Processing and Modeling Text Mining for Data Science Time Series Analysis and Forecasting Applied Management Control Systems Ciberlaw Project Design for Data Science
CATÓLICA LISBON	MSc in Business Analytics	"Analytics provides students with specialized and comprehensive training in data science and business analytics. Students will learn how to structure data analytics problems and use data analytics tools and techniques to design solutions for real business challenges. Beyond the acquisition of technical knowledge, the program also focuses on developing the critical skills necessary for successful leadership in organizations."	**Mandatory courses** Introduction to Programming Foundations of Statistics with Applications in R Database Management Causality and Randomized Experiments Data Visualization Time Series Econometrics Predictive Analytics Advanced Topics in Predictive Analytics Data Privacy and Ethics Managing People **Electives** Marketing Analytics Operations Analytics Accounting Economics of Business and Markets Strategy and Change Finance Mathematical Programming and Simulation Big Data Technologies Decision Theory
ISEG	MSc in Data Analytics for Business	"Aims to prepare data scientists, managers, economists, statisticians, and information management professionals."	Business Intelligence and Decision Support Data Platforms for Analytics Decision Making and Optimization Enterprise Analytics Programming Foundations in Python Statistical Methods and Visualization Advanced Statistical Methods Big Data Tools and Analytics Forecasting Methods Machine Learning and Data Mining Programming for Data Science Research & Developement Seminars

Appendix 3 – Job Growth in Analytics and Data Science according to LinkedIn Analytics (2. Literature Review Section 2.3. How companies' needs have changed)



Appendix 3 - Analytics and Data Science Job Growth up to 2010. Source: LinkedIn Analytics

Appendix 4: Interview Guide (3. Methodology)

WARM UP- Hello! My name is Carolina Monteiro and I am a Master's student at Nova SBE currently researching the data science job and in particular what is expected from recent graduates coming into this position. For this end, I am conducting interviews with managers from different companies to understand what is their perception On the topic. As so, your input will be extremely helpful for my analysis and I want to thank you in advance for agreeing to participate. These will be open ended questions, so feel free to answer however you'd like, also there are no right or wrong answers.

For the purpose of analyzing these interviews later, I will record them in audio format and later will transcribe them. All data will be anonymized, meaning no one will be able to track any answers to you. That being said, I'll ask one introductory question to register your consent: Do you consent to recording this interview?

Thank you, let's start.

QUESTIONS

GENERAL

0. Can you tell me your current role and how long you've been in it?

1. How important is the data science function in your organization? Is it a role, a team, a department?

2. Do you think this function will continue to become more relevant in the future?

3. Are there data science needs in your company that have yet to be met?

HIRING

4. As a recruiter, or participating in recruitment, have you found it challenging to hire data scientists?

4.1. If so, what would you say are the biggest constraints?

COMPETENCIES

5. In general, how satisfied are employers with the skills of data science graduates?

6. Which skills do you consider important when hiring a new data science graduate?

6.1 And which important skills are recent graduates falling short in?

7. Have you found it necessary to provide additional training in any of these competencies?

8. In later years, have you noticed any improvement in the skillsets of recent graduates?

ROLE OF (BUSINESS) SCHOOLS

9. What study areas do most talents come from? What % comes from a business background?

10. What do you think about the integration of data science in business schools?

11. What do you think a business school could do to train better data scientists?

Appendix 5 – Content Grid Excerpt ([3. Methodology](#))

10. What do you think about the integration of data science in business schools?					
I1	I2	I3	I4	I5	I6
Yes, definitely. Business schools are the owners of the business knowledge. Right? So I think, well, this is my opinion, I'm not saying is based on any kind of data, just based on my experience and the knowledge I have. But, I think I wouldn't say something stupid when I say that **in the future any manager needs to have at least some quite mid-level knowledge of data science**. Because in a data driven world, where we are already a little bit but not as far as we'd like, but I think that we are already a little bit, but in the future it would be even more and more**. Most of decisions will be taken based on data, and if you don't know how to deal with your data is not the basic excel from the 90's that will help you. So in the future, of course business schools have a great role on this topic.** Because if they assume data science and data as something all managers need to have, either they're learning how to be managers in retail, finance or whatever other main management topics, well these topics will be learned by all of them and these gaps will be bridged. So definitely, without the business schools this gaps will not be bridged. Yes, I've hired people coming from business schools. Hmm, but the last one that I remember hiring came from a business school but it was for a sales position. But directly for the data science team, I hired someone that came from a business school but then he did a master in data science. So he had both educations. BS + DS. Usually, at least at the moment**, if I want to hire a data science manager, I would probably look more into graduates from a business school. But if I'm looking for a pure data scientist, someone to really develop my models, a business schools will not be my main place. I would prefer someone from a traditional course like math, physics, engineering, information management**. But this is not a rule, it's something that so far we have been following because of what we've described. But I think that this can definitely change.	Yeah, I think they should have done it yesterday. No. To be honest I don't think so. Because data science is not about being a pure tech developer, it is about- you don't need to know how a computer works to do good data science, you just need to be really keen to learn and always learning all the time, because it is always evolving and yeah, you just need to have the right profile. Because I don't think having an engineering degree brings a lot or a big advantage for this specific position.	I don't understand how today you have a business school that doesn't have digital competencies and digital programs. As I said in the beginning, it is not the economy that is digital, we have a digital economy. Hmm, so I don't understand how you can have master programs on management, marketing, commercial. Because all of them always have a component of digital expertise that people need to have. And to be frank with you, if you're speaking about business schools where theoretically you're training and forming the leaders of the future, without any doubts that you need to put on the basic program or core curriculum some competences on digital skills and precisely to have some trainings in that part. To be frank, I think we are already lagging behind. I think universities should already be in that level. I think universities take too long of a time to adapt to the market needs and if we're speaking about bs, bs should think ahead, should think about the Lt, should think about the strategic analysis and how the economy will change in the future. So today I'll tell you that I feel that bs they should be adapting to skills that the market will need in 2030. We are speaking today about schools adapting and bringing digital skills to today's business practices which is something that the market is already lagging and will take considerable efforts to adapt.	I would say there are 2 things almost from a business perspective. **1st) make sure obviously to focus on the commercial side of it and real life applications**. Specially from me, I come from a theoretical background try to understand what is the application of different techniques in the real work I think that's quite a leap and something that you obviously learn with experience but that if maybe there was some understanding or some knowledge prior to starting and getting out into the job /real world that would be beneficial. 2nd) thing, idk how it is in the Nova business school, but I think here, idk what are the requirements for getting into DS for business Masters so making sure the **technical knowledge is** also there. I think that is quite important for students that have quite different backgrounds. 'Cause if you have some who comes from a stat undergrad and then they're doing a DS for business masters then you can almost assume they'd have the right technical background. If you have someone from eng they will know a lot about maths and numbers but they might not know the stats. The one that has the Stats degree, the engineering? Because for me its easier to teach commercial and communication skills on the job, than to teach technical skills.	Yes. Definitely. It's like a product manager role as well, you can come from product management as a developer as a manger or as a designer as a manager. In this data science positon any base of academics both development or business can provide its own synergies. Like for example for a startup, I can see maybe the other way around being a lot more useful. Being a manager role, that then fortified their developer skills to fill in that role. Hmm, I can also see it the other way around, there's a lot of fields that maybe require a bunch of technical know-how and maybe not so much managerial know how. So I can see a balance between both. Yeah, yeah I can see it happening and work well. You can also say the same for product management, due to poor product management you get to a point where some of this ds do not meet the expectations because they don't know what to ask for. Maybe product managers what to get answers without looking for the right questions to their teams, that includes data scientists.	Definitely, right? So.. hm.. that's pretty much what -, I mean if you think of me, with an economics, business-y background. I mean, I know I would never be able to compete with the astrophysics to understand things that I can't even grasp, that I would never grasp. But I'm bridging between the MBA (btw the CEO and the highly technical person. Because someone needs to do the translation, and the stakeholder management and spotting of opportunities. So yeah, there's definitely the case of more data scientists with business acumen that would be useful. - Hmm, that's a good question. To be honest with you I haven't had many applicants who come straight from a business school for data science. What always helps is if they have any type of internships experience in companies. We've hired many people coming to bootcamps, so bootcamps are like 8 weeks, or 2 months-long projects in a company and then you can talk about those projects in interviews. If you manage to place people in companies maybe justeats, who knows, we can always use more hands. So if you manage to place people in internships that is certainly useful. I have seen this before. I have people who hmm coming straight from school, it was useful to see they had done their master thesis in a company, that

Appendix 6 – Questionnaire ([3. Methodology](#))

Q0

Hello! My name is Carolina Monteiro, I am a Master student at Nova SBE researching the data science job and in particular what is expected from recent graduates coming into this position. For this end, I'm doing this questionnaire to get data scientists' perspective on this topic.

To participate in this study, you need to be older than 18 years old. Your participation is voluntary and all answers will be anonymized, confidential and will only be used for the purpose of this research. If you have any further questions or comments, please contact me at 29216@novasbe.pt.

- I confirm that I am older than 18 years old and want to participate in this study.
- I do not want to participate or do not belong to the defined age slot.

Condition: I do not want to participat... Is Selected. Skip To: End of Survey.

Q1

Are you currently working or have you worked in data science or in related areas (AI, Analytics, ML)?

- Yes
- No

Condition: No Is Selected. Skip To: End of Survey.

Q2

For how long have you been working in the area?

- <= 1 year
- 2 years
- 3 years
- 4 years
- => 5 years

Q3

What level of seniority is your current job?

- Entry-level
- Junior
- Senior
- Director
- Other (Please specify)

Q4 How many people report to you?

- ○ None
- ○ 1 to 2
- ○ 2 to 5
- ○ More than 5

Q5 What is your first academic degree?

- ○ Computer Science
- ○ Engineering
- ○ Mathematics/ Statistics/ Physics
- ○ Information Management/ Information Systems
- ○ Business/ Economics/ Finance
- ○ Other (Please specify)

Q6 Which best describes the additional training you took to work in this field? (Do not include any in-job training)

- ○ None
- ○ I did not need additional training
- ○ I took courses in specific topics, but not post graduate or masters program
- ○ I completed a post graduate or masters program

Q7 In what industry do you currently work in?

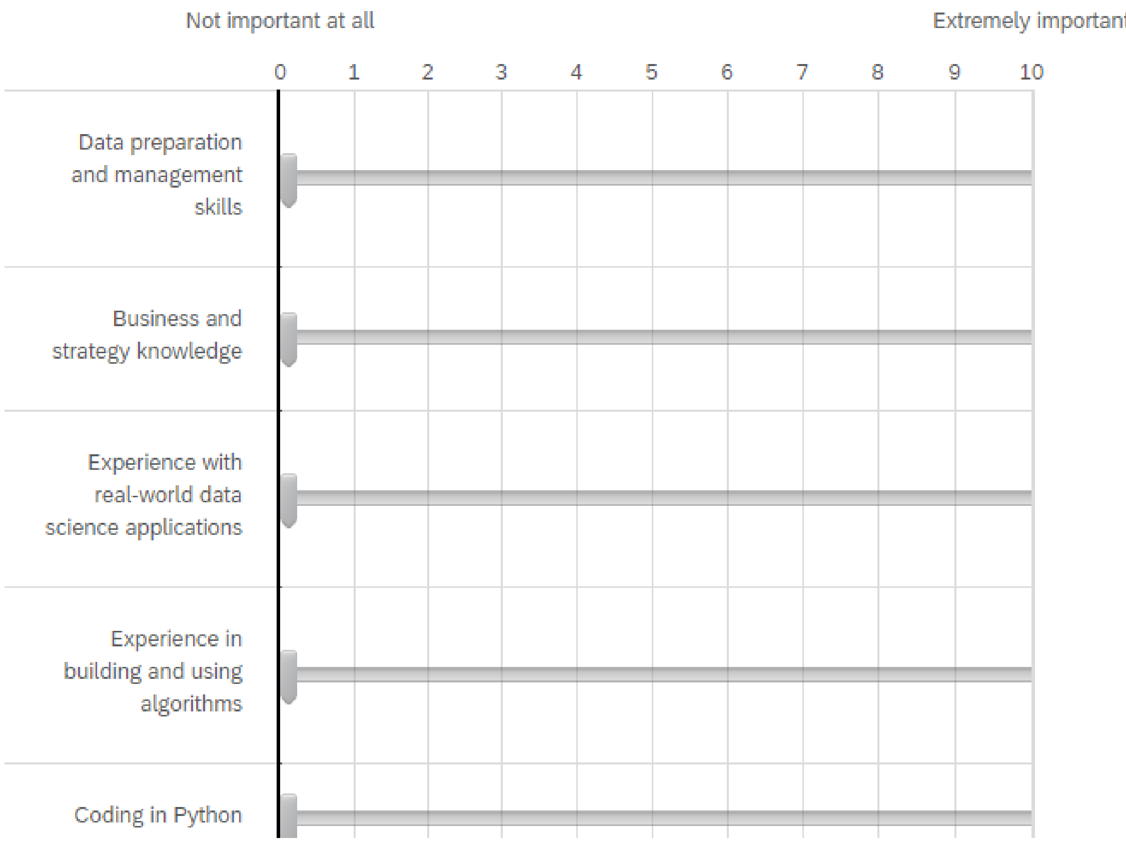
- ○ Retail
- ○ IT
- ○ Telecommunications
- ○ Consulting
- ○ Banking
- ○ Insurance
- ○ Health Care
- ○ Other (Please specify)

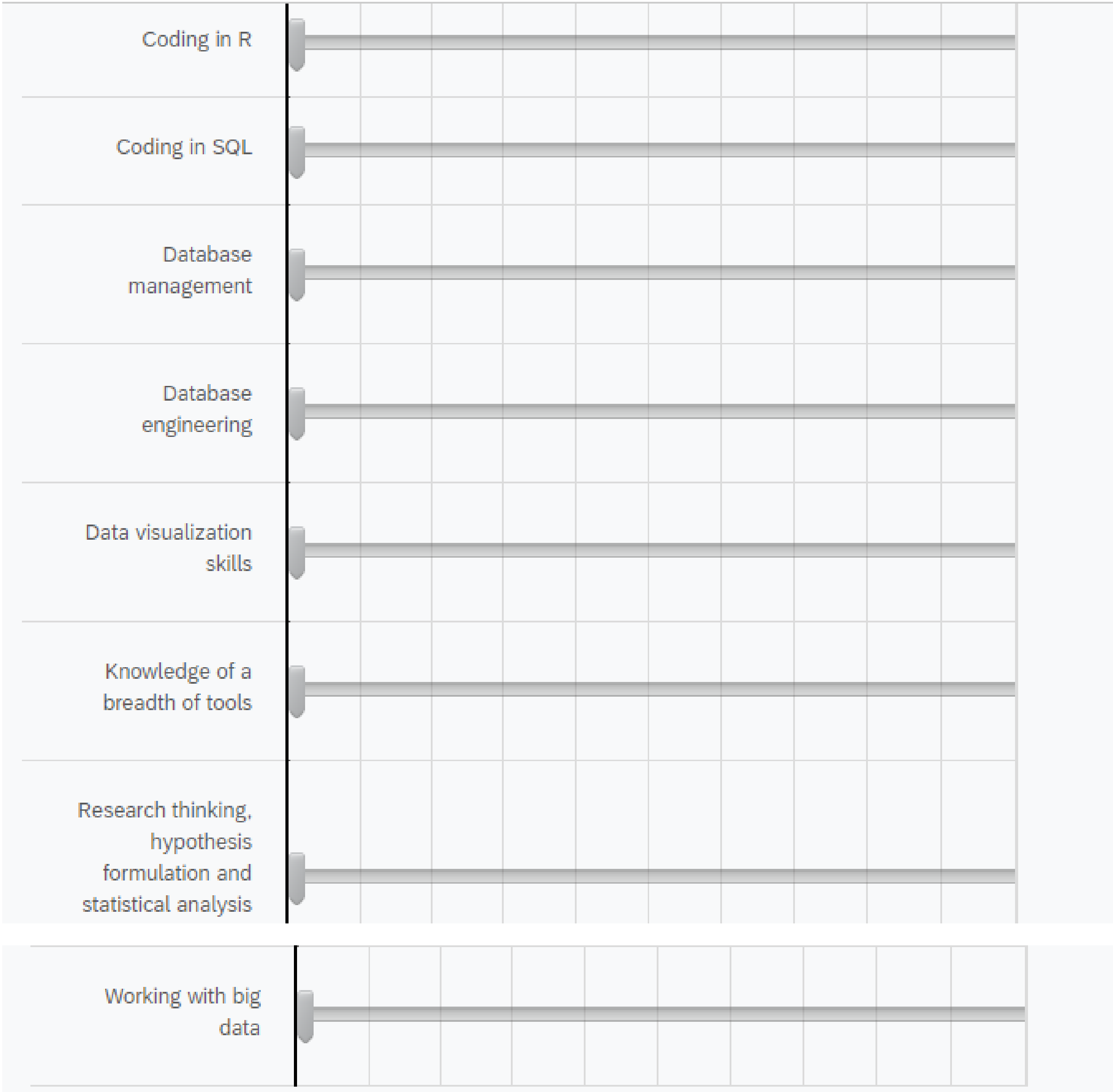
For the next question, please state on a scale from 0-10, in which 0 represents "Completely disagree" and 10 represents "Completely agree", to what extent you agree with the following statements:

	Completely disagree										Completely agree
	0	1	2	3	4	5	6	7	8	9	10
Data Science is an important role/team/department for my company/organization.											
Data Science will increasingly gain relevance within companies/organizations in the years to come.											
The most important skills for a data scientist are the data ones (eg.: Data Engineering, Data Analysis including Data visualization, Modelling).											
Having well established data skills (eg.: Data Engineering, Data Analysis including Data visualization, Modelling) is enough for a data scientist to be successful.											
Business skills (eg.: Marketing concepts, Communication, Customer experience) are equally as important as data-related skills for a successful data scientist.											

Q9

Please rank the importance, on a scale from 0-10, in which 0 represents "Not important at all" and 10 represents "Extremely important", of each of the next **skills** for a data scientist:




Coding in R
Coding in SQL
Database management
Database engineering
Data visualization skills
Knowledge of a breadth of tools
Research thinking, hypothesis formulation and statistical analysis
Working with big data

Q10

Please rank the importance, on a scale from 0-10, in which 0 represents "Not important at all" and 10 represents "Extremely important", of each of the next **soft skills** for a data scientist:

Not important at all — Completely important

0 1 2 3 4 5 6 7 8 9 10

Resilience

Analytical thinking

Critical thinking

Communication

Interpersonal relationships

Leadership

Problem solving

Teamwork

Creativity

Time Management

Ethical behaviour

Q11

From the next skills/competencies please select those that you wish you had learned more in depth.

- [ ] Data preparation and management skills
- [ ] Business and strategy knowledge
- [ ] Experience with real-world data science applications
- [ ] Experience in building and using algorithms
- [ ] Coding in Python
- [ ] Coding in R
- [ ] Coding in SQL
- [ ] Database management
- [ ] Database engineering
- [ ] Knowledge of a breadth of tools
- [ ] Research thinking, hypothesis formulation and statistical analysis methodologies
- [ ] Working with big data
- [ ] Other (Please specify)

Q12

From the next skills/competencies please select those you think most companies are struggling with the most.

- [ ] Data preparation and management skills
- [ ] Business and strategy knowledge
- [ ] Experience with real-world data science applications
- [ ] Experience in building and using algorithms
- [ ] Coding in Python
- [ ] Coding in R
- [ ] Coding in SQL
- [ ] Database management
- [ ] Database engineering
- [ ] Knowledge of a breadth of tools
- [ ] Research thinking, hypothesis formulation and statistical analysis methodologies
- [ ] Working with big data
- [ ] Other (Please specify)

Display This Question:

If Which best describes the additional training you took to work in this field? (Do not include any... I completed a post graduate or masters program Is Selected

Q13

For the next question, please state, on a scale from 0-10, in which 0 represents "Completely disagree" and 10 represents "Completely agree", to what extent you agree with the following statements related to how well university prepared you for the data job:

Completely disagree / Completely agree

0 1 2 3 4 5 6 7 8 9 10

University prepared me well enough for the data job in terms of data related skills (eg.: Data Engineering, Data Analysis including Data visualization, Modelling).

Univeristy prepared me well enough for the data job in terms of business skills (eg.: Marketing concepts, Communication, Customer experience).

Overall university prepared me well enough for the data job.

The university curriculum addressed well enough the data needs that my company/organization experiences.

Display This Question:

If Which best describes the additional training you took to work in this field? (Do not include any... I completed a post graduate or masters program Is Selected

Q14

In which areas did you look for additional training?

- None
- Business skills (eg.: Marketing concepts, Communication, Customer experience)
- Data related skills (eg.: Data Engineering, Data Analysis including Data visualization, Modelling)
- Both

Q15

What gender do you identify as?

- Male
- Female
- Prefer not to say
- Other

Q16

What is your age?

Under 18
18 - 24
25 - 34
35 - 44
45 - 54
55 - 64

Q17

What is your nationality?

Afghan

Q18

What is your highest level of education?

- High School
- Bachelor's degree
- Master's degree
- PhD or higher

Q19

Which of these describes your yearly gross income?

- Less than €10,000
- €10,000 - €19,999
- €20,000 - €29,999
- €30,000 - €39,999
- €40,000 - €49,999
- €50,000 - €59,999
- €60,000 - €69,999
- €70,000 - €79,999
- €80,000 - €89,999
- €90,000 - €99,999

Appendix 7 - (4. Analysis and Discussion, 4.1. Qualitative Findings)

Table 3: *Interviewees profiles in terms of position, company, industry and location*

ID	Position	Company	Type	Location
I1	Lead Data Scientist	BiLD Analytics	Analytics Consulting start up	Portugal based
I2	Business Intelligence and Analytics manager	JustPark	Start up	UK based
I3	Digital and Innovation advisor	European Commission	-	-
I4	Director of Data Science	Wyoming Interactive	Digital Consulting agency	UK based
I5	Product Manager	Impossible/Bondtouch	Consulting start up/Wearable company	Portugal based
I6	Senior Data Science manager	JustEats	Start up	International
I7	Lead Data Scientist	Everies	Business and IT Consulting	Portugal based
I8	Director of Advanced Analytics	Fidelidade	Insurance	Portugal based
I9	Lead Data Scientist	Amazon Web Services	Cloud technology services	International
I10	Lead Data Scientist	Merkle	Marketing Consulting agency	International

Appendix 8 – (4. Analysis and Discussion, 4.2. Quantitative Findings)

Table 4: *Summary of the results concerning statements related to data science in organizations and skills necessary to succeed. Analysis of variance between the statements and gender, significant relationships are marked with * at a confidence level of 95%. Correlation analysis between statements.*

Item	Mean	Std. Deviation	Gender	Gender (Mean, Std. Dev.)	Data Science will increasingly gain relevance within companies/organizations in the years to come.
Data science is an important role/team/department for my company/organization.	7.69	2.48	*	Female (6.40; 2.91) Male (8.50; 1.79)	r = 0.535**
Data Science will increasingly gain relevance within companies/organizations in the years to come.	8.69	1.31			
The most important skills for a data scientist are the data ones (e.g.: Data Engineering, Data Analysis including Data Visualization, Modelling)	6.57	1.56			
Having well established data skills (e.g.: Data Engineering, Data Analysis including Data Visualization, Modelling) is enough for a data scientist to be successful.	4.72	1.98			
Business skills (e.g.: Marketing concepts, Communication, Customer experience) are equally as important as data-related skills for a successful data scientist	7.18	2.17			

***Appendix 9* – (4. Analysis and Discussion, 4.2. Quantitative Findings)**

Table 5*:* *Skills ranked by importance for a data scientist, on a scale from 0 (not important at all) to 10 (extremely important). Analysis of variance between skills and academic background and industry, significant relationships are marked with * at a confidence level of 95%.*

SKILL	Mean	Std. deviation	Academic Background	Academic Background (mean, std dev.)	Industry	Industry (mean, std dev.)
Data preparation and management skills	8.45	1.39				
Data Visualization skills	8.15	1.41			*	Retail (4.00) IT (8.73, 1.10) Telecomunications (8.00) Consulting (7.83, 0.40) Banking (8.37, 1.30) Insurance (7.00) Other (8.00, 2.16)
Experience with real world data science applications	8.12	1.47				
Coding in Python	7.78	1.74				
Business and Strategy knowledge	7.68	1.30				
Coding in SQL	7.62	1.68				
Experience in building and using algorithms	7.54	1.45	*	• Business/Economics/Finance (7.37, 1.30) • Infomartion management/Information Systems (7.00, 1.15) • Computer science, Engineering, Math/Statistics/Physics (10.9, 8.10)		
Working with big data	7.47	1.96				
Research thinking, hypothesis formulation and statistical analysis methodologies	7.28	1.81				
Knowledge of a breadth of tools	7.03	1.59				
Database management	6.65	1.69				
Database engineering	6.15	1.86				
Coding in R	6.09	1.89				

Appendix 10 – (4. Analysis and Discussion, 4.2. Quantitative Findings)

Table 6*: Soft Skills ranked by importance for a data scientist, on a scale from 0 (not important at all) to 10 (extremely important)***.** *Analysis of variance between soft skills and gender and type of training***,** *significant relationships are marked with * at a confidence level of 95%.*

SKILL	Mean	Std. deviation	Gender	Gender (mean, std dev.)	Type of training	Type of training (mean, std dev.)
Problem Solving	9.00	0.95				
Analytical Thinking	8.96	0.99				
Critical Thinking	8.84	1.08			*	"I did not need additional training" (7.00, 0.94) "I completed a post graduate or master's program" (8.57, 1.01) "I took courses in specific topics, but not postgraduate or master's program" (9.30, 0.94)
Resilience	8.46	1.39				
Ethichal Behaviour	8.16	1.59				
Team Work	8.06	1.36				
Communicatio n	7.96	1.30	*	Female (7.20, 0.78) Male (8.31, 1.35)		
Team Work	8.06	1.36				
Time Management	7.56	1.41				
Creativity	7.40	1.66				
Leadership	6.31	1.67				

Appendix 11 – (4. Analysis and Discusion, 4.2. Quantitative Findings)

Table 7*: Skills recent data scientists wish they had learned more in depth, by order of most selected*

SKILL	FREQUENCY (%)
Experience with real world data science applications	52.9%
Coding in Python	50%
Experience in building and using algorithms	41.2%
Working with Big Data	41.2%
Business and Strategy knowledge	35.3%
Data preparation and management skills	23.5%
Coding in SQL	17.6%
Database management	17.6%
Database engineering	17.6%
Research thinking, hypothesis formulation and statistical analysis methodologies	17.6%
Coding in R	8.8%
Knowledge of a breadth of tools	8.8%

Appendix 12 – (4. Analysis and Discusion, 4.2. Quantitative Findings)

Table 8*: Skills recent data scientists think companies are struggling with the most, by order of most selected*

SKILL	FREQUENCY (%)
Experience with real world data science applications	52.9%
Working with Big Data	50%
Data preparation and management skills	47.1%
Experience in building and using algorithms	38.2%
Database management	26.5%
Database engineering	20.6%
Research thinking, hypothesis formulation and statistical analysis methodologies	20.6%
Coding in Python	17.6%
Coding in SQL	17.6%
Knowledge of a breadth of tools	14.7%
Coding in R	11.8%
Business and Strategy knowledge	8.8%

Appendix 13 – (4. Analysis and Discussion, 4.2. Quantitative Findings)

Table 9: *Participant's rating of their preparedness for the data science job*

Item	Mean	Std. Deviation
University prepared me well enough for the data job in terms of data related skills	7.08	1.56
University prepared me well enough for the data job in terms of business skills	5.57	2.06
Overall university prepared me well enough for the data job	6.47	1.98
The university curriculum addressed well enough the data needs that my company/organization experiences.	6.68	1.41