SKILL UP

LEARN MORE TO EARN MORE

DATA SCIENCE & BI

SALARY & SKILLS REPORT

[PACKT] PUBLISHING

www.packtpub.com/skillup

‘WHAT YOU NEED TO KNOW TO EARN MORE IN DATA SCIENCE AND BUSINESS INTELLIGENCE’

THE MOST COMPREHENSIVE GLOBAL IT SKILLS AND SALARY SURVEY EVER.

CONTENTS

The world of Data Science is rapidly growing, with data becoming increasingly vital to a huge range of organizations. Undoubtedly tied to the much-discussed rise of 'Big Data' over the past decade, data-oriented roles are today some of the most prominent technical roles in the economy.

This report, focusing on the data science respondents to Packt's Skill Up survey explores where data science is most valuable, what tools are being used, and what the trends and challenges will be in the future.

- Data science is immensely valuable to SMEs evidenced by their investment in young talent.
- Finance is still a lucrative sector for data science.
- R and Python are still neck and neck as the key data science languages.
- Distributed Computing and machine learning are still on the ascendancy.
- IOT is one of the hottest trends for data scientists that promises to bring new challenges and opportunities.
- Excel will never die!

The need to answer these questions led us to look at the community as a whole, and so we decided to launch our Skill Up campaign.

WHAT IS SKILL UP?

With our Skill Up survey we wanted to look at the tech community as a whole to identify upcoming trends over the next few years and share what you can do to ensure you get the most out of your career and skills. We divided our survey into 4 segments, Web Development & Design, Application Development, Security & System Administration, and Data Science & Business Intelligence, making this one of the most comprehensive surveys in recent years.

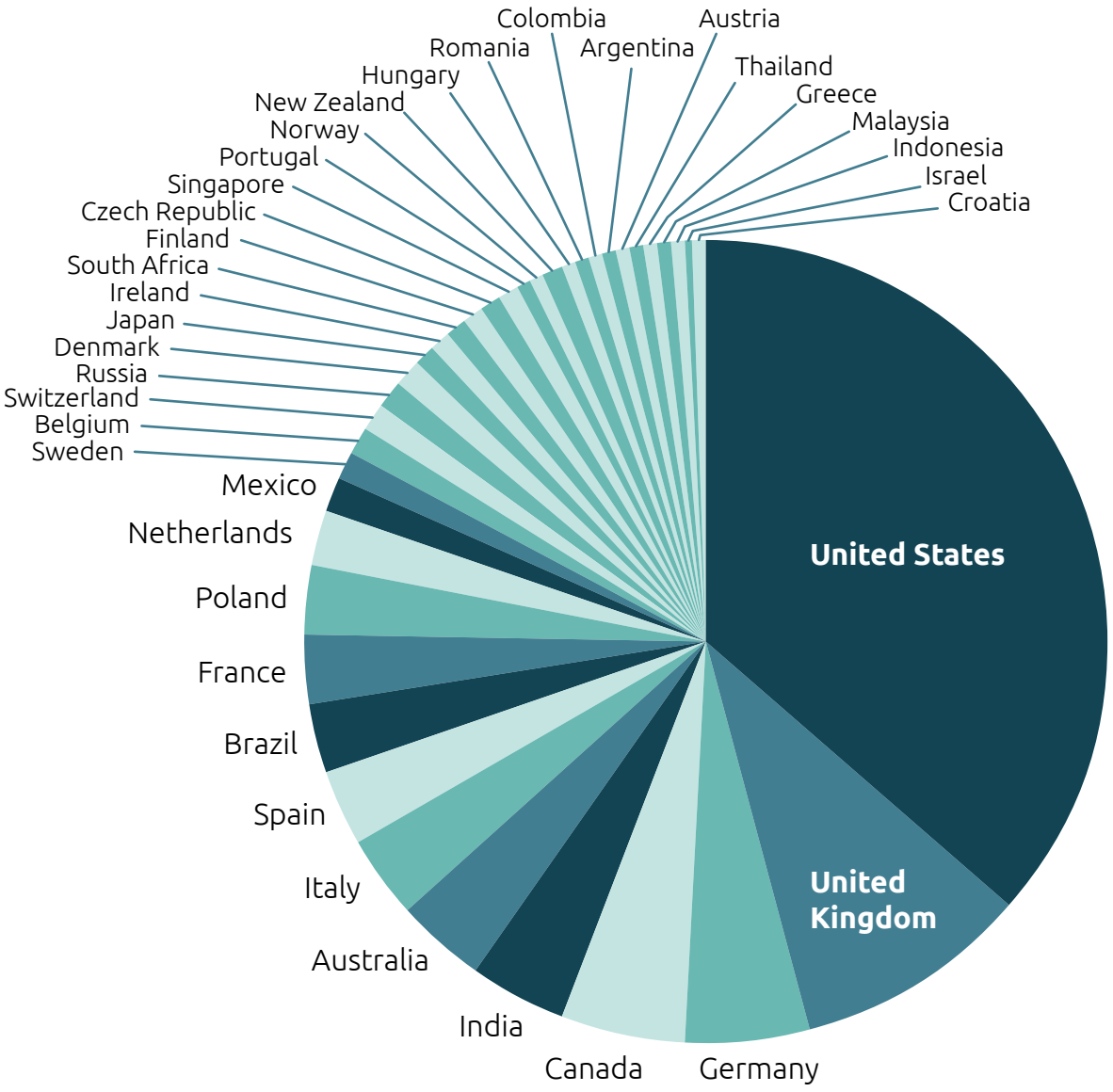
Specifically we asked:

- What skills lead to a higher salary?
- What skills/technologies are most highly valued by different industries?
- What cutting edge technologies are really worth you spending your time learning?

To get a better idea of the community's thoughts we asked you all to fill in our survey, the results of which you can find compiled here in this report, giving you the facts, the figures, and more importantly – the knowledge and skills you need to make the best career decisions.

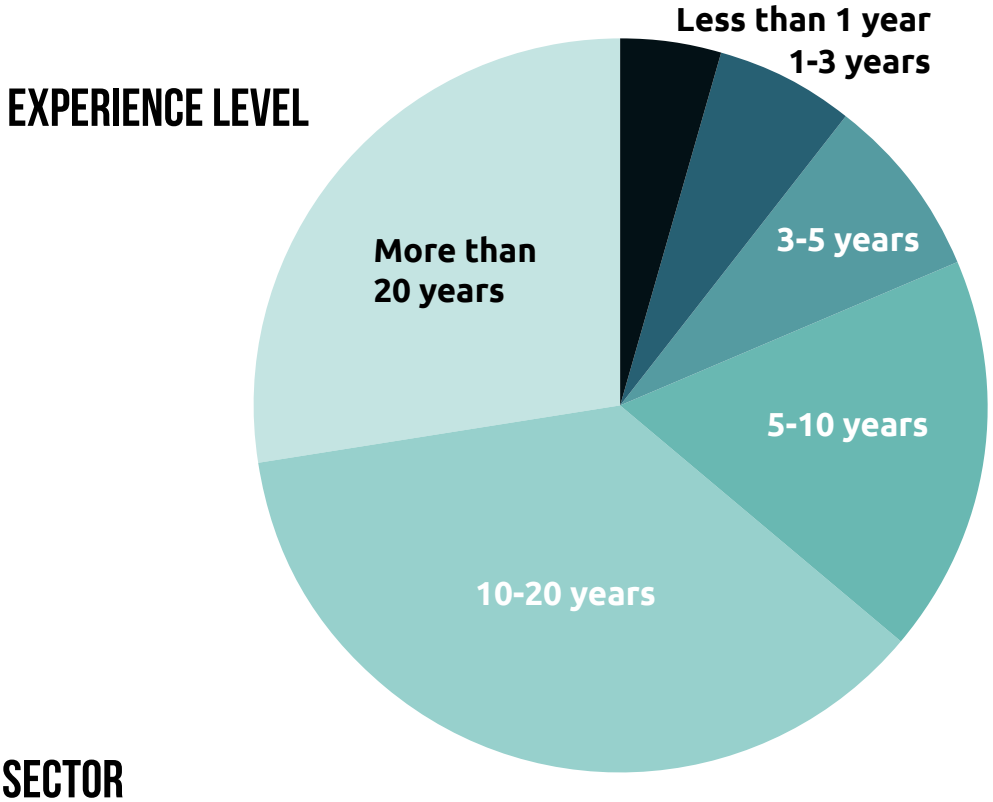
Let's look at the results in more detail.

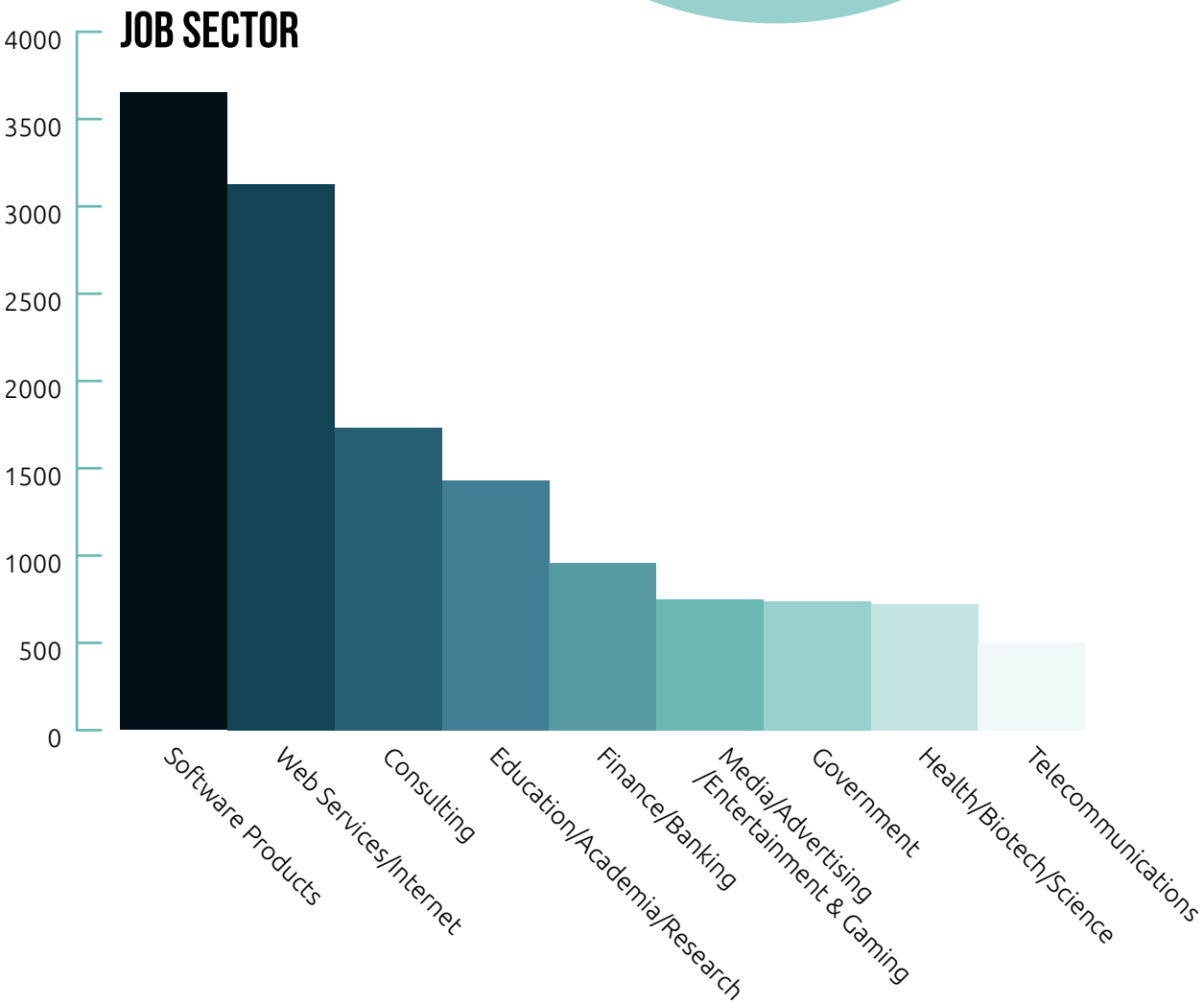
RESPONDENTS



RESPONDENTS BY COUNTRY

The Data Science stream, from which the data in this report is drawn, received over 3,800 responses from individuals with a wide range of experience levels, working in a diverse set of industries.



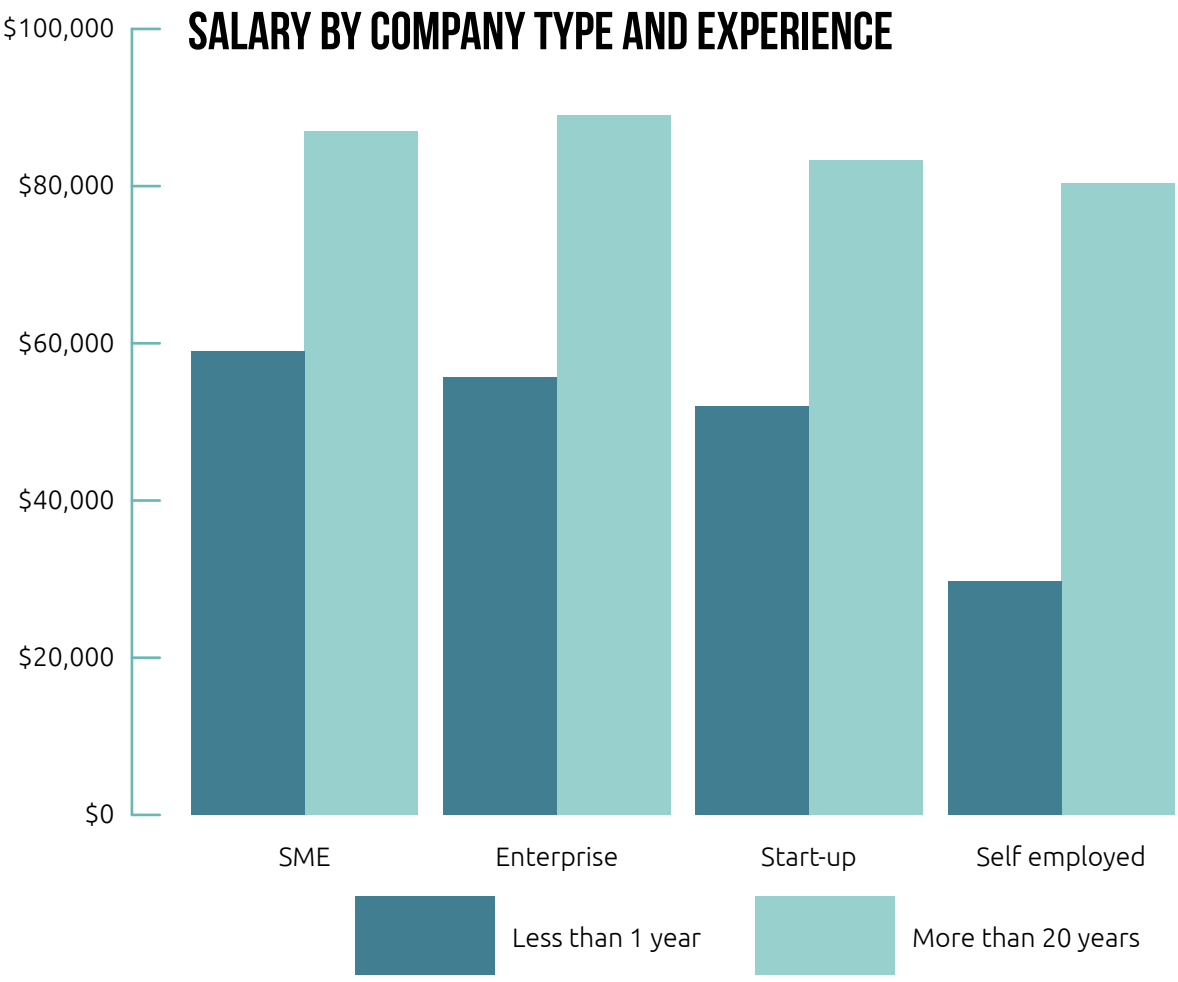


WHICH INDUSTRIES ARE BEST FOR DATA JOBS – AND WHICH ROLES ARE MOST VALUABLE?

There's a consensus that data-oriented roles are some of the most valuable around, especially in industries where data is so vital (there are very few industries where you could say it isn't). But there are trends that indicate where it's having the biggest impact, and what's going to become more and more important for anyone working in data.

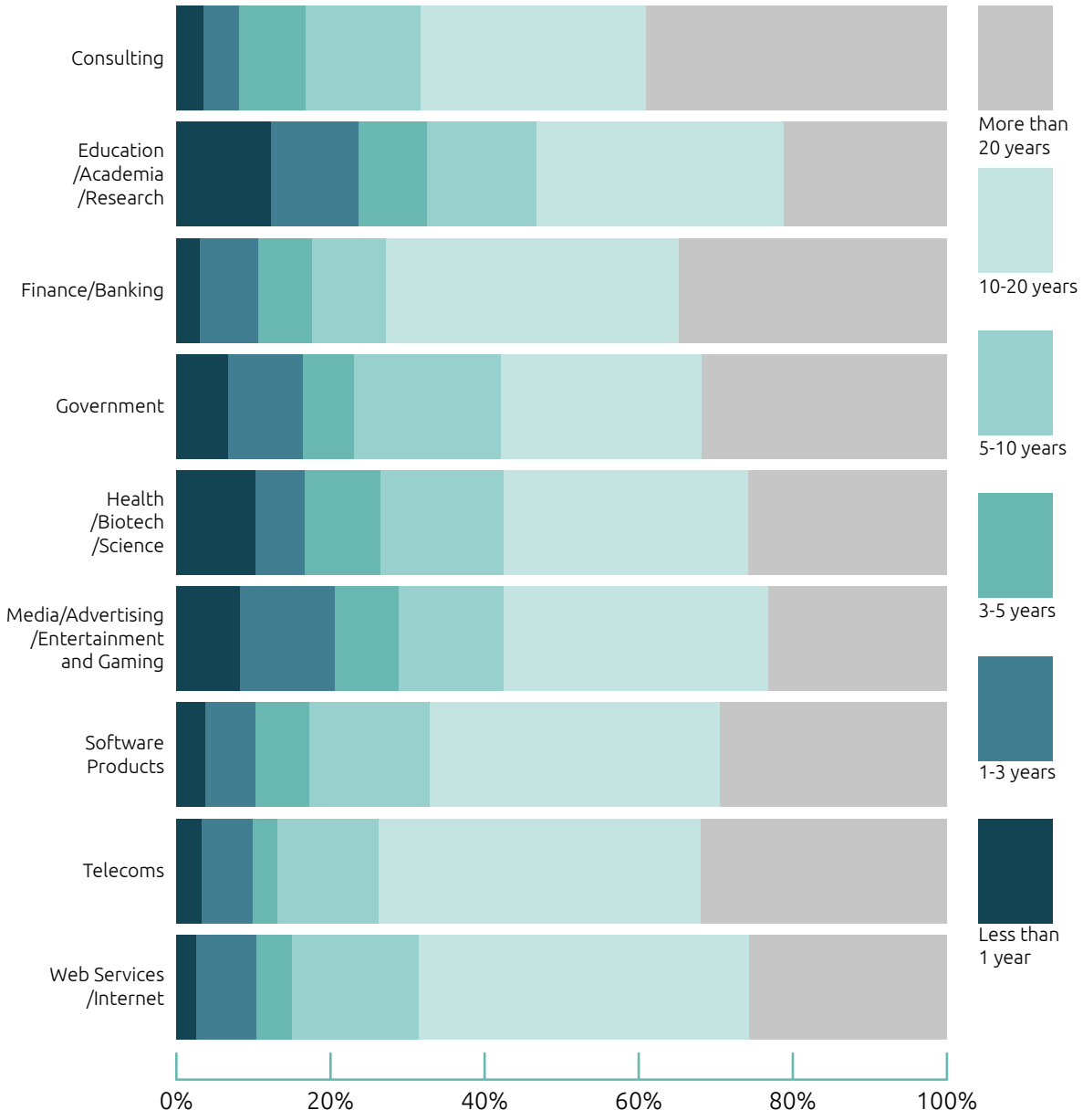
- SMEs pay inexperienced people the best.
- In terms of industries, Finance still offers the best pay for inexperienced employees.
- Data Architect is one of the most important roles in dynamic and fast-paced industries such as Media and Entertainment. This proves that the ability to build and implement business critical solutions is essential.

Our research shows that SMEs are a great place for inexperienced data scientists and analysts to begin their careers, offering a higher starting salary than Enterprise organizations. Even those working for start-ups earn only slightly less than Enterprise. With the opportunities available in small, rapid growth organizations, Start-ups are a great option for anyone ambitious and eager to prove themselves!



INDUSTRY BREAKDOWN

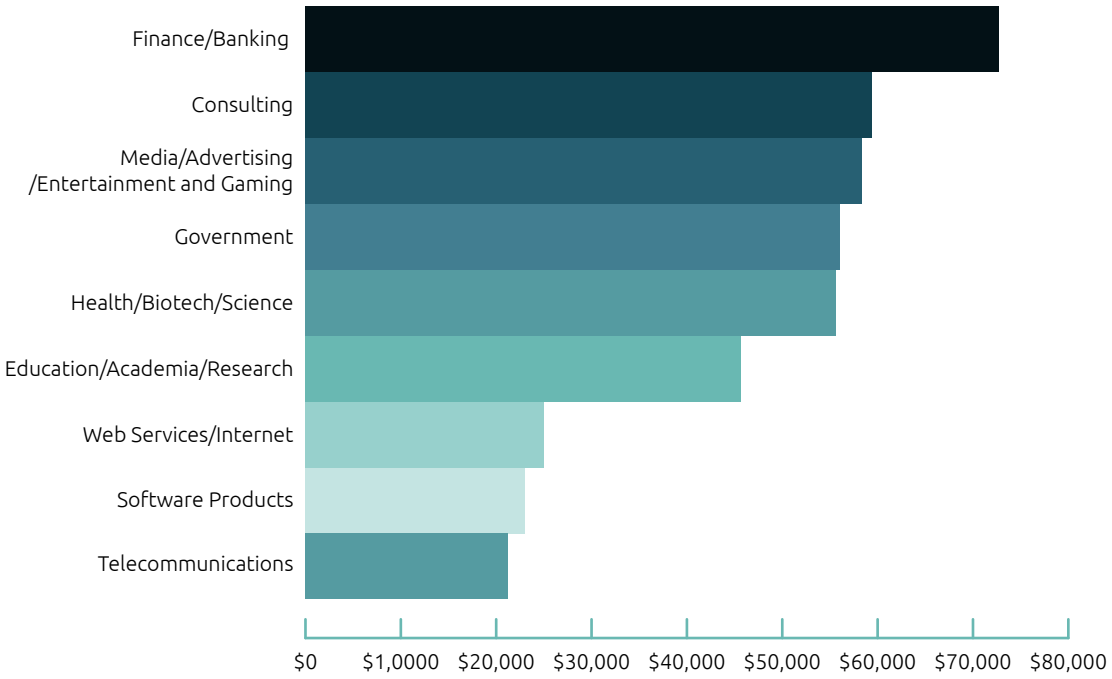
Our research also investigated how data science is faring in each sector. The responses provided a good indication of where data science is most critical.



The largest number of respondents with more than 20 years experience are working in consulting. Moreover, the commonality of this suggests that external expertise is something that is very much in demand.

Those respondents with least experience are predominantly working in Education/Academia and Research. However, as the graph overleaf shows, it does not appear to be one of the best paid industries.

WELL PAID INDUSTRIES FOR LESS EXPERIENCED PEOPLE



Finance and Banking comes out as the best paid industry according to our data. With the rise of algorithmic trading, and Big Data in general playing a large part in just about every component of Finance, there are huge opportunities for data fluent people to command high salaries, even without a great deal of experience.

Want to get to grips with quantitative finance? Pick up this bundle today.

- Mastering R for Quantitative Finance
- Introduction to R for Quantitative Finance
- Python for Finance
- Mastering Python for Finance
- Advanced Quantitative Finance with C++

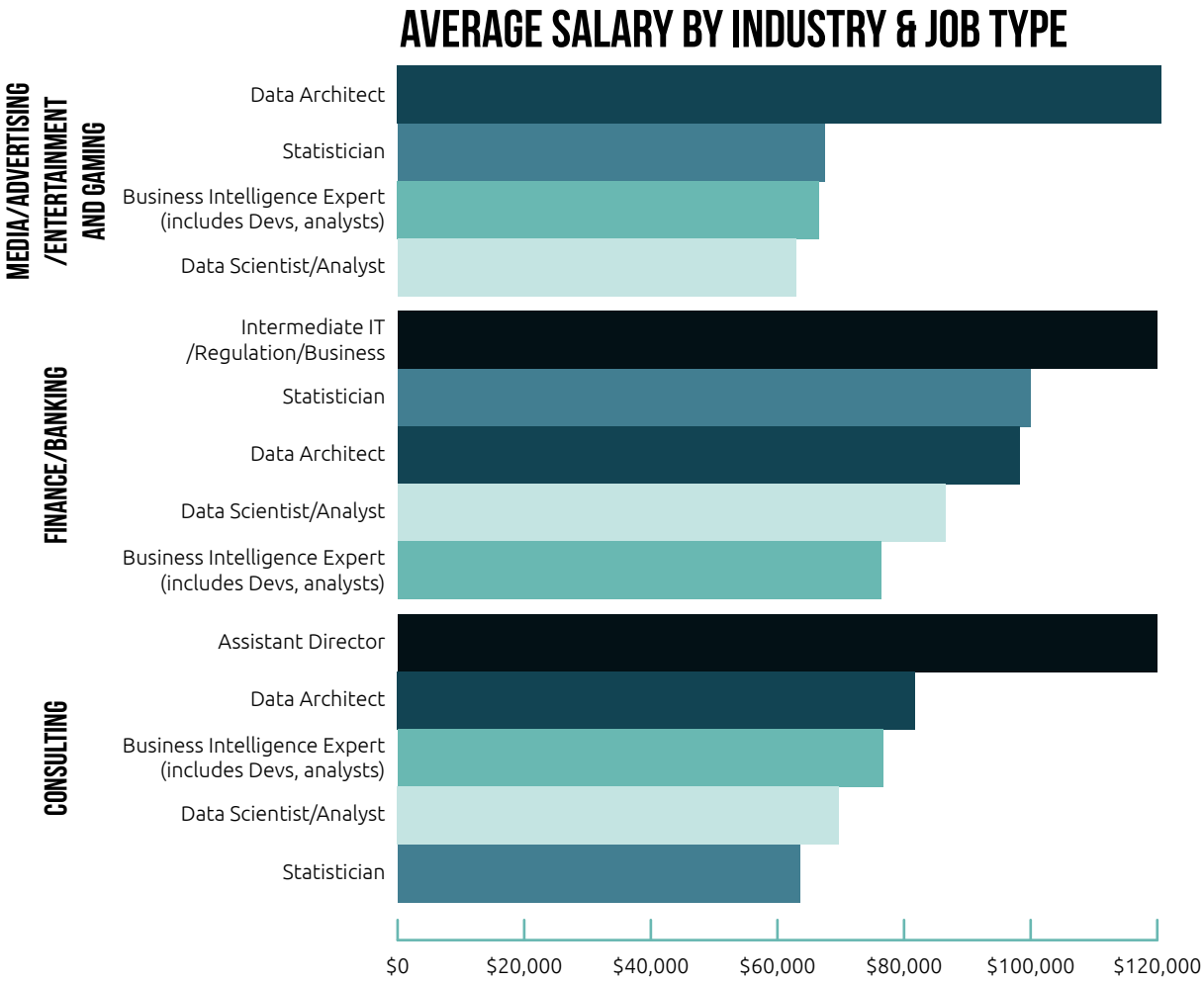
Media / Advertising / Entertainment and Gaming, while certainly not as lucrative as the Financial sector, appears to be a sector offering a substantial salary to inexperienced people. If we consider the fact that this sector, taken generally, is very competitive for inexperienced people and known for low salaries, the data provides a clear indication that these industries are willing to invest in inexperienced people with technical and numerical skills.

This is an indication that these industries are relying on data-driven strategies to remain competitive in tough areas of the economy. While they may lack the funds and cashflow to invest in established data professionals, inexperienced people with the right skills might well fill this role perfectly, without commanding such high salaries.

CAREER DEVELOPMENT AND DATA-ORIENTED ROLES

We've seen indicators of what's best for inexperienced data professionals, but what about career development?

We also wanted to see how the different job types fare in different industries. The graph below shows which specific roles command the highest salaries, and in which industry these roles are most valuable.



- Data Architect is one of the most important roles within Media/ Advertising/Entertainment and Gaming, commanding a higher salary in other areas.
- Comparing the roles of Data Architect to Statistician, we can infer some key differences about how data science and data-driven strategies are playing out between industries. In Finance, the statistician earns slightly more than the data architect whilst in the Media category, the Data Architect earns significantly more.
- In Media and Entertainment, where agility and organizational change is essential for rapid responses to change, there is a high value placed on someone who is able to develop a solution, such as a Data Architect. Whereas, in Finance these architectures are already in place.

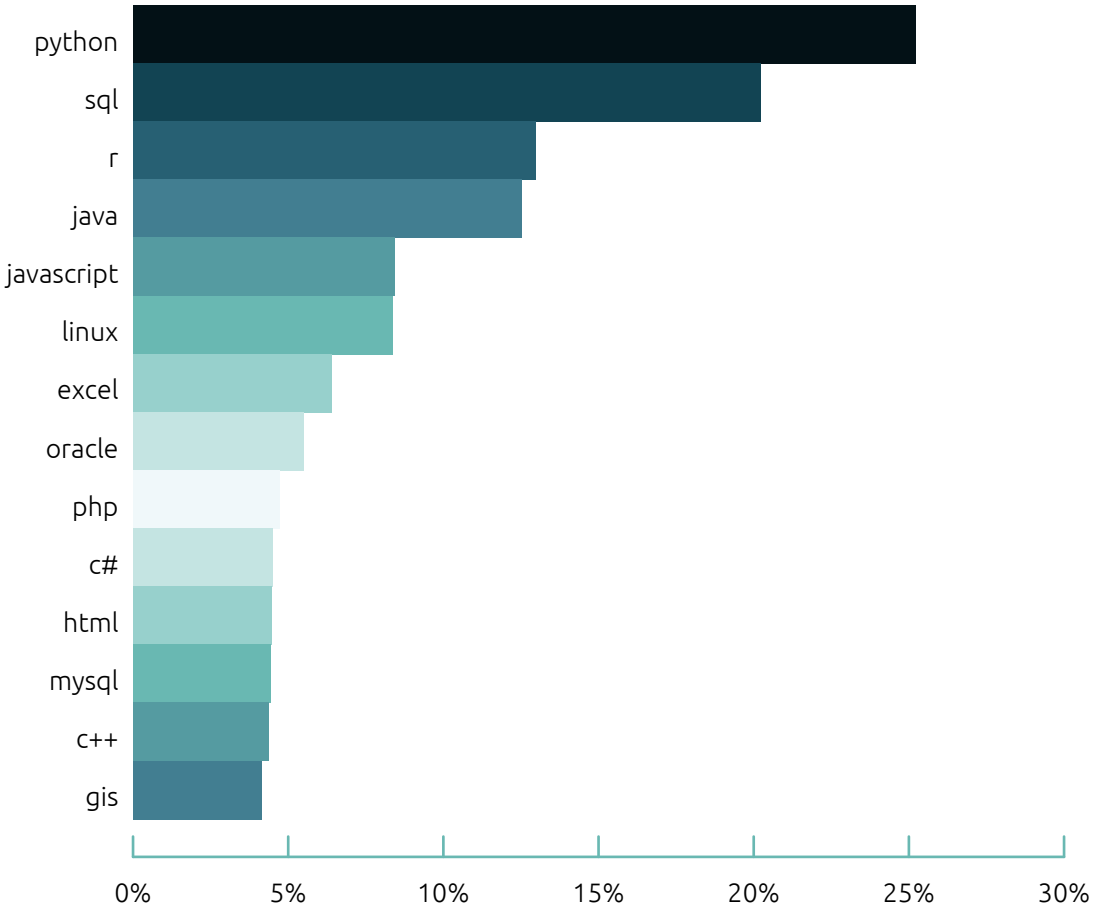
TECHNOLOGY USAGE ANALYSIS

- More than 25% of respondents use Python on a daily basis, but almost the same number use R.
- Distributed computing and machine learning tools are becoming more and more important.
- Augmented Reality and the Internet of Things are poised to change how we think about data.

The last two points spell out the next 5 years in data. The volume of data available from all sources will only continue to grow exponentially. The technology to deal with that data is still being developed, so don't let yourself get left behind!

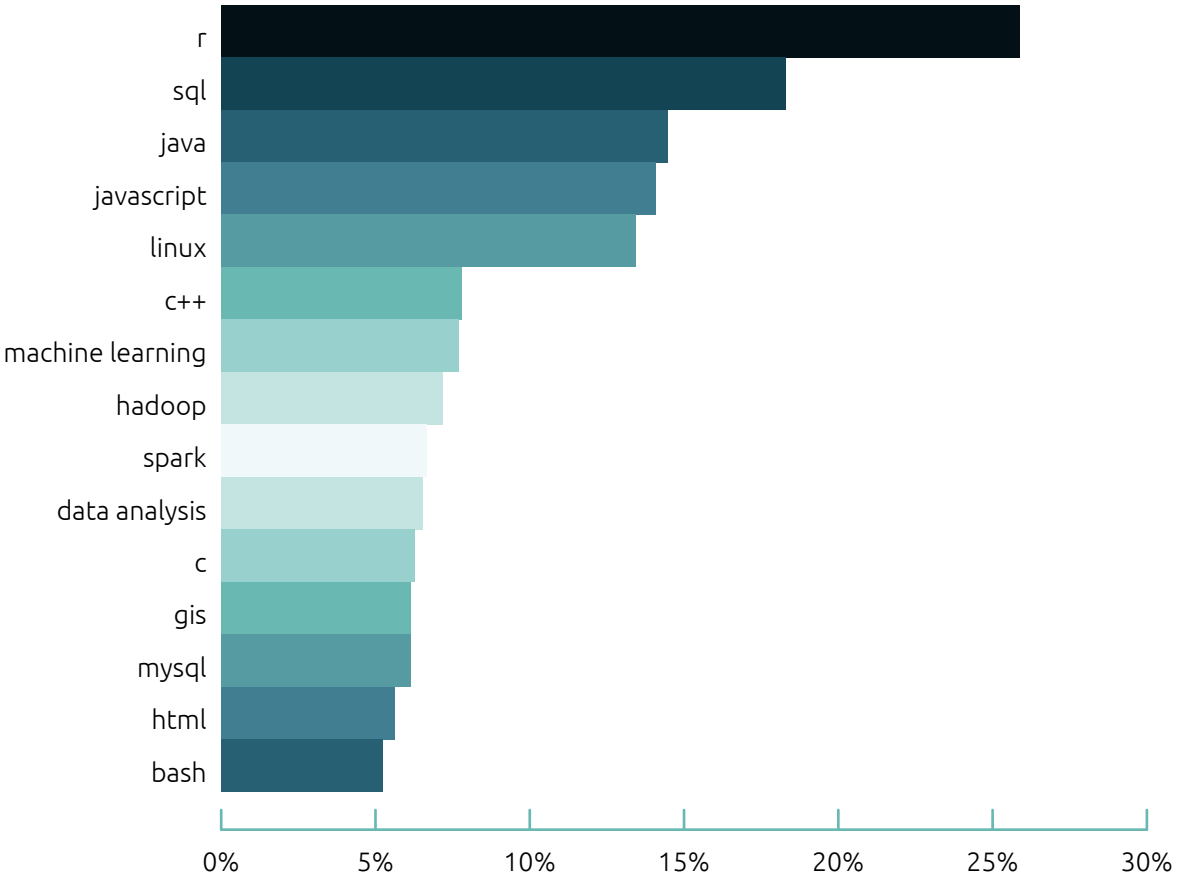
WHAT EXACTLY ARE PEOPLE USING?

We asked what tools people use on a daily basis, and here's what they said:



Python is definitely top dog when it comes to data. Its versatility and wealth of easy to use third-party libraries for everything from machine learning to web scraping, combined with its low barrier to entry, make it the ideal choice.

WHAT ELSE ARE DATA PYTHONISTAS USING ON A DAILY BASIS?



A large number of respondents are using Python and R, Python and Java, or Python and C++.

Get to grips with both Python and R to broaden your fluency and become a more flexible data scientist:

- Practical Data Science Cookbook
- R Data Analysis Cookbook
- Python Data Analysis
- R for Data Science
- Python Data Science Essentials

TECH STACKS

To get a better idea of what people are commonly using together, let's do some clustering. We put all the responses into a graph database, and ran a clustering algorithm across the techs people use every day, and from this we've identified some coherent clusters of people based on their stack:

1. DATA VISUALIZERS

These are people working at the design end of the data science spectrum. As you can see from the cluster of tools below, they have as much in common with a front end web developer as any of our groups. JavaScript comes top here – unsurprising when you consider how much it dominates web development today – and other design tools and plugins such as CSS, HTML5 and jQuery serve to underline its dominance.

The very presence of this group of tools in our data emphasises the importance of communicating insight via the web, and highlights just how important design is when trying to understand and interpret data.

1. JAVASCRIPT
2. HTML
3. CSS
4. PHP
5. HTML5
6. JQUERY
7. JAVA
8. XML

2. PROGRAMMATIC DATA WRANGLERS

This group of tools are used by those people that squeeze insight out of data. They are primarily responsible for mining, cleaning and manipulating data very quickly, in order to answer specific questions about everything from customer behaviour to financial planning.

Python here comes out as the most important tool – perhaps unsurprising given how easy it is to prototype and its much-vaunted flexibility.

The presence of pandas serves to underline Python's dominance – indeed, it might consolidate it precisely because of the way in which pandas improve Python's data analysis capabilities.

But it's also interesting to see C++ here – while Python offers flexibility and ease, the impressive speed that C++ offers is still unrivalled.

1. PYTHON
2. C++
3. LINUX
4. BASH
5. PANDAS
6. MATLAB
7. MACHINE LEARNING
8. POSTGRESQL

3. BIG DATA EXPERTS

This cluster of tools is used by Big Data specialists, interested in scalability and robustness.

Hadoop here dominates the Big Data world – however, Scala and Spark are also growing in prominence, as their presence in this cluster indicates. As demand grows for faster processing (and by extension real-time analytics), it's likely that we'll see more from them.

It's also interesting to see web tools such as JavaScript and Spring included here. As with the first cluster, this indicates the need to communicate effectively and quickly through web based applications.

1. JAVA
2. HADOOP
3. MYSQL
4. JAVASCRIPT
5. SCALA
6. SPRING
7. SPARK
8. MAVEN

4. DATA ARCHITECTS

This group of tools reflects the need to organize and communicate data insights in effective and intelligent ways – the main challenge of a Data Architect. Clearly, Enterprise-Ready tools dominate this cluster, indicating that Microsoft and Oracle are still regarded as go-to brands when it comes to these business-critical tools.

1. SQL
2. MS
3. ORACLE
4. SERVER
5. SSIS
6. SSAS
7. DATABASE
8. SSRS

This provides a useful insight into how the world of data breaks down, and how different roles appear to be built around 'ecosystems' of tools. One of the interesting questions that comes out of this is how this might change over the coming years. Is it possible that these clusters will become more fluid?

WHAT COMES NEXT?

We asked people what tools they were planning on learning over the next 6 months. We wanted to know what's hot, and what people in the know (those earning the money) are using, to help you decide in which area you'd like to enhance your skills, or learn new ones.

In the tag cloud below word frequencies are weighted by salary. So what are people looking to learn?

There are a few key points our research suggests here:

- Spark and Hadoop are well represented, suggesting the growth of cluster computing.
- Web Based technologies are also appearing here indicating people are seeing a value in sharing data analysis on the web.
- NoSQL databases are going to keep rising in the data world.

WHAT TRENDS ARE EMERGING?

We asked respondents what they think is the most important trend emerging in their field in the next 12 months:

There are some clear messages here:

- Machine learning is going to become one of the focal points for everyone working in data, driven by a demand for predictive insights and statistical analysis in a range of sectors and industries.
- Augmented Reality and Internet of Things are going to be key challenges for data scientists over the next few years.
- The prominence of mobile in the tag cloud suggests an increased emphasis on mobile analytics as users move further away from desktop.
- Distributed Computing (both on clusters and on the Cloud) is going to change the way we even think about data, suggesting increasing anxiety about how to manage resources for Big Data projects. This is possibly symptomatic of the two outcomes above. As IoT and mobile become more dominant, managing larger datasets is going to become a greater challenge.

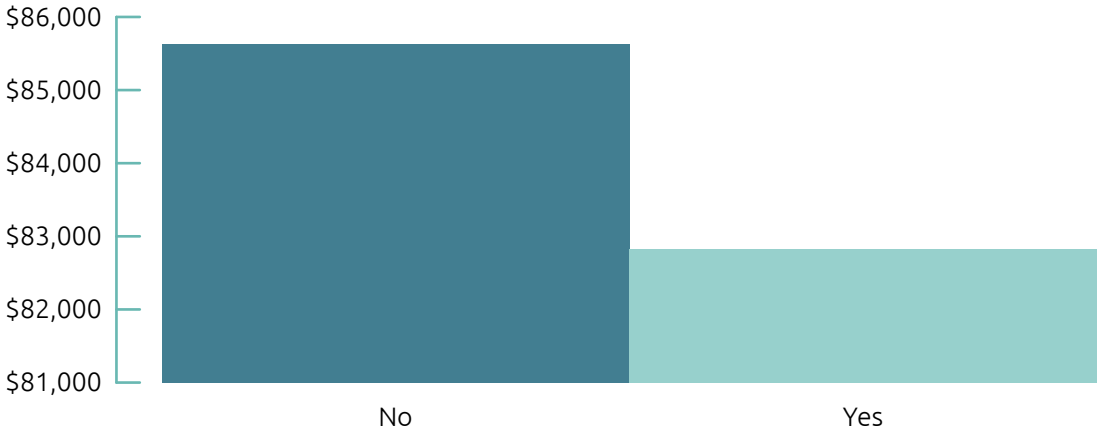
If you're involved with a Big Data project, you're going to need Machine Learning. Pick up this bundle and start exploring machine learning and predictive analytics today.

- Machine Learning with Spark
- Scala for Machine Learning
- Machine Learning with R
- Machine Learning with R Cookbook
- Building Machine Learning Systems with Python - Second Edition

HOT TOPICS

To complete our survey, we asked respondents some simple questions about hot topics, trends and challenges in the data.

DO YOU THINK JULIA WILL REPLACE R AND PYTHON AS THE DATA SCIENCE LANGUAGE OF CHOICE IN THE NEXT 12 MONTHS?

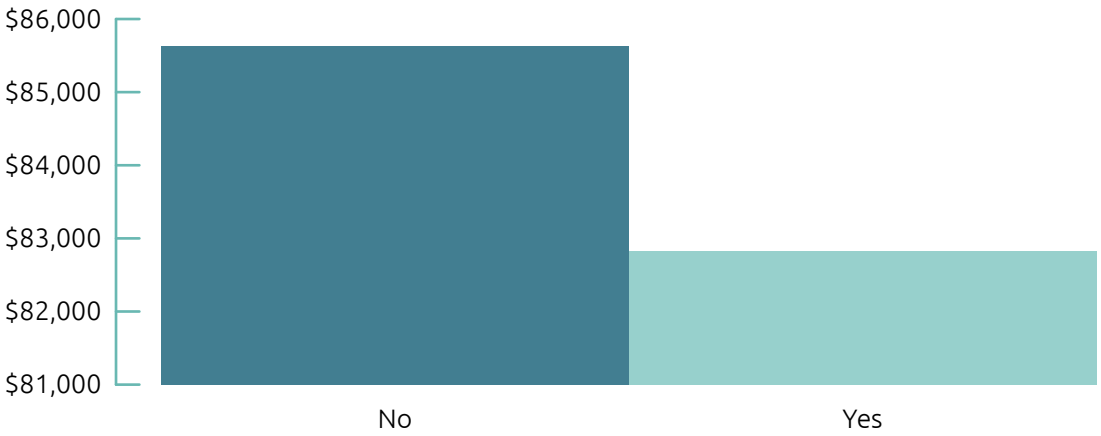


According to the highest earning respondents, Julia is on the ascent. It's easy to see why as it is designed specifically for technical computing, and boasting interesting features such as multiple dispatch, useful libraries for graphing, and impressive JIT compiler benchmarks. Julia is one to watch!

Stay ahead of the trend and start learning Julia with this essential selection of Julia books.

- Getting started with Julia Programming Language
- Mastering Julia
- Getting Started with LLVM Core Libraries
- Python High Performance Programming
- R High Performance Programming

DO YOU THINK APACHE SPARK IS LIKELY TO REPLACE HADOOP IN THE NEXT 12 MONTHS?



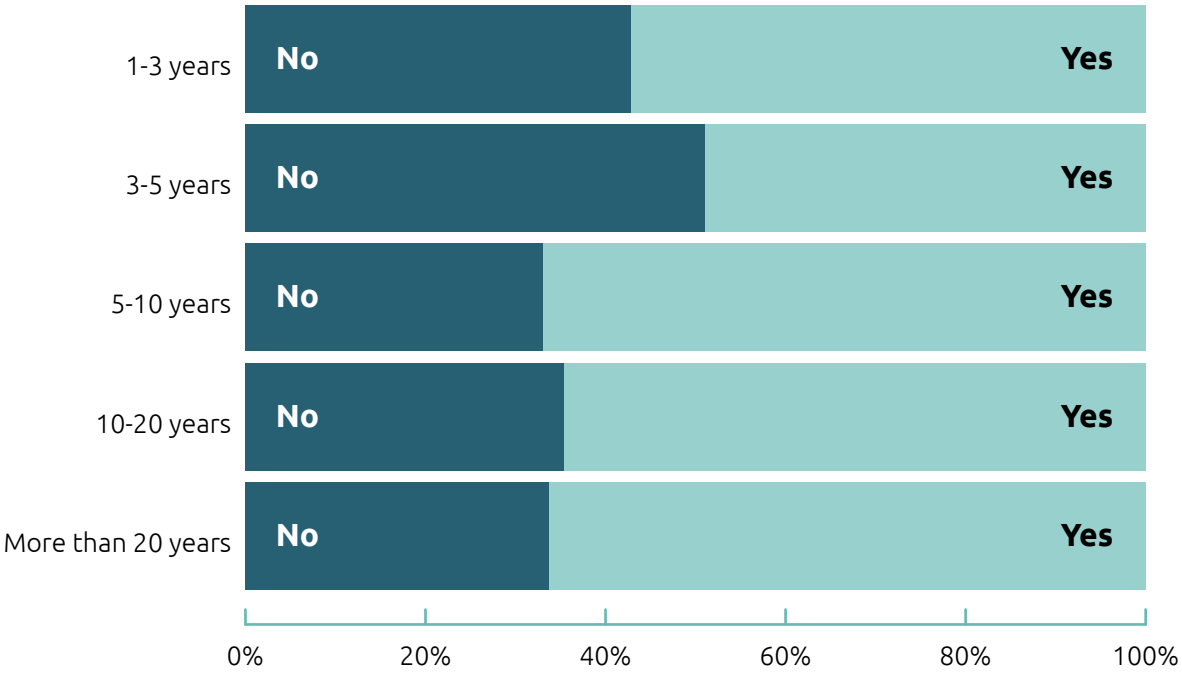
Our survey says you're safe for now!

However, if you're not already on the Hadoop train, there's never been a better time to get on it…

- Learning Hadoop 2
- Mastering Hadoop
- Big Data Analytics with R and Hadoop
- Fast Data Processing with Spark - Second Edition
- Apache Mesos Essentials

IS THE LINE BETWEEN DATA ANALYSIS AND DATA RETRIEVAL BEING BLURRED?

Across the board, this was contentious.



- 64% agree that it is.

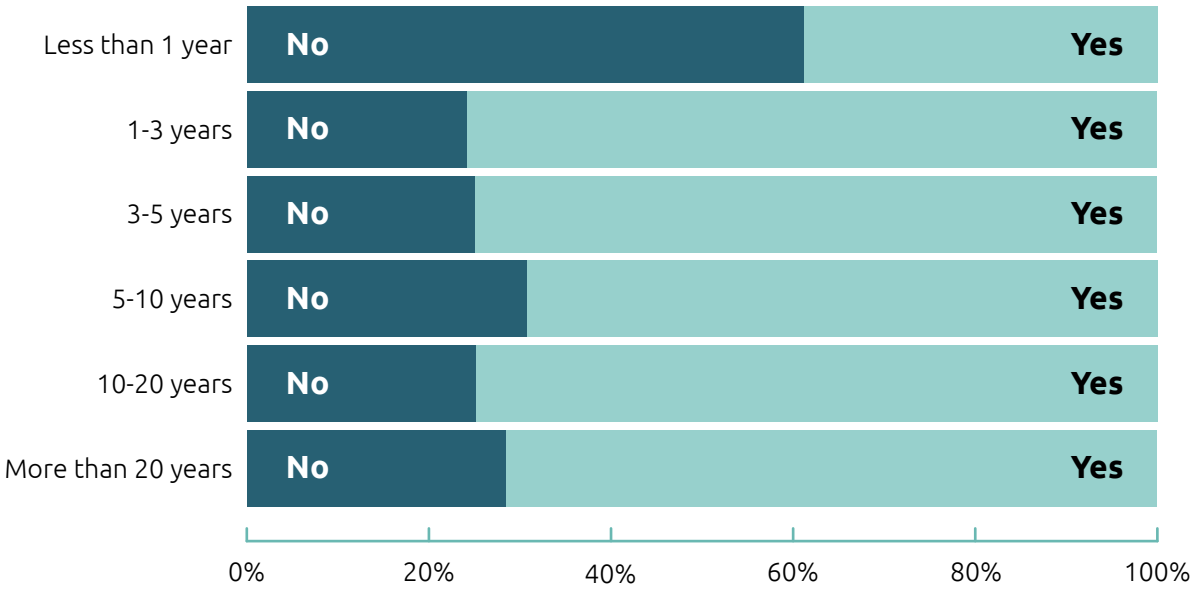
This suggests tools such as BigQuery could become more prominent. It's certainly a tool to watch over the next 12 months.

As Big Data becomes ubiquitous, the aim of the game is no longer to simply have the most effective strategy, but also the most efficient and fast.

Learn how to master the art of data analysis and retrieval with this bundle of popular books:

- Clean Data
- Practical Data Analysis
- Mastering Predictive Analytics with R
- Learning Data Mining with R
- Learning Pandas

IS YOUR COMPANY PLANNING TO IMPLEMENT A BIG DATA PROJECT OVER THE NEXT 12 MONTHS?

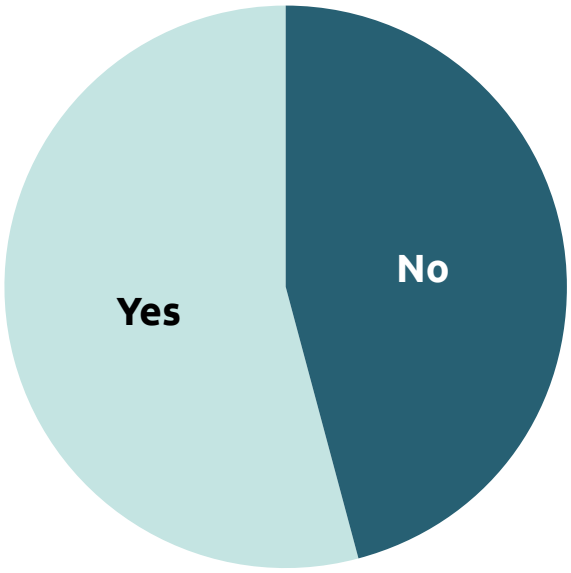


It appears that most respondents are working in organizations looking to implement Big Data projects. The significant anomaly to that are those with very little experience. The reasons for this aren't immediately obvious, but there are a number of possible explanations.

It could be that those with very little experience simply aren't privy to organizational strategy and decision-making. Conversely, it could be that those people who are just starting out have been employed precisely because a Big Data project has been implemented.

AND FINALLY...

DOES EXCEL STILL HOLD A PLACE IN YOUR HEART?



Excel is eternal, Excel 2013 is awesome…all of you know it deep down!

SUMMARY

It may be a truism, but it's clear, and perhaps it has been for years, that Data Science and Big Data are not simply trends, but are instead symptomatic of a wider social, cultural and economic change.

It's time we stopped talking about the 'Big Data revolution' or how 'data scientist' is the 'sexiest job of the twenty-first century', and instead look at the different ways data is being used in different areas. For SMEs, data is crucial for making companies more responsive and open to changes in the market. The prominence of machine learning underlines this further, making it clear that there is a real onus on delivering rapid insight and fast! For larger organizations there is a drive towards creating faster Big Data solutions. The apparent rise of distributed and cluster computing is evidence of this, as data scientists and analysts look for new ways to put tools such as Hadoop and Spark to work.

But even more interesting is how our understanding of data looks set to change, thanks to emerging trends such as the Internet of Things and Augmented Reality. It's possible that IOT will become the buzzword to replace Big Data. How Data Scientists, analysts and architects tackle it day to day isn't clear yet, but it will almost certainly be a challenge that will offer exciting opportunities for data literate people everywhere.

What you should be doing if you're a data scientist:

- Broadening the range of languages you know is essential. It will help you become more flexible when working on a range of different projects and also provides you with more solutions. If you know R, why not learn Python?
- You need to get to grips with Machine Learning. If you want to get started or investigate it further, grab our Machine Learning bundle!
- If you're just starting your career, you could do a lot worse than working in Finance or for an SME. You might command a higher salary working for an established enterprise organization, but the difference is likely to be small with more opportunities and responsibility at an SME.
- If you're interested in working in popular industries such as Media, work towards becoming a data architect, and learn how to develop and implement large-scale data solutions that can deliver benefits across an organization.
- Getting to grips with Big Data tools such as Hadoop and Spark will be valuable, but learning how to use them in the context of distributed networks will be even more valuable as resources become stretched.
- Pay attention to IoT – we still don't quite know where it will lead the data world!

[PACKT]
PUBLISHING

HELPING IT PROFESSIONALS TO PUT SOFTWARE TO WORK IN NEW WAYS

Founded in 2004 in Birmingham, UK, Packt's mission is to help the world put software to work in new ways, through the delivery of effective learning and information services to IT professionals.

Working towards that vision, we have published over 3000 books and videos so far, providing IT professionals with the actionable knowledge they need to get the job done –whether that's specific learning on an emerging technology or optimizing key skills in more established tools.

As part of our mission, we have also awarded over $1,000,000 through our Open Source Project Royalty scheme, helping numerous projects become household names along the way.

DATA ANALYST
Greg Roberts

EDITOR
Richard Gall

DESIGN
Chris Murray

PROJECT MANAGER
Sarah Cullington

TECHNICAL ADVISOR
Akram Hussain

www.packtpub.com **Tel: +44 (0)121 265 6484** **Published 2015**