Insight Report

Data Science in the New Economy

A new race for talent in the Fourth Industrial Revolution

Centre for the New Economy and Society

July 2019

This Report is produced by the World Economic Forum's Centre for the New Economy and Society as part of its New Metrics CoLab project. For more information, or to get involved, please contact cnes@weforum.org.

This Report has been published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum, but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

Contents

Key Findings

Through three industrial revolutions, technology has led to significant changes across economies, societies and businesses. Steam engines jump-started the transition of societies from agriculture to industrial production. The use of fossil fuels in engines and innovation in business models such as the assembly line rapidly scaled production. More recently, the digital revolution brought computing power and information technology. Each successive industrial revolution has involved significant shifts in the way people live and work, in how value is created in the economy, and demand for the highest-value skills.

As the Fourth Industrial Revolution unfolds, led by advances in technologies such as data science and artificial intelligence, the labour market is again changing in a fundamental fashion. In 2018 the Future of Jobs Survey and Report revealed that business leaders believe that by 2022, human workers and automated processes are set to share the workload of current tasks equally, while a range of new roles is expected to emerge simultaneously as digital innovation is absorbed across industries and regions. In particular, in many large advanced and emerging markets, growth is expected in sectors that will experience the bulk of these new roles, such as information technology, renewable energy, education and the care economy, and in occupations such as data science, healthcare work and human resources.

While the new labour market is changing at a rapid pace, emerging data sources are shedding light on its composition with a new depth and dynamism that has not previously existed. Online platforms and specialized insight firms are now offering new and complementary ways to understand how specific skills, tasks and occupations are changing across industries and geographies. While many of these remain limited to specific populations—and difficult to compare and contrast—when coupled with traditional and qualitative sources of data, they can help businesses, policy-makers and workers have greater analytic capacity about the present and future of work and adopt better informed and coordinated business strategies and policies.

The World Economic Forum's Centre for the New Economy and Society is a platform for insights and action on emerging socio-economic issues. As part of its agenda, the Centre offers an opportunity for companies to partner with the Forum to share insights on emerging trends in a diverse set of issues such as education, skills, jobs, gender and inclusive growth. This Report is among a series of such collaborations aimed at developing new metrics and deploying data for shedding light on public good challenges.

It has become commonplace to refer to data as the 'new oil' of the global economy. Data scientists are the talent that provide the ability to extract, refine and deploy this new source of value in the global economy. This Report focuses on data science, among the most competitive skills of the Fourth Industrial Revolution, in collaboration with Burning Glass Technologies, LinkedIn and Coursera to shed light on how data science talent is being developed and deployed across today's labour market.

In the first wave of digitization, data could be seen as purely a by-product of the functioning of digital applications, operating systems and platforms. Data is now increasingly recognized as a significant asset enabling further innovation across ancillary fields such as artificial intelligence, which can drive the improvement of services through process efficiency and deliver better results for customers. The dividends of new sources of data and methods of processing are not limited to the private sector; the public sector increasingly uses data to improve government services and academia applies new methods to enhance research. Yet the rapid rise in the demand for workers with skills in data science has led to a shortfall in data science skills supply and intense competition between industry, academia and the public sector for such talent. This has created a high premium on such skills and has reduced the capacity of businesses, industries and entire economies to leverage fully the dividends of innovation.

What do we know about how recent demand for this talent is emerging across industries, how data science skills are becoming part of the core composition of roles and how workers are expanding their data science skillset through new learning across industries and regions? What do we know about how data science demand might be expected to grow in the future?

In this paper, we look at three complementary ways in which leaders can understand the market for data science skills across the new economy: monitoring the demand for data science skills through job posting analysis from Burning Glass Technologies; the distribution and quality of data science

talent across industries and regions based on learner skills insights from Coursera; and analysing the rising prevalence of data science skills within the core composition of selected roles through user profile analysis from LinkedIn. Finally, we conclude with a look towards the future demand for data science skills across industries, drawing from the insights of executives of the largest companies in the world surveyed through the World Economic Forum's Future of Jobs Survey.

Key insights

1. While data science roles and skills form a relatively small part of the workforce, recent trends indicate that these are currently among the highest in-demand roles in the labour market.

2. The demand for data science skills is not limited to the Information Technology sector as data's importance grows across multiple sectors, including Media and Entertainment, Financial Services and Professional Services.

3. Data science skills are particularly critical to a distinctive set of growing roles. For example, in the United States those roles are Machine Learning Engineers and Data Science Specialists. These skills are only nominally in demand across more traditional roles such as Relationship Consultants, but those roles are also facing major churn in skills.

4. The data science skillset is not fixed and is rapidly evolving as new opportunities in data analysis and further technological advances redefine the specific skills composition of data scientist roles.

5. The disparities in achievement of data science learners point to varying levels of data science talent across industries and economies:

   a. The Information Communication and Technology (ICT), Media and Entertainment, Financial Services and Professional Services industries are currently taking the lead both in hiring data science talent and in the achievements of online learners who are actively updating their skillsets across industries.

   b. Across most industries, online learners based in Europe demonstrate higher proficiency in data science skills than in North America, followed by emerging regions. Exceptions to such trends exist in sectors such as Telecommunications and Technology, where learners in the Asia Pacific region and the Middle East and Africa outperform regional averages across industries.

6. Jobs such as Artificial Intelligence and Machine Learning Specialists or Data Scientists, in which data science skills are perhaps most profoundly applicable, are forecasted to be among the most in-demand roles across most industries by 2022.

Implications for decision-makers

Overall, the rapid growth and evolution of data science roles and skills stresses the need for appropriate business strategies and education and training policies that can match this demand, in quantity and quality, so that skills shortages do not hinder the transformation potential unveiled by vast sources of data and improved data analysis techniques. Industries and countries that fail to understand and address these dynamics risk slower growth and dynamism.

More precisely, the insights included in this Report point to the following implications:

— In the Fourth Industrial Revolution all sectors will need to undergo a fundamental transformation to fully absorb the potential dividends of the data economy. Such transformations will need to be accompanied by appropriate talent investments in data science skills.

— Industries that have been able to capture a large share of high-skilled talent in more traditional data science skills such as statistics or data management cannot be complacent, and they need to make fresh investments in newer skills, such as data visualization or statistical programming, if they wish to meet their innovation potential.

— A one-shot investment in reskilling will not be sufficient. Given the rapid pace of change within data science professions, maintaining skillset relevance will require responsive and dynamic upskilling systems that respond to fast-changing technologies and associated skills demand.

— Differences in achievements of online learners across industries and regions showcase potential data science skills capacity gaps which, if left unaddressed, may reduce the innovation and competitiveness potential of specific businesses, industries and regions. Public and private sector stakeholders in these regions will need to consider greater investment in data science. Given the rising demand for such skills, this investment is likely to generate significant returns for individuals and companies and contribute to generating new pathways for socio-economic mobility.

The sections that follow present three new metric scorecards that individually and collectively shed new light on data science roles and skills in the labour market of the Fourth Industrial Revolution. The collection provides one starting point to what could be further efforts aimed at tracking skills demand and capacity across emerging sectors such as renewable energy and the care economy. This exercise can set the foundations to analyse skills dynamics in other sectors, building on the potential of multi-source data collaboration to create coherent frames of analysis and common taxonomies that can provide business and policy leaders with a common frame of reference. Finally, this Report presents a forecast on the importance of data analysis jobs across multiple industries.

Scorecard 1
Emerging Demand for Data Science Skills Across Industries

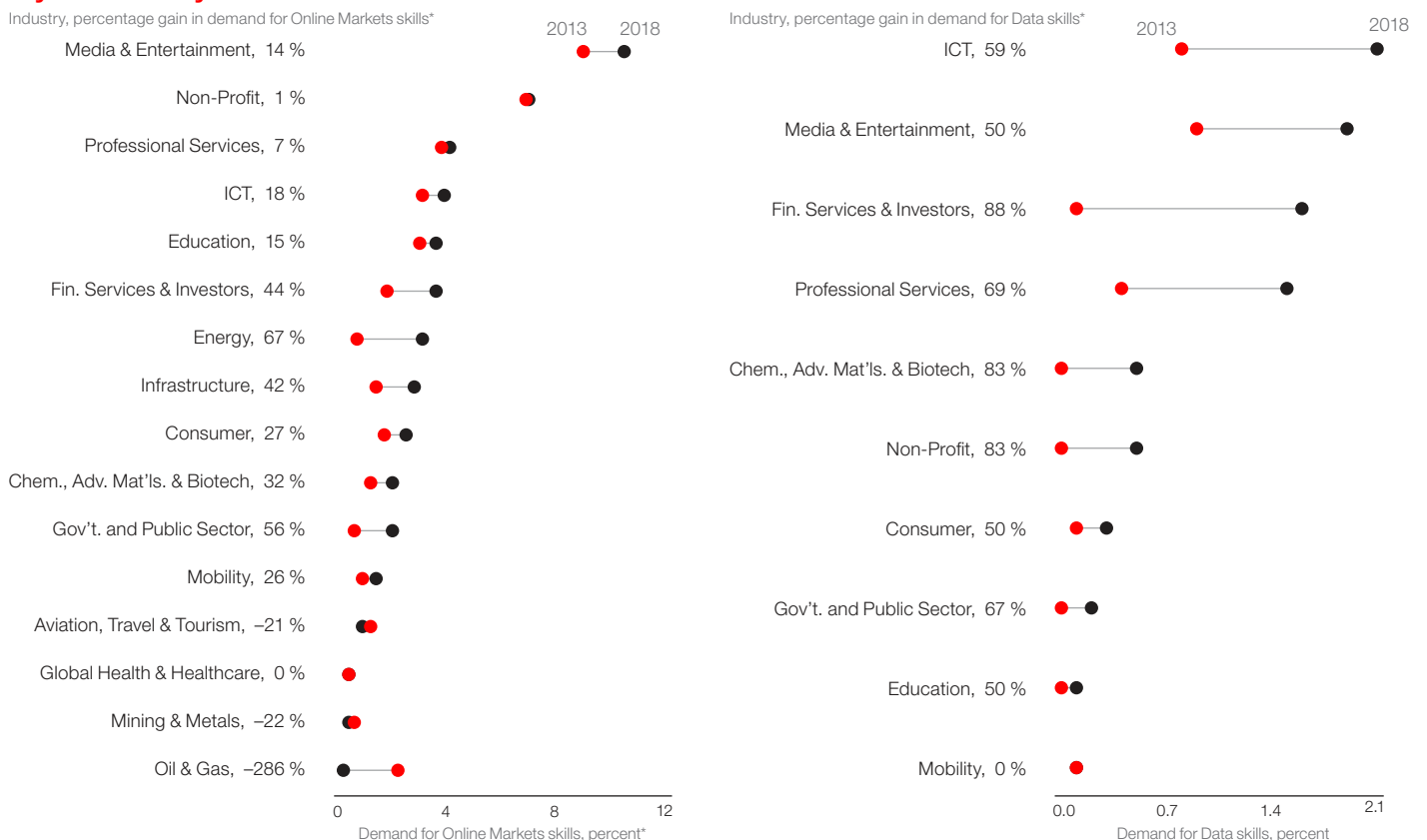# Online Markets and Data skillsets, five-year trend, by industry
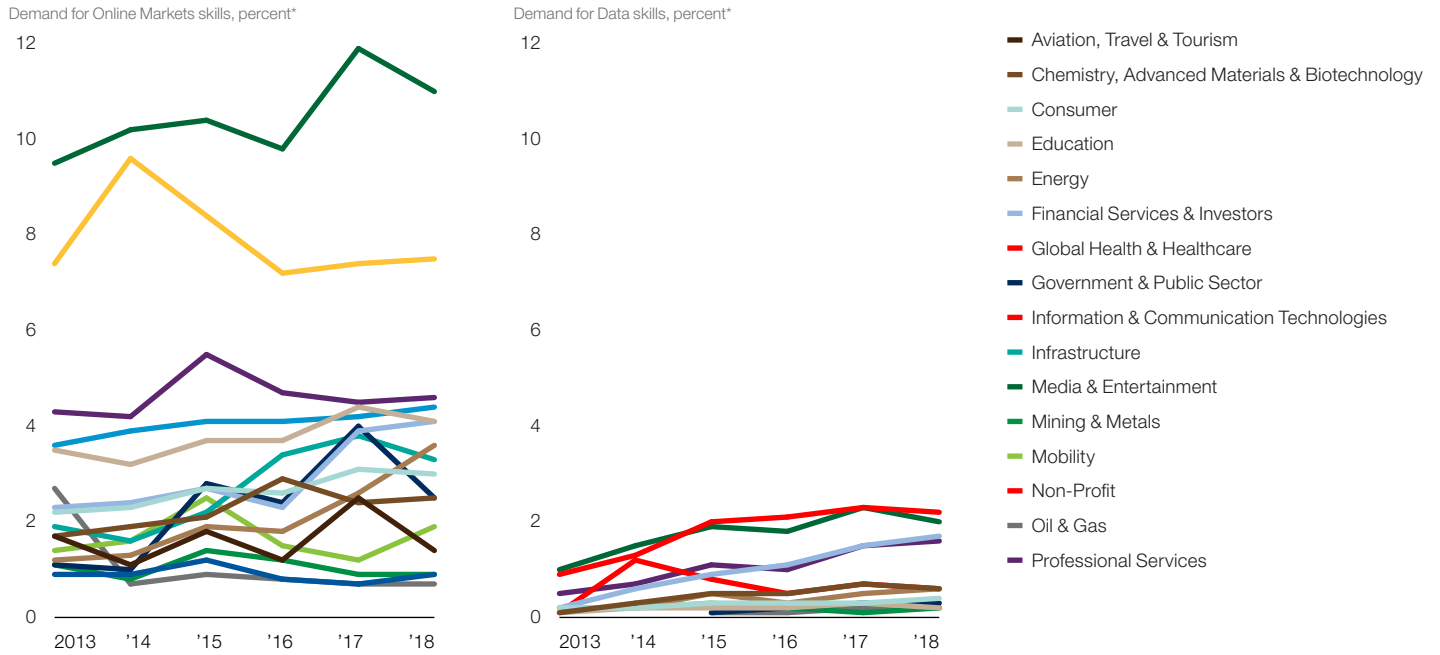
Demand for Online Markets skills, percent*

Demand for Data skills, percent*

Legend:
- Aviation, Travel & Tourism
- Chemistry, Advanced Materials & Biotechnology
- Consumer
- Education
- Energy
- Financial Services & Investors
- Global Health & Healthcare
- Government & Public Sector
- Information & Communication Technologies
- Infrastructure
- Media & Entertainment
- Mining & Metals
- Mobility
- Non-Profit
- Oil & Gas
- Professional Services

Source: Burning Glass Technologies.
Note: Demand is calculated as the number of times that skills in the Online Markets and Data skillsets are mentioned in job postings as a share of all skills that are mentioned.

Online Markets and Data skillsets relative position and growth, by industry

Industry, percentage gain in demand for Online Markets skills*

2013 2018

- Media & Entertainment, 14 %
- Non-Profit, 1 %
- Professional Services, 7 %
- ICT, 18 %
- Education, 15 %
- Fin. Services & Investors, 44 %
- Energy, 67 %
- Infrastructure, 42 %
- Consumer, 27 %
- Chem., Adv. Mat'ls. & Biotech, 32 %
- Gov't. and Public Sector, 56 %
- Mobility, 26 %
- Aviation, Travel & Tourism, −21 %
- Global Health & Healthcare, 0 %
- Mining & Metals, −22 %
- Oil & Gas, −286 %

Demand for Online Markets skills, percent*

Industry, percentage gain in demand for Data skills*

2013 2018

- ICT, 59 %
- Media & Entertainment, 50 %
- Fin. Services & Investors, 88 %
- Professional Services, 69 %
- Chem., Adv. Mat'ls. & Biotech, 83 %
- Non-Profit, 83 %
- Consumer, 50 %
- Gov't. and Public Sector, 67 %
- Education, 50 %
- Mobility, 0 %

Demand for Data skills, percent

Source: Burning Glass Technologies.
Note: Demand is calculated as the number of times that skills in the Online Markets and Data skillsets are mentioned in job postings as a share of all skills that are mentioned.

Emerging Demand for Data Science Skills Across Industries

Actionable Insight

The metric

The Burning Glass Skills Metric tracks demand for the skills that accompany the adoption of new technologies. When a new job is posted online, an algorithm distils relevant data, such as the type of role advertised and the skills required for the role. The changing share of skills mentioned in job postings viewed as a trend can be analysed in an aggregate fashion to indicate rising and falling demand for specialized talent. Such changes in business' search for specialized human capital can also serve as a proxy for business investment in technologies.

This metric offers new insights into the demand for skills across industries. A job posting communicates a required skill that is important, new or unusual. The information that is scraped and distilled from such a source corresponds to the skills that are either fundamental or new requirements for roles in the labour market.

In a labour market in transition, the skills requirements detailed in newly advertised roles offer an early indication of the adoption of new technologies and associated changes in process. They showcase the intentions of business leaders to match investment in new technologies with appropriate human capital.

Key insights

This scorecard examines changes in industry demand for a cluster of skills related to user and entity big data analytics. That data is then contrasted against the demand for a set of skills associated with app and web-enabled markets. These two skills clusters correspond to two of the technologies that were identified in the *Future of Jobs Report 2018* as most likely to be adopted in the horizon up to 2022. The cluster of skills included in the App- and Web-enabled Markets skillset—abbreviated to Online Markets throughout this Report—includes skills such as: social media, paid search, e-commerce and mobile application design. The skills cluster included in the User and Entity Big Data Analytics skillset (or Data skillset) includes skills such as: Apache Hadoop, Hives, Big Data, Oracle Big Data and AWS Elastic MapReduce.

The comparison between both skillsets for the period between 2013 and 2018 across different industries show some common features. Notably, the industries making more investments are the Media and Entertainment (ME), Professional Services (PS), Information and Communications Technologies (ICT) and Financial Services (FS) industries. At the same time, the data also reveals sharp differences. More precisely, it shows that the demand for data skills is still relatively modest in comparison to online markets skills. However, the figures also show that the demand for data skills is growing quickly. More precisely, between 2013 and 2015 the demand for data skills increased by 59%, 50%, 69% and 88% for the Information and Communications Technologies (ICT), the Media and Entertainment (MEI), Professional Services (PS), and Financial Services (FS) industries, respectively. In contrast, demand for skills related to online markets rose by 18%, 14%, 7% and 44% for the same industries during that same period.

A closer examination of the trends for the Online Markets skillset reveals some industries have experienced demand peaks during particular years. Such peaks in demand can be observed in the Not-for-Profit industry in 2014 and the Government and Public Sector industry in 2017. Those peaks can be interpreted as an acute demand for specialized human capital as new technologies are adopted, followed by relative slowdown in demand for related skills as those technologies start to mature.

The Burning Glass Skills Metric offers an opportunity to understand the ebbs and flows of demand for specialized talent across the labour market in near real time and in granular detail. Such data allows decision-makers to track skills demand and technology adoption across industries and regions. The metric can be further expanded by complementing the understanding of the scale of demand for specific skills in the labour market with an understanding of the incentives for workers to upskill in the form of wages.

Implications for decision-makers

The value of this data lies in allowing individuals, businesses and policy-makers to understand the upsurge in demand for new skills across industries with a view to more dynamically addressing skills shortages or mismatches. It also supports leaders in targeting future investment in specialized human capital within and across industries.

Source Data and Model

The dataset compiled by Burning Glass Technologies is based on online job postings. This information is sourced by 'scraping' detailed data for a role from various online sources such as job boards and employer sites. Posts are then analysed, and key information is distilled from each posting. That key information includes the skills listed within the advertisement and the role advertised. Categories such as roles and skills are then further refined through methods such as clustering algorithms to arrive at data points organized by a coherent occupation and skills taxonomy, such as the O*NET occupational classification. For the current analysis the Burning Glass Technologies skills taxonomy is further mapped to a list of new technologies defined in the World Economic Forum's *Future of Jobs Report 2018*.

The resulting measures of skills prevalence represent the share of job postings that mention a specific set of skills during any one year and can be interpreted as skills demand— as in this scorecard. Across industries, a skills cluster is rarely mentioned in more than 10% of job postings. Of the combinations of cluster and role combinations examined, about 95% list a skills prevalence measure under 10%. It's best to focus interpretations of the listed figures on the distribution of values by industry and on the growth trend over time.

Scorecard 2
Learning Achievements in Data Science Skills Across Industries and Regions

Data science skills proficiency by industry, grouped by skills cluster

Data Management
1. Technology
2. Professional Services
3. Healthcare
4. Finance
5. Media and entertainment
6. Telecommunications
7. Consumer
8. Insurance
9. Manufacturing
10. Automotive

Data Visualization
1. Technology
2. Healthcare
3. Telecommunications
4. Professional Services
5. Consumer
6. Automotive
7. Manufacturing
8. Insurance
9. Media and entertainment
10. Finance

Machine Learning
1. Technology
2. Telecommunications
3. Professional Services
4. Manufacturing
5. Consumer
6. Finance
7. Media and entertainment
8. Insurance
9. Automotive
10. Healthcare

Mathematics
1. Technology
2. Professional Services
3. Telecommunications
4. Insurance
5. Finance
6. Manufacturing
7. Healthcare
8. Automotive
9. Media and entertainment
10. Consumer

Stat. Programming
1. Technology
2. Professional Services
3. Telecommunications
4. Manufacturing
5. Finance
6. Media and entertainment
7. Healthcare
8. Consumer
9. Insurance
10. Automotive

Statistics
1. Telecommunications
2. Finance
3. Technology
4. Automotive
5. Manufacturing
6. Media and entertainment
7. Professional Services
8. Insurance
9. Healthcare
10. Consumer

Source: Coursera.

Data science skills proficiency, by industry and region

Skills Clusters

Automotive, Consumer, Finance, Healthcare, Insurance

Data Management, Data Visualization, Machine Learning, Mathematics, Stat. Programming, Statistics

Proficiency (0–1, where 1 is best in class)

Skills Clusters

Manufacturing, Media & Entertainment, Professional Services, Technology, Telecommunications

Data Management, Data Visualization, Machine Learning, Mathematics, Stat. Programming, Statistics

Proficiency (0–1, where 1 is best in class)

Asia Pacific   Europe   Latin America   Middle East and Africa   North America

Source: Coursera.
Note: Regional data points might be excluded in cases where the data is inconclusive.

Learning Achievements in Data Science Skills Across Industries and Regions

Actionable Insight

The metric

The Coursera Global Skills Index evaluates learners across economies and industries and presents an indicative measure of their talent capacity across a range of critical skills. Each country, region or industry is evaluated relative to its peers and labelled with one of four categories – "Cutting edge" (score 0.76 to 1.00), "Competitive" (score 0.51 to 0.75), "Emerging" (score 0.26 to 0.50) and "Lagging" (score 0.00 to 0.25).

The courses and associated assessments that learners take on Coursera's online learning platform offers valuable insights into skills capacity across industries and economies. The data provides an indicative measure of skills proficiency for talent across industries and regions.

This is particularly important in the midst of the Fourth industrial Revolution, when regions and industries need to understand, develop and deploy specialized talent in order to ensure long-term labour and economic competitiveness.

Key insights

This scorecard presents differences in the assessed ability of online learners who specialize in data science skills. The results reflect the skills investment made by individuals, classified by industry and region, and reveal their degree of reskilling and upskilling achieved through online learning.

The figure above illustrates the comparative strengths and weaknesses of learners across industries and regions against the six skill clusters contained in the Data Science skillset: Mathematics, Statistics, Statistical Programming, Machine Learning, Data Management and Data Visualization. At the most granular level, these six clusters can be further disaggregated into more specific skills such as: calculus within Mathematics; linear regression within Statistics; programming languages such as R within Statistical Programming; neural networks within Machine learning; Hadoop within Data Management; and charting within Data Visualization.

Among these six clusters, three have introduced significant innovation in recent years: Statistical Programming, Data Visualization and Machine learning.

Overall, the Technology industry, which is at the forefront of the Fourth Industrial Revolution, shows the most competitive talent base in Data Science skills, both at the aggregate level and across skills clusters. Two industries follow: Professional Services and the Telecommunications industries.

Industries that have performed well in more traditional data skills, such as statistics or data management, cannot be complacent and need to make fresh investments in more innovative data skills, such as data visualization or statistical programming, if they wish to fulfil their innovation potential. The Finance industry, for example, falls under this category. Historically, it has maintained strong talent bases in quantitative skills and, especially statistics. Learner achievements indicate that the industry continues to be competitive in these traditional areas of expertise despite increasing competition for talent from the Technology industry. However, the achievements of Finance industry learners showcase lower skills proficiency across other skills such as Statistical Programming and Machine Learning where the industry ranks 5th and 6th, respectively.

The skills proficiency of learners varies by region. For example, in the Finance industry, learners across Europe are assessed as being at the cutting-edge of talent, while in Latin America they are considered emerging.

On average, online learners based in Europe demonstrate higher proficiency in Data Science skills than in North America, followed by emerging regions. For example, the Telecommunications and Technology sectors learners exhibit comparatively strong skills proficiency in the Asia Pacific and Middle East and Africa regions. Similarly, learners in the Middle East and Africa in the Automotive industries demonstrate cutting-edge data science skills.

When comparing skills achievements *across* industries and regions, some results are highly polarized. For example, in the European Healthcare sector, learners achieve cutting-edge results in Data Management skills while learners in North America lag behind. For Data Visualization learners in the Media and Entertainment industry, the same pattern follows. For mathematics skills, that pattern is reversed: learners in the Consumer industry in North America outperform Europeans learners.

Differences in learner assessments may reflect both supply- and demand-side factors, such as quality of education systems or workplace demand for skills. Over the long term, persistent gaps in data science skills capacity across regions and industries may exacerbate innovation and competitiveness divides and hinder the development of specific countries, regions and industries. As data becomes an increasingly important asset and labour markets demand more data science skills, policy-makers and business leaders must consider better development and deployment of competitive talent for the new labour market. The metric presented here can be expanded to further examine the relationship between output measures such as stock market indices and productivity figures for industries in the context of the varying achievements of learners.

Implications for decision-makers

The granular analysis of skills proficiency for online learners by industry and region can help identify opportunities to prioritize future talent specialization and target cross-regional upskilling and reskilling opportunities. This can guarantee the long-term competitiveness of regions and industries in the new global economy.

Keeping pace with the fundamental market shifts already underway in the Fourth Industrial Revolution will demand coordinated investments in skill development—not just by individuals, but also by companies and governments around the world.

Source Data and Model

The Coursera Global Skills Index is built on a mapping of skills to the content of online courses teaching them, and the assessments used to asses performance within them. It leverages machine learning models trained on skill tags crowdsourced from the tens of millions of learners and thousands of instructors on Coursera's platform.

Coursera builds a skill profile for each learner that is based on its performance across all attempted assessments, adjusting for the difficulty of those assessments. A skill proficiency value specific to an industry within a region is the average of the skill profiles of learners in that industry and region, weighted by the level of certainty in the per-learner skill profile. An industry-by-region proficiency ranking is eligible for inclusion if at least 100 learners from that industry and region have taken related assessments on the Coursera platform. Each proficiency value is ranked according to its skills index score, which is higher when learners in that group perform better on assessments within any single skill cluster.

To generate the industry-by-region proficiency rankings within each skills cluster, Coursera ranks each industry-by-region combination against the set with a sufficient number of learners for that cluster. This produces a series of relative percentile rankings. The percentile rankings are then organized into four categories: Lagging, Emerging, Competitive, and Cutting-Edge. Each industry-by-region ranking falls into one of these four categories based on relative performance in that cluster. The result can serve as a proxy for the relative quality of human capital capacity across regions and industries.

Out of the skills profiles compiled for each learner, Coursera is able to calculate a skills index score for 60 countries. The average number of assessed learners in each cluster-industry-regional group in this Report is 2,000 learners. Please note a series of data points listed as 'NA' (or missing). These reflect cluster-industry-regional groups for which Coursera is not able to calculate sufficiently robust measures of proficiency.

Scorecard 3
Changing Composition of Data Science Skills Within Roles

Top 10 emerging roles and skills genome

Change in rank of skills genome of selected emerging roles

rank	2015	change in rank	2018
1	Machine Learning	[new]	TensorFlow
2	Python (Programming Language)	[−1]	Machine Learning
3	Apache Spark	[+1]	Deep Learning
4	Deep Learning	[new]	Keras
5	Algorithms	[−2]	Apache Spark
6	Java	[new]	Natural Language Processing (NLP)
7	Big Data	[new]	Computer Vision
8	Hadoop	[−6]	Python (Programming Language)
9	Data Science	[—]	Data Science
10	C++	[new]	Amazon Web Services (AWS)

Top 10 emerging roles in the US

1	Blockchain Developer
2	**Machine Learning Engineer**
3	Application Sales Executive
4	Machine Learning Specialist
5	Professional Medical Representative
6	**Relationship Consultant**
7	**Data Science Specialist**
8	Assurance Staff
9	**Sales Development Rep**
10	Business Support Consultant

rank	2015	change in rank	2018
1	Financial Services	[+3]	Banking
2	Credit	[new]	Loans
3	Investments	[+2]	Retail Banking
4	Banking	[new]	Consumer Lending
5	Retail Banking	[+4]	Finance
6	Sales	[-]	Sales
7	Portfolio Management	[-5]	Credit
8	Team Building	[-7]	Financial Services
9	Finance	[-1]	Team Building
10	Leadership	[new]	Financial Analysis

rank	2015	change in rank	2018
1	Python (Programming Language)	[+1]	Machine Learning
2	Machine Learning	[+2]	Data Science
3	R	[−2]	Python (Programming Language)
4	Data Science	[−1]	R
5	Apache Spark	[+5]	SQL
6	Data Mining	[+3]	Statistics
7	Hadoop	[new]	Tableau
8	Data Analysis	[new]	Data Visualization
9	Statistics	[new]	Natural Language Processing (NLP)
10	SQL	[−5]	Apache Spark

rank	2015	change in rank	2018
1	Salesforce.com	[—]	Salesforce.com
2	Cold Calling	[—]	Cold Calling
3	Lead Generation	[+1]	Software as a Service (SaaS)
4	Software as a Service (SaaS)	[+1]	Sales
5	Sales	[new]	Sales Prospecting
6	Business Development	[—]	Business Development
7	Leadership	[−4]	Lead Generation
8	Sales Process	[+1]	Customer Relationship Management
9	Customer Relationship Management	[new]	SalesLoft
10	Account Management	[−2]	Sales Process

Source: LinkedIn.

Changing Composition of Data Science Skills Within Roles

Actionable Insight

The metric

The LinkedIn Skills Genome provides an opportunity to track the skills content of jobs on the basis of information about the skills which LinkedIn members feature on their professional profiles. A role is associated with a specific LinkedIn user when that user indicates that they have switched positions on their profile. The historical trend of such switches can be interpreted as hiring patterns. Through such data it is possible to trace the positions that users have occupied at different points in time to establish a comparative trend analysis. The set of emerging roles listed above is based on analysing hiring trends among the 177 million LinkedIn members in the United States between 2014 and 2018.

A metric based on the self-declared skills of individuals on the LinkedIn professional networking platform offers new insights into emerging roles and the evolving skills of professionals employed across the global economy. Each role requires a distinctive set of skills, and that skills profile shifts in tandem with the adoption of new technologies and business models. In the midst of the Fourth Industrial Revolution, changes in the labour market can be observed across two axes. First, the prevalence of different roles is shifting, with some roles declining in importance and another set of roles increasing in importance—these roles can be considered 'emerging'. Second, as technology modifies the tasks within professions the importance of different skills within roles also changes.

The LinkedIn Skills Genome is a measure of the unique set of skills which are prevalent in any one segment of the labour market. The type of segments that might provide insightful analysis include a particular geography such as a city or municipality, an industry, a job type such as data scientists, or a population such as women in the workforce.

Key insights

This scorecard examines the changing skills genome of two emerging roles which are known to require data science skills in the United States labour market between 2015 and 2018: Machine Learning Engineers and Data Science Specialists. This analysis is complemented by comparing these results with those of two emerging roles that do not traditionally require such skills: Sales Development Representatives and Relationship Consultants.

This allows a closer review of two types of roles which are emerging across the labour market today. On the one hand are a set of roles which contribute to building, operating and maintaining new technologies. These roles include a range of technical professions such as Blockchain Developers, Machine Learning Engineers, Machine Learning Specialists and Data Science Specialists. On the other hand are roles which engage customers and sell products and services: Application Sales Executives, Relationship Consultants, Sales Development Representatives and Business Support Consultants.

In 2015 Machine Learning Engineers typically indicated they used skills such as the programming language Apache Spark, Data Science, Deep Learning, Algorithms, Hadoop and Big Data. This skillset is similar to the skillset of a Data Science Specialist, although these roles place larger emphasis on additional skills such as R, Statistics and SQL. In the three years up to 2018, both roles have seen a significant change in their skills profiles. Overall, both have seen the emergence of Natural Language Processing Skills. The relevance of the programming language Python has declined across the two professions, notably for Machine Learning Engineers. As a point of differentiation, Data Science Specialists place higher emphasis on Data Visualization and SQL skills, while their Machine Learning counterparts are more likely to place emphasis on skills that make use of new programming tools specifically designed for machine learning such as TensorFlow and Keras, alongside an understanding of Computer Vision.

In contrast to these roles, Sales Development Representatives require a different skillset. Technological skills such as mastery of Salesforce.com are paramount, but the base skills requirements for the role also include customer relationship management and business development—skills that are considered 'soft' and human-centric. Moreover, the roles appear to be less affected by accelerated technological changes, and the skillset required for Sales Development Representatives demonstrates stronger stability than the Machine Learning Engineers or Data Scientist Specialists. One key difference in the analysed period, however, is the declining importance on skills in lead generation as opposed to sales prospecting. This change reflects a shift in emphasis among Sales Development Representatives when it comes to how potential customers are engaged. Sales prospecting—

one skill rising in prominence—prioritizes placing potential customers in touch with human sales representatives for an interactive dialogue over routinized company communication.

Changes in the profile of Relationship Consultants suggest a reorientation of this role to reflect the adoption of new technologies. Relationship Consultants are typically financial institution representatives. The change in required skills sees decreasing emphasis on financial services skills and on assessments of credit rating—both areas in which technological advances have supplemented human labour. However, there is renewed emphasis on industry-specific knowledge and skills: Banking, Finance and Lending, as well as a stronger emphasis on specialized analysis skills: Financial Skills.

Too often, labour market analysis is focused on the overall gains and losses in jobs. The changing skills genome of roles points to another vector of disruption experienced by workers today: the changing requirements of workers whose roles are undergoing significant transformation. Tracking the skills content of jobs provides more accurate insights into the requirements of labour market than job roles alone. To track the most significant vectors of change in today's labour markets, skills need to become the new unit of analysis in the labour market.

Beyond these examples, LinkedIn's skills genome analysis can track the skills composition of clusters of roles, industries, cities and regions for the purposes of comparing historical trends and contrasting skills profiles. Such measures can therefore provide the tools to identify reskilling and upskilling priorities across industries. The metric can be expanded to match emerging jobs to populations with the right skillsets to fill those new roles such as migrant populations or women. It can also be expanded to understand entry into the labour market for new entrants to professions and industries today and in the recent past.

Implications for decision-makers

Making skills the key unit of analysis in the labour market provides policy-makers with the tools to proactively prepare for a disruptive future of work. An in-depth, granular and dynamic understanding of unfolding changes in the labour market requires us to reach beyond analysing occupations and educational qualifications achieved at a young age.

Policy-makers and business leaders can use such granular, dynamic data to target skills gaps that lead to socio-economic inequality between citizens by characteristics such as gender, race and ethnicity. In addition, such skills comparisons can be used to identify areas in which similar skills profiles suggest potential synergies—such as the similarity between occupations that can determine future career paths for workers and occupations with declining prospects. Finally, such data can support education and training providers in iterative efforts to align professional qualifications for the skills requirements observed in the labour market in a more agile fashion.

Source Data and Model

The data used in developing the LinkedIn Skills Genome is drawn from the member profiles of professionals on the LinkedIn platform. Specifically, the data is sourced from the skills listed on a member's profile, the positions that they hold and the locations where they work. The metric is calculated using a term frequency–inverse document frequency (TF-IDF) model.

In order to compute this metric, LinkedIn first calculates a weight for each skill based on the prevalence of that skill in a particular segment, such as a particular geography, industry or occupation, and compares it to other segments of the labour market. For example, one may wish to determine the top representative skills of an occupation relative to other occupations during a set period. First, all members who hold the occupation during the relevant period are included in the analysis. Then a frequency measure is assigned to each skill by calculating the number of times members list the skill under the "skills" section of their LinkedIn profile. Note that skills are only included in the analysis if they were specifically added during the period for which the individual has held a that position. The skills that are added by fewer than or equal to 10 members during the pre-defined period are dropped to reduce 'noise' in the skills data. This ensures that skills are only captured if they are relevant to the role and enables a comparison between skills profiles over time. Finally, each occupation-skill pair is weighted; skills that are generic and appear in multiple occupations are down-weighted. The result is a list of skills that are most representative of that occupation.

A Look at Future Demand for Data Science Jobs

In 2018, through the Future of Jobs survey, Chief Human Resources Officers and other business leaders were asked to identify the technologies that they were looking to integrate across their enterprises and the roles that would be important and new. The results show the relevance of data science jobs and skills. Out of the business leaders surveyed, 85% indicated that in the period up to 2022 they planned to integrate User Entity and Big Data Analytics, making this technology the most prioritized across all industries.

In addition, leaders identified a range of newly important roles that are forecasted to become more prominent in that same period. The table below shows the rank order placed on Data Analyst and Scientist and on Artificial Intelligence and Machine Learning Specialists across industries. Most industries' business leaders planned to match their investment in User and Entity Big Data Analytics technologies with the creation of new data science related positions. More precisely, for seven out of the 12 surveyed industries, Data Analysts and Scientists was the top-ranked fully new type of role that business leaders planned to hire leading up to 2022.

To match this demand, policy-makers and business leaders will need to work together to expand the capacity of data science talent today. The data collaborations in this Report reveal the need for dynamic and responsive reskilling and upskilling programmes in addition to fundamental education around data science skills. Such investment is set to generate rapid returns for individuals, companies and economies in the form of higher wages, innovation, competitiveness and socio-economic mobility.

Ranking priority order of roles which require data science skills, by industry

Industry	AI and Machine Learning Specialists	Data Analysts and Scientists
Automotive, Aerospace, Supply Chain and Transport	2	1
Aviation, Travel and Tourism	4	2
Chemistry, Advanced Materials and Biotechnology	2	6
Consumer	3	1
Energy Utilities and Technologies	8	1
Financial Services and Investors	2	1
Global Health and Healthcare	3	1
Information and Communication Technologies	2	1
Infrastructure	33	2
Mining and Metals	8	4
Oil and Gas	14	1
Professional Services	11	5

Source: *Future of Jobs Report 2018*, World Economic Forum.

Acknowledgements

The Centre for the New Economy and Society is the global platform for building prosperous, resilient and inclusive economies and societies that provide opportunity for all. The Centre helps leaders understand emerging socio-economic trends, shape the future through new models and standards, and act together collaboratively to address key challenges. The Centre's New Metrics CoLab aims to empower decision-making among leaders in business and policy by providing fresh, actionable insight into key questions on the socio-economic agenda through collaboration with leading experts and data-holding companies. The CoLab is advised by the Global Future Council on New Metrics in 2018–2019.