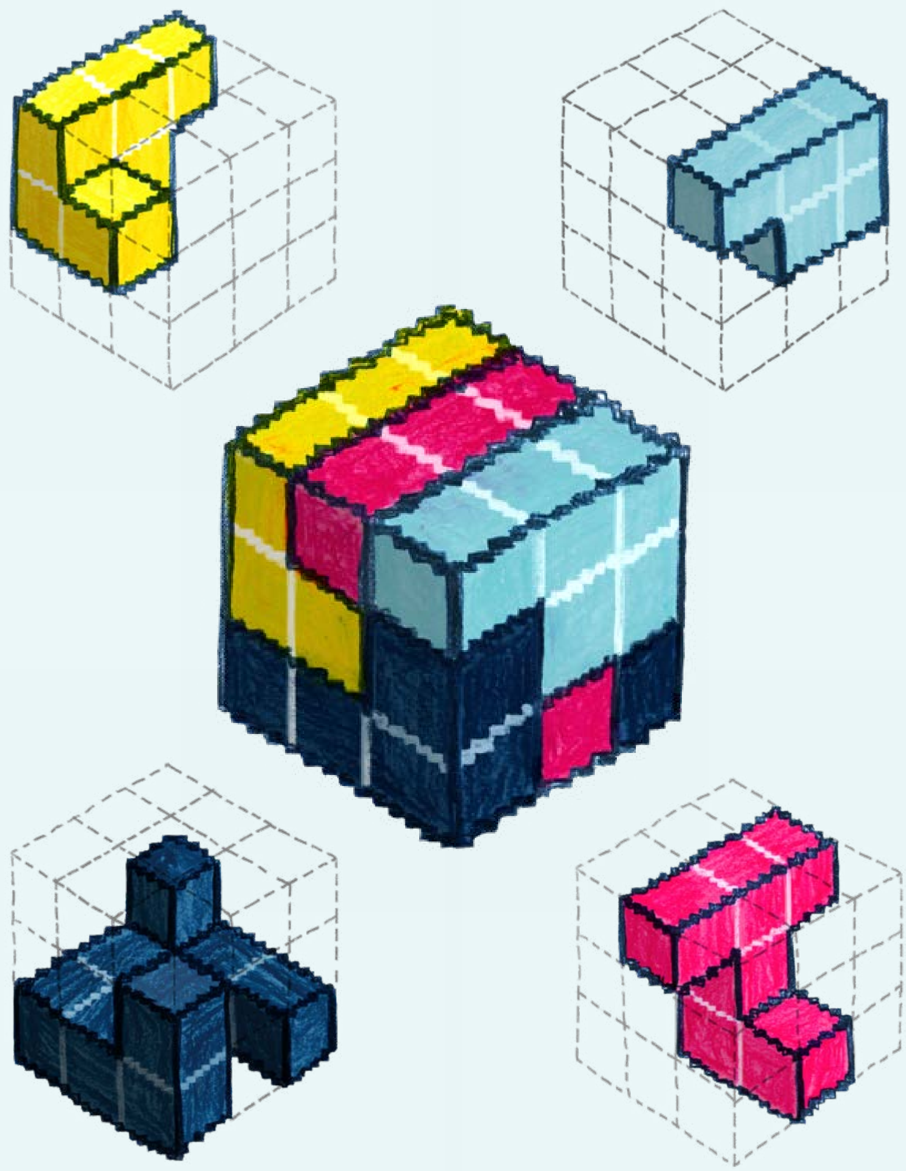
Digital, Defined:

Understanding the demand for digital skills in Canada

December 2019



brookfield institute for innovation + entrepreneurship

burningglass TECHNOLOGIES®

THE ANATOMY OF DIGITAL SKILLS DEMANDS IN CANADA

INTRODUCTION

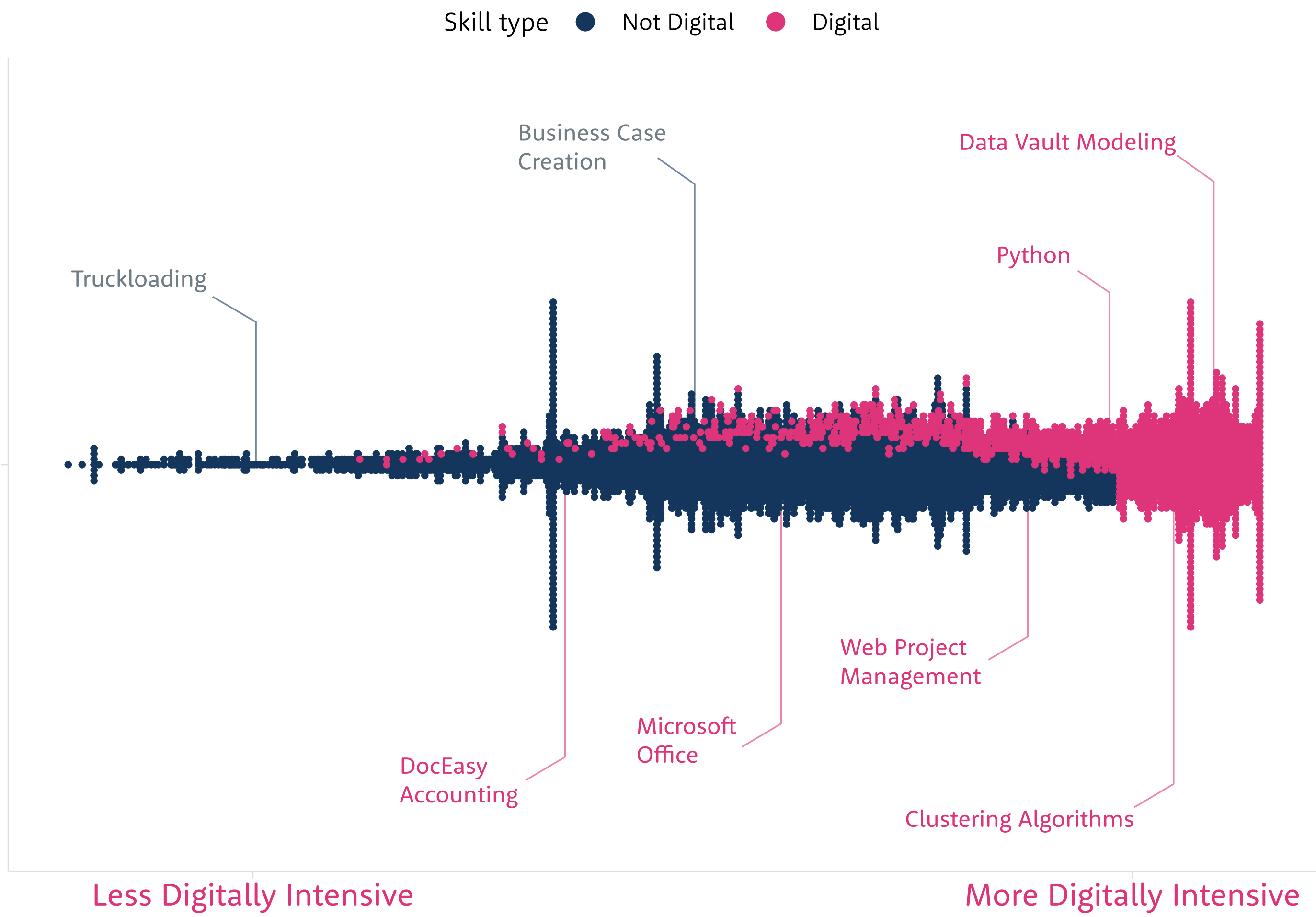
In the following pages, we outline our approach to defining digital skills based on a skill's digital intensity. Using this novel approach, we were able to define 3,600 unique digital skills. Each of these skills exists on a spectrum depending on their overall digital intensity. This spectrum ranges from less digitally-intensive skills, such as general spreadsheet skills, to much more digitally-intensive skills, such as those involved in machine learning.

We also examine the digital skills Canadian employers ask for the most.

Unsurprisingly, the most in-demand digital skills are less digitally-intensive, in particular those associated with the Microsoft Office Suite software. However, many employers are also looking for much more digitally-intensive skills related to data analysis, including SQL and SAP, indicating the importance of data and data analysis skills in today's economy. Canada's growing digital economy is also reflected in the large number of times employers asked for general software skills, as well as specific programming languages such as Java.

Figure 1

Identified Digital Skills



Source: Author Calculations
Note: Each dot is a skill

THE DIGITAL SKILLS SPECTRUM

Previous Brookfield Institute studies articulate digital literacy as a spectrum of skills, ranging from the baseline digital skills that all Canadians need to participate in an increasingly digital economy, to much more specialized digital skills involved in the creation of new digital products and services.[1] We build off this framework, creating a new demand-driven taxonomy using job postings from Burning Glass. This allowed us to define digital skills along a spectrum based on their digital intensity. Out of the 13,000 unique skills that show up in Canadian job advertisements, 3,600 met our digital threshold.

Figure 1 illustrates the spectrum of digital skills. At the far right are the skills that are the most digitally-intensive digital skills. These include highly technical digital skills and knowledge, such as those associated with: data vault modeling, for long-term storage of data; clustering algorithms, a machine learning technique designed to group data points; and programming languages, such as Python. The least digitally-intensive skills, which still fit within our definition of digital skills, have wider use and include, for example, proficiency in skills related to web-based project management, Microsoft Office, and accounting software.

THE DIGITAL SKILLS EMPLOYERS ASK FOR THE MOST

In addition to examining the digital intensity of skills, we also highlight the digital skills that are most in-demand from Canadian employers. Unsurprisingly, by sheer volume of mention, employers are looking for software skills associated with general office tasks, in particular use of the Microsoft Office Suite. These skills have a low digital intensity. However, other more digitally-intensive general office skills are also in high demand, including SAP, an enterprise resourcing software.

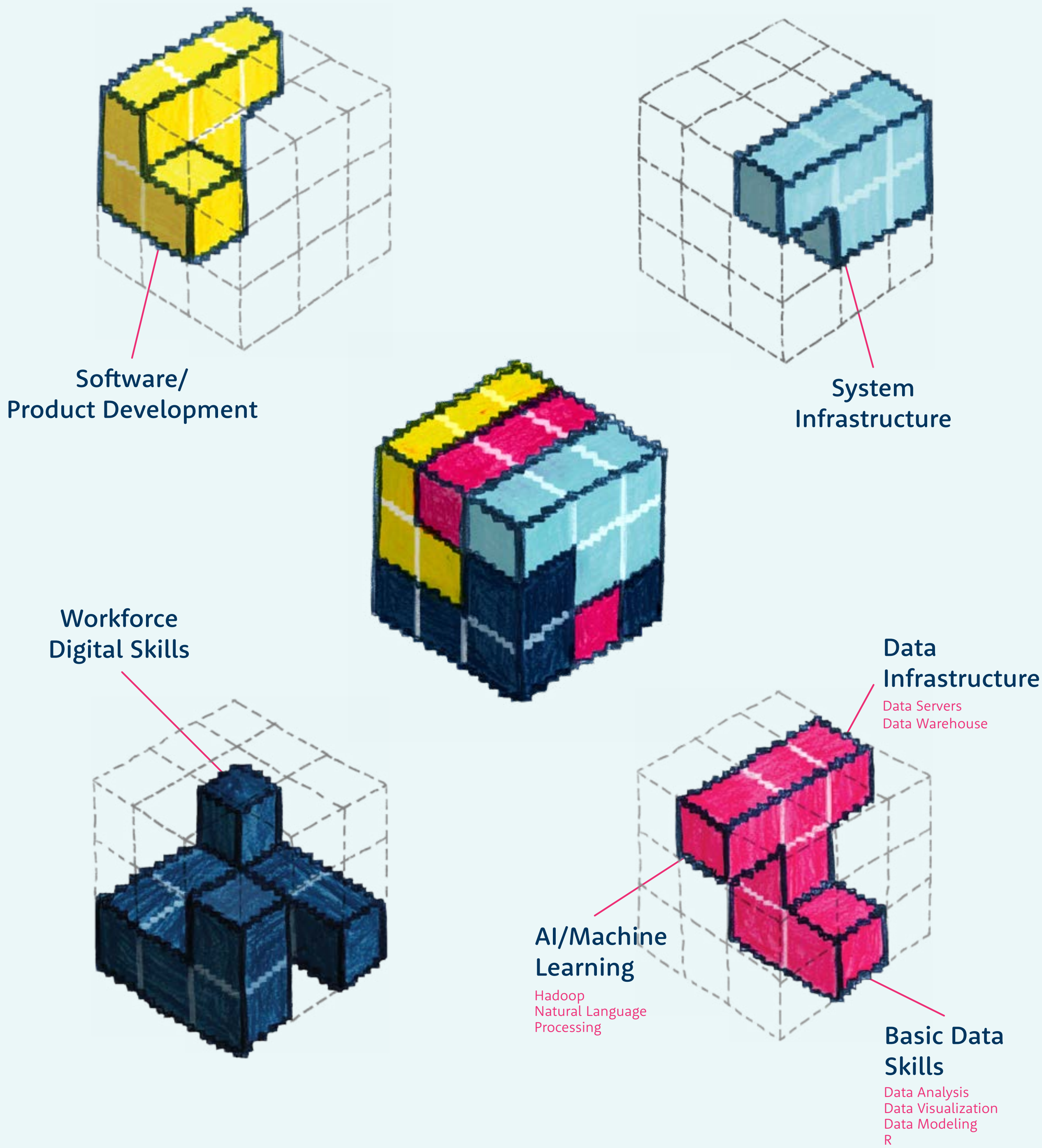
Interestingly, Microsoft Excel was mentioned almost three times more than Microsoft Word. This is likely driven by employers' assumptions about a worker's baseline level of knowledge in specific roles, but also reflects the importance of data skills across sectors — a conclusion that is reinforced by the high frequency at which general spreadsheet skills and the more digitally-intensive data software, SQL, was mentioned.

The importance of more digitally-intensive skills across Canada's economy is also reflected in the number of times employers asked for general software skills, as well as knowledge of specific programming languages such as Java. [2]

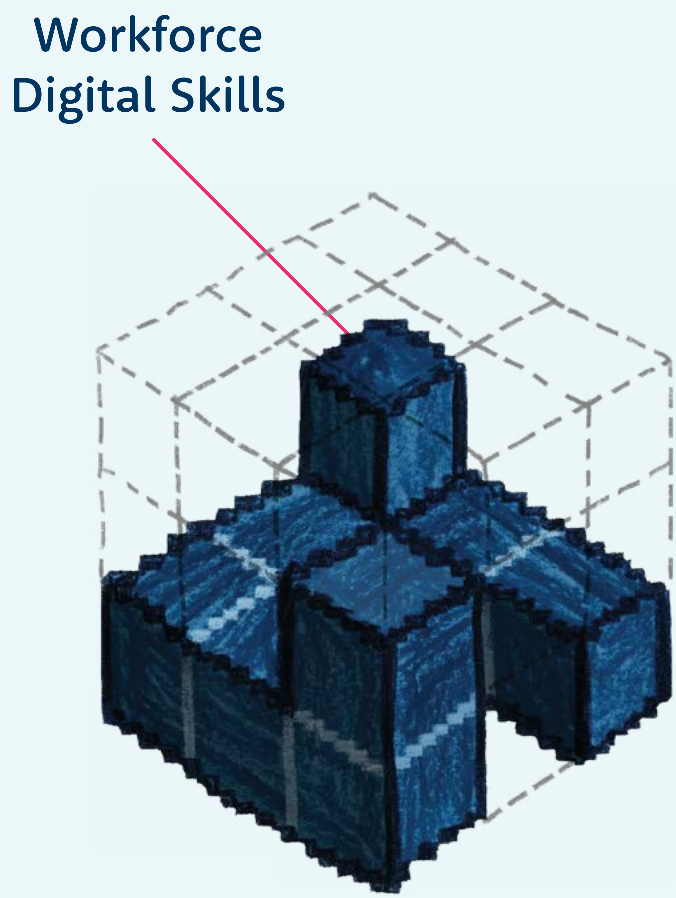
Table 1: **Top 10 Digital Skills by Number of Job Postings That Mention Them**

Skill	Description	Number of mentions	Digital intensity classification
Microsoft Excel	Spreadsheet software	741,191	Low
Microsoft Office	Work productivity software suite	621,690	Low
Microsoft Word	Document composing software	296,992	Low
Microsoft PowerPoint	Presentation creation software	266,748	Low
SQL	Database management software	163,000	High
Spreadsheets	Data organization	151,719	Low
Software Development	General skill	133,681	High
Technical Support	General skill	130,540	Medium
SAP	Business enterprise software	126,787	Medium
Java	Programming language	112,680	High

DIGITAL SKILLS FRAMEWORK



Within the broad category of digital skills exist four unique clusters that show up in job postings together: general workforce digital skills, data skills, system infrastructure skills, and, finally, software and product development skills. While these skills form unique clusters, data skills in particular are so frequently requested in different occupations with varying digital intensity that they also act as a connector between the three other digital skill clusters.

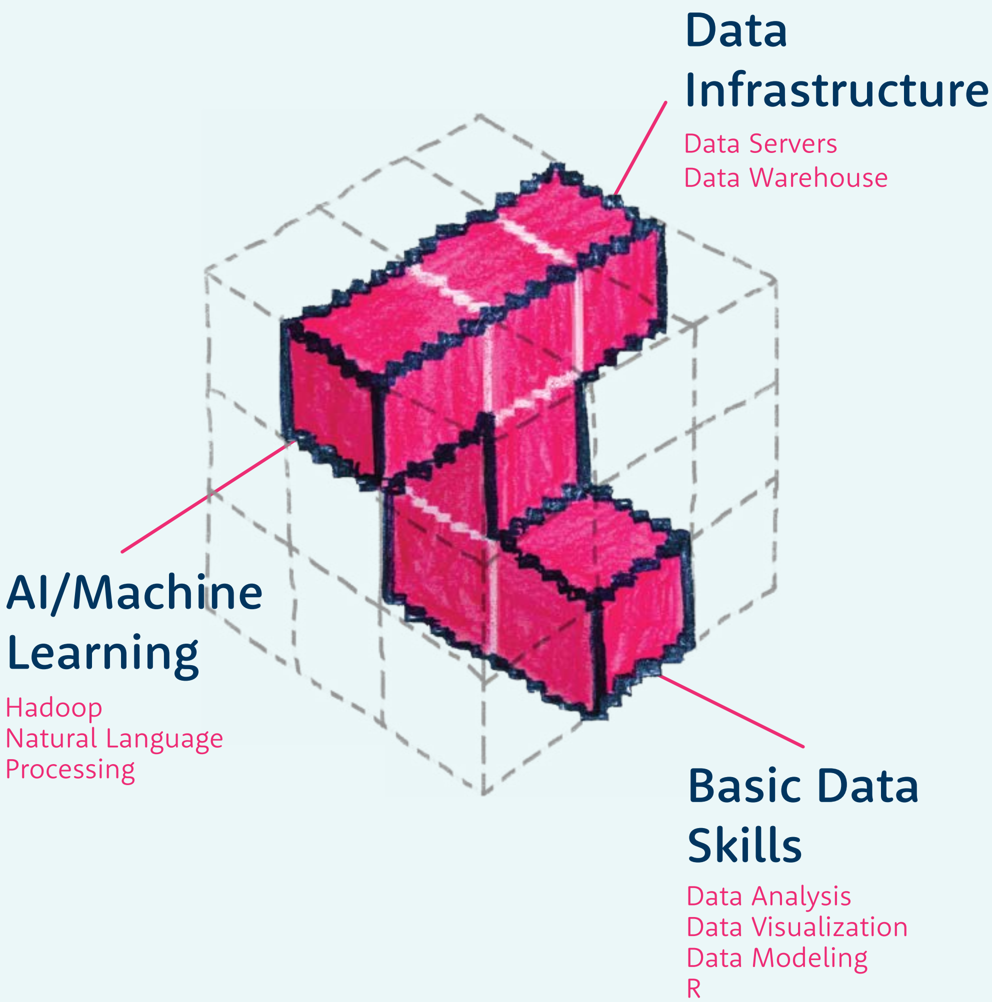


WORKFORCE DIGITAL SKILLS

This skill cluster is the most general and least digitally-intensive, consisting of 997 unique skills used by workers across many industries. The skills range from those associated with general office tasks to those associated with specific professions, such as use of architectural and engineering-based software to augment existing job tasks and business processes.

Some prominent skills in this cluster include Microsoft Excel, Word, PowerPoint, and Office (741,191, 296,992, 266,792, and 621,690 mentions respectively). This cluster also includes skills associated with some general-use design software, such as Adobe Photoshop (53,855 mentions), as well as general data analysis skills (mentioned 60,256 times) and use of tools such as SAS (21,130 mentions).

Based on the average weight of our previously defined digital intensity scores, skills in this cluster are the least digitally-intensive, and most in-demand across the economy.

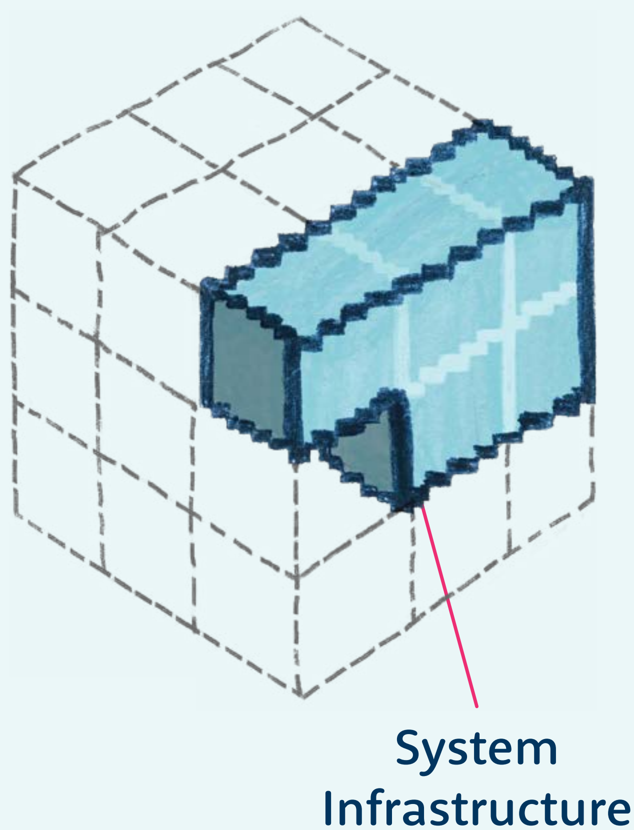


DATA SKILLS

This skill cluster consists of 507 unique skills focussed primarily on data gathering and analysis, especially in large-scale enterprise analytics. Some prominent skills in this cluster include “Data Modeling” (20,252 mentions), “Big Data” (13,173 mentions), and “Business Intelligence” (35,361 mentions), as well as skills associated with specific data analytics tools, such as Apache Hadoop (10,509 mentions), Tableau (9,121 mentions), and R (4,132 mentions).

Data skills have a unique sub-cluster structure, illustrated above. There is a strong separation of basic data skills, which straddle the Data and Workforce Digital Skills clusters, and advanced data skills, where two further sub-clusters of skills exist: Data Infrastructure, which exists alongside the System Infrastructure cluster, and AI/Machine Learning skills, which appears alongside the Software/Product Development cluster.

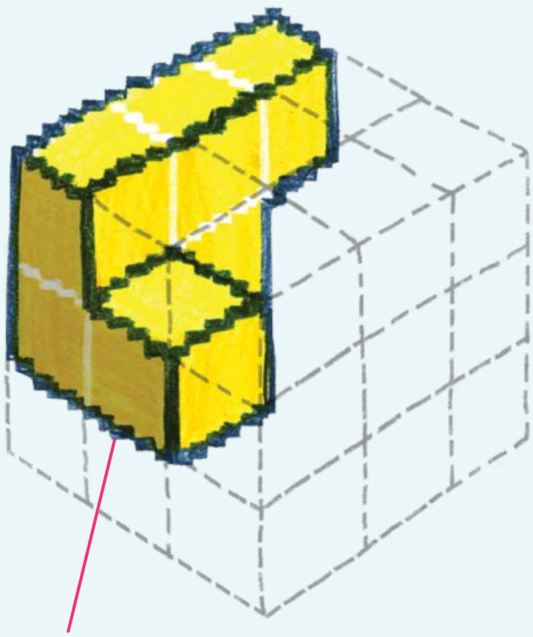
Data skills are important across the economy and have a wide variance in their digital intensity. As a result, the overall digital intensity of this cluster exceeds only the Workforce Digital skills cluster.



SYSTEM INFRASTRUCTURE SKILLS

This skill cluster consists of 985 unique skills that pertain to digital infrastructure management, ranging from setting up and managing cloud computing services to more general IT support. Some prominent skills in this group include proficiency with specific platforms such as VMWare (25,319 mentions) or Windows Server (21,094 mentions), and general support skills such as 'system administration' (33,459) and 'hardware and software installation' (23,940 mentions).

This cluster is the second most digitally-intensive. The overall digital intensity of this cluster is lower than that of the Software/Product Development Skills cluster because of the general IT support skills that are included.



Software/
Product Development

SOFTWARE / PRODUCT DEVELOPMENT SKILLS

This skill cluster consists of 1,109 skills that pertain to the generation of new digital products, both web- and software-based. Some prominent skills in this group include proficiency in specific programming languages, such as Java (112,680) and Python (43,137 mentions), and general skills such as 'software development' (133,681 mentions), 'software engineering' (47,775 mentions), and 'web development' (41,184 mentions). Some technical design skills, pertaining specifically to web development, are also a part of this cluster. On average, skills in this cluster are the most digitally-intensive.

DATA OVERVIEW

This study relied on data from Burning Glass, an analytics software company that analyzes data from hundreds of millions of online job postings across the globe to provide real-time information on the jobs and skills that employers are looking for. This data can help to close many of the gaps left by traditional sources of labour market information (LMI), in particular when it comes to timely data on skills demands across the economy.

For this report, we used six years of online job postings covering nearly all English-language online job postings in Canada. The first job posting considered for this report was posted on January 5th, 2012, and the last job posting considered for this report was posted on December 31st, 2018.

In total, there were more than 7 million job postings, 6.6 million of which were directly connected to an existing occupational group. These 6.6 million job postings were used as the main unit of analysis in the report.

Anatomy of a job posting

Throughout this report, we relied on job postings data gathered from Burning Glass as the basis of our research. This is a unique dataset that has several important features. Most importantly, job postings characterize a flow (potential new workers), as opposed to the stock of current workers. As a result, the number of job postings may or may not reflect the number of people working in such roles.

Secondly, a job posting is not a perfect characterization of the skills involved in performing a specific role. It is a reflection of what the hiring party (such as employers, human resource departments, and recruiters) believes the role entails.

Therefore, it must be stressed that any insights coming from job postings data reflect employer beliefs about jobs and skills needed for their organization at the time when the posting is made — or their wish list. Employers may not expect that the successful candidate will possess all of the listed skills. Additionally, the content of job postings varies; some may be very detailed, while others offer comparatively limited information.

The Opportunity:
The Brookfield Institute for Innovation and Entrepreneurship is seeking an ***Economist*** to produce quantitative and qualitative research, analysis and written products to help support building a robust, data driven and user experience driven picture of Canada's innovation and entrepreneurship policy landscape and of how Canadians from different communities interacts with this landscape. The Economist will work with the Director and the broader team to develop actionable policy recommendations and to inform the design of pilots, programs and other initiatives related to innovation and entrepreneurship.

> The Job Posting states the company, as well as the position being hired.

Qualifications:
Successful completion of a post-secondary ***degree program in Economics*** or relevant field of study required.
A minimum of ***2 years of relevant professional work experience***, including:
Making a major contribution to designing and conducting quantitative research, analysis and writing in a full-time role.
Directly or indirectly supporting the communication of research results to external audiences
Working in a fast-paced environment with multiple and overlapping deadlines
Experience working with large datasets
Using statistical software, such as ***Stata, SPSS, Python, R***, or ***other equivalents***
Experience working with large data sets and using statistical software such as Stata, SPSS, Python, or other equivalents
Proven ability to gather, synthesize and analyze complex information from a variety of sources, to identify trends in data and research, and to develop specific recommendations; strong logical reasoning and intellectual curiosity
Ability to apply sophisticated statistical, mathematical and computational methods to clean and analyze quantitative data
Excellent ***time management, organizational and problem-solving skills***
Strong understanding of the policy process, and of government players, processes and programs
Ability to work professionally with a wide range of stakeholders, including government officials, business leaders, academics, entrepreneurs and others
Ability to clearly communicate complex issues, and ***make research accessible*** to maximize the impact of the Institute's work
Knowledge of key issues relating to innovation and entrepreneurship in Canada and an understanding of the Institute's mission
Knowledge of and ability to apply the principles of macro and microeconomics
Advanced statistical and econometrics knowledge
Knowledge of available Canadian economic data including
Computer proficiency, including the use of the ***Microsoft Office Suite.***

> It also lists educational credentials, and previous experience.

> Specialized digital tools/softwares are listed.

> Specific non-digital skills are enumerated.

> Some skills may be implied but not explicitly listed (e.g. Data Visualization)

> Specialized knowledge may also be required.

Representativeness of the data

Job postings were collected for all 13 provinces and territories in Canada; however, only data from English-language postings was collected, due to the platform — at this stage — being optimized only for processing job postings in English. This is reflected in the distribution of job postings. For Quebec, the inability to parse job data in French meant that the share of job postings data was much lower than the province's overall share of employment.

Table 2: Distribution of Job Postings Captured by Burning Glass across Canada

Province/Territory	Number of Job Postings (cumulative 2012-2018)	Share of Burning Glass Job Postings	Share of Employment [12]
British Columbia	1,071,217	14.9%	13.5%
Ontario	2,946,740	41%	39.0%
Manitoba	193,378	2.7%	3.4%
Alberta	1,103,817	15.3%	12.3%
Nova Scotia	203,508	2.8%	2.4%
Quebec	1,044,653	14.5%	22.8%
Saskatchewan	363,185	5%	3%
New Brunswick	129,694	1.8%	1.9%
Newfoundland and Labrador	85,515	1.2%	1.2%
Northwest Territories	9,117	0.1%	<0.1%
Yukon Territories	7,510	0.1%	<0.1%
Prince Edward Islands	28,395	0.4%	0.4%
Nunavut	6,254	0.1%	<0.1%

To test the representativeness of this dataset, we examined how Burning Glass job postings data corresponds to Statistics Canada's Job Vacancy and Wages Survey (JVWS). The JVWS captures job openings regardless of whether they are posted online or offline. However, some sample variation is inevitable as the JVWS has a smaller (random) sample size compared to Burning Glass. The JVWS also does not capture the kinds of skills employers are looking for. JVWS was first conducted in 2015 and is conducted quarterly, every January, April, July, and October. As such, we also compiled job postings data captured by Burning Glass in those four months, for four years: 2015, 2016, 2017, and 2018. Specifically, we looked at the share of job postings in both Burning Glass and JVWS that came from one of Canada's major occupational groups.[3]

Overall, the job postings data in Burning Glass corresponds well with the JVWS. However, there are some discrepancies. For example, Burning Glass has considerably fewer job postings, in terms of proportion, from sales and service occupations, while having a higher share of jobs from manufacturing, natural and applied sciences, and management occupations. The correlation between the two series is 0.8775.

We also compared the share of job postings in an occupation to the number of people working in that occupation according to the 2016 Census. Burning Glass data has a lower share of postings in sales and service occupations, and a higher share of postings in natural and applied sciences and related occupations. Apart from these discrepancies, the two sources are remarkably similar.

We also examined whether job postings disproportionately represent more- or less-skilled workers in Canada. As a proxy, we compared the education credentials asked for by employers in job postings to the educational attainment of individuals working in those same occupations using the 2016 census. As a blunt instrument, we calculated the share of workers and job postings with a bachelor's degree or above.

Looking at the distribution of degrees, most occupational groups are similar between the two data sources. However, three occupational groups stand out, all of which have a higher share of postings in Burning Glass asking for bachelor's degree or above. These are management occupations, natural and applied sciences and related occupations, and business, finance, and administration occupations.

From these series of representativeness checks, we can conclude that, overall, Burning Glass is a fairly representative snapshot of the Canadian labour market, despite missing French language job postings. There are some areas of overrepresentation, in particular when it comes to sales and service and natural and applied science occupations, and those that require degree credentials. The technical appendix gives further details surrounding representativeness checks on the data.

DEFINING DIGITAL SKILLS

To understand the universe of skills that Canadian employers are looking for, we first sought to refine our understanding of what makes a skill digital. It's clear that when an employer asks for proficiency in a specific software, they are asking for a digital skill. However, other skills, which are decidedly digital, such as machine learning, might not list specific programming languages, but knowledge in at least one is implied. Other cases are even more complex — for example, should 'social media', a skill in our sample that is highly associated with the use of digital tools, but isn't a digital tool itself, be considered digital?

To address this challenge, we began by examining skills that were clearly digital in the data, using existing taxonomies in Burning Glass data. We then augmented this analysis by identifying the skills that consistently show up in digitally-intensive occupations,[4] labelling those above a certain threshold as digital. We devised several robustness checks to minimize the risk of non-digital skills being classified as digital skills. Consistent with previous studies, this enabled us to examine digital skills along a spectrum based on their digital intensity.[5]

Digital skills as a spectrum

First, we leveraged existing skills taxonomies devised by Burning Glass to define a set of core digital skills, which included software, as well as skills in the following pre-assigned clusters: Information Technology, Analysis, E-Commerce, Web Analytics, and Bioinformatics.[6] The clusters were chosen through manual examination of the skills that were clearly digital.

Second, to construct a digital intensity measure for every skill, we leveraged the existing measure of digital skill intensity across 4-digit National Occupation Classifications (NOCs), used to define tech occupations in Vu, Zafar, and Lamb (2019). This digital intensity measure ranks occupations based on six job attributes that relate to technology use or production from the US O*NET database. If a skill frequently appears only in highly-digital jobs, it will have a high digital intensity score, and vice versa. Specifically, we found the average technical rank of occupations that each particular skill appeared in. We then tested this approach in various ways, detailed in full in the appendix of the report: *I, Human*, which is available as a separate download.

Once we identified the digital intensity for each of the 13,000 skills in the database, we defined a cutoff that would create two distinct categories: digital skills and non-digital skills. To do so, we built a logistic regression that estimated the probability of a skill, with a particular digital intensity, appearing within one of Burning Glass's previously defined core digital skills clusters and our additions. If a skill had a 50 percent or more probability of appearing within one of these defined digital skills clusters, it was classified as a digital skill. We also tested our analysis with a lower cut-off; none of the results of this report were affected by that change.

Skills clustering

Using our definition, we examined what combinations of digital skills employers are looking for and how they overlap. Here, we examined only connections between digital skills; non-digital skills were not yet considered. A common approach in analyzing networks involves understanding the network's community structure.[7] A community is a group of skills that have many connections to one another in the same community, and not many connections to skills in other communities.

The communities in a network allow us to understand important divisions between elements in that network — in our case, each skill represents an element. In practice, there are many different conceptual frameworks and specific algorithms that can be used to define communities. [8] However, we restricted our choice based on the data and our research focus.

Specifically, in the network of skills available through Burning Glass, the strength by which two skills are connected depends on how frequently these skills show up in each job advertisement together. As a result, connections between general skills such as 'communications' showed up many more times than between two specialized, but related, skills, such as 'Python' and 'machine learning'. Ideally, our choice of community detection algorithm should deemphasize the importance of these often-mentioned general skills, and avoid letting them define the communities.

Additionally, as our network was fairly large, algorithms with a straightforward computational complexity were also prioritized.[9] This led us to select a class of algorithms known as modularity, which measure how likely it is that a given community structure is observed if all the skills in the network are connected at random. However, modularity-based measures have a number of limitations, such as a resolution problem,[10] which describes their inability to recognize communities smaller than a specified size.

We have addressed this issue in two ways. [11] First, we performed the community detection algorithm separately for the group of digital and non-digital skills identified. As the resolution problem depends on the size of the input network, this division reduced the size of the individual network being clustered, allowing us to examine smaller communities. Second, we used two different modularity-based algorithms that employ different strategies to find the communities. Using two algorithms enables a more granular understanding of different skill clusters and how they relate to each other.

ENDNOTES

1. Huynh, A., & A. Do. 2017. "Digital Literacy in a Digital Age." Brookfield Institute.

2. The inclusion of SQL and Java in the list of most commonly cited digital skills runs contrary to previous research that indicated the most common language developers in Canada use is Javascript, with 69% of professional developers surveyed by Stackoverflow in Canada using that language. SQL and Java were second (54.6%) and fourth (39.0%) respectively. However, this could simply reflect a discrepancy in the data, since in a job advertisement that asks for developers proficient in Javascript, it is much more common to list skills associated with different ways of implementing Javascript, such as jQuery or node.js, as opposed to just asking explicitly for Javascript. This could also reflect trends in the demand for skills, since Burning Glass data shows that Java demand is stable or slightly increasing, whereas Javascript is increasing in demand.

3. Further disaggregation of occupational group is possible, though not advisable, as the JVWS becomes less precise due to sample size.

4. Vu, V., A. Zafar, & C. Lamb. 2019. "Who Are Canada's Tech Workers?" *Brookfield Institute.* Muro, M., S. Liu, J. Whiton, & S. Kulkarni. 2017. "Digitization and the American Workforce." *Brookings Institute* (November). Gallipoli, G., & A. Makridis. 2018. "Structural Transformation and the Rise of Information Technology." *Journal of Monetary Economics* 97: 91–110.

5. Huynh, A., & A. Do. 2017. "Digital Literacy in a Digital Age." *Brookfield Institute.*

6. For more information on Burning Glass's taxonomy of skills, see: https://www.burning-glass.com/research-project/skills-taxonomy/

7. Fortunato, S. 2010. "Community detection in graphs." *Physics Reports* 486(3–5): 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

8. Reichardt, J., & S. Bornholdt. 2006. "Statistical mechanics of community detection." Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 74(1): 1–16. https://doi.org/10.1103/PhysRevE.74.016110
   Newman, M. E. J., & M. Girvan. 2004. "Finding and evaluating community structure in networks." Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 69(22): 1–15. https://doi.org/10.1103/PhysRevE.69.026113
   Clauset, A., M. E. J. Newman, & C. Moore. 2004. "Finding community structure in very large networks." Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 70(6): 6. https://doi.org/10.1103/PhysRevE.70.066111
   Blondel, V. D., J. L. Guillaume, R. Lambiotte, & E. Lefebvre. 2008. "Fast unfolding of communities in large networks." Journal of Statistical Mechanics: Theory and Experiment 2008(10): 1–12. https://doi.org/10.1088/1742-5468/2008/10/P10008

9. Fan, R., K. Xu, & J. Zhao. 2017. "A GPU-based solution for fast calculation of the betweenness centrality in large weighted networks." PeerJ Computer Science 3(2007). https://doi.org/10.7717/peerj-cs.140

10. Fortunato, S., & M. Barthe. 2006. "Resolution Limit in Community Detection." *Pnas* 104(1): 36–41. Retrieved from www.pnas.orgcgidoi10.1073pnas.0605965104.

11. Fortunato, S., & M. Barthe. 2006. "Resolution Limit in Community Detection." *Pnas* 104(1): 36–41. Retrieved from www.pnas.orgcgidoi10.1073pnas.0605965104.