RESEARCH PAPER SERIES NO. 2020-04

Assessing the Alignment of Data Science and Analytics (DSA)-Related Undergraduate Programs with the Emerging Demands for DSA Workforce

Brenda A. Quismorio
Maria A.D. Pasquin
Claire S. Tayco

Research Paper Series No. 2020-04

Assessing the Alignment of Data Science and Analytics (DSA)-Related Undergraduate Programs with the Emerging Demands for DSA Workforce

Brenda A. Quismorio, Maria Antonette D. Pasquin, and Claire S. Tayco

Philippine Institute for Development Studies
Surian sa mga Pag-aaral Pangkaunlaran ng Pilipinas

Published by
Philippine Institute for Development Studies

Printed in the Philippines. Some rights reserved.

Please address all inquiries to:

Philippine Institute for Development Studies
18th Floor, Three Cyberpod Centris - North Tower
EDSA corner Quezon Avenue, 1100 Quezon City
Telephone: (63-2) 8877-4000
Fax: (63-2) 8877-4099
E-mail: publications@mail.pids.gov.ph
Website: https://www.pids.gov.ph

ISSN 1908-3297
ISSN 2508-0830 (electronic)
RP 10-20-600

Editorial and production team:
Sheila V. Siar, Maria Judith L. Sablan, Carla P. San Diego, and Rejinel G. Valencia

Table of Contents

List of tables and figures

List of Acronyms

AAP	Analytics Association of the Philippines
APEC	Asia-Pacific Economic Cooperation
BHEF	Business Higher Education Forum
CHED	Commission on Higher Education
CMO	CHED memorandum order
DICT	Department of Information and Communications Technology
DOLE	Department of Labor and Employment
DSA	data science and analytics
DTI	Department of Trade and Industry
EDSF	EDISON Data Science Framework
E-PJN	enhanced Phil-Job.net
ETL	extract, transform, load
FGD	focus group discussion
FIRe	Fourth Industrial Revolution
GDP	gross domestic product
HEI	higher education institution
IT	information technology
IT- BPO	information technology - business process outsourcing
IT-BPM	information technology - business process management
NMF	nonnegative matrix factorization
NPC	National Privacy Commission
PIDS	Philippine Institute for Development Studies
PQF	Philippine Qualifications Framework

PSA	Philippine Statistics Authority
PSIC	Philippine Standard Industry Code
PwC	Pricewaterhouse Cooper
STEM	science, technology, engineering, and math
TF	term frequency
TF-IDF	term frequency - inverse document frequency
UA&P	University of Asia and the Pacific
UPD	University of the Philippines - Diliman
US	United States

Abstract

The rapid advancement in technology has facilitated far-reaching use of data. This has consequently led to an increasing demand for data science and analytics (DSA) professionals. However, recent studies show that such demand is often not met in many economies. This shortage is claimed to be rooted in the mismatch between the skills the industry demands and the skills academic institutions supply. This is particularly true in the Philippines, where studies also reveal the challenges faced by Philippine education and training in meeting the level of competencies required to do high-skilled jobs. An indicator of this weak point is the persistent high youth unemployment and underemployment rate, wherein graduates land jobs which their completed education did not intend for them.

As the first step in addressing this shortage, it is necessary to know the DSA skills demanded by the industry and those which academic institutions equip their students. To do this, the study employed the Analytics Association of the Philippines' (AAP) Professional Maturity Model as analytical method. This model is based on the 10 DSA competencies recommended by the Asia-Pacific Economic Cooperation. Through online job scraping, the study estimated the percentage distribution of professionals equipped with DSA competencies for DSA job roles sought by employers. This was then juxtaposed with the percentage distribution of higher education institution (HEI) graduates equipped with DSA competencies for DSA job roles, which, on the other hand, was estimated using available statistics from the Commission on Higher Education (CHED) and the sample curriculum evaluations of academic experts.

In terms of the demand side, the functional analyst job role emerged as the most sought role of employers among the four DSA-related job roles. The demand for these roles also mostly came from the information and communication industry group under the services sector. For each demanded job role, the employers required differing DSA competencies and levels of proficiency. One challenge for employers in hiring DSA talents was the inadequate 21st century skills, more than other technical skills.

Regarding the supply side, the top 10 undergraduate programs were identified to be DSA-related programs based on the common degrees of surveyed analytics practitioners. The extent to which these DSA-related programs prepare their respective graduates with the basic proficiency

of each competency varied. However, the curricula assessments revealed that these degree programs mostly equipped their graduates with competencies on data engineering and statistical techniques and correspondingly enabled them to perform data engineer job role.

With these results, the study revealed a misalignment between the demand and the supply of DSA workforce in the country. Specifically, there were DSA competencies—both sought by employers and required by the analytical framework of the study—that were poorly supplied by these DSA-related undergraduate programs. Moreover, while employers mostly looked for functional analysts, most of the graduates of the degree programs were equipped to perform the data engineer job role.

The study offers two recommendations. First, it recommends the use of the AAP Professional Maturity Model to define the DSA profession. This addresses the lack of a common definition of the analytics profession among stakeholders. Having a common understanding can facilitate efforts to align supply and demand and eventually the maturity of the industry. With the first recommendation as starting point, the study also recommends the promotion of government-industry-academe linkages to expand the existing market for DSA workforce in the country.

Introduction

The rapid and continuous advancement in technology allowed big data[1] to be more accessible. McKinsey notes that "data [have] become widespread across industries that [they] can now be considered as an important factor of production [along] with labor and capital" (Tableau 2016, p. 2). Because of this explosion of data, the field of data science and analytics (DSA) has also gained much relevance and attention. "Companies of all sizes rely on [DSA] as key transformational components to their core operations (BHEF 2014, p. 1)." The recent study of Microsoft about the digital transformation in the Asia Pacific showed that big data and analytics topped as the third most invested technology in 2018 (Microsoft 2018).

This growing interest in DSA consequently causes an uptrend in demand for professionals from this field. However, such demand is often not met. Developing a workforce with DSA skills continues to be a challenge for many economies (APEC Human Resource Development Working Group 2017; PwC and BHEF 2017). Next to cybersecurity skills, DSA skills were identified by 95 percent of employers as most difficult to find among prospective employees in the United States (US) (PwC and BHEF 2017).

The DSA skills shortage can be traced to the mismatch between the skills that the industry demands and the skills that academic institutions supply. Most higher education data science programs are lodged under the school of engineering or the department of computer science (BHEF 2014). What makes this practice quite problematic is that it may prepare data science experts who, while possessing the quantitative skills, still lack the domain-specific knowledge in health, transportation, economics, business, and public policy that companies need (BHEF 2014). At the same time, this may contribute to having a workforce who may only know the business domain but lack the technical skills. Hence, employers usually encounter difficulties looking for the right DSA talents, perpetuating the current problem of DSA skills shortage.

As this skills mismatch may lead to economic losses, initiatives have been started to equip the future workforce with DSA skills (APEC Human Resources Development Working Group 2019). For instance, the

[1] As cited in Dadios et al. (2018), a scoping study on the Fourth Industrial Revolution, there is yet no universal consensus on the definition of "big data". Nonetheless, it is understood by most players in the industry as "the digital data sets that are so large (either in terms of scale, i.e., 'volume', or streams across time, i.e., 'velocity') or complex (in terms of variety)", which are mostly sourced from use of digital gadgets or social media, etc.

US has started to gear up its universities with as many DSA degree programs (Tableau 2016). The Asia-Pacific Economic Cooperation (APEC) Human Resource Development Working Group (2017) recommended a set of DSA competencies to serve as a reference for the academe, the industry, and government sectors in institutionalizing DSA.

With the increasing demand for DSA, the lack of employment opportunities for graduates of tertiary education in the Philippines (WB 2016) can be addressed. However, this idea still posts difficulties since DSA skills are still underdeveloped among the workforce in the Philippines (Tan and Tang 2016). In addition, the level of competencies of high-skilled jobs required by industries is not met by current Philippine education and training. Specifically, the Philippines suffers from "outdated curriculum, weakness in STEM [science, technology, engineering, and math] education and lack of soft skills training (including proficiency in English), insufficient collaboration between the educational institutions and industry, inadequate instructional facilities, poor quality of instructors who often do not possess the necessary knowledge or industry experience to impart industry-relevant skills to the students, lack of industry exposure for students, and the absence of a rigorous skill competency certification program" (Tan and Tang 2016, p. 100).

While existing studies recognize the DSA skills mismatch in the Philippines, there continues to be a dearth of assessment studies about the current state of DSA competencies in the country. This, in return, makes policy formulation challenging across relevant stakeholders. This paper thereby addresses such literature gap by examining how the Philippines fares in equipping the workforce for future jobs. Specifically, it provides an introspection of the Philippine undergraduate education curricula on DSA-related undergraduate degree programs insofar as they equip graduates with DSA competencies needed in their future workplace.

Objectives

This paper aimed to assess existing Philippine education curricula of DSA-related undergraduate degree programs insofar as they meet the DSA competency requirements of the industry. In doing so, the paper addressed the main question, *Are the DSA-related undergraduate degree programs able to supply the current demands for DSA talents?,* with the following specific questions:

1. What is the current demand for DSA job roles in the Philippines?
   a. To what extent do current work activities require DSA competencies?
   b. What DSA job roles do companies look for?
   c. What are the profiles of these DSA job roles following employers' expectations and the framework's requirements?
2. What is the current supply of DSA job roles in the Philippines?
   a. What are the DSA-related undergraduate programs?
   b. Which DSA competencies do these DSA-related undergraduate programs equip their graduates with basic proficiency?
   c. Which DSA job roles do these DSA-related undergraduate programs enable their graduates to perform?

Significance of the Study

How can the Philippines prepare its workforce for the Fourth Industrial Revolution (FIRe)[2] or future jobs? This study took the first step in assessing the current ability of Philippine educational institutions in enabling the future workforce for the DSA job roles needed by industries. This study then proposed courses of actions to ensure the availability of adequate supply of DSA workforce. In addition, this paper set the stage for further studies on DSA in the Philippines given the dearth of literature on this subject.

Scope and Limitation

Demand side

Data on the demand for DSA workforce is very scarce given that the field is very new. The data used were mostly sourced from government agencies' published reports on employment statistics and employers' online job postings, which point to the demand coming mostly from the information technology and business process management (IT-BPM) industry. Other industries, though they may also demand for DSA workforce, may have not been captured by the study.

[2] Refer to Dadios et al. (2018) for further explanation of the phenomenon.

Supply side

The 10 DSA-related undergraduate programs were identified from the top undergraduate programs of surveyed DSA practitioners. The possible sampling bias caused by limiting the surveyed DSA practitioners to members of Analytics Association of the Philippines (AAP) and those who attended AAP's meetups was considered minimal given that the analytics profession was not yet defined. At the time of the survey, the people involved in AAP were those who, despite the absence of a formal definition, had a good understanding of the DSA work. Most of them were organizers or attendees of other non-AAP meetups. The 10 DSA-related undergraduate programs that emerged from the surveyed DSA practitioners excluded other undergraduate programs, which likewise equip students with any of the DSA competencies.

Each DSA-related undergraduate program was evaluated by academe experts using the sample curriculum found in the program's corresponding Commission on Higher Education (CHED) Memorandum Order (CMO). The following biases emerged:

1. These DSA-related undergraduate programs did not intend to produce DSA professionals. Specifically, their curricula were designed to equip their students with the competencies needed to perform the tasks of their respective professions. However, this bias was mitigated by the fact that these competencies are shared by many professions, not exclusive to any programs.
2. The academe experts who evaluated the sample curricula were not necessarily DSA professionals. Thus, their evaluation of the curricula on DSA requirements can be biased by their own respective professions. To mitigate this possible bias, experts were guided by the AAP's framework and definitions as they evaluated the sample curricula.
3. The sample curriculum contained the minimum requirements of the degree program and served as a guide for higher education institutions (HEIs) in developing their actual curriculum. However, the paper evaluation of the sample curriculum did not consider other factors that affect the delivery of instruction, such as teacher qualifications, teaching materials, facilities, evaluation tools, and university culture.

Because of this scope and limitation, it is best to treat the findings and values arrived at in this study as estimates and aimed to steer further studies of the same or related topic. However, these findings sufficed for this research, which is intended as an exploratory study.

Review of Related Literature

Digital transformation, big data, and skills mismatch in the Philippines
With the wave of digital transformation crashing across economies, the Philippines is not exempt from such phenomenon. Microsoft (2018) highlighted the economic impact that digital transformation may bring to the country. It projected that about 40 percent of the Philippine gross domestic product (GDP) by 2021 would be credited to digital transformation. Specifically, it would contribute 0.4 percent of the projected 6.9 percent growth in GDP. Microsoft (2018) also enumerated other positive socioeconomic consequences, such as an increase in income and higher-value jobs in the future, and a rise in opportunities for individuals through better access to education and training. Furthermore, it presented five potential benefits of digital transformation for the companies of Southeast Asian countries, namely, improved profit margin, increased revenue from new products and services, improved customer advocacy, loyalty and retention, improved productivity, and reduced costs.

Rather than displacement, Microsoft (2018) mentioned that job transformation would take place. In the next three years, 92 percent of the Philippine jobs would be transformed. Of these jobs, 34 percent would be retained and upskilled, 31 percent would be new roles due to digital transformation investment, and 27 percent would be outsourced, automated, or made redundant. Meanwhile, only 8 percent of the jobs would remain unchanged. The sectors most vulnerable to automation, according to Francisco et al., (2019), are the youth (age 15–24), those who have received less education, those with lower incomes, those employed in casual or irregular and seasonal jobs or working without pay, and those belonging to the agriculture, forestry, and fishing sectors.

With digital transformation comes the exponential growth of data and consequently, of a data-savvy workforce. The Business-Higher Education Forum (2014, p. 1) quoted Gartner Inc.'s projection that "in less than 12 months, 4.4 million information technology (IT) jobs to support big data will be created globally. About 1.9 million of those jobs will be within [US], and big data has the potential to create three times that number of jobs outside of IT." A 2011 report by McKinsey Global

Institute mentioned that the growth of big data in the US was projected at 40 percent each year with a potential value of USD 300 billion to the nation's healthcare industry alone (Manyika et al. 2011).

This increasing demand for DSA workforce was likewise observed in APEC (APEC Human Resource Development Working Group 2017; BHEF 2014; BHEF 2017). At the top of skills shortage among the APEC economies are DSA skills (APEC Human Resource Development Working Group 2017). The current and projected demand for DSA-related workers in select APEC economies is presented in Table 1.

Table 1. Projected data science and analytics (DSA) workforce demand in select economies

Economy	Current DSA Workers	Projected DSA Workers Needed	Change (%)
Malaysia	4,000 (2016)	20,000 (2020)	400
Philippines	147,420 (2016)	340,880 (2022)	131
Singapore	9,300 (2015)	15,000 (2018)	61
Canada	33,600 (2016)	43,300 (2020)	33
United States	2,350,000 (2015)	2,720,000 (2020)	16

Source: APEC Human Resource Development Working Group (2017)

The skills mismatch that gives rise to difficulties in recruitment processes (Iredale et al. 2014) signifies the inability of educational institutions to meet industry demands (Elemia 2016; Tan and Tang 2016). This "school-industry gap", arguably the most serious hurdle in developing an effective workforce, is felt across industries (Tan and Tang 2016, p. 108).

The Philippine Statistics Authority (PSA) (2012; 2013a) especially highlighted the shortage of high skills, such as those in the field of science and technology (Ramos 2016). Specifically, graduates who joined the IT-business process outsourcing (IT-BPO) industry and emerging infrastructure management services lack the necessary IT skills and infrared software proficiency (Tan and Tang 2016). Moreover, skills gaps that include "numerical competence, verbal and report writing skills, familiarity with different business models and terms, industry-specific knowledge and processes codes and terms" also impair the growing nonvoice segments, such as knowledge processing outsourcing of business analytics, insurance services, and health information management (Tan and Tang 2016, p. 95).

Earlier efforts[3] have been done throughout the years to augment the gap. These were manifested in the proliferation of data science degree programs in various universities and online courses, such as Coursera, a massive open online course and Microsoft's Professional Degree in Data Science (Tan and Tan n.d.). Hence, this rising demand must be addressed by encouraging schools with existing programs in statistics and computer science degrees to also consider offering a program in data science (Tan and Tan n.d.).

Despite these initiatives, the problem persists as reflected in the large share of youth unemployment to overall unemployment. Such a situation warrants the assessment of current Philippine education and its ability to equip its graduates with the set of competencies needed by industry sectors.

Data science and analytics

DSA, as a new field, has no universally accepted definition. One reason provided for this is its involvement of a wide range of skills or competencies that could not easily be identified or narrowed down to a single definition (Demchenko 2017). Nonetheless, existing literature on DSA provides a host of definitions.

Mike Driscoll, chief executive officer of Metamarket, defined data science as "the civil engineering of data (Schutt and O'Neil 2013, p. 7). He further claimed that definitions following this idea of a data science, such as Conway's Venn Diagram of Data Science, possess both the practical knowledge of tools and materials and theoretical understanding of the multidisciplinary nature of the field (Schutt and O'Neil 2013).

Meanwhile, the National Institute of Standards and Technology distinguished the meaning of DSA. "Analytics" is defined as a step in the data life cycle that involves the collection of raw data, preparation

[3] In 2014, it can be recalled how the Aquino administration, through the Department of Labor and Employment, made its Thrust and Priorities centered on addressing the job-skill mismatch in the country. The four identified programs under the policy action of the Human Development and Poverty Reduction Cluster were education and curriculum review; development of a Philippine Qualifications Framework (PQF); implementation of career guidance advocacy; and optimizing the utilization of the enhanced Phil-Job.Net (E-PJN). Earlier, in 2013, the implementation of the K-12 Program was also a major step of the government to address the problem of skill mismatches in the Philippines. More recently, Senate Bill 1456 or the PQF Act of 2017, which just passed in its 3rd and final reading, also served as a policy action toward job-skill alignment (Pasion 2017).

of information then analytics, visualization, and access. On the other hand, “data science” refers to the “extraction of actionable knowledge directly from data through either a process of discovery or hypothesis formulation and hypothesis testing” (National Institute of Standards and Technology as cited in PwC and BHEF 2017, p. 3).

Using these two concepts “analytics” and “data science”, the landscape of DSA job roles was formulated as seen in Table 2. It classifies DSA job roles in analytics-enabled jobs and data science jobs. Professionals who perform analytics-enabled jobs are data-driven decisionmakers who “leverage data to inform strategic and operational decisions” and functional analysts who “utilize data and analytical models to inform domain-specific functions and business decisions” (PwC and BHEF 2017, p. 4). On the other hand, professionals who perform data science jobs are the data engineers who “design, build, and maintain an organization’s data and analytical infrastructure”, data analysts who “leverage data analysis and modeling techniques to solve problems and glean insight across functional domains”, and data scientists who “create sophisticated analytical models used to build new data sets and derive new insights from data” (PwC and BHEF 2017, p. 4). Job titles currently used in organizations were mapped based on these five DSA job roles.

Demchenko (2017) extended the discussion and provided a more detailed and comprehensive definition of the Data Science Profession Profiles encapsulated in the EDISON Data Science Framework (EDSF). EDSF presents five core data science competencies and skills groups as follows:

- Data science analytics, including statistical analysis, machine learning, data mining, business analytics, among others;
- Data science engineering, including software and applications engineering, data warehousing, big data infrastructure, and tools;
- Domain knowledge and expertise, such as scientific domain-related;
- Data management and governance, including data stewardship, curation, and preservation; and
- Research methods for research-related professions and business process management for business-related professions.

Table 2. Data science and analytics job roles

Data Science and Analytics Job Roles				
Analytics-enabled Jobs		Data Science Jobs		
Data-driven Decisionmakers	Functional Analysts	Data Engineer	Data Analyst	Data Scientist
Common job roles • Chief Executive Officer • Chief Data Officer • Chief Information Officer • Director of Information Technology • Financial Manager • Human Resources Manager • Marketing Manager	Common job roles: • Actuary • Business/ Management Analyst • Compensation /Benefits Analyst • Financial Analyst • Geographer/ Geographic Information System Specialist • HRIS Analyst • Operations Analyst • Researcher	Common job roles: • Business Intelligence Architect • Computer Systems Engineer • Data Warehousing Specialist • Data Administrator • Database Architect • Systems Analyst	Common job roles: • Data mining Analyst • Business Intelligence Analyst	Common job roles: • Biostatistician • Data Engineer • Data Scientist • Financial Quantitative Analyst • Statistician

HRIS = human resource information system
Source: Pricewaterhouse Cooper (PwC) and Business Higher Education Forum (BHEF) (2017)

The APEC Human Resource Development Working Group (2017, p. 3) adopted the definition of DSA as "the ability to gather, analyze, and draw practical conclusions from data, as well as communicate data findings to others". Likewise, the definitions of three basic concepts in DSA, namely, data science, big data, and data analytics were given as follows (APEC Human Resource Development Working Group, 2017, p. 5):

- Data science: The field addresses both structured and unstructured data in terms of data cleansing, preparation, and analysis. Overall, it is a grouping of techniques that enable insight and information extraction from data. Internet searches, search recommenders, and digital advertisements and profiles all fall under the application of data science.

- Big data: This field addresses the massive volumes of data generated across industries. The majority of this data is simply too large to be processed or interpreted effectively with traditional approaches. The analysis of big data can deliver insights to support better decisionmaking and business strategies. Financial services, retail, and communication all generate big data and thus require skilled techniques.
- Data analytics: This field examines raw data to draw conclusions from the information. Inference skills are necessary within data analytics, as the researcher must derive conclusions from what they already know.

Two definitions in the Philippine literature come from Tan and Tan (n.d, p. 4) and the Analytics Association of the Philippines (AAP). For Tan and Tan (n.d, p. 4), DSA is “a systematic study of digital data using statistical techniques and applications of computer science. Its goal is to make sense of vast amounts of dynamic data and extract information that will lead to new knowledge that can provide actionable insights for decisionmakers.” On the other hand, AAP provides the operational definition of “analytics” as “a process of progressing data along the value chain as it transforms data to information to insight to imperatives (or actionable insights) with the purpose of delivering the right decision support to the right people and digital processes at the right time for the good” (Pelayo 2019).

Definition of Terms

In the absence of an agreed definition, this study took on the operational definition of “analytics” by AAP for DSA as defined in the previous section. It is deemed that with this definition, the specific attributes identified by varying references about DSA are captured.

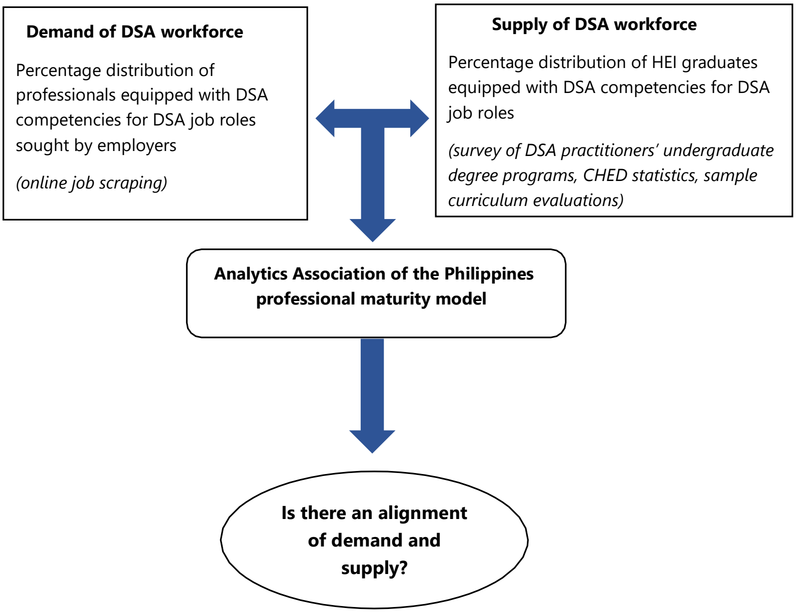
In this paper, the discourse about DSA was explored especially concerning the set of competencies supplied by HEIs and demanded by companies in the Philippines. It is through this approach that the paper attempted to contribute to the ongoing development of DSA and narrowing the gap between what DSA is in theory and practice.

Research Design and Methodology

Conceptual framework

The conceptual framework (Figure 1) makes use of the Professional Maturity Model (Table 3) of AAP as the reference point in determining the alignment between two variables—the percentage distributions of professionals equipped with DSA competencies for DSA job roles sought by employers, and of HEI graduates equipped with DSA competencies for DSA job roles. The demand for DSA workforce was determined through online job scraping, while the supply was sourced from the sample curriculum evaluations of academe experts and relevant statistics from CHED.

Figure 1. Conceptual framework: The alignment between demand and supply of DSA competencies in the Philippines



DSA = data science and analytics; CHED = Commission on Higher Education; HEI = higher education institution
Source: Authors' illustration

The professional maturity model defines four DSA job roles, namely, data steward, data engineer, data scientist, and functional analyst (Table 3):

- Data steward: Develops, enforces, and maintains an organization's data governance process to ensure that data assets provide the organization with high-quality data
- Data engineer: Designs, constructs, tests, and maintains data infrastructures, including applications that extract, transform, and load data from transactional systems to centralized data repositories
- Data scientist: Leverages statistical techniques and creates analytical models to derive new insights from quantitative and qualitative data
- Functional analyst: Utilizes data and leverages on derived insights to help organizations make better decisions in a specific functional domain

Table 3. AAP's professional maturity model

	DSA Professional Job Roles			
DSA Competencies	Data Steward	Data Engineer	Data Scientist	Functional Analyst
Domain knowledge	3	1	2	3
Data governance	3	2	2	2
Operational analytics	3	3	3	3
Data visualization	2	1	2	3
Research methods	1	1	3	1
Data engineering	-	3	1	-
Statistical techniques	-	1	3	-
Data methods and algorithms	-	1	3	-
Computing	1	2	3	1
21st century skills	3	3	3	3

AAP = Analytics Association of the Philippines; DSA = data science and analytics
Note: Level 1 – basic proficiency; Level 2 – intermediate/average proficiency; Level 3 – advanced/expert-level proficiency
Source: AAP (2018)

Each job role is characterized by the required proficiency level of the 10 APEC-recommended DSA competencies. These sets of DSA competencies were developed by an advisory group composed of representatives from the industry, academe, and governments of 14 various APEC-member countries (APEC Human Resource Development Working Group 2017). Taking off from these defined DSA competencies by APEC, AAP's professional maturity model sets the definition or criteria for each level of proficiency per competency. In this study, the definition of a basic proficiency or minimum competency provided by AAP for each competency was adopted. The academe experts were asked to evaluate to what extent their respective program's sample curriculum equips graduates with the minimum competencies.

- Operational analytics - use general and specialized business analytics techniques for the investigation of all relevant data to derive insight for decisionmaking.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Perform business analysis for specified tasks and data sets
2- Intermediate/average	Identify business impact from trends and patterns
3- Advanced/expert	Identify new opportunities to use historical data for organizational processes optimization

- Data visualization and presentation - create and communicate compelling and actionable insights from data using visualization and presentation tools and technologies.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Prepare data visualization reports or narratives based on provided specifications
2- Intermediate/average	Create infographics for effective presentation and communication of actionable outcome
3- Advanced/expert	Select appropriate and develop new visualization methods used in a specific industry

- Data management and governance - develop and implement data management strategies, incorporating privacy and data security, policies and regulations, and ethical considerations.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Be aware and always apply policies and measures to ensure data security, privacy, intellectual property, and ethics
2- Intermediate/average	Enforce policies and procedures for data security, privacy, intellectual property, and ethics
3- Advanced/expert	Develop policies on data security, privacy, intellectual property, and ethics

- Domain knowledge and application - apply domain-related knowledge and insights to effectively contextualize data, achieved by practical experience and exposure to emerging innovations.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Understand collected data, and how they are handled and applied in the specific industry domain
2- Intermediate/average	Develop content strategy and information architecture to support a given industry domain and its audiences
3- Advanced/expert	Make business cases to improve domain-related procedures through data-driven decisionmaking

- Statistical techniques - apply statistical concepts and methodologies to data analysis.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Know and use statistical methods such as sampling, ANOVA, hypothesis testing, descriptive statistics, regression analysis, and others
2- Intermediate/average	Select and recommend appropriate statistical methods and tools for specific tasks and data
3- Advanced/expert	Identify problems with collected data and suggest corrective measures, including additional data collection, inspection, and preprocessing

- Computing - apply information technology and computational thinking, and utilize programming languages and software and hardware solutions for data analysis.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Perform basic data manipulation, analysis, and visualization
2- Intermediate/average	Apply computational thinking to transform formal data models and process algorithms into program code
3- Advanced/expert	Select appropriate application and statistical programming languages, and development platforms for specific processes and data sets

- Data analytics methods and algorithms - implement and evaluate machine learning methods and algorithms on the data to derive insights for decisionmaking.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Demonstrate understanding and perform statistical hypothesis testing, and explain statistical significance of collected data
2- Intermediate/average	Apply quantitative techniques (e.g., time series analysis, optimization, simulation) to deploy appropriate models for analysis and prediction
3- Advanced/expert	Assess data on reliability and appropriateness, and select appropriate approaches and their impact on analysis and the quality of the results

- Research methods - utilize scientific and engineering methods to discover and create new knowledge and insights.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Understand and use the four-step research model, namely, hypothesis, research methods, artifact, and evaluation
2- Intermediate/average	Develop research questions around identified issues within existing research or business process models
3- Advanced/expert	Design experiments which include data collection (passive and active) for hypothesis testing and problem solving

- Data science engineering principles - use software and system engineering principles and modern computer technologies to develop data analytics applications.

Proficiency levels (1,2,3)	Behavioral Indicators
1- Basic	Program selected Structured Query Language (SQL) and not only SQL platform for data storage and access, in particular write extract, transform, load (ETL) scripts
2- Intermediate/average	Design and build relational and nonrelational databases, ensure effective ETL processes for large datasets
3- Advanced/expert	Employ advanced knowledge and experience of using modern big data technologies to process different data types from multiple sources

- 21st century skills - exhibit cross-cutting skills essential for analytics at all levels, including but not limited to, collaboration, ethical mindset, empathy, social and societal awareness, dynamic (self) reskilling, and entrepreneurship.[4]

21st century skills	Behavioral indicators for Basic Proficiency (1)
Collaboration	Work with others to accomplish something
Communication and storytelling	Communicate narratives appropriately to the target audience
Ethical mindset	Discern what is right or wrong in every situation
Organizational awareness	Understand organization from top to bottom and from inside to out and relate one's role to the larger organization and industry in which one works
Critical thinking	Objectively analyze and evaluate an issue to form a judgement
Planning and organizing	Create and use logical and systematic processes to achieve goals
Problem solving	Find solutions to difficult or complex issues

[4] The definitions of the 21st century skills were lifted from Glossary of Education Forum (2016) to define the behavior indicators for the basic proficiency level.

21st century skills	Behavioral indicators for Basic Proficiency (1)
Decisionmaking	Take a stand over an issue
Customer focus	Pay attention to the needs and wants of the customer
Flexibility	Adjust to changes with ease
Business fundamentals	Address the different aspects of running an organization. This includes understanding of business structures and functions that apply to any organization.
Cross-culture awareness	Stand back from oneself and become aware of one's and others' cultural values, beliefs, and perceptions.
Social consciousness	Be conscious or aware of the problems within a society or community.
Dynamic (self) reskilling	Acquire new skills
Professional networking	Build, reinforce, and maintain relationships
Entrepreneurship	Start and run a new business/initiative/ solution along with its risks to generate profit/value

Source: Glossary of Education Forum 2016

Research design, instruments, and participants

Demand side

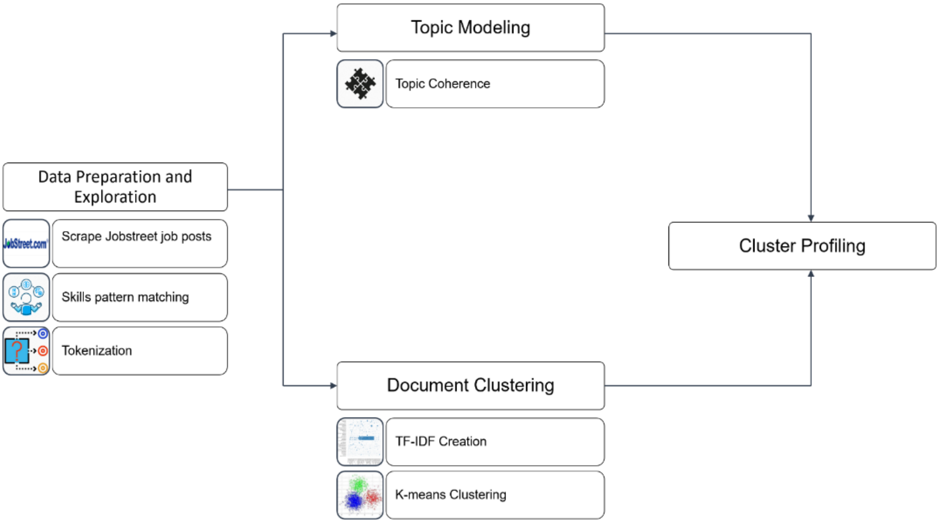
DSA-related activities were identified from the list of activities in the Philippine Standard Industry Code (PSIC) of PSA (2009).[5] The workers who performed these DSA-related activities were assumed to possess the DSA competencies. Employment data from the 2013–2016 Annual Survey of Philippine Business and Industry on various industries and the 2018 Compilation of Industry Statistics on Labor and Employment were used to determine the number of these workers who possessed any of the DSA competencies (PSA 2013b–2016; 2018). Key informant interviews with relevant stakeholders, such as those from the Department of Labor and Employment (DOLE), were conducted for additional and supplementary information on the demand for DSA workforce.

[5]The study manually identified which industries do analytics-related activities are mostly lodged in accordance with the industry definitions provided by PSA, specifically in the 2009 PSIC and the "analytics" definition used by the framework of this study.

The demand for the four DSA job roles and DSA competencies was determined by extracting skills and segmenting job roles found in the online job listings (see Appendix A for the detailed online job scraping methodology). The online job scraping methodology followed four stages (Figure 2):

- Stage 1: Data preparation and exploration
- Stage 2: Topic modeling
- Stage 3: Document clustering
- Stage 4: Cluster profiling

Figure 2. Skills extraction and job segmentation process



TF-IDF = term frequency - inverse document frequency
Source: Authors' illustration

Supply side

The top 10 undergraduate degree programs of current DSA practitioners were considered as the 10 DSA-related undergraduate degree programs in this study (see Appendix B for the list and count). These DSA practitioners were AAP members or those who attended AAP monthly meetups. The key characteristics and attributes of these programs were defined in the respective CMOs (Table 4).[6]

[6] For further details of the program description, the respective CMO for each DSA-related degree program is accessible through CHED's official website (www.ched.gov.ph)

Table 4. Program description of DSA-related degree programs

DSA-related Degree Program	CMO	Program Description
Computer Science	CMO 25, series 2015	Includes the study of computing concepts and theories, algorithmic foundations, and new developments in computing
Business Administration	CMO 17, series 2017	Utilizes an integrated approach to study the interrelationships among the different functional areas of business and examines how the effective orchestration of these different components of business operations can lead to organizational success
Statistics	CMO 42, series 2017	Aims to characterize and understand (provide prognosis) the random process that leads to uncertainty and consequently facilitates decisionmaking using tools that aggregate data into meaningful information
Applied Math/ Math	CMO 48, series 2017	Mathematics program: engages in pure mathematics or mathematics for its own sake, without having, at least initially or intentionally, application or utility in mind
Information Technology	CMO 25, series 2015	Includes the study of the utilization of both hardware and software technologies involving planning, installing, customizing, operating, managing, and administering and maintaining information technology infrastructure
Information Science	CMO 24. series 2015	The study of the development, deployment, and management of information resources in print, nonprint, electronics, and digital formats and services
Economics	CMO 32, series 2017	As the scientific study of how people allocate and use scarce resources, is designed to equip students with knowledge in economic theory and its applications, and with essential skills for undertaking economic analysis, incorporate observed trends in economics education, align teaching methodologies to specific learning philosophies, and promote research competencies

Table 4. (continued)

DSA-related Degree Program	CMO	Program Description
Physics/Applied Physics	CMO 20, series 2007	Generally provides comprehensive and rigorous training in physics as a foundation for careers in pure and applied physics or interdisciplinary sciences
Industrial Engineering	CMO 96, series 2017	Intends to prepare students for a professional industrial career in the design, improvement, and installation of integrated systems of people, materials, information, equipment, and energy with specialized knowledge and skills in mathematical, physical, and social sciences along with principles and methods of engineering analysis and design
Civil Engineering	CMO 92, series 2017	Applies the basic principles of science in conjunction with mathematical and computational tools to solve problems associated with developing and sustaining civilized life on the planet; one of the broadest engineering disciplines both in terms of the range of problems that fall within its purview and in the range of knowledge required to solve those problems

DSA = data science and analytics; CMO = CHED memorandum order
Source: CHED (n.d.)

Purposive sampling, specifically, expert sampling, was employed given the exploratory nature of the study.[7] Academe experts from the University of the Philippines Diliman (UPD) and University of Asia and the Pacific (UA&P) who were either program administrators or faculty members of these 10 DSA-related undergraduate programs were convened

[7] Aside from the exploratory nature of the study, the authors deemed purposive sampling method ("expert sampling") as the most appropriate given resource constraints in obtaining the needed empirical data to employ a probabilistic sampling method. Invitations to the FGDs were also sent to all HEIs. The authors obtained official email addresses of all registered HEIs from CHED. However, only 18 HEIs participated, and given such low number of participants relative to the population of Philippine higher education, the authors diverted to purposive sampling. Given this, the authors acknowledge the limitations in the representativeness of the employed sample. Readers are then also advised to take results from this study within its scope and limitations.

in focus group discussions (FGDs) to evaluate the sample curriculum in the CMO of their respective program of expertise.[8] These two HEIs represent the two types of academic institutions in the Philippines, namely, public (UPD) and private (UA&P).

Each sample curriculum was evaluated on the extent it can equip the graduates with the basic proficiency of the 10 analytics competencies following this scale:

0 - does not equip graduates with the analytics competency
1 - minimally equips graduates with the analytics competency
2 - moderately equips graduates with the analytics competency
3 - mostly equips graduates with the analytics competency
4 - fully equips graduates with the analytics competency

Likewise, each sample curriculum was evaluated on its ability to prepare the graduates to perform the four DSA job role–data steward, data engineer, data scientist, and functional analyst following this scale:

0 - does not enable graduates to perform the tasks of the DSA job role
1 - minimally enables graduates to perform the tasks of the DSA job role
2 - moderately enables graduates to perform the tasks of the DSA job role
3 - mostly enables graduates to perform the tasks of the DSA job role
4 - fully enables graduates to perform the tasks of the DSA job role

Through the evaluations of these experts, the study was able to analyze the supply of DSA workforce.

Results and Discussion

Demand side

Current availability of DSA competencies

Out of 1,271 activities listed in the PSIC, 22 activities (Table 5) were identified to require any of the 10 DSA competencies. The majority of these DSA-related activities belonged to the information and communications

[8] Except for the degree programs Economics, Math/Applied Math, and Information Technology, which were assessed by UA&P faculty/program directors, all the other identified DSA-related degree programs were assessed by experts of the field in UPD. Eleven program directors/ faculty members submitted their curriculum evaluations for this study (8 from UPD and 3 from UA&P) while 20 academe experts (including the 11) attended the FGDs/KIIs conducted in these two universities.

subsector (41%), followed by the manufacturing subsector (33%), and professional, scientific and technical activities (24%). The study estimated that, in 2016, there were about 174,000 workers who performed these activities and, thus, possessed any of the DSA competencies[9] (Table 5). This figure was only 0.4 percent[10] of the total labor force in the Philippines in the said year (see Appendix C for employment share per major industry group). Almost half of these DSA-competent workers (42%)[11] belonged to the IT-BPM industry[12] (Table 5).

Table 5. Number of employees working on activities that require any of the DSA competencies

Industry subsectors	Activities	2016	
		Employees	In %
Manufacturing		**58,224**	**34.00**
C26200	Manufacture of computers and peripheral equipment and accessories	58,224	33.52
Financial and insurance services		191	0.11
K66290	Other activities auxiliary to insurance and pension funding	191	0.11
Professional, scientific, and technical activities		**42,518**	**24.00**
M70200	Management consultancy activities	27,396	15.77
M71109	Other technical activities related to architectural and engineering	2,902	1.67

[9] The researchers made use of the keywords also used in the online job scraping. This is for the purpose of consistency with regard to the definition of DSA used in this study.

[10] Authors' computation given that there are about 41 million workers in the country in 2016 (PSA 2013–2017, Labor Force Survey as cited in the JobsFit 2022 Report of the Bureau of Local Employment, Department of Labor and Employment; PSA CISLE Report 2018).

[11] This is also supported by the data gathered through online job scraping. Job postings online can be a reliable source of pinning down current demand for DSA workforce as affirmed in our interviews with some IT-BPM companies in the country. It is because companies activate all possible channels including online platforms, such as JobStreet, in hiring the right DSA talents, which can reflect actual demand by the IT-BPM industry. Due to lack of available data specific to the analytics workforce, the study could not estimate the proportion of these online job postings to the total demand. Moreover, while companies use online platforms to make known their need, they still get to successfully hire the needed DSA employees mostly through their respective networks and headhunters.

[12] The Annual Survey of Philippine Business and Industry reports of PSA identify business process management (BPM) industries as determined by two industries, namely, Section J: Information and Communication and Section N: Administrative and Support services (see Appendix D for the complete list of industry subclasses under BPM).

Table 5. (continued)

Industry subsectors	Activities	2016	
		Employees	In %
M72101	Research and experimental development in natural sciences	723	0.42
M72102	Research and experimental development in engineering and technology	342	0.20
M72103	Research and experimental development in health sciences	465	0.27
M72104	Research and experimental development in agricultural sciences	257	0.15
M72200	Research and experimental development on social sciences and humanities	1,034	0.60
M72300	Research and experimental development in information technology	3,686	2.12
M72400	Research and experimental development services, nec	s	NA
M73200	Market research and public opinion polling	5,713	3.29
M749	Other professional, scientific, and technical activities, nec	nad	NA
Information and communication		**71,321**	**41**
J58200	Software publishing	1,852	1.07
J62010	Computer programming activities	38,808	22.34
J62020	Computer consultancy and computer facilities management activities	11,476	6.61
J62090	Other information technology and computer service activities	3,780	2.18
J63111	Data processing	14,052	8.09
J63112	Website hosting services	905	0.52
J63113	Application hosting services	448	0.26
Administrative and support services		**1,432**	**1.0**
N82294	Research and analysis activities	291	0.17
N82299	Other back office operations activities, nec	1,141	0.66
	TOTAL	**173,686**	**100.00**

DSA = data science and analytics; nec = not elsewhere classified; s = suppressed by the publisher; nad = no available data
Note: See Appendix E for the employment sizes of these industry subclasses for 2013–2015
Source: PSA (2016)

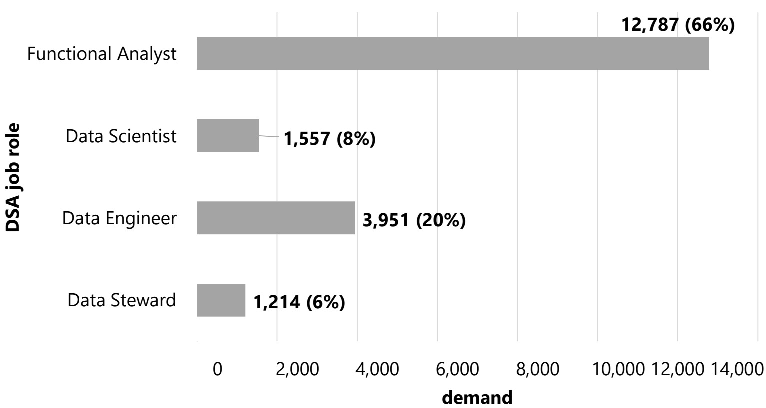
Current demand for DSA job roles

The online scraping of job postings in early 2019, from May 1 to July 21, through Jobstreet.com.ph revealed the skills employers sought from the DSA professionals they wanted to hire.

DSA is at the very nascent stage in the Philippines. Its journey, as reflected in the operational definition of "analytics", begins with convincing leadership that data is the new "oil" of the organization. This challenging task of getting the buy-in of top executives certainly falls on those who operate the business as they can articulate how data can deliver their operational targets. Consistently, the bulk (66%) of the demand for DSA professionals in early 2019 was for the functional analyst role (Figure 3).

Once the top management has decided to embrace DSA, the next important DSA job role is the data engineer, who ensures that data at their right form are easily available for analytics. The data engineer builds and maintains the right data infrastructures and populates their central databases with data from transactional systems in the form ready for model building and analysis. Rightly, the second (20%) most sought role was the data engineer. This demand for the data engineer, however,

Figure 3. Demand for data science and analytics (DSA) professionals



Note: Total demand = 19,509 DSA professionals
Source: Authors' compilation

may be conflated with the software engineer/developer role leading to a lower in demand for the data engineer.

Only with clean data in data warehouses can algorithms be applied on data for insights. The low demand for data scientists (8%) can imply that companies did not have their data ready for the DSA process.

The Facebook and Cambridge Analytica scandal in 2018 revealed the dark side of analytics. It showed how the data of millions of Facebook users were used without their consent for political advertisement in the 2016 US election and Brexit campaign in the United Kingdom (MacASkill 2018). When data are misused, ethical issues on human rights arise and put society at risk. Since then, governments have tightened the implementation of data privacy and protection laws. In the Philippines, the Data Privacy Act was passed in 2012 with the National Privacy Commission as the mandated implementing agency (NPC n.d.). Hence, the demand for data steward has been slightly evident (6%) in early 2019.

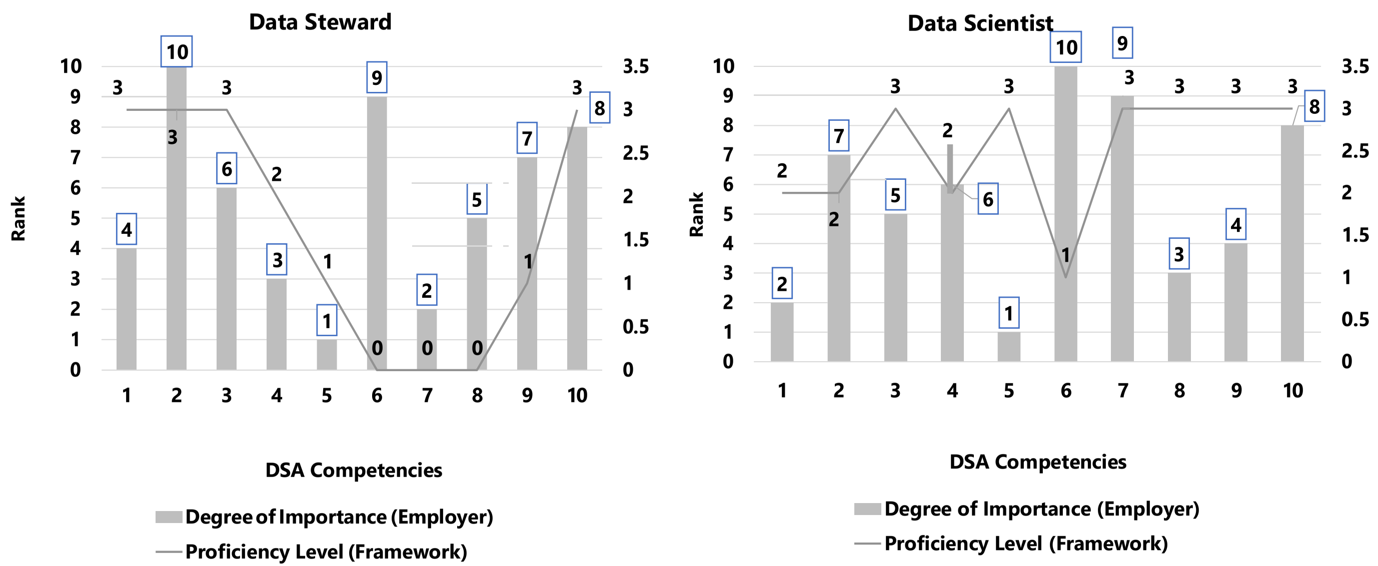
Work analysis of in-demand DSA job roles

The work of each demanded DSA professionals was analyzed based on the competencies' importance and level of proficiency as perceived by employers and as required by the AAP Model respectively (Table 3). While research methods did not come out as an important competency, the study considered it as incorporated with statistical techniques.

The AAP Model requires an advanced level of proficiency for domain knowledge, data governance, and 21st century skills. Meanwhile, it does not require three competencies (data engineering, statistical techniques, and methods and algorithms competencies) for the work of a data steward (Figure 4). Employers, however, considered these three as important, together with all the other competencies for the data steward. The top three competencies that employers considered to be needed were data governance (rank 10), data engineering (rank 9), and 21st century skills (rank 8). The employer's inclusion of data engineering as a needed competency could be attributed to the fact that the task of drafting and implementing data governance policies requires an understanding of data systems.

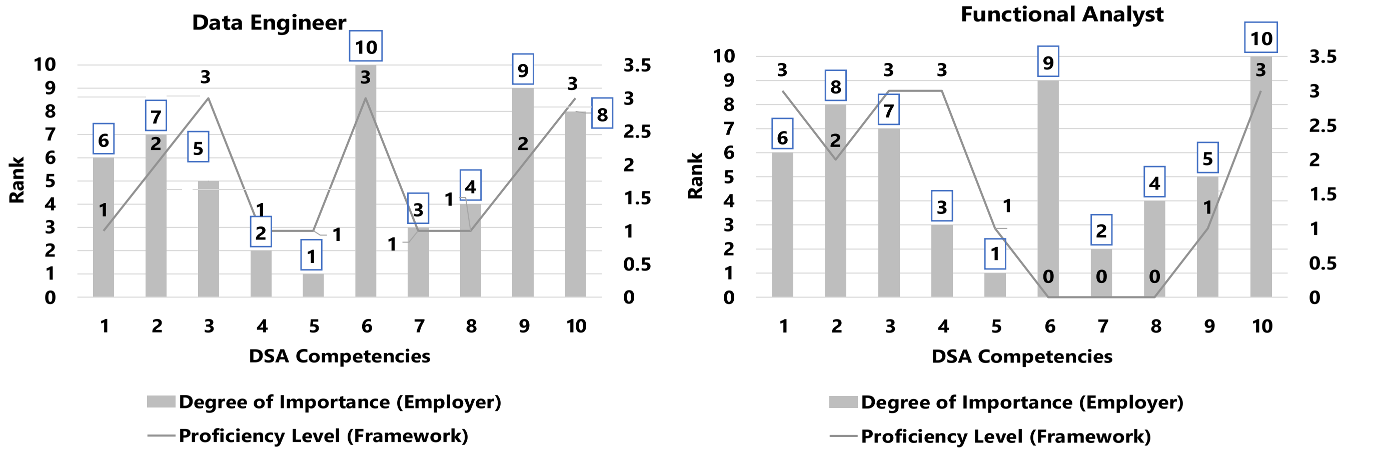
The AAP Model and employers agreed that all the competencies are needed by a data engineer. Among these, employers ranked data engineering (rank 10), computing (rank 9), and 21st century skills (rank 8) as the most important ones. The AAP Model requires an advanced level of

Figure 4. Data science and analytics (DSA) competencies' degree of importance and level of proficiency for each DSA job role



Notes: *X-axis (DSA Competencies):* 1- Domain knowledge; 2 - Data governance; 3- Operational analytics; 4- Data visualization; 5- Research methods; 6- Data engineering; 7- Statistical techniques; 8- Methods and algorithm; 9- Computing; 10- 21st century skills; *Y-axis [Degree of Importance (Rank)]*: 10 – most important; ….; 1 – least important. The ranking presents the extent employers expect the competency to be needed to complete the work of the specific job role; *Y-axis (Level of Proficiency):* 1 – basic/beginner proficiency; 2 – intermediate/average proficiency; 3 – advanced/expert proficiency). Proficiency presents the minimum level of mastery required by the AAP Model for the competency to complete the work of the specific job role
Source: Authors' compilation

Figure 4. (continued)



Notes: *X-axis (DSA Competencies):* 1- Domain knowledge; 2 - Data governance; 3- Operational analytics; 4- Data visualization; 5- Research methods; 6- Data engineering; 7- Statistical techniques; 8- Methods and algorithm; 9- Computing; 10- 21st century skills; *Y-axis [Degree of Importance (Rank)]*: 10 – most important; ....; 1 – least important. The ranking presents the extent employers expect the competency to be needed to complete the work of the specific job role; *Y-axis (Level of Proficiency):* 1 – basic/beginner proficiency; 2 – intermediate/average proficiency; 3 – advanced/expert proficiency). Proficiency presents the minimum level of mastery required by the AAP Model for the competency to complete the work of the specific job role
Source: Authors' compilation

proficiency for data engineering and 21st century skills, while only an intermediate level for computing. A close match was observed between the importance of the employers and proficiency levels of the AAP Model on the competencies for the data engineer role (Figure 4).

Like the data engineer role, the AAP Model and employers agreed that all the competencies are needed by a data scientist. The three most needed competencies for the data scientist role according to employers were data engineering (rank 10), statistical techniques [which incorporates research methods] (rank 9), and 21st century skills (rank 8). On the other hand, the AAP Model requires advanced expertise in statistical techniques, research methods, and 21st century skills, and only a basic proficiency in data engineering. The treatment of data engineering as a key competency among employers reflected the current work of data scientists, which includes data preparation, model building, and implementation of algorithms. Two other competencies that require an advanced level of proficiency are operational analytics (rank 5) and methods and algorithms (rank 3) per the AAP Model. Employers, however, did not perceive these competencies as crucial to the work of a data scientist.

Like the data steward role, the AAP Model excludes some competencies for the functional analyst role, namely, data engineering, statistical techniques, and methods and algorithms. The model further prescribes that the functional analyst be an expert in the 21st century skills and with intermediate proficiency in data governance. For employers, all the competencies were needed, especially 21st century skills (rank 10), data engineering (rank 9), and data governance (rank 8). The necessity of data engineering for the work of a functional analyst, according to employers, may indicate that functional analysts were expected to perform a few tasks of the data engineer and the data scientist.

In summary, the AAP Model and employers agreed that all DSA professionals must be very competent in 21st century skills. However, they did not agree on data engineering—employers found it needed by all four DSA job roles, while the AAP Framework requires it only for data engineers and data scientists.

The observed discrepancies between the AAP Model and employers' expectations on the job roles may be attributed to the fact that DSA as a profession is yet to be clearly defined. In practice, these job roles are not particularized within the DSA process. This was evident in the responses of hiring managers, who said that they often end up hiring the wrong candidate based on incorrect expectations. For example, there were job postings for the data science role with the tasks of a data engineer. These discrepancies can also serve as inputs to the AAP's professional maturity model, which is continuously being updated to adequately capture the nature of DSA work in the country.

Current supply side

DSA-related undergraduate degree programs

The undergraduate degrees of current DSA practitioners were surveyed (see Appendix C for specific breakdown).[13] The top 10 undergraduate degrees were considered as DSA-related programs in this study. The most popular or common DSA-related programs among surveyed DSA practitioners were computer science, business administration, and statistics (Table 6).

Meanwhile, the undergraduate programs mostly offered by HEIs in 2016 were business administration (57%), information technology (56%), and computer science (38%) (CHED 2019). Aside from these programs, very few HEIs offer other DSA-related programs.

Table 6. Top 10 DSA-related undergraduate degree programs

Degree Program	Count	Percent of HEIs (in %)
Computer Science	11	38.39
Business Administration	11	56.51
Statistics	10	0.82
Applied Math/Mathematics	9	8.00
Information Technology	8	56.05
Library and Information Science	7	3.81
Economics	6	5.56
Physics/Applied Physics	6	1.00
Industrial Engineering	4	5.61
Civil Engineering	4	12.51

DSA = data science and analytics; HEI = higher education institution
Note: Mathematics includes Applied Mathematics program;
Physics includes Applied Physics program
Source: Survey of DSA practitioners and authors' computations based on CHED (2019)

[13] The respondents of this survey were those who attended the monthly meetups of AAP and R-User's group. Both groups are run by pioneering DSA industry leaders and frequent resource persons in DSA-related fora.

Current supply of DSA competencies

As mentioned earlier, this study made use of the assessments on the sample curricula as HEIs were assumed to comply with them to deliver their respective degree programs. The discussion on the 21st century skills is provided in a separate subsection.

The curriculum assessments revealed that these DSA-related programs equipped the students with the basic proficiency of the 10 DSA competencies in varying degrees (Table 7). Among these programs, the computer science program performed best by equipping its students with seven DSA competencies. Meanwhile, the mathematics program equipped their respective students with six competencies, physics and industrial engineering programs with four competencies, IT program with three competencies, and statistics with only one competency. The business administration, library and information science, economics, and civil engineering programs did not equip the students with the basic proficiency of any of the competencies.

The DSA competencies, which most programs equipped their graduates with, were statistical techniques (5 of 10 DSA-related programs) and data engineering (4 of 10 DSA-related programs). Meanwhile, 21st century skills were the hardest to transmit to students. During FGDs, hiring managers highlighted their concern regarding job candidates not possessing the 21st century Skills.

21st century skills

A key feature among the APEC-recommended DSA competencies is the 21st century skills. This competency is often contrasted with the other competencies as it pertains to a set of nontechnical skills. Hiring managers indicated that such skills were the most difficult to develop and most valuable in the workplace. Hence, the discussion on the 21st century skills merits a special section.

Employers expect that students acquire these 21st century skills while gaining their undergraduate degrees.[14] It is then important to determine the extent HEIs can incorporate these 21st century skills in the curricula of the DSA-related programs. The 21st century skills refer to a broad set of 16 knowledge, skills, habits, and character traits (Glossary of Education Forum 2016) otherwise called “soft skills”.

[14] This is based on the insights gathered from the FGD with hiring managers for DSA job roles on February 2, 2019 at the UA&P.

Table 7. DSA-related programs' ability to equip graduates with DSA competencies

DSA Competencies	DSA-related degree program (based on CMO's sample curriculum)									
	Computer Science	Business Administration	Statistics	Mathematics	Information Technology	Library and Information Science	Economics	Physics	Industrial Engineering	Civil Engineering
Domain knowledge	3	2	3	4	1	2	2	4	4	1
Data governance	4	1	3	4	3	2	0	2	3	0
Operational analytics	3	0	3	4	1	1	2	2	4	0
Data visualization	4	2	4	4	0	1	1	3	3	2
Research methods	4	2	3	4	3	2	3	4	3	1
Data engineering	4	0	3	3	4	3	0	4	4	0
Statistical techniques	4	2	2	4	4	2	3	4	4	1
Methods and algorithms	4	0	2	3	2	2	3	3	3	0
Computing	4	1	2	3	4	1	0	2	3	2
21st century skills										

DSA = data science and analytics; CMO = CHED memorandum order

Note: Rating 4 – fully equips graduates with the DSA competency; 3 – mostly equips; 2 – moderately equips; 1 – minimally equips; 0 – does not equip Source: Survey responses of participating Program Directors or HEI representatives and CHED (2019)

An assessment of the sample CMO curricula of the nine DSA- related programs[15] revealed that their graduates possessed these 16 soft skills in varying degrees (Table 8): problem-solving skills (7 of 9 DSA-related programs); job collaboration and critical thinking (5 out of 9 DSA-related programs); organizational awareness, planning and organizing skills, and decisionmaking (4 out of 9 DSA-related programs); communication and storytelling and job flexibility (3 out of 9 DSA-related programs); ethical mindset and customer focus (3 out of 9 DSA-related programs); business fundamentals, cross-cultural awareness, and dynamic (self) reskilling (1 out of 9 DSA-related programs); and social awareness, professional networking, and entrepreneurship (0 out of 9 DSA-related programs).

Overall, these DSA-related programs can be improved specifically on seven soft skills, namely, entrepreneurship, professional networking, social awareness, dynamic (self) reskilling, cross-cultural awareness, business fundamentals, and customer focus.

Among the sample CMO curricula, the industrial engineering program best educated its graduates with 11 of 16 soft skills. Meanwhile, the math and applied math program was able to equip its graduates with 9 of 16 soft skills, computer science with 8 of 16 soft skills, physics and applied physics with 6 of 16 soft skills, library information science with 5 of 16 soft skills, statistics with 2 of 16 soft skills, civil engineering with 1 of 16 soft skills, and business administration and IT programs with 0 of 16 soft skills (see Table 8).

Overall, the business administration, IT, and statistics can be improved significantly by incorporating most of the 16 soft skills in their respective sample CMO curricula.

[15] Respondent opted not to assess the economics program's sample CMO curriculum on the 21st century skills.

Table 8. DSA-related programs' ability to equip graduates with 21st century skills

DSA Competencies	DSA-related degree program (based on CMO's sample curriculum)								
	Computer Science	Business Administration	Statistics	Mathematics	Information Technology	Library and Information Science	Physics	Industrial Engineering	Civil Engineering
Job collaboration	4	2	4	4	3	2	4	4	2
Communication and storytelling	3	2	3	4	1	2	4	4	3
Ethical mindset	4	3	3	4	2	2	2	3	3
Organizational awareness	4	3	2	4	1	4	1	4	2
Critical thinking	4	2	3	4	3	4	4	4	2
Planning and organizing	4	3	3	4	2	4	3	4	2
Problem solving	4	2	4	4	3	4	4	4	4
Decisionmaking	4	2	3	4	2	3	4	4	3
Customer focus	2	3	1	3	1	4	1	4	1
Job flexibility	2	3	2	4	1	3	4	4	2
Business fundamentals	1	3	1	2	1	2	0	4	1
Cross-cultural awareness	2	2	1	3	0	3	1	4	2
Social awareness	2	3	1	3	3	3	1	3	2
Dynamic (self) reskilling	4	2	1	2	0	3	2	3	2
Professional networking	2	3	1	3	0	3	3	3	2
Entrepreneurship	2	3	1	3	0	2	0	3	2

DSA = data science and analytics; CMO = CHED memorandum order
Source: Survey responses of participating program directors or representatives from higher education institutions

Additional insights can be drawn from the findings of DOLE and SFI Group (2018), which profiled the Philippine labor force on a set of soft skills (see Table 9). The soft skills considered in the PTMI study and AAP Model (Table 3) are decisionmaking, planning and organizing, and problem-solving skills. The PTMI found these soft skills to be poor among Filipino students and trainees. This study, likewise, finds the necessity to enhance the sample CMO curricula to better equip the graduates with these soft skills rating below "4", namely, decisionmaking (5 out of 9 programs), planning and organizing (5 out of 9 programs), and creative problem-solving (2 out of 8 programs).

Table 9. Mean scores of the Philippines for 21st century skills

Philippine Overall Mean Scores		Student Overall Mean Scores		Trainee Overall Mean Score	
English functional skill	72	English functional skill	74	English functional skill	64
Math functional skill	72	Math functional Skill	73	Workplace ethics	61
Workplace ethics	72	Workplace ethics	72	Social perceptiveness	60
multitasking	70	English comprehension	71	Multitasking	60
English comprehension	69	Multitasking	70	Stress tolerance	59
Critical thinking	66	Critical thinking	68	Math functional skill	58
Stress tolerance	65	Stress tolerance	65	English comprehension	57
Social perceptiveness	64	Social Perceptiveness	65	Critical thinking	55
Self-motivation	61	Problem sensitivity	62	Self-motivation	52
Teamwork	61	Self-motivation	62	Teamwork	52
Problem sensitivity	61	Teamwork	62	Problem sensitivity	50
Decisionmaking	57	Innovation	59	Creative problem Solving	49
Innovation	57	Decisionmaking	59	Decisionmaking	48
Planning and organizing	57	Planning and Organizing	57	Planning and organizing	48
Creative problem solving	54	Creative problem Solving	56	Innovation	45

Source: DOLE and SFI Group of Companies (2018)

Current supply of DSA-job ready graduates

In 2019, the 10 DSA-related undergraduate programs produced 176,597 professionals of 10 various disciplines (Table 10). Among these graduates, 62,583 (38%) were assessed to be "ready" to shift to DSA since they have been prepared to perform at least one DSA job role.

The industrial engineering program best prepared students for DSA job roles. New industrial engineers may immediately work as data stewards, data engineers, or functional analysts. Research methods, methods, and algorithms and computing competencies needed to be strengthened to also enable its graduates for the work of a data scientist (Table 10).

The next best DSA-related program was the computer science program for qualifying its graduates for the data engineer and data scientist roles. Still, it needed upskilling on the domain knowledge and operational analytics competencies to be ready to work as data stewards or functional analysts (Table 10).

Immediate data scientist employment may be offered to new mathematicians and physicists (Table 10). They were not deemed to perform readily as data stewards and functional analysts because they needed to learn more topics in data management and widen the application of their techniques outside of their respective fields, i.e., math (finance and actuarial science) and physics (Table 7).

The IT program enabled its graduates to be data engineers (Table 10). IT graduates have been educated to handle technology architectures, infrastructures, and security, which require the foundational skills of data architectures, infrastructures, and security.

New economists were assessed to be functional analysts with their deep training in conducting research, i.e., collect and analyze data to draw logical conclusions regarding the economy (Table 10). To be prepared to work as data scientists, they just need a bit more skills upgrading especially in computing (Table 7).

Even though only 16 HEIs (0.82%) in the country offered the statistics program, there were relatively a good number of data science practitioners who had this degree (Table 6). While the sample curriculum in statistics should provide students with fundamental knowledge and skills related to data science, the evaluators were not able to ascertain the sample curriculum's ability to equip the students with

Table 10. Supply of DSA job roles from the DSA-related programs

Job roles	Computer Science	Business Administration	Statistics	Math	IT	Library and Information Science	Economics	Physics	Industrial Engineering	Civil Engineering	Total DSA Job roles	% DSA Job Roles
Data Steward	2	1	2	3	2	2	1	2	4	0	4,002	5
Data Engineer	4	1	3	2	4	2	1	2	4	0	58,438	72
Data Scientist	4	1	3	4	1	1	3	4	3	3	13,150	16
Functional Analyst	2	1	1	3	2	2	4	3	4	0	5,488	7
											81,078	1.00
Total graduates*	10,491	98,842	500	2,423	43,945	636	1,486	236	4,002	14,036		176,597
Total "DSA ready" graduates	10,491	X	X	2,423	43,945	X	1,486	236	4,002	X		62,583

DSA = data science and analytics; IT = information technology
Note: Rating 4 – fully enables graduates to perform tasks of the DSA job role; 3 – mostly enables; 2 – moderately enables; 1 – minimally enables; 0 – does not enable; X – does not fully enable; Survey responses of participating Program Directors or HEI representatives; Commission on Higher Education. (15 March 2019).
* = Data used for number of graduates were based on authors' computed projection of 2019 graduates. Meanwhile, the basis of such computations was CHED's 2012–2017 actual number of graduates for all baccalaureate degree programs. As of the writing of this paper, CHED was only able to provide the number of graduates for these academic years.
Source: Authors' computation

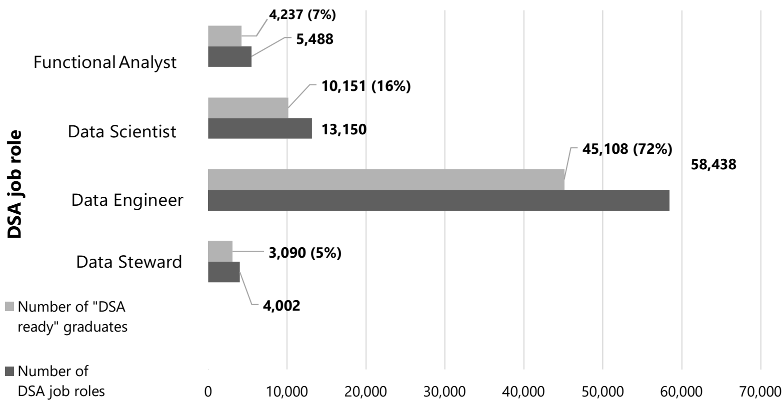
the basic proficiency of all the competencies (except data visualization) due to lack of details in the courses' descriptions. Its graduates need to gain the lacking competencies after graduation to adequately work as data scientists.

Similarly, DSA practitioners with undergraduate degrees in library and information science, business administration, and civil engineering must acquire the basic proficiency of all the DSA competencies through other means (Table 7 and 10).

Graduates assessed to be ready to perform any DSA roles upon graduation are said to be "DSA ready" graduates. In 2019, they were estimated to be 62,583 (Table 10). Since some programs enable their graduates to take on multiple DSA job roles, these "DSA-ready" graduates can fill in 81,078 DSA job roles. Of these available DSA job roles that "DSA-ready" graduates can fill in, 4,002 (5%) were data steward roles of industrial engineering graduates; 58,438 (72%) were data engineer roles of computer science, library and information science, and industrial engineering graduates; 13,150 (16%) were data scientist roles of computer science, mathematics, and physics graduates; and 5,488 (7%) were functional analyst roles of economics and industrial engineering graduates.

Once a "DSA-ready" graduate decides to get into DSA, they must choose only one of the DSA job roles they are qualified to do. Given that there were only 62,583 "DSA ready" graduates in 2019, the supply of DSA professionals from the 10 DSA-related programs was estimated to be 5 percent for data steward, 72 percent for data engineers, 16 percent for data scientists, and 7 percent for data stewards (Figure 5).

Figure 5. Supply of "DSA ready" graduates from the DSA-related programs



DSA = data science and analytics
Source: Authors' computations

Summary of supply-side discussion

Comparing the percentage distributions of demand and supply of DSA professionals (Table 11), the study found a misalignment between the type of DSA workers sought by employers and the type of DSA graduates produced by HEIs in 2019. The majority of employers were looking for functional analysts (66%) while HEIs were mostly producing data engineers (72%).

Table 11. Demand and supply of data science and analytics professionals (in %)

	Data Steward	Data Engineer	Data Scientist	Functional Analyst
Demand	6	20	8	66
Supply	5	72	16	7

Source: Authors' compilation

It is suggested that future research be done to investigate the impact of the mismatch to business (e.g., opportunity loss), especially since hiring managers revealed that the average time to fill a DSA position was 46 days[16]. This delay in filling in a position certainly means opportunity loss to the company. Another future investigation can be on the impact of this mismatch on the country's unemployment, specifically youth unemployment.

It is likewise recommended that an investigation be made on whether this misalignment points to an over/undersupply of DSA professionals. Since the DSA-related programs were not explicitly designed to produce DSA professionals, the graduates who completed these non-DSA undergraduate programs can take many other non-DSA jobs. Moreover, it is also impossible to say how many among those "ready" to take those DSA job roles are interested in doing so. Thus, only a small portion of them will go to DSA. Should more companies decide to be data-driven and become analytically competitive organizations and if the IT-BPO industry expands to absorb more analytics work, the demand for DSA jobs will likewise expand. It would be impossible, however, to be explicit about over/undersupply of DSA professionals unless there is a count (or estimate) of the number of DSA jobs available (which is beyond the scope of the study).

Conclusion: Demand and Supply

The findings indicated a scarcity of DSA competencies in the current workforce and a misalignment between the demand and supply of DSA professionals in the country. While employers were looking for graduates enabled to perform the work of a functional analyst, HEIs were producing graduates fit to be data engineers. Consistently, DSA-related degree programs equipped their graduates with the basic proficiency of the competencies required to perform the tasks of a data engineer or data scientist.

Such misalignment can be attributed to the fact that DSA is still at its infancy in the Philippines. Hence, job roles were not particularized. Employers also run the risk of failing to hire the right worker for the right set of tasks. This was evident in the discrepancies between employers' expectations and AAP's definition of the DSA job roles. With

[16] Information is based on the companies that responded to the survey questionnaire for hiring managers and participants of the FGD among hiring managers on February 2, 2019 at UA&P.

this misalignment comes the impetus for the government to work on appropriate mechanisms. If not, youth unemployment may exacerbate.

Recommendations

Based on the foregoing, this study advances two main recommendations: the use and further development of AAP's framework to define the DSA profession and the promotion of industry-academe-government linkages that can help analytics activities in the country mature and expand into an industry and contribute more to the economy.

Use and further development of AAP's framework

Various strategies have been presented in relevant literature and recent conferences about FIRe. However, many of these strategies can successfully take off by first having a common understanding of this emerging market for DSA workforce.

Other economies have already begun such initiative and have been making huge steps in developing their respective DSA workforce. Examples of these initiatives are EDISON Data Science framework in Europe, BHEF in the US, and SkillsFuture in Singapore. Indeed, a framework that can facilitate a common understanding among various stakeholders can help in aligning the demand and supply of DSA workforce in the country. Currently, AAP's professional maturity model is continuously being enhanced and so can be an appropriate starting point especially in profiling the demand and supply of DSA competencies and job roles in the country.

1. In terms of demand, analytics companies or associations can make use of the framework as a basis for hiring and developing DSA talents. Moreover, relevant government agencies, such as the Department of Trade and Industry (DTI) and the Department of Information and Communications Technology (DICT) can also collaborate more efficiently with industry players through policies and programs fitted to their DSA needs.
2. In terms of the supply, CHED can make use of the framework in creating new standards for DSA degrees or updating existing CMOs of DSA-related degree programs to accommodate the DSA industry requirements.

HEIs can likewise use the framework to improve their existing DSA-related degree programs to increase their capacity in enabling the graduates to be industry-ready DSA workers. In a specific way, they can incorporate courses that particularly develop DSA competencies and aim to enable their graduates to perform any of the DSA job roles. In so doing, these degree programs can better educate more enabled DSA workforce.

HEIs can likewise use the framework to develop and offer undergraduate programs especially aimed for graduate data stewards, data engineers, data scientists, and functional analysts with DSA competencies.

3. This study puts forward the use of a common understanding of analytics. However, it covered only the demand of the IT-BPM sector and the supply of DSA workforce from 10 DSA-related undergraduate programs. Hence, this study recommends that this scope be expanded to capture a more inclusive picture of the demand and supply of DSA workforce in the country. Future studies can include more sectors and more undergraduate programs. In addition, they also consider other sources of supply of DSA workforce such as the online platforms (e.g., Coursera, Moot, etc.) or training centers that offer DSA courses.

Promotion of industry-academe-government linkages

A common definition of the DSA profession can facilitate industry-academe-government linkages aimed to align the demand and supply of DSA workforce.

1. Academe and government, specifically through CHED, can work together in establishing standards for degree programs in DSA or updating existing CMOs of degree programs.
2. The academe can involve the industry in improving curriculum, course design, and delivery of instructions or in codeveloping teaching materials. Companies can share their use cases and data.
3. Industry players can encourage their employees to teach and impart a practical background of DSA to students.
4. Companies can share the cost of DSA education, such as data laboratories, hardware, and software licenses.

5. Industry players can help intensify faculty development through faculty exchange programs and research collaborations aimed to increase teachers' theoretical and practical background of the field. An example of this enrichment currently being done is the Data Incubator Graduate Program of the Binghamton University. The program is an intensive eight-week fellowship that prepares master's students, PhDs, and postdoctoral students in STEM and social science fields for industry careers as data scientists. However, instead of fishing them out of the universities, companies can partner with universities for the expertise of their professors.
6. The government working closely with the industry and academe can play a major role in formulating policies and programs that promote DSA. The government's most relevant agencies are DTI to facilitate the evolution of analytics as an industry, DICT to enable the needed information and communications technology infrastructure for DSA, DOLE to facilitate the healthy balance between the demand and supply of DSA workforce, Department of Science and Technology to encourage the proliferation of research projects in DSA, PSA to include DSA in its databases and studies, and CHED to promote DSA education among the HEIs.

References

Aggarwal, C. and C. Zhai. 2012. *Mining text data.* New York, NY: Springer.

Analytics Association of the Philippines (AAP). 2018. 10 analytics competencies. Taguig City, Philippines: AAP.

Asia Pacific Economic Co-operation (APEC) Human Resource Development Working Group. 2017. Data science and analytics skills shortage: Equipping the APEC workforce with the competencies demanded by employers. Singapore, Singapore: APEC Secretariat. https://www.apec.org/Publications/2017/11/Data-Science-and-Analytics-Skills-Shortage (accessed on August 11, 2018).

———. 2019. Close the digital skills gap by 2025 through collaboration. Singapore, Singapore: APEC Secretariat. https://www.apec.org/Press/News-Releases/2019/0719_Digital (accessed August 1, 2019).

Arthur, D. and S. Vassilvitskii. 2007. k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Philadelphia. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Business Higher-Education Forum (BHEF). 2014. Data science emerges as essential tool for decisionmaking and innovation across industry sectors. Washington D.C.: BHEF.

Commission on Higher Education (CHED). 2019. Graduates by Specific Baccalaureate Programs_2013-2016 as of 15 March 2019. Unpublished (data request). Quezon City, Philippines: CHED.

———. n.d. CHED Memorandum orders. http://ched.gov.ph (accessed on January 2019).

Chen, Y., J. Bordes, and D. Filliat. 2016. An experimental comparison between NMF and LDA for active cross-situational object-word learning. Sixth Joint IEEE International Conference Developmental Learning and Epigenetic Robotics (ICDL-EPIROB), September 2016, Cergy-Pontoise, France.

Cheng, X., J. Guo, S. Liu, Y. Wang, and X. Yan. 2013. Learning topics in short texts by nonnegative matrix factorization on term correlation matrix. SDM. https://pdfs.semanticscholar.org/b5d0/36429877568a648389531e323ea0983a5148.pdf (accessed on April 2019).

Dadios, E.P. A.B. Culaba, J.R.G. Albert, V.B. Paqueo, A.C. Orbeta Jr., R.B. Serafica, A.A. Bandala, and J.C.A.C. Bairan. 2018. Preparing the Philippines for the Fourth Industrial Revolution: A Scoping Study: A Scoping Study. PIDS Discussion Paper Series No. 2018-11. Quezon City, Philippines: Philippine Institute for Development Studies. https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1811.pdf (accessed on February 2019).

Demchenko, Y. 2017. EDISON data science framework: Part 4. Data science professional profiles (DSPP) release 2. Education for Data Intensive Science to Open New science frontiers (EDISON) Project. Working Document. Workshop "Wissenschaft im Digitalen Wandel", University of Amsterdam, Mannheim, Germany, June 6. https://www.kfa-juelich.de/SharedDocs/Downloads/UE/DE/Controlling/Workshop_WiDW/17-07-07Demchenko.pdf?blob=publicationFile (accessed on August 27, 2018).

Department of Labor and Employment (DOLE) and SFI Group of Companies. 2018. Philippine talent map Initiative: Profiling the 21st century skills of the Philippine labor force. Manila, Philippines and Pasig City, Philippines: DOLE and SFI Group of Companies.

Elemia, C. 2016. Villanueva: We lack graduates who match jobs. *Rappler*. https://www.rappler.com/nation/146475-villanueva-we-lack-graduates-match-jobs (accessed on August 12, 2018)

Francisco et al. 2019. Mapping Philippine workers at risk of automation in the Fourth Industrial Revolution. Working Paper 2019-01. Makati City, Philippines: Asian Institute for Management Rizalino S. Navarro Policy Center for Competitiveness. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3366809 (accessed on March 23, 2019).

Iredale, R., P. Toner, T. Turpin, and M. Fernández-Esquinas. 2014. A report on the APEC Region Labour Market: Evidence of skills shortages and general trends in employment and the value of better labour market information systems. APEC Human Resources Development Working Group. http://skillsmap.apec.org/home/download_document_overview/09568860014023801442ASkillsMappingAPECregionreport-final.pdf (accessed on August 12, 2018).

MacASkill, A. 2018. What are the links between Cambridge Analytica and a Brexit campaign group? https://www.reuters.com/article/us-facebook-cambridge-analytica-leave-eu/what-are-the-links-between-cambridge-analytica-and-a-brexit-campaign-group-idUSKBN1GX2IO (accessed on December 16, 2019)

National Privacy Commission (NPC). n.d. A brief primer on Republic Act 10173. Manila, Philippines: NPC. https://www.privacy.gov.ph/data-privacy-act-primer/ (accessed on December 16, 2019)

Manning, C. and S. Hinrich, S. 1999. *Foundations of statistical natural language Processing*. Cambridge, MA: MIT Press.

Microsoft. 2018. Unlocking the economic impact of digital transformation in Asia Pacific: Microsoft Asia digital transformation study 2018 - Southeast Asia and the Philippines. Singapore, Singapore: Microsoft.

Pasion, P. 2017. Bill on job-skills mismatch hurdles final reading in Senate. *Rappler*. December 12.

Pelayo, S. 2019. Getting grounded on analytics. https://www.linkedin.com/pulse/getting-grounded-analytics-sherwin-pelayo/ (accessed on August 11, 2018).

Philippine Statistics Authority (PSA). 2009. Philippine Standard for Industrial Classification (PSIC). Quezon City, Philippines: PSA. https://psa.gov.ph/sites/default/files/PSA_PSIC_2009.pdf (accessed on March 20, 2019).

———. 2012. Labor and employment survey. Quezon City, Philippines: PSA.

———. 2013a. Labor and employment survey. Quezon City, Philippines: PSA.

———. 2013b. Annual Survey of Philippine Business and Industry. Quezon City, Philippines: PSA.

———. 2014. Annual Survey of Philippine Business and Industry. Quezon City, Philippines: PSA.

———. 2015. Annual Survey of Philippine Business and Industry. Quezon City, Philippines: PSA.

———. 2016. Annual Survey of Philippine Business and Industry. Quezon City, Philippines: PSA.

———. 2018. Compilation of Industry Statistics on Labor and Employment. Quezon City, Philippines: PSA. https://psa.gov.ph/sites/default/files/2018%20CISLE%20Final.pdf (accessed on March 18, 2019).

Pricewaterhouse Cooper and Business-Higher Education Forum (PwC and BHEF). 2017. Investing in America's data science and analytics talent: the case for action. Retrieved from http://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf.

Ramos, C. 2016. Skills issues, sources of skills issues, and policy responses in five ASEAN member-countries: Indonesia, the Philippines, Thailand, and Vietnam. Singapore, Singapore: The HEAD Foundation.

Röder, M., A. Both, and A. Hinneburg. 2015. Exploring the space of topic coherence measures. Shanghai, China: Association for Computing Machinery.

Schutt, R. & C. O'Neil. 2013. *Doing data science: straight talk from the frontline*. Newton, MA: O'Reilly Media Inc.

Tableau. 2016. The state of data education in 2016: How U.S. higher education responds to the data skills gap. https://www.tableau.com/data-education-2016 (accessed on August 11, 2018).

Tan, K.S. and J. Tang. 2016. New skills at work: Managing skills challenges in ASEAN-5. Research Collection School of Economics. Singapore, Singapore: Singapore Management University.

Tan, R. and L. Tan. n.d. Need for B.S. Data Science degree program in the Philippines. Philippine Statistics Authority. https://psa.gov.ph/sites/default/files/Session%203-8%20Need%20for%20B.S.%20Data%20Science%20Degree%20Program.pdf (accessed on January 2019).

The Glossary of Education Reform. 2016. 21st century skills. Portland, ME: Great Schools Partnerships. https://www.edglossary.org/21st-century-skills/ (accessed on December 16, 2019).

Wang, W. 2017. Job skills extraction. https://github.com/2dubs/Job-Skills-Extraction (accessed on April 2019).

Weiss, S., N. Indurkhya, and T. Zang. 2015. *Fundamentals of predictive text mining*. London, UK: Springer-Verlag.

World Bank (WB). 2016. Republic of the Philippines labor market review: Employment and poverty. Washington, DC: WB https://openknowledge.worldbank.org/handle/10986/24768 (accessed on March 18, 2019).

Xu, R. 2005. Survey of clustering algorithms. *IEEE Transactions on Neural Networks* 16(3):645–678.

Appendices

Appendix A. Job scraping methodology

A.1 Definition

- Definition 1: Tokenization is the process of segmenting a running text into elements of vocabulary, called words or tokens. (Weiss et al. 2015).

- Definition 2: Topic Modeling is a dimensionality reduction technique that links words of original texts with the same context and discovers themes that run through them, called topics (Aggarwal and Zhai 2012).

- Definition 3: Document Clustering is the segmentation of documents in a corpus into homogenous groups, using the words or terms within each document as features in the model. (Manning and Hinrich 1999).

A.2. Methodology overview

The online job scraping methodology followed four stages:

- Stage 1: Data Preparation and Exploration – involved a) scraping Jobstreet job posts and b) skills pattern matching and tokenization[17]. The industry-required skills were sourced from the job descriptions found in these job posts. Job posts from May 1, 2019 to July 21, 2019 were scraped from Jobstreet.com.ph, using keywords that are related to the data profession (e.g., roles, tools, and APEC DSA competencies). The online job scraping lasted for about three months and gathered a total of 150,869 unique job posts. For faster data recovery, 13.6 percent of these unique job posts (20,574) were used as a representative sample where the algorithms were performed.

[17] Tokenization, the process of segmenting a running text into elements of vocabulary, called words or tokens (Weiss et al. 2015) was first performed on the extracted skill sets. To remove the effect of noise, due to inclusion of equal opportunity statements, application instructions, and company benefits, a predetermined set of skills from Wang (2017) was used as vocabulary. Moreover, to enable classification of these skills into similar groups based on the pattern by which they appear in the job posts, their term frequency-inverse document frequency (TF-IDF) statistics were calculated.

- Stage 2: Topic Modeling - a dimensionality reduction technique that links words of original texts with the same context and discovers themes that run through them, called topics, followed afterwards.[18] This process extracted 30 skills groups from the job posts, which were then manually mapped into the 10 APEC-recommended DSA competencies by select industry experts.[19]
- Stage 3: Document Clustering – involved grouping the job posts into the four DSA job roles–data steward, data engineer, data scientist, and functional analyst.[20]
- Stage 4: Cluster Profiling – included holding meetings among field experts to refine the skills and job role groupings from the online job scraping. A survey questionnaire was also used to conduct key informant interviews with industry practitioners and human resource personnel of select IT-BPM companies to gain more insights about the job roles and competencies which employers mostly look for.

A.3. Methodology details

- Stage 1: Data Preparation and Exploration
  Model building procedure
  To determine the demand and industry-required competencies for the four DSA roles, the study performed the following tasks were performed:
  (1) Skills sets or competencies were extracted from the descriptions of online job posts.
  (2) Extracted jobs were categorized into groups with similar skills requirements or role responsibilities.

[18] Nonnegative matrix factorization (NMF), a nonprobabilistic topic modelling technique was employed in the analysis because it works better and faster when learning keywords or short texts (Cheng et al. 2013; Chen et al. 2016). Since NMF will not automatically provide the optimal number of skills groups, the selection of skills group was based on topic coherence score. This was used in the study since it is the most aligned with human interpretability (Röder et al. 2015).

[19] Weights per competency of each job post were calculated. This weight signifies the portion of work that requires the competency. The sum of the competency weights should equal to 100 percent.

[20] Document clustering is the segmentation of documents in a corpus into homogenous groups with the use of words or terms within each document as features in the model (Manning and Hinrich 1999). While there are several algorithms to implement clustering, K-means method was the one used in the analysis.

Online job post scraping

Job posts from May 1, 2019 to July 21, 2019 were scraped from Jobstreet.com.ph, using keywords related to the data profession (e.g., roles, tools, and APEC DSA competencies). Specifically, the lists of keywords are given in Table A.1, A.2, and A.3.

Table A.1. Keywords per APEC data science and analytics competency

Competency	Keyword
Domain knowledge and application	domain
	business
	procedure
Data management and governance	data security
	policy
	MDM
Operational analytics	process
	operations
	six sigma
	MIS
Data visualization and presentation	data storytelling
	charts
	report
Research methods	research
	experiment
Data engineering principles	data warehouse
	data lake
Statistical techniques	statistics
	qualitative
	ANOVA
Data analytics methods and algorithms	machine learning
	mathematical
	data mining
Computing	SQL
	software
	computing

APEC = Asia Pacific Economic Cooperation
Source: Authors' list based on APEC's definitions of the 10 data science and analytics competencies

Table A.2. Keywords per data science and analytics role

Job Role	Keyword
Data Steward	steward
	governance
	policy
	data security
Data Engineer	ETL
	data warehouse
	big data
	database
	data engineer
	programmer
	developer
	solution engineer
	solution architect
Data Scientist	statistic
	math
	statistical
	mathematical
	predictive
	algorithm
	data science
	actuarial
	statistician
Functional Analyst	industry
	domain
	operations
	business
	analyst

Source: Authors' list based on the Asia Pacific Economic Cooperation's definitions of the 10 data science and analytics competencies

Table A.3. Keywords for other categories

Category	Keyword
Tools	Python
	SAS
	SPSS
	Rapid Miner
	Knime
	R
	Eviews
	MInitab
	Informatica
	Oracle
	Teradata
	Tableau
	Power BI
	Stata
	Excel
	Access
	Qlik
	Matlab
	VBA
	analytics
Others	data

Source: Authors' list

Overall, 150,869 unique posts were scraped. The fields extracted include:

Table B. List of fields

Column Name	Description
job_post_id	Job Post ID
job_title	Job Title
company_name	Company Name
job_salary	Job Salary
yrs_exp	Years of Experience
job_location	Job Location
avg_processing_time	Average Processing Time
company_industry	Company Industry
company_size	Company Size
job_desc	Job Description
job_post_dt	Job Post Dt
job_close_dt	Job_Close Dt

Source: Authors' list

For skills extraction and job segmentation, job description was used.

For faster data discovery, a random sample of size 20,574 (13.6%) was selected from the 150,869 unique posts. The algorithms described were performed on this representative sample.

Skills pattern matching and tokenization

Based on a similar study done by Wang (2017), there was a need for a supervised approach in the extraction of features (e.g., skills, courses, responsibilities, tools) from the job descriptions. Most descriptions had a lot of noise, due to the inclusion of equal opportunity statements, application instructions, and company benefits. To remove the effect of these noises on the results of subsequent analyses, a predetermined set of skills from Wang (2017) was used as vocabulary (see Appendix B). From the job description column, another field for skills set was created, which only retained words or phrases that match the items in the predetermined skills list.

Figure A. Illustration of skills pattern matching

job_post_id	job_desc	skills_set
8767788	Strategic **Analytics** is a new, growing team at Arch. The team develops innovative predictive models and analytical tools to improve profitability, growth, and **operational efficiency**. This position will help support **predictive modeling** needs across a growing portfolio of high-profile **advanced analytics** projects with the Strategic **Analytics** team. As a key member of the team you will also play a role in advancing **predictive modeling** capabilities, enabling Arch to make better decisions. REQUIREMENTS:Predictive modeling experience in a professional setting, using tools such as **R**, **SQL**, **Python** or SASWilling to work on Mid Shift (3pm-12am | Mon-Fri)	['Predictive modeling', 'SQL', 'R', 'advanced analytics', 'analytical', 'analytics', 'Analytics', 'Python', 'operational efficiency', 'predictive modeling', 'modeling']

Source: Results from Authors' online job scraping

Unique skills were extracted from the skills sets (i.e., tokenization). For easy classification of these skills into similar groups based on the pattern by which they appear in the job posts, their term frequency-inverse document frequency (TF-IDF) statistics were calculated. TF-IDF measures the importance of a skill to a job post in a collection of job posts. It consists of two components (Weiss et al. 2015):

1) Term frequency (TF) is defined as:

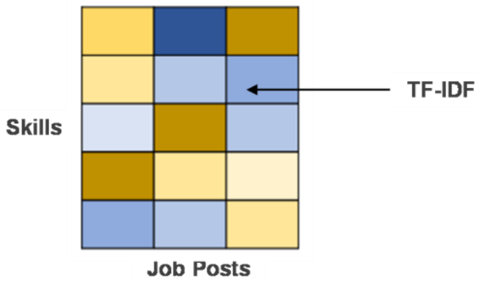
$$TF_{skill} = \frac{\textbf{Number of times the skill appears in a job post}}{\textbf{Total number of skills in a job post}}$$

2) Inverse Document Frequency is a scaling factor computed as:

$$IDF_{skill} = \log\left(\frac{\textbf{Total number of job posts}}{\textbf{Number of job posts with the skill in it}}\right)$$

Higher values for TF indicate that the skill appears more frequently in the job post. Lower values for IDF indicate that the skill is common across job posts. Hence, the higher the TF-IDF value, the more the skill can distinguish the job posts from each other. The structure of the table that was used in the succeeding analyses is as follows:

Figure B. Skill-job post matrix



TF-IDF = term frequency - inverse document frequency
Source: Results from authors' online job scraping

- Stage 2: Topic Modeling

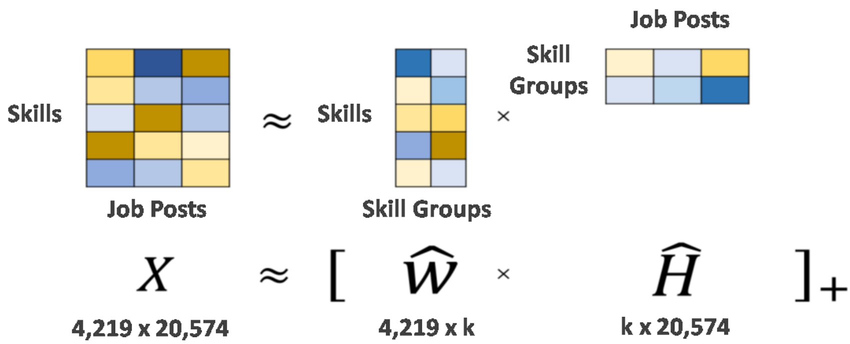
The 4,219 extracted skills still needed to be categorized into more interpretable groups or sets. This was done by topic modeling, where the underlying themes unifying the different skills were explored.

There were two classifications of topic modeling techniques, namely, probabilistic and nonprobabilistic. The most popular method under each was latent Dirichlet allocation and nonnegative matrix factorization (NFM), respectively (Cheng et al. 2013). The latter was employed in the analysis because it works better and faster than the former when learning keywords or short texts (Cheng et al. 2013; Chen et al. 2016).

NMF is a matrix factorization method that decomposes the skill-job post matrix $\mathbf{X}_{(m \times n)}$, into 2 matrices, $\hat{\mathbf{W}}_{(m \times k)}$ and $\hat{\mathbf{H}}_{(k \times n)}$, where m = number of skills (4,219), n = number of job descriptions (20,574), and k = number of skills group (refer to Figure C) [Chen et al. (2016)]. The matrix $\hat{\mathbf{H}}$ contains the mapping of skills groups to different job posts.

Since NMF will not automatically provide the optimal number of skills groups, the selection was based on the topic coherence score. Coherence score measures whether pairs of top skills in a skills group tend to cooccur together relative to

Figure C. NMF algorithm



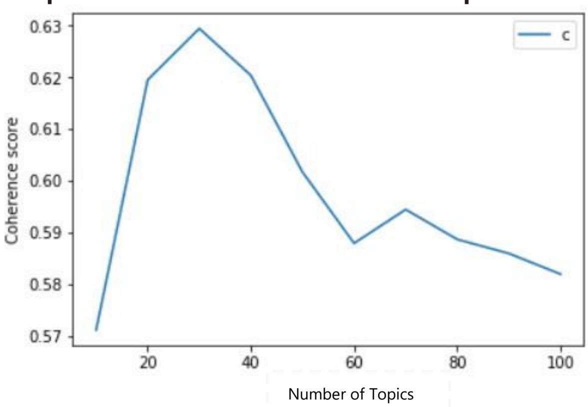
NMF = nonnegative matrix factorization
Source: Results from authors' online job scraping

the whole set of job posts. Several iterations of NMF, varying the number of skills groups from 10 to 100, were ran. For each iteration, the coherence score was computed per skill group, then aggregated. The iteration with the highest aggregated score determines the optimal number of groups.

Röder et al. (2015) reviewed different variants of coherence measures and found that Cv is the one most aligned with human interpretability. Hence, this was the score method used.

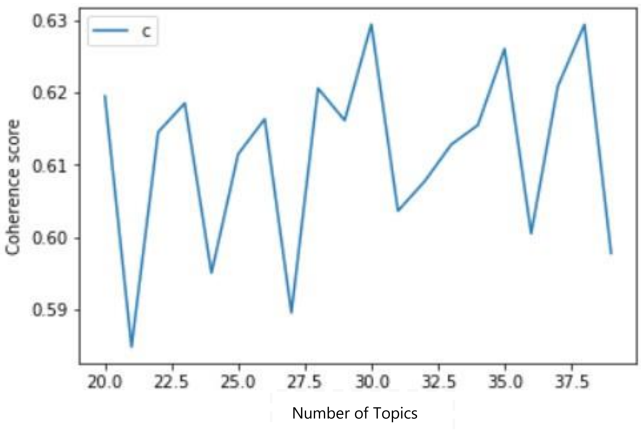
Results suggest that the topic coherence is highest when the number of skills groups chosen is 30 (Figure D).

Figure D.1. Topic coherence score for number of topics 10 to 100 by 10



Source: Results from authors' online job scraping

Figure D.2: Topic coherence score for number of topics 20 to 40 by 1



Source: Results from authors' online job scraping

For a better understanding of what the skills groups represent, they were visualized using word clouds. The size of the words or phrases shows the weight of the skills in characterizing the group. Refer to Figure E for a sample and to Appendix B for all the complete set of skills groups wordclouds.

Figure E. Sample skills groups word clouds

Skill group 14

Skill group 15

Source: Results from authors' online job scraping

The skills group-job post matrix (recall matrix $\hat{H}$) resulting from the NMF procedure shows the percentage of each skills group in the requirements of job positions. The sum of these weights is equal to 100 percent.

Figure F. Sample skills group weights by job post

job_post_id	Skills_Group_0	Skills_Group_1	Skills_Group_2	Skills_Group_3	Skills_Group_4	Skills_Group_5	Skills_Group_6	Skills_Group_7	Skills_Group_8	Skills_Group_9	Skills_Group_10	Skills_Group_11	Skills_Group_12	Skills_Group_13	Skills_Group_14	Skills_Group_15	Skills_Group_16	Skills_Group_17	Skills_Group_18	Skills_Group_19	Skills_Group_20	Skills_Group_21	Skills_Group_22	Skills_Group_23	Skills_Group_24	Skills_Group_25	Skills_Group_26	Skills_Group_27	Skills_Group_28	Skills_Group_29	
8782678	0.0%	9.8%	0.0%	3.2%	0.0%	18.1%	0.0%	4.3%	0.0%	0.0%	0.0%	0.0%	2.3%	11.1%	0.0%	0.0%	0.0%	0.0%	4.0%	22.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.6%	19.9%	0.0%	0.0%	100.0%
8781741	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.1%	0.0%	0.0%	0.0%	0.0%	6.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.0%	54.1%	0.0%	29.9%	100.0%
8781262	0.0%	0.0%	0.0%	0.0%	9.3%	0.0%	0.0%	0.0%	0.0%	4.1%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	0.0%	0.0%	0.0%	5.5%	0.0%	0.0%	14.8%	0.0%	0.0%	10.1%	0.0%	49.5%	0.0%	0.0%	100.0%
8779097	0.0%	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	0.0%	5.2%	0.0%	0.0%	0.0%	0.0%	0.0%	13.2%	1.4%	7.8%	0.0%	0.0%	12.3%	0.0%	4.3%	17.5%	0.0%	0.0%	0.0%	0.0%	33.5%	0.0%	1.9%	100.0%
8776774	24.7%	0.0%	0.0%	0.0%	14.3%	0.0%	0.0%	0.0%	4.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	41.5%	0.0%	0.0%	100.0%

Source: Results from authors' online job scraping

These 30 skills groups are then mapped to the 10 APEC DSA competencies.

Table C. Skills group–competency mapping

DSA Competency	Skills Group Number
Domain knowledge	0, 12, 21
Data governance	6, 11, 13, 16
Operational analytics	2, 8, 20
Data visualization	17
Data engineering	3, 4, 7, 14, 25, 27, 29
Statistical techniques	23
Methods and algorithms	15
Computing	1, 22, 24, 26
21st century skills	5, 9, 10, 18, 19, 28

DSA = data science and analytics
Source: Results from authors' online job scraping

To get the competency weights to be obtained per job post, the weights of the skills groups mapped to the same competency were combined (Figure G).

As in the skills group level, the sum of the competency weights should also equal to 100 percent. The weights represented the contribution of the competencies to the requirements of the job positions.

- Stage 3: Document Clustering
  Instead of grouping skills, the job posts were grouped to find among them the DSA roles—data steward, data engineer, data scientist, functional analyst, and analytics manager. The grouping was done by applying clustering analysis on the skill-job post-TF-IDF matrix.

  There are several algorithms to implement clustering (Xu 2005), but the k-means method was the one used in the analysis.

Figure G. Illustration of derivation of competency weights

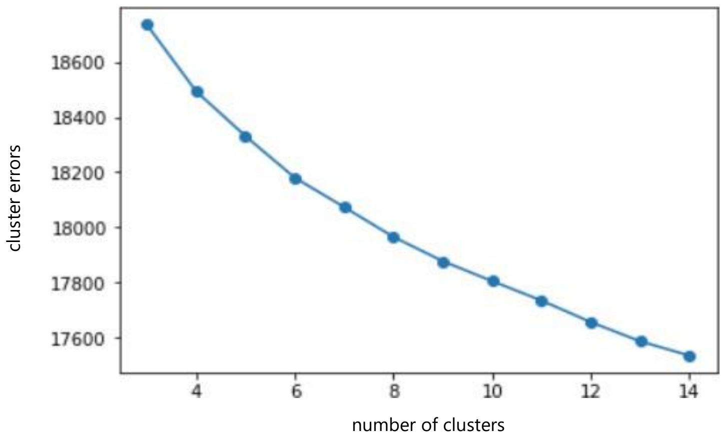
job_post_id	Skills_Group_21	Skills_Group_22	Comp_Domain_Knowledge
8782678	0.0%	0.0%	0.0%
8781741	0.0%	0.0%	0.0%
8781262	0.0%	14.8%	14.8%
8779097	4.3%	17.5%	21.8%
8776774	0.0%	0.0%	0.0%

job_post_id	Comp_Domain_Knowledge	Comp_Data_Governance	Comp_Ops_Analytics	Comp_Data_Viz	Comp_Data_Engg	Comp_Stat_Techniques	Comp_Algorithms	Comp_Computing	Comp_21stCentury_Skills	TOTAL
8782678	0.0%	11.1%	44.9%	9.8%	4.6%	4.3%	19.9%	3.2%	2.3%	100.0%
8781741	0.0%	5.6%	10.4%	0.0%	29.9%	0.0%	54.1%	0.0%	0.0%	100.0%
8781262	14.8%	10.1%	14.8%	0.0%	0.0%	6.7%	49.5%	4.1%	0.0%	100.0%
8779097	21.8%	0.0%	28.3%	0.0%	7.1%	9.2%	33.5%	0.0%	0.0%	100.0%
8776774	0.0%	0.0%	14.3%	24.7%	4.5%	15.0%	41.5%	0.0%	0.0%	100.0%

Source: Results from authors' online job scraping

As with NMF, the k-means algorithm does not automatically suggest the best number of clusters. Hence, the number of groups from 2 to 15 was tested. Inertia, the within-cluster sum of squares, was obtained from each iteration. Plotting the clustering errors against the number of clusters, the point at which the improvement starts to become marginal is ideally chosen as the "elbow criterion" (stopping criterion) (Arthur and Vassilvitskii 2007).

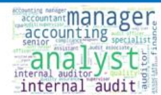
Figure H. Inertia by number of clusters



Source: Results from authors' online job scraping

Since the results did not show a clear elbow, iterations with the number of groups from 4 to 10 were revisited (close to the number of data roles the study wants to discover). The number of job groups was subjectively selected such that the top skills per group are as related as possible and the skills across job groups are as different as possible. The differentiation of the job titles across job groups was also considered in the selection of the number of groups. Ten job groups were generated.

Figure I. Top terms of the final set of clusters

					
	DATA SCIENTIST	LICENSED MANAGER / ANALYST	DATA STEWARD	DATA ENGINEER	AUDIT MANAGER / ANALYST
Top Terms (Terms Closest to Cluster Centroids)					
1	analysis	professional license	security	computer	audit
2	report	license	infrastructure	information technology	report
3	analytics	bar	management	science	compliance
4	research	pass	computer	computer science	internal audit
5	statistical	board	troubleshoot	technology	management
6	management	report	information security	telecommunication	accountancy
7	analytical	finance	technology	software	finance
8	statistics	management	software	database	internal control
9	data analysis	budget	configuration	sql	cpa
10	intelligence	accountancy	cloud	design	analysis

Source: Results from authors' online job scraping

Figure I. (continued)

Top Terms (Terms Closest to Cluster Centroids)	FINANCE ANALYST	SOFTWARE / APP DEVELOPER	INVENTORY / WAREHOUSE MANAGER	TECH SUPPORT SPECIALIST	ANALYTICS MANAGER
1	finance	design	inventory	customer service	management
2	accountancy	software	warehouse	communication	report
3	report	javascript	management	communication skills	communication
4	management	java	report	english	coordinate
5	analysis	computer	warehouse management	management	design
6	payroll	sql	coordinate	report	leadership
7	budget	developer	logistics	interpersonal	compliance
8	payments	html	inventory management	computer	analysis
9	cpa	css	inventory control	technical support	implementation
10	reconciliation	software development	sales	bpo	research

Source: Results from authors' online job scraping

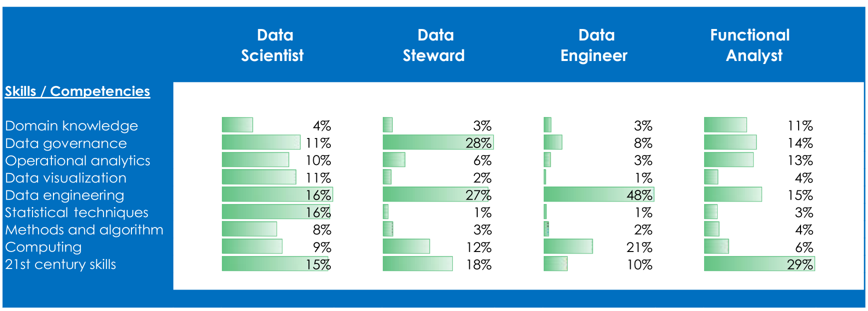
- Stage 4: Cluster Profiling
  These clusters were further refined, following the reclassifications in Table D:

Table D. Job group reclassification of data science and analytics job roles

New Groupings	Original Clusters
Data Scientist	Data Scientist
Licensed Manager/Analyst	Dropped
Data Steward	Data Steward
Data Engineer	Data Engineer, Software/ Application Developer
Functional Analyst	Audit Manager/Analyst, Finance Analyst, Inventory/Warehouse Manager, Technical Support Specialist, Analytics Manager

Source: Results from authors' online job scraping

Figure J. Employers' data science and analytics job role profiles



Source: Results from authors' online job scraping

Appendix B. Distribution table of DSA employees' degree programs

DSA Degree Program	Count	Percent Count (in %)
Computer Science	11	10
Business Administration	11	10
Statistics	10	9
Applied Math/Math	9	8
Information Technology	8	7
Information Science	7	6
Economics	6	5
Physics/Applied Physics	6	5
Industrial Engineering	4	4
Civil Engineering	4	4
Electrical Engineering	3	3
Electronics and Communications Engineering	3	3
Accountancy	3	3
Geodetic Engineering	3	3
Communications-related	3	3
Chemistry	2	2
Humanities	2	2
Political Science	2	2
Psychology	2	2
Biology	1	1
Commerce	1	1
Computer Applications	1	1
Computer Engineering	1	1
Fine Arts	1	1
Legal Management	1	1
Mechanical Engineering	1	1
Public Health	1	1
Sociology	1	1
Tourism	1	1
Geography	1	1
Total Participants	**110**	**100%**

DSA = data science and analytics
Source: Survey responses from data science and analytics-related practitioners

Appendix C. Employment share per major industry group (2013–2017)

By Major Industry Group	Employment Share (in thousands)					Compound Annual Growth rate* (in %)
	2013	2014	2015	2016	2017	
Agriculture	11,836	11,801	11,294	11,064	10,261	-3
Industry	5,936	6,166	6,275	7,159	7,371	4
Services	20,345	20,682	21,172	22,775	22,703	2
Total	**38,117**	**38,649**	**41 228**	**40,998**	**40,335**	

Note: *computed by authors based on standard compound annual growth rate formula
Source: PSA (various years) as cited from JobsFit 2022 Report by Bureau of Local Employment, Department of Labor and Employment; PSA CISLE Report, 2018

Appendix D. Employment size of industry subclasses of BPM subsector

Section	Code	PSIC Industry subclasses (non-BPM)	Employees			
			2013	2014	2015	2016
BPM (sections I and J)	J58190	Other publishing activities	492	466	349	394
I: Information and Communication	J58200	Software publishing	924	1,039	2,281	1,852
N: Admin Support and Services	J59110	Motion picture, video and television programme activities	67	284	445	418
	J59120	Motion picture, video and television programme post-production activities	935	846	939	708
	J62010	Computer programming activities	36,023	35,951	39,503	38,808
	J62020	Computer consultancy and computer facilities management activities	12,041	10,766	14,609	11,476
	J62090	Other information technology and computer service activities	3,833	4,372	4,049	3,780

Appendix D. (continued)

Section	Code	PSIC Industry subclasses (non-BPM)	Employees 2013	2014	2015	2016
	J63111	Data processing	19,163	18,212	17,684	14,052
	J63112	Website hosting services	533	738	294	905
	J63113	Application hosting services	721	564	588	448
	N78103	Online employment placement agencies	106	104	73	78
	N82211	Customer relationship management activities	298,485	354,140	390,799	420,763
	N82212	Sales and marketing (including telemarketing) activities	52,940	72,539	72,546	80,442
	N82219	Other call centers activities (voice), nec	7,366	11,070	9,265	9,512
	N82221	Finance and accounting activities	6,948	6,699	6,809	7,227
	N82222	Human resources and training activities	404	512	2,742	2,434
	N82223	Administrative support activities	591	443	1,484	1,691
	N82224	Document processes activities	189	119	173	928
	N82225	Payroll maintenance and other transaction processing activities	39	42	297	700
	N82226	Medical transcription activities	2,066	2,058	1,886	2,945

Appendix D. (continued)

Section	Code	PSIC Industry subclasses (non-BPM)	Employees 2013	2014	2015	2016
	N82227	Legal services activities	s	s	s	s
	N82228	Supply chain management activities	s		s	s
	N82229	Other nonvoice related activities, nec	2,008	2,290	2,850	6,627
	N82291	Engineering outsourcing activities	806	937	822	1,562
	N82292	Product development activities	s	s		
	N82293	Publishing outsourcing activities	2,802	596	824	495
	N82294	Research and analysis activities	332	162	183	291
	N82295	Intellectual property research and documentation activities		s	s	144
	N82296	Security outsourcing activities	91		s	
	N82299	Other back office operations activities, nec	nad	823	1,618	1,141
		TOTAL	**449,905**	**525,772**	**573,112**	**609,821**

PSIC = Philippine Standard Industrial Classification; BPM = business process management;
s = suppressed by the publisher; nec = not elsewhere classified; nad = no available data
Source: PSA (2018)

Appendix E. Employment size of analytics-related activities from PSIC's list and definition of industries (2014–2016)

Industry (based on PSIC)	Code	Industry subclasses	2014		2015		2016	
			Employees	Percent (%)	Employees	Percent (%)	Employees	Percent (%)
C: Manufacturing	C26200	Manufacture of computers and peripheral equipment and accessories	60,096	34.41	73,462	36.80	58,224	33.52
K: Financial and Insurance Services	K66290	Other activities auxiliary to insurance and pension funding	aggregated	NA	255	0.13	191	0.11
M: Professional, Scientific and Technical Activities	M70200	Management consultancy activities	24,795	14.20	26,862	13.45	27,396	15.77
	M71109	Other technical activities related to architectural and engineering	nad	NA	2,530	1.27	2,902	1.67

Appendix E. (continued)

Industry (based on PSIC)	Code	Industry subclasses	2014		2015		2016	
			Employees	Percent (%)	Employees	Percent (%)	Employees	Percent (%)
	M72101	Research and experimental development in natural sciences	1,557	0.89	610	0.31	723	0.42
	M72102	Research and experimental development in engineering and technology	nad	NA	688	0.34	342	0.20
	M72103	Research and experimental development in health sciences	nad	NA	497	0.25	465	0.27
	M72104	Research and experimental development in agricultural sciences	nad	NA	245	0.12	257	0.15

Appendix E. (continued)

Industry (based on PSIC)	Code	Industry subclasses	2014		2015		2016	
			Employees	Percent (%)	Employees	Percent (%)	Employees	Percent (%)
	M72200	Research and experimental development on social sciences and humanities	886	0.51	531	0.27	1,034	0.60
	M72300	Research and experimental development in information technology	2,604	1.49	3,492	1.75	3,686	2.12
	M72400	Research and experimental development services, nec	56	0.03	18	0.01	s	NA
	M73200	Market research and public opinion polling	8,648	4.95	6,728	3.37	5,713	3.29
	M749	Other professional, scientific, and technical activities, nec	3,400	1.95	2,921	1.46	nad	NA

Appendix E. (continued)

Industry (based on PSIC)	Code	Industry subclasses	2014		2015		2016	
			Employees	Percent (%)	Employees	Percent (%)	Employees	Percent (%)
J: Information and Communication	J58200	Software publishing	1,039	0.59	2,281	1.14	1,852	1.07
	J62010	Computer programming activities	35,951	20.58	39,503	19.79	38,808	22.34
	J62020	Computer consultancy and computer facilities management activities	10,766	6.16	14,609	7.32	11,476	6.61
	J62090	Other information technology and computer service activities	4,372	2.50	4,049	2.03	3,780	2.18
	J63111	Data processing	18,212	10.43	17,684	8.86	14,052	8.09
	J63112	Website hosting services	738	0.42	294	0.15	905	0.52

Appendix E. (continued)

Industry (based on PSIC)	Code	Industry subclasses	2014		2015		2016	
			Employees	Percent (%)	Employees	Percent (%)	Employees	Percent (%)
	J63113	Application hosting services	564	0.32	588	0.29	448	0.26
N: Administrative and Support Services	N82294	Research and analysis activities	162	0.09	183	0.09	291	0.17
	N82299	Other back office operations activities, nec	823	0.47	1,618	0.81	1,141	0.66
		TOTAL	174,669	100.00	199,648	100.00	173,686	100.00

PSIC = Philippine Standard Industry Code; s = suppressed by the publisher; NA = not applicable; nec = not elsewhere classified; nad = no available data
Source: PSA (2013b; 2014; 2015; 2016)

The Authors

Brenda A. Quismorio is a professor and the program director of the Master in Applied Business Analytics program of the University of Asia and the Pacific (UA&P). She also co-chairs the Professional Development and Education Committee of the Analytics Association of the Philippines. She holds a PhD in Business Administration from the University of the Philippines (UP) Diliman. Her areas of specialization include process re-engineering, application development and maintenance, quality management, and data services work for global and local firms.

Maria Antonette D. Pasquin is a junior researcher at the UA&P. She holds an MA degree in Political Economy, with specialization in International Relations and Development, from the same university. She has worked on several research projects related to her research interests, particularly the labor market.

Claire S. Tayco is head of research and analytics at CirroLytix Research Services, where she focuses on driving social impact through analytics, in areas such as public health, education, and human rights. She obtained her MS in Statistics degree from UP Diliman.

The rapid advancement in technology has facilitated far-reaching use of data. This has consequently led to an increasing demand for data science and analytics (DSA) professionals. However, recent studies show that such demand is often not met in many economies. Such DSA skills shortage is claimed to be rooted in the mismatch between the skills the industry demands and the skills academic institutions supply. As the first step in addressing this shortage of industry-ready DSA workers, it is necessary to know the DSA skills demanded by the industry and the DSA skills with which academic institutions equip their students. To do this, the study employed the Analytics Association of the Philippines' Professional Maturity Model, which is based on the 10 APEC-recommended DSA competencies, as analytical framework. The study reveals the misalignment between the demand and the supply of the DSA workforce in the country. Specifically, there are DSA competencies sought by employers that are poorly supplied by existing DSA-related undergraduate programs. To allow the growth and maturity of the still-infant analytics industry, there is a need to address the current lack of a common definition of the analytics profession among stakeholders. The study further endorses the promotion of government-industry-academe linkages to expand the existing market for DSA workforce in the country.

Philippine Institute for Development Studies
Surian sa mga Pag-aaral Pangkaunlaran ng Pilipinas

18th Floor, Three Cyberpod Centris - North Tower
EDSA corner Quezon Avenue, Quezon City, Philippines
Tel.: (+632) 8877-4000
Email: publications@mail.pids.gov.ph
Website: http://www.pids.gov.ph

facebook.com/PIDS.PH   twitter.com/PIDS_PH