DECEMBER 2019

DISCUSSION PAPER SERIES NO. 2019-34

Assessing the Alignment of Philippine Higher Education with the Emerging Demands for Data Science and Analytics Workforce

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

CONTACT US:

RESEARCH INFORMATION DEPARTMENT
Philippine Institute for Development Studies

18th Floor, Three Cyberpod Centris - North Tower
EDSA corner Quezon Avenue, Quezon City, Philippines

publications@mail.pids.gov.ph
(+632) 8877-4000

https://www.pids.gov.ph

Assessing the Alignment of Philippine Higher Education with the Emerging Demands for Data Science and Analytics Workforce

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

PHILIPPINE INSTITUTE FOR DEVELOPMENT STUDIES

December 2019

Abstract

Rapid advancement in technology has allowed for 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. This mismatch is evident in the Philippines where studies also reveal certain difficulties of current Philippine education and training to meet 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 in the Philippines where graduates land jobs which their completed education did not intend for them.

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' (AAP) Professional Maturity Model, which is based on the ten APEC-recommended DSA competencies, as analytical framework. It estimated the current availability of DSA competencies using available data from the Labor Force Survey (LFS) and interviews with companies engaged in analytics activities. Meanwhile, the profiling into four DSA job roles (Data Steward, Data Engineer, Data Scientist and Functional Analyst) of these workers was done with the use of relevant data from online job postings. On the other hand, the current supply for DSA job roles were mainly determined from survey interviews of analytics practitioners and undergraduate program administrators and CHED's databases.

Findings showed that in terms of the demand-side, the Functional Analyst job role emerges as the most sought role of employers among the four DSA- related job roles. The demand for these DSA-roles also mostly come from the Information and Communication industry group under the Services sector. For each demanded job role, the employers' require differing DSA competencies and levels of proficiency. The findings also revealed that one challenge for employers in hiring DSA talents is the inadequate 21st century skills more than other technical skills.

With regard to the supply-side, top ten undergraduate programs were identified to be DSA – related programs based on the common degrees of current analytics practitioners. The extent to which these DSA-related programs prepare their respective graduates with the basic proficiency of each competency varies. However, the curricula assessments reveal that these degree programs mostly equip their graduates with data engineering and statistical techniques competencies and correspondingly enable them to perform the Data Engineer job role.

With these results, the study evinced a misalignment between the demand and supply of the DSA workforce in the country. Specifically, there are DSA competencies – both sought by employers and required by the analytical framework of the study – that are poorly supplied by these DSA-related undergraduate programs. Moreover, while employers mostly look for Functional Analysts, most of the graduates of the degree programs are enabled to perform the Data Engineer job role. Because of these findings, the study offers two main recommendations. First, it recommends the use of the AAP Professional Maturity Model to define the DSA profession. To allow the growth and maturity of the still an emerging analytics industry, there is a need to address the current 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 further endorses the promotion of government-industry-academe linkages to expand the existing market for DSA workforce in the country.

Keywords: data science and analytics, analytics, digital transformation, skills mismatch

Table of Contents

List of Tables

List of Figures

Assessing the alignment of Philippine higher education with the emerging demands for Data Science and Analytics workforce

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

1. Introduction

The rapid and continuous advancement in technology allowed big data[1] to be more accessible nowadays for decision-makers in governments, businesses or any organization. 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 (McKinsey as cited in 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 data science and analytics as key transformational components to their core operations (Business Higher Education Forum 2014 p. 1)." The recent study of Microsoft about Digital transformation in the Asia Pacific shows that big data and analytics topped as the third mostly invested technology in 2018 next to Security and Mobility (Microsoft, 2018).

This growing interest in DSA consequently causes an uptrend in demand for professionals from this field. However, recent studies reveal that such demand is often not met. Developing a workforce with DSA skills continue to be a challenge for many economies (Asia Pacific Economic Cooperation Human Resource Development Working Group, 2017; Pricewaterhouse Cooper and Business-Higher Education Forum, 2017). In fact, next to Cybersecurity skills, DSA skills are identified by 95% of the employers as most difficult to find among prospective employees in the United States (PwC & 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 and/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 often 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, thus, 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 the skills of data science and analytics (APEC Human Resources Development Working Group, 2019). For instance, the United States has started to gear up its universities with as many DSA degree programs (Tableau, 2016). The APEC Human Resource Development Working Group (2017) recommended a set of DSA competencies to serve as a reference for academe, the industry and government sectors in institutionalizing DSA.

* Assistant Professor (University of Asia and the Pacific), Junior Researcher (University of Asia and the Pacific) and Senior Analytics Researcher (Cirrolytix), respectively.

[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.

With the advent of increasing demand for Data Science and Analytics, the lack of employment opportunities, often claimed for graduates of tertiary education in the Philippines, can be addressed. However, this idea still posts difficulties since DSA skills are still poor or underdeveloped among the workforce in the Philippines (Tan & 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 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 & Tang 2016, p. 100).

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

2. Objective/s

This paper aims to assess existing Philippine education specifically the curricula of Data Science and Analytics-related undergraduate degree programs insofar as they meet the DSA competency requirements of the industry. In doing so, the paper augments the main question with the following specific questions:

> Are Philippine undergraduate programs able to supply the current demands for Data Science and Analytics (DSA) talents?

i. What is the current demand for DSA-related jobs 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 jobs in accordance with employers' expectations and framework's requirements?

ii. What is the current supply of DSA-related jobs in the Philippines?

a. What are the DSA-related undergraduate programs?
b. Which DSA competencies do these DSA-related programs equip their graduates with basic proficiency?
c. Which DSA jobs do these DSA-related programs enable their graduates to perform?

3. Significance of the study

How can the Philippines prepare its workforce for the "Fourth Industrial Revolution[2]" or future jobs? This study takes the first step in assessing the current ability of Philippine educational institutions in enabling the future workforce for the Data Science and Analytics (DSA) jobs needed by industries. This study then aims to recommend courses of actions that education, industry and government sectors can do to ensure the availability of adequate supply of DSA workforce. In addition, this paper sets the stage for further studies on DSA in the Philippines given the dearth in the literature on this subject.

4. Scope and limitation

Ten DSA-related undergraduate programs were identified from the ten top undergraduate programs of surveyed DSA practitioners. Each DSA-related undergraduate program was evaluated by academe experts using the sample curriculum found in the respective Memorandum Order (CMO) by the Commission on Higher Education (CHED). The sample curriculum contains the minimum requirements of the degree program and serves as a guide for Higher Education Institutions (HEIs) in developing their actual curriculum. 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 and evaluation tools, etc. This study used only the experts' assessment on the sample curriculum i.e., excluded the expert's assessment on their actual curriculum as it did not represent the programs delivered by the other HEIs.

Data on the demand for DSA workforce is very scarce given that the field is just beginning. The data used were mostly sourced from government agencies' published reports on employment statistics and employers' online job postings which points to the demand coming mostly from the Information Technology and Business Processing Management (IT-BPM) industry. Other industries, though they may also demand for DSA workforce, are not included in the study.

5. Review of related literature

5.1 Digital transformation, big data and skills mismatch in the Philippines

The wave of digital transformation crashes across various economies and the Philippines is not exempt from such phenomenon. In fact, the recent study of Microsoft (2018) highlights the economic impact that digital transformation may bring to the country. It projects that by 2021; about 40% of the Philippine Gross Domestic Product (GDP) would be credited to digital transformation. Specifically, it would contribute 0.4% of the projected 6.9% growth in GDP. The same study nonetheless enumerates other positive socio-economic consequences such as 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, Microsoft (2018) presents 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.

[2] Refer to Dadios et. al (2018), PIDS' Scoping Study on the Fourth Industrial Revolution for further explanation of the phenomenon.

Microsoft (2018) mentions that rather than displacement, job transformation would take place. In the next three years, 92% of the Philippine jobs would be transformed. Of these transformed jobs, 34% would be retained and upskilled, 31% would be new roles due to digital transformation investment and 27% would be outsourced, automated or made redundant. Meanwhile, only 8% of the jobs would remain unchanged. The sectors most vulnerable to automation, according to Francisco et. al (2019) are the 1.) youth (age 15-24), 2.) those who have received less education, 3.) those with lower incomes, 4.) those employed in casual or irregular and seasonal jobs or working without pay and 5.) those belonging to the agriculture, forestry and fishing sectors.

With digital transformation comes the exponential growth of data. BHEF (2014, p. 1) quotes 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 the United States, and big data has the potential to create three times that number of jobs outside of IT." In 2017, the growth of big data in the United States was projected at 40% each year with potential value of $300 billion to the nation's health care industry alone (BHEF, 2017).

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

Table 1. Projected DSA workforce demand in select economies

Economy	Current DSA Workers	Projected DSA Workers Needed	Percent 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 between the demand and the supply of needed skills (skill shortage or gap) that gives rise to difficulties in recruitment processes (Iredale et. al as cited in Ramos 2016) signifies the inability of educational institutions to meet industry demands (Tan & Tang, 2016; Elemia, 2016). This "school-industry gap," an arguably most serious hurdle in developing an effective workforce, is felt across industries (Tan & Tang 2016, p. 108).

The Labor and Employment Statistics Survey of the Philippines (PSA, 2012/2013) especially highlights the shortage of high skills, as those in the field of Science and Technology (Ramos, 2016). Specifically, graduates who join the Information Technology-Business Processing Outsourcing (IT-BPO) industry and emerging infrastructure management services lack the necessary IT skills and Infrared (IR) software proficiency (Tan & 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 non-voice segments such as Knowledge Processing Outsourcing (KPO) of business analytics, insurance services, and health information management (Tan& Tang, p. 95).

Earlier efforts and interventions[3] have been done throughout the years. Recent efforts to augment the gap were manifested in the proliferation of Data Science degree programs in various universities and online courses on Data Science such as Coursera, a MOOC (Massive Open Online Course) and Microsoft's Professional Degree in Data Science (Tan & Tan, n.d.). Hence, it is further suggested that this rising demand be addressed by encouraging schools with existing B.S. Statistics and B.S. Computer Science degrees to also consider offering a B.S. Data Science (Tan & Tan (n.d.).

Despite these initiatives, the problem persists as reflected in the rising youth unemployment. Such situation, thus, warrants the assessment of current Philippine education on its ability to equip its graduates with the set of competencies needed by industry sectors. This paper, thereby, seeks to address this gap by assessing, in particular, the DSA -related degree programs currently offered by various Higher Education Institutions to find out whether or not there is indeed skills shortage on DSA competencies.

5.2. Data Science and Analytics

Data Science and Analytics (DSA), as a new field, has no universally accepted definition. One reason provided for this is that DSA involves 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.

Metamarket CEO Mike Driscoll defines Data Science as "the civil engineering of data (Schutt & O'Neil 2013, p. 7). He further claims that definitions following this idea of a Data Science, such as Conway's Venn Diagram of Data Science (2010), possess both the practical knowledge of tools and materials and theoretical understanding of the multidisciplinary nature of the field (Schutt & O'Neil 2013).

Meanwhile, the National Institute of Standards and Technology (NIST) distinguishes the meaning of Data Science and Analytics. "Analytics" is defined as a step in the data life cycle that involves the collection of raw data, preparation of information then analytics, visualization and access ("synthesis of knowledge from information"). On the other hand, "Data Science"

[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 Human Development and Poverty Reduction Cluster namely, 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) were the four identified programs under such policy action. Earlier on, 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, another Senate Bill 1456 or the Philippine Qualifications Framework (PQF) Act of 2017, which just passed in its 3rd and final reading also served as a policy action towards job-skill alignment (Pasion, 2017).

refers to the “extraction of actionable knowledge directly from data through either a process of discovery, or hypothesis formulation and hypothesis testing (NIST as cited in PwC and BHEF 2017, p. 3).”

Using these two concepts “analytics” and “data science, PwC and BHEF (2017) formulate the landscape of DSA job roles as seen in Table 2 below”. It classifies DSA job roles into analytics-enabled jobs and data science jobs. Professionals who perform analytics-enabled jobs are Data-driven Decision Makers 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,” the Data Analysts who “leverage data analysis and modeling techniques to solve problems and glean insight across functional domains” and the 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 to these five Data Science and Analytics Job roles.

Demchenko (2017) extends the discussion and provides 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, others)
- Data Science Engineering (including Software and Applications Engineering, Data Warehousing, Big Data Infrastructure and Tools)
- Domain Knowledge and Expertise (Subject/Scientific domain related)
- Data Management and Governance (including data stewardship, curation, and preservation)
- 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 decision-makers	**Functional Analysts**	**Data Engineer**	**Data Analyst**	**Data Scientist**
Common job titles: Chief Executive Officer Chief Data Officer Chief Information Officer Director of IT Financial manager Human resources manager Marketing manager	Common job titles: Actuary Business/Management analyst Compensation/Benefits analyst Financial analyst Geographer/GIS specialist HRIS analyst Operations analyst Researcher	Common job titles: Business intelligence architect Computer systems engineer Data warehousing specialist Data administrator Database architect Systems analyst	Common job titles: Data mining analyst Business intelligence analyst	Common job titles: Biostatistician Data engineer Data scientist Financial quantitative analyst Statistician

Source: PwC and BHEF (2017, p. 4)

The APEC Human Resource Development Working Group (2017, p. 3) adopts the definition of Data Science and Analytics 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 (APEC Human Resource Development Working Group, 2017, p. 5):

- Data Science: The data science field addresses both structured and unstructured data in terms of data cleansing, preparation, and analysis. Overall, data science 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: The big data field addresses the massive volumes of data that are 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 decision-making and business strategies. Financial services, retail, and communication all generate big data and thus require skilled techniques.
- Data Analytics: The data analytics 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 he or she already knows.

Two definitions in the Philippine literature come from Tan & Tan (n.d, p.4) and the Analytics Association of the Philippines. For Tan & 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 decision makers." On the other hand, the recently established Analytics Association of the Philippines (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 (AAP 2018).”

6. Definition of terms

In the absence of an agreed definition, this study takes on the operational definition of “Analytics” by AAP for Data Science and Analytics (DSA). It is deemed that with this definition, the specific attributes identified by varying references about DSA are captured. In this paper, the discourse about the field of DSA is further explored especially with regard to the set of competencies supplied by HEIs and demanded by companies in the Philippines. It is through this approach that the paper attempts to contribute to the on-going development of the field of DSA and narrowing the gap between what DSA is in theory and practice.

Analytics is “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 of society (AAP 2018).”

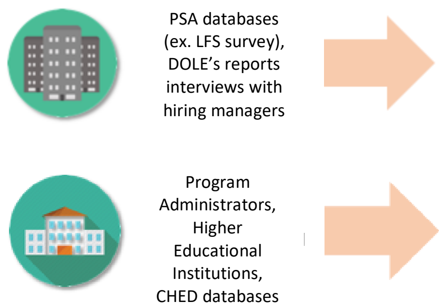
7. Research design and methodology

7.1. Conceptual framework

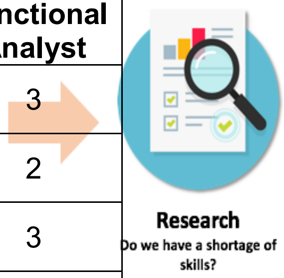
The conceptual framework (Figure 1) makes use of the Professional Maturity Model of the Analytics Association of the Philippines (Table 3) as the point of alignment between the supply and demand for the DSA workforce.

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

Analytics Association of the Philippines' (AAP's) professional maturity model[4]



DSA Competencies	DSA Professional Job Roles			
	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



Source: AAP (2018); Authors' illustration

The 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.

Each job role is characterized by the required proficiency level of the 10 APEC Recommended DSA competencies. These set of DSA competencies were developed by a 50-person Advisory Group composed of representatives from the industry, academe, and governments of 14 various APEC-member countries (APEC Human Resource Development Working Group 2017).

[4] For the updated AAP Model i.e., v1.0, see Pelayo (2019).

- Operational Analytics - use general and specialized business analytics/intelligence techniques for the investigation of all relevant data to derive insight for decision-making.
- Data Visualization and Presentation - create and communicate compelling and actionable insights from data using visualization and presentation tools and technologies.
- Data Management and Governance - develop and implement data management strategies, incorporating privacy and data security, policies and regulations, and ethical considerations.
- Domain Knowledge and Application - apply domain-related knowledge and insights to effectively contextualize data, achieved by practical experience and exposure to emerging innovations.
- Statistical Techniques - apply statistical concepts and methodologies to data analysis.
- Computing - apply information technology and computational thinking, and utilize programming languages and software and hardware solutions for data analysis.
- Data Analytics Methods and Algorithms - Implement and evaluate machine learning methods and algorithms on the data to derive insights for decision-making.
- Research Methods - utilize the scientific and engineering methods to discover and create new knowledge and insights.
- Data Science Engineering Principles - use software and system engineering principles and modern computer technologies to develop data analytics applications.
- 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) re-skilling, and entrepreneurship.

Table 3. Analytics Association of the Philippines' (AAP's) professional maturity model

DSA Competencies	**DSA Professional Job Roles**			
	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

Note: Level 1 – basic proficiency; Level 2 – intermediate/average proficiency; Level 3 – advanced/expert-level proficiency
Source: AAP (2018)

7.2. Research Design, Instruments, and Participants

7.2.1. Demand-side

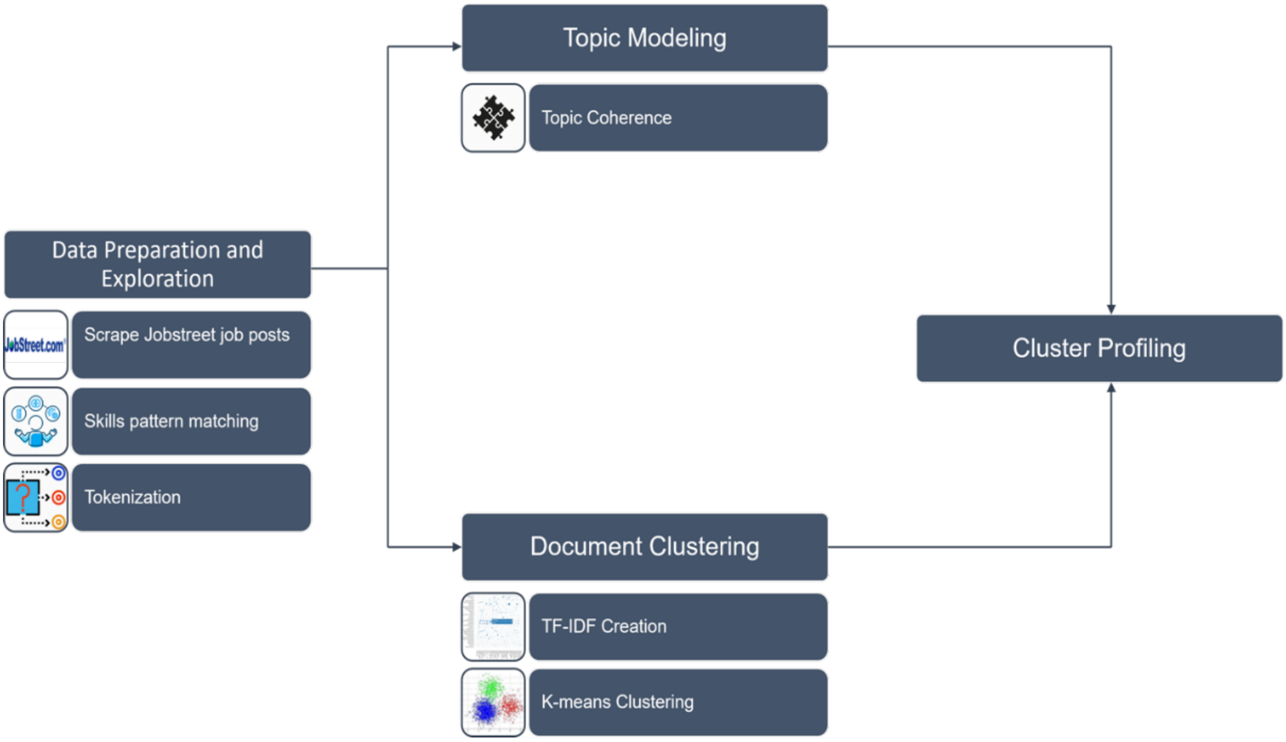
DSA-related activities were identified from the list of activities listed in the Philippine Standard Industry Code (PSIC) of the Philippine Statistics Authority (2009)[5]. The workers who performed these DSA- related activities were assumed to possess the DSA competencies. Employment data from the 2013-2016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

The demand for the four (4) DSA job roles and DSA competencies were determined by these tasks: 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

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

Figure 2. Skills Extraction and Job Segmentation Process



Source: Authors' illustration

Stage 1: Data Preparation and Exploration.

Scrape Jobstreet job posts. 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 3 months and gathered a total of 150, 869 unique job posts. For faster data recovery, 13.6% of these unique job posts (20, 574) were used as a representative sample where the algorithms were performed.

Skills pattern matching and tokenization. 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. In an effort 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 (see Appendix B). 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

Topic modelling, a dimensionality reduction technique which links words of original texts with the same context and discover themes that run through them, called topics, followed afterwards. Non-negative matrix factorization (NMF), a non-probabilistic topic modelling technique was employed in the analysis because it works better and faster when learning keywords or short texts (Cheng et. al, 2013, and 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). The extracted 30 skills groups were manually mapped into the 10 APEC-Recommended DSA competencies by select industry experts. 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%.

Stage 3: Document Clustering

To group the job posts into the four DSA job roles - data steward, data engineer, data scientist and functional analyst-, document clustering was performed. 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.

Stage 4: Cluster Profiling

Meetings among field experts were held to refine the 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.

7.2.2. Supply-side

The top ten undergraduate degree programs of current DSA practitioners were considered as the ten DSA-related undergraduate degree programs in this study (see Appendix C for the list and count). These DSA practitioners were AAP members or those who attended AAP's monthly meetups.

Academe experts who were either Program Administrators or faculty members of these ten DSA-related undergraduate programs were convened in Focus Group Discussions. The respondents were asked to evaluate two types of curricula of each DSA-related degree program: the actual curriculum of their university and the sample curriculum[6] found in the respective CHED Memorandum Order (CMO). The actual curriculum represents an innovative program while the sample curriculum represents the minimum requirements for the program. Since HEIs must comply with CHED's CMO, this *discussion paper* makes use of the assessments on the sample curriculum.

Each curriculum was evaluated on the extent it is able to equip the graduates with the basic proficiency of the ten 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

[6] The study made use of the curricula from the University of the Philippines Diliman (UPD) and the University of Asia and the Pacific (UA&P). 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.

4 - fully equips graduates with the analytics competency

Each curriculum was likewise evaluation 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.

8. Results and Discussion

8.1. Demand-side

8.1.1. Current availability of DSA competencies

Out of 1,271 activities listed in the Philippine Standard Industry Code (PSIC), twenty-two (22) activities (Table 4) were identified to require any of the 10 DSA competencies. Majority of these DSA related activities belongs to the information and communications subsector (41%), followed by the manufacturing (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[7] (Table 4). This figure was only 0.4 percent[8] of the total labor force in the Philippines in the said year (see Appendix D for employment share per major industry group). Almost half of these DSA competent workers (i.e., 42%)[9] belonged to the information technology and business process management (IT-BPM) industry[10] (Table 4).

[7] 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 Data Science and Analytics used in this study.
[8] Authors' computations given that there are about 41 million workers in the country in 2016 (Philippine Statistics Authority 2013-2017, Labor Force Survey as cited from the JobsFit 2022 Report of the Bureau of Local Employment, Department of Labor and Employment; PSA CISLE Report, 2018).
[9] 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 in 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 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 needed DSA employees mostly through respective networks and headhunters.
[10] The Annual Survey of Philippine Business and Industry report identifies 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 E for complete list of industry subclasses under BPM).

Table 4. Number of employees working on activities that require any of the data science and analytics competencies

Industry subsectors	Activities	2016[11]	
		Employees	%
Manufacturing		58,224	34%
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%
M70200	Management consultancy activities	27,396	15.77%
M71109	Other technical activities related to architectural and engineering	2,902	1.67%
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, n.e.c. [12]	s[13]	NA
M73200	Market research and public opinion polling	5,713	3.29%
M749	Other professional, scientific and technical activities, n.e.c.	nad[14]	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%

[11] See Appendix F for the employment sizes of these industry subclasses for 2013-2015
[12] n.e.c. = not elsewhere classified
[13] s = suppressed by the Publisher
[14] nad = no available data

Industry subsectors	Activities	2016[11]	
		Employees	%
N82294	Research and analysis activities	291	0.17%
N82299	Other back office operations activities, n.e.c	1,141	0.66%
TOTAL		**173,686**	100.00%

Source: Philippine Statistics Authority. (2016). *Annual Survey of Philippine Business and Industry (ASPBI)*

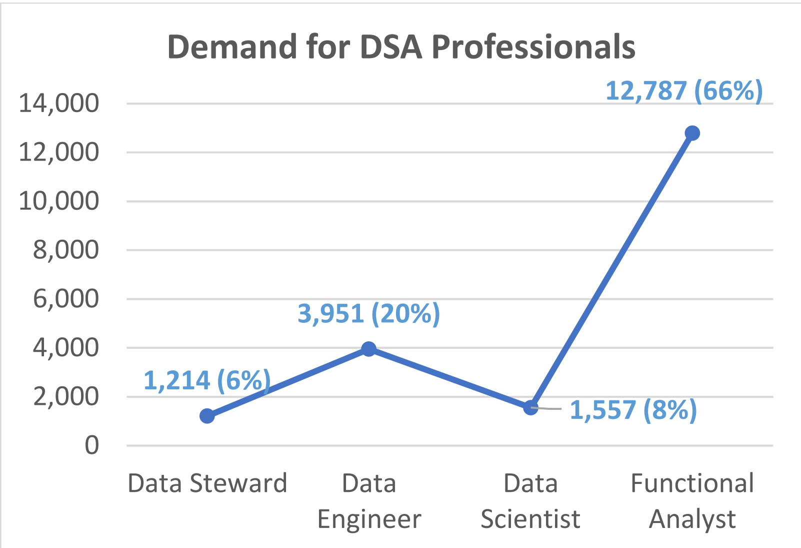
8.1.2. Current demand for DSA job roles

The online scraping of job postings in early 2019[15] revealed the skills that employers sought from the DSA professionals they wanted to hire.

DSA is at the very nascent stage in the Philippines. Its journey 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, this study revealed that the bulk (66%) of the demand for DSA professionals in early 2019 was for the *functional analyst* role

DSA is at the very nascent stage in the Philippines. Its journey 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 are able to articulate how data can deliver their operational targets. Consistently, this study revealed that the bulk (66%) of the demand for DSA professionals in early 2019 was for the functional analyst role (Figure 3).

Figure 3. Demand for DSA professionals



Data Steward	Data Engineer	Data Scientist	Functional Analyst	Total
1,214	3,951	1,557	12,787	19,509
6%	20%	8%	66%	100%

Source: Authors' Online Job Scraping

[15] May 1, 2019 to July 21, 2019 from Jobstreet.com.ph

Once the top management has decided to embrace DSA, the next important DSA job role is the *data engineer*, who ensures that data at its right form is 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, may be conflated with software engineer/developer role leading to a lower in demand for the *data engineer.*

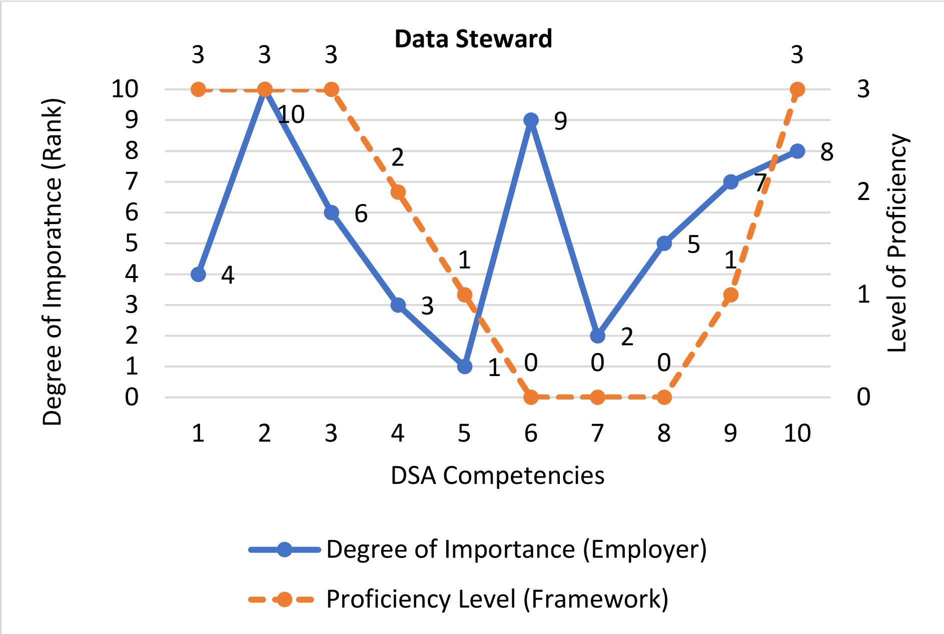
Only with clean data in data warehouses can algorithms be applied on data for insights. The low demand for *data scientist* (8%) showed that companies did not have their data ready for the DSA process.

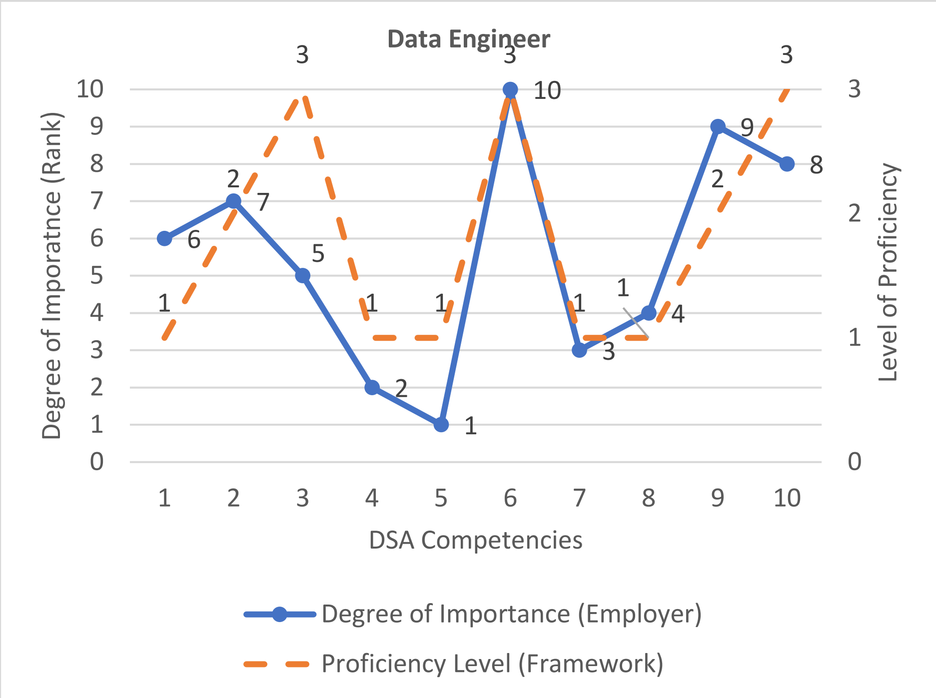
The *Facebook and Cambridge Analytica* scandal in 2018 revealed the dark side of analytics. It showed how the data of million Facebook users were used without their consent for political advertisement in the 2016 election in the United States and Brexit campaign in the United Kingdom (MacASkill 2018). 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.

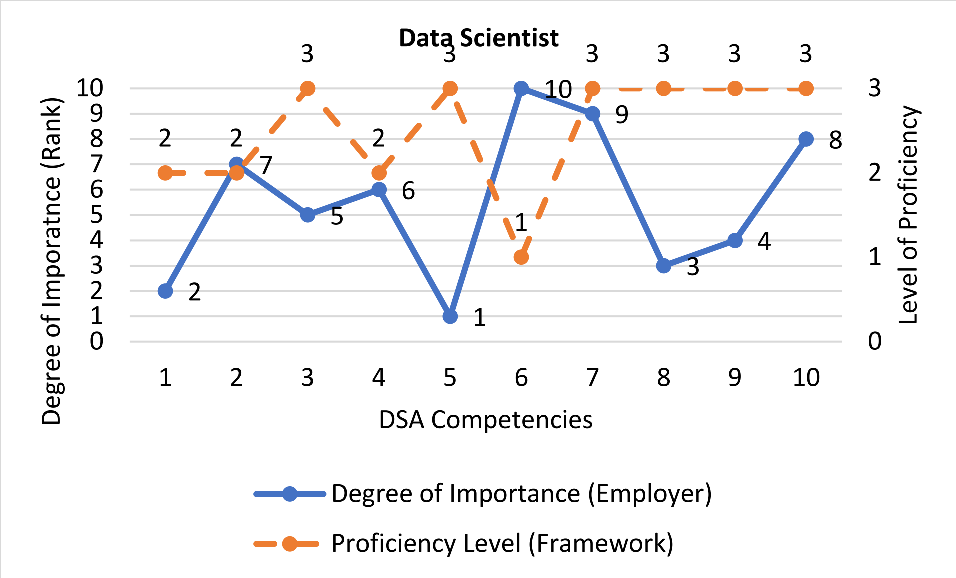
8.1.3. Work analysis of in-demand DSA job roles

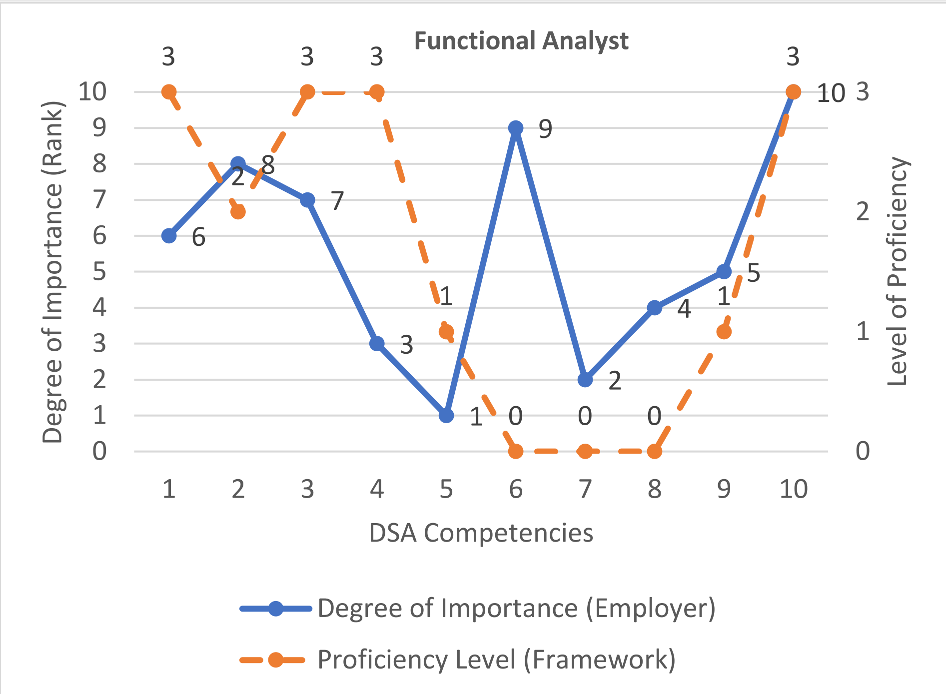
The work of each demanded DSA professionals was analyzed based on the competencies' importance and level of proficiency as perceived by employers and as required by the AAP Model respectively (Figure 2). While research methods did not come out as an important competency, the study considered it as incorporated with statistical techniques.

Figure 4. DSA Competencies' degree of importance and level of proficiency for each DSA Job Role









Legends:
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' Online Job Scraping

The AAP Model does not require 3 competencies for the work of a *data steward.* These are data engineering, statistical techniques, and methods and algorithms competencies. Employers, however, consider 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. AAP Model requires advance level of proficiency for data governance and 21st century skills (Figure 4).

AAP Model and employers agree that all the competencies are needed by a *Data Engineer.* Among these, employers rank data engineering (rank 10), computing (rank 9) and 21st Century Skills (rank 8) as the most important ones. The AAP Model requires an advance level of proficiency for data engineering and 21st Century Skills while only an intermediate level for computing. A close match is observed between the importance of the employers and proficiency levels of AAP Model on the competencies for the *data engineer* role (Figure 4).

Like the *data engineer* role, AAP Model and employers agree that all the competencies are needed by a d*ata scientist.* The three most needed competencies for the *data scientist role* according to employers are data engineering (rank 10), statistical techniques [which incorporates research methods] (rank 9), and 21st century skills (rank 8). 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 reflects the current work of *data scientists*, which includes data preparation, model building, and implementation of algorithms. Two other competencies that require an advance level of proficiency are operational analytics (rank 5) and methods and algorithms (rank 3) as per AAP Model. Employers, however, did not perceive these competencies as crucial to the work of a *data scientist* (Figure 4).

Like the *data steward role,* the AAP Model excludes some competencies for the *functional analyst* role. These are data engineering, statistical techniques, and methods and algorithms. For employers, all the competencies are needed, specially 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* are expected to perform a few tasks of the *data engineer* and the *data scientist*. The AAP Model prescribes that the *functional analyst* be an expert in the 21st Century Skills and with intermediate proficiency in data governance (Figure 4).

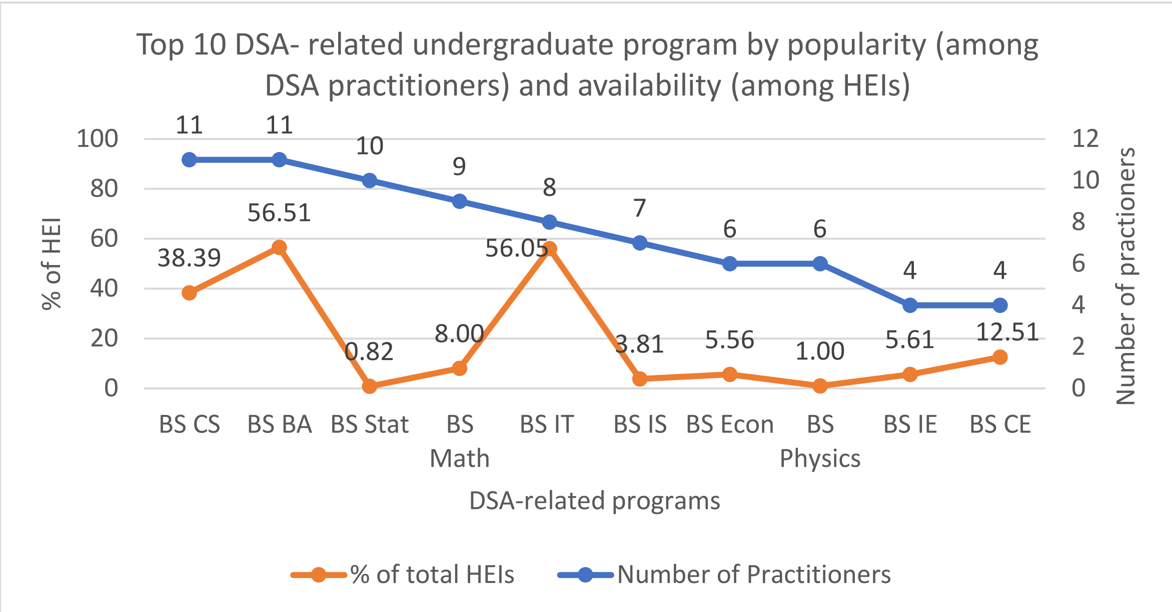
In summary, the AAP Model and employers agree that all DSA professionals must be very competent in the 21st Century Skills. However, they don't converge with *data engineering* as employers find it mandatory for all DSA professional while AAP Model requires it at the advanced level for *data engineers* but only at the entry level *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 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.

8. 2. Current Supply side

8.2.1. DSA-related undergraduate degree programs

The undergraduate degrees of current DSA practitioners were surveyed[16] (see Appendix C for specific breakdown). The top ten of these 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 (Figure 3).

Figure 5. Top ten data science and analytics-related undergraduate degree programs



Note: BS Math includes BS Applied Mathematics; BS Physics include BS Applied Physics
Source: Survey of DSA practitioners and Commission on Higher Education. (15 March 2019).

The undergraduate programs that were mostly offered by higher education institutions (HEIs) in 2016 were business administration (57%), information technology (56%), and computer science (38%) (CHED 2019). Except for BS Business Administration (57%), BS Information Technology (56%) and BS Computer Science (38%), only very few HEIs offer these DSA-related programs.

8.2.2. Current supply of DSA competencies

As mentioned earlier, this *discussion paper* makes use of the assessments on the sample curricula since actual curricula of HEIs must comply to the CHED's CMOs to deliver their respective degree programs. The discussion on the 21st Century Skills is provided in a separate sub-section.

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 5). Among these programs, the Computer Science program performed best by equipping its students with seven DSA competencies. Meanwhile, the Mathematics program equipped their respective

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

students with six competencies, the Physics and Industrial Engineering programs with four competencies, the Information Technology 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, the 21st century skills were the hardest to transmit to students. During focus group discussions, hiring managers highlighted their concern regarding job candidates not possessing the 21st century skills.

Table 5: DSA- related programs' ability to equip graduates with DSA competencies

DSA Competencies	DSA-related degree program									
	BS CS	BS BA	BS STAT	BS MATH	BS IT	BLIS	BS ECON	BS PHYSICS	BS IE	BS CE
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										

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 Commission on Higher Education. (15 March 2019).

21st Century Skills

A key feature among the APEC-Recommended Data Science and Analytics competencies is the 21st Century Skills. This competency is often contrasted with the other competencies as it pertains to a set of non-technical skills. Hiring managers indicated that such skills are 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[17]. It is then important to determine the extent HEIs are able to incorporate these 21st Century Skills in the curricula of the DSA-related programs. The *21st Century Skills* refer to a broad set of sixteen (16) "knowledge, skills, habits and character traits" (Glossary of Education Forum 2016) otherwise called "soft skills".

[17] Insights from the Focus Group Discussion with hiring managers for DSA job roles on February 2, 2019 at the University of Asia and the Pacific

Assessing the sample CMO curricula of the nine DSA-related programs[18] reveal that their graduates possess these eighteen soft skills in varying degrees (Table 6): problem solving skills (7 of 9 DSA-related programs); Job Collaboration and Critical Thinking (5 of 9 DSA-related programs); Organizational Awareness, Planning and Organizing skills and Decision-making (4 of 9 DSA-related programs); Communication and Storytelling and Job Flexibility (3 of 9 DSA-related programs); Ethical Mindset and Customer Focus (3 of 9 DSA-related programs); Business Fundamentals, Cross-cultural Awareness and Dynamic (self) Re-skilling (1 of 9 DSA-related programs); Social Awareness, Professional Networking and Entrepreneurship (0 of 9 DSA-related programs).

Overall, these DSA-related programs can be improved specifically on seven "soft skills": Entrepreneurship, Professional Networking, Social Awareness, Dynamic (self) Re-skilling, Cross-cultural Awareness, Business Fundamentals and Customer Focus.

Among the sample CMO curricula, the BS Industrial Engineering program best educates its graduates with 11 of 16 "soft skills"; BS Math and Applied Math with 9 of 16 "soft skills"; BS Computer Science with 8 of 16 "soft skills"; BS Physics and Applied Physics with 6 of 16 "soft skills"; BLIS with 5 of 16 "soft skills"; BS Statistics with 2 of 16 "soft skills"; BS Civil Engineering with 1 of 16 "soft skills"; BS Business Administration and BS Information Technology with 0 of 16 "soft skills" (see Table 6).

Overall, BS Business Administration, BS Information Technology and BS Statistics can be improved significantly by incorporating most of the 16 "soft skills" in their respective sample CMO curricula.

[18] Respondent opted not to assess the BS Economics' sample CMO curriculum on the 21st Century Skills.

Table 6. DSA- related programs' ability to equip graduates with 21st century skills

21st Century Skills	DSA-related programs (based on CMO's sample curriculum)								
	BS CS	BS BA	BS Stat	BS Math/ Applied Math	BS IT	BLIS	BS Physics/ Applied Physics	BS IE	BS CE
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
Decision Making	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) re-skilling	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

Source: Survey responses of participating Program Directors or HEI representatives

Additional insights can be drawn from the findings of the 2018 Philippine Talent Map Initiative (PTMI)[19] (see Table 7) which profiled the Philippine labor force on a set of "soft skills". The "soft skills" considered in the PTMI study and AAP Model (Table 6) are decision-making, planning and organizing and problem-solving skills. The PTMI found these "soft skills" to be poor among Filipino students and trainees i.e., low mean scores (PH overall, students overall, trainee overall) - decision-making (57, 59, 48); planning and organizing (57, 57, 48); creative problem-solving (54, 56,49). This study, likewise, finds the necessity to enhance the sample CMO curricula to better equip the graduates with these "soft skills" i.e., rating below "4" - decision-making (5 out of 9 programs); planning and organizing (5 out of 9 programs); creative problem-solving (2 out of 8 programs).

[19] A Department of Labor and Employment (DOLE) and SFI Group Joint Initiative: Profiling the 21st Century Skills of the Philippine Labor Force in 2018

Table 7. Mean Scores of the Philippines for 21st Century Skills

PHILIPPINES OVERALL MEAN SCORES		STUDENTS OVERALL MEAN SCORES		TRAINEE OVERALL MEAN SCORES	
ENGLISH FUNCTIONAL SKILL	72	ENGLISH FUNCTIONAL SKILL	74	ENGLISH FUNCTIONAL SKILL	62
MATH FUNCTIONAL SKILL	72	MATH FUNCTIONAL SKILL	73	WORKPLACE ETHICS	61
WORKPLACE ETHICS	72	WORKPLACE ETHICS	72	SOCIAL PERCEPTIVENESS	60
MULTI-TASKING	70	ENGLISH COMPREHENSION	71	MULTI-TASKING	60
ENGLISH COMPREHENSION	69	MULTI-TASKING	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
DECISION MAKING	57	INNOVATION	59	CREATIVE PROBLEM SOLVING	49
INNOVATION	57	DECISION MAKING	59	DECISION MAKING	48
PLANNING AND ORGANIZING	57	PLANNING AND ORGANIZING	57	PLANNING AND ORGANIZING	48
CREATIVE PROBLEM SOLVING	54	CREATIVE PROBLEM SOLVING	56	INNOVATION	45

Source: Department of Labor and Employment and SFI Group of Companies (DOLE and SFI Group). (2018). *Philippine Talent Map Initiative: Profiling the 21st Century Skills of the Philippine Labor Force.*

8.2.3. Current supply of DSA-job ready graduates

In 2019, the 10 DSA- related undergraduate programs graduated 176,597 professionals of 10 various disciplines (Table 8). 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 BS Industrial Engineering program best prepares students for DSA jobs. New industrial engineers may immediately work as data stewards, data engineers or functional analysts. Research methods, methods and algorithms and computing competencies need to be strengthened to also enable its graduates for the work of a data scientist (Table 5).

The next best DSA-related program is the BS Computer Science program for qualifying its graduates for the data engineer and data scientist roles. They need upskilling on the domain knowledge and operational analytics competencies to be ready to work as data stewards or functional analysts (Table 5).

Immediate data scientist employment may be offered to new mathematicians and physicists (Table 8). They are not deemed to perform readily as data stewards and functional analysts because they need 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 5).

The BS Information Technology program enables graduates to be data engineers (Table 8). 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 have been 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 8). To be prepared to work as data scientists, they just need a bit more skills upgrading specially in computing (Table 5).

Even though only 16 HEIs (0.82%) in the country offer BS Statistics, there are relatively a good number of data science practitioners who have this degree (Figure 5.) 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 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 science, business administration, and civil engineering must acquire the basic proficiency of all the DSA competencies through other means (Table 5 and 8).

While academic evaluators were asked to assess their respective program's curricula using the AAP's framework (and definitions), it is inevitable that they evaluated with their disciplines in mind. Therefore, their assessment of the DSA competencies would have been in the context of their field of expertise. Given that DSA is just being defined and this new field is different from theirs, it is likely that the evaluators had varying understanding of DSA. Thus, their assessment of their respective program's ability to prepare its graduates to take on DSA jobs maybe incomparable to each other.

Table 8. Supply of DSA Job Roles from the DSA-related programs

DSA Job roles	BS CS	BS BA	BS STAT	BS MATH	BS IT	BLIS	BS ECON	BS PHYSICS	BS IE	BS CE	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[20]	**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**	

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

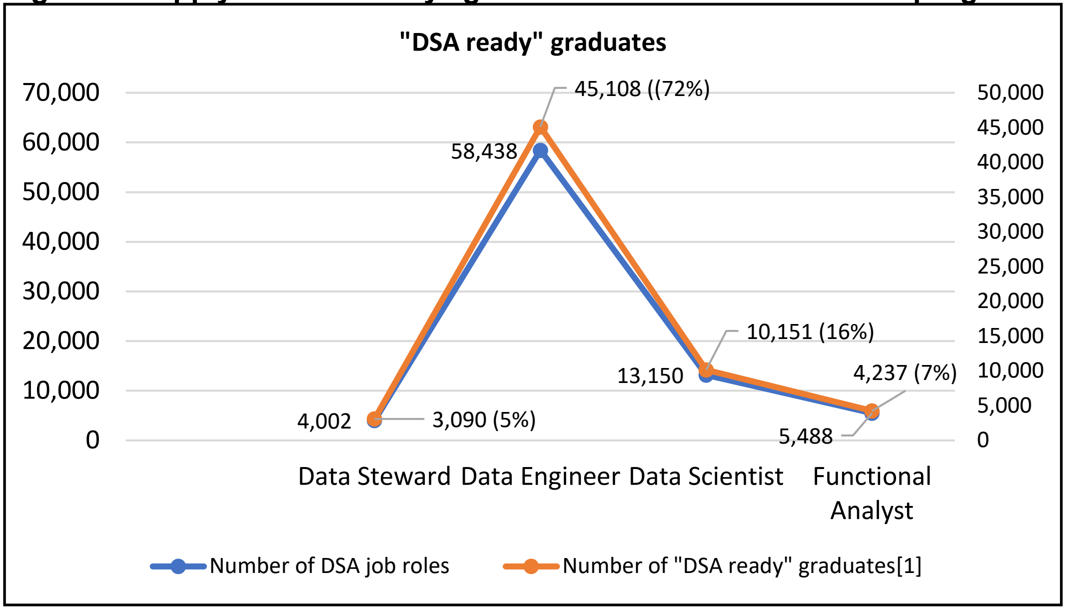
Source: Survey responses of participating Program Directors or HEI representatives; Commission on Higher Education. (15 March 2019); Authors' computations.

Graduates who have been 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 (Figure 4). 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%) are *data steward* roles of industrial engineering graduates; 58,438 (73%) are *data engineer* roles of Computer Science, information science and industrial engineering graduates; 13,150 (16%) are *data scientist* roles of Computer Science, mathematics and physics graduates and 5,488 (7%) are *functional analyst* roles of economics and industrial engineering graduates.

[20] Data used for number of graduates were based on Authors' computed projection of 2019 graduates. Meanwhile, basis of such computations were 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 for the number of graduates for these academic years.

Once a “DSA ready” graduate decides to get into DSA, he/she must choose only one of the DSA job roles he/she is 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% for *data steward*; 72% for *data engineers*; 16% for *data scientists*; 7% for *data stewards* (Figure 6).

Figure 6. Supply of “DSA ready” graduates from the DSA-related programs



8.2.4. Summary of Supply-side discussion

	Data Steward	Data Engineer	Data Scientist	Functional Analyst	Totals
Number of DSA job roles	4,002	58,438	13,150	5,488	81,078
Number of "DSA ready" graduates	3,090 (5%)	45,108 (72%)	10,151 (16%)	4,237 (7%)	62,586

Comparing the percentage distributions of demand and supply of DSA professionals (Table 9), 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. Majority of employers were looking for functional analysts (66%) while HEIs were mostly producing data engineers (72%).

Table 9. Demand and supply of DSA professionals

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

Source: From Figure 3 and Figure 6

It is suggested that future research be done to investigate the impact of the mismatch to business (e.g. opportunity loss). Hiring managers revealed that the average time-to-fill a DSA position is 46 days[21]. This delay in filling in a position certain means opportunity loss to the company. Another future investigation can be on the impact of this mismatch to the country's unemployment, specifically youth unemployment.

It is likewise recommended that an investigation be made on whether this misalignment points to an over/under supply of DSA professionals. Since the DSA-related programs were not explicitly designed to graduate 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 jobs 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 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/under supply 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).

9. 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 apt to be data engineers. Consistently, DSA-related degree programs equipped their graduates with the basic proficiency of the competencies that are 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 are 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 this misalignment comes the impetus for the government to work on appropriate mechanisms. If not, youth unemployment may exacerbate.

It must be noted, however, that the undergraduate programs evaluated in this study were not intended to produce DSA professionals. Consequently, it can be assumed that the graduates of these programs did not intend to be DSA professionals when they decided to take these programs.

Moreover, while the sample curriculum contains the minimum requirements of the degree program and serves as a guide for Higher Education Institutions (HEIs) in developing their actual curriculum, 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 and evaluation tools, etc.

[21] Information is based on the companies that responded to the survey questionnaire for hiring managers and participants of the Focus Group Discussion among hiring managers on
February 2, 2019 at the University of Asia and the Pacific.

10. Recommendations

Based on the foregoing, this study advances two main recommendations: a.) the use of AAP's framework to define the Data Science and Analytics profession, and b.) 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.

1.1. Use of AAP's Framework

Various strategies have been presented in relevant literature and recent conferences about the Fourth Industrial Revolution (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 United States 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 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, Communication and Technology (DICT) can also collaborate more efficiently with industry players through policies and programs fitted to the industry needs.

2. In terms of the supply, CHED can make use of the framework in creating new standards for Data Science and Analytics 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. For instance, BS Statistics programs educate Statisticians who are industry-ready to perform the tasks of a Data Scientist. This curriculum updating can be a potential solution to augment the current DSA skills shortage in the short run.

   HEIs can likewise use the framework to develop and offer undergraduate programs specially aimed to 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 (ex. Coursera, Moot, etc.) or training centers that offer DSA courses.

1.2. Other industry-academe-government linkages

A common definition of Data Science and Analytics 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 data science and analytics 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 co-developing teaching materials. Companies can share their use cases and data.

3. Industry players can encourage their own 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 that is currently being done is the Data Incubator Graduate Program of the Binghamton University. The program is an intensive 8-week fellowship that prepares master's students, PhDs, and postdocs in STEM and social science fields for industry careers as Data Scientists. But instead of fishing them out of the universities, companies can partner with the 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 Department of Trade and Industries (DTI) to facilitate the evolution of analytics as an industry, Department of Information and Communication Technology (DICT) to enable the needed information and communication technology infrastructure for DSA, Department of Labor and Employment (DOLE) to facilitate the healthy balance between the demand and supply of DSA workforce, Department of Science and Technology (DOST) to encourage the proliferation of research projects in DSA, Philippine Statistical Authority (PSA) to include DSA in its databases and studies and Commission on Higher Education (CHED) to promote DSA education among the Higher Education Institutions (HEIs).

REFERENCES

Aggarwal, C. and Zhai, C. (2012). *Mining Text Data*. New York, NY: Springer.

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

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)

APEC Human Resources Development Working Group. 2019 July 19. Close the Digital Skills Gap by 2025 through Collaboration. Singapore. Retrieved from https://www.apec.org/Press/News-Releases/2019/0719_Digital (accessed August 1, 2019)

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

Business Higher-Education Forum (BHEF). (2014). *Data Science Emerges as Essential Tool for Decision Making and Innovation Across Industry Sectors.* Business Higher-Education Forum.

Commission on Higher Education. Graduates by Specific Baccalaureate Programs_2013-2016 as of 15 March 2019. *Unpublished (data request)*

Chen, Y., Bordes, J., and Filliat, D. (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), Cergy-Pontoise, France, Sep 2016.

Cheng, X., Guo, J., Liu, S., Wang, Y., & Yan, X. (2013). *Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix*. SDM.

Elemia, C. (2016, September 17). Villanueva: We lack graduates who match jobs. Retrieved from Rappler: https://www.rappler.com/nation/146475-villanueva-we-lack-graduates-match-jobs (accessed August 12, 2018)

Dadios et al.(2018). Preparing the Philippines for the Fourth Industrial Revolution: A Scoping Study: A Scoping Study. *Discussion Paper Series No. 2018-11.* Retrieved from https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidsdps1811.pdf

Department of Labor and Employment and SFI Group of Companies (DOLE and SFI Group). (2018). *Philippine Talent Map Initiative: Profiling the 21st Century Skills of the Philippine Labor Force.*

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. Retrieved from https://www.kfa-juelich.de/SharedDocs/Downloads/UE/DE/Controlling/Workshop_WiDW/17-07-07_Demchenko.pdf?__blob=publicationFile

Francisco et. al. (2019). Mapping Philippine Workers at Risk of Automation in the Fourth Industrial Revolution. *Working Paper 2019-01.* Asian Institute for Management Rizalino S. Navarro Policy Center for Competitiveness. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3366809

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, n.d. A brief primer on republic act 10173. https://www.privacy.gov.ph/data-privacy-act-primer/ (accessed on December 16, 2019)

Manning, C. and Hinrich, S. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

McKinsey Global Institute. (2016). *People on the Move: Global Migration's Impact and Opportunity.* McKinsey&Company.

Microsoft. (2017 February 6). *Unlocking the Economic Impact of Digital Transformation in Asia Pacific: Microsoft Asia Digital Transformation Study 2018 - Southeast Asia and the Philippines.*

Philippine Statistics Authority. (2009). Philippine Standard for Industrial Classification (PSIC). Retrieved from https://psa.gov.ph/sites/default/files/PSA_PSIC_2009.pdf

Philippine Statistics Authority. (2012). Philippine Standard for Occupational Classification (PSOC). Retrieved from https://psa.gov.ph/sites/default/files/4publication.pdf

Philippine Statistics Authority. (2016). *Annual Survey of Philippine Business and Industry (ASPBI)*. Retrieved from https://psa.gov.ph/statistics/survey/business-and-industry/index

Philippine Statistics Authority. (2018). *Compilation of Industry Statistics on Labor and Employment*. Retrieved from https://psa.gov.ph/sites/default/files/2018%20CISLE%20Final.pdf

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 May). Skills Issues, Sources of Skills Issues, and Policy Responses in Five ASEAN- Member Countries: Indonesia, the Philippines, Thailand, and Vietnam. The HEAD Foundation

Röder, M., Both, A., and Hinneburg, A. (2015). *Exploring the Space of Topic Coherence Measures*. Shanghai, China: ACM.

Schutt, R. & O'Neil, C. 2013. Doing data science: straight talk from the frontline. O'Reilly Media Inc. United States of America

Tableau. (2016). *The State of Data Education in 2016: How U.S. higher education responds to the data skills gap.* Retrieved from https://www.tableau.com/data-education-2016

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

Tan & Tan. n.d. *Need For B.S. Data Science Degree Program in the Philippines.* Philippine Statistics Authority. Retrieved from https://psa.gov.ph/sites/default/files/Session%203-8%20Need%20for%20B.S.%20Data%20Science%20Degree%20Program.pdf

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

Wang, W. (2017, August 19). *Job Skills Extraction*. Retrieved from https://github.com/2dubs/Job-Skills-Extraction.

Weiss, S., Indurkhya, N., and Zang, T. (2015). *Fundamentals of Predictive Text Mining.* London: Springer-Verlag.

Xu, R. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, Vol. 16, No. 3, May 2005.

APPENDICES

Appendix A. Detailed Online Job Scraping Methodology

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

- ***Definition 2: Topic Modeling*** is a dimensionality reduction technique which links words of original texts with the same context and discover 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)]

Model Building Procedure

To determine the demand and industry-required competencies for the four (4) data science and analytics 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.

Online Job Post Scraping

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). 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
Domain Knowledge and Application	industry
Domain Knowledge and Application	business
Domain Knowledge and Application	practice
Domain Knowledge and Application	procedure
Data Management and Governance	governance
Data Management and Governance	data security
Data Management and Governance	information security
Data Management and Governance	policy
Data Management and Governance	privacy
Data Management and Governance	MDM
Operational Analytics	analysis
Operational Analytics	process
Operational Analytics	performance
Operational Analytics	operations
Operational Analytics	metrics
Operational Analytics	six sigma
Operational Analytics	business Intelligence
Operational Analytics	MIS
Data Visualization and Presentation	visualization
Data Visualization and Presentation	data storytelling
Data Visualization and Presentation	graph
Data Visualization and Presentation	charts
Data Visualization and Presentation	insights
Data Visualization and Presentation	report
Data Visualization and Presentation	dashboard
Research Methods	research
Research Methods	hypothesis
Research Methods	experiment
Data Engineering Principles	ETL
Data Engineering Principles	data warehouse
Data Engineering Principles	big data
Data Engineering Principles	data lake
Data Engineering Principles	architecture
Statistical Techniques	statistics
Statistical Techniques	quantitative
Statistical Techniques	qualitative
Statistical Techniques	regression
Statistical Techniques	ANOVA
Data Analytics Methods and Algorithms	algorithm
Data Analytics Methods and Algorithms	machine learning
Data Analytics Methods and Algorithms	statistical
Data Analytics Methods and Algorithms	mathematical
Data Analytics Methods and Algorithms	predictive
Data Analytics Methods and Algorithms	data mining
Data Analytics Methods and Algorithms	econometrics
Computing	SQL
Computing	programming
Computing	software
Computing	hardware
Computing	computing
Computing	IoT
Computing	cloud

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

Source: Authors' list based on APEC's definitions of the 10 Data Science and Analytics Competencies

Table A.3: Keywords for Other Categories

Category	Keyword
Tools	Python
Tools	SAS
Tools	SPSS
Tools	RapidMiner
Tools	Knime
Tools	R
Tools	Eviews
Tools	Minitab
Tools	Informatica
Tools	Oracle
Tools	Teradata
Tools	Tableau
Tools	Power BI
Tools	Stata
Tools	Excel
Tools	Access
Tools	Qlik
Tools	Matlab
Tools	VBA
Others	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 will mainly be 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 is a need for a supervised approach in extraction of features (e.g. skills, courses, responsibilities, tools) from the job descriptions. Most descriptions have a lot of noise, due to inclusion of equal opportunity statements, application instructions, and company benefits. In an effort 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 2 components [Weiss et. all (2015)]:

(1) *Term frequency (TF)* is defined as:

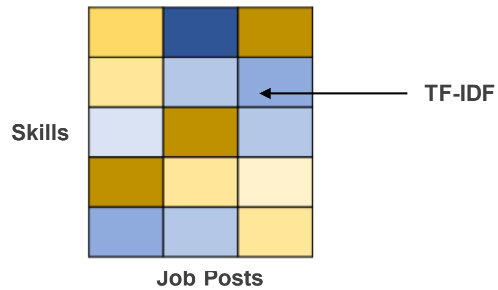
$$TF_{skill} = \frac{\text{No. of times the skill appears in a job post}}{\text{Total no. of skills in a job post}}$$

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

$$IDF_{skill} = \log \left(\frac{\text{Total no. of job posts}}{\text{No. of job posts with the skill in it}} \right)$$

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 were used in the succeeding analyses is given below:

Figure B. Skill-Job Post Matrix



Source: Results from Authors' online job scraping

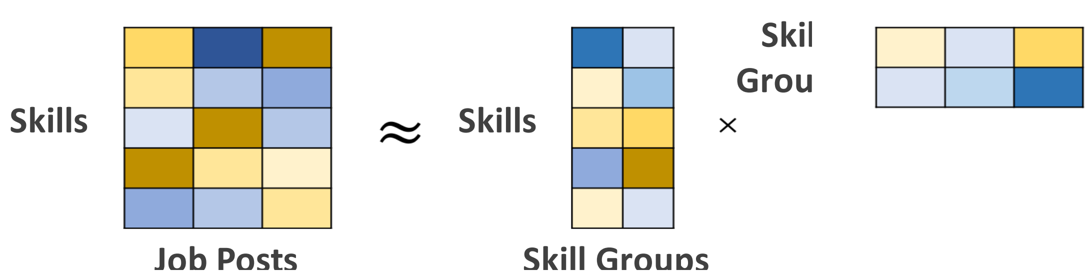
Topic Modeling

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 are explored.

There are two classifications of topic modeling techniques—probabilistic and non-probabilistic. The most popular method under each is latent Dirichlet allocation (LDA) and non-negative matrix factorization (NMF), 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 $\mathbf{X}_{(m\ x\ n)}$, into 2 matrices, $\hat{\mathbf{W}}_{(m\ x\ k)}$ and $\hat{\mathbf{H}}_{(k\ x\ n)}$, 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 the different job posts.

Figure C. NMF Algorithm



$$X \approx [\ \hat{W} \times \hat{H}\ ]_{+}$$

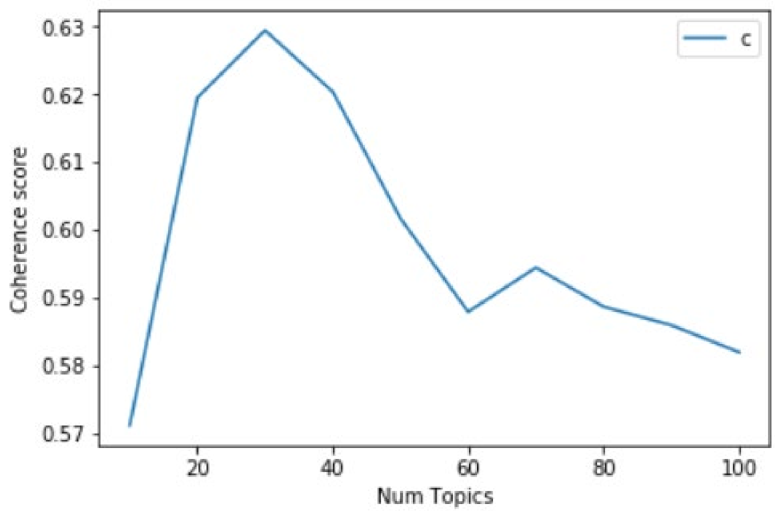
Source: Results from Authors' online job scraping

Since NMF will not automatically provide the optimal number of skills groups, the selection was based on topic coherence score. Coherence score measures whether pairs of top skills in a skills group tend to co-occur together relative to 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 C_v is the one most aligned with human interpretability. Hence, this was the score method used.

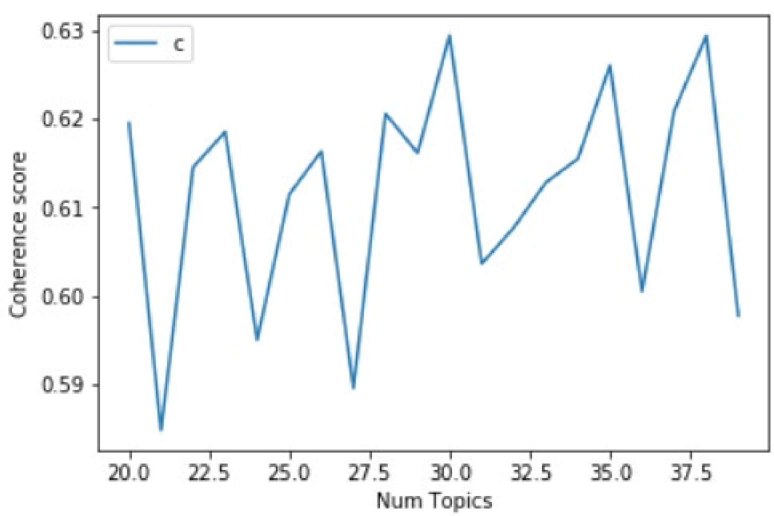
Results suggest that the topic coherence is highest when the number of skills groups chosen is 30.

Figure D.1: Topic Coherence Score for Number of Topics 10 to 100 by 10



Source: Results from Authors' online job scraping

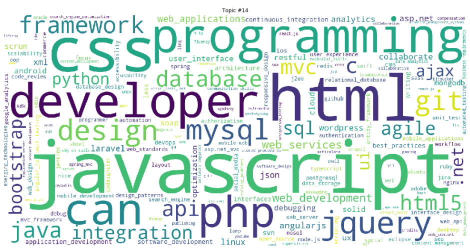
Figure D.2: Topic Coherence Score for Number of Topics 20 to 40 by 1



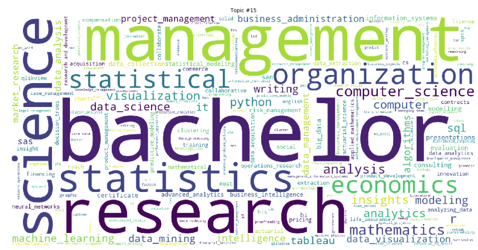
Source: Results from Authors' online job scraping

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 word clouds.

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 $\hat{\mathbf{H}}$) 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%.

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
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%
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%
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%
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%
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%

job_post_id	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	TOTAL
8782678	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%	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%	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	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	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 group are then mapped to the 10 APEC Data Science and Analytics (DSA) competencies.

Table C. Skills Group−Competency Mapping

DSA Competency	Skills Group No.
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 & Algorithms	15
Computing	1, 22, 24, 26
21st Century Skills	5, 9, 10, 18, 19, 28

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 are combined (refer to Figure G for illustration).

Figure G. Illustration of Derivation of Competency Weights

Source: Results from Authors' online job scraping

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%

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

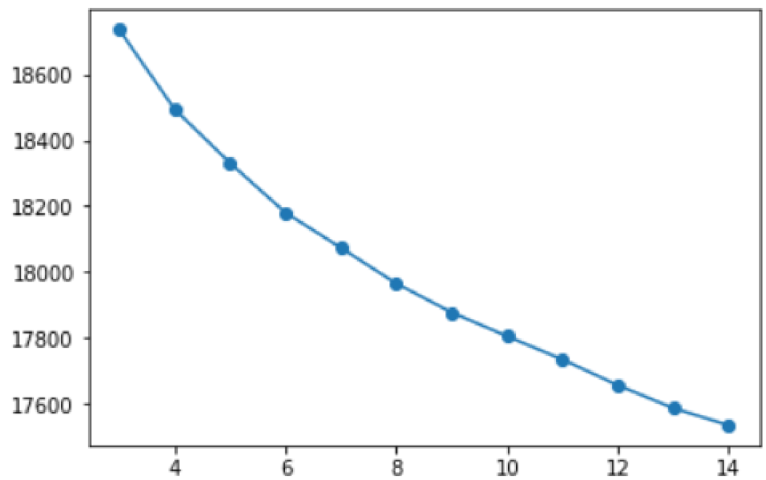
Document Clustering

Instead of grouping skills, the job posts were grouped in order to find among them the DSA roles—data steward, data engineer, data scientist, functional analyst, and analytics manager. 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 k-means method was the one used in the analysis.

As with NMF, k-means algorithm does not automatically suggest the best number of clusters. Hence, the number of groups from 2 to 15 was tested. Inertia, which is 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 were 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

Top Terms (Terms Closest to Cluster Centroids)	DATA SCIENTIST	LICENSED MANAGER / ANALYST	DATA STEWARD	DATA ENGINEER	AUDIT MANAGER / ANALYST	FINANCE ANALYST	SOFTWARE / APP DEVELOPER	INVENTORY / WAREHOUSE MANAGER	TECH SUPPORT SPECIALIST	ANALYTICS MANAGER
1	analysis	professional license	security	computer	audit	finance	design	inventory	customer service	management
2	report	license	infrastructure	information technology	report	accountancy	software	warehouse	communication	report
3	analytics	bar	management	science	compliance	report	javascript	management	communication skills	communication
4	research	pass	computer	computer science	internal audit	management	java	report	english	coordinate
5	statistical	board	troubleshoot	technology	management	analysis	computer	warehouse management	management	design
6	management	report	information security	telecommunication	accountancy	payroll	sql	coordinate	report	leadership
7	analytical	finance	technology	software	finance	budget	developer	logistics	interpersonal	compliance
8	statistics	management	software	database	internal control	payments	html	inventory management	computer	analysis
9	data analysis	budget	configuration	sql	cpa	cpa	css	inventory control	technical support	implementation
10	intelligence	accountancy	cloud	design	analysis	reconciliation	software development	sales	bpo	research

Source: Results from Authors' online job scraping

These clusters were further refined, following the below reclassification:

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

Source: Results from Authors' online job scraping

Figure J. Employers' DSA Role Profiles

Skills / Competencies	DATA SCIENTIST	DATA STEWARD	DATA ENGINEER	FUNCTIONAL ANALYST
Domain Knowledge	4%	3%	3%	11%
Data Governance	11%	28%	8%	14%
Operational Analytics	10%	6%	3%	13%
Data Visualization	11%	2%	1%	4%
Data Engineering	16%	27%	48%	15%
Statistical Techniques	16%	1%	1%	3%
Methods and Algorithm	8%	3%	2%	4%
Computing	9%	12%	21%	6%
21st Century Skills	15%	18%	10%	29%

Source: Results from Authors' online job scraping

Appendix B. List of pre-determined set of skills by Wang (2017)

vocab.txt

Appendix C. Distribution table of DSA employees' degree programs

	DSA Degree Program	COUNT	% Count
1.	Computer Science	11	10%
2.	Business Administration	11	10%
3.	Statistics	10	9%
4.	Applied Math/Math	9	8%
5.	Information Technology	8	7%
6.	Information Science	7	6%
7.	Economics	6	5%
8.	Physics/Applied Physics	6	5%
9.	Industrial Engineering	4	4%
10.	Civil Engineering	4	4%
11.	Electrical Engineering	3	3%
12.	Electronics and Communications Engineering	3	3%
13.	Accountancy	3	3%
14.	Geodetic Engineering	3	3%
15.	Communications-related	3	3%
16.	Chemistry	2	2%
17.	Humanities	2	2%
18.	Political Science	2	2%
19.	Psychology	2	2%
20.	Biology	1	1%
21.	Commerce	1	1%
22.	Computer Applications	1	1%
23.	Computer Engineering	1	1%
24.	Fine Arts	1	1%
25.	Legal Management	1	1%
26.	Mechanical Engineering	1	1%
27.	Public Health	1	1%
28.	Sociology	1	1%
29.	Tourism	1	1%
30.	Geography	1	1%
	TOTAL PARTICIPANTS	**110**	**100%**

Source: Survey Responses from Data Science and Analytics-related practitioners

Appendix D. Employment Share per Major Industry Group (2013-2017)

By Major Industry Group	Employment Share (in thousands)					Compound Annual Growth rate[22]
	2013	2014	2015	2016	2017	
Agriculture	11836	11801	11294	11064	10261	-3%
Industry	5936	6166	6275	7159	7371	4%
Services	20345	20682	21172	22775	22703	2%
Total	38, 117	38, 649	41, 228	40, 998	40, 335	

Source: Philippine Statistics Authority 2013-2017, Labor Force Survey as cited from *JobsFit 2022 Report by Bureau of Local Employment, Department of Labor and Employment; PSA CISLE Report, 2018*

Appendix E. Employment Size of Industry subclasses of BPM subsector

Section	Code	PSIC Industry sub-classes (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	12,041	10,766	14,609	11,476

[22] Computed by Authors based on standard Compound Annual Growth Rate (CAGR) formula

Section	Code	PSIC Industry sub-classes (non-BPM)	Employees			
			2013	2014	2015	2016
		facilities management activities				
	J62090	Other information technology and computer service activities	3,833	4,372	4,049	3,780
	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	On-line 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), n.e.c.	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

Section	Code	PSIC Industry sub-classes (non-BPM)	Employees			
			2013	2014	2015	2016
	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
	N82227	Legal services activities	s	s	s	s
	N82228	Supply chain management activities	s		s	s
	N82229	Other non-voice related activities, n.e.c.	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, n.e.c	nad	823	1,618	1,141
		TOTAL	**449,905**	**525,772**	**573,112**	**609,821**

Source: Philippine Statistics Authority 2013-2017, Labor Force Survey as cited from *JobsFit 2022 Report by Bureau of Local Employment, Department of Labor and Employment; PSA CISLE Report, 2018*

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

Industry (based on PSIC)	Code	Industry sub-classes	2014		2015		2016	
			Employees	%	Employees	%	Employees	%
C: Manufacturing	C26200	Manufacture of computers and peripheral equipment and accessories	60096	34.41%	73,462	36.80%	58,224	33.52%
K: Financial and Insurance Services	K66290	Other activities auxilary 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%
	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%

Industry (based on PSIC)	Code	Industry sub-classes	2014		2015		2016	
			Employees	%	Employees	%	Employees	%
	M72103	Research and experimental development in health sciences	nad	NA	497	0.25%	465	0.27%
	M72104	Research and experimental development in agricutural sciences	nad	NA	245	0.12%	257	0.15%
	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, n.e.c.	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, n.e.c.	3,400	1.95%	2,921	1.46%	nad	NA
J: Information and Communication	J58200	Software publishing	1,039	0.59%	2,281	1.14%	1,852	1.07%

Industry (based on PSIC)	Code	Industry sub-classes	2014		2015		2016	
			Employees	%	Employees	%	Employees	%
	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%
	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, n.e.c	823	0.47%	1,618	0.81%	1,141	0.66%
		TOTAL	**174,669**	100.00%	**199,648**	100.00%	**173,686**	100.00%

Source: Philippine Statistics Authority. (2013-2016). *Annual Survey of Philippine Business and Industry (ASPBI).* Retrieved from https://psa.gov.ph/statistics/survey/business-and-industry/index