Innovative skill Typology and Toolboxes of Data Experts

[1] Alon Hasgall, PhD., [2] Prof. Niv Ahituv, PhD

[1] The Israel Academic College Ramat Gan, Israel

[2] Tel Aviv University and Peres academic Center, Israel

Abstract: *Data Science (DS) is a rapidly growing profession and field of research, now producing some of the most important products and outputs for modern and innovative industries, businesses and for many research disciplines. DS covers multiple fields and issues (defined in the "data cycle")and they are required to deal with business disruptions in the market resulting from innovative products. The article lists out the innovative tools required by data analysts and data scientists foreach stage of the Data Cycle. It, then, divides the data related professions into three levels of expertise. Itfinally relates data experts' tools and skills, along the relevant stages of the data cycle, and the types ofusers*.

Keywords: *Data Cycle, Web Analytics, Data Analysis, Data Science*

1. Introduction: Data Professionals

The professional practice of data management is a rapidly growing field of research, which today enables the creation of innovative products, work processes and innovative methods for modern industries and businesses. This practice covers a number of areas and topics (defined in the "data cycle" model below) from high-level computing tools capable of sifting and combining scattered "big data" repositories into clear and useful information by identifying trends, patterns and insights that can then be used to address business disruptions in the market, as well as In the process of making intelligent and informed decisions, in the development of organizational innovation. Those professionals have been found valuable to managers, decision makers, researchers, publishers and policy makers in a wide variety of organizational, social and economic fields (McAfee & Brynjolfsson, 2012).

This rapid development of data professions is the result of many factors, including the widespread use of social media and the World Wide Web, the development of the ability to innovate, through big data, artificial intelligence, human/computer interfaces, visualization and related fields (Knowles-Cutler and Lewis, 2016). DS is of increasing importance for all sectors of society and government as the decision-making systems used by individuals, organizations and governments are constantly updated with massive amounts of data collected from many and varied sources and are required to cope with changing demands of citizens. The amount of data produced in the world exceeds 2.5 Exabyte per day, and the amount of information collected increases by about 20% per year (Harcourt, 2014).

Although it has been proven that the use of advanced data analysis tools and big data applications can help organizations improve their innovative business model development processes and increase profits, many organizations still do not make effective use of the information they have (Brock & Khan, 2017). A study commissioned by KPMG International Data and Analytics (Thomas et al., 2016) found that only 35% of executives surveyed use data analytics to improve their services, organizational processes and business models. This is due, at least in part, to a lack of data experts with knowledge relevant to the organization's needs. A longitudinal survey found that graduates with degrees in data analysis are in demand to fill a wide variety of roles and positions in every organizational field (Vossen, 2014).

Some preliminary literature has studied the data expert profession to characterize the roles of this growing field (Ahituv&Hasgall, 2017) and identified essential types of data experts. However, previous discussions have not distinguished between the various skills that each type of data expert is required to exercise. It is obvious, for example, that an expert in data analysis and visualization does not necessarily has to master deep learning and algorithmic, and vice versa (Anderson & Rainie, 2012). Our study is unique by distinguish the skills and tools needed by the various levels of data specialists.

The introductory section attempts to present "what constitutes Data Sciences?" by portraying the generic model of Data Cycle. The second section of the article briefly describes the methods employed in the study. The third section elaborates about the results. The last sections raise some conclusions.

1.1 The Data Cycle Model

The way data is handled should be very methodical and well structured. Otherwise, data gathered from various sources (e.g., Big Data) shall not be integrated to comprehensive processing and analysis. Apparently, the model of handling is quite similar for most of the decision-making processes and the research inquiries data (Davenport, 2017). Every decision-making process is based on a **data cycle**which culminates in a decision being made (Kearney, 2018). The cycle can be short and based on few data items, such as when we decide whether it is safe to cross the street. In such a simple case, we first identify the problem or the mission (crossing the street safely). We collect data (number and speed of cars passing by, width of the street), and estimate our walking speed. We integrate this data, operate an algorithm based on our past experiences, analyze the results, make a decision, then store and communicate feedback for future similar activities (Ahituv,2019).Obviously, most decisions made by organizational bodies and research teams are far more complicated. However, the stages of the Data Cycle (DC) model (Fig. 1) are essentially the same at any degree of complexity, and for every sector and field.

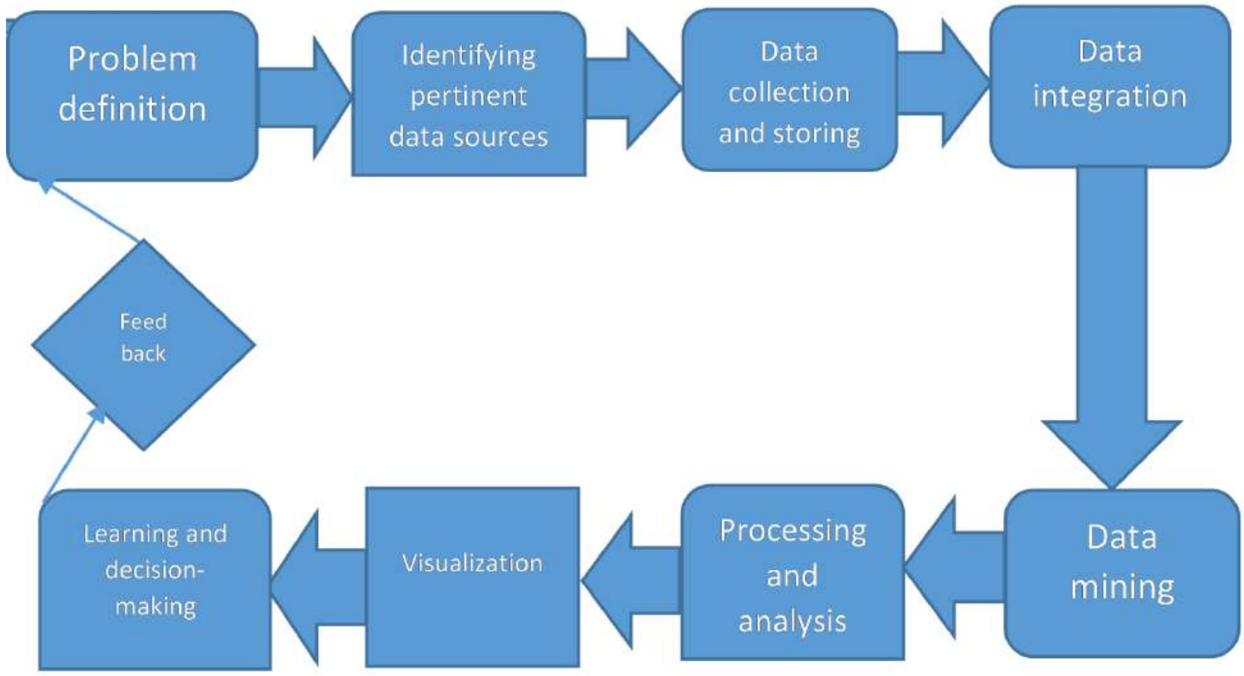

Figure 1: The data cycle.

The Data Cycle delineated in Figure 1 (Ahituv,2019)indicates that handling data from their origin and up to final exploitation for decision making should be a well-structured process. It commences with the definition of the decision problem (or the research questions), continues through the collection, integration, and analysis of the pertinent data, and culminates in visualization of the results and making the decision (Favi, et.al, 2018).). Each of the Data Cycle steps can be performed by a different expert or by the same professional. This implies that the list of skills should be associated with the various requirement along the Data Cycle. This is the purpose of the research presented here.

2. Methods

This meta-study presents analysis of information gathered by the authors and by other researchers. The proposed typology is based on a preliminary study of surveys pertaining to the skills and tools of various specializations of data experts, which had been published in academic journals and on professional websites (Ahituv & Hasgall, 2017). The study also presents a comparative analysis based on the findings of interviews with 20 data experts working in various specializations of the field. Some of the experts are employed by organizations that need heavy data analysis; some of them are independent consultants. The companies they are employed by do data mining, data analysis, algorithmic development and use, business intelligence, social network analysis, digital marketing, and the like. The survey was based on the same questionnaire presented to each participant. The results were analyzed qualitatively and quantitatively (as far as such a small sample enables).

The interviews were conducted over the course of six months. The authors of the article interviewed a range of participants, including company CEOs and analysts working in the government, military, and academia. The interviewees ranged between five and twenty years of experience.

To understand the needs of organizations, an analysis was conducted on employment advertisements for data analysts that were placed on major employment website sites in Israel (Alljobs, SQLink, John Bryce) and a survey conducted by KDNuggets (2018). The typology reflects classification according to four main factors: customer needs, types of tasks, characteristics of data, and required tools.

3. Results

3.1 Levels of Expertise in Data Analysis

According to the preliminary study (Ahituv&Hasgall, 2017) there are three main experts in the field of data. However, we have found that their skills and tools level them out.

Level 1: Web Analysts

Web analysts primarily use data collected from online social activity, such as databases of companies that maintain online social networks, e-commerce sites, and other websites. Analysis of this type of data enables identification of trends and development of processes to promote engagement between users and various communities, products, and services.

Previous research has found that personalize and tailor services can be based on customer behavior through the commercialization of data pertaining to individuals.

Level 2: Data Analysts– DA

During our interviews with experts, we found that Data analysts are unique in their abilities to understand the customer's needs, the business model, the organization's work processes, to organize and express the needs qualitatively, and after that, to guide the customer to focus on relevant data (Weiser, et al, 2022). According to the analysts, what distinguishes them from data scientists is their ability to draw from the data alone, to understand the background and surrounding influences, and to adapt data tables (made by the scientists) to the customer needs. For this work, the analyst needs to comprehend the data as well as the business or organization. Therefore, analysts must understand economics, marketing, management, and information systems. They must be proficient in thinking in an interdisciplinary manner and consider context (Evgeniou&Niessing, 2018).

Among the analysts we surveyed, 70% said that critical thinking skills, management, multidisciplinary and economic understanding, and quantitative thinking are needed, to understand the context and adapt data to the actual needs. The skills they need include some knowledge of mathematics, statistics, and programming, and sufficient understanding of algorithmic studies, models of complexity, segmentation, machine learning, and artificial intelligence. Usually, they need to locate DA tools rather than to develop them. Data analysts address complex questions and problems, develop models to analyze the existing situation, and perform predictive analytics. They often specialize in data analysis for various sectors of business and industry and are less focused on deep learning or DS research (Brock & Khan, 2017).Considering the foregoing, the skills required for data analysis may be referred to as "creating quantitative mental models" (as opposed to designing mathematical models, as discussed below). As such, a highly skilled data analyst DA must have the capabilities to understand the needs of the specific client thoroughly and deeply before making a work plan: familiarizing yourself with the organizational environment and market; formulating questions and sub-questions. At that time the DA have to know how to carry out a professional work process: identify a relevant database, organize it, select data from it in accordance with the needs of the question to be researched; create and present clear PivotTables and graphs that confirm a reality of the market (this differs from academic research questions, which are designed to uncover new knowledge), then Design a detailed and professional methodology of the work process; clearly articulate the relevance of the data to be analyzed and what can be proven by it. It seems that his main expertise lies on the ability to investigate quantitative evidence that is validated and integrated with qualitative information which taken together describe the existing situation, then optimize the data in order to find an answer or response to the analytic question(s) and then present and visualize insights in a manner that will have an immediate and convincing influence on the customer/decision-maker

Level 3: Data Scientists– DS

DS is the highest-level specialization that involves the design of complex models and development of the necessary original algorithms for modeling, computer learning, and machine learning that form the basis for artificial intelligence (Kaggle, 2017). Only those who studied computer science, high-level statistics, and DS are qualified to work at this level. Data scientists must be able to handle a large amount of chaotic, unsynchronized, and apparently incomprehensible data and to create data matrices (clustering) from them. They develop and implement advanced statistical methods to create a new structure for the data that show the changing situation and can be used in forecasting and predictive analytics.

3.2 Data Expert Types

Table 1 summarizes the type of data experts and the skills required by each, thus illustrating the differences between them.

Table 1: *Characterizations of the Three Levels of Data Experts*

Data Expert (level)	Characteristics of the Data	Type of Tasks	Work Processes	Customer Needs
Web analysts	Information on websites and the internet	Marketing Tailoring products and services to customers Customer service	Extracting information on social networks Customer profiles History and characteristics of online activity	Marketing managers E-commerce managers Customer service managers
Data analysts	Medium-sized data mass Economic, business and budgetary data (sales, income, expenses, etc.) Data from online social activity	Business models Databases on competitive business practices and social trends Realizing the potential for immediate application of micro-data Application of algorithmic or mechanical processes to existing knowledge	Statistical analysis on existing data related to a business strategy Design and maintenance of data systems and databases Demonstrating patterns in data Predictive analytics Preparing reports Visualization of results	Senior executives Department heads Consulting firms Startups and developing businesses Production managers
Data scientists	Large mass of unstructured, unsynchronized data Digital signals	Level of strategizing New trends Making predictions for business and social issues	Management of large-scale structured and unstructured data Production of databases Modeling databases using algorithms and scientific observations Creating new models AI-based analysis methods, deep learning machine learning Developing search engines and content-based recommendation systems	Managers Senior executives Research directors Consulting and forecasting firms Academic and research institutes

The findings presented in Table 1 indicate that the data experts clearly distinguish among various skills related to the Data Cycle. One level of skills relates to the depth of analysis of the data; another level of skills relates to the collection, cleansing and integration of data; and a third level centers of the insights that can be derived from the data with relation to business, organizational and research issues. There is also an attempt to get insights about type of searches in search engines and type of activities popular in social networks.

In sum, the first level: web analysts, focus on online Information and social activity. In contrast the DA focus on the emerging needs of businesses in various sectors and content worlds. The third level – DS, use intelligent algorithms and machine learning to deal with a large mass of data.

3.3 Core Skills and Tools of Types of Data Experts

This section shows the connections between the demands and needs of organizations, the challenges facing the various data experts, and the tools each type uses.As displayed in Table 2, the skills required of each expert canbe divided into three main levels which employ specific tools (Pergl, et al, 2019).. During the study, the experts were asked about the tools needed to fulfill their tasks. According to their opinion, the classification of the tools corresponds to the specialization level required of the expert. The basic level characterizes the tools of web analysis while the advanced level is needed by those who specialize in data science. The intermediate level is required of the data analysts.

Table 2: *Data Experts' Tools and Skills, Stages in the Data Cycle, and Types of Clients*

Category		Subjects of Study	Skill Level			Remarks
			Beginner (1)	Intermediate(2)	Advanced (3)	
Background	Goals and business values	Information about the organization	√			
		Data Analytics and introduction to BI	√			Distinction between the operating system and the administrative / analytical system / data warehouse; also data modeling
Data Cycle	Understanding the business environment and definingthe problem	Tools and methods for understanding the business environment	√			
		Tools for defining and formulating a focused problem	√			
	Collecting relevant data	Relational database	√			
		Non-relational database and information objects in the cloud			√	NOSQL database types, interactions with cloud information objects (EMR and more); this differs from retrieving information from relational databases
	Cleaning, validating and inspecting the	Basic data inspection	√			Importance of the issue of information quality and types of quality issues
		Inspecting		√		Outliers; handling missing

	data	data and detecting biases in the data				information
	Data integration, processing and analysis	Descriptive statistics	√			
		Advanced Excel	√			
		Basic SQL	√			
		Advanced SQL		√		New generation for integrating BI data including procedures; use of analytics tools
		Advanced Python			√	Data processing libraries
		Dealing with the world of machine learning (ML)		√		Familiarity with key methods and terms in the world of ML, use ofAutomated Machine Learning tools
		Developing ML models based on DL, NLP, Python/R			√	This is the most significant addition in the advanced rank (3) in relation to the intermediate rank(2)
	Extraction and visualization of insights	Data visualization and Data storytelling	√			
		Reporting and analysis tools	√			Tableau, SAP BO, Power BI, Qliksense; General presentation of the market and familiarity of different tools and their progress
		Integrating information products in the business processes		√		
Supplementary Information	Additional information	Introduction to Data Science and big data technologies	√			Technologies for distributed storage and parallel processing(Hadoop, Spark)
		Informatics and data regulation	√			

Table 2 relates to the various tools that are required to each data experts when facing a certain task in any of the Data Cycle steps. It appears that every steps and every activity related to each steps needs its own tools. The "division" of tools is more associated with the step of the Data Cycle than with the area of the business, the research study or the organization conducting the data analysis.

Skills and Tools of Web Analysts

The main tasks of web analysts are: identifying trends, usage patterns, mapping ongoing processes (patterns of online interaction or shopping), analyzing, categorizing, and ordering quantitative data and continuous variables. They must understand how to utilize statistical indices (means, standard deviations, medians, percentages), data distributions and histograms, correlations, and causalities. They use website management tools such as change monitors, aggregators from social media, content websites, and news sites.

Skills and Tools of Data Analysts

Thefocused on actual questions need to be solved. They must be proficient in retrieving data using search engines and targeted search tools for locating online information, databases, profiles, companies, images, and videos. They use organizational diagnostic tools for strategic diagnostics and business models. They must have strong ability for managing databases using Excel, and experience with Structured Query Language (SQL), Python / R, and JavaScript.

For dealing with business issues, data analysts should be able to operate Business Intelligence (BI) systems. They use open-source artificial intelligence algorithms to sort and filter data and present it in the appropriate context. Data analysts present their integrated insights to clients using display systems, infographics, symbols, charts, and BI visualization tools such as Tableau, Power BI, Qlik and microservices tools and technologies. Each BI tool has its own strengths, though there are high-quality, freely available tools such as Power BI, which can create end products that are equivalent in quality to those produced by Tableau. BI tools are useful in evaluating the quality and features of data, including identifying trends and generating insights. These products are designed for technical and non-technical users, and include "drag-and-drop" interfaces to enable users to easily and seamlessly create accessible visualizations.

At each stage of the DC, data analysts must consider the needs of the project's clients. For example, data visualization gives clients a view of the data that would otherwise be impossible. Therefore, effective data analysts know how to appropriately leverage technologies such as dashboards and other platforms for visually integrating insights from the data. In comparison to traditional flat files, dashboards make data more tangible and easily understandable for decision-makers and other end users. Another important aspect of data analysts' work involves merging and integrating data from various sources based on shared keys, group aggregations, and statistical calculations. Data analysts must be able to program, write, and use various scripts. They manage databases, address challenges in storing and processing multiple databases, including classical relational databases, and conduct analysis of qualitative data.

Skills and Tools of Data Scientists

Data scientists are expected to have a deep understanding of the field in which they are working so they can quickly understand the data and the relevant issues. For example, a data scientist who works for a retailer would know that December is their busiest month, and thus will be able to foresee logistical problems or irregular sales behavior. Data scientists must communicate with team members and stakeholders. Data scientists need strong technical knowledge, problem-solving skills, and be proficient in using code languages. By understanding the mathematics (especially statistics) and the algorithmic on which machine learning models are based, data scientists are able to make adjustments and improve model performance.

Prospective employers want candidates who are skilled in data wrangling and data preparation, since the datasets they work with are often unruly or chaotic and require pre-processing. Skills in machine learning and modeling are core skills of data scientists, enabling them to offer high-quality solutions to organizations.

We have also examined the variety of challenges and skills required by the various types of experts involved along the Data Cycle. Apparently, there are significant differences among the various skills and expertise that should be mastered by each professional in each stage.

4. Conclusions

In order to make informed decisions in today's world, business organizations, public and governmental agencies, and individuals all rely on data experts who are able to understand and make sense of vast amounts of data. Professional data experts have a wide range of skills, tools, and resources that are used in each stage of the data cycle, including data collection, cleansing, integration, and analysis, which enable them to offer recommendations. They also assist in long-term storage of data and management of organizational knowledge. The data analyses are used to draw conclusions and make predictions. Data experts (analysts and scientists) who have abilities in collecting, sorting, storing, and managing data, and creating useful knowledge from them, are key in every organization in the business, public, or academic sectors. This investigation of the skills, methods, and tools needed for each Data professional level, led the authors to conclude that through their participation in daily processes and business activities, Data analysts have the most appropriate opportunity and skill to connect the results of their analyzes to managers, without having to use complex mathematical and algorithmic methods. This makes them the only professionals capable of producing data-driven organizations.

References

[1]Ahituv, N.(2019). What should be taught in an academic program of data sciences? Proceedings of the Digital Presentation and Preservation of Cultural and Scientific Heritage, vol. 9. Sofia, Bulgaria: Institute of Mathematics and Informatics.

[2] Ahituv, N., & Hasgall, A. (2017). ADAPTING UNDERGRADUATE PROGRAMS OF DATA SCIENCE TO THE PROFESSIONAL REQUIREMENTS OF THE INDUSTRY. *Исследования по геоинформатике: труды Геофизического центра РАН*, *5*(1), 157-157.

[3] Anderson, J. Q., & Rainie, L. (2012). Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well. *The future of the internet*.

[4] Evgeniou, T., & Niessing, J. (2018). Data Analytics: A Marketing Segmentation Case Study. INSEAD

https://inseaddataanalytics.github.io/INSEADAnalytics/BoatsSegmentationCaseSlides.pdf

[5] KDNuggets. (2018). The 13th annual KDNuggets software poll.

https://www.kdnuggets.com/polls/2012/analytics-data-mining-big-data-software.html

[6] Kearney, A. T. (2018). The data value chain. GSM

https://www.gsma.com/publicpolicy/wpcontent/uploads/2018/06/GSMA_Data_Value_Chain_June_2018.pdf

[7] Knowles-Cutler, A., & Lewis, H. (2016). Talent for survival | Essential skills for humans working in the machine age. Deloitte LLP.

[8] Kaggle. (2017). Machine learning & data science survey. Kaggle.

https://www.kaggle.com/kaggle/kaggle-survey-2017

[9] Brock, V., & Khan, H. U. (2017). Big data analytics: does organizational factor matters impact technology acceptance?. *Journal of Big Data*, *4*, 1-28.

https://doi.org/10.1186/s40537-017-0081-8.

[10] Harcourt, B. E. (2014). Governing, exchanging, securing: Big Data and the production of digital knowledge. *Columbia Public Law Research Paper*, (14-390).

[11] Davenport, T. H. (2017). How analytics has changed in the last 10 years (and how it's stayed the same). *Harvard Business Review*, *28*(08), 2017.

[12] Favi, C., Campi, F., Germani, M., & Manieri, S. (2018). Using design information to create a data framework and tool for life cycle analysis of complex maritime vessels. *Journal of Cleaner Production*, *192*, 887-905.

[13] Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., & Slifka, J. (2019). "Data Stewardship Wizard": A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal*, *18*(1).

https://datascience.codata.org/article/10.5334/dsj-2019-059/

[14] McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review 90, 10, 60-68.

[15] Thomas E., Murray R., Brad F., & Anthony C. (2016). Building trust in analytics: Breaking the cycle of mistrust in D&A. KPMG.

[16] Vossen, G. (2014). Big data as the new enabler in business and other intelligence. *Vietnam Journal of Computer Science*, *1*(1), 3-14.

[17] Weiser, O., Kalman, Y. M., Kent, C., & Ravid, G. (2022). 65 competencies: which ones should your data analytics experts have?. *Communications of the ACM*, *65*(3), 58-66..