ASSESSING THE ENTRY COMPETENCES OF DATA SCIENCE MASTER PROGRAM POTENTIAL STUDENTS

K. Rasheva-Yordanova[1], S. Toleva-Stoimenova[1], I. Kostadinova[1], D. Christozov[2], B. Nikolova[1]

[1]University of Library Studies and Information Technologies (BULGARIA)
[2]American University in Bulgaria (BULGARIA)

Abstract

Recent evolution of information technologies (IT), computing power and Internet have resulted in a tremendous growth in the amount of data generated by humans and machines. Big Data paradigm has revolutionized the entire IT landscape and changed all and every facet of human culture and behavior. Nowadays we have many new technologies, methods and tools to manage and process efficiently the large data volumes available. As a result of the information age, Data Science has emerged as a new inter-multi and even transdisciplinary area of knowledge. The Data Scientist is "a unicorn" facing the complexity of Big Data challenges, responsible for integrated diverse aspects of data management from data capturing to uncovering patterns and insights hidden in huge volumes of heterogeneous data to drive decision-making.

A Data Scientist is a professional with analytic skills, abilities in bringing disparate areas together in a creative way, asking the right questions driven by solving problems, curious, creative and IT savvy. In our previous studies, we identified analytical thinking and analytical skills as the core competences of data analysts and researchers. This study addresses Data Science and the Data Scientist's skills from the point of view of designing a curriculum for developing competences for successful career in the area of Data Science.

We are sharing the results of an empirical study conducted to assess the entry competences of students, potential candidates for a Data Science MA Program. The survey emphasizes abilities rather than knowledge, stressing on analytical, critical thinking of current students in BA programs.

The constructed questionnaire is based on known tests to assess analytical thinking, critical thinking, problem solving attitude, etc. adapted to the objectives of designing a Data Science curriculum. It represents logical problems in three different formats: math questions, text assignment and figures pattern recognition. The questionnaire also includes self-assessment of analytical thinking skills and dispositions.

The survey was done in the Fall 2018. About 250 BSc students in their senior year were approached by an on-line or paper questionnaire. Students were originally classified in two categories according to their majors as "technical" or "non-technical". Statistical data analysis was conducted by using SPSS. Our analysis reveals the level of analytical, creative and critical thinking of the respondents. The assumptions that both categories of potential students have similar potential for success in studying at the Data Science Master program and to become successful Data Scientists was justified.

Keywords: Data Science, Data Scientist, Data Science skills, analytical skills.

1 INTRODUCTION

The constant generation of data and its accumulation in different media requires competent specialists able to explore them. The volume of captured and recorded data have led to major challenges induced by the need for meaningful data processing and acquiring the knowledge they may bring [5]. Building up skills for working with Big Data can respond to those challenges and to help making more evidence-based solutions and to provide opportunity to convert complex, often unstructured data, into applicable information. This is a prerequisite for the business to seek experienced Data Scientists who have the ability not only to manipulate big amounts of data with advanced statistical and visualization techniques but also to have a solid insight from which they can derive prospects. These insights help to predict potential outcomes and mitigate potential threats to the business.

In our previous studies we investigate the range of expected by the industry competencies of Data Scientist's [1]. We have shown that Data Scientist's analytical competences represent the cross-point

of all other hard (technical) and soft (non-technical e.g. communication, collaboration, curiosity etc.) skills, especially in the Big Data context. Analytical thinking, in the course of our study, was considered as a core competency [9]. Undoubtedly, analytical thinking is a mandatory element in determining relations of priority, hierarchy, causes, effects exposed by data. Variety of tools for analyzing data, redefining the entry assertions, and presenting conclusions [2, 6, 8] are applicable. The ability of thinking analytically, closely related to logical and critical thinking is the necessary requisite in activities such as decision making, issue analysis, inquiry, problem solving and self-assessment [9].

Critical thinking is an act, based on knowledge and by exploring available evidences and information, to arrive at a reasonable conclusion. Analyticity, in turn, allows the components of complex situations, practices, problems, allegations, ideas, theories, arguments to be identified, contrasted and evaluated [9]. This raises the role of human skills in data processing. This idea is particularly important in the context of Big Data, where skills are not only needed in exploitation of technology, such as software and infrastructure, but mainly in generation of insights leading to organizational decisions [7, 10].

The breadth of knowledge, required by a data specialist, has prompted discussions on how academic programs should be adjusted to meet the growing demand of skilled and experienced professionals [3]. Along with hard and soft skills, a data specialist should possess skills in analytical, critical, and logical thinking [4]. Understanding the importance of analytical skills and their application when working with data in problem-solving is the starting point of our research.

In this paper, we are sharing the results of an empirical study conducted to assess the entry competences of students, potential candidates for a Data Science Master Program. The main research questions can be summarized as follows:

- RQ 1: What is the extent of acquisition and application of analytical thinking among undergraduate students in bachelor degree programs, contrasting IT and not IT degrees?
- RQ 2: Does the professional area of a person affects the extent to which analyticity, critical thinking and logic are applied?
- RQ 3: Which analytical skills are missing by potential candidates for Data Science Master program?

The entry hypotheses of the study are as follows:

- *H1: Students studying computer science or IT related majors have a higher degree of knowledge and application of analytical skills than students who are trained in not technical or natural science programs.*
- *H2: People with a higher education degree, have a higher level of knowledge and application of analytical, critical and logical skills.*

This paper is structured into four parts. The first section presents an overview of the study. Second one details the methodology used in the study. The third section presents the results obtained. Finally, the fourth section, summarizes and discusses the findings and includes concluding remarks.

2 METHODOLOGY

2.1 Questionnaire design

The survey was conducted through a questionnaire. It consists of three groups of questions:

a) questions about the profile of the respondents and their association with a particular professional field;

b) self-assessment questions, which measure how students themselves rate their analytical skills and dispositions (Likert scale);

c) logical problems and tasks in three different formats: math questions, text assignment and figures pattern recognition.

Development of individual questions were based on literature review and published questionnaires [9]. Some were adopted or adapted from existing tests; others were developed especially for this questionnaire. We analyzed items such as difficulty, reliability, etc. and adjusted them to the target respondents and their likely level of analytical skills. Questions were written so that respondents understand them correctly, to ensure students' active involvement, motivation, and precise work.

The essential part required student to answer 20 multiple choice questions indicated four categories of competences (five questions per category): [1]

1 Identifying analytical skills - checking ability to identify similarity, difference, progression in a sequence, contradiction (AS);

2 Identifying abstract thinking – answering by following definition regarding the objects instead of intuitive well-known attributes (AT);

3 Distinguishing between facts and opinions (FO);

4 Quantitative reasoning – ability to compare quantities and to apply basic mathematical facts as, for example, relationship between the radius and the circle length (QR).

The questions were randomly mixed among different categories.

2.2 Respondents

The majority of the respondents are undergraduate students who are potential candidates for the Data Science master program. They are studying in different universities in the Republic of Bulgaria offering IT-related bachelor degree programs. Students from majors as Computer Science, Information Technologies, and Information Brokerage, who have passed training emphasizing different aspects of IT, were the primary target, but also students from other non-technical areas as Library Studies and Public Communications were approached.

Most of the students surveyed are specializing in Informatics (37%), while the rest are distributed in the specialties "Information Technologies" (24%), "Computer Science" (11%), "Information Security" (3%) "IT in the Judicial Administration" (3%), "Information Brokerage" (2%) and "Communication and Informing" (2%). The largest share is formed by the students, trained in the professional field "Informatics and Computer Sciences". To a lesser extent, respondents are trained in the Professional field in "Communications and Information Sciences" (3%) and "National Security" (1%).

Respondents were divided into two groups according to their majors as possessing “technical” and “non-technical” background. Our target group is consisted of students who are training in bachelor degree programs within the professional field "Informatics and Computer Sciences" (or technical specialties). All respondents training in a non-technical professional field was joined the second (control) group (Table 1).

Table 1. Distribution of the respondents

Competence Category	Total	Percent
Technical	160	78.8
Non-technical	43	21.2
Total	203	100

82.8% of respondents are currently studying in a bachelor degree program. The remaining 17.2% are ranging from possessing bachelor degree to PhD (Table 2).

Table 2. Distribution of students undergoing training by degree

Distribution of students by degree	Frequency	Percentage
BSc students	168	82,8
MSc students	17	8,4
PhD students	18	8,9
Total	203	100,0

3 RESULTS

To answer the research questions in this study, we drew upon two separate samples. The first sample was conducted fully online. We contacted the respondents by e-mail, which included a link to the e-survey. About 200 emails were distributed online with a link to the questionnaire. The final number of registered responses was 187, 21 out of them were counted as invalid.

For the qualitative aspect of our study and in order to increase the representative sample, a second sample was carried out. A further 50 questionnaires were distributed (in hard copy), a total of 48 were filled, and 11 were removed after a validation (Table 3).

Table 3. Sample distribution

Samples	Number of questionnaires	Number of completed questionnaires	Removed invalid data	Valid data
First	200	187	21	166
Second	50	48	11	37
Total	250	235	32	203

Based on the two samples, we collected a total of 235 completed questionnaires.

3.1 Self-assessment

Issues of self-assessment and application of the analytical skills included in the questionnaire allowed us to assess the confidence of the respondents in their own knowledge and skills. We used a three-tier scale - High, Middle, and Low level of applying the Analytical skills.

Self-assessment questions were included both at the input and at the output of the questionnaire. 22.1% initially assessed themselves as possessing the highest degree of analytical thinking. 7.95% rated their skills, at the end of the test, with a lower score than the initial one. The remaining respondents retained the same self assessment as at the beginning of the test.

3.2 Applying of analytical skills

Data processing of the essential part of the questionnaire about four categories of competences includes three steps:

1 In the preliminary processing the inconsistent or invalid responses were removed. Of these, 32 were counted as invalid, and the remaining 203 were used for analysis in the study. The available sample provided us with a valid response rate of 81.2%.

2 Next, every student was described according to the number of correct answers in each of this category. For example, if a student X answered correctly two out five questions in the AS category, two out of five in AT, five of five in FO, and one of five in QR, the student X = {AS {2}, AT (2), FO(5), QR(1)} is represented as the vector X = {2, 2, 5, 1}.

3 The profile of an expected student (generalized) for the program was defined as the average of each of these categories. Or the profile of expected students' entry competences is represented by the vectors SP = {AS, AT, FO, QR}.

In the Table 4 are summarized the results after data processing according to the number of correct answers in each of the four categories.

Table 4. Summary of results

Competence Category	Analytical skills	Abstract thinking	Facts and opinions	Quantitative reasoning
Technical	55,0 %	55,4 %	56,6 %	81,4 %
Non-technical	59,3 %	59,5 %	53,5 %	45,3 %

Based on the available data, we were able to reject the first hypothesis (H1). It turns out that the dependence between the specialty of the respondents and the level of applying the analytical skills in solving problems is negligible. Data show that students in a professional field "Informatics and computer science" show higher results than the control group with regard to questions requiring quantitative reasoning and questions expressing facts or opinion. This is an expected result. On the other hand, an unexpected result was the higher success rate of non-technical specialties in terms of abstract thinking and application of analytical skills.

This gives us a reason to believe that analytics and criticism are not bound skills with a particular professional profile or specialty. We have found a relationship between Quantitative reasoning and the profile of the specialist who owns them. As can be seen from the data, more respondents with a Computer Science profile have developed higher Quantitative reasoning skills (by 36,1%) than respondents from other areas.

Other question, answered by the survey, was the link between the level of education and the analytical skills. The hypothesis arises that the specialists who have a higher education have also a higher degree of analytical skills than specialists with a lower level of education. Data comparing the analyticity of the respondents when reporting their education level are shown in Table 5.

Table 5. Link between the level of education and the analyticity

Level of education	Analytical skills	Abstract thinking	Facts and opinions	Quantitative reasoning
Secondary education	56,4 %	56,0 %	55,1 %	89,9 %
Higher education (BSc, MSc, PhD)	53,6 %	57,1 %	60,0 %	35,7 %

The available data rejected our second hypothesis (H2). Available data show that there is no relationship between the level of education and analyticity.

Data showed more significant differences between the two groups (technical and non-technical) mainly in the Quantitative reasoning skills category, where the professionals with an "Informatics and Computer Science" profile show better results than the other respondents. The most important thing for us was to evaluate the analytical skills of BSc students in the field of Computer Science. Respondents with a profile of "Informatics and computer science" have a satisfying level of command of analytical skills, such as abstract thinking, logical and critical thinking. This makes them suitable candidates for a Master program in "Data Science".

4 CONCLUSION

To summarize the results:

- Technical competences are not critical for success in studying Data Science;
- Analytical thinking and problem solving attitude are not properties of longevity of study.

Results of this survey has several implication in designing Data Science Master Program Curriculum:

- To open the program for students with humanitarian or social science background, not to limit it to candidates with technical expertise;
- To include elective courses to train non-technical students to obtain the needed technical competences;
- To stress on developing those of analytical skills requisite for indirect data exploration;
- The program will benefit on such diverse audience by encouraging different view-points.

ACKNOWLEDGMENT

This work has been supported by National Science Fund at the Ministry of Education and Science, Republic of Bulgaria, within the Project DM 12/4 - 20/12/2017.

REFERENCES

[1] D. Christozov, K. Rasheva-Yordanova, S. Toleva-Stoimenova. “Designing data science curriculum in a way to address expected students entry competences,” *Proceedings of INTED2019 Conference*, pp. 2635 – 2640, 2019.

[2] D. Willingham, “Critical thinking. Why Is It So Hard to Teach?” *American Educator*, pp. 8-19, 2007. Retrieved from https://www.aft.org/sites/default/files/periodicals/Crit_Thinking.pdf

[3] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51-59, 2013.

[4] K. Rasheva-Yordanova, E. Iliev, B. Nikolova, “Analytical Thinking as a Key Competence for Overcoming the Data Science Divide,” *Proceedings of EDULEARN18 Conference*, pp. 7892 – 7898, 2018.

[5] K. Rasheva-Yordanova, S. Toleva-Stoimenova, D. Christozov, I. Kostadinova, “Road Map in Developing Data Science Competences,” Proceedings of INTED2019 Conference, 2019, pp. 6643 – 6650.

[6] P. A. Facione, “The California Critical Thinking Skills Test-college level,” Technical report #2. Factors predictive of CT skills. Millbrae, CA: California Academic Press, 1990.

[7] P. Mikalef, M. Giannakos, I. Pappas, J. Krogstie, “The Human Side of Big Data: Understanding the skills of the data scientist in education and industry,” *Proceedings of IEEE EDUCON 2018 Global Engineering Education Conference*, pp.503-512, 2018.

[8] R. H. Ennis, “Critical thinking and subject specificity: Clarification and needed research,” *Educational Researcher*, vol.18, no.3, pp.4-10, 1989.

[9] S. Toleva-Stoimenova, D. Christozov, K. Rasheva-Yordanova, Entry Competences Assessment of Data Science Potential Students. *Proceedings of INTED2019 Conference*, pp. 4248 – 4256.

[10] T. Schoenherr, C. Speier-Pero, "Data science, predictive analytics, and big data in supply chain management: Current state and future potential," *Journal of Business Logistics*, vol. 36, no. 1, pp. 120-132, 2015.