Manchester Metropolitan University

Comparative analysis of the skills among IT specialists in Russia and USA

Sergey Sosnovskikh, PhD
Manchester Metropolitan University (UK)
s.sosnovskikh@mmu.ac.uk

BASEES (April 2024)

Research background

Part of the continuous project studying the skills structure of the data analyst profession in Russia.

- Skhvediani, A., Sosnovskikh, S., Rudskaia, I., & Kudryavtseva, T. (2022). Identification and comparative analysis of the skills structure of the data analyst profession in Russia. *Journal of Education for Business*, *97*(5), 295–304. https://doi.org/10.1080/08832323.2021.1937018

JOURNAL OF EDUCATION FOR BUSINESS
2022, VOL. 97, NO. 5, 295–304
https://doi.org/10.1080/08832323.2021.1937018

Routledge
Taylor & Francis Group

Check for updates

Identification and comparative analysis of the skills structure of the data analyst profession in Russia

Angi Skhvediani[a], Sergey Sosnovskikh[b], Irina Rudskaia[a], and Tatiana Kudryavtseva[a]

[a]Graduate School of Industrial Economics, Institute of Industrial Management, Economics and Trade, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation; [b]Management and Entrepreneurship Department, Leicester Castle Business School, De Montfort University, Leicester, UK

ABSTRACT
The development of digital technologies has created a market need for specialists working with the big data that is necessary for making management decisions. This study aims to identify the skills structure of the data analyst profession (DAP) in Russia. The authors used a program code written in Python to examine relevant vacancies extracted from a recruitment website and employed a social network analysis method to identify skill clusters. Findings suggest that the DAP consists of predominately hard skills, that is, specialists must have the technical skills required to collect and process information. These results could be used to develop a Higher Education curriculum in data analysis.

KEYWORDS
Big data; data analyst; digital economy; Russia; skill gap; skills structure

Introduction

- Businesses have acknowledged the strategic benefits and potential value derivable from integrating big data capabilities into their structural framework (Wang et al., 2015).

- Viewing data as a quintessential business asset, these entities have concentrated on procuring the tools and developing the competencies essential for big data analytics (BDA) (Abbasi et al., 2016; Halwani et al., 2022).

- BDA intersects with many disciplines: information science, engineering, computer science, mathematics, social sciences, systems science, psychology, management, business, and economics (Jin et al., 2015).

- This field leverages methodologies from diverse areas, including probability theory, machine learning, statistical learning, computer programming, data engineering, pattern recognition, data visualisation, data warehousing, and high-performance computing (Jin et al., 2015).

- Additionally, it encompasses a broad spectrum of contemporary technological hardware, software applications, and services.

Research context and aim

- This issue is particularly pronounced in Russia, where there is a notable demand for IT managers (Anisimova et al., 2023; Matraeva et al., 2020), yet there exists a marked lack of clarity in delineating the distinctions between data analysts (DAs) and business analysts (BAs).

- **This study aims to bridge the existing knowledge gap by systematically collecting and analysing job descriptions for two professions situated at the intersection of big data and data science.**

- This investigation is contextualised within the COVID period and subsequent start of the military conflict between Russia and Ukraine in February 2022—a period marked by a heightened demand for highly-skilled professionals in digital analysis in Russia (Basu, 2023; Grizovski, 2022; Skhvediani et al., 2022).

- **Our research aims to examine the dynamics and evolving skillset structures associated with these two roles.**

- The selection of these professions for our study is predicated on their frequent co-mentioning within industry job descriptions, as well as the observation by several scholars that these roles often cluster together in discussions regarding job preparedness offered by existing curricula (Halwani et al., 2022; Persaud, 2020; Verma et al., 2019).

BDA paradigm

- In the realm of computing professions, data science is commonly conceptualised as an interdisciplinary amalgamation encompassing statistics, business intelligence, sociology, computer science, and communication (Sajid et al., 2021).

- Data science is characterised as a discipline dedicated to discovering, extracting, and analysing data for informed decision-making and predictive analysis.

- A more exhaustive inventory of tools and competencies expected of data scientists might encompass a wide array of elements, including knowledge of data management and storage tools like SQL, contemporary computing and manipulation tools capable of merging, aggregating, and iteratively processing data, proficiency in data visualisation and the principles of visual perception, understanding of confidence intervals through bootstrap methods, simulation, regression, variable selection, data mining/machine learning, classification, cross-validation, text mining, mapping, regular expressions, network science, MapReduce, among other subjects (Hardin et al., 2015).

- Differentiating DAs from data scientists presents a challenge. The academic groundwork for DAs begins with obtaining a degree in statistics, mathematics, computer science, management science, biological sciences, economics, information management, or business information systems (Abbott, 2014).

- The designation of a BAs professional traditionally refers to an individual who primarily examines existing business processes and the formulation of technological solutions (Vidgen et al., 2017).

- From an IT perspective, BA is delineated as an ensemble of solutions employed to construct analytical models and simulations for creating scenarios, comprehending realities, and projecting future states. At its core, business analytics represents a synthesis of business acumen and data science disciplines (Russom, 2013).

- Professionals in business analytics are tasked with generating novel business insights and formulating recommendations predicated on data analysis. They are expected to possess a background in IT and statistical knowledge, augmented by relevant business experience.

Theoretical Framework

- A competency framework is conventionally regarded as an analytical instrument that delineates the requisite skills, knowledge, personal attributes, and behaviours essential for the efficacious execution of a role within an organisation (Le Deist & Winterton, 2005; Lucia & Lepsinger, 1999).

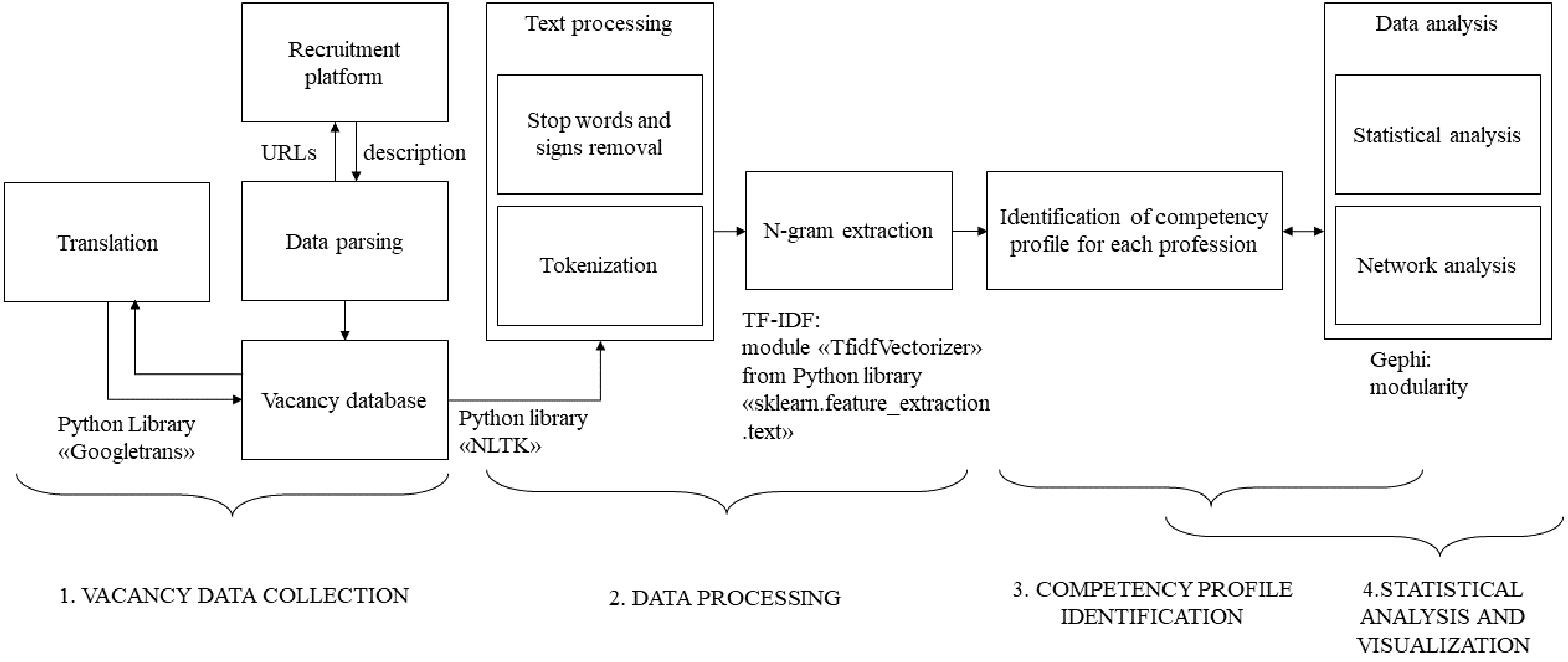
- Moreover, Le Deist & Winterton (2005) observe that competency models have been extensively utilised to synchronise individual capabilities with an organisation's core competencies (Rothwell & Lindholm, 1999).

- Competency transcends mere knowledge and skills to encompass attitudes, behaviours, work habits, abilities, and personal characteristics (Gangani et al., 2006; Le Deist & Winterton, 2005; Russ-Eft, 1995), and is manifested not in isolation but within contexts pivotal to professional practice (Hager & Gonczi, 1996; Russ-Eft, 1995).

- The adoption of a competency approach is motivated by its potential to signal labour market demands, assist individuals in navigating career mobility (Le Deist & Winterton, 2005), enhance the capacity of educational providers to more effectively bridge education and training with labour market requirements (Hager & Gonczi, 1996), and demonstrate the convergence between formal education and experiential learning in the cultivation of professional competence (Cheetham & Chivers, 1996).

Methodology

Algorithm and tools for identification, statistical and graph analysis of key competencies presented in vacancies from the online job postings.



- In the first stage, we collected data using web scraping of vacancies on the HeadHunter website and parsing the collected data.
- The data collection (web scraping and data parsing) was performed in Python in **July 2020 and July 2023**. Our sample size comprises **2357 vacancies.**
- First, the URLs of all vacancies for each position were collected. Next, each URL was parsed for positions: job title, job description, salary, address, and work experience.
- In the second stage, we processed the data. We tokenised text descriptions of vacancies; we removed stop words and signs from them using the Nltk library, which helped eliminate words of little significance.
- To extract unigrams, bigrams and trigrams from job descriptions, we used the TF-IDF (Term Frequency-Inverse Document Frequency) method, implemented using the TfidfVectorizer module of the Scikit-Learn library.
- For each n-gram, we calculated how many job descriptions they appeared in and arranged the ranking of n-grams for each group. The TF-IDF method allowed the conversion of text into numeric vectors:

$$TF_{t,v} = \frac{f_{t,v}}{max\{f_{t',v} : t' \epsilon\ v\}} ,$$

- where TF is the frequency of word/term *t* in vacancy *v*
- t – word/term
- v – vacancy
- $f_{t,v}$ – number of mentions of the word/term *t* in vacancy *v*
- $max\{f_{t',v} : t' \epsilon\ v\}$ – total number of words/terms in the vacancy.

$$IDF_{t,V} = log \frac{V}{\{v\ \epsilon\ V : t\ \epsilon\ v\}},$$

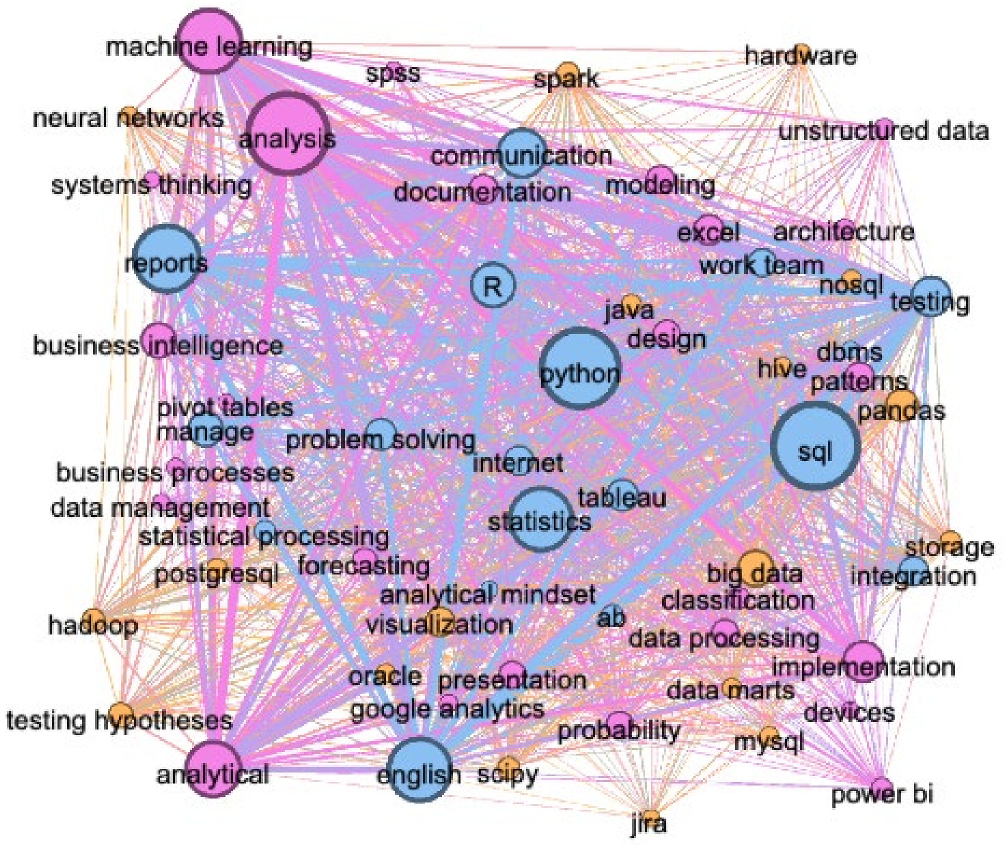
- where IDF – inverse vacancy frequency
- V – total number of vacancies
- $\{v\ \epsilon\ V : t\ \epsilon\ v\}$ – number of vacancies in which the word/term is mentioned.

$$TF - IDF = TF_{t,v} \cdot IDF_{t,V}$$

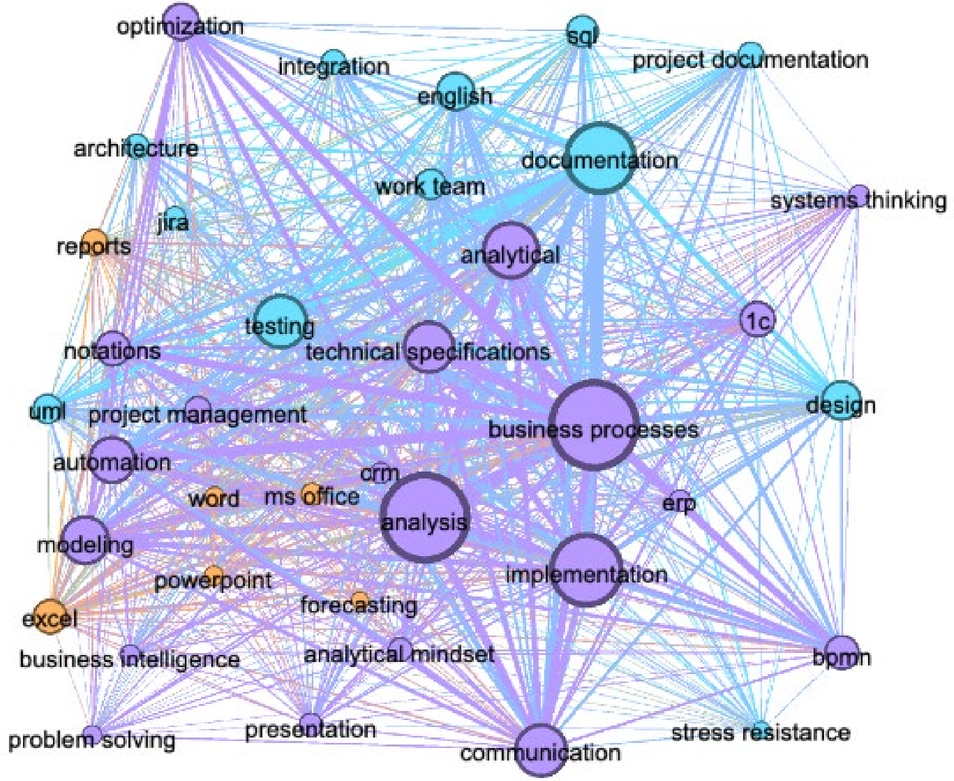
- In the third stage, based on the selected n-grams, we graphically determined the structure of competencies. In this case, categorisation occurred using graph construction.
- In the fourth stage, we carried out a statistical analysis of data on vacancies and their visualisation. We analysed the most mentioned competencies in the competency structure and assessed their relationship.
- To assess the connection, we constructed the graph using Gephi software, where the selected competencies were represented as nodes and the connections between them (joint appearance in the same vacancy) as edges. To divide competencies into groups/clusters, we used a measure of network structure - modularity.
- This graph allowed us to identify relationships between competencies and define clusters with similar competencies.

Results 2020

Data Analyst

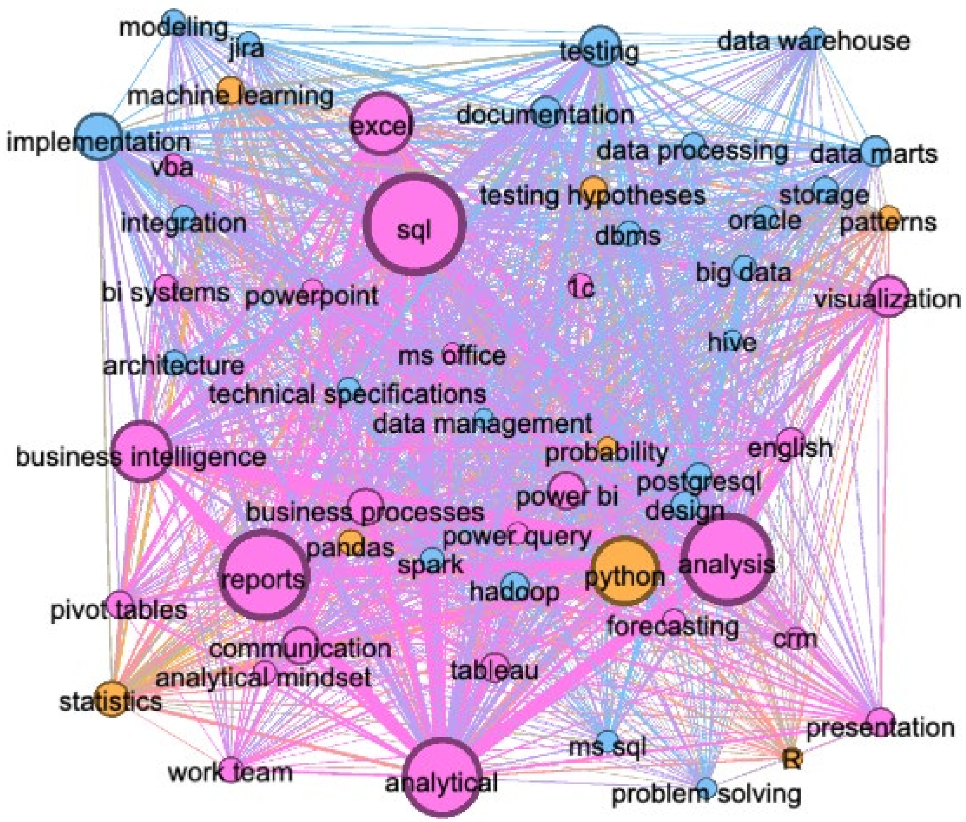


Business Analyst

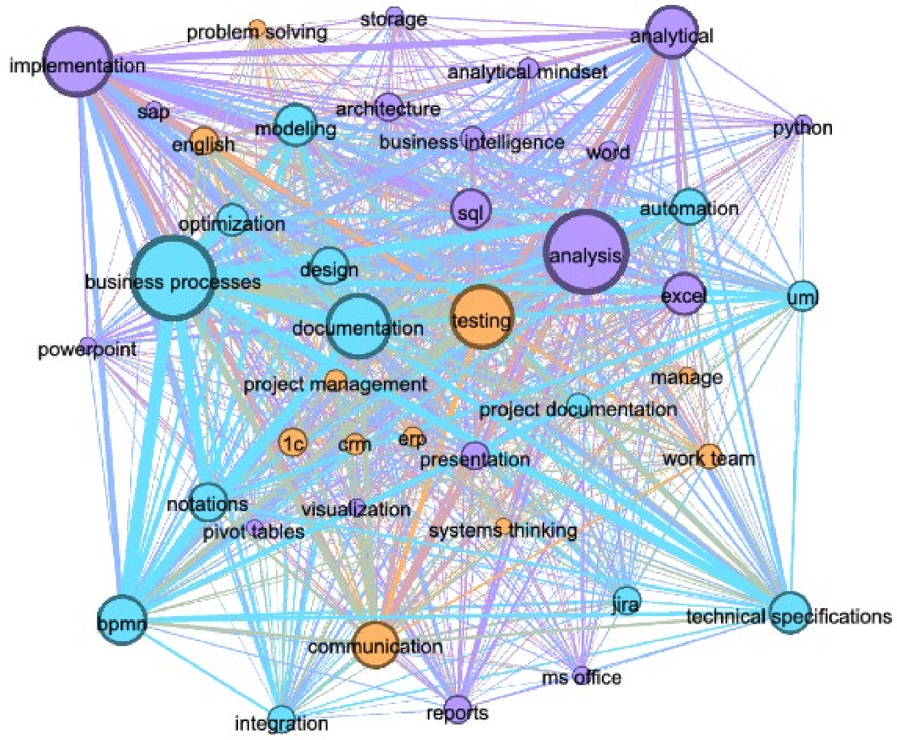


Results 2023

Data Analyst



Business Analyst



Comparison of the DA competences between 2020 and 2023

Competence	Category	2020	2023	2020, position	2023, position	Change	Change, position
Excel	Statistics	18.6%	43.3%	19	6	24.7%	13
Business intelligence	Business process management	23.3%	42.1%	14	7	18.8%	7
Business processes	Business process management	5.8%	20.4%	49	13	14.6%	36
Analytical	Decision making skills	44.2%	57.6%	8	4	13.4%	4
Power BI	Visualization solutions	9.3%	21.3%	36	11	12.0%	25
Reports	Decision making skills	55.8%	67.6%	4	3	11.8%	1
1C	Enterprise systems software	0.0%	9.2%	70	31	9.2%	39
Data marts	Structured data management	7.0%	14.8%	44	16	7.8%	28
Pivot tables	Structured data management	4.7%	12.1%	54	22	7.5%	32
Power query	Structured data management	0.0%	7.0%	70	46	7.0%	24
Visualization	Visualization solutions	17.4%	24.5%	20	9	7.0%	11
Technical specifications	Business process management	1.2%	8.1%	68	42	6.9%	26
PowerPoint	Communication skills	1.2%	7.4%	68	44	6.2%	24
BI systems	Visualization solutions	3.5%	9.2%	62	31	5.7%	31
MS SQL	Structured data management	2.3%	7.2%	64	45	4.9%	19
DBMS	Structured data management	4.7%	9.5%	54	30	4.9%	24
Data warehouse	Structured data management	2.3%	7.0%	64	46	4.7%	18
VBA	Programming skills	2.3%	6.9%	64	49	4.5%	15
Analytical mindset	Decision making skills	4.7%	8.5%	54	40	3.8%	14
CRM	Enterprise systems software	3.5%	7.0%	62	46	3.6%	16
Postgresql	Structured data management	8.1%	11.1%	41	25	3.0%	16
MS office	Communication skills	2.3%	5.3%	64	52	3.0%	12
Jira	Project management	4.7%	7.6%	54	43	2.9%	11
Hadoop	Big data management	10.5%	12.9%	33	20	2.4%	13
SQL	Structured data management	75.6%	77.6%	1	1	2.1%	0
Oracle	Structured data management	7.0%	8.8%	44	37	1.8%	7
Architecture	Hardware	8.1%	9.7%	41	29	1.5%	12
Storage	Hardware	9.3%	10.7%	36	26	1.4%	10

Competence	Category	2020	2023	2020, position	2023, position	Change	Change, position
Implementation	Decision making skills	29.1%	29.8%	11	8	0.7%	3
Data management	Structured data management	4.7%	4.9%	54	54	0.3%	0
Hive	Big data management	5.8%	5.1%	49	53	-0.7%	-4
Testing hypotheses	Statistics	12.8%	11.8%	31	24	-1.0%	7
Systems thinking	Decision making skills	4.7%	3.2%	54	57	-1.5%	-3
Spark	Big data management	10.5%	8.8%	33	37	-1.7%	-4
Presentation	Communication skills	15.1%	13.4%	27	19	-1.7%	8
Analysis	Decision making skills	70.9%	68.8%	2	2	-2.1%	0
Documentation	Communication skills	17.4%	15.3%	20	15	-2.1%	5
Design	Decision making skills	15.1%	12.9%	27	20	-2.3%	7
Testing	Decision making skills	27.9%	23.9%	12	10	-4.0%	2
Forecasting	Data mining	12.8%	8.8%	31	37	-4.0%	-6
Hardware	Hardware	5.8%	1.6%	49	61	-4.2%	-12
Google analytics	Specialized analytics solutions	5.8%	1.6%	49	61	-4.2%	-12
Devices	Hardware	4.7%	0.4%	54	69	-4.3%	-15
SPSS	Statistics	4.7%		54		-4.7%	
Neural networks	Advanced modelling/analytics techniques	5.8%	1.1%	49	66	-4.8%	-17
Unstructured data	Big data management	7.0%	1.4%	44	63	-5.6%	-19
MySQL	Structured data management	7.0%	1.4%	44	63	-5.6%	-19
Tableau	Visualization solutions	19.8%	13.9%	17	17	-5.9%	0
Data processing	Data mining	15.1%	9.2%	27	31	-6.0%	-4
NoSQL	Big data management	7.0%	0.7%	44	68	-6.3%	-24
Work team	Organization skills	16.3%	9.9%	25	28	-6.4%	-3
Java	Programming skills	8.1%	1.1%	41	66	-7.1%	-25
Modelling	Decision making skills	16.3%	9.2%	25	31	-7.1%	-6
AB-test	Advanced modelling/analytics techniques	9.3%	1.8%	36	60	-7.5%	-24
SciPy	Statistics	10.5%	2.6%	33	59	-7.8%	-26
Classification	Data mining	9.3%	1.4%	36	63	-7.9%	-27
Integration	Enterprise systems software	17.4%	9.0%	20	35	-8.5%	-15
Patterns	Data mining	17.4%	8.3%	20	41	-9.2%	-21

Competence	Category	2020	2023	2020, position	2023, position	Change	Change, position
Statistical processing	Statistics	9.3%		36		-9.3%	
Probability	Statistics	15.1%	4.9%	27	54	-10.2%	-27
Pandas	Statistics	20.9%	10.0%	16	27	-10.9%	-11
Internet	Networking	17.4%	3.9%	20	56	-13.6%	-36
Problem solving	Decision making skills	19.8%	6.0%	17	50	-13.8%	-33
Big data	Big data management	24.4%	9.0%	13	35	-15.4%	-22
Communication	Communication skills	37.2%	21.0%	9	12	-16.3%	-3
Python	Programming skills	68.6%	48.8%	3	5	-19.8%	-2
Manage	Organization skills	23.3%	3.0%	14	58	-20.3%	-44
R	Statistics	32.6%	6.0%	10	50	-26.6%	-40
Statistics	Statistics	52.3%	19.0%	6	14	-33.3%	-8
English	Language skills	52.3%	13.6%	6	18	-38.8%	-12
Machine learning	Advanced modelling/analytics techniques	53.5%	12.1%	5	22	-41.3%	-17

Comparison of the BA competences between 2020 and 2023

Competence	Category	2020	2023	2020, position	2023, position	Change	Change, position
BPMN	Business process management	20.77%	34.70%	17	7	13.93%	10
Testing	Decision making skills	38.83%	47.60%	6	5	8.77%	1
SQL	Structured data management	18.35%	26.90%	18	12	8.55%	6
Excel	Statistics	20.91%	28.40%	16	10	7.49%	6
Jira	Project management	9.53%	15.60%	28	20	6.07%	8
Presentation	Communication skills	10.67%	15.60%	27	20	4.93%	7
Business intelligence	Business process management	7.82%	12.00%	33	27	4.18%	6
Architecture	Hardware	11.81%	15.80%	24	19	3.99%	5
Notations	Business process management	21.34%	25.30%	15	13	3.96%	2
Python	Programming skills	2.56%	6.20%	42	34	3.64%	8
Integration	Enterprise systems software	11.81%	15.00%	24	24	3.19%	0
Visualization	Visualization solutions	2.84%	6.00%	40	35	3.16%	5
Pivot tables	Structured data management	1.71%	4.80%	44	41	3.09%	3
Storage	Hardware	2.42%	5.50%	43	36	3.08%	7
Manage	Organization skills	2.84%	5.00%	40	40	2.16%	0
Reports	Decision making skills	13.37%	15.40%	22	23	2.03%	-1
UML	Business process management	16.22%	17.80%	20	17	1.58%	3
SAP	Enterprise systems software	3.98%	5.10%	39	37	1.12%	2
Word	Communication skills	6.83%	7.80%	35	32	0.97%	3
CRM	Enterprise systems software	8.39%	9.30%	32	29	0.91%	3
Forecasting	Data mining	4.69%	4.50%	38	43	-0.19%	-5
Project documentation	Project management	13.23%	12.40%	23	26	-0.83%	-3
ERP	Enterprise systems software	9.25%	8.30%	30	31	-0.95%	-1
Design	Decision making skills	25.75%	24.30%	11	14	-1.45%	-3
PowerPoint	Communication skills	6.69%	5.10%	36	37	-1.59%	-1
Stress resistance	Organization skills	5.12%	3.00%	37	44	-2.12%	-7
Analytical mindset	Decision making skills	11.10%	8.80%	26	30	-2.30%	-4

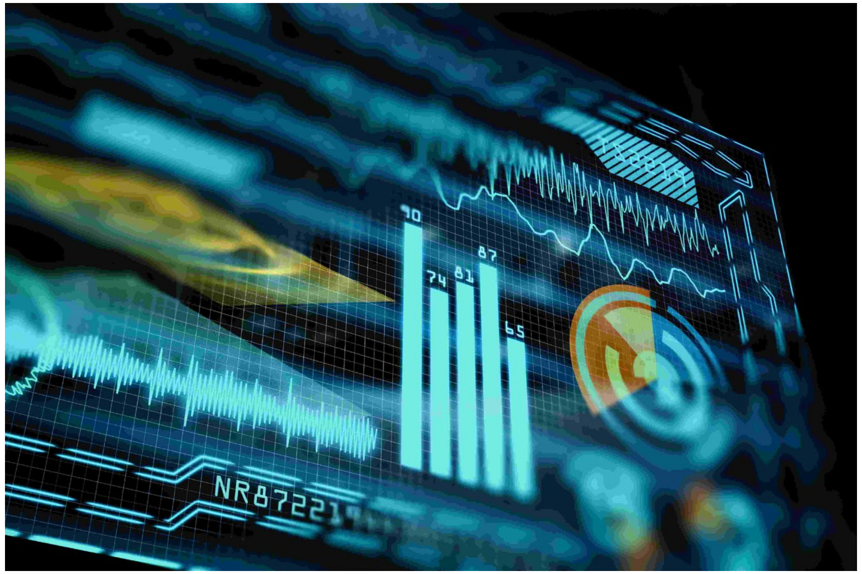
Competence	Category	2020	2023	2020, position	2023, position	Change	Change, position
Problem solving	Decision making skills	7.25%	4.80%	34	41	-2.45%	-7
MS office	Communication skills	9.39%	6.60%	29	33	-2.79%	-4
Implementation	Decision making skills	56.61%	52.90%	3	3	-3.71%	0
Systems thinking	Decision making skills	8.96%	5.10%	31	37	-3.86%	-6
Optimization	Business process management	23.33%	19.30%	13	16	-4.03%	-3
Analytical	Decision making skills	41.68%	37.60%	5	6	-4.08%	-1
Project management	Project management	13.80%	9.60%	21	28	-4.20%	-7
Work team	Organization skills	17.35%	12.60%	19	25	-4.75%	-6
Business processes	Business process management	71.83%	66.80%	1	1	-5.03%	0
Modeling	Decision making skills	33.85%	28.60%	9	9	-5.25%	0
Analysis	Decision making skills	71.55%	66.00%	2	2	-5.55%	0
Communication	Communication skills	37.98%	32.30%	7	8	-5.68%	-1
1C	Enterprise systems software	22.05%	16.10%	14	18	-5.95%	-4
Documentation	Communication skills	55.19%	48.60%	4	4	-6.59%	0
Automation	Business process management	32.29%	23.40%	10	15	-8.89%	-5
English	Language skills	24.61%	15.60%	12	20	-9.01%	-8
Technical specifications	Business process management	36.84%	27.70%	8	11	-9.14%	-3

Conclusions

- This study meticulously delineates the distinct differences between these two professions.
- This suggests a shift towards recognising the importance of diverse skill sets within the data science field in Russia.
- The observed transformation in the skills framework and requisite competencies presents a fascinating dynamic in studies years: **DAs are increasingly expected to exhibit soft skills, including management, communication, and teamwork capabilities, whereas BAs are more frequently required to possess hard skills, such as proficiency in data analysis, analytics, and specific language skills (coding).**

- Comparing our results to similar research about the skills structures in the USA (Halwani et al., 2022; Fareri et al., 2020; Sousa & Rocha, 2019; Verma et al., 2019), we see that in both countries, most frequently mentioned skill categories coincide: **statistical packages, structured data management, and decision making**. However, they occupy different places in the ranked list.
- In the US, **the emphasis is given to soft skills represented by the decision making and organisation categories**, while in Russia they are ranked lower, respectively.
- On the other hand, **programming skills are quite important in Russia**, while in the USA, they are not included in the top most frequently cited skills in relevant vacancies.

Manchester
Metropolitan
University



Thank you for your attention!

Sergey Sosnovskikh, PhD

Manchester Metropolitan University (UK)

s.sosnovskikh@mmu.ac.uk