Table 2: relationship between data experts’ tools, skills, relevant stage in the data cycle, and types of clients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** |  | **Subjects of Study** | **Level of Certification** | **Remarks** |
|  |  |  | Beginner (1)  | Intermediate(2) | Advanced(3)  |  |
| Background and business value | Background and business value | Information about the organization (background, goals, business value) | √ |  |  |  |
| 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 defining the 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 |
| Cleansing, validating and inspecting the data | Basic data inspection | √ |  |  | Importance of the issue of information quality and types of quality issues |
| Inspecting data and detecting biases in the data |  | √ |  | Outliers; handling missing 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 of Automated 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,  |
| 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 | √ |  |  |  |