Journal of Information Systems Education

Volume 35
Issue 3
Summer 2024

Data Analyst Competencies: A Theory-Driven Investigation of Industry Requirements in the Field of Data Analytics

Cassandra Artman Collier and Anne L. Powell

Recommended Citation: Collier, C. A., & Powell, A. L. (2024). Data Analyst Competencies: A Theory-Driven Investigation of Industry Requirements in the Field of Data Analytics. *Journal of Information Systems Education*, 35(3), 325-376. https://doi.org/10.62273/SPYC4248

Article Link: https://jise.org/Volume35/n3/JISE2024v35n3pp325-376.html

Received:	September 26, 2023
First Decision:	January 4, 2024
Accepted:	January 31, 2024
Published:	September 15, 2024

Find archived papers, submission instructions, terms of use, and much more at the JISE website:
https://jise.org

ISSN: 2574-3872 (Online) 1055-3096 (Print)

Journal of Information Systems Education, 35(3), 325-376, Summer 2024
https://doi.org/10.62273/SPYC4248

Data Analyst Competencies: A Theory-Driven Investigation of Industry Requirements in the Field of Data Analytics

Cassandra Artman Collier
Anne L. Powell
Department of Computer Management and Information Systems
Southern Illinois University – Edwardsville
Edwardsville, IL 62025, USA
cassaco@siue.edu, apowell@siue.edu

ABSTRACT

As organizations' reliance on data increases, the prevalence of data analytics programs in universities likewise increases. However, despite this specialized education, scholars still report a gap between the knowledge and skills students graduate with and those required by industry upon beginning work as an entry-level data analyst. We draw on theories of data analysis and curriculum frameworks to create an integrated theoretical model to drive our work. We then conduct an extensive analysis to identify relevant languages and tools in data analysis today and collect data from hiring managers seeking data analysts through a survey-based research method. We report the major knowledge, skills, and dispositions desired in the industry today for entry-level data analysts, including specific software platforms and applications. Our findings highlight several leading tools and a better understanding of how well data analysts are expected to know each tool and when those tools are used throughout the knowledge discovery via the data analytics lifecycle. This produces important contributions, particularly to academics working to keep data analytics programs competitive and up-to-date in today's rapidly changing landscape.

Keywords: Data analytics, Information systems education, IS education research, Careers

1. INTRODUCTION

As the amount of data generated in our world increases, so does the importance of data analysis in our organizations. Data analytics has a significant impact on a wide variety of industries and applications, rendering the skills of those working with data highly important (Dwivedi et al., 2021). To help provide these skill sets, universities offer undergraduate and graduate programs intended to train entry-level information systems (IS) analysts to take on data-related challenges (*The Best Business Analytics Programs, Ranked*, n.d.). Quickly, "the number of university-based analytics programs… exploded" with at least 400 analytics-degree programs being offered across 220 international business schools (LeClair, 2013).

Despite these specialized programs, scholars report a gap between the skills demonstrated by graduates of these programs and the requirements of the data analyst positions in which graduates begin their careers (Setor & Joseph, 2021). This gap translates into expensive, time- and resource-consuming training that organizations must provide to their entry-level employees. In 2021, U.S. training expenditures totaled $92.3 billion, with companies spending an average of over $1,000 per learner and new employees spending upward of 60 hours in training courses ("2021 Training Industry Report," 2021).

Given the existence of a gap between university curricula and industry requirements in the field of data analytics, there is a clear motivation for university business schools to understand the needs of industry and provide these within the data analytics curricula developed to train students. Therefore, our work aims to answer the research question: *What competencies – that is, knowledge, skills, and dispositions – are expected of entry-level data analysts today?* By addressing this question, we can contribute beneficial outcomes for multiple stakeholders, including:

- *Academics:* Academics can understand industry needs, thereby offering more relevant and real-world content in courses, and ensuring adherence to standards such as those required by the AACSB (AACSB International, 2021). Academic programs will be more well-rounded and appropriate, allowing universities to compete with other services that offer to fill the gap observed between education and career (e.g., by MOOCs, online bootcamps, and non-academic training). This will help address the known challenge of curriculum development in IS (Cummings & Janicki, 2021) and extend prior research on data analytics curriculum development (Gupta et al., 2015). While we apply this method to the field of data analytics, our work may also provide a method that could be replicated in the future to 1) re-evaluate the rapidly changing data analytics field and 2) apply it to other IS fields such as cybersecurity, networking, or IT support. This is not only relevant to traditional education, but also reskilling and upskilling – necessary processes considering the rapid change in industry today (Li, 2022).
- *Industry practitioners:* As a result of the contribution to academics described above, practitioners will have a more qualified base of individuals from whom to hire

and can save money, time, and other resources currently devoted to training entry-level data analysts on missing skill sets.

- *Students*: Students will experience more relevant, real-world content in their courses. Not only will this contribute to a higher level of preparation for their careers, but it will also help students evaluate whether they are on the right career path and forecast their enjoyment in their ultimate career. Students can also draw on these findings to supplement their own education and experiences to be more competitive in the job market for entry-level data analyst roles.

2. LITERATURE REVIEW

2.1 General Information Systems Skill Sets

Several scholars have studied the skill sets that will make students graduating from IS programs more employable. Many of these focus on IS students as a homogeneous group, e.g., by studying curricular guidelines for the IS field (Leidig & Salmela, 2022). For example, Setor and Joseph (2021) report that cooperative education, internship, and mentorship experiences increase the likelihood of initial IT employment. In their biennial report of knowledge and skills required in the IS field, Cummings and Janicki (2021) report an increased need for students to have experience with the MacOS platform, cloud and virtualization technologies, network security, a wider variety of database platforms, and a working knowledge of different programming languages. Some scholars highlight the importance of a balance between business and technical content for IS students (Elrod et al., 2022; Plice & Reinig, 2007; Qiu et al., 2020) while others indicate that technical skills are less important than project management, business domain, or sourcing capabilities (Dubey & Tiwari, 2020; Goles et al., 2008). In their Information Systems Job Index, Mandviwalla et al. (2022) report that students graduating with bachelor's degrees in IS have an 80% placement, earn higher salaries than other business majors, and are almost twice as likely to get a job offer if they have done an internship. Despite a high level of optimism from students about their job offers (Mandviwalla et al., 2022), scholars report a gap between the skills obtained by graduates and the expectations of employers (Sahin & Celikkan, 2020).

2.2 Specialized Information Systems Skill Sets

Developers of IS curricula may benefit from a narrower view of IS student employability. The IS field is broad, and graduates of university IS programs may work in any number of specialized fields from networking to cybersecurity to project management. Each of these specializations requires unique knowledge and skills, which is reflected in specialized programs and certificates developing in the academy now. A minority of research in this area has focused on specific IT roles such as the programmer/analyst role (Lee, 2008) or the software engineer role (Assyne et al., 2022; Colomo-Palacios et al., 2013).

Another example of specific IT training lies in the data analytics specialization, intended to prepare IS students for a role as a data analyst, data scientist, data visualization analyst, analytics analyst, or similar role (Clayton & Clopton, 2019). Of the jobs accepted by students with bachelor's degrees in IS in 2021-2022, 20% of them were in Data/Analytics – up from 16% in 2019 (Mandviwalla et al., 2022).

In 2020, Dong and Triche published a study investigating longitudinal trends of skills for entry-level data analysts between 2014 and 2018. By scraping and analyzing data from Indeed.com, the authors report which tools gained popularity in job postings over the four-year span (SQL Server, NoSQL, Tableau, Power BI, Python, Pig, Hadoop, Salesforce, Azure, Hive, Google Analytics, R, SAS, and SPSS), which grew less popular (Microsoft Access, Cognos, and SAP), and which stayed the same (Oracle, Microsoft Office) (Dong & Triche, 2020).

Dong and Triche's work gives IS curricula designers important insight regarding which technical skills are most frequently cited in job postings for data analysts. However, a few gaps remain in our understanding. First, the job postings utilized in Dong and Triche's study were not comprehensive but rather only those captured by the Common Crawl tool to facilitate the longitudinal nature of their study. Second, the wildcard searches used by Dong and Triche, which included titles like data analyst and business intelligence analyst, excluded some other valuable titles such as analytics analyst and data visualization analyst. Finally, their results tell us the frequency with which certain skills were listed in job postings but cannot shed light on which of the listed technical skills were most important to any given job.

In another study of job descriptions, Verma et al. (2019) analyzed descriptions for four analytics-related job titles: business analyst, business intelligence analyst, data scientist, and data analyst in four U.S. states. They utilized content analysis to rank skill categories required for each job. Their findings allow for comparison across the job types: business analyst is the least technical and requires a high level of domain knowledge, while business intelligence analyst jobs focus on data management and statistics. The authors report some overlap between data analysts and data scientists regarding decision-making and organization skills, but data scientist roles tend to additionally require statistical and programming skills. Verma et al.'s study is helpful in comparing expectations across job types but is limited by including only a few U.S. states and only a few job titles.

2.3 Skill Transferability

Another gap in the extant studies on job descriptions lies in what we have conceptualized as skill transferability. In extant literature, researchers have examined how skills transfer from one context to another. For example, Lee (2005) examined how skills transfer across different IT jobs: programmer/analyst, systems analyst, and IT manager. Lee (2005) defined transferability as "the degree to which a person can move from one place to another" (p. 85) and conceptualized this as individuals moving between jobs. Lee (2005) summarized other studies that have focused on transferring skills from other countries to the U.S., from the military to other employment, from college to a corporate job, and from public to private organizations. Other scholars have called for investigations on how skills such as programming transfer to other non-IS domains (Scherer, 2016). However, we are interested in how skills in a specific tool (or language, software, platform) transfer to similar tools.

We refer to this concept as *skill transferability among tools*. Extant literature has demonstrated that skill transferability occurs among programming languages (Scholtz, 1993), but the research is dated, and, to our knowledge, no such investigations have been made for software platforms or skills associated with data analysis. However, it stands to reason that as much as a student who learns how to create loops in Java will be able to apply that knowledge to Python, a student who learns to develop dashboards in Tableau will be able to apply that knowledge to Power BI. Therefore, one goal of our data collection was to understand if hiring managers subscribe to this belief for the languages, tools, and platforms used by entry-level data analysts.

2.4 Literature Review Conclusion

In general, job description studies like those done by Dong and Triche (2020) and Verma et al. (2019) suffer from a limitation that findings indicate what industry professionals *ask for* in their job descriptions, but these descriptions may not describe the actual positions accurately and may be limited by what industry professionals have encountered in applicants on the market so far. To build on this existing body of work, our project asks hiring managers to describe their *ideal* entry-level data analyst candidate. To structure this data collection, we drew on the following theories.

3. THEORY

The primary activity required of data analysts is knowledge discovery via data analytics (KDDA), which makes Li et al.'s (2016) snail shell process model for KDDA a logical theoretical model on which to base our research. The model integrates "eight key phases and related tasks at the meta-level," allowing us insight into the activities of a data analyst that contribute to their organization. The phases, descriptions, and example tasks in each are presented in Table 1.

We integrate the snail shell process model for KDDA with a competency-based lens of individual preparedness as recommended by the Computing Curricula 2020 Paradigms for Global Computing Education (CC2020 Task Force, 2020). According to the CC2020 Task Force, "there is a general agreement in educational circles that career success requires three things: *knowledge* – "know-what" – a proficiency in core concepts and content and the application of learning to new situations; *skills* – "know-how" – the ability to carry out tasks with determined results; and *dispositions* – "know-why" – intellectual, social, or moral tendencies" (CC2020 Task Force, 2020, p. 13, emphasis added). Together, these three elements form a competency, observed within the performance of a task.

Finally, we incorporate one more element from the CC2020 Paradigms for Global Computing Education: the revised Bloom's Cognitive Skills list (CC2020 Task Force, 2020). Based on Bloom's Taxonomy of the Cognitive Domain (Huitt, 2011), this list allows for differentiation among levels of skills as observable knowledge in action. The cognitive skills list identifies six levels of understanding, their definitions, and example verbs to be used in writing competency statements for each. We present these in Table 2, adapted from Table D.4 in the CC2020 Paradigms for Global Computing Education.

We integrate the above-described theories to produce an integrated model (depicted in Figure 1) to describe 1) the elements of a competency, 2) the phases of the KDDA when those competencies would be utilized, 3) the levels an individual may possess of these competencies, and 4) the mechanisms by which an individual can achieve various levels of competencies. In the figure, we zoom in on just one phase of the snail shell model (Data Preparation) but note that each of the eight phases will have associated tasks, knowledge, skills, dispositions, and competencies.

Utilizing this integrated model, we designed a survey to investigate 1) which competencies 2) at which levels are most valued/required by hiring managers, and 3) when those competencies are used in the KDDA.

4. METHODS

To address our research question, we undertook an analysis to identify the major programming languages, tools, and software programs utilized in data analytics. We drew upon both academic sources and practitioner sources to identify relevant languages and tools (this process, the sources, and a complete list of tools are included in Appendix A). Through this process, we identified 119 programming languages and software tools relevant to the field of data analytics.

Many of the practitioner and academic sources we utilized to identify relevant tools categorized them based on their purpose. However, nearly all tools in data analytics are multifunctional; for example, tools with functions for data visualization almost always also have data storage and cleaning capabilities and enterprise software often has functionality for data analysis. While the academic community may benefit from a research-based typology of data analytic languages and tools, such efforts are highly complex and challenging (Nickerson et al., 2013) and lie outside the scope of this paper. However, given the sheer number of tools of interest, loose categories provided a necessary structure for our data collection method.

Drawing on the categorizations in the academic and practitioner sources we utilized, we categorized tools into four loose groups: Programming languages (25), Databases/Data Management (25), Software – Analysis (43), and Software – Enterprise, Statistics, and Visualization (26). However, we have provided all raw data in the appendices of this manuscript. Scholars or practitioners who would prefer to see the tools grouped differently can easily reconfigure this raw data and conclude differently-formed categories of tools. The loose categories we utilized are presented and defined in Table 3.

Having identified a wide array of tools (for the sake of brevity, in the rest of the paper we will simply refer to these as "tools," although they might more accurately be described as languages, tools, software programs, or platforms) likely to be relevant for entry-level data analysts, we developed a survey to investigate which of these are expected of entry-level data analysts by the managers who hire them. We partnered with Qualtrics to have this survey distributed to a panel of hiring managers involved in the hiring of entry-level data analysts. Survey respondents answered two screening questions to ensure they were eligible for the study and then provided various demographic information about both the respondent and their organization. All survey questions are available in Appendix B.

Phase	Description	Example Tasks
Problem Formulation (PF)	Formulating the business problems that a given project should address and transforming it into an actionable analytic problem statement.	• Determine business objectives and success measures • Deploy problem formulation strategies • Define business problem • Determine KDDA problem, goals, and success measures
Business Understanding (BU)	Business requirement elicitation that ultimately helps to translate high-level executive requirements into specific analytic needs.	• Establish business case: costs and benefits, requirements, assumptions, constraints, risks, contingencies, inventory of resources, etc. • Analytics capability maturity assessment • Enterprise knowledge acquisition from existing documentation, business processes, ETL processes, queries, BI reports, etc. • Determine project management methodology • Select initial tools and techniques
Data Understanding (DU)	Familiarizing oneself with data from various sources that are relevant to solving the analytic problem.	• Within-DBMS data exploration (e.g., writing SQL or other NoSQL languages) • Out-DBMS exploration (advanced visualization tools) • DU for business requirements – many business requirements and business logic reside in the data or data related processes • DU for modeling requirements: depending on the selected modeling technique, different types of DU need to be performed • Verify data quality: depends on business requirements as well as the analytic techniques selected • Describe data: data description should include source, owner, update frequency, and other relevant attributes
Data Preparation (DP)	Based on outputs from the PF, BU, and DU phases, an initial data integration requirement shall first be created, including how each data element for modeling would be sourced or transformed.	• Create data integration requirements • Data transformation based on quality, business, or modeling requirements • Data integration
Modeling (MO)	Selecting applicable modeling techniques and building analytic models to provide most desirable outcomes for the stated analytic goal.	• Select modeling techniques • Describe modeling rules for the modeling technique • Defining training and testing strategy • Build models • Assess models
Evaluation (E)	Candidate models are evaluated against business objectives and business problems formulated in the PF phase.	• Evaluate result • Conduct field test • Review analytic process • Communicate results
Deployment (D)	Project is deployed.	• Create deployment plan • Produce final project report and final presentation • Review project
Maintenance (MA)	Project is maintained.	• Describe and store analytic results • Create a model maintenance process • Define change initiation • Monitor model usage

Table 1. The Phases and Tasks of the Knowledge Discovery by Data Analytics Process Model, Adapted from Li et al. (2016)

Level	Definition	Example Verbs
1. Remembering	Exhibit memory of previously learned materials by recalling facts, terms, basic concepts, and answers.	Choose, define, find, label, list, recall
2. Understanding	Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions.	Classify, compare, explain, illustrate, infer, interpret, summarize
3. Applying	Solve problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.	Apply, build, choose, construct, develop, model, organize, solve
4. Analyzing	Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support those inferences.	Analyze, assume, conclude, discover, examine, test for
5. Evaluating	Present and defend opinions by making judgments about information, validity of ideas, or quality of ideas.	Agree, appraise, determine, decide, explain, measure, prioritize, recommend
6. Creating	Compile information together in a different way by combining elements in a new pattern or proposing alternatives.	Change, choose, compile, create, design, elaborate, formulate, solve, test

Table 2. Revised Bloom's Cognitive Skill List, Adapted from CC2020 Paradigms for Global Computing Education Table D.4

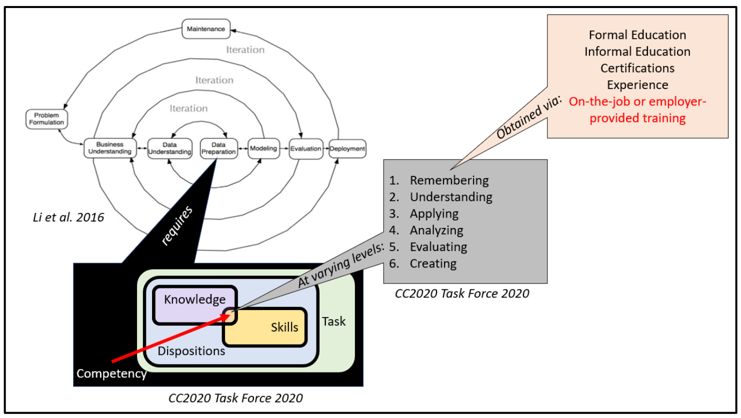


Figure 1. Theoretical Model Composed of the KDDA Snail Shell Model (Li et al. 2016), the Competency Viewpoint (CC 2020 Task Force, 2020), and the Elements of Computing Education (CC 2020 Task Force, 2020)

Category	Definition	Count	Examples
Programming languages	Methods of communicating with computers, typically by writing code interpreted by a compiler that commits actions within an automated computer program.	25	Python R Java
Databases/Data Management	Software packages and/or online platforms utilized for storage and access to data.	25	Oracle MySQL SQL Server
Software – Analysis	Software packages and/or online platforms utilized for the interrogation of data to generate findings and actionable conclusions.	43	Excel Qualtrics Talend
Software – Enterprise	Software packages and/or online platforms utilized for purposes related to the running and governance of the organization or business.	6	Salesforce SAP Watson
Software – Statistics	Software packages and/or online platforms utilized to conduct statistical analysis and models.	7	STATA SAS MATLAB
Software – Visualization	Software packages and/or online platforms utilized to develop visual analysis of data to inform decision-making and communication.	13	Tableau Cognos Power BI

Table 3. Tool Categories

Next, respondents were presented with the definitions of the levels of understanding from the Computing Curricula 2020: Paradigms for Global Computing Education (Table 2) and definitions of each phase from the KDDA process model (Table 1). Afterwards, respondents were presented with the following prompt:

- *For each of the following, indicate the level of understanding you would desire from the ideal entry-level data analyst. It may be that you would desire an applying understanding at some times/for some tasks and a creating understanding at other times/in other tasks; in this case, select the highest level of understanding (in this example, creating).*

Respondents had the opportunity to select one of the levels of understanding (None, Understanding, Applying, Analyzing, Evaluating, or Creating) for each of the tools identified. Respondents could hover over each level to see a brief definition as needed.

For each tool that a respondent selected "Understanding" or higher, they were asked to indicate which phase(s) of the data analytics lifecycle they expected entry-level data analysts to use that tool. Respondents could hover over each phase to see a brief definition as needed.

After each section of tools, participants were given the opportunity to enter optional free-form text to suggest missing tools or add any other details or explanations. Participants also faced an attention question as part of this sequence.

Following the identification of 1) what level of understanding was expected for each tool and 2) which phases that tool would be used in, respondents answered questions regarding:

- Skill transferability. Participants were asked to rank their agreement on a 7-point Likert scale with statements such as, "Skills are transferable across software platforms" and "Skills are transferable across language."
- Dispositions. We asked respondents to rate the desirability and importance of the dispositions identified in the Computing Curricula Task Force 2020 Paradigms (CC2020 Task Force, 2020).

5. RESULTS

A total of 68 hiring managers responded to the survey facilitated by Qualtrics. Only complete responses with correct answers to the attention check question were collected. Six responses were removed because they indicated that they expected entry-level data analysts to have some level of skill and knowledge with every single one of the 119 tools included in the survey, which seems unlikely and indicates a lack of understanding of or attention to the survey. We removed six additional responses whose answer to the prompt, "Briefly describe your organization's process for hiring an entry-level data analyst" were not relevant and did not demonstrate understanding of or attention to the survey. Finally, we removed one more response that answered the same phase to every single question as this repetition did not reflect attention to the survey. Thus, our analysis is based on 55 valid responses from hiring managers involved in the hiring of entry-level data analysts in the United States. The full survey responses are presented in the supplemental appendices; we summarize and present responses here.

5.1 Respondent Demographics

Our respondents were mostly IT Managers (25%), HR Managers (24%) or Data Analytics Managers (22%), who had been in their role for 1-5 years (36%) or 6-10 years (42%). Their companies ranged in size, but most (42%) were in the 250-999 employee bracket. Respondents represented a variety of industries. Most respondents worked in the Information (27%), Data Processing (16%), or Computer and Electronic Product Manufacturing (15%) industries. Most respondents (69%) called the position of interest a "data analyst" or an IT analyst (47%), although all job titles had some responses and many respondents indicated that they use more than one of the job titles we provided. Organizations were located primarily in urban (64%) or suburban (27%) areas across the United States (see Figure 2).

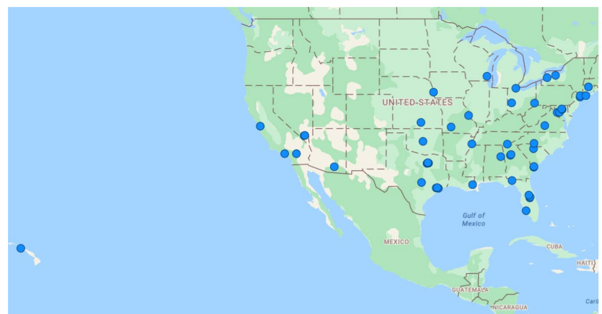


Figure 2. Geographical Locations of Survey Respondents

5.2 Which Tools Are Requested by Hiring Managers for Entry-Level Data Analysts and at Which Levels of Understanding?

5.2.1 Programming Languages. Over 90% of respondents indicated a desire for entry-level data analysts to have some level of understanding of HTML (98%), C++ (93%), Python (93%), SQL (93%), and Java (91%). For each of these, there was a wide diversity in the specific level of understanding that hiring managers requested, as shown in Appendix C. In other words, some hiring managers wanted entry-level data analysts to simply *understand* these tools, while others required a higher level of knowledge of the language. HTML and Java were the tools hiring managers wanted the highest level of understanding for hiring managers mostly wanted entry-level data analysts to be able to write code in these languages, while many others were acceptable just understanding existing code.

The least popular programming languages were the ones which respondents indicated no level of knowledge was needed for entry-level hires. These included Pig (65%), Ruby (55%), S (55%), Julia (49%), and R (49%). The inclusion of R here is particularly surprising – while still used by a little over half the hiring managers surveyed, it lags far behind Python as a favored tool by hiring managers. Of the hiring managers who did select these tools, most indicated that they would want entry-level data analysts to have an understanding level at most – only 2-3 hiring managers indicated requiring a creating level for any of these tools.

5.2.2 Enterprise Software. Because our focus was specifically on data analytics tools, we only identified six tools that fell in the category of Enterprise software. All six were selected as important by more than half of the responding hiring managers, but the most-agreed-upon platform was Google Analytics (95% of respondents indicate use). Three tools fell roughly together: Azure (89%), Salesforce and SAP (both 84%). The least popular enterprise tools were Hive and Watson. In general, hiring managers preferred an applying, analyzing, or evaluating level of understanding for these tools, although basic understanding was the preferred level for Watson, Hive, and Azure.

5.2.3 Statistics Software. Of the seven tools surveyed that are used primarily for statistics, SAS was a clear front-runner selected by 82% of hiring managers, mostly at the applying, analyzing, or evaluating levels. However, all seven of the tools were selected by over half the respondents: Matlab (71%), H2O (64%), R Studio (62%), STATA (60%), Splunk and SPSS (both 58%). R Studio had the highest count of hiring managers requiring a *creating* level of understanding, indicating that hiring managers are more likely to expect entry-level data analysts to be able to create new code and models in R Studio than other software packages.

We noted that, as a category, statistical software was least often selected by hiring managers as something they would desire from their entry-level data analysts. Thus, we can conclude that for this sample, statistical software in general was less of a concern for hiring managers.

5.2.4 Visualization Software. One visualization tool was selected by more than 90% of hiring managers: JavaScript (93%). Notably, 23% of hiring managers indicated that they would require a creating level of understanding for JavaScript.

The lowest-selected visualization tool still garnered nearly half of the survey responses: QGIS with 49% of hiring managers. Every other tool (of the 13 visualization tools surveyed) fell in the 55%-89% range (see Appendix C for full information).

This tells us that many tools are accepted in the market for data visualization. Two leaders for data visualization include Tableau and Microsoft Power BI. We found that Power BI edged out Tableau with 85% of respondents selecting some type of knowledge of the tool compared to 67% of respondents indicating a need for Tableau. Hiring managers also seemed to require a higher level of understanding of Power BI compared to Tableau, such that 13% of hiring managers requiring a creating level of understanding of Power BI compared with only 9% requiring a creating level of understanding of Tableau.

5.2.5 Analysis Software. We found a similar level of diversity in hiring managers' selections for analysis software. The least common tools (Erwin Data Modeler and KNIME) still garnered selection by 47% of hiring managers. Two other tools (Redash and Talend) garnered 49% of selections; all others were selected for some level of understanding by over half the responding hiring managers. The most popular tool, unsurprisingly, was Microsoft Excel (98%). Other widely selected tools included Amazon Web Services (95%) and Google Data Studio (89%).

Thirty-five percent of hiring managers expected entry-level data analysts to have a creating level of understanding of Microsoft Excel – the highest of any tool in our survey. Twenty percent required a creating level of understanding of Qualtrics, but generally hiring managers only required understanding, applying, or analyzing for Amazon Web Services and Google Data Studio.

5.2.6 Database Software. In contrast to the other categories of tools, we saw more agreement on popular database tools. The most selected database tool was Oracle (93%), followed closely by SQL Server and Microsoft Access (each 87%), and MySQL (84%). Eleven of the 25 database tools we surveyed on were selected by less than half of the hiring managers, including highly publicized modern tools such as Hadoop (58% of hiring managers indicating no knowledge required), Cassandra (56% indicating no knowledge needed), and Teradata (53% indicating no knowledge needed).

Of those expecting some understanding of the tools, 18-20% of hiring managers expect entry-level data analysts to have a creating level of understanding with those most widely selected database tools: Oracle, SQL Server, Microsoft Access, and MySQL. The other database tools were rarely selected for a creating level of understanding, with most falling in the understanding or applying range.

5.3 In Which KDDA Phases Are the Tools Used?

In the phases portion of the survey, hiring managers could select multiple KDDA phases for each tool (unlike the level of understanding expected, when hiring managers were asked to pick just one – the highest level they would expect). Thus, the total counts for each of these are higher than just the 55 respondents in the survey. The highest possible count for phases was 440 (that is, if a tool was selected by all 55 hiring managers and then indicated for use in each of the 8 phases, the total count would be 440). Hiring managers were only asked to indicate the phases they would expect a tool to be used if they had previously selected an option other than "none" for the level of understanding they would expect for that tool. In other words, if a hiring manager indicated that they would expect no level of understanding for a tool, they were not asked to indicate the phases when a tool would be used (since, presumably, it would not be used at all). The highest count for KDDA phases that we observed was 186 (Microsoft Excel).

5.3.1 Programming Languages. To our surprise, on average, hiring managers indicated that programming languages were used more often in the earlier phases of the KDDA (Problem Formulation, Business Understanding, and Data Understanding).

In the earlier section, we identified the top 5 programming languages that hiring managers selected for data analysts to have some level of understanding of: C++ (93% of respondents), HTML (98%), Java (91%), Python (93%), and SQL (93%). As might be expected, all these also had the highest counts for the number of phases they were expected to be used within; however, the order changed. HTML (selected by 98% of managers) was selected 122 times – notably less often than SQL (selected 151 times, but only by 93% of hiring managers). This indicates that while SQL is used by somewhat fewer organizations in our sample, those organizations that do use it expect it to be used more often throughout the KDDA phases.

Of the most popular programming languages, C++ is used most often in during the early stages of the lifecycle process, although multiple respondents listed C++ as used in every phase. This indicates that C++ is used throughout the data analytics lifecycle, which may be unexpected. Coding is frequently required to obtain relevant data to analyze; thus, programming language proficiency is required in the earlier stage of understanding data (Bharati, 2019). C++ was selected frequently by hiring managers to require some level of understanding and was also used in more phases. As shown in Figure 3, there is some variation in where programming languages were reported to be used throughout the KDDA lifecycle. We display the top few programming languages. Note that programming languages in general seem to peak in the Data Understanding phase, but C++ is used more than the other top programming languages in the Modeling and Maintenance phases. We also note an interesting peak indicating the use of SQL (but not other top programming tools) in the Evaluation phase.

5.3.2 Enterprise Tools. On average, hiring managers indicated that the enterprise tools we surveyed for are used throughout the KDDA, but most often in the Business Understanding (selected by an average of 20 hiring managers), Data Understanding (17), and Problem Formulation (15.5) phases. Enterprise software was indicated to be used least often in the Modeling (11) phase.

Google Analytics, the most popular tool selected by hiring managers in the Enterprise category, seems to be used widely in all phases except for Modeling. Azure followed this trend but with additional lower selections for the Evaluation phase. Azure was selected often in the Deployment phase, more so than other enterprise software.

These findings support a perspective that enterprise tools are used by data analysts to gain an understanding of how the business operates, where and in what format the data resides, and how problems can be formulated, with additional possible benefits of evaluating a model, deploying a solution, and maintaining that solution.

Figure 4 graphically represents the use of the top enterprise tools by phase. This reflects the reliance on enterprise tools earlier in the KDDA lifecycle (Problem Formulation, Business Understanding, and Data Understanding). We can identify a spike in the use of Azure over other enterprise tools in the Business Understanding and Deployment phases.

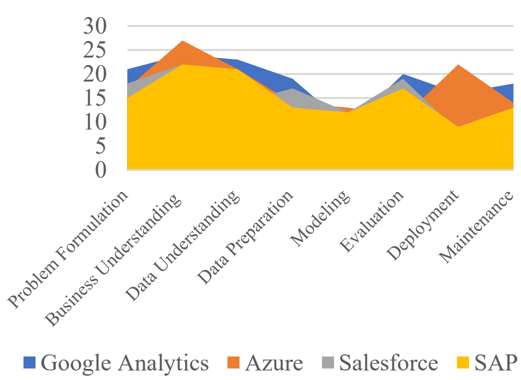


Figure 3. Top Enterprise Tools by Phase

5.3.3 Statistics Software. Perhaps surprisingly, our responding hiring managers indicated that the statistics software we surveyed for is likewise used throughout the KDDA phases. In fact, on average, hiring managers most often indicated the use of statistics tools in the Business Understanding (15) and Data Understanding (14) phases, followed closely by Problem Formulation (12), then Data Preparation, Modeling, Evaluation, and Deployment (all 11), and finally Maintenance (9). This average-level pattern is reflected for most individual tools as well, as shown in full in Appendix D. Hiring managers indicated that SAS, for example, is used heavily in all phases.

While educators might have considered statistics tools to be limited to the phases focused on preparing and modeling data and evaluating models, our survey responses indicate that data analysts should be prepared to use statistics tools throughout the KDDA phases. This may require a better understanding of the "soft" skills needed in the earlier phases to be paired with the "hard" skills of understanding and utilizing statistics software tools. Soft skills are in more demand during the early stages of the life cycle because of higher uncertainty and risk (Wood & Ellis, 2003).

In Figure 5, we display the top statistics tools by phase. Note the popularity of SAS, which shows higher usage in nearly every phase than the other statistics tools, and the general uniformity as tools are used throughout the KDDA.

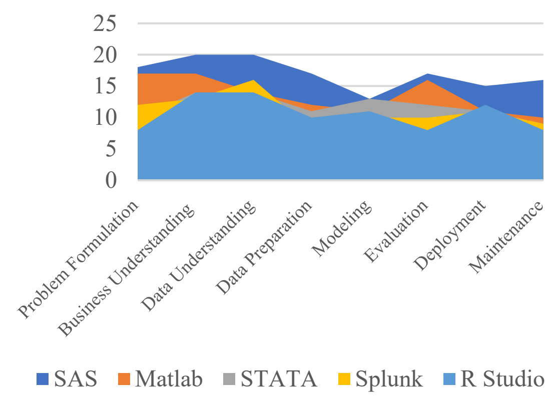


Figure 4. Top Statistics Software by Phase

5.3.4 Visualization Software. On average, respondents indicated that visualization tools are used most heavily in the Business Understanding (16) and Data Understanding (16) phases, with use in the Problem Formulation (12), Evaluation and Maintenance (both 11), Data Preparation (10) and Modeling (9.5) phases.

Interestingly, hiring managers indicated that different tools are used in different KDDA phases. For example, Tableau and Lumira were only indicated to be used in the Problem Formulation phase by 8% and 9% of respondents respectively, while 22% of respondents who had selected ArcGIS indicated it would be used for Problem Formulation. These numbers indicate the percentage of hiring managers who already selected a non-none level of understanding for the given tool in the earlier stage of the survey. For example, 37 hiring managers indicated that they would want a level of understanding other than none for Tableau; thus, here we describe that 8% of those

(3 hiring managers) subsequently indicated that they would expect Tableau to be utilized in the Problem Formulation stage.

In comparing two leading visualization tools – Tableau and Microsoft Power BI – we see very similar patterns of use throughout the KDDA with some exceptions. Hiring managers indicated that Tableau is used less often than Power BI for Problem Formulation and that Tableau is used more often than Power BI for Modeling and Deployment. This may reflect the higher level of maturity of Tableau in the workplace – making it a preferred solution for modeling and deploying actual projects with Power BI relegated to mere data exploration activities. Tableau may also be more likely to be used in the later phases of modeling and deployment because of its ability to process larger volumes of data than Power BI. At a lower price point, Power BI may be more likely to be used if the need is to only analyze limited amounts of data, contrasted with implementing the more expensive Tableau for bigger projects (Pedamkar, 2018).

In Figure 6, we can observe the popularity of data visualization tools particularly in the earlier phases of the lifecycle (Business Understanding and Data Understanding). This also serves to display interesting insights in the battle between Power BI and Tableau: note that Power BI overshadows Tableau use in Problem Formulation, Business Understanding, and Evaluation, but not necessarily in the other phases.

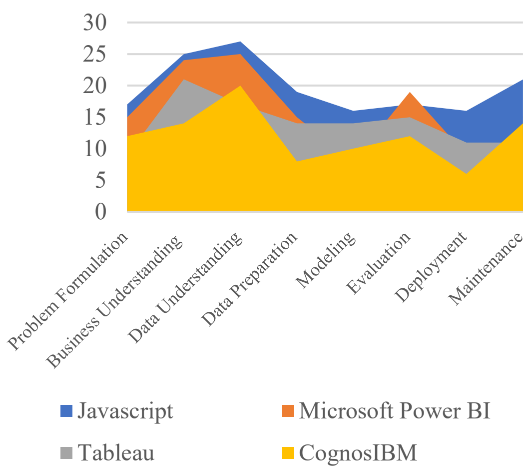


Figure 5. Top Data Visualization Software by Phase

5.3.5 Analysis Software. Analysis software was our largest category with 43 tools included. On average, hiring managers indicated that these tools were used most heavily in the earlier stages of the KDDA (Problem Formulation, Business Understanding, and Data Understanding) with 9-10% of respondents indicating use in the back half of the KDDA lifecycle (Data Preparation, Modeling, Evaluation, Deployment, and Maintenance).

Respondents indicated that the most popular tool, Microsoft Excel, was used widely throughout the KDDA – least often in the Deployment and Maintenance phases, but these phases still accounted for about 10% each of respondents who had selected Excel.

In Figure 7, we display analysis software by phase. Note the general negative trend: analysis tools are used more in the earlier phases and less as the KDDA lifecycle proceeds.

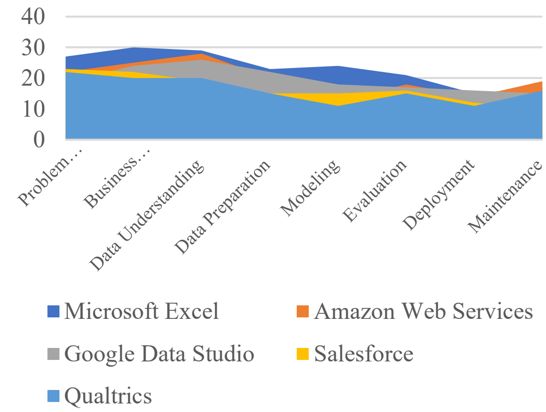


Figure 6. Analysis Software by Phase

5.3.6 Database Software. As a whole, respondents indicated that database tools would be used most frequently in the beginning of the KDDA – Problem Formulation, Business Understanding, and Data Understanding (15%, 18%, and 17%, respectively) and less frequently in Data Preparation (10%), Modeling (9%), Evaluation (11%), Deployment (10%) and Maintenance (10%).

Of the top four tools we described earlier (Oracle, Microsoft Access, SQL Server, and MySQL), all followed this similar pattern, with the highest use being reported in Business Understanding and Data Understanding, and the lowest use reported in Modeling. These findings are reflected in Figure 8, which also demonstrates the spike of use of Oracle in earlier KDDA phases.

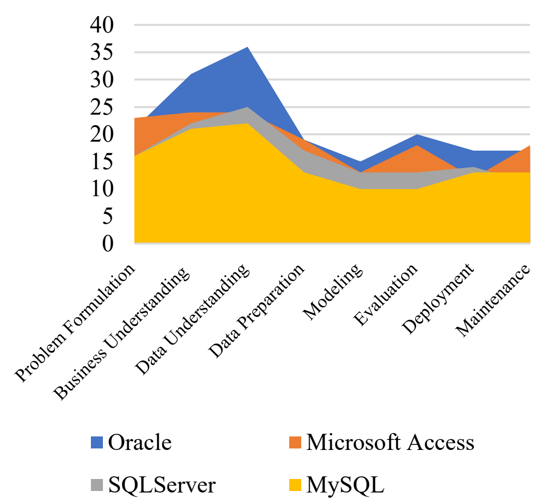


Figure 7. Database Software by Phase

5.4 How Does Tool Popularity Vary by Industry?

Tool popularity likely varies by industry. We presented respondents with a list of industries provided by the U.S.

Bureau of Labor Statistics (Industries at a Glance, n.d.) and asked them to indicate the industry in which they work. The highest percentage of respondents (27%) were affiliated with the Information industry, with other represented industries including Data Processing, Hosting, and Related Services (16%), Computer and Electronic Product Manufacturing (15%), and Professional, Scientific, and Technical Services (7%). The remaining 35% of respondents were split among 17 diverse other industries, each representing 2-4% of respondents. Descriptions and full details of the 21 industries represented by our survey respondents are presented in Table E1 in Appendix E, but this analysis will focus on the top 4 industries that each represent 7% or more of the respondents. The remaining tables in Appendix E provide the raw data of responses for each software tool broken down by industry. We interpret and summarize that raw data here. In the survey, hiring managers indicated the level of understanding they would prefer from an entry-level data analyst: none, understanding, applying, analyzing, evaluating, or creating. For portions of this analysis, we will refer to the *non-none percentage*: the percentage of hiring managers who indicated some level of understanding other than "none." We utilized a non-none percentage cutoff of 90%; in other words, if 90% of hiring managers in a given industry indicated a level of understanding other than "none" for a tool or language, we considered it desirable for that industry.

5.4.1 Programming Languages by Industry. Across the top four industries represented in our survey, we observed some similarities in preferred programming languages. These are presented in Table 4.

Programming Language	Industries (non-null percentage)
HTML	• Information (100%) • Data Processing (100%) • Professional, Scientific, and Technical Services (100%)
Python	• Information (93%) • Data Processing (100%) • Computer and Electronic Product Manufacturing (100%)
SQL	• Data Processing (100%) • Computer and Electronic Product Manufacturing (100%) • Professional, Scientific, and Technical Services (100%)
C	• Data Processing (100%) • Professional, Scientific, and Technical Services (100%)

Table 4. Desirable Programming Languages Across Industries

Some languages were highly valued by some industries but not others. These are presented in Table 5.

In summary, languages that might generally be recommended to aspiring data analysts regardless of the industry in which they plan to work include HTML, Python, SQL, and C. For data analysts planning to work in the Data Processing industry, ASP.NET and VBA would be good languages to learn, while those aiming to land in the Information industry should consider picking up C++ and Java. Those interested in the Computer and Electronic Product Manufacturing industry would do well to learn Linux.

Programming Language	Industry (non-null percentage)	Average non-null percentage of other industries
ASP.NET	Data Processing (100%)	62%
C++	Information (93%)	84%
Java	Information (93%)	80%
Linux	Computer and Electronic Product Manufacturing (100%)	73%
VBA	Data Processing (100%)	54%

Table 5. Desirable Programming Languages Between Industries

5.4.2 Enterprise Tools by Industry. Across the four top industries from our survey, there is considerable diversity in preferred enterprise software programs for entry-level data analysts. Google Analytics was preferred by hiring managers in both the Information (93%) and Data Processing (100%) industries, but it is the only tool to be preferred by more than one industry. The preferred enterprise tool for Computer and Electronics Product Manufacturing was Salesforce (100%), and no single tool stood out as a leader in the Professional, Scientific, and Technical Services industry (Azure, Google Analytics, and Salesforce all achieved a 75% non-null percentage). The Data Processing industry, on the other hand, indicated three different preferred tools: Azure, Google Analytics, and SAP.

5.4.3 Statistics Software by Industry. Across industries, hiring managers seem to be split on desired statistics software. In the Data Processing industry, hiring managers indicated SAS as a leading tool (100%) but this was closely followed by Matlab and STATA (both 89%). Hiring managers in the Information industry most often selected Matlab (87%) while those in Computer and Electronic Product Manufacturing most frequently selected SAS (88%). In the Professional, Scientific, and Technical Services industry, there was little agreement: the most popular tool was SAS, but with only a 75% non-null percentage (while Matlab, RStudio, and STATA all earned 0 votes). This may indicate that aspiring data analysts interested in entering the Professional, Scientific, and Technical Services industry should focus less on statistical software in general, while those entering Data Processing should build skills in one or more programs.

5.4.4 Visualization Software by Industry. There was little agreement across industries in our survey regarding visualization software and tools. JavaScript was preferred by both the Information and Data Processing industries (100%) but no other tools were preferred by more than one industry. Hiring managers from the Data Processing industry seemed to be particularly interested in visualization tools. Of the 13 tools we surveyed for, every single one had a non-none percentage of 78% or higher in the Data Processing industry for an overall average non-none percentage of 85%. This is an interesting contrast with the Professional, Technical, and Scientific Services industry, where 11 of the 13 tools received less than a 50% non-none percentage and the average non-none percentage was only 19%.

Comparing direct competitors Microsoft Power BI and Tableau, we can observe differing preferences across industries. Power BI is preferred in the Information industry (80% to Tableau's 47%) and the Professional, Technical, and Scientific Services industry (75% to Tableau's 25%). Tableau (100%) edges out Power BI (89%) in the Data Processing industry, and the tools are tied in Computer and Electronics Product Manufacturing (both 75%).

Overall, it may be prudent for aspiring data analysts to focus on JavaScript, but those aiming to land in the Professional, Technical, and Scientific Services industry likely need not focus on visualization as much as those interested in the Data Processing industry.

5.4.5 Analysis Software by Industry. Unsurprisingly, Microsoft Excel was preferred at 100% by three of the four top industries (Information, Data Processing, and Professional, Technical, and Scientific Services) and only just outside the 90% threshold for the Computer and Electronic Product Manufacturing industry (88%). Other analysis tools demonstrating agreement across industries were Amazon Web Services (preferred at 100% for both Information and Data Processing) and Google Data Studio (preferred at 100% for both Data Processing and Computer and Electronic Product Manufacturing).

5.4.6 Database Software by Industry. Microsoft Access (preferred by hiring managers in the Information industry and the Data Processing industry at 93% and 100% respectively) and Oracle (preferred by hiring managers in Data Processing and Computer and Electronic Product Manufacturing at 100%) were the two most popular database tools across industries. In Data Processing, four different tools (Microsoft Access, MySQL, Oracle, and SQL Server) were all preferred by 100% of hiring managers, contrasted with no tools reaching the 90% preferred threshold in Professional, Technical, and Scientific Services.

It is not surprising that the Data Processing industry would seek data analysts with skills in multiple database platforms, as is reflected in our findings. Data analysts looking to enter this industry should consider self-training in multiple database tools, while others might do better to focus their efforts on experience with Microsoft Access and Oracle. Although they did not all reach the 90% non-none preference threshold, the results of this industry analysis reflect the same findings as our overall analysis: that Microsoft Access, Oracle, SQL Server, and MySQL were generally the most popular tools across industries.

5.4.7 Summary of Tool Popularity by Industry. Due to the diversity of industries represented in our sample, we can only compare four industries; however, even across a limited number of industries, we can observe that tool popularity varies notably. In general, hiring managers in the Data Processing industry want data analysts to have some level of understanding of more tools, while those in the other industries we examined tend to land on only 1-2 specific tools per category. In the Professional, Technical, and Scientific Services industry, hiring managers tended to agree less, resulting in fewer clearly preferred tools and a wider variety of mid-rated tools. Academics can utilize this analysis to help inform decisions around tool selection for courses, particularly if their university tends to feed students into a specific industry.

5.5 Do Hiring Managers Subscribe to the Concept of Skill Transferability?

Our analysis of the four skill transferability questions led to some interesting findings. We found strong evidence that hiring managers do subscribe to the concept of skill transferability for both software platforms (mean 5.55) and languages (mean 5.45). Hiring managers did agree that they would hire someone for a role even if they did not have experience in a specific tool, if they had experience with something similar (mean 4.75) although this was weaker than their agreement with the concept of skill transferability in general. Hiring managers weakly agreed with the statement that they are willing to train candidates (mean 4.15).

The most interesting finding from the skill transferability section was that hiring managers also indicated that they use automated systems to scan resumes (mean 4.69) meaning that if a candidate does not list a specific tool, their resume will not be seen by an individual.

The skill transferability answers are summarized in Table 6 and are discussed in more depth in the Discussion.

#	Question Text	Mean	Median
1	Skills are transferable across software platforms – for example, someone experienced in Tableau can transfer their skills to Power BI.	5.55	6
2	Skills are transferable across languages – for example, someone experienced in Java can transfer their skills to Python.	5.45	5
3	I would hire someone for a role even if they didn't have experience in a specific language/tool (e.g., Tableau) as long as they had experience in something similar (e.g., Power BI).	4.75	5
4	I/we use an automated system that scans resumes, so if the required software/language isn't listed, I never see the resume.	4.69	5
5	I am unconcerned about the skills an applicant has at the time of applying because we will train them ourselves.	4.15	4

Table 6. Summary of Skill Transferability Agreement from Hiring Managers

5.6 What Dispositions Are Most Sought by Hiring Managers for Entry-Level Data Analysts?

Not surprisingly, all the dispositions we surveyed for were seen as mostly important and mostly desirable by the survey respondents. We coded the 5-point Likert scale for importance (1 = very unimportant, 5 = very important) and the 3-point scale for desirability (-1 = undesirable, 0 = neutral, 1 = desirable) and calculated averages for each. Those aggregated values are presented in Table 7.

Disposition	Average Importance *(max possible: 5.0)*	Average Desirability *(max possible: 1.0; min possible -1.0)*
Adaptable	**4.38**	0.75
Analytical	4.29	0.84
Collaborative	4.31	0.71
Communicative	4.35	0.73
Inventive	4.16	0.67
Meticulous	4.24	0.76
Passionate	4.11	0.73
Proactive	4.18	*0.64*
Professional	4.33	0.69
Purpose Driven	4.24	0.80
Responsible	4.20	**0.85**
Responsive	4.20	0.65
Self-Directed	*3.96*	0.69

*Highest values in **bold**, lowest values in *italics*

Table 7. Summary of Disposition Importance and Desirability

Although all the dispositions included in the survey were seen as important and desirable by the respondents, we did observe some variation. Respondents rated *adaptable* the most important disposition and *responsible* the most desirable. *Self-directed* was the least important disposition (though still relatively important at m = 3.96/5.00) and *proactive* was the least desirable (although again relatively desirable at an average of 0.69 on a scale from -1 to 1).

Figure 9 presents a scatter plot of average desirability and importance rankings for each disposition. Our analysis revealed a weak correlation (r = 0.24) between importance and desirability.

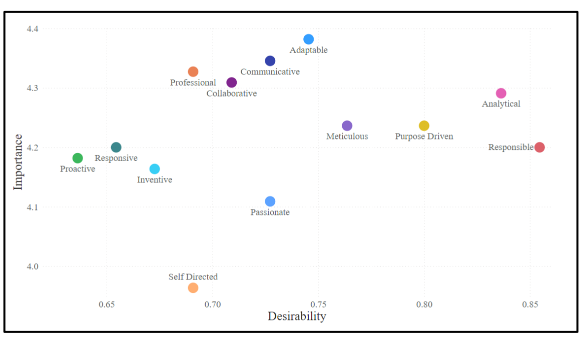


Figure 8. Scatter Plot Displaying the Average Importance and Desirability of Dispositions

While respondents chose not to participate in most of the optional open-text opportunities in the survey, the open-ended question asking for any more ideas around disposition was exercised by multiple participants. A theme that emerged was remote work. One respondent wrote, "Able to work in office, we are getting away from the hybrid setup" and another contributed, "When a candidate is able to work remotely and still deliver expectedly." Thus, it seems that in a post-COVID world, hiring managers are still being impacted by varying remote work policies and value a data analyst's ability to contribute regardless of their remote work policy. Another disposition overlooked was punctuality, which one hiring manager wrote in. Finally, one hiring manager summed up their ideas around dispositions with this: "Everyone should be judged as individuals. There isn't a one size fits all answer to hiring."

6. DISCUSSION

6.1 A Wide Variety of Tools Are Utilized in Data Analytics

Our survey included 119 data analytics tools, and we found that every single tool was selected as a desired competency by at least one organization included in our survey. This indicates that a wide variety of tools are utilized in data analytics. Thus, for scholars or students interested in a specific organization, it is likely worth the effort to investigate which tools that organization utilizes, since there exists such a wide landscape of possibilities.

6.2 There Are Clear Leading Database Tools

Unlike most of the other categories, there do seem to be clear leading tools in the database space: Oracle, SQL Server, Microsoft Access, and MySQL. Notably, these are all relational database tools.

6.3 Power BI Edges Out Tableau in Visualization

Microsoft Power BI and Tableau are accepted to be the top two data visualization tools on the market (Haan, 2023); however, disagreement exists about which will lead in the future. In our data, Microsoft Power BI edges out Tableau as the leading data visualization tool. Currently, Power BI is offered as part of the Office 365 platform that many organizations already pay for (Microsoft, 2023). This creates cost savings that may be motivating many organizations to move away from Tableau to adopt Power BI instead.

6.4 Summary of Most Essential Tools Across Categories

One contribution of our work is to identify those tools that are valued by hiring managers looking for entry-level data analysts. We summarize tools selected by 85% or more of the hiring managers in our sample and present these in Table 8. Note that none of the statistical tools were selected by 85% or more of hiring managers, so they do not appear here.

Category	Tool	% Selected
Programming Languages	HTML	98%
	C++	93%
	Python	93%
	SQL	93%
	Java	91%
	C#	87%
	C	87%
	Linux	85%
Enterprise Software	Google Analytics	95%
	Azure	89%
Visualization Software	JavaScript	93%
	Microsoft Power BI	85%
Analysis Software	Microsoft Excel	98%
	Amazon Web Services	95%
	Google Data Studio	89%
Database Software	Oracle	93%
	Microsoft Access	87%
	SQL Server	87%
	*MySQL**	*84%*

**Note: MySQL, at 84% selected, did not meet the 85%-selected threshold for this table. However, the next most-selected database tool dropped to 71% selected, so it seemed a natural break to include MySQL here.*

Table 8. Summary of Most Frequently Selected Tools

6.5 Comparison With Prior Research

One major work prior to ours is Dong and Triche (2020). As described in our literature review, Dong and Triche conducted a longitudinal analysis of trends of skills for entry-level data analysts between 2014 and 2018 based on online job postings. The authors reported which tools gained popularity over their timespan, which became less popular, and which stayed the same. We have summarized their findings in Table 9 by the major software tools they discussed, along with a summary of our own findings to contrast the works.

Our work extends Dong and Triche's work. We compare their predictions about which data analytics tools were increasing or decreasing in popularity in 2018 with the popularity based on our survey in 2022. Direct comparisons between the two works are challenging, since we used different methodologies, but we still have several interesting findings. Chief among these is the idea that software trends may not continue; that is, several tools that were declining in popularity based on the trend from 2014-2018 are listed as highly popular in 2022 (Microsoft Access, SAP) while some that were increasing in popularity based on the trend from 2014-2018 are now relatively unpopular based on our sample (Tableau, Hadoop). This may also point to a difference in source data. Dong and Triche analyzed online job postings while we surveyed hiring managers. Some tools may not be listed in job postings but are identified by hiring managers, and vice versa.

6.6 Many Tools Are Used Throughout the KDDA Lifecycle

We expected that tools would be used in specific KDDA lifecycle phases; however, our data indicates that many tools are used throughout the entire KDDA lifecycle. This offers a few important points. First, today's tools are highly complex and likely offer functionality that is useful in multiple KDDA phases. Second, individuals training to be data analysts should be prepared to participate in all phases of the KDDA lifecycle and make use of various tools throughout each phase.

6.7 Hiring Managers Believe in Skill Transferability, But Not Reflected in Software

In our questions surveying hiring managers on skill transferability, most indicated that they do believe an individual's competency in a tool (say, Tableau) will lead to competency or a lower learning curve in a similar tool (say, Microsoft Power BI). Interestingly, however, this does not mean that individuals can focus on a single tool in a category. While hiring managers subscribe to the concept of skill transferability, most also indicate that they utilize automated resume filtering tools and, as a result, if an individual does not have a specific tool listed on their resume, it may be filtered out despite listing experience with a similar tool. This leads to paradoxical suggestions for academics designing data analytics programs. On one hand, because hiring managers subscribe to the idea of skill transferability, it makes sense to train students with a deep understanding of a specific tool, knowing this understanding will transfer to different tools used by organizations. On the other hand, knowing that automated tools may reject resumes that do not list a specific software platform, academics may be inclined to try to teach a surface-level understanding of many tools. Students would have less deep knowledge but would be able to list more tools on their resume and might get farther in their job search.

6.8 Contributions to Research

Our work extends previous research on the tools and skills required for data analysts to better understand the profession. We focus on surveying hiring managers, rather than scraping job descriptions from the web, and thus gain a new perspective on this rapidly changing field.

We also utilize both the KDDA snail shell model (Li et al., 2016) and computing curricula standards (CC2020 Task Force, 2020) to contribute a model that demonstrates synthesis and integration between academic research and practical pedagogical guidelines. In this way, we produce research that is both rigorous and relevant (Glass, 2001).

6.9 Contributions to Practice

This work offers important contributions to a variety of practitioners. Our findings can be helpful to organizations looking to develop or reorganize their data analytics divisions, to understand which tools and competencies may be leaders in the current market. Our skill transferability information may also be useful to hiring managers, who may want to revamp their hiring practices to align with their values concerning skill transferability. Students, or individuals looking to enter the data analytics field, can supplement their education and training based on our findings to be more competitive in the entry-level data analyst job market.

Category (Dong & Triche, 2020)	Tool*	Dong & Triche (2020) Popularity Finding	Our Findings	Alignment With Dong & Triche
Database	SQL Server	Increasing	SQL Server was a highly popular tool in 2022, selected by 87% of hiring managers.	In accord
	Oracle	Increasing	Oracle was a highly popular tool in 2022, selected by 93% of hiring managers.	In accord
	Microsoft Access	Decreasing	We found Microsoft Access to still be a highly sought after tool in 2022, despite the report of it decreasing from 2014-2018.	Contrary
	NoSQL	Increasing	NoSQL was selected by 64% of our hiring manager sample.	Unclear
	DBMS	Increasing	DBMS was selected by 58% of our hiring manager sample.	Unclear
	MySQL	*Not included in Dong & Triche (2020)*	*MySQL was a leading database tool in our analysis (selected by 84% of hiring managers) but was not included in Dong & Triche's work.*	*N/A*
Personal Productivity	Microsoft Office	Staying the same	We did not examine Microsoft Office as a whole but did find Microsoft Excel to be the most popular tool we surveyed for.	N/A
	Microsoft PowerPoint	Staying the same		N/A
Business Intelligence	Tableau	Increasing	In our analysis in 2022, Power BI (listed as less important than Tableau by Dong and Triche) had overtaken Tableau – being selected by 85% of hiring managers compared to Tableau's 67%.	Contrary
	Cognos	Decreasing	We found that Cognos was not particularly sought after by hiring managers: selected by 65%.	In accord
	Power BI	Increasing	We found Power BI to be highly popular and sought after in 2022, selected by 85% of hiring managers.	In accord
	JavaScript	*Not included in Dong & Triche (2020)*	*While not included in Dong and Triche's analysis, we found Javascript to be the most highly sought tool in data visualization; selected by 93% of hiring managers.*	*N/A*
Programming Languages	Python	Increasing	We found Python to be particularly well sought – selected by 93% of hiring managers.	In accord
	Pig	Increasing	Pig was the least selected programing language we surveyed for, garnering selections from only 35% of hiring managers.	Contrary
	Other programming languages	*Not included in Dong & Triche (2020)*	*We found several other programming languages (HTML, SQL, C++, Java, C#, and C) to be highly sought after by our sample but not included in Dong and Triche's analysis.*	*N/A*
Enterprise System	SAP	Decreasing	SAP was selected by 84% of our hiring manager sample, indicating a continued high level of popularity for this tool.	Contrary
	Hadoop	Increasing	Hadoop was selected by only 42% of our hiring manager sample. We listed Hadoop as a database, and it was tied for the least popular database we surveyed for (tied with xsd).	Contrary
	Salesforce	Increasing	Salesforce was selected by 84% of our hiring manager sample.	In accord
	Azure	Increasing	Azure was selected by 89% of our hiring manager sample.	In accord
	Hive	Increasing	Hive was selected by 65% of the hiring managers in our sample.	Unclear
	Google Analytics	Increasing	Google Analytics was a highly popular tool in our survey – selected by 95% of hiring managers.	In accord
Statistical Packages	R	Increasing	R was selected by only 51% of the hiring managers in our sample.	Contrary
	SAS	Increasing	SAS was selected by 82% of hiring managers.	In accord
	SPSS	Increasing	SPSS was selected by only 58% of the hiring managers in our sample – the least popular statistical tool we surveyed for.	Contrary

Table 9. Comparison of Findings From Dong & Triche (2020) With Findings of the Current Work

Sorted by most popular first in 2018 by Dong & Triche (2020)

6.10 Contributions to Academia and Curriculum Development

One major contribution of our work is informing curricular development in data analytics programs. As academics formulate new data analytics programs to meet a demand that is "exploding" (Schroeder, 2021), they must wrestle with design decisions about which knowledge and skills to include. Based on our findings, we can make several recommendations.

6.10.1 Recommendation 1: Focus Less on Statistical Tools Relative to Other Tools. As a group, the statistical tools we surveyed were least selected as important by hiring managers. Thus, while statistics as basic knowledge is likely needed in a data analytics program, academics may elect to spend relatively more time on the other types of tools.

6.10.2 Recommendation 2: Select One (or More) of the Leading Database Tools: Microsoft Access, Oracle, SQL Server, or MySQL. While other database tools may bring value to specific job postings, these were overwhelmingly the most popular. Programs offering two more database classes may be well-positioned to train students in multiple of these leading tools.

6.10.3 Recommendation 3: Focus on HTML, C++, Python, SQL, and Java for Programming Methods. When possible, include HTML and/or Java in more advanced programming courses, since these are the ones hiring managers indicated they wanted the highest level of understanding in. Do not prioritize including Pig, Ruby, S, Julia, and R – unless these are known to be valued by your specific organizations hiring students.

6.10.4 Recommendation 4: Do Not Overlook Enterprise Software. Data analytics instructors may be tempted to focus exclusively on data analytics-specific software, but hiring managers indicated that general enterprise tools are also important. In particular, hiring managers value Google Analytics.

6.10.5 Recommendation 5: Include JavaScript in Visualization Curricula. When choosing between Microsoft Power BI and Tableau, Power BI may be the more popular choice (and may be included in a university's Office 365 subscription). A particularly powerful addition to the visualization portion of a data analytics program could be through the Power BI JavaScript API (Caplan, 2016).

6.10.6 Recommendation 6: Do Not Neglect Microsoft Excel. It was the most highly selected tool from our survey and represented one for which hiring managers desired the highest level of understanding. Other valuable additions to an analysis section of a data analytics program would be Amazon Web Services or Google Data Studio.

6.10.7 Recommendation 7: Prioritize Students' Learning Skills Throughout the KDDA Using a Variety of Tools. While it may be tempting to focus on, for example, modeling data in analysis tools, our survey shows that many tools are used throughout the KDDA. Plan to set aside some class time to investigate how tools are used for everything from problem formulation to deployment whenever possible.

6.10.8 Recommendation 8: Make a Strategic Decision Regarding the Number of Software Tools Covered in Courses and the Depth of Understanding; Recommend Student-Driven Additional Learning. This is a challenge for academics, since hiring managers indicated that they believed in skill transferability, but also that automated hiring software might not pass through a resume that did not mention a specific tool. Academics, then, are tasked with deciding whether to provide students a deeper understanding of a specific tool (knowing that, if the tool is not one selected by the organization, the student may be passed over) or a shallower understanding of several tools (knowing that the student may have a higher chance at an interview but overall a weaker understanding, since more time will be devoted to installing, setting up, and learning the basics of multiple tools). Simultaneously, academics face the challenge of ensuring that students receive a deep and meaningful education – a goal that can be thwarted when a course favors simple tool proficiency over intellectual understanding of theory and concepts. Thus, our recommendation is that instructors consider the strategy that best aligns with their program and suggest student-driven learning to gain additional tool proficiency as needed.

6.10.9 Recommendation 9: Consider the Industries Aspiring Data Analysts Want to Enter When Selecting Tools to Support Coursework. Our findings indicate that tool preference varies depending on the hiring manager's industry. Although the limitations of our data preclude us from providing specific recommendations for all industries, we report on tool preference on a few industries and encourage academics to seek guidance from industry partners as they design their own curricula.

6.10.10 Recommendation 10: Find Opportunities for Students to Demonstrate Skills Such as Being Analytical, Responsible, Adaptable, Communicative, Collaborative, and Professional. Place less emphasis on opportunities to demonstrate being self-directed and passionate.

6.11 Limitations and Future Research

Our research has several limitations that should be considered. First, we focused only on organizations operating mostly or fully in the United States; thus, future research may take a more global view to understand the landscape in other countries. We elected to ask only about the competencies expected of entry-level data analysts. Organizations may have different expectations and requirements for experienced hires or for individuals entering different areas of information systems; future research may examine this.

We utilized a survey data collection method, causing some limitations in the conclusions we can draw. First, some tools included in the survey could have been examined at a more detailed level of analysis. For example, some tools like Tibco and SAS encompass multiple platforms that could have been presented individually. However, our survey was already lengthy, and we chose to mitigate survey fatigue by asking about some higher-level platforms where it was reasonable to do so. We also utilized loosely formed, non-academic categories to group tools and reduce survey fatigue. Future research may take on the challenging task of developing a research-based taxonomy or typology of data analytic tools. In the meantime, all raw data is provided in the appendices so interested scholars and practitioners can rearrange tools among categories as they see fit.

One of our conclusions was that many tools are used throughout the KDDA lifecycle; however, our data collection methodology may have obscured the use of specific tools in specific phases by specific organizations. For example, if Organization A uses Tableau mostly for Modeling and a little for Problem Formulation and Organization B uses Tableau mostly for Problem Formulation and a little in Modeling, our data would present this as moderate use in both phases. Future research may endeavor to resolve this problem with more detailed analyses.

There are several other interesting avenues for future research. Scholars may examine how requirements for data analysts vary by geographical location or how data analytics requirements differ for entry-level professionals in other roles (e.g., accountants, supply chain managers, and other business school professionals). Future research may also contrast the competencies required for data analysts with other IS roles such as DBAs, software engineers, or cybersecurity specialists. Our analysis stopped at the point that entry-level data analysts are hired, but future research could examine whether specific competencies correlate with higher salaries and/or more career longevity. Another potential area of future work may examine fluctuations within the data analyst role – e.g., which skills are associated with higher salaries within the role, thereby extending work investigating the determinants of starting salary of more general IT graduates (Ge et al., 2015). Finally, future research may examine specific challenges that data analysts face in the workplace such as data security, incomplete data, ethical concerns, and ambiguous results.

Our research focused exclusively on the view of the hiring manager. Future research may examine which competencies are considered most important by those who recently hired into entry-level data analyst positions and/or by those developing data analytics programs in universities. Our survey data extends previous studies that have web scraped online job postings – an interesting avenue for future research may also be a comparison of our findings with those competencies most frequently listed in online job postings. Do online job postings match the values presented by hiring managers here? Related to this is the paradoxical dilemma of using automated resume scanning software while acknowledging the importance of skill transferability. An interesting area of future research can investigate how hiring managers solve this contradiction. At the same time, a researcher may examine the impact that AI will have on the need to learn specific tools and how much AI will impact the need for skill transferability.

Finally, one additional perspective for future research is from the academy. Scholars may conduct reviews of the data analytics programs offered and synthesize which competencies are valued based on the curricula available (similar to the work done by Saltz et al., 2018, on the data science field). There are also many ways for individuals to gain the competencies we have examined here. Future research should investigate which of those ways – formal education, online bootcamps, and certifications – are most valued by industry as demonstrations of the required competencies and the different pedagogical approaches educators and trainers may use to teach data analyst skill sets.

7. CONCLUSION

This study examined 55 survey responses from hiring managers involved in the hiring of entry-level data analysts in the United States. Based on a theoretical model integrating the KDDA (Li et al., 2016) with Computing Curricula guidelines (CC2020 Task Force, 2020) and an extensive search for modern data analysis tools (see Appendix A), we developed a survey involving 119 languages, tools, and platforms that may be used by entry-level data analysts. We asked about all three components of a competency: knowledge, skill, and disposition. Our results revealed clear preference for skills and knowledge in specific database tools (Oracle, MySQL, SQL Server, and Microsoft Access) and for specific analysis tools (Microsoft Excel, Amazon Web Services, and Google Data Studio), but a wide variety of preferred tools in other areas (statistics and visualization). We found that tools tend to be used throughout the KDDA lifecycle and that hiring managers value adaptability and responsibility in entry-level data analysts while being self-directed and proactive were valued less. Although hiring managers subscribe to the idea of skill transferability, automated hiring software may require specific tools to be listed on resumes. This work contributes to research by testing the KDDA lifecycle model and laying the foundation for future pedagogical research on the under-studied but timely area of data analytics education. The work also contributes to pedagogical practice for individuals developing and maintaining data analytics programs at various levels of education. Finally, it provides interesting information for data analytics practitioners who may desire to compare their own expectations of entry-level data analysts with those practitioners from our survey.

8. REFERENCES

2021 Training Industry Report. (2021, November 19). *Training Magazine*, 21-36.

AACSB International. (2021). *2020 Guiding Principles and Standards for Business Accreditation Standards* (p. 59). https://www.aacsb.edu/-/media/documents/accreditation/2020-aacsb-business-accreditation-standards-july-2021.pdf?rev=7f4c2893dc1e47eb91c472d9fc59b238&hash=833E7A4A1E094BADDACDAAB60CF2CD69

Assyne, N., Ghanbari, H., & Pulkkinen, M. (2022). The Essential Competencies of Software Professionals: A Unified Competence Framework. *Information and Software Technology*, 151, 107020. https://doi.org/10.1016/j.infsof.2022.107020

Bharati, R. (2019, September 29). *Explore the Role of Coding in Data Science & Analytics*. https://junilearning.com/blog/guide/coding-and-data-science/

Caplan, J. (2016, September 1). *Introducing the New Power BI JavaScript API*. Microsoft Power BI Blog. https://powerbi.microsoft.com/en-us/blog/intro-pbi-js-api/

CC2020 Task Force. (2020). *Computing Curricula 2020: Paradigms for Global Computing Education*. ACM. https://doi.org/10.1145/3467967

Clayton, P. R., & Clopton, J. (2019). Business Curriculum Redesign: Integrating Data Analytics. *Journal of Education* *for Business*, 94(1), 57-63. https://doi.org/10.1080/08832323.2018.1502142

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., García-Peñalvo, F. J., & Tovar-Caro, E. (2013). Competence Gaps in Software Personnel: A Multi-Organizational Study. *Computers in Human Behavior*, 29(2), 456-461. https://doi.org/10.1016/j.chb.2012.04.021

Cummings, J., & Janicki, T. (2021). Survey of Technology and Skills in Demand: 2020 Update. *Journal of Information Systems Education*, 32(2), 150-159.

Dong, T., & Triche, J. (2020). A Longitudinal Analysis of Job Skills for Entry-Level Data Analysts. *Journal of Information Systems Education*, 31(4), 312-326.

Dubey, R. S., & Tiwari, V. (2020). Operationalisation of Soft Skill Attributes and Determining the Existing Gap in Novice ICT Professionals. *International Journal of Information Management*, 50, 375-386. https://doi.org/10.1016/j.ijinfomgt.2019.09.006

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy. *International Journal of Information Management*, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Elrod, C. C., Stanley, S. M., Hilgers, M. G., & Graham, C. (2022). Management Information Systems Education: A Systematic Review. *Journal of Information Systems Education*, 33(4), 357-370.

Ge, C., Kankanhalli, A., & Huang, K.-W. (2015). Investigating the Determinants of Starting Salary of IT Graduates. *The Data Base for Advances in Information Systems*, 46(4), 9-25. https://doi.org/10.1145/2843824.2843826

Glass, R. L. (2001). Rigor vs. Relevance: A Practitioner's Eye View of an Explosion of IS Opinions. *Communications of the Association for Information Systems*, 6. https://doi.org/10.17705/1CAIS.00602

Goles, T., Hawk, S., & Kaiser, K. M. (2008). Information Technology Workforce Skills: The Software and IT Services Provider Perspective. *Information Systems Frontiers*, 10(2), 179-194. https://doi.org/10.1007/s10796-008-9072-9

Gupta, B., Goul, M., & Dinter, B. (2015). Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs. *Communications of the Association for Information Systems*, 36, 449-476. https://doi.org/10.17705/1CAIS.03623

Haan, K. (2023, May 22). *The Best Data Visualization Tools Of 2023 – Forbes Advisor*. Forbes Advisor. https://www.forbes.com/advisor/business/software/best-data-visualization-tools/

Huitt, W. (2011). *Bloom et al.'s Taxonomy of the Cognitive Domain*. Educational Psychology Interactive. http://www.edpsycinteractive.org/topics/cogsys/bloom.html

Industries at a Glance: Alphabetical Index : U.S. Bureau of Labor Statistics. (n.d.). https://www.bls.gov/iag/tgs/iag_index_alpha.htm

LeClair, D. (2013, February 7). *Analytics Demand Reshaping Business Education | AACSB* [Blog]. AACSB Insights. https://www.aacsb.edu/insights/articles/2018/02/an-insatiable-demand-for-analytics-talent-is-reshaping-business-education

Lee, C. K. (2005). Transferability of Skills Over the IT Career Path. *Proceedings of the 2005 ACM SIGMIS CPR Conference on Computer Personnel Research* (pp. 85-93).

Lee, C. K. (2008). Analysis of Skills Requirement for Entry-Level Programmer/Analysts in Fortune 500 Corporations. *Journal of Information Systems Education*, 19(1), 17-27.

Leidig, P. M., & Salmela, H. (2022). The ACM/AIS IS2020 Competency Model for Undergraduate Programs in Information Systems: A Joint ACM/AIS Task Force Report. *Communications of the Association for Information Systems*, 50(1), 470-480. https://doi.org/10.17705/1CAIS.05021

Li, L. (2022). Reskilling and Upskilling the Future-Ready Workforce for Industry 4.0 and Beyond. *Information Systems Frontiers*, 1-16. https://doi.org/10.1007/s10796-022-10308-y

Li, Y., Thomas, M. A., & Osei-Bryson, K.-M. (2016). A Snail Shell Process Model for Knowledge Discovery via Data Analytics. *Decision Support Systems*, 91, 1-12. https://doi.org/10.1016/j.dss.2016.07.003

Mandviwalla, M., Dinger, M., Anderson, B., Wyndetryst Graphic Design Studio, & Allegra, J. (2022). *Information Systems Job Index* (p. 24). Institute for Business and Information Technology, Fox School of Business, Temple University.

Microsoft. (2023). *Pricing & Product Comparison | Microsoft Power BI*. https://powerbi.microsoft.com/en-us/pricing/

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A Method for Taxonomy Development and Its Application in Information Systems. *European Journal of Information Systems*, 22(3), 336-359. https://doi.org/10.1057/ejis.2012.26

Pedamkar, P. (2018, May 25). Power BI vs. Tableau | 7 Most Valuable Differences You Should Know. *EDUCBA*. https://www.educba.com/power-bi-vs-tableau/

Plice, R. K., & Reinig, B. A. (2007). Aligning the Information Systems Curriculum With the Needs of Industry and Graduates. *Journal of Computer Information Systems*, 48(1), 22-30.

Qiu, M., Xu, Y., & Omojokun, E. O. (2020). To Close the Skills Gap, Technology and Higher-Order Thinking Skills Must Go Hand in Hand. *Journal of International Technology and Information Management*, 29(1), 98-123. https://doi.org/10.58729/1941-6679.1444

Sahin, Y. G., & Celikkan, U. (2020). Information Technology Asymmetry and Gaps Between Higher Education Institutions and Industry. *Journal of Information Technology Education: Research*, 19, 339-365. https://doi.org/10.28945/4553

Saltz, J., Armour, F., & Sharda, R. (2018). Data Science Roles and the Types of Data Science Programs. *Communications of the Association for Information Systems*, 43, 615-624. https://doi.org/10.17705/1CAIS.04333

Scherer, R. (2016). Learning From the Past–The Need for Empirical Evidence on the Transfer Effects of Computer Programming Skills. *Frontiers in Psychology*, 7, 1-4. https://doi.org/10.3389/fpsyg.2016.01390

Scholtz, J. C. (1993). A Longitudinal Study of Transfer Between Programming Languages. In *People and Computers VIII* (Vol. 7, p. 397). Cambridge University Press.

Schroeder, B. (2021, June 11). *The Data Analytics Profession and Employment Is Exploding—Three Trends That Matter*. Forbes. https://www.forbes.com/sites/bernhardschroeder/2021/06/11/the-data-analytics-profession-and-employment-is-exploding-three-trends-that-matter/

Setor, T. K., & Joseph, D. (2021). College-Based Career Interventions: Raising IT Employability and Persistence in Early Careers of IT Professionals. *Journal of Information Systems Education*, 32(4), 262-273.

The Best Business Analytics Programs, Ranked. (n.d.). https://www.usnews.com/best-graduate-schools/top-business-schools/business-analytics-rankings

Verma, A., Yurov, K. M., Lane, P. L., & Yurova, Y. V. (2019). An Investigation of Skill Requirements for Business and Data Analytics Positions: A Content Analysis of Job Advertisements. *Journal of Education for Business*, 94(4), 243-250. https://doi.org/10.1080/08832323.2018.1520685

Wood, G. D., & Ellis, R. C. T. (2003). Risk Management Practices of Leading UK Cost Consultants. *Engineering, Construction and Architectural Management*, 10(4), 254-262.

AUTHOR BIOGRAPHIES

Cassandra "Cassie" Collier is an assistant professor in the Computer Management and Information Systems department at Southern Illinois University Edwardsville. She received her Ph.D. in Business Administration with a specialization in Management Information Systems from the University of Houston after a corporate career focusing on data analytics in the oil and gas industry. Her research focuses on online decision-making, specifically nudges, and information systems pedagogy, particularly in the realm of data analytics and data visualization.

Anne Powell is a professor in the Computer Management and Information Systems department at Southern Illinois University Edwardsville. She received both her MBA and Ph.D. in Management of Information Systems from Indiana University. She has ten years corporate experience as a systems analyst and business analyst. Her research interests include virtual teams, the impact of new technology on individuals, teams, and organizations, and technology-enhanced education. Her work has been presented at many conferences and has been published in numerous Information Systems academic journals.

APPENDICES

Appendix A. Selecting Tools to Include in The Survey

To build the master set of tools used to construct our survey of hiring managers, we draw from two academic sources (Dong & Triche, 2020; Verma et al., 2019) and numerous practitioner sources. We identified practitioner sources by running three Google searches: "top programming languages in data analytics," "top software in data analytics," and "top tools for data analytics" and compiling each tool mentioned in the top five results for each search (excluding results that were sponsored or advertisements). There were some duplicates (e.g., sites that were in the top 5 results for both "top software in data analytics" and "top programming tools in data analytics"), in which case we took the top 5 non-duplicated results from each search. This resulted in a total of 15 practitioner sources, listed and described below.

Search	Source Title	URL
"top programming languages in data analytics"	CareerKarma.com: "Most Popular Programming Languages for Data Analysis"	https://careerkarma.com/blog/best-programming-languages-for-data-analysis/#:~:text=The%20most%20popular%20programming%20languages,statistical%20computing%2C%20and%20statistical%20analysis.
	DataCamp.com: "Top Programming Languages for Data Scientists in 2022"	https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022
	FlatIron School: "10 Best Data Science Programming Languages"	https://flatironschool.com/blog/data-science-programming-languages/
	edX Blog: "9 Top Programming Languages for Data Science"	https://blog.edx.org/9-top-programming-languages-for-data-science
	Analytics Insight: "10 Best Data Science Programming Languages for Data Aspirants in 2021"	https://www.analyticsinsight.net/10-best-data-science-programming-languages-for-data-aspirants-in-2021/
"top software in data analytics"	Datamation: "Top Data Analytics Tools for 2022"	https://www.datamation.com/big-data/data-analytics-tools/
	Solutions Review: "The 28 Best Data Analytics Software Tools for 2022"	https://solutionsreview.com/business-intelligence/the-best-data-analytics-software-and-top-tools/
	Forbes: "Best Data Analytics Tools & Software (2022)"*	https://www.forbes.com/advisor/business/software/best-data-analytics-tools/
	Datapine: "Essential Data Analyst Tools: Discover a List of the 17 Best Data Analysis Software & Tools on the Market"*	https://www.datapine.com/articles/data-analyst-tools-software
	Stitch: "Top 24 Tools for Data Analysis and How to Decide Between Them"*	https://www.stitchdata.com/resources/data-analysis-tools/
"top tools for data analytics"	Edureka: "Top 10 Data Analytics Tools You Need to Know in 2023"	https://www.edureka.co/blog/top-10-data-analytics-tools/
	Hackr.io: "7 Top Data Analytics Tools to Use in 2022"	https://hackr.io/blog/top-data-analytics-tools
	CareerFoundry: "The 9 Best Data Analytics Tools for Data Analysts in 2023"	https://careerfoundry.com/en/blog/data-analytics/data-analytics-tools/
	Hevo: "25 Best Data Analysis Tools in 2022"	https://hevodata.com/learn/data-analysis-tools/
	MonkeyLearn: "Top 15 Data Analysis Tools for Managing Data Like a Pro"	https://monkeylearn.com/blog/data-analysis-tools/

*Also returned as results for the search "top tools for data analytics"

Table A1. Sources Used to Compile List of Languages and Tools Relevant to Data Analytics

There was considerable overlap and repetition among the sources. Thus, although each of the 17 sources listed 7-60 different languages or tools, we identified a total of 120 unique languages or tools. We loosely grouped these into the following categories: Programming Languages, Database System/Data Management, Software – Enterprise, Visualization, and Statistics, Software – Analysis. These categories were used to break up tools in the survey to reduce survey fatigue and to structure results and understanding of survey responses. Categories were based on categories within the sources used to identify the tools and were agreed upon by both authors. A full list of the tools and languages identified is included next.

#	Tool/Language	Source(s)	Category
1.	Cassandra	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
2.	DB2	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
3.	DBMS	Dong & Triche, 2020	Database System/Data Management
4.	Flume	Verma et al. 2019	Database System/Data Management
5.	Hadoop	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
6.	Hbase	Dong & Triche, 2020	Database System/Data Management
7.	Mahoot	Verma et al. 2019	Database System/Data Management
8.	MapReduce	Verma et al. 2019	Database System/Data Management
9.	Microsoft Access	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
10.	MongoDB	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
11.	MySQL	Dong & Triche, 2020 Verma et al. 2019 Datapine	Database System/Data Management
12.	NoSQL	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
13.	Oozie	Verma et al. 2019	Database System/Data Management
14.	Oracle (incl. Oracle Analytics, Oracle Analytics Cloud)	Dong & Triche, 2020 Verma et al. 2019 SolutionsReview StitchData Hevo Data	Database System/Data Management
15.	PostgreSQL	Verma et al. 2019	Database System/Data Management
16.	Presto	Verma et al. 2019	Database System/Data Management
17.	shark	Verma et al. 2019	Database System/Data Management
18.	Spark (Apache)	Verma et al. 2019 Datapine Edureka Hacker.io CareerFoundry Hevo Data	Database System/Data Management
19.	SQL Server	Dong & Triche, 2020 Verma et al. 2019	Database System/Data Management
20.	Teradata	Dong & Triche, 2020	Database System/Data Management
21.	tsql	Dong & Triche, 2020	Database System/Data Management
22.	XML	Dong & Triche, 2020	Database System/Data Management
23.	xsd	Dong & Triche, 2020	Database System/Data Management
24.	xsl	Dong & Triche, 2020	Database System/Data Management
25.	zookeeper	Verma et al. 2019	Database System/Data Management
26.	ASP.NET	Verma et al. 2019	Programming Language
27.	BASH	Verma et al. 2019	Programming Language
28.	C	Dong & Triche, 2020 Verma et al. 2019 DataCamp.com Flatiron School	Programming Language

		edX Blog AnalyticsInsight	
29.	C#	Verma et al. 2019	Programming Language
30.	C++	Dong & Triche, 2020 Verma et al. 2019 DataCamp.com Flatiron School edX Blog AnalyticsInsight	Programming Language
31.	COBOL	Verma et al. 2019	Programming Language
32.	FORTRAN	Verma et al. 2019	Programming Language
33.	Go	DataCamp.com	Programming Language
34.	HTML	Dong & Triche, 2020	Programming Language
35.	Java	Verma et al. 2019 CareerKarma.com DataCamp.com Flatiron School AnalyticsInsight	Programming Language
36.	JBOSS	Verma et al. 2019	Programming Language
37.	JQUERY	Verma et al. 2019	Programming Language
38.	Julia	DataCamp.com Flatiron School edX Blog AnalyticsInsight	Programming Language
39.	Linux	Dong & Triche, 2020	Programming Language
40.	Pearl/Perl	Dong & Triche, 2020 Verma et al. 2019	Programming Language
41.	Pig	Dong & Triche, 2020 Verma et al. 2019	Programming Language
42.	Python	Dong & Triche, 2020 Verma et al. 2019 CareerKarma.com DataCamp.com Flatiron School edX Blog AnalyticsInsight Datapine StitchData Edureka Hacker.io CareerFoundry Hevo Data MonkeyLearn	Programming Language
43.	Ruby	Dong & Triche, 2020	Programming Language
44.	S	Verma et al. 2019	Programming Language
45.	Scala	Verma et al. 2019 CareerKarma.com DataCamp.com Flatiron School edX Blog AnalyticsInsight	Programming Language
46.	SPLUS	Verma et al. 2019	Programming Language
47.	SQL	Verma et al. 2019 CareerKarma.com DataCamp.com	Programming Language

		Flatiron School edX Blog AnalyticsInsight MonkeyLearn	
48.	Swift	DataCamp.com	Programming Language
49.	VBA	Dong & Triche, 2020	Programming Language
50.	Visual Basic	Dong & Triche, 2020 Verma et al. 2019	Programming Language
51.	AirTable	MonkeyLearn	Software - Analysis
52.	Altair	SolutionsReview	Software - Analysis
53.	Alteryx	SolutionsReview	Software - Analysis
54.	Amazon Web Services	SolutionsReview	Software - Analysis
55.	AnswerRocket	SolutionsReview	Software - Analysis
56.	Board	SolutionsReview	Software - Analysis
57.	Chartio	StitchData	Software - Analysis
58.	ClicData	MonkeyLearn	Software - Analysis
59.	Datapine	Datapine	Software - Analysis
60.	Domo	SolutionsReview Forbes StitchData	Software - Analysis
61.	Erwin Data Modeler	Datapine	Software - Analysis
62.	Google Data Studio	StitchData Hevo Data	Software - Analysis
63.	Hitachi Vantara	SolutionsReview	Software - Analysis
64.	Incorta	SolutionsReview	Software - Analysis
65.	Info Birst	SolutionsReview	Software - Analysis
66.	Jenkins	Datapine	Software - Analysis
67.	Jupyter Notebook	StitchData CareerFoundry Hevo Data	Software - Analysis
68.	Klipfolio	Forbes	Software - Analysis
69.	KNIME	Edureka CareerFoundry MonkeyLearn	Software - Analysis
70.	Looker	SolutionsReview Forbes StitchData Hevo Data MonkeyLearn	Software - Analysis
71.	Metabase	StitchData Hevo Data	Software - Analysis
72.	MicroStrategy	Datamation SolutionsReview	Software - Analysis
73.	Mode	StitchData	Software – Analysis
74.	MonkeyLearn	MonkeyLearn	Software - Analysis
75.	MS Excel	Dong & Triche, 2020 Verma et al. 2019 edX Blog Datapine StitchData	Software - Analysis

		Edureka Hacker.io CareerFoundry Hevo Data MonkeyLearn	
76.	OpenRefine	Datapine	Software - Analysis
77.	Periscope Data	StitchData	Software - Analysis
78.	Pyramid Analytics	SolutionsReview	Software - Analysis
79.	Qlik/QlikView/QlikSense	Datamation SolutionsReview Forbes StitchData Edureka Hevo Data MonkeyLearn	Software - Analysis
80.	Qualtrics	Datapine	Software - Analysis
81.	Query.me	Hevo Data	Software - Analysis
82.	RapidMiner	Datapine StitchData Edureka Hevo Data MonkeyLearn	Software - Analysis
83.	Redash	StitchData	Software - Analysis
84.	Salesforce	SolutionsReview	Software - Analysis
85.	Sigma Computing	SolutionsReview	Software - Analysis
86.	Sisense	Datamation SolutionsReview StitchData Hevo Data	Software - Analysis
87.	Talend	Datapine Edureka MonkeyLearn	Software - Analysis
88.	Targit	SolutionsReview	Software - Analysis
89.	Tellius	SolutionsReview	Software - Analysis
90.	Thoughtspot	Datamation SolutionsReview StitchData Hevo Data	Software - Analysis
91.	Whatagraph	Hevo Data	Software - Analysis
92.	Yellowfin BI	SolutionsReview	Software - Analysis
93.	Zoho	SolutionsReview Forbes Hevo Data	Software - Analysis
94.	Azure	Dong & Triche, 2020	Software - Enterprise
95.	Google Analytics	Dong & Triche, 2020 Verma et al. 2019 Hevo Data	Software - Enterprise
96.	Hive	Dong & Triche, 2020 Verma et al. 2019	Software - Enterprise
97.	Salesforce	Dong & Triche, 2020	Software - Enterprise
98.	SAP (and SAP Analytics Cloud)	Dong & Triche, 2020 Datamation SolutionsReview	Software - Enterprise

		StitchData Hevo Data	
99.	Watson	Dong & Triche, 2020	Software - Enterprise
100.	H2o	Verma et al. 2019	Software - Statistics
101.	Matlab	Verma et al. 2019 DataCamp.com Flatiron School edX Blog AnalyticsInsight	Software - Statistics
102.	R Studio	Dong & Triche, 2020 Verma et al. 2019 CareerKarma.com DataCamp.com Flatiron School edX Blog AnalyticsInsight Datapine StitchData Edureka Hacker.io CareerFoundry Hevo Data MonkeyLearn	Software - Statistics
103.	SAS (SAS Forecasting, SAS Business Intelligence)	Dong & Triche, 2020 Verma et al. 2019 DataCamp.com Flatiron School AnalyticsInsight Datamation SolutionsReview Datapine StitchData Hacker.io CareerFoundry Hevo Data MonkeyLearn	Software - Statistics
104.	Splunk	Verma et al. 2019 Edureka Hevo Data	Software - Statistics
105.	SPSS	Dong & Triche, 2020 Verma et al. 2019	Software - Statistics
106.	STATA	Dong & Triche, 2020 Verma et al. 2019	Software - Statistics
107.	ArcGIS	Verma et al. 2019	Software - Visualization
108.	Cognos (IBM)	Dong & Triche, 2020 Datamation SolutionsReview StitchData Hevo Data	Software - Visualization
109.	Crystal Reports	Verma et al. 2019	Software - Visualization
110.	fixml	Dong & Triche, 2020	Software - Visualization
111.	GIS	Verma et al. 2019	Software - Visualization
112.	HighCharts	Datapine Hevo Data	Software - Visualization
113.	Javascript/Javascript D3	Dong & Triche, 2020 Verma et al. 2019 DataCamp.com Flatiron School	Software - Visualization

		edX Blog AnalyticsInsight	
114.	Lumira	Verma et al. 2019	Software - Visualization
115.	Microsoft Power BI	Dong & Triche, 2020 Datamation SolutionsReview Forbes StitchData Edureka Hacker.io CareerFoundry Hevo Data MonkeyLearn	Software - Visualization
116.	Pentaho	Dong & Triche, 2020	Software - Visualization
117.	QGIS	Verma et al. 2019	Software - Visualization
118.	Tableau	Dong & Triche, 2020 Verma et al. 2019 Datamation SolutionsReview Forbes StitchData Edureka Hacker.io CareerFoundry Hevo Data MonkeyLearn	Software - Visualization
119.	TIBCO (incl Spotfire)	Verma et al. 2019 Datamation SolutionsReview StitchData Hevo Data	Software - Visualization

Table A1. List of All Tools and Languages Included in the Survey, Their Categories and the Sources That Suggested Them

Appendix B. Hiring Manager Survey Questions

Category	Questions
Screening Questions	1. Are you involved in the hiring of entry-level data analysts? (Yes/No) 2. Is your organization located in the United States, such that most or all data analysts at the company work, live, and were educated in the U.S.? (Yes/No)
Demographics	1. What is your job title? (radio buttons of options including HR Manager, Recruiting Manager, etc. with an Other/open text field) 2. What is the name of the department in which your data analysts work? (radio buttons with IT, IS, etc. with an Other/open text field) 3. Some organizations may have slightly different terminology for "data analysts." In your organization, what is the title for an individual who works with data, including but not limited to data procurement, acquisition, cleansing, wrangling, analysis, and visualization? If you have multiple descriptors for this individual, check all that apply. (radio buttons with data analyst, analytics analyst, etc. with an Other/open text field). 4. How long have you been in this role in which you are involved with hiring entry-level data analysts? (radio buttons with less than 1 year, 1-5 years, etc. up to Over 20 years). 5. How long have you been working in general, in any role? (radio buttons with less than 1 year, 1-5 years, etc. up to Over 20 years). 6. How big is your company? (radio buttons with 10 or fewer employees, 10-49 employees, 50-249 employees, 250-999 employees, 1000-4999 employees, 5000-9999 employees, and 10,000 or more employees). 7. What industry is your company in? (drop down field with industries from (*Industries at a Glance*, n.d.) 8. Where is your organization located? Select all states in which you have offices where a data analyst may work, as well as the state in which your organization is headquartered. (Checkboxes with all U.S. states and territories as well as an option for all data analyst jobs being remote). 9. Describe the location of your company's headquarters or main offices. Check all that apply. (Checkboxes – urban, rural, suburban, headquartered outside the U.S., Other/open text field). 10. Briefly describe your organization's process for hiring an entry-level data analyst and your involvement in it. Consider the process end-to-end: conception of the job description, posting/marketing of the position, collecting applications, evaluating applications, scheduling and conducting multiple rounds of interviews, and making an official offer. Be sure to include what steps are automated or completed using software vs. those steps completed by individuals. (Open text field).
Directions	1. For the following questions, consider what you would like to see from your **ideal** entry-level data analyst candidate. There are two aspects to consider for the following questions: the level of understanding and the phase of the data analytics lifecycle in which the competencies are used. The levels of understanding are defined as follows (these definitions will be available to refer back to in the coming questions): 0. None: the individual has no knowledge or understanding of the material. 1. Remembering: the individual can **recall** facts, terms, and basic concepts of previously learned material. 2. Understanding: the individual can **organize, compare, translate, interpret, and describe** the material 3. Applying: the individual can **solve problems** in new situations by applying acquired knowledge, facts, techniques, and rules in a different way 4. Analyzing: the individual can **examine and break information into parts** by identifying motives or causes and can **make inferences and find evidence** to support solutions 5. Evaluating: the individual can **present and defend opinions** by making judgments about information, the validity of ideas, or quality of material 6. Creating: the individual can **compile information** together in a different way by combining elements in a new pattern or by **proposing alternative solutions.** Each level builds on the one before it; that is, an individual capable of *analyzing* material is also capable of *applying, understanding, and remembering* it. 2. The second aspect to consider is the phase(s) of data analytics knowledge development in which the skill is utilized. The following phases have been identified as stages in the data analytics lifecycle. Most data analytics projects will encompass tasks in most of the stages, which are described below. To the best of your ability, indicate in which phases you would expect an entry-level data analyst to participate and which skills would be utilized in each one, even if your organization does not formally recognize these stages of analytical development. These descriptions will be available to refer back to in the coming

	questions. 1. Problem formulation. In this stage, a project is given motivation by **formulating the business problems that it will address and transforming them into actionable problem statements**. Tasks include determining business objectives and success measures, deploying problem formulation strategies, and defining a business problem. 2. Business understanding. In this stage, high-level **executive requirements are translated into specific analytic needs**. Some tasks individuals may be involved in in this stage include establishing a business case, assessing analytics capability maturity, enterprise knowledge acquisition (from existing documentation regarding business processes, queries, ETL processes, BI reports, etc.), determining a project management methodology, and selecting initial tools and techniques. 3. Data understanding. In this stage, **individuals become familiar with the data** from various sources that are relevant to solving the analytic problem. Tasks include exploration of within-DBMS and out-DBMS tools, considering business requirements and modeling requirements, verifying data quality, and describing the data/documentation. 4. Data preparation. Based on the outputs from the first three stages, an initial **data integration requirement is created**, indicating how each data element for modeling will be sourced or transformed. Tasks include creating data integration requirements, transforming data based on quality, business, or modeling requirements, and integrating data. 5. Modeling. In this stage, **applicable modeling techniques are selected, and analytic models are built** to provide the most desirable outcomes for the stated analytic goal. Tasks include selecting modeling techniques, describing modeling rules for the modeling technique, defining the training and testing strategy, building models, and assessing models. 6. Evaluation: In this stage, **candidate models are evaluated against business objectives and business problems formulated in the earlier stages**. Tasks include evaluating results, conducting field tests, reviewing the analytic process, and communicating results. 7. Deployment. In this stage, **the project is deployed**. Tasks include creating a deployment plan, producing a final project report, presentation, or other documentation, and reviewing the project. 8. Maintenance. In this stage, the **project is available in production and requiring maintenance**. Tasks include describing and storing analytic results, creating a model maintenance process, defining change initiation, monitoring the model usage, and initiating changes when required.
General skills	First, we will ask about the desirability and importance of several broad sets of skills. Consider each of the following skills and answer appropriately. Next, we will ask about more specific languages, software platforms, and tools. *Note: A skill could be both highly undesirable and very important - in this case, you would prioritize understanding whether the individual had that skill so that you could avoid hiring them.* (Options included Hardware Skills, Programming, Software, Systems Architecture & Infrastructure, Systems Modeling, Users and Organizations, Statistics, Visualization, and Data Management – drawn from (CC2020 Task Force, 2020). For each category, respondents were asked to respond regarding desirability, importance, and the level of understanding desired. See Figure B1 for an example.
Specific skills – level of understanding	For each of the following, indicate the level of understanding you would desire from the ideal entry-level data analyst. It may be that you would desire an applying understanding at some times/for some tasks and a creating understanding at other times/in other tasks; in this case, select the highest level of understanding (in this example, creating). Respondents were presented with the tools identified in Appendix A, and each was provided a drop-down with the levels of understanding. See Figure B2 for an example of the Programming Languages section.
Specific skills - phases	For each of the following, indicate the phase(s) in which a skill would be used in your organization by an entry-level data analyst. Select all that apply. Respondents were presented with the specific skills they selected in the previous section and checkboxes for each phase. See Figure B3 for an example of the Programming Languages section.
Optional specific skills notes	OPTIONAL: Any other notes or thoughts on programming languages/skills for entry-level data analysts? If there is a programming language you use heavily and value that we did NOT include above, please list it here. After each section (programming languages, analysis software, visualization software, and data management) respondents had the opportunity to answer this question.
Transferability	Each question was answered with a 7-point Likert scale ranging from “Strongly disagree” to “Strongly agree.”

	1. Skills are transferable across software platforms – for example, someone experienced in Tableau can transfer their skills to Power BI. 2. Skills are transferable across languages – for example, someone experienced in Java can transfer their skills to Python. 3. I would hire someone for a role even if they didn't have experience in a specific language/tool (e.g., Tableau) as long as they had experience in something similar (e.g., Power BI). 4. I/we use an automated system that scans resumes, so if the required software/language isn't listed, I never see the resume. 5. I am unconcerned about the skills an applicant has at the time of applying because we will train them ourselves. Open text answer: Do you have any other thoughts or notes on transferability of skills between software programs, tools, programming languages, etc. or about the role of training in preparing applicants who may not already have the necessary skills?
Dispositions	Read each of the following descriptions of dispositions that individuals might have, and answer how desirable/undesirable and important/unimportant these would be in your ideal entry-level data analyst candidate. A disposition may be both highly undesirable and very important - in this case, you would prioritize understanding whether the individual had that disposition so that you could avoid hiring them. *For each disposition, respondents were asked to rate its importance (5-point Likert scale from "very unimportant" to "very important") and its desirability (undesirable, neutral, or desirable). The list of dispositions was drawn from (CC2020 Task Force, 2020). See Figure B4 for an example.* 1. Adaptable: flexible, agile, adjusts in the response of change 2. Analytical: able to collect and analyze information, problem-solve, and make decisions using logical reasoning 3. Collaborative: team player; willing to work with others 4. Communicative: ready and willing to talk or impart information 5. Inventive: exploratory, able to look beyond simple solutions, creative 6. Meticulous: strong attention to detail, thorough, accurate 7. Passionate: conviction, strong commitment, compelling 8. Proactive: with initiative, self-starter, independent 9. Professional: discrete, ethical, astute 10. Purpose-driven: goal-driven, high goal achiever, strong business acumen 11. Responsible: use good judgment, discretion, acts appropriately 12. Responsive: respectful, reacts quickly and positively 13. Self-directed: self-motivated, determined, independent OPTIONAL: Any other notes or thoughts on dispositions for entry-level data analysts? If there are any specific dispositions or personality traits that you value or avoid that we did NOT list above, please list them here. (open text).

Table B1. Survey Measures

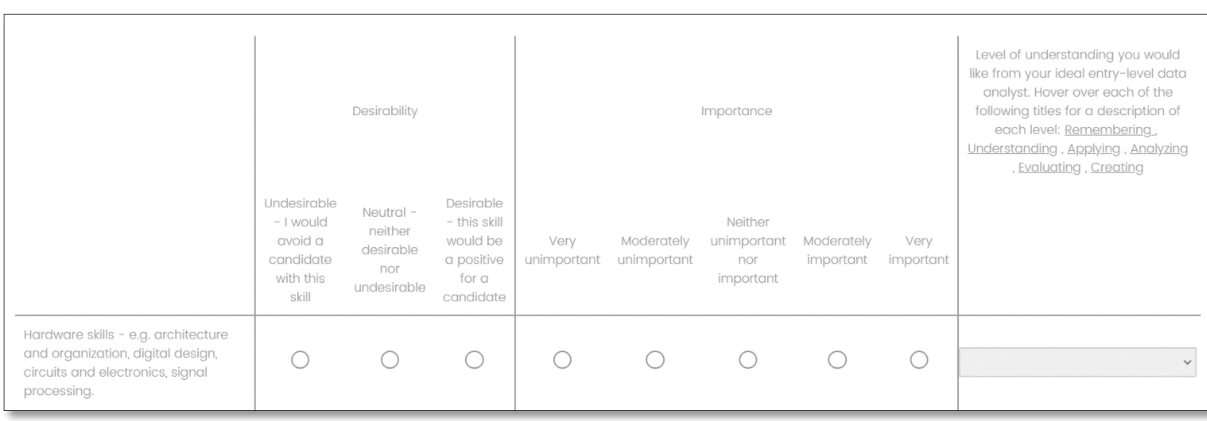
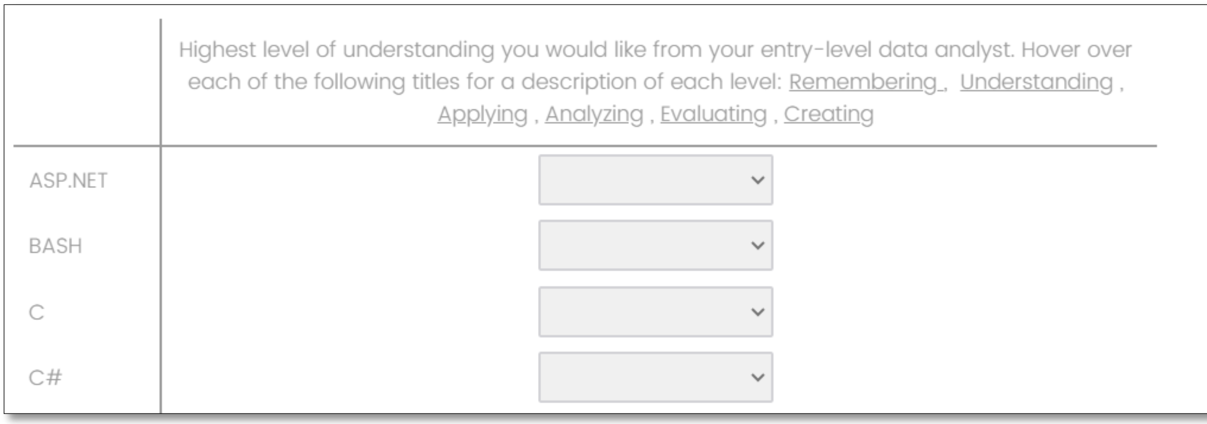
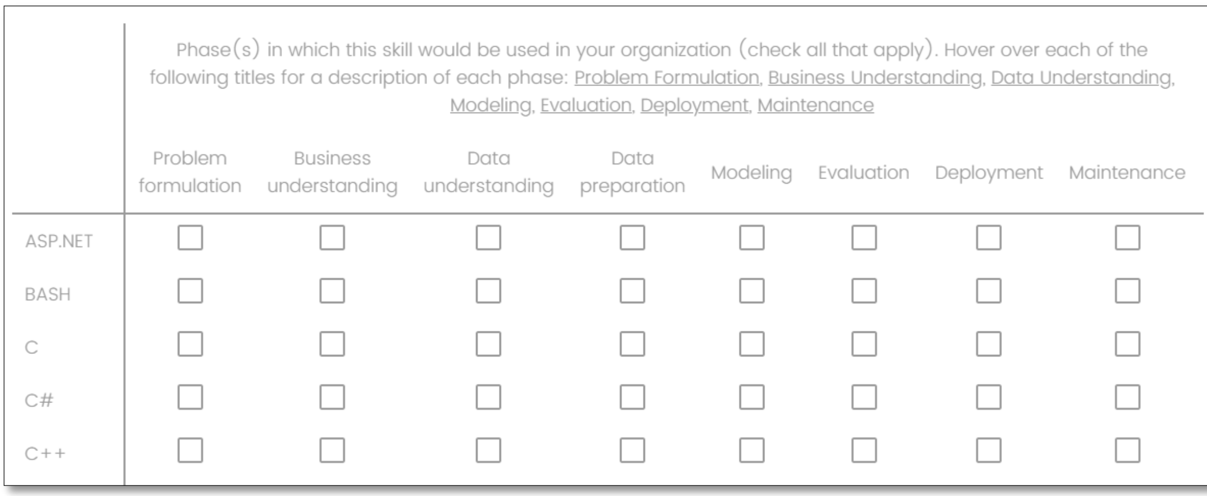


Figure B1. Example of General Skills Question



	Highest level of understanding you would like from your entry-level data analyst. Hover over each of the following titles for a description of each level: Remembering , Understanding , Applying , Analyzing , Evaluating , Creating
ASP.NET	
BASH	
C	
C#	

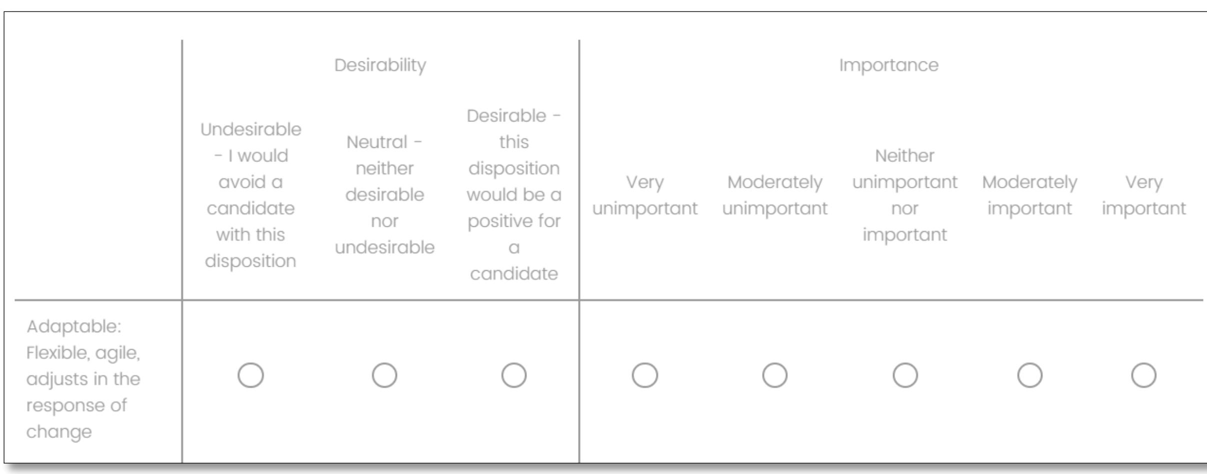
Figure B2. Example of Specific Skills Level of Understanding Questions (Programming Languages)



Phase(s) in which this skill would be used in your organization (check all that apply). Hover over each of the following titles for a description of each phase: Problem Formulation, Business Understanding, Data Understanding, Modeling, Evaluation, Deployment, Maintenance

	Problem formulation	Business understanding	Data understanding	Data preparation	Modeling	Evaluation	Deployment	Maintenance
ASP.NET								
BASH								
C								
C#								
C++								

Figure B3. Example of Specific Skills Phases (Programming Languages)



	Desirability			Importance				
	Undesirable - I would avoid a candidate with this disposition	Neutral - neither desirable nor undesirable	Desirable - this disposition would be a positive for a candidate	Very unimportant	Moderately unimportant	Neither unimportant nor important	Moderately important	Very important
Adaptable: Flexible, agile, adjusts in the response of change								

Figure B4. Example of Dispositions Question

Appendix C. Summary of Levels of Understanding for Each Tool Surveyed

Level	ASP.NET	BASH	C	C#	C++	COBOL	FORTRAN	GO	HTML	Java	JBOSS	jQuery
None	11	19	7	7	4	17	24	22	1	5	21	10
Understanding	20	16	15	14	10	17	12	9	13	10	10	15
Applying	3	7	9	10	11	7	7	9	7	13	7	10
Analyzing	11	6	8	10	10	7	8	7	11	5	7	10
Evaluating	7	4	11	8	10	3	1	5	10	10	7	6
Creating	3	3	5	6	10	4	3	3	13	12	3	4
Total	55	55	55	55	55	55	55	55	55	55	55	55

Table C1a. Levels of Understanding Required for Programming Languages, Part 1

Level	Julia	Linux	Pearl	Pig	Python	R	Ruby	S	Scala	Splus	SQL	Swift	VBA
None	27	8	22	36	4	27	30	30	22	25	4	20	12
Understanding	12	17	15	7	10	11	7	12	16	10	8	11	7
Applying	3	8	2	3	14	4	6	8	5	7	14	11	15
Analyzing	8	6	8	3	7	8	3	3	5	6	11	2	8
Evaluating	2	14	5	5	11	3	6	0	5	5	10	7	7
Creating	3	2	3	1	9	2	3	2	2	2	8	4	6
Total	55	55	55	55	55	55	55	55	55	55	55	55	55

Table C1b. Levels of Understanding Required for Programming Languages, Part 2

Level	Azure	Google Analytics	Hive	Salesforce	Watson
None	6	3	19	9	22
Understanding	15	4	11	4	14
Applying	12	12	5	10	3
Analyzing	7	16	9	14	8
Evaluating	7	12	6	12	3
Creating	8	8	5	6	5
Total	55	55	55	55	55

Table C2. Levels of Understanding Required for Enterprise Software and Platforms

Level	H2O	Matlab	Rstudio	SAS	Splunk	SPSS	STATA
None	20	16	21	10	23	23	22
Understanding	15	12	9	5	11	9	7
Applying	7	8	11	14	6	6	8
Analyzing	3	6	3	12	9	8	10
Evaluating	5	9	4	10	5	7	6
Creating	5	4	7	4	1	2	2
Total	55	55	55	55	55	55	55

Table C3. Levels of Understanding Required for Statistics Software and Platforms

Level	ArcGIS	CognosIBM	Crystal Reports	fixml	GIS	High Charts	Javascript
None	20	19	25	25	23	22	4
Understanding	19	7	5	11	8	6	7
Applying	3	7	5	6	11	4	12
Analyzing	11	8	12	5	7	12	8
Evaluating	1	12	7	4	5	7	11
Creating	1	2	1	4	1	4	13
Total	55	55	55	55	55	55	55

Table C4a. Levels of Understanding Required for Visualization Software and Platforms, Part 1

https://doi.org/10.62273/SPYC4248

Level	Lumira	Microsoft Power BI	Pentaho	QGIS	Tableau	TIBCO
None	23	8	24	28	18	19
Understanding	8	9	14	5	6	16
Applying	4	15	6	6	10	5
Analyzing	13	5	4	9	7	5
Evaluating	4	11	5	4	9	5
Creating	3	7	2	3	5	5
Total	55	55	55	55	55	55

Table C4b. Levels of Understanding Required for Visualization Software and Platforms, Part 2

Level	AirTable	Altair	Alteryx	Amazon Web Services	Answer Rocket	Board	Chartio	ClicData	Datapine	Domo	Erwin Data Modeler
N	21	25	23	3	22	22	24	21	18	25	29
U	14	12	9	13	10	7	10	8	6	8	6
Ap	5	6	5	18	8	7	11	8	6	3	6
An	6	7	12	6	5	7	3	10	10	11	5
E	6	3	4	8	9	8	5	5	12	6	6
C	3	2	2	7	1	4	2	3	3	2	3
T	55	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C5a. Levels of Understanding Required for Analysis Software, Part 1

Level	Google Data Studio	Hitachi Vantara	Incorta	InfoBirst	Jenkins	Jupyter Notebook	Klipfolio	KNIME	Looker	Meta-base	Micro-Strategy
N	6	24	23	20	26	23	27	29	24	12	20
U	11	12	9	10	9	10	4	7	9	17	8
Ap	8	5	5	6	8	6	11	8	8	5	7
An	16	4	8	12	6	5	4	7	7	7	5
E	7	7	7	4	5	7	5	3	2	9	10
C	7	3	3	3	1	4	4	1	5	5	5
T	55	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C5b. Levels of Understanding Required for Analysis Software, Part 2

Level	Mode	Monkey Learn	Microsoft Excel	Open Refine	Periscope Data	Pyramid Analytics	Qlik	Qualtrics	QueryMe	Rapid-Miner	Redash
N	25	26	1	23	17	17	24	11	24	26	28
U	3	10	7	6	6	9	5	10	3	7	8
Ap	13	7	10	4	7	6	11	9	7	6	3
An	4	3	9	11	10	12	8	7	7	9	9
E	4	7	9	9	8	7	6	7	8	6	4
C	6	2	19	2	7	4	1	11	6	1	3
T	55	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C5c. Levels of Understanding Required for Analysis Software, Part 3

Level	Salesforce	Sigma-Computing	Sisense	Talend	Targit	Tilius	ThoughtSpot	Whatagraph	Yellowfin BI	Zoho
N	10	22	21	28	27	27	24	25	23	18
U	4	10	9	11	10	14	9	12	8	8
Ap	10	6	7	5	7	2	7	5	8	8
An	10	3	6	5	1	5	6	4	5	8
E	14	7	4	5	7	3	6	7	6	9
C	7	7	8	1	3	4	3	2	5	4
T	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C5d. Levels of Understanding Required for Analysis Software, Part 4

Level	Cassandra	DB2	DBMS	flume	Hadoop	Hbase	Mahoot	Map-Reduce	Micro-soft Access	Mongo-DB	My-SQL	No-SQL
N	31	26	23	30	32	29	28	24	7	22	9	20
U	13	13	12	6	8	6	10	11	8	11	7	9
Ap	1	6	11	6	5	2	4	6	17	6	10	6
An	8	2	3	7	4	11	5	8	6	6	12	10
E	2	8	3	5	2	5	5	4	7	7	7	7
C	0	0	3	1	4	2	3	2	10	3	10	3
T	55	55	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C6a. Levels of Understanding Required for Database and Data Management, Part 1

Level	oozie	Oracle	Postgre-SQL	Presto	shark	Spark	SQL-Server	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
N	28	4	22	26	29	26	7	29	30	16	32	23	31
U	7	7	10	9	13	9	14	10	6	7	7	8	5
Ap	10	13	12	11	5	6	9	5	6	9	2	11	6
An	6	11	5	3	3	4	12	7	3	7	5	6	5
E	1	10	3	3	4	4	5	2	5	8	6	4	5
C	3	10	3	3	1	6	8	2	5	8	3	3	3
T	55	55	55	55	55	55	55	55	55	55	55	55	55

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total

Table C6b. Levels of Understanding Required for Database and Data Management, Part 2

Appendix D. Summary of Phases When Tools Are Used

Phase	ASP.NET	BASH	C	C#	C++	COBOL	FORTRAN	GO	HTML	Java	JBOSS	jQuery
PF	22	10	19	15	19	10	12	9	16	20	7	17
BU	15	17	19	17	20	16	9	14	15	22	10	16
DU	19	18	23	21	29	15	13	16	26	30	11	17
DP	6	4	14	12	15	15	5	9	10	19	5	10
MA	7	6	12	7	20	8	3	4	13	9	12	8
E	7	5	8	9	16	10	5	8	14	13	8	9
D	11	6	12	15	11	8	3	8	14	13	6	8
MO	13	6	9	9	21	4	5	6	14	10	7	8
T	100	72	116	105	151	86	55	74	122	136	66	93

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D1a. Summary of Phases When Programming Languages Are Used, Part 1

Phase	Julia	Linux	Pearl	Pig	Python	R	Ruby	S	Scala	Splus	SQL	Swift	VBA
PF	8	13	15	6	19	6	12	6	8	7	15	14	14
BU	11	19	11	9	18	10	7	10	11	10	20	12	18
DU	14	21	13	4	28	9	10	9	16	16	25	12	25
DP	6	9	8	4	16	5	7	4	9	5	15	6	11
M	7	15	6	4	13	7	3	4	5	7	14	5	13
E	8	13	10	3	13	4	5	6	4	7	25	7	11
D	6	9	8	7	14	5	4	6	3	5	13	7	10
M	3	17	6	2	14	7	4	5	9	7	12	6	13
T	63	116	77	39	135	53	52	50	65	64	139	69	115

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D1b. Summary of Phases When Programming Languages Are Used, Part 2

Phase	Azure	Google Analytics	Hive	Salesforce	SAP	Watson
Problem Formulation	17	21	10	18	15	12
Business Understanding	27	24	13	22	22	16
Data Understanding	21	23	16	14	21	11
Data Preparation	14	19	13	17	13	10
Modeling	13	9	9	12	12	13
Evaluation	11	20	9	19	17	9
Deployment	22	16	8	8	9	16
Maintenance	14	18	11	13	13	9
Total	139	150	89	123	122	96

Table D2. Summary of Phases When Enterprise Tools Are Used

Phase	H2O	Matlab	R Studio	SAS	Splunk	SPSS	STATA
Problem Formulation	11	17	8	18	12	10	10
Business Understanding	13	17	14	20	13	13	14
Data Understanding	9	14	14	20	16	12	14
Data Preparation	9	12	10	17	8	7	11
Modeling	8	11	11	13	10	13	13
Evaluation	8	16	8	17	10	8	12
Deployment	8	11	12	15	11	7	11
Maintenance	7	10	8	16	9	7	9
Total	73	108	85	136	89	77	94

Table D3. Summary of Phases When Statistics Tools Are Used

Phase	Arc-GIS	Cog-nos IBM	Cry-stal Re-ports	fixml	GIS	High Charts	Java-script	Lu-mira	Micro-soft Power BI	Pen-taho	QGIS	Tab-leau	TIB-CO
Problem Formula-tion	19	12	12	12	8	13	17	7	15	12	9	9	14
Business Under-standing	14	14	13	13	18	10	25	14	24	15	9	21	14
Data Under-standing	14	20	11	13	10	10	27	9	25	12	13	17	18
Data Preparation	6	8	13	7	9	10	19	10	15	5	10	14	9
Modeling	6	10	9	8	11	8	16	11	8	9	6	14	8
Evaluation	9	12	6	7	10	9	17	9	19	6	13	15	8
Deploy-ment	5	6	7	9	7	8	16	7	9	9	8	11	11
Mainte-nance	15	14	6	8	9	13	21	11	12	8	5	11	10
Total	88	96	77	77	82	81	158	78	127	76	73	112	92

Table D4. Summary of Phases When Visualization Tools Are Used

Phase	Air-Table	Altair	Alteryx	Amazon Web Services	Answer Rocket	Board	Chartio	Clic-Data	Data-pine	Domo	Erwin Data Modeler
PF	18	10	15	22	14	9	11	10	11	9	14
BU	14	16	14	25	11	12	13	16	13	15	11
DU	12	12	17	28	13	11	14	13	15	12	13
DP	6	6	6	20	11	14	9	12	8	9	8
M	9	10	7	13	11	12	9	8	11	7	4
E	8	5	9	18	12	11	7	13	8	6	10
D	7	8	8	14	9	7	8	10	13	8	10
M	5	9	6	19	11	4	8	10	7	7	7
T	79	76	82	159	92	80	79	92	86	73	77

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D5a. Summary of Phases When Analysis Software Is Used, Part 1

Phase	Google Data Studio	Hitachi Vantara	Incorta	Info-Birst	Jenkins	Jupyter Note-book	Klip-folio	KNIME	Looker	Meta-base	Micro-Strategy
PF	19	13	13	14	12	9	10	13	9	15	14
BU	24	14	13	11	13	11	14	9	14	13	16
DU	26	15	12	17	13	15	12	10	12	18	15
DP	22	7	10	12	9	9	6	6	8	14	8
M	18	7	9	8	6	7	6	7	8	14	12
E	17	6	9	12	6	11	10	7	6	15	8
D	16	6	10	10	6	4	6	8	12	11	10
M	15	7	7	6	5	11	11	8	8	9	8
T	157	75	83	90	70	77	75	68	77	109	91

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D5b. Summary of Phases When Analysis Software Is Used, Part 2

Phase	Mode	Monkey Learn	Microsoft Excel	Open Refine	Periscope Data	Pyramid Analytics	Qlik	Qual-trics	Query-Me	Rapid-Miner	Re-dash
PF	12	11	27	16	13	19	12	22	14	13	12
BU	16	16	30	17	15	12	13	20	16	17	15
DU	15	10	29	12	15	19	11	20	10	11	6
DP	10	6	23	7	8	11	8	15	4	8	8
M	6	8	24	9	10	10	7	11	7	10	7
E	8	5	21	3	13	12	10	15	12	9	5
D	8	7	15	9	11	9	9	11	8	9	6
M	8	5	17	10	9	8	5	16	7	6	7
T	83	68	186	83	94	100	75	130	78	83	66

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D5c. Summary of Phases When Analysis Software Is Used, Part 3

Phase	Salesforce	Sigma-Computing	Sisense	Talend	Targit	Tilius	Thought-Spot	Whatagraph	Yellow-fin BI	Zoho
PF	23	13	15	12	17	11	16	19	12	15
BU	22	16	16	14	9	15	14	13	13	16
DU	19	17	14	10	14	9	15	12	14	17
DP	15	7	13	3	7	8	11	8	6	14
M	15	9	9	6	8	7	9	5	11	9
E	16	7	11	7	9	9	6	14	10	11
D	12	15	9	4	9	7	5	7	10	14
M	10	7	10	6	6	10	8	4	6	8
T	132	91	97	62	79	76	84	82	82	104

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D5d. Summary of Phases When Analysis Software Is Used, Part 4

Phase	Cassandra	DB2	DBMS	flume	Ha-doop	H-base	Mahoot	Map-Reduce	Micro-soft Access	Mongo-DB	My-SQL	No-SQL
PF	13	16	12	12	8	14	9	9	23	11	16	14
BU	11	14	17	13	8	9	9	14	24	20	21	20
DU	5	16	21	12	11	11	14	12	24	17	22	19
DP	6	5	12	9	7	8	8	10	19	9	13	8
M	3	7	9	9	5	8	4	8	13	10	10	11
E	9	8	13	7	5	8	9	9	18	9	10	11
D	3	8	9	6	3	6	4	6	12	13	13	12
M	7	8	7	9	3	9	4	7	18	9	13	9
T	57	82	100	77	50	73	61	75	151	98	118	104

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D6a. Summary of Phases When Database Tools Are Used, Part 1

Phase	oozie	Oracle	Postgre-SQL	Presto	shark	Spark	SQL-Server	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
PF	9	21	17	10	11	10	16	8	8	14	10	13	9
BU	17	31	17	19	10	22	22	8	11	18	11	17	14
DU	12	36	20	11	10	13	25	13	11	17	12	13	12
DP	11	19	7	6	6	7	17	8	6	10	5	5	6
M	7	15	8	8	4	7	13	9	6	10	5	10	5
E	7	20	7	9	7	10	13	9	8	7	6	13	4
D	8	17	7	4	10	9	14	7	8	8	9	9	11
M	9	17	5	8	8	11	11	13	5	15	6	6	6
T	80	176	88	75	66	89	131	75	63	99	64	86	67

PF = Problem formulation; BU = Business Understanding; DU = Data Understanding; DP = Data Preparation; M = Modeling; E = Evaluation; D = Deployment; M = Maintenance; T = Total

Table D6b. Summary of Phases When Database Tools Are Used, Part 2

Appendix E. Tool Popularity by Industry

Industry	Description	Count	%
Information (NAICS 51)	Establishments engaged in the following processes: (a) producing and distributing information and cultural products, (b) providing the means to transmit or distribute these products as well as data or communications, and (c) processing data.	15	27%
Data Processing, Hosting, and Related Services (NAICS 518)	A subsector of the Information sector: establishments that provide the infrastructure for hosting and/or data processing services.	9	16%
Computer and Electronic Product Manufacturing (NAICS 334)	A subsector of the Manufacturing sector: establishments that manufacture computers, computer peripherals, communications equipment, and similar electronic products, and establishments that manufacture components for such products.	8	15%
Professional, Scientific, and Technical Services (NAICS 54)	Establishments that specialize in performing professional, scientific, and technical activities for others. These activities require a high degree of expertise and training. The establishments in this sector specialize according to expertise and provide these services to clients in a variety of industries and, in some cases, to households. Activities performed include: legal advice and representation; accounting, bookkeeping, and payroll services; architectural, engineering, and specialized design services; computer services; consulting services; research services; advertising services; photographic services; translation and interpretation services; veterinary services, and other professional, scientific, and technical services.	4	7%
Health Care and Social Assistance (NAICS 62)	Establishments providing health care and social assistance for individuals.	2	4%
Finance and Insurance (NAICS 52)	Establishments primarily engaged in financial transactions (transactions involving the creation, liquidation, or change in ownership of financial assets) and/or in facilitating financial transactions.	2	4%
Health and Personal Care Stores (NAICS 446)	Retail health and personal care merchandise from fixed point-of-sale locations. Establishments are characterized principally by the products they retail, and some health and personal care stores may have specialized staff trained in dealing with the products.	1	2%
Goods-Producing Industries	A supersector group consisting of: natural resources and mining (agriculture, forestry, fishing and hunting; mining, quarrying, and oil and gas extraction), construction, and manufacturing.	1	2%
Administrative and Support and Waste Management and Remediation Services (NAICS 56)	Establishments performing routine support activities for the day-to-day operations of other organizations, e.g., office administration, hiring and placing of personnel, document preparation, solicitation, collection, security and surveillance services, cleaning, and waste disposal services.	1	2%
Electronics and Appliance Stores (NAICS 443)	Retail new electronics and appliances from point-of-sale locations.	1	2%
Accommodation (NAICS 721)	Provide lodging or short-term accommodations for travelers, vacationers, and others.	1	2%
Private Households (NAICS 814)	Private households that engage in employing workers on or about the premises in activities primarily concerned with the operation of the household.	1	2%
Administrative and Support Services (NAICS 561)	Establishments engaged in activities that support the day-to-day operations of other organizations, e.g., general management, personnel administration, clerical activities, cleaning activities.	1	2%
Retail Trade (NAICS 44-45)	Establishments engaged in retailing merchandise, generally without transformation, and rendering services incidental to the sale of the merchandise .	1	2%

Service-Providing Industries	A supersector group consisting of trade, transportation and utilities; information; financial activities; professional and business services; education and health services; leisure and hospitality; other services (except Public Administration); and government.	1	2%
Internet Publishing and Broadcasting (NAICS 516)	Now part of Other Information Services: establishments supplying information, storing and providing access to information, searching and retrieving information, operating Web sites that use search engines to allow for searching information on the Internet, or publishing and/or broadcasting content exclusively on the Internet.	1	2%
Educational Services (NAICS 61)	Establishments that provide instruction and training in a wide variety of subjects.	1	2%
Management of Companies and Enterprises (NAICS 55)	Comprises 1) establishments that hold the securities of (or other equity interests in) companies and enterprises for the purpose of owning a controlling interest or influencing management decisions or 2) establishments (except government establishments) that administer, oversee, and manage establishments of the company or enterprise and that normally undertake the strategic or organizational planning and decision making role of the company or enterprise.	1	2%
Professional and Business Services	A supersector of the service-providing industries supersector, composed of: professional, scientific, and technical services; management of companies and enterprises; administrative and support and waste management and remediation services.	1	2%
Transportation and Warehousing (NAICS 48-49)	Industries providing transportation of passengers and cargo, warehousing and storage for goods, scenic and sightseeing transportation, and support activities related to modes of transportation.	1	2%
Other Information Services (NAICS 519)	Establishments supplying information, storing and providing access to information, searching and retrieving information, operating Web sites that use search engines to allow for searching information on the Internet, or publishing and/or broadcasting content exclusively on the internet.	1	2%

Table E1. Industries Represented by the Sample

Level	ASP. NET	BASH	C	C#	C++	CO-BOL	FOR-TRAN	GO	**HTML**	**Java**	J-BOSS	J-Query
None	4	6	4	3	1	4	7	6	**0**	**1**	5	3
Under-standing	5	4	3	3	1	6	2	5	**4**	**0**	3	3
Applying	1	1	2	3	2	2	2	1	**2**	**5**	0	2
Analyzing	2	1	1	2	4	2	3	1	**3**	**2**	2	1
Evaluating	2	2	3	2	3	0	0	1	**2**	**4**	2	4
Creating	1	1	2	2	4	1	1	1	**4**	**3**	3	2
Total	15	15	15	15	15	15	15	15	**15**	**15**	15	15
Non-None %	73%	60%	73%	80%	93%	73%	53%	60%	**100%**	**93%**	67%	80%

Table E2a1. Programming Language Popularity: Information Industry, Part 1

Level	Julia	Linux	Pearl	Pig	**Py-thon**	R	Ru-by	S	Scala	Splus	SQL	Swift	VBA
None	8	3	6	11	**1**	9	8	11	7	7	2	6	2
Under-standing	2	5	4	1	**4**	3	1	2	4	4	2	2	3
Applying	1	3	0	1	**3**	1	2	1	0	0	3	1	3
Analyzing	1	1	2	1	**2**	1	1	0	1	2	2	0	4
Evaluating	1	3	1	0	**2**	1	2	0	3	2	4	6	1
Creating	2	0	2	1	**3**	0	1	1	0	0	2	0	2
Total	15	15	15	15	**15**	15	15	15	15	15	15	15	15
Non-None %	47%	80%	60%	27%	**93%**	40%	47%	27%	53%	53%	87%	60%	87%

Table E2a2. Programming Language Popularity: Information Industry, Part 2

Level	**ASP. NET**	BASH	**C**	C#	C++	COB OL	FOR-TRAN	GO	**HTML**	Java	J-BOSS	J-Query
None	**0**	1	**0**	1	1	2	3	4	**0**	1	2	1
Under-standing	**5**	4	**3**	3	1	1	2	1	**1**	1	4	3
Applying	**1**	1	**3**	1	2	2	2	1	**1**	2	1	3
Analyzing	**3**	2	**1**	2	2	3	2	0	**2**	1	1	2
Evaluating	**0**	0	**2**	2	1	0	0	2	**3**	3	1	0
Creating	**0**	1	**0**	0	2	1	0	1	**2**	1	0	0
Total	**9**	9	**9**	9	9	9	9	9	**9**	9	9	9
Non-None %	**100%**	89%	**100%**	89%	89%	78%	67%	56%	**100%**	89%	78%	89%

Table E2b1. Programming Language Popularity: Data Processing Industry, Part 1

Level	Julia	Linux	Pearl	Pig	**Py-thon**	R	Ruby	S	Scala	Splus	**SQL**	Swift	**VBA**
None	3	1	2	6	**0**	5	3	5	2	2	**0**	2	**0**
Under-standing	4	4	3	2	**2**	1	3	1	4	2	**2**	2	**0**
Applying	1	0	1	1	**1**	0	2	1	1	2	**2**	4	**5**
Analyzing	1	1	2	0	**2**	1	0	2	2	1	**2**	0	**1**
Evaluating	0	3	1	0	**4**	1	1	0	0	1	**3**	0	**1**
Creating	0	0	0	0	**0**	1	0	0	0	1	**0**	1	**2**
Total	9	9	9	9	**9**	9	9	9	9	9	**9**	9	**9**
Non-None %	67%	89%	78%	33%	**100%**	44%	67%	44%	78%	78%	**100%**	78%	**100%**

Table E2b2. Programming Language Popularity: Data Processing Industry, Part 2

Level	ASP. NET	BASH	C	C#	C++	CO-BOL	FOR-TRAN	GO	HTML	Java	JBOSS	JQuery
None	3	3	1	1	1	4	4	3	1	2	3	1
Under-standing	2	1	1	0	1	3	1	1	1	3	1	3
Applying	0	2	2	2	1	0	1	2	1	1	2	1
Analyzing	2	0	1	4	3	0	0	1	0	0	2	3
Evaluating	0	1	1	0	1	0	0	0	3	0	0	0
Creating	1	1	2	1	1	1	2	1	2	2	0	0
Total	8	8	8	8	8	8	8	8	8	8	8	8
Non-None %	63%	63%	88%	88%	88%	50%	50%	63%	88%	75%	63%	88%

Table E2c1. Programming Language Popularity: Computer and Electronic Product Manufacturing Industry, Part 1

Level	Julia	**Linux**	Pearl	Pig	**Py-thon**	R	Ruby	S	Scala	Splus	**SQL**	Swift	VBA
None	4	**0**	3	6	**0**	3	4	4	5	5	**0**	4	4
Under-standing	0	**4**	3	1	**1**	1	0	3	0	0	**1**	2	0
Apply-ing	0	**1**	0	0	**2**	1	2	1	1	1	**3**	0	1
Analyz-ing	3	**0**	1	0	**0**	2	0	0	2	1	**1**	0	2
Evaluat-ing	0	**2**	0	1	**3**	0	2	0	0	1	**1**	1	1
Creating	1	**1**	1	0	**2**	1	0	0	0	0	**2**	1	0
Total	8	**8**	8	8	**8**	8	8	8	8	8	**8**	8	8
Non-None %	50%	**100%**	63%	25%	**100%**	63%	50%	50%	38%	38%	**100%**	50%	50%

Table E2c2. Programming Language Popularity: Computer and Electronic Product Manufacturing Industry, Part 2

Level	ASP.NET	BASH	**C**	C#	C++	CO-BOL	FOR-TRAN	GO	**HTML**	Java	J-BOSS	J-Query
None	2	4	**0**	1	1	2	4	3	**0**	1	3	1
Under-standing	1	0	**2**	1	1	2	0	0	**3**	2	0	1
Applying	0	0	**1**	1	1	0	0	0	**0**	1	0	2
Analyzing	0	0	**0**	0	0	0	0	0	**1**	0	0	0
Evaluating	1	0	**1**	0	0	0	0	1	**0**	0	1	0
Creating	0	0	**0**	1	1	0	0	0	**0**	0	0	0
Total	4	4	**4**	4	4	4	4	4	**4**	4	4	4
Non-None %	50%	0%	**100%**	75%	75%	50%	0%	25%	**100%**	75%	25%	75%

Table E2d1. Programming Language Popularity: Professional and Technical Services Industry, Part 1

Level	Julia	Linux	Pearl	Pig	Py-thon	R	Ruby	S	Scala	Splus	**SQL**	Swift	VBA
None	4	2	4	4	2	4	4	3	3	4	**0**	2	3
Under-standing	0	0	0	0	0	0	0	1	0	0	**1**	1	0
Applying	0	1	0	0	1	0	0	0	1	0	**0**	0	0
Analyzing	0	0	0	0	0	0	0	0	0	0	**3**	0	0
Evaluating	0	1	0	0	0	0	0	0	0	0	**0**	0	1
Creating	0	0	0	0	1	0	0	0	0	0	**0**	1	0
Total	4	4	4	4	4	4	4	4	4	4	**4**	4	4
Non-None %	0%	50%	0%	0%	50%	0%	0%	25%	25%	0%	**100%**	50%	25%

Table E2d2. Programming Language Popularity: Professional and Technical Services Industry, Part 2

Level	Azure	**Google Analytics**	Hive	Salesforce	SAP	Watson
None	2	**1**	4	4	4	7
Understanding	3	**3**	5	0	1	4
Applying	3	**2**	1	2	3	0
Analyzing	1	**4**	3	5	3	2
Evaluating	3	**3**	2	2	2	0
Creating	3	**2**	0	2	2	2
Total	15	**15**	15	15	15	15
Non-None %	87%	**93%**	73%	73%	73%	53%

Table E3a. Enterprise Software Popularity: Information Industry

Level	**Azure**	**Google Analytics**	Hive	Salesforce	**SAP**	Watson
None	**0**	**0**	2	1	**0**	2
Understanding	**3**	**0**	2	2	**1**	4
Applying	**1**	**2**	0	1	**0**	0
Analyzing	**2**	**2**	3	3	**4**	0
Evaluating	**3**	**5**	1	2	**2**	2
Creating	**0**	**0**	1	0	**2**	1
Total	**9**	**9**	9	9	**9**	9
Non-None %	**100%**	**100%**	78%	89%	**100%**	78%

Table E3b. Enterprise Software Popularity: Data Processing Industry

Level	Azure	Google Analytics	Hive	**Salesforce**	SAP	Watson
None	1	1	2	**0**	1	3
Understanding	4	0	2	**1**	2	0
Applying	0	1	0	**4**	1	1
Analyzing	1	5	0	**1**	3	3
Evaluating	0	0	1	**1**	1	0
Creating	2	1	3	**1**	0	1
Total	8	8	8	**8**	8	8
Non-None %	88%	88%	75%	**100%**	88%	63%

Table E3c. Enterprise Software Popularity: Computer and Electronics Manufacturing Industry

Level	Azure	Google Analytics	Hive	Salesforce	SAP	Watson
None	1	1	4	1	2	4
Understanding	0	0	0	0	0	0
Applying	3	1	0	0	1	0
Analyzing	0	1	0	1	1	0
Evaluating	0	0	0	1	0	0
Creating	0	1	0	1	0	0
Total	4	4	4	4	4	4
Non-None %	75%	75%	0%	75%	50%	0%

Table E3d. Enterprise Software Popularity: Professional, Scientific, and Technical Services Industry

Level	H2O	Matlab	Rstudio	SAS	Splunk	SPSS	STATA
None	9	2	5	4	6	7	7
Understanding	3	5	4	0	3	2	1
Applying	1	0	1	5	1	1	2
Analyzing	0	3	1	2	4	3	2
Evaluating	1	4	1	3	1	1	3
Creating	1	1	3	1	0	1	0
Total	15	15	15	15	15	15	15
Non-None %	40%	87%	67%	73%	60%	53%	53%

Table E4a. Statistics Software Popularity: Information Industry

Level	H2O	Matlab	Rstudio	**SAS**	Splunk	SPSS	STATA
None	3	1	3	**0**	2	2	1
Understanding	2	2	1	**3**	4	1	1
Applying	1	3	2	**1**	1	2	3
Analyzing	1	1	0	**2**	0	3	2
Evaluating	0	2	2	**3**	1	1	2
Creating	2	0	1	**0**	1	0	0
Total	9	9	9	**9**	9	9	9
Non-None %	67%	89%	67%	**100%**	78%	78%	89%

Table E4b. Statistics Software Popularity: Data Processing Industry

Level	H2O	Matlab	Rstudio	SAS	Splunk	SPSS	STATA
None	2	3	4	1	5	4	3
Understanding	2	0	1	1	0	1	1
Applying	2	2	2	1	0	1	1
Analyzing	1	1	0	3	2	0	3
Evaluating	0	1	0	0	1	1	0
Creating	1	1	1	2	0	1	0
Total	8	8	8	8	8	8	8
Non-None %	67%	75%	83%	67%	43%	86%	88%

Table E4c. Statistics Software Popularity: Computer and Electronics Manufacturing Industry

Level	H2O	Matlab	Rstudio	SAS	Splunk	SPSS	STATA
None	3	4	4	1	2	3	4
Understanding	1	0	0	1	0	1	0
Applying	0	0	0	2	1	0	0
Analyzing	0	0	0	0	1	0	0
Evaluating	0	0	0	0	0	0	0
Creating	0	0	0	0	0	0	0
Total	4	4	4	4	4	4	4
Non-None %	25%	0%	0%	75%	50%	25%	0%

Table E4d. Statistics Software Popularity: Professional, Scientific, and Technical Services Industry

Level	ArcGIS	Cognos IBM	Crystal Reports	fixml	GIS	High Charts	**JavaScript**
None	5	5	9	9	7	6	**0**
Understanding	5	3	1	2	1	1	**3**
Applying	1	1	1	0	1	0	**4**
Analyzing	4	1	3	2	3	3	**3**
Evaluating	0	4	1	1	3	3	**1**
Creating	0	1	0	1	0	2	**4**
Total	15	15	15	15	15	15	**15**
Non-None %	67%	67%	40%	40%	53%	60%	**100%**

Table E5a1. Visualization Software Popularity: Information Industry, Part 1

Level	Lumira	Microsoft Power BI	Pentaho	QGIS	Tableau	TIBCO
None	5	3	7	8	8	5
Understanding	4	2	5	1	1	2
Applying	0	3	1	0	2	1
Analyzing	3	0	0	4	2	1
Evaluating	1	6	2	1	1	3
Creating	2	1	0	1	1	3
Total	15	15	15	15	15	15
Non-None %	67%	80%	53%	47%	47%	67%

Table E5a2. Visualization Software Popularity: Information Industry, Part 2

Level	ArcGIS	Cognos IBM	Crystal Reports	fixml	GIS	High Charts	**Java-Script**
None	1	1	1	2	2	2	**0**
Understanding	4	0	2	1	2	2	**0**
Applying	1	2	1	3	3	2	**1**
Analyzing	3	3	1	0	1	2	**1**
Evaluating	0	3	3	1	1	1	**5**
Creating	0	0	1	2	0	0	**2**
Total	9	9	9	9	9	9	**9**
Non-None %	89%	89%	89%	78%	78%	78%	**100%**

Table E5b1. Visualization Software Popularity: Data Processing Industry, Part 1

Level	Lumira	Microsoft Power BI	Pentaho	QGIS	**Tableau**	TIBCO
None	2	1	2	2	**0**	1
Understanding	0	1	4	1	**2**	4
Applying	2	3	1	2	**0**	1
Analyzing	3	2	0	2	**2**	2
Evaluating	2	1	0	1	**4**	0
Creating	0	1	2	1	**1**	1
Total	9	9	9	9	**9**	9
Non-None %	78%	89%	78%	78%	**100%**	89%

Table E5b2. Visualization Software Popularity: Data Processing Industry, Part 2

Level	ArcGIS	Cognos IBM	Crystal Reports	fixml	GIS	High Charts	Java-Script
None	4	2	3	4	4	2	2
Understanding	2	3	1	3	1	1	0
Applying	0	1	0	1	2	1	0
Analyzing	1	1	3	0	0	3	1
Evaluating	0	0	1	0	0	1	4
Creating	1	1	0	0	1	0	1
Total	8	8	8	8	8	8	8
Non-None %	50%	75%	63%	50%	50%	75%	75%

Table E5c1. Visualization Software Popularity: Computer and Electronics Manufacturing Industry, Part 1

Level	Lumira	Microsoft Power BI	Pentaho	QGIS	Tableau	TIBCO
None	4	2	4	4	2	3
Understanding	1	0	1	1	1	1
Applying	0	3	1	2	2	2
Analyzing	2	2	1	1	1	1
Evaluating	1	0	1	0	1	0
Creating	0	1	0	0	1	1
Total	8	8	8	8	8	8
Non-None %	50%	75%	50%	50%	75%	63%

Table E5c2. Visualization Software Popularity: Computer and Electronics Manufacturing Industry, Part 2

Level	ArcGIS	Cognos IBM	Crystal Reports	fixml	GIS	High Charts	Java-Script
None	4	3	3	4	4	4	2
Understanding	0	0	0	0	0	0	0
Applying	0	1	1	0	0	0	2
Analyzing	0	0	0	0	0	0	0
Evaluating	0	0	0	0	0	0	0
Creating	0	0	0	0	0	0	0
Total	4	4	4	4	4	4	4
Non-None %	0%	25%	25%	0%	0%	0%	50%

Table E5d1. Visualization Software Popularity: Professional, Scientific, and Technical Services Industry, Part 1

Level	Lumira	Microsoft Power BI	Pentaho	QGIS	Tableau	TIBCO
None	3	1	3	4	3	4
Understanding	0	0	0	0	0	0
Applying	0	1	1	0	0	0
Analyzing	1	0	0	0	0	0
Evaluating	0	1	0	0	0	0
Creating	0	1	0	0	1	0
Total	4	4	4	4	4	4
Non-None %	25%	75%	25%	0%	25%	0%

Table E5c2. Visualization Software Popularity: Professional, Scientific, and Technical Services Industry, Part 2

Level	AirTable	Altair	Alteryx	**Amazon Web Services**	Answer Rocket	Board	Chartio	ClicData	Datapine	Domo	Erwin Data Modeler
N	7	7	7	**0**	6	5	6	7	5	6	8
U	4	4	2	**5**	3	2	6	2	2	2	1
Ap	1	1	0	**1**	2	1	2	3	0	1	2
An	2	2	4	**2**	1	3	0	1	1	1	1
E	1	0	2	**3**	3	3	0	2	6	3	2
C	0	1	0	**4**	0	1	1	0	1	2	1
T	15	15	15	**15**	15	15	15	15	15	15	15
NN%	53%	53%	53%	**100%**	60%	67%	60%	53%	67%	60%	47%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6a1. Analysis Software Popularity: Information Industry, Part 1

Level	Google Data Studio	Hitachi Vantara	Incorta	InfoBirst	Jenkins	Jupyter Notebook	Klipfolio	KNIME	Looker	Meta-base	Micro-Strategy
N	2	8	6	5	5	6	7	7	7	2	4
U	4	3	4	4	6	3	1	3	3	6	3
Ap	1	0	0	0	1	2	2	2	1	1	2
An	4	2	3	4	1	3	2	3	0	2	2
E	2	0	1	2	2	0	1	0	1	3	2
C	2	2	1	0	0	1	2	0	3	1	2
T	15	15	15	15	15	15	15	15	15	15	15
NN%	87%	47%	60%	67%	67%	60%	53%	53%	53%	87%	73%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6a2. Analysis Software Popularity: Information Industry, Part 2

Level	Mode	Monkey Learn	**Micro-soft Excel**	Open Refine	Peri-scope Data	Pyramid Analytics	Qlik	Qualtrics	Query-Me	Rapid-Miner	Re-dash
N	6	6	**0**	7	6	5	6	4	4	6	7
U	2	5	**2**	1	3	2	1	1	2	3	2
Ap	5	3	**3**	1	3	0	4	2	1	2	0
An	0	0	**2**	3	0	3	1	2	4	2	2
E	0	1	**5**	3	3	3	3	2	1	2	3
C	2	0	**3**	0	0	2	0	4	3	0	1
T	15	15	**15**	15	15	15	15	15	15	15	15
NN%	60%	60%	**100%**	53%	60%	67%	60%	73%	73%	60%	53%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6a3. Analysis Software Popularity: Information Industry, Part 3

Level	Sales-force	Sigma-Computing	Sisense	Talend	Targit	Tilius	Thought-Spot	Whata-graph	Yellow-fin BI	Zoho
N	3	5	6	9	6	7	5	5	6	4
U	2	3	2	1	5	4	3	5	4	2
Ap	3	1	3	1	0	1	4	1	2	3
An	2	0	0	1	0	1	1	1	1	0
E	4	3	1	3	3	0	2	3	1	4
C	1	3	3	0	1	2	0	0	1	2
T	15	15	15	15	15	15	15	15	15	15
NN%	80%	67%	60%	40%	60%	53%	67%	67%	60%	73%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6a4. Analysis Software Popularity: Information Industry, Part 4

Level	AirTable	Altair	**Alteryx**	**Amazon Web Services**	Answer Rocket	Board	Chartio	**ClicData**	**Datapine**	Domo	Erwin Data Modeler
N	2	3	**0**	**0**	1	1	1	**0**	**0**	2	2
U	2	1	**1**	**4**	1	1	1	**2**	**1**	1	2
Ap	3	2	**1**	**0**	5	2	5	**1**	**1**	0	1
An	2	2	**4**	**2**	0	1	1	**3**	**4**	5	2
E	0	1	**2**	**2**	2	2	1	**1**	**2**	1	2
C	0	0	**1**	**1**	0	2	0	**2**	**1**	0	0
T	9	9	**9**	**9**	9	9	9	**9**	**9**	9	9
NN%	78%	67%	**100%**	**100%**	89%	89%	89%	**100%**	**100%**	78%	78%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6b1. Analysis Software Popularity: Data Processing Industry, Part 1

Level	**Google Data Studio**	Hitachi Vantara	Incorta	Info-Birst	Jenkins	Jupyter Note-book	Klip-folio	KNIME	Looker	Meta-base	Micro-Strategy
N	**0**	1	1	1	3	2	1	1	1	1	1
U	**3**	1	1	1	2	3	1	3	3	2	1
Ap	**0**	4	1	3	1	1	4	0	0	0	1
An	**3**	0	2	0	1	0	1	4	2	2	3
E	**2**	3	4	2	1	1	1	0	1	4	2
C	**1**	0	0	2	1	2	1	1	2	0	1
T	**9**	9	9	9	9	9	9	9	9	9	9
NN%	**100%**	89%	89%	89%	67%	78%	89%	89%	89%	89%	89%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6b2. Analysis Software Popularity: Data Processing Industry, Part 2

Level	Mode	Monkey Learn	**Micro-soft Excel**	Open Refine	Peri-scope Data	Pyramid Analytics	Qlik	**Qualtrics**	Query-Me	Rapid-Miner	Re-dash
N	2	2	**0**	2	1	1	3	**0**	2	3	3
U	0	2	**2**	1	0	1	1	**3**	0	2	2
Ap	4	1	**0**	1	1	1	2	**2**	2	1	1
An	1	2	**2**	2	3	5	1	**2**	1	1	2
E	0	1	**2**	3	2	1	1	**1**	3	2	0
C	2	1	**3**	0	2	0	1	**1**	1	0	1
T	9	9	**9**	9	9	9	9	**9**	9	9	9
NN%	78%	78%	**100%**	78%	89%	89%	67%	**100%**	78%	67%	67%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6b3. Analysis Software Popularity: Data Processing Industry, Part 3

Level	Salesforce	Sigma-Computing	Sisense	Talend	Targit	Tilius	ThoughtSpot	Whata-graph	Yellowfin BI	Zoho
N	2	3	3	3	3	4	3	3	2	1
U	0	1	1	2	1	3	3	3	1	1
Ap	1	0	1	1	3	0	0	0	1	2
An	2	2	1	1	0	1	1	0	3	4
E	3	2	2	1	2	1	1	2	0	1
C	1	1	1	1	0	0	1	1	2	0
T	9	9	9	9	9	9	9	9	9	9
NN%	78%	67%	67%	67%	67%	56%	67%	67%	78%	89%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6b4. Analysis Software Popularity: Data Processing Industry, Part 4

Level	AirTable	Altair	Alteryx	Amazon Web Services	Answer Rocket	Board	Chartio	ClicData	Datapine	Domo	Erwin Data Modeler
N	4	3	3	1	5	4	5	3	2	5	4
U	1	2	1	1	1	0	1	0	0	0	0
Ap	0	1	1	5	0	2	1	1	2	0	0
An	1	1	3	0	0	1	0	4	3	1	1
E	2	0	0	1	2	1	1	0	1	2	2
C	0	1	0	0	0	0	0	0	0	0	1
T	8	8	8	8	8	8	8	8	8	8	8
NN%	50%	63%	63%	88%	38%	50%	38%	63%	75%	38%	50%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6c1. Analysis Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 1

Level	**Google Data Studio**	Hitachi Vantara	Incorta	InfoBirst	Jenkins	Jupyter Notebook	Klip-folio	KNIME	Looker	Meta-base	Micro-Strategy
N	**0**	3	3	3	5	3	5	6	4	2	4
U	**1**	1	1	3	0	0	0	0	0	2	1
Ap	**1**	1	2	1	1	2	1	1	2	0	1
An	**3**	1	1	1	1	1	1	0	2	1	0
E	**2**	1	0	0	1	1	1	1	0	1	2
C	**1**	1	1	0	0	1	0	0	0	2	0
T	**8**	8	8	8	8	8	8	8	8	8	8
NN%	**100%**	63%	63%	63%	38%	63%	38%	25%	50%	75%	50%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6c2. Analysis Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 2

Level	Mode	Monkey Learn	Micro-soft Excel	Open Refine	Peri-scope Data	Pyramid Analytics	Qlik	Qualtrics	Query-Me	Rapid-Miner	Re-dash
N	5	6	1	3	1	2	5	3	4	4	5
U	0	0	1	1	0	3	1	2	0	1	1
Ap	1	1	1	1	1	2	1	2	0	0	1
An	1	0	2	1	4	0	1	0	2	2	0
E	1	1	1	1	1	1	0	0	1	0	1
C	0	0	2	1	1	0	0	1	1	1	0
T	8	8	8	8	8	8	8	8	8	8	8
NN%	38%	25%	88%	63%	88%	75%	38%	63%	50%	50%	38%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6c3. Analysis Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 3

Level	Salesforce	Sigma-Computing	Sisense	Talend	Targit	Tilius	Thought-Spot	Whata-graph	Yellow-fin BI	Zoho
N	1	5	3	4	4	4	4	5	3	4
U	0	0	2	1	1	0	1	1	1	1
Ap	1	1	0	1	1	1	1	1	0	1
An	3	1	2	2	1	1	2	1	0	0
E	2	1	0	0	0	1	0	0	2	2
C	1	0	1	0	1	1	0	0	2	0
T	8	8	8	8	8	8	8	8	8	8
NN%	88%	38%	63%	50%	50%	50%	50%	38%	63%	50%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6c4. Analysis Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 4

Level	Air-Table	Altair	Alteryx	Amazon Web Services	Answer Rocket	Board	Chartio	Clic-Data	Data-pine	Domo	Erwin Data Modeler
N	4	3	3	1	3	4	4	4	3	4	4
U	0	1	0	1	0	0	0	0	0	0	0
Ap	0	0	1	2	0	0	0	0	0	0	0
An	0	0	0	0	1	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	1	0	0
C	0	0	0	0	0	0	0	0	0	0	0
T	4	4	4	4	4	4	4	4	4	4	4
NN%	0%	25%	25%	75%	25%	0%	0%	0%	25%	0%	0%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6d1. Analysis Software Popularity: Professional, Scientific, and Technical Services Industry, Part 1

Level	Google Data Studio	Hitachi Vantara	Incorta	Info-Birst	Jenkins	Jupyter Notebook	Klip-folio	KNIME	Looker	Meta-base	Micro-Strategy
N	1	4	4	4	3	4	4	4	4	3	3
U	1	0	0	0	0	0	0	0	0	0	0
Ap	0	0	0	0	0	0	0	0	0	0	1
An	2	0	0	0	1	0	0	0	0	1	0
E	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0
T	4	4	4	4	4	4	4	4	4	4	4
NN%	75%	0%	0%	0%	25%	0%	0%	0%	0%	25%	25%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6d2. Analysis Software Popularity: Professional, Scientific, and Technical Services Industry, Part 2

Journal of Information Systems Education, 35(3), 325-376, Summer 2024
https://doi.org/10.62273/SPYC4248

Level	Mode	Monkey Learn	**Micro-soft Excel**	Open Refine	Peri-scope Data	Pyramid Analytics	Qlik	Qualtrics	Query-Me	Rapid-Miner	Redash
N	4	3	**0**	4	3	3	4	1	4	3	4
U	0	1	**0**	0	1	1	0	0	0	0	0
Ap	0	0	**1**	0	0	0	0	0	0	0	0
An	0	0	**1**	0	0	0	0	1	0	1	0
E	0	0	**0**	0	0	0	0	1	0	0	0
C	0	0	**2**	0	0	0	0	1	0	0	0
T	4	4	**4**	4	4	4	4	4	4	4	4
NN%	0%	25%	**100%**	0%	25%	25%	0%	75%	0%	25%	0%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6c3. Analysis Software Popularity: Professional, Scientific, and Technical Services Industry, Part 3

Level	Salesforce	Sigma-Computing	Sisense	Talend	Targit	Tilius	Thought-Spot	Whata-graph	Yellowfin BI	Zoho
N	2	3	4	4	3	4	4	4	4	4
U	0	1	0	0	1	0	0	0	0	0
Ap	1	0	0	0	0	0	0	0	0	0
An	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0
T	4	4	4	4	4	4	4	4	4	4
NN%	50%	25%	0%	0%	25%	0%	0%	0%	0%	0%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E6d4. Analysis Software Popularity: Professional, Scientific, and Technical Services Industry, Part 4

Level	Cassandra	DB2	DBMS	flume	Ha-doop	Hbase	Mahoot	Map-Reduce	**Microsoft Access**	Mongo-DB	My-SQL	No-SQL
N	10	8	7	9	9	9	9	7	**1**	7	3	4
U	2	3	4	1	1	1	2	2	**4**	1	3	3
Ap	0	0	4	2	1	1	0	2	**3**	1	3	0
An	3	1	0	2	1	2	2	2	**2**	2	3	4
E	0	3	0	1	1	1	1	1	**2**	3	1	3
C	0	0	0	0	2	1	1	1	**3**	1	2	1
T	15	15	15	15	15	15	15	15	**15**	15	15	15
NN%	33%	47%	53%	40%	40%	40%	40%	53%	**93%**	53%	80%	73%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7a1. Database Software Popularity: Information Industry, Part 1

Level	oozie	Oracle	Postgre-SQL	Presto	shark	Spark	SQL-Server	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
N	8	2	7	7	9	9	2	8	8	4	9	8	8
U	2	2	1	3	4	2	5	3	1	2	3	1	2
Ap	4	3	3	2	1	0	4	0	2	0	0	2	0
An	1	3	2	1	0	3	1	3	0	2	0	0	3
E	0	3	2	1	1	0	1	1	1	3	3	3	2
C	0	2	0	1	0	1	2	0	3	4	0	1	0
T	15	15	15	15	15	15	15	15	15	15	15	15	15
NN%	87%	53%	53%	40%	40%	87%	47%	47%	73%	40%	47%	47%	87%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7a2. Database Software Popularity: Information Industry, Part 2

Level	Cassandra	DB2	DBMS	flume	Hadoop	Hbase	Mahoot	Map-Reduce	**Micro-soft Access**	Mongo-DB	**My-SQL**	No-SQL
N	5	3	2	2	4	3	4	3	**0**	1	**0**	2
U	2	3	2	2	1	1	2	1	**1**	0	**0**	1
Ap	0	1	2	2	1	0	0	0	**3**	2	**2**	1
An	1	0	2	1	2	3	0	3	**1**	4	**3**	2
E	1	2	1	2	1	2	3	1	**2**	1	**2**	2
C	0	0	0	0	0	0	0	1	**2**	1	**2**	1
T	9	9	9	9	9	9	9	9	**9**	9	**9**	9
NN%	44%	67%	78%	78%	56%	67%	56%	67%	**100%**	89%	**100%**	78%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7b1. Database Software Popularity: Data Processing Industry, Part 1

Level	oozie	**Oracle**	Postgre-SQL	Presto	shark	Spark	**SQL-Server**	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
N	4	**0**	2	3	4	2	**0**	4	5	1	4	2	4
U	0	**2**	3	1	3	2	**2**	2	0	2	1	3	1
Ap	3	**0**	0	2	0	2	**1**	0	0	2	0	2	3
An	1	**4**	1	0	1	0	**4**	1	0	1	2	1	0
E	0	**2**	1	2	1	1	**1**	0	2	3	1	0	1
C	1	**1**	2	1	0	2	**1**	2	2	0	1	1	0
T	9	**9**	9	9	9	9	**9**	9	9	9	9	9	9
NN%	56%	**100%**	78%	67%	56%	78%	**100%**	56%	44%	89%	56%	78%	56%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7b2. Database Software Popularity: Data Processing Industry, Part 2

Level	Cassandra	DB2	DBMS	flume	Hadoop	Hbase	Mahoot	Map-Reduce	Microsoft Access	Mongo-DB	My-SQL	No-SQL
N	4	4	4	5	4	5	3	2	2	4	2	4
U	3	0	1	0	1	1	1	3	1	2	1	1
Ap	0	2	1	1	1	0	1	1	2	0	1	2
An	1	0	0	1	1	1	2	2	1	0	1	0
E	0	2	0	1	0	1	0	0	0	2	2	1
C	0	0	2	0	1	0	1	0	2	0	1	0
T	8	8	8	8	8	8	8	8	8	8	8	8
NN%	50%	50%	50%	38%	50%	38%	63%	75%	75%	50%	75%	50%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7c1. Database Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 1

Level	oozie	**Oracle**	Postgre-SQL	Presto	shark	Spark	SQL-Server	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
N	4	**0**	3	3	4	3	2	4	5	4	5	3	5
U	1	**0**	1	2	2	1	1	0	1	0	1	1	0
Ap	1	**3**	4	2	1	1	1	2	0	2	0	2	0
An	0	**2**	0	0	1	0	2	1	1	0	0	2	1
E	0	**2**	0	0	0	1	0	1	1	2	1	0	1
C	2	**1**	0	1	0	2	2	0	0	0	1	0	1
T	8	**8**	8	8	8	8	8	8	8	8	8	8	8
NN%	50%	**100%**	63%	63%	50%	63%	75%	50%	38%	50%	38%	63%	38%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7c2. Database Software Popularity: Computer and Electronic Product Manufacturing Industry, Part 2

Level	Cassandra	DB2	DBMS	flume	Hadoop	Hbase	Mahoot	Map-Reduce	Microsoft Access	Mongo-DB	My-SQL	No-SQL
N	3	3	3	3	3	3	3	3	1	3	1	3
U	0	1	0	0	1	0	1	0	1	1	0	1
Ap	1	0	1	0	0	0	0	1	2	0	0	0
An	0	0	0	1	0	1	0	0	0	0	2	0
E	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	1	0
T	4	4	4	4	4	4	4	4	4	4	4	4
NN%	25%	25%	25%	25%	25%	25%	25%	25%	75%	25%	75%	25%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7d1. Database Software Popularity: Professional, Scientific, and Technical Services Industry, Part 1

Level	oozie	Oracle	Postgre-SQL	Presto	shark	Spark	SQL-Server	Tera-data	tsql	XML	xsd	xsl	Zoo-keeper
N	2	1	1	3	4	3	1	3	3	2	4	3	4
U	0	0	2	0	0	1	1	1	0	1	0	0	0
Ap	1	1	1	1	0	0	0	0	1	0	0	1	0
An	1	1	0	0	0	0	1	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	1	0	0	0	0	1	0	0	1	0	0	0
T	4	4	4	4	4	4	4	4	4	4	4	4	4
NN%	50%	75%	75%	25%	0%	25%	75%	25%	25%	50%	0%	25%	0%

N = None; U = Understanding; Ap = Applying; An = Analyzing; E = Evaluating; C = Creating; T = Total; NN% = Non-Null Percentage

Table E7d2. Database Software Popularity: Professional, Scientific, and Technical Services Industry, Part 2

INFORMATION SYSTEMS & COMPUTING ACADEMIC PROFESSIONALS

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the ***Journal of Information Systems Education*** have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN: 2574-3872 (Online) 1055-3096 (Print)