"Data Analytics and Predictive Modeling"

Job Cluster

Acknowledgements

The development and publication of these skill standards has been a joint and collaborative effort between business and industry representatives and the education community. We are grateful to the industry personnel who participated in the development and validation process. Industry subject matter experts, technical executives, supervisors and technicians donated their time and effort to assure the relevancy of the standards 12 to 36 months into the future.

We gratefully acknowledge funding from the National Science Foundation and the leadership by the team on the IT Skill Standards 2020 and Beyond grant, based at Collin College.

Our leaders are strategically divided into Central, Western, and Eastern teams.

Central

Dr. Ann Beheler, Principal Investigator

Christina Titus, Program Director

Deborah Roberts, Co-Principal Investigator

Helen Sullivan, Senior Staff

West Coast

Terryll Bailey, Co-Principal Investigator

Dr. Suzanne Ames, Co-Principal Investigator

East Coast

Peter Maritato, Co-Principal Investigator

Gordon Snyder, Senior Staff

Data Analytics and Predictive Modeling

The definition for Data Analytics and Predictive Modeling as developed by approximately 100 Thought Leaders (mostly Chief Technology Officers and Chief Information Officers) through three meetings and follow-up surveys to gain consensus is:

> Data Analytics and Predictive Modeling includes inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Business intelligence (BI) specifically focuses on extracting business information for use by decision makers. Common functions of business intelligence include reporting, data mining, process mining, benchmarking, and text mining. This definition was adapted from Wikipedia with input from IT Thought Leaders.

This packet includes…

Job skills as developed by subject matter experts (SMEs) via multiple synchronous meetings (Page 5).

The tasks, knowledge, skills and abilities (KSAs) were developed with a focus 12 to 36 months in the future for an entry-level employee working in that specific cluster.

More specific definitions can be found within the KSA list.

The average was calculated from the subject matter expert votes.

- A vote of "4" indicated the item must be covered in the curriculum.
- A vote of "3" indicated the item should be covered in the curriculum.
- A vote of "2" indicated that it would be nice for the item to be covered in the curriculum.
- A vote of "1" indicated the item should not be covered in the curriculum.

Employability Skills as developed by SMEs via multiple synchronous meetings (Page 9).

Employability competencies are essential for every IT job and are based on what the work requires. SMEs were offered three clearly-defined "levels of proficiency" for each employability skill. The proficiency scale is defined as Level 1 – basic: Level 2- intermediate; and Level 3 - advanced. The levels are cumulative, so a "Level 3" assumes the employee can perform all characteristics of "Level 1" and "Level 2."

For each employability skill, SMEs selected the competency level that best aligned with what would be expected from an entry-level worker for the job cluster in question.

Key Performance Indicators (KPIs) as developed by SMEs (Page 10).

Key Performance Indicators answer the question, "How do we know when a task is performed well?"

A search was performed to locate validated/verified KPIs for technician level work in IT fields. Sources included the Texas Skill Standards System, National Skill Standards Board, National Institute of Standards and Technology and other sources. The identified KPIs were then cross-referenced to the tasks for the

ITSS 2020 job clusters. They were reviewed and revised by a group of the same subject matter experts who developed the tasks and KSAs for the cluster in a structured, facilitated verification session.

Student Learning Outcomes (SLOs) as identified by educators attending the KSA meetings (Page 12).

The SLOs are for use in the creation of curriculum to help define what the students will know and be able to demonstrate. Each of these SLOs can be observed, measured, and demonstrated.

Degree Expectations as identified by educators (Page 16).
A pool of 23 community college and four-year university faculty members from across the country were asked to categorize each knowledge, skill, ability, and task below. The question posed to them: would these KSA+Ts be reasonably included in a two-year AAS program, a four-year Bachelor's program, both, or neither? These results provide another tool for educators to use in assessing how to best incorporate each knowledge, skill, ability, and task.

Data Analytics and Predictive Modeling Tasks and KSAs

	Task SPECIFIC THINGS an entry level person would BE EXPECTED TO PERFORM on the job WITH LITTLE SUPERVISION.	AVG
	Business Problem (Question) Framing	
T-1	Assist in obtaining or receiving problem statement and usability requirements.	2.9
T-2	Assist in determining if the problem is amenable to an analytics solution.	3.1
	Analytics Problem Framing	
T-3	Assist in reformulating the problem statement as an analytics problem.	3.4
T-4	Assist in developing a proposed set of drivers and relationships to outputs.	2.8
T-5	Assist in stating the set of assumptions related to the problem.	3.3
T-6	Assist with collecting metrics and trending data.	3.3
T-7	Assist in obtaining stakeholder agreement on analytical approach by providing detailed analysis.	2.6
	Data	
T-8	Assist with identifying and prioritizing data needs and sources.	3.2
T-9	Assist with assessing the validity of source data and subsequent findings.	3.4
T-10	Assist in acquiring data.	2.9
T-11	Assist in harmonizing, rescaling, cleaning, and sharing data.	3.6
T-12	Assist with identifying relationships in the data.	3.6
T-13	Assist with documenting and reporting findings (e.g., insights, results, business performance).	3.4
T-14	Assist with refining the business and analytics problem statements.	2.9
	Methodology (Approach) Selection	
T-15	Assist with identifying available problem-solving approaches (methods).	3.2
T-16	Assist in conferring with systems analysts, engineers, programmers, and others to design applications.	2.9
T-17	Assist in using basic and contemporary software tools.	3.1
T-18	Assist in reading, interpreting, writing, modifying, and executing simple scripts/code.	3.2
T-19	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	3.3
T-20	Assist in utilizing open-source language or software applications to apply quantitative techniques.	3.3
T-21	Assist with developing and implementing data mining and data programs.	3.4
T-22	Assist with testing approaches (methods).	3.0
T-23	Assist in conducting hypothesis testing using statistical processes.	3.4
T-24	Assist with providing analyses and support for effectiveness assessment.	3.2
	Model Building	
T-25	Assist with selecting approaches (methods).	3.0
T-26	Assist with identifying model structures.	3.2
T-27	Assist in running and evaluating the models.	3.1
T-28	Assist with integrating the models.	2.6
T-29	Assist with documenting and communicating findings (including assumptions, limitations, and constraints).	3.2
T-30	Assist with performing internal business verification and validation of test cases.	2.8
	Deployment	
T-31	Assist with publishing validation and verification report.	2.6
T-32	Assist in developing recommendations to the supervisor based on data analysis and findings.	3.1
T-33	Assist with deploying application codes and analytical models.	2.3
T-34	Assist with presenting technical information to technical and non-technical audiences.	2.8
T-35	Assist with presenting data in creative formats.	3.0
T-36	Assist with delivering reports with findings.	3.1
	Model Lifecycle Management	
T-37	Assist with creating modules and usability via experience and exposure.	2.6
T-38	Assist in supporting deployment.	2.8
T-39	Assist with documenting initial structure (data modeling).	2.8

T-40	Assist in tracking model quality.	2.9
T-41	Assist with providing input and assist in post-action effectiveness assessments.	2.6
T-42	Assist in the identification of information collection shortfalls.	3.1
T-43	Assist with evaluating the business benefit of the model over time.	2.5
T-44	Assist with developing strategic insights from large data sets.	3.3
Knowledge Knowledge focuses on the understanding of concepts. It is theoretical. An individual may have an understanding of a topic or tool or some textbook knowledge of it but have no experience applying it. For example, someone might have read hundreds of articles on health and nutrition, many of them in scientific journals, but that doesn't make that person qualified to dispense advice on nutrition.		
K-1	Knowledge of risk management processes as part of Software Development Life Cycle.	2.7
K-2	Knowledge of computer algorithms.	2.9
K-3	Knowledge of computer programming principles.	3.4
K-4	Knowledge of data administration and data standardization policies.	2.8
K-5	Knowledge of data mining and data management principles.	3.2
K-6	Knowledge of database management systems, query languages, table relationships, and views.	3.7
K-7	Knowledge of a broad and just-in-time understanding of mathematical concepts (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	2.9
K-8	Knowledge of programming language structures and logic.	3.2
K-9	Knowledge of query languages such as SQL (structured query language) and NOSQL.	3.5
K-10	Knowledge of sources, characteristics, and data assets.	2.7
K-11	Knowledge of the various technologies for organizing and managing information (e.g., databases, bookmarking engines).	2.8
K-12	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	2.5
K-13	Knowledge of interpreted and compiled computer languages.	2.7
K-14	Knowledge of how to utilize current popular frameworks and languages.	3.1
K-15	Awareness of machine learning and AI.	2.9
K-16	Knowledge of Personally Identifiable Information (PII) data security standards and how to mask the data.	3.0
K-17	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	3.3
K-18	Knowledge of data mining techniques.	3.0
K-19	Knowledge of relational database framework and why relation databases are used.	3.2
K-20	Knowledge of how to extract, analyze, and use metadata.	3.3
K-21	Awareness of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	3.0
K-22	Knowledge of advanced statistical techniques and concepts and when to apply them.	3.3
K-23	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	3.5
K-24	Knowledge of data analysis concepts.	3.5
K-25	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	3.1
K-26	Knowledge in implementing/developing basic data modules using existing tools.	3.0
K-27	Knowledge in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	3.0
K-28	Knowledge of refining the problem statement and delineate.	2.7
K-29	Knowledge of tuning models and data.	2.9
K-30	Knowledge of how to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	2.9
K-31	Knowledge of ethics as it applies to data analytics and how to apply ethical judgment when policies are not well-defined.	2.8
K-32	Knowledge in determining if a difference in values is significant (statistical and common sense) or not.	3.6

K-33	Knowledge of the landscape of BI tools (Power BI, Google) and data preparation tools and understanding of the data platform associated with each.	3.0

Skills
The capabilities or proficiencies developed through training or hands-on experience. Skills are the practical application of theoretical knowledge. Someone can take a course to gain knowledge of concepts without developing the skills to apply those concepts. Development of skills requires hands-on application of the concepts.

S-1	Skill in conducting queries and developing algorithms to analyze data structures.	3.4
S-2	Skill in creating and utilizing mathematical or statistical models.	3.3
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	3.1
S-4	Skill in using and contributing content to data dictionaries and documentation.	3.0
S-5	Skill in generating queries and reports.	3.6
S-6	Skill in writing code in a currently supported programming language (e.g., Python).	3.3
S-7	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	3.4
S-8	Skill in identifying patterns or relationships.	3.5
S-9	Skill in performing sentiment analysis.	2.8
S-10	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	3.5
S-11	Skill in using data analysis tools (e.g., Excel, Python).	3.7
S-12	Skill in using data mapping tools.	2.9
S-13	Skill in using outlier identification and removal techniques.	3.1
S-14	Skill in writing scripts in contemporary/popular languages.	3.3
S-15	Skill to identify sources, characteristics, and uses of the data assets.	2.9
S-16	Skill in conducting information searches.	3.0
S-17	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	2.7
S-18	Skill in evaluating information for reliability, validity, and relevance.	3.1
S-19	Skill in preparing and presenting briefings.	2.8
S-20	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	2.9
S-21	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools such as ChatGPT in conducting open-source searches.	2.7
S-22	Skill in utilizing feedback to improve processes, products, and services.	3.1
S-23	Skill in performing data analysis including applying statistics.	3.5
S-24	Skill in using statistical / other popular computer languages and frameworks to manipulate.	3.3
S-25	Skill in Visualization using R, Python, or other languages and frameworks.	3.3
S-26	Skill in problem-solving skills and critical thinking ability.	3.7
S-27	Skill in collaboration and communication skills within and across teams.	3.3
S-28	Skill in analytics problem framing (e.g., define geometric sets).	3.1

Abilities
Abilities have historically been used to describe the innate traits or talents that a person brings to a task or situation. Many people can learn to negotiate competently by acquiring knowledge about it and practicing the skills it requires. A few are brilliant negotiators because they have the innate ability to persuade. In reality, abilities may be included under skills or may be separated out.

A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	3.3
A-2	Ability to identify basic common coding flaws at a high level.	2.9
A-3	Ability to use popular/contemporary data visualization tools.	3.2
A-4	Ability to source data used in information, assessment, and/or planning products.	2.9
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	3.2
A-6	Ability to effectively collaborate via virtual teams.	3.3

A-7	Ability to evaluate information for reliability, validity, and relevance.	3.3
A-8	Ability to focus research efforts to meet the customer's decision-making needs.	3.0
A-9	Ability to adapt to a dynamic environment.	3.3
A-10	Ability to function in a collaborative environment, seeking continuous consultation with other analysts and experts—both internal and external to the organization—to leverage analytical and technical expertise.	3.3
A-11	Ability to identify information gaps.	3.2
A-12	Ability to recognize and assist in mitigating cognitive biases which may affect analysis.	2.9
A-13	Ability to recognize and assist in mitigating deception in reporting and analysis.	3.1
A-14	Ability to think critically.	3.7
A-15	Ability to understand objectives and effects.	3.3
A-16	Ability to utilize multiple information sources across all information disciplines.	3.3
A-17	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	3.1
A-18	Ability to understand a business problem.	3.3
A-19	Ability to understand and use the databases and tools to run queries to solve the business problem.	3.5
A-20	Ability to identify patterns.	3.5
A-21	Ability to present and tell the story with data.	3.7
A-22	Ability to drive confidence in numbers they are presenting by indicating probabilities of the numbers being accurate.	3.7

Data Analytics and Predictive Modeling Employability Skills

Workplace Professionalism and Work Ethics	Level 1 - Employee learns expectations of workplace environment (professional behavior and ethics) and adheres to practices with some guidance. Level 2 - Employee exhibits sound professionalism, judgment, and integrity and accepts responsibility for own behavior. Employee exhibits these qualities without guidance but occasionally refers to policies as needed.
Written Communication	Level 1 - Employee understands written instructions and executes tasks with guidance and feedback from supervisor. Employee clearly communicates concepts in writing. Level 2 - Employee comprehends and executes written instructions with minimal guidance. Employee composes well-organized written documents.
Oral Communication	Level 1 - Employee understands oral instructions and executes tasks with guidance and feedback from supervisor. Employee communicates concepts orally while clarifying for meaning. Employee develops listening skills. Level 2 - Employee comprehends and executes oral instructions with minimal guidance and exhibits good listening skills. Employee clarifies for meaning without needing prompting from supervisor.
Teamwork	Level 1 - With guidance and feedback from supervisor, employee obeys team rules and understands team member roles. Employee actively participates in team activities, volunteers for special tasks, and establishes rapport with co-workers.
Problem Solving & Critical Thinking	Level 1 - Employee identifies the problem and relevant facts and principles with guidance and feedback from supervisor. Employee summarizes existing ideas and demonstrates creative thinking process while problem solving.
Organization and Planning	Level 1 - Employee prepares schedule for self, monitors and adjusts task sequence, and analyzes work assignments with guidance from supervisor. Level 2 - Employee manages timelines and recommends timeline adjustments. Employee escalates timeline-impacting issues as appropriate.
Adaptability and Flexibility	Level 1 - With guidance and feedback from supervisor, employee is able to adjust ways of doing work based on changing dynamics. Working under pressure is difficult, but employee makes it through the project with guidance and oversight.
Initiative	Level 1 - Employee finishes a step in a project and waits for direction before going on to the next step. Level 2 - Employee finishes multiple steps in a project and appropriately begins working on the next step without being asked.
Accuracy	Level 1 - Employee makes mistakes routinely but is committed to learning to adjust work habits to prevent them in the future. Level 2 - Employee occasionally makes mistakes but quickly makes adjustments to work habits to avoid making the same mistake twice.
Cultural Competence	Level 1 - Employee is inexperienced with working with diverse teams. With support and guidance and getting to know team members, employee develops working relationships. Level 2 - Employee is committed to working with diverse teams but struggles when differences arise. Employee identifies those challenges and works with colleagues to find ways to work effectively.
Self and Career Development	Level 1 - Employee requires feedback and direction from supervisor regarding improvement needed in professional and technical skills. Employee follows through with skills development with monitoring by supervisor.

Data Analytics and Predictive Modeling Key Performance Indicators

For the entry-level employee, all tasks are typically done under supervision for much of the first year and then with some

	Task	Key Performance Indicators
Business Problem (Question) Framing		
T-1	Assist in obtaining or receiving problem statement and usability requirements.	Problem statement and usability requirements are obtained in a timely manner and properly documented.
T-2	Assist in determining if the problem is amenable to an analytics solution.	Determination of the applicability of an analytics solution is accurate. Business and analytics problem statements are clear, and are continuously refined.
Analytics Problem Framing		
T-3	Assist in reformulating the problem statement as an analytics problem.	The alternatives to the analytics problem statement are documented and ranked according to best match with current problem and rationale for choices is clearly stated. Assumptions related to the problem are stated clearly and concisely. Criteria for success are clearly identified. Accurate and relevant analysis is provided regarding analytical approach.
T-4	Assist in developing a proposed set of drivers and relationships to outputs.	
T-5	Assist in stating the set of assumptions related to the problem.	
T-6	Assist with collecting metrics and trending data.	
T-7	Assist in obtaining stakeholder agreement on analytical approach by providing detailed analysis.	
Data		
T-8	Assist with identifying and prioritizing data needs and sources.	Sources and methods for acquiring data are efficient and information is accurate and complete. Data is secured from reliable and respected sources. Data is correctly harmonized, rescaled, and cleaned and relationships in the data are correctly identified. Findings are documented and communicated in a clear and timely manner; company procedures may vary. Data definitions are fully developed and agreed upon in accordance with company procedures.
T-9	Assist with assessing the validity of source data and subsequent findings.	
T-10	Assist in acquiring data.	
T-11	Assist in harmonizing, rescaling, cleaning, and sharing data.	
T-12	Assist with identifying relationships in the data.	
T-13	Assist with documenting and reporting findings (e.g., insights, results, business performance).	
T-14	Assist with refining the business and analytics problem statements.	
Methodology (Approach) Selection		
T-15	Assist with identifying available problem-solving approaches (methods).	Sources and methods for acquiring data are efficient and information is accurate and complete. The alternatives to the methodology are documented and ranked. Data is secured from reliable and respected sources. Findings are documented and communicated in a clear and timely manner; company procedures may vary. Data definitions are fully developed and agreed upon in accordance with company procedures. Problem-solving approaches and methods are affordable and relevant. Analysis processes and conclusions are clearly and concisely documented. Effective software tools and problem-solving methods are used. Scripts are complete, relevant, and congruent. Appropriate testing methodology is identified and planned, and scope of testing is clearly identified. Algorithms, programming principles, statistical processes are used correctly.
T-16	Assist in conferring with systems analysts, engineers, programmers, and others to design applications.	
T-17	Assist in using basic and contemporary software tools.	
T-18	Assist in reading, interpreting, writing, modifying, and executing simple scripts/code.	
T-19	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	
T-20	Assist in utilizing open-source language or software applications to apply quantitative techniques.	
T-21	Assist with developing and implementing data mining and data programs.	
T-22	Assist with testing approaches (methods).	
T-23	Assist in conducting hypothesis testing using statistical processes.	
T-24	Assist with providing analyses and support for effectiveness assessment.	
Model Building		
T-25	Assist with selecting approaches (methods).	Models are evaluated, and integrated using the proper procedures. Data model is laid out clearly. Performance criteria for the data model have verifiable assumptions. Scope and purpose of model are defined. Code is developed using efficient software design processes. Reusable components are employed whenever possible. Code is well documented so that it can be understood by others. Tests accurately assess the functions the module is designed to perform. Ethics reviews are routinely accomplished.
T-26	Assist with identifying model structures.	
T-27	Assist in running and evaluating the models.	
T-28	Assist with integrating the models.	
T-29	Assist with documenting and communicating findings (including assumptions, limitations, and constraints).	
T-30	Assist with performing internal business verification and validation of test cases.	

	Deployment	
T-31	Assist with publishing validation and verification report.	Presentations are well-organized, utilize creative formats, and meet the needs of technical and non-technical audiences. Enterprise goals are taken into account when drawing conclusions from data analysis and making recommendations to supervisor. Model and usability are evaluated in accordance with company procedures; company procedures may vary.
T-32	Assist in developing recommendations to the supervisor based on data analysis and findings.	
T-33	Assist with deploying application codes and analytical models.	
T-34	Assist with presenting technical information to technical and non-technical audiences.	
T-35	Assist with presenting data in creative formats.	
T-36	Assist with delivering reports with findings.	
	Model Lifecycle Management	
T-37	Assist with creating modules and usability via experience and exposure.	Initial structure of the model (data modeling) is documented in accordance with company standards. Tracking of model quality and model recalibration and maintenance are conducted in a timely manner. Effectiveness testing is based on specification criteria. Recommendations are fed back into the modeling process. Computer data administration, data standardization, data mining, and data management are conducted in accordance with industry and company procedures and standards.
T-38	Assist in supporting deployment.	
T-39	Assist with documenting initial structure (data modeling).	
T-40	Assist in tracking model quality.	
T-41	Assist with providing input and assist in post-action effectiveness assessments.	
T-42	Assist in the identification of information collection shortfalls.	
T-43	Assist with evaluating the business benefit of the model over time.	
T-44	Assist with developing strategic insights from large data sets.	

Data Analytics and Predictive Modeling Student Learning Outcomes

	Knowledge	Student Learning Outcomes
K-1	Knowledge of risk management processes as part of Software Development Life Cycle.	Explain information security fundamentals. Demonstrate an understanding of the importance of ethics and privacy with data. Describe the functions of database recovery, security and administration, and basic data warehousing concepts.
K-16	Knowledge of Personally Identifiable Information (PII) data security standards and how to mask the data.	
K-4	Knowledge of data administration and data standardization policies.	Describe the principles, techniques, and business policies for collecting, organizing, managing, analyzing, and reporting data. Describe the process of data science analytics from data acquisition to recommendations based on data. Describe different methods and tools for data collection and their impact on analysis of data. Identify the concepts of the relational model, normalization, dependencies, integrity, and constraints.
K-11	Knowledge of the various technologies for organizing and managing information (e.g., databases, bookmarking engines).	
K-5	Knowledge of data mining and data management principles.	
K-18	Knowledge of data mining techniques.	
K-24	Knowledge of data analysis concepts.	
K-2	Knowledge of computer algorithms.	Apply the basics of programming principles. Demonstrate problem-solving skills by developing and implementing algorithms to solve problems.
K-3	Knowledge of computer programming principles.	
K-8	Knowledge of programming language structures and logic.	
K-12	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	
K-13	Knowledge of interpreted and compiled computer languages.	
K-6	Knowledge of database management systems, query languages, table relationships, and views.	Identify and explain concepts and frameworks relating to different major types of database systems, including major query languages. Understand query languages and how to use some of the popular query languages such as MySQL.
K-19	Knowledge of relational database framework and why relational databases are used.	
K-23	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	
K-9	Knowledge of query languages such as SQL (structured query language) and NOSQL.	
K-7	Knowledge of a broad and just-in-time understanding of mathematical concepts (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	Select appropriate mathematical and statistical tools used for data analytics.
K-22	Knowledge of advanced statistical techniques and concepts and when to apply them.	
K-10	Knowledge of sources, characteristics, and data assets.	Describe the data acquisition process. Explain data warehousing architectures, processes, and operations.
K-20	Knowledge of how to extract, analyze, and use metadata.	
K-14	Knowledge of how to utilize current popular frameworks and languages.	Describe tools and techniques to store and process data.
K-17	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	Compare and contrast different tools available for gathering and analyzing data. Explain existing tools available for developing data modules.
K-33	Knowledge of the landscape of BI tools (Power BI, Google) and data preparation tools and understanding of the data platform associated with each.	
K-26	Knowledge in implementing/developing basic data modules using existing tools.	
K-29	Knowledge of tuning models and data.	Discuss various aspects of data including how to determine when data is statistically significant. Demonstrate how to use various tuning models and how data tuning works.
K-32	Knowledge in determining if a difference in values is significant (statistical and common sense) or not.	
K-25	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	Select AI tools to analyze data and differentiate between ethical and unethical use of AI. With respect to data analysis, explain what is meant by ethical judgment and the factors to consider when applying ethical judgment. Differentiate knowing right from wrong.
K-31	Knowledge of ethics as it applies to data analytics and how to apply ethical judgment when policies are not well-defined.	

K-27	Knowledge in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	Explain the differences between various types of Regression Analysis Tools (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).
K-28	Knowledge of refining the problem statement and delineate.	Explain how to refine a problem statement, delineating the steps required to do so.
K-30	Knowledge of how to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	Summarize project proposal analysis and experience writing project design documents as well as project life cycle tracking.
K-21	Awareness of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	Discuss the pros and cons of different machine learning techniques in real-world applications.

Skills		**Student Learning Outcomes**
S-1	Skill in conducting queries and developing algorithms to analyze data structures.	Perform queries and develop reports.
S-5	Skill in generating queries and reports.	
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	Create data models and use data mining techniques, models, and tools using data dictionaries, mapping tools, and search tools—document results. Cleanse and prepare data for analysis. Identify and address outliers.
S-4	Skill in using and contributing content to data dictionaries and documentation.	
S-12	Skill in using data mapping tools.	
S-13	Skill in using outlier identification and removal techniques.	
S-7	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	
S-8	Skill in identifying patterns or relationships.	Develop or discover analytical patterns from data models.
S-9	Skill in performing sentiment analysis.	
S-10	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	Perform data analysis using statistical models and tools. Demonstrate knowledge of statistical and mathematical models by analyzing a problem and formulating solution using geometric sets. Develop data analysis models using statistics for optimal solution and provide documentation explaining your results.
S-23	Skill in performing data analysis including applying statistics.	
S-2	Skill in creating and utilizing mathematical or statistical models.	
S-28	Skill in analytics problem framing (e.g., define geometric sets).	
S-20	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	
S-11	Skill in using data analysis tools (e.g., Excel, Python).	Demonstrate coding and scripting techniques using data analytics programming languages. Use appropriate programming language, data structures, and concepts to solve data analysis problems. Build arrays, data frames, dictionaries, and perform basic calculations using programming languages. Develop a software statistical modeling project and present the solution. Use data visualization tools to analyze data and produce dashboards.
S-6	Skill in writing code in a currently supported programming language (e.g., Python).	
S-14	Skill in writing scripts in contemporary/popular languages.	
S-24	Skill in using statistical/other popular computer languages and frameworks to manipulate.	
S-25	Skill in Visualization using R, Python, or other languages and frameworks.	
S-15	Skill to identify sources, characteristics, and uses of the data assets.	Interpret and analyze the data relevance, reliability, and validity from multiple sources. Integrate the business intelligence methodology and concepts and relate them to decision support.
S-21	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools such as ChatGPT in conducting open-source searches.	
S-16	Skill in conducting information searches.	
S-17	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	
S-18	Skill in evaluating information for reliability, validity, and relevance.	
S-22	Skill in utilizing feedback to improve processes, products, and services.	Demonstrate effective collaboration and communication skills to improve team productivity. Derive problem specifications from problem statements.
S-26	Skill in problem-solving skills and critical thinking ability.	
S-27	Skill in collaboration and communication skills within and across teams.	

	Abilities	Student Learning Outcomes
S-19	Skill in preparing and presenting briefings.	Demonstrate presentation preparation and delivery skill by preparing a well-balanced briefing of a project showing comprehensive analysis and outcomes.
A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	Solve business problems by identifying data gaps and synthesizing data to deliver quality output. Structure SQL and QBE commands to define, query, and manipulate a relational database. Apply databases to actual situations and business problems.
A-4	Ability to source data used in information, assessment, and/or planning products.	
A-7	Ability to evaluate information for reliability, validity, and relevance.	
A-11	Ability to identify information gaps.	
A-18	Ability to understand a business problem.	
A-20	Ability to identify patterns.	
A-13	Ability to recognize and mitigate deception in reporting and analysis.	
A-16	Ability to utilize multiple information sources across all information disciplines.	
A-2	Ability to identify basic common coding flaws at a high level.	Use data analytics programming and visualization tools. Use data visualization tools to analyze data and produce reports. Develop steps to design and implement a dashboard.
A-3	Ability to use popular/contemporary data visualization tools.	
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	Collaborate, communicate, and listen effectively to explain and solve business problems to a diverse audience. Develop algorithms using modular design principles to meet stated specifications. Identify, evaluate, and suggest solutions to problems encountered in a team communication context.
A-6	Ability to effectively collaborate via virtual teams.	
A-9	Ability to adapt to a dynamic environment.	
A-10	Ability to function in a collaborative environment, seeking continuous consultation with other analysts and experts—both internal and external to the organization—to leverage analytical and technical expertise.	
A-17	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	
A-14	Ability to think critically.	
A-12	Ability to recognize and mitigate cognitive biases which may affect analysis.	Maintain high standards of professional competence, conduct, and ethical practice.
A-8	Ability to focus research efforts to meet the customer's decision-making needs.	Competently identify and evaluate data relevance, reliability, and validity from multiple sources to meet customer's needs. Research and utilize validated data to logically construct a report based on customer's needs.
A-15	Ability to understand objectives and effects.	
A-19	Ability to understand and use the databases and tools to run queries to solve the business problem.	Demonstrate understanding of databases and query tools by completing multiple projects using realistic data sets (i.e., formulate a project story).
A-21	Ability to present and tell the story with data.	Run various queries and present results in story format so that concepts and results can be easily understood. Explain the problem to be solved, how it was solved, and what are the results.
A-22	Ability to drive confidence in numbers they are presenting by indicating probabilities of the numbers being accurate.	Confidently use presentation skills that provide an accurate depiction of your analysis. Demonstrate your confidence in your analysis by providing an accuracy analysis (What is the probability of your analysis being accurate. It could be in the form of a percentage).

Data Analytics Degree Expectations

A pool of 23 community college and four-year university faculty members from across the country were asked to categorize each knowledge, skill, ability, and task below. The question posed to them: would these KSA+Ts be reasonably included in a two-year AAS program, a four-year Bachelor's program, both, or neither? These results provide another tool for educators to use in assessing how to best incorporate each knowledge, skill, ability, and task.

	Tasks	2 Year AAS	Both 2 yr AAS and 4 yr Academic Degree	4 Year Academic Degree	Number of responses
		% Best Estimate			
T-1	Assist in obtaining or receiving problem statement and usability requirements.	35%	57%	9%	23
T-2	Assist in determining if the problem is amenable to an analytics solution.	22%	52%	26%	23
T-3	Assist in reformulating the problem statement as an analytics problem.	22%	35%	44%	23
T-4	Assist in developing a proposed set of drivers and relationships to outputs.	14%	41%	46%	22
T-5	Assist in stating the set of assumptions related to the problem.	23%	55%	23%	22
T-6	Assist with collecting metrics and trending data.	30%	57%	13%	23
T-7	Assist in obtaining stakeholder agreement on analytical approach by providing detailed analysis.	9%	36%	55%	22
T-8	Assist with identifying and prioritizing data needs and sources.	17%	61%	22%	23
T-9	Assist with assessing the validity of source data and subsequent findings.	17%	61%	22%	23
T-10	Assist in acquiring data.	30%	65%	4%	23
T-11	Assist in harmonizing, rescaling, cleaning, and sharing data.	23%	50%	27%	22
T-12	Assist with identifying relationships in the data.	18%	55%	27%	22
T-13	Assist with documenting and reporting findings (e.g., insights, results, business performance).	13%	57%	30%	23
T-14	Assist with refining the business and analytics problem statements.	13%	22%	65%	23
T-15	Assist with identifying available problem-solving approaches (methods).	17%	65%	17%	23
T-16	Assist in conferring with systems analysts, engineers, programmers, and others to design applications.	5%	29%	67%	21
T-17	Assist in using basic and contemporary software tools.	26%	70%	4%	23
T-18	Assist in reading, interpreting, writing, modifying, and executing simple scripts/code.	35%	61%	4%	23
T-19	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	26%	61%	13%	23
T-20	Assist in utilizing open-source language or software applications to apply quantitative techniques.	14%	41%	46%	22
T-21	Assist with developing and implementing data mining and data programs.	24%	29%	48%	21
T-22	Assist with testing approaches (methods).	14%	57%	29%	21
T-23	Assist in conducting hypothesis testing using statistical processes.	14%	23%	64%	22
T-24	Assist with providing analyses and support for effectiveness assessment.	15%	25%	60%	20
T-25	Assist with selecting approaches (methods).	10%	48%	43%	21
T-26	Assist with identifying model structures.	14%	32%	55%	22
T-27	Assist in running and evaluating the models.	18%	46%	36%	22
T-28	Assist with integrating the models.	9%	46%	46%	22
T-29	Assist with documenting and communicating findings (including assumptions, limitations and constraints).	18%	55%	27%	23

T-30	Assist with performing internal business verification and validation of test cases.	9%	44%	48%	23
T-31	Assist with publishing validation and verification report.	9%	35%	57%	23
T-32	Assist in developing recommendations to the supervisor based on data analysis and findings.	14%	55%	32%	23
T-33	Assist with deploying application codes and analytical models.	13%	35%	52%	23
T-34	Assist with presenting technical information to technical and non-technical audiences.	10%	62%	29%	21
T-35	Assist with presenting data in creative formats.	13%	70%	17%	23
T-36	Assist with delivering reports with findings.	22%	70%	9%	23
T-37	Assist with creating modules and usability via experience and exposure.	10%	52%	38%	21
T-38	Assist in supporting deployment.	22%	74%	4%	23
T-39	Assist with documenting initial structure (data modeling).	14%	55%	32%	22
T-40	Assist in tracking model quality.	14%	36%	50%	22
T-41	Assist with providing input and assist in post-action effectiveness assessments.	14%	29%	57%	21
T-42	Assist in the identification of information collection shortfalls.	17%	35%	48%	23
T-43	Assist with evaluating the business benefit of the model over time.	13%	26%	61%	23
T-44	Assist with developing strategic insights from large data sets.	14%	32%	55%	22
Knowledge					
K-1	Knowledge of risk management processes as part of Software Development Life Cycle.	4%	61%	35%	23
K-2	Knowledge of computer algorithms.	9%	65%	26%	23
K-3	Knowledge of computer programming principles.	17%	70%	13%	23
K-4	Knowledge of data administration and data standardization policies.	9%	59%	32%	22
K-5	Knowledge of data mining and data management principles.	9%	55%	36%	22
K-6	Knowledge of database management systems, query languages, table relationships, and views.	13%	83%	4%	23
K-7	Knowledge of a broad and just-in-time understanding of mathematical concepts (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	5%	23%	73%	22
K-8	Knowledge of programming language structures and logic.	13%	74%	13%	23
K-9	Knowledge of query languages such as SQL (structured query language) and NOSQL.	17%	70%	13%	23
K-10	Knowledge of sources, characteristics, and data assets.	13%	70%	17%	23
K-11	Knowledge of the various technologies for organizing and managing information (e.g., databases, bookmarking engines).	17%	78%	4%	23
K-12	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	13%	70%	17%	23
K-13	Knowledge of interpreted and compiled computer languages.	5%	68%	27%	22
K-14	Knowledge of how to utilize current popular frameworks and languages.	14%	77%	9%	22
K-15	Awareness of machine learning and AI.	14%	68%	18%	22
K-16	Knowledge of Personally Identifiable Information (PII) data security standards and how to mask the data.	9%	73%	18%	22
K-17	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	13%	70%	17%	23
K-18	Knowledge of data mining techniques.	10%	57%	33%	21
K-19	Knowledge of relational database framework and why relation databases are used.	13%	83%	4%	23
K-20	Knowledge of how to extract, analyze, and use metadata.	14%	67%	19%	21
K-21	Awareness of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	5%	24%	71%	21

K-22	Knowledge of advanced statistical techniques and concepts and when to apply them.	0%	4%	96%	23
K-23	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	18%	41%	41%	22
K-24	Knowledge of data analysis concepts.	17%	74%	9%	23
K-25	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	9%	61%	30%	23
K-26	Knowledge in implementing/developing basic data modules using existing tools.	18%	64%	18%	22
K-27	Knowledge in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	0%	24%	76%	21
K-28	Knowledge of refining the problem statement and delineate.	4%	70%	26%	23
K-29	Knowledge of tuning models and data.	9%	36%	55%	22
K-30	Knowledge of how to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	0%	46%	55%	22
K-31	Knowledge of ethics as it applies to data analytics and how to apply ethical judgment when policies are not well-defined.	9%	64%	27%	22
K-32	Knowledge in determining if a difference in values is significant (statistical and common sense) or not.	5%	68%	27%	22
K-33	Knowledge of the landscape of BI tools (power BI, Google) and data preparation tools (pull from recording) and understanding of the data platform associated with each.	14%	62%	24%	21
Skills					
S-1	Skill in conducting queries and developing algorithms to analyze data structures.	14%	68%	18%	22
S-2	Skill in creating and utilizing mathematical or statistical models.	9%	35%	57%	23
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	17%	52%	30%	23
S-4	Skill in using and contributing content to data dictionaries and documentation.	9%	65%	26%	23
S-5	Skill in generating queries and reports.	22%	78%	0%	23
S-6	Skill in writing code in a currently supported programming language (e.g., Python).	13%	83%	4%	23
S-7	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	5%	57%	38%	21
S-8	Skill in identifying patterns or relationships.	9%	74%	17%	23
S-9	Skill in performing sentiment analysis.	5%	35%	60%	20
S-10	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	17%	61%	22%	23
S-11	Skill in using data analysis tools (e.g., Excel, Python).	22%	70%	9%	23
S-12	Skill in using data mapping tools.	10%	52%	38%	21
S-13	Skill in using outlier identification and removal techniques.	14%	41%	46%	22
S-14	Skill in writing scripts in contemporary/popular languages.	18%	77%	5%	22
S-15	Skill to identify sources, characteristics, and uses of the data assets.	9%	68%	23%	22
S-16	Skill in conducting information searches.	13%	78%	9%	23
S-17	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	4%	22%	74%	23
S-18	Skill in evaluating information for reliability, validity, and relevance.	4%	65%	30%	23
S-19	Skill in preparing and presenting briefings.	9%	78%	13%	23
S-20	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	10%	50%	40%	20

S-21	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools such as ChatGPT in conducting open-source searches.	17%	74%	9%	23
S-22	Skill in utilizing feedback to improve processes, products, and services.	10%	71%	19%	21
S-23	Skill in performing data analysis including applying statistics.	13%	48%	39%	23
S-24	Skill in using statistical / other popular computer languages and frameworks to manipulate.	9%	57%	35%	23
S-25	Skill in Visualization using R, Python, or other languages and frameworks.	9%	73%	18%	22
S-26	Skill in problem-solving skills and critical thinking ability.	4%	83%	13%	23
S-27	Skill in collaboration and communication skills within and across teams.	0%	87%	13%	23
S-28	Skill in analytics problem framing (e.g., define geometric sets).	0%	33%	67%	21
Abilities					
A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	5%	41%	55%	22
A-2	Ability to identify basic common coding flaws at a high level.	4%	57%	39%	23
A-3	Ability to use popular/contemporary data visualization tools.	9%	70%	22%	23
A-4	Ability to source data used in information, assessment, and/or planning products.	10%	48%	43%	21
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	4%	48%	48%	23
A-6	Ability to effectively collaborate via virtual teams.	9%	87%	4%	23
A-7	Ability to evaluate information for reliability, validity, and relevance.	0%	73%	27%	22
A-8	Ability to focus research efforts to meet the customer's decision-making needs.	5%	50%	46%	22
A-9	Ability to adapt to a dynamic environment.	9%	78%	13%	23
A-10	Ability to function in a collaborative environment, seeking continuous consultation with other analysts and experts—both internal and external to the organization—to leverage analytical and technical expertise.	9%	57%	35%	23
A-11	Ability to identify information gaps.	9%	68%	23%	22
A-12	Ability to recognize and assist in mitigating cognitive biases which may affect analysis.	5%	38%	57%	21
A-13	Ability to recognize and assist in mitigating deception in reporting and analysis.	5%	48%	48%	21
A-14	Ability to think critically.	9%	91%	0%	23
A-15	Ability to understand objectives and effects.	9%	91%	0%	22
A-16	Ability to utilize multiple information sources across all information disciplines.	5%	82%	14%	22
A-17	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	0%	78%	22%	23
A-18	Ability to understand a business problem.	13%	70%	17%	23
A-19	Ability to understand and use the databases and tools to run queries to solve the business problem.	22%	70%	9%	23
A-20	Ability to identify patterns.	13%	74%	13%	23
A-21	Ability to present and tell the story with data.	13%	61%	26%	23
A-22	Ability to drive confidence in numbers they are presenting by indicating probabilities of the numbers being accurate.	9%	55%	36%	22