

## Data Analytics and Predictive Modeling Tasks and KSAs

Task		AVG
SPECIFIC THINGS an entry level person would BE EXPECTED TO PERFORM on the job WITH LITTLE SUPERVISION.		
<b>Business Problem (Question) Framing</b>		
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
<b>Analytics Problem Framing</b>		
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
<b>Data</b>		
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
<b>Methodology (Approach) Selection</b>		
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
<b>Model Building</b>		
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
<b>Deployment</b>		
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
<b>Model Lifecycle Management</b>		
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
<b>Knowledge</b>		
<p>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.</p>		
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
<b>Skills</b>		
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
<b>Abilities</b>		
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



This material is based upon work supported by the National Science Foundation under Grant No. 1838535. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.