

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.		
Business Problem (Question) Framing		
T-1	Assist in obtaining or receiving problem statement and usability requirements.	3.6
T-2	Assist in identifying stakeholders.	3.0
T-3	Assist in determining if the problem is amenable to an analytics solution.	3.3
T-4	Assist in refining the problem statement and delineate.	3.2
T-5	Assist in defining an initial set of business benefits.	3.2
T-6	Assist in obtaining stakeholder agreement on the problem.	3.1
Analytics Problem Framing		
T-7	Assist in reformulating the problem statement as an analytics problem.	3.4
T-8	Assist in developing a proposed set of drivers and relationships to outputs.	3.1
T-9	Assist in stating the set of assumptions related to the problem.	3.3
T-10	Assist in defining key metrics of success.	3.4
T-11	Assist with collecting metrics and trending data.	3.2
T-12	Assist in obtaining stakeholder agreement on analytical approach.	2.9
Data		
T-13	Assist with identifying and prioritizing data needs and sources.	3.3
T-14	Assist with assessing the validity of source data and subsequent findings.	3.0
T-15	Assist in acquiring data.	3.3
T-16	Assist in harmonizing, rescaling, cleaning, and sharing data.	3.5
T-17	Assist with identifying relationships in the data.	3.3
T-18	Assist with documenting and reporting findings (e.g., insights, results, business performance).	3.5
T-19	Assist with refining the business and analytics problem statements.	3.2
Methodology (Approach) Selection		
T-20	Assist with identifying available problem solving approaches (methods).	3.3
T-21	Assist in conferring with systems analysts, engineers, programmers, and others to design application.	2.8
T-22	Assist in using software tools.	2.8
T-23	Assist in reading, interpreting, writing, modifying, and executing simple scripts (e.g., Perl, VBScript) on Windows and UNIX systems (e.g., those that perform tasks such as: parsing large data files, automating manual tasks, and fetching/processing remote data).	3.2
T-24	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	3.2
T-25	Assist in utilizing open source language such as R and apply quantitative techniques (e.g., descriptive and inferential statistics, sampling, experimental design, parametric and non-parametric tests of difference, ordinary least squares regression, general line).	3.2
T-26	Assist with developing and implementing data mining and data programs.	2.7
T-27	Assist with testing approaches (methods).	3.4
T-28	Assist in conducting hypothesis testing using statistical processes.	3.1
T-29	Assist with providing analyses and support for effectiveness assessment.	2.8
T-30	Assist with selecting approaches (methods).	3.3
Model Building		
T-31	Assist with identifying model structures.	3.1
T-32	Assist in running and evaluating the models.	3.4
T-33	Assist with tuning models and data.	3.2
T-34	Assist with integrating the models.	2.9
T-35	Assist with documenting and communicating findings (including assumptions, limitations and constraints).	3.7
T-36	Assist with performing internal business verification and validation of the model.	3.0
T-37	Assist with publishing validation and verification report.	2.9
T-38	Assist in developing recommendations to the supervisor based on data analysis and findings.	3.1
Deployment		

T-39	Assist with deploying application codes and analytical models using CI/CD tools and techniques and provides support for deployed data applications and analytical models.	2.6
T-40	Assist with performing business validation of the model.	3.1
T-41	Assist with presenting technical information to technical and nontechnical audiences.	3.3
T-42	Assist with presenting data in creative formats.	2.8
T-43	Assist with delivering reports with findings.	3.5
T-44	Assist with creating model, usability, and system requirements for production.	2.8
T-45	Assist in supporting deployment.	2.8
Model Lifecycle Management		
T-46	Assist with documenting initial structure.	3.0
T-47	Assist in tracking model quality.	3.3
T-48	Assist with providing input and assist in post-action effectiveness assessments.	2.7
T-49	Assist in the identification of information collection shortfalls.	2.8
T-50	Assist with recalibrating and maintaining the model.	3.0
T-51	Assist with evaluating the business benefit of the model over time.	2.9
T-52	Assist with developing strategic insights from large data sets.	2.9
Knowledge		
<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 (e.g., methods for assessing and mitigating risk).	2.7
K-2	Knowledge of computer algorithms.	3.3
K-3	Knowledge of computer programming principles.	3.2
K-4	Knowledge of data administration and data standardization policies.	2.9
K-5	Knowledge of data mining and data management principles.	3.0
K-6	Knowledge of database management systems, query languages, table relationships, and views.	3.1
K-7	Knowledge of mathematics (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	3.3
K-8	Knowledge of programming language structures and logic.	3.2
K-9	Knowledge of query languages such as SQL (structured query language).	3.5
K-10	Knowledge of sources, characteristics, and uses of the organization's data assets.	2.9
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).	2.7
K-13	Knowledge of interpreted and compiled computer languages.	2.7
K-14	Knowledge of how to utilize Hadoop, Java, Python, SQL, Hive, and Pig to explore data.	3.2
K-15	Knowledge of machine learning theory and principles.	3.3
K-16	Knowledge of data classification standards and methodologies based on sensitivity and other risk factors.	2.7
K-17	Knowledge of Personally Identifiable Information (PII) data security standards.	2.9
K-18	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	3.1
K-19	Knowledge of data mining techniques.	3.4
K-20	Knowledge of database theory.	2.8
K-21	Knowledge of how to extract, analyze, and use metadata.	3.0
K-22	Knowledge on ETL techniques, Hadoop, Data analytics, Big data is an advantage.	2.9
K-23	Knowledge of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	3.2
K-24	Knowledge of advanced statistical techniques and concepts (regression, properties of distributions, statistical tests and proper usage, etc.) and experience with applications.	3.2
K-25	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	2.8
K-26	Knowledge of Decision Science Game theory.	2.9

K-27	Knowledge of the use of simulation.	3.1
K-28	Knowledge of optimization.	3.3
K-29	Knowledge of data analysis concepts.	3.6
K-30	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	3.1
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.	3.5
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.	3.3
S-4	Skill in using and contributing content to data dictionaries.	2.7
S-5	Skill in developing data models.	3.0
S-6	Skill in generating queries and reports.	3.5
S-7	Skill in writing code in a currently supported programming language (e.g., Python).	2.9
S-8	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	3.0
S-9	Skill in identifying patterns or relationships.	3.1
S-10	Skill in performing sentiment analysis.	3.3
S-11	Skill in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	3.1
S-12	Skill in supporting transformation analytics to invoke a business shift.	2.9
S-13	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	3.4
S-14	Skill in using data analysis tools (e.g., Excel, Python).	3.3
S-15	Skill in using data mapping tools.	2.9
S-16	Skill in using outlier identification and removal techniques.	3.3
S-17	Skill in writing scripts using R, Python, PIG, HIVE, SQL, etc.	3.5
S-18	Skill to identify sources, characteristics, and uses of the organization's data assets.	2.8
S-19	Skill in conducting information searches.	3.0
S-20	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	2.8
S-21	Skill in evaluating information for reliability, validity, and relevance.	2.9
S-22	Skill in preparing and presenting briefings.	3.1
S-23	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	2.9
S-24	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools in conducting open-source searches.	2.8
S-25	Skill in utilizing feedback to improve processes, products, and services.	3.1
S-26	Skill in performing data analysis including applying statistics.	3.6
S-27	Skill in using statistical computer languages (R, Python, etc.) to manipulate data and draw insights from large data sets.	3.5
S-28	Skill in Visualization using R, Python, or other languages and frameworks.	3.4
S-29	Skill in problem-solving skills and critical thinking ability.	3.6
S-30	Skill in collaboration and communication skills within and across teams.	3.6
S-31	Skill in analytics problem framing (e.g., define geometric sets).	3.5
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.	3.2
A-2	Ability to identify basic common coding flaws at a high level.	3.0
A-3	Ability to use data visualization tools (e.g., Flare, HighCharts, AmCharts, D3.js, Processing, Google Visualization API, Tableau, Raphael.js).	3.3

A-4	Ability to source data used in information, assessment, and/or planning products.	2.7
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	3.5
A-6	Ability to develop or recommend analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	3.0
A-7	Ability to evaluate, analyze, and synthesize large quantities of data (which may be fragmented and contradictory) into quality, fused targeting/information products.	2.9
A-8	Ability to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	2.9
A-9	Ability to effectively collaborate via virtual teams.	3.3
A-10	Ability to evaluate information for reliability, validity, and relevance.	3.3
A-11	Ability to exercise strong ethical judgment when policies are not well-defined.	2.8
A-12	Ability to focus research efforts to meet the customer's decision-making needs.	3.3
A-13	Ability to adapt to a dynamic environment.	3.2
A-14	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.	3.6
A-15	Ability to identify information gaps.	3.1
A-16	Ability to recognize and mitigate cognitive biases which may affect analysis.	3.1
A-17	Ability to recognize and mitigate deception in reporting and analysis.	2.9
A-18	Ability to think critically.	3.8
A-19	Ability to understand objectives and effects.	3.3
A-20	Ability to utilize multiple information sources across all information disciplines.	3.2
A-21	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	3.7
A-22	Ability to understand a business problem.	3.7
A-23	Ability to understand and use the databases and tools to run queries to solve the business problem.	3.7
A-24	Ability to identify patterns.	3.3