

Data Analytics and Predictive Modeling KSA

	Task	AVG
Domain I Business Problem (Question) Framing		
T-1	Assist in obtaining or receiving problem statement and usability requirements	3.62
T-2	Assist in identifying stakeholders	3.04
T-3	Assist in determining if the problem is amenable to an analytics solution	3.31
T-4	Assist in refining the problem statement and delineate	3.24
T-5	Assist in defining an initial set of business benefits	3.17
T-6	Assist in obtaining stakeholder agreement on the problem	3.07
Domain II Analytics Problem Framing		
T-7	Assist in reformulating the problem statement as an analytics problem	3.38
T-8	Assist in developing a proposed set of drivers and relationships to outputs	3.14
T-9	Assist in stating the set of assumptions related to the problem	3.28
T-10	Assist in defining key metrics of success	3.39
T-11	Assist in obtaining stakeholder agreement on analytical approach	2.86
Domain III Data		
T-12	Assist with identifying and prioritizing data needs and sources	3.28
T-13	Assist in acquiring data	3.34
T-14	Assist in harmonizing, rescaling, cleaning and sharing data	3.45
T-15	Assist with identifying relationships in the data	3.31
T-16	performance)	3.48
T-17	Assist with refining the business and analytics problem statements	3.17
Domain IV Methodology (Approach) Selection		
T-18	Assist with identifying available problem solving approaches (methods)	3.31
T-19	Assist in using software tools	2.83
T-20	Assist with testing approaches (methods)	3.43
T-21	Assist with selecting approaches (methods)	3.29
Domain V Model Building		
T-22	Assist with identifying model structures	3.07
T-23	Assist in running and evaluating the models	3.41
T-24	Assist with tuning models and data	3.21
T-25	Assist with integrating the models	2.86
T-26	limitations and constraints)	3.69
Domain VI Deployment		
T-27	Assist with performing business validation of the model	3.10
T-28	Assist with delivering report with findings	3.45
T-29	Assist with creating model, usability and system requirements for production	2.83
T-30	Assist in supporting deployment	2.83
Domain VII Model Lifecycle Management		
T-31	Assist with documenting initial structure	3.03
T-32	Assist in tracking model quality	3.28
T-33	Assist with recalibrating and maintaining the model	2.97
T-34	Assist with evaluating the business benefit of the model over time	2.90

T-35	Assist with assessing the validity of source data and subsequent findings.	2.97
T-36	Assist with collecting metrics and trending data.	3.24
T-37	Assist in conducting hypothesis testing using statistical processes.	3.14
T-38	Assist in conferring with systems analysts, engineers, programmers, and others to design application.	2.79
T-39	Assist with developing strategic insights from large data sets.	2.90
T-40	Assist with presenting technical information to technical and nontechnical audiences.	3.31
T-41	Assist with presenting data in creative formats.	2.83
T-42	Assist in developing recommendations to the supervisor based on data analysis and findings.	3.07
T-43	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).	3.17
T-44	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	3.21
T-45	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).	3.24
T-46	Assist with developing and implementing data mining and data programs.	2.69
T-47	Assist in the identification of information collection shortfalls.	2.83
T-48	Assist with providing analyses and support for effectiveness assessment.	2.79
T-49	Assist with providing input and assist in post-action effectiveness assessments.	2.66
T-50	Assist with deploying application codes and analytical models using CI/CD tools and techniques and provides support for deployed data applications and analytical models.	2.64
T-51	Assist with performing internal business Verification and Validation of the model	3.00
T-52	Assist with publishing Validation and Verification report	2.86
Knowledge		
K-2	Knowledge of risk management processes (e.g., methods for assessing and mitigating risk).	2.72
K-3	Knowledge of computer algorithms.	3.28
K-4	Knowledge of computer programming principles	3.24
K-5	Knowledge of data administration and data standardization policies.	2.86
K-6	Knowledge of data mining and data management principles.	3.00
K-7	Knowledge of database management systems, query languages, table relationships, and views.	3.10
K-11	Knowledge of mathematics (e.g. logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	3.34

K-15	Knowledge of programming language structures and logic.	3.21
K-16	Knowledge of query languages such as SQL (structured query language).	3.50
K-17	Knowledge of sources, characteristics, and uses of the organization's data assets.	2.86
K-18	Knowledge of the with various technologies for organizing and managing information (e.g., databases, bookmarking engines).	2.83
K-19	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	2.68
K-20	Knowledge of interpreted and compiled computer languages.	2.72
K-21	Knowledge of how to utilize Hadoop, Java, Python, SQL, Hive, and Pig to explore data.	3.17
K-22	Knowledge of machine learning theory and principles.	3.34
K-23	Knowledge of data classification standards and methodologies based on sensitivity and other risk factors.	2.71
K-25	Knowledge of Personally Identifiable Information (PII) data security standards.	2.86
K-26	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	3.07
K-27	Knowledge of data mining techniques.	3.38
K-28	Knowledge of database theory.	2.76
K-32	Knowledge of how to extract, analyze, and use metadata.	2.97
K-33	Knowledge on ETL techniques, Hadoop, Data analytics, Big data is an advantage	2.93
K-34	Knowledge of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	3.24
K-35	Knowledge of advanced statistical techniques and concepts (regression, properties of distributions, statistical tests and proper usage, etc.) and experience with applications.	3.21
K-36	Knowledge of the underlying theory and concepts of Relation Databases. (e.g. Microsoft SQL Server, Oracle, Teradata MySQL, etc.)	2.76
K-37	Knowledge of decision science game theory	2.89
K-38	Knowledge of the use of simulation	3.10
K-39	knowledge of optimization	3.30
K-40	Knowledge of data analysis concepts	3.60
K-41	Knowledge of how to identify and document potential ethical concerns for application of model outputs	3.14
Skills		
S-1	Skill in conducting queries and developing algorithms to analyze data structures.	3.48
S-2	Skill in creating and utilizing mathematical or statistical models.	3.31
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	3.28
S-4	Skill in using and contributing content to data dictionaries.	2.66
S-5	Skill in developing data models.	3.04
S-6	Skill in generating queries and reports.	3.45

S-7	Skill in writing code in a currently supported programming language (e.g., Python).	2.93
S-8	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	2.96
S-9	Skill in identifying patterns or relationships.	3.07
S-10	Skill in performing sentiment analysis.	3.33
S-11	Skill in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	3.14
S-12	Skill in supporting transformation analytics to invoke a business shift.	2.89
S-13	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	3.43
S-14	Skill in using data analysis tools (e.g., Excel, Python).	3.34
S-15	Skill in using data mapping tools.	2.93
S-16	Skill in using outlier identification and removal techniques.	3.25
S-17	Skill in writing scripts using R, Python, PIG, HIVE, SQL, etc.	3.48
S-18	Skill to identify sources, characteristics, and uses of the organization's data assets.	2.76
S-19	Skill in conducting information searches.	2.96
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.	2.79
S-21	Skill in evaluating information for reliability, validity, and relevance.	2.93
S-22	Skill in preparing and presenting briefings.	3.14
S-23	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	2.93
S-24	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools in conducting open-source searches.	2.83
S-25	Skill in utilizing feedback to improve processes, products, and services.	3.10
S-26	Skill in performing data analysis including applying statistics.	3.57
S-27	Skill in using statistical computer languages (R, Python, etc.) to manipulate data and draw insights from large data sets.	3.48
S-28	Skill in Visualization using R, Python or other languages and frameworks.	3.36
S-29	Skill in problem-solving skills and critical thinking ability.	3.62
S-30	Skill in collaboration and communication skills within and across teams.	3.59
S-31	Skill in analytics problem framing, ex. define geometric sets	3.50
Abilities		
A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	3.21
A-2	Ability to identify basic common coding flaws at a high level.	3.00
A-3	Ability to use data visualization tools (e.g., Flare, HighCharts, AmCharts, D3.js, Processing, Google Visualization API, Tableau, Raphael.js).	3.34
A-4	Ability to source data used in information, assessment and/or planning products.	2.66
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	3.50

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.	3.00
A-7	Ability to evaluate, analyze, and synthesize large quantities of data (which may be fragmented and contradictory) into quality, fused targeting/information products.	2.89
A-8	Ability to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	2.93
A-9	Ability to effectively collaborate via virtual teams.	3.32
A-10	Ability to evaluate information for reliability, validity, and relevance.	3.29
A-11	Ability to exercise strong ethical judgment when policies are not well-defined.	2.83
A-12	Ability to focus research efforts to meet the customer's decision-making needs.	3.31
A-13	Ability to adapt to a dynamic environment.	3.21
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.	3.57
A-15	Ability to identify information gaps.	3.14
A-16	Ability to recognize and mitigate cognitive biases which may affect analysis.	3.07
A-17	Ability to recognize and mitigate deception in reporting and analysis.	2.86
A-18	Ability to think critically.	3.76
A-19	Ability to understand objectives and effects.	3.32
A-20	Ability to utilize multiple information sources across all information disciplines.	3.17
A-21	Ability to effectively communicate ideas to team members with varying levels of technical expertise	3.66
A-22	Ability to understand a business problem	3.70
A-23	Ability to understand and use the databases and tools to run queries to solve the business problem	3.70
A-24	Ability to identify patterns	3.30