**ExxonMobil Functional Skills Catalog (Detail) – Computational Sciences**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Skill Definition** | **Level 1** | **Level 2** | **Level 3** | **Level 4** |
| **Data Analytics**  Our data is becoming increasingly large and complex, and our problems usually involve substantial uncertainty and variability. It is unreasonable to expect that standard workflows or generic software will always extract every bit of valuable information from our data, much less provide us with significant competitive advantage. It is therefore important that our technical professionals build and maintain the capability to perform specialized data analyses themselves. Data analytics encompasses the computational and statistical skills required to use data in support of scientific enquiry and sound business decision-making. This is typically a cyclic process involving most of the following:   1. Translating a scientific or business question into one that can be investigated using data 2. Generating new data in support of the study 3. Retrieving existing data from various sources and preparing it for analysis 4. Exploring and visualizing data for quality control and improved understanding 5. Building, analyzing, and comparing models for the data in a reproducible manner 6. Contextualizing the results and synthesizing with existing knowledge in order to gain an improved understanding and/or make a more informed decision | Critical Consumer   1. Understands the concepts of accuracy vs. precision; aware of the differences between anecdotal evidence, observational data, and designed experiments; recognizes the importance of representative sampling 2. Familiar with: relational databases, data historians, geodatabases, and other data stores; CSV, spreadsheets, and other file formats; basic principles of scientific computing and programming (loops, functions, etc.) 3. Able to interpret numerical summaries (e.g. means, medians, standard deviations, percentiles) and graphical displays (e.g. histograms, scatter plots, heat maps) of data 4. Understands basic concepts of probability and statistics, including: variability and data distributions; association vs. causation; dependent vs independent variables; interpolation vs. extrapolation; and statistical vs. practical significance | Individual Practitioner   1. Familiar with the statistical design of experiments, including replication, randomization, blocking, confounding, main effects, and interactions 2. Proficient in a data-oriented scripting language (e.g. Python, Matlab, or R) and associated libraries for data processing and analysis; able to write scripts to import data from Excel, text (familiarity with regular expressions), and relational databases (familiarity with SQL) 3. Proficient with visualization tools in a scripting language, commercial business intelligence software, and/or geographic information system; familiar with clustering, principal component analysis, and outlier detection methods 4. Proficient with applied probability (distributions, Bayes’ theorem), statistics (hypothesis testing, confidence intervals, linear regression), and Monte Carlo simulation; familiar with basic concepts of machine learning (supervised vs. unsupervised learning, training data vs. test data); demonstrated dedication to reproducible analysis | Local Expert   1. Proficient in statistical design and analysis of experiments, including adaptive experiments and response surface methodology 2. Familiar with time series analysis, geospatial analysis, digital signal processing, image processing, natural language processing, and other specialized areas often required for feature engineering 3. Proficient with methods for dimensionality reduction, clustering, and anomaly detection 4. Deeply and broadly knowledgeable in the theory, algorithms, numerics, and applications of statistical analysis, machine learning, and pattern recognition; able to adapt and extend known methods; understands research publications on statistics, machine learning, and pattern recognition while distinguishing between academic and practical concerns | Strategic Consultant   * In-depth understanding of the state-of-the-art applications of statistics, machine learning, and pattern recognition in the oil and gas industry * Widely experienced in integrating advanced data analytics and extensive industry knowledge on multiple business-related problems * Internally and externally recognized as an advanced expert in data analytics; directs company-wide strategies and influences external decisions |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Potential Meaningful Experiences**  *Examples of opportunities to exhibit the Functional Skill* | 1. Identified a business problem to investigate and a data set to use for the analysis 2. Performed a sequence of laboratory measurements or computational simulations following a given experimental design 3. Isolated, exported, and quality checked data from an internal database; added new features to the data set by performing computations on the raw data (using, e.g., Excel) or by performing basic spatial analyses (using, e.g., ArcGIS) 4. Built an informative visualization of business-relevant data using a commercial business intelligence program (e.g. Tableau or Spotfire) or geographic information system (e.g. ArcGIS) 5. Used commercial software or internally developed applications to analyze noisy data 6. Used the results of a data-driven analysis, including a careful consideration of the uncertainties involved, when making a business-relevant decision | 1. Identified problems amenable to statistical, predictive, or geospatial modeling 2. Developed a fractional factorial design for a laboratory or computational experiment 3. Written scripts to automate a process of importing data sets from multiple sources, merging them together, and cleaning the results for further analysis 4. Created an interactive dashboard or GUI for exploratory analysis; used off-the-shelf clustering algorithms for routine analysis 5. Successfully completed undergraduate-level courses in probability and statistics; used commercial statistical software (e.g. JMP, Minitab), a geospatial analysis platform (e.g. ArcGIS), or a scripting language to perform an analysis; used knitr, Jupyter, or similar tools to ensure reproducibility of an analysis 6. Worked with a domain expert to effectively communicate results and uncertainties back to the problem owner | 1. Worked iteratively with domain experts to help them to define a problem of value 2. Rigorously designed complex and/or adaptive experiments 3. Used a nontrivial, domain-specific algorithm to derive features for inclusion into a bigger model; for example, spectral analysis or wavelets for time series, image segmentation, kriging to build a spatial distribution, etc. 4. Uncovered useful structure by applying a nonlinear dimensionality reduction algorithm to a complex data set 5. Used combinations of advanced methods from statistical analysis, Monte Carlo simulation, pattern recognition, and machine learning to build models for multiple complex problems 6. Demonstrated publishable or patentable applications of data-analytic methods that have solved important business problems | * Active in the statistics, pattern recognition, and/or machine learning communities (e.g. invited talks, conference session chairs, peer reviewer for scientific journals) * Provided multidisciplinary leadership, mentorship, and technical support for data analytics across ExxonMobil’s organizations * Recognized innovator in the field – authored multiple external publications / internal reports / patent applications, developed new techniques and/or applications |