CareerWise

Data Analyst Rubric with Option for Data Scientist

COMPETENCY	NOVICE	EMERGING	PROFICIENT	ADVANCED
Data Analyst Required Competencies				
Data Analysis *Acquires, organizes, and interprets data from various data sources.* **Note**: For the purpose of this competency, query is referring to the process for producing data outputs. As an example, query is used in SQL, other programs may use their own code or syntax to produce data outputs.	Navigates system to access data on a basic level. Identifies best practices & concepts regarding how to organize data based on the data analysis request. Begins to generate questions & observations about query results. Demonstrates an ability to apply and comprehend basic queries or formulas in a program (such as Excel).	Develops simple **formulas**/**queries** to gather necessary data outputs, with support from a supervisor. Organizes data for analysis, based on best practices for the data analysis business request, with support from a supervisor. Analyzes **data output/query results** by **performing basic calculations** and begins development of the narrative, with support from supervisor.	Develops simple **formulas**/**queries** to gather necessary data outputs. Organizes data for analysis, based on best practice for the data analysis business request. Analyzes **data output/query results** to begin development of the narrative.	Develops complex **formulas/queries** with support from a supervisor. Anticipates how data will need to be organized and what information will need to be displayed, based on the data analysis business request. Generates insights related to the business request, based on **data output/query results.**

Programming for Data Analysis *Uses programming language such as SQL or Microsoft Excel program to produce data output.*	Understands the basic purpose and functions for accessing data. Uses a tool, such as SQL or Microsoft Excel, that helps generate the **data outputs.**	Writes basic Microsoft Excel or SQL (or other relevant language), commands, with support from supervisor.	Writes basic Microsoft Excel or SQL (or other relevant language) commands.	Independently writes basic SQL (or other relevant language) statements without having to rely on a tool that generates SQL. Knows how to use different languages and tools (such as Stata, SASS, SPSS) to create outputs for certain data sets.
Data Visualization *Translates data into basic visualizations to support business analytics and problem solving.* *Data Visualization tools may include Tableau, Power BI, Salesforce, Looker.*	Understands basic visualization functions (ex. drag & drop, tables) inside a graphic user interface (such as Tableau). Explores basic visualization options to support the narrative about the data. Reviews and provides feedback of visualization tools created by team members and demonstrates an ability to apply basic principles of layout & design.	Creates new visualization reports or customer self-service tools for stakeholders, with support from supervisor. Identifies best visualization to support the narrative about the data, with support from a supervisor. Creates basic visualizations that implement best practices of layout & design, with support from a supervisor.	Preemptively determines the visualizations needed by stakeholders and proactively creates those visualizations or self-service tools. Identifies best visualization to support the narrative about the data. Creates basic visualizations that **consistently** implement best practices of layout & design.	Creates and uses visualizations to generate new insights or hypotheses, beyond what the stakeholder requested, that drive towards current business goals or create new value propositions. Communicates how data was gathered, filtered, and displayed in the visualization. Acts as a resource for implementing best practices of layout and design.

Inquiry-Driven Solutions *Identifies key business challenges and performs preliminary data collection and root cause analysis to recommend and implement solutions.*	Generates research questions that initiate the problem solving process.	Develops strong inquiry skills regarding which questions to ask, who to ask, and what resources are available, with support from their supervisor.	Develops **simple** solutions to implement and resolve problems, using information gathered in inquiry.	Develops **complex** solutions to implement and resolve **complex** problems, using information gathered in inquiry.
Customer Service *Respond to customer requests for support (such as answering a business or technical question, or generating reports).*	Familiar with common questions or the types of support customers need. Generates a bank of solutions and processes for commonly asked questions or requests.	Responds to customers' low-level questions or requests for support, such as delivering a packaged analysis, with support from the supervisor.	Independently responds to customers' questions or requests for support and seeks solutions, such as creating a data visualization tool or completing new data analysis, as needed. Follows up with general respondents to clarify certain questions in the data, such as from survey results.	Provides responses and/or solutions quickly and independently to customers; well-versed in customer needs. Follows up with general respondents to clarify certain questions in the data, such as from survey results, and is able to offer an explanation as to why some survey results might be different.
Troubleshooting *In a technical setting, performs problem resolution and root cause analysis and recommends and implements solutions.*	Identifies challenges with data and proactively reaches out for support for the specific challenge. Recognizes patterns of the challenges encountered in the data analysis process.	Identifies people and resources that can help with problem resolution and can clearly communicate the challenge. Discovers basic solutions to common challenges in	Recognizes patterns of challenges in the data analysis process and applies learned solutions to overcome the obstacle.	Acts as a resource who is called on to support others when resolving problems.

		the data analysis process.		
Communicates Data Insights Tells and adapts a story from the results of a model through a clear, concise explanation in a verbal or written method, that is uniquely framed to the intended audience.	Elaborates on basic insights to the team in easy to read explanations.	Produces a one pager that summarizes insights and results and justifies recommendations, with support from supervisor. Engages with teammates about the results and can answer questions about the results.	Produces a short memo that elaborates on the insights and the results they got from a certain output and that explains their rationale for their insights. Utilizes simple data visualization tools, such as tables or graphs, within memo.	Produces a research brief with clear and detailed results, insights, and recommendations, and which explains how they got the results and how the results were affected by the certain variables. Utilizes complex data visualization tools within memo.
Business Acumen *Uses understanding of the organization's priorities, financial model, and operations to guide daily work.*	Describes the general purpose and goals of the business. Discusses the crucial metrics being measured. Displays detailed understanding of the particular customer. Understands how data supports the organization in making decisions and achieving its goals.	Generates data reports that support the organization in achieving their business goals, as requested by the supervisor.	Anticipates what data or reports would support the organization with decision-making around priorities and goals; efficiently provides the information needed.	Intuits a customer's needs and develops an end result that is novel to what the customer requested that may more effectively support educating the customer for decision making.

Optional Competencies for Data Science Specialization				
Problem Formulation *Identifies opportunities to improve **processes** through use of predictive analytics. Effectively frames and states the problem and how data science may add value.*	With support from a supervisor, understands a customer need, problem, or opportunity and how data science may add value. Observes team members interviewing stakeholders and asking clarifying questions to **create requirements for a data collection/analysis tool.**	With support from a supervisor, clearly articulates the customer need, problem, or opportunity and how data science may add value. Participates with other team members to interview stakeholders and ask clarifying questions.	Independently articulates the customer need, problem, or opportunity and how data science may add value. Co-leads conversations with stakeholders and **creates requirements for a data collection/analysis tool.**	Independently identifies new opportunities to apply data science in operational processes. Effectively articulates the customer need, problem, or opportunity to enable others to understand. Leads conversations with stakeholders and **creates requirements for a data collection/analysis tool.**
Data Acquisition *Demonstrates ability to identify relevant data for a business use case and acquire it for analysis and modeling activities.* **Note**: The language used in this competency is more specific to SQL, but an apprentice could develop proficiency in comparable functions in other statistical tools.	Knows the process to **assemble and run queries** to pull data for a specific problem or analysis. Articulates when and how to use **joins, unions**, **and basic filtering functions**.	With support and direction from supervisor, **assembles and runs queries** that combine multiple sources to produce a comprehensive data set. Correctly applies **joins, unions, and filtering functions.** With support from a supervisor, validates data with subject matter experts to ensure the	**Assembles and runs queries** that combine multiple sources to produce a comprehensive data set. Correctly applies **complex code objects, including triggers, procedures, and functions.** Validates data with subject matter experts to ensure the data meets the needs of the use case.	Independently identifies data needed for a specific use case. Runs **complex queries** that combine multiple sources to produce a comprehensive data set. Validates data with subject matter experts to ensure that the data meets the needs of the use case.

		data meets the needs of the use case.		
Exploratory Data Analysis *Uses a notebook to summarize and describe a data set. Looks for patterns, missing values, and anomalies. Identifies appropriate remediation to prepare data set for model building.*	Understands the function of exploratory data analysis and how it fits into the data science process. Understands the concepts related to familiarizing with a data set, cleaning a data set, and analyzing relationships between variables. For example, redundant variables or outliers.	With support from a supervisor, accesses data set from a notebook and completes basic summarization and descriptive tasks. With support from a supervisor, identifies and addresses issues in a data set (e.g. null value treatment or data type conversions). With support from a supervisor, uses techniques like a correlation matrix or scatter plot to visualize/identify relationships/patterns in the data, and begins to craft any potential targets.	Independently accesses data set from a notebook and completes summarization and descriptive tasks. Identifies and addresses a variety of possible issues in a data set. For example, redundant variables or outliers. Uses techniques like a correlation matrix or scatter plot to visualize and identify relationships and patterns in the data. Partners with supervisor or other team members to review findings, incorporate feedback, and identify appropriate next steps.	Independently and thoroughly explores and cleans a data set using relevant techniques for the specific use case. Correctly interprets findings in the context of the use case and communicates them to relevant team members. Identifies appropriate next steps based on analysis, including modeling approach and experiment design.
Statistics and Probability *Uses basic statistics and probability to perform job functions.*	Understands basic concepts of statistics	With support from a supervisor, applies statistics and probability concepts to perform job functions.	Independently applies statistics and probability concepts to perform job functions.	Is considered highly proficient in applying statistics and probability concepts to job functions.

Model Building *Selects and builds a variety of predictive models to perform job functions and support business goals.*	Demonstrates awareness of a variety of available predictive methods and their uses, such as linear and logistic regression, decision trees, and naive bayes. Understands the difference between supervised and unsupervised machine learning and when each might be used.	With support from a supervisor, selects a predictive modeling approach that fits the use case. With support from a supervisor, creates an appropriate data split, trains, and tests a predictive model. With support from a supervisor, evaluates model results.	Independently selects a predictive modeling approach that fits the use case. Crafts a target, creates an appropriate data split, trains, and tests a predictive model. Evaluates model results and identifies performance enhancement opportunities. Effectively communicates outcomes to appropriate stakeholders.	Is considered highly proficient in selecting or building predictive models, based on the needs of the business or project.
Programming language *Uses a programming language such as Python or R to facilitate data science processes.*	Demonstrates knowledge and ability to use a programming language for data science processes.	With support from a supervisor, applies programming language to data science processes.	Independently applies programming language to data science processes.	Applies advanced programming concepts to data science processes.

Prerequisites: Must have completed Statistics or Calculus *plus* Introduction to Computer Science by the end of year 2.

Certifications that Support:

Summer between years 1 and 2: Python Training

Summer between years 2 and 3: IBM Data Science Certification

Note: Version 1 of this rubric has been developed by the team at Pinnacol Assurance.