

THE REGISTERED APPRENTICESHIP OCCUPATIONS AND STANDARDS CENTER OF EXCELLENCE (AOSC)

Data Analyst National Occupational Framework

ONET Code: 15-2041.00

RAPIDS Code: 2099

Created: April 2025

This project has been funded, either wholly or in part, with federal funds from the Department of Labor, Employment and Training Administration under Cooperative Grant Number AP-36653-21-75-A-11. The contents of this publication do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of the same by the US Government.





ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization that provides data and evidence to help advance upward mobility and equity. We are a trusted source for changemakers who seek to strengthen decisionmaking, create inclusive economic growth, and improve the well-being of families and communities. For more than 50 years, Urban has delivered facts that inspire solutions—and this remains our charge today.

Acknowledgments

We would like to thank several people who have contributed to the development and vetting of this National Occupational Framework. We especially thank Diane Jones, Leslee Haisma, and Fernando Hernandez for their research, support, and contributions to developing the National Occupational Framework. Additionally, we had terrific experts lend their time, review, and support to the framework. They include Donna LaLonde, associate executive director at the American Statistical Association; Kirsten Rohde, vice president of technology solutions at the American Institutes of Research (AIR); Nathan Watson, president of Contemporary Analysis; Carlos Vazquez, president and CEO of Miami EdTech and GoSprout; Jess McClain, principal analytical lead at Google; Dr. Donald S. Walker, training solutions director at Apprenticely (Arkansas Center for Data Sciences); Kanisha Vaughn, director of analytics at All Voting is Local; and Alena Stern, director of data science and chief data scientist at Urban Institute. Finally, we thank Bhavani Arabandi and Lexi Mills, our editor, for their thoughtful review.

Introduction to Using This Document

Under the Registered Apprenticeship Technical Assistance Centers of Excellence award, the Urban Institute leads the Occupations and Standards work. One of the main objectives of Urban's project is to create high-quality, well-researched, consensus-based work process schedules that are nonproprietary and widely available. This document is a product of that work and contains three sections: the occupational overview, the work process schedule, and the related technical instruction.

The **occupational overview** is a general introduction, including alternative job titles, any prerequisites, and, if applicable, the total number of hours needed to complete a time-based or hybrid program.

The **work process schedule** outlines the major job functions, competencies, and/or hours an apprentice completes in a registered apprenticeship program. It outlines what apprentices are expected to learn on the job with the support of a mentor or journeyworker (a worker mastering the competencies of an occupation in a particular industry), including both core competencies and those deemed optional by experts in the field. The work process schedule is the foundational document guiding a program.

Urban works with numerous experts to ensure the content is thoroughly researched and vetted to reflect the expectations of industry, educators, labor unions, employers, and others involved in apprenticeship for this occupation. Sponsors and employers can use the work process schedule as their program standards with assurances it has been approved by experts in the field.

The **related technical instruction** presents considerations for the coursework that apprentices will undertake to supplement on-the-job learning. It is intended to serve as a reference to sponsors exploring their options for the accompanying classroom, virtual, or hybrid training.

How to Use the Work Process Schedule

Sponsors can adapt the work process schedule to accommodate their needs for competency- or time-based or hybrid programs. In a **competency-based** apprenticeship, sponsors assess apprentices' progress across core and optional competencies listed in the work process schedule. In a **time-based** apprenticeship, apprentices complete a predetermined number of hours across major job functions and the program overall. In a **hybrid** apprenticeship, sponsors monitor apprentices' hours spent on major job functions and assess their proficiency across competencies.

Each program type has a different method of assessment:

- **For a competency-based program**, apprentices engage in activities and make progress toward proficiency in the identified competencies. Sponsors overseeing apprentices' work assess their mastery of the outlined competencies using the following rating scale:

4—Competent/proficient (able to perform all elements of the task successfully and independently)

3—Satisfactory performance (able to perform elements of the task with minimal assistance)

2—Completed the task with significant assistance

1—Unsuccessfully attempted the task

0—No exposure (note the reason—absence, skill isn't covered, etc.)

The competencies may be completed in any order. Apprentices must perform at a level 4 or 3 in all competencies listed as “core” to complete the apprenticeship program successfully.

- **For a time-based program**, sponsors monitor apprentices' completion of hours in training across major job functions. The total number of hours recommended for this occupation is listed in the occupational overview and is based on guidance from the US Department of Labor. Generally, apprentices must have at least 2,000 hours overall for on-the-job learning, but occupations of greater complexity may require more hours. Sponsors will provide apprentices with supervised work experience and allocate the total number of hours across the major job functions to adequately train their apprentices.
- **The hybrid approach** blends both competency- and time-based strategies. Sponsors measure apprentices' skills acquisition through a combination of completing the minimum number of hours of on-the-job learning successfully demonstrating identified competencies. Sponsors will assess apprentices' proficiencies as described for competency-based programs with a rating scale of 0–4 for every core competency. Generally, apprentices have at least 2,000 hours overall for on-the-job learning, but occupations of greater complexity may require more hours. Sponsors will document apprentices' completion within a minimum and maximum range of hours assigned for each major job function.

Data Analyst Occupational Overview

Occupational Purpose and Context

Data analysts apply quantitative methods to data (numerical, structured, and unstructured) to transform it into meaningful insights. Data analysts use data-oriented programming languages and visualization software to bring out these insights from the data. Part of the process of creating insights using data includes cleaning existing data, collecting survey data, collapsing data into convenient levels, or other activities related to data collection. Data analysts may also provide support to researchers on the appropriate modeling methods needed for inferential analysis. They also provide descriptive statistics using basic mathematical concepts in order to support research.

Potential Job Titles

Business intelligence analyst, competitive intelligence analyst, data analyst, market intelligence analyst, market intelligence consultant, strategic business and technology intelligence consultant, strategist, clinical data management director (CDM Director), clinical data management manager (CDM manager), clinical data manager, clinical informatics manager, data deliverables manager, data management manager, clinical research assistant, graduate assistant, graduate research assistant, research aide, research assistant, research associate, research technician, social research assistant, financial analysts, freelance data analysts, hospitality analysts, business analysts, research analysts, marketing analysts, database developers, social media analysts

Apprenticeship Prerequisites

A high school diploma or equivalent is generally required. Candidates should possess a solid foundation in basic mathematics, including algebra and probability. Candidates should also have foundational computer literacy knowledge and skills.

Recommended Length of Apprenticeship (Time/Hybrid Programs Only)

The recommended length of on-the-job learning in a data analyst apprenticeship program is 2,000 to 4,000 hours

Work Process Schedule

Data Analyst

ONET Code: 15-2041.00

RAPIDS Code: 2099

Instructions for Use:

Competency-based programs: In the “performance level achieved” column of the work process schedule (see examples starting on the next page), assess apprentices’ performances on each competency with the scale below. No monitoring of hours is required for this approach. See “Guidelines for Competency-Based, Hybrid and Time-Based Apprenticeship Training Approaches,” US Department of Labor, Employment and Training Administration, Office of Apprenticeship, October 20, 2015,

<https://www.apprenticeship.gov/sites/default/files/bulletins/Cir2016-01.pdf>.

- 4—Competent/proficient (able to perform all elements of the task successfully and independently)
- 3—Satisfactory performance (able to perform elements of the task with minimal assistance)
- 2—Completed the task with significant assistance
- 1—Unsuccessfully attempted the task
- 0—No exposure (note the reason—absence, skill isn’t covered, etc.)

Time-based programs: In the “hours” row, specify the number of hours apprentices will fulfill for each job function. No assessment of competencies is required for this approach.

Hybrid programs: In the “performance level achieved” column, assess apprentices’ performances on each competency using the 0–4 scale above. In the “hours” row, identify a range of hours apprentices should spend working on each major job function.

Job Function 1: Effectively communicates and collaborates in a team environment		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Interacts effectively with company personnel and other stakeholders, deferring to experts as needed	Core	
B. Demonstrates professional communication etiquette when interacting with colleagues or other stakeholders both verbally and in writing	Core	
C. Demonstrates the ability to convey in a cogent manner their ideas or findings through written and spoken communication to both technical and nontechnical stakeholders	Core	
D. Communicates issues impacting their deliverables clearly and in a timely fashion	Core	
E. Demonstrates the ability to self-manage their work time, including prioritization and calendar management, such that they meet identified deadlines	Core	
F. Collaborates with other analysts and researchers as part of projects or teams, communicating clearly, effectively, while maintaining respect and inclusivity	Core	
G. Complies with and demonstrates understanding of company policies about inclusivity and equity	Core	
H. Demonstrates the ability to give constructive upward feedback to supervisors and peers in a respectful and actionable manner	Core	
I. Leverages appropriate project management and version control software and tools, including AI tools, technology, and expertise to strengthen team coordination and accelerate project progress	Optional	

Job Function 2: Assists team in developing plans of analysis, proposals, and other documents delineating data and research activities

Hours (time-based and hybrid programs only):

Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Under supervision of a lead researcher, identifies research questions and methodologies for analysis products	Core	
B. Follows a formulated work plan, including timeline, data life cycle, and work process schedule (as needed) for analysis plans or proposals	Core	
C. Under the supervision of a researcher, collaborates in producing non-research documents needed for proposals	Core	
D. Assists team in determining what data to collect to measure the identified conceptual problems	Core	
E. Evaluates the appropriateness of prospective data for the identified conceptual problems, including assessing data quality	Core	
F. Identifies relationships and trends in the data or any factors that could affect the results of research	Core	
G. Under the supervision of a researcher, applies sampling techniques to determine groups to be surveyed or uses complete enumeration methods	Core	
H. Participates in the soliciting of feedback using generative interviewing throughout the development period from internal and external stakeholders	Core	
I. Uses AI technology to assist in the development of the work plan and work process optimization	Optional	

Job Function 3: Collecting and cleaning secondary data		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Collaborates in collecting data and documentation, and storing it in an effective manner for a project's needs	Core	
B. Demonstrates the ethical handling of secondary data outlined in organizational protocols and data sharing agreements as relevant	Core	
C. Writes well-documented, understandable, and efficient code which produces reproducible data-cleaning procedures using Structured Query Language (SQL), or typical statistical software (such as Python, R, and Stata)	Core	
D. Documents work along the way, using organizational protocols, and following guidelines from supervising researchers	Core	
E. Demonstrates the ability to transform, merge, and append existing datasets as needed	Core	
F. Creates new columns of data (variables) in a structured data table from existing datasets	Core	
G. Implements methods for working with missing data and documenting sources of missing data	Core	
H. Collaborates in creating an efficient workflow	Core	
I. Demonstrates knowledge of major relevant datasets for the area of focus, including the ability to explain basic information on a relevant dataset requested by a lead researcher	Core	
J. Identifies and extracts pertinent information from a dataset when given a general question or goal, ensuring that the selected data aligns with the research objectives and analytical needs	Core	
K. Extracts structured and unstructured data from websites, servers (using SQL), and form documents, and formats them into usable structured data	Optional	

Job Function 4: Assists in primary data (survey) collection efforts		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. With guidance from supervisor, works with survey collection software (e.g., Qualtrics, SurveyMonkey), effectively building surveys and troubleshooting as needed	Core	
B. Collaborates in collecting survey data from respondents, communicating survey questions, and capturing responses effectively, either by phone or in person, as needed	Core	
C. Under the supervision of a researcher, collaborates in executing sampling procedures, determining sample respondents, and maintaining outreach logs	Core	
D. Abides by internal and external data security procedures to secure the safety of sensitive information collected during the survey process	Core	
E. Ensures proper protection of data by following established security protocols	Core	
F. Using company best practices, conducts and maintains organizational quality control procedures of data	Core	
G. Solicits feedback using generative interviewing throughout the development period from internal and external stakeholders	Core	
H. Combines multiple external data sources and primary survey data together using statistical software or SQL	Core	
I. Under the supervision of researcher, evaluates potential bias in survey structure to allow for the identification of potentially biased data	Optional	
J. Under the supervision of a researcher, supports collecting information through interviews, such as from individuals, focus groups, or through computer assisted telephone interviews	Optional	
K. Monitor the collection of survey data, ensuring the proper implementation of instruments and representativeness of the survey respondents, and addressing any questions respondents or researchers may have regarding the survey and its implementation	Optional	

Job Function 5: Analyzes data		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Conducts descriptive analyses of results, producing basic statistical metrics such as mean, median, etc., in a format desired by the client or stakeholder	Core	
B. Analyzes, manipulates, or processes large sets of data using statistical software in an efficient manner to produce insights in the form of data points to clients or stakeholders	Core	
C. Produces, summarizes, and visualizes key metrics, according to the needs of clients and stakeholders, using statistical and data visualization and dashboard tools	Core	
D. Writes new functions or applications in programming language to collect and transform data into analyses, using both traditional methods and generative AI	Core	
E. Under guidance and supervision of a researcher, produces code and runs regression and other computational methods on statistical software	Optional	
F. Under guidance and supervision from a researcher, produces code and runs hypothesis tests to address research questions or for robustness checks for any estimates they produce	Optional	
G. Under the supervision of a researcher, assists in analyzing interview data for themes/codes	Optional	

Job Function 6: Assesses and interprets results		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Uses data and sense checks, compares against previous research, or consults with experts to assess validity	Core	
B. Assesses reliability of results by examining the context in which the data originated, reviewing the source data's documentation, and creating data codebooks for others to understand the data tables they produce	Core	

C. Analyzes data from dashboards, metrics, or other estimates based on their own analysis to identify potential solutions to key organizational questions	Core	
D. Analyzes data from graphs and tables (descriptive statistics, regression results), based on their own analysis, to identify trends and relationship among variables	Core	
E. Consults the literature on the topic as needed to interpret and contextualize their results	Optional	
F. Under guidance and supervision of a researcher, reports possible problems with any applied computation methods	Optional	
G. In collaboration with researcher, addresses special concerns with time series analyses	Optional	

Job Function 7: Conveys findings to key stakeholders		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Using data visualization tools, creates graphic data visualizations of analytical findings to convey results	Core	
B. Writes reports on analytical findings	Core	
C. Prepares and delivers presentations of analytical findings to management or other end users using data storytelling practices to ensure understanding by both technical and nontechnical parties	Core	
D. Conveys issues and concerns to supervisor and team leads	Core	
E. Conveys findings to stakeholders to inform data-driven decisionmaking process	Core	
F. Recommends data-driven solutions to stakeholders	Core	
G. Develops interactive data tools on the web	Optional	
H. Identifies appropriate AI technology for use and generates effective prompts	Optional	

Job Function 8: Participates in ongoing professional development		
Hours (time-based and hybrid programs only):		
Competencies	Core or optional	Performance level achieved (0–4) (competency-based and hybrid programs only)
A. Develops and enhances analytical and computing skills	Core	
B. Develops and enhances personal skills	Core	
C. Develops and enhances workplace skills	Core	
D. Reads scientific articles, conference papers, or other sources of research to identify emerging analytic and industry trends and technologies	Core	
E. Stays up to date on new tools and techniques, including generative AI	Core	
F. Identifies and attends trainings to enhance work performance	Core	
G. Seeks and incorporates feedback on technical and communication skills to foster personal and professional growth	Core	

Related Technical Instruction

Data Analyst

ONET Code: 15-2041.00

RAPIDS Code: 2099

Instructions for Use:

Registered apprenticeships must include at least 144 hours of related technical instruction (RTI). Courses offered by accredited colleges and universities may be assigned a credit hour determination rather than a contact hour determination. In general, an academic credit unit is the equivalent of 15 clock hours of instruction.

Development and Use of This RTI Outline: Employers and academic institutions may approach RTI in markedly different ways. Our goal was not to identify the single best way to provide RTI or to identify a single provider whose content we deemed to be superior. Instead, our goal was to survey numerous education providers, including employers, institutions of higher education, high schools, private continuing education providers, labor organizations, professional associations and, in some cases, municipalities that provide worker training, to identify topics or courses common among those providers that align with the job functions included in this work process schedule. Those common topics or courses are reflected in the RTI outline provided below, which may be useful in developing your RTI program or communicating your needs to an educational partner.

Licensure or certification requirements: Professional licensure or certification is not required to work in this field. That said, employers may give hiring preference to individuals who have earned industry certifications, such as Google Analytics Professional Certificate, Python Data Analysts Certification or IBM Data Analytics certification.

Degree requirements for licensure or certification, if applicable: None

Accreditation requirements of instructional provider for licensure or certification, if applicable: Not applicable

Anticipated changes in licensure or certification requirements, if known: None known

Examples of state licensure or certification requirements: Not applicable

Examples of RTI providers for this occupation

Professional associations and labor organizations: The Institute for Operations Research and the Management Sciences (INFORM) provides courses (such as their Certified Analytics Professional prep course or Leading in Analytics course) designed to cover skills that are necessary for operations, research, and other analytics professionals (<https://www.informs.org/Professional-Development/Continuing-Education>).

The Data Science Council of America offers two levels of certification on a Data Scientist track and two on a Data Analytics track. The lower level of each is geared toward those with little professional experience beyond a degree in a related discipline, and the higher level of each is ideal for more experienced data professionals (<https://www.dasca.org/data-science-certifications>).

Military: The military provides instruction designed to train and certify candidates for appointment and initial assignment as a Data Operations Warrant Officer. In three phases of courses, the instruction emphasizes designing, building, operating, and maintaining information systems within the military cyberspace domains (<https://cybercoe.army.mil/Cyber-Center-of-Excellence/Schools/Signal-School/Signal-Courses/Warrant-Officer-Courses/255A-Data-Operations-Warrant-Officer-WOBC>).

Federal: The USUS CIO Council periodically hosts educational events and trainings focused on improvements to IT practices (<https://www.cio.gov/programs-and-events>).

States/municipalities: Not applicable

Colleges and universities: There are many accredited colleges and universities that offer coursework and educational programming leading to relevant degrees and certifications for IT Generalists.

No-cost online providers: Amazon Web Services (AWS) provides a wide range of online professional development and training, including a data analytics training that involves free virtual training webinars, and individual courses. The focus of the training is data collection and visualization, storage, and processing (<https://aws.amazon.com/training/learn-about/data-analytics>).

Continuing education or specialty education providers: In the tech industry, there are many providers that offer education and training programs, as well as certification programs, for users competent in particular software, systems or services.

Prerequisite knowledge, skills or experience typically required by RTI providers for this occupation

RTI providers are likely to require individuals to demonstrate competency in English and mathematics, which can include achieving passing scores on institution-based placement tests, nationally administered placement exams or college admissions exams (such as the SAT or ACT tests). Colleges and universities serving as RTI providers may require students to complete a specific mathematics course sequence, such as Algebra I, Algebra II, Geometry, and Trigonometry prior to enrolling in calculus classes. For this reason, apprenticeship sponsors may give priority to candidates who are transitioning from another field which required advanced mathematics knowledge and skills.

Employer Onboarding

Hours: 1–3

Sample learning objectives

- Explain the mission of the employer’s organization.
- Explain the organizational structure of the employer’s organization.
- Explain the chain-of-command in the employer’s organization, as well as the steps an employee should take to report concerns about practices or behaviors in the workplace or to share ideas about how to improve processes or efficiencies.
- Explain the employer’s policies for workplace conduct and ethics.
- Explain the employer’s policies regarding medical leave and vacation.
- Explain the requirements of the apprenticeship program and the role of the apprentice in the organization’s business activities.

Computer Basics

Hours: 20–30

Sample learning objectives

- Demonstrate the ability to use desktop computers, laptop computers, and smart devices to send an email, schedule a meeting, share a document, participate in group editing, use track changes to edit a document, maintain version control of documents, convert Word documents to PDF format, create e-fillable and e-signable forms, and create/verify electronic signatures.
- Describe the purpose of network computing and demonstrate the ability to connect devices such as computers, printers, telephones, and tablets to a secure network.
- Explain the significance of document security and the use of passwords and other document storage strategies to maintain document privacy and security.
- Explain the importance of Internet security and identify examples of cybersecurity threats, such as malware, Trojan horses, phishing, botnet, spyware, and social engineering.
- Describe effective strategies to prevent cyberattacks, to create and maintain secure passwords, and to take the appropriate action when potential cyber threats are identified.
- Demonstrate the ability to use Excel spreadsheets to tabulate, manipulate, and create visualizations of data.

Calculus for Data Analysis

Hours: 30–40

Sample learning objectives

- Define the word calculus and discuss the importance of calculus in studying rates of change as well as the various properties of objects.
- Explain the differences between and typical uses of differential and integral calculus.
- Explain the likely sources of error in designing and using data models, and describe techniques that are used to increase the accuracy of models.
- Describe the techniques used to modify the weights and biases of a data model, and demonstrate the impact of small changes on data outcomes.
- Demonstrate the ability to determine the slope of a line as well as the cost or error level of a line.
- Differentiate between points and vectors.
- Differentiate between single variate and multivariate functions.
- Differentiate between parametric surfaces and vector fields.
- Define vectors and demonstrate the ability to write vectors in different forms.
- Explain the concept of scalar multiplication of vectors.
- Explain the use of dot products, cross products, and matrices in developing equations and models.
- Explain how determinants are used as scaling factors.
- Explain the way in which gradient descent is used in machine learning.

Probability and Statistics

Hours: 30–40

Sample learning objectives

- Define probability and its principal concepts.
- Explain the concept of a random event and express probability definitions.
- Explain the concept of a random variable.
- Calculate expected values and moments.
- Explain binomial distribution.
- Explain hypergeometrical distribution.
- Explain Poisson distribution.
- Explain uniform, exponential, and normal distribution.
- Calculate the expected value of a random variable.

- Calculate the location of data (quartiles and percentiles).
- Calculate the measures of the center of data (mean, median, and mode).
- Calculate measures of the spread of data (variance, standard deviation, and range).
- Calculate probabilities using the addition and multiplication rules.
- Construct and interpret contingency tables.
- Construct and interpret Venn diagrams.
- Construct and interpret tree diagrams.
- Explain the significance of confidence intervals.
- Calculate the sample size required to estimate a population mean and a population proportion given a desired confidence level and margin of error.
- Explain the relevance and method of hypothesis testing with one or two samples.
- Explain the significance of Chi Square distribution.
- Explain linear regression and correlation, create and analyze scatter plots, and interpret a line of best fit and outliers.

Foundations in Data Collection

Hours: 25–35

Sample learning objectives

- Explain the ways in which data are collected by government agencies, businesses, academic researchers, community-based organizations, public health organizations.
- Describe the methods used by data scientists to determine how many and which samples are required to generate data that meets expected levels of accuracy and reliability.
- Discuss the role of spreadsheets, query languages, and data visualization tools in data analytics.
- Describe the role of the data analyst in various occupations or industries.
- Differentiate between biased and unbiased data.
- Discuss the use of data in decisionmaking, and how the significance of the decision could determine the level of accuracy and reliability required of the data analysis exercise.
- Describe ways in which data scientists provide context about the accuracy of their data collection methodology when responsibly interpreting and reporting results.
- Identify instances where data reports exaggerate the level of accuracy, reproducibility, or reliability of the data and discuss ways in which the authors should have been more honest about the limits of their interpretations.
- Demonstrate the ability to develop a questionnaire that will collect relevant data needed to answer a specific question.

- Demonstrate the ability to develop a methodology for distributing a survey such that survey results will be representative of the intended sample population, will be accurate and reliable, and will be meaningful.

Preparation and Cleaning of Data

Hours: 25–35

Sample learning objectives

- Demonstrate the ability to record and organize data for its intended use.
- Define different types of data integrity and identify risks to data integrity.
- Explain the problem of confirmation bias and the potential for statisticians to continue manipulating inputs and formulas until they obtain the result they desire.
- Differentiate between legitimate practices to clean datasets and the less desirable practice of eliminating data points that interfere with confirmation bias.
- Apply SQL functions to clean string variables in a database.
- Develop basic SQL queries for use on databases.
- Describe the process of verifying data cleaning results.

Analyzing and Visualizing Data

Hours: 45–55

Sample learning objectives

- Discuss the role of sorts and filters in analyzing data, and explain the importance of maintaining data integrity, reliability, and accuracy when applying sorts and filters.
- Convert and format data.
- Combine data from multiple databases or tables.
- Demonstrate the ability to use basic functions to perform calculations with data in spreadsheets.
- Discuss the role of data visualizations to share the results of data analysis.
- Demonstrate the ability to use Tableau as a data visualization tool.
- Demonstrate the ability to use SPSS statistical software to analyze data.
- Demonstrate the ability to use STATA statistical software to analyze data.
- Demonstrate the ability to use Statistical Analysis Software (SAS) to create and query databases and run statistical analyses.
- Demonstrate the ability to develop effective data presentations that incorporate clear and concise data visualizations.

- Explain the significance of data-driven stories and demonstrate the ability to create compelling presentations using data.

Data Analysis with R Programming

Hours: 30–40

Sample learning objectives

- Explain the purpose of R language and compare and contrast it with Python programming language.
- Demonstrate the ability to use R and RStudio to analyze statistical data.
- Describe the functions and variables available in R.
- Demonstrate the ability to structure, organize, and clean data using R.
- Discuss ways that R can be used to reduce or eliminate data bias.
- Use R to create data visualizations.
- Demonstrate the ability to annotate and save visualizations developed in R.

Python for Data Science

Hours: 30–40

Sample learning objectives

- Demonstrate the ability to use the Python interpreter.
- Explain the principles of binary computation and discuss input/output models.
- Explain the different data types used in Python, such as variables, expressions, and statements.
- Explain the purpose of control structures.
- Explain the use of modularization and classes to package data and subject to functions.
- Discuss the different types of data structures used in Python.
- Explain the differences between object-oriented programming and object-oriented design.
- Demonstrate the ability to construct expressions in Python.
- Demonstrate the ability to write custom python functions, classes, and workflows.
- Demonstrate the ability to create data science applications.
- Demonstrate the ability to create data arrays and tables in Python.
- Demonstrate the ability to organize data with Python.
- Demonstrate the ability to analyze and visualize data with Python.
- Demonstrate the ability to build predictive machine learning models.

Relevant military experience

255A—Data Operations Warrant Officer

16KXS—Data Scientist

Occupational insights

Most data analysts are white, making up 64.9 percent of the population; 14.5 percent of data analysts are Asian; 8.5 percent are Hispanic/Latinx; 7.5 percent are Black; and 4.6 percent are other. Just over half of data analysts are men, at about 50.2 percent. In 2022, women earned \$0.95 for every dollar that their male counterparts made (<https://www.zippia.com/data-analyst-jobs/demographics>).

Works Consulted

Amazon Web Services, “Data analytics,” accessed November 15, 2024, <https://aws.amazon.com/training/learn-about/data-analytics>.

Apprenticeship USA, “Data Scientist.” accessed November 15, 2024, <https://www.apprenticeship.gov/apprenticeship-occupations/listings?occupationCOde-15-2051.00>.

Data Camp, “What is Data Analysis? An Expert Guide With Examples,” November 10, 2024, <https://www.datacamp.com/blog/what-is-data-analysis-expert-guide>.

Data Science Council of America, “Big Data. Greater Promise, Superior Certifications.” accessed November 15, 2024, <https://dasca.org/data-science-certifications>.

Hillier, Will. “A Step-by-Step Guide to the Data Analysis Process,” Career Foundry, May 31, 2023, <https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step>.

Informa, “INFORMS Continuing Education,” accessed November 15, 2024, <https://www.informs.org/Professional-Development/Continuing-Education>.

O*Net Online, “Data Scientist,” accessed November 15, 2024, <https://www.onetonline.org/link/summary/15-2051.00>.

US Army Cyber Center of Excellence, “Data Operations Warrant Officer- Warrant Officer Basic Course (WOBC) 255A,” accessed November 15, 2024, <https://cybercoe.army.mil/Cyber-Center-of-Excellence/Schools/Signal-School/Signal-Courses/Warrant-Officer-Courses/255A-Data-Operations-Warrant-Officer-WOBC>.

Zippia, “Data Analyst Demographics and Statistics in the US,” April 5, 2024, <https://www.zippia.com/data-analyst-jobs/demographics>.

STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org