



DATA SCIENCE FUNDAMENTALS

COURSE OUTLINE

Want to be a Data Scientist? This is a starting point to give yourself the knowledge and skills to elevate and drive forward the work of the NLM! By learning data science, you can make powerful predictions through analytics, machine learning, visualization, and more, taking your skillset and your work to new heights.

COURSE STRUCTURE

Throughout this course, you'll work through four section units that culminate in a capstone project – a presentation of a model you build based on a topic of your choosing – which you'll compile over the last two units. Each unit consists of three to four topics, and each topic has theory, demonstration, and independent practice opportunities. By the end of the course, you will have practiced several exercises with your instructor, worked through your own problem statement and model, and have a detailed understanding of the opportunities of data science in your field.

UNIT ACTIVITIES

These activities test your understanding of each unit's most important concepts with in-class independent practice and instructor support. For those of you who want to go above and beyond, we will also offer “stretch” options, enrichment activities, and additional resources with opportunities for further reading and practice.

DATASETS

Datasets used throughout the course include information about sales, product users, order history, etc. These are all datasets we've vetted and used in the past, so they are optimized for teaching students new to Python and data science. Once you begin learning how to model data, you will then have the chance to select your own dataset to begin working with for your capstone project.

CAPSTONE PROJECT

The capstone project is an opportunity for you to apply what you are learning to a problem statement of your creation, using a dataset of your choosing. You will get structured guidance and designated time to work on the capstone project throughout the course. Project deliverables on the last day of the course are:

- Cleaned Jupyter notebook documenting your process for technical peers
- Brief written summary that can be read by a general audience
- Presentation for a mixed audience

COURSE REQUIREMENTS & GRADING

To pass Data Science Fundamentals and receive your post-course “Certificate of Completion” from General Assembly, each participant must meet three essential requirements:

- **Attendance:** participants should miss no more than one class. However, if an absence is unavoidable and known in advance, participants will be expected to proactively communicate this with GA instructors, while also working with them to make up for the instructional time that was missed.
- **Projects/Capstones:** participants must submit all course projects and capstones, with an average score across all project work of 80% or better. GA projects are graded on a 0-2 point scale, where 0 = Does Not Meet Expectations; 1 = Needs Revisions, Must Resubmit; and 2 = Meets Expectations. Participants can also earn a 3rd point on each project by answering a bonus question that your instructors will share.
- **Surveys & Assessments:** all participants must complete an End-Course Survey, which is administered during the last 1-2 classes. In addition, participants must complete a reassessment of the Data Science Skills Assessment, which will be completed on GA’s [myGA learning platform](#).

If all the above requirements are met, participants can expect to receive the aforementioned Certificate of Completion within 1-3 weeks after the course ends.

COURSE SCHEDULE

Below, please find the course schedule for Data Science Fundamentals, including the topics and skills that will be covered during each class session. However, *please note* that specific skills and topics covered are subject to change, based on overall course pacing and other contingencies.

Lesson # & Dates	Topics Covered	Skills Covered
Lesson One (Thurs, 6/20)	<ul style="list-style-type: none">• Welcome to Data Science• Your Development Environment• Python Foundations	<ul style="list-style-type: none">• Set up development environment, define data science workflow and common machine learning concepts, python fundamentals• Using the command line, introduction to Git and GitHub• Python data types, iteration, control flows and functions
Lesson Two (Thurs, 6/27)	<ul style="list-style-type: none">• Exploratory Data Analysis in Pandas• Data Visualization in Python• Review & Recap / Project Workshop	<ul style="list-style-type: none">• Definition and purpose of Pandas in data science, manipulate data in Data Frames and Series, Filter and sort, handling null/missing values, join multiple data sources

		<ul style="list-style-type: none"> • Data visualization overview and best practices, guided practice lab • Designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
Lesson Three (Thurs, 7/11)	<ul style="list-style-type: none"> • Statistics in Python • Experiments & Hypothesis Testing • Review & Recap / Project Workshop 	<ul style="list-style-type: none"> • Descriptive statistics fundamentals using NumPy and Pandas, basic modeling • Causation and correlation, determine causality and sampling bias, identifying and handling missing data, testing a hypothesis • Designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
Lesson Four (Thurs, 7/18)	<ul style="list-style-type: none"> • Linear Regression • Train-Test Split & Bias-Variance • Review & Recap / Project Workshop 	<ul style="list-style-type: none"> • Define data modeling, build a linear regression model • Bias and variance trade-off, train-test-split, k-folds cross-validation • Designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
Lesson Five (Thurs, 7/25)	<ul style="list-style-type: none"> • KNN / Classification • Logistic Regression • Review & Recap / Project Workshop 	<ul style="list-style-type: none"> • K-Nearest Neighbors Classification model, tuning, standardizing features • Using logistic regression for classification, interpret

		<p>logistic regression coefficients, construct confusion matrix</p> <ul style="list-style-type: none"> Designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
<p>Lesson Six (Thurs, 8/1)</p>	<ul style="list-style-type: none"> Working with Data: APIs Intro to Natural Language Processing Review & Recap / Project Workshop 	<ul style="list-style-type: none"> Intro to APIs, pulling data from APIs, building a webscraper Intro to Natural Language Processing, text classification, preprocessing techniques designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
<p>Lesson Seven (Thurs, 8/8)</p>	<ul style="list-style-type: none"> Intro to Time Series Flex: Instructor Choice Review & Recap / Project Workshop 	<ul style="list-style-type: none"> Time series data, create rolling means and plotting time series data, perform autocorrelation, decompose data into trend and residual component. Optional topics to choose from: clustering, decision trees, robust regression, deploying models with flask Designed to provide your class with time to catch up, review materials, ask questions, work on projects, or go deeper into specific topic areas.
<p>Lesson Eight (Thurs, 8/15)</p>	<ul style="list-style-type: none"> Final Project/Capstone Presentations Data Science Re-Assessment 	